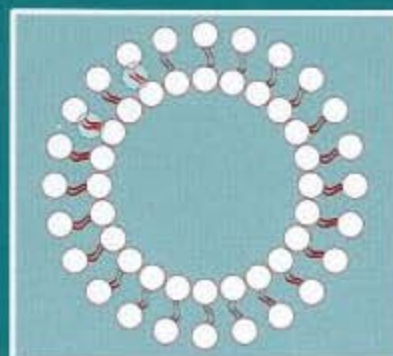
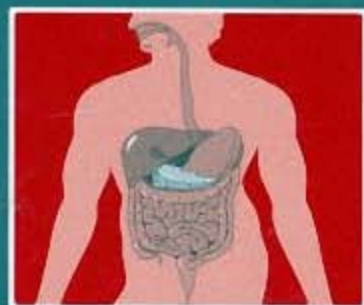
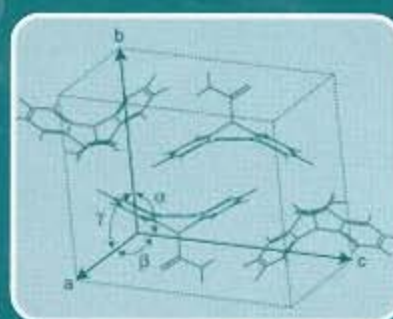


2004 Update Supplement

Encyclopedia of Pharmaceutical Technology

Second Edition



edited by
James Swarbrick

ABSORPTION ENHANCERS

Jean Paul Remon

Ghent University, Ghent, Belgium

INTRODUCTION

There is enormous literature on the use of absorption enhancers. Here, the most important absorption enhancers for topical, transdermal and mucosal drug delivery are reviewed.

TOPICAL AND TRANSDERMAL

It is generally accepted that the bioavailability of most topically applied drugs remains low. Various methods to increase this bioavailability have been used. One of the approaches is the use of absorption enhancers, and over the years, there has been a great interest in new chemical absorption enhancers. An absorption enhancer should be pharmacologically inert, nontoxic, have a rapid and reversible onset of action, be chemically and physically compatible with other formulation compounds, and be cosmetically acceptable (1). Of course not all absorption enhancers possess all of these characteristics, and a benefit-to-risk evaluation will determine the choice of a molecule as an absorption enhancer. The range of absorption enhancers that has been researched is large. Thus, overview of the most researched compounds is presented.

Alcohols and Polyols

Some solvents are able to remove lipids from the stratum corneum, and several topical preparations (e.g., gels) and transdermal reservoir systems contain high concentrations of ethanol capable of modifying the lipid content of the skin (2). Solvents, such as ethanol, but also others such as propylene glycol, *N*-methylpyrrolidone, and Transcutol™, might also increase the drug flux through the skin by increasing the solubility of the permeant in the skin. It has also been suggested that the activity of propylene glycol results from solvation of α -keratin within the stratum corneum, hereby promoting permeation by reducing drug-tissue binding.

Amines and Amides

Some excipients might intercalate into the structure of lipids of the skin and disrupt the ordered packing making so the structure more fluid and influencing positively the diffusion coefficient. Azone® and its analogues have been widely studied in that respect, and it has been shown that the hydrogen bonding between the polar head group in Azone® probably interacts with the skin ceramides (3). Godwin et al. (4) compared the penetration-enhancing ability of a wide range of pyrrolidone compounds, including those with different chain lengths and functional groups. Using hydrocortisone as a model drug, these authors suggested that *N*-dodecyl-2-pyrrolidone and its acetate analogue were the two most effective penetration enhancers using in vitro hairless mouse skin model. Several studies dealt also with the mechanism of action of Azone® and its analogues. Compounds with short alkyl chains, such as *N*-methylpyrrolidone, seemed to have no effect on the phase transition temperature and probably work through its action of solvency rather than through a structural change of the skin barrier function. Using multilamellar DDPE liposomes, Hadgraft et al. (3) showed that their phase transition temperature was lowered by the Azone and its analogues in the same rank order as their enhancing abilities. This indicates that the modifier activity might be related to the fluidising effect on the lipid lamellae.

Studies involving the structure activity relationship of several groups of enhancers showed that the presence of a cyclic structure in the molecule plays an important role in the activity determination of the enhancers. In addition, the greatest barrier disruption activity was recorded for compounds with long alkyl chains between C₈–C₁₆ (5). Unfortunately, these molecules show also irritating potentials (6). Recently, Hadgraft (7) described some new molecules with similar structures but with low irritation potential.

Urea promotes transdermal permeation by facilitating hydration of the stratum corneum and the formation of hydrophilic channels (8).

Fatty Acids

The perturbation of the intercellular lipid bilayers in the stratum corneum seems to be the most important reason

for the enhancing activity of fatty acids such as oleic acid. Oleic acid has been described to decrease the phase transition temperatures of the skin lipids with a resultant increase in motational freedom—or fluidity—of these lipids (9).

Terpenes

Mono- and sesquiterpenes are known to increase percutaneous resorption by increasing the diffusion in the stratum corneum and/or disruption of the intercellular lipid barrier (10, 11). It has been shown that there is a major difference between different types of terpenes: e.g., it was shown that *d*-limonene did not disrupt the intercellular bilayers, whereas 1-8-cineole seemed lipid disruptive at physiological temperatures (12).

Menthol also has been described as a potential penetration enhancer due to its preferential distribution into the intercellular spaces of the stratum corneum and its possible reversible disruption of the intercellular lipid domain (13).

Esters

A typical example of an ester acting as a penetration enhancer is isopropyl myristate. Isopropyl myristate might show a double action: influence on the partition between vehicles and skin by solubilization and disruption of lipid packing, thus increasing the lipid fluidity (14, 15).

Sulfoxides

Dimethylsulfoxide (DMSO) has been found to be a potent enhancer, but unfortunately high concentrations which produce irreversible skin damage, erythema, and wheals, are required to obtain a desired effect. Recently, novel molecules were produced by modifying DMSO, by replacing the oxygen atom with a nitrogen atom that was substituted with an arylsulfonyl, aroyl, or aryl group. The *S*, *S*,-dimethyl-*N*-(4-bromobenzoyl)iminosulfurane produced the highest activity. But these compounds require more activity and toxicity studies, especially in less permeable models such as the human skin (16).

Cyclodextrins

Cyclodextrins can form inclusion compounds with an increase in solubility of lipophilic compounds, but they

seemed less effective alone than in combination with fatty acids and propyleneglycol (17).

Surface Active Agents

The effect of surface active agents on the skin barrier function depends on the agent's chemical structure. In general, anionic surfactants tend to be more effective than cationic ones, whereas nonionic surfactants are considerably less effective. Most anionic surfactants can induce swelling of the stratum corneum, as well as uncoiling and stretching of α -keratin helices, thereby opening up the protein controlled polar pathways (18).

The impact of anionic surfactants is a function of the alkyl chain length of the molecule. A maximum was observed for surfactants having a linear alkyl chain of 12 carbon atoms (e.g., sodium lauryl sulphate). Unfortunately, anionic surfactants are reported to be irritative. Nonionic surfactants might increase the membrane fluidity of the intercellular regions of the stratum corneum (e.g., Brij®) and may extract lipid components and additionally, though of minor importance, they might interfere with keratin filaments and create a disorder within the corneocytes (19). It should be emphasized that surfactant form micelles which, if used above their CMC, might negatively influence the drug bioavailability.

Other Enhancers

Other potential penetration enhancers have also been described, such as *N*-acetylprolineesters (20) and glyceryl monocaprylate/caprate (21).

It should be emphasized that the activity of any enhancer should be evaluated in terms of function of the vehicle used and that the selection of the combination enhancer-vehicle is a function of the final therapeutic objectives.

(TRANS)MUCOSAL PERMEATION ENHANCERS

At all mucosal sites, the coadministration of absorption enhancers is normally necessary to achieve therapeutic relevant plasma levels of (large) hydrophilic molecules such as peptides or proteins. Table 1 gives an overview of the most used and researched absorption enhancers and their possible mechanism of action.

Next, an overview is given of the most used transmucosal drug delivery routes and the use of permeation enhancers in each of them.

Table 1 Most used and researched mucosal permeation enhancers

Type	Examples	Mechanism of action
Synthetic surfactants	Laureth-9 sodium lauryl sulphate polysorbate 20 and 80 PEG-8 laurate sorbitan laurate glyceryl monolaurate saponins (e.g., Quillaja saponins)	membrane interaction extraction of membrane proteins and lipids solubilization of peptides
Bile salts	sodium deoxycholate sodium glycocholate sodium fusidate sodium taurodihydrofusidate	denaturation of proteins decrease of mucus viscosity decrease of peptidase activity solubilization of peptides formation of reversed micelles
Fatty acids and derivatives	oleic acid caprylic acid lauric acid palmitoylcarnitine	fosfolipid acylchain disruption
Chelators	Na ₂ EDTA citric acid salicylates	Ca ²⁺ complexation (influencing tight junctions)
Inclusion complexes	cyclodextrins and derivatives	increasing peptide stability increasing solubility enzyme inhibition
Other agents	Azone [®]	lipid structure disruption

Oral and Buccal Mucosa

There are clear differences between the oral mucosal membrane and other epithelial membranes of the intestine, nasal cavity and rectum. The oral mucosal membranes are less keratinized than the skin membranes and show a more loosely packed intercellular lipid domain. In terms of function of the absorption enhancement through the oral mucosal membrane, it can be said that it occurs principally through the lipid-filled intercellular spaces. One could suggest that the mechanism of increasing lipid fluidity of intercellular lipids, as indicated previously for the skin, should also apply for the oral mucosal membranes (22). As has been reported in the case of skin, other mechanisms can be applied here. For example, sodium deoxycholate appeared to denature and extract proteins from rabbit buccal mucosa and affected membrane lipids and inhibited proteases.

There are only a limited number of studies comparing the systematic changes in the structure of enhancers and

their influence on the oral mucosal membranes. For example, for insulin absorption in rats, it was shown that sodium glycocholate, laureth-9, sodium laurate, and sodium lauryl sulphate were approximately equipotent. Several nonionic surfactants having a C₁₂ hydrophobic tail were much less effective (23, 24).

A study related to the buccal bioavailability of testosterone indicated the absorption enhancing effect of hydroxypropyl-β-cyclodextrine with a relative bioavailability of 165% versus the administration without absorption enhancers. This effect was probably due to an increased solubility of testosterone, although cyclodextrins might also extract lipids from the intercellular matrix (25). In the same study, sodium tauro-24,25-dihydrofusidate and sodium deoxycholate did not show any enhancing properties.

Nasal Mucosa

Many papers have been published on the use and efficacy of absorption enhancers for nasal peptide and protein

delivery. The enhancing effect of bile salt seemed dependent on its lipophilicity: The bioavailability of gentamicin increased with increasing lipophilicity of trihydroxy bile salts (cholate > glycocholate > taurocholate), and the enhancement of nasal insulin bioavailability followed the rank order of deoxycholate, chenodeoxycholate, and cholate. However, most studies reported severe damage of bile salts to the mucosa. Deoxycholate had the most ciliotoxic effect, whereas taurocholate had the least ciliotoxic effect (26). In the case of dihydrofusidates, a dose-dependent increase in bioavailability was reported for peptides such as insulin.

A number of dihydrofusidate derivatives have been synthesized in order to evaluate the structure-enhancement relationship. Acidic derivatives achieved a higher enhancement than basic derivatives, but the safety of dihydrofusidates remains a contradictory issue and some structural damage to the mucosa has been reported (27).

In the past years, much research has concentrated on the use of cyclodextrins to enhance bioavailability of peptides and proteins especially because of their mild and reversible effect on the nasal mucociliary clearance (28).

Among the cyclodextrins, the use of DM β CD was shown to have the highest effect on the transnasal bioavailability of insulin in rats. Several studies reported on their concentration-dependent effect. Besides for peptides, the methylated β -cyclodextrins have shown to be useful in nasal delivery of lipophilic drugs. The toxicological profile of dimethyl β -cyclodextrins and of randomly methylated β -cyclodextrins appeared excellent. Attention should be paid, if possible, on bioavailability differences between animal and human models.

Vaginal Mucosa

Laureth-9, lysophosphatidylcholine and palmitylcarnitine chloride were found to be highly effective absorption enhancers, but all induced epithelial damage (29). Insulin was also administered to ovariectomized rats, and the coadministration of sodium taurodihydrofusidate, laureth-9, lysophosphatidylcholine, and -glycerol significantly increased hypoglycemia. Lysophosphatidylglycerol showed only minor damage of the vaginal epithelium, in contrast to the other absorption enhancers used (30). The combination of lysophosphatidylcholine and starch microspheres showed promising insulin bioavailability results (31).

Deoxycholate and quillajasaponins were reported to have a positive effect on the vaginal absorption of calcitonin (32).

Rectal Mucosa

Due to a combination of poor membrane permeability and metabolism at the site of absorption, rectal bioavailability of peptide and proteins is low. As in other mucosal bioavailability testing, insulin is the most studied polypeptide with respect to rectal absorption.

Sodium salicylate and 5-methoxysalicylate increased the absorption of insulin (33). Sodium glycocholate was more effective than sodium taurocholate but less effective than sodium-deoxycholate and PE-9-lauryl ether in enhancing rectal insulin absorption in rabbits (34). The role of disodium EDTA in the enhancement of rectal drug absorption, along with the damaging effects on the rectal mucosa, has been described for several drugs (35, 36).

Bile salts were also used for the enhancement of drug absorption, but several studies indicated severe damage due to their use in rectal drug delivery (37). Sodium tauro-24, 25-dihydrofusidate (STDHF) had a positive effect on the availability of cefoxitin, vasopressine, and insulin in rats (38).

The possible use of mixed micelles (e.g., made of unsaturated fatty acids and monoglycerides) has been shown for the enhanced rectal absorption of several compounds, including α and β interferon and insulin.

Pulmonary Absorption Enhancers

Only a few studies are available related to the effect of known absorption enhancers on the pulmonary absorption of poorly absorbable drugs, including peptides and proteins.

It was reported that oleic acid, oleyl alcohol, and Span 85 can increase the transfer rate of disodium fluorescein in isolated rat lungs (39). Pulmonary insulin absorption was reported to be increased in the presence of glycocholate and Span 85 (40). Fluorescein isothiocyanate, insulin, and a calcitonin analogue were better absorbed when coadministered with *n*-lauryl β -D-maltopyranoside, sodium glycocholate, and linoleic acid mixed micelles (41). The same authors, however, reported on the toxicity of *n*-lauryl β -D-maltopyranoside. A large number of questions are still remaining, such as why sodium caprate enhances the bioavailability of phenol red and isothiocyanate-labeled dextrans but not of insulin and a calcitonin analogue.

Hydroxypropyl- β -cyclodextrin and especially dimethyl- β -cyclodextrin have been shown to enhance the pulmonary bioavailability of insulin in rats, and indications were found of a relatively low acute mucotoxicity (42).

Table 2 Some of the commonly used intestinal absorption enhancers

Bile salts	Sodium cholate, sodium deoxycholate
Nonionic surfactants	Polysorbates and polyoxyethylene alkyl esters and ethers
Ionic surfactants	Sodium lauryl sulphate and dioctyl sulfosuccinate
Fatty acids	Sodium caprate, oleic acid
Glycerides	Medium-chain glycerides, phospholipids
Acyl carnitines	Palmitoylcarnitine
Chelating agents	EDTA
Swellable polymers	Polycarbophil and chitosan

Intestinal Absorption Enhancers

The optimization of oral bioavailability is of common interest because low bioavailability is often the cause of variable and poorly controlled clinical and toxic effects. This is of major importance for polar molecules such as peptides and proteins (43).

Table 2 reviews the most commonly described compounds to enhance intestinal absorption and indicates some examples (44).

It should be emphasized that absorption enhancers might act selectively on some parts of the GI tract, and this fact implicates that the formulation will play a major role in the optimal delivery of drug and absorption enhancers.

Bile salts have proven to act very differently on the intestinal absorption of drugs. In some cases, the drug absorption was reduced due to micelle formation, whereas in other cases, the absorption was enhanced due to intestinal membrane disruption caused by the solubilization of phospholipids or by Ca^{2+} complexation (45–47).

Nonionic and anionic surfactants haven been shown to be able to enhance the intestinal absorption of drugs. Some studies have shown that in the area of nonionic surfactants ethers were more effective than esters, but this phenomenon was not always confirmed. There is an indication that the surfactants cause membrane damage, which can be correlated with their enhancement activity (48). It has been shown that some tensioactive agents might influence tight junction permeability (49).

Fatty Acids

Fatty acids increase intestinal absorption via their influence on the paracellular and transcellular transport route. Most interesting results were obtained with lauric

acid, palmitic acid, caprylic acid, and oleic acid or their salts. Cytotoxic effects of fatty acids are concentration dependent long-chain unsaturated fatty acids especially can cause epithelial cell damage (50–52).

Glycerides

Medium-chain glycerides (mainly C_8 – C_{10}) are known to increase the intestinal absorption of poorly permeable drugs, mono- and diglycerides, especially, improve bioavailability, and it is believed that mainly transcellular permeation is increased.

It should be emphasized that the formulation plays an important role in the effect of these glycerides (emulsification, enteric coating, etc.) The main advantage of these products is their general acceptance for use in oral drug administration (53, 54).

Finally, it should be noted that during the last decade both weakly crosslinked poly(acrylic acid) derivatives and chitosan derivatives were described as safe penetration enhancers for hydrophilic compounds especially as they can trigger mechanisms of tight junction opening of mucosal tissues and did not show acute toxicity. Poly(acrylic acid) derivatives were shown to have excellent mucoadhesive properties and can inhibit the activity of gut enzymes, such as trypsin, chymotrypsin, and carboxypepsidases (55, 56). Chitosan salts and N_1 -trimethylchitosan chloride revealed to be potential absorption enhancers for nasal absorption of calcitonin and insulin and for the intestinal absorption of buserilin (57).

REFERENCES

1. Chattaraj, S.C.; Walker, R.B. Penetration Enhancer Classification. *Percutaneous Absorption Enhancers*; Smith, E.W., Maibach, H.I., Eds.; CRC Press: Boca Raton: Florida, 1995; 5–20.
2. Bommannan, D.; Potts, R.O.; Guy, R.H. Examination of the Effect of Ethanol on Human Stratum-corneum In Vivo Using Infrared-spectroscopy. *J. Controlled Release* **1991**, *16*, 299–304.
3. Hadgraft, J.; Peck, J.; Williams, D.G.; Pugh, W.J.; Allan, G. Mechanisms of Action of Skin Penetration Enhancer Retarders : Azone and Analogues. *Int. J. Pharm.* **1996**, *141*, 17–25.
4. Godwin, D.A.; Michniak, B.B.; Player, M.R.; Sowell, J.W., Sr. Transdermal and Dermal Enhancing Activity of Pyrrolidinones in Hairless Mouse Skin. *Int. J. Pharm.* **1997**, *155*, 241–250.
5. Bouwstra, J.A.; Boddé, H.E. Human Stratum Corneum Barrier Impairment By *N*-alkylazacycloheptanones: A Mechanistic Study of Drug Flux Enhancement, Azone® Mobility and Protein and Lipid Perturbation. *Percutaneous*

- Penetration Enhancers*; Smith, E.W., Maibach, H.I., Eds.; CRC Press: Boca Raton Florida, 1995; 137–169.
6. Bouwstra, J.A.; Peschier, L.C.; Brussee, J.; Boddé, H.E. Effect Of *N*-alkyl-azocycloheptan-2-ones Including Azone® On the Thermal Behavior of Human Stratum Corneum. *Int. J. Pharm.* **1989**, *52*, 47–54.
 7. Hadgraft, J. Passive Enhancement Strategies in Topical and Transdermal Drug Delivery. *Int. J. Pharm.* **1999**, *184*, 1–6.
 8. Kim, C.K.; Kim, J.J.; Chi, S.C.; Shim, C.K. Effect of Fatty Acids and Urea on the Penetration of Ketoprofen Through Rat Skin. *Int. J. Pharm.* **1993**, *99*, 109–118.
 9. Golden, G.M.; McKie, J.E.; Potts, R.O. Role of S.C. Lipid Fluidity in Transdermal Flux. *J. Pharm. Sci.* **1987**, *76*, 25–28.
 10. William, A.C.; Barry, B.W. Terpenes and the Lipid-protein Partitioning Theory of Skin Penetration Enhancement. *Pharm. Res.* **1991**, *8*, 17–24.
 11. Cornwell, P.A.; Barry, B.W. The Routes of Penetration of Ions and 5-fluorouracil Across Human Skin and the Mechanism of Action of Terpene Skin Penetration Enhancers. *Int. J. Pharm.* **1993**, *94*, 189–194.
 12. Cornwell, P.A.; Barry, B.W.; Bouwstra, J.A.; Gooris, G.S. Modes of Action of Terpene Penetration Enhancers in Human Skin : Differential Scanning Calorimetry, Small-angle X-ray Diffraction and Enhancer Uptake Studies. *Int. J. Pharm.* **1996**, *127*, 9–26.
 13. Kunta, J.R.; Goskonda, V.R.; Brotherton, H.O.; Khan, M.A.; Reddy, I.K. Effect of Menthol and Related Terpenes on the Percutaneous Absorption of Propranolol Across Excised Hairless Mouse Skin. *J. Pharm. Sci.* **1997**, *86*(12), 1369–1372.
 14. Friend, D.; Catz, P.; Heller, J.; Reid, J.; Baker, R. Transdermal Delivery of Levenorgestol 3. Simple Alkyl Esters as Skin Permeation Enhancers. *J. Control. Rel.* **1989**, *9*, 33–41.
 15. Sato, K.; Sugibayashi, K.; Morimoto, Y. Effect and Mode of Action of Aliphatic Esters on In Vitro Skin Permeation of Nicorandil. *Int. J. Pharm.* **1988**, *43*, 31–40.
 16. Kim, N.; El-Khalili, M.; Henary, M.M.; Strekowski, G.; Michniak, B.B. Percutaneous Penetration Enhancement Activity of Aromatic S, S,-dimethyliminosulfuranes. *Int. J. Pharm.* **1999**, *187*, 219–229.
 17. Vollmer, U.; Muller, B.W.; Mesens, J.; Willfert, B.; Peters, T. In Vitro Skin Pharmacokinetics of Liarozole: Percutaneous Absorption Studies with Different Formulations of Cyclodextrin Derivatives in Rats. *Int. J. Pharm.* **1993**, *99*, 51–58.
 18. Rhein, L.D.; Robbins, C.R.; Fernee, K.; Cantore, R. Surfactant Structure Effects on Swelling of Isolated Human S.c. *J. Soc. Cosmet.* **1986**, *37*, 125–139.
 19. Walters, K.A.; Walker, M.; Olejnik, O. Non Ionic Surfactant Effects on Hairless Mouse Skin Permeability Characteristics. *J. Pharm. Pharmacol.* **1988**, *40*, 525–529.
 20. Harris, W.T.; Tenjarla, S.N.; Holbrook, J.M.; Smith, J.; Mead, C.; Entekin, J. *N*-pentyl *N*-acetyl Proline. A New Skin Penetration Enhancer. *J. Pharm. Sci.* **1995**, *84*, 640–642.
 21. Cornwell, P.A.; Tubek, J.; Van Gompel, H.A.H.P.; Little, C.J.; Wiechers, J.W. Glyceryl Monocaprylate/Caprate as a Moderate Skin Penetration Enhancer. *Int. J. Pharm.* **1998**, *171*, 243–255.
 22. Wertz, P.W.; Squier, C.A. Cellular and Molecular Basis of Barrier Function in Oral Epithelium. *Crit. Rev. Ther. Drug Carr. Systems* **1991**, *8*, 237–269.
 23. Aungst, B.J.; Rogers, N.J. Comparison of the Effect of Various Transmucosal Absorption Promoters on Buccal Insulin Delivery. *Int. J. Pharm.* **1989**, *53*, 227–235.
 24. Aungst, B.J. Site Dependence and Structure-effect Relationships for Alkylglycosides as Transmucosal Absorption Promoters for Insulin. *Int. J. Pharm.* **1994**, *105*, 219–225.
 25. Voorspoels, J.; Remon, J.P.; Eechaute, W.; De Sy, W. Buccal Absorption of Testosterone and its Esters Using a Bioadhesive Tablet in Dogs. *Pharm. Res.* **1996**, *13*, 1228–1232.
 26. Ennis, R.D.; Borden, L.; Lee, W.A. The Effects of Permeation Enhancers in the Surface Morphology of the Rat Nasal Mucosa: A Scanning Electron Microscopy Study. *Pharm. Res.* **1990**, *7*, 468–475.
 27. Kissel, T.; Drewe, J.; Bantle, S.; Rummelt, A.; Beglinger, C. Tolerability and Absorption Enhancement of Intranasally Administered Octreotide Subject. *Pharm. Res.* **1992**, *9*, 52–57.
 28. Merkus, F.W.H.M.; Verhoef, J.C.; Martinn, E.; Romeijn, S.G.; van der Kuy, P.H.M.; Hermens, W.A.J.J.; Schipper, N.G.M. Cyclodextrins in Nasal Drug Delivery. *Adv. Drug Delivery Rev.* **1999**, *36*, 41–57.
 29. Richardson, J.L.; Minhas, P.S.; Thomas, N.W.; Illum, L. Vaginal Administration of Gentamycin to Rats. *Pharmaceutical and Morphological Studies Using Absorption Enhancers*. *Int. J. Pharm.* **1989**, *56*, 29–35.
 30. Richardson, J.L.; Illum, L.; Thomas, N.W. Vaginal Absorption of Insulin in the Rat—Effect of Penetration Enhancers on Insulin Uptake and Mucosal Histology. *Pharm. Res.* **1992**, *9*(7), 878–883.
 31. Richardson, J.L.; Farraj, N.F.; Illum, L. Enhanced Vaginal Absorption of Insulin in Sheep Using Lysophosphatidylcholine and a Bioadhesive Microsphere Delivery System. *Int. J. Pharm.* **1992**, *88*(1–3), 319–325.
 32. Nakada, Y.; Miyake, M.; Awata, N. Some Factors Affecting the Vaginal Absorption of Human Calcitonin in Rats. *Int. J. Pharm.* **1993**, *89*, 169–175.
 33. Nishihata, T.; Rytting, J.H.; Kamada, A.; Higuchi, T.; Routh, M.; Caldwell, L. Enhancement of Rectal Absorption of Insulin Using Salicylates in Dogs. *J. Pharm. Pharmacol.* **1983**, *35*, 148–151.
 34. Yamamoto, A.; Hayakawa, E.; Kato, Y.; Nishiura, A.; Lee, V.H.L. A Mechanistic Study on Enhancement of Rectal Permeability to Insulin in Albino Rabbits. *J. Pharmacol. Exp. Ther.* **1992**, *263*, 25–31.
 35. Nishihata, T.; Tomida, H.; Frederick, G.; Rytting, J.H.; Higuchi, T. Comparison of the Effects of Sodium Salicylate, Disodium EDTA, and Polyoxyethylene 23-lauryl Ether as Adjuvants for the Rectal Absorption of Sodium Cefoxitin. *J. Pharm. Pharmacol.* **1985**, *37*, 159–163.
 36. Nishihata, T.; Lee, C.S.; Rytting, J.H.; Higuchi, T. The Synergistic Effects of Concurrent Administration to Rats of EDTA and Sodium Salicylate on the Rectal Absorption of Sodium Cefoxitin and the Effect of Inhibitors. *J. Pharm. Pharmacol.* **1987**, *39*, 180–184.
 37. Murakami, T.; Sasaki, Y.; Yamajo, R.; Yata, N. Effects of Bile Salts on the Rectal Absorption of Sodium Ampicillin in Rats. *Chem. Pharm. Bull.* **1984**, *32*, 1948–1955.

38. Van Hoogdalem, E.; Heijligers-Feijen, C.D.; Mathot, R.A.A.; Wackwitz, A.T.E.; Van Bree, J.B.M.M.; Verhoef, J.C.; de Boer, A.G.; Breimer, D.D. Rectal Absorption Enhancement of Cefoxitin and Desglycinamide Arginine Vasopressin by Sodium Tauro-24, 25-dihydrofusidate in Conscious Rats. *J. Pharmacol. Exp. Ther.* **1989**, *251*, 741–744.
39. Niven, R.W.; Byron, P.R. Solute Absorption from the Airways of the Isolated Rat Lung. 2. Effect of Surfactants on Absorption of Fluorescein. *Pharm. Res.* **1990**, *7*, 8–13.
40. Okumura, K.; Iwakawa, S.; Yoshida, T.; Seki, T.; Komada, F. Intratracheal Delivery of Insulin: Absorption from Solution and Aerosol by Rat Lung. *Int. J. Pharm.* **1992**, *88*, 63–73.
41. Yamamoto, A.; Fujita, T.; Muranishi, S. Pulmonary Absorption Enhancement of Peptides by Absorption Enhancers and Protease Inhibitors. *J. Controlled Release* **1996**, *41*, 57–67.
42. Shao, Z.Z.; Li, Y.P.; Mitra, A.K. Cyclodextrins as Mucosal Absorption Promoters of Insulin. 3. Pulmonary Route of Delivery. *Eur. J. Pharm. Biopharm.* **1994**, *40*, 283–288.
43. Swenson, E.S.; Curatolo, W.J. Intestinal Permeability Enhancement for Proteins, Peptides and Other Polar Drugs — Mechanisms and Potential Toxicity. *Adv. Drug Delivery Rev.* **1992**, *8*, 39–92.
44. Aungst, B.J.; Saitoh, H.; Burcham, D.L.; Huang, S.M.; Mousa, S.A.; Hussain, M.A. Enhancement of the Intestinal Absorption of Peptides and Nonpeptides. *J. Controlled Release* **1996**, *41*, 19–31.
45. Kakemi, K.; Sezaki, H.; Konishi, R.; Kimura, T.; Murakami, M. Effect of Bile Salts on the Gastrointestinal Absorption of Drugs 1. *Chem. Pharm. Bull.* **1970**, *18*, 275–280.
46. Kidron, M.; Baron, H.; Berry, E.M.; Ziv, E. The Absorption of Insulin from Various Regions of the Rat Intestine. *Life Sci.* **1982**, *31*, 2837–2841.
47. Feldman, S.; Gibaldi, M. Physiological Surface-active Agent and Drug Absorption 1. Effect of Sodium Taurodeoxycholate on Salicylate Transfer Across the Everted Rat Intestine. *J. Pharm. Sci.* **1969**, *58*, 425–428.
48. Swenson, E.S.; Milisen, W.B.; Curatolo, W. Intestinal Permeability Enhancement: Efficacy, Acute Local Toxicity and Reversibility. *Pharm. Res.* **1994**, *11*, 1132–1142.
49. Anderberg, E.K.; Artursson, P. Epithelial Transport of Drugs in Cell Culture 8. Effects of Sodium Dodecylsulfate on Cell Membrane and Tight Junction Permeability in Human Intestinal Epithelial (Caco-2) Cells. *J. Pharm. Sci.* **1993**, *82*, 392–398.
50. Muranushi, N.; Mack, E.; Kim, S.W. The Effects of Fatty Acids and Their Derivatives on the Intestinal Absorption of Insulin in Rats. *Drug Dev. Ind. Pharm.* **1993**, *19*, 929–941.
51. Anderberg, E.K.; Lindmark, T.; Artursson, P. Sodium Caprate Elicits Dilatations in Human Intestinal Tight Junctions and Enhances Drug Absorption by the Paracellular Route. *Pharm. Res.* **1993**, *10*, 857–864.
52. Sawada, T.; Ogawa, T.; Tomita, M.; Hayashi, M.; Awazu, S. Role of Paracellular Pathway in Nonelectrolyte Permeations Across Rat Colon Epithelium Enhanced by Sodium Caprate and Sodium Caprylate. *Pharm. Res.* **1991**, *8*, 1365–1371.
53. Unowsky, J.; Behl, C.R.; Beskid, G.; Sattler, J.; Halpern, J.; Cleeland, R. Effect of Medium Chain Glycerides on Enteral and Rectal Absorption of Beta-lactam and Aminoglycoside Antibiotics. *Chemotherapy* **1988**, *34*, 272–276.
54. Constantinides, P.P.; Scalart, J.P.; Lancaster, C.; Mareello, J.; Marks, G.; Ellens, H.; Smith, P.L. Formulation and Intestinal Absorption Enhancement Evaluation of Water-in-oil Microemulsions Incorporating Medium-chain Glycerides. *Pharm. Res.* **1994**, *11*, 1385–1390.
55. Luessen, H.L.; Lehr, C.M.; Rentel, C.O.; Noach, A.B.; Junginger, H.E.; de Boer, J.C.; Verhoef, H.E. Bioadhesive Polymers for the Peroral Delivery of Peptide Drugs. *J. Controlled Release* **1994**, *29*, 329–338.
56. Luessen, H.L.; de Leeuw, B.J.; Langemeyer, M.W.; de Boer, A.G.; Verhoef, J.C.; Junginger, H.E. Mucoadhesive Polymers in Peroral Peptide Drug Delivery. Part 6. Carbomer and Chitosan Improve the Intestinal Absorption of the Peptide Drug Buserelin In Vivo. *Pharm. Res.* **1996**, *13*, 1668–1672.
57. Kotzé, A.F.; Luessen, H.L.; de Leeuw, B.J.; de Boer, A.G.; Junginger, H.E.; Verhoef, J.C. Comparison of the Effect of Different Chitosan Salts and *N*-trimethyl Chitosan Chloride on the Permeability of Intestinal Epithelial Cells (Caco-2). *J. Controlled Release* **1998**, *51*, 35–46.

ABSORPTION OF DRUGS

Peter G. Welling

University of Strathclyde, Glasgow, Scotland

INTRODUCTION

In the first edition of this encyclopedia the section on absorption covered a range of topics that included discussion of the cell membrane, parenteral and enteral absorption, clinical factors, and pharmacokinetic characterization of absorption.

During the intervening time many of these areas have changed, some more than others. There also have been changes in emphasis, reflected particularly in the rapidly expanding interest and discoveries in membrane penetration, exploitation of various dosage routes, formulation factors, and absorption enhancers.

Emphasis in this section will reflect these activities. In order to conserve space, the pharmacokinetic treatment of drug absorption has been omitted.

OBJECTIVES OF DRUG ABSORPTION

Absorption may be defined as the process by which a compound penetrates one or more biological membranes to gain entry into the body. Absorption is not to be confused with bioavailability, which describes entry of administered compounds into the systemic circulation. For some drugs and dosage routes, absorption and bioavailability may be identical, i.e., after intravenous (IV) dosing. However, in many cases they are not. For a drug that does not undergo any metabolic transformation between an immediate postabsorption site and entry into the systemic circulation, absorption and bioavailability are likely to be the same. All of the absorbed drug enters the systemic circulation. This is regardless of any drug that may be degraded or changed in some other way, i.e., preabsorption.

On the other hand, for any drug that is degraded at a point between the postabsorption site and entry into the systemic circulation, the systemic availability—bioavailability—will be less than the absorption. An orally administered drug that undergoes extensive first-pass hepatic clearance may give rise to poor oral bioavailability despite being efficiently absorbed from

the gastrointestinal (GI) tract into the splanchnic circulation.

The pharmacologic activity profile of a systemically active drug is a function of its intrinsic activity and of the concentration profile that is achieved in the circulation. The speed of onset of action and the intensity and duration of activity are functions of the drug concentration profile.

The speed of onset of drug action is determined by the rate of drug absorption. Extreme cases are the use of bolus IV injection, which yields immediate and usually maximal pharmacologic effect, and slow controlled release, not necessarily by the oral route, where the onset of action is deliberately prolonged to achieve a desired therapeutic profile.

The intensity of pharmacologic effect is generally a function of the concentration of drug achieved in the circulation. Actual pharmacokinetic/pharmacodynamic relationships are often complex, but it is reasonable to generalize that higher circulating drug concentrations yield greater effect.

The levels of circulating drug that are achieved are a function of dose, absorption efficiency, overall bioavailability, distribution, and also clearance. The major determinant of drug distribution volume is its lipophilicity. As lipophilicity increases, so does the ability of the drug to cross biological membranes and move into extravascular environments, particularly into fatty tissue and the central nervous system (CNS).

Many drugs bind to plasma proteins, in particular to plasma albumin. Although binding of drugs to plasma proteins is dynamic and reversible, any drug that is bound at a particular time is necessarily confined to the plasma volume and thus cannot participate in extravascular distribution.

The last factor affecting circulating drug levels is clearance. The faster a drug is cleared from the circulation as a result of metabolism or any other process (i.e., the shorter its elimination half-life) the lower are its circulating levels. High circulating levels are less likely to be achieved with a high clearance drug than with a low clearance drug, and accumulation of a high clearance drug in the circulation with repeated dosing is unlikely.

Thus, the phenomenon of drug absorption is only one, albeit an important one, of several factors that determine

a drug profile in the circulation. It is important to understand all of these factors before drug profiles, and in particular pharmacokinetic/pharmacodynamic relationships, can be fully characterized, particularly in a predictive sense.

All of the above factors are functions of the physical and chemical properties of a drug. While distribution and clearance are affected only by drug properties and cannot generally be altered except by introducing some kind of interaction, drug absorption and bioavailability are often markedly influenced by route of administration, dosage form, and coadministration of other substances. Some of the major thrusts of pharmaceutical research during the last two decades have been devoted to these latter issues.

DOSAGE ROUTES

A required drug absorption profile is achieved by a variety of dosage routes. These routes may be divided into parenteral and enteral. Due to the importance of the various routes of administration in drug delivery, and of recent advances in optimizing route-dependent drug delivery, they are briefly reviewed here.

PARENTERAL ROUTES

Parenteral delivery routes are those that do not give rise to drug absorption into the splanchnic circulation. Thus, they avoid the possibility of hepatic first-pass metabolism. It should be noted that some parenteral routes do not avoid other first-pass metabolism effects (e.g., pleural metabolism for some inhaled drugs). Some major parenteral drug delivery routes are intra-arterial, intrathecal, intravenous, intramuscular, transdermal, intranasal, buccal, inhalation, intraperitoneal, vaginal, and rectal.

Intra-arterial

Intra-arterial injection is used to deliver drugs directly to organs, for example, in cancer chemotherapy, and in the use of vasopressin for GI bleeding. Intra-arterial carmustine is effective to treat brain tumors (1) and pelvic intra-arterial actinomycin D is used for malignant trophoblastic disease (2).

Intra-arterial drug administration has potential safety implications. Embolization, arterial occlusion, and localized drug toxicity have been reported.

Intrathecal

Injection directly into the cerebrospinal fluid (CSF) ensures complete CNS bioavailability for drugs that cannot cross the blood-brain barrier. This dosage route is used to treat serious CNS infections such as meningitis and ventriculitis, and with such agents as mepivacaine and prilocaine for spinal anesthesia. Drugs injected intrathecally initially distribute into approximately 140 ml of CSF, thus producing high concentrations in the CNS with low risk of systemic toxicity.

Intravenous (IV)

IV administration introduces drug directly into the venous circulation. The shape of the resulting circulating drug profile is determined by the size, rate, and duration of injection. IV bolus is used for immediate therapeutic effect, typically for general anesthesia and for treatment of cardiac arrhythmia. IV dosing is popular for preclinical testing of compounds during drug development and also as a standard to determine absolute bioavailability from other dosage routes.

Intramuscular (IM)

Following intramuscular (IM) administration, drugs must cross one or more biological membranes in order to enter the systemic circulation. Intramuscular injection is used mainly for drugs and vaccines that are not absorbed orally, for example, aminoglycosides, insulin, and hepatitis vaccine. The IM route is often used for sustained medication and specialized vehicles, such as aqueous suspensions, oily vehicles, complexes and microencapsulation, which has been developed for slow delivery of drugs by this route (3).

Transdermal

Since the introduction of transdermal scopolamine (4), many transdermal delivery systems have been developed for systemic activity. Major advantages claimed for this drug delivery route include continuous release of drug over a specified period, low presystemic clearance, facile drug withdrawal by simply removing the device, and good patient convenience and compliance.

Some disadvantages relate to barrier properties of the skin, skin reactions, and the relatively large dose size. Transdermal delivery is a realistic option only for drugs generally given in small doses (<10 mg) and which have good membrane penetration. Drugs currently approved for transdermal delivery include clonidine, estradiol, nicotine, nitroglycerin, and scopolamine.

Intranasal

Intranasal administration may be used for local or systemic effects. Local effects include treatment of nasal allergies, rhinitis, and nasal congestion. Nasal delivery for systemic effects is established for a small number of drugs and is being examined for many others.

The sophisticated structure and specialized function of airways and membranes in the nasal cavity, and also the small surface area of this region, may limit its capacity for drug delivery. The effect of chronic drug exposure on the integrity of nasal membranes must also be considered. This problem may be compounded by the evident need for surfactants to achieve good systemic penetration with this dosage route.

Notwithstanding these factors, the physical characteristics of compounds for optimal intranasal absorption are the same as for other absorption routes. The drug must dissolve in the fluids of the nasal mucosa and must be sufficiently lipophilic to cross the membranes of the nasal epithelium. Nasal absorption is facilitated by the high permeability of small venules and capillaries associated with the nasal mucosa.

A variety of delivery systems have been described for nasal drug delivery, including drops, aerosols, nebulizers, and soluble matrices (5).

Small peptide molecules seem to be ideally suited for intranasal drug delivery. Vasopressin analogues and oxytocin are commercially available for intranasal dosage. Thyrotropin-releasing hormone agonists and antagonists, other vasopressin analogs, and peptides are being examined. Intranasal delivery of sex hormones has produced interesting results in animals. Intranasal delivery of insulin has been examined, but only with moderate success.

Buccal

Early recognition of buccal and sublingual absorption was manifested in the use of nitroglycerin by these dosage routes to treat severe headache and to relieve angina pectoris.

Drugs can be absorbed from the oral cavity itself or sublingually. Absorption from either route is rapid,

sublingual more so apparently because of greater permeability of sublingual membranes and rich blood supply. The mean pH of saliva is approximately 6 so that drug absorption, predominantly passive in nature, is favored for unchanged molecules, acids with pK_a values >3, and bases with pK_a values <9.

Compounds that are currently marketed or are being considered for buccal or sublingual routes include organic nitrates, barbiturates, papaverine, trypsin, prochlorperazine, benzodiazepines, buprenorphine, captopril, isoprenaline, oxytocin, and nifedipine. Oxytocin is currently the only peptide marketed in sublingual form. Sublingual steroids have been examined with moderate success.

Inhalation

When a substance is inhaled, it is exposed to membranes of the mouth or nose, pharynx, trachea, bronchi, bronchioles, alveolar sacs, and alveoli (Fig. 1). The lung has a potential absorption surface of some 70 m², a much larger surface than the small intestine. However, the lungs and their associated airways are designed to deny access of administered compounds to the highly absorptive peripheral lung surfaces. The system is designed to deny access to particulate matter. However, if compounds can reach the peripheral region of the lung, absorption can be very efficient.

Particle (droplet) size and velocity of application control the extent to which inhaled substances penetrate into airway spaces. Optimum size for deep airway penetration is 3–5 μ M. Large particles tend to deposit in upper airways.

Most inhalation devices deliver approximately 10% of the administered dose to the lower respiratory tract. A number of devices have been developed to increase lung delivery, and delivery of up to 21% has been

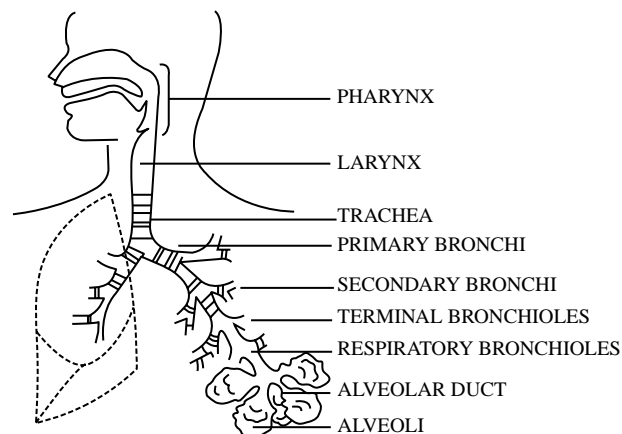


Fig. 1 The human respiratory tract.

reported with a pressurized metered-dose inhaler (6). Despite these advances, drug delivery via the lung is still inefficient.

Systemic availability of inhaled drugs may be inhibited by first-pass pulmonary metabolism. The lungs contain many drug metabolizing enzymes, including mixed function oxidases, monoamine oxidase, and esterases.

Several animal models principally the rat, rabbit, and dog, are used to study drug inhalation.

Intraperitoneal

Intraperitoneal drug administration is not common. It is used predominantly to administer compounds during preclinical discovery and development. Its clinical use is generally limited to chemotherapy for tumors with peritoneal involvement.

Peritonitis occurs frequently in renally impaired patients who are receiving continuous ambulatory peritoneal dialysis (CAPD). Peritonitis is often accompanied by systemic infection so that therapeutic levels of antibiotic are needed both in the peritoneal cavity and in the systemic circulation.

Drugs may be given orally or by injection in order to achieve adequate systemic levels in the hope of also achieving therapeutic levels in the peritoneal cavity. Alternatively, drugs may be administered directly into the peritoneal cavity with the objective of also achieving systemic levels by intraperitoneal absorption (7). To the author's knowledge, there is little definitive information on the relative merits of these alternatives.

Vaginal

The human vagina, a fibromuscular tube 10–15 cm long, extends upwards and backwards from the vulva to the lower uterine cervix. Blood is supplied to the vagina via the uterine and pudendal arteries, and is drained from the vagina by a rich plexus, which flows into the internal iliac veins. The surface of the vaginal epithelium is kept moist by cervical secretions. The pH of vaginal fluid is 4–5.

Vaginal drug delivery is used mostly for local effects, but vaginal absorption can give rise to rapid and efficient systemic delivery. Good systemic absorption, and also the ability of the vagina to retain delivery devices, has given rise to many vaginal dosage forms, in particular for steroid contraceptives. A large number of vaginal controlled release dosage forms are available, including vaginal rings and biodegradable microspheres.

Rectal

The human rectum is 15–20 cm long. It is normally empty and contains 2–3 ml of mucous fluid with pH 7–8. There are no villi and only a limited surface area of 200–400 cm² is available for absorption.

Blood and lymph vessels are abundant in the rectal submucosa. Veins from the upper rectum drain into the portal circulation, while veins from the middle and lower rectum drain directly into the inferior vena cava. However, there are extensive anastomoses among these veins so that precise anatomical differentiation is difficult. It appears that compounds absorbed from the lower rectum, in contrast to those absorbed from the upper rectum, avoid hepatic first-pass metabolism.

Rectal absorption is generally slower than oral absorption, but for some drugs, rectal absorption exceeds oral absorption presumably due to avoidance of first-pass metabolism after rectal delivery. This has been reported for morphine, metoclopramide, ergotamine, lidocaine, and propranolol. Human rectal systemic availability of the extensively metabolized drug lidocaine is 65% as compared to 30% after oral administration (8).

Rectal absorption of drugs from aqueous or alcoholic solutions is generally much faster than from suppositories. Nonsurfactant adjuvants, such as salicylates, increase rectal absorption of water-soluble drugs and also of high molecular weight compounds, such as insulin, heparin, and gastrin.

ENTERAL ROUTES

Enteral routes of drug absorption are from the stomach and the small and large intestine. Substances absorbed from these areas enter the splanchnic circulation and pass through the liver before entering the systemic circulation.

The GI tract is the site of absorption for most nutrients. Thus, the GI tract has evolved to facilitate absorption of substances. The peristaltic action of the stomach, secretion of enzymes and hydrochloric acid, the villi and microvilli of the intestine, as well as the rich blood supply and lymphatics in this region, all facilitate absorption. Enteral absorption is generally by far the most effective drug delivery route and, whenever possible, drugs are administered in this way.

Any orally absorbed compound is exposed to an absorption environment that is both friendly and hostile, depending on the compound and the patient. A brief review of the structure and physiology of the GI tract relative to drug absorption follows.

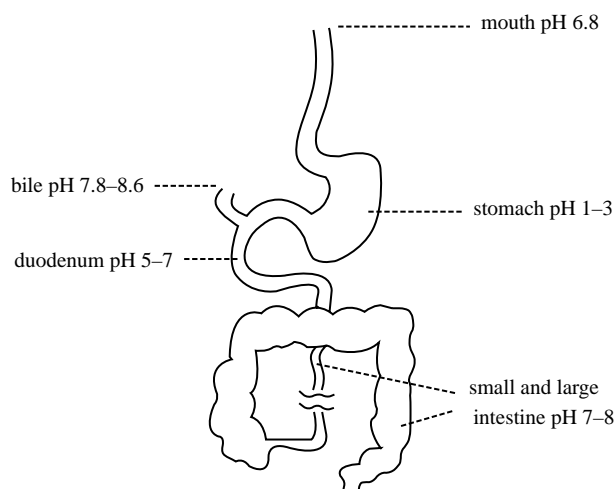


Fig. 2 Approximate pH values in the human gastrointestinal tract.

Physiology of the GI Tract

Figure 2 shows the pH of various regions of the GI tract. From the slightly acidic region of the mouth, a compound enters the more acidic region of the stomach. Acidity is a consequence of hydrochloric acid secretion by the parietal cells of the stomach. This plays an important role in food digestion by facilitating conversion of pepsinogens and zymogens to active proteolytic enzymes.

The acidic environment in the stomach tends to favor gastric absorption of acidic drugs provided the drugs are in solution. On the other hand, basic drugs tend to dissolve readily in the stomach but absorption may be prevented because the drug will be ionized and therefore not sufficiently fat soluble for efficient membrane penetration. The acidic environment in the stomach may give rise to reduced drug absorption due to acid-catalyzed degradation.

If a drug dissolves in the stomach or is a liquid, and if it is fat-soluble and acid stable, then it is likely to be absorbed efficiently from the stomach. Ethyl alcohol is a liquid that is completely miscible with water and sufficiently lipophilic to cross biological membranes. It is efficiently absorbed from the stomach.

After passing through the pyloric sphincter, a compound reaches the duodenum, jejunum, and ileum. These regions of the small intestine differ from the stomach with respect to pH, the presence of digestive enzymes, and the absorptive surface area. Excretion of alkaline bile into the duodenum raises the pH of the duodenal and more distal intestinal contents to 5–7. The change in pH from acidic to essentially neutral causes many changes that affect drug absorption. Enteric coatings that were impermeable in the

stomach will dissolve. Acidic drugs will dissolve more rapidly and yet the pH will not be sufficiently high to prevent dissolution or cause precipitation of weakly basic drugs.

GI Structure and Motility

The stomach is a pouch-like organ lined with a relatively smooth epithelial surface. Although compounds can be absorbed from the stomach, the contribution of this organ to overall enteral drug absorption is modest. The absorptive properties of the proximal small intestine are superior to those of the stomach or any other region of the GI tract. The rate at which compounds pass from the stomach into the small intestine is a rate-limiting step controlling drug absorption. Stomach motility is complex and is influenced by nervous and hormonal stimuli. Stomach emptying rate is a function of rhythmic contractions that have a frequency of approximately three per minute in a fasted person, and less when food enters the stomach.

Food passes from the stomach into the duodenum as a result of these rhythmic contractions. The heavier the meal and the higher the fat content, the longer it will take for a meal, and any drug that may be ingested with it, to pass into the small intestine. This process acts as a defense mechanism by which substances are prevented from entering the proximal small intestine and injuring the delicate absorptive surface of this region until they are reduced to a suitable consistency in the stomach.

Whereas solid food delays stomach emptying, liquids tend to accelerate the process. Acceleration results from activation of stretch receptors in the stomach wall. When the fluid is water, activation of the inhibitory receptors is stopped. This results in rapid emptying of stomach contents into the duodenum.

By far the most important difference between the proximal small intestine and the stomach is the nature of the mucosal surface of the epithelium. The mucosal surface of the small intestine is increased by finger-like projections, or villi, that arise from the folds of Kerckring, and in turn by microvilli that arise from the villi. These invaginations increase the surface area of the intestinal mucosa some 600-fold to approximately 200 m², and 1.0–1.5 L of blood passes through intestinal capillaries each minute. Corresponding values for the stomach are only 100 m² of surface area and a blood flow rate of 150 ml/min. Thus, the small intestine has a surface area approximately double that of the stomach and a blood perfusion rate 6–10 times faster. Both factors strongly favor more efficient absorption from the small intestine.

The villi and microvilli of the small intestine are lined by a sulphated mucoprotein, glycocalyx. Fluid trapped within the glycocalyx is stationary, and a series of thin layers, each progressively more stirred, extends to the bulk phase of the intestinal lumen. This series of unstirred layers has an effective thickness of 0.01–1.0 mm.

Molecules move within the unstirred layers by diffusion at a rate inversely proportional to the square root of molecular weight below 450, and inversely proportional to the cube root of molecular weight above 450. The glycocalyx is negatively charged, with counterions in the unstirred layer. If a substantial proportion of these cations is composed of hydrogen ions, as is often the case, then the microclimate within the brush border of the epithelium is likely to be acidic relative to the bulk phase. This may influence drug ionization at the membrane surface, and provides a basis for the “acid microclimate” frequently associated with the GI mucosa.

The length of time during which material stays in the small intestine is approximately 5 min in the duodenum, 2 h in the jejunum, and 3–6 h in the ileum. Material then enters the large intestine.

The large intestine does not have villi or microvilli at its mucosal surface. Its contents are neutral or alkaline. Therefore, absorption of drugs from the large intestine is less efficient than from the small intestine. The large intestine and colon contain an active bacterial microflora that can degrade foreign molecules that also tends to reduce absorption of drugs from this region of the GI tract.

GI Secretions

The rate of acid secretion into the stomach is controlled by parietal cells. Acetylcholine, histamine, and gastrin are important for regulation of hydrochloric acid secretion, and they act directly on parietal cells to enhance acid secretion rate.

Phases of gastric acid secretion

The basal rate of hydrochloric acid secretion varies diurnally, being highest in the evening and lowest in the morning. After ingestion of a meal, the rate of acid secretion in the stomach increases. The three phases of increased acid secretion in response to food are the cephalic phase (before food reaches the stomach), the gastric phase (elicited by the presence of food in the stomach), and the intestinal phase (elicited by input from the duodenum and upper jejunum).

Cephalic phase

The sight, smell, and taste of food elicit this phase. Acid secretion during this phase can be as much as 40% of the

maximum rate. Other stimuli sensed in the brain, in addition to those related to the presence of food, may evoke acid secretion through vagal impulses.

Gastric phase

The presence of food in the stomach evokes gastric secretion. The principal stimuli include distension of the stomach and the presence of amino acids and peptides. Distension of the stomach stimulates mechanoreceptors that bring about secretion of acetylcholine, hydrochloric acid, and gastrin.

Intestinal phase

The presence of chyme in the duodenum stimulates neuronal and endocrine responses that stimulate and later inhibit secretion of acid into the stomach. The stimulatory influences dominate when the pH of gastric chyme is above 3. However, when the buffer capacity is exhausted and the pH falls below 2, inhibitory influences dominate.

Gastric acid secretion may be regulated by several brain peptides, some of which may enhance secretion, while others may act centrally to inhibit secretion.

Gastric Juice

Gastric juice contains a mixture of secretions from surface epithelial cells and gastric glands. Salts, water, pepsins, intrinsic factor, and mucus are main components of gastric juice. Gastric secretions increase after a meal. Ionic composition of gastric juice is related to the rate of secretion. The higher the secretory rate, the higher the hydrogen ion concentration. The rate of gastric acid secretion varies among individuals. In humans, the basal rate is 1–5 mEq/h. With histamine or pentagastrin stimulation, acid output rises to 5–40 mEq/h.

Other Secretions

Bile, pH 7.8–8.6, is produced continuously in humans. Hepatic bile is concentrated and stored in the gall bladder between meals. It is ejected from the gall bladder and flows into the duodenum when food enters the intestine. The main constituents of bile are bile salts, bilirubin, end products of hemoglobin breakdown, the electrolytes sodium, chloride, and bicarbonate, cholesterol, phospholipids, and lecithin. The gall bladder contracts within 30 min after eating due to liberation of cholecystokinin. The most effective stimulus to this is food high in fat.

Bile salts, which are surface active, promote dissolution of lipophilic drugs and lipophilic drug formulations, enteric coatings, and waxy drug matrices. Bile salts may

also promote membrane permeability of lipophilic molecules through micelle formation and solubilization.

Pancreatic juice contains an alkaline fluid and enzymes, both of which empty into the duodenum. The alkaline pH contributes to neutralization of the acid that empties from the stomach. The enzymes amylase, lipase, trypsin, and chymotrypsin play major roles in the digestion of carbohydrates, fats, and proteins. Trypsin and chymotrypsin are secreted as inactive precursors and are converted to the active forms enzymatically.

As a result of proteolytic enzyme secretion into the duodenum, protein or peptide drugs, such as corticotropin, vasopressin, and insulin, are rapidly degraded and generally cannot be given orally. Secretory activity of the pancreas is under hormonal and neuronal control.

Intestinal secretions do not exist in the same sense as gastric, pancreatic, or biliary secretions. Nonetheless, large fluid fluxes take place throughout the intestine. Any secretions from the intestinal mucosa appear to have a lubricant and protective effect.

GI Blood Flow in Relation to Drug Absorption

Drugs may be transported away from the serosal side of the GI tract by one or both of two mechanisms. The GI tract is supplied by a blood capillary network from the splanchnic circulation. Drugs may also be taken up by the lymph vessels in the GI epithelium and carried by the lymphatic system that drains the abdominal area into the thoracic duct. Any drug that is absorbed via this system enters the systemic circulation directly and is not susceptible to first-pass hepatic metabolism. Despite the presence of both the blood capillary network and the lymphatic system, absorption of the great majority of drugs appears to occur predominantly via the capillary system associated with the splanchnic circulation.

The reason for this appears to lie in the relative flow rates of blood and lymph. The rate of blood flow in the splanchnic circulation is 1.0–1.5 L/min, or 30% of cardiac output. This rate may increase to 2 L/min after a meal. Lymph flow through the same region is only 1–2 ml/min, but may increase to 5–20 ml/min after a meal. Lymph flow in this region is thus 500–700 times slower than blood flow. Relatively fast splanchnic blood flow establishes virtual sink conditions on the serosal side of the GI epithelium and ensures a steep concentration gradient. These conditions promote efficient absorption into the bloodstream rather than into lymph.

Only a small number of drugs are absorbed via the lymph system. These include drugs with high molecular weights that cannot enter the capillaries and specific molecules such as steroids.

Hepatic First-Pass Metabolism

The majority of compounds absorbed from the stomach and intestines enter the splanchnic circulation. This leads to the portal vein, the liver, and then to the general circulation. Compounds absorbed via this route must therefore pass through the liver and will do so initially at a higher concentration relative to when they eventually distribute into the general circulation and elsewhere.

As hepatic metabolism is generally first order in nature, a large proportion of any orally administered drug that is highly and efficiently metabolized in the liver will be cleared during the initial first pass. A drug could be efficiently absorbed from the GI tract and yet poorly available to the general circulation as a consequence of first-pass hepatic clearance. Such high extraction drugs include acebutolol, alprenolol, desipramine, isoproterenol, and lidocaine.

ABSORPTION MECHANISMS

An orally administered drug must pass through a number of membranes in order to be absorbed into the systemic circulation. Many physiological membranes differ in structure and function. Despite this, there is general consensus regarding the basic structure of the cell membrane.

The Cell Membrane

The primary structure of the cell membrane, shown in Fig. 3 (9) is a 5-nm thick bimolecular lipid film that separates intracellular and extracellular fluids. The lipid is composed mainly of the phospholipids phosphatidylserine and phosphatidylinositol, and contains saturated and unsaturated fatty acids and sterols. The bilayer exhibits high permeability to hydrophobic molecules and low permeability to hydrophilic molecules.

The cell membrane is associated with intrinsic and extrinsic proteins. Intrinsic proteins are globular proteins that generally span the bilayer and are held within the membrane by hydrophobic and electrostatic interactions. The proteins can form channels, carriers, or pumps that enable polar molecules to cross the membrane.

Membrane Transport

Several mechanisms have been identified for drug transport across membranes. One of them is passive. The remainder utilize some type of carrier mechanism.

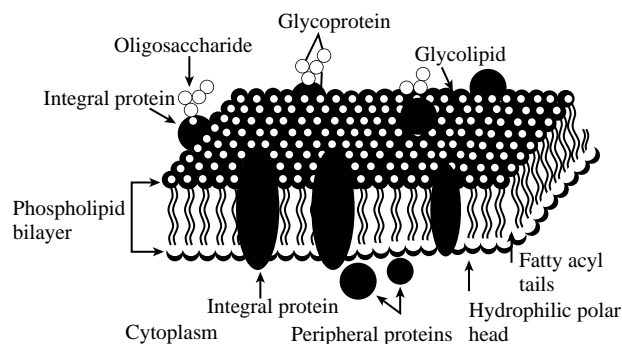


Fig. 3 Structural model of the cell membrane. The membrane is composed of a bimolecular leaflet of phospholipid with the polar head groups facing the extracellular and cytosolic compartments and the acyl groups in the middle of the bilayer. Integral membrane proteins are embedded in the lipid bilayer. Integral proteins are glycosylated on the exterior surface and may be phosphorylated on the cytoplasmic surface. Extrinsic membrane proteins, peripheral proteins, are linked to the cytosolic surface of the intrinsic proteins by electrostatic interactions. (From Ref. 9.)

Simple or Passive Membrane Transport

This mechanism, based primarily on lipid solubility and concentration gradient, is responsible for membrane transport of the great majority of drugs. The range of membrane permeabilities is very high. Typically, hydrophobic molecules have high partition coefficients, while hydrophilic molecules have low partition coefficients. The direction and rate of transport are determined by the concentration gradient of a substance across the membrane.

Molecules that are weak acids or bases cross membranes more readily when they are in the nonionized form. However, aqueous solubility is favored for the ionized form. In order to be available to cross any membrane, a drug must be in solution. This paradoxical requirement of both aqueous and lipid solubility is of particular concern in the area of drug absorption and presents a constant challenge in pharmaceutical formulation.

Transport Proteins

Many substances, particularly polar molecules, cross membranes at rates greater than those predicted from solubility and permeability data. Some can cross membranes against a concentration gradient. Unexpectedly high membrane permeability is related to transport proteins. Many transport proteins have been identified, cloned, and sequenced. Current knowledge has permitted

an operational definition of carrier proteins as channels, carriers, and pumps. The current state of the art in identification and characterization of these systems has been described by Wright (9).

Channels

Initial speculation on the existence of small aqueous pores in membranes was based on high membrane permeability of small polar molecules. For example, the permeability of water is 1000-fold more, and that of urea 10- to 100-fold more, than predicted. These types of observations led to the prediction of aqueous channels with radii of approximately 4 Å.

Water channels

The presence of water channels has been demonstrated by successful cloning of proteins that increase membrane water permeability. These have been expressed in erythrocytes and in cells of the renal tubule.

Ion channels

Evidence for ion channels in biological membranes was introduced in the 1970s. Possibly the most significant experiment was by Neher and Sakmann (10), who recorded single ion channel currents in muscle fibers. Many types of channels for sodium, potassium, calcium, and chloride ions have been described, and each has a specific conductance, ion selectivity, and probability of opening. Ion channel opening may be controlled by voltage or by ligand, and channels are thus designated voltage-gated or ligand-gated. Each of the ion channels has a specific pharmacology. Sodium channels are 12 times more selective for sodium than for other cations. Calcium channels are 1000 times more selective for calcium than for other cations.

Facilitated Diffusion

Facilitated diffusion is a simple mechanism proposed to explain transport of water soluble compounds. The main characteristics of this transport system are that membrane permeability exceeds that predicted from partition coefficients, transport occurs down a concentration gradient, transport is saturable, and competition occurs between isomers. Facilitated diffusion has been used to explain cellular uptake of sugars and amino acids.

Six human sugar transporters with different tissue distributions, substrate kinetics, and specificities have been identified. A number of facilitated amino acid transporters have also been identified in mammalian cells. System L,

which transports neutral amino acids, such as leucine and phenylalanine, is probably the best known of these.

Pumps

Pumps are proteins that can transport ions against electrochemical potential gradients using adenosine-5-triphosphate (ATP) as an energy source. Sodium–potassium pumps maintain intracellular sodium and potassium concentrations in animal cells and also control salt and water absorption by the epithelial cells in the intestine and kidney. The sodium–potassium pump transports three sodium ions out of the cell and two potassium ions into the cell at the cost of one molecule of ATP. The 3:2 coupling ratio results in net loss of sodium ions into the cell down an electrochemical gradient and maintains cell volume. Currently, considerable research is attempting to elucidate the structures of the various isoforms and subunits of sodium potassium pumps.

Cotransporters and Exchangers

There are many other examples of ions and ionized molecules accumulating in cells against their concentration gradient, such as uptake of iodine by the thyroid gland, accumulation of acids in liver cells, and absorption of sugars and phosphate by the small intestine. Recent studies have shown that these are governed by cotransport mechanisms. In all cases tested to date, sodium or hydrogen ion gradients are used to drive cotransporters, and these gradients are maintained by ion pumps. Glucose transport across the brush border of the small intestine is coupled with sodium transport, and uphill sugar transport is driven by the sodium gradient.

Cotransporters

Cotransporters use the sodium or hydrogen ion gradient to drive transport of a substrate. Many cotransporters have been described, cloned, sequenced, and expressed. The sodium–glucose cotransporter just described is one of these.

Other cotransporters facilitate the transport of other sugars, osmolytes, and amino acids. In humans, a disorder of intestinal glucose and galactose absorption is due to a defective sodium–glucose transporter.

Antiporters

The best characterized antiporters, or exchangers, are the chloride–bicarbonate, sodium–hydrogen ion, and sodium–calcium exchangers. The cellular sodium–hydrogen ion exchanger controls cell volume, pH, growth,

and sodium transport. Mammalian isoforms of these exchangers have been cloned and sequenced. Sodium–hydrogen exchangers play a dominant role in the regulation of intracellular calcium, and thus the force of contraction of the heart. The therapeutic effect of cardiac glycosides is probably related to decreased sodium–calcium exchange in the heart caused by a decreased sodium–hydrogen gradient across the cell membrane.

Permeability Glycoprotein

Permeability glycoprotein (p-glycoprotein) is an ATP-dependent efflux pump responsible for pumping substances out of cells. It is implicated in development of drug resistance in tumor cells. Localization of p-glycoprotein in the apical membranes of intestinal, liver, and kidney cells, and also at the blood brain barrier, provides potential for this pump to have a profound effect on drug absorption, distribution, and elimination, as well as in drug–drug interactions.

Pinocytosis and Endocytosis

Pinocytosis is a nonspecific process whereby a substrate enters a cell by invagination to form an intracellular vesicle. Receptor-mediated endocytosis occurs when substrate binds to a specific membrane receptor. Substrates ingested by cells in this way are stored in vesicles or degraded. Receptor-mediated endocytosis is involved in cellular uptake of immunoglobulin and low density lipoprotein.

FORMULATION FACTORS AFFECTING DRUG ABSORPTION AND ABSORPTION ENHANCERS

The chemical and physical properties of a drug and its formulation can affect drug stability and absorption characteristics.

Chemical Factors

A variety of chemical options can be used to improve the stability and systemic availability of drugs. For example, esters can be prepared of both acids and bases to produce more stable derivatives, which hydrolyse to the active parent once absorbed. The stability and solubility of both acids and bases tend to increase when they are in the form of salts. Typically, administration of soluble salts of penicillin give rise to higher circulating antibiotic levels

than the free acid. When the salt of a weak acid dissolves in the stomach, it generates a diffusion layer of relatively high pH which, in turn, promotes further dissolution. The same argument could theoretically be used for basic drugs. However, the pH effect in this case is swamped by the very low pH present in stomach fluids. Thus, salts of basic drugs are used primarily for handling and solubility rather than for improved dissolution.

Physical Factors

Different physical forms of a drug can affect its absorption. Typically, the crystal or polymorphic form, the state or nature of hydration or solvation, and physical size of drug particles may have considerable impact on the rate and extent of drug absorption.

Polymorphism and amorphism

Many compounds form crystals with different molecular arrangements, or polymorphs. These polymorphs may have different physical properties, such as dissolution rate and solubility. The vitamin riboflavin exists in several polymorphic forms, and these have a 20-fold range in aqueous solubility. Polymorphs that have no crystal structure, or amorphic forms, have different physical properties from the crystalline forms.

Absorption of many orally administered drugs is controlled by dissolution rate. Amorphous forms generally dissolve faster than crystalline forms because no energy is needed to break up the crystal lattice. For this reason, the amorphous form is often preferred over the crystalline form and several drugs, including hydrocortisone and prednisolone, are marketed in the amorphic form.

Solvation

During their preparation, drug crystals may incorporate one or more solvent molecules to form solvates. The most common solvate is water. If water molecules are already present in a crystal structure, the tendency of the crystal to attract additional water to initiate the dissolution process is reduced, and solvated (hydrated) crystals tend to dissolve more slowly than anhydrous forms. Significant differences have been reported in the dissolution rate of hydrated and anhydrous forms of ampicillin, caffeine, theophylline, glutethimide, and mercaptopurine. The clinical significance of these differences has not been examined but is likely to be slight.

Particle size

Particle size may play a major role in drug absorption. Dissolution rate of solid particles is proportional to surface

area, and hence to particle fineness. Particle size reduction has been used to increase the absorption of a large number of poorly soluble drugs, such as bishydroxycoumarin, digoxin, griseofulvin, nitrofurantoin, and tolbutamide.

Griseofulvin has extremely low aqueous solubility, and material of normal particle size gave rise to poor and erratic absorption. Microsize particles improve absorption, but it is improved even more when it is formulated in ultramicrosize particles as a monomolecular dispersion in polyethylene glycol.

Formulation Factors

Drug formulations are designed to provide an attractive, stable, and convenient method to use products. Conventional dosage forms may be broadly characterized in order of decreasing dissolution rate as solutions, solid solutions, suspensions, capsules and tablets, coated capsules and tablets, and controlled release formulations.

Solutions

Aqueous solutions, syrups, elixirs, and emulsions do not present a dissolution problem and generally result in fast and often complete absorption as compared to solid dosage forms. Due to their generally good systemic availability, solutions are frequently used as bioavailability standards against which other dosage forms are compared.

Solid solutions

The solid solution is a formulation in which drug is trapped as a solid solution or monomolecular dispersion in a water-soluble matrix. Although the solid solution is an attractive approach to increase drug absorption, only one drug, griseofulvin, is currently marketed in this form.

Suspensions

A drug in a suspension is in solid form, but is finely divided and has a large surface area. Drug particles can diffuse readily between the stomach and small intestine so that absorption is relatively insensitive to stomach emptying rate.

Similar to solutions, suspensions are useful for patients who have difficulty taking solid medication. Adjusting the dose to a patient's needs is easier with solutions and suspensions than with solid dosage forms. Liquid dosage forms, therefore, have several practical advantages besides simple dissolution rate. However, they also have some disadvantages, including greater bulk, difficulty in handling, and perhaps reduced stability.

Capsules and tablets

Capsules and tablets are the most common oral dosage forms. These formulations differ from each other in that material in capsules is less impacted than in compressed tablets. Once a capsule dissolves, the contents generally disperse quickly. The capsule material, although water soluble, can impede drug dissolution by interacting with the drug, but this is uncommon.

Tablets generally disintegrate in stages, first into granules and then into primary particles. As particle size decreases, dissolution rate increases due to increased surface area.

Tablet disintegration was once considered a sufficient criterion to predict *in vivo* absorption. This was proven inadequate, however, and dissolution is now recognized as a better criterion. Regulatory agencies now require dissolution rate data for all new oral formulations. The increasingly wide acceptance of dissolution as the best available *in vitro* parameter to predict *in vivo* absorption is reflected in the proliferation of such tests in official compendia.

Excipients: Along with active material contained in tablets and capsules, a variety of so-called inert ingredients are present, for example, starch, magnesium aluminum silicate, methylcellulose, carboxymethylcellulose, lactose, kaolin, talc, calcium sulfate, and magnesium stearate. Tablets may also have a variety of coatings to improve stability, taste, appearance, and drug release characteristics. Although considered to be inert, these additives can affect drug dissolution and absorption. Changing an excipient from calcium sulfate to lactose and increasing the proportion of magnesium silicate, increases the activity of oral phenytoin. Systemic availability of thiamine and riboflavin is reduced by the presence of Fuller's earth. Absorption of tetracycline from capsules is reduced by calcium phosphate due to complexation.

Most of these types of interactions were reported some time ago and are unlikely to occur in the current environment of rigorous testing of new dosage forms and formulations.

Coated tablets

Tablets may be formulated with coatings such as shellac, resin, or styrene-maleic acid copolymer. These coatings are insoluble in acid but dissolve readily at neutral or alkaline pH. Thus they are ideally suited to prevent drug release until the formulation has passed from the stomach into the small intestine. Preventing drug release in the stomach may protect drugs that are acid labile. It may also protect the patient from irritant substances like iron salts, diethylstilboestrol, and some anti-inflammatory agents.

Release, and subsequent systemic availability of drugs from these formulations is likely to be highly sensitive to stomach emptying patterns.

In vitro–in vivo correlations: The relationship between *in vitro* dissolution and *in vivo* bioavailability is of considerable interest today. The U.S. Food and Drug Administration (US FDA) has spearheaded a research program to examine these relationships with the objectives to gain a better understanding of their interdependence and to use the relationships as a means to predict *in vivo* performance from *in vitro* data. The cost savings of realization of this second objective would be significant. The main thrust of research in this area is based on differentiation of drugs or formulations in terms of solubility and membrane permeability (11).

Drugs or formulations can be considered in four groups: 1) high solubility and low permeability; 2) high solubility and high permeability; 3) low solubility and high permeability and 4) low solubility and low permeability. For drugs in group 3, dissolution is likely to be rate-limiting for absorption and *in vitro* dissolution data may be useful. For drugs in group 1, on the other hand, permeability is probably rate-limiting and *in vitro* data are less likely to be useful. For groups 2 and 4, the picture is less clear and *in vitro*–*in vivo* relationships would need to be determined experimentally.

Controlled release formulations

Appreciation of the advantages of controlled drug release, development of many novel controlled release systems, and also the interest of major pharmaceutical houses in protecting marketed drug products, have led to increased interest in this type of dosage form. Most controlled release products currently marketed include diuretic agents, cardiovascular and respiratory drugs, and compounds acting on the CNS. Little attention has been paid to antimicrobial agents.

Advantages of controlled drug release: Due to of their generally higher cost, controlled release dosage forms can be justified only if they offer therapeutic advantages, i.e., improved maintenance of therapeutic drug levels in the circulation, reduced dosing frequency, reduced fluctuation in circulating drug levels, increased convenience to the patient, reduced patient care time, less nighttime dosing, more uniform pharmacologic response, reduced GI irritation, and reduced side effects. The second of these, reduced dosing frequency, has often been claimed as a sufficient rationale for development of a controlled release dosage form, but has become unacceptable as a sole criterion. This is understandable given the current emphasis on cost containment in health care.

Disadvantages of controlled drug release: Potential disadvantages of controlled release dosage forms include the possibility of dose dumping, less facile dose adjustment, increased potential for hepatic first-pass metabolism, possible delay in onset of action, possibly lower system availability, and time of drug release limited to residence time of formulation in the optimum absorption region(s) of GI tract.

Dose dumping, or inadvertent rapid release of drug, is important for potent drugs that have a narrow therapeutic index. Good manufacturing practice (GMP) generally reduces the probability of this happening. Fine dose adjustment is often difficult with controlled release formulations. Controlled release tablets that use a granule matrix may be subdivided in order to reduce the dose, but repeat action tablets or osmotic pump devices lose their controlled release properties once the dosage form is fractured. Increased first-pass metabolism may occur with drugs that are cleared by the liver, but only if hepatic clearance is saturable following rapid absorption from conventional dosage forms. Reduced systemic availability is common with controlled release dosage form, availability generally being 80–85% of that from conventional formulations. Limited residence time in the GI tract is a potential disadvantage of oral controlled release products, and this distinguishes oral from other controlled release dosage forms (e.g., skin patches, which can provide slow release of drug over a prolonged period).

Drugs that are unsuitable for controlled release: Some drugs are unsuitable for controlled release formulations. Typical characteristics of such drugs include short biological half-life, long biological half-life, potent drug with narrow therapeutic index, large dose, poorly absorbed, low or slow dissolution, active absorption, time course of activity not the same as that of circulating drug levels, and extensive first-pass metabolism.

A controlled release form of a drug that has a short biological half-life, <2 h, or is administered in large doses may need to contain a prohibitively large amount of drug. Drugs with long biological half-lives (>8 h) are generally sufficiently sustained in the body from conventional doses, and prolonged release is unnecessary. Incorporating slowly dissolving compounds into a controlled release formulation is likely to be counterproductive since dissolution is rate-limiting anyway. Administering drugs like warfarin, whose pharmacologic effect is prolonged relative to its blood profile, offers no therapeutic advantage. Incorporating such compounds as some beta-lactam antibiotics, fluorouracil, and some amino acids, which appear to be absorbed predominantly from the proximal intestine, is likely to reduce their efficacy and

achieve little or no prolongation of effect. As stated earlier, if a drug undergoes saturable first-pass metabolism from conventional doses, its systemic availability may be decreased after controlled release.

Although the above arguments provide useful general rules, there are many exceptions. Nitroglycerin has a biological half-life of less than 0.5 h. It is generally considered to be poorly absorbed and is rapidly metabolized by the liver, with obvious first-pass implications. However, a large number of oral controlled nitroglycerin products are marketed. Low circulating levels of nitroglycerin obtained from these products appear to provide adequate prophylaxis against angina attacks, but not against acute angina episodes. Some established and more recently introduced controlled release dosage forms are given in Table 1 (12).

Absorption Enhancers

Although oral dosing is generally more convenient than other dosage routes, oral absorption of many drugs is poor. As molecules become larger, more complex, and generally more lipophilic in the quest for new or improved efficacy, their absorption tends to decline. To address this problem, absorption enhancers continue to be examined, so far with variable success. Some compounds that have been shown, largely in animal studies, to increase absorption of drugs are shown in Table 2 (13). Little or no information is available in humans for most of these compounds.

Nonsteroidal antiinflammatory agents (NSAIDs)

NSAIDs, in particular indomethacin, diclofenac, mepirazole, phenylbutazone, and salicylate, can promote absorption of other drugs, including insulin, ampicillin, cephalothin, cefoxitin, and cefmetazole. Most of these observations were made in the rat and frequently after rectal administration. Several mechanisms by which NSAIDs promote drug absorption have been postulated, but exact mechanisms are not known. As NSAIDs are often irritating to the GI mucosa, and this may well relate to their absorption enhancing ability, the feasibility of their use to promote drug absorption is uncertain.

Surfactants

In view of their solubilizing effects and also their potential to change membrane permeability, surfactants have been considered as absorption enhancers, again mostly in animals. Polyoxyethylene ethers have been shown to enhance gastric or rectal absorption of lincomycin, penicillin, cephalosporins, and fosfomycin in rats and

Table 1 Some oral controlled-release dosage forms

Category	Product	Active ingredient
Slow erosion with initial fast release dose	Tedral SA	Theophylline, ephedrine, phenobarbital
Erosion core only	Tenuate Dospan	Diethylpropion
Repeat action tablets	Chlor-trimeton repetabs	Pseudoephedrine, chlorpheniramine
Pellets in capsules	Combid spansule	Isopropamide, prochlorperazine
Pellets in tablets	Theo Dur	Theophylline
Leaching	Desbutal gradumet	Methamphetamine, pentobarbital
Ion exchange resin	Biphetamine	Amphetamine, Dextroamphetamine
Complexation	Rynata	Chlorpheniramine, phenylephrine, pyrilamine
Microencapsulation	Nitrospan	Nitroglycerin
Flotation–diffusion	Valrelease	Diazepam
Osmotic pump	Acutrim	Phenylpropanolamine
	ProcardiaXL	Nifedipine

(From Ref. 12.)

rabbits. In rats, colonic absorption of interferon- α is increased from 3 to 8% by polyoxyethylene esters of oleic acid and oleic acid glycerides.

Some studies have examined the effects of surfactants on intestinal absorption of insulin, with variable results. Both rectal and jejunal absorption of insulin was increased by anionic and cationic surfactants. However, in humans, oral polyoxyethylene-20-oleyl ether resulted in poor and variable insulin absorption (14).

Any enhancing effect of surfactants on drug absorption appears to be related to increased drug solubilization, modification of mucosal permeability, or reduction of resistance of the unstirred water layer at the GI membrane surface. In general, unionic surfactants have little effect on membrane structure but cationic surfactants have been associated with reversible cell loss and loss of goblet cells. These effects must limit consideration of surfactants as absorption promoters, particularly for long term treatment.

Table 2 Some types of oral drug absorption enhancers

Nonsteroidal antiInflammatory agents
Surfactants
Bile salts
Medium chain fatty acids
Mixed micelles
Liposomes
Azone
Cell permeation enhancers
Nanoparticles

(From Ref. 13.)

Bile salts

Bile contains conjugates of cholic acid and chenodeoxycholic acid, which emulsify dietary fat, facilitate lipolysis, and transport lipid molecules through the unstirred layer of the intestinal mucosa by micellar solubilization. The ability of bile salts to promote lipid absorption has prompted their investigation as absorption enhancers for drugs, with modest success. Studies in animals have demonstrated increased intestinal absorption of heparin and interferon- α . Absorption of insulin can be increased by bile salts, both in experimental animals and in humans. The effect on drug absorption appears to correlate with mucosal damage. This, together with possible cocarcinogenic and comutagenic properties of secondary bile salts, reduces the attractiveness of bile salts as absorption enhancers.

Medium-chain fatty acids and glycerides

The presence of medium chain fatty acids and glycerides in food products has stimulated interest in their potential utility as absorption enhancers. Some fatty acids and glycerides have been shown to increase drug absorption under a variety of conditions, almost always in animals and in most cases after rectal dosing. However, some studies have yielded positive results after oral dosing. Oral insulin bioavailability was increased to 9–13% relative to IM administration by a mixture of sodium dodecanoate and cetyl alcohol (15). Afiraxone absorption was enhanced by glyceryl-1-monooctanoate after oral, duodenal, and rectal administration to animals.

Despite the potential of these classes of compounds as absorption enhancers, they have been shown to have negative effects on mucosal membrane integrity. Additional research is needed to evaluate risks and benefits.

Mixed micelles

Mixed micelles consist of fatty acids solubilized by surfactants or bile salts. The effects of mixed micelles on drug absorption were reviewed by Muranishi et al. (16). Mixed micelles are effective absorption enhancers for compounds such as heparin, streptomycin, gentamycin, and insulin. The effect of mixed micelles on drug absorption tends to be greater at the distal region of the GI tract. The mechanism for increased absorption is not known. Some publications claim that they are safe to use. Others report a disordering effect on intestinal epithelial cells.

Liposomes

Liposomes consist of vesicles composed of bilayers or multilayers that contain phospholipids and cholesterol surrounding an aqueous compartment. Drug is entrapped within the liposome and is released from the liposome for absorption at the intestinal membrane surface. This dosage form received considerable attention during the 1970s and 1980s, and several animal studies demonstrated potential for absorption enhancement. However, lack of effect in other studies, and also stability problems, have resulted in reduced interest in liposomes as absorption enhancers.

Azone

Azone (1-Dodecylazacycloheptan-2-one) and related compounds have been studied as transdermal penetration and oral absorption enhancers. Although some efficacy has been shown, an emulsifying agent appears to be necessary for azone to penetrate the intestinal mucosal membrane in order to promote drug absorption. One study reported the absence of gross morphological damage after exposure of mucosa to azone (17) but additional information on the effect of azone on overall mucosa structure is not available.

Cell permeation enhancers

Although the objective of most absorption enhancers is to avoid direct interaction with the mucosal membrane, cell permeation enhancers use this as a means to increase drug absorption. One form of enhancer currently of interest consists of glycosylated molecules, or facial amphiphiles. It is claimed that these compounds temporarily increase membrane permeability. Molecules are designed to self-assemble in membranes to form transient pores that permit hydrophilic compounds to cross the membrane. This technology has considerable potential for absorption

enhancement. No adverse effects have been reported to date (18).

Nanoparticles

From known relationships between surface area and dissolution, it is reasonable to predict that ultrafine particles may increase the dissolution rate of relatively insoluble compounds. If these particles are then stabilized to avoid aggregation and agglomeration, and yet retain fluidity, then a useful drug product could be obtained.

This concept has found expression in a proprietary nanoparticle technology in which a drug is reduced to nanometer-size particles in the presence of stabilizers. Originally developed for IV computer imaging (19), this technology shows considerable promise to increase absorption of poorly water soluble compounds. As the nanoparticle system is purely "formulation" in nature, it is unlikely to affect GI mucosal integrity. Nanoparticles are sufficiently small that they can be used parenterally, apparently without ill effects. More is likely to be heard about this novel absorption enhancer technology.

CONCLUSIONS

The rate and extent of drug absorption into the systemic circulation are key factors that influence drug pharmacologic activity. Drugs may be administered by a variety of routes, each with its own advantages and disadvantages. The route of administration for a particular drug is dictated by the properties of the drug and the systemic activity profile required.

Increased knowledge of membrane structure and of transmembrane transport has improved understanding of the mechanisms of drug absorption and of ways in which this may be modulated.

Formulation continues to play a pivotal role in drug absorption. Many enhancer technologies have been examined, with varying success. However, some recent technologies based on formulation or membrane effects show considerable potential to increase absorption of orally administered compounds.

REFERENCES

1. Yamada, K.; Bremer, A.M.; West, C.R.; Ghoorah, J.; Park, H.C.; Takita, H. Intra-arterial BCNU Therapy in the Treatment of Metastatic Brain Tumor from Lung

- Carcinoma. A Preliminary Report. *Cancer* **1979**, *44*, 2000–2007.
2. Goldman, J.A.; Peleg, D.; Agmon, M.; Shapiro, G. Arteriography and Chemotherapy in Localized Trophoblastic Disease by Means of Local-Pelvic-Intra-arterial Infusion. *Acta Obstet. Gynecol. Scand.* **1979**, *58*, 415–416.
3. Florence, A.T.; Jenkins, A.W.; Loveless, A.H. Approaches to the Formulation of a Long-acting Intramuscular Injection. *J. Pharm. Pharmacol.* **1973**, *25* (Suppl.), 120–121.
4. Chandrasekaran, S.K. Controlled Release of Scopolamine for Prophylaxis of Motion Sickness. *Drug Dev. Ind. Pharm.* **1983**, *9*, 627–646.
5. Su, K.S.E.; Campanale, K.M. Nasal Drug Delivery Systems Requirements, Development and Evaluations. *Transnasal Systemic Medications*; Chien, Y.W., Ed.; Elsevier Scientific: Amsterdam, 1985; 139–159.
6. Newman, S.P.; Miller, A.B.; Lennard-Jones, T.R.; Moren, E.; Clarke, S.W. Improvement of Pressurised Aerosol Deposition with Nebuhaler Space Device. *Thorax*. **1984**, *39*, 935–941.
7. Rogge, M.C.; Johnson, C.A.; Zimmermann, S.W.; Welling, P.G. Vancomycin Disposition During Continuous Ambulatory Peritoneal Dialysis: A Pharmacokinetic Analysis of Peritoneal Drug Transport. *Antimicrob. Ag. Chemother.* **1985**, *27*, 578–582.
8. Ritschel, W.A.; Elconin, H.; Alcorn, G.J.; Denson, D.D. First-pass Elimination of Lidocaine in the Rabbit after Peroral and Rectal of Administration. *Biopharm. Drug Dispos.* **1985**, *6*, 281–290.
9. Wright, E.M. Membrane Transport. *Pharmacokinetics: Regulatory, Industrial, and Academic Perspectives*, 2nd Ed., Welling, P.G., Tse, F.L.S., Eds.; Marcel Dekker, Inc.: New York, 1995; 89–118.
10. Neher, E.; Sakmann, B. Single-channel Currents Recorded from Membrane of Denervated Frog Muscle Fibres. *Nature* **1976**, *260*, 799–802.
11. Amidon, G.L.; Lennernas, H.; Shah, V.P.; Crison, J.R. A Theoretical Basis for a Biopharmaceutics Drug Classification: The Correlation of In Vitro Drug Product Dissolution and In Vivo Bioavailability. *Pharm. Res.* **1995**, *420*, 413–420.
12. Welling, P.G. Physicochemical and Formulation Factors Affecting Drug Absorption. *Pharmacokinetics: Processes, Mathematics, and Applications*, 2nd Ed., American Chemical Society: Washington, DC, 1997; 88.
13. van Hoogdalem, E.J.; de Boer, A.G.; Breimer, D.D. Intestinal Drug Absorption Enhancement. An Overview. *Pharmacol. Ther.* **1989**, *44*, 407–433.
14. Galloway, J.A.; Root, M.A. New Forms of Insulin. *Diabetes* **1972**, *21* (Suppl 2), 637–648.
15. Touitou, E.; Rubinstein, A.; Donbrow, M.; Azaz, E. New Hydrophilic Vehicle Enabling Rectal and Vaginal Absorption of Insulin, Heparin, Phenol Red and Gentamicin. *J. Pharm. Pharmacol.* **1978**, *30*, 662–663.
16. Muranishi, S. Modification of Intestinal Absorption of Drugs by Lipoidal Adjuvants. *Pharm. Res.* **1985**, *2*, 108–118.
17. Murakami, M.; Takado, K.; Muranashi, S. Promoting Effect of Azone on Intestinal Absorption of Poorly Absorbable Drugs in Rats. *Int. J. Pharm.* **1986**, *31*, 231–238.
18. Cheng, Y.; Ho, D.M.; Gottlieb, C.R.; Kahne, D. Facial Amphiphiles. *J. Am. Chem. Soc.* **1992**, *114*, 7319–7320.
19. Gazelle, G.S.; Wolf, G.L.; Bacon, E.R.; McIntire, G.L.; Cooper, E.R.; Toner, J.L. Nanocrystalline Computed Tomography Contrast Agents for Blood-Pool and Liver-Spleen Imaging. *Invest. Radiol.* **1994**, *29*, S268–S288.

ABUSE OF DRUGS

Richard B. Seymour

David E. Smith

Haight Ashbury Free Clinics, San Francisco, California, U.S.A.

INTRODUCTION

Drug abuse may manifest at one of several different levels. Habituation involves a distinct and possibly harmful pattern of use. Drug abuse has been defined as a pattern of problem use that results in health consequences, social problems, or both. Drug addiction is a chronic disease of the brain that involves relapse, progressive development, and the potential for fatality if not treated. Addiction cannot be cured but can be brought into remission through a program of treatment, abstinence from all psychoactive substances, and supported recovery. In general, the drugs involved in abuse of drugs are within the grouping of “psychoactive” drugs. These are substances that have their primary effect on the brain and central nervous system (CNS) and include opioids, sedative–hypnotics, stimulants, and hallucinogens. Recent additions include performance-enhancing drugs, such as steroids, and combinations producing the effects of several drug groups.

LEVEL OF DRUG USE AND ABUSE

Inaba and Cohen (1) list six levels of drug use and abuse. These are: 1) abstinence, 2) experimentation, 3) social/recreational, 4) habituation, 5) drug abuse, and 6) addiction. These levels may not be progressive from one to the next, but will indicate in a progression context if the individual is developing a drug problem.

Abstinence

Abstinence means no use of psychoactive substances. However, in our culture, or for that matter in any culture, psychoactive substances are virtually impossible to avoid. Nearly all of us have experienced pain medication. Most soft drinks, tea, and, of course, coffee contain caffeine, a potent stimulant. Yet, individuals who partake of these will consider themselves abstinent and with few

exceptions this use is marginal to the point of not being worth consideration.

In terms of potential drug abuse or addiction, the nominally abstinent person is in little danger. Development of abuse requires the initial use of potent drugs. Addiction may involve a genetic vulnerability, but it also requires an environment that is conducive to initial use, and that will be lacking in the consistently abstinent individual, so long as he or she remains abstinent.

Drug Experimentation

Experimentation usually ensues when the individual becomes curious about a drug's effects or is influenced to try a drug by relatives, friends, coworkers or such cues as advertising and word of mouth. Many young people in the late 1950s and early 1960s tried marijuana when it was offered to them. Although all they may have heard about the drug in school, from parents, teachers, and other authority figures was highly negative regarding the drug, they were curious as to its effects and that curiosity overrode the warnings that they had received.

Experimentation, however, involves no pattern of use and usually minimal negative consequences. The individual may use on occasion as the occasion presents itself, but there is no drug-seeking behavior, and the consequences are minimal. There are exceptions, however, and these may include:

- Using a large amount, such as binge drinking in high school or college under peer pressure, that results in accident, injury, or illness
- An extreme physical reaction to a small amount of drug via an allergy or idiosyncratic reaction
- Aggravation or triggering of a pre-existing physical or mental condition
- Use during pregnancy
- Use resulting in legal problems such as an arrest for possession or loss of a job after failing a drug test or
- An addictive response from individuals with a very high susceptibility to compulsive use triggering immediate abuse or addiction.

Social or Recreational Use

With social or recreational use, the individual does seek out the drug or the situation in which the drug is used. However, there is no established pattern. Drug use is irregular, infrequent, and still has minimal impact on the individual's life. All of the possible exceptions listed in the preceding paragraph apply, with perhaps a greater chance given that drug-seeking behavior has become part of the pattern.

Drug Habituation

A definite pattern of use is apparent in habituation. This category includes the individual who needs a cup of coffee to kick-start the day, or a martini on arriving home to relax and end the day. So long as no real harm to others or oneself is taking place, this sort of consistent drug use can be considered habituation. This is the point, however, where use is apt to slide over into abuse. Habituation indicates that the user has lost some control over the drug, that use has become a habit, and with a nudge could become a bad habit.

Drug Abuse

Drug abuse can be transient: the product of stress, the acting out of social and cultural patterns, the response to overwhelming circumstances, or the need to perform. Examples may include taking a sedative to turn off and go to sleep after a hard day's work with a lot of mental but no physical exercise; doing cocaine or other stimulants to cope with a 24/7 lifestyle; having a weekly night out with the guys or staying over at the pub with coworkers on the way home; getting drunk to deal with a crisis one does not really want to face; or using steroids to bulk up for a professional sporting competition.

On the other hand, it can become chronic. Abuse behavior takes place on a regular basis, no matter what the circumstances. The behavior has become ingrained and part of the abuser's routine, such as the individual who must have a cigarette as soon as he or she wakes up in the morning, talks on the phone, and sits down to a meal, etc. As of yet, the problem is not addiction.

Drug Addiction

Drug addiction is a disease of the brain that is characterized by compulsion, loss of control, and continued use despite adverse consequences. Inaba and Cohen (1) define compulsion as "H spent most of their time either using, getting, or thinking about the drug" as the step between abuse and addiction. Seymour and Smith (2) consider "loss of control" the pivotal point. In the

recovering community, that is the point where the user turns from a cucumber into a pickle.

Addictive disease is characterized by compulsive use of one or more drugs and loss of control over that use. Here, we have the individual who goes out, swears he or she will have no more than two beers and instead gets very drunk, has a blackout, and cannot remember getting home. Continued use despite adverse consequences can include the businessman who continues to use cocaine even after losing family, possessions, and business, or the person who smokes two to three packs of cigarettes a day despite having lung cancer.

Addiction is progressive and can be fatal if not treated. It is incurable in that the addict cannot go back to noncompulsive, non-out-of-control use. In the recovering community, it is said that a cucumber can continue to be a cucumber, but once it becomes a pickle, it cannot go back to being a cucumber. Although incurable, the disease can, however, be brought into remission via abstinence from all psychoactive substances and a program of supported recovery.

Abstinence on its own is not enough for the addict. Addiction is a function of the lower brain and cannot be controlled by force of will. Craving will wear down the addict's resolve and in an attempt to resist can create a rigidity known in the field as "white-knuckle sobriety," in which the abstinent addict is clinging so hard that it is like the driver with such a grip on the wheel that his or her knuckles are bloodless. That is no way to drive a car or to maintain recovery from addiction. Recovery for the addict is a lifelong undertaking and requires help (3).

THE NATURE OF PSYCHOACTIVE DRUGS

Psychoactive drugs are substances that are chemically similar to chemicals that occur naturally in the human brain and act as neurotransmitters of information between brain cells. Because of the similarity, these chemicals are able to pass through the blood-brain barrier that exists to protect the brain from foreign materials. Once in the brain, all psychoactive drugs produce their effects by stimulating the release, inhibiting the release, blocking the reuptake, or imitating the brain's own neurotransmitters. Often, in such actions as producing pleasure, relief from pain or inhibitions, or anxiety, these drugs perform more rapidly and effectively than the endogenous substances. That is the primary appeal of these drugs. People use them because they work. They produce the desired results, at least for the short term.

THE PSYCHOACTIVE DRUG CATEGORIES

With a few notable exceptions, the bulk of psychoactive drugs that are abused by human beings fall into four general categories. These categories are: 1) opioid/analgesic drugs; 2) sedative–hypnotic drugs; 3) stimulant drugs; and 4) hallucinogenic drugs. There are drugs of abuse that either fall outside these basic categories, such as ether and other general anesthetics and steroids, or are considered to have attributes of more than one category, and these include the stimulant hallucinogens. Although the effects of psychoactive drugs may vary, they do have several things in common:

- Their principle action is in the brain and CNS, and they are therefore also referred to in the literature as “CNS” drugs.
- They are able to cross the blood–brain barrier that usually protects the brain from foreign substances because they resemble chemicals that are indigenous to the brain.
- They all act by stimulating, depressing, or imitating neurotransmitters that are native to the human brain and
- They all produce some form of disinhibition euphoria while they are active in the brain. It is this disinhibition euphoria that tends to be the most alluring general quality of psychoactive drugs.

Opioid Drugs

The narcotic/analgesic drugs have been used medically for pain relief and have been abused primarily for their ability to induce a state of euphoria and control pain. Historical accounts of opium extend to Assyrian depiction’s of goddesses with poppy pods growing out of their heads from around 4000 BC. Opium smoking became endemic in China in the 19th century after the British began exporting the drug from large holdings in India. In Britain, opium pills of 2–3 grains were easily available from apothecaries well into the 20th century. These and the tincture of opium, laudanum, are thought to have addicted many British writers and artists of the “romantic” and “pre-Raphaelite” periods. In the United States, opium and cocaine were often combined in patent medicine and tonics sold by traveling “snake-oil salesmen” in rural areas. The typical turn of the century opioid abuser was “a middle-aged, middle-class white woman with children.”

Natural opioids, that is, opioids extracted directly from opium include codeine—used for dental and other postoperative pain; laudanum, paregoric—a mild tincture of opium mixed with camphor and used primarily for

control of diarrhea; and morphine. Heroin is a partial synthetic that combines morphine and diacetyl acid. There are a number of synthetic opioids, including the highly powerful fentanyl (Sublimaze[®]), methadone—used in morphine and heroin addiction treatment, and such pain control mainstays as meperidine (Demerol[®]), hydromorphone (Dilaudid[®]), and oxycodone (Percodan[®]).

Medical use of opioids

Both natural and synthetic opioids are now, as they have been throughout medical history, the primary means of providing relief from pain and anticipatory anxiety. Along with analgesia, they induce a corresponding state of well-being or euphoria and at high doses somnolence, sometimes referred to as twilight sleep. They can also provide a sense of being immune to the effects of environmental and psychic distress, what street users refer to as “being in the wicker basket.” Opioid drugs can also be effective in controlling diarrhea and coughing.

How opioids work

The molecular structure of opioids is similar to that of certain neurotransmitters that occur naturally in the brain. Because of the similarities, these drugs are able to cross the blood–brain barrier and able to occupy receptor sites used by these neurotransmitters. The brain substances are called endorphins, which is short for endogenous morphines. The endorphins are what provide our natural pain control.

If we consider pain to be a signal that something is wrong, then endorphins are the internal means of mediating that signal. The subjective sequence is more or less as follows: Say you hit your thumb with a hammer. Intense pain. The brain receives the message, “Stop hitting yourself on the thumb with that hammer!” You jump around and yell a bit. It really hurts. After a while, however, you may still feel some surface pain from damaged thumb tissue, but the intense initial pain is gone. The endorphins that the pain released in the brain have attached to receptor sites that have disconnected the acute pain signal to your CNS, and even given you a little sense of euphoria.

Nonmedical use and abuse

Opioid drugs provide a vastly amplified version of what the internal pain management messengers provide. Beyond that, the use of opioid drugs gives the addict access to the reinforcement reward system, normally reserved to reward the performance of species-specific survival behaviors. That access provides the user with an experience that the brain equates with profoundly important events like eating, drinking, and sex. As a

consequence, opioid use becomes an acquired drive state that permeates all aspects of human life. This quality makes these drugs prime candidates for nonmedical use and abuse. Nonmedical use often involves self-medication and can be a result of medical misprescribing. Chronic pain sufferers, for example, may seek out street opioids to provide ongoing relief in situations in which they have been underprescribed for pain medication.

Whether iatrogenic in nature or developed on the street within a drug subculture, addiction to opioid drugs can occur with any drug in this category. Street users generally gravitate toward morphine and heroin, available through illicit dealers. Middle-class addicts and health professionals find prescription opioids more available to them. However, that can change over time and with changes in user status and drug availability.

Ingestion of opioids

Opioids may be taken orally in pill or liquid form, such as codeine or the many opioid-based prescription cough and diarrhea medications. They may be injected under the skin (skin popping), intramuscularly or intravenously. Injection has the added attraction of producing a “rush,” i.e., a relatively immediate drug reaction that has been described by users as being like a full body orgasm. Intravenous injection is said to produce the most intense rush. Given the expense and the frequent difficulty in obtaining opioids and the often low potency of street drugs, economy of delivery is often a consideration. Injection provides the least waste of drug in that the substance is introduced quickly into the bloodstream without previous evaporation or metabolism taking place. With higher potency heroin, however, smoking or “chasing the dragon” is often the choice. Smoking actually is the most rapid system for delivering opioids or any other drug to the brain, even faster than intravenous injection. Further, in light of AIDS, hepatitis C, and other illnesses that can be communicated by needle sharing, smoking is seen by users who can afford high-quality opioids as the safest use—and often seen by them as nonaddicting.

Physical dependence

Opioid users are subject to the classic symptoms of physical dependence. These are increasing tolerance and the onset of physical withdrawal symptoms. Tolerance involves needing more of the drug as time passes to achieve the same desired results. Physical withdrawal can initiate within hours of the last use and consists of a cluster of flu-like symptoms. Withdrawal is mediated by neural pathways separate from those involving the reward system, causing withdrawal events to be perceived as life-threatening, and subsequent physiological and

psychological reactions often lead to renewed opioid use. Withdrawal can be a tremendous force for continuing use, often at any cost (4).

Sedative–Hypnotic Drugs, Including Alcohol

Sedative–hypnotic drugs and anxiolytic drugs are CNS depressants that are used medically to reduce anxiety and/or induce sleep. They may also be used as anticonvulsants. Phenobarbital, for example, is often the maintenance drug of choice for seizure-prone individuals. In general, the sedative–hypnotic family of drugs includes alcohol, barbiturates, benzodiazepines, and such barbiturate-like drugs as chloral hydrate, glutethimide, meprobamate, and methaqualone.

The history of sedative–hypnotic drugs in medicine

The history of sedative–hypnotic drugs is one of attempts to find a drug or family of drugs that produces the desired effects without the risk of dependence and debilitating or life-threatening side effects and overdoses. In the 19th century, anxiety and insomnia were treated with opiates, bromide salts, chloral hydrate (developed in 1869), paraldehyde (developed in 1882), and alcohol. Each of these substances had its problems. The bromides could cause chronic bromide poisoning; many patients refused to take alcohol; and chloral hydrate and paraldehyde had objectionable taste and smell. As a result, the development of barbiturates was hailed as a major breakthrough.

Barbiturates

Barbiturates are all derived from barbituric acid, first obtained from uric acid and synthesized in Germany by Dr. Adolf van Baeyer in 1864. Conrad and Guthzeit synthesized the first barbiturate, 5,5-diethylbarbituric acid (barbital) in 1882. In 1903, Emile Fischer and Baron Josef von Mering introduced barbital into clinical medicine under the trade name Veronal[®]. Phenobarbital, which has remained the “model T” of barbiturates, first appeared on the market in 1912 as Luminal[®]. Unfortunately, intoxication with barbiturates is qualitatively similar to intoxication with alcohol and produced similar problems of abuse.

Benzodiazepines

A family of CNS depressants that has gained wide acceptance and use in the medical community is the benzodiazepines. These drugs, also called the minor tranquilizers, have been developed over the past 30 years, starting with chlordiazepoxide (Librium[®]), quickly followed by diazepam (Valium[®]). Since then, a variety

of benzodiazepines has been synthesized, including alprazolam (Xanax[®]) and triazolam (Halcion[®]). Benzodiazepines may differ in duration of effects and specific indications, but they are all cross-tolerant and chemically similar.

Major tranquilizers

Major tranquilizers like the phenothiazines, which include chlorpromazine (Thorazine[®]), are not usually subject to recreational-type abuse. Problems with these drugs most often involve misprescription or lack of understanding of their effects by health professionals. They are not considered to be addictive, and although a few deaths have been attributed to the ingestion of them at high doses, it is difficult to use them to commit suicide. The most notable problem with these drugs is the development of extrapyramidal symptoms, including facial and other abnormal movements of the head and neck, as well as such Parkinson syndrome-like symptoms as tremor at rest, rigidity, and shuffling walk.

Alcohol

Although its systemically administered medical uses have been limited to the treatment of methanol and ethylene glycol poisoning, alcohol is an excellent solvent and is used as a vehicle in many pharmaceutical formulations. It is also used topically as a disinfectant and to reduce fever through evaporation. Medieval alchemists considered it to be the “elixir of life,” a title that has survived in certain European fruit brandies called collectively *eau de vie*.

Although some cultures have expressly forbidden the use of alcohol (particularly some but not all Muslim cultures), most peoples have embraced this drug, giving themselves permission to use it ceremonially and recreationally, at least in moderate quantities. At the same time, alcoholism or alcohol addiction is considered to be a worldwide problem, and most cultures invoke sanctions against behavior related to alcohol overuse, such as drunk driving.

As a recreational substance, alcohol is second only to caffeine in worldwide use and second only to tobacco in health costs from abuse. In recent years, the American public has received a mixed message on alcohol's health benefits and deficits. Wine is said to help protect “moderate” drinkers from heart disease, but at the same time, alcoholism is responsible for more substance-related deaths than all other psychoactive drugs combined, with the exception of tobacco. Although there are few pharmacotherapies for alcoholism and alcohol abuse, a multimillion dollar project has been funded at the University of California at San Francisco by Gallo Wine

and the state of California to study alcohol neurochemistry and, if possible, to develop a cure for alcoholism. Remission from alcohol addiction was the aim of a fellowship developed in the mid-1930s, Alcoholics Anonymous, which today has a worldwide membership numbered in the millions.

Alcohol is usually imbibed in liquid forms such as beer, wine, brandy, and hard liquor, etc. The type of alcohol commonly consumed is known as “ethanol.” It is rapidly and efficiently absorbed into the bloodstream from the stomach, small intestine, and colon. Recent studies have suggested that women have a more efficient absorption than men. In the bloodstream, alcohol is distributed to all parts of the body, including the fetus(es) of pregnant women. Alcohol is metabolized in the liver and converted to acetaldehyde by the action of alcohol dehydrogenase (ADH) and other oxidizing agents at a relatively constant rate.

Adverse effects

The effects of sedative–hypnotic overdose or intoxication are similar for all drugs in this class. Ethanol acts as a classic sedative hypnotic drug, although the quality of sleep may be reduced by its ingestion. Intoxication works to decrease most mental and physical acuity, causing lapses in judgment, unsteady gait, slurred speech, slowed reactions, and mechanical difficulty. Blackouts, that is, continuing to function physically while being mentally disengaged, can occur as tissue dependence develops. Blackouts can be particularly dangerous in that users may forget how many pills they have taken and dose themselves into inadvertent overdoses. The degree of disinhibition euphoria can rapidly shift to dysphoria or even rage reactions with violent acting-out. In advanced stages, the intoxicated individual may pass out or, in extreme cases, lapse into a coma requiring emergency resuscitation. Acute intoxication to any sedative–hypnotic can be a life-threatening event.

Chronic abuse

The effects of chronic abuse can include memory impairment and chronic cognitive and psychomotor impairment. Tolerance develops to these drugs as the liver becomes more efficient in processing these drugs; however, the potential for a fatal overdose remains the same for these drugs. That means, as the sedative–hypnotic abuser needs and uses more of the drug, he or she comes closer and closer to a potential fatal overdose. Further, as a user gets older, age-dependent tolerance also occurs, in that the effect of a sedative–hypnotic on a 50-yr-old can be 5–10 times stronger than the same dose on a 20-yr-old.

Cross-tolerance and cross-dependence

Cross-tolerance means that tolerance to any sedative–hypnotic drug will extend to other drugs in the same class. Cross-dependency means that use of any drug in this class, or any opioid drug will enhance the effects and abusers may turn to other drugs in either category to either supplement their drugs of choice or stand in for them if they are not readily available.

Synergy and its dangers

Synergism can occur when more than one depressant drug, including alcohol, is used at the same time. In combination, that can cause a much greater reaction than the simple sum of effects. The liver tends to be choosy about what it metabolizes first. For example, diazepam is considered a relatively safe drug from the standpoint of being difficult to overdose on. However, if alcohol and diazepam (Valium) are taken together, the liver becomes busy metabolizing the alcohol, and the diazepam passes through to the brain at full strength. The result can be blackouts—resulting in even more use if the individual is medicating and forgets having already taken his or her medication, and extreme respiratory depression. These synergistic effects result in more than 4000 deaths a year and almost 50,000 emergency room visits for adverse multiple drug reactions (1).

How they work

Benzodiazepines, barbiturates, and alcohol act by stereospecifically binding to recently discovered receptors in the CNS. The effects of CNS-effective sedative–hypnotics have generally been linked to this complex, which also contains the receptor for γ -aminobutyric acid (GABA), the major inhibitory neurotransmitter in the brain and the chloride ion channel, through which chloride ions pass (5).

GABA receptors are the primary site of action for benzodiazepines in a highly complex process, but one that gives rise to the possibility of developing benzodiazepine agonists, antagonists, and inverse agonists. Benzodiazepine antagonists, such as RO15-1788 or flumazenil, may provide treatment options for both overdose and chronic abuse.

The Nature of Stimulant Drugs

Beyond the obvious of being stimulants rather than depressants, CNS stimulants have some basic differences from the two preceding groups of psychoactive substances. Although CNS stimulants can produce addiction, their users do not develop a steadily increasing tolerance.

Instead, as “presynaptic” drugs, they exhaust the brain’s own supply of stimulant neurotransmitters within the sympathetic system. This results in a binge pattern of use in which the user is forced to stop intense use periodically when the drugs no longer produce their desired effects so that the brain can replenish its supply of sympathetic transmitters.

The most common stimulant drugs are caffeine and nicotine, and their use is virtually worldwide. In recent years, the deleterious effects of nicotine have come under increasing scrutiny, particularly in the United States, where it has been recognized that tobacco is responsible for at least 400,000 deaths per year. At the same time that nicotine is being increasingly censured, caffeine is enjoying what seems to be an ever-increasing popularity. Perhaps this is because with increasing public health attention to the dangers of both tobacco and alcohol, it is the one remaining CNS drug that most people feel okay about using. Coffee shops have become the social centers of our society, and the market is increasingly dominated by chains that provide a wide variety of coffee products.

Caffeine

In general, coffee, tea, maté, Coca-Cola[®], and other sodas are so ubiquitous that people rarely think of them as drugs. Aside from individuals who are hypersensitive to caffeine, the controversy continues on whether caffeine itself is harmful or helpful to the people who use it. There is no doubt that caffeine is a CNS drug. It is well known that many individuals are physically dependent on its daily use and will exhibit withdrawal symptoms, including headache and disorientation, if their use is abruptly stopped. On the other hand, aside from the spiraling cost of cappuccinos, it may be hard to specify adverse consequences to the use of caffeine.

Writers such as Andrew Weil, M.D. (6), in his germinal book *The Natural Mind*, often contend that indigenous psychoactive substances used within cultural boundaries enhance rather than endanger the lives of their users. Plant stimulants, such as khat in Africa and the Near East and coca leaf-chewing in the South American highlands, are the most often cited as providing vitamins missing in the regional meager diets and needed stimulation for the hardscrabble existence of their users.

In teaching courses on drug abuse treatment to health professionals who may have had no personal experience with drug use and have a hard time understanding the compulsion involved, Richard Seymour asks how many are habitual coffee drinkers. Most hands usually go up. He then says, “Think about how you feel if you can’t get your first cup of coffee in the morning and then multiply that. That’s how the compulsive stimulant drug user feels.”

Other CNS stimulants

A drug similar in structure and effect to the amphetamines, methylphenidate (Ritalin[®]), and phenylpropanolamine, a stimulant vasoconstrictor that shows up in many cough and cold remedies, should also be mentioned. The most insidiously dangerous stimulant, nicotine as found in tobacco, is discussed briefly as well.

Cocaine

Cocaine is derived from the coca leaf, which has been chewed for its stimulant qualities by dwellers in the South American highlands since prehistoric times. When Spanish conquistadors first encountered the Inca Empire of Peru, coca leaves were a means of exchange controlled by the emperor himself. It is something of a miracle that coca leaf-chewing was not imported to Europe along with tobacco use at that time. And cocaine was not medically extracted from the leaf until 1860. Once the strong stimulant was isolated, however, it came into multiple use throughout Euro-American culture. Cocaine formed the original basis for coca-cola and could be found by itself or in combination with opium in a variety of quasimedical elixirs and tonics. Its use was recommended for the treatment of asthma, hay fever, fatigue, and at least a dozen other ailments. Sigmund Freud made frequent use of it, both personally and in his practice, and was involved in what may have been the first case of iatrogenic cocaine addiction. The most common use by serious abusers was by injection. Often cocaine was injected in a combination with morphine or heroin, called a “speedball.” By the early years of the 20th century, cocaine abuse had become serious enough in the United States for that drug to be included with heroin in the 1914 Harrison Narcotic Act.

Today, cocaine appears in several forms: coca leaf, liquid, powdered cocaine hydrochloride, purified freebase, and crack and can be chewed, insufflated, or snorted into the nose, injected (with or without opioids), or smoked.

Amphetamine and methamphetamine

Amphetamines are a 20th century development that first came into general medical use in the 1930s for a wide variety of medical conditions. During World War II, amphetamines were provided in large quantities to combat troops and bomber crews who had to stay awake and alert for long periods. After the war, production of these drugs remained high in most of the combatant countries, and they were readily prescribed by physicians for everything from depression to prefinals fatigue in college students. The first serious outbreaks of amphetamine abuse occurred in Japan, where stockpiles of the drug remained at the end of the

war, and in Sweden. Although some abuse had existed in the United States, the first postwar outbreak of stimulant abuse took the form of high-dose intravenous methamphetamine abuse between 1968 and 1969.

Stimulant drug pharmacology

Unlike opioid drugs, which work by imitating the indigenous morphines (endorphins) and attaching directly to the endorphin receptor sites, stimulant drugs produce their effects by acting as sympathomimetic agents and thereby stimulating the release of sympathetic neurotransmitters in the brain. The normal function of these sympathetic agents is to implement our “fight-or-flight” response by constricting blood vessels (vasoconstriction), increasing pulse rate and heart rate, increasing temperature (hyperthermia), and in general increasing alertness and response. These energy agents are also directly connected with the brain’s reward/pleasure center; thus, the satisfaction from using stimulants can be intense. One professional ball player who was introduced to cocaine at the height of his career said that the feeling from the drug was the same feeling he got when an entire stadium was on its feet shouting his name.

Nicotine

Tobacco, the primary source of nicotine, was used ceremonially in both pre- and post-Columbian America, imported to Europe where it was both embraced and reviled as a recreational drug, condemned by the court of James I of England, and today may be responsible for more than 400,000 deaths a year in the United States alone. Contrary to popular belief, although nicotine may help focus attention, it interferes with complex brain functions including access to long-term memory and the performing of multiple attention tasks.

Nicotine and the other ingredients in tobacco have been cited as causing a variety of fatal illnesses. A study by the Centers for Disease Control and Prevention (CDC) in 1991 listed the causes of death related to smoking with annual death toll as follows: (cardiovascular) heart disease, 150,000; stroke, 26,000; other, 24,000; (cancer) lung, 112,000; other, 31,000; (nonmalignant pulmonary disease) chronic obstructive pulmonary disease, 62,000; other, including pneumonia and influenza, 21,000; for a total of 426,000 fatalities a year directly attributable to tobacco. Further, the CDC points out that tobacco is also responsible for an annual death rate of 53,000 per year among nonsmokers affected by smoke in their immediate environment.

Pharmacologically, tolerance to tobacco develops quickly but once established, levels of smoking may remain about the same throughout one’s smoking career.

A withdrawal syndrome has been well established. Withdrawal symptoms may vary but can include craving for nicotine, irritability, frustration, anger, anxiety, depression, difficulty in concentrating, restlessness, and increased appetite. Although nicotine withdrawal is highly distressing and may continue for weeks, the compulsion to resume use may remain high for an extended period of time, and weight gain may be daunting, but it is not life threatening. However, detoxification can be an attenuated process of reversing neuronal adaptation to nicotine.

Although nicotine is also absorbed into the bloodstream through chewing and the use of snuff, inhaling cigarette smoke provides the most rapid brain access. Nicotine can also be absorbed through the skin, facilitating the use of skin patches. It is readily absorbed through the stomach, but first-pass digestion in the liver greatly decreases the amount reaching the brain from the stomach. Patients using nicotine gum are, therefore, now advised to mix the gum with saliva and lodge it between cheek and gum to facilitate absorption through the buccal mucosa.

Use of tobacco is bolstered by the positive reinforcement of producing euphoria and maintained by the negative of rapid-onset withdrawal symptoms as soon as nicotine levels decline below the brain's accustomed levels that are quickly relieved by the ingestion of nicotine.

The effects of amphetamine/methamphetamine

Stimulants promote the release of the brain's energy chemicals. On the short term, this can result in increased wakefulness and alertness, giving the occasional or situational user a performance edge. It was that edge that led science writers in the late 1940s to laud amphetamines as a wonder drug. Unfortunately, these drugs also deplete the available energy chemicals, induce a drug-based paranoia, and trigger intense cravings for more of the drug. Cocaine, in particular, blocks the reuptake of energy chemicals by the brain cells in which they are usually stored, creating a cerebral chain reaction until the chemicals are metabolized.

Prevention efforts

In keeping with their stimulant nature, cocaine and the stimulants produce a very rapid onset of abuse. Prevention efforts brought the slogan "Speed kills." Rock musician Frank Zappa filmed a TV commercial in which he said, "Kids, if you keep using speed you'll end up just like your parents."

The most effective amphetamine/methamphetamine prevention agent, however, proved to be the intravenous users themselves. These individuals tended to be walking, acting-out, negative advertisements of their drug. Anorexia

left them skeletal, whereas stimulant psychosis turned them into violent victims of delusional paranoia, a danger to themselves and others.

Although these efforts were effective to some extent, it became evident in the late 1960s that not only illicit manufacture of methamphetamine but the production and subsequent diversion of pharmaceutical psychoactive drugs were out of control, and the government and industry took steps to remedy the growing problem.

Hallucinogens

Although opioids and sedative-hypnotic drugs evolved primarily as medical substances for dealing with physical and psychic pain, and stimulants developed as recreational and performance-enhancing substances, hallucinogens had their role primarily within the realms of religion and magic. Throughout prehistory, history, and on into our own century, hallucinogenic substances have been used, often depending on the degree of sophistication of the culture in which they are being used, as a means of establishing contact with the spirit world, the realm of the gods, or the deeper reaches of the human subconscious. Shamans have used plant and mineral hallucinogens, often within the context of highly complex ritual to establish a point of contact between their people and their people's deities, or at least the supernatural forces that may affect their individual and collective lives.

Ethnobotanists have classified hundreds of plant hallucinogens, the majority of these originating in the rainforests of South America. In this article, however, the focus is on the five categories of hallucinogens classified by Goodman and Gilman (7):

1. Lysergic acid diethylamide (LSD)-like drugs, including mescaline, psilocybin, and psilocin;
2. Drugs that probably are LSD-like, such as DMA, DOM, and DMT;
3. Drugs that probably are LSD-like and have other properties, such as MDMA, MDA, and other amphetamine derivatives;
4. Drugs that probably are not LSD-like, such as 5-hydroxytryptophan; and
5. Drugs that are not LSD-like, such as scopolamine and δ -9-THC.

The history and nature of hallucinogens

The Goodman and Gilman classification obviously uses LSD as the base measure of hallucinogens. This LSD centrality most likely relates to the status of that drug as the most widely discussed and the most notorious of hallucinogens.

Although LSD was a relatively recent discovery, dating from 1943 when Dr. Albert Hofmann, a chemist at Sandoz Laboratories in Basel, Switzerland, accidentally ingested a small quantity of a substance he had first synthesized in 1938, its most active component, ergotomine, had a long history as a psychoactive agent. Occurring naturally as a rye-grain mold, ergotomine was featured in mystic potions in the classical world. In the Middle Ages, when its applied use had been forgotten, the hallucinogenic effects of ergotomine contamination in the bread supplies of entire communities was blamed on witchcraft and demonic possession.

Western scientific interest in hallucinogens was rekindled in the 19th century by poets and anthropologists observing and then participating in ceremonial rites involving psychoactive substances in a variety of cultures. Mescaline, the active ingredient in the peyote cactus used by religious sects in Mexico and the U.S. Southwest, was isolated in 1856 and by the turn of the century was available for research by the likes of Sigmund Freud, William James, and Havelock Ellis. It was, however, the discovery of what Dr. Hofmann considered the most powerful psychic drug (LSD) that induced tremendous scientific and popular interest in hallucinogen research.

With the advent of LSD availability, perception of these drugs underwent a process of evolutionary models. The first of these models was the psychotomimetic. This treated the drug experience as a form of psychosis, permitting researchers to study psychotic symptoms in nonpsychotic subjects. The psychotomimetic model was followed, although not necessarily superseded, by the hallucinogenic model, which treated LSD and mescaline as tools for studying the mechanisms of perception, and the therapeutic model, which involved the use of these drugs in the treatment of alcoholism, other forms of addiction, and mental health problems. Finally, there came the psychedelic model, which maintained that under proper conditions, the drug experience would be one of enlightening and productive consciousness expansion. It was with the psychedelic model that the use of LSD and other hallucinogens spread from the laboratory into the community.

Acute and chronic effects

The adverse effects of hallucinogens are generally divided between acute and chronic or long term. The acute effects, often referred to as “bad trips,” occur as direct negative results of hallucinogen ingestion and involve such elements as frightening images and thoughts, fear surrounding loss of control, and fear of losing one’s mind.

Acute effects: In 1967, David E. Smith, M.D., identified the adverse effects of hallucinogens as “largely psychological in nature,” and classified them as acute

toxicity, effects occurring during the use of the drug, or chronic after-effects (8). Although there have been some occurrences of physiological consequences, particularly with MDMA, these have been primarily of an idiosyncratic nature, although in most cases the adverse effects of these drugs still appear to be psychological in nature.

The acute toxic effects take many forms. Often individuals knowingly take a hallucinogenic drug and find themselves in a state of anxiety as the powerful hallucinogen begins to take effect. They were aware that they had taken a drug, but felt that they could not control its effects. This condition is similar to that of not being able to wake up from a threatening dream. Some users experience a bad trip and try to physically flee the situation, giving rise to potential physical danger. Others may become paranoid and suspicious of their companions or other individuals.

Not all acute toxicity is based on anxiety or loss of control. Some people taking hallucinogens display decided changes in cognition and demonstrate poor judgment. They may decide that they can fly, and jump out of a window. Some users are reported to have walked into the sea, feeling that they were “at one with the universe.” Such physical mishaps have been described within the acid culture as “being God, but tripping over the furniture.” Susceptibility to bad trips is not necessarily dose related but can depend on the experience, maturity and personality of the user, and “set and setting,” (i.e., the circumstances and the environment in which the trip takes place). Sometimes, the individual will complain of unpleasant symptoms while intoxicated and later speak in glowing terms of the experience. Negative psychological set and environmental setting are the most significant contributing factors to bad hallucinogenic trips (8).

Talkdowns of most acute toxicity reactions can be accomplished without medication or hospitalization. Paraprofessionals with psychedelic drug experience have been particularly effective at sites such as large rock concerts. In the talkdown approach, one should maintain a relaxed, conversational tone aimed at putting the individual at ease. Quick movements should be avoided. One should make the patient comfortable but not impede their freedom of movement. Let them walk around, stand, sit, or lie down. At times, such physical movement and activity may be enough to break the anxiety reaction. Gentle suggestion should be used to divert patients from any activity that seems to be adding to their agitation. Getting the individual’s mind off the frightening elements of a bad trip and onto positive elements is the key to the talkdown.

An understanding of the phases generally experienced in a hallucinogenic drug trip is most helpful, in treating acute reactions. After orally ingesting an average dose of

100–250 µg of LSD, the user experiences sympathomimetic, or stimulant responses, including elevated heart rate and respiration. Adverse reactions in this phase are primarily managed by reassurances that these are normal and expected effects of psychedelic drugs. This reassurance is usually sufficient to override a potentially frightening situation.

From the first to the sixth hour, visual imagery becomes vivid and may take on frightening content. The patient may have forgotten taking the drug, and given acute time distortion, may believe this effect will go on forever. Such fears can be dispelled by reminding the individual that these effects are drug-induced by suggesting alternative images and by distracting the individual from those images that are frightening.

In the later stages, philosophical insights and ideas predominate. Adverse experiences here are most frequently attributable to recurring unpleasant thoughts or feelings that can become overwhelming in their impact. The therapist can be most effective by being supportive and by suggesting new trains of thought.

The therapist's attitude toward hallucinogens and their use is very important. Empathy and self-confidence are essential. Anxiety and fear in the therapist will be perceived in an amplified manner by the client. Physical contact with the individual is often reassuring, but can be misinterpreted. Ideally, the therapist should rely on intuition rather than on preconceptions.

Wesson and Smith (9) noted that medication may be necessary and should be given either after the talkdown has failed or as a supplement to the talkdown process. During the first phase of intervention, oral administration of a sedative, such as 25 mg of chlorthalidopoxide (Librium) or 10 mg of diazepam (Valium), can have an important pharmacological and re-assuring effect.

During the second and third phases, a toxic psychosis or major break with reality may occur in which one can no longer communicate with the individual. If the individual begins acting in such a way as to be an immediate danger, antipsychotic drugs may be used. Only if the individual refuses oral medication and is out of behavioral control should antipsychotics be administered by injection. Haloperidol (Haldol) (2.0–4.0 mg administered intramuscularly every hour) is the current drug of choice. Any medication, however, should only be given by qualified personnel. If antipsychotic drugs are required, hospitalization is usually indicated. It has been found at the Haight Ashbury Free Clinics, however, that most bad acid trips can be handled on an outpatient basis by talkdown alone.

As soon as rapport and verbal contact are established, further medication is generally unnecessary. Occasionally, an individual fails to respond to the above regimen and

must be referred to an inpatient psychiatric facility. Such a decision must be weighed carefully, however, because transfer to a hospital itself may have an aggravating and threatening effect. Hospitalization should only be used as a last resort.

Chronic hallucinogenic drug aftereffects: These present situations wherein a condition that may be attributable to the ingestion of a toxic substance occurs or continues long after the metabolization of that substance. With the use of hallucinogens, four recognized chronic reactions have been reported: 1) prolonged psychotic reactions; 2) depression sufficiently severe so as to be life threatening; 3) flashbacks; and 4) exacerbation of pre-existing psychiatric illness. Recently, a fifth chronic reaction has been listed in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, posthallucinogen perceptual disorder (PHPD).

Some people who have taken many hallucinogenic drug trips, especially those who have had acute toxic reactions, show what appear to be serious long-term personality disruptions. These prolonged psychotic reactions have similarities to schizophrenic reactions and appear to occur most often in people with pre-existing psychological difficulties, such as primarily prepsychotic or psychotic personalities. Hallucinogenic drug-induced personality disorganization can be quite severe and prolonged. Appropriate treatment often requires antipsychotic medication and residential care in a mental health facility, followed by outpatient counseling.

At the Haight Ashbury Free Clinics, it has been noted that some of the clients self-medicated their hallucinogenic-precipitated psychotic episodes with amphetamines (10). Often, this self-medication with amphetamines resulted in the development of amphetamine abuse, followed by secondary heroin, barbiturate, or alcohol abuse patterns, to ameliorate the side effects of the amphetamines. Thus, in certain patients, chronic psychological problems induced by LSD and other hallucinogenic drugs led to complicated patterns of polydrug abuse that required additional treatment approaches (9).

Flashbacks: By far the most ubiquitous chronic reaction to hallucinogens is the flashback. Flashbacks are transient spontaneous occurrences of some aspect of the hallucinogenic drug effect after a period of normalcy that follows the original intoxication. This period of normalcy distinguishes flashbacks from prolonged psychotic reactions. Flashbacks may occur after a single ingestion of a psychedelic drug, but more commonly occur after multiple psychedelic drug ingestion.

Flashbacks are a symptom, not a specific disease entity. They may well have multiple causes, and many cases called flashbacks may have occurred although the

individual had never ingested a psychedelic drug. Some investigators have suggested that flashbacks may be attributable to a residue of the drug retained in the body and released into the brain at a later time. Although this is known to happen with phencyclidine (PCP) and drugs similar to it, there is no direct evidence of retention or prolonged storage of such psychedelics as LSD.

Individuals who have used psychedelic drugs several times a month have indicated that fleeting flashes of light and afterimage prolongation occurring in the periphery of vision commonly occur for days or weeks after ingestion. Active and chronic psychedelic drug users tend to accept these occurrences as part of the psychedelic experience, are unlikely to seek medical or psychiatric treatment, and frequently view them as "free trips." It is the inexperienced user and the individual who attaches a negative interpretation to these visual phenomena who are likely to be disturbed by them and seek medical or psychiatric help. Although emotional reactions to the flashback are generally contained within the period of the flashback itself, prolonged anxiety states or psychotic breaks have occurred after a frightening flashback. There is no record of flashback activity specifically attributable to hallucinogenic drug use occurring more than a year after the individual's last use of a psychedelic drug (10).

Chronic consequences of hallucinogen use: The long-term study of adverse hallucinogenic drug reactions has revealed the existence of low prevalence, but quite disabling chronic consequences of LSD use. Of particular concern is PHPD. With PHPD, individuals describe a persistent perceptual disorder that they describe as being like living in a bubble underwater. They also describe trails of light and images after moving their hands, and they often describe living in a purple haze. This perceptual disorder is aggravated by any psychoactive drug use, including alcohol and marijuana, and is distinguished from flashbacks, which are episodic rather than chronic phenomena. With PHPD, the individual often experiences anxiety, even panic, and becomes phobic and depressed. With PHPD sufferers, our experience has been that individuals do not have a disturbed psychiatric history before the onset of psychedelic drug use and that PHPD can occur even after a single dose.

With more severe, prolonged hallucinogen reactions, such as an LSD-precipitated schizophrenic reaction or severe depressive disorder, individuals almost always have a premorbid psychiatric history and require inpatient treatment. With the prolonged psychotic reactions, antipsychotic medication is required, and with the

prolonged depressive reactions, antidepressant medication is required. A major concern involves teenagers with depressive reactions to psychedelic drug use that may result in severe depression culminating in suicide.

With PHPD, drug-free recovery with supportive counseling is often adequate treatment, although recovery may take several months, and anti-anxiety medication may be needed to treat the secondary anxiety and panic disorder that develops when the individuals feel that they are irreversibly brain-damaged and will never see normally again.

Other concerns with hallucinogen abuse: There are many variations on the conditions addressed above, particularly with PCP, a mind-body disassociative drug that can act as a stimulant, a depressant, and a hallucinogen, depending on the dosage. These are discussed in another article on treatment and assessment. Suffice it to say, the etiology and pharmacology of hallucinogenic drugs is varied and involves a number of differing symptoms and sequelae.

REFERENCES

1. Inaba, D.; Cohen, W. *Uppers, Downers, All Arounders*; CNS Publications, Inc.: Ashland Oregon, 2000; 61–63.
2. Seymour, R.; Smith, D. *Drugfree*; Facts on File: New York, 1987.
3. Smith, D.; Seymour, R. *The Clinician's Guide to Substance Abuse*; McGraw Hill: New York, 2000.
4. Gold, M.; Herkov, M. The Pharmacology of Cocaine, Crack, and Other Stimulants. *Principles of Addiction Medicine*, Graham, A., Schultz, T., Wilford, B., Eds.; 2nd Ed.; American Society of Addiction Medicine: Chevy Chase Maryland, 1998; 137.
5. Juergens, S.M.; Cowley, D. The Pharmacology of Sedative-Hypnotics. *Principles of Addiction Medicine*, Graham, A., Schultz, T., Wilford, B., Eds.; 2nd Ed.; American Society of Addiction Medicine: Chevy Chase Maryland, 1998; 117.
6. Weil, A. *The Natural Mind*; Houghton Mifflin Company: Boston, 1972.
7. Goodman, L.; Gilman, A. *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 8th Ed.; Macmillan: New York, 1990.
8. Smith, D.; Seymour, R. Dream Becomes Nightmare: Adverse Reactions to LSD. *J. Psychoactive Drugs* **1985**, *17* (4), 297–303.
9. Wesson, D.; Smith, D. Psychedelics. *Treatment Aspects of Drug Dependence*; Schecter, A., Ed.; CRC Press: West Palm Beach, 1978.
10. Seymour, R.; Smith, D. *The Psychedelic Resurgence: Treatment, Support, and Recovery Options*; Hazelden: Center City, Minnesota, 1993.

ADVERSE DRUG REACTIONS

Therese I. Poirier

Robert L. Maher, Jr.

Duquesne University, Pittsburgh, Pennsylvania, U.S.A.

INTRODUCTION

Adverse drug reactions (ADRs) are types of adverse drug events (ADEs) (1). ADEs include ADRs, medication errors, and other drug-related problems. ADEs are the negative consequences of drug misadventures. Henri Manasse defined drug misadventure as the iatrogenic hazard that is an inherent risk when drug therapy is indicated. This chapter will focus on ADRs.

DEFINITIONS

The World Health Organization's (WHO) and Karch and Lasagna's definitions of an ADR are quite similar. An ADR is any response to a drug that is noxious and unintended, and occurs at doses used for prophylaxis, diagnosis, or therapy, excluding failure to accomplish the intended purpose (2). The Food and Drug Administration (FDA) focuses on ADRs that have unexpected reactions and/or those of more significant morbidity. These ADRs would include those where the patient outcome is death, life-threatening, hospitalization, disability, congenital anomaly, or required intervention to prevent permanent impairment or damage (3). The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) is concerned with the reporting of significant ADRs. Those that result in morbidity, require additional treatment, require an increased length of stay, temporarily or permanently cause disability, or cause death must be reported to the FDA (4). The American Society of Health-System Pharmacists (ASHP) defines significant ADRs as any unexpected, unintended, undesired, or excessive response to a drug that includes the following:

- Requires discontinuing the drug
- Requires changing the drug therapy
- Requires modifying the dose
- Necessitates admission to the hospital
- Prolongs stay in a health care facility
- Necessitates supportive treatment
- Significantly complicates diagnosis

- Negatively affects prognosis or results in temporary or permanent harm, disability, or death (5)

The ASHP definition does not include reactions due to drug withdrawal, drug abuse, poisoning, or drug complications.

Other terms that may be included as ADRs are side effects, drug intolerance, idiosyncratic reactions, toxic reactions, allergic reactions, or hypersensitivity reactions (6). *Side effects* are reactions that are unintended and unwanted but are known pharmacologic effects of the drug and occur with predictable frequency. *Drug intolerance* is a mild reaction to a drug that results in little or no change in patient management. *Idiosyncratic reaction* is an unexpected response that occurs with usual dose of a drug. *Toxic reaction* is a predictable response that results from greater than recommended drug dosages or drug concentration in the body. *Allergic or hypersensitivity reaction* is an unusual sensitivity to a drug of an immunologic nature.

CLASSIFICATION SYSTEMS

Four classification systems are used to describe ADRs (1, 7). ADRs can be classified according to the pharmacologic effect of the drug—Type A, B, C, and D reactions. Type A reactions are exaggerated but normal pharmacologic actions of a drug. They are predictable and dose dependent. Type B reactions are not predictable given the known pharmacologic action of a drug and are not dose related. Many of these Type B reactions are hypersensitivity or immune-based. These reactions can be further subdivided into type I (IgE-mediated reaction), II (IgG or IgM-mediated cytotoxic reaction), III (IgG-mediated immune complex reactions), and IV (cell-mediated immune reaction). Type C reactions are those due to long-term use of a drug. Type D reactions are delayed drug effects, such as due to carcinogenicity or teratogenicity.

ADRs can also be classified according to the dose relationship, i.e., dose-related and non-dose-related reactions. Another classification system is based on the causal relationship between the reaction and the drug. One of the

most widely used causality classifications is based on Naranjo's descriptions. These categories include definite (drug is likely the true cause), probable (drug is the apparent cause), possible (drug appears to be associated), and remote (drug is not likely to be the cause). The fourth classification system is based on degree of injury or severity of reaction. There are mild reactions (temporary discomfort and tolerable), moderate (significant discomfort), and severe (potentially life threatening or causing permanent disability or death).

INCIDENCE

The frequency of ADRs in the general population is unknown. However, the reported rates of new occurrences for ADRs are noted for selected patient populations. A meta-analysis of 39 prospective studies reported an overall incidence of serious ADRs in hospitalized patients of 6.7% and of fatal ADRs of 0.32% (8). The fatality rate makes ADRs the fourth to sixth leading cause of death in the United States. Another meta-analysis of 36 studies indicated that approximately 5% of hospital admissions are due to ADRs (9). The costs of ADRs are estimated to be \$1.56–\$4 billion in direct hospital costs per year in the United States (10).

FACTORS PREDISPOSING TO ADRS

Two major factors predispose to adverse drug reactions: the drug itself and patient factors. Factors related to the drug include its dose, dosage form and delivery system, and interactions between drugs. Patient-related factors include age, disease states, genetics, gender, nutrition, multidrug therapy use, and use of herbal therapies.

Drug-Related Factors

Dose

ADRs may be the result of ingestion of increased amounts of a drug. Dosing issues are especially likely with narrow therapeutic index drugs. Examples of these types of drugs include digoxin, anticoagulants, anticonvulsants, antiarrhythmics, antineoplastic agents, bronchodilators, sedatives, and hypnotics (11).

Dosage form and delivery system

Many of the ADRs related to the dosage form and delivery system are the result of local irritation or

hypersensitivity reactions (12). Local irritation to the gastrointestinal (GI) tract can occur with oral dosages. For example, toxicity resulting in mouth ulcerations is associated with antineoplastic drugs. In addition, the use of certain formulations, such as sustained release preparations, can increase esophageal injury if esophageal transit is delayed. For example, a controlled release wax matrix of potassium chloride has been associated with significant esophageal erosions. Factors identified to predispose to esophageal injury include large film-coated tablets, capsules, large sustained-release preparations, rapidly dissolving formulations, and ingestion of solid oral dosage forms before bed rest with very little water intake (12).

Localized tissue irritation can be seen from the intramuscular (IM) route. This is especially an issue when the formulation pH differs from the pH of the surrounding tissue or when precipitation of poorly soluble drugs occurs (12). Incorrect administration of IM injections is probably the most important factor that causes local adverse effects. Local skin irritation can also be seen with transdermal delivery systems due to the alcohols, nonionic surfactants, and adhesives.

Hypersensitivity reactions can occur due to the presence of contaminants or excipients in pharmaceutical dosage forms (e.g., outbreaks of eosinophilia-myalgia syndrome associated with oral tryptophan contaminants in various drugs) (12). Another example is the anaphylactoid reactions to the surfactant Cremaophor EL, which is used in paclitaxel (Taxol).

Direct toxicity effects related to use of preservatives also has been documented. For example, severe metabolic acidosis and death in infants was attributed to the presence of benzyl alcohol, a preservative used in bacterostatic normal saline that was used to flush catheters (12).

The use of specific intravenous (IV) delivery devices also can cause ADRs. For instance, use of plastic infusion sets for IV administration of nitroglycerin has resulted in subtherapeutic effects due to diffusion of the drug into the plastic tubes (12).

Formulation effects, such as bioavailability differences, can cause ADRs when patients are switched to generic products. For example, significant adverse effects have occurred with anticonvulsants and thyroid preparations (12).

Interactions between drugs

It has been estimated that 6.9% of ADRs are due to drug–drug interactions (6). The most likely reason for an adverse drug interaction is the pharmacokinetic changes that result in altered metabolism or excretion of drugs, or the

pharmacodynamic changes that result in synergistic or additive effects of drugs.

Patient-Related Factors

Age, disease states, genetics, gender, nutrition, multidrug therapy use, and herbal therapies use are patient-related factors that influence the likelihood of adverse drug reactions.

Age—geriatrics

Age-related alterations in pharmacokinetics and pharmacodynamics may affect the response of elderly patients to certain medications, and may increase the susceptibility for ADRs among elderly patients (13–15) (Table 1). The risk of ADRs among elderly patients is probably not due to age alone. ADRs may be related more to the degree of frailty and medical conditions of the patient (15). On average, older persons have five or more coexisting diseases that may increase the risk of adverse events. Polypharmacy seems to be more of a common problem among the elderly. The average elderly patient takes 4.5 chronic medications and fills 13 prescriptions yearly (15). Elderly patients appear to have a decline in homeostatic mechanisms. The imbalance of homeostatic mechanisms and the decline in function reserves may put a patient at greater risk for ADEs due to decreased tolerance of medications and the ability to handle stressful situations (16).

Age—pediatrics

The two factors responsible for increasing risks of ADRs in children are pharmacokinetic changes and dose delivery issues. Age-related differences in pharmacokinetics in children are documented (17). However, the data on both efficacy and safety are often limited or not studied at all in this population. Thus, it is unclear

whether an increased risk for ADRs exists in this group. However, there is a potential risk for increased ADRs if appropriate considerations are not taken into account in view of pharmacokinetic changes (18).

It is important to note that only one-fourth of the drugs approved by the FDA have indications specific for use in a pediatric population (17). Medications used in adults are often given to children without FDA safety and efficacy data. Compatibility and stability issues with dosage forms intended for adults that have been altered (e.g., dilution or reformulation) can increase risks for ADRs.

Information on pediatric age-related difference in neonates, children, and adolescents may aid in prevention of pediatric ADRs (18) (Table 2). Further studies of drug use in pediatrics are needed in order to prevent ADRs.

Concurrent diseases

Diseases such as hepatic or renal diseases can influence the incidence of ADRs by altering the pharmacokinetics of drugs, such as absorption, distribution, metabolism, or excretion (6).

Hepatic disease

Patients with liver disease have an increased susceptibility to certain drugs due to decreased hepatic clearance for drugs metabolized by the liver or due to enhanced sensitivity (6). For example, impaired hepatic metabolism can precipitate central nervous system (CNS) toxicity in patients on theophylline, phenytoin, or lidocaine; or ergot poisoning on ergotamine (19).

Increased sensitivity to drugs is also encountered in liver disease(19). The use of anticoagulants increases the risk of bleeding due to the reduced absorption of vitamin K or decreased production of vitamin K-dependent clotting factors. There is an enhanced risk for respiratory depression and hepatic encephalopathy due to morphine

Table 1 Geriatric age-related changes in pharmacokinetics

Pharmacokinetic phase	Pharmacokinetic parameters
Gastrointestinal absorption	Unchanged passive diffusion and no change in bioavailability for most drugs ↓ Active transport and ↑ bioavailability for some drugs ↓ First-pass effect and ↑ bioavailability
Distribution	↓ Volume of distribution and ↑ concentration of water soluble drugs ↑ Volume of distribution and ↑ half-life for fat soluble drugs ↑ or ↓ free fraction of highly plasma protein-bound drugs ↓ Clearance and ↑ half-life for some Phase I
Oxidation drugs	↓ Clearance and ↑ half-life of drugs with high extraction ratio
Renal excretion	↓ Clearance and ↑ half-life of renally eliminated drugs

↓ = Decreased; ↑ = Increased.

Table 2 Pediatric age-related risk factors and causes of ADRs

<i>Neonates:</i>
Placental transfer of drug before birth
Differing drug action
Altered pharmacokinetics
Increased percutaneous absorption
Decreased renal/hepatic function
Decreased plasma protein binding
Use of multiple drugs
Limited information on drug action in critically ill and premature neonates
<i>Children:</i>
Paradoxical effect of medications (excitability rather than sedation from antihistamines)
Excipients of liquid dosage forms
Sugar as sweeteners
Propylene glycol as solvent
Large volume intravenous solutions
Treatment of viral infections with antibiotics
Disruption of neurologic and somatic development
<i>Adolescents:</i>
Autonomy seeking
Use and misuse of devices (e.g., tampons)
Use and misuse of prescription and nonprescription medications
Poor compliance with instructions
Use of multiple medications
Recreational use of alcohol and illicit drugs
Effects of changing hormone levels on drugs

(From Ref. 7.)

or barbiturates in patients with severe liver disease. Vigorous use of diuretics can precipitate hepatic coma due to potassium loss in liver disease. There is an increased risk of hypoglycemia with sulphonylurea antidiabetic drugs due to decreased glycogenesis in liver disease.

Liver disease can also cause hypoalbuminemia due to decreased liver synthesis of albumin. For drugs that are extensively bound to albumin, such as phenytoin, an enhanced risk of drug toxicity could occur because of the increase in free drug concentration.

There are no useful methods to quantify the degree of liver disease that can assist in dosage adjustment. A practical approach involves checking patients for elevated prothrombin time, rising bilirubin levels, and/or falling albumin levels. In such instances, drugs that have an altered response in liver disease or cause hepatotoxicity need to be avoided.

Renal disease

Impaired renal function increases the incidence of ADRs for drugs that depend on the kidney for their elimination.

Unlike liver disease, use of pharmacokinetic dosing principles can minimize the risk for adverse effects.

Mechanisms responsible for enhanced ADRs in renal disease include delayed drug excretion, decreased protein binding due to hypoalbuminemia, and increased drug sensitivity (6). Delayed renal excretion is responsible for enhanced toxicity with drugs such as aminoglycosides, digoxin, vancomycin, chlorpropamide, H2-antagonists, allopurinol, lithium, insulin, and methotrexate (20). For some drugs, the accumulation of a toxic metabolite during renal failure is responsible for ADRs. This is the case with meperidine, where a toxic metabolite, normeperidine, accumulates in renal failure (20).

Patients with accumulation of uremic toxins have increased sensitivity to certain drugs. There may be an enhanced response to CNS depressants (such as barbiturates and benzodiazepines), hemorrhagic effects from aspirin or warfarin, and other bleeding effects from antibiotics that inhibit platelet aggregation, such as carbenicillin, ticarcillin, and piperacillin.

Other diseases

On theoretical grounds, other diseases associated with hypoalbuminemia could predispose patients to adverse reactions and to altered responses to drugs that are highly protein bound (21) (Table 3).

The presence of other diseases can influence the risk for ADRs. Many of these adverse effects are related to an extension of the pharmacologic effects of the drug in the presence of certain pathophysiology. Numerous examples are given in Table 4 (6).

Patients who have had a previous reaction to drugs are also more likely to experience an ADR (22). Patients with history of allergic diseases also have an increased risk due to a genetically related ability to form immunoglobulin E.

Genetic factors

Genetic factors account for some ADRs due to either altered pharmacokinetics or by altering tissue responsiveness. Altered metabolism of drugs occurs due to

Table 3 Conditions associated with hypoalbuminemia

Aging	Liver disease
Burns	Nephrotic syndrome
Cancer	Nutritional deficiency
Cardiac failure	Pregnancy
Protein-losing enteropathy	Renal failure
Inflammatory diseases	Sepsis
Injury	Stress
Immobilization	Surgery

Table 4 Influence of diseases on adverse drug reactions

Disease	Drug	Adverse reactions
Gastrointestinal		
Peptic ulcer	Aspirin, corticosteroids, nonsteroidal antiinflammatory drugs	Risk of bleeding or perforation of ulcer
Cardiovascular		
Heart failure	β -Blockers Lidocaine, theophylline	Aggravate or precipitate heart failure Enhanced toxicity—seizures
Myocardial ischemia	Tricyclic antidepressants Digoxin	Disturbances of cardiac rate, rhythm, and conduction Arrhythmias
Bradycardia	β -Blockers Quinidine	Cardiac standstill
Hypertension	Oral contraceptives, vasoconstrictors Phenothiazines, nitrates Tricyclic antidepressants	Increased blood pressure Decreased blood pressure
Hematologic		
Bleeding disorders—hemophilia	Aspirin	Increased risk of hemorrhage
Neurological disorders		
Myasthenia gravis	Aminoglycosides Quinidine, quinine	Aggravate muscle weakness Paralysis
Epilepsy	Phenothiazines Tricyclic antidepressants	Lower seizure threshold
Cerebrovascular	Ergotamine	Ischemic episodes
Rheumatic		
Systemic lupus	Drugs	Increased incidence of drug reactions in general
Hyperuricemia	Thiazide diuretics, furosemide	Gouty attack
Respiratory		
Asthma	β -Blockers	Acute bronchospasms
Respiratory insufficiency	Narcotic analgesics	Hypoventilation, respiratory arrest
Endocrine disorders		
Diabetes mellitus	Thiazide diuretics, furosemide, corticosteroids, oral contraceptives	Hyperglycemia; aggravates diabetic control
Hypothyroidism	Digoxin	Enhanced response
Hyperthyroidism	Oral anticoagulants Digoxin	Enhanced response Decreased response
Ocular		
Narrow-angle glaucoma	Anticholinergics	Glaucoma attack

differences in hydrolysis, acetylation, and hepatic oxidation of drugs. Altered pharmacodynamic reactions could be either an exaggerated response or a qualitative response. These types of reactions are unpredictable. Examples of altered drug response due to genetic factors are found in Table 5 (6).

Gender

A higher incidence of ADRs has been reported for women in comparison to men (6). One reason for this observation is that women take more drugs than men. Yet, no sex-linked differences in drug pharmacokinetics have been documented. Other reports have not supported a higher incidence of ADRs in women as compared to men. Thus, sex alone is unlikely to be a major determinant of ADRs.

Nutrition

Nutritional factors are also responsible for ADRs. These factors include the interaction of drugs and nutrients, and altered pharmacokinetics related to nutritional status.

One study reported a very low incidence (0.4%) of clinically significant drug–nutrient interactions in a teaching hospital (23). Three mechanisms postulated for drug–nutrient interactions are interference with drug absorption, alteration of drug excretion, and affecting drug activity. For example, the absorption of tetracycline is reduced by chelation with iron, calcium, and magnesium. Foods that acidify or alkalinize the urine can affect drug excretion. Foods that contain a large amount of vitamin K can inhibit the activity of warfarin. A listing of important drug–nutrient interactions is found in Table 6 (23). A review article on drug–food interactions in clinical practice is found in Ref. 24.

Drug–nutrient interactions may be more highly significant in renal failure patients. A review article of drug–nutrient interactions in renal failure has been published (25).

Nutritional status can affect drug pharmacokinetics. Malnutrition states can cause the following: 1) the liver and kidneys changes affect drug elimination; 2) GI system changes affect drug absorption; 3) changes in the heart affect blood flow; 4) hormone changes affect metabolic enzymes and drug binding proteins; 5) plasma, tissue proteins, and body composition changes affect protein binding and elimination; 6) mineral and electrolyte changes affect drug metabolism and protein binding; and 7) tissue changes affect uptake of drugs and drug–receptor interactions (26).

Multidrug use

According to several epidemiological studies, multiple drug use has a strong association in the causality of ADRs.

It has been suggested that the more medications used, the higher the risk for ADRs (27). Consistent drug regimen reviews by healthcare providers in order to reduce polypharmacy may decrease the risk of ADRs.

Herbal therapies use

The use of herbal therapies increased dramatically during the 1990s. Herbal therapy sales are estimated to be \$4 billion a year, with sales increasing at 20% per year since the early 1990s (28). Patients often mistakenly believe that since these products are natural, they do not possess the potential harm as in prescription medications. Since herbal medications are sold and marketed without stringent FDA approval and guidelines, limited evidence-based data on efficacy, adverse effects, and drug interactions exist. Recently, two review articles examined available data on ADRs for the most common herbal medications (28, 29). Many of these available reports fall short on documentation of temporal relationship with the specific ADR and the herbal drug.

For most conditions, herbal products are not a replacement for proven prescription or nonprescription drugs. Patients should be aware that health care practitioners cannot guarantee the safety and consistency of herbal products. Patients should start with the recommended effective doses and report any unusual side effects to their health care practitioner. Patients should always consult with their pharmacist for possible drug–herbal interactions. Side effects and possible drug interactions for the ten most commonly used herbals are listed in Table 7.

ADVERSE DRUG REACTION REPORTING SYSTEMS

The WHO, the FDA, the JCAHO, and the Health Care Financing Administration (HCFA) have all addressed and mandated the need for health care institutions to implement an ADE detection and reporting system. Detection systems are instrumental in postmarketing surveillance of ADRs. The JCAHO requires all accredited health care institutions to have an ongoing drug surveillance program (4). The goals of ADR detecting and reporting systems are to aid in postmarketing surveillance of FDA approved medications and to identify ways to decrease ADR risks. The main focus of all of these reporting systems is to aid in promoting improvements in the medication use process.

Table 5 Genetic factors and altered drug responses

Genetic mechanism	Drug(s)	Adverse drug response
Pharmacokinetic		
Low plasma pseudocholinesterase	Succinylcholine	Prolonged neuromuscular blockade leading to apnea
Slow acetylator	Isoniazid	Increased incidence of peripheral neuropathy; SLE-like syndrome; and more prone to phenytoin toxicity
	Hydralazine, procainamide	Increased incidence of SLE-like syndrome
	Phenelzine, sulfasalazine	More prone to side effects
Rapid acetylator	Isoniazid	More prone to hepatitis
Deficiency of epoxide hydrolase	Phenytoin, carbamazepine, phenobarbital	Life threatening hypersensitivity syndrome due to accumulation of toxic intermediates
Pharmacodynamic		
Glucose 6-phosphate dehydrogenase deficiency (G-6-PD)	Aspirin, BAL (dimercaprol), chloroquine, chloramphenicol, dapsone hydroxychloroquine, nalidixic acid, nitrofurantoin, primaquine, probenecid, quinine, quinidine, sulfonamides	Hemolytic anemia
Methemoglobin reductase deficiency	Acetaminophen, anesthetics, topical, benzocaine, chloroquine, dapsone, nitrites, primaquine, sulfonamides	Methemoglobinemia
Abnormality of calcium regulation	Anesthetics, general, (halothane), muscle relaxants (succinylcholine)	Malignant hyperpyrexia

Table 6 Important drug-nutrient interactions

Drug	Nutrient	Interaction
Phenytoin	Alcohol	Enhanced metabolism of phenytoin
	Enteral feedings	Decreased phenytoin absorption
Tetracycline	Dairy products	Impaired drug absorption
Theophylline	Caffeine	Potential for toxic effects
Warfarin	Foods high in vitamin K	Decreases anticoagulant response
Chlorpropamide, tolbutamide, tolazamide, acetohexamide, metronidazole	Alcohol	Disulfiram-like reaction
Trancylcypromide	Foods high in tyramine	Hypertensive crisis
Disulfiram	Alcohol	Nausea, blurred vision, chest pain, dizziness, fainting
Spirolactone	Foods high in potassium	Hyperkalemia

(Adapted from Ref. 23.)

ADR Screening Methods

The best methodology for screening for ADRs has not been determined. However, several screening methods have been proposed. In particular, the literature has highlighted five screening methods using clinical data (30–34). The five include screening for: 1) “tracer drugs,” e.g., antidotes such as vitamin K and diphenhydramine; 2) “narrow therapeutic range drugs,” e.g., follow-up of computer lab values for warfarin and digoxin; 3) change in medications, e.g., documentation of discontinued medications or decreased dose; 4) diagnosed ADRs documented in the medical record, e.g., chart review or reviewing ICD-9 CM (International Classification of Diseases, Ninth Revision, Clinical Modification) codes; and 5) ADR computer report tracking systems. Although each of these ADR screening methods has been described in detail, limited data are available on the productivity of these screens.

Systems for Pharmaco-epidemiologic Studies

Pharmacoepidemiology is used to detect ADRs (35, 36). Several types of systems use pharmacoepidemiologic methods. These include spontaneous reporting, studies of therapeutic classes, and studies of specific medical syndromes.

Spontaneous reporting

Spontaneous reporting is currently the major backbone for the detection of ADRs (37). It occurs in one of three ways:

1. Reporting to the FDA as part of clinical trials;

2. Reporting by practitioners to medical journals; or
3. Patients’ self-reporting to either manufacturers or the FDA (38).

Clinical trials in new drug development cannot detect all the possibilities for drug safety. Limitations in Phase III clinical trials include a relatively small sample size, short duration of the trial, restricted populations (e.g., geriatrics and pediatrics), uncomplicated patients, (e.g., limited disease states), and limited power for adverse drug reaction detection (30). Thus, the FDA relies heavily on spontaneous reporting of suspected ADRs (39). Spontaneous reporting is important in early market history of the drug to determine previously unidentified drug reactions. This has been particularly true in the last few years because of numerous new medications that have entered the market and now carry a black box warning. For example, Rezulin[®] and Trovan[®] are associated with hepatotoxicity and carry black box warnings.

Additional advantages of spontaneous reporting systems include the detection of extremely rare ADRs and ability to identify at-risk subgroups. In order to enhance the spontaneous reporting system approach, the FDA developed the MedWatch form. This form can be faxed to the agency (1-800-FDA-1078) or called in (1-800-FDA-1088) (40). The forms also can be obtained by the “MedWatch Online” internet-based website (<http://www.fda.gov/medwatch/>).

Limitations of FDA spontaneous reporting include both under-reporting and over-reporting.

An example of over-reporting occurs with recently approved drugs. This is partly due to enhanced publicity about these drugs.

Table 7 ADRs for the top ten herbal medicines

Herbal	Common use	Side effects and interactions
Echinacea	Treatment and prevention of upper respiratory infections, common cold	Rash, pruritis, dizziness, unclear long-term effects on the immune system.
St. John's wort	Mild to moderate depression	Gastrointestinal upset, photo-sensitivity. Mild serotonin syndrome with the following medications: paroxetine, trazodone, sertraline, and nefazodone. May decrease digoxin levels. May decrease cyclosporine serum concentrations. Combined oral contraceptives—breakthrough bleeding.
Ginkgo biloba	Dementia	Mild gastrointestinal distress, headache, may affect warfarin (increase INR). Interaction with aspirin (spontaneous hyphema)
Garlic	Hypertension, hypercholesterolemia	Gastrointestinal upset, gas, reflux, nausea, allergic reactions, and antiplatelet effects. May effect warfarin (increase INR)
Saw palmetto	Benign prostatic hyperplasia	Uncommon
Ginseng	General health promotion, sexual function, athletic ability, energy, fertility	High doses may cause diarrhea, hypertension, insomnia, nervousness, may affect warfarin (decreased INR)
Goldenseal	Upper respiratory infections, common cold	Diarrhea, hypertension, vasoconstriction
Aloe	Topical application for dermatitis, herpes, wound healing, and psoriasis, orally for constipation	May delay wound healing after topical application. Diarrhea, and hypokalemia with oral use
Siberian ginseng	Similar to ginseng	May raise digoxin levels. May affect warfarin (increased INR)
Valerian	Insomnia, anxiety	Fatigue, tremor, headache, paradoxical insomnia (not advised with other sedative-hypnotics)

Studies of therapeutic classes

Observational cohort or case control designs have been used to determine ADR relationships with specific therapeutic classes (36, 41). Medical claims data are often used in these studies and caution should be warranted due to lack of definite confirmation of drug exposure and the potential for confounding variables (38). However, these studies have been beneficial in determining risk of ADRs with specific classes (e.g., NSAIDs and the risk of peptic ulcer disease) (42).

Studies of specific medical syndromes

Observational cohort or case control designs can also be useful to study possible causality relationships of specific medical conditions or syndromes due to drug exposure (36, 41). These types of studies have been particularly useful in examining ADRs in a specific population, such as geriatric

or pediatric patients. These groups of patients are often excluded in Phase III trials. However, a disadvantage of these studies is that they also often use administrative data. These data can warrant risk of problems in determining causality due to potential confounding variables (38).

Assessing Adverse Drug Reactions

After detection of a possible ADE, causality assessment needs to be performed. It is important to be able to rank the likelihood of an ADR as unlikely, possible, probable, or definite. A major problem with determining causality is that confounding variables can contribute to the complexity of causality assessment (43). In order to determine causality, several important points of data are required. These include the nature of the adverse event, name of the putative drug, other potential causes, and the temporal relationship

Table 8 ADR Naranjo causality algorithm

	Yes	No	Do not know	Score
1. Are there previous conclusive reports on this reaction?	+1	0	0	
2. Did the adverse event appear after the suspected drug was administered?	+2	-1	0	
3. Did the adverse reaction improve when the drug was discontinued, or a specific antagonist was administered?	+1	0	0	
4. Did the adverse reaction reappear when the drug was readministered?	+2	-1	0	
5. Are there alternative causes (other than drug) that could on their own caused this reaction?	-1	+2	0	
6. Did the reaction reappear when a placebo was given?	-1	+1	0	
7. Was the drug detected in the blood (or other fluids) in concentrations known to be toxic?	+1	0	0	
8. Was the reaction more severe when the dose was increased, or less severe when the dose was decreased?	+1	0	0	
9. Did the patient have a similar reaction to the same or similar drugs in any previous exposure?	+1	0	0	
10. Was the adverse event confirmed by any objective evidence?	+1	0	0	
			Total score	

Probability category scores: Definite ≥ 9 ; Probable 5–8; Possible 1–4; Doubtful ≤ 0 .

between the drug and adverse event. Potential causes are obtained by examining the medical history, physical examination findings, and directed diagnostic tests.

Identification of causality can be performed simply by using a health care provider's clinical reasoning and judgment. The main disadvantage to this approach is a low inter-rater and intra-rater agreement for ADR causality (44, 45).

An ADR causality algorithm addresses the issue of inter-rater and intra-rater reliability with a series of clinical questions. For example, the Naranjo algorithm consists of a series of clinical questions that focus on temporal and dose–response relationships, consistency of the ADR with previous clinical reports or patient experiences, placebo response, drug dechallenge and rechallenge, toxic blood drug concentrations, alternative causes of the reaction, and whether the event was confirmed by objective evidence (44) (Table 8). Numerous health care institutions and the FDA use some type of causality algorithm to minimize disagreement among different evaluators and improve inter-rater and intra-rate agreement.

PREVENTING ADVERSE DRUG REACTIONS

ADRs are problematic in that they cause significant morbidity and mortality. Almost 95% of ADRs are Type A

(predictable) reactions, and thus with quality improvement measures, ADRs can be avoided and prevented (46). Knowledge of causative factors and an increase in patient education may help prevent ADRs. Improvements in the documentation of allergic reactions (e.g., via computer tracking), development of tools to enhance compliance, and application of tools to improve prescribing and administration of drugs are other preventative approaches to ADRs.

In 1994, the ASHP, the American Medical Association (AMA), and the American Nurses Association (ANA) generated the following system of recommendations to prevent ADRs in health care systems:

1. Health care systems should establish processes in which prescribers enter medication orders directly into computer systems.
2. Health care systems should evaluate the use of machine-readable coding (e.g., bar coding) in their medication use processes.
3. Health care systems should develop better systems for monitoring and reporting adverse drug events.
4. Health care systems should use unit dose medication distribution and pharmacy-based intravenous medication admixture systems.
5. Health care systems should assign pharmacists to work in patient care areas in direct collaboration with prescribers and those administering medications.

6. Health care systems should approach medication errors as system failures and seek system solutions in preventing them.
7. Health care systems should ensure that medication orders are routinely reviewed by the pharmacist before first doses and should ensure that prescribers, pharmacists, nurses, and other workers seek resolution whenever there is any question of safety with respect to medication use (47).

SUMMARY

Adverse drug reactions are of significant concern in the pharmaceutical technology arena. Various drug and patient factors that predispose to ADRs have been identified. Reporting systems used to screen and assess ADRs facilitate the understanding of risk factors and contribute to the development of systematic improvement in the prevention of ADRs.

REFERENCES

1. Schumock, G.; Guenette, A. Adverse Drug Events. *Pharmacotherapy Self-Assessment Program*; Carter, B., Ed.; ACCP: Kansas City, MO, 1999; 5, 103–130.
2. Karch, F.E.; Lasagna, L. Adverse Drug Reactions: A Critical Review. *JAMA* **1975**, *234*, 1236–1241.
3. Rossi, A.; Knapp, D. Discovery of New Adverse Drug Reactions. A Review of the Food and Drug Administration's Spontaneous Reporting System. *JAMA* **1984**, *252*, 1030–1033.
4. *Joint Commission on Accreditation of Health Care Organizations. 1998 Comprehensive Accreditation Manual for Hospitals*; A.M.H: Oakbrook Terrace, IL, 1997.
5. American Society of Health-System Pharmacists. Suggested Definitions and Relationships Among Medication Misadventures, Medication Errors, Adverse Drug Events and Adverse Drug Reactions. *Am. J. Health-Syst. Pharm.* **1998**, *55*, 165–166.
6. Edwards, I.R. Pharmacological Basis of Adverse Drug Reactions. *Avery's Drug Treatment*; Speight, T., Holford, N., Eds.; ADIS International LTD: Auckland, New Zealand, 1997; 261–299.
7. Young, L.R.; Wurtzbacher, J.D.; Blankenship, C.S. Adverse Drug Reactions: A Review for Healthcare Practitioners. *Am. J. Managed Care* **1997**, *12*, 1884–1906.
8. Lazarou, J.; Pomeranz, B.; Corey, P. Incidence of Adverse Drug Reactions in Hospitalized Patients. *JAMA* **1998**, *279* (15), 1200–1205.
9. Einarson, T. Drug-Related Hospital Admissions. *Ann. Pharmacother.* **1993**, *27*, 832–840.
10. Clasen, D.C.; Pestonik, S.L.; Evans, R.S.; Lloyd, J.F.; Burke, J.P. Adverse Drug Events in Hospitalized Patients: Excess Length of Stay, Extra Costs, and Attributable Mortality. *JAMA* **1997**, *277*, 301–306.
11. *Applied Biopharmaceutics and Pharmacokinetics*; Shargel, L., Yu, A., Eds.; Appleton and Lange: Stamford, CT, 1999.
12. Uchegbu, I.; Florence, A. Adverse Drug Events Related to Dosage Forms and Delivery Systems. *Drug Safety Concepts* **1996**, *14* (1), 39–67.
13. Swift, C.G. Pharmacodynamics: Changes in Homeostatic Mechanisms, Receptor and Target Organ Sensitivity in the Elderly. *Br. Med. Bull.* **1990**, *46*, 36–52.
14. Parker, B.M.; Cusack, B.J.; Vestal, R.E. Pharmacokinetic Optimisation of Drug Therapy in Elderly Patients. *Drugs Aging* **1995**, *7*, 10–18.
15. Gurwitz, J.H.; Avorn, J. The Ambiguous Relation Between Aging and Adverse Drug Reactions. *Ann. Intern. Med.* **1991**, *114*, 956–966.
16. Taffet, G.E. Age-Related Physiologic Changes. *Geriatric Review Syllabus: A Core Curriculum in Geriatric Medicine*; Reuben, D.B., Yoshikawa, T.T., Besdine, R.W., Eds.; Kendall/Hunt for the American Geriatric Society: Dubuque, IA, 1996; 11–24.
17. Nahata, M.C. Pediatrics. *Pharmacotherapy A Pathophysiologic Approach*; DiPiro, J.T., Talbert, R., Yee, G.C., Matzke, G.R., Wells, B.G., Posey, M.L., Eds.; Appleton & Lange: Stamford, CT, 1999; 44–51.
18. Gupta, A.; Waldhauser, L.K. Adverse Drug Reactions from Birth to Early Childhood. *Pediatr. Clin. North. Am.* **1997**, *44*, 79–92.
19. Piper, D.W.; deCarle, D.J.; Talley, N.J.; Gallagher, N.D.; Wilson, J.S.; Powell, L.W.; Crawford, D.; Gibson, P.R.; Sorrell, T.C.; Kellow, J.E.; Roberts, R.K. Gastrointestinal and Hepatic Diseases. *Avery's Drug Treatment*; Speight, T., Holford, N., Eds.; ADIS International LTD: Auckland, New Zealand, 1997; 1010–1012.
20. Critchley, J.A.; Chan, T.Y.K.; Cumming, A.D. Renal Diseases. *Avery's Drug Treatment*; Speight, T., Holford, N., Eds.; ADIS International LTD: Auckland, New Zealand, 1997; 1107–1109.
21. Tillement, J.P.; Lhoste, F.; Giudicelli, J.F. Diseases and Drug Protein Binding. *Clin. Pharmacokinet* **1978**, *3*, 144–154.
22. Smith, J.W.; Seidl, L.G.; Cluff, L.E. Studies on the Epidemiology and Adverse Drug Reactions. *Ann. Intern. Med.* **1966**, *65*, 629–640.
23. Franse, V.; Stark, N.; Powers, T. Drug-Nutrient Interactions in a Veterans Administration Medical Center Teaching Hospital. *Nutrition in Clinical Practice* **1988**, *3* (4), 145–147.
24. Yamreudeewong, W.; Henann, N.; Fazio, A.; Lower, D.; Cassidy, T. Drug-Food Interactions in Clinical Practice. *J. Fam. Pract.* **1995**, *40* (4), 376–384.
25. Mason, N.; Boyd, S. Drug-Nutrient Interactions in Renal Failure. *J. Renal Nutrition* **1995**, *5* (4), 214–222.
26. Krishnaswamy, K. Drug Metabolism and Pharmacokinetics in Malnutrition. *Clin. Pharmacokinet* **1978**, *3*, 216–240.
27. Grymonpre, R.E.; Mitenko, P.A.; Sitar, D.S.; Aoki, F.Y.; Montgomery, P.R. Drug-Associated Hospital Admissions in Older Medical Patients. *J. Am. Geriatr. Soc.* **1988**, *36*, 1092–1098.

28. Mar, C.; Bent, S. An Evidence-Based Review of the 10 Most Commonly Used Herbs. *West. J. Med.* **1999**, *171*, 168–171.
29. Fugh-Berman, A. Herb-Drug Interactions. *Lancet* **2000**, *355*, 134–38.
30. American Society of Health-System Pharmacists. ASHP Guidelines on Adverse Drug Reaction Monitoring and Reporting. *Am. J. Health-Syst. Pharm.* **1995**, *52*, 417–419.
31. Johnston, P.E.; Morrow, J.D.; Branch, R. Use of a Database Computer Program to Identify Trends in Reporting Adverse Drug Reactions. *Am. J. Hosp. Pharm.* **1990**, *47*, 1321–1327.
32. Koch, K.E. Use of Standard Screening Procedures to Identify Adverse Drug Reactions. *Am. J. Hosp. Pharm.* **1990**, *47*, 1314–1320.
33. O'Neil, A.C.; Petersen, L.A.; Cook, E.F.; Bates, D.Q.; Lee, T.H.; Berman, T.A. Physician Reporting Compared With Medical-record Review to Identify Adverse Medical Events. *Ann. Intern. Med.* **1993**, *119*, 370–376.
34. Prosser, T.R.; Kamysz, P.L. Multidisciplinary Adverse Drug Reaction Surveillance Program. *Am. J. Hosp. Pharm.* **1990**, *47*, 1334–1339.
35. Porta, M.S.; Hartzema, A.G. The Contribution of Epidemiology to the Study of Drugs. *Drug. Intell. Clin. Pharm.* **1987**, *21*, 41–47.
36. Strom, B.L., Ed. *Pharmacoepidemiology*, 2nd Ed.; John Wiley: Chichester, 1994.
37. *The Detection of New Adverse Drug Reactions*; Stephens, M.D.B., Ed.; The MacMillan Press: Basingstoke, 1988.
38. Hanlon, J.T.; Schmader, K.; Lewis, I. Adverse Drug Reactions. *Therapeutics in the Elderly*; Delafuente, J.C., Stewart, R.B., Eds.; Harvey Whitney Books: Cincinnati, 1995; 212–227.
39. Kennedy, D.L.; McGinnis, T. Monitoring Adverse Drug Reactions: The FDA's New MedWatch Program. *Pharmacol. Ther.* **1993**, *833–834*, 839–842.
40. Kessler, D.A. Introducing MEDWatch: New Approach to Reporting Medication and Device Adverse Effects and Product Problems. *J.A.M.A.* **1993**, *269*, 2765–2768.
41. *Pharmacoepidemiology: An Introduction*; Hartzema, A.G., Porta, M.S., Tilson, H.H., Eds.; Harvey Whitney Books: Cincinnati, 1991.
42. Guess, H.A.; West, R.; Strand, L.M.; Helston, D.; Lydick, E.G.; Bergman, C.A.; Wolski, K. Fatal Upper Gastrointestinal Hemorrhage of Perforation Among Users and Nonusers of Nonsteroidal Anti-Inflammatory Drugs in Saskatchewan, Canada 1983. *J. Clin. Epidemiol.* **1988**, *41*, 35–45.
43. Meyboom, R.H.B.; Hekster, Y.A.; Antoine, C.G.; Gribnau, F.W.J.; Edward, I.R. Causal or Casual? The Role of Causality Assessment in Pharmacovigilance. *Drug Saf.* **1997**, *6*, 374–389.
44. Naranjo, C.A.; Busto, U.; Sellers, E.M.; Sandor, P.; Ruiz, I.; Roberts, E.A.; Janecek, E.; Domecq, C.; Greenblatt, D.J. A Method for Estimating the Probability of Adverse Drug Reactions. *Clin. Pharmacol. Ther.* **1981**, *30*, 239–245.
45. Karch, F.E.; Lasagna, L. Adverse Drug Reactions: A Critical Review. *JAMA* **1975**, *234*, 1236–1241.
46. Rawlins, M.D. Adverse Reactions to Drugs. *Br. Med. J.* **1981**, *82*, 974–976.
47. American Society of Health System Pharmacists. Top-Priority Actions for Preventing Adverse Drug Events in Hospitals: "Recommendations of An Expert Panel." *Am. J. Health-Syst. Pharm.* **1996**, *53*, 747–751.

ADVERTISING AND PROMOTION OF PRESCRIPTION AND OVER-THE-COUNTER DRUG PRODUCTS

Wayne L. Pines

Pharmaceutical Consultant, Washington, DC., U.S.A.

INTRODUCTION

Pharmaceutical companies aggressively market their products to assure that everyone who needs to know about them receives information in a timely way. The marketing of drugs has many aspects that are unique in the world of product marketing:

1. Pharmaceutical marketing is directed to a large extent at audiences who are not the ultimate users of the products (i.e., physicians, pharmacists, HMO managers).
2. Even in an age when mass media predominate, marketing to certain audiences such as physicians remains very personal.
3. The very fact that pharmaceutical companies market their products is controversial, as there are critics who believe that essential health care products should not be marketed.
4. The federal government dedicates significant resources and has established far-reaching and complicated rules to assure that pharmaceutical promotion is accurate and balanced.

A number of significant issues linger, such as how companies can communicate scientific information that is not within FDA-approved labeling; how to regulate the Internet; and the role of the brief summary. These and other issues make the field of advertising and promotion of pharmaceuticals dynamic and ever-changing.

BACKGROUND

The advertising and promotion of prescription drugs is inherently controversial. The drug industry maintains that such promotion provides an essential educational benefit and is needed to assure that the benefits of new drugs are made available to patients through their physician, pharmacist, or other health care provider. The industry stresses that promotion encompasses a variety of objectives beyond just selling more products. These

objectives include informing consumers about new and existing products that may be of value in improving their health; educating health care providers about new and existing products and how to use them; and providing information needed by those who finance or pay directly for medical products (1).

Further, the industry and advertising/promotion advocates contend that in the modern age there is an inherent need for pharmaceutical companies to communicate information about their research and new products. The public demands information about health care products; everyone wants to know about medical progress. Investment capital for new biotechnology companies is dependent on understanding the potential value and research progress of new products. Health care providers need to know, on a continuing basis, what kinds of advances are being made in their specialty.

When a new drug is introduced, pharmaceutical companies dedicate substantial resources to providing information to the health care provider through the sales force, and through advertising, direct mail, seminars, dinner meetings, presentations at medical meetings, exhibits, etc. In the 1990s promoting new drugs directly to the ultimate user, the patient, became common, and companies began to invest in magazine, newspaper, radio, television, and Internet promotional activities.

Critics of drug marketing say that such efforts are inappropriate ways to communicate information about prescription drugs. They say that pharmaceuticals are different from other heavily-promoted products in that they are necessary for health, can be harmful if misused or overused, and that they can mean the difference between life and death.

The critics say that advertising and promotion can endanger health by leading consumers to seek out and take certain prescription drugs that are inappropriate for their illness. They also say that promotional activities by drug companies are too expensive, adding to the price of prescription drugs and thereby denying access to those who cannot afford the drugs they need. They say that promotional activities are not needed, as medical information can best be conveyed through the medical

literature to physicians who then can determine what is best for their patients (2).

Finally, the critics have expressed concern about the accuracy and truthfulness of information issued by prescription drug manufacturers, fearing that such information will be biased because it comes from companies whose principal interest is in increasing the sales of its products.

The debate over the appropriateness of pharmaceutical promotion is one that will go on forever. The reality is that, at least in the United States, pharmaceutical promotion is an integral part of the process of health care delivery. It is heavily regulated by FDA to assure accuracy and balance. For the foreseeable future pharmaceutical companies will continue to use communications to capture the attention of various audiences.

EXTENT OF PHARMACEUTICAL ADVERTISING AND PROMOTION

There is no entirely accurate way to calculate how much is spent on advertising and promoting prescription drugs. The reason is that the term encompasses so many different activities. It includes, for example, traditional advertising in medical journals, visits to physicians and other health professionals by the company's sales force, dinner meetings, extensive exhibits at medical meetings, CME sponsored by drug companies, public relations programs, promotion directly to consumers using all media available, and outreach to professional and patient organizations.

Within drug companies the amount spent on promoting a particular product comes from many different budgets. Estimates on how much is spent in the U.S. on promoting prescription drugs range as high as \$13.9 billion. In 1999 the amount spent on direct-to-consumer television advertising alone was \$1.1 billion (3). Direct-to-consumer advertising of prescription drugs in consumer magazines, newspapers and on television has raised the visibility of prescription drug promotional activities.

Drug companies generally spend most on drug promotion when a new drug is being launched. Very often the promotional budget is cut dramatically, if not altogether, when a drug faces generic competition at the conclusion of its patent life.

REGULATORY HISTORY

FDA's role is to assure that all advertising and promotion of prescription drugs is accurate, provides full disclosure

of risks and is fairly balanced. There are two legal bases for FDA's jurisdiction over all advertising and promotional materials and programs issued by or sponsored (i.e. funded) by drug companies.

One legal basis of FDA jurisdiction is the 1962 Drug Amendments to the Food, Drug & Cosmetic (FD&C) Act. In 1962, the Congress added major amendments to the FD&C Act. The thalidomide incident—in which Americans were spared birth defects by a vigilant FDA that kept this potent teratogenic drug off the market—spurred enactment of the 1962 Amendments. But the Congress paid considerable attention in its legislative hearings to alleged advertising and promotion abuses by pharmaceutical companies.

Section 502(n) of the 1962 Drug Amendments dealt specifically with advertising. It assigned responsibility for prescription drug advertising to the FDA; previously, the jurisdiction had resided with the Federal Trade Commission (FTC). Section 502(n) was narrow in scope and vague in requirements. It said a prescription drug is misbranded

... unless the manufacturer, packer or distributor thereof includes in all advertisements and other descriptive printed matter issued or caused to be issued by the manufacturer, packer or distributor with respect to that drug a true statement of 1) the established name as defined in Section 502(e), printed prominently and in type at least half as large as that used for any trade or brand name thereof, 2) the formula showing quantitatively each ingredient of such drug to the extent required for labels under Section 502(e), and 3) such other information in brief summary relating to side effects, contraindications, and effectiveness as shall be required in regulations which shall be issued by the Secretary in accordance with the procedure specified in Section 701(e) of this Act: *Provided*, that a) except in extraordinary circumstances, no regulation issued under this paragraph shall require prior approval by the Secretary of the content of any advertisement, b) no advertisement of a prescription drug, published after the effective date of regulations issued under this paragraph applicable to advertisements of prescription drugs, shall...be subject to the provisions of sections 12 through 17 of the Federal Trade Commission Act, as amended... (4).

Basically, the law said that for a drug to not be misbranded, the generic name must be in the advertisement; the drug's formula must be included; and a "brief summary" of risks must accompany each advertisement. FDA was barred from requiring that advertising be approved in advance of use, a provision that addressed and

accommodated First Amendment (“free speech”) issues. In the 1960s FDA promulgated detailed regulations to implement Section 502(n).^a

The second legal base for FDA’s current regulatory authority over prescription drug advertising and promotion stems from the definition of “labeling” in Section 201(m) of the FD&C Act. This section defines “labeling” as printed or graphic materials accompanying a drug product. FDA regulations further define “labeling” to mean

brochures, booklets, mailing pieces, detailing pieces, file cards, bulletins, calendars, price lists, catalogs, house organs, letters, motion picture films, film strips, lantern slides, sound recordings, exhibits, literature and reprints and similar pieces of printed audio or visual matter descriptive of a drug and references published for use by medical practitioners, pharmacists or nurses, containing drug information supplied by the manufacturer, packer, or distributor and which are disseminated by or on behalf of its manufacturer, packer or distributor (5).

In effect, FDA combines Sections 502(n) and 201(m) and their implementing regulations to declare that any materials issued by or sponsored by a drug company about its prescription drugs are subject to regulation. In effect, promotional materials that are not deemed to be “advertising” are regulated as “labeling”. FDA has established virtually the same requirements for all promotional material, whether it comes under the definition of “advertising” or under the “labeling” definition (6).

Even though the FD&C Act as amended in 1962 did not envision FDA taking broad authority over all communications and marketing materials issued by drug companies, the drug industry has acquiesced to FDA’s increasingly broad definition of its authority. The industry has taken the view that just so long as FDA enforces the regulations even-handedly, the regulations help provide an underlying confidence in the accuracy of their materials. Further, the industry over the years has chosen not to challenge FDA’s assumption of authority in this area, choosing instead to wage its jurisdictional battles with FDA over other matters.

Over the years, a huge array of formal regulations, guidances, informal policies, and enforcement actions through warning letters has evolved from FDA. Combined, they establish a highly intricate system for regulating the advertising of prescription drugs.

The FDA has established authority over every word, footnote, picture and nuance of all promotional activities

issued by or sponsored by drug companies. That includes promotion to the consumer, press materials, detailing aids, continuing medical education, seminars, materials for managed care organizations. The agency has jurisdiction over oral statements made by a company representative to a physician.

Similar requirements exist for other medical products regulated by FDA. While pharmaceutical products are regulated by FDA’s Center for Drug Research and Evaluation (CDER), the comparable center that regulates biological products—the Center for Biologics Research and Evaluation (CBER)—basically applies the same standards. The regulation of the advertising of medical devices has its own legal basis, Section 502 (r) of the FD&C Act.

FDA regulations that apply to the promotion of medical products are spread out in several sections of the Code of Federal Regulations (Title 21). Among the applicable sections are Part 200 and, most particularly, Part 201 for pharmaceutical promotion; Part 510 for veterinary drugs; Part 601 for biological products; and Part 801 for medical devices.

FDA’s regulation of advertising and promotion was stepped up dramatically during the 1990s (7). The agency recognized that new forms of communications about prescription drugs had become more common, such as press releases, brochures for consumers, and cable television shows directed at physicians. The surge in promotional activity directed both at health professionals and consumers led FDA to increase significantly the resources dedicated to monitoring such activities and enforcing the regulations more aggressively, as well as the issuance of new policies to cover the new forms of communication. The increase in staff and enforcement occurred in every product category of FDA’s regulatory jurisdiction but was most pronounced in the pharmaceuticals area.

BASIC PRINCIPLES OF FDA REGULATION

The agency’s policies continue to evolve, in response to newer forms of communications and new needs of the medical marketplace. However, a few basic principles have been at the foundation of FDA’s regulatory approach (8).

All information in prescription drug advertising and promotional materials issued by or sponsored by drug companies must be truthful and not misleading. Only information consistent with FDA-approved labeling may be disseminated in promotional materials.

^a 21 C.F.R., Part 201 contains these regulations.

Any claims made about the product must be supported by scientific studies that have been reviewed by FDA or that are subject to FDA review. If a company wants to make, for example, a cost-effectiveness claim or a quality of life claim for a product, the claim must be substantiated by studies.

All benefits information must be fairly balanced with risk information. FDA defines “fair balance” as a “balanced presentation of benefits and risks.” This means that, in an advertisement, for example, to the extent that a particular benefit is described, the medical risks associated with that benefit also must be described. “Fair balance” also may mean that a product’s limitations must be delineated. “Fair balance” differs for each individual drug and communication technique. FDA requires that the presentation of all side effects and contraindications must have a prominence and readability that is reasonable, comparable to the presentation of benefits information.

All relevant material facts about the product must be disclosed. If an advertising or promotional piece omits to mention a material fact that would influence the prescribing, use or purchase of a product, then the piece would be rendered violative.

Whenever a product and its medical claim are promoted, there must be provision for the recipient of the information to obtain additional comprehensive information about the product. Advertising in print must be accompanied by the “brief summary” of risks. FDA permits print advertising directed at the consumer to be accompanied by a consumer-directed “brief summary,” which can be in the form of a Q&A. If a product is promoted directly to the consumer through a television commercial, then provision must be made to assure that the consumer can obtain substantial additional information needed to assess the benefits and risks of the product. Under an FDA guidance issued in August 1999, this requirement can be met through four mechanisms: an 800 number that the consumer can call to ask that the information be mailed or read; referral to an internet site; information in an advertisement in a current issue of a magazine; or materials that can be obtained from a health care provider (9).

The rules apply to any product-specific information issue or caused to be issued by a company, either directly or through any agent. Anyone with whom a prescription drug company has a financial relationship is, in effect, bound by the same communications rules as the company. A physician hired by a drug company to speak on its behalf is obligated to restrict promotional claims to approved labeling. The drug company is required to assure that this occurs.

SPECIFIC FDA ISSUES

FDA has expressed, through its guidances, policies and enforcement actions, particular concerns about certain issues.

Launch materials: These are defined as materials that are used to promote a product or a new indication when it first comes into the marketplace. FDA regards such materials as its highest review priority, because claims made in an introductory campaign for a new product or new indication establish certain expectations in the minds of health professionals or patients. The review of launch materials is FDA’s top advertising/promotion priority (10). While FDA by law cannot require the submission of such materials in advance of use, most companies submit launch materials before using them. FDA will not officially “approve” such materials, but will provide detailed comments on them. The failure by a company to change materials to address FDA’s comments creates a risk of an enforcement action.

Off-label promotion: This has become a highly controversial area involving FDA regulation. Promoting a marketed product for an unapproved use poses, in the view of the FDA, a significant potential threat to the public health. If health care providers and consumers are led to believe that a product is safe and effective for a use for which it is not approved, then adverse health consequences may occur. Further, if a product is approved for one use and then can be promoted for other uses, there is no incentive for companies to conduct the needed additional research.

FDA defines an unapproved use very simply: it is a use that is not in the FDA-approved labeling. It makes no difference to the FDA whether a use is widely practiced or is universally accepted by the medical community. The sole criterion for whether a claim can be promoted is whether it is included in the FDA-approved labeling. Thus, FDA can take action against a company when it promotes an indication not included in the approved labeling, even when medical scientists believe that a particular use is safe and effective. For example, in a 1993 consent decree signed by one company, FDA objected to its efforts to promote a drug for a use that was approved and widely accepted overseas, but not included in the product’s U.S. labeling (11).

FDA’s efforts to prohibit or restrict the provision of off-label information has been at the center of debate and controversy since the early 1990s. The Washington Legal Foundation, a conservative legal organization, filed a petition with FDA in October 1993 challenging the constitutionality of FDA’s rules regulating educational activities such as Continuing Medical Education (CME)

programs. The petition led to a case eventually wound up in the federal courts with rulings that, for most of the 1990s, left the issue unresolved (12).

While the Washington Legal Foundation case was wending its way through the courts, the Congress sought to resolve some of the central issues. Section 401 of the FDA Modernization Act of 1997 set forth a procedure for companies to secure FDA approval when they wanted to disseminate reprints and textbooks containing off-label to health care professionals. This procedure requires that the company be investigating the off-label use with the intention of seeking approval; that records be kept of who received the off-label information; that FDA clearance for the dissemination be sought; and that the information include a physician package insert, a bibliography and a disclaimer that the information is off-label (13).

As the debate over off-label information dissemination continues, FDA has other legal remedies to pursue if it believes that a company is issuing information that could lead to mis-prescribing or jeopardize the health of patients. FDA has said it can and will invoke a section of its regulations that prohibit the dissemination of information that is "false or misleading" (14).

The issue of off-label information is certain to remain controversial for some time, with the FDA seeking ways to restrict the dissemination of off-label information about prescription drugs and the legal community and industry seeking ways to provide such information to health professionals without stimulating enforcement actions.

Pre-approval promotion: FDA objects to the promotion of a product before approval. The agency's view was set forth in 1994 correspondence with a company, expressing the agency's interpretation of Section 312.7 of the regulations:

Prior to approval, promotional materials provided by the sponsor are often inaccurate with respect to the indications and risk information that ultimately appear in the approved product labeling. This inaccuracy occurs because the sponsor does not know what indications and other information the final product labeling will contain. Additionally, the sponsor's assessment of the drug may be overly optimistic, exaggerating efficacy while minimizing risk. Thus, the sponsor's biases may be incorporated into the promotional materials it provides prior to approval of final product labeling. Such promotional materials do not fulfill the needs of health care providers or benefits managers to obtain balanced, accurate information about new drugs, but instead disseminate misinformation and create potential risk and misuse (15).

Comparative claims: FDA requires that any comparative claim made by a company for its pharmaceutical product be substantiated with scientific evidence. Examples of comparative claims are "drug of choice," "unsurpassed," or "more effective." The supporting studies must be "head-to-head"—that is, they must be designed prospectively to compare the two products directly. FDA prohibits a claim of superiority based solely on data derived from two or more studies that are compared, even when those studies have similar protocols. Further, FDA requires that any comparative claim be clinically relevant to patients. And, if a comparative claim is made, it cannot be false or misleading and cannot leave out other measurements by which the company's product is inferior to its competitor.

Pharmacoeconomic and quality of life claims: Claims that a product is cost-effective, or that it will improve an individual's quality of life, may be made so long as there is proof. There is no firm standard for how such proof must be documented. In the case of pharmacoeconomic claims, the Food and Drug Administration Modernization Act of 1997 said that companies can provide cost-effectiveness information to managed care and other purchasing or reimbursing organizations, so long as the claims were based on "competent and reliable" information. This was intended to be a lesser standard than the usual one that FDA applies to claims of safety and effectiveness (which must be based on studies that are "adequate and well-controlled") (16). Quality of life claims, on the other hand, must be demonstrated in clinical studies.

Reminder advertisements: Companies often sponsor advertising that does not mention the indications for a product, but that is intended to "remind" readers or viewers of the name. Such advertising is called "reminder" advertising. Under an FDA regulation, such advertising may delineate the name (including the generic name), the ingredients, the dosage form and the name and address of the manufacturer, packager or distributor. The advertisement may not mention what the product is used for, and may make no safety or efficacy claim or any other representation about the product. Such advertising is exempt from the requirement for a "brief summary." "Reminder advertisements" may not be sponsored for prescription drugs for which there is a "boxed" warning in the FDA-approved labeling, or, in the case of older drugs, whose efficacy has not been reaffirmed by the FDA.

Press releases: FDA takes jurisdiction over the content of company-issued press releases by declaring them to be "labeling."^b As such, press releases must contain fair

^b FDA initially asserted jurisdiction over the content of product-specific press releases in 1983.

balance of risk information and must be accompanied by the full package insert. FDA also expects press releases to contain disclaimers that will help put the product into regulatory context (e.g., in a research announcement for a product not yet approved, the release should make clear that the product is not approved, or that an application is pending before FDA.) Press releases issued before product approval must avoid stating that the product has been shown to be safe and effective and should avoid any statements that might be construed as promotional, such as through use of words like “promising” or “break-through.”

Video news releases: Video news releases are 75- to 90-second electronic versions of press releases. They are regulated by FDA as promotional labeling, and therefore must contain balancing information. In addition, FDA expects the recipients of video news releases to have easy access to a physician package insert (17).

Publications intended for stockholders and investors: FDA regulates publications and other materials intended for stockholders and investors with the same rules as apply to materials for other audiences. However, the agency generally does not enforce the rules very aggressively when it is clear that the materials are for financial purposes rather than to promote a product’s use among physicians and patients. For example, FDA has never required annual reports issued to investors that mention ongoing research or newly approved products in a cursory way to include the labeling as an attachment.

Scientific exchange: FDA has adopted the view that “scientific exchange”—the well recognized need for scientists to communicate with each other about research findings—should be protected. When such “exchange” crosses over into “promotion”—and the line is hard to define—then FDA regulations are invoked.

Any scientific meeting whose content is fully under the control of a company and that is product-specific must meet all criteria for fair balance and full disclosure. This means that the materials and speeches must be within approved labeling.

A mechanism for drug companies to see that health professionals receive the latest results of medical research is for drug companies to sponsor CME programs. FDA has set forth detailed rules for how companies may influence and interact with legitimate CME programs about their products or product category. FDA permits such company-sponsored CME so long as the financial grant is provided to an accredited third party and the CME is conducted independently of the company. Companies that sponsor CME generally have standard contracts with providers that set forth the specific relationship that is intended to assure independence of the program, the presenters and the

derivative material such as monographs or publications (18).

Exhibits: Exhibits sponsored by drug companies at medical or other meetings are regulated by the FDA. The regulations require that all materials at the exhibit booth, including the headings in the booth itself, must be within approved labeling. There is provision within FDA’s regulations for scientific exchange outside of approved labeling so long as such discussions are not promotional, are conducted between health professionals, and so long as there is a clear distinction between approved indications and new research or off-label uses of a product. Companies often sponsor multiple booths at meetings, one for promotion of approved indications and the other for scientific exchange where new research is discussed.

Internet: The Internet has posed some unprecedented regulatory challenges. FDA has declined to issue a written guidance on Internet activities sponsored by drug companies, saying instead that it will apply the same standards to Internet advertising and promotional statements as it does to other means of communication (19). For the most part, corporate information that is not product-specific may be put on the Internet without regulation. So too, can information about approved products, so long as the information is consistent with labeling and provides ready and easy access to product labeling. The challenges arise with links to other Internet sites or with chat rooms, especially those that may contain off-label information, and how to manage information that emanates from overseas.

Requests for information: Companies may respond to requests for information from any external source. However, FDA differentiates between requests that are solicited by the company—that is, that the company has encouraged in any way—and those that are totally unsolicited.

If a request for information is solicited, then the company must stay within labeling in responding. If a request is totally unsolicited, then questions about off-label uses can be answered with information that is not restricted to approved label information. Companies can respond by providing research studies or other information. Generally, the medical affairs departments within companies handle responses to unsolicited requests for information. Companies usually keep records of such requests, and those records are subject to FDA inspection. If off-label information is sent to the requester, then the cover letter must make clear FDA does not approve the indication.

Accelerated approval drugs: These are drugs that receive priority attention by FDA and that, because they are for life-threatening conditions, are approved on the

basis of limited or surrogate marker data. FDAs regulations require that any advertising or promotional pieces for drugs approved through the accelerated process be approved prior to use. This requirement includes pieces used during the launch period as well as pieces used thereafter, for as long as the drug approval comes under the accelerated approval status (20).

Direct to consumer promotion: The promotion of pharmaceuticals directly to consumers began in the 1980s but came of age in the 1990s. In 1996, the amount spent on advertising prescription drugs to consumers exceeded the amount spent on advertising to physicians (though a company's promotional budget directed at physicians is still much higher than the consumer budget, since the highest spending is on sales representatives and direct promotion to physicians).

Print advertising directed at consumers must meet the same standard as print advertising directed at physicians. It must not only be accurate; it also must include a "fair balance" of benefits and risks, and it must include a "brief summary" of risks. FDA encourages companies to print "brief summary" information in consumer-friendly language.

Under a guidance issued in August 1999, FDA permits direct to consumer advertising on television. TV advertising must include a "major statement" of risk information. The TV commercial must also make provision for the consumer to obtain full labeling via the Internet or a toll-free number, and must also state that additional information is available from physicians, pharmacists or in an advertisement in a current publication (21).

FDA also permits advertising to consumers that does not mention the product name, but that identifies a disease condition and urges consumers to see their doctors. These are called "help-seeking" advertisements. However, FDA is sensitive to efforts to make implicit product claims that, while not mentioning the product by name, might be understood by a reasonable consumer as referring to a particular product.

FDA ENFORCEMENT

FDA monitors the promotional marketplace in a variety of ways. Staffers attend medical meetings, read consumer and medical publications, and watch TV. They are also asked to review materials in advance of use by drug companies. In addition, FDA learns of marketplace violations from competitor complaints. Finally, all advertising and promotional materials for prescription drugs must be submitted to FDA at the time of their first

use. Materials are submitted to the Division of Drug Marketing, Advertising and Communications (DDMAC) in CDER, or to the Advertising and Promotional Labeling Staff (APLS) in CBER. These are the staff that regulate the advertising and promotion of pharmaceuticals and biological products, respectively.

The growing prominence of advertising and promotion regulation during the 1990s made it necessary for pharmaceutical companies to dedicate more resources to the development and review of such materials. Virtually all companies have written policies and procedures for reviewing promotional materials and programs, including an internal review process. They also have dedicated regulatory staff that oversee the process. In addition, many companies regularly train staff on regulatory requirements. Periodic training has become a major focus for companies that want to be sure that everyone in the company, from top management to the sales force, is aware of the company's philosophy towards promotion and of FDA's rules.

The most common actions that FDA takes when it detects a violation are to send the offending company a letter (once known as a "Notice of Violation" or NOV, now known as an untitled letter) or a warning letter. An untitled letter identifies FDA's concern and seeks remedial action such as discontinuation of the promotional piece at issue.

Warning letters, which are sent far less frequently than untitled letters (the average is less than a dozen per year from DDMAC) are reserved for more serious situations. The issuance of a warning letter means that the agency has concluded that a legal violation has occurred and remedial action is needed. Usually the remedial action includes withdrawal of the violative materials and the company-wide dissemination of the warning letter. FDA may also seek a "Dear Health Professional" letter or other corrective measures such as mandatory training or pre-clearance of any new promotional materials for a period of time. In one instance, FDA required a company to produce corrective television commercials to address issues raised in a TV commercial that FDA found violative.

Untitled and warning letters provide 10 to 15 days for the company to respond and present a plan of action. Companies may appeal the FDA's finding and discuss the appropriateness of the remedial action, but FDA always retains the authority to seek more stringent action such as a consent decree. Warning letters are made public immediately upon issuance by FDA, and thus become a potential source of negative publicity. Untitled letters also are publicly available.

FDA has other enforcement options if it finds that a company has violated the advertising regulations. It can seek to negotiate a consent decree with a company if it feels that such a decree is needed to keep the company in

compliance. Three consent decrees involving advertising violations were signed in the 1990s. FDA may also seek a seizure, injunction or criminal action (misdemeanor or felony) against violative companies or products, but as a practical matter these more severe and time-consuming legal actions are unlikely. The most dramatic enforcement action taken by FDA in this area involved a fine of \$50 million against a company that allegedly overpromoted a product over a substantial period of years.^c

FTC REGULATION

The FDA has jurisdiction over prescription drug advertising and promotion, while the FTC has jurisdiction over the advertising and promotion of over-the-counter (OTC) medicines, as well as dietary supplements and medical devices (other than restricted devices as defined in the FD&C Act). FDA retains full jurisdiction over the labeling of all medical products, including OTC drugs and dietary supplements.

The FTC's authority over advertising, which predates FDAs, stems from Section 5 of the FTC Act, which gives the Commission authority to prevent companies from "using unfair methods of competition...and unfair or deceptive acts or practices in or affecting commerce." (FTC technically retains some legal jurisdiction over prescription drug advertising, but as a practical matter defers entirely to FDA, even for direct-to-consumer advertising.^d)

The FTC's organizational structure, regulatory approach and enforcement tools differ considerably from those of the FDA. This means that, in effect, there are different systems in place for regulating prescription drug advertising and OTC drug advertising. FDA and FTC work together in cases involving medical products. FTC does not have a medical staff, and any FTC action must be based on FDA-approved labeling. FDA and FTC operate under a Memorandum of Understanding dating back to 1971 and amended in 1992 that sets forth how the relationship between the two agencies will work.^e

FTC requires that a company have substantiation of any advertising or promotional claim. At any point, FTC may launch an investigation of an advertisement campaign for a

specific product, or, more commonly, for a category of products. An FTC investigation can originate from multiple sources—from any of the five commissioners or the semi-independent FTC staff. The commissioners must approve any enforcement actions recommended by the staff.

FTC procedures permit companies to engage in formal discussions with its staff before any matter is resolved. The usual resolution of an FTC investigation is the signing of a consent decree in which the company admits no wrongdoing but agrees never to do it again. FTC also has the authority, however, to seek a temporary restraining order or an injunction.

Over the years, FTC's enforcement has been directed to a large extent at advertising for fringe products, such as medical devices sold directly to the consumer that promise dramatic weight loss or some other body enhancement.

PRESCRIPTION DRUG MARKETING ACT

An important aspect of drug marketing is the provision of samples to physicians to provide to their patients. The Prescription Drug Marketing Act of 1987 is the statute that provides FDA with the authority to regulate drug marketing practices. It prohibits the diversion of prescription drug or biological products into illegitimate commercial channels. It also prohibits the sale of drug samples. The law provides that companies must obtain the signature of a physician for all drug samples. The law also requires record-keeping and careful storage of samples. Whenever a manufacturer discovers any diversion, it must report this to FDA.

Under the PDMA, physician requests for samples must be provided in writing. There must also be a signed receipt for the delivery of the samples. FDA regulations require drug companies to establish detailed, written procedures for product sampling programs, and to keep records of its sample distribution.

"ANTI-KICKBACK"

Another Federal law that applies to the marketing of medical products is the so-called "anti-kickback" law.^f This law, which dates from 1972, is not part of the FD&C Act; instead, it relates to Medicaid and Medicare statutes.

The "anti-kickback" statutes are intended to prevent overuse of and overcharging for medical products that are subject to government reimbursement either under

^c The three consent decrees were with Syntax, Kabi, and ICN. The \$50 million fine involved Genentech. See FDA Advertising and Promotion Manual for details.

^d FTC regulations are contained in 16 C.F.R. Part 251 and 255.

^e A copy of the Memorandum of Understanding, as amended, is in FDA Advertising and Promotion Manual.

^f 42 U.S.C., §1320a-7b.

Medicaid or Medicare (or other government-reimbursed health care programs). They prohibit any activity in which a medical products company may try to persuade, other than with sound therapeutic arguments, a physician or another health care provider or facility to use a particular product.

Specifically, the law prohibits a company from offering “any remuneration (including any kickback, bribe or rebate) directly or indirectly, overtly or covertly, in cash or in kind” to anyone to induce them to purchase a product or service for which reimbursement may be sought under Medicaid or Medicare. Regulations provide for “safe harbors.”

Many companies, in their internal training, include information about the “anti-kickback” statute. The Office of the Inspector General in the Department of Health and Human Services (HHS) enforces the “anti-kickback” law. HHS administers the Medicaid and Medicare programs. Any serious violations are referred for potential criminal prosecution to the Department of Justice; administrative sanctions also are possible.

LINGERING ISSUES

Many of the issues raised by FDA’s regulation of advertising and promotion remain subject to continued debate and lack resolution. These are issues that are important to resolve because until they are, there will be uncertainty in the marketplace and an uneven playing field from a regulatory standpoint.

Among the FDA’s policies that are most controversial are those that seek to restrict the dissemination by pharmaceutical companies of information from the scientific literature that is outside of FDA labeling. The issue revolves around the extent to which FDA can and should restrict such information, in an environment in which so much information is available from other sources such as the news media and the Internet.

Another issue is what are appropriate penalties for issuing violative materials. If a company issues information that is misleading, what are the appropriate remedies—e.g., letters to health professionals, corrective advertising? Further, and perhaps more importantly, what corrective measures are most effective? There has never been a study of the effectiveness of corrective measures imposed by FDA.

Still another issue is how FDA should regulate DTC advertising. How much information should the consumer receive about risks, and what is the best way to convey that information? A related issue is what role the “brief

summary” should play in the future. Does it still serve a useful purpose, or is it just an additional advertising expense for drug companies?

The Internet poses still another challenge to FDA regulation. FDA’s jurisdiction extends to promotional activities in the United States or directed at United States citizens. The Internet eliminates political borders and makes it impossible for FDA to regulate everything issued by drug companies on an international basis.

Issues like these are likely to remain for some time to come, as new means of communication evolve and as FDA, the Congress and the industry reevaluate the proper role and dimensions of regulation.

SUMMARY

The regulation of marketing practices by medical product companies—specifically their promotional programs—has become a highly specialized field within food and drug law. The rules are set forth in a variety of ways—in FDA regulations, in guidances and in enforcement actions, and are refined in public discussions by the regulators themselves.

The rules continue to evolve as new products are introduced and new means of communications become available, but basically a few principles apply. In essence, programs and materials must not only be truthful, they must also contain prominently appropriate risk information, and there may be no efforts by medical products manufacturers to persuade or induce health care providers to use products for uses not approved by FDA.

REFERENCES

1. Holmer, A.F. Direct-to-Consumer Prescription Drug Advertising Builds Bridges Between Patients and Physicians. *JAMA* **Jan. 1999**, 380–382.
2. Hollon, M.F. Direct-to-Consumer Marketing of Prescription Drugs. *JAMA* **Jan. 1999**, 382–393.
3. IMS Health, press release dated April 20, 2000; available at www.imshealth.com.
4. Pub. L. No. 75-717, 52 Stat. 1040 (1938).
5. *Supra*, note 4.
6. Pines, W.L., Ed. *FDA Advertising and Promotion Manual*; www.thompson.com Thompson Publishing Group: Washington, DC.
7. Kessler, D.A.; Pines, W.L. The Federal Regulation of Prescription Drug Advertising and Promotion. *JAMA* **Nov. 1990**, 2404–2415.
8. Pines, W.L. Three Principles that Govern FDA Advertising and Promotion Regulation. *Drug Information Journal* **1997**, 31 (1).

9. FDA Guidance for Industry: Consumer-Directed Broadcast Advertisement. August 1999; www.fda.gov/cder/guidance/index.htm.
10. Abrams, T.W. Director, FDA Division of Drug Marketing Advertising and Communication at Drug Information Association meeting Feb. 28, 2000; New York
11. Consent Decree Signed by Kabi Pharmaceuticals. In *FDA Advertising and Promotion Manual*; www.thompson.com Thompson Publishing Group: Washington, DC.
12. Washington Legal Foundation vs. Henney, 56 F Supp. 2d 81, CD D.C., 1999.
13. Pub. L. No. 105-115, 111 Stat. 2296 (codified at 21 U.S.C. §360.999, 403 (1998)).
14. Abrams, Supra note 12.
15. Correspondence Available in FDA Advertising and Promotion Manual. www.thompson.com Thompson Publishing Group: Washington, DC.
16. Pub. L. No. 105-115, 111 Stat. 2296 (codified at 21 U.S.C. §360.999, 403 (1998)).
17. Peck, C. Letter to New Drug Application Holders. In *FDA Advertising and Promotion Manual*; www.thompson.com Thompson Publishing Group: Washington, DC, July 1991.
18. Final Guidance on Industry-Supported Scientific and Educational Activities, 62 Fed. Reg. 64,075 (Dec. 3, 1997), 61 Fed. Reg. 52,800 (Oct. 8, 1996).
19. Abrams, Supra note 10.
20. 21 C.F.R. §314.550, 601.45.
21. FDA Supra note 9.

FURTHER READING

FDA Advertising and Promotion Manual, (published monthly); Thompson Publishing Group: Washington, DC.

Pina, K.R.; Pines, W.L. *A Practical Guide to Food and Drug Law and Regulation*; The Food and Drug Law Institute: Washington, DC, 1999.



Alternative Medicines

Kristi L. Lenz

Medical University of South Carolina, Charleston, South Carolina, U.S.A.

INTRODUCTION

Although alternative medicine has the oldest healing practices, it is regaining popularity against a background of rapidly increasing technology in conventional medicine. For conventional health care practitioners, accepting the validity of alternative practices can be difficult, particularly because so little alternative medical training is provided at conventional medicine colleges and universities.

For pharmacy practitioners, herbal medicine presents a contradiction. Pharmacognosy, the study of natural product medicines, is a historical field of study in pharmacy. In addition, many conventional medicines are derived from herbs and other plants. However, in recent years, pharmacy has abandoned its “roots” in favor of clinical practice, and most pharmacists have little knowledge of herbals. No matter the level of pharmacist knowledge or bias on whether herbals can be helpful, the fact of the matter is that more patients are using them. This is sometimes at the risk of significant toxicity. All health care providers, but particularly pharmacists, need to develop a knowledge base of herbal medicines in order to best care for patients.

DEFINITIONS

Alternative Medicine

Often called “complementary and alternative medicine,” this group of medical practices has also been termed “unconventional,” “unorthodox,” “unproven,” and even “quackery.”^[1] Because these terms have significant negative connotations, terms such as “alternative medicine” are preferred. Alternative medicine is not one form of medicine, but rather a diverse group of health practices that are outside of what is considered usual or conventional by the medical establishment.^[2] Alternative medicine spans the range of practices, from home remedies to manufactured products, from patient self-treatment to care by a skilled practitioner, from efficacious to potentially dangerous. Specific definitions of more common alternative medicine practices used in the United States are listed in Table 1.^[1]

Herbal Medicine

Although alternative medicine encompasses a very broad range of practices (more than 150 in fact), the area of most interest to pharmacy practitioners is herbal medicine. Interestingly, there is no definition of “herb” in any federal legislation or in any Food and Drug Administration (FDA) regulation.^[3] Definitions vary considerably depending upon the source. For example, botanists define an herb as a plant whose stem dies back in winter (vs. trees or shrubs). On the contrary, pharmacognosists define herbs as the aerial parts of plants (vs. seeds or roots).^[3] The Herbal Trade Association, a group that has economic interests in this definition, defines an herb as a plant, plant part, or extract thereof used for flavor, fragrance, or medicinal purpose.^[3] In any case, herbal medicine implies the use of the *whole plant or plant part* as a remedy, rather than a single, active constituent derived from a plant.

SIGNIFICANCE OF HERBAL USE

Prevalence of Use

Phone surveys were done in 1990 and 1997 by Eisenberg et al. to determine the prevalence of alternative medicine use in the United States. Over 1500 participants were surveyed in 1990, as well as over 2000 in the follow-up survey in 1997. The percentage of participants surveyed who reported using at least one form of alternative medicine in the preceding 12 mo was 34% in 1990 and 42% in 1997.^[4,5] This is the oft-quoted “1 in 3 Americans use alternative medicine” statistic. Herbal medicine use rose from only 2.5% in 1990 to 12% in 1997, making it one of the fastest growing alternative medical practices used in the United States.^[4,5]

Demographics of Users

According to the surveys conducted by Eisenberg et al., alternative medicine users tend to be educated, younger to middle-aged, female, and have chronic medical conditions. They self-refer 90% of the time, meaning

**Table 1** Complementary and alternative medical practices

Practice	Definition
Acupuncture	Ancient Chinese technique that uses needles to pierce the skin
Aromatherapy	Taps into a grid of flowing energy ("qi") that controls organ function
Ayurveda	Uses botanical oils and essences to treat both physical and psychological disorders "Life knowledge"
Bioelectromagnetics	Ancient Indian practice that uses diet, exercise, yoga, meditation, herbs, and massage to treat imbalances in physical, emotional, and spiritual harmony
Chiropractic	Study of living organisms and their interaction with electromagnetic fields
Herbal medicine	Belief that magnetic fields penetrate the body and heal damaged tissues
Homeopathy	Practitioners use manipulation to treat disorders of the spine, joints, and muscles Plants that are made into pills or extracts to prevent and cure physical and psychological disorders "Like cures like"
Mindfulness meditation	Belief that very small doses of substances that would at high doses cause adverse effects can be used to cure those effects
Naturopathy	Preparations may be so dilute that the active ingredient no longer remains
Osteopathy	Belief that the mind can influence health and control physiologic responses
Reflexology	Relies on diet, fasting, massage, herbs, homeopathy, and other natural treatments
Therapeutic touch	Practitioners use manipulation to expedite recovery from disease or injury Practitioners also receive conventional medical training and prescribe drugs Spots on the foot are massaged to stimulate specific organs Caregiver moves hands inches above the patient's body to realign disturbed energy fields or remove "blockages"

(From Ref. 1.)

that their primary care practitioners are not involved in the decisions to pursue alternative therapies.^[4-7]

The use of alternative medicine varies with the patient population. For example, up to 80% of cancer patients report use of alternative medicine vs. one-third in the general population. Surveys also find that two-thirds use herbal medicine.^[7-10] Interestingly, cancer patients do not abandon conventional therapies when using alternative medicine, with close to 90% using both conventional and alternative medicines together.^[4,6,9] Particularly with herbal medicine use, this creates the potential for conventional drug interactions. In fact, in Eisenberg's^[5] 1997 survey, 20% of participants admitted to using conventional prescription medications with herbal supplements. Because less than 10% of herbal medicine users are under the care of an herbalist, pharmacists are often the only health care providers who can help to avoid potentially dangerous adverse effects and interactions.^[10] Similar to cancer patients, a large proportion (45%) of AIDS patients also report alternative medicine use.^[11] Of those patients, two-thirds use herbal supplements, and most use alternative practices with their conventional HIV treatments.^[11]

Reasons given by patients as to why alternative medicine is used include the following: the perception that conventional therapies are ineffective and/or toxic; frustration when no effective conventional therapy exists; the desire to take a more active role in their own care;

distrust of conventional practitioners; and the belief that alternative practitioners focus on the whole patient.^[6,10] Although most patients still use conventional medical practices with alternative medicine, only one-third tell their conventional practitioner that they are doing so.^[4,5,8] This again means that the pharmacist may be the only conventional health care practitioner who is aware of both patients' conventional and alternative medicine use.

Cost

It is estimated that over \$20 billion are spent each year on visits to alternative practitioners.^[5] Sixty percent of patients pay all of these costs out-of-pocket, with only 20% of health maintenance organizations (HMOs) and third-party payers supplementing at least some of these health care costs.^[5] It is also estimated that \$1-5 billion are spent on herbal supplements per year, and interestingly, less than 5% of patients who use herbals report that they bought those products in a pharmacy.^[5]

What Pharmacists Know

Given the recent rise in herbal supplement use and the potential dangers of misuse, one would expect that the pharmacy profession would be prepared to deal with public's need for reliable information on herbs.



Unfortunately, this is generally not the case. A 1998 survey done by the University of Mississippi found that only 2% of pharmacists felt confident in their herbal medicine knowledge.^[12] Another survey in 1998 performed in North Carolina and Virginia included a 15-point test on five of the most common herbal supplements.^[13] Of the 164 participants, 68% worked in a community setting and 74% sold herbals. However, the mean score on the 15-point test was only 6.3, indicating that pharmacists have inadequate knowledge with which to advise patients taking herbals.^[13]

One would also hope that Colleges of Pharmacy would have increased the education they give to pharmacy students on herbal medicine, to prepare them better for the needs of consumers. This has also been quite slow to change. There is very little detailed literature in this area; however, a 1997 survey was done and asked Colleges about alternative medicine courses in the curricula. Although three-fourths responded that they taught about alternative medicine, only one-third of the courses were required, and less than 40% of content taught was about herbals.^[12] Comparison of this statistic to a survey of breast cancer patients found that 71% of herbal users thought that herbals were “perfectly safe,” a fact known to be untrue.^[10] Pharmacists will not be able to help patients avoid harmful effects from herbals unless they are educated about uses, toxicities, and potential herb–drug interactions.

REGULATION AND OVERSIGHT

National Center for Complementary and Alternative Medicine

The federal government has had some insight into the need for a federal office to oversee dissemination of information on alternative medicine. Previously known as the Office of Alternative Medicine (established in 1992), the National Center for Complementary and Alternative Medicine (NCCAM) was established in 1998 as part of the National Institutes of Health.^[1,2] The mission of NCCAM is to “give the public reliable information on the safety and effectiveness of complementary and alternative medicine,” with an emphasis on clinical trial sponsorship.^[2] The budget of the NCCAM has grown from \$2 million in 1992 to almost \$70 million in 2000, and almost half the budget is mandated to fund peer-reviewed grants.^[2] The NCCAM also helps to disseminate information to the public, sponsors literature evaluations, and acts as an international liaison for alternative medicine. The NCCAM actually divides alternative medicine into seven broad categories:

1) alternative systems of medical practice (such as ayurveda and Chinese medicine); 2) bioelectromagnetics; 3) diet and nutrition (such as macrobiotics); 4) herbal medicine; 5) manual healing methods (such as chiropractic and therapeutic touch); 6) mind/body techniques (such as yoga and meditation); and 7) pharmacological and biological treatments (such as shark cartilage).^[2] More specific information about NCCAM may be found at the website: <http://nccam.nih.gov/>.

Dietary Supplement Health and Education Act

One would assume that the federal government would also take a significant role in regulating substances (e.g., herbal therapies) that have the potential to cause significant harm. Although the FDA does want to hold many herbs to a higher standard, political pressure has led to the passage of legislation that makes herbal supplements more widely available with less oversight. This legislation is the Dietary Supplement Health and Education Act (DSHEA).

Prior to DSHEA, herbs were inconsistently regulated as drugs, foods, and/or food additives. Several events led to the passage of DSHEA. In 1990, 258 ingredients in over-the-counter drug products were banned for sale by the FDA due to inadequate efficacy data. Eighty-five of these ingredients were of herbal origin. Manufacturers responded by either pulling the products from the market or, more importantly, selling the products as “dietary” supplements.^[14] This caught the attention of the FDA, and in 1993, the then Commissioner David Kessler publicly proposed that herbal supplements be held to the same standards as drugs, that is, be proven safe and effective or be removed from the market. These comments led to much public concern about continued availability of herbals, and many letters were written to members of Congress. The political pressure was great enough that in 1994, DSHEA was passed.

The DSHEA was a bipartisan bill cosponsored by Orrin Hatch of Utah and Tom Harkin of Iowa. First, it broadened the definition of a “dietary supplement” to include any product designed to supplement the diet that contained one or more of the following: a vitamin, mineral, herb, botanical, amino acid, or any metabolic constituent or extract thereof. To be removed from the market, supplements must be proven unsafe by the FDA. This is in contrast to drugs, where the burden of proof (both safety and efficacy) is on the manufacturer. Supplements must contain the labeling “Not evaluated by the FDA. This product is not intended to diagnose, treat, cure, or prevent disease.” They must also be labeled either “dietary supplement” or “herbal supplement,” as well as have the common name, Latin binomial, quantity, and plant



part used on the label. Products may be sold as dietary supplements as long as no health or therapeutic claims are made. So, structure/function claims are permitted, and the label may state things such as “elevates mood” or “maintains cardiovascular health.” Disease claims, however, are prohibited. For example, supplement labels may not state “treats depression” or “lowers cholesterol.”^[3,14]

Limitations of Regulations

There are several limitations that lead to potential for patient harm with the current regulation of herbals under DSHEA. First, botanical nomenclature is not standardized. Common names for herbals may vary depending on the region of the country. A single herb may have more than a dozen common names (e.g., echinacea), or one common herbal name may refer to several different species (e.g., yellowroot and snakeroot).^[14] To avoid confusion, the American Herbal Products Association (AHPA) has published *Herbs of Commerce*, a text of more than 500 herbs with the preferred common name, Latin binomial, and appropriate synonyms.

Second, good manufacturing practices (GMPs) of herbal supplement manufacturers are not regulated by the FDA. This means that there is no guarantee that what is on the label of the supplement is actually what is in the bottle. An example of this is a study published by Gurley et al.^[15] examining the ephedra alkaloid content of 20 herbal weight-loss supplements. Ten of the 20 supplements assayed (50%) had more than a 20% discrepancy between the actual ephedra content and the labeling. Four products also had significant lot-to-lot variation, up to 1000% of labeled ephedra content in one. Interestingly, one product contained no ephedra alkaloids at all.^[15]

Third, safety assurance is the responsibility of the FDA, not the manufacturer. This means that the FDA must accumulate a significant amount of proof that something is unsafe before pulling it from the market. In addition, to help professionals, the AHPA has published *The Botanical Safety Handbook* that summarizes safety data of more than 600 herbs. However, there is no guarantee that any particular manufactured product is safe. An example is a recent recall of PC-SPES supplements by the FDA.^[16] The PC-SPES is a combination of eight herbs marketed as a treatment for prostate cancer. It is known to have estrogenic activity, which is presumably how it affects prostate cancer.^[17] However, laboratory analysis performed at the California Department of Health found that the recalled products were contaminated with warfarin.^[16] The implications of this are that even if PC-SPES is thought to be safe, any given product could contain other potentially harmful substances.

Finally, efficacy of herbal supplements is difficult to establish. Many reported uses for herbs are hundreds of years old and anecdotal in nature. Because most herbal supplements cannot be patented, large pharmaceutical firms are reluctant to spend the money that is necessary to conduct randomized, placebo-controlled trials. To overcome this problem, in 1978, the German government established Commission E to evaluate the safety and efficacy of herbs. To date, more than 300 have been evaluated, and the monographs were translated into English in 1998. Unfortunately, most consumers do not have access to these monographs, and the language in DSHEA simply adds to the potential for patient harm, with structure/function claims leading to vague and potentially dangerous messages to consumers. For example, echinacea is often marketed as a supplement to “boost” the immune system. This may lead patients with HIV disease to believe echinacea would enhance their impaired immunity. Actually, the opposite is true: effects of echinacea may actually *decrease* CD4 cell counts in HIV patients, leading to increased risk of infection.^[18]

COMMONLY USED HERBAL SUPPLEMENTS

Many dozens of herbal supplements are available to be purchased over-the-counter. Discussion of all these herbal supplements is beyond the scope of this article. Listed below is a brief discussion of each of the more commonly available and used supplements, including highlights of known active constituents, pharmacology, dosing, supporting clinical literature, adverse effects, and interactions with conventional medications. The known and potential drug–herb interactions are summarized in Table 2.^[18–22]

Black Cohosh (*Cimicifuga racemosa*)

Black cohosh is also known as black snakeroot, bugbane, bugwort, rattleroot, and rattleweed. It should not be confused with blue cohosh, a uterine stimulant historically used to induce labor.^[18] Black cohosh is used to treat menopausal symptoms, including hot flashes, excessive sweating, emotional lability, and sleep changes.^[23] The exact active constituents are not known but are thought to include the *triterpene glycosides* (such as actein, 27-deoxyactein, deoxyacetylactol, and cimicifugoside), *phytosterins*, and *isoflavones* found in root extracts.^[18,23] For many years, practitioners thought that black cohosh must be estrogenic, but more recent animal and human data suggest that its pharmacology does not involve acting as an estrogen.^[23] This was confirmed by a recent controlled trial in breast cancer patients with

**Table 2** Reported and potential herb–drug interactions

Herb	Interacting drug or drug class	Effect
DHEA	Antidiabetic agents ^a	Decreased hypoglycemic effects
Garlic	ASA, NSAIDs	Additive antiplatelet effects
	Clopidogrel, ticlopidine	Additive antiplatelet effects
	Warfarin	Increased risk of bleeding
Ginger	ASA, NSAIDs	Additive antiplatelet effects
	Clopidogrel, ticlopidine	Additive antiplatelet effects
	Warfarin	Increased risk of bleeding
Ginseng	ASA, NSAIDs	Additive antiplatelet effects
	Antidiabetic agents	Additive hypoglycemia
	Clopidogrel, ticlopidine	Additive antiplatelet effects
	CNS stimulants, caffeine	Additive CNS toxicity
	Corticosteroids	Additive CNS toxicity
	Digoxin	Falsely elevated levels
	MAO inhibitors	Increased toxicity
Ginkgo	Warfarin	Increased risk of bleeding
	Anticonvulsants	Decreased antiseizure effects
	ASA, NSAIDs	Additive antiplatelet effects
	Clopidogrel, ticlopidine	Additive antiplatelet effects
	Warfarin	Increased risk of bleeding
Green tea	Warfarin	Decreased anticoagulant effects
Hawthorne	Antihypertensives	Additive hypotension
	Digoxin	Potential of (+) inotropic effects
Kava	CNS depressants, ethanol	Additive sedation, risk of coma
	Hepatotoxins	Additive hepatotoxicity
Licorice	Antihypertensives	Antagonism of hypotensive effects
	Corticosteroids	Additive mineralocorticoid effects
	Digoxin	Risk of toxicity due to hypokalemia
	Diuretics	Additive hypokalemia
Ma huang	CNS stimulants, caffeine	Additive CNS stimulation
	Digoxin	Additive toxicity
	MAO inhibitors	Hypertensive crisis
Melatonin	CNS depressants, ethanol	Additive sedation
St. John's wort	Cyclosporine	Decreased levels and decreased effect
	Digoxin	Decreased levels and decreased effect
	Indinavir, nevirapine	Decreased levels and decreased effect
	MAO inhibitors	Increased risk of MAO toxicity
	Oral contraceptives	Decreased levels and decreased effect
	Simvastatin	Decreased levels and decreased effect
	SSRIs	Increased risk of serotonin syndrome
Valerian	Warfarin	Decreased levels and decreased effect
	CNS depressants, ethanol	Additive sedation

^aAntidiabetic agents include drugs such as insulin, glipizide, glyburide, and metformin.

ASA = aspirin; NSAIDs = nonsteroidal anti-inflammatory drugs, such as ibuprofen, naproxen, and diclofenac;

CNS stimulants include drugs such as pseudoephedrine, dextroamphetamine, theophylline, and caffeine;

MAO = monoamine oxidase; CNS depressants include drugs such as benzodiazepines, barbiturates, and ethanol;

SSRIs = selective serotonin reuptake inhibitors, such as fluoxetine, sertraline, and paroxetine.

(From Refs. 18–22.)

treatment-related hot flashes found no difference in efficacy when compared to placebo.^[24] However, the compound was considered to be safe in these patients. No changes in luteinizing hormone (LH) or follicle stimulating

hormone (FSH) levels were noted between the two groups, indicating limited estrogenic activity.^[24] The most studied doses range from 20 mg to 40 mg twice daily.^[18,23] Initial effects may be seen within 2 weeks, but maximal benefit



generally takes 8 weeks of continued therapy.^[23] Black cohosh is quite well tolerated, with only mild gastrointestinal (GI) effects noted. Long-term data is not available, and patients should be advised to limit use to six months.^[18] Again, the concern that black cohosh may stimulate breast or endometrial cancer cells by acting as an estrogen is not founded based on recent data.^[23,24] However, this herb should be avoided in pregnancy as miscarriages have been reported. Black cohosh has no known herb–drug interactions.^[18] However, postmenopausal patients should be advised that black cohosh is not a direct substitute for estrogen replacement therapy, as it has undetermined benefits for osteoporosis and cardiovascular disease.^[23]

Dehydroepiandrosterone (DHEA)

Dehydroepiandrosterone is not truly an herb but rather a natural steroid product of the adrenal glands. In addition, DHEA is the steroid precursor to 50% of androgens in men, 75% of estrogens in premenopausal women, and close to 100% of estrogens in postmenopausal women. Levels peak between age 20 and 30, then decline approximately 2% per year thereafter.^[25] Because levels are known to decline with age, DHEA is often marketed as an “antiaging” supplement. There are absolutely no clinical trials published examining this effect. In addition, DHEA levels are known to be lower in many chronic disease states, including some cancers, cardiovascular disease, systemic lupus erythematosus (SLE), Alzheimer’s, and progression of HIV disease. Dehydroepiandrosterone has been studied in small trials as a treatment for fatigue in HIV patients as well as for treatment of depression in middle-aged patients and for SLE.^[25] Replacement doses in patients known to be deficient (due to long-term corticosteroid use or chronic disease) are 20 mg–50 mg per day in men and 10 mg–30 mg per day in women.^[25] Doses for the other indications listed above are much higher, ranging from 200 mg to 500 mg per day.^[25] Adverse effects are directly related to increased androgen production and include acne, insomnia, irritability, and hirsutism.^[18,25] Perhaps, the most serious potential adverse effect of DHEA is stimulation of hormone-dependent cancers such as prostate, breast, and endometrial. Because of this potential, patients with known risk factors or a personal history of these tumors should not take DHEA.^[18,25] It has not been reported to cause herb–drug interactions. However, it is a mild inhibitor of the cytochrome P450 3A4 isoenzyme system and may increase concentrations of metabolized drugs to a minor extent.^[18]

Echinacea (*Echinacea angustifolia*, *pallida*, and *purpurea*)

Echinacea has a host of other common names, including American coneflower, black Sampson, black Susans, comb flower, Indian head, Kansas snakeroot, purple coneflower, red sunflower, survey root, and Sampson root.^[18] It is used as a nonspecific immune stimulant and has several active constituents. Caffeic acid derivatives and high-molecular weight polysaccharides stimulate phagocytosis of macrophages and natural killer cells. They also increase production of tumor necrosis factor-alpha (TNF-alpha) and interleukin-1. In addition, echinacea contains alkylamides that have anti-inflammatory properties.^[26] The portions of the herb used are the dried root or the fresh juice from the root and aerial parts. The form most commercially available is the dried root in capsules. Because the active constituents are not water soluble, tinctures and teas are not likely to be as effective.^[18] The best form of echinacea is likely an alcoholic extract of the root, either 1:1 or 1:5 in 45% ethanol. Doses are usually 1 ml–2 ml taken 3–4 times a day. The most studied indications include prevention of viral upper respiratory infections (URIs) in patients at high risk, as well as adjunct treatment of URIs, to decrease severity of infection. The strongest data are with treatment, rather than prevention, and a recent review of randomized trials published since 1997 found two negative and three positive. The positive trials found that echinacea decreased the frequency, duration, and magnitude of URI symptoms.^[27] For this indication, patients should be advised to begin echinacea at the onset of symptoms and continue therapy until 24 hr after symptom resolution.^[26] For prevention, studies have examined taking echinacea daily for 8 weeks during the “cold and flu” season; however, this indication is not generally recommended due to inadequate data.^[18] Adverse effects are rarely reported, although patients may experience allergic reactions.^[26] There are also a few reports of immunosuppression with more than 8 weeks of continuous use, and long-term use should be avoided. There are no known herb–drug interactions with echinacea, but there are several herb–disease state contraindications. HIV patients should avoid using echinacea, because it may increase TNF-alpha levels, which in turn decreases CD4 counts. Also, due to the potential for immunostimulation, patients with autoimmune diseases should not take echinacea (e.g., multiple sclerosis, collagen–vascular diseases).^[18,26] Of interest, there is one controlled trial in women who took echinacea during pregnancy. More than 400 women were monitored for pregnancy outcome: 200 who took echinacea and 200 who



Alternative Medicines

did not. There was no difference in the rate of spontaneous abortion or the rate of fetal malformations, indicating echinacea may be safe for limited use during pregnancy.^[28] No other herbal product has such conventional literature published for use and exposure during pregnancy.

Evening Primrose Oil (EPO, *Oenothera biennis*)

Evening primrose oil is named for the flowers on the plant that open in the late afternoon. It is also known as king's cure-all.^[18] The seeds contain 7%–10% gamma-linoleic acid (GLA), an essential omega-6 fatty acid. The body relies on the metabolic conversion of linoleic acid to GLA, and this conversion is defective in a number of disease states, including fibrocystic breast disease, diabetic neuropathies, and various skin disorders (such as atopic dermatitis and eczema).^[18] The EPO has been studied in numerous small trials for these indications as well as for premenstrual syndrome. Trials in various breast diseases have been positive, although it may take 6 mo or more to see effects. Doses range from 2 g to 4 g EPO (which is 200 mg–400 mg GLA) per day in 2–3 divided doses. Doses for neuropathy treatment are generally higher.^[18] At all these doses, EPO is well tolerated, producing only mild GI upset and headache. Patients should take doses with food. In addition, EPO has no known contraindications and no reported herb–drug interactions.^[18]

Garlic (*Allium sativum*)

The active component of garlic is formed when alliin, a compound in the bulbs, is converted by the enzyme allinase to allicin.^[18] During food preparation, this conversion takes place when the bulbs/cloves are crushed. The conversion may also take place in the GI tract, although the conversion is reduced in the presence of stomach acid. Fresh garlic intake is considered to be the most efficacious way to consume this herb for disease–treatment indications, although many forms of tablets, capsules, tincture, and juices are available. For nonfresh intake, tinctures and oils should be avoided due to instability of allicin. Enteric-coated tablets, prepared by drying the crushed bulbs and then compressing them into tablets, are also likely effective, as they bypass the effect of stomach acid.^[18] Garlic has been most studied for the treatment of dyslipidemias and hypertension. Its effects on blood pressure are not well substantiated; however, in a recent meta-analysis of 13 randomized trials, garlic was found to produce modest reductions (20%) in total cholesterol and triglycerides.^[29] Doses used range from

600 mg to 900 mg per day, which is roughly equivalent to 4 g of fresh garlic cloves per day.^[29] In addition, a recent review of 45 trials found not only reductions in total cholesterol, low density lipoprotein (LDL) cholesterol, and triglycerides with garlic therapy, but also noted decreases in platelet aggregation. No effects on blood glucose were seen.^[30] Garlic has minimal adverse effects, with the exception of GI effects (such as heartburn and flatulence) and the characteristic body odor. Contact dermatitis has been reported with extensive handling of the fresh cloves.^[18] Again, garlic inhibits platelet aggregation that may cause herb–drug interactions with other antiplatelet and anticoagulant drugs.^[18]

Ginger (*Zingiber officinale*)

The part of this herb used is the root that contains 1%–3% gingerols. When the root is dried, the gingerols are converted to shogaol and zingerone, but the clinical significance of this conversion is not known.^[31] The gingerols are thought to stimulate upper GI motility, similar to the effects of metoclopramide.^[31] This gives rise to ginger's most common indication, that of an antiemetic. Ginger has no known central effects on the chemoreceptor trigger zone in the brain. Dosing for motion sickness and morning sickness is 1 g–4 g per day fresh ginger or 500 mg–1000 mg as dry powdered root. Daily intake is generally divided into 2–4 doses. For motion sickness, ginger has been shown to be superior to placebo and even to dimenhydrinate in some small trials.^[31] A recent randomized trial in 70 pregnant women with morning sickness also found ginger to be superior to placebo for reducing both nausea and vomiting.^[32] Ginger has also been studied in postoperative nausea/vomiting but with less positive results. It has also been marketed as a treatment for osteoarthritis, but the data for this indication are weak at best.^[18,31] No significant adverse effects have been observed with ginger. It may be a mild inhibitor of platelet aggregation, and therefore, the potential for herb–drug interactions exist.^[18] These antiplatelet data are not as strong as for garlic, ginseng, and ginkgo.

Ginkgo (*Ginkgo biloba*)

The ginkgo tree is also called the kew tree, maidenhair tree, and fossil tree.^[18] It is the oldest known tree species, and the largest commercial farm of ginkgo trees is in Sumter, South Carolina. The most common indications for ginkgo are related to increasing mental capacity, as well as to treat cerebral and peripheral vascular disease. This may be due to touted benefits of flavonoids and terpenoids (ginkgolides and others) that may inhibit

platelet-activating factor, increase cerebral circulation, inhibit arterial spasms, decrease capillary permeability and fragility, and improve brain tolerance to hypoxia. They may also act as free radical scavengers.^[18] It is recommended that extracts from the seeds and leaves be used that have been standardized to 24% flavonoids and 6% terpenoids. The most studied tablets contain 40 mg of this extract. Use for general improvements in cognitive function are not well studied; however, there have been several trials for Alzheimer's-type dementia. Doses from 40 mg to 80 mg 3 times daily have been shown to improve cognitive skills, although results may take 3 mo or longer to be seen.^[18,33] In addition, a recent meta-analysis examined the results of eight randomized trials treating intermittent claudication with ginkgo 40 mg 3–4 times daily. This dose significantly improved pain-free walking distances.^[34] The reported adverse effects of ginkgo include nausea/vomiting, diarrhea, and seizures with overdose or excessive intake of the seeds (rather than the leaves).^[18] There have also been case reports of excessive postoperative bleeding, likely due to the antiplatelet effects of ginkgo.^[35] Due to these antiplatelet effects, there is the potential for herb–drug interactions with conventional anticoagulant and antiplatelet agents.^[18]

Ginseng

The herb known as “ginseng” is representative of the dangers of inconsistent nomenclature. There are actually three herbs that are known by this common name: American or western ginseng (*Panax quinquefolius*), Asian or Chinese ginseng (*Panax ginseng*), and Siberian ginseng (*Eleutherococcus senticosus*), a very different albeit related plant species.^[18] For the purposes of this text, ginseng refers to both American and Asian, as they have very similar therapeutic and adverse effects. The active components in ginseng are 12 ginsenosides isolated from the root.^[18] Ginseng is marketed to enhance both physical and mental performance, and certainly the ginsenosides are CNS (central nervous system) stimulants. However, data for the therapeutic effectiveness of these compounds are limited. A recent trial in 83 otherwise young healthy adults (mean age 26 years) compared two doses of ginseng (200 mg and 400 mg per day) to placebo.^[36] There were no differences noted in psychological well-being.^[36] Adverse effects reported for ginseng are numerous and include chest pain, palpitations, hypertension, headache, insomnia, irritability, nervousness, hypoglycemia, impotence, and GI effects (such as nausea, vomiting, and diarrhea). Reported and potential drug interactions are related to ginseng's antiplatelet and

CNS stimulant effects.^[18] Due to these adverse effects that occur more frequently than with other previously discussed herbals, ginseng should be used cautiously, particularly in patients with a history of cardiovascular or CNS disease.

Glucosamine/Chondroitin

Glucosamine is an endogenous aminomonosaccharide used in the synthesis of proteoglycans in cartilage, which are depleted in osteoarthritis.^[37] Current conventional treatment of osteoarthritis is supportive only and includes treatment with acetaminophen or nonsteroidal anti-inflammatory drugs (NSAIDs). These drugs decrease the pain and inflammation associated with the disease, but they do not change the overall disease process.^[37] Glucosamine, on the other hand, may increase cartilage production, or at least slow breakdown, as well as provide mild anti-inflammatory effects.^[37] Chondroitin is extracted from the cartilage in bovine trachea and may increase collagen synthesis in cartilage.^[18] Short-term trials with each agent used individually have shown modest results compared to placebo and NSAIDs.^[18] A meta-analysis of 15 randomized, placebo-controlled trials was positive overall in favor of glucosamine/chondroitin, but the authors remarked that bias was likely as almost all the trials were sponsored by the manufacturer.^[38] They conceded that efficacy in osteoarthritis appears “probable.”^[38] Unfortunately, the additional benefit of chondroitin to oral tablets is minimal at best, because glucosamine is 95% absorbed but chondroitin has minimal oral bioavailability.^[18] The combination is well tolerated, producing only mild headache, nausea/vomiting, and occasional rash.^[18] The Arthritis Foundation does not recommend the use of this supplement, but due to few adverse effects, a trial in patients intolerant to other conventional analgesics would not be unreasonable. Doses are often weight-based and range from 1000 mg to 2000 mg per day, divided.^[18] Unfortunately, there are no long-term studies published with this long-term condition. Animal data have reported hyperglycemic effects, so diabetic patients should monitor blood glucose levels closely.^[18]

Grapeseed (*Vitis vinifera*)

Grapeseed extracts from the seeds of the herb contain flavonoids, as well as some essential fatty acids and tocopherols. The flavonoids are considered as the primary active constituents, inhibiting lipid peroxidation. Tocopherols, related to vitamin E, are also antioxidant in



Alternative Medicines

nature.^[18] The most studied patient population with grape-seed is cardiovascular disease patients, particularly patients with dyslipidemias. Doses range from 25 mg to 300 mg daily. No adverse effects or drug interactions have been reported for grape-seed, although conventional clinical data are limited.^[18]

Green Tea (*Camellia sinensis*)

Green tea is made in a curing process that withers and then ferments the leaves of this herb before drying them. This process is thought to preserve the activity of the polyphenols contained in the fresh leaves. Most teas are standardized to contain 60% polyphenols, which are purported to have antioxidant, lipid-lowering, and potentially anticancer effects.^[18] Green tea also contains 1%–4% methylxanthines, including caffeine.^[18] Clinical trials looking at the efficacy of green tea are few, but it is thought that significant consumption (6–10 cups per day) is necessary to see effects.^[18] A recent phase 1 trial in patients with refractory solid tumors found no objective responses with 5% stable disease.^[39] The dose-limiting toxicities were caffeine-related, and the maximum tolerated dose was equivalent to 7–8 cups per day of green tea.^[39] Dairy products may inhibit the polyphenols in green tea, so concurrent consumption should be avoided. The only reported adverse effects are rare allergic reactions. Patients should be advised, however, of the modest caffeine content of green tea, particularly because many herbal teas are “caffeine-free.”^[18] Green tea may contain significant vitamin K content as well; therefore, the effects of warfarin may be inhibited.^[40]

Kava (*Piper methysticum*)

Kava is a member of the black pepper family and is also known as ava, awa, kew, sakaw, tonga, and yagona.^[18] The root contains kavapyrones that inhibit the limbic system, and they are sedating and may elevate mood to a minor degree. Kava is marketed as an anxiolytic and as treatment for depression and insomnia.^[41] Doses used range from 60 mg to 120 mg daily at bedtime.^[18] Unfortunately, several potentially dangerous effects have been recently attributed to kava. Clearly, sedation, visual changes, and decreased motor reflexes are potential toxicities. Coma has even been reported with concurrent use of kava and alprazolam.^[41] In addition, the FDA is investigating at least 25 cases in Europe of hepatotoxicity attributed to kava.^[16] Caution should be exercised when using kava with other CNS depressants, especially benzodiazepines, or hepatotoxins. Kava does not appear to cause psychological dependence.^[18]

Ma Huang (*Ephedra sinica* and *nevadensis*)

Ephedra has also been called Brigham tea, desert tea, herbal fen-phen, joint fir, Mexican tea, Mormon tea, natural ecstasy, popotillo, sea grape, squaw tea, teamster tea, and yellow horse.^[18] The active component, ephedrine, was isolated in the 19th century and was used extensively in the 20th century in nasal decongestants and as a CNS stimulant. Extracts of ephedra seeds and stems contain 0.5%–2.5% alkaloids, not only including ephedrine, but also pseudoephedrine, methylephedrine, and norephedrine. All are CNS stimulants, with effects similar to amphetamines and caffeine.^[18] Although ephedra is no longer available in conventional nasal decongestants due to the risk of toxicity, it is commonly found in many herbal weight-loss products. This is due to its ability to suppress the appetite and increase metabolic rate. Ephedra is available as capsules, tablets, teas, and tinctures. The FDA advises consumers not to take more than 24 mg in 24 hr and to limit use to no more than 7 days.^[18] The primary problem with ephedra is the high potential for toxicity. It is a CNS stimulant, and as such, can cause anxiety, confusion, headache, insomnia, and irritability. These effects are greatly increased when ephedra is taken with caffeine. In addition, ephedra has significant cardiovascular effects, including reported hypertension, tachycardia, myocardial infarction, and cardiac arrest, usually when the labeled maximum daily dose is exceeded. These potentially fatal effects have been reported even in previously healthy young adults. A recent review of 98 events thought to be possibly or probably related to ephedra intake in a 2-yr period (1997–1999) found 47% cardiovascular effects (primarily hypertension, as well as palpitations and tachycardia), as well as 18% CNS effects, including strokes and seizures. Ten deaths were reported.^[42] All of these adverse effects also lead to potential herb–drug interactions with other stimulant drugs. In addition, patients with glaucoma, cardiovascular disease, thyroid disease, psychiatric conditions, diabetes, or prostate disease should avoid using ephedra. Pregnant women should also not take this supplement due to the potential for uterine contractions and preterm labor.^[18] The FDA is still investigating all reports of serious effects (more than 800 to date) due to ephedra and may ban this product from the market in the future.^[18]

Melatonin (N-Acetyl-5-methoxytryptamine)

Melatonin is not an herb but rather is a naturally occurring hormone produced in the pineal gland from tryptophan. It is also commercially synthesized. Endogenously,



melatonin is released during sleep periods, and levels are low during the day.^[18,43] As a supplement, melatonin has been most studied as a preventative/treatment for jet lag and as a sedative-hypnotic. For jet lag, patients are advised to take 5 mg daily at bedtime beginning three days prior to travel and continuing for three days after travel is complete.^[18,43] Levels of melatonin decrease with age, so it has been studied in elderly patients with sleep disorders. This patient population is more sensitive to the effects of melatonin; therefore, they should start with lower doses, usually 1 mg–2 mg at bedtime.^[43] Patients with liver disease should use melatonin cautiously, as clearance of the compound is impaired.^[43] Adverse effects reported with melatonin use include headache, confusion, sedation, and mild hypothermia. Concurrent use of CNS depressants, including alcohol, should be avoided.^[18,43]

Milk Thistle (*Silybum marianum*)

Milk thistle is also known as Our Lady's thistle, Mary thistle, Marian thistle, and St. Mary's thistle.^[18] The active constituent, silymarin, has been isolated and the structure determined for many years. Only the fruit of the herb contains the active ingredient, 1%–4% as a mixture of three related compounds: silibinin, silidianin, and silychristin.^[44] Silymarin may act to stabilize hepatocyte membranes, as well as stimulate RNA polymerase, aiding liver regeneration after cell damage. It is also thought to be a free radical scavenger.^[18] Interestingly, milk thistle is the only known antidote for poisoning by the "death cap" mushroom, *Amanita phalloides*. The dosing for this poisoning is intravenous silymarin, which is only available in Europe.^[18] Milk thistle is also marketed and studied as a treatment for acute and chronic alcoholic and viral hepatitis, as well as a treatment for hepatotoxicity from drugs such as haloperidol and prochlorperazine.^[44] Decreased complication rates have been observed in acute viral hepatitis, as well as improvement in liver function tests in alcoholic hepatitis. However, no survival benefit with milk thistle has been observed.^[44] The most studied dose for liver diseases is 140 mg silymarin three times daily, which are 200 mg milk thistle extract doses standardized to 70% silymarin.^[18] Teas should be avoided as a dosage form due to silymarin being practically water-insoluble.^[44] Milk thistle has no known herb-drug interactions and causes minimal toxicity, including diarrhea and occasional allergic reactions. Milk thistle should be avoided in pregnancy due to its ability to cause uterine contractions.^[18]

S-Adenosylmethionine (SAME)

This is an endogenous substance produced from adenosine triphosphate and the amino acid methionine. It is naturally involved in a range of biological processes.^[45] As a supplement, it is most studied in the treatment of depression, as well as treatment of liver disease. In a recent meta-analysis, doses of 400 mg–1600 mg per day were found to be superior to placebo and as efficacious as moderate-dose tricyclic antidepressants.^[45] A few small trials in patients with liver disease have demonstrated improvement in liver enzymes with doses of 1600 mg per day or more.^[45] Due to limited bioavailability, SAME was previously only available intravenously. Enteric-coated tablets are now available as well, but bioavailability studies are lacking. Doses should be titrated, starting with 200 mg once daily and increasing by 200 mg increments every 1–2 weeks.^[45] The SAME is well tolerated, with only mild GI distress noted as an adverse effect and no known drug interactions.^[45]

Saw Palmetto (*Serenoa repens*)

Saw palmetto is also known as the cabbage palm and American dwarf palm tree.^[18] The ripe berries are thought to contain the active ingredients (perhaps fatty acids and sterols); however, the exact active components are not known.^[18] Saw palmetto is mainly used in the treatment of benign prostatic hypertrophy^[46] due to postulated inhibition of 5-alpha-reductase, the same mechanism as the conventional drug finasteride. The dose most studied is 160 mg twice daily. In a recent meta-analysis of 18 randomized trials, saw palmetto improved urologic symptoms to a rate similar to finasteride and better than placebo.^[46] Saw palmetto is well tolerated, causing mild abdominal pain, nausea/vomiting, and occasional decreased libido.^[46] Due to the risk of birth defects, pregnant women should avoid the use of saw palmetto.^[18] No known herb-drug interactions are reported; however, saw palmetto will reduce prostate-specific antigen (PSA) levels. This test is used to screen for prostate cancer; therefore, men at risk of this disease should avoid saw palmetto due to the risk of a false negative result.^[18]

St. John's Wort (*Hypericum perforatum*)

This herb is also known as St. Joan's wort, klamath weed, and goatweed.^[18] It has historically been used for many purposes, but most recently it is marketed as an antidepressant. In fact, it outsells all conventional antidepressants in Germany. The active constituent is hypericin that seems to act as a weak monoamine oxidase



MAO inhibitor and a selective serotonin reuptake inhibitor (SSRI). Dopamine and norepinephrine uptakes are also mildly inhibited.^[18] St. John's wort is available in many forms, as a tablet, tea, tincture, and the raw dried herb. For best results, a tablet standardized to contain 0.3% hypericin should be taken; Kira[®] by Lichtwer Pharma is the most extensively studied.^[18] Randomized, placebo-controlled trials using 300 mg of St. John's wort three times daily have found it to be superior to placebo in mild to moderate depression. Response rates are generally regarded as inferior to conventional antidepressants, including tricyclic antidepressants and SSRIs.^[47,48] In addition, St. John's wort is not without toxicity. Reported adverse effects include dizziness, headache, sleep changes, restlessness, dry mouth, and photosensitivity.^[18] Perhaps most significant are the reported herb-drug interactions. St. John's wort is a significant inducer of the cytochrome P450 3A4 isoenzyme system, which is responsible for metabolizing up to 60% of conventional drugs. There are several case reports of clinically significant decreases in serum cyclosporine concentrations, leading to transplanted organ rejection.^[49,50] Also, decreases in indinavir concentrations have been reported, which could potentially lead to HIV treatment failures and resistance.^[51] Other drugs metabolized via 3A4, including warfarin, digoxin, oral contraceptives, and simvastatin could potentially have levels decreased by St. John's Wort.^[18] Finally, because of its weak MAO inhibition and serotonin re-uptake effects, St. John's wort should not be taken concurrently by patients on MAO inhibitors or SSRIs.^[18]

Valerian (*Valeriana officinalis*)

Valerian has also been called All Heal, amantilla, Baldrianwurzel, and setwell.^[18] The root extract contains iridoid triesters (valepotriates) that stimulate the release of gamma-aminobutyric acid (GABA).^[52] This pharmacology is similar to that of the benzodiazepine sedatives.^[18] Animal studies confirm this pharmacology, as valerian attenuates benzodiazepine withdrawal symptoms in rats.^[52] As a sedative/hypnotic, valerian doses range from 400 mg to 900 mg taken at bedtime. Placebo-controlled trials are small and have mixed results. Some have shown increased quality of sleep and decreased sleep latency, while others have shown no difference vs. placebo.^[52] Data for using valerian as an anxiolytic are equally weak. Patients desiring to take valerian prior to an anxiety-producing event should be advised of the questionable efficacy and that the doses are much smaller than as a sedative, 100 mg taken 90 min prior to the event.^[52] Adverse effects include sedation, visual changes,

headache, rare allergic reactions, nausea/vomiting, and case reports of hepatotoxicity. Use with other CNS depressants, including alcohol, should be avoided.^[18]

Potentially Unsafe Herbs

Any herb may be unsafe or cause serious adverse effects when used incorrectly. The herbs listed in Table 3, however, have been more commonly reported to cause serious or life-threatening effects. Many are on the FDA's official list of unsafe herbs.^[18]

Many unsafe herbs are anticholinergic in nature, often due to significant hyoscyamine (and to a lesser extent, scopolamine) content. This includes belladonna, the nightshades, henbane, jimsonweed, and mandrake. These herbs cause a constellation of symptoms often referred to as "Hot as a hare, blind as a bat, dry as a bone, red as a beet, mad as a hatter." This includes confusion, hallucinations, agitation, elevated temperature, hypertension, tachycardia, mydriasis, dry mucous membranes, dry/flushed skin, and nausea/vomiting. Ultimately, respiratory arrest, seizures, and life-threatening arrhythmias can occur.^[53]

One herb, the calabar bean, actually causes cholinergic toxicity (as seen with pesticide overdoses) due to the physostigmine content in the ripe seeds. This toxicity includes bradycardia and hypotension, potentially leading to cardiac and respiratory arrest.^[18]

Several herbs have significant content of cardiac (digitalis) glycosides, including A Scotch broom, Canadian hemp, hedge mustard, Lily of the Valley, monkshood, wallflower, and foxglove, of which the conventional medication digoxin is derived. These herbs can cause bradyarrhythmias and heart block.^[18]

Other potentially unsafe herbs have varying degrees of neurotoxicity, GI toxicity, and hepatotoxicity. Many cause multiorgan toxicity. Nux vomica contains strychnine in the seeds and bark. Jalap and castor bean are cathartic laxatives. Callamus, chapparal, and comfrey are potential hepatocarcinogens.^[54] Autumn crocus contains colchicine, and lobelia has nicotine-like effects. All of these herbs should be avoided, and their high potential for toxicity re-enforces the idea that "natural" certainly does not always mean "safe."^[18,53,55,56]

PATIENT ASSESSMENT AND COUNSELING TIPS

When assessing a patient who wants to begin an herbal supplement, both past medical history as well as concurrent conventional medication used should be taken into account for potential interactions, as discussed

**Table 3** Potentially unsafe herbs

Harmful effects	Common name	Latin binomial	Toxic constituents	Serious adverse effects	Comments
Anticholinergic	Belladonna	<i>Atropa belladonna</i>	0.3%–0.6% hyoscyamine in leaves and root	Anticholinergic toxicity	Toxic effects with 5 mg–50 mg
	Bittersweet nightshade	<i>Solanum dulcamara</i>	Solasonine in stem and unripe berries	Anticholinergic toxicity	Toxic effects with ≥ 10 berries; fatal adult dose ~ 200 berries
	Black nightshade	<i>Solanum nigrum</i>	2% solasonine in stem, root, unripe berries	Anticholinergic toxicity	Less toxic than belladonna
	Henbane	<i>Hyoscyamus niger</i>	0.04%–0.28% hyoscyamine in leaves	Anticholinergic toxicity	More sedating than belladonna
	Jimsonweed	<i>Datura stramonium</i>	0.1%–0.6% hyoscyamine in ripe seeds, leaves, flowers	Anticholinergic toxicity	Seeds may be chewed or leaves smoked as cigarettes
	Mandrake	<i>Mandragora vernalis</i>	0.4% hyoscyamine in root	Anticholinergic toxicity	More sedating than belladonna
Cholingeric	Calabar bean	<i>Physostigma venenosum</i>	Physostigmine in ripe seeds	Cholinergic toxicity	Chewing seeds releases more physostigmine
Cardiotoxicity	(Scotch) broom	<i>Cytisus scoparius</i>	2% tyramine and 0.01%–0.22% sparteine in aerial parts	(–) inotrope, quinidine-like antiarrhythmic	Toxic effects with ≥ 30 g raw herb
	Canadian hemp	<i>Apocynum cannabinum</i>	Cardioactive glycosides in root and aerial parts	Bradycardia, A–V block	Less cardiotoxic than foxglove
	Foxglove	<i>Digitalis purpurea</i>	Cardioactive glycosides in ripe seeds, leaves, flowers	Bradycardia, A–V block, miosis	<i>Digitalis ianata</i> is major source of digoxin in the United States.
	Hedge mustard	<i>Sisymbrium officinale</i>	Cardioactive glycosides in aerial parts	Bradycardia, A–V block	Avoid confusion with other mustard species
	Lily of the valley	<i>Convallaria majalis</i>	Cardioactive glycosides in root, flowers, leaves	Bradycardia, A–V block	Water from the cut flowers also toxic
	Monkshood	<i>Aconitum napellus</i>	Aconitine in root, leaves, flowers	Hypothermia, bradycardia, respiratory arrest	Fatal adult dose ≥ 2 mg; topical use also toxic
	Wallflower	<i>Cheiranthus cheiri</i>	Cardioactive glycosides in ripe seeds and flowers	Bradycardia, A–V block	
Neurotoxicity	Nux vomica	<i>Strychnos nuxvomica</i>	2%–5% strychnine in ripe seeds and bark	Hyperthermia, agitation, seizures	Toxic effects with 30 mg–50 mg; fatal adult dose 1 g–2 g
	Wormwood	<i>Artemisia absinthium</i>	Thujone in the volatile oil from leaves and flowers	Delirium, psychosis, renal failure, xanthopsia	Thujone related to camphor



GI Toxicity	Bloodroot	<i>Sanguinaria Canadensis</i>	4%–7% sanguinarine in root	Tissue irritation and necrosis with topical use	Poorly absorbed
	Castor bean	<i>Ricinus communis</i>	42%–55% castor oil and 0.1%–0.7% ricin in ripe seeds	Cathartic laxative	Chewing seeds releases the ricin
	Jalap	<i>Ipomoea orizabensis</i>	12%–15% polymeric ester glycosides in root	Cathartic laxative	
	Marsh marigold	<i>Caltha palustris</i>	Protoanemonin in fresh flowers	Skin blistering, mucous membrane irritation	Dried flowers have little toxic effect
	Queen's delight	<i>Stillingia sylvatica</i>	Diterpene esters in juice of root	Inflammation and mucous membrane irritation	
Hepatotoxicity	Calamus	<i>Acorus</i> spp: <i>A. calamus</i> , <i>A. americanus</i> , <i>A. angustatinus</i>	β -Isoasarone in oil from root: calamus <10%, americanus 0%, angustatinus > 80%	Potential hepatocarcinogen	Active component related to reserpine
	Chaparral	<i>Larrea tridentata</i>	Nordihydro-guaiaretic acid in leaves	Cholestasis, potential hepatocarcinogen	
	Comfrey	<i>Symphytum officinale</i>	0.03%–0.6% UPAs in root and leaves	Veno-occlusive disease, potential hepatocarcinogen	Safe for use topically to unbroken skin for <10 days
	Germander	<i>Teucrium chamaedrys</i>	Teucrin A in aerial parts	Liver necrosis	Safe for use in small amounts to flavor beverages
	Life root	<i>Senecio nemorensis</i>	0.01%–0.1% UPAs in aerial parts	Veno-occlusive disease, potential hepatocarcinogen	
Multi-organ Toxicity	Arnica	<i>Arnica montana</i>	Sesquiterpene lactones in flowers	Hypertension, arrhythmias, muscle paralysis	May be safe for topical use
	Autumn crocus	<i>Colchicum autumnale</i>	$\geq 0.4\%$ colchicine in flowers and ripe seeds	Renal and bone marrow failure, hepatotoxicity	Fatal adult dose 5 g of seeds
	Cotton	<i>Gossypium herbaceum</i>	Gossypol in seeds and root	Heart failure, inhibition of spermatogenesis	Small amounts of cotton seed oil in foods are safe
	Daffodil	<i>Narcissus pseudonarcissus</i>	Lycorine in bulb, leaves, flowers	Respiratory arrest, topical irritation	
	Lobelia	<i>Lobelia inflata</i>	6% lobeline in leaves and seeds	Hypothermia, anxiety, seizures, arrhythmias, respiratory arrest	Nicotine-like, toxic effects with 1 g; fatal adult dose 4 g
	Mayapple	<i>Podophyllum peltatum</i>	20% podophyllotoxin in root and resin	Ataxia, seizures, psychosis, coma, mucous membrane irritation, bloody diarrhea	Ripe fruits are not toxic; etoposide is a semisynthetic derivative

(Continued)

**Table 3** Potentially unsafe herbs (Continued)

Harmful effects	Common name	Latin binomial	Toxic constituents	Serious adverse effects	Comments
	Mistletoe	<i>Viscum album</i>	2% viscotoxins in leaves, berries, branches	Delirium, seizures, hypotension, cardiac arrest	Lives as a parasite on the branches of other trees
	Poison hemlock	<i>Conium maculatum</i>	Coniine in leaves, berries, flowers	Mydriasis, dizziness, respiratory arrest	Fatal adult dose 30 g berries or 100 g leaves
	Wahoo	<i>Euonymus atropurpureus</i>	Unknown toxic compound in root bark, seeds, berries	Tonic-clonic spasms, coma, hypotension, cardiac arrest	
	Wild cherry	<i>Prunus virginiana</i>	Amygdalin in leaves, seeds, stem	Headache, muscle spasms, coma, respiratory arrest	Converted to cyanide in the GI tract
	Wormseed	<i>Chenopodium ambrosioides</i>	80% ascaridole in distilled oil of seeds, berries, aerial parts	Vertigo, seizures, respiratory arrest, hypotension, cardiac arrest	Oil may explode when heated
	Yellow jessamine	<i>Gelsemium sempervirens</i>	Gelsamine alkaloids in root	Dizziness, mydriasis, seizures, bradycardia, muscle paralysis, respiratory arrest	Fatal adult dose 4 ml extract or 2 g–3 g root
	Yohimbe	<i>Pausinystalia yohimbe</i>	6% yohimbine in bark	Mydriasis, anxiety, tremor, α -2 adrenergic block with hypertension, cardiac arrest	

UPAs = unsaturated pyrrolizidine alkaloids.

(From Refs. 18, 53–56.)



before. Patients should be counseled to inform the pharmacist of all medications being taken, both conventional and otherwise. They should be told that herbals may be helpful or harmful and that limited efficacy data are usually available that is done in a controlled, scientific manner. If the pharmacy practitioner feels that the supplement is safe to be taken by the patient, then several counseling points should be stressed to maximize the potential of taking a quality supplement product. Multi-ingredient products should generally be avoided unless the patient is under the care of an herbalist. Labels should list both the common and Latin names, as well as the name and address of the manufacturer. It is likely better to purchase a supplement made by a reputable, well-known manufacturer who has a history of conforming to GMPs. The patient should be told to be aware of the different dosage forms available and that not all herbals work best when taken as a tablet. "Whole herb" products are generally ground plant parts and are not standardized at all. They should be avoided. Of course, patients should be cautioned to not believe every efficacy claim made for supplements. Finally, patients should be told to promptly report any adverse effects they think may be due to the supplement.

ALTERNATIVE MEDICINE RESOURCES

Of course, published review articles, clinical trials, and case reports can be found on Medline. To aid the pharmacy practitioner in evaluating the potential safety and efficacy of various herbal products, the following is a list of additional reputable resources.^[57,58]

Textbooks

- *The American Cancer Society's Guide to Complementary and Alternative Cancer Methods* (American Cancer Society, 2000).
- *The Botanical Safety Handbook* (McGuffin, CRC Press, 1997).
- *Herb Contraindications and Drug Interactions* (second edition, Brinker, Eclectic Medical Publications, 1998).
- *Herbal Medicinals: A Clinician's Guide* (Miller & Murray, Haworth Press, 1998).
- *Herbal Medicine: Expanded Commission E Monographs* (Blumenthal, Brinckmann, Goldberg, Integrative Medicine Communication, 2000).
- *The Professional's Handbook of Complementary and Alternative Medicine* (Fetrow & Avila, Springhouse, 1999).
- *Rational Phytotherapy: A Physician's Guide to Herbal Medicine* (fourth edition, Schulz, Springer-Verlag, 2001).
- *Tyler's Herbs of Choice: The Therapeutic Use of Phytomedicinals* (second edition, Robbers & Tyler, Haworth Press, 1999).
- *Tyler's Honest Herbal: A Sensible Guide to the Use of Herbs and Related Remedies* (fourth edition, Foster & Tyler, Haworth Press, 1999).

Periodicals

- Alternative Medicine Alert.
- Alternative Medicine Journal.
- HerbalGram (American Botanical Council).
- Journal of Alternative and Complementary Medicine.
- Journal of Natural Products (American Society of Pharmacognosy).
- Natural Medicines Comprehensive Database (The Pharmacists' Letter).
- The Review of Natural Products (Facts and Comparisons).

Web Resources

- Alternative Medicine Foundation Herbal Medicine (<http://www.herbmed.org>).
- American Botanical Council (<http://www.herbalgram.org>).
- American Cancer Society's Alternative and Complementary Therapies.
- Herbal Research Foundation (<http://www.herbs.org>).
- HerbNet (<http://www.herbnet.com>).
- NAPRALERT at the University of Illinois at Chicago (<http://www.pmpu.uic.edu>).
- National Cancer Institute PDQs on Alternative Cancer Therapies (<http://www.cancer.gov>).
- National Center for Complementary and Alternative Medicine (<http://nccam.nih.gov>).
- The Natural Pharmacist (<http://www.tnp.com/home.asp>).

CONCLUSION

Use of alternative medicine is on the rise, with herbal medicine being one of the fastest growing practices. Herbal medicines are not held to the same efficacy and safety standards as conventional medicines, but are rather sold as dietary supplements under the DSHEA of 1994. Limi-



tations of this legislation include: nonstandard botanical nomenclature, little guarantee of GMPs, the burden of safety being on the FDA rather than the manufacturer, and efficacy data lacking. Nonetheless, dangerous adverse effects and herb–drug interactions are being reported with increasing frequency. Current pharmacy practitioners have not received formal training in herbal medicine, yet they need to be familiar with herbal uses, dosing, toxicities, contraindications, and potential drug interactions. This is necessary to help patients who choose to use supplements safely. Echinacea, garlic, ginseng, ginkgo, St. John's wort, ma huang, and valerian are among the most commonly used supplements, but there are many others readily available with potentially harmful effects. Fortunately, in addition to conventional periodicals, a number of reputable herbal texts and journals are available to provide the pharmacy professional with reliable herbal drug information.

REFERENCES

- Cassileth, B.R.; Chapman, C.C. Alternative and Complementary Cancer Therapies. *Cancer* **1996**, *77* (6), 1026–1034.
- <http://nccam.nih.gov/> (accessed February 2002).
- Blumenthal, M.; Israelsen, L.D. The History of Herbs in the United States: Legal and Regulatory Perspectives. In *Herbal Medicinals: A Clinician's Guide*; Miller, L.G., Murray, W.J., Eds.; The Haworth Press: New York, 1998; 325–353.
- Eisenberg, D.M.; Kessler, R.C.; Foster, C.; Norlock, F.E.; Calkins, D.R.; Delbanco, T.L. Unconventional Medicine in the United States. *N. Engl. J. Med.* **1993**, *328* (4), 246–252.
- Eisenberg, D.M.; Davis, R.B.; Ettner, S.L.; Appel, S.; Wilkey, S.; Van Rompey, M.; Kessler, R.C. Trends in Alternative Medicine Use in the United States, 1990–1997: Results of a Follow-Up National Survey. *J. Am. Med. Assoc.* **1998**, *280* (18), 1569–1575.
- Astin, J.A. Why Patients Use Alternative Medicine: Results of a National Study. *J. Am. Med. Assoc.* **1998**, *279* (19), 1548–1553.
- Paltiel, O.; Avitzour, M.; Peretz, T.; Cherny, N.; Kaduri, L.; Pfeffer, R.M.; Wagner, N.; Soskolne, V. Determinants of the Use of Complementary Therapies by Patients with Cancer. *J. Clin. Oncol.* **2001**, *19* (9), 2439–2448.
- Ernst, E.; Bassileth, B.R. The Prevalence of Complementary/Alternative Medicine in Cancer: A Systemic Review. *Cancer* **1998**, *83* (4), 777–782.
- Richardson, M.A.; Sanders, T.; Palmer, J.L.; Greisinger, A.; Singletary, S.E. Complementary/Alternative Medicine Use in a Comprehensive Cancer Center and the Implications for Oncology. *J. Clin. Oncol.* **2000**, *18* (13), 2505–2514.
- Boon, H.; Stewart, M.; Kennard, M.A.; Gray, R.; Sawka, C.; Brown, J.B.; McWilliams, C.; Gavin, A.; Baron, R.A.; Aaron, D.; Haines-Kanka, T. Use of Complementary/Alternative Medicine by Breast Cancer Survivors in Ontario: Prevalence and Perceptions. *J. Clin. Oncol.* **2000**, *18* (13), 2515–2521.
- Fairfield, K.M.; Eisenberg, D.M.; Davis, R.B.; Libman, H.; Phillips, R.S. Patterns of Use, Expenditures, and Perceived Efficacy of Complementary and Alternative Therapies in HIV-Infected Patients. *Arch. Intern. Med.* **1998**, *158*, 2257–2264.
- Lenz, K.L. Developing a Pharmacy Curriculum in Complementary and Alternative Medicine. *J. Pharm. Pract.* **1999**, *12* (3), 178–186.
- Chang, Z.G.; Kennedy, D.T.; Holdford, D.A.; Small, R.E. Pharmacists' Knowledge and Attitudes Toward Herbal Medicine. *Ann. Pharmacother.* **2000**, *34*, 710–715.
- Robbers, J.E.; Tyler, V.E. *Tyler's Herbs of Choice: The Therapeutic Use of Phytomedicinals*, 2nd Ed.; The Haworth Press: New York, 1999.
- Gurley, B.J.; Gardner, S.F.; Hubbard, M.A. Content Versus Label Claims in Ephedra-Containing Dietary Supplements. *Am. J. Health-Syst. Pharm.* **2000**, *57*, 963–969.
- <http://www.fda.gov/medwatch> (accessed February 2002).
- DiPaola, R.S.; Zhang, H.; Lambert, G.H.; Meekr, R.; Licitra, E.; Rafi, M.M.; Zhu, B.T.; Spaulding, H.; Goodin, S.; Toledano, M.B.; Hait, W.N.; Gallo, M.A. Clinical and Biologic Activity of an Estrogenic Herbal Combination (PC-SPES) in Prostate Cancer. *N. Engl. J. Med.* **1998**, *339* (12), 785–791.
- Fetrow, C.W.; Avila, J.R. *Professional's Handbook of Complementary and Alternative Medicines*; Springhouse: Springhouse, PA, 1999.
- Miller, L.G. Herbal Medicinals: Selected Clinical Considerations Focusing on Known or Potential Drug–Herb Interactions. *Arch. Intern. Med.* **1998**, *158*, 2200–2211.
- Vaes, L.P.J.; Chyka, P.A. Interactions of Warfarin with Garlic, Ginger, Ginkgo, or Ginseng: Nature of the Evidence. *Ann. Pharmacother.* **2000**, *34* (12), 1478–1482.
- Heck, A.M.; DeWitt, B.A.; Lukes, A.L. Potential Interactions Between Alternative Therapies and Warfarin. *Am. J. Health-Syst. Pharm.* **2000**, *57*, 1221–1230.
- Fugh-Berman, A. Herb–drug Interactions. *Lancet* **2000**, *355*, 134–138.
- Pepping, J. Black Cohosh: *Cimicifuga racemosa*. *Am. J. Health-Syst. Pharm.* **1999**, *56*, 1400–1402.
- Jacobson, J.S.; Troxel, A.B.; Evans, J.; Klaus, L.; Vahdat, L.; Kinne, D.; Steve Lo, K.M.; Moore, A.; Rosenman, P.J.; Kaufman, E.L.; Neugut, A.I.; Grann, V.R. Randomized Trial of Black Cohosh for the Treatment of Hot Flashes Among Women with a History of Breast Cancer. *J. Clin. Oncol.* **2001**, *19* (10), 2739–2745.
- Pepping, J. DHEA: Dehydroepiandrosterone. *Am. J. Health-Syst. Pharm.* **2000**, *57*, 2048–2056.
- Pepping, J. Echinacea. *Am. J. Health-Syst. Pharm.* **1999**, *56*, 121–122.



27. Giles, J.T.; Palat, C.T.; Chien, S.H.; Chang, Z.G.; Kennedy, D.T. Evaluation of Echinacea for Treatment of the Common Cold. *Pharmacotherapy*. **2000**, *20* (6), 690–697.
28. Gallo, M.; Sarkar, M.; Au, W.; Pietrzak, K.; Comas, B.; Smith, M.; Jaeger, T.V.; Einarson, A.; Koren, G. Pregnancy Outcome Following Gestational Exposure to Echinacea: A Prospective Controlled Study. *Arch. Intern. Med.* **2000**, *160*, 3141–3143.
29. Stevinson, C.; Pittler, M.H.; Ernst, E. Garlic for Treating Hypercholesterolemia: A Meta-analysis of Randomized Clinical Trials. *Ann. Intern. Med.* **2000**, *133* (6), 420–429.
30. Ackermann, R.T.; Mulrow, C.D.; Ramirez, G.; Gardner, C.D.; Morbidoni, L.; Lawrence, V.A. Garlic Shows Promise for Improving Some Cardiovascular Risk Factors. *Arch. Intern. Med.* **2001**, *161*, 813–824.
31. Grant, K.L.; Lutz, R.B. Ginger. *Am. J. Health-Syst. Pharm.* **2000**, *57*, 945–947.
32. Vutyavanich, T.; Kraissarin, T.; Ruangsri, R.A. Ginger for Nausea and Vomiting in Pregnancy: Randomized Double-Masked, Placebo-Controlled Trial. *Obstet. Gynecol.* **2001**, *97* (4), 577–582.
33. Le Bars, P.L.; Katz, M.M.; Berman, N.; Itil, T.M.; Freedman, A.M.; Schatzberg, A.F. A Placebo-Controlled, Double-Blind, Randomized Trial of an Extract of Ginkgo Biloba for Dementia. *J. Am. Med. Assoc.* **1997**, *278* (16), 1327–1332.
34. Pittler, M.H.; Ernst, E. Ginkgo Biloba Extract for the Treatment of Intermittent Claudication: A Meta-analysis of Randomized Trials. *Am. J. Med.* **2000**, *108* (3), 276–281.
35. Fessenden, J.M.; Wittenborn, W.; Clarke, L. Ginkgo Biloba: A Case Report of Herbal Medicine and Bleeding Postoperatively from a Laparoscopic Cholecystectomy. *Am. Surgeon* **2001**, *67* (1), 33–35.
36. Cardinal, B.J.; Engels, H.J. Ginseng Does Not Enhance Psychological Well-being in Healthy Young Adults: Results of a Double-Blind, Placebo-Controlled Randomized Clinical Trial. *J. Am. Diet. Assoc.* **2001**, *101* (6), 655–660.
37. Runkel, D.R.; Cupp, M.J. Glucosamine Sulfate Use in Osteoarthritis. *Am. J. Health-Syst. Pharm.* **1999**, *56*, 267–269.
38. McAlindon, T.E.; LaValley, M.P.; Gulin, J.P.; Felson, D.T. Glucosamine and Chondroitin for Treatment of Osteoarthritis: A Systemic Quality Assessment and Meta-analysis. *J. Am. Med. Assoc.* **2000**, *283* (11), 1469–1475.
39. Pisters, K.M.W.; Newman, R.A.; Coldman, B.; Shin, D.M.; Khuri, F.R.; Hong, W.K.; Glisson, B.S.; Lee, J.S.; Phase I Trial of Oral Green Tea Extract in Adult Patients with Solid Tumors. *J. Clin. Oncol.* **2001**, *19* (6), 1830–1838.
40. Taylor, J.R.; Wilt, V.M. Probable Antagonism of Warfarin by Green Tea. *Ann. Pharmacother.* **1999**, *33*, 426–428.
41. Almeida, J.C.; Grimsley, E.W. Coma from the Health Food Store: Interaction Between Kava and Alprazolam. *Ann. Intern. Med.* **1996**, *125*, 940–941.
42. Haller, C.A.; Benowitz, N.L. Adverse Cardiovascular and Central Nervous System Events Associated with Dietary Supplements Containing Ephedra Alkaloids. *N. Engl. J. Med.* **2000**, *343* (25), 1833–1838.
43. Pepping, J. Melatonin. *Am. J. Health-Syst. Pharm.* **1999**, *56*, 2520–2527.
44. Pepping, J. Milk Thistle: *Silybum marianum*. *Am. J. Health-Syst. Pharm.* **1999**, *56*, 1195–1197.
45. Chavez, M. S-Adenosylmethionine. *Am. J. Health-Syst. Pharm.* **2000**, *57*, 119–123.
46. Wilt, T.J.; Ishani, A.; Stark, G.; MacDonald, R.; Lau, J.; Mulrow, C. Saw Palmetto Extracts for Treatment of Benign Prostatic Hypertrophy: A Systemic Review. *J. Am. Med. Assoc.* **1998**, *280* (18), 1604–1609.
47. Shelton, R.C.; Keller, M.B.; Gelenberg, A.; Dunner, D.L.; Hirschfeld, R.; Thase, M.E.; Russell, J.; Lydiard, R.B.; Crits-Cristoph, P.; Gallop, R.; Todd, L.; Hellerstein, D.; Goodnick, P.; Keitner, G.; Stahl, S.M.; Halbrich, U. Effectiveness of St. John's Wort in Major Depression: A Randomized Clinical Trial. *J. Am. Med. Assoc.* **2001**, *285* (15), 1978–1986.
48. Gaster, B.; Holroyd, J. St. John's Wort for Depression: A Systemic Review. *Arch. Intern. Med.* **2000**, *160*, 152–156.
49. Barone, G.W.; Gurley, B.J.; Ketel, B.L.; Lightfoot, M.L.; Abul-Ezz, S.R. Drug Interactions Between St. John's Wort and Cytosporine. *Ann. Pharmacother.* **2000**, *34*, 1013–1016.
50. Ruschitzka, F.; Meier, P.J.; Turina, M.; Luscher, T.F.; Noll, G. Acute Heart Transplant Rejection Due to Saint John's Wort. *Lancet* **2000**, *355*, 548–549.
51. Piscitelli, S.C.; Burstein, A.H.; Chait, D.; Alfaro, R.M.; Falloon, J. Indinavir Concentrations and St. John's Wort. *Lancet* **2000**, *355*, 547–548.
52. Plushner, S.L. Valerian: *Valeriana officinalis*. *Am. J. Health-Syst. Pharm.* **2000**, *57*, 328–335.
53. Vanderhoff, B.T.; Mosser, K.H. Jimson Weed Toxicity: Management of Anticholinergic Plant Ingestion. *Am. Fam. Physician* **1992**, *46* (2), 526–530.
54. Sheikh, N.M.; Philen, R.M.; Love, L.A. Chaparral-Associated Hepatotoxicity. *Arch. Intern. Med.* **1997**, *157*, 913–919.
55. Weisbord, S.D.; Soule, J.B.; Kimmel, P.L. Poison On-line: Acute Renal Failure Caused by Oil of Wormwood Purchased Through the Internet. *N. Engl. J. Med.* **1997**, *337* (12), 825–827.
56. Tai, Y.T.; But, P.P.H.; Young, K.; Lau, C.P. Cardiotoxicity After Accidental Herb-Induced Aconite Poisoning. *Lancet* **1992**, *340*, 1254–1256.
57. Kiefer, D.; Shah, S.; Gardiner, P.; Wechkin, H. Finding information on Herbal Therapy: A Guide to Useful Sources for Clinicians. *Altern. Ther.* **2001**, *7* (6), 74–78.
58. Eisenberg, D.M. Advising Patients Who Seek Alternative Medical Therapies. *Ann. Intern. Med.* **1997**, *127* (1), 61–69.

ASEPTIC PROCESSING, VALIDATION OF

James P. Agalloco

Agalloco & Associates, Belle Mead, New Jersey, U.S.A.

James E. Akers

Akers Kennedy & Associates, Kansas City, Missouri, U.S.A.

INTRODUCTION

Aseptic processing is a widely used methodology in the health care industry for the preparation of sterile materials. The term aseptic processing as it is applied in the pharmaceutical industry refers to the assembly of sterilized components and product in a specialized clean environment. The clean environment may be a conventional human scale classified clean rooms or an environment engineered to further reduce the likelihood of contamination by reducing (or as much as is possible eliminating) direct human contact with the product and components being assembled "aseptically." The idea of sterile products manufactured aseptically is inherently contradictory, a demonstrably sterile product cannot be produced aseptically using even the most advanced technology available today. Nevertheless, on any given day millions of putatively sterile dosage form units are produced using aseptic techniques that in the literal sense are inadequate to achieve sterility. A sterile product is one that is free from all living organisms, whether in a vegetative or spore state. This is an absolute condition, something cannot be partially or nearly sterile, the presence of a single viable organism represents a failure of the product, and the systems (environment, equipment, and procedures) used to produce it. Asepsis, that state in which all aseptically filled sterile products are manufactured, cannot be established as "sterile." Asepsis is commonly defined as a condition in which living pathogenic organisms are absent.

Putting aside the classical definitions, one must consider the real difficulty in establishing an aseptic environment, let alone a sterile one. The practitioner is left with an insurmountable task, to somehow create an environment free of any organisms, but also one (with the exception of isolators) in which personnel must be present to perform critical functions. The problem is further compounded if it is recalled that personnel are considered the single greatest source of microbial contamination in aseptic processing. Recent experiments

have shown that personnel clothed in new, sterile clean room garments slough viable contamination at a rate of roughly one viable particulate to 10,000 nonviable particles. During slow deliberate movements with the best possible clothing, operators will slough particulate and viable organisms. Therefore, the probability of human borne microbial contamination being released in the conventional clean room is one over the course of any reasonably long operational shift. With this fact in mind, how then is one to accomplish a truly sterile or even aseptic environment? Especially when we must consider that many organisms that are normally nonpathogenic, can under certain circumstances become opportunistically pathogenic. Among those circumstances are a debilitated condition of general health in the patient, or, as is increasingly common, immunological insufficiency due to age or pre-existing condition.

Other than the obvious considerations of proper facility design, sterilization validation, and sanitization procedures (all of which are discussed elsewhere in this encyclopedia), the focus of attention must be on the personnel and the activities which they must perform. These actions are broadly termed, aseptic technique, and like any other human activity they can be accomplished in a variety of ways. In order to better understand aseptic technique, some general guidance and examples of good and bad technique can be used to delineate what should and should not be permitted.

The fundamental concept behind every aseptic processing activity is that nonsterile objects must never touch sterile objects. This is often accomplished by the establishment of a "sterile field" in which the core activities are performed. All of the surfaces of the gownned human operator must be always considered nonsterile. Nonsterile objects including the operators hands must never be placed between the source of the air and a sterile object. The operators' hands and arms must always be kept at a level beneath that of open product containers. Sterile components should under no circumstances be touched directly with gloved hands, a sterilized tool should always

used for this purpose. Since gloved hands and arms will enter the sterile field they must never touch walls, floors, doors, etc. Strenuous lifting and moving of tanks, trolleys, etc. must not be done by operators assigned to work within or near the sterile field, because the more strenuous the activity the higher the level of particle generation, and at least some of the particulate released by the operator will surely be viable microorganisms.

Some of the techniques to avoid include: reaching over exposed sterile objects to make adjustments beyond them; correcting a stopper feed problem with a gloved hand; touching face, eye shield, or any other nonsterile object with gloves; taking an air sample directly over open containers; continuously standing inside flexible partitions that mark the boundary of the sterile field; breaking up clumps of components with gloved hands. Each of these actions exposes the sterile objects to undue risk of contamination from the personnel. Certainly there are more ways to contaminate the “sterile field” than we can imagine. For this reason, the procedures used in and around the “sterile field” must be carefully defined and followed closely by all personnel. These procedures should follow the general principles outlined above and are evaluated in a media fill simulation and performed in an identical fashion during aseptic processing. It is beneficial to define in writing how each procedure is to be performed and train the operators in these exact procedures.

Worst Case

No discussion of aseptic processing validation can be considered complete without some mention of “worst case.” As initially defined by the FDA, worst case included consideration of numbers of personnel, temperature, relative humidity, and other aspects (1). This aspect has been adopted with some degree of modification by industry which has included some, but certainly not all of the FDA guidance. Some of the more common worst case situations which industry employs include: number of personnel, maximum hold time for containers and other items prior to filling, number and type of interventions performed. It is important to note that the determination of worst case conditions in a aseptic processing has been largely intuitive and highly subjective. Quantitative risk analysis is rarely if ever undertaken to establish and categorize actual modes of failure. Thus, worst case conditions for tests are established largely by precedence rather than actual data. The most significant worst case condition that is employed is the use of a microbiological growth media itself. Because the majority of aseptic formulations have either a preservative system and some

products are inherently inhibitory or nonsupportive of microbial growth, the media represents a substantially more favorable environment for the survival of microorganisms. The protocol prepared for the aseptic processing validation effort should delineate where worst case type considerations have been incorporated into the experimental plan. Throughout this document, recommendations will be made to worst case assumptions where choices in the definition of the validation effort must be made.

PREREQUISITES

The validation of aseptic processing should be preceded by the formal validation of the various systems, which contribute to the sterility assurance of the materials to be produced. In essence that mandates that the facility, HVAC system, sterilization procedures for the product contact surfaces, equipment, components and product, sanitization/disinfection procedures for the suite, and personnel gowning. Merely listing the activities, which must be completed, serves to indicate the magnitude of the effort required to prepare for the aseptic processing validation effort. It is sometimes tempting to begin the validation of aseptic processing while these tasks are still underway, especially when one considers that it is universally accepted that the primary source of microbial contamination in an aseptic process are those activities which are performed by gowned personnel (Table 1).

It should be evident that virtually all of the items that top the list are either performed by or corrected by the human operator. The impact of the remaining factors is widely acknowledged to be of secondary consideration.

Table 1 Most likely sources of microbial contamination in aseptic processing^a

1. Personnel borne contaminants
2. Human error
3. Non-routine operations during aseptic process
4. Assembly of sterile equipment prior to use
5. Mechanical failure
6. Inadequate or improper sanitization
7. Transfer of materials within APA
8. Routine operations during aseptic process
9. Airborne contaminants
10. Surface contaminants
11. Failure of sterilizing filter
12. Failure of HEPA filter
13. Inadequate or improper sterilization

^aFrom most to least likely. (From Ref. 2.)

Yet proceeding with the validation of aseptic processing before completing the validation of the supportive processes and systems raises the risk of failure unnecessarily and makes failure resolution well nigh impossible. Obviously, given the significance of human borne contamination as a risk factor, training and qualification of operators is a significant prerequisite to aseptic process validation. However, in an effort to move validation along quickly many firms do not emphasize training or even take short cuts with personnel education. Fortunately the various validations, which are required, are well documented in the literature and the practitioner should have no difficulty finding information on their execution (3).

REGULATORY AND HISTORICAL PERSPECTIVES

Aseptic processing activities are evaluated through process simulations in which a microbiological growth medium is utilized in the process in lieu of the product. The media is incubated after completion of the process to evaluate the procedures utilized. When utilized to evaluate aseptic filling, it is more narrowly defined as “a means of validating the aseptic assembly process that involves the use of a microbiological growth nutrient medium to simulate sterile product filling operations” (4). This technique was first applied in the late 1940s by Rhode, and incorporated into a World Health Organization guideline in the mid 1970s (5). In the late 1980’s, the PDA developed one of the first guides to the execution of media fills for the evaluation of aseptic processing (6). Several years afterwards the U.S. Food and Drug Administration defined its aseptic processing requirements for the first time (7). Additional guidance has been developed by other regulatory and pharmacopeial sources, which have defined the required activities (8–12). The PDA latest guidance documents provide perhaps the most comprehensive source of information on process simulation tests (13, 14). The desire to evaluate aseptic processing activities other than sterile drug filling has made for some adaptation of the classical media fill test, and for this reason the term process simulation test has come into vogue to embrace a wider range of aseptic processing activities. Process simulation tests can be defined as a means for “evaluating an aseptic process employing methods that closely approximate those used for sterile materials using an appropriate placebo material” (15). As mentioned previously, central to the evaluation of an aseptic process is the inclusion of the required interventions which must be performed. Any process

simulation that does not properly evaluate employee aseptic technique and does not fully consider the process human interface is technically invalid.

THE MECHANICS OF MEDIA FILL (PROCESS SIMULATION) EXECUTION

The conduct of aseptic processing validation ordinarily requires the use of a microbiological growth medium in lieu of the product. The PDA/PhRMA joint task force of Validation of Sterile Bulk Processes has outlined some process simulation methodologies which do not require the use of media (see later discussion on this subject), but aside from bulk applications process simulation testing has become essentially synonymous with media fill (16).

Media Sterilization

The execution of a media fill begins with the sterilization of the liquid media. This can be accomplished using either bulk sterilization of the liquid media in a large (glass or stainless steel) container or by filtration with a sterilizing grade filter. The choice of sterilization method is based on considerations of the volume of media required, growth promotion requirements, and filtration rate for the media. Provided that the media is introduced into the process at or before the point in which the production process being simulated becomes sterile, the choice of sterilization method is open. Whether the media be sterilized by filtration or by steam in bulk, there is no benefit to be gained from requiring that the sterilization method used be identical to that for the production materials. It is often suggested that a media fill test can be used to verify formulation bioburden control and process filter validation; however, this is not the case because as already pointed out the growth promoting and physical characteristics of media are far different from those of nearly all pharmaceutical preparations. Sterilization validation is established independent of the process simulation, and is the appropriate activity in which to confirm the appropriateness of the methods employed.

Manufacturing Activities

Once the sterile medium is in place the simulation can begin in which the aseptic processing steps are executed through the conclusion of the process. In many cases the simulation entails the filling of the product into its final container and explains its more commonly used name—media filling. To conclude however that media filling is all

that is required in the validation of aseptic processing ignores the many possible human interventions which can take place during the manufacturing and prefilling activities. Many sterile dosage forms require the execution of a significant number of complex aseptic manipulations after the materials become sterile. Suspensions, creams, ointments, implants, and liposome formulations are among the more common examples of processes where aseptic processing involves activities other than filling of containers. In the more ordinary production of sterile solutions there are manufacturing activities such as sampling and integrity testing of filters which can potentially expose sterile materials to contamination. Thus, a process simulation must include all of these prefilling activities in order to mimic the actions routinely performed in production of sterile materials. Understanding these additional requirements makes it evident why process simulation has come into vogue as a more appropriate description for this activity as opposed to more restrictive media filling. Evaluation of aseptic manufacturing activities can be achieved independently of the filling process using a variety of methods (17). If performed as part of an integrated activity with filling, the evaluation of the filled containers serves as verification of both aseptic manufacturing and filling practices.

The bulk production of sterile drug products such as antibiotics, corticosteroids, insulin, and certain biotechnology products requires that a number of processes be carried out under aseptic conditions. These processes can be evaluated in a manner adapted from those employed for aseptic filling processes. A joint PDA/PhRMA task force has developed the definitive guidance document on this subject (18).

Aseptic Filling

Media is filled into sterile containers using methods identical to those required for production of the sterile product. This activity has been the subject of numerous papers and several surveys of industry practice (19–25). Attention must be paid to the specifics of the aseptic filling process itself. The range of sterile dosage forms, which can be produced, encompasses variations in container type, container size, formulation, lot size, filling speed, and other aspects which should be embraced in the design of the validation program. Each of these must be given careful consideration in the definition of the program requirements and a sound rationale for the selection of the test conditions included in the validation protocol. Presented later are brief discussions of some of the issues to be addressed and some recommendations for execution. While the focus of these points is on activities during the

simulated filling of containers, many of these are relevant to both the manufacturing of sterile bulk materials and the compounding of bulk sterile formulations.

Product-Related Considerations

Type of container

Sterile products are aseptically filled into a variety of containers including glass and plastic bottles, metal and plastic tubes, ampules, and plastic bags. The variety of these containers is matched by the variety of methods required to prepare them for use in the aseptic filling process. Where a filling line is used to fill different types of containers, the differences in sterilization and handling suggest that each type should be assessed independently of the others. Attempting to identify a worst case situation when different sterilization methods, handling issues, and sealing mechanisms are employed is tenuous at best. One container substitution that is always valid is the use of a container that allows the most effective and less intrusive reading of the results. For example, clear containers of identical dimensions should always be substituted for opaque containers provided closure feeding and/or sealing are not affected.

Type of product

The process simulation should encompass the procedures used in the entire filling process. Thus, for lyophilized product, the aseptic loading of the freeze dryer should be a part of the simulation. A suspension product that utilizes a recycle loop around the filling machine would be validated using an identical set-up even though the media being filled does not require such a set-up. Similarly, the validation of a powder filling process should include a placebo material passed through the powder handling system and then filled into the container. When placebo materials are used development tests to ensure that the ratio of placebo to media does not result in a failure of the placebo/media mixture to support microbial growth are necessary. Any special filling or handling activities that are specific to the product being simulated should be a part of the media fill. It is acceptable to add additional steps, such as liquid filling, to a powder fill process to allow for direct incubation of the filled units. Such additional steps may increase the potential for contamination of the filled units in the process simulation relative to the production filling process, but their inclusion is often unavoidable and represents an additional worst case challenge to the process. In some instances, the use of filling of control units, i.e., vials filled with liquid media but not the placebo powder may be beneficial as a means of assessing these

add-on activities, which are not a routine part of the aseptic filling process (26).

In complex processes, the process simulation may be divided into steps. Provided that the steps overlap, they can cover the entire process and allow for isolation of contamination to a specific portion of the overall aseptic process. This practice is employed commonly for freeze drying, where an number of vials can be filled and sealed without transfer to the freeze dryer as a means of distinguishing between contamination derived from the aseptic filling and contamination derived from the lyophilizer loading and freeze drying process (27). Detailed advice on the more common product types can be found in PDA's most recent document of process simulation testing (28).

Filling speed

The extremes of filling speed on the line should be considered in the validation planning. The use of the slowest normal filling speed may increase the potential for contamination ingress via deposition from the surrounding environment. The use of the fastest normal speed may increase the potential for human intervention by increasing the number of routine and nonroutine line interventions. The likely impact of fill speed will depend upon personnel population, proximity of personnel to the sterile field, and number of interventions. A rationale should be developed for the process simulation strategy chose. For example, in the initial validation of a filling line, one fill might be performed at the slowest speed, and two at the highest speed. In routine evaluation of the line, the speeds would be alternated.

Container size

The size of the containers being filled is viewed in a manner similar to fill speed. The largest container (often filled at the slowest speed because of its large fill volume) often has the largest opening, so the potential for microbial entry from the environment should be the greatest for that size. At the other extreme, the smallest container (often filled at the highest speed by virtue of its lower fill volume), represents the greatest handling difficulty. Smaller containers are generally more fragile, and less stable, and thus would be more prone to breakage and jamming in the equipment. Any container that presents additional handling steps or is more prone to breakage or instability should be included in the validation program, as it may represent a greater challenge to the aseptic process than either the largest or smallest container processed on the line. As such it may represent an additional worst case to be addressed.

Closure system

One of the more common differences between products is the closure system. Closure systems are selected for compatibility with the formulation, and in some cases differences in formulation may be create differences in handling difficulty. A stopper that is more prone to clumping or jamming in the tracks of the stopper bowl will necessitate additional interventions not present with other stoppers of similar size and thus would be considered worst case situations. There are a number of specialized closure systems designed to facilitate the delivery of an aseptically filled product. As these systems have sealed interstitial spaces where product contact can occur during administration of the drug product, the simulation procedure for these product configurations should include this space and media should be allowed to contact these surfaces during the incubation.

Fill volume

The volume of media filled into the containers need not be the routine fill volume for the container. It should be of sufficient volume to contact the container-closure seal surfaces and sufficiently large to allow for easy inspection of the filled units postincubation. Despite the lower fill volume, the speed of filling should match that used for the routine filling of the product being simulated. Smaller containers should not be over-filled as sufficient air must be available in the container headspace to support the growth of aerobic organisms and problems have been encountered where the liquid media essentially fills the entire container.

Filling Process Related Considerations

Filling lines

Considering the number of permutations of container, closure system, and other product attributes that must be encompassed in a process simulation program, it should be evident that only in the simplest of situations would a single set of media fills be adequate to provide coverage of all aseptic processes performed. Where multiple lines are present in the facility, each should be considered independently. Process simulation results of one line are not predictive of results on another because the contamination rate is primarily dependent upon human performance. Even identical equipment in two clean rooms designed to the same standard will not give uniform results unless the aseptic technique of the operators is at the same level of performance.

Duration of fill

The duration of the media fill is one of the more contentious issues. In general, media fills should be sufficiently long to include all of the required interventions. Using that requirement alone, a typical media fill might be at least 3–4 h long. Ideally a media fill should utilize more units than are in the product being simulated. This approach is normally followed for all batches up to 5000 units. As the number of units in batch increases current practice is to fill at least 5000 units, and increase the number as the batch size increases. For very large batches or long the campaigns common in blow/fill/seal or isolator systems, media fills interspersed with blank units (either empty or water filled) are used to maintain operating conditions during the simulation. Where this is done, media is filled before and after all planned routine and nonroutine interventions, and conversion to media is performed after any unplanned nonroutine interventions (see following section). Media filled units interspersed with blank units has been a technique used to validate processes that may run for several days. In these cases media is generally filled at the beginning to evaluate set up and then again at the end to evaluate the ability of the process to maintain asepsis for the full length of the longest approved campaign. Filling a large number of units as may be possible on a high speed filling machine is not a substitute for a realistic simulation of the process. A high speed filler could fill 5000 or more units in less than 15 min of filling time, yet that would hardly be considered an acceptable practice.

Interventions

In virtually all aseptic processing activities, operator interventions are required to complete the process. Understanding the types of interventions required and how they are incorporated into the validation program is essential to protocol development.

Aseptic assembly: The first interventions performed are those that prepare the equipment for the aseptic process. This entails the removal of sterilized materials and equipment items from the autoclave and transfer to the location where the aseptic processing activities will be performed. This is ordinarily followed by the assembly/preparation of the equipment for the process. Aseptic assembly in which sterilized parts are removed from protective materials, installed and adjusted in preparation for the aseptic process are perhaps the most potentially invasive of all of the activities which must be performed. The operator must be meticulous in their execution of these tasks to prevent the inadvertent contamination of product contact surfaces. Strict adherence to the principles of aseptic technique described earlier is essential. These

interventions are a necessary part of every aseptic activity, and it is common to identify the first containers filled as they may be more indicative of potential problems with the aseptic assembly. For this reason, the validation program should include process simulations that include containers filled immediately after the set-up of the equipment.

Routine interventions: The execution of the aseptic process ordinarily requires a number of repetitive activities such as: product and component replenishment, weight checking, operator breaks, and environmental monitoring. Each of these is a required part of the process, and cannot be eliminated. They should be included in the process simulation and performed by the operators in a consistent fashion using defined methods and practices.

Nonroutine interventions: During the course of the aseptic filling process there may be instances where a nonroutine or corrective intervention is required. These usually occur in relation to difficulties with components or equipment aspects. Containers can break, jam in the conveyor, or fall over on a turntable. Stoppers can jam in the stopper track, clump in the bowl, or fail to seat properly. Problems with the equipment can include: weight adjustments, minor leakage, sensor failure, or rail adjustment. Each of these will require a corrective operation by the line operator to return the line to proper operation. Unlike routine interventions, these are not a required part of every process, but in order to assess their potential impact on the aseptic process the more prevalent of these should be included in every media fill. To the extent that these nonroutine interventions can be considered repetitive, i.e., weight adjustments, stopper jams, etc. their proper execution should be described in a procedure and adhered to by the operators. Nonroutine interventions may occur randomly or not at all during an aseptic filling process. To ensure that they are a part of the process simulation, they should be performed as if they were a required part of the simulation. Thus, even if the fill weights during the simulation are correct, the line should be stopped and an adjustment made to demonstrate the acceptability of the methods employed. Given the breadth of possible nonroutine interventions, which may be possible during a batch, it may not be possible to simulate all of them. The simulation protocol should address those the firm has identified as more commonly required. Firms should make a concerted effort to minimize the number and extent of these nonroutine interventions. It may be possible to reduce the need to perform them by improving component quality by tighter AQLs on incoming components, tighter controls on preparatory activities, repairs or upgrades to equipment, and similar activities. Such measures can contribute

substantially to the reliability of the process and patient safety.

New interventions: If during the conduct of a batch or a process simulation, the need for an intervention not previously evaluated in a process simulation may occur. The firm should allow for this eventuality and assess the intervention during its execution via supervisory observation. Further evaluation of the new intervention in a follow-up media fill should also be considered.

Documenting interventions: It can be beneficial to define in some detail the permitted interventions for a given aseptic process in an SOP. This practice eliminates any subjectivity regarding what is permitted, and also allows for the establishment of a defined method for performing the intervention. The SOP can be employed in training of the operators and used as guidance in both routine aseptic processing and process simulation. Documentation of routine processing and process simulation should include details on the interventions (routine and nonroutine) performed. The list of interventions required during routine filling can help to define the media fill program by establishing which are more common and should be given precedence in the process simulation. The required interventions can also be used as the initial justification for improvements to the procedures, components, and equipment used in the aseptic process. In the absence of such documentation during the process simulation, it is difficult to defend the acceptability of the intervention during routine processing. The use of video tapes as a means of both documenting interventions during media fills, and as a training tool for operators in the proper execution of an intervention is becoming more prevalent.

Environmental Considerations

Environmental monitoring

An aseptic processing activity is ordinarily supported by monitoring of the environmental air and surfaces in proximity to the process. The purpose of this monitoring is to confirm the acceptability of the environment during the process execution. There are a number of environmental sampling methodologies that are appropriate for this purpose each having particular advantages and disadvantages (29, 30). There are also a number of regulatory and pharmacopoeial references that delineate microbial levels considered acceptable for aseptic processing environments (31–34). Each of these documents has defined the microbial conditions under which aseptic processing operations should be conducted in a slightly different manner. This situation could prove problematic for those endeavoring to comply with all of

the requirements simultaneously except that in the years since the FDA's guideline on aseptic processing was published, aseptic processing capabilities have improved substantially. The microbial limits which may have once proved so daunting are now routinely observed in the majority of aseptic processing applications. Twenty years ago in this industry it was in vogue to speak of the importance of identifying trends in environmental monitoring results. It is recognized today, that trends no longer exist and that the presence of a detectable microorganism in a critical location is a rare event (35). With such stellar performance near routine, some change in the paradigms relative to the performance of environmental monitoring relative to aseptic processing are necessary:

- The sensitivity of environmental sampling systems may be insufficient to detect microorganisms in critical environments with any degree of accuracy.
- Increased sampling in a effort to detect the already low levels of microorganisms is unlikely to be successful and can actually put the product at risk by increasing personnel incursions into the sterile field.
- Equipment and components have improved in quality to the extent, that environmental monitoring may be the most invasive intervention during an aseptic process (an undesirable situation).
- The detection of contamination of any type in an environmental sample from a critical environment, sterility testing or a filled container during a process simulation has become a rare event.

With these views in mind, environmental monitoring must be viewed in a new light. The following insights may prove useful to the practitioner:

- To paraphrase the Hippocratic Oath, the first rule of environmental monitoring should be "to do no harm." Sampling in a effort to detect microorganisms should not increase the potential for contamination of sterile materials.
- No amount of sampling could ever establish the acceptability of filled containers of sterile product.
- The sampling and enumeration of microorganisms is perhaps more prone to inadvertent contamination than the aseptic process itself.
- The identification of a recoverable organism in the environment is a random event and may have no relationship to the integrity of the "sterile field," or the sterility of the goods being produced. In fact, given the presence of human operators (including the individual performing the sampling) how could one not expect to find occasional contamination?

- There are no “smoking guns,” establishing linkage between sterility test failure isolates, media fill contaminants and environmental isolates is extremely difficult.

Despite these somewhat negative perspectives, environments must still be monitored, and the levels of microorganisms controlled at levels that reconfirms the continued acceptability of the environmental conditions. Air sampling, using either active or passive sampling methods should be performed during the execution of the process. Surface sampling is best performed after the completion of the aseptic process to prevent the inadvertent contamination of product contact surfaces during the process. The vast majority of the samples taken should be devoid of contamination; however, the incidental detection of a recoverable organism from even product contact surfaces postprocess should not be cause for undue alarm. The sampling of the environment is also an aseptic process, and subject to its own flaws.

Personnel monitoring

The evaluation of microbial contamination on operating personnel is a necessary part of the overall program. Sampling should be considered from a perspective similar to that described above for environmental air and surfaces. In this context samples should be taken from hands and other gown surfaces only at the conclusion of the aseptic process. It is suggested that sampling of the operators be performed on every exit from the aseptic area. Unlike environmental air and surface samples, it is unrealistic to expect that all of these samples will be free of detectable organisms, nevertheless the individuals should be able to consistently meet the levels established for them. Individuals who demonstrate a repeated pattern of nonconformance with the expected microbial levels should be subjected to corrective measures. The actions taken could include retraining, aseptic processing reevaluation and gowning recertification (see following section). Personnel undergoing these corrective measures would be assigned duties outside the aseptic processing area until they have reestablished acceptable performance.

Personnel Considerations

Preparatory training

Some firms have adopted specialized aseptic processing exercises to evaluate and prepare personnel being introduced to aseptic processing activities for the first time. These tests can take the form of hand filling, media transfers, and other procedures designed to challenge the aseptic technique of the individual in the absence of the

mechanical equipment. Only after successful completion of the hands-on manipulations would an individual be considered for further training as an aseptic operator. Whether this type of evaluation is performed or not, it is generally accepted that new personnel should actively participate in a media fill before they would be allowed to perform those same activities during a production batch. Additional lecture and demonstration type training of personnel is also necessary in aspects of microbiology, aseptic technique, gowning, equipment operation, and of course CGMP.

Gowning certification

Personnel assigned work in aseptic processing areas are ordinarily subjected to initial and periodic certification of their ability to gown in the prescribed manner. Gowning certification includes sampling of a variety of gown surfaces in addition to those normally evaluated during routine monitoring (36). For new personnel, this might be successfully performed three times before they are permitted access to the aseptic corp. Annual or semiannual sampling reconfirms that personnel are still able to gown properly. Gowning certification is generally extended to include other individuals, i.e., supervisors, maintenance workers, housekeeping personnel who must access the aseptic area, but who do not perform any activities related directly to the sterility of the products being manufactured.

Personnel participation

Virtually all aseptic processing operations require the active participation of human operators who are required to perform important tasks during the process in a manner that avoids the contamination of sterile materials, components, and surfaces. Their success in performing these tasks is assessed in the process simulation as they perform the routine and nonroutine interventions, which comprise their participation in the aseptic process. In order to establish that each of the operators is capable of successfully performing their duties their periodic participation in a process simulation is required. Set-up and line operators should be a part of not less than one process simulation per year. Individuals such as line mechanics and environmental samplers should be managed in a similar manner. Participation requires more than mere presence in proximity to the aseptic process, it must include active execution of the interventions to which they are normally assigned. Supervisory staff and others, such as maintenance personnel who do not perform process related aseptic interventions should not be considered in the validation program.

One of the personnel concerns addressed by the FDA in its guideline on aseptic processing concerned the

maximum occupancy in the aseptic processing room (37). A simple means of accomplishing this is to designate a maximum occupancy for each room. Process simulations are then conducted at that level of staffing, and procedures are followed during routine operation in which a person must exit the room before another can enter whenever the maximum number of personnel are present in the room.

Where a firm operates on multiple shifts, the second and third shift should be included in the program as well to demonstrate the acceptability of their performance as well. Managing a large operation with large numbers of personnel, multiple filling lines and insuring that only personnel who have participated in a related media fill are allowed to perform aseptic processing can be a complex task.

Media Considerations

Media selection

The choice of a growth medium for use in the process simulation requires consideration of the organisms expected to be found in the environment, with emphasis on human derived contaminants from the operating personnel. With this in mind, the conventional choice is a general purpose medium capable of growing a wide range of common aerobic microorganisms (38). The usual choice is Soybean–Casein Digest Medium, also known as Tryptic Soy Broth, the same medium used in sterility testing. In some instances, other media might be appropriate. If filling is performed in an isolator under a total nitrogen environment, then an anaerobic medium such as Fluid Thioglycollate Medium might be more appropriate for the purposes of the simulation.

Use of anaerobic media

The sterility testing of parenteral products includes testing with both aerobic and anaerobic media. When media fills became common in industry during the mid-1970s the use of anaerobic media was considered a standard part of every aseptic processing validation program (39). Operational experience over recent years has indicated that the difficulty in establishing a truly anaerobic environment on a ordinary manned filling line are such that their execution is no longer common (40). Recognition that the predominant source of microorganisms in a clean room are personnel, the conduct of anaerobic media fills for other than special circumstances is unwarranted. Furthermore, anaerobic media fills are not required by any current regulatory body, as they also recognize that the focus of the effort should be on human borne contamination which can survive in air.

Media incubation conditions

An area of particular divergence in industry practice in the execution of media fills is that of incubation conditions (41). Cogent arguments can be made for incubation at a variety of temperatures. Some firms use only a single temperature in the range of 20–35°C. Other firms have chosen to incubate for 7 days at 20–25°C, and then move the filled containers to 30–35°C. An almost equal number have chosen to incubate for 7 days at 30–35°C, and then move the filled containers to 20–25°C. The lack of consensus suggests that the selection of incubation conditions is likely a minor concern. The suitability of the conditions employed, whatever they might be is established by the conduct of growth promotion studies (see following section). The one constant in this area is the duration of the incubation period which is almost universally set at 14 days (42).

Media growth promotion

The central issue in any discussion of appropriate media to use, the need to perform anaerobic media fills, selection of incubation conditions and the one that supports all of the decisions made in these areas, is the growth promotion studies. These establish that the medium used in the process simulation can successfully support the growth of microorganisms. Conventional practice is to test the media against a panel of organisms such as *Bacillus subtilis*, *Candida albicans*, *Pseudomonas aeruginosa*, *Staphylococcus epidermidis*, and *Aspergillus niger*. The use of one or perhaps two of the most common environmental isolates in the growth promotion studies is highly recommended as these may be the most likely to be encountered in a contaminated unit. Failures of the challenge organisms to grow voids the media fill and requires a repeat of the media fill. In media challenges where a powder is added to the media, studies demonstrating acceptable growth promotion are required (43). Process simulations using media performed in antibiotic product facilities will sometimes require the addition of inactivating enzymes to the media in order to obtain acceptable growth promotion results as even trace amounts of antibiotics can inhibit the growth of some organisms.

ACCEPTANCE CRITERIA

Over the last 20–30 years, nothing has proven more controversial in the validation of aseptic processing than the selection of a acceptance criterion to use. The first regulatory requirements for media fills was published by

WHO in the 1960s and had a criterion of 0.3%. Thus, a media fill of more than 1000 filled units would be considered acceptable if no more than three of those units were found contaminated after incubation (44). This limit lasted until the early 1980s when PDA prepared its first guidance on the validation of aseptic processing (45). The PDA document drew upon what was at that time the first survey on aseptic processing practice which had been conducted by PMA, and suggested a criterion of 0.1% of the units filled (46). This suggestion was at the time considered quite radical at the time as it seemingly raised the bar substantially. Interpretations of limit definition were made by Ronald Tetzlaff, at that time an employee of the Food and Drug Administration (47, 48). These drew upon suggestions made in PDA's second document on aseptic processing validation where a Poisson distribution was used to project the contamination over a large batch (49). FDA's first official statement on acceptance criteria for process simulations was provided in their 1987 aseptic processing guideline (50). The FDA guideline stated, "Test results should show, with a high degree of confidence, that the probability of a product becoming contaminated during aseptic processing is very low. In general, test results showing a probability of contamination of not more than one in 1000 are acceptable." With the publication of this document, PDA's 1980 limit had been echoed by a regulatory agency, and this limit remains FDA's current official position. PDA continued its work on aseptic processing through the conduct of industry surveys in 1986 and again in 1992 (51, 52). These efforts established that industry performance was continually improving and that 0.1% was almost universally accepted by industry on a global basis. The Parenteral Society (TPS) developed its first guidance on the validation of aseptic processing, and this placed greater emphasis on statistical treatment of the acceptance criteria than prior documents (53). This document profoundly influenced a ISO document that was then in preparation and a second controversy was born (54). The emphasis placed on statistics in both these documents made many uncomfortable. While mathematically correct the extension of the acceptance criterion tables to include larger numbers of filled units created an awareness of the statistical approach that had not existed previously. Inherent in the statistics, as the number of units filled in a media fill increased, the allowed number of contaminated units increased as well. While one contaminated unit in 1000 containers seemed acceptable, it is statistically equivalent to 10 contaminated units in 16,970 units. This seemed to many as an unacceptably large number of positives units in any size batch. Official EU guidance was provided for the first time in 1996 with the publication of their annex 1 on sterile

medicinal products (55). The annex suggested that, "The contamination rate should be less than 0.1% with 95% confidence level." This was provided without elaboration, thus while the limit was clearly statistical the implications of the TPS and ISO efforts were not addressed. The clearest response to the TPS and ISO statistical data treatment was developed by the PDA (56). The PDA guidance states the following: "Despite the number of units filled during a process simulation test or the number of positives allowed, the ultimate goal for the number of positives in any process simulation test should be zero. A sterile product is, after all, one that contains no viable organisms." The PDA that originally introduced the Poisson approach for evaluation of media fills, has revised its perspective, and now believes that statistical treatment of the data is invalid. The Poisson method is appropriate for the evaluation of a low incidence of a random event, which is no longer believed to be consistent with how sterile materials become contaminated during aseptic processing. When first introduced it was assumed that contamination in an aseptic process could be derived from a variety of sources, and that any statistical approach that addressed that possibility as a random event was considered appropriate. With the execution of many more media fills, substantial improvements in facilities and equipment, and an increased awareness of the human contribution to microbial contamination views have changed substantially. It is now a widely held belief that contamination in aseptic processing is the result of human derived contamination resulting from improper technique in the execution of a required intervention. Thus, there is nothing random about the contamination, its source is known and its elimination certainly more possible than ever. This PDA view represents the latest thinking relative to acceptance criteria. The tenants of PDA's latest effort are provided in Table 2.

Additional publications which have provided acceptance criteria for aseptic processing validation have been prepared by both CEN and PIC (58, 59). These have attempted to reconcile the differences between the ISO document and the PDA guidance. The essence of these documents is the following. "Ideally the contamination rate should be zero. However, currently the accepted contamination rate should be less than 0.1% with a 95% confidence level." This appears to be an effort to have it both ways, and it is unclear whether the two perspectives can really be reconciled so easily.

The most recent industry survey on aseptic processing was published by PDA in 1997 (60). It attempted to address the statistical nature of the limits as actually practiced, however the response to questions in that area are inconclusive. It did include evidence that several firms

Table 2 PDA views on aseptic processing validation acceptance criterion

The test methodology must simulate the process as closely as possible.
Rationale for the chosen methodology and limits must be justifiable and documented.
Test methodology should be sensitive enough to confirm a low process simulation test contamination rate, and the selected limit must be routinely achievable.
Any positive unit indicates a potential problem, regardless of run size. All positives should be identified and should result in a thorough, documented investigation.
Process simulation test contamination rates approaching zero should be achievable using automated production lines in well-designed aseptic processing facilities, blow-fill- seal and form-fill-seal and in isolator-based systems.
Processes conducted in older facilities or employing considerable product handling or manual operation may not be capable of achieving near-zero contamination rates. Nevertheless, such processes must be capable of a process simulation test contamination rate not exceeding one in 1000 when 3000 units are filled.
For batch sizes smaller than 3000 units, process simulation tests should at least equal the batch size. No positives should be allowed due to the low sensitivity of small runs.
When more than 3000 units are filled, caution should be used when deciding to increase the allowable number of positives based on arithmetic extrapolation.

(From Ref. 57.)

had adopted acceptance criteria tighter than 0.1%, suggesting that another round of acceptance criterion definition might be in the offing. The most recent commentary on acceptance criteria is that provided by USP in draft chapter, 1116 (61). The USP expanded upon PDA’s 1996 position, and established ever tighter requirements, “The goal is zero contamination. In an individual run not more than one positive unit in 5000 filled units. In a series of three media fills, two of the three media fills should have no contamination present.”

Perhaps the simplest means of establishing an acceptance criteria (and perhaps to define one’s entire program) is to follow the pack, and develop a firm’s entire aseptic processing validation program based upon the latest survey information (62). What is certainly clear is that aseptic processing performance has improved substantially over the last 20 odd years and that firms should monitor their practices against their peers and regulatory expectation on a continuing basis. What was acceptable in the past is not acceptable today, and there can be little doubt that further tightening of the criteria can be expected in the future.

Technological Advances

The production of sterile products has benefited from at least two novel production methods: blow-fill-seal and isolation technology. Each of these can offer a substantial reduction in the amount of operator interaction with sterile materials. In blow-fill-seal (and the closely related form-fill-seal), the product container is created only seconds before it is filled and sealed. This has distinct advantages over more conventional filling where containers are

exposed to the manned aseptic filling environment. A number of process simulation studies have established that the blow-fill-seal method is capable of protecting the contents of the sealed container to a greater extent than a ordinary clean room (63–65). Isolation technology enjoys similar results using an entirely different approach in which personnel are removed from the operating environment (66, 67). Isolation technology, which is a direct evolution of glove boxes, places operating personnel outside a sealed (physically or via an air pressure differential) enclosure in which the aseptic process is conducted. The enclosure can be treated with a sterilizing gas which can render the interior surfaces free of microorganisms. This treatment when combined with the elimination of direct personnel presence in the aseptic environment, makes the isolator perhaps the ideal tool for sterile drug production (68).

STERILITY ASSURANCE FOR ASEPTIC PROCESSING

To this point in this effort, the validation of aseptic processing has been described as closely related to media filling or process simulation. In reality the relationship between simulation and routine production is not a direct one. A completely successful media fill program does not establish the sterility of anything other than itself. The next lot, or for that matter the previous one, may be sterile or not. The absence of contaminated units in a media fill merely demonstrate that the facility, personnel, and procedures are capable of preventing contamination in that media fill. Demonstrating that capability for routine

filling is quite a different thing, and at the present time cannot be accomplished with other than a destructive sterility test of every filled unit. Ultimately the practitioner can only infer that because media fills are successful, that similar success is also possible during routine production. There is a common misunderstanding that the 0.1% or maximum of one in 1000 units is a sterility assurance level. That is most definitely not the case, it is nothing more than a maximum allowable contamination rate during the process simulation. There is no accepted means for establishing the sterility assurance level of aseptically filled products, it might be termed the "holy grail" of sterile production. We perform media fills to demonstrate a capability, and with significant limitations we infer from that effort that we can produce sterile drug products using aseptic processing. At the present time, this is perhaps as close as we can get to the "validation of aseptic processing." Despite this most basic of constraints, media fills represent the only means of even approximating what occurs when aseptic processing is performed. Demonstrating the capability of producing sterile products will have to be sufficient for industry needs.

CONCLUSION

Demonstrating success with aseptic processing requires process simulation studies closely matching the routine production activities. As the range of sterile products manufactured by aseptic processing is quite extensive, this mandates that the practitioner be prepared to adapt the general guidance provided in this effort to their particular situation. The more closely the simulation matches the production activities, the clearer indication that success with the simulation means success in routine operation. A well founded process simulation program affords the firm confidence in their routine operation that cannot be obtained by any other means.

It is important to consider however, that at the current level of technology, particularly manned aseptic processing, uncertainty is an inherent feature. The fact that process simulation tests are merely a snapshot in time has been recognized since media fills became a standard feature of aseptic process validation in the late 1970s. Guaranteeing safety to the end user is not as simple as successful media fill tests, "good" environmental monitoring results, and successfully completed sterility tests. Even when these data appear satisfactory, uncertainty exists. Although human nature detests uncertainty, particularly in fields where numerical values are widely stated, scientific rationality requires us to recognize that

we cannot test or monitor uncertainty away. Only by thorough training, supervision and reduction of human borne contamination hazards can the likelihood of contamination be controlled. Fortunately, it appears certain from the absence of data to the contrary that aseptically produced health care products are very safe when produced in accordance with current industry standards. Certainly as the industry process capabilities continue to evolve our products will become safer still, and perhaps someday the uncertainty associated with aseptic processing will be so low that we can consider these products truly sterile.

REFERENCES

1. *Guideline on Sterile Drug Products Produced by Aseptic Processing*; FDA, 1987.
2. Current Practices in the Validation of Aseptic Processing—1996, PDA Technical Report #24. PDA J. Pharm. Sci. Technol. **1997**, 51 (2).
3. Carleton, F., Agalloco, J., Eds. *Validation of Pharmaceutical Processes: Sterile Products*; Marcel Dekker, Inc.: New York, 1998.
4. Process Simulation Testing for Aseptically Filled Products, PDA Technical Report #22. PDA J. Pharm. Sci. Technol. **1996**, 50 (6).
5. *General Requirements for the Sterility of Biological Substances*; Part A, Section 2, Annex 4, WHO, 1973.
6. *Validation of Aseptic Filling for Solution Drug Products Technical Monograph*; No. 2, 1980 PDA.
7. *Guide to Sterile Drug Products Produced by Aseptic Processing*; FDA, 1987.
8. *Sterilization of Health Care Products—Aseptic Processing*; Part 1 General Requirements ISO-13408-1. International Standards Organization, 1996.
9. *European Union Guide to Good Manufacturing Practice*; Annex 1 on the Manufacture of Sterile Medicinal Products: European Commission, 1996.
10. *The Use of Process Simulation Tests in the Evaluation of Processes for the Manufacture of Sterile Products*; Technical Monograph No. 4. The Parenteral Society, 1993.
11. *Recommendations on Validation of Aseptic Processes*; CEN /TC 204WG 8 N 38. 1998.
12. *Recommendations on Validation of Aseptic Processes*; PIC/S, PE002-1; 1999.
13. Process Simulation Testing for Aseptically Filled Products PDA Technical Report #22. PDA J. Pharm. Sci. Technol. **1996**, 50 (6).
14. Process Simulation Testing for Sterile Bulk Pharmaceutical Chemicals PDA Technical Report #28. PDA J. Pharm. Sci. Technol. **1998**, 52 (4).
15. Process Simulation Testing for Sterile Bulk Pharmaceutical Chemicals PDA Technical Report #28. PDA J. Pharm. Sci. Technol. **1998**, 52 (4).
16. Process Simulation Testing for Sterile Bulk Pharmaceutical Chemicals PDA Technical Report #28. PDA J. Pharm. Sci. Technol. **1998**, 52 (4).

17. Process Simulation Testing for Sterile Bulk Pharmaceutical Chemicals PDA Technical Report #28. PDA J. Pharm. Sci. Technol. **1998**, 52 (4).
18. Process Simulation Testing for Sterile Bulk Pharmaceutical Chemicals PDA Technical Report #28. PDA J. Pharm. Sci. Technol. **1998**, 52 (4).
19. Prout, G. Validation and Routine Operation of a Sterile Dry Powder Facility. J. Parenteral Sci. Technol. **1982**, 36 (5), 199–204.
20. Tetzlaff, R. Aseptic Process Validation. Particulate & Microb. Control **1983**, 2 (5), 24–38.
21. Tetzlaff, R. Regulatory Aspects of Aseptic Processing. Pharm. Technol. **1984**, 8 (11), 36, 40–44.
22. Korczynski, M. Validation of Aseptic Process by Media Fills—Survey Report and Discussion Proceedings of the Second PMA Seminar Program on Validation of Sterile Manufacturing Processes, Aseptic Processing, PMA, 1979; 186–213
23. Agalloco, J.; Gordon, B. Current Practices in the Use of Media Fills in the Validation of Aseptic Processing. J. Parenteral Sci. Technol. **1987**, 41 (4).
24. Agalloco, J.; Akers, J. Current Practices in the Validation of Aseptic Processing—1992. J. Parenteral Sci. Technol. **1993**, 47 (2).
25. Agalloco, J.; Akers, J. Current Practices in the Validation of Aseptic Processing—1996, PDA Technical Report #24. PDA J. Pharm. Sci. Technol. **1997**, 51 (2).
26. Process Simulation Testing for Aseptically Filled Products, PDA Technical Report #22. PDA J. Pharm. Sci. Technol. **1996**, 50 (6).
27. Process Simulation Testing for Aseptically Filled Products, PDA Technical Report #22. PDA J. Pharm. Sci. Technol. **1996**, 50 (6).
28. Process Simulation Testing for Aseptically Filled Products, PDA Technical Report #22. PDA Journal of Pharmaceutical Science and Technology **1996**, 50 (6).
29. *Fundamentals of a Microbiological Environmental Monitoring Program*; Technical Report #13, PDA, 1991.
30. Akers, J.; Agalloco, J. Aseptic Processing—A Current Perspective. In *Sterilization Technology*; Morrissey, R., Phillips, G.B., Eds.; Van Nostrand Reinhold: New York, 1993.
31. *United States Pharmacopoeial Convention*. In-Process Revision (1116) Microbiological Evaluation of Clean Rooms and Other Controlled Environments, Pharmacopoeial Forum, 1999.
32. *Guideline on Sterile Drug Products Produced by Aseptic Processing*, Food and Drug Administration, 1987.
33. *European Union Guide to Good Manufacturing Practice*; Annex 1 on the Manufacture of Sterile Medicinal Products, European Commission, 1995.
34. *Aseptic Processing of Health Care Products*; ISO/CD 13408.3; International Standards Organization, 1995.
35. *Proceedings on the August 21, 2000 Environmental Monitoring & Aseptic Processing Symposium*; PDA, 2000.
36. *Proceedings on the August 21, 2000 Environmental Monitoring & Aseptic Processing Symposium*; PDA, 2000.
37. *Guideline on Sterile Drug Products Produced by Aseptic Processing*; Food and Drug Administration, 1987.
38. Process Simulation Testing for Aseptically Filled Products, PDA Technical Report #22. PDA J. Pharm. Sci. Technol. **1996**, 50 (6).
39. Korczynski, M. *Validation of Aseptic Process by Media Fills—Survey Report and Discussion*; Proceedings of the Second PMA Seminar Program on Validation of Sterile Manufacturing Processes: Aseptic Processing, PMA, 1979; 186–213
40. Agalloco, J.; Akers, J. Current Practices in the Validation of Aseptic Processing—1996, PDA Technical Report #24. PDA J. of Pharm. Science and Technol. **1997**, 51 (2).
41. Agalloco, J.; Akers, J. Current Practices in the Validation of Aseptic Processing—1996, PDA Technical Report #24. PDA J. Pharm. Sci. Technol. **1997**, 51 (2).
42. Agalloco, J.; Akers, J. Current Practices in the Validation of Aseptic Processing—1996, PDA Technical Report #24. PDA J. Pharm. Sci. Technol. **1997**, 51 (2).
43. Process Simulation Testing for Aseptically Filled Products, PDA Technical Report #22. PDA J. Pharm. Sci. Technol. **1996**, 50 (6).
44. *General Requirements for the Sterility of Biological Substances*; Part A, Section 2, Annex 4, WHO, Geneva, 1973.
45. *Validation of Aseptic Filling for Solution Drug Products*; Technical Monograph No. 2, PDA, Philadelphia, 1980.
46. Korczynski, M. *Validation of Aseptic Process by Media Fills—Survey Report and Discussion*; Proceedings of the Second PMA Seminar Program on Validation of Sterile Manufacturing Processes, Aseptic Processing, PMA, 1979; Washington, DC, 186–213.
47. Tetzlaff, R. Aseptic Process Validation. Particulate & Microbial Control **1983**, 2 (5), 24–38.
48. Tetzlaff, R. Regulatory Aspects of Aseptic Processing. Pharmaceutical Technology **1984**, 8 (11), 36, 40–44.
49. *Validation of Aseptic Drug Powder Filling Processes*; Technical Report No. 6, PDA, Philadelphia, 1984.
50. *Guideline on Sterile Drug Products Produced by Aseptic Processing*; FDA, 1987.
51. Agalloco, J.; Gordon, B. Current Practices in the Use of Media Fills in the Validation of Aseptic Processing. J. Parenteral Sci. Technol. **1987**, 41 (4), 128–141.
52. Agalloco, J.; Akers, J. Current Practices in the Validation of Aseptic Processing—1992. J. Parenteral Sci. Technol. **1993**, 47 (2).
53. *The Use of Process Simulation Tests in the Evaluation of Processes for the Manufacture of Sterile Products*; Technical Monograph No. 4, The Parenteral Society, Swindon, Wiltshire, UK, 1993.
54. *Aseptic Processing of Health Care Products—Part I: General Requirements*; ISO/DIS 13408-1, ISO, Geneva, 1996.
55. *EU Guide to Good Manufacturing Practice, Annex 1 on the Manufacture of Sterile Medicinal Products*; European Commission, 1996.
56. Process Simulation Testing for Aseptically Filled Products, PDA Technical Report #22. PDA J. Pharm. Sci. Technol. **1996**, 50 (6).
57. Process Simulation Testing for Aseptically Filled Products, PDA Technical Report #22. PDA J. Pharm. Sci. Technol. **1996**, 50 (6).
58. *Recommendations on Validation of Aseptic Processes*; CEN /TC 204WG 8 N 38; 1998.
59. *Recommendations on Validation of Aseptic Processes*; PIC/S PE002-1, 1999.

60. Agalloco, J.; Akers, J. Current Practices in the Validation of Aseptic Processing—1996, PDA Technical Report #24. PDA J. Pharm. Sci. Technol. **1997**, 51 (2).
61. United States Pharmacopeial Convention. In *Microbiological Evaluation of Clean Rooms and Other Controlled Environments*; 1116, Pharmacopeial Forum, 1999.
62. Agalloco, J.; Akers, J. Current Practices in the Validation of Aseptic Processing—1996, PDA Technical Report #24. PDA J. Pharm. Sci. Technol. **1997**, 51 (2).
63. Sharp, J. Manufacture of Sterile Pharmaceutical Products Using ‘Blow-Fill-Seal’ Technology. The Pharm. J. **1987**, 239 (106), 22.
64. Sharp, J. Validation of a New Form-Fill-Seal Installation. Manufacturing Chemist **1988**, 22.
65. Kvarnström A.C.; Ernerot L.; Mattsson K. In Form-Fill-Seal: Experience with the Aseptic Filling and Terminal Sterilization of Small Volume Parenterals; Proceedings of PDA International Congress, 1992, 180.
66. Edwards L.; Porter M. In *Microbiological and Physical Limits Testing of a Locally Controlled Environment (LCE) Prototype Filling System*; Proceedings of PDA/ISPE Conference on Advanced Barrier Technology, 1995.
67. Sweeney M.; Davenport S.; Edwards L. *Validation Issues for a Product Barrier/Isolator Sterile Liquid Filling System in a Controlled Environment*; Proceedings of PDA/ISPE Conference on Advanced Barrier Technology, 1995.
68. Agalloco, J. Barriers, Isolators and Microbial Control. PDA J. Pharm. Sci. Technol. **1999**, 53 (1).

ANIMALS IN DRUG DEVELOPMENT

Farrel L. Fort

TAP Pharmaceutical Products, Inc., Lake Forest, Illinois, U.S.A.

ANIMALS IN DRUG DEVELOPMENT

History of Animal Use in Drug Research

The use of animals in scientific investigation has been traced back to several centuries BC. For instance, the writings of Aristotle (384–322 BC) and Erasistratus (304–258 BC) indicate that they had studied the anatomy of various animals (1). Early investigations such as these were the beginning of the basic sciences that today form the foundations for new drug development.

Until the end of the last century, experiments using living animals were carried out on domestic or easily captured wild species. The choice was usually limited and based on availability. By the end of the nineteenth century, the concept of the laboratory animal had begun to emerge as an animal deliberately chosen for its inherent suitability for the purpose; it was especially bred in captivity or obtained from its environment not merely on grounds of convenience but rather for its usefulness for the particular investigation at hand (2).

Need and Rationale for Animal Use

Tests in intact animals are necessary to understand how a drug will work in the context of the myriad metabolic and homeostatic mechanisms that are active in vivo. Screening tests are commonly conducted with in vitro systems and isolated tissues or organs to identify and, in some cases, to act as a bioassay to help purify pharmacologically active agents. However, the variable processes of absorption, distribution within the organism, metabolism to either inactive or more active products, and excretion will modulate the expression of pharmacologic activity in vivo. The only way this modulation can be estimated is by studying the new drug in intact animals.

Aside from studies of pharmacologic activity, side effects of new drugs must be identified and an initial assessment made of their risk-to-benefit ratio. Again, mechanisms of action and effects on specific organs can be studied by using in vitro techniques. However, to identify unexpected adverse effects and to estimate the dosages that may be pharmacologically active without producing unwanted effects, in vivo studies must be conducted.

Extrapolation from Animal to Human

For most substances, the mechanism of action will be the same in humans and other mammals. Therefore, quantitative rather than qualitative differences in response are most common. Humans may be more sensitive to some drugs than certain laboratory animals are, but in many cases some animal species are more sensitive than humans are. For example, the mouse is most sensitive to atropine, the cat is less sensitive, and the dog and the rabbit tolerate atropine at doses 100 times higher than does the human.

Species differences in sensitivity can often be explained by differences in metabolism, including quantitative and qualitative differences in the ability to detoxify drugs and also differences in the rates of absorption, transport, distribution, and elimination of chemicals. After oral administration, absorption in laboratory animals is generally considered to be similar to that in humans, although there are quantitative differences for some compounds. For example, species differences in the absorption and action of some compounds are related to differences in the bacterial flora of the gastrointestinal tract. The distribution and storage of drugs are reasonably consistent among mammalian species, including humans, although plasma binding tends to be more extensive in humans than in small mammals. Urinary excretion in different animal species depends to some extent on their different diets, because diet influences urinary pH and thus the extent of ionization of compounds. Biliary excretion is quite variable from species to species and is apparently more extensive in mice and rabbits than in rats and humans. Species differences in response to drugs appear to be related mainly to rates of biotransformation, which are generally more rapid in small laboratory animals than in humans (3).

ANIMAL MODELS IN DRUG DISCOVERY

Twenty years ago, the initial screening of new compounds for pharmacologic activity was conducted using whole animals or tissues and organs isolated from animals. Today most of the initial screening for new drugs is done in vitro using the techniques of biochemistry and molecular biology. Only after a new drug candidate has been

identified and studied in vitro are studies in animals initiated. The purpose of the animal studies is to verify the pharmacologic activity of the new drug, identify any unexpected pharmacologic activity, and develop an initial data base on the action of the drug in vivo.

An ideal approach would be to have an in vivo screening program designed to allow for the detection of unique profiles of activity or combinations of activities. Therefore, in addition to a set of initial screening models, relevant secondary tests would be conducted to generate additional information on specificity, mechanism of action, and possible side effects.

A complete in vivo pharmacology screening program would include models for the detection of the major therapeutic classes of compounds shown in Table 1.

The following sections describe some representative models currently in use in screening for new

pharmacological compounds. Descriptions of most of these models were derived with permission from standard operating procedures of the former Pharmakon Laboratories, Waverly, PA 18471. Many other models, as well as variations on the models to be described, are available (4–6).

CNS

A convenient test for initial screening of compounds for CNS activity would be the *neuropharmacologic profile* or open field behavior test in mice or rats. In this test, animals are dosed with the compound and placed individually in an open area consisting of a square box about 2–4 feet on each side with an open top. The animals are then observed continuously for alertness, depression, spontaneous motor activity, defecation, urination, grooming behavior, ataxia, and convulsions. The animals are systematically observed to measure time of onset, peak effect, duration, and character of drug action. This kind of test can provide an initial pharmacological appraisal of chemical compounds and an estimate of acute toxicity. CNS depressants, tranquilizers, antidepressants, central and peripheral skeletal muscle relaxants, psychostimulants, analgesics, anticonvulsants, and diuretics may be detected by this test. Other classes of drugs, however, will generally not be detected.

The *Randall–Selitto assay* is a specific test for analgesia. In this test, a suspension of 20% brewer's yeast in water is injected into one hind foot of a rat. This causes an edematous condition that increases the rat's sensitivity to pain. The test compound is administered 2 h after the yeast injection, and after an additional hour, the pain thresholds of the inflamed and noninflamed paws are measured. This is done with a device, the Analgesy Meter (Ugo Basile, Milan, Italy), which exerts a force that increases at a constant rate. The rat paw is positioned on a teflon platform, and the force is applied to the volar surface with a blunt teflon cone.

The *phenylquinone writhing assay* is also a test for analgesic activity. In this test, mice are given an intraperitoneal injection of 0.25 ml of 0.02% phenylquinone in 5% ethanol. This will cause 100% of the mice to writhe within 10 min. Characteristic patterns of writhing consist of torsion of the abdomen and thorax, drawing the hind legs close to the body, and raising the heels off the floor. Complete blocking of the writhing syndrome for 10 min after the phenylquinone injection is considered a positive response.

The ability of a test compound to cause reversal of tetrabenazine-induced ptosis in mice is an indication of

Table 1 Major therapeutic classes

CNS	Sedative–hypnotic
	Anticonvulsant
	Muscle relaxant
	Anesthetic
	Analgesic (narcotic and non-narcotic)
	Anxiolytic
	Dementia/cognition
	Antidepressant
	Neuroleptic/antipsychotic
Cardiovascular	Antihypertensive
	Anti-arrhythmic
	Cardiotonic
	Antithrombic
	Anti-anginal
	Cerebrovasodilator
Metabolic	Hypolipidemic
	Hypoglycemic
	Diuretic
Immunopharmacologic	Anti-inflammatory
	Anti-edema
	Anti-allergenic
	Anti-asthmatic
	Bronchodilator
Gastrointestinal	Immunostimulant
	Anti-ulcer
	Gastrokinesis
Antimicrobial	Antisecretory
	Antibacterial (Gram positive and negative)
	Antiviral
	Antifungal
	Antitrichomal

CNS stimulant effect. Mice are dosed with test compound and then given an intraperitoneal injection of 32 mg/kg of tetrabenazine methane sulfonate. Thirty minutes after the tetrabenazine injection, the mice are graded for the degree of ptosis. Lack or partial reversal of ptosis is considered a positive response.

Inhibition of pentylenetetrazol-induced seizures is used as an indication of hypnotic, tranquilizer, or anti-convulsive activity. Mice or rats are given the test drug, and after 30 min a dose of pentylenetetrazol, which will cause 50% of the animals to convulse (CD50), is administered intramuscularly. A drug that increases the CD50 may have activity of the type mentioned above.

Cardiovascular

Spontaneously hypertensive rats can be used to screen compounds for antihypertensive effects and for effects on heart rate. The animals are dosed for one or a few days. Blood pressure and heart rate are measured by means of an inflatable cuff around the tail. Most classes of anti-hypertensives will be detected. Agents such as beta-adrenergic antagonists will be detected by decreased heart rate. Other rat models include deoxycorticosterone acetate (DOCA)-induced hypertensive, renal hypertensive (one or both renal arteries clamped), and stroke-prone spontaneously hypertensive rats. Hypertensive dogs produced by clamping one or both renal arteries may also be used to test or verify antihypertensive activity in a second species.

Because of the ease of introducing intravenous and intra-arterial catheters and measuring blood flow and blood pressure, dogs are commonly used to conduct hemodynamic studies. These studies evaluate the effect of the test compound on systolic and diastolic blood pressure, heart rate, cardiac output, dp/dt , respiration, ECG, and ventricular pressure. From these data, effects desirable for treating angina pectoris, congestive heart failure, coronary vasospasm, and myocardial infarction can be detected.

Anti-Inflammatory

Two commonly used models to detect anti-inflammatory activity are carrageenan-induced paw edema and adjuvant-induced polyarthritis in rats. The former represents an acute and the latter a chronic inflammatory process. Nonsteroidal anti-inflammatory agents inhibit the formation of carrageenan-induced paw edema. However, to detect activity in the developing and established phases of chronic inflammation, the polyarthritis model is well accepted. In both tests the measured endpoint is volume of

the hind paw. This is done by immersing the hind paw in the well of a mercury displacement plethysmograph. In the carrageenan test, 0.1 ml of a 1% suspension of carrageenan is injected into the plantar surface of one hind paw of a rat. After 1 h the test compound is administered, and after a further 3 h the volumes of both the carrageenan-treated and nontreated hindpaws are measured. The net difference between the two measurements is an indication of the amount of edema remaining after test compound treatment. The smaller the net difference, the greater the anti-inflammatory activity of the test compound. In the polyarthritis model, 0.1 ml of a sterile 0.75% suspension of finely ground *Mycobacterium tuberculosis* or *Mycobacterium butyricum* (Freund's adjuvant) in paraffin oil is injected into the hind paw as done for carrageenan. The polyarthritis disease state takes about 20 days to develop. Therefore, to test a compound for activity in the developing stages of chronic inflammation, the test compound is administered during the first 20 days after the single injection of Freund's adjuvant. To test a compound for activity against fully developed chronic inflammation, the test compound is administered for 4–10 days, starting about 20 days after the Freund's adjuvant injection. In both the carrageenan and polyarthritis models, drug effects are expressed as the percentage inhibition of paw edema in drug-treated animals compared to non-drug-treated controls.

In the *croton oil topical inflammation* test in mice, 20 μ l of a 3% solution of croton oil is applied topically to the anterior and posterior surfaces of one ear and 20 μ l of the test drug is applied to the same ear 30 min later. Inhibition of ear swelling by more than 50% relative to vehicle-treated controls measured 2 h after croton oil application indicates acute topical anti-inflammatory activity.

Immunomodulation

Inhibition of oxazolone-induced delayed-type hypersensitivity is used as an *in vivo* test for inhibition of the cellular immune response. The shaved abdominal surfaces of groups of mice are sensitized by topical application of 0.1 ml of 5% oxazolone. The test drug is given intraperitoneally an hour later and then daily for five consecutive doses. After an additional 4 days, the animals are challenged by application of 0.5 ml oxazolone to one ear. Ear thickness is measured 24 h later. A decrease in ear thickness relative to untreated controls is an indication of immunosuppression.

Inhibition of sheep red blood cell lysis by serum from sensitized mice is an indication of suppression of the humoral immune response. Groups of mice are sensitized

by the intravenous injection of 0.2 ml of a 2% sheep red blood cell suspension. The test drug or vehicle is administered intraperitoneally to groups of mice for 3 consecutive days beginning 2 h after sensitization to sheep red blood cells. Nine days after sensitization, blood samples are collected, complement-inactivated serum prepared, and serial dilutions made. Sheep red blood cells and complement are added to aliquots of the serum dilutions. The serum titer is then expressed as the reciprocal of the dilution causing complete hemolysis. A lower than normal titer may indicate immunosuppression, and a higher than normal titer may indicate immunostimulation.

Antitussive

Dogs are surgically prepared by implantation of a small iron slug in the lumen of the trachea (7). After a 2-week recovery period, the cough response of the dogs is determined in response to electromagnetic stimulation of the implanted tracheal slug. The stimulus is imposed for 3- to 5-s periods, and the total number of coughs elicited during and 30 s after stimulation is recorded. After establishing each dog's control cough response, test substances can be administered at weekly intervals and compared with the effects observed after administration of a standard antitussive agent.

Endocrine

A simple assay for antidiabetic activity is the determination of changes that occur in blood glucose levels in rats after oral administration of a test compound followed by a subcutaneous injection of glucose. Rats are fasted overnight, administered the test compound, usually by gavage, and then immediately given a subcutaneous injection of 125 mg glucose. Two hours after test compound administration, the rats are anesthetized and blood drawn for plasma glucose determination. Decreased plasma glucose relative to controls not treated with test compound is an indication of antidiabetic activity.

Antimicrobial

Many animal models use infections artificially introduced into the animal either systemically or into specific organs or tissues. The behavior of infections and the efficacy of antibiotics in eliminating those infections differ, depending on the organ or tissue involved. The type of local pathology, the penetration and pharmacokinetics of antibiotics, the local host defense system, and the clearance rates for the bacterial inocula differ among the

various organs and tissues of an animal. Therefore, to get a reliable prediction of how effective an antibiotic will be clinically it is mandatory to test new antibiotics *in vivo*. There are a great variety of such models. For descriptions of many of these, the reader is referred to a three-volume series (8).

Anticancer

New drugs suspected of having anticancer activity are tested in standard animal models in which tumors are transplanted from one generation of animals, generally mice, to the next and are thus perpetuated *in vivo*. *In vitro* tumor cell culture systems are commonly used to determine if a new drug has cytotoxic or growth-inhibiting activity. However, the *in vivo* test is considered to be a better predictor of activity in humans.

The L1210 lymphoid leukemia model is commonly used. This tumor is carried as an ascites tumor in BDF1 or CDF1 mice. Ascites fluid from carrier mice containing 100,000 cells is implanted in the test mice. Twenty-four hours later, treatment with the test drug is started. After 6 or 7 days of treatment, the numbers of survivors are determined and the animals weighed. The results are evaluated by comparing drug-treated groups to untreated ones. Drugs resulting in a greater proportion of surviving mice compared with non-drug-treated controls may have anticancer activity. Many other similar models are available (9).

ANIMALS IN DRUG SAFETY TESTING

General Principles

Before a new drug can be given to people, it must be tested in animals to determine its side effects and at what dosage those side effects will appear. This testing must be done *in vivo* because the effects of the processes of absorption, distribution, metabolism, excretion, the interactions among these processes, and the interactions among the various organs and neuroendocrine systems within the whole animal cannot be duplicated *in vitro*. To characterize the nature of the side effects to be expected from a new drug, it is usually necessary to give much higher dosages than would be given clinically and sometimes to give the drug over prolonged periods of time. This is because the dosages of drugs needed to elicit pharmacological or toxicological effects are often higher in laboratory animals than in humans, and because the side effects of drugs suitable for clinical use are usually

provoked only by exaggerated dosages. In some cases the drug may accumulate within the body or within particular organs or tissues and thus give rise to toxic manifestations. Some side effects appear only after long periods of repeated administration.

In addition to the requirement to test new drugs in animals, it is also necessary to test in more than one species of animal, because extrapolation of the results of testing in animals to humans is less than perfect. Certain species may be more sensitive or predictive than others. To ensure that the safety of a new drug has not been overpredicted and that any potential side effects have not been overlooked, it is routine practice to test new drugs for safety in at least two species. In most cases, these will be a rodent and a nonrodent species.

The variety of animal tests and the length of the studies required for a new drug depend on the nature of the drug (pharmacological or chemical class), the intended clinical use of the drug (for example, length of the usual course of treatment), and, to some extent, on the requirements of the countries in which the new drug will be registered for marketing. These factors will be discussed further in the following sections.

Acute Toxicology

In acute toxicology testing, animals are given single doses of a drug. The most common study design is to give groups of rats or mice (such as 5 per sex) single treatments over a wide range of dosages and then observe them over a period of 7–14 days for survival and for physical or behavioral signs of toxicity. In most cases a LD₅₀ (the dosage at which 50% of the animals die) will be determined; however, this is not always feasible or necessary. The acute toxicity test may be conducted for several reasons. This is the first toxicity test that will be done for a new drug, and it provides the basis for choosing dosages that will be used in subsequent toxicity studies involving repeat dosing. This test provides information regarding the effects of single large doses of a new drug. This information may be used to predict the effects of overdosing (either accidental or purposeful) in humans. The acute toxicity test is a convenient safety check on new batches of drug. Even though extensive chemical analyses are conducted on new batches of any drug, a quick test in live animals to assure that there have been no dramatic changes in acute toxicity provides a desirable level of comfort before making the new drug batch available for human use. The acute toxicity test is also a convenient method to determine any interaction between two drugs that would cause unexpected toxicity in clinical use. Thus, if a new drug will be commonly used clinically in

combination with other drugs or in patients who commonly receive certain medications, it might be prudent to test the new drug in combination with these other drugs in an acute toxicity test in order to predict such an interaction before using the new drug clinically. A positive finding in such a combination study would likely be followed up by longer-term, repeat-dosing studies.

Acute toxicity studies may be conducted by administering the drug by any of several routes (oral, intravenous, intramuscular, intraperitoneal, subcutaneous, or dermal). The route chosen usually is that intended to be used clinically. Rats and mice are generally used for acute toxicity work; however, rabbits are commonly used when the route of administration is dermal.

If a drug is to be given parenterally, acute vein and muscle irritation studies are commonly conducted in rabbits. These studies help assure that a new parenteral formulation will not cause unnecessary pain or tissue damage in people. They are also useful in developing new formulations for parenteral use with minimal potential for such effects. In a vein irritation study, groups of drug-treated and vehicle-treated rabbits are given injections via the lateral ear veins. The reason rabbits are used for this test is because the lateral ear veins are large enough, long enough, and easily visible, so as to facilitate making the injections and identifying the injected portion for subsequent sampling for histological examination. The animals are observed during the injection and at intervals (such as 2–4 h) afterward for indications of discomfort during the injection and of swelling, erythema, or tenderness of the injection site. Twenty-four hours after injection, the animals are killed, and the injection site veins are examined histologically for evidence of tissue damage or inflammation.

In a typical muscle irritation study, groups of male rabbits are given intramuscular injections of 1.0 ml of drug formulation or vehicle with a 23-gauge needle inserted vertically to a depth of about 15–20 mm into the sacrospinalis muscle. The injection site is marked by drawing a circle of ink around the site. The animals are observed during the injection and daily thereafter for 7 days for discomfort during injection and for signs of swelling, erythema, or tenderness at the injection site. On 1, 3, and 7 days after injections, blood samples are taken for measurement of creatine phosphokinase activity. Elevation of this enzyme in blood is an indication of muscle damage. Also, 1 and 7 days after injection, rabbits from each group are sacrificed and the muscle tissue surrounding the injection site taken for histological examination for evidence of tissue damage or inflammation.

Subchronic and Chronic Toxicology

Repeat-dosing toxicity studies are conducted to determine what side effects will arise from repeated administration of a drug at lower dosages than those used in acute toxicity studies and to determine safe dosages to be used in the initial human clinical trials. These studies range in duration from 1 to 2 weeks to 1 to 2 years. The length of studies required for a drug depends mainly upon the duration of treatment and the intended clinical dosing regimen. (This is discussed further under Governmental Guidelines). These studies are conducted in stages so that the results of one study can be used to design the subsequent study of longer duration. The first are usually 2 weeks in length followed by 1-month, 3-month, 6-month, and then 1-year studies. Parallel subchronic and chronic studies are almost always conducted in two species, usually the rat and dog, because there is a large historical data base for these species and they are easy to work with, relatively economical to house, and readily available from commercial vendors. However, special circumstances may dictate that other species be used. For instance, if a drug causes excessive vomiting in dogs, which are known to be particularly sensitive to such effects, then some other nonrodent species such as the monkey may have to be used. Also, if the absorption or metabolic handling of a drug in the rat or dog is found to be markedly different from that in humans such that one of these species would not be a reasonable predictor of toxicity in humans, then another species would be used.

The usual protocol for subchronic and chronic studies includes groups of animals containing equal numbers of both sexes receiving at least three dosage levels of drug plus vehicle or other control groups. These animals are observed daily for clinical signs of toxicity. Their body weights and food consumption are measured frequently. These three parameters—clinical signs, body weight, and food consumption—can be very sensitive indicators of toxicity. Complete hematology and serum chemistry profiles are determined at least at the end of the administration period and in some cases at intervals during the period of administration. Thorough physical examinations by a veterinarian or a trained technician are conducted at regular intervals. Periodic electrocardiograms are commonly recorded in studies with dogs. At the end of the period of drug administration, all the animals are subjected to a complete necropsy under the supervision of a veterinary pathologist. The tissues are then subjected to complete microscopic examination by a veterinary pathologist to detect morphologic alterations in the tissues that may have resulted from drug administration. In some cases it may be desirable to allow some

animals from the drug-treated groups to live for a period of time after the end of drug administration to determine if any drug-related changes will disappear upon withdrawal of the drug.

As mentioned, the results of the subchronic and chronic studies are used to help determine the dosage to be used in the initial human clinical trials. To do this, the lowest dosage causing no toxicity (the *no-toxic-effect dosage*) is determined for each study, and a safety factor is applied depending on the species. For instance, for rats the no-toxic-effect dosage is divided by 10 and for dogs by 6 to arrive at an estimate for the initial human dosage. These factors are derived from the observations that laboratory animals can usually tolerate higher dosages of drugs and other chemicals without exhibiting toxicity than can humans and that the differences in tolerance vary with differences in basal metabolic rate, which in turn varies with body surface area to weight ratio. This ratio varies by a factor of approximately 10 in rats and 6 in dogs relative to humans. The descriptions of toxic effects elicited in the subchronic and chronic studies allow clinicians conducting clinical trials to know which side effects to anticipate so as to protect the patient volunteers.

Reproductive Toxicology

Reproductive toxicology covers the entire process of reproduction from mating to pregnancy, birth, weaning and, sometimes, the reproductive function of subsequent generations. The species most commonly used in these studies are rats and rabbits, although in special circumstances, other species such as mice, dogs, or monkeys may be used. A series of studies is usually conducted so as to cover all the phases reproduction. Within each study, groups of animals corresponding to untreated controls and two or three drug-treated groups are used.

One type of typical study, usually conducted in rats, provides information on fertility and general reproductive performance, particularly on gonadal function, estrous cycles, mating behavior, conception rates, and early stages of gestation. This study is usually referred to as a fertility study. Mature male rats are treated with the drug for 28 days prior to mating. This period is considered to be sufficient to detect effects on male reproductive organs when used in conjunction with thorough histological evaluation of testes. This fertility study in male animals is conducted prior to large-scale (Phase III) clinical trials in men. Female rats are treated for 14 days and then mated to treated males. Daily vaginal inspections are made and the finding of sperm or a copulation plug is considered as Day 0 of pregnancy. The females are sacrificed on Day 13,

and the number of live and dead embryos, the number of implantation sites, and the number of corpora lutea on the ovaries determined. Some of the dams may be allowed to deliver normally, and the newborn pups are counted, examined, sexed, and weighed. The pups are again weighed and counted on Days 4 and 21 (weaning). From this test, preimplantation death, derived from the number of implantation sites in the uterine horns compared to the numbers of corpora lutea on the respective ovaries, postimplantation death, derived from the number of resorption sites compared to the number of implantation sites in the uterus, and the survival and growth of the offspring are determined. This study is generally conducted prior to any clinical trials involving women of childbearing potential. However, in the USA and Europe, women of childbearing potential confirmed to be nonpregnant and using effective birth control may be enrolled in Phase I and Phase II (usually small-scale) studies prior to completion of the female fertility study.

The second typical study, commonly called the teratology study, is usually conducted in two species, rats and rabbits. Rabbits are commonly used as a second species in Segment 2 studies because after the thalidomide tragedy in the early 1960s, rabbits were found to be more susceptible than were other commonly used laboratory species to the teratogenic effects of thalidomide, one of the few proven human teratogens. In the Segment 2 test, inseminated females are treated only during the organogenesis period, which for rats and mice are Days 6–15 and for rabbits are Days 6–18 of gestation. One day prior to birth, the dams are sacrificed and the fetuses delivered by cesarean section. The fetuses are weighed, sexed, and examined for gross abnormalities. The fetuses from each treatment are then randomly divided into two groups. One group is preserved in fixative for subsequent soft tissue examination. In the other group, the tissues are cleared and the skeletons stained for skeletal examination. The main endpoints in this study are numbers of live and dead fetuses, sex ratio of the fetuses, fetus weights, grossly observable abnormalities, and soft tissue and skeletal abnormalities. This test procedure was developed based on the well-established principle that the embryo is most susceptible to induction of birth defects during the organogenesis period. Previous exposure of the dam, as in the Segment 1 test, can lead to infertility or elevation of detoxification mechanisms, and thus to underestimation of the potential of a compound to cause birth defects. A separate phase of testing was developed to specifically assess the occurrence of birth defects. This study is usually completed prior to any human studies involving women of childbearing potential (proven nonpregnant and using

effective birth control) in Europe and Japan and prior to large-scale Phase III studies involving women in the USA.

In a third type of study, called the perinatal/postnatal test, pregnant females are treated during the last quarter of pregnancy and through lactation until weaning. The weanlings are necropsied to detect any internal anomalies. The purpose of this study is to determine the effects of the drug on late fetal development, labor, delivery, lactation, neonatal viability, and growth. It also assesses potential effects on neonates due to drug being excreted in the mother's milk. The endpoints for this study are the numbers of live and dead pups, the sex ratio of the pups, the numbers of surviving pups at 4 and 21 days after birth, and the growth and development of the pups. Postnatal testing of pups commonly includes developmental indices such as days on which the following events occur: balanopreputial separation in males, vaginal opening in females, eye opening, pinna unfolding, and incisor eruption. Tests for reflex development and learning and memory are also commonly included. The reproductive competence of the offspring may also be tested. The potential of a compound to induce changes in these endpoints can be masked by infertility in the fertility and general reproductive performance study, which would lead to insufficient numbers of offspring to detect a change. This study must be completed prior to applying for marketing approval.

Mutagenicity

New drugs are routinely screened for their potential to cause gene mutations and/or chromosome aberrations by *in vitro* tests. An *in vivo* test is also required by most regulatory agencies. Further mutagenicity tests in whole animals are usually conducted only if the *in vitro* tests prove inconclusive or questionably positive. A new drug showing a strong positive response in an *in vitro* mutagenicity test probably would not be developed. *In vivo* tests serve to determine whether a drug has potential to cause genetic alterations under conditions closer to those that one would obtain under clinical use. The significance of these studies to the safety of a drug is twofold. First, a drug that causes genetic damage could produce such damage in the sperm or ovaries of the patient and thus potentially cause genetic abnormalities in the children of that patient. Second, it is widely accepted that any agent capable of causing genetic damage is also likely to be a carcinogen. The most common *in vivo* tests for genetic damage are chromosome damage tests and the dominant lethal test in mice or rats.

Chromosome damage can be assessed in mice and rats by bone marrow cytogenetic analysis and by bone

marrow micronucleus assay. In both methods, bone marrow is examined after animals have been treated with either a single high dose or after several (2–4) daily doses of the drug. After the single dose, bone marrow is collected 6, 24, and 48 h after dosing for cytogenetic analysis or 24, 48, and 72 h after dosing for micronucleus determination. After the repeat dosing schedule, bone marrow is collected 6 h or 24–30 h after the last dose for cytogenetic or micronucleus assay, respectively. For cytogenetic analysis, the animals are treated with colchicine 2–3 h prior to euthanasia to block dividing bone marrow cells in metaphase, femoral bone marrow is collected and fixed, and metaphase spreads are prepared on microscope slides. After Giemsa staining, the numbers of chromosomes and their normal or abnormal appearance in metaphase spreads are determined. In the micronucleus test, femoral bone marrow smears are stained with Giemsa or a fluorescent dye such as acridine orange, and the numbers of polychromatic erythrocytes containing micronuclei are determined. Normal polychromatic erythrocytes contain no nuclei. In the process of red blood cell maturation, the nucleus is expelled just prior to the polychromatic erythrocyte stage. Micronuclei represent chromosomes or chromosome fragments that are left behind in the cell after expulsion of the nucleus.

The dominant lethal assay is used to detect mutagenic damage to male germ cells. The principle behind this assay is that theoretically there are a large number of chromosomes that, when damaged, could lead to lethality during early gestation if included in a fertilized egg. Hence, the target size relative to the genome is large for dominant lethality, and therefore the test should be a sensitive indicator of genetic damage. In this test, male animals (usually rats or mice) are treated with the drug daily for 5 days. They are then mated with separate groups of females each week for 8 weeks. The 8-week interval covers the entire spermatogenic cycle. The females are killed at midpregnancy, about 14 days after the midweek of mating. At necropsy the uteri are examined and the numbers of corpora lutea, living implantations, and dead implantations are counted. The endpoints for this assay are (1) the numbers of preimplantation losses, determined by subtracting the numbers of implants from the numbers of corpora lutei, and (2) the numbers of dead implants. Agents that cause increased numbers of preimplantation losses or dead implants compared to vehicle or historical control levels are considered positive in the dominant lethal assay. Because of the serial mating scheme, it is possible to know which stage of spermatogenesis is affected if a positive result is obtained.

Several transgenic animal models are available for mutagenicity testing *in vivo*. The most widely used are the Big BlueTM and MutaTM Mouse models. These animals have portions of the *E. coli* galactosidase operon inserted into their genome. This bacterial DNA is used as the target for mutagenicity testing when the animals are treated with test chemicals. After chemical treatment, the bacterial DNA is removed from the tissues and packaged into bacterial viral (phage) particles. The phages are used to infect bacteria growing on agar plates. The bacteria multiply to form colonies that can be counted. Because of a color reaction that can be catalyzed by the galactosidase protein, a color reaction formed in the bacterial colonies can be used to assess the presence or absence of mutations formed in the target DNA from virtually any tissue within the treated mice.

Carcinogenicity

Carcinogenicity bioassays are long-term studies in mice and rats conducted according to standard guidelines established by the National Cancer Institute (10) and the International Conference on Harmonization (11). In these studies the animals are treated repeatedly for periods of 18–24 months for mice and of 24 months for rats. Treatment is by a route consistent with the intended clinical route. The study design usually includes vehicle control and two or three drug-treated groups. The highest dosage is selected as the maximum dosage that will be tolerated by the animals. This dosage may cause minimal signs of toxicity but should cause no more than a 10% decrement in body weight gain compared with control groups and should not cause toxicity, other than that related to a neoplastic response, that would be predicted to shorten the animal's natural life span. This high dosage is necessary to give a maximum test of the potential of a new drug to cause cancer. The animals are examined frequently for palpable masses. At the end of the treatment period, the animals are necropsied and histologic examinations of the tissues are conducted with particular attention given to the detection and identification of tumors. Recently, transgenic strains of mice that allow detection of a carcinogenic response within 6 months of treatment rather than 2 years have become available. Examples of these are the P53^{+/-} (P53 knockout), TG.AC and the rasH2 mouse models. These animals contain deletions of genes involved in suppressing the formation of tumors. The use of these animal models is justified based upon the fact that these same genes are known to be involved in many human tumors. Thus, the mechanisms leading to enhanced tumorigenic response in these animals is known to be relevant to human tumorigenesis.

Primary Irritation Testing

If a new drug is intended to be applied to the skin or eyes, one of the first tests to be conducted would be to determine if the drug, or the formulation containing the drug, will cause irritation of the skin or eyes. Even if a drug is intended only for dermal application, eye irritation testing may also be required because of the possibility of inadvertent exposure to the eyes. For instance, an antibiotic intended for topical treatment of acne would be applied on the face, and the potential for eye contact would be considerable. These tests are usually conducted in rabbits because of their widely accepted sensitivity to dermal and ocular irritation and because of the scientific literature describing such tests in rabbits. Other species, such as guinea pigs or mice, may be used for dermal irritation studies under special circumstances.

A typical dermal irritation assay is conducted as follows. Six male albino rabbits are clipped free of hair on the back. One area of skin is left intact, whereas another is abraded in a tic-tac-toe pattern with the point of a hypodermic needle so as to incise the superficial epidermis layer without causing bleeding. The test material, 0.5 ml of liquid or 0.5 g of solid or semisolid is applied to each site under a 1 × 1 in. gauze pad. The entire trunk of the animal is wrapped with an impervious material and held in place with tape for 24 h. The patches are then removed and excessive material wiped off. The skin reactions are scored at 24 and 72 h after the initial application according to a scheme such as that listed in Table 2.

Table 2 Dermal irritation scoring system

Skin reaction	Value ^a
Erythema and eschar formation	
No erythema	0
Very slight erythema (barely perceptible)	1
Well-defined erythema	2
Moderate to severe erythema	3
Severe erythema (beet redness) to slight eschar formation (injuries in depth)	4
Edema formation	
No edema	0
Very slight edema (barely perceptible)	1
Slight edema (edges of area well defined by definite raising)	2
Moderate edema (raised approximately 1 mm)	3
Severe edema (raised more than 1 mm and extending beyond the area of exposure)	4

^aThe value recorded for each reading is the average value of the six or more animals subjected to the test.

The mean values of the six rabbits for erythema and eschar formation at 24 and 72 h for both intact and abraded skin (four values) are added. The mean values of the six rabbits for edema at 24 and 72 h (four values) are also added. The total of eight values is divided by 4 to give the primary irritation index. Values of 5 or greater are considered indicative of a positive irritant (12).

A typical ocular irritation test is conducted as follows. Six albino rabbits of the same sex are given thorough ophthalmological examinations within 24 h before use to ensure absence of preexisting ocular damage. The animals would be firmly but gently restrained. The test formulation is placed in one eye of each animal by gently pulling the lower lid away from the eyeball (conjunctival cul-de-sac) to form a cup into which the test substance is dropped. The lids are then gently held together for one second and the animal released. The other eye, remaining untreated, serves as a control. For testing liquids, 0.1 ml is used. For solid, paste, or particulate substances, the amount used must have a volume of 0.1 ml weighing not more than 100 mg. The eyes of each rabbit would be examined 24, 48, and 72 h after treatment. After the 24-h examination, the treated eyes may be rinsed with tap water or saline. At this time the eyes may also be examined with the aid of fluorescein stain. One drop of fluorescein sodium ophthalmic solution USP is dropped directly on the cornea. After flushing out the excess fluorescein with tap water or saline, injured areas of the cornea appear yellow and are best seen under ultraviolet illumination. At each examination period, the eyes of each rabbit are scored for ocular reaction according to a scheme such as that listed in Table 3.

The test may be considered positive if three or more animals exhibit positive reactions at any observation period. Equivocal tests are repeated. For humane reasons, formulations known to be corrosive or to be severely irritating in the dermal irritation test are assumed to be ocular irritants and not tested (13).

Antigenicity Testing

New drugs may undergo testing for antigenicity depending upon the chemical structure (suspected of being a sensitizer because of similarity to known sensitizers) or because of the intended use of the drug. Formulations intended for topical use are usually tested for delayed contact sensitivity. However, new drugs intended for systemic use are sometimes tested for their potential to induce anaphylaxis. These tests are commonly conducted in guinea pigs, but rats or mice are sometimes used.

Table 3 Ocular irritation scoring system

Ocular reaction	Score
Cornea	
No ulceration or opacity	0
Scattered or diffuse areas of opacity (other than slight dulling of normal luster, details of iris clearly visible)	1 ^a
Easily discernible translucent areas, details of iris slightly obscured	2
Nacreous areas, no details of iris, size of pupil barely discernible	3
Opaque cornea, iris not discernible through the opacity	4
Iris	
Normal	0
Markedly deepened rugae, congestion, swelling, moderate circumcorneal hyperemia, or injection, any of these or any combination thereof, iris still reacting to light (sluggish reaction is positive)	1 ^a
No reaction to light, hemorrhage, gross destruction (any or all of these)	2
Conjunctivae	
Blood vessels normal	0
Some blood vessels definitely hyperemic (injected)	1
Diffuse, crimson color, individual vessels not easily discernible	2 ^a
Diffuse beefy red	3
No swelling	0
Any swelling above normal (includes nictitating membranes)	1
Obvious swelling with partial eversion of lids	2 ^a
Swelling with lids about half closed	3
Swelling with lids more than half closed	4

^aReadings at these numerical values or greater indicate positive response.

Delayed contact hypersensitivity can be evaluated by several standard protocols in guinea pigs. A typical procedure is to sensitize groups of 10 guinea pigs each by injecting 0.1 ml (0.05 ml for the first injection) of a 0.1% suspension or solution of test material intradermally in the upper dorsal area three times per week for a total of 10 injections. Additional groups are similarly treated with a positive control compound such as dinitrochlorobenzene

Table 4 Dermal hypersensitivity scoring scheme

Skin reaction	Score
Erythema and eschar formation	
No erythema	0
Very slight erythema (barely perceptible)	1
Well-defined erythema	2
Moderate to severe erythema	3
Severe erythema (beet redness) to slight eschar formation (injuries in depth)	4
Edema	
No edema	0
Very slight edema (barely perceptible)	1
Slight edema (edges of area well defined by definite raising)	2
Moderate edema (raised approximately 1 mm)	3
Severe edema (raised more than 1 mm and extending beyond area of exposure)	4

and a negative control (saline). About 2 weeks after the tenth injection, the animals are challenged by intradermal injection of 0.05 ml of the respective test compound, positive control, or saline in the dorsal lumbar area. The animals are shaved at least 18 prior to each injection. At 24 and 48 h after each sensitizing and challenge injection, the area of skin around the injection site is graded for erythema and edema according to the scheme shown in Table 4.

The erythema and edema grades are recorded separately. A dermal irritation score for each sensitizing injection and the challenge injection for each animal are calculated by adding the 24-h and 48-h scores for both erythema and edema (four scores) and dividing by 4. The dermal irritation scores obtained from the initial sensitizing injection and the challenge injection for each animal are used to determine whether sensitization had been produced. When the challenge score exceeds the initial irritation score for a given animal by 1 or more, the animal is considered sensitized.

The potential of a compound to induce anaphylaxis can be evaluated in both active systemic anaphylaxis (ASA) and passive cutaneous anaphylaxis (PCA) tests in guinea pigs or rats (14, 15). In both tests, the test compounds and positive and negative control substances are injected subcutaneously or intraperitoneally three times per week to groups of 6–10 animals for a total of 6 injections. In the ASA test, the animals are challenged 2 weeks after the final sensitization injection intravenously with the respective test compound or positive or negative control and observed for signs of anaphylaxis. These would be graded as:

Negative: No anaphylactic symptoms

Mild: Restlessness, rubbing of the nose or ears, coughing, expiratory acceleration, or piloerection

Moderate: Urination, defecation, dyspnea, or ataxic gait

Severe: Convulsion followed by recovery

Fatal: Death by a fatal anaphylactic reaction

In the PCA test, serum is collected from the sensitized animals 2 weeks after the final sensitization injection. This serum is diluted with saline to a dilution of 1/8 or more and injected intradermally on the backs of previously untreated guinea pigs. Four hours later, these guinea pigs are given intravenous injections of 1.5–2.0 ml of saline containing 0.5% Evans Blue plus the respective test compound or positive or negative control. Thirty minutes after the intravenous injection, the animals are examined for the presence of blue areas of skin surrounding the earlier intradermal injections. A blue spot having a diameter of 5–10 mm is considered a positive reaction.

Governmental Requirements

Governmental regulatory agencies such as the USFDA have established guidelines describing the kind of safety tests that should be conducted in animals in order to

Table 5 Duration of repeated dose toxicity studies to support phase I and II trials in the EU and phases I, II and III trials in the United States and Japan^a

Duration of clinical trials	Minimum duration of repeated dose toxicity studies	
	Rodents	Nonrodents
Single dose	2–3 weeks ^b	2 weeks
Up to 2 weeks	2–4 weeks ^b	2 weeks
Up to 1 month	1 month	1 month
Up to 3 months	3 months	3 months
Up to 6 months	6 months	6 months ^c
6 months	6 months	Chronic ^c

^aIn Japan, if there are no Phase II clinical trials of duration equivalent to the planned Phase III trials, conducting longer duration toxicity studies should be considered, as given in Table 6.

^bIn the EU and United States, 2-week studies are the minimum duration. In Japan, 2-week nonrodent and 4-week rodent studies are needed. In the United States, as an alternative to 2-week studies, single dose toxicity studies with extended examinations can support single dose human trials.

^cData from 6 months of administration in nonrodents should be available before the initiation of clinical trials longer than 3 months. Alternatively, if applicable, data from a 9-month nonrodent study should be available before the treatment duration exceeds that which is supported by the available toxicity studies.

Table 6 Duration of repeated dose toxicity studies to support phase III trials in the EU and marketing in all regions^a

Duration of clinical trials	Minimum duration of repeated dose toxicity studies	
	Rodents	Nonrodents
Up to 2 weeks	1 month	1 month
Up to 1 month	3 months	3 months
Up to 3 months	6 months	3 months
3 months	6 months	Chronic

^aAlso reflects the marketing recommendations in the three regions except that a chronic nonrodent study is recommended for clinical use >1 month.

have a new drug approved for use in clinical trials and in order to get approval of a new drug application (NDA) for marketing. The rationale and circumstances for conducting reproductive, mutagenicity, carcinogenicity, irritation, and sensitization studies have already been mentioned. Tables 5 and 6 summarize the requirements for acute, subacute, and chronic toxicity studies for pharmaceutical products intended for use in humans. These descriptions are limited to the requirements of the United States, Japan, and Europe because these areas represent the largest pharmaceutical markets in the world today. These requirements have been developed at the International Conference on Harmonization to provide uniformity among the three regions (16). Phases I, II, and III refer to the different phases of human clinical trials. Phase I refers to the initial trials, limited to one or a few doses to determine absorption, pharmacokinetics, and an initial estimate of safety. Phase II refers to larger-scale studies to establish safety and to get an initial estimate of clinical efficacy. Phase III refers to the final, large-scale, multicenter trials aimed at establishing efficacy.

CARE OF LABORATORY ANIMALS

Proper care of laboratory animals used in research is a basic requirement to assure the validity and reproducibility of the results obtained. Animals used in drug research are subject to the strictest standards of care beginning with the animal supplier. For the most commonly used laboratory animals, these standards of care often apply for the entire life of the animal. Guidelines for proper care of research animals are provided by the Department of Health and Human Services. The American Association for the Accreditation

of Laboratory Animal Care provides the service of certifying laboratories complying with those guidelines. All reputable industrial drug development houses strive to achieve and maintain certification by the association.

Sources of Laboratory Animals

Laboratory animals for drug research in industry are virtually always obtained from reputable animal suppliers who observe strict standards of care of the animals they provide. Animals that are not properly cared for often result in lost time—and therefore lost money—in drug development.

Examples of suppliers of laboratory animals are listed below.

Charles River Labs
251 Ballardvale St.
Wilmington, MA 01887-1000
www.criver.com

COVANCE Research Products
R.D. No. 2
Swampridge Rd.
Denver, PA 17517
www.covance.com

Harlan Sprague Dawley
P.O. Box 29176
Indianapolis, IN 46229-0176
www.harlan.com

Jackson Laboratories
600 Main St.
Bar Harbor, ME 04609
www.jax.org

Kuiper Rabbit Ranch
5317 W 41st St.
Gary, IN 46468

Marshall Farms USA, Inc.
5800 Lake Bluff Rd.
North Rose, NY 14516

Taconic Farms
273 Hover Ave.
Germantown, NY 12526
www.taconic.com

A complete listing of licensed animal dealers in the United States can be obtained from the Lab Animal website (<http://guide.labanimal.com>.)

Governmental Guidelines

The Animal Welfare Act enacted in 1966 (Public Law 89-544) and amended in 1970 (Public Law 91-579) and 1976 (Public Law 94-279) contains provisions to ensure that animals intended for use in research receive humane care and treatment. Implementing rules and regulations are published in the Code of Federal Regulations (CFR), Title 9 (Animals and Animal Products), Subchapter A-Animal Welfare, Parts 1, 2, and 3.

The U.S. Department of Health and Human Services publishes the *Guide for the Care and Use of Laboratory Animals*, Publication No. (NIH) 78-23. This was prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources of the National Research Council. The “guide” contains recommendations regarding housing, sanitation, husbandry, veterinary care, personnel qualifications, personal hygiene, occupational health, and physical plant.

The National Institutes of Health (NIH) requires that grantees and contractors using live, vertebrate animals in projects supported by NIH follow the guidelines prescribed in the guide. The Public Health Service further requires that grant-seeking institutions either be accredited by the American Association for Accreditation of Laboratory Animal Care (see later) or have an institutional committee that reviews its animal facilities and practices for compliance with the guide.

American Association for Accreditation of Laboratory Animal Care (AAALAC)

AAALAC is a nonprofit corporation directed by representatives of 24 scientific and professional organizations that are members of the corporation. It was organized in 1965 to conduct a voluntary program for the accreditation of laboratory animal care facilities and programs. AAALAC encourages optimal care for laboratory animals by providing a mechanism for peer evaluation of animal care programs by the scientific community. Humane treatment of laboratory animals, protection of personnel from hazards associated with the use of animals, and control of variables that could affect animal research adversely are among the principal objectives of the accreditation program. Animal care facilities of applicant institutions are visited and a thoroughly evaluated by two experts in laboratory animal science, who submit a detailed report to the Council on Accreditation. Following the standard listed in the *Guide for the Care and Use of Laboratory*

Animals, the Council determines whether AAALAC accreditation should be granted. Accredited facilities submit annual reports on the status of their animal facilities to AAALAC, and site visits to accredited facilities are conducted at least every 3 years. These annual reports and site visits determine whether accreditation will be continued. Full accreditation by AAALAC is accepted by the NIH as assurance that the animal facilities are evaluated in accordance with their policy on laboratory animals.

REFERENCES

1. Cohen, B.J.; Loew, F.M. Laboratory Animal Medicine: Historical Perspectives. *Laboratory Animal Medicine*; Fox, J.G., Cohen, B.L., Loew, F.M., Eds.; Academic Press: Orlando, 1984; 1–17.
2. Lane-Petter, W.; Pearson, A.E.G. *The Laboratory Animal: Principles and Practice*; Academic Press: London, 1971; 1.
3. Principles and Methods for Evaluating the Toxicity of Chemicals, Part 1. Environmental Health Criteria 6, World Health Organization, Geneva, 1978; 35–36.
4. Turner, R.A. *Screening Methods in Pharmacology*; Academic Press: New York, 1965.
5. Turner, R.A.; Hebborn, R. *Screening Methods in Pharmacology*; Academic Press: New York, 1971; II.
6. Nodine, J.H.; Siegler, P.E. *Animal and Clinical Pharmacologic Techniques in Drug Evaluation*; Year Book Medical Publishers: Chicago, 1964.
7. Tedeschi, R.E.; Tedeschi, D.H.; Kitchens, J.T.; Cook, L.; Mattis, P.A.; Fellows, E.J. *J. Pharmacol Exp. Ther.* **1959**, 126, 338–344.
8. Zak, O.; Sande, M.A. *Experimental Models in Antimicrobial Chemotherapy*; Academic Press: New York, 1986; 1–3.
9. Godlin, A.; Venditti, J.M.; Carter, S.K. Screening at the National Cancer Institute. *Methods of Development of New Anticancer Drugs*; Saunders, J.F., Carter, S.K., Gregoric, F.I., Eds.; NCI Monograph No. 45, DHFW Publication No. (NIH): 1977; 76-1037, 37–48.
10. Sontag, J.M.; Page, N.P.; Saffiotti, U. *Guidelines for Carcinogen Bioassay in Small Rodents*, National Cancer Institute Carcinogenesis Technical Report Series No. 1, DHEW Publication No. (NIH) 76–801, 1976.
11. International Conference on Harmonisation; Guidance on Testing for Carcinogenicity of Pharmaceuticals. Federal Register **1998**, 63 (35), 8983–8986.
12. Oilman, M.R. Skin and Eye Testing in Animals. *Principles and Methods of Toxicology*; Hayes, A.W., Ed.; Raven Press: New York, 1982; 209–222.
13. Interagency Regulatory Liaison Group, Testing Standards and Guidelines Work Group. *Recommended Guideline for Acute Eye Irritation Testing*; 1981.
14. Broder, I. Anaphylaxis. *Inflammation, Immunity and Hypersensitivity*; Movat, H.Z., Ed.; Harper & Row: New York, 1979; 320–335.
15. Ovary, Z. Passive Cutaneous Anaphylaxis. *Immunological Methods*; Ackroyd, J.R.A., Ed.; F A. Davis: Philadelphia, 1964; 259–283.
16. International Conference on Harmonisation Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals. Federal Register **1997**, 62 (227), 62922–62925.
17. American Association for Accreditation of Laboratory Animal Care (AAALAC), 11300 Rockville Pike, Suite 1211, Rockville, MD 20852. website: <http://www.aaalac.org>.
18. Brusick, D. *Principles of Genetic Toxicology*; Plenum Press: New York, 1980.
19. Gorrod, J.W. *Testing for Toxicity*; Taylor & Francis: London, 1981.
20. Harmison, L.T. *Research Animals in Medicine*; DHEW Publication No. (NIH): 1973, 72–333.
21. Hayes, A.W. *Principles and Methods of Toxicology*; 3rd Ed.; Raven Press: New York, 1994.
22. Heddle, J.A.; Dean, S.; Hohmi, T.; Boerrigter, M.; Casciano, D.; Douglas, G.R.; Glickman, B.W.; Gorelick, N.J.; Mirsalis, J.C.; Martus, H.J.; Skopek, T.R.; Thybaud, V.; Tindall, K.R.; Yajima, N. In Vivo Transgenic Mutation Assays. *Environ. Mol. Mutagen.* 2000; 35, 253–259.
23. Kawamata, J.; Melby, B.C., Jr. *Animal Models: Assessing the Scope of their Use in Biomedical Research*; Alan R. Liss: New York, 1987.
24. Lane-Petter, W.; Pearson, A.E.G. *The Laboratory Animal: Principles and Practices*; Academic Press: New York, 1971.
25. Marzulli, F.N.; Maibach, H.I. *Dermatotoxicology*, 5th Ed.; Taylor & Francis: Washington, DC, 1996.
26. Mirsalis, J.C.; Monforte, J.A.; Winegar, R.A. Transgenic Animal Models for Measuring Mutations In Vivo. *Crit. Rev. Toxicol.* **1994**, 24 (3), 255–280.
27. Provost, G.S.; Kretz, P.L.; Hamner, R.T.; Matthews, C.D.; Rogers, B.J.; Lundberg, K.S.; Dyaico, M.J.; Short, J.M. Transgenic Systems for In Vivo Mutation Analysis. *Mut. Res.* **1993**, 288, 133–149.
28. Yoshida, H.; Hagehara, Y.; Ebashi, S. *Advances in Pharmacology and Therapeutics*; II Pergamon Press: New York, 1982; 5.

ANALYSIS OF BIOLOGICAL FLUIDS

Stephen G. Schulman

Judith A. Davis

Gayle A. Brazeau

University of Florida, Gainesville, Florida, U.S.A.

INTRODUCTION

Biomedical drug analysis has been important for many years and has become the cornerstone for the development and formulation of new chemical entities (1). However, escalating interest, most recently as a component of patient care, has launched drug assay to the forefront of applied analytical chemistry (2–6). The topic covered herein is the accurate assessment of the concentration of a known drug, or of a particular metabolite of a known drug, in biological material such as blood or urine or tissue fluid. These techniques have also become critical in the new areas of forensic and clinical toxicology (7). The strategy is inextricably linked to knowledge of the metabolism and pharmacokinetics of the drug in question (8–13), because knowledge of metabolism governs the techniques used to ensure specificity and selectivity. The pharmacokinetic properties of the governed drug, in turn, determine the detection limits that are required.

Drugs and their metabolites are found in complex biologic matrices such as blood, urine, saliva, cerebrospinal fluid (CSF) and solid tissues. In many cases, the concentrations to be measured are in the microgram to nanogram or picogram levels. These matrices normally contain large amounts of endogenous compounds. Such compounds can interfere in the chemical and physical analytical methods used to detect and determine the materials of pharmacological interest. Consequently, unless some ultraspecific method of analysis is available for the substance of interest, physical separation of that substance from other interfering substances is usually necessary before quantitative determination can be achieved. To this end, most drug analyses must involve a separation step and a detection step. The separation step removes the drug from the biologic matrix. This separation can employ columns, solid phase extraction, solvent extraction, or a simple deproteinization of plasma with a liquid chromatography solvent such as acetonitrile. In more recent years, supercritical fluid extraction has been found to be useful for the extraction of low to moderately polar compounds. The requirements in any particular case

depend on the characteristics of the matrix, the drug, and the drug metabolites. The concentration range of the drug is also obviously important, as it will determine the methods selected for the separation and analytical procedures.

CHEMICAL BASIS FOR DRUG AND METABOLITE EXTRACTION

Many drugs are more lipophilic than the constituents of the matrices in which the compound needs to be separated from in the analyte of interest. This is the basis of Brodie's principle, which uses the least polar solvent that adequately extracts the drug from the biologic matrix (2). However, in contrast, most drug metabolites are more polar than their precursors, so that simple differential solvent extraction systems can be used to separate drugs and their metabolites. For example, nitroglycerin can be quantitatively removed from plasma by extraction into heptane, hexane, or pentane. However, its metabolites, glycerol-dinitrate and glycerol-mononitrate, are not extractable into these solvents.

Sometimes drugs and their metabolites are extractable into the same solvents. For example, chlorpromazine, demonomethylchlorpromazine, dedimethylchlorpromazine, and chlorpromazine sulfoxide are extractable into heptane (11). However, adjustment of the pH of the aqueous medium to various values between 4 and 11 permits selective extraction of one or another of the compounds. An example of the application of this principle is that chlorpromazine analysis utilizes a pH 4.6 backwash, which leaves most of the chlorpromazine in the heptane and all of the metabolites in the aqueous medium and permits a specific chlorpromazine assay (Fig. 1).

An important step forward in drug detection analysis utilizes techniques as the Bratton–Marshall approach in the sulfadimidine assay. This drug is converted to an azo dye in the Bratton–Marshall assay. However, the acetyl metabolite, if it is to be assayed, requires hydrolysis back to sulfadimidine and application of a different assay

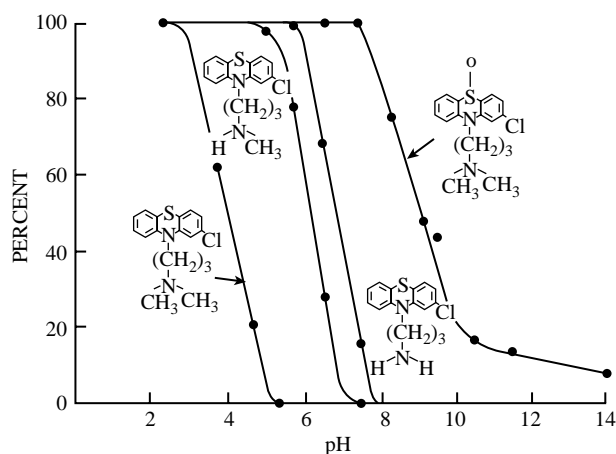


Fig. 1 Dependence of extraction of chlorpromazine and three of its model metabolites into heptane containing 1.5% isoamyl alcohol on pH. The graph shows the percentage remaining in the aqueous solutions when equilibrated with equal volumes of the solvent mixture. (From Ref. 11.)

calculation. Unfortunately, the necessary hydrolysis conditions cause some decomposition of the unmetabolized drug, which leads to analytical errors (12). These errors, however, can be overcome with liquid chromatography approaches.

While analytical derivatizations are an effective way for extracting compounds, these often require additional steps in the analytical procedure and can introduce side products that may interfere with the analysis. Solid phase extraction has provided an alternative method to this process. The advantage of solid phase extraction is that the reagents, derivatives, and side products are maintained on the solid phase. As needed, these derivatives and side products can be selectively eluted after the desired derivative has been formed on the column. In addition, this method can eliminate potential problems associated with emulsion formulation that may occur with liquid-liquid extraction of compounds from the biological matrix. Finally, solid phase extraction is easily amenable to automation with other analytical detection methods such as gas and liquid chromatography. The phases used in solid phase extraction are the standard ones employed in other extraction methods (14).

Supercritical fluid extraction is an alternative method for the extraction of low to moderately polar compounds. This method involves the use of super- and subcritical carbon dioxide for the extraction of compounds from various matrices. An advantage of this method is that it reduces the number of hazardous or expensive chemical solvents during the extraction procedures. In recent studies (15), supercritical fluids have been suggested to be a

powerful alternative to other extraction methods. The use of supercritical fluid extraction for isolating drugs from various dosage forms and biological matrixes has been reported for fluconazole, benzodiazepines, acyclovir, and morphine. The use of superfluid extraction can result in reduced sample handling, protection from conditions that can cause degradation, such as light, heat or oxygen, high loadability of samples, ability to conduct trace analysis, and a greater assurance against undesired chemical reactions during the extraction procedure. While carbon dioxide is the most commonly used agent for supercritical fluid extraction, its nonpolar characteristics may limit its usefulness. To offset this limitation, polar organic solvents have been added to aid in the extraction efficiencies. There are several considerations that must be taken into consideration when using supercritical fluid extraction methods. While the solubility of the analyte in the supercritical fluid is important, it does not always ensure extractability. It is important to consider the location of the analyte in the matrix, as the interaction of the analyte with the active sites of the matrix must be disrupted for extractability. As such, this might require a longer extraction time, higher extraction pressure, higher extraction temperature, and a higher concentration of a modifier. Recovery is best assessed in terms of the recovery from native samples rather than from spiked samples.

During the last 25 years, gas chromatography and, even more so, liquid chromatography and liquid chromatography-mass spectrometry have become selective detection steps of choice, such that nonspecific extractions are satisfactory, with specificity introduced by the chromatographic column and quantitative detection achieved with the chromatography detector. There is now a vast literature on chromatographic drug assays, with a seemingly endless supply of separation data. Many separations involve derivatization steps following conventional chromatography work-up principles.

In the last 10 years, developments in biomedical drug analysis have mostly involved gas or high-pressure liquid chromatography, the latter using ultraviolet (UV) or visible light absorption or emission, or electrochemical oxidation or reduction as a means of detection or immunoanalysis. The introduction of liquid chromatography-mass spectrometry is quickly becoming the gold standard method due to the fact that these methods are specific, sensitive, precise, and fast. These approaches probably account for more than 90% of methods currently in use. Both chromatographic and immunoassay approaches have their advantages and disadvantages. For example, chromatography prior to detection can be time-consuming. Immunoanalysis that relies on high specificity of mammalian antibodies for specific molecules

for which they are tailored often obviates the need for separation steps.

CHROMATOGRAPHY

Chromatography is characterized by a mobile phase that moves through an open bed or a column with a stationary phase. The components of mixtures brought into the chromatographic system are separated because of their different relative affinities for the mobile and stationary phases. Solutes with a low affinity migrate with greater velocity through the system (column or bed) than components with a high affinity for the stationary phase. The migration rates depend on the structures of the solute molecules, the compositions of the stationary and mobile phases, and environmental factors, such as the temperature.

Chromatographic systems can be classified according to whether the mobile phase is a gas or a liquid, hence the distinction between gas chromatography (GC) and liquid chromatography (LC). The stationary phase can be a liquid or a solid. Frequently applied systems are gas–liquid chromatography (GLC), liquid–liquid chromatography (LLC), and liquid–solid chromatography (LSC). In paper chromatography (PC), paper is used as a support for the (liquid) stationary phase. In thin-layer chromatography (TLC), an adsorbent layer, such as silica gel, is fixed to a suitable, inert plate, such as glass; the adsorbent can function either as the stationary phase or as a support for a liquid stationary phase. In column chromatography (CC) the system is closed—the stationary phase is contained in a glass or metal column through which the mobile phase moves.

The great success of chromatography is due largely to the development of chromatographic systems with a very high degree of separation power and extremely sensitive detectors.

Gas–Liquid Chromatography (GLC)

GLC is a separation process used for analysis of volatile substances and some nonvolatile substances that can be made volatile by chemical derivatization. Historically, it was the first chromatographic method applied to trace analysis in biologic matrices. The sample to be analyzed is volatilized by flash evaporation and moved, in a relatively inert carrier gas, along a column that contains a liquid stationary phase coated onto an inert, solid support. Due to the variety of differential affinities, the various analytes in the sample have for the mobile and stationary phases, separation occurs and some components are eluted from

the column before others. Each analyte is detected as it emerges from the column by a detector that operates as a transducer.

Only in relatively few instances can a biologic sample be injected directly into a gas chromatograph. Ordinarily, some preliminary purification must be accomplished. Depending on the compound to be analyzed, the purification procedure may be simple or rigorous. Analysis of steroids requires multiple extractions with solvents, but some methods for the determination of blood alcohol allow direct injection of whole blood into the gas chromatograph. The sample preparation depends on the number of compounds in the unknown, their concentrations, the presence of interfering substances, and the column material used.

After purification, this sample may require further manipulation before injection into the gas chromatograph. Some substances in a family of compounds are so similar that they cannot be well separated from one another unless they are first converted into derivatives such as silyl ethers. By converting the compounds into ethers, the molecules are made larger and are more readily separated from one another.

A suitable solvent must be used for sample injection. The solvent most frequently used in GC are acetone, alcohol, chloroform, hexane, and other volatile organic solvents. Aqueous solutions are very rarely used in gas chromatographic analysis.

The sample (about 1–15 μ) is injected from a microsyringe into the gas chromatograph through an injection port sealed with a rubber septum. The injection port is surrounded by a metal block that is heated by an independent heating unit to a temperature considerably higher than that of the column. The temperature of the injection block is determined by the boiling point of the least volatile compound in the mixture, usually about 50–100°C above the column temperature. The evaporated sample is moved through the column by an inert carrier gas, such as argon or nitrogen. The speed with which the main mass of the injected sample moves along the column depends upon the pressure of the gas flow and the temperature of the column. The column temperature is maintained by an enclosure in an oven that operates independently of the heating element for the injection port and the oven used to heat the detector chamber.

The time required for an individual component to emerge from the column is referred to as that compound's retention time under the conditions of temperature and pressure stated. Each compound has its own characteristic retention time, forming the basis for gas chromatographic separation. The separation of the various components is governed by the partition coefficient between the gas

phase in which the sample travels and the liquid phase, which is the column coating.

Columns may be constructed of packed metal or glass, or of empty or coated glass or nylon capillaries. Metal and glass columns may be straight, coiled, or U-shaped, range in length from 1 to 6 m long, and have a bore of about 0.5 cm. Capillary columns are usually of very narrow bore (0.1 cm) and about 25 m in length. The longer the column, the more efficient, in general, is the separation.

Column packing consists of two essential ingredients—the inert supporting phase and the stationary liquid phase. The solid supporting phase is usually an inert material of uniform particle size. Diatomaceous earth is a frequently used solid support, although celite, firebrick, and glass beads are also used. The particle size is important in achieving the maximum separation or efficiency. However, smaller particles inhibit the flow rate of gas due to the increased resistance of the denser medium.

The liquid phase of the column packing is actually responsible for separating the various components of the mixture. The principal characteristics of the compounds to be considered are polarity and volatility. The liquid column coating should be nonvolatile and thermostable. Its boiling point should be approximately 250–300°C higher than the optimum temperature at which the analysis will be conducted, and it should have low viscosity. High-viscosity liquids decrease column efficiency. The melting point should ensure that the coating becomes liquid at the optimum temperature of analysis.

Nonpolar column coatings include silicone oil, hydrocarbons, and esters of high molecular weight alcohols, and dibasic acids. Polyethylene glycols, polyesters, ethers, carbohydrate esters, and derivatives of ethylenediamine are widely used polar liquid phases.

Column-packing material may be purchased already prepared. In fact, columns can be purchased already packed with the desired inert and liquid phases ready for insertion into the chromatograph.

As each component in the mixture is eluted from the column, it is detected by one of several detecting devices. The signal recognized by the detector is converted into electrical energy in the electrometer. This small signal is amplified and made to drive a pen on the recorder, where it is registered as a peak.

The recorder tracing gives two different kinds of information: 1) identification of a compound by its retention time (time it takes for the peak to appear after injection of the sample); and 2) quantitation of the compound by comparison of the area under the unknown peak with the area under a standard peak corresponding to a sample of the same analyte of known concentration.

The function of the detector is to identify and quantitate the various components that have been separated by the column and that are carried through the detector by the carrier gas. Although many different types of detectors are used in GC, only a few are used in biomedical applications. All of these produce a very small electrical current that varies in proportion to the quantity of the compound in the effluent carrier gas. This small electrical current is then amplified by the electrometer to a level sufficient to drive the recorder pen. Because very small amounts of solute (in the range of 10^{-10} mol) are present in biologic analysis, the detectors used must be extremely sensitive and stable.

The thermal conductivity detector responds to all types of compounds and has a sensitivity of approximately 10^{-8} mol solute. However, it is sensitive to temperature changes as well as to changes in the flow rate of the carrier gas. In this detector, a thin filament of wire is placed at the end of the column and is heated by passing an electrical current through it. When the effluent gas passes over the wire, a temperature change occurs, which changes the resistance in the wire. When the carrier gas contains various components of the sample, a greater temperature change is caused than with the carrier gas alone, and a greater change in resistance in the wire and a greater change in the flow of current through the wire are produced. In actual practice, two detectors are used. One detector is the reference cell through which only carrier gas passes, and the other is the cell through which the carrier gas plus the vaporized sample components pass. These two detectors are placed in a balanced electrical circuit that is adjusted by various resistors with only carrier gas flowing through both detectors so that no flow of electricity occurs. When solute is in the carrier gas passing through one detector, there is a change in resistance and current flows through the electrometer. The amplified signal is then recorded as a peak on the chart.

Organic compounds ionize when burned in a hydrogen air flame. If two electrodes at a potential difference of approximately 150 V are inserted into this flame, differences in conductivity of the flame can be measured as the solutes elute from the column and are burned. This is the principle on which the flame ionization detector is based. In the usual flame detector, the column effluent is mixed with hydrogen. This mixture is fed into the flame jet of the detector. The jet is a thin-walled stainless steel tube that also acts as one electrode. The other electrode is a fine platinum wire held above the jet. The response of this detector is practically instantaneous. It is not affected as much as the thermal conductivity detector is by changes in temperature and carrier gas flow rate. It is very sensitive and can detect approximately 10^{-10} – 10^{-15} mol solute.

In the argon ionization detector, argon is used as the carrier gas. Atoms of argon are ionized by β particles from a radioactive source as they enter the detector. The ionized argon atoms collide with solute organic molecules, whereupon secondary electrons are produced, giving rise to a current proportional to the concentration of solute in the carrier gas. The current is amplified and recorded in the usual way. The secondary electrons also generate more argon ions to replace those deactivated in the collisions with solute molecules. This type of detector is relatively unaffected by variations in temperature and carrier gas flow rate. Its sensitivity is approximately the same as that of the hydrogen flame detector (i.e., 10^{-10} – 10^{-15} mol solute).

The electron capture detector has a source of low-energy ionizing radiation. The electrons that strike the solute molecules have just enough energy to penetrate the electrical field of the molecule and be captured, but not enough energy to break the molecule into ions. The original electrical signal is decreased by electron capture. Compounds that contain halogen atoms and certain polar functional groups are most easily detected by electron capture. If the compounds to be analyzed do not contain halogen atoms, they must be halogenated prior to chromatography. The detector is more sensitive to temperature fluctuations than the argon ionization detector, but is relatively insensitive to changes in carrier gas flow rate. This detector is extremely sensitive and can detect 10^{-15} – 10^{-20} mol solute.

Occasionally, as in the case of extremely complex biologic samples, GC alone does not provide sufficient specificity or sensitivity for desired analysis. In these cases, using the ultimate detector, the mass spectrometer (MS) may be advantageous. In GC–MS, the effluent gas from the chromatograph is fed directly into the vacuum chamber of the mass spectrometer, where the separated solutes it contains are ionized either by impact with a beam of electrons that have about 70 eV energy or by collision with ionized atoms or molecules of a reactant gas (chemical ionization), which itself is ionized as a result of electron impact. The ionized analyte molecules are accelerated by a strong magnetic or electric field and after traveling some distance over a curved path, are allowed to fall on a photographic plate or are collected by an ionization detector. The locations of darkened areas (analyte ion impact regions) on a photographic plate depend upon the ratio of ionic mass to charge (m/z) and, with some knowledge of the natural abundances of isotopes of the atoms that comprise the analyte molecules, can be used to identify and quantitate the analytes. In some cases, an ionization detector equipped with a time discriminator can be used to distinguish between the arrival times of the ions of various mass-to-charge ratios (time-of-flight MS) and thereby identify the various

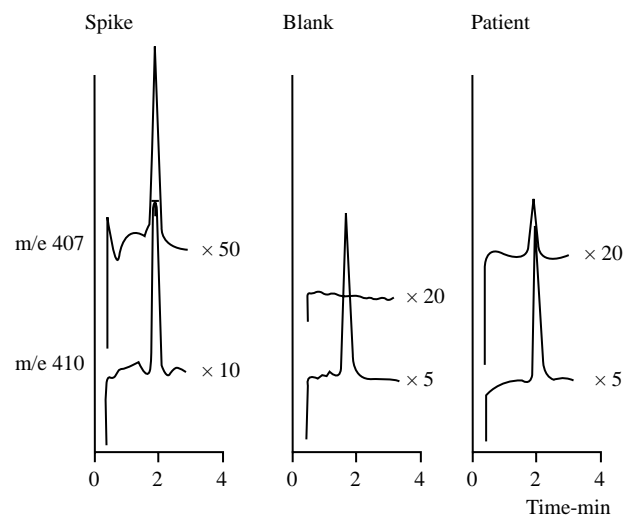


Fig. 2 GC–MS analysis of trifluoperazine. The diagram shows traces at mass to charge ratios (m/e) of 407 and 410, at various attenuation settings ($\times 5$, $\times 10$, $\times 20$, and $\times 50$) for blank plasma, blank plasma with trifluoperazine added, and plasma of a treated patient. The method detects “endogenous” trifluoperazine (m/e 407) and deuterated trifluoperazine internal standard. (Reproduced from the *Journal of Pharmacy and Pharmacology*).

species. Current GC–MS occupies a position of great importance in pharmacological research laboratories as it makes possible the identification and structure determination of metabolites that are present in quantities too small to be detected by other means.

The application of the combined GC–MS approach is illustrated by trifluoperazine data (Fig. 2) (11). Plasma of treated patients contains trifluoperazine itself. In the assay method, a trace (less than 1% of total) of radio-labeled trifluoperazine is added to the sample. Also added is an excess of a deuterated version of trifluoperazine. The deuterated version acts as both “cold carrier” (ensuring adequate extraction of trifluoperazine) and as the mass spectrometer internal standard. The radio-labeled trifluoperazine is used to evaluate extraction. The mass spectrometer monitors at one or more of the m/z values for both “cold” original trifluoperazine and deuterated trifluoperazine (generally 407 and 410, the parent ion m/z values). Quantitation is by standard MS ratio techniques.

HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC)

Notwithstanding the great analytical successes brought about by GLC, the method does have serious limitations.

In particular, the necessity of application to volatilized, thermally stable compounds is very restrictive, as a majority of pharmaceuticals and their metabolites are charged species at intermediate pH (aliphatic amines encompass more drugs than any other chemical class). Large molecules, such as peptides, are unstable in the injection port of the gas chromatograph and are also difficult to analyze by GLC. However, liquid-liquid partition chromatography is amenable to the analysis of ionic and large molecules as heating is not necessary. LLC, moreover, has the advantage over GLC in that the chemical natures of both the mobile and stationary liquid phases affect separability and thus analytical selectivity, whereas in GLC, only the stationary phase influences separation chemically. In LLC, either the more or less polar liquid phase may be used as the mobile phase. The latter is called *normal-phase chromatography* and the former *reverse-phase chromatography*.

The principal reason why LLC did not take precedence over GLC in the early days of analytical chromatography was the time-consuming aspect of the former due to slower mass transfer in the liquid phase. Because transfer of solute molecules between mobile and stationary phase occurs by diffusion, the much higher viscosity of liquid relative to that of a gaseous mobile phase results in diffusion rates about 10^5 times slower in the liquid; that is, partition equilibrium in the column is established very slowly. Additionally, because separation efficiency increases with increasing column length and the higher viscosity of a liquid results in slower longitudinal movement than in the case of a gas, increasing the efficiency of the separation by using a longer column increases the time of analysis.

The only way to increase the rate of diffusion of solutes is to raise the temperature substantially, but this would lead to problems in thermally unstable analytes. The alternative is to reduce the distance through which the molecules diffuse. Efficient separation then requires the use of smaller particles for column packings. This apparently simple expedient has evolved into a new practice of LC that is competitive with GLC in speed and resolution of complex mixtures and applicable to many more materials than GLC (16–19). Packing materials comprised of particles as small as $5\text{ }\mu\text{m}$ are currently available. Smaller particles are extremely difficult to handle and give an almost impermeable column. To solve this problem, solid glass beads of $30\text{--}50\text{ }\mu\text{m}$ in diameter can be coated with a layer of porous material. These are called *pellicular beads*. The porous layer may serve as a solid stationary phase or be coated with a very thin layer of liquid stationary phase with an extremely large surface area.

The thin film of stationary phase, if untreated, may be rinsed away by mobile phase under the high pressures used. One method of dealing with this is chemically bonding the liquid phase to the solid support. Porous silica beads are esterified with various alcohols that form the corresponding silicate esters. These esterified, siliceous packings are not thermally stable and are subject to hydrolysis and exchange with lower alcohols. Another type of chemically bonded packing utilizes silicone polymers, which are more stable because of their 3D cross-linked structure. For charged analytes, ion-exchange resins or celluloses are often used as packing materials.

Columns packed with small particles require high inlet pressures in order to give a reasonable flow rate. Pressures of $30\text{--}300\text{ atm}$ are commonly employed in the small columns used ($2\text{--}3\text{ mm}$ in diameter) and flow rates are typically $0.5\text{--}3\text{ ml/min}$.

HPLC apparatus consists of a mobile-phase reservoir, a sample injection system, a column, a detector, and a recorder. The operation of most of these components is self-evident, and a schematic diagram of an HPLC apparatus is shown in Fig. 3.

The smaller columns and faster flow rates place rigid requirements on the detection system. Flow-through detectors with low dead volumes and high sensitivity are necessary. The dead volume within the detector should be no more than a tenth of the analyte peak volume. Because highly efficient LC columns can give peak volumes of the order of the $50\text{ }\mu\text{l}$, a detector volume of about $5\text{ }\mu\text{l}$ is desirable.

For biomedical trace analysis, three types of detectors are currently popular—the absorption photometric detector, the fluorescence detector, and the electrochemical detector. Although there are other kinds of detectors, only these have the ability to detect $10^{-9}\text{--}10^{-12}\text{ g}$ of analyte, the kind of detectability needed in biomedical analysis, especially where small amounts of drugs are concerned.

The most widely used detectors, the absorption photometric detectors, are those based on the absorbance of UV light. They are not universal in application (19), but

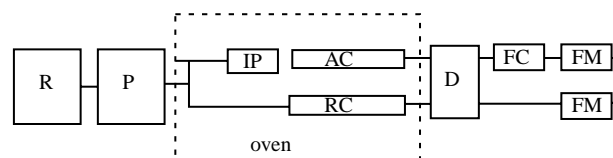


Fig. 3 Block diagram of a high-performance liquid chromatograph; R is the reservoir for the mobile phase, P the pump, IP the sample injection part, AC the analytical column, RC the reference column, D the detector, FC a fraction collector (for preparative work), and FM mobile-phase, flow-rate meters.

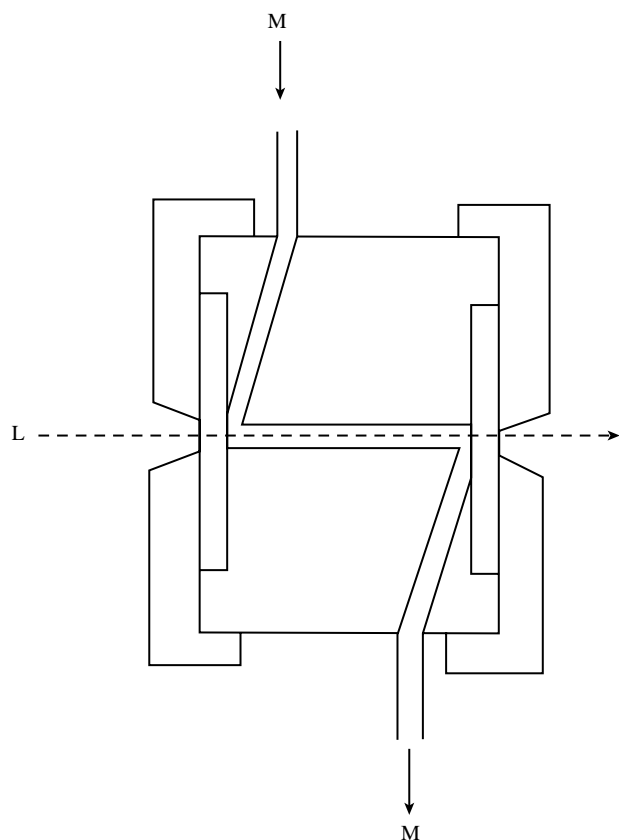


Fig. 4 Schematic diagram of the Z cell: the Z-cell-M is the path of the mobile phase, and L is the optical path through the quartz window and the sample in the mobile phase.

a great many substances do absorb UV radiation, including all substances having π -bonding electrons and also those with unshared (nonbonded) electrons, such as olefins, aromatics, and compounds containing C=O, C-S, -N=O, and N=N. In UV detectors, the absorbance A is directly proportional to sample concentration and obeys Beer's Law: $A = \mu Cl$ where μ is the molar absorptivity, l is the cell path length, and C is the sample concentration. UV detectors are essentially nondestructive of the sample. The short residence time of the sample in the detector flow cells, a few seconds or less, minimizes damage to compounds sensitive to UV light. Spectrophotometric detector cells of two different types have been most commonly used in commercial HPLC systems (16). In the Z-configuration type (Fig. 4), the mobile phase moves halfway along one window of the cell, travels the distance between the two windows in line with the light beam, and goes out along the other window. A split-stream or H-flow pattern (Fig. 5) is utilized in the second type of flow cell to minimize noise and drift due to possible flow

fluctuations. The mobile phase flows into the center of the optical path, is split, and goes in opposite directions. At both ends of this optical path, the mobile phase sweeps the cell windows and is recombined in an upper bore. This cell has an optical path with dimensions of about 1 mm in diameter and 10 mm in length, with an internal volume of about 10 μ l.

Two types of photometers are used in HPLC—the fixed-wavelength filter photometer and the variable-wavelength spectrophotometer. These use a low-pressure mercury arc lamp as a light source, a transmission or interference filter to isolate a narrow band of wavelengths of light with which to excite the sample, and a photodiode to detect the light transmitted through the sample.

Most commercial fixed-wavelength UV detectors take advantage of the intense line source of 254-nm radiation in the low-pressure mercury arc lamp. The high intensity of

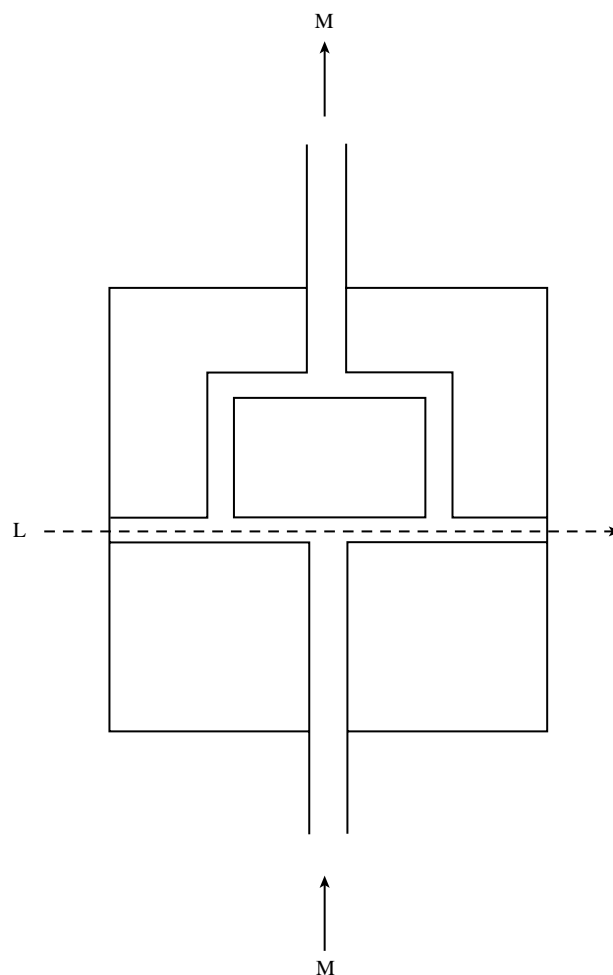


Fig. 5 Schematic diagram of the H cell: M is the path of the mobile phase and L is the optical path.

the radiation provides excellent detectability for the small-aperture microvolume flow cells required in HPLC. Concentration of most of the radiation in a narrow-wavelength band places less demand on optical filters and enhances the linear range of the detector.

The 254 nm wavelength has a wide range of applications. Biologically important compounds, such as aromatic amino acids, proteins, enzymes, and nucleic acid constituents, absorb strongly at 254 nm. Due to its inherently high sensitivity and the broadness of most UV absorption bands, the 254 nm detector is useful for many compounds whose maximum absorption is not at 254 nm. Nevertheless, UV detectors that offer detection at either 254 or 280 nm or at both wavelengths are also available. The 280 nm detectors employ a phosphor screen to convert 254 nm mercury arc radiation to 280 nm and generally sacrifice some linear dynamic range due to the relatively large bandwidth of the phosphor emission.

In recent years, advances in pump and column technology greatly expanded the usefulness of HPLC. As chromatographers applied HPLC to increasingly complex separations, the need arose for an absorbance detector with selectable UV wavelengths to optimize sample response and discriminate against interferences and also with sub-254 nm detection capability to allow detection of compounds such as carbohydrates and fatty acids. This need led to the development of the more sophisticated variable-wavelength UV detector. These detectors use a continuum source, such as a deuterium lamp (190- to 400 nm output), and a monochromator to isolate the narrow-wavelength bands desired. Usually, one can isolate a 2- to 16 nm bandwidth centered about the nominal wavelength. The deuterium lamp-monochromator-photomultiplier tube (or diode array) system is more expensive than the mercury lamp-filter-photodiode system of a fixed-wavelength detector.

The emission of light by molecules in solution that have been excited by the absorption of UV or visible light is the basis of the fluorescence detector (20).

The fluorimeter is a very sensitive and selective detector that is applicable to many compounds, including certain metabolites, amino acids, vitamins, and many drugs excited by UV radiation. In addition, fluorescent derivatives of many nonfluorescent substances, such as steroids, can be prepared (19). HPLC fluorescence detectors are similar to normal fluorimeters. Most fluorescence detectors are filter instruments, although variable-wavelength grating fluorimeters have been developed specifically for HPLC. Filters generally pass light in a wider band than do the monochromators, and this frequently turns out to be an advantage because the detector does not need to be specifically tuned for each compound as it elutes.

Furthermore, the filter detector is much less expensive than the variable-wavelength grating detector.

At excitation wavelengths and concentration ranges where the simple absorbance fluorescence is linear with concentration, the fluorimetric detector is susceptible to the usual interferences that hinder fluorescence measurements, mainly background fluorescence and quenching.

In the operation of a fluorimeter, the light from a UV source is filtered and focused on the cell. Fluorescence is emitted by the sample in all directions so that the emitted light can be measured with the detector at right angles to or (rarely) in line with the path of exciting light (Fig. 6). The excitation wavelength is then blocked by a filter, and the intensity of the emitted energy is measured by a photocell. Almost all of the non-halogen-containing solvents that are used with the UV-absorbance detectors can be used with the fluorescence detector. Halogenated solvents, such as CH_2Cl_2 or CHCl_3 , should be used with care because they tend to "quench" or diminish fluorescence. The solvent has a strong effect on the intensity of fluorescence. For instance, quinoline is nonfluorescent in hexane but fluorescent in ethanol.

Flow cells that can be used for both fluorescence and absorption measurements have been designed. Quantitation

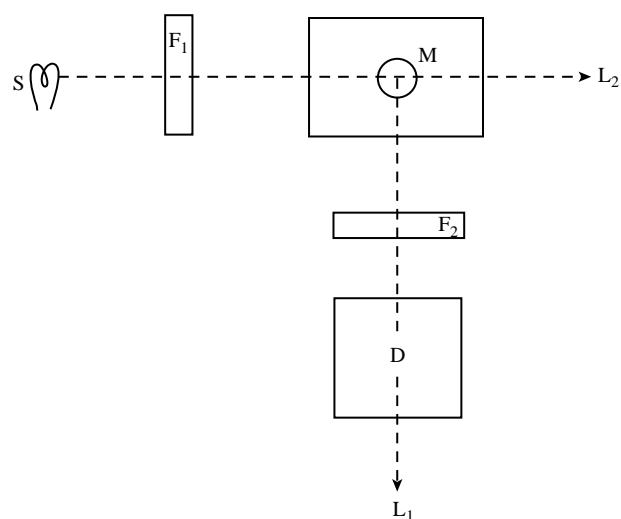


Fig. 6 Schematic diagram of a fluorescence detector: S is the source of exciting light. F_1 and F_2 are filters for the exciting and emitted light, respectively; M is the position of the mobile phase containing tube with respect to the excitation-emission arrangement. The path of the mobile phase is perpendicular to the plane of the page. D is the detector, and L_1 and L_2 , are the paths of the fluorescent light detected using the right angle and in-line geometries, respectively. When the in-line arrangement is employed, F_2 and D would be colinear with S, F_1 , and L_2 .

can be considerably improved by simultaneously monitoring the absorbance and fluorescence signals. Such concurrent measurement extends the linear dynamic range for the fluorescent samples. At high sample concentrations, where the absorbance at the excitation wavelength is greater than 0.01, fluorescence response becomes nonlinear. At these higher concentrations, however, the light absorbance of the sample is often measurable and linear with concentration. Fluorescence, however, is often detectable at concentrations up to 10^{-6} times lower than those where absorbance is detectable. Native fluorescence is less common than absorption of UV or visible light. In this regard, fluorescence may be thought to be more selective or less applicable in range than absorption, depending on one's point of view.

Electrochemical detectors are based upon the voltametric oxidation or reduction of separated analytes at a micro- or thin-film electrode. A number of pharmacologically active compounds that are aldehydes, ketones, or quinones (such as doxorubicin), or nitro compounds (such as nitrofurantoin) are amenable to reduction at a mercury or platinum electrode; electron-rich indole derivatives and catecholamines can be oxidized at these electrodes. An important condition that must be fulfilled for electrochemical detection to be practicable is that the mobile phase must be capable of conducting an electrical current. This makes electrochemical detection particularly useful in reversed-phase liquid chromatography, where buffered water mixed with one or more organic cosolvents is usually the mobile phase.

The selectivity inherent to electrochemical detection is derived from the differences between the oxidation or reduction half-wave potentials exhibited by different analytes. Even when two or more analytes have nearly the same half-wave potentials, complexing agents or alterations in mobile-phase composition can be used to differentiate between analytes. In order to carry out the electrochemical quantitation of an analyte, the potential difference between the working microelectrode and the reference electrode is maintained at a value that lies on the plateau of the oxidation or reduction wave (voltammogram) of the analyte of interest. The diffusion current thus measured, which is due to the oxidation or reduction of the analyte, is proportional to the area under the analyte peak eluted. In order for absolute quantitation to be effected, the diffusion current of a standard sample of the analyte must also be measured for comparison with that of the unknown sample.

Commercial electrochemical detectors whose cells are directly connectable to the postcolumn efflux of the HPLC apparatus are available from a number of manufacturers. This method of detection is comparable

in sensitivity to detection by absorption spectroscopy when the amperometric circuit is operated as a DC pulse polarograph.

Liquid Chromatography–Mass Spectrometry

Liquid chromatography–mass spectrometry (LC–MS) is quickly becoming a very common analytical method due to its selectivity and sensitivity, and the increased availability of bench top instruments at reasonable costs for research groups (7). Since the 1970s, there have been numerous ways to remove the mobile phase in order to ionize the analyte, including fast atom bombardment, as well as particle beam, thermospray, electrospray, and atmospheric-pressure chemical ionization. In recent years, the most common and universal technique appears to be LC–MS that employs electrospray and atmospheric-pressure ionization techniques. The advantage of electrospray is that it enables the determination of compounds with large masses (e.g., peptides) or those that are very polar (e.g., quaternary amines, phospholipids, etc.). The key is that these compounds must be ionizable in solution. As such, the mobile phase often includes a small quantity of a volatile acid or base. It may be necessary to add this volatile acid or base just prior to the electrospray detection process. Atmospheric-pressure ionization techniques, in contrast, are more suitable for compounds with lower molecular mass and a moderate polarity.

IMMUNOASSAY

Immunoassay methods have become increasingly important in the field of biomedical analysis (21, 22). Analyses of nanomolar and picomolar amounts of large biopolymers present in biologic matrices, unassayable by other techniques, have provided biochemists with much essential information. Additionally, diagnostic methodologies have evolved, with immunoassays becoming centrally important in the analysis of drugs, pesticides, hormones, and proteins. Their low cost and adaptability to automation, coupled with their sensitivity and specificity, have made them competitive with chromatographic methods in diagnostics.

Immunoanalytic methods are largely based upon the competitive binding that occurs between a labeled and unlabeled ligand for highly specific receptor sites on antibodies. The analysis of this competitive binding, effected by measuring some physical or chemical property associated with the label, allows the construction of a standard curve that represents a measured physical signal that is altered by changes in distribution of bound labeled

ligand as a function of the concentration of the unlabeled ligand. Unknown ligand (analyte) concentrations are extracted from this calibration curve.

Separation of the signal corresponding to either the bound or free-labeled analyte from that of the total labeled analyte population can be accomplished in two ways. The first involves physical separation of the protein (antibody) bound fraction from the free fraction of labeled analyte. This can be accomplished by a salting-out procedure, using a salt such as ammonium sulfate or a polymer such as polyethylene glycol, to precipitate the excess and analyte-complexed antibody, followed by centrifugation. Alternatively, solid phase techniques are possible, in which drug or antibody is attached to a solid surface (beads, tube wall, dip sticks, etc.) and competition between reactants in the liquid phase and the solid phase is followed by their physical separation. These techniques are known as *heterogeneous immunoassays* and are required in radioimmunoassay, where the labels employed are radioactive, usually either ^{125}I or ^3H , and there is no way to distinguish, in situ, between the radioactivities of the free and antibody-bound labeled material. Also, the majority of immunoassay methods in which the label is an enzyme, e.g., enzyme-linked immunosorbent assays (ELISA), are heterogeneous. Homogeneous immunoassays make up the second category, in which physical separation of bound from free labeled ligand is not required. Signals that are different for the bound and free-labeled ligands are obtained from the solution, which contains all the participating analytic species. The signal-producing species may be derived enzymatically when the label is an enzyme whose substrate turnover rate is reduced upon ligand-antibody association, or the label can be a fluorophore. In this case, environmental effects imposed on specific label populations can lead to signal modification.

Separation-free, homogeneous immunoassay protocols offer several advantages in comparison to heterogeneous methods. Because no separation is involved, the number of procedural steps is decreased, which decreases the time required per assay. Additionally, because the physical transfer step is avoided, potential sample loss related to this step is eliminated. Drugs with low molecular weights (amphetamines, digoxin) are commonly measured by separation-free homogeneous immunoassay protocols (23).

Most immunoassays currently employed in the biomedical field are either radioimmunoassays, enzyme immunoassays, or luminescence immunoassays (including fluorescence immunoassays [FIA] and chemiluminescence immunoassays). Although radioimmunoassay is currently the most sensitive of these (10^{-12} – 10^{-15} M concentrations are often detectable), due to the problems inherent to

dealing with radioactive materials, such as licensing, radiation hazard, short shelf-life of expensive radioisotopes, the expense of the counting equipment, and the tedium associated with heterogeneous immunoassay, it has fallen, in popularity, behind the nonisotopic methods of analysis.

In the enzyme-multiplied immunoassay techniques (EMIT), the antigen or antibody is labeled with an enzyme (e.g., isozyme, alkaline phosphatase, horseradish peroxidase, or glucose-6-phosphate dehydrogenase) instead of a radioisotope. For example, an alkaline phosphatase-labeled drug can be made to compete with an unlabeled drug for binding sites on a drug-directed antibody. When the enzyme-labeled drug is bound to the antibody, the enzyme loses its activity (ability is bound to the antibody, the enzyme loses its activity i.e., ability to hydrolyze phosphates). However, the free enzyme-labeled drug retains its enzymatic activity. If a potentially absorbing or fluorescing organic phosphate whose optical properties are altered by the condition of esterification by phosphate is put into the solution that contains enzyme-labeled drug, drug-directed antibody, and unlabeled drug, only that fraction of the labeled drug population not bound to the antibody is capable of generating the absorption of the fluorescence spectrum of the hydrolyzed phosphate. In turn, the amount of absorptiometrically or fluorimetrically measured, hydrolyzed phosphate depends upon the concentration of unlabeled drug added to the test solution, as it is this concentration that ultimately determines how much enzymic activity is released to the solution. Commercial EMIT kits based upon absorptiometric (colorimetric) estimation of enzymatically oxidized NADH to quantitate a variety of drugs have been popular for several years. The sensitivity is not especially high, with drug concentrations down to about $0.5\text{ }\mu\text{M}$ being detectable. The measurement of the light emitted by a fluorescent or chemiluminescent enzyme substrate is capable of extending the limits of detection of the analyte down to 10^{-9} – 10^{-12} M.

FIA involves the measurement of the fluorescence of a luminescent label (a fluorophore) that participates at some level of a competitive immunochemical binding system, and whose spectral properties in most homogeneous systems vary with differences in the concentrations of the analyte. Fluorescent labels are used in homogeneous and heterogeneous immunoassay systems; they may be bound to antigens, antibodies, or solid phases, or they may exist free in solution as enzyme substrates. A fluorescent label to be used in homogeneous immunoanalysis should fulfill several requirements. Since the fluorescent signal must be measured in a serum matrix, the probe should have a high fluorescence quantum yield, and the excitation and

emission maxima of the probe should occur at wavelengths longer than those of the serum. Excitation and emission spectra of dilute serum occur at 280 and 340 nm, respectively. For a fluorescent label to be useful in a serum solution, it must emit at 50 nm or more to longer wavelengths or at wavelengths greater than 400 nm.

To date, the most popular fluorescent labels for FIA have been those derived from the long-wavelength, strongly emitting xanthene dyes fluorescein isothiocyanate (FITC) and lissamine rhodamine B (RB200) (23). The isothiocyanates or isocyanates of these fluorophores can be used to label primary and secondary aliphatic amines in aqueous solutions by simple procedures. Consequently, they can be used to label antibiotics or alkylamine-substituted drug molecules. Even those drugs that do not have indigeneous alkylamino groups can often be labeled by introducing bridging groups (e.g., aminoethyl) that are amenable to coupling with the isothiocyanate or isocyanate functions.

Heterogeneous FIAs can be carried out with the aid of the same separation procedures used in radioimmunoassay. The more expedient homogeneous FIAs usually require quenching, enhancement, or shifting of the fluorescence of the label upon binding of the labeled drug to its antibody. Occasionally, a second antibody, directed at the antidrug antibody, is used in a "double-antibody" method to precipitate the bound labeled and unlabeled drug or to alter the optical properties of the label in such a way as to make the analysis more sensitive (24, 25).

Homogeneous FIAs often can be affected even when there is no obvious change in the intensity or spectral position of fluorescence of the label upon binding to the antibody. If light-polarizing polymer films are used to polarize the exciting light and analyze the fluorescence of the sample excited by polarized light, it will generally be observed that the amount of fluorescence that reaches the detector is considerably smaller in those samples where a greater amount of antibody binding is extant, all other things being equal. This is a result of the higher degree of polarized fluorescence emitted from the labels affixed to the slowly rotating antibody. The polarized emission is more efficiently attenuated by the analyzer-polarizing film than by the unpolarized light emitted by the rapidly rotating, labeled drug molecules that are not bound to macromolecules. This phenomenon forms the basis of fluorescence polarization immunoassay. The decrease in fluorescence measured with decreasing labeled drug binding occurs as a result of increasing unlabeled drug (analyte) concentration and can be used to construct a calibration curve from which the concentrations of unknown drug samples can be determined when their

polarized fluorescences are measured. Fluorescence polarization immunoassay probably accounts for most of the FIAs currently performed.

Chemiluminescence immunoassay (25, 26), a technique that has rapidly gained popularity because its sensitivity is comparable to that of radioimmunoassay, is in a sense a variation of FIA. In the 1930s, the first work on chemiluminophores was published, but it was not until the 1980s that chemiluminescence was first tried in immunoassays. Due to the increased use of automated immunoassay analyzers, chemiluminescence has become one of the most common immunoassay detection methods used in the clinical laboratory setting (27). Chemiluminescence (also called *bioluminescence* when it occurs in fireflies and some dinoflagellates, coelenterates, and fungi) is fluorescence. However, what is usually thought of as fluorescence is light emission caused by prior light absorption by the emitting molecule. This is properly termed *photofluorescence* or *photoluminescence*. Chemiluminescence occurs in the oxidation products of some highly strained, highly reduced molecules (e.g., peroxoxalate esters and amino-substituted cyclic hydrazides of phthalic acid). The oxidation of the latter results in initial products that possess such great quantities of thermal energy from the reaction that they spontaneously (without photoexcitation) become electronically excited and subsequently fluoresce as a means of achieving the state of lowest energy. The oxidations are frequently catalyzed by metal ions and occur at appreciable rates only when the precursors of the chemiluminescent species (the labels) are freely diffusible in solution (i.e., they will not generate light rapidly when bound to antibodies or perhaps prior to release in an enzymatic reaction). In this sense, they are analogous to many of the fluorescent labels. However, photofluorescence analysis entails the measurement of light emitted by a fraction of an even smaller fraction of molecules that absorb light for an instant. In chemiluminescence, it is possible to gather and integrate the light output associated with the chemiluminescent reaction over the entire course of the reaction. Consequently, very low detection limits can be attained if the luminescence efficiency of the chemiluminescent reaction is reasonably good (0.1). Unfortunately, very few chemiluminescent reactions have high luminescence efficiencies so that the choice of labels is much more restricted than in FIA. Older chemiluminophores (luminol, isoluminol, and dioxetanes) were insoluble in aqueous buffers, which complicated the labeling process. However, the new acridinium chemiluminophores incorporate a unique sulfopropyl substituent that conveys high aqueous solubility that enhances sensitivity and overall assay robustness (27). Chemiluminescence immunoassay is a

field whose popularity is on the rise, and this may inspire the discovery of new labels.

CAPILLARY ELECTROPHORESIS

Capillary electrophoresis is an exciting, new, high resolution separation technique useful for the determination of drugs and their metabolites in body fluids. The first commercial capillary electrophoresis instruments began to emerge on the market in 1988. Today approximately a dozen companies manufacture electrokinetic capillary instrumentation, with many of these fully automated, that comprise auto samplers with computerized data evaluation (28). Capillary electrophoresis involves the electrophoretic separations of minute quantities of molecules in solution according to their different velocities in an applied electrical field. The velocity of these molecules in solution is directly proportional to their charge; therefore, the more positively charged molecules are detected first (23, 29). Capillary electrophoresis is capable of separating and quantifying cations, anions, and uncharged molecules simultaneously after the injection of very small sample sizes (1 nL) that are about 1000 times smaller than typical HPLC samples (28, 30). The separation process takes place in a narrow-bore, electrolyte filled, fused silica capillary, typically 25–75 μm in diameter by 100 cm in length, with a detection window about 50 cm from the sample loading port. Samples are introduced into the capillary either by hydrostatic or electrokinetic injection, after which electro-osmotic flow acts as a pump that sucks the sample volume into the capillary. Buffers are used to maintain a constant pH in the capillaries electrophoretic solution, and may also be used to adjust the separation parameters. The capillary is filled with the buffer solution and each end of the capillary is immersed in vials that

contain the same buffer solution along with the systems driving electrodes (Fig. 7). Sample separation occurs when an electrical field (electrophoresis) is applied. The migration rate of the sample components depends on the pH of the buffer and the voltage applied. Capillary electrophoresis employs a high voltage DC (5–30 kV) electrical current to induce the electrophoretic transport and separation of the sample components. The combined action of electrophoresis and electro-osmosis transports the sample through the capillary tube to the systems detector (23, 28–31). Various UV absorbance, fluorescence, and mass spectroscopy detectors have been used in capillary electrophoresis systems. Since UV detection techniques require the presence of a relatively high drug concentration in the sample, they have not demonstrated the sensitivity to quantify drugs in biologic fluids (28, 30). Hempel reported two methods that have been tried to remedy these problems (30). Bubble cells that extend the UV light path in the capillary detection window have only slightly increased the detection sensitivity. Another technique, the Z-shaped capillary flow cell is composed of a bent capillary tube (Fig. 8) similar to the

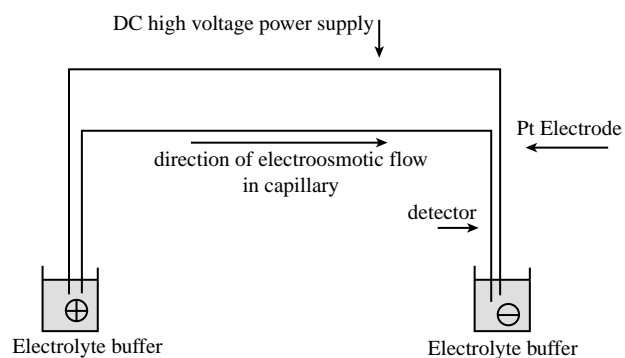


Fig. 7 Schematic diagram of a capillary electrophoresis system.

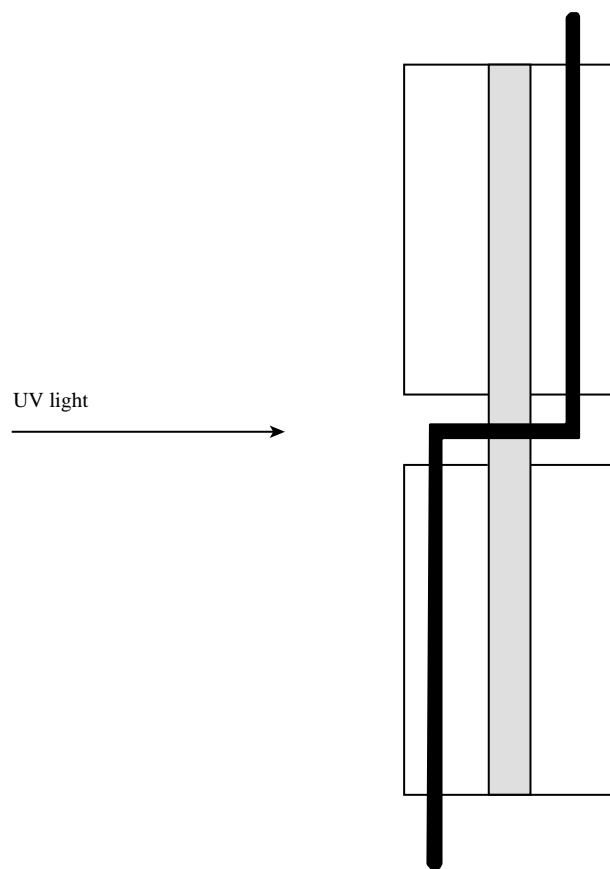


Fig. 8 Schematic diagram of a Z-shaped capillary flow cell.

spectrophotometric Z-configuration detector cell (Fig. 4). The Z-shaped capillary flow cell allows a few millimeters of the capillary to be parallel to the source of light. This appears to have resolved some of the earlier detector problems of scattered light. Typically, the length of the detector cell path in capillary electrophoresis is about 100 times shorter than that of HPLC (30).

Several different electrophoretic separation techniques have been used to separate an array of large and small molecules. The two most useful methods for the analysis of drugs and their metabolites in biologic fluids are micellar electrokinetic capillary chromatography (MECC) and capillary zone electrophoresis (CZE) (28, 31). MECC is an electrokinetic capillary chromatography technique that utilizes surfactants (such as sodium dodecyl sulfate) in the buffer solution at sufficient concentrations to form micelles. This technique is dependent on the formation of these micelles, which will then migrate to an electrode under the influence of an applied electrical field. There are two distinct phases in this separation process—an aqueous phase and a semistationary phase. The capillary exhibits electro-osmosis where a migration of the charged micelles and a movement of the entire liquid occur concurrently. The two phases migrate at different velocities, which allows chromatographic separation of the sample (28). Electro-osmotic flow and capillary wall chemistry are critical factors in many of these separations (32). MECC is a useful technique for the separation of both charged and uncharged drugs based on their ability to partition between the micelles and the surrounding buffer. In addition, it offers the possibility of directly injecting serum or other proteinaceous fluids (28, 31). CZE is a nonchromatographic separation technique. Although CZE does not separate neutral compounds, it is probably the simplest and most common of the capillary electrophoresis separation techniques in use today (23, 28). The separation techniques utilized in CZE are based on differences in the charge/mass/ratios of the analytes with the most positively charged analytes detected first. Separation efficiencies in CZE are typically 10–100 times that of HPLC separations, with detection sensitivity at the attomole (10^{-18} mol) level or less (23).

Laser induced fluorescence (LIF) detection may be used for fluorescent analytes, such as the fluoroquinolone antibiotics that demonstrate native fluorescence when excited at 325 nm. However, at the present time, LIF has seen rather limited use due to the fact that lasers are only commercially available with a limited range of wavelengths. This method of detection is still much too expensive for routine laboratory use, which is unfortunate considering that LIF has pushed detection limits to the zeptomole level (10^{-21} mol) (23, 30). Capillary

electrophoresis is an emerging technology that offers exciting possibilities for the determination of drugs in biological fluids.

Assessment of Drug Assay Methods

Criteria for assessment include 1) blank signal, 2) specificity, 3) precision and accuracy, 4) sensitivity, and 5) cost. Quantitation is a major issue in drug assay (10). It has been said that it is better to be roughly accurate than precisely wrong. There are few, if any, unique statistical aspects of drug assay, so there is little need to consider statistical matters such as coefficients of variation, in general, in this review. Generally speaking, drug analysts prefer the mean of two estimates over single assays. Triplicates are rarely of great value. Use of single estimates is commonly dictated by the conditions of the biologic investigation; this is obviously the case when all available biologic sample is utilized in obtaining a single estimate. All reports should include data on precision and accuracy of a method in the hands of those who conducted the work, even with extensively published, validated, and applied methods.

Several special aspects of drug assay quantitation are connected with signal-to-blank ratios and with pharmacokinetic experiments in particular. First, blank plasma (in samples from untreated patients) contains materials that give positive signals for the drug of interest. These signals obviously have to be minimized (reduction of “blank”). The signals of interest typically ought to be significantly higher than the blank signals. However, the most interesting scientific conclusions often depend on the signals close to blank. In the past, signals less than three times that of the blank were not used. In fact, this rule of thumb is an oversimplification. Signals much smaller than three times blank are usable, provided they are scientifically and statistically larger than blank. Whether or not this is so depends on the precision of the estimate of the blank and the lowest calibration point. Only a thorough review of the replication characteristics of the assay method can test this adequately. Incidentally, as in all assay work, conclusions should never be drawn on the basis of signals above blank but below the lowest calibration point.

Drug analysts must be careful with sample identification and handling. For example, plasma, serum, and blood are likely to contain different concentrations of drugs. Extraction characteristics from these three matrices also vary. This has implications for standardization and data reporting. The absolute rule in setting up standards is to compare like with like. Of course, setting up standards with exactly the same mix of contaminants and drug metabolites as in the test samples is never possible, but it

can be ensured that the same matrix is used. Standards should also be subjected to the same handling conditions as those for experimental samples in regard to collection and storage, transmission by carriers between laboratories, and so forth. Some drug analysts have set up quality control schemes to test reproducibility of their methods. Control charts such as those used in clinical laboratories are obviously useful. Multicenter quality control schemes have produced some alarming and perhaps erroneous results. Drug assays can be adversely affected by a variety of nonspecific factors, such as hemolysis in sample handling, bacterial contamination of blood collection tubes that lead to enzymatic losses of drug, anticoagulant interference, centrifugation conditions, assay contaminants entering samples from syringes and tubes, and use of antioxidants that cause reversion of drug metabolites to parent drugs. Only the application of the principle of comparison of like with like can overcome these problems.

REFERENCES

1. Brodie, B.B. Physicochemical and Biochemical Aspects of Pharmacology. *JAMA* **1967**, 202 (7), 600–609.
2. Brodie, B.B.; Udenfriend, S.; Bear, J.E. The Estimation of Basic Organic Compounds in Biological Material. *J. Biol. Chem.* **1947**, 168, 299–309.
3. Rowland, N.; Tozer, T.N. *Clinical Pharmacokinetics*; Lea & Febiger: Philadelphia, 1980.
4. Evans, W.E., Shentag, J.J., Jusko, W.J., Eds. *Applied Pharmacokinetics, Applied Therapeutics*; San Francisco, 1980.
5. Benet, L.Z.; Massoud, N.; Gambertoglio, J.E. *Pharmacokinetic Basis for Drug Treatment*; Raven Press: New York, 1983.
6. Fedeniuk, R.W.; Shand, P.J. Theory and Methodology of Antibiotic Extraction from Biomatrices. *J. Chromat. A.* **1998**, 812, 3–15.
7. Maurer, H.H. Liquid Chromatography–Mass Spectrometry in Forensic and Clinical Toxicology. *J. Chromat. B* **1998**, 713, 3–25.
8. Smith, R.V. Determination of Drugs and Metabolites in Biological Fluids. *Trac-Trend Anal. Chem.* **1984**, 3 (7), 178–181.
9. Curry, S.H. Drug Assay in Therapeutic Monitoring. *Trac-Trend Anal. Chem.* **1986**, 5 (4), 102–105.
10. Curry, S.H.; Whelpton, R.I. Statistics of Drug Analysis and the Role of Internal Standards. *Blood Drugs and Other Analytical Challenges*; Reid, E., Ed.; 1978; 29–41.
11. Curry, S.H. Determination of Nanogram Quantities of Chlorpromazine and Some of Its Metabolites in Plasma Using Gas–Liquid Chromatography With An Electron Capture Detector. *Anal. Chem.* **1968**, 40, 1251–1255.
12. Whelpton, R.; Watkins, G.M.; Curry, S.H. Bratton–Marshall and Liquid-Chromatographic Methods Compared for Determination of Sulfamethazine Acetylator Status. *Clin. Chem.* **1981**, 11, 1911–1914.
13. Whelpton, R.; Curry, S.H.; Watkins, G.M. Analysis of Plasma Trifluoperazine by Gas Chromatography and Selected Ion Monitoring. *J. Chromatogr.* **1982**, 228, 321–326.
14. Rosenfeld, J.M. Solid-Phase Analytical Derivatization: Enhancement of Sensitivity and Selectivity of Analysis. *J. Chromat.* **1999**, 843, 19–27.
15. Dean, J.R.; Khunder, S.; Extraction of Pharmaceuticals Using Pressurized Carbon Dioxide. *J. Pharm. Biomed. Anal.* **1997**, 15, 875–886.
16. Byrne, S.H., Jr. *Modern Practice of Liquid Chromatography*; Kirkland, J.J., Ed.; Wiley-Interscience: New York, 1971; 195.
17. Synder, L.R.; Kirkland, J.J. *Introduction to Modern Liquid Chromatography*; Wiley-Interscience: New York, 1974; 144.
18. Hamilton, R.J.; Sewell, P.A. *Introduction to HPLC*; Chapman and Hall: London, 1977; 55.
19. Seiler, N.; Demisch, L. *Handbook of Derivatives for Chromatography*; Heyden: London, 1978; 346.
20. Hulshoff, A.; Lingeman, H.; Fluorescence Detection in Chromatography. *Molecular Luminescence Spectroscopy: Methods and Application*; Part I, Schulman, S.G., Ed.; New York, 1985; Chap. 7.
21. Theorell, J.I.; Larson, S.M. *Radioimmunoassay and Related Techniques*; Mosby, C.V., Ed.; St. Louis, 1978.
22. Odell, W.D.; Daughaday, W.H. *Principles of Competitive Protein Binding Assays*; Lippincott: Philadelphia, 1971.
23. Christian, G.D. Clinical Chemistry. *Analytical Chemistry*; John Wiley and Sons Inc.: New York, 1994; 611–628.
24. Ullman, E.E.; Langen, J.; Clapp, J.J., Eds.; *Liquid Assay: Analysis of International Development on Isotopic and Non-Isotopic Immunoassay*; Masson: New York, 1981; 113.
25. Sharma, A.; Schulman, S.G. Fluorescence Analytical Methods and Their Applications. *Introduction to Fluorescence Spectroscopy*; Wiley-Interscience: New York, 1999; 123–158.
26. Karnes, H.T.; O'Neal, J.S.; Schulman, S.G. Luminescence Immunoassay. *Molecular Luminescence Spectroscopy: Methods and Applications*; Part 1, Schulman, S.G., Ed.; Wiley-Interscience: New York, 1985; Chap. 8.
27. Messeri, G. Chemiluminescence in Immunoassays. *Amer. Clin. Lab.* **1998**, 17 (1), 6–7.
28. Thormann, W.; Zhang, C.; Schmutz, A. Capillary Electrophoresis for Drug Analysis in Body Fluids. *Ther. Drug Monit.* **1996**, 18, 506–520.
29. Brewer, J.M. Electrophoresis. *Clinical Chemistry*; Kaplan, L.A., Pesce, A.J., Eds.; Mosby: St. Louis, 1996; 199–212.
30. Hempel, G. Strategies to Improve the Sensitivity in Capillary Electrophoresis for the Analysis of Drugs in Biological Fluids. *Electrophoresis* **2000**, 21, 691–698.
31. Perrett, D.; Ross, G. Capillary Electrophoresis of Drugs in Biological Fluids. *Bioanalytical Approaches for Drugs, Including Anti-Asthmatics and Metabolites*; Reid, E., Wilson, D., Eds.; The Royal Society of Chemistry: Letchworth, 1992; 22 (A), 269–278.
32. Baryl, N.E.; Lucy, C.A. Simultaneous Separation of Cationic and Anionic Proteins Using Zwitterionic Surfactants in Capillary Electrophoresis. *Anal. Chem.* **2000**, 72 (10), 2280–2289.

BIOABSORBABLE POLYMERS

Shalaby W. Shalaby

Poly-Med, Inc., Pendleton, South Carolina, U.S.A.

SCOPE

Addressed in this article are polymers that, because of their current and future applications, require that they interface with living tissues for predetermined periods. Obviously, the physical presence of these polymers in the biological environment must be transient, and their physicochemical properties are expected to meet the in-use requirements and allow their transport or clearance from the application site when the intended efficacy is achieved. The transport or clearance of those polymers while in contact with tissues is termed *bioabsorption*. Hence, the term *bioabsorbable polymers* is used to denote all transient polymeric materials, regardless of their chemical type, origin, and mode of clearance or absorption from the application site. Bioabsorbable polymers can be derived from naturally occurring polysaccharides and proteins, totally synthetic polymers, or combinations of natural and synthetic components, and they may be transported from the application site through the following: 1) dissolution followed by diffusion, with or without a simple ion-exchange process to achieve solubility; 2) chemical chain dissociation of insoluble materials and their conversion to small particles suitable for phagocytosis or to soluble chain fragments that may also be metabolizable; or 3) enzymatic dissociation of the insoluble polymers to soluble moieties. Major types of bioabsorbable polymers and possible modes of bioabsorption are discussed in the next section. This section also provides some technological details on bioabsorbable materials based on soluble or solubilizable chain molecules and on polymers that can be converted to soluble forms without undergoing scission of any covalent bonds. Then an important class of polymers that undergo bioabsorption through depolymerization and chain scission is discussed. This section also includes a discussion on the newly available absorbable gel-former liquids. Throughout both sections, brief discussions relevant to the application of bioabsorbable polymers address topics such as polymer processing and applications.

MAJOR TYPES OF BIOABSORBABLE POLYMERS AND POSSIBLE MODES OF BIOABSORPTION

The chemical structure of the chain molecules is the major factor that determines the mode of bioabsorption. About two decades ago, most bioabsorbable polymers were natural polymers or derivatives thereof. Scientific interest in totally synthetic bioabsorbable polymers has grown considerably since the early seventies because of their relatively low tissue reaction and because of their more predictable in vitro and in vivo properties as compared with the natural materials. Bioabsorbable polymers can be classified into three major groups: soluble, solubilizable, and depolymerizable. The structural features and possible mode of bioabsorption of these polymers are outlined as follows.

Soluble Polymers

Soluble chain molecules are usually water-soluble, natural, modified natural, or synthetic materials. These polymers are characterized by having hydrogen-bonding, polar groups that are responsible for their water-solubility and their transport in the biologic environment. The solubility of these polymers is dependent, to a great extent, upon the type of these polar groups (hydroxyl, carboxyl, amido, or amino) and the location in the chain (as part of the main chain or side groups). Molecular weight (1) and branching of the polymer chain also play a key role in determining the polymer solubility. Examples of water-soluble polymers and the type of polar groups they carry are given in Tables 1 and 2.

Some of the most important water-soluble polymers belong to the family of polysaccharides. Soluble polysaccharides were discussed in a review by Franz (2) on polysaccharides in pharmaceutical formulations. Due to their industrial importance, key characteristics and typical applications of those polysaccharides are listed in Table 3. In addition to those noted in Table 3, the oxidation of cellulose (to convert 16–24% of its primary hydroxyl groups to carboxyl groups) produces oxidized cellulose, a

Table 1 Types of polar groups in typical examples of water soluble polymers

Polar group	Natural polymers	Modified natural polymers	Synthetic polymers
–OH	Dextran and alginic acid	Hydroxyethyl cellulose and hydroxypropyl cellulose	Polyvinyl alcohol, polyhydroxyethyl methacrylate, and low-molecular-weight polyethylene oxide
–C–O–C	Dextran	Hydroxyethyl cellulose and hydroxypropyl cellulose	Polyethylene oxide and copolymers of ethylene and propylene oxide
–COOH	Alginic acid	Carboxymethyl cellulose	Polyacrylic acid
–CO–NH ₂	—	—	Polyacrylamide
–CO–NH–	—	Gelatin	Polylysine
–C–NH ₂	—	—	Polylysine

fibrous white powder. It is used as a bioabsorbable hemostatic agent. The bioabsorption may be partly attributed to solubilization through the carboxylate moieties.

Blends of different polysaccharides and plasticizers are commonly used in pharmaceutical formulations, particularly those geared for controlled release. Due to its importance, the physical compatibility of these blends was examined by a number of investigators. Sakellariou and associates (3) studied the polymer–polymer interaction in blends of ethyl cellulose with a few cellulose derivatives (hydroxypropyl methyl-cellulose, hydroxypropyl cellulose, hydropropyl methylcellulose phthalate, and cellulose acetate phthalate) and polyethylene glycol-6000. They noted that hydroxypropyl methylcellulose phthalate and cellulose acetate phthalate exhibited initial interaction and mutual dispersion owing to the presence of the phthalyl groups.

The growing interest in new or modified water-soluble polymers or combinations thereof to meet the needs of contemporary pharmaceutical technology is illustrated by a few of the recent studies outlined below. McCormick et al. (4) have relied on copolymerization as an effective means to modify water-soluble polymers, as illustrated in a recent communication. The effect of alkyl side groups on the properties of polyelectrolytes was examined earlier by Dubin and Strauss (5, 6). Pantar and Rao (7), who were also interested in developing polyacrylamide (PAAm) having enhanced dissolution rates, noted the following: 1) the rate of dissolution of this polymer increased when it is prepared in the presence of polyethylene glycols (PEG) 200 and 400; 2) heterogeneous PAAms were found to have occluded

the PEG; and 3) incorporation of PEG decreased with the increase in its molecular weight. Gelation and properties of hydrogels based on water-soluble polymers remain important areas, as stressed in a review by Ross-Murphy and McEvoy (8). The gamma ray-induced cross-linking of PAAm with a dose of over 50 kg produces hydrogels that absorb water by 1000–1500 wt% (9). Barbucci et al. (10) synthesized new polymers with amido and amino groups in the side chains and demonstrated the importance of hydrophobic interaction on the protonation of these polymers leading to their solubilization. Association of water-soluble polymers and formation of hetero-pairs by intermolecular complex formation have been the subject of early investigations and hold great promise for future research (5, 6, 11–15).

A review by Drobnik and Rypáček (16) summarized key findings and some theories on the fate of water-soluble polymers in living organisms, using a compartmentalized model to discuss the polymers' pharmacokinetics and their movement across several biologic barrier membranes in an organism. Compartments of this model are characterized by similarity in the mechanism of crossing the biologic barriers. Thus, the plasma circulation is viewed as the central and the only compartment across which the exchange of compounds between remote parts of the body may be accomplished. Intracellular exchanges were associated with a large intracellular compartment entailing all cells in the body. The authors discussed in some detail the participation of cells of the reticuloendothelial system (RES), kidney tubular epithelium, and liver hepatocytes. The lymph system and interstitium were denoted as important compartments in dealing with synthetic, water-soluble

Table 2 Structural features of typical water-soluble polymers

Name	Class of polymer or source	Solubilizing groups	Other features
Dextran (DXT)	Natural polysaccharide; a bacterial fermentation product	-C-O-C-	Substituted pyranose chain sequences
Carboxymethyl cellulose and its alkali metal salts (CMC)	Carboxyalkylated cellulose	-C-OH -C-O-CH ₂ COOH	Substituted pyranose chain sequences
Hydroxyethyl cellulose (HEC)	Ethoxylated cellulose	-C-O-C	Substituted pyranose chain sequences
Hydroxypropyl cellulose (HPC)	Propoxylated cellulose	-C-O-CH ₂ -CHCH ₃ -OH -C-O-C	Substituted pyranose chain sequences
Hydroxypropyl methylcellulose (HPMC)	Methylated HPC	Similar to HPC	Substituted pyranose chain sequences
Poly(β -hydroxyethyl methacrylate) (PHEMA)	Synthetic, by direct polymerization	-CO-O-CH ₂ CH ₂ OH	Substituted polyethylene chain
Polyvinyl alcohol (PVA)	Synthetic, hydrolysis product of polyvinylacetate	CH-OH	Substituted polyethylene chain
Polyacrylamide (PA)	Synthetic, by direct polymerization	-CONH ₂	Substituted polyethylene chain
Polyacrylic acid and its alkali metal salts (PAA)	Synthetic, by direct polymerization	-COOH, -COO-Na ⁺ and/or -COO-K ⁺	Substituted polyethylene chain
Polyvinylpyrrolidone (PVP)	Synthetic by direct polymerization	-CO-NH-	Substituted polyethylene chain
Polyethylene oxide (PEO or PEG)	Synthetic polyether, by acyclization of ethylene oxide	-CH ₂ -O-CH ₂ -	Chains carry 2 (OH) end groups, at low and moderate molecular weights
Poly(ethylene oxide- <i>b</i> -propylene oxide) (PEO-PPO)	Synthetic block copolymers of ethylene and propylene oxide	-CH ₂ -OH -CH ₂ -OCHCH ₃ -	Chains carry up to 2 (OH) end groups, depending on molecular weight
Amylose (soluble starch)	Natural, linear component of starch	-CH ₂ -O-CH ₂ -CH ₂ OH -CHCH ₃ -OH	Substituted pyranose sequences
Alginate acid and its alkali metal salts	Natural, polysaccharide from seaweed	-C-O-C -C-OH C-O-C	Substituted pyranose sequences of <i>d</i> -mannuronic acid
Polylysine	Synthetic	-C-COOH -C-OH -CO-NH- -C-NH ₂ -C-O-C-	A basic polypeptide
Hyaluronic acid	Natural polysaccharide (e.g., in rooster comb); also a bacterial fermentation product	-C-OH -C-NHCO-CH ₃ -COOH	Substituted pyranose sequences of acetyl Glucosamine and glucuronic acid

Table 3 Soluble polysaccharides: key features and typical applications

Polysaccharide	Structural features	Source and properties	Applications
β -Cyclodextrin	Based on 7 glucopyranose units; the α -form has only 6 units	Amylose derived, cyclic heptasaccharide, capable of forming inclusion complexes of drugs	Inclusion complexes with drugs to mask odor to taste
Methylcellulose	Methylated cellulose with a methoxy content of about 6 – 15%	Solutions are stable at pH 2–12 but coagulate or precipitate in presence of SO_4^{2-} , CO_3^{2-} or PO_4^{3-}	Ophthalmic and burn preparations, nose drops and ointments
Cellulose acetate phthalate	Esterified cellulose with about 20 and 35% acetate and phthalate groups, respectively	Free acid form dissolves in organic solvents and only the alkali metal salts are water soluble	Protective coating of tablets, insoluble in stomach but soluble in intestine fluids
Carboxymethyl cellulose	Usually prepared as the carboxylic salt of polycarboxymethyl ether of cellulose	Dispersion of the free acid or its sodium salt is stable at a wide pH range (2–10)	Thickening agent and tablet excipient
Acacia gum	A mixture of structurally related polysaccharides, the main one is based on β -D-galactopyranose residues	Natural plant exudates solutions are stable at pH 2–7 and display relatively low viscosity	Emulsifying or suspending agents, adhesive and binders (in tablets)
Tragacanthin	A complex mixture of polysaccharides containing some galacturonic acid residues	The soluble component of the natural (plant origin) tragacanth gum	Binder (in tablets) and demulcent
Pectin	Made of partly methoxylated 1,4-linked polygalacturonic acid	At pH 3 pectin forms a thermoreversible gel. It is extracted from citrus peel	Suspending agent, antidiarrheal formulations
Alginate acid	Linear polymer made of β -1, 4-D-mannuronic and L-glucuronic acids	Extracted from brown algae, the polymer sodium salt and not the free acid or calcium salt dissolves in water	Thickening, emulsifying, or gel-forming agent
Agar	Based primarily on galactose sequences	Extracted from red algae	Suspending or emulsifying agent, surgical lubricant, and tablet disintegrant
Galactomannans	Made primarily of a linear mannose chain with galactose side groups	Obtained from the powdered endosperm of certain seeds, guar gum	Thickening agent and tablet binder
Xanthan	Made primarily of linear glucose chain with side chains based on mannose and glucuronic acid residues	A microbial gum made by glucose fermentation, forms viscous pseudoplastic solutions	Emulsifying or suspending agent, particularly in toothpaste and ointments
Dextran	The polysaccharide chain is made primarily (based on the bacterial strain used in fermentation) of glucan units having α -1,6-linkages and fewer α -1,2- and 1,3 or 1,4-linkages to give a highly branched structure	A class of exocellular bacterial glucans, commercial dextrans have molecular weights of 40 to 110×10^3 . The low molecular wt. dextran (40,000) was reported to clear readily through the kidneys	Blood extender
Hyaluronic acid	Made of acetyl glucosamine and glucuronic acid (partially neutralized to dissolve) sequences	Obtained by the extraction of rooster comb or as a bacterial fermentation product, molecular weight ranges from 0.5 to 3×10^6	Exceptional lubricant for living tissues

(From Ref. 2.)

polymers. Ways in which water-soluble, endogenous and exogenous polymers cross compartmental barriers include transendothelial passage, transport into the lymphatic system, glomerular filtration, tubular secretion, intestinal transport, and biliary transport. In their discussion of compartmental barrier crossing, the authors indicated that water-soluble polymers cannot pass across a lipoprotein membrane by any type of diffusion process without the interruption of the membrane integrity and the development of defects in the cell surface. These defects can be caused by polyelectrolytes or a high concentration of polyethylene glycol. On the other hand, a common process by which polymers can cross a biologic membrane is invagination of the membrane, that is, the formation of a vesicle that buds inward and separates from the membrane. This process, in which the macromolecules enter the cell completely enclosed in a membrane vesicle containing some extracellular fluid, is generally termed *endocytosis*. The authors indicated that vesicle transport seems to be the only way in which synthetic polymers can overcome the lipoprotein membrane barrier. Drobnik and Rypáček (16) also

discussed the mechanism of storage in cells and elimination via the respiratory system. Although the authors stressed the role of endocytosis in transporting large molecules across biologic barriers, the passive diffusion of soluble polymers of less than 20,000 molecular weight was not excluded. Polymers addressed in this section may or may not undergo limited changes during their residence in the organism. The changes may consist of modifications that have minor effects on their mode of transport, absorption, and eventual elimination from the organism. The chain modification may be associated with hydrolytic reactions, oxidative processes, and/or conjugation (usually through esterification, acylation, or alkylation) but are not expected to introduce major changes in the chain molecular weight or solubility.

Solubilizable Polymers

The second group of bioabsorbable polymers, namely, the solubilizable macromolecules, may be viewed as derivatives of the water-soluble polymers as outlined in

Table 4 Soluble depolymerizable polymers

Name	Class of polymer of source	Solubilizing groups	Other features
Hyaluronic acid (HA) and its alkali metal salts	Natural polysaccharide from rooster comb or obtained as a fermentation product	C—O—C —C—COOH —C—NH—CO—CH ₃ —C—OH	Substituted pyranose cosequences of glucuronic acid and acetyl glucosamine
Pectin and its alkali metal salts	Natural polysaccharide from citrus fruits	—C—O—C —C—COOH —C—OH	Substituted pyranose sequences of galacturonic acid
Gelatin	A hydrolysis product of collagen (a natural polymer)	Mostly —CO—NH—	Major constituent amino acids are glycine, proline, and hydroxyproline
Protamine (as a sulfate)	Natural polymer from fish eggs	—COHN— —COHN— —NH—C—NH ₂ NH	A basic low-molecular-weight protein with arginine as a dominant amino acid constituent

Tables 2 and 3 or as a subgroup of the depolymerizable soluble polymers, which are described in Table 4. The solubilizable macromolecules can be insoluble calcium or magnesium salts of carboxylic or sulfonic acid bearing synthetic chains that undergo dissolution in the organism by cation exchange with alkalic metal salts. From this point on, the bioabsorption of these materials follows the same pattern as those noted for the water-soluble polymers. In addition, a solubilizable polymer can conceptually be a synthetic insoluble macromolecule that undergoes limited in-chain or side-chain scission to produce water-soluble, lower molecular weight fractions. For instance, a polyether-ester copolymer of polyethylene glycol diol with glycolic acid and/or an aliphatic diacid may undergo hydrolysis to a water-soluble polyethylene oxide and low molecular weight polyester fragments that undergo further hydrolysis to water-soluble monomeric species. The bioabsorbability of polymethyl cyanoacrylate may also in part be attributed to hydrolysis of the ester and/or the cyano group attached to the polyethylene main chain to create water-soluble polymeric moieties, which are transported and eliminated from the living organism earlier for intrinsically water-soluble polymers as discussed.

Depolymerizable Polymers

The third class of bioabsorbable polymers are those made of chains that undergo depolymerization to simple organic compounds in the living organism. The depolymerization may take place via an enzyme-catalyzed or chemically induced scission of the polymer chains, as in the case of collagen or polyglycolic acid, respectively. Most, if not all, the enzymatically

depolymerizable (or simply degradable) polymers are naturally occurring chains such as collagen or fermentation products, as in the case of poly- β -hydroxybutyric acid (PHB) (17). In addition, enzyme-degradable (or depolymerizable) polymers can be water-soluble (e.g., gelatin, pectin, hyaluronic acid, and protamine sulfate) or water-insoluble (e.g., collagen). A list of the water-soluble depolymerizable polymers is given in Table 4. Depending on their molecular weight, movement of the water-soluble polymers across the biologic membrane can take place by passive diffusion or endocytosis (or pinocytosis) as for low and high molecular weight chains, respectively. In most cases, however, water-soluble polymers undergo enzyme-catalyzed chain degradation to essentially monomeric species that can be metabolized further or excreted as such. The water-insoluble depolymerizable polymers are represented by three major naturally occurring polymers and derivatives thereof. The most common types of water-insoluble depolymerizable polymers are those based on natural proteins such as collagen and its derivatives (through complex formation with transition metals such as chromium ions). These are degraded by collagenase to soluble low molecular or monomeric species that may be excreted as such or metabolized further to ammonia water and carbon dioxide. Chitin, a polysaccharide, and its deacylated derivative, chitosan, represent the second important type of insoluble enzyme-degradable polymers. Poly- β -hydroxybutyric acid and copolymers β -hydroxybutyric and β -hydroxyvaleric acid, which are fermentation products, represent the third type, the relatively new water-insoluble enzyme-depolymerizable polymers (see Table 5). The polysaccharide and polyester types of insoluble materials are likely to undergo enzyme

Table 5 Typical examples of natural enzyme depolymerizable waterinsoluble polymers

Polymer	Class of polymer or source	Special comments
Collagen	A protein, the main constituent of connective tissues and the organic component of bones	It contains high concentrations of glycine (33%) and proline (13%); it also contains hydroxyproline (10%) and the uncommon hydroxylysine (10%)
Chitin	A natural polysaccharide obtained from crab shells	The chain is based on acetylated glucosamine units; the natural polymer can be converted to the acid-soluble partially deacetylated product chitosan
Poly- β -hydroxybutyrate and copolymers	Aliphatic polyesters of 3-hydroxybutyric acid (HB) and copolymers of HB and 3-hydroxyvaleric acid (HV)	Prepared by fermentation; most polymers are crystalline thermoplastic film and fiber-forming materials and dissolve in certain organic solvents

degradation first to water-soluble low-molecular-weight or monomeric species for their transport across biologic membranes.

SYNTHETIC BIOCHEMICALLY ABSORBABLE POLYMERS

Until the late 1960s, collagen-based surgical products (mainly sutures) dominated the field of absorbable polymers. However, a few undesirable features associated with the natural origin of these products motivated the health care community to develop synthetic, absorbable polymers with more predictable and superior properties. In the past two decades, four major synthetic, biochemically absorbable polymers were introduced in the form of surgical sutures and allied surgical devices. These are polyglycolide (PGA); 90/10 poly(l-lactide-co-glycolide) (90/10 PLG); poly-*p*-dioxanone (PDS); and copolymers of poly(trimethylene carbonate and glycolide) (18–20). Other forms of lactide–glycolide copolymers and their blends have been converted to absorbable staples (21, 22). Although several forms of polylactides, including the optically pure poly-l-lactide, have been discussed extensively by many investigators (23, 25), their marketing as useful health-care products is hardly existent. A major potential use of these polymers is likely to be in the area of bone augmentation. The degradation of the lactide-glycolide systems and PDS was examined by a few authors (18, 26–28). Although the bioabsorption of these polymers is considered to be chemically driven, a few authors advocate that lactide-glycolide polymers (29, 30) and caprolactone polymers (31) undergo enzymatic degradation.

Synthetic bioabsorbable polymers other than those made mostly from glycolide, lactide, or *p*-dioxanone have been described in the technical and patent literature (18, 32–35). Dominant among these polymers are the polyoxalates (18, 34, 35), poly(carbalkoxyalkyl 2-cyanoacrylates) (36), polyanhydrides (37, 38), and absorbable organometallic or inorganic polymers such as polyphosphazenes (39) and phosphate glass as composites in absorbable organic matrices (40). Different types of polyalkylene oxalates and isomorphous copolyoxalates (based on 1,6-hexanediol and 1,4-*trans*-cyclohexanedi-methanol) have been patented as useful bioabsorbable materials with tailored bioabsorption profiles, depending on their chemical structure, molecular weight, and morphology (34, 35). These polymers were noted as useful for the production of many surgical implants (including sutures) and surface lubricants. Polyanhydrides

such as poly(trimethylene-bis-*p*-oxybenzoic anhydride), poly(terephthalic anhydride), and their copolymers with poly(sebacic anhydride) have been synthesized, and their degradation as bioabsorbable materials was studied by Leong and co-workers (37). These polymers were indicated to be nontoxic and nonmutagenic (38). Other polymers that have been described as bioabsorbable include copolymers of poly(trimethylene malonate and *p*-dioxanone (41), polyalkylene oxamates (35), polyester-amides (42), poly-*p*-malolactone (43), copolymers of dl-lactide and ethylene oxide (44), copolymers of substituted glycolic acid and glycine (45), and poly-orthoesters (46). Modulation of the properties of biochemically absorbable polymers to impart certain desirable properties has been achieved by Shalaby and his co-workers (18, 47, 48). To increase the bioabsorption of poly-*p*-dioxane, a small proportion of codimeric sequences based on morpholinedione were introduced into the polymer main chain (49). In order to improve the radiation stability of polyglycolide and poly-*p*-dioxanone, copolymers and/or melt-blends of these polymers with a polyester based on phenylene-diglycolic acid or carboxymethylated *p*-hydroxy benzoic acid were prepared and converted to radiation-sterilizable surgical articles (47, 48, 50–53). Typical examples of the biochemically absorbable polymers and an outline of their key properties and/or applications are given in Table 6. The newly developed class of absorbable gel-former polymers is addressed below.

ABSORBABLE GEL-FORMERS AND THEIR USE AS INJECTABLE CARRIERS

Growing interest in developing absorbable pharmaceutical surgical products that degrade in the biologic environment to safe byproducts and leave no residual mass at the application site (54–60), justified the search for novel, absorbable gels. In a recent disclosure (61), a novel gel-former was described to be based on absorbable copolymers which, upon hydration, result in hydrogels that are stabilized by pseudo-crosslinks provided by a hydrophobic polyester component covalently linked to a hydrophilic component made of pharmaceutically acceptable polymer, such as polyoxyethylene. The polyester component is made of safe monomers, such as *p*-dioxanone, ϵ -caprolactone, glycolide, lactide, and mixtures thereof. Contrary to a related study (62) that describes in situ formation of biodegradable, microporous, solid implants in a living body through coagulation of a

Table 6 Typical examples of waterinsoluble chemically depolymerizable polymers

General class	Specific examples	Key properties and/or applications ^a
Poly-2-hydroxy acids and copolymers	a. Polyglycolic acid (PGA)	A, B, F, G
	b. 10/90 l-lactide/glycolide copolymer (910 polyglactin)	A, B, F
	c. Poly-l-lactide	A, B, C, D, F, G
	d. Glycolide/ε-caprolactone copolymers	A, B, G, H, I
	e. Glycolide/1,5-dioxepan-2-one copolymers	A, B, G, H, I
	f. Glycolide/trimethylene carbonate copolymers	A, B, F, G
	g. Copolymers of glycolide and polyethylene- <i>p</i> -phenylene diglycolate	A, B, G, J
Poly- <i>p</i> -dioxanone and copolymers	a. Poly- <i>p</i> -dioxanone (PDS)	A, B, F, G
	b. Copolymers of PDS and l-lactide	A, B, G
	c. Copolymers of PDS and glycolide	A, B, G
	d. Poly- <i>p</i> -dioxanone-co-morpholine-2,5 dione	A, B, G
	e. Copolymers of <i>p</i> -dioxanone and polyethylene- <i>p</i> -phenylene diglycolate	A, B, G, J
	f. Copolymers of polytrimethylene malonate and <i>p</i> -dioxanone	A, B, G, H
Polyalkylene oxalates	a. Polyesters of C4 to C16 and copolymers	A, B, C, D, E, G, H, I
	b. Isomorphous copolyoxalate of cyclic and alicyclic diols	A, B, C, D, G
Polyester-amides	Polyalkylene oxamate based on β-hydroxy-l-hexanol	A, B, G
Polyanhydrides	Copolyanhydrides of sebacic and 1,3-propane-bis-(4-oxybenzoate)	C, H

^a A = Thermoplastic crystalline polymer; B = Melt processable into fibers and films; C = Soluble in certain common organic solvents; D = Film-former by solution dipping or casting; E = Low-T_m polymer (below 100°C); F = Used to produce surgical devices; G = Patented as useful polymers for surgical and allied devices; H = Described as useful matrices for drug delivery; I = Patented as a surface coating; J = Sterilizable by gamma radiation.

solution of a polymer in an organic solvent such as *N*-methyl-2-pyrrolidone, the new hydrogel-former does not require the use of solvents. Such solvents did include low molecular organic ones that can migrate from the application site and cause damage to living tissue, such as cell dehydration and necrosis. Equally important is the fact that previously known systems are solid implants that can elicit mechanical incompatibility and, hence, patient discomfort in contrast to the new compliant, swollen, mechanically compatible hydrogels (61).

The use of absorbable gel-formers may very well lead to some of the most important applications of absorbable polymers in the pharmaceutical and biomedical industries. Among the recent activities in this area are uses of the gel-formers in 1) periodontal application of antibiotics; 2) antibiotic formulations for osteomyelitis; 3) intraocular drug delivery; 4) wound healing and hemostasis; 5) controlling the release of insulin; 6) controlling the bioavailability of ricin A-chain; 7) wound repair as a suture adjuvant; 8) modifying absorbable tissue adhesives;

9) intravaginal controlled delivery of misoprostol; and 10) sealing microporous vascular grafts. These uses are discussed next.

Periodontal Application

This entails the use of injectable gel-forming formulations for controlled delivery of antibiotics, such as tetracycline or doxycycline, for combating periodontal infections for periods of 1–4 weeks (61).

Antibiotic Formulations for Bone Infection

In a Phase I study of an NIH-SBIR program addressing osteomyelitis, available results (63) indicate that 1) selected gel-formers are capable of controlling the *in vitro* release of gentamicin and vancomycin for at least 2 weeks; 2) two types of gel-formers can be formulated, with clinically relevant doses of vancomycin, into

injectable forms; 3) injection of the vancomycin formulation about the periosteum of the goat tibia for localized drug delivery; and 4) controlled release of the vancomycin formulation is feasible without leading to toxic blood levels.

Injectable Intraocular Delivery Systems

In an SBIR (Phase I) supported by the DOD, the feasibility of using tailored gel-formers to develop an injectable, controlled release system for intraocular delivery of key drugs is being investigated. The available data indicate that:

1. Injectable gel-formers containing pilocarpine, naproxen, cyclosporin, and ganciclovir in clinically relevant doses can be prepared.
2. A continued release in a buffered medium for at least one week can be achieved.
3. Active formulations of the four drugs and a placebo can be readily injected into the vitreous cavity of the rabbit eye without eliciting unacceptable, gross tissue reactions.

Burn Wound Healing and Hemostatic Application

Preliminary results of a study supported by a DOD grant on wound healing and hemostatic agents (using hairless rats and rabbits) indicate that:

1. Certain gel-forming formulations can be used for the controlled delivery of antibiotics to incisional and burn wounds in hairless rats.
2. Incisional wound repair in hairless rats can be improved when placebo gel-formers are used.
3. Selected gel-forming formulations can induce hemostasis in a rabbit animal model.

Insulin Controlled Release Systems

Preliminary study on the use of certain gel-formers for the controlled release of insulin demonstrates the feasibility of this concept (64).

Controlled Release of Ricin A-Chain and Other Vaccines

This has been the subject of a Phase I SBIR program supported by the DOD. Results and relevant conclusions of the study (65) are summarized below. Available results

on subcutaneously (sc) administered active formulations do verify that:

1. Gel formulations can be easily prepared and appear suitable for scale-up.
2. One sc formulation is capable of releasing sufficient amounts of ricin A-chain (RAC) to elicit IgG formation at protective levels over a period of 4–6 weeks.
3. One formulation provides persistent protection for at least 6 weeks postimmunization.
4. A correlation can be established between IgG formation and the composition of the polymeric carriers.

Available Phase I results do not only fulfill these criteria but also suggest that:

1. A single-shot, absorbable sc formulation, GF-II, exhibits potentially unique in vivo performance as it comprises a microparticulate cation-exchanger.
2. Upon comparing commercial RAC solution (RAC-L) with GF-II, the latter elicits a more gradual antibody response that peaks at 10 weeks, and it exceeds a fast-decaying, initially higher response to RAC-L.
3. In terms of antibody response, GF-II is associated with higher durability over the 10 – 20 week period.
4. GF-II elicits a higher response of IgG-2A than RAC-L at 6 weeks.

Skin Wound Repair

In a study of skin wound repair, using sutures or staples on rats, it was shown that covering the wound with gel-formers allows the use of half the number of interrupted suture stitches or staples while minimizing scar formation.

Tissue Adhesive Formulation

The addition of certain members of the family of gel-formers to methoxypropyl cyanoacrylate yields a series of tissue adhesive formulations, which can be used effectively as:

1. Sutures and staples in repairing skin wounds.
2. Affixing elastin-based patches in repairing defects at certain sites of the gastrointestinal tract.

Intravaginal Controlled Delivery of Misoprostol

In a recent study, gel-formers have shown to be effective carriers for the controlled release of misoprostol in a

newly developed animal model for studying induced cervical ripening.

Sealing of Microporous Vascular Grafts

The preliminary results of a recent study on expanded Teflon® vascular grafts show that certain gel-formers can be used as sealants for these microporous implants. As sealants, they can also be used as carriers for the controlled delivery of bioactive agents to prolong patency of the grafts.

REFERENCES

- Molyneux, P. Br. Polym. J. **1986**, 18 (32).
- Franz, G. Adv. Polym. Sci. **1986**, 76 (3).
- Sakellariou, P.; Rowe, R.C.; White, E.E.T. Inter. J. Pharm. **1986**, 34 (93).
- McCormick, C.L.; Blackman, K.P.; Elliott, D.L. Angew. Makromol. Chem. **1986**, 144 (87).
- Dubin, P.L.; Strauss, U.P. J. Phys. Chem. **1973**, 77 (1427).
- Dubin, E.L.; Strauss, U.E. Hypercoiling in Hydrophobic Polyacids. *Polyelectrolytes and Their Applications*; Rembaum, A., Selegny, E., Eds.; D. Reidel Pub. Co.: Holland, 1975; 3.
- Pantar, A.V.; Rao, V.M.R. Makromol. Chem., Rapid Commun. **1986**, 7 (77).
- Ross-Murphy, S.B.; McEvoy, H. Br. Polym. J. **1986**, 18 (2).
- Burfillo, S.G.; Ogawa, T.J. Appl. Polym. Sci. **1986**, 32 (3783).
- Barbucci, R.; Casolaro, M.; Nocentini, M.; Reginato, G. Makromol. Chem. **1986**, 187 (1953).
- Kataoka, K.; Tsuruta, T. Makromol. Chem. **1980**, 181 (1363).
- Salamone, J.C.; Mahmud, N.A.; Mahmud, M.U.; Nagabhushanam, T.; Watterson, A.C. Polymer **1982**, 23 (843).
- Tsuchida, E.; Abe, K. Adv. Polym. Sci. **1982**, 45 (1).
- Abe, K.; Senoh, S. J. Polym. Sci. Part A; Polym. Chem. **1986**, 24 (3461).
- Saito, S. Coll. Surf. **1986**, 19 (351).
- Drobnik, J.; Rypátcek, E. Adv. Polym. Sci. **1984**, 57 (1).
- Miller, N.D.; Williams, D.E. Biomaterials **1987**, 8 (129).
- Shalaby, S.W. Fibrous Materials for Biomedical Applications. *High Technology Fibers*; Lewin, M., Preston, J., Eds.; Marcel Dekker, Inc.: New York, 1985; Ch. 3.
- Doddi, N.; Versfelt, C.C.; Wasserman, D. US Patent (to ETHICON, Inc.) 4,052,988, 1980.
- Katz, A.R.; Mukherjee, D.P.; Kaganov, A.L.; Gordon, S. Surg. Gynecol. Obstet. **1985**, 161 (216).
- Adzick, N.S.; Harrison, M.R.; Flake, A.W.; Glick, P.L.; Bottles, K. Surg. Forum **1985**, 36 (479).
- Steckel, R.R.; Jann, H.W.; Kaplan, D.; Jakowski, R.M. Obstet. Gynecol. **1986**, 68 (40).
- Leenslag, J.W.; Gogolewski, S.; Pennings, A.J. J. Appl. Polym. Sci. **1984**, 29 (2829).
- Leenslag, J.W.; Pennings, A.J.; Bos, R.R.M.; Rozema, E.R.; Boering, G. Biomaterials **1987**, 8 (70).
- Tune, D.C.; Rohovsky, M.W.; Jadhav, B.; Lehman, W.B.; Strongwater, A.; Kummer, E. Proceed. A. C. S. Div. Polym. Mater. Sci. Eng. **1986**, 53 (502).
- Gupta, M.C.; Deshmukh, V.G. Polymer **1983**, 24 (827).
- Chu, C.C.; Louie, M.J. Appl. Polym. Sci. **1985**, 30 (3133).
- Pitt, C.G.; Gu, Z.W. J. Controlled Release **1987**, 4 (283).
- Williams, D.E.; Chu, C.C. J. Biomed. Mater. Res. **1983**, 17 (1029).
- Vainionpää, S.; Kilpikari, J.; Liahio, J.; Helevirta, P.; Rokkanen, P.; Tormala, P. Biomaterials **1987**, 8 (46).
- Pitt, C.G.; Hendren, R.W.; Schindler, A. J. Controlled Release **1984**, 1 (3).
- Devi, K.S.; Vasudevan, E. J. Macromol. Sci. Rev. Macromol. Chem. Phys. **1985**, c-25 (315).
- Devi, K.S.; Sinha, J.M.; Vasudevan, P. Biomaterials **1986**, 7 (193).
- Shalaby, S.W.; Jamiolkowski, D.D. US Patent (to ETHICON, Inc.) 4,140,678; 4,141,087, 1979.
- Shalaby, S.W.; Jamiolkowski, D.D. US Patent (to ETHICON, Inc.) 4,205,399; 4,226,243, 1980.
- Kronenthal, R.L.; Schipper, E. US Patent (to ETHICON, Inc.) 3,995,641, 1976.
- Leong, K.W.; Brott, B.C.; Langer, R. J. Biomed. Mater. Res. **1985**, 19 (194).
- Leong, K.W.; Amore, P.D.; Marietta, M.; Langer, R. J. Biomed. Mater. Res. **1986**, 20 (51).
- Allcock, H.R. Macromol. Chem., Macromol. Symp. **1986**, 6 (101).
- Kronenthal, R.L. US Patent (to ETHICON, Inc.) 4,612,923, 1985.
- Koelmel, D.E.; Shalaby, S.W. US Patent (to ETHICON, Inc.) 4,559,945, 1985.
- Barrows, T.H.; Grussing, D.M.; Hegdahl, D.W. Trans. Soc. Biomater. **1983**, 6 (109).
- Lenz, R.W.; Vert, M. US Patent 4,265,247, 1981.
- Zhu, K.J.; Xiangzhou, L.; Shilin, Y. J. Polym. Sci., Part C, Polym. Lett. **1986**, 24 (331).
- Goodman, M.; Kirshenbaum, G.S. US Patent (to Suture, Inc.) 3,773,737, 1973.
- Choi, N.S. U.S. Pat. 4,093,709, 1978.
- Bezwada, R.S.; Shalaby, S.W.; Kafrawy, A. Newman, H. US Patent (to ETHICON, Inc.) 4,643,191, 1987.
- Bezwada, R.S.; Shalaby, S.W.; Newman, H. US Patent (to ETHICON, Inc.) 4,653,497, 1987.
- Shalaby, S.W.; Koelmel, D.E. US Patent (to ETHICON, Inc.) 4,441,496, 1984.
- Bezwada, R.S.; Shalaby, S.W.; Jamiolkowski, D.D. US Pat. (to ETHICON, Inc.) 4,510,295, 1985.
- Koelmel, D.E.; Jamiolkowski, D.D.; Shalaby, S.W. US Patent (to ETHICON, Inc.) 4,546,152, 1985.

52. Koelmel, D.E.; Jamiolkowski, D.D.; Shalaby, S.W.; Bezwada, R.S. US Patent (to ETHICON, Inc.) 4,649,921, 1987.
53. Shalaby, S.W.; Jamiolkowski, D.D. US Pat. (to ETHICON, Inc.) 4,435,590, 1984.
54. Shalaby, S.W.; Thermoreversible Gels. *Water-Soluble Polymers*; ACS Symp. Ser. American Chemical Society: Washington, DC, 1991; 467, Ch. 33.
55. Shalaby, S.W. Polym. News **1991**, 16 (238).
56. Shalaby, S.W. J. Appl. Biomater. **1992**, 3 (73).
57. Shalaby, S.W., Ed. *Biomedical Polymers: Design to Degrade Systems* : Hanser Publ.: New York, 1994.
58. Shalaby, S.W. et al. *Polymers of Biological & Biomedical Significance*; ACS Symp. Ser. American Chemical Society: Washington DC, 1994; 540.
59. Shalaby, S.W. et al. Irish Patent (to Kinerton, Ltd.) S-61251, 1994.
60. Shalaby, S.W. Indian J. Tech. **1993**, 31 (464).
61. Shalaby, S.W. US Patent (to Poly-Med, Inc.) 5,610,052, 1997.
62. Dunn, R.L. et al. US Patent 4,938,763, 1990.
63. Corbett, J.T. et al. Trans. Soc. Biom. **1998**, 21 (341).
64. Corbett, J.T. Poly-Med. Internal Report, 1996.
65. Corbett, J.T. et al. Trans. Soc. Biom. **1997**, 20 (362).

Bio-validation of Steam Sterilization

Nigel A. Halls

NHC—Nigel Halls Consulting, Chorleywood, Herts, U.K.

INTRODUCTION

The science that underpins steam sterilization is well known and has been long established. It is the preferred method of sterilization in the pharmaceutical industry; it is used for sterilization of aqueous products in a wide variety of presentations, for sterilization of equipment and porous materials required in aseptic manufacture, in microbiology laboratories for sterilizing media and other materials, and for sterilization of “massive” systems of vessels and pipework [steam-in-place (SIP) systems]. Numerous rules and guidelines have been published on the topic, yet steam sterilization and particularly bio-validation of steam sterilization is still a subject for controversy and debate.

The purpose of this article is to reexamine the bio-validation of steam sterilization, to clarify what is needed and why it is needed, and to distinguish the scientific need from the regulatory need in areas where they may appear to differ.

PRINCIPLES

Micro-organisms are inactivated when metabolically irreversible deleterious intracellular reactions occur. At high temperatures and in the presence of moisture, as in steam sterilization, the energy input from the steam inactivates micro-organisms by denaturation of intracellular proteins.

Although these reactions are complex at a biochemical level, their kinetics approximate to reactions of the first order. Thus, the kinetics of inactivation of populations of pure cultures of micro-organisms take the typical exponential form of reactions of the first order. What this means in experimental practice is that there is a linear relationship^[1] when numbers of micro-organisms held at high temperatures are plotted on a logarithmic scale against time plotted on an arithmetic scale (Fig. 1).

There are two highly significant points to be drawn from the kinetics of inactivation of micro-organisms.

First, logarithmic scales never reach zero. This means that there can never be any specifications for temperature and time which can guarantee that all micro-organisms contaminating items are going to be inactivated. However, the consequences to patients of micro-organisms surviving in allegedly sterile pharmaceutical preparations can easily be fatal. Thus, sterilization processes must be specified to ensure that the probability of micro-organisms surviving in treated items is low enough to ensure patient safety. The accepted low probability indicated in the pharmacopeias is that there should be not more than one chance in one million of viable micro-organisms surviving on a treated item. This is called a probability of nonsterility of 10^{-6} or a sterility assurance level (SAL) of 10^{-6} .

Second, the inactivation curve takes a regular form. This means that steam sterilization is a predictable process as long as some information is available (or can be safely assumed) about the numbers and thermal resistances (D_T -values, Fig. 1) of the micro-organisms contaminating items before treatment. This is important because there is no practical way to test for the achievement of SALs of 10^{-6} . The sterility or nonsterility of items cannot sensibly be confirmed in a treated item except by sacrificing the item. The pharmacopeial *test for sterility* is a sacrificial test with statistical limitations which have been so extensively criticized^[2–6] over so many decades that they should now be well understood. For instance, the sample of 20 items which is generally required in the test would allow a batch containing nonsterile items at a frequency of 1:100 to be passed on four out of every five occasions. This falls a long way short of being able to detect deviations from a standard of not more than one nonsterile item in one million.

Justification of the reliable achievement of SALs of 10^{-6} for particular pharmaceutical items treated according to particular specifications of temperature and time in particular sterilizers is predicated on the regularity and predictability of steam sterilization processes. The means of justification are through scientifically based development of sterilization specifications and sterilizer parameters,

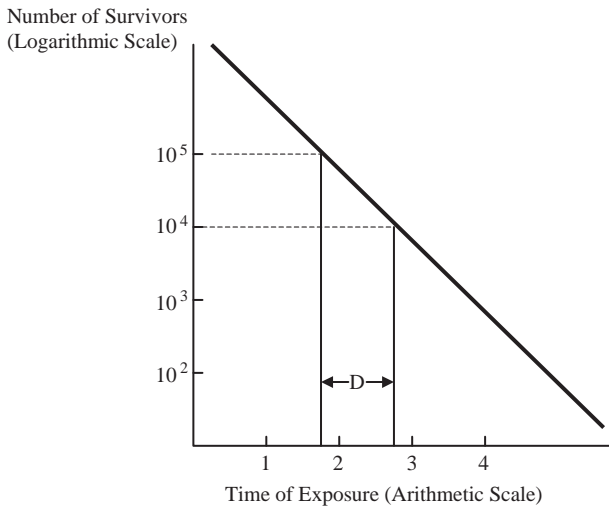


Fig. 1 Exponential inactivation of micro-organisms (the survival curve).

and through subsequent validation of the specified processes.

DEVELOPMENT OF STERILIZATION SPECIFICATIONS AND STERILIZER PARAMETERS

The development of sterilization specifications differs from the development of sterilizer parameters. Both differ from validation (Fig. 2).

Pharmaceutical Products and Materials for Aseptic Manufacture—Sterilization Specifications

For pharmaceutical products and materials used in connection with aseptic manufacture, sterilization specifications apply to conditions of temperature and time, or F_0 , or combinations of F_0 , temperature and time to which the contaminating micro-organisms themselves must be exposed over the “hold” period of the sterilization process. In practice, this means actually within aqueous products, on the surfaces of rubber stoppers or metal machine parts, or within the folds of cartridge filters, etc.

The sterilizer parameters are the practical criteria that must be specified to ensure that the sterilization specification is delivered to all parts of the load. They always include specifications for temperature and time, but it is important to recognize the distinction between sterilizer parameters applying to the machine settings on the autoclave console, and sterilization specifications applying to actual conditions within the load. Essential sterilizer parameters also include other specifications, e.g., for load configuration, number and depth of prevacuums, cooling characteristics, etc.

Sterilization specifications are product specific. Sterilizer parameters are specific to combinations of product, presentation, and autoclave.

Sterilization specifications may be determined from theoretical considerations or from laboratory data and are within reason transferable from presentation to

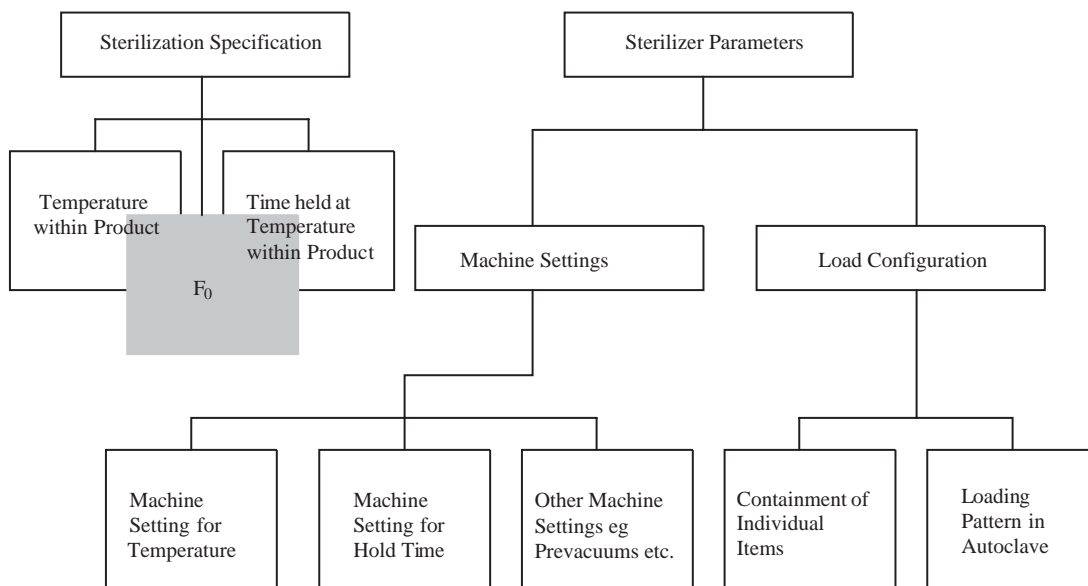


Fig. 2 Sterilization specifications and sterilizer parameters.

presentation, e.g., from 1 ml ampules to 5 ml vials to 50 ml bags. The sterilizer parameters required to deliver the sterilization specification to these presentations differ within the same autoclave and from one autoclave to another according to differences in load configurations, chamber size, steam entry points, control systems, etc. Sterilizer parameters are not transferable and must be developed empirically for each autoclave.

Sterilization specifications should be easy to develop. The pharmacopeias allow sterilization specifications to be developed from a basis of no actual data concerning the numbers and thermal resistances of micro-organisms actually contaminating the items to be sterilized. Although this statement may appear initially to be barren of scientific reason, this is in fact not the case. What the pharmacopeias provide are either a recommended overkill specification (*PhEur*) or principles for specification development (*USP*) that incorporate amounts of thermal lethality well in excess of that which could ever be practically required to obtain SALs of 10^{-6} —for this reason they are called “overkill” specifications.

In the *European Pharmacopoeia (PhEur)*, a specification of 121°C for 15 min is given as the reference condition for overkill sterilization of aqueous preparations. The *United States Pharmacopeia (USP)* defines a lethality input of 12D. These specifications merit some examination in detail.

It is worth considering the thermal resistances of micro-organisms found in pharmaceutical manufacturing environments. It is extremely rare for anyone to have isolated thermally resistant bacteria with D_{121} -values (in water) of greater than 0.3 min. The author of this paper has experience of having determined a D_{121} -value (in water) of 0.8 min for an environmental isolate of *Bacillus coagulans*, but this was several decades ago and was done with what would now be considered fairly primitive equipment. It is therefore probably quite reasonable to assume a worst case D_{121} -value of 1 min. Given this “worst case,” the *PhEur* overkill specification of 121°C would deliver 15 decimal reductions which are equivalent to assuring a 10^{-6} SAL for contaminating populations per item of up to 10^9 micro-organisms each with a D_{121} -value of 1 min. The *USP* specification of 12D under the same assumption ensures an SAL of 10^{-6} for populations of up to 10^6 micro-organisms per item.

Thus, the pharmacopeial overkill specifications provide considerable degrees of assurance that SALs of 10^{-6} will be achieved. However, these high theoretical levels of overkill are contingent upon D -values in water being reflected by D -values in or on product. Most pharmaceutical products depress the thermal resistance of micro-organisms relative to their D -values in water, but this is not universally true. Some other materials (e.g., rubber) are

known to increase the thermal resistance of micro-organisms (this may as likely be due to physical characteristics of heat transference as to biochemical protection). These product effects on thermal resistance can only be determined empirically, and are usually done in the laboratory using thermally resistant bacterial endospores, often spores of *Bacillus stearothermophilus*. Some previously unpublished guidance values on the effects of materials and pharmaceutical products on D -values of *B. stearothermophilus* relative to water are given in Table 1, and a detailed analysis has been published by Berger et al.^[7] The use of *B. stearothermophilus* for this purpose and their frequent use as biological indicators (BIs) in bio-validation have contributed to a belief that steam sterilization must be defined in terms of being able to kill this micro-organism. It is not; spores of *B. stearothermophilus* are used with steam sterilization because of the convenience of their high resistance to steam sterilization and the unique and distinctive conditions required for their recovery and growth. Indeed, some major companies use *Clostridium sporogenes* or other species of *Bacillus* as reference or indicator organisms for this purpose.

The use of “overkill” specifications is not mandatory. In some instances, there may be pharmaceutical products which are unable to withstand the temperatures or energy inputs of overkill specifications. In these cases, specifications can be developed by calculating SALs of 10^{-6} from data characterizing the number of micro-organisms actually contaminating items before sterilization treatment, or from data characterizing the actual numbers and thermal resistances of the contaminating micro-organisms. The question is this—Is this exercise worth doing or would it be better and simpler to opt for aseptic manufacture of such heat-sensitive products?

Let us consider the number of micro-organisms contaminating pharmaceutical products prior to sterilization.

Table 1 D -values of spores of *B. stearothermophilus* on various substrates relative to water (D -value approximately 4 min)

Substrate	% relative to D -value in water (100%)
Stainless steel	60
Hydrophobic filter media	90
Silicone tubing	100
Rubber stoppers (various types)	85–150
Polycarbonate	120
Two pharmaceutical products pH 3.4–3.7	15–40
Pharmaceutical product pH 10.5–10.7	115

What are the highest and the lowest numbers which could be expected? For sterile parenteral products, the highest tolerable number of micro-organisms would be expected to be on the order of 10^2 . This is because 10^3 or more per item is likely to begin to incur a risk of pyrogenicity. The lowest number which could be inferred from even an extensive number of zero counts would be one micro-organism.

Achievement of a 10^{-6} SAL from an initial bioburden of 10^2 would require eight log reductions. Applying these eight log reductions to an assumed worst case thermal resistance of D_{121} -value in water of 1 min gives a sterilization specification of 121°C for 8 min.

Achievement of a 10^{-6} SAL from an initial bioburden of 1 would require six log reductions. Applying these six log reductions to an assumed worst case thermal resistance of D_{121} -value in water of 1 min gives a sterilization specification of 121°C for 6 min.

The determination of thermal resistances is technically complex and requires special equipment (BIER Vessels). Since it is unlikely that *Bacillus* spp. can be excluded from any survey of microbiological contamination, it is reasonable to assume that spores with D_{121} -values on the order of 0.3 min will be isolated. Using this figure, SALs of 10^{-6} can be calculated at 121°C for 2.4 min for bioburdens of 10^2 , and at 121°C for 1.8 min for bioburdens of one micro-organism per item.

The range of sterilization specifications calculable by these various approaches is summarized in Table 2. It is apparent that very brief sterilization specifications (on the order of 2–3 min holding time at 121°C) are obtainable when the microbiological contamination is completely characterized in terms of numbers and thermal resistances. In practice, such limits on hold times could be difficult to control precisely, are probably insignificant in terms of thermal lethality compared with heat-up and cool-down times, and could prove difficult to “sell” to regulators. Without complete thermal characterization of thermal resistances, specifications calculable by the “bioburden”

approach are hardly significantly shorter than “overkill” specifications. Thus, it probably makes practical sense in most cases to choose only between overkill cycles for thermally resistant products and aseptic manufacture for heat-sensitive products.

Some products may be heat sensitive only above a threshold temperature; for those that can withstand temperatures in the range of 110–118°C but cannot withstand 121°C it is possible to apply the F_0 concept to the principles above and derive equivalent sterilization specifications to those given in Table 2. These specifications are summarized for 116°C in Table 3. As can be seen, if there is a requirement to sterilize at (say) 116°C, there are considerable time savings to be obtained by characterization of the contaminating micro-organisms.

Pharmaceutical Products and Materials for Aseptic Manufacture—Sterilizer Parameters

Sterilizer parameters are specific to combinations of product, presentation, and autoclave. They must be established empirically. Heat penetration studies done prior to the performance qualification phase of validation serve the purpose of determining the loading patterns, prevacuums, and temperature and pressure settings, etc. which ensure that the sterilization specification is delivered to the product and that it is delivered uniformly throughout the load.

For instance, a particular proposed loading pattern may never allow for uniform conditions (within specified limits) to be achieved throughout the load. In this case the pattern would have to be changed. Or, in a particular autoclave it may be necessary to set the temperature at 122°C for 121°C to be achieved within the load.

Air removal is particularly important in porous and equipment loads, but is usually of little importance in the sterilization of aqueous pharmaceutical products. Air removal can be important to the specification of new

Table 2 The range of sterilization specifications at 121°C calculable according to various approaches

	“Holding” time at 121°C (min)	
	Shortest possibility	Longest possibility
“Overkill” <i>PhEur</i>	15	Not specified
“Overkill” <i>USP</i>	12	Not specified
With bioburden ^a data only	6	8
With bioburden ^a and thermal resistance ^b data	1.8	2.4

^a Assuming a “worst case” D_{121} -value of 1 min vs. bioburdens of one micro-organism (least) to 100 micro-organisms (most) per item.

^b Calculated for a D_{121} -value of 0.3 min vs. bioburdens of one micro-organism (least) to 100 micro-organisms (most) per item.

Table 3 The range of sterilization specifications at 116°C calculable according to various approaches

	“Holding” time at 116°C (min)	
	Shortest possibility	Longest possibility
“Overkill” <i>PhEur</i>	48	Not specified
“Overkill” <i>USP</i>	38	Not specified
With bioburden ^a data only	19	26
With bioburden ^a and thermal resistance ^b data	5.7	7.6

^a Assuming a “worst case” D_{121} -value of 1 min vs. bioburdens of one micro-organism (least) to 100 micro-organisms (most) per item. For equivalence a Z-value of 10 K has been used.

^b Calculated for a D_{121} -value of 0.3 min vs. bioburdens of one micro-organism (least) to 100 micro-organisms (most) per item. For equivalence a Z-value of 10 K has been used.

autoclaves—those which are to be designated only for aqueous product sterilization have no need for the pumps and ancillary equipment required to pull deep vacuums.

The involvement of steam in the sterilization of different types of product is an important consideration in understanding and controlling autoclaves.

For aqueous products, steam is solely a means of raising the product to the specified sterilizing temperature; the steam does not come into contact with the contaminating micro-organisms. The transfer of heat energy (lethality) to the contaminating micro-organisms is from the product itself. To all intents and purposes any suitable form of energy source could be used to raise the temperature of the product. For instance, if ampules of aqueous products were to be sterilized in a hot air oven, the mechanisms of microbial inactivation would still be by coagulation of intracellular proteins. However, heat transfer from hot air is much slower than heat transfer from steam, which is why this is not seen as a practical process. Microwave irradiation could be an alternative means of sterilizing aqueous pharmaceutical products utilizing the same anti-microbial mechanisms as steam; certainly there is evidence that microwave killing patterns are mainly due to heat transfer with very little direct energy being absorbed from the microwaves.^[8,9]

For porous and equipment loads, the steam comes into direct contact with the contaminating micro-organisms on the materials being sterilized and there is no intermediary in the transfer of heat. The energy content of steam is defined by its latent heat. If the steam is pure in the sense that it contains neither entrained gas nor moisture, an amount of energy defined by its latent heat at the pressure of the steam will be transferred to the micro-organisms by condensation on their surfaces.

There are many potential pitfalls in equipment and porous load sterilization, mainly concerned with air or other noncondensable gas. First, the purity of the steam is important; if it is carrying moisture, or noncondensable gas,

it will not contain the same amount of energy as pure steam and its lethality will be less than that predicted for pure steam. Second, any residual air around the contaminating micro-organisms may insulate them from contact with the steam and thus reduce the amount of energy (lethality) transferred. In this type of sterilization, steam quality becomes very important and so also do the materials and manner in which the products are contained in steam-permeable wrapping or perforated trays, etc., within the autoclave, and the number and depth of evacuations of the autoclave prior to the temperature-hold phases. Thermal monitoring alone gives little information on the adequacy of the measures put in place to control these complex factors, and it is therefore generally thought essential that some empirical studies be done with BIs as part of process development to ensure that the thermal lethality being imparted by the steam is not being impeded. These development studies may be rolled into bio-validation.

Sterilization of Microbiological Media in the Laboratory

The various suppliers of microbiological media include recommendations for sterilization in their catalog under “Directions for Use,” for instance, “sterilize by autoclaving at 121°C for 15–18 min.” The question that must be asked is—What do these specifications mean?

Are they intended to apply within the media as are the sterilization specifications for pharmaceutical products and materials for aseptic manufacture? Or are they sterilizer parameters? There may be some indication in some of the older suppliers’ manuals which expand their recommendations along the lines of “sterilize by autoclaving at 15 psi (121°C) for 15–18 min.” Since pressures of 15 psi are not achievable within media, it is clear that the intention was that the recommendations be applied to sterilizer parameters.

In most cases, it is probably immaterial how these recommendations are interpreted. For media, “overcooking” is bad because of deleterious effects on growth-support characteristics, and “undercooking” is generally self-disclosing through evident contamination.

SIP Systems

Systems that are sterilized in-place are often immensely complex. The initial challenges to their sterilization are the removal of air and the elevation of the temperature of the pipework to prevent heat losses and condensation. As such, most work in the development of sterilization specifications for SIP systems is concerned with the heat-up phase. Appropriate questions are: Is the sterilization temperature achieved throughout the system? Where is the slowest location to achieve temperature? Where should the control probe be located?

Often vast amounts of thermal lethality calculated as F_0 -values are delivered in these prehold stages of SIP. However, because these temperatures are being achieved in the presence of steam–air mixtures, it is not correct to assume that the biological lethality during the heat-up phase of SIP systems is equivalent to that achieved with pure steam.

The time for which the system must be held at temperature (the sterilization specification) is often relegated to a minor consideration compared with this earlier development work. Typically, it is decided arbitrarily to use, 121°C for 15, 20, or 30 min, with no real scientific basis.

Perhaps, a basis parallel to that of the pharmacopeial overkill specifications could be developed. For instance, if the actual maximum number of micro-organisms within an SIP system is assumed to be 10^{12} (since this would amount to a few grams of biomass it certainly should be maximal), then 18 log reductions would be required to ensure not more than one chance in a million of a survivor. An overkill cycle of 121°C for 20 min could be proposed by adding two log reductions as a safety factor and assuming each micro-organism to have a D_{121} -value of 1 min.

BIO-VALIDATION

The performance qualification (PQ) phase of validation follows the development of the sterilization specifications and of the sterilizer parameters which will deliver them. The purpose of PQ in steam sterilization of pharmaceutical products, equipment, laboratory media, and SIP systems is to confirm that the sterilization specification consistently achieves its intended purpose. The process is run using the parameters derived from process development on (usually)

three separate occasions and tested for compliance with a variety of predetermined acceptance criteria. As a subset of PQ, the purpose of bio-validation is to confirm that the lethality expected from the process does not significantly deviate from what is expected. Bio-validation is a “test” of consistency. If the acceptance criteria are not achieved, there may be need for more process development.

In consideration of the extent, thoroughness, and history of the research evidence that micro-organisms are inactivated in a regular fashion in response to temperature and time, it is periodically suggested that bio-validation should not be necessary where there is evidence of adequate heat penetration. In practice, however, the expected lethality may not always be achieved. Most frequently, such deviations from ideality occur in equipment and porous load sterilization because of inadequate air removal. Where deviations from ideality occur for aqueous pharmaceutical products, they most likely arise from inadequate knowledge of how the product affects the thermal resistances of micro-organisms, but this is best determined in the laboratory at an earlier stage of process development, not at the bio-validation “milestone” later in the critical path of product introduction.

Acceptance criteria for bio-validation of steam sterilization processes are usually (but not invariably) defined along the following lines:

- n BIs will be placed in the load at locations defined in a drawing.
- Each BI will contain at least 10^6 viable spores of *B. stearothermophilus*.
- The load will be exposed to a defined autoclave treatment (the validation cycle).
- Bio-validation will be considered satisfactory if no viable spores are recovered from the BIs after x days of incubation at 55–60°C.

Because this approach is the common practice, there is a widely held belief within the pharmaceutical QA community that the ability to inactivate 10^6 spores of *B. stearothermophilus* is a synonym for achieving an SAL of 10^{-6} . It is not. It is true, however, that inactivation of 10^6 spores of *B. stearothermophilus* with the pharmacopeially approved minimum D -value of 1.5 min in 10–100 replicates guarantees achievement of better than 10^{-6} SALs for worst case bioburdens (Table 4). However, Table 4 also shows that the converse, i.e., failing to inactivate 10^6 spores of *B. stearothermophilus*, does not necessarily mean that a 10^{-6} SAL has not been achieved.

Another area of confusion is that the USP definition of an overkill specification—“a lethality input of $12D$ ”—can be demonstrated directly in bio-validation. It should be

Table 4 Sterility assurance levels indicated by inactivation of 10^6 spores of *B. stearothermophilus*

	Spore <i>D</i> value			
	1.5 min	2 min	3 min	4 min
Bioburden of 10^2 micro-organisms per item each with <i>D</i> -values of 1 min	10^{-10}	10^{-14}	10^{-22}	10^{-30}
Bioburden of 10^2 micro-organisms per item each with <i>D</i> -values of 0.3 min	10^{-38}	10^{-51}	10^{-78}	10^{-104}
Bioburden of 1 micro-organism per item each with <i>D</i> -values of 1 min	10^{-12}	10^{-16}	10^{-24}	10^{-32}
Bioburden of 1 micro-organism per item each with <i>D</i> -values of 0.3 min	10^{-40}	10^{-53}	10^{-80}	10^{-106}

Inactivation of 10^6 spores in 10–100 replicates is assumed to be equivalent to 8 log inactivations.

understood that the maximum number of log inactivations of any bacterial population is technically limited to about 9 or 10 *D*-values. The maximum number of micro-organisms that can be handled as a BI is about 10^7 – 10^8 , the sensitivity of recovery of micro-organisms is restricted to more than 10^{-2} . An indirect demonstration of 12 log inactivations of a micro-organism with a *D*-value of 1 min can be achieved by showing inactivation of 10–100 replicate BIs each carrying 10^6 spores with *D*-values of 1.5 min, or by inactivation of 10–100 replicate BIs each carrying 10^4 spores with *D*-values of 2 min. Direct demonstration of 12*D* is technically impossible.

In bio-validation, the spore of *B. stearothermophilus* is akin to an end-point analytical reagent. For instance, when litmus changes from blue to red at pH levels below 7, it shows only that the pH is not higher than 7. By killing all of 10–100 replicate BIs with 10^6 spores having *D*-value 1.5 min, all that is proven is that the thermal lethality delivered is not less than an F_0 of 12 min. The *PhEur* overkill sterilization specification of 121°C for 15 min should meet this requirement easily, and so should any other longer specification at 121°C, or any specification for longer times at lower temperatures taking into account of the F_0 concept.

Numbers and Locations of BIs for Bio-validation

It is usual for bio-validation to be done with an arbitrary number of BIs between 10 and 100. Both limits are based on practical considerations.

The lower number of BIs is defined in terms of ensuring that bio-validation addresses sufficient parts of the load for confidence that items in all parts of the autoclave are receiving the required lethality. Normal practice is to

define this number in terms of placing at least as many BIs as the number of thermal probes used for thermal qualification. It is sensible to place one BI alongside each thermal probe in order to be able to relate thermal data to biological data. In addition to this, some BIs should be placed in other nonprobed locations in consideration of the possibility that the leads to the thermal probes may be acting as conduits for air removal or steam penetration, and thus provide falsely high levels of lethality.

More often than not the number of BIs used is about 20–30. Larger numbers up to 100 may be necessary to address very large autoclaves or in thermal mapping studies, but in validation there is little extra statistical confidence to be gained by doing so. In most microbiology QA laboratories, 20–30 BIs can be handled conveniently.

Periodically in the bio-validation of sterilization of porous or equipment “minimum” loads, it is not practical to locate 20 or 30 BIs. For instance, a minimum load may be one cartridge filter, one mop head, or one machine manifold. In such cases, it is important to avoid too much distortion of the statistics of bio-validation. At least five BIs are recommended no matter how difficult it may be to place them.

Choice of BIs

Spores of *B. stearothermophilus* are most commonly used for bio-validation of steam sterilization processes. This is not to say that it is mandatory to use *B. stearothermophilus* nor that it is used exclusively. Other micro-organisms, e.g., *sporogenes*, are used by some companies and accepted by the regulatory agencies. Use of *B. subtilis* spores with resistances to steam sterilization in the higher range of that found in natural bioburden has, in recent years, been criticized by European regulatory agencies.

The principles underlying the choice of micro-organism used as BIs are quite well known:

- The micro-organisms must have high resistance to the sterilization treatment which they are being used to validate. This does not mean that they must be the most resistant micro-organism known to man. *B. stearothermophilus* has D_{121} -values of 1–4 min according to conditions of culture and the substrate upon which they are mounted. This is higher than most spores of *Bacillus* spp., which tend to have D_{121} -values below 0.5 min.
- The micro-organisms must have stable resistances to the sterilization treatment which they are being used to validate. There are data from commercial suppliers of BIs to show that spores of *B. stearothermophilus* survive and retain stable resistances over long periods of crudely controlled storage.
- The micro-organisms must be easily culturable and preferably be easily identifiable in culture. Very few micro-organisms share with *B. stearothermophilus* the ability to grow in simple culture media at 55–60°C.

It is customary to use 10^6 (in practical terms 10^5 – 10^7) spores per BI. This number is based on custom, practice, and convenience rather than on science. Larger numbers than this are difficult to handle in culture and result in large errors in counting. Smaller numbers reduce the sensitivity of the test. Unfortunately, the widespread use of 10^6 spores for bio-validation has (as described before) led to a confusion between 10^{-6} SALs and 6 log reductions of *B. stearothermophilus*.

More complex decisions surround the choice of substrate within which spores are suspended or on which they are mounted for use as BIs.

The decision tree shown in Fig. 3 may be used to help choose the spore substrate used in bio-validation of aqueous pharmaceutical products. To use this decision tree, it is essential to have some knowledge of the effects of product on the resistance of spores; as mentioned before, this requires special equipment and experience.

In all circumstances water is the preferred substrate for bio-validation of aqueous pharmaceutical products. Where water would give deceptive results, it should not be used.

- If the D_{121} -value of *B. stearothermophilus* is higher in the product than it would be in water (i.e., the product makes the spores more resistant to steam sterilization), then bio-validation must be done with the spores suspended in product. Otherwise falsely favorable results may occur.
- If the D_{121} -value in product of *B. stearothermophilus* is equal to or less than its D_{121} -value in water, and the

sterilization specification is based on overkill, then bio-validation must be done with spores suspended in water. However, if the sterilization specification has been “tailored” specifically to the resistance of micro-organisms in the product, then bio-validation must be done with the spores suspended in product. Otherwise falsely unfavorable results may occur.

Under no circumstances must spores be suspended in product if that product is sporicidal.

For other materials (for instance, equipment and supplies for aseptic manufacture) where the mechanisms of inactivation rely on direct contact between the steam and the item being sterilized and therefore where air removal is matter of importance, the choice of substrate for BIs generally lies between using commercially available paper spore strips and the material itself. The decision tree in Fig. 4 may be helpful. Regulatory pressure is currently toward use of inoculated product, but commercially available spore strips are more convenient. “Tailor-made” inoculated product requires substantial amounts of microbiological expertise. The decision tree in Fig. 4 may be helpful in selecting which approach is best in particular circumstances.

Use of commercially available BIs transfers much of the responsibility for assuring quality in manufacture to the supplier. Regardless of this, their quality must be controlled on receipt and prior to use in bio-validation. There is no reason why any microbiology laboratory should not verify the numbers of spores per commercial BI. On the other hand, the determination of resistance requires special equipment and expertise and is probably best accepted on the basis of the supplier’s certification. If this is done, the user of the BIs is responsible for knowing what the certified measures of resistance mean, how they were determined (on the strips, in aqueous suspension, or in or on something else), and that they were determined correctly and in compliance with applicable standards and legislation.

Validation Cycle

Bio-validation is usually done against a sterilization specification which delivers less lethality than the lower limit of lethality allowed by the sterilization specification defined for the material being sterilized. It is clearly intellectually flawed to choose to validate something different to that which is ever to be used in practice. So what is the reasoning behind this practice?

Sterilization specifications in the “hold” period are presented in terms of temperature and time with upper and

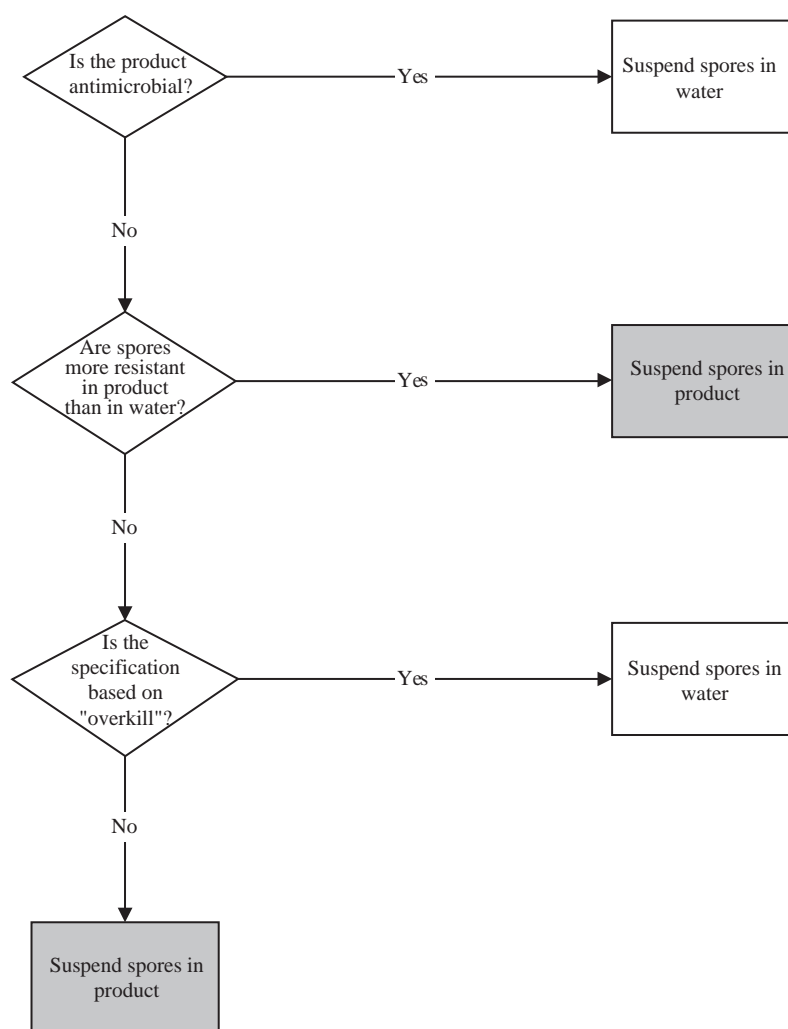


Fig. 3 Recommended substrates for BIs used in bio-validation of aqueous fluid loads.

lower tolerances set around them. The lower specification limits are critical to sterilization.

Time is generally easily controllable to quite high levels of accuracy and precision: A steam valve allows steam to enter the autoclave until the hold temperature is reached; the valve is then closed and the process is controlled by a timer which, at the end of the specified hold period, sends a control signal to activate the exhaust valves and cooling sequences. The hold time is usually specified in terms of whole minutes—well within the accuracy and precision of all but the most inappropriate of timers.

Temperature is less easy to control precisely. The temperature in the hold period in autoclaves is generally maintained by modulating valves which open to allow steam entry when the temperature (or pressure, because these valves are more often than not controlled through pressure transducers) begins to drop toward the critical lower limit

of the specification. It is generally not possible to control an autoclave to run through a complete hold period at the lower limit of its temperature specification. However, even quite apparently trivial errors in temperature above or below the limit can make significant differences in the amount of lethality delivered. For instance, at a nominal temperature of 121°C, an error of 1 K can increase or decrease the amount of lethality by 25%.

Bio-validation cycles are therefore designed to ensure that no more lethality is delivered than that specified by the lower limits of lethality of the sterilization specification used in routine practice. Ideally this is done by reducing the time of the hold period, but sometimes, when quite short cycles are being used, it may also be necessary to reduce the temperature set point on the autoclave as well. The risk in all of this is that the bio-validation cycle is used as a justification for the release of sterilized items when

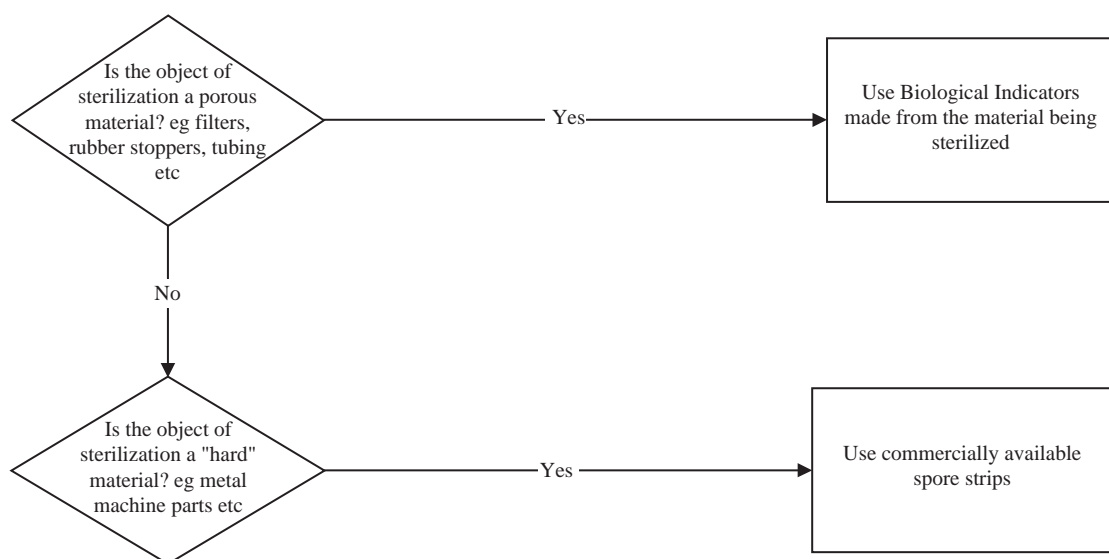


Fig. 4 Recommended substrates for BIs used in bio-validation of porous loads.

specifications are not complied with under atypical production conditions. This idea should not be entertained.

Acceptance Criteria

Bio-validation is a limit test, which at best produces quantal data. Each BI should be tested separately for survivors or absence of survivors. The acceptance criterion should be that there are no survivors.

Detection of survivors in all exposed BIs is clearly unacceptable; such a result would be obtainable if the BIs had never been near a sterilizer.

Data showing survival on some but not all of the BIs may be valuable in process development (particularly in the development of sterilization specifications and sterilizer parameters for porous and equipment loads), but would likely raise issues at regulatory inspection. The inference taken from having some survival could be that each item in the autoclave load is not being exposed to the same treatment.

For initial validation of a newly developed process, complete inactivation of all BIs should not be difficult to achieve.

The range of D_{121} -values acceptable to USP for spores of *B. stearothermophilus* allowed as BIs for use in steam sterilization is 1.5–3.0 min. An overkill sterilization or bio-validation specification delivering an F_0 of 15 min would deliver 10 log inactivations or, if there were 10^6 spores per BI, one chance in one million of finding a survivor on any one BI, one chance in one thousand of finding a survivor in 10 BIs, one chance in one hundred of finding a survivor in 100 BIs, and so on. In other words,

there are pretty long odds against failing the acceptance criteria.

However, spores of *B. stearothermophilus* may have D_{121} -values of 3 min. In such a case there would be practically no chance of meeting the acceptance criteria of killing 10^6 spores with an F_0 of 15 min. What are the implications of this?

On the face of it, the implication is that this sterilization specification/sterilizer parameters combination is invalid. However, remember that this same sterilization specification/sterilizer parameter combination would have been valid if the BIs used had D_{121} -values of 1.5 min. In practical terms the pharmacopeial specification for thermal resistance in BIs has been set naively. Many companies purchasing spores of *B. stearothermophilus*, either for preparing BIs or as commercial strips, order against their own specifications which include upper D_{121} -value limits (in water) of around 2 or 2.5 min. The author of this paper has published recommendations^[10] for determining bio-validation cycles appropriate to challenge numbers and D_{121} -values of the spores available, but in the long run it is far more convenient and easier to justify compliance to regulatory agencies when there is bio-validation to show that 10^6 spores have been inactivated in 10–100 replicates.

Requalification

Periodically it is wise to repeat bio-validation. Changes do occur in autoclaves and no change control procedure, no matter how rigorously implemented, is infallible. The

purpose of requalification is to determine if any unforeseen change has arisen to affect the sterility assurance provided to the items being sterilized.

It is important for requalification that the numbers, resistances, and substrates for the BIs are closely similar to those used in the initial validation. For the same reasons as resistance variation within BIs as discussed before, it is possible if these factors are not well controlled to emerge from requalification with either a false confidence in the security of the process (use of BIs which are less resistant than those used in initial validation), or with the incorrect opinion that the process has failed (use of BIs which are more resistant than those used in initial validation).

Biological requalification is usually done following significant process changes or on an annual frequency. The establishment of a frequency should, in principle, be based on business risk; in fact, however, with well-designed and controlled autoclaves, the risk to the business of extending the interval beyond 1 yr is probably more one of incurring regulatory criticism at inspection than of releasing nonsterile products to market.

Bio-validation of Laboratory Autoclaves Used for Sterilizing Microbiological Media

It is not difficult to argue that the effectiveness of sterilizing microbiological media is self-disclosing and should not therefore merit bio-validation. The pertinence of bio-validation to the qualification of laboratory autoclaves is more to do with having confidence in the sterility of media before or after it is used in the laboratory (and risk false positive results if the media is not sterile), or take it into (say) aseptic manufacturing areas for environmental monitoring (and if it is nonsterile contaminate areas and products which may otherwise have been secure). Many regulatory bodies see bio-validation of laboratory autoclaves as mandatory.

Bio-validation of Steam-in-Place (SIP) Systems

SIP systems range from very small systems (say, a mixing tank) where all parts may reach temperature within 2 or 3 min, to absolutely massive arrangements of vessels, valves, and pipework in which the expulsion of air, the drainage of condensate, and the attainment of the

sterilization specification temperature at the “slowest point” can take 20 or 30 min.

Bio-validation is essential. The placement of BIs is largely a matter of judgment. Certainly the “slowest point” to reach the sterilization specification temperature must be challenged. Certainly vent filters and low points where condensate could accumulate must be challenged. Thereafter, it is a regulatory expectation that there should be sufficient BIs placed to give coverage to the whole system, which in effect may mean placing BIs in locations which, due to the heat-up time of the system, have been exposed to the sterilization specification temperature for two, three, or four times longer than the “slowest point.” Undue confidence should not be taken from favorable results from these locations.

REFERENCES

1. Kelsey, J.C. The myth of surgical sterility. *Lancet* **1972**, 1301–1303.
2. Savage, R.M. Sterility tests on surgical dressings. *Quart. J. Pharm. Pharmacol.* **1940**, *13*, 237–254.
3. Bryce, D.M. Tests for the sterility of pharmaceutical preparations. *J. Pharm. Pharmacol.* **1956**, *8*, 561–572.
4. Savage, R.M. Interpreting the Results of Sterility Tests. In *Recent Developments in the Sterilization of Surgical Materials*; The Pharmaceutical Press: London, 1961.
5. Ernst, R.R.; West, K.L.; Doyle, J.E. Problem areas in sterility testing. *Bull. Parent. Drug Assoc.* **1969**, *23*, 29–38.
6. Brown, M.R.W.; Gilbert, P. Increasing the probability of sterility of medicinal products. *J. Pharm. Pharmacol.* **1977**, *29*, 517–523.
7. Berger, T.J.; Chavez, C.; Tew, R.D.; Navasca, F.T.; Ostrow, D.H. Biological indicator comparative analysis in various product formulations and closure sites. *PDA J. Pharmaceut. Sci. Technol.* **2000**, *54* (2), 101–109.
8. Yeo, C.B.A.; Watson, I.A.; Stewart-Tull, D.E.S.; Koh, V.H.H. Heat transfer analysis of *Staphylococcus aureus* on stainless steel with microwave radiation. *J. Appl. Microbiol.* **1999**, *87*, 396–401.
9. Watanabe, K.; Kakita, Y.; Kashige, N.; Miake, F.; Tsukiji, T. Effect of ionic strength on the inactivation of microorganisms by microwave irradiation. *Lett. Appl. Microbiol.* **2000**, *31*, 52–56.
10. Halls, N.A. Resistance Creep of Biological Indicators. In *Sterilization of Medical Products*; Morrissey, R.F., Kowalski, J.B., Eds.; Polysciences Publications Inc.: Champlain, NY, 1998; Vol. VII.

AUTOXIDATION AND ANTIOXIDANTS

John M. Pezzuto

Eun Jung Park

University of Illinois at Chicago, Chicago, Illinois, U.S.A.

INTRODUCTION

By definition, oxygen is an absolute requirement for aerobic life, but it may also be viewed as toxic under certain conditions. A chemical reaction that usually takes place at ambient temperature between atmospheric oxygen and an organic compound is generally defined as autoxidation. The phenomena of autoxidation are commonly observed in everyday life. For example, the browning of fruit, deterioration of edible oils, and degeneration of old rubber bands are the results of autoxidation. In the human body, reactive oxygen species (ROS) produced during autoxidation processes are also commonly encountered. This is a normal part of human physiology, as long as defense systems are effective, but overproduction or failure to scavenge free radicals can result in toxic biological responses that can yield deterioration and degeneration of cells, tissues, or organisms.

There are numerous reports suggesting that free radicals are involved in various human disease states (Table 1). Pathological phenomena related to autoxidation are due to reactive oxygen species, and most age-related diseases can be explained through reactions of free radicals with biological substances. A notable case is cancer or, more specifically, carcinogenesis. Free radicals play a critical role in the initiation of carcinogenesis by damaging nucleic acids and producing a variety of lesions. For example, 8-hydroxyguanine, 5-hydroxymethyluracil, and thymine glycol are formed by the attack of hydroxyl free radicals on deoxyribonucleic acid (DNA), and these mutations may be viewed as a cause of carcinogenesis. Another free radical-related disease state is stroke. Damage to the brain due to central nervous system (CNS) ischemia is caused by injury resulting from the interruption of blood flow (i.e., lack of oxygenation) which is then followed by reoxygenation of the brain (ischemia/reperfusion). It appears that all or most aerobic tissue suffers damage once it undergoes an ischemia/reperfusion insult. The severity depends on many factors, one of which is the length of the ischemic period. Considerable evidence is now accumulating that injury occurs almost exclusively during the reperfusion

phase, and that the injury is due to oxygen free radical-mediated oxidative stress. Furthermore, some neuronal pathologies, such as Alzheimer's disease, may relate to lipid peroxidation of cell membranes by free radicals.

Autoxidation phenomena could be completely prevented by total exclusion of oxygen or other oxidizing substances from a biological system. This is generally not practical, but changes in endogenous factors, such as addition of inhibitors, may decrease the reaction rate or prolong the induction period. However complete prevention of autoxidation is unlikely. Substances that can suppress autoxidation are termed inhibitors or antioxidants. Preventive inhibitors decrease the rate of autoxidation by suppressing the rate of initiation reactions. Antioxidants in the true sense are substances that can inhibit propagation steps; that is, they interrupt autoxidation chain reactions. Other types of inhibitors, for instance, antiozonants, are not biologically relevant but may be important in industry (e.g., protective coatings). Antioxidants are known to inhibit carcinogenesis or atherosclerosis. For example, vitamins and phenolic compounds can function as chemopreventive agents. The mechanism of action and the biological role of autoxidation and antioxidants are discussed in this article.

AUTOXIDATION

Autoxidation Reactions

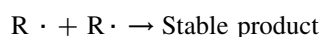
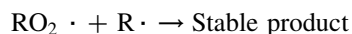
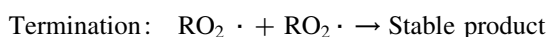
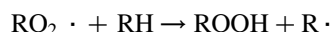
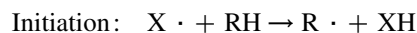
The majority of compounds that are subject to autoxidation are unsaturated substances or highly condensed polymers, and the target of ROS is commonly the diene functionality—or double bonds. Autoxidation of unsaturated lipids is a good model to demonstrate this mechanism. The initial reaction between molecular oxygen and a polyunsaturated fatty acid (PUFA) occurs as $\text{RH} + \text{O}_2 \rightarrow \text{ROOH}$, which involves the movement of a double bond as well as the insertion of oxygen. In general, there is a “spin barrier” that prevents the direct addition of ground-state molecular oxygen in a single step to an organic compound. In the case of autoxidation, since direct addition is eliminated by the

Table 1 Structural and functional alterations of biological entities produced by radicals or ROS

Structure alterations	Functional alterations	Disease
Lipid peroxidation	Increased permeability of membrane	Cancer
Protein adducts	Impaired response of membrane to signals	Stroke
Hemoglobin adducts	Altered enzyme activity	Alzheimer's disease
	Reduced affinity of binding proteins	Rheumatoid arthritis
	Altered uptake of oxidized entities	Cataract

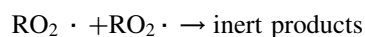
spin barrier, the alternatives are i) electron transfer (i.e., redox reactions) involving, for example, a transition metal ion, or ii) the participation of free radicals, whereby the addition of a radical to molecular oxygen can give rise to another radical (1).

The important features of autoxidation are autocatalytic and free radical chain reactions. The rate of oxidation is initially slow and increases as the reaction progresses. However, once autoxidation is initiated, the reaction continues until the reaction substrate or catalytic factor becomes extinct. In short, unsaturated lipids undergo three reaction phases: initiation, propagation, and termination. The participation of reactive oxygen radicals in autoxidation reactions is summarized in the following reaction steps.



Initiation is perhaps the process most difficult to define. This is because of the very low concentration of radicals and the likelihood of there being more than one process, since a large number of different radicals can abstract hydrogen from RH to form $R \cdot$. For example, $X \cdot$ may be a transition-metal ion, a radical generated by photolysis or high-energy irradiation, a radical obtained by decomposition of a hydroperoxide (e.g., $RO \cdot$), or a radical formed from an exogenous initiator. The two propagation reactions form the basis of the chain-reaction process. The autoxidation reaction is generally assumed to be a very fast reaction with almost no activation energy. A major consequence of this phase is that the concentration of $R \cdot$ is much smaller than that of $RO_2 \cdot$. Like initiation processes, termination reactions may also be divided into

those involving organic free radicals exclusively, and those in which metal ions participate. Of the termination processes involving $R \cdot$ or $RO_2 \cdot$, the biomolecular reaction



has received the most attention since this is likely to be the most important termination reaction under physiological conditions.

Autoxidation of unsaturated lipids is affected by many factors, so that none can be considered exclusively prooxidative or antioxidative. The rate of autoxidation is increased with reactivity of the autoxidizing substrate, concentration of reactants (the number of active sites and the concentration of oxygen), modified physical factors (e.g., a rise in temperature or by irradiation), and especially with an increase in the rate of initiation reactions (1). The rate of chain initiation reactions is increased mainly by factors that increase free radical function in autoxidizing systems, for example, ultraviolet, X-Ray, or ionizing radiation. In various reactant systems, such as fats, oils, and biological membranes, heavy metals and their derivatives are important initiators of autoxidation reactions. The reaction rate, however, is suppressed by factors such as a reduction in the number of active sites, a decrease in the partial pressure of oxygen by a suitable selection of physical factors, and a decrease in the reaction temperature. The most important, however, is the reduction of initiation rate (i.e., the level of free radicals capable of chain initiation). Substances actively suppressing the concentration of the free radicals are antioxidants. Other substances are also very effective in suppressing autoxidation, particularly those eliminating free radical precursors (e.g., sulfur compounds that cause reduction of lipid hydroperoxide) (2). Substances deactivating ozone (antiozonants) or singlet oxygen, which protect against irradiation, or heavy metals (metal scavengers), also act as inhibitors of lipid autoxidation.

Lipid Peroxidation

Oxidative damage to membrane lipids (i.e., lipid peroxidation in biological systems) has been studied for many years. Lipid peroxidation is a primary event produced by oxidative stress or as a consequence of tissue damage, which can exacerbate tissue injury, due to the potential cytotoxicity and genotoxicity of the end products of lipid peroxidation. Membrane lipids with double bonds are most susceptible to oxidation. Lipid peroxidation can reduce membrane fluidity, leading to increased rigidity throughout the hydrophobic space of membranes, decreased permeability, osmotic fragility, and altered activity of certain membrane-bound enzymes and transport systems (3).

Membrane lipid peroxidation can change the activity of essential membrane proteins such as Na^+/K^+ -ATPase. As a consequence, rates of ion pumping may be altered (4). Two well-characterized products of lipid peroxidation, malondialdehyde and 4-hydroxynonenal, have been shown to react with critical biomolecules that may have a key role in the development of certain pathological states. Microsomal lipid peroxidation forms mainly 4-hydroxy-2-nonenal with minor amounts of 4-hydroxy-2-octenal, 4-hydroxy-2-decenal, and 4-hydroxy-2-undecenal. α,β -Unsaturated aldehydes, such as 4-hydroxynonenal, are highly reactive electrophilic reagents that react easily with thiols by Michael addition to the $\text{CH}=\text{CH}$ double bond (1,5). Aldehyde adducts of lipid peroxidation have been shown to induce glutathione depletion in hepatocytes, and corresponding abnormal liver function (6, 7). Free hemoglobin, in the presence of xanthine/xanthine oxidase, will also promote the peroxidation of arachidonic acid and unsaturated fatty acids within normal cellular membranes. Furthermore, hemoglobin or red cell lysates cause brisk peroxidation of crude murine brain homogenates. This hemoglobin-driven peroxidation is blocked by an iron chelator (i.e., desferrioxamine) which indicates that iron released from heme is responsible (8).

Peroxides can play a physiological role in the cell but also mediate processes leading to heart disease and carcinogenesis (9–11). Fatty acid peroxidation may also be related to free radical-mediated metabolic activation of carcinogens or drugs, which lead to the initiation of carcinogenesis or cytotoxicity. Modification of membrane function as a consequence of lipid peroxidation includes uncoupling of oxidative phosphorylation in mitochondria, alteration of liver endoplasmic reticulum functions, and modification of the ionic permeability of phospholipid membranes. Radiation-catalyzed lipid peroxidation of erythrocyte membranes increases

permeability to Na^+ , K^+ , and Ca^{++} . Ultimately, peroxidation leads to gross destruction of membranes, as demonstrated by release of lysosomal enzymes in irradiated lysosomes, hemolysis in irradiated erythrocytes, and release of various enzymes by disruption of the liver plasma membrane.

FREE RADICALS AND DEFENSE SYSTEMS

Free Radicals and Defense Systems

A free radical may be broadly defined as a molecule or ion containing an unpaired electron (Table 2). Although most radicals are reactive and undergo dimerization or other reactions in which the unpaired electron becomes paired some radicals such as nitroxide radicals ($\text{R}_2\text{NO}\cdot$) are relatively stable. A major source of radicals in biological systems is molecular oxygen. This very reactive molecule is essential to the life of higher organisms, but nevertheless can be considered dangerous in excess.

Superoxide anion ($\text{O}_2^{\cdot-}$) and hydrogen peroxide (H_2O_2) are normal metabolites in mammalian cells produced during the biological reduction of oxygen. The occurrence of oxidative radical reactions in biological systems is usually associated with cellular electron transfer chains of the mitochondria and certain enzyme activities. Free radicals are further generated during the course of specialized physiological reactions, such as the release of $\text{O}_2^{\cdot-}$ and H_2O_2 by endothelium noninflammatory cells, which might serve as cell signals promoting growth responses. Furthermore, environmental factors as well as the metabolism of xenobiotics are significant sources of free radicals, although their actual contribution to the redox state of the cell is difficult to assess.

Although of modest chemical reactivity, $\text{O}_2^{\cdot-}$ and H_2O_2 contribute to the formation of more reactive species via various redox reactions. It is currently accepted that the reaction of both $\text{O}_2^{\cdot-}$ and H_2O_2 with suitable metal complexes yields $\text{HO}\cdot$. Fenton-type reactions requires both $\text{O}_2^{\cdot-}$ and H_2O_2 as precursors of $\text{HO}\cdot$. This proceeds via an intermediate catalyst, such as a transition metal chelate, which reacts to produce $\text{HO}\cdot$ (12). $\text{HO}\cdot$ may account for some aspects of mitochondrial damage, such as oxidative impairment of mitochondrial DNA. The requisite conditions for $\text{HO}\cdot$ formation by mitochondria are met during mitochondrial electron transfer: mitochondria are a major source of $\text{HO}\cdot$ generated by H_2O_2 cleavage.

Haber – Weiss reactions: $\text{Fe}^{3+} + \text{O}_2^{\cdot-} \rightarrow \text{Fe}^{2+} + \text{H}_2\text{O}_2$

Fenton reaction: $\text{Fe}^{2+} + \text{H}_2\text{O}_2 \rightarrow \text{Fe}^{3+} + \text{HO}\cdot + \text{HO}^-$

Table 2 Principal reactive oxygen intermediates formed by cells

ROS	Function
Superoxide anion, $O_2^{\cdot-}$	One-electron reduction products of O_2 , produced by phagocytes; formed in autoxidation reactions; generated by oxidases (heme protein).
Hydrogen peroxide, H_2O_2	Two-electron reduction product of O_2 ; formed from $O_2^{\cdot-}$ by dismutation or directly from O_2 . Reactivity of $O_2^{\cdot-}$ and H_2O_2 amplified in the presence of heme protein.
Hydroxyl radical, $HO\cdot$	Three-electron reduction product of O_2 generated by Fenton reaction; transition metal (iron, copper)-catalyzed Haber–Weiss reaction; formed by decomposition of peroxynitrite produced by the reaction of $O_2^{\cdot-}$ with $NO\cdot$.
Nitric oxide, $NO\cdot$	Generated from arginine via NO synthase in an O_2 -dependent reaction; in the absence of O_2 , formed by Fe^{2+} -mediated nonenzymatic generation from tissue NO_2^- .

(Adapted in part from Ref. 15.)

Another type of one-electron transfer reaction that contributes to the formation of oxyradicals involves the quenching of carbon center radicals ($R\cdot$) by molecular oxygen. This reaction leads to the formation of peroxy radicals ($R-OO\cdot$), which generally have quite different reactivities from those of the parent $R\cdot$ species. As a result of the reactivity of these species toward unsaturated fatty acids, the propagation steps of lipid peroxidation follow the initiation step. These propagation steps occur at membrane hydrophobic sites, and the length of the chain reaction is determined by the availability of reactants, PUFA and O_2 , and of chain-breaking antioxidants such as α -tocopherol, carotenoids, and ubiquinone.

In cells, the first line of defense against these reactive species is represented by primary antioxidant defenses involving enzymes that specifically remove free radicals or oxidants, such as superoxide dismutase, glutathione peroxidase, and catalase. In general, aerobes do not express an excess of antioxidant defenses, although these systems are often inducible by elevated O_2 if sufficient time for adaptation is allowed. Moreover, endogenous antioxidants may not prevent damage by ROS at ambient O_2 . Thus, animals rely on a second line of defense in the form of repair systems, the most important of which removes mutagenic lesions in DNA induced by ROS. Superimposed on such defenses are inducible proteins such as heme oxygenase-1. Heme oxygenases remove the prooxidant heme and produce the antioxidant bilirubin in the process (13). Additional protection is provided by dietary antioxidants. The physiological role of some of these is well established (e.g., vitamin E and ascorbate), whereas the role of others (e.g., flavonoids, carotenoids) is currently uncertain. However, dietary antioxidants appear to be important in delaying/preventing certain human diseases, especially cardiovascular disease and some types of cancer (14).

Antioxidants

Natural antioxidants such as nonenzymatic dietary components are not specific but can scavenge organic and inorganic radicals. These agents are found in numerous plant materials and commonly include an aromatic ring as part of their molecular structure. There are a variety of cyclic ring structures that are generally associated with one or more hydroxyl groups to provide a labile hydrogen and a basis for free radical formation. These antioxidants can be classified as water soluble or lipid-soluble, depending on whether they act primarily in the aqueous phase or in the lipophilic region of cell membranes. Hydrophilic antioxidants include ascorbic acid and urate. Ubiquinols, retinoids, carotenoids, flavonoids, and tocopherol are representative lipid-soluble antioxidants (Fig. 1). Plasma proteins, glutathione, and urate are endogenous, whereas ascorbic acid, carotenoids, retinoids, flavonoids, and tocopherols constitute some of the dietary antioxidants. These compounds possess the potential to scavenge and quench various radicals and ROS. Certain radical scavengers are not recyclable, however, others are recycled through the intervention of a series of enzyme systems or other nonenzymic antioxidant systems.

Based on safety and other pragmatic reasons, the number of active substances used as antioxidants is restricted to a few phenolic substances. The flavonoids are an unusually large group of naturally occurring phenolic compounds ubiquitously distributed in the plant kingdom (Fig. 2). These aromatic compounds are formed in plants from the aromatic amino acids, phenylalanine, tyrosine, and acetate units. Phenylalanine and tyrosine are converted to cinnamic acid and *p*-coumaric acid that condense with acetate units to form the cinnamoyl structure of the flavonoid (15). Flavonoids are generally

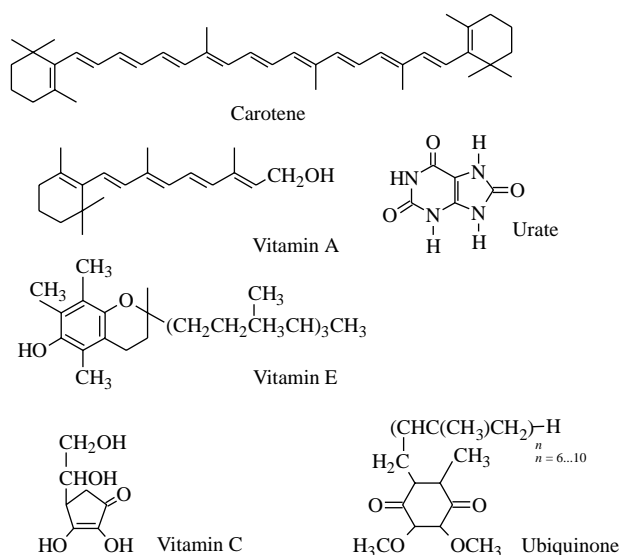


Fig. 1 Natural antioxidants.

known to be plant, flower, leaf, and fruit pigments. They are responsible for the brilliant shades of blue, scarlet, orange, etc., in flowers, fruits, and leaves. They are found in various fruits, vegetables, nuts, seeds, grains, spices, and herbs, as well as in beverages such as tea, cocoa, and wine. Dietary exposure to flavonoids is significant. The average diet in the United Kingdom and United States may contain up to 1 g of mixed flavonoids per day. Their dietary intake far exceeds that of vitamin E (a monophenolic antioxidant) and β -carotene (15).

Flavonoids act as potent metal chelators and free radical scavengers. They are powerful chain-breaking antioxidants (16). Moreover, flavonoids are known to possess vitamin C stabilizing and antioxidant-dependent vitamin C sparing activities. They are also known to increase the absorption of vitamin C. In addition, flavonoids are known to modify the activities of a host of enzyme systems including protein kinase C, protein tyrosine kinase and various other kinases, aldose reductase, myeloperoxidase, NADPH oxidase, xanthine oxidase, phospholipase, reverse transcriptases, ornithine decarboxylase, lipoxxygenase, cyclooxygenase, and so on. Some of these enzyme systems are critically involved in immune function, carcinogenesis, cellular transformation, and tumor growth and metastasis. The physiologic and pathologic processes affected by flavonoids are diverse and numerous, and include secretion, mitogenesis, platelet aggregation and adhesion to endothelial surface, cell motility and malignant cell proliferation, cancer metastasis, and function/expression of adhesion molecules in various mammalian cell types. The antioxidant function

and enzyme-modifying actions of flavonoids could account for many of their pharmacological activities (15).

Quercetin and other flavonoids are effective inhibitors of $O_2^{\cdot -}$ -production by cells. Quercetin is a potent inhibitor of human neutrophil degranulation and $O_2^{\cdot -}$ -production, and also inhibits the phosphorylation of neutrophil proteins accompanying neutrophil activation by phorbol myristate acetate (17). Quercetin can also suppress lipid peroxidation in several biological systems, such as mitochondria, microsomes, chloroplasts, and erythrocytes. Silymarin, a 3-OH flavone present in *Silybum marianum* (the European milk thistle), protects rat liver mitochondria and microsomes from lipid peroxide formation induced by Fe^{2+} -ascorbate and NADPH- Fe^{3+} -ADP systems (18). Soybean isoflavonoids have shown antioxidative potency and prevent peroxidative hemolysis of sheep, rat, and rabbit erythrocytes (15). Quercetin and silybin (19) were reported to exert a protective effect by the decrease in the xanthine dehydrogenase/oxygenase ratio observed during ischemia/reperfusion in the rat. The enzyme (xanthine oxidase) implicated in tissue oxidative injury after ischemia/reperfusion is a source of ROS that is formed from a dehydrogenase during ischemia. The protective effect of quercetin and silybin on the xanthine dehydrogenase/oxidase ratio is due to inhibition of the dehydrogenase.

PROOXIDANTS

A prooxidant is an agent that can induce oxidative stress, which is defined as a shift in the prooxidant-antioxidant balance toward oxidant activity. Oxidative stress induced by a prooxidant in a biological system manifests itself as increased production of bioactive free radical species, a decrease or modulation of antioxidant defenses, and/or an increase in oxidative damage. The fine balance between the oxygen center radicals and antioxidants may be dependent on the concentration of prooxidant, oxygen tension, and interactions with other antioxidants.

Vitamins

Carotenoids

Carotenoids are a well-characterized class of pigments widely distributed in nature and responsible for the bright colors of various fruits and vegetables. Some of the more than 600 different carotenoids are well known, such as β -carotene, which is widely used as a precursor of vitamin A, a food colorant, and a food additive (22). These

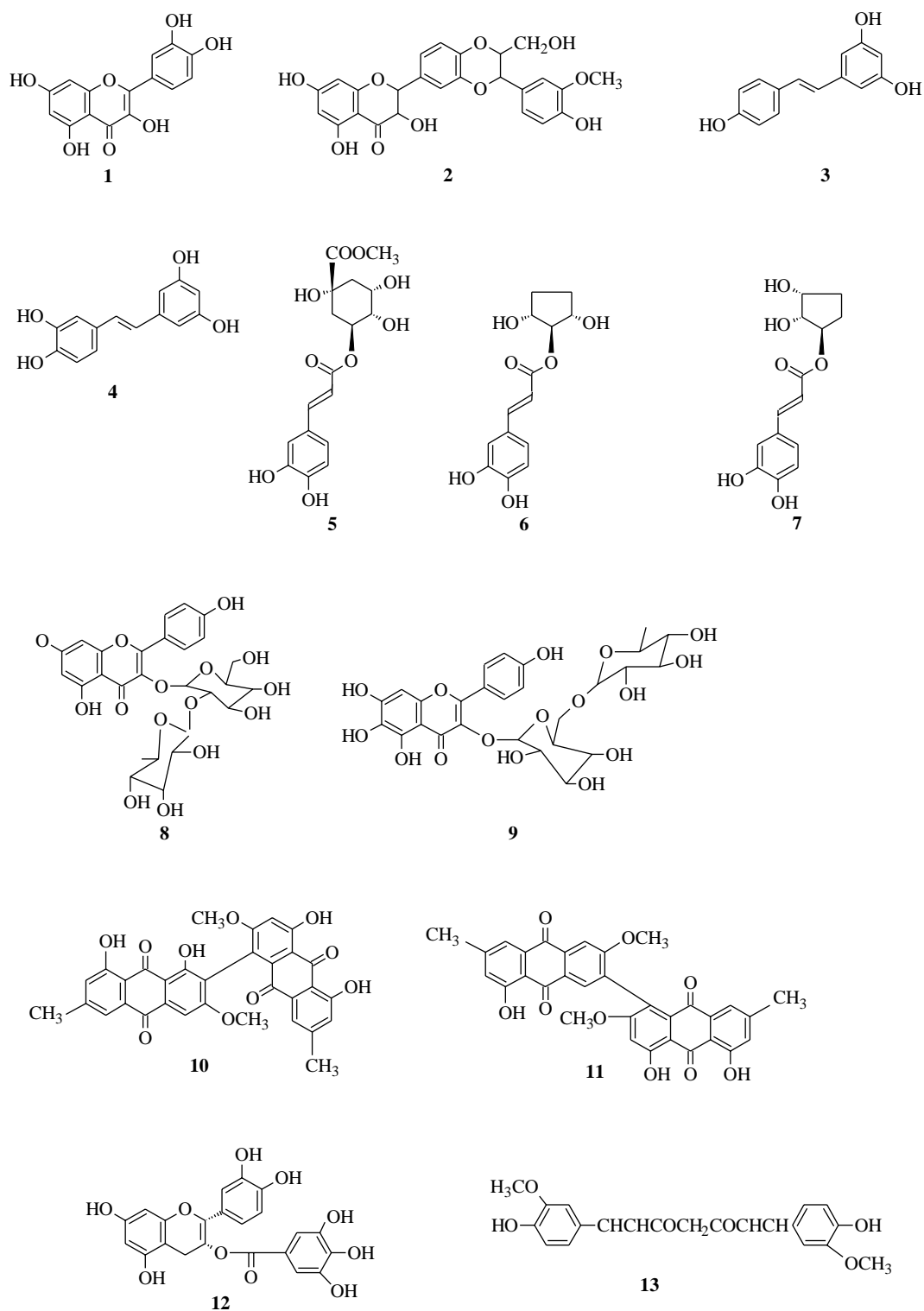


Fig. 2 Phenolic antioxidants isolated from plants; **1.** quercetin; **2.** silymarin; **3.** *trans*-resveratrol; **4.** piceatanmol; **5.** chlorogenic acid methyl ester (16, 17); **6.** 1-(3',4'-Dihydroxycinnamoyl)-cyclopenta-2,5-Diol (16); **7.** 1-(3',4'-Dihydroxycinnamoyl)-cyclopenta-2,3-Diol (16); **8.** kaempferol-3-*O*-neohesperidoside (17); **9.** 5,6,7,4'-Tetrahydroxyflavonol-3-*O*-rutinoside (22); **10.** floribundones I; **11.** floribundones II, **12.** (-)-epicatechin-3-*O*-gallate; **13.** curcumine.

compounds have been shown to function as enhancers of gap–junction communication, stimulants of immune responses, and quenchers of electronically excited species, such as singlet oxygen and triplet sensitizers (23–25). The action of β -carotene and other carotenoids as antioxidants has recently attracted widespread attention. Carotenoids are thought to scavenge free radicals, and the antioxidant action of β -carotene and other carotenoids has been observed with in vitro and in vivo systems (26, 27). However, carotenoids do not have structural features commonly associated with chain-breaking antioxidants. The extensive system of conjugated double bonds in their molecules imparts a prooxidant character and makes them very susceptible to attack by free radical species (Fig. 3).

The development of either a harmful or beneficial cellular response by carotenoids will depend on their antioxidant or prooxidant characteristics, which are determined by various factors in the intra- and extracellular environments such as oxygen tension and β -carotene concentration. When an inappropriate prooxidant activity of carotenoids develops in normal cells, the reactive oxygen metabolites generated could induce damage to lipids, proteins, and DNA. This effect alters normal regulatory functions and can damage cellular integrity or induce neoplastic transformation. Some human intervention trials indicate that carotenoid supplements are of little or no value in preventing chronic disease, such as cardiovascular disease and cancer, and may actually increase lung cancer incidence in smokers (22). In contrast, when carotenoids act as prooxidants in already transformed cells, they could induce beneficial effects, such as inhibiting the growth and development of malignant lesions and/or producing tumor cytotoxic effects (28).

Although it has been reported that carotenoids may prevent normal cells from becoming transformed through their antioxidant activity, there is much evidence indicating that they may also block the growth of cells already

transformed through their prooxidant action. Carotenoid autoxidation has been suggested to occur at a higher level in tumor cells than in normal cells. In tumor cells, therefore, it can be hypothesized that the prooxidant properties of carotenoids prevail over the antioxidant properties (29). In addition, some studies have shown that carotenoids act as prooxidant agents selectively in tumor cells by increasing the expression of heat-shock proteins (30) and enhancing the formation of lipid peroxidation products, further reducing the levels of oxygen-protective enzymes and stimulating the expression of tumor necrosis α -factor.

Ascorbates

Ascorbate is an essential vitamin that must be taken from external sources. It is marketed as a dietary supplement because of its antioxidant properties. Several epidemiologic studies suggest that antioxidant vitamins in sufficient concentrations inhibit heart disease and cancer. However, there is considerable uncertainty about the optimal level of intake. Substantial evidence shows that ascorbate can also act as an oxidant, depending on the environment in which the molecule is present. It can induce cell death, nuclear fragmentation, and internucleosomal DNA cleavage in human myelogenous leukemic cell lines (29). More recently, it was reported that dietary supplementation of 500 mg/day of vitamin C to healthy volunteers for 6 weeks results in significant prooxidant effects. This is exemplified by an increase in lymphocytes with typical markers of DNA damage, mediated by oxygen radicals such as 8-oxoguanine and 8-oxoadenine (31).

Vitamin C supplementation is able to significantly affect CYP2E1-catalyzed drug metabolism in the rat and is linked to an overgeneration of the superoxide anion in hepatic microsomes. Generation of a superoxide anion induced by vitamin C supplementation provides a plausible explanation for the observed damage to DNA in peripheral blood lymphocytes (32). Ascorbate is a reducing agent in brain tissue homogenate but has an oxidizing effect in brain slices. A hypothesis put forth to explain the oxidative effects of ascorbate in cortical slices proposes that extracellular ascorbate is oxidized to dehydroascorbate which is rapidly carried into cells via a glucose transporter. The dehydroascorbate in cytosol is then reduced back to ascorbate, and, during the reduction process, cellular components are oxidized (Fig. 4) (33).

Tocopherols

Vitamin E, the major lipophilic antioxidant of exogenous origin in tissues, is the collective name for the eight major naturally occurring molecules, four tocopherols and four tocotrienols, that qualitatively exhibit the biological

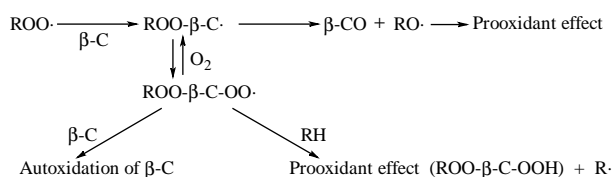


Fig. 3 Proposed reaction pathway of β -Carotene prooxidant activity. β -C, β -carotene; $\text{ROO}\cdot$, peroxy radical; $\text{ROO-}\beta\text{-C}\cdot$, β -carotene radical; $\text{ROO-}\beta\text{-C-OO}\cdot$, β -carotene peroxy radical; $\beta\text{-CO}$, β -carotene epoxide; $\text{RO}\cdot$, alkoxy radical. (Adapted in part from Ref. 28.)

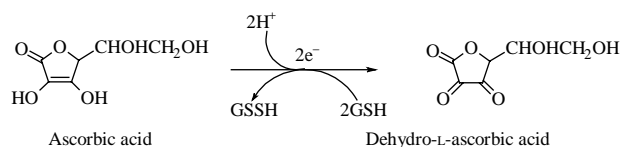


Fig. 4 Structure of the reduced and oxidized forms of vitamin C.

activity of α -tocopherol. α -Tocopherol is generally regarded as the most important lipid-soluble antioxidant in plasma, circulating lipoproteins (34), and tissues, whereas γ -tocopherol and tocotrienols are present in these tissues at much lower concentrations (Fig. 5).

In an investigation of the antitumor activities of α -tocopherol acid succinate and acetate, the acid succinate form inhibited the growth of oral carcinoma cells, while stimulating the growth and differentiation of normal keratinocytes (35). The acetate form increased thymidine incorporation and mutant p53 expression in cancer cells, thereby increasing proliferation and expression of the cyclin regulator p34cdc. Vitamin E treatment reduced the time required for wound healing in the oral cavity and increased the in vitro growth of endothelial cells (36). Platelet adhesion was also inhibited, resulting in a possible reduction in thrombosis. The mechanism of action for this response appears to be blockage of the prostaglandin pathway at the site of cyclooxygenase activity, and it is a plausible mechanism that supports the use of vitamin E in periodontal disease (29).

The antioxidative activity of vitamin E is converted to prooxidant activity when mild conditions were used to initiate oxidation. A prooxidant effect of α -tocopherol on lipid peroxidation was found only when the samples were virtually free of ascorbate, or if the final concentration of ascorbate in the samples was physiologically low. Adding ascorbate to a near-physiological final concentration restored the antioxidant activity of α -tocopherol under mild oxidative conditions (37).

Retinoids

Retinoids, metabolic and synthetic derivatives of vitamin A, have been shown to function as effective antioxidants and inhibit the peroxidation of PUFA in lipid bilayers. For example, several retinoids inhibit ascorbate-dependent, iron-catalyzed lipid peroxidation in rat liver microsomes and brain mitochondria. Retinol palmitate was shown to function as an antioxidant in rat heart and brain tissues (38). However, retinoic acid was shown to stimulate the rate of 2,2'-Azobis(2-Amidinopropane)-initiated autoxidation of linoleic acid in sodium dodecyl sulfate micelles, and this observation may account for the prooxidant effect of retinoic acid in this system (39). In addition, possible

detrimental effects of retinoids (e.g. promotion of tumor growth), increase in low-density lipoproteins (LDL) and triglyceride, and exacerbation of preexisting autoimmune disease have all been reported (40).

Phenolics

Phenolics are one of the major groups of nonessential dietary components that have been associated with the inhibition of atherosclerosis and cancer. The bioactivity of phenolics may be related to their antioxidant behavior, which is attributed to their ability to chelate metals, inhibit lipoxygenase, and scavenge free radicals. However, phenolics can also function as prooxidants by chelating metals in a manner that maintains or increases their catalytic activity (Fig. 6). Also, polyphenolics reduce metals, thereby increasing their ability to form free radicals from peroxide.

Flavonoids

Flavonoids, especially, those with catechol or pyrogallol groups, obviously are prone to autoxidation reactions (40). In a recent paper by Morgan et al. (41), the prooxidative and antioxidative properties of phenolics from soybeans and other legumes were documented. The primary phenolics (phenolic acids) and flavonoids were able to reduce ferric to ferrous ions and were able to chelate and alter the catalytic activity of iron. Most of the phenolics tested were also able to inhibit oxidation of linoleic acid micelles and ferrous ion-catalyzed oxidation of glutamine synthase, presumably through free radical scavenging and removal of iron from catalytic sites via chelation. Although flavonoids inhibited oxidation in certain systems, they did not protect against all forms of oxidative damage. The phenolics chelated iron, but this metal ion was still catalytically active and able to oxidize both deoxyribose and DNA. Prooxidant activity of phenolics has also been observed for carnosol, carnosic acid, quercetin, rutin, and luteolin (41).

It was also found that pH was essential in determining the oxidative role of phenolics. In general, a decrease in pH increased iron-reducing activity and reduced the ability of phenolics to chelate and inhibit the catalytic activity of iron. Increasing pH increased deoxyribose and DNA oxidation. Inhibition of lipid oxidation was also influenced by pH, with γ -resorcylic acid being antioxidative at pH 5.8 and prooxidative at pH 7.4. Hydrobenzoic acid was antioxidative, and apigenin-7-glucoside was prooxidative at pH 7.4, yet neither had an effect on lipid oxidation at pH 5.8. These results suggest that the pH of biological tissues could also influence the antioxidative/prooxidative activity of phenolics (42).

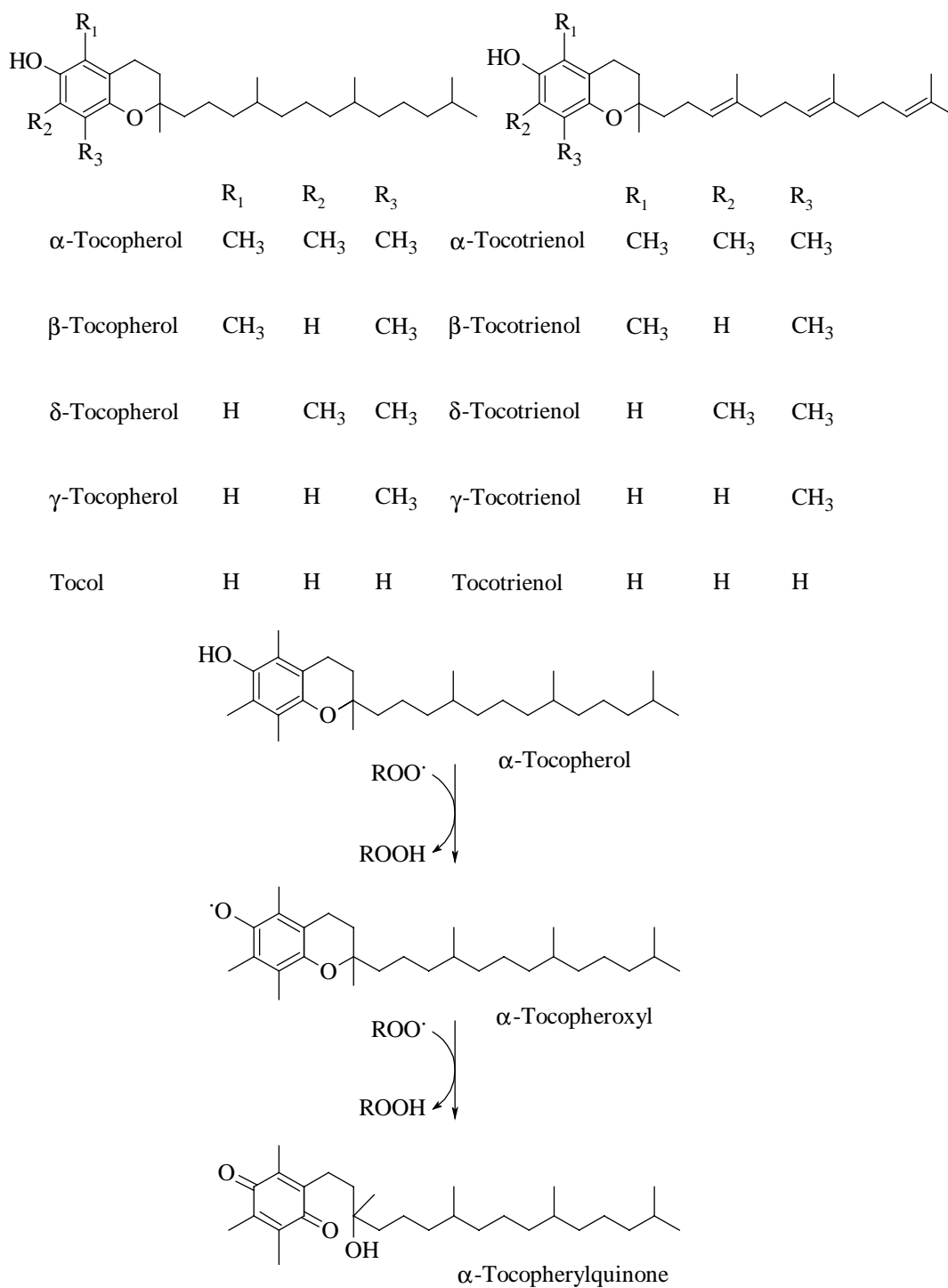


Fig. 5 Chemical structures of the vitamin E groups and metabolic pathway of α -Tocopherol.

A possible mechanism of cytotoxicity in polyphenols may be related to their prooxidant properties (43). Flavonoids autoxidize in an aqueous medium and may form highly reactive HO^\bullet radicals in the presence of

transition metals. In addition, polyphenols and flavonoids may act as substrates for peroxidase and other metalloenzymes, yielding quinone- or quinomethide-type prooxidant and/or alkylating products. The prooxidant character

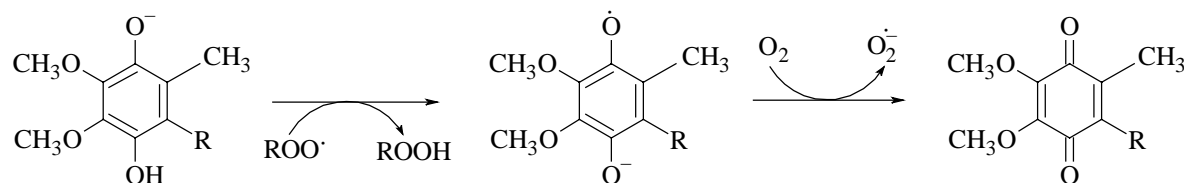


Fig. 6 Antioxidant and prooxidant functions of ubiquinone.

of polyphenol cytotoxicity is supported by the formation of activated oxygen species during gallic acid induced apoptosis, and by the enhancement of gallic and caffeic acid induced apoptosis by non toxic concentrations of copper ions (44).

Catechins, such as (–)-epicatechin and (–)-epigallocatechin abundant in green tea, possess the antioxidative and prooxidative characteristics of Cu^{2+} -induced LDL oxidation. In the initiation phase, LDL oxidation was inhibited by addition of catechin. In contrast, during the propagation phase of LDL oxidation, catechins served as accelerators of oxidation. Depending on redox status, they might form reactive oxidation products such as semi-quinones and quinones and function to stimulate oxidative reactions (44).

Quercetin, a highly studied antioxidant flavonoid has the potential to inhibit free radical processes in cells by a) scavenging $\text{O}_2^{\cdot-}$, b) blocking lipid peroxidation, c) reacting with peroxy or lipid peroxy radicals, d) inhibiting formation of HO^{\cdot} , and e) chelating iron

ions. The biological effects of quercetin are believed to result from its antioxidant properties. Recently, it was clearly demonstrated that quercetin could function both as an antioxidant and a prooxidant, depending on concentration and free radical sources and their location in the cell. Also, quercetin was observed to be cytotoxic in a dose-dependent manner. Although the exact mechanism of cytotoxicity has not yet been fully elucidated, it may involve formation of $\text{O}_2^{\cdot-}$ or its metabolite *o*-quinone (Fig. 7). Such species are known to be toxic and to bind irreversibly to various cell constituents by covalent binding with sulfhydryl groups or other essential groups (45).

Catecholestrogens

The antioxidant properties of estrogens have been demonstrated in many in vitro and in vivo studies (46–48). For instance, estrogens inhibit the oxidation of LDL, the peroxidation of lipids, and the oxidation of cholesterol. The administration of 17β -estradiol to ovariectomized pigs inhibits the oxidation of LDL. This

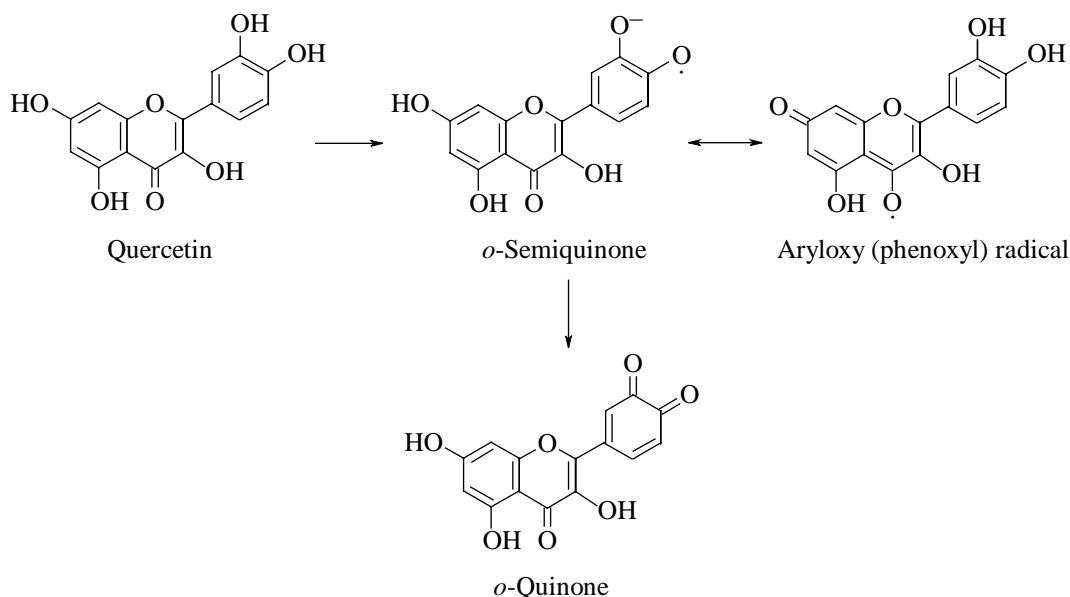


Fig. 7 A simple scheme of quercetin oxidoreductive activation.

effect has also been observed in postmenopausal women after administration of 17 β -estradiol. These antioxidant activities of estrogens are believed to explain the lower rate of heart disease in premenopausal women, or postmenopausal women treated with estrogen, compared to that of men (49).

In contrast, prooxidant effects of estrogens have been established in other model systems (50). 17 β -estradiol or other estrogens induce single-strand breaks or 8-hydroxylation of guanine bases of DNA in Syrian hamsters treated with these hormones. Moreover, metabolites of estrogen or diethylstilbestrol in the presence of peroxidase and DNA induce 8-hydroxylation of guanine bases. In addition to oxidant-induced damage of DNA, estrogens have also been shown to generate lipid peroxidation and oxygen-radical-mediated oxidation of amino acid residues of proteins, resulting in carbonyl-containing moieties (50).

Estrogens may exhibit either pro- or antioxidant activities depending on the nature of their metabolites and concentrations (Fig. 8). Pharmacological concentrations of parent hormones or their metabolites clearly have antioxidant properties. This inhibitory mechanism may be based on the free radical scavenging action of the phenol moiety of estrogen. In contrast, 2- or

4-hydroxyestradiol enhances the oxidation of LDL by decreasing lag times of lipid peroxidation by 40–50% compared to control values in the absence of estrogen. The catechol structure of catecholestrogens appears to be necessary for this prooxidant activity, as the parent hormones and other estrogen metabolites do not possess any detectable oxidant activity. Therefore, the *in vivo* prooxidant activity of estrogens may be dependent on their conversion to catecholestrogen metabolites in a specific organ or species. The mechanism of prooxidant activity of catecholestrogens is likely based on the reduction of metal ions, specifically Cu²⁺ to Cu⁺, by the catecholestrogen metabolites. The lipid oxidation of LDL by Cu⁺ generated hydroxyl radicals, which further initiate oxidation of lipids (51).

DISEASES ASSOCIATED WITH FREE RADICALS

Ischemia/Reperfusion Injury

The syndrome of ischemia/reperfusion (I/R) injury has been characterized in recent years for the heart, brain, intestine, kidney, and other organs. This phenomenon

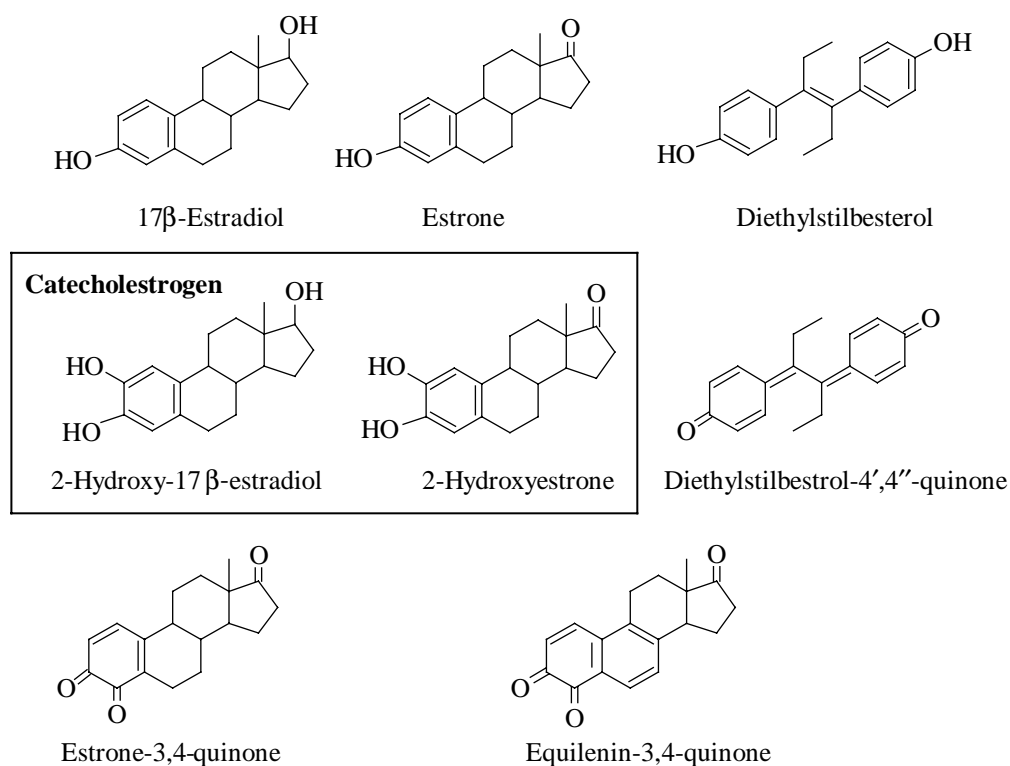


Fig. 8 The metabolites of estrogen.

consists of a paradoxical increase in tissue injury during the reperfusion period in an organ that has sustained relatively minor damage during a period of ischemia. It is now evident that reperfusion tissue injury is mediated through oxidant mechanisms associated with the generation of oxygen-based radicals. ROS have been implicated in both the myocardial dysfunctions that are observed during reperfusion following short periods of ischemia (the stunned myocardium) and the irreversible injury to cardiac myocytes that occurs during reperfusion after longer periods of ischemia (52).

Infusions of high concentrations of the catecholamines, epinephrine, or norepinephrine into experimental animals are known to produce myocellular mitochondrial swelling, myofibrillar disruption, plasma membrane blebbing, and myocardial necrosis. It has been suggested that these cardiotoxic effects result not from the catecholamines themselves but from the production of $O_2^{\cdot-}$ and H_2O_2 formed by a complicated series of reactions during the autoxidation of catecholamines. It was observed that vitamin E deficient rats were more sensitive to the cardiotoxic effects of isoproterenol, whereas myocardial damage induced by this synthetic catecholamine was reduced when the diet was supplemented with vitamin E. However, the results of studies demonstrating protection by antioxidants against catecholamine-induced myocardial necrosis must be interpreted with caution since accumulation of neutrophils, a major source of oxygen radicals, has been observed in such model (53).

Carvedilol, a potent antioxidant, prevents the lipoperoxidation of mitochondrial membranes, which suggests a strong contribution to the known cardioprotective activity of this compound through protection of mitochondrial function (54).

A similar cardioprotective benefit is achieved by agents and antioxidant enzymes that scavenge hydroxyl radicals (or reduce their formation), but not by agents that reduce superoxide anion production. Some compounds of plant origin that have been shown to protect against ischemic injury are procyanidine from *Vitis vinifera* (55), resveratrol from red wine (56), and ginseng extract (57).

Cancer-Carcinogenesis

The progression of tumor formation may be slow, often taking 10 years or more. It is important that carcinogenesis can be viewed as a multistage, microevolutionary process. It is generally agreed that cancer can be derived from a single abnormal cell, and work with experimental systems shows that carcinogenesis is divisible into three major stages: initiation, promotion, and progression. Initiation is

a heritable aberration of a cell. Such initiated cells can undergo transformation to malignancy if promotion and progression follow. Initiation appears to be irreversible and can result from DNA damage. Promotion, however, is affected by factors that do not alter DNA sequences; it involves the selection and clonal expansion of initiated cells. This process is partly reversible and accounts for a major portion of the lengthy latent period of carcinogenesis. The final stage of tumor formation is the progression of a benign growth to a malignant neoplasm. There is loss of growth control, an escape from the host defense mechanism, and metastasis.

Certain initiators, such as radiation or chemical carcinogens, can induce the production of various free radicals and subsequent DNA base sequence alteration. In addition, cells of the immune system (e.g., neutrophils and macrophages) produce $O_2^{\cdot-}$ and H_2O_2 , which have been associated with the induction of experimental cancers. Oxygen free radicals and methyl radicals are known to damage DNA. In some cases, such free radicals may arise in reactions catalyzed by ferric and cupric ions localized in the vicinity of cellular DNA. Free radical mediated DNA damage can have serious consequences on an organism unless the damage is repaired. Although oxygen free radical effects can lead to DNA damage, they may also directly affect the protein components of the DNA repair apparatus. Unrepaired DNA alterations are inherited as mutations.

Oxygen radicals and related species may also be involved in tumor-promotion. Tumor-promoting phorbol esters not only can induce changes in cellular genes leading to some of the phenotypic characteristics of tumor cells, but they also can stimulate inflammatory leukocytes to release superoxide. The release of superoxide by phagocytic cells following stimulation with phorbol esters is proportional to their tumor-promoting activity. Low levels of both $O_2^{\cdot-}$ and H_2O_2 , products of the "respiratory burst," can promote fibroblast growth, possibly fibroblasts that harbor an oncogene or a mutated protooncogene. Also, low levels of superoxide can stimulate growth or growth responses in a variety of cell types when added exogenously to culture medium. In particular, these species stimulate the activation and translocation of protein kinase C, as well as the expression of early growth-regulated genes such as the protooncogenes *c-fos* and *c-myc*. Superoxide and/or hydrogen peroxides might function as mitogenic stimuli through biochemical processes common to natural growth factors. Thus, signaling of growth responses involving released superoxide or hydrogen peroxide may be mediated through the oxidative modification of components of the signal transduction pathway. It is also possible that oxidative

inactivation of serum protein inhibitors allows proteases to remodel the cell surface, thereby facilitating (modulating) the action of normal growth factors (58).

A final and decisive step in carcinogenesis is the invasion and metastatic spread of the tumor to various body spaces and cavities. This appears to be facilitated by the activation of genes for the release of proteolytic enzymes. Although high levels of immune cells appear to favor cell killing, lower numbers of immune cells can favor metastasis. Again, the released superoxide may serve to promote metastatic growth. Alternatively, superoxide could inactivate serum antiproteases, some of which are extremely sensitive to oxidative inactivation (59).

In addition, lipid peroxidation is associated with some phases of carcinogenesis. There is increasing evidence that covalent binding of carcinogens or toxic substances to cellular macromolecules, particularly those carrying genetic information, is the primary event in the initiation of carcinogenesis. Thus, covalent binding to macromolecules could be the basis of many pathological changes induced by toxic substances. The ultimate forms of xenobiotics are believed to be reactive electrophilic metabolites, which combine with nucleophilic groups of macromolecules. It is also possible that miscoding or mutagenesis may be of minor importance in the initial events of chemical carcinogenesis, and that genetic transpositions, including relatively large regions of the genome, may be more relevant (60).

The DNA adducts, deoxyadenosine and deoxyguanosine, which are induced by malondialdehyde, the end-product of lipid peroxidation, accumulate in human breast tissues. These adducts are present at relatively higher concentrations in breast cancer cells compared to normal breast cells (61). In a recent study, serum antioxidative vitamin levels and lipid peroxidation were compared in gastric cancer patients (62). The level of serum ascorbic acid, α -tocopherol, β -carotene, and retinol were assessed. The levels of ascorbic acid in patients with gastric carcinoma were less than one-fifth of that in the control group, and the production of β -carotene and α -tocopherol were decreased, as well.

Neuronal Disease

Lipid peroxidation of biological membranes gives rise to degeneration of synapses and neurons, and may be observed in stroke, or neuronal disorders such as Alzheimer's, Parkinson's and Huntington's diseases. Oxidative stress and damage are accepted features of neural degeneration. The pathological presentation of Alzheimer's disease, the leading cause of senile dementia,

involves regionalized neuronal death and accumulation of intraneuronal and extracellular lesions (63). 4-Hydroxynonenal mediates oxidation-induced impairment of glutamate transport and mitochondrial function in synapses (64). Amyloid β -protein may be involved in modulation of membrane lipid peroxidation. Amyloid β -protein fragments 25–35 [A-beta (25–35)] inhibit lipid peroxidation at low concentrations as a result of physicochemical interactions with the membrane lipid layer (65). Further, there is close association between increased levels of the antioxidant enzymes superoxide dismutase and heme oxygenase-1 and cytoskeleton abnormalities found in Alzheimer's disease (66).

DETECTION AND CHARACTERIZATION OF FREE RADICALS

ROS have been involved in the pathogenesis of a variety of human diseases. Their injury potential and pathologic role demand quantitative methods that are diagnostic of the process and that meet basic analytical criteria regarding accuracy, reliability, sensitivity and specificity. In addition, the determination and quantification of antioxidant activity is necessary for the discovery and evaluation of drug candidates (67). However, because of their reactive nature and short half-lives, it is difficult to quantify ROS. Alternatively, analyses involving secondary or end products produced by the attack of ROS on lipids, enzymes, or other cellular components are generally preferred. In spite of recent advances in technology, however, these indirect methods often give misleading results due to their poor specificity and sensitivity.

Chemiluminescence Measurements

Chemiluminescence is the production of light generated from chemical sources. The quantum yield of photons for intrinsic (nonstimulated) or native reactions are low but organic substances are able to undergo an oxidative reaction that can be sufficiently exothermic to produce an emitting state. Generally, the light produced is in the visible range (400–600 nm), but UV or infrared emission is possible.

Due to potential variability and low intensity of native chemiluminescence, enhancer compounds have been introduced (68). These compounds were selected primarily as a result of their high quantum efficiency and photon yield after oxidation. Luminol and lucigenin are enhancers for oxygenation when added to an *in vitro* biological

system and form high levels of excited-state products and chemiluminescence. These compounds react with all species of oxidants to form 3-aminophthalate and *N*-methylacridone, respectively. The excited electrons in these compounds revert to their ground state with the emission of energy as light, which can be detected by photomultipliers. The sensitivities of luminol and lucigenin vary. Luminol detects H_2O_2 , $\text{HO}\cdot$, hypochloride, peroxy nitrile and lipid peroxy radicals, whereas lucigenin is particularly sensitive to the superoxide radicals. Chemiluminescence may be utilized as a direct noninvasive method for measuring ROS, and for detection of lipid hydroperoxide, phospholipid and cholesterol hydroperoxide if cytochrome *c*-heme is added prior to luminol.

Thiobarbituric Acid (TBA) Assays

The TBA test is perhaps the most widely used method for determining lipid peroxidation. The representative adduct of lipid peroxidation, malondialdehyde, forms a 1:2 adduct with TBA that can be measured by spectroscopy or fluorometry. The general procedure, of which there are numerous variations, simply involves heating a small quantity of the test substance for a defined period of time in an aqueous acidic solution of TBA, and then measuring the absorbance (535 nm) of the red color which is produced in the TBA reaction. It should be considered as an index of oxidative stress that represents primarily lipid peroxidation (69).

Chemical Measurement of Superoxide Radical

The reduction of yellow nitroblue tetrazolium (NBT) to blue formazan is applied as a probe of $\text{O}_2^{\cdot -}$ generation in biological systems. This reaction is utilized in demonstrating the role of phagocytes (neutrophils and monocytes/macrophages) in the host response to infection and inflammation. Oxygen is rapidly reduced via a complex NADPH-oxidase system composed of a flavoprotein and a cytochrome. The cytochrome has a sufficiently low midpoint potential to allow the direct catalytic transfer of electrons from NADPH to oxygen resulting in the production of superoxide.

The respiratory burst of phagocytic cells can be assessed by incubating a suspension of cells in an isotonic solution of the yellow oxidized nitroblue tetrazolium dye. During this process, the soluble dye interacts with the cytoplasmic components associating with the oxidant species generated. NBT reduction by activation cells in the presence of superoxide dismutase has been shown to be markedly reduced (70), which suggests the major oxidant

species responsible for the reduction of dye to a black-blue deposit called formazan is superoxide. The overall degree of NBT reduction in a given cell population can be quantified by measuring the concentration of reduced NBT or formazan spectrophotometrically.

Cytochrome *c* Reduction Assay Using a HL-60 Cell Culture System

HL-60 is an acute human premyelocytic leukemia cell line derived by Collins et al. (71), of which about 10% spontaneously differentiate. Various differentiation inducers, such as DMSO, TPA (12-*O*-tetradecanoylphorbol 13-acetate), or retinoic acid, lead to the differentiation of HL-60 cells through the monocyte or granulocyte pathways. HL-60 cells treated with DMSO appear as granulocytes with morphological and functional changes including production of superoxide anion and phagocytosis. Respiratory burst due to phagocytosis produce ROS such superoxide anion in the nonmitochondria oxidase system. In a simple assay procedure, HL-60 cells differentiated by treatment with 1.3% DMSO are stimulated to produce superoxide anion by addition of TPA, and cytochrome *c* oxidized by superoxide anion is measured by absorbance at 550 nm (72).

Xanthine/Xanthine Oxidase Assay

Xanthine oxidase is an obvious candidate for the production of oxygen free radicals. This enzyme is localized in the liver, small intestinal mucosa, and vascular endothelial cells, and catalyzes the hydroxylation of many purine substrates. It converts hypoxanthine to xanthine and then to uric acid in the presence of molecular oxygen to yield superoxide anion. During ischemia, xanthine dehydrogenase is converted to the oxidase form. At the same time, ATP is degraded to hypoxanthine that accumulates in ischemic tissue. On reperfusion, with the readmission of large quantities of molecular oxygen in the presence of high concentrations of hypoxanthine that is the other substrate for xanthine oxidase, there may be a burst of superoxide anion production. With in vitro systems, antioxidant activity can be measured by the absorbance of uric acid (292 nm) which is produced by xanthine oxidase. If a test sample inhibits the enzyme, xanthine oxidase cannot produce uric acid from xanthine (73).

DPPH Assay

DPPH (1,1-diphenyl-2-picrylhydrazyl) is a purple-colored stable free radical that is reduced to the yellow-colored

diphenylpicrylhydrazine by free radicals. The DPPH assay measures one electron, such as hydrogen atom donating activity and hence provides a measure of free radical scavenging activity. This assay is suitable for the initial screening of multiple samples, such as plant extracts. Reaction mixtures containing test samples dissolved in DMSO and DPPH in absolute ethanol are incubated at 37°C for 30 min in a 96-well plate and absorbance measured at 515 nm (74).

CONCLUSIONS

As summarized in this article, ROS are involved in reactions of importance in human disease states. A mechanistic understanding of these pathological processes is beginning to emerge. As a result, intervention strategies can be devised and antioxidants are receiving a great attention as potential drugs. While it is clear that prevention of oxidative damage should have a beneficial effect on human health in general, the possibility of prooxidant activity leading to adverse health effects must be strongly borne in mind. A number of assay systems are currently available for thorough characterization of the large number of potential antioxidants which are known, or for the discovery of new chemical entities. With these tools and prudent preclinical and clinical studies, it should be possible to devise dietary strategies or pharmaceutical preparations that will reduce morbidity and mortality.

ACKNOWLEDGMENT

The authors extend great thanks to Dr. Ghee T. Tan for review of this manuscript. Laboratory work conducted in this area is supported by program project P01 CA48112 awarded by the National Cancer Institute.

REFERENCES

1. Chan, H.W.-S. The Mechanism of Autoxidation. In *Autoxidation of Unsaturated Lipids*; Chan, H.W.-S., Ed.; Academic Press: London, 1987; 1–16.
2. Gardner, H.W.; Kleiman, R.; Weisleder, D.; Inglett, G.E. Cysteine Adds to Lipid Hydroperoxide. *Lipids* **1977**, *12*, 655–660.
3. Dobretsov, G.E.; Borschevskaya, T.A.; Petrov, V.A.; Vladimirov, Y.A. The Increase of Phospholipid Bilayer Rigidity after Lipid Peroxidation. *FEBS Lett.* **1977**, *84*, 125–134.
4. Lenz, M.L.; Michael, L.H.; Smith, C.V.; Hughes, H.; Shappell, S.B.; Taylor, A.A.; Entman, M.L.; Mitchell, J.R. Glutathione Disulfide Formation and Lipid Peroxidation During Cardiac Ischemia and Reflow in the Dog in Vivo. *Biochem. Biophys. Res. Commun.* **1989**, *164*, 722–727.
5. Kuo, C.-L.; Vaz, A.D.N.; Coon, M.J. Metabolic Activation of Trans-4-hydroxy-2-nonenal, a Toxic Product of Membrane Lipid Peroxidation and Inhibitor of P450 Cytochromes. *J. Biol. Chem.* **1997**, *272*, 22611–22616.
6. Cadenas, E.; Müller, A.; Brigelius, R.; Esterbauer, H.; Sies, H. Effects of 4-Hydroxynonenal on Isolated Hepatocytes. *Biochem. J.* **1983**, *214*, 479–487.
7. Alin, P.; Danielson, U.H.; Mannervi, K.B. 4-hydroxyalk-2-enals are Substrates for Glutathione Transferase. *FEBS Lett.* **1985**, *179*, 267–270.
8. Sadzadeh, S.M.H.; Graf, E.; Panter, S.S.; Hallaway, P.E.; Eaton, J.W. Hemoglobin: A Biological Fenton Reagent. *J. Biol. Chem.* **1984**, *259*, 14354–14356.
9. Egan, R.W.; Paxton, J.; Kuehl, F.A., Jr. Mechanism for Irreversible Self-Deactivation of Prostaglandin Synthetase. *J. Biol. Chem.* **1976**, *251*, 7329–7335.
10. Harlan, J.M.; Callahan, K.S. Role of Hydrogen Peroxide in the Neutrophil-Mediated Release of Prostacyclin from Cultured Endothelial Cells. *J. Clin. Invest.* **1984**, *74*, 442–448.
11. Whorton, A.R.; Montgomery, M.E.; Kent, R.S. Effect of Hydrogen Peroxide on Prostaglandin Production and Cellular Integrity in Cultured Porcine Aortic Endothelial Cells. *J. Clin. Invest.* **1985**, *76*, 295–302.
12. Lesnefsky, E.J. Tissue Iron Overload and Mechanisms Of Iron-Catalyzed Oxidative Injury. In *Free Radicals in Diagnostic Medicine. A Systems Approach to Laboratory Technology, Clinical Correlations, and Antioxidant Therapy, Advances in Experimental Medicine and Biology*; Armstrong, D., Ed.; Plenum Press: New York, 1994; 366, 129–146.
13. Gutteridge, J.M. Iron in Free Radical Reactions and Antioxidant Protection. In *Free Radicals, Oxidative Stress, and Antioxidants. Pathological and Physiological Significance*; Özben, T., Ed.; Plenum Press: New York, 1998; 296, 1–14.
14. Jacob, R.A.; Burri, B. Oxidative Damage and Defense. *Am. J. Clin. Nutr.* **1996**, *63*, 985s–990s.
15. Kandaswami, C.; Middleton, E., Jr. Free Radical Scavenging and Antioxidant Activity of Plant Flavonoids. In *Free Radicals in Diagnostic Medicine. A Systems Approach to Laboratory Technology, Clinical Correlations, and Antioxidant Therapy, Advances in Experimental Medicine and Biology*; Armstrong, D., Ed.; Plenum Press: New York, 1994; 366, 351–376.
16. Santos, A.C.; Uyemura, S.A.; Lopes, J.L.; Bazon, J.N.; Mingatto, F.E.; Cutti, C. Effect of Naturally Occurring Flavonoids on Lipid Peroxidation and Membrane Permeability Transition in Mitochondria. *Free Radical Biol. Med.* **1998**, *24*, 1455–1461.
17. Blackburn, W.D.; Heck, L.W.; Wallace, R.W. The Bioflavonoid Quercetin Inhibits Neutrophil Degranulation, Superoxide Production, and the Phosphorylation of Specific Neutrophil Proteins. *Biochem. Biophys. Res. Commun.* **1987**, *144*, 1229–1236.
18. Valenzuela, A.; Guerra, R.; Garrido, A. Silybin Dihemisuccinate Protects Rat Erythrocytes Against Phenylhydrazine-induced Lipid Peroxidation and Hemolysis. *Planta Med.* **1987**, *53*, 402–405.

19. Valenzuela, A.; Lagos, C.; Schmidt, K.; Videla, L.A. Silymarin Protection Against Hepatic Lipid Peroxidation Induced by Acute Ethanol Intoxication in the Rat. *Biochem. Pharmacol.* **1985**, *34*, 2209–2212.
20. Wang, H.; Nair, M.G.; Strasburg, G.M.; Booren, A.M.; Gray, J.I. Novel Antioxidant Compounds from Tart Cherries (*Prunus cerasus*). *J. Nat. Prod.* **1999**, *62*, 86–88.
21. Gamez, E.J.C.; Luyengi, L.; Lee, S.K.; Zhu, L.-F.; Zhou, B.-N.; Fong, H.H.S.; Pezzuto, J.M.; Kinghorn, A.D. Antioxidant Flavonoid Glycosides From *Daphniphyllum calycinum*. *J. Nat. Prod.* **1998**, *61*, 706–708.
22. Mayne, S.T. β -Carotene, Carotenoids, and Disease Prevention in Humans. *FASEB J.* **1996**, *10*, 690–701.
23. Zhang, L.-X.; Cooney, R.V.; Bertram, J.S. Carotenoids Up-regulate *Connexin 43* Gene Expression Independent of Their Provitamin A or Antioxidant Properties. *Cancer Res.* **1992**, *52*, 5707–5712.
24. Bendich, A.; Shapiro, S.S. Effect of β -Carotene and Canthaxanthin on the Immune Responses of the Rat. *J. Nutr.* **1986**, *116*, 2254–2262.
25. Foote, C.S.; Denny, R.W. Chemistry of Singlet Oxygen. VII. Quenching by β -carotene. *J. Am. Chem. Soc.* **1968**, *90*, 6233–6235.
26. Krinsky, N.I. Actions of Carotenoids in Biological Systems. *Annu. Rev. Nutr.* **1993**, *13*, 561–587.
27. Sies, H.; Stahl, W. Vitamins E and C, β -carotene, and Other Carotenoids as Antioxidants. *Am. J. Clin. Nutr.* **1995**, *62* (Suppl.), 1315S–1321S.
28. Palozza, P. Prooxidant Actions of Carotenoids in Biological Systems. *Nutr. Rev.* **1998**, *56*, 257–265.
29. Schwartz, J.L. The Dual Roles of Nutrients as Antioxidants and Prooxidants: Their Effects on Tumor Cell Growth. *J. Nutr.* **1996**, *126*, 1221S–1227S.
30. Iftikhar, S.; Lietz, H.; Mobarhan, S.; Frommel, T.O. In Vitro β -carotene Toxicity for Human Colon Cancer Cells. *Nutr. Cancer* **1996**, *25*, 221–230.
31. Paolini, M.; Pozzetti, L.; Pedulli, G.F.; Marchesi, E.; Cantelli-Forti, G. The Nature of Prooxidant Activity of Vitamin C. *Life Sci.* **1999**, *64* (PL), 273–278.
32. Paolini, M.; Pozzetti, L.; Pedulli, G.F.; Cipollone, M.; Mesirca, R.; Cantelli-Forti, G. Paramagnetic Resonance in Detecting Carcinogenic Risk from Cytochrome P450 Overexpression. *J. Invest. Med.* **1996**, *44*, 470–473.
33. Song, J.H.; Shin, S.H.; Ross, G.M. Prooxidant Effects of Ascorbate in Rat Brain Slices. *J. Neurosci. Res.* **1999**, *58*, 328–336.
34. Zhang, D.; Okada, S.; Yu, Y.; Zheng, P.; Yamaguchi, R.; Kasai, H. Vitamin E Inhibits Apoptosis, DNA Modification, and Cancer Incidence Induced by Iron-mediated Peroxidation in Wistar Rat Kidney. *Cancer Res.* **1997**, *57*, 2410–2414.
35. Ottino, P.; Duncan, J.R. Effect of α -Tocopherol Succinate on Free Radical and Lipid Peroxidation Levels in BL6 Melanoma Cells. *Free Radical Biol. Med.* **1997**, *22*, 1145–1151.
36. Kim, J.E.; Shklar, G. The Effect of Vitamin E on the Healing of Gingival Wounds in Rats. *J. Periodontol.* **1983**, *54*, 305–308.
37. Kontush, A.; Finckh, B.; Karten, B.; Kohlschütter, Beisiegel, U. Antioxidant and Prooxidant Activity of α -tocopherol in Human Plasma and Low Density Lipoprotein. *J. Lipid. Res.* **1996**, *37*, 1436–1448.
38. Ciaccio, M.; Valenza, M.; Tesoriere, L.; Bongiorno, A.; Albiero, R.; Livrea, M.A. Vitamin A Inhibits Doxorubicin-Induced Membrane Lipid Peroxidation in Rat Tissues in Vivo. *Arch. Biochem. Biophys.* **1993**, *302*, 103–108.
39. Freyaldenhoven, M.A.; Lehman, P.A.; Franz, T.J.; Lloyd, R.V.; Samokyszyn, V.M. Retinoic Acid-Dependent Stimulation of 2,2'-Azo bis(2-Aminopropane)-Initiated Autoxidation of Linoleic Acid in Sodium Dodecyl Sulfate Micelles: A Novel Prooxidant Effect of Retinoic Acid. *Chem. Res. Toxicol.* **1998**, *11*, 102–110.
40. Bolton, J.L.; Pisha, E.; Shen, L.; Krol, E.S.; Iverson, S.L.; Huang, Z.; van Breemen, R.B.; Pezzuto, J.M. The Reactivity Of *O*-quinones Which Do Not Isomerize to Quinone Methides Correlates with Alkylcatechol-Induced Toxicity in Human Melanoma Cells. *Chem. Biol. Interact.* **1997**, *106*, 133–148.
41. Morgan, J.F.; Klucas, R.V.; Grayer, R.J.; Abian, J.; Becana, M. Complexes of Iron With Phenolic Compounds from Soybean Nodules and Other Legume Tissues: Prooxidant and Antioxidant Properties. *Free Radical Biol. Med.* **1997**, *22*, 861–870.
42. Decker, E.A. Phenolics: Prooxidants or Antioxidants. *Nutr. Rev.* **1997**, *55*, 396–407.
43. Sergediene, E.; Jönsson, K.; Szymusiak, H.; Tyrakowska, B.; Rietjens, I.M.C.M.; Cénas, N. Prooxidant Toxicity of Polyphenolic Antioxidants to HL-60 Cells: Description of Quantitative Structure-activity Relationships. *FEBS Lett.* **1999**, *462*, 392–396.
44. Yamanaka, N.; Oda, O.; Nagao, S. Prooxidant Activity of Caffeic Acid, Dietary Non-flavonoid Phenolic Acid, on Cu^{2+} -Induced Low Density Lipoprotein Oxidation. *FEBS Lett.* **1997**, *405*, 186–190.
45. Metodiewa, D.; Jaiswal, A.K.; Cenas, N.; Dickancaité, E.; Segura-Aguilar, J. Quercetin May Act as a Cytotoxic Prooxidant after its Metabolic Activation to Semiquinone and Quinoidal Product. *Free Radical Biol. Med.* **1999**, *26*, 107–116.
46. Dubey, R.K.; Tyurina, Y.Y.; Tyurin, V.A.; Gillespie, D.G.; Branch, R.A.; Jackson, E.K.; Kagan, V.E. Estrogen and Tamoxifen Metabolites Protect Smooth Muscle Cell Membrane Phospholipids Against Peroxidation and Inhibit Cell Growth. *Circ. Res.* **1999**, *84*, 229–239.
47. Kühn-Velten, W.N.; Pippirs, U. Novel Connections Between NADPH-Induced Lipid Peroxidation and Cytochrome P450 Inactivation, and Antioxidant and Enzyme Protective Properties of Estradiol in Gonadal Membranes. *Free Radical Res.* **1997**, *26*, 125–133.
48. Bolton, J.L.; Shen, L. *P*-quinone Methides Are the Major Decomposition Products of Catechol Estrogen *O*-quinone. *Carcinogenesis* **1996**, *17*, 925–929.
49. Lacort, M.; Leal, A.M.; Liza, M.; Martín, C.; Martínez, R.; Riuz-Larrea, M.B. Protective Effect of Estrogens and Catecholestrogens Against Peroxidative Membrane Damage in Vitro. *Lipids* **1995**, *30*, 141–146.
50. Han, X.; Liehr, J.G. Microsome-mediated 8-Hydroxylation of Guanine Bases of DNA by Steroid Estrogens: Correlation of DNA Damage by Free Radicals With Metabolic Activation to Quinones. *Carcinogenesis* **1995**, *16*, 2571–2574.
51. Markides, C.S.A.; Roy, D.; Liehr, J.G. Concentration Dependence of Prooxidant and Antioxidant Properties of

- Catecholestrogens. Arch. Biochem. Biophys. **1998**, *360*, 105–112.
52. Ferrari, R. Oxygen-free Radicals at Myocardial Level: Effects of Ischemia and Reperfusion. In *Free Radicals in Diagnostic Medicine. A Systems Approach to Laboratory Technology, Clinical Correlations, and Antioxidant Therapy*; Armstrong, D., Ed.; Plenum Press: New York, 1994; 366, 99–111.
53. Singal, P.K.; Kapur, N.; Dhillon, K.S.; Beamish, R.E.; Dhalla, N.S. Role of Free Radicals in Catecholamine-induced Cardiomyopathy. Can. J. Physiol. Pharmacol. **1982**, *60*, 1390–1397.
54. Moreno, A.J.; Santos, D.J.; Palmeira, C.M. Ischemic Heart Disease: The Role of Mitochondria-carvedilol Prevents Lipid Peroxidation of Mitochondrial Membranes. Rev. Port. Cardiol. **1998**, *17* (Suppl. 2), II63–77.
55. Facinó, R.M.; Carini, M.; Aldini, G.; Berti, F.; Rossoni, G.; Bombardelli, E.; Morazzoni, P. Procyanidines From *Vitis vinifera* Seeds Protect Rabbit Heart from Ischemia/Reperfusion Injury: Antioxidant Intervention And/or Iron and Copper Sequestering Ability. Planta Med. **1996**, *62*, 495–502.
56. Ray, P.S.; Maulik, G.; Cordis, G.A.; Bertelli, A.A.E.; Bertelli, A.; Das, D.K. The Red Wine Antioxidant Resveratrol Protects Isolated Rat Hearts from Ischemia Reperfusion Injury. Free Radical Biol. Med. **1999**, *27*, 160–169.
57. Maffei, F.R.; Carini, M.; Aldini, G.; Berti, F.; Rossoni, G. Panax Ginseng Administration in the Rat Prevents Myocardial Ischemia-reperfusion Damage Induced by Hyperbaric Oxygen: Evidence for an Antioxidant Intervention. Planta Med. **1999**, *65*, 614–619.
58. Burdon, R.H. Carcinogenesis and Free Radicals. In *Free Radicals: From Basic Science To Medicine, Molecular and Cell Biology Updates*; Poli, G., Albano, E., Dianzani, M.U. Eds.; Birkhäuser Verlag: Basel, 1993; 187–198.
59. Floyd, R.A. Role of Oxygen Free Radicals in Carcinogenesis and Brain Ischemia. FASEB J. **1990**, *4*, 2587–2597.
60. O'Brien, P.J. Antioxidants and Cancer: Molecular Mechanisms. In *Free Radicals in Diagnostic Medicine. A Systems Approach to Laboratory Technology, Clinical Correlations, and Antioxidant Therapy, Advances in Experimental Medicine and Biology*; Armstrong, D., Ed.; Plenum Press: New York, 1994; 366, 215–239.
61. Wang, M.; Dhingra, K.; Hittelman, W.N.; Liehr, J.G.; de Andrade, M.; Li, D. Lipid Peroxidation-induced Putative Malondialdehyde-DNA Adducts in Human Breast Tissues. Cancer Epidemiol. Biomarkers Prev. **1996**, *5*, 705–710.
62. Choi, M-A.; Kim, B-S.; Yu, R. Serum Antioxidative Vitamin Levels and Lipid Peroxidation in Gastric Carcinoma Patients. Cancer Lett. **1999**, *136*, 89–93.
63. Smith, M.A.; Harris, P.L.; Sayre, L.M.; Perry, G. Iron Accumulation in Alzheimer's Disease is a Source of Redox-Generated Free Radicals. Proc. Natl. Acad. Sci. U.S.A. **1997**, *94*, 9866–9868.
64. Keller, J.N.; Mark, R.J.; Bruce, A.J.; Blanc, E.; Rothstein, J.D.; Uchida, K.; Waeg, G.; Mattson, M.P. 4-Hydroxynonenal, an Aldehydic Product of Membrane Lipid Peroxidation, Impairs Glutamate Transport and Mitochondrial Function in Synaptosomes. Neuroscience **1997**, *80*, 685–696.
65. Walter, M.F.; Mason, P.E.; Mason, R.P. Alzheimer's Disease Amyloid β Peptide 25–35 Inhibits Lipid Peroxidation As a Result of Its Membrane Interactions. Biochem. Biophys. Res. Commun. **1997**, *233*, 760–764.
66. Smith, M.A.; Perry, G. The Role of Oxidative Stress in the Pathological Sequelae of Alzheimer's Disease. In *Free Radicals, Oxidative Stress, and Antioxidants. Pathological and Physiological Significance*; Özben, T., Ed.; Plenum Press: New York, 1998; 296, 195–204.
67. Lee, S.K.; Mbwambo, Z.H.; Chung, H.; Luyengi, L.; Gamez, E.J.; Meththa, R.G.; Kinghorn, A.D.; Pezzuto, J.M. Evaluation of the Antioxidant Potential of Natural Products. Comb. Chem. High Throughput Screen **1998**, *1*, 35–46.
68. Allen, R.C.; Loose, L.D. Phagocytic Activation of a Luminol-Dependent Chemiluminescence in Rabbit Alveolar and Peritoneal Macrophages. Biochem. Biophys. Res. Commun. **1976**, *69*, 245–253.
69. Yagi, K. A Simple Fluorometric Assay for Lipoperoxide in Blood Plasma. Biochem. Med. **1976**, *15*, 212–216.
70. Maly, F-E.; Nakamura, M.; Gauchat, J-F.; Urwyler, A.; Walker, C.; Dahinden, C.A.; Cross, A.R.; Jones, O.T.G.; Weck, A.L. Superoxide-dependent Nitroblue Tetrazolium Reduction and Expression of Cytochrome b_{245} Components by Human Tonsillar Lymphocytes and B Cell Lines. J. Immunol. **1989**, *142*, 1260–1267.
71. Collins, S.J.; Gallo, R.C.; Gallagher, R.E. Continuous Growth and Differentiation of Human Myeloid Leukemic Cells in Suspension Culture. Nature **1977**, *270*, 347–349.
72. Suh, N.; Luyengi, L.; Fong, H.; Kinghorn, A.D.; Pezzuto, J.M. Discovery of Natural Product Chemopreventive Agents Utilizing HL-60 Cell Differentiation as a Model. Anticancer Res. **1995**, *15*, 233–240.
73. Sheu, S-Y.; Lai, C-H.; Chiang, H-C. Inhibition of Xanthine Oxidase by Purpurogallin and Silymarin Group. Anticancer Res. **1998**, *18*, 263–267.
74. Kirby, A.J.; Schmidt, R.J. The Antioxidant Activity of Chinese Herbs for Eczema and of Placebo Herbs-I. J. Ethnopharmacol. **1997**, *56*, 103–108.

BIOPHARMACEUTICS

Leon Shargel

Eon Labs Manufacturing, Inc., Laurelton, New York, U.S.A.

Andrew B.C. Yu^a

U.S. Food and Drug Administration, Rockville, Maryland, U.S.A.

INTRODUCTION

Biopharmaceutics is the study of the interrelationship of the physicochemical properties of the drug [active pharmaceutical ingredient, (API)] and the drug product (dosage form in which the drug is fabricated) based on the biological performance of the drug (Table 1).

Biopharmaceutics also considers the impact of the various manufacturing methods and technologies on the intended performance of the drug product. Biopharmaceutics uses quantitative methods and theoretical models (1) to evaluate the effect of the drug substance, dosage form, and routes of drug administration on the therapeutic requirements of the drug and drug product in a physiological environment.

Bioavailability is often used as a measure of the biological performance of the drug and is defined as a measure of the rate and extent (amount) to which the active ingredient or active moiety becomes available at the site of action. Bioavailability is also a measure of the rate and extent of therapeutically active drug that is systemically absorbed.

Biopharmaceutics allows for rational design of drug products to deliver the drug at a specific rate to the body in order to optimize the therapeutic effect and minimize any adverse effects. As shown in Table 1, biopharmaceutics is based on the physicochemical characteristics of the active drug substance, the desired drug product, and considerations of the anatomy and physiology of the human body (1). Inherent in the design of a suitable drug product is knowledge of the pharmacodynamics of the drug, including the desired onset time, duration, and intensity of clinical response, and the pharmacokinetics of the drug including absorption, distribution, elimination, and target drug concentration.

Thus, biopharmaceutics involves factors that influence the: 1) protection and stability of the drug within the drug product; 2) the rate of drug release from the

drug product; 3) the rate of dissolution of the drug at the absorption site; and 4) the availability of the drug at its site of action (Fig. 1).

BIOPHARMACEUTIC CONSIDERATIONS IN DRUG PRODUCT DESIGN

Drugs are generally given to a patient as a manufactured drug product (finished dosage form) that includes the active drug and selected ingredients (excipients) that make up the dosage form. Common pharmaceutical dosage forms include liquids, tablets, capsules, injections, suppositories, transdermal systems, and topical drug products. The formulation and manufacture of a drug product requires a thorough understanding of the biopharmaceutics.

Each route of drug application presents special biopharmaceutic considerations in drug product design (Table 2). Systemic drug absorption from an extravascular site is influenced by the anatomic and physiologic properties of the site and the physicochemical properties of the drug and the drug product. The anatomy, physiology, and the contents of the gastrointestinal tract (GI) are considered in the design of a drug product for oral administration. For example, considerations in the design of a vaginal tablet formulation for the treatment of a fungus infection include whether the ingredients are compatible with vaginal anatomy and physiology, whether the drug is systemically absorbed from the vagina and how the vaginal tablet is to be properly inserted and placed in the appropriate area for optimum efficacy. Requirements for an eye medication include pH, isotonicity, sterility, local irritation to the cornea, draining of the drug by tears, and concern for systemic drug absorption. An additional consideration might be the contact time of the medication with the cornea. Although, increased eye contact time might be achieved by an increase in viscosity of the ophthalmic solution, the patient may lose some visual acuity when a viscous product is administered. Biopharmaceutic considerations for a drug administered

^aThe content in this article reflects the view of the authors and does not represent the view of FDA.

Table 1 Biopharmaceutic considerations in drug product design

Active pharmaceutical ingredient (API)	Stability Solubility pH and pKa Crystalline form (polymorph) Excipient interaction and compatability	Impurities Salt form Particle size Complexation
Drug product	Type of drug product (capsule, tablet, solution, etc.) Immediate or modified release Dosage strength Bioavailability	Stability Excipients Manufacturing variables
Physiologic factors	Route of administration Permeation of drug across cell membranes Binding to macromolecules	Blood flow Surface area Biotransformation
Pharmacodynamic and pharmacokinetic considerations	Bioavailability Therapeutic objective Adverse reactions	Pharmacokinetics Dose Toxic effects
Manufacturing considerations	Production methodology and technology Quality control/quality assurance Specification of raw materials	Cost Stability testing
Patient considerations	Compliance, labeling, and product acceptance	Cost

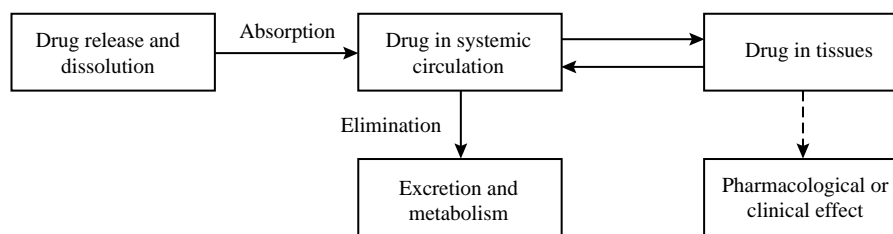


Fig. 1 Scheme demonstrating the dynamic relationships among the drug, the product, and pharmacologic effect. (From Ref. 1.)

by intramuscular injection include, local irritation, drug dissolution, and drug absorption from the injection site.

Biopharmaceutic studies may be performed using *in vitro* or *in vivo* methods (Table 3). *In vitro* methods are useful (2–6) to understand the physico-chemical properties of the drug and drug product and to evaluate the quality of the manufacturing process. Ultimately, the drug must be studied *in vivo*, in humans to assess drug efficacy, including the pharmacodynamic, pharmacokinetic, therapeutic and toxic profiles. Drug dissolution, absorption, metabolism, and potential interaction with food and other components in the GI tract are major biopharmaceutic topics for research and regulatory considerations in drug development.

A drug given by intravenous administration is considered complete or 100% bioavailable because the drug is placed directly into the systemic circulation. By carefully choosing the route of drug administration and proper design of the drug product, drug bioavailability can be varied from rapid and complete systemic drug absorption to a slow, sustained rate of absorption or even virtually no absorption, depending on the therapeutic objective. Once the drug is systemically absorbed, normal physiologic processes for distribution and elimination occur, which usually is not influenced by the specific formulation of the drug. The rate of drug release from the product, and the rate of drug absorption, are important in determining the onset, intensity, and duration of drug action of the drug.

RATE-LIMITING STEPS IN ORAL DRUG ABSORPTION

Systemic drug absorption from a drug product consists of a succession of rate processes (Fig. 2). For solid oral, immediate release drug products (e.g., tablet, capsule), the rate processes include 1) disintegration of the drug product and subsequent release of the drug; 2) dissolution of the drug in an aqueous environment; and 3) absorption across cell membranes into the systemic circulation. In the

process of drug disintegration, dissolution, and absorption, the rate at which drug reaches the circulatory system is determined by the slowest step in the sequence.

The slowest step in a kinetic process is the rate-limiting step. Except for controlled release products, disintegration of a solid oral drug product is usually more rapid than drug dissolution and drug absorption. For drugs that have very poor aqueous solubility, the rate at which the drug dissolves (dissolution) is often the slowest step, and therefore exerts a rate-limiting effect on drug bioavailability. In contrast, for a drug that has a high aqueous solubility, the dissolution rate is rapid and the rate at which the drug crosses or permeates cell membranes is the slowest or rate-limiting step.

PHYSIOLOGIC FACTORS AFFECTING DRUG ABSORPTION

Passage of Drugs Across Cell Membranes

For systemic absorption, a drug must pass from the absorption site through or around one or more layers of cells to gain access into the general circulation. The permeability of a drug at the absorption site into the systemic circulation is intimately related to the molecular structure of the drug and the physical and biochemical properties of the cell membranes. For absorption into the cell, a drug must traverse the cell membrane. Transcellular absorption is the process of a drug movement across a cell. Some polar molecules may not be able to traverse the cell membrane, but instead, go through gaps or “tight junctions” between cells, a process known as paracellular drug absorption. Some drugs are probably absorbed by a mixed mechanism involving one or more processes.

Passive diffusion

Passive diffusion is the process by which molecules spontaneously diffuse from a region of higher concentration to a region of lower concentration. This process is passive because no external energy is expended. Drug

Table 2 Common routes of drug administration

Route	Bioavailability	Advantages	Disadvantages
Parenteral routes			
Intravenous bolus (IV)	Complete (100%) systemic drug absorption. Rate of bioavailability considered instantaneous.	Drug is given for immediate effect.	Increased chance for adverse reaction. Possible anaphylaxis.
Intravenous infusion (IV inf)	Complete (100%) systemic drug absorption.	Plasma drug levels more precisely controlled. May inject large fluid volumes.	Requires skill in insertion of infusion set.
Intramuscular injection (IM)	Rate of drug absorption controlled by infusion pump. Rapid from aqueous solution.	May use drugs with poor lipid solubility and/or irritating drugs. Easier to inject than intravenous injection.	Tissue damage at site of injection (infiltration, necrosis, or sterile abscess). Irritating drugs may be very painful.
Subcutaneous injection (SC)	Slow absorption from nonaqueous (oil) solutions. Prompt from aqueous solution.	Larger volumes may be used compared to subcutaneous solution. Generally, used for insulin injection.	Different rates of absorption depending upon muscle group injected and blood flow. Rate of drug absorption depends upon blood flow and injection volume.
Enteral Routes			
Buccal or sublingual (SL)	Slow absorption from repository formulations. Rapid absorption from lipid-soluble drugs.	No "first-pass" effects.	Some drug may be swallowed. Not for most drugs or drugs with high doses.
Oral (PO)	Absorption may vary. Generally slower absorption rate compared to IV bolus or IM injection.	Safest and easiest route of drug administration. May use immediate-release and modified-release drug products.	Some drugs may have erratic absorption, be unstable in the gastrointestinal tract, or be metabolized by liver prior to systemic absorption.
Rectal (PR)	Absorption may vary from suppository.	Useful when patient cannot swallow medication.	Absorption may be erratic. Suppository may migrate to different position.
Other routes			
Transdermal	More reliable absorption from enema (solution). Slow absorption, rate may vary.	Used for local and systemic effects.	Some patient discomfort.
Inhalation	Increased absorption with occlusive dressing. Rapid absorption. Total dose absorbed is variable.	Transdermal delivery system (patch) is easy to use. Used for lipid-soluble drugs with low dose and low MW. May be used for local or systemic effects.	Some irritation by patch or drug. Permeability of skin variable with condition, anatomic site, age, and gender. Type of cream or ointment base affects drug release and absorption. Particle size of drug determines anatomic placement in respiratory tract. May stimulate cough reflex. Some drug may be swallowed.

(From Ref. 1.)

Table 3 Examples of in vitro and in vivo biopharmaceutic studies

Biopharmaceutic studies (in vivo)	Bioavailability study	Measurement of drug in plasma, urine or other tissues
	Acute pharmacologic effect	Measurement of a pharmacodynamic effect, e.g., FEV ₁ , blood pressure, heart rate, skin blanching
Biopharmaceutic studies (in vitro)	Clinical study	Measurement of drug efficacy
	Drug release/dissolution	Measurement of the rate of drug dissolved under specified conditions
	Drug permeability	Use of CACO2 cells (an isolated colon cell line) are grown into membranes to study the intestinal permeability and gut metabolism of drugs.
	Drug biotransformation (metabolism)	Use of liver cells, homogenates or isolated cytochrome P450 isozymes to drug study biotransformation.

molecules move randomly forward and back across a membrane (Fig. 3). If the two regions have the same drug concentration, forward-moving drug molecules will be balanced by molecules moving back, resulting in no net transfer of drug. For a region that has a higher drug concentration, the number of forward-moving drug molecules will be higher than the number of backward-moving molecules, resulting in a transfer of molecules to the region with the lower drug concentration, as indicated by the big arrow. Flux is the rate of drug transfer and is represented by a vector to show its direction. Molecules tend to move randomly in all directions because molecules possess kinetic energy and constantly collide with each other in space. Only left and right molecule movements are shown in Fig. 3, because movement of

molecules in other directions would not result in concentration changes because of the limitation of the container wall.

Passive diffusion is the major transmembrane process for most drugs. The driving force for passive diffusion is the difference in drug concentrations on either side of the cell membrane. According to Fick's Law of Diffusion, drug molecules diffuse from a region of high drug concentration to a region of low drug concentration

$$dQ/dt = \{DAK/h\}(C_{GI} - C_p)$$

where dQ/dt = rate of diffusion; D = diffusion coefficient; K = partition coefficient; A = surface area

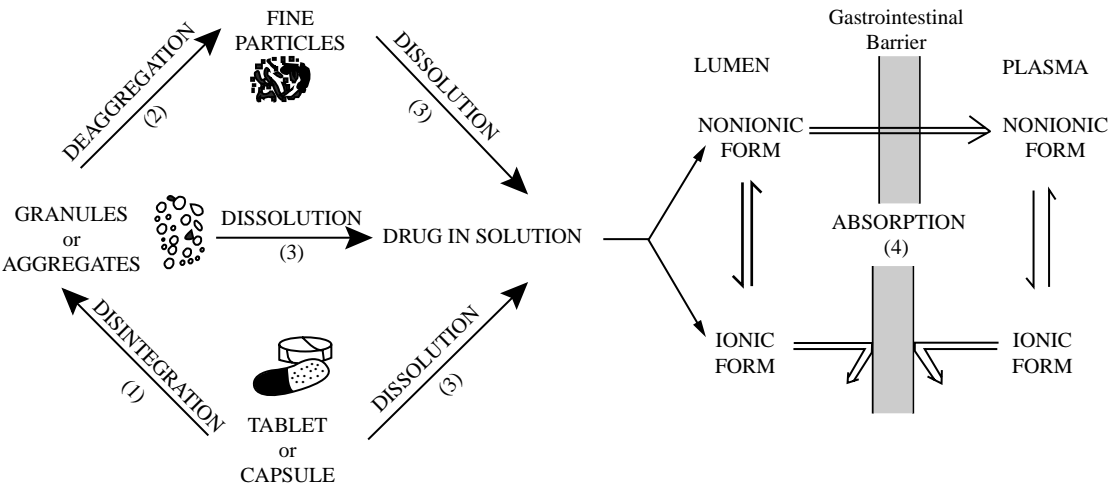


Fig. 2 Summary of processes involved following the oral administration of a drug in tablet or capsule form. (From Blanchard, J. Gastrointestinal absorption. II. Formulation factors affecting bioavailability. *Am. J. Pharm.* **1978**, *150*, 132–151.)

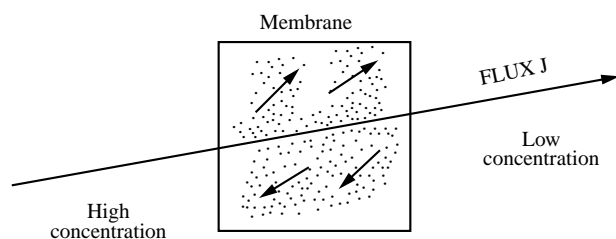


Fig. 3 Passive diffusion of molecules. Molecules in solution diffuse randomly in all directions. As molecules diffuse from left to right and vice versa (small arrows), a net diffusion from the high-concentration side to the low-concentration side results. This results in a net flux (J) to the right side. Flux is measured in mass per unit area (e.g., mg/cm^2). (From Ref. 1.)

of membrane; h = membrane thickness; and $C_{\text{GI}} - C_{\text{p}}$ = difference between the concentrations of drug in the GI tract and in the plasma.

Drug distributes rapidly into a large volume after entering the blood resulting in a very low plasma drug concentration with respect to the concentration at the site of drug administration. Drug is usually given in milligram doses, whereas plasma drug concentrations are often in the microgram per milliliter or nanogram per milliliter range. For drugs given orally, $C_{\text{GI}} \gg C_{\text{p}}$. A large concentration gradient is maintained driving drug molecules into the plasma from the GI tract.

As shown by Fick's Law of Diffusion, lipid solubility of the drug and the surface area and the thickness of the membrane influence the rate of passive diffusion of drugs. The partition coefficient, K , represents the lipid-water partitioning of a drug. More lipid soluble drugs have larger K values that theoretically increase the rate of systemic drug absorption. In practice, drug absorption is influenced by other physical factors of the drug, limiting its practical application of K . The surface area of the membrane through which the drug is absorbed directly influences the rate of drug absorption. Drugs may be absorbed from most areas of the GI tract. However, the duodenal area of the small intestine shows the most rapid drug absorption due to such anatomic features as villi and microvilli, which provide a large surface area. These villi are not found in such numbers in other areas of the GI tract.

The membrane thickness, h , is a constant at the absorption site but may be altered by disease. Drugs usually diffuse very rapidly into tissues through capillary cell membranes in the vascular compartments. In the brain, the capillaries are densely lined with glial cells creating a thicker lipid barrier (blood-brain barrier) causing a drug to diffuse more slowly into brain. In certain disease states (e.g., meningitis) the cell

membranes may be disrupted or become more permeable to drug diffusion.

Many drugs have lipophilic and hydrophilic substituents. More lipid soluble drug molecules traverse cell membranes more easily than less lipid-soluble (i.e., more water-soluble) molecules. For weak electrolyte drugs (i.e., weak acids, bases), the extent of ionization influences drug solubility and the rate of drug transport. Ionized drugs are more water soluble than nonionized drugs which are more lipid soluble. The extent of ionization of a weak electrolyte depends on the pK_a of the drug and the partition hypothesis (pH) of the medium in which the drug is dissolved. The Henderson and Hasselbalch equation describes the ratio of ionized (charged) to unionized form of the drug and is dependent on the pH conditions and the pK_a of the drug:

For weak acids,

$$\text{Ratio} = -\frac{(\text{salt})}{(\text{acid})} = \frac{(\text{A}^-)}{(\text{HA})} = 10^{(\text{pH}-\text{pK}_a)}$$

For weak bases,

$$\text{Ratio} = -(\text{base}) (\text{salt}) = (\text{RNH})_2 (\text{RNH}^{+3}) = 10^{(\text{pH}-\text{pK}_a)}$$

According to the pH , a weak acid (e.g., salicylic acid) should be rapidly absorbed from the stomach (pH 1.2) due to a favorable concentration gradient of the unionized (more lipid soluble) drug from the stomach to the blood, because practically all the drug in the blood compartment is dissociated (ionized) at pH 7.4. A weak base (e.g., quinidine) is highly ionized in acid pH and is poorly absorbed from the stomach. Although many drugs obey by the pH , in practice, the major site of absorption of most drugs is usually in the small intestine (duodenum) due presence of a large surface area and high blood flow.

The drug concentration on either side of a membrane is also influenced by the affinity of the drug for a tissue component, which prevents the drug from freely moving back across the cell membrane. For example, drug that binds plasma or tissue proteins causes the drug to concentrate in that region. Dicumarol and sulfonamides strongly bind plasma proteins; whereas, chlordane, a lipid-soluble insecticide, partitions and concentrates into adipose (fat) tissue. Tetracycline forms a complex with calcium and concentrates in the bones and teeth. Drugs may concentrate in a tissue due to a specific uptake or active transport process. Such processes have been demonstrated for iodide in thyroid tissue, potassium in the intracellular water, and certain catecholamines in adrenergic storage sites.

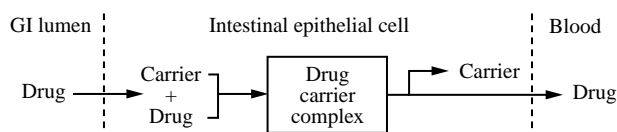


Fig. 4 Hypothetical carrier-mediated transport process. (From Ref. 1.)

Carrier-mediated transport

Theoretically, a lipophilic drug may pass through the cell or go around it. If drug has a low molecular weight and is lipophilic, the lipid cell membrane is not a barrier to drug diffusion and absorption. In the intestine, molecules smaller than 500 MW may be absorbed by paracellular drug absorption. Numerous specialized carrier-mediated transport systems are present in the body especially in the intestine for the absorption of ions and nutrients required by the body.

Active transport: Active transport is a carrier-mediated transmembrane process that is important for GI absorption of some drugs and also involved in the renal and biliary secretion of many drugs and metabolites. A carrier binds the drug to form a carrier–drug complex that shuttles the drug across the membrane and then dissociates the drug on the other side of the membrane (Fig. 4). Active transport is an energy-consuming system characterized by the transport of drug against a concentration gradient, that is, from regions of low drug concentrations to regions of high concentrations.

A drug may be actively transported, if the drug molecule structurally resembles a natural substrate that is actively transported. A few lipid-insoluble drugs that resemble natural physiologic metabolites (e.g., 5-fluorouracil) are absorbed from the GI tract by this process. Drugs of similar structure may compete for adsorption sites on the carrier. Because only a certain amount of carrier is available, the binding sites on the carrier may become saturated at high drug concentrations. In contrast, passive diffusion is not saturable.

Facilitated diffusion: Facilitated diffusion is a non-energy requiring, carrier-mediated transport system in which the drug moves along a concentration gradient (i.e., moves from a region of high drug concentration to a region of low drug concentration). Facilitated diffusion is saturable, structurally selective for the drug and shows competition kinetics for drugs of similar structure. Facilitated diffusion seems to play a very minor role in drug absorption.

Carrier-mediated intestinal transport: Various carrier mediated systems (transporters) are present at the intestinal brush border and basolateral membrane for

the absorption of specific ions and nutrients essential for the body. Many drugs are absorbed by these carriers because of the structural similarity to natural substrates. An intestinal transmembrane protein, *P*-Glycoprotein (*P*-Gp) appears to reduce apparent intestinal epithelial cell permeability from lumen to blood for various lipophilic or cytotoxic drugs. Other transporters are present in the intestines. For example, many oral cephalosporins are absorbed through the amino acid transporter.

Vesicular transport

Vesicular transport is the process of engulfing particles or dissolved materials by the cell. Pinocytosis refers to the engulfment of small solutes or fluid, whereas phagocytosis refers to the engulfment of larger particles or macromolecules generally by macrophages. Endocytosis and exocytosis are the processes of moving macromolecules into and out of a cell, respectively.

During pinocytosis or phagocytosis, the cell membrane invaginates to surround the material, and then engulfs the material into the cell. Subsequently, the cell membrane containing the material forms a vesicle or vacuole within the cell. Vesicular transport is the proposed process for the absorption of orally administered sabin polio vaccine and various large proteins. An example of exocytosis is the transport of a protein such as insulin from insulin-producing cells of the pancreas into the extracellular space. The insulin molecules are first packaged into intracellular vesicles, which then fuse with the plasma membrane to release the insulin outside the cell.

ORAL DRUG ABSORPTION

Physiologic Considerations

Drugs may be administered by various routes of administration (Table 2). Except for intravenous drug administration, drugs are absorbed into the systemic circulation from the site of administration and are greatly affected by conditions at the administration site.

Oral administration is the most common route of drug administration. Major physiologic processes in the GI system include secretion, digestion, and absorption. Secretion includes the transport of fluid, electrolytes, peptides, and proteins into the lumen of the alimentary canal. Enzymes in saliva and pancreatic secretions are involved in the digestion of carbohydrates and proteins. Other secretions such as mucus protect the linings of the lumen of the GI tract. Digestion is the breakdown of food

constituents into smaller structures in preparation for absorption. Both drug and food constituents are mostly absorbed in the proximal area (duodenum) of the small intestine. The process of absorption is the entry of constituents from the lumen of the gut into the body. Absorption may be considered as the net result of both lumen-to-blood and blood-to-lumen transport movements.

Drugs administered orally pass through various parts of the enteral canal including the oral cavity, esophagus, and various parts of the GI tract. Residues eventually exit the body through the anus. Drugs may be absorbed by passive diffusion from all parts of the alimentary canal including sublingual, buccal, GI, and rectal absorption. For most drugs, the optimum site for drug absorption after oral administration is the upper portion of the small intestine or duodenum region. The unique anatomy of the duodenum provides an immense surface area for the drug to passively diffuse (Table 4). In addition, the duodenal region is highly perfused with a network of capillaries, which helps to maintain a concentration gradient from the intestinal lumen and plasma circulation.

The total transit time, including gastric emptying, small intestinal transit, and colonic transit ranges from 0.4 to 5 days. Small intestine transit time (SITT) ranges from 3 to 4 h for most healthy subjects. If absorption is not completed by the time a drug leaves the small intestine, drug absorption may be erratic or incomplete. The small intestine is normally filled with digestive juices and liquids, keeping the lumen contents fluid. In contrast, the fluid in the colon is reabsorbed, and the lumen content in the colon is either semisolid or solid, making further drug dissolution erratic and difficult.

Gastrointestinal motility

Once the drug is given orally, the exact location and/or environment of the drug product within the GI tract is difficult to discern. GI motility tends to move the drug through the alimentary canal so that it may not stay at the absorption site. For drugs given orally, an anatomic absorption window may exist within the GI tract in which the drug is efficiently absorbed. Drugs contained in a nonbiodegradable controlled-release dosage form must be completely released into this absorption window prior to the movement of the dosage form into the large bowel. The transit time of the drug in the GI tract depends upon the pharmacologic properties of the drug, type of dosage form, and various physiologic factors. Physiologic movement of the drug within the GI tract depends upon whether the alimentary canal contains recently ingested food (digestive or fed state) or is in the fasted or interdigestive state.

Gastric emptying time

After oral administration, the swallowed drug rapidly reaches the stomach. Because the duodenum has the greatest capacity for the absorption of drugs from the GI tract, a delay in the gastric emptying time will slow the rate and possibly the extent of drug absorption from the duodenum, thereby prolonging the onset time for the drug. Drugs, such as penicillin, that are unstable in acid, may decompose if stomach emptying is delayed. Other drugs, (e.g., aspirin) may irritate the gastric mucosa during prolonged contact.

Factors that tend to delay gastric emptying include consumption of meals high in fat, cold beverages, and anticholinergic drugs. Liquids and small particles less than 1 mm are generally not retained in the stomach. These small particles are believed to be emptied due to a slightly higher basal pressure in the stomach over the duodenum. Different constituents of a meal will empty from the stomach at different rates. For example, liquids are generally emptied faster than digested solids from the stomach. Large particles, including tablets and capsules, are delayed from emptying for 3–6 h by the presence of food in the stomach. Indigestible solids empty very slowly, probably during the interdigestive phase, a phase in which food is not present and the stomach is less motile but periodically empties its content due to housekeeper wave contraction.

Intestinal motility

Normal peristaltic movements mix the contents of the duodenum, bringing the drug particles into intimate contact with the intestinal mucosal cells. The drug must have a sufficient time (residence time) at the absorption site for optimum absorption. In the case of high motility in the intestinal tract, as in diarrhea, the drug has a very brief residence time and less opportunity for adequate absorption.

Blood perfusion of the gastrointestinal tract

The blood flow is important in carrying the absorbed drug from the absorption site to the systemic circulation. A large network of capillaries and lymphatic vessels perfuse the duodenal region and peritoneum. The splanchnic circulation receives about 28% of the cardiac output and is increased after meals. Drugs are absorbed from the small intestine into the mesenteric vessels which flows to the hepatic-portal vein and then to the liver prior to reaching the systemic circulation. Any decrease in mesenteric blood flow, as in the case of congestive heart failure, will decrease the rate of systemic drug absorption from the intestinal tract.

Table 4 Drug absorption in the gastrointestinal tract

Anatomic area	Function	Affect on drug absorption
Oral cavity	Saliva, pH 7, contains ptyalin (salivary amylase), digests starches. Mucin, a glycoprotein, lubricates food and may interact with drugs.	Buccal and sublingual absorption occurs for lipid-soluble drugs.
Esophagus	The esophagus connects the pharynx and the cardiac orifice of the stomach. The pH is 5–6. The lower part of the esophagus ends with the esophageal sphincter, which prevents acid reflux from the stomach.	Tablets or capsules may lodge in this area, causing local irritation. Very little drug dissolution occurs in the esophagus.
Stomach	The fasting stomach pH is about 2 to 6. In the fed state, the stomach pH is about 1.5 to 2, due to hydrochloric acid secreted by parietal cells. Stomach acid secretion is stimulated by gastrin and histamine. Mixing is intense and pressurized in the antral part of the stomach, a process of breaking down large food particles described as antral milling. Food and liquid are emptied by opening the pyloric sphincter into the duodenum.	Drugs are not efficiently absorbed in the stomach. Basic drugs are solubilized rapidly in acid. Stomach emptying influences the time for drug reaching the small intestine. The food content and osmolality influenced by stomach emptying. Fatty acids delay gastric emptying. High-density foods generally are emptied more slowly from the stomach.
Duodenum	A common duct from the pancreas and gall bladder enters the duodenum. Duodenal pH is 6 to 6.5 due to the presence of bicarbonate that neutralizes the acidic chyme emptied from the stomach. The pH is optimum for enzymatic digestion of protein and peptide food. Pancreatic juice containing enzymes is secreted into the duodenum from the bile duct. Trypsin, chymotrypsin, and carboxypeptidase are involved in the hydrolysis of proteins into amino acids. Amylase is involved in the digestion of carbohydrates. Pancreatic lipase secretion hydrolyzes fats into fatty acids.	The main site for drug absorption. An immense surface area for the passive diffusion of drug to due to the presence of villi and microvilli forming a brush border. A high blood perfusion maintains a drug concentration gradient from the intestinal lumen and plasma circulation. The complex fluid medium in the duodenum dissolves many drugs with limited aqueous solubility. Ester prodrugs are hydrolyzed during absorption. Proteolytic enzymes degrade many protein drugs in the duodenum, preventing adequate absorption. Acid drugs dissolve in the alkaline pH. Bile secretion helps to dissolve fats and hydrophobic drugs
Jejunum	The jejunum is the middle portion of the small intestine in between the duodenum and the ileum. Digestion of protein and carbohydrates continues after receiving pancreatic juice and bile in the duodenum, this portion of the small intestine generally has less contraction than the duodenum and is preferred for in vivo drug absorption studies.	Drugs generally absorbed by passive diffusion.
Ileum	The ileum, pH about 7, with the distal part as high as 8, is the terminal part of the small intestine and has fewer contractions than the duodenum. The ileocecal valve separates the small intestine with the colon.	Drugs generally absorbed by passive diffusion.
Colon	The colon, pH 5.5–7, is lined with mucin functioning as lubricant and protectant. The colon contains both aerobic and anaerobic micro-organisms that may metabolize some drugs. Crohn's disease affects the colon and thickens the bowel wall. The microflora may also become more anaerobic. Absorption of clindamycin and propranolol are increased, whereas other drugs have reduced absorption with this disease (Rubinstein et al. 1988).	Very limited drug absorption due to the lack of microvilli and the more viscous and semisolid nature of the lumen contents. A few drugs such as theophylline and metoprolol are absorbed in this region. Drugs that are absorbed well in this region are good candidates for an oral sustained-release dosage form.

(Continued)

Table 4 Drug absorption in the gastrointestinal tract (*Continued*)

Anatomic area	Function	Affect on drug absorption
Rectum	<p>The rectum is about 15 cm long, ending at the anus.</p> <p>In the absence of fecal material, the rectum has a small amount of fluid, (about 2 m) with a pH about 7.</p> <p>The rectum is perfused by the superior, middle, and inferior hemorrhoidal veins. The inferior hemorrhoidal vein (closest to the anal sphincter) and the middle hemorrhoidal vein feed into the vena cava and back to the heart. The superior hemorrhoidal vein joins the mesenteric circulation, which feeds into the hepatic portal vein and then to the liver.</p>	<p>Drug absorption may be variable depending upon the placement of the suppository or drug solution within the rectum. A portion of the drug dose may be absorbed via the lower hemorrhoidal veins, from which the drug feeds directly into the systemic circulation; some drug may be absorbed via the superior hemorrhoidal veins, which feeds into the mesenteric veins to the hepatic portal vein to the liver, and metabolized prior to systemic absorption.</p>

Some drugs may be absorbed into the lymphatic circulation through the lacteal or lymphatic vessels under the microvilli. Absorption of drugs through the lymphatic system bypasses the first-pass effect due to liver metabolism, because drug absorption through the hepatic portal vein is avoided. The lymphatics are important in the absorption of dietary lipids and may be partially responsible for the absorption for some lipophilic drugs such as bleomycin or aclarubicin which may dissolve in chylomicrons and be systemically absorbed via the lymphatic system.

Effect of food and other factors on GI drug absorption

Digested foods may affect intestinal pH and solubility of drugs. Food effects are not always predictable. The absorption of some antibiotics (e.g., penicillin, tetracycline) is decreased with food, whereas other drugs (e.g., griseofulvin) are better absorbed when given with food containing a high fat content. Food in the GI lumen stimulates the flow of bile. Bile contains bile acids. Bile acids are surfactants are involved in the digestion and solubilization of fats, and increases the solubility of fat-soluble drugs through micelle formation. For some basic drugs (e.g., cinnarizine) with limited aqueous solubility, the presence of food in the stomach stimulates hydrochloric acid secretion, which lowers the pH, causing more rapid dissolution of the drug and better absorption.

Generally, the bioavailability of drugs is better in patients in the fasted state and with a large volume of water (Fig. 5). However, to reduce GI mucosal irritation, drugs such as erythromycin, iron salts, aspirin, and nonsteroidal anti-inflammatory agents (NSAIDs) are given with food. The rate of absorption for these drugs may be reduced in the presence of food, but the extent of absorption may be the same.

The drug dosage form may also be affected by food. For example, enteric-coated tablets may stay in the stomach for a longer period of time because food delays stomach emptying. If the enteric-coated tablet does not reach the duodenum rapidly, drug release and subsequent systemic drug absorption are delayed. In contrast, enteric-coated beads or microparticles disperse in the stomach, are less affected by food, and demonstrate more consistent drug absorption from the duodenum.

Food may also affect the integrity of the dosage form, causing an alteration in the release rate of the drug. For example, theophylline bioavailability from Theo-24 controlled-release tablets is much more rapid (7) when given to a subject in the fed rather than fasted state (Fig. 6).

Some drugs, such as ranitidine, cimetidine, and dipyridamole, after oral administration produce a blood concentration curve consisting of two peaks. This double-peak phenomenon is generally observed after the administration of a single dose to fasted patients. The rationale for the double-peak phenomenon has been attributed to variability in stomach emptying, variable intestinal motility, presence of food, enterohepatic recycling, or failure of a tablet dosage form. For a drug with high water solubility, dissolution of the drug occurs in the stomach, and partial emptying of the drug into the duodenum will result in the first absorption peak. A delay in stomach emptying results in a second absorption peak as the remainder of the dose is emptied into the duodenum.

Diseases such as Crohn's disease that alter GI physiology and corrective surgery involving peptic ulcer, antrectomy with gastroduodenostomy and selective vagotomy may potentially affect drug absorption. Drug absorption may be unpredictable in many disease conditions. Drugs or nutrients or both may also affect the absorption of other drugs. For example, propantheline

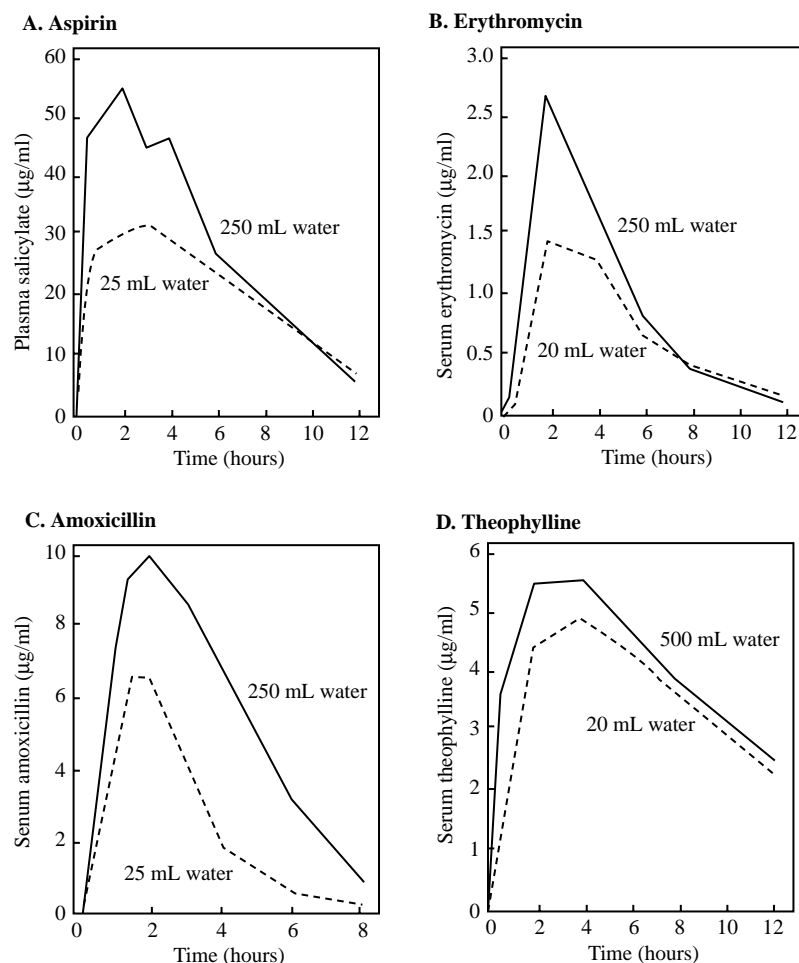


Fig. 5 Mean plasma or serum drug levels in healthy, fasting human volunteers ($n = 6$ in each case) who received single oral doses of aspirin (650 mg) tablets, erythromycin stearate (500 mg) tablets, amoxicillin (500 mg) capsules, and theophylline (260 mg) tablets, together with large. (From Welling P.G.; Drug Bioavailability and Its Clinical Significance. *Progress in Drug Metabolism*, Vol. 4; Bridges K.W.; Chassea, VD LF. Eds.; Wiley; London, 1980.)

bromide is an anticholinergic drug that slows stomach emptying and motility of the small intestine and may reduce stomach acid secretion. Grapefruit juice was found to increase the plasma level of many drugs due to inhibition of their metabolism in the liver.

PHARMACEUTICAL FACTORS AFFECTING DRUG BIOAVAILABILITY

Biopharmaceutic considerations in the design and manufacture of a drug product to deliver the active drug with the desired bioavailability characteristics include: 1) the type of drug product (e.g., solution,

suspension, suppository), 2) the nature of the excipients in the drug product, 3) the physicochemical properties of the drug molecule, and 4) the route of drug administration.

Disintegration

Immediate release, solid oral drug products must rapidly disintegrate into small particles and release the drug. The United States Pharmacopoeia (USP) describes an official tablet disintegration test. The process of disintegration does not imply complete dissolution of the tablet and/or the drug. Complete disintegration is defined by the USP as "that state in which any residue of the tablet, except fragments of insoluble coating, remaining on the screen of

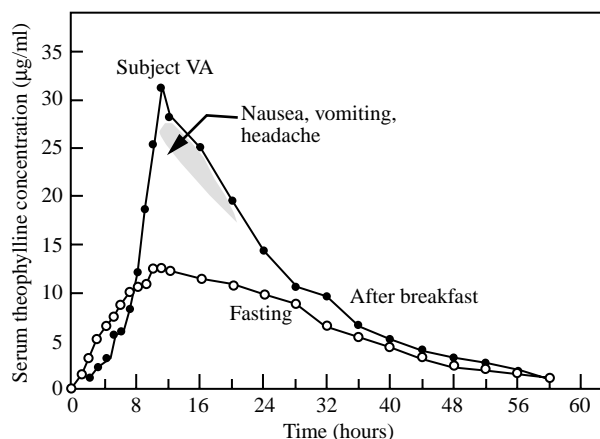


Fig. 6 Theophylline serum concentration in an individual subject after a single 1500 mg dose of Theo-24 taken during fasting, period during which this patient experienced nausea, repeated vomiting, or severe throbbing headache. The pattern of drug release during the food regimen is consistent with “dose-dumping.” (From Ref. 7.)

the test apparatus in the soft mass have no palpably firm core.” The USP provides specifications for uncoated tablets, plain coated tablets, enteric tablets, buccal tablets, and sublingual tablets. Exempted from USP disintegration tests are troches, tablets which are intended to be chewed, and drug products intended for sustained release or prolonged or repeat action.

Disintegration tests allow for precise measurement of the formation of fragments, granules, or aggregates from solid dosage forms, but do not provide information on the dissolution rate of the active drug. The disintegration test serves as a component in the overall quality control of tablet manufacture.

Dissolution

Dissolution is the process by which a chemical or drug becomes dissolved in a solvent. In biologic systems, drug dissolution in an aqueous medium is an important prior condition of systemic absorption. The rate at which drugs with poor aqueous solubility dissolve from an intact or disintegrated solid dosage form in the GI tract often controls the rate of systemic absorption of the drug. Thus, dissolution tests are discriminating of formulation factors that may affect drug bioavailability.

As the drug particle dissolves, a saturated solution (stagnant layer) is formed at the immediate surface around the particle. The dissolved drug in the saturated solution gradually diffuses to the surrounding regions. The overall rate of drug dissolution may be described by the Noyes–Whitney equation which models drug dissolution in terms

of the rate of drug diffusion from the surface to the bulk of the solution. In general, drug concentration at the surface is assumed to be the highest possible, i.e., the solubility of the drug in the dissolution medium. The drug concentration C is the homogeneous concentration in the bulk solution which is generally lower than that in the stagnant layer immediate to the surface of the solid. The decrease in concentration across the stagnant layer is called the diffusion gradient

$$dC/dt = DA(CS - C)h$$

where, dC/dt = rate of drug dissolution, D = diffusion rate constant, A = surface area of the particle, CS = drug concentration in the stagnant layer, C = drug concentration in the bulk solvent, and h = thickness of the stagnant layer.

The rate of dissolution, $(dC/dt) \times (1/A)$, is the amount of drug dissolved per unit area per time (e.g., g/cm² per min).

The Noyes–Whitney equation shows that dissolution rate is influenced by the physicochemical characteristics of the drug, the formulation, and the solvent. In addition, the temperature of the medium also affects drug solubility and dissolution rate.

PHYSICOCHEMICAL NATURE OF THE DRUG

Solubility, pH, and Drug Absorption

The natural pH environment of the GI tract varies from acidic in the stomach to slightly alkaline in the small intestine. Drug solubility may be improved with the addition of acidic or basic excipients. Solubilization of aspirin, for example, may be increased by the addition of an alkaline buffer. Controlled release drug products are nondisintegrating dosage forms. Buffering agents may be added to slow or modify the release rate of a fast-dissolving drug in the formulation of a controlled release drug product. The buffering agent is released slowly rather than rapidly so that the drug does not dissolve immediately in the surrounding GI fluid. Intravenous drug solutions are difficult to prepare with drugs that have poor aqueous solubility. Drugs that are physically or chemically unstable may require special excipients, coating or manufacturing process to protect the drug from degradation.

Stability, pH, and Drug Absorption

The pH-stability profile is a plot of reaction rate constant for drug degradation versus pH and may help to predict if

some of the drug will decompose in the GI tract. The stability of erythromycin is pH-dependent. In acidic medium, erythromycin decomposition occurs rapidly, whereas at neutral or alkaline pH the drug is relatively stable. Consequently, erythromycin tablets are enteric coated to protect against acid degradation in the stomach. In addition, less soluble erythromycin salts that are more stable in the stomach have been prepared.

Particle Size and Drug Absorption

The effective surface area of the drug is increased enormously by a reduction in the particle size. Because drug dissolution is thought to take place at the surface of the solute, the greater the surface area, the more rapid the rate of drug dissolution. The geometric shape of the drug particle also affects the surface area, and during dissolution the surface is constantly changing. In dissolution calculations, the solute particle is usually assumed to have retained its geometric shape.

Particle size and particle size distribution studies are important for drugs that have low water solubility. Particle size reduction by milling to a micronized form increased the absorption of low aqueous solubility drugs such as griseofulvin, nitrofurantoin, and many steroids. Smaller particle size results in an increase in the total surface area of the particles, enhances water penetration into the particles, and increases the dissolution rates. With poorly soluble drugs, a disintegrant may be added to the formulation to ensure rapid disintegration of the tablet and release of the particles.

Polymorphic Crystals, Solvates, and Drug Absorption

Polymorphism refers to the arrangement of a drug in various crystal forms (polymorphs). Polymorphs have the same chemical structure but different physical properties, such as solubility, density, hardness, and compression characteristics. Some polymorphic crystals may have much lower aqueous solubility than the amorphous forms, causing a product to be incompletely absorbed. Chloramphenicol (9), for example, has several crystal forms, and when given orally as a suspension, the drug concentration in the body depended on the percentage of β -Polymorph in the suspension. The β -form is more soluble and better absorbed (Fig. 7). In general, the crystal form that has the lowest free energy is the most stable polymorph. Polymorphs that are metastable may convert to a more stable form over time. A crystal form change may cause problems in manufacturing the product. For example, a

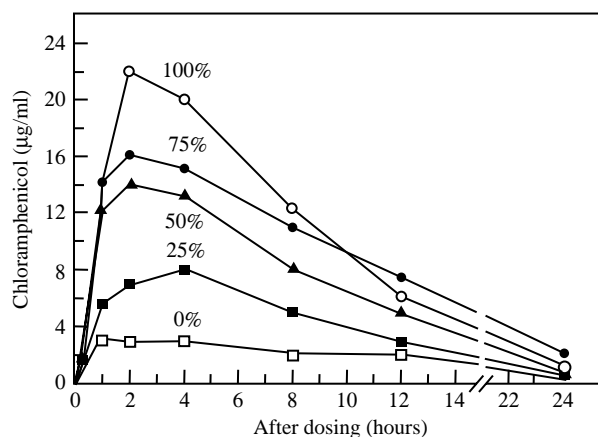


Fig. 7 Comparison of mean blood serum levels obtained with chloramphenicol palmitate suspensions containing varying ratios of α and β polymorphs, following single oral dose equivalent. (From Ref. 9.)

change in crystal structure of the drug may cause cracking in a tablet or even prevent a granulation to be compressed into a tablet requiring reformulation of the product. Some drugs interact with solvent during preparation to form a crystal called solvate. Water may form a special crystal with drugs called hydrates, for example, erythromycin forms different hydrates (8) which may have quite different solubility compared to the anhydrous form of the drug (Fig. 8). Ampicillin trihydrate, for example, was reported

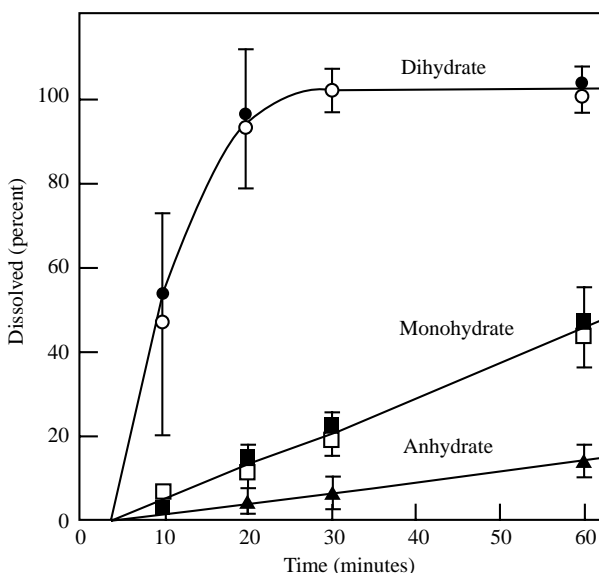


Fig. 8 Dissolution behavior of erythromycin dihydrate, monohydrate, and anhydrate in phosphate buffer (pH 7.5) at 37°C. (From Ref. 8.)

Table 5 Common excipients used in solid drug products

Excipient	Property in dosage form
Lactose	Diluent
Dibasic calcium phosphate	Diluent
Starch	Disintegrant, diluent
Microcrystalline cellulose	Disintegrant, diluent
Magnesium stearate	Lubricant
Stearic acid	Lubricant
Hydrogenated vegetable oil	Lubricant
Talc	Lubricant
Sucrose (solution)	Granulating agent
Polyvinyl pyrrolidone (solution)	Granulating agent
Hydroxypropylmethylcellulose	Tablet-coating agent
Titanium dioxide	Combined with dye as colored coating
Methylcellulose	Coating or granulating agent
Cellulose acetate phthalate	Enteric coating agent

(From Ref. 1.)

to be less absorbed than the anhydrous form of ampicillin due to faster dissolution of the latter.

FORMULATION FACTORS AFFECTING DRUG DISSOLUTION

Excipients are pharmacodynamically inactive substances that are added to a formulation to provide certain functional properties to the drug and dosage form. Excipients may be added to improve the compressibility of the active drug, stabilize the drug from degradation, decrease gastric irritation, control the rate of drug absorption from the absorption site, increase drug bioavailability, etc. Some excipients used in the

Table 6 Common excipients used in oral liquid drug products

Excipient	Property in dosage form
Sodium carboxymethylcellulose	Suspending agent
Tragacanth	Suspending agent
Sodium alginate	Suspending agent
Xanthan gum	Thixotropic suspending agent
Veegum	Thixotropic suspending agent
Sorbitol	Sweetener
Alcohol	Solubilizing agent, preservative
Propylene glycol	Solubilizing agent
Methyl propylparaben	Preservative
Sucrose	Sweetener
Polysorbates	Surfactant
Sesame oil	For emulsion vehicle
Corn oil	For emulsion vehicle

(From Ref. 1.)

manufacture of solid and liquid drug products are listed in Tables 5 and 6. For solid oral dosage forms such as compressed tablets, excipients may include 1) diluent (e.g., lactose), 2) disintegrant (e.g., starch), 3) lubricant (e.g., magnesium stearate), and 4) other components such as binding and stabilizing agents. When improperly used in the formulation, excipients may alter drug bioavailability and possibly pharmacodynamic activity.

Excipients may affect the drug dissolution rate by altering the medium in which the drug is dissolving or by reacting with the drug itself. Some common manufacturing problems that affect drug dissolution and bioavailability are listed in Table 7. For example,

Table 7 Effect of excipients on the pharmacokinetic parameters of oral drug product^a

Excipients	Example	k_a	t_{max}	AUC
Disintegrants	Avicel, Explotab	↑	←	↑/—
Lubricants	Talc, hydrogenated vegetable oil	←	↑	←/—
Coating agent	Hydroxypropylmethyl cellulose	—	—	—
Enteric coat	Cellulose acetate phthalate	←	↑	←/—
Sustained-release agents	Methylcellulose, ethylcellulose	←	↑	←/—
Sustained-release agents (waxy agents)	Castorwax, Carbowax	←	↑	←/—
Sustained-release agents (gum/viscous)	Veegum, Keltrol	←	↑	←/—

^aThis may be concentration and drug dependent.

↑ = Increase, ← = decrease, — = no effect. k_a = absorption rate constant, t_{max} = time for peak drug concentration in plasma, AUC = area under the plasma drug concentration time curve.

(From Ref. 1.)

suspending agents increase the viscosity of the drug vehicle, but may decrease the drug dissolution rate from the suspension. An excessive quantity of magnesium stearate (a hydrophobic lubricant) in the formulation may retard drug dissolution and slow the rate of drug absorption. The total amount of drug absorbed may also be reduced. To prevent this problem, the lubricant level should be decreased or a different lubricant selected. Sometimes, increasing the amount of disintegrant may overcome the retarding effect of lubricants on dissolution. However, with some poorly soluble drugs an increase in disintegrant level has little or no effect on drug dissolution because the fine drug particles are not wetted. The general influence of some common excipients on drug bioavailability parameters for typical oral drug products is summarized in Table 7.

Excipients may enhance or diminish the rate and extent of systemic drug absorption. Excipients that increase the aqueous solubility of the drug generally increase the rate of drug dissolution and absorption. For example, sodium bicarbonate in the formulation may change the pH of the medium surrounding the active drug substance. Aspirin, a weak acid, in an alkaline medium will form a water-soluble salt in which the drug rapidly dissolves. This process is known as dissolution in a reactive medium. The solid drug dissolves rapidly in the reactive solvent surrounding the solid particle. As the dissolved drug molecules diffuse outward into the bulk solvent, the drug may precipitate out of solution with a very fine particle size. The small particles have enormous collective surface area and disperse and redissolve readily for more rapid absorption on contact with the mucosal surface.

Excipients may interact directly with the drug to form a water-soluble or water-insoluble complex. If tetracycline is formulated with calcium carbonate, an insoluble complex of calcium tetracycline is formed that has a slow rate of dissolution and poor absorption.

Excipients may increase the retention time of the drug in the GI tract and therefore increase the amount of drug absorbed. Excipients may act as carriers to increase drug diffusion across the intestinal wall. The addition of surface-active agents may increase wetting as well as solubility of drugs. In contrast, many excipients may retard drug dissolution and thus reduce drug absorption.

Shellac used as a tablet coating, upon aging, can slow the drug dissolution rate. Surfactants may affect drug dissolution in an unpredictable fashion. Low concentrations of surfactants lower the surface tension and increase the rate of drug dissolution, whereas higher concentrations of surfactants tend to form micelles with

the drug and thus decrease the dissolution rate. High tablet compression without sufficient disintegrant may cause poor disintegration in vivo of a compressed tablet.

IN VITRO DISSOLUTION TESTING

A dissolution test in vitro measures the rate and extent of dissolution of the drug in an aqueous medium in the presence of one or more excipients contained in the drug product. A potential bioavailability problem may be uncovered by a suitable dissolution method. The optimum dissolution testing conditions differ with each drug formulation. Different agitation rates, different medium (including different pH), and different dissolution apparatus should be tried to distinguish which dissolution method is optimum for the drug product and discriminating for drug formulation changes. The appropriate dissolution test condition for the drug product is then used to determine acceptable dissolution specifications.

The size and shape of the dissolution vessel may affect the rate and extent of dissolution. For example, the vessel may range in size from several milliliters to several liters. The shape may be round-bottomed or flat, so that the tablet might lie in a different position in different experiments. The amount of agitation and the nature of the stirrer affect the dissolution rate. Stirring rates must be controlled, and specifications differ between drug products. Low stirring rates (50–100 rpm) are more discriminating of formulation factors affecting dissolution than higher stirring rates. The temperature of the dissolution medium must be controlled and variations in temperature must be avoided. Most dissolution tests are performed at 37°C.

The nature of the dissolution medium, the solubility of the drug and the amount of drug in the dosage form will affect the dissolution test. The dissolution medium should not be saturated by the drug. Usually, a volume of medium larger than the amount of solvent needed to completely dissolve the drug is used in such tests. The usual volume of the medium is 500–1000 ml. Drugs that are not very water soluble may require use of a very-large-capacity vessel (up to 2000 ml) to observe significant dissolution. Sink conditions is a term referring to an excess volume of medium that allows the solid drug to continuously dissolve. If the drug solution becomes saturated, no further net drug dissolution will take place. According to the USP, “the quantity of medium used should be not less than three times that required to form a saturated solution of the drug substance.”

Which medium is best is a matter of considerable controversy. The preferred dissolution medium in USP dissolution tests is deaerated water or if substantiated by the solubility characteristics of the drug or formulation, a buffered aqueous solution (typically pH 4–8) or dilute HCl may be used. The significance of deaeration of the medium should be determined. Various investigators have used 0.1 N HCl, 0.01 N HCl, phosphate buffer, simulated gastric juice, water, and simulated intestinal juice, depending on the nature of the drug product and the location in the GI tract where the drug is expected to dissolve. No single apparatus and test can be used for all drug products. Each drug product must be tested individually with the dissolution test that best correlates to in vivo bioavailability.

The dissolution test usually states that a certain percentage of the labeled amount of drug in the drug product must dissolve within a specified period of time. In practice, the absolute amount of drug in the drug product may vary from tablet to tablet. Therefore, a number of tablets from each lot are usually tested to get a representative dissolution rate for the product. The USP provides several official (compendia) methods for carrying out dissolution tests of tablets, capsules and other special products such as transdermal preparations. The selection of a particular method for a drug is usually specified in the monograph for a particular drug product.

BIOAVAILABILITY AND BIOEQUIVALENCE

Bioavailability and bioequivalence may be determined directly using plasma drug concentration vs. time profiles, urinary drug excretion studies, measurements of an acute pharmacologic effect, clinical studies, or in vitro studies. Bioavailability studies are performed for both approved active drug ingredients or therapeutic moieties not yet approved for marketing by the FDA. New formulations of active drug ingredients or therapeutic moieties must be approved, prior to marketing, by the FDA. In approving a drug product for marketing, the FDA must ensure that the drug product is safe and effective for its labeled indications for use. To ensure that the drug product meets all applicable standards of identity, strength, quality, and purity, the FDA requires bioavailability/pharmacokinetic studies and where necessary bioequivalence studies for all drug products.

For unmarketed drugs which do not have full New Drug Application (NDA) approval by the FDA, in vivo bioavailability studies must be performed on the

drug formulation proposed for marketing. Essential pharmacokinetic parameters of the active drug ingredient or therapeutic moiety is also characterized. Essential pharmacokinetic parameters include the rate and extent of systemic absorption, elimination half-life, and rates of excretion and metabolism should be established after single- and multiple-dose administration. Data from these in vivo bioavailability studies are important to establish recommended dosage regimens and to support drug labeling.

In vivo bioavailability studies are performed also for new formulations of active drug ingredients or therapeutic moieties that have full NDA approval and are approved for marketing. The purpose of these studies is to determine the bioavailability and characterize the pharmacokinetics of the new formulation, new dosage form, or new salt or ester relative to a reference formulation. After the bioavailability and essential pharmacokinetic parameters of the active ingredient or therapeutic moiety are established, dosage regimens may be recommended in support of drug labeling.

Bioequivalent Drug Products

Bioequivalent drug products are pharmaceutical equivalents whose bioavailability (i.e., rate and extent of systemic drug absorption) does not show a significant difference when administered at the same molar dose of the therapeutic moiety under similar experimental conditions, either single or multiple dose. Some pharmaceutical equivalents or may be equivalent in the extent of their absorption but not in their rate of absorption and yet may be considered bioequivalent because such differences in the rate of absorption are intentional and are reflected in the labeling, are not essential to the attainment of effective body drug concentrations on chronic use, or are considered medically insignificant for the particular drug product studied [21 CFR 320.1(e)].

Generic Drug Products

A generic drug product is considered bioequivalent to the reference listed drug product (generally the currently marketed, brand-name product with a full (NDA) approved by the FDA) if both products are pharmaceutical equivalents and its rate and extent of systemic drug absorption (bioavailability) do not show a statistically significant difference when administered in the same dose of the active ingredient, in the same chemical form, in a similar dosage form, by the same route of administration, and under the same experimental conditions.

Pharmaceutical equivalents are drug products that contain the same therapeutically active drug ingredient(s), same salt, ester, or chemical form; are of the same dosage form; and are identical in strength and concentration and route of administration. Pharmaceutical equivalents may differ in characteristics such as shape, scoring configuration, release mechanisms, packaging, and excipients (including colors, flavoring, preservatives).

Therapeutic equivalent drug products are pharmaceutical equivalents that can be expected to have the same clinical effect and safety profile when administered to patients under the same conditions specified in the labeling. Therapeutic equivalent drug products have the following criteria: 1) The products are safe and effective; 2) The products are pharmaceutical equivalents containing the same active drug ingredient in the same dosage form, given by the same route of administration, meet compendia or other applicable standards of strength, quality, purity, and identity and meet an acceptable in vitro standard; 3) The drug products are bioequivalent in that they do not present a known potential problem and are shown to meet an appropriate bioequivalence standard; 4) The drug products are adequately labeled; 5) The drug products are manufactured in compliance with current good manufacturing practice (GMP) regulations.

The generic drug product requires an abbreviated new drug application (ANDA) for approval by the FDA and may be marketed after patent expiration of the reference listed drug product. The generic drug product must be a therapeutic equivalent to the Reference drug product but may differ in certain characteristics including shape, scoring configuration, packaging, and excipients (includes colors, flavors, preservatives, expiration date, and minor aspects of labeling).

Pharmaceutical alternatives are drug products that contain same therapeutic moiety but are different salts, esters or complexes (e.g., tetracycline hydrochloride versus tetracycline phosphate) or are different dosage forms (e.g., tablet versus capsule; immediate release dosage form versus controlled release dosage form) or strengths.

In summary, clinical studies are useful in determining the safety and efficacy of the drug product. Bioavailability studies are used to define the affect of changes in the physico chemical properties of the drug substance and the affect of the drug product (dosage form) on the pharmacokinetics of the drug; whereas, bioequivalence studies are used to compare the bioavailability of the same drug (same salt or ester) from various drug products. If the drug products are bioequivalent and therapeutically equivalent, then the

clinical efficacy and safety profile of these drug products are assumed to be similar and may be substituted for each other.

DRUG PRODUCT PERFORMANCE IN VITRO AS A MEASURE OF IN VIVO DRUG BIOAVAILABILITY

The best measure of a drug product's performance is to give the drug product to human volunteers or patients and then determine the in vivo bioavailability of the drug using a pharmacokinetic or clinical study. For some well characterized drug products and for certain drug products where bioavailability is self-evident (e.g., sterile solutions for injection), in vivo bioavailability studies may be unnecessary. In these cases, the performance of the drug product in vitro is used as a surrogate to predict the in vivo drug bioavailability. Because these products have predictable in vivo performance as judged by the in vitro characterization of the drug and drug product, the FDA may waive the requirement for performing an in vivo bioavailability study (Table 8).

Drug Products for which Bioavailability is Self-Evident

Drug bioavailability from a true solution is generally considered self-evident. Thus, sterile solutions, lyophilized powders for reconstitution, ophthalmic solutions do not need bioequivalence studies but still must be manufactured according to current GMPs. However, highly viscous solutions may have bioavailability problems due to slow diffusion of the active drug.

In Vitro–In Vivo Correlation (IVIVC)

In vitro bioavailability data may be used to predict the performance of a dosage provided that the dissolution method selected is appropriate for the solid oral dosage form and prior information has been collected showing that the dissolution method will result in optimum drug absorption from the drug product. In general, IVIVC is best for well absorbed drugs for which the dissolution rate is the rate-limiting step. Some drugs are poorly absorbed and dissolution is not predictive of absorption (1). The objectives of IVIVC are to use rate of dissolution as a discriminating (i.e., sensitive to changes in formulation or manufacturing process), as an aid in setting dissolution specifications. When properly applied, IVIVC may be used to facilitate the evaluation

Table 8 Examples of drug products for which in vivo bioavailability studies may be waived

Condition	Example	Comment
Drug products for which bioavailability is self-evident	Drug solution (e.g., parenteral ophthalmic, oral solutions)	Drug bioavailability from a true solution is considered self-evident. However, highly viscous solutions may have bioavailability problems.
In vivo–in vitro correlation (IVIVC)	Modified release drug products	The dissolution of the drug from the drug product in vitro must be highly correlated to the in vivo bioavailability of the drug.
Biopharmaceutic classification (BCS) system	Immediate release solid oral drug products	Drug must be a highly soluble and highly permeable substance that is in a rapidly dissolving dosage form.
Biowaiver	Drug product containing a lower dose strength	Drug product is in the same dosage form, but lower strength and is proportionally similar in its active and inactive ingredients.

of drug products with manufacturing changes including minor changes in formulation, equipment, process, manufacturing site, and batch size. (see section on SUPAC) (2, 3, 10).

Three levels of IVIVC are generally recognized by the FDA (10). Level A correlation is usually estimated by deconvolution followed by comparison of the fraction of drug absorbed to the fraction of drug dissolved. A correlation of this type is the highest level of correlation and best predictor of bioavailability from the dosage form. A Level A correlation is generally linear and represents a point-to-point relationship between in vitro dissolution rate and the in vivo input rate. The Level A correlation should predict the entire in vivo time course from the in vitro dissolution data. Level B correlation utilizes the principles of statistical moment analysis. Various dissolution IVIVC methods were discussed by Shargel and Yu in 1985, 1993, 1999 (1). The mean in vitro dissolution time is compared to either the mean residence time or the mean in vivo dissolution time. Level B correlation, like Level A correlation, uses all of the in vitro and in vivo data but is not considered to be a point-to-point correlation and does not uniquely reflect the actual in vivo plasma level curve, since several different in vivo plasma level-time curves will produce similar residence times. A Level C correlation is the weakest IVIVC and establishes a single point relationship between a dissolution parameter (e.g., time for 50% of drug to dissolve, or percent drug dissolved in two hours, etc.) and a pharmacokinetic parameter (e.g., AUC, C_{max}, T_{max}). Level C correlation does not reflect the complete shape of the plasma drug concentration-time curve of dissolution profile.

BIOPHARMACEUTICS CLASSIFICATION SYSTEM (BCS)

The FDA may waive the requirement for performing an in vivo bioavailability or bioequivalence study for certain immediate release solid oral drug products that meets very specific criteria, namely, the permeability, solubility, and dissolution of the drug. These characteristics include the in vitro dissolution of the drug product in various media, drug permeability information, and assuming ideal behavior of the drug product, drug dissolution and absorption in the GI tract. For regulatory purpose, drugs are classified according to BCS in accordance the solubility, permeability and dissolution characteristics of the drug (FDA Draft Guidance for Industry, January, 1999, see FDA website for guidance) (11). Based on drug solubility and permeability, Amidon et al. (10, 12) recommended the following BCS in 1995 (Table 9).

This classification can be used as a basis for setting in vitro dissolution specifications and can also provide a basis for predicting the likelihood of achieving a successful in IVIVC. The solubility of a drug is determined by dissolving the highest unit dose of the drug in 250 ml of buffer adjusted between pH 1.0 and 8.0. A drug substance is considered highly soluble when the dose/solubility volume of solution are less than or equal to 250 ml. High-permeability drugs are generally those with an extent of absorption that is greater than 90%.

Solubility

An objective of the BCS approach is to determine the equilibrium solubility of a drug under approximate

Table 9 Biopharmaceutics classification system (BCS)

Condition	Comments
Solubility	A drug substance is considered highly soluble when the highest dose strength is soluble in 250 ml or less of water over a pH range of 1–8.
Dissolution	An immediate release (IR) drug product is considered rapidly dissolving when not less than 85% of the label amount of the drug substance dissolves within 30 min using the USP apparatus I at 100 rpm (or apparatus II at 50 rpm) in a volume of 900 ml or less. ^a
Permeability	A drug substance is considered highly permeable when the extent of absorption in humans is to be >90% of an administered dose based on mass balance determination.

^aMedia include: acidic media (e.g., 0.1 N HCl) or simulated gastric fluid, USP without enzymes, pH 4.5 buffer and pH 6.8 buffer of simulated intestinal fluid, USP without enzymes (From FDA Draft Guidance, Jan, 1999.)

physiological conditions. For this purpose, determination of pH-solubility profiles over a pH range of 1–8 is suggested. Preferably eight or more pH conditions should be evaluated. Buffers that react with the drug should not be used. An acid or base titration method can also be used for determining drug solubility. The solubility class is determined by calculating what volume of an aqueous media is sufficient to dissolve the highest anticipated dose strength. A drug substance is considered highly soluble when the highest dose strength is soluble in 250 ml or less of aqueous media over the pH range of 1–8. The volume estimate of 250 ml is derived from typical bioequivalence study protocols that prescribe administration of a drug product to fasting human volunteers with a glass (8 ounces) of water.

Solution stability of a test drug in selected buffers (or pH conditions) should be documented using a validated stability-indicating assay. Data collected on both pH-solubility and pH-stability should be submitted in the biowaiver application along with information on the ionization characteristics, such as pKa(s), of a drug.

Determining Permeability Class

Studies of the extent of absorption in humans, or intestinal permeability methods, can be used to determine the permeability class membership of a drug. To be classified as highly permeable, a test drug should have an extent of absorption >90% in humans. Supportive information on permeability characteristics of the drug substance should also be derived from its physical–chemical properties (e.g., octanol:water partition coefficient).

Some methods to determine the permeability of a drug from the GI tract include 1) in vivo intestinal perfusion studies in humans, 2) in vivo or in situ intestinal perfusion studies in animals, 3) in vitro permeation experiments using excised human or animal intestinal tissues, and 4) in vitro permeation experiments across a monolayer of cultured human intestinal cells. When using these methods, the experimental permeability data should correlate with the known extent-of-absorption data in humans.

Table 10 Postapproval change levels

Change level	Example	Comment
Level 1	Deletion or partial deletion of an ingredient to affect the color or flavor of the drug product	Level 1 changes are those that are unlikely to have any detectable impact on formulation quality and performance.
Level 2	Quantitative change in excipients greater than allowed in a Level 1 change.	Level 2 changes are those that could have a significant impact on formulation quality and performance
Level 3	Qualitative change in excipients	Level 3 changes are those that are likely to have a significant impact on formulation quality and performance. A Level 3 change may require in vivo bioequivalence testing.

Dissolution

The dissolution class is based on the in vitro dissolution rate of an immediate release drug product under specified test conditions and is intended to indicate rapid in vivo dissolution in relation to the average rate of gastric emptying in humans under fasting conditions. An immediate release drug product is considered rapidly dissolving when not less than 85% of the label amount of drug substance dissolves within 30 min using the USP apparatus I at 100 rpm or apparatus II at 50 rpm in a volume of 900 ml or less in each of the following media 1) acidic media such as 0.1 N HCl or Simulated Gastric Fluid USP without enzymes; 2) a pH 4.5 buffer; and 3) a pH 6.8 buffer or Simulated Intestinal Fluid USP without enzymes.

BIOWAIVERS

In addition to routine quality control tests, comparative dissolution tests have been used to waive bioequivalence requirements (biowaivers) for lower strengths of a dosage form. The drug products containing the lower dose strengths should be compositionally proportional or qualitatively the same as the higher dose strengths and have the same release mechanism. For biowaivers, a dissolution profile should be generated and evaluated using one of the methods described under Section V in this guidance, "Dissolution Profile Comparisons." Biowaivers are generally provided for multiple strengths after approval of a bioequivalence study performed on one strength, using the following criteria: For multiple strengths of IR products with linear kinetics, the bioequivalence study may be performed at the highest strength and waivers of in vivo studies may be granted on lower strengths, based on an adequate dissolution test, provided the lower strengths are proportionately similar in composition [21 CFR 320.22(d)(2)]. Similar may also be interpreted to mean that the different strengths of the products are within the scope of changes permitted under the category "Components and Composition," discussed in the SUPAC-IR guidance.

SCALE-UP AND POSTAPPROVAL CHANGES (SUPAC)

After a drug product is approved for marketing by the FDA, the manufacturer may want to make a manufacturing change. The pharmaceutical industry, academia and

the FDA developed (2, 3, 5, 3, 10, 3, 12–17) a series of guidances for the industry that discuss scale-up and postapproval changes, generally termed, SUPAC guidances (11). The FDA SUPAC guidances are for manufacturers of approved drug products who want to change 1) a component and composition of the drug product; 2) the batch size; 3) the manufacturing site; 4) the manufacturing process or equipment; and/or 5) packaging. These guidances describe various levels of postapproval changes according to whether the change is likely to impact on the quality and performance of the drug product. The level of change as classified by the FDA as to the likelihood that a change in the drug product might affect the quality of the product (Table 10).

REFERENCES

1. Shargel, L.; Yu, A.B.C. *Applied Biopharmaceutics and Pharmacokinetics*; McGraw-Hill Medical Publishing: 1999.
2. SUPAC-MR: Modified Release Solid Oral Dosage Forms; Scale-Up and Post-Approval Changes: Chemistry, Manufacturing and Controls, In Vitro Dissolution Testing, and In Vivo Bioequivalence Documentation, FDA, Guidance for Industry, Sept. 1997.
3. Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate Release Solid Oral Dosage Forms Containing Certain Active Moieties/Active Ingredients Based on a Biopharmaceutics Classification System, FDA Draft Guidance for Industry, Jan. 1999.
4. Skelly, J.P.; Shah, V.P.; Konecny, J.J.; Everett, R.L.; McCullouen, B.; Noorizadeh, A.C. Report of the Workshop on CR Dosage Forms: Issues and Controversies. *Pharmaceutical Research* **1987**, 4 (1), 75–78.
5. Shah, V.P.; Konecny, J.J.; Everett, R.L.; McCullouen, B.; Noorizadeh, A.C.; Shah, V.P. In Vitro Dissolution Profile of Water Insoluble Drug Dosage Forms in the Presence of Surfactants. *Pharmaceutical Research* **1989**, 6, 612–618.
6. Moore, J.W.; Flanner, H.H. Mathematical Comparison of Dissolution Profiles. *Pharmaceutical Technology* **1996**, 20 (6), 64–74.
7. Hendeles, L.; Weinberger, M.; Milavetz, G.; Hill, M.; Vaughan, L. Food Induced Dumping from "Once-A-Day" Theophylline Product as Cause of Theophylline Toxicity. *Chest* **1985**, 87, 758–785.
8. Allen, P.V.; Rahn, P.D.; Sarapu, A.C.; Vandewielen, A.J. Physical Characteristics of Erythromycin Anhydrate and Dihydrate Crystalline Solids. *J. Pharm. Sci.* **1978**, 67, 1087–1093.
9. Aguiar, A.J.; Krc, J.; Kinkel, A.W.; Samyn, J.C. Effect of Polymorphism on the Absorption of Chloramphenicol from Chloramphenicol Palmitate. *J. Pharm. Sci.* **1967**, 56, 847–853.
10. SUPAC-IR Immediate Release Solid Oral Dosage Forms. Scale-up and Post-Approval Changes: Chemistry, Manufacturing and Controls, In Vitro Dissolution Testing, and

- In Vivo Bioequivalence Documentation, FDA Guidance for Industry, Nov. 1995.
11. FDA Regulatory Guidances FDA Website for Regulatory Guidances. www.fda.gov/cder/guidance/index.htm).
 12. Amidon, G.L.; Lennernas, H.; Shah, V.P.; Crison, J.R. A Theoretical Basis For a Biopharmaceutic Drug Classification: The Correlation of In Vitro Drug Product Dissolution and In Vivo Bioavailability. *Pharmaceutical Research* **1995**, *12*, 413–420.
 13. Skelly, J.P.; Amidon, G.L.; Barr, W.H.; Benet, L.Z.; Carter, J.E.; Robinson, J.R.; Shah, V.P.; Yacobi, A. In Vitro and In Vivo Testing and Correlation for Oral Controlled/Modified-Release Dosage Forms. *Pharmaceutical Research* **1990**, *7*, 975–982.
 14. In Vitro-In Vivo Correlation for Extended Release Oral Dosage Forms, Pharmacopeial Forum Stimuli Article, United States Pharmacopeial Convention, Inc.: July 1988; 4160–4161.
 15. In Vitro In Vivo Evaluation of Dosage Forms, U.S.P. XXIV<1088> United States Pharmacopeial Convention, Inc. 2051–2056.
 16. Shah, V.P.; Skelly, J.P.; Barr, W.H.; Malinowski, H.; Amidon, G.H. Scale-up of Controlled Release Products — Preliminary Considerations. *Pharmaceutical Technology* **1992**, *16* (5), 35–40.
 17. Skelly, J.P. Report of Workshop on In Vitro and In Vivo Testing and Correlation for Oral Controlled/Modified-Release Dosage Forms. *Journal of Pharmaceutical Sciences* **1990**, *79* (9), 849–854.
 18. Cadwallader, D.E. *Biopharmaceutics and Drug Interactions*; Raven Press: New York, 1983.
 19. Gibaldi, M. *Biopharmaceutics and Clinical Pharmacokinetics*; Lea & Febiger: Philadelphia, 1984.
 20. Gibaldi, M.; Perrier, D. *Pharmacokinetics*; Marcel Dekker, Inc.: New York, 1982.
 21. McGinity, J.W.; Stavchansky, S.A.; Martin, A. Bioavailability in Tablet Technology. *Pharmaceutical Dosage Forms: Tablets*; Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1981; 2.
 22. Rowland, M.; Tozer, T.N. *Clinical Pharmacokinetics. Concepts and Applications*; Lea & Febiger: Philadelphia, 1995.

BIODEGRADABLE POLYMERS AS DRUG CARRIERS

Peter Markland

Southern Research Institute, Birmingham, Alabama, U.S.A.

Victor C. Yang

The University of Michigan, Ann Arbor, Michigan, U.S.A.

OVERVIEW

Polymers first developed in the search for biodegradable suture materials have proven to be useful and successful for long-term drug delivery applications. Biodegradable polymers are highly desirable in these situations because they degrade in the body to biologically inert and compatible molecules. By incorporating drug into biodegradable polymers, dosage forms that release the drug over a prolonged length of time can be prepared in a variety of shapes and sizes. No secondary surgical procedures are needed after completion of the dosing regimen since the remaining polymer dosage form will be degraded and cleared by the body. As a result, biodegradable polymers offer a novel approach for developing sustained release drug delivery systems that are simple and convenient to the patient.

Many different biodegradable polymer chemistries have been proposed for these applications; however, the most common and successful are the polyesters that were first investigated as degradable sutures. These polymers include poly(glycolide), poly(D,L-lactide), and their related copolymers poly(D,L-lactide-co-glycolide). Many commercial products based on these materials are currently on the market, including Decapeptyl[®], Lupron Depot[®], and Sandostatin LAR[®]. In traditional parenteral depot applications, biodegradable polymers have been used simply as inert carrier vehicles. Preparation of the dosage form was carried out, by various means, by simply incorporating the drug into the polymer matrix. However, the effective delivery of new drug therapies, including peptides, proteins, and genetic- and cell-based drugs, places greater demands on the performance of the polymer platform. Consequently, new polymer chemistries and novel dosage form design are being investigated in attempts to produce tailor-made dosage forms that are capable of enhancing the delivery and efficacy of these drugs.

INTRODUCTION

A wide variety of delivery systems have been developed for the purpose of prolonging the release and, ultimately, bioavailability of drugs to the body. Examples include the transdermal patch, oral dosage forms such as osmotic pumps and swellable hydrophilic polymer matrices, and various types of polymer-based parenteral depot formulations (1, 2). Depot sustained release formulations can overcome limitations associated with oral or transdermal administration routes. Several of these limitations include poor drug stability in the GI tract, low drug permeability through the GI mucosa or stratum cornea, rapid clearance from first-pass metabolism, and the need for delivery for more than a few days. Many interesting and successful parenteral depot systems for sustained release applications have been developed. Such systems can be distinguished into degradable and nondegradable delivery systems based on the properties of the polymers.

One example of a nondegradable delivery system is Norplant[®] (Wyeth-Ayerst) which has been shown successful in prevention of pregnancy for up to five years (3). This delivery system consists of six small, closed tubes made of a silicone rubber copolymer of dimethylsiloxane and methylvinylsiloxane (Silastic[®]) and is implanted subcutaneously by scalpel incision or via trocar. Each tube contains 36 mg of the progestin hormone levonorgestrel. At termination of the dosing regimen, a nondegradable delivery system like this requires a secondary surgical procedure to retrieve the implanted device from the body. Such a retrieval procedure can be undesirable for several reasons including the added cost, the possibility of complications during retrieval, the risk of infection, and the lack of patient compliance. Although Norplant[®] is widely considered a safe and highly cost-effective method of contraception (4), there nevertheless has been a precipitous drop in demand for this product in

the United States since 1994. This decrease in interest has been attributed, in large part, to the difficulties associated with removal of the implanted rods along with the publicity created by these clinical problems (5).

Consequently, polymers that can degrade into biologically compatible components under physiologic conditions present a far more attractive alternative for the preparation of drug delivery systems. The use of biodegradable polymers precludes the need for retrieval at the conclusion of the dosing regimen, thereby avoiding the potential complications associated with the use of nondegradable systems. Degradation may take place by a variety of mechanisms, although it generally relies on either erosion or chemical changes to the polymer. Degradation by erosion normally takes place in devices that are prepared from soluble polymers. In such instances, the device erodes as water is absorbed into the system causing the polymer chains to hydrate, swell, disentangle, and, ultimately, dissolve away from the dosage form. Alternatively, degradation could result from chemical changes to the polymer including, for example, cleavage of covalent bonds or ionization/protonation either along the polymer backbone or on pendant side-chains. A number of degradation schemes have been described that characterize how chemical degradation of the polymer or of polymer-drug conjugates can be utilized to achieve drug release (6, 7).

The most widely studied biodegradable polymers include those which undergo chemical degradation through random hydrolysis of the covalent bonds constituting the backbone of the polymer chains. Random chain scission results in a reduction in the molecular weight of the polymer. As this process continues over time, polymer chains become progressively shorter resulting, at some point in time, in the loss of mechanical integrity in the dosage form. Ultimately, the degradation process proceeds until polymer fragments are degraded to soluble oligomers or individual monomers. As a necessity due to this process, biodegradable polymers and their degradation products must be biologically compatible and nontoxic. Consequently, the monomers typically used in the preparation of biodegradable polymers are often molecules that are endogenous to biological systems. The most common class of biodegradable polymers is the hydrolytically labile polyesters prepared from lactic acid, glycolic acid, or combinations of these two molecules. Polymers prepared from these individual monomers include poly(D,L-lactide) (PLA), poly(glycolide) (PGA), and the copolymer poly(D,L-lactide-co-glycolide) (PLG).

The distinctions between chemical and erosion degradation mechanisms, however, are by no means

absolute, and there are instances where both processes contribute to the overall degradation and resorption of a drug delivery system. For instance, copolymers of glutamic acid and ethyl glutamic acid have been studied for the release of the narcotic antagonist naltrexone (8). In this case, hydrolysis of the ethyl ester side-chain converts this copolymer to the soluble polymer poly(glutamic acid). Following hydrolysis, the polymer device erodes as the soluble polymer chains dissolve away from the device. In addition, there are many other examples where both chemical degradation and physical erosion are involved in the final disintegration and resorption of the dosage form. In this regard, the terms such as "biodegradable" and "bioerodible" are sometimes used interchangeably in the literature while, at other times, they can be used to refer to distinct degradation processes. Consequently, care should be taken when reviewing the literature to interpret how these terms are being used. In some cases, the term "biodegradation" is limited to the description of chemical processes (chemical changes that alter either the molecular weight or solubility of the polymer) while "bioerosion" may be restricted to refer to physical processes that result in weight loss of a polymer device.

HISTORY

The use of biodegradable polymers in drug delivery applications grew from the search for polymers that could be employed as degradable sutures. Synthetic polymers such as poly(glycolic acid) were first developed in the 1950s (9); however, their poor hydrolytic stability made them unsuitable for permanent applications. This attribute, however, made these materials useful for applications such as sutures that could benefit from their ability to degrade in the presence of moisture (10, 11). Examples include Dexon[®] and Vicryl[®] sutures prepared from poly(glycolic acid) and poly(lactide-co-glycolide), respectively (10, 12). The utility of these materials as degradable sutures further led to their application in the development of sustained release drug delivery formulations. In 1970, Yolles et al. reported the use of the poly(lactic acid) biodegradable system for delivery of the narcotic antagonist cyclazocine (13). At about the same time, a number of other drugs including anticancer agents (14), steroids (15), and other narcotic antagonists (16–18) were reported to be delivered from biodegradable formulations. Many reviews catalog the early development of this technology for the sustained delivery of a variety of small molecule drugs (19–21).

More recently, the growth of biotechnology has led to the identification of many potent and powerful protein- and

gene-based macromolecular drugs. However, delivery of these drugs to the body in an efficacious manner without sacrificing the quality of life of the patient continues to be a major obstacle. Gastric proteases and low permeability across the gastrointestinal epithelium mean that macromolecular drugs have poor bioavailability via the oral route. Instead, these drugs are normally dosed by injection (e.g., subcutaneous, intramuscular, or intravenous). However, the short biological half-lives of these molecules requires that frequent parenteral injections be performed in order to maintain treatment efficacy. Despite the obvious need to develop long-term delivery systems for these highly valuable drugs, the application of this technology has been fraught with obstacles (22–24). Polymer–drug interactions, processing conditions, and the internal pH, temperature, and moisture levels within the implanted device have contributed to difficulties retaining drug stability before being released from the implanted device.

In spite of these complications, considerable achievements have been made in the development of biodegradable delivery systems for proteins and, in particular, bioactive peptides. Most of the commercial success over the past decade has been achieved using the polyester PLA and the various copolymers of PLG. In particular, a number of biodegradable delivery systems have been developed for synthetic analogs of luteinizing hormone-releasing hormone (LHRH). Sanders et al., for instance, demonstrated the release of nafarelin acetate for over 30 days from microsphere formulations of PLG with a

50:50 molar ratio of the lactide and glycolide monomers (50:50 PLG) (25). The first such system to the market was Decapeptyl® (Ipsen Biotech), a microsphere depot formulation for treatment of prostate cancer (Fig. 1) (26). This product delivers 3.75 mg of an LHRH analog over a 30-day period. More recently, TAP Pharmaceuticals (Deerfield, IL) commercialized a series of depot microsphere formulations of the LHRH analog leuprolide acetate under the trade name Lupron Depot®. Products prepared from either PLG or PLA are presently available that can provide peptide release over 1, 3, and 4 months (3). Lupron Depot® formulations are indicated for the treatment of endometriosis (27), prostate cancer (28), and precocious puberty in children (29).

In contrast to these microsphere formulations, Zoladex® (AstraZeneca) is a small implantable cylinder (approximately 1 mm in diameter and 10 mm in length) containing goserelin acetate in a polymer matrix of 50:50 PLG. This system delivers approximately 3.6 mg drug over a 1-month period and is indicated for the treatment of prostate cancer. A second Zoladex® system containing 10.6 mg goserelin has also been developed to release the drug for over 3 months (3).

The synthetic somatotropin analog octreotide acetate has been successfully formulated into a microsphere formulation composed of PLG. This product, Sandostatin LAR® (Novartis), has been used for acromegaly treatment (30) as well as the treatment of diarrhea and flushing episodes associated with metastatic carcinoid (31).

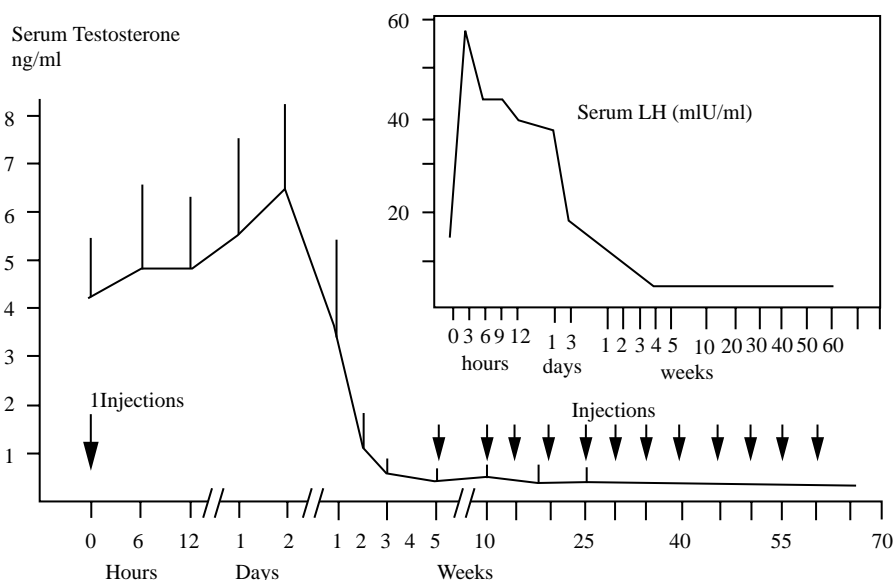


Fig. 1 Serum LH and testosterone concentrations of 22 human subjects treated at 5 week intervals with the biodegradable depot formulation Decapeptyl (3 mg LHRH per dose). (From Ref. 26.)

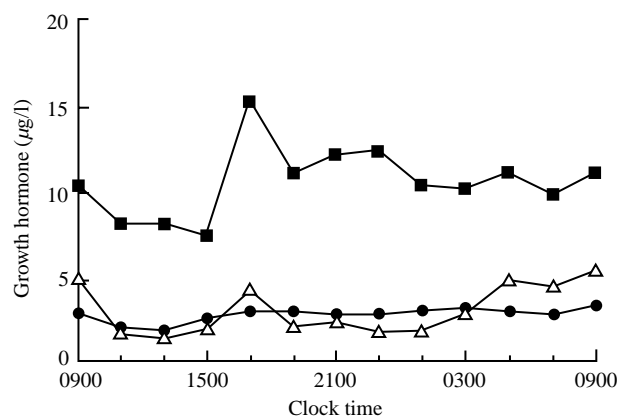


Fig. 2 The influence of Sandostatin LAR biodegradable depot formulation on the mean plasma growth hormone concentrations in humans. Plasma concentrations are shown over a 24 hour period 28 days after administration of a second monthly 20 mg dose of Sandostatin LAR, ●. For comparison, plasma concentrations are provided for untreated controls, ■, and for patients receiving multiple daily subcutaneous octreotide injections, △. (From Ref. 30.)

The depot formulation, given once monthly, is reported to be as well tolerated and effective as octreotide solutions that require long-term subcutaneous dosing 2–3 times daily (Fig. 2). This example clearly illustrates the potential for biodegradable sustained release formulations to significantly improve quality of life.

In December 1999, the first protein biodegradable depot formulation received regulatory approval. Nutropin-Depot® (Genentech) contains somatotropin, a recombi-

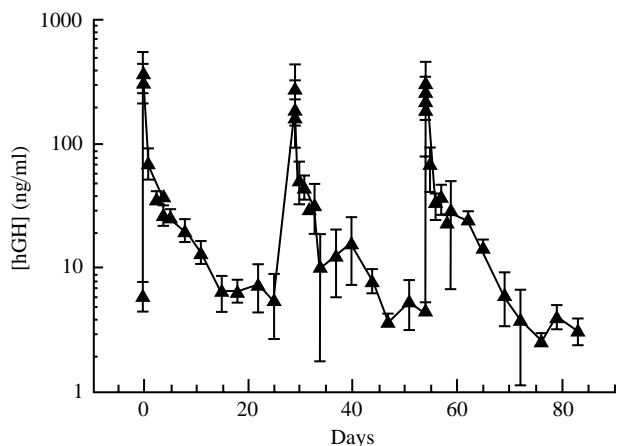


Fig. 3 Serum human growth hormone (hGH) levels in immunosuppressed rats receiving three monthly doses of a biodegradable depot formulation containing 7.5 mg rhGH. (From Ref. 32.)

nant human growth hormone (rhGH) having a molecular weight of 22,125 D, within PLG microspheres. Protein activity was preserved by preparing the microspheres from an insoluble zinc–hormone complex (Fig. 3) (32).

While many different polymer chemistries have been developed for drug delivery applications, only one class of polymer beside the polyesters has received regulatory approval. Gliadel® is a thin wafer containing the chemotherapeutic agent carmustine (BCNU) in a poly-anhydride polymer matrix. Gliadel® received conditional approval in 1996 as an adjunct therapy following surgical removal of recurrent glioblastoma multiforme (GBM) tumors (3).

Obviously, the regulatory status of polyesters means that these materials dominate the field as far as commercial development is concerned. However, there is a myriad assortment of polymers that continue to be developed for drug delivery applications. These efforts are aimed at developing new polymeric materials designed specifically to overcome obstacles in drug delivery, bioavailability, or stability. Additionally, new engineering or processing techniques are also being developed to aid the design of novel dosage forms that can enhance drug delivery characteristics and the use of these polymeric materials in a variety of new biomedical applications.

FACTORS AFFECTING POLYMER SELECTION

Both synthetic polymers and those derived from naturally occurring sources have been evaluated in biodegradable drug delivery applications. Examples of several classes of biodegradable polymers are presented in Table 1. There are a variety of polymer properties or attributes to be considered when selecting a biodegradable polymer. Many of these are listed in Table 2. One of the most critical considerations is the regulatory requirement for a particular application. If an application requires rapid development and commercialization, then the polymer selection will most likely be made from among those polyesters that have already received regulatory approval. Another factor to consider is whether to use homopolymers consisting of a single monomeric repeat unit or copolymers containing multiple monomer species. If copolymers are to be employed, then the relative ratio of the different monomers may be manipulated to change polymer properties. It should be noted that polymer composition would directly dictate many of the polymer physicochemical properties listed in Table 2 including bulk hydrophilicity, morphology, structure, and the extent of drug–polymer interactions (e.g., drug solubility in the

Table 1 Common classes and examples of biodegradable polymers

Type	Class	Examples
Synthetic polymers	Polyamide	Polyamino acids, Polypeptides
	Polyester	Poly(glycolide)
		Poly(D,L-lactide)
		Poly(D,L-lactide-co-glycolide)
		Poly(ϵ -caprolactone)
		Poly(dioxanone)
		Poly(hydroxybutyrate)
	Polyanhydride	
	Polyorthoester	
	Polyphosphazene	
Naturally occurring polymers	Polyphosphoester	Polyphosphate
		Polyphosphonate
		Polyphosphite
	Polysaccharides	Dextran
		Chitosan
		Alginate
		Starch
	Polypeptides, Proteins	Hyaluronic acid
		Collagen
		Gelatin
		Bovine serum albumin (BSA)
		Human serum albumin (HSA)

polymer). Ultimately, these properties will all influence the performance of the drug delivery system via changes to the relative rates of mass transport (e.g., water in and solute or drug out of the system) and the degradation rate of both the polymer and the device.

In addition to polymer composition, the thermal attributes of the polymer, as described by the glass transition temperature (T_g) and melting temperature (T_m), can also affect the mass transport rates through the polymer as well as the polymer processing characteristics and the stability of the dosage form. Below the glass transition temperature, the polymer will exist in an amorphous, glassy state. When exposed to temperatures above T_g , the polymer will experience an increase in free volume that permits greater local segmental chain mobility along the polymer backbone. Consequently, mass transport through the polymer is faster at temperatures above T_g . Often, polymer processing, such as extrusion or high-shear mixing, is performed above T_g . On the other hand, the greatest stability during storage of a polymer device may be obtained at temperatures below T_g , where solute diffusion is much slower and more subtle changes in polymer properties (e.g., tackiness) are reduced.

The presence of plasticizers, such as residual solvent or dissolved solutes including the drug or other additives, will tend to lower polymer T_g . Conversely, features that hinder segmental motion along the polymer, such as greater chain rigidity, bulky side groups, and ring structures, would tend to increase T_g . Finally, the presence of charged groups on a polymer can also influence drug release from the device. The number and density of ionized groups along the polymer backbone, on the side-chain groups, or at the terminal end-groups of the polymer chains can all vary the extent of polymer–polymer and polymer–drug interactions. In this manner, ionizable groups can affect drug solubility in the polymer and, correspondingly, the release rate from the polymer. A number of reviews which describe in detail of the relationship between polymer properties and performance in drug delivery applications have been published (33, 34).

FACTORS AFFECTING DRUG RELEASE

Aside from the physicochemical properties of the polymer itself, there are additional factors that can influence drug

Table 2 Properties affecting polymer selection, manufacture, and performance

Property	Examples
Regulatory and toxicology status	
Monomer or copolymer composition	
Molecular weight	M_w , M_n
Molecular weight distribution	Polydispersity ratio (M_w/M_n)
Molecular architecture	Linear polymers Branched polymers Crosslinked network
Tacticity	Isotactic Syndiotactic Atactic
Secondary structural attributes	Helicity beta-structure
Morphology	Amorphous Semicrystalline Crystalline
Thermal transition temperatures	Melting temperature, T_m Glass transition temperature, T_g
Ionization	Side-chains Main-chain end groups

release and the performance characteristics of the dosage form. Examples of these factors are highlighted in Table 3. Dosage forms can be formulated into a wide range of geometries and physical forms and sizes. The appropriate selection would depend on the flexibility with which a polymer can be processed, the desired route of administration, the duration of action, and the stability

of the drug under processing conditions. Additionally, the drug distribution in the dosage form can be either homogeneous (a monolithic or matrix system) or heterogeneous (reservoir system). Finally, the rate and mechanism by which the polymer degrades can also influence drug release and, potentially, drug stability following administration. If necessary, systems can be designed so that release of the drug load is completed before polymer degradation begins to affect the integrity of the dosage form. Conversely, a dosage form can also be designed in a way that polymer degradation and device erosion can both contribute to drug release.

Based on the above reasons, polymers possessing a variety of degradation rates and mechanisms have been developed; however, hydrolysis still remains the predominant degradation mechanism for polymers that are most commonly used in drug delivery applications. Many polymers that are susceptible to hydrolysis, for example, the polyesters PLA and PLG, degrade by random hydrolysis that takes place homogeneously throughout the bulk of the polymer device. In contrast, other classes of polymers, such as the polyanhydrides and polyorthoesters, have been developed in an attempt to yield hydrolysis only at the outer surface of the device that is exposed directly to the aqueous environment. When this so-called surface-erosion process is achieved, hydrolysis is believed to take place in a heterogeneous manner, and the polymer device degrades from the outside toward the center.

Reservoir systems are designed to have the drug deposited inside of a polymer membrane. In such systems where the polymer membrane serves as the barrier, drug release is controlled by Fickian diffusion of the drug

Table 3 Device attributes affecting performance

Property	Attribute	Examples
Design of delivery device	Shape/geometry	Cylinder or rod
		Microparticles (microsphere, microcapsule, nanoparticle)
	Drug distribution	Film or sheet
		Viscous gel or liquid
Degradation	Formulation	Homogeneous (matrix or monolithic system)
		Heterogeneous (reservoir system)
		Drug loading
	Rate	Excipients
		Porosity
	Mechanism	Homogeneous degradation (bulk)
		Heterogeneous degradation (surface-erosion)

through the membrane, and the release rate can be described by the following relationship:

$$\frac{dM_t}{dt} = \frac{ADC_s}{t_m} \quad (1)$$

where M_t is the mass of drug released at time t , A is the surface area of the barrier membrane, D is the diffusion coefficient of the drug in the membrane, C_s is the solubility of the solute in the polymer, and t_m is the membrane thickness (35). As long as the drug concentration in the reservoir remains well above saturation and the membrane thickness is small relative to the other dimensions of the device, this relationship indicates that the drug release rate should be constant over time (zero-order) so long as release was diffusion controlled. Because biodegradable polymers are chemically unstable, they are generally not used to prepare reservoir delivery systems. The potential for these polymers to degrade prematurely thereby releasing the remaining contents of the drug reservoir presents a safety concern.

More typically, therefore, biodegradable polymers are used to prepare matrix (monolithic) systems in which the drug is dispersed or dissolved homogeneously throughout the polymer. Matrix systems can be prepared in a multitude of forms including films, sheets, cylinders, rods, and microparticles (e.g., microspheres and nanospheres). Drug is initially released from the exposed, outer surfaces of this type of dosage form. As these regions become depleted of drug, release continues as drug is transported from increasingly deeper regions of the dosage form. Because the diffusional path length gets continuously longer during the release process, diffusion-controlled release from matrix systems is not zero-order. Based on a thin-slab geometry, Higuchi derived a relationship which, when differentiated, yields the following prediction for the diffusion-controlled release rate from a matrix system (35):

$$\frac{dM_t}{dt} = \frac{A}{2} \left(\frac{DS_s(2C_L - C_s)}{t} \right)^{\frac{1}{2}} \quad (2)$$

where A is the area of the polymer slab, C_L is the drug loading in the device and the remaining variables are the same as defined in Eq. (1). This relationship suggests that when diffusion-controlled release is achieved, the drug release rate will decrease proportionately with $t^{-1/2}$, assuming that polymer degradation does not contribute to drug release.

Results obtained by quantitatively monitoring drug release from monolithic matrix systems indicate that low-molecular-weight drugs can be released at rates that are consistent with a diffusion-controlled mechanism. High-

molecular-weight drugs such as peptides and proteins, however, generally have little permeability through the polymer phase due to their low polymer solubility and diffusion rates. Consequently, these molecules are unlikely to be released from monolithic systems by purely diffusion-controlled mechanisms. Instead, release of peptides and proteins from polymer matrices needs to be aided by additional mechanisms that can facilitate mass transport out of the device (36). For example, release can be enhanced by the formation of pores or channels that are created by the continuous dissolution and removal of soluble components, such as the drug or other formulation excipients, from the polymer matrix. Alternatively, drug release can also be enhanced by polymer degradation and the subsequent erosion of the device itself. A variety of mathematical models have been proposed to correlate the effects of both polymer degradation and drug diffusion on the overall release rate (20, 37, 38).

In a special case, novel systems have been developed that are capable of providing zero-order drug release profiles. Surface-eroding polymers in which polymer degradation takes place only at the outer surfaces of the device have been prepared. When drug release can be limited to regions undergoing degradation, a constant drug release profile is possible. In this case, the polymer degradation rate can be used to control the release rate of the drug. Hopfenberg developed a mathematical model to correlate this type of drug release system so long as the surface area remains constant during the degradation process (39). The cumulative fraction of drug released at time t was described by:

$$\frac{M_t}{M_\infty} = 1 - \left(1 - \frac{k_0 t}{C_L a} \right)^n \quad (3)$$

where k_0 is the zero-order rate constant describing the polymer degradation (surface-erosion) process, C_L is the initial drug loading throughout the system, a is the system half-thickness (i.e., the radius for a sphere or cylinder), and n is an exponent that varies with geometry [$n = 1, 2$, and 3 for slab (flat), cylindrical, and spherical geometry, respectively] (39).

COMMON CLASSES OF BIODEGRADABLE POLYMERS

Polyesters

A variety of hydrolytically labile polyesters have been evaluated in drug delivery applications. Several

Table 4 Common polyester structures

Polyester linkage	$\left(\text{O}-\text{R}-\overset{\text{O}}{\parallel}{\text{C}} \right)_n$
poly(glycolide), PGA	$\left(\text{O}-\text{CH}_2-\overset{\text{O}}{\parallel}{\text{C}} \right)_n$
poly(D,L-lactide), PLA	$\left(\text{O}-\overset{\text{CH}_3}{\underset{\text{~}}{\text{CH}}}-\overset{\text{O}}{\parallel}{\text{C}} \right)_n$
poly(D,L-lactide-co-glycolide), PLG	$\left[\left(\text{O}-\overset{\text{CH}_3}{\underset{\text{~}}{\text{CH}}}-\overset{\text{O}}{\parallel}{\text{C}} \right)_l \left(\text{O}-\text{CH}_2-\overset{\text{O}}{\parallel}{\text{C}} \right)_m \right]_n$
poly(ϵ -caprolactone)	$\left(\text{O}-\left(\text{CH}_2 \right)_5-\overset{\text{O}}{\parallel}{\text{C}} \right)_n$
poly(dioxanone)	$\left(\text{O}-\left(\text{CH}_2 \right)_2-\text{O}-\text{CH}_2-\overset{\text{O}}{\parallel}{\text{C}} \right)_n$
poly(hydroxybutyrate)	$\left(\text{O}-\overset{\text{CH}_3}{\underset{\text{~}}{\text{CH}}}-\text{CH}_2-\text{O}-\text{CH}_2-\overset{\text{O}}{\parallel}{\text{C}} \right)_n$

examples are listed in Table 4. Among these, however, poly(glycolide), poly(lactide), and various copolymers of poly(lactide-co-glycolide) are the ubiquitous choice because of their proven safety and lack of toxicity, their wide range of physicochemical properties, and their flexibility to be processed into a variety of physical dosage forms. These polymers and copolymers are prepared by anionic ring-opening reaction of highly purified glycolide and lactide monomers, the cyclic dimers of glycolic acid and lactic acid, respectively. Stannous octoate (tin(II) 2-ethyl hexanoate) is the chain initiator most commonly used in the synthesis of polymers for pharmaceutical applications. Because the lactide monomer possesses two chiral carbons, polymerization may be performed using D-lactide (the D-,D-cyclic dimer), L-lactide (the L-,L- cyclic dimer), or the meso-lactide (the D-,L- cyclic dimer). Synthesis of poly(D,L-lactide) (PLA), though, is most commonly performed by copolymerization of a racemic mixture of the D- and L-lactide monomers. In a similar manner, the copolymer poly(D,L-lactide-co-glycolide) (PLG) is generally prepared using varying ratios of the glycolide to a racemic blend of D-/L-lactides. Because PLA and PLG are prepared from distinct monomer species, there exists the possibility that polymerization may result in a nonrandom sequence of monomer species (21). Compositional heterogeneity can lead to variability in properties between polymer lots. Historically, this has been problematic for PLG polymers containing 50% glycolide (50:50 PLG).

Homopolymers of poly(D-lactide) and poly(L-lactide) tend to be semi-crystalline. As a result, water transport into these polymers is slow. Because of the slow uptake of water and the structural integrity introduced by crystallites, degradation rates of these polymers tend to be relatively slow (i.e., 18–24 months) (21). In contrast, poly(D,L-lactide) (PLA) is amorphous and is observed to degrade somewhat faster (i.e., 12–16 months). Adding increasing proportions of glycolide into PLA lowers T_g and generally increases polymer hydrophilicity. These PLG copolymers generally remain amorphous as long as the glycolide content remains within the range of about 0–70 mole%. In contrast, poly(L-lactide-co-glycolide) is amorphous when the glycolide content is 25–70 mole%. The most rapid degradation rate (i.e., 2 months) is observed in PLG copolymers containing 50% glycolide. Poly(glycolide), despite being semi-crystalline, is found to degrade relatively fast (i.e., 2–4 months) even compared to the amorphous PLA. This is attributed to the much greater hydrophilicity of the glycolide over the lactide. Actual degradation times, though, will depend on environmental conditions, polymer molecular weight,

system geometry and morphology, and processing conditions (21, 40, 41).

After exposure to and equilibration in an aqueous environment, polyesters degrade by hydrolysis that occurs homogeneously throughout the bulk of the polymer device. Evidence commonly used to support such a conclusion is that the bulk degradation of the polymer device (as indicated by weight loss) lags behind the decrease in polymer molecular weight over time (Fig. 4). However, other studies have provided evidence of nonhomogeneous hydrolytic degradation under certain circumstances. Devices prepared from amorphous polyesters including PLA and 75:25 PLG have been reported to exhibit accelerated degradation within the interior of the device (42). This phenomenon has been attributed to acid-catalyzed polymer hydrolysis resulting from the build-up in these regions of acidic oligomers and monomers during the degradation process. In these instances, the devices are either large in physical dimensions or they are nonporous, meaning that the acidic by-products cannot be readily washed away from the interior of the device. Alternatively, heterogeneous degradation in semi-crystalline polymers such as poly(glycolide) results from the preferential diffusion of water into the amorphous regions. As a result, degradation occurs in the amorphous regions leaving behind crystalline-rich portions that hydrolyze more slowly (43, 44). Initially amorphous polymers such as PLA and 75:25 PLG have also been found to exhibit heterogeneous degradation, reportedly through the crystallization of polymer fragments rich in L-lactide during the course of hydrolysis (43).

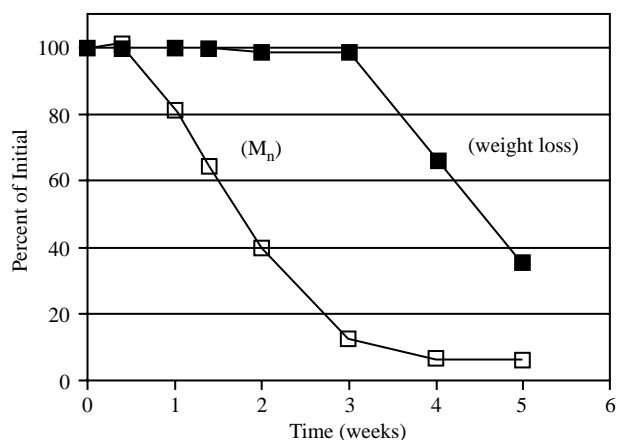


Fig. 4 Degradation profiles of cylindrical PLG rods incubated at 37°C in vitro. Degradation is exhibited by the percent change over time of the sample weight, ■, and number-average molecular weight (M_n), □. (Adapted from Ref. 25.)

Many reviews describing the development of polyester-based biodegradable drug delivery systems have been published (21, 22, 40, 22, 45). Much work has been performed in evaluating these materials for delivery of proteins (36) and vaccines (46). More recently, novel chemical and engineering approaches have been discovered to make crosslinked hydrogels and three-dimensional porous matrices for a wide range of biomedical applications including, protein and gene delivery as well as cell therapy.

Polyanhydrides

Historically, polyanhydrides were developed in the textile industry during the first half of the 20th century as alternate fiber materials (47, 48). However, the modern polyanhydrides that are currently under investigation as drug delivery platforms represent a novel class of polymer that, unlike the polyesters, has been specifically developed for biodegradable applications. In particular, these polyanhydrides were specifically prepared in attempts to produce surface-eroding dosage forms.

The anhydride linkages of these polymers are, in general, more hydrolytically labile than the polyester bond. In order to achieve a surface-eroding mechanism, polymers are generally prepared from very hydrophobic monomers in order to minimize water penetration into the bulk of the device. By doing this, hydrolysis of the labile anhydride linkages would be restricted to the outer exposed surfaces of the polymer device. A wide variety of aliphatic and aromatic monomers have been used to prepare surface-eroding polyanhydride polymers (Table 5). Aliphatic polyanhydrides are normally prepared from dicarboxylic acids such as adipic acid, sebacic acid (SA), dodecanoic acid, and fumaric acid (FA). Additionally, diacidic fatty acid monomers (fatty acid dimers, or FAD) have also been used. Polymers with increasing hydrophobicity can be made from aromatic monomers including phthalic acid and various carboxyphenoxyalkanes such as CPM, CPP, and CPH (Table 5). High-molecular-weight polyanhydrides are usually synthesized by first converting the dicarboxylic acid monomer to mixed anhydride prepolymers using acetic anhydride followed by polymerization of prepolymers using polycondensation reaction in the melt.

Typically, homopolymers are not studied because they possess unfavorable characteristics rendering their handling and manufacture difficult. Poly(SA), poly(CPP), and poly(FA) are semicrystalline and thus suffer from either being brittle or having high T_m . Conversely, poly(FAD) is a liquid. Therefore, polyanhydrides are often prepared as copolymers of aliphatic and/or aromatic monomers. The

most common copolymers under investigation in drug delivery applications include poly(FAD-SA) and poly(CPP-SA).

Studies on aliphatic polyanhydrides have shown that increasing the alkyl chain length (e.g., from $n = 4$ to 12) of the dicarboxylic acid monomers increases polymer hydrophobicity resulting in a decrease in both polymer degradation and drug release rates (49). These trends in degradation were observed both in vitro and in vivo. While poly(FAD) is amorphous, the degree of crystallinity observed in poly(FAD-SA) copolymers increased directly once the SA content was raised above 30 mole%. In spite of this, raising the SA content has been shown to increase the degradation rate. This can be attributed to the greater hydrophilicity of SA relative to FAD (50). In contrast, poly(CPP-SA) copolymers are semicrystalline across the entire range of composition, yielding the lowest degree of crystallinity when the SA content is between 15 and 70 mole%. While having little influence on polymer crystallinity, changes in copolymer composition can affect polymer degradation rates. Figure 5 shows the relative in vitro degradation rates of poly(CPP-SA) copolymers containing 0–79 mole% SA. In these samples, degradation of 100% poly(CPP) was estimated (by extrapolation) to require over 3 years, whereas copolymers containing the highest amounts of SA degraded as quickly as 1–2 weeks. The initial motivation for synthesizing polyanhydrides was to prepare devices that degrade at a constant rate by a surface-erosion process. To this end, nearly zero-order degradation profiles were achieved for over a 6-day period for discs prepared from poly(SA) and also from poly(CPP-SA) containing 80 and 20% SA. Similarly, poly(CPP) and 85:15 poly(CPP-SA) also showed constant degradation profiles that lasted for a period of several months.

Degradation of poly(FAD-SA) has also been reported to be nearly zero-order based on the rate of SA release from the polymer (50). However, evidence suggests that devices prepared from these polymers do not necessarily degrade by a purely surface-erosion mechanism. Cumulative release profiles of individual monomers released from a degrading poly(CPP-SA) matrix indicated that SA was released at a much faster rate than was CPP. Other studies have shown that, similar to certain polyesters, amorphous regions in poly(CPP-SA) will preferentially hydrolyze first, leaving behind a network of pores surrounding the remaining crystalline-rich structures (48). These findings suggest that while polyanhydrides are able to provide degradation behavior that is consistent with a surface-erosion mechanism, the actual processes involved in polymer degradation and erosion of the device can be more complicated.

Table 5 Common polyanhydride structures

Polyanhydride linkage	$\left(\text{C}(=\text{O})-\text{R}_1-\text{C}(=\text{O})-\text{O} \right)_n$
poly(sebacic acid), SA	$\left(\text{C}(=\text{O})-(\text{CH}_2)_8-\text{C}(=\text{O})-\text{O} \right)_n$
poly(fumaric acid), FA	$\left(\text{C}(=\text{O})-\text{CH}=\text{CH}-\text{C}(=\text{O})-\text{O} \right)_n$
poly(erucic acid dimer) or poly(FAD), (FAD, fatty acid dimer)	$\left(\text{C}(=\text{O})-(\text{CH}_2)_{12}-\text{CH}\left(\text{CH}_2\right)_7\text{CH}\left(\text{CH}_2\right)_7-\text{C}(=\text{O})-\text{O} \right)_n$
poly(terephthalic acid), TA (para-) poly(isophthalic acid), IPA (meta-)	$\left(\text{C}(=\text{O})-\text{C}_6\text{H}_4-\text{C}(=\text{O})-\text{O} \right)_n$
poly[bis(<i>p</i> -carboxyphenoxy) alkanes] <div> <div>$m = 1$</div> <div>poly[1-bis(<i>p</i>-carboxyphenoxy)methane],</div> <div>CPM</div> </div> <div> <div>$m = 3$</div> <div>poly[1,3-bis(<i>p</i>-carboxyphenoxy)propane],</div> <div>CPP</div> </div> <div> <div>$m = 6$</div> <div>poly[1,6-bis(<i>p</i>-carboxyphenoxy)hexane],</div> <div>CPH</div> </div>	$\left(\text{C}(=\text{O})-\text{C}_6\text{H}_4-\text{O}-(\text{CH}_2)_m-\text{O}-\text{C}_6\text{H}_4-\text{C}(=\text{O})-\text{O} \right)_n$

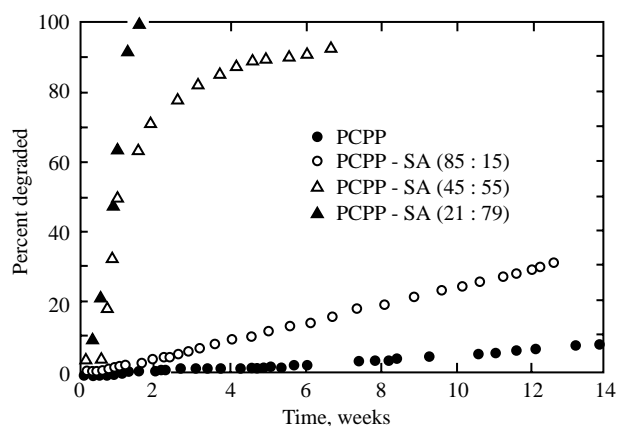


Fig. 5 Degradation profiles (percent weight loss) of compression-molded poly[bis(*p*-carboxyphenoxy) propane anhydride] (PCPP) containing varying ratios of sebacic acid (SA). The copolymers were incubated in 0.1M pH 7.4 phosphate buffer at 37°C. (From Ref. 47.)

The release of a number of drugs from polyanhydride matrices has been studied including ciprofloxacin (49), *p*-nitroaniline (47, 51), cortisone acetate (51), insulin (51), and a variety of proteins (50). In many instances, drug release was reported to coincide with polymer degradation. The biocompatibility of polyanhydrides has also been assessed in a number of studies (51–54). The only commercial product that has received regulatory approval is Gliadel[®]. This product is a thin wafer (1.45 cm in diameter, 1 mm in thickness) containing 7.7 mg carmustine(BCNU), and is prepared from 20:80 poly(CPP-SA). As many as 8 wafers are implanted in the cavity created after surgical removal of recurrent glioblastoma multiforme, an aggressive form of brain cancer. Studies show that BCNU release from Gliadel[®] wafers is controlled by both drug diffusion and erosion of the polyanhydride matrix (55).

Polyorthoesters

A series of polyorthoesters has been under development since 1970 in efforts to prepare surface-eroding biodegradable polymers specifically for drug delivery applications (56–61). The first hydrolytically labile polyorthoesters were synthesized by polycondensation of a diol (either 1,6-hexanediol, HD, or *cis/trans*-1,4-cyclohexane dimethanol, CHDM) with an orthoester (diethoxy tetrahydrofuran, DETHF). This polymer was initially called Chronomer[™] and was developed at the Alza Corporation (Palo Alto, CA) by Choi and Heller (Table 6). Later the name was changed to Alzamer[®];

development of these polymers has since been discontinued. Hydrolysis of Alzamer[®] produces the corresponding diol along with γ -hydroxybutyrolactone which rapidly opens its ring structure to form γ -hydroxybutyric acid. Because polymer hydrolysis was acid catalyzed, production of this by-product accelerated the rate of degradation of the remaining polymer. In order to avoid catastrophic self-accelerated hydrolysis, the polymer was usually stabilized by the addition of an inorganic base such as sodium carbonate.

In attempts to overcome the limitations of these polymers, Heller developed alternate polyorthoesters at SRI International (Palo Alto, CA). The second generation polyorthoesters were prepared by the addition of a diol with the cyclic diketene acetal, 3,9-diethylidene-2,4,8,10-tetraoxaspiro [5.5] undecane (DETOU) (Table 5). When prepared with varying ratios of these two diols (i.e., HD and CHDM), polyorthoesters possessing a wide range of mechanical properties could be obtained. Increasing the HD composition from 0 to 100% lowered the polymer T_g from over 100°C to about 20°C accordingly (57). In addition, the T_g of DETOU polyorthoesters can be further reduced by synthesis using linear diols of successively longer alkyl chain length. Increasing the alkyl length from $n = 6$ to $n = 12$ lowers T_g from 20 to 0°C (59). The reduction in glass transition temperatures of these polymers reflects the increased chain flexibility of the diol component.

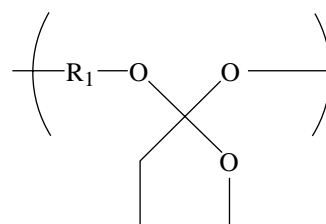
DETOU polyorthoesters, because of the hydrophobic nature of the monomers are found to degrade relatively slowly in vitro with only 15% weight loss after about 150 days and 60% weight loss after 325 days. These polymers are also shown to hydrate slowly (i.e., adsorbing about 0.3–0.75 wt% of water). While degradation of the DETOU polyorthoester is also acid-catalyzed, the degradation products do not generate any acidic species. Heller suggested that a faster polymer degradation could be achieved by the addition of acidic excipients into the polymer matrix. When 1 wt% suberic acid was added to a DETOU polyorthoester polymer prepared with 1,6-hexanediol, a complete release of naltrexone pamoate was achieved in vitro in 30 days (57). Conversely, the polymer can be stabilized against hydrolysis by the addition of a base such as magnesium hydroxide.

One potential limitation to the use of blended excipients to regulate polymer degradation rates is the likelihood that the excipient could diffuse out of the device before polymer degradation has terminated. For acidic additives, this would cause polymer erosion to slow down significantly, meaning that polymer residues would remain in the implanted tissue site for a long time. To avoid this possibility, novel approaches have been developed to

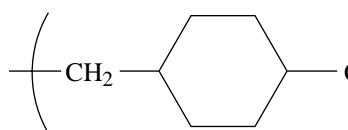
Polyorthoester linkage

$$\left(\text{R}_1 - \text{O} - \underset{\begin{array}{c} | \\ \text{O}-\text{R}_2 \\ | \end{array}}{\text{CH}} - \text{O} \right)_n$$

Polyorthoester I, with:



$\text{R}_1 = \textit{cis/trans}$ -cyclohexyl dimethanol, CHDM:



$\text{R}_1 = 1,6\text{-hexanediol}, \text{HD}: \quad \left(\text{CH}_2 \right)_6$

retain acidic species within the polymer. For example, hydrolytically labile ester bonds have been incorporated into the polymer network. Degradation of these groups produced acidic species that remained covalently attached to that polymer system over time.

Reaction of a triol such as 1,2,6-hexanetriol with a DETOU prepolymer permits the formation of a crosslinked network. Rod-shape polymer systems (2.4 mm in diameter) were fabricated containing 30 wt% levonorgestrel and 7 wt% micronized $\text{Mg}(\text{OH})_2$ as a stabilizer. These devices contained 1 wt% 9,10-dihydroxystearic acid to modify polymer erosion rate. SEM photomicrographs of the rods show evidence of surface erosion following implantation for up to 16 weeks. A degradation zone is observed to move progressively toward the center of the rod while, at the same time, erosion causes this region to become increasingly more porous as the polymer continues to degrade (56).

Heller and colleagues at Advanced Polymer Systems (Redwood City, CA) continue to develop additional polyorthoesters with a variety of physical properties and potential applications (60,61). Generally speaking, polyorthoesters do possess the potential to exhibit surface-eroding behavior. However, several issues that may limit the commercial application of this class of polymers are still present. One issue is that synthesis of these polymers involves complicated monomers and polymerization chemistry. The second issue is, to date, the toxicology and biocompatibility of these polymers and their degradation products have not yet been fully characterized. Finally, the requirement for pH-regulating additives such as acids and bases can be undesirable in applications where the drug stability is affected.

Phosphorus-Containing Polymers

There are two common classes of phosphorus-containing polymers, the phosphazenes (Table 7) and phosphoesters (Table 8). In both cases, polymers possessing a range of chemical, physical, and mechanical properties can be synthesized through simple changes in monomer side-chain substitution. Polyphosphazenes contain alternating phosphorus-nitrogen double and single bonds and are synthesized by reaction of poly(dichlorophosphazene) with organic nucleophiles such as alkoxides, aryloxides, or amines (62, 63). Polymer prodrugs have been prepared in which the drug entity is covalently attached to the polyphosphazene. Examples of these systems include the drugs procaine, benzocaine, and heparin. More typically, however, side-chain substituents are selected in order to vary polymer hydrophobicity as well as the hydrolytic stability of the phosphazene bond. Copolymers containing

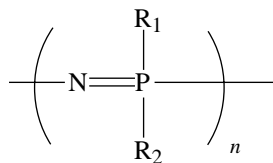
varying ratios of imidazole and methylphenoxy side-chains have been prepared (Table 7) (62). Increasing the imidazole content of these polymers increases the reactivity toward hydrolysis of the phosphazene linkages. After 600 hours in phosphate buffer, a polyphosphazene containing 20% imidazole lost only 4% weight due to degradation, whereas another polymer containing 45% imidazole lost over 30% weight. The suitability of these polymers as drug delivery platforms was evaluated using *p*-nitroaniline, progesterone, and bovine serum albumin as model drugs (62).

Polyphosphazenes have also been prepared using modified amino acid ester side-chain substituents (63) (Table 7). Studies show that the polymer containing the smallest hydrophobic side-chain constituent, glycine ethyl ester, exhibited the most rapid degradation. This was attributed to the lower steric hindrance that the small side-chain group could provide in protecting the phosphazene bond from hydrolytic attack. Cast films of this polymer were degraded in vitro by 40% (as determined by percent mass loss) after 1200 hours (50 days), whereas polymers containing longer amino acid side-chain constituents such as alanine ethyl and benzylalanine ethyl esters lost only 10–15% weight during the same time interval. Substitution at the α -carbon of the amino acid side-chain, therefore, seemed to provide a greater influence on the rate of polymer hydrolysis than the overall hydrophobicity of the ester substituent. Characterization of polymer degradation products indicated that these polymers degraded to phosphates, glycine or alanine, ethanol or benzyl alcohol, and ammonia. Drug release, as was demonstrated using ethacrynic acid and the azo dye Biebrich Scarlet, followed the same trends as did polymer degradation whereby the fastest release rates were exhibited by the glycine ethyl ester polymer. Polyphosphazenes have been studied for the delivery of proteins (64), naproxen (65,66) and colchicine (67), as well as in periodontal treatments (68).

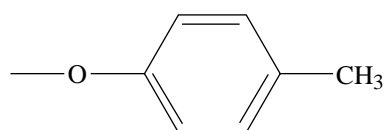
The class of phosphoester-based polymers includes polyphosphates, polyphosphonates, and polyphosphites (Table 8). A series of phosphoesters based on bisphenol A (BPA) have been prepared and evaluated in drug delivery applications (69, 70). Polymerization was carried out by interfacial polycondensation reaction of the diol (BPA) with a phosphodichloridate. Selection of either an alkyl- or alkoxide-substituted dichloridate resulted in the synthesis of either the phosphonate or phosphate polymer, respectively. High-molecular-weight polymers (M_w between 20–40 kDa) with T_g of 103–115°C were synthesized. The uptake of water and swelling of these polymers, both in vitro and in vivo, increased with the increasing relative

Table 7 Common phosphorus-containing polymers — polyphosphazenes

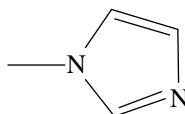
Polyphosphazene linkage

R₁ and R₂ side-chain substituents:

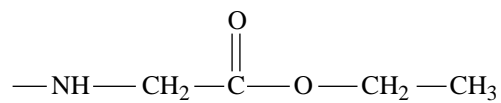
methoxyphenoxy



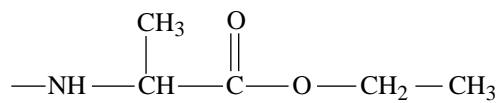
imidazole



ethyl glycine



ethyl alanine



ethyl benzylalanine

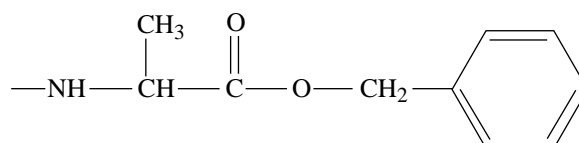
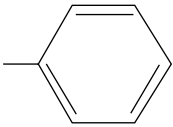
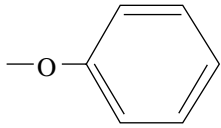


Table 8 Common phosphorus-containing polymers — polyphosphoesters

Polyphosphate linkage	$\left(\text{O} - \text{R}_1 - \text{O} - \underset{\text{R}_2 - \text{O}}{\overset{\text{O}}{\text{P}}} \right)_n$
Polyphosphonate linkage	$\left(\text{O} - \text{R}_1 - \text{O} - \underset{\text{R}_2}{\overset{\text{O}}{\text{P}}} \right)_n$
Polyphosphite linkage	$\left(\text{O} - \text{R}_1 - \text{O} - \underset{\text{R}_2}{\overset{\text{O}}{\text{P}}} \right)_n$
bisphenol A (BPA) polyphosphoesters	
	$\left(\text{O} - \text{C}_6\text{H}_4 - \underset{\text{CH}_3}{\overset{\text{CH}_3}{\text{C}}} - \text{C}_6\text{H}_4 - \text{O} - \underset{\text{R}_1}{\overset{\text{O}}{\text{P}}} \right)_n$
BPA-EP, poly(ethyl phosphonic acid-BPA)	$\text{R}_1 = \text{—CH}_2\text{—CH}_3$
BPA-EOP, poly(ethyl phosphate acid-BPA)	$\text{R}_1 = \text{—O—CH}_2\text{—CH}_3$
BPA-PP, poly(phenyl phosphonic acid-BPA)	$\text{R}_1 = $ 
BPA-POP, poly(phenyl phosphate acid-BPA)	$\text{R}_1 = $ 

hydrophilicity of the phosphate substituent. The aliphatic polyphosphate BPA-EOP exhibited the greatest water uptake and swelling during degradation, whereas the aromatic polyphosphonate BPA-PP exhibited the least. Degradation rates increased with polymer hydrophilicity. For example, the percentage weight loss in BPA-EOP and BPA-PP samples was 35 and 10%, respectively, following implantation in rabbits for 30 weeks.

Other diols have been used to prepare polyphosphoesters including 1,4-bis(hydroxyethyl)-terephthalate (BHET) and *trans*-cyclohexane dimethanol (T-CHDM). The synthesis and characterization of various polyphosphoesters have recently been reviewed (71). Of particular interest is a polyphosphoester prepared by polycondensation of an oligomeric polyester diol with ethylphosphodichloridate (EOP). In this case, the diol pre-polymer was synthesized using ethylene glycol or 1,2-propylene glycol to initiate the ring-opening condensation reaction of a cyclic lactide. This polymerization was carried out in order to prepare a low-molecular-weight oligomeric polyester diol that was then polymerized with EOP. The product, actually a poly(lactide-co-phosphate), exhibited faster degradation as the phosphate content increased. This polymer is currently being investigated by Guilford Pharmaceuticals (Baltimore, MD) for delivery of the chemotherapeutic agent paclitaxel (72).

NEW DIRECTIONS

A wide variety of approaches are continuously being pursued in the quest for improved biodegradable drug delivery systems. New and modified polymer chemistries that offer distinctive degradation and drug delivery attributes are being identified and evaluated. At the same time, innovative engineering and manufacturing methods are under development to fabricate devices and physical platforms having novel three-dimensional structural attributes. In certain instances, these products make it possible to treat entirely new therapeutic applications using biodegradable devices. One approach is through the preparation of biodegradable crosslinked polymer networks. When crosslinking is carried out using hydrophilic polymers, hydrogels that absorb a significant amount of water (for example, greater than 10 wt%) can be produced. Crosslinking can generally be achieved via three different means: by formation of covalent bonds, by ionic interactions, and by formation of highly entangled chain networks. Degradation of crosslinked polymer networks can take place by either breaking the crosslinks or the polymer backbone bonds, or by slow disentanglement and

dissolution of the polymer network. The highly swollen, three-dimensional structure of hydrogels has been an attractive feature to researchers in developing delivery systems for macromolecular drugs. A number of polymers derived from naturally-occurring constituents including alginate (73), dextran (74), collagen (75), gelatin (76), and albumin (77) have been employed to prepare hydrogels. Alternatively, proven biocompatible polymers such as PLA and PLG (78) and poly(γ -benzyl-L-glutamic acid) (79) have also been used to prepare crosslinked, synthetic hydrogels. In addition, a variety of biodegradable hydrogels based on other types of constituents have also been synthesized (80).

New approaches have been taken to design and engineer novel biodegradable drug delivery platforms. Highly porous polymer platforms with controlled pore density and size have recently been synthesized by the formation of either sponge-like or entangled fibrous matrices (81, 82). These systems are currently being investigated as extracellular matrices for cell therapy. Matrices loaded with viable cells capable of releasing bioactive agents are being investigated as implantable artificial organs. In certain studies, hepatocytes incorporated into PLA matrices were shown to remain viable for 14 days. In other studies, transplanted hepatocytes were shown to be capable of forming new liver-like tissue (83). Another example is the artificial pancreas in which the islet cells were incorporated into fibrous PGA matrices for the glucose-responsive release of insulin (84). Other polymers, such as polyphosphazene, have also been investigated for cell therapy (85), while alginate hydrogels have been explored for islet transplantation (86). The implantation of biodegradable polymer platforms seeded with osteoblast cells (87) or incorporated with bone-derived growth factors (88) have been studied for induction of skeletal tissue growth. The growing applications of biodegradable polymer scaffolds in cell therapy and tissue engineering have been recently reviewed in several articles (81, 83, 89, 90).

SUMMARY

This is an exciting time for the development of biodegradable drug delivery platforms. The sustained delivery of drugs to the body has, practically speaking, become a reality as numerous products have passed regulatory review and proven to be commercially successful. New polymers continue to be developed while, at the same time, new therapeutic applications are being increasingly recognized.

REFERENCES

- Robinson, J.R.; Lee, V.H.L. *Controlled Drug Delivery*; From Series: Drugs and the Pharmaceutical Sciences Marcel Dekker, Inc.: New York, 1987; 29.
- Banker, G.S. Pharmaceutical Applications of Controlled Release – An Overview of the Past, Present, and Future. *Medical Applications of Controlled Release*; Langer, R.S., Wise, D.L., Eds.; CRC Press: Boca Raton, Florida, 1984; II.
- Physicians Desk Reference*, 53rd Ed.; Medical Economics Company, Inc.: Montvale, New Jersey, 1999.
- Fraser, I.S.; Tiitinen, A.; Affandi, B.; Brache, V.; Croxatto, H.B.; Diaz, S.; Ginsburg, J.; Gu, S.; Holma, P.; Johansson, E.; Meirik, O.; Mishell, D.R.; Nash, H.A.; von Schoultz, I.; Sivin, I. Norplant® Consensus Statement and Background Review. *Contraception* **1998**, *57*, 1–9.
- Harrison, P.F.; Rosenfield, A. Research, Introduction, and Use: Advancing from Norplant. *Contraception* **1998**, *58*, 323–334.
- Heller, J. Biodegradable Polymers in Controlled Drug Delivery. Critical Reviews in Therapeutic Drug Carrier Systems **1984**, *1* (1), 39–90.
- Pitt, C.G.; Schindler, A. Biodegradation of Polymers. *Controlled Drug Delivery*; Bruck, S.D., Ed.; CRC Press: Boca Raton, Florida, 1983; 1.
- Sidman, K.R.; Schwöpe, A.D.; Steber, W.D.; Rudolph, S.E. Use of Synthetic Polypeptides in the Preparation of Biodegradable Delivery Systems for Narcotic Antagonists. NIDA Research Monograph **1981**, *28*, 214–231.
- Charles, E.L.; Buffalo, N.Y. U.S. Patent 2,668,162, 1954(to E.I. DuPont de Nemours).
- Schmitt, E.E.; Polistina, R.A. U.S. Patent 3,297,033, 1967(to American Cyanamid).
- Kulkarni, R.K.; Pani, K.C.; Neuman, C.; Leonard, F.; Polylactic Acid for Surgical Implants, Technical Report 6608, Walter Reed Army Medical Center (Project 3A013001A9-1C), 1966.
- Wasserman, D. U.S. Patent, 3,375,008, 1971.
- Yolles, S.; Eldridge, J.E.; Woodland, J.H.R. Sustained Delivery of Drugs from Polymer/Drug Mixtures. *Polym. News* **1970**, *1*, 9–15.
- Yolles, S.; Leafe, T.D.; Meyer, F.J. Timed-release Depot for Anticancer Agents. *J. Pharm. Sci.* **1975**, *64* (1), 115–116.
- Yolles, S.; Leafe, T.; Ward, L.; Boettner, F. Controlled Release of Biologically Active Drugs. *Bull. Parent. Drug Assoc.* **1976**, *30* (6), 306–312.
- Yolles, S.; Woodland, J.H.R. Long-acting Delivery Systems for Narcotic Antagonists. I. **1973**, *16* (8), 897–901.
- Wise, D.L. Controlled Release for Use in Treatment of Narcotic Addiction. *Medical Applications of Controlled Release*; Langer, R., Wise, D.L., Eds.; CRC Press: Boca Raton, Florida, 1984; II.
- Schwöpe, A.D.; Wise, D.L.; Howes, J.F. Development of Polylactic/glycolic Acid Delivery Systems for Use in Treatment of Narcotic Addiction. NIDA Research Monograph **1976**, *4*, 13–18.
- Kronenthal, R.L. Biodegradable Polymers in Medicine and Surgery. *Polymers in Medicine and Surgery*; From series, Kronenthal, R.L., Oser, Z., Martin, E., Eds.; Polymer Science and Technology; Plenum Press: New York, 1975; 8.
- Heller, J.; Baker, R.W. Theory and Practice of Controlled Drug Delivery from Bioerodible Polymers. *Controlled Release of Bioactive Materials*; Baker, R., Ed.; Academic Press: New York, 1980; 1–17.
- Lewis, D.H. Controlled Release of Bioactive Agents from Lactide/Glycolide Polymers. *Drugs Pharm. Sci.* **1990**, *45*, 1–41.
- Holland, S.J.; Tighe, B.J.; Gould, P.L. Polymers for Biodegradable Medical Devices. 1. The Potential of Polyesters As Controlled Macromolecular Release Systems. *J. Control. Rel.* **1986**, *4*, 155–180.
- Crotts, G.; Park, T.G. Protein Delivery from Poly(lactic-co-glycolic Acid) Biodegradable Microspheres: Release Kinetics and Stability Issues. *J. of Microencap.* **1998**, *15* (6), 699–713.
- Putney, S.D.; Burke, P.A. Improving Protein Therapeutics with Sustained-Release Formulations. *Nat. Biotechnol.* **1998**, *16* (2), 153–157.
- Sanders, L.M.; Kent, J.S.; McRae, G.I.; Vicery, B.H.; Tice, T.R.; Lewis, D.H. Controlled Release of a Luteinizing Hormone-releasing Hormone Analog from Poly(D,L-lactide-co-glycolide) Microspheres. *J. Pharm. Sci.* **1984**, *73* (9), 1294–1294.
- Jacobi, G.H.; Wenderoth, U.K.; Ehrenthal, W.; von Wallenberg, H.; Spindler, H.W.; Hohenfellner, R. Endocrine and Clinical Evaluation of 107 Patients With Advanced Prostatic Carcinoma Under Long-term Parnasal Buserelin or Intramuscular Decapeptyl Depot Treatment. *Am. J. Clin. Oncol.* **1988**, *11* (Suppl.1), S36–S43.
- Dlugi, A.M.; Miller, J.D.; Knittle, J. Lupron Depot (Leuprolide Acetate for Depot Suspension) in the Treatment of Endometriosis: A Randomized, Placebo-controlled, Double-blind Study. *Lupron Study Group. Fertil. Steril.* **1990**, *54* (3), 419–427.
- Akaza, H.; Usami, M.; Koiso, K.; Kotake, T.; Aso, Y.; Nijima, T. Long-term Clinical Study on Luteinising Hormone-releasing Hormone Agonist Depot Formulation in the Treatment of Stage D Prostatic Cancer. The TAP-144-SR Study Group. *Jpn. J. Clin. Oncol.* **1992**, *22* (3), 177–184.
- Neely, E.K.; Hintz, R.L.; Parker, B.; Bachrach, L.K.; Cohen, P.; Olney, R.; Wilson, D.M. Two-year Results of Treatment With Depot Leuprolide Acetate for Central Precocious Puberty. *J. Pediatr.* **1992**, *121* (4), 634–640.
- Hunter, S.J.; Shaw, J.A.M.; Lee, K.O.; Woods, P.J.; Atkinson, A.B.; Bevan, J.S. Comparison of Monthly Intramuscular Injections of Sandostatin LAR with Multiple Subcutaneous Injections of Octreotide in the Treatment of Acromegaly; Growth Hormone Secretion. *Clin. Endocrinol.(Oxf.)* **1998**, *50*, 245–251.
- Davies, P.H.; Stewart, S.E.; Lancranjan, L.; Sheppard, M.C.; Stewart, P.M. Long-term Therapy with Long-acting Octreotide (Sandostatin-LAR) for the Management of Acromegaly. *Clin. Endocrinol.(Oxf.)* **1998**, *48* (3), 311–316.
- Rubin, J.; Ajani, J.; Schirmer, W.; Venook, A.P.; Bukowski, R.; Pommier, R.; Saltz, L.; Dandona, P.; Anthony, L. Octreotide Acetate Long-acting Formulation Versus Open-label Subcutaneous Octreotide Acetate in

- Malignant Carcinoid Syndrome. *J. Clin. Oncol.* **1999**, *17* (2), 600–606.
32. Johnson, O.L.; Jaworowicz, W.; Cleland, J.L.; Bailey, L.; Charnis, M.; Duenas, E.; Wu, C.; Shepard, D.; Magil, S.; Last, T.; Jones, A.J.; Putney, S.D. The Stabilization and Encapsulation of Human Growth Hormone into Biodegradable Microspheres. *Pharm. Res.* **1997**, *14* (6), 730–735.
 33. Berner, B.; Dinh, S. Fundamental Concepts in Controlled Release. *Treatise on Controlled Drug Delivery*; Kydonieus, A., Ed.; Marcel Dekker, Inc.: New York, 1992.
 34. Heller, J. Fundamentals of Polymer Science. *Controlled Drug Delivery*; Robinson, J.R., Lee, V.H.L., Swarbrick, J., Eds.; From Series: Drugs and the Pharmaceutical Sciences, Chapter 3; Marcel Dekker, Inc.: New York, 1987; 29.
 35. Bruck, S.D.; Mueller, E.P. Materials and Biological Aspects of Synthetic Polymers in Controlled Drug Release Systems: Problems and Challenges. *Critical Reviews in Therapeutic Drug Carrier Systems* **1988**, *5* (3), 171–187.
 36. Cohen, S.; Yoshioka, T.; Lucarelli, M.; Hwang, L.H.; Langer, R. Controlled Delivery Systems for Proteins Based on Poly(Lactic/glycolic Acid) Microspheres. *Pharm. Res.* **1991**, *8* (6), 713–720.
 37. Joshi, A.; Himmelstein, K.J. Dynamics of Controlled Release from Bioerodible Matrices. *J. Control. Rel.* **1991**, *15*, 95–104.
 38. Lee, P.I. Diffusional Release of a Solute from a Polymeric Matrix — Approximate Analytical Solutions. *J. Membr. Sci.* **1980**, *7*, 255–275.
 39. Hopfenberg, H.B. Controlled Release from Eroderible Slabs, Cylinders, and Spheres. *Controlled Release Polymeric Formulations*; Paul, D.R., Harris, F.W., Gould, R.F., Eds.; American Chemical Society: Washington DC, 1976.
 40. Kopecek, J.; Ulbrich, K. Biodegradation of Biomedical Polymers. *Prog. Polym. Sci.* **1983**, *9*, 1–58.
 41. Gilding, D.K.; Reed, A.M. Biodegradable Polymers for Use in Surgery — Polyglycolic/Poly(Lactic Acid) Homo- and Copolymers: 1. Polymer **1979**, *20*, 1459–1464.
 42. Li, S.M.; Garreau, H.; Vert, M. Structure –property Relationships in the Case of the Degradation of Massive Poly-(α -hydroxy Acids) in Aqueous Media, Part 1. Poly(DL-Lactic Acid). *J. Mater. Sci. Mater. in Med.* **1990**, *1*, 123–130.
 43. Reed, A.M.; Gilding, D.K. Biodegradable Polymers for Use in Surgery — Polyglycolic/Poly(Lactic Acid) Homo- and Copolymers: 2. In Vitro Degradation. *Polymer* **1981**, *22*, 494–498.
 44. Fukuzaki, H.; Yoshida, M.; Asano, M.; Kumakura, M. Synthesis of Copoly(D,L-Lactic Acid) With Relatively Low Molecular Weight and In Vitro Degradation. *Eur. Polym. J.* **1989**, *25* (10), 1019–1026.
 45. Jain, R.; Shah, N.H.; Malick, A.W.; Rhodes, C.T. Controlled Drug Delivery by Biodegradable Poly(ester) Devices: Different Preparative Approaches. *Drug Dev. and Ind. Pharm.* **1998**, *24* (8), 703–727.
 46. O'Hagan, D.T. Recent Advances in Vaccine Adjuvants for Systemic and Mucosal Administration. *J. Pharm. Pharmacol.* **1997**, *49*, 1–10.
 47. Leong, K.W.; Brott, B.C.; Langer, R. Bioerodible Polyanhydrides as Drug-carrier Matrices. I: Characterization, Degradation and Release Characteristics. *J. Biomed. Mat. Res.* **1985**, *19*, 941–955.
 48. Gopferich, A. Biodegradable Polymers: Polyanhydrides. In *The Encyclopedia of Controlled Release*; Mathiowitz, E., Ed.; Wiley: New York, 1999; 60–71.
 49. Domb, A.J.; Nudelman, R. In Vivo and In Vitro Elimination of Aliphatic Polyanhydrides. *Biomaterials* **1995**, *16* (4), 319–323.
 50. Tabata, Y.; Gutta, S.; Langer, R. Controlled Delivery Systems for Proteins Using Polyanhydride Microspheres. *Pharm. Res.* **1993**, *10* (4), 487–496.
 51. Leong, K.W.; Kost, J.; Mathiowitz, E.; Langer, R. Polyanhydrides for Controlled Release of Bioactive Agents. *Biomaterials* **1986**, *7*, 364–371.
 52. Laurencin, C.T.; Pierre-Jacques, H.M.; Langer, R. Toxicology and Biocompatibility Considerations in the Evaluation of Polymeric Materials for Biomedical Applications. **1990**, *10* (3), 549–570.
 53. Leong, K.W.; D'Amore, P.; Marletta, M.; Langer, R. Bioerodible Polyanhydrides as Drug-Carrier Matrices. II: Biocompatibility And Chemical Reactivity. *J. Biomed. Mat. Res.* **1986**, *20*, 51–64.
 54. Tamada, J.; Langer, R. The Development of Polyanhydrides for Drug Delivery Applications. *J. Biomater. Sci., Polym. Ed.* **1992**, *3* (4), 315–353.
 55. Dang, W.; Daviau, T.; Brem, H. Morphological Characterization of Polyanhydride Biodegradable Implant Gliadel During In Vitro and In Vivo Erosion Using Scanning Electron Microscopy. *Pharm. Res.* **1996**, *13* (5), 683–691.
 56. Heller, J. Controlled Drug Release from Poly(Ortho Esters) — A Surface Eroding Polymer. *J. Control. Rel.* **1985**, *2*, 167–177.
 57. Heller, J. Development of Poly(Ortho Esters): A Historical Overview. *Biomaterials* **1990**, *11*, 659–665.
 58. Heller, J. Poly(Ortho Esters). *Adv. Polym. Sci.* **1993**, *107*, 41–92.
 59. Heller, J. Poly(Ortho Esters). *The Encyclopedia of Controlled Release*; Mathiowitz, E., Ed.; Wiley: New York, 1999; 852–874.
 60. Heller, J.; Barr, J.; Ng, S.Y.; Shen, H-R.; Schwach-Abdellaoui, K.; Emmahl, S.; Rothen-Weinhold, A.; Gurny, R. Poly(Ortho Esters)- Their Development and Some Recent Applications. *Eur. J. Pharm. Biopharm.* **2000**, *50* (1), 122–128.
 61. Rothen-Weinhold, A.; Heller, J.; Barr, J.; Gurny, R. Poly(Ortho Esters) Implants for the Sustained Delivery of a Protein: Factors Influencing the Release Behavior of BSA In Vitro. *Proc Intl. Symp. Control. Rel. Bioact. Mater.* **2000**, *27*, 8004.
 62. Laurencin, C.T.; Koh, H.J.; Neenan, T.X.; Allcock, H.R.; Langer, R. Controlled Release Using a New Bioerodible Polyphosphazene Matrix System. *J. Biomed. Mat. Res.* **1987**, *21*, 1231–1246.
 63. Allcock, H.R.; Pucher, S.R.; Scopelianos, A.G. Poly [(Amino Acid Ester)phosphazenes] as Substrates for the Controlled Release of Small Molecules. *Biomaterials* **1994**, *15* (8), 563–569.
 64. Payne, L.G.; Andrianov, A.K. Protein Release from Polyphosphazene Matrices. *Adv. Drug Deliv. Rev.* **1998**, *31* (3), 185–196.
 65. Caliceti, P.; Lora, S.; Marsilio, F.; Veronese, F.M. Preparation and Characterization of Polyphosphazene-based Controlled Release Systems for Naproxen. *Farmaco* **1995**, *50* (12), 867–874.

66. Veronese, F.M.; Marsilio, F.; Caliceti, P.; De Filippis, P.; Giunchedi, P.; Lora, S. Polyorganophosphazene Microspheres for Drug Release: Polymer Synthesis, Microsphere Preparation, In Vitro and In Vivo Naproxen Release. *J. Control. Rel.* **1998**, *52* (3), 227–237.
67. Ibim, S.M.; el-Amin, S.F.; Goad, M.E.; Ambrosio, A.M.; Allcock, H.R.; Laurencin, C.T. In Vitro Release of Colchicine Using Poly(Phosphazenes): The Development of Delivery Systems for Musculoskeletal Use. *Pharm. Dev. Technol.* **1998**, *3* (1), 55–62.
68. Veronese, F.M.; Marsilio, F.; Lora, S.; Caliceti, P.; Passi, P.; Orsolini, P. Polyphosphazene Membranes and Microspheres in Periodontal Diseases and Implant Surgery. *Biomaterials* **1999**, *20* (1), 91–98.
69. Richards, M.; Dahivat, B.I.; Arm, D.M.; Brown, P.R.; Leong, K.W. Evaluation of Polyphosphates and Polyphosphonates as Degradable Biomaterials. *J. Biomed. Mat. Res.* **1991**, *25*, 1151–1167.
70. Saltzman, W.M.; Parsons-Wingerter, P.; Leong, K.W.; Lin, S. Fibroblast and Hepatocyte Behavior on Synthetic Polymer Surfaces. *J. Biomed. Mat. Res.* **1991**, *25*, 741–759.
71. Mao, H.-Q.; Kadivala, I.; Leong, K.W.; Zhao, Z.; Dang, W. Biodegradable Polymers: Poly(Phosphoester)s. *The Encyclopedia of Controlled Release*; Mathiowitz, E., Ed.; Wiley: New York, 1999; 45–60.
72. Harper, E.; Dang, W.; Lapidus, R.G.; Garver, R.I. Enhanced Efficacy of a Novel Controlled Release Paclitaxel Formulation (PACLIMER[®] Delivery System) for Local-regional Therapy of Lung Cancer Tumor Nodules in Mice. *Clin. Cancer Res.* **1999**, *5*, 4242–4248.
73. Wee, S.; Gombotz, W.R. Protein Release from Alginate Matrices. *Adv. Drug Deliv. Rev.* **1998**, *31* (3), 267–285.
74. Cadée, J.A.; van Luyn, M.J.; Brouwer, L.A.; Plantinga, J.A.; van Wachem, P.B.; deGroot, C.J.; den Otter, W.; Hennink, W.E. In Vivo Biocompatibility of Dextran-based Hydrogels. *J. Biomed. Mater. Res.* **2000**, *50* (3), 397–404.
75. Rao, K.P. Recent Developments of Collagen-based Materials for Medical Applications and Drug Delivery Systems. *J. Biomater. Sci., Polym. Ed.* **1995**, *7* (7), 623–645.
76. Ikada, Y.; Tabata, Y. Protein Release from Gelatin Matrices. *Adv. Drug Deliv. Rev.* **1998**, *31* (3), 287–301.
77. D'Urso, E.M.; Jean-Francois, J.; Doillon, C.J.; Fortier, G. Poly(Ethylene Glycol)-Serum Albumin Hydrogel as Matrix for Enzyme Immobilization: Biomedical Applications. *Artif. Cells Blood Substit. Immobil. Biotechnol.* **1995**, *23* (5), 587–595.
78. Sawhney, A.S.; Pathak, C.P.; van Rensburg, J.J.; Dunn, R.C.; Hubbell, J.A. Optimization of Photopolymerized Bioerodible Hydrogel Properties for Adhesion Prevention. *J. Biomed. Mat. Res.* **1994**, *28*, 831–838.
79. Markland, P.; Zhang, Y.; Amidon, G.L.; Yang, V.C. A pH- and Ionic Strength-responsive Polypeptide Hydrogel: Synthesis, Characterization, and Preliminary Protein Release Studies. *J. Biomed. Mat. Res.* **1999**, *47* (4), 595–602.
80. Kamath, K.R.; Park, K. Biodegradable Hydrogels in Drug Delivery. *Adv. Drug Deliv. Rev.* **1993**, *11*, 59–84.
81. Kim, B.-S.; Mooney, D.J. Development of Biocompatible Synthetic Extracellular Matrices for Tissue Engineering. *Tibtech* **1998**, *16*, 224–230.
82. Wald, H.L.; Sarakinos, G.; Lyman, M.D.; Mikos, A.G.; Vacanti, J.P.; Langer, R. Cell Seeding in Porous Transplantation Devices. *Biomaterials* **1993**, *14* (4), 270–278.
83. Murphy, W.L.; Mooney, D.J. Controlled Delivery of Inductive Proteins, Plasmid DNA, and Cells from Tissue Engineering Matrices. *J. Periodont. Res.* **1999**, *34*, 413–419.
84. Juang, J.-H.; Bonner-Weir, S.; Vacanti, J.P.; Weir, G.C. Outcome of Subcutaneous Islet Transplantation Improved by a Polymer Device. *Transplantation Proceedings* **1995**, *27* (6), 3215–3217.
85. Cohen, S.; Bano, M.C.; Cima, L.G.; Allcock, H.R.; Vacanti, J.P.; Vacanti, C.A.; Langer, R. Design of Synthetic Polymeric Structures for Cell Transplantation and Tissue Engineering. *Clin. Mater.* **1993**, *13* (1–4), 3–10.
86. Lim, F.; Sun, A.M. Microencapsulated Islets As Bioartificial Endocrine Pancreas. *Science* **1980**, *210* (21), 908–910.
87. Laurencin, C.T.; El-Amin, S.F.; Ibim, S.E.; Willoughby, D.A.; Attawia, M.; Allcock, H.R.; Ambrosio, A.A. A Highly Porous 3-Dimensional Polyphosphazene Polymer Matrix for Skeletal Tissue Regeneration. *J. Biomed. Mat. Res.* **1996**, *30*, 133–138.
88. Meikle, M.C.; Papaioannou, S.; Ratledge, T.J.; Speight, P.M.; Watt-Smith, S.R.; Hill, P.A.; Reynolds, J.J. Effect of Poly D-lactide-co-glycolide Implants and Xenogeneic Bone Matrix-derived Growth Factors on Calvarial Bone Repair in the Rabbit. *Biomaterials* **1994**, *15* (7), 513–521.
89. Peter, S.J.; Miller, M.J.; Yaszemski, M.J.; Mikos, A.G. Polymer Concepts in Tissue Engineering. *J. Biomed. Mat. Res.* **1998**, *43*, 422–427.
90. Kim, S.S.; Vacanti, J.P. The Current Status of Tissue Engineering As Potential Therapy. *Semin. Pediatr. Surg.* **1999**, *8* (3), 119–123.

BIOAVAILABILITY OF DRUGS AND BIOEQUIVALENCE

James T. Dalton

The Ohio State University, Columbus, Ohio, U.S.A.

Marvin C. Meyer

University of Tennessee, Memphis, Tennessee, U.S.A.

INTRODUCTION

The U.S. Food and Drug Administration (FDA) defines *bioavailability* as “the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action” (1–3). Because in practice it is rare that drug concentrations can be determined at the site of action (e.g., at a receptor site), bioavailability is more commonly defined as “the rate and extent that the active drug is absorbed from a dosage form and becomes available in the systemic circulation.” Usually bioavailability refers to the absorption of a drug from the gastrointestinal tract following oral administration of a dosage form. The dosage form may be of any type, including a solution, suspension, tablet, capsule, powder, or elixir. Bioavailability can also refer to the absorption of a drug from other routes of administration, such as intramuscular (IM) injection, transdermal patches, ointments and other topical preparations, and implants, which also require absorption prior to reaching the systemic circulation. As these routes of administration (e.g., oral, IM, and topical) deliver the drug to a site outside the vascular system, they are often referred to as routes of extravascular administration. The only route of drug administration that will always result in a bioavailability of 100% is an intravenous injection, in which the amount of drug reaching the systemic circulation is equal to the total administered dose.

The term *relative bioavailability* refers to a comparison of two or more dosage forms in terms of their relative rate and extent of absorption. If an intravenous injection is employed as the reference dose, one can determine the *absolute bioavailability* of the test dosage form. Two dosage forms that do not differ significantly in their rate and extent of absorption are termed *bioequivalent*.

In general, bioequivalence evaluations involve comparisons of dosage forms that are *pharmaceutical equivalents*. Such dosage forms are defined as “drug products that contain identical amounts of the identical active drug ingredient, i.e., the same salt or ester of the same therapeutic moiety, in identical dosage forms, but not

necessarily containing the same inactive ingredients, and that meet the identical compendial or other applicable standard of identity, strength, quality, and purity, including potency and, where applicable, content uniformity, disintegration times and/or dissolution rates” (1–3). Bioequivalence determinations may also be made for *pharmaceutical alternatives*, defined as “drug products that contain the identical therapeutic moiety, or its precursor, but not necessarily in the same amount or dosage form or as the same salt or ester. Each such drug product individually meets either the identical or its own respective compendial or other applicable standard of identity, strength, quality, and purity, including potency and, where applicable, content uniformity, disintegration times and/or dissolution rates.” In some instances, two pharmaceutical alternatives exhibit markedly different bioavailability, for example, a rapidly absorbed elixir versus a more slowly absorbed capsule. In other cases, two different dosage forms (e.g., a tablet and a capsule) may or may not exhibit very similar bioavailability.

FACTORS AFFECTING DRUG BIOAVAILABILITY

Extravascularly administered drugs must traverse several barriers to reach the systemic circulation and/or their site of action. Many studies illustrate that differences in manufacturing procedures, as well as the composition of the dosage form, can affect the bioavailability of a drug product. In addition, the bioavailability of a drug product can also be influenced by the physiology of the patient and other factors, such as the content of the gastrointestinal tract. The steps involved in the disposition of an orally administered solid dosage form and the experimental approaches that may be used to characterize them are summarized in Table 1. Factors affecting the bioavailability of a dosage form, discussed briefly here, are discussed in detail elsewhere in this encyclopedia.^a

^aSee *Absorption of Drugs*, page 8; *Biopharmaceutics*, page 156.

Table 1 Disposition and evaluation of orally administered solid dosage forms

Steps involved in disposition	Experimental approaches
Dosage form reaches stomach/intestine	Measurement of pH or scintigraphy
Dosage form disintegrates into small particles ^a	In vitro disintegration testing
Drug dissolves in gastrointestinal fluids	In vitro dissolution testing
Drug reaches gastrointestinal wall/membrane	
Drug is returned to gastrointestinal lumen by P-glycoprotein efflux pump	In vitro drug transport studies
Drug is metabolized by intestinal enzymes	In vitro drug metabolism studies
Drug is absorbed into hepatic circulation	In situ/in vivo hepatic perfusion studies
Drug reaches systemic circulation	Assay of drug in blood, plasma, or serum
Drug is excreted in urine	Measurement of drug excretion in urine
Drug is metabolized	Measurement of metabolite(s) in blood and/or urine
Drug and/or active metabolite reaches its site of action and causes a pharmacologic response	
Response is unrelated to desired therapeutic activity	Measurement of onset, duration, and intensity of pharmacologic response
Response related to desired clinical response	Determination of clinical efficacy in patients

^aSome dosage forms are designed to remain intact (e.g., certain controlled-release products).

A major factor determining the bioavailability of an orally administered drug product is the dissolution rate of the drug. A drug must be in solution in order to be absorbed from the gastrointestinal tract. Even if the drug product is administered as a solution, some dissolution process may be required in the event that the drug precipitates as a result of low solubility in the fluids of the gastrointestinal tract.

Drug Product Formulation

Most drugs are not taken as pure chemicals but are formulated into a pharmaceutical dosage form. Such drug products may be a relatively simple solution or a compressed tablet containing binders, fillers, lubricants, a coloring agent, and the like; or a controlled-release product. The following are some of the formulation and manufacturing variables that could influence the bioavailability of a drug product: 1) the properties of the drug (salt form, crystalline structure, formation of solvates, solubility); 2) the composition of the finished dosage form (presence or absence of excipients, special coatings); 3) manufacturing variables (tablet compression force, processing variables, particle size of drug or excipients, environmental conditions); and 4) rate and/or site of dissolution in the gastrointestinal tract.

Physiologic and Other Factors Affecting Bioavailability

The rate and extent of drug absorption can also be affected by a wide variety of factors related to the

characteristics of the subject/patient receiving the drug product. These factors are important to consider because they can contribute to intrasubject and intersubject variability during treatment, and, if not well controlled during the course of a bioavailability study, they can bias the results and confound interpretation of the data. Examples include: 1) contents of the gastrointestinal tract (fluid volume and pH, diet, presence or absence of food, bacterial activity, presence of other drugs); 2) rate of gastrointestinal tract transit (influenced by disease, physical activity, drugs, emotional status of subject, and composition of the gastrointestinal tract contents); 3) pre-systemic drug metabolism and/or degradation (influenced by local blood flow; condition of the gastrointestinal tract membranes; drug transport, metabolism or degradation in the gastrointestinal tract or during the first pass of the drug through the liver); 4) age, sex, race, disease, body size, time of day, and physical activity.

Other factors related to the subject, if not recognized or controlled, can also influence the assessment of drug bioavailability and product bioequivalence. For example, bioavailability studies typically involve the collection of blood and/or urine specimens to determine drug appearance in the systemic circulation. Thus, physiologic and pharmacokinetic perturbations that affect the concentration of drug measured in these fluids can potentially influence the results and interpretation of the study. Examples include changes that alter 1) the rate, extent, and/or route of metabolism (e.g., coadministered drugs that compete for or induce drug metabolizing

enzymes); 2) the rate and/or extent of drug elimination by the kidney (e.g., kidney disease and/or competition and changes in urine pH that affect renal transport mechanisms); 3) the degree of binding of the drug to plasma or tissue proteins (e.g., age-related changes in plasma binding proteins or protein binding displacements); and 4) distribution of drug into the erythrocytes. Genetic polymorphisms in the drug metabolizing enzymes of the liver may also contribute to large differences in the pharmacokinetics of a drug and the interpretation of bioavailability studies (4). A well-designed bioavailability study must either control or account for the influence of such variables.

CHARACTERISTICS OF DRUGS WITH THE GREATEST POTENTIAL FOR A BIOAVAILABILITY PROBLEM

The total number of marketed drug products known to exhibit a significant bioavailability problem is relatively small. Thus, one view is that bioavailability has been overemphasized and that for most drug products it is not a matter of concern. Another view is that those products that exhibit a bioavailability deficiency in a carefully controlled study provide ample evidence of the potential of a bioavailability problem for many drug products not yet studied.

With minor exceptions, the U.S. FDA has required that bioavailability and bioequivalence of a drug product be demonstrated through in vivo studies. However, a Biopharmaceutics Classification System (BCS) was recently proposed that divides drugs into classes based on their solubility, permeability, and in vitro dissolution rate (5). This classification system could be used to justify the waiver of the requirement for in vivo studies for “rapidly dissolving drug products containing active moieties/active ingredients that are highly soluble and highly permeable.” If adopted by the FDA, the bioavailability and bioequivalence of drug products meeting these requirements could be demonstrated using in vitro solubility, permeability, and dissolution studies. However, drugs that are poorly permeable, poorly soluble, and/or formulated in slowly dissolving dosage forms would be considered more likely to demonstrate a bioavailability problem and thus, would not be candidates for the waiver of in vivo bioavailability studies.

The U.S. FDA published a summary of evidence that may be employed to assess the importance of establishing the bioavailability of a given drug (3). The summary is as follow:

1. Data from clinical trials or bioequivalence studies indicate a bioequivalence problem.
2. The drug has a narrow therapeutic ratio, and the drug concentrations in the patient must be carefully adjusted.
3. A lack of bioequivalence could have serious medical consequences.
4. Physicochemical evidence indicates that:
 - a. the drug has low solubility in water and/or the dissolution rate of the dosage form is slow;
 - b. the particle size, crystalline structure, and other factors of the drug can affect the dissolution and bioavailability; and
 - c. the drug product contains a high ratio of excipients to active ingredients, or the product may require excipients to enhance absorption or contain excipients that inhibit absorption.
5. Pharmacokinetic evidence indicates that:
 - a. the absorption of the active drug is limited to a specific region of the gastrointestinal tract;
 - b. the extent of absorption is low;
 - c. there is rapid metabolism such that rapid dissolution and absorption are required for effectiveness;
 - d. the product required special formulations to stabilize the drug in the gastrointestinal fluids; and
 - e. the drug exhibits dose-dependent pharmacokinetics.

Drugs that meet one or more of the above criteria and have been shown to exhibit significant differences in the bioavailability of marketed dosage forms include: digoxin, quinidine, furosemide, nitrofurantoin, prednisone, chloramphenicol, theophylline, chlorpromazine, phenytoin, amitriptyline, and phenylbutazone.

EXPERIMENTAL DETERMINATION OF BIOAVAILABILITY

Types of Studies

Several methods can be used to determine the bioavailability or bioequivalence of a drug product. The vast majority of bioavailability studies involve the administration of the test dosage form to a group of healthy human subjects, followed by the collection and assay of the drug concentration in blood (plasma or serum) samples. The second most frequent method utilizes urinary excretion measurements. Occasionally, other types of biological

material, such as saliva, cerebrospinal fluid, bile, or feces are also collected. For a few drugs, for which assay methods are not available for the determination of drug concentrations in biological fluids, a pharmacologic response may be measured. Finally, some bioavailability assessments have been made on the basis of a determination of the therapeutic response of patients to a given dosage form. However, this type of study is usually restricted to drugs that are active at the site of administration (e.g., topical) but are not intended to be available in the systemic circulation. For approval by the U.S. FDA, pharmacokinetic, pharmacodynamic, clinical, and in vitro studies are recognized (in descending order of preference) as acceptable approaches to document the bioavailability or bioequivalence of a drug product.

Blood level studies

The primary basis for blood concentration studies is the assumption that two dosage forms that exhibit superimposable blood concentration time profiles in a group of subjects should result in identical therapeutic activity in patients. Figure 1 illustrates a theoretical blood concentration time curve after the oral administration of a drug product. The key parameters to note from this figure are the maximum blood concentration (C_{\max}), the time (T_{\max}) of occurrence of the maximum blood concentration, and the total area under the blood concentration time curve (AUC). The value of T_{\max} provides a means to assess the rate of absorption of the drug. The T_{\max} is independent of the amount of drug absorbed but is inversely related to the absorption rate. Thus, the faster the absorption of a drug, the shorter will be the T_{\max} . The value of T_{\max} is also influenced by the rate of elimination of the drug from the body. However, if one assumes elimination rate does not change during the period when two or more dosage forms are being tested in a given subject, then observed differences in T_{\max} will reflect absorption rate differences among the test products. The interpretation of C_{\max} is somewhat more complicated because it is a function of both the rate of absorption and the extent of absorption, as well as the elimination rate. Thus, as the amount of the drug absorbed increases and/or the rate of absorption increases, the C_{\max} also increases, assuming no change in elimination rate. A determination of the extent of drug absorption is usually based on a measure of AUC, which is directly proportional to the fraction of the administered dose that reaches the systemic circulation and is independent of the rate of absorption. The calculation of AUC is commonly accomplished using the linear trapezoidal rule (Fig. 1). The blood concentration time curve is divided into a series of geometric sections, and the

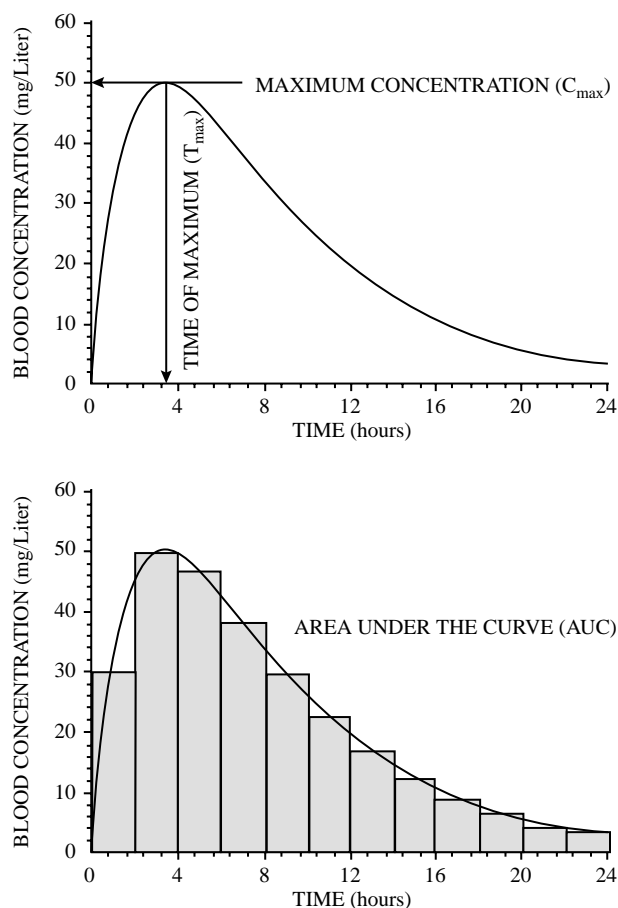


Fig. 1 Drug concentration in blood time plots illustrating calculation of C_{\max} , T_{\max} , and AUC.

area encompassed by each section is determined from the trapezoidal rule

$$\text{AUC} = 12 * (\Delta t) * (C_1 + C_2) \quad (1)$$

where Δt is the time interval between the collection of two blood samples of concentrations C_1 and C_2 . The units of AUC are given as the product of concentration and time (e.g., mg/Liter \times hr). If blood samples are not obtained for a sufficient period of time to result in a zero drug concentration in the final sample, it is necessary to estimate the portion of the AUC remaining after the final sample. Eq. 2 gives the relationship between the AUC ($0 - t$) for the portion of the curve to the last sample taken at time t , and the total AUC($0 - \infty$)

$$\text{AUC}(0 - \infty) = \text{AUC}(0 - t_{\text{last}}) + C_{\text{last}}/K \quad (2)$$

where C_{last} is the last measurable drug concentration, t_{last} is the time at which it was collected, and K is the apparent first-order elimination rate constant estimated

from the terminal slope of the log-linear plot of concentration versus time. In studies involving blood sampling intervals after the peak that are relatively long compared with the half-life of the drug, the use of the logarithmic trapezoidal method has been recommended to estimate the postpeak AUC (6).

The value of $AUC(0 - \infty)$ may also be as expressed in Eq. 3:

$$AUC(0 - \infty) = F * D / CL \quad (3)$$

where F is the fraction of dose absorbed, D is the administered dose, and CL is the total body clearance of the drug. Thus, a comparison may be made between two different dosage forms on the basis of the ratio of their respective $AUC(0 - \infty)$ values. Assuming equivalent doses are given and the clearance of the drug remains constant during the time the two doses are administered, the ratio of the AUCs will be directly proportional to the ratio of the fraction of each dose absorbed, that is, the extent of absorption.

Figure 2 illustrates the types of plasma concentration time data that might be obtained during the testing of three different bioinequivalent formulations of a drug. Products A and C are absorbed at the same rate, based on identical T_{max} values, but B is more slowly absorbed, as shown by the longer time required to achieve C_{max} . Products A and B appear to be absorbed to the same extent on the basis of very similar values for AUC. However, product C is obviously less completely absorbed, as shown by the lower AUC.

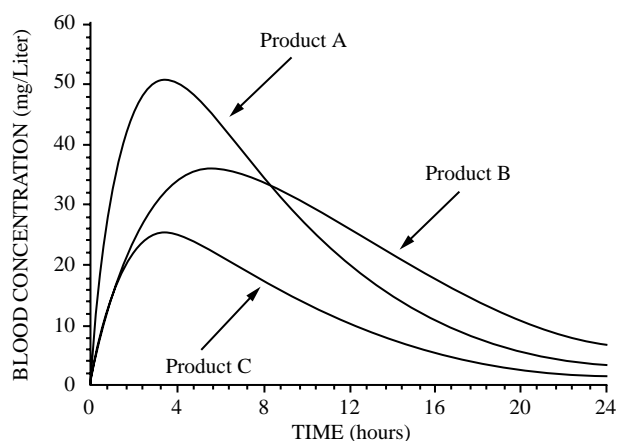


Fig. 2 Drug concentration in blood versus time after single doses of product A (rapidly and completely absorbed), product B (slowly and completely absorbed), and product C (rapidly absorbed, but only 50% as completely absorbed as products A and B).

All of the foregoing discussion assumes the absorption and elimination of the drug do not exhibit dose-dependent pharmacokinetics. For example, if the metabolism of a drug is a saturable process, the total AUC may be significantly increased for a rapidly absorbed dosage form because the initial blood drug concentrations exceed the metabolic capacity of the body to eliminate the drug. Nonlinear pharmacokinetics may also be caused by changes in the clearance of a drug due to factors such as nonlinear drug binding to plasma and/or tissue proteins. Thus, for a drug such as disopyramide, the AUC does not increase in proportion to the administered dose (7, 8). This is the result of a decrease in drug binding as drug concentration increases. Thus, it is necessary to measure free drug concentration in the plasma, rather than total drug concentration (free + bound), for disopyramide bioequivalence studies. Other mechanisms that can contribute to nonlinear pharmacokinetics and complicate the interpretation of bioequivalence data have been reviewed by Tozer and Tang-Liu (9). The interpretation of non-superimposable blood concentration time profiles for drugs that exhibit nonlinearity must be done with caution. Further, bioavailability and bioequivalence studies of such drugs may need to include multiple-dose protocols to further define the effect of differences in the rate and extent of absorption on the steady-state drug concentrations.

Urinary excretion studies

The estimation of bioavailability on the basis of the appearance of drugs in the urine is an attractive alternative to blood sampling because it represents a noninvasive method. This approach is particularly useful for drugs that have the urine as the site of activity (e.g., urinary tract antiseptics, such as nitrofurantoin and methenamine). This method is also useful for drugs that are extensively excreted unmetabolized in the urine, such as certain thiazide diuretics and sulfonamides. Often a less sensitive analytic method is required for urine concentrations compared with blood concentrations. If the urine concentrations are low, assaying larger sample volumes is relatively easy. The primary disadvantage to urinary excretion studies is that they require the collection of samples for a longer period of time to ensure the complete recovery of the absorbed drug. In addition, the subjects must be careful to completely void at each collection time and to avoid accidentally discarding any samples. As with the assessment of bioavailability from blood levels, urinary excretion studies also generally assume the pharmacokinetics are not dose dependent. If renal excretion is a saturable process, the percentage of the

drug excreted unmetabolized in the urine may not reflect the rate and extent of the drug absorption.

The three major parameters examined in urinary excretion bioavailability studies are: 1) the cumulative amount of drug excreted unmetabolized in the urine (ΣXu); 2) the maximum urinary excretion rate (ER_{\max}); and 3) the time of maximum excretion rate (T_{\max}). In simple pharmacokinetic models, the rate of appearance of drug in the urine is proportional to the concentration of drug in the systemic circulation. Thus, the values for T_{\max} and ER_{\max} for urine studies are analogous to the T_{\max} and C_{\max} values derived from blood level studies. The value of T_{\max} decreases as the absorption rate of the drug increases, and ER_{\max} increases as the rate and/or the extent of absorption increases. The value for ΣXu is related to the AUC and increases as the extent of absorption increases.

The calculation of excretion rate (ER) is based on Eq. 4

$$ER = \left(\sum Xu_2 - \sum Xu_1 \right) (t_2 - t_1) \quad (4)$$

where ΣXu_1 and ΣXu_2 represent the cumulative amount of drug recovered in the urine samples obtained at sampling times up to t_1 and t_2 , respectively. When constructing a plot of ER versus time, or for the determination of ER_{\max} , the values for time are taken to be the midpoint of the urine collection period, that is, the midpoint between t_1 and t_2 . Thus, estimates of T_{\max} and ER_{\max} from urinary excretion data provide less information on the rate of drug absorption than can be obtained from analysis of the blood concentration time profile, largely due to the fact that there is a limit to the frequency at which urine can be readily collected.

When sufficient urine samples have been collected to ensure that no significant amount of drug remains to be excreted, the cumulative urinary recovery is symbolized as $\Sigma Xu \infty$. The relative extent of absorption of drug 5 from two dosage forms may then be expressed as the ratio of the $\Sigma Xu \infty$ values. However, the value of $\Sigma Xu \infty$ is a function of the fraction (F) of administered dose (D) absorbed, the renal elimination rate constant (ke), and the rate constant for overall elimination (K) from the systemic circulation, as expressed in Eq. 5:

$$\sum Xu_{\infty} = F * D * ke / K \quad (5)$$

Assay of other biological material

For a few drugs such as theophylline, saliva drug concentrations have been employed to supplement the collection of blood samples. However, the intersubject and intrasubject variability in saliva/plasma ratios have generally precluded the sole use of saliva drug

concentrations to assess bioavailability. For some drugs such as cephalosporin antibiotics, clinical studies may also include a determination of the appearance of the drug in other body fluids, such as the cerebrospinal fluid and bile.

One might initially think that assessing the extent of drug absorption after oral administration would be possible by simply quantitating the amount of drug excreted in the feces. Such determinations occasionally provide useful data. For example, if subjects receive the drug as an enteric-coated tablet or some other solid dosage form and the product is recovered intact in the feces, there is little question regarding the lack of bioavailability. However, the data obtained from fecal recovery studies must be carefully interpreted. If the drug is analytically measured in a fecal sample, this does not establish that the drug was not absorbed. For example, certain drugs undergo extensive enterohepatic recycling and/or excretion in the saliva. Thus, a drug could be fully bioavailable, and, yet, a portion of the administered dose could be found in the feces. Further, the absence of intact drug in the feces is not proof of absorption because the drug may be degraded during its transit through the gastrointestinal tract.

Assessment of bioavailability from pharmacologic response

Topical application of a corticosteroid does not generally result in measurable blood concentrations of the drug. Thus, bioavailability and bioequivalence determinations for these drug products may involve measurement of dermatologic vasoconstriction (i.e., skin blanching), a pharmacodynamic response. A few studies have attempted to relate quantitatively a pharmacologic response to the oral bioavailability of a drug. For example, a relationship between the extent and duration of serum glucose concentration reduction and the bioavailability of two dosage forms of tolbutamide was demonstrated (10, 11). Others have employed pharmacologic end points that were not necessarily related to the therapeutic activity of the test drug. For example, attempts have been made to relate pharmacologic responses, such as changes in pupil diameter, electrocardiogram readings, or electroencephalogram readings, to the time course of a given drug in humans and animals. However, pharmacologic data tend to be more variable, and demonstrating a good correlation between the measured response and the amount of drug available from the dosage form may be difficult. Further, the potential exists that the measured response may be due to a metabolite whose concentration is not proportional to the concentration of the parent drug responsible for therapeutic activity.

Assessment of bioavailability from therapeutic response

Because the ultimate goal of drug therapy is to achieve some therapeutic response in a patient, ideally the assessment of drug product efficacy should be studied in patients requiring the drug. Nonetheless there are good reasons to utilize healthy volunteers rather than patients. First, the quantitation of patient clinical response is too imprecise to permit a reasonable estimation of the relative bioavailability of two dosage forms of the same drug. Second, bioequivalence studies are usually conducted using a crossover design in which each subject receives each of the test dosage forms, and it is assumed that the physiologic status of the subject does not change significantly over the duration of the study. If patients, however, were utilized, this could be an invalid assumption because of changes in a patient's disease state. Third, unless multiple-dose protocols were employed, a patient who required the drug for a disease would be able to receive only a single dose of the drug every few days or perhaps each week. To avoid such problems, one could test each product in different groups of patients, but this would require both the use of a large number of patients and careful matching of the various patient groups. Fourth, many patients receive more than one drug, and the results obtained from a bioavailability study could be compromised because of a drug-drug interaction. Finally, an ethical question would arise in the case in which a particular product was believed to be defective. Thus, a patient requiring treatment with a given drug would need to consent to receive a product that might not provide sufficient drug for adequate treatment. Because of these considerations, the general conclusion is that most bioequivalence studies should be carried out with healthy subjects. However, for drugs that are not designed to be absorbed into the systemic circulation and are active at the site of administration, clinical studies in patients are the only means to determine bioequivalence. Such studies are usually conducted using a parallel, rather than a crossover, design. Examples include studies of topical antifungal agents, drugs used in the treatment of acne, and agents such as sucralfate used in ulcer therapy.

Other experimental approaches

The current bioequivalence regulations of the U.S. FDA describe several types of experimental approaches that do not involve human testing (3).

Use of experimental animals: Animal studies are not acceptable for bioequivalence determinations unless the data obtained with animals have been correlated with the data obtained in human studies, ensuring that the

bioavailability of a dosage form in animals is closely related to that in humans. Animals are known to differ from humans in terms of gastrointestinal tract characteristics, metabolism, distribution, and excretion. For the study of solid dosage forms, relatively large animals, such as, dogs or monkeys, must be employed. Although such studies can provide useful data during dosage form development and may serve as an alternative to human subjects for drugs that are quite toxic to humans (e.g., cancer drugs), in general, animal studies are not acceptable as the final assessment of the bioavailability of a dosage form.

In vitro methods: In recent years, there has been great interest in the development of laboratory test systems that can simulate the disintegration and dissolution of a drug product in the human gastrointestinal tract. The development of such devices is to reduce the need for human testing. One of the early approaches to relate in vivo bioavailability data to in vitro measurements employed testing based on the time required for a solid dosage form to disintegrate in a particular solvent. The official apparatus employed for such testing is described in the U.S. Pharmacopoeia XXIV (USP) (12). However, the problem with this method is that the measurement of the time required for a dosage form to break into small particles may not necessarily relate to the dissolution rate of the drug. The current XXIV describes one official in vitro disintegration apparatus (i.e., basket-rack assembly) and two official dissolution apparatus (i.e., one with a paddle and one with a basket stirring element) for the evaluation of solid dosage forms (12). Although these methods are well-established and used extensively, few in vitro/in vivo correlations between dissolution data and human bioavailability data have been established. In vitro dissolution testing is useful as a standard for monitoring product quality and for distinguishing between dosage forms for which a bioavailability problem is known to exist. However, the USP acknowledges that many of the formulation factors that affect the performance of a drug product during in vitro dissolution testing may only sometimes affect the in vivo bioavailability of the drug (i.e., dissolution testing may identify subtle differences in the characteristics of the dosage form that are not relevant to its in vivo performance) (12). Thus, in vitro dissolution testing cannot be assumed to relate to the in vivo bioavailability of a given dosage form.

In conducting dissolution studies, the choice of an appropriate solvent is very important. Dissolution experiments should be conducted using conditions that mimic the environment in the gastrointestinal tract. Typical dissolution studies employ 0.1 *N* hydrochloric acid, water, or a buffer as the dissolution media. Simulated gastric

fluid (pH 1.2), with or without pepsin and simulated intestinal fluid (pH 6.8), with or without pancreatin, are also commonly employed for in vitro dissolution testing. It is widely recognized that the dissolution of controlled-release dosage forms may be pH-dependent. Thus, the US FDA recommends that dissolution testing for controlled-release dosage forms be conducted over a wide range of pH values, including 1–1.5, 4–4.5, 6–6.5, and 7–7.5, with multiple time determinations to better characterize the dissolution properties of the dosage form (1–3). Bioequivalence determinations for products containing cholestyramine resin (used to control cholesterol) represent a novel in vitro approach to bioequivalence testing. Generic versions of such drug products are evaluated in vitro by determining the rate and extent of the interaction of different bile salts with the resin.

Experimental Design

The proper design of a bioavailability or bioequivalence study is essential to the collection of meaningful data. Studies must include a sufficient number of subjects, and blood or urine samples must be collected at appropriately spaced intervals to characterize accurately the pharmacokinetics of the drug product(s) and make statistically relevant conclusions regarding its bioavailability and bioequivalence. Amongst other factors, study design must consider the characteristics of the study population (e.g., age, weight, gender, race, and health), the timing of dose administration, meals, and blood sample collection, and the time interval (i.e., washout period) between consecutive administrations of the drug products. In a bioequivalence study involving two or more dosage forms, the sequence of product administration must also be carefully considered to minimize experimental bias. The purpose of these rigorous controls on experimental design and conduct is to minimize the variability associated with pharmacokinetic (e.g., clearance, volume of distribution, and absorption) and physiologic (e.g., gastric emptying and pH) factors, such that the variability observed during the study is more closely related to the performance of the drug product(s) under consideration.

Crossover designs versus other designs

The most common type of study uses a crossover design in which each subject receives each of the test products. In such a design, differences among dosage forms, subjects, and sequences of administration can be readily estimated. In essence, each subject serves as his own control. Crossover designs have also been developed to minimize the effects of residual or carryover effects, which could occur if the administration of a given dosage form had an

Table 2 Three-way crossover design for bioequivalence study

Subject	Dosing period:		
	Period 1	Period 2	Period 3
1 and 2	A	B	C
3 and 4	A	C	B
5 and 6	B	A	C
7 and 8	B	C	A
9 and 10	C	A	B
11 and 12	C	B	A

influence on the bioavailability of a subsequently administered product. Table 2 illustrates a crossover design that could be employed to evaluate the relative bioavailability of three dosage forms (A, B, and C) in a group of 12 human subjects. Note that there are six possible sequences for the administration of each of the three products. Further, each subject receives each of the three products, and each dosing period contains all three products. The value of such a design is that it minimizes any bias relating to subject and dosing sequence effects. Replicate study designs, in which each subject receives the test and reference drug products on more than one occasion, are currently being evaluated as an alternative method to examine the bioequivalence of drug products.

Single-dose versus multiple-dose studies

Bioavailability studies intended to determine the disposition of a drug, particularly those involving new chemical entities, must include both single- and multiple-dose administration. However, most bioequivalence studies, which compare the bioavailability of two or more dosage forms, usually employ only single-dose administration for each product, under fasting conditions. One major exception is bioequivalence studies of controlled-release products. The U.S. FDA requires both single- and multiple-dose administration, as well as a determination of the effect of food on the absorption of the drug from the dosage form (1–3). However, the requirement of multiple dose studies in assessing the bioequivalence of controlled-release products while receiving much attention recently, may be abandoned, with the thought that single dose studies provide more sensitive information to assess the performance of these products.

Multiple-dose studies are more difficult to control, and they expose the subject to more drug. However, multiple-dose study designs also have advantages. They are more representative of how drug products are usually used by patients, and they also require fewer blood

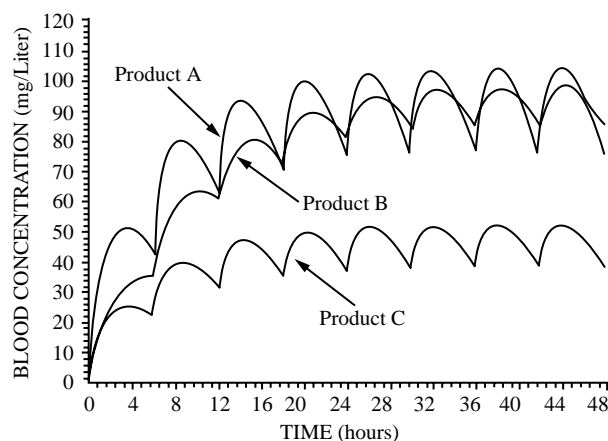


Fig. 3 Drug concentrations in blood vs. time after multiple doses of three products administered every 6h. The products (A, B, and C) are the same as those illustrated in Fig. 2.

samples and less sensitive analytical methods. Such studies require a sufficient number of doses to permit the achievement of steady state, which may be defined as the point where the amount of drug being absorbed into the body is equal to the amount being eliminated from the body. A general rule is that dosing must continue for approximately five biological half-lives in order to be within approximately 95% of steady state. Once steady state has been reached, the area under the blood concentration time curve during a single dosing interval should be equal to the value of the AUC ($0-\infty$) from a single dose (assuming dose-independent clearance of the drug). Thus, it is necessary to obtain blood samples over only a single dosing interval at steady state in order to determine the AUC. The dosing interval selected for sampling (e.g., 7 A.M. through 7 P.M. at steady state, if dosing occurs every 12 h) should be identical for each study phase. Because the disposition of the drug could vary as a function of time of day, comparing an AUC determined during the period 7 A.M. through 7 P.M. for one product and the AUC determined from 7 P.M. through 7 A.M. for a second product would not be valid.

An example of the results of a steady-state study, with dosing every 6 h, is illustrated in Fig. 3. The pharmacokinetic data employed to generate the results shown in Fig. 3 were identical to those used for Fig. 2. The results demonstrate the influence of the rate and extent of absorption on the steady-state plasma concentrations. The lower plasma concentrations shown for product C reflect the lower extent of absorption for this product. Products A and B have the same extent of absorption but differ in rate of absorption. Product A is more rapidly absorbed than product B, and, thus, there is a greater fluctuation between the maximum and minimum concentrations at steady state.

Other study design considerations

In the design or evaluation of the results of a bioavailability or bioequivalence study, it is important to establish that an adequate number of subjects were studied and that an adequate number of blood and/or urine samples were collected. As in most scientific studies, the use of too few subjects precludes reaching meaningful decisions regarding the significance of differences that may be observed. Moreover, values for T_{\max} , C_{\max} , or ER_{\max} cannot be accurately determined if the time interval between samples is too great. Further, accurate estimates of AUC and/or total urinary recovery of drug are not possible if blood and/or urine collections are terminated prematurely. As a general rule, collecting blood samples for at least two to three biological half-lives is desirable, and urinary excretion studies should include urine collections for five to seven biologic half-lives. Some extension of these guidelines may be necessary if the study involves controlled-release dosage forms that may result in absorption for a prolonged period. If data are not available for the half-life of a drug, a reasonable guideline is to continue blood sampling until blood concentrations have declined to less than 10% of the peak concentrations. Similarly, urine collections should be continued until significant quantities of drug are no longer being excreted. A typical bioavailability study utilizes from 12 to 24 or more healthy subjects who have no history of any disease that could affect the disposition of the test drug. The subjects are required to refrain from taking any drug other than the test compound for a period (usually one week or more) prior to the study and throughout the course of the study. Certain drugs, such as, enzyme inducers, could affect the disposition of the test drug. Further, depending on the specificity of the assay, other drugs may interfere in the quantitation of the drug of interest. Usually, only subjects between the ages of 18 and 40 are employed, unless there are specific reasons for employing other subject populations. Both male and female subjects should be included in the study, as well as subjects of differing race, realizing, of course, that the heterogeneity that one may accommodate in a study of 12 subjects is limited if statistically meaningful comparisons are to be made between the groups. Finally, for typical single-dose studies, the subjects are required to fast from approximately 12 h before the dose until 4 h after the dose, unless the objective of the study is to determine the influence of food on absorption of the drug.

Analytical methodology

One of the most important considerations in any bioavailability study is the validity of the analytical

method. In addition to being reproducible, it must be sufficiently sensitive to permit the detection of low drug concentrations in the biologic sample. After single-dose administration, drug concentrations in plasma are frequently in the low microgram or nanogram per milliliter range. Further, the method must be specific for unmetabolized drug and be capable of accurately determining drug concentrations in the presence of metabolites of the drug and the constituents of blood and/or urine. Most analytical methods involve some type of clean-up step, such as, solvent extraction, to separate the drug from the biological fluid. In addition, most assays employ some type of chromatography, often using either gas chromatography or high-performance liquid chromatography. Assay validation is a critical component of any bioavailability or bioequivalence study and should consider the accuracy, precision, sensitivity, specificity, linearity, and reproducibility of the analytical method. In the biologic sample, the stability of the drug during storage must also be given careful consideration.

Analysis and Interpretation of Data

Bioequivalence studies are usually intended to demonstrate that two or more formulations do not differ significantly (i.e., that the products are therapeutically equivalent and interchangeable). Once the data from a bioequivalence study are collected, statistical methods must be applied to determine the level of significance of any observed differences. Statistical comparisons of the C_{\max} and AUC are performed after log transformation. Log transformation is appropriate because 1) many biological and pharmacokinetic parameters are log-normally distributed, and 2) it is the ratio of C_{\max} and AUC between drug products, and not the absolute difference between the mean values, that is most relevant for comparison. Also, one-sided statistical tests at a 0.05 level of significance are performed using the log-transformed data from the bioequivalence study. It is important to note that 90% confidence intervals (CIs), and not the mean values, of C_{\max} and AUC are employed to make this comparison and assure the bioequivalence of the drug products. In other words, the 90% CI for the mean C_{\max} and mean AUC for the drug product must be between 80 and 125% of the respective mean values for the reference dosage form (e.g., a drug product with a mean C_{\max} of 83% of the reference product with a 90% CI of 79–87% would not be considered bioequivalent). In fact, if the mean C_{\max} or mean AUC of the drug product differs by more than 10–15%, the CI is likely to fall outside the 80–125%

range. Typically, an analysis of variance (ANOVA) method appropriate for the study design is applied to evaluate differences among dosage forms, subjects, and treatment periods. Mean T_{\max} values are also computed, but unless large differences are observed, they are generally not used for the bioequivalence determination, largely due to the fact that T_{\max} values are highly dependent on study design and the time at which blood samples were collected.

SUMMARY

The measurement of the bioavailability and bioequivalence of drug dosage forms is commonplace. The numerous reasons for the increased use of such studies include: 1) rapid growth of the generic drug industry; 2) an increased awareness of the effect of dosage form on the rate and extent of drug absorption; 3) the development of more sensitive and reliable analytic methods to quantitate drugs and metabolites in the body; and 4) the need to have a means to evaluate the vivo performance of a dosage form by a means other than clinical trials in patients. Clearly, bioavailability and bioequivalence studies have become a routine part of the drug product approval process by governmental agencies. Such studies provide valuable information regarding the disposition of a drug in humans and the factors that may influence its performance. These data are useful in the development of new dosage forms and assist in the preparation of appropriate labeling for a drug product. Finally, bioequivalence studies have become an essential part of the comparison of pharmaceutically equivalent products manufactured by different pharmaceutical firms.

REFERENCES

1. United States Food and Drug Administration, Center for Drug Evaluation and Research, <http://www.fda.gov/cder/regguide.htm/> (accessed 12/2/99).
2. National Archives and Records Administration Code of Federal Regulations, Title 21, Food and Drugs, Food and Drug Administration, Department of Health and Human Services, Part 320, Bioavailability and Bioequivalence Requirements, <http://www.access.gpo.gov/cgi-bin/cfrassemble.cgi?title=199921> (accessed 12/2/99).
3. Federal Register, January 7, 1977; Part III, Department of Health, Education and Welfare, U.S. Food and Drug Administration, Washington, DC, 1624–1653.
4. Kidd, R.; Blaisdell, J.; Straughn, A.; Meyer, M.C.; Goldstein, J.; Dalton, J.T. Pharmacokinetics of Phenytoin,

- Glipizide and Nifedipine in a Homozygous CYP2C9-Leu 359 Individual. *Pharmacogenetics* **1999**, 9 (1), 71–80.
5. Amidon, G.L.; Linnerhas, H.; Shah, V.P.; Crison, J.R. A Theoretical Basis for a Biopharmaceutic Drug Classification: The Correlation of In Vitro Drug Product Dissolution and In Vivo Bioavailability. *Pharm. Res.* **1995**, 12 (3), 413–420.
6. Yeh, K.C.; Kwan, K.C. A Comparison of Numerical Integrating Algorithms by Trapezoidal, Lagrange, and Spline Approximation. *J. Pharmacokinet. Biopharm.* **1978**, 6 (1), 79–98.
7. Haughey, D.B.; Lima, J.J. Influence of Concentration-dependent Protein Binding on Serum Concentrations and Urinary Excretion of Disopyramide and its Metabolite Following Oral Administration. *Biopharm. Drug Dispos.* **1983**, 4 (2), 103–112.
8. Upton, R.A.; Williams, R.L. The Impact Of Neglecting Nonlinear Plasma-protein Binding on Disopyramide Bioavailability Studies. *J. Pharmacokinet. Biopharm.* **1986**, 14 (4), 365–379.
9. Tozer, T.N.; Tang-Liu, D.D.S. Linear versus Nonlinear Kinetics. *Topics in Pharmaceutical Sciences*; Breimer, D.D., Speiser, P., Eds.; Elsevier: New York, 1981; 3–17.
10. Varley, A.B. The Generic Inequivalence Of Drugs. *J. Am. Med. Assoc.* **1968**, 206 (8), 1745–1748.
11. Olson, S.C.; Ayres, J.W.; Antal, E.J.; Albert, K.S. Effect of Food and Tablet Age on Relative Bioavailability and Pharmacodynamics of Two Tolbutamide Products. *J. Pharm. Sci.* **1985**, 74 (7), 735–740.
12. Committee of Revision. Monograph 701, Disintegration; Monograph 711, Dissolution. *United States Pharmacopeia XXIV/National Formulary XIX*; United States Pharmacopeial Convention, Inc.: Rockville, MD, 2000; 1941–1943.
13. Welling, P.G.; Tse, F.L.S.; Dighe, S.V. Pharmaceutical Bioequivalence. *Drugs and the Pharmaceutical Sciences*; Marcel Dekker, Inc.: New York, 1991; 48, 1–467.
14. Abdou, H.M. Dissolution, Bioavailability & Bioequivalence; Mack Publishing: Easton, PA, 1989; 1–554.

BIOTRANSFORMATION OF DRUGS

L.F. Chasseaud

D.R. Hawkins

Huntingdon Life Sciences, Huntingdon, United Kingdom

INTRODUCTION

The term *biotransformation of drugs* can be defined as the chemical conversion of drugs to other compounds in the body. However, this definition excludes degradation due to any inherent chemical instability of drugs in biological media. These conversions are usually mediated by enzymes in the body's organs, tissues, or biofluids, but occasionally they may occur by nonenzymic reactions and even by a combination of both enzymic and nonenzymic processes. Synonymous with biotransformation in this context is the word *metabolism*, and indeed the terms *drug biotransformation* and *drug metabolism* are used to describe the same events and are often interchanged, seemingly somewhat randomly. If anything, the latter is the more popular term.

Compared to the established scientific disciplines, the study of the biotransformation of drugs constitutes a relatively new subject area: the earliest scientific papers are little more than a century old. Regarded by many as a branch of pharmacology, drug biotransformation more properly belongs to biochemistry. However, knowledge of the biotransformation of drugs often contributes, in no small way, to a better understanding of pharmacologic, toxicologic, and clinical phenomena, and papers on drug biotransformation are to be found in journals specializing in these other subjects as well as those of biochemistry and chemical analysis. In the last three decades, exponents of drug biotransformation have created their own journals and review series.

A leading pioneer in drug biotransformation research was the late R.T. Williams, who wrote *Detoxication Mechanisms*, the first major text on the subject published over 40 years ago, and a classic to this day. Subsequently, the literature on drug biotransformation grew steadily until recent years, when it seems to have expanded almost exponentially. The major driving forces for this proliferation include the realization of the importance of drug biotransformation knowledge in promoting a better understanding of certain pharmacologic, toxicologic, and/or clinical findings; governmental legislation requiring drug biotransformation studies as part of new chemical entity safety evaluation programs; major advances in the

necessary scientific instrumentation; and the commercial availability of isotopically labeled compounds whose use can greatly facilitate drug biotransformation studies.

Because this article deals mainly with concepts, citation of specific scientific references is deliberately avoided. The reader is referred, instead, to the monographs referenced in the bibliography for a lead into more detailed or earlier scientific literature.

NEED FOR DRUG BIOTRANSFORMATION

During their lives, humans take a wide, almost infinite variety of substances into their bodies, some of which are necessary for their well-being and provide the raw materials essential for the processes of intermediary metabolism. However, other substances which the body does not require are also ingested, inhaled, or absorbed; these substances have been described as *exogenous compounds*, *foreign compounds*, or *xenobiotics* (Greek *xenos* = foreign) and can be either naturally occurring or man-made. The physicochemical properties (e.g., lipid solubility) that enable these compounds to be absorbed readily into the body would tend to preclude their facile excretion. Consequently, they would accumulate in the body and eventually cause deleterious effects. As a counter, metabolic processes have been adapted during evolution to deal with foreign compounds by converting (biotransforming) them into derivatives of great polarity (i.e., water solubility) that are consequently more easily excreted. Drugs are one type of foreign compound that the body usually has to biotransform in order to effect their removal. The action of many drugs is curtailed by biotransformation. With some drugs, however, biotransformation results in metabolites that share the pharmacologic action of the parent drug or may even exhibit other properties. The polynitrated organic nitrate vasodilators, such as isosorbide dinitrate, have metabolites that, as long as they contain nitrate groups, can exert the antianginal action of the parent drug, albeit at lower potency. Metabolites of benzodiazepines, such as diazepam, also possess its anxiolytic properties, and one of its metabolites, temazepam, is used

as a hypnotic. Yet others (prodrugs) rely on biotransformation to generate the pharmacologically active agent. The depot antipsychotics are of this type; haloperidol decanoate relies on ester hydrolysis in the body to release the active drug, haloperidol.

PATHWAYS OF BIOTRANSFORMATION

Biotransformation pathways have been classified under the headings of Phase I and Phase II reactions. The former includes processes involving oxidation, reduction, and hydrolysis that result in the formation of new functional groups in the drug molecule. Many of the metabolites formed are then able to participate in a subsequent biotransformation process (Phase II) in which a polar hydrophilic moiety such as glucuronic acid, sulfate, or an amino acid is covalently linked (conjugated) to the functional group introduced in the Phase I reaction. The usual overall result of these processes is the production of metabolites with progressively greater water solubility that facilitates their removal from the body. Biotransformations are generally catalyzed by enzymes present in the liver and other tissues. Some researchers consider that the biotransformation by the gut flora of drug metabolites excreted in the bile represents a third phase.

Many drugs undergo a series of biotransformations, and the combinations of consecutive Phase I and Phase II reactions can produce a complex array of metabolites (Fig. 1). Drugs already containing appropriate functional

groups can undergo Phase II reactions without the need for Phase I biotransformation; examples are the arylacetic and arylpropionic acid anti-inflammatory drugs (e.g., indomethacin and ibuprofen respectively), in which the carboxylic acid function can conjugate with glucuronic acid or amino acids such as glycine.

In some cases, Phase I metabolites may not be detected, owing to their instability or high chemical reactivity. The latter type are often electrophilic substances, called *reactive intermediates*, which frequently react nonenzymically as well as enzymically with conjugating nucleophiles to produce a Phase II metabolite. A common example of this type is the oxidative biotransformation of an aromatic ring and conjugation of the resulting arene oxide (epoxide) with the tripeptide glutathione. Detection of metabolites derived from this pathway often points to the formation of precursor reactive electrophilic Phase I metabolites, whose existence is nonetheless only inferred.

Certain biotransformation processes are reversible, and formation of an inactive metabolite that can be converted back to the active drug delays the removal of the drug from the body and probably prolongs the duration of exposure of the target tissues to the drug. The common processes that can contribute to this phenomenon are oxidation/reduction of secondary alcohols/ketones, sulfides/sulfoxides, and tertiary amines/N-oxides, all of which are reversible processes.

The pathways of biotransformation that a given drug follows are determined by many factors, including lipophilicity, molecular size and shape, steric and electronic characteristics, stereochemistry, acidity, and basicity. The interrelationships between these factors are not well understood, as judged by the present inability to predict biotransformations accurately except perhaps within a narrow group of closely related compounds. This inability is primarily due to the limited depth of available background information and insufficient knowledge of the factors affecting the interaction of compounds with the active sites of enzymes responsible for biotransformation. The knowledge of biotransformation processes is continually increasing, and new pathways are still being discovered.

OXIDATION

Oxidation is the most commonly encountered biotransformation process, and few drugs undergo a Phase I reaction that does not involve it. Oxidation occurs most frequently at saturated (sp^3) or unsaturated (sp^2) carbon

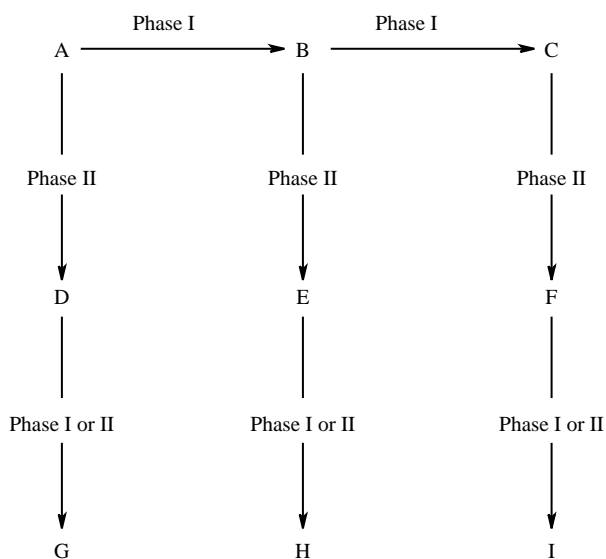


Fig. 1 Biotransformation of drug A by Phase I and Phase II reactions to several types of metabolite (B through I).

and with different valency states of nitrogen and sulfur present in either acyclic or cyclic structural environments.

Oxidation tends to decrease lipophilicity by the introduction of hydrophilic functional groups forming metabolites that may be more readily excreted. In addition, many of these functionalized metabolites are also substrates for Phase II conjugation reactions.

The nonspecific oxidation of drugs comprising many different chemical classes is catalyzed by a multienzyme mixed-function oxygenase system involving cytochrome P450 as the terminal oxidase. It is located in the endoplasmic reticulum of cells. Cytochrome P450 is a multigenic family of heme proteins, that can be distinguished from each other by their chromatographic and immunologic behavior, as well as by substrate specificity. The mechanisms of cytochrome P450 catalysis has not been established unequivocally, but a generally accepted scheme is shown in Fig. 2. The cytochromes P450 are the most versatile and extensively studied of the enzyme systems responsible for drug biotransformation.

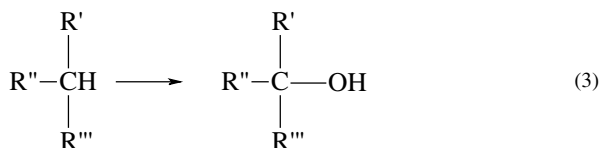
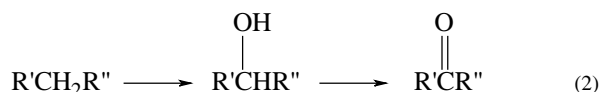
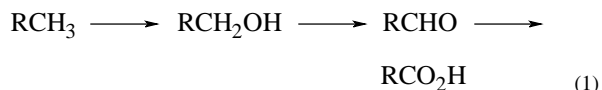
After binding of the substrate to oxidized cytochrome P450, a one-electron transfer from NADPH is catalyzed by a flavoprotein reductase. Molecular oxygen combines with the reduced enzyme-substrate complex to form a ternary complex that accepts a second electron, possibly derived from NADH via cytochrome b_5 . One oxygen atom is transferred from the activated oxygen to the substrate to

form the oxidized product. Water and oxidized cytochrome P450 also result.

Other mono-oxygenases are not cytochrome P450 dependent, such as flavoproteins located in the endoplasmic reticulum that are involved in the oxidation of tertiary amines to N-oxides and of various sulfur compounds. Yet other oxidative enzymes, including alcohol and aldehyde dehydrogenases and monoamine oxidases, are located in the mitochondria or cytosol.

Aliphatic Carbon

Oxidation of primary (Eq. 1), secondary (Eq. 2), or tertiary (Eq. 3) carbon atoms all occur to give the corresponding alcohols. Further oxidation of a primary alcohol yields the aldehyde, which is usually rapidly converted to the carboxylic acid. The oxidation of secondary alcohols to ketones generally leads to less hydrophilic metabolites and is less common.



Metabolites derived by loss of an alkyl or arylalkyl group from ethers (Eq. 4), thioethers (Eq. 5), amines (Eq. 6), and amides (Eq. 7) represent common biotransformation pathways ($\text{R}', \text{R}'' = \text{H}$, alkyl or aryl). These processes involve oxidation on carbon adjacent to the heteroatom. The intermediates are generally unstable and readily decompose to the corresponding alcohol, thiol, amine, or amide and an aldehyde. Intermediates formed from amides (Eq. 7) are more stable and may be detected as excreted metabolites. If a secondary carbon atom is adjacent to the heteroatom, then this portion of the molecule is released as a ketone. The heteroatom may also be located in a cyclic structure (e.g., morpholine, piperazine). Two processes have been adopted for amines, namely, N-dealkylation or deamination, that are essentially the same event. In general, which of the two terms applies depends on the side of the C—N bond

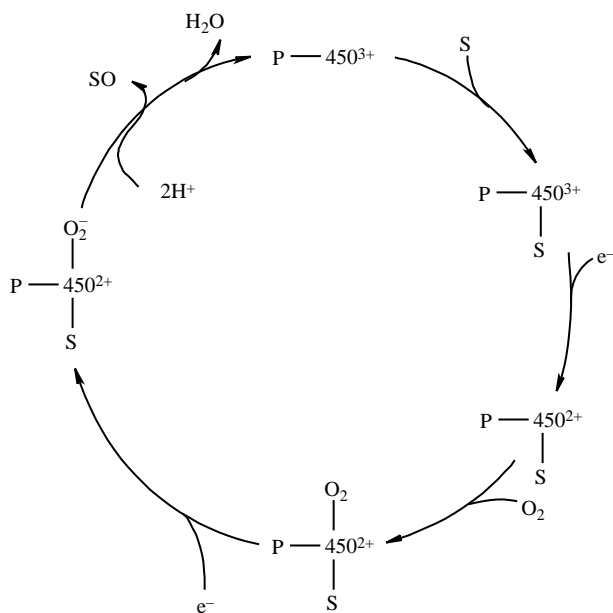
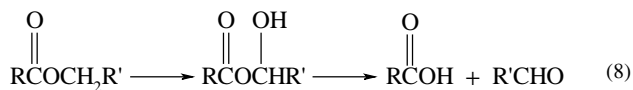
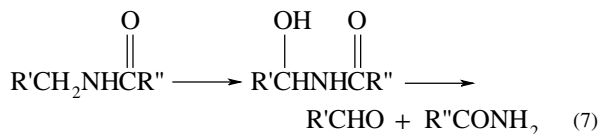
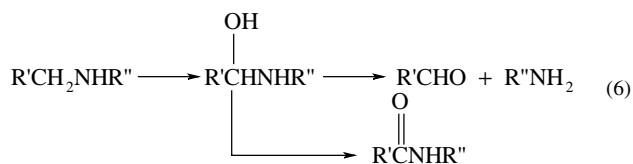
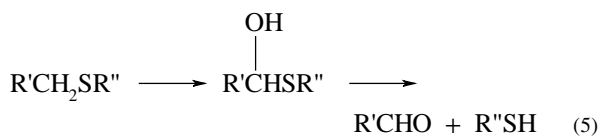
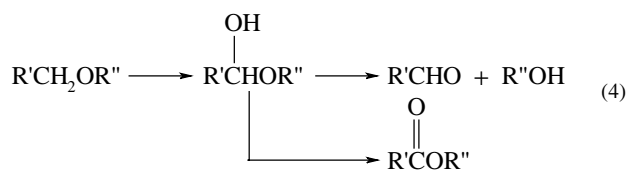


Fig. 2 Postulated mechanism of cytochrome P450 catalysis. Net reaction: $\text{Substrate} + 2\text{e}^- + 2\text{H}^+ + \text{O}_2 \rightarrow \text{substrate} - \text{O} + \text{H}_2\text{O}$.

that is of greatest concern, that is, whether interest is focused on the formation of the carbonyl compound or the amine. For a primary amine, therefore, the process is termed *deamination* because the fate of the ammonia generated is of no interest. For amines, the intermediate hydroxylated products may be oxidized further to a carbonyl function, resulting in formation of amides (Eq. 6). Cleavage of esters can also occur by an oxidative process of this type (Eq. 8). The only difference compared to hydrolysis is that instead of forming an alcohol, this part of the molecule is released as an aldehyde.



Aromatic and Unsaturated Hydrocarbons

Double bonds in unsaturated hydrocarbons are oxidized by epoxidation. The resultant epoxides are generally reactive substrates that are biotransformed further to alcohols or glycols through the action of epoxide hydrolases.

Hydroxylation of aromatic compounds to phenols has long been considered an important detoxication process, particularly as phenols can conjugate readily with glucuronic acid or sulfate to give highly water-soluble

metabolites, that are excreted. The oxidation process can involve formation of arene oxide intermediates in which an aromatic double bond has undergone epoxidation. These epoxides are generally unstable and considered to be reactive intermediates. Phenols, nominally the major product of aromatic hydroxylation, do not necessarily require formation of intermediate epoxides. Epoxides can give rise to several different types of metabolites by spontaneous isomerization to phenols, enzymic hydration to dihydrodiols and conjugation with glutathione (GSH) (Fig. 3). The pattern of metabolites depends in part on the intrinsic reactivity of the epoxide. Epoxides can also react with a variety of other biologic nucleophilic centres in cellular macromolecules, such as DNA, RNA, and proteins; this may ultimately produce toxic effects. Other aromatic ring systems are also hydroxylated via postulated epoxide intermediates, the best known being furans. The site of oxidation in an aromatic ring is dependent on the type of cytochrome P450 involved and on electronic and steric factors associated with the substrate structure. For substituted aromatic compounds, each epoxide can give rise to two isomeric phenols. Thus, methsuximide forms both a dihydrodiol and two isomeric phenols in animals and humans.

The dihydrodiols may be reduced to catechols through the action of dihydrodioldehydrogenase. Catechols can also be formed by two sequential hydroxylations. These catechols can be monomethylated by catechol

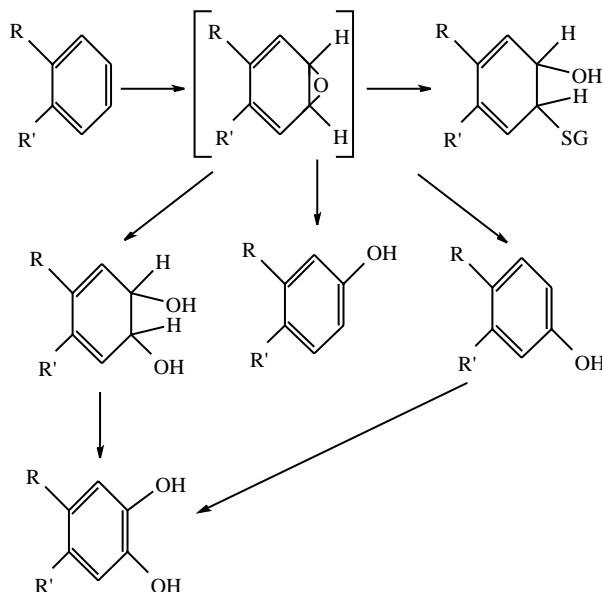


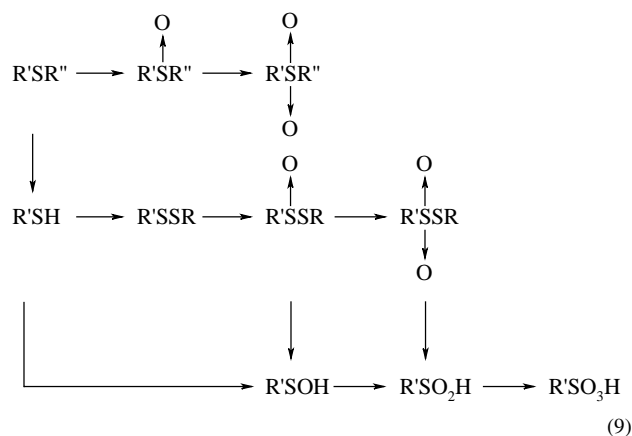
Fig. 3 Biotransformation of an aromatic hydrocarbon to an epoxide that undergoes various other reactions.

O-methyltransferase, and are often the terminal metabolites formed from this pathway.

A rather novel biotransformation that can occur as part of an aromatic hydroxylation process has been called the NIH shift whereby a substituent such as bromo, chloro, and fluoro at the position of hydroxylation undergoes migration to the adjacent position. Relevant examples are relatively few but this may be due to the lack of rigorous identification of ring-hydroxylated metabolites.

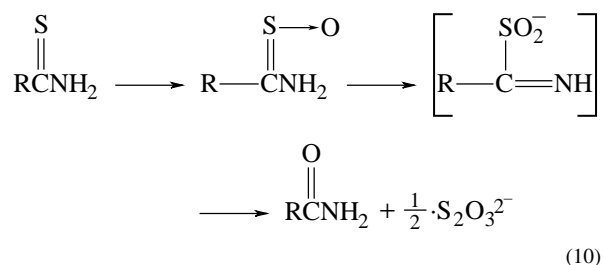
Sulfur

Oxidation of divalent sulfur atoms in thioethers is a common biotransformation of sulfur-containing compounds (Eq. 9). Oxidation proceeds in two stages, first to the sulfoxide and then to the sulfone. Sulfoxides have increased polarity and are often observed as excreted metabolites, but they can also be reduced back to the sulfide. Formation of the sulfoxide creates an asymmetric center, and stereospecific oxidation can occur. Sulfones tend to be terminal metabolites; no evidence for their reduction exists.



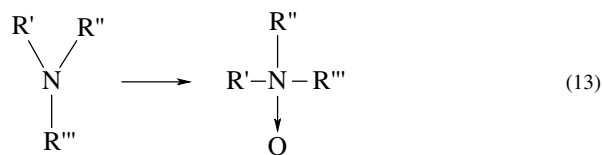
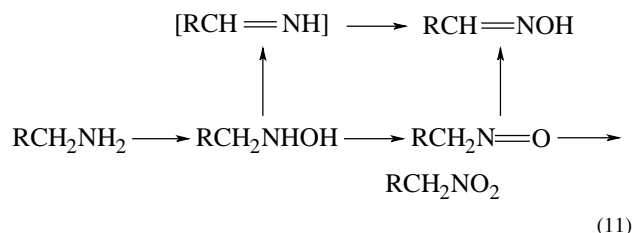
Thioethers can be cleaved to thiols that may be converted to stable oxygenated compounds, such as sulfinic and sulfonic acids, that are highly water-soluble and readily excreted. Thiols may be oxidized first to disulfides, in which further oxidation of sulfur then occurs before cleavage to sulfinic and sulfonic acids (Eq. 9).

The thiocarbonyl group in compounds such as thioamides is biotransformed by oxidative attack on sulfur. Sulfoxidation of the thioamide occurs first and is followed by further S-oxidation to give an unstable intermediate that is readily hydrolyzed to the amide and thiosulfate (Eq. 10).



Nitrogen

N-oxidation can occur in a number of ways to give either hydroxylamines from primary and secondary amines (Eqs. 11 and 12), hydroxamic acids from amides, or N-oxides from tertiary amines (Eq. 13). The enzyme systems involved are either cytochrome P450 or a flavoprotein oxygenase. Hydroxylamines may be further oxidized to a nitro compound via a nitroso intermediate (Eq. 11). Oximes can be formed by rearrangement of the nitroso intermediate or N-hydroxylation of an imine, that could in turn be derived by dehydration of a hydroxylamine (Eq. 11). N-Oxides may be formed from both tertiary arylamines and alkylamines and from nitrogen in heterocyclic aromatic systems, such as a pyridine ring.



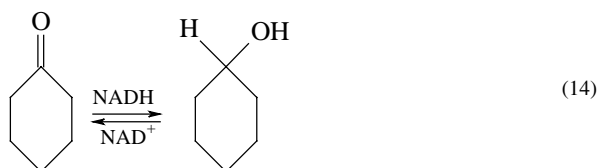
REDUCTION

Reductive biotransformation processes are less common than those involving oxidation, but a number of functional groups are susceptible to reduction. Some may occur non-enzymically through the action of biologic reducing agents. Certain processes are part of a reversible oxidoreductase

system, and the direction in which the equilibrium lies depends on the properties of the oxidized and reduced products. Reductive processes are favored by anaerobic conditions, and besides the action of reductases in tissues such as liver, kidney, and lung, micro-organisms in the anaerobic environment of the intestinal tract may produce reduced metabolites that are subsequently absorbed. Because reduction may often lead to metabolites of equivalent or greater lipophilicity than the precursor, it is seldom the only process occurring before a metabolite is excreted.

Carbon

Numerous ketones serve as substrates for alcohol dehydrogenases. These include alicyclic and arylalkyl ketones (Eq. 14).



Unlike aldehydes, ketones are essentially unreactive, apart from being reduced when the equilibrium shifts predominantly to the secondary alcohol that is readily conjugated, thereby encouraging further reduction of the ketone (Eq. 14). If ketone reduction results in the formation of an asymmetric carbon, a high degree of stereospecificity often occurs, even though the enzyme systems demonstrate a low degree of substrate specificity.

Alcohol dehydrogenase catalyzes the oxidation of primary alcohols to aldehydes, as well as the reverse reaction. However, due to the facile oxidation of aldehydes to acids, the reaction equilibrium is driven to the right (Eq. 15).



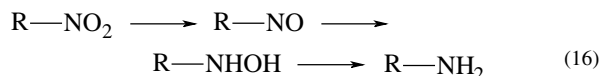
Thus administration of an aldehyde invariably results in the appearance of the corresponding acid as an excreted metabolite, although in some instances both the acid and alcohol are formed. Thus the aldehyde formed by deamination of the antihistamine chlorpheniramine is metabolized to a mixture of the corresponding acid and alcohol.

Reduction of carbon-carbon double bonds that may be instigated by gut flora has been observed.

Nitrogen

The types of nitrogen-containing compounds that are most frequently involved in reductive biotransformation are

those containing nitro, azo, and N-oxide functional groups. Similar enzymes are involved that are generally located in the endoplasmic reticulum or cytosol of the liver or in the intestinal microflora. Complete reduction of a nitro compound to the primary amine involves a six-electron transfer and proceeds through nitroso and hydroxylamine intermediates (Eq. 16).



Additional free-radical intermediates, such as the nitro anion radical and hydronitroxide radical, involving one-electron transfers have also been proposed. The benzodiazepine nitrazepam is reduced by liver microsomes to the corresponding amine. However, reductions involving NADPH-cytochrome c reductase may form only the hydroxylamine. A high level of nitroreductase activity is associated with anaerobic bacteria in the gut, that are capable of reducing aromatic nitro compounds to the corresponding amines. When the nitro group is located adjacent to a heteroatom in a heterocyclic ring, reduction to the nitroso or hydroxylamine intermediates often leads to cleavage of the ring. The hydroxylamine function in hydroxamic acids can also be reduced to the corresponding amides.

Aliphatic and aromatic tertiary amine N-oxides can be reduced to the corresponding amines. Reduction can occur either in liver cytosol, endoplasmic reticulum, or mitochondria, in extrahepatic tissues such as the kidney and lung, or by the gut flora.

Azoreductase activity is present in both the endoplasmic reticulum and the cytosol of the liver. It is thought to produce the hydrazo intermediate that is cleaved to give two primary amines. The best known example is that of the azo dye Prontosil that is cleaved to form sulfanilamide, an active antibacterial compound, a reaction that involves a four-electron transfer.

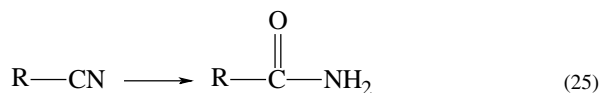
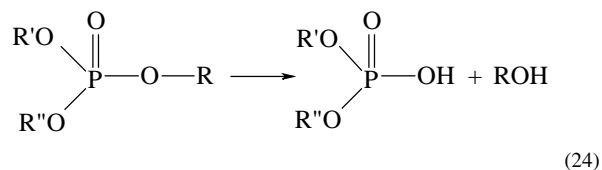
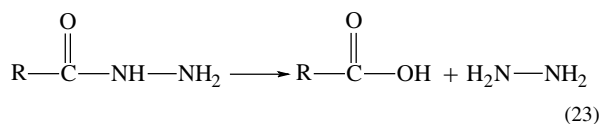
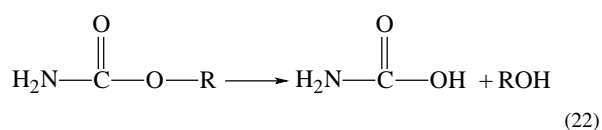
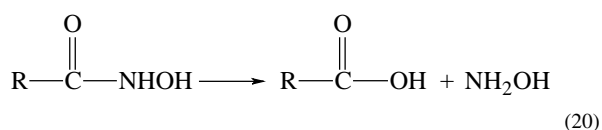
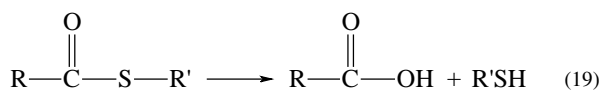
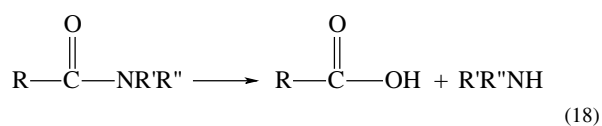
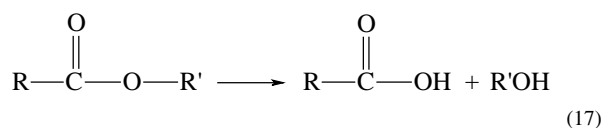
Certain halogenated hydrocarbons such as halothane, undergo reductive dehalogenation.

Sulfur

Most functional groups bearing sulfur, other than thiols, are primarily metabolized by oxidative processes. However, sulfoxides can be reduced to sulfides. The anti-inflammatory drug sulindac is reduced to the sulfide, which is an active metabolite, and it can be oxidized back to the sulfoxide. Disulfides are reduced to the corresponding sulfides.

HYDROLYSIS

Other important Phase I drug biotransformation reactions are those involving hydrolytic processes. The types of groups hydrolyzed include esters (Eq. 17), amides (Eq. 18), thioesters (Eq. 19), hydroxamates (Eq. 20), oximes (Eq. 21), carbamates (Eq. 22), hydrazides (Eq. 23), organophosphates (Eq. 24), and perhaps even nitriles (Eq. 25).

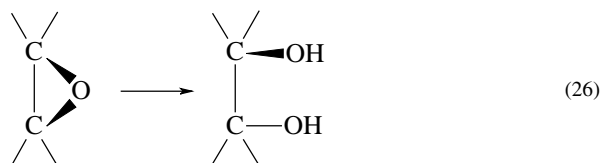


The hydrolytic reactions are mostly catalyzed by carboxylesterases although other names may be used for the enzymes involved depending on the substrates concerned, for example, amidases (Eq. 18), or phosphatases (Eq. 24).

With the increasing number of peptides being synthesized as potential drugs, mention should be made of peptidases that cleave (by hydrolysis) the amide (peptide) linkage between adjacent amino acids.

The carboxylesterases are a family of somewhat nonspecific enzymes that are not easily classified. They have an endogenous role in, for example, transport. B-esterases (serine hydrolases) are those sensitive to inhibition by organophosphates such as paraoxon that esterify a serine residue at the enzymes' active site; they are present in most mammalian tissues and biofluids such as the plasma. Other esterases are A-esterases (arylesterases) and the C-esterases (acetyl esterases). Esterase activity is particularly high in the endoplasmic reticulum of the liver. Some esterase activity appears to be tissue and species specific. The latter can lead to marked species differences in ester half-lives in the body, ranging, for example, from only a few minutes in the rat to several hours in the dog. Because both these species are commonly used in drug evaluation, their comparison with humans becomes of paramount importance when ester-containing drugs are being assessed. The specificity of carboxylesterases appears to depend on the nature of the substituents (R, R', R'') rather than on the type of heteroatom (O, S, or N) attached to the carbonyl group. Usually esters are hydrolyzed more rapidly than the corresponding amides, and this difference can be exploited in drug design; the duration of action of ester-containing drugs eliminated by hydrolysis could be prolonged by replacement of the ester with the amide link. The hydrolysis of an ester or amide link is usually associated with detoxication of the drug because of the greater water solubility and thus more facile excretion of the resultant hydrolysis products. However, exceptions exist; for example, salicylic acid produced by the hydrolysis of acetylsalicylic acid is still pharmacologically active and has a longer half-life than the parent drug.

The hydration of epoxides (Eq. 26) can be viewed as a special form of hydrolysis in which the elements of water are added to the epoxide without the resultant formation of more than one product, such as occurs in most hydrolytic reactions. This reaction is catalyzed by epoxide hydrolase.



Epoxides are usually chemically quite reactive as a consequence of the strain inherent in the 3-membered cyclic ether ring and the electrophilic nature of the ring's carbon atoms. Through the action of epoxide hydrolase (epoxide hydrase, epoxide hydratase), they react stereoselectively with water to produce vicinal diols that are predominantly in the *trans* configuration (Eq. 26) because nucleophilic attack by OH^- , at the least hindered carbon atom, occurs from the opposite side of the molecule to where the epoxide ring is situated. Although the diol is less reactive than the parent epoxide, it may serve as a substrate for further epoxidation elsewhere, particularly in large hydrophobic molecules such as polycyclic aromatic hydrocarbons in which it leads, for example, to the formation of highly toxic dihydrodiol epoxides resistant to epoxide hydrolase.

Epoxide hydrolase activity is present in most mammalian tissues and is high in the endoplasmic

reticulum and cytosol of the liver. Humans appear to have considerable intersubject variability in the activity of epoxide hydrolase that may impact on individual susceptibility to the toxic action of certain epoxides: This variability argues for the existence of several epoxide hydrolases.

Some epoxides are resistant to epoxide hydrolase: Carbamazepine, for example, is metabolized to an epoxide that survives biotransformation and is excreted in urine.

CONJUGATION

Phase II drug biotransformation involves reaction (conjugation) of a range of functional groups with endogenous compounds (Table 1). The resultant product is usually more water soluble than its precursor and is thus more readily excreted in the urine or bile. However, some fatty acid conjugates are more lipid soluble than the parent compound. Ester conjugates of the alcohol function of tetrahydrocannabinol and etofenamate with palmitic and other fatty acids have been identified as metabolites in feces and tissues. Although they are only minor metabolites in excreta,

Table 1 Common conjugation reactions

Reaction	Enzyme system	Functional group involved
Glucuronidation	Glucuronyltransferase	—OH —COOH —NH ₂ —SO ₂ NH ₂ —NHOH —SH
Sulfation	Sulfotransferase	—OH —NH ₂ —SO ₂ NH ₂ —NHOH
Glutathione conjugation	Glutathione S-transferase	Electrophilic center
Acetylation	Acetyltransferase	—OH —NH ₂ —SO ₂ NH ₂ —NHNH ₂ —NHOH
Amino acid conjugation	Acyltransferase	—COOH
Methylation	Methyltransferase	—OH —NH ₂ —SH

their high lipid solubility may make them of greater importance in tissues.

Conjugation reactions are catalyzed by group-transferring enzymes.

Glucuronidation

Glucuronidation is quantitatively the most important of the conjugation reactions. It can take place in most body tissues, presumably because the endogenous substrate required—uridine diphosphate *D*-glucuronic acid (UDP-GA)—is produced during intermediary metabolism by processes related to glycogen synthesis (Fig. 4). Therefore, this substrate is less likely to be depleted than those serving other conjugation reactions. Glucuronidation is catalyzed by a family of enzymes, the UDP-glucuronosyl-transferases (glucuronyltransferases); these enzymes are located mainly in the endoplasmic reticulum and are mostly tightly associated with it. An unusual property of these enzymes is their apparent latency, whereby full enzymic activity (in vitro) is expressed only after treatment with agents (e.g., surfactants) that disrupt membrane structure.

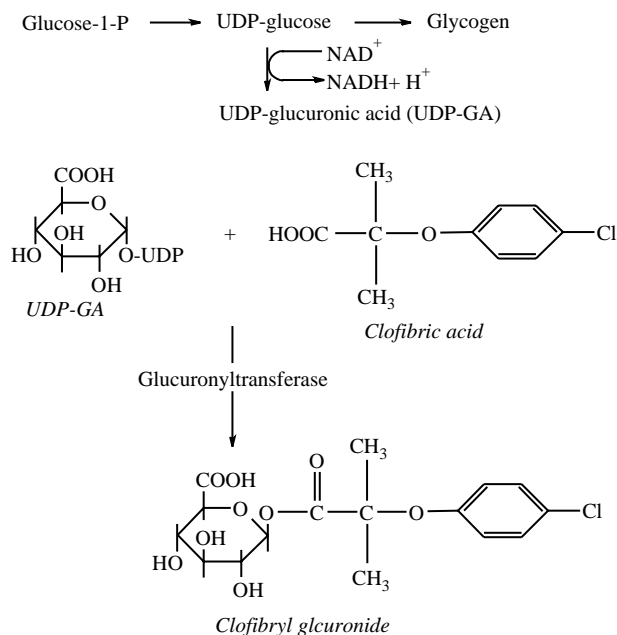


Fig. 4 Formation of uridine diphosphate glucuronic acid (UDP-GA) and its conjugation with clofibric acid. Inversion takes place during conjugation, the α -*D*-glucuronic acid forming a β -*D*-glucuronide. At a more alkaline pH, such glucuronides may undergo intramolecular rearrangement to forms resistant to enzymic hydrolysis by β -glucuronidase.

Glucuronidation can occur with alcohols and phenols (ether glucuronides are formed), carboxylic acids (ester glucuronides are formed), amines, sulfonamides, hydroxylamines, thiols, amides, carbamates, and even certain compounds containing nucleophilic carbon atoms, such as phenylbutazone. Glucuronidation of tertiary amines yields conjugates containing a quaternary nitrogen. Besides foreign compounds, glucuronidation of endogenous substrates, such as steroids, catechols, bilirubin, and thyroxine also takes place.

Less commonly, conjugation with other sugars such as glucose, ribose or xylose can occur, and for example, glucosides are formed following conjugation with UDP-glucose.

Sulfation

Sulfation is an important conjugation reaction that, like glucuronidation, requires a “high-energy” or “activated” endogenous substrate, 3′-phosphoadenosine-5′-phosphosulfate (Fig. 5); however, owing to its lower cellular concentrations, this substrate is more easily depleted than that utilized in glucuronidation. Thus, sulfation is a readily saturable biotransformation process. Where sulfation and

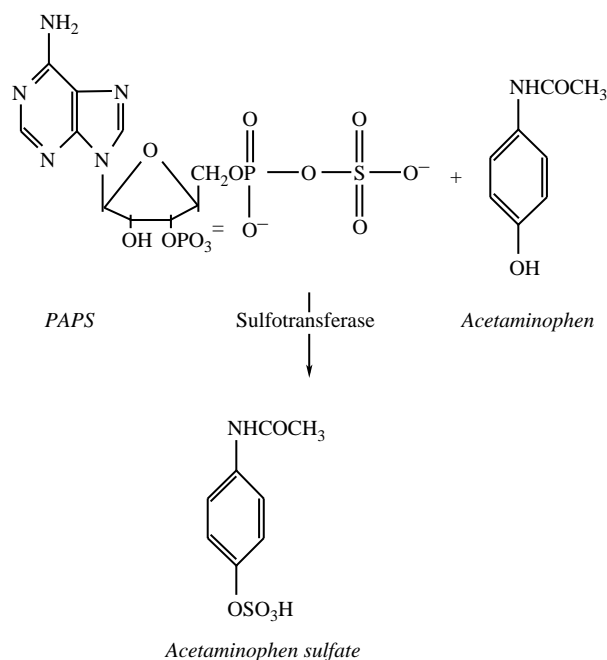


Fig. 5 Formation of 3′-phosphoadenosine-5′-phosphosulfate (PAPS) and its conjugation with acetaminophen (paracetamol). Strictly, the process is one of sulfonation because the $-\text{SO}_3^-$ group is being transferred.

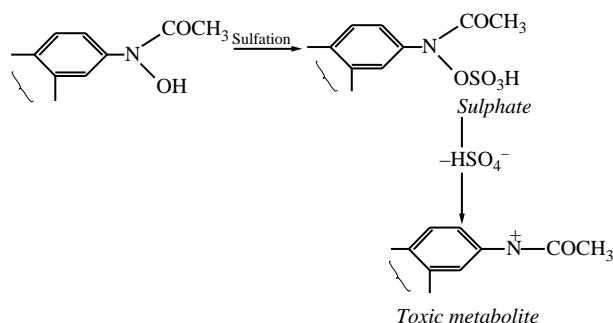


Fig. 6 Toxic metabolite formation mediated via sulfation.

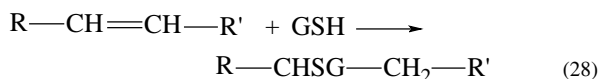
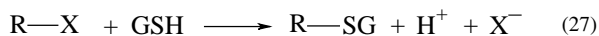
glucuronidation compete, the former tends to predominate at low substrate concentrations and the latter at high concentrations. Depending on the substrate, sulfation occurs in most tissues in the body and is catalyzed by a family of enzymes, the sulfotransferases, that are located in the cytosol.

Alcohols, phenols, hydroxylamines, and amines can be sulfated, as can endogenous substrates such as steroids, polysaccharides, and biogenic amines. Sulfation can lead to metabolites that are more toxic than their precursors (Fig. 6).

Glutathione Conjugation

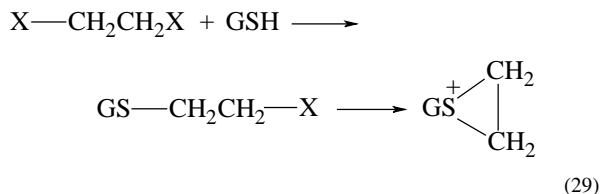
Compounds possessing a sufficiently electrophilic center can conjugate readily with the endogenous nucleophile and tripeptide, glutathione (GSH). Although this reaction can proceed nonenzymically, it is usually catalyzed by the glutathione S-transferases, a group of nonspecific enzymes with overlapping substrate specificities that can also serve as binding proteins.

The electrophilic center may already be present in the compound, or it may arise as a consequence of Phase I biotransformation, as in epoxides. Conjugations with GSH may be classified broadly as replacement (substitution; Eq. 27) or addition (Eq. 28; R' = electron-withdrawing group) reactions. With the latter type in particular, new asymmetric centers are created, but only recently has the stereochemistry and stereospecificity of the reaction begun to attract interest.



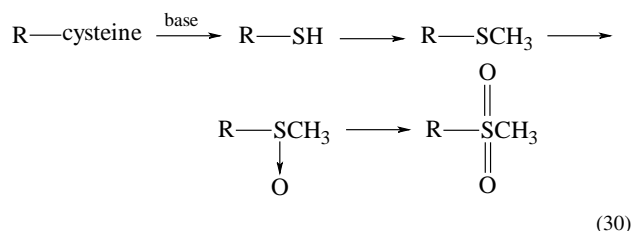
GSH exerts an important protective function in the body, and conjugation with GSH generally results in

detoxication, although the chemical nature of certain substrates can facilitate the formation of a toxic conjugate (for example, Eq. 29; X = halogen).



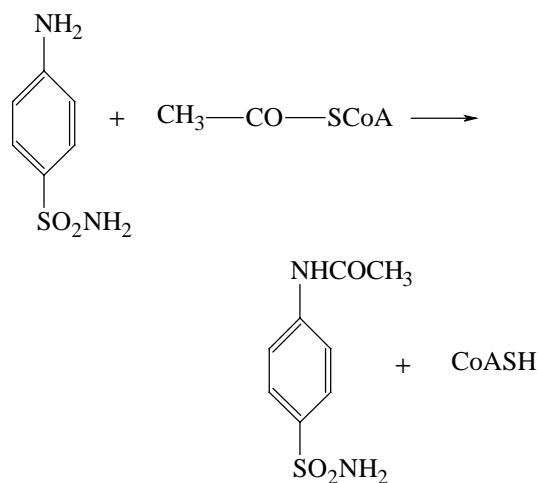
The glutathione S-transferases appear to be present in almost all vertebrate species and their tissues, although certain isoenzymes may exhibit tissue and/or species specificity. Concentrations of the enzymes in certain tissues, such as the liver, are relatively high. They are located primarily in the cytosol but may also be present in other cell structures; those in the endoplasmic reticulum are activated by sulfhydryl reagents. Although GSH is relatively abundant in tissues (for example, about 5mM in liver), it can be severely depleted by large doses of certain drugs; in the case of acetaminophen, depletion leads to hepatotoxicity through a reactive intermediate (a quinoneimine) that is detoxified at clinical doses by conjugation with available GSH. Thus GSH conjugation is a saturable pathway.

GSH conjugates are excreted in bile but almost never in urine. Instead, they are catabolized by removal of the glutamyl and glycine moieties to yield cysteine conjugates, that may be N-acetylated and excreted in urine or in bile as N-acetylcysteine conjugates (mercapturic acids) or metabolites thereof, such as sulfoxides. The conjugates may be subjected to the action of a C-S lyase that enables methylation of the resulting thiol and subsequent oxidation of sulfur to yield the sulfoxide and sulphone. (Eq. 30).



Acetylation

The acetylation of amino groups is a fairly common reaction utilizing acetylcoenzyme A; it is catalyzed by cytosolic N-acetyltransferases. With a substrate such as sulfanilamide (Eq. 31), even diacetylation is possible. Acetylation occurs in the liver and in many other tissues.



(31)

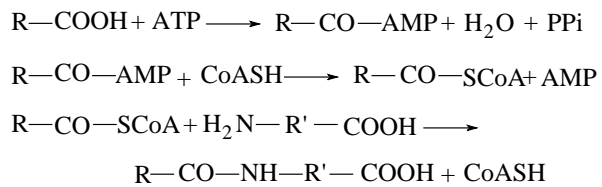
The products of acetylation may be more or less toxic than the precursor. In the case of certain sulfonamides, the less water-soluble acetylsulfonamides can cause renal toxicity owing to their precipitation in the kidney.

Acetylation exhibits a well-known pharmacogenetic difference in the human population that is bimodal (fast or slow) and perhaps even trimodal. This difference accounts for the greater toxicity of certain drugs in either fast or slow acetylators, depending on whether the parent drug or its acetyl derivative is the more toxic entity.

Transacetylation to an acceptor containing an amino group has also been reported; it is enzyme catalyzed.

Amino Acid Conjugation

Another type of acylation involves the conjugation of drugs containing the carboxyl group with amino acids; this process occurs through a coupled system involving the formation of a high-energy intermediate of the carboxylic acid catalyzed by acyl-CoA synthetases, followed by acyl transfer to the amino acid (Eq. 32) catalyzed by acyltransferases. In mammalian species, the amino acids most utilized are glycine, taurine, glutamine, and carnitine. This conjugation occurs extensively in the hepatic mitochondria, but the enzyme systems involved have yet to receive the sort of detailed attention accorded other pathways.

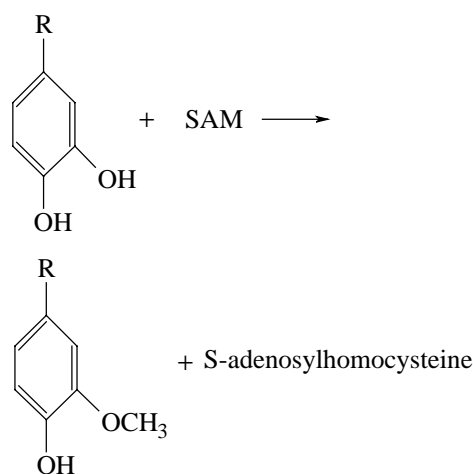


(32)

Amino acid conjugation is complementary to glucuronidation. Some carboxylic acids undergo both types of reaction, whereas others undergo predominantly one type.

Methylation

Although methylation reactions are mainly concerned with endogenous substrates, a number of drugs are also methylated (Eq. 33) by nonspecific methyltransferases that are present in the liver, lung, and other tissues. The enzymes are located mostly in the cytosol, but they may also be membrane-bound. The principal methyl donor is S-adenosylmethionine (SAM) which is formed from ATP and L-methionine.



(33)

The relative importance of O-methylation in the meta or para positions (Eq. 33) appears to depend on the nature of the substituent R. Methylation usually results in the formation of a less water-soluble product, but it is often regarded as a detoxication process. S-methylation is common and, N-methylation, and the methylation of certain metals (e.g., mercury) also occur.

STEREOSELECTIVE AND STEREOSPECIFIC BIOTRANSFORMATION

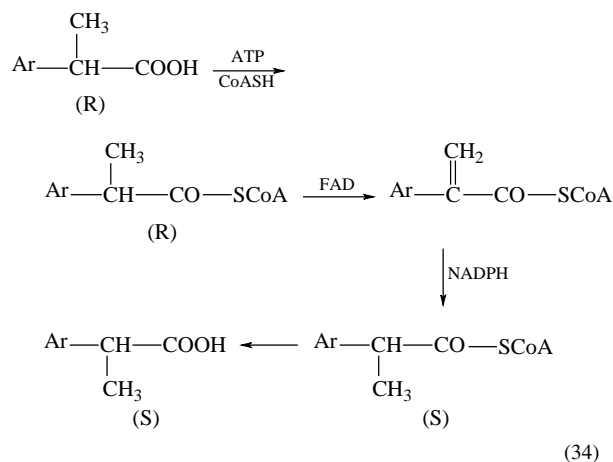
Many drugs contain one or more asymmetric centers, and thus consist of a mixture of enantiomers or diastereoisomers. For each asymmetric center, there is a pair of enantiomers in a mirror-image relationship. Both pharmacologic activity and biotransformation usually require interaction of the drug with specific receptors or enzymes that can be highly stereoselective according to

the steric configuration at the site of interaction. Hence, a pair of enantiomers may differ in biologic activity and rates of biotransformation. The latter results in stereoselective pharmacokinetics, thereby also influencing biologic activity. Due to this fact, there has been a drive toward the development of single enantiomers as new drugs that provides the opportunity to optimise the therapeutic ratio (index). In fact, single enantiomers of existing racemic drugs such as ibuprofen and bupivacaine have been developed as new products. Development of racemic drugs now requires detailed investigations of the pharmacokinetics and biotransformation of the individual enantiomers.

The cardiovascular drug propranolol is used as a racemate (i.e., an equal mixture of two enantiomers), with most of the therapeutic activity mediated by the (–)isomer. One of its major biotransformation pathways in the dog is glucuronidation, and urinary excretion of the glucuronic acid conjugate of (–)propranolol is about fourfold greater than that of the (+)isomer; thus, plasma concentrations of the latter are greater than those of the former. Ring hydroxylation is another important pathway, and for this process in humans, there is a preference for the (+)isomer, leading to greater plasma concentrations of the active (–)isomer.

Some biotransformations introduce an asymmetric center into a drug and these often proceed stereospecifically. The most common examples are hydroxylation of a secondary carbon and the reduction of ketones to secondary alcohols. Ibuprofen undergoes both ω and $\omega-1$ oxidation of the isobutyl side chain, and formation of the resulting carboxylic acid metabolite introduces a second asymmetric center into the molecule. Both ibuprofen enantiomers have been shown to undergo stereospecific oxidation to give a metabolite with the same configuration at the new asymmetric center.

Stereochemical inversion is another aspect of stereospecific biotransformation. The 2-arylpropionic acids containing an asymmetric center represent a major class of the nonsteroidal anti-inflammatory drugs. A notable aspect of their biotransformation is the conversion of one enantiomer to the other. Although most of these drugs were originally developed and used as a racemic mixture, only one of the isomers is usually regarded as being biologically active. The inactive (R) isomer of ibuprofen is converted to the active (S) isomer in vivo. A possible mechanism is dehydration to an α -methylene intermediate; this stage, involving a flavin-dependent dehydrogenase (FAD), is stereospecific (Eq. 34).



FACTORS AFFECTING DRUG BIOTRANSFORMATION

The preceding sections have illustrated that drugs can be biotransformed by a variety of reactions that are generally enzyme catalyzed. The rates of these reactions and their relative importance may be modulated by a number of potential interacting factors that can be broadly classified as physiologic or environmental. The influence of most of the factors mentioned has been revealed by laboratory animal studies, often involving high dose levels and exaggerated systemic exposure. In humans, where a single factor may not be seen to exert any notable effect, the interplay of several factors could contribute to important alterations in drug biotransformation and thereby to drug action.

Physiologic

Age

The activities of the enzymes responsible for drug biotransformation generally seem relatively low in the fetus, then increase rapidly after birth to reach adult levels, and then eventually decline in old age. However, the developmental profiles of the various enzymes probably depend on the species and the enzyme system. The decline of drug biotransformation capability in old age may have important clinical consequences because the elderly comprise much of the patient population, whereas the pharmacokinetics and biotransformation of new drugs are mostly studied and first defined in young, healthy subjects. Similarly, limited drug biotransformation capability in the very young requires careful selection of drug dose regimens in pediatrics.

Disease

In modern drug development, a knowledge of drug biotransformation and pharmacokinetics in certain disease states is essential, especially disease states of the liver and kidney because of their key role in drug elimination from the body. The impairment of these organs by disease could result in altered pharmacokinetics and drug and/or metabolite accumulation in the body. Cardiovascular disease is also of importance because of possible alterations in blood flow that could affect drug transport to the eliminating organs.

Hormones

A deficiency or surfeit of certain hormones can influence the activity of the enzymes involved in drug biotransformation. Apart from the sex hormones, those of the adrenal, pancreas, pituitary, and thyroid also exert effects on the regulation of certain enzyme activities. Some hormonal disease states, such as diabetes and hyperthyroidism, could have important implications for drug biotransformation. Alterations in drug biotransformation can occur in pregnancy, and this subject area requires much more study.

Pharmacogenetics

Intersubject variability in drug biotransformation can be very large primarily due to genetic factors modulated by other factors such as those mentioned elsewhere in this section. Thus the ability of certain individuals to biotransform particular drugs is greatly diminished owing to a genetically inherited deficiency in the enzyme(s) mainly responsible for the biotransformation(s) of those drugs. Although enzyme systems, such as the cytochromes P450 (CYP), have broad and overlapping substrate specificities, one enzyme alone (because of high enzyme-substrate affinity) may be dominant in the biotransformation of a particular drug. Thus a deficiency in this enzyme (or its inhibition by concomitant medication) can lead to exaggerated pharmacologic or toxicologic effects. For example, about 10% of the Caucasian population is deficient in CYP2D6, whereas this deficiency is less than 1% in the Japanese population where the major known CYP deficiency of about 20% is in CYP2C19 (about 5% in Caucasians).

Drugs that have been shown to be subject to pharmacogenetic differences in biotransformation include debrisoquine, perhexiline, phenformin, mephenytoin, tolbutamide, dapsone, isoniazid and sulfadimidine. The

latter three drugs are primarily biotransformed by N-acetylation, a pathway for which about 50% of the United Kingdom population can be classified as slow acetylators and the rest as fast acetylators.

In modern drug development, it is prudent to avoid developing those compounds likely to be exclusively biotransformed by an enzyme known to be deficient in a notable proportion of the population.

Sex (gender)

The biotransformation of certain drugs may differ in males and females. Such sex differences are fairly common in the rat, which is probably the most widely utilized laboratory animal in drug safety evaluation, but have been infrequently reported for humans or even for other laboratory animal species. When encountered, these sex differences appear to arise as a result of the regulation of drug biotransformation by hormones such as growth hormone.

ENVIRONMENTAL

Chemical Exposure

Exposure to chemicals can be intentional or unintentional. The former includes the ingestion of drugs and food additives, alcohol consumption, and tobacco smoking; the latter includes ingestion of agrochemical residues present in food and chemical exposure in the workplace or environment. The extent of exposure may be sufficient to affect the activities of the enzymes involved in drug biotransformation to cause either an increase (induction) or a decrease (inhibition) of these activities. Barbiturates, for example, can cause an increase, and cimetidine can cause a decrease. Induction caused by the organochlorine pesticides DDT and dieldrin and the consequences thereof have been well publicized. The consequences of inhibition of biotransformation are usually more serious than those of induction, and drugs, such as terfenadine and mibefradil, have been withdrawn from clinical use owing to unacceptable drug-drug interactions arising from enzyme inhibition. Paradoxically, others, such as ketoconazole and quinidine, that are potent inhibitors of cytochrome P450 enzymes, remain in clinical use.

Diet

The diet contains an almost infinite number of foreign chemicals, and the enzymes of drug biotransformation

probably evolved to cope with these chemicals. Consequently they were already available to deal with subsequently developed drugs. The activities of these enzymes can be affected by dietary constituents that may serve to increase or decrease them or cause both effects; in alcohol ingestion, enzyme activities show a short-term decrease (inhibition) and then increase (induction). They can also be affected by nutritional status and reflect protein, fat, carbohydrate, mineral, and vitamin intake.

Stress

Stress can perturb the well being of the body and result in alterations in the rate of drug biotransformation. Such stress factors include abnormal temperature, violent exercise, and disease states.

SITES OF DRUG BIOTRANSFORMATION

The liver is usually considered to be the main site of drug biotransformation. Other organs and tissues, including those located at portals of entry or exit of drugs into or from the body, such as the intestinal tract, lungs, skin, and kidneys, also make major contributions. Certain other organs and tissues may be important to the biotransformation of particular drugs; for example, nitroglycerin has been shown to be biotransformed in vascular tissue. Within an individual organ or tissue, the enzyme systems responsible for drug biotransformation are not uniformly distributed but vary in activity according to cell type and differ in subcellular localization. In the lung, for example, cytochrome P450 activity is relatively high in Clara cells.

Although most types of organs and tissues contain the enzyme systems involved in drug biotransformation, the enzyme activities may be expressed by a family of enzymes, whose individual distribution may be either organ or tissue specific. The profile of these enzymes can be altered, for example, during treatment with certain compounds.

The biotransformation of many drugs is often slower in humans than in the laboratory animal species used in drug development and safety evaluation. Lower activities of the respective enzyme systems may be the cause, and indeed the rates of drug biotransformation and the body weights of animal species appear to be correlated to some extent.

IMPACT ON ROUTES OF EXCRETION

The two main routes of systemic excretion of drugs and their metabolites are renal (urinary) and hepatic (biliary). Which of these occurs or predominates depends mainly on the physicochemical properties of the metabolites. Sulfate conjugates are more likely to be present in urine and glucuronic acid conjugates appear more often in bile. Glutathione conjugates are almost exclusively excreted in bile. However, this is both compound- and species-dependent due to molecular-weight thresholds for biliary excretion. Thus, many glucuronides are almost always actively excreted in the bile of rats but excreted primarily in urine of humans and nonhuman primates. Conjugation with sulfate, glucuronic acid, and glutathione increases the molecular-weight by as much as 80, 177, and 307 mass units, respectively. The molecular-weight thresholds for active biliary excretion are about 330 for the rat, and dog; 450 for the rabbit and nonhuman primate; and 500 for humans.

METHODS FOR THE STUDY OF DRUG BIOTRANSFORMATION

In Vivo

Investigations of biotransformation pathways in vivo require the collection and analysis of appropriate biologic samples. The types of sample collected include urine, feces, expired air, blood and/or plasma, bile, milk, saliva, synovial fluid, and tissues. These samples can be divided into two groups: 1) those requiring complete collection (e.g., urine and feces) in order to provide quantitative as well as qualitative information on the excretion of the drug and its metabolites, and 2) those such as blood and milk that are subsampled at specific times to yield information on the identity and time-related concentrations of the drug and its metabolites that contribute to systemic exposure. Samples of the above types can be obtained from all the common laboratory animals used in biomedical research. In addition, with the exception of bile and tissues, similar samples can usually be obtained from humans without great difficulty.

One approach to the identification of metabolites is to investigate the presence of specific compounds in biologic samples when reference compounds are available. Quantitative data will require validated bioanalytic methods for each specific compound and may not be readily obtained from certain biological matrices (e.g.,

feces). In addition, the possible presence of, and interference from, unknown metabolites would frequently create uncertainty.

Another approach is to attempt isolation and identification of all major metabolites, regardless of the availability of reference compounds. This approach requires an unambiguous method of detecting metabolites, such as use of a radiolabeled form of the drug. Provided that the drug is radiolabeled in a metabolically stable position, this approach allows accurate measurement of metabolites in terms of percentage of administered dose and thence in concentration or quantity units. Procedures are well established for the detection and measurement of radiolabeled metabolites after chromatographic separation mainly by high-performance liquid chromatography and thin-layer chromatography.

Although chromatographic comparison with reference compounds may provide reasonable evidence for the identity of metabolites, more rigorous characterization by application of spectroscopic techniques to isolated samples is highly desirable. Isolation and purification of metabolites are an important prelude to successful structural elucidation. The higher the state of chemical purity, the greater the chance of success in identification. Skillful selection of procedures according to the physicochemical properties of the metabolite can greatly facilitate generation of a viable sample. The most common technique for structure elucidation is mass spectrometry due to the very high sensitivity and versatility of this technique for the introduction of samples into the mass spectrometer. Further developments to increase sensitivity and specificity, new ionization techniques, and direct combination with high-performance liquid chromatography are ensuring the prime role of mass spectrometry. Other techniques that complement mass spectrometry are nuclear magnetic resonance spectroscopy (particularly the emerging technique of LC-NMR), that generally requires larger samples of greater purity, and less commonly, infrared and ultraviolet spectroscopy.

In Vitro

Several biologic systems can be utilized for the study of drug biotransformation in vitro. These range in remoteness from the living animal in the following order: perfused organs, tissue slices, isolated cells, subcellular preparations, crude enzyme preparations, and purified enzymes. In complexity, these systems range from the relatively simple to the exceedingly awkward; each system has its advantages and disadvantages, and none is ideal. Undoubtedly the most popular system has

been the subcellular preparation because of its ease of production and convenience of use, and because expensive equipment is not always necessary. Indeed, much of the earlier biotransformation knowledge in the literature has been gleaned from the use of subcellular preparations. However, for a better understanding of catalytic mechanism, substrate specificity, and enzyme kinetics, purified enzymes are necessary. In recent years, the increasing availability of cDNA-expressed human enzymes responsible for drug biotransformation is greatly facilitating drug discovery and development. Several of the human cytochrome P450 enzymes and some others are commercially available in this form. Perfused organs and isolated cells are often regarded as providing a better reflection of what might occur in the intact animal in vivo when the overall biotransformation of a compound is being considered; the latter are the much more popular.

BIBLIOGRAPHY^a

Journals

Chemical Research in Toxicology, Drug Metabolism and Disposition, European Journal of Drug Metabolism and Pharmacokinetics, Xenobiotica.

Review Series

Hawkins, D.R., Ed. *Biotransformations*; Royal Society of Chemistry: London, 1989–1996; 1–7.
Hinson, J.A., Ed. *Drug Metabolism Reviews*; Marcel Dekker, Inc.: New York.

Monographs

Gibson, G.G.; Skett, P. *Introduction to Drug Metabolism*, 2nd Ed.; Chapman and Hall: London, 1993.
Hansch, C., Sammes, P.G., Taylor, J.B., Eds. *Comprehensive Medicinal Chemistry*; Pergamon: Oxford, 1990; 5.
Testa, B.; Caldwell, J. *The Metabolism of Drugs and Other Xenobiotics: Biochemistry of Redox Reactions*; Academic: London, 1995.
Timbrell, J. *Principles of Biochemical Toxicology*, 3rd Ed.; Taylor & Francis: London, 1999.
Woolf, T.F. *Handbook of Drug Metabolism*; Marcel Dekker, Inc.: New York, 1999.

^aFor earlier literature, the reader is referred to titles cited in the first edition (1990) of this article on the Biotransformation of Drugs.

BIOSYNTHESIS OF DRUGS

Geoffrey A. Cordell

Kyung Hee Lee

University of Illinois at Chicago, Chicago, Illinois, U.S.A.

BIOSYNTHESIS AND BIOGENESIS

Metabolism is the series of pathways operating when biological systems synthesize their constituents. *Biosynthesis* is the experimentally established pathway of formation of secondary metabolites; where experimental proof is absent, the term *biogenesis* is used.

SECONDARY METABOLITES

Reaction products that are necessary for the generative functions (respiration and catabolism) of an organism are termed primary metabolites; those resulting in products used for other functions are known as *secondary metabolites*. Such compounds typically characterize the individuality of an organism or a group of organisms; this study is termed chemotaxonomy. Consequently, the biosynthetic pathways of secondary metabolism are not random but are highly conserved. Thus, a given plant family may produce substantial numbers of a certain type of metabolite (e.g., quassinoids in the Simaroubaceae), whereas another family produces quite different metabolites (e.g., monoterpene indole alkaloids in the Apocynaceae).

THE PRECURSORS

A select group of primary metabolites, predominantly acetate, shikimic acid, isopentenyl pyrophosphate, and a few amino acids, is responsible for the diversity of the 135,000 plant-derived secondary metabolites in 12 major classes (Fig. 1). Only those classes with some social or economic significance as bioactive agents are discussed here.

SIGNIFICANCE OF BIOSYNTHESIS

Biosynthetic knowledge is an integral and essential aspect of natural products chemistry and has recently assumed

high-profile academic and commercial significance for biotechnological reasons. It is the foundation permitting a systematic and rational framework for organizing the bewildering structural diversity of mammalian, arthropod, insect, plant, microbial, and marine secondary metabolites. It provides structural clues for new metabolites through biogenetic possibilities based on established biosynthetic schemes. It permits the bioengineering of metabolic pathways for enhanced yields or the altering of desired product profiles for greater economic gain. Manipulations at the genetic level may afford a substantially new array of metabolites for future drug discovery. Finally, it is of fundamental human curiosity to discern how secondary metabolites are produced in living systems. The modification of such processes at the enzyme or gene level may be of critical importance in the treatment of mammalian processes involved in disease states; cholesterol synthesis-inhibiting drugs, such as mevinolin, are an example.

METHODS IN BIOSYNTHESIS

Biogenetic theories arose as the need to classify diverse secondary metabolites became apparent. Biosynthetic experimentation tested hypotheses when the precursors became available in radio-, and subsequently stable, isotope labeled forms. It focuses on the study of precursor relationships (including the stereochemistry of specific processes) and of the enzymes involved in the succinct steps in the pathway. A typical experiment involves the administration of a potential precursor in labeled form to the organism, at a time when it is known that the organism is actively producing the metabolite of interest. After an appropriate period of time, the organism is processed for the metabolite of interest and the isotope content located and measured. Most of the fundamental pathways of natural product biosynthesis were established using the radioactive isotopes of hydrogen and carbon, ^3H and ^{14}C . Extensive use is now made of the stable isotopes of the key atoms present in natural products, ^{13}C , ^{18}O , ^{15}N , and ^2H .

Evaluation of the structural complexity of natural products, coupled with prior biosynthetic knowledge,

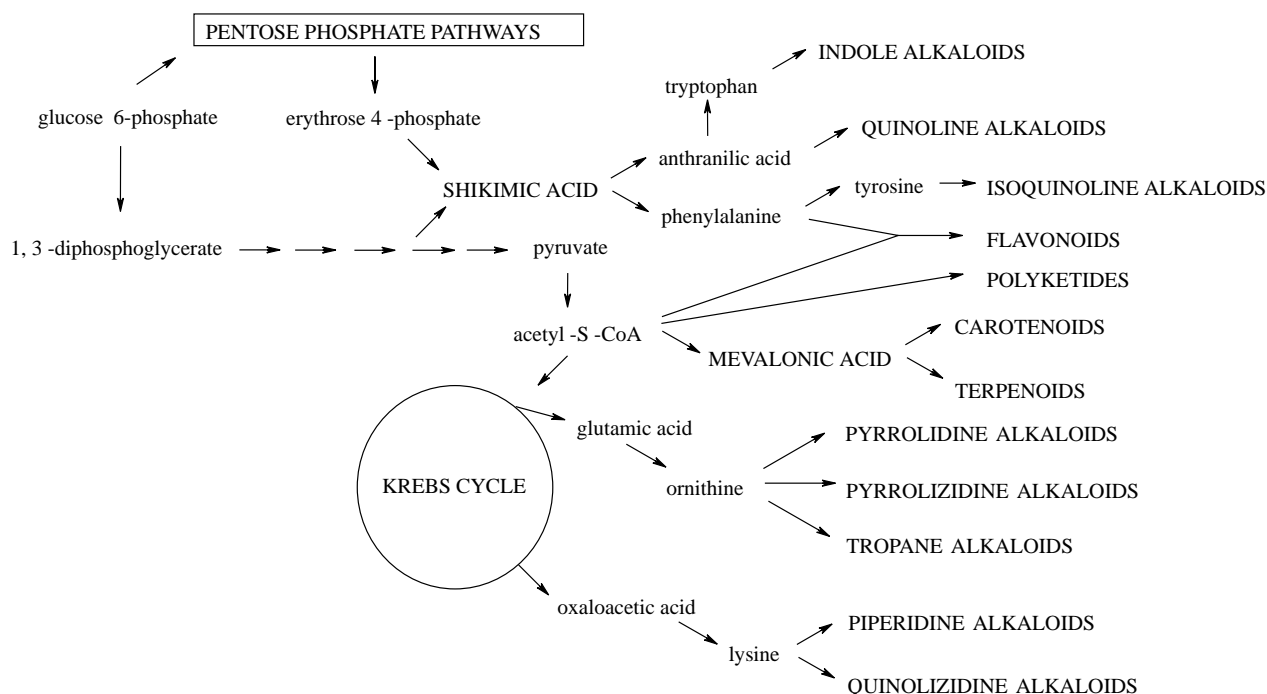


Fig. 1 The biosynthetic origin of secondary metabolites.

frequently offers a rational overview of the necessary chemical transformations. This may lead to isolation of the enzymes and perhaps the involvement of unanticipated intermediates. Although frequently these pathways are conserved, given the diversity of metabolic systems, it should not be assumed that the same compound will be produced by the same biosynthetic pathway in a different organism, i.e., in one plant family compared with another.

For biosynthetic experiments to be meaningful, the precursor must reach the site of synthesis at a time when the enzyme systems mediating metabolite formation are both present and active. Secondary metabolism occurs at a discontinuous rate in a given organism and at different points in the growth cycle in different organisms. Consequently, establishing a time for precursor administration to the organism is critical, if substantial degradation is to be avoided. Transportation and permeability factors may result in very low incorporations, even if the organism is known to be producing secondary metabolites at the time of the feeding and a known precursor is being used. Translocation of the precursor to the site of synthesis may be important in studies with whole plants or callus tissue. For microorganisms, metabolic degradation may occur because the growth of cell mass frequently precedes the initiation of secondary metabolite production.

Use of Radioisotopes

Although a single radiolabel may be adequate to demonstrate a preliminary precursor relationship through incorporation, it is preferable to use a precursor containing two, different, strategically placed labels. There are two types of these experiments, one in which two labeled precursors are physically mixed and the ratio of labels monitored. An example is a feeding experiment with $[2-^{14}\text{C}, 4-^3\text{H}_2]$ -mevalonic acid. A second experiment involves using a precursor in which the two labels are in the same molecule; the use of $[1,2-^{13}\text{C}_2]$ -acetate is an example. Double- or multiple-labeled substrates are used to examine bond-forming and bond-breaking reactions and to examine the stereospecificity of enzymatic processes if the precursor is labeled stereotopically (e.g., $[4R-^3\text{H}]$ -mevalonic acid) and is selectively retained in the product or if the label in the product can be assigned stereotopically. Techniques are available for distinguishing between the prochiral hydrogens on a methylene or a methyl group.

Use and Detection of Stable Isotopes

The common stable isotopes used in biosynthetic studies are ^{13}C , ^2H , ^{15}N , and ^{18}O . Stable isotope-labeled

precursors have replaced radiolabeled precursors in many biosynthetic studies for the following reasons: 1) no appropriate radiolabeled isotope is available (e.g., N and O); 2) the detection methods frequently permit location of the label in the product directly; and 3) radiocontamination and safety issues are reduced. The negative aspects of stable isotope studies are: 1) detection methods are relatively insensitive (higher incorporation levels needed); 2) high levels of enrichment of the label in the precursor are required; and 3) reasonable quantities of the precursor are necessary, which can be expensive.

Mass spectrometry and NMR spectroscopy are the dominant techniques for detecting stable isotopes. MS offers the advantage that less sample is needed to establish incorporation, whereas NMR typically permits direct determination of the labeled site.

Administration of Precursors

When the system under examination is microbial, i.e., a plant in tissue culture or a cell-free system, administration of the precursor is straightforward. For meaningful conclusions, a profile of formation of the compound of interest over time is necessary so that feeding is conducted when there is active product formation. When intact plants are used, precursor feeding is more difficult; options are wick feeding through the stem, root feeding, isolated leaf feeding, or even direct injection into the stem.

Examining Intermediates

Conclusively establishing the role of potential intermediates in a biosynthetic pathway is a difficult aspect of biosynthesis. Typically, intermediates accumulate because subsequent enzymatic reactions are slow. Organisms also produce shunt metabolites that are off the main pathway and may not be further metabolized; these will also accumulate. Isolation of an "intermediate" does not, therefore, establish intermediacy. Trapping experiments are sometimes used to overcome these problems. In the pathway $A \Rightarrow B \Rightarrow C$, where A is a known precursor of C , labeled A and nonlabeled B are fed at the same time. The latter is metabolized to C and labeled B is produced from A ; B is then temporarily available for isolation. An alternative approach for microbial metabolites is to mutate the organism or add specific enzyme inhibitors. This may allow intermediates to accumulate. Incorporation of a labeled, potential intermediate into a product does not prove that the intermediate lies on the main biosynthetic pathway. It may simply serve as a substrate for the enzymes involved. Only when each of the enzymes in a

pathway has been isolated and characterized, and the substrate specificity determined, can the intermediates in a biosynthetic route be characterized.

Enzymes and Genes

Biosynthetic pathways are characteristically under enzymatic control and proceed with a very high degree of stereospecificity. Compared with the number of steps in the pathways of significant natural products, very few enzymes have been isolated and characterized and even fewer cloned and expressed. In the future, it will be very important to be able to express these enzymes heterologously in more productive systems so that these biocatalysts can be used for both known metabolite production and new metabolite generation.

One dream of the biosynthetic chemist is to develop a system of stabilized enzymes on solid supports, permitting a continuous flow process from precursors to products. With no variability due to climate or soil conditions, yields would be totally controllable and reproducible, and product clean-up would be greatly simplified, or ideally, unnecessary. With the isolation, characterization, cloning, and expression of more enzymes in biosynthetic pathways, the reality of the dream moves inexorably closer. Already the use of enzyme systems for directing stereospecific reactions in organic synthesis has risen dramatically, with a corresponding increase in efficiency and enantioselectivity.

Combinatorial Biosynthesis

Another biosynthetic dream is the ability to modulate predictably the product profile of an organism. In the microbial and plant tissue culture areas, this can be achieved randomly by modifying the growth medium or by challenging the organism with a chemical or other external agent (such as a fungus), producing metabolic stress metabolites (allelochemicals). A more controllable route to altering a metabolic profile in the polyketide area can be achieved through selectively modifying the gene sequence of the biosynthetic pathway. Known collectively as combinatorial biosynthesis, this way allows new products to be formed for chemical analysis and biological evaluation.

COMPOUNDS DERIVED FROM ACETATE

Acetyl coenzyme A is the biosynthetically active form of the two-carbon building block, acetate. It is of central

importance in mammalian, plant, and microbial biochemistry, giving rise to the fatty acids, the polyketides, and through mevalonic acid, the terpenes.

Fatty Acid Biosynthesis

The common fatty acids, such as palmitic (C_{16}), stearic (C_{18}), and arachidonic (C_{20}), have an even number of carbons. Their chain building process initially involves a reaction of acetyl CoA with carbon dioxide to afford the more chemically reactive malonyl CoA, which condenses with a second acetate unit. The carbon dioxide subsequently lost is the same carbon that was added. It is this specificity that allows correlative experiments with $[1,2-^{13}C_2]$ -acetate. Reduction of the beta carbonyl group is followed by dehydration and reduction of the *cis*-olefin to the saturated fatty acid. Repetition causes chain extension by two-carbon fragments (Fig. 2). Unsaturated fatty acids can also result from dehydrogenation, as shown in the conversion of oleic acid to arachidonic acid. The latter compound is the precursor of the prostaglandins (Fig. 3). Branching in fatty acids may occur either through the initiating acid, through acylation with a preformed fatty acid, or through reaction of an intermediate olefin with methionine.

The Polyketide Pathway

Aromatic compounds are predominantly formed through either the shikimate (vide infra) or the polyketide pathway. Collie first suggested the polyketide pathway in 1893, and this was extended theoretically (acetate hypothesis) and experimentally by Birch. Factors involved in the diversity of products include the chain-initiating unit, the number of units in the cyclizing chain, condensation reactions occurring between separately formed polyketide chains,

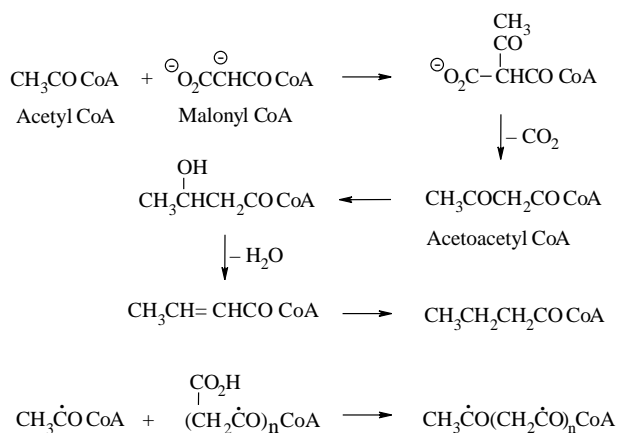


Fig. 2 The biosynthesis of fatty acids.

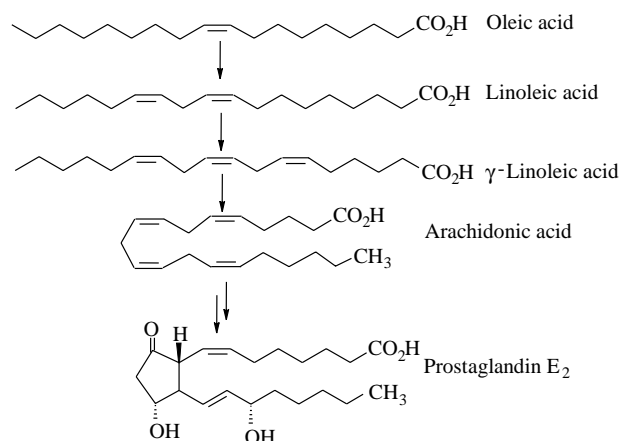


Fig. 3 The biosynthesis of prostaglandins.

and secondary processes, such as alkylation or halogenation. The 1,3-diketone nature of the intermediate chain leads to a characteristic *meta*-relationship between ether or phenolic groups; in shikimate-derived metabolites these groups are typically *ortho*-related.

Penicillic acid provides an example of a simple tetraketide whose aromatic ring is cleaved and cyclized, as shown by experiments with $[1,2-^{13}C_2]$ -acetate (Fig. 4). On the other hand, the pentaketide citrinin is the result of the cyclization of a linear polyketide chain wherein three methyl groups are introduced from methionine (Fig. 5). Hexaketide derivatives are rare. The naphthoquinone plumbagin is an example where cyclization followed by decarboxylation occurs. Griseofulvin is a heptaketide (Fig. 6) and was one of Birch's very early demonstrations of the accuracy of the acetate hypothesis. Anthraquinones, such as islandicin, are octaketide derivatives; the two alternative modes of cyclization of the polyketide chain

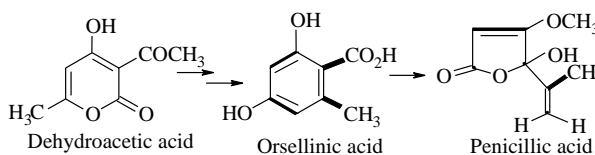


Fig. 4 The biosynthesis of penicillic acid.

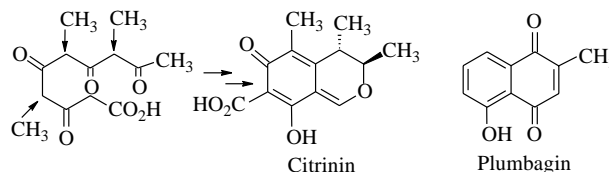


Fig. 5 The biosynthesis of citrinin.

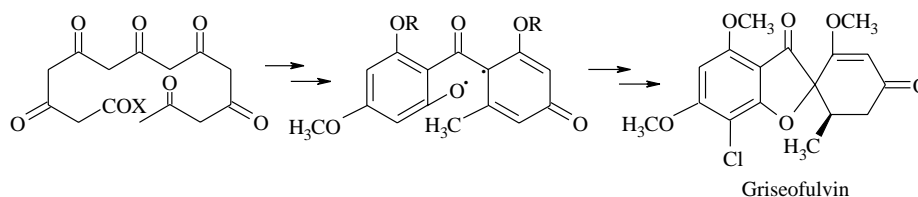


Fig. 6 The biosynthesis of griseofulvin.

can be distinguished through the use of $[1,2-^{13}\text{C}_2]$ -acetate. Xanthenes can be produced through oxidative cleavage of the quinone ring, cyclization and decarboxylation (Fig. 7).

The most clinically significant polyketides are the anthracycline and tetracycline antibiotics produced in *Streptomyces* cultures. The tetracyclines demonstrate that chain initiation can occur with a malonamide unit and that a wide range of reactions can occur after the initial aromatic cyclization (Fig. 8). Mixed biosynthesis is very evident in the macrolide antibiotics, where various combinations of acetate and propionate form the chain (e.g., tylosin and nystatin) or only propionate (e.g., erythromycin) and cyclize to a ring of varying size. Several plant-derived anthraquinones are of clinical significance, including the sennosides of senna (*Cassia angustifolia*), the aloins of aloe (*Aloe vera* and related species), the cascarosides of cascara sagrada (*Rhamnus purshiana*), and hypericin of Saint John's Wort (*Hypericum perforatum*; Fig. 9).

SHIKIMATE PATHWAY

The majority of aromatic compounds, including most alkaloids, are derived through the shikimate pathway. At

the branching point of chorismic acid, either anthranilic acid, the precursor of tryptophan, or prephenic acid, the precursor of phenylalanine, itself the precursor of tyrosine and dopa (3,4-dihydroxy-phenylalanine), is formed (Fig. 10). Phosphorylation at the 3-position, condensation with phosphoenolpyruvate, and elimination of phosphoric acid yields chorismate from shikimate. Chorismate is also the precursor of a number of simple, and very important, aromatic compounds, including salicylic acid, 4-amino-benzoic acid (PABA), a constituent of folic acid, and 2,3-dihydroxybenzoic acid, a key acylating group of enterobactin.

Amination at the 2-position and loss of pyruvate yields anthranilic acid, the precursor of the quinoline alkaloids, distributed widely in the Rutaceae, and tryptophan, the precursor of the indole alkaloids (Fig. 10). Internal Claisen rearrangement on chorismic acid yields prephenic acid en route to phenylalanine. During the course of the hydroxylation of phenylalanine to tyrosine, there is a characteristic NIH shift of the proton at C-4' (Fig. 11). Further hydroxylation yields dopa, used in the treatment of Parkinson's disease. Dopa can oxidize and polymerize to yield the melanin group of hair, skin, and eye pigments. 4'-amination of chorismic acid, a Claisen rearrangement, and amination yields 4'-aminophenylalanine, whose importance is as a precursor of chloramphenicol (Fig. 12).

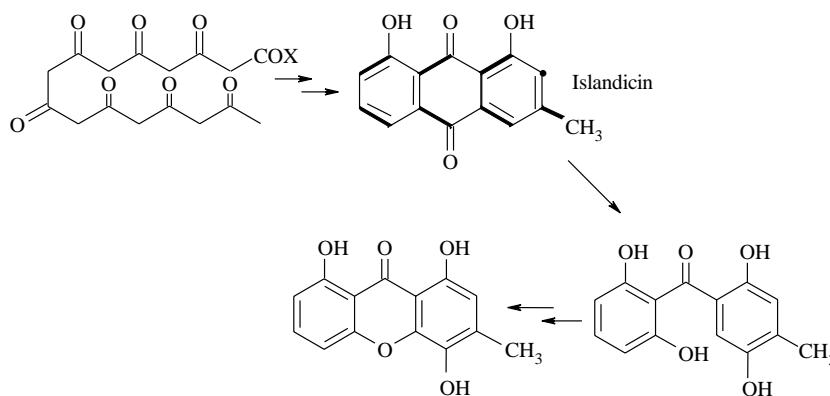


Fig. 7 The biosynthesis of anthraquinones and xanthenes.

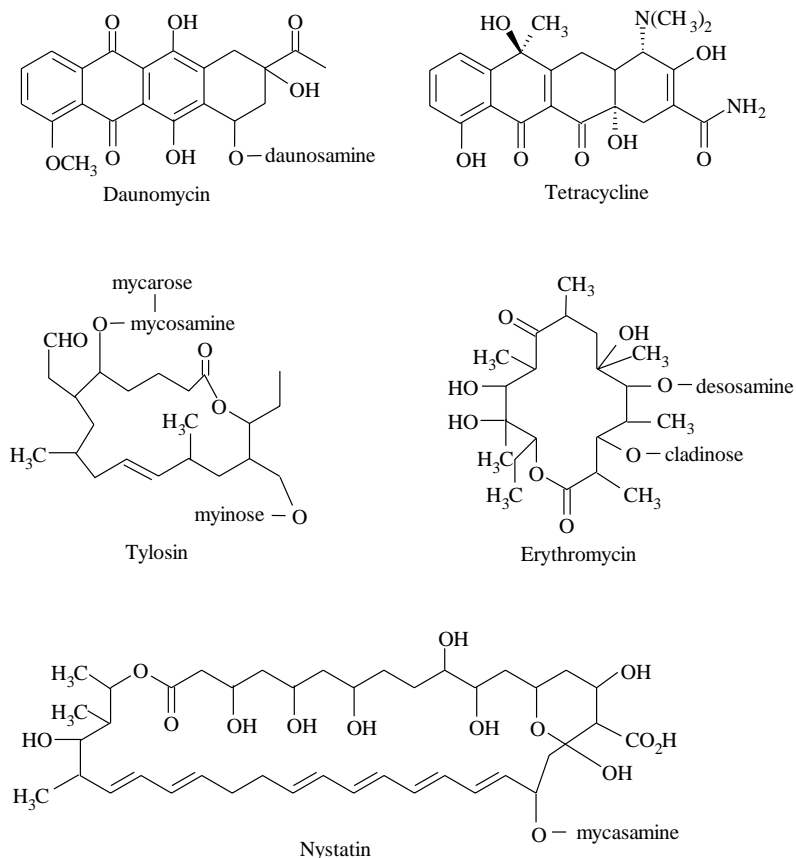


Fig. 8 Representative polyketide antibiotics.

Oxidative deamination of phenylalanine by phenylalanine ammonia lyase (PAL) and 4-hydroxylation affords *p*-coumaric acid, whose derivatives are the fundamental building blocks of lignin, as well as the lignans, such as the

potent anticancer agent podophyllotoxin (Fig. 13). The latter is the template for the drugs, teniposide and etoposide.

2'-Hydroxylation of *p*-coumaric acid, followed by photocatalyzed isomerization of the double bond and

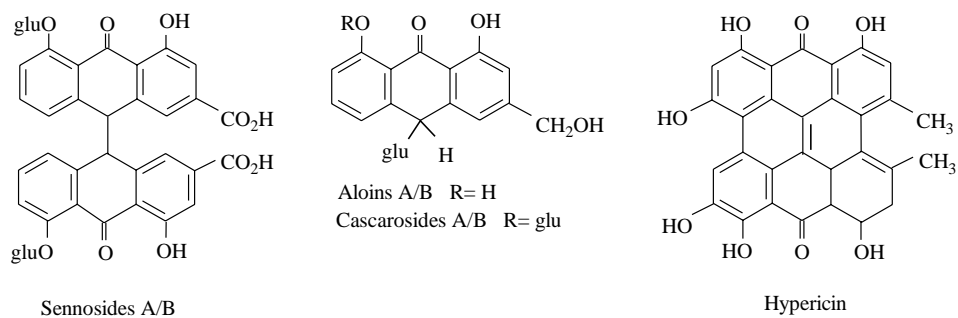


Fig. 9 Representative plant-derived anthraquinones.

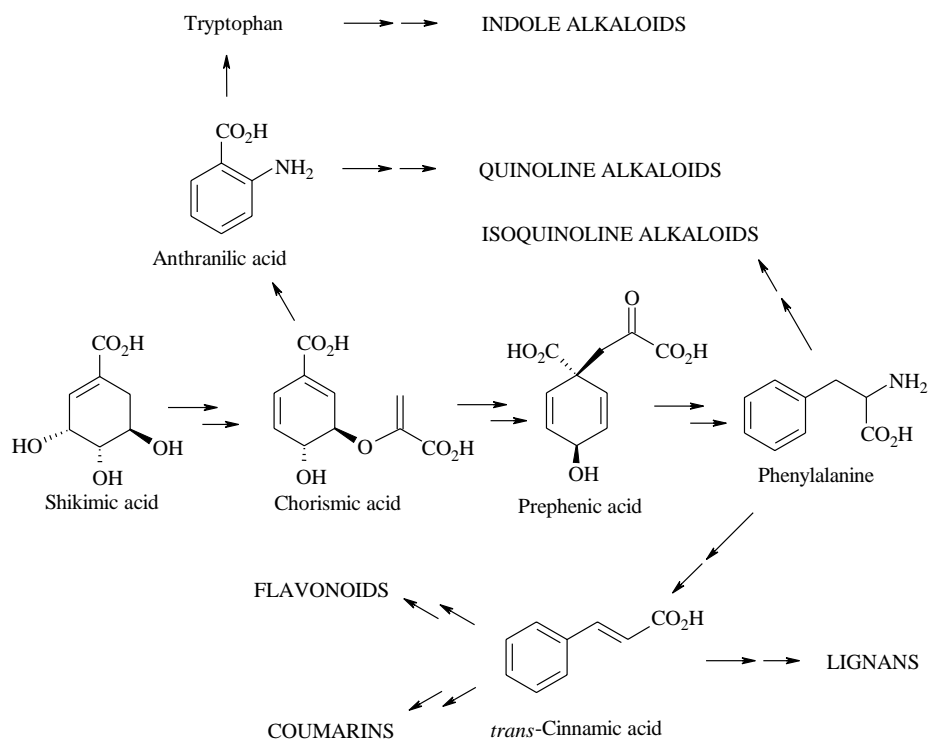


Fig. 10 The relationships of shikimate-derived compounds.

lactonization, affords 7-hydroxy coumarin (umbelliferone). Prenylation at the 6-position, epoxidation, cyclization, and a retro-aldol reaction afford the furanocoumarins (Fig. 14); some (psoralen, bergapten) are known as photosensitizers and find use in the treatment of vitiligo and psoriasis. Coumarin stimulates the reticulo-endothelial system and served as a model for anticoagulant drugs, such as dicoumarol and warfarin.

Flavonoids are essentially universal plant pigments and exist in at least nine different structure classes, frequently with attached sugar units. They are derived from a mixed acetate-shikimate biosynthesis. 4-Coumaroyl CoA reacts with a triketide unit to afford 4,2',4',6'-tetrahydroxychalcone, which cyclizes to naringenin under the influence of chalcone isomerase. 3-Hydroxylation and dehydrogenation leads to the

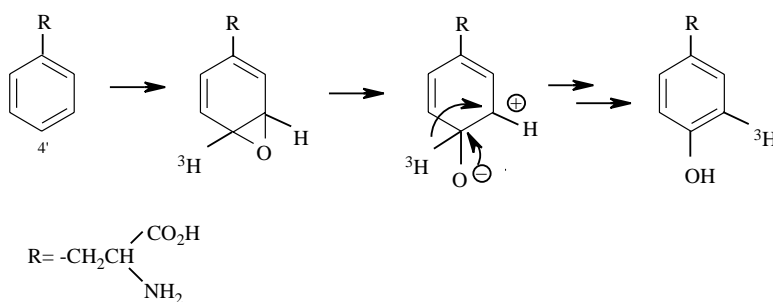
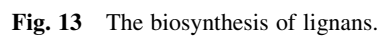
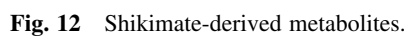
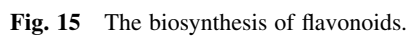
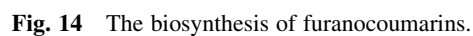


Fig. 11 The NIH shift in the 4'-hydroxylation of phenylalanine.





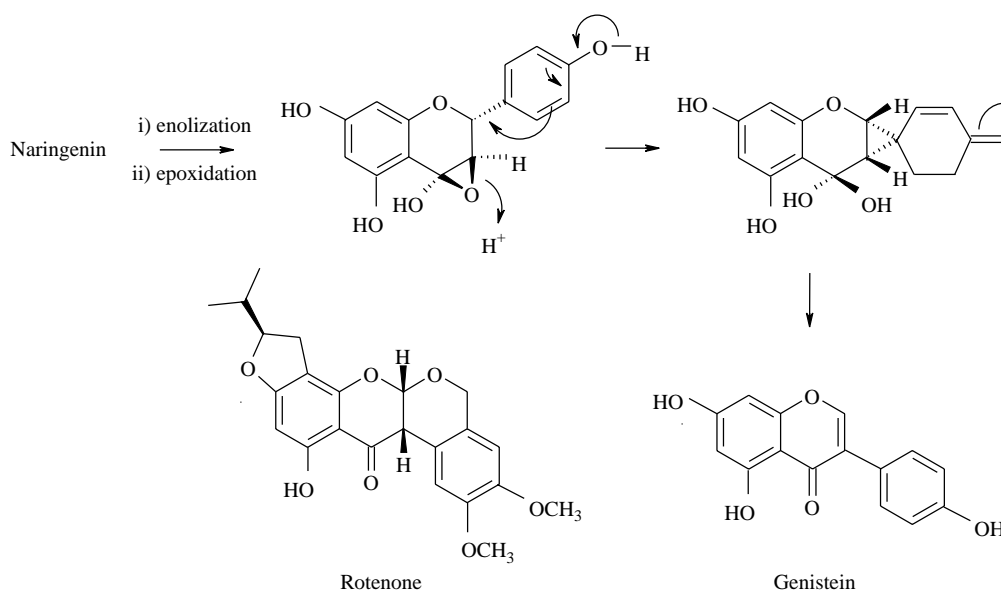


Fig. 16 The biosynthesis of isoflavonoids.

flavanol, kaempferol (Fig. 15), which, along with quercetin (3'-hydroxykaempferol), is widely distributed. There is very substantial interest in the flavonoids present in the diet for their wide range of *in vitro* activities. *Silybum marianum* (milk thistle) is used in Europe as an antihepatotoxic agent for mushroom poisoning, where the active ingredient is silymarin (a mixture of flavanolignans).

Isoflavonoids are of limited distribution (Fabaceae, bean family) and are probably formed through the epoxidation and rearrangement of the enol form of a flavone (e.g., the conversion of naringenin to genistein; Fig. 16). Several isoflavonoids are associated with strong estrogenic activity, and there is interest in their potential in the prevention of hormone-dependent breast cancer. Rotenoids (e.g., rotenone) from *Derris* and *Tephrosia* species are noted for their insecticidal and cytotoxic activity.

COMPOUNDS DERIVED FROM ISOPENTENYL PYROPHOSPHATE

Numerous natural products contain units derived from a terpene precursor. Built up of five carbon "isoprene" units, they are successively known as hemiterpenes (C_5),

monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), sesterterpenes (C_{25}), triterpenes (C_{30}), and tetraterpenes (C_{40}). Successive units may join through head-to-tail (e.g., farnesol) or tail-to-tail linkages (e.g., squalene). Isopentenyl pyrophosphate (IPP), derived from either mevalonic acid or 1-Deoxyxylulose, is the moiety isomerizing to dimethylallyl pyrophosphate (DMAPP) and is also the chain extending unit. Fig. 17 shows the relationships between these terpenes and how the steroids are derived from a triterpene precursor. The formation and occurrence of these metabolites is widespread, and many derivatives are essential for mammalian functions (e.g., cholesterol, steroid hormones, bile acids, vitamin D, and retinols).

Geraniol is the primordial monoterpene, and its simple derivatives and cyclization products occur in the oils of many plants, used as flavoring and aromatic agents (e.g., caraway, coriander, dill, eucalyptus, lavender, orange, peppermint, rose, and sandalwood), as well as drugs (e.g., camphor, menthol), and insecticides (e.g., pyrethrins). Chain extension leads to farnesol, which can dimerize to squalene or undergo its own molecular modifications to yield the diverse sesquiterpenes. One of these, artemisinin from *Artemisia annua*, is of significance as an antimalarial agent, and another, (–)-gossypol, is a male contraceptive agent (Fig. 18).

Several diterpene derivatives are commercially significant drugs. Forskolin, from the Indian medicinal plant

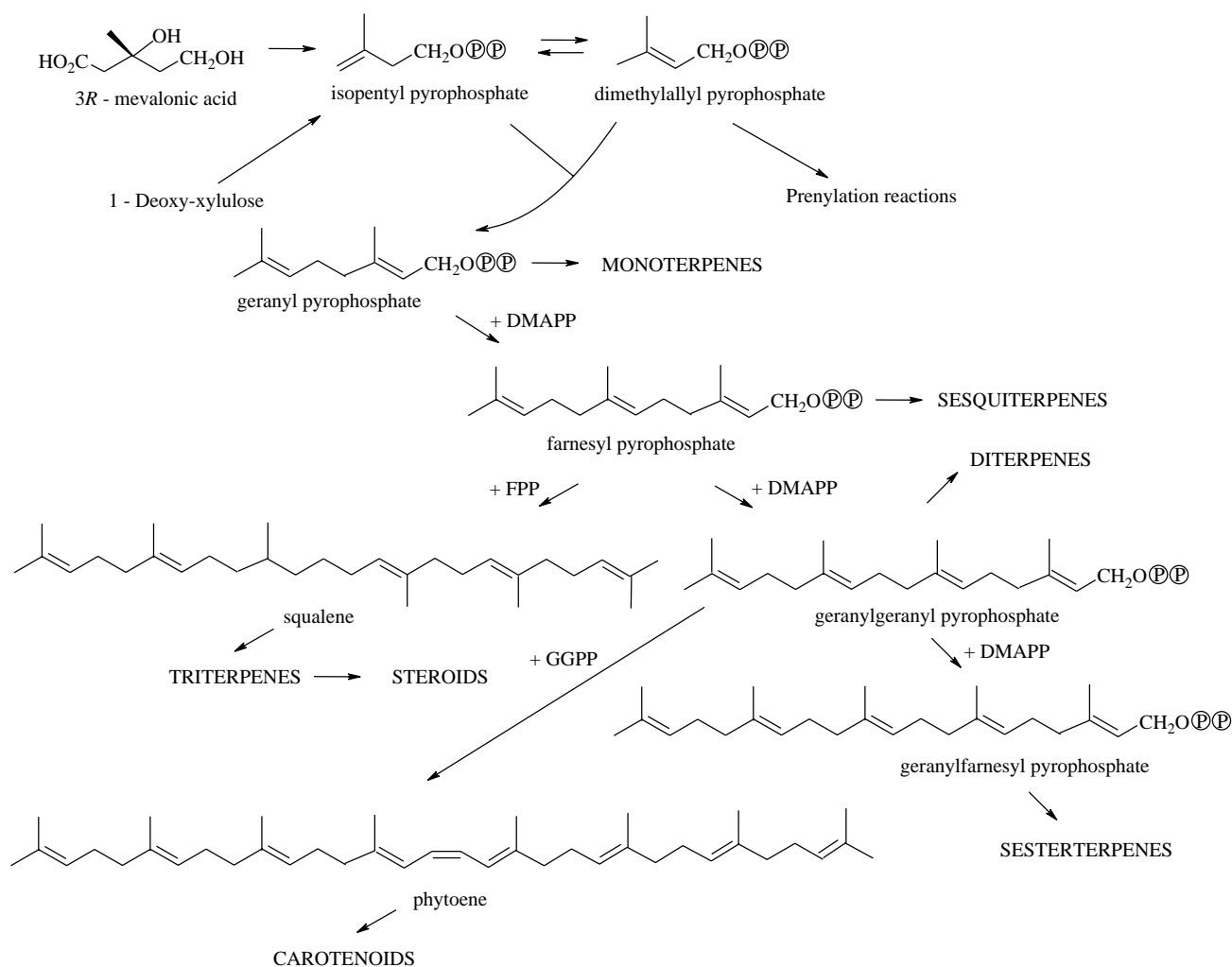


Fig. 17 The relationships of the isoprene-derived compounds.

Coleus forskohlii, is a potent inhibitor of adenylate cyclase and shows promise for congestive heart failure and bronchial asthma. The ginkgolides, from the leaves of *Ginkgo biloba*, are potent inhibitors of platelet activating factor, and in a specified mixture with associated flavonoids, they improve peripheral and cerebrovascular function, hence, their wide use for senile dementia and memory loss. Taxol, originally isolated from the yew *Taxus brevifolia*, is a potent agent against many forms of cancer, including ovarian and breast cancer. It acts by promoting the assembly of microtubules (compare podophyllotoxin, colchicine, and vincristine). Geranylgeranyl units are also found in the side chains of chlorophyll a and vitamin K₁. Many *Euphorbia* species have a latex containing tiglane esters noted for their powerful skin irritant and cocarcinogenic activity. Derivatives of an

ent-kaurene alcohol (stevioside, rebaudioside) are non-cariogenic sweetening agents.

Through a series of cyclizations, squalene oxide (C_{30}) affords lanosterol in animals and fungi and cycloartenol in plants (Fig. 19). In both instances, the intermediate is a protosteryl cation that can also undergo a series of Wagner-Meerwein rearrangements to afford the cytotoxic cucurbitacins of melons and cucumbers. Squalene oxide in a chair-chair-chair-boat conformation yields the dammar-enyl cation, a parent of numerous triterpene skeleta (e.g., lupane, oleanane, ursane, and taraxerane) contained in the saponins found in many foodstuffs, in soaps, and in several drugs from complementary systems of medicine (e.g., ginseng, liquorice, *Bupleurum*, and horsechestnut).

Steroids are degraded triterpene derivatives, and the different nuclei are classified based on carbon number.

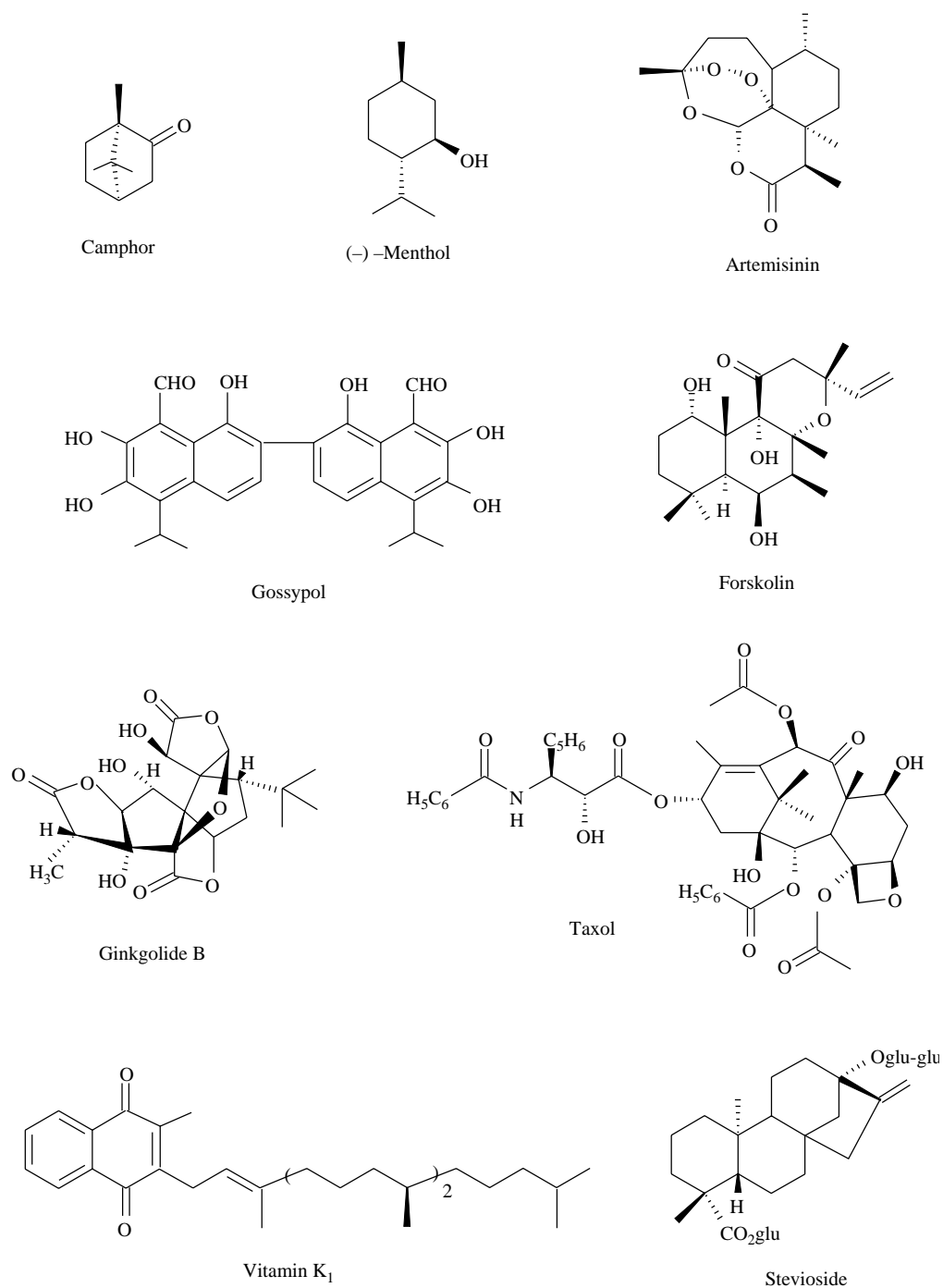


Fig. 18 Representative mono-, sesqui- and diterpene derivatives.

The most significant, from a drug perspective, are the cholane, pregnane, androstane, and estrane systems. When a carbon, such as a methyl group, is lost from the nucleus, the term *nor* is used.

In animals, lanosterol undergoes a series of degradative steps (Fig. 20), whose sequence depends on the organism to afford cholesterol. In photosynthetic organisms, this role is played by cycloartenol where the cyclopropane ring

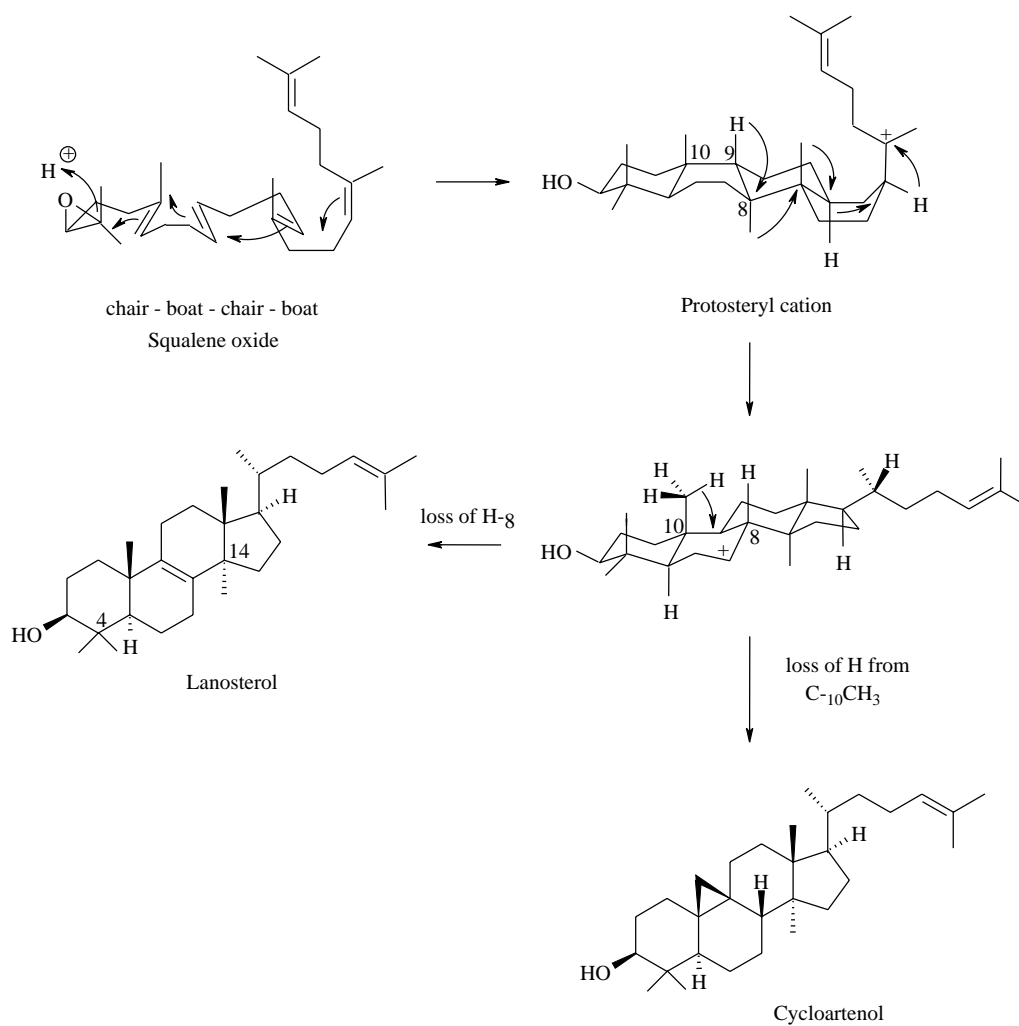


Fig. 19 The biosynthesis of lanosterol and cycloartenol.

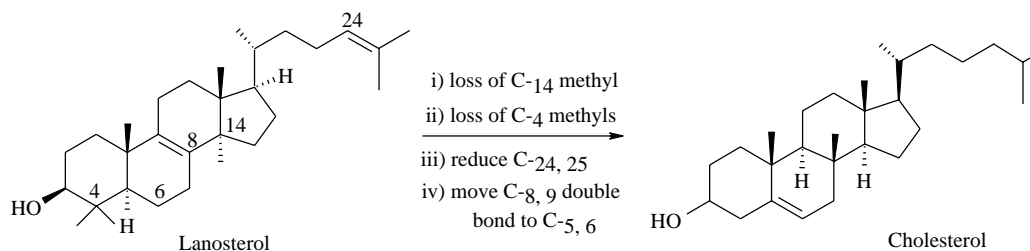


Fig. 20 Steps in the pathway from lanosterol to cholesterol in mammals.

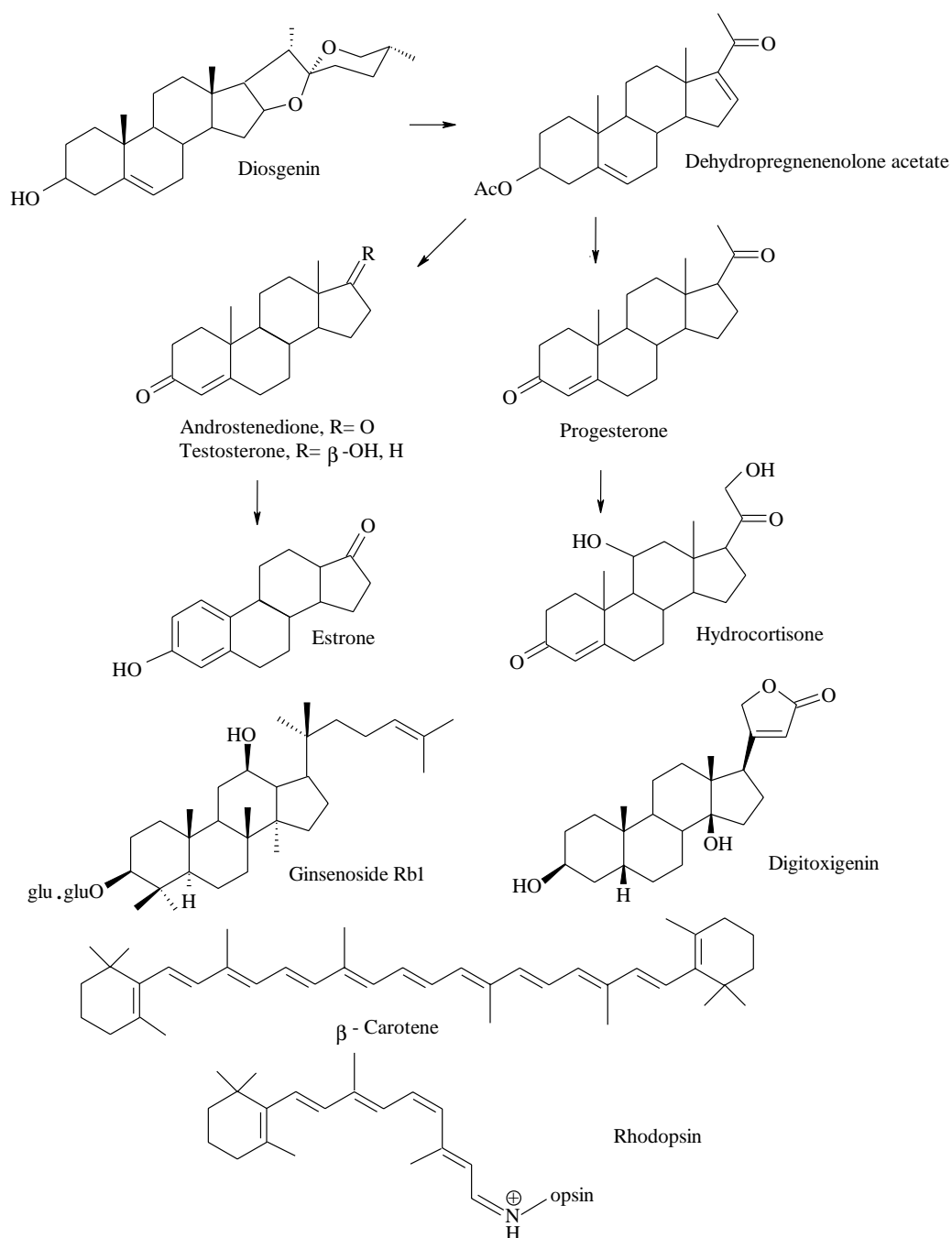


Fig. 21 Representative degraded triterpenes and steroids, and the carotenoids.

is opened. Cholesterol is in almost every animal tissue, and is derived from cattle brains and spinal chords, as well as lanolin from sheep wool. High levels of blood cholesterol are correlated with a high risk of heart disease and atherosclerosis (through deposition of cholesterol and its esters in the artery wall). Thus, an agent that can selectively inhibit an early stage (HMG-CoA reductase) in cholesterol biosynthesis in humans, such as mevinolin (lovastatin), has the effect of reducing serum cholesterol levels.

The steroidal saponins, typically based on a C_{27} sterol nucleus, are distributed in the Dioscoreaceae, the Agavaceae, and the Liliaceae, and have a spiroketal at C-22 (e.g., diosgenin). They are also characterized by numerous sugar units attached at C-3 and sometimes elsewhere. Although not employed as drugs, steroidal saponins are critical for the semisynthesis of important hormones (estrogens, androgens, and progestins) and selected anti-inflammatory agents. An example is the conversion of diosgenin to a dehydropregnenolone acetate for elaboration to the hormones (progesterone, testosterone, androstenedione, and estrone) and the corticosteroids. Microbial transformations are important in several of the reaction sequences. Ginsenosides, the adaptogenic principles of Korean ginseng (*Panax ginseng*), are polysaccharide derivatives of a trihydroxylated (3β , 12β , and $20S$) dammarane nucleus, and sugar variation occurs at C-3 and C-20 (Fig. 21).

Cardiac glycosides, such as those of *Digitalis lanata*, are composed of a polysaccharide unit of three or four sugars, including some 2,6-dideoxyhexoses, linked at C-3 to a modified polyhydroxy ($C-3\beta$, $C-12\beta$, and $C-14\beta$ in the case of digitoxigenin) steroid nucleus. The modification takes place on a 20-keto-pregnane through hydroxylation, the addition of acetate, and cyclization to yield an α,β -unsaturated butyrolactone. Cucurbitacins, withanolides, ecdysones, guggulsterone, limonoids, and quassinoids are also modified triterpene derivatives with potent biological effects, including high cytotoxicity (Fig. 21).

The carotenoids are tetraterpenes and are formed through the tail-to-tail coupling of geranyl pyrophosphate, followed by cyclizations at each terminal (e.g., β -Carotene in carrots). They are widely used as coloring agents for foods, confectionery, and drugs. β -Carotene is under investigation as an antioxidant for the prevention of cancer. Cleavage of β -carotene yields retinol (vitamin A_1 ; Fig. 21). The retinoids are important signalling agents and regulate many aspects of cell differentiation, embryonic development, growth, and vision (rhodopsin is a derivative of 11-*cis*-retinal with the protein opsin).

The steroidal saponins, typically based on a C_{27} sterol nucleus, are distributed in the Dioscoreaceae, the Agavaceae, and the Liliaceae, and have a spiroketal at C-22 (e.g., diosgenin). They are also characterized by numerous sugar units attached at C-3 and sometimes elsewhere. Although not employed as drugs, steroidal saponins are critical for the semisynthesis of important hormones (estrogens, androgens, and progestins) and selected anti-inflammatory agents. An example is the conversion of diosgenin to a dehydropregnenolone acetate for elaboration to the hormones (progesterone, testosterone, androstenedione, and estrone) and the corticosteroids. Microbial transformations are important in several of the reaction sequences. Ginsenosides, the adaptogenic principles of Korean ginseng (*Panax ginseng*), are polysaccharide derivatives of a trihydroxylated (3β , 12β , and $20S$) dammarane nucleus, and sugar variation occurs at C-3 and C-20 (Fig. 21).

Cardiac glycosides, such as those of *Digitalis lanata*, are comprised of a polysaccharide unit of three or four sugars, including some 2,6-dideoxyhexoses, linked at C-3 to a modified polyhydroxy ($C-3\beta$, $C-12\beta$, and $C-14\beta$ in the case of digitoxigenin) steroid nucleus. The modification takes place on a 20-Keto-pregnane through hydroxylation, the addition of acetate, and cyclization to yield an α,β -unsaturated butyrolactone. Cucurbitacins, withanolides, ecdysones, guggulsterone, limonoids, and quassinoids are also modified triterpene derivatives with potent biological effects, including high cytotoxicity (Fig. 21).

The carotenoids are tetraterpenes and are formed through the tail-to-tail coupling of geranyl pyrophosphate, followed by cyclizations at each terminal (e.g., β -carotene in carrots). They are widely used as coloring agents for foods, confectionery, and drugs. β -carotene is under investigation as an antioxidant for the prevention of cancer. Cleavage of β -carotene yields retinol (vitamin A_1 ; Fig. 21). The retinoids are important signalling agents and regulate many aspects of cell differentiation, embryonic development, growth, and vision (rhodopsin is a derivative of 11-*cis*-retinal with the protein opsin).

ALKALOIDS

Alkaloids are nitrogenous secondary metabolites primarily derived from amino acids for both their nitrogen content and a portion of their carbon framework. However, the approximately 27,000 known alkaloids defy a simple definition. Many alkaloids and their derivatives display

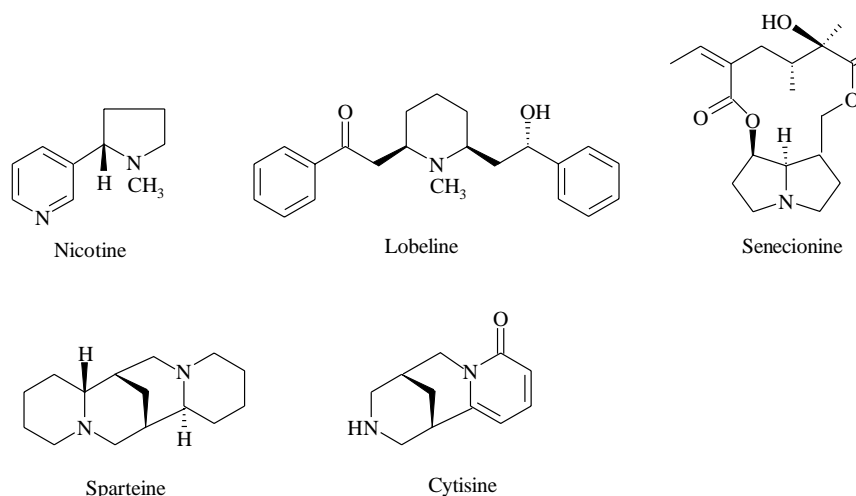


Fig. 22 Representative ornithine- and lysine-derived alkaloids.

profound biological effects and are of enormous commercial, pharmaceutical, and social significance. Ornithine, lysine, phenylalanine, and tryptophan are the principal amino acid precursors.

ALKALOIDS DERIVED FROM ORNITHINE AND LYSINE

There are three principal groups of alkaloids derived from ornithine: nicotine, the pyrrolizidines, and the tropanes, all having significant biological effects. From lysine are derived the piperidine alkaloids (e.g., lobeline, an antismoking agent) and the quinolizidine alkaloids, such as sparteine (an oxytocic agent) and cytisine (a teratogenic agent; Fig. 22). Only the alkaloids derived from ornithine will be discussed here.

Two very different taxa, the Asterales (Solanaceae) and Geraniales (Erythroxylaceae), formulate the tropane nucleus through similar, but distinct, pathways. Using [2-¹⁴C]-ornithine in solanaceous plants, the bridgehead carbon C-1 was labeled, precluding an unbound symmetrical intermediate. In *Atropa belladonna*, ornithine is *N*-methylated prior to decarboxylation to *N*-methylputrescine. By contrast, in *Erythroxylum coca*, the bridgehead carbons C-1 and C-5 were equally labeled, suggesting that an unbound putrescine is methylated. Oxidative deamination affords an aldehyde that undergoes Mannich closure to yield *N*-methyl-pyrrolinium; condensation with acetoacetate affords hygrine-1'-carboxylic acid.

Decarboxylation is followed by oxidative cyclization to tropinone, followed by stereospecific reduction to the α -Hydroxy group, which is esterified to hyoscyamine. The esterifying ester, tropic acid, is an intramolecularly rearranged phenyllactic acid (derived from phenylalanine). Further elaboration of hyoscyamine yields scopolamine (Fig. 23). The enzymes for this transformation are known.

In the biosynthesis of cocaine, the carboxylic acid is retained as the methyl ester, and after stereospecific reduction to afford the β -alcohol, benzylation affords cocaine (Fig. 23). Although cocaine is widely recognized as a drug of abuse, for many populations in South America, the chewing of coca leaves is a routine aspect of the working day and has been for thousands of years.

There are no drugs based on the pyrrolizidine alkaloids of the Asteraceae (e.g., *Senecio* and *Symphytum*) and Boraginaceae (*Crotolaria*). However, these alkaloids pose a great threat to human and animal health because of their potential for inadvertent consumption. In the case of 1,2-dehydro derivatives, such as senecionine, ingestion leads to nonreversible hepatotoxicity. Pyrrolizidine nucleus formation from two units of ornithine is shown (Fig. 24).

ALKALOIDS DERIVED FROM PHENYLALANINE AND TYROSINE

The isoquinoline alkaloids are the second largest group of alkaloids, numbering about 6000, and can be viewed as five

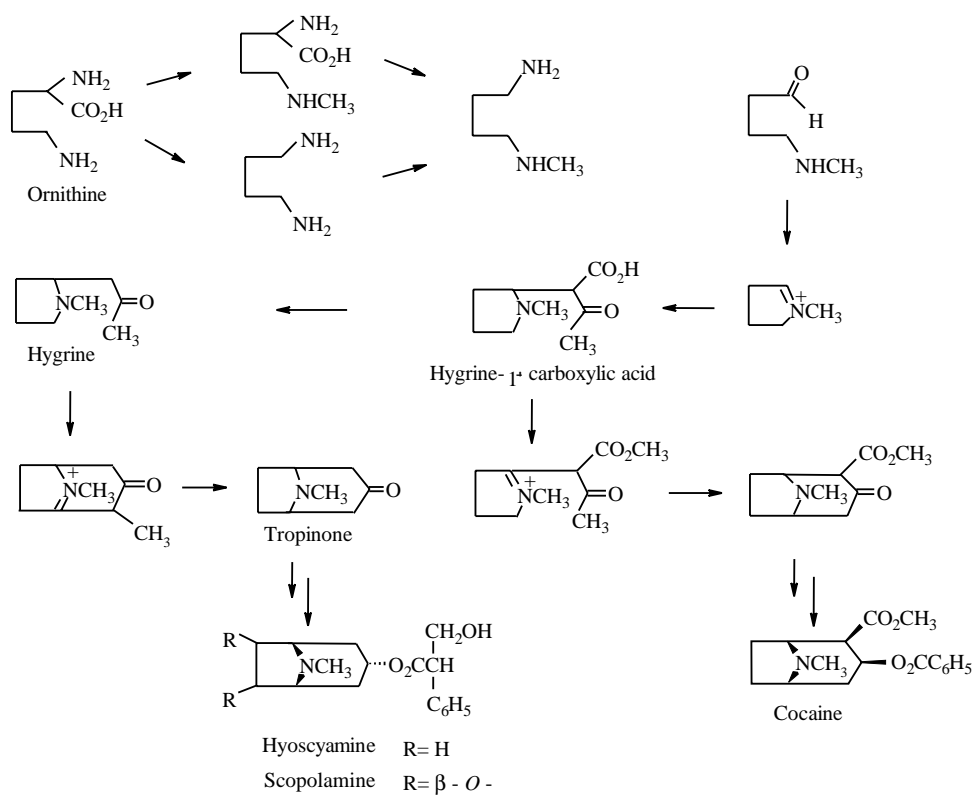


Fig. 23 The biosynthesis of the tropane alkaloids.

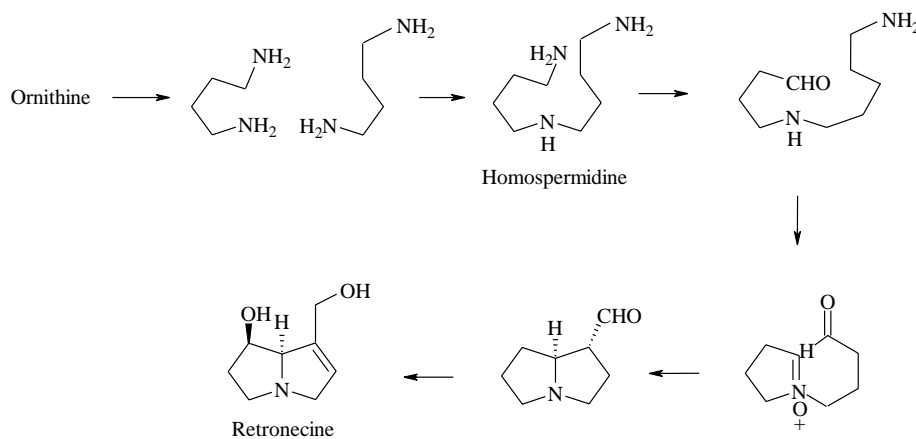


Fig. 24 The biosynthesis of the pyrrolizidine alkaloids.

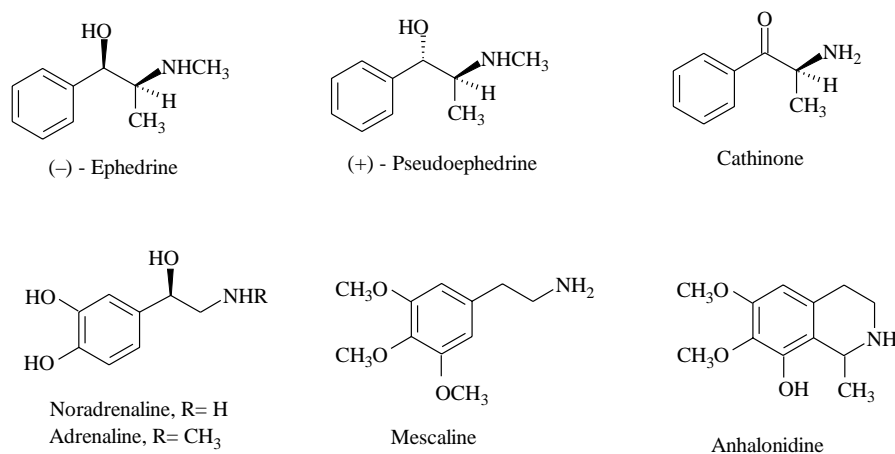


Fig. 25 Simple alkaloids derived from phenylalanine and tyrosine.

subgroups—the simple tetrahydroisoquinolines, the benzyloisoquinolines, the phenethylisoquinolines, the Amaryllidaceae alkaloids, and the monoterpene isoquinolines. In addition, there are a number of simple phenethylamine derivatives, including ephedrine (originally from *Ephedra* species, but now synthesized) and pseudoephedrine, used for asthma and nasal decongestion, respectively. Khat (*Catha edulis*) is widely used as a stimulant in the southeastern Arabian peninsula and contains cathinone. Tyrosine is the precursor of the neurotransmitter noradrenaline and the hormone adrenaline. The hallucinogen mescaline, from the mushroom *Lophophora williamsii*, is also a member of this series (Fig. 25).

An aldehyde, 4-hydroxy-phenylacetaldehyde, operating under the influence of the enzyme norcoclaurine synthase, condenses with dopamine to afford (S)-norcoclaurine,

the progenitor of all benzyloisoquinoline alkaloids (Fig. 26). Robinson, in 1917, was the first to suggest the biogenetic derivation of the pavines, aporphines, morphinans, and protoberberines from a benzyloisoquinoline precursor. These ideas led to a correct proposal for the structure of morphine and were expanded to embrace numerous alkaloid classes (Fig. 27).

Papaverine is one of the few commercial alkaloids synthesized, rather than isolated. It is used either alone for various vascular disorders or in combination as an antispasmodic. The several classes of bisbenzyloisoquinoline alkaloid are based on the number of bridges between the units and their orientation. The alkaloids are common in the Menispermaceae and the Ranunculaceae. One member of the series, tubocurarine, is the prototype for several neuromuscular blocking agents; an activity based

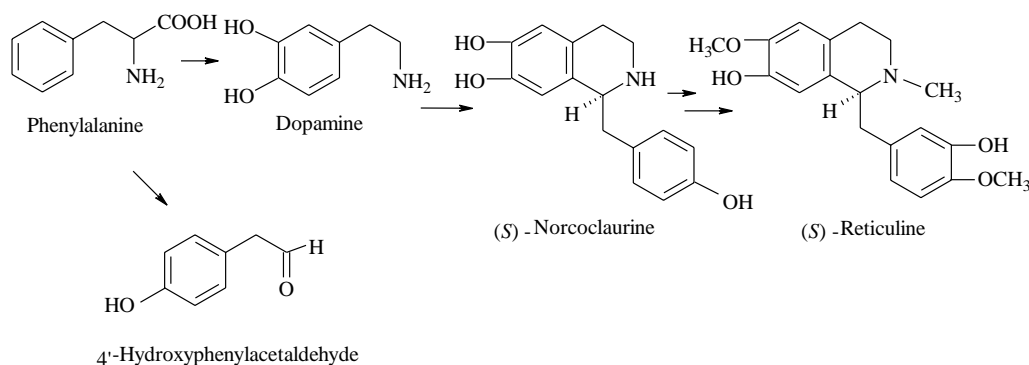


Fig. 26 The biosynthesis of (S)-norcoclaurine and (S)-reticuline.

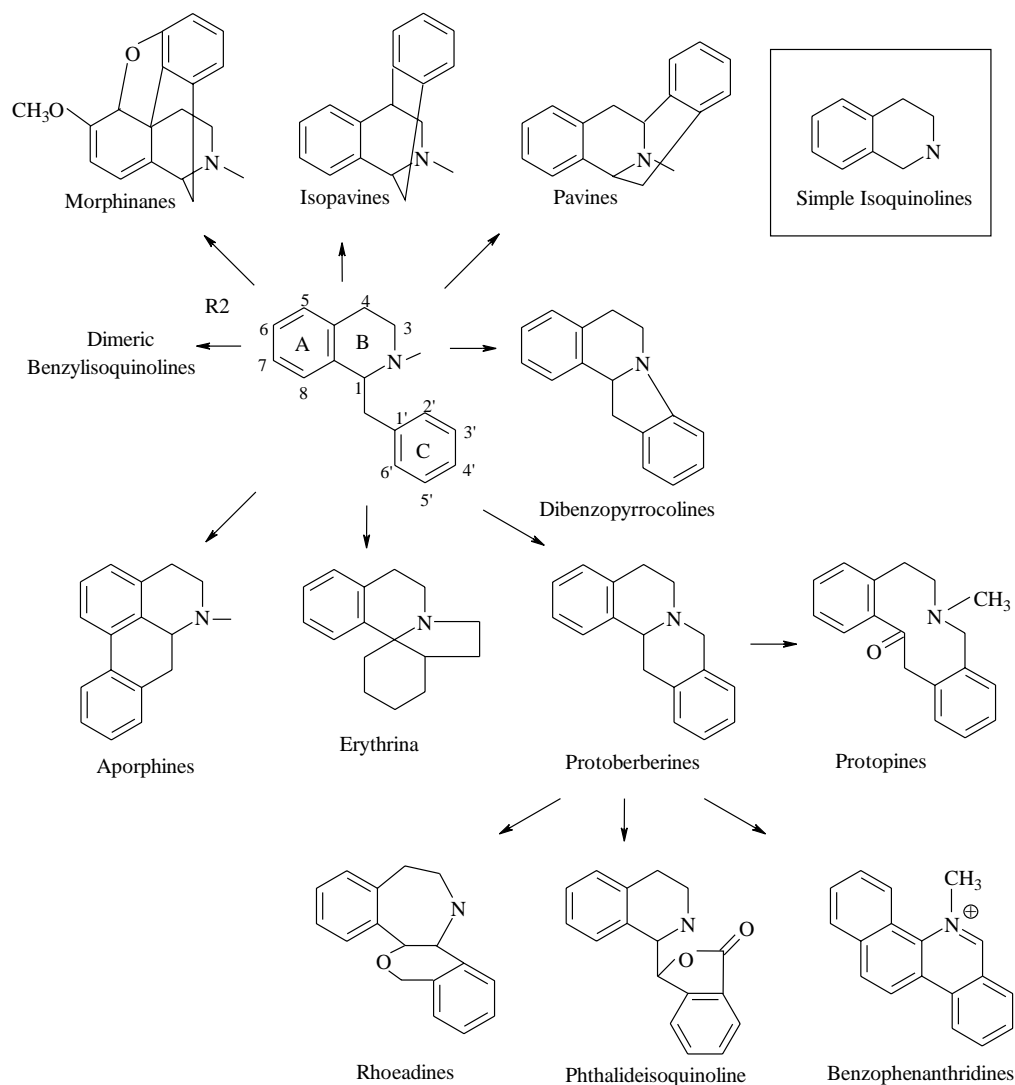


Fig. 27 Biosynthetic interrelationships of the major classes of benzyloquinoline alkaloids.

on the ethnobotanical use of the tube curares (*Chondrodendron* species) is its use as arrow poisons in the upper Amazon.

Phenol oxidative coupling was proposed by Barton and Cohen in 1957 to enumerate the relationships between many of the benzyloquinoline alkaloid groups (Fig. 27) and accounts for the importance of reticuline as a precursor. In the case of berberine, phenolic oxidative coupling occurs between the *ortho* benzylic position and the *N*-methyl carbon (berberine bridge) to afford scoulerine (Fig. 28). Many of the enzymes in this pathway have been elucidated by Zenk and coworkers. The protoberberines themselves are the precursors of several

groups of alkaloids, including the protopines, the benzophenanthridines (e.g., sanguinarine, used as an antiplaque agent) and the phthalideisoquinolines (e.g., β -Hydrastine, a constituent of *Hydrastis canadensis*; Fig. 29). Berberine is a widely used antimicrobial agent, being active against *Staphylococcus*, *Streptococcus*, *Proteus*, *Vibrio*, etc. (Fig. 28).

Morphine and related alkaloids are specific to the genus *Papaver* (Berberidaceae), although the antipodal series of alkaloids is distributed in the Menispermaceae. Early in the biosynthesis of morphine, an inversion at C-1 of (*S*)-reticuline occurs, followed by *ortho*-*para* benzylic coupling to afford salutaridine. Stereospecific

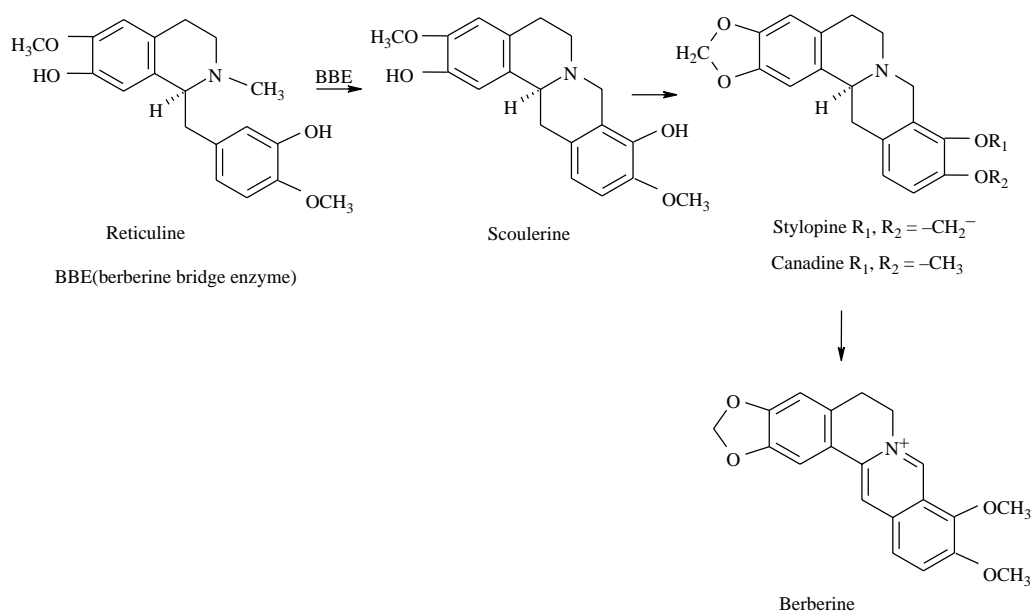


Fig. 28 The biosynthesis of protoberberine alkaloids.

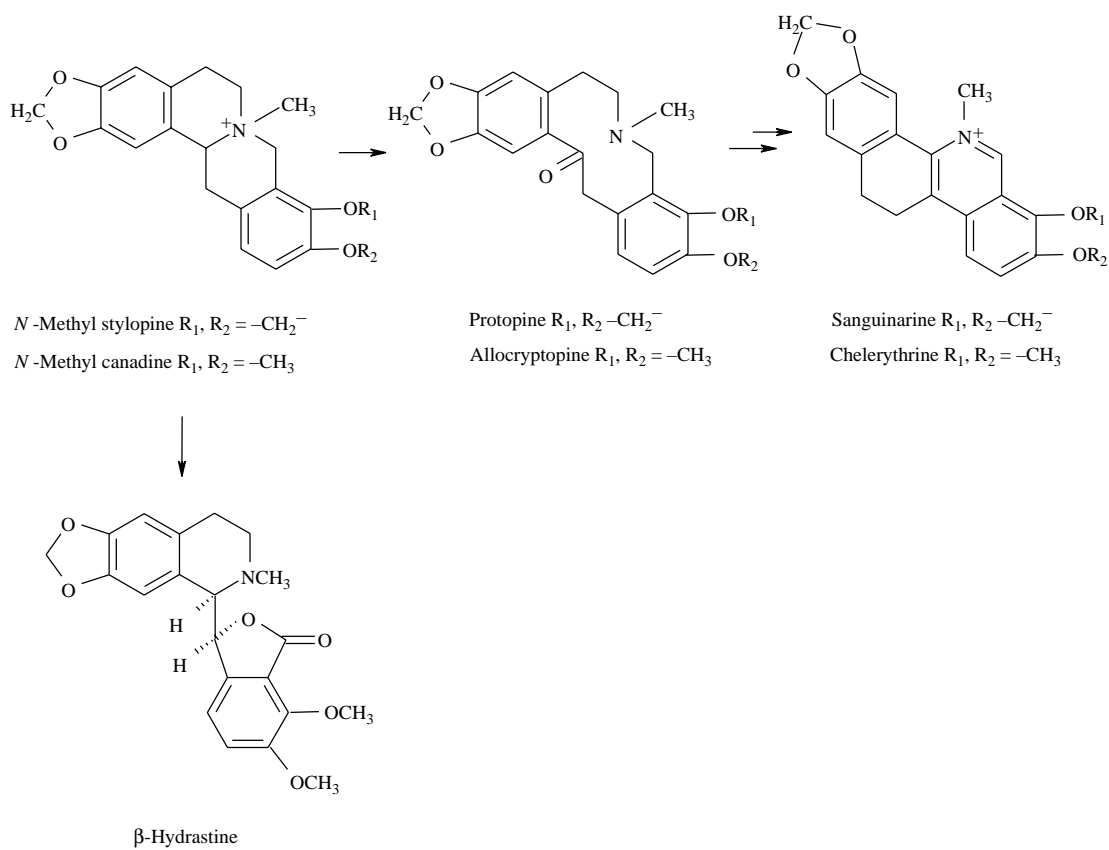
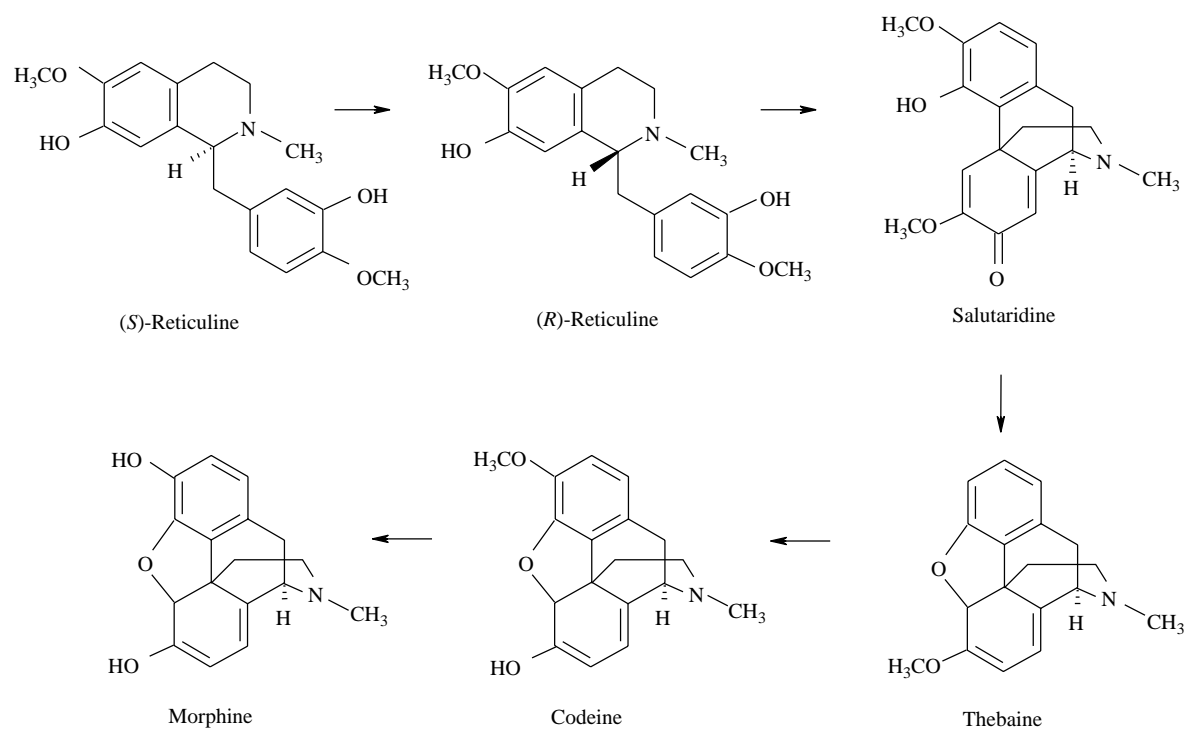
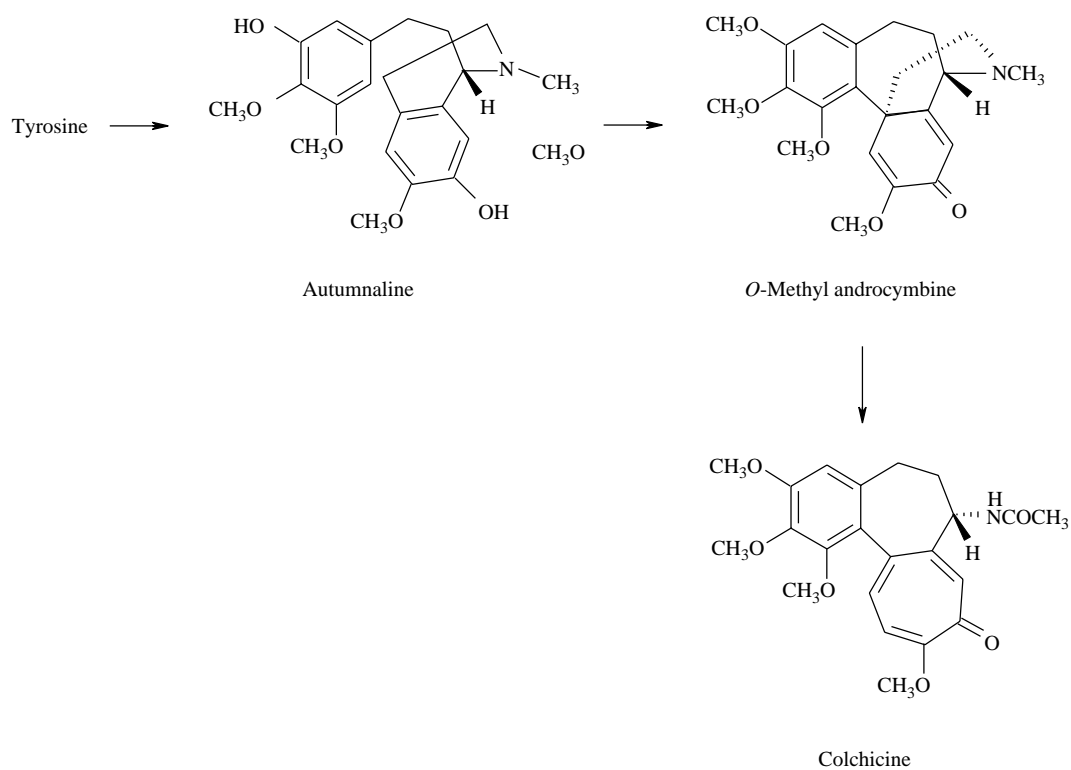


Fig. 29 The biosynthesis of β -hydrastine and sanguinarine.

**Fig. 30** The biosynthesis of morphine.**Fig. 31** The biosynthesis of colchicine.

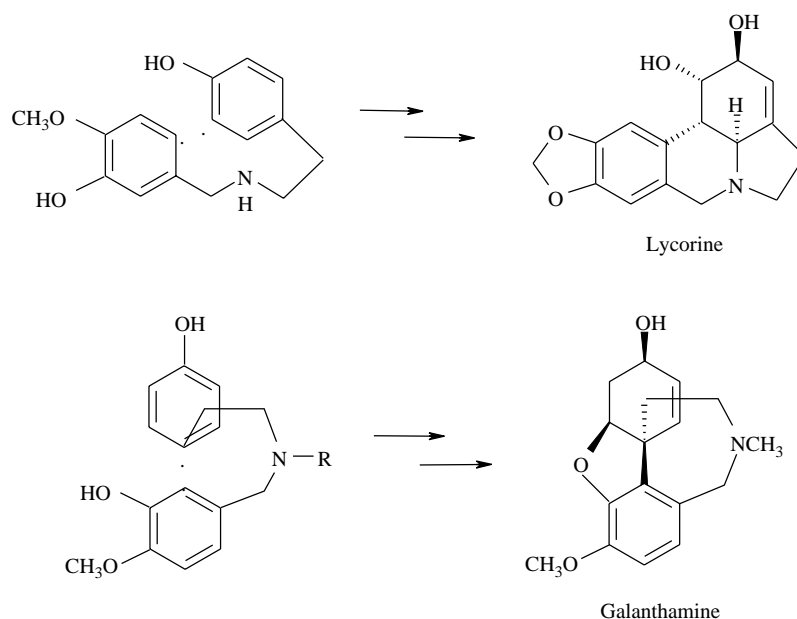


Fig. 32 The biosynthesis of some Amaryllidaceae alkaloids.

reduction and cyclization-elimination affords the 4,5-Ether bridge and thebaine. The dominant pathway from this point involves neopinone, codeinone, codeine, and morphine. Again, most of the enzymes in this sequence were isolated and characterized by Zenk's group (Fig. 30).

Morphine binds with very high affinity to several receptors in the CNS and is a potent analgesic and central depressant. It is also the prototype for many semisynthetic derivatives (e.g., naloxone, oxycodone, ethorphine, nalbuphine, and buprenorphine) with various degrees of analgesic and narcotic properties. Heroin is diacetyl morphine.

Colchicine is a phenethylisoquinoline alkaloid from the autumn crocus, *Colchicum autumnale* (Liliaceae), a plant used since at least the fifth century to treat gout. *Gloriosa superba* (Liliaceae) is an alternative source. Colchicine inhibits microtubule formation at a specific site, and at a dose of 10 mg, it causes fatal respiratory arrest and renal insufficiency. It is used for the prevention and treatment of gout. The biosynthesis remains unclear; autumnaline has been proposed as an intermediate which undergoes *para-para'* coupling to afford *O*-methyl androcymbine. Hydroxylation, cyclopropane ring formation, and ring expansion affords colchicine (Fig. 31).

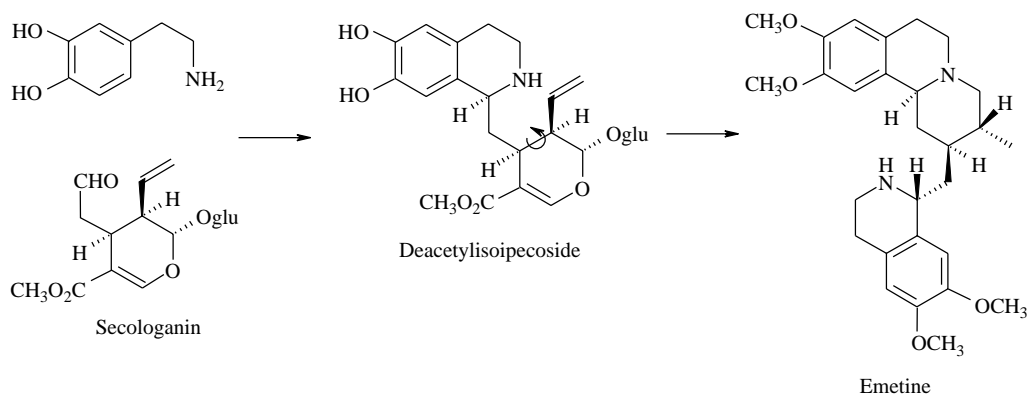


Fig. 33 The biosynthesis of emetine.

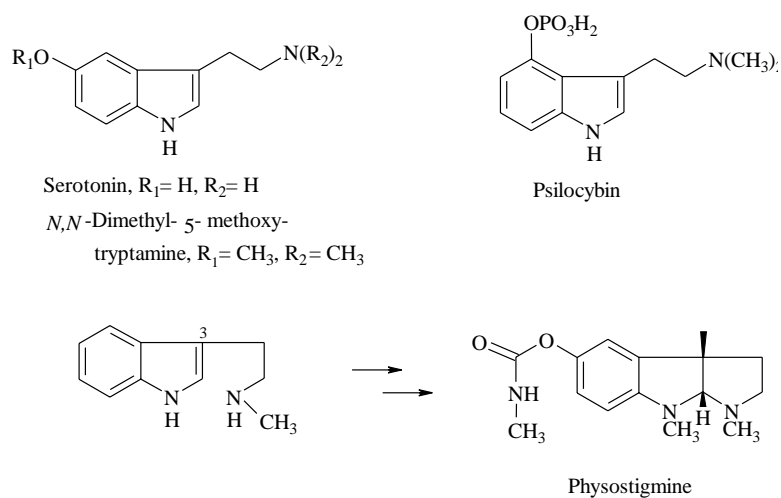


Fig. 34 Simple alkaloids derived from tryptophan.

The Amaryllidaceae alkaloids (e.g., lycorine) are derived from the phenol oxidative coupling of a $\text{C}_6\text{C}_2\text{NC}_6\text{C}_1$ unit. One unit ($\text{C}_6\text{C}_2\text{N}$) is derived from

tyrosine, whereas the other (C_6C_1) is projected to be derived from phenylalanine, followed by oxidative deamination, hydroxylation, and cleavage of a two-carbon

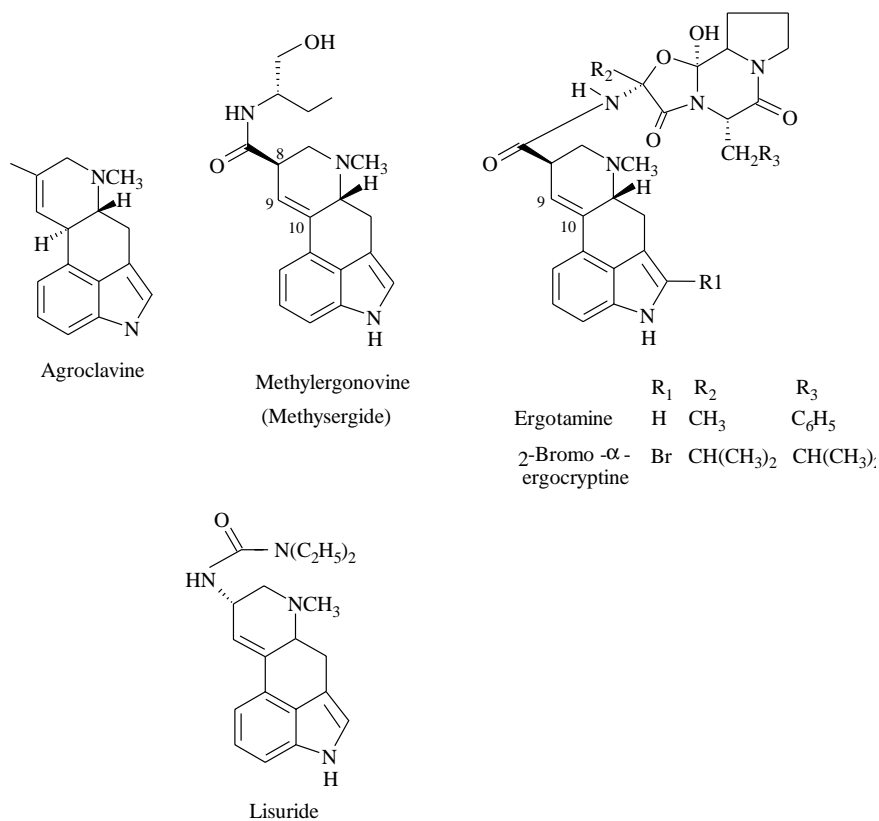


Fig. 35 Representative ergot alkaloids.

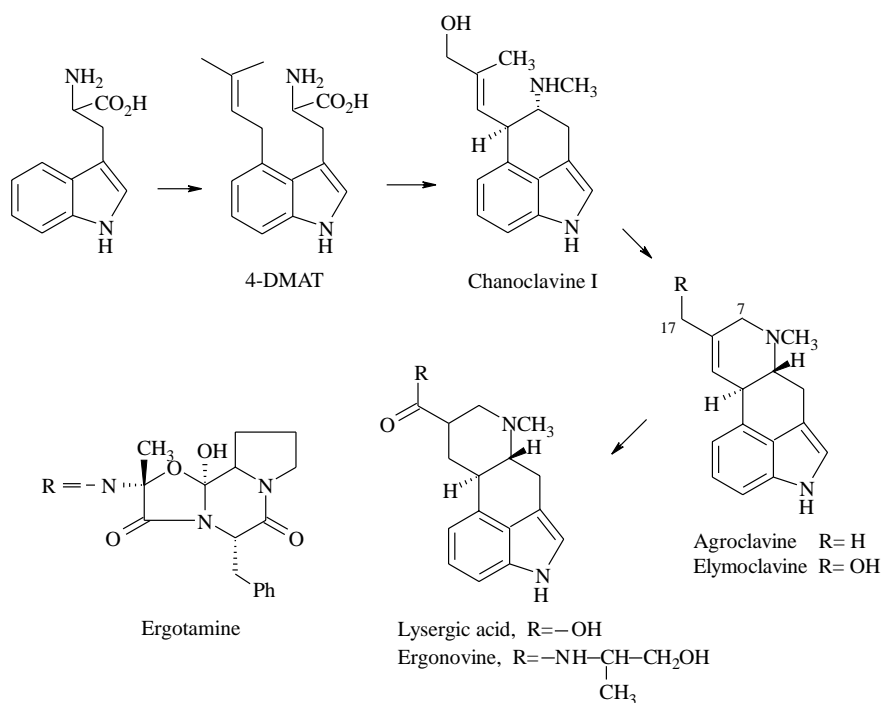


Fig. 36 The biosynthesis of the ergot alkaloids.

unit to afford a dihydroxylated benzaldehyde. *Para-ortho'* coupling leads to lycorine, whereas *para-para'* coupling affords galanthamine, of interest as a cholinesterase inhibitor (Fig. 32).

The monoterpene isoquinoline alkaloids are constituents of the genus *Cephaelis* and selected other Rubiaceae

species. *C. ipecacuanha* (ipecac) is a powerful emetic whose active principle is emetine, derived through the condensation of dopamine and secologanin (Fig. 33). Emetine is also a powerful amebicide, antiviral, and inhibitor of protein synthesis. It is now largely replaced by synthetic dehydroemetine.

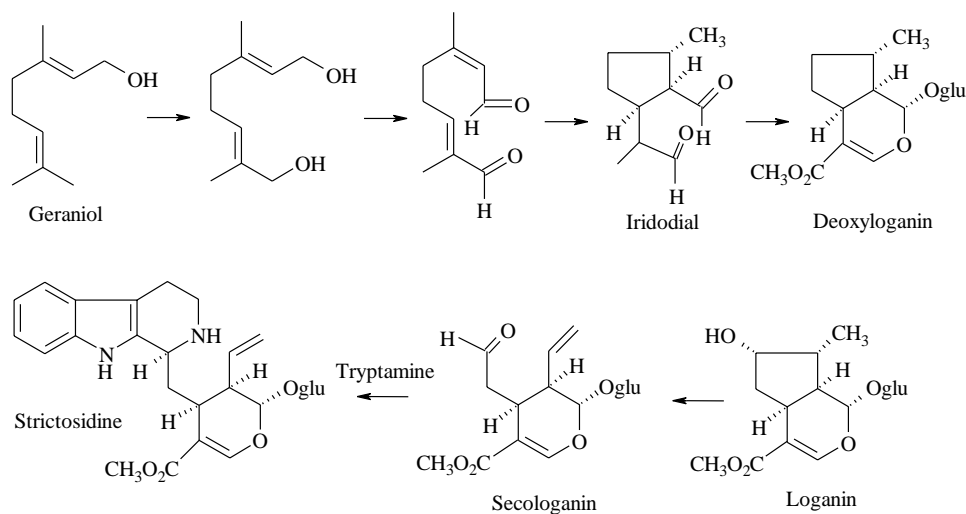


Fig. 37 The biosynthesis of strictosidine.

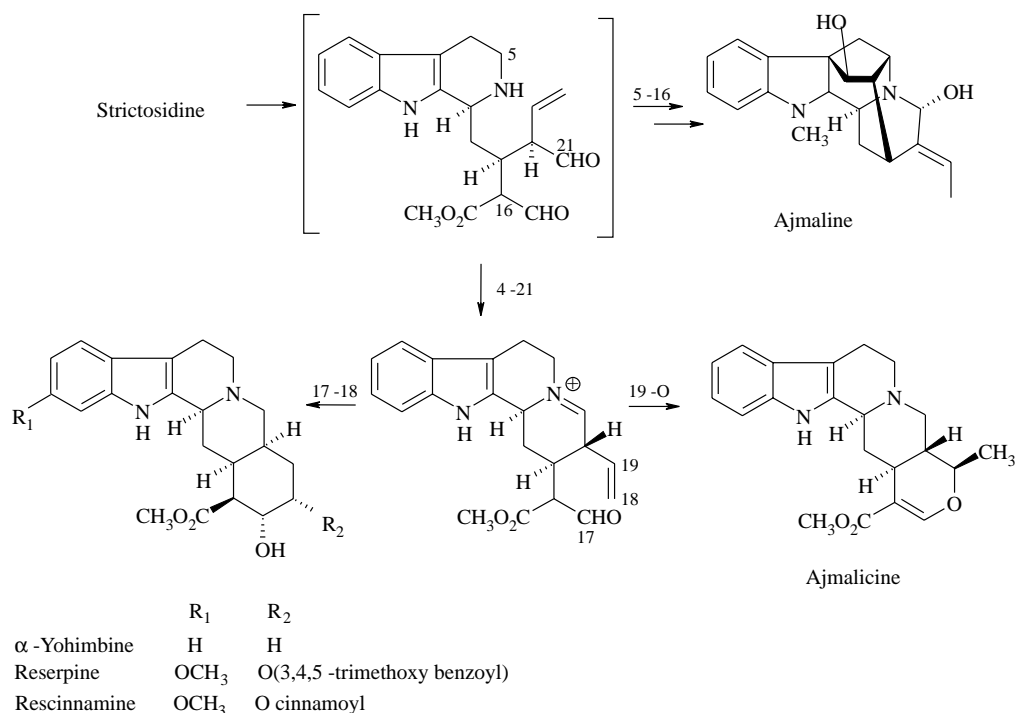


Fig. 38 The biosynthesis of ajmaline, ajmalicine, α-yohimbine, reserpine, and rescinnamine.

ALKALOIDS DERIVED FROM TRYPTOPHAN

Numerous clinically important alkaloids are also derived from-tryptophan and are frequently referred to as “indole alkaloids.” They range in molecular complexity from the mammalian hormone serotonin to the complex bisindolic anticancer alkaloid vincristine. In addition, a number of indole alkaloids, particularly those in the carbazole series, are found in peppers (*Murraya* sp.), and simple derivatives of tryptamine (e.g., *N,N*-dimethyl-5-methoxy-tryptamine and psilocybin) are found in several sources (e.g., snuffs, mushrooms, and toad skins) with attributed hallucinogenic properties.

Physostigmine, under investigation for potential use in Alzheimer’s disease, is possibly formed from *N_b*-methyltryptamine through a radical mechanism involving C-3 methylation and concomitant C-2 cyclization, followed by *N_b*-methylation (Fig. 34).

The powerful biological effects of ergot have been known for over 1000 years. Ergot is a fungus (*Claviceps purpurea* and species) that grows parasitically on rye and some other grains, and it produces three types of ergot alkaloids—the clavines, the simple lysergic acid-derived, and the peptide lysergic acid-derived alkaloids. The

alkaloids find therapeutic use as agents for postpartum hemorrhage (methylergonovine) and as vasoconstrictors and vasoregulators (ergotamine and 9,10-dihydroergotamine) for migraine headaches. More elaborate synthetic derivatives are also available, including lisuride (for Parkinsonism) and 2-bromo-α-ergocryptine (for prolactin-secreting adenomas and Parkinsonism; Fig. 35).

The biosynthetic pathway to the ergoline nucleus proceeds through 4-dimethylallyl tryptophan (4-DMAT), chanoclavine-I, agroclavine, and lysergic acid. Two *cis*, *trans* isomerizations occur: one before chanoclavine-I and the other before agroclavine, as shown by experiments with [2-¹⁴C]-mevalonic acid and [Z-CH₃]-4-DMAT (Fig. 36). The peptide unit is derived from a combination of three amino acids, one of which is always proline. Several genera in the plant family Convolvulaceae (*Rivea*, *Ipomoea*, etc.) also produce ergot alkaloids.

The largest group of alkaloids is the monoterpene indole alkaloids, distributed in the Apocynaceae (mutual exclusion with cardenolides) and in the Rubiaceae and Loganiaceae. The molecular acrobatics of the various systems derived from deglycosylation of the primordial alkaloid strictosidine accounts for this stunning structural diversity.

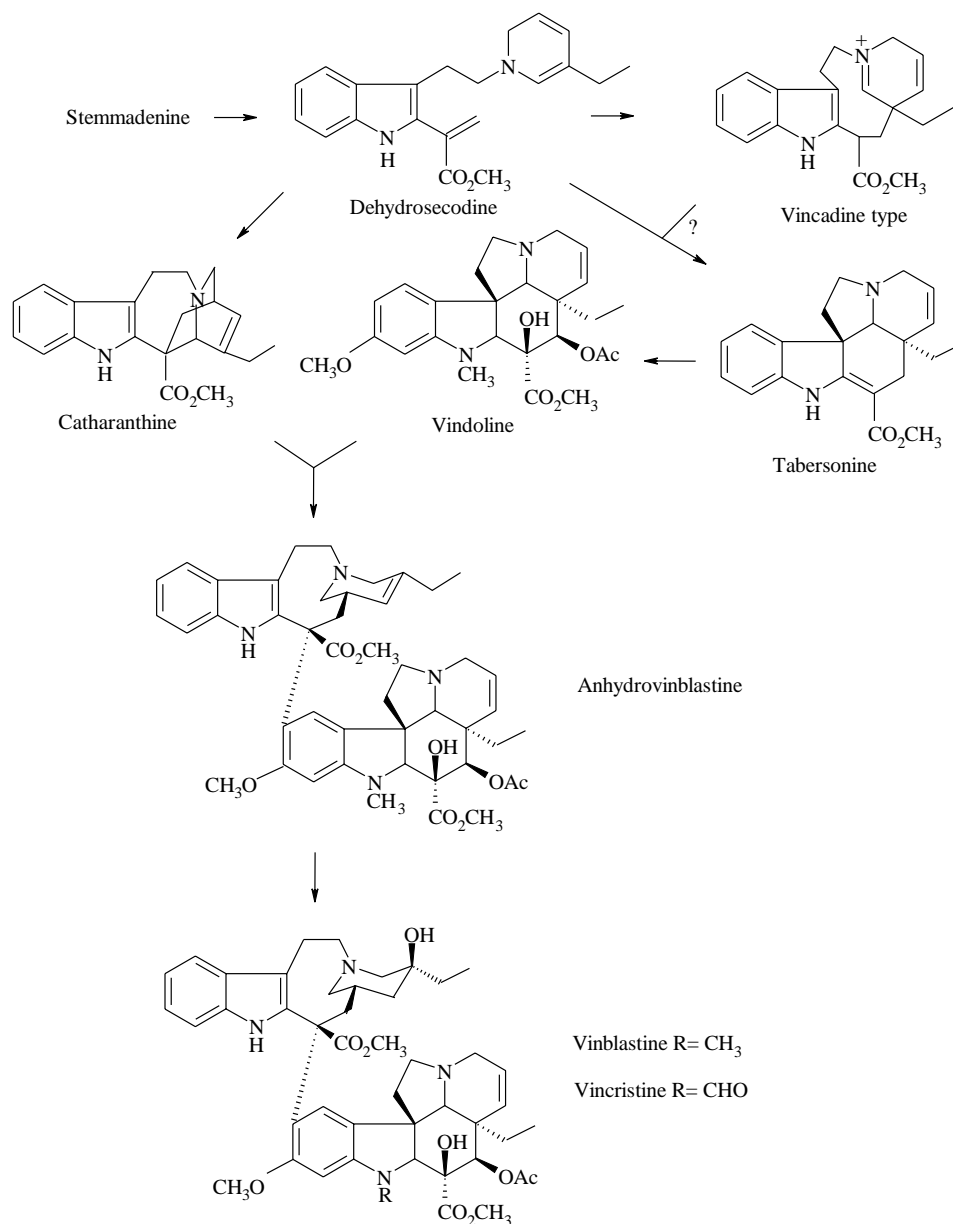


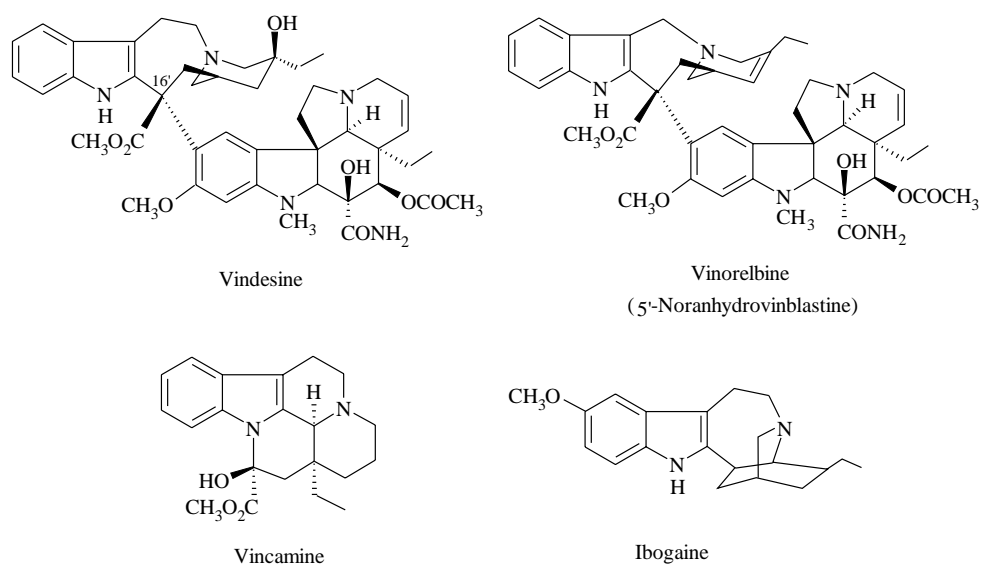
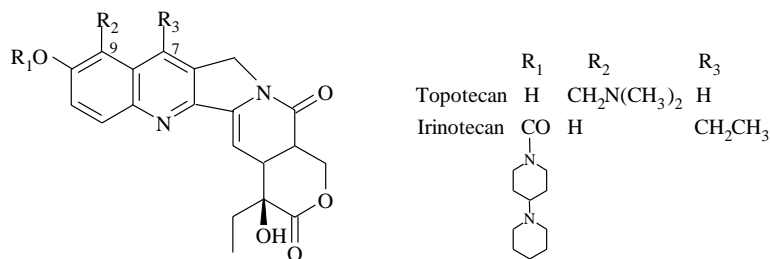
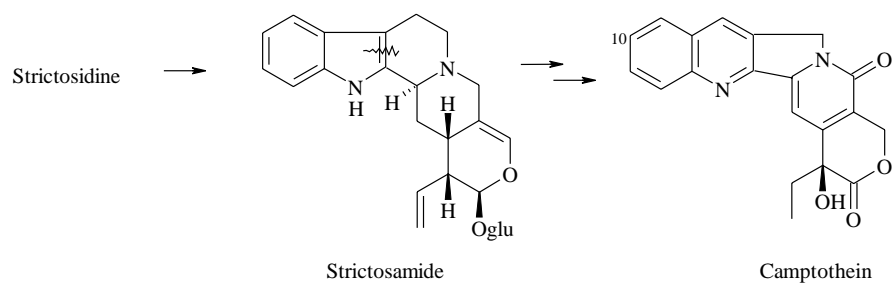
Fig. 39 The biosynthesis of vinblastine.

Strictosidine is produced, stereospecifically, from tryptamine and secologanin by strictosidine synthase, isolated from several species producing monoterpene indole alkaloids. The enzyme was cloned and can be expressed in large quantity (Fig. 37).

After deglycosylation, the pathway proceeds through a 4,21-Dehydrogeissoschizine derivative to ajmalicine (an α -Blocking spasmolytic agent, used for tinnitus and cranial trauma with an ergot derivative). If cyclization occurs between C-17 and C-18, the yohimbine nucleus is

produced, whose derivatives include the *Rauvolfia* alkaloids reserpine and rescinnamine (antihypertensive activity). Ajmaline, formerly used as an antiarrhythmic, also occurs in *Rauvolfia* species, and several of the enzymes in the pathway have been isolated. Recent considerations suggest that the C-16–C-5 bond may be formed before the N-4–C-21 bond (Fig. 38).

The subsequent steps from geissoschizine to form the *Strychnos* alkaloids, the secodines, the *Aspidosperma* alkaloids and the iboga alkaloids remain speculative, based

**Fig. 40** Representative advanced indole alkaloids.**Fig. 41** The biosynthesis of camptothecin.

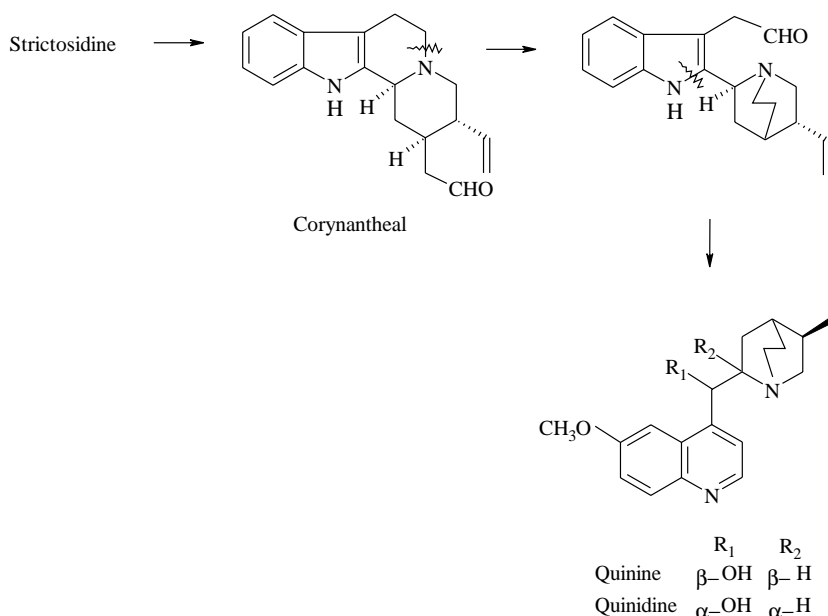


Fig. 42 The biosynthesis of quinine and quinidine.

on low levels of incorporation of early precursors or alkaloid time course studies. Joining C-2 and C-16 while moving C-3 to C-7, yields the strychnan skeleton of strychnine (lethal dose 0.2 mg/kg), still used as a rodenticide in some countries.

If the C-15,C-16 bond is oxidatively cleaved, the secodine skeleton results (the proposed progenitor of the

Aspidosperma and the iboga systems) through alternative Diels–Alder type cyclizations to afford tabersonine and catharanthine. The bisindole alkaloids of *Catharanthus roseus* reflect the union of vindoline and catharanthine to afford anhydrovinblastine; modification affords the clinically significant alkaloids, vinblastine (VLB) and vincristine (VCR; Fig. 39). The alkaloids, particularly VCR, are important as anticancer agents and have led to the development of the semisynthetic derivatives vindesine and vinorelbine (Fig. 40). Synthetic approaches are available to join the monomeric precursors. The enzymatically controlled sequence of reactions from tabersonine to vindoline has been elucidated.

Through a biomimetic approach, tabersonine is also the semisynthetic precursor of vincamine, a Eburna alkaloid isolated from *Vinca minor*, and is used for cerebral insufficiency in Europe. *Tabernanthe iboga* has a long history of use as a stimulant in tropical Africa; its main active principle is ibogaine, a controlled substance in many countries (Fig. 40). It is being actively investigated in the United States for its potential to induce opium addiction withdrawal.

Camptothecin, a quinoline alkaloid from *Camptotheca acuminata* (Nyssaceae), is derived from strictosidine through strictosamide (Fig. 41). Originally isolated in 1966, it biologically inhibits topoisomerase I, and in 1996, two derivatives, topotecan and irinotecan, were approved for the treatment of ovarian cancer and

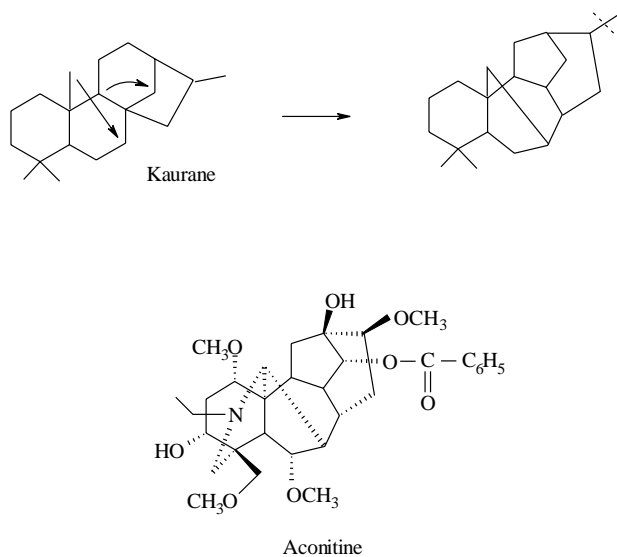


Fig. 43 The biogenesis of aconitine.

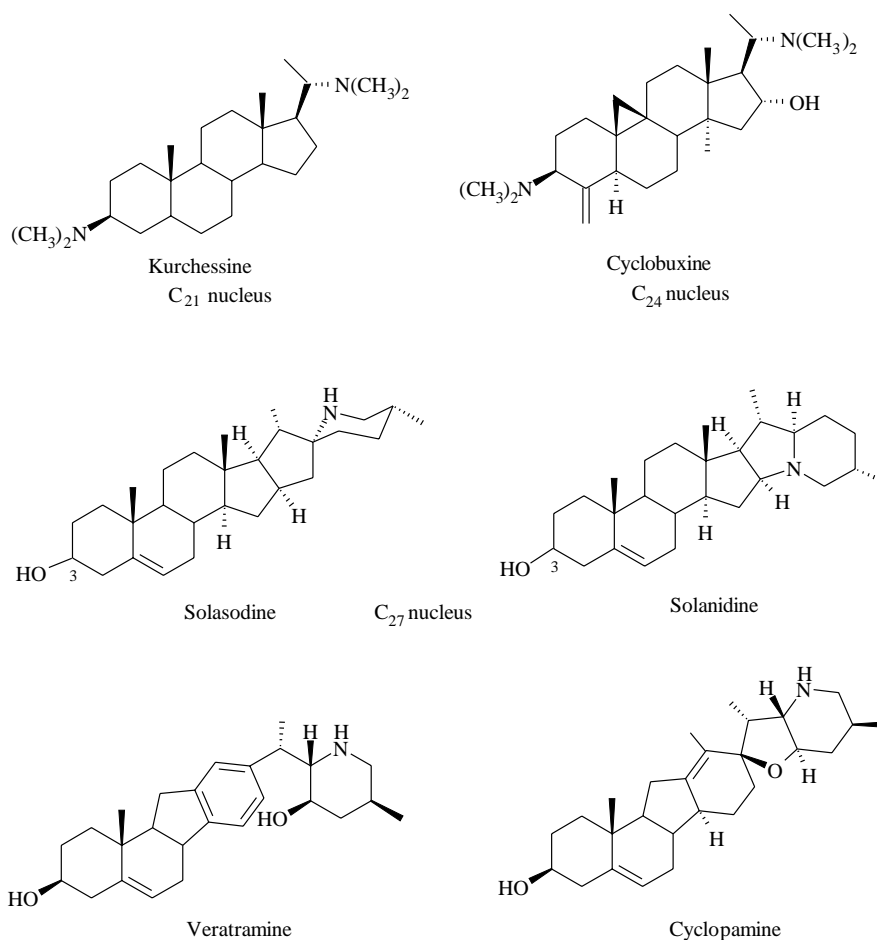


Fig. 44 Representative steroid alkaloids.

colon cancer, respectively. Other derivatives are in clinical trial.

Cinchona species (Rubiaceae) are sources of quinine and quinidine, containing a quinoline nucleus and derived through the extensive elaboration of strictosidine (Fig. 42). The intriguing history of the antimalarial quinine and its role in world politics over the past 350 years are legendary. It is frequently the only antimalarial drug to which patients are not resistant. Its widest use, however, is in the beverage industry in tonic water. Quinidine, an isomer of quinine, is used to treat cardiac arrhythmias.

ALKALOIDS BASED ON A TERPENE SKELETON

Although a variety of monoterpene alkaloids and some sesquiterpene alkaloids are known, they are of little

biological interest. By contrast, the diterpene alkaloids of the Ranunculaceae, for example from *Aconitum* and *Delphinium* species, have profound biological effects. The principal alkaloids of interest are those related to aconitine, acting by exciting and paralyzing peripheral nerve endings. The plants are some of the most toxic known, with merely 10 g of aconite root being lethal. Detoxified root preparations are used as drugs in several major systems of traditional medicine. Formation of the aconitine nucleus is thought to occur through a rearrangement of the kaurane skeleton (Fig. 43).

The steroidal alkaloids have a nucleus based on 21, 24, or 27 carbon atoms (Fig. 44). The C_{21} alkaloids are pregnane-derived with nitrogen inserted at C-3, at C-20, or at both positions. They are characteristic of the Apocynaceae (*Funtumia* and *Holarrhena* species) and the Buxaceae (*Buxus* species). The Buxaceae also produces C_{24} alkaloids based on the cycloartane skeleton.

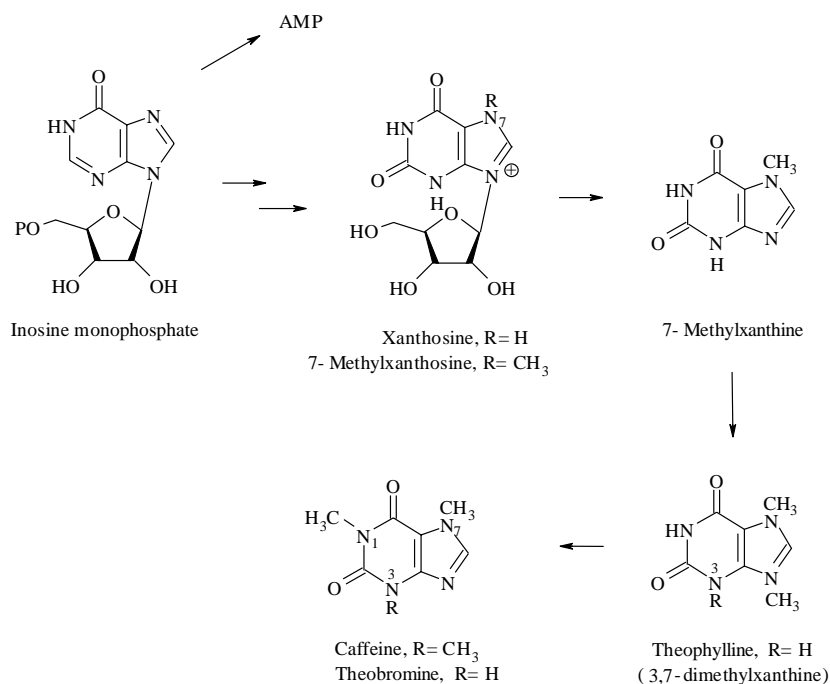


Fig. 45 The biosynthesis of purine alkaloids.

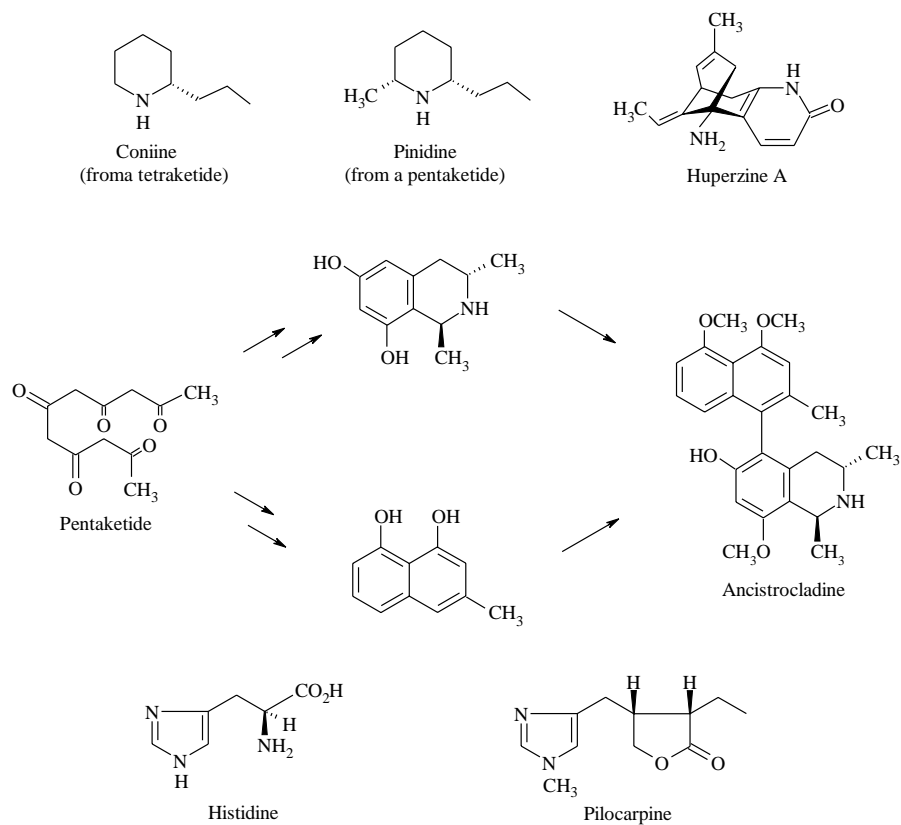


Fig. 46 Some alkaloids derived from acetate and histidine.

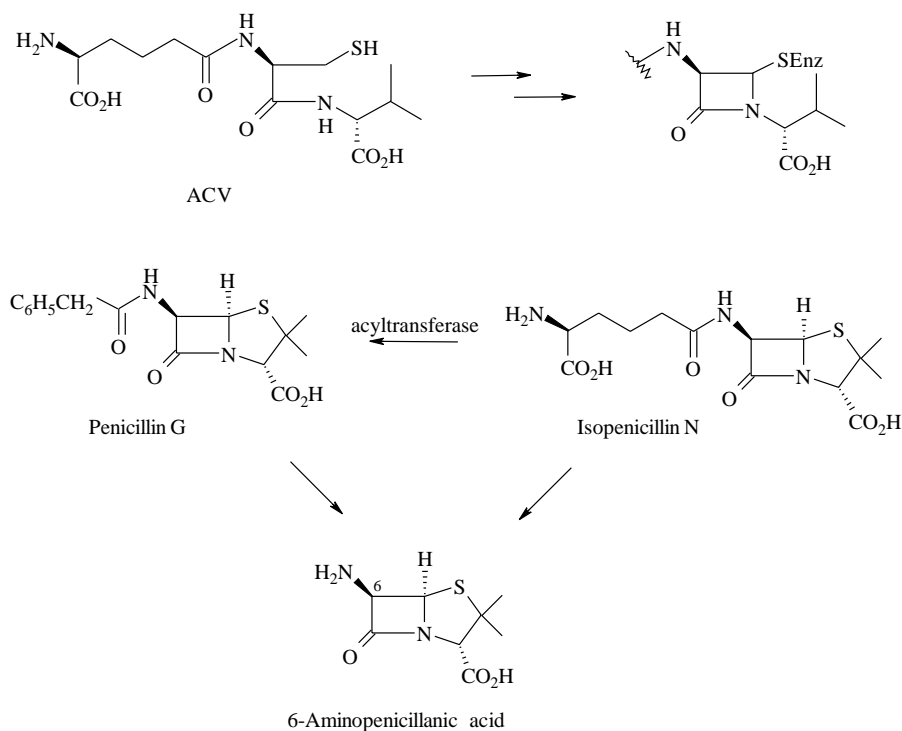


Fig. 47 The biosynthesis of penicillin G.

The most interesting alkaloids are those in the Solanaceae and the Liliaceae. These are C_{27} alkaloids, and examples include solasodine and solanidine; many derivatives are glycosylated. The alkaloids from the Liliaceae, such as veratramine of the white hellebore (*Veratrum album*),

were formerly used for cardiac insufficiency. Other alkaloids, for example cyclopamine, are potent teratogens. Biogenetically, they are derived through nitrogen insertion at C-22, followed by a rearrangement generating a C-*nor*-D-*homo* steroid nucleus (Fig. 44). They are not used

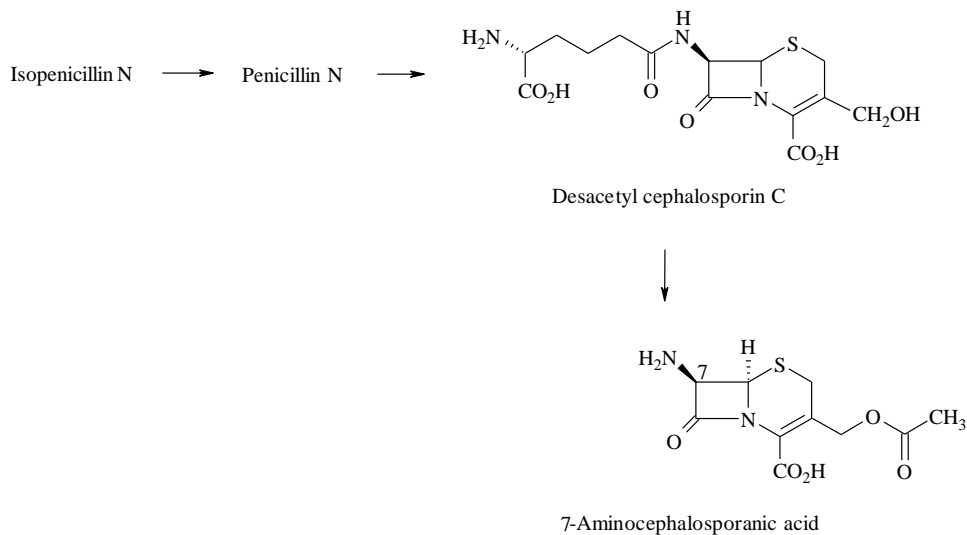


Fig. 48 The biosynthesis of desacetylcephalosporin C.

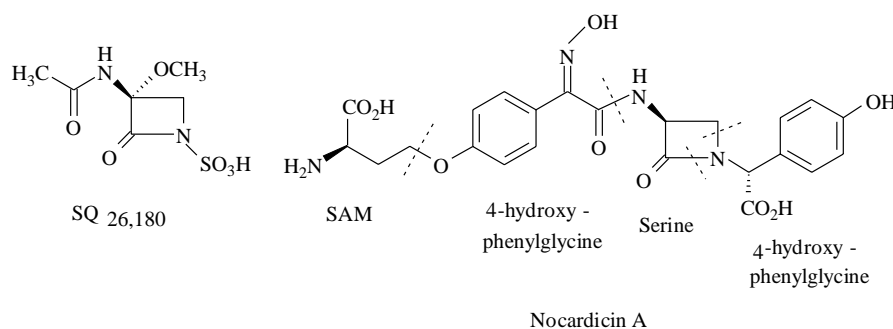


Fig. 49 The biogenesis of nocardicin A.

therapeutically, but the glycoalkaloids of *Solanum tuberosum* (potato) are very toxic and are not destroyed in cooking.

ALKALOIDS DERIVED FROM A NUCLEOTIDE PRECURSOR

Caffeine is one of most widely consumed alkaloids on a daily basis. As well as being a significant constituent of coffee (*Coffea arabica*) and tea (*Camellia sinensis*), caffeine is also present in kola (*Kola* species), guarana (*Paullinia cupana*), and maté (*Ilex paraguariensis*). All of these species are used in various parts of the world to produce beverages that reduce fatigue.

Although some steps remain to be fully elucidated, caffeine is probably derived through a pathway beginning with inosine 5'-monophosphate and proceeds through xanthosine, 7-methylxanthosine, 7-methylxanthine, 3,7-dimethylxanthine (theobromine) to caffeine (Fig. 45). The final methyltransferase has been characterized in coffee and tea, whereas the methylation of xanthosine has only been studied in tea.

Caffeine is noted for its ability to stimulate the CNS and also has positive inotropic and mild diuretic activity. Theophylline (1,7-Dimethylxanthine) is noted for its smooth muscle relaxant activity and its use for chronic asthma.

ALKALOIDS DERIVED FROM OTHER PRECURSORS

Acetate is also a precursor of several groups of alkaloids in the form of a polyketide chain that interacts with an

unknown nitrogen source (as in the terpene alkaloids). Examples of acetate-derived alkaloids are coniine—the toxic principle of *Conium maculatum*, pinidine—from several *Pinus* species, and the naphthylisoquinoline alkaloids (e.g., ancistrocladine)—showing antimalarial and anti-HIV activity. The latter alkaloids are apparently derived from the oxidative coupling of two pentaketide units. Huperzine A, currently in clinical trials for the treatment of Alzheimer's disease and isolated from the club moss (*Serrata huperzia*), is derived from a polyacetate precursor (Fig. 46).

Histidine is a precursor of a very limited number of alkaloids. The most well-known is pilocarpine from *Pilocarpus jaborandi*. The plant was formerly used as a truth serum (diaphoretic activity), and the alkaloid is used to counter the mydriatic effects of atropine.

The penicillins, from the fungus *Penicillium chrysogenum*, are the oldest and most widely used antibiotics. They are formed through stepwise build-up from a tripeptide (ACV) derived from α -Amino adipic acid, cysteine, and valine. Successive oxidation steps form the β -lactam and close the thiazolidine ring to form isopenicillin N. Action of an acyltransferase then yields penicillin G (Fig. 47). Alternatively, hydrolysis of isopenicillin N (or penicillin G) yields 6-aminopenicillanic acid, a key precursor for the wide range of semisynthetic penicillins used therapeutically.

Cephalosporins are modified penicillin derivatives produced by *Cephalosporium acremonium*, wherein isomerization occurs to afford penicillin N followed by a ring expansion involving one of the methyl groups and hydroxylation to produce descetylcephalosporin C (Fig. 48). Once again, removal of the acylating side chain to afford 7-Aminocephalosporanic acid was the key to generating the diversity of available cephalosporin derivatives. The monobactams, such as SQ26,180 from *Chromobacterium violaceum*, contain a 3-Methoxy group

and an *N*-Sulphonate moiety. The carbon framework is derived from serine. More complex derivatives, such as nocardicin A, are formed through a tripeptide pathway similarly to the penicillins (Fig. 49).

BIBLIOGRAPHY

- Baldwin, J.E. The Biosynthesis of Penicillins and Cephalosporins. *J. Heterocyclic. Chem.* **1990**, 27, 71–78.
- Beale, M.H. The Biosynthesis of C₅–C₂₀ Terpenoid Compounds. *Nat. Prod. Rep.* **1991**, 8, 441–454; **1990**, 7, 387–407; **1990**, 7, 25–39.
- Brown, G.D. The Biosynthesis of Steroids and Triterpenoids. *Nat. Prod. Rep.* **1998**, 15, 653–696.
- Brown, S.A. Methodology. *Biosynthesis*; Geissman, T.A., Ed.; The Chemical Society: London, 1972; 1, 1–40.
- Bruneton, J. *Pharmacognosy, Phytochemistry, Medicinal Plants*; Lavoisier: Paris, 1995; 915.
- Chang, S.-I.; Hammes, G.G. Structure and Mechanism of Action of a Multifunction Enzyme: Fatty Acid Synthase. *Acc. Chem. Res.* **1990**, 23, 363–369.
- Cordell, G.A. *Introduction to Alkaloids, A Biogenetic Approach*; Wiley Interscience: New York, 1981; 1055.
- Cordell, G.A. The Monoterpene Alkaloids. *The Alkaloids: Chemistry and Biology*; Cordell, G.A., Ed.; Academic Press: San Diego CA, 1999; 52, 261–376.
- Daly, J.W.; Jerina, D.M.; Witkop, B. Arene Oxides and the NIH Shift: The Metabolism, Toxicity and Carcinogenicity of Aromatic Compounds. *Experientia* **1972**, 28, 1129–1149.
- Dewick, P.M. The Biosynthesis of Shikimate Metabolites. *Nat. Prod. Rep.* **1998**, 15, 17–58, 1995, 12, 579–607; 1995, 12, 101–133; 1994, 11, 173–203; 1993, 10, 233–263; 1992, 9, 153–181.
- Dewick, P.M. *Medicinal Natural Products. A Biosynthetic Approach*; John Wiley & Sons: Chichester, 1997, 466.
- Dewick, P.M. The Biosynthesis of C₅–C₂₅ Terpenoid Compounds. *Nat. Prod. Rep.* **1999**, 16, 97–130, 1997, 14, 111–144; 1995, 12, 507–534.
- Estévez-Braun, A.; González, A.G. Coumarins. *Nat. Prod. Rep.* **1997**, 14, 465–475.
- Gröger, D.; Floss, H.G. Biochemistry of Ergot Alkaloids—achievements and Challenges. *The Alkaloids: Chemistry and Biology*; Cordell, G.A. Ed.; Academic Press: San Diego CA, 1998; 50, 171–218.
- Herbert, R.B. *The Biosynthesis of Secondary Metabolites*; Chapman and Hall: London, 1981; 178.
- Herbert, R.B. The Biosynthesis of Plant Alkaloids and Nitrogenous Microbial Metabolites. *Nat. Prod. Rep.* **1999**, 16, 199–208, **1997**, 14, 359–372; **1996**, 13, 45–58; **1995**, 12, 445–464; **1995**, 12, 55–68; **1993**, 10, 575–592; **1992**, 9, 507–529; **1991**, 8, 185–209.
- Hutchinson, C.R. The Use of Isotopic Hydrogen and Nuclear Magnetic Resonance Spectroscopic Techniques for the Analysis of Biosynthetic Pathways. *J. Nat. Prod.* **1982**, 45, 27–37.
- Harrison, D.M. The Biosynthesis of Triterpenoids, Steroids, and Carotenoids. *Nat. Prod. Rep.* **1990**, 7, 459–484.
- Iwasa, K. The Biotransformation of Protoberberine Alkaloids by Plant Tissue Cultures. *The Alkaloids: Chemistry and Pharmacology*; Cordell, G.A., Ed.; Academic Press: San Diego CA, 1993; 46, 273–346.
- Jones, D.H. Phenylalanine Ammonia-lyase: Regulation of Its Induction, and Its Role in Plant Development. *Phytochemistry* **1984**, 23, 1349–1359.
- Knaggs, A.R. The Biosynthesis of Shikimate Metabolites. *Nat. Prod. Rep.* **1999**, 16, 525–560.
- Kutchan, T.M. Strictosidine: From Alkaloid to Enzyme to Gene. *Phytochemistry* **1993**, 32, 493–506.
- Kutchan, T.M. Alkaloid Biosynthesis—The Basis for Metabolic Engineering of Medicinal Plants. *Plant Cell* **1995**, 7, 1059–1070.
- Mann, J. *Secondary Metabolism*, 2nd Ed.; Clarendon Press: Oxford, 1987; 374.
- Mann, J.; Davidson, R.S.; Hobbs, J.B.; Banthorpe, D.V.; Harborne, J.B. *Natural Products*; Longman Scientific and Technical: Harlow, 1994; 455.
- Murray, R.D.H. Coumarins. *Nat. Prod. Rep.* **1995**, 12, 477–505.
- Nagabhushan, T.; Miller, G.H.; Varma, K.J. Antibiotics (Chloramphenicol and Analogues). *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th Ed.; John Wiley: New York, 1992; 961–978.
- Nasreen, A.; Gundlach, H.; Zenk, M.H. Incorporation of Phenethylisoquinolines into Colchicine in Isolated Seeds of *Colchicum autumnale*. *Phytochemistry* **1997**, 46, 107–115.
- O'Hagan, D. Biosynthesis of Fatty Acid and Polyketide Metabolites. *Nat. Prod. Rep.* **1995**, 12, 1–32, **1993**, 10, 593–624; **1992**, 9, 447–479.
- Rawlings, B.J. Biosynthesis of Polyketides (Other than Actinomycete Macrolides). *Nat. Prod. Rep.* **1999**, 16, 425–484, **1998**, 15, 275–308; **1997**, 14, 523–556; **1997**, 14, 335–358.
- Rohmer, M. The Discovery of a Mevalonate-independent Pathway for Isoprenoid Biosynthesis in Bacteria, Algae and Higher Plants. *Nat. Prod. Rep.* **1999**, 16, 565–574.
- Robins, R.J.; Walton, N.J. The Biosynthesis of Tropane Alkaloids. *The Alkaloids: Chemistry and Pharmacology*; Cordell, G.A., Ed.; Academic Press: San Diego CA, 1993; 44, 1116–1187.
- Robins, D.J. Biosynthesis of Pyrrolizidine and Quinolizidine Alkaloids. *The Alkaloids: Chemistry and Pharmacology*; Cordell, G.A., Ed.; Academic Press: San Diego CA, 1995; 46, 1–61.
- Sandmann, G. Carotenoid Biosynthesis in Microorganisms and Plants. *Eur. J. Biochem.* **1994**, 223, 7–24.
- Stadler, R.; Zenk, M.H. A Revision of the Generally Accepted Pathway for the Biosynthesis of the Benzyltetrahydroisoquinoline Alkaloid Reticuline. *Liebigs Ann. Chem.* **1990**, 555–562.
- Stöckigt, J. Biosynthesis in *Rauwolfia Serpentina*. Modern Aspects of An Old Medicinal Plant. *The Alkaloids: Chemistry and Pharmacology*; Cordell, G.A., Ed.; Academic Press: San Diego CA, 1995; 47, 115–172.

- Suzuki, T.; Ashihara, H.; Waller, G.R. Purine and Purine Alkaloid Metabolism in Camelia and Coffea Plants. *Phytochemistry* **1992**, *31*, 2575–2584.
- Torssell, K.B.G. *Natural Product Chemistry*, 2nd Ed.; Apotekarsocieteten: Stockholm, 1997; 480.
- Townsend, C.A. Oxidative Amino Acid Processing in β -Lactam Antibiotic Biosynthesis. *Biochem. Soc. Trans.* **1993**, *21*, 208–213.
- Veders, J.C. The Use of Stable Isotopes in Biosynthetic Studies. *Nat. Prod. Rep.* **1987**, *4*, 277–337.
- Verpoorte, R.; van der Heijden, R.; Moreno, P.R.H. Biosynthesis of Terpenoid Indole Alkaloids in *Catharanthus roseus* Cells. *The Alkaloids: Chemistry and Pharmacology*; Cordell, G.A., Ed.; Academic Press: San Diego CA, 1997; 49, 221–299.
- Zenk, M.H.; Gerardy, R.; Stadler, R. Phenol Oxidative Coupling of Benzyloquinoline Alkaloids is Catalyzed by Regio- and Stero-selective Cytochrome P-450 Linked Plant Enzymes: Salutaridine and Berbaminine. *J. Chem. Soc. Chem. Commun.* **1989**, 1725–1727.

BIOTECHNOLOGY AND BIOLOGICAL PREPARATIONS

Ronald P. Evens

Clinical Research, Amgen Inc., Thousand Oaks, California, U.S.A.

INTRODUCTION

A scientific revolution in drug discovery and development occurred in the last 2–3 decades of the 20th century with the advent and full realization of biotechnology. This new discipline resulted in the discovery, development, production, and commercialization of over 60 innovative biological products from 1982 to 1999 that treat over 50 human disease conditions, many of which were major medical breakthroughs. Biotechnology is the use of a living system(s) to discover and produce a pharmaceutical product, often called a biological product because of its structural and biologic similarity to naturally occurring substances in the human body. Biological products have been predominantly proteins, but new biologicals have expanded to include DNA/RNA derivatives, such as m-RNA analogues and genes, peptides, cell therapies, and biologic carriers, such as liposomes. Biotechnology further encompasses biological products that have agricultural uses and environmental applications.

Biotechnology is a collection of biologic techniques and drug development technologies that permit whole new biologic discoveries and products. The techniques and technologies include recombinant DNA (r-DNA) technology, monoclonal antibodies (Mab), antisense, genomics, polymerase chain reactions (PCR), combinatorial chemistry, high throughput screening (HTS), transgenics, proteomics, X-ray crystallography, cell therapy, and gene therapy. These technologies help to elucidate new biologic mechanisms of disease, identify naturally occurring substances responsible for a biologic effect, create duplicates of the natural substances that often are found only in minute amounts in the body, determine new products to enhance natural products or block their functions, and permit mass production of these rare products for commercialization. Fig. 1 displays these technologies, along with traditional drug development techniques (i.e., medicinal chemistry, pharmacology, pharmaceuticals, toxicology, and pharmacokinetics) and good clinical (research) and manufacturing practices, all of which are collectively employed to discover, develop, and produce biological products. More than 60 biological products comprise a new “biological” method to treat human

disease, that is, biotherapy, a new armamentarium for health care providers, to be used in conjunction with drugs, devices, radiology, physical therapy, and psychotherapy.

HISTORY OF BIOTECHNOLOGY

Some science historians will cite that the origins of biotechnology go back 6000–8000 years to the Sumerian culture. Fermentation is an age-old, basic biologic process whereby a living organism, a yeast, will react with carbohydrate materials, such as wheat, in a vessel to produce alcohol. This Sumerian product was beer. Later, this basic biotechnology process (fermentation) was employed in the preparation of bread and cheese, food staples, over the millennia.

The modern era of biotechnology is thought to have started in the 1950s with the discovery by Watson and Crick of the DNA double helix—the matched pairs of four nucleic acids in a specific sequence and in a 3D spatial configuration. Several key discoveries in biology in the 1960s underpin biotechnology; namely, the genetic code is universal in nature among all living things; for the 20 amino acids, 64 specific nucleic acid triplet codes are responsible for interpretation of genes into proteins; and genetic material is transferable among different organisms.

The early 1970s saw the development of the two core technologies of biotechnology, that is, r-DNA and Mab, which account for 48 of the 61 commercially available products in 2000. This DNA process is sometimes called genetic engineering. The r-DNA technology is a process whereby human DNA (genes) is identified, a human gene is inserted into bacterial DNA (plasmid), and the plasmid is placed into the nonhuman host cells (e.g., *E. Coli* bacteria), which manufacture their typical variety of proteins, plus produce a human protein from the human gene. Monoclonal antibodies are proteins that are produced by the body's plasma cells in response to foreign substances, and Mabs serve to attach to and neutralize the foreign substance. A series of discoveries have enhanced the r-DNA and Mab processes including PCR; where genetic material can be accessed and reproduced even a millionfold; sophisticated analytical processes for genes

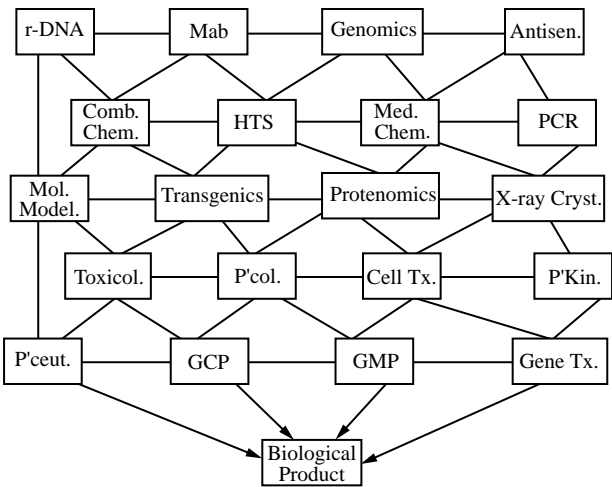


Fig. 1 Antisen: antisense; Cell tx: cell therapy; Comb Chem: combinatorial chemistry; GCP: good clinical practices; Gene Tx: gene therapy; GMP: good manufacturing practices; HTS: high throughput screening; Mab: monoclonal antibodies; Med Chem: medicinal chemistry; Mol Model: molecular modeling; P'ceut: pharmaceuticals; P'Kin: pharmacokinetics; PCR: polymerase chain reaction; P'col: pharmacology; r-DNA: recombinant deoxyribonucleic acid; X-ray Cryst: X-ray crystallography; Toxicol: toxicology.

and proteins; and accelerated gene sequencing techniques. The first commercial product derived from biotechnology was recombinant human insulin.

DRUG DEVELOPMENT TECHNOLOGIES

The products produced through biotechnology, also called biological products, have been primarily proteins, which have been isolated from the human body or created ex vivo to interact with human physiologic systems. About 50 of the 60 commercially available products are proteins. Proteins are complex, large molecules with several key structural features required for their activity. Each protein has a specific amino acid sequence, from the 20 amino acids in nature, which must be preserved intact. Special structural features include disulfide bridges, specific terminal amino acid species, glycosylation (carbohydrate species attached to the protein backbone), and 3D configurations.

Recombinant DNA Technology

The predominant technology in biotechnology during the 1980s and 1990s was r-DNA technology, which was the process used to create more than 35 products, that is, about

60% of the commercially available products. Five steps comprise r-DNA technology: protein identification, gene isolation, cloning and expression, manufacturing (scaleup processes), and quality assurance for protein and process integrity (Table 1). Step 1 involves finding a protein responsible for some biological effect in the human body that has therapeutic potential. The protein needs to be isolated from its normal milieu, usually a body fluid or cell. The structure of the protein is determined, including amino acid sequence, glycosylation (carbohydrate content), disulfide bridging, and possibly 3D configuration. Step 2 requires the isolation of the human gene responsible for the protein, which entails one of three mechanisms. First, if we know the protein's full amino acid sequence, and we do know the 64 nucleic acid triplets that code for the 20 amino acids, we can construct many combinations of those triplet codes that may be genetic representations of the target protein. These genetic constructs are genes, one of which will be identified through screening as the correct gene with the capacity to produce the target protein. Second, we may be able to find the human cell that produces our target protein. In this cell, we can ferret out the messenger RNA (m-RNA) that is responsible through

Table 1 Recombinant DNA technology

Step 1	Protein identification Protein isolation Biological property description Protein sequencing and mapping
Step 2	Gene isolation Nucleic acid triplet sequencing for amino acid sequence mRNA isolation, with reverse transcriptase and complementary DNA DNA probes for genetic library
Step 3	Cloning and expression Gene (human) insertion into plasmids Plasmid incorporation into host cells Host cell production of proteins Master working cell bank
Step 4	Manufacturing scale-up Inoculum stage Fermentation or cell culture stage Protein purification stage Formulation stage
Step 5	Quality assurance Genetic testing Bulk product testing Process validation Final product testing

the process of translation for producing the target protein. The viral enzyme, reverse transcriptase, is capable of creating the target complementary DNA, or gene, from this specific m-RNA. This method was used to find the gene for insulin, which led to marketed products. Third, the human gene could be fished out of the human genome using nucleic acid probes, which is a complex, daunting task. In this method, we identify several peptides in the amino acid sequence of the protein. Using the triplet codes for amino acids, we build a specific nucleic acid combination (probe) for each peptide. Then, we break the chromosomes (3–4 billion pairs of nucleic acids) into thousands of pieces of DNA. Through many, many serial experiments, we try to match the first nucleic acid probe to the DNA mixture, which does create a subset of DNA pieces (hundreds or thousands). In another series of experiments, a second, different nucleic acid probe for a different peptide is matched against this DNA subset for matches. Matches do occur, resulting in a series of possible genes. Each gene must be evaluated by genetics to produce proteins, which then are tested to be sure that the protein has the structural features and pharmacological properties in test animals of the targeted, naturally occurring protein.

Step 3 in r-DNA technology involves cloning of the gene and expression of the protein by the gene. Cloning is the reproduction of the target human gene in a nonhuman cell. Expression is the production of the target human protein by a nonhuman cell containing the human gene. These processes require a vector for the DNA (genes) so that the gene can be carried from cell to cell. A bacterial plasmid is a circular piece of DNA that is transferable between cells (a carrier), will accept the insertion of a human gene, and will allow the human gene to be turned on. The plasmid must be cut open to accept the human DNA (gene) by unique enzymes (restriction endonucleases), each of which is highly specific to a certain nucleic acid sequence that can match a terminal end of the human gene. These “sticky” ends of the opened bacterial plasmid and the human gene permit recombination of the DNA, under the influence of a ligase enzyme, resulting in an r-DNA molecule containing a human gene inserted into a bacterial plasmid. Next, the r-DNA molecule is inserted into a host cell that serves to produce all of its routine proteins, plus the cell manufactures the human protein from the human gene that it carries. The host cell needs to possess several important characteristics: a short reproductive life cycle, a long-term viability in an *in vitro* setting, the ability to accept bacterial plasmids, a good productive capacity for proteins, and the ability to produce the human protein consistently without its alteration. This unique new cell and its offspring, created in the laboratory, are called the master working cell bank.

Step 4 in r-DNA technology is scale-up manufacturing and is comprised of four phases: inoculum, fermentation, purification, and formulation. Inoculum phase entails the use of daughter cells from the new host cell (master working cell bank), actually removed from storage in a -70°C freezer. The daughter cells are grown in a specific medium in serially larger flasks and assessed for normal growth characteristics. The growth medium is a unique and specific mixture of minerals and compounds to enhance cell viability (lifespan) *in vitro* and functional ability of cells to produce proteins. The fermentation or cell culture phase involves inoculating thousand of containers, or large fermenters, with cells from the inoculum phase and adding the growth media. The host cells will proceed to produce proteins, either intracellularly in storage vacuoles for bacterial host cells or extracellularly for mammalian cells. Purification of the protein follows, which varies between bacterial vs. mammalian systems. For bacteria, the cells are removed from the fermenters in a cell paste, which is centrifuged to obtain proteins out of the cells. The protein mixture is then run through an extraction process, often high-pressure liquid chromatography to separate the target protein from all other proteins. For mammalian cells, the culture media containing proteins is collected periodically and replenished. Extraction is basically the same process. A pure bulk protein is the result of purification. The final phase is formulation, wherein a diluent is chosen for the protein, incorporating the best mix of fluids, stabilizers, and minerals to achieve optimal protein stability and maximal shelf life. Formulation of proteins is confounded by the general delicate nature of proteins and the many degrading processes that can occur with them. Most vials of biological products that are proteins do not contain preservatives because of their reactivity with proteins. As can be readily observed from this description of manufacturing r-DNA biotechnology, manufacturing is quite multi-faceted and complex.

Step 5 in r-DNA technology is quality control for the final product, components, and processes throughout the manufacturing process. Proteins are quite complex molecules and their manufacturing is very complex, both of which create high potential for degradation, alteration, or contamination. Protein breakdown can occur through many processes, many of which are listed in Table 2. Contamination can result from typical methods for drugs, e.g., microbial or chemical means, and because of the genetic material employed, through oncogenic or viral DNA incorporation or alteration. Therefore, quality control ensures final product integrity through an extensive series of tests, which involve four key areas: genetic material (plasmids and genes), bulk protein product, final

Table 2 Biological product—stability/degradation

Precipitation
Clumping/aggregation
Cross-linkage
Unlinkage (disulfides)
Amino acid mutation
Glycosylation/deglycosylation
Conjugation
Amino acid deletions/additions
Reduction
Oxidation
Folding/unfolding of protein
Deamidation
Proteolysis
Protein inclusions
Terminal amino acids variations

product, and the manufacturing process. Exemplary tests are listed in Table 3.

Monoclonal Antibody Production

Mabs are complex proteins that have a uniform basic structure, comprised of four subunits that are divided into two matched pairs of protein material—heavy chains and light chains. These Mabs are highly specific proteins that the plasma cells in the human produce against a single antigenic foreign material. Historically, Mabs are produced in biotechnology in a mouse, and are targeted against human antigens. We are developing a product (Mab) to attack and eliminate these target antigens. The mouse will create highly specific Mabs against the human antigen (target); however, the quantity of Mabs (protein) produced by even very large numbers of mice is very small. Hence, biotechnology for Mabs uses a myeloma cell and fuses it with the mouse plasma cell to create a murine hydridoma cell. Myeloma cells impart several additional characteristics to the hybridoma, that is, very high production of proteins (Mabs) and long lifespan, which make Mab manufacturing commercially feasible. These hydridoma cells are the new master cells to produce large amounts of specifically targeted Mabs. The murine origin of Mabs creates significant limits for these products, because administration of murine Mabs to humans leads to the body’s rejection against the murine nature of the protein. Two limits of murine Mabs are a toxic Human Antibody Mouse Antigen response with fever and chills and less binding of the Mabs to the target cell, thus limiting their activity. Therefore, scientists now manipulate Mabs by substituting human subunits for the four murine subunits, creating chimeric molecules, part murine and part human.

Table 3 Quality control tests in r-DNA technology

1. Genetic material (5–10 tests)	Karyotypic analysis
	Oncogene screening
	Gene stability
	Infection DNA screens
2. Bulk protein products (20 tests)	Amino acid sequence
	Peptide maps
	High-pressure liquid chromatography
	Radio immunoassay
	Western blot chromatography
	Bio assay
3. Process validation (10 tests)	Protein yield
	Protein challenge
	Endotoxin spiking
4. Final product (30 tests)	Protein analyses (repeat of #2 above)
	DNA contamination
	Stability tests
	Freeze-thaw tests

This humanization lessens toxicity and could increase activity. We now have eight Mabs available to treat inflammatory conditions or bind to proteins to arrest a process, as in slowing kidney rejection in transplant patients.

Genetic Technologies

In drug discovery and development, genetic materials (DNA, m-RNA, genes, and RNA enzymes) have come to the forefront as biological products and, more so, as primary tools of product discovery. Three primary areas are covered here: antisense, gene therapy, and genomics. Antisense is a RNA molecule that is complementary to, or a mirror image of, a segment of an aberrant or mutated m-RNA molecule, involved in the pathogenesis of a disease. The antisense RNA molecule will bind to the noxious m-RNA molecule, preventing the disease from manifesting. One product is currently available, fornivirsen, used to treat cytomegalovirus associated with retinitis that can occur in AIDS patients.

Gene therapy is a technology employing a gene as a therapeutic agent to treat a disease. The potential goals of gene therapy include replacement of an inactive gene, reactivating inactive genes, turning off genes causing disease, as in oncogenes, turning on further naturally protective genes, or adding a gene characteristic to cells, such as increased susceptibility of cancer cells to chemotherapy drugs. Development challenges in gene

therapy include finding the target gene that is causing the disease or needing enhancements, identifying the optimal target human cell for insertion of an additional gene, insertion (delivery) of an extra gene reliably into a cell, and causing the gene to be functional in a reasonably physiologic manner (extent and duration of activity). We are rapidly isolating and identifying all the genes in the human genome; however, gene delivery has not been achieved in a reliable, reproducible manner sufficient for routine therapy.

Genomics is a technology in which we search for a gene that is responsible for some process or product in the human body. The gene may lead to a new process or disease target for which drugs or biologicals could be developed to favorably alter. For example, the gene for cell immortality was discovered that led to the protein, telomerase, an enzyme. Telomerase is responsible for adding telomeres, short nucleic acid sequences, to the end of all chromosomes, protecting chromosomes from mutations and, ultimately, cell death. Alternatively, genomics may lead to a gene that produces a protein that can be used as a therapeutic agent. For example, osteoprotegerin is a key protein that is the natural substance that turns off osteoclasts, which break down bones and could lead to osteoporosis. Genomics requires the collection of massive amounts of information regarding the human genome or genetics, including protein and peptides, receptors for activity of proteins or drugs, mechanisms of drug activity, and subunits of drugs, proteins, RNA, or DNA responsible for physiologic or pharmacologic actions. The science of informatics now exists to share, integrate, and manipulate these massive amounts of data to create new products.

BIOLOGICAL PROCESSES

Polymerase Chain Reaction

PCR is a critical core process in biotechnology that permits the expansion of the amount of genetic material, starting from minute amounts. The process involves first denaturing DNA with high heat, that is, unraveling the DNA double helix so that the genetic code (sequence) can be read and possibly duplicated. Second, a leader sequence for DNA is used to initiate reading of the genetic code at a specific point, and both helixes or strands of DNA can be read. Third, the enzyme, DNA polymerase, catalyses the reading of the genetic code. By sequential repetition of these three steps, the genetic material is magnified; for example, 20

replications yield a millionfold increase in the DNA material.

High Throughput Screening

HTS is intended to obtain faster more high-quality product leads from large volumes of genetic or peptide molecules. HTS has improved the number of molecules that can be screened for activity by 10- to 1,000-fold. The process of HTS is dependent on improved analytical processes, miniaturization of equipment, and automation. Currently, over 100,000 samples can be tested in a day.

Combinatorial Chemistry

Combinatorial chemistry involves the use of the basic building blocks, either the 20 amino acids or 4 nucleic acids, to build new molecules. All of the different combinations of a set number of building blocks are created; for example, 10 amino acids can result in over 3.5 million compounds. Huge libraries of compounds are produced, which require screening through HTS and the use of informatics to help sort out the structures and actions of all these new compounds.

Cell Therapy

A newer technique to treat disease is based on obtaining healthy cells from a specific tissue, selecting out a specific subset of cells with certain desirable properties, and enhancing the activity of these cells through ex vivo manipulation. We then return these specifically selected, enhanced, and activated cells to patients whose cells are not sufficiently functional, thereby ameliorating a disease. Currently, chondrocytes responsible for cartilage production are taken from a patient's knee that has serious damage and is repairing poorly. These chondrocytes are manipulated ex vivo and returned to the patient to normalize cartilage production. Bone marrow progenitor cells are collected from peripheral blood, bone marrow cells, or cord blood, and the cells with greatest regenerative potential are selected out through various cell-tagging processes. Following life-threatening chemotherapy in a cancer patient that destroys almost all bone marrow, these selected progenitor cells are administered to the patient to accelerate regeneration of bone marrow and white blood cell production, thereby preventing infections. Foreskins of newborns are collected and placed in a 3D construct that permits growth of dermis and epidermis. This construct is used later in venous leg ulcers to accelerate wound healing, thereby reducing healthcare needs.

Molecular Manipulation of Biologicals

A new generation of biological products are derivatives of existing molecules, which are altered to change their properties resulting in new molecules. Pegylation is a process in which polyethylene glycol (PEG) is added to the protein structure, which is being done for alfa-interferon. The addition of PEG may extend the product's half-life or duration of effect, yet the desired activity of the molecule can be maintained.

Glycosylation entails adding carbohydrate species, e.g., sialic acid residues, to the amino acid backbone of a protein. The altered structure in some proteins, such as epoetin alfa, can extend the product's half-life while sustaining its pharmacologic properties.

through r-DNA technology into 11 products (see Table 4). Insulin was the first product from biotechnology to be approved in 1982, and two products are now available. One parent molecule can result in more than one product, because manipulations of the molecule can be done (for example shifting terminal amino acids), without altering the biological activity of the protein. Each molecule may have a related but different indication. Seven growth hormone products are on the market. The generic names of products often will be different from similar products produced by other companies, e.g., follitropin beta from Organon and Follitropin alfa from Ares-Serono. Each table of approved biologicals includes the generic and brand names, company marketing the product, and therapeutic areas of use.

BIOLOGICAL PRODUCT CATEGORIES

Over 60 products (biologicals) were marketed by over 30 companies in the United States from 1982 to 1999. In many diseases, biological products often have been major breakthroughs offering the first treatments where nothing previously was effective for serious diseases, i.e., dornase for cystic fibrosis or beta-interferon for multiple sclerosis. Now, we have biological products in more than 10 distinct categories.

Hormones

Three naturally occurring, protein-based hormones—insulin, growth hormones (somatotropin), and follicle stimulating hormone (follitropin)—have been duplicated

Interferons

These proteins called interferons are produced by many cells in the human body and serve to provide protection against foreign substances, such as infectious material. Interferons have several beneficial properties, such as stimulating the immune system and possessing cytolytic activity. Three families of interferons exist—alfa, gamma, and beta; and eight products have been created (Table 5). The indications (10) for interferons are broad, including oncology (e.g., malignant melanoma and chronic myelogenous leukemia), viral infection (e.g., hepatitis C), and immune diseases (e.g., multiple sclerosis).

Indications are added often to a molecule over time following the original product approval. Alpha-interferon is a very good example of this drug development process. The first indication was hairy cell leukemia in 1986,

Table 4 Approved biologicals (hormones)

Generic name	Brand name (company)	Therapeutic area
Human insulin	Humulin (Eli Lilly)	Insulin-dependent diabetes mellitus
	Novolin (NovoNordisk)	
Human growth hormone	Protropin (Genentech)	Growth hormone deficiency
	Humatrope (Eli Lilly)	Growth retardation in chronic renal disease
	Nutropin (Genentech)	AIDS wasting
	Saisen (serono)	
	Serostim (Serono)	
	Gentropin (General Biotechnology)	
	Norditropin (NovoNordisk)	
Follitropin beta (FSH)	Follistim (Organon)	Ovulatory failure
Follitropin alfa (FSH)	Gonal-F (Ares-Serono)	Ovulatory failure
Growth hormone releasing hormone	Geref (Serono)	GH deficiency in children

Table 5 Approved biologicals (interferons)

Generic name	Brand name (company)	Therapeutic area
Interferon alfa	Wellferon (Glaxo Wellcome)	Chronic hepatitis C
Interferon alfa-2a	Roferon-A (Hoffmann-La Roche)	Hairy cell leukemia; AIDS-related Kaposi's sarcoma; chronic myelogenous leukemia
Interferon alfa-2b	Intron A (Schering-Plough)	Hairy cell leukemia; AIDS-related Kaposi's sarcoma; Chronic hepatitis, types B and C; Condylomata acuminata; Malignant melanoma; nonhodgkins lymphoma
Interferon alfa-n3	Alferon N (Interferon Sciences)	Condylomata acuminata
Interferon gamma-1b	Actimmune (Genentech)	Chronic granulomatous disease
Interferon beta-1b	Betaseron (Berlex)	Acute relapsing-remitting multiple sclerosis
Interferon beta-1a	Avonex (Biogen)	Acute relapsing-remitting multiple sclerosis
Interferon alfa con-1	Infergen (Amgen)	Hepatitis C (Naive and Relapse)

Table 6 Approved biologicals (growth factors and interleukins)

Generic name	Brand name	Therapeutic area
Epoetin alfa	Epogen (Amgen) Procrit (Ortho Biotech)	Certain anemias from chronic renal disease; Zidovudine-induced anemia; cancer chemotherapy; anemia in surgery patients
Filgrastim (r-metHuG-CSF)	Neupogen (Amgen)	Neutropenias due to myelosuppressive chemotherapy; myeloid reconstitution after bone marrow transplantation; severe chronic neutropenia; peripheral blood progenitor cell transplant; induction and consolidation therapy in AML
Sargramostim (r-HuGM-CSF)	Leukine (Immunex)	Myeloid reconstitution after bone marrow transplantation; bone marrow transplant failure; adjunct to chemotherapy in AML; peripheral blood progenitor cell transplant
Becaplermin (PDGF)	Regranex (Ortho-McNeil)	Diabetic foot ulcer
Aldesleukin (IL-2)	Proleukin (Chiron)	Metastatic renal cell carcinoma; metastatic melanoma
Oprelvekin (IL-11)	Neumega (Genetics Institute)	Thrombocytopenia due to chemotherapy

Table 7 Approved biologicals (monoclonal antibodies)

Generic name	Brand name	Therapeutic area
Trastuzumab	Herceptin (Genentech)	Metastatic breast cancer (Her 2 Neu+)
Palivizumab	Synagis (MedImmune/Abbott)	Prevention of respiratory syncytia viral and fatal pneumonia in children
Infliximab	Remicade (Centocor)	Crohn's disease
Muromonab- CD3	Orthoclone OKT 3 (Ortho Biotech)	Acute allograft rejection in renal transplant patients; heart and liver transplant rejection
Abciximab	ReoPro (Centocor/Eli Lilly)	Prevention of blood clots after PTCA, unstable angina prior to PTCA
Rituximab	Rituxan (IDEC/Genentech)	Low grade non-Hodgkins lymphoma
Daclizumab	Zenapax (PDL/Roche)	Kidney transplant, acute rejection
Basiliximab	Simulect (Ligand/Novartis)	Acute organ transplant rejection

which was followed by seven additional uses. Extensive clinical trials (Phases 2 and 3) are required to establish each use.

Growth Factors and Interleukins

These proteins can be divided into three areas: blood cell growth factors (GFs), also known as colony stimulating factors (CSFs), tissue growth factors, and interleukins, which help communicate between cells. Eight CSFs are available worldwide: filgrastim, sargramostim, molgramostim, regimostim, lenograstim, and nartograstim, which stimulate white blood cell production and activity. Epoetin products (epoetin alfa and epoetin beta) reverse anemia by stimulating red blood cell production. Becaplermin is a tissue growth factor for the epidermis, being used to accelerate wound healing in diabetic ulcers (see Table 6).

Two interleukins are in use for renal cell carcinoma (IL-2) and for thrombocytopenia (IL-11). These products carry substantial multiorgan, especially cardiovascular, toxicity.

Monoclonal Antibodies

In the late 1990s, the monoclonal antibody proteins greatly expanded with six product approvals in a wide range of indications. In addition to kidney transplants, new uses include prevention of blood clots, metastatic breast cancer, non-Hodgkins lymphoma, Crohn's disease, and viral pneumonia in children. The growth in products with major new indications is predicated on humanization of the murine antibodies leading to less side effects and more affinity for the target receptors (more potential desired activity; see Table 7).

Enzymes and Blood Coagulation Factors

The first protein enzyme developed in the late 1980s was alteplase (t-PA), used to minimize complications due to blood coagulation in acute myocardial infarction. Three further enzymes were developed for similar indications. A unique enzyme was discovered for cystic fibrosis, dornase alfa, which is the enzyme deficiency responsible for the etiology of this disastrous disease. Blood factors (F.8, F.9, and F.7) are involved in normal blood coagulation. Their

Table 8 Approved biologicals (enzymes and blood factors)

Generic name	Brand name	Therapeutic area
Alteplase	Activase (Genentech)	Acute myocardial infarction; pulmonary embolism; stroke
Dornase alfa	Pulmozyme (Genentech)	Respiratory complications of cystic fibrosis
Imiglucerase (recombinant)	Cerezyme (Genzyme)	Type 1 Gaucher's disease
Retepase	Retevas (Centocor)	Acute myocardial infarction
Eptifibatide	Integrelin (Cor/Schering)	Acute coronary syndromes; angioplasties
Tirofiban HCl	Aggrastat (Merck)	Acute coronary syndromes
Factor 7	Novo-Seven (NovoNordisk)	Hemophilia (F.7 deficiency)
Factor 8	KoGENate (Bayer); Recombinate (Baxter)	Hemophilia A
Factor 9	Benefix (Genetics Institute)	Hemophilia B

Table 9 Approved biologicals (others)

Generic name	Brand name	Therapeutic area
Hepatitis B vaccine	Engerix-B (SmithKline Beecham) Recombivax HB (MSD)	Hepatitis B prophylaxis
Lyme disease vaccine	LymErix (SKB)	Prevention of Lyme disease
Doxorubicin-liposomal	DOXIL (Alza)	Kaposi's sarcoma
Amphotericin-liposomal	Abelcet (Liposome) Amphotec (Alza) Ambisome (Gilead)	Aspergillosis infection Systemic fungal infections
Daunorubicin-liposomal	DaunoXome (Gilead)	Kaposi's sarcoma
BCNU-polymer	Gliadel (Guilford)	Recurrent glioblastoma multiforme

Table 10 Approved biologicals (others)

Generic name	Brand name	Therapeutic area
Hyaluronic acid membrane	Seprafilm (Genzyme)	Prevention of adhesions after surgery
Hyaluronic acid gel	Adcon-L (Gliatech)	Prevention of adhesions after lumbar surgery
Glatiramer	Copaxone (Teva/HMR)	Relapsing multiple sclerosis
Cartilage culturing service	Carticel (Genzyme)	Cartilage damage in knees
Skin graft product	Apligraf (Organogenesis/Novartis)	Wound healing of venous leg ulcers
Vitravene	Fornivirsen (Isis)	CMV retinitis
Etanercept	Enbrel (Immunex)	Rheumatoid arthritis
Denileukin diftitox	Ontak (Ligand)	Cutaneous T-cell lymphoma

deficiency leads to bleeding disorders, but it is fully correctable through replacement therapy with these r-DNA proteins (see Table 8).

Other Biological Products

Tables 9 and 10 list a myriad of different product types for many unique indications. Two vaccines exist for hepatitis B and Lyme disease. Five liposomal products are available for oncologic and fungal uses. Cell therapies involve harvesting human cells or tissue, growing them *ex vivo*, and returning them to patients to heal damaged tissues, e.g., for wound healing and cartilage damage. Hyaluronic acid products are used during surgery, being placed between or around tissues preventing adhesions. An antisense molecule (fornivirsen) is used to treat CMV retinitis. The destructive tumor necrosis factor is blocked by etanercept to improve rheumatoid arthritis treatment. A fusion molecule combines interleukin-2 with a toxin to treat a rare cutaneous cancer. A peptide of four amino acids (glatiramer) is used to treat multiple sclerosis. A biological carrier (wafer) is used to administer the cancer drug, BCNU, for brain tumors.

FORMULATION AND DISTRIBUTION ISSUES

Most biological products (75%) are proteins, which as described earlier are large, complex, delicate molecules. Breakdown of proteins can occur through many mechanisms, also noted earlier. Furthermore, the final formulations often are sensitive to temperature extremes, which can cause precipitation. Restrictions in diluents exist because of potential adverse changes in stability. The protein formulations usually do not contain preservatives because of their interactions with

Table 11 Types of products in research

• Antisense 9	• Interferons 12
• Cellular Therapy 20	• Interleukins 9
• Clotting Factors 3	• Monoclonal Antibodies 74
• Colony Stimulating Factors 3	• Nucleotide Analogues 3
• Erythropoietins 3	• Signaling 3
• Fusion Proteins 3	• Soluble Receptors 3
• Gene Therapy 38	• Tissue Plasmin Activators 3
• Growth Factors 21	• Vaccines 77
• Growth Hormones 5	• Others 58

Table 12 Diseases being studied

• AIDS/HIV 29	• Heart Disease 28
• Autoimmune Disorders 29	• Infectious Disease 36
• Blood Disorders 8	• Infertility 4
• Cancer 151	• Neurologic Disorders 26
• Diabetes 13	• Respiratory Diseases 20
• Digestive Disorders 9	• Skin Disorders 14
• Eye Conditions 3	• Transplantation 14
• Genetic Disorders 10	• Other Diseases 22
• Growth Disorders 4	

proteins, leading to single-use vials. Refrigeration usually is needed to achieve maximal shelf life. Prior to administration, vials should be warmed to room temperature to lessen local reactions.

PRODUCT PIPELINE

The future of biotechnology is expected to involve many product approvals. As of the year 2000, over 700 molecules are undergoing clinical studies, with many more in basic research. The approximately 1250 biotechnology companies in the United States are investing about \$10 billion annually in research. The types of products in development are quite varied, with monoclonal antibodies and vaccines comprising the largest categories (Table 11). The therapeutic areas being treated cover most diseases, with cancer uses comprising the most common research area (Table 12).

Table 13 Biotechnology companies

• Alza	• Immunomedics
• Amgen	• Immunex
• Advanced Tissue Sciences	• Interferon Sciences
• Biogen	• Ligand
• Bio-Technology General	• Liposome Company
• Centocor	• MedImmune
• Chiron	• Nabi
• Cytogen	• NeoRx
• Enzon	• Organogenesis
• Genentech	• Organon
• Genetics Institute	• Ortho Biotech
• Gilead	• Protein Diagnostics
• Gliatech	Laboratory
• Guilford	• Seragen
• IDEC	• Serono
	• Zymogenetics

These biological products are being evaluated by many biotechnology companies, beyond the 32 companies who already have at least one marketed product (Table 13). The top 10 companies in order of their sales/revenue success in 1999 are Amgen, Genentech, Chiron, Genzyme, Biogen, Agouron, Centocor, Immunex, Nabi, and MedImmune.

BIBLIOGRAPHY

- Andrade, M.A.; Sander, C. Bioinformatics: From Genome Data to Biological Knowledge. *Curr. Opin. Biotechnol.* **1997**, *8* (6), 675–683.
- Buckel, P. Recombinant Proteins for Therapy. *Trends Pharm. Sci.* **1996**, *17* (12), 450–456.
- Bush, P. Pharmacotherapeutics of Biotechnology—Derived Products. *J. Pharm. Practice* **1998**, *11* (1), 54–71.
- Carroll, W.L. Introduction to Recombinant-DNA Technology. *Am. J. Clin. Nutr.* **1993**, *58* (Suppl. 2), 249S–258S.
- Crooke, S.T. Antisense Therapeutics. *Biotechnol. Genetic Engineering Rev.* **1998**, *15*, 121–157.
- Demain, A.L. An Overview of Biotechnology. *Occup. Med.* **1991**, *6* (2), 157–168.
- Evens, R.P.; Dinarello, C.A.; Browne, J.; Fenton, D. Biotechnology and Clinical Medicine: Part I. *Hosp. Physician* **1995**, *31* (1), 27–36.
- Evens, R.P.; Dinarello, C.A.; Browne, J.; Fenton, D. Biotechnology and Clinical Medicine: Part II. *Hosp. Physician* **1995**, *31* (2), 26–31.
- Georgiou, G. Recombinant DNA Technology. *Trends Biotechnol.* **1995**, *13* (3), 79–80.
- Glossary of Terms, Biotechnology at Work. Biotechnology Industry Organization*; 1625 K Street NW: Washington, DC, 1995; 2006.
- Kipriyanov, S.M.; Little, M. Generation of Recombinant Antibodies. *Mol. Biotechnol.* **1999**, *12* (2), 173–201.
- Lipner, M. The Polymerase Chain Reaction: Amplifying its role in Research and Beyond. *Oncol. Times* **Dec. 1992**, 14–16.
- Mountain, A. Gene Therapy: The First Decade. *Trends Biotechnol.* **2000**, *18* (3), 119–128.
- Myers, P.L. Will Combinatorial Chemistry Deliver Real Medicines? *Curr. Opin. Biotechnol.* **1997**, *8* (6), 701–707.
- Morrison, S.W.; Giovannetti, G.T. *Biotech 99: Bridging the Gap, Ernst & Young's 13th Biotechnology Industry Annual Report 1998*; Ernst & Young LLP; Palo Alto, CA, Dec. 1998.
- Muellner, S.; Neumann, T.; Löffspeich, F. Proteomics —A New Way for Drug Target Discovery. *Arzneimittel-Forschung* **1998**, *48* (1), 93–95.
- Seamon, K.B. Specifications for Biotechnology-derived Protein Drugs. *Curr. Opin. Biotechnol.* **1998**, *9* (3), 319–325.
- Spurr, N.; Darvasi, A.; Terrett, J.; Jazwinska, L. New Technologies and DNA Resources for High Throughput Biology. *Br. Med. Bull.* **1999**, *55* (2), 309–324.
- Survey, New Medicines in Development, Biotechnology Pharmaceutical Research and Manufacturers of America*; 1100 15th St. NW: Washington, DC, 2006; 1–44, February 2000.
- Tami, J.A. The Major Techniques of Biotechnology. *J. Pharm. Practice* **1998**, *11* (1), 28–37.
- Wiley, S.R. Genomics in the Real World. *Curr. Pharm. Design* **1998**, *4* (5), 417–422.

BLOW-FILL-SEAL—ADVANCED ASEPTIC PROCESSING

Deborah J. Jones

Norton Steripak, Cheshire, United Kingdom

INTRODUCTION

Blow-Fill-Seal (BFS) technology was developed in the early 1960s and was initially used for filling many liquid product categories, such as nonsterile medical devices, foods, and cosmetics. The technology has now developed to an extent that BFS systems are used today throughout the world to successfully aseptically produce sterile pharmaceutical products, such as respiratory solutions, ophthalmics, and wound care products.

BFS is an advanced aseptic processing technique within which plastic containers are formed by means of molded extruded polymer granules that are filled and sealed in one continuous process. This differs from conventional aseptic processing where container formation, preparation, sterilization, and container filling and closure are all separate processes.

Due to the level of automation of the entire process, very little human intervention is necessary during manufacture as compared to traditional aseptic filling. This is considered an advanced aseptic filling process. It is therefore possible to achieve very high levels of sterility confidence with a properly configured BFS machine designed to fill aseptically.

OUTLINE OF THE BFS PROCESS

The pharmaceutical BFS process combines the formation of plastic containers by blow/vacuum molding extruded pharmaceutical grade polymers, with an aseptic solution filling system.

Polymer granules are continuously fed to a machine hopper through an adiabatic screw extruder. Within the extruder the polymer is subjected to high temperature (generally greater than 160°C) and pressure (up to 350 bar) and becomes molten. It is then extruded through a die and pin set to form an open-ended tube of molten polymer known as a parison. The parison is supported by sterile air (parison support air) that is fed into the center of the parison through a sterilizing grade air filter fed with oil free compressed air. The parison is held in position by a parison clamp, which on some machines also serves to seal

the bottom of the parison. A mold set in two halves then moves over to the parison and closes around it. Molding is facilitated by vacuum slots in the mold. The molded plastic is severed from the continuously extruding parison by a hot knife, and is then shuttled within the mold set to the filling position.

The filling mandrels are comprised of a set of filling tips that are held within a protective air shower; this is a small area within the filling machine that is typically fed with sterile filtered air. When the molds are beneath the air shower, the filling tips are lowered into the neck of the partially formed container and the containers are filled. The mandrels then return to the protective air shower, and the containers are sealed by a second mold set (head mold), which forms the neck and closure of the BFS containers.

The entire cycle takes only a few seconds and therefore, results in minimal exposure of the open container to the surrounding clean room/air shower environment. The mold then opens and the filled containers surrounded by excess polymer are released. Excess plastic is then removed (typically this is done on-line by means of a mold specific cropping tool).

Liquid product is fed to the BFS machine from a holding tank or vessel. The product pathway is sterilized in place prior to receiving product, and product is sterilized by means of in-line sterilizing grade filters. Usually more than one stage of sterile filtration is on the product pathway.

FILLING ENVIRONMENT

Aseptic BFS machines are housed within classified clean areas of a minimum specification of class M5.5 (Federal Standard 209E) for 0.5 μm particles and greater (or equivalent), at rest. The new generation of BFS machines also is capable of operating with significantly decreased particle levels. The localized filling environment, or “air shower,” is of a higher classification, which meets the specification of class M3.5 (FS 209E) for 0.5 μm particles and greater.

Total particle levels should meet the required specifications and be measured, with the machine at rest, at defined

intervals by means of a laser particle counter (or other suitable instrument) to demonstrate continued compliance.

Levels of viable contamination, however, are of importance in operation. Microbiological monitoring for viable contaminants should be carried out to coincide with routine manufacture with normal levels of dynamic activity. As with traditional aseptic filling, viable contamination within the clean area should be controlled by means of an effective routine cleaning and disinfecting program and the adoption of appropriate clean room behaviors and practices by trained personnel. BFS technology has the advantage of being able to operate without continuous personnel presence within the clean area. However, operators will need to enter the area to start up the machinery and to attend to the machine as necessary to make routine adjustments. It is a requirement within the European forum that clean room garments worn to enter the class M5.5 (FS209E) clean room be of a standard appropriate for a higher (M3.5) classification clean room.

A routine microbiological environmental monitoring program should be established and documented based on historical and operational data to demonstrate continued compliance with specifications, as well as to monitor trends. A typical monitoring regime within the clean room would include quantitative air and surface monitoring. Semiquantitative air monitoring by the use of settle plates also is useful in supplying data associated with a longer period of time in operation (up to 4 h exposure). Recommended limits for viable contaminants (not specific to BFS processes) in clean rooms are quoted in various guidelines, including the current United States Pharmacopoeia (USP) and directive 91/356/EEC (MCA rules and Guidance for Pharmaceutical Manufacturers and Distributors 1997, Annex 1) (see Table 1). Alert and Action levels should be clearly defined based upon both operational data and published recommendations.

Consideration should also be given to monitoring the localized filling zone (air shower). Although access to this area will be prohibited (and also extremely dangerous) during operation, some monitoring for viable and nonviable contaminants may be possible at rest (e.g., at the end of a product batch). It may also be feasible to install a remote means of obtaining samples during operation.

MEANS OF BFS CONTAINER CONTAMINATION FROM THE ENVIRONMENT

As previously stated, for aseptic BFS, container filling occurs in a localized air shower provided with sterile filtered air. However, there is a short period of time between container formation and filling when the open container is transferred from the parison formation position to the filling position, and when the open container is exposed to the clean room environment. Therefore, it may be possible for contaminants from the room environment to enter the container during this shuttling period.

Air used to form the parison (parison support air) is typically sterile filtered air. If this is not the case, non-sterile air may be able to enter the parison during parison formation.

It was demonstrated during a simple practical experiment that broth filled units (totaling over 44,000) manufactured over several days in a highly contaminated environment remained sterile (1). The environment was contaminated by means of high levels of personnel activity in order to generate contaminants in keeping with those generated under normal conditions (albeit at grossly elevated levels).

During a more controlled study carried out within an environment artificially contaminated with high levels of individual nebulized spores of *Bacillus subtilis* (2), a level of contamination within the environment was achieved that led to the contamination of broth filled units. The results were extrapolated to suggest a contamination rate of 1 unit in 4×10^6 , with a surrounding environmental contamination of 1 cfu/m³.

Routes of air-borne contamination into BFS containers were investigated during a study using Sulfur hexafluoride (SF₆) tracer gas (3). During this experiment, the tracer gas was released at a known concentration into a clean room that housed an aseptic BFS machine. Levels of the tracer gas were then measured within subsequently filled BFS units. The study concluded that the container was effectively protected by the localized air shower. Although not necessarily representative of deposition of microbial

Table 1 EU and USP guidelines for clean room microbial limits

Classification	Air (cfu/m ³)	Settle plates (cfu/4 h)	Surface samples (cfu/55 mm diameter plate)
EC grade C	100	50	25
USP grade M5.5	20	not specified	10 (floors) 5 (other surfaces)

contaminants, there also was conclusive evidence of some room air within the BFS containers. The control of environmental contamination within the clean room is therefore important.

Extensive process simulation (broth fill) results for BFS effectively demonstrate that high levels of sterility confidence can be obtained with a properly configured and validated machine. However, in order to maintain high levels of sterility assurance, it is important that levels of microbial contamination are controlled within the filling environment.

CONTAMINATION FROM PRODUCT COMPONENTS

As with traditional aseptic filling, in order to comply with pharmaceutical good manufacturing practices (GMP), it is important to minimize contamination at all stages of manufacture. Raw materials should be of a high quality and tested for microbial contamination. Water used for product manufacture should be of low bioburden and high purity (preferably water for injection quality, although this requirement is dependent upon the nature of the product being manufactured).

A program of bioburden testing for each product batch at various stages of manufacture should be established and documented. This will be dependent upon the manufacturing process, but as a minimum should include bioburden analysis of bulk solutions prior to any sterile filtration. The maximum life of the bulk solution in a nonsterile environment (generally within a mixing tank) should be limited to prevent increase in bioburden beyond an acceptable level. Bioburden testing at this stage should be carried out on samples taken at the end of the holding period to give “worst case” data.

BFS technology often results in considerable machine down time, especially as associated with activities such as Clean In Place (CIP) and Steam In Place (SIP), in order to prepare a machine for manufacture. Initial machine adjustments will then be necessary in order for integral and cosmetically acceptable units of the correct fill volume to be consistently produced. It, therefore, can be advantageous to fill larger product batches once this is achieved. In order to facilitate this with respect to maintaining a low bioburden throughout all stages of liquid processing, it is a common practice to have a sterilized storage vessel into which bulk product is filtered through a sterilizing grade filter. This sterilized bulk solution can then be used to feed the filling machine without escalation of microbial levels. Further stages of

sterile filtration are required on the filling machine closer to the point of fill. A facility for sampling products during the course of the filling stage prior to further filtration can be incorporated. This will give data to confirm the low/zero bioburden of the product prior to the final stages of filtration, during the course of a longer batch.

The BFS container is produced from high-grade virgin polymer granules. Studies have investigated the lethality of the extrusion process with respect to container sterilization, the most recent of which is discussed in the validation section. Bioburden testing of polymer granules can be carried out in order to establish base line data. Virgin polymer granules, if handled and stored correctly, should be of very low bioburden.

EQUIPMENT—INTERVENTIONS AND MAINTENANCE

In order to produce sterile pharmaceutical products with a high degree of sterility confidence, it is of key importance that the equipment be operated by experienced and trained personnel with a full understanding of both the technology and aseptic processing. Operator intervention during machine operation is limited due to the nature of the technology; however, BFS machines are complex and some operator activity will be required from time to time during normal manufacture. Clearly documented rules are imperative in order to clarify which activities are prohibited during batch manufacture and which are permitted. For example, if a fault occurs that requires immediate corrective action involving the sterile product pathway, or within the direct vicinity of the filling zone, these would typically be prohibited activities that would lead to termination of the product batch. Activities such as parison and fill volume adjustments are part of the normal operation of the machinery and are permitted. A proceduralized means of documenting these activities should exist, however routine they may be.

Interventions should be categorized according to their potential for affecting the product being manufactured, and only those with no risk to product sterility should be permitted during operations.

As with all machinery, BFS machines must be properly maintained in order to maintain effective operation with the minimum of operator activity. A documented preventative maintenance program should be in place and specify appropriate frequencies for all machine components and associated systems and services. Maintenance activities should ensure that moving parts are sufficiently (but not overly) lubricated,

and that excess lubricants are removed at regular intervals to maintain the cleanliness of the machine. Abrasion among moving parts, particularly hoses and flexible pipe work, can be a problem with BFS machines and can cause undesirable particle generation and leaks that lead to unplanned maintenance and downtime. Moving parts should be inspected at regular intervals to avoid abrasion and to check for wear and tear. Regular seal changes with reconciliation of new/old seals should also be included.

Coolant systems are an integral part of container formation and serve to cool the molds and, if applicable, the parison clamp assembly. Coolant, although not in direct contact with product pathways, is in close proximity to the containers, and maintenance should be carried out to prevent coolant leakage. Coolant systems are prone to microbiological contamination and should be routinely treated to keep the bioburden under control. Coolant systems should be regularly sampled and tested for bioburden to ensure continuous compliance to a predefined specification.

VALIDATION OF BFS SYSTEMS

BFS machinery and associated equipment for aseptic manufacture should be constructed in such a way that the product pathways are of hygienic design with hygienic valves and minimal joints to facilitate cleaning and sterilizing in place.

Clean in Place (CIP)

As for all machinery involved in aseptic manufacture, CIP is necessary for all equipment that has product contact. This would typically include a bulk mixing tank, transfer lines, and the BFS machine itself, and may also include a holding vessel with associated transfer lines. CIP validation should be carried out to establish routine CIP practices that will clean the manufacturing equipment so that no contamination of subsequent products manufactured that would alter the safety, identity, quality, or purity of the drug beyond the predetermined requirements can occur. CIP procedures should be established by cleaning validation following the manufacture of worst case products (i.e., those that are most difficult to remove down to acceptable levels due to their solubility or activity). Means of measuring CIP efficacy include analysis of swabs taken directly from product contact machine parts and analysis of rinse waters. When establishing areas for swabbing, the specific equipment design needs to be taken into account, and those areas that

are potentially most problematic should be selected for analysis (e.g., filter housings or areas that may cause product hold-up).

Steam in Place (SIP)

Aseptic BFS machines are subject to SIP sterilization following standard CIP cycles. SIP cycles are routinely measured by thermocouples located in fixed positions along the product pathway. Validation of SIP cycles should be carried out to demonstrate that consistent sterilization temperatures are achieved throughout the equipment in order to prove that the system can be effectively sterilized. Validation should also identify suitable positions for routine use, or justify the fixed probe positions already in place. SIP validation is generally carried out using additional thermocouples and should include the use of Biological Indicators (appropriate for moist heat sterilization). Test locations should include areas that may be prone to air or condensation entrapment. An accurate engineering line drawing of the system in order to aid identification of suitable test locations and to document test locations selected should be available.

Qualification of Aseptic Filling

The standard and most appropriate method for the qualification of aseptic filling is by means of a broth fill (or media fill). Using this method, units of liquid microbiological growth media (usually a full strength general-purpose media, such as Tryptone Soy Broth), are filled and incubated. Following an appropriate incubation period, the units are inspected for contamination. In this way, an indication of the level of contamination during the filling process can be evaluated.

There is no appropriate defined sterility confidence level that can be translated directly into acceptance criteria for broth fill contamination for BFS processes. The most commonly recognized acceptance criteria is a sterility assurance level (SAL) of 10^{-3} , although it is accepted that modern aseptic filling techniques such as BFS can achieve a higher SAL and that this should be reflected by broth fill results and acceptance criteria for this recognized advanced technology.

Broth fills should be a major part of the operational qualification of a new BFS machine to demonstrate aseptic processing capability prior to product manufacture (typically three successful consecutive broth fills are required) and should be carried out at defined intervals thereafter.

Broth fills should be carried out under conditions that are representative of those during normal operation. If there is to be a deviation from routine processes, it should only be in the direction of presenting a greater, rather than a lesser, challenge to the process. Due to the level of automation of BFS technology, it is extremely difficult to take “extra care” in order to reduce the chance of container contamination during a broth fill; therefore, results are not as operator-dependent as other less automated aseptic manufacturing processes.

New facilities should contain some background environmental monitoring data. It is important that environmental monitoring data be obtained during the course of broth fill batches to demonstrate a normal level of environmental contamination—the validity of broth fill results carried out in an environment having consistently lower contamination levels than those obtained during routine batch manufacture could be questioned.

Batch manufacture, storage, and transfer should be carried out in accordance with routine procedures and with the same operators. The machine should be cleaned and sterilized as normal, although if an overkill cycle is used routinely for sterilization, a partial sterilization (although still meeting standard sterilization parameters) may be chosen as “worst case.”

Broth filled BFS units should generally meet all of the necessary product acceptance criteria, such as fill volume, wall thickness, container integrity, and cosmetic acceptability. The necessary operator activity at the start of a product batch is arguably more intrusive than at any other stage of manufacture. Product units routinely produced at the very start of a batch will usually be discarded due to fill volume, cosmetic, or other deficiencies as the machine set-up is adjusted. However, during a broth fill, it is a good practice to retain and incubate all start-up units (except any leaking units) to demonstrate that start-up activities have not affected product sterility. Such units should be segregated from the subsequent units that meet the acceptance criteria and labeled accordingly.

In addition, it can be useful to retain and incubate reject units filled during the course of a broth fill batch (again, excluding leaking units) for additional information. Again, these should be segregated from acceptable units and labeled accordingly. Although such units would be rejected during normal production, microbial contamination found in such units can be indicative of a problem that requires attention.

During the course of a broth fill, operator activity will be necessary as with routine manufacture. However, additional activities can be carried out to cover all permissible activities in order to provide evidence that

product sterility is not affected. Such interventions should be planned and documented with the batch documentation.

Frequency and size of broth fills must be clearly defined. Size of fill is usually based upon the statistical probability of detecting an acceptably low incidence of microbial contamination. Tables have been published to this effect (4), but the BFS operator must decide both the size and frequency of broth fills based upon their specific facility, routine product batch sizes, and operation. For high-speed BFS machines, filling routine product batches in excess of 100,000 units, relatively large broth fill batches, in comparison with traditional aseptic filling lines, are both feasible and appropriate.

The internal surfaces of broth filled units should be fully wetted to ensure capture of any contaminants within the broth. This is commonly achieved by agitation or inversion of the units either prior to or during the incubation period.

Incubation time and temperature should be such that macroscopic microbial growth of a wide range of common isolates will be detected. This should be routinely demonstrated by including positive control units inoculated with a low level of compendial microorganisms. It is desirable to perform additional testing to demonstrate that the incubation time and temperature selected will promote the growth of isolates obtained from machine operating environments. The Pharmaceutical BFS Operators Association recommends incubation of 14 days at 25–32°C.

Broth fill data from various BFS users were put together during a survey carried out in 1998 by the Pharmaceutical BFS Operators Association. These results, together with some more recent data, can be found in Table 2.

Some of the media fills carried out were full production batch volumes with hundreds of thousands of units filled in a single batch. In addition to the figures within the table, a run of over 1,500,000 units was recorded with the detection of a single contaminated unit.

It is clearly impractical to carry out very high numbers of broth filled units on a routine basis, but if unpreserved products are manufactured, and if practicable, it is good practice to fill broth directly following product batches with no further machine flushing or sterilization.

Given the high performance demonstrated during media fills, acceptance criteria should be based upon what can be realistically achieved. During broth fills of a standard size, any incidence of contamination among the units filled should lead to an investigation. In the absence of a cause, even with very low levels of contamination, consideration should be given to machine recommissioning.

Machine recommissioning should also be carried out if modifications to a filling machine have been made that

Table 2 BFS broth fill data

Company	A	B	C	D	E	F
Total number of units filled	6,462,570	222,900	2,697,496	1,534,626	1,042,254	31,600
Total number of nonsterile units detected	3	0	0	0	0	0
Contamination rate	0.0005%	0%	0%	0%	0%	0%

Positive units, assigned a definitive cause and unrelated to the BFS process as practiced during routine product manufacture, have not been included.

EC grade C is the closest classification to Federal Standard grade M5.5 (FS209E). In both cases, the limits specified are guidelines only and not regulatory requirements.

may have an effect on process capability (e.g., changes to the sterile product pathway or air shower).

BFS Containers

The BFS container is formed as an integral part of the process from medical grade virgin polymer granules. A recent study (5) investigated the lethality of the extrusion process when challenged with a high bioburden of spores. The spores of the test organism *Bacillus subtilis* var. *niger* were selected as they are known to be resistant to dry heat; the same strain was selected as the organism of choice for Biological Indicators used in dry heat sterilization processes. A series of broth fills were carried out using polymer batches inoculated with various levels of spores between 2×10^1 and 2×10^5 spores per gram. The broth filled units were then incubated in line with the company's routine broth fill procedure (25–32°C for 14 days). Spore contamination of units was observed with batches of polymer inoculated with high spore levels. The experiment demonstrated a relationship between polymer contamination and product contamination that was dependent upon both the level of contamination in the polymer and the resistance of the contaminant (in terms of D-value) to dry heat sterilization. The study also demonstrated inactivation of the spores on the granules with strong evidence of lethality associated with the extrusion process.

Routine bioburden testing of virgin pharmaceutical grade polymer granules tends to give very low or zero counts per gram of polymer tested, with contaminants generally much more heat labile than *Bacillus subtilis* spores. The study detailed was also carried out using a BFS machine adjusted to extrude at the lower end of the operating temperature range for extrusion. Therefore, it can be concluded that the extrusion process renders the contaminants unavailable, with sufficient bioburden reduction/inactivation for it to be appropriate for aseptic formation of BFS containers. This is further endorsed by routine broth fill data.

The closures of BFS containers are formed within the automated process by the head mold set which closes around the top of the severed section of parison following filling. The integrity of the container and closure is generally tested by a manual or automated method of leak detection performed outside of the filling environment following removal of excess plastic (deflashing) from the filled product units.

In order to minimize the number of leaking units produced, it is important that mold sets are correctly aligned. Very slight misalignment of molds may potentially lead to the production of units with very slight leaks that may be difficult to detect by routine methods. Therefore, correct molding is of key importance and usually can be checked easily by careful and experienced visual examination of units.

Container integrity testing can be carried out very effectively by a bacterial challenge test. Using this method, sterile broth filled units are submerged for a period of time (e.g., 24 hours) within a buffered solution that contains a high level bacterial challenge. (There are no regulations or guidelines that specify which organism to use, but it would seem logical to use a factory isolate or a relatively small organism such as a *Pseudomonas spp.*) Units are then removed, incubated, and checked for growth of the challenge organism. An absence of growth shows an integral unit and closure. This method is extremely sensitive and although this is not a test that is practical to perform on a routine basis, it can be a useful tool for infrequent use.

Filtration

Hydrophilic and hydrophobic sterilization grade filters are used throughout the BFS process for the sterilization of product and air, respectively. Filters should be purchased from an approved supplier and should be certified as meeting the regulatory requirements for sterilizing grade filters. By definition this means that the filter will have

full bacterial retention when subjected to an aqueous challenge of *Brevundimonas diminuta* (ATCC 19146) at a minimum concentration of 1×10^7 cfu/cm² of filter surface area.

Hydrophobic filters do not come into direct product contact and, therefore, the standard bacterial retention test alone generally is sufficient validation. However, as hydrophilic filters are in direct product contact, additional validation will be necessary for each product type in order to demonstrate that the filters selected for product sterilization do not alter the safety, identity, strength, quality, or purity of the drug product. Qualification of hydrophilic filters will also be necessary in order to demonstrate that the specific product type, in conjunction with a bacterial challenge, does not affect the efficacy of the filter. Validation of filters by means of bacterial retention tests requires specialist equipment and is often arranged between the filter manufacturer and the BFS operator.

SUMMARY

Aseptic pharmaceutical BFS technology for the manufacture of sterile liquid products demonstrates high levels of sterility assurance when correctly operated and configured. The technology is continually improving as more expertise is developed.

However, an understanding of the means of potential container contamination and the implementation of systems operating to minimize these means is important in order to maintain the high standards achievable with this technology.

REFERENCES

1. Jones, D.; Topping, P.; Sharp, J. Environmental Challenges to an Aseptic Blow-Fill-Seal Process: A Practical Study. *J. Pharm. Sci. Tech.* **1995**, 49 (5), 226–234.
2. Bradley, A.; Probert, S.; Sinclair, C.S.; Tallentire, A. Airborne Microbial Challenges of Blow-Fill-Seal Equipment: A Case Study. *J. Parenter. Sci. Tech.* **1990** July/Aug.
3. Whyte, W.; Matheis, M.; Dean-Netcher, M.; Edwards, A. Airborne Contamination During Blow-Fill-Seal Pharmaceutical Production. *PDA J. Pharm. Sci. Tech.* **1998**, May/June.
4. The Use of Process Simulation Tests for the Manufacture of Sterile Products. UK Parenteral Society Technical Monograph Number 4.
5. Birch, C.J.; Sinclair, C.S. Blow-Fill-Seal Extrusion of Spore Contaminated Polymer: An Exploratory Study. *BFS News* **1998**, Sept.

FURTHER READING

- Points to Consider for Pharmaceutical Blow/Fill/Seal Manufacturing, Sept. 1998; Draft.
- MCA Rules and Guidance for Pharmaceutical Manufacturers and Distributors, 1997.
- Revision to Annex 1 to the EU Guide to Good Manufacturing Practice, June 5, 1996.
- Sharp, J.R. Manufacture of Sterile Pharmaceutical Products Using Blow-Fill-Seal Technology. *Pharm. J.* **1997**, 239.
- Sharp, J.R. Validation of a New Form-Fill-Seal Installation: Principles and Practices. *J. Parenter. Sci. Tech.* **1990**, 44 (5).
- Sinclair, C.S.; Tallentire, A. Performance of Blow-Fill-Seal Equipment Under Controlled Airbourne Microbial Challenges. *J. Pharm. Sci. Tech.* **1995**, 49 (6).
- USPXXIII. *Microbiological Evaluation of Clean Rooms and Other Controlled Environments*; Ch. 1116.
- PDA Technical Report No. 22, Process Simulation Testing for Aseptically Filled Products. *J. Pharm. Sci. Tech. Suppl.* **1996**, 50 (S1).

CALORIMETRY IN PHARMACEUTICAL RESEARCH AND DEVELOPMENT

Sophie-Dorothée Clas

Chad R. Dalton

Merck Frosst Canada & Co., Pointe Claire-Dorval, Quebec City, Canada

Bruno C. Hancock

Pfizer Inc., Groton, Connecticut, U.S.A.

INTRODUCTION

Calorimetry is the measurement of energy changes within a material that are either manifested as exothermic (heat liberating) or endothermic (heat consuming) events (Table 1). Changes in energy (not absolute energies) are conventionally determined, and quantitative measurements may be made if the mass of the sample(s) is accurately known. Advanced and some custom-built instruments may permit the simultaneous monitoring of energetic changes along with other physicochemical events (e.g., changes in mass). These include, for example, microscopy Differential Scanning Calorimetry (microscopy-DSC) (1).

The most common applications of calorimetry in the pharmaceutical sciences are found in the “sub-fields” of DSC and Microcalorimetry. State-of-the-art DSC instruments and microcalorimeters are extremely sensitive and are powerful analytical tools for the pharmaceutical scientist.

DSC usually involves heating and/or cooling samples in a controlled manner, whereas microcalorimetry maintains a constant sample temperature. DSC instruments are considered to be part of the “Thermal Analysis” armamentarium; for additional information the reader should refer to the Thermal Analysis section of this book.

The beginning of this article gives a brief introduction into thermodynamics. A description of differential scanning calorimetry, which includes instrumentation, calibration, and applications, follows. A section on microcalorimetry is next, with a brief introduction into microcalorimetry, instrumentation, calibration, and applications. The article ends with a general comment on the regulatory aspects of calorimetry. A general description of the underlying physical or chemical transitions/reactions can be found in the section on DSC.

THERMODYNAMICS

The field of calorimetry relies on the principles of thermodynamics; the next section provides a brief overview of the general principles. References are provided for those who are unfamiliar with thermodynamics (2, 3).

A calorimeter consists of a container that is isolated from its exterior surroundings, where the heat exchange that occurs between the system and the environment can be measured. The “environment” is defined as the calorimeter and its contents, and the “system” is either a chemical reaction or physical change of state. The system can either absorb (endothermic) or lose energy (exothermic) to or from the environment. Exothermic changes will require the temperature of the environment to increase since it is the environment that is receiving the energy lost by the system. The energy of an isolated system remains constant and the energy exchange of the system must be equal but opposite in sign to the energy of the environment (First Law of Thermodynamics, Conservation of Energy). Endothermic changes in the system will involve a decrease in temperature of the environment since the environment is providing the energy absorbed by the system.

Based on the assumption that the system is closed, which is usually the case in differential scanning calorimetry and microcalorimetry, any reaction or change in state is independent of the path and can be subdivided into small reversible steps (Hess’ Law of Summation) (2, 3). The First Law of Thermodynamics states that energy may neither be created nor destroyed. It defines the internal energy, dU , as the sum of the change in heat that has been transferred to the system, dq , and the work done on the system, dw .

$$dU = dq + dw \quad (1)$$

When operated at constant pressure, Equation (1) can be written in terms of the enthalpy, H . The total energy exchange between the system and the environment, the

Table 1 Common thermal events that can be detected using calorimetric techniques

Event	Example
<i>Endothermic</i>	
Fusion	Melting of drug substances; purity evaluations
Vaporization	Evaporation of liquid or semisolid excipients
Sublimation	Removal of frozen water during lyophilization
Desorption	Drying of wet granulated formulations
Desolvation	Removal of stoichiometric water from crystalline hydrates
<i>Exothermic</i>	
Crystallization	Solvent vapor induced crystallization of amorphous excipients
Precipitation	Formation of salt forms of drug substances
Solidification	Melt granulation with semisolid excipients
Adsorption	Solvent vapor sorption by drug substances
<i>Chemisorption</i>	
Solvation	Water vapor sorption by excipients
Curing of resins	Curing of polymeric packaging materials
<i>Other</i>	
Glass transition	Variation of glass transition temperature with water content
Relaxation of glasses	Enthalpic recovery of amorphous drug substance upon storage or annealing
Decomposition	Thermal decomposition of drug substance
Dissolution	Dissolving drug substance in dissolution media
Complexation	Complex formation between drug and cyclodextrin

enthalpy change (dH), is the sum of the change in internal energy of the system, dU , and the change in the amount of work, PV :

$$dH = dU + PdV \quad (\text{at constant pressure}) \quad (2)$$

At zero net work and negligible change in volume (a close approximation for solids and liquids), the equation reduces to

$$(dU)_p = (dH)_p = (dq)_p \quad (3)$$

Thus, the enthalpy is effectively equal to the heat added or lost from the system, and changes in enthalpy can be measured directly in a calorimeter as dq (heat flow).

The heat exchange, dq , entering or exiting the system is equal to the change in enthalpy, dH , which is related to the heat capacity, C_p :

$$dq = dH = \int_{T_2}^{T_1} C_p dT \quad (4)$$

The increase in temperature of the system (from T_1 to T_2) is a function of its heat capacity. If C_p is large, then the transfer of a given amount of heat to a system results in only a small temperature increase.

DIFFERENTIAL SCANNING CALORIMETRY

During DSC, differences in heat flow between a sample and a reference are measured as a function of time and sample temperature. DSC analysis allows quantitative and qualitative information to be obtained about the physical and chemical changes that occur in a sample. DSCs are used extensively in the pharmaceutical industry to determine the melting points, purity, and glass transition temperatures of materials.

Instrumentation

Two types of differential thermal instruments, similar in practice but different in concept, are commercially available—differential thermal analyzers (DTA) and differential scanning calorimeters (DSC). Both techniques consist of a sample and reference pan, and provide quantitative and qualitative information on important physical phenomena of the sample (melting, purity, glass transition, etc.). These techniques differ both in sensitivity and the type of information that is recorded. DSC is a calorimetric technique where energy differences are measured as a function of time and temperature. DTA is solely a thermal technique where temperature differences and not energy are recorded as a function of time and temperature (Fig. 1a) (4, 5).

Two principal DSC designs are commercially available—power compensated DSC and heat flux DSC. The two instruments provide the same information but are fundamentally different. Power compensated DSCs heat the sample and reference material in separate furnaces while their temperatures are kept equal to one another (Fig. 1b). The difference in power required to “compensate” for equal temperature readings in both sample and reference pans are recorded as a function of sample temperature. Heat flux DSCs measure the difference in heat flow into the sample and reference as the temperature is changed. The differential heat flow to

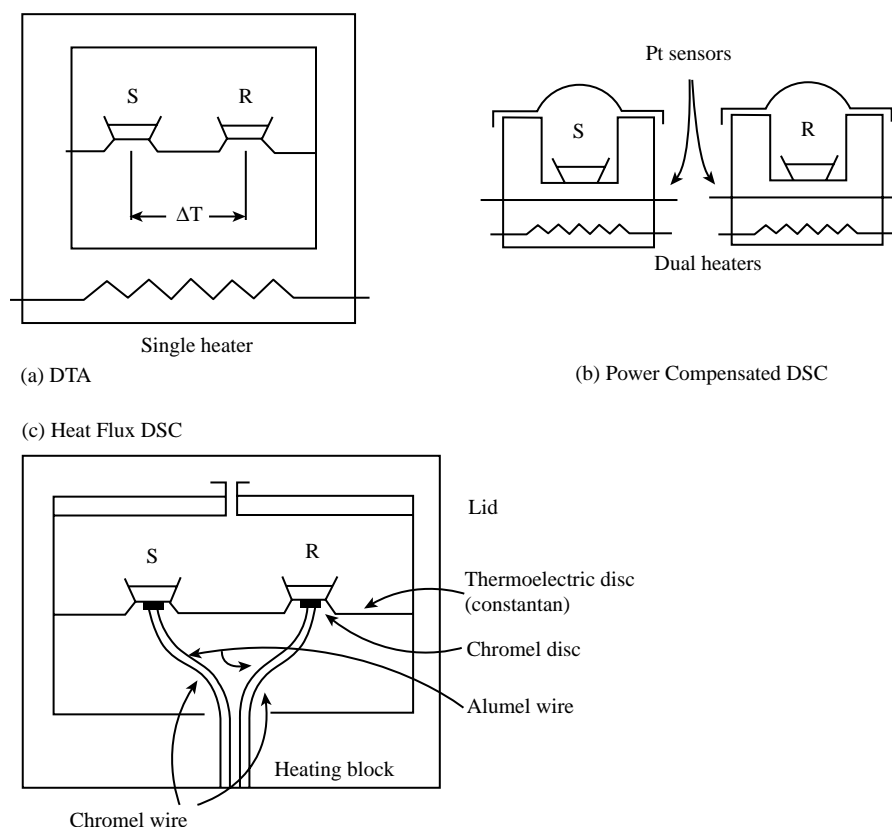


Fig. 1 Schematic diagrams of the (a) differential thermal analysis (DTA), (b) power-compensated differential scanning calorimeter, and (c) heat-flux differential scanning calorimeter cells (DSC) (adapted from DuPont Instruments Systems Brochure). (From Ref. 5.)

the sample and reference is monitored by chromel/constantan area thermocouples (Fig. 1c) (4, 5).

Modulated Differential Scanning Calorimetry/Dynamic Differential Scanning Calorimetry

Conventional DSC measures a sample's total heat flow. This total heat flow is comprised of a heat capacity component and a kinetic component (Eq. 5)

$$\begin{aligned} \text{total heat flow} &= \text{heat capacity component} \\ &+ \text{kinetic component} \\ dq/dt &= C_p dT/dt + f(T, t) \end{aligned} \quad (5)$$

where dq/dt = heat flow, C_p = heat capacity, dT/dt = temperature rate, $f(T, t)$ = heat flow from kinetic component as a function of temperature and time. Variations of conventional DSC have been used in order to extract additional information from these experiments. Two such techniques are Modulated Differential Scanning

Calorimetry (MDSC) and Dynamic Differential Scanning Calorimetry (DDSC). Unlike conventional DSC, MDSC and DDSC determine the total heat flow, dq/dt , and the heat capacity component of heat flow, $C_p dT/dt$. Equation (5) then leads to the indirect determination of the kinetic component of heat flow, $f(T, t)$.

MDSC, developed by TA instruments (New Castle, DE), is based on the conventional heat flux DSC furnace design. The temperature programs differ in that a sinusoidal modulation is overlaid on the conventional linear heating or cooling rate to produce a continuously changing nonlinear sample temperature. This can be viewed as running two experiments at once. The first experiment consists of heating the sample at a constant linear rate to obtain the total heat flow much like conventional DSC. During the second experiment, the heat capacity component of the heat flow is obtained by continuously varying the temperature sinusoidally with a zero net temperature change during the course of the modulation. The experimental parameters may be optimized by modifying three variables—the average heating rate, the period of modulation, and the temperature

amplitude of modulation. Fourier Transformation of the modulated heat flow signal is used to calculate an average heat flow value, which is similar to the total heat flow obtained by conventional DSC. The heat capacity is determined by the ratio of the heat flow amplitude to the modulated heating rate amplitude. The heat capacity heat flow is then obtained by multiplying the heat capacity by the average heating rate. The kinetic component heat flow is obtained by the difference between the total heat flow and the heat capacity component.

$$\begin{aligned} dq/dt = & C_p(dT/dt + A_T w^* \cdot \cos wt) \\ & + f'(t, T) + A_K(\sin wt) \end{aligned} \quad (6)$$

where $(dT/dt + A_T w^* \cdot \cos wt)$ = measured heating rate, $f'(t, T)$ = kinetic response without temperature modulation, A_K = amplitude of kinetic response to temperature modulation.

DDSC provides heat capacity and kinetic component information differently from MDSC. The temperature program consists of an "Iso-Scan" whereby the traditional heating rate program is combined with several isothermal holds or a "Heat-Cool" program, which consists of combined heating and cooling temperature programs. The user selects the appropriate method depending on the type of experiment being performed (6). From the dynamic component of the sample response, the complex heat capacity can be calculated. The complex heat capacity, C_p^* , is the vector sum of the storage, C_p' , and loss heat capacity, C_p'' . It is generally the same as the storage heat capacity except in the melting region where heat losses dominate. The storage heat capacity is associated with molecular motions within the sample in a manner similar to the storage modulus in dynamic mechanical measurements. The out-of-phase component, the loss heat capacity, C_p'' is associated with the dissipative properties

of the material. The loss heat capacity is out-of-phase with the temperature change because heat flow has resulted in molecular structural changes in the material. The loss tangent is the ratio of the loss heat capacity to the storage capacity and is a measure of the relative importance of each component (6).

$$C_p^* = C_p' + C_p'' \quad (7)$$

where C_p^* = complex heat capacity, C_p' = storage heat capacity, C_p'' = loss heat capacity. Some of the advantages and disadvantages of using MDSC are given in Table 2.

Sample Preparation and Calibration

DSC samples are generally analyzed in small metal pans that consist of inert or treated metals (aluminum, platinum, silver, stainless steel, etc.). Several pan configurations exist, such as open, pinhole, covered, or sealed. Reference pans should be made of the same material as the sample pan and in identical configurations. Typical DSC sample sizes are 3–5 mg for pharmaceutical materials. The material should completely cover the bottom of the pan to ensure good thermal contact. The pan should not be overfilled to prevent thermal lag from the bulk of the material to the sensor. Physically stable compounds that consist of large granular particles should be ground to reduce unwanted thermal effects. Accurate weights are imperative if quantitative data of the sample's energetic parameters are desired (5).

DSC is a scanning technique that usually is used for relative rather than absolute measurements. The meaningfulness of the results depends in the care taken in calibrating the instrument as close to the transition temperatures of interest as possible. The accuracy of any thermo-analytical instrument is strongly dependent on the

Table 2 Advantages and disadvantages of modulated DSC

Advantages	Disadvantages
Ability to differentiate overlapping transitions	More complex thermal lag effects
Increased resolution without loss of sensitivity	Not as precise linear heating rate
Measurement of heat capacity and heat flow in a single experiment	Many experimental parameters
Measurement of initial crystallinity	Not recommended for melting transitions.
Ability to study previous thermal history	Gives sample a complex thermal history
Ability to distinguish between reversible and nonreversible transitions	Sometimes difficult to interpret

use of high purity calibration standards. Well-defined standards are especially important when analyses are carried out using different instruments and at different times. In general, metal calibration standards, such as indium, tin, bismuth, and lead, are utilized due to ready availability and ease of use. Low melting metals, such as mercury and gallium, are used to a lesser extent because of toxicity and handling problems. Organic compounds have been recommended as standards when studying organic material to minimize differences in thermal conductivity, heat capacity, and heat of fusion (7–9). It is likely that metals will continue to be popular temperature and enthalpy standards due to availability and ease of use, and organic standards may be used predominantly at temperatures below 300 K (10).

DSC results are dependent on the calibration of the instrument, sample preparation, and sample configuration. Some researchers argue that power compensated DSCs need to be properly calibrated upon both heating and cooling at the same rates in order to maintain a high level of accuracy (11, 12). Standard procedures can be obtained from the American Society for Testing of Materials (ASTM). In addition, all results obtained by DSC are a function of the scanning rate used and should be reported with the scanning rate. The shape, the area of the transition or change of baseline, and particularly the temperature of the transition will be dependent on the scanning rate; it will move to higher temperatures with increasing heating rate. Increasing the scanning rate increases sensitivity, while decreasing the scanning rate increases resolution. To obtain thermal event temperatures close to the true thermodynamic value, slow scanning rates should be used (e.g., 1–5 K/min).

Definitions and Applications of DSC

The purpose of this section is to define the various parameters that are measured by DSC. The types of thermal events, exothermic or endothermic, that can be measured by DSC are reported in Table 1. The following sections will describe some of the more fundamental thermal events. Examples from the pharmaceutical field will be given to illustrate the techniques. The examples will be based on either single components, such as drug substance and bulk excipients, or on a mixture of components, such as physical blends of drugs and excipients, solid dispersions, formulated drugs after granulation, and/or compression.

Melting

Melting is a first order endothermic process by which the compound takes in a net quantity of heat (molar heat of

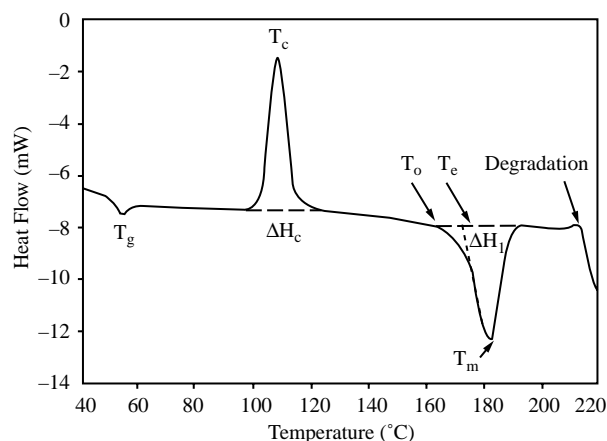


Fig. 2 DSC scan of sucrose showing the glass transition temperature, (T_g), recrystallization exotherm temperature (T_c) and enthalpy (ΔH_c), onset of melting (T_o), extrapolated melting onset (T_e), peak melting temperature (T_m), enthalpy of fusion (ΔH_f), and onset of degradation at 10 K/min. Endothermic transitions are shown down.

fusion). Through DSC, melting can be seen as an endothermic peak (Fig. 2). The broadness of the peak defines the purity of the crystalline compound undergoing melting, with the less pure and less perfect smaller crystals melting first followed by melting of the purer larger crystals (5). The melting temperature is the temperature at which the 3-dimensionally ordered crystalline state changes to the disordered liquid state. It is defined either as an extrapolated melting temperature onset, T_e , obtained at the intersection of the extrapolated baseline prior to the transition with the extrapolated leading edge, or as the peak melting temperature, T_m . Other temperatures that describe the melting process are the onset of melting, T_o , and the extrapolated end of the transition (5).

The enthalpy of fusion, ΔH_f , is obtained from the area of the endothermic transition. The area of the transition is affected by the selection of the baseline. The baseline is generally obtained by connecting the point at which the transition deviates from the baseline of the scan to where it rejoins the baseline after melting is completed. For some materials that undergo a significant change in heat capacity change on melting, other baseline approximations (such as a sigmoidal baseline) are used (5).

Purity

The purity of crystalline compounds can be calculated using the van't Hoff equation from the enthalpy of fusion and melting temperature obtained by DSC.

$$T_{s(i)} = T_e - RT_e^2 X / (\Delta H_f F_i) \quad (8)$$

where $T_{s(i)}$ is the sample temperature at equilibrium corrected for thermal lag effects (K), T_e is the melting temperature of the pure compound (K), R is the gas constant (8.314 J/mol/K), X is the molar fraction of impurity, ΔH_f is the enthalpy of fusion of the pure compound (J/mol), and F is the fraction of the sample that is molten at $T_{s(i)}$. The melted fraction is equal to the area of the section melted (A_i) divided by the total area of the melting endotherm (A_T) as shown in Fig. 3. The melting depression, $(T_e - T_{s(i)})$ is equal to the slope, $(RT_e^2/\Delta H_f)X$, of the straight line obtained when $T_{s(i)}$ is plotted as a function of $1/F_i$. The theoretical melting temperature is obtained on extrapolation to $1/F_i = 0$. A straight line may not be obtained due to thermal lag, sensitivity, lack of a eutectic point detection, and formation of solid solutions. In addition, a significant amount of material may have melted before a measurable heat flow is observed by DSC. As a result, a correction constant K_{corr} is added to the measured areas (each fraction) to correct the curvature of the plot of $T_{s(i)}$ as a function of $1/F_i$ (Eq. 9). The melting depression $(T_e - T_{s(i)})$ is then obtained when $F_i = 1$ (5).

$$1/F_i = (A_T + K_{\text{corr}})/(A_i + K_{\text{corr}}) \quad (9)$$

It is necessary that the melting curve is obtained with a calibrated DSC using small samples (1–3 mg) and slow scanning speeds ($<5 \text{ K/min}$, preferably 2 K/min). The

purity of a compound should be determined at several scanning speeds to ensure that the compound does not undergo any solid-solid transitions, such as polymorphic conversion or degradation (5).

The three advantages of obtaining purity by DSC are: 1) its speed of measurement; 2) the type of impurity does not have to be known; and 3) a minimal amount of sample is required. However, in the case of salts, excess base or acid is counted as an impurity.

DSC calculations of drug purity were used to assess the quality of progesterone and lipoic acid, and to define specifications for the drugs (13). The enantiomeric purity of (–)ephedrinium 2-naphthalenesulphonate has also been determined using DSC (14).

Crystallization

Crystallization can occur on cooling from the melt and/or heating above the glass transition temperature of amorphous materials. The temperature at which this occurs is the crystallization temperature, T_c . Through DSC, crystallization is observed as an exothermic transition with an enthalpy of crystallization, ΔH_c (Fig. 2). The energy released when the molecules, atoms, or ions organize into a 3D solid state is related to the crystal lattice energy. Some compounds can crystallize into different molecular arrangements called polymorphs, discussed later.

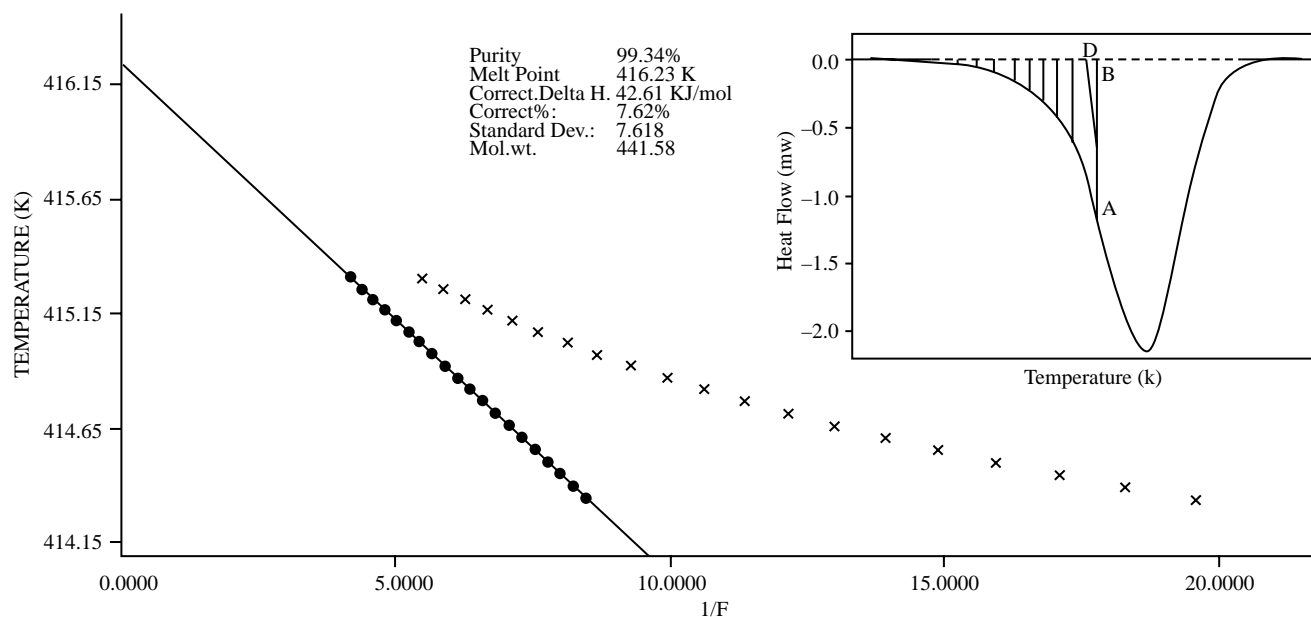


Fig. 3 DSC scan of drug substance divided into segments, A_i , for purity calculations of a compound of total enthalpy or area, A_T (inset, only a few segments are shown). The van't Hoff plot of the temperature of each segment as a function of $1/F = 1/(A_i/A_T)$, and with correction (straight line) for determination of purity of the drug substance.

Quantification of Crystallinity

The crystallinity of drugs and excipients before and after formulation processing can be determined using calorimetry. In some cases, crystalline compounds can convert during pharmaceutical processing to the amorphous form, which is a thermodynamically less stable form. Amorphous compounds consist of nonordered molecules (see the section on *Glass Transition*). This can have important implications for the chemical and physical stability of the formulations. The effect of grinding on the crystallinity of cefixime trihydrate was evaluated using DSC and other thermal techniques. The dehydration temperature of the ground sample decreased linearly with decreasing crystallinity (15).

An exothermic transition can sometimes be observed by DSC on crystallization of the amorphous form. This can be used to quantify the amorphous content of crystalline drugs. A calibration curve that consists of a plot of the enthalpy of crystallization as a function of crystalline content was used to determine if the lyophilized MK-0591 drug substance was completely amorphous or contained some crystalline compound (16).

Polymorphism

Polymorphs are crystalline compounds of the same molecular structure that have a different arrangement of molecules in the unit cell.^a Polymorphs have the same chemical composition but have unique cell parameters. Therefore, polymorphs can have very different melting temperatures, densities, solubilities, chemical and physical stabilities, dissolution rates, and bioavailabilities (17).

Polymorphs are either enantiotropic or monotropic. Enantiotropic polymorphs have a thermodynamic conversion temperature where one form is more stable above this temperature while the other is more stable below. Processing the least stable form, dissolution/recrystallization, and certain storage conditions might cause enantiotropic polymorphs to later convert (5). If there is no conversion temperature below the melting temperatures of the polymorphic pair, then the different crystal forms are monotropic. That is, there is only one crystal form that is thermodynamically stable at all temperatures and pressures. Calorimetry can be used to determine which polymorph is the more stable form. DSC can provide accurate unambiguous melting temperatures and enthalpies of fusion. Based on the melting temperature and the enthalpy of fusion, the relative thermodynamic stability of the polymorphic

pair can be determined [e.g., using the Heat of Fusion Rule (5)].

DSC and complimentary thermal techniques, such as temperature X-Ray powder diffraction, were used to determine the thermodynamic relationship of the six anhydrous polymorphs of tetracaine hydrochloride (17). The phase diagram of the polymorphic conversion of diflunisal in polyethylene glycol (PEG) 4000 solid dispersions was obtained as a function of polymer content (18).

Heat Capacity

Accurate heat capacity, C_p , measurements may be obtained by DSC under strict experimental conditions, which include the use of calibration standards of known heat capacity, such as sapphire, slow accurate heating rates (0.5–2.0 K/min), and similar sample and reference pan weights (19, 20). MDSC or DDSC also have been used to determine the heat capacity of several pharmaceutical materials (6, 21).

Glass Transition

By the use of various pharmaceutical manufacturing processes, (e.g., lyophilization or comminution techniques), drugs or excipients may be made amorphous. Amorphous compounds are defined by their lack of long-range molecular order and structural periodicity. Their high energy state is of great interest to the pharmaceutical industry as it can lead to fast dissolution rates and increased bioavailabilities. However, amorphous compounds are thermodynamically unstable, although depending on their glass transition temperature, they may be kinetically stable for extended times. Amorphous compounds are characterized by a glass transition, which by DSC is seen as an increase in heat capacity change (Fig. 2):

$$\Delta C_p = C_{p_{\text{liq}}} - C_{p_{\text{glass}}} \quad (10)$$

where $C_{p_{\text{liq}}}$ is the heat capacity of the liquid, and $C_{p_{\text{glass}}}$ is the heat capacity of the glassy phase. The glass transition temperature is measured either at its onset or midpoint as shown in Fig. 4. Structural relaxation can occur due to the restricted but finite mobility of the molecules below the glass transition. This gradual volume or enthalpy change is observed by DSC as an endothermic peak superimposed on the glass transition (Fig. 4), and this may lead to difficulties in interpretation of the transition. Modulated DSC can sometimes be used to separate the enthalpic overshoot from the glass transition temperature (5).

^aSee *Polymorphism: Pharmaceutical Aspects*, page 2249.

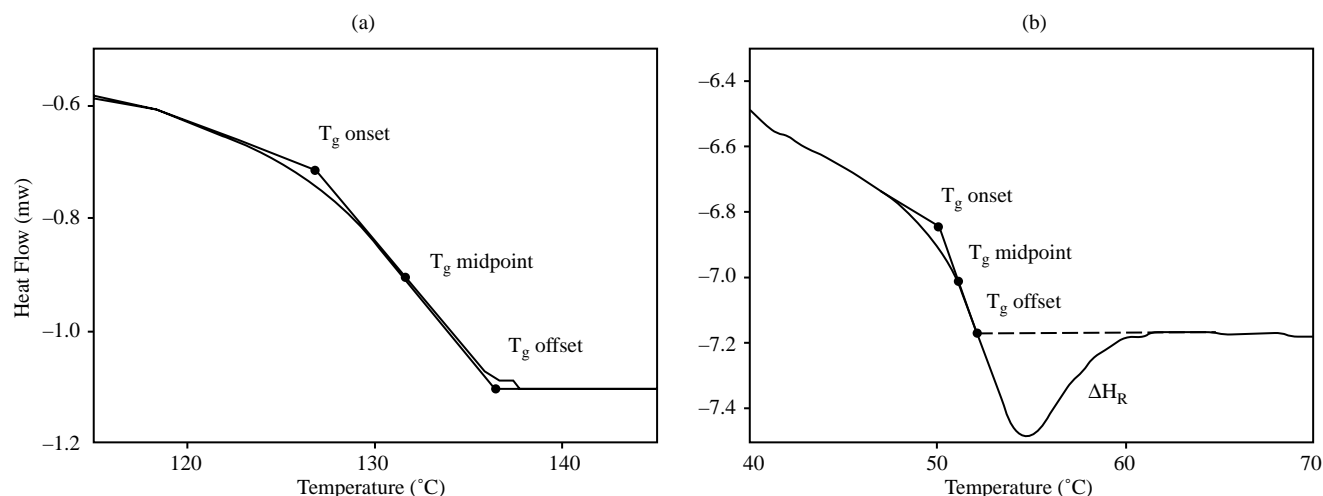


Fig. 4 (a) DSC scan of the glass transition temperature of a miscible blend of a MK-0591 with 10% PVP, showing the onset, midpoint, and offset glass transition temperatures. (b) DSC scan of the glass transition of sucrose with the enthalpic relaxation endotherm and enthalpy (ΔH_R). (From Ref. 5.)

Defining the glass transition temperature is important to the development of stable amorphous pharmaceutical materials. A leukotriene biosynthesis inhibitor, MK-0591, has been shown to be kinetically stable in the amorphous phase at normal storage temperatures if protected from moisture, due to its elevated glass transition temperature of 125°C (16). In lyophilized systems, a high T_g' , defined as the apparent glass transition temperature and observed as the change of the heat capacity of the lyophilized formulations, is important to define the stability of such formulations (22–24). MDSC has been used to select optimal freeze-drying conditions to avoid cake collapse (25, 26).

The glass transition temperature of amorphous multicomponent mixtures can be used to determine the miscibility of the components. If the mixture is miscible, then a single glass transition temperature is usually obtained. Various equations can be used to predict the glass transition temperature of miscible mixtures. Examples include the Gordon-Taylor equation (Eq. 11) or the Fox-Flory equation (Eq. 12).

$$T_{g\text{ mix}} = [w_1 T_{g1} + K w_2 T_{g2}] / (w_1 + K w_2); \quad (11)$$

$$K = \rho_1 T_{g1} / \rho_2 T_{g2}$$

$$1/T_{g\text{ mix}} = 1/T_{g1} + 1/T_{g2} \quad (12)$$

where ρ_1 and ρ_2 are the densities of the two components and T_{g1} and T_{g2} are their respective glass transition temperatures (27).

Temperature Dependence of Molecular Motions in Amorphous Materials

A critical attribute that dictates the stability and performance of any amorphous material is the manner in which its rate of molecular motions (τ) varies with changing temperature (T) (i.e., $d\tau/dT$). At temperatures which are approximately 0–100 K above the calorimetric glass transition temperature (T_g), this property is known as the *fragility* of the material (27). Several workers have suggested that $d\tau/dT$ below T_g (in the nonequilibrium glassy state) is the most appropriate descriptor of amorphous pharmaceutical materials since this is the normal state for the storage and processing of such systems (27, 28). A simple graphical plot of τ versus T can be constructed at temperatures below T_g from the results of enthalpy relaxation experiments. These measurements can be performed using either a conventional DSC or microcalorimeter (28, 29). Alternate calorimetric methods of estimating $d\tau/dT$ at T_g have been described in the literature (30), but the applicability of these methods to pharmaceutical materials has not yet been clearly demonstrated (31).

Degradation, Decomposition, Stability Determinations, and Drug-Excipient Compatibility

The degradation, decomposition, and stability of drugs or formulations can be determined by DSC or microcalorimetry. The advantages of the techniques are their speed of measurement and the small amounts of sample required.

At times, interpretation of the results can be difficult, particularly when simultaneous reactions occur. Decomposition kinetics are generally determined using the Arrhenius equation. The samples are stored at elevated temperatures for known periods of time and analyzed by DSC. Alternatively, they can be held isothermally in the DSC at different temperatures, followed by scanning at heating rates sufficiently fast to avoid additional decomposition. A rate constant is calculated for each storage condition by plotting the logarithms of the areas of the transitions (e.g., the decomposition endotherm, etc.) as a function of time. The natural logarithm of the reaction rates, k , are then plotted as a function of $1/T$ as per the Arrhenius equation.

$$k = Ze^{-(E_a/RT)}; \quad \ln k = \ln Z - E_a/RT \quad (13)$$

where Z is the Arrhenius frequency or pre-exponential factor, E_a is the Arrhenius activation energy (J/mol) for the reaction, and R is the gas constant. The activation energy and pre-exponential factor are assumed to be constant and independent of temperature.

Alternatively, the reaction peak maxima may be determined at different heating rates (φ) and used to calculate the activation energy, assuming first order kinetics.

$$E_a = -2.19 T \frac{d \log \varphi}{d(1/T)} \quad (14)$$

The energy of activation is obtained from the slope of the log of the heating rate (φ) as a function of $1/T$ (31). It is assumed that only one reaction occurs during the transition and that the peak maximum represents a point of constant conversion for each heating rate. The method cannot be used with compounds that decompose on melting or undergo isomerizations at the reaction temperature or any other simultaneous reaction (32). Some modification of DSC may be needed to determine the degradation kinetics of compounds under different environmental conditions.

DSC is often used for the rapid screening of excipients for drug-excipient compatibility studies. Certain assumptions have to be made, which include that the thermal properties of these mixtures are the sum of the individual components when there are no interactions between the components. The method does not take into consideration: 1) effects due to thermal conductivity (thermal lag effects); 2) mixing effects that can lower the purity of each component resulting in slightly broader, lower melting temperatures; or 3) sample geometry effects that result in variations in peak shapes and peak temperatures. In addition, reduction in enthalpies of fusion can occur as a result of the solubilization of the drug in molten excipients. This latter phenomenon can be used in

part to determine the solubility in different molten excipients (5, 33).

Interactions with Water/Solvents, Hydrates

Water can have a significant impact on the physical and chemical stability of drugs. Water may be present as part of the crystalline lattice (hydrate), or it may be on the surface ("free") or more tightly incorporated ("bound"). The evaluation of the type of water present in a pharmaceutical material has been determined using subambient DSC (thermoporosimetry), such as in the case of magnesium stearate hydrates (34, 35), as well as thermogravimetric techniques. Free or surface water can crystallize and the melting enthalpy of this free water can be used to calculate the surface water content of compounds from the melting enthalpy of pure water (36). The state of water in HPMC gels with and without drugs, such as propranolol hydrochloride or diclofenac, was determined in this way by DSC (37).

MICROCALORIMETRY

Microcalorimetry is used to monitor thermal changes associated with physical and/or chemical events that do not require heating or cooling for their initiation. Such events include dissolution, precipitation, reaction, and crystallization. In a typical microcalorimetry experiment, these events are "triggered" in a controlled manner by mixing two pre-equilibrated and separate phases (e.g., water vapor and amorphous drug, solvent and crystalline drug, or protein and carbohydrate solutions). Microcalorimetry techniques are sometimes referred to by the processes that are monitored (e.g., immersion calorimetry, solution calorimetry, titration calorimetry, etc.) (38). Differential scanning calorimeters may be operated in isothermal mode; however, for highly accurate and reliable isothermal measurements, specially designed microcalorimeters are required.

Thermodynamics

Solution microcalorimetry may be used to determine the free energy of dissolution of a solid compound, which is particularly important in pharmaceutical research for dissolution studies and in the determination of the relative thermodynamic stability of polymorphs (38). The change in the Gibbs–Helmholtz free energy, ΔG_{sol} , on dissolution is

$$\Delta G_{\text{sol}} = -RT \ln K_{\text{eq}} \quad (15)$$

where T is the temperature (Kelvin, K), R is the gas constant ($8.314 \text{ J mol}^{-1} \text{ K}^{-1}$), and K_{eq} is the equilibrium constant for the change of the compound from the solid state to the dissolved liquid state. The equilibrium constant can be determined, at low concentrations, from the ratio of the concentration of the compound in the solution, or its solubility, to that in the solid state (where by definition, $[C]_{\text{solid}} = 1$).

$$K_{\text{eq}} = [C]_{\text{soln}}/[C]_{\text{solid}} = [C]_{\text{soln}} \quad (16)$$

The ΔH_{sol} is the enthalpy change that occurs on dissolution of one mole of compound in a solvent. The solution microcalorimeter may be used to obtain the enthalpy of solution directly. The change of free energy can be calculated from the concentration using the enthalpy obtained. The change in the entropy of solution ΔS_{soln} can then be determined from the Gibbs–Helmholtz equation (2, 3).

Alternatively, the change in free energy of solution ΔG_{soln} can be calculated from the van't Hoff equation:

$$\delta(\Delta G_{\text{soln}}/T)/\delta T = -\Delta H/T^2 \quad (17)$$

It may be more accurate to obtain the enthalpy directly from microcalorimetry (39).

Instrumentation

The simplest type of nonscanning calorimeter is the isoperibol instrument (Fig. 5). In this type of calorimeter, a constant environment is maintained inside an insulated reaction vessel. Typically, a silvered dewar is used, and the interacting components (e.g., solvent and solute) are held in sub-containers. Liquid phases are usually stirred and the temperature is accurately recorded using a thermometer or thermocouple. At the start of the experiment, the reaction vessel is allowed to reach a steady state, and a baseline temperature or temperature drift is recorded. The interaction of interest is then initiated by permitting the two components to mix, and the resulting temperature increase from baseline is recorded. The system is calibrated by monitoring a standard reaction (e.g., neutralization of hydrochloric acid) or by applying a controlled amount of electrical energy via a heating coil. Isoperibol calorimeters can be easily constructed from their individual components, and several different instruments are commercially available at a modest cost.

Due to the high degree of accuracy and sensitivity that is required for pharmaceutical analysis, more sophisticated microcalorimeters are frequently used for studying

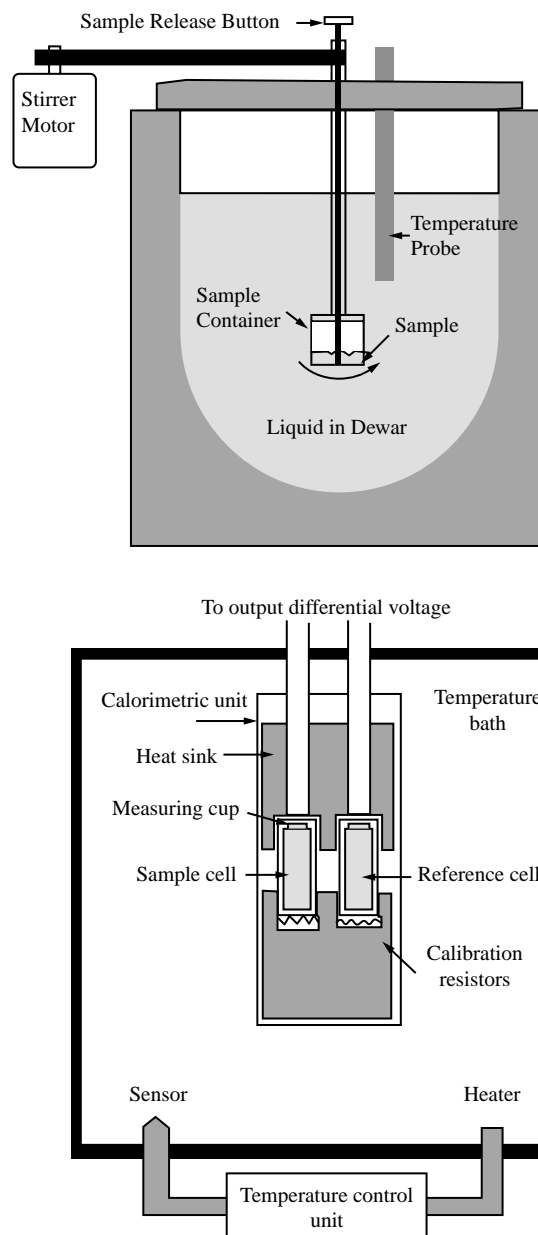


Fig. 5 Schematic diagram of an isoperibol calorimeter (top) and a thermal activity monitor (bottom).

pharmaceutical systems. An example of such an instrument is the Thermal Activity Monitor (TAM) manufactured by Thermometrics (Sweden). In this instrument, twin sample cells are used (one for the sample and one for a reference) to achieve greater signal stability and to minimize the effects of spurious thermal fluctuations (Fig. 5, bottom). The cells are housed in a constant temperature environment maintained via a sophisticated heater and water jacket system. Minor changes in sample heat flow are detected with relative ease using this

arrangement. The TAM is calibrated electrically, and the commercially available sample configurations allow the mixing of solids, liquids, and gases in various proportions. Controlled gas and liquid flow rates, and changing sample environments (e.g., relative humidities) can also be achieved with appropriate accessories. A major practical advantage of this type of calorimeter when compared with less sophisticated instruments is the small sample size requirement of only a few tens or hundreds of milligrams per determination. The TAM microcalorimeter has been used for detection and monitoring of crystallization events, sorption and desorption of organic and inorganic vapors, chemical reactions (drug degradation and drug interactions with excipients), molecular motions in amorphous pharmaceutical materials, ligand binding phenomena, microbiological growth, and the determination of solid heat capacities.

Sample Preparation and Calibration

Microcalorimeters are usually operated in a similar way irrespective of the source of the energy change that is being monitored. Specimens are pre-equilibrated at the desired measuring temperature for several hours and then introduced into the calorimeter chamber. After a short delay, the external stimulus is applied to trigger the event of interest, and then the energy that is liberated or consumed is measured. The reaction is isolated from the environment by a jacket, which serves as a thermal shield to minimize the absorption and emission of radiant heat (Fig. 5). Calibration of microcalorimeters is usually achieved by direct heating using an electric heating element (as mentioned previously). Extreme care is required to achieve consistent sample preparation and to maintain constant experimental procedures since the interpretation of results can be confused easily by experimental artifacts. Simultaneous thermal events of opposite sign (exothermic and endothermic) are quite common and may often confound the interpretation of data. In all experiments, an appropriate thermal reference is required since the energy changes that are measured are simply energy changes relative to the reference specimen. Common references include an empty sample container, or a sample container filled with an inert material which has a similar heat capacity and mass to the sample.

Definitions and Applications of Microcalorimetry

Microcalorimeters have found widespread use in the pharmaceutical sciences in recent years for applications as diverse as determining the degradation rate of drugs,

estimating the strength of binding between proteins and receptor sites, and monitoring metabolic processes in microorganisms.

Hollenbeck used immersional calorimetry to investigate interactions between microcrystalline cellulose and water (39). Interactions of hydroxypropyl methylcellulose and cholestyramine with water also have been investigated by microcalorimetry (40, 41). In addition, the desorption of water from theophylline monohydrate has been investigated using microcalorimetric approaches (42). The properties of surfactants and surface active drugs in solution were studied by Attwood et al. (43) using calorimetry, while titration microcalorimetry has been utilized to elucidate the nature of specific interactions in several pharmaceutical polymer-surfactants systems (44, 45). A more unusual pharmaceutical use of microcalorimetry is to study energetic changes that occur during tablet compaction (46, 47). The compression calorimeter is a custom-made research instrument that appears to have many potential applications for the pharmaceutical scientist.

Microcalorimetry has proven to be a particularly useful tool to detect different levels of disorder in pharmaceutical materials (48). Pikal et al. used a solution calorimetry approach to measure differences in the crystallinity of a wide range of antibiotic samples (49). Gao and Rytting demonstrated the validity of this approach for a wider range of materials (50). Other workers have used elevated vapor pressures to trigger crystallization of disordered materials in the calorimeter and have been able to use the measured energy output to directly quantify the levels of disorder crystallinity in their samples (51).

Stability studies using microcalorimetry are widely reported in the literature (38, 52–54). Some authors have monitored exothermic degradation reactions over several days or weeks and have projected the degradation extent and rate over the shelf-life of the drug or drug product. Other authors have used microcalorimetry to monitor relaxation of amorphous pharmaceutical materials and have then calculated relaxation time constants from these data for use in shelf life predictions (29). The use of microcalorimetry for preformulation stability screening of a drug with potentially reactive excipients has also been described (55).

REGULATORY CONSIDERATIONS

Calorimetric methods are infrequently used for routine quality control purposes because of their nonspecific nature and relatively slow speed. However, data from calorimetry experiments are commonly presented in applications for

new product licenses and in support of patent applications. To ensure the integrity of all calorimetry data, normal procedures for good laboratory practices, standard operating procedures, appropriate calibration methods, and regular instrument servicing are necessary. The use of DSC for the measurement of transition temperatures and sample purity is described in the United States Pharmacopoeia, and standard procedures for DSC analyses are also suggested by the ASTM (100 Barr Harbor Dr., West Conshohocken, Pennsylvania 19428).

ACKNOWLEDGMENTS

The authors wish to thank Dr. Karine Khougaz for the figures on the glass transition temperature. The assistance of the Merck Frosst Canada & Co. librarians, Ms. M.-L. Gaal and C. Taylor, is gratefully acknowledged.

REFERENCES

- Giron, D. Thermal Analysis, Microcalorimetry and Combined Techniques for the Study of Pharmaceuticals. *J. Therm. Anal. Calorim.* **1999**, *56*, 1285–1304.
- Yu, L. Inferring Thermodynamic Stability Relationship of Polymorphs from Melting Data. *J. Pharm. Sci.* **1995**, *84*, 966–974.
- Wunderlich, B. The Basis of Thermal Analysis. *Thermal Characterization of Polymeric Materials*, 2nd Ed.; Turi, E., Ed.; Academic Press: San Diego, CA, 1997; 206–484.
- Skoog, D.A.; Leary, J.J. *Principles of Instrumental Analysis* 4th Ed.; Saunders College Publishing: New York, 1992; 568–578.
- Clas, S.-D.; Dalton, C.R.; Hancock, B.C. Differential Scanning Calorimetry: Applications in Drug Development. *Pharm. Sci. Technol. Today* **1999**, *2*, 311–320.
- Varma-Nair, M.; Wunderlich, B. Nonisothermal Heat Capacities and Chemical Reactions Using a Modulated Differential Scanning Calorimeter. *J. Therm. Anal.* **1996**, *46*, 879–892.
- Barker, A.D. A New Approach to Differential Scanning Calorimetry in Wax/Polymer Blending. *J. Therm. Anal.* **1993**, *40*, 799–805.
- Price, D.M. Temperature Calibration of Differential Scanning Calorimeters. *J. Therm. Anal.* **1995**, *45*, 1285–1296.
- Tan, Z.C.; Sabbah, R. Reference Materials for Energy and Temperature Calibrations of DTA and DSC instrumentation in the range 100–300 K. *J. Therm. Anal.* **1994**, *41*, 1577–1592.
- Skoglund, P.; Fransson, A. Accurate Temperature Calibration of Differential Scanning Calorimeters. *Thermochim. Acta* **1996**, *276*, 27–39.
- Sarge, S.M.; Hemminger, W.; Gmelin, E. Metrologically Based Procedures for the Temperature, Heat and Heat Flow Rate Calibration of DSC. *J. Therm. Anal.* **1997**, *49*, 1125–1134.
- Gehenot, A.; Rao, R.C.; Maire, G. Value of Thermal Analysis in the Critical Evaluation of Classical Methods of Melting Point Determination. *Int. J. Pharm.* **1988**, *45*, 13–17.
- Lvova, M.S.; Kozlov, E.I.; Wishnevezkaya, I.A. Thermal Analysis in the Quality Control and Standardization of Some Drugs. *J. Therm. Anal.* **1993**, *40*, 405–411.
- Duddu, S.P.; Grant, D.J.W. The Use of Thermal Analysis in the Assessment of Crystal Disruption. *Thermochim. Acta* **1995**, *248*, 131–145.
- Kitamura, S.; Miyamae, A.; Koda, S.; Morimoto, Y. Effect of Grinding on the Solid-State Stability of Cefixime Trihydrate. *Int. J. Pharm.* **1989**, *56*, 125–134.
- Clas, S.-D.; Cotton, M.; Moran, E.; Spagnoli, S.; Zografi, G.; Vadas, E.B. Assessment of the Physical Stability of Lyophilized MK-0591 by Differential Scanning Calorimetry. *Thermochim. Acta* **1996**, *288*, 83–96.
- Giron, D.; Draghi, M.; Goldbronn, C.; Pfeffer, S.; Piechon, P. Study of the Polymorphic Behavior of Some Local Anesthetic Drugs. *J. Therm. Anal.* **1997**, *49*, 913–927.
- Martinez-Oharritz, M.C.; Martin, C.; Goni, M.M.; Rodriguez-Espinosa, C.; Tros-Iarduya, M.C.; Zornoza, A. Influence of Polyethylene Glycol 4000 on the Polymorphic Forms of Diflunisal. *Eur. J. Pharm. Sci.* **1999**, *8*, 127–132.
- Nishimoto, Y.; Ichimura, Y.; Kinoshita, R.; Teramoto, Y.; Yoshida, H. Heat Capacity Measurements by Heat-Flux Type DSC on Cooling and Heating Cycles at Low Scan Rates. *Thermochim. Acta* **1991**, *179*, 117–124.
- Richardson, M.J. Quantitative Differential Scanning Calorimetry from Heat Capacities to Free Energies. *Anal. Proc.* **1988**, *25*, 16–18.
- Hill, V.L.; Craig, D.Q.M.; Feely, L.C. Characterisation of Spray-Dried Lactose Using Modulated Differential Scanning Calorimetry. *Int. J. Pharm.* **1998**, *161*, 95–107.
- Martini, A.; Kume, S.; Crivellente, M.; Artico, R. Use of Subambient Differential Scanning Calorimetry to Monitor the Frozen-State Behaviour of Blends of Excipients for Freeze-Drying. *PDA. J. Pharm. Sci. Technol.* **1997**, *51*, 62–67.
- Her, L.-M.; Nail, S.L. Measurement of Glass Transition Temperatures of Freeze-Concentrated Solutes by Differential Scanning Calorimetry. *Pharm. Res.* **1994**, *11*, 54–59.
- Miller, D.P.; de Pablo, J.J.; Corti, H. Thermophysical Properties of Trehalose and its Concentrated Aqueous Solutions. *Pharm. Res.* **1997**, *14*, 578–590.
- van Winden, E.C.A.; Talsma, H.; Crommelin, D.J.A. Thermal Analysis of Freeze-Dried Liposome-Carbohydrate Mixtures with Modulated Temperature Differential Scanning Calorimetry. *J. Pharm. Sci.* **1998**, *87*, 231–237.
- Shalae, E.Y.; Franks, F. Structural Glass Transitions and Thermophysical Processes in Amorphous Carbohydrates and Their Supersaturated Solutions. *J. Chem. Soc. Faraday Trans.* **1995**, *91*, 1511–1517.
- Hancock, B.C.; Zografi, G. Characteristics and Significance of the Amorphous State in Pharmaceutical Systems. *J. Pharm. Sci.* **1997**, *86*, 1–12.
- Hancock, B.C.; Shamblin, S.L.; Zografi, G. Molecular Mobility of Amorphous Pharmaceutical Solids Below Their Glass Transition Temperatures. *Pharm. Res.* **1995**, *12* (6), 799–806.
- Shamblin, S.L.; Tang, X.; Chang, L.; Hancock, B.C.; Pikal, M.J. Characterization of the Time Scales of Molecular

- Motion in Pharmaceutically Important Glasses. *J. Phys. Chem. B* **1999**, *103*, 4113–4121.
30. Hatley, R.H.M. Glass Fragility and the Stability of Pharmaceutical Preparations: Excipient Selection. *Pharm. Dev. Technol.* **1997**, *2* (7), 257–264.
 31. Hancock, B.C.; Dalton, C.R.; Pikal, M.J.; Shamblyn, S.L. A Pragmatic Test of a Simple Calorimetric Method for Determining the Fragility of Some Amorphous Pharmaceutical Materials. *Pharm. Res.* **1998**, *15* (5), 762–767.
 32. Wesolowski, M.; Konarski, T. General Remarks on the Thermal Decomposition of Some Drugs. *J. Therm. Anal.* **1995**, *43*, 279–289.
 33. Schwarz, E.; Pfeffer, S. Use of Subambient DSC for Liquid and Semisolid Dosage Form: Pharmaceutical Product Development and Quality Control. *J. Therm. Anal.* **1997**, *48*, 557–567.
 34. Rajala, R.; Laine, W.E. The Effect of Moisture on the Structure of Magnesium Stearate. *Thermochim. Acta.* **1995**, *248*, 177–188.
 35. Han, J.; Suryanarayanan, R. Applications of Pressure DSC in the Study of Pharmaceutical Hydrates. I. Carbamazepine Dihydrate. *Int. J. Pharm.* **1997**, *157*, 209–218.
 36. Danjo, K.; Nishio, F.; Zhou, B.D.; Otsuka, A. Interactions Between Water and Pharmaceutical Polymers Determined by Water-Vapor Sorption Measurements and Differential Scanning Calorimetry. *Chem. Pharm. Bull.* **1995**, *43*, 1958–1960.
 37. McCrystal, C.B.; Ford, J.L.; Rajabi-Siahboomi, A.R. Water Distribution Within Cellulose Ethers Using Differential Scanning Calorimetry. I. Effect of Polymer Molecular Weight and Drug Addition. *J. Pharm. Sci.* **1999**, *88*, 792–796.
 38. Phipps, M.A.; Mackin, L.A. Application of Isothermal Microcalorimetry in Solid State Drug Development. *Pharm. Sci. Technol. Today* **2000**, *3*, 9–17.
 39. Hollenbeck, R.G.; Peck, G.E.; Kildsig, D.O. Application of Immersional Calorimetry to Investigation of Solid–Liquid Interactions: Microcrystalline Cellulose–Water System. *J. Pharm. Sci.* **1978**, *67*, 1599–1606.
 40. Joshi, H.N.; Wilson, T.D. Calorimetric Evaluation of Hydration of Cholestyramine. *J. Pharm. Pharmacol.* **1995**, *47*, 188–192.
 41. Joshi, H.N.; Wilson, T.D. Calorimetric Studies of Dissolution of Hydroxypropyl Methylcellulose E5 (HPMC E5) in Water. *J. Pharm. Sci.* **1993**, *82* (10), 1033–1038.
 42. Duddu, S.P.; Das, N.G.; Kelly, T.P.; Sokoloski, T.D. Microcalorimetric Investigation of Phase Transitions: I. Is Water Desorption from Theophylline Monohydrate a Single-Step Process? *Int. J. Pharm.* **1995**, *114*, 247–256.
 43. Attwood, D.; Boitard, E.; Dubes, J.P.; Tachoire, H. Calorimetric Study of the Influence of Electrolyte on the Micellization of Phenothiazine Drugs in Aqueous Solution. *J. Phys. Chem. B* **1997**, *101*, 9586–9592.
 44. Blandamer, M.J.; Briggs, B.; Cullis, P.M.; Irlam, K.D.; Engberts, J.B.F.N.; Kevelam, J. Titration Microcalorimetry of Adsorption Processes in Aqueous Systems: Interaction of Sodium Dodecylsulfate and Sodium Decylsulfate With Poly(*N*-vinylpyrrolidone). *J. Chem. Soc. Faraday Trans.* **1998**, *94* (2), 259–266.
 45. Bloor, D.M.; Holzwarth, J.F.; Wyn-Jones, E. Polymer/Surfactant Interactions: The use of isothermal titration calorimetry and EMF measurements in the Sodium Dodecyl Sulphate/Poly(*N*-vinylpyrrolidone) system. *Langmuir* **1995**, *11*, 2312–2313.
 46. Rowlings, C.E.; Wurster, D.E.; Ramsey, P.J. Calorimetric Analysis of Powder Compression: II. The Relationship Between Energy Terms Measured with a Compression Calorimeter and Tableting Behavior. *Int. J. Pharm.* **1995**, *116*, 191–200.
 47. Wurster, D.E.; Rowlings, C.E.; Creekmore, J.R. Calorimetric Analysis of Powder Compression: I. Design and Development of a Compression Calorimeter. *Int. J. Pharm.* **1995**, *116*, 179–189.
 48. Salvetti, G.; Tognoni, E.; Tombari, E.; Johari, G.P. Excess Energy of Polymorphic States or Glass Over the Crystal State by Heat of Solution Measurement. *Thermochim. Acta.* **1996**, *285*, 243–252.
 49. Pikal, M.; Lukes, A.; Lang, J.; Gaines, K. Quantitative Crystallinity Determinations for Beta-Lactam Antibiotics by Solution Calorimetry: Correlations with Stability. *J. Pharm. Sci.* **1978**, *67*, 767–773.
 50. Gao, D.; Rytting, J.H. Use of Solution Calorimetry to Determine the Extent of Crystallinity of Drugs and Excipients. *Int. J. Pharm.* **1997**, *151*, 183–192.
 51. Briggner, L.E.; Buckton, G.; Bystrom, K.; Darcy, P. The Use of Isothermal Microcalorimetry in the Study of Changes in Crystallinity During the Processing of Powders. *Int. J. Pharm.* **1994**, *105*, 125–135.
 52. Beezer, A.E.; Gaisford, S.; Hills, A.K.; Willson, R.J.; Mitchell, J.C. Pharmaceutical Microcalorimetry: Applications to Long-Term Stability Studies. *Int. J. Pharm.* **1999**, *179* (2), 159–165.
 53. Lehto, V.P.; Salonen, J.; Laine, E. Real Time Detection of Photoreactivity in Pharmaceutical Solids and Solutions with Isothermal Microcalorimetry. *Pharm. Res.* **1999**, *16* (3), 368–373.
 54. Pikal, M.J.; Dellerman, K.M. Stability Testing of Pharmaceuticals by High-Sensitivity Isothermal Calorimetry At 25°C: Cephalosporins in the Solid and Aqueous Solution States. *Int. J. Pharm.* **1989**, *50*, 233–252.
 55. Heidemann, D.R.; Jarosz, P.J. Preformulation Studies Involving Moisture Uptake in Solid Dosage Forms. *Pharm. Res.* **1991**, *8*, 292–297.

BIBLIOGRAPHY

- Ford, J.L.; Timmins, P. *Pharmaceutical Thermal Analysis: Techniques and Applications*; Halsted Press: New York, NY, 1989.
- McCauley, J.A.; Brittain, H.G. *Thermal Methods of Analysis. Physical Characterization of Pharmaceutical Solids*; Brittain, H.G., Ed.; Marcel Dekker, Inc.: New York, NY, 1995; 223–251.
- Turi, E.A., Ed. *Thermal Characterization of Polymeric Materials*, 2nd Ed.; Academic Press: San Diego, CA, 1997.

CAPSULES, SOFT

David H. Bergstrom

R.P. Waranis

M.S. Rahman

Cardinal Health Pharmaceutical Technologies & Services Center, Somerset, New Jersey, U.S.A.

J.C. Ferdinando

R.P. Scherer Limited, Swindon, United Kingdom

INTRODUCTION

Soft gelatin capsules (also referred to as soft elastic gelatin capsules, Liqui-gels®, or softgels) are a unique drug delivery system that can provide distinct advantages over traditional dosage forms such as tablets, hard-shell capsules, and liquids. However, due to economic, technological, and patent constraints, there are relatively few manufacturers of softgels in the world (1).

Some of the major advantages of softgels include:

- improved bioavailability (increased drug absorption (2))
- speed of product development
- enhanced drug stability (protection against oxidation, photodegradation, and hydrolysis in lipophilic systems)
- superior patient compliance/consumer preference (ease of swallowing, appealing appearance, absence of objectionable taste, and convenience) and pharmaceutical elegance
- excellent dose uniformity (less than $\pm 1\%$ for solution fills; $\pm 1-3\%$ for suspension fills)
- better tamper evidence (tampering leads to puncturing and visible leakage)
- safer handling of highly potent or cytotoxic drug compounds
- product differentiation (through selection of novel shapes, colors, and sizes)
- excellent product life-cycle management.

In comparison, the disadvantages of softgels are relatively few. These include:

- specialized manufacturing equipment requirements
- higher cost for manufacturing as compared to tablets
- limited availability of technical experts

In surveys comparing various pharmaceutical dosage forms, softgels are rated as a high tech dosage form with strong customer preference. They are formed, filled, and sealed in a single operation. Once production for a specific

product begins, the manufacturing process normally proceeds 24 h per day until the lot of product is completed. This results in a manufacturing environment that operates around the clock, 7 days a week.

The standard softgel shapes for oral pharmaceutical products are oval, oblong, and round, though softgels can be easily manufactured in any shape. A recent survey has shown that smaller sized softgels are preferred within each shape category, with oval being the most popular shape.

DESCRIPTION

The softgel (Fig. 1) is a hermetically-sealed, one-piece capsule with a liquid or semisolid fill. The softgel consists of two major components, the gelatin shell and the fill (Fig. 2). In the finished product, the gelatin shell is primarily composed of gelatin, plasticizer, and water. The fill materials can include a wide variety of vehicles and can be either a solution or a suspension. Softgels may be coated with suitable enteric coating agents, such as cellulose acetate phthalate, to obtain enteric release of encapsulated material.

Because of their special properties and advantages, softgels are used extensively in many pharmaceutical, cosmetic, and nutritional products. The primary pharmaceutical applications include oral dosage forms, chewable softgels, suppositories, and topical products. The size of a softgel represents its nominal capacity in minims (1 cc = 16.23 minims). For example, an 11 oblong softgel can be filled with 8.5–11.0 minims of fill formulation.

FORMULATION DEVELOPMENT

Having selected the softgel dosage formulation of a drug, the desired performance characteristics will influence the



Fig. 1 Examples of softgels.

nature of the development of that formulation. The first consideration in this process is whether the softgel should be transparent, containing a solution of drug, or opaque, containing a drug suspension. Next, it must be determined whether the formulation is intended to provide bioequivalence with another product or to have enhanced performance *in vivo*; for example, faster or more complete absorption. This section will discuss the formulation principles of softgels, including gelatin shell and fill formulations.

Gelatin Shell Formulation

Typical softgel shells consist of gelatin, plasticizer, water, and materials that impart the desired appearance (colorants and/or opacifiers) and, on occasion, flavors and/or preservatives. A description of the functions, types, and amounts of materials most often used in manufacturing softgel shell formulations is detailed in the following paragraphs.

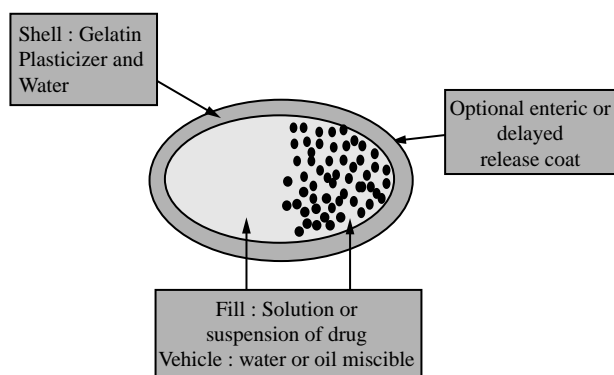


Fig. 2 Softgel components.

Gelatin

Gelatin provides the structural support for the shell of the softgel. It is typically 40–50% of the wet gel formulation and can be either Type A (acid processed) or Type B (alkali processed). The selection for the type of gelatin for a particular softgel formulation is based on compatibility with the other ingredients (both active and inactive) within the softgel. The steps involved in the gelatin manufacturing process include extraction, neutralization, drying, and grinding (3). The physicochemical properties of gelatin are largely controlled by the source of collagen, extraction method, pH, thermal history, and electrolyte content.

Plasticizers

Plasticizers are used to make the softgel shell elastic and pliable. The ratio of plasticizer to gelatin determines the hardness of the shell, assuming there is no effect from the fill. Plasticizers generally account for 20–30% of the wet gel formulation and are commonly glycerin, sorbitol, or propylene glycol, either individually or in combination. Several proprietary blends of sugar mixtures with sorbitol anhydrides can also be used and are available from excipient suppliers (4, 5). The amount and choice of the plasticizer help to determine the hardness of the final product, and may also affect the dissolution or disintegration of the softgel, as well as its physical and chemical stability. Plasticizers are selected on the basis of their compatibility with the fill formulation, processing (drying) time, and desired properties of the final softgels, including hardness, appearance, handling characteristics, stability, and even the geographical location in which the product will be sold.

Water

Water usually accounts for 30–40% of the wet gel formulation and is critical to ensure proper processing during gel preparation and softgel encapsulation. Following encapsulation, excess water is removed from the softgels through controlled drying, leaving the equilibrium water content typically at less than 10%.

Colorants/Opacifiers

Colorants and opacifiers are typically used at low concentrations in the wet gel formulation. A wide range of colorants such as FD&C and D&C water-soluble dyes, certified lakes, pigments, and vegetable colors have been incorporated into gelatin shells alone or in combination to produce the desired color, tint, or hue for product identification. A general rule in color selection is that the color of the capsule shell should be similar to or darker than the fill material.

An opacifier is sometimes added to the gelatin shell to obtain an opaque shell for suspension fills or to protect light sensitive fill ingredients. Titanium dioxide is the most commonly used opacifier. Flavors such as ethyl vanillin and essential oils are sometimes included in the capsule shell to impart desirable odors or flavors or to offset odoriferous materials that may be contained within the softgel itself.

Fill Formulation

Because of the migration of components (water, plasticizers, drugs, etc.) within the softgel both during and following encapsulation, formulation of the fill material must be conducted concurrently with formulation of the shell for maximum product quality. Without this simultaneous development, it is not uncommon for problems to arise.

The viscosity of fills can range from mobile liquids to thick suspensions or pastes. The fill material in a soft gelatin capsule can be a liquid, a combination of miscible liquids, a solution of a solid(s) in a liquid(s), or a suspension. These formulations are designed to produce the smallest possible capsule consistent with acceptable chemical and physical stability, therapeutic effectiveness, and production efficiency.

The large groups of liquids that can be encapsulated into softgels fall into one of two categories: water-miscible liquids and water-immiscible liquids (6).

Water-miscible liquids include polyethylene glycols and nonionic surfactants, such as the polysorbates. Low molecular weight grades of polyethylene glycol (e.g., PEG 400) are used most commonly since they remain liquid at ambient temperatures. Small amounts (up to 5–10%) of other water-miscible liquids, such as propylene glycol, ethanol, and glycerin, can also be used.

Water-immiscible liquids include vegetable and aromatic oils, aliphatic, aromatic and chlorinated hydrocarbons, ethers, esters, high molecular weight organic acids, and some alcohols.

Liquids that are likely to cause problems following encapsulation are low molecular weight water-soluble and volatile organic compounds, such as some alcohols, acids, ketones, and esters; water (above 5%); emulsions (whether oil in water or water in oil); liquids with extremes of pH; and aldehydes.

Drugs that are not sufficiently soluble in the solvent or combinations of solvents can be formulated into suspensions and encapsulated. The particle size of insoluble drugs should be 80 mesh or finer for maximum suspension homogeneity and capsulation equipment requirements. Examples of suspending agents include paraffin wax, beeswax, and hydrogenated vegetable oil for oily vehicles,

and solid glycol esters (such as higher molecular weight PEG) for nonoily vehicles. Surfactants, such as polysorbates, are often added to the dispersion to promote wetting of the ingredients and/or dispersion of the fill *in vivo*.

In general, many different materials may be encapsulated; however, limitations do exist for some compounds due to high solubility in water and/or inherent chemical reactivity and the resultant effect on the shell. These compounds include strong acids and alkalis and their salts, as well as ammonium salts. Some compounds, such as aldehydes, can react with gelatin, causing crosslinking and resulting in a product that lacks bioavailability. In addition, any substance (such as aspirin) that is unstable in the presence of moisture may also exhibit unacceptable chemical stability in softgels.

PRODUCT DEVELOPMENT

This section will discuss the rationale of softgel product development, including solubility screening, gel compatibility, process development, and trial batch manufacture, concluding with a review of the manufacturing process and specialized formulation approaches to enhance pharmacokinetic performance.

The flow diagram in (Fig. 3) details the activities normally undertaken in the development of a softgel. The major steps involved are described in the following paragraphs.

The first step in developing a solution containing softgel is to determine the solubility of the drug in a range of pharmaceutically acceptable solvents. After the solubilities are measured, the solvents are then selected on the basis of their regulatory acceptability and known compatibility with softgel dosage forms. The types of excipient typically include:

- hydrophilic solvents
- lipophilic materials
- hydrophilic surfactants
- lipophilic surfactant
- co-solvents.

Solvents that provide adequate solubility of the drug can be selected, though it may be necessary to combine them to achieve the desired *in vitro* or *in vivo* characteristics and to ensure good physical stability. In addition to characterizing the *in vitro* and *in vivo* performance of the preliminary formulation, it is important to evaluate the drug solubility in the mixtures for the following conditions:

- physical stability under accelerated conditions
- chemical stability under accelerated conditions
- excipient compatibility.

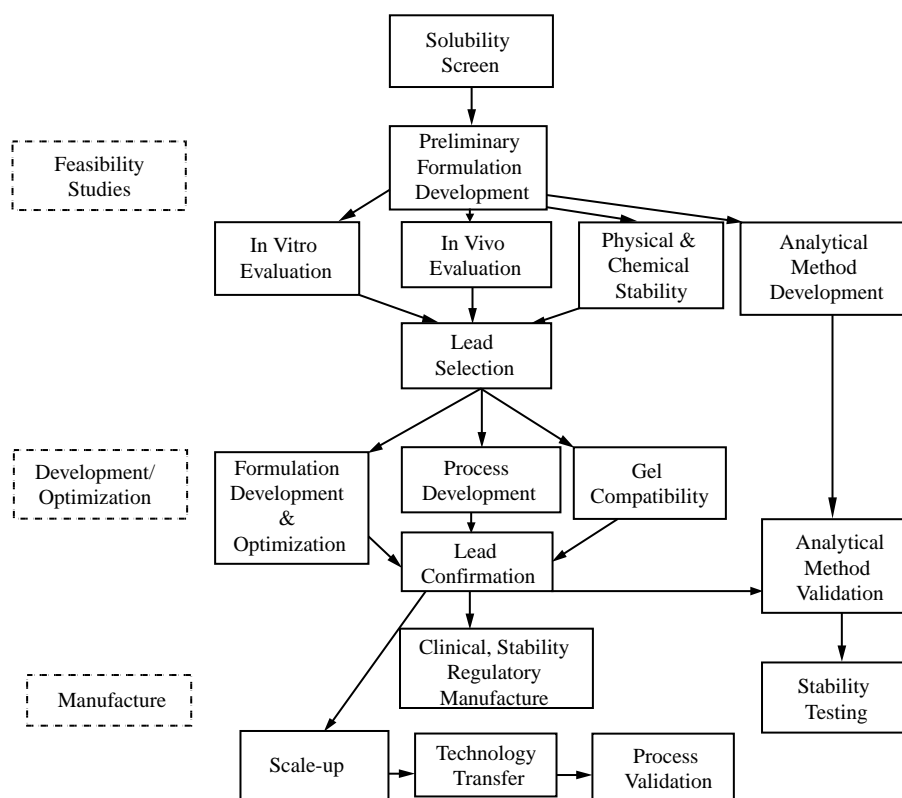


Fig. 3 Typical softgel development flow diagram.

For softgels containing suspension fills, the solubility of the drug in a range of pharmaceutically acceptable solvents is also measured and excipients in which the drug shows little or no solubility are then selected. These formulations generally require viscosity enhancers in order to provide adequate suspending characteristics for the drug during processing. This is vital in maintaining drug homogeneity during manufacture. The type and level of viscosity enhancer is optimized to provide the best manufacturability.

Gel Compatibility

Thorough gel compatibility testing between the fill and shell formulation is an important part of the development process. A variety of problems may result if the fill is not well matched to the proper shell formulation. These may be observed either immediately after encapsulation or after prolonged storage.

Process Development/Trial Manufacture

Having identified potential fill and shell formulations on the laboratory scale, a suitable manufacturing process that

will enable successful preparation of the trial batch materials required for regulatory and clinical studies must be developed. Such process development includes investigating several processing parameters, such as the order of addition, temperature, mixing condition, and selection of equipment.

Formulations selected for stability and clinical evaluation will be prepared as trial (pilot) batches. During manufacture, the process and product will be evaluated to provide valuable information for later process ranging and validation studies. For example, the fill moisture and hardness of the capsules during the drying stages will be monitored to optimize the drying process and resulting product stability. Fig. 4 shows the drying profile of a softgel product. Note that the reduction in fill moisture is accompanied by an increase in capsule hardness.

METHOD OF MANUFACTURE

As early as the 1830s, softgels were used as a method of drug delivery. Early manufacturing included both a hand-dipping method and a plate-press method. The hand-dipping

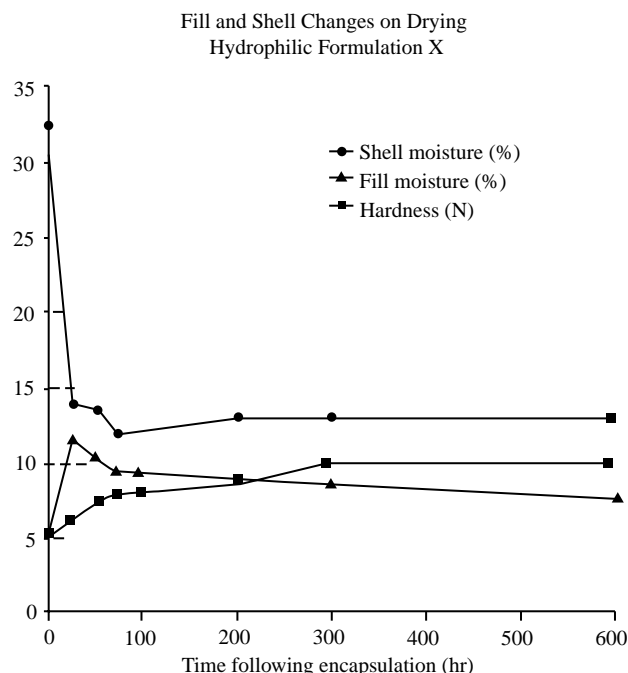


Fig. 4 Typical drying profile.

method created individual empty softgel shells that were subsequently filled with a syringe or dropper. The plate-press method was a batch process that involved pressing two sheets of wet gelatin together between two molds. The molds provided depressions in the gelatin sheet into which active fill was then placed. A second gelatin sheet was laid

over the first and both were pressed together with fill material sandwiched between. The pressure of the plate dies sealed the top and bottom sheets of gelatin together and cut out the individual softgels for subsequent drying.

Almost every softgel on the market today is made using the rotary die process patented by Scherer in 1933 (7). The equipment and manufacturing process has improved dramatically over the years, but the underlying manufacturing principle remains essentially unchanged. In this method, two independent processes take place, often simultaneously, yielding two different materials, the gel mass and the fill material. Both are united in the encapsulation process that produces wet softgels.

The wet gel mass is manufactured by mixing together and melting, under vacuum, the gelatin shell ingredients (gelatin, plasticizer, water, colorants and sometimes opacifiers, flavors, and preservatives). At the encapsulation machine (Fig. 5), molten gel mass flows through heated transfer tubes and is cast onto chilled drums, forming two separate ribbons, each approximately 6 in. wide. The thickness of the ribbons (usually 0.02–0.04 in.) is carefully controlled and checked periodically throughout manufacture. The gel ribbons traverse through rollers that provide proper alignment of the ribbons and apply lubricant to both surfaces of the ribbons. Each gel ribbon forms one half of the softgel. Two-toned softgels are made using two different colored gel ribbons.

Active fill materials are manufactured in a process separate from the gel mass manufacture. The viscosity of

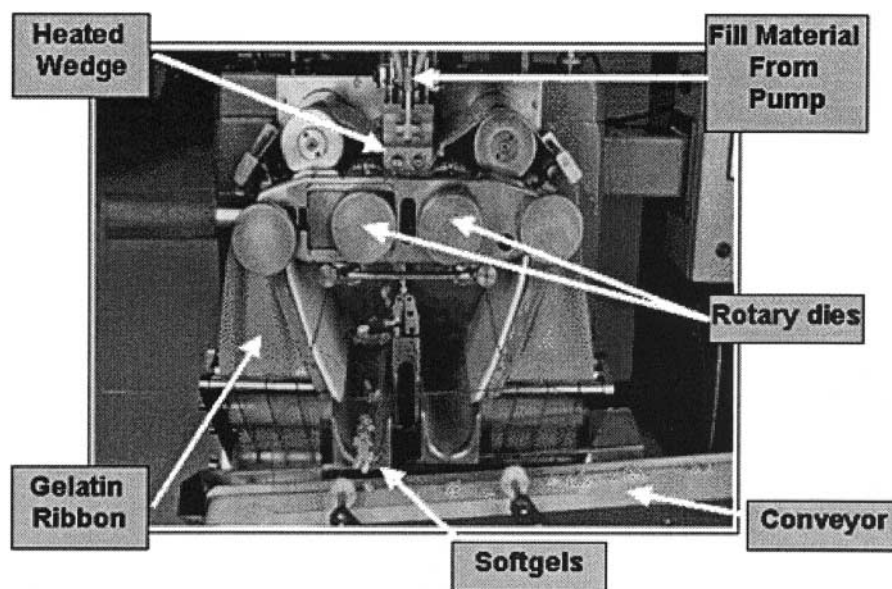


Fig. 5 Encapsulation equipment.

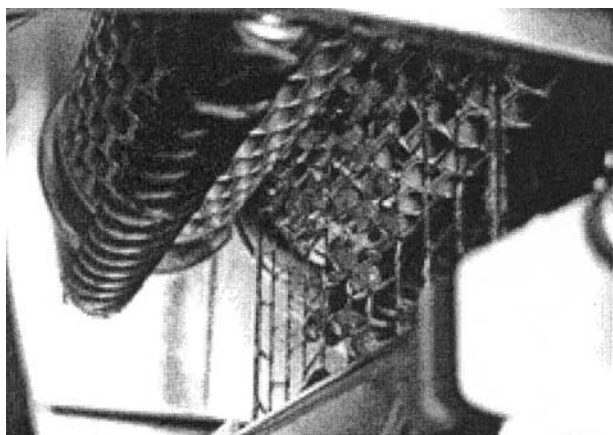


Fig. 6 Newly formed softgels.

all fills and the particle size of suspended materials are important parameters established during development and controlled throughout manufacture.

Softgels are formed during the encapsulation part of the process, using the two gel ribbons and the fill material. Lubricated gel ribbons are fed between a pair of counter-rotating dies, the surface of which contains matching pockets of appropriate size and shape that serve as molds for forming the softgels. The die pockets also seal both sides of the softgel and cut the formed softgel away from the residual gel ribbon. Fig. 6 shows softgels immediately following encapsulation as they are being separated from the ribbons. The softgels are then conveyed to a tumble dryer to initiate drying.

Situated between the ribbons and rotating dies is the wedge as shown in Figs. 7 and 8. The wedge serves three separate functions during the encapsulation process. First, it heats the gel ribbons close to the gel-sol transition temperature to ensure that melting (welding) of the two gel ribbons occurs when the ribbons are pressed together between the dies. Second, the wedge is part of the system that distributes the fill material from a positive displacement pump to each of the die pockets. Finally, the wedge, in conjunction with the lubricant, provides a sealing surface against the ribbons to eliminate air and allows a seal to be formed between the shell and fill material without the introduction of air into the product.

In order to properly manufacture the gel mass and form the gel ribbons, the gel mass formulation contains excess water. Following encapsulation, softgels must be dried to obtain a final product that will be durable enough to withstand subsequent processing, packaging, and shipping, and possess good long term physical stability. Drying occurs in two stages. Initial drying takes place in a rotating basket dryer that tumbles the softgels in temperature and humidity controlled air. This removes approximately half of the excess water. The balance of the excess water is removed during the secondary drying stage, when the softgels are spread in a single layer on shallow trays. The trays are designed and stacked to allow air to pass through the rack and around the softgels (Fig. 9). Secondary drying proceeds under controlled conditions of temperature and humidity until the appropriate level of hardness or fill moisture is achieved. Complete drying can

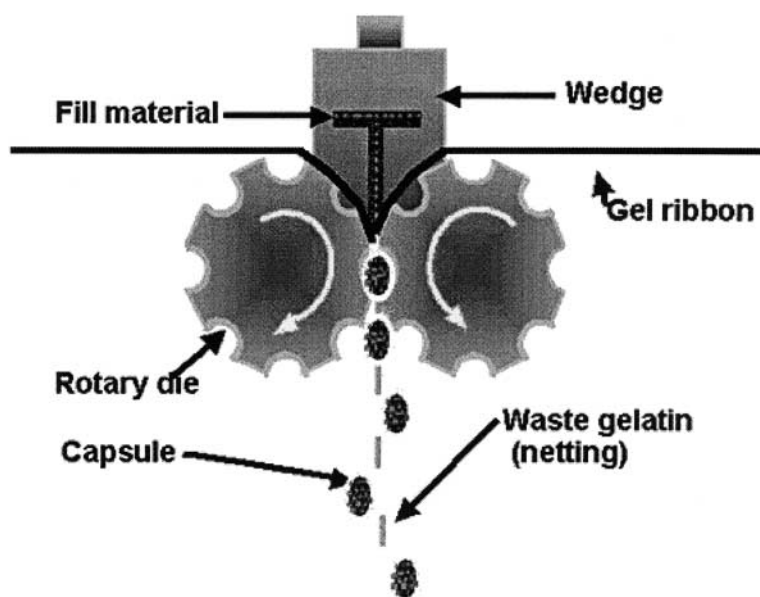


Fig. 7 Softgel encapsulation process.

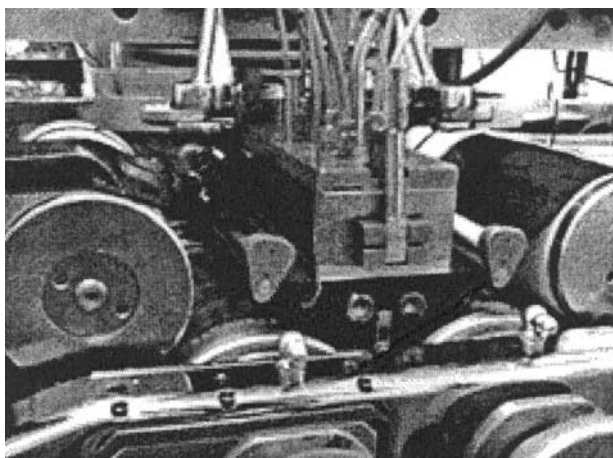


Fig. 8 Close-up of die-wedge equipment.

take from 3 days to 3 weeks depending on shell and fill formulations and the size of the softgel.

Once the softgels have reached the desired drying endpoint, they are placed into bulk holding containers to prevent further drying. At this point, several additional operations may be performed, including washing, printing, inspecting, and packaging.

THERAPEUTIC PERFORMANCE

The pharmacokinetic performance of drugs can be enhanced by softgel dosage forms, the exact formulation of which depends on the desired pharmacokinetic improvement. The two most common requirements for

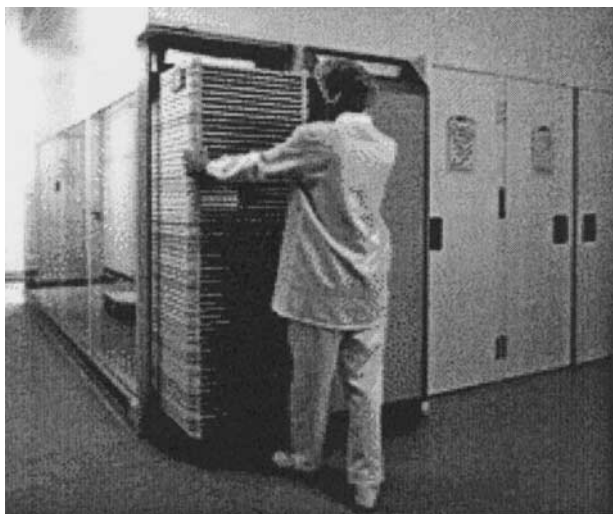


Fig. 9 Tray drying of softgels in controlled drying tunnels.

this formulation are faster and more complete absorption. In both cases, the ideal situation is for the drug to be dosed in solution and formulated to remain in solution after dispersion in gastrointestinal media, possibly as a nanoemulsion. Formulation of nanoemulsion preconcentrates for softgel encapsulation requires the drug to be in solution in a mixture of oils, surfactants, cosurfactants, and possibly cosolvents.

Rate of Absorption

Noteworthy advances recently have been made in the development of softgel formulations to address particular performance issues in vivo. These include presentation of the drug to the gastrointestinal tract in a solution from which the drug can be absorbed significantly faster than that from a solid oral dosage form, which may be rate-limited by the need for disintegration followed by drug dissolution. With the solution-softgel approach, the shell ruptures within minutes to release the drug solution, usually in a hydrophilic or highly dispersing vehicle that aids the rate of absorption. This can be a valuable attribute for treatments such as migraine or acute pain, or where there is a limited absorption window in the gastrointestinal tract. Fig. 10 compares the absorption rates between a solution softgel formulation and a tablet of ibuprofen (8). The data are based on a pharmacokinetic comparison of 400 mg ibuprofen in 12 human volunteers.

Increased Bioavailability

In addition to increasing the rate of absorption, softgels may also improve the extent of absorption. This can be particularly effective for large hydrophobic drugs.

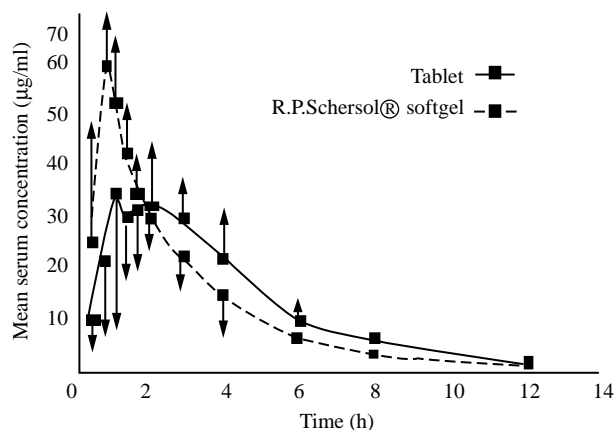


Fig. 10 Comparison of absorption rates of ibuprofen from softgel and tablet.

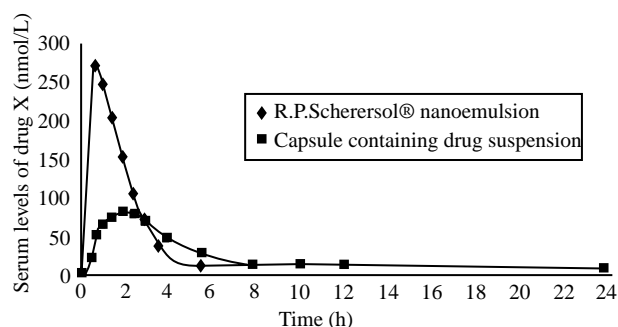


Fig. 11 Comparison of nanoemulsion softgel to a drug suspension capsule.

Recently, the protease inhibitor saquinavir has been relaunched in a patented solution–softgel formulation, providing approximately three times the bioavailability as the original hard-shell formulation (9).

In some cases, drugs may be solubilized in vehicles capable of spontaneously producing a microemulsion or nanoemulsion on contact with gastrointestinal fluids. This particular vehicle consists of oils and surfactants in appropriate proportions which, on contact with aqueous fluids, produce an emulsion preferably with an average droplet size less than 100 nm. The solubility of the drug should be maintained as long as possible, delivering solubilized drug directly to the enterocyte membrane. It may even be possible to utilize the body's own systems for oil digestion to produce micelles containing solubilized drug. Fig. 11 depicts the enhancement in plasma levels achieved in 12 human volunteers when a nanoemulsion softgel was used to dose a hydrophobic drug as compared

to a capsule containing a suspension of micronized drug particles (9).

Decreased Plasma Variability

High variability in drug plasma levels is a common characteristic of drugs with limited bioavailability. By dosing the drug optimally in solution, the variability of such drug plasma levels can often be reduced. Cyclosporin benefits from such an approach (10). Fig. 12 depicts the administration of a 10 mg/kg dose of Cyclosporin A (Sandimmune®) softgel solution formulation in eight human volunteers (11). Fig. 13 depicts the administration of a 10 mg/kg dose of Cyclosporin A (Neoral®) microemulsion preconcentrate softgel formulation in eight fasting human volunteers (11).

PRODUCT QUALITY CONSIDERATIONS

Ingredient Specifications

Numerous specifications and control measures are employed to determine final product quality, the first of which is ensuring adequate quality of excipients and active ingredients. Excipient testing ensures compliance with compendial specifications, as well as specifications determined during development of the fill material and/or shell formulation. Among these are limiting values for trace impurities, especially peroxides, aldehydes, some metals, and ionic salts. Presence of these impurities can result in gelatin crosslinking and possible dissolution

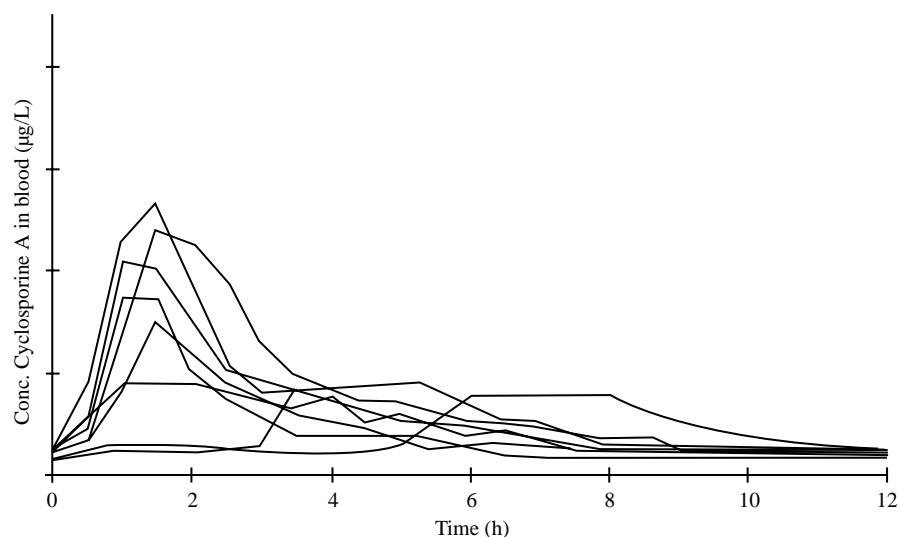


Fig. 12 Cyclosporine A (Sandimmune®) softgel formulation.

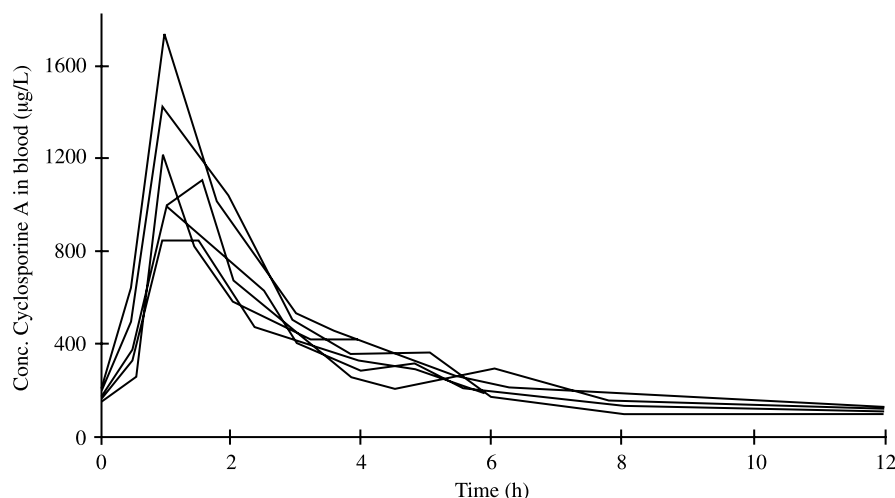


Fig. 13 Cyclosporine A (Neoral[®]) microemulsion preconcentrate softgel formulation.

problems or in undesired changes in the product appearance over time.

Since gelatin is the key ingredient for the shell and is present in larger quantities than other excipients, it is important to ensure that the gelatin meets not only current USP specifications, but the additional controls of particle size, viscosity, and bloom strength, all of which are significant for manufacturing process as well as final product stability. Other specifications, such as the quantity of certain ionic materials, are necessary to ensure stable product appearance during storage. Furthermore, it is essential to specify or limit other gelatin properties, such as color or even the source of the gelatin (bovine, porcine, bone, hide), depending on the formulation and intended market of the final product.

In-Process Testing

Several tests are conducted on a regular basis throughout the encapsulation portion of the softgel manufacturing process. These include weight determinations for both the fill material and the shell, and measurements of the thickness of the seals of the softgels themselves.

Fill moisture and/or hardness measurements are performed during the drying process, the results of which are used to determine the drying endpoint for each lot. Specifications for fill weight, shell weight, seam thickness, and drying endpoints are based on the softgel size, amount and type of fill, and the results obtained during previous process development studies.

Final Product Testing

Once the softgels have completed all required processing steps, the lot is inspected and sampled for final product

release testing. Tests required for final product release are dependent on regulatory requirements for the product and usually include microbiological testing, assay and identity of actives, physical appearance, fill weight, dissolution or disintegration, and dosage uniformity.

RECENT ADVANCES IN NEW TECHNOLOGY

Recent advances in the development of softgel formulations and manufacturing have lead to exciting improvements, including the capability of imprinting the product logo directly on the wet gelatin ribbon prior to encapsulation, the ability to encapsulate small quantities of softgels in the laboratory setting for early stability or quick in vivo evaluations, and the ability to encapsulate microspheres (controlled or immediate release) suspended in the fill.

TRENDS IN PATENT ACTIVITY

A review of the United States patent activity over the past decade reveals some interesting trends within the softgel technology arena. While the review is limited to U.S. patents, it most likely represents the patent activity on a worldwide basis. A listing of the more significant patents, sorted into groups relating to either formulations (12–21), manufacturing technology (22–29) or softgel design innovations (30–40), has been included in the reference section.

In 1990, there were 210 patents issued citing soft gelatin capsules as either specific claims or examples of

possible dosage forms. Over the past 12 months, approximately 760 patents were issued. Pharmaceutical application of soft gelatin capsule technology has increased over the past decade. This fourfold increase in the number of patents involving softgels may suggest a broader understanding of the benefits of this technology both in clinical performance as well as patient and consumer appeal.

Looking more closely at patents over the same period, there were 45 U.S. patents issued in 1990 where the soft gelatin capsule was a specific claim. The past 12 months yielded approximately 300 patents with softgels as a specific claim. This analysis is probably a better indication of patent activity specifically related to the softgel technology. This sixfold increase in the number of patents specifically involving softgel formulations may reflect greater and more widespread expertise with regard to softgel formulation processes. It may also be an indication of the greater proportion of "difficult" to formulate drugs currently coming out of basic research centers, that is, low aqueous solubility and/or poor or variable gastrointestinal absorption.

Examination of the patent activity of the top 20 pharmaceutical companies, or their predecessor companies, in the year 2000 vs. 1990 suggests an industry sector shift in the use of soft gelatin capsules. In 1990, the top 20 pharmaceutical firms obtained 85% of patents. This decreased to 57% over the past 12 months. Since overall pharmaceutical application of softgel technology has increased, a reasonable inference would be that the comparatively young and smaller biopharmaceutical industry sector is coming of age as compounds begin to move from basic research to development and commercialization.

Over the 1990 and 2000 periods, 95% of patents granted for pharmaceutical softgel products relate to drugs or fill formulations and not to specific claims or improvements regarding shell formulations or manufacturing processes. On the surface, this would appear to indicate a mature technology but since patents are public domain and process patents are difficult to enforce it is more likely that industry leaders are reluctant to pursue patents except in unusual circumstances. In this process-critical industry it is more reasonable to expect that companies prefer to maintain technological advances as internal know-how for competitive reasons.

CONCLUSION

Despite the specialized manufacturing process, softgels provide a versatile and efficient drug delivery system with

distinct advantages over conventional dosage forms, including improved bioavailability, shorter development times, superior patient preference, and enhanced dose uniformity. The inherent nature of the softgel offers a wide variety of usage and fill options. In any softgel product development effort, formulation of the fill and gelatin shell should be considered concurrently in order to optimize product quality and performance.

REFERENCES

1. Stanley, J.P. Soft Gelatin Capsules. *The Theory and Practice of Industrial Pharmacy*, 2nd Ed.; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea Febiger: Philadelphia, 1976; 359.
2. Ferdinando, J.C. Formulation Solutions—Softgels. *Pharmaceutical Manufacturing and Packaging* **2000**, March, 69–73.
3. Ward, A.G.; Courts, A. *The Science and Technology of Gelatin*; Academic Press: New York, 1977.
4. SPI Polylols, Inc. 321 Cherry Lane New Castle, DE 19720.
5. Rouquette America, Inc., Corporate Headquarters: 1417 Exchange Street, P. O. Box 6647 Keokuk, IA 52632-6647.
6. Jimerson, R.F. Softgel (Soft Gelatin Capsule) Update. *Drug Development and Industrial Pharmacy* **1986**, *12* (8/9), 1133–1137.
7. Scherer, R.P. US Patent 1,970,396, 1934.
8. Saano, V.; Paronen, P.; Peura, P.; Vidgen, M. Relative Pharmacokinetics of Three Oral 400 mg Ibuprofen Dosage Forms in Healthy Volunteers. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1991**, *29* (10), 381–385.
9. Perry, C.M.; Noble, S. Saquinavir Softgel Capsule Formulation. *Drugs* **1988**, *55* (3), 461–486.
10. Kovarik, J.M.; Mueller, E.A.; Van Bree, J.B.; Tetzloff, W.; Kutz, K. Reduced Inter- and Intraindividual Variability in Cyclosporine Pharmacokinetics from a Microemulsion Formulation. *J. Pharm. Sci.* **1994**, *83* (3), 444–446.
11. Trull, A.; Tan, K.K.C.; Tan, L.; Alexander, G.J.M.; Jamieson, N.V. Absorption of Cyclosporin from Conventional and New Microemulsion Oral Formulations in Liver Transplant Recipients with External Biliary Diversion. *Br. J. Clin. Pharmacol.* **1995**, *39*, 627–631.
12. Benet, L.Z.; Wacher, V.J. Use of Essential Oils to Increase Bioavailability of Orally Administered Pharmaceutical Compounds. US Patent 6,121,234, Sept 19, 2000.
13. Lacy, J.E.; Embleton J.K.; Perry E.A. Delivery Systems for Hydrophobic Drugs. US Patent 6,096,338, Aug 1, 2000.
14. Wacher, V.J.; Benet, L.Z. Propyl Gallate to Increase Bioavailability of Orally Administered Pharmaceutical Compounds. US Patent 5,962,522, Oct 5, 1999.
15. Makino, H.; Hakamada, I.; Yamada, O. Solvent System to be Enclosed in Capsules. US Patent 5,912,011, June 15, 1999.
16. Rinaldi, M.A.; Saxena, S.J.; Tutschek, P.C. Softgel Formulation Containing Retinol. US Patent 5,891,470, April 6, 1999.
17. Adesunloye, A.T.; Stach, P.E. Filled Gelatin Capsules. US Patent 5,874,106, Feb 23, 1999.

18. Bayless, R.E.; Laba, D.; Reynolds, W. Polyoxyalkylene Glycol Gelatin Capsule Fill Formulations Comprising Crosslinked Carboxylic Copolymers. US Patent 5,885,671, March 23, 1999.
19. Punto, L.; Potini, C.; Gould, E. Topical Application Emulsions. US Patent 5,587,149, Dec 24, 1996.
20. Lambrechts, J. Shampoo Conditioner and Softgel Filled Therewith. US Patent 5,573,756, Nov 12, 1996.
21. Morton, F.S.S.; Shelley, R.S.; Patel, M.S. Enhanced Solubility Pharmaceutical Solutions. US Patent 5,376,688, Dec 27, 1994.
22. Yamamoto, T.; Yagyu, M. Appearance Inspecting Device for Solid Formulation. US Patent 6,108,030, Aug 22, 2000.
23. Schurig, G.A.; Bezaire, B.; Ratko, B.; Baker, L.A.; Mikalian, R. Method for Striping or Marbleizing Capsules. US Patent 6,022,499, Feb 8, 2000.
24. Schmidt, W. Method for the Purification and Recovery of Waste Gelatin Using Diafilters. US Patent 5,945,001, Aug 31, 1999.
25. Bierdel-Willkommen, E.; Sambale, C.; Hahnlein, W. Soft Gelatin Capsules. US Patent 5,916,591, June 29, 1999.
26. Julius, K.; Bagusche, C. Method and Apparatus for Sorting Capsules. US Patent 5,819,953, Oct 13, 1998.
27. Rowe, D. Method and Apparatus for Forming Soft Gelatin Capsules Containing Particulate Material. US Patent 5,740,660, April 21, 1998.
28. Stroud, N.S.; Walker, D.J.; Garnett, K.R.; Burnett, A.J.; Wood, A.J. Encapsulation Apparatus and Process. US Patent 5,735,105, April 7, 1998.
29. Steele, D.; Dietel, G. Softgel Manufacturing Process. US Patent 5,200,191, April 6, 1993.
30. Morton, F.S.S.; Duque, P.P.; Chiprich, T.B.; Stroud, N.S. Methods for Preparing Gelatin Capsules Containing Fragrances. US Patent 6,099,858, Aug 8, 2000.
31. Hoy, M.R. Multiphase Soft Gelatin Dosage Form. US Patent 6,024,980, Feb 15, 2000.
32. Brox, W.; Meinzer, A.; Zande, H. Soft Gelatin Capsule Manufacture. US Patent 5,985,321, Nov 16, 1999.
33. Cody, S.L.; Devlin, B.T.; Dubek, J.; Hoy, M.R. Fill Material for Soft Gelatin Pharmaceutical Dosage Form. US Patent 5,919,481, July 6, 1999.
34. Hutchinson, K.G.; Garnett, K.R.; Fischer, G.; Page, N.S. Soft Gelatin Capsule Shell Compositions. US Patent 5,817,323, Oct 6, 1998.
35. Stone, D.D. Graphically Impressed Softgel and Method for Making Same. US Patent 5,827,535, Oct 27, 1998.
36. Hutchison, K.G.; Garnett, K.R. Method of Preparing a Water-Based Beverage. US Patent 5,681,606, Oct 28, 1997.
37. Tanner, K.; Orange, K. Tamper Evident Pharmaceutical Dosage Form. US Patent 5,624,681, April 29, 1997.
38. Chiprich, T.B.; Hoyleman, M.T.; Stroud, N.S. Capsule Shell Formulation to Produce Brittle Capsules. US Patent 5,641,217, March 25, 1997.
39. Stroud, N.S. High Amylose Starch Substituted Gelatin Capsules. US Patent 5,554,385, Sept 10, 1996.
40. Brox, W.; Block, P.; Schafer, R. Printed Shaped Bodies and Sheets Containing Gelatin. US Patent 5,529,767, June 25, 1996.

BLOOD SUBSTITUTES: THE FLUOROCARBON APPROACH

Jean G. Riess^a

University of California at San Diego, San Diego, California, U.S.A.

INTRODUCTION

Fluorocarbon emulsions provide a safe, efficient, and cost-effective passive means of delivering oxygen in vivo that does not depend on the collection of human or animal blood. Perfluorocarbons (PFCs, also designated as fluorocarbons or perfluorochemicals) are inert materials that have no ability to bind or react with oxygen, carbon monoxide, carbon dioxide, or nitric oxide. However, they can physically dissolve large amounts of these gases. PFCs are administered in the form of submicron size emulsions. The emulsifier utilized is egg yolk phospholipids (EYP), the same as that used for manufacturing fat emulsions for parenteral nutrition, with which PFC emulsions bear a definite similarity. Oxygen delivery to tissues by PFCs is facilitated by high extraction rates and ratios, the ability to increase oxygen content proportionally to oxygen partial pressure, and, in case of hemodilution, by increased cardiac output, resulting in large contributions to oxygen consumption relative to oxygen content. Phase 3 clinical trials have established the aptitude of such an emulsion to alleviate or reduce exposure to donor blood transfusions during surgery when used as part of the augmented acute normovolemic hemodilution procedure, which is one of the potential uses of blood substitutes. PFC emulsions thus have the potential to improve oxygen delivery, patient safety, and relieve blood shortages.

This entry will first survey the reasons for developing blood substitutes and outline the principles of oxygen delivery by PFC emulsions. It will then focus on the main challenges encountered in the development of such emulsions, namely the selection of an appropriate excretable PFC and the preparation of a stable, biocompatible emulsion. It will also allude to questions related to raw material procurement, product manufacture, and cost. Further sections will concern the pharmacokinetics, efficacy, and side effects of these oxygen carriers.

^aThe author is also a member of the Board of Directors of Alliance Pharmaceutical Corp. in San Diego; however, the opinions expressed here are solely his and not necessarily those of the Corporation.

Finally, the potential applications of these products will be outlined, including the status of their clinical trials, and some forward looking comments will be made.

OBJECTIVES—WHY BLOOD SUBSTITUTES?

Issues with Banked Blood

The blood supply of developed countries is now safer than it ever has been (1–3) although this is not the perception of the public. However, the risk associated with blood transfusion will never be completely eliminated. Further issues concern the immunodepressant effects of transfusions, and the lack of immediate efficacy of red blood cells (RBCs) stored for more than a few days. New legal questions have also arisen concerning, in particular, the extent of testing (4). Finally, blood shortages have become chronic in all areas of the world.

Safety

Acute anxiety about the safety of transfusion has developed following the AIDS tragedy. AIDS is only one of several infectious diseases transmitted by blood transfusion. Hepatitis is transmitted to recipients far more commonly than AIDS. Septic shock from bacterial contamination, although rare, does occur in certain areas (2). Malaria and Chagas disease are common in certain areas of the world (5). Other, as yet unknown, infectious agents are likely to emerge. The recent occurrence of a new variant of human Creutzfeldt–Jakob disease (nvCJD), which closely resembles bovine spongiform encephalopathy, raises serious public health concern (6, 7). Whether such encephalopathies are transmissible by transfusion is being debated (8, 9).

Blood transfusion is frequently associated with mild allergic reactions; more serious hemolytic transfusion reactions, transfusion-related acute respiratory distress syndrome, and even fatal acute hemolytic transfusion reactions are rare but do exist (2, 3). Finally, administrative errors remain one of the main causes of transfusion-related morbidity and mortality.

Additionally, there is increasingly strong evidence that allogeneic blood transfusion reduces the immune

responsiveness (or immunocompetence) of the organism, thus predisposing a transfusion recipient to infectious complications and septicemia (10–12). Fewer infections were observed in patients receiving autologous rather than allogeneic blood (13, 14). Mortality in patients with sepsis was associated with the age of the RBCs transfused (15). Intraoperative blood transfusion was also associated with an enhanced inflammatory response, increased concentrations of inflammatory mediators, and increased post-operative morbidity in patients undergoing cardiac surgery (16). Blood transfusion has been identified as a risk factor for postinjury multiple organ failure in trauma patients (17).

Blood transfusions also appear to increase the risks of recurrence and spread of certain cancers (10, 18, 19), and to reduce long-term survival following surgery (20). This risk may again be lower when autologous rather than allogeneic blood is used (21).

Whatever the level of risk associated with a unit of blood, this risk is cumulative with the number of units transfused; conversely, each unit spared to a patient reduces the risk of transfusion-related side effects.

Efficacy

Refrigeration and storage of RBCs results in so-called storage lesions, which include changes in the affinity of hemoglobin for O₂, decrease in pH, hemolysis, changes in RBC deformability, formation of microaggregates, release of vasoactive substances, and denaturation of proteins (22). After 2 weeks, little 2,3-diphosphoglycerate (2, 3-DPG), the allosteric factor that facilitates O₂ release to tissues, is left. As a result, transfusion of stored blood is not immediately effective in delivering O₂. It takes actually about 24 h for banked RBCs to restore their 2, 3-DPG level to about half of normal. In contrast to fresh RBCs, 28-day-old RBCs failed to improve tissue oxygenation in severely hemodiluted rats (23). A new important awareness is thus emerging: Banked blood (including predonated autologous blood) is not equivalent to *fresh* blood.

Availability

In developed countries blood collection is declining as a result of the aging of the populations, the development of aggressive new therapies, and the increase in the number of patients undergoing elective surgery (3). Blood shortages are increasingly frequent, leading to increasingly frequent postponement of elective surgeries. Further substantial increases in demand and decrease in donation are anticipated (22, 24, 25). Any additional steps to improve safety will invariably result in deferral of donors, hence directly impacting blood availability. If safety increases at the expense of availability, people may eventually die from lack of blood.

In the emerging countries, blood is chronically short due, among other things, to the limited number of suitable and/or willing donors. Viral or parasitic infection with blood-borne diseases can lead to the rejection of a large proportion of the units collected. In addition, the infrastructure and training required to collect, store, and deliver the available supply are sometimes lacking.

What is Expected from Blood Substitutes?

The products under development are not substitutes for blood; they only fulfill the O₂/CO₂ exchange function of blood and only for a short duration. They assume none of the regulation, defense, and coagulation functions of blood. Effective O₂ carriers would nevertheless be precious for the duration of a surgical procedure. In addition, they could ensure tissue oxygenation in situations where blood is no longer able to do so.

To be of value, O₂ carriers need to be able to load O₂ rapidly during passage of the blood through the lungs, reach the tissues at risk of ischemia, and deliver the O₂ load rapidly and maximally to these tissues. They must be free of bacteria and viruses, and should not promote the development of pathogens. They must be devoid of antigenic effects, and their side effects must be clinically minimal. They should preferably be devoid of physiological activity other than O₂ delivery. It is also essential that they preserve the increase in cardiac output that normally follows hemodilution. Ideally, they should have prolonged circulation lives. Use of O₂ carriers should eliminate clerical errors and the risk of transfusion incompatibilities. It is essential that the products be ready for use, easy to handle, and immediately effective. They must also be manufacturable on a large scale from safe raw materials, preferably be heat sterilized and stable enough to allow long-term storage under standard conditions. Finally, they need to be cost effective.

All the present O₂ carriers, whether fluorocarbon- or hemoglobin-based, have short circulation half-lives, a reality that needs to be taken into account in their use. This limitation, however, does not restrict use during surgery (which consumes about 60% of all blood collected), potentially reducing or avoiding exposure to allogeneic blood. An artificial O₂ carrier could also serve as a bridge to transfusion in case of emergency or during the time required for administered stored RBCs to restore their O₂-delivery capacity. A blood substitute that does not rely on human blood collection could play a considerable role in helping to relieve blood supply shortages. When combined with appropriate surgical procedures, blood substitutes could make the patient become his or her own blood

donor. For the developing countries, the availability of a blood substitute may be the best hope to meet future health care needs.

One difficulty concerning the evaluation of injectable O₂ carriers from a regulatory standpoint comes from the absence of a standard with which to compare them. Blood or RBCs have never been subjected to a controlled clinical trial for demonstrating their efficacy and have never been formally approved by a regulatory agency (26, 27). The same holds for the autologous transfusion alternatives, including preoperative donation and normovolemic hemodilution. Substantial efforts have been devoted to defining proper clinical endpoints that allow determining whether or not administration of an O₂ carrier translates into clinical benefit. Avoidance of transfusion of allogeneic blood during surgery appears now to be an accepted endpoint for such products.

PRINCIPLES OF OXYGEN DELIVERY BY PERFLUOROchemicals (26)

Stable and Passive Gas Solvents

PFCs are essentially made of carbon and fluorine instead of carbon and hydrogen as are most organic compounds. Transport of O₂ by PFCs relies on dissolution of the gas in a biocompatible liquid solvent, rather than on chemical binding to the carrier molecule as in the case of hemoglobin. The reason for using PFCs for in vivo O₂ delivery is their unique combination of high gas-dissolving capacity (the highest known among liquids), and exceptional chemical and biological inertness (they are among the most stable organic materials ever invented). The first of these properties is a direct consequence of the weakness of the cohesive forces (van der Waals interactions) between molecules in liquid PFCs, which facilitates the insertion of gas molecules within the liquid. The PFCs' inertness, on the other hand, reflects the strength of the intramolecular covalent bonds, and the high ionization potential and low polarizability of the fluorine atoms. The C—F bond is the strongest single bond encountered in organic chemistry. The bonds between perfluoroalkyl (*F*-alkyl; the prefix *F*-indicates that all hydrogen atoms are replaced by fluorine atoms in the structure) chains and oxygen, nitrogen, chlorine, or bromine are also usually strengthened. The high electron density of the fluorine atoms results in a compact, repellent electron shield that ensures effective protection of the molecule's backbone. PFCs are not subject to oxidation and do not undergo any reaction under the conditions of processing, storage,

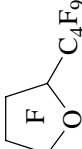
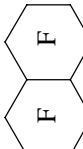
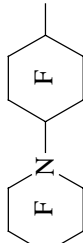
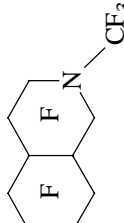
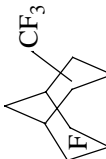
and use relevant to therapeutic O₂ delivery. Neat PFCs typically resist heating to 300°C, and PFC emulsions can be heat sterilized at 121°C. No enzymatic system is known to digest PFCs, and no microorganism is known to feed on them. PFCs are extremely hydrophobic, as well as lipophobic, which also contributes to their biocompatibility and largely determines their excretion rates. Table 1 lists the PFCs that have been most thoroughly investigated for the purpose of developing an injectable O₂ carrier.

The amount of gas dissolved in a PFC obeys Henry's law, i.e., is directly proportional to the partial pressure of the gas (Fig. 1). Since no chemical bonding is involved, there is no saturation. Uptake and release of O₂ by PFCs is essentially insensitive to temperature and environment. Oxygen content can be adjusted by simply controlling pO₂. The O₂ solubilities of the PFCs pertinent to intravascular use range from 40 to 50 vol% under one atmosphere (Table 1); i.e., they are larger than in water by a factor of 20 or more (28). Linear PFCs have an advantage over cyclic and polycyclic ones (29). The CO₂ solubilities of PFCs range from 140 to 240 vol%.

Pure Synthetic Carriers

Two major strategies are available for manufacturing PFCs, which consist either of substituting fluorine atoms for hydrogen atoms in the direct hydrocarbon analog of the desired PFC, or of combining smaller, already fluorinated, reactive building blocks to form the target PFC (30). Electrochemical fluorination in hydrogen fluoride, fluorination by high-valency metal fluorides (usually cobalt trifluoride) and direct fluorination by elemental fluorine, belong to the first category, while telomerization of tetrafluoroethylene (TFE) belongs to the second. Because of the higher strength of the C—F bond as compared to the C—H bond, substituting fluorine for hydrogen on carbon involves the release of large amounts of energy and results usually in complex mixtures. Yet several PFCs prepared by substitution methods have been used in O₂ carrier development (Tables 1 and 2). Telomerization, on the other hand, readily provides highly pure linear PFCs. Telomerization involves the reaction of a telogen, such as C₂F₅I, with an olefin, such as TFE, to form a series of longer PFC products, the telomers, in high yields. Examples of PFCs derived from telomerization of TFE include *F*-octyl bromide **1** (PFOB, also known as perflubron) and *F*-decyl bromide **2** (PFDB), the O₂-carrying components of *Oxygent*TM AF0144, an emulsion currently developed by Alliance Pharmaceutical Corp. (San Diego, California)

Table 1 Perfluorocarbons most thoroughly investigated for injectable oxygen carriers development^a

Structural formula	Code name (MW)	Preparation procedure (Purity)	Boiling point (°C), Vapor pressure (torr, 37°)	Solubility O ₂ , CO ₂ (vol%, 37°)	RES half-life (days)
CF ₃ (CF ₂) ₇ Br	1 PFOB Perflubron (499)	Telomerization (>99%)	143, 11	50, 210	~4
CF ₃ (CF ₂) ₉ Br	2 PFDB (599)	Telomerization (>98%)	180, 1.5		20–25
CF ₃ (CF ₂) ₃ CH=CH(CF ₂) ₃ CF ₃	3 F-44E (464)	Telomerization (>99%)	–, 12.5	50, 247	~7
	4 FX-80	Electrochemical fluorination	~103, 58	52, 190	
N[(CF ₂) ₃ CF ₃] ₃	5 FTBA	Electrochemical fluorination (80–85%)	178, 1.1	38, 140	>500
	6 FDC (462)	CoF ₃ (~97%, <i>cis</i> + <i>trans</i>)	142, 12.5	42, 142	~7
N[(CF ₂) ₂ CF ₃] ₃	7 FTPA (521)	Electrochemical, fluorination (>95%)	131, 18	45, 166	~65
	8 FMCP (596)	Electrochemical fluorination (~55%)	168, 2	40	~90
Cl(CF ₂) ₈ Cl	9 PFDCO (471)	Electrochemical fluorination	155, 6	43	~7
	10 FMIQ (495)	Electrochemical fluorination (~95%)	153–154	42	~11
	11 FDN (512)	CoF ₃ (50–55%)	157, 6	40	~14

^aThe symbol F within a cycle indicates that all carbons in the cycle are perfluorinated.

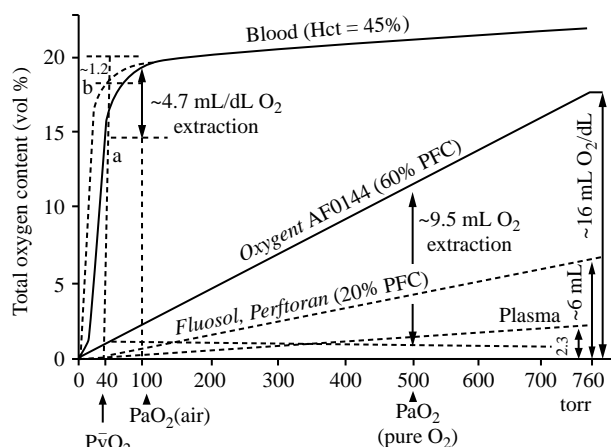


Fig. 1 Oxygen content of the fluorocarbon emulsions *Oxygent* AF0144, *Fluosol*, and *Perftoran* as compared to fresh (a) and stored (b) blood and plasma, as a function of oxygen partial pressure.

and 1,2-bis(*F*-butyl)ethene **3** (code-named F-44E), which is utilized by Neuron Therapeutics, Inc. (Malvern, Pennsylvania) for the development of an emulsion for the treatment of stroke (31).

Injectable Emulsions

Because they are virtually insoluble in water, PFCs are formulated as emulsions for parenteral administration. Obtaining stable, biocompatible, small-sized, narrowly distributed emulsions of an excretable PFC was crucial. The extreme hydrophobicity of PFCs can, however, render the obtaining of a stable emulsion difficult. Further critical aspects of PFC emulsion development involve emulsion formulation, physical and biological characteristics, scale-up, sterilization, and packaging, as well as its user-friendliness and strategy of use.

The submicron-size PFC droplets are coated with a thin layer of a surfactant that serves as an emulsifier and an emulsion stabilizer. A range of emulsion concentrations—from 20–100% w/v (i.e., about 11–52 vol%)—has been investigated (31–34). The present PFC emulsions use EYP as the emulsifier. These emulsions are, therefore, very similar to the Intralipid®-type parenteral lipid emulsions. Emulsion osmolality is independent of PFC concentration and is adjusted with salts or other tonicity agents. Particles of less than 0.2 μm evade phagocytosis more easily than larger ones, resulting in longer intravascular persistence and lesser side effects. Contrary to Hb solutions, PFC droplets do not normally filter out of the circulation.

Modest Oxygen-Transport Capacity that Translates into High Oxygen-Delivery Efficacy

Introducing a PFC emulsion into the circulation is akin to increasing the O_2 solubility of the plasma compartment of blood. The principles that underlie O_2 transport by the PFC and plasma are essentially the same. In both cases dissolution is proportional to pO_2 ; simply, the solubility of O_2 in PFCs is typically 20 times larger than for the plasma (28). The amount of O_2 dissolved in a given PFC emulsion is the product of the PFC's O_2 solubility coefficient, emulsion concentration, and O_2 partial pressure (Fig. 1). It is, for example, ~ 3 vol% at 37°C in a 60% w/v emulsion of PFOB in room air, and ~ 16 vol% in pure O_2 . In the blood of a patient breathing pure O_2 , the arterial O_2 tension is increased to about 500 torr and the emulsion will carry about 10.5 vol% of O_2 .

Oxygen release by PFCs to tissues is not dependent on a change in conformation of the PFC molecule or any cooperative effect, and does not require the assistance of an allosteric effector. It is greatly facilitated as compared to hemoglobin because the van der Waals interactions between O_2 molecules and PFCs are an order of magnitude weaker than the covalent $\text{O}_2\text{—Fe(II)}$ coordination bond, resulting in much higher extraction rates and ratios. The latter typically reach 90% with PFC emulsions as compared to about 25% for hemoglobin in normal conditions (35,36). This amounts to about 9.5 vol% O_2 for the above-mentioned *F*-octyl bromide emulsion. Oxygen release from PFCs is effective at any physiologically relevant partial pressure, rendering a cooperativity-like effect unnecessary. Likewise, O_2 release by PFCs is not dependent on pH and is not adversely affected by temperature. Since PFCs do not undergo oxidation or other modification over time, their O_2 uptake and release characteristics are not affected by storage. When hemoglobin and a PFC are present in the circulation simultaneously, the PFC will release its O_2 load first, thus conserving the O_2 bound to the hemoglobin.

Hemoglobin is exquisitely well adapted to supporting life in earth's atmosphere, but the conditions available in the operating room or critical care unit are different, and adjustable. A valuable consequence of O_2 dissolution in PFCs following Henry's law is that the transport capacity of a PFC emulsion can be increased by a factor of five by just increasing the fraction of O_2 in the air inspired by the patient (FiO_2).

Hemoglobin solutions and PFC emulsions thus cannot be compared on the sole basis of their static O_2 -binding or O_2 -dissolving capacities on a gram-per-gram basis in air. Such an approach would overlook the approximately fourfold higher tissue O_2 extraction ratio that is seen with

Table 2 Perfluorocarbon emulsions having reached some degree of commercial development

Trade name	Company	Perfluorocarbon ^a	Concentration v/v (w/v)	Surfactants	Remarks	Status (Sept. 2000)
Fluosol [®]	Green Cross Corp. (Japan)	FDC/FTPA 7:3 6/7	11% (20%)	Pluronic F68 EYP ^b K oleate	Frozen stem emulsion Reconstitute dilute	Approved in the U.S. for PTCA 1989. Discontinued
Oxypherol [®] (Fluosol-43)	Green Cross Corp. (Japan)	FTBA 5	11% (20%)	Pluronic F68	High organ retention	For experimental use. Discontinued
Perftoran	Perftoran Co. (Russia)	FDC/FMCP 7:3 6/8	11% (20%)	Poloxamer EYP	High organ retention	Approved in Russia 1996
FMIQ emulsion	Green Cross Corp. (Japan)	FMIQ 10	13% (25%)	EYP K oleate		Not developed
Addox [®]	Adamantech (USA)	"FMA" ^c FDN 11	21% (40%)	EYP	Low PFC definition	Abandoned
Therox [®]	DuPont (USA)	F-44E 3	40% (78%)	EYP	For research only	Discontinued
Oxyfluor [®]	HemaGen-PFC (USA)	PFDCO 9	40% (78%)	EYP	Stabilized with saffoil	Phase 2 clinical trials in CPB. Abandoned
Oxygent [®] (AF0144)	Alliance Pharmaceutical Corp./Baxter (USA)	PFOB/PFDB 1/2	32% (60%)	EYP	Stabilized with PFDB	Completed one Phase 3 in general surgery.

^aSee Table 1 for abbreviations.^bEYP = egg yolk phospholipids.^c"FMA" = perfluoromethyladamantane.

PFCs under normal conditions, and ignores the possibility of increasing O₂ dissolution in PFCs by a factor of almost five simply by giving the patient pure O₂ instead of air to breathe; i.e., ignores the actual conditions in which PFC products are to be utilized.

Eventually it is the amount of O₂ that is delivered to the tissues and the carrier's contribution to tissue O₂ consumption, $\dot{V}O_2$, that determines its effectiveness. Oxygen delivery, $\dot{D}O_2$ (the amount of O₂ offered to the tissues within a defined period of time), is the product of arterial blood O₂ content, CaO₂, and the cardiac output, CO (Eq. 1):

$$\dot{D}O_2 = (CaO_2) \times CO \quad (1)$$

When normovolemic hemodilution is performed, there is normally a substantial increase in cardiac output as a result of increased fluidity of the diluted blood (37–39). This increase in cardiac output is preserved when PFC emulsions are administered, which further enhances their O₂-delivery capacity (39, 40), while cell-free Hb products, because of vasoconstrictive effects, classically display an unchanged or reduced cardiac output that can negate the benefit of increased fluidity (41, 42).

The large pO₂ gradients that set in place at the high FiO₂ at which PFCs are utilized provide a strong driving force for O₂ diffusion from the PFC droplets to the tissues. The emulsion droplets being typically in the 0.1–0.2 μm size range, i.e., 30–70 times smaller than RBCs, circulate more easily in the capillary beds than RBCs. They will be present in large numbers in the plasma gaps that exist between red cells in the microcirculation, thereby increasing O₂ content. These plasma gaps are particularly large when the patient suffers from acute anemia or when hemodilution is practiced. It is likely that PFC particles not only transport O₂, but also facilitate its diffusion from RBCs to the tissues by providing numerous “stepping stones” or dynamic chains of particles for O₂ to travel on (36, 43). The effectiveness of PFC emulsions is expected to be greatest in the capillary beds, at low RBC concentrations, and when FiO₂ is high. Altogether, the contribution of PFCs to O₂ delivery and to tissue pO₂ can thus be highly significant, even though the amount of PFC-dissolved O₂ in the circulation is smaller than the amount of Hb-dissolved O₂.

The O₂ tension of the mixed venous blood, P $\bar{v}O_2$, i.e., the residual O₂ tension after the tissues have extracted the O₂ they needed, is often used as a global indicator of tissue oxygenation. A substantial increase in P $\bar{v}O_2$ was consistently observed upon administration of PFC emulsions and was paralleled by an increase in tissue pO₂ (36, 39, 44, 45). A physiological model was

developed that allows calculation of the contribution of a given dose of PFC to a given patient and was validated using clinical data (46). The contribution of the PFC to O₂ consumption could thus be described in terms of an “hemoglobin equivalent” value (47).

Early Emulsions—Fluosol®

The first physiologically adjusted emulsion of a PFC (*F*-butyltetrahydrofuran, FX-80, **4**) was used in 1967 (48) for oxygenating isolated rat brain. Shortly after, virtually all the RBCs of O₂-breathing rats were successfully replaced with a poloxamer-stabilized emulsion of *F*-tributylamine (FTBA, **5**) (49). FTBA was, however, retained in the animal's organs for an exceedingly long time. Fortunately, it was discovered that *F*-decalin (FDC, **6**) was excreted within a few weeks of its administration (organ half-life of about 7 days) (50, 51). However, it was realized that PFCs that gave stable emulsions were retained in the organism for an unacceptably long period of time, while PFCs that were excreted rapidly did not produce emulsions stable enough for practical use (29).

The first commercially developed emulsion, Fluosol-DA (Green Cross Corp. Osaka, Japan), was stabilized using a mixture of **6** and *F*-tripropylamine (FTPA, **7**, 30% of total PFC) (Table 2) (52, 53). However, the product's stability was still poor: Fluosol had to be frozen for shipment and storage, and reconstituted prior to use. Fluosol utilized a surfactant system that included poloxamer-188 (Pluronic® F-68) with smaller amounts of EYP and potassium oleate. The poloxamer provided steric stabilization, and potassium oleate introduced negative charges on the droplets, likely to oppose flocculation.

Another 20% w/v emulsion, trade-name Perftoran, based on an FDC and *F*-[1-(4-methylcyclohexyl)piperidine] (FMCP, **8**) mixture emulsified with a poloxamer was developed in Russia (Russian Academy of Sciences/Perftoran Company, Pushchino, Russia) (43, 54). Perftoran is filter-sterilized and stable for about 1 month at 4–8°C but requires freezing for longer storage. The Russian health authorities licensed it in 1997 for a wide range of indications. A highly stable 20% w/v *F*-tributylamine/Pluronic F-68 emulsion, Fluosol-43 (later renamed Oxypherol) (52), was commercially available for many years for experimental use.

SECOND GENERATION EMULSIONS (26)

Fluosol's shortcomings included prolonged organ retention of *F*-tripropylamine (reticuloendothelial system

(RES) half-life 65 days), complement activation, and hemodynamic effects due to Pluronic, excessive dilution, limited intravascular persistence, insufficient stability, and lack of user-friendliness. The product came as three separate preparations: the frozen stem emulsion and two annex salt solutions. The stem emulsion had to be carefully thawed, then admixed sequentially to the annex solutions, and the reconstituted product had to be used within 8 h. This cumbersome procedure, the short window for use, the further need for administering a small-test dose to patients prior to infusion in order to identify those patients who were sensitive to *Pluronic*, contributed to compromising the product's commercial success.

Assessment of Fluosol established that the emulsion did carry and deliver the expected amount of O₂ to tissues. It also established that large doses of PFCs could be administered without significant side effects and provided valuable information on the pharmacology of PFCs. An essential point learned from the development of Fluosol was that PFC emulsions needed to be ready for use, hence significantly more stable. At least 1 year of shelf stability under standard refrigeration storage conditions was indispensable. Particles needed to be small (0.1–0.2 μm) and remain as such during heat sterilization and throughout the product's shelf life, as small particle sizes translate into longer intravascular persistence and reduced side effects (55–57). It was also deemed necessary that the emulsion be relatively concentrated, yet fluid. The choice of a 60% w/v PFC concentration was determined by convenience of use in surgical applications, and marketing and manufacturing considerations.

SELECTING THE FLUOROCARBON—PERFLUOROOCTYL BROMIDE

Selection of an appropriate PFC and the production of a stable emulsion are key to the development of a safe and effective PFC-based O₂ carrier. The principal selection criteria for the PFC are rapid excretion and the aptitude at giving stable emulsions. Candidate PFCs also need to have high O₂ solubility and be well tolerated. Additionally, the selection process must take into account the PFC's manufacturability and cost.

Excretion was determined to be primarily an exponential function of the PFC's molecular weight (MW, Fig. 2). Structural features such as cyclization, branching, or the presence of heteroatoms have little influence other than changing the PFC's MW. The dependence of excretion rate on MW is consistent with the realization that excretion requires a certain degree of volatility and solubility of the

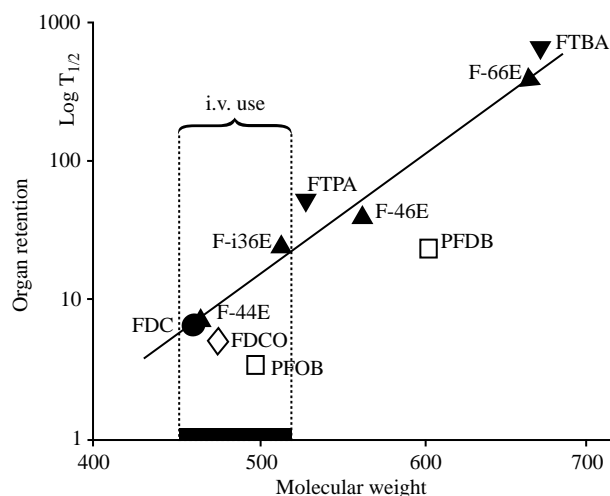


Fig. 2 Semilogarithmic plot of organ retention half-times for PFCs, as a function of molecular weight. Open symbols indicate PFCs with lipophilic character; F-nn'E: n and n' represent the number of carbon atoms in the homologous series C_nF_{2n+1}-CH=CHC_nF_{2n'+1}, i = iso; for the other symbols see Table 1.

PFC in lipids. However, as MW diminishes, vapor pressure tends to increase, which can eventually cause pulmonary complications (52, 58), thus setting an upper limit (around 20 torr) to the PFC's acceptable vapor pressure. The range of MWs adequate for intravenous administration was eventually established to be narrow (approximately 460–520) (29).

A candidate PFC was eventually identified that had a shorter organ-retention time than would have been predicted on the sole basis of its MW and was amenable to producing stable emulsions. This PFC was 1-*n*-*F*-octyl bromide (PFOB, 1). PFOB had initially been investigated because the radiopacity provided by its bromine atom allowed its use as a contrast agent (59).

PFOB's exceptionally fast excretion rate (~3 days in humans for a 2.7 g/kg dose) was attributed to the slightly lipophilic character induced by its well-exposed terminal bromine atom. This lipophilic touch facilitates uptake of the PFC by circulating lipids and transit through the organism (60, 62). In addition, this PFC can be manufactured in better than 99.9% purity. A 100-ton/year production capacity is already in place, which can easily be scaled higher. PFOB also ranks amongst the PFCs that have the highest O₂ and CO₂ solubilities (Table 1). Its emulsions show improved stability when phospholipids are the emulsifier. Its very low surface tension and positive spreading coefficient are advantageous in liquid ventilation, an application of PFOB that is currently in advanced clinical evaluation. Table 3 collects some of the physicochemical features of PFOB and FDC, and

Table 3 Physical properties of *F*-octyl bromide (PFOB, **1**) and *F*-decalin (FDC, **6**)

Property (units)	Symbol	PFOB	FDC (<i>cis</i> + <i>trans</i>)
Molecular formula		C ₈ F ₁₇ Br	C ₁₀ F ₁₈
Molecular weight (g/mol)	M_w	499	462
Molar volume (cm ³ /mol)	V_m	261	237
Molecular volume (Å ³)	V	432	393
Density (g/cm ³ , 25°C)	ρ	1.92	1.94
Melting point (°C)	m.p.	5	−10
Boiling point (°C)	b.p.	143	142
Vapor pressure (torr, 37°C)	v.p.	10.5	14
Heat of vaporization (kJ/mol)	ΔH_v	4.83	45.8
Refractive index (25°C)	n_D	1.30	1.313
Kinematic viscosity (centistokes, 25°C)	ν	1.0	2.9
Surface tension (mN/m)	σ	18.0	15
Interfacial tension vs. saline (mN/m)	σ^i	51.3	~60
Spreading coefficient (mN/m)	S (o/w)	+2.7	−1.5
O ₂ solubility (vol%, 25°C)	[O ₂]	50	40
CO ₂ solubility (vol%, 25°C)	[CO ₂]	~210	~140
Critical solution temperature (<i>n</i> -hexane, °C)	CST (hexane)	−20	+22
Solubility in water (mol/L)		5×10 ^{−9}	10×10 ^{−9}
Solubility in olive oil (mmol/L)		37	4.6

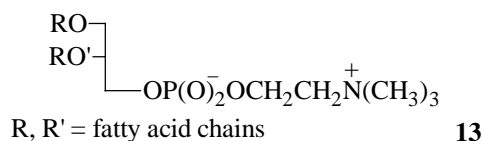
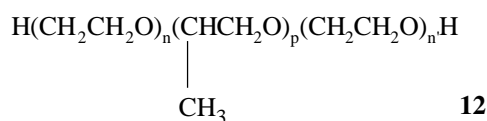
illustrates the differences that can exist among PFCs, including O₂ solubility, lipophilicity, viscosity, surface properties, and solubility in water (31). Another somewhat lipophilic fluorocarbon, α,ω -dichloro-*F*-octane (PFDCO, **9**), has been selected for development of another concentrated EYP-based PFC emulsion, Oxy-fluor[®] (HemaGen/PFC, St. Louis, Missouri) (63).

Selecting the Emulsifier—EYP

Proper selection of a biocompatible emulsifier or emulsifier system was also essential for successful PFC emulsion development. One of the emulsifier's roles is to reduce the large interfacial tension (σ^i) that opposes the dispersion of the very hydrophobic PFC ($\sigma^S \leq 20$ mN/m) into water ($\sigma^S = 72$ mN/m, 20°C). Another is to stabilize the emulsion once it is formed. The only two surfactants used in PFC emulsion development so far are poloxamers and phospholipids.

Poloxamers are neutral block copolymers such as **12**, consisting of two terminal hydrophilic polyoxyethylene blocks flanking a central hydrophobic polyoxypropylene block. Poloxamer 188 (e.g., Pluronic F-68) was used in the first generation PFC emulsions, but was far from adequate: Its surface activity is relatively poor, translating into low emulsions stability; the purity of the commercial products is usually rather low; its cloud point (~ 110 – 115°C)

prevents sterilization at the standard temperature of 121°C; its tendency to form gels limits the PFC concentration in the emulsions; and, finally, Pluronic F-68 has been found to be responsible for the unpredictable transient complement activation-mediated anaphylactic reaction observed in some patients in response to the injection of Fluosol (64–66).



EYP have been chosen as the emulsifier in all second generation PFC emulsions. EYP, whose major constituents are the amphiphilic amphoteric phosphatidylcholines **13**, provide significantly better emulsion stability than poloxamers. The stabilization effect is particularly remarkable with PFOB (31, 67, 68). Emulsions of PFCs prepared with EYP do not cause complement activation (56, 65). Phospholipids have a long history of use in pharmaceuticals; their pharmacology is well documented and there exist reliable commercial sources

of pharmaceutical grade EYP. The hydrolysis of EYP in PFC emulsions is minimal when pH is close to neutral; its oxidation can be minimized by adding small amounts of a metal chelator and an antioxidant, and through manufacture and packaging under nitrogen.

Stabilizing the Emulsion—A Lipophilic “Heavier” Perfluorocarbon

Stability is an essential condition for PFC emulsions to be of practical use. The principal mechanism for irreversible droplet growth in submicronic PFC emulsions during storage is molecular diffusion (also known as Ostwald ripening or isothermal distillation) (29,31,69,70). Coalescence may contribute to instability when mechanical stress is applied and at higher temperatures, as during heat sterilization. Sedimentation and flocculation are fully reversible and pose no problem. Molecular diffusion involves the transfer of individual PFC molecules from the smaller droplets, where the chemical potential is higher due to the higher curvatures of the particles (Kelvin effect), through the continuous aqueous phase to join larger ones, resulting in irreversible increase of droplet size over time. The Lifshitz–Slezov–Wagner theory predicts a linear increase in number average particle radius \bar{a} versus time (Eq. 2) (31,70,71), which was consistently observed experimentally with PFC emulsions (Fig. 2):

$$\frac{d}{dt}(\bar{a}^3) = \omega = \frac{8\sigma^i DC(\infty)V_m}{9RT} \gamma(\phi) \quad (2)$$

Of the parameters of Eq. 2, only $C(\infty)$, the solubility of the PFC in the aqueous phase, varies significantly among PFCs, resulting in a dramatic effect on particle size growth over time (Fig. 3).

Molecular diffusion in emulsions can be effectively slowed down by including in the dispersed phase a small amount of a component with lesser water solubility (71), in the case of a PFC emulsion, a secondary, higher MW (“heavier”) PFC (31,69,70). This is the role of **7** in Fluosol and of **8** in Perftoran, however, at the cost of longer organ half-lives of ~65 and ~90 days, respectively.

In *Oxygent* AF0144, the inevitable increase in organ retention with increasing MW was mitigated by choosing a *lipophilic* homolog of PFOB, PFDB (RES half-life of about 25 days), as the “heavy” stabilizing additive (31,62).

Fluorinated surfactants (or fluorosurfactants, i.e., surfactants with hydrophobic tails comprising a fluorocarbon moiety) provide an alternative means of achieving

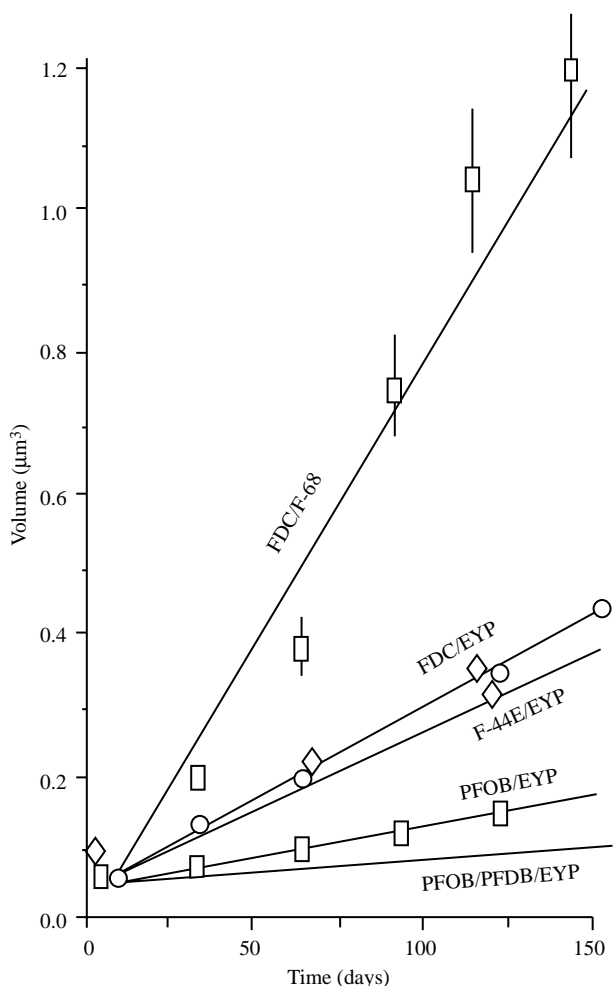
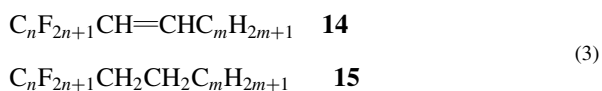


Fig. 3 Particle size increase over time for emulsions made of diverse PFCs and surfactants (F-68: Pluronic F-68; for the other symbols see Table 1). (Adapted from Ref. 32.)

extremely stable PFC emulsions, as they can provide very low PFC/water interfacial tensions (σ^i , another factor in Eq. 2) (72). As yet, this option has not been developed, in part because of the added cost involved in the evaluation for approval of a novel active excipient. A further means of effectively increasing the stability of EYP-based PFC emulsion consists of supplementing standard phospholipids with mixed fluorocarbon–hydrocarbon diblock compounds, such as **14** or **15** (73). Such diblocks, which have fluorophilic–lipophilic amphiphilic properties, are expected to improve the adhesion of the phospholipid film onto the PFC droplet.



Raw Material Procurement and Emulsion Manufacture

Reliable access to raw material, ease of manufacture, and cost-effectiveness will play a critical role in determining the degree of acceptance, breadth of application, and commercial success of an O₂ carrier. To have a significant impact on medical practice, O₂ carriers must be able to replace a fair proportion of the blood and packed RBCs that are presently being transfused. A mere 10% of this amount represents over one million RBC units in the United States alone and several million units worldwide. Only O₂ carriers that can be produced in that order of magnitude at a cost that remains in the vicinity of that of blood, can potentially mitigate the current and projected blood shortages.

PFCs, being synthetic, have obvious advantages from the standpoints of raw material procurement, purity, safety, and cost-effectiveness. Highly pure PFCs can be manufactured, in any desired amount, within tight specifications, using well-established procedures. PFOB, for example, is directly derived from the production line that leads to polytetrafluoroethylene (e.g., Teflon®) and to diverse large-tonnage industrial surfactants.

Technology for large-scale production of injectable emulsions in compliance with good manufacturing practices is well established in the pharmaceutical industry. Manufacture of a sterile PFC emulsion for parenteral use, although it relies on existing technologies, required the development of specific know-how. Egg phospholipids have a long history of use in pharmaceuticals, including preparation injectable lipid emulsions and, more recently, liposomes. The formulation process for PFC emulsions being additive, its yield is essentially quantitative with respect to the raw materials utilized. The present products are terminally heat-sterilized in standard conditions. The investment in a given size production unit may probably be as much as an order of magnitude lower for PFC emulsions than for any hemoglobin product. Small-size PFC particles and narrow-size distributions have been achieved reproducibly.

For EYP-based PFC emulsions, the first step of the process involves dispersing of the water-insoluble phospholipid in a saline solution. The PFC is then added to this saline phase, where it is broken down into fairly large droplets (av. $\sim 5 \mu\text{m}$) with a high shear rotor-stator type homogenizer. This premix undergoes final emulsification, using a high-pressure homogenizer that provides a high-energy density (31). Minimal exposure to oxygen (through nitrogen sparging and blanketing), pyrogen-free water-for-injection and a particulate-free environment are used throughout processing. The Gaulin-type high-pressure homogenizer is the equipment of choice for

industrial production of pharmaceutical emulsions as it is easy to control, gives narrow, consistent particle size distributions and can be operated on very large scales.

The Current *Oxygent* AF0144 Emulsion

A concentrated (60% w/v) PFOB/PFDB emulsion (*Oxygent* AF0144) is currently being produced in a commercial-scale facility (31). The emulsion is steam-sterilized in a rotary autoclave at or above 121°C, using a procedure that achieves uniform heat penetration, maintains emulsion integrity, and provides the required probability of less than one nonsterile unit in 10⁶. As compared to Fluosol, *Oxygent* is characterized by use of PFCs having some lipophilic character, use of phospholipids as the emulsifier instead of poloxamer, a several-fold increase in PFC concentration, simplification of the overall formulation, considerable increase in stability, and consequently, far superior convenience of use.

Oxygent AF0144 is a ready-for-use, pH-balanced, isotonic fluid emulsion of *F*-alkyl bromides. It comprises 60 wt% PFC (about 32 vol%) consisting primarily of PFOB (AtoFina, Pierre Bénite, France), a small percentage of PFDB as the stabilizing additive and EYP as the emulsifier. Osmolarity is adjusted with NaCl, and pH with a phosphate buffer. Minute amounts of α -D-tocopherol and EDTA are included to protect the phospholipids against oxidation. Average droplet size, after heat sterilization, is about 0.16 μm and viscosity around 5 cP (extrapolated at zero shear rate). The product has a shelf life of at least 18 months at 5–10°C (31). Other post-Fluosol formulations with EYP as the emulsifier are listed in Table 2.

PHARMACOKINETICS

The PFC droplets in the circulation are opsonized and progressively cleared from the blood stream by circulating monocytes and fixed macrophages of the RES, the larger droplets being removed first. This mechanism is responsible for the limited intravascular persistence of the emulsion. The PFC is temporarily stored in the primary organs of the RES, the liver, spleen, and bone marrow. Subsequently, the PFC molecules slowly diffuse across cell membranes (the rate determining step) from the RES organs back into the blood, where they are taken up by circulating lipid carriers (lipoproteins and chylomicrons) (31, 56, 60, 74). The rate of these steps depends critically on the PFC's MW and lipophilicity. The PFC is eventually excreted unchanged from the blood through the alveoli in the expired air. Some of the PFC also distributes into

adipose tissue and other lipid-containing tissues, before being excreted by the same pathway as above. There are no reports of PFC metabolism or enzymatic degradation.

The emulsion's circulatory half-life, $t_{1/2}$, is particle size-, species-, and dose-dependent. It increases significantly when the size of the droplets decreases (52, 57). Dose-dependence in humans is illustrated by $t_{1/2}$ values of 7.5 and 22 h for doses of 4 and 6 g of PFC/kg, respectively, for Fluosol (52), and 6.1 and 9.4 h for doses of 1.2 and 1.8 g/kg, respectively, for *Oxygent* (75). Particles of less than 0.2 μm are preferable, as they facilitate O_2 diffusion and are less rapidly phagocytized than larger particles, resulting in longer intravascular persistence and fewer side effects. The effects of PFCs on RES organ morphology and function have been thoroughly investigated, found fully reversible, and considered nondetrimental (56, 76, 77).

Four-compartment pharmacokinetic models (PFC emulsion in blood, PFC in RES tissues, PFC in non-RES tissues, PFC dissolved in blood lipoproteins) indicated that the rate-determining step in PFC excretion was the dissolution of the PFC into lipid carriers, confirming the critical importance of the lipid solubility of the PFC in the excretion process (31).

EFFICACY

The efficacy of PFC emulsions at delivering O_2 in vivo has been established in numerous experimental animal models and in clinical trials with first and second generation products. Oxygen-breathing animals, exchange-perfused with PFC emulsions to an hematocrit of 1%, have been shown to survive the exchange (49, 67). Mixed venous O_2 tension, $\text{P}\bar{\text{v}}\text{O}_2$, increased significantly in spite of decreased total arterial O_2 content. When RBCs were present, hemoglobin remained close to fully saturated with O_2 .

Clinical trials conducted with Fluosol have established that the product could contribute meaningfully to O_2 consumption in anemic patients and had an immediate beneficial impact on the condition of these patients (78–81). For example, administration of a 4 g/kg body weight dose of PFC to severely anemic O_2 -breathing surgical patients, while only increasing arterial O_2 content by a modest 0.8 vol%, provided 7% of the O_2 delivered and a significant 24% of the patient's O_2 consumption (78). $\text{P}\bar{\text{v}}\text{O}_2$ was seen to increase by approximately 60%, and the mixed venous Hb saturation ($\text{S}\bar{\text{v}}\text{O}_2$) reached 90%. The amount of O_2 provided by Fluosol was deemed clinically important. Another study with highly anemic surgical patients confirmed that Fluosol unloaded O_2 very effectively and contributed at least as much to O_2

consumption ($28 \pm 7\%$) as the patients' own RBCs (79). However, the benefit being short-lived, the outcome for patients who refused blood transfusion on religious grounds could not be improved, indicating that treatment of sustained anemia was not a valid indication for such a product. As a logical result, the Food and Drug Administration did not approve Fluosol in the United States when submitted for this indication in the early 1980s. Subsequent approval of Fluosol in 1989 for use during an angiography procedure in high-risk patients implies that both efficacy (reduced ischemia of the myocardium (82, 83)) and safety had been demonstrated for this application. Fluosol did not gain acceptance, not for lack of efficacy in delivering O_2 or because of side effects, but because of poor stability, inappropriate indication and strategy of use, and because of the development of autoperfusion catheters for the percutaneous transluminal coronary angioplasty procedure.

The efficacy of *Oxygent* AF0144 in terms of improving systemic oxygenation status has been demonstrated in various animal models (39, 84) as well as in surgical patients (85). The capacity of PFCs to contribute significantly to O_2 delivery was in agreement with computer simulation. For example, a 2.7 g PFC/kg dose administered to O_2 -breathing dogs, while contributing only 8–10% to total blood O_2 content, accounted for 25–30% of total O_2 consumption, $\dot{\text{V}}\text{O}_2$ (84) allowing O_2 consumption to be maintained while Hb levels were decreased to values

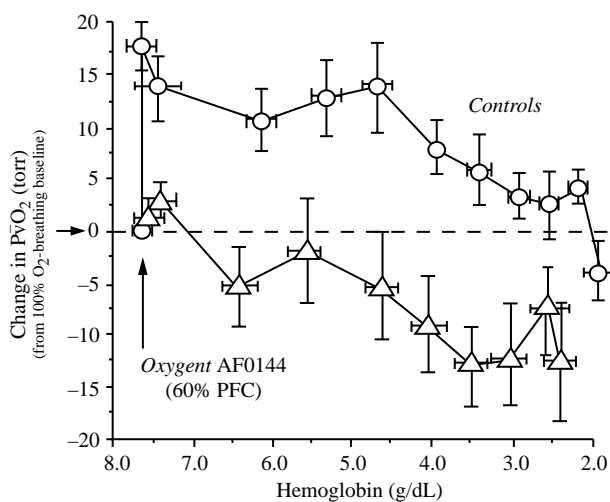


Fig. 4 Efficacy of *Oxygent* in a canine model mimicking surgical blood loss after acute normovolemic hemodilution. Both the treatment group and the control group were hemodiluted to a hemoglobin level of 8 g/dl, breathed oxygen and lost blood, hence hemoglobin (x axis). Change in $\text{P}\bar{\text{v}}\text{O}_2$ relative to O_2 -breathing baseline during volume-compensated blood loss (y axis) is significantly different between the two groups. (From Ref. 45.)

as low as 2.0 g/dl (Fig. 4). The PFC-treated dogs could lose almost 70 ml of blood per kg (about two-thirds of their blood), compared to 10 ml/kg in controls, before $P\bar{v}O_2$ fell below the initial 100% O_2 -breathing baseline. Tissue pO_2 of skeletal muscle, gut, and brain was seen to increase substantially. Another study of this type concluded, on the basis of both indirect global indicators ($P\bar{v}O_2$, $\dot{V}O_2$) and direct measurements of tissue pO_2 on the surface of liver and skeletal muscle, that the emulsion was as effective as fresh autologous RBC transfusion in maintaining tissue oxygenation during volume-compensated blood loss designed to simulate surgical bleeding (39). Hemodilution of the dogs could be extended to a 3 g/dl Hb level without impairment of tissue oxygenation. Likewise, dogs that had been hemorrhaged until their $P\bar{v}O_2$ was equal to or less than 25 torr, when resuscitated with Oxyfluor, restored their $P\bar{v}O_2$ and all survived, while only 62% of the Ringer's solution-treated control animals treated survived, and the latter had significantly lower $P\bar{v}O_2$, i.e., less adequate tissue oxygenation (86). In a model of surgical hemodilution to a hematocrit of 20%, the PFC provided 40% of the O_2 consumption (63). Clinical efficacy of *Oxygent* was demonstrated in a Phase 3 clinical study in surgical patients. (See section on Augmented Acute Normovolemic Hemodilution (A-ANHSM).

Using *Oxygent*, further demonstration that small doses of PFCs can contribute significantly to tissue oxygenation (26, 32, 33, 45) includes: increase in maximal O_2 consumption by an isolated working skeletal muscle in the presence of PFC; effective cardiac muscle function preservation in a canine model of low-flow coronary ischemia; sustainment of cardiac tissue oxygenation for a prolonged period following arrest of coronary perfusion of isolated rat hearts; improved myocardial O_2 delivery and aerobic metabolism during cold cardioplegic arrest of isolated rabbit heart; graded increase in $P\bar{v}O_2$ during cardiopulmonary bypass in dogs in response to graded increases in PFC dosing; preservation of cerebral function following experimental brain stem ischemia in dogs; significant increase in retinal O_2 tension and brain oxygenation in cats; improvement in the cat's primary visual cortex tissue pO_2 upon intravenous infusion of the emulsion; restoration of cerebral O_2 delivery in severely hemorrhaged rats; improvement of systolic function and reduction of myocardial edema and acidosis with PFC-supplemented cardioplegia in dogs that underwent CPB, global myocardial ischemia, followed by cold cardioplegic arrest, and eventually normothermic reperfusion; preservation of regional organ perfusion in myocardium, intestinal mucosa, liver, lung, kidney, and skeletal muscle; improvement of myocardial oxygenation and tolerance to low-flow ischemia (as found with

certain surgical procedures) in isolated rabbit hearts; improved tumor oxygenation and increased tumor radio- and chemosensitivity.

SAFETY

The acute toxicity of properly prepared PFC emulsions is low. The intravenous LD_{50} of Fluosol was estimated at 26–29 and 35 g PFC/kg body weight in rats for 20% and 35% w/v concentrated emulsions, respectively, indicating that the volume injected (130–145 and 101 ml/kg) may have had a part in these LD_{50} values (52). Similar figures were reported for *Perftoran* (43). For PFOB the LD_{50} reached 41 g/kg body weight when a 100% w/v emulsion (52 vol%) was administered to rats (87). No carcinogenic, mutagenic, teratogenic effects, or immunogenicity have been found for properly purified and formulated PFC emulsions. The anaphylactoid reaction observed with some patients with Fluosol (64) was determined to be due to complement activation by the poloxamer used as a surfactant in the formulation. It was no longer observed with the EYP-based emulsions (56, 65, 88).

The side effects observed in the clinic with the early 100% and 90% w/v concentrated PFOB formulations (56) and with Oxyfluor (63) consisted of early effects, during or shortly after infusion, including headache and occasional lower backache, and delayed effects (2–12 h), referred to as flu-like symptoms; e.g., fever, occasional chills and nausea (56, 75). These reactions, generally categorized as mild, were transient and fully reversible within 4–12 h. A transient, moderate drop (about 15%) in platelet count was seen about 3 days after dosing. Similar effects have been documented for parenteral fat emulsions and liposomes, indicating that these effects were likely related to the particulate nature of the emulsion.

The mechanism of the flu-like side effects was shown to be, for most part, related to the progressive clearance of the droplets from the blood stream. The effects are considered the natural consequence of macrophage activation during phagocytosis, which is accompanied by the release of products from the arachidonic acid cascade, including diverse prostaglandins and pyrogenically active cytokines (56, 63).

The magnitude and frequency of the particle-related side effects depend strongly on particle sizes (52, 57). Adjustments in the *Oxygent* emulsion formulation and processing parameters (and, in particular, the addition of PFDB, which helped reduce particle size and narrow particle size distribution) resulted in significant attenuation of the side-effect profile (57). During clinical safety

studies in conscious volunteers with the optimized 60% w/v *Oxygent* formulation AF0144, fever was less frequent than seen with an earlier formulation and seldom exceeded 1°C, and platelet count, although temporarily depressed with respect to baseline, remained within normal range (75, 89). In the normal volunteers, there were no effects on platelet function and coagulation parameters, complement activation, immunologic or allergic reactions, vasoconstriction or microcirculatory disturbances, no abnormal changes in liver function, pulmonary or renal function, or clinically meaningful effects on blood chemistry at the doses administered (90). No complement activation was found either with *Oxyfluor* (88). No emulsion-related serious adverse clinical events were reported in a subsequent Phase 2 efficacy study in orthopedic surgery patients (91).

PFCs have no capability to react with and scavenge nitric oxide, and clinical trials with EYP-based PFC emulsions have shown no perturbation of hemodynamics, increased vascular resistance and arterial tension, reduced cardiac output or heart rate. Cardiac output increased normally in response to hemodilution. Coagulation function, including bleeding times, were unaffected by the administration of *Oxygent* (89).

The observation that the lungs of rabbits given Fluosol or emulsions containing volatile PFCs, such as FX-80 or FDC, did not deflate normally at autopsy raised concerns about a possible toxic effect of these PFCs in the lungs (58). This phenomenon, which relates to increased pulmonary residual volume (IPRV) due to retention of air in the alveoli (92), is species-dependent and dependent on the PFC's vapor pressure. No such effect was reported for the patients who had received Fluosol, although it contained FDC. Addition of a small percentage of PFDB to PFOB in *Oxygent* further reduced the vapor pressure of the PFC phase to less than 8 torr, which sufficed to eliminate IPRV in most of the sensitive animal species (56).

PRINCIPAL INDICATIONS

Surgery

A recently defined, novel strategy for reducing the need for donor blood transfusions and increasing patient safety during surgery involves the use of an O₂ carrier in conjunction with acute normovolemic hemodilution in the so-called Augmented-ANH (or A-ANH) procedure (26, 91, 93). The A-ANH procedure (Fig. 5) consists first in collecting a portion of the patient's blood, usually 2 to 4 units, immediately before surgery begins. This blood is set aside in the operating room, and is replaced by a

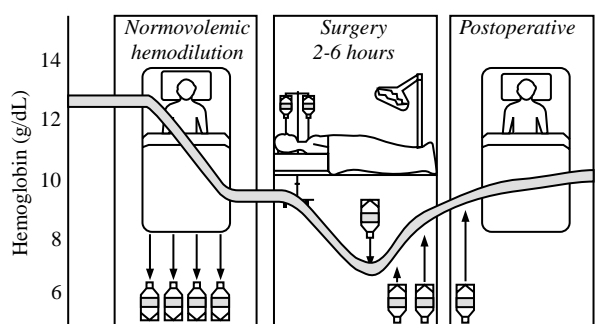


Fig. 5 In the augmented acute normovolemic hemodilution (A-ANH) procedure, two to four units of blood are withdrawn from the patient just before undergoing surgery and are replaced by a volume-expanding solution. When, during surgery, the need for transfusion is determined, a PFC emulsion is administered instead of RBCs, thus preventing tissue hypoxia from occurring. When the hemoglobin concentration reaches a level that is considered unsafe, or at the end of surgery, the patient receives his/her own fully effective fresh blood back. The A-ANH procedure is expected to provide increased safety and reduced exposure to allogeneic blood, and to help reduce blood shortages. (Courtesy of Alliance Pharmaceutical Corp.)

plasma expander to maintain constant circulatory volume. When, during the surgery, the patient would normally be transfused, the patient receives the O₂ carrier instead. Tissue oxygenation thus remains in safe control and, by operating at a lower hematocrit, fewer RBCs are lost during surgical bleeding. If the hemoglobin reaches a level that is deemed too low, or at the end of the surgery, the patient receives back his or her own safe, fresh, and fully functional blood.

The A-ANH procedure is expected to allow more profound hemodilution without impairment of tissue oxygenation, resulting in reduced exposure to allogeneic blood transfusion (26, 40, 93, 94). Because the emulsion enables the physician to perform ANH in a safer and more effective way, it should help make ANH available to a broad population of surgical patients (about 6 million a year in the United States). Additionally, by allowing each patient to be his or her own donor, this method also has the potential to improve the overall availability of blood and substantially relieve banked blood shortages. This strategy has become a primary target indication for O₂ carriers.

Clinical trial data with *Oxygent* confirmed the expectations. Even a low 0.9 g/kg dose of PFC, administered after ANH, allowed P \bar{v} O₂ (85) values to remain at or above predosing levels while Hb levels decreased substantially due to surgical blood loss. Multicenter randomized Phase 2 trials involving patients undergoing elective surgery with ANH, determined that a 1.8 g/kg dose of PFC was significantly more effective than

fresh autologous blood (which is more effective than 2,3-DPG-depleted stored RBCs) at reversing physiological transfusion triggers (40, 75, 91). Use of the PFC emulsion also substantially delayed the need for transfusion of the stored autologous blood. The benefit of reduced viscosity on cardiac output was fully preserved. An hemoglobin equivalency for PFOB was established from the clinical data to be about 1.5 g of hemoglobin for 1 g of PFOB in the above conditions of use (47). Alliance Pharmaceutical Corp. recently announced that its Phase 3 clinical study with *Oxygent* in surgical patients in Europe (33 centers in 8 countries) was successful in reducing the need for donor blood transfusion. The intent of the study was to determine whether the use of *Oxygent*, administered according to the A-ANH procedure, would reduce the need for donor blood when compared to standard transfusion therapy. Statistically significant reduction in donor blood use was found in the entire study population. The target population consisted of those patients (86% of all patients enrolled) for whom a transfusion is most likely needed (i.e., experiencing modest to high surgical blood loss). In this population *Oxygent* provided highly significant avoidance ($p = 0.002$) or reduction ($p < 0.001$) of donor blood transfusion versus control patients receiving standard transfusion therapy. No safety issues were reported.

Cardiopulmonary Bypass Surgery and Neuroprotection

A-ANH with PFC emulsions is expected to have significant value in cardiopulmonary bypass (CPB) surgery (95, 96). Addition of the O₂ carrier to the solution that is used to prime the extracorporeal circuit should reduce the need for allogeneic RBC transfusion while increasing O₂ supply. Phase 3 clinical trials are currently underway in Europe and the United States in surgery patients with CPB support. In addition, because PFCs are able to dissolve nitrogen as well as oxygen, they could dissolve the tiny air bubbles that may be introduced in the CPB circuit and may cause microemboli. They could thus reduce the incidence and severity of the neurological deficits observed in a significant proportion of the patients undergoing cardiovascular surgery with CPB.

A Bridge to Transfusion

There are a number of instances where administration of a PFC emulsion could help bridge the time gap between a critical need for increased tissue oxygenation and transfusion of compatible blood, or between the time of transfusion and that when the transfused banked blood

becomes fully effective. Trauma is one such situation, in particular during the prehospital “golden hour” period that largely determines the outcome for the patient. During this period, blood is usually not available and transfused stored blood not yet fully operative. Fluorocarbon emulsions may provide a means of stabilizing the patient waiting for an intervention, and could therefore find their place in any ambulance or rescue vehicle.

Further Indications for PFC Emulsions (26, 32–34, 40, 34, 97–99)

Potential cardiovascular therapeutic uses for PFC emulsions, other than CPB, include treatment of acute myocardial infarction, cardioplegia, reperfusion, coronary angioplasty and preservation of donor hearts for transplantation. High O₂-delivery capacity, small particle sizes, and low viscosity may improve tissue perfusion and oxygenation. Treatment of cardiac arrest is also being explored.

Stroke is one of the leading causes of death in the industrialized countries. The brain is extremely sensitive to O₂ deprivation. In the case of vessel obstruction, the small size PFC droplets could improve perfusion and provide support to ischemic tissues by using smaller collateral vessels. Contrary to blood, the emulsions do not tend to have increased viscosity at low shear rate. Perfusion of the brain with PFC emulsions through the ventriculo-cisternal route is also being investigated.

Solid tumors, because of poor vascularization, generally contain hypoxic cancer cells that are resistant to treatment. PFC emulsions can deliver O₂ deep into tumor regions that would otherwise be hypoxic, thereby improving the response of tumor cells to radio and chemotherapy without compromising the tolerance of normal tissues.

Use of PFC-enriched perfusates has the potential for increasing the availability and quality of organs suitable for transplantation. Kidney, heart, liver, lung, pancreas, testis, and multiple organ blocks have been preserved using such preparations.

Various types of PFC-based contrast agents for diagnosis using X-ray computed tomography (CT), magnetic resonance (MR), and ultrasound (US) imaging have been investigated. PFC emulsions also allowed mapping O₂ in tissues by exploiting the ability of the paramagnetic O₂ molecule to perturb the ¹⁹F nuclear magnetic resonance signals.

Further potential applications for O₂-carrying PFC emulsions include treatment of sickle cell disease; use of intraperitoneal perfusion as an alternative to pulmonary oxygenation during certain forms of respiratory failure;

treatment of CO intoxication; treatment of decompression sickness; "total body washout" for removal of toxins, viruses, drug overdoses, etc.; treatment of acute pancreatitis; protection of the gastric mucosa against damage provoked by hemorrhagic shock; stabilization and control of animal models, organs, and tissues; and use as drug-delivery systems.

Medical Applications of Perfluorochemicals in Other Forms than Emulsions (72, 98, 99, 100)

PFC-stabilized gaseous microbubbles are being developed as contrast agents for ultrasound imaging. Externally applied PFC-filled pads are commercially available (*SatPad*[®], Alliance Pharmaceutical Corp.) that improve magnetic homogeneity, hence image quality, when fat saturation techniques are utilized during MR imaging. Neat PFCs are used as an ocular tamponade for treating complicated retinal detachments or managing the dislocated crystalline lens. Further potential applications include use in blood oxygenators; prevention of the bends by liquid breathing; treatment of ischemic ulcers; preservation of plant and animal semen, tissues and transplants; increasing growth rates of animal and plant cell cultures; and delivery of drugs to the lung. Fluorinated amphiphiles are critical components of a variety of colloidal and supramolecular systems with potential in drug delivery.

PROSPECTS

Although use of hemoglobin as an O₂ carrier for developing a blood substitute appears to be a *natural* approach and therefore attracted much interest and effort, an increasing body of experimental and clinical data indicates that the PFC approach has its advantages as well. These advantages include greater simplicity, chemical and biological inertness, and large contribution to O₂ consumption relative to O₂ content due to dissolution proportional to FiO₂, effective extraction and preservation of increased cardiac output following normovolemic hemodilution. The fact that PFC-based O₂ carriers do not rely on the collection of human or animal blood or on the genetic engineering of a mutant hemoglobin is another definite advantage. PFCs can be produced in virtually any amount to satisfy the needs. The emulsion manufacturing process being additive results in high yields. It is also very cost-effective. PFC emulsions are heat sterilizable, ready for use and have a long shelf life. They elicit no vasoconstriction or other

pharmaceutical activities that could potentially compromise O₂ delivery to tissues.

PFC emulsions, together with augmented acute normovolemic hemodilution potentially provide a straightforward means of better managing a patient's *own* fresh blood during surgery, thereby contributing to improved safety and relief from blood shortages. Blood substitutes are thus expected to profoundly change the practice of transfusion medicine.

Future research will certainly aim at further increasing emulsion stability, reducing side effects, prolonging circulation life, and exploring novel therapeutic indications for PFC-based O₂ delivery systems.

REFERENCES

- Schreiber, G.B.; Murphy, E.L.; Horton, J.A.; Wright, D.J.; Garfein, R.; Chien, H.C.; Nass, C.C. Risk Factors for Human T-cell Lymphotropic Virus Types I and II (HTLV-I and -II) in Blood Donors: The Retrovirus Epidemiology Donor Study. *J. Acquir. Defic. Syndr. Hum. Retrovirus* **1997**, *14*, 263–271.
- Fiebig, E. Safety of the Blood Supply. *Clin. Orthopaed. Rel. Res.* **1998**, *357*, 6–18.
- Goodnough, L.T.; Brecher, M.E.; Kanter, M.H.; Aubuchon, J.P. Transfusion Medicine: Blood Transfusion. *New Engl. J. Med.* **1999**, *340*, 438–447.
- Zuck, T.F. Legal Liability for Transfusion Injury in the Acquired Immune Deficiency Syndrome Era. *Arch. Pathol. Lab. Med.* **1990**, *114*, 309–315.
- Dodd, R.Y. Transmission of Parasites by Blood Transfusion. *Vox Sang.* **1998**, *74*, 161–163.
- Scott, M.R.; Will, R.; Ironside, J.; Nguyen, H.-O.B.; Tremblay, P.; DeArmond, S.J.; Prusiner, S.B. Compelling Transgenic Evidence for Transmission of Bovine Spongiform Encephalopathy Prions to Humans. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 15137–15142.
- Andrews, N.J.; Farrington, C.P.; Cousens, S.N.; Smith, P.G.; Ward, H.; Knight, R.S.G.; Ironside, J.W.; Will, R.G. Incidence of Variant Creutzfeldt–Jakob Disease in the UK. *Lancet* **2000**, *356*, 481–482.
- Ricketts, M.N.; Cashman, N.R.; Stratton, E.E.; ElSaadany, S. Is Creutzfeld–Jakob Disease Transmitted in Blood. *Emerg. Infect. Dis.* **1997**, *3*, 155–163.
- Mitka, M. Will TSEs Taint the U.S. Blood Supply. *J. Am. Med. Assoc.* **1999**, *281*, 1157–1158.
- Landers, D.F.; Hill, G.E.; Wong, K.C.; Fox, I.J. Blood Transfusion-Induced Immunomodulation. *Anesth. Analg.* **1996**, 187–204.
- Wheatley, T.; Veitch, P.S. Effect of Blood Transfusion on Postoperative Immunocompetence. *Brit. J. Anaesth.* **1997**, *78*, 489–492.
- Houbiers, J.G.; van de Velde, C.J.; van de Watering, L.M.; Hermans, J.; Schreuder, S.; Bijnen, A.B.; Pahlplatz, P.; Schattenkerk, M.E.; Wobbes, T.; de Vries, J.E.; Klementsich, P.; van de Maas, A.H.; Brand, A. Transfusion of Red Blood Cells is Associated With

- Increased Incidence of Bacterial Infection After Colorectal Surgery: A Prospective Study. *Transfusion* **1997**, *37*, 126–134.
13. Murphy, P.; Heal, J.M.; Blumberg, N. Infection or Suspected Infection After Hip Replacement Surgery With Autologous or Homologous Blood Transfusions. *Transfusion* **1991**, *31*, 312–317.
 14. Heiss, M.M.; Mempel, W.; Jauch, K.-W.; Delanoff, C.; Mayer, G.; Mempel, M.; Eissner, H.-J.; Schildberg, F.-W. Beneficial Effect of Autologous Blood Transfusion on Infectious Complications After Colorectal Cancer Surgery. *Lancet* **1993**, *342*, 1328–1333.
 15. Purdy, F.R.; Tweeddale, M.G.; Merrick, P.M. Association of Mortality With Age of Blood Transfused in Septic ICU Patients. *Can. J. Anaesth.* **1997**, *44*, 1256–1261.
 16. Fransen, E.; Maessen, J.; Dentener, M.; Senden, N.; Buurman, W. Impact of Blood Transfusions on Inflammatory Mediator Release in Patients Undergoing Cardiac Surgery. *Chest* **1999**, 1233–1239.
 17. Moore, F.A.; Moore, E.; Sauaia, A. Blood Transfusion. An Independent Risk Factor for Post-injury Multiple Organ Failure. *Arch. Surg.* **1997**, *132*, 620–625.
 18. Blumberg, N.; Heal, J.M. Immunomodulation by Blood Transfusion: An Evolving Scientific and Clinical Challenge. *Am. J. Med.* **1996**, *101*, 299–308.
 19. Vamvakas, E.C. Transfusion-Associated Cancer Recurrence and Postoperative Infection: Meta-Analysis of Randomized, Controlled Clinical Trials. *Transfusion*. **1996**, *36*, 175–186.
 20. Edna, T.; Bjerkeset, T. Perioperative Blood Transfusions Reduce Long-Term Survival Following Surgery for Colorectal Cancer. *Dis. Colon Rectum* **1998**, *41*, 451–459.
 21. Heiss, M.M.; Mempel, W.; Delanoff, C.; Jauch, K.-W.; Gabka, C.; Mempel, M.; Dieterich, H.-J.; Eissner, H.-J.; Schildberg, F.-W. Blood Transfusion-Modulated Tumor Recurrence: First Results of a Randomized Study of Autologous Vs Allogeneic Blood Transfusion in Colorectal Cancer Surgery. *J. Clin. Oncol.* **1994**, *12*, 1859–1867.
 22. Fakhry, S.M.; Sheldon, G.F. Blood Administration, Risks, and Substitutes. *Adv. Surg.* **1995**, *28*, 71–92.
 23. Fitzgerald, R.D.; Martin, C.M.; Dietz, G.E.; Doig, G.S.; Potter, R.F.; Sibbald, W.J. Transfusing Red Blood Cells Stored in Citrate Phosphate Dextrose Adenine-1 for 28 Days Fails to Improve Tissue Oxygenation in Rats. *Crit. Care Med.* **1997**, *25*, 726–732.
 24. Vamvakas, E.C. Epidemiology of Red Blood Cell Utilization. *Transf. Med. Rev.* **1996**, *10*, 44–61.
 25. Wilkinson, S. *Proceeding of the Advisory Committee on Blood Safety and Availability*; US Public Health Service; 1999; 58–64.
 26. Zuck, T.F.; Riess, J.G. Current Status of Injectable Oxygen Carriers. *Crit. Rev. Clin. Lab. Sci.* **1994**, *31*, 295–324.
 27. Fratanoni, J.C. Red Cell Substitutes: Evolution of Approaches for Demonstrating Efficacy. *Blood Substitutes—Present and Future Perspectives*; Tsuchida, E., Ed.; Elsevier: Amsterdam; 1998; 33–39.
 28. Riess, J.G.; Le Blanc, M. Solubility and Transport Phenomena in Perfluorochemicals Relevant to Blood Substitution and Other Biomedical Applications. *Pure Appl. Chem.* **1982**, *54*, 2383–2406.
 29. Riess, J.G. Reassessment of Criteria for the Selection of Perfluorochemicals for Second-Generation Blood Substitutes: Analysis of Structure/Property Relationships. *Artif. Organs* **1984**, *8*, 44–56.
 30. Riess, J.G.; Le Blanc, M. Preparation of Perfluorochemical Emulsions for Biomedical Use: Principles, Materials and Methods. *Blood Substitutes: Preparation, Physiology, and Medical Applications*; Lowe, K.C., Ed.; Ellis Horwood Ltd.: Chichester, 1988; 94–129.
 31. Krafft, M.P.; Riess, J.G.; Weers, J.G. The Design and Engineering of Oxygen-Delivering Fluorocarbon Emulsions. *Submicronic Emulsions in Drug Targeting and Delivery*; Benita, S., Ed.; Harwood Academic Publ. Amsterdam, 1998; 235–333.
 32. Chang, T.M.S.; Riess, J.G.; Winslow, R.M., Eds.; Blood Substitutes: General. Proc. Vth Int. Symp. On Blood Substitutes, Vol 1. *Artif. Cells Blood Subst. Immobil. Biotechnol.* 1994; 22, 123–360.
 33. Riess, J.G., Ed. Blood Substitutes, the Fluorocarbon Approach Proc. Vth Intl. Symp. On Blood Substitutes, Vol 3. *Artif. Cells Blood Subst. Immobil. Biotech.* 1994; 22, 945–1543.
 34. Riess, J.G. Fluorocarbon Emulsions—Designing An Efficient Shuttle Service for the Respiratory Gases—The So-Called “Blood Substitutes”. *Fluorine at the Millenium*; Banks, R.E., Ed.; Elsevier: Amsterdam, 2000; 385–431.
 35. Lutz, J.; Decke, B.; Bauml, M.; Schultze, H.G. High Oxygen Extraction Combined With Extensive Oxygen Consumption in the Rat Liver Perfused With Fluosol-DA, a New Perfluoro Compound Emulsion. *Pflügers Arch. (Eur. J. Physiol.)* **1978**, *376*, 1–6.
 36. Faithfull, N.S. Mechanisms and Efficacy of Fluorochemical Oxygen Transport and Delivery. *Artif. Cells Blood Subst. Immobil. Biotech.* **1994**, *22*, 181–197.
 37. Stehling, L.; Zauder, H.L. Acute Normovolemic Hemodilution. *Transfusion* **1991**, *31*, 857–868.
 38. Kreimeier, U.; Messmer, K. Hemodilution in Surgery: State of the Art 1996. *World J. Surg.* **1996**, *20*, 1208–1217.
 39. Habler, O.P.; Kleen, M.S.; Hutter, J.W.; Podtschaske, A.H.; Tiede, M.; Kemming, G.I.; Welte, M.V.; Corso, C.O.; Batra, S.; Keipert, P.E.; Faithfull, N.S.; Messmer, K.F.W. Hemodilution and Intravenous Perflubron Emulsion As An Alternative to Blood Transfusion: Effects on Tissue Oxygenation During Profound Hemodilution in Anesthetized Dogs. *Transfusion* **1998**, *38*, 145–155.
 40. Spahn, D.R. Current Status of Artificial Oxygen Carriers. *Adv. Drug Deliv. Rev.* **2000**, *40*, 143–151.
 41. Poli de Figueiredo, L.F.; Mathru, M.; Solanki, D.; W., M.V.; Hess, J.; Kramer, G.C. Pulmonary Hypertension and Systemic Vasoconstriction May Offset the Benefits of Acellular Hemoglobin Blood Substitutes. *J. Trauma, Inj. Infect. Crit. Care.* **1997**, *42*, 847–856.
 42. Kasper, S.; Grune, F.; Walter, M.; Amr, N.; Erasmi, H.; Buzello, W. The Effects of Increased Doses of Bovine Hemoglobin on Hemodynamics and Oxygen Transport in Patients Undergoing Preoperative Hemodilution for Elective Abdominal Aortic Surgery. *Anesth. Analg.* **1998**, *87*, 284–291.
 43. Ivanitskii, G.R.; Vorob'yev, S.I. Organization of Motile Structures in Blood Flow—The Basis of the Functioning of Perfluorocarbon “Artificial Blood”. *Biophysics* **1996**, *41*, 189–201.
 44. Padnick, L.; Linsenmeier, R.; Goldstick, T.; McRipley, M.; Ahmed, J. Perflubron Emulsion Infusion Improves

- Tissue Oxygenation in the Cat Brain Under Hyperoxic Conditions. *Artif. Cells Blood Subst. Immob. Biotech.* **1996**, *24*, 398.
45. Flaim, S.F. Efficacy of Perflubron-Based Emulsion As a Temporary Oxygen Carrier. *Red Blood Cell Substitutes*; Rudolph, A.S., Rabinovici, R., Feuerstein, G.Z., Eds.; Marcel Dekker, Inc.: New York, 1998; 79–117.
 46. Faithfull, N.S.; Rhoades, G.E.; Keipert, P.E.; Ringle, A.S.; Trouwborst, A. A Program to Calculate Mixed Venous Oxygen Tension—A Guide to Transfusion. *Adv. Exp. Med. Biol.* **1994**, *361*, 41–48.
 47. Faithfull, N.S. The Concept of Hemoglobin Equivalency of Perfluorochemical Emulsions. *Adv. Exp. Med. Biol.*, **in press** (2001).
 48. Slovis, H.A.; Kamimoto, T. Erythrocyte Substitute for Perfusion of Brain. *Nature* **1967**, *216*, 458–460.
 49. Geyer, R.P.; Taylor, K.; Duffett, E.B.; Eccles, R. Successful Complete Replacement of the Blood of Living Rats With Artificial Substitutes. *Fed. Proc.* **1973**, *32*, 927.
 50. Clark, L.C.; Becattini, F.; Kaplan, S.; Obrock, V.; Cohen, D.; Becker, C. Perfluorocarbons Having a Short Dwell Time in the Liver. *Science* **1973**, *181*, 680–682.
 51. Okamoto, H.; Yamanouchi, K.; Imagawa, T.; Murashima, R.; Yokoyama, K.; Watanabe, R.; Naito, R. Persistence of Fluorocarbons in Circulating Blood and Organs. *Proc. 11th Intercompany Conf*; Osaka, 1973.
 52. Naito, R.; Yokoyama, K. Perfluorochemical Blood Substitutes. *FC-43 Emulsion, Fluosol-DA, 20% and 35%*; Green Cross Corp. Osaka, Japan, 1978, 1981.
 53. Alpha Therapeutic Corporation. *FLUOSOL 20%*; Product Monograph Alpha Therapeutic Corp.: Los Angeles, 1990.
 54. Ivanitsky, G.R.; Vorobyev, S.I. *Perftoran Blood Substitute with Gas-Transporting Function*; Product Monograph Perftoran Co.: Pushchino, Russia, 1997.
 55. Tsuda, Y.; Yamanouchi, K.; Okamoto, H.; Yokoyama, K.; Heldebrandt, C. Intravascular Behavior of a Perfluorochemical Emulsion. *J. Pharmacobio-Dyn.* **1990**, *13*, 165–171.
 56. Flaim, S.F. Pharmacokinetics and Side Effects of Perfluorocarbon-Based Blood Substitutes. *Artif. Cells, Blood Subst., Immob. Biotech.* **1994**, *22*, 1043–1054.
 57. Keipert, P.E.; Otto, S.; Flaim, S.F.; Weers, J.G.; Schutt, E.A.; Pelura, T.J.; Klein, D.H.; Yaksh, T.L. Influence of Perflubron Emulsion Particle Size on Blood Half-Life and Febrile Response in Rats. *Artif. Cells Blood Subst. Immob. Biotech.* **1994**, *22*, 1169–1174.
 58. Clark, L.C.; Hoffmann, R.E.; Davis, S.L. Response of the Rabbit Lung As a Criterion of Safety for Fluorocarbon Breathing and Blood Substitutes. *Biomater. Artif. Cells Immob. Biotech.* **1992**, *20*, 1085–1099.
 59. Long, D.M.; Higgins, C.B.; Mattrey, R.F.; Mitten, R.M.; Multer, F.K. Is there a Time and Place for Radiopaque Fluorocarbons. *Preparation, Properties, and Industrial Applications of Organofluorine Compounds*; Banks, R., Ed.; Ellis Horwood: New York, 1982; 139–156.
 60. Riess, J.G. Post-Fluosol Progress in Fluorocarbon Emulsions for in Vivo Oxygen Delivery. *La Trasf. del Sangue* **1987**, *32*, 316–334.
 61. Kabalnov, A.S.; Makarov, K.N.; Shchukin, E.D. Stability of Perfluoroalkyl Halide Emulsions. *Colloids Surf.* **1992**, *62*, 101–104.
 62. Weers, J.G.; Liu, J.; Fields, T.; Resch, P.; Cavin, J.; Arlauskas, R.A. Room Temperature Stable Perfluorocarbon Emulsions With Acceptable Half-Lives in the Reticuloendothelial System. *Artif. Cells Blood Subst. Immob. Biotech.* **1994**, *22*, 1175–1182.
 63. Kaufman, R.J. Clinical Development of Perfluorocarbon-Based Emulsions As Red Cell Substitutes. *Blood Substitutes: Physiological Basis of Efficacy*; Winslow, R.M., Vandegriff, K.D., Intaglietta, M., Eds.; Birkhäuser: Boston, 1995; 53–75.
 64. Tremper, K.K.; Vercellotti, G.M.; Hammerschmidt, D.E. Hemodynamic Profile of Adverse Clinical Reactions to Fluosol-DA 20%. *Crit. Care Med.* **1984**, *12*, 428–431.
 65. Hammerschmidt, D.E.; Vercellotti, G.M. Limitation of Complement Activation by Perfluorocarbon Emulsions: Superiority of Lecithin-Emulsified Preparations. *Biomater. Artif. Cells Artif. Organs* **1988**, *16*, 431–438.
 66. Ingram, D.A.; Forman, M.B.; Murray, J.J. Activation of Complement by Fluosol Attributable to the Pluronic Detergent Micelle Structure. *J. Cardiovasc. Pharmacol.* **1993**, *22*, 456–461.
 67. Long, D.N.; Riess, J.G.; Follana, R.; Burgan, A.; Mattrey, R.F. Preparation and Application of Highly Concentrated Perfluorooctylbromide Fluorocarbon Emulsions. *Biomater. Artif. Cells Artif. Organs* **1988**, *16*, 441–442.
 68. Riess, J.G.; Dalfors, J.L.; Hanna, G.K.; Klein, D.H.; Krafft, M.P.; Pelura, T.J.; Schutt, E.G. Development of Highly Fluid, Concentrated and Stable Fluorocarbon Emulsions for Diagnosis and Therapy. *Biomater. Artif. Cells Artif. Organs* **1992**, *20*, 839–842.
 69. Davis, S.S.; Round, H.P.; Purewal, T.S. Ostwald Ripening and the Stability of Emulsion Systems: An Explanation for the Effect of An Added Third Component. *J. Colloid Interf. Sci.* **1981**, *80*, 508–511.
 70. Kabalnov, A.S.; Shchukin, E.D. Ostwald Ripening Theory: Applications to Fluorocarbon Emulsion Stability. *Adv. Colloid Interf. Sci.* **1992**, *38*, 69–97.
 71. Higuchi, W.I.; Misra, J. Physical Degradation of Emulsions via the Molecular Diffusion Route and The Possible Prevention Thereof. *J. Pharm. Sci.* **1962**, *51*, 459–466.
 72. Riess, J.G.; Krafft, M.P. Fluorinated Materials for in Vivo Oxygen Transport (Blood Substitutes), Diagnosis and Drug Delivery. *Biomaterials* **1998**, *19*, 1529–1539.
 73. Riess, J.G.; Cornélus, C.; Follana, R.; Krafft, M.P.; Mahé, A.M.; Postel, M.; Zarif, L. Novel Fluorocarbon-Based Injectable Oxygen-Carrying Formulations With Long-Term Room-Temperature Storage Stability. *Adv. Exp. Biol. Med.* **1994**, *345*, 227–234.
 74. Yamanouchi, K.; Tanaka, M.; Tsuda, Y.; Yokoyama, K.; Awazu, S.; Kobayashi, Y. Quantitative Structure—In Vivo Half-Life Relationships of Perfluorochemicals for Use As Oxygen Transporters. *Chem. Pharm. Bull.* **1985**, *33*, 1221–1231.
 75. Riess, J.G.; Keipert, P.E. Update on Perfluorocarbon-Based Oxygen Delivery Systems. *Blood Substitutes—Present and Future Perspectives*; Tsuchida, E., Ed.; Elsevier: Amsterdam, 1998; 91–101.
 76. Lutz, J. Effect of Perfluorochemicals on Host Defense, Especially on the Reticuloendothelial System. *Int. Anesth. Clin.* **1985**, *23*, 63–93.
 77. West, L.; McIntosh, N.; Gendler, S.; Seymour, C.; Wisdom, C. Effects of Intravenously Infused Fluosol-DA

- 20% In Rats. *Int. J. Radiat. Oncol. Biol. Phys.* **1986**, *12*, 1319–1323.
78. Tremper, K.K.; Friedman, A.E.; Levine, E.M.; Lapin, R.; Camarillo, D. The Preoperative Treatment of Severely Anemic Patients With a Perfluorochemical Oxygen-Transport Fluid, Fluosol-DA. *New Engl. J. Med.* **1982**, *307*, 277–283.
 79. Gould, S.A.; Rosen, A.L.; Sehgal, L.R.; Sehgal, H.L.; Langdale, L.A.; Krause, L.M.; Rice, C.L.; Chamberlain, W.H.; Moss, G.S. Fluosol-DA As a Red-Cell Substitute in Acute Anemia. *New Engl. J. Med.* **1986**, *314*, 1653–1656.
 80. Mitsuno, T.; Ohyanagi, H.; Yokoyama, K.; Suyama, T. Recent Studies on Perfluorochemical (PFC) Emulsion As An Oxygen Carrier in Japan. *Biomat. Artif. Cells Artif. Organs* **1988**, *16*, 365–373.
 81. Spence, R.K.; McCoy, S.; Costabile, J.; Norcross, E.D.; Pello, M.J.; Alexander, J.B.; Wisdom, C.; Camishion, R.C. Fluosol DA-20 in the Treatment of Severe Anemia: Randomized, Controlled Study of 46 Patients. *Crit. Care Med.* **1990**, *18*, 1127–1230.
 82. Kent, K.M.; Cleman, M.W.; Cowley, M.J.; Forman, M.; Jaffe, C.C.; Kaplan, M.; King, S.B.; Krucoff, M.; Lassar, T.; McAuley, B.; Smith, R.; Wisdom, C.; Wohlgelemler, D. Reduction of Myocardial Ischemia During Percutaneous Transluminal Coronary Angioplasty With Oxygenated Fluosol. *Am. J. Cardiol.* **1990**, *66*, 279–284.
 83. Forman, M.B.; Perry, J.M.; Wilson, B.H.; Verani, M.S.; Kaplan, P.R.; Shawl, F.A.; Friesinger, G.C. Demonstration of Myocardial Reperfusion Injury in Humans: Results of a Pilot Study Utilizing Acute Coronary Angioplasty With Perfluorochemical in Anterior Myocardial Infarction. *J. Am. Coll. Cardiol.* **1991**, *18*, 911–918.
 84. Keipert, P.E.; Conlan, M.G. Advances in Perflubron Emulsion Development: Potential Use During Surgery and Cardiopulmonary Bypass to Avoid Donor Blood Transfusion and Prevent Tissue Hypoxia. *Artif. Cells Blood Subst. Immob. Biotech.* **1996**, *24*, 359.
 85. Wahr, J.A.; Trouwborst, A.; Spence, R.K.; Henny, C.P.; Cernaianu, A.C.; Graziano, G.P.; Tremper, K.K.; Flaim, K.E.; Keipert, P.E.; Faithfull, N.S.; Clymer, J.J. A Pilot Study of the Effects of a Perflubron Emulsion, AF0104, on Mixed Venous Oxygen Tension in Anesthetized Surgical Patients. *Anesth. Analg.* **1996**, *82*, 103–107.
 86. Goodin, T.H.; Grossbard, E.B.; Kaufman, R.J.; Richard, T.J.; Kolata, R.J.; Allen, J.S.; Layton, T.E. A Perfluorochemical Emulsion for Prehospital Resuscitation of Experimental Hemorrhagic Shock: A Prospective, Randomized, Controlled Study. *Crit. Care Med.* **1994**, *22*, 680–689.
 87. Burgan, A.R.; Herrick, W.C.; Long, D.M.; Long, D.C. Acute and Subacute Toxicity of 100% PFOB Emulsion. *Biomat. Artif. Cells Artif. Organs* **1988**, *16*, 681–682.
 88. Rosoff, J.D.; Soltow, L.O.; Vocolka, C.R.; Schmer, G.; Chandler, W.L.; Cochran, R.P.; Kunzelman, K.S.; Spiess, B.D. A Second-Generation Blood Substitute (Perfluorodichlorooctane Emulsion) Does Not Activate Complement During Ex Vivo Circulation Model of Bypass. *J. Cardiothorac. Vasc. Anesth.* **1998**, *12*, 397–401.
 89. Leese, P.T.; Noveck, R.J.; Shorr, J.S.; Woods, C.M.; Flaim, K.E.; Keipert, P.E. Randomized Safety Studies of Intravenous Perflubron Emulsion I. Effects on Coagulation Function in Healthy Volunteers. *Anesth. Analg.* **2000**, *91*, 804–811.
 90. Noveck, R.J.; Shannon, E.J.; Leese, P.T.; Shorr, J.S.; Flaim, K.E.; Keipert, P.E.; Woods, C.M. Randomized Safety Studies of Intravenous Perflubron Emulsion II. Effects on Immune Function in Healthy Volunteers. *Anesth. Analg.* **2000**, *91*, 812–822.
 91. Spahn, D.R.; van Bremp, R.; Theilmeier, G.; Reibold, J.-P.; Welte, M.; Heinzerling, H.; Birck, K.M.; Keipert, P.E.; Messmer, K. Perflubron Emulsion Delays Blood Transfusions in Orthopedic Surgery. *Anesthesiology* **1999**, *91*, 1195–1208.
 92. Schutt, E.; Barber, P.; Fields, T.; Flaim, S.; Horodniak, J.; Keipert, P.; Kinner, R.; Kornbrust, L.; Leakakos, T.; Pelura, T.; Weers, J.; Houmes, R.; Lachmann, B. Proposed Mechanism of Pulmonary Gas Trapping (PGT) Following Intravenous Perfluorocarbon Emulsion Administration. *Artif. Cells Blood Subst. Immob. Biotech.* **1994**, *22*, 1205–1214.
 93. Keipert, P.E.; Faithfull, N.S.; Roth, D.J.; Bradley, J.D.; Batra, S.; Jochelson, P.; Flaim, K.E. Supporting Tissue Oxygenation During Acute Surgical Bleeding Using a Perfluorochemical-Based Oxygen Carrier. *Adv. Exp. Med. Biol.* **1996**, *388*, 603–609.
 94. Habler, O.P.; Messmer, K.F. Tissue Perfusion and Oxygenation With Blood Substitutes. *Adv. Drug Deliv. Rev.* **2000**, *40*, 171–184.
 95. Holman, W.L.; Spruell, R.D.; Ferguson, E.R.; Clymer, J.J.; Vicente, W.V.A.; Murrah, C.P.; Pacifico, A.D. Tissue Oxygenation With Graded Dissolved Oxygen Delivery During Cardiopulmonary Bypass. *J. Thorac. Cardiovasc. Surg.* **1995**, *110*, 774–785.
 96. Cochran, R.P.; Kunzelman, K.S.; Vocolka, C.R.; Aki-moto, H.; Thomas, R.; Soltow, L.O.; Spiess, B.D. Perfluorocarbon Emulsion in the Cardiopulmonary Bypass Prime Reduces Neurologic Injury. *Ann. Thorac. Surg.* **1997**, *63*, 1326–1332.
 97. Flaim, S.F. Medical and Therapeutic Applications of Perfluorocarbon-Based Products. *Red Blood Cell Substitutes*; Rudolph, A.S., Rabinovici, R., Feuerstein, G.Z., Eds.; Marcel Dekker, Inc.: New York, 1998; 437–464.
 98. Krafft, M.P.; Riess, J.G. Highly Fluorinated Amphiphiles and Colloidal Systems, and Their Applications in the Biomedical Field—A Contribution. *Biochimie* **1998**, *80*, 489–514.
 99. Lowe, K.C.; Davey, M.R.; Power, J.B. Perfluorochemicals: Their Applications and Benefits to Cell Culture. *TIB Tech.* **1998**, *16*, 272–277.
 100. Riess, J.R. Blood Substitutes and Other Potential Biomedical Applications of Fluorinated Colloids. *J. Fluorine Chem.* **2002**, *112*, in press.

BLOOD SUBSTITUTES: HEMOGLOBIN-BASED OXYGEN CARRIERS

Deanna J. Nelson

BioLink Technologies, Inc., Cary, North Carolina, U.S.A.

INTRODUCTION

Successful allogeneic blood transfusion, one of the most important life-saving procedures of modern medicine, was first reported in 1818 (1). Its use grew steadily until the 1990s, when it peaked and reached a plateau. Today, more than 12 million units of packed red blood cells (RBCs) in the United States and at least 20 million red cell units worldwide are transfused into patients annually (2, 3). The chief indication for RBC transfusion is anemia, and the therapeutic goal is to increase oxygen delivery (4).

As transfusions began to occur in greater numbers, risks associated with this therapy began to be recognized. The hemolytic reactions that were a consequence of many of the early transfusions have largely been eliminated by blood typing and cross-matching. In contrast, the possibility of disease transmission by blood-borne pathogens, first recognized in the early 1940s, has not declined (5). Today, donated blood is tested for the presence of a broad spectrum of infectious agents, including hepatitis B and C, human immunodeficiency virus (HIV) 1 and 2, and human T-cell lymphotropic virus (HTLV) 1 and 2. In addition, blood may be contaminated with agents that are not monitored, including herpes viruses, such as cytomegalovirus or Epstein-Barr virus, parvoviruses, bacteria or parasites, such as *Yersinia* or *Trypanosoma cruzi*, respectively, or poorly characterized pathogens, such as those responsible for Creutzfeld-Jakob disease.

Other undesirable shortcomings of stored RBCs include an increase in the oxygen affinity of the intracellular hemoglobin due to losses in intracellular effectors (primarily 2,3-diphosphoglycerate) (6). Adverse changes in RBC shape and deformability, associated with the loss of cellular adenosine triphosphate (ATP) during storage, further compromise the RBC capability to deliver oxygen, because poorly deformable RBCs can become entrapped and impede local blood flow in the microcirculation (7, 8). In addition, immunogenic reactions, particularly those that occur after multiple transfusions or the transfusion of multiple units of allogeneic RBCs, are of continuing concern (9). Finally, several studies have shown depressed immune function post-transfusion of RBCs; i.e., patients with different tumors had a shorter disease-free survival and a higher incidence of recurrent or metastatic cancers or

a higher rate of postoperative infections after homologous blood transfusions (10–12).

A number of forces have combined to stimulate the development of fluids, so-called blood substitutes, that could be used in place of allogeneic blood. For example, the pressing need for the ready availability of a completely compatible “blood” has been recognized for many years (13) and continues to be important today. The observation that there are ischemic states where blood volume is not depleted but a requirement for the restoration of oxygen delivery is present, has focused attention on oxygen-carrying fluids that might be infused where RBC transfusion clearly is not indicated. All of these factors, together with a perceived commercial potential, have combined to promote the development of blood substitutes.

The term “blood substitute” is, however, a misnomer. Blood has dynamic, metabolic, regulatory, coagulation, and immunologic functions not mimicked by the plasma volume expanders or electrolyte solutions that are used clinically for volume replacement following blood loss. Nor are these properties embodied in the two classes of oxygen-carrying fluids, hemoglobins and perfluorocarbons, that are being developed as proposed alternatives to one component of blood, the red cell. As a result, it is widely recognized that the therapeutic use of blood and its various components will continue to expand; consequently, voluntary donations and advances in transfusion therapy must be continued in order to meet clinical needs.

This review will focus on hemoglobin therapeutics, that is, those hemoglobin-based oxygen carriers under development to complement the oxygen-carrying properties of blood. Given this focus on the requirement that the fluid transport and deliver oxygen, the review neglects consideration and discussion of colloidal plasma expanders (e.g., albumin, dextran, gelatin, and hydroxyethyl starch) and electrolyte solutions [Ringer’s lactate (RL) solution, for example]. Selected properties of these resuscitation fluids are compared in Table 1.

HISTORICAL OVERVIEW

Over a century ago, hemoglobin (Hb), the protein in RBCs, was discovered to be the means of oxygen transport

Table 1 A summary comparison of the properties of blood, blood substitutes, and volume expanders

Property	Blood	Blood substitutes	Volume expanders
Volume expansion	Yes	Yes	Yes
Oxygen carrying capacity	Yes	Yes	No
Other therapeutic proteins	Yes	No	No
Therapeutic life	1–2 months	1–2 days	Hours (varies with dose and species)
Storage life	1 month	6–24 months	2 years
Changes during storage	Yes	No	No
Type specific	Yes	No	No
Viral inactivation	No	Yes	Yes
Size	Large	Small	Small
Viscosity	High	Low/moderate	Low

by blood (14). Thus, it is not surprising that preparations of this protein have been repeatedly evaluated as the active principal of temporary blood replacement solutions (15).

Human Hb, a typical mammalian Hb, is a protein made up of four polypeptide subunits—two α -chains and two β -chains. Each subunit contains a heme prosthetic group, consisting of a porphyrin ring and an iron ligand. In the form of Hb capable of reversibly binding oxygen, the iron is in its +2 oxidation state, the ferrous form, and is sequestered by the porphyrin ring within the protein. If the iron is oxidized to its ferric or +3 form, Hb is oxidized to methemoglobin (methHb), which is incapable of binding oxygen. The molecular weight of human Hb is about 64,500 Daltons, similar to that of bovine Hb, another widely studied mammalian Hb (10–12).

Whereas anecdotal results obtained with historically prepared Hb solutions were somewhat encouraging, it became apparent that purified Hb is rapidly excreted through the kidneys after intravenous infusion and does not effectively release oxygen to the tissues (16). The former results when, once outside the red cell, normal tetrameric Hb dissociates into dimeric subunits readily filtered through the kidney glomerulus. The inability to efficiently release oxygen to the tissues stems from the fact that purified, acellular human Hb assumes a conformation that binds oxygen with a higher affinity than Hb in RBCs. Human RBCs contain high concentrations of 2,3-diphosphoglycerate (2,3-DPG), an organic polyanion that binds to Hb in a manner such that the affinity of oxygen binding is decreased. Unfortunately, 2,3-DPG readily hydrolyzes during the storage of blood for transfusion purposes or is lost during Hb purification.

Analysis of these historical observations on the properties of purified native human Hb has led to the conclusion that Hb must be modified in order to be clinically useful. Generally, Hb modification achieves two goals: first, the oxygen affinity of the Hb is decreased, and

second, its intravascular retention is prolonged. Beyond these general objectives, however, there is little agreement. Historically, a plasma expander is expected to be highly hydrophilic, nontoxic, and nonantigenic, have a colloid osmotic pressure close to that of blood, have a viscosity compatible with physiological conditions, and be biodegradable or excretable, but with a sufficient half-life (17). In addition, the Hb should not cause a foreign-body reaction, should be immunologically inert, should not simulate the coagulation cascade, should not stimulate the complement system, and should neither stimulate nor depress other defense systems inside the body (18). However, these properties are insufficient to define the optimal oxygen affinity or molecular weight profile of a therapeutically suitable Hb. Other criteria, vague at the present time, include the appropriate balance between oxygen-carrying capacity and colloid osmotic activity of the Hb and the corresponding balance between the intravascular retention as a functional Hb and optimal metabolism of the oxidized Hb. Nor is there agreement on the maximum tolerable methHb content of a therapeutic Hb, although it is recognized that methHb levels greater than 10% may significantly reduce tissue oxygenation (19).

Historical results obtained with Hb solutions were also confounded by the fact that many of these preparations were impure. One common impurity was fragments of the red cell plasma membrane, denoted as “stroma.” This stromal detritus was subsequently shown to be responsible, at least in part, for the kidney toxicity frequently observed after Hb infusion (20). For this reason, stroma must be rigorously excluded from Hb preparations. In addition to removing a source of nephrotoxicity, rigorous stroma removal eliminates the blood type specific antigens, thereby enabling these purified Hb formulations to be administered to any patient without the necessity of blood typing or cross-matching.

Endotoxin contamination has been encountered frequently in Hb solutions. Historically, the production of Hb solutions low in endotoxin content has proven challenging, in part because much of the early work was performed in academic research laboratories that were not familiar with the techniques used in the pharmaceutical industry to prevent bacterial contamination of the process stream. In addition, Hb has been reported to bind endotoxin, making it even more difficult to completely remove once present (21). Therefore, the production of Hb solutions must be carefully designed to minimize the introduction of endotoxin into the process stream.

THERAPEUTIC HBS UNDER DEVELOPMENT

Hb Sources

Hb suitable for modification may be obtained from any one of a variety of natural and biotechnological sources. In terms of quantities available, mammalian red cells are the most useful source at the present time. Over 90% of the protein in a typical mammalian red cell is Hb, and the protein is readily isolated by lysis of the red cell membrane. Red cell-sourced Hbs in commercial development are isolated from either human or bovine red cells.

If the Hb is isolated from mammalian red cells, both freshly collected and stored red cells may be used as sources. The red cell membrane is the site of the primary deterioration that occurs in RBCs during storage for transfusion, whereas the quality of Hb contained within cells remains unchanged (22). As a consequence, the quality of Hb obtained from expired RBC units is high, and both freshly collected and expired but otherwise transfusable RBC units are suitable as Hb sources.

Genetic engineering has enabled production of Hb in microorganisms, transgenic animals, and plants (23). A mutant, cross-linked human Hb has been expressed in *E. coli* at a large scale (24). An Hb expression system in yeast has also been described (25). Human Hb has been expressed at high levels in both mice and swine (26). Clearly, the latter animal is one from which suitably high quantities of Hb could be obtained for commercial development.

Classes of Modified Hbs

At present, four classes of modified Hbs are under development: 1) intramolecularly cross-linked Hb; 2) conjugated Hb; 3) polymerized Hb; and 4) Hbs produced by a combination of these processes (Fig. 1).

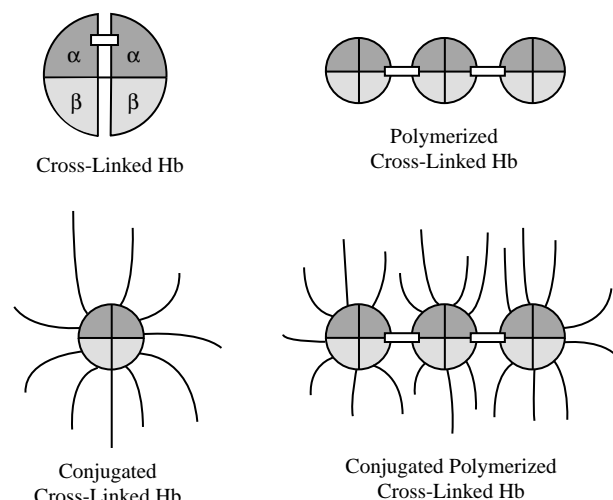


Fig. 1 Types of modified Hb.

Hb may be intramolecularly cross-linked between the α - or the β -subunits of the protein by chemical or genetic engineering methods. For example, if a polyanionic cross-linking agent such as bis(3,5-dibromosalicyl) fumarate (DBBF) or 2-norformyl- pyridoxal 5'-phosphate is used, Hb is selectively cross-linked in the so-called DPG-binding site between the beta subunits. If this site is blocked by an unreactive polyanion, Hb may be selectively cross-linked between the epsilon-amino groups of the lysine-99 residues on the alpha subunits, as it is in Diaspirin Cross-Linked Hemoglobin (DCLHb; Baxter). Alternatively, by insertion of the appropriate DNA constructs, the expressed Hb is intramolecularly cross-linked between the amino- and carboxyl termini of the alpha subunits. Intramolecularly cross-linked Hbs have a molecular weight and structure nearly identical to the native Hb from which they are derived.

Hemoglobin may be conjugated with a wide variety of polymers, including monomethyl polyethylene glycol (MPEG),^a dextran, and inulin. Conjugation is effected specifically only if the polymer is monofunctionally activated, i.e., if there is only a single site for reaction with Hb. However, multiple polymer chains may be attached to a single Hb molecule. Attachment typically occurs to the ϵ -amino groups of lysine residues located at the surface of the protein, although other amino acids (histidine, arginine, cysteine, or tyrosine, for example) may be modified under specific reaction conditions or through the use of specific reagents. Because the reaction is typically

^aAlthough MPEG is generally viewed as a monofunctional reagent, most commercial MPEGs are contaminated with bifunctional polyethylene glycol (PEG). Therefore, both conjugated and polymerized Hbs are obtained from modification reactions using these reagents.

Table 2 Products in commercial development and their potential clinical applications

Company	Modified Hb	Clinical study focus
Apex Biosciences Inc. (USA)	PEG-conjugated human Hb (Hb-PHP)	Septic shock
Baxter Healthcare Corp. (USA)	Modified rHb	No current trials ^a
Biopure Corp. (USA)	Glutaraldehyde-polymerized bovine Hb (HBOC-201 or Hemopure™)	Surgery
Enzon Corp. (USA)	PEG-conjugated bovine Hb (PEG-Hb)	Adjunct to radiation therapy
Hemosol (Canada)	Oxidized raffinose-polymerized and cross-linked human Hb (o-Raff-Hb or Hemolink)	Orthopedic or cardiac surgery; use as an adjunct to erythropoietin in treatment of chronic renal failure
Northfield Laboratories (USA)	Pyridoxalated, glutaraldehyde-polymerized human Hb (PolySFH-P or PolyHeme)	Trauma, surgery

^aDevelopment of DCLHb, an intramolecularly cross-linked human Hb, was discontinued in 1998.

one of surface modification, the modified Hbs are sometimes described as “decorated” Hbs. The conjugated Hbs obtained by this type of modification typically have a range of molecular weights defined by the molecular weight of the conjugating reagent and the number of molecules of that reagent conjugated to the protein. In addition, if the conjugating reagent is contaminated with precursors having more than one reactive site (as may be true of MPEG, for example), polymerized Hbs may constitute a part of the final product.

Likewise, Hb may be polymerized by a wide variety of reagents; the most widely used is glutaraldehyde. Other polymerization agents include oxidized adenosine, oxidized raffinose, and bifunctional, activated PEGs. A typical polymerization agent is bifunctional, offering two sites for reaction. Attachment occurs to the ϵ -amino groups of lysine residues located at the surface of the protein. The modified Hbs obtained by polymerization are polymers with about 2–10 Hbs linked linearly. Although branched polymers may also be obtained, the use of short polymerization agents significantly favors linear polymerization of Hb. In addition, during the reaction, Hb may be surface-modified (“decorated”) by the polymerization agent. The polymerized Hb products are mixtures of decorated polymers, ranging in molecular weight from the mass of the native Hb to 10-fold multiples of that molecular weight. (Higher molecular weight entities tend to be unstable and precipitate from solution.)

Finally, modified Hbs may be produced by a combination of these processes. For example, human Hb

may be cross-linked and then conjugated to another macromolecule, or the Hb may be cross-linked and then polymerized. (Although the processes may be reversed, cross-linking first ensures that tetrameric Hb will be further modified.)

The information in Fig. 1 and Table 2 illustrates the various approaches used by manufacturers of therapeutic Hbs now in clinical trials.

Formulation

Modified Hbs are most frequently formulated in electrolyte solutions iso-osmotic with human plasma. The Hb concentration typically is in the range of about 5–15%. Among the vehicles used in the formulation of the current generation of Hb therapeutics are phosphate-buffered saline, Ringer's acetate, and RL solution. Formulation in bicarbonate buffer has also been reported (27).

The effects of other additives, particularly anti-oxidants, have been investigated in an effort to extend the storage stability of the therapeutic Hb (28). However, not all reducing agents can reduce Hb, and some may even act as Hb oxidants. For example, ascorbate and glutathione, two of the more widely employed pharmaceutical anti-oxidants, are effective anti-oxidants only for deoxyhemoglobin solutions and are oxidants when Hb is oxygenated. Moreover, some anti-oxidants or reducing agents are unsuitable for therapeutic use or may be effective only within a pH range that is not tolerated well

physiologically, or are toxic in the concentrations required to show effectiveness as an anti-oxidant.

Encapsulation in liposomes or polylactide nanoparticles has been studied as a means for increasing the Hb concentration of a formulation while maintaining a low colloid oncotic pressure, prolonging the lifetime in the circulation, and reducing direct Hb exposure to cells and tissues in the body. To date, Hb encapsulation has met with limited success. Data from early studies of liposome-encapsulated Hb (LEH) focused attention on the shortcomings of these preparations as follows: immune system suppression, complement activation, difficulties in controlling endotoxin contamination, and the relatively low (1–2 g/dl) Hb concentration typically captured in liposomes (29). More recent work has benefited from the progress in liposome encapsulation made in other areas (30). Although low Hb-encapsulation efficiency remains an issue, several groups have demonstrated that improved encapsulation technology (31), coupled with surface modification of the liposome with PEG, further increases the circulation persistence of the LEH, presumably by decreasing recognition and reticulo-endothelial system (RES) uptake (32–36). Moreover, the ability to co-encapsulate systems for preventing Hb oxidation or effecting metHb reduction offers the advantage of prolonging the circulation of a functional oxyHb (37). Finally, the recognition that PEG-modification also allows for greater drug delivery to tumors may redirect clinical applications of LEH to this area as well as to resuscitation (38).

Packaging and Storage

Gradual Hb oxidation to metHb is a stability-limiting parameter during long-term storage of modified Hb solutions. In contrast, both native and modified Hbs are remarkably stable when stored deoxygenated or completely ligated by carbon monoxide. Moreover, following exposure to room air, deoxyhemoglobin is rapidly oxygenated. Consequently, most firms developing therapeutic Hbs store their preparations deoxygenated. Alternatively, an oxygenated Hb solution may be stored frozen; storage at -80°C ensures long-term stability. Likewise, the protein may be lyophilized and reconstituted at the time of use.

Although glass containers are very suitable for Hb packaging, flexible plastic containers constructed from materials such as poly(ethylene-vinyl acetate) (EVA) are most widely used. The plastic packaging material must be carefully selected. Hemoglobin, like other proteins, is lipophilic and can facilitate extraction of low molecular

weight lipophilic substances, such as the plasticizers used in poly(vinyl chloride) containers.

Technical Issues Associated with Commercial Production

When mammalian red cell Hb is used as the raw material for production of a modified Hb, the requirement for the minimization of plasma proteins and red cell stroma is a rigorous and general one. Some manufacturers meet this requirement by extensive red cell washing to reduce contamination by residual plasma, controlled lysis, and careful filtration of the hemolysate, followed by ultrafiltration. Other manufacturers add a chromatographic purification step to this procedure. Published data indicate that the phospholipid content of the so-called “stroma-free” Hb resulting from either process is very low ($<2\text{ }\mu\text{g/ml}$) (39–41).

A second issue associated with the commercial production of modified Hb from mammalian red cells is the exclusion of endotoxin contamination from the process stream and the product. Once introduced, endotoxin contamination is difficult to eliminate and causes a pyrogen response or a more severe, adverse effect in humans. Several chromatographic methods have been proposed as means for endotoxin removal (e.g., DEAE-Sephacrose or commercial affinity resins such as DetoxigelTM or Acticlean EtoxTM), but demonstrations are lacking as to their effectiveness for endotoxin removal sufficient to completely eliminate LAL-reactivity in Hbs produced at commercial scales (42).

Another microbiological purity issue that has been associated with human blood is the possibility of adventitious contamination with viral pathogens. Likewise, if animal blood is used as the red cell source, a spectrum of viruses have been identified as potential contaminants and possibly as zoonotic agents that may be infectious to humans (43). Therefore, in addition to the careful control of the blood used as the Hb source, the process for Hb solution preparation should employ robust virus inactivation and removal steps to reduce the risk of disease transmission.

When genetic engineering is used to produce the therapeutic Hb, three challenges present potential limitations to utility. First, the production scale itself presents a significant challenge. A second challenge is related to the expression system and its current limitations; and third, bacterial endotoxins or other pyrogens must be removed.

Genetic engineering enabling expression in microbial systems has been used successfully to produce proteins having annual production requirements of a few tons or

less. However, annual production requirements for a therapeutic Hb will likely reach thousands of tons following commercialization. Thus, matching the production capabilities to commercial requirements is a major challenge.

Enhancing expression efficiency of the desired Hb is a second challenge to be addressed for therapeutic Hb production (44). At present, the typical biotechnological process incorporates microbiological expression of each of the requisite globin proteins and insertion of preformed heme into the protein subunits. (Hemin isolated from bovine Hb is added to the growth medium in order to provide adequate amounts of heme.) In many expression systems, recombinant α - and β -chains are produced with an extra methionine at their respective amino termini. Since the native *N*-terminal valine residues of the α - and β -chains are known to play important roles in regulating oxygen affinity, in interactions with allosteric effectors, and in the Bohr effect, the extra methionine or another mutation at these *N*-terminals can alter both the protein conformation and the functional properties of the recombinant Hb (rHb). Moreover, Ho has reported that as much as 25% of the purified rHb exhibits an incorrect conformation of the heme group in the β -chain (45). Whether enhanced endogenous heme production would eliminate this defect is not known. At present, however, microbiological heme synthesis may not be sufficiently efficient to permit its use in commercial applications.

Finally, if *E. coli* is used as the production system, the therapeutic Hb must be purified rigorously. Both elimination of all non-Hb proteins and elimination of endotoxins are stringent requirements, because the latter lipoproteins are recognized pyrogens. Extensive chromatographic purification is used to accomplish this.

Some key issues are common to the development of any practical modified Hb solution. For example, the propensity of this protein to oxidize must be addressed. In order to remain functional in the delivery of oxygen, the iron in the prosthetic heme group of Hb must be maintained in the +2 oxidation state. Therefore, processing and long-term storage conditions must minimize the formation of metHb. This is accomplished in a number of ways. Low processing temperatures are used as much as possible. In addition, some processing steps that expose Hb to particularly adverse conditions may involve conversion of (oxy)Hb to deoxyhemoglobin or carboxyhemoglobin, completion of the process step, and subsequent reconversion to (oxy)Hb.

Irrespective of the manner in which the therapeutic Hb is manufactured, there is a general requirement

for conformance with current Good Manufacturing Practices.

POTENTIAL THERAPEUTIC BENEFITS

Many historical expectations regarding the potential safety and efficacy of Hb-based oxygen carriers have been realized, and these proteins are being increasingly recognized as therapeutic hemoglobins (tHbs). For example, tHbs may be administered without blood typing or cross-matching and apparently do not activate the complement cascade in the recipient or elicit an immune response (46). (However, trials of tHbs produced from bovine Hb may require testing for existing antibodies in patients prior to Hb infusion.)

The properties of the current generation of tHb solutions (Tables 2 and 3) that are expected to contribute directly to clinical potency are the abilities to transport oxygen, redistribute blood flow, and support the maintenance of vascular volume. The ability to transport oxygen is likely to be beneficial as a consequence of three properties of a therapeutic Hb: 1) the favorable oxygen binding characteristics of the reduced heme groups; 2) the small size of an acellular Hb compared to that of a red cell; and 3) the low viscosity of some tHb solutions. Given these properties, a tHb is likely to perfuse microcapillary beds and areas where red cell perfusion has been restricted by partial blockage or severe constriction. Potentially, these properties will prove beneficial in the following clinical indications. The three primary and most widely recognized potential indications are discussed first, followed by other, more speculative uses of a tHb.

Reduction of Acute Global Oxygen Deficit

Other than blood itself, tHb solutions are the only colloidal plasma volume expanders that will transport sufficient oxygen to maintain life. In general, preclinical studies in a spectrum of animal models of hemorrhagic shock show tHb solutions restore mean arterial pressure (MAP), promote adequate tissue oxygenation, more effectively resolve base deficit and ischemic acidosis, and improve survival in the test animals receiving the Hb formulation. In contrast, test groups receiving electrolyte [e.g., Ringer's lactate (RL)] or colloid (e.g., hetastarch or albumin) solutions responded less positively (47–50). Consequently, the most widely investigated potential therapeutic applications of the current tHb formulations comprise studies in which a tHb solution is infused as a substitute for

Table 3 Acronyms used in this review

Therapeutic Hb	Manufacturer	Acronym
Pyridoxalated, PEG-conjugated Hb	Apex	Hb-PHP
Intramolecularly $\alpha\alpha$ -cross-linked human Hb	Baxter	DCLHb
Glutaraldehyde-polymerized bovine Hb	Biopure	HBOC-201
PEG-Modified bovine Hb	Enzon	PEG-Hb
Oxidized raffinose-cross-linked and polymerized human Hb	Hemosol	o-Raff-Hb
Pyridoxalated glutaraldehyde-polymerized human Hb	Northfield	PolySFH-P

transfusion of blood or packed red cells or is used as a hemodiluent.

Resuscitation from acute hypovolemia or hemorrhagic shock

Infusion of a tHb in resuscitation from acute hypovolemia or hemorrhagic shock, clinical indications where blood or packed red cells typically are transfused, constitutes one of three general categories of clinical indications recognized by the FDA (51). Clinical trials assessing the efficacy and safety of a polymerized human Hb, PolySFH-P (Poly-Heme; Northfield), in these indications are in progress. Data concerning their safety and efficacy in reducing acute global oxygen deficit are not disclosed until patient enrollment is completed and rigorous, multi-faceted analyses are performed. Thus, it is difficult to predict when summaries will be presented that confirm the expectations based on preclinical findings.

Preliminary reports suggest these tHbs may be of value in resuscitation following trauma or unanticipated hemorrhage, where the Hb solution could be infused as an immediate response to blood loss. Under these conditions, the tHb is a bridge to transfusion: data accumulated during transport, in the emergency room or recovery room, peri-operatively, or in the intensive care unit, enable broader assessment of the physiologic need for a subsequent blood transfusion.

For example, a recent report discusses the assessment of clinical utility of PolySFH-P as a resuscitation fluid following acute trauma or urgent surgery (i.e., aortic aneurysm repair, hepatic resection, portosystemic shunt, and hip replacement) (53). Forty-four trauma patients with an average Injury Severity Score of 21 ± 10 were randomized to receive red cells (control; $n = 23$) or up to 6 units (300 g) of PolySFH-P ($n = 21$) as their initial blood replacement. Investigators blinded to the treatment made the decision to transfuse based on clinical signs, symptoms, and laboratory data, and the clinical circumstances. (PolySFH-P was given pre-operatively

and intra-operatively to both awake and anesthetized patients but only during the actual period of hemorrhage.) Samples were obtained before infusion, at the end of infusion, at 12 h, and at days 1, 2, and 3. Vital signs (temperature, blood pressure, and heart rate), liver function [aspartate aminotransferase (AST), alanine aminotransferase (ALT), and bilirubin], amylase, and renal function (creatinine) were measured.

No important safety issues or significant adverse events related to PolySFH-P were observed, either during infusion or in the monitoring period that followed. It is noteworthy that a PolySFH-P infusion of 4.4 ± 2.0 units resulted in a plasma Hb of 3.9 ± 1.3 g/dl, which accounted for approximately 40% of the total circulating Hb. The plasma Hb concentration decreased to 2.2 ± 1.3 g/dl one-half day following infusion and to 1.4 ± 0.9 , 0.6 ± 0.6 , and 0.3 ± 0.3 g/dl on days 1, 2, and 3 postinfusion, respectively. No significant differences between groups were observed in temperature, MAP, heart rate (HR), and creatinine, documenting the absence of vasoactive properties, renal dysfunction, or fever because of PolySFH-P. Nor were differences in other measures of organ function observed, other than an increase in total bilirubin to concentrations in the range 2.4 ± 1.7 g/dl through the third day postinfusion. The total number of allogeneic red cell transfusions for the control group was significantly greater than that of the experimental group on days 1 and 3, post-infusion. Beyond 3 days post-infusion, all the PolySFH-P had left the circulation, and investigators transfused RBCs to address a further need for oxygen.

A subsequent Phase II study evaluated use of up to 10 units (500 g Hb) of PolySFH-P in treatment of acute blood loss (53). Early results (17 patients) suggested safety. Phase III trials in fifteen North American centers are assessing the therapeutic value of infusion of PolySFH-P in patients undergoing surgery for aortic aneurysm repair.

The efficacy of DCLHb, an intramolecularly cross-linked human Hb (Baxter) was assessed with respect to reducing or preventing the postoperative use of blood

transfusions (54). More than 1500 patients undergoing cardiac surgery were screened to enter this study. Of these, 209 were judged to be in need of transfusion after surgery, had met the study inclusion/exclusion criteria, and had signed informed consent. The patients were randomized to receive up to 3 units (a total of 750 ml) of 10% DCLHb or up to 3 units of packed RBCs within 24 h of surgery. Following this period, patients could receive additional transfusion of RBCs or other treatment as necessary. On the day of surgery, 59% of the patients treated with DCLHb did not require RBC transfusion. At day 7 after surgery, 19% of the DCLHb-treated patients had completely avoided a transfusion of RBCs. The mean number of units of RBCs transfused through 2 days postsurgery was significantly lower in DCLHb-treated patients than in controls ($p < 0.05$). After this period, there was no difference between the mean number of RBC units transfused in both groups. It was noted that DCLHb was well tolerated in this patient population and that there were no significant differences in morbidity or mortality between the groups.

In two other trials, one in the United States and the other in Europe, DCLHb was tested in trauma and hemorrhagic shock patients as an adjunct to the current therapies for enhancing oxygen delivery: fluids, blood, and operative intervention. In the multicenter, randomized, controlled, single-blinded efficacy trial conducted at 18 U.S. trauma centers from Feb 1997 to Jan 1998, patients with presumed or proven hemorrhage and persistent hypoperfusion were treated with DCLHb or normal saline (control) solution (55). Although there were no restrictions in the use of fluids, blood, or any other intervention prior to enrollment in this study, once enrolled, the patient received 500 ml of the treatment solution no later than 30 min after first meeting the entry criteria in the study hospital and within 60 min of hospital arrival. Critically ill patients who continued to meet entry criteria after treatment received up to an additional 500 ml of treatment solution during the 1-h infusion period. A key objective of the trial was a DCLHb-related reduction in the 28-day mortality in traumatic, hemorrhagic shock patients by 25%, i.e., from 40 to 30%. A reduction in 28-day morbidity or 48-h mortality and a normalization of 24-h lactate levels from base deficits of more than 15 mEq/l were corollary objectives. Of the 112 patients enrolled in the U.S. study before it was terminated in early 1998, 98 (88%) were infused with DCLHb or saline solution. At 28 days, 24 (46%) of the 52 patients infused with DCLHb died, and 8 (17%) of the 46 patients infused with the saline solution died ($p = 0.003$). At 48 h, 20 (38%) of the 52 patients infused with DCLHb died and 7 (15%) of the 46 patients infused with the saline solution died ($P = 0.01$).

The 28-day morbidity rate, as measured by the multiple organ dysfunction score, was 72% higher in the DCLHb group ($P = 0.03$). There was no difference in adverse event rates or the 24-h lactate levels. No subgroup analysis or covariate adjustment altered this mortality imbalance, nor did the 28-day multiple organ dysfunction score and the 24-h lactate clearance end points demonstrate a beneficial DCLHb effect. The study director noted that the absence of morbidity and perfusion marker improvements were due, in part, to the higher mortality rate in the DCLHb group, which limited the ability to measure morbidity independently using these two clinical markers. Moreover, it was noted that the study group believed it is not possible to conclude definitively that the mortality imbalance was due solely to a DCLHb-treatment effect. The study was terminated by the sponsor on March 17, 1998, based on the recommendation of the data monitoring committee. The concurrent second study in Europe was terminated a few months later.

Perioperative applications of therapeutic hemoglobins

Techniques for conserving the blood supply include not only substituting a tHb for blood itself, but also limiting the number of allogeneic blood transfusions in surgical patients through the use of autologous blood donations and intra-operative blood salvage. Late Phase II and Phase III trials are in progress, assessing the utility of two polymerized Hbs in these clinical applications.

No adverse effects were found during completion of a multicenter, Phase IIa trial in 18 orthopedic (hip replacement) surgical patients in which up to 500 ml of o-Raff-Hb (Hemosol), an oxidized raffinose-polymerized human Hb, were used for pre-operative hemodilution (56). A similar assessment is now being completed in cardiac surgical patients.

Four reports summarize clinical findings regarding similar hemodiluent applications of HBOC-201, a 13 g/dl solution of highly polymerized, purified bovine Hb (Biopure). The first of these evaluated the effects of low doses (ca. 0.4 g/kg) of HBOC-201 in 13 patients undergoing pre-operative hemodilution before elective abdominal aortic (repair) surgery (57) and the second, 12 surgical patients undergoing liver resection (58). After induction of anesthesia, autologous donation of a liter of blood, and hemodilution with either 1 or 2 L of RL solution, patients were randomly assigned to receive, within 30 min, a 3 ml/kg dose of either HBOC-201 or 6% hydroxyethyl starch (HES). Blood chemistry, co-oximetry, hematology, coagulation profiles, and immunological examinations were completed prior to surgery, on the day of surgery, on days 2, 3, 4, and 7 after the operation, on

the discharge day; 2, 3, 4 days postdischarge, and 3 months after operation. During surgery, monitored variables included invasive arterial and pulmonary artery pressures and arterial and mixed venous blood gases. The cardiac index (CI), systemic and pulmonary vascular resistance indices (SVRI and PVRI, respectively), oxygen delivery index (DO_2I), oxygen consumption index (VO_2I), and oxygen extraction ratio (O_2ER) were calculated using standard equations.

In general, the data indicated the low, single-dose administration of HBOC-201 was well-tolerated. However, in contrast to the patients receiving HES, patients who received HBOC-201 showed a progressive increase in MAP and SVRI and a significant decrease in CI, starting immediately after the beginning of the infusion; maximal changes were observed about 30 min after the infusion of the Hb solution ended. The investigators had anticipated clear signs of improved tissue oxygenation, as evidenced by an increase in arterial oxygen content and a maintenance or increase in DO_2I . However, the decreases in CI were sufficiently large that this calculated increase was not seen. The authors concluded that a 3 ml/kg dose of HBOC-201 was insufficient to augment the oxygen-carrying capacity of the blood.

Based on investigators' recommendations, the effects of increasing doses of HBOC-201 were assessed in two subsequent studies (59). After the induction of anesthesia, collection of 1 L of blood, and the infusion of 1 L of RL solution, 24 patients (12 in each study) were randomly assigned to receive, within 30 min, a 6.9- or 9.2-ml/kg dose of either HBOC-201 or 6% HES. (Patients infused with Hb received either 0.9 or 1.2 g Hb/kg, respectively.) Then, patients received an additional 500 ml of RL solution to maintain intravascular volume. As before, patient data relative to systemic and pulmonary arterial pressures and arterial and mixed venous blood gases were used to calculate hemodynamic and oxygen transport indices. The autologous blood withdrawn for preoperative hemodilution was reinfused intra-operatively after aortic unclamping.

In these two studies, as in the preceding ones, the infusion of HBOC-201 was associated with sizeable increases in SVRI (121 and 71%, respectively) and PVRI (70 and 53%, respectively). A concomitant influence on cerebral hemodynamics was not detected (60). Hemodilution with HBOC-201 increased the plasma Hb concentration to as high as 3 g/dl—Hb concentrations sufficient to maintain the arterial oxygen content at levels significantly higher than hemodilution with HES. However, the advantage of a greater oxygen-carrying capacity was offset by the increases in SVRI and PVR, and the resulting, sizeable decrease in CI. Although whole-body oxygen

consumption was maintained by increased oxygen extraction, in the absence of evidence of increased tissue oxygenation, the increase in oxygen consumption was viewed as a compensatory response to the decreased oxygen delivery, rather than a marker of enhanced unloading of oxygen. In terms of hemodynamics and oxygen transport, hemodilution with HBOC-201, either in low or clinically relevant doses, provided no apparent benefit over hemodilution with HES.

If a tHb is used perioperatively, studies must be completed to address concerns that cell-free Hb could alter red cell fragility or that shear forces typically found at the air/fluid interface of a cell-saving device could alter the structure of the tHb. To this end, 16 blood units were collected from healthy volunteers into CPD (blood preservation) solution and diluted with either HBOC-201 (1500 mg/dl; $N = 10$) or normal saline (equivalent volume; $N = 6$) (61). Automated, interoperative blood salvage was mimicked by suctioning the units into the reservoir of an autologous blood salvage system and processing the resulting solution. (The unit was operated manually, because the optical sensor that monitors the interface of red cells and supernatant in the centrifuge bowl of the unit responded to the red color of the bovine Hb in the supernatant and initiated the wash cycle prematurely and inappropriately.) Samples were analyzed for the concentration and molecular weight distribution of plasma hemoglobin and red cell morphology presalvage and following processing and washing. Five of the HBOC-201 units underwent a second 1000 ml wash. As a result of processing, the plasma Hb concentration of the HBOC units decreased from 1311 ± 265 mg/dl prior to "salvage" to 27.8 ± 19.6 mg/dl afterwards. (Processing the solution a second time further reduced the plasma Hb concentration to 6.5 ± 2.2 mg/dl.) Exposure to HBOC-201 did not alter red cell or platelet morphology, either pre- or postprocessing. Moreover, processing the HBOC-containing solution did not alter red cell fragility.

Hemodilution with a tHb as a part of cardiac surgery using cardiopulmonary bypass offers additional opportunity for conserving autologous blood. With the onset of bypass, a portion of the autologous blood volume could be collected via the venous cannulae and stored in a "set aside" reservoir. Simultaneously, a tHb solution could be included in the pump "prime" to replace the volume and oxygen-carrying capacity of the "set aside" erythrocyte mass. As a result, blood loss during the operation would include a lower percentage of the original erythrocyte mass, leading to net conservation of the autologous erythrocyte mass. After completion of cardiopulmonary bypass, the stored autologous whole blood could be returned to the patient with its associated clotting function,

thus contributing to decreasing homologous blood requirements. Additionally, setting aside a significant part of the autologous blood volume just before bypass may preserve platelet function.

Neya et al. studied the properties and stability of HBOC-201 in the bypass environment (62). One test preparation consisted of HBOC-201 solution diluted 1:1 with normal saline, and the second comprised 500 ml of blood added to 500 ml normal saline and 1 L of the HBOC-201 solution. Two liters of each test solution, warmed to 37°C, were evaluated in two types of cardiopulmonary bypass circuit models—a bubble oxygenator with an integral reservoir (BO; $n = 5$ preparations) and a membrane oxygenator/venous cardiectomy reservoir (MO; $n = 5$ preparations). A 50-ml sample of each diluted HBOC, oxygenated for 3 min in the bypass circuit and then stored at 37°C for 5 h, served as a control. Extracorporeal circulation in a continuous loop circuit was maintained at 2 L/min throughout the 5-h experiment. Samples were obtained after 30 min of bypass and hourly, thereafter, for determination of THb, metHb, the size-exclusion chromatographic and oxygen-binding profiles of the plasma Hb, and infusate pH. (Bicarbonate was added, as needed, to maintain physiological pH.) The Hb concentration did not change throughout the experiment. However, the MetHb content increased in all groups not containing RBCs and in the control from an initial value of about 3% to values of 22–28% at the end of the 5-h experiment. After 2–3 h circulation, the time required for most bypass surgeries, the metHb content was 15–20% in these HBOC-201 solutions. As McGown et al. found (63), RBCs significantly attenuated the rate and extent of Hb oxidation; in the presence of red cells, the plasma metHb content was about 8–10% after 2–3 h circulation and about 12% at the end of the experiment. As expected, increases in metHb content decreased the P_{50} value of the tHb (64).

In summary, the use of tHb solutions to counter acute hypovolemia and concurrent global ischemia attendant to trauma or surgery has met with widely divergent results. The reports available at the time of this writing suggest that when limited volumes of either DCLHb or PolySFH-P are infused in surgical patients, the infusion is tolerated well. Moreover, tHb infusion may reduce the requirement for blood transfusion. In contrast, when larger volumes of these tHbs are administered, only PolySFH-P continues to show potential utility in this indication. Development of DCLHb was terminated after unusually high mortality was observed in some trials.

Likewise, clinical assessment of the use of tHb solutions as a hemodiluent continues. If therapeutic benefit is confirmed in larger numbers of patients and

efficacy is shown, it is reasonable to anticipate regulatory approval and clinical acceptance.

Normalization of Hemodynamics and Tissue Oxygenation in Septicemia

Sepsis and its complications—hypotension, shock, and multiple organ system failure—represent a significant cause of mortality among hospitalized patients (65). Moreover, recent studies in a rodent septicemia model suggest that transfusion of stored red cells fails to effect acute improvement in tissue oxygenation in septic animals (66). The effects of tHb solutions have been assessed in a variety of animal models of septicemia (67–69) and in patients. The studies described below used specific tHbs. However, if the restoration of both adequate hemodynamics and tissue oxygenation is key to clinical utility in this indication, it is likely that similar results will be observed following the infusion of other tHbs that have these properties.

Clinical studies of the safety and efficacy of Hb-PHP (Apex) as a septicemia therapy are continuing. In a Phase I/II open label, ascending-dose study, patients with volume-refractory, vasopressor-dependent shock secondary to sepsis or presumed sepsis received a bolus infusion of Hb-PHP (25, 50, or 100 mg Hb/kg body weight) over 30 min and were observed for cardiovascular changes and adverse effects (70). Hb infusion was well tolerated and permitted a decrease in vasopressor utilization while maintaining an increasing MAP. No adverse effects on pulmonary, cardiac, renal, or hepatic functions have been noted at this dose level; Hb-PHP infusion increased SVR and decreased HR. These results supported further study of Hb-PHP in this indication.

In a subsequent Phase I/II study that allowed for dose escalation and extended duration of dosing, patients with shock secondary to sepsis or presumed sepsis received a dose of Hb-PHP of 160, 320, 640, or 2560 mg Hb/kg body weight (70). Dosing was initiated at a rate of 10 mg Hb/kg/h and increased in increments of 10 mg/kg/h at 15 min intervals to a maximum rate of 20, 40, or 80 mg Hb/kg/h for the lowest, middle, or two highest doses, respectively. At the highest dose, dosing was stopped if shock resolved (i.e., MAP >70 mm Hg without vasopressor). The dose rate was determined by titration to hemodynamic effect (i.e., the standard of care guideline of >70 mm Hg). Again, Hb-PHP infusion was well tolerated. The increased MAP and SVR observed following Hb infusion permitted a decrease in vasopressor utilization. The results supported a proposal that Hb-PHP be indicated for the treatment of shock associated with systemic inflammatory response syndrome (SIRS) in

fluid-refractory, pressor-dependent patients. A Phase III pivotal trial is planned.

Likewise, the hemodynamic effects of DCLHb (an intramolecularly cross-linked human Hb; Baxter) in 14 critically ill patients were evaluated in a prospective, observational study (71). All of the subjects required vasopressor therapy to maintain adequate MAP and had secondary organ dysfunction prior to DCLHb treatment. Following a decision by the investigator to infuse, boluses (100 ml) of DCLHb were given up to a maximum of 500 ml. Each infusion was separated by 60–90 min. Hemodynamic parameters, norepinephrine and inotropic requirements, arterial and mixed venous blood gases, and urine output were monitored, and biochemical and hematological analyses were made before DCLHb administration and at multiple times up to 72 h postadministration. The main end-point employed to assess the efficacy of DCLHb as a vasopressor was maintenance of the MAP at approximately pre-infusion values concomitant with a reduction in norepinephrine requirements. This objective was met, and reductions in norepinephrine requirements were maintained at 24, 48, and 72 h ($p < 0.01$ at all time points). Thus, even in these critically ill patients, whose prognosis of survival was very poor, DCLHb seemed to have a beneficial effect.

Use as an Adjunct to Cancer Treatment

Local effects of a tHb comprise the second of three general categories for indications recognized by the FDA (51). For example, the concomitant use of a tHb and a cancer therapy, such as ionizing radiation or chemotherapeutic agents, potentially has two benefits for the patient. First, cancer patients are usually anemic, and evidence has been presented that anemia is an important prognostic factor. Increasing Hb levels into the normal range can improve prognosis and treatment outcome (72, 73). Second, the small size (relative to a red cell) and excellent oxygen delivery capabilities of a tHb have the potential to improve tumor oxygenation, a factor which has been associated with increased tumor cell killing and delays in tumor growth following radiation or chemotherapy. Moreover, the Hb may be a source of heme to promote hematopoiesis (see below).

Although clinical study data have not yet been reported, preclinical studies in rodent models suggest that the administration of tHbs may demonstrate many of these benefits. For example, the effects of PEG-Hb on tumor oxygenation and response to chemotherapy were monitored in a study in which two rodent models of solid tumor cancer were used (74). When a 6 ml/kg (360 mg/kg) dose of 6 g/dl PEG-Hb solution was administered to rats bearing

the 13762 mammary carcinoma immediately prior to measurements of tissue oxygenation, the level of hypoxia in the tumor was significantly decreased, particularly when the animals were exposed to an oxygen-enriched atmosphere. Similar observations were made when the measurements were made 24 h after the administration of cyclophosphamide, melphalan, taxol, or cisplatin—chemotherapeutic agents that render the tumors severely hypoxic. Likewise, the co-administration of a 6 ml/kg dose of PEG-Hb to mice bearing the EMT-6 murine mammary carcinoma immediately prior to the administration of cyclophosphamide, adriamycin, 5-fluorouracil, 1,3-bis (2-chloroethyl)-1-nitrosourea (BCNU), or taxol on each of 3–5 days of treatment (depending on the chemotherapeutic agent) increased the effectiveness of tumor cell killing, delayed tumor cell growth, and increased the response of the metastatic disease without changing the toxicity to the bone marrow, as monitored by changes in the colony stimulating activity (CFU-GM). Taken together, the authors concluded that further investigation of PEG-Hb as an oxygen delivery agent in oncology is warranted.

In a parallel study, it was found that the administration of an Hb solution decreases hypoxia and increases radiation response in a rat mammary carcinoma model (73, 75).

In addition to these three applications, more exploratory studies are assessing the use of tHbs for clinical indications where blood typically is not used. Several of these applications are described below.

Support for Hematopoiesis/Erythropoiesis

Hematopoiesis is the process of blood cell production that takes place in the bone marrow. Through a complex series of regulatory events, stem cells are differentiated into various types of cells, including red blood cells. Stem cell differentiation is responsive to exogenous stimuli and can be upregulated to resupply a deficient cell population. However, only a limited number of therapeutics (iron, hemin, or erythropoietin, for example) can stimulate Hb-replete red blood cell formation. Recognition of the poor bioavailability of many iron compounds or hemin compositions and the improved safety profile of a tHb relative to unmodified Hb has renewed the interest in Hb administration as a stimulant for red cell production first shown in the 1920s. The two studies described below reference the use of specific Hbs. However, it is likely that many other tHbs will provide support for hematopoiesis.

A recent report summarized the effect of administration of a recombinant Hb, hrHb 1.1, on *in vivo* hematopoietic recovery in rodent models from azidothymidine (AZT)

toxicity (77). The results suggest that this Hb can alleviate AZT toxicity in normal and immunodeficient mice and that it may be clinically useful in preventing severe bone marrow depression brought about by various drugs or agents, such as AZT. Similarly, Lutton et al. have reported the hematopoietic effects of 50 ml infusions of 6% β - β cross-linked human and bovine Hb (DECA-Hb and XLBV-Hb, respectively) and noncross-linked (HbA) into rabbits following a 50 ml ear-bleed (78). All rabbits receiving DECA-Hb or XLBV-Hb tolerated the Hbs well, whereas 50% of the animals transfused with similar doses of noncross-linked HbA died. Analysis of peripheral blood and bone marrow BFU-E and CFU-GM production 2 days following Hb-infusion revealed no significant variation in the generation of BFU-E and CFU-GM numbers for each cross-linked Hb group, but significant reductions were observed in the HbA group.

Approval has been obtained for Phase II trials in chronic renal failure patients to use o-Raff-Hb (Hemosol) as an adjunct to erythropoietin (79). To date, no preliminary reports are available.

Maintenance of Tissue Oxygenation Resulting from Focal Ischemia

Oxygen deprivation follows when blood flow is compromised by a local, temporary circulatory blockade that may occur naturally as a result of stenosis, stroke, or a myocardial infarction (heart attack) or may accompany therapeutic interventions, such as percutaneous transluminal coronary angioplasty (PTCA). Packed RBCs are not useful for restoration of oxygen delivery in these situations, because they are too large to traverse the constricted vessels or too viscous and fragile to be pumped through a perfusion balloon catheter. A number of preclinical studies in this area are summarized below.

Cerebral ischemia

In patients at risk for focal cerebral ischemia, hemodilution has been proposed as a prophylactic and resuscitative therapy to ameliorate brain injury. The rationale for hemodilution therapy is based on two facts: 1) blood viscosity decreases when blood is diluted with many crystalloid and colloid solutions; and 2) an inverse correlation exists between blood viscosity and cerebral blood flow (CBF). However, the beneficial effect on viscosity by hemodilution with nonoxygen-carrying fluids is countered by a loss in oxygen-transport capacity. A tHb, by decreasing viscosity while maintaining oxygen delivery capacity, theoretically may improve oxygen delivery to ischemic areas of the brain, compared to the oxygen delivery achieved with typical clinically utilized

intravenous solutions. This hypothesis has been tested in several animal models of cerebral ischemia and ischemic neuronal damage (80, 81).

Cole et al. have performed several investigations using an instrumented, spontaneously hypertensive rat model to evaluate the therapeutic effect of DCLHb on focal cerebral ischemia (82–85). In these studies, middle cerebral artery occlusion (MCAO) followed by 10–90 min of ischemia resulted in consistent ischemia in both cortical and subcortical tissue. In both MCAO models, Cole observed that hemodilution with DCLHb augmented CBF and decreased infarct volume in a dose-dependent manner. Moreover, infarct-size reduction was greater and brain edema was less when the blood pressure was allowed to increase. Because the data suggested that in the ischemic brain a reduction in viscosity is the predominant mechanism for a hemodilution-induced increase in CBF, the effects of dilution with DCLHb were compared to those of like doses of oncologically matched albumin. The data demonstrated that in these models of cerebral ischemia, DCLHb decreased ischemic brain injury more proficiently than did albumin. Moreover, since the volumes of hemodiluent that may be used are limited, the observation that 20 g/dl DCLHb performed better than 10 g/dl DCLHb suggested that a hyperoncotic oxygen-carrying hemodiluent, such as 20 g/dl DCLHb, may be preferable to a relatively normo-oncotic fluid.

There is some evidence, however, that the potential utility of a tHb as a hemodiluent and tissue-oxygenating solution may be dependent on the structure and properties of that tHb. For example, the results of a study of the effects of partial blood replacement with pyridoxalated Hb polyoxyethylene conjugate (PHP) solution (Ajinomoto) in a gerbil model of transient cerebral ischemia suggested that PHP solution holds less promise as a clinical treatment for this type of focal ischemia than does DCLHb (86).

Therapeutic Hbs such as DCLHb potentially are unique hemodiluents in that in addition to reducing viscosity, cerebral ischemic injury may also be reduced by the maintenance of oxygen delivery by the tHb and by the inactivation of nitric oxide (NO) (87). This hypothesis was supported by the results of a study of cerebral ischemic injury in rats in which the effects of DCLHb infusion were compared to those of DCLHb infusion with L-NAME (an NO synthase inhibitor) or L-arginine (an NO synthase substrate) (88). It was found that the beneficial effects of DCLHb in reducing infarct size were not compromised by the co-administration of a NOS inhibitor. In contrast, co-administration of an NOS substrate significantly offset the benefits of hemodilution with DCLHb.

In patients surviving the initial hemorrhagic insult of a ruptured cerebral aneurysm, delayed cerebral ischemia

with infarction, due to vasospasm of cerebral arteries, is a major cause of death and disability (89). OxyHb, among other erythrocyte breakdown products, is thought to be a major factor in the pathogenesis of cerebral vasospasm after subarachnoid hemorrhage (90). Therefore, administration of a tHb might have a detrimental effect on cerebral perfusion after subarachnoid hemorrhage. Cole evaluated the effect of subarachnoid administration of DCLHb, blood, or an artificial cerebrospinal fluid on cerebral blood flow in a spontaneously hypertensive rat model (91). Although subarachnoid Hb statistically increased the area of hypoperfusion compared with artificial cerebrospinal fluid, the area of hypoperfusion was only about 20% of that observed after a similar volume of blood was given in the subarachnoid space. Significantly, despite the observation that DCLHb slightly increased the area of hypoperfusion, infarct size was lower in the presence of this tHb.

Taken together, all of these preclinical studies supported completion of a randomized, prospective trial of DCLHb administration to stroke patients (92). In this Phase II study, increasing doses of DCLHb (25, 50, and 100 mg/kg, $n = 8$, 8, and 11, respectively) or placebo (normal saline; $n = 26$) were infused intravenously every 6 h for 72 h to patients with an acute ischemic stroke, consistent with localization in the anterior cerebral circulation. Patients were selected who had suffered the present stroke symptoms for less than 18 h, in whom a brain-computed tomography scan was normal or compatible with a recent infarction, and who were likely to survive for at least 3 months. Blood pressure (MAP) and HR were measured every 15 min, and plasma concentrations of endothelin-1 (ET-1), catecholamines, renin, vasopressin, and atrial natriuretic peptide were measured before and 24 h and 77 h after the start of infusions. In addition, blood was sampled for routine laboratory evaluations of liver enzymes, blood urea nitrogen, serum creatinine, creatine phosphokinase, and a complete blood count. In the placebo group, MAP was 112 (109–115) mm Hg at baseline and decreased spontaneously with time, a common observation in stroke patients. This decrease in MAP was attenuated in patients treated with DCLHb, reaching statistical significance in the highest dose group. (Nonetheless, there was no need for antihypertensive treatment in comparison with the control group, nor were there any signs or symptoms of hypertensive encephalopathy or hemorrhage into the infarct.) In contrast to the marked fall in MAP, HR did not change in the placebo group during the 66 h after start of infusions, but it had decreased at both time points after infusion of the highest dose of DCLHb. The plasma ET-1 concentration decreased slightly in the placebo group, from a value near the high extreme of the normal range toward the

midpoint of that range after 24 h and 66 h, but increased dose-dependently in response to DCLHb infusion. With the highest dose of DCLHb, the plasma ET-1 concentration rose from the high normal range at baseline to 4- or 5-fold higher values after 24 h and 66 h ($P < 0.001$ at both intervals). The increases in the plasma ET-1 concentration and in MAP were correlated ($r = 0.30$, $P = 0.02$). Other vasoactive hormones were not affected by the DCLHb infusion, and their concentrations did not correlate with the observed increases in plasma ET-1 concentration (93). Side effects considered directly related to DCLHb included a transient yellow discoloration of the skin and sclerae (0/8, 1/8, and 11/11) and hemoglobinuria (2/8, 6/8, and 11/11). The hemoglobinuria was caused by excretion of DCLHb in the urine and was not associated with any impairment of renal function as assessed by serum creatinine. In a 3-month monitoring period that followed, 3/26 control patients died, all of whom were in the high-saline group and 7/27 DCLHb patients died (3/8, 1/8, and 3/11) ($P = 0.16$). Treatment with DCLHb did not have a favorable effect on the neurological outcome as assessed by the Rankin disability score. In contrast to the favorable effects associated with expression of Hbs pressor activity in the preclinical studies, the absence of a fall in blood pressure in the DCLHb group, while well tolerated, did not have a beneficial effect on survival or degree of disability. It was concluded that administration of DCLHb in this patient population resulted in significantly worse neurological outcomes and more deaths than conventional treatment.

Cardiac ischemia

PTCA is an established therapeutic procedure for the treatment of obstructive coronary artery disease. Balloon inflation during PTCA produces a transient interruption of coronary blood flow to the vascular bed distal to the inflation, giving rise to temporary regional myocardial ischemia and increased risk of ventricular arrhythmias. Perfusion of the distal bed with an oxygen-carrying solution during PTCA has been shown to have beneficial effects (94). However, the use of a perfluorocarbon emulsion approved in the United States for this indication, FluosolTM DA 20%, was relatively inconvenient and has had questionable results (95). (This product is no longer marketed.) Since a tHb transports oxygen in a similar way to whole blood and can be perfused through an angioplasty catheter during balloon occlusion, it was hypothesized that use of a tHb may increase myocardial oxygenation and reduce myocardial ischemia during PTCA. This hypothesis has been tested in a swine model, and it was found that perfusion with DCLHb during PTCA allowed the extension of balloon occlusion times to 4 min or longer

without a significant decrease in cardiac function (96). Although these preliminary findings are encouraging, no additional preclinical or clinical trials of a tHb have been reported for this indication.

Improvement of Oxygen Delivery in Sick Cell Anemia Patients

Sickle cell anemia is a genetic disorder characterized by chronic hemolysis and intermittent vaso-occlusive crises. An element of the pathophysiology of these vaso-occlusive episodes is tissue ischemia and/or infarction subsequent to an impairment in oxygen delivery. Recently, a Phase I/II clinical study was completed in which a single dose of 0.2, 0.4, or 0.6 g/kg of HBOC-201 was administered to adult patients with sickle cell disease (97). Normal saline solution was used as the control. The subjects were not in crisis at the time of the study. The parameters used to assess the safety of the hemoglobin included vital signs, clinical laboratory tests (complete blood counts, urinalysis, coagulation profile, serum electrolytes, and tests of renal and liver function), transcutaneous oximetry, and an adverse experience profile. In addition, each subject underwent an assessment of selected aspects of muscular function consisting of a hand-grip strength test and an aerobic capacity evaluation. The latter tests were performed at baseline (16 h prior to the administration of study treatment) and 4 h and 48 h after the initiation of HBOC-201 infusion.

The HBOC-201 infusions were well tolerated by the subjects, and no evidence of toxicity was noted. There was no appreciable change in the circulating hemoglobin level, the blood pressure, or any of the other hematological or clinical laboratory parameters. Hemoglobin was not detected in the urine, nor were signs of liver or kidney toxicity observed. The hand-grip testing showed no significant differences between the study treatment groups at any of the testing sessions. Although the exercise intensity for all of the patient groups during each of the aerobic exercise tests was not different, HR responses of the HBOC-201 patients differed significantly from those of the placebo patients at the various exercise levels. At +4 h, the HR of the subjects who had received HBOC-201 were significantly lower than the HR that these same subjects had exhibited during their baseline exercise period. At +48 h and in the saline group, these differences were not observed. Likewise, at +4 h there was a smaller rise in the lactate level of the HBOC-201 subjects as compared to the baseline period; this difference was not statistically significant, nor was it observed at +48 h or in the placebo group.

Thus, this study in sickle cell disease patients appeared to support the findings of a parallel study in normal human volunteers, where it was concluded that HBOC-201 apparently supported submaximal exercise capacity in a manner quite similar to autologous blood, but at lower pulse rate, CI, and lactate concentration (98). The results of both studies suggest that the provision of acellular Hb may be an efficient way to deliver oxygen to the tissues.

Resuscitation Following Pressure Contusions

The administration of small volumes of a therapeutic hemoglobin solution has also been proposed for the treatment of two types of lethal, pressure-related injuries. The first type of injury is a contusion or barotrauma-like injury to the lungs and gastrointestinal tract resulting from blast overpressure (the abrupt, rapid rise in atmospheric pressure resulting from explosive detonation) or nonpenetrating trauma (99). The second type of injury is severe head injury and associated hypotension. For example, pulmonary contusion occurs in almost 20% of nonpenetrating trauma admissions with multiple injuries (Injury Severity Score of >15); (Yale Trauma Registry), and is associated with a marked increase in the frequency of pneumonia, respiratory failure, adult respiratory distress syndrome, and death. In survivors, the probability of chronic, disabling respiratory problems is increased. Likewise, head injury is the leading cause of traumatic death in the United States; when associated with hypotension, the incidence of adverse outcome (severe disablement, vegetative quality of life, or death) is increased significantly. Injuries of these types are complicated by secondary damage, often at sites remote from the area of primary injury. Aggressive resuscitation to restore tissue oxygenation can prevent secondary insults and improve outcome. However, since these injuries are accompanied by edema, aggressive fluid resuscitation with large volumes of electrolyte solutions may be contraindicated. In contrast, low volume resuscitation with a hyperoncotic, oxygen-carrying hemoglobin solution has the potential to restore tissue perfusion and oxygenation and minimize water accumulation at the site of injury and the surrounding tissues.

Cohn et al. used a swine model to determine the impact of a vasoactive red cell substitute on respiratory derangements after traumatic lung injury (100). Young Yorkshire pigs ($n = 6/\text{group}$) were anesthetized and mechanically ventilated. Each animal then received pneumatic blasts to the right thoracic cage at baseline and was hemorrhaged a total of 30 ml/kg over 20 min prior to resuscitation with 0.9% saline (90 ml/kg) or diaspirin cross-linked hemoglobin (DCLHb; 15 ml/kg) from $t = 20$

to 40 min. The infusion of maintenance fluids (in the saline group and 4 ml/kg/ in the DCLHb group) was continued until the end of the 4-h observation period following the initial injury. Serial pulmonary and systemic hemodynamic measurements were used to monitor pulmonary and hemodynamic status, and total thoracic compliance assessment, spiral three-dimensional computed tomography scans, and lung weights ($n = 3/\text{group}$) were used to assess lesion size and lung water. MAP was restored in both groups but was significantly higher in the animals receiving the hemoglobin solution. Oxygenation worsened somewhat in both groups. Compliance worsened in both groups but was significantly worse at the end of the experiment in animals receiving DCLHb. The authors also found that resuscitation with DCLHb led to greater contusion lesion size. Thus, this therapeutic hemoglobin does not appear to have benefit in this model of barotrauma-like injury.

On the other hand, Shackford investigated small-volume resuscitation (4 ml/kg) using RL, hypertonic saline-dextran (HSD) and DCLHb in a porcine model of cryogenic brain injury and shock and found that both the HSD and the DCLHb had benefit in this model (101). In these experiments, swine were subjected to a cryogenic injury and a Wiggers-type hemorrhage and then resuscitated with a small-volume bolus of an asanguinous solution. The first regimen consisted of a bolus of RL followed by a constant infusion to return the MAP to baseline (RL/RL). The second regimen consisted of an initial bolus of 7.5% hypertonic saline in 6% dextran solution followed by a continuous infusion of RL to restore the MAP to baseline (HSD/RL). The third regimen consisted of a bolus of HSD followed by a constant infusion of hypertonic sodium lactate to restore the MAP to baseline (HSD/HSL). Shed blood was returned 1 h after the initiation of resuscitation in all three regimens. In a second experiment a test group received a bolus of DCLHb; no shed blood was returned. In summary, small-volume resuscitation with either hypertonic solutions or with hemoglobin solutions had favorable effects on the cerebral perfusion pressure with decreasing intracranial compliance.

The clinical management of brain injury includes the use of osmotic and other diuretics to attempt to control tissue edema and agents which modify the rheology of blood in order to better perfuse the brain and counter ischemia. Two recent reports summarize the effects of hemodilution with DCLHb or oncologically matched plasma protein solution following traumatic brain injury in rodents. Piper monitored intraventricular intracranial pressure (ICP) and CBF in the region of the sensorimotor cortex 4 h after a severe weight-drop injury to the brain

and found that DCLHb significantly reduced ICP from mean 13 ± 2 to 3 ± 1 mm Hg ($P < 0.001$) and increased CBF from 21 ± 2 to 29 ± 2 ml/100 g/min ($P = 0.032$). It was concluded that DCLHb improved coronary perfusion pressure without a reduction in CBF in this rodent model of posttraumatic brain swelling (102). Likewise, Muldoon examined coronal samples of the hippocampus for the presence of necrotic neurons following traumatic brain injury to anesthetized rats treated by hemodilution with DCLHb or oncologically matched human serum albumin (103). Whereas treatment with albumin showed a trend toward a reduction in the number of necrotic neurons in the hippocampus, treatment with DCLHb significantly reduced neuronal necrosis at this site, suggesting that hemodilution with the tHb decreases hippocampal neuronal ischemia when administered following traumatic brain injury.

These preclinical results suggest further investigation into the use of therapeutic hemoglobins after severe head trauma may be warranted.

SAFETY

With the infusion of an acellular tHb, the Hb concentration in the plasma rises from 1 to 3 mg/dl (normal values) to concentrations as high as several grams per deciliter. Concentrations such as these are sufficiently high (≥ 15 mg/kg) to overwhelm the ability of haptoglobin, the physiological hemoglobin scavenger, to bind and remove it from the circulation (14). In addition, some tHbs do not bind to haptoglobin, negating its value for Hb removal.

The protein is administered and circulates for varying lengths of time as oxyHb. However, with time, the ferrous iron contained in each heme ligand undergoes either auto- or chemically induced oxidation, yielding metHb (104, 105). The rate at which Hb oxidation occurs depends on many factors, including the structure of the Hb, the concentrations of natural or exogenous oxidizing agents and perhaps, the pathophysiological state of the patient (106). Oxidation of acellular Hb to metHb is a continuous process, in contrast to the condition within the red cell, where the metHb concentration is normally maintained at 2–3% of the total Hb by the methemoglobin reductase system (14). Therefore, the metHb content of an acellular Hb can rise to very high percentages of the total circulating Hb; the maximum plasma metHb concentration, however, is typically several hundred mg/dl.

Within the red cell, enzymatic systems reduce the metHb to Hb (14). In the plasma, however, the

concentrations of metHb-reducing agents are very low. The only effective *in vivo* reducing agent system for acellular metHb apparently is exported from the red cell; the ability of this unidentified agent to reduce metHb *in vitro* has been demonstrated in several laboratories (62, 63) but is difficult to quantify *in vivo*.

Finally, the possible presence of high concentrations of metHb in the plasma raises concerns about inappropriate release of heme and iron. Whereas normal plasma concentrations of heme are negligible, heme-protein interactions are weakened in metHb, as a result of which heme may be transferred to lipophilic surfaces, such as RBC or endothelial cell membranes. Heme-association at these sites has been reported to promote cell lysis, either as a direct effect or through synergistic effects with hydrogen peroxide (108). Similarly, scientists and clinicians have long recognized the ability of iron, both “free” and sequestered within enzymes, to catalyze oxidation reactions *in vitro* and *in vivo*. Thus, the safety of administration of large quantities of a redox-active Hb continues to be actively debated (109–112).

Given these possibilities (summarized in Fig. 2.), if a redox-active, acellular Hb was to exacerbate injury, it is reasonable to anticipate that cell and tissue exposure *in vivo* to large quantities of Hb would have both directly and indirectly observed consequences. For example, if generation of reactive oxygen species (e.g., superoxide radical anion, hydrogen peroxide, or peroxynitrite) were a consequence of Hb exposure, increased leukocyte adhesion, peripheral blood mononuclear cell activation, and changes in cell-wall integrity would typically be observed. Likewise, indirect effects of Hb-related injury would also be expected, such as increased sensitivity of protein-overloaded cells to ischemic damage (in the renal

tubular epithelium, for example, where uptake of acellular Hb is known to occur) or cell-surface obstruction caused by Hb binding or deposition (e.g., cast formation in the renal tubules, a frequent observation following infusion of native, unmodified Hbs).

Numerous concerns about the safety of hemoglobin preparations have been presented in the literature. Some of the historical safety concerns have been addressed and are no longer issues with respect to the hemoglobin formulations now in advanced clinical trials. For example, effects such as neutrophil and macrophage activation, the formation of microthrombi, and platelet aggregation appear to have been eliminated from these Hb formulations. However, new safety concerns, centering on effects that have been observed clinically, have been raised. Three principal concerns will be discussed here: 1) Hb's pressor activity, 2) the potential for acellular Hb to exacerbate injury, and 3) the potential for Hb to potentiate or exacerbate infection.

Hemoglobin's Pressor Activity

Following the infusion into mammals of most, if not all, tHb solutions, increases in mean arterial and pulmonary artery pressures and systemic and pulmonary vascular resistances are observed. The observations vary with the species and the preclinical or clinical model being studied. For the most part, the increases in MAP are rapid but apparently self-limiting (i.e., there is no continuous, dose-related escalation). Their duration apparently depends on the structure or some other property of the Hb infusate and the dose administered.

A growing body of evidence indicates that the interplay of multiple mechanisms contributes to Hb's pressor activity. Some of these mechanisms, namely, the blood pressure increases related to Hb's oncotic properties and viscosity, appear to be of lesser importance. For example, significant pressor activity is not observed following infusion of oncotically matched albumin solutions or molecular weight-paired dextran or hydroxyethylstarch solutions. Likewise, the vasoconstriction elicited by the Hb solutions appeared to override the vasodilation associated with decreases in viscosity due to hemodilution (113).

Both NO and endothelin-1 (ET-1), acting in concert with each other and in synergy with other vasoactive substances, are believed to mediate a broad spectrum of physiological actions, including homeostatic regulation of blood pressure, vascular tone, and blood flow to various tissues, as well as enzyme and fluid secretion in organs such as the kidney and pancreas (114–116). Since the Hb-related effects on blood pressure and hemodynamics appear to involve both of these mediators, two key

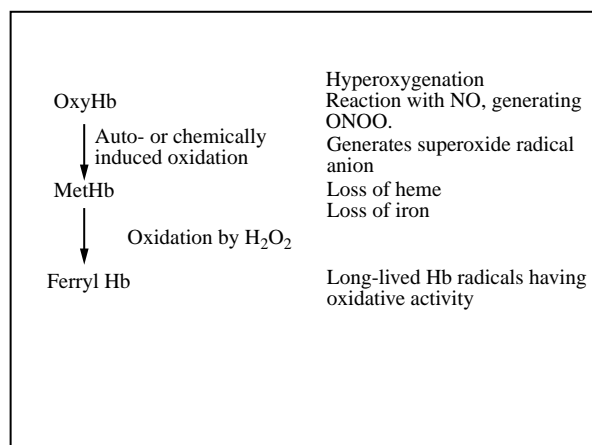


Fig. 2 Potential mechanisms for Hb-mediated injury.

questions have been raised. One, does the Hb-related increase in blood pressure, commonly viewed as a sign of sufficient resuscitation following hemorrhage, mask on-going, decreased organ perfusion and local ischemia in the microcirculation of key organs such as the heart, liver, kidney, or gut? In hypovolemic or ischemic states, tissue perfusion and oxygenation already may be compromised in these critical areas, and augmented vasoconstriction could exacerbate injury. Two, do the mechanisms underlying the pressor effect alter the functionality of other critical physiological systems, particularly those in the secretory organs (e.g., the kidney or pancreas)? Both questions remain at the forefront of clinical and regulatory concerns.

To address the first of these concerns, many preclinical studies on the effects of Hb administration in hypervolemia or resuscitation from hemorrhagic shock have been completed using a variety of techniques. For example, the effects of tHb on the perfusion of microcirculation have been assessed using hemodynamic measurements, acid-base status, radioactive microsphere measurements, blood flow measurement with ultrasonic flow probes, and intravital microscopy (39). Oxygenation of key tissues has been assessed via palladium porphyrin phosphorescence techniques, microplatinum electrodes placed in muscle tissue, oxygen-sensing electrodes, and measurement of whole body oxygen consumption. A spectrum of studies in a variety of animal models of hemorrhagic shock have demonstrated that resuscitation with a tHb typically restores MAP, base deficit, and subcutaneous and mucosal pO_2 to baseline levels (118–120). Studies of the microcirculation in organs susceptible to post-ischemic organ damage, such as the gut or pancreas, have shown a restoration of microcirculatory function by a tHb comparable to that achieved by blood transfusion (121, 122). Likewise, the administration of a tHb to septic animals has been shown to restore MAP and systemic vascular resistance and maintain blood flow to key organs (123). Taken together, the experimental data suggest that the pressor effect induced by a tHb may alter blood flow in the microcirculation and key organs in beneficial ways, although the redistribution of blood flow may vary from Hb to Hb (124, 125).

With respect to the second question, numerous recent reports suggest that the scavenging of NO by Hb or the increases in ET-1 may have both local and remote systemic effects. Thus, NO-scavenging by acellular Hb has been proposed as a cause of the frequent occurrence of gastrointestinal symptoms reported following the infusion of a genetically engineered, intramolecularly cross-linked Hb (126, 127) and o-Raff-Hb (56) in humans. Similarly, it has been proposed that the elevations in serum amylase

and lipase observed following infusion of many tHbs may also relate to an inhibition of NO-mediated activity or increased plasma ET-1 (128). For example, ET-1 is known to reduce pancreatic perfusion (129), and its Hb-augmented synthesis or release may adversely influence microcirculatory oxygenation.

Although the renal safety profile of the current generation of tHbs appears to be satisfactory, Hb modification greatly reduces, but does not eliminate, Hb uptake and filtration in the kidney. Hypothetically, therefore, the presence of high concentrations of acellular Hb in the tubular cells kidney could interfere with the processing of other proteins by this organ. Interestingly, Gburek et al. have recently identified Hb binding sites in the distal tubule of the rat kidney that recognized a motif common to rat, human, and swine hemoglobins (130). If similar binding sites are identified in human Hb-processing organs, it is reasonable to consider that the slow pinocytotic uptake of bound Hb could interfere with the binding of other proteins and peptides, an interference that would effectively elevate plasma concentrations of these moieties.

Preliminary data concerning clinical trials in which a tHb has been infused in resuscitation from hemorrhagic shock or as a substitute for blood have shown widely disparate findings. The available data are likely too sparse to draw conclusions as to whether the pressor activity of a tHb or its absence is a meaningful factor in the therapeutic utility of a tHb.

The overproduction of NO in septic shock has made this molecule a major target for therapeutic intervention in endotoxemia (131, 132). Since nitric oxide synthase (NOS) inhibitors such as arginine derivatives (e.g., L-NMMA, L-NAME, and NNLA) or methylene blue appear to improve blood pressure and vascular function, the potential usefulness of these agents for treatment of septic shock has been advocated. However, clinical studies using these inhibitors, although limited in number, indicate that these unselective NOS inhibitors may not improve survival, even though they increase MAP and vascular resistance and partially restore vascular reactivity to noradrenaline (133). The NO-scavenging actions of Hb have prompted both preclinical and clinical studies of its therapeutic benefit in treatment of septic shock. For example, a recent study using a rodent model of endotoxemia compared the effects on hemodynamics and renal function of polymerized bovine Hb, HBOC-201, with those of NOS inhibitors (67). The results clearly indicated that treatment with HBOC-201, but not with unselective NOS inhibitors, could improve both renal and cardiovascular function in rats suffering from septic shock.

Clinical trial data for this indication also are sparse. An observational study of the hemodynamic effects of DCLHb in 14 critically ill patients with septicemic shock or systemic inflammatory response syndrome suggested a beneficial effect (71). Likewise, patient stabilization and a decrease in the requirement for vasopressor therapy has been observed following the administration of PHP-Hb in volume-refractory, vasopressor-dependent shock patients (70).

The Potential for Acellular Hb to Exacerbate Injury

Whereas the adverse findings described above have sometimes been reported following completion of in vitro experiments, in general, they have not been demonstrated in preclinical and clinical studies of tHbs. Thus, preclinical studies in a sizeable number of animal species have generally shown no signs of inappropriately rapid Hb degradation (i.e., unusually rapid oxidation to metHb or unanticipated heme loss), no changes in the concentration of non-transferrin bound iron in the serum, and an absence of toxicities related to hemin or bilirubin-induced red cell hemolysis following volume-load, or partial or complete exchange of the test animal's blood for a tHb (134). Similarly, intravital microscopic studies of the microcirculation following hemorrhagic shock and resuscitation with a tHb solution have generally revealed a restoration of microcirculatory function not realized following resuscitation with crystalloid or colloid solutions (135). Moreover, resuscitation with a tHb solution has tended to restore functional capillary density to prehemorrhage conditions and is generally complemented by an absence of signs of damage to the endothelium, such as increased leukocyte adhesion or increased endothelial cell-wall permeability to dye-labeled macromolecules. Although some groups have observed an elevation in the concentration of thiobarbituric acid-reactive materials in the tissue or conjugated diene release following Hb exposure, histopathologic changes (edema, necrosis, cell infiltration, and hemorrhage) or significant reductions in organ function were generally absent (136). Moreover, surrogate markers of effective resuscitation, such as a restoration of physiological, acid-base balance or intestinal mucosal pH_i, suggest that microcirculatory function has been restored.

In the various trials conducted to date, hundreds of patients have received low-to-moderate doses of a tHb solution in clinical tests. A limited number of patients have received larger doses of 100–500 g of a tHb. For the most part, the infusions have been well tolerated. However,

several observations have been made relatively consistently and will be discussed below.

Pseudojaundice

A tHb cleared from the systemic circulation will undergo catabolism to bilirubin in several organs and tissues, including the kidney, liver, and spleen. Following the administration of a tHb to animals, increases in liver weight at necropsy and microscopic observations of pigmented Kupffer cells are common observations. Liver sections stain with Prussian blue, suggesting that an iron-containing pigment is present within hepatocytes and Kupffer cells. The absence of elevations in ALT indicates these changes are not associated with liver hepatotoxicity. The pigmentation decreases markedly during the recovery period. All of these observations point to increased acellular Hb catabolism by this organ. Consistent with this catabolism, during preclinical studies in animal models, serum bilirubin concentrations increased following the infusion of a tHb and gradually returned to normal values during the recovery period.

This pseudo-jaundice has also been observed clinically. In a study in critically ill patients, the most common observation following the administration of DCLHb was the observation of a yellowing of the patient's skin, which was either of new appearance or was an increase in the intensity of a pre-existing color change (71). The effect was documented in 6 of the 14 patients in the study; 3 of the affected patients had received a total of 20 g of DCLHb (200 ml), and 3 a total of 50 g (500 ml). Uniformly, the effect was transient: color peaked at 24 h post-infusion and decreased to pre-infusion levels within 5 days. The pseudo-jaundice was associated with an increase in bilirubin but with no significant change in any other liver function test. A similar incidence of transient pseudo-jaundice was also reported for 39 of 104 patients in the cardiac surgery trial of DCLHb (54). These patients generally received 25 to 75 g (250–750 ml) of DCLHb. No signs indicating liver dysfunction were observed. A review of the data indicates that it is likely an expected result of the intracellular metabolism of the plasma Hb.

Likewise, in the first 3 days following the infusion of PolySFH-P, the total bilirubin concentration increased to 2.4 ± 1.7 g/dl and then gradually declined (53). Although not cited as an adverse event in the report, the consequences of bilirubin concentrations as high as these are controversial. In an otherwise healthy infant, for example, these high bilirubin concentrations would cause concern about bilirubin neurotoxicity (137). However, under certain circumstances, it has been hypothesized that bilirubin may act as a natural antioxidant (138), and perhaps, that is the case here.

Although preliminary data from clinical trials of Hb-PHP or HBOC-201 are not available, it is reasonable to anticipate that as these tHbs are cleared from the circulation and metabolized, total bilirubin in the plasma will increase and then decline.

Enzyme and autocoid peptide elevation following hemoglobin infusion

A reversible elevation in the concentrations of several serum enzymes or autocoid peptides has been seen in preclinical studies following single and repeated dose infusions of some tHbs. For example, in various studies, the infusion of a single dose of DCLHb (100–4000 mg/kg) in rats, dogs, and monkeys caused mild-to-moderate increases in the concentrations of AST and sorbitol dehydrogenase (SDH) (139). No histopathological changes were associated with these increases, and the enzyme concentrations returned to baseline values during the recovery period. Since similar AST elevations were seen following infusion of albumin solution, it was hypothesized that the increases might be related to the volume and protein load that was administered.

Likewise, following the repeated infusion of doses of 1000–2000 mg/kg DCLHb daily for 7 days or of 400 mg/kg every 6 h for 3 days, the concentrations of AST, lactate dehydrogenase (LDH) and creatine kinase (CK) were elevated in monkeys (139). Isoenzyme profiles for CK and LDH revealed predominant increases in the MM form of CK and the LD-5 form of LDH. The MM-CK originates predominantly from skeletal muscle and may also derive from the myocardium; however, the MB isoenzyme, which emanates only from myocardium, was not elevated. The elevation of LD-5 was also consistent with a skeletal muscle source.

Similarly, transient elevations of AST, ALT, γ -glutamyl transferase, and LDH were observed in dogs following the administration of 0.6 g/kg doses of o-Raff-Hb (56). The changes were related to liver catabolism of this acellular Hb.

Hemorrhage leads to an increase in the plasma ET-1 in rats and dogs and may be a part of a natural compensatory response (140, 141). The higher concentrations likely result from increases in the synthesis or release of ET-1 as a consequence of hemodynamic changes (142), because of the activation of stress hormones or the coagulation cascade (136), or as a result of a decreased clearance in the liver, kidneys, and lungs, organs where ET-1 is metabolized (117). In vivo studies have recently confirmed a significant role for the autocoid peptide ET-1 in the cardiovascular pressor effects observed following the administration of DCLHb (143). An increase in the plasma ET-1 concentration has been correlated with the DCLHb

pressor activity in rats and swine and with constriction of isolated pig pulmonary vessels (144–147). However, maximizing the predictive value of this information has been difficult, because the ET-receptor types and their density show considerable variability, both within tissues and organ systems and from species to species (142). Moreover, no information is available about the relationship between Hb structure or properties and increases in ET-1 concentrations.

Transient elevations of these same serum enzymes have been observed in clinical trials of the various tHbs. For example, at the 0.6 g/kg body weight dose of o-Raff-Hb, changes in total bilirubin, AST, ALT, γ -glutamyl transferase, and LDH levels were noted (148). The changes had no clinical significance and were ascribed to Hb catabolism by the liver. In addition, transient increases in serum amylase and lipase, markers of pancreatic function, were observed at the 24 h time point in patients receiving this dose. The magnitude and time course of these latter changes were not consistent with the pattern observed in acute pancreatitis and were not considered to be clinically significant.

Transient rises in serum AST also were observed in clinical trials of DCLHb in surgical and critically ill populations (54). Serum amylase elevations were observed in clinical studies of DCLHb after single or cumulative doses of 20 g or more of DCLHb. These elevations are typically 2–3 times the upper limits of normal serum enzyme concentrations. The low incidence of acute pancreatitis observed in these trials was at the expected level for these patient populations (148).

Finally, transient rises in serum AST and amylase were observed in a study of the clinical utility of PolySFH-P in acute trauma and urgent surgery patients (53). In the 24 h following infusion of this tHb, the concentrations of AST nearly doubled but declined toward pre-infusion levels thereafter. Similarly, for 3 days following infusion of PolySFH-P, the concentration was about double the pre-infusion level. The changes were not considered to be clinically significant.

Clinical trials of DCLHb have revealed that ET-1 plays a role in the hemodynamic effects induced by at least one tHb (92). Peripheral circulating levels of ET-1 in human plasma (1.6–4.9 pg/ml) are normally far below the concentrations associated with pathological conditions (142). However, infusion of repeated, 100 mg/kg doses of DCLHb was associated with a 4- to 5-fold increase in the plasma ET-1 concentration (92) to values equal to or higher than the ET-1 concentrations determined following an acute myocardial infarction, cardiogenic shock, or heart failure (149, 150, 151). Although an increase in the ET-1 concentration of this magnitude may be a response

peculiar to DCLHb, a clinical concern remains that other tHbs may also induce ET-1 release, albeit in lower quantities. In vivo, even low levels of the peptide may amplify the constrictor effect of other circulating hormones or autocoids such as NO (152).

The Potential for Hemoglobin to Potentiate or Exacerbate Infection

A major anticipated use of acellular Hb would be for the emergency treatment of patients with traumatic hemorrhagic shock. The probability of associated bacterial infections in this patient population is high. Thus, the potential for hemoglobin to potentiate or exacerbate bacterial infection has been discussed widely.

A lethal synergy of bacteria and some component of blood was first described in 1890. Initially Hb was identified as the causative agent in blood (153), but subsequent studies implicated hemin and iron as the likely factors in promoting mortality (154, 155). An acellular (met)Hb is potentially a source of both.

Moreover, it has been hypothesized that exposure to an acellular Hb may injure the epithelium of the gastrointestinal tract (156). Under normal conditions, this tissue acts as a highly selective barrier, which permits the absorption of nutrients but restricts the passage of microbes and potentially toxic macromolecules from within the lumen into the systemic compartment. Derangements in the barrier function of the gut have been implicated as being important in the increases in intestinal permeability in patients with sepsis or in burn victims (157), in patients with multiple trauma (158), and in patients undergoing cardiopulmonary bypass or major surgery (159), and in the pathogenesis of multiple organ system dysfunction in critically ill patients (160). In addition, the development of GI mucosal acidosis, a predictor of multiple organ system dysfunction and/or mortality in critically ill patients (161, 162), is strongly associated with the development of increased mucosal permeability (163). Compromised morphology of the intestinal barrier is a consistent finding in hemorrhage without resuscitation (164). However, the administration of a tHb frequently restores the pH_i in the intestine (165–168). Moreover, following the use of a representative tHb, DCLHb, histopathologic examination of intestinal tissue samples indicated cell wall integrity was maintained (169). In addition, tissue culture supported the conclusion that bacterial translocation was attenuated.

Whether or not Hb accentuates lethal effects of endotoxin is also not resolved. When inoculation with living *E. coli* or endotoxin has been used to induce sepsis

in mice, some Hb preparations have been found to demonstrate deleterious effects (170–172). In contrast, other studies using better characterized, modified Hb solutions have failed to confirm these findings, both in mice and in other animal models (172, 173).

For example, the hypothesis that DCLHb would improve blood pressure, organ perfusion, and mortality was tested in a rodent model of sepsis (169). Administration of this tHb to moribund, septic rats immediately reversed the decreased MAP, increased systemic vascular resistance (SVR) and by 24 h, significantly elevated perfusion to vital areas (intestines, heart, and brain) as compared to albumin-treated animals. In addition, areas that did not display an increase in perfusion also did not demonstrate any deficits, suggesting that they were being adequately perfused.

Similar observations have been made following the infusion of bolus doses of 50, 100, or 200 mg/kg of Hb-PHP in septic sheep (175). All three doses of Hb reversed the hyperdynamic circulation that had developed during sepsis by increasing MAP and SVR while decreasing CI. Although pulmonary arterial pressure increased after Hb infusion, it neither adversely affected arterial oxygen saturation nor resulted in the development of pulmonary edema. Moreover, there were no significant increases of conjugated dienes in the lung tissue, suggesting that increased free radical production is unlikely after the infusion of low doses of this tHb. Following administration of the 100 mg/kg dose, the increase in serum iron concentrations, while significant, never exceeded the iron-binding capacity, suggesting that no free iron was available to increase bacterial growth. No differences were seen between the Hb and control group in pulmonary bacterial clearance (the lung is the major organ of the RES in sheep) or in white blood cell counts and white blood cell differentiation, suggesting an absence of Hb-related interference with the activity of the reticuloendothelial system in this model. None of the animals treated with Hb died after treatment.

Although preliminary, the data from clinical trials of the various tHbs parallel the observations in the preclinical studies described above. For example, DCLHb was administered to critically ill patients with septicemic shock or systemic inflammatory response syndrome in two studies. In the first of these studies, serum iron concentrations were increased following administration of the tHb; however, no obvious toxicity or development of overwhelming sepsis occurred (71). On the contrary, measurable patient benefit appeared to derive from its administration. Likewise, in the second study, where DCLHb was used in the treatment of acute anemia in 23 critically ill, predominantly septic patients, oxygen

delivery, oxygen consumption, as well as overall organ function and metabolic status were maintained or improved during the 24 h following treatment (176).

Moreover, tHbs have been administered to patient populations that included sizeable fractions of elderly and anemic patients or patients having co-morbidity. Enhanced susceptibility to infection or its exacerbation following the administration of a tHb to these patients has not been reported. Taken together, the preliminary data from clinical trials suggest that a tHb may provide significant benefit in a variety of clinical indications where sepsis may develop or is present.

STATUS OF CLINICAL TRIALS

Given the hurdles and the costs associated with the development of a clinical product that is a new biologic, it is remarkable that at least five companies continue to be engaged in clinical trials of hemoglobin-based oxygen carriers (Table 2). Most of the companies are developing modified hemoglobins derived from natural sources (human or bovine red cells), but at least one firm is using a biotechnological approach. Each of the hemoglobins under study is unique but has properties that are common to hemoglobin (e.g., its oxygen-transport capability and color) and properties that are unique to the particular therapeutic hemoglobin (e.g., its composition and formulation).

When this review was written, a spectrum of clinical trials were being completed at various locations around the world. Many of the trials were safety studies and dose-escalation studies that are best described as Phase I and II clinical studies. However, a few trials were definitive Phase III efficacy studies that had a reasonable probability of culminating in regulatory filings. In these various trials, hundreds of patients have received low doses of hemoglobin solutions in clinical tests, and a more limited number of patients have received larger doses of 100–300 g of therapeutic hemoglobin (as much as 6 units of 10 g/dl hemoglobin solution). Lastly, development of two tHbs, DCLHb and rHb 1.1, had been discontinued following clinical trials that failed to demonstrate significant benefit. As of this writing, none of the current generation of therapeutic hemoglobins has been approved by a regulatory body for clinical use.

SUMMARY OF THE FIELD

Surprisingly, development of a therapeutic hemoglobin solution from one of the most familiar and best-

characterized proteins, hemoglobin, has been considerably slower than had been predicted or anticipated. Progress, although considerable, has been influenced by two major factors. One, clinical understanding of the maintenance of general and local circulatory homeostasis and the effects of oxygen deprivation is expanding and changing rapidly. Two, superimposed on this backdrop of expansion and change is the recognition that the biological and physiological consequences of the infusion of acellular hemoglobin solutions are incompletely understood. Each new report on clinical trials of the various tHbs makes it clear that effective therapeutic use of the current generation of hemoglobins will require a determination of the appropriate balance between benefits and risks, coupled with a careful match of the beneficial properties of a tHb to the clinical indication(s) to which it is best suited. The goal of the trials, however, remains constant. As Sloan et al. stated, the clinical objective is “a reflection of the earnest desire of the investigators to increase the survival chances of patients for whom standard therapy appeared to offer little hope” (94).

REFERENCES

1. Blundell, J. Experiments on the Transfusion of Blood by the Syringe. *Medicochir. Trans.* **1818**, 9, 56–92.
2. Surgenor, D.M.; Wallace, E.L.; Hao, E.S.; Chapman, R.H. Collection and Transfusion of Blood in the United States, 1982–1988; *New Engl. J. Med.* **1990**, 322, 1646–1651.
3. Wallace, E.L.; Churchill, W.H.; Surgenor, D.M.; An, J.; Cho, G.; McGurk, S.; Murphy, L. Collection and Transfusion of Blood and Blood Components in the United States, 1992. *Transfusion* **1995**, 35, 802–812.
4. Hébert, P.C.; Schweitzer, I.; Calder, L.; Blajchman, M.; Giulivi, A. Review of the Clinical Practice Literature on Allogeneic Red Blood Cell Transfusion. *Can. Med. Assoc. J.* **1997**, 156 (Suppl 11), S9–S26.
5. Schreiber, G.B.; Busch, M.P.; Kleinman, S.H.; Korelitz, J.J.; The Risk of Transfusion-Transmitted Viral Infections. *New Engl. J. Med.* **1996**, 334(26), 1685–1690.
6. Peek, C.C.; Moore, G.L.; Bolin, R.B. Adenine in Blood Preservation. *Crit. Rev. Clin. Lab. Sci.* **1981**, 13, 173–212.
7. Simchon, S.; Jan, K.M.; Chien, S. Influence of Reduced Red Cell Deformability on Regional Flow. *Am. J. Physiol.* **1987**, 253, H895–H903.
8. Doyle, M.P.; Walker, B.R. Stiffened Erythrocytes Augment the Pulmonary Hyperdynamic Response to Hypoxia. *J. Appl. Physiol.* **1990**, 69, 1270–1275.
9. Fakhry, S.M.; Sheldon, G.F. Blood Administration, Risks, and Substitutes. *Adv. Surg.* **1995**, 28, 71–92.
10. Tartter, P.I. The Association of Perioperative Blood Transfusion with Colorectal Cancer Recurrence. *Annals of Surgery* **1992**, 216, 633–638.
11. Tartter, P.I.; Burrows, L.; Kirschner, P.; Perioperative Blood Transfusion Adversely Affects Prognosis After

- Resection of Stage I (subset NO) Non-oat Cell Lung Cancer. *J. Thorac. Cardiovasc. Surg.* **1984**, 88, 659–662.
12. Heiss, M.M.; Mempel, W.; Jauch, K.W.; Delanoff, C.; Mayer, G.; Mempel, M.; Eissner, H.J.; Schildberg, F.W. Beneficial Effect of Autologous Blood Transfusion on Infectious Complications after Colorectal Cancer Surgery. *Lancet* **1993**, 342, 1328–1333.
 13. Davidsohn, I. Reflections on the Past, Present, and Future of Blood Transfusion. *Practical Blood Transfusion*; Huestis, D.W., Bove, J.R., Busch, S., Eds.; Little, Brown and Company: Boston, 1969.
 14. Bunn, H.F.; Forget, B.G. *Hemoglobin: Molecular, Genetic and Clinical Aspects*; W.B. Saunders Co.: Philadelphia, 1986.
 15. Winslow, R.M. *Hemoglobin-based Red Cell Substitutes*; Johns Hopkins University Press: Baltimore and London, 1992.
 16. DeVenuto, F. Appraisal of Hemoglobin Solution As a Blood Substitute. *Surg. Gynec. Obstet.* **1979**, 149, 417–436.
 17. Willinger, C.C.; Schramek, H.; Pfaller, K.; Joannidis, M.; Deetjen, P.; Pfaller, W. Ultrapure Polymerized Bovine Hemoglobin Improves Structural and Functional Integrity of the Isolated Perfused Rat Kidney. *Renal Physiol. Biochem.* **1995**, 18, 288–305.
 18. Nosé, Y. Oxygen-Carrying Macromolecules: Therapeutic Agents for the Treatment of Hypoxia. *Art. Organs* **1998**, 22(7), 618–622.
 19. Linberg, R.; Conover, C.D.; Shum, K.L.; Shorr, R.G. Hemoglobin Based Oxygen Carriers: How Much Methemoglobin is Too Much? *Art. Cells Blood Subs. Immob. Biotech.* **1998**, 26, 133–148.
 20. Rabiner, S.F.; Helbert, J.R.; Lopas, H.; Friedman, L.H. Evaluation of a Stroma-free Hemoglobin Solution for Use As a Plasma Expander. *J. Exp. Med.* **1967**, 126(6), 1127–1142.
 21. Kaca, W. Hemoglobin, a Newly Recognized Lipopolysaccharide (LPS)-Binding Protein That Enhances LPS Biological Activity. *J. Biol. Chem.* **1994**, 269(40), 25078–25084.
 22. Bunn, H.F. Hemoglobin Function in Stored Blood. *J. Clin. Invest.* **1969**, 48, 311–321.
 23. Sanders, K.E.; Ackers, G.; Sligar, S. Engineering and Design of Blood Substitutes. *Curr. Opin. Struct. Biol.* **1996**, 6, 534–540.
 24. Looker, D.; Abbott-Brown, D.; Cozart, P.; Durfee, S.; Hoffman, S.; Mathews, A.; Miller-Roehrich, J.; Shoemaker, S.; Trimble, S.; Fermi, G. A Human Recombinant Haemoglobin Designed for Use as a Blood Substitute. *Nature* **1992**, 356, 258–260. Although development of this Hb has been halted, the commercial development of other recombinant Hbs continues at Baxter Healthcare Corp.
 25. Ogden, J.E.; Coghlan, D.; Jones, G.; Denton, K.A.; Harris, R.; Chan, B.; Woodrow, J.; Wilson, M.T.; Expression and Assembly of Functional Human Hemoglobin In *S. Cerevisiae*. *Art. Cells Blood Subs. Immob. Biotech.* **1992**, 20(2–4), 473–475.
 26. Kumar, R.; Manjula, B.N. Human Hemoglobin Derived from Transgenic Swine: A Starting Material for “Blood Substitute” Formulations. *Red Blood Cell Substitutes*, Rudolph, A.S., Rabinovici, R., Feuerstein, G.Z., Eds.; Marcel Dekker, Inc.: New York, 1998; 309–324.
 27. Conover, C.D.; Lejeune, L.; Shum, K.; Gilbert, C.; Shorr, R.G.L. Physiological Effect of Polyethylene Glycol Conjugation on Stroma-free Bovine Hemoglobin in the Conscious Dog After Partial Exchange Transfusion. *Artif. Organs* **1997**, 21(5), 369–378.
 28. Lips, V.; Celedón, G.; Escobar, J.; Lissi, E.A. Thiol-induced Hemoglobin Oxidation. *Redox Report* **1996**, 2(3), 205–212.
 29. Rudolph, A.S. Encapsulation of Hemoglobin in Liposomes. *Blood Substitutes: Physiological Basis of Efficacy*, Winslow, R.M., Vandegriff, K.D., Intaglietta, M., Eds.; Birkhäuser: Boston 1995; 90–104.
 30. Torchilin, V.P. Modification of Molecules and Particles With Polyethylene Glycol (PEG): Long-Circulating Pharmaceuticals. *Advances in Blood Substitutes: Industrial Opportunities and Medical Challenges*, Winslow, R.M., Vandegriff, K.D., Intaglietta, M., Eds.; Birkhäuser: Boston, 1997; 251–297.
 31. Brandl, M.; Gregoriadis, G. Entrapment of Haemoglobin into Liposomes by the Dehydration-rehydration Method: Vesicle Characterization and In Vivo Behavior. *Biochim. Biophys. Acta* **1994**, 1196, 65–75.
 32. Phillips, W.T.; Klipper, R.W.; Awasthi, V.D.; Rudolph, A.S.; Cliff, R.; Kwasiborski, V.; Goins, B.A. Polyethylene Glycol-Modified Liposome Encapsulated Hemoglobin: A Long Circulating Red Cell Substitute. *J. Pharmacol. Exp. Ther.* **1999**, 288, 665–670.
 33. Nakai, K.; Usuba, A.; Ohta, T.; Kuwabara, M.; Nakazato, Y.; Motoki, R.; Takahashi, T.A. Coronary Vascular Bed Perfusion with a Polyethylene Glycol-Modified Hemoglobin-Encapsulated Liposome, Neo Red Cell, in Rats. *Artif. Organs* **1998**, 22(4), 320–325.
 34. Tsuchida, E.; Takeoka, S. Stabilized Hemoglobin Vesicles. *Artificial Red Cells: Materials, Performances and Clinical Study as Blood Substitutes* Tsuchida, E., Ed.; John Wiley & Sons: New York, 1997; 35–64.
 35. Usuba, A.; Motoki, R. Safety and Efficacy of Encapsulated Hemoglobin in Hemorrhagic Shock. *Artificial Red Cells: Materials, Performances and Clinical Study as Blood Substitutes*; Tsuchida, E., Ed.; John Wiley & Sons: New York, 1997; 65–92.
 36. Sakai, H.; Tsai, A.G.; Intaglietta, M.; Tsuchida, E. Hemoglobin Encapsulation with Polyethylene Glycol-modified and Unmodified Vesicles: Systemic and Microvascular Hemodynamics At 80% Blood Substitution. *Advances in Blood Substitutes: Industrial Opportunities and Medical Challenges*; Winslow, R.M., Vandegriff, K.D., Intaglietta, M., Eds.; Birkhäuser: Boston, 1997; 151–166.
 37. Phillips, W.T.; Lemen, L.; Goins, B.; Rudolph, A.S.; Klipper, R.; Fresne, D.; Jerabek, P.A.; Emch, M.E.; Martin, C.; Fox, P.T.; McMahan, C.A. Use of Oxygen-15 to Measure Oxygen-carrying Capacity of Blood Substitutes In Vivo. *Am. J. Physiol.* **1997**, 272, H2492–H2499.
 38. Gabizon, A.; Price, D.C.; Huberty, J.; Bresalier, R.S.; Papahadjopoulos, D. Effect of Liposome Composition and Other Factors on the Targeting of Liposomes to Experimental Tumors: Biodistribution and Imaging Studies. *Cancer Res.* **1990**, 50, 6371–6378.

39. Nelson, D.J. HemAssist: Development and Clinical Profile. *Red Blood Cell Substitutes: Basic Principles and Clinical Applications*; Rudolph, A.S., Rabinovici, R., Feuerstein, G.Z., Eds.; Marcel Dekker, Inc.: New York, 1998; 353–400.
40. Bakker, J.C.; Bleeker, W.K.; DenBoer, P.J.; Biessels, P.T.M.; Agterberg, J.; Rigter, G.; Holthuis, J.C.M.; Mertens, K. Prevention of Side Effects by Hemoglobin Solutions: The Selection of Optimal Test Models, Especially Concerning Thrombogenicity. *Art. Cells Blood Subs. Immob. Biotech.* **1994**, *22*(3), 577–585.
41. Feola, M.; Simoni, J.; Canizaro, P.C. Quality Control of Hemoglobin Solutions. I. The Purity of Hemoglobin Before Modification. *Art. Organs* **1991**, *15*(3), 243–248.
42. Kang, Y.; Luo, R.G. Chromatographic Removal of Endotoxin from Hemoglobin Preparations: Effect of Solution Conditions on Endotoxin Removal Efficiency and Protein Recovery. *J. Chromatog. A* **1998**, *809*, 13–20.
43. Potts, B.J.; Kivens, W.J. Detecting and Removing Porcine Viruses to Prevent Contamination of Biopharmaceutical Raw Materials. *BioPharm* **2000**, *13*(5), 26–31.
44. Shen, T.J.; Ho, N.T.; Zou, M.; Sun, D.P.; Cottan, P.F.; Simplaceanu, V.; Tam, M.F.; Bell, D.A., Jr.; Ho, C. Production of Human Normal Adult and Fetal Hemoglobin in *Escherichia Coli*. *Protein Eng* **1997**, *10*(9), 1085–1097.
45. Nosé, Y. The Oxygenator. *Manual on Artificial Organs*; Mosby: St. Louis, 1973; Vol. 2.
46. Patel, M.J.; Webb, E.J.; Shelbourn, T.E.; Mattia-Goldberg, C.; George, A.J.T.; Zhang, F.; Moore, E.G.; Nelson, D.J. Absence of Immunogenicity of Diaspirin Cross-linked Hemoglobin in Humans. *Blood* **1998**, *91*(2), 710–716.
47. DeAngeles, D.A.; Scott, A.M.; McGrath, A.M.; Korent, V.A.; Rodenkirch, L.A.; Conhaim, R.L.; Harms, B.A. Resuscitation from Hemorrhagic Shock With Diaspirin Cross-linked Hemoglobin, Blood, or Hetastarch. *J. Trauma* **1997**, *42*(3), 406–414.
48. Siegel, J.H.; Fabian, M.; Smith, J.A.; Costantino, D. Use of Recombinant Hemoglobin Solution in Reversing Lethal Hemorrhagic Hypovolemic Oxygen Debt Shock. *J. Trauma* **1997**, *42*(2), 199–212.
49. Kerger, H.; Tsai, A.G.; Saltzman, D.J.; Winslow, R.M.; Intaglietta, M. Fluid Resuscitation with O₂ Vs. Non-O₂ Carriers After 2 H of Hemorrhagic Shock in Conscious Hamsters. *Am. J. Physiol.* **1997**, *272*, H525–H537, (Heart Circ Physiol 41).
50. Conover, C.D.; Linberg, R.; Gilbert, C.W.; Shum, K.L.; Shorr, R.G.L. Effect of Polyethylene Glycol Conjugated Bovine Hemoglobin in Both Top-load and Exchange Transfusion Rat Models. *Artif. Organs* **1997**, *21*(10), 1066–1075.
51. Draft Guidance for Industry (1): Efficacy Evaluation of Hemoglobin- and Perfluorocarbon-based Oxygen Carriers, Sept 8, 1997; U.S. Food and Drug Administration.
52. Gould, S.A.; Moore, E.E.; Moore, F.A.; Haenel, J.B.; Burch, J.M.; Sehgal, H.; Sehgal, L.; DeWoskin, R.; Moss, G.S.; Clinical Utility of Human Polymerized Hemoglobin As a Blood Substitute After Acute Trauma and Urgent Surgery. *J. Trauma* **1997**, *43*(2), 325–332.
53. Gould, S.A.; Moore, E.E.; Hoyt, D.B.; Burch, J.M.; Haenel, J.B.; Garcia, J.; DeWoskin, R.; Moss, G.S. The First Randomized Trial of Human Polymerized Hemoglobin As a Blood Substitute in Acute Trauma and Emergent Surgery. *J. Am. Coll. Surg.* **1998**, *187*, 113–122.
54. Baron J.F. Berridge J. Brichant J.F. Demeyere R. Lamy M. Larbuisson R. Lehot J.J. Parsloe M. Sinclair C.J. Vandermeersch E. The DCLHb Collaborative Study Group, a Randomized Trial to Assess the Efficacy of Diaspirin Cross-linked Hemoglobin (DCLHb) Solution as an Alternative to Blood Transfusion Following Cardiac Surgery ASA Meeting San Diego October 20–22, 1997.
55. Sloan, E.P.; Doenigsberg, M.; Gens, D.; Cipolle, M.; Runge, J.; Mallory, M.N.; Rodman, G., Jr. DCLHb Traumatic Hemorrhagic Shock Study Group. Diaspirin Cross-linked Hemoglobin (DCLHb) in the Treatment of Severe Traumatic Hemorrhagic Shock: A Randomized Controlled Efficacy Trial. *JAMA* **Nov 17, 1999**, *282*(19), 1857–1864.
56. Adamson, J.G.; Bonaventura, B.J.; Er, S.S.; Langlois, S.F.; MacDonald, I.D.; Moore, C.; Pluira, D.H.; Rydall, J.R.; Wicks, D.G.; Wiffen, D.E.; Wojciechowski, P.W.; Wong, L.T.; Jones, R.T. Productions, Characterization, and Clinical Evaluation of Hemolink™, An Oxidized Raffinose Cross-linked Hemoglobin-based Blood Substitute. *Red Blood Cell Substitutes*, Rudolph, A.S., Rabinovici, R., Feuerstein, G.Z., Eds.; Marcel Dekker, Inc.; New York, 1998; 335–352.
57. Kasper, S.-M.; Walter, M.; Grüne, F.; Bischoff, A.; Erasmi, H.; Buzello, W. Effects of a Hemoglobin-based Oxygen Carrier (HBOC-201) on Hemodynamics and Oxygen Transport in Patients Undergoing Preoperative Hemodilution for Elective Abdominal Aortic Surgery. *Anesth. Analg.* **1996**, *83*, 921–927.
58. Standl, T.; Burmeister, M.-A.; Horn, E.-P.; Wilhelm, S.; Knoefel, W.T.; Schulte am Esch, J. Bovine Haemoglobin-based Oxygen Carrier for Patients Undergoing Haemodilution Before Liver Resection. *Brit. J. Anaesth* **1998**, *80*, 189–194.
59. Kasper, S.-M.; Grüne, F.; Walter, M.; Amr, N.; Erasmi, H.; Buzello, W. The Effects of Increased Doses of Bovine Hemoglobin on Hemodynamics and Oxygen Transport in Patients Undergoing Preoperative Hemodilution for Elective Abdominal Aortic Surgery. *Anesth. Analg.* **1998**, *87*, 284–291.
60. Brauer, P.; Standl, T.; Wilhelm, S.; Burmeister, M.-A.; Schulte am Esch, J. Transcranial Doppler Ultrasonography Mean Flow Velocity During Infusion of Ultrapurified Bovine Hemoglobin. *J. Neurosurg. Anesthes* **1998**, *10*(3), 146–152.
61. Wahr, J.A.; Anderson, M.M.; Giacherio, D.A.; Hallock, L.; Gawryl, M.S.; Lansden, C.; Tremper, K.K. The Effects of a Blood-salvaging Device on Blood Containing a Hemoglobin-based Oxygen Carrier, HBOC-201. *J. Cardiothorac. Vasc. Anesthes* **1997**, *11*(1), 10–12.
62. Neya, K.; Lee, R.; Vlahakes, G.J. Hemoglobin Based Oxygen Carrying Solution Stability in Extracorporeal Circulation: An *In Vitro* Evaluation and Implications for Clinical Use. *ASAIO J.* **1998**, *44*, 166–170.

63. McGown, E.L.; Lyons, M.F.; Marini, M.A.; Zegna, A. Reduction of Extracellular Methemoglobin by Erythrocytes. *Biochim. Biophys. Acta* **1990**, *1036*, 202–206.
64. Darling, R.; Roughton, F.J.W. The Effect of Methemoglobin on the Equilibrium Between Oxygen and Hemoglobin. *Am. J. Physiol.* **1942**, *137*, 56–68.
65. Bone, R.C. Sepsis and Its Complications: The Clinical Problem. *Crit. Care Med.* **1994**, *22*, S8–S11.
66. Fitzgerald, R.D.; Martin, C.M.; Dietz, G.E.; Doig, G.S.; Potter, R.F.; Sibbald, W.J. Transfusing Red Blood Cells Stored in Citrate Phosphate Dextrose Adenine-1 for 28 Days Fails to Improve Tissue Oxygenation in Rats. *Crit. Care Med.* **1997**, *25*(5), 726–732.
67. Heneka, M.T.; Löschmann, P.-A.; Osswald, H. Polymerized Hemoglobin Restores Cardiovascular and Kidney Function in Endotoxin-induced Shock in the Rat. *J. Clin. Invest.* **1997**, *99*, 47–54.
68. Fischer, S.R.; Bone, H.G.; Powell, W.C.; McGuire, R.; Traber, L.D.; Traber, D.L. Pyridoxalated Hemoglobin Polyoxyethylene Conjugate Does Not Restore Hypoxic Pulmonary Vasoconstriction in Ovine Sepsis. *Crit. Care Med.* **1997**, *25*(9), 1551–1559.
69. Bone, H.G.; Schenarts, P.J.; Booke, M.; McGuire, R.; Harper, D.; Traber, L.D.; Traber, D.L. Oxalated Pyridoxalated Hemoglobin Polyoxyethylene Conjugate Normalizes the Hyperdynamic Circulation in Septic Sheep. *Crit. Care Med.* **1997**, *25*(6), 1010–1018.
70. DeAngelo, J. Nitric Oxide Scavengers in the Treatment of Shock Associated With Systemic Inflammatory Response Syndrome. *Exp. Opin. Pharmacother* **1999**, *1*(1), 19–29.
71. Reah, G.; Bodenham, A.R.; Mallick, A.; Daily, E.K.; Przybelski, R.J. Initial Evaluation of Diaspirin Cross-linked Hemoglobin (DCLHbTM) as a Vasopressor in Critically Ill Patients. *Crit. Care Med.* **1997**, *25*(9), 1480–1488.
72. Vaupel, P. Oxygenation of Solid Tumors. *Drug Resistance in Oncology*, Teicher, B.A., Ed.; Marcel Dekker, Inc.: New York, 1993; 53–85.
73. Vaupel, P. Oxygenation of Human Tumors. *Strahlenther. Onkol.* **1990**, *166*, 377–386.
74. Teicher, B.A.; Ara, G.; Herbst, R.; Takeuchi, H.; Keyes, S.; Northey, D. PEG-Hemoglobin Effects On Tumor Oxygenation and Response to Chemotherapy In Vivo **1997**, *11*, 301–312.
75. Teicher, B.A.; Dupuis, N.P.; Liu, F.; Robinson, M.F.; Menon, K. Administration of a Hemoglobin Solution Decreases Hypoxia and Increases Radiation Response in a Rat Mammary Carcinoma Model. *Drug Development Res.* **1995**, *34*, 231–240.
76. Teicher, B.; Ara, G.; Chen, Y.; Emi, Y.; Kakeji, Y.; Ikebe, M.; Machara, Y. PEG-Hemoglobin Effects on Tumor Oxygenation and Radiosensitization. *Rad. Oncol. Invest* **1996**, *4*(5), 200–210.
77. Moqattash, S.; Lutton, J.D.; Rosenthal, G.; Abu-Hijleh, M.F.; Abraham, N.G. Effect of Blood Substitute, Recombinant Hemoglobin, on in Vivo Hematopoietic Recovery from AZT Toxicity. *Acta Haematol.* **1997**, *98*, 76–82.
78. Lutton, J.D.; Jiang, S.; Drummond, G.S.; Abraham, N.G.; Levere, R.D.; Kappas, A. Pharmacologic Effects of the Red Blood Cell Substitutes Cross-linked and Non-cross-linked Hemoglobins on Hematopoiesis in Rabbits. *Pharmacol.* **1999**, *58*, 319–324.
79. Cohn, S.M. Blood Substitutes. *New Horiz.* **1999**, *7*(1), 54–60.
80. Bowes, M.P.; Burhop, K.E.; Ziven, J.A. Diaspirin Cross-linked Hemoglobin Improves Neurological Outcome Following Reversible But Not Irreversible CNS Ischemia in Rabbits. *Stroke* **1994**, *25*(11), 2253–2257.
81. Aronowski, J.; Strong, R.; Grotta, J.C. Combined Neuroprotection and Reperfusion Therapy for Stroke. Effect of Lubeluzole and Diaspirin Cross-linked Hemoglobin in Experimental Focal Ischemia. *Stroke* **1996**, *27*(9), 1571–1577.
82. Cole, D.J.; Schell, R.M.; Przybelski, R.J.; Drummond, J.C.; Bradley, K. Focal Cerebral Ischemia in Rats: Effect of Hemodilution With α - α Cross-linked Hemoglobin on CBF. *J. Cereb. Blood Flow Metab.* **1992**, *12*, 971–976.
83. Cole, D.J.; Drummond, J.C.; Patel, P.M.; Marcantonio, S. Effects of Viscosity and Oxygen Content on Cerebral Blood Flow in Ischemic and Normal Rat Brain. *J. Neurol. Sci.* **1994**, *124*, 15–20.
84. Cole, D.J.; Schell, R.M.; Drummond, J.C.; Reynolds, L. Focal Cerebral Ischemia in Rats: Effect of Hypervolemic Hemodilution With Diaspirin Cross-linked Hemoglobin Versus Albumin on Brain Injury and Edema. *Anesthesiology* **1993**, *78*, 335–342.
85. Cole, D.J.; Drummond, J.C.; Patel, P.M.; Nary, J.C.; Applegate, R.L., II. Effect of Oncotic Pressure of Diaspirin Cross-linked Hemoglobin (DCLHbTM) on Brain Injury After Temporary Focal Cerebral Ischemia in Rats. *Anesth. Analg.* **1996**, *83*, 342–347.
86. Hanzawa, K.; Ohzeki, H.; Moro, H.; Eguchi, S.; Nakajima, T.; Makifuchi, T.; Miyashita, K.; Nushiura, M.; Naritomi, H. Effects of Partial Blood Replacement With Pyridoxalated Hemoglobin Polyoxyethylene Conjugate Solution on Transient Cerebral Ischemia in Gerbil. *Art. Cells Blood Subs. Immob. Biotech.* **1997**, *25*(1, 2), 105–114.
87. Medele, R.J.; Stummer, W.; Reulen, H.J.; Steiger, H.J. Evidence for Peroxidative Damage by Nitric Oxide in Experimental Chronic Cerebral Vasospasm. *Neurol. Res.* **1996**, *18*, 277–280.
88. Cole, D.J.; Nary, J.C.; Drummond, J.C.; Patel, P.M.; Jacobsen, W.K. α - α Diaspirin Cross-linked Hemoglobin, Nitric Oxide and Cerebral Ischemic Injury in Rats. *Art. Cells Blood Subs. Immob. Biotech.* **1997**, *25*(1, 2), 141–152.
89. Kassell, N.F.; Torner, J.C.; Hale, E.C., Jr.; Jane, J.A.; Adams, H.P.; Kongable, G.L. The International Cooperative Study on the Timing of Aneurysm Surgery. Part 1: Overall Management Results. *J. Neurosurg.* **1990**, *73*, 18–36.
90. Macdonald, R.L.; Weir, B.K.A. A Review of Hemoglobin and the Pathogenesis of Cerebral Vasospasm. *Stroke* **1991**, *22*, 971–982.
91. Cole, D.J.; McKay, L.; Jacobsen, W.K.; Drummond, J.C.; Patel, P.M. Effect of Subarachnoid Administration of α - α Diaspirin Cross-linked Hemoglobin on Cerebral Blood Flow in Rats. *Art. Cells Blood Subs. Immob. Biotech.* **1997**, *25*(1, 2), 95–104.

92. Saxena, R.; Wijnhoud, A.D.; Man in 't Veld, A.J.; van den Meiracker, A.H.; Boomsma, F.; Przybelski, R.J.; Koudstaal, P.J. Effect of Diaspirin Cross-linked Hemoglobin on Endothelin-1 and Blood Pressure in Acute Ischemic Stroke in Man. *J. Hypertens.* **1998**, *16*, 1459–1465.
93. Arent, R.M.; Schmoekel, M.; Wilbert-Lampen, U.; Plasse, A.; Heucke, L.; Werdan, K. Bidirectional Effects of Endogenous Opioid Peptides on Endothelin Release Rates in Porcine Aortic Endothelial Cell Culture: Mediation by Delta Opioid Receptor Antagonist-Insensitive Mechanisms. *J. Pharmacol. Exp. Ther.* **1995**, *272*, 1–7.
94. Anderson, V.H.; Nelson, D.I.; Leingruber, P.P.; Roubin, G.S.; Gruentzig, A.R. Coronary Artery Perfusion during Percutaneous Transluminal Coronary Angioplasty. *Transfusion Medicine: Recent Technological Advances*; Alan R Liss: New York, 1986; 3–20.
95. Christensen, C.W.; Reeves, W.C.; Lassar, T.A.; Schmidt, D.H.; Inadequate Subendocardial Oxygen Delivery during Perfluorocarbon Perfusion in a Canine Model of Ischemia. *Am. Heart J.* **1988**, *115*, 30–37.
96. McKenzie, J.E.; Cost, E.A.; Scandling, D.M.; Ahle, N.W.; Savage, R.W. Effects of Diaspirin Cross-linked Haemoglobin during Coronary Angioplasty in the Swine. *Cardiovasc. Res.* **1994**, *28*, 1188–1192.
97. Gonzalez, P.; Hackney, A.C.; Jones, S.; Strayhorn, D.; Hoffman, E.B.; Hughes, G.; Jacobs, E.E.; Orringer, E.P. A Phase I/II Clinical Study of Polymerized Bovine Hemoglobin in Adult Patients with Sickle Cell Disease not in Crisis at the Time of Study. *J. Invest. Med.* **1997**, *45*(5), 258–264.
98. Hughes, G.P.; Yancey, E.P.; Albrecht, R.; Locker, P.K.; Francom, S.F.; Orringer, E.P.; Antal, E.J.; Jacobs, E.E. Hemoglobin-based Oxygen Carrier Preserves Submaximal Exercise Capacity in Humans. *Clin. Pharmacol. Ther.* **1995**, *58*, 434–443.
99. Elsayed, N.M.; Gorbunov, N.V.; Kagan, V.E. A Proposed Biochemical Mechanism Involving Hemoglobin for Blast Overpressure-induced Injury. *Toxicology* **1997**, *121*, 81–90.
100. Cohn, S.M.; Zieg, P.M.; Rosenfield, A.T.; Fisher, B.T. Resuscitation of Pulmonary Contusion: Effects of a Red Cell Substitute. *Crit. Care Med.* **1997**, *25*(3), 484–491.
101. Shackford, S.R. Effect of Small-Volume Resuscitation on Intracranial Pressure and Related Cerebral Variables. *J. Trauma* **1997**, *42*(5), S48–S53.
102. Piper, I.R.; Garrioch, M.A.; Souter, M.J.; Andrews, P.J.D.; Thomson, D. Effects of Diaspirin Cross-linked Haemoglobin on Post-traumatic Cerebral Perfusion Pressure and Blood Flow in a Rodent Model of Diffuse Brain Injury. *Brit. J. Anaesthet.* **1998**, *80*, 639–643.
103. Nary, J.; Ling, G.; Lojeski, E.; Muldoon, S. Traumatic Brain Injury in Rats: Histopathologic Effects of Hemodilution with Diaspirin Crosslinked Hemoglobin. *Anesthesiol* **1997**, *87*(3A), A 681.
104. Wallace, W.J.; Houtchens, R.A.; Maxwell, J.C.; Caughey, W.S. Mechanism of Autoxidation for Hemoglobins and Myoglobins. *J. Biol. Chem.* **1982**, *257*(9), 4966–4976.
105. Watkins, J.A.; Kawanishi, S.; Caughey, W.S. Autoxidation Reactions of Hemoglobin a Free from Other Red Cell Components a Minimal Mechanism. *Biochem. Biophys. Res. Commun.* **1985**, *132*(2), 742–748.
106. Ortiz, M.C.; Fortepiani, L.A.; de Rycker, M.; Atucha, N.M.; Romero, J.C.; Garcia-Estañ, J. Pressor and Renal Effects of Cross-linked Hemoglobin in Anesthetized Cirrhotic Rats. *J. Hepatol.* **2000**, *32*, 32–37.
107. Chou, A.C.; Fitch, C.D. Mechanism of Hemolysis Induced by Ferriprotoporphyrin IX. *J. Clin. Invest.* **1981**, *68*, 672–677.
108. Balla, G.; Vercellotti, G.M.; Muller-Eberhard, U.; Eaton, J.; Jacob, H.S. Exposure of Endothelial Cells to Free Heme Potentiates Damage Mediated by Granulocytes and Toxic Oxygen Species. *Lab. Invest.* **1991**, *64*, 648–655.
109. D'Agnillo, F.; Alayash, A.I. Site-Specific Modifications and Toxicity of Blood Substitutes. The Case of Diaspirin Cross-linked Hemoglobin. *Adv. Drug Deliv. Rev.* **2000**, *40*, 199–212.
110. Alayash, A.I. Hemoglobin-based Blood Substitutes: Oxygen Carriers, Pressor Agents, or Oxidants? *Nature Biotech.* **1999**, *17*, 545–549.
111. Alayash, A.I.; Ryan, B.A.B.; McLeod, L.L.; Goldman, D.W.; Cashion, R.E. Cell-free Hemoglobin and Tissue Oxidants: Probing the Mechanisms of Hemoglobin Cytotoxicity. *Blood Substitutes: Principles, Methods, Products and Clinical Trials*; Chang, T.M.S., Ed.; Karger Landes Systems: Basel, 1998; 157–177.
112. Everse, J.; Nelson, H. The Toxicities of Native and Modified Hemoglobins. *Free Rad. Biol. Med.* **1997**, *22*(6), 1075–1099.
113. Caron, A.; Menu, P.; Faivre-Fiorina, B.; Labrude, P.; Alayash, A.I.; Vigneron, C. Cardiovascular and Hemorheological Effects of Three Modified Human Hemoglobin in Hemodiluted Rabbits. *J. Appl. Physiol.* **1999**, *86*(2), 541–548.
114. Loscalzo, J.; Welch, G. Nitric Oxide and the Cardiovascular System. *Prog. Cardiovasc. Dis.* **1995**, *38*, 87–104.
115. Davies, M.G.; Fulton, G.J.; Hagen, P.-O. Clinical Biology of Nitric Oxide. *Brit. J. Surg.* **1995**, *82*, 1598–1610.
116. Takaori, K.; Inoue, K.; Kogire, M.; Higashide, S.; Tun, T.; Aung, T.; Doi, R.; Fujii, N.; Tobe, T. Effects of Endothelin on Microcirculation of the Pancreas. *Life Sci.* **1992**, *51*, 615–622.
117. Kohan, D.E. Endothelins in the Normal and Diseased Kidney. *Am. J. Kidney Dis.* **1997**, *29*(1), 2–26.
118. Kerger, H.; Tsai, A.G.; Saltzman, D.J.; Winslow, R.M.; Intaglietta, M. Fluid Resuscitation with O₂ Vs. Non-O₂ Carriers After 2 H of Hemorrhagic Shock in Conscious Hamsters. *Am. J. Physiol.* **272**, H525–H537.
119. Xu, L.; Sun, L.; Rollwagen, F.M.; Li, Y.; Pacheco, N.D.; Pikoulis, E.; Leppäniemi, A.; Soltero, R.; Burris, D.; Malcolm, D.; Nielsen, T.B. Cellular Responses to Surgical Trauma, Hemorrhage, and Resuscitation with Diaspirin Cross-linked Hemoglobin in Rats. *J. Trauma* **1997**, *42*(1), 32–41.
120. Noone, R.B., Jr.; Mythen, M.G.; Vaslef, S.N. Effect of $\alpha\alpha$ -cross-linked Hemoglobin and Pyridoxalated Hemoglobin Polyoxyethylene Conjugate Solutions on Gastrointestinal Regional Perfusion in Hemorrhagic Shock. *J. Trauma* **1998**, *45*(3), 457–469.
121. Von Dobschuetz, E.; Hoffmann, T.; Messmer, K. Diaspirin Cross-linked Hemoglobin Effectively Restores

- Pancreatic Microcirculatory Failure in Hemorrhagic Shock. *Anesthesiol.* **1999**, *91*, 1754–1762.
122. Schultz, S.C.; Powell, C.C.; Bernard, E. Diaspirin Cross-linked Hemoglobin (DCLHbTM) Attenuates Bacterial Translocation in Rats. *Art. Cells Blood Subs. Immob. Biotech.* **1995**, *23*, 647.
123. Bone, H.G.; Schenarts, P.J.; Fischer, S.R.; McGuire, R.; Traber, L.D.; Traber, D.L. Pyridoxalated Hemoglobin Polyoxyethylene Conjugate Reverses Hyperdynamic Circulation in Septic Sheep. *J. Appl. Physiol.* **1998**, *84*(6), 1991–1999.
124. Gulati, A.; Sharma, A.C.; Singh, G. Role of Endothelin in the Cardiovascular Effects of Diaspirin Cross-linked and Stroma-reduced Hemoglobin. *Crit. Care Med.* **1996**, *24*(1), 137–147.
125. Loeb, A.; McIntosh, L.J.; Raj, N.R.; Longnecker, D. Regional Vascular Effects of RHB 1.1, a Hemoglobin-based Oxygen Carrier. *J. Cardiovasc. Pharm.* **1997**, *30*, 703–710.
126. Viele, M.K.; Weiskopf, R.B.; Fisher, D. Recombinant Human Hemoglobin does not Affect Renal Function in Humans: Analysis of Safety and Pharmacokinetics. *Anesthesiol.* **1997**, *86*(4), 48–58.
127. Murray, J.A.; Ledlow, A.; Launspach, J.; Evans, D.; Loveday, M.; Conklin, J.L. The Effects of Recombinant Human Hemoglobin on Esophageal Motor Function in Humans. *Gastroenterol.* **1995**, *109*, 1241–1248.
128. For Example, See Refs. (41) and (42).
129. Takaori, K.; Inoue, K.; Kogire, M.; Higashide, S.; Tun, T.; Aung, T.; Doi, R.; Fujii, N.; Tobe, T. Effects of Endothelin on Microcirculation of the Pancreas. *Life Sci.* **1992**, *51*, 615–622.
130. Gburek, J.; Zabel, M.; Osada, J. Immunohistochemical Localization of Haemoglobin Binding Sites in the Distal Tubule of the Rat Kidney. *Histochem. J.* **1998**, *30*, 421–424.
131. Lorente, J.A.; Landin, L.; Renes, E.; DePablo, R.; Jorge, P.; Rodena, E.; Liste, D. Role of Nitric Oxide in the Hemodynamic Changes of Sepsis. *Crit. Care Med.* **1993**, *21*, 759–767.
132. Thiernemann, C.; Szabo, C.; Mitchell, J.A.; Vane, J.R. Vascular Hyporeactivity to Vasoconstrictor Agents and Hemodynamic Decompensation in Hemorrhagic Shock is Mediated by Nitric Oxide. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 267–271.
133. Petros, A.; Lamb, G.; Leone, A.; Moncada, S.; Bennett, D.; Valance, P. Effects of a Nitric Oxide Synthase Inhibitor in Humans With Septic Shock. *Cardiovasc. Res.* **1994**, *28*, 34–39.
134. Bush, S.; Marshall, T.; Spicuzza, J.; Nelson, D. Diaspirin Cross-linked Hemoglobin (DCLHbTM): Bioanalytical Studies in Swine. *Art. Cells Blood Subs. Immob. Biotech.* **1994**, *22*(3), 917–922.
135. Sibbald, W.J.; Sielenkamper, A. Blood Substitutes—Effects on the Microcirculation. *Update Intensive Care Emerg. Med.* **1998**, *33*, 318–331.
136. Pincemail, J.; Detry, O.; Philippart, C.; Defraigne, J.O.; Franssen, C.; Burhop, K.; Deby, C.; Meurisse, M.; Lamy, M. Diaspirin Crosslinked Hemoglobin (DCLHbTM): Absence of Increased Free Radical Generation Following Administration in a Rabbit Model of Renal Ischemia and Reperfusion. *Free Rad. Biol. Med.* **1995**, *19*(1), 1–9.
137. Bratlid, D. Bilirubin Toxicity: Pathophysiology and Assessment of Risk Factors. *New York State J. Med.* **1991**, *91*(11), 489–492.
138. McDounagh, A.F. Is Bilirubin Good for You? *Clin. Perinatol.* **1990**, *17*, 359–369.
139. Tithof, P.; Estep, T.; Burhop, K. Diaspirin Cross-linked Hemoglobin: Safety and Pre-clinical Profile of a Novel Hemoglobin Therapeutic in Preparation.
140. Vemulapalli, S.; Chiu, P.J.S.; Griscti, K.; Brown, A.; Kurowski, S.; Sybertz, E.J. Phosphoramidon Does Not Inhibit Endogenous Endothelin-1 Release Stimulated by Hemorrhage, Cytokines and Hypoxia in Rats. *Eur. J. Pharmacol.* **1994**, *257*, 95–102.
141. Chang, H.; Wu, G.-J.; Wang, S.-M.; Hung, C. Plasma Endothelin Level Changes During Hemorrhagic Shock. *J. Trauma* **1993**, *35*, 825–833.
142. Hlaváček, J.; Marcof, R. Biological Function and Chemistry of Endothelin. A Review. *Collect Czech. Chem. Commun.* **1999**, *64*, 1211–1252.
143. Gulati, A.; Sen, A.P.; Sharma, A.C.; Singh, G. Role of ET and NO in Resuscitative Effect of Diaspirin Cross-linked Hemoglobin after Hemorrhage in Rat. *Am. J. Physiol.* **1997**, *273*, H817–H836.
144. Ledvina, M.A.; Hart, J.; Bina, S.; Jing, M.; Muldoon, S. Endothelin Plays a Role in Contractions of Isolated Pig Pulmonary Vessels Induced by Diaspirin Cross-linked Hemoglobin. *J. Lab. Clin. Med.* **1999**, *133*, 478–487.
145. Gulati, A.; Singh, G.; Rebello, S.; Sharma, A.C. Effect of Diaspirin Cross-linked and Stroma-reduced Hemoglobin on Mean Arterial Pressure and Endothelin-1 Concentration in Rats. *Life Sci.* **1995**, *56*, 1433–1442.
146. Gulati, A.; Sharma, A.C.; Singh, G. Role of Endothelin in the Cardiovascular Effects of Diaspirin Cross-linked and Stroma Reduced Hemoglobin. *Crit. Care Med.* **1996**, *24*, 137–147.
147. Schultz, S.C.; Grady, B.; Cole, F.; Hamilton, I.; Burhop, K.; Malcolm, D.S. A Role for Endothelin and Nitric Oxide in the Pressor Response to Diaspirin Cross-linked Hemoglobin. *J. Lab. Clin. Med.* **1993**, *122*, 301–308.
148. Warshaw, A.L.; O'Hara, P.J. Susceptibility of the Pancreas to Ischemic Injury in Shock. *Ann. Surg.* **1978**, *188*, 197–201.
149. Hasdai, D.; Komowski, R.; Battler, A. Endothelin and Myocardial Ischemia. *Cardiovasc. Drugs. Ther.* **1994**, *8*, 589–599.
150. Wieszczored, I.; Haynes, W.G.; Webb, D.J.; Ludlam, C.A.; Fox, K.A. Raised Plasma Endothelin in Unstable Angina and Non-Q Wave Myocardial Infarction: Relation to Cardiovascular Outcome. *Br. Heart J.* **1994**, *72*, 436–441.
151. McMurray, J.J.; Ray, S.G.; Abdullah, I.; Dargie, H.J.; Morton, J.J. Plasma Endothelin in Chronic Heart Failure. *Circulation* **1992**, *85*, 1374–1379.
152. Godfraind, T.; Salomone, S. Calcium Antagonists and Endothelial Function: Focus on Nitric Oxide and Endothelin. *Cardiovasc. Drugs Ther.* **1996**, *10*(4), 439–446.
153. Davis, J.H.; Yull, A.B. A Possible Toxic Factor in Abdominal Injury. *J. Trauma* **1961**, *2*, 291–300.

154. Eaton, J.R. Haptoglobin: A Natural Bacteriostat. *Science* **1982**, *215*, 691–692.
155. Lee, J.T., Jr.; Ahrenholz, D.H.; Nelson, R.D.; Simmons, R.L.; Mechanisms of the Adjuvant Effect of Hemoglobin in Experimental Peritonitis. V. The Significance of the Coordinated Iron Component. *Surgery* **1979**, *86*(1), 41–48.
156. Baldwin, A. Blood Substitutes and the Intestinal Microcirculation: Extravasation and Ultrastructural Alterations. *Advances in Blood Substitutes: Industrial Opportunities and Medical Challenges*; Winslow, R.M., Vandegriff, K.D., Intaglietta, M., Eds.; Birkhäuser: Boston, 1997; 19–37.
157. Ziegler, T.R.; Smith, R.J.; O'Dwyer, S.T.; Demling, R.H.; Wilmore, D.W. Increased Intestinal Permeability Associated with Infection in Burn Patients. *Arch. Surg.* **1988**, *123*, 1313–1319.
158. Langkamp-Henken, B.; Donovan, T.B.; Pate, L.M.; Maull, C.D.; Kudsk, K.A. Increased Intestinal Permeability Following Blunt and Penetrating Trauma. *Crit. Care Med.* **1995**, *23*, 660–664.
159. Riddington, D.W.; Venkatesh, B.; Boivin, C.M. Intestinal Permeability, Gastric Intramucosal PH, and Systemic Endotoxemia in Patients Undergoing Cardiopulmonary Bypass. *J. Am. Med. Assoc.* **1996**, *275*, 1007–1012.
160. Deitch, E.A. The Role of Intestinal Barrier Failure and Bacterial Translocation in the Development of Systemic Infection and Multiple Organ Failure. *Arch. Surg.* **1990**, *125*, 403–404.
161. Doglio, G.R.; Pusajo, J.F.; Egurrola, M.A. Gastric Mucosal PH as a Prognostic Index of Mortality in Critically Ill Patients. *Crit. Care Med.* **1991**, *19*, 1037–1040.
162. Ivatury, R.R.; Simon, R.J.; Islam, S.; Fuego, A.; Rohman, M.; Stahl, W.M. A Prospective Randomized Study of End Points of Resuscitation after Major Trauma: Global Oxygen Transport Indices Versus Organ-specific Gastric Mucosal PH. *J. Am. Coll. Surg.* **1996**, *183*, 145–154.
163. Salzman, A.I.; Wang, H.; Wollert, P.S. Endotoxin-Induced Ileal Mucosal Hyperpermeability in Pigs: Role of Tissue Acidosis. *Am. J. Physiol.* **1994**, *266*, G633–G646.
164. Bark, T.; Katouli, M.; Ljungqvist, O. Bacterial Translocation after Non-Lethal Hemorrhage in the Rat. *Circ. Shock* **1993**, *41*, 60.
165. Fitzgerald, R.D.; Martin, C.M.; Dietz, G.E.; Doig, G.S.; Potter, R.F.; Sibbald, W.J. Transfusing Red Blood Cells Stored in Citrate Phosphate Dextrose Adenine-1 for 28 Days Fails to Improve Tissue Oxygenation in Rats. *Crit. Care Med.* **1997**, *25*(5), 726–732.
166. Xu, L.; Sun, L.; Rollwagen, F.M.; Li, Y.; Pacheco, N.D.; Pikoulis, E.; Leppäniemi, A.; Soltero, R.; Burris, D.; Malcolm, D.; Nielsen, T.B. Cellular Responses to Surgical Trauma, Hemorrhage, and Resuscitation with Diaspirin Cross-linked Hemoglobin in Rats/Trauma **1997**, 32–41.
167. Langermans, J.A.M.; Bleeker, W.K. Haemoglobin-based Blood Substitutes and Infection. *Lancet* **1995**, *345*, 863–864.
168. Sinassappel, M.; Ince, C. Low Volume Resuscitation With a Hemoglobin-Based Oxygen Carrier Following Hemorrhage Improves Gut Microvascular Oxygenation in Pigs. *J. Appl. Physiol.* **1996**, *81*, 2297–2303.
169. Bornside, G.H.; Bouis, P.J.; Cohn, L., Jr. Enhancement Of *E. Coli* Infection and Endotoxic Activity by Hemoglobin and Ferric Ammonium Citrate. *Surgery* **1970**, *68*, 350–355.
170. Griffiths, E.; Cortes, A.; Gilbert, N.; Stevenson, P.; MacDonald, S.; Pepper, D. Haemoglobin-based Blood Substitutes and Sepsis. *Lancet* **1995**, *345*, 158–160.
171. Su, D.; Roth, R.I.; Yoshida, M.; Levin, J. Hemoglobin Increases Mortality from Bacterial Endotoxin. *Infect. Immun.* **1997**, *65*(4), 1258–1266.
172. Langermans, J.A.M.; van Vuren-van der Hulst, M.E.B.; Bleeker, W.K. Safety Evaluation of a Polymerized Hemoglobin Solution in a Murine Infection. *Model. J. Lab. Clin. Med.* **1996**, *127*, 428–434.
173. Mourelatos, M.G.; Enzer, N.; Ferguson, J.L. The Effects of Diaspirin Cross-linked Hemoglobin in Sepsis. *Shock* **1996**, *5*, 141–148.
174. Sielenkämper, A.W.; Chin-ye, I.H.; Martin, C.M.; Sibbald, W.J. Diaspirin Cross-linked Hemoglobin Improves Systemic Oxygen Uptake in Oxygen-supply-dependent Septic Rats. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 1066–1072.
175. Bone, H.G.; Booke, M.; Schenarts, P.J.; O'Connor, M.; Traber, L.D.; Traber, D.L. Comparison of Different Doses of Pyridoxalated Hemoglobin Polyoxyethylene Conjugate in the Treatment of Hyperdynamic Sepsis. *Crit. Care Med.* **1996**, *24* (Suppl), A27.
176. Birnbaum, M.L.B.; Lipman, J.; Garrioch, M. Diaspirin Cross-linked Hemoglobin (DCLHb) Vs. Blood in Acutely Anemic Patients. *Int. Care Med.* **1997**, *23* (Suppl. 1), S85, Abstract 263.
177. Bunn, H.F.; Forget, B.G. *Hemoglobin: Molecular, Genetic and Clinical Aspects*; W.B. Saunders Co: Philadelphia, 1986.
178. Rudolph, A.S.; Rabinovici, R.; Feuerstein, G.Z., Eds. *Red Blood Cell Substitutes: Basic Principles and Clinical Applications*; Marcel Dekker, Inc.: New York, 1997.
179. Nelson, D.J. Blood Substitutes: Hemoglobin-based Oxygen Carriers. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 2000; Vol. 19, 1–33.

CAPSULES, HARD

Brian E. Jones

Shionogi Qualicaps, Alcobendas, Spain

INTRODUCTION

Hard or two-piece capsules have been produced since the 1840s (1). They were initially proposed for filling with oils and later were used for powders and other dry forms. The first patent was issued in France; however, the first industrial-scale manufacture took place in the United States in the 1870s. By the turn of the 20th century, millions of capsules per year were being filled in U.S. pharmacies because they met the need for an easily manufactured unit dosage form. Even though the use of capsules spread to other countries, they were considered an American specialty, and the manufacture of capsules did not spread to Europe and the rest of the world until the 1950s. The use of hard capsules increased greatly at this time because of the types of medicine such as antibiotics and antidepressants that were in vogue. This increase stimulated the production of fully automatic filling machines and self-locking capsules to meet the demands for high-speed industrial production. Hard capsules are welcomed by consumers because of their elegant appearance and shape, which is easy to swallow. The majority of capsule fills are dry powder blends, which are typically simple mixtures. The processing and filling of materials involve minimum stress and are the reasons why products are presented in this form. The formulator is able to prepare products that have the desired release characteristics, rapid, controlled, or modified, because of the limited number of factors involved. Hard capsules can be filled with formulations that have a wide range of physical properties from dry solids to nonaqueous solutions, thus enabling the formulator to use many different types of excipients to achieve the desired effect. Capsule products can be formulated to release their active ingredients at many sites along the gastrointestinal tract and to deliver them to the lungs.

RAW MATERIALS

Gelatin and Alternative Materials

Gelatin is a material derived from collagen, a natural protein, which is a fibrous material that occurs in the skin,

bones, and connective tissues of animals (2). It is insoluble in water and is solubilized by hydrolysis. The raw materials used for its manufacture are obtained primarily from bovine bones or porcine skins. The reaction can be carried out at an acid pH level, yielding a type A gelatin (which is primarily produced from skins) and at a basic pH level giving a type B gelatin (which is primarily produced from bovine bones). Gelatin is a heterogeneous product that is a mixture of molecular species, α -, β -, and γ -peptides. The proportions and molecular weights are dependent on the nature of the chemical process. Gelatin was the first material used for the manufacture of capsules because of its unique properties. It is ideal because it is edible, soluble at body temperature, forms strong thin films, and undergoes a gelation process at temperatures just above ambient (1). It has a few technical drawbacks. It is of animal origin, and thus there are certain religious or dietary restrictions on its use. During the 1990s, there had been concerns over the use of bovine materials because of Bovine Spongiform Encephalopathy, which originated in the United Kingdom. The situation in Europe was brought under control by the European Commission (EC), which brought into effect rules that reduced risks to a minimum (3). The parts of the animal that theoretically could be the most infectious, such as the brain, are removed at slaughter and excluded from further processing. All the animals, from which bones are used in gelatin manufacture in Europe, are subjected to pre- and postmortem veterinary inspections. The alkaline hydrolysis process was made the method of choice for bones by the EC because of the pH levels and temperatures used in the process. There are a limited number of producers of pharmaceutical-grade gelatin worldwide thus enabling the problem to be addressed efficiently.

Hard capsules have traditionally been made from gelatin using a dipping process. Gelatin is very well-suited to this because it is an excellent film former and changes in state from liquid to solid at temperatures just above ambient (2). The film produced is homogenous and very robust, and gelatin capsules can readily withstand the mechanical stresses of the filling and packaging operations. The primary drawback in the use of gelatin is that it contains water, which acts as a plasticizer to the film. Thus, if they are not stored properly, their properties

will change. When water is lost from the shells, they become brittle, and thus they are not suitable for hygroscopic materials. Moisture-labile substances cannot be filled into them. For certain markets, there are consumer requirements for a capsule of vegetable origin. Since the last century, people have been searching for gelatin alternatives. The primary problem to be overcome has been the need to obtain a system that gels in a manner similar to that of gelatin so that the same manufacturing process and machines can be used. In the 1950s, capsules of methylcellulose were produced; however, their use was soon discontinued because of poor *in vivo* solubility (1). More recently, hard capsules have been produced from hydroxypropyl methylcellulose (HPMC), either by using a modified production process with heated mold pins (4a, 4b, 4c) or by the use of additives to make a true gelling system (5a, 5b, 5c, 5d). Hard capsules made from HPMC have similar but different properties from gelatin capsules. Their primary advantage is that their moisture content is much lower, and even if this is removed, they retain their mechanical strength (6).

CAPSULE MANUFACTURE

Gelatin capsules are manufactured using a dipping process, which was first proposed by M. Lehuby in 1846 (1). The same process is in use today, but it has been improved and automated. It requires a large investment in machinery and a plant, which need to be operated on a continuous basis to make the process economic. This has resulted in capsules being made by only a limited number of companies worldwide.

The process starts by the preparation of a concentrated solution of gelatin in hot demineralized water. This solution is subjected to a low pressure to remove entrapped air bubbles. Small aliquots (20–30 L) of this solution are taken. To this colorants, either solutions of soluble dyes or suspensions of pigments, process aids such as sodium lauryl sulfate solution, and water to adjust the viscosity are added. The final solution has a concentration of 25–30 wt% of gelatin. This solution is then delivered to the capsule-manufacturing machine.

The manufacturing machines are housed in rooms supplied with filtered air, conditioned to 40–45% relative humidity (RH) and 22–25°C. The most commonly used machines are approximately 12 m long and 3 m high and are divided down the midline into two parts that are mirror images of each other (Fig. 1). The machines are divided lengthways into two levels, an top and a bottom. The caps are made on the left side and the bodies on the

right side of the machine. The gelatin solutions are kept in temperature-controlled, jacketed hoppers (Fig. 1A). From there, they are fed to the capsule-forming container, variously known as a dip pan or dip pot (Fig. 1B). This is a stainless steel jacketed vessel that is oblong with a box in the center. The gelatin solution is pumped into the box and overflows the edge, thus maintaining a constant height. The stainless steel mold pins are mounted in a row of 30 on mild steel bars. Sets of bars are held in a device operated by a cam, which raises and lowers them. The mold pins, which are at 22°C, are lowered into the gelatin solution, which is at 50–55°C. The gelatin immediately gels on the mold. The molds are slowly raised, and, as they do, the excess gelatin runs off (Fig. 2). The quantity picked up by the mold is proportional to the viscosity. The higher the viscosity of the gelatin solution, the more gelatin is picked up. Thus, the viscosity of the solution is used to control the thickness of the gelatin film. As the mold breaks the surface, a blob of gelatin forms on the end of the mold. The sets of bars are transferred from the bottom to the top level of the machine and, as they do so, the bars are rotated to spread the film evenly over the end of the mold pin. The gelatin film is completely set by the time the molds reach the top level of the machine. Sets of bars are grouped together and mechanically transferred through a series of drying kilns (Fig. 1C). In these, air at controlled temperature and humidity is blown over them. When the bars reach the end of the machine, they are transferred to the bottom level and pushed back toward the front of the machine. When the bars emerge from the drying kilns, the moisture content of the gelatin films has been reduced from 70% at dipping to approximately 16 wt%. The molds, which had been warmed at the start of the drying process, have returned to ambient machine room temperature.

The dried gelatin films are removed from the mold pins and cut to the correct lengths, and the cap and body pieces are joined together. This is done in the automatic section of the machine (Fig. 1D). Pairs of bars, one with bodies and one with caps, are passed into the central section. Metal jaws pull the films off the mold pins into collets, which grip them. The collets rotate against a knife, and the gelatin film is cut. The excess is sucked away and recycled. The two pieces are transferred to a central joining block and are closed to a set length, called the unclosed joined length. The capsules are not fully closed because the filling machines would have difficulty separating them. They are closed so that the “prelock” indentations on the cap are engaged by the body, which provides sufficient holding strength so that they will not separate in handling.

The manufacture of HPMC capsules uses a similar process, and there are two different approaches. The viscosity of HPMC solutions increases with increasing

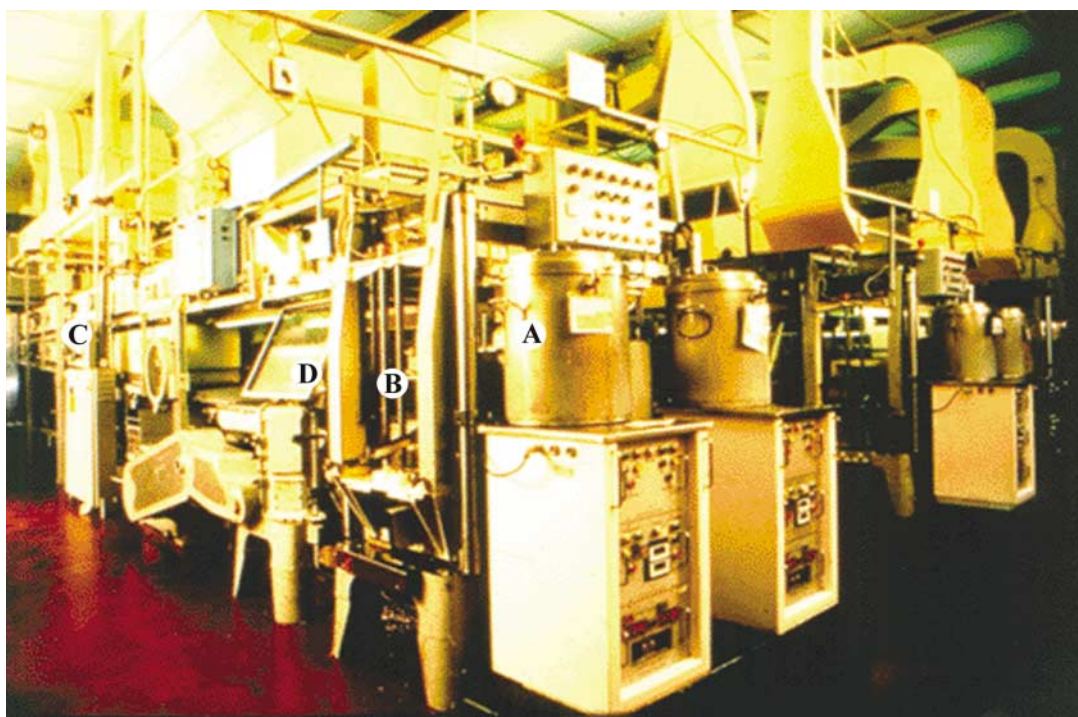


Fig. 1 A capsule manufacturing machine: A) Gelatin solution storage tank; B) dip pan; C) drying kilns; D) automatic section. (From Refs. 5a, 5b, 5c, 5d.)

temperatures. Thus, if hot mold pins at $>60^{\circ}\text{C}$ are lowered into HPMC solutions at $20\text{--}25^{\circ}\text{C}$ material will gel on the molds. The manufacturing machines have to be modified to carry out this process (4a, 4b, 4c). The mold pins are also heated post dipping to hold the material in place before drying. Capsules made in this manner have a wall thickness twice that of standard gelatin capsules to give them sufficient mechanical strength for handling. In the other process using a gelling system, an HPMC solution is prepared that contains carrageenan USNF as a gelation aid and potassium chloride as a gelation promoter (5a, 5b, 5c, 5d). Capsules are manufactured in the same manner and to the same dimensions as gelatin capsules. The only difference in the process is that the rate of output is slower because the rate of gelation is slower than that of gelatin.

STANDARDS FOR EMPTY CAPSULES

Two sets of standards are set for empty capsules: analytical and functional (6). Capsules, like all other pharmaceutical preparations, must comply with good manufacturing practice (GMP) norms and must be made of materials that comply with pharmacopeial chemical and

microbiological standards. However, these tests do not indicate whether a capsule will run well on a filling machine. A series of functional tests are applied by the manufacturers. The critical dimensions of a capsule (the lengths and diameters of the caps and bodies) are checked. It is a continuous production process, and there will be a very small proportion of visually defective capsules. Standard statistical sampling methods are used to estimate quality from samples. The manufacturers and users agree on acceptable quality levels (AQL). The faults are divided into categories depending on the likely impact on capsule performance or the filling process. A different AQL is assigned to each category of fault.

STORAGE OF EMPTY CAPSULES

Empty gelatin capsules are designed to have a moisture content between 13 and 16%. The water acts as a plasticizer and is essential to maintain the flexibility and strength of the film. If falls below the limit, the capsules will become brittle; if its rises above the limit, they will soften. Empty HPMC capsules have a moisture content of 3–7%. These capsules can be dried to less than 1% moisture without losing their mechanical strength and

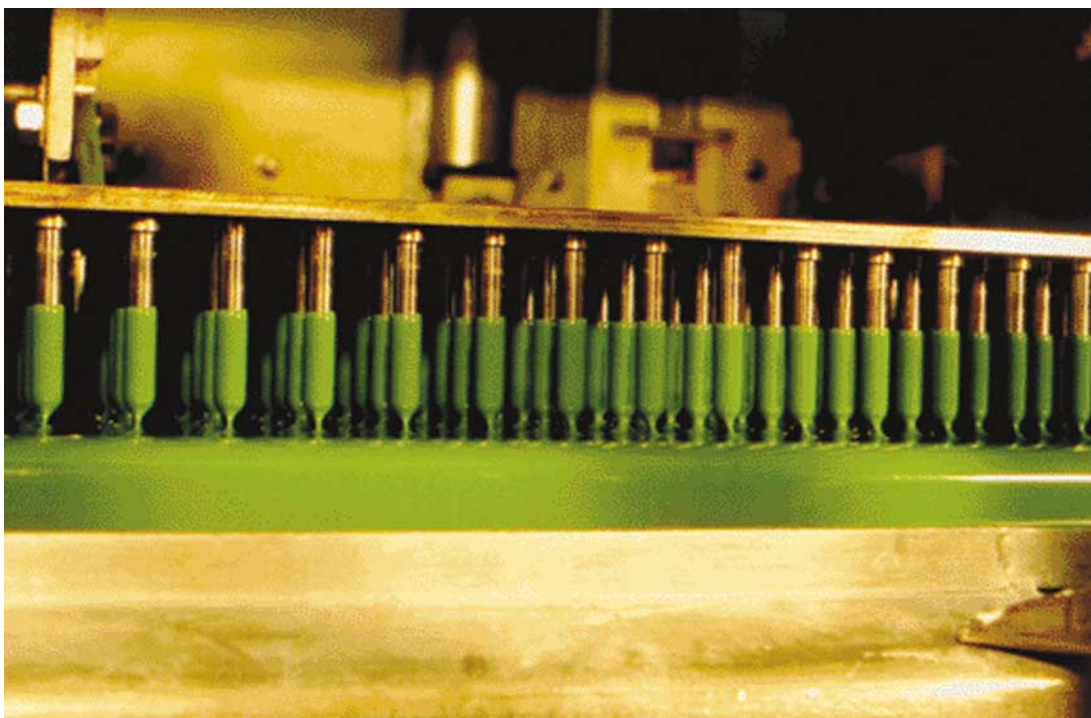


Fig. 2 Dip pan, capsule shell formation: (From Refs. 5a, 5b, 5c, 5d.)

becoming brittle (7). The water in hard capsules is tightly bound into the polymer structure and is insufficient for active bacterial growth (2).

The water content influences capsule dimensions as well. The dimensions vary slightly as moisture is lost or gained. As a general rule, the dimensions of gelatin capsules change by 0.5% for every 1% change in moisture in the range of 13.0–16.0% (1). HPMC capsules contain less water than do gelatin capsules, and the change in dimensions with moisture content is less. Thus, higher-speed filling machines must be operated in air-conditioned areas to achieve maximum performance. The moisture content of capsules depends on the conditions to which they are exposed (8, 9). Water will be lost or gained, and the absorption/desorption isotherm follows a marked hysteresis. In practice, this means that if capsules lose excessive amounts of water, they will not fully rehydrate when exposed to standard conditions, RH levels between 35 and 50%.

CAPSULE FILLING

The hard two-piece capsule can be filled with materials that have a wide range of physical properties. The types of formulations that have been filled into capsules are shown

in Table 1. This is possible because of the manner in which filling machines handle empty capsules. First, capsules are orientated so that they are all pointing in the same direction, with the body downward. The capsules are transferred into pairs of bushes: the opening in the base of the top bush only allows the passage of the body, thus trapping the cap. The body is separated from the cap by means of suction. The open end of the body is then presented to a dosing mechanism and material transferred into it. The cap is then replaced on the body and the capsule closed to the correct closed joined length. This is an important dimension because it ensures that the self-locking mechanism, a series of indentations on the cap and body, is engaged correctly. This allows filled capsules to be transported and packaged on automatic equipment without separating.

Filling machines are differentiated by the means by which they measure the dose of material. They are available with a range of outputs, from bench scale to high-output industrial scale and from manual to fully automatic.

Powder Filling

The majority of formulations that are filled into capsules are dry powder mixtures. The methods of measuring the

Table 1 Formulation types for filling into hard capsules

Dry solids	Semisolids	Liquids
Powders	Thixotropic mixtures	Oily liquids
Granules	Thermosoftening mixtures	Nonaqueous solutions and suspensions
Pellets	Pastes	
Tablets		
Capsules		

dose can be divided into two groups: dependent and independent. The dependent machines use the capsule body directly to measure the dose of powder, whereas the independent machines use a separate device. The literature available on the mechanics of capsule-filling is limited compared with that available for tableting. Part of the reason for this is that tablets, unlike capsules, are used in a wide range of industries outside the healthcare sector, and thus there have been many more workers in the field.

The first industrial filling machines were of the dependent type. Powder is transferred from a hopper directly to the capsule body. The flow of the powder is aided either by a revolving auger or by a vibrating plate. The powder mass inside these capsules is a loose fill. The fill weights achievable on these machines is often higher than that obtained on automatic independent-type machines because the body is overfilled, and thus the total internal volume of the capsule shell is used (10). The first successful industrial filling machine was the ubiquitous machine designed by the doyen of pharmaceutical engineering, Arthur Colton, and the most popular version was the Model No. 8. This is a semiautomatic auger-filling machine (Fig. 3). The empty capsules are fed, aided by suction, into a pair of doughnut-shaped plates, which separate them. The plate containing the bodies is transferred manually to a turntable, and the powder hopper is pulled over the top of it. Powder is forced by the auger into the bodies as their plate revolves under the hopper. The fill weight is controlled by the speeds of rotation of the turntable and the auger. The only way to achieve good uniformity of fill weight on these machines is to completely fill the bodies. Partial filling is not an option. More recently, modifications have been made to the design to bring them in line with GMP requirements. The first was the Quali-fillTM Model 8S developed by Eli Lilly & Co. in the 1980s. Since then, additional improvements have been made: the Quali-fill Model 10 from Schaefer Technologies, Inc. and the Ultra 8 II[®] from Capsugel. A fully automatic rotary auger-filling machine has been developed by Shionogi Qualicaps, the LIQFIL^{super} JCF40/80TM. This machine has a three-roller system for capsule orientation

and continuous feeding to a revolving disk assembly where the capsules are separated, filled, and rejoined. Other automatic dependent filling machines use vibration to aid powder flow into the bodies. The first such machine, the Osaka OCF 120, used a vibrating plate in the powder hopper. The bodies in their holders passed under the hopper. The cavities with the bodies were overfilled and raised up against a metal plate to push the excess inside powder inside. This system works well with dense, free-flowing material. Shionogi Qualicaps have made a modification to this method, the LIQFIL^{super} 40 and 80TM, which uses spring-loaded fingers to compress the powder into the capsule body after it has been filled using a vibrating plate system.

Most automatic machines used in industry are of the independent type and compress a measured amount of powder to form a plug. There are two types of mechanisms: the dosing tube (or dosator) and the dosing disk and tamping finger.

The dosing tube is the most widely used, and it originated in Italy. Current manufacturers are IMA (Zanazi & Farmatic) (IMA North America, Inc.), MG2 (MG America, Inc.), Macophar (Romaco, Inc.), and Bonapace. The plug is formed inside a tube with a moveable piston that controls the dosing volume and applies a force to form the plug (Fig. 4). The lower-output machines have an intermittent motion, whereas the higher-output machines are rotary. The intermittent machines

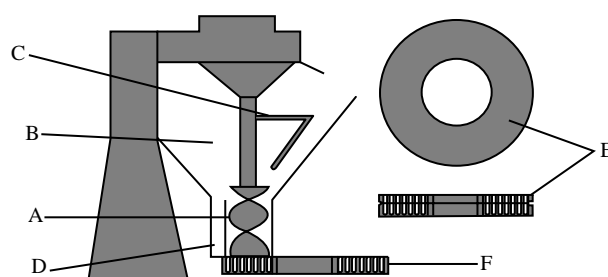


Fig. 3 Schematic diagram of auger filling system (Model No. 8): A) Auger; B) powder hopper; C) stirrer arm; D) pressure relief hole; E) capsule carrying rings; F) body ring holder.

tend to apply greater stresses to the powder than does the rotary machines because there is less time to form the plug, and thus formulations tend to require a higher level of lubricant. These machines are very versatile because the fill weight can be varied over a wide range by a simple adjustment to the position of the piston. The rotary machines can be linked to weighing devices and have automatic weight-control adjustment, which allows them to operate unattended.

The dosing disk and tamping finger machines form a plug in a similar but different manner. They are produced

by a number of companies: Bosch (T.L. Systems Corp.), A.W. Bohanan Co., and Index Manufacturing Co., Inc. The dosing disk, which forms the base of the powder hopper, has up to six sets of machined holes (Fig. 5). In a holder, above the powder hopper, there are sets of stainless steel tamping fingers corresponding to the holes in the disk. These machines have an intermittent motion. After the machine has indexed, and the turret is stationary, the tamping fingers are lowered into the powder bed. The fingers are set to different levels, and they penetrate into the plate and consolidate the powder

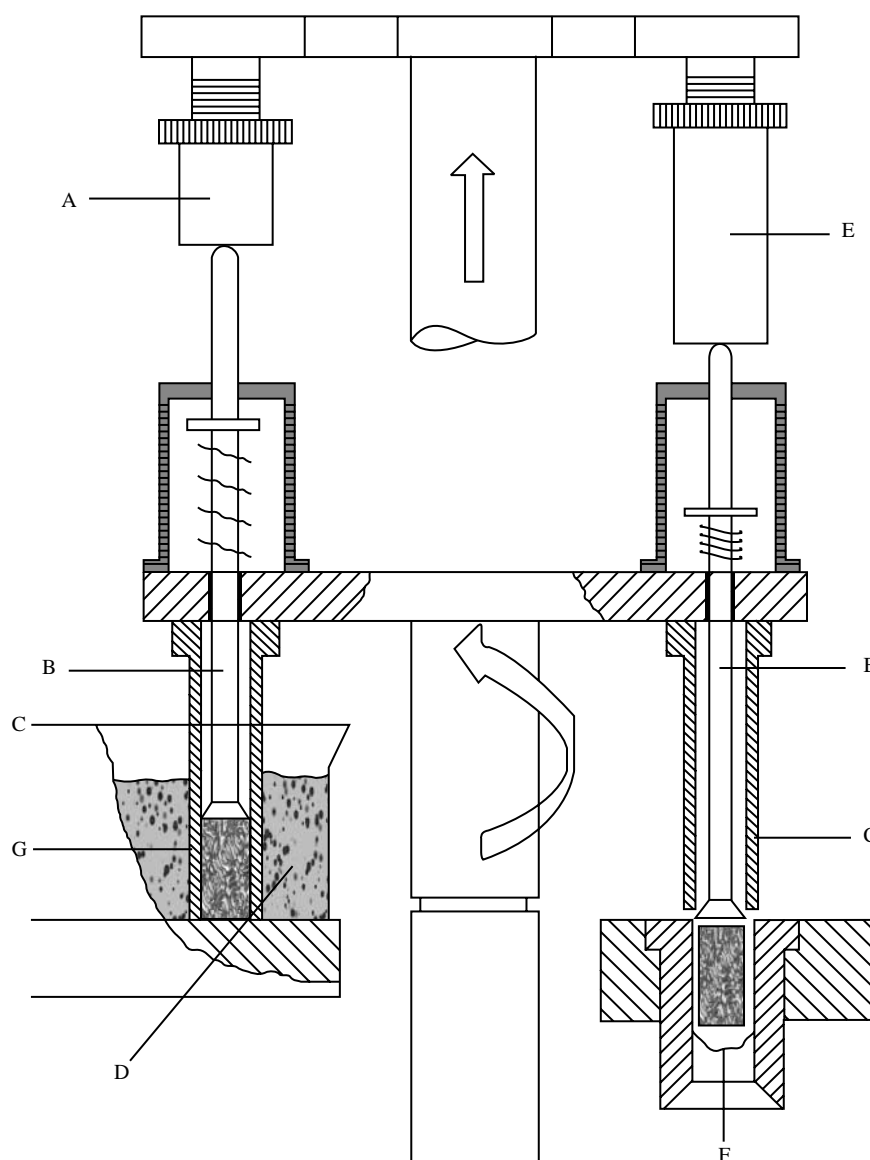


Fig. 4 Diagram of a dosator or dosing tube system (Zanasi RM63): A) Compression force platen; B) piston; C) dosing tube; D) powder hopper; E) plug ejection platen; F) capsule body in bush; G) powder plug.

in the cavities into plugs. Thus, the plug is formed in a series of tamps and not in a single motion as on the dosator machines. The dosing disks are produced in a range of thicknesses for each size of capsule. Thus, the selection of the correct thickness of disk is important because if the fill weight cannot be achieved, the machine has to be dismantled to change it. The selection of the optimum disk thickness for a formulation can be made either pragmatically using a simple test rig (11) or systematically by using an Instron tester to determine plug density and strength of a formulation at known compression forces (12).

Bench-scale filling machines

There are a variety of devices for the manual filling of small numbers of capsules. These typically consist of sets of plastic plates that have sets of holes drilled in them corresponding to the size of the capsule that can be filled. The capsules are fed into the plates, either manually one at a time or in groups using a feeding device. The bodies are clamped in the bottom plate and the top plate removed, which separates the caps from the body. The bodies are released so that they sit below the top of the bottom plate. Powder or pellets are filled into the capsules by spreading material over the body plate. Normally, a spatula is used to aid material flow. A larger version of this machine is

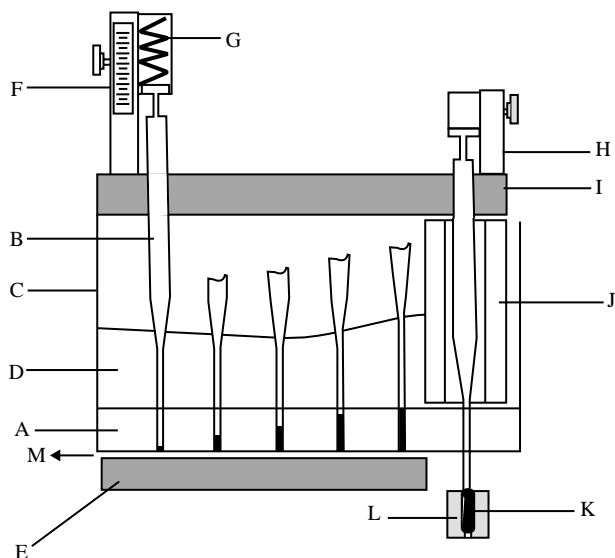


Fig. 5 Schematic diagram of a dosing disk and tamping finger system (Bosch GKF machine): A) dosing disk; B) tamping finger; C) powder hopper; D) powder bed; E) support plate; F) tamp depth adjuster; G) over load relief spring; H) ejection adjuster; I) guide block; J) transfer block; K) powder plug; L) capsule body in bush; M) suction.

available, the Model 301, made in stainless steel by Chemical and Pharmaceutical Industry Co., Inc. This has a tamping device, which enables higher fill weights to be achieved.

A small-scale automatic machine, the In-cap[®], Dott. Bonapace (Schaefer Technologies, Inc.), is available. This machine measures the dose using a tamping finger and dosing disk device. The output is up to 3000 capsules per hour.

Instrumented Filling Machines

Instrumented capsule-filling machines are not in widespread use, unlike tableting machines (13). This is partly owing to the facts that less basic work has been done, that there is an inherent problem in measuring the low forces, (1–100 N) used in forming and ejecting plugs, and that the powder bed is less controllable than that a tablet machine (10). Most of the published studies have been on intermittent motion dosing tube machines. Strain gauges have been applied to the piston and LVDTs to the moving parts of the system to measure the work involved in plug formation (13). Only one group has published work on an instrumented dosing disk and tamping finger machine (14). The problem with these machines is that the plug is formed at up to five different positions, and full instrumentation would be difficult. Capsule-filling machine simulators have been constructed to overcome some of the problems inherent in putting instrumentation on actual machines. Rotary operation machines present the biggest problem because of the movement of the dosing parts. One novel solution was to use the machine turret, with a single dosator, held stationary, and to construct a rig that moved the powder hopper around the dosator, simulating machine-running (15). A conventional simulator for an intermittent motion dosator machine has been built that, in addition to the forces of axial compression and ejection, can measure the radial compression force (16). This has been used to study the consolidation and elastic properties of excipients. A tablet compaction simulator has been used to investigate plug formation at low forces and the results analyzed using standard tableting physics (17).

Dry Solid Filling

Granules, pellets, and tablets can be filled into capsules using automatic filling machines. Products are prepared in these forms to modify the release rates of active ingredients, to separate incompatible components, or to

densify a product to achieve the fill weight in a specific size of capsule.

The machines that can be used to fill these products can be divided into the direct and indirect categories. In the former, the material is fed into the body until it is completely full, e.g., Quali-fill Pellet Filler (Schaefer Technologies, Inc.) or any auger-filling machine with the auger removed. In the latter, the dose of material is measured in a chamber with a variable volume, which can be adjusted to give the required weight. Machines have modified dosators that either use suction to hold the material in the tube during transfer or are filled when they pushed up through the material bed. Other machines have special chambers with sliding plates to measure and release the material, e.g., Bosch GKF machines. After measurement, material is transferred to the capsule bodies either using gravity or assisted by air pressure, e.g., IMA Farmatic 2090 and MG2 G60.

The physical properties of these formulations are similar. Each type must be preferably nonfriable; tablets are usually film-coated. Two types of tablets are filled into capsules. Generally, the tablets are convex and have diameters that enable them to be introduced easily into the body and with sufficient clearance so that they do not tip onto their side (18). In the United States, after the Tylenol[®] incident in the 1980s, there is a need to fill large single tablets into a capsule so that there is no room for movement in the shell. The capsule shell is either banded or shrunk onto the tablet to prevent its removal. Granules and pellets should be regular in shape so that they flow and pack well. Their size should be related to the size of the capsule. Smaller-diameter pellets should be used for smaller capsules; otherwise, lower than expected fill weight will occur because of the “wall effect” of particle-packing (19).

Liquid Filling

All the major machine manufacturers have made machines that can fill capsules with liquids. There are two types of liquid fills: formulations of nonaqueous solutions and suspensions and formulations that are liquefied only for the filling process by either heat or shear stress (20). If the formulation is mobile at ambient temperatures, then the capsules will need to be sealed after filling.

The dose of material is measured using volumetric pumps, and thus the uniformity of fill is in most cases better than what can be achieved normally on a powder-filling machine. Typically, coefficients of variation of fill weight less than 1.0% are routinely achievable. This value will depend on the physical properties of the liquid,

particularly its viscosity (22). Filling machines have been made that can handle materials with viscosities from 100 cp to 20,000 cp. Liquid-filling machines operate mostly at slower speeds than do the equivalent powder-filling machines. This is because the liquid has to pass through a much smaller orifice than that for a powder and thus takes longer. The rates are typically 50–66% of the rated output of the same size powder-filling machine.

Capsule fill capacity

The fill capacity of a hard capsule is dependent on the physical size of the capsule, the type of formulation, and the dosing mechanism on the filling device (Table 2) (10). The fill weight for powders has historically been calculated by multiplying a powder density value by the capsule volume as provided by the capsule manufacturers. This capsule volume is in fact the volume of the capsule body only, because the value was derived from work at the beginning of the 20th century when capsules were filled by hand in pharmacies. The relationship gives a reasonably accurate forecast machine filling if this volume number is multiplied by the tapped bulk density (TBD) of the powder. The reason for this is the dosing mechanisms on filling machines. The dependant machines, which can fill the total internal volume of the capsule, are able only to pack the powder at densities less than the TBD of the fill. The independent machines, which are able to apply a higher compressive force to the powder, form plugs whose dimensions must be less than the internal diameter and length of the capsule (Table 2). Thus, although the density of the plug will be higher than the TBD of the fill, the machines are unable to fill the total internal capsule volume.

The same rules of packing apply to pellet- and granule-filling. The size of the particles is important because of the increase in voidage caused by large particles in a small diameter tube. The smaller the capsule size, the smaller the corresponding size of the particles should be to achieve uniform fill weights.

There is a restriction in the fill capacity for liquids in capsules. To prevent the spillage of product, the maximum fill volume should not exceed 80% of the body volume.

Capsule Sealing

Many methods have been proposed for the hermetic sealing of capsules to prevent the leakage of liquids. The method proven to be the most successful is gelatin-banding (20). Two bands of gelatin solution are applied around the center of the filled capsule e.g., HicapsealTM 40/100

Table 2 Capsule fill volume data

Size	Body volume (ml) ^a	Internal volume ^b (ml)	Max. plug length ^c (mm)	Max. plug volume ^d (ml)
0E	0.78	0.87	21.9	0.68
0	0.68	0.78	19.7	0.61
1	0.50	0.56	17.7	0.44
2	0.37	0.44	16.1	0.34
3	0.30	0.32	14.3	0.26
4	0.21	0.25	13.2	0.19

^aFrom Capsugel Multistate File and Shionogi Qualicaps Capsules Technical Information Manual.

^bCalculated from optimum theoretical capsule shape, each part a cylinder with a hemispherical end.

^cCalculated from size of nondeforming plug, which fits inside a capsule closed to length specification of Shionogi Qualicaps Inc.

^dBased on hole dimensions of Bosch (H&K) machine dosing disk.

(Adapted from Ref. 10.)

(Schaefer Technologies, Inc.). This band is dried using air at ambient conditions to prevent moisture loss from the shells, which would make them brittle. The band can be colored, permitting a more complicated appearance for product branding. This band complies with the requirements of the FDA *Tamper-Evident Packaging Requirements for Over-the-Counter Human Drug Products* (21) for tamper-evident sealed capsules.

Multiple Contents

Automatic filling machines are available that can have more than one product dosing device. Therefore, combinations of materials can be filled into the same capsule, such a mixture of a powder and a semisolid formulation or a powder and a tablet. The same formulation rules apply as to single forms. Combinations of materials allow the formulator to achieve specific goals in terms of product stability and types of release.

FORMULATION

Powder Properties

Powder formulations for capsule-filling must have good flow properties, be nonadhesive, and be cohesive enough to form plugs at low compression forces. In addition, they must be stable and release the active ingredient in the desired manner. The filling properties can be assessed on the bench scale by using a variety of tests ranging from simple to complex (11). Successful correlations between

powders and filling performance have been made in several reports by determining various powder property constants calculated from tapped bulk-density volumetry (23–25). For various microcrystalline celluloses, Lüdde–Kawakita's constant a and Hausner's ratio were shown to be good indicators of machine-filling performance, especially when judging interchangeability of materials from various sources (23). Investigations on the packing properties of binary mixtures of different-shaped particles have shown that Lüdde–Kawakita's constant a can be used as an indicator of the maximum volume reduction (24). For microcrystalline cellulose, an angular particle, lactose monohydrate, improved packing, whereas spherical or needle-shaped particles tended to decrease the packing properties. The same methodology was used to investigate the bulk-volume changes of powders after granulation or low compression (25). This showed that capsule fill weights could be increased by high-shear granulation or by the use of machine compression and that the outcome was related directly to the initial powder properties. The filling of capsules with powdered herbs presents additional challenges because of the range of tissue materials used. The flow properties of these materials are poor, and a range of powder property constants was determined to try to find a parameter that correlated with filling-machine performance (26). It was found that tamp-filling machines were able to handle a greater variety of herbs than were dosator machines. The flow of powders under active conditions can be measured using specially constructed rheometers, and these data can be related to other powder properties (27).

The flow of powders on filling machine is aided by machine design. Most machines have devices to assist flow in the form of moving mechanical parts, vibration, or

suction pads. Another hindrance to obtaining good fill-weight uniformity on machines is adhesion of material to moving parts, particularly to the dose-measuring devices. It has been shown that the nature of the surface texture of the dosator is an important factor (28). The surface of the dosing parts can be coated with different metal finishes, similar to those used for tablet punches and dies, to reduce adhesion (29). The type of coating is related to the physical nature of the powder formulation.

In Vitro Testing

Pharmacopeias require that hard capsules be tested in the same apparatus as tablets, even though they have very different physical properties. Filled capsules contain entrapped air, and most formulations will float on water. Devices are required to ensure that they sink, and these can influence the results obtained (30). Gelatin and HPMC are adhesive materials and tend to block wire meshes that form part of the standard equipment. The manner in which capsules disintegrate and dissolve is dependent on several factors such as temperature and nature of the test media (31). The literature makes reference to the hard capsule effect; however, the literature shows that the rate-controlling step is the nature of the contents and not of the shell (32).

When capsules are placed in an aqueous solution at body temperature (37°C), the walls absorb water and swell (31). The rate of penetration is proportional to the thickness of the wall. In gelatin capsules, water droplets can be observed on the inside surface of the shell after 30–40 s. The wall ruptures first at the shoulders of the cap and body, which are the thinnest parts of the shell. The rate of gelatin solubility is dependent on the temperature of the solution (33). There is a significant decrease as the temperature falls below 35°C, and below approximately 30°C, they are completely insoluble and merely swell and distort. HPMC capsules, on the other hand, have a slower but uniform solubility between 10 and 55°C (34). The results for both types of capsules are influenced by the nature of the test media, e.g., the ionic strength of the ions present and the pH level (6, 31).

The rotating paddle method is the most frequently prescribed apparatus for measuring the dissolution rate of products in hard capsules. The test is used for manufacturing-control purposes and for assessing product stability. When gelatin capsules are stored at high temperatures and humidities (45°C, 75% RH), their solubility in water decreases with time. This is owing to the formation of a “pellicle” that slows release (35). This effect is called cross-linking and can be caused either by

interaction between gelatin and compounds containing reactive groups such as an aldehyde (36) or by reorientation of the gelatin molecules to a more collagen-like structure (2). In the early 1990s, the FDA became concerned with this and initiated a test program to measure whether this had an effect on product efficacy (37). They filled acetaminophen into capsules that had been stressed by treatment with formaldehyde at two levels and into unstressed shells. They measured the dissolution in water and in simulated gastric fluid (SGF), with and without pepsin. This produced three sets of results, those that passed in all media (unstressed shells), those that failed in water but passed in the SGF (low-stress formaldehyde), and those that failed in all media (high-stress formaldehyde). The capsules were tested in human volunteers. The pharmacokinetic parameters C_{\max} , T_{\max} , and lag times could be ranked in order and the Area Under the Curve (AUC) were identical. However, the products were not considered bioequivalent because the results from the capsules that failed all the dissolution tests were outside the 80–125% confidence limits compared with the unstressed capsules. From this study, the *U.S. Pharmacopeia* introduced a two-tier test for hard-capsule dissolution (36). If the capsule product fails in water, then the test can be repeated using either a solution at pH 1.2, containing pepsin, or one at pH 7.2, containing pancreatin. An additional study using γ -scintigraphy showed that there was no difference in disintegration in vivo between untreated and medium-stressed capsules (38).

HPMC does not react with aldehydes or other agents that cause cross-linking of gelatin (6). HPMC capsules start to release their contents slightly slower than do gelatin capsules because of the slower rate of diffusion of water through the shell walls. However, once dissolution has begun, rates are similar, and the results are comparable. There are only minor changes in their dissolution after storage under accelerated conditions (6).

In Vivo Performance

Capsule products can be formulated to deliver active ingredients to various sites along the gastrointestinal tract or to the respiratory system (39). Buccal products can be made by filling standard capsules with semisolid matrix formulations, which give the product good sensory characteristics that allow them to be chewed or sucked and the contents retained in the mouth for absorption or action (40). The capsule shape is good for swallowing because one axis is longer than the other. This enables the tongue to line it up like a torpedo for entry into the throat. Many large tablets are capsule-shaped, the so-called

caplet, to take advantage of this. The literature shows that, provided the patient takes the capsules with water while up right, they do not stick in the throat any more than and, in fact, probably less than any other solid dosage form (41, 42). Capsules can be visualized inside the patient either using radio-opaque markers and x-rays or using radioisotopes such as technetium-99 and γ -scintigraphy. In the stomach, they disintegrate, and the contents spread depending on the patient's feeding state, fed or fasting (43). Capsule products can be retained in the stomach by the use of floating formulations. These are based on the use of hydrocolloids that swell on contact with water, forming a gel that releases the active drug by diffusion (44). Enteric products can be made either by coating the capsule shell with a polymer, which has the correct pH solubility characteristics, or by filling the capsule with coated particles (39). The challenge facing many formulators is the delivery of small peptides and proteins to the colon. This can be achieved by coating capsules with polymers that will only be broken down in the colon, e.g., mixtures of an azopolymer and a methacrylate polymer (45). Delivery to the colon can also be achieved by using a fill that includes an organic acid and a combination of pH-sensitive coatings, which together deliver the active drug to the proximal colon (46). Capsules can be administered rectally. They can be formulated to give immediate or prolonged release (47). The administration technique is different for other solid rectal forms, and they need to be coated with a glidant such as liquid paraffin.

Powders for inhalation products have been filled into capsules, which function as an inert biodegradable package. The active ingredient is in a micronized state, and it is filled either directly into the capsule or, more frequently, is attached to a carrier particle such as lactose (48). The formulations are filled using automatic machines and because the fill weight is small, i.e., less than 40 mg, microdosators are used. The product is taken using a special inhalation device. Powder is released from the shell either through holes, which are punctured into it by the device, or by the capsule parts being separated inside the device. The inhalers are breath-actuated. When the patient inhales, there is a turbulent airflow through the device that carries the active particles into the lung (49).

Formulation for Release

Most products are formulated to release their contents into the stomach. The rate-controlling step for release is the nature of the contents inside the capsule (32). In preparing a formulation, a formulator needs to take into account the

physicochemical properties of the active ingredient, the nature and type of excipients required, and the filling process (10).

The properties of the active drug that are most significant are its aqueous solubility and particle size. The particle size needs to be chosen carefully. Smaller particles should dissolve faster because of their greater surface area but when filled inside a capsule they may aggregate, and the dissolving liquid may not be able to reach the individual particles (Fig. 6) (50). The available surface area of the active ingredient is more important than the actual surface area. Usually, the excipient that is the largest single quantity in the formulation is the filler (diluent), which functions both to increase the amount of fill material for potent actives and to aid in the formation of the powder plug. They can also play a role in the release of the active drug. People were first alerted to this in the late 1960s by the diphenylhydantoin incident in Australia, which showed that fillers need to be selected with solubility properties complementary to those of the active (51). Poorly soluble actives are best formulated with soluble excipients. The overall aim should be to make a powder mass that is as hydrophilic as possible. This can be done easily with potent actives because there is space available inside the capsule to accommodate excipients with the necessary properties, both in terms of flow and solubility. For higher-dose actives, excipients must be chosen that are active at low concentrations. Thus, disintegration and wetting agents need to be added. Excipients such as starch do not function as disintegrants in capsules as they do in tablets because the powder fill is much more porous. Sodium starch glycolate and croscarmellose, the so-called superdisintegrants, are used because of their greater swelling and wicking capability (52).

Certain excipients added to formulations to improve filling-machine performance can have an adverse effect on release because they are hydrophobic in nature. This is true of lubricants, which are added to formulations to prevent adhesion and to improve flow. The most used excipient in capsule formulations in both the United States and Europe is magnesium stearate (52, 53). This is hydrophobic, and there are many reports in the literature concerning its adverse effect on dissolution rates. However, the relationship between the concentration of magnesium stearate and release rate is not quite as simple as for tablets, in which an increase in amount brings a proportional decrease in release. The reason for this is the very different nature of tablets and capsules. A tablet is compressed using high forces to form a solid compact of relatively low porosity and must be if it is to survive subsequent handling. A hard capsule product, on the other hand, contains a powder mass of high porosity, which may

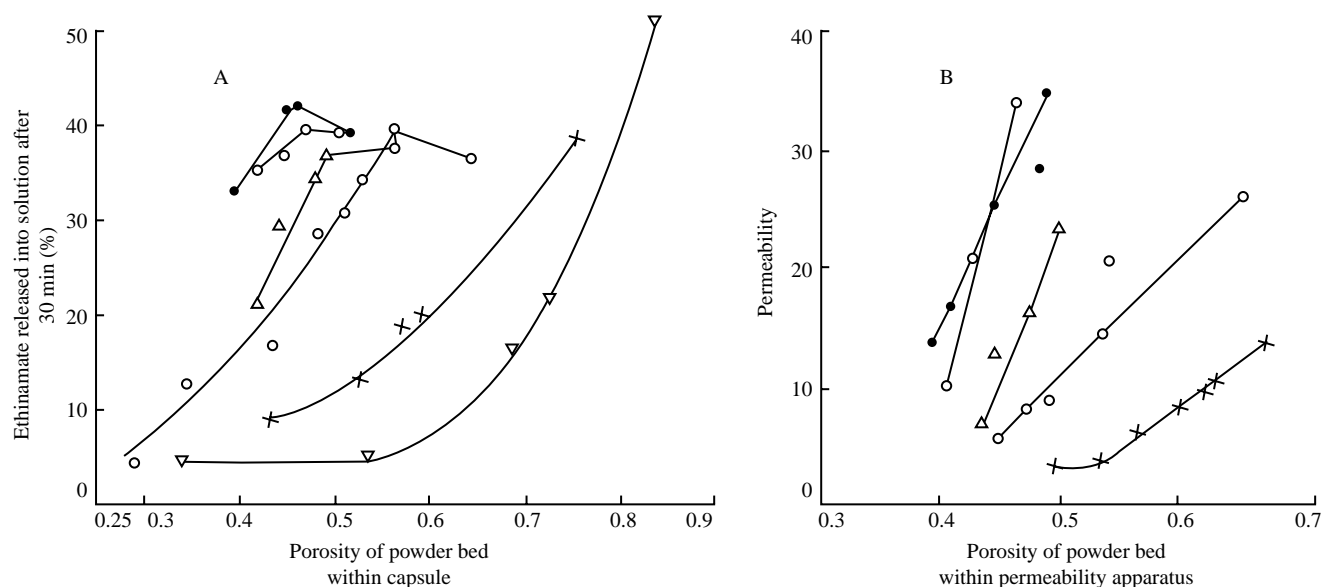


Fig. 6 The release of ethinamate from capsules containing different particle size fractions packed to give different porosities. A) The percentage of drug release after 30 min; B) The liquid permeability $\text{m}^2 \times 10^{11}$ of equivalent size fractions at known porosities. ●—● = 251–420 μm ; □—□ = 177–251 μm ; ○—○ = 152–177 μm ; +—+ = 66–76 μm ; ▽—▽ 8.3 μm . (Adapted from Ref. 50.)

or may not have been compressed into a plug, and is contained within the shell that can withstand handling. Magnesium stearate functions as a lubricant when it is dispersed on the surfaces of other particles. At this site, it also reduces the cohesion of particles, and thus as its concentration increases, the powder mass will become weaker. Several workers have shown that an increase in magnesium stearate concentration has increased dissolution rates: small particles are made less cohesive (Fig. 7) (54), and powder plugs are weakened, thus breaking apart more readily when the capsule shell has dissolved (Fig. 8) (55). If the level in the formulation is not optimized, then there is a possibility that during the filling operation, the magnesium stearate will be gradually dispersed to a greater extent, resulting in changes in dissolution or weight uniformity (10, 56).

The method to improve the release rate of poorly soluble actives by dissolving or suspending them in polyethylene glycol was first suggested in 1970 (57). Since then, the filling of semisolid matrix formulations for filling into hard capsules has been developed, which enables this simple concept to be turned into a practical application (20). This formulation technique gives a different means to control the release of active drug from a capsule, either improving or delaying release. The technique involves dispersing or dissolving the active drug in excipients that are available in a range of melting points and Hydrophile-Lipophile Balance values (20). It is possible to modify the release rate of an active ingredient from such a matrix

capsule by simply changing the properties of the single excipient (57). This technique has the added advantage that when working with potent and toxic material, it significantly reduces cross-contamination within an area (58). The actives once dispersed in a semisolid matrix are safe to handle without resorting to the use of expensive containment areas, i.e., any material that is spilt does not spread through the local environment, unlike a powdered material.

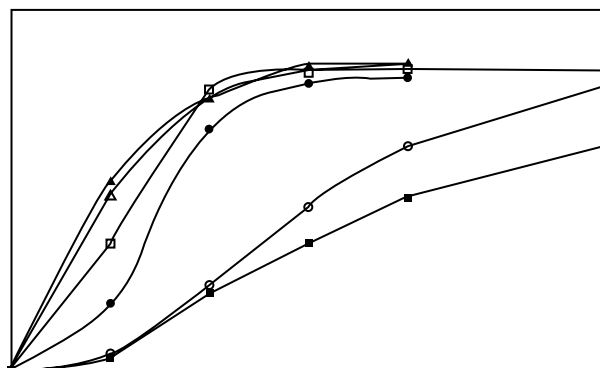


Fig. 7 Effect of particle size of rifampicin on its dissolution profile from capsules in the first fluid of JP IX by USP XVIII method. Rifampicin particle size: □—■, 42–80 mesh; △—▲, 80–100 mesh; ○—●, <200 mesh. Closed symbols with 0.5% magnesium stearate; open symbols without any addition. (Redrawn from Ref. 54.)

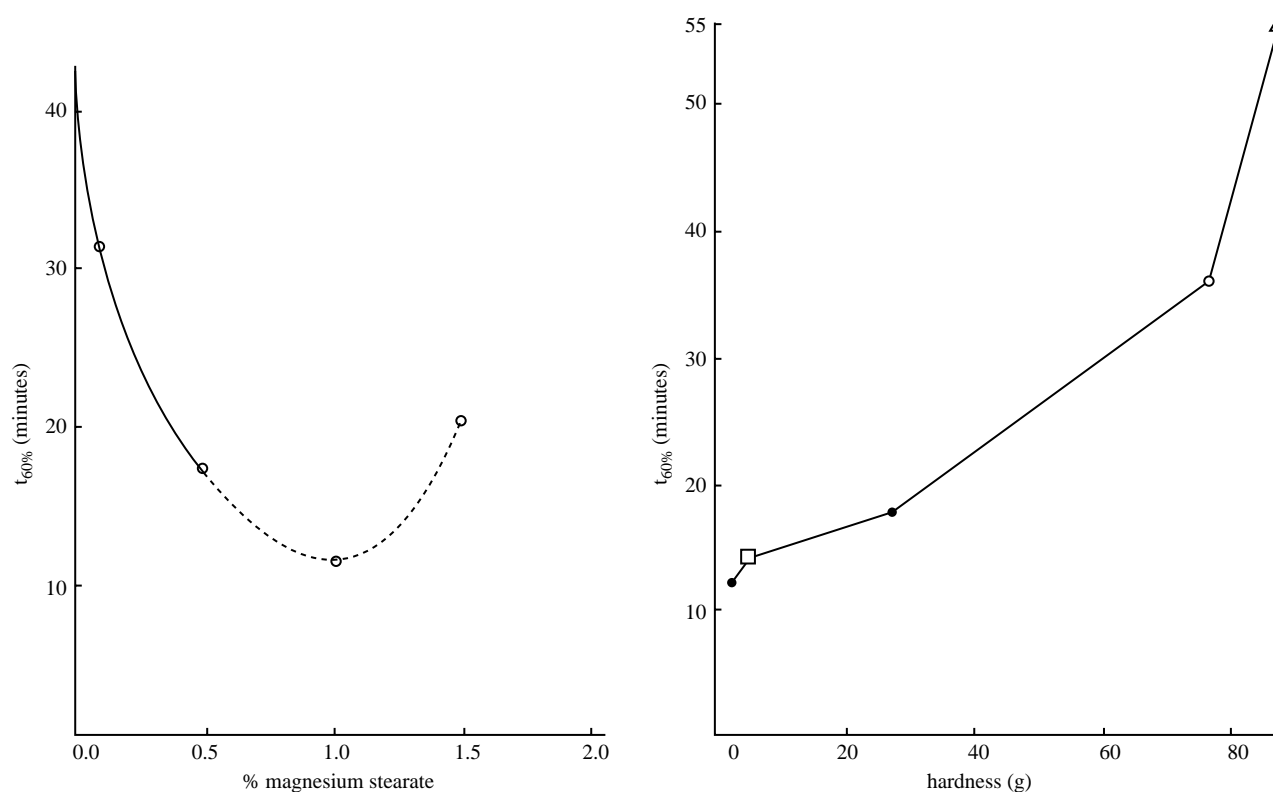


Fig. 8 Release of hydrochlorthiazide content from capsules, time for 60% to dissolve. All formulations filled on an instrumented Zanasi model LZ64 at the same compression force: A) effect of magnesium stearate concentration, microcrystalline cellulose filler, compression force 15 Kg; B) effect of plug hardness, microcrystalline cellulose filler, compression force 21.7 Kg. (Adapted from Ref. 55.)

Formulation Optimization and Expert Systems

Product formulations must meet a number of goals. They must be able to be filled by machines to give a uniform, stable product. They must release the active ingredients in a manner to give the desired therapeutic effects. They must comply with the regulatory and compendial specifications. The excipients used in formulations often have properties that aid in compliance with one aspect but, at the same time, can have a negative effect on another goal. The relationship among the factors is complex. There are a variety of statistical tools that can be used to optimize formulations to achieve the best values of all the factors (59).

Another method of obtaining the best formulation is to use a so-called expert system to devise a formulation. The computer software is based on the use of neural networks and knowledge-based systems (60, 61). They serve two functions. First, they are able to reduce development time by suggesting the probable formulations, and second, they act as a teaching tool to pass on the knowledge of experts in the field.

REFERENCES

1. Jones, B.E. *Hard Capsules, Development and Technology*; Ridgway, K., Ed.; Chaps. 1, 4, and 6, Pharmaceutical Press: London, 1987.
2. Jones, R.T. *Hard Capsules, Development and Technology*; Ridgway, K., Ed.; Chaps. 2 and 3, Pharmaceutical Press: London, 1987.
3. *BSE Vademecum, Information for Consumers*, 3rd Ed.; Directorate General XXIV, European Commission: Brussels, 1998; 63.
- 4a. G.S. Technologies Inc., Method for the Manufacture of Pharmaceutical Cellulose Capsules US Patent 5,698,155 Dec. 16, 1997.
- 4b. Apparatus for the Manufacture of Pharmaceutical Cellulose Capsules US Patent 5,750,157 May 12, 1998.
- 4c. Method for the Manufacture of Pharmaceutical Cellulose Capsules US Patent 5,576,036 May 26, 1998.
- 5a. Shionogi Qualicaps Co., Ltd., Hard Capsule for Pharmaceutical Drugs and Method for Producing Same US Patent 5,264,223 Nov. 23, 1993.
- 5b. Hard Capsule for Pharmaceutical Drugs and Method for Producing Same US Patent 5,431,917 Jul. 11, 1995.
- 5c. Capsule Shell US Patent 5,756,123 May 26, 1998.
- 5d. Hard Capsules European Patent 0 592 130 B1.

6. American Federal Standard No. 285A. *Capsules (for Medicinal Purposes)*; General Services Administration: Washington, DC, 1976; 20.
7. Ogura, T.; Matsuura, S. HPMC Capsules—An Alternative to Gelatin. *Pharm. Technol. Europe* **1998**, *10* (11), 32, 34, 36, 40, 42.
8. York, P. Analysis of Moisture Sorption Hysteresis in Hard Gelatin Capsules, Maize Starch, and Maize Starch:Drug Mixtures. *J. Pharm. Pharmacol.* **1981**, *33*, 269–273.
9. Chang, R.-K.; Raghavan, K.S.; Hussain, M.A. A Study on Gelatin Capsule Brittleness: Moisture Transfer between the Capsule Shell and its Content. *J. Pharm. Sci.* **1998**, *87*, 556–55.
10. Jones, B.E. New Thoughts on Capsule Filling. *S.T.P. Pharma Sci.* **1998**, *8*, 277–283.
11. Jones, B.E. Powder Formulations for Capsule Filling. *Mfg. Chem.* **1988**, *59* (7), 28–30 & 33.
12. Davar, N.; Shah, R.; Pope, D.G.; Augsburger, L.L. The Selection of a Dosing Disk on a Hofliger—Karg Capsule-Filling Machine. *Pharm. Tech.* **1997**, *21* (2), 32, 34, 36, 38, 40, 42, 44 and 46.
13. Augsburger, L.L. Instrumented Capsule Filling Machines. Methodology and Application to Product Development. *S.T.P. Pharma* **1988**, *4*, 116–122.
14. Cropp, J.W.; Augsburger, L.L.; Marshall, K.M. Simultaneous Monitoring of Tamping Force and Pin Displacement (F-D) on an Hofliger and Karg Capsule Filling Machine. *Int. J. Pharmaceut.* **1991**, *71*, 127–136.
15. Jolliffe, I.G.; Newton, J.M.; Cooper, D. The Design and Use of an Instrumented mG2 Capsule Filling Machine Simulator. *J. Pharm. Pharmacol.* **1982**, *34*, 230–235.
16. Britten, J.R.; Barnett, M.I.; Armstrong, N.A. Studies on Powder Plug Formation Using a Simulated Capsule Filling Machine. *J. Pharm. Pharmacol.* **1996**, *48*, 249–254.
17. Heda, P.K.; Muller, F.X.; Augsburger, L.L. Capsule Filling Machine Simulation. I. Low Force—Force Powder Compression Physics Relevant to Plug Formation. *Pharm. Dev. Technol.* **1999**, *4*, 209–219.
18. Hofer, U. Trockene Füllgüter. *Die Kapsel, Grundlagen, Technologie und Biopharmazie einer moderne Arzneiform*; Fahrig, W., Hofer, U., Eds.; Wissenschaftliche Verlagsgesellschaft mbH: Stuttgart, 1983; 83–111, Ch. V.
19. Cartensen, J.T. *Pharmaceutics of Solids and Solid Dosage Forms*; John Wiley & Sons: New York, 1977; 138–142.
20. Bowtle, W.J. Liquid Filling of Hard Gelatin Capsules—New Technology for Alternative Formulations. *Pharm. Technol. Europe* **1998**, *10* (10), 84, 86, 88–90.
21. 21 CFR Part 211, Docket No. 92N-0314, Federal Register, Vol. 63, Nov. 4 1998, 59463–59471.
22. Shah, N.H.; Phuapradit, W.; Ahmed, H. Liquid/Semi-Solid Filling in Hard Gelatin Capsules: Formulation and Processing Considerations. *Bull. Tech. Gattefossé* **1996**, *89*, 27–37.
23. Patel, R.; Podczek, F. Investigation of the Effect of Type and Source of Microcrystalline Cellulose on Capsule Filling. *Int. J. Pharm.* **1996**, *128*, 123–127.
24. Podczek, F.; Sharma, M. The Influence of Particle Size and Shape of Components of Binary Powder Mixtures on the Maximum Volume Reduction Due to Packing. *Int. J. Pharm.* **1996**, *137*, 41–47.
25. Podczek, F.; Lee-Amies, G. The Bulk Volume Changes of Powders by Granulation and Compression with Respect to Capsule Filling. *Int. J. Pharm.* **1996**, *142*, 97–102.
26. Podczek, F.; Claes, L.; Newton, J.M. Filling Powdered Herbs into Gelatine Capsules. *Mfg. Chem.* **1999**, *60* (2), 29–31 & 33.
27. Podczek, F. Rheological Studies of the Physical Properties of Powders Used in Capsule Filling. I, II. *Pharm. Technol. Europe* **1999**, *11* (9), 16, 18, 20–21 & 24 and 11(10), 34, 36, 38, 40 & 42.
28. Tan, S.B.; Newton, J.M. Influence of Capsule Dosator Wall Texture and Powder Properties on the Angle of Wall Friction and Powder-Wall Adhesion, and Capsule Filling Performance of Powders with Dosator Nozzles of Different Wall Texture. *Int. J. Pharm.* **1990**, *64*, 227–234 & 66:207–211.
29. Podczek, F. Investigations into the Reduction of Powder Adhesion to Stainless Steel Surfaces by Surface Modification to Aid Capsule Filling. *Int. J. Pharm.* **1999**, *178*, 93–100.
30. Avgoustakis, K.; Anthanasiou, A.; Georgakopoulos, P.P. Effect of Helix Characteristics on the Dissolution Rate of Hard Gelatin Capsules. *Int. J. Pharm.* **1992**, *79*, 67–69.
31. Ludwig, A.; Van Ooteghem, M. Disintegration of Hard Gelatin Capsules. III. Penetration and Diffusion of Liquid through the Capsule Wall Investigated by Scanning Electron Microscopy. *Pharm. Ind. Berl.* **1980**, *42*, 1040–1043.
- 31a. Method of Measuring the Time of Rupture of the Capsule. *Pharm. Ind. Berl.* **1980**, *42*, 1140–1141.
32. Jones, B.E. Disintegration of Hard Gelatin Capsules. *Chem. Pharm. Bull.* **2000**, *48*, 951–956.
33. Jones, B.E.; Cole, W.V.J. The Influence of Test Conditions on the Disintegration Time of Gelatin Capsules. *J. Pharm. Pharmacol.* **1971**, *23*, 438–443.
34. Chiwele, I.; Jones, B.E.; Podczek, F. S. S. M. Formulation and Capsules: An Improved Way to Handle Toxic Compounds.
35. Cartensen, J.T.; Rhodes, C.T. Pellicule Formation in Gelatin Capsules. *Drug Dev. Ind. Pharm.* **1993**, *19*, 2709–2712.
36. Digenis, G.A.; Gold, T.B.; Shah, V.P. Cross-Linking of Gelatin Capsules and its Relevance to their In Vitro–In Vivo Performance. *J. Pharm. Sci.* **1994**, *83*, 915–921.
37. Hard and Soft Gelatin Capsules: Issues, Research and Outcomes. *Pharm. Res.* **1997**, *14* (12), Podium Session, AAPS Convention, Boston, 1997.
38. Brown, J.; Madit, N.; Cole, E.T.; Wilding, I.R.; Cadé, D. The Effect of Cross-Linking on the In Vivo Disintegration of Hard Gelatin Capsules. *Pharm. Res.* **1998**, *15*, 1026–1030.
39. Jones, B.E. The Two-Piece Gelatin Capsule and the Gastro-Intestinal Tract. *S.T.P. Pharma Pratiques* **1991**, *1*, 128–134.
40. Wilson, C.G. Washington, N. Assessment of Disintegration and Dissolution Dosage Forms In Vivo Using Gamma Scintigraphy. *Drug Dev. Ind. Pharm.* **1988**, *14*, 211–281.
41. Evans, K.T.; Roberts, G.M. Where do all the Tablets Go? *Lancet* **1976**, *ii*, 1231–1239.
42. Hey, H.; Jørgensen, F.; Sørensen, K.; Hasselbach, H.; Wamberg, T. Oesophageal Transit of Six Commonly Used Tablets and Capsules. *Br. Med. J.* **1982**, *285*, 1717–1719.

43. Hunter, E.; Fell, J.T.; Calvert, R.T.; Sharma, H. "In Vivo" Disintegration of Hard Gelatin Capsules in Fasting and Non-Fasting Subjects. *Int. J. Pharm.* **1980**, *4*, 175–183.
44. Timmermans, J.; Möes, A.J. Factors Controlling the Buoyancy and Gastric Retention Capabilities of Floating Matrix Capsules: New Data for Reconsidering the Controversy. *J. Pharm. Sci.* **1994**, *83*, 18–24.
45. Kalala, W.; Kinget, R.; Van den Mooter, G.; Samyn, C. Colonic-Drug Targeting: In Vitro Release of Ibuprofen from Capsules Coated with Poly(Ether-Ester) Azopolymers. *Int. J. Pharm.* **1996**, *139*, 187–195.
46. Ishibashi, T.; Hatano, H.; Kobayashi, M.; Mizobe, M.; Yoshino, H. Design and Evaluation of a New Capsule-Type Dosage Form for Colon-Targeted Delivery of Drugs. *Int. J. Pharm.* **1998**, *168*, 31–40.
47. Leino, J.; Haavisto, H.; Tomminen, T.; Heinilä, K.; Eerikäinen, S.; Klinge, E.; Marvola, M. Development of Rectally Administered Prolonged Release Hard Gelatin Capsules Using Different Polymers as Diluents. *S.T.P. Pharma Sci.* **1997**, *7*, 348–353.
48. Srichana, T.; Martin, G.P.; Marriott, C. On the Relationship between Drug and Carrier Deposition from Dry Powder Inhalers In Vitro. *Int. J. Pharm.* **1998**, *167*, 13–23.
49. Srichana, T.; Martin, G.P.; Marriott, C. Dry Powder Inhalers: The Influence of Device Resistance and Powder Formulation on Drug and Lactose Deposition. *Eur. J. Pharm. Sci.* **1998**, *7*, 73–80.
50. Newton, J.M.; Rowley, G.M. On the Release of Drug From Hard Gelatin Capsules. *J. Pharm. Pharmacol.* **1970**, *22* (Suppl.), 163S–168S.
51. Tyrer, J.M.; Eadie, M.J.; Sutherland, J.M.; Hopper, W.D. Outbreak of Anticonvulsant Intoxication in an Australian City. *Br. Med. J.* **1970**, *iv*, 271–273.
52. Shangraw, R.F.; Demarest, D.A. A Survey of Current Industrial Practices in the Formulation and Manufacture of Tablets and Capsules. *Pharm. Technol.* **1993**, *17* (1), 32–44.
53. Jones, B.E. Two-Piece Gelatin Capsules: Excipients for Powder Products, European Practice. *Pharm. Technol. Europe* **1995**, *7* (10), 25, 28–30 & 34.
54. Nakagawa, H.; Mohri, K.; Nakashima, K.; Sugimoto, I. Effects of Particle Size of Rifampicin and the Addition of Magnesium Stearate on Release of Rifampicin from Hard Gelatin Capsules. *Yakugaku Zasshi* **1980**, *100*, 1111–1117.
55. Mehta, A.M.; Augsburger, L.L. A Preliminary Study of the Effect of Slug Hardness on Drug Dissolution from Hard Gelatin Capsules Filled on an Automatic Capsule-Filling Machine. *Int. J. Pharm.* **1981**, *7*, 327–334.
56. Harding, V.D.; Higginson, S.J.; Wells, J.I. Predictive Stress Tests in the Scale-Up of Capsule Formulations. *Drug Dev. Pharm.* **1989**, *15*, 2315–2338.
57. Chiou, W.I.; Riegelman, S. Oral Absorption of Griseofulvin in Dogs: Increased Absorption Via Solid Dispersions in Polyethylene Glycol 6000. *J. Pharm. Sci.* **1970**, *59*, 937–942.
58. Duclos, R.; Bourret, E.; Brossard, C. Rheology of Polyol Behenates and Drug Release from Matrix Capsules. *Int. J. Pharm.* **1999**, *182*, 145–154.
59. Bowtle, W.J.; Burrows, S.L.; Holmes, E.D.; Jones, B.E. *Pharm. Res.* **1989**, *5* (12), PT612.
60. Hogan, J.; Shue, P-I.; Podczek, F.; Newton, J.M. Investigations into the Relationship between Drug Properties, Filling, and the Release of Drugs from Hard Gelatin Capsules Using Multivariate Statistical Analysis. *Pharm. Res.* **1996**, *13*, 944–949.
61. Bateman, S.D.; Verlin, J.; Russo, M.; Guillot, M.; Laughlin, S.M. The Development and Validation of a Capsule Formulation Knowledge-Based System. *Pharm. Technol.* **1996**, *20* (3), 174, 176, 178, 180, 182 & 184.
62. Lai, S.; Podczek, F.; Newton, J.M.; Daumesnil, R. An Expert System to Aid the Development of Capsule Formulations. *Pharm. Technol. Europe* **1996**, *8* (9), 60, 62, 64, 66 & 68.

CHROMATOGRAPHIC METHODS OF ANALYSIS— GAS CHROMATOGRAPHY

Isadore Kanfer
Roderick B. Walker
Michael F. Skinner

Rhodes University, Grahamstown, South Africa

INTRODUCTION

History

The word “chromatography” is derived from the Greek words “chroma” and “graphein,” which mean “color” and “to write,” respectively, or “color writing.” The initial use of the term is attributed to Tswett (1), who separated colored bands of plant pigments on a chromatographic column that consisted of an adsorbant powder that was washed with a liquid solvent termed the mobile phase. Substitution of this liquid mobile phase by a gas constitutes the fundamentals of gas chromatography (GC), where the solute to be separated is vaporized and carried down the length of a tube that contains an immobile solid or liquid phase, i.e., the stationary phase. The gaseous mobile phase serves to move the solute vapors along the column at rates dependent on several factors, the most important of these being temperature.

The first reported use of a vapor as the mobile phase is attributed to Martin and Synge (2) in 1941. They used the principles of partition chromatography, whereas James and Martin, in 1952, described the first application of this method, gas–liquid chromatography (GLC), for the analysis of fatty acids and amines (3). Gas adsorption chromatography (GSC), on the other hand, involves the use of a solid stationary phase and separation is based on an adsorptive mechanism. This technique was first described in 1947 in a doctoral thesis by Prior (4), under the supervision of Professor Cremer, and subsequently in their 1951 publication (5).

It was not, however, until 1955–1956 that the first commercial gas chromatographs were built (6, 7). Rapid progress in instrumentation and increased use of GC followed with the introduction of novel detectors, such as the flame ionization detector (8, 9), development of the capillary column (10), and the introduction of temperature programming and microsyringes for sample injection (6).

Utility

The introduction of GC as an analytical technique has had a profound impact on both qualitative and quantitative analysis of organic compounds. Identification of compounds by GC can be accomplished by their retention times on the column as compared to known reference standards, by inference from sample treatment prior to chromatography (11), or by the concept of retention index (12). The latter method and tables of retention indices (13) with associated conditions have been reported (14). Although qualitative data and analytical techniques for identification of compounds are well-established (15, 16) and relative retention data for over 600 substances also have been published (17), the main utility of GC undoubtedly lies in its powerful combination of separation and quantitative capabilities. Use in quantitative analysis involves the implementation of two techniques being performed concurrently, i.e., separation of components and subsequent quantitative measurement.

The use of GC was first included in the United States Pharmacopoeia (USP) in the sixteenth edition (18) in 1960, and became an official method of the British Pharmacopoeia (BP) in 1968 (19). GC has found widespread use in pharmaceutical analysis by virtue of its applications to purity and control analysis of raw materials, content and quality assessment of dosage forms (including product stability), and in the quantitative measurement of drugs in biological fluids. The latter application is important for therapeutic drug monitoring, pharmacokinetic studies, and bioavailability assessments. In fact, in a survey on GC use (20), a major application of this technique was in the field of pharmaceuticals.

When this article was first written several years ago, it appeared that the advent and establishment of high-performance liquid chromatography (HPLC) in pharmaceutical analysis had somewhat diminished with the utilization of GC. However, new regulatory requirements for drug approvals by the Federal Drug Administration

(FDA) and other regulatory agencies around the world, more particularly with respect to the determination of organic volatile impurities (OVI's) as well as other impurities and related substances (20), has resulted in more extensive use being made of GC in modern compendia, such as the USP 24th edition (21) and the 1999 edition of the BP (22). Perusal of the current USP (21) indicates that many more GC applications have been introduced since the 22nd edition of the USP. New inclusions have been incorporated in the tables in this article. Similarly, the recent edition of the BP (22) also includes numerous new applications. A list of compounds in the BP (which includes the European Pharmacopoeia) that use GC is included separately as an Appendix.

GC remains the chromatographic method of choice for thermally stable volatile compounds and for drugs, which are difficult to measure by HPLC due to detector insufficiency and/or inadequate resolution by the HPLC technique. The use of capillary columns in GC makes the method particularly attractive for difficult multicomponent analysis since extremely high resolution can be readily attained.

Modus Operandi

As previously mentioned, GC is a two-phase system that consists primarily of a stationary (solid and/or liquid) and mobile (gas) phase. When a liquid stationary phase is used (GLC), the liquid is immobilized as a thin film supported on a finely divided, inert solid support usually consisting of siliceous earth, crushed firebrick, glass beads, or in some cases, the inner wall of a glass tube. In GSC, the stationary phase is an active adsorbent, such as alumina, silica gel, or carbon, which is tightly packed into a tube.

Separation of components takes place in this tube (chromatographic column) following the introduction of sample at the tube inlet, which is subsequently swept through the column, partitioning or being dynamically adsorbed (or both) between the stationary and mobile phases during transit. The degree and speed of separation of components is governed by several factors, such as temperature, gas flow rate, and the physicochemical properties of the individual components being separated, as well as those of the stationary and to a lesser extent, mobile phase. Obviously, therefore, molecules with greater affinity for the stationary phase will spend more time either adsorbed to or partitioned within that phase and thus take longer to emerge from the column.

On emerging at the outlet, each component passes into a detector system that produces a signal that can be related to the mass of the individual component being detected.

This signal is usually amplified electronically and subsequently recorded on a chart-recorder, integrator, or captured by an online data system. The resulting response is in the form of a signal–time plot or chromatogram, and is subsequently evaluated for either qualitative or quantitative use.

THEORETICAL PRINCIPLES AND RATE THEORY

A general account of chromatographic theory was presented in volume 2 of *Encyclopedia of Pharmaceutical Technology* (23). Therefore, the following discussion will focus specifically on GC theory. The separation of the component of a mixture depends upon the column performance (efficacy) and the relative retention capability of the stationary phase (selectivity). The former determines the width of the peaks relative to the length of time a component spends in the column, while the latter determines the relative position of each emerging component (resolution).

When the sample is introduced into the column, usually in the form of a zone of vapor, it takes the form of a narrow band. During transit through the column, various factors influence the width of this band, which is continuously increased due to various dispersion processes. These include diffusion of the solute, resistance to mass transfer between and within phases, and the influence of flow irregularities and perturbations (24). A simple concept, the “theoretical plate,” carried over from distillation processes, has been used to compare columns and account for the degree of dispersion that influences bandwidth. A chromatographic column may be considered to consist of numerous theoretical plates where the distribution of sample components between the stationary and mobile phase occurs. Hence, a measure of the efficiency of a GC column may be obtained by calculating the number of theoretical plates, N , in the column from:

$$N = 16 \left(\frac{t}{w} \right)^2 \text{ OR } N = 5.54 \left(\frac{t^2}{w_{1/2}} \right) \quad (1)$$

where t is the retention time of the substance, w is the width of the base of the peak obtained by extrapolation (tangential extension of the sides of the relevant peak) of the relevant peak to the baseline, and $w_{1/2}$ is the peak width at half height (Fig. 1).

The higher the value of N for a column, the more efficient it will be. Columns with high efficiency allow smaller samples to be injected into shorter columns at

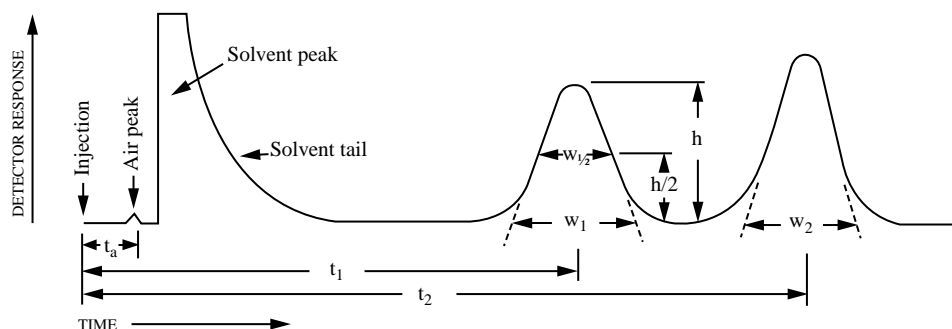


Fig. 1 Chromatographic separation of two substances.

lower temperatures, which results in high resolution of components in less time.

Column performance under different conditions or the comparison of different columns may be assessed by considering the height equivalent of a theoretical plate (HETP). Thus, $HETP = L/N$ where L is the length of the column. Van Deemter et al. (25) derived an equation for dispersion in chromatography:

$$HETP = 2\lambda d_p + \frac{2\gamma D_g}{\mu} + \frac{8k' d_f^2}{\mu^2 (1 + k')^2 D_1} \mu \quad (2)$$

where λ is a constant related to the geometry of the column packing particles, d_p is the average diameter of the solid support particles, γ is a factor to correct for the "tortuosity" of the column's gas channels, D_g and D_1 are solute diffusion coefficients in the gas and liquid phases respectively, d_f is the liquid film (stationary phase) thickness, k' is the partition coefficient of the solute, and μ is the linear gas velocity (26). Therefore, the Van Deemter equation expresses HETP as a function of the average mobile phase velocity μ and for a specific column, the equation has the general form:

$$HETP = A + \frac{B}{\mu} + C\mu \quad (3)$$

where A is the eddy diffusion term that results from flow inequalities in the column packing, B is the molecular diffusion term (when divided by μ it reflects axial diffusion in the mobile phase), and C reflects resistance to mass transfer from the stationary phase. The linear gas velocity, μ may be obtained from:

$$\mu = \frac{\text{Length of column}}{\text{Retention time of an unretained component (e.g., air)}} \quad (4)$$

The flow dependence of the three terms in the Van Deemter equation gives rise to a hyperbola (Fig. 2) when HETP is plotted against μ for a single component. The

minimum is the flow rate at which the column will function at optimum efficiency. Fig. 2 also depicts how the A , B , and C terms contribute to the HETP. For maximum efficiency, these terms must be minimized, i.e., keep HETP as small as possible. Minimization of the A parameter is readily accomplished by using small uniform packing material particles in small diameter columns. Decreasing the particle diameter, d_p , lowers the HETP. However, below a certain particle size, flow of carrier gas through the column is restricted and results in pressure increases, which limits the reduction in particle size. Since λ is a measure of the column packing particle irregularities, the more uniform the size and shape of these particles, the smaller the value of λ .

The B parameter relates to the diffusivity of the solute in the carrier gas. Increases in molecular diffusion will result in increases in band broadening, which may, however, be controlled to some extent by increasing the

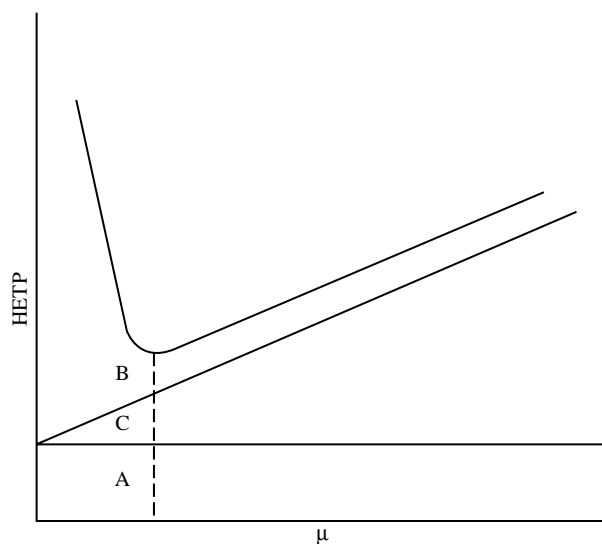


Fig. 2 Plot of HETP vs. flow velocity.

gas pressure. The use of a high-molecular-weight carrier gas will retard diffusivity and result in the best efficiency. However, detector function may be affected by the type of carrier gas used; hence, for the purposes of expediency, a compromise is often made by employing hydrogen or helium, which allows rapid analysis albeit with somewhat reduced efficiency. These aforementioned gases, which can rapidly diffuse into the stationary phase, are used at higher flow rates than nitrogen, another commonly used carrier gas.

The remaining parameter, C , is a measure of the mass transfer of the solute molecules from the stationary into the gas phase and depends upon several variables. These include the transfer of solute from liquid to gas and vice versa. Use of thin films of liquid phase will thus allow faster analysis at lower operating temperatures, although sample capacity will be reduced. Viscosity of the liquid stationary phase also affects mass transfer and therefore, should be kept as low as possible at the lowest possible temperature.

Golay (10) and Giddings (27), respectively, described a modification of the rate theory for capillary columns (hollow tube with inner wall coated with liquid phase) and the random walk, nonequilibrium theory. The former derived an equation to describe the efficiency of an open tubular column, while the random walk theory describes chromatographic separation in terms of statistical moments. The nonequilibrium theory involves a rigorous mathematical treatment to account for incomplete equilibrium between the two phases (28).

Selectivity is a function of the efficiency of the stationary phase with respect to its interactions with the solute vapor. Selection of an appropriate liquid stationary phase will even allow the separation of compounds that have the same vapor pressure. Separation is thus determined by the solubilities of the respective solutes in the stationary phase. Hence, the partition coefficient, k , is an extremely important parameter and is given by the following relationship:

$$k = \frac{\text{Concentration of solute in liquid phase}}{\text{Concentration of solute in gas phase}} \quad (5)$$

The efficiency of a stationary phase for a particular separation is measured by α , the relative retention, which is the ratio of two adjusted retention times (Fig. 1):

$$\alpha = \frac{t_2 - t_a}{t_1 - t_a} \quad (6)$$

where t_2 is the retention time of one of the components, t_1 is the retention time of a second or reference component in the mixture determined on the same column using the

same separation conditions, and t_a is the retention time for an unretained compound, such as air. It is thus seen that α reflects the ratio of the partition coefficients for two components being separated under identical conditions and is a useful parameter for the identification of compounds when one of the components is a reference standard material (21). In order to express how well two peaks are actually separated, a resolution term, R , may be determined from Fig. 1, i.e.,

$$R = \frac{2(t_2 - t_1)}{W_2 + W_1} \quad (7)$$

where t_2 and t_1 are the retention times of the two components, and W_2 and W_1 are the corresponding widths of the bases of the peaks. Resolution is a measure of both column and stationary phase efficiency and relates peak width and maximal separation. In order to obtain complete separation (baseline resolution) between two peaks, the value of R must be a minimum of 1.5.

SYSTEM COMPONENTS/EQUIPMENT

Gases

While in principle any gas may be used in GC as the carrier, a prerequisite stipulates that the gas be inert with respect to both sample and stationary phase at the operating temperature. The carrier gas plays a critical role in the separation process and indeed, contributes to the efficiency of the system, as was shown in the Van Deemter equation where HETP depends on solute diffusivity in the gas phase. In practice, however, the importance of this role is relegated to a somewhat lower priority since the choice of carrier gas is usually dictated by the detector requirements. Helium is the gas of choice for use with the thermal conductivity detector (TCD), and allows greater sensitivity as compared to nitrogen. The electron capture detectors (ECD), on the other hand, are more efficient when nitrogen or argon–methane mixtures are used as carrier gas, while no noticeable difference in sensitivity is evident between nitrogen and helium when using the flame ionization detector (FID) (29). Thermionic detectors (TD), such as the nitrogen–phosphorus detector (NPD) utilize nitrogen or helium as the carrier gas. Similarly, the photoionization detector (PID) uses oxygen-free nitrogen or helium, while nitrogen is used as carrier gas with the flame photometric detector (FPD). All gases used as carriers in GC should be of high purity. A report on carrier gas purity in GC has been comprehensively discussed by Perretta (30), and procedures for the

preparation of “clean” gases were published previously (31). Traces of hydrocarbons can lower detector sensitivity (FID), trace amounts of water can desorb contaminants in the column, which leads to high background signals and/or “ghost peaks,” while traces of oxygen can cause degradation of certain liquid phases, such as polyglycol and polyamides, which results in changes of solute retention times. Moisture can be removed by placing cartridges that contain an appropriate molecular sieve fitted in-line between the gas cylinder and the instrument. These type of filters also serve to remove other small, trace level contaminants, such as low-molecular weight hydrocarbons, and may be regenerated by heating with a slow flow of nitrogen for a few hours. Oxygen traps also should be used to protect stationary phases from oxidative degradation (32).

Flow Control

The carrier gas is fed into the GC via a pressure regulator, while flow controllers are used to control the mass flow rate. Maintenance of an accurate and constant carrier gas flow rate is essential for solute elution reproducibility in both qualitative and quantitative analysis. Normally, gas flow rates will decrease due to an increase in gas viscosity and column back pressure, with an increase in temperature, especially during temperature programmed work. Differential flow controllers are thus essential to assure a constant mass flow rate independent of the resistance of the column. In addition, detectors usually require gas flow control, and this can be accomplished using pressure regulators operating against flow restrictors. Gas flow rates can be simply measured at the end of the column with a soap bubble flow meter or by using rotometers. While flow control was previously adjusted manually, various manufacturers now offer software and associated hardware to effect such changes.

Sample Inlets

Various sample inlet systems have been designed with a primary objective of facilitating satisfactory vaporization of samples and subsequent transfer to the column as a compact “plug” in the shortest possible time and in an accurate and reproducible manner. Additional considerations for efficient sample introduction include maintenance of constant carrier gas flow rate and temperature during sample injection. Considerable differences, however, exist between the manner of sample injection and the actual injecting system, depending on whether packed columns or capillary columns are used. Therefore, sample volume considerations must be taken into account; whereas 1–10 μl is usual for packed columns, several orders of magnitude less is used with capillary columns. Inlet systems for packed columns usually consist of a heated injection block (Fig. 3) with a minimum dead volume port (to reduce band spreading) which is sealed with a special rubber septum through which the injection syringe needle may be inserted. Compounds which are thermally sensitive and unstable when in contact with metal surfaces may be protected by using glass liners that minimize the sample contact time with the metal injection block.

The foregoing discussion relates to the flash vaporization sample introduction technique that involves injection of sample into a precolumn zone that is kept at a temperature of 30–50°C higher than that of the column. This facilitates instantaneous sample vaporization. Samples also may be introduced by on-column injection where the sample is injected directly into the head of the column, which results in better precision than flash vaporization (33, 34).

Inlet systems for packed columns can often be used with capillary columns as well. However, the much smaller injection volumes and slower gas flow rates used with capillary columns, especially small-bore open tubular

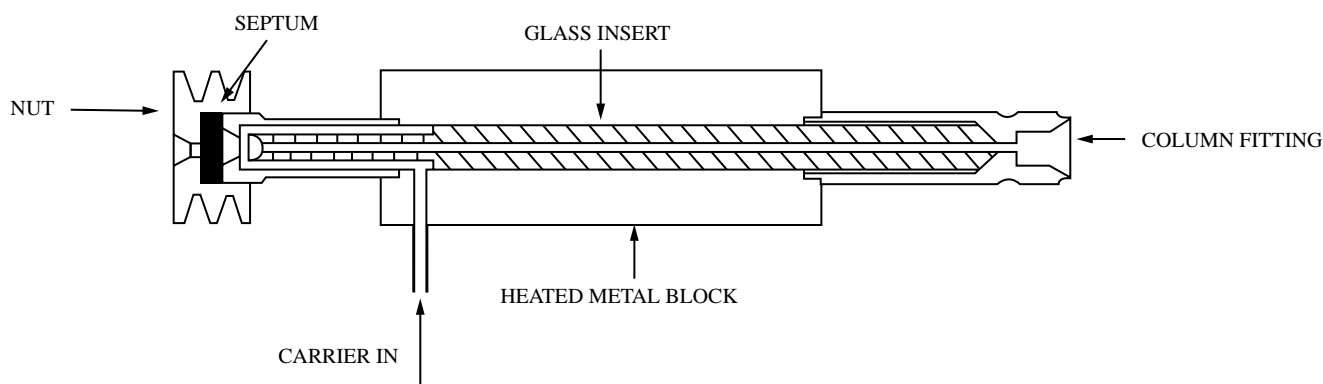


Fig. 3 Injection block.

capillaries (0.25 mm i.d.), require different sampling techniques.

Split injection

Early capillary inlets utilize an inlet splitter, which splits the sample into two unequal portions, the smaller of which goes into the column. The major function of the inlet splitter is not only to redirect the amount of sample placed on the column but also to permit rapid flushing of the injection chamber so that the sample in the column is followed by pure carrier gas, thereby avoiding sample dilution (35). The larger portion of sample is vented out of the system and the ratio of the two flows, the split ratio, typically ranges from 1:10 to 1:500. Since split injection is a flash vaporization technique, the possibility of sample discrimination exists. All sample components must be divided in the same ratio irrespective of differences in molecular weight, component concentration, polarity, injected volume, and inlet temperature for optimum reproducibility. Although the discriminatory effect can be minimized through the use of different inlet configurations (36), quantitative results by sample splitting are often not as good as by splitless and on-column techniques.

Splitless injection

Splitless injection utilizes a “solvent effect” (37) and allows a relatively large amount of dilute sample (1–5 μl) to be injected. The sample is vaporized and then carried onto the column on which it must be reconcentrated prior to analysis. This is essential in order to prevent band broadening. In order to prevent column overloading, the amounts of components being separated should be less than 50 ng. The large excess of solvent used to prepare the sample is backflushed 30–60 s after injection in order to minimize the occurrence of a long solvent tail, which can obscure any early eluting peaks. There are two mechanisms for reconcentrating the solutes at the head of the column. Grob and Grob (38) utilized the “solvent effect,” whereby the solvent acts as a barrier to the sample components, which facilitates their condensation and concentration at the head of the column. This is due to the fact that when the sample components encounter a liquid phase mixed with retained solvent, the front of the sample plug undergoes stronger retention than the rear of the plug. In order to minimize column deterioration that can result from solvent overloading, a solvent in which the liquid phase is not readily soluble should be used. Both dichloromethane and hexane have been widely used and care should be taken to see that the initial column temperature is 10–30°C below the boiling point of the solvent selected.

Another method of reconcentrating the components at the head of the column is to keep the column temperature low enough to condense the solutes (cold trapping) (35). A general guideline for the use of this precolumn concentration technique is that compounds with boiling points 100°C higher than the column temperature will be cold trapped. Therefore, splitless injection should be used when component concentrations are too low for detection by split injection ($< 0.1\%$ sample) or when only a very limited amount of sample is available.

On-column injection

When dealing with thermolabile compounds, vaporization of the sample can result in degradation during this process. Schomburg et al. (39) described an on-column injection technique whereby the sample never encounters temperatures higher than the column temperature. This method also has been shown to be extremely useful for the separation of compounds that have low volatility and samples that have a wide boiling range. Using very fine long fused silica needles attached to a microsyringe and inserted into the capillary column bore, on-column systems have been described (40). The fine needles are too fragile for normal septum piercing, hence other methods have been devised in which a septum-free valved inlet is used (41).

Inlet modifications that incorporated air-cooling of the column inlet were later designed to overcome vaporization in the needle that resulted from the slow injections necessary to achieve a narrow band of injected sample (42).

Automatic injection

The injection process has been automated, thus facilitating batch processing of large numbers of samples that are completely unattended. Various automatic injector systems are commercially available for use with both packed and capillary columns. These are mainly based on the use of syringe injection and pneumatically operated under microprocessor control. Injector loops are largely used for the introduction of gases into the column, while manual injection continues to be extensively used. In the latter instance, the operator's injection technique can dramatically influence the quality of the analysis (35). A skilled operator may achieve precisions of the order of $\pm 1\%$ with manual syringes by careful debubbling of the syringe, and using sample sizes at least 50% of the syringe capacity to minimize needle hold up and setting errors, as well as using a very reproducible injection technique. However, the advent of automatic samplers considerably enhances injection precision and accuracy. In addition, the tedious process of carefully cleaning and flushing the syringe

between samples to avoid cross-contamination during manual injections is readily accomplished automatically.

Oven and Temperature Controls

The column is usually suspended in an insulated thermostatically controlled air oven through which the air is very rapidly circulated by means of fans or pumps. This allows accurate temperature control to within 0.1°C and minimizes thermal gradients. Provision is also made for the temperature to be rapidly increased and for the equally rapid cooling required with temperature-programmed work.

Injection port temperatures for packed columns include thermostatically controlled heating. These parts should be hot enough to rapidly vaporize the sample in order to prevent a loss in efficiency from the injection process. Heated detector systems also are used depending on the type of detector, to prevent sample condensation, which inevitably will result in peak broadening and loss of component peaks. Temperature control imparts stability to the detection system, often reducing noise and enhancing the detection limit. When the FID is used, its temperature must be kept high enough to avoid any water or by-products formed during the combustion process.

Detectors

A large variety of detectors have been designed for use in GC. In addition to numerous publications, several books and reviews that discuss the design and operating principles of GC detectors have been published (43–46).

The chromatographic detector, placed at the column exit, constantly monitors the emitted gas, and generates an electrical signal that is amplified and appears as a plot of detector response versus time, i.e., the chromatogram. Detectors may be “universal,” which responds to every eluted component (TCD), “selective,” which responds only to certain functional or elemental characteristics of the analyte (ECD and FID), or “specific,” which provides qualitative information concerning the structure of the eluting component (FPD). However, for classification purposes, GC detectors generally fall into one of two groups: concentration-dependent and mass rate-dependent detectors. The former, which includes the TCD and ECD, produces a signal that is proportional to the concentration of the sample in the carrier gas. In the latter (e.g., FID), the detector signal is dependent on the mass of sample that flows through the detector per unit time (g/s). Some of the most important properties relating to detectors are: 1) their sensitivity, which is a function of the amount of

component present in the injected sample; 2) their signal noise, which refers to random, short-term detector response and which combined with sensitivity, determines the detection limit for a given component; 3) their linearity of response, which indicates the region over which the detector signal is directly proportional to sample concentration or mass flow rate. The dynamic linear range of the detector is the range of sample size for which a signal is detected as a linear function of the sample size. Thus, a wide linear range is useful for quantitative analysis of multicomponent mixtures. In contrast to short-term noise, which depends upon electrical factors, temperature sensitivity, or flow variations, long-term noise is manifested by baseline drift in the chromatogram.

Detectors commonly used in GC and specified in the USP (21) include FID, alkali FID (NPD, TD), ECD, and TCD. A description of these detectors, including their operational principles and relative performance, was presented in a previous volume of this encyclopedia (23). Various other useful detectors for GC include photoionization (PID), flame photometric (FPD), electrolytic conductivity (ELCD), redox (RCD) and sulfur chemiluminescence (SCD), and helium ionization (HID) (47). Table 1 summarizes some of the features of detectors used in GC.

In addition to the above, several newer and highly sophisticated detection techniques that involve the coupling of various types of spectrometers with GC have emerged. These “hyphenated” techniques include the on-line interfacing of mass spectrometers (GC–MS) (57–60), infrared spectrometers that incorporate Fourier transformation techniques (FTIR–GC) (61), and FTIR–GC–MS (62). Generally, these are considered specific detectors mainly used to obtain qualitative information, although quantitative data can be obtained when operating a GC–MS system in the selected-ion-monitoring mode (SIM). This mode, in contrast to the normal scanning mode used for qualitative purposes, allows a single or a few characteristic ions of an analyte to be monitored and subsequently determined quantitatively (63, 64). Triple–quadrupole MS/MS spectrometers are becoming more prevalent and these are being increasingly coupled to GC's that provide enhanced quantitation capabilities (60).

Columns

The suitability of a column for a particular use depends on various factors, such as stationary phase, solid support, column tubing material, inside diameter, percent liquid loading, and temperature. Columns may be prepared in

Table 1 Detector features

Detector	Classification	Response	Recommended carrier gas	Applications	Approximate detection capability (g)	Comments
TCD	Concentration	Universal	Hydrogen or Helium	Most compounds including water	$10^{-6} - 10^{-8}$	Although less sensitive than other detectors, it is useful for moisture determinations (48), can detect some compounds
FID	Mass-flow	Universal	Argon, Helium or Nitrogen	Most compounds excluding water	$10^{-10} - 10^{-11}$	unresponsive to FID and is non destructive and thus useful for preparative work Lack of response to water and carbon disulfide allows these to be used as solvents without interference with the analysis
ECD	Concentration	Selective	Hydrogen, Nitrogen or Argon/5–10% Methane	Compounds with high electron affinities, e.g. halogens	$10^{-12} - 10^{-14}$	Good for electronegative compounds such as those containing extensively conjugated pi-electron systems, nitro- and highly conjugated aromatic compounds and compounds which can be readily derivatized to respond (49). Response depends also upon carrier gas flow rate
NPD	Mass-flow	Selective	Hydrogen, Helium or Nitrogen	Compounds containing nitrogen or phosphorus	$10^{-12} - 10^{-13}$	Can be used in the flame or flameless mode. In the latter instance, a very low fuel (hydrogen) flow is used to form a plasma around a heated bead of potassium or rubidium salts. This results in a reduced response to hydrocarbons and subsequently less interference. Halogens as well as organolead compounds respond to the NPD detector in the flame mode. Phosphates (from cleaning detergents), chlorinated solvents and silanizing reagents can deplete the alkali beads and should thus be avoided (50)
PID	Concentration	Universal	Helium or Nitrogen	Compounds with ionization potentials <12eV	2×10^{-12}	Carrier gas must be free of oxygen and hydrocarbons to avoid a loss in sensitivity and interference, respectively. Water, sulfur dioxide, saturated hydrocarbons smaller than hexane, chloroform, methylene chloride, ethylene chloride and acetonitrile are not detected. The detector is non-destructive and can detect low levels of inorganic compounds. (51)

(Continued)

Table 1 Detector features (Continued)

Detector	Classification	Response	Recommended carrier gas	Applications	Approximate detection capability (g)	Comments
FPD	Mass-flow	Selective	Hydrogen or Nitrogen	Phosphorus and sulfur compounds	2×10^{-10}	Operated with a flame, photomultiplier tube and either a 393 or 526 nm bandpass filter for sulfur or phosphorus detection, respectively (52)
ELCD	Concentration	Selective	Helium or Nitrogen	Sulfur, nitrogen and halogen compounds	10^{-12} – 10^{-13}	Following pyrolysis of column effluent, unwanted species are removed and the relevant products mixed with a liquid (alcohol/water) and differential conductivity subsequently measured. Depending upon the reaction gas used and whether the process is reductive or oxidative, specificity for nitrogen, sulfur or halogen compounds can be obtained (53)
RCD	Concentration	Selective	Helium	Compounds that react chemically to emit photons	2×10^{-11}	As with the ELCD, post-column treatment of the effluent is involved. The RCD responds to compounds that serve as reducing agents, such as alcohols, aldehydes, olefins, and carboxylic acids, and it is insensitive to many potentially interfering compounds such as water, oxygen, hydrocarbons and carbon dioxide (54)
SCD	Concentration	Specific	Helium	Compounds containing a sulfur—carbon bond	10^{-11} – 10^{-12}	Amenable sulfur compounds are reacted with fluorine to produce HF and the chemiluminescence is subsequently measured. Saturated hydrocarbons, methylene chloride, acetonitrile, methanol and carbon tetrachloride give little or no response (55)
HID	Mass-flow	Universal	Helium	Trace impurities in bulk gases and liquids	10^{-12} – 10^{-14}	The helium carrier gas must be at least 99.9999% pure for optimum sensitivity. Useful to detect nitrogen oxides, sulfur gases, alcohols aldehydes, ketones, hydrocarbons and water (56)

various lengths and diameters depending on the particular objective. Preparative columns may range from 0.95 to 10 cm (3/8–4') in diameter or larger for the collection of quantities of individual components when volumes between 0.5 ml and more are injected. Analytical (or packed) columns generally have outside diameters of 3, 4.7, or 6.25 mm (1/8, 3/16, or 1/4'), and inside diameters of 1–4 mm, while capillary columns with very narrow inside diameters are used for applications that require very high resolution. The inside diameter of the tubing is, in fact, one of the most critical column dimensions in determining the efficiency of separation.

Although packed GC columns may be made from various materials, such as glass, nickel, stainless steel, copper, aluminum, or even Teflon[®], the USP (21) and BP (22) recommend that glass or stainless steel columns be used for pharmaceutical analyses unless otherwise specified. The advantage of using glass lies in its relative inertness as compared to metal columns, although its fragile nature is certainly a disadvantage. In order to assure further the inertness of glass, silanization of the inside walls with 5–10 vol% dimethyldichlorosilane in toluene is often performed (65).

Capillary columns are usually fabricated from fused silica, with a polyamide outer coating to impart flexibility and reduce breakage during handling. These columns can be classified into three categories according to the size of the internal diameter. Typical inside diameters are 0.53 mm, 0.32–0.22 mm, and 0.2–0.1 mm for megabore (wide-bore), normal bore (high-resolution), and microbore (high-speed), respectively.

The selection of a stationary phase is extremely important in GC since it is the major controllable variable of selectivity in the separation process. Stationary phases can be nonpolar, polar, or of intermediate polarity. Cyclodextrins, cyclic oligosaccharides composed of varying numbers of glucopyranose units, were found recently to be extremely useful for the separation of chiral compounds (66). Three types of derivatives, 5-hydroxypropyl (hydrophilic), dialkyl (hydrophobic), and trifluoroacetyl (intermediate) have been used, each of these phases having a selected area of specificity.

Capillary columns offer many advantages in terms of speed of analysis, high resolution, and overall very high separation efficiency. New applications that involve the use of capillary columns are included in the USP (21). In particular, methods for OVI analysis prescribe, almost exclusively, capillary columns, whereas approximately 20% of the other GC methods also prescribe such columns.

Solid supports should be chemically inert and exhibit a large surface area. Support materials used are diatomaceous earths, Teflon[®], glass beads, and various polymers.

Since the surface of the diatomaceous materials consist of silanol (Si—O—H) and siloxane (Si—O—Si) groups, compounds capable of hydrogen bonding (alcohols, acids, amines, etc.) can interact with these media, which results in tailing. This problem may, however, be minimized by silanization (65) following extensive acid washing to remove inorganic impurities. Acid-washed flux-calcined diatomaceous earth is often used for drug analysis.

Specially treated glass beads have been used as supports for high-molecular-weight compounds. The beads are usually etched and silanized prior to coating with the liquid phase. Teflon[®] is a useful support material for the analysis of short chain polar substances, which tail on diatomaceous supports, and is particularly indicated for the analysis of corrosive substances such as halogenated acids.

When a separation can be effected by a purely adsorptive mechanism (GSC), various adsorbants and porous polymers are used. Commercially available adsorbants include silica gel, activated charcoal, and molecular sieve materials. These adsorbants are used mainly for the analysis of gases and low-molecular-weight, low-boiling compounds (67).

Porous polymers also are useful for gas analyses and for very polar molecules, such as amines, glycols, and acids. Copolymers of styrene and divinylbenzene and others, such as ethylvinylbenzene–divinylbenzene, cross-linked acrylic ester, vinylpyridine, pyrrolidone, and ethylene glycol dimethacrylate, are commercially available (29).

Open tubular columns are simply capillary tubes in which the inside of the column wall is used as the support for the liquid phase. These wall-coated open tubular columns (WCOT) have the stationary phase distributed in the form of a thin film on the inside surface of the open capillary tube, the walls thus serving as the support. In order to reduce the thickness of the liquid phase film, a porous layer may be formed on the inside wall of the capillary tubing and then coated with the liquid phase to produce a support-coated open tubular column (SCOT). Porous-layer open tubular columns (PLOT) are similar to SCOT columns, the difference being that in the former, the stationary phase is deposited on fine crystalline particles or glass powder which is adsorbed onto the walls of the tube. In both cases, the available surface area of the wall is increased, and allows an increased amount of liquid phase to be accommodated in the same length and diameter of tubing. The whisker-walled (WW) column consists of whiskers chemically etched on the surface of the wall, which also result in a significant increase in the available surface area. Wall-coated, porous-layer, and support-coated capillary columns are all available as whisker-walled, i.e., WWCOT, WWPLOT, and WWSCOT, respectively.

The stationary phase film thickness of capillary columns range from about 0.1–10 μm and can be divided into three film thickness ranges. Thin-film columns are usually 0.1–0.2 μm and offer the greatest stability. They have smaller sample capacity as compared to the thicker films but are the best for use with high temperatures. Thick films are usually 0.6–10 μm and allow higher sample loading, better retention for volatile compounds, and a high degree of inertness. Their main drawback, however, is larger bleed at high temperature as compared to the thin film type. The medium film thickness is about 0.3–0.6 μm and is a useful compromise in terms of sample capacity, retention properties, and phase stability.

In addition to the open tubular capillary columns, packed capillaries, and even micropacked capillaries, are commercially available. These columns contain support material, have internal diameters that range from 0.6–1.0 mm, and have the main advantage of being able to handle larger sample loadings. This is useful since direct analysis, as opposed to split-analysis, can be used with very short columns that operate at relatively high efficiency.

Column lengths of 0.3–6.1 m (1–20 ft) are commonly used and configurations can be straight, U-shaped, spiral, or flat coils. Straight and U-shaped columns are purported to be slightly more efficient than the coiled types. However, the dimensions of the GC oven usually dictate the choice. Capillary columns are generally much longer than packed columns and range in lengths from 10–100 m for the open tubular type and 1–6 m for packed type.

SEPARATION TECHNIQUES

Gas chromatographic analysis can be performed at constant temperature (isothermal mode) or with the column temperature increasing with time (temperature programming).

Isothermal

In isothermal GC, peak width increases linearly with retention time (68) and retention time increases exponentially with carbon number in a homologous series. In contrast, the peak width remains constant in programmed-temperature gas chromatography (PTGC) and retention times increase only linearly with carbon number in a homologous series (69).

Gas chromatographic methods for the analysis of impurities in pharmaceutical dosage forms and raw materials generally make use of isothermal techniques

where the temperature of the instrument is maintained constant throughout the run. However, when complex mixtures that contain components with widely different boiling ranges need to be analyzed, PTGC is undoubtedly the method of choice. Separation of components in such a mixture may prove difficult if at all possible under isothermal conditions. Using a high temperature for the analysis may result in poor resolution between the rapidly eluting volatile components. On the other hand, operating at a lower temperature could cause excessively long retention times for the less volatile compounds, with resulting peak broadening leading to poor detectability and prolonged running times.

Temperature Programming

The advent in 1952 of PTGC (70) and the subsequent introduction of commercial equipment for temperature programming provided the necessary means to analyze complex mixtures that contained components of widely differing boiling points and solved some of the problems previously described in the section dealing with isothermal separations.

The PTGC technique involves increasing the column temperature at a preset rate during the elution process. This rate may be constant throughout the run, or periods of isothermal operation may be automatically programmed at set times between temperature increases. Generally, the electronically controlled ovens are designed to increase temperature at rates from 0.5–30°C per minute. The initial temperature should be chosen to minimize the retention time for the least retained solute, while the final temperature must be sufficient to elute the least volatile compound in a reasonable time. The instrument then automatically resets the temperature to the initial value in preparation for the next sample.

A major problem with PTGC, however, is that column bleed may cause baseline perturbations as the temperature increases, which results in interference with the analysis. Compensation for this effect is usually accomplished by using a dual column/dual detector system or replacing the liquid stationary phase with another less volatile coating. In the former instance, the output signal from the reference column is used to cancel out the bleed from the analytical column. Electronic compensation using single-column, single-detector systems are also available, as are GC's with electronically controlled programmed gas flows. In the latter case, the carrier flow rate is increased during the analysis, which results in reduced baseline drift by avoiding or reducing column bleed. Since lower temperatures can be used, the analysis of thermolabile

compounds is facilitated and a wider range of liquid phases can be used.

Special Techniques

Various compounds do not readily lend themselves to analysis by GC by virtue of several factors, such as nonvolatility, instability, elicitation of poor detector response, or high adsorptive properties (presence of polar groups). These problems sometimes can be overcome by the use of pyrolysis or by derivatization. The former technique involves high temperature decomposition of high-molecular-weight, nonvolatile substances to lighter, and more volatile compounds.

Derivatization is a valuable aid in GC. Suitable derivatives may be produced using synthetic organic reactions such as esterification, acylation, and silylation (29). These methods serve to increase thermal stability in unstable compounds, improve detectability in some instances (e.g., derivatives for electron capture detection) and often sensitivity, improve volatility in instances where the parent compound is relatively non-volatile, and mask polar groups to reduce adsorption. Several comprehensive reports and reviews on derivatization have been published over the years (71–74).

Quantitative and Qualitative Analysis

GC constitutes an analytical technique whereby the separation of components in a mixture and their quantitative and qualitative assessment can be performed simultaneously.

The area under the component elution peak is proportional to the concentration of that particular component. Various methods can be used to measure this area and are based on the assumption that the shape of the peak is Gaussian. Electronic integration is the preferred method since very accurate and precise measurements are obtained this way ($RSD \leq 0.5\%$).

Peak heights may be used but are less reliable since any variability in temperature and/or flow rate will affect this measurement. Peak areas may also be calculated by multiplication of the peak height and the width at half-height. This gives a value equal to 84% of the true area (75). Triangulation, planimetry, and cutting and weighing also can be used to measure peak areas.

In order to minimize the possibility of errors as a result of variable injection volumes, the internal standard (IS) method should be used. It involves the addition of a

compound, the IS, which is not already present in the sample. This is normally a substance that elutes at a position near the sample component of interest and should be well resolved and readily detected under the given chromatographic conditions. The IS is added in constant amount to solutions that contain varying amounts of the analytical standard. Calibration curves are then constructed by plotting the ratio of either the areas or peak heights of the two peaks (analyte/IS) versus amount or concentration of analyte. The amount or concentration of each chromatographed sample can then be obtained by interpolation of the calibration curve. The calibration plot should be a straight line with an intercept of zero. However, nonlinear standard curves may result when the linear range of the particular detector is exceeded due to excess analyte concentration. The presence of high background and/or interfering compounds results in plots that have a positive intercept, while negative intercepts are usually indicative of sample loss during handling.

In contrast to the IS method, external standardization may be used in which several standard solutions of varying concentrations of the sample are prepared. Following constant volume injection of each standard solution, a plot of peak area (or height) versus concentration is made, and unknown sample concentrations are obtained from interpolation of the calibration curve. The success of this technique, however, is dependent upon the precision of injection volume, readily accomplished with automatic injection but less so when manual microliter syringes are used.

A further quantitative measure, normalization, is sometimes utilized to determine the proportion of one or more components in a mixture. This involves calculating the ratio of the individual component peak area to the sum of the areas of all component peaks in the chromatogram. It assumes that all the components have identical response factors to the detector. This is a reasonable assumption when all the components of the mixture are chemically similar. When structurally dissimilar components are analyzed, a response factor correction should be used by measuring the peak area for a known quantity of pure material and calculating the respective response factor (F) from:

$$F = \frac{\text{Component amount}}{\text{Peak area}} \quad (8)$$

Acquisition of qualitative data may be obtained by using specific detectors such as GC–MS, FTIR–GC, and FTIR–GC–MS, as well as from relative retention times or by the concept of retention index (vide infra).

Headspace Analysis

Specifications for pharmaceutical compounds, in addition to the actual active ingredient itself, include information and limits relating to the intermediates, residual solvents, and other volatile impurities (21) associated with their preparation.

Headspace analysis is a very effective technique for the analyses of volatile compounds, and is particularly valuable when direct injection would ruin the column due to corrosive or highly nonvolatile components present in the sample. Headspace analysis obviates extensive sample preparation, eliminates the possibility of unwanted component interference, and avoids degradation of susceptible components in the injection port or on the column (76).

The liquid or solid sample is placed in a vial, which is sealed with a septum and heated to a predetermined temperature for a period of time. Equilibrium between the sample and vapor phase is then established and a portion of the volatiles in the gas phase (headspace) is subsequently injected onto the column. Several different methods have been used to transfer headspace volatiles into the GC, from manual withdrawal that uses a gas syringe, to sophisticated automatic sampling that involves transfer lines, and valves that lead directly onto the column.

Two main techniques, however, may be used. These are static and dynamic sampling. The static method allows the temperature of the sample container to be held for a sufficiently long period to allow the gas-phase and sample-phase to equilibrate. The dynamic sampling uses an inert gas that is swept over or through the thermostated sample for a period of time sufficient to extract most or all of the volatile components. A general chapter on OVI's, in which five OVI's are specifically mentioned (chloroform, dichloromethane, benzene, trichloroethylene, and dioxane), was introduced into the USP (21). Previously, the USP included methods II and III, which provided an alternative to direct injection and specified dynamic headspace sampling. These, however, have now been excluded and method IV provides for the general use of static headspace sampling (77). Method VI, while still included, prescribes analysis by direct injection into a gas chromatograph, but none of the monographs in the USP or National Formulary (NF) specify use of method VI for OVI testing. The trend, it appears, has been toward using method IV as a replacement for other methods in individual monographs and that monographs that now specify method IV, originally specified method VI (78). The advantages, sampling techniques, and problems associated with headspace sampling were previously discussed by Hinshaw (79).

PHARMACEUTICAL APPLICATIONS

Applications of GC to the analysis of pharmaceuticals have been described in several books, compendia, and the review of Jack (80) to which the reader is referred. In addition, the reader might want to review the three-part multi-authored series (81) that includes HPLC determinations, as well as the comprehensive listing of published chromatographic methods in a book edited by Adamovics (82).

Although HPLC has largely superseded GC as the compendial chromatographic method of choice for the assay of pharmaceuticals, the application of GC continues to be an important and valuable analytical method for monitoring certain impurities and for the determination of various related substances and OVI's in many pharmaceutical dosage forms, as well as in raw materials. GC, furthermore, continues to be a valuable tool for the analysis of drugs in biological fluids for the purposes of therapeutic drug monitoring, pharmacokinetic studies, and in the assessment of bioavailability/bioequivalence.

A large number of pharmaceutical compounds can, however, be analyzed by either GC or HPLC. In terms of cost effectiveness, given that the equipment is available, GC often may be preferred due to its advantage of avoiding the use of expensive solvents and associated subsequent disposal problems.

Raw Materials (Bulk Drugs) and Dosage Forms

Both the USP/NF (21) and the BP (22) utilize GC for the following types of determinations:

1. Assay.
2. Chromatographic purity.
3. Identification.
4. Presence of volatile matter, intermediates, and related substances.
5. Water.
6. Presence of isomers, isomeric purity, and racemate ratios.
7. Alcohol content.

Tables 2–11 list the various compounds and the associated USP tests, together with their relevant chromatographic conditions. Drugs and related dosage forms have been arranged alphabetically, while descriptions of the various GC supports and liquid (coating) phases largely correspond to the abbreviations used in the USP. These are given starting on page 409:

Table 2 Compendial applications of GC for the assay of pharmaceutical raw materials and dosage forms

Material/dosage form	Column	Carrier gas	Temp. (°C)	Detector	Internal standard	Reference
Acetone	1.8 M × 3 mm I.D. S4	Helium	110–220 8/min	FID	None	NF (19, p. 2409)
Acetyltributyl Citrate	30 M × 0.32 mm I.D. Column Bonded with G42 (0.5µm)	Helium	80–220 20/min	FID	None	NF (19, p. 2409)
Amantadine Hydrochloride capsules	Glass, 1.22 M × 2 mm I.D. 10% G1/S1A	None	115	FID	Naphthalene	USP (24, pp. 103 and 104)
Amantadine Hydrochloride syrup	(100–120 mesh)	Specified				
Amitraz Amitraz concentrate for dip	1.5 M × 4 mm I.D. 3% G1/S1A	Nitrogen	250	FID	Squalane	USP (24, p. 121)
Amylene Hydrate	Glass, 2 M × 4 mm I.D. S2	Helium	190	TCD	None	NF (19, p. 2414)
Atropine Sulphate ophthalmic ointment	Glass, 1.8 M × 2 mm I.D. 3% G3/S1AB	Nitrogen	225	FID	Homatropine	USP (24, p. 179)
Atropine Sulphate ophthalmic solution					Hydrobromide	USP (24, p. 179)
Atropine Sulphate tablets	Fused Silica 25 M × 0.32 mm Capillary coated with G1	Helium	200–280 4/min	FID	None	USP (24, p. 181)
Avobenzene	Glass, 1.2 M × 4 mm I.D. 3% G3/S1AB	Helium	215	FID	Homatropine	USP (24, pp. 199, 200 and 201)
Belladonna extract						
Belladonna extract tablets						
Belladonna leaf						
Belladonna tincture						
Benzocaine and Menthol topical aerosol	1.8 M × 2 mm I.D. 10% G16/S1AB (for Menthol)	Helium	170	FID	Decanol	USP (24, p. 208)
Benzyl Alcohol	Glass or Stainless Steel, 1.8 M × 3 mm I.D. 5% G16/S1	Helium or Nitrogen	140	FID	Phenol	USP (24, p. 1865)
Butabarbital	1.8 M × 4 mm I.D. 10% G37/S1AB		260	FID	Tetracosane	USP (24, p. 261)
Butabarbital Sodium elixir	Glass, 0.9 M × 4 mm I.D. 3% G10/S1A (80–100 mesh)	Nitrogen	200	FID	Secobarbital	USP (24, p. 262)
Butabarbital Sodium tablets						
Butane	Aluminum, 6 M × 3 mm I.D. 10% liquid G 30/S1C	Helium	33	TCD	None	NF (19, p. 2422)
Burylated Hydroxyanisole	Stainless Steel, 1.8 M × 2 mm I.D. 10% liquid G26/S1A	Helium	175–185	FID	4-tert-butylphenol	NF (19, p. 2422)
Castor Oil emulsion	Glass, 1.8 M × 4 mm I.D. G25/S1	Helium	245	FID	Di(2-ethylhexyl) phthalate	USP (24, p. 323)
Cetostearyl Alcohol	2 M × 3 mm I.D. 10% liquid G2/S1	Helium	205	FID	None	NF (19, p. 2434)

(Continued)

Table 2 Compendial applications of GC for the assay of pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage form	Column	Carrier gas	Temp. (°C)	Detector	Internal standard	Reference
Cetyl Alcohol	2 M × 3 mm I.D. 10% liquid G2/S1A	Helium	205	FID	None	NF (19, p. 2435)
Chlorobutanol	Glass or Stainless Steel, 1.2 M × 3 mm I.D. 5% G16/S1A	Helium or Nitrogen	110	FID	Benzaldehyde	USP 24, p. 1865
Chloroxylenol	Glass, 1.8 M × 4 mm I.D. 3% G16/S1A	Nitrogen	210	FID	None	USP (24, p. 391)
Clindamycin Palmitate Hydrochloride	Glass, 0.6 M × 3 mm I.D. 1% G36/S1AB (80–100 mesh)	Helium	290	FID	Cholesteryl Benzoate	USP (24, pp. 431 and 432)
Clindamycin Palmitate Hydrochloride fororal solution						
Clioquinol	Glass, 1.83 M × 2 mm I.D. 3% G3/S1AB (80–100 mesh)	Helium	165	FID	Pyrene	USP (24, pp. 435 and 436)
Clioquinol cream						
Clioquinol ointment						
Clioquinol and Hydrocortisone cream						
Clioquinol and Hydrocortisone ointment						
Cyclomethicone						
Desflurane	3.66 M × 3 mm I.D. 20% liquid G1/S1A (60–80 mesh)	Helium	125–320	8/min	None	NF (19, p. 2443)
Dibutyl Sebaccate	Stainless Steel, 3.7 M × 2.4 mm I.D. 10% G31 and 15% G18/S1A	Helium	80–88	2/min	Halothane	USP (24, p. 504)
Dichlorodi-fluoromethane	1.8 M × 2 mm I.D. 10% liquid G41/S1A (100–120 mesh)	Helium	150–280	4/min	None	NF (19, p. 2445)
	Stainless Steel, 1.8 M × 2 mm I.D. 1% G25/S12	Helium	70–170	10/min	None	NF (19, pp. 2445 and 2446)
Dichlorotetrafluoroethane						
Dicyclomine Hydrochloride capsules	Fused Silica 15 M × 0.5 mm coated with G3 (1 μm)	Nitrogen	160° – 240°	20°/min	Phenacetin	USP (24, pp. 549, 550 and 551)
Dicyclomine Hydrochloride injection						
Dicyclomine Hydrochloride syrup	30 M × 0.32 mm I.D. fused Silica column with G46(1 μm)	Helium	120–225	12/min	None	NF (19, p. 2447)
Dicyclomine Hydrochloride tablets	1.8 M × 4 mm I.D. 15% G39/S1A	Helium	160	FID	Dimethyl-formamide	USP (24, p. 580)
Dimethyl Glycol Monoethyl ether						
Dimethyl Sulfoxide gel						
Dimethyl Sulfoxide irrigation	Glass, 1.5 M × 3 mm I.D. 10% liquid phase G25 on packing S1A	Helium	100–170	10/min	Dibenzyl	USP (24, p. 581)

(Continued)

Table 2 Compendial applications of GC for the assay of pharmaceutical raw materials and dosage forms (Continued)

Material/dosage form	Column	Carrier gas	Temp. (°C)	Detector	Internal standard	Reference
Dimethyl Sulfoxide topical solution	1.8 M × 4 mm I.D. 10% liquid phase G16 on packing S1A	Helium	170	FID	Dimethyl-formamide	USP (24, p. 581)
Diphenoxylate Hydrochloride and Atropine Sulfate oral solution	Glass, 1.2 M × 4 mm I.D. 3% G3/S1 (for Atropine Sulfate)	Helium	230	FID	Homatropine Hydrobromide	USP (24, p. 585)
Enflurane	Stainless Steel, 3 M × 4 mm I.D. 20% G4/S1A (60–80 mesh)	Helium	60–125 6/min	TCD	None	USP (24, p. 642)
Conjugated Estrogens	15 M × 0.25 mm I.D. fused Silica capillary bonded with G19 (0.25 µm)	Hydrogen	220	FID	3-O-Methylsterone	USP (24, pp. 681, 682, 683 and 684)
Conjugated Estrogen tablets						
Esterified Estrogens						
Esterified Estrogen tablets						
Ethchlorvynol	Glass (pre-treated with 10% dimethyldi chlorosilane in toluene), 1.8 M × 4 mm I.D. 10% G16/S1AB (60–80 mesh)	Helium	160	TCD	None	USP (24, p. 691)
Ethchlorvynol capsules	Glass (pre-treated with 10% dimethyldi chlorosilane in toluene), 1.8 M × 4 mm I.D. 10% G16/S1AB (60–80 mesh)	Helium	160	FID	Chlorobutanol	USP (24, p. 691)
Eucalyptol	60 M × 0.32 mm I.D. Fused Silica capillary G16	Helium	60–200 6/min	FID	None	USP (24, p. 705)
Guaifenesin and Codeine Phosphate syrup	0.6 M × 2 mm I.D. 3% liquid G3/S1A (100–120 mesh) (For Codeine Phosphate)	Helium	210	FID	Hydrocodone Bitartrate	USP (24, p. 793)
Guaifenesin and Codeine Phosphate syrup	1.2 M × 4 mm I.D. 3% liquid G6/S1A (100–120 mesh) (For Guaifenesin)	Helium	170	FID	Dipropyl Phthalate	USP (24, p. 793)
Homosalate	30 M × 0.53 mm I.D. G27 (1 µm)	Hydrogen	70–220 6/min	FID	None	USP (24, p. 816)
Hydroxypropyl Methylcellulose	Glass, 1.8 M × 4 mm I.D. 20% G28/S1C (100–120 mesh)	Helium	130	TCD	Toluene	USP (24, p. 843)
Hyoscyamine tablets	Glass, 1.8 M × 2 mm I.D. 3% liquid G3/S1AB	Nitrogen	225	FID	Homatropine Hydrobromide	USP (24, p. 850)
Hyoscyamine Sulphate elixir	Glass, 1.8 M × 2 mm I.D. 3% liquid G3/S1AB	Nitrogen	225	FID	Hydrobromide	USP (24, pp. 851, 852 and 853)
Hyoscyamine Sulfate injection						
Hyoscyamine Sulphate oral solution						
Hyoscyamine Sulphate tablets						
Isobutane	Aluminum, 6 M × 3 mm I.D. 10% liquid G30 on nonacid washed S1C	Helium	33	TCD	None	NF (19, p. 2467)

(Continued)

Table 2 Compendial applications of GC for the assay of pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage form	Column	Carrier gas	Temp. (°C)	Detector	Internal standard	Reference
Isoamyl Methoxycinnamate	25 M × 0.32 mm I.D. coated G1 (0.1 μm)	Helium	60–240 8/min	FID	None	USP (24, p. 918)
Isoflurane	Nickel or Stainless Steel, 3.7 M × 2.4 mm I.D. 10% G31 + 15% G18 on S1C (60–80 mesh)	Helium	65–110 4/min	FID	Butyl Acetate	USP (24, p. 920)
Isoflurophate Ophthalmic solution	Glass, 1.8 M × 4 mm I.D. 5% G33/S1AB (80–100 mesh)	Helium	75–80	FID	Cyclohexanone	USP (24, pp. 921 and 922)
Isopropyl Alcohol	Stainless Steel, 1.8 M × 6.4 mm I.D. 10% liquid G20/S1A	Helium	55	TCD	None	USP (24, p. 927 and NF 19, p. 2468)
Isopropyl Myristate	1.8 M × 4 mm I.D. 10% liquid G8/S1 (100–120 mesh)	Nitrogen	90–210 2/min	FID	None	NF (19, p. 2468)
Isopropyl Palmitate	1.8 M × 4 mm I.D. 10% liquid G8/S1 (100–120 mesh)	Nitrogen	90–210 2/min	FID	None	NF (19, p. 2468)
Isosorbide	Glass, 0.6 M × 3 mm I.D. S9	Nitrogen	230	FID	Triethylene Glycol	USP (24, pp. 934 and 935)
Isosorbide oral solution						
Lindane cream	Glass, 1.8 M × 3 mm I.D. G3/S1A	Nitrogen	195	FID	n-Docosane	USP (24, pp. 976 and 977)
Lindane lotion						
Lindane shampoo						
Malathion lotion	Glass, 1.8 M × 2 mm I.D. 5% liquid G6/S1A (110–120 mesh)	Nitrogen	190	FID	Parathion	USP (24, p. 1012)
Menthol lozenges	30 M × 0.53 mm I.D. fused Silica coated with G16 (μm)	Helium	125	FID	Anethole in Hexanes	USP (24, p. 1038)
Menthyl Anthranilate	25 M × 0.32 mm I.D. Column coated with G1 (0.1 μm)	Helium	60–240 8/min	FID	None	USP (24, p. 1039)
Methadone Hydrochloride injection	Glass, 1.2 M × 4 mm I.D. 3% G2/S1A (10–20 mesh)	Helium	170	FID	Procaine	USP (24, p. 1057)
Methohexital Sodium for injection	1.2 M × 4 mm I.D. 3% G10/S1AB	Helium	230	FID	Aprobarbital	USP (24, p. 1069)
Methoxyflurane	Stainless Steel, 3 M × 4 mm I.D. G11/S1A	Helium	100–110	TCD	None	USP (24, p. 1074)
Methyl Alcohol	Stainless Steel, 2 M × 3 mm O.D. S4 (50–80 mesh)	Nitrogen	140	FID	None	NF (19, p. 2479)
Methyl Benzylidene Camphor	30 M × 0.32 mm I.D. fused Silica capillary coated with G1 (0.25 μm)	Helium	100–230 10/min	FID	None	USP (24, p. 1077)

(Continued)

Table 2 Compendial applications of GC for the assay of pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage form	Column	Carrier gas	Temp. (°C)	Detector	Internal standard	Reference
Methylcellulose	1.8 M × 4 mm I.D. 10% liquid G1/S1A (100–120 mesh)	Helium	100	FID	Toluene	USP (24, p. 1079)
Methylene Chloride	1.8 M × 4 mm I.D. 15% liquid G18/Unsilanized SIC (30–60 mesh)	Helium	60	TCD	None	NF (19, p. 2480)
Methylparaben, Ethylparaben, Propylparaben, Butylparaben	Glass or Stainless Steel, 1.8 M × 2 mm I.D. 5% G2/S1A	Helium or Nitrogen	150	FID	Benzophenone	USP (24, p. 1865)
Mibolerone oral solution	0.61 M × 3 mm I.D. 1% liquid G6/S1AB	Helium	175	FID	1,3,5-Triphenylbenzene	USP (24, p. 1112)
Miconazole Nitrate cream	1.2 M × 2 mm I.D. 3% G32/S1A	Helium	250	FID	Cholestane	USP (24, pp. 1114 and 1115)
Miconazole Nitrate topical powder						
Miconazole Nitrate vaginal suppositories						
Nitrogen Nitrogen 97 percent	3 M × 4 mm I.D. synthetic Alkali metal Aluminosilicate as a molecular sieve	Helium	Controlled	TCD	None	NF (19, p. 2485)
Octocrylene	60 M × 0.32 m I.D. column coated with G1 (0.25 µm)	Helium	80–280 4/min	FID	None	USP (24, p. 1213)
Octyldodecanol	2 M × 2 mm I.D. 3% G2/S1A	Nitrogen	80–300 6 min	FID	None	NF (19, p. 2286)
Octyl Methoxycinnamate	25 M × 0.32 mm I.D. column with G1	Helium	None	FID	Benzyl Benzoate	USP (24, p. 1213)
Octyl Salicylate	25 M × 0.32 mm I.D. column coated with G1 (0.1 µm)	Helium	60–240 8/min	FID	None	USP (24, p. 1214)
Oxandrolone tablets	Glass, 2 M × 4 mm I.D. 3% Methyl Silicone oil on 80- to 10- mesh acid-, base- and water-washed, flux calcined, silanized sliceaceous earth	Helium	250	FID	n-Octacosane	USP (24, p. 1225)
Penicillin G Procaine, Dihydrostreptomycin Sulfate, Chlorpheniramine Maleate, and Deamethasone injectable suspension (for Chlorpheniramine)	Glass, 1.8 M × 4 mm I.D. 1.2% G16 amd 0.5% KOH/S1A (10–120 mesh)	Nitrogen	180	FID	Brompheniramine Maleate	USP (24, p. 1281)
Pentobarbital elixir	Glass, 0.9 M × 4 mm I.D. 3 % liquid G10/S1A (80–10 mesh)	Nitrogen	190–210	FID	n-Tricosane	USP (24, p. 1293)
Pentobarbital Sodium capsules	Glass, 0.9 M × 4 mm I.D. 3 % liquid G10/S1A (80–10 mesh)	Nitrogen	190–210	FID	n-Tricosane	USP (24, p. 1295)
Perflubron	60 M × 0.25 mm I.D. G2 (1 µm)	Hydrogen	35–185 20/min	FID	None	USP (24, p. 1295)
Phendimetrazine Tartrate capsules	Glass, 1 M × 1 mm I.D. 3% G3/S1A	Helium	160	FID	Benzocaine	USP (p. 1055)

(Continued)

Table 2 Compendial applications of GC for the assay of pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage form	Column	Carrier gas	Temp. (°C)	Detector	Internal standard	Reference
Phendimetrazine Tartrate tablets	Glass, 1.2 M × 4 mm I.D. 10% liquid G3/S1AB	Helium	260	FID	Androsterone	USP (24, pp. 1392 and 1393)
Primidone oral suspension						
Primidone tablets						
Propane	Aluminum, 6 M × 3 mm I.D. 10% liquid G30 on non-acid washed S1C	Helium	33	TCD	None	NF (19, p. 2505)
Propoxyphene Hydrochloride, Aspirin and Caffeine capsules (for Propoxyphene and Caffeine)	0.6 M × 3 mm I.D. 3% Methyl Phenyl Silicone liquid/80–10 mesh siliceous earth	Nitrogen	175	FID	n-Tricosane	USP (24, p. 1423)
Propoxyphene Napsylate oral suspension	Glass or Stainless Steel, 0.6 M × 3 mm I.D. 3% G2/S1AB	Helium	160	FID	n-Tricosane	USP (24, p. 1425)
Propoxyphene Napsylate tablets						
Propoxyphene Napsylate and Aspirin tablets (for Propoxyphene)	1.2 M × 3 mm I.D. 3% G3/ S1A	Nitrogen	175	FID	n-Tricosane	USP (24, p. 1427)
Propylene Glycol	Glass or Stainless Steel, 1 M × 4 mm I.D. 5% G16/S5	Helium	122–200	TCD	None	USP (24, p. 1434)
Scopolamine Hydrobromide injection	Glass, 1.8 M × 2 mm I.D. 3% liquid G3/S1AB	Nitrogen	225	FID	Homatropine Hydrobromide	USP (24, pp. 1507, 1508 and 1509)
Scopolamine Hydrobromide ophthalmic solution						
Scopolamine Hydrobromide ophthalmic ointment						
Scopolamine Hydrobromide tablets						
Secobarbital elixir	Glass, 0.9 M × 4 mm I.D. 3% G10/S1A (80–100 mesh)	Nitrogen	200	FID	Butabarbital	USP (24, pp. 1509 and 1510)
Secobarbital Sodium capsules						
Secobarbital Sodium and Amobarbital	Glass, 0.6 M × 3.5 mm I.D. 3% liquid G10/S1AB (100–120 mesh)	Helium	175	FID	Aprobarbital	USP (24, p. 1511)
Sodium capsules						
Spectinomycin Hydrochloride	Glass, 0.6 M × 3 mm I.D. 5% G27/S1AB (80–100 mesh)	Helium	190	FID	Triphenyl-lantimony	USP (24, p. 1545)
Spectinomycin for injectable suspension	Glass, 0.6 M × 3 mm I.D. 5% G27/S1AB (80–100 mesh)	Helium	190	FID	Hexamethylsilazane	USP (24, p. 1545)
Stearic Acid	Glass, 1.5 M × 3 mm I.D. 15% G4/S1A	Helium	165	FID	None	NF (19, p. 2525)

(Continued)

Table 2 Compendial applications of GC for the assay of pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage form	Column	Carrier gas	Temp. (°C)	Detector	Internal standard	Reference
Purified Stearic Acid	2 M × 3 mm I.D. 10% G2/S1A	Helium	205	FID	None	NF (19, p. 2526)
Stearyl Alcohol						USP (24, pp. 1607 and 1608)
Terpin Hydrate	Glass, 1.2 M × 3.5 mm I.D. 6% G1/S1A	Nitrogen	120	FID	Biphenyl	
Terpin Hydrate elixir						
Terpin Hydrate and Codeine elixir	Glass, 1.2 M × 3.5 mm I.D. 6% G1/S1A	Nitrogen	230	FID	Biphenyl and N-Phenylcarbazole	USP (24, p. 1608)
Testosterone Cypionate	Glass, 1.2 M × 3 mm I.D. 1% G6/S1AB	Helium	260	FID	Cholesteryl Caprylate	USP (24, pp. 1610 and 1611)
Testosterone Cypionate injection						
Tetracaine and Menthol ointment	Glass, 1.8 M × 2 mm I.D. 10% G16/S1AB	Helium	170	FID	Decanol	USP (24, p. 1615)
Tiletamine and Zolazepam for injection	1.24 M × 4 mm I.D. 3% G2/S1AB (100–120 mesh)	Helium	150–230 10/min	FID	Tetraphenylethylene	USP (24, p. 1660)
Tocopherols excipient	Glass, 2 M × 4 mm I.D. 2–5% liquid G2/S1AB (80–1000 mesh)	Nitrogen	245–265	FID	Hexadecyl hexadecanoate	NF (19, p. 2531)
Tributyl Citrate	30 M × 0.32 mm I.D. G42 (0.5 µm)	Helium	80–225 20/min	FID	None	NF (19, p. 2532)
Trichloromono-fluormethane	Stainless Steel, 1.8 M × 2 mm I.D. 1% G25/S12	Helium	70–170 10/min	FID	None	NF (19, p. 2532)
Compound Undecylenic Acid ointment	Glass, 1.8 M × 2 mm I.D. 3% G1/S1A (100–20 mesh)	Helium	165	FID	Tridecanoic Acid	USP (24, p. 1729)
Valproic Acid	1.8 M × 2.0 mm I.D. 10% G34/S1A (80–100 mesh)	Helium	155	FID	Nonanoic Acid	USP (24, p. 1732)
Valproic Acid capsules	Glass, 1.8 M × 2 mm I.D. 10% G34/S1A (80–100 mesh)	Helium	150	FID	Biphenyl	USP (24, pp. 1733 and 1734)
Valproic Acid Syrup						
Vitamin E	Borosilicate glass, 2 M × 4 mm I.D. 2–5% liquid G2/S1AB 80–10 mesh	Nitrogen	245–265	FID	Hexadecyl Hexadecanoate	USP (24, pp. 1747 and 1749)
Vitamin E preparation						
Vitamin E capsules						
Vitamin E Polyethylene Glycol Succinate	5 M × 0.25 mm I.D. fused Silica capillary with G27 (0.25 µm)	Helium	260–340 20/min	FID	Ethyl Aracitate	NF (19, p. 2535)
Xylitol	30 M × 0.25 mm I.D. capillary with G46 (0.25 µm)	Helium	170°–215° 6/min 215–270 10/min	FID	Erythritol	NF (19, p. 2538)

Table 3 Compendial applications of GC for the chromatographic and radiochemical purity of pharmaceutical raw materials and dosage forms

Material/dosage form	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Alprazolam	Glass, 1.2 M × 3 mm I.D. 3% G6/S1AB	Helium	240	FID	None	USP (24, p. 64)
Amantadine Hydrochloride	Glass, 1.8 M × 2 mm I.D. G44/S1A (100–120 mesh)	None	100–200	FID	None	USP (24, p. 103)
Avobenzone	Fused Silica 25 M × 0.32 mm Capillary Coated with G1	Specified Helium	6/min 200–280 4/min	FID	None	USP (24, p. 181)
Chloroxylenol	Glass, 1.8 M × 4 mm I.D. 3% G16/S1A	Nitrogen	180	FID	None	USP (24, p. 391)
Clofibrate	15 M × 0.53 mm I.D. Capillary Coated with G1 (1.5 µm)	Helium	120–180 5/min	FID	None	USP (24, p. 442)
Dichlorodi-fluoromethane	Stainless Steel, 1.8 M × 2 mm I.D. 1% G25/S12	Helium	70–170 10/min	FID	None	NF (19, pp. 2445 and 2446)
Dichlorotetrafluoroethane	Glass (pre-treated with 10% dimethyldichlorosilane in toluene), 1.8 M × 4 mm I.D. 10% G16/S1AB (60–80 mesh)	Helium	160	TCD	None	USP (24, p. 691)
Ethchlorvynol	Glass, 1.8 M × 4 mm I.D. S11	Nitrogen	115–200 16/min	FID	None	NF (19, p. 2450)
Ethyl Acetate	30 M × 0.53 mm I.D. Fused Silica G43 (3 µm) with a 5 M × 0.53 mm I.D. Silica Guard Column (deactivated with phenylmethyl siloxane)	Helium	40–220 5/min	FID	None	USP (24, p. 780)
Glycerin	Stainless Steel, 3 M × 2 mm I.D. 20% G24/S1AB	Nitrogen	60	FID	1,1,2-Trichloro-1,2,2-trifluoroethane	USP (24, p. 806)
Haloethane					Decanol	USP (24, p. 1038) and NF (19, p. 2477)
Menthol	1.8 M × 2 mm I.D. 10% G16/S1AB	Helium	170	FID		

(Continued)

Table 3 Compendial applications of GC for the chromatographic and radiochemical purity of pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage form	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Meperidine Hydrochloride	Glass, 2 M × 2 mm I.D. 10% G3/S1A	Helium	190	FID	None	USP (24, p. 1039)
Mepivacaine Hydrochloride	Glass, 1.8 M × 4 mm I.D. 3% G19/S1A	Helium	230	FID	None	USP (24, p. 1041)
Nicotine	0.3 M × 0.53 mm I.D. fused Silica column bonded with G1 (1.5 μm)	Helium	50–250 5/min	FID	None	USP (24, p. 1179)
Octyl Methoxycinnamate	25 M × 0.32 mm I.D. Column with G1	Helium	80–300 Linearly	FID	None	USP (24, p. 1213)
Water O 15 Injection (radiochemical purity)	Stainless Steel, 1.8 M × 3 mm I.D. S3 (80–100 mesh)	Helium	150	TC and Radioactivity	None	USP (24, p. 1240)
Padimate O	Stainless Steel, 1.8 M × 3 mm I.D. 10% liquid G9/S1A	Helium	150–250 10/min	FID	None	USP (24, p. 1252)
Perflubron	60 M × 0.25 mm I.D. G2 (1 μm)	Hydrogen	35–185 20/min	FID	None	USP (24, p. 1295)
Squalane	1.8 M × 3 mm I.D. 3% G1/S1A	Nitrogen	130–270 6/min	FID	None	NF (19, p. 2524)
Triazolam	Glass, 1.2 M × 3 mm I.D. 3% G6/S1AB	Helium	240	FID	None	USP (24, p. 1695)
Trichloromono-fluoromethane	Stainless Steel, 1.8 M × 2 mm I.D. 1% G25/S12	Helium	70–170 10 min	FID	None	NF (19, p. 2532)
Triclosan	15 M × 0.53 mm I.D. capillary with G3	Helium	34	FID	None	USP (24, p. 1700)
Valproic Acid	60 M × 0.32 mm I.D. coated with G25	Helium	145	FID	None	USP (24, p. 1732)

Table 4 Compendial applications of GC for the identification of pharmaceutical raw materials and dosage forms

Material/dosage form	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Acetyltributyl Citrate	30 M × 0.32 mm I.D. column bonded with G42 (0.5 μm)	Helium	80–220 20/min	FID	None	NF (19, p. 2409)
Acetyltriethyl Citrate	1.5 M × 4 mm I.D. 3% G1/S1A	Nitrogen	250	FID	Squalane	USP (24, p. 121)
Amitraz						
Amitraz Concentrate for dip	2 M × 3 mm I.D. 10% liquid G2/S1	Helium	205	FID	None	NF (19, p. 2434)
Cetostearyl Alcohol	2 M × 3 mm I.D. 10% liquid G2/S1A	Helium	205	FID	None	NF (19, p. 2435)
Cetyl Alcohol	Glass, 1.83 M × 2 mm I.D. 3% G3/S1AB (80–100 mesh)	Helium	165	FID	Pyrene	USP (24, pp. 435, 436 and 437)
Clioquinol						
Clioquinol cream						
Clioquinol ointment						
Clioquinol and Hydrocortisone cream						
Clioquinol and Hydrocortisone ointment						
Colestipol Hydrochloride	Glass, 1.8 M × 3 mm I.D. 0.25% potassium hydroxide and 5% G16/S1A (80–100 mesh)	Helium	85	FID	None	USP (24, p. 465)
Dicyclomine Hydrochloride capsules	fused Silica 15 M × 0.5 mm coated with G3(1 μm)	Nitrogen	160–240 20/min	FID	Phenacetin	USP (24, pp. 549, 550 and 551)
Dicyclomine Hydrochloride injection						
Dicyclomine Hydrochloride syrup						
Dicyclomine Hydrochloride tablets						
Conjugated Estrogens	15 M × 0.25 mm I.D. fused Silica capillary bonded with G19 (0.25 μm)	Hydrogen	220	FID	3-O-Methylsterone	USP (24, pp. 681, 582, 683 and 684)
Conjugated Estrogen tablets						
Esterified Estrogens						
Esterified Estrogen tablets						
Glyceryl Behenate	1.8 M × 4 mm I.D. 10% liquid G27/S1A	Not specified	225	FID	None	NF (19, p. 2462)

(Continued)

Table 4 Compendial applications of GC for the identification of pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage form	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Guaifenesin and Codeine Phosphate syrup	0.6 M × 2 mm I.D. 3% liquid G3/S1A (100–120 mesh) (For Codeine Phosphate)	Helium	210	FID	Hydrocodone Bitartrate	USP (24, p. 793)
Guaifenesin and Codeine Phosphate syrup	1.2 M × 4 mm I.D. 37% liquid G6/S1A (100–120 mesh) (For Guaifenesin)	Helium	170	FID	Dipropyl Phthalate	USP (24, p. 793)
Hydrocortisone and Acetic Acid Otic solution	Glass, 1.8 M × 2 mm I.D. 20% liquid G35/S1A	Nitrogen	115–190 35/min	FID	Anisole	USP (24, p. 827)
Isosorbide oral solution	Glass, 0.6 M × 3 mm I.D. S9	Nitrogen	230	FID	Triethylene Glycol	USP (24, p. 935)
Perflubron	60 M × 0.25 mm I.D. G2 (1 μm)	Hydrogen	35–185 20/min	FID	None	USP (24, p. 1295)
Triclosan	15 M × 0.53 mm I.D. capillary with G3	Helium	34	FID	None	USP (24, p. 1700)
Valproic Acid	1.8 M × 2.0 mm I.D. 10% G34/S1A (80–100 mesh)	Helium	155	FID	Nonanoic Acid	USP (24, p. 1732)
Valproic Acid capsules	Glass, 1.8 M × 7 × 2 mm I.D. 10% G34/S1A (80–100 mesh)	Helium	150	FID	Biphenyl	USP (24, pp. 1733 and 1734)
Valproic Acid syrup	Glass, 2 M × 4 mm I.D. 2–5% liquid G2/S1AB (80–10 mesh)	Nitrogen	245–265	FID	Hexadecyl Hexadecanoate	USP (24, pp. 1747 and 1749)
Vitamin E Preparation	Glass, 2 M × 4 mm I.D. 2–5% liquid G2/S1AB (80–10 mesh)	Nitrogen	245–265	FID	Hexadecyl Hexadecanoate	USP (24, pp. 1747 and 1749)
Vitamin E capsules						
Vitamin E Polyethylene Glycol Succinate	15 M × 0.25 mm I.D. fused Silica capillary with G27 (0.25 μm)	Helium	260–340 20/min	FID	Ethyl Aracideate	NF (19, p. 25s35)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Amoxicillin	Dimethylaniline	2 M X 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FID	Napthalene	USP (24, p. 129)
Ampicillin	Dimethylaniline	2 M X 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FID	Napthalene	USP (24, p. 136)
Ampicillin Sodium	Dimethylaniline	2 M X 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FIDN	N-Diethylaniline	USP (24, p. 140)
Ampicillin Sodium	Methylene Chloride	glass, 1.8 M X 4 mm I.D. 10% 39/S1A	Nitrogen	65	FID	Dioxane	USP (24, p. 140)
Amyl Nitrite	Total Nitrites	2 M X 3 mm I.D. 25% methyl polysiloxane on suitable calcined diatomate	Helium	80	TCD	None	USP (24, pp. 144 and 145)
Amyl Nitrite inhalant							
Bacampicillin	Dimethylaniline	2 M X 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FID	Napthalene	USP (24, p. 189)
Hydrochloride Benzethonium	Acetone and Alcohol	1.2 M X 4 mm S3 (80–100 mesh)	Helium	120	FID	Methanol	USP (24, p. 204)
Chloride Tincture Brompheniramine	Related substances	Glass, 1.2 M X 4 mm I.D. 3% G3/S1AB	Helium	190	FID	None	USP (24, p. 251)
Maleate Bupivacaine	Residual solvents	2 M X 6 mm I.D. S3	Nitrogen	175	FID	None	USP (24, p. 256)
Hydrochloride Butyl Alcohol	Butyl Ether	Stainless steel, 2 M X 6 mm I.D. 25% liquid G29 3,3' thiodipropionitrile/SIC (30–40 mesh)	Helium	85	TCD	None	NF (19, p. 2422)
Carbomer 910	Benzene	30 M X 0.53 mm I.D. fused Silica analytical column with G34 (3.0 µm) and a 5 M X 0.53 mm I.D. guard column deactivated with Phenylmethyl Siloxane	Helium	40 – 260 50/min	FID	None	NF (19, pp. 2426, 2427 and 2428)

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Carbomer 934							
Carbomer 934P							
Carbomer 940							
Carbomer 941							
Carbomer 1342							
Cefadroxil	Dimethylaniline	2 M × 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FID	Napthalene	USP (24, p. 326)
Cefoxitin Sodium	Acetone and Methanol	Glass, 1.8 M × 6.3 mm I.D. S2, silane treated 60–80 mesh glass beads precolumn	Nitrogen	110	FID	None	USP (24, p. 345)
Cephalexin	Dimethylaniline	2 M × 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FID	Napthalene	USP (24, p. 361)
Cephalexin Hydrochloride							
Maleate	Related substances	Glass, 1.2 M × 4 mm I.D. 3% G3/S1AB	Helium	190	FID	None	USP (24, p. 392)
Maleate							
Ciclopirox Oxamine	Benzyl Alcohol	Glass, 2 M × 4 mm I.D. 3% G3/S1AB(100–120 mesh)	Nitrogen	100	FID	1-Nonyl Alcohol	USP (24, pp. 410 and 411)
Cream							
Ciclopirox Oxamine Topical suspension							
Cilastatin Sodium	Acetone, Methanol and Mesityl Oxide	30 M × 0.53 mm capillary, G16 1 μ film	Helium	50–70 8/min	FID	n-Propyl Alcohol	USP (24, p. 411)
Clavulanate	Methanol and tert-butylamine	30 M × 0.32 mm capillary, G1	Nitrogen	40 – 200 55/min	FID	None	USP (24, p. 426)
Potassium Clofibrate	p-chlorophenol	15 M × 0.53 mm I.D. capillary coated with G1 (1.5 m)	Helium	120 – 180 5/min	FID	None	USP (24, p. 442)
Cloxacillin Sodium	Dimethylaniline	2 M × 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FID	Napthalene	USP (24, p. 456)
Colchicine	Residual Chloroform and Ethyl Acetate	Glass or Stainless Steel, 1.5 M × 4 mm I.D. 20% G14/S1	Nitrogen	75	FID	n-Propanol	USP (24, p. 464)

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Collodion	Alcohol	Glass, 1.8 M × 3.5 mm I.D. S3	Helium	150	TCD	1,2-Dichloroethane	USP (24,P. 469)
Flexible collodion							
Cyclosporine injection	Alcohol	Glass, 2 M × 2 mm I.D. S3	Nitrogen	145–270 32/min	FID	n-Propyl Alcohol	USP (24, pp. 488 and 489)
Cyclosporine oral solution							
Cyclosporine injection Cyclosporine oral solution	Alcohol	Glass, 2 M × 2 mm I.D. S3	Nitrogen	145–270 32/min	FID	n-Propyl Alcohol	USP (24, pp. 488 and 489)
Desflurane	Related compounds and Trichlorofluoromethane, Dichlorofluoromethane, Methylene Chloride, Chloroform, Trichlorotrifluoroethane and Isoflurane	Stainless Steel, 6.1 M × 2.4 mm I.D. 25% G16/S1A (80–100 mesh)	Helium	75	FID	None	USP (24, p. 504)
Dexbrom-pheniramine Maleate	Related substances	Glass, 1.2 M × 4 mm I.D. 3% G3/S1AB	Helium	190	FID	None	USP (24, p. 520)
Maleate	Related substances	Glass, 1.2 M × 4 mm I.D. 3% G3/S1AB	Helium	190	FID	None	USP (24, p. 520)
Dexpanthenol	Pantolactone	1.8 M × 2 mm I.D. 5% G2/S1A	Helium or Nitrogen	170	FID	2,6-Dimethylphenol	USP (24, p. 523)
Preparation							
Dextran 40 Dextran 70	Alcohol and related impurities	1.8 M × 2 mm I.D. S3	Nitrogen	160	FID	None	USP (24, pp. 523 and 526)
Dextran 70							
Dicloxacillin Sodium	Dimethylaniline	2 M × 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FID	Napthalene	USP (24, p. 548)
Diethylene Glycol Monoethyl Ether	Ethylene Oxide	Glass or Quartz, 30 M × 0.32 mm I.D. capillary with G1 (1.0 µm)	Helium	50 – 180 5/min 180–230 30/min	FID	None	NF (19, p. 2447)

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Diethylene Glycol	Ethylene Oxide	Glass or Quartz, 30 M × 0.32 mm I.D. capillary with G1 (1.0 μm)	Helium	50 – 180 5/min 180–230 30/min	FID	None	NF (19, p. 2447)
Monoethyl Ether Dihydroxyaluminum Sodium Carbonate Dimethyl	Isopropyl Alcohol	0.9 M × 3 mm I.D. S3		180	FID	None	USP (24, p. 572)
	Dimethyl Sulfone	1.5 M × 3 mm I.D. 10% liquid phase G25/ S1A	Helium	100–170 10/min	FID	Dibenzyl	USP (24, p. 579 and USP 24-NF 19 First Suppt, p. 2608)
Sulfoxide Doxylamine Succinate	Secondary peaks	2 M × 4 mm I.D. 5%G16, 5% G12/S1A (60–80 mesh)	Helium	212	FID	None	USP (24, p. 612)
Dyphylline elixir Dyphylline and Guaifenesin elixir Conjugated Estrogens	Alcohol	Glass, 0.75 M × 4 mm I.D. 20% G20/S1AB	Nitrogen	85	FID	None	USP (24, pp. 618 and 619)
	Estrone, Equilin and 17α-dihydroequilin (free steroids)	15 M × 0.25 mm I.D. fused Silica capillary bonded with G19(0.25 μm)	Hydrogen	220	FID	3-O-Methylestrone	USP (24, pp. 681 and 683)
Esterified Estrogens Ether	Low-boiling hydrocarbons	Stainless Steel, 3.7 M × 2 mm I.D 30%G22/S1C (30–60 mesh)	Nitrogen	80	FID	None	USP (24, p. 692)
Ethosuximide	2-Ethyl-2-methylsuccinic Anhydride and other impurities	1.8 M × 6.4 mm I.D. 5% G5/ S1A (60–80 mesh)	Helium	140	HFD	None	USP (24, p. 695)
Etodolac	Alcohol and Methanol	25 M × 0.32 mm I.D. fused Silica capillary G36 (5 μm)	Helium	45–280 30/min	FID	Isopropyl Alcohol	USP 244, p. 701)

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Fluocinonide topical solution	Alcohol content	Glass, 1.8 M × 2 mm I.D. S3 (80–120 mesh)	Nitrogen or Helium	130	FID	Isopropyl Alcohol	USP (24, p. 729)
Gadodiamide	Acetone, Ethyl Alcohol, Isopropyl Alcohol	30 M × 0.32 mm I.D. capillary with G43 (1.8 µm)	Helium	40	FID	Methyl Ethyl Ketone	USP (24 - NF 19 First Suppl.P. 2618)
Gentamicin Sulfate	Methanol	1.5 M × 4 mm I.D. S3	Nitrogen	120 – 140	FID	n-Propyl Alcohol	USP (24, p. 765)
Glycerin	Methylene Chloride, Benzene, Trichloroethylene, 1,4-Dioxane, Chloroform, Heuylene Glycol	30 M × 0.53 mm I.D. fused Silica G43 (3 µm) with a 5 M × 0.53 mm I.D. Sillica guard column (deactivated with phenylmethyl siloxane)	Helium	40–220 5/mm	FID	None	USP (24, p. 780 and NF 19, p. 2461)
Glycerol	Free Glycerin	Glass, 2.4 M × 4 mm I.D. 2% liquid G16/S1A (80–100 mesh)	Helium	190 – 200	FID	Tributyrin	NF (19, p. 2463)
Monostearate Glycerol	Monoglycerides	Glass, 2.4 M × 4 mm I.D. 2% liquid G27/S1A (80–100 mesh)	Helium	270 – 280	FID	Hexadecyl Hexadecanoate	NF (19, p. 2463)
Monostearate	Alcohol content	1.8 M × 2 mm I.D. 5% liquid G16/S1A (100–120 mesh)	Helium	50–100 20/min	FID	Acetone	USP (24, p. 793)
Guaifenesin and Codeine Phosphate syrup	2,6-dichlorobenzaldehyde	Glass, 1.8 M × 3 mm I.D. 20% G1/S1A (80–100 mesh)	Nitrogen	190	FID	p-Chlorobenzaldehyde	USP (24, p. 796)
Guanabenz Acetate	2,3,7,8-tetrachlorodibenzo-p-dioxin	Glass, 1 M × 2 mm I.D. G1/S1	Helium	250	Mass Spectrograph	None	USP (24, p. 810)
Hexachlorophene	Acetic Acid	Glass, 1.8 M × 2 mm I.D. 20% liquid G35/S1A	Nitrogen	115–190 35/min	FID	Anisole	USP (24, p. 827)

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Ifosfamide	2-chloroethylamine hydrochloride	1.8 M × 2 mm I.D. 10% liquid G16, 2% KOH/S1A (80–100 mesh)	Nitrogen	140°	FID	None	USP (24, p. 861)
Imipenem	Acetone, Isopropyl Alcohol	1.8 M × 3 mm I.D. 10% G16/S5	Helium	70–170 32/min	FID	n-Propyl Alcohol	USP (24, p. 863)
Indomethacin	Acetone	Glass, 1.8 M × 3 mm I.D. S3	Nitrogen	165	FID	None	USP (24, p. 878)
Sodium Iohexol	Methanol, Isopropyl Alcohol, Methoxyethanol	Glass, 2 M × 2 mm I.D. S6	Nitrogen	125–225 Maximum Rate	FID	None	USP (24, p. 900)
Iopromide	Alcohol (limit)	30 M × 0.25 mm I.D. capillary coated liquid G43 (1.4 µm)	Helium	40–70 5/min 70–220 30/min	FID	None	USP (24, p. 904)
Ioxilan	Residual Methanol	10 M × 0.53 mm I.D. fused Silica capillary/S2	Helium	45–80 5/min	FID	Dehydrated Alcohol	USP (24, p. 911 and 912)
Ioxilan injection Iron dextran injection	Phenol (limit)	1.2 M × 3 mm I.D. 5% G16/S1A	Helium or Nitrogen	145	FID	Benzyl Alcohol	USP (24, p. 916)
Isoflurane	Acetone, 1-chloro-2,2,2-trifluoroethyl-chlorodifluoro-methyl ether 2,2,2-trifluoroethyl-difluoro-methyl ether	Nickel or Stainless Steel, 3.7 M × 2.4 mm I.D. 10% G31 + 15% G18 on S1C (60–80–mesh)	Helium	65–110 4/min	FID	Butyl Acetate	USP 24, p. 920
Azeotropic Isopropyl Alcohol	Air, Diethyl Ether, Diisopropyl Ether, Acetone, Isopropyl Alcohol, 2-Butanol, n-Propyl Alcohol, Water	Stainless Steel, 1.8 M × 6.4 mm I.D. 10% liquid G20/S1A	Helium	55	FID	None	USP 24, p. 927
Isosorbide	Methyl-ethyl Ketone	Glass or Stainless Steel, 0.6 M × 2 mm I.D. 25% G16 On Unsilanized S1A	Nitrogen	70	FID	Methyl-isobutyl Ketone	USP 24, p. 934
Isoxsuprine	Related compounds	Glass, 2 M × 3 mm I.D. 3% liquid G3/S1A	Nitrogen	215	FID	None	USP 24, p. 940

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Hydrochloride Lanolin	Diazinon, Dicrofenthion, Bromophos ethyl, lindane, Dieldrin	System I: 30 M × 0.53 mm I.D. fused Silica bonded with G1 (1.5 μm) and a 6 M × 0.53 mm I.D. uncoated guard column System II: 30 M × 0.53 mm I.D. Fused Silica Capillary Bonded with G3 (1 μm) and a 6 M × 0.53 mm I.D. uncoated guard column	Helium	200°	ECD Flamephoto-metric Detector	Chlorpyrifos	USP 24, p. 953
Modified Lanolin	Free Lanolin Alcohols	50 M × 0.33 mm I.D. fused Silica capillary bonded with G2 (0.5 μm) with a 0.5 M × 0.32 mm I.D. uncoated guard column	Nitrogen	210 – 280 3/min	FID	None	USP 24, p. 955 and NF 19, P. 2471
Levopropoxyphene Napsylate	Acetoxy Analog Napsylate	Glass, 0.6 M × 3 mm I.D. 3% G2/SIAB	Helium	160	FID	n-Tricosane	USP, p. 706
Magnesium Stearate	Stearic Acid, Palmitic Acid	30 M × 0.32 mm I.D. fused Silica with G16 (0.5 μm)	Helium	70 – 240 5/min	FID	None	NF 19, p. 2474
Malathion lotion	Isopropyl Alcohol	Glass, 1.8 M × 2 mm I.D. S2(110–120 mesh)	Nitrogen	130	FID	Ethylacetate	USP 24, p. 1012
Mesalamine	Aniline, 4-Aminophenol, 2-Aminophenol	10 M × 0.53 mm I.D. fused Silica capillary coated with G27 (2.65 μm)	Helium	70 – 150 30/min	FID	None	USP 24, p. 1045 and USP 24-NF 19 first Suppl, p. 2640
Metaproterenol	Isopropyl Alcohol, Methanol	2 M × 2 mm I.D. 0.1% liquid G25/S7 (80–10 mesh)	Helium	40–200 15/min	FID	None	USP 24, p. 1051
Sulfate Mitoxantrone	Alcohol	Glass, 3 M × 2 mm I.D. 20% G1 and 0.1%	Helium	50–140 30/min	FID	n-Propyl Alcohol	USP 24, p. 1121

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Hydrochloride							
Mono- and Diglycerides	Free Glycerin	G39/Silanized SIA Glass, 2.4 M × 4 mm I.D. 2% liquid G16/SIA (80–100 mesh)	Helium	190–200	FID	Tributyrin	NF 19, p. 2483
Mono- and Diglycerides	Monoglycerides	Glass, 2.4 M × 4 mm I.D. 2% liquid G27/SIA (80–100 mesh)	Helium	270–280	FID	Hexadecyl Hexadecanoate	NF 19, p. 2483
Moricizine Hydrochloride	Alcohol	Glass, 1.8 M × 4 mm I.D. S2	Helium	150	FID	None	USP 24, p. 1129
Myristyl Alcohol	Hydroxyl value	Glass, 2 M × 3 mm I.D. 10% G2/SIA	Helium	205	FID	None	NF, p. 1952
Naftifine Hydrochloride Gel	Alcohol	1.5 M × 3.2 mm I.D. S3 (80–100 mesh)	Nitrogen	170	FID	n-Propyl Alcohol	USP 24, p. 1139
Naltrexone Hydrochloride	Total Solvents (Alcohol and Methanol)	Glass, 1.8 M × 4 mm I.D. S3 (80–100 mesh)		150	FID	Isopropyl Alcohol	USP 24, p. 1143
Nonoxynol 9	Free ethylene oxide	Nickel, 6.4 M × 2.1 mm I.D. S9 (60–80 mesh)	Helium	100	FID	None	USP 24, p. 1194 and NF 19, p. 2486
Nonoxynol 9	Dioxane	Glass, 1.8 M × 2 mm I.D. S10	Nitrogen or Helium	140	FID	None	USP 24, p. 1194, NF 19, p. 2486 and USP 24-NF 19 First Suppt, P. 2725
Octoxynol	Free ethylene oxide	Nickel, 6.4 M × 2.1 mm I.D. S9 (60–80 mesh)	Helium	100	FID	None	NF 19, p. 2486 and USP 24-NF 19 First Suppt, P. 2725
Octoxynol 9	Dioxane	Glass, 1.8 M × 2 mm I.D. S10	Nitrogen or Helium	140	FID	None	NF 19, p. 2486 and USP 24-NF 19 First Suppt, P. 2725
Ofloxacin	Methanol, Ethanol	30 M × 0.53 mm I.D. fused Silica column coated with G43 (3 μm)	Helium	35–90 20/min 90–200 40/min	FID	n-Propyl Alcohol	USP 24, p. 1215
Oxycodone Hydrochloride	Alcohol	Glass, 1.8 M × 4 mm I.D. S3 (10–120 mesh)	Helium	150	FID	Isopropyl Alcohol	USP 24, p. 1233
Phenytol Sodium injection	Propylene Glycol and Alcohol	Glass, 1.8 M × 2.0 mm I.D. silanised S3 (50–80 mesh)	Helium	140–190 6/min	FID	Ethylene Glycol and Methanol	USP 24, p. 1326

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Polycarbophil	Ethyl Acetate	10 ft × 2 mm I.D. 1% G25/S12	Helium	160	FID	Methyl Ethyl Ketone	USP 24, p. 1348
Poloxamer	Limit of Free Ethylene Oxide Propylene Oxide, 1,4 dioxane	50 M × 0.32 mm I.D. fused Silica with G27 (5 μm)	Helium	70–240 10/min	FID	None	NF 19, p. 2492 and USP 24-NF 19 First Suppt.P. 2725
Polyethylene Glycol	Limit of Free Ethylene Oxide and 1,4 dioxane	50 M × 0.32 mm I.D. fused Silica with G27 (5 μm)	Helium	70–250 10/min	FID	None	NF 19, p. 2493 and USP 24-NF 19 First Suppt.P. 2726
Polyethylene Glycol	Ethylene Glycol and Diethylene Glycol	Stainless Steel, 1.5 M × 3 mm I.D. 12% G13/SINS	Nitrogen	165	FID	None	NF 19, p. 2493
Polyethylene Glycol Monomethyl Ether	Ethylene Oxide and 1,4 dioxane	50 M × 0.32 mm I.D. fused Silica with G27 (5 xμm)	Helium	70 – 250 10/min	FID	None	NF 19, p. 2495
Polyethylene Glycol Monomethyl Ether	Ethylene Glycol and Diethylene Glycol	1M × 3mm I.D. S2 (60–80 mesh)	Nitrogen	200	FID	None	NF 19, p. 2495
Polyethylene Glycol Monomethyl Ether	2-methoxyethanol	15 M × 0.53 mm I.D. fused Silica capillary column with G16 (1 μm)	Helium	70–250 10/min	FID	None	NF 19, p. 2495
Polyethylene Oxide	Free Ethylene Oxide	10 M × 0.53 mm I.D. Capillary Column with G45 (20 μm)	Helium	70–200 10/min	FID	None	NF 19, p. 2497
Polyoxyl 10 Oleyl Ether	Free Ethylene Oxide	Stainless Steel, 1.8 M × 3 mm I.D. (OD) S3	Helium	160	FID	n-Butyl Chloride	NF 19, p. 2498
Polyoxyl 20 Cetostearyl Ether	Free Ethylene Oxide	Stainless Steel, 1.8 M × 3 mm I.D. (OD) S3	Helium	160	FID	n-Butyl Chloride	NF 19, p. 2499
Procyclidine Hydrochloride	Related compounds	Glass 1 M × 2 mm I.D. 10% PEG 20,000 and 2% KOH/SIA	Helium	180	FID	None	USP 24, p. 1406
Propafenone Hydrochloride	Methanol, Acetone	30 M × 0.53 mm I.D. fused Silica column with G43 (3 μm) and a 5 M × 0.53 mm I.D. Silica guard column deactivated with Phenylmethyl Siloxane	Helium	40–240 Rapidly	FID	None	USP 24, p. 1414

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Propoxyphene Hydrochloride	α-d-2-Acetoxy-4-dimethylamino-1,2-diphenyl-3-methylbutane, Carbinol hydrochloride, α-d-4-Dimethylamino-1,2-diphenyl-3-methyl-2-butanol hydrochloride	Glass or Stainless Steel, 0.6 M × 3 mm I.D. 3% G2/ S1 AB	Helium	160	FID	n-Tricosane	USP 24, p. 1420
Propoxyphene Napsylate	α-d-2-Acetoxy-4-dimethylamino-1,2-diphenyl-3-methylbutane, Carbinol hydrochloride, α-d-4-Dimethylamino-1,2-diphenyl-3-methyl-2-butanol hydrochloride	Glass or Stainless Steel, 0.6 M × 3 mm. I.D. 3% G2/S1 AB	Helium	160	FID	n-Tricosane	USP 24, p. 1424
Saccharin	Toluenesul-fonamides	Glass, 1.8 M × 3.2 mm I.D. 10% liquid G3/S1AB (100–120 mesh)	Helium	210	FID	n-Tricosane	NF 19, p. 2509
Saccharin Calcium	Toluenesul-fonamides	Glass, 1.8 M × 3.2 mm I.D. 10% liquid G3/S1AB (100–120 mesh)	Helium	210	FID	n-Tricosane	USP 24, p. 1497
Saccharin Sodium	Toluenesul-fonamides	Glass, 1.8 M × 3.2 mm I.D. 10% liquid G3/S1AB (100–120 mesh)	Helium	210	FID	n-Tricosane	USP 24, p. 1498 and NF19, p. 2509
Salsalate	Dimethylaniline	30 M × 0.53 mm I.D. capillary coated with G42 (1 μm)	Helium	105	FID	Indene	USP 24, p. 1502
Salsalate	Isopropyl, ethyl and methyl salicylates	30 M × 0.53mm I.D. capillary coated with G42 (1 μm)	Helium	120	FID	None	USP 24, p. 1502
Sucralfate	Pyridine, 2-methylpyridine	10 M × 0.53 mm I.D. capillary coated with G27 (2.65 μm)	Helium	50	FID	3-Methylpyridine	USP 24, p. 1555
Sucralose	Methanol	Glass, 2 M × 4 mm I.D. silanized S6 (80–100 mesh)	Helium	150	FID	n-Propyl Alcohol	NF 19, p. 2527

(Continued)

Table 5 Compendial applications of GC for the presence of volatile matter, intermediates, and related substances in raw materials and dosage forms (*Continued*)

Material/dosage forms	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Sufentanil Citrate	Acetone	Glass, 1.83 M × 4 mm I.D. S2	Nitrogen	175	FID	None	USP 24 - NF 19 First Suppl, p. 2658
Tamoxifen Citrate	Related impurities	Glass, 1 M × 4 mm I.D. 5% liquid G17/S1AB (100–120 mesh)	Helium	260	FID	None	USP 24, p. 1586
Ticarcillin Disodium Ticarcillin injection	Dimethylaniline	2 M × 2 mm I.D. 3% G3/silanized S1A	Nitrogen	120	FID	Napthalene	USP 24, p. 1657 and 1659
Propylene Glycol	Propylene Glycol	15 M × 0.53 mm I.D. fused Silica with liquid G16 (1 µm)	Helium	100	FID	Pentadecane	USP 24, p. 1662
Triclosan	Related compounds	15 M × 0.53 mm I.D. capillary with G3	Helium	34	FID	None	USP 24, p. 1700
Triclosan	2,3,7,8-tetrachlorodibenzo-p-dioxin, 2,3,7,8-tetrachlorodibenzofuran	30 M × 0.53 mm I.D. capillary with G3	Helium	80–180 20/min 180–270 4/min	Mass Spec with electron impact ionization	C Tetrachloro-dibenzo-p-13 labelled 2,3,7,8-dioxin, C13 Labelled 2,3,7,8-Tetrachloro-dibenzofuran	USP 24, p. 1700
Warfarin Sodium	Isopropyl Alcohol	Glass, 1.8 M × 4 mm I.D. S2 (80–100 mesh)	Nitrogen	140	FID	n-Propyl Alcohol	USP 24, p. 1750
Xanthan gum	Isopropyl Alcohol	Stainless Steel, 1.8 M × 3.2 mm I.D. silanized S3 (80–100 mesh)	Helium	165	FID	Tertiary Butyl Alcohol	NF 19, p. 2537
Xylazine	Acetone, Isopropyl Alcohol	1.8 M × 2mm I.D. 0.1% G25/ S7 (80–10 mesh)	Helium	30–100 10/min 100–220 15/min	FID	None	USP 24, p. 1756
Xylitol	Other Polyols	30 M × 0.25 mm I.D. capillary with G 46 (0.25 µm)	Helium	170–215 6/min 215–270 10/min	FID	Erythritol	NF 19, p.2538

Table 6 Compendial applications of GC for the determination of water in pharmaceutical raw materials and dosage forms

Material/dosage form	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Acetone	Stainless Steel, 1.5 <i>M</i> × 6 mm I.D. S4	Helium	180	TCD	None	NF (19, p. 2409)
Echothiophate Iodide for Ophthalmic Solution	Glass, 1.8 <i>M</i> × 2 mm I.D. silanised S3 (80–100 mesh)	Helium	115	TCD	Anhydrous Methanol	USP (24, p. 621)
Gonadorelin Hydrochloride	Glass, 1.8 <i>M</i> × 2 mm I.D. S3 (80–100 mesh)	Helium	100	TCD	Anhydrous Methanol	USP (24, p. 784)

Table 7 Compendial applications of GC for the presence of isomers, isomeric purity, and racemate ratios in pharmaceutical raw materials

Material	Analyte	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Butorphanol Tartrate	Purity and presence of α -isomer	Glass, 1.8 <i>M</i> × 4 mm I.D. 3% G3/S1	Nitrogen	250	FID	None	USP (24,p.271)
Ethchlorvynol	E-Ethchlorvynol	Glass (pretreated with 10% dimethyldichlorosilane in toluene), 1.8 <i>M</i> × 4 mm I.D. 10% G16/S1AB (60–80 mesh)	Helium	160	TCD	None	USP (24, p. 691)
Fludeoxyglucose F 18 injection	Isomeric purity	1.8 <i>M</i> × 3 mm I.D. 4% G2 (SF-30),6% G6 (OV 210)/S1A (80–100 mesh) (Chromosorb W-HP)	Helium	150	FID	None	USP (24, p. 733)
Labetalol Hydrochloride	Racemate ratio	Glass, 1.8 <i>M</i> times; 2 mm I.D. 10% G3/S1AB (10–120 mesh)	Nitrogen	320	FID	None	USP (24, p. 949)
Anhydrous Lactose	Content of alpha and beta anomers	Glass, 0.9 <i>M</i> × 4 mm I.D. 3% liquid G19/S1A	Helium	215	FID	None	NF (19, p. 2470)
Phendimetrazine Tartrate	L-erythro isomer	25 <i>M</i> × 0.25 mm I.D. capillary column G1 (0.4 μ m)	Helium	140	FID	None	USP (24, p. 1301)

Table 8 Compendial applications of GC for the determination of alcohol content in raw materials and dosage forms^a

Dosage form	Reference
Acetaminophen oral solution	USP (24, p. 18)
Oral solution containing at least three of the following—Acetaminophen and salts of Chlorpheniramine, Dextromethorphan, and Phenylpropanolamine	USP (24, p. 24)
Oral solution containing at least three of the following—Acetaminophen and salts of Chlorpheniramine, Dextromethorphan, and Pseudoephedrine	USP (24, p. 31)
Acetaminophen and Codeine Phosphate oral solution	USP (24, p. 35)
Acetaminophen, Dextromethorphan, Hydrobromide, Doxylamine Succinate, and Pseudoephedrine Hydrochloride oral solution	USP (24, p. 37)
Aminobenzoic Acid gel and topical solution	USP (24, p. 110 and, p. 111)
Aromatic elixir	NF (p. 1901)
Belladonna tincture	USP (24, p. 201) ^b
Compound Benzoin tincture	USP (24, p. 211)
Butabarbital Sodium elixir	USP (24, p. 262)
Camphor spirit	USP (24, p. 293)
Aromatic Cascara fluid extract	USP (24, p. 322)
Aromatic elixir	NF (19, p. 2415)
Cinoxate Lotion	USP (24, p. 416)
Desoximetasone gel	USP (24, p. 508) ^c
Dexamethasone elixir	USP (24, p. 512)
Dexamethasone Sodium Phosphate	USP (24, p. 516) ^d
Dexamethasone Sodium Phosphate inhalation aerosol	USP (24, p. 517)
Ergoloid Mesylates oral solution	USP (24, p. 656)
Erythromycin topical solution	USP (24, p. 667)
Etoposide injection	USP (24, p. 704)
Green Soap tincture	USP (24, p. 788)
Isoetharine Mesylate inhalation aerosol	USP (24, p. 919)
Isoproterenol Hydrochloride inhalation	USP (24, p. 928) ^d
Methdilazine Hydrochloride syrup	USP (24, p. 1061)
Methylcellulose oral solution	USP (24, p. 1080)
Nitroglycerin injection	USP (24, p. 1190)
Nortriptyline Hydrochloride oral solution	USP (24, p. 1206)
Opium tincture	USP (24, p. 1221)
Oxycodone Hydrochloride oral solution	USP (24, p. 1234) ^b
Paregoric	USP (24, p. 1263)
Peppermint spirit	USP (24, p. 1295)
Phenobarbital elixir	USP (24, p. 1306)
Potassium Chloride oral solution	USP (24, p. 1358) ^b
Potassium Gluconate elixir	USP (24, p. 1365)
Prednisilone syrup	USP (24, p. 1381)
Prednisone oral solution	USP (24, p. 1388)
Prednisone syrup	USP (24, p. 1389)
Propoxyphene Napsylate oral suspension	USP (24, p. 1425)
Pseudoephedrine Hydrochloride, Carbinoxamine Maleate, and Dextromethorphan Hydrobromide oral solution	USP (24) NF 19 First Suppl, p. 2654)
Resorcinol and Sulfur lotion	USP (24, p. 1478)
Sulfamethoxazole and Trimethoprim oral suspension	USP (24, p. 1573)
Terpin Hydrate elixir Terpin Hydrate and Codeine elixir	USP (24, pp. 1607 and 1608)
Theophylline and Guaifenesin oral solution	USP (24, p. 1634) ^b
Theophylline Sodium Glycinate elixir	USP (24, p. 1635)
Thiamine Hydrochloride elixir	USP (24, p. 1639) ^b
Thiamine Mononitrate elixir	USP (24, p. 1641) ^b

(Continued)

Dosage form	Reference
Thimerosal Topical aerosol	USP (24, p. 1644) ^d
Thimerosal tincture	USP (24, p. 1645)
Thiothixene Hydrochloride oral solution	USP (24, p. 1654)
Triprolidine Hydrochloride syrup	USP (24, p. 1717)
Powdered Valerian extract	USP (24), NF (19 First Suppl, p. 2729)

^aChromatographic conditions: Column—Glass 1.8 m × 4 mm I.D. S3 (100–120 mesh) or with S8
^dCarrier Gas – Nitrogen or Helium; Temperature – 120°C; Detector—FID
^bInternal Standard – Acetonitrile, Acetone
^cIsopropyl Alcohol
^dMethyl Ethyl Ketone

Table 9 Compendial applications of GC for the presence of organic volatile impurities in pharmaceutical raw materials and dosage forms

Material/dosage forms	Chapter ≤467≥	Reference	Material/dosage forms	Chapter ≤467≥	Reference
Acacia	Method I ^b	NF (19, p. 2408)	Arginine	Method I	USP (24, p. 158) and NF (19, p. 2415)
Acetaminophen	Method V ^b	USP (24, p. 17)	Arginine Hydrochloride	Method I	USP (24, p. 158)
Acetazolamide	Method V	USP (24, p. 40)	Aromatic Elixir	Method I	NF (19, p. 2415)
Acetic Acid	Method IV ^b	NF (19, p. 2408)	Ascorbic Acid	Method IVx	USP (24, p. 160) and NF (19, p. 2415)
Acetohexamide	Method V	USP (24, p. 42)	Ascorbyl Palmitate	Method IV	NF (19, p. 2415)
Acetohydroxamic Acid	Method I	USP (24, p. 43)	Aspirin	Method IV	USP (24, p. 161)
Acetylcholine Chloride	Method I	USP (24, p. 44)	Aspartame	Method IV	NF (19, p. 2415)
Acetylcysteine	Method I	USP (24, p. 45) and NF (19 p. 2409)	Atropine	Method V	USP (24, p. 177)
Acyclovir	Method V	USP (24, p. 46)	Atropine Sulfate	Method I	USP (24, p. 178)
Adenine	Method V	USP (24, p. 50)	Activated Attapulgate	Method IV	USP (24, p. 180)
Agar	Method IV	NF (19, p. 2410)	Colloidal Activated Attapulgate		
Alanine	Method I	USP (24, p. 52)	Azathioprine	Method V	USP (24, p. 184)
Albuterol Sulfate	Method I	USP (24, p. 55)	Baclofen	Method IV	USP (24, p. 194)
Allopurinol	Method V	USP (24, p. 62)	Belladonna Extract	Method IV	USP (24, p. 199)
Aluminium Monostearate	Method IV	NF (19, p. 2412)	Bendroflumethiazide	Method V	USP (24, p. 201)
Aluminium Sulfate	Method IV	USP (24, p. 91)	Bentonite Purified	Method IV	NF (19, pp. 2416, 2417 and 2418)
Amantadine	Method I	USP (p. 103)	Bentonite Bentonite Magma		
Hydrochloride			Benzaldehyde	Method V	NF (19, p. 2418)
Amiloride Hydrochloride	Method V	USP (24, p. 107)	Compound Benzaldehyde Elixir	Method I	NF (19, p. 2419)
Aminoglutehimide	Method V	USP (24, p. 112)	Benzonatate	Method I	USP (24, p. 211)
Aminophylline	Method I	USP (24, p. 115)	Benzotropine Mesylate	Method I	USP (24, p. 213)
Amitriptyline	Method I	USP (24, p. 122)	Benzyl Alcohol	Method V	NF (19, p. 2420)
Hydrochloride			Benzyl Benzoate	Method V	USP (24, p. 215) and NF (19, p. 2421)
Amobarbital Sodium	Method I	USP (24, p. 126)			
Amodiaquine Amodiaquine Hydrochloride	Method V	USP (24, pp. 126 and 127)			
Amphetamine Sulfate	Method I	USP (24, p. 134)			
Amyl Nitrate	Method V	USP (24, p. 145)			
Amylene Hydrate	Method I	NF (19, p. 2414)			
Anethole	Method V	NF (19, p. 2414)			

(Continued)

(Continued)

Table 9 Compendial applications of GC for the presence of organic volatile impurities in pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage forms	Chapter ≤467≥	Reference
Beta Carotene	Method IV	USP (24, p. 216)
Betamethasone	Method V	USP (24, p. 217)
Bethanechol Chloride	Method I	USP (24, p. 229)
Biotin	Method V	USP (24, p. 234)
Biperiden Biperiden Hydrochloride	Method IV	USP (24, p. 234)
Bromocriptine Mesylate	Method V	USP (24, p. 248)
Bromodiphenhydramine Hydrochloride	Method I	USP (24, p. 251)
Brompheniramine Maleate	Method I	USP (24, p. 251)
Busulfan	Method V	USP (24, p. 260)
Butabarbital	Method V	USP (24, p. 261)
Butabarbital Sodium	Method I	USP (24, p. 261)
Butalbital	Method V	USP (24, p. 263)
Butylated Hydroxyanisole	Method V	NF (19, p. 2422)
Butylated Hydroxytoluene	Method IV	NF (19, p. 2423)
Butyl Paraben	Method IV	NF (19, p. 2423)
Caffeine	Method I	USP (24, p. 272)
Calcifediol	Method V	USP (24, p. 274)
Calcium Acetate	Method I	USP (24, p. 275)
Calcium Carbonate	Method IV	USP (24, p. 277) and NF (19, p. 2423)
Calcium Chloride	Method I	USP (24, p. 282) and NF (19, p. 2423)
Calcium Citrate	Method IV	USP (24, p. 282)
Calcium Gluceptate	Method I	USP (24, p. 283)
Calcium Gluconate	Method I	USP (24, p. 284)
Calcium Hydroxide	Method IV	USP (24, p. 286)
Calcium Lactate	Method I	USP (24, p. 287)
Calcium Lactobionate	Method I	USP (24, p. 288)
Calcium Levulinate	Method I	USP (24, p. 288)
Calcium Pantothenate	Method I	USP (24, pp. 289 and 290)
Racemic Calcium Pantothenate		
Dibasic Calcium Phosphate	Method IV	USP (24, p. 290) and NF (19, p. 2423)
Calcium Polycarbophil	Method IV	USP (24, p. 291)
Calcium Saccharate	Method IV	USP (24, p. 291) and NF (19, p. 2424)
Calcium Stearate	Method IV	NF (19, p. 2425)
Captopril	Method I	USP (24, p. 296)
Carbamazepine	Method V	USP (24, p. 299)
Carbamide Peroxide	Method I	USP (24, p. 301)

(Continued)

Material/dosage forms	Chapter ≤467≥	Reference
Carbidopa	Method IV	USP (24, p. 304)
Carbinoxamine Maleate	Method I	USP (24, p. 305)
Carbomer 934P	Method IV	NF (19, p. 2427)
Carboxymethylcellulose	Method IV	NF (19, p. 2428)
Calcium Carboxymethylcellulose Sodium	Method IV	USP (24, p. 313) and NF (19, p. 2429)
Carboxymethylcellulose Sodium 12	Method IV	NF (19, p. 2429)
Carisoprodol	Method V	USP (24, p. 314)
Cellacefat	Method IV	NF (19, p. 2431)
Microcrystalline Cellulose	Method IV	NF (19, p. 2432)
Microcrystalline Cellulose and Carboxymethylcellulose Sodium	Method IV	NF (19, p. 2433)
Powdered Cellulose	Method IV	NF (19, p. 2433)
Cellulose Acetate	Method IV	NF (19, p. 2434)
Cetylpyridium Chloride	Method V	USP (24, p. 370)
Chloral Hydrate	Method I	USP (24, p. 371)
Chlorambucil	Method V	USP (24, p. 372)
Chlordiazepoxide	Method IV	USP (24, pp. 381 and 384)
Chlordiazepoxide Hydrochloride		
Chlorobutanol	Method IV	NF (19, p. 2437)
Chloroquine	Method V	USP (24, p. 388)
Chloroquine Phosphate	Method I	USP (24, p. 389)
Chlorothiazide	Method V	USP (24, p. 389)
Chlorpheniramine Maleate	Method IV	USP (24, p. 392)
Chlorpromazine	Method V	USP (24, p. 395)
Chlorpromazine Hydrochloride	Method I	USP (24, p. 396)
Chlorpropamide	Method IV	USP (24, p. 398)
Chlorzoxazone	Method V	USP (24, p. 403)
Cholecalciferol	Method IV	USP (24, p. 404)
Cholesterol	Method IV	NF (19, p. 2438)
Cholestyramine Resin	Method IV	USP (24, p. 406)
Cimetidine	Method IV	USP (24, p. 412)
Cinoxacin	Method IV	USP (24, p. 415)
Citric Acid	Method IV	USP (24, p. 423) and NF (19, p. 2438)
Clidinium Bromide	Method I	USP (24, p. 429)
Clofibrate	Method V	USP (24, p. 442)
Clomiphene Citrate	Method V	USP (24, p. 444)
Clonazepam	Method V	USP (24, p. 445)
Clorazepate Dipotassium	Method I	USP (24, p. 448)
Cocoa Butter	Method IV	NF (19, p. 2438)

(Continued)

Table 9 Compendial applications of GC for the presence of organic volatile impurities in pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage forms	Chapter ≤467≥	Reference
Colchicine	Method I	USP (24, p. 464)
Corn Oil	Method IV	NF (19, p. 2439)
Cortisone Acetate	Method IV	USP (24, p. 473)
Cottonseed oil	Method IV	NF (19, p. 2439)
Cromolyn Sodium	Method I	USP (24, p. 475)
Croscarmellose Sodium	Method IV	NF (19, p. 2441)
Cupric Chloride	Method V	USP (24, p. 477)
Cupric Sulfate	Method IV	USP (24, p. 479)
Cyclizine Hydrochloride	Method IV	USP (24, p. 481)
Cyclobenzaprine	Method I	USP (24, p. 481)
Hydrochloride		
Cyproheptadine	Method V	USP (24, p. 489)
Hydrochloride		
Cysteine Hydrochloride	Method IV	USP (24, p. 490)
Danazol	Method V	USP (24, p. 495)
Dapsone	Method V	USP (24, p. 496)
Dehydrocholic Acid	Method V	USP (24, p. 500)
Desipramine	Method I	USP (24, p. 505)
Dexamethasone	Method IV	USP (24, p. 512)
Dexamethasone Acetate	Method V	USP (24, p. 515)
Dexamethasone Sodium Phosphate	Method IV	USP (24, p. 516)
Dexbrompheniramine Maleate	Method I	USP (24, p. 520)
Dexchlorpheniramine Maleate	Method I	USP (24, p. 521)
Dexpanthenol	Method IV	USP (24, p. 523)
Dextrates	Method I	NF (19, p. 2444)
Dextrin	Method I	NF (19, p. 2444)
Dextroamphetamine Sulfate	Method I	USP (24, p. 528)
Dextrose Excipient	Method I	NF (19, p. 2445)
Diazepam	Method V	USP (24, p. 538)
Dichlorphenamide	Method V	USP (24, p. 545)
Dicyclomine Hydrochloride	Method I	USP (24, p. 549)
Diethanolamine	Method IV	NF (19, p. 2446)
Diethylpropion Hydrochloride	Method I	USP (24, p. 552)
Diethylstilbestrol	Method V	USP (24, p. 554)
Diethylstilbestrol Diphosphate	Method V	USP (24, p. 555)
Diffunisal	Method V	USP (24, p. 558)
Digitalis Powdered Digitalis	Method IV	USP (24, p. 560)
Dihydrotachysterol	Method IV	USP (24, p. 568)
Dihydroxyaluminum Aminoacetate	Method IV	USP (24, p. 570)

(Continued)

Material/dosage forms	Chapter ≤467≥	Reference
Dihydroxyaluminum Sodium Carbonate	Method IV	USP (24, p. 572)
Diltiazem Hydrochloride	Method IV	USP (24, p. 573)
Dimenhydrinate	Method V	USP (24, p. 576)
Diphenhydramine Citrate	Method V	USP (24, p. 582)
Diphenhydramine Hydrochloride	Method I	USP (24, p. 583)
Dipyridamole	Method IV	USP (24, p. 590)
Disopyramide Phosphate	Method I	USP (24, p. 593)
Disulfiram	Method V	USP (24, p. 594)
Doxepin Hydrochloride	Method I	USP (24, p. 604)
Doxylamine Succinate	Method I	USP (24, p. 612)
Dyphylline	Method I	USP (24, p. 618)
Enalapril Maleate	Method IV	USP (24, p. 638)
Ephedrine Ephedrine Hydrochloride Ephedrine Sulfate	Method I	USP (24, pp. 642 and 643)
Ergocalciferol	Method V	USP (24, p. 651)
Conjugated Estrogens Esterified Estrogens	Method V	USP (24, pp. 681 and 683)
Estropipate	Method V	USP (24, p. 686)
Ethacrynic Acid	Method V	USP (24, p. 688)
Ethambutol Hydrochloride	Method I	USP (24, p. 689)
Ethionamide	Method V	USP (24, p. 694)
Ethosuximide	Method I	USP (24, p. 695)
Ethyl Acetate	Method I 30 M × 0.53 mm I.D. fused Silica column with G16 (1 µm)	NF (19, p. 2450)
Ethyl Vanillin	Method IV	NF (19, p. 2450)
Ethylcellulose	Method IV	NF (19, p. 2451)
Ethylcellulose aqueous dispersion	Method V	NF (19, p. 2451)
Ethylparaben	Method IV	NF (19, p. 2452)
Ethylenediamine	Method V	USP (24, p. 698)
Ethynodiol Diacetate	Method IV	USP (24, p. 698)
Etidronate Disodium	Method I	USP (24, p. 700)
Eucatropine Hydrochloride	Method I	USP (24, p. 706)
Famotidine	Method V	USP (24, p. 707)
Fenoprofen Calcium	Method V	USP (24, p. 708)
Ferrous Fumarate	Method IV	USP (24, p. 7111)
Ferrous Gluconate	Method I	USP (24, p. 712)
Ferrous Sulfate Dried Ferrous Sulfate	Method IV	USP (24, p. 715)

(Continued)

Table 9 Compendial applications of GC for the presence of organic volatile impurities in pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage forms	Chapter ≤467≥	Reference
Flucytosine	Method V	USP (24, p. 719)
Fluoxetine Hydrochloride	Method IV	USP (24, p. 738)
Fluoxymesterone	Method V	USP (24, p. 740)
Fluphenazine Hydrochloride	Method I	USP (24, p. 743)
Flurazepam Hydrochloride	Method I	USP (24, p. 746)
Flurbiprofen Sodium	Method I	USP (24, p. 749)
Folic Acid	Method IV	USP (24, p. 752)
Fumaric Acid	Method V	NF (19, p. 2453)
Furosemide	Method V	USP (24, p. 756)
Galageenan	Method IV	NF (19, p. 2454)
Gemfibrozil	Method V	USP (24, p. 763)
Liquid Glucose	Method I	NF (19, p. 2461)
Glyceryl Behenate	Method IV	NF (19, p. 2462)
Glyceryl Monostearate	Method IV	NF (19, p. 2463)
Glycine	Method I	USP (24, p. 782) and NF (19, p. 2464)
Griseofulvin	Method V	USP (24, p. 788)
Guaifenesin	Method IV	USP (24, p. 791)
Guar gum	Method IV	NF (19, p. 2464)
Guanabenz Acetate	Method V	USP (24, p. 796)
Guanadrel Sulfate	Method I	USP (24, p. 797)
Guanethidine Monosulfate	Method I	USP (24, p. 799)
Haloperidol	Method V	USP (24, p. 804)
Homatropine Methylbromide	Method I	USP (24, p. 815)
Hydralazine Hydrochloride	Method I	USP (24, p. 818)
Hydrochlorothiazide	Method V	USP (24, p. 820)
Hydrocodone Bitartrate	Method I	USP (24, p. 821)
Hydrocortisone	Method IV	USP (24, p. 824)
Hydrocortisone Sodium Phosphate	Method I	USP (24, p. 831)
Hydroflumethiazide	Method V	USP (24, p. 835)
Hydromorphone Hydrochloride	Method I	USP (24, p. 836)
Hydroquinone	Method I	USP (24, p. 838)
Hydroxychloroquine Sulfate	Method I	USP (24, p. 841)
Hydroxyethyl Cellulose	Method IV	NF (19, p. 2465)
Hydroxypropyl Cellulose	Method IV	NF (19, p. 2465)
Hydroxypropyl Methylcellulose	Method IV	USP (24, p. 843) and NF (19, p. 2466)
Hydroxyurea	Method I	USP (24, p. 844)
Hydroxyzine Hydrochloride	Method I	USP (24, p. 846)

(Continued)

Material/dosage forms	Chapter ≤467≥	Reference
Hydroxyzine Pamoate	Method V	USP (24, p. 848)
Hypromellose Phthalate	Method IV	NF (19, p. 2467)
Hyoscyamine Sulfate	Method I	USP (24, p. 851)
Ibuprofen	Method V	USP (24, p. 854)
Imidurea	Method I	NF (19, p. 2467)
Imipramine Hydrochloride	Method I	USP (24, p. 865)
Indapamide	Method IV	USP (24, p. 867)
Indomethacin	Method IV	USP (24, p. 874)
Iopromide	Method IV	USP (24, p. 904)
Isoleucine	Method I	USP (24, p. 923)
Isoniazid	Method I	USP (24, p. 924)
Isopropamide Iodide	Method I	USP (24, p. 926)
Isopropyl Myristate	Method IV	NF (19, p. 2468)
Isopropyl Palmitate	Method IV	NF (19, p. 2468)
Isoproterenol Hydrochloride	Method I	USP (24, p. 928)
Isoproterenol Sulfate	Method I	USP (24, p. 932)
Isotretinon	Method V	USP (24, p. 938)
Isoxsuprine Hydrochloride	Method V	USP (24, p. 940)
Ketoconazole	Method IV	USP (24, p. 945)
Ketoprofen	Method IV	USP (24, p. 946)
Ketorolac Tromethamine	Method V	USP (24, p. 947)
Labetalol Hydrochloride	Method I	USP (24, p. 949)
Lactitol	Method IV	NF (19, p. 2469)
Lecithin	Method IV	NF (19, p. 2471)
Levmetamfetamine	Method I	USP (24, p. 959)
Levodopa	Method IV	USP (24, p. 963)
Lime	Method IV	USP (24, p. 974)
Lithium Carbonate	Method I	USP (24, p. 981)
Lithium Citrate	Method V	USP (24, p. 983)
Lithium Hydroxide	Method V	USP (24, p. 984)
Loxapine Succinate	Method V	USP (24, p. 993)
Lysine Acetate	Method I	USP (24, p. 994)
Lysine Hydrochloride	Method I	USP (24, p. 994)
Mafenide Acetate	Method V	USP (24, p. 995)
Magaldrate	Method V	USP (24, p. 996)
Magnesium Chloride	Method I	USP (24, p. 1002)
Magnesium Citrate	Method IV	USP (24, p. 1003)
Magnesium Gluconate	Method I	USP (24, p. 1004)
Magnesium Salicylate	Method I	USP (24, p. 1008)
Magnesium Silicate	Method IV	NF (19, p. 2473)
Magnesium Stearate	Method IV	NF (19, p. 2473)
Magnesium Sulfate	Method I	USP (24, p. 1009)
Malic Acid	Method I	NF (19, p. 2475)
Mandelic Acid	Method IV	USP (24, p. 1012)
Manganese Chloride	Method I	USP (24, p. 1012)
Manganese Gluconate	Method I	USP (24, p. 1013)

(Continued)

Table 9 Compendial applications of GC for the presence of organic volatile impurities in pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage forms	Chapter ≤467≥	Reference
Manganese Sulfate	Method I	USP (24, p. 1014)
Maprotiline Hydrochloride	Method V	USP (24, p. 1016)
Mecamylamine Hydrochloride	Method I	USP (24, p. 1023)
Meclizine Hydrochloride	Method IV	USP (24, p. 1025)
Meclofenamate Sodium	Method I	USP (24, p. 1027)
Megestrol Acetate	Method IV	USP (24, p. 1030)
Melphalan	Method IV	USP (24, p. 1032)
Menthol	Method V	USP (24, p. 1038)
Meperidine Hydrochloride	Method I	USP (24, p. 1039)
Mephénytoin	Method V	USP (24 - NF 19) First Suppl, (p. 2639)
Meprobamate	Method V	USP (24, p. 1043)
Mercaptopurine	Method V	USP (24, p. 1044)
Mesoridazine Besylate	Method I	USP (24, p. 1048)
Metaproterenol Sulfate	Method IV	USP (24, p. 1051)
Methacrylic Acid copolymer	Method V	NF (19, p. 2477)
Methadone Hydrochloride	Method I	USP (24, p. 1056)
Methamphetamine Hydrochloride	Method I	USP (24, p. 1058)
Methazolamide	Method I	USP (24, p. 1060)
Methdilazine Hydrochloride	Method I	USP (24, p. 1061)
Methenamine Methenamine Hippurate Methenamine Mandelate	Method I	USP (24, pp. 1062, 1063 and 1064)
Methimazole	Method I	USP (24, p. 1066)
Methionine	Method I	USP (24, p. 1067)
Methocarbamol	Method V	USP (24, p. 1067)
Methotrexate	Method V	USP (24, p. 1070)
Methoxsalen	Method V	USP (24, p. 1073)
Methsuximide	Method V	USP (24, p. 1075)
Methyl Salicylate	Method IV	NF (19, p. 2479)
Methylcellulose	Method V	USP (24, p.1079) and NF (19, p. 2480)
Methyldopa	Method V	USP (24, p.1080)
Methylene Blue	Method I	USP (24, p.1085)
Methylene Chloride in coated tablets	Method V	USP (24, p.1878)
Methylparaben	Method IV	NF (19, p. 2480)
Methylparaben Sodium	Method I	NF (19, p. 2481)
Methylphenidate Hydrochloride	Method I	USP (24, p. 1088)
Methyltestosterone	Method V	USP (24, p. 1094)

(Continued)

Material/dosage forms	Chapter ≤467≥	Reference
Metoclopramide Hydrochloride	Method I	USP (24, p. 1097)
Metoprolol Fumarate	Method I	USP (24, p. 1100)
Metoprolol Tartrate	Method I	USP (24, p. 1101)
Metyrosine	Method IV	USP (24, p. 1108)
Mexiletine Hydrochloride	Method I	USP (24, p. 1109)
Minoxidil	Method IV	USP (24, p. 1118)
Mititane	Method V	USP (24, p. 1120)
Monoethanolamine	Method I	NF (19, p. 2484)
Monosodium Glutamate	Method I	NF (19, p. 2484)
Monothioglycerol	Method IV	NF (19, p. 2484)
Moricizine Hydrochloride	Method V	USP (24, p. 1129)
Morphine Sulfate	Method I	USP (24, p. 1131)
Nadolol	Method IV	USP (24, p. 1134)
Nandrolone Decanoate	Method V	USP (24, p. 1144)
Nandrolone Phenpropionate	Method V	USP (24, p. 1145)
Naproxen	Method V	USP (24, p. 1147)
Naproxen Sodium	Method I	USP (24, p. 1149)
Niacin	Method IV	USP (24, p. 1176)
Niacinamide	Method I	USP (24, p. 1179)
Nifedipine	Method V	USP (24, p. 1183)
Norethindrone	Method IV	USP (24, p. 1196)
Norethindrone Acetate	Method IV	USP (24, p. 1199)
Norethynodrel	Method V	USP (24, p. 1202)
Nortriptyline Hydrochloride	Method I	USP (24, p. 1206)
Octoxynol	Method IV	NF (19, p. 2486)
Octyldodecanol	Method V	NF (19, p. 2486)
Oleic Acid	Method IV	NF (19, p. 2486)
Oleovitamin A and D	Method V	USP (24, p. 1217)
Olive Oil	Method IV	NF (19, p. 2487)
Omeprazole	Method IV	USP (24, p. 1217)
Oxandrolone	Method V	USP (24, p. 1225)
Oxazepam	Method V	USP (24, p. 1226)
Oxprenolol Hydrochloride	Method I	USP (24, p. 1228)
Oxtriphylline	Method I	USP (24, p. 1229)
Oxybutynin Chloride	Method I	USP (24, p. 1232)
Oxymetholone	Method V	USP (24, p. 1242)
Oxyquinoline Sulfate	Method I	NF (19, p. 2487)
Panthenol	Method I	USP (24, p. 1258)
Papaverine Hydrochloride	Method IV	USP (24, p. 1260)
Peanut Oil	Method IV	NF (19, p. 2488)
Pectin	Method IV	USP (24, p. 1265) and NF (19, p. 2489)
Pentobarbital	Method V	USP (24, p. 1293)
Pentobarbital Sodium	Method I	USP (24, p. 1294)
Peppermint	Method IV	NF (19, p. 2489)

(Continued)

Table 9 Compendial applications of GC for the presence of organic volatile impurities in pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage forms	Chapter ≤467≥	Reference
Peppermint water	Method I	NF (19, p. 2490)
Perphenazine	Method V	USP (24, p. 1296)
Phendimetrazine Tartrate	Method I	USP (24, p. 1301)
Phenelzine Sulfate	Method I	USP (24, p. 1303)
Pheniramine Maleate	Method IV	USP (24, p. 1304)
Phenmetrazine Hydrochloride	Method I	USP (24, p. 1304)
Phenobarbital	Method V	USP (24, p. 1305)
Phenobarbital Sodium	Method I	USP (24, p. 1307)
Phenol Liquefied Phenol	Method I	USP (24, p. 1308)
Phenoxybenzamine Hydrochloride	Method V	USP (24, p. 1308)
Phensuximide	Method V	USP (24, p. 1309)
Phentermine Hydrochloride	Method I	USP (24, p. 1310)
Phenylalanine	Method I	USP (24, p. 1313)
Phenylbutazone	Method V	USP (24, p. 1313)
Phenylethyl Alcohol	Method IV	USP (24, p. 1318) and NF (19, p. 2490)
Phenylmercuric Acetate	Method IV	NF (19, p. 2490)
Phenylmercuric Nitrate	Method IV	NF (19, p. 2490)
Phenylpropanolamine Bitartrate	Method I	USP (24, pp. 1318 and 1319)
Phenylpropanolamine Hydrochloride		
Phenytoin Phenytoin Sodium	Method V	USP (24, pp. 1322 and 1324)
Pimozide	Method V	USP (24, p. 1335)
Pindolol	Method V	USP (24, p. 1336)
Piroxicam	Method V	USP (24, p. 1342)
Polacrillin Potassium	Method IV	NF (19, p. 2492)
Polycarbophil	Method IV	USP (24, p. 1348)
Poloxamer	Method V	NF (19, p. 2492)
Polyethylene Glycol	Method IV	NF (19, p. 2493)
Polyethylene Oxide	Method I	NF (19, p. 2497)
Polyoxyl 10 Oleyl Ether	Method V	NF (19, p. 2498)
Polyoxyl 20 Cetostearyl Ether	Method I	NF (19, p. 2499)
Polyoxyl 35 Castor oil	Method I	NF (19, p. 2500)
Polyoxyl 40 Hydrogenated Castor oil	Method I	NF (19, p. 2501)
Polyoxyl 40 Stearate	Method I	NF (19, p. 2501)
Polysorbate 20 Polysorbate 40 Polysorbate 80	Method IV	NF (19, pp. 2501 and 2502)
Polyvinyl Acetate Phthalate	Method IV	NF (19, p. 2502)

(Continued)

Material/dosage forms	Chapter ≤467≥	Reference
Polyvinyl Alcohol	Method I	USP (24, p. 1352) and NF (19, p. 2503)
Sulfurated Potash	Method IV	USP (24, p. 1353)
Potassium Benzoate	Method I	NF (19, p. 2503)
Potassium Bicarbonate	Method IV	USP (24, p. 1354)
Potassium Carbonate	Method I	USP (24, p. 1357) and NF (19, p. 2503)
Potassium Chloride	Method I	USP (24, p. 1357) and NF (19, p. 2503)
Potassium Citrate	Method I	USP (24, p. 1362) and NF (19, p. 2503)
Potassium Gluconate	Method I	USP (24, p. 1364)
Potassium Iodide	Method I	USP (24, p. 1368)
Potassium Metabisulfite	Method V	NF (19, p. 2503)
Potassium Perchlorate	Method I	USP (24, p. 1370)
Monobasic Potassium Phosphate	Method I	NF (19, p. 2504)
Potassium Sorbate	Method I	NF (19, p. 2504)
Prazosin Hydrochloride	Method IV	USP (24, p. 1379)
Primaquine Phosphate	Method I	USP (24, p. 1391)
Primidone	Method V	USP (24, p. 1392)
Probenecid	Method V	USP (24, p. 1393)
Probucol	Method V	USP (24, p. 1395)
Procainamide Hydrochloride	Method I	USP (24, p. 1397)
Procabazine Hydrochloride	Method I	USP (24, p. 1403)
Prochlorperazine Edisylate	Method I	USP (24, p. 1405)
Prochlorperazine Maleate	Method V	USP (24, p. 1406)
Procyclidine Hydrochloride	Method V	USP (24, p. 1406)
Proline	Method I	USP (24, p. 1409)
Promazine Hydrochloride	Method I	USP (24, p. 1410)
Propantheline Bromide	Method I	USP (24, p. 1415)
Propionic Acid	Method I	NF (19, p. 2505)
Propoxyphene Hydrochloride	Method I	USP (24, p. 1420)
Propoxyphene Napsylate	Method V	USP (24, p. 1424)
Propranolol Hydrochloride	Method I	USP (24, p. 1428)
Propyl Gallate	Method V	NF (19, p. 2506)
Propylene Carbonate	Method I	NF (19, p. 2506)

(Continued)

Table 9 Compendial applications of GC for the presence of organic volatile impurities in pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage forms	Chapter ≤467≥	Reference	Material/dosage forms	Chapter ≤467≥	Reference
Propylene Glycol	Method IV	USP (24, p. 1434) and NF (19, p. 2506)	Silicon Dioxide Colloidal	Method IV	NF (19, p. 2514)
Propylene Glycol Monostearate	Method IV	NF (19, p. 2507)	Silicon Dioxide		
Propylparaben	Method IV	NF (19, p. 2508)	Simethicone	Method IV	USP (24, p. 1518) and NF (19, p. 2515)
Propylparaben sodium	Method I	NF (19, p. 2508)	Sodium Acetate	Method IV	USP (24, p. 1524)
Propylthiouracil	Method V	USP (24, p. 1436)	Sodium Ascorbate	Method I	USP (24, p. 1525) and NF (19, p. 2515)
Protriptyline Hydrochloride	Method I	USP (24, p. 1439)	Sodium Benzoate	Method IV	NF (19, p. 2516)
Pseudoephedrine Hydrochloride	Method V	USP (24, p. 1439)	Sodium Bicarbonate	Method IV	USP (24, p. 1525) and NF (19, p. 2516)
Pyrazinamide	Method I	USP (24, p. 1444)	Sodium Borate	Method I	NF (19, p. 2516)
Pyridostigmine Bromide	Method I	USP (24, p. 1446)	Sodium Butyrate	Method I	USP (24 - NF 19) First Suppl (p. 2657)
Pyridoxine Hydrochloride	Method I	USP (24, p. 1447)	Sodium Carbonate	Method I	NF (19, p. 2516)
Pyrilamine Maleate	Method I	USP (24, p. 1449)	Sodium Dehydroacetate	Method I	NF (19, p. 2516)
Pyrimethamine	Method V	USP (24, p. 1450)	Sodium Fluoride	Method I	USP (24, p. 1532)
Pyroxylin	Method V	USP (24, p. 1451)	Sodium Formaldehyde Sulfoxylate	Method I	NF (19, p. 2517)
Quinidine Gluconate	Method I	USP (24, pp. 1453 and 1456)	Sodium Iodide	Method I	USP (24, p. 1535)
Quinidine Sulfate			Sodium Lauryl Sulfate	Method IV	NF (19, p. 2517)
Quinine Sulfate	Method IV	USP (24, p. 1458)	Sodium Monofluorophosphate	Method I	USP (24, p. 1536)
Racinephrine	Method V	USP (24, p. 1461)	Monobasic Sodium Phosphate	Method I	USP (24, p. 1540) and NF (19, p. 2518)
Racinephrine Hydrochloride	Method I	USP (24, p. 1462)	Sodium Propionate	Method I	NF (19, p. 2518)
Ranitidine Hydrochloride	Method I	USP (24, p. 1462)	Sodium Salicylate	Method I	USP (24, p. 1542)
Rauwolfia Serpentina	Method IV	USP (24, p. 1466)	Sodium Stearate	Method I	NF (19, p. 2519)
Resorcinol	Method IV	USP (24, p. 1478)	Sodium Stearyl Fumurate	Method IV	NF (19, p. 2519)
Resorcinol Monoacetate	Method I	USP (24, p. 1479)	Sorbic Acid	Method IV	NF (19, p. 2520)
Riboflavin	Method IV	USP (24, p. 1480)	Sorbitan Monolaurate	Method V	NF (19, p. 2520)
Riboflavin 5'-Phosphate	Method IV	USP (24, p. 1482)	Sorbitan Monooleate	Method IV	NF (19, p. 2521)
Sodium			Sorbitan Monopalmitate	Method IV	NF (19, p. 2521)
Rimexolone	Method V	USP (24, p. 1487)	Sorbitan Monostearate	Method IV	NF (19, p. 2522)
Ritodrine Hydrochloride	Method I	USP (24, p. 1493)	Sorbitol	Method IV	NF (19, p. 2522)
Strong Rose water	Method I	NF (19, p. 2508)	Spironolactone	Method V	USP (24, p. 1546)
Saccharin	Method V	NF (19, p. 2509)	Stanozolol	Method V	USP (24, p. 1549)
Saccharin Calcium	Method I	USP (24, p. 1497) and NF (19, p. 2509)	Starch Pregelatinized starch	Method IV	NF (19, pp. 2524 and 2525)
Saccharin Sodium	Method IV	USP (24, p. 1498)	Stearic Acid Purified	Method V	NF (19, p. 2525)
Salicylamide	Method V	USP (24, p. 1499)	Stearic Acid		
Salsalate	Method V	USP (24, p. 1502)	Storax	Method IV	USP (24, p. 1551)
Scopolamine Hydrobromide	Method I	USP (24, p. 1507)	Sucrose	Method IV	NF (19, p. 2527)
Secobarbital	Method V	USP (24, p. 1509)	Compressible Sugar	Method IV	NF (19, pp. 2528 and 2529)
Secobarbital Sodium	Method I	USP (24, p. 1510)	Confectioner's Sugar		
Selenious Acid	Method I	USP (24, p. 1514)	Sugar spheres		
Serine	Method I	USP (24, p. 1517)			
Sesame oil	Method IV	NF (19, p. 2512)			

(Continued)

(Continued)

Table 9 Compendial applications of GC for the presence of organic volatile impurities in pharmaceutical raw materials and dosage forms (*Continued*)

Material/dosage forms	Chapter ≤467≥	Reference	Material/dosage forms	Chapter ≤467≥	Reference
Sulfamethoxazole	Method IV	USP (24, p. 1571)	Triflupromazine	Method V	USP (24, p. 1704)
Sulfapyridine	Method V	USP (24, p. 1575)	Triflupromazine Hydrochloride	Method I	USP (24, p. 1705)
Sulfasalazine	Method V	USP (24, p. 1576)	Trihexyphenidyl Hydrochloride	Method IV	USP (24, p. 1707)
Sulfinpyrazone	Method V	USP (24, p. 1578)	Trioxsalen	Method V	USP (24, p. 1715)
Sulfisoxazole Acetyl	Method IV	USP (24, p. 1580)	Tripelethamine Hydrochloride	Method I	USP (24, p. 1716)
Sulindac	Method V	USP (24, p. 1581)	Tripolidine Hydrochloride	Method V	USP (24, p. 1716)
Syrup	Method I	NF (19, p. 2530)	Trolamine	Method I	NF (19, p. 2533)
Tamoxifen Citrate	Method V	USP (24, p. 1586)	Tromethamine	Method V	USP (24, p. 1721) and NF (19, p. 2533)
Tannic Acid	Method I	USP (24, p. 1588)	Tryptophan	Method IV	USP (24, p. 1724)
Tartaric Acid	Method I	NF (19, p. 2530)	Tyloxapol	Method I	USP (24, p. 1727) and NF (19, p. 2533)
Terbutaline Sulfate	Method I	USP (24, p. 1604)	Tyrosine	Method IV	USP (24, p. 1728)
Testolactone	Method V	USP (24, p. 1608)	Valine	Method I	USP (24, p. 1732)
Testosterone	Method V	USP (24, p. 1610)	Valproic Acid	Method V	USP (24, p. 1732)
Testosterone Cypionate	Method V	USP (24, p. 1610)	Vanillin	Method IV	NF (19, p. 2534)
Testosterone Enanthate	Method V	USP (24, p. 1611)	Hydrogenated vegetable oil	Method IV	NF (19, p. 2534)
Testosterone Propionate	Method IV	USP (24, p. 1612)	Verapamil Hydrochloride	Method V	USP (24, p. 1739)
Theophylline	Method V	USP (24, p. 1628)	Vitamin E	Method IV	USP (24, p. 1747)
Theophylline Sodium Glycinate	Method I	USP (24, p. 1635)	Vitamin E Polyethylene Glycol Succinate	Method I	NF (19, p. 2535)
Thiamine Hydrochloride	Method IV	USP (24, p. 1639)	Warfarin Sodium	Method I	USP (24, p. 1750)
Thiamine Mononitrate	Method IV	USP (24, p. 1641)	Powdered Valerian extract	Method I	USP (24 - NF 19) First Suppl, (p. 2728)
Thiethylperazine Maleate	Method V	USP (24, p. 1642)	Carnauba wax	Method IV	NF (19, p. 2536)
Thioguanine	Method V	USP (24, p. 1646)	Microcrystalline wax	Method IV	NF (19, p. 2536)
Thioridazine	Method V	USP (24, p. 1648)	Xanthan gum	Method IV	NF (19, p. 2537)
Thioridazine Hydrochloride	Method IV	USP (24, p. 1649)	Xylitol	Method I	NF (19, p. 2538)
Thiothixene	Method V	USP (24, p. 1651)	Xylose	Method I	USP (24, p. 1760) and NF (19, p. 2539)
Thiothixene Hydrochloride	Method I	USP (24, p. 1653)	Zein	Method IV	NF (19, p. 2539)
Threonine	Method V	USP (24, p. 1654)	Zidovudine	Method V	USP (24, p. 1763)
Thymol	Method IV	NF (19, p. 2530)	Zinc Acetate	Method I	USP (24, p. 1766)
Tmolol Maleate	Method I	USP (24, p. 1663)	Zinc Chloride	Method I	USP (24, p. 1766)
Titanium Dioxide	Method IV	USP (24, p. 1666)	Zinc Gluconate	Method I	USP (24, p. 1767)
Tocainide Hydrochloride	Method I	USP (24, p. 1672)	Zinc Stearate	Method IV	USP (24, p. 1769)
Tocopherols excipient	Method IV	NF (19, p. 2531)			
Tolazamide	Method V	USP (24, p. 1674)			
Tolbutamide	Method IV	USP (24, p. 1676)			
Tolmetin Sodium	Method I	USP (24, p. 1677)			
Tragacanth	Method IV	NF (19, p. 2531)			
Trenbolone Acetate	Method IV	USP (24, p. 1683)			
Triamterine	Method IV	USP (24, p. 1692)			
Trientine Hydrochloride	Method I	USP (24, p. 1701)			
Trifluoperazine Hydrochloride	Method I	USP (24, p. 1703)			

(Continued)

Table 10 Chromatographic conditions for different methods

	Method I	Method IV	Method V	Method VI
Analytes	Benzene, Chloroform, 1,4-Dioxane, Methylene Chloride, Trichloroethylene	Benzene, Chloroform, 1,4-Dioxane, Methylene Chloride, Trichloroethylene	Benzene, Chloroform, 1,4-Dioxane, Methylene Chloride, Trichloroethylene	Benzene, Chloroform, 1,4-Dioxane, Methylene Chloride, Tri-chloroethylene
Column	30 M × 0.53 mm I.D. fused Silica analytical column with G27(5 µm) and a 5 M × 0.53 mm I.D. guard column deactivated with Phenyl-methyl Siloxane	30 M × 0.53 mm I.D. fused Silica analytical column with G43 (3.0 µm) and a 5 M × 0.53 mm I.D. guard column deactivated with Phenyl-methyl Siloxane	30 M × 0.53 mm I.D. fused Silica analytical column with G43 (3.0 µm) and a 5 M × 0.53 mm I.D. guard column deactivated with Phenyl-methyl Siloxane	A: 2 M × 3 mm I.D. S3: temperature: 190°C B: 2.1 M × 3 mm I.D. S2: temperature: 160°C C: 30 M × 0.53 mm I.D. G16: temperature: 40°C D: 2 M × 3 mm I.D. G39: temperature: 65°C E: 2 M × 3 mm I.D. G16: temperature: 70°C F: 2.5 M × 2 mm I.D. S4: temperature: 120°C–200°C(2°C/min) H: 2.5 M × 2 mm I.D. G14: temperature: 45°C–120°C(8°C/min) I: 30 M × 0.53 mm I.D. G27: temperature: 35°C–175°C (8°C/min) 175°C - 260°C (35°C/min) J: 30 M × 0.33 mm I.D. G16: temperature: 50°C–165°C(6°C/min) As appropriate for the column dimensions and temperature
Carrier gas	Helium or Nitrogen	Helium	Helium	
Temperature (°C)	35–175 (8/min), 175–260 (35/min)	40–240 (Rapidly)	40–240 (Rapidly)	
Detector	FID	FID	FID	FID
Internal Standard	None	None	None	None
procedure	Inject 1 µl	Inject using a heated gas-tight syringe, 1mL of headspace	Inject 1 µl	Inject 1 µl

Table 11 Miscellaneous compendial applications of GC for pharmaceutical raw materials and dosage forms

Material/dosage form	Column	Carrier gas	Temp (°C)	Detector	Internal standard	Reference
Cocoa butter (fatty acid composition)	15 M × 0.25 mm fused Silica capillary with G 19 (0.25 μm)	Helium	180–240 10/min	FID	None	NF (19, p. 2438)
Corn oil (fatty acid composition)	Glass, 1.8 M × 4 mm I.D. 10% liquid G 4/S1A	Nitrogen	175	FID	None	NF (19, p. 2439)
Dexchlorpheniramine Maleate tablets (dissolution rate)	1.8 M × 2 mm I.D. 1.2% G16, 0.5% KOH/S1AB	Helium	205	FID	Dexbrom-pheniramine Maleate	USP (24, p. 522)
Fatty acid composition	30 M × 0.53 mm I.D. Fused Silica with G16 (1 μm)	Helium	70–260 5/min	FID	None	USP (24, p. 1871)
Mecamylamine Hydrochloride tablets (dissolution rate)	30 M × 0.53 mm I.D. capillary coated with G 27 (1–5 μm)	Helium	150	FID	Biphenyl	USP (24, p. 1023)
Organophosphorous insecticides (See Table 5 USP (24 Page 1890))	30 M × 0.53 mm I.D. fused Silica with G1 (0.25 μm)	Hydrogen, Helium or Nitrogen	80–150 30 /min 150– 280 4/min	Alkali Fid or FPD	None	USP (24, p. 1889)
Organochlorine and Pyrethroid insecticides (See Table 6 USP 24 Page 1890)	30 M × 0.32 mm I.D. fused Silica with G1 (0.25 μm)	Hydrogen, Helium or Nitrogen	80–150 30/min 150– 280 4/min	ECD	None	USP (24, p. 1890)
Saw Palmetto (fatty acid content) Powdered Saw Palmetto (fatty acid content)	30 M × 0.25 mm I.D. fused Silica capillary with G16(0.25 μm)	Helium	120–220 50/min	FID	Nonadecane	NF 19, pp. 2510 and 2512)
Safflower oil (fatty acid composition)	Glass, 1.5 M × 4 mm I.D. 10% liquid G4/S1A	Nitrogen	175	FID	None	USP (24, p. 1499)
Soybean Oil (fatty acid composition)	Glass, 1.5 M × 4 mm I.D. 10% liquid G4/S1A	Nitrogen	175	FID	None	USP (24, p. 1544 and NF 19, p. 2524)

USP 24 NF-19 2000 (21)**Phases**

G1	Dimethylpolysiloxane oil	G14	Polyethylene glycol (average molecular weight of 950–1050)
G2	Dimethylpolysiloxane gum	G15	Polyethylene glycol (average molecular weight of 3000–3700)
G3	50% Phenyl-50% methylpolysiloxane	G16	Polyethylene glycol compound (average molecular weight about 15,000). A high-molecular-weight compound of polyethylene glycol with a diepoxide linker. Available commercially as polyethylene glycol compound 20M, or as Carbowax 20M, from suppliers of chromatographic reagents.
G4	Diethylene glycol succinate polyester		
G5	3-Cyanopropylpolysiloxane		
G6	Trifluoropropylmethylpolysiloxane		
G7	50% 3-Cyanopropyl-50% phenylmethylsilicone		
G8	90% 3-Cyanopropyl-10% phenylmethylsilicone		
G9	Methylvinylpolysiloxane	G17	75% Phenyl-25% methylpolysiloxane
G10	Polyamide formed by reacting a C ₃₆ dicarboxylic acid with 1,3-di-4-piperdylpropane and piperidine in the respective mole ratios of 1.00:0.90:0.20	G18	Polyalkylene glycol
		G19	25% Phenyl–25% cyanopropyl-50% methylsilicone
G11	Bis(2-ethylhexyl) sebacate polyester	G20	Polyethylene glycol (average molecular weight of 380–420)
G12	Phenyldiethanolamine succinate polyester		
G13	Sorbitol	G21	Neopentyl glycol succinate

- G22 Bis(2-ethylhexyl) phthalate
 G23 Polyethylene glycol adipate
 G24 Diisodecyl phthalate
 G25 Polyethylene glycol compound TPA. A high-molecular-weight compound of a polyethylene glycol and a diepoxide that is esterified with terephthalic acid. Available commercially as Carbowax 20M-TPA from suppliers of chromatographic reagents.
 G26 25% 2-Cyanoethyl-75% methylpolysiloxane
 G27 5% Phenyl-95% methylpolysiloxane
 G28 25% Phenyl-75% methylpolysiloxane
 G29 3-3'-Thiodipropionitrile
 G30 Tetraethylene glycol dimethyl ether
 G31 Nonylphenoxypoly(ethyleneoxy)ethanol (average ethyleneoxy chain length is 30); Nonoxynol 30
 G32 20% Phenylmethyl-80% dimethylpolysiloxane
 G33 20% Carborane-80% methylsilicone
 G34 Diethylene glycol succinate polyester stabilized with phosphoric acid
 G35 A high molecular weight compound of a polyethylene glycol and a diepoxide that is esterified with nitroterephthalic acid
 G36 1% Vinyl-5% phenylmethylpolysiloxane
 G37 Polyimide
 G38 Phase G1 containing a small percentage of tailing inhibitor (Commercially available as SP2100/0.1% Carbowax 1500 from Supelco)
 G39 Polyethylene glycol (average molecular weight about 1500)
 G40 Ethylene glycol adipate
 G41 Phenylmethyldimethylsilicone (10% phenyl-substituted)
 G42 35% phenyl-65% dimethylpolysiloxane (percentages refer to molar substitution)
 G43 6% cyanopropylphenyl-94% dimethylpolysiloxane (percentages refer to molar substitution)
 G44 2% low-molecular-weight petrolatum hydrocarbon grease and 1% solution of potassium hydroxide
 G45 Divinylbenzene-ethylene glycol-dimethylacrylate
 G46 14% cyanopropylphenol-86% methylpolysiloxane

Supports USP 24 NF-19 2000 (21) (NOTE:
Unless otherwise specified, mesh sizes of 80–100 or, alternatively, 100–120 are intended.)

- S1A Siliceous earth for gas chromatography has been flux calcined by mixing diatomite with Na_2CO_3

flux and calcining above 900° . The siliceous earth is acid-washed, then water-washed until neutral, but not base-washed. The siliceous earth may be silanized by treating with an agent such as dimethyldichlorosilane (unless otherwise specified in the individual monograph, silanized support is intended) to mask surface silanol groups.

- S1AB The siliceous earth as described above is both acid- and base-washed (unless otherwise specified in the individual monograph, silanized support is intended).
 S1C A support prepared from crushed firebrick and calcined or burned with a clay binder above 900° with subsequent acid-wash. It may be silanized.
 S1NS The siliceous earth is untreated.
 S2 Styrene-divinylbenzene copolymer having a nominal surface area of less than 50 m^2 per g and an average pore diameter of $0.3\text{--}0.4 \mu\text{m}$.
 S3 Copolymer of ethylvinylbenzene and divinylbenzene, having a nominal surface area of $500\text{--}600 \text{ m}^2/\text{g}$ and an average pore diameter of $0.0075 \mu\text{m}$.
 S4 styrene-divinylbenzene copolymer with aromatic -O and -N groups, having a nominal surface area of $400\text{--}600 \text{ m}^2/\text{g}$ and an average pore diameter of $0.0076 \mu\text{m}$.
 S5 40- to 60 mesh, high molecular weight tetrafluorethylene polymer.
 S6 Styrene-divinylbenzene copolymer, having a nominal surface area of $250\text{--}350 \text{ m}^2/\text{g}$ and an average pore diameter of $0.0091 \mu\text{m}$.
 S7 Graphitized carbon having a nominal surface area of $12 \text{ m}^2/\text{g}$.
 S8 Copolymer of 4-vinyl-pyridine and styrene-divinylbenzene.
 S9 A porous polymer based on 2,6-diphenyl-p-phenylene oxide.
 S10 A highly polar cross-linked copolymer of acrylonitrile and divinylbenzene.
 S11 Graphitized carbon having a nominal surface area of $9 \text{ m}^2/\text{g}$ modified with small amounts of petrolatum and polyethylene glycol compound (commercially available as SP1500 on Carbopack B from Supelco).
 S12 Graphitized carbon having a nominal surface area of $100 \text{ m}^2/\text{g}$.

Biological Fluids

While it is clearly apparent that packed columns are mainly used for the various compendial tests, the use of capillary columns for the determination of therapeutic agents in biological fluids has become increasingly

popular. The high resolution capability of capillary columns often overcomes the problems of interference from the biological sample matrix and, coupled with specificity and increased sensitivity has made possible the quantitative analysis of many, formerly undeterminable drugs in biological fluids. In particular, drugs with very poor ultraviolet (UV) molar absorptivities, as well as drugs which by virtue of their specific physicochemical properties result in extremely low recovery values when extracted from biological fluids, have been successfully determined by GC using capillary columns.

The literature is full of such examples and continues to expand at an ever-increasing rate. In addition to the several texts that contain examples of such determinations (3, 81, 82), the reader is referred to several periodicals in which accounts and details of GC assays for drugs (and/or metabolites) in biological fluids are regularly published. These include: *Chromatographia*, *International Journal of Pharmaceutics*, *Journal of Chromatographic Sciences*, *Journal of Chromatography—Biomedical Applications*, *Journal of Pharmaceutical and Biomedical Analysis*, *Journal of Pharmaceutical Sciences*, *Pharmaceutical Research*, *Therapeutic Drug Monitoring*, and LC–GC

FUTURE TRENDS

Although the introduction of HPLC has often been perceived as ultimately replacing GC for use in pharmaceutical analysis, perusal of the current literature and new official compendia clearly indicate that GC is “here to stay.” The notion of an imminent demise of GC appears unrealistic in the light of innovations and applications that continue to expand. Other innovative quantitative techniques with potential for use in pharmaceutical analysis are certainly looming on the horizon. In particular is the technique of high voltage capillary zone electrophoresis. These newer methods, coupled with the now well-established TLC and HPLC methods, will undoubtedly gain more importance and widespread use in the future. However, all these techniques are unlikely to oust GC, since each have their strengths and weaknesses and together compliment the array of techniques and methods for use in pharmaceutical analysis.

Advances in the manufacture of flexible fused silica WCOT columns will almost certainly extend the applications of GC by virtue of their high-resolution capability, while the advent of sophisticated, computerized detectors forecast the improvement in sensitivity and specificity. The use of chiral stationary phases for the resolution of enantiomers is becoming an increasingly

important topic in pharmaceutical analyses. The range and availability of various liquid phases for chiral analysis by GC is bound to make this technique extremely valuable for the assay of pharmaceutical raw materials as well as for use in biological fluids.

Multidimensional and multihyphenated techniques may become increasingly useful, particularly for the analysis of drugs in biological fluids where LC–GC interfacing has a great deal of promise with respect to sample cleanup and preparation time (84,85). GC–MS applications continue to grow in number from the qualitative structural identification point of view, for quantitative analysis that uses SIM (86), and for other quantitative applications of GC (in particular, the increasing use of triple-quad MS/MS spectrometers).

Thus, it is apparent that GC has firmly established itself as a valuable technique for the qualitative and in particular, quantitative determination of drugs. Its application in monitoring impurities, volatile matter, intermediates, and related substances in pharmaceutical raw materials and dosage forms makes it currently the method of choice in this respect. Meanwhile, its increasing use for the quantitative determination of some “hard-to-measure” drugs and metabolites in biological fluids suggest that it is likely to remain an important tool in this arena in the future.

APPENDIX

British Pharmacopoeia (BP) and European Pharmacopoeia (EP) Applications of GC for the Assay; Chromatographic Purity; Identification; Presence of Volatile Matter, Intermediates and Related Substances; Organic Volatile Impurities; Determination of Water; Presence of Isomers and Racemate Ratios; Determination of Alcohol and Miscellaneous Uses of GC in Pharmaceutical Raw Materials and Dosage Forms.

REFERENCES

1. Tswett, M.; Ber. Dtsch. Bot. Ges. **1906**, 24, 316.
2. Martin, A.J.P.; Synge, R.L.M. A New Form of Chromatogram Employing Two Liquid Phases. 1. A Theory of Chromatography. 2. Application to the Micro-Determination of the Higher Monoamino-Acids in Proteins. *Biochem. J.* **1941**, 35, 1358.
3. James, A.T.; Martin, A.J.P. Gas–Liquid Partition Chromatography: The Separation and Micro-Estimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid. *Biochem. J.* **1952**, 50, 679.
4. Prior, F. Doctoral Thesis, University of Innsbruck, 1947.

5. Cremer, E.; Prior, F.Z.; *Electrochem.* **1951**, 55, 66.
6. Ettre, L.S. Early Evolution of Gas Chromatography and Gas Chromatographic Instrumentation. *LC-GC* **1990**, 8, 716.
7. Dimick, K.P. The Emergence of a Commercial Gas Chromatograph. *LC-GC* **1990**, 8, 782.
8. McWilliam, I.G.; Dewar, R.D. Flame Ionization Detector for Gas Chromatography. *Nature* **1958**, 181, 760.
9. Harley, J.; Nel, W.; Pretorius, V. Flame Ionization Detector for Gas Chromatography. *Nature* **1958**, 181, 177.
10. Golay, M.J.E. *Gas Chromatography*; Desty, D.H., Ed.; Butterworths: London, 1958; 36.
11. Perry, J.A. *Introduction to Analytical Gas Chromatography*; Chapter 13 Marcel Dekker, Inc.: New York, 1981.
12. Kovats, E.Z.; *Analyt. Chem.* **1961**, 181, 351.
13. Ardrey, R.E.; Moffat, A.C. Gas-liquid Chromatographic Retention Indices of 1318 Substances of Toxicological Interest on SE-30 or OV-1 Stationary Phase. *J. Chromatogr.* **1981**, 220; 195.
14. In *Clarke's Isolation and Identification of Drugs*; The Pharmaceutical Press: London, 1986; 193–1112.
15. Leathard, D.A.; Schurlock, B.C. *Identification Techniques in Gas Chromatography*; Wiley: New York, 1970.
16. Leathard, D.A. *Advances in Chromatography*; Giddings, J.C., Grushka, E., Keller, R.A., Cazes, J., Eds.; Marcel Dekker, Inc.: New York, 1975; 13, 265.
17. Finkle, B.; Cherry, E.J.; Taylor, D.M. A GLC Based System for the Detection of Poisons, Drugs and Human Metabolites Encountered in Forensic Toxicology. *J. Chromatogr. Sci.* **1971**, 9, 393.
18. *The United States Pharmacopeia*; United States Pharmacopeial Convention, Inc.: Rockville, MD, 1960.
19. *The British Pharmacopoeia*. The Pharmaceutical Press: London, 1968; 1268.
20. Ahuja, S. *Impurities Evaluation of Pharmaceuticals*; Marcel Dekker, Inc.: New York, 1998.
21. *The United States Pharmacopeia/National Formulary (USP 24-NF 19)*; United States Pharmacopeial Convention Inc.: Rockville, MD, 2000.
22. *The British Pharmacopoeia (1999) incorporating the 3rd edition of the European Pharmacopoeia (1997)*; The Stationery Office: London, 1999.
23. Pollack, G.M. Chromatographic Analysis. In *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1989; 2, 419.
24. Jönsson, J.A. Chromatographic Theory and Basic Principles. *Chromatographic Science Series*; Jönsson, J.A., Ed.; Marcel Dekker, Inc.: New York, 1987; 38, 27.
25. Van Deemter, J.J.; Zuiderweg, F.J.; Klinkenberg, A. Longitudinal Diffusion and Resistance to Mass Transfer as Causes of Non-Ideality in Chromatography. *Chem. Eng. Sci.* **1956**, 5, 271.
26. Souter, R.W. *Modern Methods of Pharmaceutical Analysis*. Schirmer, R.E., Ed.; CRC Press Inc.: Boca Raton Florida, 1982; III, 5.
27. Giddings, J.C. The Random Downstream Migration of Molecules in Chromatography. *J. Chem. Ed.* **1958**, 35, 588.
28. Giddings, J.C. *Dynamics of Chromatography, Principles and Theory*; Marcel Dekker, Inc.: New York, 1965.
29. Kline, B.J.; Soine, W.H. *Pharmaceutical Analysis, Modern Methods*; Munson, J.W., Ed.; Chapter I Marcel Dekker, Inc.: New York, 1981.
30. Perretta, A.T.; *Am. Lab.* **1976**, 35, 6.
31. Prescott, B.O.; Wise, H.L. Clean Gases for Gas Chromatography. *J. Gas. Chromatogr.* **1966**, 4, 80.
32. *A User's Guide to Chromatography*; Regis Chemical Co.: Morton Grov, IL, 1976; 23.
33. Grant, D.W.; Clarke, A. A Study of the Effects of Instrument Variables on Accuracy and Precision in Gas Chromatography. II. Errors Associated with the Injection of Liquid Samples by Microlitre Syringes. *J. Chromatogr.* **1974**, 97, 115.
34. Cramers, C.A.; McNair, H.M. *Chromatography. Part A. Fundamentals and Techniques*; Heftmann, E., Ed.; Journal of Chromatography Library; Elsevier Scientific Publishing Company: Amsterdam, 1983; 22A, A203.
35. Jennings, W. *Gas Chromatography with Glass Capillary Columns*; Academic Press: New York, 1980; 53.
36. Freeman, R.P. *High Resolution Gas Chromatography*; Hewlett Packard: California, 1979; 47.
37. Harris, W.E. Some Aspects of Injection of Large Samples in Gas Chromatography. *J. Chromatogr. Sci.* **1973**, 11, 184.
38. Grob, K.; Grob, K., Jr. Isothermal Analysis on Capillary Columns without Stream Splitting: The Role of the Solvent. *J. Chromatogr.* **1974**, 94, 53.
39. Schomburg, G.; Diehlmann, R.; Husmann, H.; Weeke, F. New Developments with Glass Capillary Column Production and Sampling Techniques. *Chromatographia* **1977**, 10, 580.
40. Grob, K.; Grob, K., Jr. On-Column Injection onto Glass Capillary Columns. *J. Chromatogr.* **1978**, 151, 311.
41. Grob, K.; HRC & CC. **1978**, 1, 263.
42. Galli, M.; Trestianu, S.; Grob, K., Jr. HRC & CC. **1979**, 2, 366.
43. David, D.J. *Gas Chromatographic Detectors*; Wiley-Interscience: New York, 1974.
44. Adlard, E.R. A Review of Detectors for Gas Chromatography. Part 1: Universal Detectors. Part 2: Selective Detectors. *Crit. Rev. Anal. Chem.* **1975**, 5, 1.
45. Grob, R.L., Ed.; *Modern Practice of Gas Chromatography*; John Wiley & Sons: New York, 1985.
46. Clement, R.E.; Onuska, F.I.; Eiceman, G.A.; Hill, H.H., Jr. *Gas Chromatography*. *Anal. Chem.* **1990**, 62, 414R.
47. Both, D. *Chromatographic Analysis of Pharmaceuticals*; Adamovics, J.A., Ed.; Marcel Dekker, Inc.: New York, 1990; 49, 119.
48. Elvidge, D.A.; Proctor, K.A. The Use of Gas Chromatography for Determining Water in Pharmaceutical Preparations. *The Analyst* **1959**, 84, 463.
49. Zlatkis, A.; Poole, C. Derivatization Techniques for the Electron-Capture Detector. *Anal. Chem.* **1980**, 52, 1002a.
50. Paterson, P. Recent Advances in Thermionic Ionization Detection for Gas Chromatography. *J. Chromatogr. Sci.* **1986**, 24, 41.
51. Driscoll, J. Review of Photoionization Detection in Gas Chromatography. *J. Chromatogr. Sci.* **1985**, 23, 488.
52. Farwell, S.; Barinaga, C. Sulfur-Selective Detection with the FPD: Current Enigmas, Practical Usage and Future Directions. *J. Chromatogr. Sci.* **1986**, 24, 483.
53. Anderson, R.J.; Hall, R.C. *Chromatography*; International Scientific Communications Inc., 1981; 1, 64.

54. Nyarady, S.A.; Barkley, R.M.; Sievers, R.E. Redox Chemiluminescence Detector: Application to Gas Chromatography. *Anal. Chem.* **1985**, *57*, 2074.
55. Hutte, R.; Sievers, R.; et al. Gas Chromatography Detectors Based on Chemiluminescence. *J. Chromatogr. Sci.* **1986**, *24*, 499.
56. Andrawes, F.; Ramsey, R. The Helium Ionization Detector. *J. Chromatogr. Sci.* **1986**, *24*, 513.
57. Grayson, M. The Mass Spectrometer as a Detector for Gas Chromatography. *J. Chromatogr. Sci.* **1986**, *24*, 529.
58. Biemann, K. The Mass Spectrometer as a Detector in Chromatography. *J. Chromatogr. Libr.* **1985**, *32*, 43.
59. Evershed, R.P. Analysis of Mixtures by Mass Spectrometry. Part 1. Development and New Applications of Gas Chromatography/Mass Spectrometry. *Mass. Spectrom.* **1987**, *9*, 196.
60. McMaster, M.; McMaster, L. *GC/MS: A Practical Users' Guide*; Wiley-VCH: New York, 1998.
61. Smith, S.; Adams, G. Chromatographic Performance and Capillary Gas Chromatography: Fourier Transform Infrared Spectroscopy. *J. Chromatogr.* **1983**, *279*, 1358.
62. Demirgian, J.C. Gas Chromatography: Fourier Transform Infrared Spectroscopy: Mass Spectrometry: A Powerful Tool for Component Identification in Complex Organic Mixtures. *Trends Anal. Chem.* **1987**, *6*, 58.
63. Cone, E.J. *Pharmaceutical Analysis, Modern Methods*; Munson, J.W., Ed.; Marcel Dekker, Inc.: New York, 1981; 143.
64. Coldwell, R.D.; Porteous, C.E.; Trafford, D.J.H.; Makin, H.L.J. Gas Chromatography-Mass Spectrometry and the Measurement of Vitamin D Metabolites in Human Serum or Plasma. *Steroids* **1987**, *49*, 155.
65. Supina, W.R. *The Packed Column in Gas Chromatography*; Supelco, Inc., 1974; 106.
66. *Chiraldex Capillary GC Columns: A Guide to Using Cyclodextrin Bonded Phases for Gas Chromatography*; Advanced Separation Technologies, Inc., 1990.
67. Baiulescu, G.E.; Illie, V.A. *Stationary Phases in Gas Chromatography*; Chapter 4 Pergamon Press: Oxford, 1975.
68. Dal Norgare, S.; Juvet, R.S. *Gas-Liquid Chromatography*; Wiley-Interscience: New York, 1962; 63.
69. Martin, A.J.; Bennett, C.E.; Martinez, F.W., Jr. *Gas Chromatography*; Noebels, H.J., Wall, R.F., Brenner, N. Eds.; Academic Press: New York, 1961; 363.
70. Griffiths, J.H.; James, D.H.; Phillips, C.S.G. Section 8. Adsorption and Partition Methods. *Gas Chromatography. The Analyst* **1952**, *77*, 897.
71. Blau, K.; King, G.S. *Handbook of Derivatives for Chromatography*; Heyden and Sons, 1977.
72. Knapp, D.R. *Handbook of Analytical Derivatization Reactions*; Wiley-Interscience: New York, 1979.
73. Ahuja, S. Derivatization in Gas Chromatography. *J. Pharm. Sci.* **1976**, *65*, 163.
74. Nicholson, J.D. Derivative Formation in Quantitative Gas-Chromatographic Analysis of Pharmaceuticals. Part II, A Review. *The Analyst* **1978**, *103*, 193.
75. Bailey, L.C. *Remington's Pharmaceutical Sciences*; Mack Publishing Company, 1985; 17, 603.
76. Barnes, R.L. *Chromatographic Analysis of Pharmaceuticals*; Adamovics, J.A., Ed.; Chapter 6 Marcel Dekker, Inc.: New York, 1990; 49.
77. McClure, G.L. Improved Determination of Organic Volatile Impurities in Pharmaceutical Materials by <USP-467> Using Automated Static Headspace GC/MS. *PDA J. Pharm. Sci. Tech.* **1999**, *53*, 129.
78. Paul, L. *Personal Communication, Associate Director, General Policies, Nomenclature and Labeling Division*; United States Pharmacopoeia: Rockville, MD, 2000.
79. Hinshaw, J.V. Headspace Sampling. *LC-GC* **1990**, *8*, 362.
80. Jack, D.B. *Drug Analysis by Gas Chromatography*; Academic Press: New York, 1984.
81. Tsuji, K. *GLC and HPLC Determination of Therapeutic Agents*; Tsuji, K., Morozowich, W., Tsuji, K., Eds.; Marcel Dekker, Inc.: New York, 1979; 9.
82. Adamovics, J.A. Ed.; *Chromatographic Analysis of Pharmaceuticals*; Marcel Dekker, Inc.: New York, 1990; 49, 225.
83. Schirmer, R.E. Ed. *Modern Methods of Pharmaceutical Analysis*; CRC Press Inc.: Boca Raton Florida, 1990; II, 227.
84. Grob, K. On-line Coupled High-Performance Liquid Chromatography: Gas Chromatography. *Trends Anal. Chem.* **1989**, *8*, 162.
85. Riekkola, M.L. Applications of On-line Coupled Liquid Chromatography: Gas Chromatography. *J. Chromatogr.* **1989**, *473*, 315.
86. Cone, E.J. *Pharmaceutical Analysis, Modern Methods*; Munson, J.W., Ed.; Chapter 3 Marcel Dekker, Inc.: New York, 1981.

CHIROPTICAL ANALYTICAL METHODS

Neil Purdie

Oklahoma State University, Stillwater, Oklahoma, U.S.A.

INTRODUCTION

The word chiroptical is descriptive of the techniques that use optical detection devices that are selective toward optically active (chiral) materials and/or molecules. They are used for structural investigation and analytical determination. There are three chiroptical techniques:

1. Polarimetry, which deals with the angular rotation of plane polarized light, usually at a single wavelength.
2. Optical rotatory dispersion (ORD), in which the angular rotation of the plane polarized light is measured as a function of the wavelength.
3. Circular dichroism (CD), in which the angular rotation is measured as a function of wavelength, but the light is circularly polarized.

Absorption of light energy is not essential to either polarimetry or ORD. It is, however, an integral part of the CD phenomenon making this method the most selective detector for chiral substrates.

THEORY

Theories of optical activity are described in detail in a number of studies (1–6). The physical phenomenon was first observed during experimental investigations of the transmission of solar radiation through Iceland spar, a natural form of CaCO_3 , by the French astronomer Arago. One year later, Biot (1812) was the first to demonstrate that solutions of certain organic compounds also rotate a beam of incident polarized light. Biot and Fresnel, working independently (1817), reported that the rotatory power of a substance increases as the wavelength is decreased, the phenomenon now called ORD. By 1846, Haidinger had reported differences in the measured absorptions of left and right circularly polarized light, which is the origin of CD.

The first experimental interpretation of the physical basis for optical activity was provided by Pasteur, who observed the hemihedrism of tartrate crystals, which was visually manifest by tetrahedral facets oriented either right

or left with respect to the main crystal surfaces for two crystalline forms. His observation that a linearly polarized beam of light was rotated in opposite angular directions by aqueous solutions prepared from the separated crystal forms demonstrated the first direct connection between macroscopic and microscopic, or molecular, asymmetry.

The first theoretical model of optical activity was proposed by Drude in 1896. It postulates that charged particles (i.e., electrons), if present in a dissymmetric environment, are constrained to move in a helical path. Optical activity was a physical consequence of the interaction between electromagnetic radiation and the helical electronic field. Early theoretical attempts to combine molecular geometric models, such as the tetrahedral carbon atom, with the physical model of Drude were based on the use of coupled oscillators and molecular polarizabilities to explain optical activity. All subsequent quantum mechanical approaches were, and still are, based on perturbation theory. Most theoretical treatments are really semiclassical because quantum theories require so many simplifications and assumptions that their practical applications are limited to the point that there is still no comprehensive theory that allows for the predetermination of the sign and magnitude of molecular optical activity.

A chiral substance is defined by the International Union of Pure and Applied Chemistry (IUPAC) as one that interacts differently with left and right circularly polarized light. Two types of molecular optical activity are recognized: inherent dissymmetry characterized by large rotational strengths and inherently symmetrical, but asymmetrically perturbed, molecules for which rotational strengths are less by a factor of a thousand or so.

The first group is characterized by the absence of a plane of symmetry in the molecule, e.g., hexahelicene. The latter type requires the existence of a chromophore in close proximity to an asymmetric center, such as a carbonyl group adjacent to an asymmetric carbon atom.

Polarimetry

Unpolarized light is thought to consist of an infinite number of time-dependent electric and magnetic fields that vibrate in

phase and at right angles to each other in planes that are orthogonal to the axis of propagation. Only the electric vector is considered in theoretical discussions of optical activity. Linearly polarized light is represented by only one of these planes and is given by the vector sum of two in-phase components of equal intensity that are circularly polarized in opposite directional senses (Fig. 1). The components actually propagate in a helical manner with time; however, the polarization projection on the plane, which is orthogonal to the axis of propagation, is circular, thus, the acquired description of light as circularly polarized.

As linearly polarized light is transmitted through an achiral medium, a single refractive index η is seen by both circularly polarized components, and their rates of propagation are equal. The result is that the vector sum is always a linearly polarized beam oriented along the direction of the incident beam, OO' in Fig. 1a. In contrast, because of the distinctly different interactions that occur between the two helical electromagnetic fields and the helical electronic motion in a chiral medium, two different refractive indices, η_L and η_R , are presented to the coherent beam (birefringence). There being two refractive indices, the left and right components propagate out of phase. On summing the vectors for the transmitted beam, represented for instance by the diagonal of a parallelogram for which OL and OR are adjacent sites, the polarization is still linear; however, a net rotation from the incident direction by an angle equal to α will have occurred (Fig. 1b). Rotational strengths are equal

and opposite for optically active molecular or mirror-image pairs (enantiomers) of equal purity.

The magnitude of the optical rotation α (in degrees) is directly proportional to the refractive index difference and to the sample pathlength d , indicative of the fact that rotation is an extensive property, as shown in Eq. 1:

$$\begin{aligned}\alpha &= (\pi d/\lambda)(\eta_L - \eta_R)(1800/\pi) \\ &= (1800d/\lambda)(\eta_L - \eta_R)\end{aligned}\quad (1)$$

and inversely proportional to the wavelength, in keeping with the observed increase in rotatory power with decreasing wavelength. The quantity $(1800/\pi)$ is included to convert radians to degrees. The magnitude of the birefringence $(\eta_L - \eta_R)$ that produces an angle of rotation equal to 1.0° at the Na-D line (590 nm), for a sample with a 10-cm pathlength, is only 3.2×10^{-8} . Because refractive indices are typically approximately 1.0, it is obvious that the absolute size of the birefringence effect is extremely small.

To normalize rotational values when comparing solutions of different concentrations, the specific rotation $[\alpha] = \alpha/c'd$ was defined, where c' was expressed in g/cm^3 . This unit is an improper choice for making comparisons among substances with different molar masses M , and therefore $[\alpha]$ was replaced by the molecular rotation term $[\Phi] = [\alpha], M/100$. In the older literature $[\Phi]$ was expressed in degrees $\times \text{cm}^2/\text{decimole}$. Division by 100 had no physical meaning whatsoever and was introduced only to keep numbers small. IUPAC has determined that the term molecular rotation is improper and recommended that it be replaced with the more accurately descriptive molar rotation.

OPTICAL ROTATORY DISPERSION

Two types of ORD were first described in 1852 by Biot. In his earliest quantitative experiments on quartz, he demonstrated that the optical rotatory power α varies inversely with the square of the wavelength: $\alpha = \kappa/\lambda^2$. Measurements on a large number of chiral organic compounds, dissolved in solvents both chemically and optically inactive, showed that most of these appeared to obey this law. Originally referred to as the "orthodox" class, these compounds are now thought to produce a plain or normal ORD curve. The distinctive property of the plain curve is that it is always concave to the $\alpha = 0$ -axis, regardless of whether the dispersion is positive (Fig. 2a) or negative. A substantial number of organic molecules, however, were found that did not appear to obey this law but had enormously large rotational powers compared with plain curves, which were limited to relatively narrow ranges

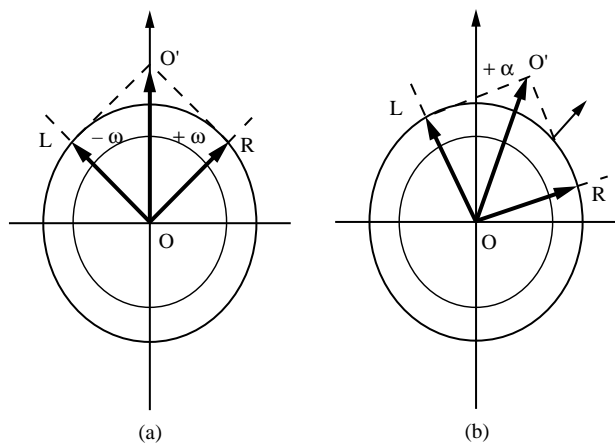


Fig. 1 The incident linearly polarized light OO' is composed of left and right linearly polarized rotating components OL and OR of equal length. In a chiral medium (a), the rotations are in phase, and the result of the two components is always in the same plane as the incident light. In a chiral medium (b), the components rotate out of phase. The resultant, represented by the diagonal of the parallelogram $OLO'R$, is rotated through the angle $+\alpha$ from the incident plane.

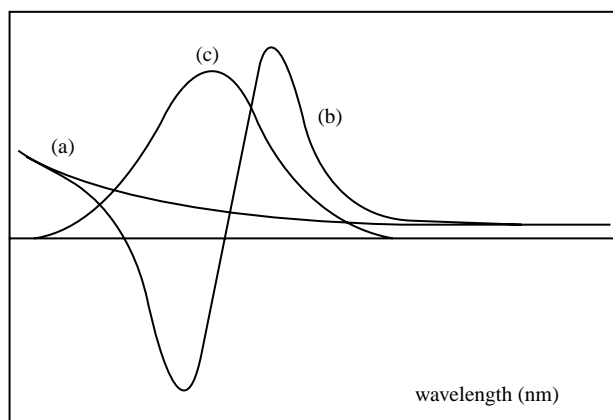


Fig. 2 Composite diagram of the three chiroptical dispersion spectra: (a) A positive plain ORD curve; (b) a positive anomalous ORD curve (Cotton effect); (c) a positive CD curve for a single Cotton effect.

in the spectra. Biot referred to this smaller group as the “heretical” class because of their anomalous behavior (Fig. 2b). Tartaric acid was the seminal example. In an effort to more accurately distinguish between the two types, Lowry specified that for normal ORD behavior, the specific rotation α and its first and second derivatives with respect to wavelength must all maintain the same sign throughout the wavelength range over which the medium is transparent. In other words, it is a mathematical statement to the effect that there should be no zero, no maximum, and no reversal of sign for α as the spectrum is scanned, that is, the curve is always concave to the axis (Fig. 2a).

Among the first experimental discoveries regarding the origins of anomalous dispersion was the observation that the effect could be created by mixing pairs of natural products that generate plain ORD curves, provided they were of opposite signs and unequal rotatory strengths, for example, (l)-turpentine and (d)-camphor. This observation was to be of fundamental importance in the subsequent development of theories for anomalous dispersions. For solutions of tartaric acid, a single pure substance, the existence of an anomalous dispersion was more difficult to explain. It was first assumed that an equilibrium mixture of two molecular forms that generate plain curves of opposite signs must exist in solution. Eventually, however, the effect was correctly interpreted as being a consequence of the molecule having two asymmetric centers that give rise to the “required” unequal and opposite rotations.

The development of several mathematical models and interpretations followed, with the best interpretation being proposed by Drude in Eq. 2, where $i = 1, 2, 3 \dots m$:

$$\alpha = \sum_i \left\{ k_i / \lambda^2 - \lambda_i^2 \right\} \quad (2)$$

In modern terms, this is rewritten using molar rotational values by replacing α with $[\Phi]$ and k_i with A_i . Drude originally referred to λ_i as the “characteristic vibrational” wavelength, meaning that there were periods of vibration of the charged particles that, when close to the vibrational period of the incident light, would produce the anomalous effect. Again, in modern terms, these are identified with wavelengths of maximum absorbance in the electronic absorption spectra. Whenever $\lambda > \lambda_i$, that is, the wavelength of observation is outside the range of an absorption band, the Drude equation is reduced to the one-term Biot expression for a plain curve. As the value of λ approaches λ_i , α increases asymptotically, reaching infinity at $\lambda = \lambda_i$. Immediately past the maximum wavelength, α is numerically close to minus infinity, and as λ continues to decrease, the curve follows along an inverse asymptotic path towards zero (Fig. 2b).

The anomalous positive ORD curve in Fig. 2b is rounded at finite values for the maximum (peak) and minimum (trough) extremities. The crossover wavelength, where $\alpha = 0$, generally coincides with the wavelength of the maximum absorbance. An anomalous curve is always superimposed on a fundamental plain curve that is alluded to as the background rotation. Media confirmed to have just a single anomalous dispersion can be solved for λ_i . Historically, this procedure was used to predict the wavelength maximum for an incomplete absorbance band that could not be observed in its entirety because of instrumental limitations. This particular application of ORD is now obsolete.

An anomalous curve is referred to as a Cotton effect in honor of the French physicist Aime Cotton (7) who, in 1892, was the first to point out that the absorption of light energy was the other physical property behind anomalous ORD and CD. Positive and negative anomalous dispersions are equally evident in practice. In ORD, the sign of a Cotton effect, by convention, is defined to be positive when the peak precedes the trough as the wavelength decreases and vice versa. A simple structural way of looking at the origins of ORD in a single molecule is to imagine that the fixed asymmetry of a saturated chiral group induces a degree of dissymmetry into an unsaturated and therefore symmetrical functional group or chromophore. In theory $\sigma \rightarrow \sigma^*$ electronic excitations are possible for saturated molecules for which the theoretical limit of λ_i was determined to be ≈ 150 nm. Chromophores, on the other hand, absorb at much longer wavelengths in the easily accessible range of modern instrumentation. The closer 150 nm is approached, the harder it is to measure a spectrum. Whenever the induced dissymmetry is opposite in sign to the fixed asymmetry, anomalous dispersion is produced. The mutual proximity of the asymmetric center

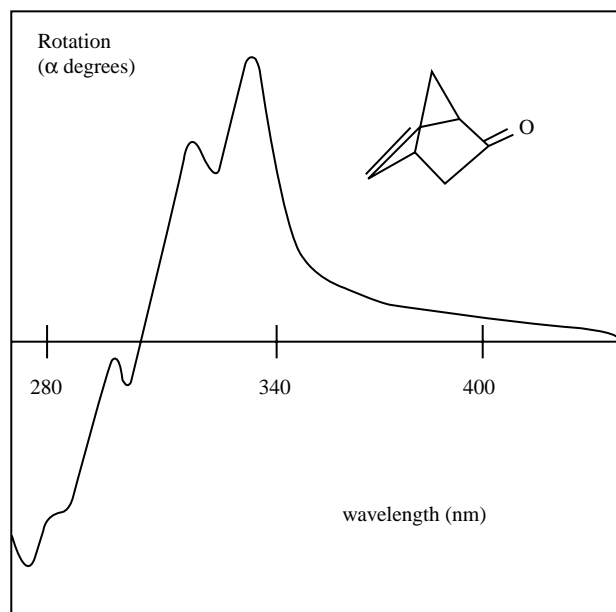


Fig. 3 The anomalous ORD spectrum consisting of several overlapping Cotton bands of the unsaturated rigid ketone, bicyclo(2,2,1)hept-5-enone, which shows how complex the spectrum can be, even for a single molecule.

and the chromophore is a necessary prerequisite to CD induction. Dissymmetry is the preferred description for the induced chirality because the chromophore might well have a high degree of axial symmetry. How this perturbation might occur has never been satisfactorily explained.

Fig. 2b is an idealized illustration of a single, uncomplicated Cotton effect. In reality, the occurrence of a complete curve in the electronic spectrum is rare. Complete dispersions are more likely to be observed in the vibrational spectral range because of the increased spectral resolution. However, even there, dispersions are too often complicated by extensive band overlap. The same is true for electronic spectra where hidden absorption bands coupled vibronic excitations and interferences from bands associated with other chiral chromophores contribute to producing anomalous ORD curves that are so complex they have little utility in quantitative analytical applications (Fig. 3).

CIRCULAR DICHROISM

Because absorption is a prerequisite to CD activity, the phenomenon is limited to only those wavelength ranges that encompass an absorption band in any part of the

electromagnetic spectrum. Outside the range of absorption, the CD signal is zero, which is the first important advantage CD has over ORD as an analytical detector. It should be emphasized, however, that the absence of a band is not evidence of the lack of chirality in the substrate.

At the time that Cotton was correctly interpreting the physical origins of anomalous ORD behavior, he proposed that there is also a difference between the absolute absorbances of the two circular polarized beams by a chiral medium (dichroism) and that the magnitude of the dichroism is proportional to the absorbance difference. Convention has dictated that the difference is always written as the absorbance of the left rotating beam minus the absorbance of the right, $\Delta A = A_L - A_R \neq 0$. Using the Beer–Lambert law to convert A to molar units, the dichroism expression can be rewritten as $\Delta \epsilon = \epsilon_L - \epsilon_R$, where ϵ has the units of L/mol cm.

A single positive CD band is shown superimposed on the anomalous ORD spectrum in Fig. 2c. The wavelength of the maximum CD coincides with the crossover wavelength of the ORD dispersion. Shorter lengths for the vectors OL and OR compared with the incident vector in Fig. 1 are used to convey the fact that absorption has occurred. The absorbance difference at a given wavelength is then represented by unequal vector lengths $OL \neq OR$. The resultant of OL and OR, given by the instantaneous diagonal vector OO' of the parallelogram OLO'R, no longer oscillates in a single plane, but traces out the perimeter of an ellipse as OL and OR rotate around an angle 2π . The transmitted beam is rotated by the angle α from the original plane of polarization (owing to birefringence) and is elliptically polarized owing to dichroism (Fig. 4).

The eccentricity of the elliptically polarized light is characterized by the term ellipticity Ψ equal to the arctangent of the ratio of the minor to the major axis of the ellipse and given by OA/OB in Fig. 4. Because the ratio $(\Delta \epsilon / \epsilon)$ necessary to produce an observable CD signal is as small as 1 part in 10^7 , the ellipticity is approximated almost exactly by the expression $\Psi = \pi(\epsilon_L - \epsilon_R) / \lambda$, which is entirely analogous to Fresnel's equation that relates birefringence to α and points up the common origins of anomalous ORD and CD.

In keeping with the older definitions of terms that are part of polarimetry, there are definitions for specific ellipticity $[\Psi] = \Psi \cdot c \cdot d$, and molecular ellipticity $[\Theta] = [\Psi] M / 100$, where M is the molar mass. With appropriate substitutions, the molecular ellipticity can be expressed in terms of ϵ , namely $[\Theta] = 3300(\epsilon_L - \epsilon_R) = 3300\Delta \epsilon$. The numerical constant is the result of all the physical conversion factors. The survival of these arcane units is a

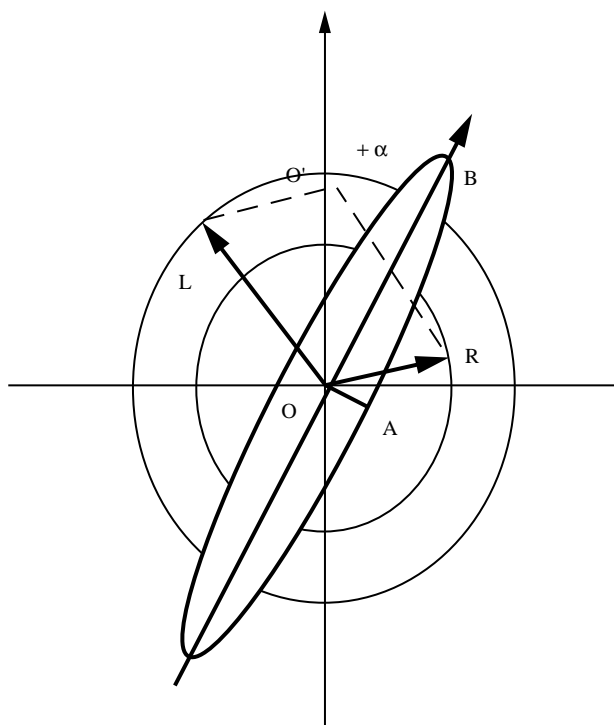


Fig. 4 Production of elliptically polarized light in CD. The direction of polarization of the incident beam is OO' . The unequal vectors OL and OR represent the difference in absorbances of the left and right rotating component beams. The angular rotation is $+\alpha$, and the ellipticity is the arctan (OA/OB) .

consequence of the wealth of informational data already in the literature. The disclosure that CD is no more than a modified absorbance technique should ultimately motivate investigators to adopt the term molar ellipticity, Θ_M , in the CD analog of the Beer–Lambert law, that is, $y = \Psi = \Theta M c d$.

Anomalous ORD and CD both originate from light absorption by a chiral species and as such contain the same information. A mathematical equation, the Kronig–Kramers transform, relates one to the other over the wavelength range of the absorption, namely, $[\Theta(\lambda)] = -2/\pi [\Theta(\lambda')](\lambda'^2/\lambda^2 - \lambda'^2)d\lambda'$. When the appropriate substitutions are made, the equation relating ORD to CD reduces to $\Theta = 40.28\Delta\epsilon$.

Because all the rules that apply to absorbance detection apply equally well to CD, it is convenient to think of CD as a modified form of absorption spectrophotometry. Spectra are temperature- and pH-dependent; nonlinear correlations of signal versus concentration are commonplace and are produced for the same reasons, such as chemical equilibria, polychromatic radiation, stray light, etc. Fluorescence emission CD (FDCD) spectroscopy is observed whenever an

analyte meets all three of the structural prerequisites simultaneously, either intrinsically or extrinsically. Therefore, anyone with experience in absorption and emission spectrophotometries can easily become acquainted with the experimental capabilities of CD. Similarities end there, however, and the differences are what make CD detection unique, especially the enhanced selectivity that arises from the fact that CD bands can be positive or negative in sign. Electronic absorption bands are generally broad and lack the kind of resolution associated with the infrared range. In contrast to visible-UV absorptions, however, exciton coupling can divide CD bands into two sub-bands of opposite sign and unequal intensity separated by a characteristic crossover wavelength where the signal is zero (6), resulting in narrower bands than those given by absorption. Circular dichroism is used most often for the analysis of bulk samples and has seen limited use in liquid chromatography and capillary electrophoresis.

CD activity can be induced into molecules that are either chiral or achiral and is generally referred to as extrinsic CD. For chiral species, intrinsic and extrinsic CD effects are additive. Compared with intrinsic CD, the extent of extrinsic or induced chiroptical effects is small. One way to induce activity is to apply a static magnetic field whose strength is on the order of 10–20 kGauss. Magnetically induced CD (MCD) was originally described by Verdet and correctly interpreted by Faraday in what has become known as the Faraday effect. A magnetic field of sufficient strength splits the degeneracy of the electronic ground and/or excited states (the Zeeman effect), resulting in absorbance differences between the two circularly polarized beams. The effect is entirely general and can be observed in every dielectric substance that transmits light. The magnitude of the effect depends on the relative orientations of the light path and the magnetic field strength and is at maximum when the fields are parallel. For a fixed geometry, the maximum signal is proportional to the sample pathlength and the analyte concentration. Poor sensitivities and even poorer selectivities associated with MORD and MCD make them unacceptable as analytical detectors.

A second way to induce chiroptical behavior is to associate a chiral center on one molecule with a chromophore on another by some aggregation or complexation reaction. If the chiral moiety is CD-inactive, only the resultant complex exhibits CD activity. The intensity of the induced CD signal is determined by two factors: the concentration of the complex that is formed and the magnitude of the induced $\Delta\epsilon$ term. Magnitudes and selectivities of chemically induced CD are much greater than those of MCD and have correspondingly

higher potential for analytical applications. Typically, the correlations of signal amplitudes with analyte concentrations are nonlinear.

INSTRUMENTATION

The basic instrumental needs for chiroptical methods are virtually the same as for other spectroscopic methods, namely, a stable unpolarized illuminating source of sufficient intensity, a wavelength-selection device, sample holder, and detector; polarizing elements are essential. Because the only parameter measured in polarimetry and ORD is rotation, the polarizing elements are common to both. A monochromatic source, such as an Na or Hg lamp, is all that is required for polarimetry. Deuterium or halogen lamps are of sufficient intensity for ORD, but highly intense (150–450 W) Xe arc lamps are needed for CD.

Polarizing elements are transparent rhombs constructed by joining together two triangular prisms cut from a single crystal of calcite or quartz. The junction between the two parts may be just air or a light-weight balsam cement. The purpose of the junction is to physically separate the ordinary and extraordinary rays of the linearly polarized light beam, allowing only one ray to pass while the other is selectively reflected in a direction at a right angle to the first. Often, the reflected ray is fully absorbed to eliminate any interference with the transmitted ray. An excellent historic account of the assembly of the parts into working polarimeters is given by Lowry (1).

In polarimetry and ORD, the sample is placed between the first polarizing element (the polarizer), which remains fixed, and the second element (the analyzer), which can be rotated about the axis of propagation. Maximum intensity of the transmitted light is observed when the principal axis of the polarizer and analyzer are colinear and exactly parallel. The intensity is zero when they are crossed; that is, when the principal axes are orthogonal to each other. The most accurate way to determine the rotation angle α is to set the polarizer and analyzer in the crossed position using an achiral substrate and to measure the extent to which the analyzer has to be turned to restore the optical null position when the achiral sample is replaced by a chiral substrate.

Optical rotations are temperature-dependent. For the most accurate work, sample cells must be thermostatted. Solution concentrations are typically above 0.2 M for polarimetric detection, and pathlengths range from 1.0 to 100 mm; volumes vary from 0.1 to 50 ml. Because rotations increase in magnitude with decreasing wavelength, the best sensitivities using conventional light

sources are achieved in the UV. Accuracies are reported to be on the order of $\pm 0.2\%$ for rotations $> 1.0^\circ$. Only the most sophisticated high-sensitivity polarimeters meet the requirements for chromatographic detection. With a stable laser system as the illuminating source, rotations as small as 10^{-10} to 10^{-11} radians can be measured fairly accurately. High sensitivities are critically important in chromatography because concentrations of eluted components are very low, being limited by the retention capacity of the column materials, and because pathlengths, viewed across the eluant exit tubes, are very short (8).

The first commercial ORD spectropolarimeters appeared in the 1950s but are no longer available. The ORD capability is typically offered as an add-on to a CD spectropolarimeter.

At present, technical difficulties associated with scanning chiroptical methods prevent the use of diode-array detection, and therefore wavelength is selected in CD with a scanning double monochromator set-up (Fig. 5). The block diagram for CD differs from ORD instrumentation by the addition of an electro-optic modulator, placed immediately after the linear polarizer, to generate the phase-separated left and right circularly polarized component beams that are the origins of the elliptically polarized light beam. The physical parameter that was measured in the first CD instruments was the ellipticity of the transmitted beam (Fig. 4). Greater accuracy and greatly improved sensitivities are achieved if the absorbance difference is measured, which is the procedure preferred by every contemporary CD instrument manufacturer.

Because of significant losses of radiant power on polarization and transmission through the double monochromator system needed to keep stray light to an absolute minimum, the light sources for CD detection must be intense. As a consequence, instrument compartments are purged with nitrogen to remove ozone that might be produced by the high-intensity radiation of oxygen. The detector is a photomultiplier tube. Wavelength ranges on commercial instruments extend from 180 to 850 nm. Instruments for the vibrational spectroscopy range are still only custom built. The ellipticity, or $\Delta\epsilon$, scale should be calibrated daily against selected standards. Scale calibration is wavelength-dependent, and whenever the range of study is very broad, the use of more than one standard is recommended. Those most commonly used are androsterone, pantoylactone, (+)-camphor-10-sulfonic acid, and ammonium camphor-10-sulfonate for the near-UV, and alkaline nickel(II) tartrate in the visible.

With absorbance differences on the order of only 1 part in 10^7 for CD activity, the ratio of transmitted intensities for the left and right circularly polarized beams (I_L/I_R)

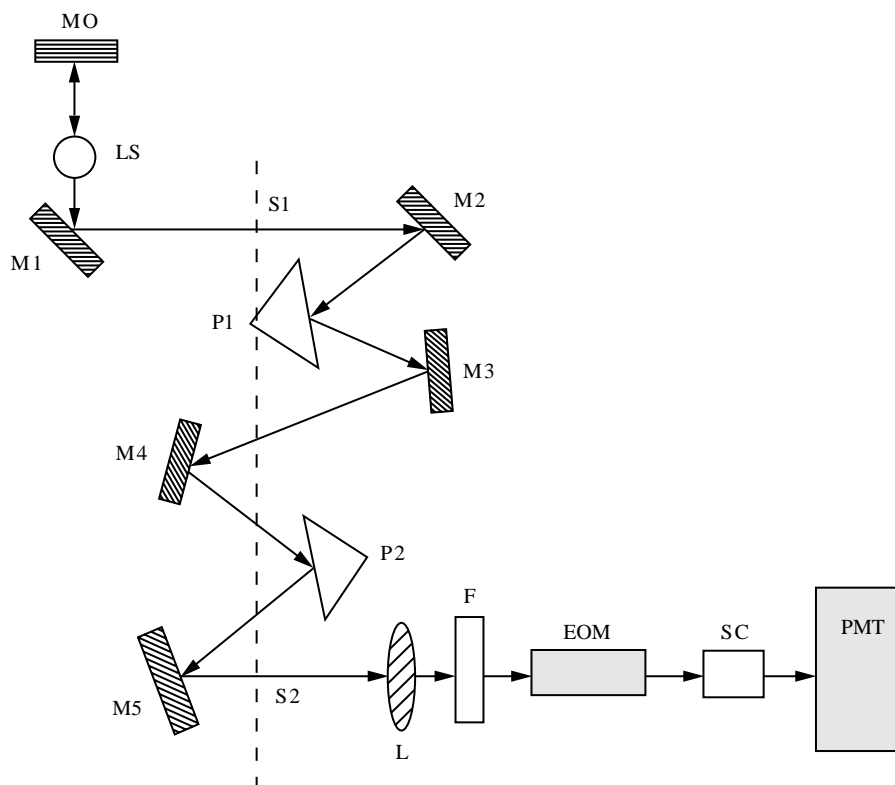


Fig. 5 Block diagram for a commercial double-monochromator CD spectropolarimeter. M = various mirrors; S = slits; P = polarizing elements; LS = light source; SC = sample; PMT = photomultiplier tube; L and F = lens and filter, respectively; EOM = electro-optic modulator.

is essentially one because the errors in $\Delta\epsilon$ would be very large if I_L and I_R are measured directly.

$$\Delta\epsilon = (\epsilon_L - \epsilon_R) = [(1/c \cdot d) \log(I_L/I_R)] \quad (3)$$

The problem is overcome instrumentally by measuring the intensities I_L and I_R separately, at a frequency of 5 kHz. This has the effect of producing an AC voltage proportional to the CD signal, riding on top of a steady-state DC component proportional to the total absorbance at each wavelength. The independence of I_L and I_R from the steady-state DC voltage indicates that the size of a CD signal is not dependent on the absolute magnitude of the absorption, and relatively strong CD signals can be obtained from overall weak absorbers. Although an absorbance difference is measured in CD detection, the total absorbance by the substrate and matrix is still a limiting feature because it can affect the intensities of the transmitted beams to be measured. Excessive amplification of very weak signals increases the noise level and adversely affects the quality of the CD signal.

The principal electronic excitations in the accessible UV range that lead to absorption by organic molecules are

the $\pi - \pi^*$ transitions associated with the aromatic ring and the $n - \pi^*$ transitions of carbonyl functional groups. Excitations associated with $\pi - \pi^*$ transitions have a high probability, and absorbances are highly intense. To preserve the signal quality, solutions must be very dilute and/or pathlengths must be short, which is the second advantage that CD has over polarimetry and ORD. The photomultiplier (PMT) measures the total transmitted intensity and is incapable of discriminating between chiral and achiral species. Besides affecting the signal/noise ratio, excessive absorptions reduce the linear dynamic range of the detector. At the low concentration end, the determining factor is the very small size of the CD signal, and the upper limit is determined by the total absorption. Ranges are often much narrower than they normally are using absorption detection.

Current CD instruments commercially available for analytical applications are limited to the electronic excitation range of the electromagnetic spectrum. The more sophisticated of these have the added capability of pulling an eluate from a chromatographic separation off-line into a microcell attachment where, instead of limiting

detection to just one wavelength, a partial CD spectrum can be measured. Volumes can be as small as 10 nL (8). Instrumentation for the measurement of vibrational CD (VCD) and Raman optical activity (ROA) are still custom-built, although the prospects for their commercial development in the not too distant future are bright. A major disadvantage is, of course, that the emission intensities of tunable IR sources are generally weak.

An intriguing recent development in CD detection is its extension to the wavelength range of soft X-Rays using a synchrotron source (9). Although this might never become a routine analytical method, it has been speculated that from CD measurements made in this range, it will at last be possible to indisputably determine the absolute conformation of a chiral molecule of any size in solution. This would make it superior to NMR detection, which is limited to small molecules and single-crystal X-ray structure analyses, in which, for the want of other methods, structures are usually assumed to be the same in solution.

ANALYTE SELECTION

In deciding whether analytes are CD-active, it is not always a simple matter to inspect a molecular formula and be certain that the chromophore and the chiral center are mutually located in a manner that produces activity. Even if the molecular structure suggests that a chirally perturbed chromophore is present, the substance might only be available as an achiral racemic mixture and therefore is not detected by CD.

Optimum wavelength ranges for CD detection are those where the absorption is minimum and the CD signal maximum. Absorptions for $n - \pi^*$ and $\pi - \pi^*$ transitions are generally weaker at wavelengths longer than 230 nm, where they appear as shoulders on the edges of the intense bands that reach a maximum at shorter wavelengths. Frequently, CD bands in the range of 230–340 nm are intense enough to allow quantitative analysis, e.g., for testosterone and dihydrotestosterone (Fig. 6).

Electronic excitation energies are shifted to longer wavelengths as the extent of molecular conjugation increases. The molecular symmetry that accompanies the structural planarity created by conjugation might reduce the number of potential achiral centers and the chances of observing intrinsic CD activity, for example, in organic dye molecules. Strong absorptions by dyes, however, are exploited by associating them with a chiral molecule to induce an extrinsic CD activity in the longer wavelength range, where passive absorption by the matrix is less of an interference. Other absorbers of this capability are colored

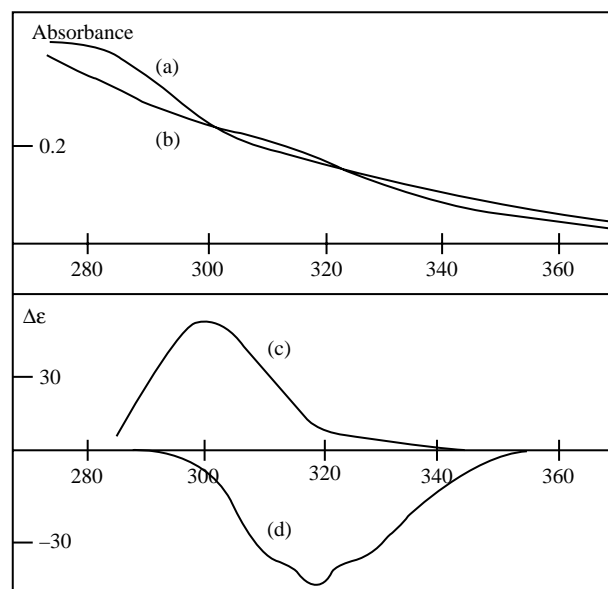


Fig. 6 Demonstration of the strong similarity between the absorbance spectra for (a) dihydrotestosterone and (b) testosterone in methylene chloride in contrast to their CD spectra (c) and (d), respectively.

chiral metal complexes that have the special advantage that their absorbances are generally of much lower intensity than those of organic dyes.

Similarities between CD and absorbance methods are also found between CD and fluorescence and CD and circularly polarized luminescence (CPL). Three prerequisites are needed to produce FDCD and CPL activities (8). Intense emission signals normally associated with fluorescence are attractive because limits of detection are lowered considerably. FDCD finds more uses as a chromatographic detection device. A CD signal is usually induced by some kind of molecular complexation reaction. Association can be with a simple molecule or with an aggregate of molecules, such as chiral micelles, which are known to be fluorescence enhancers. In cases of color induction combined with fluorescence induction, FDCD can lead to even higher levels of selectivity among analytes that have been derivatized by the same color reagent.

Selectivity enhancement is a result of a number of circumstances. For example, not all of the sub-bands in the absorption or fluorescence spectrum of the derivative are necessarily CD-active; because only the chromophore next to the chiral center must fluoresce for CD activity, all other fluorescence that centers on the analyte, or in the matrix, will not interfere with the FDCD signal, and if a chiral center is lost in the derivatization process, that molecule will be removed from the list of interferences.

CHIROPTICAL DETECTION IN CHROMATOGRAPHY

Only polarimetry and CD find practical use as chromatographic detectors (8). Important parameters to consider in modifying chiroptical detectors for use in chromatography are the very small sample volumes involved, which today can be handled with relative ease, and the very short time intervals that separate consecutive peaks, which has not been, and probably never will be, totally resolved. Peak overlap remains a significant problem. In molar terms, limits of detection are not particularly impressive, for example, micromolar levels, but in volumes as little as 1.0 μl , the limits of detection are actually at the nano- to picomole level and sometimes lower (10–14). Engineering priorities are to develop the technology to focus the beam on such small targets while maintaining the high level of radiance needed for chiroptic detection. CW lasers are an obvious place to start, but source noise and instability are problems to contend with (14). By exploiting the added radiance of pulsed lasers, limits of detection can be stretched to even lower levels. Laser sources, however, are limited in the number of their output wavelengths. Dye lasers offer the best, albeit still very narrow, ranges (approximately 60 nm). Current CD instruments, in which the source is laser illumination, really do operate at just a single wavelength, depriving the detector of its ability to identify an analyte.

Options for multichannel LC–CD detection do exist (14–16). Stopped-flow accessories for commercial instruments are available that allow part of an eluted fraction to be taken off-line into a microcell placed in the regular sample compartment where data are measured in the normal way. The method still requires rapid scanning capabilities. Repeated injections and multiple scans can be averaged to improve the quality of the signal. A major deterrent to the progress in the early development of HPLC–CD detection was the lack of a dedicated instrument at a reasonable cost, the only option being a fully equipped CD instrument.

In 1998, Jasco International Co. introduced the first dedicated commercial polarimetric detector, the OR-990, and the first dedicated CD cum absorbance detector, the CD-1595, for HPLC. The CD detector operates in the 220–420 nm range, with a 20 nm bandwidth. The illuminating source is a 150 W Hg–Xe lamp. Injections are typically in the microgram range. Minimum detectable amounts are on the 0.1-ng scale. The sensitivity of the CD detector is typically 200 times higher than that of the OR-990 detector but of a factor of four lower than the absorbance detector of the CD-1595. The latter is the

limiting factor in the combination of CD with absorbance detection when applied to enantiomeric purity measurements. This last assay has taken on considerable new meaning with the obsessive focus on measuring the enantiomeric purities of chiral drugs in the biotechnology and pharmaceutical industries.

Eventually these devices may turn out to be the starting point for the development of CD diode-array detectors. Adjusting the scanning speed for on-line, wide-spectrum CD measurements is a formidable problem. A major reason for the problem is the incongruity between the time it takes to accumulate CD data, even for just one spectral “pass” using the very best currently available diode array technology, and the typical dispersion time between chromatographic peaks. The situation may very well change as faster electronic detection devices become available.

If all the components of a sample loaded on an HPLC column are baseline-separated, any conventional detector will work, unless the object of the separation is to determine or confirm the stereochemical conformation of an enantiomer. In achiral systems, (solvent and/or stationary phase) enantiomers have identical retention times and are not separable. The problem has a solution if two detectors are used in series, e.g., CD and absorbance (17–19). Because the enantiomers elute together, the absorbance detector measures the sum of their concentrations, and the CD detector measures the difference ΔA . Solving the simultaneous equations gives the concentrations for both enantiomers.

Historically, the experimental limitations of this procedure give totally meaningless results when enantiomeric ratios are greater than 95:5, or within 5% of being racemic (20). Concepts that have evolved as potential solutions to these experimental limitations making them capable of improving on the accuracies of enantiomeric purity determinations are the *g*-factor and principal component analysis (PCA) treatments of eluted band intensities as a function of time (21, 22). These are especially useful in cases in which bands are asymmetrical (22), which is frequent. The *g*-factor is defined as the ratio of the CD intensity to the absorbance intensity $\Delta A/A$. One attribute of this factor is that for an enantiomerically pure material, the *g*-ratio does not change with concentration. Should a change in the *g*-factor occur during the elution of a band, it is clear evidence for an enantiomeric impurity. Ideal traces documenting the total resolution of bands for two pure enantiomers in a racemic mixture would consist of two horizontal lines with constant *g*-values of equal and opposite signs. Over the time interval between the elutions, the traces are separated by signals that are excessively noisy. This occurs because calculated

g-factors in the ranges in which no CD-active species is being eluted correspond to zero divided by zero. With PCA, the number of components that are coeluted can be derived by reduction of a matrix of signal intensities versus concentrations versus time data for a series of solutions with prepared compositions (21).

Ostensibly, the better alternative is to separate the enantiomers on a chiral HPLC system, typically done by reacting both enantiomers of a racemic mixture with a third chiral species. The chiral derivatizing agent is an integral part of either the mobile phase or the stationary phase (19). The products are two diastereoisomer derivatives with different retention times. The same principle was used in classic experiments in which, for example, (–)brucine was added to separate enantiomers by fractional crystallization. It is still the only viable option to discriminate among enantiomeric forms by NMR. The numerous problems associated with chiral chromatographic methods are familiar (20, 21):

- The number of chiral derivatizing agents that are 100% enantiomerically pure is extremely small.
- Differences in retention times are very small if the material has several chiral centers.
- In practice, chiral solvents can be used as mobile phases only once.
- Even if derivatization is accomplished, baseline separation is not guaranteed.
- Racemization of the analyte may occur on the column during elution.

The protocol for separating a partial racemic mixture calls for the chromatographic conditions to be modified in such a way that the minor component elutes first (23) and is not lost in the trailing edge of the band for the major component. Errors encountered in the determination of enantiomeric excesses when they are in the range of 98–100%, or close to unique protocol is required for every chiral analyte assayed by chiral-HPLC, requiring considerable development time and constant review of the procedure.

DIRECT CHIROPTICAL DETECTION

In this context, direct means separation of the substrate, except solvent extraction, is not a part of the analytical work up (24–26). Only CD has the necessary selectivity to function as a direct detector. Chiral molecules that do not absorb (e.g., most simple sugars) do not interfere. Achiral molecules that absorb interfere to the extent that their absorption lowers the signal/noise ratio and the limits of

detection. Naturally occurring pigments and coloring agents added to pharmaceuticals are among the worst interferences.

Overlapping bands from multiple CD-active analytes are also a concern, although there is less of a tendency for this to happen with CD compared with absorption because bands are generally narrower and often have opposite signs. Curve-fitting algorithms might be used to resolve overlapping bands, but these kinds of solutions often lead to ambiguous results. More and more attention will be given to pattern-recognition strategies that involve data analyses that use chemometric methods such as principal component analyses and artificial neural networking (27).

Molar ellipticities in the preferred wavelength range of 230 to 340-nm for underivatized analyses typically differ by only a factor of two or three. By comparison, linear dynamic ranges are much greater than this, and the limiting property in discriminating among CD signals is the analyte concentration rather than the rotational strength of the chiral chromophore. Limits of direct CD measurements made on bulk samples using direct transmission detection are similar to those for absorbance, approximately 100 nM for a 1.0-cm pathlength.

REFERENCE CD SPECTRA

There are no comprehensive data files for CD spectra for standard reference materials (SRM) that compare with the exhaustive libraries which have been compiled for absorbance data in the electronic and vibrational spectroscopy ranges. Analysts are required to create their own CD spectral files using SRM prepared by the usual purveyors of fine chemicals. A significant problem with an SRM is that although it might meet the industry specifications for chemical purity, its enantiomeric purity is open to question. The few cases in which absolute enantiomeric purity might be assured involve natural products whose syntheses are under total enzymatic control. To prove 100% enantiomeric purity is beyond current capabilities (20). The problem is compounded even more with the risk that the material might racemize after its extraction from its natural environment. Therefore, it is not possible to assume absolute enantiomeric purity with firm conviction.

The superficial observation that CD spectra for enantiomers are exact mirror images of each other is only true if the two SRM used to calibrate the CD have equivalent enantiomeric purities. And even if the spectra are exact images, the evidence is not irrefutable proof that both SRM are 100% enantiomerically pure.

Added complications arise when an analyte molecule has two asymmetric centers for which there are a total of four optical isomers (R,R; R,S; S,R; and S,S). Together they constitute two pairs of diastereoisomers for which there are two pairs of "equivalent" CD spectra. It is conceivable then that the wrong analyte could be identified and assayed. The practical solution of these disconcerting uncertainties is to run regular checks on the reproducibility of the spectrum for a chemically pure SRM that has been "defined" to be enantiomerically pure. This is done by adding spectral data for every new issue of an SRM, supplied from different product lots by different manufacturers, to an ever-increasing data pool and periodically updating the statistically averaged spectrum as the reference spectrum. The inability to get standards of absolute enantiomeric purities takes on an even greater practical significance when attempts are made to assay enantiomeric excesses or enantiomeric purities in mixtures of isomers that may have been produced synthetically.

APPLICATIONS

The ubiquity of the aromatic ring and carbonyl chromophores in the molecular structure of natural products means that the number of potential analyses is enormously large. Djerrasi (2) pointed out the analytical potential of chiroptical methods as long ago as 1960, but even now, the number of investigations is small, which is explained in part by the enormity of the field of separation sciences. Most analysts would argue that the obsession with separation is because problems with interferences are minimized. On the other hand, for many of these processes, their potential was illustrated using carefully chosen synthetic laboratory mixtures, most of which were so simple they did not even begin to address the complexities that are encountered in the analysis of real samples.

The emphases of this section reflect the author's own special interests in using CD detection to directly determine chiral substrates. The majority of the systems described are drug substances. Direct analytical applications over the last 20 or so years have clearly demonstrated that a priori expectations of serious interference problems are ill-founded. Analytical sensitivities similar to those for absorbance spectrophotometry are readily accessible, and a high degree of analytical selectivity is obtained because of that very same property that makes ORD and CD such useful structural tools, namely, the sensitivity of a chromophore to its chiral

environment (3, 6). Substrates are organized into three groups:

1. Those with chiral chromophores that absorb in the near-UV
2. Those that are either chiral or achiral but do not absorb and are derivatized to absorb in the visible
3. Those that are achiral and absorb and have optical activity induced by interaction with a chiral host.

CHIRAL CHROMOPHORES THAT ABSORB IN THE NEAR-UV

The major analytes in this category are the alicyclic compounds (alkaloids and terpenes); heterocyclic compounds (barbiturates, benzodiazapams, indole alkaloids, quinolines, nucleic acids, and nucleotides); aminoacids and peptides; oligopeptides; and proteins (globular, nucleo-, and lipo-); saccharides and polysaccharides; and condensation products of saccharides with all the other analytes, e.g., glucuronides and glycoproteins (26). Thus far, most analyses have been done on solid and solution forms of the drug substances. A few illustrations are reported in which CD was used in the direct analysis of biological extracts.

Morphine Alkaloids

The first series of compounds assayed directly by CD detection were the morphine alkaloids. They were supported in aqueous solutions (28), in a chiral cholesteric liquid crystal solvent (29), and mixed in pellet form with solid KBr (30). Contrary to expectations, the homogeneous aqueous solution medium gave the best selectivity among 10 related opiates and the most quantitative results. The pH-dependence of phenol substituted analogs, which in some instances caused the sign of the CD signal to invert, enhanced the selectivity. Heroin was assayed both directly and as the morphine hydrolysate (31). Direct multicomponent analyses were made for prepared mixtures of morphine, codeine, thebaine, noscapine, and opium extracts (32).

Aromatic Amines

The strong structural similarities and the proliferation of enantiomeric forms in the phenethylamine and catecholamine series are major reasons for the considerable difficulties encountered in their analyses (20). Absorbance bands in the 250- to 320-nm range are identified with the aromatic ring and are nondiscriminatory.

Chiroptical detection methods have a slight edge over absorbance and a large advantage over electrochemical detection because of the birefringence factor. An ORD detection assay was developed to analyze mixtures of ephedrine and pseudoephedrine (33). An unprecedented advantage was found in the determinations of amphetamine and methamphetamine in cases in which achiral excipients such as lidocaine, procaine, and benzocaine had been added to deliberately confuse the assay by absorbance detection (34). These additives have absorbance spectra and retention times in achiral liquid chromatography that are too similar to the analytes.

Antibiotics

All the tetracyclines have intense visible CD spectra, with some degree of discrimination among them possible (35). The β -lactam antibiotics have very similar near-UV absorbance spectra, making some kind of separation the method of choice for their determination. Discrimination between the penicillin and cephalothin groups by CD detection, however, turned out to be an elementary exercise when mixtures of Pen-V and cephalothin were simultaneously determined with equal imprecisions in prepared laboratory mixtures (36). In contrast, discriminations among individual members of either β -lactam group is a very difficult prospect that will, in all likelihood, require a prior chemical derivatization. An ORD study reported the discrimination between the neomycin B and C aminoglycoside antibiotics. However, the strong similarities between the absorbance and CD spectra for the polymixin and bacitracin antibiotics (30) make their discriminations by direct assay impractical unless there is first a derivatization step.

Alkaloids

Other alkaloids assayed with varying degrees of success include the quinine–quinidine, cinchonine–cinchonidine, digoxin–digitoxin, L-Hyoscyamine–atropine, and pilocarpine–isopilocarpine diastereoisomeric forms. Being diastereoisomers, these have different chromatographic retention times, yet their assays are confused when compositions of the enantiomeric mixtures change. Prepared binary mixtures of the first two pairs of diastereoisomers were easily quantified using direct CD detection (37). Observed signals for the digoxins and pilocarps, on the other hand, are so weak that the best possible analysis was qualitative identification. The CD spectra for the colchicine, strychnine, brucine, and tubocurarine alkaloids, all

potent poisons, have been characterized in strong aqueous acids (38, 39). No reports of their being assayed by chiroptical methods have appeared.

Vitamins

In the area of vitamin analyses, CD spectra have been characterized for the water-soluble vitamins B2 and C. When they occur together, their distinction by direct measurement is an elementary procedure. Both were successfully assayed in the extracts of pharmaceutical preparation, as was B12 (40). Analysis of the fat-soluble D2 and D3 vitamins (ergocalciferol and cholecalciferol) has not been equally successful (38). Vitamin D extracted from natural sources has a single conformational stereochemistry that is one of several isomers produced in synthetic preparations. To certify that the natural form is present in a synthetic product, where it can be accurately assayed in the presence of the other isomers, is a formidable analytical task. Whether direct CD detection can satisfactorily solve it is currently unknown. A prior nonselective derivatization reaction might be required on all isomers. The A and E vitamins are achiral and not subject to chiroptical detection unless first derivatized by reaction with a chiral host.

Steroids

The seminal work on steroid analyses using chiroptical detection was done by Djerrasi by the determination of hecogenin acetate in the presence of tigonenin acetate (2). Every steroid is chiral and therefore amenable to polarimetric detection after chromatographic separation. Chromophores are fairly uncommon, and analysis by ORD or CD is therefore less suitable. The only unsaturation in the cholesterol molecule, for example, is the isolated Δ^5 -double bond, which has an absorbance maximum at 205 nm. Unsaturation coupled with chirality provides some selectivity, as ably demonstrated by the work of Potapov for analogs of progesterone (41). Even simpler than that is the direct discrimination between the ketosteroids testosterone and dihydrotestosterone, which have opposite signs in methylene chloride solution (Fig. 6).

Gergely promoted the development of ORD methods for the Δ^4 -3-Ketosteroids and the 17-Keto- and 17-Ethynyl derivatives (42). The 17-Keto derivative is often present as an impurity in the manufacture of 17-Ethynyl-substituted steroids and is easily quantitated by mathematically fitting the spectrum for the mixture using weighted spectra for the components or by measuring the spectra in two solvents. In the second option, data at two wavelengths are used to

prepare simultaneous equations that are solved for the concentrations of both components. The latter was used to quantitate mixtures of corticosteroids and Δ^4 -3-Ketosteroids (42). For the most accurate results, however, chemometrics methods are recommended with full spectral data.

Carbohydrates

Although included in this subsection, the only carbohydrates that meet the condition of absorbing in the near-UV are the keto-, amido-, and carboxylate-substituted sugars (43). Fully saturated sugars absorb only at wavelengths less than 200 nm and in general have incompletely developed spectra, with many not reaching a maximum signal. In the near-UV, excellent analytical data have been obtained for the in situ determination of D-fructose in honey (44) and of the *N*-Acetyl content of chitosan in crustacean shells (45). Simple sugars commonly exist in the form of equilibrium mixtures of open-chain and cyclic anomers (46), in which equilibrium must be established and the temperature controlled for the most accurate and reproducible measurements. Aldoses are typically determined by high-performance liquid chromatography (HPLC) using absorbance detection at approximately 300 nm. Polarimetry can be, and has been, used whenever information on the enantiomeric forms of the eluants is needed. Kuo and Yeung combined both these detectors for the analysis of several saccharides in laboratory mixtures. An advantage is that the focus is narrowed to cover only chiral absorbers, which simplifies the analytical identification of the eluates. Of the systems in which HPLC with CD detection was used (47), the most common are the simplest ketoses, D-fructose, D-tagatose, D-sorbose, turanose, D-ribose, and vitamin C.

Aminoacids, Peptides, and Proteins

Most of the findings related to CD detection of steroids and carbohydrates apply equally well to these analyses. Without derivatization, only aminoacids with aromatic side-chain substituents are CD-active in the near-UV. Signals are generally weak, and enantiomeric purity measurements using polarimetry detection are not quantitative. Peptides and proteins have stronger rotatory powers with obvious potential for clinical analyses. Nevertheless, the major exploitations of these data are toward elucidating secondary and tertiary structural information in aqueous media (20, 48, 49). With respect to analytical applications, there is a larger role for these macromolecules as auxiliary or host substrates in determining low-molecular-weight

substances that bind to the hosts in stereocontrolled ways, such as warfarin to human serum albumin for which FDCCD is the preferred detector (50).

The ubiquitous involvement of these materials in chirality induction for the purposes of assaying small molecules, determining enantiomeric purities, quality control, and quantitative structure-activity relationships (QSAR) is reviewed later in this article. The detector that is common to all these applications is CD.

Natural Products in Plant Extracts

A special example of the analytical selectivity of CD is its ability to directly assay natural products in plant extracts. Because there are no reference standards for plant materials, an assay is deemed to be successful if the results lie within the expected compositional ranges for that material (51). Analyses are not fully quantitative. Direct assays have been described for tetrahydrocannabinol and cannabidiol in marijuana extracts (52); S-Nicotine in leaf extracts from tobacco and tobacco products (53); Pen-V extracted from a crude fermentation broth (36); vitamin C from a variety of whole fruits, fruit juices, and whole vegetables (40); reserpine alkaloids from *Rauwolfia* (54); pyrethroid insecticides (11); and amaryllidacea alkaloids (17); atropine from *digitalis* (24); and humulone from hops (24). Obviously, the most serious interferences would be from the absorbance by the plant pigments and light-scattering from suspended materials. The CD assays offer the advantage that the pigments are not fully extracted into the selected solvents, leaving the near-UV virtually transparent to absorption.

Exceptions to this last observation are encountered when the colored materials happen to be present in the same phase where, because of their excessive absorbances, the signal-to-noise ratio is decreased. This kind of complication was successfully handled in direct analytical assays devised for lysergic acid diamide (LSD) (55) and phencyclidine (PCP) in illicit drugs spiked with intensely colored dyes (56); for L-Cocaine, morphine, and methadone in the pharmaceutical product commonly referred to as Brompton's cocktails (57); and for D-Pseudoephedrine in children's Sudafed (37).

SUBSTRATES MADE CD-ACTIVE BY COLOR INDUCTION

From what has been learned from the near-UV studies, the selectivity and the sensitivity of CD detection are greatly enhanced if the CD-active absorption bands are shifted

from the wavelength range where the matrix absorbances are highest. With a few exceptions, the range of least interference is the visible. Wavelengths are shifted by using selective color or fluorescence derivatization reactions on chiral analytes as they exist in the matrix.

Color derivatizations can be broadly divided between reversible and irreversible reactions. Reversible reactions typically involve some kind of complex formation equilibrium. The color originates on the host and is imparted to the analyte on the formation of the complex. The combination of chirality and absorbance that produces CD activity is limited to only the complex. Any absorbance by uncomplexed host or any residual chirality on the uncomplexed analyte is not detected and are therefore not interfering. Because these are equilibrium reactions, the correlation between the experimental ellipticity and the analyte concentration is nonlinear. An elegant and simple illustration of the capabilities of this kind of procedure is the determination of cholesterol in human gallstones in which association between the chiral cholesterol and colored bilirubin produced the CD-active complex (Fig. 7) (20, 50, 58, 59). Analogous reactions could be exploited for other naturally occurring pigments.

Other colored host molecules with obvious potential as reversible derivatizing agents are organic dyes and metal

complexes of the first-row transition metals (26). Dyes, of course, are inherently strong absorbers, whereas transition metal complexes are not; however, the CD signals for the complexes are of similar intensities. Dyes are used as prosthetic binding groups in the CD analysis of peptides, proteins, and oligo- and polysaccharides, although more often, the object of the study is to discover stereochemical information about macromolecular structures in solution and at binding sites (6, 60). Their applications as hosts for the chemical analysis of simple molecules, oligopeptides, and nucleic acids using CD detection will ultimately follow.

An example of a colored metal complex as an analytical prosthetic reagent is alkaline Cu(II)-tartrate, which is used routinely for the determination of total-plasma protein. Detection is by absorbance. If racemic tartrate is replaced by the L-Enantiomer, the analytical reagent itself is CD-active. Exchanges of analyte ligands with the L-Tartrate induce significant changes from the CD spectrum of the host complex—changes that can be used not only for structural information about interactions occurring between the ligands in the first coordination sphere, but also as an analytical method selective toward the incoming ligand (26, 61, 62). The assay is a take-off from the analytical discrimination between neomycin B and neomycin C using ORD detection (26). The discrimination power of CD detection allied with the ligand exchange reaction on copper-L-tartrate was demonstrated for the amikacin, gentamycin, kanamycin, neomycin, and streptomycin antibiotics (26) (Fig. 8). The same host complex used to measure enantiomeric excesses or ratios (61, 62) in difficult-to-measure compositional ranges is addressed below.

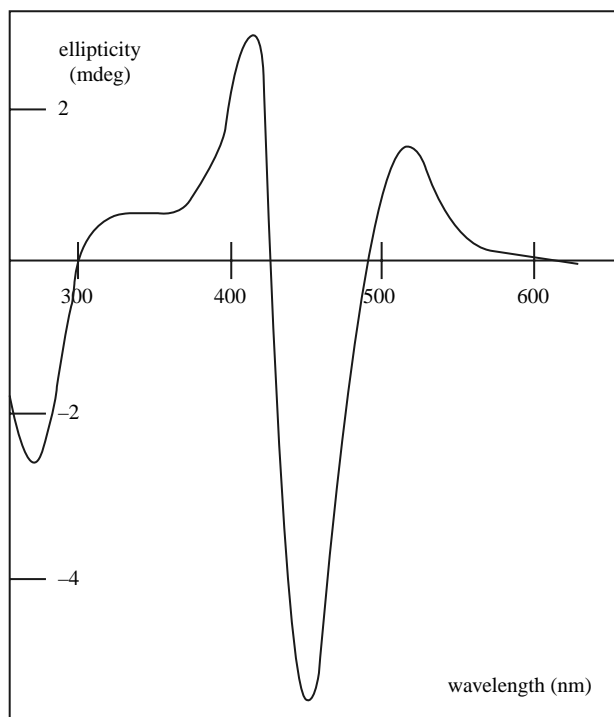


Fig. 7 Extrinsic (induced) CD spectrum for the complex formed between cholesterol and bilirubin in chloroform solution.

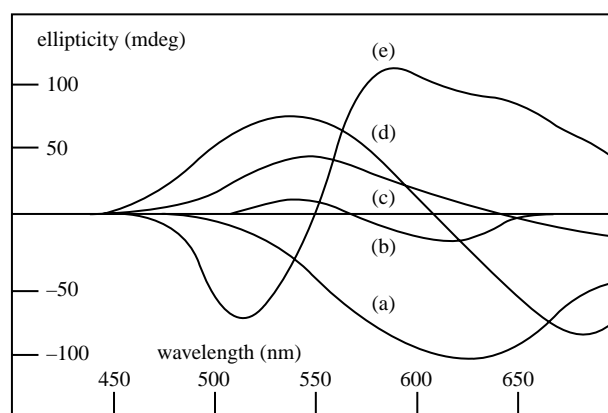


Fig. 8 CD spectra for (a) the chiral Cu(II)-L-tartrate metal complex and for mixed complexes of this host with equimolar amounts of (b) kanamycin, (c) amikacin, (d) gentamycin, and (e) streptomycin.

The choice of an irreversible color-induction reaction requires more ingenuity and greater care in execution. If the extended molecular unsaturation required to produce the color is too exhaustive, the chiral centers could be systematically eliminated, and must be avoided. Reaction conditions are much more unfavorable. Reagents are generally toxic and corrosive, and reaction conditions are anhydrous, e.g., the measurement of plasma cholesterol using a modified Chugaev reagent (63).

SUBSTRATES MADE CD-ACTIVE BY CHIRALITY INDUCTION

Although the heading implies that the analyte is achiral, it does not have to be. Greater analytical selectivity can be engendered by specific mutual influences of intrinsic and extrinsic chirality properties. Chirality induction reactions are generally reversible. Many are used extensively in chiral liquid chromatography (12). Developments presented here are the changes toward chiroptical detection and the applications of direct spectral measurements. Optimal conditions are achieved when chirality resides only on the derivatizing agent, and the chromophore is limited to the analyte. On reaction, the CD activity is exclusive to the molecular complex. Potential interferences from host and guest are inconsequential. The best hosts are linear and cyclic forms of oligo- and polysaccharides, such as the oligomaltoses and the cyclodextrins, which absorb only in the far-UV. Oligomaltoses are more soluble but less selective than are cyclodextrins. Virtually any material that has been used as a stationary phase in chiral chromatography is a candidate for direct homogeneous association reactions, and many others will ultimately appear, e.g., cryptands, vesicles, micelles, peptides, proteins, enzymes, antibodies, and nucleic acids.

Complexation is an equilibrium process. It is recommended that the host molecule be kept in large compositional excess over the analyte, thereby maximizing the mass-action effect and complexing as much of the analyte as possible. The larger the formation constant, the straighter the correlation line between the experimental ellipticity ΔA and the analyte concentration. The β - and γ -Cyclodextrins were used in exploratory investigations as analytical reagents (64) and to determine achiral forms of barbiturates (65), phenethylamines (64), benzodiazepin-2-ones (66), and phencyclidine and its analogs (5). Initial assays were done on prepared laboratory mixtures, but successful assays were also reported for secobarbital in secondal suppositories, meperidine in demerol dispensary

products, and diazepam and flurazepam in pharmaceutical preparations (66).

The CD spectra for chiral complexes, formed by chirality induction on organic dyes by oligomaltoses and oligocelluloses, were used to elucidate the rotational direction of the helical structures of the hosts in aqueous media (60). CD spectra of chiral derivatives formed by the association of aromatic residues, such as 9-Anthroate and p-Hydroxycinnamate with analogous oligosaccharides, were used to probe the local stereochemistry of the ring linkages and conformational arrangements of adjacent groups in acyclic polyols (26).

Protein hosts, whose absorbance and CD spectra are dominated by intense signals in the far-UV range (190–230 nm) are appropriate choices for introducing chirality into any molecule that absorbs at longer wavelengths, e.g., associative complexations of proteins, enzymes, and/or oligopeptides with warfarin (50); bilirubin and its analogs (26, 50, 58, 59); and a few dye molecules (63). Beside the generation of analytical data, structural modifications at the active sites can be monitored over two spectral ranges: the far-UV, where the active group is the peptide bond, and the near-UV, where the absorber is the guest molecule. In addition, CD ought to be seriously considered as an alternative detector for immunoassays because the experimental selectivity might overcome some of the limitations associated with polyclonal antibodies.

DETERMINATION OF ENANTIOMERIC EXCESS IN PARTIAL RACEMIC MIXTURES

For all types of chemical analysis, the quality of the results ultimately relates to the chemical purity of the best available SRM. For naturally chiral substances, there is the additional more serious concern over what constitutes absolute enantiomeric purity. Not even mass spectroscopy, which provides assurance that a substance is chemically pure, can be used to report absolute enantiomeric purities. To actually report an enantiomeric purity higher than 99% is truly beyond the capability of current analytical methodology (20). As noted previously, the fact is that results are measured relative to an enantiopurity defined to be 100%. Chemical purities aside, the measurement of enantiomeric purity and enantiomeric excess is technically the same, the difference being the extent of racemization. There are only two experimental options, either enantiomeric separations or multivariate spectroscopic analyses, that involve either two distinct detectors or multiple-wavelength detection for a single detector, as noted above.

The newly described derivatization reactions fulfill the second option.

If the chosen derivatization reaction is chirality induction, a simple two-step process is to measure the CD spectrum for the underivatized partial racemic mixture, followed by a measure of the net spectrum after the addition of the derivatizing agent to the mixture. Fundamentally, the reaction is the simple competitive instantaneous complexation of the enantiomers with, for example, β -Cyclodextrin, in which two diastereoisomers are formed. These might have different formation constants, or different induced spectra, or both. Regardless, the result is a change in the original CD spectrum. The two unknown enantiomeric concentrations are calculated by solving the simultaneous equations that describe the additivity of the two enantiomers in the case of the original solution and the two diastereoisomers in the case of the derivatized solution. The method was used with limited success in the measurement of enantiomeric distribution for prepared nonracemic mixtures of R- and S-Nicotine (26). The poor results could in part be attributed to the spectral changes being very small.

More accurate results were achieved by exchanging both the enantiomers in prepared nonracemic mixtures of ephedrine and pseudoephedrine with the coordinated L-Tartrate ligand of the Cu(II)-L-Tartrate host complex dissolved in 0.10 M aqueous base (62). The first CD spectrum is that for the parent complex and the second for the mixed ligand complexes, where only one of the two bonded L-Tartrate ligands on Cu(II) is exchanged by either the (+)- or the (–)-ephedrine or pseudoephedrine enantiomers. The total signal is the sum of three terms, the CD signal from the decreased concentration of the parent complex, plus the CD signal from the 1:1 complex of Cu(II) with one L-tartrate and the (+) enantiomer, plus the CD signal from the 1:1 complex of Cu(II) with one L-Tartrate and the (–) enantiomer. With careful control of the reaction conditions, the detection limit for the enantiomeric purity was on the order of $\pm 2\%$. In fact, changes in the CD signal on complexation was so large that almost equivalent results were more easily obtained using polarimetric measurements at only four wavelengths.

CD DETECTION/LIGAND EXCHANGE FOR ASSAYS OF PEPTIDES, OLIGOPEPTIDES, AND PROTEINS

With literally thousands of potential new drug substances in the combinatorial chemistry pipeline and the expanding emphasis on chiral drugs, the development of low-cost,

routine quality-control procedures is becoming a priority in pharmaceuticals and biotechnology. Ligand-exchange derivatization, coupled to CD detection, has the potential to do that for peptides, oligopeptides, and proteins.

Regulatory agencies have already set the standards for quality control (QC) for chiral drug substances. If it is a company's decision to market a chiral drug as a single enantiomer (the eutomer), the submission for regulatory approval must also include the equivalent chirality information for the other, nontherapeutic enantiomeric form (the distomer). An accurate determination of the enantiomeric purity of both forms is essential. Furthermore, chemical racemization will alter the eutomer to distomer ratio with time, diminishing the therapeutic property of the eutomer. Another very significant factor in QC, therefore, is to be able to accurately measure the rate of change of enantiomeric purity and decide when the meaningful therapeutic value of the eutomer has expired.

By replacing the L-tartrate ligand of the Cu(II)-derivatizing agent described above with D-histidine, a very selective host complex for ligand exchange was created (67–71). Much of the analytical selectivity accomplished by full spectrum visible-range CD detection is attributable to the specifics of the ligand–ligand interactions that ostensibly occur within the first coordination sphere of the complex. The extent of the selectivity that is accomplished for peptides and proteins is extraordinarily high (Fig. 9).

It was also determined that the CD spectral changes are extremely sensitive to changes in the amino acid sequence among residues that are far removed from the binding sites. A case in point is the individualized CD spectral changes for the exchange of D-histidine with human, human LysPro, porcine, and bovine insulin forms, all of which are 51 amino acid residue proteins

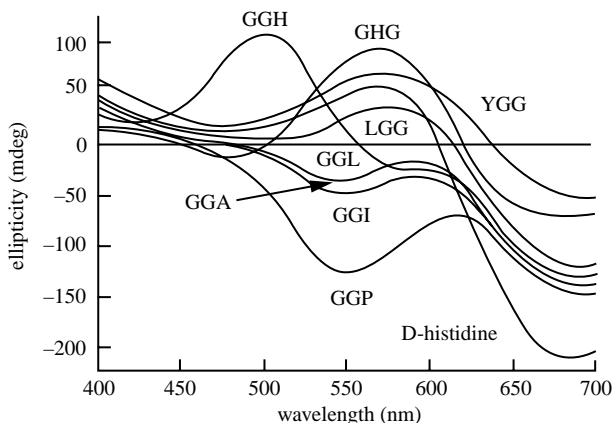


Fig. 9 CD spectra for the host Cu(II)-D-histidine spectrum and the mixed complexes with a series of tripeptides.

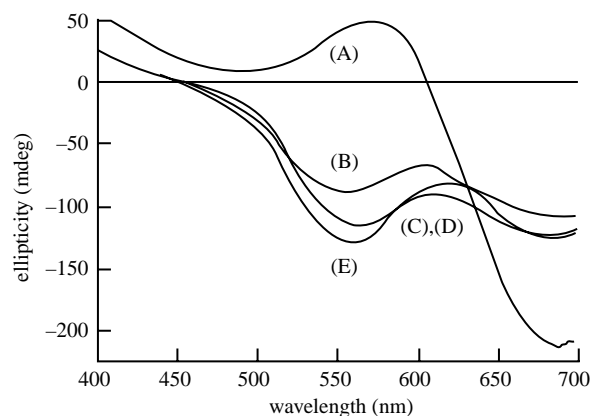


Fig. 10 CD spectra for (A) the Cu(II)-D-histidine host complex and for the mixed ligand complexes with (B) bovine insulin, (C) human insulin, (D) porcine insulin, and (E) human LysPro insulin.

(70). It is a known fact that proteins bind to Cu(II) ion in aqueous base by first substitution via the *N*-atom at the amine terminus (72). Chelation is completed by second and third substitutions into the metal first coordination sphere through the *N*-atoms of the first and second peptide functional groups. Because all four insulins have identical initial sequences, the observed CD spectral changes on ligand exchange are caused by some other structural variations. The only differences in the insulin residue sequences occur either at or adjacent to the acid terminus of the B-chain. Human insulin differs from human LysPro by the simple exchange of lysine and proline residues at positions B-28 and B-29. Porcine and human insulins differ by one residue, alanine for threonine, at position B-30 (70). The only explanation that accounts for this is that the metal ion is enclosed by an extensive, flexible, chiral, three-dimensional architecture and that subtle differences in the conformational properties of the enclosure bring different residues into the realm of the ligand to metal ion electronic transitions. With the D-Histidine kept in very large excess over the insulin analytes, there is a good quantitative linear correlation of the CD signal with concentration. What many researchers may have failed to realize in drug modeling is that when two chiral molecules interact, the stereochemical conformations of both change, and these changes may be just as significant in the mechanism of drug action as the absolute conformations of the participants themselves.

In other tests of the ligand exchange/CD detection assay procedure, it was demonstrated that 51 of a total of 53 di- and tripeptides could be uniquely identified by the character of the changes in the CD spectra for the mixed

Cu(II)-peptide-D-histidine complexes (67, 68, 70). A series of 19 neuropeptides was the focus of another investigation (69, 71). Neuropeptides are invariably chiral molecules and, because of the peptide bond, are CD-active. Once again, all but two of the analogs were individually distinguishable by the CD spectra for the mixed ligand complexes. For the two exceptions, ICI 174,864 and PLO17, the terminal amine group is substituted and, as such, it is not competitive with D-histidine in binding to the Cu(II) ion. In other words, substitution did not occur. To determine whether the CD spectra had characteristic properties that could be associated with the structure of first coordination spheres of the metal complexes, a Y-correlation matrix of the CD data was subjected to principal component analysis (PCA). All of the spectra can be accounted for mathematically by only four factors (71). The spectral data for each neuropeptide analyte can be graphically represented by a single point in an X-Y plot of the first two principal components, PC1 and PC2. When the points are subjected to a hierarchical clustering algorithm, the neuropeptides

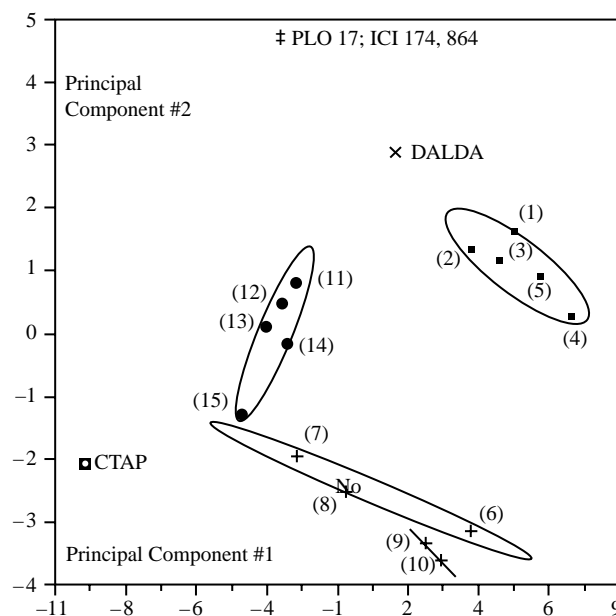


Fig. 11 Cluster plot of the first and second principal components derived by PCA from spectral data. The δ -receptor cluster covers (1) DTLET, (2) DSLET, (3) DADLE, (4) α^2 -Leu⁵-enkephalin, and (5) DPDPE. The μ -receptor cluster includes (6) DAGO, (7) Met⁵-enkephalin, and β -endorphin. The alternate δ -receptor cluster is composed of (9) Leu⁵-enkephalin and (10) Leu⁵-enkephalin amide. The κ -cluster of the dynorphins contains (11) B(1-13), (12) A(1-13); (13) A(1-9); (14) A(1-11); and (15) A(1-13) amide. No receptor preference was reported for PLO 17 or ICI 174 864.

are observed to aggregate according to their preferences for the δ , μ , or κ protein receptors (Fig. 11). This calibration model has the potential to become a prototypical-predictive in vitro model for correlating CD spectroscopic data with quantitative structure-activity relationships (QSAR) and, on further substantiation, may become a viable procedure for new drug forms (71). Despite the fact that the system is a mixed equilibrium reaction, in quantitative terms, a strong linear correlation exists between PC1 and analyte concentrations.

A recurring chiral property in many of the neuropeptides is the presence of at least one D-enantiomeric residue, e.g., natural and designer enkephalins such as DALDA, DAGO, and DPDPE. Locating the position of the D-form in a sequence is a challenging endeavor that is being systematically studied on two series of model penta- and hexapeptides. The role of D-enantiomeric forms in biotechnology drug substances is a very real interest and more especially in view of the recent recognition of the existence of D-Serine, which functions as neurotransmitter in mammalian brain tissue (73). The D-Enantiomer is synthesized in the brain from the natural L-Form catalyzed by serine racemase.

SUMMARY

The intent of this article was to demonstrate, based on a wealth of relatively new experimental data, that there is sufficient analytical selectivity and sensitivity to accept polarimetry and CD as viable and easy-to-use analytical detection methods. In contrast to other detectors, they provide the capability of making direct analytical assays after a sample work-up that is a simple solvent extraction and of measuring enantiomeric purities in the ranges specified by the FDA for the process and quality control for new chiral drug substances. In the future, broader applications will be developed, especially as the current analytical emphasis turns toward nucleic acid protein interactions and the CD properties of intact single cells.

REFERENCES

- Lowry, T.M. *Optical Rotatory Power*; Dover Publications: New York, 1935.
- Djerrasi, C. *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry*; McGraw-Hill: New York, 1960.
- Charney, E. *The Molecular Basis of Optical Activity. Optical Rotatory Dispersion and Circular Dichroism*; Wiley: New York, 1979.
- Schnatzke, G. *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry*; Heyden: London, 1967.
- Crabbe, P. *ORD and CD in Chemistry and Biochemistry. An Introduction*; Academic Press: New York, 1972.
- Harada, N.; Nakanishi, K. *Circular Dichroic Spectroscopy: Exciton Coupling in Organic Stereochemistry*; University Science Books: Mill Valley, CA, 1983.
- Cotton, A. *Compt. Rend.* **1895**, 120, 989.
- Bobbitt, D.R. Analytical Applications of Circular Dichroism. *Technical Instrumentation in Analytical Chemistry*; Purdie, N., Brittain, H.G., Eds.; Elsevier: Amsterdam, 1994; 14, Chap. 2.
- Sutherland, J.C. *Circular Dichroism and the Conformational Analysis of Biomolecules*; Fasman, G.D., Ed.; Plenum Press: New York, 1996.
- Synovec, R.E.; Yeung, E.S. Fluorescence Detected Circular Dichroism as a Detection Principle in High-Performance Liquid Chromatography. *J. Chromatogr.* **1986**, 368, 85–93.
- Lloyd, D.K.; Goodall, D.M.; Scrivener, H. Diode-Laser-Based Optical Rotation Detector for High-Performance Liquid Chromatography and On-Line Polarimetric Analysis. *Anal. Chem.* **1989**, 61, 1238–1243.
- Christensen, P.L.; Yeung, E.S. Fluorescence-Detected Circular Dichroism for On-Column Detection in Capillary Electrophoresis. *Anal. Chem.* **1989**, 61, 1344–1347.
- Yeung, E.S.; Synovec, R.E. Detectors for Liquid Chromatography. *Anal. Chem.* **1986**, 58, 1237–1256.
- Takakuwa, T.; Kurosu, Y.; Sakayanagi, N.; Kaneuchi, F.; Takeuchi, N.; Wada, A.; Senda, M. Direct Combination of a High Performance Liquid Chromatograph and a Circular Dichroism Spectrometer for Separation and Structural Analysis of Proteins. *J. Liq. Chromatogr.* **1987**, 10, 2759–2769.
- Westwood, S.A.; Games, D.E.; Sheen, L. Use of Circular Dichroism as a High-Performance Liquid Chromatography Detector. *Chromatography* **1981**, 204, 103–107.
- Mannschreck, A. On-Line Measurement of Circular Dichroism Spectra During Enantioselective Liquid Chromatography. *Trends. Anal. Chem.* **1993**, 12, 220–225.
- Boehme, W.; Wagner, G.; Oehme, U. Spectrophotometric and Polarimetric Detectors in Liquid Chromatography for the Determination of Enantiomer Ratios in Complex Mixtures. *Anal. Chem.* **1982**, 54, 709–711.
- Meinard, C.; Bruneau, P.; Perronnet, J. High-Performance Liquid Chromatograph Coupled with Two Detectors: A UV Spectrometer and a Polarimeter. *J. Chromatogr.* **1985**, 349, 109–116.
- Ward, T.J.; Armstrong, D.W. *Chromatographic Science Series Chromatographic Chiral Separations*; Zief, M., Crane, L.J., Eds.; Marcel Dekker, Inc.: New York, 1988; 40.
- Zukowski, J.; Tang, Y.; Berthod, A.; Armstrong, D.W. Investigation of a Circular Dichroism Spectrophotometer as a Liquid Chromatography Detector for Enantiomers: Sensitivity, Advantages, and Limitations. *Anal. Chim. Acta.* **1992**, 258, 83–92.

21. Kudo, K.; Ajima, K.; Sakamoto, M.; Saito, M.; Morris, S.; Castiglioni, E. New Circular Dichroism (CD) Based Detection for Monitoring Chiral Compounds in HPLC. *Chromatography* **1999**, *20*, 59–64.
22. Horvath, P.; Gergely, A.; Noszal, B. Determination of Enantiomeric Purity by Simultaneous Dual Circular Dichroism and Ultraviolet Spectroscopy. *Talanta* **1997**, *44*, 1479–1485.
23. Lodevico, R.G.; Bobbitt, D.R.; Edkins, T.J. Determination of Enantiomeric Excess Using a Chiral Selective Separation Mode and Polarimetric Detection. *Talanta* **1997**, *44*, 1353–1363.
24. Purdie, N.; Swallows, K.A. Analytical Applications of Polarimetry, ORD, and CD. *Anal. Chem.* **1989**, *61*, 77–86.
25. Purdie, N.; Swallows, K.A.; Murphy, L.B.; Purdie, R.B. Circular Dichroism. II. Analytical Applications. *Trends Anal. Chem.* **1990**, *9*, 136–142.
26. Purdie, N. Analytical Applications of Circular Dichroism. *Technical Instrumentation in Analytical Chemistry*; Purdie, N., Brittain, H.G., Eds.; Elsevier: Amsterdam, 1994; 14, Chap. 8.
27. Martens, H.; Naes, T. *Multivariate Calibration*; Wiley: New York, 1989.
28. Crone, T.A.; Purdie, N. Circular Dichroism Spectra of Opium Alkaloids in Aqueous Media. *Anal. Chem.* **1981**, *53*, 17–21.
29. Bowen, J.M.; Crone, T.A.; Hermann, A.O.; Purdie, N. Circular Dichroism Spectra of Opium Alkaloids in a Cholesteric Liquid Crystalline Solvent System. *Anal. Chem.* **1980**, *52*, 2436–2440.
30. Bowen, J.M.; Purdie, N. Circular Dichroism Spectra of Opium Alkaloids in the Solid State. *Anal. Chem.* **1980**, *52*, 573–575.
31. Bowen, J.M.; Crone, T.A.; Kennedy, R.K.; Purdie, N. Determination of Heroin by Circular Dichroism. *Anal. Chem.* **1982**, *54*, 66–68.
32. Han, S.M.; Purdie, N. Simultaneous Determination of Opiates by Circular Dichroism. *Anal. Chem.* **1986**, *58*, 113–116.
33. Potapov, V.M.; Demianovich, V.M.; Terentiev, A.P. Spectropolarimetric Analysis. IV. Polarimetric Analysis of Ephedrine and Pseudoephedrine in Their Mixtures. *Zh. Anal. Khim.* **1964**, *19*, 254–257.
34. Bowen, J.M.; Crone, T.A.; Head, V.L.; McMorrow, H.A.; Kennedy, R.K.; Purdie, N. Circular Dichroism: An Alternate Method for Drug Analysis. *J. Forensic. Sci.* **1981**, *26*, 664–670.
35. Mitscher, L.A. *The Chemistry of the Tetracycline Antibiotics*; Marcel Dekker, Inc.: New York, 1978; 13.
36. Purdie, N.; Swallows, K.A. Direct Determination of Beta-Lactam Antibiotics by Circular Dichroism. *Anal. Chem.* **1987**, *59*, 1349–1351.
37. Engle, A.R., Ph.D. Dissertation. Oklahoma State University: Stillwater, OK, 1993.
38. Han, S.M.; Purdie, N. Determination of Stereoisomers by Circular Dichroism. *Anal. Chem.* **1986**, *58*, 455–458.
39. Purdie, N. Determination of Drugs of Abuse and their Stereoisomers by Circular Dichroism. *Forensic. Sci. Rev.* **1991**, *3*, 1–16.
40. Purdie, N.; Swallows, K.A. Direct Determination of Water Soluble Vitamins by Circular Dichroism. *J. Agr. Food Sci.* **1991**, *39*, 2171–2175.
41. Potapov, V.M.; Demianovich, V.M.; Rakovska, R.S. Analytical Applications of Spectropolarimetry. *Zh. Anal. Khim.* **1984**, *39*, 983–996.
42. Gergely, A. A Review of the Application of Chiroptical Methods to Analytical Chemistry. *J. Pharm. Biomed. Anal.* **1989**, *7*, 523–541.
43. Johnson, W.C., Jr. The Circular Dichroism of Carbohydrates. *Adv. Carbohydr. Chem. Biochem.* **1987**, *45*, 73–124.
44. Hayward, L.D.; Angyal, S.J. A Symmetry Rule for the Circular Dichroism of Reducing Sugars, and the Proportion of Carbonyl Forms in Aqueous Solutions Thereof. *Carbohydr. Res.* **1977**, *53*, 13–20.
45. Domard, A. pH and CD Measurements on a Fully Deacetylated Chitosan. *Int. J. Biol. Macromol.* **1987**, *9*, 333.
46. Kuo, J.C.; Yeung, E.S. Determination of Carbohydrates in Urine by High-Performance Liquid Chromatography and Optical Activity Detection. *J. Chromatogr.* **1981**, *223*, 321–329.
47. Kimura, A.; Chiba, S.; Yoneyama, M. A Simple and Rapid Determination of Ketoses by Circular Dichroism. *Carbohydr. Res.* **1988**, *175*, 17–23.
48. Fasman, G.D. *Circular Dichroism and the Conformational Analysis of Biomolecules*; Plenum Press: New York, 1996.
49. Jirgensons, B. *Optical Activity of Proteins and Other Macromolecules*, 2nd Ed.; Springer Verlag: New York, 1973.
50. Thomas, M.P.; Patonay, G.; Warner, I.M. Fluorescence-Detected Circular Dichroism Studies of Serum Albumins. *Anal. Biochem.* **1987**, *164*, 466–473.
51. Souci, S.W.; Fachmann, W.; Kraut, H. *Food Composition and Nutrition Tables*, 2nd Ed.; Wissenschaftliche Verlagsgesellschaft mbH: Stuttgart, 1981.
52. Han, S.M.; Purdie, N. Determination of Cannabinoids by Circular Dichroism. *Anal. Chem.* **1985**, *57*, 1068–1072.
53. Atkinson, W.M.; Han, S.M.; Purdie, N. Determination of Nicotine in Tobacco by CD Spectropolarimetry. *Anal. Chem.* **1984**, *56*, 1947–1950.
54. Purdie, N.; Swallows, K.A. Determination of Reserpine by Circular Dichroism. *Pharm. Res.* **1989**, *6*, 521–524.
55. Bowen, J.M.; McMorrow, H.A.; Purdie, N. Determination of LSD in Drug Confiscates by Circular Dichroism Spectropolarimetry. *J. Forensic. Sci.* **1982**, *27*, 822–826.
56. Bowen, J.M.; Purdie, N. Identification of Cocaine and Phencyclidine by Solute-Induced Circular Dichroism. *Anal. Chem.* **1981**, *53*, 2239–2242.
57. Atkinson, W.M.; Bowen, J.M.; Purdie, N. Determination of Brompton's Cocktails by CD. *J. Pharm. Sci.* **1984**, *173*, 1827–1828.
58. Blauer, G. Complexes of Bilirubin with Proteins. *Biochim. Biophys. Acta.* **1986**, *884*, 602–605.

59. Blauer, G. Optical Activity of Bile Pigments. *Israel J. Chem.* **1983**, 23, 201–209.
60. Engle, A.R.; Hyatt, J.A.; Purdie, N. Induced Circular Dichroism Study of the Aqueous Solution Complexation of Cello-Oligosaccharides and Related Polysaccharides with Aromatic Dyes. *Carbohydr. Res.* **1994**, 265, 181–195.
61. Engle, A.R.; Lucas, E.A.; Purdie, N. Determination of Enantiomers of Ephedrine Mixtures by Polarimetry. *J. Pharm. Sci.* **1994**, 83, 1310–1314.
62. Engle, A.R.; Purdie, N. Determination of Enantiomeric Purities Using CD/CD Detection. *Anal. Chim. Acta.* **1994**, 298, 175–182.
63. Purdie, N.; Murphy, L.H.; Purdie, R.B. Serum Cholesterol: A Direct Measure of the Low Density Fractions. *Anal. Chem.* **1991**, 63, 2947–2951.
64. Han, S.M.; Atkinson, W.M.; Purdie, N. Solute Induced CD: Drug Discrimination by Cyclodextrin. *Anal. Chem.* **1984**, 56, 2827–2830.
65. Han, S.M.; Purdie, N. Cyclodextrin Complexation of Barbiturates in Aqueous Solution. *Anal. Chem.* **1984**, 56, 2825–2827.
66. Han, S.M.; Purdie, N.; Swallows, K.A. Determination of the Benzodiazepin-2-Ones by Circular Dichroism. *Anal. Chim. Acta* **1987**, 197, 57–64.
67. Purdie, N.; Province, D.W. Algorithms for the Quantitative Validation of Chiral Properties of Peptides. *Chirality* **1999**, 11, 546–553.
68. Purdie, N.; Province, D.W.; Johnson, E.A. Tripeptide Discriminations Using Circular Dichroism Detection. *J. Pharm. Sci.* **1999**, 88, 715–721.
69. Purdie, N.; Province, D.W.; Johnson, E.A. A Convenient Assay Method for the Quality Control of Peptides and Proteins. *J. Pharm. Sci.* **1999**, 88, 1242–1248.
70. Purdie, N.; Province, D.W.; Layloff, T.P.; Nasr, M.M. Algorithms for Validating Chiral Properties of Insulins. *Anal. Chem.* **1999**, 7, 3341–3346.
71. Purdie, N.; Province, D.W. Clustering of CD Spectral Data as a Prototype QSAR Model for Neuropeptides. *J. Pharm. Sci.* **1999**, 88, 1249–1253.
72. Sigel, H.; Martin, R.B. Coordinating Properties of the Amide Bond. Stability and Structure of Metal Complexes of Peptides and Related Ligands. *Chem. Rev.* **1982**, 82, 384–426.
73. Snyder, S.H.; Wolosker, H.; Blackshaw, S. Serine Racemase: A Glial Enzyme Synthesizing D-Serine to Regulate Glutamate-*N*-methyl-D-Aspartate Neuro-Transmission. *Proc. Natl. Acad. Sci. USA* **1999**, 96, 13409–13414.

CARCINOGENICITY TESTING: PAST, PRESENT, AND FUTURE

Koen Van Deun

Janssen Pharmaceutica N.V., Beerse, Belgium

INTRODUCTION

This chapter presents an historical overview of cancer, carcinogenicity testing, and human cancer causes. Cancer has been known for a very long time, but the awareness of human carcinogenicity caused by chemicals is a phenomenon of the 20th century. This in turn has produced legislation that prohibits the use of carcinogens in the food chain and has provided guidelines for carcinogenicity testing in animals. Lifetime studies (18–24 months) in two main rodent species (rat and mouse), also known as the “Standard Chronic Bioassay,” have been conducted since the 1960s.

Meanwhile, various deficiencies have been detected in the Chronic Bioassay; over-sensitivity is the major one. Hundreds of compounds have been tested with the Chronic Bioassay method, and about 50% have yielded (false) positive results. Lack of relevance to man has often been demonstrated by additional mechanistic studies. Additionally, more mechanistic and molecular knowledge has been gained in regards to the human carcinogenicity concept, including genotoxic versus epigenetic carcinogens, the multi-stage cancer theory, and human life style factors involved in carcinogenesis.

The above evolutions have opened up new opportunities for carcinogenicity testing, including short-term alternative carcinogenicity models. In addition, carcinogenicity testing is evolving from a standard chronic bioassay to a weight-of-evidence approach, where the mechanisms involved in rodent and human carcinogenesis are considered, and where communication between industry and regulatory authorities is encouraged.

CARCINOGENICITY TESTING IN THE PAST

History of Cancer, Carcinogenicity Testing, and Human Causes of Cancer

Cancer has probably existed as long as multicellular organisms have. Paleontologists have shown the presence of tumors in the bones of dinosaurs, long before the advent of *Homo sapiens*. The Ancient Egyptians knew of the

existence of cancer in man. A papyrus with a hieroglyph that shows a clinical tumor was discovered. Autopsies of mummies have shown bone tumors in Egyptians. The first real descriptions of various tumors were those presented by Hippocrates in the 4th century B.C. He used the term “*carcinoma*,” by which he meant a tumor that spreads and kills the patient. In the 1st century B.C., Galen made the distinction between “*natural tumors*,” such as those present during the development of breasts in female adolescence and “*tumors, which go beyond the bounds of nature*,” such as those related to bone healing following a fracture. Finally, he described “*unnatural tumors*,” which today can be defined as a “neoplastic” growth of tissue. The views of Hippocrates and Galen dominated medicine for many centuries. In the 19th century, cancer knowledge expanded enormously, primarily through intensified studies on anatomy and histology. Bichat, and later Müller, confirmed Galen’s theory. They postulated the hypothesis that cancer arises from accidental tissue formation. This in turn prompted Pasteur to propose his theory of “*omnis cellula et cellula*.” Every cell originates from another cell.

The discovery that exposure to exogenous chemicals could lead to cancer in humans was first made in the late 18th century, when Percival Pott demonstrated the relationship between cancer of the scrotum and the occupation of chimney sweepers exposed to coal tar/soot. Other examples noted later were scrotal cancers in cotton spinners exposed to unrefined mineral oils, and cancers of the urinary bladder in men who worked in textile dye and rubber industries due to their exposure to certain aromatic amines used as antioxidants. Experimental induction of cancer by chemicals was first reported in detail by Yamagiwa and Ichikawa in 1918, when repeated application of coal tar to the ear of rabbits resulted in skin carcinomas. Over the next few years, Kennaway and Leitch confirmed this finding and demonstrated similar effects in mice and rabbits from the application of soot extracts, other types of tar (e.g., acetylene or isoprene), and some heated mineral oils. These researchers also observed skin “irritation” sometimes accompanied by ulcers at the site of application of the test material. “Irritation” was thought to be an important factor in skin tumor development. However, not all irritants (e.g., acridine) induced skin cancer in mice and conversely, some purified

chemicals isolated from these crude materials produced a high incidence of skin tumors with little or no irritation. These observations suggested that intracellular changes induced by carcinogens might lead to cancer, independent of any lesions observable visually or by light microscopy. As a result, the hypothesis that cancer stems from the interaction of a chemical with a “sensitive” site in the cell was proposed. Further reading on the early history and concepts in cancer is provided by Van Cauteren et al. (1), Roe (2), and Grasso et al. (3).

By the mid 20th century, causes of human cancer related to exposure of chemical substances was of growing concern. Until the 1950s, the known causes of cancer were those associated with chemical exposure at the work place. These findings led to the concept that most human cancers were caused by environmental chemicals. For example, Hueper noted that the increasing rate of lung cancer coincided with the remarkable growth of industry and motorized transportation, as well as the use of industrial carcinogenic products and their release as effluents and exhausts into the occupational and general atmosphere. However, Wynder and Doll later showed that a major type of cancer, lung cancer, which was increasing sharply in men, was due to the chronic use of cigarettes and exposure to tobacco smoke.

One of the first indications that factors other than chemical interaction with specific intracellular sites could be responsible for tumor induction was the discovery of fibrosarcomas, observed around large bakelite discs implanted subcutaneously in rats. A similar observation was seen after wrapping pieces of cellophane film around the kidneys in rats. After several months, a large tumor developed around the cellophane film in some of the rats. The finding of “solid-state” carcinogenesis was confirmed later in the 1960s by implantation of various solids (including noble metals such as gold, silver, or platinum), which led to the development of sarcomas. The shape, and particularly the size of the implant, had a determining influence on sarcoma induction. In addition, iron-dextran was found to produce this type of tumor following repeated subcutaneous injections in rats and mice. Subsequent investigations revealed that some food colorings evoked local sarcomas in rats and mice after repeated subcutaneous administration. The principal mechanism believed to be involved in sarcoma induction was the continued activation of reparative processes in connective tissues by the physical, tissue-damaging properties. Further examples of chemical and solid-state carcinogenesis that lead to tumor development in rodents are provided by Grasso et al. (3) and Weissburger (4). Other recent evolutions in carcinogenicity concepts will be described further in this review.

History of Carcinogenicity Testing Legislation

One area of great contemporary concern was the entry of carcinogens into the human environment through food in the form of food additives and food contaminants, such as pesticides. Thus, hearings were held before the U.S. Congress relative to desirable modifications of the pure Food and Drug Laws that had been part of legislation since 1908. In 1958 and 1960, Congressman Delaney introduced the following two pieces of legislation, also known as the “Delaney Clause:”

1. Food Additives Amendment of 1958: Federal Food, Drug and Cosmetic Act. Section 409(c) (3) (A): “. . . no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal, or if it is found, after tests which are appropriate for the evaluation of the safety of food additives, to induce cancer in man or animal. . . .”
2. Color Additive Amendment of 1960; Federal Food, Drug and Cosmetic Act. Section 706 (b) (5) (B): “A color additive that will result in ingestion shall be deemed unsafe, and shall not be listed, . . . if, after tests which are appropriate for the evaluation of the safety of additives for such use or after other relevant exposure of man or animal to such additive; it is found by the Secretary to induce cancer in man or animal.”

This legislation consisted of the absolute prohibition of entry of “carcinogens” into the food chain, and established for the first time the view that any substance that could be considered carcinogenic was not to be added to foods or cosmetics. There was no room for quantification and calculation of risk. More detailed information on this clause can be found in the article by Weissburger (4).

From 1961 onward, the “Bioassay Program” of the National Cancer Institute (NCI) was developed. This program was based upon the need for a more systematic investigation of chemical carcinogenesis in animals (rats and mice). Each test also involved the use of appropriate negative and positive control carcinogens to verify that the models actually gave the appropriate results. Preliminary toxicology testing was performed using five rats or mice per group, and applying the dosage for 6 weeks, with a further holding period of 2 weeks without chemical administration to detect delayed toxicity. Animals that survived provided information on survival and weight gain (loss), the highest possible dose, and the maximum tolerated dose (MTD). The final NCI Chronic Bioassay consisted of a life-time (18–24 month) oral administration to rats and mice up to the MTD, with a second group at 50 or 30% of the selected MTD. The principle of the MTD was chosen since a number of bioassays in the literature

with the known human carcinogens gave a good incidence of cancer only at high dose levels. Between the 1960s and the 1980s, the NCI Bioassay Program tested many types of chemicals that humans were potentially exposed to, including chemicals used in cancer chemotherapy, hair dye formulations, and a series of agricultural chemicals, including pesticides, insecticides, and fungicides.

The NCI Chronic Bioassay became the basis for current guidelines on carcinogenicity testing. Protocols and reports for the large-scale bioassays were standardized in order to increase scientific acceptability.

A number of the chemicals administered were found to induce cancer in rats and mice. Some chemicals, particularly agricultural products, were negative in the rat models but induced liver tumors in mice. With either type of response, a chemical that induced cancer in any test was labeled a carcinogen, and by extrapolation was considered to be a human cancer risk without questioning its relevance or the underlying mechanism.

Towards the end of the 1980s, the NCI carcinogenesis branch merged with other agencies that shared similar efforts to form the National Toxicology Program (NTP). The NTP placed much more emphasis on mutagenicity and other short-term tests, teratology, behavioral effects, neurotoxicity, and various subchronic studies.

More information on the history of the Chronic Bioassay can be found in the reviews by Weissburger and Weissburger (5, 6).

Carcinogenicity Testing Guidelines

In the 1970s and 1980s, the US, European, and Japanese Registration Authorities established guidelines for carcinogenicity testing in animals for the various chemicals

characterized by possible long-term intake by man. These chemicals included food and color additives, agrochemicals, industrial chemicals, solvents, human pharmaceuticals, and veterinary products. The guidelines were based upon the Chronic Bioassay of the NTP and gave indications for route and frequency of dosing, dose levels, group sizes, duration of the study, and observations during the study. A summary of the various guidelines is provided in Table 1. More detailed information on the various guidelines is provided by an overview of Inveresk Research International (7). Meanwhile newer versions of most of the guidelines have become effective.

CURRENT CARCINOGENICITY TESTING

A description of the standard approaches in carcinogenicity testing for the safety of chemicals is provided.

Standard Rodent Chronic Bioassay

Species and strains

According to the previously described guidelines, carcinogenicity studies have to be performed in two rodent species, usually the rat and the mouse. Ideally, the strains should have a low spontaneous incidence of cancer, but they should also be sensitive to induction of cancer by human carcinogens. Commonly used species are the Sprague Dawley, Fisher F344, or Wistar strains in rats, and the CD-1 or C57BL-based strains in mice. At least 50 animals are included per sex in each dose group. Today, at least three dose groups are used as well as at least one

Table 1 Establishment of guidelines for carcinogenicity testing in animals

Continent	Scope	Authority, year of approval
General	Chemicals	OECD, 1981
US	Food and color additives	FDA, 1982
	Agrochemicals	EPA–FIFRA, 1984
	Chemicals	EPA–TSCA, 1985
EC	Pharmaceuticals	EC directives, 1983
	Chemicals	EC directives, 1988
Japan	Agrochemicals	MAFF, 1985
	Veterinary products	MAFF, 1988
	Pharmaceuticals	MHW, 1989

EPA: Environmental Protection Agency; FDA: Food and Drug Administration; FIFRA: Federal Insecticide Fungicide and Rodenticide Act; MAFF: Ministry Agricultural Forestry and Fisheries; MHW: Ministry for Health and Welfare; OECD: Organisation for Economic Co-Operation and Development; TSCA: Toxic Substances Control Act.

negative control group, which results in minimally 200 males and 200 females for one study.

Doses and route of administration

The animals are exposed daily to the test compound from the age of 6 weeks onward. The administration procedure should simulate human exposure as closely as possible; oral intake is the most common route.

In the past, the test item was often mixed in the drinking water or the feed, either in a fixed concentration in the feed during the entire study or with regular adaptations to maintain a steady ratio of mg/kg of body weight intake during the entire life expectancy. Nowadays, oral gavage administration in the stomach is used, except for agrochemicals and food additives, where feed administration is still applicable. Oral gavage administration provides more certainty of test item intake, but also leads to another pattern of test item exposure in the body (peak concentrations after dosing). Dose-selection for the various dose groups has been based mainly upon a MTD, which is defined to elicit slight target organ toxicity but will not shorten the treated animals' survivability from any toxic effects other than the induction of neoplasms. For the most part, a body weight gain loss of 10% is considered acceptable as evidence of minimal toxicity. The medium dose may elicit minimal toxicity; however, the low dose should be free of any toxicity.

The MTD is mostly defined in a 3-month dose-range-finding study, allowing optimal dose setting for the pivotal chronic bioassay.

Duration

The studies are designed to last for at least 24 months and survival should be at least 25 animals per sex in the control and low dose groups, both in males and females. In the past, the studies were often extended beyond 24 months because survival (especially in the control and low dose group) was still above 25 animals/group/sex, and because of concern that the carcinogenic effect might become visible only at a later end point. Currently, most of the studies are not extended since geriatric pathology increases, which can complicate and obscure the assessment of carcinogenicity.

Experimental conditions

Experimental conditions are of utmost importance because they can influence the results of the study. Factors such as hygiene, temperature, relative humidity, number of air replacements, and light have to be maintained and monitored consistently during the study.

Today, carcinogenicity studies are performed under Specific Pathogen Free (SPF) conditions. This means

that SPF animals are obtained from commercial breeders and are placed in SPF rooms after arrival in the experimental unit. Other factors taken into account are quarantine, health monitoring, and hygienic measures during handling, such as sterile gloves and mouth masks. Good Laboratory Practice (GLP) has also contributed to improved test conditions. GLP not only applies to the way animals are handled, but also to appropriate documentation and recording of all actions during a study. This leads to better traceability, reconstruction, and interpretation of study data and results. All the improvements in experimental conditions have led to increased survivability in the animals. In the past, various deaths occurred due to respiratory or other infectious diseases; these are almost totally excluded within the current improved health condition of the animals.

Parameters examined in the study

During the 24-month study, various study parameters are examined as shown in Table 2. The daily follow-up is of utmost importance in order to pick up unexpected findings. If problems arise, the study director's, veterinarian's, or pathologist's attention is drawn, and immediate and appropriate actions are requested. After necropsy during or at the end of the study, a mean list of 30 tissues is sampled and examined macroscopically. This may lead to a total number of 12,000 or more tissue samples for a single carcinogenicity study. All tissues are fixed and processed for further microscopic examination for neoplastic and nonneoplastic changes. These examinations are done by pathologists specialized in rodent pathology. The final aim is to detect the number of animals with tumors, but also multiplicity of tumors and whether the tumors caused death of the animal.

Histopathological evaluation

As indicated in Table 2, histopathological examination is performed on all animals to detect "nonneoplastic" and especially "neoplastic" changes induced by the test compound. Nonneoplastic changes may include inflammatory, degenerative, or other changes in various tissues, either caused by the test item or by geriatric pathology. Neoplastic changes, or tumors, can be divided by "benign" and "malignant" neoplasms. Benign neoplasms are well defined, often encapsulated, noninvasive, and well differentiated. They grow relative slowly, display relative few mitoses, and are not metastatic. Malignant tumors are less well defined and usually not well encapsulated. They are invasive and relatively undifferentiated; they grow rapidly, display abundant mitosis, and finally undergo metastasis).

Table 2 Parameters and frequency of observation or measurement in carcinogenicity studies

Frequency	Parameter	Examples of intended endpoint(s)
Daily	Mortality	Week of death or sacrifice Death or sacrifice status: Natural death or sacrificed in bad condition Terminal sacrifice
	Clinical observations	Drug-related signs of toxicity Bad condition, preceding mortality
Weekly	Body weight & weight gain	Drug-related body weight and weight gain or loss (10–15% body weight gain loss is accepted at MTD)
	Food consumption	Drug-related changes in food intake
After 6 or 12 months	Toxicokinetics	Exposure of parent compound and metabolites and relation to human exposure
After 6, 12, 18 months and terminally	Hematology	Drug-related haematological deviations
	Clinical biochemistry	Drug-related biochemical deviations
Terminally (also on all animals dying or sacrificed during the study except organ weights)	Gross pathology	Drug-related macroscopic findings
	Organ weights	Drug-related organ weight changes
	Histopathology	Nonneoplastic changes (including “pre-neoplastic” changes, such as hyperplasia) Neoplastic changes (tumors) Malignancy status (malignant or benign) Cause of death (tumors that caused death = probably fatal and fatal tumors; tumors that did not cause death = probably incidental and incidental tumors, and tumors in terminally sacrificed animals) Multiplicity of tumors (number of tumors/animal and/tissue)

Preceding conditions to neoplasia, such as “hyperplasia,” (which refers to an increase in the number of cells) and “hypertrophy,” (which refers to an increase in cell dimensions), can also be found.

Histopathological examination is a very intensive job; months may go by before all tissues are evaluated.

In addition, the histopathologist will relate the histopathological findings with the other nonmorphological findings during the study, such as hematological or biochemical results, either for individual animals or for dosed groups. As such, diagnoses and mechanisms for the findings are established. For the most part, independent pathologists

perform peer reviews in order to have confirmation on the exact diagnosis or mechanism.

Historical control data

Historical control data can be used to interpret the changes seen in carcinogenicity data. These control data can apply to the various parameters studied, such as hematology, biochemistry, and the incidences of tumors. They may be used when differences are seen between the incidences of tumors in the dosed groups and spontaneous incidences in concurrent control animals, where coincidence is suspected, or for tumors with very sporadic incidences.

Spontaneous incidences in tumors are commonly seen in untreated rats and mice, and vary from strain to strain. Examples include pituitary and mammary tumors in rats, and liver and lung tumors in mice. The incidences of tumors can vary, and even today, there is no clear understanding of their etiology, except for ad-libitum feeding. On the other hand, caloric restriction could retard aging (associated with a reduction in the rate of cell replication), and reduce the incidence of degenerative diseases and tumor incidences (2).

Statistical analysis

Statistical analysis is performed on all parameters in the study. Its most fundamental objective is to determine whether administration of the test agent results in an increase in tumor incidence rates as compared to those in unexposed controls. Various statistical methods can be used. Tests for increased tumor occurrence rates between dosages may be based on "pair-wise comparisons," such as the Chi-square test for 2×2 tables, the Fisher's exact test, or the Cochran-Armitage test. These tests are most appropriate when survival rates do not differ appreciably in the various dose groups.

If the treatment results in reduced survival, early mortality in the high-dose groups may preclude the development of tumors and other statistical methods are required. Peto proposed a test for differences in tumor occurrence rates due to treatment, taking into account differences in survival and the times at which tumors were observed. This procedure requires information on the cause of death of each animal, and is based on a time-stratified contingency table analysis of the prevalence of incidental tumors that did not kill their host and a similar analysis of fatal tumors that resulted in death prior to the study. These two analyses are then combined to arrive at an overall test for increase in trend in tumor occurrence rates allowing for differential survival rates among the treatment groups. More information on study design and statistical analysis is

provided by Feron et al. (8), Portier (9), Gart et al. (10), and Ciminera and Allen (11).

Cancer Risk Assessment

Once carcinogenicity testing has been performed, carcinogenicity risk assessment must be performed. Regulatory agencies have the responsibility to identify and assess compounds that are administered in food, provided as pharmaceuticals, or have the potential to be released in the environment at levels that warrant concern. Various topics have to be addressed when characterizing the carcinogenic risk. These include hazard identification (i.e., the likelihood to be a human carcinogen), dose-response, and extent of human exposure. Each of these assessments involve the use of many assumptions and estimations, the magnitude of which may be decreased by the incorporation of more information (e.g., mechanistic studies, pharmacokinetic data, and improved low dose extrapolation models). Velasquez et al. (12) have provided an overview of cancer risk assessment by the US Environmental Protection Agency (EPA) and biological issues in cancer risk assessment. In addition, the International Agency for Cancer Research (IARC) has evaluated and published carcinogenic risk to humans for hundreds of chemicals (13, 14).

In both systems, chemicals, including pharmaceuticals, are assigned to five groups: 1) carcinogenic to humans; 2) probably carcinogenic to humans; 3) possibly carcinogenic to humans; 4) not classifiable for human carcinogenicity; and 5) probably not carcinogenic to humans. Assignment to one of these groups is based on scientific judgement of data derived from studies in humans and animals as well as supporting data. Data are estimated as providing sufficient, limited, or inadequate evidence for carcinogenicity in humans and rodents.

For example, agents assigned as probably carcinogenic to humans have shown a positive association between exposure and human cancer, and have also shown sufficient evidence of carcinogenicity in animals. For the possible human carcinogens, only sufficient evidence for carcinogenicity in animals is established, but inadequate or no data in humans.

Evolution in Carcinogenicity Testing and Concepts Between 1960 and 2000

Mechanistic knowledge of carcinogenicity has increased enormously since the introduction of the Standard Chronic Bioassay. First, "genotoxic" carcinogens were distinguished from "nongenotoxic" carcinogens, which

Table 3 Main classification of chemical carcinogens

Category	Classification (+ example)	Features
Genotoxic	DNA-alkylating	Transspecies occurrence
	activation-independent (mustard gas)	At low dose levels
Epigenetic	activation-dependent (nitrosamines)	Short latency period
	Interfering with DNA (inorganic metals)	No threshold for extrapolation to humans
	Promoters (phenobarbital)	Mostly in one species, strain or sex
	Cytotoxic/mitogenic agents (<i>d</i> -limonene, saccharin)	At high dose levels (MTD)
	Hormone-modifiers (estrogen)	Long latency period
	Immunosuppressors (cyclophosphamide)	Threshold considerable for extrapolation to humans
	Peroxisome proliferators (clofibrate)	
	Receptor-mediated (dioxins)	
	Miscellaneous mechanisms (sodium nitrilotriacetic acid)	

led to a separate test battery for genotoxicity testing. Second, a “multistage” concept in carcinogenesis was discovered up to a molecular level. Third, “lifestyle factors” were found to play a major role in human cancer causes.

Genotoxic vs. nongenotoxic carcinogens

During the 1970s, Ames postulated that most human carcinogens were “genotoxic.” Genotoxic compounds may react, either directly or indirectly (after metabolic activation) with DNA, which leads to alteration of the genetic material, and mostly to carcinogenicity in rodents and humans. On the other hand, many rodent carcinogens are “nongenotoxic” or “epigenetic,” and act through a different mechanism.

The knowledge of these two basic mechanisms in carcinogenicity has led to the current classification of chemical carcinogens in “genotoxic carcinogens” and “nongenotoxic” or “epigenetic carcinogens.” The second category comprises agents that exert primary cellular effects, which further result in carcinogenicity in rodents (“rodent carcinogens”). Genotoxic and nongenotoxic carcinogens, each with their own typical features, are further classified in main categories, as shown in Table 3. An example of an epigenetic rodent carcinogen is the artificial sweetener saccharin. In the past, when saccharin was shown to produce bladder tumors in rats, it was considered to be a likely (possible) human carcinogen. Today, it has been removed from the list of suspected human carcinogens since a large body of experimental data indicates that the rat bladder tumors arise from mechanisms that are not relevant to the human situation, which in turn supports the conclusion that saccharin is not related to bladder cancer in humans. Phenobarbital, a sedative, is an example of an agent that extensively

induces drug-metabolizing enzymes in rodents. This results in liver tumors, but also in enhanced catabolism of thyroid hormones by the liver enzymes, which leads to thyroid stimulation by compensatory mechanisms and further to thyroid hypertrophy and hyperplasia, and eventually to thyroid tumors. Epidemiological studies, however, have indicated that it is of no concern for humans. More detailed explanations and examples of epigenetic carcinogens have been published by Purchase (15), Shaw and Jones (16), Van Cauteren et al. (17), and Williams and Whysner (18).

Due to the increase in knowledge of genotoxic and nongenotoxic mechanisms, test batteries were designed to evaluate “genotoxicity” in the early phases of drug development and prior to the start of chronic bioassays. These assays varied from test systems such as bacteria, yeast cells, and *Drosophila*-flies, to cell cultures of mouse lymphoma cells or human lymphocytes and in vivo systems, such as rodents for exposure of bone marrow. An overview of the genotoxicity toxicity assays that are currently performed for regulatory testing is shown in Table 4.

As indicated in Table 4, the potential to induce gene mutations or chromosome aberrations is investigated in a battery of genotoxicity studies. Additionally, DNA-damage and repair can be investigated separately. For the in vitro assays, some chemicals only become genotoxic after bioactivation to an electrophilic reactant. Therefore, in some in vitro assays, a liver enzymatic fraction is added in order to mimic the normal in vivo metabolic (de)activation route of compounds. The knowledge of the genotoxic versus the nongenotoxic carcinogens and of the mechanism of epigenetic carcinogenesis in rodents has gained significant importance in the extrapolation or relevance to humans. For example, humans have been

Table 4 Overview of genotoxicity assays

Genotoxicity assays	Genetic damage type
Assays according to current ICH-4 Standard battery:	
Ames reversion test in salmonella typhimurium or <i>Escherichia coli</i> (in vitro)	Gene mutations
Chromosomal aberration test in cell culture of human lymphocytes (in vitro) or	Chromosome aberrations
Mouse lymphoma test in mouse lymphoma cells (in vitro)	Chromosome aberrations and gene mutations
Micronucleus test in rodent bone marrow (in vivo)	Chromosome aberrations
Additional assays if cause for concern:	
Unscheduled DNA synthesis (UDS) test in rodent hepatocytes (in vitro/in vivo)	Primary DNA-damage
32P-postlabeling test	Primary DNA-damage
Test with transgenic animals	Gene mutations
Chromosome aberration test in human lymphocytes (in vitro) or in rodent bone marrow (in vivo)	Chromosome aberrations

exposed to many pesticides known to cause cancer in experimental animals by epigenetic mechanisms, but none has been definitely linked to cancer in humans. The absence of effects in humans is usually because exposure is below the threshold for the epigenetic effects. In addition, some cancer mechanisms may be specific to the rodent test animals and not relevant to humans at any possible exposure.

Multistage cancer concept

Armitage and Doll postulated the multistage cancer concept in the 1960s. According to this theory, tumors can arise from single cells that obtain proliferative capacity after several consecutive independent events at cellular level. Cancer of the colon is one of the best-understood examples of multistage carcinogenesis; it reveals at least six stages in its development. Recently, it also has become possible to identify the molecular events that underlie the development of these tumors. Some examples of somatic alterations in this model are gene mutations of a “ras” proto-oncogen, detected in approximately 50%, and loss of a “P 53” tumor suppressor gene in approximately 75% of the colorectal carcinomas. Similar alterations have also been detected in other human tumors, such as lung, breast, and brain tumors. More detailed information and examples of multistage carcinogenesis are provided by Fearon and Vogelstein (19), Couch (20), and Purchase (15). Based on the above theories, the main steps involved in carcinogenesis are: 1) “initiation” (first stimulus leading to alteration of the DNA in the cell nucleus); 2) “promotion” (further expression of the genetic change, leading to benign tumors); and 3) “progression” (further evolution of the tumor leading to malignancy and clinical manifestation).

“Initiation” is the first phase of carcinogenesis whereby cells are exposed to a carcinogenic agent. In this stage, an irrevocable step is taken in which daughter cells may later

acquire relative autonomy with regard to cell division. “Promotion” is the next phase, in which initiated cells, not yet recognizable, are stimulated to divide and become clinically or pathologically detectable neoplasms. “Progression” is the phase in which the tumor increasingly damages the host and finally destroys it. Widespread invasion and metastasis, with destruction of the original normal tissue, are predominant in this phase.

Lifestyle factors

Finally, associated causes of major human cancers have been discovered. Most cancers are associated with lifestyle (specifically tobacco and excessive alcohol use), inappropriate nutritional traditions, and lack of exercise. These lifestyle components involve currently known genotoxic carcinogens, and importantly, nongenotoxic carcinogens. The effect of nongenotoxic carcinogens is highly dose-dependent and also reversible upon lowering the dose below a threshold. Thus, it is quite possible to lower human cancer risk as well as the risk of related chronic diseases, such as coronary heart disease, hypertension and stroke, and adult on-set diabetes, by proper lifestyle adjustments. Clearly, the Delaney Clause (requiring total absence of carcinogens in the food chain) plays no role in disease prevention (5). Misconceptions about the relationship between environmental pollution and human disease, particularly cancer, have also been highlighted by Ames and Gold (21). Smoking, dietary imbalances, chronic infections, and hormonal factors, all influenced by lifestyle, are considered major causes of cancer. Reduction in synthetic pesticides in the food chain does not effectively prevent diet related cancer, whereas high consumption of fruits and vegetables (containing anti-oxidants) does. On the other hand, humans do ingest many natural chemicals (in fruits and plants). Even though only a small proportion has been tested for carcinogenicity, half

of them are rodent carcinogens. Therefore, prevention of cancer will not only be based upon carcinogenicity testing and risk assessment of chemicals, but from knowledge obtained from biomedical research, education of the public, and lifestyle changes (21).

Need for Changes

Deficiencies of the standard chronic bioassay

If clear-cut evidence of genotoxic potential is obtained in the genotoxicity battery, especially via *in vivo* assays, then the potential for carcinogenicity in humans is highly suspected. In that case, drug development is discontinued. For compounds with negative or equivocal results in the genotoxicity battery and positive results in the Standard Chronic Bioassay, interpretation can be difficult.

The concept of the bioassay for detecting carcinogenic potential was developed at a time when relatively few agents were recognized as being carcinogens. However, during the past 20 years, many investigations have shown that it is possible to provoke a carcinogenic response in rodents by a wide diversity of experimental procedures, many of which are considered to have little or no relevance for human risk assessment. The outcome of the Standard Chronic Bioassay has been shown to be positive for about 50% of a random selection of chemicals in the US NTP program (22). These bioassays have frequently identified different target organs in rats and mice and even in different strains of one species. Further, the data from bioassays were often contradictory, with simultaneous increases and decreases of tumors in different tissues (23).

Drugs have to be administered at very high doses (MTD). The paradox is that the safer the chemical, the higher the MTD, and the more likely that biochemical distortions will lead to cellular injury or cell death, abnormal compensatory cell replication, toxic hyperplasia, and toxicity-induced cancer, which is not considered relevant to man (24). A comparative study between animal experiments and epidemiological studies on chemicals also showed that animal studies with test doses above the MTD resulted in carcinogenic effects in multiple organs that had no or limited predictive value for man (25).

The IARC has provided additional indications for the lack of human relevance by the chronic bioassay. IARC work groups have evaluated hundreds of pharmaceuticals for their carcinogenic properties. However, only 20 pharmaceuticals were conclusively carcinogenic to humans and 52 were "probably" or "possibly" carcinogenic (13, 14). The pharmaceuticals known to cause cancer in humans each possess at least one of the four

properties: genotoxicity, immunosuppression, hormonal activity, or chronic irritation (cytotoxicity/mitogenic activity). These properties can be identified by genotoxicity studies *in vivo* and in well-designed toxicology studies of up to 6 months duration in rats (26). The absence of effects in humans was also considered to be due to an exposure in man below the threshold dose for the epigenetic effects or due to cancer mechanisms that were specific to the rodent test animal.

Therefore, the findings from nongenotoxic compounds can be shown to lack relevance for humans because either the animal exposure is excessive as compared to human exposure, or the response to a carcinogenic challenge is qualitatively different in rodents from that in humans (26). In case of a different mechanism, additional mechanistic studies often have to be performed to prove the lack of relevance to man.

International Conference on Harmonization (1991–1997)

Since 1991, the International Congress of Harmonization (ICH 1, 2, 3, and 4, respectively, taking place in 1991, 1993, 1995, and 1997) has met to promote harmonization of regulatory requirements between its regions (Japan, US, and Europe) on Safety, Quality, and Efficacy of human pharmaceuticals. Both regulatory authorities and industry associations were involved in these discussions.

Based upon the deficiencies of the standard chronic bioassay as well as on increased mechanistic knowledge, the following needed to be defined: 1) when carcinogenicity studies should be performed; 2) the use of new methods in carcinogenicity testing, and 3) guidance for dose selection (27). These three safety topics are described in more detail in Table 5.

As indicated in Table 5, carcinogenicity studies (see S1A) are only required for long-term human exposure of chemicals, but may also be recommended if there is cause for concern even with compounds lacking long-term exposure. ICH topic S1B takes into account the recent scientific developments, rather than harmonizing between nations. Several large databases of long-term rodent carcinogenicity studies conducted over the past 25 years indicate consistently that approximately 50% of chemicals yield a positive result. These pharmaceutical databases derived from pharmaceuticals only were based upon the shortcomings of the Standard Chronic Bioassay previously described. The pharmaceutical databases are summarized in Table 6.

The ICH addressed several main questions. These included: "Would the use of rats, but not mice, result in loss of information on carcinogenicity that would be relevant to human risk assessment?" and "Has a positive

Table 5 ICH topics on carcinogenicity testing

S1A. “Need for carcinogenicity studies”
Required for any pharmaceutical with expected “clinical use of at least 6 months”
Recommendation if there is “cause for concern:”
1. Previous demonstration of carcinogenic potential in a product class that is considered relevant to humans
2. Structure-activity relationship (SAR) suggesting carcinogenic risk
3. Evidence of preneoplastic lesions in repeated dose toxicity studies and
4. Long-term tissue retention of parent compound or metabolite(s) resulting in local tissue reactions or other pathological responses
S1B. “Testing for carcinogenicity of pharmaceuticals”
Flexibility and judgement should be exercised in the choice of approach.
Either 2 long-term carcinogenicity studies (one in the rat and one in the mouse) or one long-term study plus one other study with a shorter duration (“one plus approach”).
S1C. “Dose selection in carcinogenicity studies”
Criteria for dose selection in carcinogenicity studies:
1. Toxicity-based endpoints (MTD)
2. Pharmacokinetic endpoints (25-fold AUC ratio: rodent-human)
3. Saturation of absorption (highest exposure reached)
4. Pharmacodynamic endpoints (e.g. sedation at high dose levels)
5. Maximum feasible dose (e.g. maximal solubility reached)
6. Other additional endpoints.

result in mice (but not in rats) correctly prevented a nongenotoxic drug from being marketed”? (32).

Results of the database surveys proposed that rats were more “sensitive” than mice and that tumorigenicity detected in mice only was never the sole reason for regulatory action. Furthermore, findings in rats only were twice as frequent as in mice only; all known human carcinogens were positive in rats (32). Therefore, it was proposed that normally one long-term study in one rodent species would suffice. The species should be the most appropriate and on practical convenience, the rat would be preferred. From the European point of view, one chronic bioassay in the rat would be sufficient. However from the US perspective, an additional short-term study was requested. The short-term models are further explained later in this article.

Finally, more criteria than the sole MTD were acceptable for dose selection (see Table 5, S1C). The availability of multiple acceptable criteria for dose selection provides greater flexibility in optimizing the design of carcinogenicity studies.

CARCINOGENICITY TESTING IN THE FUTURE

Alternative Carcinogenicity Models

According to ICH S1B, a “one plus approach” might be possible instead of the 2-year bioassay in rats and mice. In this approach, the species to be used in the one long-term study should first be selected, and secondly, the

Table 6 Overview of pharmaceutical databases

Continent	Scope	Reference
IARC (France)	International Agency for Cancer Research	13, 14
FDA (US)	Food and Drug Administration	28
PDR (US)	Physicians Desk Reference	29
JPMA (Japan)	Japanese Pharmaceutical Manufacturing Association	30
CPMP (EC)	Committee for Proprietary Medicinal Products	31
CMR (UK)	Center for Medicines Research	30

alternative model should be chosen, both based upon various selection criteria. Selection criteria for both the species and strain to be used in the long-term study and in the alternative model are summarized in Table 7. A description of the various alternative short-term models is given.

Transgenic mice

Transgenic mice may provide advantages in developing a more specific model and reducing the number of animals and the time required for bioassays (6–9 months daily dosing). Such models include transgenic mice that carry reporter genes that may serve as targets for genotoxic events or mice carrying specific oncogenes or inactivated tumor suppressor genes that are important factors that contribute to the multistage process of carcinogenesis.

Table 7 ICH S1B “testing for carcinogenicity of pharmaceuticals”: Rationale

Select most appropriate species for the long-term study:

Based on comparative studies in two or more rodent species:

1. Pharmacology
2. Repeated dose toxicity studies
3. Metabolism
4. Toxicokinetics
5. Route of administration
6. In the absence of clear evidence favoring one species, the rat is recommended.

Select most appropriate alternative model:

Models:

1. Transgenic rodents (in vivo)
2. Neonatal rodents (in vivo)
3. Initiation–Promotion models (in vivo)
4. Syrian Hamster Embryonic cell (SHE) assay (in vitro)

Prior to inclusion of any new method, it is critical that the method be evaluated for:

1. Is new, additional information expected?
2. Will concerns be addressed?
3. Is there a comparable metabolism in man?
4. Is there a comparable exposure in man?
5. Is there literature or other evidence on the value and relevance of the model?

Mouse lines with defined genetic alterations that result in over-expression or inactivation of a gene intrinsic to carcinogenesis, but that are insufficient alone for neoplastic conversion, are promising models for identification and evaluation of rodent and/or human carcinogens. Likely models are:

1. Transgenic animals that over-express (proto-) oncogenes, such as:
 - a. “TG.AC” mouse line (a “skin-painting model”), which expresses a mutation of the “v-ras” proto-oncogene (33).
 - b. “Tg-rasH2” mouse line, which expresses a mutation and amplification of the human “c-Ha-ras” proto-oncogenes in different tissues (34).
2. Transgenic animals that lack certain genes (“knockout” animals), such as:
 - a. “P53 +/-” mouse line, which expresses heterozygous inactivation of the P53 “tumor suppressor gene,” which is critical to cell cycle control and DNA repair (33).
 - b. “XPA-/-” mouse line, which expresses homozygous absence of “DNA-repair genes” (35).

Neonatal mice

Neonatal rodents have been studied since the 1960s. The test on newborn rodents (12 months of duration) has been extensively investigated in the United States and in Japan, and has been demonstrated to be highly sensitive to genotoxic carcinogens (36).

Meanwhile, further investigations have revealed that dosing on days 8 and 15 of age at the MTD and at half the MTD appears to be the current dosing regime, followed by a 12-month observation period.

The susceptibility of neonatal mice to tumorigenicity has been demonstrated and is explained by the fact that metabolic activity of a chemical in neonates is in the developing stage. Delay of excretion of a chemical via metabolic pathways results in prolongation of the presence of a chemical in the body. Such prolongation of the presence of a carcinogen would have a greater chance of changing normal cells into tumor cells.

Initiation—promotion models (in vivo)

The principle of this model is that in order to evaluate a substance as an “initiator,” it will be administered in a single dose or over a period of several days or weeks. After several weeks of washout, a “promotor” (e.g., phenobarbital) is administered, and several months later, the number of pre-neoplastic or neoplastic changes are examined.

To evaluate whether an agent acts as a “promotor,” the procedure is reversed. Following initiation with a known initiator (e.g., diethylnitrosamine), the substance under investigation is administered intermittently, at different dosages, over a number of months. In retrospect, this model might be valuable in establishing the mechanism underlying carcinogenicity (17).

Syrian hamster embryo (SHE) cell transformation assay (in vitro)

In this system, normal cells are isolated from 13-day-old embryos, and these cells are treated with carcinogenic agents capable of inducing genetic alterations that can produce changes in cellular and colony morphology that result in “morphologically transformed” SHE cells. These cells will senesce unless they undergo the additional genetic alterations necessary to acquire immortality and tumorigenicity. Several intermediate stages in this system, including morphological transformation, immortality, acquisition of tumorigenicity, and malignant progression, shape the alterations observed during *in vivo* neoplastic transformation and possibly explain why the SHE assay is able to identify rodent carcinogens (37). In the SHE assay, compounds are cultured over a period of 24 h–7 days, and colony formation is examined as the endpoint of clonal transformation.

Evaluation, Validation, and Implementation of the Alternative Models

Further description of alternative models and their evaluation and validation is given by Blain et al. (38). Validation exercises are taking place at the American National Institute of Environmental Health (NIEH), the International Life Science Institute/Health and Environmental Sciences Institute (ILSI/HESI) (39), the Japanese Central Institute for Experimental Animals/National Institute of Health Sciences (CIEA/NIHS), and the Dutch Public Institute for Safety and Environment (RIVM).

The NTP is also evaluating several lines of genetically-altered mice for possible use in identifying and assessing carcinogens. A number of comments and concerns were raised, offering some thoughts on future directions for this line of research as well as for the possible ways in which genetically altered mice might be integrated into a comprehensive testing strategy (40).

In the industry, the ICH S1B guidelines have been implemented, and have led to changes in carcinogenicity

testing approaches, as well as changes in evaluation by the regulatory authorities. Guidelines increase flexibility, which in turn obligates industry and regulatory authorities to use more scientific, evidence-based decision-making. The changes are anticipated to significantly improve the relevance of the assessment of carcinogenic risk for humans, but might also lead to confusion and occasional disagreement on appropriate test strategies for specific drugs (41). Alternative test models might be used in various scenarios as follows: 1) as an alternative to a second 2-year rodent carcinogenicity study; 2) as a complementary confirmatory study for drugs with equivocal carcinogenicity findings in 2-year studies; 3) as a preliminary screen to set priorities for full carcinogenicity testing; 4) as an alternative to repeating a 2-year rodent carcinogenicity study; or 5) to assess the carcinogenic potential of genotoxic contaminants or degradation. Based upon available information, there is sufficient experience with some *in vivo* transgenic rodent carcinogenicity models to support their application as complementary second species studies in conjunction with a single 2-year rodent carcinogenicity study when appropriately justified (42). Therefore, a weight-of-evidence approach is justified.

Weight-of-Evidence Approach

A weight-of-evidence analysis on the level of suspicion for a carcinogenic risk should be performed based upon all available pre-clinical information that could raise concern for carcinogenicity. These concerns include carcinogenicity results, structure-activity-relationship (SAR), class evaluation, genotoxicity findings, repeated dose toxicity findings, interspecies comparison of pharmacodynamics, exposure and metabolism, and information on the intended clinical use, patient population, dose regimen, and pharmacokinetics and pharmacodynamics. SAR consists of finding the relationship between different (usually computable) characteristics of the molecules and the induced biological activity. The resulting model can be used for predicting (estimating) the biological activity, and in this case, carcinogenicity. Various commercial computerized systems are available for SAR, as described by Richard (43). However, the performance of SAR is still limited and affected by over sensitivity (43, 44).

Schwetz and Gaylor proposed a potential strategy for the weight-of-evidence approach that includes alternative carcinogenicity models (45). This strategy proposes alternate approaches, according to SAR, or findings in the toxicity and genotoxicity studies that would provide more mechanistic information.

Elements that should be included in the weight-of-evidence approach, including alternative models, are provided by Contrera and DeGeorge (42). These include evaluation of the results of carcinogenicity studies (including alternative models) and genotoxicity studies, as well as assessment factors in the consideration of the adequacy of the test models and risk-benefit considerations. It is apparent that different tests might be warranted for the various regulatory areas, such as pharmaceuticals, food additives, environmental, or industrial chemicals (42, 45).

Another option of performing a weight-of-evidence approach “before” starting the classical and/or alternative carcinogenicity bioassays is to submit a “Carcinogenicity Testing Strategy Paper” to the authorities (46). In this way, a dialogue between authorities and industry can be reached in order to effect an agreement on the approach and dose selection prior to initiation of pivotal studies.

In the United States, the Center for Drug Evaluation and Research (CDER) engages in dialogue with industry to reach agreements on “approach” and “dose selection” prior to initiation of pivotal studies (41). In addition, a procedure of scientific advice can be found in the European Medicines Evaluation Agency (EMA)/Committee for Pharmaceutical and Medical Products (CPMP). This agency is open for questions regarding the requirement for a rationale. Companies are encouraged to inform the CPMP of the proposed carcinogenicity testing strategy (47).

Other Future Opportunities

Data from short to medium-term toxicity studies that precede carcinogenicity studies reveal that most of the nongenotoxic agents which induce tumors in rodents also produce other pathological changes in the tissues in which the tumors develop and at dose levels at which tumors are observed. These early changes range from altered hormone levels, impaired ion balance, and organ enlargement to specific and marked histopathological changes (48). These findings may be used for early detection of nongenotoxic carcinogens, and may also be extremely valuable for designing protocols for long-term bioassays. Furthermore, a thorough understanding of such early indicators will lead to the elucidation of specific mechanisms involved in carcinogenesis. Together with examination of possible thresholds for underlying toxic events, this confirms the basis for assessment of carcinogenic risk and for the regulation of human exposure.

Based upon the above rationale, a “tier approach in carcinogenicity testing and assessment” of pharmaceuticals can be followed, possibly with refinement, reduction,

or replacement of test methodologies in carcinogenicity testing.

A first approach, according to the ICH S1A guideline scenario on the need for carcinogenicity testing, prescribes long-term carcinogenicity testing (in one or two species) for compounds with continuous or intermittent exposure to humans and compounds with cause for concern. If there is no long-term exposure to the compound, and if there is no cause for concern, no further action is recommended, whereas short- or long-term studies may be warranted for suspicious findings (46).

A second approach postulates that much of the information necessary to assess the carcinogenic potential of a new drug without a bioassay is usually available by the end of the first clinical studies in patients. (Suspicious findings from *in vivo* genotoxicity studies and 3–6 month toxicology studies aimed at assessing risk factors associated with carcinogenicity in humans include: genotoxicity, immune suppression, hormonal activity, and chronic irritation/mitogenic activity.) Evaluation of this package will, therefore, identify the presence or absence of the known causes of cancer from pharmaceuticals in humans, under conditions relevant to the use of the drug in question. If cause for concern remains at this stage, useful information on long-term adverse effects that might represent a carcinogenic hazard to humans may be obtained (e.g., from a 12-month study, usually in rats, conducted at clinically relevant dose levels) (49).

Finally, a third approach has been proposed with five stages that focus on the chemical structure, DNA-reactivity, epigenetic effects, limited bioassays, and finally, the application of “accelerated bioassays.” These accelerated bioassays require 40 weeks and apply to the use of sensitive markers for induction of neoplasia in comparison to positive control compounds for important organs in human carcinogenesis. It enables data acquisition of the entire carcinogenesis process directed toward developing mechanistic information. This system would have the potential to replace the chronic bioassay in rodents in some circumstances and could serve as an alternative to a chronic bioassay in a second species (50).

CONCLUSION

In the 20th century, the concept of carcinogenesis and carcinogenicity testing has evolved enormously, although the standard Chronic Bioassay still contains many of deficiencies. New carcinogenicity testing strategies, however, are to be expected. Also, validation results

with regards to the alternative carcinogenicity models will become available and lead into new insights in the most appropriate short-term carcinogenicity studies.

A weight-of-evidence or tier approach to the level of suspicion of carcinogenic risk will be the main objective, and will include SAR and findings from sub-chronic to chronic toxicity testing (genotoxicity, reproductive toxicity, and organ toxicity). In addition, communication between industry and regulatory authorities will be formulated early in order to allow the most optimal scientific approach.

REFERENCES

1. Van Cauteren, H.; de Kok, Th.M.C.M.; van Schooten, F.-J. Study Unit 12: Introduction to Carcinogenesis. *Toxicology: Principles and Applications*; Niesink, R.J.M., de Vries, J., Hollinger, M.A., Eds.; CRC Press, Inc. and Open University of the Netherlands: Boca Raton, FL, 1996; 347–383.
2. Roe, F.J.C. A Brief History of the Use of Laboratory Animals for the Prediction of Carcinogenic Risk for Man with a Note on Needs for the Future. *Exp. Toxic. Pathol.* **1998**, *50*, 271–276.
3. Grasso, P.; Sharratt, M.; Cohen, A.J. Role of Persistent, Non-Genotoxic Tissue Damage in Rodent Cancer and Relevance to Humans. *Annu. Rev. Pharmacol. Toxicol.* **1991**, *31*, 253–287.
4. Weissburger, J.H. The 37-Year History of the Delaney Clause. *Exp. Toxic. Pathol.* **1996**, *48*, 183–188.
5. Weissburger, J.H. Human Protection against Non-Genotoxic Carcinogens in the US without the Delaney Clause. *Exp. Toxic. Pathol.* **1996**, *48*, 201–208.
6. Weissburger, E.K. History of the Bioassay Program of the National Cancer Institute. *Progr. Exp. Tumor Res.* **1983**, *26*, 187–201.
7. Inveresk Research International. Rodent Carcinogenicity and Chronic Toxicity. A Review of Test Protocols for Pharmaceuticals, Agrochemicals, Food Additives and Industrial Chemicals According to European, American and Japanese Guidelines. *Regulatory Guidelines* **1990**, *1*, 1–13.
8. Feron, V.J.; Schwarz, M.; Hemminki, K.; Krewski, D. Long- and Medium-Term Carcinogenicity Studies in Animals and Short-Term Genotoxicity Tests, *IARC Scientific Publications*, No. 131; International Agency for Research on Cancer: Lyon, 1999; 103–129.
9. Portier, C.J. Design of Two-Year Carcinogenicity Experiments: Dose Allocation, Animal Allocation and Sacrifice Times. *Stat. Toxicol.* **1991**, 457–469.
10. Statistical Methods in Cancer Research. *The Design and Analysis of Long-Term Animal Experiments*; Gart, J.J., Krewski, D., Lee, P.N., Tarone, R.E., Wahrendorf, J., Eds.; No. 79 IARC Scientific Publications; 1986; 3.
11. Ciminera, J.L.; Allen, H.L. Carcinogenicity Testing. *Encyclopedia of Pharmaceutical Technology*; Marcel Dekker, Inc.: New York, 1998; 285–317.
12. Velasquez, S.F.; Schoeny, R.; Rice, G.E.; Cogliano, V.J. Cancer Risk Assessment: Historical Perspectives, Current Issues, and Future Directions. *Drug Chem. Toxicol.* **1996**, *19* (3), 161–185.
13. Marselos, M.; Vainio, H. Carcinogenic Properties of Pharmaceutical Agents Evaluated in the IARC Monographs Programme. *Carcinogenesis* **1991**, *12* (10), 1751–1766.
14. Vainio, H.; Coleman, M.; Wilbourn, J. Carcinogenicity Evaluations and Ongoing Studies: The IARC Databases. *Environ. Health Perspect.* **1991**, *96*, 5–9.
15. Purchase, I.F.H. Current Knowledge of Mechanisms of Carcinogenicity: Genotoxins Versus Non-Genotoxins. *Human Exp. Toxicol.* **1994**, *13*, 17–28.
16. Shaw, I.C.; Jones, H.B. Mechanisms of Non-Genotoxic Carcinogenesis. *TIPS* **March 1994**, *15*, 89–93.
17. Van Cauteren, H.; de Kok, Th.M.C.M.; van Schooten, F.-J. Study Unit 13. Cancer Risk Evaluation. *Toxicology: Principles and Applications*; Niesink, R.J.M., de Vries, J., Hollinger, A., Eds.; CRC Press, Inc. and Open University of the Netherlands, 1996; 384–412.
18. Williams, G.M.; Whysner, J. Epigenetic Carcinogens: Evaluation and Risk Assessment. *Exp. Toxic. Pathol.* **1996**, *48*, 189–195.
19. Fearon, E.R.; Vogelstein, B. A Genetic Model for Colorectal Igenesis. *Cell* **1990**, *61*, 759–767.
20. Couch, D.B. Carcinogenesis: Basic Principles. *Drug Chem. Toxicol.* **1996**, *19* (3), 133–148.
21. Ames, B.N.; Gold, L.S. Environmental Pollution, Pesticides, and the Prevention of Cancer: Misconceptions. *FASEB J.* **1997**, *11*, 1042–1052.
22. Ashby, J.; Tennant, R.W. Prediction of Rodent Carcinogenicity for 44 Chemicals: Results. *Mutagenesis* **1994**, *9* (1), 7–15.
23. Davies, T.S.; Monro, A.M. The Rodent Carcinogenicity Bioassay Produces a Similar Frequency of Increases and Decreases: Implications for Risk Assessment. *Reg. Toxicol. Pharmacol.* **1994**, *20*, 281–301.
24. Carr, C.J.; Kolbye, A.C. A Critique of the Use of the Maximum Tolerated Dose in Bioassays to Assess Cancer Risks from Chemicals. *Reg. Toxicol. Pharmacol.* **1991**, *14*, 78–87.
25. Meijers, J.M.M.; Swaen, G.M.H.; Bloemen, L.J.N. The Predictive Value of Animal Data in Human Cancer Risk Assessment. *Reg. Toxicol. Pharmacol.* **1997**, *25*, 94–102.
26. Monro, A. How Useful are Chronic (Life-Span) Toxicology Studies in Rodents in Identifying Pharmaceuticals that Pose a Carcinogenic Risk to Humans? *Adverse Drug React. Toxicol. Rev.* **1993**, *12* (1), 5–34.
27. Usui, T.; Masuda, H. Testing for Carcinogenicity of Chemicals: S1, A—Conditions which Require Carcinogenicity Studies for Pharmaceuticals; S1, B—Use of Two Rodent Species; S1, C—Dose Selection for Carcinogenicity Studies of Pharmaceuticals. *J. Toxicol. Sci.* **1996**, *21*, 475–478.
28. Contrera, J.F.; Jacobs, A.C.; DeGeorge, J.J. Carcinogenicity Testing and the Evaluation of the Regulatory Requirements for Pharmaceuticals. *Reg. Toxicol. Pharmacol.* **1997**, *25*, 130–145.
29. Davies, T.S.; Monro, A. Marketed Human Pharmaceuticals Reported to be Tumorigenic in Rodents. *J. Am. Coll. Toxicol.* **1995**, *14* (2), 90–107.
30. Usui, R.; Griffiths, S.A.; Lumley, C.E. Industry Viewpoint: The Utility of the Mouse for the Assessment of the Carcinogenic Potential of Pharmaceuticals. *Proceedings of*

- the Third International Conference of Harmonization; Yokohama, D'Arcy, P.F. Harron, D.W.G., Eds.; The Queen's University, Belfast, 1996; 279–284.
31. Van Oosterhout, J.P.J.; Van der Laan, J.W.; De Waal, E.J.; Olejniczak, K.; Hilgenfeld, M.; Schmidt, V.; Bass, R. The Utility of Two Rodent Species in Carcinogenic Risk Assessment of Pharmaceuticals in Europe. *Reg. Toxicol. Pharmacol.* **1997**, *25*, 6–17.
 32. Monro, A. Testing for Carcinogenic Potential. Rapporteur's Report Proceedings of the Third International Conference of Harmonization Yokohama, D'Arcy, P.F.; Harron D.W.G., Eds.; The Queen's University: Belfast 1996; 261–268.
 33. Tennant, R.W.; Spalding, J.; French, J.E. Evaluation of Transgenic Mouse Bioassay for Identifying Carcinogens and Non-Carcinogens. *Mutat. Res.* **1996**, *365*, 119–127.
 34. Yamamoto, S.; Mitsumori, K.; Kodama, Y.; Matsunuma, N.; Manabe, S.; Okamiya, H.; Suzuki, H.; Fukuda, T.; Sakamaki, Y.; Sunaga, M.; Nomura, G.; Hioki, K.; Wakana, S.; Nomura, T.; Hayashi, Y. Rapid Induction of More Malignant Tumors by Various Genotoxic Carcinogens in Transgenic Mice Harboring a Human Prototype c-Ha-ras Gene than in Control Non-Transgenic Mice. *Carcinogenesis* **1996**, *17*, 2455–2461.
 35. De Vries, A.; Van Oostrom, C.T.M.; Dortant, P.M.; Beems, R.B.; Van Kreijl, C.F.; Capel, P.J.A.; van Steeg, H. Spontaneous Liver Tumors and Benzo(a)pyrene-Induced Lymphomas in XPA-Deficient Mice. *Mol. Carcinog.* **1997**, *19* (1), 46–53.
 36. Flammang, T.J.; Von Tungeln, L.S.; Kadlubar, F.F.; Fu, P.P. Neonatal Mouse Assay for Tumorigenicity: Alternative to the Chronic Rodent Bioassay. *Reg. Toxicol. Pharmacol.* **1997**, *26*, 230–240.
 37. LeBoeuf, R.A.; Kerckaert, K.A.; Aadema, M.J.; Isfort, R.J. Use of Syrian Hamster Embryo and BALB/c 3T3 Cell Transformation for Assessing the Carcinogenic Potential of Chemicals. *IARC Scientific Publications*; No. 146; McGregaro, D.B., Rice, J.M., Venitt, S., Eds.; Lyon, 1999; 409–425.
 38. Blain, P.G.; Battershill, J.M.; Venitt, S.; Cooper, C.C.; Fielder, R.J. Consideration of Short-Term Carcinogenicity Tests Using Transgenic Mouse Models. *Mut. Res.* **1998**, *403*, 259–263.
 39. Robinson, D. The International Life Sciences Institute's Role in the Evaluation of Alternative Methodologies for the Assessment of Carcinogenic Risk. *Toxicol. Pathol.* **1998**, *26* (4), 474–475.
 40. Bucher, J.R. Update on the National Toxicology Program (NTP) Assays with Genetically Altered "Transgenic" Mice. *Environ. Health Perspect.* **1998**, *106* (10), 619–621.
 41. DeGeorge, J. Challenges in Application of New Approaches to Carcinogenicity Testing for Pharmaceuticals. *Toxicol. Lett.* **1998**, *102–103*, 565–568.
 42. Contrera, J.F.; DeGeorge, J.J. In Vivo Transgenic Bioassays and Assessment of the Carcinogenic Potential of Pharmaceuticals. *Environ. Health Perspect.* **1998**, *106* (1), 71–80.
 43. Richard, A.M. Structure-Based Methods for Predicting Mutagenicity and Carcinogenicity: Are we there yet?. *Mutat. Res.* **1998**, *400*, 493–507.
 44. Benigni, R. The First US National Toxicology Program Exercise on the Prediction of Rodent Carcinogenicity: Definitive Results. *Mutat. Res.* **1997**, *387*, 35–45.
 45. Schwetz, B.; Gaylor, D. New Directions for Predicting Carcinogenesis. *Mol. Carcinog.* **1997**, *20*, 275–279.
 46. Van Deun, K.; Van Cauteren, H.; Vandenberghe, J.; Canning, M.; Vanparys, P.; Coussement, W. Review of Alternative Methods of Carcinogenicity Testing and Evaluation of Human Pharmaceuticals. *Adverse Drug React. Toxicol. Rev.* **1997**, *16* (4), 215–233.
 47. Van der Laan, J.W. New Perspectives for Alternative Approaches to Carcinogenicity Testing: A Regular Viewpoint. *Toxicol. Lett.* **1998**, *102–103*, 561–564.
 48. Hildebrand, B.; Grasso, P.; Ashby, J.; Chamberlain, M.; Jung, R.; van Kolfschoten, A.; Loeser, E.; Smith, E.; Bontinck, W.J. Validity of Considering that Early Changes may Act as Indicators for Non-Genotoxic Carcinogenesis. *Mutat. Res.* **1991**, *248*, 217–220.
 49. Monro, A. Are Life-Span Rodent Carcinogenicity Studies Defensible for Pharmaceutical Agents?. *Exp. Toxic. Pathol.* **1996**, *48*, 155–166.
 50. Williams, G.M.; Iatropoulos, M.J.; Weisburger, J.H. Chemical Carcinogen Mechanisms of Action and Implications for Testing Methodology. *Exp. Toxic. Pathol.* **1996**, *48*, 101–111.

CLINICAL DATA MANAGEMENT SYSTEMS

Samuel Givens

Hoffmann-LaRoche, Nutley, New Jersey, U.S.A.

Debra Barnes

Hoffmann-LaRoche, Palo Alto, California, U.S.A.

Victoria Imber

CastleRock Technologies, Inc., Sausalito, California, U.S.A.

Barbara Perry

Hoffmann-LaRoche, Welwyn, United Kingdom

INTRODUCTION

Overview

This article is intended to be an overview of clinical data management systems and the processes they support. Data management systems are highly dependent on the size and complexity of the organization using them. Systems can range from a set of SAS data sets to a fully integrated, distributed set of applications using a relational database. The entire data management process may employ a variety of technical solutions.

The information presented here is based on experience with processes and technology in a large international pharmaceutical company. Many of these concepts are employed in smaller pharmaceutical companies and contract research organizations (CROs) but on a lesser scale.

The core data management processes are represented in the study lifecycle illustrated in Figs. 1 and 2.

Re-Engineering

The pharmaceutical industry is under constant pressure to bring drugs to market more quickly and less expensively, without compromising the quality of the products. The desire to achieve a profitable balance between these three objectives—speed, cost, and quality—results in perpetual “re-engineering.” The investment put into developing and implementing a solid database and streamlining data management tools and processes can contribute greatly to the success of this effort.

How Technology is Driving Changes in Data Management

The introduction and proliferation of the Internet and web-based applications is having a profound impact on the

conduct of clinical trials. The Internet provides us with the ability to communicate easily with CROs and investigators; and them with us, without compromising corporate security. Data sets can easily be placed on a secure web site for being reviewed and updated by a partner. Remote data entry will reach its full potential as a result of the introduction and acceptance of internet technology.

STUDY SET-UP

Standardization

To avoid redundancy, many parts of the study and set-up processes can be standardized and reused (Fig. 1). For example, as case report forms (CRFs) are developed for use across studies, the corresponding components of the study definition (questions, response values), the validation checks, reports, and extract data structures can be reused as well, especially within a drug project. This standardization allows for more efficient use of resources and systems, as well as realization of benefits in CRF design, study definition, validation, analysis, reporting, metrics, and training. Standards for data collection and processing are also necessary to fulfill the reporting requirements for pooled analyses across a project (e.g., safety updates, integrated reports, summaries).

Protocol Development

The protocol must contain a clear statement of the objectives of the investigation (primary and secondary endpoints) and the methods of analysis to be used. The various sections may be written by the appropriate team members and assembled by the clinical representative who is the protocol author. Once the protocol is complete, the

CRF is designed. The CRF is the principal document used to collect the study data and guide the study definition in the data management system.

In-House vs. Outsourcing to CROs

Once the project team completes a protocol synopsis, in-house resources should be evaluated. Consideration is given to the type of study required along with the study's priority within the project. The advantages of doing the data management in-house include:

- All data for project resides on one database
- Full control over study set-up
- Re-use of tools already developed
- Real-time access to data
- Control over study costs.

If in-house resources are limited and the decision is made to outsource data management activities, the planning process begins by creating a detailed scope of work statement and issuing the request for proposal (RFP). Bids should be solicited from several CROs.

In choosing a CRO, consider the following:

- Depth of experience in data management
- Compatibility of computer systems
- Qualified technical support for data transfer
- SOPs and policies that meet GCP criteria
- Interpersonal compatibility
- Price

Previous sponsor experience with the CRO, the CRO's track record of delivering, and the CRO reputation within the industry are more important than price.

When a CRO has been chosen, the contract should include a detailed definition of the scope of work required along with a clear understanding of when the data is considered clean and ready for analysis. A test run of data transfer from the CRO to the sponsor should be done early in the study to identify any problems in data formatting and transmission. The milestones and deliverables must be tracked closely during study conduct in order to ensure that appropriate progress payments are made. The importance of regular communication cannot be overemphasized. The early identification and resolution of technical or process problems is necessary for a smooth database closure and transfer of data. Transfer of data often contains more issues and surprises than anticipated.

The sponsor must take the time to gain a thorough understanding of the CRO's organization, work processes, and needs with respect to the project. In turn, the CRO must understand the sponsor's structure, organiz-

ation of the project management team, and the role the CRO is expected to fill in the overall operational process of the clinical study. This building of mutual understanding takes time and effort but is crucial for project success.

The project team requires reports to track the progress of the study including patient enrollment data, discrepancy counts, outstanding CRF pages, and terminated patients (including dropouts). If the CRO already has adequate tracking systems, the reports should be evaluated and adapted as needed.

After study completion, a review of the CRO's performance and a written report of lessons learned will provide information for future planning of projects and outsourcing needs.

There are several options for the transfer of CRO-processed data back to the sponsor. Most often, the final data are entered into the CRO's database to produce the final study reports with corresponding datasets. However, with the advances in the Internet and in distributed study conduct, it is possible, and generally desirable from the sponsor's viewpoint, for the CRO to enter data directly into the sponsor's database.

CRF Development

The CRF design process can begin either following or concurrent with the protocol development. Well-designed data collection forms are critical to achieve the objectives of the clinical trial. Consideration should be given to the content, format, and layout of the forms since all these factors contribute to the overall quality and accuracy of the data that will be collected, processed, and reported.

Many disciplines should participate in the CRF design stage. The core team will generally consist of representatives from statistics, data management, forms design, medical and clinical monitoring groups, with other specialists, consulting, as necessary. The primary objective of the team in this process is to optimize and balance the following requirements for the CRF:

- To facilitate the investigational site in filling out the forms correctly
- To allow for quick and accurate data entry
- To ensure that data can be analyzed and that they represent the patient's experience for statistical and clinical reporting
- To facilitate the pooling of data across a project for safety updates and integrated safety and efficacy reporting
- Consistency within project or area (e.g., pharmacoeconomics, clinical pharmacology) to allow reuse of tools.

The forms should only collect data that are needed for reporting purposes and avoid collecting unnecessary or redundant data.

The standardization of CRFs for use across multiple studies results in significant savings in the design, processing, reporting, and training resources required for a clinical study. Forms can be further broken down into modules (e.g., physical exam, vital signs, demography) or groups of questions. The use of these modules allows for greater flexibility when constructing the forms/pages while retaining the standard use of the question groups. A library of forms can be centrally gathered and maintained that includes “global” forms (those that can be used across projects—e.g., adverse events, demography), project or therapeutic standard forms (used across a project), and study specific forms.

There are many applications and systems that can be used to design and generate the CRFs. These range from word processing and desktop publishing packages to customized systems that facilitate the maintenance and use of a library of modules/forms and enforce standards. The CRF may also be in an electronic format rather than on paper, as in the case of remote data entry systems.

Randomization

Randomization is the process by which patients are randomly assigned to a treatment group. It is used to reduce the possibility for investigators and study personnel to bias the results (consciously or unconsciously) in favor of one treatment over another in a study. Randomization also allows for maintaining the blinding of a study when the blind must be broken for an individual patient. In most trials, the randomization data will be kept blinded until data are considered clean and after exclusions are decided to avoid influencing the results of a study. At the end of the trial, it is a requirement to confirm the integrity of the blinding. There must be documentation or an audit trail of all blind breaks and of all data changes post unblinding.

The statistician plays an important role in specifying appropriate parameters to be utilized in generating the randomization. The randomization specifications include treatments, centers, block size, study design, blinding requirements, and stratification factors. It is through the usage of stratification or grouping criteria that patient differences can be minimized between treatment groups.

Randomization codes are used for packaging and labeling study medications. The systems that are used to generate labels for the treatment bottles may be independent of those that produce the randomization.

Study Definition

A study definition is used to identify the system characteristics about the data fields stored within, criteria for acceptance of those data, as well as extract formats and file structures. These characteristics may include database variable name, data type (numeric, character, date/time), question name and label, short reference name (e.g., SAS^a variable name), format (e.g., ddmmyy), field length, acceptable response values (e.g., male/female), coding formats (e.g., 1 = yes, 2 = no), and validation information (e.g., validate against specific thesaurus).

Data entry screen layout is also part of the study definition. The ease (or difficulty) of data entry must be balanced with the utility of the data extract file structure, both of which are greatly determined by this process. Therefore, good database design requires the close cooperation and compromise between data management and statistics to ensure quality and efficiency throughout the data processing, management, analysis, and reporting lifecycle.

Many database systems have a catalog of questions or other global library capabilities to facilitate the storage and retrieval of data definition objects. Using these objects as “building blocks,” standard modules can be established and used across many studies.

This standardization saves considerable resources, not only in the study set-up process but also in training data entry personnel, coding validation checks, producing monitor reports, and for analysis/reporting. The continued usage of these standards can also allow for constant streamlining and improvement based on experience, with appropriate maintenance and controls.

Related to the study definition process, most systems require a schedule of events to be defined to instruct the system when to expect certain forms for tracking purposes and to associate date/visit with the form.

Data Quality Specifications

A data quality plan is a tool to aid in the implementation of data quality. The plan should be developed as soon as the protocol is finalized. Data quality is a shared responsibility across all functions. For example, the monitor assures quality by source document verification (SDV), and the clinician reviews listings of individual patient profiles and study “outliers.”

New data and corrections to data are usually processed nightly through a batch validation program in the clinical database. The batch validation program will identify new discrepancies that have appeared since the last execution

^aSAS is a registered trademark of SAS Institute, Inc.

of the validation. The program will also resolve any previously generated discrepancies that are no longer valid because either the data or the associated validation criteria have changed. Batch validation may also be run “on demand” if immediate validation of data is required.

With the help of the study team, data management usually prepares the validation procedures document to identify specific variables that must be validated. Edit checks may be defined as part of the data structure and executed during data entry. Programmed checks are user-defined checks executed off-line during batch validation.

These programmed checks include

- Standard checks developed for all standard CRF pages
- Project specific checks used within each project
- Study specific checks used for study specific pages of the CRF.

The completed validation checks should be run against test data to ensure they are written correctly. As the data is received and validated by these procedures, it is important to review the output and add or delete edit checks as appropriate.

STUDY CONDUCT

Receipt of Data

Technology is providing a number of options for the transmission of data from the investigator, CRO, or lab, back to the sponsor site. Imaging technology allows for the capture and efficient storage of all CRFs for use in data tracking and electronic submissions. Current and near future imaging technology will allow us to store an electronic copy of signed CRFs as well as easily archive all study-related documents. Images can be read using optical character recognition and bar coding techniques. These technologies, once perfected, will greatly reduce the manpower required to index and enter the data on CRFs into the data management systems. Imaging technology is currently being employed to route documents through the appropriate study conduct workflow (Fig. 2).

There have been numerous advances in the area of remote data acquisition. Data can be collected at the site via an electronic CRF or a hand held electronic device. These data can then be transmitted back to the sponsoring company and batch loaded into the sponsor’s clinical trial database. Remote data entry technology currently allows for the easy definition and distribution of the electronic CRFs to the investigator site. Some online cleaning can be performed as the data are entered before transmission to the sponsor site, where additional quality checks are

applied and transmitted back to the investigator site. This iterative process allows for collection and generally faster cleaning of clinical data. Data can also be transmitted from the CRO, investigator, or lab via electronic data transfer. Laboratory data are most often transmitted this way due to the volume of the data. The data are then batch loaded into the sponsor’s clinical trial database. Fax transmissions are often received from the investigator. The fax transmission can be printed out and then data entered, or the fax can go directly to a fax server or be passed through a scanner and an electronic image of the form/document can be created. This image can then be stored, or data entered either by optical character recognition, manual data entry, or a combination of the two.

Many studies are still conducted by traditional paper-based methods. CRFs and documents are sent by post (often overnight) to the sponsor site where they are data entered and filed. Today’s technology allows for the conduct at multiple sites, with the ability to pool data for interim analyses and integrated safety summaries. The size and complexity of a study should determine which technology should be employed. Most large pharmaceutical companies have a portfolio of study conduct technologies to employ.

Data Entry

In a paper-based data flow, as CRFs are received by the sponsor, the pages or forms can be “logged in” or identified to the system. A document number may be used to uniquely identify a page for further tracking within the database. This document identifier can be scanned from a barcode printed on a form, created using a document number generator, or manually entered.

Once the form is recognized as received by the system, a data entry operator can start entry into the database. In the data definition process, screen layouts will have been defined to facilitate the accurate and speedy entry. Some validation or discrepancy checks can be designed to trigger at entry. For example, if a data entry operator attempts to violate the criteria defined to the system for a particular data field (e.g., entering character information into a numerically defined field), a discrepancy can be raised to alert the operator for acceptance or correction to the data. If the entry correctly reflects what is written on the CRF, the value can be accepted and a discrepancy noted for later follow up.

Many systems allow for the option to perform an independent second pass of data entry to ensure that data that is recorded on the CRF matches what is entered into the database. Second pass (double key) should be performed by a different data entry operator than the first pass. In the cases where the first pass and second pass

do not match, the data entry operator is prompted and can accept either entry. An audit trail is kept by the system, and reports may be generated to document changes performed during the second-pass process.

Data entry conventions are recommended to assist in the consistent handling of the data. These conventions should include guidelines and rules for dealing with expected (and unexpected) issues arising on the forms. Some examples include handling missing data, illegible text or data, investigator comments, acceptable abbreviations, etc.

Not all data are received by the sponsor site on CRFs or paper. For example, data may be entered remotely at the investigator site or generated as an output file from instrumentation and then electronically transferred to the sponsor site via the Web/Internet, other connections, or even diskettes. Alternatives to traditional data entry also include using optical character recognition (OCR) technology. This scanning technique used to populate the database may require the CRFs to be designed with special considerations as to the density of the forms, increased use of coded fields, and legibility of the completed forms.

Discrepancy Management

In addition to the discrepancies generated as a result of study definition (univariate discrepancies), discrepancies may also arise when a batch validation detects data inconsistencies (univariate and multivariate discrepancies). Discrepancies are also identified by a visual review of the data, e.g., monitoring lists, SDV review. Discrepancies may also be created by people responsible for data analysis (e.g., statisticians, pharmacoeconomists, clinical pharmacologists). All discrepancies and data fields requiring verification or clarification are tracked using the clinical database.

Quality control for clinical data within data management includes computerized validation of data in the database and second-pass data entry. These activities are performed to ensure that data are complete, accurate, and

compliant with the protocol. In addition to discrepancy reports, verification of randomly selected fields may be used to assess the data quality.

Discrepancy reports are prepared for investigator review and correction. The sponsor translates the computer output into user-friendly reports. There is direct communication between the investigational site coordinator and the data manager for any error messages that may need clarification.

Corrections are made by the site representative directly onto the CRF page and then faxed back to the sponsor. If fax technology is not used, a copy of the corrected CRF page is made and sent to the sponsor by mail or courier. Good clinical practice requires that all corrections must be dated and initialed by the site representative.

Once the corrected copies are received, the data manager makes the change in the clinical database. An electronic audit trail is maintained in the clinical database of all data entered and changed. This audit trail tracks the date and time stamp and the identification of the person making the entry correction or change.

Ongoing Monitoring

A number of query tools may be used to track the quality and completeness of CRF and non-CRF data. Many of the tracking reports reside within the data management system, but tracking may also be done using simple ad hoc query tools such as Brio or even SAS. An example of on-going monitoring is the tracking of study enrollment by investigators (Table 1).

Inclusion and exclusion criteria are usually listed on the CRF. The investigator reviews the criteria and either admits or excludes the subject from continuing in the study. This may be reviewed and monitored manually by the monitor reviewing the subjects' medical records to confirm eligibility during source document verification or through reports/listing of this particular patient data highlighting any irregularities.

Table 1 Patient enrollment statistics, by investigator

Investigator	Inv ID	Site	Total enrolled	Total active	Total completed	Early terms
Smith, Michael	205000_P	15103_P	7	6	1	0
Sole, Thomas	205250_P	22609_P	1	0	1	0
Kay, James	205636_P	22938_P	12	10	2	0
Burn, Alan	205707_P	22997_P	4	4	0	0
Chub, Andrew	205708_P	22998_P	3	3	0	0
Door, Robert	205709_P	22999_P	3	3	0	0
Field, Roy	205713_P	23001_P	4	0	4	0
			34	26	8	0

Database Closure

At study completion, the data manager is responsible for assuring that the data are clean and then prepares to lock the study/close the database. The purpose of locking the study is to ensure that a full audit trail of any changes exists once the study/patients have been unblinded. Database closure marks the end of the study conduct phase and the beginning of the analysis and reporting phase. The standard definition of clean data is:

- All outstanding data in-house
- All outstanding discrepancies resolved
- SDV completed
- Clinical review of data complete
- The allocation of preferred terms to CRF verbatim terms reviewed and complete
- All non-CRF data revised and processed.

The data management system through a series of reports and internal checks provides the documentation and verification that data have been completely cleaned.

When the criteria for clean data are met, a formal sign-off meeting is held for team members and ad hoc functional representatives. With the database closure form signed off, the data manager locks the database. Locking limits the ability to change values for specific privileged users. It also starts a new audit trail of any changes made after locking.

The randomization codes may now be entered allowing the statistician to review the data in an unblinded fashion. Any pharmacokinetic data are also loaded at this time. Data management then freezes the database. No changes may be made to the existing database and no new data may be added.

Study Performance Metrics

Performance metrics are used by the study team to track and manage the study. The metrics will aid in the early

identification and resolution of problems that may affect data quality and study timelines.

For example, metrics involving patient enrollment, visits, forms flow, and discrepancies may be tracked using the clinical database. An example of a tracking report for time from patient visit to receipt in-house is given in Table 2.

Laboratory Data

It is common practice to employ outside laboratories to perform testing for safety and efficacy measures in clinical trials. Along with the results, these laboratories will also provide the units and normal ranges for the tests performed. Since the laboratories are typically utilized by many patients in a study or even across studies, it is practical for the units and ranges to be received and entered once in the system and then linked internally to the patient data to which they apply. This principle of centrally storing values that can be shared across the system is also desirable for maintaining the conversion factors used in deriving lab results into standardized units.

Thesaurus

Medical dictionaries are utilized extensively in clinical trials to assign common terminology to medical events such as adverse events reporting and clinical diagnoses, as well as to link medication trade names to their generic components. Thesaurus management systems facilitate both the ongoing maintenance of base dictionaries (e.g., COSTART, WHOART, MEDDRA) and the linkages to the reported and entered data.

Pharmacokinetic Data

In blinded studies, entering of pharmacokinetic (PK) data on an ongoing basis could jeopardize the blinding of the

Table 2 Weeks from patient visit to login, by investigator

Investigator	Inv ID	Site	Num DCIs logged	Min wks to log	Max wks to log	Avg wks to log
Smith, Michael	205000_P	15103_P	640	0.0	32.0	4.8
Sole, Thomas	205250_P	22609_P	294	1.6	35.9	12.5
Kay, James	205636_P	22938_P	959	0.1	24.1	5.7
Burn, Alan	205707_P	22997_P	731	0.1	28.1	3.7
Chub, Andrew	205708_P	22998_P	1050	0.4	32.7	3.7
Door, Robert	205709_P	22999_P	679	0.1	23.6	1.9
Field, Roy	205713_P	23001_P	637	0.0	36.6	4.8
			4990			

study. Consequently, the PK data are often entered into a separate database. The data is usually loaded into the data management system only after the study is closed and ready for analysis.

ANALYSIS AND REPORTING

Extracting Data

A well-designed data management system typically will focus on the primary objective to facilitate the collection and cleaning of clinical data. Although it must also support analysis and reporting, it is not always possible to achieve an equal balance across all these requirements; therefore, data are usually analyzed outside of the clinical database.

Data extraction is the process of selecting and copying data fields to an external file. Data extraction procedures generally produce files that are simply a reflection of the database. The data can then be manipulated and/or transposed to achieve an optimal structure for analysis and reporting requirements. Additional fields can be derived, response values standardized or decoded, and variables labeled more clearly.

Data may be organized by type of data, such as adverse events, laboratory data, demography, physical exam, etc. Since many of these categories of data exist across clinical trials, standard file structures can be designed and implemented. This standardization allows for the reuse of validated software as well as facilitates the pooling of data across studies for use in project safety summaries and other data reporting across studies.

Derivations

The derivation of data points can be conducted in a number of different ways. Usually they are calculated either in the clinical trials database or as part of the creation of the analysis ready, value added data sets.

It is advisable to store derivations for values that are not likely to change, and for which the derivation algorithm is commonly accepted in the clinical trials database. Derivations that are a result of a constantly changing database, or of a complex algorithm particular to a given study, should be conducted outside the clinical trials database and as part of the creation of the analysis ready, value added data sets.

Reporting Tools

Reporting and analysis is usually a continuous process throughout the life of a study. The “final report” is the

culmination of the efforts involved in conducting a clinical trial.

For ongoing reporting during the life of a study, there are a large number of reporting tools available on the market for the querying of clinical trials data. Each database has a number of tools that are appropriate for creating easy to mildly complex reports against the clinical trials database (i.e., Oracle^{®a} has several reporting tools). There are also a number of user-friendly query tools that are designed to retrieve data from a number of different databases. Brio^{®b} can generate query results that join multiple tables and give quite a bit of flexibility over report format and features such as sorting. More complex reports, such as a “missing and overdue forms report,” are usually written in third generation language (3gl) such as C++, or taking the data outside the database and using external programming tools. For the final statistical listings and tables, SAS is the industry standard. Where reporting is concerned, the tool that best performs the job should be the one selected.

Electronic Submission of CRFs to Regulatory Agencies

Sponsors may be required to provide selected CRFs as part of the overall package submitted to the regulatory authorities. Recently, regulatory agencies have been encouraging the electronic submission of these CRFs. Given a comprehensive and well-indexed imaging system, it may be possible to subset the requested images and electronically transfer the file with relative ease. For the situations where these files must be manually compiled, a different process may be employed. As CRFs are identified for inclusion, a scanner can be used to produce .pdf files (via Adobe Acrobat Exchange^{TMc}). An index is required to facilitate the retrieval of the forms, as desired. The collection of indexed images is then transferred onto CD-ROM for the electronic submission.

SYSTEM ISSUES

Year 2000

The approach of 2000 A.D. had caused the technology industry to take an in-depth look at all of the automated solutions employed in the industry. Every place where a date was used in an application was examined. Two digit dates were particularly troublesome as we approached the new millennium. The validation effort consumed an

^aOracle is a registered trademark of Oracle Corporation.

^bBrio is a registered trademark of Brio Technology, Inc.

^cAdobe Acrobat Exchange is a trademark of Adobe Systems, Inc.

enormous amount of resources, both in-house and at the software vendors.

Upgrades

Whether your clinical trials management system was developed in-house or purchased from a vendor, eventually you will have the opportunity to experience an upgrade. At some point you will probably need to upgrade the operating system on the PC or server, the version of the database that your application is built on, or the application software itself. Worst case is when you have to upgrade all of these at once.

Ideally your application environment consists of a fully functional and separate test environment. It is in this area that you would test any upgrades. Testing should consist of executing documented test scripts with the goal of proving that existing functionality still works and any advertised new functionality also works.

Ideally you would try to avoid upgrading multiple pieces of your environment at the same time, as in the worst case example above. Although multiple rounds of testing is resource intensive, it is much easier to determine the source of any problems and resolve them in a controlled environment. This is a point to be aware of when choosing clinical trial software: Will the vendor support multiple versions of an operating system and database? This will give you the time to test the worst case scenario in a two-phase approach.

Conversion vs. Migration vs. Upgrade

Over time software becomes obsolete, as does hardware. Upgrading to the latest version of the software or hardware is probably the easiest path. But when you find that you must move to a completely new hardware or application environment, there are several things to consider. Often software and hardware vendors can provide the service of migrating or converting your existing data from one system to another. One should carefully investigate what this process would entail. Conversion can be a painful and extremely resource intensive operation. You should realistically look at whether it is feasible to let ongoing trials complete in the legacy system or whether it is feasible to re-enter data into the new system for smaller studies. These strategies are often much more straightforward and less error prone than a conversion would be.

System Validation

Computer systems validation (CSV) is an ongoing process that involves the evaluation and documentation of all

components of a system during its life cycle to ensure compliance with approved user requirements and quality standards. A system is defined not only by its hardware and software, but also by the processes surrounding its use. CSV is applicable to a system used to collect, process, capture, or manipulate data that may be included in a submission to a regulatory authority. Validation requires establishing documented evidence that a system meets its predefined specifications and quality attributes. Validation seeks to assure that a system has been developed, tested, and implemented in a controlled manner, performs and will continue to perform accurately and reliably, and is secure from unauthorized or accidental change. In addition to documenting the development and implementation of system components, validation includes documenting hardware and software change control, security management, and training.

Audit Trails and Change Control

It is important to be able to track the reason and source for any changes to data in your clinical trials database. Many applications have built in audit trail capabilities that track the date, time, and ID of the person entering or changing data through the application. Some applications will even prompt for a data change reason. Any changes or deletion of data should be done through the application whenever possible. Sometimes however, the volume of the data to be modified or complexity of the changes requires external intervention. If you plan to modify data in the clinical trials database from outside of the application, the process should be very carefully documented. As in any software development, the program or script that will be run to enter or update data should have a design document, that outlines the modules purpose and expected performance, as well as a set of fully executed test cases. It is common to keep all requests for manual data changes and data change scripts with their respective documents in one directory as backup for a data change log.

BIBLIOGRAPHY

- Barkat, Mohammed. Quality Assurance/Quality Control Issues in GMP Regulatory Compliance. *Drug Inf. J.* **1997**, 31, 765–769.
- Oracle Clinical Corp. *Oracle Clinical User Reference*; Oracle Clinical Corp.: Redwood Shores, CA, 1–10.
- Shapiro, S.H. *Clinical Trials: Issues and Approaches*; Marcel Dekker, Inc.: New York, 1983.
- Spilker, B. *Guide to Clinical Trials*; Raven Press: New York, 1991.

CHROMATOGRAPHIC METHODS OF ANALYSIS— THIN LAYER CHROMATOGRAPHY

Joseph Sherma

Lafayette College, Easton, Pennsylvania, U.S.A.

INTRODUCTION

Thin layer chromatography (TLC) consists of the sample solution being applied as a spot or band on the origin of a layer spread on a support (the plate). After evaporation of the sample solvent, the plate is placed in a sealed chamber or tank that contains a solvent mixture chosen as the mobile phase. Development occurs as the mobile phase moves up the layer by capillary forces. Instrumental development methods, such as overpressured layer chromatography (OPLC) or automated multiple development (AMD), can provide separations with increased resolution. The plate is removed from the chamber, and the separated zones are detected by physical or chemical methods, identified by comparison of their R_f values (R_f = distance of migration of the sample zone/distance of the mobile phase front) and colors to standard zones on the same plate, and quantified by visual or instrumental densitometry based on measurement of zone sizes and intensities. Zone identification is confirmed by off- or on-line coupling of TLC with visible/ultraviolet (UV), Fourier transform infrared (FTIR), Raman, and mass spectrometry (MS). Pharmaceutical applications of TLC include analysis of starting raw materials, intermediates, pharmaceutical raw materials, formulated products, and drugs and their metabolites in biological media (1).

TLC is a flexible, versatile, and economical process in which the various stages (Fig. 1) are carried out independently. The advantages of this off-line arrangement as compared with an on-line process, such as column high performance liquid chromatography (HPLC), have been outlined (2) and include the following:

1. Availability of a great range of stationary phases with unique selectivities for mixture components
2. Ability to choose solvents for the mobile phase is not restricted by low UV transparency or the need for ultra-high purity
3. Repetition of densitometric evaluation can be achieved under different conditions without repeating the chromatography in order to optimize quantification since all sample fractions are stored on the plate

4. High sample throughput since many samples can be chromatographed simultaneously (3)
5. Minimal cost of solvent purchase and disposal since the required amount of mobile phase per sample is small
6. Accuracy and precision of quantification is high because samples and standards are chromatographed and measured under the same conditions on a single TLC plate
7. Sensitivity limits of analysis are typically at nanogram (ng) to picogram (pg) levels.

Comparative studies have often found that high performance TLC (HPTLC) is superior to HPLC in terms of total cost and time required for pharmaceutical analyses (4). The Bibliography contains sources of general information on the principles, theory, practice, instrumentation, and applications of TLC and HPTLC. Detailed information on the subjects mentioned above as well as on additional topics and applications that could not be covered because of lack of space will be found in these references.

EXPERIMENTAL PROCEDURES

Sample Preparation

Sample extraction and cleanup procedures for TLC are similar to those for gas chromatography (GC) and HPLC. If the analyte concentration is sufficiently high, pharmaceutical dosage forms can often be simply dissolved in a solvent that will completely solubilize the analyte and leave excipients or extraneous compounds undissolved to yield a test solution that can be directly spotted for TLC analysis. Grinding of the sample and application of heat and/or sonication may be required to assure solubility of the analyte, as well as filtration or centrifugation to remove undissolved excipients. If the analyte is present in low concentration in a complex sample, solvent extraction, cleanup (purification), and concentration procedures usually precede TLC in order to maximize the analyte and minimize interfering extraneous components in the test solution. Since layers are not reused, it is often

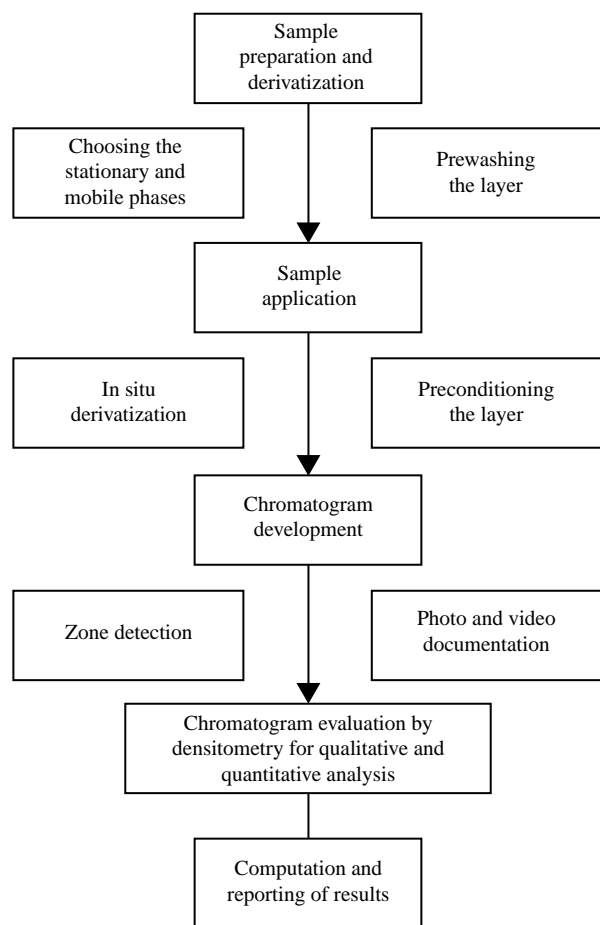


Fig. 1 Schematic diagram of the steps in a thin layer chromatographic analysis. (Adapted from Camag Scientific, Inc.)

possible to apply cruder samples than could be injected into a GC or HPLC column, including samples with irreversibly sorbed impurities.

Traditional procedures that are still widely used include liquid–liquid partitioning, column chromatography, desalting, and deproteinization (5), but the newer microwave extraction, solid phase extraction (SPE) (6), and supercritical fluid extraction (SFE) methods are being increasingly applied for isolation and cleanup of samples prior to TLC. Special plates with pre-adsorbent or concentrating zone composed of adsorption-inactive kieselguhr or silicon dioxide may provide sample cleanup by retaining some interfering substances (7).

Assay protocols for samples, such as tablets, usually involve taking multiple tablets (e.g., 10–20), grinding and mixing thoroughly, and weighing a test sample equivalent to one tablet, rather than analyzing only one tablet, so that the sample will be more representative of the batch being sampled.

Stationary Phases

Commercial precoated layers on glass support are used in virtually all analyses. HPTLC uses plates that are smaller (10×10 or 10×20 cm), have a thinner (0.1–0.2 mm) layer composed of sorbent with a finer mean particle size (5–6 μm) and are developed over shorter distances (ca. 3–7 cm), as compared to classical 20×20 cm TLC plates which are generally 20×20 cm with a 0.25-mm layer and developed for 10–12 cm. HP plates provide improved resolution, shorter analysis time, higher detection sensitivity, and improved in situ quantification and are used for industrial pharmaceutical densitometric quantitative analyses. TLC plates are usually used for qualitative identification and purity studies as contained in pharmacopeias.

Normal phase adsorption TLC on silica gel with a less polar mobile phase, such as chloroform-methanol, has been used for more than 90% of reported analyses of pharmaceuticals and drugs. Lipophilic C-18 (8), C-8, C-2; phenyl chemically-modified silica gel phases; and hydrocarbon-impregnated silica gel plates (9) developed with a more polar aqueous mobile phase, such as methanol–water or dioxane–water (10) are used for reversed phase TLC. Other precoated layers that are used include aluminum oxide (11), magnesium silicate, magnesium oxide, polyamide (12), cellulose, kieselguhr, ion exchangers (e.g., PEI cellulose anion exchanger), and polar modified silica gel layers that contain bonded amino (13), cyano, diol, and thiol groups. The polar bonded phases, in which the functional groups are bonded to silica gel by means of a hydrophobic spacer (e.g., *n*-propyl), can function with multimodal mechanisms, depending on the composition of the mobile phase. Silica gel can be impregnated with various reagents to improve separations (e.g., EDTA for tetracycline analysis) (14).

Optical isomer separations that are carried out on a chiral layer produced from C-18 modified silica gel impregnated with a Cu(II) salt and an optically active enantiomerically pure hydroxyproline derivative (15), on a silica layer impregnated with a chiral selector such as brucine (16), on molecularly imprinted polymers of α -agonists (17), or on cellulose with mobile phases having added chiral selectors such as cyclodextrins (18) have been reported mostly for amino acids and their derivatives. Mixtures of sorbents have been used to prepare layers with special selectivity properties.

Layers are often cleaned by predevelopment with the mobile phase or methylene-chloride–methanol (1:1) or immersion in methanol (19) prior to sample application, especially for quantitative TLC.

Mobile Phases

The mobile phase for a particular separation is usually selected empirically using prior personal experience and literature reports of similar separations as a guide. In addition, various computer-assisted mobile phase optimization schemes (20) have been described for selecting the mobile phase components and their relative concentrations, most notably the PRISMA model.

General mobile phases systems that are used based on their diverse selectivity properties are diethyl ether, methylene chloride, and chloroform combined individually or together with hexane as the strength-adjusting solvent for normal-phase TLC, and methanol, acetonitrile, and tetrahydrofuran mixed with water for strength adjustment in reversed phase TLC. Separations by ion pairing on C-18 layers are done with a mobile phase such as methanol–0.1 M acetate buffer (pH 3.5) containing 25 mM sodium pentanesulfonate (15.5:4.5) (21). Specific mobile phases for pharmaceutical and drug analysis are listed in the Applications of TLC in Pharmaceutical and Drug Analysis Section.

Application of Samples

The method used for application of sample solutions is determined by whether HPTLC, TLC, or preparative layer chromatography (PLC) and qualitative or quantitative analysis are being performed. Sample volumes of 0.5–5 μl for TLC and 0.1–1 μl for HPTLC are applied manually to the layer origin as spots using fixed volume glass micropipets, such as Drummond Microcaps or selectable volume 10 or 25 μl digital microdispensers. In addition, many manual and automated instruments are available for sample application, especially for quantitative HPTLC.

The partially automated Linomat IV (Fig. 2) can apply 2–99 μl volumes to HPTLC plates (5–490 μl for PLC) as bands of controlled length [1 mm (spot) to 190 mm] by spraying from a glass syringe. This instrument has been used more than any other for densitometric quantification in pharmaceutical analysis. Compact bands are also produced when 1–25 μl samples are applied manually with a digital microdispenser to plates that contain a preadsorbent zone. Band application is advantageous for obtaining the highest resolution separations and precise quantitative results by scanning densitometry.

A fully automated, personal computer-controlled spotter (e.g., the Camag Automatic TLC Sampler III), which consists of a stainless steel capillary connected to a dosage syringe operated by a stepper motor, can sequentially apply constant or variable volume samples, chosen from a



Fig. 2 Camag Linomat IV Band Applicator. (Photo courtesy of Camag Scientific, Inc.)

rack of vials, within the range of 10 nl to 50 μl as spots or bands.

Chromatogram Development

In addition to the stationary and mobile phases, separations obtained in TLC are affected by the vapor phase, which depends on the type, size, and saturation condition of the chamber during development. The interactions of these three phases as well as other factors, such as temperature and relative humidity, must be controlled to obtain reproducible TLC separations. The development process with a single (isocratic) mobile phase is complicated because of progressive equilibration between the layer and mobile phase and separation of the solvent components of the mobile phase as a result of differential interactions with the layer, which leads to the formation of an undefined but reproducible mobile phase gradient.

The important development methods in pharmaceutical and drug analysis include classical linear ascending development, horizontal development (22), gradient TLC with AMD, OPLC, and two-dimensional (2D) development. These development methods will be described briefly. Other development methods, such as circular, anticircular, continuous, and rotational, will not be covered.

In the classical method of linear, ascending development TLC and HPTLC, the mobile phase is contained in a large volume, covered glass chamber (N-chamber). The spotted plate is inclined against an inside wall of the tank with its lower edge immersed in the developing solvent below the starting line. The solvent begins to rise

immediately through the initial zones due to capillary flow. The space inside the tank is more or less equilibrated (saturated) with solvent vapors, depending on the presence or absence of a paper liner and the period of time the tank is allowed to stand before the plate is inserted. Unsaturated chambers can provide different, often higher resolution, separations as compared to saturated chambers with the same mobile phase. The twin trough chamber (Camag) (23) is an N-chamber modified with an inverted V-shaped ridge on the bottom that divides the tank into two sections. These divisions allow development with low volumes of mobile phase in one and easy pre-equilibration of the layer with vapors of the mobile phase or another conditioning liquid (e.g., a sulfuric acid-water mixture to control humidity) or volatile reagent in the other. A computer-controlled automatic developing chamber (Camag) offers programmable, reproducible linear ascending development without operator attention.

The use of a horizontal developing chamber (Camag) permits simultaneous development from both ends to the center of up to 70 samples on a 20 × 10 cm HPTLC plate, or 35 samples from one end to the other. The developing solvent, held in narrow troughs, is carried to the layer through capillary slits formed between the trough walls and glass slides. The chamber is covered with a glass plate during pre-equilibration and development, and it can be operated with controlled levels of vapor saturation and relative humidity.

Unidimensional multiple development, in which the layer is developed repeatedly for the same distance with the same solvent system or two different systems (24), is a manual method for improving the resolution of zones that migrate in the lower half of the layer (13). Multiple development has been improved by the use of AMD instruments (25), one of which is shown in Fig. 3. AMD generally involves 1–25 individual linear ascending developments of an HPTLC plate performed with a mobile phase gradient of decreasing strength (i.e., decreasing polarity for silica gel) over distances that increase by 3–5 mm for each stage. The layer is dried under vacuum and conditioned with the vapor phase of the next batch of fresh solvent before each incremental run. The repeated movement of the solvent front through the chromatographic zones causes compression into narrow bands (width about 1 mm) during AMD, leading to a spot capacity (the number of zones that can be completely separated in the available layer distance) of more than 50 for an 80-mm run. Typical “universal gradients” are produced from the solvents methanol or acetonitrile (polar), methylene chloride, diisopropyl ether, or *t*-butylmethyl ether (medium polarity), and hexane (nonpolar). Mixtures containing compounds with widely

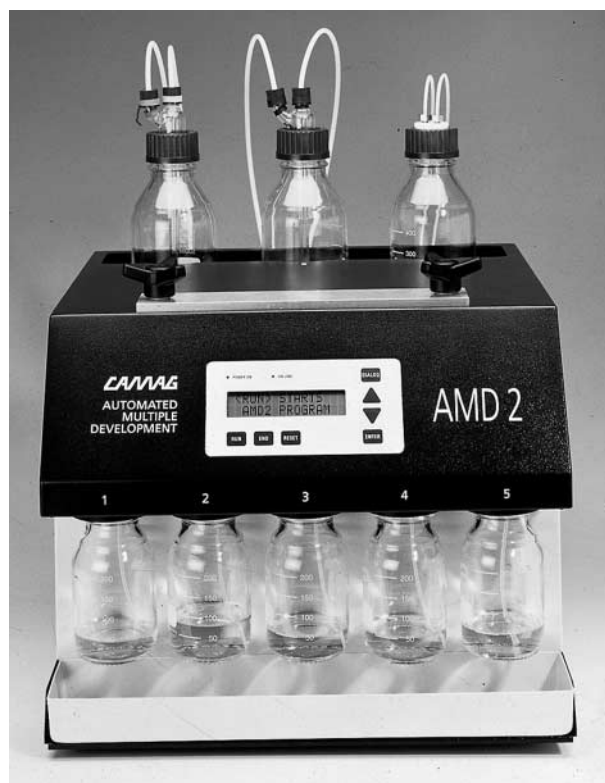


Fig. 3 Automated multiple development instrument (AMD 2). (Photo courtesy of Camag Scientific, Inc.)

different polarities can be separated by AMD on one chromatogram, and migration distances of individual components are largely independent of the sample matrix. AMD is the most promising development method available for modern TLC and will certainly receive increased use in the future for pharmaceutical analysis.

OPLC is a method in which the mobile phase is pumped through a layer that is sandwiched between a rigid support block and a flexible plastic membrane under external pressure of 10 or 25 atm (Chrompres 10 and 25). Mobile phase flows through the layer at a constant linear flow rate velocity in the range of 1–12 ml/min, leading to higher separation efficiency than is possible with capillary flow where the mobile phase velocity is variable. To carry out linear chromatography, the layer must be specially prepared by scraping the edges and treating with a polymer sealant to eliminate leaks during development, and by cutting mobile phase inlet and solvent outlet channels at appropriate positions. A newer OPLC instrument (BS 50) provides linear isocratic or three-step gradient development, on-line or off-line modes, analytical or preparative separations, and ready-to-use presealed 20 × 20 cm or 10 × 20 cm silica gel or C-18 TLC and HPTLC layers. Of all TLC methods, OPLC most closely

simulates HPLC. Determination of deramciciane by HPTLC-OPLC (26) and forensic and clinical drug screening (27) are examples of applications. In addition to the usual elution-type development, forced flow displacement TLC was reported for pharmaceutical densitometric analysis (28).

In 2D TLC, the sample mixture is applied to one corner of the TLC plate, which is developed with the first mobile phase, dried, and developed with a second mobile phase that provides different selectivity in a perpendicular direction. Greatly increased spot capacities of 250–400 for capillary flow development and 500–2000 for forced flow have been reported for 2D TLC. The analyses of amphetamine derivatives (13) and thyreostatic drugs (29) by 2D TLC were reported.

Zone Visualization (Detection)

After removal of the mobile phase from the developed plate by heating, zones are detected on the layer by their natural color, natural fluorescence, quenching of fluorescence, or as colored, UV-absorbing, or fluorescent zones after reaction with a reagent (postchromatographic derivatization). Zones with fluorescence or quench fluorescence are viewed in cabinets that incorporate shortwave (254 nm) and longwave (366 nm) UV lamps.

Fluorescence quenching occurs on an “F-layer” that contains a fluorescent indicator or phosphor. Compounds that absorb 254 nm UV light, particularly those with aromatic rings and conjugated double bonds, appear as dark violet spots on a green or pale blue background because the absorbing compounds diminish (quench) the uniform layer fluorescence. Many drugs quench fluorescence, and this method is very widely used for detection and quantification by scanning (30).

Universal or selective chromogenic (dyeing) and fluorogenic detection reagents are applied by spraying onto the layer, dipping the layer into the reagent, exposing the layer to reagent vapors (e.g., iodine), incorporating the reagent in the mobile phase or in the layer, or by pressing an adsorbent polymeric pad soaked with the reagent against the layer (overpressure derivatization). Examples of universal reagents that react with many compound classes are vanillin-sulfuric acid (31), anisaldehyde (32), and iodine, while ninhydrin is a selective reagent for detection of amino acids (33) and Dragendorff reagent is widely used to detect alkaloids (34). Although spray application is most widely used, dipping is more reproducible, especially when carried out in a mechanized chromatogram immersion instrument (Camag). Layers must frequently be heated in an oven or on a flat heating plate after applying the detection reagent in order to accelerate the reaction upon which

detection is based. Biological-physiological methods of detection, such as bioautography, are also employed for medicinal compounds (35).

An important advantage of the off-line operation of TLC is the flexibility afforded by the use of multiple methods for zone detection and identification. For example, the layer can be viewed under long- and short-wave UV light, followed by one or more chromogenic, fluorogenic, or biological detection methods.

Many hundreds of reagents and detection methods have been described in various literature sources (36–38).

Documentation of Chromatograms

TLC separations can be documented by photography or video recording (39). Commercial documentation systems that incorporate standard and instant photographic cameras and charge coupled device (CCD) video cameras are suitable for chromatograms with colored, fluorescent, and quenched zones. The latest approach for copying TLC plates is by use of computer imaging, and a system that incorporates a computer, scanner, and black-and-white or color printer was described (40). Digital cameras (e.g., Casio QV 200) are widely used for photographing TLC plates, but their use for quality control purposes is in question because of the potential for manipulating the file with software.

Quantitative Analysis

Quantification of thin layer chromatograms can be performed indirectly after scraping off the separated zones of samples and standards, and elution of the substances from the layer material with an appropriate solvent. The volumes of the eluates are adjusted and the solutions analyzed by use of a spectrometric method, GC, or HPLC. Scraping and elution are usually performed manually. Although direct quantification has become increasingly important, indirect analysis is still widely used (e.g., for assay of some drugs according to the USP).

Direct quantification is carried out *in situ* rather than after spot elution. The simplest direct method involves visual comparison of sample zone size and/or intensity (color) variation according to concentration against reference standards developed on the same plate (11). This qualitative/semiquantitative approach is specified in various pharmacopeias for the purity analysis of active raw materials and formulated products. These pharmacopeial methods are designed for analyses at several levels: 1) simple detection of impurities as additional spots; 2) detection and identification of impurities by comparison to the R_f values distances of standards (41); or 3) detection,

identification, and estimation of amounts of impurities by comparing intensities between samples and standard dilutions of the same compounds (42).

Most modern HPTLC quantitative analyses are performed in situ by measuring the zones of samples and standards using a chromatogram spectrophotometer (usually called a densitometer or scanner) with a fixed sample light beam in the form of a rectangular slit. A schematic diagram of single beam densitometer arranged for linear reflection (most used) or transmission scanning is shown in Fig. 4. A tungsten or halogen lamp is used as the source for scanning colored zones (visible absorption) and a deuterium lamp is used for scanning UV-absorbing zones directly or as quenched zones on F-layers. Of all possible densitometric modes, most quantitative pharmaceutical assays have been carried out by UV absorption scanning of fluorescence-quenched zones. The monochromator is a prism or, more often, a grating, and the detector is a photomultiplier or photodiode. For normal fluorescence scanning, a high intensity xenon or mercury lamp would be used as the source and a cutoff filter would be placed between the plate and detector to block the exciting UV radiation and transmit the visible emitted fluorescence. Zigzag, dual wavelength reflection scanning with a point source also has been used for pharmaceutical analysis (43).

Many modern scanners have a computer-controlled motor-driven monochromator that allows automatic

recording of in situ absorption and fluorescence excitation spectra. These spectra can aid compound identification by comparison of unknown spectra with stored standard spectra obtained under identical conditions or spectra of standards measured on the same plate. The spectral maximum determined from the in situ absorption spectrum is usually the optimal wavelength for scanning standard and sample areas for quantitative analysis. The densitometer is usually connected to a computer with software designed specifically for data handling and automation of the scanning process in modern instruments. With a fully automated system, the computer can perform the following: 1) data acquisition; 2) automated peak searching and optimization of scanning for each fraction located; 3) multiple wavelength scanning; 4) baseline location and correction; 5) computation of peak areas and/or heights of samples and co-developed standards, calculation of calibration curves by linear or polynomial (44) regression, interpolation of sample concentrations, statistical analysis of reproducibility, and presentation of a complete analysis report; and 6) storage of data on disk. In general, external standardization is employed for quantification with a calibration curve generated from a series of standards that covers the full concentration range of the analysis. Although the internal standard method is sometimes used (45), it is not normally needed, unless losses during the sample preparation steps are anticipated, since samples and standards are run on the same plate under essentially identical conditions.

Image processing with a video densitometer is an alternative for colored or fluorescence-quenched zones to the use of a slit-scanning densitometer (8). A video scanner consists of a transilluminator system for totally lighting the plate, CCD camera, computer and printer, and chromatogram evaluation software. Video densitometers cannot measure UV-absorbing zones or fluorescent zones directly; cannot scan a layer with uniform, monochromatic light of selectable wavelength; and are not as accurate, precise, or sensitive as slit scanning densitometers in their present state of development (46). One of their advantages as compared to slit-scanning densitometers is for quantification of 2D chromatograms (47). TLC with video technology is being used successfully for fingerprint analysis of herbal supplements and medicines.

Special Techniques

Transfer of mobile phases to HPLC

An important application of TLC is to serve as a pilot method for HPLC, the most widely used analytical method for pharmaceutical analysis. If the stationary phases are

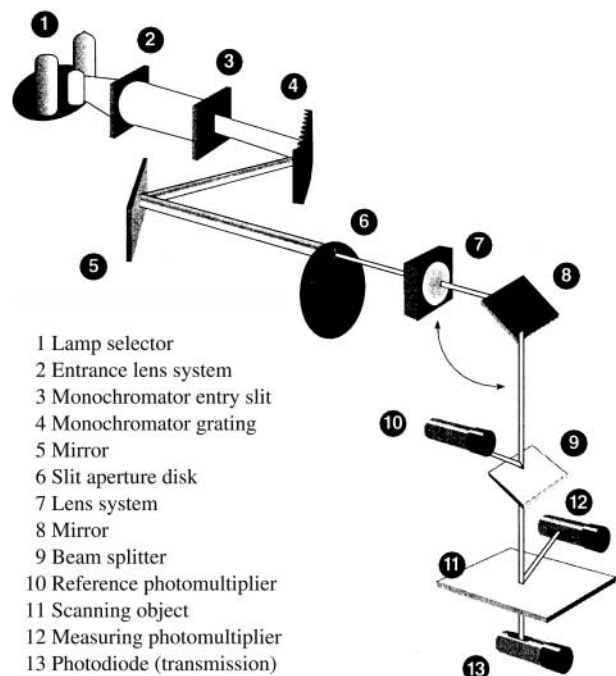


Fig. 4 Light path of TLC Scanner 3. (Diagram courtesy of Camag Scientific, Inc.)

similar, TLC can predict solute retention behavior and suitability of a particular mobile phase through correlation of $\log k'$ in HPLC and R_f data in TLC. Particularly useful is detection of compounds that migrate minimally in the mobile phase and can contaminate the HPLC column during subsequent runs.

Determination of lipophilicity

The determination of the lipophilicity of drugs is extremely important because the biological activity of a molecule can generally be correlated with its ability to penetrate the different hydrophobic barriers (membranes) (i.e., with its lipophilicity or hydrophobicity). One of the best ways to determine lipophilicity is by measurement of retention characteristics [R_M values, $R_M = (1 - R_f - 1)$] of the compounds of interest by reversed phase TLC on a silica gel layer impregnated with paraffin oil or a C-18 chemically bonded silica gel layer. The study of anti-inflammatory drugs is an example of an application (48). The TLC determination of lipophilicity and other molecular parameters was reviewed (49).

Preparative layer chromatography

Analytical TLC differs from PLC in that larger weights and volumes of samples are applied as a band across the entire layer width to thicker (0.5–2 mm) and sometimes larger layers, the purpose of which is the isolation and purification of 10–1000 mg of sample for further analysis. Multiple development of the plate is commonly used (50), and the separated substances are detected by a nondestructive method (e.g., under UV light and iodine vapors), and recovered by extraction from scraped layer material. PLC can be used to isolate sufficient pure drug compounds for confirmation by spectrometry in cases where analytical TLC is not adequate for identification. Examples of pharmaceutical applications of PLC include a new sesquiterpene trimer (51) and phenylpropanoid glycosides (52).

Combined TLC–spectrometry methods

Compound identification is initially made by comparing sample and standard zones based on R_f values and colors produced by selective detection reagents. Identification is confirmed by off- or on-line combination of TLC and spectrometric methods such as visible/UV, fluorescence, FT-Raman, FTIR (53, 54), solid state (NMR) (55), MS (56), and MS–MS. Coupled TLC-FTIR (57) and other combined methods (58) have been reviewed.

HPTLC coupled on-line with spectrometric methods has been proposed as a reference method in clinical

chemistry for identification prior to quantitative analysis. This is particularly important for unequivocal diagnosis as the basis for further clinical therapeutic measures (59).

Thin layer radiochromatography

Location and quantification of radioisotope-labeled substances on a thin layer requires the use of autoradiography, zonal analysis with scintillation counting (60), or direct scanning with a digital autoradiograph (26) or a bio-imaging analyzer (61). Thin layer radiochromatography is widely used for metabolism studies of pesticides and drugs in plant, animal, and human samples, and in quality control and development of radio-pharmaceuticals (62).

Method Validation

Validation procedures are performed according to the recommendations of regulatory agencies, such as the Committee for Proprietary Medicinal Products (CPMP) of the European Economic Community (EEC) and with consideration for the special features of the TLC procedure. The following validation parameters are typically monitored: 1) selectivity; 2) stability before, during, and after TLC development; 3) linearity of the calibration graph; 4) range of levels within which the analyte can be quantified; 5) limits of detection and accurate and precise quantification; 6) accuracy (indication of systematic errors), precision (indication of random errors), sensitivity (ability to measure small variations in concentration), and ruggedness (results of the method when used by different analysts in a variety of locations). Each step of the analysis must be validated through error analysis and a suitability test, and includes sample preparation, application of samples, TLC separation, detection procedures, and quantification. Definitions, general principles, and practical approaches for validation of pharmaceutical TLC analysis are described by Szepesi and Nyiredy (42). Benchmarking studies of the assay and purity testing of phospholipids by silica gel HPTLC with copper (II) sulfate-phosphoric acid detection reagent and scanning of the brown-violet zones at 365 nm showed that HPTLC provided a cost reduction of 1:2.5 as compared to HPLC (63).

Applications of TLC in Pharmaceutical and Drug Analysis

Gas chromatography, HPLC, and TLC are complimentary methods with their own advantages and disadvantages for pharmaceutical and drug analysis. Regular reviews of the

TLC literature (64) have shown that applications to pharmaceuticals and drugs are more prevalent than for any other class of compounds.

Applications of TLC include analysis of the following sample types (65):

- Starting raw materials (plant extracts, extracts of animal origin, fermentation mixtures)
- Intermediates (crude products, reaction mixtures, mother liquors and secondary products)
- Pharmaceutical raw materials [identification, purity testing, assay, separation of closely related compounds, stability testing (10)]
- Formulated products [identification, purity testing, assay, stability testing under storage and stress, content uniformity test (66), dissolution test]
- Drugs and their metabolites in biological media such as urine, plasma, or gastric fluid [pharmacological, toxicological (67), pharmacokinetic (68), metabolic, bioequivalence, forensic, and compliance and pharmacodynamic studies].

TLC analyses are performed in a wide variety of laboratories, including government, pharmaceutical manufacturer, hospital, police, and contract testing laboratories dealing with illicit drug detection in sports, horse racing, and employment screening. For identification of known compounds and impurities, purity testing, and obtaining impurity profiles in bulk raw materials and formulations, R_f values and spot sizes/intensities between samples and reference materials developed on the same plate are compared by noninstrumental or instrumental methods using TLC systems that can separate compounds from different classes or closely related compounds within a single class. Confirmation of identity often requires use of an on-line or off-line ancillary method, such as IR, NMR, or mass spectrometry, GC, or HPLC.

The following eight general, standardized TLC systems are recommended for the analysis of drugs:

For basic drugs: Silica gel layer dipped in 0.1 M KOH and dried; mobile phases: 1) methanol–ammonia (100:1.5), 2) cyclohexane–toluene–diethylamine (75:15:10), 3) chloroform–methanol (9:1), 4) acetone.

For acidic and neutral drugs: Silica gel layer; mobile phases: 1) chloroform–methanol (4:1), 2) ethyl acetate–methanol–ammonia (85:10:5), 3) ethyl acetate, 4) chloroform–methanol (9:1).

Migration data for many drugs in these systems have been tabulated (69). Retention data for 443 drugs were reported for four other standardized silica gel systems: 1) ethyl acetate–methanol–30% ammonia (85:10:15); 2) cyclohexane–toluene–diethylamine

(65:25:10); 3) ethyl acetate–chloroform (1:1); and 4) acetone. The plate was dipped in KOH solution before development with acetone (70). The following screening system (UniTox) employs three mobile phases for normal and reversed phase TLC:

For acidic and neutral drugs: 1) methanol–water (65:35), C-18 silica gel.

For basic, amphoteric, and quaternary drugs: 1) toluene–acetone–ethanol–conc. ammonia (45:45:7:3), silica gel, and 2) methanol–water–conc. HCl (50:50:1), C-18 silica gel (71).

The USP 24/NF 19, section 201 (72, p. 1856), contains a general TLC identification test that involves a nonhigh-performance silica gel layer with fluorescent indicator, chloroform–methanol–water (180:15:1) mobile phase, and detection under 254 nm UV light for verification of the identities of compendial drugs in dosage form test solutions prepared according to the individual monographs. Section 621 of the USP 24/NF 19 (72, pp. 1916–1917) presents brief information on the equipment and procedures of classical ascending and continuous development TLC.

Analytical information and data were presented on the TLC analysis of the most widely prescribed human and animal drugs as well as illicit drugs (69). In addition, applications of quantitative TLC in pharmaceutical analysis were described (73, 74). Biennial reviews of TLC typically contain more than 75 references that describe applications to drug and pharmaceutical analysis (64).

The following are brief descriptions of TLC analyses of drugs in pharmaceutical dosage forms and biological samples that were selected as typical examples.

Analysis of Biological Fluids Using Visual Zone Comparison

Clenbuterol and salbutamol residues in animal urine (75)

1. *Sample preparation.* Solid phase extraction on C-18 column, elution with 0.1% triethylamine in methanol.
2. *TLC.* Silica gel 60 layer with concentrating zone, ethyl acetate–methanol–acetic acid (8:1:1) mobile phase.
3. *Detection.* *N*-chlorination with chlorine vapors and detection of the *N*-chloro derivatives as blue spots with iodide–*o*-tolidine solution.
4. *Qualitative screening and semiquantitative analysis.* Based on R_f value and color brightness comparison between samples and standards.

Analysis of Biological Fluids Using Fluorescence Densitometry

Cortisol in plasma and urine (76)

1. *Sample preparation.* Extraction with dichloromethane.
2. *TLC.* Extracts and standards applied in 3–6 mm bands with a Linomat IV to an aluminum-backed silica gel 60 layer, chloroform-methanol (9:1) mobile phase.
3. *Detection.* Layer dipped into isonicotinic acid hydrazide reagent for 20 s and then into chloroform-liquid paraffin (9:1) to enhance and stabilize fluorescence.
4. *Quantification.* Scanning with 366 nm excitation and 460 nm emission wavelengths.
5. *Validation.* Limit of detection 1 ng, RSD 1.4–6.3%, comparison of results to a TLC-radioimmunoassay method gave correlation coefficients of 0.97 and 0.98.

Sulfamethazine in pork tissue (77)

1. *Sample preparation.* Sulfabromomethazine added to tissue as an internal standard, extraction with water, centrifugation, and cleanup and concentration by a series of solid phase extractions using C-18 bonded silica, acidic alumina, and AG MP-1 anion exchange microcolumns.
2. *TLC.* Samples and standards applied in 6 mm bands with a Linomat IV to a silica gel 60 layer, ethyl acetate-toluene (1:1) mobile phase.
3. *Detection.* Layer dipped into fluorecamine solution.
4. *Quantification.* Fluorescence scanning of analyte and internal standard zones at 366 nm or 400 nm (Fig. 5).
5. *Validation.* Limit of detection 0.25 ppb, average recovery over analysis range (0.54–2.18 ppb) was 95.6% (standard deviation 29.4%, $n = 54$).

Analysis of Pharmaceutical Preparations Using Visual Zone Comparison

Purity test for allylestrenol bulk drug substance and tablets (78)

1. *Sample preparation.* Drug substance was dissolved in chloroform; tablets were powdered and sonicated in acetone.
2. *TLC.* 2.5 and 5 μ l aliquots of sample and allylestrenol and impurity standard solutions were manually applied in 8 mm bands to HPTLC silica gel plates, which were developed by OPLC with cyclohexane-butyl acetate-chloroform (90:12:2) mobile phase.
3. *Detection.* Spray with 10% ethanolic sulfuric acid and heat at 120°C for 2 min.

4. *Assay.* Visual comparison of sample and standard zones under 366 nm UV light.

Analysis of Pharmaceutical Preparations Using Visible Densitometry

S-Carboxymethylcysteine in syrups used to treat respiratory diseases (33)

1. *Sample preparation.* Syrup diluted with 96% alcohol-ammonia (4:1).
2. *TLC.* Samples and standards applied with a Nanomat III to a silica gel 60, 1-butanol-glacial acetic acid-water (3:1:1) mobile phase, development in a twin-trough chamber.
3. *Detection.* Plate dipped into ninhydrin reagent and heated for 3–4 min at 100°C.
4. *Quantification.* Zones scanned at 487 nm.
5. *Validation.* Detection limit 15 ng/spot, RSD 0.99–1.6%, recoveries from adult and children's syrup 100.0 and 99.5%, respectively.

Analysis of Pharmaceutical Preparations Using UV Densitometry

Salbutamol sulfate and bromhexine hydrochloride in formulations (23)

1. *Sample preparation.* Sample solutions were prepared in methanol at concentrations of 200–800 ng/ μ l.
2. *TLC.* Standards and samples applied with a Linomat IV to precoated alumina-backed silica gel 60 F layer, methanol-chloroform-triethylamine (5.5:4.5:0.05) mobile phase, development in a twin-trough chamber.
3. *Quantification.* Fluorescence quenched zones of samples and standards scanned at 276 nm.
4. *Validation.* Calibration curves linear over the range 20–580 ng/ μ l; RSDs ranged from 1.1 to 1.6% and recoveries from 98.8 to 99.6% for assay of the compounds in a syrup and tablet.

Diphenhydramine hydrochloride in tablet, gelcap, and capsule antihistamine pharmaceuticals (79)

1. *Sample preparation.* Ground powder or gel dissolved in ethanol.
2. *TLC.* HPTLC silica gel 60 F layer, samples and standards applied as 6 mm bands with Linomat IV, ethyl acetate-methanol-conc. ammonia (85:10:15) mobile phase, development in twin-trough chamber.
3. *Quantification.* Fluorescence-quenched zones scanned at 260 nm.

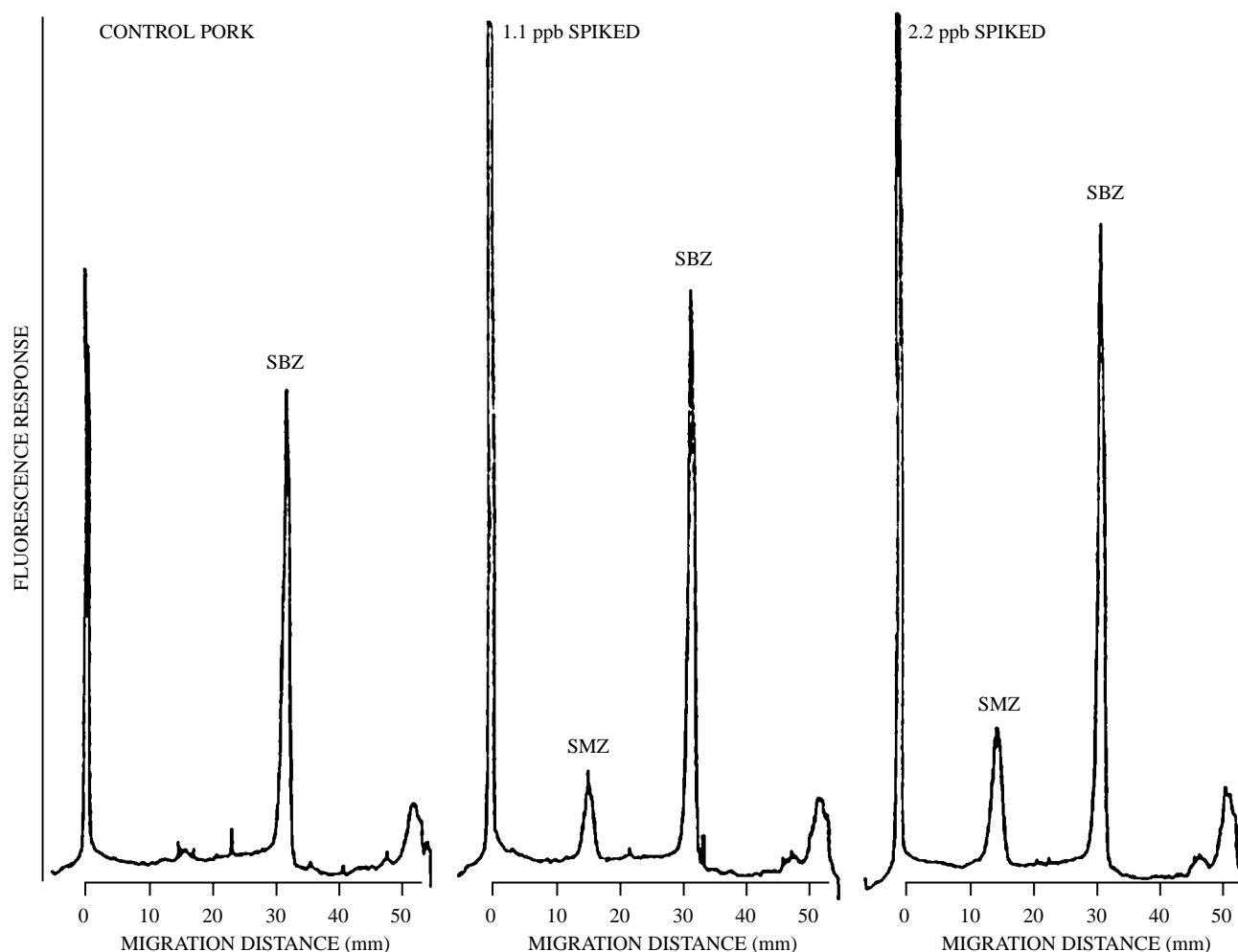


Fig. 5 Sample densitograms: Pork tissue samples fortified with 10 ppb sulfabromomethazine (SBZ) and either 0, 1.1, or 2.2 ppb of sulfamethazine (SMZ). Origin is at 0 mm, solvent front at 53 mm. (Reprinted from Ref. 77.)

4. *Validation.* Precision ranged from 1.7 to 1.9% RSD and errors in recovery analyses of spiked samples were 0.81 and 0%.

4. *Validation.* Recoveries from laboratory-made cream were 100.1 and 100.5%, respectively, and RSD ranged from 0.68 to 1.67% ($n = 6$).

Betamethasone valerate and miconazole nitrate in cream preparations (80)

1. *Sample preparation.* Creams were ultrasonicated with 96% ethanol, and insoluble material was removed by centrifugation and filtration.
2. *TLC.* 4 μ l aliquots of samples and standards applied using a Nanomat III to a silica gel 60 F layer, development in a twin-trough chamber with chloroform–acetone–glacial acetic acid (34:4:3).
3. *Quantification.* Fluorescence-quenched zones of scanned at 233 nm.

Pyridoxine hydrochloride and doxylamine succinate in tablets (81)

1. *Sample preparation.* Tablets were powdered, sonicated in methanol, and the solution filtered.
2. *TLC.* 5 μ l aliquots of samples and standards applied as 6 mm bands to HPTLC silica gel 60 F layer with the Linomat IV, acetone–chloroform–methanol–25% ammonia (7:1.5:0.3:1.2) mobile phase.
3. *Quantification.* Fluorescence-quenched zones scanned at 269 nm.

4. **Validation.** Linearity range 0.5–2.0 µg/spot; RSD 0.73 and 1.93%, and recoveries 99.3–103% and 97.7–101%, respectively.

Analysis of Pharmaceutical Preparations Using Fluorescence Densitometry

Amlodipine besylate in tablets (82)

1. **Sample preparation.** Tablets powdered, dissolved in methanol, and filtered.
2. **TLC.** Samples and standards applied to a silica gel 60 F layer with a Linomat IV, developed with chloroform–acetic acid–toluene–methanol (8:1:1:1) mobile phase in a twin-trough chamber.
3. **Quantification.** Fluorescent zones scanned at 366 nm.
4. **Validation.** Minimum detectable limit 0.2 ng; recovery from pre-analyzed tablet spiked with three different levels of the drug standard was 100.1%.

Analysis of Pharmaceutical Preparations Using Scraping and Elution of Zones

Sulfur in topical acne medications (83)

1. **Sample preparation.** Liquid and cream samples dissolved by boiling with acetone or chloroform.
2. **TLC.** Samples and standards applied as 2 cm bands, silica gel G layer, petroleum ether mobile phase.
3. **Detection.** Iodine vapor.
4. **Quantification.** Bands scraped and extracted with chloroform, UV absorption spectrometry at 265 nm.
5. **Validation.** Recovery from three spiked samples containing 3–5% sulfur averaged $99.2 \pm 2\%$, $100 \pm 2\%$, and $101 \pm 1\%$ for five replicates each.

REFERENCES

1. Arun, P.; Ravikumar, A.; Leelamma, S.; Kurup, K.A. Identification and Estimation of Endogenous Digoxin in Biological Fluids and Tissues by TLC and HPLC. *Indian J. Biochem. Biophys.* **1998**, *35*, 308–312.
2. Sherma, J. Comparison of Thin Layer Chromatography and Liquid Chromatography. *J. Assoc. Off. Anal. Chem.* **1991**, *74*, 435–437.
3. Dhanesar, S.C. Quantitation of Antibiotics by Densitometry on a Hydrocarbon-Impregnated Silica Gel HPTLC Plate. Part II. Quantitation and Evaluation of Ceftriaxone. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 258–262.
4. Morlock, G.; Charegaonkar, D. Determination of Hydrocortisone and Cinchocaine in Lanolin Ointments. *GIT Lab. J.* **1998**, *2*, 95–98.
5. Habel, D.; Guermouche, S.; Guermouche, M.H. HPTLC Determination of Isoniazid and Acetylisoniazid in Serum. *J. Planar Chromatogr.-Mod. TLC* **1997**, *10*, 453–456.
6. Popa, D.-S.; Oprean, R.; Curea, E.; Preda, N. TLC-UV Densitometric and GC-MSD Methods for Simultaneous Quantification of Morphine and Codeine in Poppy Capsules. *J. Pharm. Biomed. Anal.* **1998**, *18*, 645–650.
7. Sherma, J.; Incarvito, C.D. Analysis of Tablets and Caplets Containing Ketoprofen by Normal- and Reversed-Phase HPTLC with Ultraviolet Absorption Densitometry on Preadsorbent Plates. *Acta Chromatogr.* **1997**, *7*, 124–128.
8. Simonovska, B.; Prosek, M.; Vovk, I.; Jelen-Zmitek, A. High Performance Thin Layer Chromatographic Separation of Rantidine Hydrochloride and Two Related Compounds. *J. Chromatogr. B* **1998**, *715*, 425–430.
9. Dhanesar, S.C. Quantitation of Antibiotics by Densitometry on a Hydrocarbon-Impregnated Silica Gel HPTLC Plate. Part IV: Quantitation and Evaluation Penicillins. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 180–185.
10. Lazaric, K.; Cucek, B.; Faust, G. RPTLC Method for Determining Parabins in Oral Antacid Suspension During Stability Testing. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 86–88.
11. Radembino, N.; Loiseau, P.M.; Dessalles, M.-C.; Marchat, L.; Bories, C.; Gayral, P.; Mahuzier, G. Epoxyethane-/ethinesulfonamides with Antifilarial Activities. *Arzneim. Forsch/Drug Res.* **1998**, *48*, 294–299.
12. Zaidi, F.; Voirm, B.; Jay, M.; Viricel, M.R. Free Flavonoid Aglycones from Leaves of *Mentha Pulegium* and *M. Suaveolens* (Labiatae). *Phytochemistry* **1998**, *48*, 991–994.
13. Fater, Z.; Tasi, G.; Szabady, B.; Nyiredy, S. Identification of Amphetamine Derivatives by Uni-Dimensional Multiple Development and Two-Dimensional HPTLC Combined with Postchromatographic Derivatization. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 225–229.
14. Dong, C.; Xie, H.; Shuang, S.; Liu, C. Determination of Tetracycline and 4-Epimeric Tetracycline by TLC-Fluorescence Scanning Densitometry. *Anal. Lett.* **1999**, *32*, 1121–1130.
15. Darula, Z.; Torok, G.; Wettmann, G.; Mannekens, E.; Iterbeke, K.; Toth, G.; Tourwe, A. A Rapid, Qualitative Thin Layer Chromatographic Method for the Separation of the Enantiomers of Unusual Aromatic Amino Acids. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 346–349.
16. Bhushan, R.; Thuku, G.T. Direct Enantiomeric Resolution of some 2-Arylpropionic Acids Using (–) Brucine-Impregnated Thin Layer Chromatography. *Biomed. Chromatogr.* **1999**, *13*, 276–278.
17. Suedee, R.; Saelim, J.; Thavornpibulbut, T.; Srichana, T. Chiral Determination of Various Adrenergic Drugs by Thin Layer Chromatography Using Molecularly Imprinted Chiral Stationary Phases Prepared with Alpha-Agonists. *Analyst (Cambridge, UK)* **1999**, *124*, 1003–1009.
18. Huynh, T.K.X.; Lederer, M. Adsorption Chromatography on Cellulose. XI. Chiral Separations with Aqueous Solutions of Cyclodextrins as Eluents. *J. Chromatogr.* **1994**, *659*, 191–197.
19. Maxwell, R.J.; Lightfield, A.R. An Improved Method for Efficient Predevelopment Washing of HPTLC Plates. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 109–113.

20. Muller, D.; Ebel, S. Two-Dimensional Thin Layer Chromatographic Separation of H₁-Antihistamines. *J. Planar Chromatogr.-Mod. TLC* **1997**, *10*, 420–426.
21. Xu, X.; Stewart, J.T. Quantitative Analysis of Propantheline Bromide and Related Compounds in Drug Substance and Tablet Dosage Form by HPTLC And Scanning Densitometry. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 222–224.
22. Misztal, G. Thin Layer Reversed Phase Ion Pair Chromatography of some Cephalosporins. *J. Liq. Chromatogr. & Relat. Technol.* **1999**, *22*, 1589–1598.
23. Argekar, A.P.; Powar, S.G. Simultaneous Determination of Salbutamol Sulfate and Bromhexine Hydrochloride in Formulations by Quantitative Thin Layer Chromatography. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 254–257.
24. Krishnamurthy, R.; Ashwini, K. Simultaneous Determination of Adulterants and Coextractants in Illicit Heroin by HPTLC with Two Successive Mobile Phases. *J. Planar Chromatogr.-Mod. TLC* **1997**, *10*, 388–390.
25. Szabady, B.; Fater, Z.; Nyiredy, S. Comparative Study of Automated Development Chambers. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 82–88.
26. Szunyog, J.; Mincsovcics, E.; Hazai, I.; Klebovich, I. A New Tool in Planar Chromatography: Combination of OPLC and DAR for Fast Separation and Detection of Metabolites in Biological Samples. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 25–29.
27. Ojanpera, I.; Goebel, K.; Vuori, E. Toxicological Drug Screening by Overpressured Layer Chromatography. *J. Liq. Chromatogr. & Relat. Technol.* **1999**, *22*, 161–171.
28. Csermely, T.; Kalasz, H.; Rischak, K.; Bathori, M.; Tarjanyi, Z.; Gyarmati, Z.; Furst, S. Planar Chromatography of (–)-Deprenyl and some Structurally Related Compounds. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 247–253.
29. de Wasch, K.; de Brabander, H.F.; van Ginkel, L.A.; Spaan, A.; Sterk, S.S.; Meiring, H.D. Confirmation of Residues of Thyreostatic Drugs in Thyroid Glands by Multiple Mass Spectrometry After Thin Layer Chromatographic Screening. *J. Chromatogr. A* **1998**, *819*, 99–111.
30. Wang, P.; Zhou, M.; Feng, Y.; Chen, L. Simultaneous Determination of Trace Ofloxacin and Enoxacin by TLC-Fluorescence Spectrodensitometry. *Anal. Lett.* **1998**, *31*, 1523–1531.
31. Rischer, M.; Adamczyk, M.; Ratz, H.; Hose, H.; Marchsan, M.; Paper, D.H.; Franz, G.; Wolf-Heuss, E.; Engel, J. Quantitative Determination of the Iridoid Glycosides and Catapol in *Plantago Lanceolata* L. Extracts by HPTLC and HPLC. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 374–378.
32. Mangle, M.S.; Jolly, C.I. HPTLC Studies on *Tribulus Terrestris* L. (Chota Gokhru) and *Petalum Murex* L. (Bada Gokhru). *Indian Drugs* **1998**, *35*, 189–194.
33. Idrayanto, G.; Widjaja, S.; Aditama, L.; Darmawan, L. Densitometric Determination of S-Carboxymethylcysteine in Syrup, and Validation of the Method. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 263–266.
34. Chia, J.-C.; Chang, F.-R.; Li, C.-M.; Wu, Y.-C. Protoberberine Alkaloids from *Fissistigma Balansae*. *Phytochemistry* **1998**, *48*, 367–369.
35. Hostettmann, K.; Terreaux, C.; Marston, A.; Potterat, O. The Role of Planar Chromatography in the Screening and Isolation of Bioactive Compounds from Medicinal Plants. *J. Planar Chromatogr.-Mod. TLC* **1997**, *10*, 251–257.
36. Jork, H.; Funk, W.; Fischer, W.; Wimmer, H. Thin Layer Chromatography, Reagents and Detection Methods, Physical and Chemical Detection Methods: Fundamentals. *Reagents I*; VCH: Weinheim Germany, 1990; 1a.
37. Jork, H.; Funk, W.; Fischer, W.; Wimmer, H. Thin Layer Chromatography, Reagents and Detection Methods, Physical and Chemical Detection Methods: Activation Reactions, Reagent Series. *Reagents II*; VCH: Weinheim Germany, 1993; 1b.
38. Zweig, G.; Sherma, J. *Handbook of Chromatography*; CRC Press: Boca Raton, FL, 1972; II, 103–189.
39. Hahn-Deinstrop, E.; Koch, A.; Mueller, M. Guidelines for the Assessment of the Traditional Herbal Medicine “Olibanum” by Application of HPTLC and Desaga ProViDoc Video Documentation. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 404–410.
40. Rozylo, J.K.; Siembida, R.; Jamrozek-Manko, A. A New Simple Method for Documentation of TLC Plates. *J. Planar Chromatogr.-Mod. TLC* **1997**, *10*, 225–228.
41. Kenyon, A.S.; Xiaoye, S.; Yan, W.; Ng, W.H.; Prestridge, R.; Sharp, K. Simple, At-Site Detection of Diethylene Glycol/Ethylene Glycol Contamination of Glycerin and Glycerin-Based Raw Materials by Thin Layer Chromatography. *J. AOAC Int.* **1998**, *81*, 44–50.
42. Szepesi, G.; Nyiredy, Sz. Pharmaceuticals and Drugs. *Handbook of Thin Layer Chromatography*, 2nd Ed.; Sherma, J., Fried, B., Eds.; Marcel Dekker, Inc.: New York, 1996, Ch. 24.
43. Fu, Q.; He, L.; Miao, Z. Analysis of Strychnine Alkaloids in Biological Samples by Thin Layer Chromatographic Densitometry and Distribution Study of Strychnine in Intoxicated Rats. *J. Xi'an Med.* **1998**, *10*, 146–149.
44. DiGregorio, D.; Sherma, J. Analysis of the Active Ingredient Magnesium Salicylate in Analgesic Tablets by HPTLC with Ultraviolet Absorption Densitometry. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 230–232.
45. Hijli, P.S.; Phadke, M.M.; Shah, M.C.; Deshpande, P.P.; Sane, R.T. A High Performance Thin Layer Chromatographic Method for Determination of Doxazosin from Pharmaceutical Preparation. *Indian Drugs* **1998**, *35*, 653–657.
46. Essig, S.; Kovar, K.-A. Impurity Test of Theophylline: Comparison of Classical Slit Scanner with Videodensitometry. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 63–65.
47. Petrovic, M.; Kastelan-Macan, M.; Babic, S. Quantitative Evaluation of 2D Chromatograms with a CCD Camera. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 353–356.
48. Forgacs, E.; Cserhati, T.; Kaliszan, R.; Haber, P.; Nasal, A. Reversed Phase Thin Layer Chromatographic Determination of the Hydrophobicity Parameters of Nonsteroidal Anti-Inflammatory Drugs. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 383–387.
49. Cserhati, T.; Forgacs, E. Structure-Retention Relationships and Physicochemical Characterization of Solutes in Thin Layer Chromatography. *J. AOAC Int.* **1998**, *81*, 1105–1107.
50. Pradhan, P.; Banjerl, A. Stereochemistry of Diastereomeric Neolignans from Piper Hookeri by Nuclear Resonance Spectroscopy. *Phytochem. Anal.* **1998**, *9*, 71–74.

51. Kawabata, J.; Fukushi, E.; Mizutani, J. Sesquiterpene Dimer and Trimer from *Chloranthus Japonicus*. *Phytochemistry* **1998**, *47*, 231–235.
52. Kernan, M.R.; Amarquaye, A.; Chen, J.L.; Mrisho, S.; Rozhou, E.J. Antiviral Phenylpropanoid Glycosides from the Medicinal Plant *Markhamia Lutea*. *J. Natural Products* **1998**, *61*, 564–570.
53. Bauer, G.K.; Pfeifer, A.M.; Hauck, H.E.; Kovar, K.-A. Development of an Optimized Sorbent for Direct HPTLC–FTIR On-Line Coupling. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 84–89.
54. Stahlmann, S.; Kovar, K.-A. Analysis of Impurities by High Performance Thin Layer Chromatography with Fourier Transform Infrared Spectroscopy and UV Absorbance Detection In Situ Measurement: Chlordiazepoxide in Bulk Powder and its Tablets. *J. Chromatogr. A* **1998**, *813*, 145–152.
55. Wilson, I.D.; Spraul, M.; Humpfer, E. Thin Layer Chromatography Combined with High Resolution Solid State NMR for Compound Identification without Substance Elution: Preliminary Results. *J. Planar Chromatogr.-Mod. TLC* **1997**, *10*, 217–219.
56. Brzezinka, H.; Dallakian, P.; Budzikiewicz, H. Thin Layer Chromatography and Mass Spectrometry for Screening of Biological Samples for Drugs and Metabolites. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 96–108.
57. Stahlmann, S.A. Ten-Year Report on HPTLC–FTIR Online Coupling. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 5–12.
58. Cserhati, T.; Forgacs, E. Hyphenated Techniques in Thin Layer Chromatography. *J. AOAC Int.* **1998**, *81*, 329–332.
59. Wagner, J.; Jork, H.; Koglin, E. HPTLC as a Reference Method in Clinical Chemistry: On-Line Coupling with Spectroscopic Methods. *J. Planar Chromatogr.-Mod. TLC* **1993**, *6*, 446–451.
60. Klebovich, I.; Mincsovcics, E.; Szunyog, J.; Ludanyi, K.; Karancsi, T.; Ujszaszy, K.; Kiss, B.D.; Vekey, K. Isolation and Identification of Metabolites of ^3H - and ^{14}C -Deramciclane by OPLC-Digital Autoradiography On-Line Sample Collection and Mass Spectrometry. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 394–399.
61. Yamaguchi, T.; Nakajima, Y.; Mizobuchi, M.; Inazawa, K.; Kanazu, T.; Kadono, K.; Ohkawa, T.; Iwatani, K. Disposition and Metabolism of the New Hypocholesterolemic Compound S-8921 in Rats and Dogs. *Arzneim.-Forsch./Drug Res.* **1998**, *48*, 995–1006.
62. Ishigai, M.; Ishitani, Y.; Orikasa, Y.; Kamiyama, H.; Kumaki, K. Metabolism of 2(R,S)-1,2-Bis(Nicotinamido)-Propane, a New Agent with Anti-Vasospasm Activity, in Rats and Rabbits. *Arzneim Forsch./Drug Res.* **1998**, *48*, 429–435.
63. Renger, B. Benchmarking HPLC and HPTLC in Pharmaceutical Analysis. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 58–62.
64. Sherma, J. Planar Chromatography. *Anal. Chem.* **2000**, *72*, 9R–25R. Biennial Reviews of TLC have been Published each Even-Numbered Year in the Review Issue of *Anal. Chem.* since 1970.
65. Szepesi, G. Some Aspects of the Validation of Planar Chromatographic Methods in Pharmaceutical Analysis. I: General Principles and Practical Approaches. *J. Planar Chromatogr.-Mod. TLC* **1993**, *6*, 187–197.
66. Novakovic, J.; Nova, H.; Filka, K. In-Process Control and Process Parametric Assessment of Isosorbide 5-Mononitrate Pharmaceuticals. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 161–164.
67. Binter-Planas, V.; Durand, J.L.; Prognon, P.; Maison, C. Toxicological Research on Selective Serotonin Re-Uptake Inhibitors in Gastric Liquid by Thin Layer Chromatography. *J. Pharm. Clin.* **1997**, *16*, 266–269.
68. Mody, V.D.; Pandya, K.; Satia, M.C.; Modi, I.A.; Modi, R.I.; Gandhi, T.P. High Performance Thin Layer Chromatographic Method for the Determination of Sparfloxacin in Human Plasma and its Use in Pharmacokinetic Studies. *Biomed. Anal.* **1998**, *16*, 1289–1294.
69. Ng, L.L. Pharmaceuticals and Drugs. *Handbook of Thin Layer Chromatography*, 2nd Ed.; Sherma, J., Fried, B., Eds.; Marcel Dekker, Inc.: New York, 1991; 717–755.
70. Romano, G.; Caruso, G.; Musumarra, G.; Pavone, D.; Cruciani, G. Qualitative Organic Analysis. Part 3. Identification of Drugs and their Metabolites by Standardized TLC Data. *J. Planar Chromatogr.-Mod. TLC* **1994**, *7*, 233–241.
71. Ojanpera, I.; Ojansivu, R.-L.; Nokua, J.; Vuori, E. Comprehensive TLC Drug Screening in Forensic Toxicology: Comparison of Findings in Urine and Liver. *J. Planar Chromatogr.-Mod. TLC* **1999**, *12*, 38–45.
72. *The United States Pharmacopeia 24/National Formulary 19*; The United States Pharmacopeial Convention: Rockville, MD, 2000.
73. Renger, B. Quantitative Planar Chromatography as a Tool in Pharmaceutical Analysis. *J. AOAC Int.* **1993**, *76*, 7–13.
74. Renger, B. Contemporary Thin Layer Chromatography in Pharmaceutical Analysis. *J. AOAC Int.* **1998**, *81*, 333–339.
75. Posyniak, A.; Niedzielska, J.; Semenjuk, S.; Zmudzki, J. Screening for Clenbuterol and Salbutamol Residues in Urine by Planar Chromatography. *J. Planar Chromatogr.-Mod. TLC* **1995**, *8*, 157–159.
76. Fenske, M. Determination of Cortisol in Plasma and Urine by Thin Layer Chromatography and Fluorescence Derivatization with Isonicotinic Acid and Hydrazine. *Chromatographia* **1998**, *47*, 695–700.
77. Unruh, J.; Schwartz, D.P.; Barford, R.A. Quantitation of Sulfamethazine in Pork Tissue by Thin Layer Chromatography. *J. AOAC Int.* **1993**, *76*, 335–341.
78. Wiszkiendzky, A.; Maho, S.; Vegh, Z.; Ferenczi-Fodor, K. Validated, Stability-Indicating Semiquantitative Purity Test for Allylestrenol Drug Substance and Tablet by Personal OPLC. *J. Planar Chromatogr.-Mod. TLC* **1998**, *11*, 463–466.
79. Muller, E.E.; Sherma, J. Quantitative HPTLC Determination of Diphenhydramine Hydrochloride in Tablet, Gelcap, and Capsule Antihistamine Pharmaceuticals. *J. Liq. Chromatogr. & Relat. Technol.* **1999**, *22*, 153–159.
80. Indrayanto, G.; Widjaja, S.; Sutiono, S. Simultaneous Densitometric Determination of Betamethasone Valerate and Miconazole Nitrate in Cream, and its Validation. *J. Liq. Chromatogr. & Relat. Technol.* **1999**, *22*, 143–152.
81. Argekar, A.P.; Sawant, J.G. Simultaneous Determination of Pyridoxine Hydrochloride and Doxylamine Succinate in Tablets by HPTLC. *J. Liq. Chromatogr. & Relat. Technol.* **1999**, *22*, 2051–2060.
82. Chandrashekar, T.G.; Rao, P.S.N.; Smrita, K.; Vyas, S.K.; Dutt, C. Analysis of Amlodipine Besylate by HPTLC with

Fluorimetric Detection: A Sensitive Method for Assay of Tablets. *J. Planar Chromatogr.-Mod. TLC* **1994**, 7, 458–460.

83. Sanyal, A.K.; Chowdhury, B.; Banerjee, A.B. Rapid Thin Layer Chromatographic Microestimation of Elemental Sulfur: Application to Complex Sulfur Ointments. *J. AOAC Int.* **1993**, 76, 1152–1155.

BIBLIOGRAPHY

- Adomovics, J.A., Ed. *Analysis of Addictive and Misused Drugs*; Marcel Dekker, Inc.: New York, 1995.
- Camag *Planar Chromatography Catalog*; Camag Scientific Inc.: Wilmington, NC, 1998/99.
- Cserhati, T.; Forgacs, E. Thin Layer Chromatography. *Chromatography Fundamentals, Applications, and Troubleshooting*; Walker, J.Q, Ed.; Ch. 3, Preston Publications: Niles, IL, 1996.
- Fried, B.; Sherma, J. *Thin Layer Chromatography-Techniques and Applications*, 4th Ed.; Marcel Dekker, Inc.: New York, 1999.
- Jaenchen, D.E. Thin Layer (Planar) Chromatography. *Handbook of Instrumental Techniques for Analytical Chemistry*; Settle, F., Ed.; Ch. 13, Prentice Hall, PTR: Upper Saddle River, NJ, 1997.
- Sherma, J. Modern High Performance Thin Layer Chromatography. *J. AOAC Int.* **1994**, 77, 297–306.
- Sherma, J., Fried, B., Eds. *Handbook of Thin Layer Chromatography*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1996.

CHROMATOGRAPHIC METHODS OF ANALYSIS—HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

R. Raghavan
Jose C. Joseph

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

INTRODUCTION

Chromatography is a technique to separate individual components in a mixture. High-performance liquid chromatographic (HPLC) methods are usually preferred over other methods of quantitative analysis. The methods are usually very specific to the analyte or analytes of interest since excellent separation of individual components are easily achieved. HPLC instruments are ubiquitous since the technique finds application in biotechnological, biomedical, clinical, and pharmaceutical analyses. Additionally, HPLC is used in many other fields including chemical, cosmetics, energy, environmental, and food industries. The availability of moderately priced, reliable, efficient, and sophisticated instrumentation has resulted in the use of HPLC as a method of choice in the pharmaceutical analysis, starting from the synthesis or isolation of a potential drug to the final stage of maintaining quality control information on a formulated dosage form.

Since the fundamental theoretical principles of HPLC were established in the 1960s, the development of HPLC instrumentation has been phenomenal. Development in column packing materials led to the development of reverse-phase chromatography in the 1970s. Development of computers and automation in the 1980s led to the ease of use of HPLC. In the 1990s the development of microcolumns, specialized columns, stable detectors, coupled with integrated data acquisition, storage, and retrieval capabilities has vastly increased the speed and efficiency of the HPLC instruments.

BASIC CONCEPTS, DEFINITIONS, AND CHROMATOGRAPHIC THEORY

Basic Concepts and Definitions

In HPLC for separation of individual components, the sample is introduced into a flowing stream of a liquid (*mobile phase*) and the analytes are allowed to pass

through a layer column of packing materials of very small diameters (large surface area), called the *stationary phase*. As the analyte molecules pass through the column, carried by the moving mobile phase, there is constant interaction of the analyte molecules (or solutes) with the stationary phases as well as with the moving mobile phase. This results in a dynamic equilibrium. The differences in the equilibrium processes of the different solute molecules result in the separation of components of the mixture. When such separation is achieved by maintaining a constant composition of all the constituents of the mobile phase, the process is known as *isocratic elution*. If the mobile phase composition is changed continuously with respect to one or more of the solvents in the mobile phase, as a function of time, it is called *gradient elution*. When the *effluent* with mobile phase zones containing the analyte molecules emerges out of the column, it is passed through a *detector*, or a series of detectors. The detector signals respond as function of the solute concentration in the mobile phase zones. These signals are fed into *data processors*, which plot signal responses as a function of time. The graphic display of signals is called a chromatogram and the individual component zones are identified as *chromatographic peaks*. These peaks are characterized by the following parameters: their *peak widths*, *peak areas* or *peak heights*, and the extent of *tailing* and the *retention time* of the peaks. The instrumental set up is called a chromatograph. A typical chromatogram is shown in Fig. 1.

In HPLC solute molecules are introduced into a moving mobile phase stream. The stream passes through an inlet and emerges through the outlet of the column. Since the particles are extremely small in size (10 μm or less) and the column is fully packed, the moving mobile phase has to be pumped through using high pressure pumps. The solute molecules are only carried by the moving mobile phase. Molecules that interact with the surface of the column will be impeded and the emerging band will elute later than a band of weakly interacting molecules.

The relative migration of the solute is dependent on the thermodynamic and kinetic properties of the solute. The

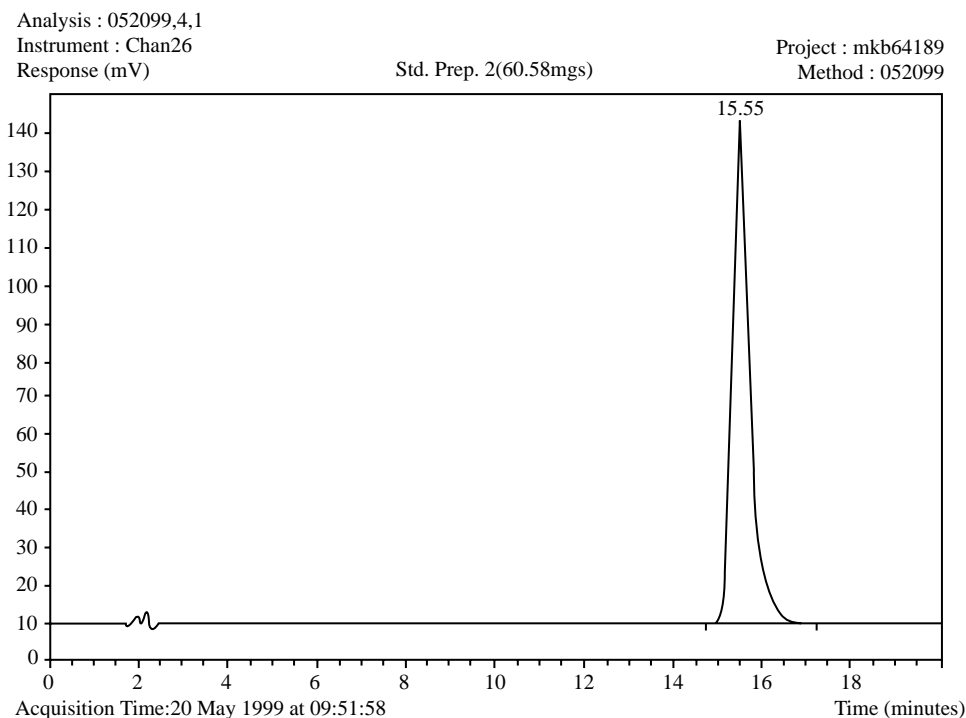


Fig. 1 Typical chromatogram.

extent and quality of separation of two closely eluting peaks are expressed by *retention (or capacity) factor* (k'), *selectivity factor* (α), and *the number of theoretical plates* (N). Capacity factor is a measure of time that the solute molecules are attached to the column particles, in comparison to that of the mobile phase. Thus greater the value of k' , the greater is the interaction with the column particles. The capacity factor is dependent on the nature of the column, the organic or aqueous strength of the mobile phase, and the temperature at which the column is maintained. Experimentally measurable parameter of *relative retention time* with respect to the retention time of an active drug is computed. Under isocratic elution conditions, a value of 2–6 for capacity factor is optimal and normally values between 1 and 10 are acceptable. Greater the value of k' , greater is the resolution between adjoining bands. However, as k' increases, there is increase in analysis time, which also results in lower detection limit, because of peak broadening. By using gradient elution these two disadvantages of isocratic elution can be overcome.

The chromatographic separation process is considered *efficient* if all the components are completely separated and the peak width is relatively narrow. Theoretically, when identical molecules enter the column head in a narrow band, the band width should be the same at the outlet. However, since solute molecules can elute at

slightly different times because not all molecules will take the same path. These differences are caused by differences in the local surface area, relative physical activity of the interacting surface, the presence of stagnant mobile phase pools in crevices and pores, and minor variations of flow rate of mobile phases through these surfaces. Not all solute molecules traverse the same path and hence they might contribute to peak broadening.

Additionally, the quality of separation is evaluated by measurement of resolution, “ R ,” between two closely eluting peaks.

$$R = 2(t_2 - t_1)/(w_1 + w_2)$$

where W_1 and W_2 are peak widths expressed in the same units as retention times, t_1 and t_2 .

The greater the value of R , greater is the separation. For an R value of 1.00, solute purity is about 97.7% if each peak is Gaussian. In practice to attain a peak purity of 99.8% or greater a resolution of 1.50 is required.

The efficiency of separation, expressed as *theoretical plate number*, N , is calculated as follows:

$$N = 16(t_r/w_B)^2 = 5.54(t_r/w_{1/2})$$

where “ t_r ” is the retention time, w_B , is the peak width at the base, and “ w ” is the band width at the peak height. A

column independent parameter, H (height equivalent to Theoretical Plate, $(HETP) = L/N$, where L is the length of the column), is more often used.

Column efficiency is inversely proportional to the particle size of the column packing. Thus the efficiency of separation will follow the following order.

$$E_{3\mu} > E_{5\mu} > E_{10\mu}$$

where E is efficiency and the subscripts denote particle size.

HPLC techniques can be used for preparative chemical separations. However, this discussion will be restricted to quantitative analytical separations. For quantitative analysis a known volume of a standard solution of known concentration is injected multiple times (most compendial methods require typically five to six injections). The average peak area of the peak of interest is computed. From a comparison of the peak area of similarly injected and separated analyte with that of the standard, the concentration of the unknown in the analyte is calculated. This procedure is known as *external calibration*. However, sometimes a known compound is added to both the standard and the analyte sample. Then the ratio of the relative peak area (or some times peak height) responses of the peak of interest, and that of the added compound are evaluated. From a comparison of the relative responses of the standard and that of the analyte injections, the concentration of the unknown is computed. This is known as *internal calibration*. Sometimes, peak height is used instead of peak area. The theoretical plates, resolution of two closely eluting peaks, percent relative standard deviation values of multiple injections, and tailing factor (extent of deviation of the chromatographic peak shape from symmetrical Gaussian peak) are used as *system suitability parameters*. USP 24 (see bibliography), and other monographs provide examples of system suitability requirements and methods of measurement to meet the corresponding requirements.

COLUMNS AND MODES OF CHROMATOGRAPHY

The stationary phase in HPLC is the solid support contained in within a specified column over which the mobile phase flows effecting the separation of the individual components. The HPLC column is normally fabricated using 100- to 300-mm long stainless steel tubes with an internal diameter of 2–5 mm. They are packed with porous, microporous, spherical, or irregularly shaped particles, or particles with specific coatings with the following characteristics.

Mean particle sizes of 3–10 μm , surface area between 150 and 400 m^2/g , specific pore volume between 0.2 and 1.5 cm^3 or mL/g , and an apparent density of 0.4–0.6 g/mL . The different modes of chromatography are distinguished based on the differences in the packing materials, coupled with the corresponding compatible mobile phase components and the differences in the nature of the interacting functional groups present in solute molecules. These functional groups selectively and specifically interact with the column support material or mobile phase leading to selectivity and specificity of separation. The stationary phase chemical characteristics are altered by using suitably modified silica particles such that the differences in the functional group properties can be selectively utilized. These are summarized in Fig. 2. A brief description of the different modes of chromatography follows.

Normal Phase Chromatography

In normal phase chromatography, a polar stationary phase and a nonpolar mobile phase is used for separation. A modulator, like methanol or acetonitrile, at a suitable concentration can be used to increase the polarity of the mobile phase. Most normal phase chromatographic columns use bare silica support, which is acidic and polar. The acidic surface silanol groups, which are hydrophilic, interact differently with different functional groups in the solute molecule. The lipophilicity of the mobile phase also affect the preferential solubility or preferential adsorption on the surfaces. The use of silica support in normal phase chromatography suffers from the following disadvantages: 1) product dependent activity of silica leading to poor separation and variations from column to column and from brand to brand; 2) irreversible adsorption of strong polar solutes on the column support; 3) the necessity to control the water content of the mobile phase; and 4) slow re-equilibration of particles to mobile phase changes. Some of these problems are over come with the use of modified silica stationary phases. The silica surfaces are modified by bonding with appropriate functional groups. Cyanoalkyl or aminoalkyl or phenyl moieties are bonded to these surfaces.

Non-polar supports like polystyrene/divinylbenzene copolymers or carbon are also used as column materials. Alumina is polar and acidic while TiO_2 , and zirconia are much more neutral. They all have good aqueous stability compared to silica. Normal phase chromatography is restricted to the separation of stereochemical isomers, diastereomers, low molecular weight aromatic compounds and functionalized long chain aliphatic compounds.

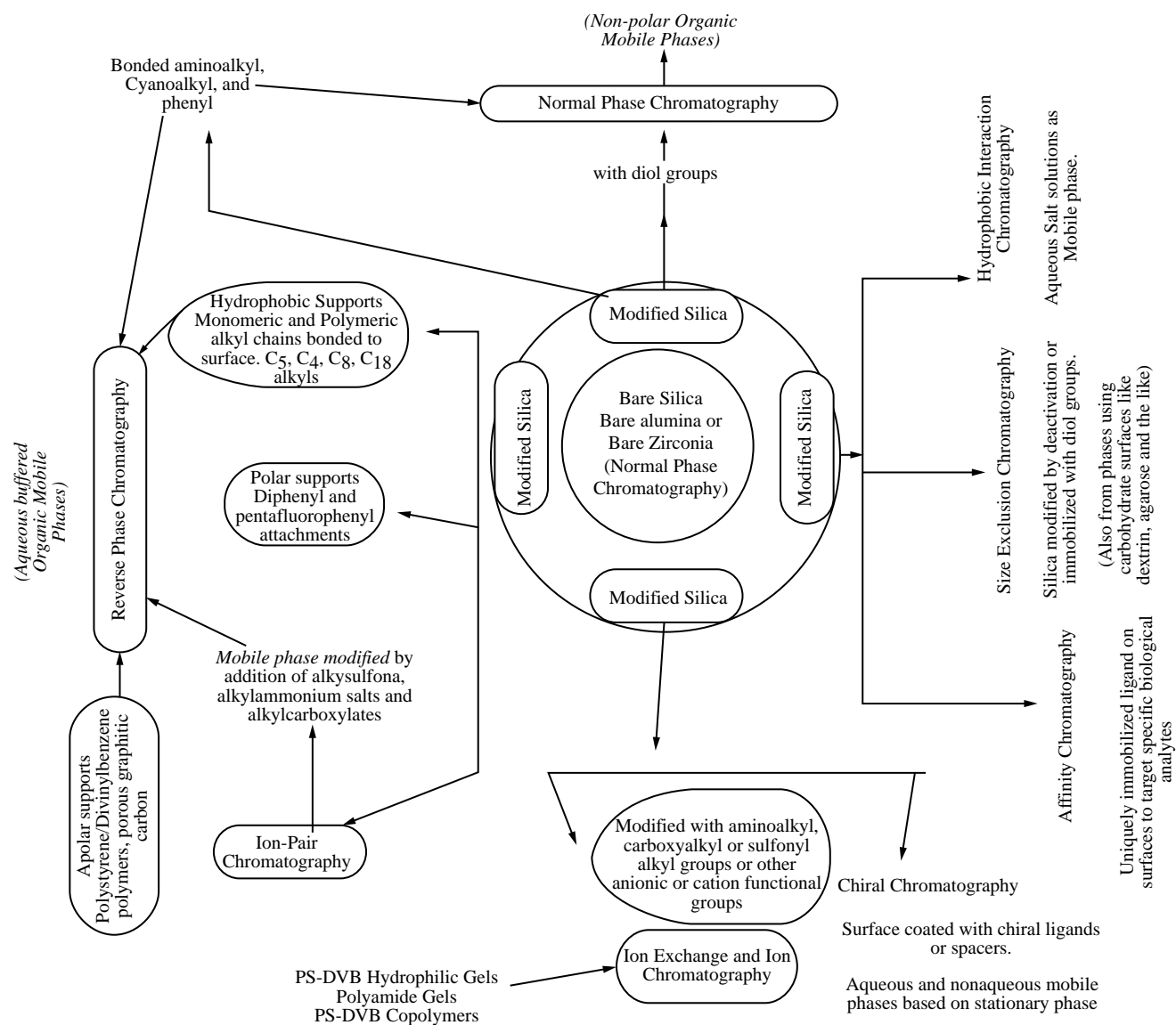


Fig. 2 Silica surface and different modes of chromatography.

Reverse Phase Chromatography

Since many compounds of pharmaceutical interest are generally polar and highly water soluble, reverse phase chromatography is extensively used. In reverse phase chromatography, separation is accomplished by the use of polar mobile phases on nonpolar stationary phase. By chemically bonding silanol groups in silica, nonpolar stationary phases are obtained. Typically C₃, C₄, C₈, C₁₈ alkyl chains are bonded to silica support surfaces. Mobile phases are usually buffered aqueous solutions containing one or more of the organic solvents like methanol, or acetonitrile, or tetrahydrofuran as modulators.

The modulators reduce the polarity and decrease retention of solutes. Depending on the organic solvent used selectivity also can be modified. Column designations like octyl (C₈), octadecyl (C₁₈), refers to the length of the carbon chain attached to the silica surface. Amino, cyano, and phenyl columns can also be used in reverse phase chromatography. To increase polarity of the stationary phases diphenyl and pentafluorophenyl columns are also used. Alkylated polystyrene/divinyl benzene polymers also can be used instead of silica based supports.

In reverse phase chromatography separation may be due to either adsorption effects or due to partitioning of the solute between the stationary phase and the mobile phase.

More often, separation is probably based on both mechanistic pathways; the relative contribution of each for a specific separation process cannot be estimated. In general, C₁₈-bonded phases yield better retention and better separation compared to C₈-bonded phases because of higher carbon content in the stationary and higher non-polar interaction with the solute.

In isocratic elution, some times compounds are not fully resolved and some compounds are highly retained. The sensitivity of the highly retained compound is reduced considerably due to peak broadening. These drawbacks are overcome when gradient elution is adopted. In gradient elution, the sample is injected when mobile phase has low organic content. The organic fraction is then increased in increments to decrease polarity. These increments are usually linear. Multiple step gradients are also adopted. The capacity factor is dependent on flow rate, slope of the gradient and column dead volume. In linear gradient elution, the bandwidth is generally constant and hence the peak becomes sharper, yielding enhanced sensitivity.

Reverse phase liquid chromatography is very versatile, fast and highly reproducible. Aqueous solutions are normally used and the modifiers used are very cheap and highly pure. Separation is predictable based on the polarity, pH profile, solubility and other physicochemical characteristics of the solute molecules. Analysis time is rather short and reequilibration is generally fast. Multiple components with minor differences in polarity can be separated by appropriate choice of gradient profiles.

Stationary phases have been modified to

1. Reduce interaction of free silanols
2. Improve the stability of phases over a wide range of pH
3. Introduce functional groups on the phases that will enable prediction of selectivity for different solutes.

Only about 50% of the free silanols in silica are bonded in octyl and octadecyl stationary phases. These residual silanols contents is further reduced by a process called *endcapping*, in which hexamethyl and isobutyl groups are additionally introduced into the matrix. Endcapped columns offer better separation and retention in addition to reduced peak tailing.

Although endcapped columns offer some advantages, the aqueous instability is still problem. To overcome these problems of silica based columns, alternative supports like alumina (Al₂O₃), zirconia (ZrO₂), and titania (TiO₂) have been developed. Alumina columns are stable in the pH range of 2–12, while zirconia columns extend the range from 0 to 14. These are basic oxides and hence silanol-like interactions are eliminated. Different bonded phases can be obtained using zirconia, however, because of poor

reactivity such bonded phases are difficult to prepare with alumina supports.

Polydivinylbenzene and polystyrene polymer based stationary phases also eliminate these effects. Porous graphitic carbon provides a highly nonpolar surface with excellent chemical stability under acidic and basic conditions. However, they suffer from lower sample loading capacity and lower efficiency than conventional columns.

In spite of these efforts, silica based modified columns are still widely used. Many column-manufacturers have introduced quality control procedures for the synthesis and characterization of silica based columns. Therefore, lot to lot variations from the same manufacturers have been considerably reduced. However, identical phases from different manufacturers can yield different separation behaviors of the same analyte under identical instrumental and mobile phase conditions. (Of 13 phenyl columns investigated for an oncology, we found only three columns providing similar separation profile for a 13-component impurity mixture.)

Ion-Pair Chromatography

In reverse phase chromatographic separations, ionic compounds, being more water soluble, are not retained in the column. To increase retention and separation a strong counter ion (an organic alkyl or aryl-substituted ion of opposite charge) is added to the mobile phase. Typically alkane sulfonic acid salts or alkyl ammonium salts are added to the mobile phase. These counter ions associate (ion-pairs) with the analyte ion, while displacing the inorganic counter ion like chloride ions. Analyte is retained since the ion-pair partitions into the stationary phase like a large non-polar neutral organic molecule. This technique, also known as *ion-interaction chromatography*, utilizes the effect of pH, ionic strength, mobile phase organic content and temperature to control retention and separation.

Ion-Exchange and Ion Chromatography

Ion-exchange stationary phases consist of solid resin particles that have positive or negative ionic bonding sites incorporated in the stationary phase. The ions of opposite charge in the mobile phase are exchanged with ions on the surface. The ions of opposite charge in the mobile phase are exchanged with ions on the surface. Cation exchange resins contain covalently bound negatively charged functional groups, while anion exchange resins have positively charged functional groups. When the charged

functional groups is a sulfonate anion, it is called strong cation exchanger. Weak cation exchange resins contain such functional groups as carboxymethyl, phosphate, sulfoalkyl groups. If strongly basic quaternary amines are on the resin, it is called a strong anion exchanger. Weak anion exchangers contain weakly basic groups like aminomethyl, diethylaminomethyl groups. If these functionalities are only on the surface of the stationary phase they are called pellicular particles. When pellicular particles are used, lower eluent concentrations are adequate. When pellicular ion-exchange resins are used, ion-exchange is the only method of separation. When ions are thus separated, particularly in the separation of inorganic ions or small organic acid anions, it is called *ion chromatography*. Most modern ion chromatographic stationary phases use polystyrene divinylbenzene copolymer resins. These stationary phases have very high pH stability and can withstand strong acids and bases.

Using ion chromatographic separation and conductivity detection the inorganic anions like halides, phosphate, nitrite, nitrate, thiocyanate, and sulfate and many cation ions including transition metal ions can be detected. When quantitation of ions are carried out in a solution matrix that is weakly conducting, conductometric detection and quantitation is possible since the total background conductivity is very small. Examples include the determination of ions in sea water or tap water or from environmental streams. However, if strongly acidic or basic eluents are used, the background conductivity is high. In order to suppress the background conductivity, special suppressor columns are used, which neutralize the acids or bases after elution and before detection. New pulsed amperometric detectors (PAD) are commercially available. With the use of PADs, parts per billion levels of metal ions can be detected. Accurate quantitation of metal ions is possible since ready to use inorganic calibration standards are commercially available.

When metal ions are used as counter ions, instead of organic quaternary ammonium ions, in the packing material, the hydroxy ($-\text{OH}-$) functional groups of the carbohydrates and other sugars interact with these metals ion. Pb^{2+} , Ca^{2+} , and Na^{+} , are typical metal ions in the packing material. Depending upon the type of counter ions used, the intensity of the interaction changes and therefore, the retention between different carbohydrates vary. Some of the carbohydrates are also retained because of the size of the molecule under these conditions.

Ion-exchange chromatography is widely used for analyses of proteins, glycoproteins, peptides and other high molecular weight compounds. These organic compounds have considerable surface charge and behave like charged anions. Hence they are amenable to ion

exchange separation. To separate nucleotides of similar molecular structures, the differences in the phosphate groups of various nucleotides and the differences in their binding characteristics are used. In addition to silica based resins, acrylic polymer based resins, dextrans, and cellulose bonded phases are used for the separation of proteins. In order to preserve the biological activity during separation, hand poured columns packed with ion-exchange materials are used. Gravity flow of eluent at low temperatures is the norm for separation. Thus, ion exchange chromatography and ion chromatography are no more used synonymously.

Hydrophobic Interaction Chromatography

In hydrophobic interaction chromatography, weakly hydrophobic sorbents are used. Gradient elution with decreasing concentration of salt is used for the separation of large biomolecules, particularly proteins, by this technique. The non-polar functional groups of large biopolymer molecules (weakly) associate with the hydrophobic ligands in the stationary phase. The stationary phase consists of a highly hydrophobic organic layer. The organic layers contain short alkyl or aryl functional groups attached at the surface. These attached groups are separated with large unattached space in between these attached functional groups. Because of this wider spacing, these “*soft*” stationary phases preserve biological activity without denaturing the proteins. High ionic strength aqueous mobile phases enhance binding between the solute and the stationary phase. Then the salt concentration is decreased to decrease the ionic strength of the mobile phase. The weak mobile phases then reduces the binding and thus separation is effected.

Typical stationary phases include the following: polyvinylpyrrolidone (PVP) coated silica sorbents, monodisperse nonporous silica columns with surface bound amides or ethers and composite agarose and polyacrylamide gels. The eluent normally consists of salts at concentrations greater than 1.0 M. Typical salts include sodium phosphate, sodium sulfate and ammonium sulfate, and organic acid salts like monosodium glutamate. Protein retention is stronger with salts that increase surface tension like phosphates, sulfates, citrates, which are solvated in water than with salts such as perchlorates and thiocyanates and the like.

Typical biological compounds that are separated by HIC include, cytochrome P-450, enzymes, DNA polymerase, epidermal growth factor, glycoprotein hormones, human immunoglobulins, human recombinant DNA and canine pancreatic juice proteins. Many HIC techniques have been used for large scale purification of proteins.

Affinity Chromatography

This chromatographic technique uses a specific binding agent. The stationary phase is prepared by immobilizing one of a pair of interacting molecules on to particles of support. The immobilized molecule is referred to as a *ligand*. These ligands selectively bind to the interacting second pair in a protein or a biomolecule. For example, an antitransferrin antibody is immobilized on the support. In this example, the antibody is the ligand; the transferrin antigen in the biomolecule will bind to the surface or release out of the surface depending on the mobile phase strength. Therefore, this technique, which utilizes the differences in the affinity of the two specific interacting groups or moieties is called *affinity chromatography*.

Typical ligands may be of biological origin like antibodies, inhibitors, substrates, coenzymes, cofactors, nucleic acids, and the like, or of nonbiological origin like triazine dyes, metal chelates, boronate salts, etc. In this technique, sample is injected on to the column using a weak mobile phase called the application buffer. Under these conditions, the only interacting component is bound to the surface and hence retained in the column. The rest are washed out of the column. Then using a stronger mobile phase, called the eluent buffer, the solute of interest is released, eluted from the column, and then quantitated or collected for later use. Elution may involve two separate steps or may be a simple step gradient. This technique is used for the separation of hormones, peptides, proteins, viruses, enzymes, glycopeptides, antibodies, metal binding amino acids, etc. Affinity chromatography is further classified as bioaffinity, bioadsorption, immunoaffinity and the like depending on the nature of the ligand on the support.

Size Exclusion Chromatography

Size exclusion chromatographic (SEC) technique is used for the separation of biomolecules based on their molecular size. Synthetic and many natural polymers like polysaccharides, celluloses, natural rubber, and some proteins have chains of differing molecular weight components. When such mixed molecular weight species are present it is said to be a *polydisperse* polymer. Otherwise, the monomer is said to be monophasic. The SEC chromatographic peak is broad indicative of the elution of the different components of the polydisperse phase. The polydisperse phase is described by up to “3” molecular weight parameters that define the distribution of species. These are 1) number average, M_n ; 2) the weight average, M_w ; and 3) z-average, M_z , molecular weights. When $M_n = M_w$, the distribution is said to “Monodisperse.” M_w/M_n is a measure of the polydispersity of the system. For large biomolecules, M_n

and M_w are different since M_w is usually higher, because it is sensitive to the presence of high molecular components in the distribution. M_n , M_w , M_z are defined as follows:

$$M_n = \sum N_i M_i / \sum W_i$$

$$M_w = \sum M_i W_i / \sum W_i$$

$$\text{and } M_z = \sum W_i M_i^2 / \sum W_i M_i$$

where N_i is the number of molecules of molecular weight M_i , and W_i refers to the weight (or concentration) of M_i . The ratio of M_w/M_n or M_z/M_w shows the width of the distribution. Size exclusion chromatography is a relative and not an absolute technique.

Gel permeation chromatography (GPC) refers to the technique in which polymers that are soluble in organic solvents are separated. In These cases, more polar organic mobile phases like tetrahydrofuran, toluene, chloroform will be used. Gel filtration chromatography (GFC) is used for separation of water soluble biopolymers. Four different calibration methods are used. If absolute known molecular weight standards are used, it is called *primary calibration method*. In *secondary calibration approach*, poly-dispersity standards of material similar to samples are used. The result is then usually specified as apparent molecular weight distribution. When M_w and M_n are obtained by use of an iteration procedure using a sophisticated software program, it is called *broad molecular weight calibration*. The iteration procedure uses calibration slopes and intercepts of broad molecular weight standards with known M_n and M_w values. *Universal calibration* is obtained from a plot of $\log (M\eta)$ vs. V_e , elution volume, where η , is the intrinsic viscosity of the polymer measured at the same temperature and in the same solvent as used for the mobile phase. This technique uses on line SEC viscometers in conjunction with universal calibration.

For organosoluble polymers cross linked polystyrene or silica based packings are used. For water soluble polymers various silica based and hydrophobic polymeric packings are used. Pore sizes of the SEC packings may range from 3 to 300 nm.

In addition to refractive index detectors, specialized detectors such as on-line SEC detectors and low angle laser light scattering detectors are used for determining the distribution of molecular weights by SEC.

INSTRUMENTATION

Solvent(s) Delivery

Solvent (mobile phase) delivery is achieved using high pressure pumps. There are several types of pumps

commercially available for delivery of mobile phase through the injector, column, detectors and then to the solvent waste container. Since the column head pressure is high, the pumps operate under high pressures (50–300 psi). Most commonly used pumps are reciprocating piston pumps with check valves. These pumps are usually computer controlled. The modern pumps provide flow rate precision which is better than 0.1% in retention time.

For gradient elution at least two solvents has to be pumped and then mixed before it passes through the column. Additionally, the solvent composition is changed in a continuous linear, continuous nonlinear, or stepwise fashion. The solvents are mixed with the use of proportioning valves, and the mixed solvents reach the pump. Since solvent mixing occurs before the pump under low pressure conditions, it is called low pressure mixing. This is the most commonly used form of mixing for gradient elution. In high pressure mixing, two or more solvents are individually pumped using different pumps and then they are mixed at high pressures. Pump head leakage is a common problem in such pumps. The pump heads has to be constantly monitored or repaired for leakage.

Dissolved oxygen from mobile phase solvents has to be removed. This is accomplished by passing helium gas through the solvent container or by passing the mobile phases through helium purging units. Currently, computers that control the pumps also control vacuum degassers, which are placed in between the pump and solvent reservoir.

Autosamplers

Autosamplers are used for unattended introduction of samples from vials that are arranged either in a rectangular or circular tray. Autosampler delivers the desired volume of 1–2.5 ml with a precision of less than or equal to 0.5% for injections greater than 10 μ l. All autosamplers use mechanized valves. Majority of HPLC autosamplers belong to one of two types. In type I autosamplers the septum cap is pierced using a syringe needle, liquid is displaced into the syringe by gas pressure or plunger action to the inlet port of a six port valve injector. The filled syringe is withdrawn, moved, and the solution deposited into flowing stream of mobile phase by the use of appropriate electromechanical devices.

Type 2 autosamplers behave in a very similar way to type I except partial loop volume can be filled and delivered. Additionally, this allows the injection of low volume with high precision compared to type I autosamplers. Both types of autosamplers suffer from carry over problems if appropriate wash cycle and wash solvents are not used between each sampling from the vials.

Detectors

A liquid chromatography detector consists of sensors and an associated electronic device to send signals to a processor. Detectors are classified as bulk property detectors or solute property detectors. Bulk property detectors measure the changes in the property of the combined eluting mobile phase and the eluting solute. For example, the refractive index is characteristic of a liquid. When a solute is dissolved, the refractive index of the solution is different from that of the solvent. These change the property of the bulk solution. Although the change is due to the presence of the solute, the refractive index of the bulk as a whole is different from that of the pure solvent. Refractive index detectors and conductivity detectors are examples of bulk property detectors. Solute property detectors detect the changes in some physical or chemical property of eluting solute component of the mobile phase.

HLPC detector

A HPLC detector should have the following characteristics.

1. Excellent linear response as a function of concentration of the solute.
2. Wide linear dynamic range; the dynamic range over which the response to concentration is linear.
3. High signal to noise ratio. The noise arises as a result of fluctuations or perturbations caused to the signal as a result of temperature, pressure, or flow rate changes in the mobile phase. Noise is also caused by the electronic circuits used in the detector system. All these combined perturbations, called noise, should be low such that very low concentrations of solute can be detected.

Refractive index detectors

The most common Refractive Index (RI) detector uses a differential refractometer, which responds to the deflection of a light beam; the deflection being caused by the differences in the refractive indices of a cell through which eluant passes and that of a reference cell in which the mobile phase is contained. The response of the detector is proportional to the mass concentration irrespective of the nature of solute being analyzed.

Conductivity detector

The conductivity detector measures the conductivity of a solution containing an electrolyte. When current is allowed to pass through two electrodes, there is resistance (or better impedance) to the flow of current through the medium. This impedance decreases if conducting electrolytes are present in the eluant. This detector is mostly used in ionchromatography. Coupled with ion

suppression technology, this has become a versatile detector for low levels of inorganic ion content in analytes of interest.

UV–vis detectors

UV–vis spectrophotometric detectors are most commonly used detectors in HPLC, since most organic compounds absorb light in the UV region (190–400 nm) and a few in the visible region (400–750 nm). Fixed wavelength, variable wavelength, and diode array detectors are commercially available. All these operate based on the ability of a solute to absorb light at defined wavelengths based on the chemical structure and functional groups present in the solute molecule. The source of UV light is a deuterium or high pressure xenon lamp while for the visible range it is a simple tungsten lamp.

A beam of light is allowed to pass through a flow cell mounted at the end of the column. As the solute molecules elute from the column and enter the flow cell, they absorb radiation. The differences in the light energy, as a result of absorption, are used as a measure of quantitation. Fixed wavelength detectors operate at a single wavelength, either at 254 or at 280 nm in the UV region. In variable wavelength detector using a monochromator light of a particular wavelength (less than ± 3 nm) can be selected, passed through the sample, and then on to a photocell for detection. Currently available detectors can be programmed to change wavelengths while analysis is in progress to get a spectrum of the eluting species. Otherwise, using appropriate software, the absorbance of the eluate can be monitored simultaneously at two to four different wavelengths. This multi-wavelength detector is less sensitive compared to fixed wavelength detectors (10^{-7} g/ml vs. 5×10^{-8} g/ml).

Photodiode array detectors

Diode array detectors acquire data over the entire range of UV–vis range 190–800 nm; in some up to 1100 nm. Two different types of photodiode array detectors are available in the market. In one, to detect over an entire spectrum, light from a continuous source is passed through the cell using a rapidly rotating or vibrating grating, which passes radiation through the cell, one wavelength at a time. The signal of the photodetector is measured as a function of time over the measuring cycle. Then, from the measuring cycle, wavelength is related to time to obtain a plot of wavelength vs. signal. In the second type polychromatic light is passed through the cell and then through a holographic grating. The light dispersed from the grating is arranged to fall on a linear photodiode array. These diode array detectors are very versatile and are used to:

1. Check peak purity using “peak overlay” (normalized) methods or by computing peak ratios at two different wavelengths
2. Identify peaks by spectral matching with accumulated and stored spectral libraries
3. Generate the spectrum of the eluting peak and determine wavelength of maximum absorption of an unknown or impurity peak
4. Quantify different peaks at different wavelengths in a single chromatographic run
5. Provide graphic 3D or contour plot presentations to regulatory agencies to show the purity of the eluting chromatographic peak
6. Identify peaks, during method optimization, when the order of elution of the compounds changes.

It should be noted that the sensitivity of these detectors is lower than fixed or variable wavelength detectors. Also, if an overlapping impurity is present either in the fronting or tailing portions of the eluting peak, it can be detected only when the concentration of the impurity is greater than 2.0% relative to that of the major peak.

Fluorescence detectors

Since the fluorescence detector is a highly sensitive, picogram levels of solute can be detected by using this detector. However, this is limited to compounds that naturally fluoresce or can be made to fluoresce by reacting with suitable derivatizing agents. Even this is restricted since appropriate functional groups that undergo such derivatization should be present in the solute molecule. In many HPLC methods where this technique is utilized, the fluorescent agent is added after separation of the components. This has the advantage that the derivatization need not be quantitative. But the reaction should be very rapid, reproducible and proportional to concentration. Because it is a unique property of the solute, it offers selectivity as well as specificity of detection.

Electrochemical detector

This is also a very specific and extremely sensitive detector. The specificity arises from the need to have an electro-oxidizable or reducible functional group present in the solute molecule. Similar to the case of fluorescence detectors, solute molecules can be derivatized to yield compounds containing oxidizable (or rarely, reducible) functional groups. A desired potential is applied between a working electrode and a reference electrode connected to the flow cell. A third electrode known as auxiliary electrode is used to control the potential. As the oxidation takes place at the working electrode surface, the current

flow changes. This is monitored, amplified and presented as response using appropriate software and hardware. When the oxidation reaction is allowed to go to completion using a high surface area of working electrode and exhausting all the reactant in the flow cell, it is called a *coulometric* detector. In this case, the total number of coulombs of charge transferred is measured. However, in the most common *amperometric* detector, the solute molecules at the surface and those are very close to the surface are oxidized by maintaining the working electrode at a constant potential. This oxidation process is diffusion controlled and is proportional to concentration. Here the increase in the current flow, “*i*” is measured, amplified and the signal is presented as a function of time. Glassy carbon electrode is the commonly used electrode for oxidation. Surface coating and the resulting contamination of the surfaces leads to a decrease in sensitivity on constant use. Therefore, this detector has to be disassembled, and cleaned very often. Also this requires long equilibration time compared to other detectors. Therefore, it is not an extensively used detector compared to other detectors described in this article. The problem of electrode pollution is overcome in pulsed amperometric detector (PAD). In this technique using a gold or platinum electrode a repeating cycle of potential pulses are applied. Typically, in a one second pulse cycle, three potential pulses are applied. In the initial negative pulse the solute is adsorbed; in the second positive potential pulse the adsorbed compounds are oxidized and the increased current as a result of oxidation is measured. In the third pulse at a high much higher positive, the electrode surface is cleaned by oxidation of the electrode itself. Thus the new surface generated is used for the repeat cycling process. This is very useful in the detection of very low levels of sugars and other polyhydroxy compounds that are otherwise not easily oxidized.

Evaporative light scattering detector

In this detector, the entire eluate from the column is atomized and evaporated to form small droplets. The solutes finally remaining form particulates suspended in the atomizing gas. When these particles are allowed to pass through a light beam, light is scattered in all directions by the particles. This is known as Rayleigh scattering. However, the light scattered at 45° angle to the incident beam is viewed using appropriate optical filters and the resultant signal is electronically processed. The detector response is sensitive to the mass of the solute particles and hence it is a universal detector. The sensitivity of this detector compares with that of the RI detector.

Computers

The computers are ubiquitous and have become indispensable component of any laboratory. It serves both data processing and process control functions. As a data processor, the computer receives, stores, archives and reprocess the input signals from various detectors. Additionally, as a system processor, the computer monitors the system detectors. The data acquired for each chromatographic run can be processed appropriately to arrive at peak area or peak height or response ratios to an added internal standard. From these it can be used to calculate the concentrations of individual components in an analyte.

Through the use of appropriate software and hardware, the computer can be programmed to do the following: 1) to command injection of the samples; 2) to control and monitor, various parameters of the pump like the flow rate, composition of the mobile phase, column pressure; 3) to monitor and control column oven, and detector temperatures; and 4) to start and stop injectors, detectors and other system units. The computer also can be used to monitor system suitability parameters and reinject samples after adjusting conditions to meet system suitability. A decision tree can be constructed to allow retesting when pre-established system suitability conditions are not met. The computer can be programmed such that if conditions are not met, the system is shut down or paused until it can be attended to. Sample preparation, derivatization and other processes also can be controlled using individual computers. With multiple computers attached to a large computer called the *server*, data from a number of detectors can be monitored, stored, and archived. The computer is also used as an excellent book-keeper storing all the information regarding samples, their results, and also the conditions under which those results are obtained. The computer has become versatile tool in the highly regulated pharmaceutical industry especially to provide traceability and data integrity to required government and compendial agencies such as FDA, USP, BP, etc.

METHOD DEVELOPMENT AND METHOD VALIDATION

The method development process involves selecting appropriate method conditions for the sample in hand. It is based on prior knowledge of the sample properties, pK_a or pK_b values of functional groups, the polarity and size of solute molecules, UV-vis spectral properties, redox behavior, concentration range, solubility behavior and the like. From a knowledge of these, suitable mode of chromatography, corresponding column(s), mobile phase

composition, flow rate, choice of detectors, gradient or isocratic conditions, and the like can be selected. Once the method has been developed with some initial trials, optimization is carried out. Optimization is necessary to accomplish best possible separation of all components within the shortest possible time or in the case of low level detection, conditions have to be optimized such that required level of detection and/or quantitation can be achieved. In general, the system suitability parameters are usually evaluated and specified before method validation is performed.

Method validation is a process by which documented evidence is prepared and provided to show that the method meets the intended need. Highly regulated pharmaceutical analytical laboratories perform method validation and generate data on the following parameters, to comply with the compliance requirements of government agencies such as FDA, EPA and/or to provide data for compendial agencies like USP, BP, etc.

The parameters that define validation are accuracy, precision, specificity, linearity, ruggedness, and robustness. Accuracy is a measure of the closeness of the measured value to the true value or an accepted reference value. This is usually measured by spiking known amounts of the analyte to a matrix called the placebo, and computing the recovery of the analyte after sample analysis. Placebo contains all the ingredients of a formulation other than the active ingredient or the ingredient being analyzed. FDA and ICH (International Committee on Harmonization of Technical Requirements for Registration of Pharmaceutical for Human use) guidelines recommend collecting data from nine determinations at (at least) three concentration levels encompassing the range of target analyte concentration.

Precision refers to the degree of repeatability under the stated conditions of the method. It is expressed as percent relative standard deviation (% RSD) for a statistically significant number of analyses of samples. Precision provides a measure of day to day, analyst to analyst and instrument to instrument variation on a routine basis. The precision data provided in support are standard deviation, % RSD, confidence intervals and may also include inter laboratory variations.

Specificity refers to the ability to determine the concentration of the analyte with a high degree of confidence that the other components in the matrix do not interfere with the target analyte. The potential interfering substances include other active and inactive ingredients, impurities, degradation products of the components and active ingredient, and extractables from the container-closure system and the like. Specificity is the currently accepted terminology by regulatory agencies. In the

literature selectivity is also used to suggest specificity. [Refer (1225) of USP XXII (1900) vs. XXIII (1995) and 24 (2000).]

Linearity refers to the linear response of the detector to the analyte concentration within a specified *range*. Range, expressed in the same units as the analytical test results, is the interval between the lower and upper levels of the analyte concentration. To show linearity, a plot of response vs. concentration is provided for at least at five different concentration levels within the range. In addition, the slope and intercept of the regression line with the correlation coefficient are also provided. Normally a single standard is recommended if the intercept is very close to zero. Otherwise, quantitation, based on linear plot of multiple standards, is generally recommended.

Ruggedness is again a measure of precision. ICH guidelines include ruggedness in precision, while USP separates it. Ruggedness according to USP is a measure of the variation in interlaboratory comparison data. It is performed to establish lack of influence of test results based on operational and environmental parameters.

Robustness is a measure of how method parameters like organic content, pH, ionic strength of the mobile phase and column temperature do not affect test results when minor variations are deliberately induced in these parameters. The quality of separation of components, the accuracy and precision of the method and the like should not change as a result of these variations. (It is the belief of the authors that robustness should be part of the method development process regarding the appropriate choice of the column and the mobile phase conditions.)

Two other terms, limit of quantitation (LOQ) and limit of detection (LOD) are used especially in the determination of the analytes at trace levels. The trace level analyte may be impurities or degradation products of ingredients in a sample or it may be a solution of active ingredient at trace levels in rinse or swab samples obtained as part of cleaning validation. LOD refers to the lowest concentration of an analyte that can be detected (but not quantitated), under a given set of chromatographic conditions. Usually, it is expressed as a concentration parameter calculated using a signal to noise ratio of 3:1. LOQ refers to the concentration of the analyte that can be quantified in a sample with a predefined value of variation in precision. Normally accepted signal to noise ratio value for LOQ is 10:1. Generally, LOD and LOQ values are dictated by specific requirements, in order to ascertain reproducibility of LOD and LOQ values. Examples of specific requirements include determination of impurities at 0.05% using a 0.1% surrogate standard, or ppm quantitation or detection requirements based on equipment cleanliness and the like. For practical purposes, it is

recommended that a control solution at a known concentration of the analyte be prepared such that a peak corresponding to this analyte concentration is always detected (LOD) or quantitated (LOQ). From one of these two parameters, if experimentally obtained, the other can be calculated.

PRACTICAL CONSIDERATIONS

Sample Preparation

For many analyses, sample preparation may simply involve dissolving a known weight or volume of sample and diluting to appropriate concentration for analysis. However, extensive sample preparation may be required, if precolumn derivatization is needed. Precolumn derivatization is carried out to increase sensitivity, to induce specificity, or to separate optical isomers and to reduce or eliminate otherwise complex and expensive clean up procedures. Solid phase extraction using appropriate solid cartridges are normally used in the clean up of biological samples. Solid phase cartridges can also be used to increase the concentration of trace analyte. These purification steps can be done in situ during analysis by appropriate plumbing using six- or ten-port valves and using computer control of the pumps and valves. No matter how it is done, attention must be paid for proper sampling and preparation of samples.

Mobile Phase Preparation

Mobile phase should be prepared using HPLC grade solvents and analytical reagent chemicals only. Mobile phases should be filtered using ($\leq 0.5 \mu\text{m}$) filters to remove any particulate matter from the solvent. To eliminate dissolved air, helium sparging or vacuum degassing is necessary. In gradient elution, in order not to alter solid phase wetting and gelling characteristics, it is advisable to use at least 5% water content instead of pure organic solvents as one of the mobile phases.

System Maintenance

Pumps and columns should be cleaned of with 50:50 methanol–water or acetonitrile–water mixture to eliminate buffers from the system. Pumps, injectors, and detectors should require periodic and scheduled maintenance to keep instruments in good operating conditions. Detector should be calibrated on a regular basis using appropriate calibration standards available from NIST or other sources traceable to NIST. FDA and other regulatory

agencies require extensive documentation regarding the “health” of these instruments. Therefore, necessary and appropriate documentation should be maintained regarding details of maintenance. If there is any instrument failure additional documentation is necessary to prove that analytical results were not impacted or compromised because of these failures, if any.

BIBLIOGRAPHY

- Barth, H.G.; Boesm, B.E.; Jackson, C. *Anal. Chem.* **1998**, *68*, 455–466.
- Berthod, A. *J. Chromatogr.* **1991**, *549*, 1.
- Bidlingmeyer, B.A. *J. Chromatogr. Sci.* **1997**, *35*, 392–400.
- Brown, P.R.; Grushka, E., Eds. *Advances in Chromatography*; Marcel Dekker, Inc.: New York, 1998; 39.
- Hanson, M.; Unger, K.K. *LC-GC* **1997**, *15*, 364–366, 368.
- Katz, E., Eksteen, R., Schoenmakers, P., Miller, N., Eds. *Chromatographic Science Series. Handbook of HPLC*; Marcel Dekker, Inc.: New York, 1998; 78.
- Katz, E.D., Ed. *High Performance Liquid Chromatography: Principles and Methods in Biotechnology*; Wiley: Chichester, UK, 1996.
- Kissinger, P.T. *J. Pharm. Biomed. Anal.* **1996**, *14*, 871–880.
- Krull, I.; Swartz, M. *LC-GC* **1997**, *15*, 534–536, 538, 540.
- LaCourse, W.R.; Dasenbrock, C.O. *Analytical Reviews* **1998**, *70*(12), 37R–52R, 251R–278R; 591R–644R.
- Lloyd, D.K. *High Perform. Liq. Chromatogra* **1996**, *114*–42.
- Lough, W.J., Wainer, I.W., Eds. *High Performance Liquid Chromatography: Fundamental Principles and Practice*; Blackie: Glasgow, UK, 1996.
- Mant, C.T., Hodges, R.S., Eds. *High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*; CRC Press: Boca Raton, FL, 1991.
- Neue, U.D. *HPLC Columns: Theory and Practice*; Wiley-VCH: New York, 1997.
- Noctor, T. *High Perform. Liq. Chromatogr.* **1996**, *97*–113.
- Raghavan, R.; Joseph, J.C. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J., Eds.; Marcel Dekker, Inc.: New York, 1993; 7, 249–298.
- Riley, C.M., Rosanske, T.W., Eds. *Progress in Pharmaceutical and Biomedical Analysis. Development and Validation of Analytical Methods*; Pergamon Press/Elsevier Science: New York, 1996; 3.
- Rothman, L.D. *Anal. Chem.* **1996**, *68*, 587–598.
- Snyder, L.R., Kirkland, J.J., Galajch, J.L., Eds. *Practical HPLC Method Development*, 2nd Ed.; Wiley: New York, 1997.
- Swadesh, J., Ed. *HPLC: Practical and Industrial Applications*; CRC Press: Boca Raton, FL, 1997.
- Swartz, M.E.; Krull, I.D. *Analytical Method Development and Validation*; Marcel Dekker, Inc.: New York, 1997.
- U.S. Pharmacopeia 24/National Formulary 19, United States Pharmacopeial Convention. Rockville, MD, 2000.
- U.S. Pharmacopeia XXIII/National Formulary XVIII, United States Pharmacopeial Convention. Rockville, MD, 1995.
- U.S. Pharmacopeia XXII/National Formulary XVII, United States Pharmacopeial Convention. Rockville, MD, 1990.
- Weston, A.; Brown, P.R. *HPLC and CE: Principle and Practice*; Academic Press: San Diego, CA, 1997.



Clinical Supplies Manufacture: GMP Considerations

David L. Chesney

KMI, Waltham, Massachusetts, U.S.A.

Mark L. Balboni

KMI, Laguna Niguel, California, U.S.A.

INTRODUCTION

Applicability of GMP to Clinical Supplies

The Federal Food, Drug, and Cosmetic Act (the Act), Title 21, U.S. Code, section 301 et. seq., establishes that a drug shall be deemed to be adulterated if "...the methods used in, or the facilities or controls used for, its manufacture, processing, packing, or holding do not conform to or are not operated or administered in conformity with current good manufacturing practice to assure that such drug meets the requirements of this article as to safety and has the identity and strength, and meets the quality and purity characteristics, which it purports or is represented to possess." [21 U.S.C. section 351(a)(2)(B)]. This requirement is generally referred to as "CGMP" (current good manufacturing practice), or simply "GMP."

The requirement as written in the law is obviously very broad. It does not define the specific steps manufacturers must take to comply, or the controls that must be in place to ensure compliance. A later section of the Act gives the Secretary of Health and Human Services authority to promulgate regulations for the efficient enforcement of the Act. [In practice, this authority is delegated to the Commissioner of the Food and Drug Administration (FDA).] The FDA, acting in accordance with this authority, has promulgated regulations found in Title 21 of the Code of Federal Regulations (CFR) that sets forth the definitions to be used in specifying GMP requirements for all drugs (21 CFR Part 210) and the procedures and controls necessary for manufacturing of finished pharmaceuticals (human and veterinary), found in 21 CFR Part 211. Part 211 of the regulations is what is generally meant by the term "CGMP" or "GMP" (the acronym GMP will be used in this text). While FDA gives consideration to prevailing practice in regulated industry before specific requirements are written into the regulations as being "current, good practices," FDA's final basis used to establish GMP requirements is whether the practice is "feasible and valuable" in assuring drug safety, quality,

and purity.^[1] We also point out that the final GMP regulations are the product of notice and comment rulemaking through publication, first as a proposal in the *Federal Register*. The affected public, including the industry, has an opportunity to comment on the proposed rule before it is finalized.

Because the regulations apply to all "finished pharmaceuticals," and because of the statutory requirement to comply with GMP, there is no question that the regulations apply in a binding manner to the manufacture of clinical supplies. Nevertheless, a debate has existed for many years, and even continues today, regarding exactly how and in what respects a company may differ in its approach to the application of GMP regulations to clinical supplies vs. full-scale commercial production.

BACKGROUND INFORMATION

Clinical supplies, also known as clinical trial materials, are those investigational new drug products intended for administration in human or veterinary (animal) patients during clinical trials.^[2] In many respects, there is no difference between the equipment and technology employed to manufacture clinical supplies and that used for commercial production. In other respects, for example, production scale, robustness of the manufacturing process, labeling for clinical trial materials, expiration period (and hence the need for supporting stability data), final container, and even formulation and dosage form, there are important differences that should be taken into account when designing appropriate GMP controls for clinical supplies.

Placebos used in medical treatment to bring about a therapeutic effect without a pharmacologically active ingredient, or placebos used as study controls in clinical trials for new drugs are also subject to the requirements in the current good manufacturing practice regulations (GMP).^[3] Because they lack any active ingredient, placebos clearly cannot be tested for potency, but their inactive



composition can be confirmed. This and other considerations unique to placebos also call for some degree of interpretation of GMP.

Clinical supply manufacturing operations are those areas involved in the manufacture of Phase I–IV clinical trial materials, and may include laboratory (or table-top) scale activities, operations performed in a pilot plant (with batch sizes generally larger than lab scale, but smaller than commercial scale), clinical supplies produced in facilities manufacturing commercially approved products, as well as clinical supplies produced at contract manufacturing sites.^[4]

The GMP regulations apply to investigational new drug products produced for clinical trials in humans or animals, whereas those activities earlier in the product development life cycle such as “basic research,” discovery or preclinical experimentation are not subject to GMP requirements.

FDA INSPECTIONS OF CLINICAL SUPPLY MANUFACTURERS

Under the Federal Food, Drug, and Cosmetic Act, the FDA has the authority to “...enter, at reasonable times, any...establishment in which food, drugs, devices or cosmetics are manufactured, processed, packed, or held...” [section 704(a), FD&C Act]. This authority clearly includes sites that manufacture or conduct analysis of clinical supplies (for lot release, stability, or in connection with failure investigations). The term “reasonable time” has been interpreted by the courts to mean any time-regulated operations are taking place, regardless of date, day of the week, or time of day or night. Inspections are typically not preannounced, but may be in certain cases.

In practice, FDA only occasionally inspects clinical supply manufacturing sites. However, that should not result in a false sense that an FDA inspection will never take place. Manufacturing sites should always be prepared to undergo FDA inspections. Factors that may cause FDA to inspect clinical supply manufacturers include: Review in connection with a preapproval inspection of the commercial manufacturing site; routine inspections of contract manufacturers who manufacture only clinical supplies; inspections resulting from observed product defects, especially when such defects are linked to adverse reactions in patients (such as contaminated parenterals); in reaction to recalls of clinical supplies; or other factors. The FDA inspections of manufacturing operations for clinical supplies used in Phase III trials are more likely than for operations producing Phase I or Phase II clinical trial material.

Each site should have ready access to a person or persons highly knowledgeable in FDA inspection procedure who can prepare personnel for the inspection and manage it from the company’s perspective when it occurs. It is not the purpose of this article to explain in detail how to manage an FDA inspection, but the authors highly recommend that key personnel at each site receive relevant training in preparation for an FDA inspection, and prepare detailed policies and procedures concerning the handling of FDA inspections.

UNDERSTANDING AND USING FDA DOCUMENTS

To properly interpret and apply GMP requirements to a specific context require some research and a good deal of judgment. This process can be greatly aided by reference to a variety of FDA documents in the public domain. Most are available via links found on the FDA website, www.fda.gov. Examples of useful documents include: *Guidelines*; the so-called *Points to Consider* documents issued by the Center for Biologics Evaluation and Research (CBER); *Compliance Policy Guides*; *Inspection Technical Guides*; *Compliance Program Guidance Manuals*; and others.

A key document for interpretation of GMP in a clinical supply setting is FDA’s 1991 *Guideline on the Preparation of Investigational New Drug Products (Human and Animal)*. This Guideline illustrates *when* drug development activities are subject to GMP requirements, the degree of compliance expected at certain stages of drug development, and FDA guidance for several important sections of the GMP regulations.

FDA’s position on the applicability of GMP to clinical supplies was clearly articulated in the Preamble to the September 29, 1978 revision of the Drug GMPs (*Federal Register*, Vol. 43, No. 190, 45013–45336) that stated: “GMP regulations apply to the preparation of any drug product for administration to humans or animals, including those still in investigational stages. It is appropriate that the process by which a drug product is manufactured in the development phase be well documented and controlled in order to assure the reproducibility of the product for further testing and for ultimate commercial production.”

Compliance Policy Guides are FDA internal documents that provide precedent for GMP (and other) enforcement decisions by FDA’s field offices. They are found on the “Field Operations” or “ORA” (Office of Regulatory Affairs) pages on FDA’s website under the heading “Compliance References.”



Points to Consider documents are topical policy statements by CBER on subjects such as viral inactivation, transgenic animals, and other specialized technology subjects.

Compliance Programs provide FDA investigators with procedural guidance for performing inspections in a wide variety of industries. Some (but not all) of these are also accessible via the ORA link of the FDA's website.

Industry publications and seminars can also provide useful information, but users should be wary of placing too much weight on "podium policy" statements by FDA officials or reading inspection citations issued to other companies. These sources may be misleading when the reader may not know all the facts surrounding any statements.

SELECTED GMP SYSTEMS AS APPLIED TO CLINICAL SUPPLIES

The Quality Organization

21 CFR 211.22 is the regulation that sets forth the responsibilities of the "Quality Control Unit" (QCU). There is no difference in the required roles and responsibilities of the QCU in the manufacture of clinical supplies vs. commercial products. The FDA considers the presence of an adequately staffed and trained QCU, empowered with authority to carry out its responsibility effectively, as a critical factor in GMP compliance.

Companies should be aware that "Quality Control Unit" is a generic term, and the company-specific terminology may differ from the term QCU. For example, in most companies, this unit is referred to as Quality Assurance (QA). Other common terms include Regulatory Compliance, QA/QC, or in some cases, the Regulatory Affairs group fulfills many, if not all, of the listed QCU functions. In very small companies or "virtual" firms, the QCU may be as small as one person, while in large companies the number of QCU employees may be quite large, and the quality organization may have several subdivisions. Thus, the name given to the QCU and its members is not as important as the authority to implement the required roles and responsibilities.

The regulation requires that the QCU be responsible for the following, at a minimum:

- Dispositioning (approval or rejection) of components, containers, closures, in-process materials, packaging, labeling, and finished products.
- Review of production records to assure that no errors have occurred.
- If errors have occurred, they have been fully investigated.
- Approval or rejection of standard operating procedures (SOPs) and specifications.
- Approval or rejection of changes.
- Oversight of contracted manufacturing operations (including testing).
- Approval or rejection of all activities having a potential impact on the safety, purity, potency, quality, and efficacy of the products being manufactured.

The FDA expects (and many other regulatory authorities require) that the QCU will have full independence from other units. This is to provide the maximum degree of assurance that the QCU's decisions will be free of conflicts of interest and other impediments. Normally, the expectation is that the head of the QCU will report to a very senior level person (CEO, President, COO, etc.) of the company, and that there will not be a situation where the head of the QCU reports to manufacturing, marketing, or other similar functions.

The GMP requires that the QCU should have a qualified laboratory. This may be an in-house laboratory that is part of the QCU itself, or it may be an outside contractor or another company site laboratory. Here again, there is an expectation of independence of the laboratory from manufacturing or other units, in order to assure the maximum degree of objectivity in the data that are generated.

Master (and Batch) Production and Control Records

The specific requirements for master production and control records, and batch production and control records outlined in sections 21 CFR 211.186 and 21 CFR 211.188, respectively, also apply to clinical supplies. Despite these requirements, the level and amount of available documentation for early developmental batches will generally be much less when compared to commercial products, until the manufacturing process becomes more fully defined. Batch production and control records for clinical batches produced will have an increased amount of notes, changes, and other information handwritten on them during execution of clinical batches. The amount of this extraneous information usually coincides with the stage of development for the drug product (e.g., early phase production = more notes and changes), and may include the following:

- Notes of any situations arising during production.
- Problems that occurred.



- Modifications made to the formulation, processing step sequence, processing parameters, or equipment set points.

This information that is annotated on executed master production and control records should be evaluated, reviewed and, if necessary, approved, and incorporated utilizing an established change control system.

In many cases, master production and control records used in the production of clinical supplies start out as “batch cards,” which are sometimes printed on a heavier stock paper that may be colored (e.g., blue or green), in order to differentiate them as related to R&D production operations, and not commercially manufactured products.

Buildings and Facilities

While buildings and facilities are required to be of suitable size and construction, as well as maintained in a good condition for both approved products and clinical supplies, the level of protection these areas are required to provide is dependent upon the activities taking place inside a respective area. For example, sterile fill operations necessitate, among other things, high-efficiency particulate air (HEPA)-filtered air, temperature and humidity controls, and Class 100 (Class A) environmental conditions.^[5] It would be inappropriate to perform aseptic filling operations for even Phase I clinical supplies in a nonenvironmentally controlled laboratory suite under a hood certified to meet Class 100 conditions. As it will be noted later in this article, there is little to no difference in matters of sterility assurance when comparing clinical trial materials to approved products.^[6] Similar discretion should also be exercised for operations such as dispensing, mixing, and packaging because FDA expect buildings and facilities used to manufacture clinical supplies are the same as those for commercial products.

Equipment

As with buildings and facilities, equipment must also be of suitable size and construction, as well as suitably located to facilitate its operation, maintenance, and cleaning (21 CFR 211.63). While there is clearly no difference in regulatory requirements for equipment used to manufacture clinical supplies, when compared to commercial manufacturing (e.g., equipment must be maintained and cleaned at appropriate intervals, measurement devices calibrated, qualified, and operating according to written and approved procedures), cleaning is especially important, in as much as many compounds encountered during drug development are of unknown toxicities that pose additional risks.^[7]

Generally, fully established and validated cleaning procedures will not be in place during Phase I and II clinical supply manufacture. However, the actual cleaning regimen should be robust enough to minimize the potential for contamination or product carry-over. This may be accomplished via visual inspection, or possibly by verification through actual sampling and analytical testing. Finally, 21 CFR Part 211.105 requirements regarding identification of equipment used in the manufacture of clinical trial material can be satisfied with a single sign, or placard, denoting the material in question, the phase of production, and batch number.^[8]

Control of Incoming Materials

General requirements related to incoming materials, which include components, drug product containers, and closures, include establishing and following written procedures for the receipt, identification, storage, handling, sampling, testing, and approval (or rejection) of said items (21 CFR 211.80). Research and development organizations will typically have receiving functions and areas that are separate from those used to receive incoming materials for commercial manufacturing operations. The quantities of components, drug product containers and closures, and the containers holding them are generally smaller when compared to the volumes and sizes of incoming materials received by commercial operations.

It is important to note that FDA expects incoming materials used in the manufacture of clinical supplies to be released by the Quality Unit (21 CFR 211.22). Attempts to circumvent the established receiving operation in order to expedite use of these materials should be avoided. Equally important is the requirement to ensure proper storage and segregation of incoming materials, such that quarantined, released, and rejected items are easily identified, and there is some level of separation among them. Different lots or item numbers of similar materials should not be comingled during storage, in order to avoid inadvertent mix-ups with their use in manufacturing (21 CFR 211.80).

Finally, incoming components (e.g., raw materials) used to manufacture clinical supplies must be tested for identity using a specific identity test (if available). Specifications should be established and confirmed for each lot of components, drug product containers, and closures received (21 CFR 211.84 and 21 CFR 211.94), and all other provisions of 21 CFR Part 211, Subpart E, Control of Components and Drug Product Containers and Closures not noted in this section also apply to clinical supply manufacturing operations.



Qualification and Validation Activities

Qualification activities are normally associated with buildings, facilities, utility systems (e.g., water, air handling, Clean-in-place/Steam-in-place (CIP/SIP), and compressed gases) major equipment (including laboratory instrumentation), whereas validation likely is in reference to those confirmatory tasks related to processes and analytical methods. In simplistic terms, validation (and qualification) can be defined as documented evidence that a process, activity, or piece of equipment can consistently meet its predetermined acceptance criteria and quality attributes.^[9] This section will be dedicated towards outlining the requirements for validation of manufacturing processes, as the requirements for buildings, facilities and equipment, and the validation of cleaning processes have been discussed in previous sections.

The FDA has indicated that during development, processes are expected be validated, and this should be completed to the “extent possible.” Challenges with earlier stages of development include fewer batches from which to establish physicochemical characteristics (including toxicity and potency), processing parameters, and commensurate equipment set points. There is also a greater reliance on in-process monitoring and testing, and final product testing for Phase I, II, and possibly early Phase III clinical supply manufacture. This more intensive monitoring and testing is needed to supplant full process validation, where multiple batches have not been produced under replicated conditions. Once, as FDA puts it, “a growing body of scientific data and documentation” reaches a point, complete validation of the manufacturing process will be required.^[8]

While blending times and tablet press parameters may not be fully established for early phase clinical supply manufacturing of solid oral dosages, those variables have a much lower potential to directly affect product safety than sterility, endotoxin contamination, or objectionable types and levels of particulates do for sterile, parenteral clinical supplies.^[6] Because clinical supplies can be incompletely characterized, and are usually given to patients already in weakened conditions,^[7] those processes and their related validation data necessary to guarantee patient and product safety (e.g., sterilization and aseptic fill) are expected to be in place as early as Phase I clinical supply manufacture.^[6, 8]

Production and Process Controls

Written procedures are required to be established and followed for production and process controls as specified in 21 CFR Part 211.100. Because the specific requirements regarding handling of changes, deviations, and equipment

identification are addressed in other sections of this article, and all other provisions of Subpart F are required in order to meet CGMP requirements for clinical supplies, this section will focus on the other aspects of Subpart F that provide unique challenges during clinical supply manufacture.

Yield calculations are not only required according to 21 CFR Part 211.103, but they also provide some measure of process, phase, or step consistency when compared on a batch-by-batch basis. This information may prove useful when continuing process development, process optimization, or in the investigation of any process problems encountered during clinical supply production. Actual yields and percentages of theoretical yield also one of many important measures used to evaluate the validation of a manufacturing process.

Time limitations on production may not be fully known or established during early clinical supply development. However, those operations that are time-sensitive such as mixing or blending times, and drying times should be supported by data that are obtained through developmental studies. Arbitrarily assigning time limits without the benefit of any data, or the working knowledge of the operation in question, should be avoided, and can result in costly developmental errors and product delays.

Reprocessing can be defined as the repeating of a defined step or sequence of steps outlined in a master production and control record, in order to achieve a predetermined endpoint. Written procedures should be established and followed for any reprocessing steps or operations that are incorporated into a developed manufacturing process. The effectiveness of any reprocessing activity should be demonstrated through validation studies and should be fully supported by data.

Sterility Assurance

During inspections, FDA places a great deal of emphasis on the following systems they feel most directly will impact the safety and efficacy of sterile products manufactured,^[10] namely:

- Sterilization processes for the drug itself, any components, container/closures, product-contact equipment, and surfaces.
- Depyrogenation processes.^[11]
- Water systems.
- Air handling systems.
- Environmental monitoring programs.
- Handling of incoming components.
- Packaging and labeling operations.
- Laboratory controls.
- Lyophilization (if applicable).



In general, the above-mentioned, including procedures, systems and, if applicable, validation should be in place for clinical supply manufacturing operations. This would include validation of processes such as sterilization, depyrogenation, and lyophilization; qualification of water and air handling systems; establishing and following procedures for: environmental monitoring programs, handling of incoming components, packaging, and labeling operations; and CGMP compliance for laboratories performing analyses of raw materials, in-process and final product samples, environmental monitoring activities, and testing in support of clinical supply manufacturing. Additionally, those processes or activities that are product-specific (e.g., sterilization, depyrogenation, the manufacturing process for the finished drug product itself, and analytical testing related to raw materials, in-process and final product testing) should be validated to what has been previously noted as the “extent possible.” For example, in the case of sterilization processes such as terminal sterilization using moist heat, all of the following validation and process control requirements would be necessary in order to avoid exposing a patient to a risk of nonsterility:

- Equipment that is qualified, maintained, and calibrated.
- Empty chamber heat distribution studies.
- Heat penetration studies that are product and load pattern-specific.
- Challenges with biological indicators.
- Process controls and monitoring typical for steam sterilization (e.g., time, temperature, and pressure).^[6]

Similar logic can be applied to sterilization by filtration or depyrogenation by means of dry heat. In these and other cases, there will be little to no difference in the level of compliance necessary when comparing clinical supply manufacturing to that for the manufacturing for commercially approved products.

Change Control

Change control is one of the GMP systems that is applied somewhat differently in clinical supply manufacture than it is in the manufacture of commercial products. Change control does not have a separate and distinct section in the GMP regulations; however, the management of change is referenced or mentioned in several GMP sections [see, for example, 21 CFR 211.100(a) and 211.160(a)].

In commercial manufacturing, the purpose of change control is twofold:

1. Keep validated systems functioning in a validated state.

2. Maintain the accuracy of approved submissions to regulatory agencies (such as NDAs, ANDAs, BLAs, etc.).

In the manufacture of clinical supplies, normally, manufacturing processes are not robust (and therefore not completely validated) until Phase III, sometimes not until the later stages of Phase III. Therefore, the goals of change control in clinical supply manufacture are somewhat different. Those goals should include:

- Assurance that equipment (including computer systems and software), which has been qualified (installation, operational, and performance qualification), is maintained in a qualified state.
- Support systems (such as water for injection, air handling systems, compressed gases, vacuum, clean-in-place systems, and others) are likewise maintained in a qualified state.
- Any significant modification of the Chemistry, Manufacturing, and Controls section of the IND is noted, and, where applicable, the approved submission is supplemented to provide for the change.
- Key changes in the development of formulation, dosage form, manufacturing process, cleaning procedures, and other key operations are identified for the purpose of documenting them as part of the development history.
- When significant numbers of batches are manufactured using replicate processes (usually in late Phase III trials), changes are evaluated in order to determine if process development is complete, and validation should commence.

Laboratory Controls

The majority of 21 CFR 211.160 (laboratory controls) and subsections can be applied in the same manner for clinical supplies as for commercial products. An exception, in many cases, is the time by which analytical methods validation is required. For new chemical entities or significant formulation changes, new analytical methods may need to be developed. As with manufacturing processes, until such methods are robust it is difficult (or impossible) to validate them to the full extent that is expected for commercial products.

The factors normally considered in analytical methods validation include:

- Specificity.
- Sensitivity.
- Linearity.



- Repeatability.
- Robustness.
- Ruggedness.
- Limit of detection.
- Limit of quantitation.

Analytical method validation should track closely to the stages of development of the method itself. However, it is not realistic to expect complete and thorough validation of the method until its development cycle is complete. An exception to this would be a situation where an accepted compendial method is applied to clinical material (such as a dissolution test or release testing of a compendial component). In these cases, companies must be prepared to demonstrate that consistent acceptable results can be obtained when using the compendial method in the company's laboratory (also known as methods verification). The obvious difference is that a compendial method is not a proprietary method, having been developed fully in a collaborative process. Therefore, transfer of a compendial method to a clinical supply context does not differ from a similar transfer to a commercial product.

Methods for the investigation of out-of-specification (OOS) results should be similar to what is required for commercial manufacturing. However, it is recognized that in many cases, specifications may be less exact and methods may still be under development; therefore, it may be considerably more difficult to determine assignable causes for OOS findings related to clinical supplies.

When considering OOS investigation procedures for clinical supplies, manufacturers must keep certain basic principles in mind. Those basic principles include:

- OOS results must not be arbitrarily dismissed simply because they are unexpected and present obstacles to release of materials.
- Consideration must be given as to whether analyst error, equipment malfunction, inappropriate reagents, or some other detectable and assignable cause was the reason for an OOS result.
- Consideration must be given as to whether a detectable process error or nonprocess related (employee) error was the reason for the OOS result.
- There must be a documented rationale for resampling or retesting, including documentation of the reasonableness of the number of retests that are needed to overcome an OOS finding.
- Averaging of results should be avoided when the purpose of the test is to reveal variability in a batch (such as blend uniformity or content uniformity testing) because averaging tends to hide variability.

- If averaging is employed, OOS results must be included along with within-specification findings, unless they may be discarded by an approved statistical outlier test (either approved in the compendia or in the IND).

Training records for analysts should demonstrate that they have the necessary combination of education, training, and experience to perform the specific methods they are assigned to perform.

Stability testing will be guided mainly by the conditions approved in the IND. However, the manner in which the stability program is managed should be similar to that required for commercial products. For example (not an all inclusive list):

- Procedures to ensure that stability samples are correctly collected and are representative of the batch.
- Procedures to ensure that chambers are qualified (temperature mapped) and monitored for environmental parameters.
- Procedures to ensure that test intervals are consistently met.
- Procedures for the development of stability-indicating analytical methods.
- Complete, clear, and accessible records of all testing, including the preservation of raw data.

Deviations and Failure Investigations

As with commercial manufacturing, product or process deviations, or the failure of any clinical batch or any of its components to meet any of its specifications must be handled by approved procedures and thoroughly investigated (21 CFR 211.192). The investigation report should include, among other things:

- Nature of the deviation or failure.
- Date.
- Affected batch or batches.
- Root cause determination.
- Impact analysis.
- Recommended corrective actions.
- Approval by the Quality Unit.

Apart from being product or process-related, deviations can also be procedural in nature, meaning that certain requirements of a specific SOP were not adhered to. Sometimes the terms "unplanned" and "planned" deviations are also used. Unplanned deviations include those events, activities, or actions that are nonintentional. An example of an unplanned deviation could include shutdown of processing equipment during the manufacture



of a batch due to safety concerns. Planned deviation, on the other hand, is a term that is not favorably looked upon by FDA. A conscious decision not to follow written procedures, manufacturing instructions, or other records required under the predicate rule is, at the very least, a CGMP issue. The FDA expects that deviations will be fully and thoroughly documented, including any decisions made and by whom.

In the clinical supply setting, systems and procedures should be established and followed for handling deviations and performing investigations. For deviations, this would also include trending (either according to defined categories such as equipment, process, etc. by product line or other predefined criteria), timely resolution of the issues, and re-evaluation of process or equipment parameters and specifications, procedural requirements, or other items bearing on the quality or purity of the finished drug product. This re-evaluation is especially important during new product development, where less information may be known about the drug product being studied, the method of manufacture, process capabilities, or sensitivities of test methods. In these and other cases, improperly or incorrectly defined product or process attributes could contribute to an artificially high deviation rate, and more importantly, could suggest that additional product development work is needed.

Complaints and Adverse Reactions

Section 211.198 of the GMP regulations sets forth the requirements for reporting and investigating complaints. This section is intended to assure that when a manufacturer is notified of a product defect, a thorough investigation will be performed to determine the cause of the defect, and whether a recall of the product from the marketplace should be initiated. Of course, these activities are all aimed at protection of patients from harm caused by defective products.

When the product in question is intended only for use in clinical trials, the fundamental purpose of the regulation, protection of patients from defective products, must still be of primary concern. However, the quantity of product at risk, and its limited distribution will impact the process of complaint investigation. For example, the generally smaller amount of product involved may make it difficult to identify low-level occurrences of manufacturing defects. The limited and more tightly controlled system of distribution of clinical trial supplies, however, will likely facilitate and limit the recall process should that be deemed necessary.

Complaints of adverse reactions to clinical supplies are typically seen as reportable events that impact the safety or

efficacy of the test article. Companies should give consideration to the possibility that an increase in the frequency or severity of adverse reactions may be linked to a manufacturing defect. For example, a product that has been formulated such that it is superpotent in its final form can be expected to produce adverse reactions of greater severity or frequency than if it was properly formulated. Labeling mix-ups can have devastating impact on patients, or, at minimum, may seriously affect the validity of a clinical trial. Therefore, the following elements should be included in the complaint assessment system:

- Complaints of observable (physical) defects should be promptly and thoroughly investigated to determine the root cause of the defect and the probable scope of its occurrence.
- Complaints of unusual or severe adverse reactions, or significant and unexpected increases in the number of adverse reactions (including marked lack of efficacy) should be evaluated to determine whether manufacturing errors may have caused or contributed to the observed effects.
- When a product defect is identified, the evaluation should include a health hazard assessment by a qualified person, usually a physician. The FDA's recall policy regulations contain a suggested rubric for performing a health hazard assessment (see 21 CFR 7.41).
- Product defect complaints and adverse reaction data and trends should be periodically compared to determine if there is any correlation between unexpected numbers, types or severity of adverse reactions, and the number and type of reported product defects.
- As with commercial products, careful records of reported complaints, their investigation and resolution are critical to GMP compliance.

Technology Transfer Plans and Reports

Technology transfer plans and reports are used by a research and development organization in order to document their official "transfer" of a newly developed or recently upgraded product/process from the developmental area and facilities to an operations unit and site, usually located in separate buildings, or at an entirely separate and different manufacturing sites. While the format and content has not been formally prescribed by FDA, companies generally include the following information^[12] in these documents:

- Definition of responsibilities for key departments; documentation review, approval, and storage requirements.
- Summary of developmental activities completed.



- Summary of scale-up activities; summary of formulation, synthesis, process, and analytical method changes; establishment of impurity specifications; establishment of critical process parameters; identification of validation plans and activities.
- Definition of component, container/closure, and product attributes.
- Analytical methods development and validation summary.
- Descriptions, specifications, design parameters and requirements of facilities, and major equipment and utility systems.
- Definition of the manufacturing process.
- Stability and expiry dating information.
- Change control.
- Reprocessing.
- Cleaning processes, including methods development and validation of the processes and methods.
- Summary of Regulatory Affairs activities, including regulatory commitments, key data and information to be summarized in regulatory filings.

Data Integrity Considerations

The integrity of GMP data should be a vital part of any company's compliance scheme. The term "data integrity" generally means that the raw (source) data and associated data summaries and reports are truthful, accurate, legible, indelible, complete, and readily accessible. This applies whether the data are in hard copy (paper) form or electronic form. The GMP rules allow for the preservation of data through photocopying or other optical reproduction process (microfilming, optically scanning, etc.); however, original documents should be preserved whenever possible.

The fundamentals of good data recording practices should be observed in clinical supply manufacturing in exactly the same manner as required for commercial materials. Those fundamentals include the following:

- Data, and the identification of those responsible for it, should be recorded contemporaneously, or as close to the time of execution of a step or an observation as is practical. Transfer of data from one form to another should be minimized, and should include one hundred percent verification for accuracy by a second person. This serves to enhance the accuracy of the data.
- Hard copy documents should be recorded in indelible ink. Errors are inevitable, and when they occur, the original entry should be struck out with a single line, initialed and dated, the correct information entered, and a brief explanation of the error should be appended

as directed by the document and associated procedures.

- Data should always be recorded on forms designed for the specific purpose. Extraneous notations on loose slips of paper (including post-it notes, scrap paper, paper towels, etc.) should be strictly prohibited in a GMP environment.
- Log books, laboratory notebooks (when used), batch production and control records, qualification and validation protocols and reports, investigations of deviations and failures, and other GMP documentation must be carefully identified, controlled, and archived in a manner that will facilitate its prompt retrieval.
- Electronic records (and electronic signatures, if used) must comply with the provisions of 21 CFR Part 11. It is beyond the scope of this article to explain "Part 11" compliance in total, however, the general principle is to construct and maintain electronic records with the same degree of accuracy and linkage to the originator as is achieved with hard copy documents. Each company choosing to use electronic data capture techniques and electronic signatures must thoroughly understand Part 11 and ensure compliance with this regulation.
- Each company should have a policy that strictly prohibits the deliberate falsification, mutilation, obliteration, or destruction of raw data and associated GMP documentation. Companies must insist on strict adherence to such policies and should take aggressive disciplinary action when lapses are detected. Failure to do so may subject the company and its corporate officers to severe regulatory sanctions, including criminal prosecution under the Federal Food, Drug, and Cosmetic Act, or the general criminal laws of the United States (Title 18, U.S. Code). Even inadvertent (non deliberate) acts that result in loss of data or records should be prevented, and, if they occur, they should be promptly and thoroughly investigated.
- Establishing and following policies and procedures for good documentation practices, numerical rounding, significant figures, and directed audits to ensure the integrity of data contained in records or other documents prior to submission of applications to FDA).

REFERENCES

1. *Human Drug CGMP Notes*; FDA, CDER, December 1996.
2. *Human Drug CGMP Notes*; FDA, CDER, September 1999.
3. Human and Veterinary Drugs, Current Good Manufacturing Practice in Manufacture, Processing, Packaging or Holding. Federal Register **1998**, 43 (190), 45013–45336.
4. *Human Drug CGMP Notes*; FDA, CDER, June 1994.



5. *Code of Federal Regulations, Food and Drugs*; Current Good Manufacturing Practice for Finished Pharmaceuticals, Title 21, Part 211; August U.S. Printing Office: Washington DC, August 1996.
6. Levchuk, J. *FDA Perspectives on Production of Parenteral Clinical Supplies*; International Good Manufacturing Practices Conference, Mar University of Georgia: Athens, GA, Mar 5, 1997.
7. European Commission, Enterprise Directorate General. *Good Manufacturing Practices, Annex 13, Manufacture of Investigational Medicinal Products*, Vol. 4; November 2001; Draft 1.
8. *Guideline on the Preparation of Investigational New Drug Products*; FDA, CDER, March 1991.
9. *Guideline on General Principles of Process Validation*; FDA, CDER, May 1990.
10. Compliance Program Guidance Manual, 7356.002A, *Sterile Drug Process Inspections*; FDA, CDER, October 1990.
11. *Guideline on Sterile Drug Products Produced by Aseptic Processing*; FDA, CDER, June 1991.
12. Tetzlaff, R. *Preparing for an FDA Pre-approval Inspection*; March Parenteral Drug Association: San Diego, CA, 1997; 20–21.

CLINICAL PHARMACOKINETICS AND PHARMACODYNAMICS

Leslie Z. Benet
Laviero Mancinelli

University of California San Francisco, San Francisco, California, U.S.A.

INTRODUCTION

Therapeutics relates drug administration to physiological effects, both beneficial and toxic, in patient populations. The important intervening steps or processes, which occur following drug dosing that lead to therapeutic effects in the body, are described in Fig. 1. This scheme illustrates the dose-effect relationship. The discipline of pharmacokinetics provides an understanding of drug absorption, disposition, and elimination, as these relate to the physiological effects of drugs. Pharmacodynamics relates the measured response, either efficacious, toxic, or both, to the drug's pharmacokinetics. Pharmacokinetics has become increasingly important, as it has proven to be a very useful tool in understanding not only drug dosage in relation to achievable drug concentrations in plasma, but also the influence of disease states on the behavior of a drug in the body. A very powerful interpretative instrument becomes available when the disciplines of pharmacokinetics and pharmacodynamics are combined. Clinical pharmacokinetics specifically deals with optimizing drug dosage in individual patients. By using plasma concentration measurements, and clinical pharmacokinetic principles, doses can be adjusted to achieve maximal therapeutic utility, with minimum toxic risks for individual patients.

Basic knowledge of pharmacokinetic principles is important not only for the clinician to understand basic drug-patient interrelationships, but also for the drug industry to facilitate the design of relevant studies for drug discovery candidates. The gain from designing proper pharmacokinetic studies can probably be counted not only in better therapy for the patient but also in long-term economic savings for the pharmaceutical industry.

In this article the basic principles of pharmacokinetics and pharmacodynamics will be addressed, and examples of how these principles can be used to increase the understanding of drug therapy and drug dosage formulation are given.

PHARMACODYNAMIC CONSIDERATIONS

The fundamental hypothesis in clinical pharmacokinetics asserts that a relationship exists between a drug concentration in some measurable biologic fluid and observed drug effects, both therapeutic and/or toxic, as illustrated in Fig. 1. Many drugs, for example, antiarrhythmics, anticonvulsants such as phenytoin, and diuretics such as furosemide, show an apparent direct relationship; that is with every change in plasma drug concentration, there is an immediate and corresponding change in effect. For other drugs the relationship is more complicated; the drug level of warfarin for instance, triggers a cascade of events leading to prolonged clotting time. Here, the mean steady-state level rather than the time course of concentrations can be most closely correlated to the therapeutic effect. In some instances, the effect itself is difficult to quantify, and demonstrating any relationship is difficult. When unbound concentrations are maintained at a particular steady-state concentration across a population, then an observed difference in response between individuals must be due to pharmacodynamic receptor-level variability across the population. However, most traditional relationships of drug effects have been developed relative to drug dose rather than steady-state unbound concentrations. When comparing the dose given to the effect, the variability in drug absorption, first-pass metabolism, protein binding, and clearance must also be included, thereby giving a significantly more variable and complicated relationship.

The full model for the concentration (C) and effect (E) relationship is the Hill equation, also called the extended E_{\max} model (1):

$$E = \frac{E_{\max} C^s}{EC_{50}^s + C^s} \quad (1)$$

where E_{\max} is the maximal effect, EC_{50} is the concentration giving 50% of maximal effect, and s is the slope factor. The Hill equation can be simplified to different extents, for example, if $s = 1$.

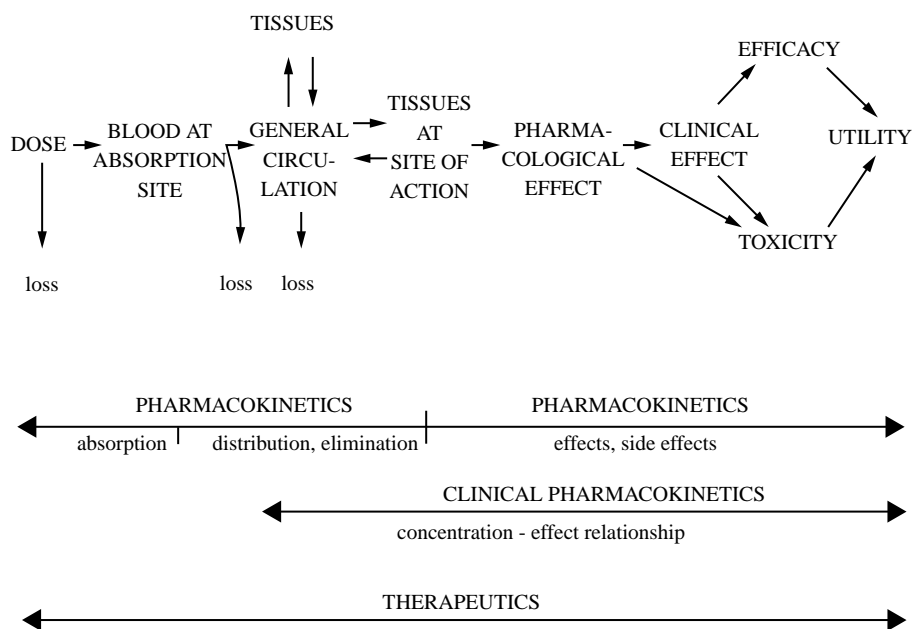


Fig. 1 Schematic representation of the dose-effect relationship for a drug.

Between 20 and 80% of the maximal response, the relationship is logarithmic:

$$E = m \log C + e \quad (2)$$

where m is the slope and e is the intercept. This equation has the limitations of not being able to predict effects outside this range of responses (i.e., <20% and >80%).

Below 20% of maximal response, the relationship is linear:

$$E = mC + e \quad (3)$$

This relationship has been used for up to 50% of maximal response; however, it then is a coarse approximation, depending upon the value of s (the lower the value of s , the better the approximation).

The Hill equation should be used only when a reasonable estimation of E_{\max} is possible and when data are gathered over the whole effect range. If data are available only from the lower part of the curve, the linear equation might just as well be used. The difference between the Hill equation and the linear or log-linear equation is that only the Hill equation gives a mechanistic-physiologic understanding of the effect. The other equations are merely descriptive.

Sheiner and associates (2) have developed the relationship where the pharmacodynamic model (Eq. 1) is integrated together with the pharmacokinetic model. This makes possible describing both the steady-state

concentration-effect relationship (i.e., EC_{20} , EC_{50} , EC_{90} , etc.) and the time lag between the measured rapidly changing plasma concentration and the corresponding steady-state effect at that concentration level. This time-lag parameter is described by the equilibration half-time. Using both these parameters, that is, the expected effect and the time required to obtain the effect, allows investigators to model the effect-concentration relationship when patients are not at steady state.

The relationship between the measured effect and steady-state plasma concentration sometimes yields bell-shaped (e.g., nortriptyline) or U-shaped (e.g., clonidine) curves. Several unusual concentration-effect curves, including the U-shaped curve describing the blood pressure lowering effect of clonidine, have been explained by Paalzow and associates (3) as being a result of multiple receptor responses. The drug then acts on several different receptors that can have opposite effects and that are triggered at different concentrations.

Tolerance to the drug effect, that is, a decrease in the effect with time, also obscures the dose-response relationship. Pharmacokinetic and/or pharmacodynamic causes for tolerance development are possible. Pharmacokinetic tolerance, for example, can be caused by induction of metabolic enzymes, thereby causing a decrease in drug concentrations. Pharmacodynamic tolerance can be characterized mainly in two different ways: the receptors down-regulate in response to the drug, giving a smaller response with time, or other physiologic

mechanisms counteract the drug effect. The blood pressure lowering effect of hydralazine that is diminished by a compensatory increase in heart rate, and the diuretic effect of furosemide that is decreased as a result of the drug's volume and salt-depleting actions are two examples.

Fig. 2 illustrates a utility curve, that is, a curve describing the clinical utility of a drug in terms of the risk of side effects from a high concentration and the risk of no effect from a low concentration. The closer the effect and toxicity curves, the more narrow is the range of plasma concentrations that can be used for therapy. The utility is obtained as the difference between the effect and toxicity.

Depending upon the steepness of the concentration–effect relationship (the size of s in Eq. 1), an increase in concentration will result in different changes in effect. An all-or-none relationship is obtained when the curve is very steep ($s > 6$). Theophylline shows a shallow relationship with plasma concentration for its antiasthmatic effect; a big increase in concentration results in a small increase in effect. However, as the side effects of theophylline show a much steeper relationship with plasma concentration, it is critical not to increase theophylline concentrations above 20 mg/L. With a lower limit for antiasthmatic effect of

10 mg/L, theophylline exhibits a narrow range of concentrations where therapy is beneficial.

Another measure of utility is in terms of the ratio between the concentration level causing an undesirable side effect to the concentration level giving the desired therapeutic effect (therapeutic index). For theophylline this ratio is 2, for digoxin it is 1.6, and for furosemide it is 40. The higher the therapeutic index, the less critical are dosing recommendations with respect to the risk of serious side effects.

CLEARANCE

Clearance is the measure of the ability of the body to eliminate a drug, and as such is one of the most important pharmacokinetic parameters, as it gives a well-defined, physiologically relevant measurement of how drugs are eliminated by the organism as a whole, or by a particular organ. Clearance relates drug concentration to the elimination rate from the body (Eq. 4), or at steady state the average concentration C_{ss} to the dosing rate because at steady state the input rate into the body will equal the output rate (Eq. (5)):

$$\text{Elimination rate} = CL C \quad (4)$$

$$\text{Dosing rate} = CL C_{ss} \quad (5)$$

The clearance concept has been used in defining the pharmacokinetics of drugs since the mid-1970s (4, 5). The clearance concept is based in physiology, where it is used as a measure of renal function (creatinine clearance). Creatinine is formed from muscle breakdown at a constant rate, and thus a constant creatinine concentration in plasma results. The magnitude of this concentration is dependent on the elimination rate of creatinine and the size of the muscle pool (formation rate). By measuring the plasma concentration and the renal excretion of creatinine, renal clearance can be estimated and thereby kidney function indicated, as creatinine is mainly filtered into the urine

$$CL_{\text{creatinine}} = \frac{\text{Urine volume} \times \text{Urine concentration}}{\text{Plasma concentration}} \quad (6)$$

Drugs are not only eliminated via the kidneys but also eliminated in the bile by the liver and metabolized in the liver and elsewhere, which makes direct measurement of the elimination rate of a drug difficult. Indeed, other routes of elimination could include loss in expired air, saliva, sweat, partition into tissue stores, efflux from the blood into the gut lumen, and gut metabolism as well as

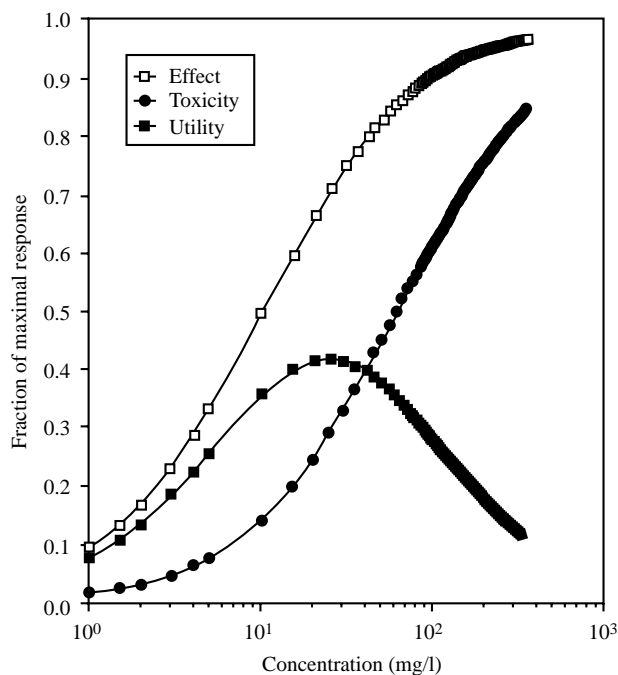


Fig. 2 Utility curve and concentration–effect/toxicity relationships for a theoretical drug according to Eq. (1). EC_{50} for efficacy = 10 mg/L, EC_{50} for toxicity = 60 mg/L, thus the therapeutic index is 6; $s = 1.0$. Utility is obtained as the difference between effect and toxicity.

other sites of metabolism such as the lung. The total clearance, CL, can be defined as

$$CL = \frac{\text{Dose}_{i.v.}}{\text{AUC}} \quad (7)$$

where $\text{Dose}_{i.v.}$ is the intravenous dose, and AUC is the resulting area under the plasma concentration time curve. Total clearance can also be measured during continuous drug therapy as dosing rate divided by C_{ss} according to Eq. 5.

Clearance is referenced to plasma (plasma clearance, CL_p), blood (blood clearance, CL_b) or plasma water (unbound clearance, CL_u), depending upon where the concentration is measured. Total clearance can be divided into the contributions of each of the eliminating organs, the most important being renal clearance, CL_R , and hepatic clearance, CL_H .

$$CL = CL_R + CL_H + CL_{\text{gut}} + CL_{\text{other}} \quad (8)$$

Renal clearance can be separately determined by measuring the excretion rate of unchanged drug into the urine, as for creatinine. The difference between total clearance and renal clearance is usually called *nonrenal clearance*, meaning the clearance that is not accounted for by excretion of unchanged drug into urine, be it metabolism in the liver or elimination by any other organ.

Clearance is measured in units of volume per time (ml/min or L/h) and thus is defined using the same units as blood or plasma flow. By definition, clearance gives the volume of plasma (blood) from which a drug is completely removed per unit time. For some drugs the liver or kidneys have the ability to clear drug from all the blood flowing through the organ; for example, *p*-aminohippuric acid (PAH) has a renal plasma clearance of 600–700 ml/min, which equals renal plasma flow. Because PAH does not partition into red blood cells, PAH renal blood clearance will equal renal blood flow (see imipramine in the next paragraph for calculations according to Eq. 9, $C_{RBC}/C_p = 0$). So, when the eliminating organ has a high capacity to eliminate a drug, the blood clearance equals the blood flow to that organ. As the organ cannot eliminate drug any faster than the rate at which drug is presented to the organ, blood flow becomes the limiting value for clearance. Thus, hepatic blood clearance cannot exceed 1.5 L/min, and renal blood clearance cannot exceed 1.2–1.3 L/min.

Some drugs show a higher plasma clearance than the corresponding plasma flow through the eliminating organ. Imipramine has a plasma clearance of 1050 ml/min (6), thus exceeding the rate of plasma flow to the liver, where it is predominantly eliminated. Looking at whole blood, the concentration of imipramine in blood is higher than its concentration in plasma because of a considerable

partitioning into red blood cells ($C_{RBC}/C_p = 2.7$). Thus, the amount of drug delivered to the liver by the blood is much higher than that assumed from measuring its plasma concentration alone. The relationship between plasma and blood clearance at steady state is given by

$$\frac{CL_p}{CL_b} = \frac{C_b}{C_p} = 1 + H \left(\frac{C_{RBC}}{C_p} - 1 \right) \quad (9)$$

Imipramine blood clearance can be calculated by substituting the red blood cell to plasma concentration ratio and the average value for hematocrit ($H = 0.45$) into Eq. 10). The resulting blood clearance is calculated to be 595 ml/min, a value within the physiologic range of liver blood flow. The higher plasma versus blood clearance for imipramine also indicates that the drug present in the red blood cells is readily available for the metabolizing enzymes by being in rapid equilibrium with the drug present in plasma water. Thus, the plasma clearance may assume values that are not “physiologic.” However, if the concentration in blood is used to define clearance, the maximal clearance possible is equal to the sum of blood flow to the various organs of elimination. If a drug shows a higher blood clearance than the combined blood flow, a probable cause is extrahepatic or extrarenal elimination, such as the metabolism of nitroglycerin in blood and tissues. The *rate of elimination* of a compound by an organ is the difference between the rate of presentation to the organ and the rate of exit from the organ. *Rate of presentation* equals the organ blood flow multiplied by the entering concentration (QC_{in}), and the *rate of exit* equals the blood flow multiplied by the exiting drug concentration (QC_{out}) (Fig. 3).

$$\text{Rate of elimination} = QC_{in} - QC_{out} \quad (10)$$

By relating the rate of elimination to the entering concentration (Eq. 4), an expression for organ clearance of drug can be obtained

$$CL_{\text{organ}} = \frac{Q(C_{in} - C_{out})}{C_{in}} = QER \quad (11)$$

in terms of blood flow and ER, the extraction ratio, which equals $(C_{in} - C_{out})/C_{in}$. A high organ clearance signifies that the organ has a high capacity to extract drug from the blood.

For a compound for which the organ has a high extraction capacity, the exiting concentration is very low relative to the entering concentration, and ER approaches unity. For an organ with a low extraction capacity, the exiting concentration is not very different from the entering concentration, and ER approaches zero. Extraction ratio can thus be expressed as the fraction of the organ blood flow

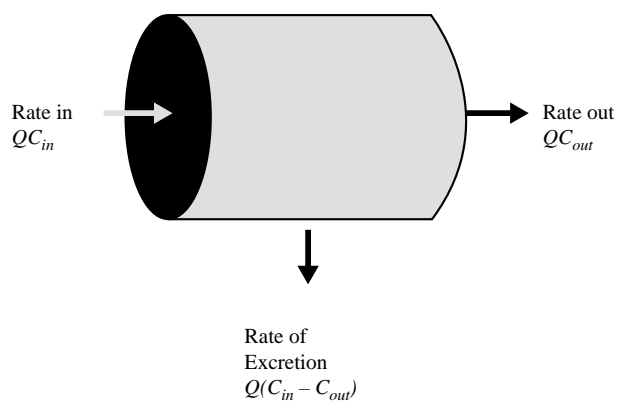


Fig. 3 Extraction of drug in an organ of elimination.

that is cleared by the organ ($ER = CL_{\text{organ}}/Q_{\text{organ}}$). Several models relating the hepatic clearance to physiologic parameters have been suggested, including the well-stirred and the parallel tube models (7, 8), the distributed model (9, 10), and the dispersion model (11–14). The well-stirred model is the most easily understood and most commonly employed model, and the following discussion will be based on this approach.

The well-stirred model assumes that the unbound drug concentration leaving the organ, for example the liver, is equal to the unbound concentration within the organ, thus assuming an instant “mixing” at the entrance of drug into the organ. The intrinsic ability to metabolize or otherwise clear unbound drug and the protein binding (unbound fraction, f_u) together with the blood flow through the liver (Q_H) all contribute to the resulting hepatic clearance, which may be mathematically expressed as

$$CL_H = Q_H \frac{f_u CL_{u,\text{int}}}{Q_H + f_u CL_{u,\text{int}}} \quad (12)$$

Thus, an example of a high-extraction drug is when the capacity of the liver to metabolize a drug is large compared to the rate at which drug enters the liver ($f_u CL_{u,\text{int}} \gg Q_H$), clearance approximates liver blood flow

$$CL \cong Q_H \quad (13)$$

An example of low-extraction drug is when the capacity of the liver to metabolize a drug is small relative to the rate of presentation ($f_u CL_{u,\text{int}} \ll Q_H$), because of a low intrinsic ability to metabolize or because of diffusion problems to the enzyme site. Under these conditions, hepatic clearance approximates

$$CL \cong f_u CL_{u,\text{int}} \quad (14)$$

Between these extremes, hepatic clearance is dependent on all three factors: Unbound fraction of drug, intrinsic

clearance, and hepatic blood flow. The unbound clearance, CL_u , equals CL/f_u .

A reasonable assumption is that the active secretion mechanism in the kidney can also be described by the well-stirred model. However, the kidneys have several mechanisms that may determine renal clearance of a drug, including passive filtration and reabsorption.

The renal clearance is the sum of filtration and secretion minus reabsorption. It can be described as

$$CL_R = CL_{\text{filtration}} + CL_{\text{secretion}} - CL_{\text{reabsorption}} \quad (15)$$

or,

$$CL_R = (CL_{\text{filtration}} + CL_{\text{secretion}}) (1 - \text{fraction reabsorbed}) \quad (16)$$

Filtration in the kidneys is determined by the glomerular filtration rate, GFR (120 ml/min in a healthy young person), and by the fraction of the drug that is dissolved in plasma water, that is, the unbound fraction, f_u . Thus,

$$CL_{\text{filtration}} = f_u \text{ GFR} \quad (17)$$

Suppositions regarding the mechanisms by which a drug is renally eliminated can be made by comparing the filtration clearance with the measured renal clearance. Drug is always passively filtered to some extent, depending of the value of f_u . If $CL_R > CL_{\text{filtration}}$, the drug is also secreted and may be reabsorbed but to a smaller extent than it is secreted. If $CL_R < CL_{\text{filtration}}$, however, this is an indication of reabsorption but does not exclude secretion. If $CL_R = CL_{\text{filtration}}$, the drug can still be both secreted and reabsorbed, but to the same extent. Conclusive evidence on whether a drug is reabsorbed cannot be made solely on the basis of clearance values.

Intestinal Metabolism or Clearance

Recently, it has been recognized that small intestinal metabolism and active efflux of orally absorbed drugs, are major determinants of oral drug bioavailability (15). Many elements, including patient specific ones, determine the extent of oral drug delivery. The observed oral bioavailability (F_{oral}) of any particular drug is dependent upon the following processes: delivery to the intestine (gastric emptying, pH, presence of food), absorption from the lumen of the intestine (dissolution, lipophilicity, particle size, active uptake), intestinal metabolism (phase I/phase II), active extrusion (drug efflux pumps), and then first pass hepatic metabolism (16).

The enzymes of the cytochrome P 4503A family are the predominant phase I drug metabolizing enzymes in man. The major isoform of the CYP3A family is CYP3A4, the

predominant form found in adult human liver and small intestine. Members of the 3A4 family are estimated to be responsible for the metabolism of more than one half of all drugs that are substrates for the P450 system of metabolic enzymes in man (17). The levels of CYP3A4 found in human liver and small intestine is highly variable, with 10–100 fold variations observed in liver and as much as 30-fold variation in the small intestine, respectively (18). Levels of CYP3A4 in the small intestine are generally 10–50% of the levels found in the human liver, and these levels as well as the activity of the enzyme decrease longitudinally along the small intestine (19). The enzymes of the CYP3A4 subfamily comprise approximately 30% of all hepatic cytochromes and at least 70% of all intestinal cytochromes responsible for drug metabolism (20).

Previously, drug absorption from the gut was assumed to occur by passive processes, and little attention was paid to the activity of counter transport systems. It has now been recognized that P-glycoprotein (P-gp), an ATP-dependent efflux transporter, is expressed at high levels on the apical surface of the columnar epithelial cells in the jejunum of the small intestine (21). P-gp represents the best studied member of the ATP binding cassette (ABC) family of transporters, and is the product of the multidrug resistance gene MDR1 in man. P-gp is expressed in a wide variety of tissues including the adrenal glands, the bladder, the cells of the blood–brain barrier, kidney, liver, lungs, pancreas, rectum, spleen, and most importantly for the purpose of oral bioavailability, in the esophagus, stomach, jejunum, and colon (21–23). In an apparent contrast to the situation noted above for CYP3A, levels of P-gp increase longitudinally along the intestine, with the lowest levels found in the stomach and the highest levels found in the colon (23).

The co-importance of CYP3A and P-gp for the oral bioavailability of drugs is suggested by their shared locations within the enterocytes of the small bowel, as well as a significant overlap in their substrate specificities (21, 24–28). Intracellularly, P-gp is found traversing the plasma membrane of enterocytes while CYP3A is found within the cytoplasm on the endoplasmic reticulum. Although gene expression of these two proteins does not appear to be coordinately regulated (29, 30), their proximal spatial relationship suggests that P-gp may act to regulate exposure of drug (substrates) to metabolism by CYP3A. Repeated absorption and extrusion processes would then, result in repeated exposure of drug to CYP3A, resulting in enhanced overall metabolism, and correspondingly lower oral bioavailability, as depicted in Fig. 4. This spatial juxtaposition, coupled with the large number of overlaps between CYP3A and P-gp substrates, suggests a strong

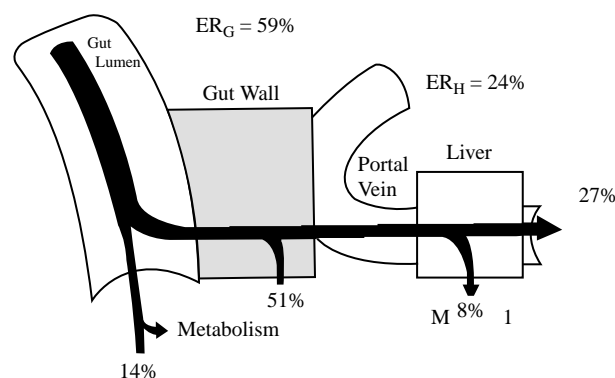


Fig. 4 This cartoon depicts the various processes leading to an oral bioavailability of 27%, following an oral dose of the Sandimmune® formulation of cyclosporine (15). The values at the bottom of the figure indicate the average fraction of the dose lost in each of the processes; i.e., 14% of the dose is either unabsorbed, counter transported (effluxed) by P-gp, or degraded in the gut lumen, 51% of the drug is metabolized in the enterocytes of the gut wall, and only 8% is lost due to hepatic first-pass metabolism.

complementary role for these proteins in the pharmacokinetics of drug absorption.

Oral bioavailability (F_{oral}) is equal to the product of the fraction of dose absorbed (F_{abs}), the fraction of the absorbed dose which passes into the hepatic portal blood flow unmetabolized (F_G), and the hepatic first-pass availability (F_H), as seen in Eq. 18.

$$F_{\text{oral}} = F_{\text{abs}} F_G F_H \quad (18)$$

Gut and hepatic availability may be defined as one minus the extraction ratio (ER) at each site.

$$F_{\text{oral}} = F_{\text{abs}}(1 - \text{ER}_G)(1 - \text{ER}_H) \quad (19)$$

The hepatic extraction ratio (ER_H) can be determined after intravenous dosing from the ratio of hepatic clearance (CL_H) to hepatic blood flow (Q_H).

$$\text{ER}_H = \frac{\text{CL}_H}{Q_H} = \frac{f_u \text{CL}_{u,\text{int}}}{Q_H + f_u \text{CL}_{u,\text{int}}} \quad (20)$$

$$\text{So that } F_H = \frac{Q_H}{Q_H + f_u \text{CL}_{u,\text{int}}} \quad (21)$$

The ability to estimate the gut extraction ratio requires further experimental manipulations and assumptions, such as evaluating the effects of an enzyme inducer, e.g., rifampin, on oral and i.v. drug dosing as first utilized by Wu et al. (31) for cyclosporine, or measurement of portal concentrations as first utilized by Thummel and co-workers (19) for midazolam.

It is apparent that the greatest impact on oral bioavailability, that would result from the concerted activity of intestinal CYP3A and P-gp, will be observed in drugs that are characterized by low to intermediate hepatic first-pass extraction. High hepatic extraction will probably obscure the gut effects. The inhibition or induction of intestinal CYP3A directly translates into changes in the oral bioavailability of drugs. Inhibition and induction of P-gp in contrast, manifests as a change in the rate of absorption, T_{\max} , which can also affect the extent of oral availability, since the T_{\max} changes reflects the access of the drug to the intestinal enzymes.

Many commonly prescribed drugs are joint substrates for CYP3A and P-gp, which as has been discussed, both reside in the human intestine, associates within the enterocytes. It is therefore very likely that changes in bioavailability for a number of drugs could result from intestinal metabolism and/or efflux counter transport of absorbed drug back into the intestinal lumen. Selective inhibition of one or both of these processes could theoretically increase bioavailability, while decreasing the variability inherent in absorption. An approach using specific modifiers of intestinal metabolic and counter transport activity could conceivably transform the therapeutic efficacy of many drugs now in use.

DISTRIBUTION

The systemic circulation transports drug molecules from the site of administration to all tissues and organs in the body. Depending upon the physicochemical properties of the drug (lipophilicity, degree of ionization), the drug partitions into different tissues to different extents.

The rate at which distribution takes place into a tissue is dependent on both the drug partition coefficient (concentration in tissue/concentration in blood at equilibrium) and the blood flow to that tissue. The higher the perfusion rate (blood flow per unit volume of tissue), the more rapid is equilibrium achieved between blood and tissue. The higher the partitioning into the tissue, the longer reaching equilibrium takes, as more drug has to be transported to the tissue.

Within the blood, drug is dissolved in the plasma water (unbound concentration, C_u). Drug can also be bound to plasma proteins and concentrated in the red blood cells. It is the unbound drug molecules that diffuse across the membranes into the tissues. At equilibrium, the unbound concentration of drug is thought to be the same throughout the whole body.

Volume of Distribution

To measure drug concentrations, a blood sample is centrifuged and the plasma concentration analyzed. The relationship between the amount of drug in the body and the concentration (C) is called volume of distribution:

$$V = \text{amount in body} / C \quad (22)$$

Depending on the fluid being measured, blood (C_b), plasma (C_p) or plasma water (C_u), different values for the volume term can be obtained as the concentrations in these fluids differ.

The volume of distribution is a proportionality constant. It can be smaller or larger than the true physiologic fluid spaces of the body, depending upon whether the affinity of the drug is highest in plasma constituents or in other tissues. For a normal 70-kg man, the plasma volume is 3 liters, blood volume is about 5.5 liters, extracellular fluid outside plasma is 12 liters, and total body water is approximately 42 liters.

To determine the volume of distribution of a drug, an i.v. dose is necessary. The volume of distribution can be calculated from the plasma concentration versus time data by means of noncompartmental methods as described by Benet and Galeazzi (32) for a bolus i.v. dose

$$V_{ss} = \frac{\text{Dose}_{i.v.} \cdot \text{AUMC}}{\text{AUC}^2} \quad (23)$$

Here V_{ss} represents the volume in which a drug would appear to be distributed during steady state, AUC is the area under the plasma concentration time curve, and AUMC is the area under the first moment of the plasma concentration time curve, that is, the area under the curve of the product of time, t , and plasma concentration, over the time span zero to infinity. The volume of distribution at steady state can also be determined by compartmental methods, that is, by using the coefficients and exponents of a multiexponential fit to the data (33).

An alternative measure of volume of distribution is V_{area} . This parameter is dependent on the terminal half-life or, expressed differently, the terminal rate constant λ_z , where λ_z is equal to $0.693 / t_{1/2}$

$$V_{\text{area}} = \frac{\text{Dose}_{i.v.}}{\lambda_z \cdot \text{AUC}} = \frac{\text{CL}}{\lambda_z} \quad (24)$$

as $\text{Dose}_{i.v.} / \text{AUC} = \text{CL}$, according to Eq. 7a. Although V_{area} is a convenient and easily calculated parameter, the value of V_{area} can show differences when the half-lives differ, for example, between different patient populations, without a true difference in the distribution space. Contrary to V_{area} , V_{ss} is theoretically independent of changes in elimination. Thus, to determine whether a particular disease state is

Table 1 Pharmacokinetic parameters for cefamandole in normals and uremics

Parameter	Normals ^a	Uremics ^b	Significance
CL (ml/min kg)	2.81 +/- 0.98	0.115 +/- 0.023	$p < 0.05$
$t_{1/2}$ (h)	1.2 +/- 0.2	13.0 +/- 4.5	$p < 0.05$
λ_z (h ⁻¹)	0.576 +/- 0.096	0.0534 +/- 0.0187	$p < 0.05$
V_{area} (L/kg)	0.298 +/- 0.104	0.138 +/- 0.048	$p < 0.05$
V_{ss} (L/kg)	0.161 +/- 0.050	0.134 +/- 0.045	n. s.

Data is +/- S.D.

^aData from Ref. (34).^bData from Ref. (35).

influencing the clearance process and/or the distribution of the drug, the V_{ss} volume term should preferentially be used. An example of the different conclusions that can be drawn, depending on which of these volume terms is used, is shown in Table 1. The pharmacokinetic parameters for cefamandole in six normal volunteers (34) and in three uremic patients (35) are compared. As cefamandole is almost exclusively eliminated via the kidneys, uremia results in a dramatic decrease in clearance (24-fold). The terminal half-life increases, but only by a factor of 11, resulting in more than a twofold difference in V_{area} between the two groups (Eq. 24). Based on these data, conclusions could be made that a decreased renal function not only influences the ability to excrete cefamandole but also results in a decrease in the distribution of the drug in the body. However, when comparing V_{ss} the conclusion is that no significant difference exists in the distribution of cefamandole between the two patient populations.

HALF-LIFE

Half-life is the "oldest" and best known of all pharmacokinetic parameters. It is a measure of the time required for the amount of drug in the body to decline to half of its value. Half-life is a useful measurement to determine the time to reach steady state for chronic dosing, or the time for the amount or concentration of drug to decline, for example, after an intoxication. To reach 90% of steady state or to eliminate 90% of the drug from the body takes 3.3 half-lives because 50% of the steady-state level is reached in one half-life, 75% in two, 87.5% in three, and 93.75% in four half-lives. The corresponding values hold for elimination of drug from the body.

Half-life can be readily determined from a plot of log plasma concentration versus time and was for many years considered to be the most important characteristic of a drug. Early studies examining drug disposition in disease states were compromised, by a reliance on half-life as a

sole measure of disposition changes. It is now appreciated that half-life is a secondary, derived parameter that relates to and depends on the primary parameters of clearance (CL) and volume of distribution (V) according to the following relationship in Eq. 25:

$$t_{1/2} \approx \frac{0.693 V}{\text{CL}} \quad (25)$$

Thus, to look at half-life only as a measure of, for instance, the effect of liver disease on drug pharmacokinetics is not sufficient, as a change in half-life can be caused by either a change in clearance or a change in volume of distribution. Furthermore, half-life may be unchanged in a particular disease state due to parallel changes in both V and CL.

Clearance and volume of distribution are two separate and independent characteristics of a drug. They are closely correlated with physiologic mechanisms in the organism (thereby the term primary parameters). Clearance defines the body's ability to remove the drug, that is, by metabolism or by renal or biliary excretion. Volume of distribution is a measure of the physical interrelationship between the drug and body constituents, such as binding to plasma proteins or partition into muscle, tissue, or fat.

PROTEIN BINDING

At steady state, the distribution of any drug in the body is dependant upon its binding to plasma proteins, blood cells and tissue receptors. Only unbound drug is capable of entering and exiting from plasma and tissue compartments. Therefore, an apparent volume of distribution can be expressed as follows (36),

$$V = V_p + V_{\text{TW}} \frac{f_u}{f_{u,T}} \quad (26)$$

where V_p represents the volume of plasma, V_{TW} is the volume of tissue fluid (nonplasma), f_u represents the

Table 2 Conditions that alter the concentration of two major plasma proteins

Plasma protein	Condition	Change in concentration
Albumin	Hepatic cirrhosis	Decrease
	Burns	
	Nephrotic syndrome	
	End stage renal disease	
	Pregnancy	
α_1 Acid glycoprotein	Myocardial infarction	Increase
	Surgery	
	Crohn's disease	
	Rheumatoid arthritis	
	Trauma	

(Adapted from Ref. 40.)

fraction unbound in plasma, and $f_{u,T}$ is the fraction unbound in tissue. Human plasma contains over 60 proteins. Albumin is the major component of this protein family responsible for the binding of most drugs in plasma. Acidic drugs bind primarily to albumin, the major drug-binding protein in plasma. Some acidic drugs bind with very high affinity, for example, furosemide, which is 98–99% bound and warfarin, which is 99.5% bound. Basic drugs bind to albumin with lower affinity but are more avidly bound to proteins like α_1 -acid glycoprotein and various plasma lipoproteins. These proteins have lower concentrations in plasma relative to albumin. Binding is therefore more easily saturable, as is the case with some drugs such as prednisolone and disopyramide, yielding fluctuations in the free fraction of drugs falling within the therapeutic plasma concentration range (6). Because the binding of drugs to plasma proteins and tissue binding sites is largely nonselective, many drugs with similar chemical properties can compete with each other for access to binding sites. The concern over the potential for adverse drug events based on competitive displacement from plasma protein binding sites has been overstated however. Indomethacin has been shown to markedly decrease warfarin binding to human serum albumin, in vitro (37). However, this in vitro drug interaction has not been confirmed by in vivo studies (38). Steady-state unbound drug concentrations in vivo are largely independent of factors which alter protein binding, unless a drug is very highly protein bound, for example, greater than 90% (39). A dynamic equilibrium exists between tissue and plasma stores of any drug. Binding at these sites is reversible and changes so rapidly that equilibrium is re-established within milliseconds. Therefore, changes in the free fraction of drugs brought about by competition with higher avidity ligands or saturable

kinetics are quickly compensated for by movement of drug from tissue stores into plasma, precluding the need for adjustment of dosage regimens.

Variation in plasma protein concentrations can occur secondary to decreased albumin concentrations associated with hepatic cirrhosis, and nephrotic syndrome (Table 2). Increased (α_1 -acid glycoprotein concentrations are associated with the stress response to disease states such as myocardial infarction, inflammatory disease, and postsurgically (41). A more relevant problem resulting from competition between drugs for plasma protein binding is the misinterpretation of the measured concentrations of drugs in plasma, as most assays are not able to differentiate between bound and unbound drug (42). Concentration-dependent binding of a drug to a plasma protein is expected when the total drug concentration approaches the protein concentration. For albumin this concentration is 0.6 mM, and for α_1 -acid glycoprotein in healthy individuals it is 0.015 mM, assuming that one drug molecule binds per protein molecule. For a drug with a molecular weight of 250, this corresponds to concentrations in plasma of 150 mg/L and 4 mg/L, for saturation of albumin and α_1 -acid glycoprotein, respectively.

**Influence of Protein-Binding
Changes on Volume of Distribution**

As is obvious from the relationship shown by Eq. 26, a drug which is characterized by a high degree of binding to plasma proteins (i.e., a low f_u) will exhibit a small volume of distribution. Decreases in albumin concentration, for example, may result in a decline in the fraction of drug bound to plasma proteins. The increased amount of unbound drug is then free to distribute to other tissues.

Unlike binding to plasma proteins however, binding to tissue receptors cannot be measured directly. This parameter is generally assumed to be constant. Again, from Eq. 26, this would appear to result in a volume of distribution increase. However, these changes are only important for drugs exhibiting a high degree of binding in plasma ($>90\%$) and an even higher degree of binding in the tissues, that is, drugs with a high volume of distribution. Changes in protein binding will not significantly affect volume of distribution for low V drugs. Note that changes in volume of distribution do not influence the steady-state relationship between dosing rate and average concentration (Eq. 5).

Influence of Protein-Binding Changes on Clearance

If we examine Eq. 12–14, we can see that for drugs with a low extraction ratio (i.e., chlordiazepoxide), Q_{organ} is much greater than $f_u \text{CL}_{\text{int}}$; clearance is then approximated by $f_u \text{CL}_{\text{int}}$. However in the case of a high extraction ratio drug (i.e. lidocaine), $f_u \text{CL}_{\text{int}}$ is much greater than Q_{organ} , and clearance approaches organ blood flow. Therefore, clearance of high extraction ratio drugs is perfusion rate limited, and not influenced by protein binding. Clearance of low extraction ratio drugs, in contrast to Eq. 14, is dependent upon both the fraction unbound and the intrinsic clearance (CL_{int}) of the metabolizing organ, for example, the liver.

THE CONCEPT OF EXPOSURE

Most pharmacokinetic theory has concentrated on explaining changes in clearance and bioavailability in terms of the parameters of intrinsic clearance, blood flow and fraction unbound (Eqs. 12 and 21). Yet the response of a patient to a dose or dosage regimen of a drug is dependent on the patient exposure to the drug, which is best characterized by AUC. Thus, it will be useful to consider the importance of individual parameters such as CL_{int} , Q and f_u in terms of exposure concepts. Consider first an intravenous dose of a drug where from Eq. 7a

$$\text{AUC} = \frac{\text{Dose}_{\text{i.v.}}}{\text{CL}} \quad (27)$$

and for a drug excreted exclusively by the liver substitution of Eq. 12 into Eq. (7a) yields

$$\text{AUC} = \frac{\text{Dose}_{\text{i.v.}}(Q_H + f_u \text{CL}_{\text{u,int}})}{Q_H f_u \text{CL}_{\text{u,int}}} \quad (28)$$

For a high extraction ratio compound ($f_u \text{CL}_{\text{u,int}} \gg Q_H$), exposure will be inversely dependent upon hepatic blood flow

$$\text{AUC} \approx \frac{\text{Dose}_{\text{i.v.}}}{Q_H} \quad (29)$$

while for a low extraction ratio compound ($Q_H \gg f_u \text{CL}_{\text{u,int}}$), exposure will be inversely related to fraction unbound and intrinsic hepatic clearance.

$$\text{AUC} \approx \frac{\text{Dose}_{\text{i.v.}}}{f_u \text{CL}_{\text{u,int}}} \quad (30)$$

However, for an orally administered compound exposure will also be a function of oral bioavailability

$$\text{AUC} = \frac{F_{\text{oral}} \text{Dose}_{\text{oral}}}{\text{CL}} \quad (31)$$

Substituting, Eq. 18 for F_{oral} , Eq. 21 for F_H , and Eq. 12 for CL into Eq. 30 above, and simplifying yields the following expression of exposure following an oral dose of a drug where elimination is via the liver.

$$\text{AUC} = \frac{F_{\text{abs}} F_G \text{Dose}_{\text{oral}}}{f_u \text{CL}_{\text{u,int}}} \quad (32)$$

Note that for oral dosing, exposure is inversely related to fraction unbound and intrinsic clearance for all drugs independent of whether they are low or high extraction ratio compounds.

Comparing Eq. 29 and 31 for a low extraction ratio drug where elimination is primarily hepatic, it is obvious that similar oral and i.v. doses will yield similar AUC values, unless the product of $F_{\text{abs}} F_G$ is low. A drug with $F_{\text{abs}} F_G \approx 1$, such as acetaminophen yields similar exposure following both oral and intravenous dosing, and also the variability of exposure will be similar for both routes, since the factors controlling variability, $f_u \text{CL}_{\text{u,int}}$, are the same in each case. Cyclosporine is an example of a low extraction ratio drug eliminated systemically by hepatic metabolism where exposure is markedly less following oral dosing vs. i.v., since $F_{\text{abs}} F_G$ is so low, in this case due to marked first pass gut metabolism as depicted in Fig. 4. For this drug we would expect variability in exposure to the oral drug to be greater than i.v., due to the addition of the $F_{\text{abs}} \cdot F_G$ differences.

Propranolol is a good example of a high extraction ratio drug where systemic elimination is almost completely hepatic. Although there is some gut metabolism for propranolol following oral dosing, it is not extensive ($F_G \gg F_H$). Fig. 5 depicts the predicted differences in exposure following i.v. (10 mg) and oral (80 mg) dosing for this high

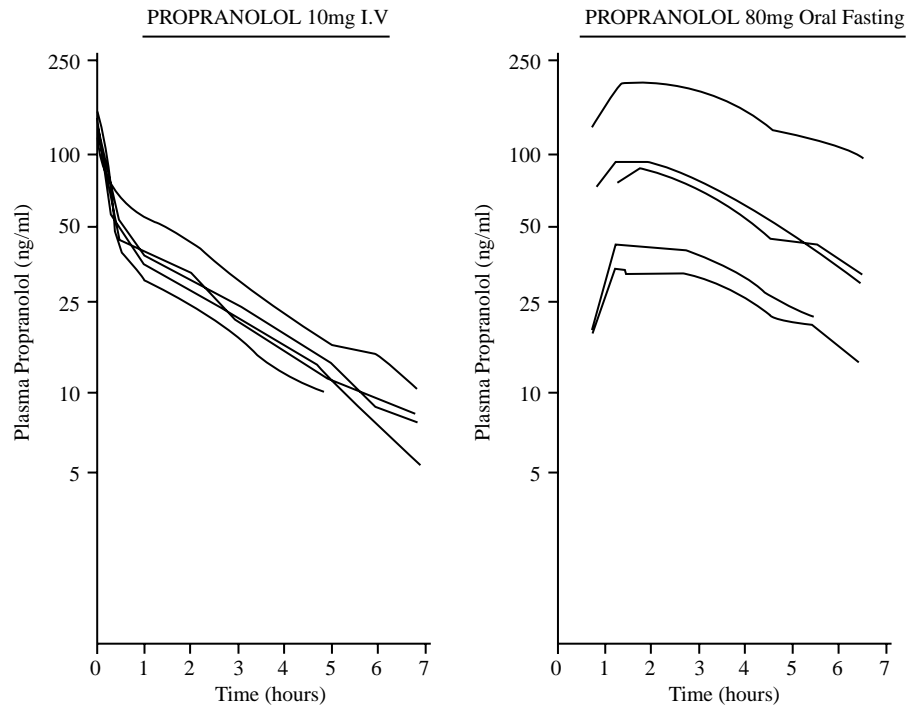


Fig. 5 Plasma propranolol levels in 5 subjects after i.v. administration of 10 mg and oral administration of 80 mg to fasting subjects. Observe the much larger variability in plasma concentrations between individuals after administration via the oral route. (From Ref. 43.)

extraction ratio drug. Note in comparing Eq. 28 and 31 that exposure is less for the oral dose since $f_u CL_{u,int} > Q_H$. In addition, it is obvious from Fig. 5 that the variability following oral dosing is much greater than that following i.v. dosing. That is, hepatic blood flow from subject to subject shows much less variability than the intrinsic clearance of the hepatic enzymes.

Returning to pharmacodynamic considerations, it is generally believed that drug responses, both efficacious and toxic, are related to exposure of the patient to unbound concentrations of the drug. For intermittent (not continuous infusion) dosing of drugs, Eq. 1 can be written as

$$E = \frac{E_{\max} AUC_u^s}{AUC_{u,50}^s + AUC_u^s} \quad (33)$$

where $AUC_{u,50}$ is the unbound AUC that yields 50% of the maximal effect over a dosing interval. That is, the effect achieved over a dosing interval for a given dose of the drug is a function of the exposure to unbound concentrations of the drug. Unbound drug exposure is related to total exposure by f_u , fraction unbound.

$$AUC_u = AUC f_u \quad (34)$$

Note that unbound drug exposure will be independent of f_u for orally administered drugs that are predominantly eliminated by hepatic mechanisms. When AUC from Eq. 31 is substituted into Eq. 33 then

$$AUC_u = \frac{F_{\text{abs}} F_G \text{Dose}_{\text{oral}}}{CL_{u,int}} \quad (35)$$

Similarly for intravenous dosing of low extraction ratio drugs, AUC_u is independent of f_u . Substituting, Eq. 29 into Eq. 33 yields

$$AUC_u \cong \frac{\text{Dose}_{i.v.}}{CL_{u,int}} \quad (36)$$

For high extraction drugs predominantly eliminated by the liver, f_u does affect AUC_u

$$AUC_u \cong \frac{f_u \text{Dose}_{i.v.}}{Q_H} \quad (37)$$

For drugs where hepatic elimination is negligible, then $F_H = 1$ and following oral or intravenous dosing we may predict AUC_u as follows

$$AUC_u = \frac{f_u F_{\text{abs}} F_G \text{Dose}}{CL} \quad (38)$$

and for a drug eliminated by renal processes

$$CL = \frac{Q_K f_u CL_{u,int}}{Q_K + f_u CL_{u,int}} \quad (39)$$

where Q_K is blood flow to the kidney. Substituting Eq. 38 into Eq. 37, then

$$AUC_u = \frac{F_{abs} F_G \text{Dose}(Q_K + f_u CL_{u,int})}{Q_K CL_{u,int}} \quad (40)$$

For a low extraction ratio drug in the kidney ($Q_K \gg f_u CL_{u,int}$), Eq. 39 becomes

$$AUC_u = \frac{F_{abs} F_G \text{Dose}}{CL_{u,int}} \quad (41)$$

whereas for a high extraction ratio drug in the kidney ($f_u CL_{u,int} \gg Q_K$), Eq. 39 becomes

$$AUC_u = \frac{F_{abs} F_G \text{Dose} f_u}{Q_K} \quad (42)$$

Note from the equations above that f_u becomes a determinant of AUC_u , and its thereby effect, only for a high extraction ratio drug given intravenously when the liver is the major route of elimination (Eq. 36), and for a high extraction ratio drug given orally or intravenously when the kidney is the major route of administration (Eq. 41). Since most drug dosings do not fall into these very limited categories, it now becomes clear why changes in protein binding as a function of disease state or drug interactions are generally of irrelevant consequences.

NONLINEAR PHARMACOKINETICS

Most drugs fortunately show linear pharmacokinetics within the therapeutic plasma concentration range (that is, with a doubling of the dose, the plasma concentration is also doubled), making possible the prediction of the impact of changes in drug dosing on the pharmacokinetic outcome.

Nonlinear pharmacokinetics are caused mainly by saturation of the metabolizing enzymes during drug elimination or during the first passage of drug through the liver, and also by saturable protein binding as previously discussed. This leads to less predictable results in drug therapy and to the risk of a higher incidence of side effects.

Nonlinear Elimination/Clearance

All metabolic processes are saturable at a certain concentration of the substrate/drug. Thus, rate of elimination of the drug by metabolism as described by

Eq. 5 can also be described by a Michaelis-Menten equation:

$$\text{Rate of metabolism} = \frac{V_m C}{K_m + C} \quad (43)$$

From these equations metabolic clearance can be described as

$$CL = \frac{V_m}{K_m + C} \quad (44)$$

where V_m is maximal velocity and K_m is the Michaelis-Menten constant. When kinetics are linear, $C \ll K_m$, and clearance equals V_m/K_m , but as the concentration approaches or exceeds K_m , clearance becomes dependent on the concentration, resulting in saturable metabolism. At steady state, when input rate (R_{in}) equals elimination rate, the steady-state concentration can be described by replacing rate of metabolism in Eq. 42 by R_{in} and rearranging

$$C_{ss} = \frac{K_m R_{in}}{V_m - R_{in}} \quad (45)$$

Drugs that show saturable metabolism within the therapeutic range include phenytoin and salicylate. Because of the serious side effects encountered with phenytoin, and because of the genetic variability in metabolic capacity with respect to individual V_m and K_m values, phenytoin therapy is closely monitored using plasma concentration analysis. Phenytoin has a therapeutic concentration range of 10–20 mg/L, which is above K_m for most individuals. A small change of the dose in this region leads to a big change in plasma concentrations. An orbit graph of the relationship between V_m , K_m , and the average steady-state concentration of phenytoin based on Eq. 44 is shown in Fig. 6.

The mean values of K_m and V_m within the population are 4 mg/L and 7 mg/kg/day, respectively. The variation within the population is also depicted in Fig. 6, where 50% of the population has K_m and V_m values within the innermost circle, 75% within the second circle, and so on. With the help of this graph, the suitable dosing rate of phenytoin for an individual patient can be determined. With two plasma concentrations at two different dosing rates of phenytoin in one patient, his/her individual V_m and K_m values can be obtained to further optimize the dosing schedule.

Saturable first-pass effect

Because of very high drug concentrations entering the gut and the liver during the absorption of a drug after oral administration, the liver may often be exposed to drug concentrations that are much higher than after i.v. administration or much higher than the concentrations

Table 3 The extent of urinary excretion of *p*-Aminobenzoic acid and its acetyl metabolite as a function of rate and route of administration in one subject

Route	Total dose Na-PABA (g)	Total PABA in urine in 24 h as % of dose excreted in 24 h (%)	Acetyl-PABA in urine as % of total PABA (%)
i.v. Bolus	1	102	51
Prolonged administration			
i.v. infusion—270 mm	0.4	90	95
10 Oral doses given every half hour	0.4	95	97
Single dose			
Oral solution	1	103	51
	2	103	47
	4	102	36
	8	102	30

(Adapted from Ref 46.)

encountered during the elimination phase after oral administration. These initial hepatic portal vein concentrations after oral administration may exceed K_m even if the plasma concentrations measured in a peripheral venous sample do not exceed K_m . The higher the dose, the more of the drug escapes first-pass metabolism, which gives higher bioavailability with increasing dose. An illustrative example of this is the availability of *p*-aminobenzoic acid (PABA) as shown in Table 3 (46). Virtually all PABA is eliminated into the urine as metabolite and unchanged drug within 24 h. However, the percent of the dose that has been metabolized to acetyl-PABA is dependent on dose size and rate of administration, with a clear decrease of the fraction of metabolite formed with increasing dose after oral administration. Also, the fraction of the dose that is metabolized increases dramatically with prolonged oral and i.v. administration. This may also have an impact on slow-release formulations, which for drugs exhibiting a saturable first-pass effect never yield portal vein concentrations as high as comparable immediate-release tablets, thus resulting in lower bioavailability for the slow-release formulation.

STEADY-STATE CONSIDERATIONS IN DESIGN OF DOSAGE REGIMENS

Most drug therapy is designed for chronic drug administration, such as for treatment of hypertension or

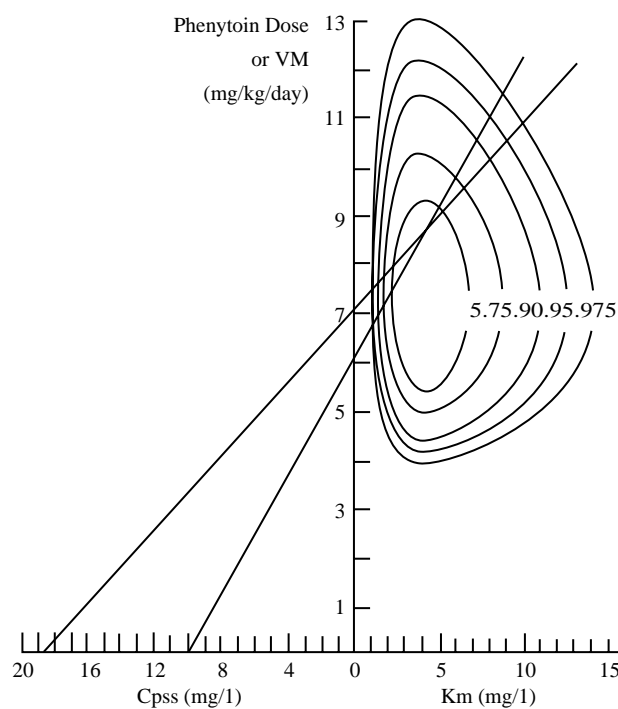


Fig. 6 The Bayesian feedback method of phenytoin dosage prediction. The eccentric circles represent the fraction of the sample population whose V_m and K_m values are within that orbit. By drawing lines from the measured C_{ss} values via the given doses of phenytoin, the most probable values of V_m and K_m can be estimated and further used in calculation of new dosing rates corresponding to a target concentration. (From Ref 44; also discussed in Ref 45.)

diabetes. At steady state, the rate of input equals the rate of elimination (cf. Eqs. 5 and 6):

$$\frac{F \text{ Dose}}{\tau} = CL C_{ss,avg} \quad (46)$$

where F is bioavailability, τ is the dosing interval, and $C_{ss,avg}$ is the average concentration during steady state. Note $C_{ss,avg}$ is another measure of exposure as discussed previously, since during the dosing interval $C_{ss,avg}$ is equivalent to AUC/τ . Multiplying both sides of Eq. 45 by τ and dividing by CL yields Eq. 30. From Eq. 45 dosing rate can be calculated knowing F and CL , or if $C_{ss,avg}$ is measured, CL/F can be estimated in a patient. Plasma drug concentration measurements as an aid in giving appropriate drug treatment may be useful when

1. The therapeutic index of the drug is low (e.g., theophylline, digoxin).
2. No pharmacodynamic measurements (e.g., blood pressure, minor side effects such as dry mouth during anticholinergic drug therapy) can be used as end points for therapy.
3. The drug shows nonlinear kinetics within the therapeutic range (e.g., phenytoin).
4. The patient is "uncommon" (e.g., newborns, small children, pregnant women, or patients with decreased kidney function).

For the average steady-state concentration, the frequency of drug administration is not important as long as the dosing rate is the same (Fig. 7). Thus, 1020 mg of theophylline can be given every 24 h or 340 mg every 8 h and still have the same average plasma concentration as when a constant infusion of 43.2 mg/h is given (17). However, the fluctuations around the mean value are influenced by the dosing interval. For drugs with a small therapeutic index, the fluctuations between maximum and minimum concentrations during a dosing interval must be kept within the therapeutic range, thus favoring shorter dosing intervals.

Maximum ($C_{ss,max}$) and minimum ($C_{ss,min}$) concentrations during a dosing interval τ can be estimated using the following equations, assuming rapid absorption compared to elimination, and a one-compartment model for the elimination, where λ is the rate constant for drug elimination:

$$C_{ss,max} = \frac{F \text{ Dose}/V}{(1 - e^{-\lambda\tau})} \quad (47)$$

In Equation (46) $(F \text{ Dose})/V$ describes what is added to the concentration from the new dose. Thus, ($C_{ss,min}$) may be

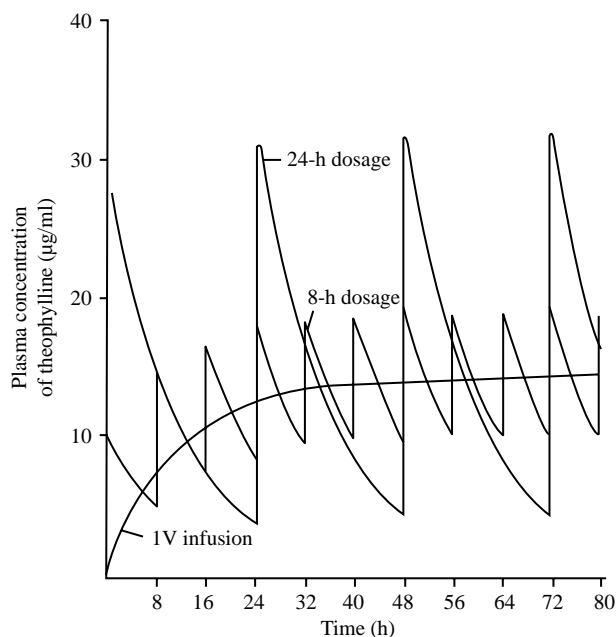


Fig. 7 Relationship between frequency of dosing and maximum and minimum plasma concentrations when a steady-state theophylline plasma level of 15 µg/ml is desired. The smoothly rising line shows the plasma concentration achieved with an intravenous infusion of 43.2 mg/h. The doses for 8-h administration are 340 mg; for 24-h administration, 1020 mg. In each of the three cases, the mean steady-state plasma concentration is 15 µg/ml. (From Ref 47.)

calculated as ($C_{ss,max}$) multiplied by the fraction remaining in the body at the end of the dosing interval:

$$C_{ss,min} = C_{ss,max} e^{-\lambda\tau} \quad (48)$$

The difference between maximum and minimum concentrations at steady state is $(F \text{ Dose}/V)$, and thus the amount that is eliminated during the dosing interval equals the available dose.

The factors that determine the average concentration at steady state are bioavailability and clearance. The maximum concentration is determined by the volume of distribution, the smaller the volume, the higher the concentration. The dosing interval compared to the half-life is most important in defining the fluctuations. A very convenient dosing interval is equal to the half-life of the drug. This results in a decline during the dosing interval to a minimum concentration that is 50% of the maximal concentration. However, administration of a drug every half-life might sometimes be difficult to achieve because of too short or too long half-lives compared to our 24-h cycle. A practical minimum dosing interval is 6 h, and the maximum is 24 h. For drugs with short half-lives,

slow-release preparations might give smoother plasma concentration curves and decrease the number of dosings required. Compliance ultimately determines the therapeutic outcome; therefore, adjusting drug dosings to the daily routine of an individual, as much as possible, is important.

Loading Dose

A loading dose is given to attain the effective target concentration rapidly. It can be given as one dose or, more commonly, be divided into several doses over a relatively short period of time, such as one per day. The size of the loading dose depends on the target concentration and the volume of distribution and bioavailability of the drug. The loading dose should give the same amount of drug as will be present in the body during steady state

$$\text{Loading dose} = C_{ss} V/F \quad (49)$$

A loading dose is used when the need for drug therapy is too urgent to wait 3–4 drug half-lives to reach the desired drug concentration in the body. The advantage is a more rapid attainment of steady state. The disadvantage is a higher risk of side effects if the drug concentrations become too high or if the individual is more sensitive to the drug than the average patient. However, in some instances, such as in dosing anti-arrhythmics following an acute myocardial infarction, the only choice is to give a loading dose.

INDIVIDUALIZING DOSAGE REGIMENS: INDIVIDUAL VARIATIONS AND DISEASE STATES

When starting a patient on a particular drug therapy, the clinician is dependent on general dosing information, such as from the PDR (Physicians Desk Reference) or other literature or experience, including an evaluation of the physical status of the patient. At that time the clinician cannot know exactly how this specific patient will react to the drug, pharmacokinetically or pharmacodynamically. For optimal therapy, the progress of therapy should be followed with measurements of relevant effect parameters, by titrating the dose to avoid certain known side effects, or by using therapeutic drug monitoring (drug plasma concentration measurements).

The need for individualized dosage recommendations has received increased attention in recent years parallel to the increased knowledge of what can cause interindividual variability in pharmacokinetic handling of drugs and in pharmacodynamic response. These causes include the

influences of genetic variations, age, and disease states. Disease states influence the pharmacokinetics of drugs mainly by impairing drug transport and/or elimination, such as in heart failure or kidney or liver disease.

Kidney function and age show a close correlation. Glomerular filtration rate (GFR) is 100–125 ml/min/70 kg at 20–30 years of age, and then declines approximately 1 ml/min per year as the patient ages. Decreased kidney function as a result of disease is usually detected by measuring serum creatinine levels and calculating the creatinine clearance as a measure of GFR, even though using creatinine as a marker for decreased kidney function has received increased criticism due to overestimation of GFR at medium–low kidney function (48). A GFR between 20 and 50 ml/min indicates moderate renal failure, and values <10 ml/min indicate severe renal failure. If a major portion of the dose of a drug is eliminated unchanged via the kidneys, or if active or toxic metabolites are renally eliminated, corrections for kidney function may be necessary in dosing recommendations.

The influence of age on metabolic clearance is less clear than its influence on renal function. Metabolic clearance is more variable between individuals because of the genetic control and the influence of environmental factors on the metabolic capacity. The term metabolism also encompasses many different enzyme reactions that might be influenced to different extents by age, liver disease, or genetic variables. Unlike renal disease, in which creatinine clearance provides a reasonable estimate of kidney function, not one good indicator exists for the degree of liver function impairment with respect to drug metabolizing capacity.

Liver disease includes many diverse conditions. A distinction can be made between acute and chronic liver disease with respect to changes in clearance. Acute states do not generally seem to influence metabolic clearance of most drugs, but chronic states, such as cirrhosis, seem to decrease metabolic function to a greater degree. Also in cirrhosis, a portion of the blood flow may be shunted past the small capillaries within the liver or bypass the liver, causing even less exposure of the drug to the metabolizing enzymes and thereby greatly increasing the bioavailability of high-extraction-ratio drugs. Different drugs show different behavior with respect to the influence of liver disease on pharmacokinetics. Oxidative pathways seem to be more highly affected by liver disease than conjugation reactions such as glucuronidation. This is evident in the benzodiazepine group, where in cirrhosis patients oxazepam and lorazepam are preferred over diazepam and chlordiazepoxide due to the smaller changes in clearance for the former drugs, which are predominantly glucuronidated (49).

Cardiac failure may influence drug disposition by decreasing the cardiac output and increasing the peripheral resistance, thereby redistributing the blood flow from peripheral organs and tissues, such as kidneys and muscles, in favor of central organs, such as brain and heart. This might cause a slower than normal distribution of drug throughout the body and higher initial concentrations in the central organs that could cause an increased risk of side effects. Because of a lower blood flow to the gastrointestinal tract, heart failure might cause a slower rate of absorption of drugs. However, this seldom results in a lower bioavailability. A decreased and redistributed blood flow in liver and kidneys could yield a slower elimination of highly extracted drugs (50).

Pharmacogenetics

The 20th century yielded an enormous expansion in the chemical arsenal available to medical professionals, for the treatment of disease. The worldwide use of these drugs, particularly over the past several decades, has revealed substantial interindividual differences in the response to drugs. These differences, confirmed and explained by pharmacokinetic analysis, appear in individual patients as variations not only in expected therapeutic responses, but often in the frequency and nature of associated adverse events. Any particular drug or drug dose may be therapeutic in some patients, but ineffective in other individuals. Along the same lines, some individuals may experience drug related adverse effects from standard therapy, while others are unaffected. Recognition of these interindividual differences in drug response is an essential first step in optimizing drug therapy. A great deal of evidence has accumulated over the past several decades indicating that a substantial portion of the observed variability in response to drugs is genetically determined, although age, nutrition, health, and environmental factors play important roles. Pharmacogenetics is primarily concerned with the genetic basis underlying this interindividual drug response, and focuses on genetic differences among individuals, and on the different patterns of drug response among geographically and ethnically distinct populations. The goal of pharmacogenetics is to make truly individualized drug therapy, one could say personalized medicine, possible, and with a reasonably predictive outcome.

The basic unit of inheritance is the *gene*, a specific DNA sequence of base pairs which codes for a particular protein. The *genotype*, is the configuration of genes in an individual, while the *phenotype*, is the outward physical expression of those genes. The mode of inheritance of any individual trait

can be either monogenic or polygenic, depending on whether it is derived from a single gene at a single locus (position), or by multiple loci on the chromosomes respectively. An *allele* is one, two or more different forms of the same gene, containing specific inheritable differences, occupying corresponding positions on paired chromosomes. *Genetic polymorphism*, refers to monogenic traits existing in the population in at least two phenotypes, each of which exists at a frequency of at least 1%. An individual possessing a pair of identical alleles for a given trait is said to be *homozygous* for that trait. Individuals who are *heterozygous* for a given trait, possess a combination of dominant and recessive alleles for that trait.

The cytochrome P450 (CYP) monooxygenase system of enzymes is responsible for the primary portion of xenobiotic (foreign substance) metabolism in man. This large family of genes is composed of numerous subtypes, among which CYP2D6, CYP3A4/3A5, CYP1A2, CYP2E1, CYP2C9 and CYP2C19 play especially important roles in genetically determined responses to a wide spectrum of drugs. Examples to date of inherited variability in pharmacokinetics have been almost entirely restricted to drug metabolism. Renal clearance for any drug tends to be very similar for age and weight matched patients in good health. Similarly, the role of genetics in determining the absorption and distribution of drugs is poorly defined.

A number of examples of genetic polymorphisms in drug metabolism have now been identified, involving oxidation, S-methylation, and acetylation., as shown in Table 4. These polymorphisms were initially identified by adverse drug reactions occurring in distinct groups, termed "poor metabolizers," following normal therapeutic doses of the initial, archetypic drugs.

Patients who are homozygous for the CYP2D6 "slow" alleles exhibit a poor metabolizer phenotype, with impaired metabolism and excretion of many drugs, such as metoprolol, nortriptyline and propafone (51). These "poor metabolizers" are more likely to experience adverse reactions to standard doses of these drugs. The frequency of this recessive trait ranges from 1–2% in Asians, approximately 5% in African Americans to 6–10% in Caucasian populations (52). More than 40 drugs now in clinical practice, particularly in the areas of cardiovascular disease and psychiatric disorders exhibit the same polymorphic pattern of metabolism by CYP2D6 (53, 54). Along similar lines, patients who are homozygous for the recessive allele of CYP2C19 are highly sensitive to omeprazole, diazepam, propranolol, mephennitoine, amitriptyline and other drugs metabolized by this isoform (51). The "poor metabolizer" phenotype of CYP2C19 comprises approximately 2–5% of Caucasians

Table 4 Polymorphic drug metabolizing enzymes

Enzyme	Variant phenotypes	Drugs	Modified response
Plasma pseudocholinesterase	Slow hydrolysis	Succinylcholine	Prolonged apnea
Acetyl transferase NAT2	Slow, rapid acetylators	Isoniazid Sulfamethazine Procainamide Sufasalazine Paraminosalicylic acid Hydralazine	Toxic neuritis, lupus erythematosus. (Slow acetylators)
Thiopurine methyltransferase	Poor TPMT methylators	6-Mercaptopurine 6-Thioguanine Azathioprin	Bone marrow toxicity, hepatotoxicity
Dihydropyrimidine dehydrogenase	Slow inactivation	5-Fluorouracil	Enhanced toxicity
Aldehyde dehydrogenase	Fast, slow metabolizers	Ethanol	Slow: facial flushing Fast: protection from liver cirrhosis
CYP 2D6	Rapid, poor metabolizers	Debrisoquine Sparteine Phenformin Nortriptyline Dextromorphan, etc.	Poor: increased toxicity Rapid: drug resistance
CYP 2C9	Poor metabolizers	Tolbutamide S-warfarin Phenytoin Nonsteroidal anti-inflammatory agents Imipramine	Increased response or toxicity
CYP 2C19	Poor and extensive hydroxylators	Mephenytoin Hexobarbital Omeprazole Proguanil, etc.	Poor: increased toxicity Extensive: drug resistance

and 3–23% of Asians. Another polymorphically expressed enzyme of the P450 superfamily is CYP2C9. This enzyme is responsible for the metabolism of a range of therapeutically important drugs such as phenytoin, tolbutamide and warfarin (5). It has been estimated, that approximately 5% of Caucasians possess the genetic variant of CYP2C9 associated with a fivefold decline in metabolic activity (6). Drugs such as phenytoin and warfarin, for example, have narrow therapeutic indices, and individual genetically based variations in metabolism could have important clinical significance.

Individuals also vary widely in their elimination kinetics of isoniazid, procainamide, and other substrates of N-acetyltransferase (NAT2). Peripheral neuropathy associated with the use of isoniazid, an antituberculosis drug, first surfaced more than 40 years ago. It is now known that the “slow acetylator” phenotype represents

approximately 40–50% of Caucasians, and results in decreased clearance of drug with increased potential for associated toxicities.

Genetic heterogeneity appears then, to be an important source of the observed variability in drug response. This strongly implies that information pertaining to interethnic and interindividual genetic differences can play an important role in drug discovery and development. New genetic tools for the rapid, inexpensive determination of patient genotypes could have important therapeutic implications. The need for a new, individualized approach to drug therapy is now, more obvious than ever. Every year approximately 3 billion prescriptions are issued in the United States, of which 2 million result in an adverse reaction, with as many as 1 million resulting in hospitalization. Some 100,000 patients die each year as a result of drug induced adverse

reactions. Early or preventive therapy, guided by individual genetic information, could significantly improve clinical outcomes, while reducing the incidence of drug associated adverse events.

USE OF PHARMACOKINETIC INFORMATION IN DRUG FORMULATION

Depending on the use of a drug, different parameters are important in designing a dosage form. For intermittent drug use, such as aspirin for headache, the ideal dosage form should give a rapid rate of absorption to yield quick relief of the pain. This objective has led to the formulation of buffered or soluble forms of acetylsalicylic acid, for which the absorption rate is limited mainly by the drug's ability to cross the GI membranes. With a different indication, such as aspirin used chronically for rheumatoid arthritis, low exposure to the stomach is desired; for this purpose, enteric-coated tablets and granules of the drug have been formulated.

For a drug intended to be used chronically, the clearance of the drug from the body, together with the extent of availability, determines the dose size for a given therapeutic drug level (Eq. (45)). If clearance and volume of distribution of the specific drug result in a short half-life and keeping drug concentrations above a certain level is important, for example, during antiarrhythmic treatment, the dosing interval is crucial, leading to the demand for a controlled-release dosage form.

However, pharmacokinetic properties of a drug may pose limitations as to the possibility of attaining an ideal therapy. The choice of digoxin over digitoxin is mainly determined by the shorter half-life and in the minds of some clinicians a safer and more predictable drug therapy with digoxin. Oxazepam is an example of a drug with a slower than wanted rate of absorption, which decreases its use as a help in starting to sleep. This slow absorption rate is a characteristic of the drug itself and cannot be increased by drug-product formulation. Thus, dosage form development is limited by the basic pharmacokinetic properties of the drug.

The oral bioavailability of a high-extraction-ratio drug can never be improved by dosage form development if the reason for low availability is high first-pass metabolism. The ability to give an oral dosing regimen of a high-extraction-ratio drug is dependent on the ratio between therapeutic concentrations of the drug and toxic concentrations of the metabolites formed during the first pass and on how critical for successful therapy staying

within the therapeutic range of drug concentrations is. Both lidocaine and propranolol have a high extraction in the liver with low bioavailability as a result (35 and 36%, respectively) (6). However, propranolol is given orally and lidocaine is not. Therapeutic concentrations of lidocaine are 1.5–6 mg/L, and side effects start to appear at concentrations of 6–10 mg/L. For lidocaine falling below the therapeutic concentration range is just as serious as going too high in concentration, as both situations may cause arrhythmias. With a half-life of only 1.8 h, oral doses would have to be given very often, which is not practical. Also, the variability in the first-pass effect between individuals (Figs. 4, 5) makes uniform dosing recommendations difficult. Propranolol also has a relatively short half-life (3.9 h), but the therapeutic index is much higher, and plasma concentrations falling below the "therapeutic concentration" are not as critical for the patient; therefore, larger fluctuations in plasma concentrations can be allowed during a dosing interval without loss of an acceptable therapeutic outcome.

The pharmacokinetic profile of a drug can in this way be used as a basis for dosage form development, together with the specific requirements of the particular drug concerning maximal and/or minimal concentrations desired, effectiveness versus toxicity, the desired rate of attaining an effect, and the convenience of drug administration for the individual.

Thus, the knowledge of basic pharmacokinetic features, such as rate of absorption, degree of absorption, degree of renal elimination versus metabolism, whether the drug is a high or a low extraction drug, and the like, and knowledge of the pharmacokinetic–pharmacodynamic relationship are tools in optimizing drug product development to the needs of the patient.

REFERENCES

1. Holford, N.H.G.; Sheiner, L.B. Understanding the Dose-Effect Relationship: Clinical Application of Pharmacokinetic-Pharmacodynamic Models. *Clin. Pharmacokinet.* **1981**, *6*, 429–453.
2. Sheiner, L.B.; Stanski, D.R.; Vozeh, S.; Miller, R.D.; Ham, J. Simultaneous Modeling of Pharmacokinetics and Pharmacodynamics: Application to d-Tubocurarine. *Clin. Pharmacol. Ther.* **1979**, *25*, 358–371.
3. Paalzow, L.K.; Paalzow, G.H.M.; Tfelt-Hansen, P. Kinetics of Drug Action. *Pharmacokinetics. A Modern View*; Benet, L.Z., Levy, G., Ferraiolo, B.L., Eds.; Plenum Press: New York, 1984; 315–343.
4. Rowland, M.; Benet, L.Z.; Graham, G.G. Clearance Concepts in Pharmacokinetics. *J. Pharmacokinet. Biopharm.* **1973**, *1*, 123–136.

5. Wilkinson, G.R.; Shand, D.G. Commentary: A Physiological Approach to Hepatic Drug Clearance. *Clin. Pharma. Ther.* **1975**, *18*, 377–399.
6. Benet, L.Z.; Oie, S.; Schwartz, J.B. Appendix II. Design and Optimization of Dosage Regimes: Pharmacokinetic Data. *The Pharmacological Basis of Therapeutics*, 9th Ed.; Hardman, J.G., Limbird, L.E., Molinoff, P.B., Ruddon, R.W., Goodman, A.G., Eds.; McGraw-Hill: New York, 1996; 1707–1792.
7. Pang, K.S.; Rowland, M. Hepatic Clearance of Drugs. 1. Theoretical Considerations of a Well-Stirred Model and a Parallel Tube Model. Influence of Hepatic Blood Flow, Plasma and Blood Cell Binding, and the Hepatocellular Enzymatic Activity on Hepatic Drug Clearance. *J. Pharmacokinet. Biopharm.* **1977**, *5*, 625–653.
8. Pang, K.S.; Rowland, M. Hepatic Clearance of Drugs. I. Theoretical Considerations of a “Well-Stirred” Model and a “Parallel Tube” Model. Influence of Hepatic Blood Flow, Plasma and Blood Cell Binding, and the Hepatocellular Enzymatic Activity on Hepatic Drug Clearance. *J. Pharmacokinet. Biopharm.* **1977**, *5*, 655–683.
9. Bass, L.; Robinson, P.; Bracken, A.J. Hepatic Elimination of Flowing Substrates: The Distributed Model. *J. Theor. Biol.* **1978**, *72*, 161–184.
10. Forker, E.L.; Luxon, B. Hepatic Transport Kinetics and Plasma Disappearance Curves: Distributed Modeling Versus Conventional Approach. *Am. J. Physiol.* **1978**, *235*, E648–E660.
11. Roberts, M.S.; Rowland, M. Hepatic Elimination—Dispersion Model. *J. Pharm. Sci.* **1985**, *74*, 585–587.
12. Roberts, M.S.; Rowland, M. A Dispersion Model of Hepatic Elimination: 1. Formulation of the Model and Bolus Considerations. *J. Pharmacokinet. Biopharm.* **1986**, *14*, 227–260.
13. Roberts, M.S.; Rowland, M. A Dispersion Model of Hepatic Elimination: 2. Steady-State Considerations—Influence of Hepatic Blood Flow, Binding within Blood, and Hepatocellular Enzyme Activity. *J. Pharmacokinet. Biopharm.* **1986**, *14*, 261–288.
14. Roberts, M.S.; Rowland, M. A Dispersion Model of Hepatic Elimination: 3. Application to Metabolite Formation and Elimination Kinetics. *J. Pharmacokinet. Biopharm.* **1986**, *14*, 289–307.
15. Benet, L.Z.; Wu, C.; Hebert, M.F.; Wachter, V.J. Intestinal Drug Metabolism and Antitransport Processes: A Potential Paradigm Shift in Oral Drug Delivery. *J. Controlled Release* **1996**, *39*, 139–143.
16. Hoener, B.; Benet, L.Z. Factors Influencing Drug Absorption and Drug Availability. *Modern Pharmaceutics*, 3rd Ed.; Banker, G.S., Rhodes, C.T., Eds.; Marcel Dekker, Inc.: New York, 1996; 121–153.
17. Benet, L.Z.; Kroetz, D.L.; Sheiner, L.B. Pharmacokinetics: The Dynamics of Drug Absorption, Distribution and Elimination. *Goodman and Gilman's The Pharmacologic Basis of Therapeutics*, 9th Ed.; Hardman, J.G., Limbird, L.E., Molinoff, P.B., Ruddon, R.W., Goodman, A.G., Eds.; McGraw-Hill: New York, 1996; 3–27.
18. Wachter, V.J.; Silverman, J.A.; Zhang, Y.; Benet, L.Z. Role of P-Glycoprotein and Cytochrome P450 3A In Limiting Oral Absorption of Peptides and Peptidomimetics. *J. Pharm. Sci.* **1998**, *87*, 1322–1330.
19. Paine, M.F.; Khalighi, M.; Fisher, J.M.; Shen, D.D.; Kunze, K.L.; Marsh, J.D.; Thummel, K.E. Characterization of Interintestinal and Intraintestinal Variations in Human CYP3A-Dependent Metabolism. *J. Pharmacol. Exp. Ther.* **1997**, *283*, 1552–1562.
20. Watkins, P.B.; Wrighton, S.A.; Schuetz, E.G.; Molowa, D.T.; Guzelian, P.S. Identification of Glucocorticoid-Inducible Cytochromes P450 in the Intestinal Mucosa of Rats and Man. *J. Clin. Invest.* **1987**, *80*, 1029–1036.
21. Thiebaut, F.; Tsuruo, T.; Hamada, H.; Gottesman, M.M.; Pastan, I.; Willingham, M.C. Cellular Localization of the Multidrug-Resistance Gene Product P-Glycoprotein in Normal Human Tissues. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7735–7738.
22. Borst, P.; Schinkel, A.H.; Smit, J.J.M.; Wagenaar, E.; Van Deemter, L.; Smith, A.J.; Eijdem, E.W.M.; Baas, G.; Zaman, G.J.R. Classical and Novel Forms of Multidrug Resistance and the Physiological Functions of P-Glycoprotein in Mammals. *Pharmacol. Therap.* **1993**, *60*, 289–299.
23. Fojo, A.T.; Ueda, K.; Slamon, D.J.; Poplack, D.G.; Gottesman, M.M.; Pastan, M.M. Expression of a Multidrug-Resistance Gene in Human Tumors and Tissues. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 265–269.
24. Wachter, V.J.; Wu, C.-Y.; Benet, L.Z. Overlapping Substrate Specificities and Tissue Distribution of Cytochrome P450 3A and P-Glycoprotein: Implications for Drug Delivery and Cancer Chemotherapy. *Mol. Carcin.* **1995**, *13*, 129–134.
25. Wachter, V.J.; Salphati, L.; Benet, L.Z. Active Secretion and Enterocytic Drug Metabolism Barriers to Drug Absorption. *Adv. Drug Delivery Rev.* **1996**, *20*, 99–112.
26. Schuetz, E.G.; Schuetz, J.D.; Strom, S.C.; Thompson, M.; Fisher, R.A.; Molowa, D.T.; Li, D.; Guzelian, P.S. Regulation of Human Liver Cytochromes P-450 in Family 3A Induction in Primary Cultures of Rat, Rabbit and Human Hepatocytes. *Hepatology* **1993**, *18*, 1254–1262.
27. Schuetz, E.G.; Beck, W.T.; Schuetz, J.D. Modulators and Substrates of P-Glycoprotein and Cytochrome P450 3A Coordinately Upregulate these Proteins in Human Colon Carcinomas. *Mol. Pharmacol.* **1996**, *49*, 311–318.
28. McKinnon, R.A.; Burgess, W.M.; Hall, P.d.I.M.; Roberts-Thomson, S.J.; Gonzalez, F.J.; McManus, M.E. Characterization of CYP3A Gene Subfamily Expression in Human Gastrointestinal Tissues. *Gut* **1995**, *36*, 259–267.
29. Lown, K.S.; Bailey, D.G.; Fontana, R.J.; Janardan, S.K.; Adair, C.H.; Fortlage, L.A.; Brown, M.B.; Guo, W.; B, W.P. Grapefruit Juice Increases Felodipine Oral Availability in Humans by Decreasing Intestinal CYP3A Protein Expression. *J. Clin. Invest.* **1997**, *99*, 2545–2553.
30. Lown, K.S.; Mayo, R.R.; Leichtman, A.B.; Hsiao, H.L.; Turgeon, D.K.; Schmiedlin-Ren, P.; Rossi, S.J.; Brown, M.; Guo, W.; Benet, L.Z.; Watkins, P.B. Role of Intestinal P-Glycoprotein (MDR1) in Interpatient Variation in the Oral Bioavailability of Cyclosporin A. *Clin. Pharm. Ther.* **1997**, *62*, 248–260.
31. Wu, C.; Benet, L.Z.; Hebert, M.F.; Gupta, S.K.; Rowland, M.; Gomez, D.Y.; Wachter, V.J. Differentiation of Absorption and First-Pass Gut and Hepatic Metabolism in Humans: Studies with Cyclosporine. *Clin. Pharm. Ther.* **1995**, *58*, 492–497.

32. Benet, L.Z.; Galeazzi, R.L. Noncompartmental Determination of the Steady-State Volume of Distribution. *J. Pharm. Sci.* **1979**, *68*, 1971–1974.
33. Wagner, J.G. Linear Pharmacokinetic Equations Allowing Direct Calculation of Many Needed Pharmacokinetic Parameters from the Coefficients and Exponents of Polyexponential Equations which have been Fitted to the Data. *J. Pharmacokinet. Biopharm.* **1976**, *4*, 443–467.
34. Aziz, N.S.; Gambertoglio, J.G.; Lin, E.T.; Grausz, H.; Benet, L.Z. Pharmacokinetics of Cefamandole Using a HPLC Assay. *J. Pharmacokinet. Biopharm.* **1978**, *6*, 153–164.
35. Gambertoglio, J.G.; Aziz, N.S.; Lin, E.T.; Grausz, H.; Naughton, J.L.; Benet, L.Z. Cefamandole Kinetics in Uremic Patients Undergoing Hemodialysis. *Clin. Pharm. Ther.* **1979**, *26*, 592–599.
36. Lin, J.H. Species Similarities and Differences in Pharmacokinetics. *Drug Metab. Disposition* **1995**, *23*, 1008–1021.
37. Zini, R.; D'Athis, P.; Barre, J.; Tillement, J.P. Binding of Indomethacin to Human Serum Albumin: Its Non-Displacement by Various Agents, Influence of Free Fatty Acid and the Unexpected Effect of Indomethacin on Warfarin. *Biochem. Pharmacol.* **1979**, *28*, 2661–2665.
38. Vesell, E.S.; Passananti, G.T.; Johnson, A.Q. Failure of Indomethacin and Warfarin to Interact in Normal Human Volunteers. *J. Clin. Pharmacokinet.* **1975**, *15*, 486–495.
39. Winter, M.E. *Basic Clinical Pharmacokinetics*, 3rd Ed.; Applied Therapeutics, Inc.: Vancouver, WA, 1994; 13–22.
40. Rowland, M.; Tozer, T.N. Distribution. *Clinical Pharmacokinetics Concepts and Applications*, 3rd Ed.; Williams & Wilkins: Baltimore, MD, 1995; 151.
41. Tozer, T.N. Implications of Altered Plasma Protein Binding in Disease States. *Pharmacokinetic Basis for Drug Treatment*; Benet, L.Z., Massoud, N., Gambertoglio, J.G., Eds.; Raven Press: New York, 1984; 173–193.
42. Pacifici, G.M.; Viani, A. Methods of Determining Plasma and Tissue Binding of Drugs: Pharmacokinetic Consequences. *Clin. Pharmacokinet.* **1994**, *23*, 449–468.
43. Shand, D.G.; Nuckolls, E.M.; Oates, J.A. Plasma Propranolol Levels in Adults with Observations in Four Children. *Clin. Pharm. Ther.* **1970**, *11*, 112–120.
44. Vozech, S.; Muir, K.T.; Sheiner, L.B.; Follath, F. Predicting Individual Phenytoin Dosage. *J. Pharmacokinet. Biopharm.* **1980**, *9*, 131–146.
45. Winter, M.E.; Tozer, T.N. Phenytoin. *Applied Pharmacokinetics. Principles of Therapeutic Drug Monitoring*, 2nd Ed.; Evans, W.E., Schentag, J.J., Jusko, W.J., Eds.; Applied Therapeutics, Inc.: Vancouver, WA, 1986; 493–539.
46. Drucker, M.M.; Blondheim, S.H.; Wislicki, L. Factors Affecting Acetylation In-Vivo of Para-Aminobenzoic Acid by Human Subjects. *Clin. Sci.* **1964**, *27*, 133–141.
47. Benet, L.Z. Pharmacokinetics: I. Absorption, Distribution and Excretion. *Basic and Clinical Pharmacology*, 7th Ed.; Katzung, B.G., Ed.; Lange Medical Publications: Stamford CT, 1995; 31.
48. Bauer, J.H.; Brooks, C.S.; Burch, R.N. Clinical Appraisal of Creatinine Clearance as a Measurement of Glomerular Filtration Rate. *Amer. J. Kidney Dis.* **1982**, *2*, 337–347.
49. Wilkinson, G.R.; Branch, R.A. Effects of Hepatic Disease on Clinical Pharmacokinetics. *Pharmacokinetic Basis for Drug Treatment*; Benet, L.Z., Massoud, N., Gambertoglio, J.G., Eds.; Raven Press: New York, 1984; 49–61.
50. Benowitz, N.L. Effects of Cardiac Disease on Pharmacokinetics: Pathological Considerations. *Pharmacokinetic Basis for Drug Treatment*; Benet, L.Z., Massoud, N., Gambertoglio, J.G., Eds.; Raven Press: New York, 1984; 89–103.
51. Touw, D.J. Clinical Implications of Genetic Polymorphisms and Drug Interactions Mediated by Cytochrome P450 Enzymes. *Drug Metab. Drug Interact.* **1997**, *14*, 55–82.
52. Evans, D.A. Genetic Factors in Drug Therapy. *Scientific Basis Med. Ann. Rev.* **1969**, *16*, 166–182.
53. Buchert, E.; Woosley, R.L. Clinical Implications of Variable Antiarrhythmic Drug Metabolism. *Pharmacogenetics* **1992**, *2*, 2–11.
54. Dahl, A.K.; Bertilsson, L. Genetically Variable Metabolism of Antidepressants and Neuroleptic Drugs in Man. *Pharmacogenetics* **1993**, *3*, 61–70.
55. Daly, A.K. Molecular Basis of Polymorphic Drug Metabolism. *J. Mol. Med.* **1995**, *73*, 539–553.
56. Gonzales, F.J. Pharmacogenetic Phenytyping and Genotyping. Present Status and Future Potential. *Clin. Pharmacokinet.* **1994**, *26*, 59–70.

CLINICAL EVALUATION OF DRUGS

Allen Cato
Lynda Sutton

Cato Research Ltd., Durham, North Carolina, U.S.A.

Allen Cato III

Cato Research Ltd., San Diego, California, U.S.A.

INTRODUCTION

The process of developing a new drug, from the identification of a potential drug candidate to postmarketing surveillance, is extremely complex. The drug development process requires input from various members of a multidisciplinary team and the conduct of numerous studies. The time from drug discovery to marketing takes an average of 13 years. Once a chemical is identified as a new drug candidate, extensive preclinical analyses must be completed before the drug can be tested in humans. The pharmacology, toxicology, and preclinical pharmacokinetics must be characterized. The formulations of the drug product that were used in the preclinical studies may be different from the formulation of the final drug product, which may require that additional formulation work and pharmacokinetic analyses be performed. If the characteristics of the new drug candidate are acceptable for all of the preclinical assessments, it may then be tested in humans. The new drug candidate, at this point, enters the clinical research stage of drug development.

Clinical research represents a vital stage in the development process a stage that is no less daunting than the preclinical research stage. The data obtained from the first-time-in-human, Phase 1 pharmacokinetic studies, and initial safety evaluations in healthy volunteers can make or break the entire developmental program for a drug candidate. The sponsoring company, of course, hopes that the data collected in these initial studies will show minimal safety concerns over an adequate dose range. The pharmacokinetic data can then be used to help design future studies in which efficacy and long-term safety are assessed and additional pharmacokinetic and pharmacodynamic data are collected.

Although the basic designs of the initial single and multiple dose-escalating studies are generally straightforward (but the starting dose is often intensely debated), it is imperative that these studies and future studies be

designed to address specific questions. The questions vary depending on numerous specific considerations, including the targeted disease characteristics (e.g., acute or chronic); desired safety, efficacy, and pharmacokinetic evaluations; and assessment of clinical pharmacology (e.g., dosage formulations or dose frequency). Basic study procedures must also be considered. Thus, the design, conduct, data reporting and analysis, and production of the final study reports can be completed only through the coordinated efforts of a multidisciplinary drug development team.

For every clinical study, input is required from multiple personnel with various areas of expertise. Members of a drug development team include physicians, scientists, pharmacists, project managers, statisticians, computer programmers, study monitors, regulatory experts, and for some studies, a representative of the formulations group. While some team members may be able to perform multiple tasks, no one team member has the expertise or the time to do everything required to conduct a clinical study. In addition, some members may have overlapping abilities, but other members with particular expertise may be called upon. For example, pharmacokineticists are the experts in pharmacokinetics, but they may also be knowledgeable in pharmaceutics, biostatistics, and clinical care. However, scientists (PhDs) are trained primarily in basic research, while physicians (MDs) are trained in clinical medicine. Since a single drug development program is derived from both of these distinct disciplines, considerable overlap, cooperation, and coordination are necessary to take a drug successfully and efficiently from discovery to market.

Clinical drug development is generally divided into four phases: Phase 1 through Phase 4. For each study conducted within a particular phase, specific information is collected according to the requirements for individual drugs being developed. Collection of safety, efficacy, and pharmacokinetic data is the focus of most clinical trials. Although these topics appear to be distinct disciplines,

they are intertwined and represent different ways of evaluating the intrinsic properties of a drug. While the safety, efficacy, and pharmacokinetics of a drug may be assessed in most studies, the team must establish the type and extent of information to be collected, which will vary based upon the specific objectives and designs of the studies.

A critical function of the drug development team is the development of the study protocol. The study protocol must clearly describe the study design and methodology that will be used to achieve the study objectives. Input from nonmedical and nonscientific members of the team, such as marketing and information technology experts, also is helpful in establishing development strategies and in designing and conducting of clinical studies. Finally, project planning efforts can synchronize team efforts, help contain the soaring costs of pharmaceutical research, and coordinate international development efforts.

The drug development team's primary goal is to gain approval to market the drug, which requires that a marketing application be submitted to a regulatory agency (e.g., a New Drug Application [NDA] is submitted to the Food and Drug Administration [FDA] in the United States and to the Health Products and Food Branch [HPFB] in Canada, while a Marketing Authorization Application [MAA] is submitted to European regulatory agencies). During the conduct of the studies and the compilation and analyses of the data, the team must consider and evaluate many issues, such as how to collect, categorize, and report adverse events. All of these decisions will affect the marketing application that is submitted and may ultimately define how the drug is to be administered.

Many of the decisions to be made by the team, and particularly by the investigators, pose ethical dilemmas. Legislation has been enacted to protect human research subjects. Recently, the most pressing ethical dilemma facing the clinical research scientist concerned biotechnology and genetic engineering research.

Frequent changes in the regulations and guidelines of various regulatory agencies, differences in interpretations of these rules, and special reporting mechanisms for adverse events represent only a few of the challenges facing a drug development team. Due to continuous advances in scientific information, understanding of disease processes, and gene therapy, change continues to be the rule in modern drug development. However, through the efficient application of sound scientific principles in an ethical manner and with a coordinated team effort, effective new therapies can continue to be developed and marketed.

ROLES OF THE DRUG DEVELOPMENT TEAM MEMBERS

Physicians

The physician's contribution to drug development and the physician's role on a drug development team have changed over the last few decades (1). Before the 1960s, medical departments of pharmaceutical companies were primarily composed of physicians who were routinely involved in responding to drug information requests rather than developing new drugs. The Kefauver–Harris Amendment, enacted in 1962, required pharmaceutical companies to demonstrate before marketing that a drug was efficacious, which necessitated that physicians increase their presence on drug development teams. Along with the advent of additional governmental regulations, the increase in complexity of medical knowledge has mandated that physicians become an integral member of any drug development team. In fact, because of the different roles of the physician within an organization, companies may now have various departments (e.g., a clinical research department and a clinical safety department) within the medical department.

Although physicians are trained in patient care, physicians who are typically employed by pharmaceutical companies have more training in scientific methodology than those in the past. The physician on the team is the one qualified to follow the progress of each patient enrolled in a clinical trial and to interpret the results. Some physicians continue to spend time treating patients at a university hospital or a specific clinic where their specialty can be utilized and practiced, which allows these physicians to maintain sharp diagnostic skills. Also, some may perform basic research in academic settings to develop or maintain their knowledge and skills in basic research.

However, much of today's clinical research is actually conducted by investigators who are not employed by the company sponsoring the development of the drug. The physician on the drug development team must help in the selection of appropriate investigators to conduct the clinical studies. Pharmaceutical physicians may rely on colleagues who are experts in their respective fields and who have appropriate patient populations and facilities for the targeted research project. The physician is also the expert who deals with emergency situations that may arise during the course of a clinical research project, such as an overdose or severe adverse experience (SAE) that might be experienced with the drug. Similarly, the physician assists investigators who are responsible for evaluating the severity of adverse experiences (AEs) and determining

the causal relationship of the AEs to the drug under development.

The physician's involvement in clinical research does not end with the completion of the clinical study. Medical reports, clinical study reports, and sections of NDAs must be written. Interactions with regulatory agencies that require the physician's input may occur frequently. Physicians in clinical research may also be called upon to promote new drugs in a scientific environment by organizing symposia and workshops and by reviewing journal advertisements and promotional material for medical validity and accuracy.

The role of the physician in a clinical drug development program has expanded and has been refined in the last 40 years. Physicians increasingly contribute clinical and scientific expertise and administrative skills. Many physicians on drug development teams today spend most of their time designing and implementing studies and interpreting and reporting data rather than being in direct contact with patients. An experienced clinician is an important member of any drug development team.

Scientists

While a drug development team may have only one primary physician, it may have multiple scientists. Pharmacokineticists, pharmacologists, toxicologists, and pharmaceutical scientists are all involved in the clinical development of drugs. The contributions of scientists to a drug development project are derived from their experience in both scientific methodology and basic research (1).

Although physicians are trained in patient care, scientists are trained in problem-solving skills related to scientific research. To obtain a doctoral degree, a scientist must conduct research and write a dissertation that covers a topic of sufficient scope and depth. During this process, the scientist learns how to solve problems from different perspectives. The scientist also collects extensive data and performs data analyses, thereby gaining valuable insight into the considerations necessary to determine the feasibility of collecting data in a clinical trial. Also, some scientists, such as pharmacokineticists with a pharmacy background, may receive some clinical experience during their training as a scientist.

Scientists help design major portions of study protocols and clinical case report forms (CRFs). The study protocol is the overall plan that the study follows, and it must contain certain types of information, including the following: 1) background data on the targeted disease; 2) the empirical and structural formula of the drug being

studied; 3) preliminary pharmacology and toxicology of the drug (specific study objectives and designs); 4) the methods and materials to be used in the study; 5) information regarding drug packaging, labeling, dosage forms, and decoding procedures; 6) overdose management; 7) patient discontinuation procedures; 8) explanation of informed consent and provisions regarding institutional review board approval; and 9) any relevant references and appendices. The CRFs are the forms on which individual patient data are recorded during a clinical trial. From these data, clinical and statistical analyses are performed. All the information that is stipulated in the study protocol must be collected on the CRFs.

In conjunction with nonscientific personnel, scientists are responsible for ensuring that the CRFs will capture the appropriate information for each study subject according to the objectives, tests, and evaluations stipulated in the protocol. Careful attention must be given to the administration of special tests or collection of samples so that the timing of the assessments or sample collections do not conflict.

Experience in basic research enables the scientist to function as an important link between the basic research labs within the company and the drug development team. Departments specializing in drug metabolism, microbiology, pharmacology, and toxicology need feedback from early human safety and pharmacokinetic studies so they can continue to plan and conduct appropriate long-term animal studies. Thus, communication between the clinical scientist and the basic scientist is important throughout the progress of the drug development program.

Because clinical research has become increasingly more scientific, experts in the methodology of science are necessary for a complete research program. The drug development team's scientists may account for much of the scientific expertise, but the roles of the research team overlap to form a scientifically sound, medically astute cohesive group. In addition to scientific expertise, use of the scientist's administrative talents, such as organizational skills and familiarity with personnel practices, enables effective drug development. Thus, scientists with these skills are often employed in management positions in many organizations.

Pharmacists

The pharmacist's role on the drug development team has greatly expanded the professional opportunities of individuals with backgrounds in pharmacy. Pharmacists can provide valuable therapeutic insight into medical research. Training of pharmacists as clinical scientists with

both clinical skills and scientific research skills continues to be an emphasis at many pharmacy schools. Several programs have been devised for the education and development of the pharmacist as clinical scientist (2). Pharmacists have a broad knowledge in both clinical medicine and pharmaceuticals, and therefore are able to bridge the gap between the clinic and the laboratory.

Pharmacists' training focuses on drug therapies in disease states, whereas physicians' training focuses on the diagnosis of disease states. Studies regarding drug interaction, positive control, or drug comparison involve drugs that have been studied and marketed. Pharmacists can help in the design of such trials because of their knowledge of marketed drugs.

Additional roles of pharmacists appear in the areas of drug information and education and training. Pharmacists have the appropriate expertise in drug therapy to answer inquiries from physicians (and other health professionals) concerning both marketed and investigational drug products. Similarly, the clinic/laboratory bridge that the pharmacist builds makes this team member especially well suited to educate and train new employees in drug development. By offering both general and special skills, the research pharmacist blends clinical medicine with pharmaceutical science and is well qualified as an educator and drug information specialist.

Nonscientific Personnel

Drug development includes many tasks that may not require the specialized expertise of a physician or a scientist. Administrative skills, creativity, and excellent communication abilities, which are qualities not necessarily emphasized within traditional medical and scientific educational curricula, may be required for many of these tasks.

The administrative skills necessary for drug development include incorporating seemingly disparate but vitally linked concepts into a single overall plan. Integration planning may mean organizing study files into a logical sequence or helping to assemble the various parts of an NDA. In the first example, files must be set up in a way that can facilitate internal quality assurance audits and FDA inspections. In the second example, knowledge of the FDA's regulations and good abstracting capabilities are required.

Creativity is a quality that cannot be developed through formal training. Creativity requires bold conjecture and it expresses itself in newer, better ways to accomplish the same goals. An example of creativity in clinical drug research might involve the development of a variable

report that could support all of the different research documents that are generated by drug research teams. With such a variable report, common information need not be recreated each time another document is generated.

Excellent communication skills may be the most important quality for individuals working in drug development, even for those with strong medical backgrounds. Clinical research requires extensive interactions with personnel within the organization and with outside vendors or clinical sites. The information flow must be both efficient and accurate. For example, marketing departments must communicate frequently with medical departments so that marketing studies, advertising, and package inserts can be planned and evaluated. Individuals who lack strong science backgrounds but who have excellent communication skills often act as liaisons in these situations.

One aspect of clinical research that requires extensive contribution by the drug development team personnel is study monitoring. Study monitors oversee the planning, initiation, conduct, and data processing of clinical studies (3). While monitoring studies, monitors must communicate frequently with investigators and help ensure the data are being collected properly, FDA regulations are being followed, and any administrative problems are resolved as quickly as possible. Although monitors traditionally have had a nonscientific background, many monitors today have training in the basic sciences, and some even have advanced degrees, which allows them to better understand the scientific aspects of the project. Effective study monitors have a wide range of talents.

The many facets of a clinical research program afford individuals with varying types of training, education, and experience, the opportunity to contribute to the drug development process. Although some tasks clearly require the clinical or scientific expertise of a physician or a scientist, other tasks are better suited to those individuals with less specialized and more general capabilities.

STAGES IN CLINICAL DRUG DEVELOPMENT

Before clinical drug development can begin, many years of preclinical development occur, millions of dollars are spent, and countless decisions are made. Basic research teams consisting of chemists, pharmacologists, biologists, and biochemists first identify promising therapeutic categories and classes of compounds. One or more compounds are selected for secondary pharmacology evaluations and for both acute and subchronic toxicology testing in animal models. A compound that is

pharmacologically active and safe in at least two nonhuman species may then be selected for study in humans. Before the drug can be tested in humans, an Investigational New Drug (IND) application, which contains supporting preclinical information and the proposed clinical study designs, must be filed with an appropriate regulatory agency.

Clinical drug development follows a sequential process. By convention, development of a new drug in humans is divided into four phases: preapproval segments (Phases 1 through 3) and a postapproval segment (Phase 4) (1–4). The definitions of the three preapproval phases have relatively clear separations. However, the different phases refer to different types of studies rather than a specific time course of studies. For example, bioequivalence studies and drug–drug interaction studies are both Phase 1 studies, but they may be conducted after Phase 3 studies have been initiated. The generalized sequence of studies may be tailored to each new drug during development.

Phase 1

After the appropriate regulatory agency has approved a potential drug for testing in humans, Phase 1 of the clinical program begins. The primary goal of Phase 1 studies is to demonstrate safety in humans and to collect sufficient pharmacokinetic and pharmacological information to permit the determination of the dose strength and regimen for Phase 2 studies.

Phase 1 studies are closely monitored, are typically conducted in healthy adult subjects, and are designed to meet the primary goal (i.e., to obtain information on the safety, pharmacokinetics, and pharmacologic effects of the drug). In addition, the metabolic profile, adverse events associated with increasing dosages, and evidence of efficacy may be obtained. Because most compounds are available for initial studies as an oral formulation, the initial pharmacokinetic profile usually includes information about absorption. Additional studies, such as drug–drug interactions, assessment of bioequivalence of various formulations, or other studies that involve normal subjects, are included in Phase 1.

Generally, the first study in humans is a rising, single-dose tolerance study. The initial dose may be based on animal pharmacology or toxicology data, such as 10% of the no-effect dose. Doses are increased gradually according to a predetermined scheme, often some modification of the Fibonacci dose escalation scheme (5), until an adverse event is observed that satisfies the predetermined criteria of a maximum tolerated dose (MTD). Although the primary objective

is the determination of acute safety in humans, the studies are designed to collect meaningful pharmacokinetic information. Efficacy information or surrogate efficacy measurements also may be collected. However, because a multitude of clinical measurements and tests must be performed to assess safety, measurements of efficacy parameters must not compromise the collection of safety and pharmacokinetic data.

Appropriate biological samples for pharmacokinetic assessment, typically blood and urine, should be collected at discrete time intervals based upon extrapolations from the pharmacokinetics of the drug in animals. Depending on the assay sensitivity, the half-life and other pharmacokinetic parameters in healthy volunteers should be able to be evaluated, particularly at the higher doses. The degree of exposure of the drug is an important factor in understanding the toxicologic results of the study. Pharmacokinetic linearity (dose linearity) or nonlinearity will be an important factor in the design of future studies.

Once the initial dose has been determined, a placebo-controlled, double-blind, escalating single-dose study is initiated. Generally, healthy male volunteers are recruited, although patients sometimes are used (e.g., when testing a potential anticancer drug that may be too toxic to administer to healthy volunteers). These studies may include two or three cohorts, with six or eight subjects receiving the active drug and two subjects receiving placebo. The groups may receive alternating dose levels, which allow assessment of dose linearity, intrasubject variability of pharmacokinetics, and dose-response (i.e., adverse events) relationship within individual subjects.

Participants in the first study are usually hospitalized or enrolled in a clinic so that clinical measurements can be performed under controlled conditions and any medical emergency can be handled in the most expeditious manner. This study is usually placebo-controlled and double-blinded so that the drug effects, such as drug-induced ataxia, can be distinguished from the nondrug effects, such as ataxia secondary to viral infection. The first study in humans is usually not considered successfully completed until an MTD has been reached. An MTD must be reached because the relationship between a clinical event (e.g., emesis) and a particular dose level observed under controlled conditions can provide information that will be extremely useful when designing future trials. Also, the dose range and route of administration should be established during Phase 1 studies.

A multiple-dose safety study typically is initiated once the first study in humans is completed. The primary goal of the second study is to define an MTD with multiple

dosing before to initiating well-controlled efficacy testing. The study design of the multiple-dose safety study should simulate actual clinical conditions in as many ways as possible; however, scientific and statistical validity must be maintained. The inclusion of a placebo group is essential to allow the determination of drug-related versus nondrug-related events. The dosing schedule, which includes dosages, frequency, dose escalations, and dose tapering, should simulate the regimen to be followed in efficacy testing.

Typically, dosing in the second study lasts for 2 weeks. The length of the study may be increased depending on the pharmacokinetics of the drug so that both drug and metabolite concentrations reach steady state. Also, if the drug is to be used to treat a chronic condition, a 4-week study duration may be appropriate. To obtain information for six dose levels with six subjects receiving active drug and two receiving placebo for each of two cohorts, a minimum enrollment of 24 subjects should be anticipated. Similar to the first study in humans, these subjects would be hospitalized for the duration of the study.

Also similar to the first study, pharmacokinetic data must be obtained. These data will be used to help determine dosage in future efficacy trials. The new pharmacokinetic information that can be gathered includes the following: 1) determination regarding whether the pharmacokinetic parameters obtained in the previous acute safety study accurately predicted the multiple dose pharmacokinetic behavior of the drug; 2) verification of pharmacokinetic linearity (i.e., dose proportionality of C_{\max} and AUC) observed in the acute study; 3) determination regarding whether the drug is subject to autoinduction of clearance upon multidosing; and 4) determination of the existence and accumulation of metabolites that could not be detected in the previous single-dose study. A number of experimental approaches can be used to gather this information, and all require frequent collection of blood and urine samples. The challenge to the clinical pharmacokineticist is to design an appropriate blood sample collection schedule that will maximize the pharmacokinetic information, yet can be gathered without biasing the primary objective—determination of clinical safety parameters.

Phase 2

After the initial introduction of a new drug into humans, Phase 2 studies are conducted. The focus of these Phase 2 studies is on efficacy, while the pharmacokinetic information obtained in Phase 1 studies is used to optimize the dosage regimen. Phase 2 studies are not as closely monitored as Phase 1 studies and are conducted in

patients. These studies are designed to obtain information on the efficacy and pharmacologic effects of the drug, in addition to the pharmacokinetics. Additional pharmacokinetic and pharmacologic information collected in Phase 2 studies may help to optimize the dose strength and regimen and may provide additional information on the drug's safety profile (e.g., determine potential drug–drug interactions).

Efficacy trials should not be initiated until the MTD has been defined. In addition, the availability of pharmacokinetic information in healthy volunteers is key to the design of successful efficacy trials. The clinical pharmacokineticist assists in the design and execution of these trials and analyzes the plasma drug concentration data upon completion of the efficacy studies.

During the planning stage of an efficacy trial, the focus is on the dosage regimen and its relationship to efficacy measurements. Plasma drug concentrations for various dosages can be simulated based upon the data collected in the first two studies in humans. The disease or physiological states of the test patients (e.g., organ dysfunction as a function of age), concurrent medications (e.g., enzyme inducers or inhibitors), and the safety data obtained earlier must be considered when choosing an optimal dosage regimen for the study. In addition, if the targeted site of the drug is in a tissue compartment, theoretical drug levels in this compartment can be simulated, which may help scientists determine the appropriate times for efficacy measurements.

On completion of the efficacy trial, a therapeutic window for plasma drug concentrations can be defined by reviewing the correlation between plasma drug concentrations and key safety and efficacy parameters. The goal is to improve efficacy and safety of the drug by individualizing the dosage based upon previous plasma drug concentration profiles in the same patient.

Phase 3

If the earlier clinical studies establish a drug's therapeutic, clinical pharmacologic, and toxicologic properties and if it is still considered to be a promising drug—Phase 3 clinical trials will be initiated. Phase 3 studies enroll many more patients and may be conducted both in a hospital or controlled setting and in general practice settings. The goals of Phase 3 studies are to confirm the therapeutic effect, establish dosage range and interval, and assess long-term safety and toxicity. Less common side effects and AEs that develop latently may be identified. In addition, studies targeted to evaluate and quantify specific effects of the drug, such as drowsiness or impaired coordination, are conducted during this phase.

Phase 3 studies are also used to identify the most appropriate population or subpopulation for the study drug and to establish a place for the drug in its therapeutic class. A drug may be developed in a therapeutic class that already has effective alternatives, but the investigative compound may have a better safety profile than its established competitors. A Phase 3 clinical study can be designed to assess relative safety profiles.

Closer inspection of drug interactions is warranted in Phase 3 clinical trials. In many disease states, the use of polytherapy is quite common, and the risk of drug–drug interactions is high, both from pharmacokinetic and pharmacodynamic perspectives. The likelihood of drug interactions and semiquantitative estimates of magnitude may be predicted from *in vitro* data (6). The potential for interactions needs to be evaluated from two perspectives: the potential that the new drug may affect the pharmacokinetics of other drugs, and the potential that other drugs may affect the pharmacokinetics of the new drug. The former generally depends on the ability of the new drug to affect various enzyme and carrier-mediated clearance processes. Most notably, this concerns the cytochrome P450 (CYP) isoforms but could also involve conjugative enzymes and transporters, such as p-glycoprotein. Drugs may be an effective inhibitor without being a substrate of a CYP isoform, as is the case for quinidine's inhibition of CYP2D6.

The potential for significant drug–drug interactions caused by other drugs requires knowledge of the components of clearance for the new drug and the likelihood that known inhibitors will be coadministered. For drugs with multiple pathways and a broad therapeutic index, the need for formal interaction studies may be limited. Population pharmacokinetic analyses of data obtained from Phase 3 studies may be used to help discover and quantify drug interactions due to classes of drugs often associated with inhibition (e.g., macrolides, systemic antifungals, calcium channel antagonists, fluoxetine, paroxetine) or induction (e.g., anticonvulsants, rifampin).

Most early clinical trials are conducted at university medical centers with physicians who specialize in a certain area of medicine. When study drugs are eventually marketed, however, general practitioners will be prescribing them as well. Therefore, it is important that family physicians are exposed to study drugs during this phase because they represent the segment of clinicians who will be writing most of the prescriptions. Similarly, to maximize the commercial return on drug development, a multi-indication strategy may be pursued (sometimes designated as Phase 5 if conducted postapproval). In addition, testing of the drug in foreign countries is

appropriate during Phase 3; however, other countries may operate under different regulatory obligations than in the United States.

Phase 4

Whereas Phase 1, 2, and 3 studies are conducted prospectively using subjects or patients whose entrance into the study depends on strict inclusion and exclusion criteria, Phase 4 studies employ mainly observational, rather than exclusionary, study designs. Postmarketing surveillance and any additional studies requested by the regulatory agency as conditional approval of the NDA are conducted during Phase 4.

Data collection in premarketing clinical trials is an extensive, scientific exercise. Detailed blood work, special laboratory tests, and careful physiologic monitoring are typical in these studies. Postmarketing studies, however, are often targeted for much larger patient populations (5000–10,000 or more), which limits extensive data collection from each patient and emphasizes collection of safety information. These studies are complemented by reports of AEs from patients not enrolled in a study. The large numbers of patients in Phase 4 studies make it easier for researchers to determine rare AEs and can help identify patient populations that are at particular risk for certain AEs. For example, demographic trends toward side effects involving geographic locus, gender, or race may be determined from postmarketing surveillance data.

PROTOCOL CONSIDERATIONS

The task of designing a clinical study cannot be undertaken until the study objective of that trial has been rigorously defined. The objective should explicitly state what is being investigated and vague language should be avoided (7). Once an unbiased and specific objective has been developed, scientists can build the study design around it and then develop and write the protocol (8).

One of the main considerations when designing an investigational study concerns the type and number of comparative groups that will be involved. A control group of subjects may be evaluated in addition to the group taking the investigational drug. Sometimes more than one control group is used in a study. The control groups take either placebo or active medication and are compared with the group taking the investigational drug. This design is used to rule out the possibility of a placebo effect or to assess the efficacy and safety of the investigational drug relative to other drugs currently marketed.

Regulatory agencies frequently require the pivotal Phase 3 studies, which will be used to support an NDA, to be placebo-controlled studies. Placebo medication should be as similar as possible to the drug being investigated (e.g., same color, taste, and shape). No statistically significant difference in response between this group and the subjects taking the investigational drug is evidence against that drug having any real effectiveness.

Similar to the placebo considerations, active medication taken by the control group also should be as similar as possible to the drug being investigated (e.g., same color, taste, and shape). If the formulations cannot be made with similar appearances (e.g., tablet, suspension, etc.), a placebo of each formulation could be made so subjects would take one active formulation and the placebo of the other formulation to maintain the blind. No statistically significant difference in response in this group relative to the subjects taking the investigational drug is evidence that active medication has no advantage therapeutically over the existing therapy. However, a higher incidence of AEs in the control group and an equal rate of efficacy relative to the subjects taking the investigational drug are evidence of the new drug's advantage over the existing therapy.

In addition to determining the types and number of control groups that should be included in a study, the drug development team must decide between a parallel and a crossover design. For example, in a placebo-controlled clinical trial, a parallel design is one in which each study group takes the same medication (i.e., either placebo or active drug) throughout the study. With a crossover design, each study group eventually receives both placebo and active drug (e.g., one group may take placebo for a 6-week period and then cross over to receive active drug for the following 6-week period).

An advantage of the crossover design is that it allows each group to be its own control, thereby allowing a demonstration of efficacy to occur during the treatment with the drug. A disadvantage of the crossover design is that residual effects from one treatment period may carry over into the other treatment period. Absolute determination of efficacy and safety of the different treatments is difficult and sometimes impossible. One way to avoid the problem of residual effects on crossover studies is to have washout periods between the different treatment phases. During the washout period, the patient is either given a placebo or no treatment for several days or weeks so that any possible metabolite or effect of the drug is "washed out" of the patient before the next treatment phase begins.

An advantage of the parallel design is that it avoids the problems associated with possible residual effects of one treatment period influencing the other treatment period(s) because each treatment group is only exposed to one drug.

Compared with a crossover study, more patients may be required for a parallel study so that statistical significance can be established between the study groups. In a parallel study, recruiting the required larger numbers of patients who fit the study criteria takes longer, but the duration of that study is usually shorter than the duration of a crossover study.

Crossover designs span greater periods of time because each group must sequentially take an active and a control medication over a period that is long enough to allow a treatment effect to emerge. When washout periods are added, the time required to conduct these studies becomes longer still, and more study subjects may drop out. These difficulties are often outweighed by the fact that statistical significance can be achieved with fewer patients in crossover studies.

Once the study design has been chosen, there are many other issues to consider when developing and writing clinical protocols. Among the topics to be considered are criteria for patient eligibility, efficacy and safety parameters, timing of the events, packaging and dispensing of the clinical trial material, and the informed consent form. Also, to be determined is how the study will be blinded. For most well-controlled studies, subjects are assigned to the various groups by using a randomization process so that biased selection is eliminated, the overall collection of the subjects' variables is comparable in each group, and statistical power is guaranteed (9). In these double-blind studies, neither the subject nor the investigating scientists know to which group the subject has been assigned. Thus, extensive input from the drug development team is required when designing studies and writing protocols.

DRUG DEVELOPMENT CONSIDERATIONS

Most drugs are tested in humans to treat a specific disease entity or some adverse clinical condition. Because the pathogenesis of diseases and the exact mechanisms of action of drugs are often poorly understood, the process of evaluating a drug's efficacy can be complicated. Upon treatment, a patient's adverse clinical condition may improve; however, for many diseases this occurrence can only be evaluated indirectly by clinical assessments (e.g., via blood pressure measurements in the treatment of hypertension). However, a drug's characteristics can also be measured directly. For example, measurement of blood concentrations of the drug enabling calculation of pharmacokinetic parameters is a direct evaluation of the drug.

Similar to efficacy assessments, evaluation of the safety of a drug may also involve indirect measurements. One of the primary methods of obtaining safety information in a

clinical trial is through a patient's reporting of AEs. Although the exact biochemical mechanisms responsible for many AEs cannot be evaluated directly, the indirect evaluation of the drug's adverse effect can be seen clinically. Because clinical assessments are indirect measures, AE reporting leads to several complex questions. The degree of drug-relatedness or causality, the effect of concomitant medication, the severity of the AE, the complications of the disease state, and the effects of other clinical conditions or diseases are usually difficult to determine, particularly early in the drug development program. Also, all reports of AEs in a clinical drug research program are recorded, tabulated, and cross-referenced to form a safety database, regardless of whether the AE is determined to be drug related. The information contained in this database is used to generate the package insert.

Although the clinical effect of a drug is perhaps the primary concern of drug development, an understanding of the drug's biochemical and physicochemical properties and mechanism of action is also desired. These direct measures are of equal concern in drug development as are the indirect evaluations of a drug's clinical effects. The primary tool used to study the intrinsic physicochemical properties of a drug is pharmacokinetics, which is a branch of biopharmaceutics. Pharmacokinetics describes the relationship between the processes of drug absorption, distribution, metabolism (biotransformation), and excretion (collectively abbreviated ADME) and the time course of therapeutic or adverse effects of drugs (10). Efficacy is determined by the drug concentration at the site of action, which generally is correlated with the drug concentration in the blood. The ultimate goal of pharmacokinetics is to characterize the sources of variability in the concentration–time profile, which may be correlated with variability in efficacy and adverse events.

Pharmacokinetics can be used to guide dosage regimen selection and thereby optimize pharmacologic effects and minimize toxicologic effects when a drug is administered to an individual patient. Thus, although the basic pharmacokinetic properties of a drug are identified during the earliest stage of clinical drug development, the many factors affecting the pharmacokinetics in the patient population must be identified throughout the drug development process to enable proper dose selection for individuals. Thus, both indirect and direct measures are used to evaluate a drug.

MARKETING INPUT

A successful pharmaceutical company has an appropriate blend of both research and marketing to enable a

symbiotic, rather than antagonistic, relationship. Because an effective scientific and clinical research team often designs and executes experiments and clinical trials that involve costly overhead expenses, it is essential for marketing decisions to be geared toward company profitability being made allow the company profitable so these expenses can be met. Therefore, both medical and marketing input are necessary if a pharmaceutical company is to be successful.

By gathering data on all facets of the needs in the marketplace from clinicians and by maintaining a profile awareness of new products under development by competitors, marketing personnel are in an excellent position to advise their colleagues in the research arena who are responsible for the drug development program. Also, a marketing expert can help identify the problems other companies are having in selling their product and thereby avoid the same difficulties. For instance, sales problems may be related to ineffective advertising or faulty packaging; therefore, they do not concern clinical research. However, problems in sales can also be related to a drug's undesirable effects. An effective drug that does not lead to the AEs associated with an already approved drug would have a marketing advantage. Someone in marketing research may suggest conducting clinical studies that would evaluate the relative incidence of the AE with the hope that the data could be used to support effective advertising.

Thus, research and marketing are mutually beneficial in a successful pharmaceutical company. Marketing groups help clinical research teams by supplying them with information about competing products, the needs of the marketplace, and suggestions for new formulations. Clinical research teams provide the data to support therapeutic and marketing claims and act as chief advisors to marketing personnel concerning drug research studies and promotional claims.

EFFECTIVE GLOBAL PLANNING

Because drugs are frequently marketed worldwide and the clinical development of drugs may involve studies that are conducted internationally effective global planning can present its own difficulties. Obviously, medical practice, regulatory guidelines, and the cultural environment may be different in various countries, but also the manner in which research is conceived can differ vastly between countries. Medical researchers in some countries may be more conservative than researchers in other countries, which could potentially lead to the underdosing of drugs.

These differences in research approaches actually stem from differences in ethical standards.

Another reason that international planning may be difficult in drug research concerns the way in which various countries view early clinical trials and drug safety. Some countries view volunteer subjects and patients differently from a regulatory perspective, making it easier to recruit and enroll subjects for Phase 1 studies than it is to recruit and enroll patients for Phase 2 or Phase 3 studies. In the United States, both patients and volunteers are viewed in the same way, and studies with patients and volunteers cannot be initiated until the FDA has authorized an IND.

In addition to regulatory guidelines, the regulatory process is still another aspect of clinical drug development that can differ widely between countries. In England, sponsoring research firms do not interact very much with the British drug regulatory agency the Committee on Safety of Medicines. This lack of direct interaction stems from the desire to keep commercial influence away from the objective evaluation of a pharmaceutical company's study data. This lack of communication results in British companies treating government guidelines for conducting clinical research as a routine checklist rather than an aid in forming the most appropriate development strategy.

In the United States, federal guidelines (Code of Federal Regulations, CFR) have been established by the FDA to help sponsoring research firms conduct good, consistent clinical studies. However, some of the items in these guidelines may not be appropriate for all clinical studies, and some items that may be appropriate to include in a clinical study may not have been incorporated into the federal guidelines. These variations occur because each drug and disease state is unique, and complete guidelines cannot be established for all cases. For these reasons, several meetings are held between clinical research teams and the FDA before an NDA submission to ensure that all appropriate methodology and experimentation is being incorporated into the overall drug development project.

Beginning in the early 1990s, the FDA participated in a collaborative effort to harmonize the technical procedures for development and regulatory approval of human pharmaceuticals internationally. Forces that led the agency in this direction included increased trade, the multinational nature of the pharmaceutical industry, trade agreements such as the North American Free Trade Agreement and the General Agreement on Tariffs and Trade by the World Trade Organization, European activism, and pressures on the industry to control costs (11). These pressures included intense competition and health care reimbursement controls. This harmonization effort is the work of the

International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH has focused on achieving harmonization of technical requirements in three major regions of the world: the United States, the European Union, and Japan. Some of the earliest ICH guidelines addressed the format and content of the Investigator's Brochure (12), stability testing (13), and genotoxicity testing (14). The FDA also works with the World Health Organization and other international organizations to set standards for health care products (11).

Clinical drug research is a complicated, multidisciplinary task that may be conducted internationally. In fact, many pharmaceutical companies are multinational, with locations in several countries. Planning and coordination become even more complex for such global drug development programs. Despite the differences among countries in medical practice, regulation, and culture, international drug development and marketing are vital parts of many organizations. The successful multinational pharmaceutical company will plan its clinical research strategy according to any differences among nations before to implementing its international development plans.

ETHICAL CONSIDERATIONS

No topic in clinical drug development is more controversial and emotionally charged than the myriad ethical dilemmas that face physicians and scientists involved in clinical research. Given that clinical research has generally proved to have moral consequences through its direct and indirect influence on alleviating suffering, steps must be taken to ensure that abuses do not occur during the course of drug development. Therefore, guidelines for the protection of human subjects have been developed, proposed, and accepted worldwide (15).

Because of the atrocities committed by Nazi medical researchers in the 1930s, the Nuremberg Code (16) was written, and highlighted the importance of obtaining all research subjects' voluntary consent to their participation in clinical studies. The Declaration of Helsinki (17), which was published by the World Medical Association in 1964 and has been updated several times since, takes the informed consent issue one step further by giving only qualified medical scientists and physicians the right to conduct clinical research. However, similar concerns go back at least to the 1830s, when Dr. William Beaumont developed a contract with a patient, and in the late 1800s, when a leprosy worker experimented on a patient without her consent (18).

Legislation that ensures the protection of human research subjects in the United States includes the 1979 publication of the Belmont Report on the Ethical Principles and Guidelines for the Protection of Human Subjects of Research (19). This report concerns the fine line between biomedical research and the routine practice of medicine and explores the criteria that determine the risk-benefit ratio in the consideration of conducting clinical research. It also addresses basic guidelines for the proper selection of human research subjects and further defines the elements of informed consent.

Other important legislation in the United States includes the FDA's Guidance for Institutional Review Boards (IRBs) (20) for further guarantees of protection for human research subjects. IRBs are independent committees that review proposed clinical research projects before the commencement of the research. These committees decide whether the risk to research subjects outweighs the potential benefit of the research; they can suggest modifications in the research proposal or disapprove the project altogether. IRBs must consist of both men and women of varying professions. At least one member must have his or her primary concern in a nonscientific area (e.g., a lawyer or clergyperson), and at least one member must not be affiliated with the institution at which the research will be conducted.

Closely related to the rights of human research subjects are the rights of routine patients involved in nonresearch medical matters. In 1973, the American Hospital Association published the Patient's Bill of Rights (21), which requires that the acting physician give his patients complete information concerning their diagnosis, treatment, and prognosis; that the patient be given respectful care; that the patient be given the opportunity to refuse treatment; and that the patient's records, condition, and medical care be treated confidentially.

Another ethical issue facing clinical research scientists concerns study design, in particular, the placebo-controlled clinical trial. The reason placebo-controlled clinical trials are conducted is quite compelling from a scientific standpoint: to ensure that the evidence supporting the efficacy of an experimental drug is actually due to the properties of the drug and not to the psychologic properties of the study subjects. In other words, if a placebo effect from the experimental drug occurs rather than a true therapeutic effect, then a comparison of the drug group with the placebo group will show statistically similar response rates. It is a way to help separate actual drug responses from placebo responses, especially in studies investigating psychiatric compounds, but also in other therapeutic areas with a clearer "physiologic" or "biochemical" basis.

One defense for conducting placebo-controlled clinical trials is that the subjects chosen for the placebo group are randomly chosen, so that no malicious withholding occurs. Also, many study protocols have provisions of study extension that guarantee subjects in placebo groups have the opportunity to take the drug as an extension of the study after they complete the original part, or they are offered the chance to receive alternative therapy. Study subjects may be given monetary compensation for their participation in studies, in addition to free, thorough physical exams, lab work, and physician visits.

Interestingly, experimental drugs have unknown side effects that can cause serious biochemical and physiologic problems, whereas placebo medication does not. This fact makes possible the contrary argument and objection, on purely ethical grounds, to giving study subjects experimental and hence unproven drugs. Of course, informed consent and careful monitoring by trained medical personnel help to alleviate the ethical problems associated with giving subjects an active, investigational drug. The most important aspect of all studies is that the patient be completely informed of all study procedures and agree to willingly participate in the study.

The most recent pressing ethical dilemma facing the clinical research scientist surrounds the increasing amount of research that is being conducted in biotechnical and genetic engineering. Ethical issues will continue to play important parts in the medical and legal worlds. Whereas pure science is value-neutral, its application is always open to debate. Undesirable extremes are likely to exist at both ends of the spectrum.

CONCLUSIONS

To conduct a clinical study for the evaluation of a new drug, a vast array of personnel is required. Physicians are largely used because of their knowledge of clinical medicine and patient care, whereas scientists are used because of their knowledge of the methodology and the science. Pharmacists serve a bridging function due to their unique training in therapeutics and the pharmaceutical sciences. Nonscientific personnel are indispensable because of their ability to coordinate the many facets of a drug development project.

The clinical evaluation of drugs involves many different levels of scrutiny before a drug product can be marketed. These levels include Phase 1 for safety testing, Phase 2 for evaluating efficacy and determining the correct therapeutic dose, Phase 3 for large-scale studies and determination of drug interactions, and Phase 4 for

postmarketing surveillance. Phase 1 studies are typically conducted in healthy volunteers, and Phase 2 through 4 studies are conducted in patients.

Study design plays a critical role in the clinical evaluation of drugs. A clinical study cannot be conducted without specifically outlined objectives and a definitive plan, which are vital components around which the study protocol is constructed. The use of placebo or active drug control groups in the study, and whether the design should be open, parallel, or crossover, must be determined. In most studies, patients are assigned to study groups randomly.

The developmental objectives facing the clinical research team include indirect evaluations of a drug's safety and efficacy, such as effects on vital signs or behavior, and direct evaluations of a drug's intrinsic properties, such as its pharmacokinetics and mode of action. Also, the marketing-medical liaison is important if research is to support future sales plans and advertising is to reflect study results. Finally, effective global planning is necessary because drugs are more frequently developed and marketed worldwide, and therefore involve differing patient populations and different government regulations. ICH guidelines have helped to standardize regulations worldwide.

The ethical dilemmas facing clinical research scientists affect much of the legislation that currently regulates the conduct of clinical trials. The goal of drug development research is to develop effective pharmacotherapy for mankind's ailments, and regulatory agencies have enacted legislation to prevent unethical research.

Although traditional medicines continue to be discovered and developed, the fields of biotechnology and gene therapy continue to advance. In addition, new methods to collect and evaluate clinical data on a real-time basis will help to speed the development process.

REFERENCES

1. Cato, A.; Cook, L. Clinical Research. *The Clinical Research Process in the Pharmaceutical Industry*; Matoren, G.M., Ed.; Marcel Dekker, Inc.: New York, 1984; 217–238.
2. Smith, R.V. Development of Clinical Scientists. *Drug Intell. Clin. Pharm.* **1987**, *21*, 101–103.
3. Spilker, B. Monitoring a Clinical Trial. *Guide to Clinical Trials*; Lippincott-Raven: Philadelphia, 1996; 430–448.
4. Burley, D.M.; Glynne, A. Clinical Trials. *Pharmaceutical Medicine*; Burley, D.M., Binns, T.B., Eds.; Edward Arnold: London, 1985; 18–38.
5. Edler, L. Statistical Requirements of Phase I studies. *Onkologie* **1990**, *13*, 90–95.
6. Bertz, R.J.; Granneman, G.R. Use of In Vitro and In Vivo Data to Estimate the Likelihood of Metabolic Pharmacokinetic Interactions. *Clin. Pharmacokinet.* **1997**, *32* (3), 210–258.
7. Cato, A. Practical Insights on Designing the Correct Protocol. *Concepts and Strategies in New Drug Development*; Praeger Publishers: New York, 1983; 82–87.
8. Spilker, B. Part II: Developing and Writing Clinical Protocols. *Guide to Clinical Trials*; Lippincott-Raven: Philadelphia, 1996; 145–272.
9. Friedman, L.M.; Furberg, C.D. Basic Study Design. *Fundamentals of Clinical Trials*; PSG Publishing Co., Inc.: Littleton, MA, 1985; 35–38.
10. Gibaldi, M.; Levy, G. Pharmacokinetics in Clinical Practice I. Concepts. *JAMA* **1976**, *235*, 1864–1867.
11. Horton, L.R. Harmonization, Regulation, and Trade: Where Do We Go from Here? [editorial]. *PDA J. Pharm. Sci. Technol.* **1996**, *50*, 61–65.
12. Cocchetto, D.M. The Investigator's Brochure: A Comparison of the Draft International Conference on Harmonisation Guideline with Current Food and Drug Administration Requirements. *Qual. Assur.* **1995**, *4*, 240–246.
13. Haase, M. Stability Testing Requirements for Vaccines: Draft Guidelines of the International Conference on Harmonization. *Dev. Biol. Stand.* **1996**, *87*, 309–318.
14. Purves, D.; Harvey, C.; Tweats, D.; Lumley, C.E. Genotoxicity Testing: Current Practices and Strategies Used by the Pharmaceutical Industry. *Mutagenesis* **1995**, *10*, 297–312.
15. Levine, R.J. *Ethics and Regulation of Clinical Research*; Urban & Schwarzenberg: Baltimore, 1981.
16. The Nuremberg Code. *JAMA* **1997**, *276*, 1691.
17. *World Medical Association Declaration of Helsinki: Recommendations Guiding Physicians in Biomedical Research Involving Human Subjects*, Adopted by the 18th World Medical Assembly, Helsinki, Finland, June 1964 and Amended by the 29th World Medical Assembly, Tokyo, Japan, October 1975, the 35th World Medical Assembly, Venice, Italy, October 1983, the 41st World Medical Assembly, Hong Kong, September 1989, and the 48th General Assembly, Somerset West, Republic of South Africa, October 1996.
18. Lock, S. Research Ethics: A Brief Historical Review to 1965. *J. Intern. Med.* **1995**, *238* (6), 513–520.
19. Department of Health, Education, and Welfare. The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research April 18, 1979. Available at: <http://ohsr.od.nih.gov/mpa/belmont.php3> (accessed March 2, 2001)
20. Food and Drug Administration Information Sheets. Guidance for Institutional Review Boards and Clinical Investigators; 1998 Update. Available at: <http://www.fda.gov/oc/oha/IRB/toc.html> (accessed March 2, 2001).
21. American Hospital Association. A Patient's Bill of Rights. First adopted by the AHA in 1973. revision approved October 21, 1992. Available at: <http://www.aha.org/resource/pbillofrights.asp> (accessed March 2, 2001).

COACERVATION/PHASE SEPARATION

Bruno Gander

Institute of Pharmaceutical Sciences, Zürich, Switzerland

M.J. Blanco-Príeto

Universidad de Navarra, Pamplona, Spain

C. Thomasin

The R.W. Johnson Pharmaceutical Research Institute, Schaffhausen, Switzerland

Ch. Wandrey

D. Hunkeler

Laboratoire des Polyélectrolytes et BioMacromolécules, Lausanne, Switzerland

INTRODUCTION

Polymer coacervation is a long established, and widely used, method for reversible gelification and microencapsulation of biological materials such as liquid and solid drug compounds or cells. Coacervation is defined, by IUPAC, as the separation of colloidal systems into two liquid phases. It is to be distinguished from precipitation, which is observed in the form of coagulum or flocs and occurs in colloiddally unstable systems (1). The term coacervation was introduced in 1929 by Bungenberg de Jong and Kruyt, for a process in which aqueous colloidal solutions separate, upon alteration of the thermodynamic condition of state, into two liquid phases, one rich in colloid, i.e. the coacervate, and the other containing little colloid (2). Accounting for different phase separation mechanisms, coacervation was subdivided into simple and complex coacervation. In simple coacervation, the polymer is salted out by electrolytes, such as sodium sulfate, or desolvated by the addition of a water miscible nonsolvent such as ethanol, or by an increase or decrease in temperature. Conversely, complex coacervation is essentially driven by the attractive forces of oppositely charged polymers.

The earliest commercial application of coacervation was for the development of “carbonless” carbon copy paper by the National Cash Register Company in the late 1950s (3). More recently, the field of polymer coacervation has developed steadily so that a more refined and complete classification of coacervation systems can be proposed here (Table 1). Other classification schemes and related principles of coacervation for microencapsulation are available in the literature with illustrated examples (4–6).

In this contribution, we discuss coacervation as a phenomenon between one or two polymers in a solvent, although similar phenomena may occur with ionic or highly polarized forms of drugs such as between the anionic heparin and the cationic gentamicin or morphine (4), the anionic surfactants sodium cholate or sodium dodecylsulfate and cationic antidepressants (7), or anionic DNA and cationic forms of gelatin and chitosan in the presence of sulfate ions (8). Further, coacervation-like phenomena are increasingly used to engineer “smart” polymers that undergo reversible strong conformational and macroscopic changes upon small changes in the environment, e.g., pH, temperature, ionic strength (9). These polymers are essentially single or associated polyions designed for stimulus-responsive drug delivery, bioseparation, biomimetic actuators, or materials with switchable hydrophilic/hydrophobic surfaces.

According to the classification proposed in this paper (Table 1), polymer coacervation is generally observed in binary or ternary systems, in either aqueous or organic liquids. Three main mechanisms govern the process of coacervation in these systems: (i) Polymer desolvation, in binary and ternary systems, (ii) Polymer 2–Polymer 3 repulsion in a common solvent of the two dissimilar polymers, i.e., a ternary system; (iii) Poly(ion)–counterion interactions such as between poly(cation) and poly(anion) in a common solvent, i.e., a ternary system; similarly, a poly(H-donor)–poly(H-acceptor) interaction may also lead to polymer coacervation. Thus, polymer coacervation is a direct consequence of changes of molecular interactions operating between polymer–polymer (same species), polymer–solvent, polymer–coacervating agent, or poly(ion)–counterion. Prior to describing the various

Table 1 Classification of common aqueous and organic systems for polymer coacervation*Binary systems: Coacervation by partial polymer desolvation:*

Solvents (component 1) ^a	Water; Organic solvents
Polymers (P; component 2) ^a	Hydrophilic ^b P ⁰ , P ⁺ , P ⁻ ; Lipophilic P
Coacervation inducing factors	Temperature, pH

Ternary systems: Coacervation induced by partial polymer desolvation:

Solvents (component 1) ^a	Water; Organic solvents
Polymers (P; component 2) ^a	Hydrophilic ^b P ⁰ , P ⁺ , P ⁻ ; Lipophilic P
Coacervating agents (component 3) ^a	Nonsolvents for the polymer; Electrolytes ("Simple coacervation")

Ternary systems: Coacervation induced by Polymer 2–Polymer 3 repulsion:

Solvents (component 1) ^a	Water or organic solvent
Polymers (P2, component 2) ^a	Hydrophilic ^b P2 ⁰ , P2 ⁺ , P2 ⁻ ; Lipophilic P2
Coacervating agents (component 3) ^a	Polymer 3, P3

Ternary systems: Coacervation induced by noncovalent polymer cross-linking:

Solvents (component 1) ^a	Water ^c
Polymers (component 2) ^a	P ⁺ or P ⁻
Cross-linking agents (component 3) ^{a,d}	P ⁻ or P ⁺ ("Complex coacervation"); Di- and trivalent counter-cations or counter-anions

^aNumbering of components: 1-designates the solvent for the polymer to be coacervated; 2-designates the polymer to be coacervated; 3-designates the coacervating agent. The numbers are also used as subscripts in the equations.

^bP⁰, P⁺, P⁻ designate nonionic, cationic, and anionic polymers, respectively.

^cNonaqueous systems are not described in pharmaceutical technology.

^dP⁻ and P⁺ or the di- and trivalent counterions are integral components of the coacervate phase.

systems in more detail, it may be helpful to consider first some basic aspects of polymer solution behavior, knowledge that may be useful for understanding coacervation phenomena as well as for optimizing microencapsulation processes based on polymer phase separation.

Polymers dissolved in a solvent are encased in a sheath of solvent molecules that solvate their functional groups, typically through hydrogen-bonding and van der Waals forces. The envelope of solvation prevents chain segments in close proximity from attracting one another by interchain H-bonds, van der Waals or opposite ionic forces. Factors that lower the solvation of dissolved polymers thin out the sheath of solvation so that, at some point, contiguous chains attract one another by secondary valence bonds, thereby forming an entangled network or even noncovalent weak cross-links. Polymer chain desolvation is one type of mechanism leading to phase separation; under certain conditions, gelification rather than phase separation occurs. Factors that lower polymer solvation include temperature change, increase in molecular weight or, for poly(ions), pH-change in binary systems, or the increase in polymer concentration in binary or ternary systems. One very effective way to increase polymer solution concentration is to lower the number of solvent molecules available for polymer solvation. This can, be achieved practically, by adding a third component

to the polymer solution (ternary system) such as an electrolyte or a second liquid, which must be a nonsolvent for the polymer. The term nonsolvent is used here for all poor solvents for the polymer to be coacervated. The added electrolyte or nonsolvent will bind part of the polymer solvent. Competition for solvent of solvation will desolvate the polymer molecules leading to phase separation in the form of coacervates or precipitates. When electrolytes are used for polymer desolvation, the phenomenon is called salting-out. In aqueous systems, the effectiveness of dehydration, i.e., a particular form of desolvation, follows the so-called Hofmeister or lyotropic series, which arranges ions in the order of increasing salting-out capacity for hydrocolloids: $\text{NH}_4^+ < \text{K}^+ < \text{Na}^+ < \text{Ca}^{2+} < \text{Mg}^{2+}$ and $\text{Cl}^- < \text{acetate}^- < \text{SO}_4^{2-} < \text{tartrate}^{2-} < \text{HPO}_4^{2-} < \text{citrate}^{3-}$ (only pharmaceutically acceptable ions are indicated here).

When at least two dissimilar non-ionic, nonpolar, or only slightly polar polymers (Polymer 2 and Polymer 3) are mixed in a common solvent, phase separation generally occurs. This event is thermodynamically controlled and can be explained as follows (10). The dissolution of a polymer in a solvent is commonly endothermic (positive enthalpy of mixing, ΔH_m), thus counteracting dissolution. It is indeed the entropy increase (positive ΔS_m) that allows a polymer to dissolve in a

solvent, i.e., the entropy increases as the arrangement (or lattice) of the solvent molecules are largely disturbed by introducing long polymer chains, which require a relatively large molar volume inside the solution. When a second polymeric species (Polymer 3) is mixed with a solution of Polymer 2 in a common solvent, the two polymer species will typically interact through van der Waals forces, and this interaction is proportional to their molecular weights. This interaction would produce a substantial endothermic energy change whereas the entropy gain by this intermixing is very small owing to the small number of polymer molecules involved. Thus, because of the positive free energy change that would occur if the dissimilar polymers would mix with one another, phase separation into two distinct phases each of them rich in one of the two polymer species is thermodynamically more favorable.

Contrary to the aforementioned mechanism, pairs of oppositely charged poly(ions), or highly polarizable polymers, or of poly(H-donor) and poly(acceptor) tend to interact favorably with one another, i.e., their free interaction energy change is negative owing to a negative ΔH_m . Therefore, they may well coexist in a common solvent, or even attract one another so strongly that the negative ΔH_m dominates over the entropy gain in the common solvent. In this case, the two polymers will form a polymeric complex separating from the solvent. Depending on the strength of enthalpic interaction between the polymeric complex and the solvent, the complex may either precipitate as solid particles or remain partly solvated (complex coacervate). A comparable mechanism operates when a poly(ion) is mixed with a low molecular weight di- or trivalent counterion, such as Ca^{2+} , Mg^{2+} , Al^{3+} , Zn^{2+} , tartrate^{2-} . This type of mechanism leads to a strong non-covalent cross-linking of the polymer chains forming a relatively tight network.

Generally, all the mechanisms of polymer coacervation involve some sort of phase separation, thereby producing more or less dense coacervate microdroplets. These microdroplets can either engulf an additional component, such as a dissolved drug, or deposit on solid surfaces, which is typically used for coating solid particles added to the system, e.g., drug particles or living cells.

THERMODYNAMIC MODELS OF POLYMER SOLUBILITY AND PHASE SEPARATION

Fundamental aspects of coacervation have been thoroughly covered for some time through the classical

studies of Bungenberg de Jong and Kruyt for ionic systems (2), and by Dobry and Boyer-Kawenoki for non-ionic systems (11). The basic thermodynamic conditions for polymer-solvent interactions and polymer phase separation have been nicely described by Flory (10). In the following, polymer phase separation processes will be briefly considered from mechanistic and thermodynamic points of view.

Classical Models to Describe Coacervation Induced by Temperature Change and Nonsolvent Addition

In the early 1940s, Flory and Huggins proposed, separately, a lattice model to describe polymer solutions and introduced the interaction parameter χ (10). This parameter increases as solvent power decreases; hence, a thermodynamically good solvent is characterized by a low interaction parameter. In practice, most polymer-solvent combinations result in χ -values ranging from 0.2 to 0.6 (12). Moreover, the theory predicts that a polymer will dissolve in a solvent only if the interaction parameter is less than a critical value χ_c , which, at a given temperature, depends on the degree of polymerization (x) of the dissolved polymer (10):

$$\chi_c \approx \frac{1}{2} + \frac{1}{\sqrt{x}} \quad (1)$$

For polymers of very high molecular weight, χ_c approaches the value of 0.5. Upon gradual addition of a nonsolvent (with a large χ value), phase separation occurs in the order of decreasing x , i.e., when $\chi < \chi_c$. Further, χ also depends on temperature. This is generally illustrated by solubility phase diagrams, where single and two-phase systems are defined as a function of polymer volume fraction and temperature. The applicability of χ in coacervation is very limited, as χ is not a constant, but depends on polymer concentration, molecular weight, and temperature. Moreover, χ cannot be determined readily for polymer-liquid pairs and is inconvenient for multi-component systems. Finally, χ is a composite term influenced by hydrogen bonding.

More commonly used descriptors of polymer solubility are the solubility parameters introduced by Hildebrand and Scott (13) for dispersive interaction forces, and extended by Hansen (14) for dispersive (δ_d), polar (δ_p), and hydrogen bonding contributions (δ_h) to interaction energies. An equation sometimes used to estimate the solubility range of Polymer 2 in a solvent (subscript 1) is

$$\delta = [4(\delta_{d1} - \delta_{d2})^2 + (\delta_{p1} - \delta_{p2})^2 + (\delta_{h1} - \delta_{h2})^2]^{0.5} \quad (2)$$

According to Van Krevelen (15), $\Delta\delta$ should not exceed 5 MPa^{0.5} for good solubility. This rule of thumb may be applied to solutions with constant polymer concentration. However, the stepwise addition of a nonsolvent, followed by solvent partitioning between coacervate and continuous liquid, changes continuously the composition of a coacervation dispersion. Under the assumptions that the continuous phase consists exclusively of solvent and nonsolvent, and the coacervate phase of solvent and polymer, the solubility parameters of both phases may be calculated by

$$\delta_{\text{continuous phase}} = \phi_{1,\text{con}}\delta_{\text{solvent}} + \phi_3\delta_{\text{nonsolvent}} \quad (3)$$

$$\delta_{\text{coacervate phase}} = \phi_{1,\text{coa}}\delta_{\text{solvent}} + \phi_2\delta_{\text{Polymer 2}} \quad (4)$$

where $\phi_{1,\text{con}}$, $\phi_{1,\text{coa}}$, ϕ_2 , and ϕ_3 are the volume fractions of solvent in the continuous phase, of solvent in the coacervate phase, of polymer, and of nonsolvent, respectively. Although Equations 3 and 4 are empirical mixing equations with limited validity, they are frequently used in pharmaceutical technology to improve solvent mixtures for coating, film formation, and microencapsulation. The mean solubility parameter of the coacervation mixture can be calculated only for a single phase system, i.e., as long as the nonsolvent is soluble in the polymer solution. As solubility parameters of polymers cannot be determined directly from vaporization energy, various calculation methods have been proposed using the cohesive properties of functional groups and assuming additivity of these properties (16). Moreover, polymer solubility experiments in a variety of solvents of known solubility parameters, may be useful to define the borderline of good solvents and nonsolvents for polymer coacervation.

The Interaction Parameter χ to Describe Coacervation Induced by Polymer 2–Polymer 3 Repulsion

The theoretical treatment of polymer phase separation based on polymer-polymer repulsion requires an extension of the χ -parameter concept on two polymers in a common solvent. For this case, Scott (17) defined the critical conditions for phase separation, provided that the rather common conditions apply that $|\chi_{1,2} - \chi_{1,3}| \ll 1$ and $\sqrt{x_2} < \sqrt{x_3} < x_2 \dots$

$$(\chi_c)_{2,3} = 0.5 \left(\frac{1}{\sqrt{x_2}} + \frac{1}{\sqrt{x_3}} \right)^2 \left(\frac{1}{1 - \phi_1} \right) \quad (5)$$

where the subscripts 1, 2, and 3 refer to the solvent, the polymer to be coacervated (Polymer 2), and the second

polymer (Polymer 3; coacervating agent). If the Polymer 2 – Polymer 3 interaction parameter $\chi_{2,3}$ is approximated from solubility parameters (Eq. 6), estimated experimentally or calculated from group contributions

$$\chi \approx \frac{V_1}{RT} (\delta_2 - \delta_3)^2 \quad (6)$$

then, the critical volume fractions of the solvent (ϕ_1) and of both polymers (ϕ_2 , ϕ_3) can be calculated from

$$\frac{\phi_2}{1 - \phi_1} = \frac{\sqrt{x_2}}{\sqrt{x_2} + \sqrt{x_3}} \quad (7)$$

$$\frac{\phi_3}{1 - \phi_1} = \frac{\sqrt{x_3}}{\sqrt{x_2} + \sqrt{x_3}} \quad (8)$$

Unless the solubility parameters of both polymers are very similar, the polymers will be incompatible, leading to polymer phase separation, as $\chi_{2,3} > (\chi_c)_{2,3}$.

New Approaches to Characterize Polymer Coacervation

Another interesting approach to describe polymer coacervation is that of Van Oss (18), who defined the conditions for simple and complex coacervation by the interfacial interactions between similar and dissimilar molecules. In this model, the solubility, s , of the Polymer 2 in a solvent (subscript 1) depends on the free energy of interfacial interaction, ΔG_{212} :

$$RT \ln s = f(G_{212}^{\text{IF}}) \quad (9)$$

According to this equation, the solubility of the polymer molecules in a liquid will increase with increasing positive values of $\Delta G_{212}^{\text{IF}}$, as the polymer molecules will increasingly repel each other and thereby tend to disperse. If $\Delta G_{212}^{\text{IF}} < 0$, solubility decreases due to molecular attraction. In most pharmaceutical systems, molecular contributions to $\Delta G_{212}^{\text{IF}}$ will be from apolar Lifshitz-van der Waals ($\Delta G_{212}^{\text{LW}}$) forces and from Lewis acid/base ($\Delta G_{212}^{\text{AB}}$) forces. Hence, $\Delta G_{212}^{\text{IF}}$ is represented by

$$G_{212}^{\text{IF}} = -2 \left(\sqrt{\gamma_2^{\text{LW}}} - \sqrt{\gamma_1^{\text{LW}}} \right)^2 - 4 \left(\sqrt{\gamma_2^+ \gamma_2^-} + \sqrt{\gamma_1^+ \gamma_1^-} - \sqrt{\gamma_2^+ \gamma_1^-} - \sqrt{\gamma_2^- \gamma_1^+} \right) \quad (10)$$

where the second part of the right hand side describes the Lewis acid/base interactions. For a ternary system with two dissimilar Polymer 2 and Polymer 3 immersed in a solvent 1, $\Delta G_{212}^{\text{IF}}$ becomes

$$\begin{aligned}
G_{213}^{\text{IF}} = & \left(\sqrt{\gamma_2^{\text{LW}}} - \sqrt{\gamma_3^{\text{LW}}} \right)^2 - \left(\sqrt{\gamma_2^{\text{LW}}} - \sqrt{\gamma_1^{\text{LW}}} \right)^2 \\
& - \left(\sqrt{\gamma_3^{\text{LW}}} - \sqrt{\gamma_1^{\text{LW}}} \right)^2 \\
& + 2\sqrt{\gamma_1^+} \left(\sqrt{\gamma_2^-} + \sqrt{\gamma_3^-} - \sqrt{\gamma_1^+} \right) \\
& + 2\sqrt{\gamma_1^-} \left(\sqrt{\gamma_2^+} + \sqrt{\gamma_3^+} - \sqrt{\gamma_1^-} \right) \\
& - 2\sqrt{\gamma_2^+ \gamma_3^-} - \sqrt{\gamma_2^- \gamma_3^+}
\end{aligned} \quad (11)$$

For all exclusive Van der Waals interactions, the interfacial interaction energy between Polymer 2 and Polymer 3 immersed in a liquid reduces to

$$G_{213}^{\text{LW}} = \gamma_{23}^{\text{LW}} - \gamma_{21}^{\text{LW}} - \gamma_{31}^{\text{LW}} \quad (12)$$

which can be rewritten as

$$G_{213}^{\text{LW}} = \left(\sqrt{\gamma_2^{\text{LW}}} - \sqrt{\gamma_1^{\text{LW}}} \right) \left(\sqrt{\gamma_1^{\text{LW}}} - \sqrt{\gamma_3^{\text{LW}}} \right) \quad (13)$$

In coacervation processes induced by Polymer 2–Polymer 3 repulsion, the conditions for coacervation are, $\Delta G_{212}^{\text{IF}} > 0$. In apolar systems, this condition is fulfilled by $\gamma_2^{\text{LW}} > \gamma_1^{\text{LW}} > \gamma_3^{\text{LW}}$ or by $\gamma_2^{\text{LW}} < \gamma_1^{\text{LW}} < \gamma_3^{\text{LW}}$. In polar systems, however, coacervation becomes only predictable by solving Eq. 11. Contrary to simple coacervation, complex coacervation in a ternary system occurs only if the polymer molecules attract each other, and hence, the interfacial free energy becomes negative.

The apparent advantage of the Van Oss theory lies in the accessibility of the surface tension parameters. For a given solute, γ_2 can be determined from contact angle and Young's equation. Moreover, these parameters can also be derived theoretically from polymer group contributions as described by Van Krevelen (15). On the other side, it remains questionable if the Van Oss model is valid for systems where hydrogen bonding plays an important role, as interfacial tension measurements are unlikely to account for H-bonding forces.

POLYMER COACERVATION INDUCED BY PARTIAL POLYMER DESOLVATION (SIMPLE COACERVATION)

Process Description

Simple polymer coacervation is based on partial polymer desolvation in binary or ternary systems. This partial polymer desolvation may be induced by changing the temperature of the polymer solution, by adding to the polymer solution a poor solvent (or nonsolvent) for

the polymer, or by “salting-out” by electrolytes. The coacervate phase may form droplets in the stirred equilibrium phase or deposit as a film on a given solid or liquid surface, such as solid drug particles or droplets of aqueous drug solutions. In both situations, the coacervate can be stabilized by intermolecular physical or covalent cross-linking, which typically can be achieved by altering the pH or the temperature, or by adding a cross-linking agent (19). In pharmaceutical technology, simple coacervation, alike other coacervation types, is very frequently used to entrap drugs into microcapsules or micromatrices; the term microspheres will be used hereafter for both types of microparticles. If a drug should be entrapped in the coacervated polymer, a liquid or solid form of the drug is dissolved or dispersed in the polymer solution. Here, the principles of simple coacervation will be illustrated for two frequently used polymer types.

A very commonly employed simple coacervation procedure utilizes gelatin. Simple coacervation of gelatin typically involves the use of water-miscible nonsolvents for gelatin, such as alcohols, or salts, such as sodium sulfate. This produces a partial dehydration of the gelatin molecules at a temperature above the gelling point and leads to separation of a liquid gelatin-rich phase and an equilibrium liquid containing only minor amounts of gelatin. (Fig. 1). Finally, the coacervate droplets in the equilibrium liquid are hardened by physical or covalent cross-links (for example, by adjusting the pH to the isoelectric point or adding glutaraldehyde or formaldehyde).

A very frequently described family of polymers subjected to simple coacervation are cellulose derivatives, particularly ethyl cellulose (EC) (20). While most cellulose ethers are soluble in water, EC and the cellulose esters are insoluble or only partly soluble in water, e.g., as a function of pH. For coacervation of EC, toluene is a preferred good solvent and cyclohexane a poor solvent (21–23). Gradual addition of cyclohexane to a solution of EC desolvates the polymer. Alternatively, EC can be dissolved in hot cyclohexane; cooling to room temperature induces polymer phase separation. In both these cases, the coacervate film or droplets can be hardened by exposing the coacervate to a large volume of cyclohexane, whereby physical cross-links are formed.

Materials

Gelatin (24, 25) and cellulose derivatives (26, 27) are probably the most widely used polymers in simple coacervation for pharmaceutical purposes, although various other polymers have been successfully employed for the production of microcapsules by this process. In principle, any polymer can be utilized as a wall-forming material, provided that partial desolvation can be achieved.

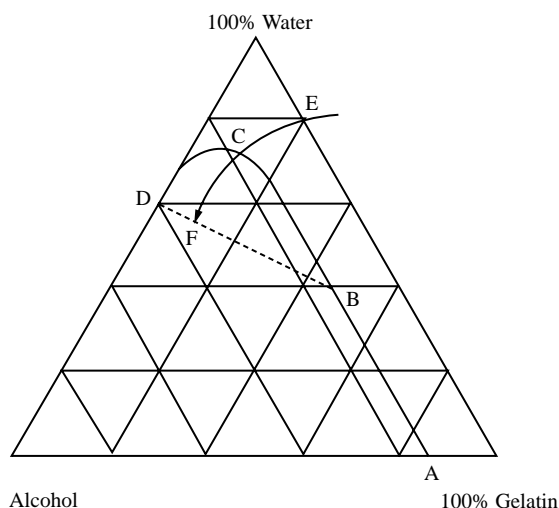


Fig. 1 Schematic ternary diagram of a gelatin–water–alcohol system. The diagram shows the zone where two phases exist, i.e., the coacervate and equilibrium phases. The two-phase systems are represented by the area under the binodal curve (points A, B, C, D), whereas the area above the binodal represents single liquid phase systems of gelatin dissolved in water–alcohol mixtures. The arrow (points E, C, F) indicates the addition of alcohol into an aqueous solution of gelatin at an initial concentration of 20% gelatin. At approx. 18% alcohol (point C), phase separation occurs. At a given composition (point F), the two phases have the compositions denoted by the points D (equilibrium phase consisting of 40% alcohol, and 60% water) and B (coacervate phase consisting of approx. 46% gelatin, 15% alcohol and 39% water) situated at both sides of the so-called tie-line (dotted line). The ratio of coacervate to equilibrium phases, on a weight basis, is given by the ratio $D-F/F-B$.

The polymers subjected to simple coacervation include, amongst others, poly(styrene), poly(vinyl acetate), poly(vinyl chloride), poly(lactide) (5), poly(vinyl alcohol) (28), poly(acrylates) (29), chitosan/PVA-blends (30, 31), albumin (32), casein (33), and also vegetable proteins like vicilin (34), and legumin (35). For more detailed information, the reader is advised to consult the extensive reviews in this field (5, 6, 36, 37).

For simple coacervation induced by nonsolvent addition in aqueous systems, ethanol, acetone, dioxane, isopropanol, and propanol are the most preferred to cause polymer desolvation and phase separation. In organic systems, mainly nonpolar solvents, such as petroleum ether, cyclohexane, and paraffin, are useful.

Critical Process Steps and Product Characteristics

Simple coacervation involves four distinct steps: i) Phase separation upon polymer desolvation; ii) Droplet formation

or deposition of the coacervate phase on a given surface; iii) Hardening of the coacervate phase; and iv) Isolation of microparticles or surface-coated material. Coacervation is quite a complex physico-chemical phenomenon, and many factors affect the properties of the resulting product. Simple coacervation in organic and aqueous systems depends on molecular interactions of the materials involved at a given temperature, the presence of solid or liquid surfaces with a high affinity for the coacervate phase, the rate of polymer desolvation and fluid dynamic processes, e.g., diffusion, laminar or turbulent movements. Therefore, for any given system, the material and process parameters must be carefully studied to control fully the process. In general, simple coacervation is a relatively slow process where diffusion and partitioning of the components between separated phases may take a considerable length of time to reach equilibrium. Therefore, the process is commonly performed under non-equilibrium conditions. This makes the rate and duration of every process step very critical for obtaining reproducible products and preventing undesired coalescence or precipitation. All these difficulties, along with the need for sometimes large amounts of organic solvents and toxicologically critical cross-linking agents have probably hampered the introduction of this process into an industrial setting.

Applications

To our knowledge, simple coacervation has essentially remained a technology described by academics and used for research rather than in pharmaceutical industry. Green (38) first demonstrated the microencapsulation of oil droplets by simple coacervation of gelatin. In this study, gelatin coacervation was induced by sodium or ammonium sulfate. Since then, simple coacervation has been used to encapsulate foods, flavors, and pharmaceuticals (19).

Simple coacervation of cellulose derivatives has been used for microencapsulation of various drugs, such as theophylline (27), ibuprofen (39), indomethacin (26), adriamycin (40), and nicardipine (41). The goal of microencapsulating these drugs was to decrease their gastric irritation, mask the bitter taste and, very importantly, to achieve sustained release.

POLYMER COACERVATION INDUCED BY POLYMER 2–POLYMER 3 REPULSION IN TERNARY SYSTEMS

Process Description

In coacervation by Polymer 2–Polymer 3 repulsion, the addition of Polymer 3 causes phase separation between the

two polymer species dissolved in a common solvent 1. This phase separation produces a viscous, liquid phase of Polymer 2, i.e., the coacervate, and a low-viscous phase of Polymer 3, often called continuous or polymer-poor phase. Under stirring, coacervate droplets are formed and dispersed in the continuous phase. The solubility of Polymer 3 in solvent 1 should be superior to that of Polymer 2 in this common solvent. For particle production, the Polymer 3 should also function as stabilizer for the coacervate droplets to prevent their aggregation. Further, for the entrapment of a biologically active material, the coacervate must have a certain degree of fluidity and a high affinity to the core material, whereas the affinity between core material and continuous phase should be low (see below). Finally, the coacervate droplets still contain a substantial amount of solvent 1 and are, therefore, too soft to isolate. Hardening of the coacervate droplets by a hardening agent is required before the product can be collected as discrete microspheres or microcapsules. A typical experimental set-up for this coacervation process yielding polymeric particles is illustrated in Fig. 2.

Materials

Coacervation by Polymer 2–Polymer 3 repulsion has been applied to various polymers, although the biodegradable poly(lactide) (PLA) and poly(lactide-co-glycolides) (PLGA) have attracted the highest interest, particularly for drug microencapsulation and controlled drug delivery (42). PLA/PLGA are commonly dissolved in dichloromethane or ethyl acetate and coacervated by adding a second, relatively low molecular weight polymer such as silicone oil (poly(dimethyl siloxane), PDMS, 100–1000 mPas). For hardening the coacervate droplets, the coacervate mixture is transferred into a nonsolvent for Polymer 2, such as hexane, heptane, or low molecular weight cyclic siloxanes, e.g., octamethylcyclotetrasiloxane (OMCTS) or decamethylcyclopentasiloxane (DMCPS).

To a very limited extent, other polymers have been coacervated by this process to form microparticles, i.e., ethylcellulose, cellulose nitrate, poly(methyl methacrylate), cellulose acetate phthalate, poly(acrylonitrile-co-styrene), and poly(styrene) (4). For some of these, liquid poly(butadiene) (8–10 kDa) was employed as a coacervating agent.

The hardening of coacervate droplets may be accomplished either by subsequent desolvation and formation of van der Waals bonds or covalent cross-linking of the coacervated polymer (4). In the first process,

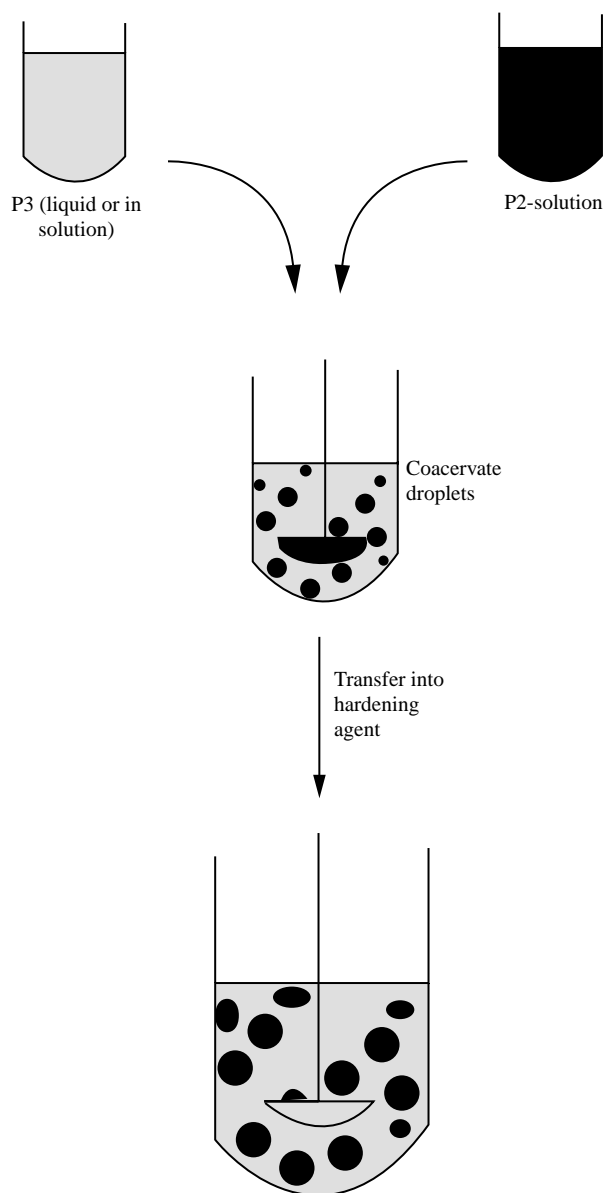


Fig. 2 Experimental set-up of a typical coacervation process by Polymer 2–Polymer 3 repulsion.

the coacervation mixture is transferred into a liquid which is a nonsolvent for Polymer 2, though it is a good solvent for Polymer 3. Examples of such hardening agents include hexane, heptane, very low molecular weight volatile siloxanes (OMCTS, DMCPS). In the chemical hardening process, a cross-linking agent, e.g., glutaraldehyde or diisocyanates, is added to the coacervate system. Typical functional groups on Polymer 2 that are useful for cross-linking reactions are hydroxyl, e.g., in cellulose ethers and esters, and amine groups.

Naturally, the aforementioned polymers may also be coacervated by desolvation upon adding a nonsolvent, such as hexane, heptane, liquid paraffin, or a vegetable oil. The advantage of using the coacervation by Polymer 2–Polymer 3 repulsion is that the viscosity and volume fraction of the coacervate phase and the stability of coacervated droplets can be controlled by the amount of added Polymer 3 (43). The control of these coacervate characteristics is important for preventing aggregation of coacervate droplets and for efficient microencapsulation of biologically active materials.

Critical Process Steps and Product Characteristics

The particular feature of coacervation by Polymer 2–Polymer 3 repulsion is that phase separation occurs already after the addition of a minute volume fraction of Polymer 3, which is in contrast to the coacervation by polymer desolvation (44). In the very first step, a dispersion of Polymer 3-in-Polymer 2 phase is formed (Fig. 3). Further Polymer 3 addition produces a phase inversion, whereupon the Polymer 2 phase (coacervate droplets) is dispersed in the Polymer 3 phase. Upon further Polymer 3 addition, the solvent is partially extracted from the coacervate droplets thereby increasing their viscosity and physical stability against coalescence. Optimal coacervate stability is generally achieved within a certain range of Polymer 3 volume fraction. This “stability window” has been determined by various authors for different PLA and PLGA types (43).

The main advantage of coacervation by Polymer 2–Polymer 3 repulsion over polymer desolvation resides in the good control of the composition and viscosity of both the coacervate and dispersing phases. This, in turn, provides a means to control particle size and prevent undesired coalescence of the coacervate droplets, as well as a way to enhance the wetting and engulfing of any biologically active material in either solid or liquid form. Indeed, encapsulation of core material requires primarily that the work of adhesion between core material and coacervate is substantially higher than that between core material and dispersing phase (44), which again depends on the respective composition of the two phases. Further, a certain degree of fluidity of the coacervate droplets will also increase qualitatively and quantitatively the encapsulation process.

Studies dedicated to drug microencapsulation by this type of coacervation method revealed that this method is quite suitable for the microencapsulation of peptide and protein drugs into PLA and PLGA in terms of

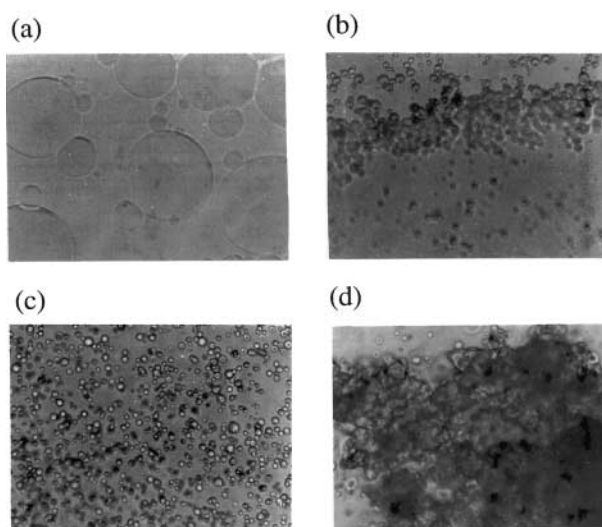


Fig. 3 Different stages of PLA coacervation by PLA–silicone oil (PDMS) repulsion in a ternary system using dichloromethane (DCM) as solvent: (a) Dispersion of PDMS in PLA/DCM; (b) Phase inversion yielding unstable droplets of PLA/DCM in PDMS; (c) Stable dispersion of well defined coacervate droplets; (d) Aggregation and precipitation of polymer particles upon exceeding PDMS-addition.

encapsulation efficiency. However, its drawback lies in the residual coacervating and hardening agents remaining in the microparticles that cannot be eliminated sufficiently and may hamper biomedical use. Therefore, efforts have been made to use coacervating and hardening agents that are relatively safe and to minimize the residues of processing liquids in the final product (45).

Applications

Coacervation by Polymer 2–Polymer 3 repulsion has been widely used in the field of drug microencapsulation into the biodegradable PLA and PLGA. (42) (see also the chapter on *Microsphere Technology and Applications* of this Encyclopedia). This type of microspheres is intended for parenteral administration and controlled delivery of low doses of very potent drugs. The microparticles produced by this method lie within a size range of 20–150 μm , depending on the process parameters, and can be injected by a conventional syringe and needle. To our knowledge, there is at least one PLGA-based microsphere product on the market for parenteral use that is produced by this particular method: Decapeptyl® Retard (Ferring AG, Kiel, Germany), which contains the LHRH-analog decapeptide triptorelin.

POLYMER COACERVATION UPON NONCOVALENT POLYMER CROSS-LINKING IN TERNARY SYSTEMS

Principles and Mechanisms

Complex coacervation in aqueous solution may be considered as a special case of network formation. Intermolecular forces such as Coulomb, van der Waals, hydrophobic, hydrogen bond, and dipole-charge transfer between polymers themselves, or polymers and low molar mass counterparts, cause phase separation, where the polymers are concentrated in a gel-like phase or a precipitate. The quality of the separated phases strongly depends on the chemical nature of the participating molecules and the separation conditions. The following combinations based on various dominating interactions have been reported:

1. Oppositely charged polyelectrolytes ("polysalt")
2. Highly polarizable polymers
3. H-donor/acceptor polymers
4. Polyelectrolytes/multivalent counterions

Fig. 4 illustrates the complex coacervation for systems 1 and 4. In system one, the complex is composed of two different polymer structures intermolecularly cross-linked (Fig. 4a), whereas in system 2, the complex is formed by the inter- and/or intramolecular bridging of a single polymer structure (Fig. 4b). Specific examples for the two complex types are sodium alginate (polyanion) with chitosan (polycation), and sodium alginate (polyanion)/calcium (divalent counterion), respectively.

From a kinetic point of view, the complex formation between a cationic and an anionic site is generally rapid, with rates in the order of fractions of seconds, though hours and days are occasionally required for the total

process (46). The rate-determining steps for complex formation are

- Diffusion processes for intermolecular contacts
- Rearrangements of the complex coacervate including both conformational changes and disentanglements.

In particular, these arrangements are relevant for polyelectrolyte components of low charge density, and/or polyelectrolyte components largely differing in molar mass, and concentrated systems (47).

The stoichiometry between polyelectrolytes has very often found to be 1:1 (48), though the optimal ratio of polyions can be as high as 20:1. Generally, as the polymer concentration decreases, the number of chains participating in the complex rises. Furthermore, the charge density required for coacervation is lower if the polymer-solvent interaction parameter (Flory-Huggins) is higher. Strong polyelectrolytes generally participate in 1:1 stoichiometries, though steric hindrance, for example via long pendant groups such as are typical in synthetic quaternary ammoniums, can disturb the so-called simplex ratio (49).

Macromolecular Characteristics for Effective Polymer-Polymer Coacervates

For both polymer-polymer complexation via ionic interactions and hydrogen bonding there is a critical chain length below which competitive binding is impossible (50). In general, polysalt formation is a function of various parameters including molar mass/chain length, charge density, chemical structure, type of the ionic group, and chain architecture. Interestingly, two polyelectrolytes of the same net charge can be complexed provided one is ampholytic, such as gelatin, and is polarized in the electric field of the other (51). In addition, medium conditions (pH, ionic strength, total concentration, temperature) influence the complex formation kinetics and complex properties.

Alginate-poly(L-lysine) and alginate-chitosan are, by far, the most common polymer-polymer coacervation systems, with the former often pregelled, in bead form, with divalent cations such as calcium or barium. As Table 2 indicates, polysaccharides with a rigid structure are generally favored for polysalt formation.

The polyanion-polycation combination generally involves one permanently charged polymer with a second whose charge density is pH-sensitive. Secondary interactions, principally hydrogen bonding, are usually required as is the flexibility of one of the chains (52).

For alginates, the copolymer composition (ratio of mannuronic to guluronic acid units) can influence the

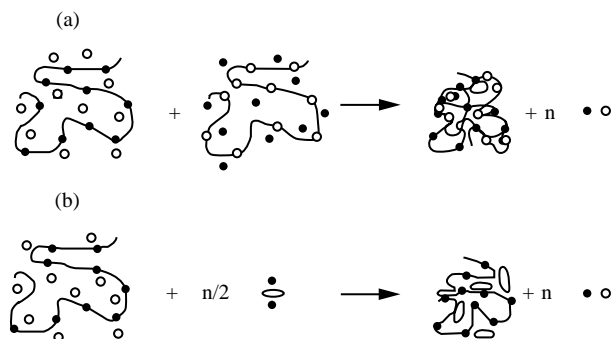


Fig. 4 Principles of complex aqueous coacervation: (a) Oppositely charged polyelectrolytes; (b) polyelectrolyte/multivalent counterions.

Table 2 Typical polymers employed in complex aqueous coacervation

Polyanions	Polycations
Alginate	Chitosan
Carrageenan	Poly(diallyldimethylammonium chloride)
Carboxymethylcellulose	Poly(L-lysine)
Chondroitin sulfate	Poly(vinylamine)
Cellulose sulfate	
Gellan	
Hyaluronic acid	
Poly(acrylic acid)	
Xanthan	

ultimate complex properties. These include elasticity as well as permeability and mechanical resistance of coacervates cast into 2D or spherical membrane structures. The type of polymer-polymer coacervate (precipitate, sol, network) will also often be highly molar mass dependent, with useful membranes found within a narrow window. This often does not correspond to the molar mass range required for bioapplications, which is dictated by factors such as cell toxicity and biocompatibility.

Characterization

The most important applications of polymer-polymer coacervates involve the formation of semipermeable membranes. They represent ionically cross-linked polymer network structures and are generally permeable to water and low molar mass solutes, though they block macromolecules and proteins above a certain molar mass. Molar mass cut-off (MMCO) experiments are carried out, for example, via inverse-size exclusion chromatography. These nontrivial experiments (53) monitor the ingress of a synthetic probe or the egress of biomacromolecules. The complex steric and electrostatic interactions involved in passing through the tortuous pores imply that the apparent MMCO is dependent on the chemistry of the macrosolute. In particular, a higher cross-linking density results in a lower MMCO, though this is rigorously true only if the network is inert, which is frequently not the case (54). Furthermore, most polyelectrolyte complex gels require characterization via transport studies (kinetics) as well as equilibrium diffusion measures (thermodynamics).

Polymer-polymer coacervates, in microcapsule form, can be characterized by a number of methods, though the most accurate involves a compression-based

micromanipulated probe connected to a sensitive transducer (55). The precision of such techniques can be as high as 10%. Generally, a 1:1 stoichiometry provides the most stable microcapsules. Other properties, including the sphericity, transparency, or membrane homogeneity are also often characterized. Novel methods based on analytical ultracentrifugation are particularly useful for 2D membranes (56).

Applications

Table 3 lists some of the typical polymer-polymer coacervation systems investigated for microcapsule formation.

The system sodium cellulose sulfate/PDADMAC permits the encapsulation of sensitive biological materials by a simple one step procedure. The broad variety of encapsulation problems so far successfully solved by this system includes the encapsulation of biocatalysts (57, 58), hepatic microsomes for extracorporeal detoxification (59), cattle embryos (47), and various drugs for targeted or controlled-release delivery (60).

Polysalts have been prepared as microcapsules for enzyme entrapment (biocatalysts), in the separation of proteins (60) as well as in surfactant binding. Protein separation is particularly selective, with biomacromolecules separated according to their isoelectric point. Coacervation is the preferred method for water soluble drugs where no modification of the absorption kinetics are warranted.

Coacervation for Cell Encapsulation

Processing

Cell encapsulation by polymer-polymer coacervation is generally performed by extrusion of a polyanion solution, seeded with the cells of interest, into a collecting batch containing a cationic simple multivalent electrolyte and/or polyelectrolyte. As an alternative method, the emulsification of polymer-cell suspension in a vegetable oil followed by the internal gelation of the polyanion has been described (62). The latter results in more homogeneous gels. Cell microencapsulation is complex (63) and involves the following steps:

- Polymer sterilization (thermal; UV- or γ -radiation; chemical, e.g., ethylene oxide)
- Polymer depyrogenation to destroy or remove potentially toxic lipopolysaccharides
- Polymer dissolution, often in culture media, to a controlled solution viscosity
- Solution dilution under sterile conditions

Table 3 Typical polymer–polymer coacervates

Polyanion	Polycation	Reference
Alginate (sodium)	Poly(L-lysine)	(76)
	Chitosan	(77)
Cellulose sulfate (sodium)	Poly(diallyldimethylammonium-chloride) (PDADMAC)	(60)
κ -Carrageenan	DEAE-dextran	(78)
Polyphosphate (sodium)	Chitosan	(79)
Chondroitin sulfate (sodium)	Collagen	(80)
	Gelatin	(81)
Dextran sulfate	Ionene	(82)
Poly(acrylic acid)	Poly(ethyleneimine)	(83)
Gelatin	Acacia	(84)
	Arabic gum	(85)
	Chitosan	(86)

- Coacervation reaction control (polyion concentrations, contact time)
- Washing, required to remove nonreacted excess polymer
- Coating to reduce permeability and/or modify surface properties

Examples of cell encapsulation

Fig. 5 shows microencapsulated mice islets intended for intraportal (liver) transplantation to achieve clinical normoglycemia. The islet's β -cells produce insulin in response to a blood glucose stimulus providing a therapeutic alternative to daily insulin injections. The capsule size is optimized to permit oxygen diffusion (64) given an O_2 -diffusion distance of 0.2 mm in hydrogels. The specific chemistry permits the simultaneous control of permeability and mechanical properties. From extended studies it has been concluded that multicomponent polymer systems can offer advantages in comparison to binary systems (65), with the thick membrane acting as an entrapment zone and permitting the de-coupling of diffusive and mechanical characteristics.

The encapsulation of hepatocytes for a bioartificial liver, and cell therapy for the treatment of other hormone deficiencies or neurodegenerative diseases, such as Alzheimer's and Parkinson's, are also under investigation. Additional examples of cell encapsulation in polymer-polymer coacervates include non-autologous gene therapy (66), blood substitutes (67) as well as the treatment of prostate cancer (68). Pharmaceutical applications of microcapsules encompass, in addition, transdermal drug delivery and protein delivery such as is required in anti-inflammatory therapy for arthritis.

Scale-Up

While it is not the purpose of this chapter to produce an extensive list of technologies related to polyelectrolyte complexes, some unique examples warrant discussion. Shioya (69) has patented an encapsulation method that controls permeability, while Dautzenberg disclosed a general method, and family of materials, for polyelectrolyte complexation (70). Recently, Vorlop has developed a mechanical cutting method to divide a fluid stream and achieve throughputs as high as 10^4 kg/h (71). Other techniques useful for high volume applications include the rotating disk atomizer.

Applications in Analytics

Polymer-polymer complexation is generally detected via conductometric or potentiometric titrations. "Colloid titration" represents an inverse-system where a polymer with known characteristics, such as potassium poly(vinylalcohol-sulfate) or poly(diallyldimethylammoniumchloride), are used to quantify the concentration of polycation or polyanion, hence relying on a 1:1 stoichiometry (47). Using the cationic dye, toluidine blue, as an indicator, a metachromatic end point is detected. Both methods are volumetric.

INDUSTRIAL VIEW OF THE USEFULNESS OF COACERVATION AND RELATED PRODUCTS

Various microencapsulation techniques have been successfully applied for several years in industry, with a large number of patent applications filed to protect market

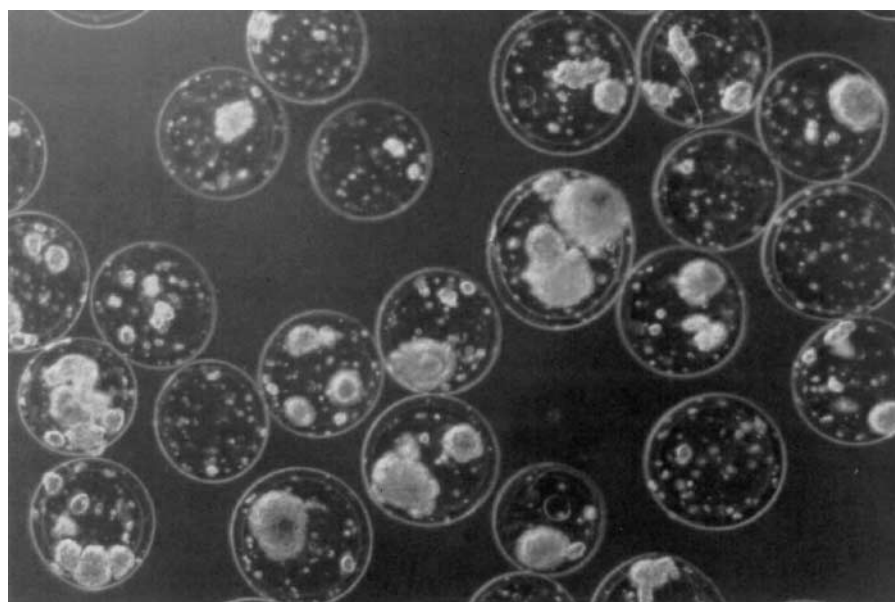


Fig. 5 Encapsulated mice islets in a 400 μm capsules made from sodium alginate/sodium cellulose sulfate/poly(methylene-co-guanidine) hydrochloride/ CaCl_2 .

shares and products, particularly in the area of contrast agents for diagnostic imaging, in agriculture, and in the food industry. In contrast to the many available consumer products based on encapsulation technology, only very few pharmaceuticals have succeeded in the market.

Primarily for safety reasons, manufacturing and marketing of pharmaceuticals is strongly controlled by regulatory guidelines, making product development and introduction especially challenging. Clearly, industrial large scale application of a technology such as coacervation primarily requires a sound and robust process and product design, which is the starting point of all full development work, but is also driven by other important criteria as outlined in Fig. 6. How do these basic considerations actually apply to products prepared by coacervation? To answer this question, the process of organic phase separation for preparing sustained release microspheres for parenteral administration is considered here as an example to illustrate the implications.

Key parameters for the quality of controlled release microspheres for parenteral use encompass the drug encapsulation efficiency, stability over time, syringeability, reproducibility of release kinetics, sterility, therapeutic efficacy, general safety, and local tolerability. Some of these aspects apply to the products made by all microencapsulation techniques, whereas others are mainly relevant to formulations prepared by coacervation. Furthermore for injection or aerosolizing, microspheres

require a compatible dispersant to wet the particles and keep them suspended for administration. The reconstitution vehicle will require developmental efforts as the surface properties of microspheres may be hydrophobic, especially when the particles have been prepared by coacervation and subsequent physical hardening in lipophilic hardening agent. From the authors' experience, insufficient attention is sometimes given to the development of the dispersant formulation, especially if

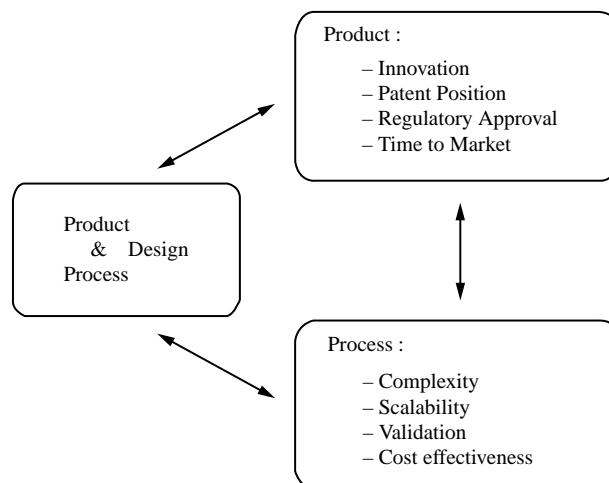


Fig. 6 Key elements for successful application of coacervation in the pharmaceutical industry.

active material and diluent are not manufactured at the same site.

Coacervation is a typical batch process requiring standard equipment with little need for costly investment, which contrasts to continuous microencapsulation processes such as aseptic spray-drying. For scalability, adequate and reproducible shear and stirring conditions, irrespective of the batch size, are very important. Selection of lab and pilot equipments suitable for scaling-up and featured with the necessary instruments to monitor critical parameters such as stirring and temperature are the basis for developmental work. Moreover, the introduction and dispersion of a given drug or bioactive material in the coacervation system also deserves particular attention. The particle size and dispersability of drug powder or drug solution are key parameters for high and consistent encapsulation efficiency and reproducible release kinetics. In organic phase separation, cleaning validation is a very challenging task, as coacervating and hardening agents require cleaning procedures with solvents and detergents suitable to remove efficaciously all coacervation components including non-entrapped drug or bioactive material. Because of this difficulty, the equipment of such critical processes are often product-dedicated to avoid any cross-contamination. In this respect, microencapsulation in aqueous systems is much easier to handle than in organic systems.

Nonetheless, organic phase separation remains an attractive method to encapsulate highly active, water soluble drugs such as peptides and proteins. The involvement of large quantities of organic solvents for production has, however, implications for the design of explosion-proof equipment and facilities. Further, residual solvents and hardening agents may hamper significant achievements for safety concerns. Efforts have therefore been devoted to replace the commonly used hydrocarbons or siloxanes by more biocompatible hardening agents such as isopropyl-myristate or propyleneglycol-octanoate/decanoate, which is used for the hardening of commercial PLGA microspheres containing triptoreline (Decapeptyl® Retard; Ferring). Evidently, such improvements are also relevant for patent protection (72, 73). Importantly, changes in processing solvents do not necessitate a new drug file application, but will be treated by the regulatory authorities as a postapproval change of an already approved process. Nowadays, where time to market is the key parameter for success, the choice of well characterized and widely accepted materials is of crucial importance. In this light, coacervation is a demanding technology because the number of useful and safe polymers, solvents and hardening agents is limited. Consequently, the quality, availability,

and cost of materials requires careful evaluation before decisions are made to introduce this technology in a pharmaceutical industry. Naturally, such a decision must also be driven by the therapeutic needs. Life saving indications, such as cancer treatment, may profit from a fast track approval by the authorities who balance carefully the pros and cons of possible formulation drawbacks and expected therapeutic benefit. Not surprisingly, microencapsulated parenteral depot formulations are found mainly in niche indications, where a limited and well controlled number of administrations in a relatively small group of patients is required.

Although coacervation is successfully used in pharmaceutical industry, the complexity of this and other microencapsulation technologies must not be underestimated, particularly in view of large scale manufacturing under aseptic conditions. Process design must consider this already at the developmental stage by keeping the number of individual manufacturing steps minimal. In this respect, sterile products prepared by organic phase separation require a significant number of unit operations:

- Sterile filtration/sterilization of excipients
- Mixing/dispersing and stirring
- Hardening by physical or chemical means
- Filtration
- Washing and sieving
- Dispensing and primary packaging (vials, syringes)
- Freeze-drying
- Terminal sterilization or full aseptic processing

For stability and safety reasons, full aseptic processing in a clean room environment is still the method of choice, although the intense involvement of personnel for a aseptic coacervation processing bears a substantial risk of contamination. By contrast, terminal sterilization by gamma or electron radiation provides a very high level of sterility assurance, but the efforts for validation and documentation of radiation sterilization are significant (74, 75). Moreover, radiation sterilization, although standard in the medical device industry, is not yet well accepted in all countries, a fact which again may hamper multinational filing and delay launch of a product.

In summary, the feasibility of coacervation has been shown on industrial scale, although it is probably fair to say that this technology is only scarcely used in pharmaceutical industry. Primarily for its physicochemical complexity and limitations with respect to residuals or processing components, efforts have been made to replace the coacervation technique by simpler technologies. To what extend new applications of coacervation technology in pharmaceutical industry will be successfully

commercialized will largely depend on the market needs and patient benefits of these new products, whereby the regulatory burden will play a major role. In this light, promising applications in the area of tissue engineering will definitely not be less challenging than those faced so far with the classical therapeutics.

CONCLUSIONS

Since the initial discovery of coacervation in the 1930s, a very large number of studies have dealt with coacervation, the underlying mechanisms, thermodynamics, materials, and applications. Nonetheless, relatively modest success and innovation in coacervation-based pharmaceutical systems have emerged. Clearly, the scientific achievements in this field, for example for the treatment of diabetes or the design of bioartificial organs, outweigh so far the commercial usefulness and benefit. Recently, however, coacervation has attracted great interest in the areas of DNA delivery from DNA-polymer complexes, in biotechnology, and in the food and agricultural fields. In the food sector, flavoring agents or agents with undesirable flavors, odors, acids, bases, artificial sweeteners, colorants, preservatives, leavening agents, among others, have been encapsulated by coacervation into materials such as starches, dextrans, alginates, proteins, and lipids. In the agricultural field, coacervation-based microencapsulation is used, e.g., for coating seeds. Finally, the ultimate relevance of coacervation may be in the biological existence and function of intracellular and extracellular complexes such as heparin-protein, DNA-protein, adrenergic agonists-protein. The association and dissociation of these complexes are very sensitive to pH, inorganic ions and electrical potential at membranes. These characteristics suit a plausible, though not elucidated process by which some pharmacological events may occur.

REFERENCES

1. Everett, D.H. IUPAC—Division of Physical Chemistry, Manual of Symbols and Terminology for Physicochemical Quantities and Units—Appendix II: Definitions, Terminology and Symbols in Colloid and Surface Chemistry. *Pure Appl. Chem.* **1992**, *31* (4), 577–638.
2. Bungenberg de Jong, H.G.; Kruyt, H.R. Koazervation (Entmischung in Kolloidalen Systemen). *Koll. Zeitsch.* **1930**, *50* (1), 39–48.
3. Bakan, J.A. Microencapsulation. *The Theory and Practice of Industrial Pharmacy*; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea & Febiger: Philadelphia, 1986; 413–429.
4. Newton, D.W. Coacervation: Principles and Applications. *Polymers for Controlled Drug Delivery*; Tarcha, P.J., Ed.; CRC Press: Boca Raton, 1991; 67–81.
5. Fong, J.W. Microencapsulation by Solvent Evaporation and Organic Phase Separation Processes. *Controlled Release Systems: Fabrication Technology*; Hsieh, D., Ed.; CRC Press: Boca Raton, 1988; 1, 81–108.
6. Arshady, R. Microspheres and Microcapsules, a Survey of Manufacturing Techniques, Part II: Coacervation. *Polym. Eng. Sci.* **1990**, *30* (15), 905–914.
7. Stott, P.W.; Williams, A.C.; Barry, B.W. Characterization of Complex Coacervates of Some Tricyclic Antidepressants and Evaluation of their Potential for Enhancing Transdermal Flux. *J. Controlled Release* **1996**, *41* (3), 215–227.
8. Leong, K.W.; Mao, H.-Q.; Truong-Le, V.L.; Roy, K.; Walsh, S.M.; August, J.T. DNA-Polycation Nanospheres as Non-Viral Gene Delivery Vehicles. *J. Controlled Release* **1998**, *53* (1–3), 183–193.
9. Galaev, I.Y.; Mattiasson, B. 'Smart' Polymers and What they Could Do in Biotechnology and Medicine. *Trends Biotechnol.* **1999**, *17* (8), 335–340.
10. Flory, P. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, New York, 1953; 495–594.
11. Dobry, A.; Boyer-Kawenoki, F. Phase Separation in Polymer Solution. *J. Polym. Sci.* **1947**, *2* (1), 90–100.
12. Robinson, D.H. Ethyl Cellulose-Solvent Phase Relationships Relevant to Coacervation Microencapsulation Processes. *Drug Dev. Ind. Pharm.* **1989**, *15* (14–16), 2597–2620.
13. Hildebrand, J.H.; Scott, R.L. *The Solubility of Nonelectrolytes*; Reinhold: New York, 1950; 424–434.
14. Hansen, C.M. The Three-Dimensional Solubility Parameter—Key to Paint Component Affinities: II And III. *J. Paint Technol.* **1967**, *39* (511), 505–514.
15. Van Krevelen, D.W. *Properties of Polymers*; Elsevier: Amsterdam, 1990; 200–220–290–297.
16. Barton, A.F.M. *Handbook of Solubility Parameters and Other Cohesion Parameters*; CRC Press: Boca Raton, FL, 1991; 95–111.
17. Scott, R.L. The Thermodynamics of High Polymer Solutions. V. Phase Equilibria in the Ternary Systems: Polymer1 – Polymer2 – Solvent. *J. Chem. Phys.* **1949**, *17* (3), 279–284.
18. Van-Oss, C.J. Acid-Base Effects At Polymer Interfaces. *Polymer Surfaces and Interfaces II*; Feast, W.J., Munro, H.S., Richards, R.W., Eds.; John Wiley: New York, 1993; 267–290.
19. Magdassi, S.; Vinetsky, Y. Microencapsulation of Oil-in-Water Emulsions by Proteins. *Microencapsulation: Methods and Industrial Applications*; Benita, S., Ed.; Marcel Dekker, Inc.: New York, 1996; 21–33.
20. Conti, B.; Giunchedi, P.; Conte, U. Cellulose Microparticles in Drug Delivery. *STP. Pharma Sci.* **1997**, *7* (5), 331–342.
21. Chen, H.; Wu, J.C.; Chen, H.Y. Preparation of Ethyl Cellulose Microcapsules Containing Teophylline by Using Emulsion Non-Solvent Addition Method. *J. Microencapsulation* **1995**, *12* (2), 137–147.
22. Kaeser-Liard, B.; Kissel, T.; Sucker, H. Manufacture of Controlled-Release Formulations by a New Microencapsulation Process, the Emulsion-Induction Technique. *Acta Pharm. Technol.* **1984**, *30* (4), 294–301.

23. Pongpaibul, Y.; Whitworth, C.W. Microencapsulation by Emulsion Non-Solvent Addition Method. *Drug Dev. Ind. Pharm.* **1986**, *12* (14), 2387–2402.
24. Madan, P.L.; Madan, D.K.; Price, J.C. Clofibrate Microcapsules: Preparation and Release Rate Studies. *J. Pharm. Sci.* **1976**, *65* (10), 1476–1479.
25. Rozenblat, J.; Magdassi, S.; Garti, N. Effect of Electrolytes, Stirring and Surfactants in the Coacervation and Microencapsulation Processes in Presence of Gelatin. *J. Microencapsulation* **1989**, *6* (4), 515–526.
26. Tirkkonen, S.; Urtti, A.; Paronen, P. Buffer Controlled Release of Indomethacin from Ethylcellulose Microcapsules. *Int. J. Pharm.* **1995**, *124* (2), 219–229.
27. Lavasanifar, A.; Ghalandari, R.; Atael, Z.; Zolfaghari, M.E.; Mortazavi, S.A. Microencapsulation of Teophylline Using Ethylcellulose: In Vitro Drug Release Kinetic Modelling. *J. Microencapsulation* **1997**, *14* (1), 91–100.
28. Bachtisi, A.R.; Kiparissides, C. Synthesis and Release Studies of Oil-Containing Poly(vinyl Alcohol) Microcapsules Prepared by Coacervation. *J. Controlled Release* **1996**, *38* (1), 49–58.
29. Genç, L.; Demirel, M.; Güler, E.; Hegazy, N. Microencapsulation of Ketorolac Tromethamine by Means of a Coacervation-Phase Separation Technique Induced by Addition of Non-Solvent. *J. Microencapsulation* **1998**, *15* (1), 45–53.
30. Miguez, M.J.B.; Rodrigues, B.C.; Sanchez, M.N.M.; Laranjeira, M.C.M. Preparation and Scanning Electronic Microscopy Study of Chitosan/Polyvinyl(alcohol)-Encapsulated Crude Urease Extract. *J. Microencapsulation* **1997**, *14* (5), 639–646.
31. Stolberg, J.; Laranjeira, M.C.M.; Sanchez, M.N.M.; Klug, M.; Fávère, V.T. Microspheres of Chitosan/Polyvinyl (alcohol) Incorporating Tetrasulphonated Copper (II) Phthalocyanine: Preparation and Characterization. *J. Microencapsulation* **1999**, *16* (4), 431–438.
32. Santhi, K.; Dhanaraj, S.A.; Rajendran, S.D.; Raja, K.; Ponnunsankar, S.; Suresh, B. Nonliposomal Approach—A Study of Preparation of Egg Albumin Nanospheres Containing Amphotericin-B. *Drug Dev. Ind. Pharm.* **1999**, *25* (4), 547–551.
33. Santinho, A.J.P.; Pereira, N.L.; de Freitas, O.; Collett, J.H. Influence of Formulation on the Physicochemical Properties of Casein Microparticles. *Int. J. Pharm.* **1999**, *186* (2), 191–198.
34. Ezpeleta, I.; Irache, J.M.; Gueguen, J.; Orecchioni, A.M. Properties of Glutaraldehyde Cross-Linked Vicilin Nano- and Microparticles. *J. Microencapsulation* **1997**, *14* (5), 557–565.
35. Irache, J.M.; Bergougnoux, L.; Ezpeleta, I.; Gueguen, J.; Orecchioni, A.M. Optimization and In Vitro Stability of Legumin Nanoparticles Obtained by a Coacervation Method. *Int. J. Pharm.* **1995**, *126* (1), 103–109.
36. Thies, C. Microcapsules As Drug Delivery Devices. *Crit. Rev. Biomed. Eng.* **1982**, *8* (4), 335–383.
37. Madan, P.L. Microencapsulation I. Phase Separation or Coacervation. *Drug Dev. Ind. Pharm.* **1978**, *4* (1), 95–116.
38. Microscopic Capsules Containing Oil, US Patent 2,800,458, July 23, 1957.
39. Weiss, G.; Knoch, A.; Laicher, A.; Stanislaus, F.; Daniels, R. Simple Coacervation of Hydroxypropyl Methylcellulose Phthalate (HPMCP) II. Microencapsulation of Ibuprofen. *Int. J. Pharm.* **1995**, *124* (1), 97–105.
40. Kawashima, Y.; Lin, S.Y.; Kasai, A.; Takenaka, H.; Matsunami, K.; Nochida, Y.; Hirose, H. Drug Release Properties of the Microcapsules of Adriamycin Hydrochloride with Ethyl Cellulose Prepared by a Phase Separation Technique. *Drug Dev. Ind. Pharm.* **1984**, *10* (3), 467–479.
41. Özyazici, M.; Sevgi, F.; Ertan, G. Micrometric Studies on Nicardipine Hydrochloride Microcapsules. *Int. J. Pharm.* **1996**, *138* (1), 25–35.
42. Thomasin, C.; Hô, Nam-Trân; Merkle, H.P.; Gander, B. Drug Microencapsulation by PLA/PLGA Coacervation in the Light of Thermodynamics. 1. Overview and Theoretical Considerations. *J. Pharm. Sci.* **1998**, *87* (3), 259–268.
43. Thomasin, C.; Merkle, H.P.; Gander, B. Drug Microencapsulation by PLA/PLGA Coacervation in the Light of Thermodynamics. 2. Parameters Determining Microsphere Formation. *J. Pharm. Sci.* **1998**, *87* (3), 269–275.
44. Thomasin, C.; Merkle, H.P.; Gander, B. Physico-Chemical Parameters Governing Protein Microencapsulation into Biodegradable Polyesters by Coacervation. *Int. J. Pharm.* **1997**, *147*, 173–186.
45. Thomasin, C.; Johansen, P.; Alder, R.; Bemsel, R.; Hottinger, G.; Altorfer, H.; Wright, T.; Wehrli, E.; Merkle, H.P.; Gander, B. Answers to the Problem of Residual Solvents in Biodegradable Microspheres Prepared by Coacervation. *Europ. J. Pharm. Biopharm.* **1996**, *42* (1), 16–24.
46. Sato, H.; Nakajima, A. Complex Coacervation in Sulfated Polyvinyl Alcohol-Amino Acetylated Polyvinyl Alcohol System. I. Conditions for Complex Coacervation. *Colloid & Polym. Sci.* **1974**, *252* (4), 294–297.
47. Dautzenberg, H.; Jaeger, W.; Kötz, J.; Philipp, B.; Seidel, C.; Stscherbina, D. *Polyelectrolytes: Formation, Characterization and Application*; Hanser: Munich, 1994.
48. Webster, L.; Huglin, M.B.; Robb, I.D. Complex Formation Between Polyelectrolytes in Dilute Aqueous-Solutions. *Polymer* **1997**, *38* (6), 1373–1380.
49. Kötz, J.; Koepke, H.; Schmidt-Naake, G.; Zarras, P.; Vogl, O. Polyanion-Polycation Complex Formation as a Function of the Functional-Groups. *Polymer* **1996**, *37* (13), 2775–2781.
50. Bekturov, E.A.; Bimendina, L.A. Complexes of Water-Soluble Polymers. *J. Macromol. Sci.* **1997**, *C37* (3), 501–518.
51. Tan, J.S.; Harrison, C.A.; Li, J.T.; Caldwell, K.D. Characterization of Soluble Polyelectrolyte-Gelatin Complexes by Differential Size-Exclusion Chromatography and Flow-Field Flow Fractionation. *J. Polym. Sci., Part B, Polym. Phys.* **1998**, *36* (3), 537–524.
52. Prokop, A.; Hunkeler, D.; DiMari, S.; Haralson, M.A.; Wang, T.G. Water Soluble Polymers for Immunoisolation I: Complex Coacervation and Cytotoxicity. *Adv. Polym. Sci.* **1998**, *136*, 1–51.
53. Bartkowiak, A.; Hunkeler, D. Alginate-Oligochitosan Microcapsules: A Mechanistic Study Relating Membrane and Capsule Properties to Reaction Conditions. *Chem. Mater.* **1999**, *11* (9), 2486–2492.
54. Gehrke, S.H.; Fisher, J.P.; Palasis, M.; Lund, M.E. Factors Determining Hydrogel Permeability. *Annal. New York Acad. Sci.* **1997**, *831*, 179–207.

55. Zhang, Z.; Saunders, R.; Thomas, C.R. Mechanical Strength of Single Microcapsules Determined by a Novel Micromanipulation Technique. *J. Microencapsulation* **1999**, *16* (1), 117–124.
56. Wandrey, C.; Bartkowiak, A. Membrane Formation At Interfaces Examined by Analytical Ultracentrifugation Techniques. *Colloids Surf. A: Physicochem. Eng. Asp.* **2001**, *180*, 141–153.
57. Stefuca, V.; Gemeiner, P.; Kurillova, L.; Dautzenberg, H.; Polakovic, M.; Bales, V. Polyelectrolyte Complex Capsules as a Material for Enzyme Immobilization. Catalytic Properties of Encapsulated Lactate Dehydrogenase. *Appl. Biochem. Biotechnol.* **1991**, *30* (3), 313–324.
58. Mansfeld, J.; Förster, M.; Schellenberger, A.; Dautzenberg, H. Immobilization of Invertase by Encapsulation in Polyelectrolyte Complexes. *Enzyme Microb. Technol.* **1991**, *13* (3), 240–244.
59. Stange, J.; Dümmler, W.; Brüggmann, E.; Falkenhagen, D.; Siegmund, E.; Ernst, B.; Dautzenberg, H. An Improved Hepatic Microsome Microencapsulation Method for Application in Extracorporeal Detoxication. *Z. Med. Laboratoriumsdiagn.* **1990**, *31* (7), 366–370.
60. Dautzenberg, H.; Loth, F.; Fechner, K.; Mehli, B.; Pommerening, K. Preparation and Performance of Symplex Capsules. *Makromol. Chem.* **1985**, (Suppl. 9), 203–210.
61. Mattison, K.W.; Brittain, I.J.; Dubin, P.L. Protein-Polyelectrolyte Phase Boundaries. *Biotechnol. Progr.* **1995**, *11* (6), 632–637.
62. Poncelet, D.; Lencki, R.; Beaulieu, C.; Halle, J.P.; Neufeld, R.J.; Fournier, A. Production of Alginate Beads by Emulsification/Internal Gelation. 1. Methodology. *Appl. Microbiol. Biotechnol.* **1992**, *38* (1), 39–45.
63. Hunkeler, D. Polymers for Bioartificial Organs. *Trends Polym. Sci.* **1997**, *5* (1), 286–293.
64. Hunkeler, D.; Rehor, A.; Canaple, L.; Bernard, P.; Renken, A.; Rindisbacher, L.; Angelova, N. Objectively assessing bioartificial organs. *Bioartificial Organs: Tissue Sourcing Immunoisolation and Clinical Trials*; Hunkeler, D.; Cherrington, A.; Prokop, A.; Rajotte, R., Eds.; New York Academy of Sciences, November, 2001.
65. Prokop, A.; Hunkeler, D.; Powers, A.C.; Whitesell, R.R.; Wang, T.G. Water Soluble Polymers for Immunoisolation II: Evaluation of Multicomponent Microencapsulation Systems. *Adv. Polym. Sci.* **1998**, *136*, 53–73.
66. Awrey, D.E.; Tse, M.; Hortelano, G.; Chang, P.L. Permeability of Alginate Microcapsules to Secretory Recombinant Gene-Products. *Biotech. Bioeng.* **1996**, *52* (4), 472–484.
67. Chang, T.M.S. Future Prospects for Artificial Blood. *Trends Biotechnol.* **1999**, *17* (2), 61–67.
68. Machluf, M.; Orsola, A.; Atala, A. Controlled Release of Therapeutic Agents: Slow Delivery and Cell Encapsulation. *World J. Urol.* **2000**, *18* (1), 80–83.
69. Shioya, T.; Hirano, R. Process for Producing Capsules Having a Permeability-Controllable Membrane, US Patent 5, 089, 272, February 18, 1992.
70. Dautzenberg, H.; Loth, F.; Pommerening, K.; Linow, K.J.; Bartsch, D. Microcapsules and Process for the Production Thereof., UK Patent Application GB 2, 135, 954, September 12, 1984.
71. Prusse, U.; Fox, B.; Kirchhoff, M.; Bruske, F.; Breford, J.; Vorlop, K.-D. New Process (Jet Cutting Method) for the Production of Spherical Beads from Highly Viscous Polymer Solutions. *Chem. Eng. Technol.* **1998**, *21* (1), 2933.
72. Komen, J.; Groenendaal, J.W. Process for Microencapsulation, US Patent 5, 066, 436, November 19, 1991.
73. Nerlich, B.; Gustafsson, J.; Mank, R.; Hörig, J.; Köchling, W. Process for the Microencapsulation of Water-Soluble Medicaments, DE Patent 42 23 169 C1, November 25, 1993.
74. The European Agency for the Evaluation of Medicinal Products. *Decision Trees for the Selection of Sterilisation Methods*, London, 1998.
75. European Guide to Good Manufacturing Practice, 1997, Annex 12, *The Use of Ionising Radiation in the Manufacture of Medicinal Products*, Brussels, 1997.
76. Lim, F.; Sun, A.M. Microencapsulated Islets As Bioartificial Endocrine Pancreas. *Science* **1980**, *210* (4472), 908–910.
77. Thu, B.; Bruheim, P.; Espevic, T.; Smidsrød, O.; Soon-Shiong, P.; Skjak-Braek, G. Alignate Polycation Microcapsules. 1. Interaction Between Alginate and Polycation. *Biomaterials* **1996**, *17* (10), 1031–1040.
78. Loeferoth, J.E.; Johansson, L.; Norman, A.C.; Wettström, K. Interaction Between Surfactants and Polymers. *Progr. Colloid Polym. Sci.* **1991**, *84*, 78–82.
79. Kawashima, Y.; Handa, T.; Kasai, A.; Takenaka, H.; Lin, S.; Ando, Y. Novel Method for the Preparation of Controlled-Release. Theophylline Granules Coated With a Polyelectrolyte Complex of Sodium Polyphosphate-Chitosan. *J. Pharm. Sci.* **1985**, *74* (3), 264–268.
80. Shao, W.; Leon, K.W. Microcapsules Obtained from Complex from Complex Coacervation of Collagen and Chondroitin Sulfate. *J. Biomater. Sci., Polym. Ed.* **1995**, *7* (5), 389–399.
81. Brown, K.E.; Leong, K.; Huang, C.H.; Dolal, K.; Green, G.D.; Hianes, H.D.; Jimenex, P.A.; Bathon, J. Gelatin/chondroitin 6-sulfate Microspheres for the Delivery of Therapeutic Proteins to the Joint. *Arthritis Rheum.* **1998**, *41* (12), 2185–2195.
82. Davison, C.J.; Smith, K.E.; Hutchinson, L.E.F.; O'Mullane, J. Physical and Biological Properties of Water Soluble Polyelectrolyte Complexes. *J. Bioact. Compat. Polym.* **1990**, *5* (3), 267–282.
83. Kono, K.; Tabata, F.; Takagishi, T. pH Responsive Permeability of Poly(acrylic Acid)-Poly(ethylenimine) Complex Capsule Membrane. *J. Membrane Sci.* **1993**, *76* (2–3), 233–243.
84. Burgess, D.J.; Carless, J.E. Microelectrophoretic Studies of Gelatin and Acacia for the Prediction of Complex Coacervation. *J. Colloid Interface Sci.* **1984**, *98* (1), 1–8.
85. Lamprecht, A.; Schäfer, U.F.; Lehr, C.M. Visualization and Quantification of Polymer Distribution in Microcapsules by Confocal Laser Scanning Microscopy (CLSM). *Int. J. Pharm.* **2000**, *196* (2), 223–226.
86. Remunanlopex, C.; Bodmeier, R. Effect of Formulation and Process Variables on the Formation of Chitosan-Gelatin Coacervates. *Int. J. Pharm.* **1996**, *135* (1–2), 63–72.

COLORING AGENTS FOR USE IN PHARMACEUTICALS

David R. Schoneker

Colorcon, West Point, Pennsylvania, U.S.A.

INTRODUCTION

The Need for Color in Pharmaceuticals

There is evidence that the coloring of pharmaceuticals has been practiced since antiquity, and although the use of colorants in medicinals produces no direct therapeutic benefit, the psychological effects of color have long been recognized.

The coloring of pharmaceutical dosage forms is extremely useful for identification during manufacturing and distribution. Many patients rely on color to recognize the prescribed drug and proper dosage. Unattractive medication can be made more acceptable to the patient by the careful selection of color. Color, in combination with flavoring agents, can be used to provide taste masking of disagreeable components of a pharmaceutical preparation. These attributes assist in improving patient compliance, which is known to be a significant problem when providing drug therapy for use in the home.

Additionally, recent reports defining the scope and magnitude of medical errors that have occurred are alarming. Although most of the medical error problems are related to practices by doctors and hospitals, some medical errors are related to mistakes made in identification of drug products by pharmacists and patients. Many drugs on the market look very similar (i.e., small, round, white, uncoated tablets) identified only by a small embossed or printed logo and identifier. This makes it very easy for a pharmacist or patient to confuse one drug for another. Obviously, this could prove to be very serious, depending on the drugs. The use of different colors for different drugs and strengths can help to eliminate this problem.

A number of pharmacy associations recently introduced a resolution at the 2000 USP Convention to develop standards for dosage form consistency that would include color, shape, size, labeling, and packaging. These types of standards would assist in developing new drug products with colors and other attributes that would make identification of the specific drug much easier. Although it would be very difficult to change the colors of existing drug products because of the amount of supporting information and regulatory review that would be needed, standards for future development might be beneficial.

History of Colorant Use in Foods, Drugs, and Cosmetics

Historical accounts describe many colors derived from animal, vegetable, or mineral sources that were used for coloring foods, drugs, and cosmetics. Until the middle of the nineteenth century, all colors were obtained from natural sources. Archaeologists date cosmetic colors as far back as 5000 BC. Ancient Egyptian writings tell of drug colorants, and historians say food colors likely emerged around 1500 BC.

The synthetic color industry dates back to the accidental discovery of the first synthetic organic dye (mauve) in 1856. Sir William Henry Perkin, in an unsuccessful attempt to synthesize quinine, succeeded in obtaining a violet dye by the oxidation of aniline. This led other scientists to experiment and discover many new colors with superior properties to the natural pigments and extracts. The use of these new and different colors in foods, drugs, and cosmetics began almost immediately because of their tinctorial value, stability, and the many shades in which they were available.

As the 1900s began, the bulk of chemically synthesized colors were derived from aniline, a petroleum product that in pure form is toxic. Originally, these were dubbed “coal-tar” colors because the starting materials were obtained from bituminous coal. Many of these colors are still used today, although with controls that ensure safe use.

Although colors from plant, animal, and mineral sources—at one time the only coloring agents available—remained in use early in this century, manufacturers had strong economic incentives to phase them out. Chemically synthesized colors simply were easier to produce, less expensive, and superior in coloring properties. Only small amounts were needed. They blended nicely and didn’t impart unwanted flavors to foods and drugs. But as their use grew, so did safety concerns.

Early uses of synthetic colors were at times a threat to health because they were used without discrimination between those that were toxic and those that were safe. Increasing public concern around the world led to early studies and regulations that produced various lists of colors found suitable for addition to foods and drugs.

The United States was one of the first countries to significantly regulate food and color additives because of a

number of incidents that occurred related to the safety of these materials. In 1906, the U.S. Congress passed the Pure Food and Drug Act. This marked the first of several laws allowing the federal government to scrutinize and control additives use. The act, however, only covered food coloring.

The Pure Food and Drug Act of 1906 listed seven colors, based on the work of Dr. Bernhard Hesse, and the Department of Agriculture made provisions for certification of all food colors on a voluntary basis.

It was not until passage of the Federal Food, Drug, and Cosmetic (FD&C) Act of 1938 that the Food and Drug Administration's (FDA's) mandate included the full range of color designations consumers still can read on product packages: "FD&C" (permitted in foods, drugs and cosmetics), "D&C" (for use in drugs and cosmetics), and "Ext. D&C" (colors for externally applied drugs and cosmetics). The 1938 FD&C Act also made certification mandatory.

Public hearings and regulations that occurred after the 1938 law gave colors the numbers that separate their hues. These letter and number combinations—FD&C blue #1 or D&C red #17, for example—make it easy to distinguish colors used in foods, drugs, or cosmetics from dyes made for textiles and other uses. Only FDA certified color additives can carry these special designations.

The law also created a listing of color "lakes." These water-insoluble forms of certain approved colors are used in coated tablets, cookie fillings, candies, and other products in which color bleeding could affect product quality or otherwise cause problems.

Although the 1938 law did much to bring color use under strict control, nagging questions lingered about tolerance levels for color additives. One incident in the 1950s, in which scores of children contracted diarrhea from Halloween candy and popcorn colored with large amounts of FD&C orange #1, led the FDA to retest food colors. As a result, in 1960, the 1938 law was amended to broaden the FDA's scope and to allow the agency to set limits on how much color could be safely added to food, drug, and cosmetic products.

The FDA also instituted a premarketing approval process that requires color producers to ensure, before marketing, that products are safe and properly labeled. Should safety questions arise later, colors can be reexamined. The 1960 measures put color additives already on the market into a "provisional" listing. This allowed continued use of the colors pending the FDA's conclusions on safety.

From the original 1960 catalog of about 200 provisionally listed colors, which included straight colors and lakes, only lakes of some colors remain on the

provisional list. Industry withdrew or the FDA banned many, while the rest became permanently listed and are still used. Some of these colors, derived from coal or petroleum sources, are subject to certification and carry the F, D, or C prefixes. Others, exempt from certification, are pigments and colors derived from plant, animal, and mineral sources. They are found in a myriad of food, drug, and cosmetic products.

Similar initiatives developed in most industrialized countries and regulations were implemented to control the use of color additives for use in foods and drugs. Although in most countries, food and drug colorants are strictly regulated, this is not the case in some regions in the Third World. Additionally, in some countries, cosmetic colorants are less regulated than the food and drug colorants. When using colorants from Third World countries, it is important to carefully assess the quality of these materials to ensure that they meet the requirements for the countries where the product will be marketed.

Worldwide Regulatory Considerations

Safety assessment

Ongoing toxicology studies are routinely being conducted worldwide by organizations such as the World Health Organization (WHO), the U.S. FDA, and the European Commission (EC) to assess the safety in various applications. Many articles have been written on assessing the safety of food colorants (1–6). Scientific and public opinion vary on the true interpretation of safety, and these views have impacted the development of food and drug regulations around the world.

Although these colors have a good safety record, consumers have voiced certain concerns about hyperactivity and allergic sensitivity regarding some of the synthetic colors.

Although this theory was popularized in the 1970s, well-controlled studies conducted since then have produced no evidence that food and drug color additives such as azo dyes (FD&C yellow #5, FD&C yellow #6, etc.) cause hyperactivity or learning disabilities in children. A Consensus Development Panel of the National Institutes of Health concluded in 1982 that there was no scientific evidence to support the claim that colorings or other food additives cause hyperactivity. The panel said that elimination diets should not be used universally to treat childhood hyperactivity, because there is no scientific evidence to predict which children may benefit. The FDA's Advisory Committee on Hypersensitivity to Food Constituents concluded in 1986 that FD&C yellow #5 (tartrazine) might cause hives in fewer than 1 out of 10,000

people. The Committee found that there was no evidence the color additive in foods provokes asthma attacks or that aspirin-intolerant individuals may have a cross-sensitivity to the color. As with other color additives certifiable for food and drug use, whenever FD&C yellow #5 is added to a food or drug, it is listed on the product label. This allows the small portion of people who may be sensitive to the color to avoid it.

During the early 1980s, before some of these reports were available, a regulation was included in Title 21 of the Code of Federal Regulation (CFR) regarding additional label requirements for FD&C yellow #5 concerning allergenicity. This regulation (201.20) required that for prescription drugs only, in addition to the label declaration showing that the product contains FD&C yellow #5, the label must also bear this warning statement: "This product contains FD&C yellow #5 (tartrazine) that may cause allergic-type reactions (including bronchial asthma) in certain susceptible persons. Although the overall incidence of FD&C yellow #5 (tartrazine) sensitivity in the general population is low, it is frequently seen in patients who also have aspirin hypersensitivity" (7).

When considering the very small potential for this type of a reaction, when compared with other products on the market with higher allergenic potential, and the conclusions of the FDA's Advisory Committee on Hypersensitivity to Food Constituents and the Consensus Development Panel of the National Institutes of Health, it appears that this statement may be overly cautious. However, this statement is still required for prescription drugs based on current U.S. regulations.

No similar type of warning statement is required for nonprescription drugs or foods. In reality, the potential exposure to FD&C yellow #5 is much greater from these products than from prescription drugs. Therefore, this regulation does not seem to make sense, given the current information. Although the industry has requested that this regulation be eliminated, the FDA has not taken any action on this as of this time. This has created much confusion concerning the actual data regarding hypersensitivity and allergenicity of FD&C yellow #5. Because of this confusion in the marketplace, most pharmaceutical companies have eliminated the use of this colorant in their prescription drug formulations to prevent the need for this misleading label requirement, which could create consumer concern.

To complicate matters more, there is also a European requirement being instituted to place allergen labeling on drug labels for a number of synthetic colors. Sound scientific data showing a need for this labeling is not apparent, and this requirement seems to be based more on nonscientific consumer concerns than on any real need. Considering the lack of evidence that FD&C yellow #5

(tartrazine) and other azo dyes such as FD&C yellow #6 (Sunset yellow FCF) cause any significant hypersensitivity or allergenicity problems, except in very rare cases, it appears that all that should be necessary is to make sure that the colorants are listed as a component on drug labels. These types of label requirements already apply in most industrialized countries and provide appropriate controls to allow those few people who are affected by these colors to manage their intake.

Nonharmonized regulations

Government regulations concerning the use of color additives change frequently, thereby making it difficult to be complete and accurate in listing colorants for use in international pharmaceutical development.

Legislation in various countries differs significantly regarding the color additives and their uses, although the principles of the laws are similar. These differences have led to certain colors being acceptable in one country and not being permitted in another. This creates a significant challenge to a pharmaceutical formulator when selecting colorants for use in global drug development.

In addition to assessing whether a particular colorant is approved for a specific use in certain countries, it is important to note that the specifications that each colorant must meet may also vary significantly from one region to another.

For instance, iron oxides are well known for being some of the most globally accepted colors. They are approved for use in drugs in most countries in the world. However, there are significant differences in the specifications for the various iron oxides listed in U.S. European, and Japanese regulations which mean that only certain iron oxides available in the marketplace can meet all of these requirements. In most cases, these grades must be subjected to many tests to provide assurance of compliance, which impacts the costs of these grades.

If a formulator is not extremely cautious in selecting an appropriate grade, the finished dosage form made from that grade may only be able to be marketed in certain countries, but not others. This would potentially mean development of multiple versions of the drug if the manufacturer determines that they would like to market the product in these other countries later on. The costs of this type of development can be very high and could potentially be eliminated by carefully selecting grades of colorants during early development of the drug that meet all of the requirements of the target market envisioned by the manufacturer. Of course, this requires that the pharmaceutical manufacturer identifies all the anticipated target markets for a particular drug product up front. Historically, many companies have not done this

until after the drug product has undergone many studies and, potentially, even regulatory reviews. Changing the colorant at this stage is very difficult.

Another interesting example that further illustrates this issue is the differences in regulations concerning D&C yellow #10 and Quinoline yellow between the United States and Europe. Although the Color Index (CI) numbers are the same, the dyes differ in composition as defined in the regulations. Quinoline yellow's purity criteria is defined in European Directive 94/45/EC as E104, which requires that the material contain not <80% of the disulfonated component of the dye and not >15% of the monosulfonated component (8). D&C yellow #10 is listed in the U.S. 21CFR (74.1710) which requires that the dye contain not <75% of the monosulfonated component and not >15% of the disulfonated component of the dye (9).

These specifications are mutually exclusive meaning that if a formulator desires the color obtained from these colorants, two versions of the drug product will be required if the product is to be marketed in both the United States and Europe.

Neither version of this dye is currently approved in Japan for use in drugs because of a lack of precedent. It would probably be possible to obtain approval in Japan if a manufacturer included one of these colorants in a new drug application and provided adequate safety data to convince Japanese regulators that the colorant is safe. However, in practice, most companies avoid using colorants that lack a precedent of previous use in Japan because this will usually delay drug approval.

Additional U.S. regulatory requirements— FDA certification

All colorants intended for use in the United States are regulated by the FDA as specified in the FD&C Act of 1938 and the 1960 Color Additives Amendment. The latter legislation defines a color additive as "any dye, pigment, or other substance made by a process of synthesis or similar artifice, or extracted, isolated, or otherwise derived, with or without intermediate or final change of identity, from a vegetable, animal, mineral or other source and that, when added or applied to a food, drug, or cosmetic or to the human body or any part thereof, is capable (alone or through reaction with another substance) of imparting a color thereto"(10).

The 1960 Color Additives Amendment also listed those color additives that need to be certified and those that are exempt from certification. The law establishes that color additives can be used under provisional listings until scientific investigations determine that they are suitable for permanent listing. See Tables 1 and 2 for a listing of permanently and provisionally listed colorants (11).

The certification procedure requires manufacturers of certified colorants to submit a sample of each batch of the color to the FDA for chemical analysis. If the results of the analysis comply with the specifications in the CFR, a certification lot number is assigned, and a certificate is issued for that batch. No color that requires certification can legally be used in the United States without being previously certified by the FDA. The certification lot number should be referenced in all documentation regarding the use of these colorants.

This procedure allows the FDA to be certain that all batches produced of a given color are chemically similar to the batches used in the original toxicology studies that were the basis for colorant approval.

TYPES OF COLORANTS APPROVED IN VARIOUS REGIONS

U.S. Certified Synthetic Colorants

The FD&C Act divided the synthetic colors into three categories: colors for foods, drugs, and cosmetics (FD&C), colors for drugs and cosmetics (D&C), and colors for externally applied drugs and cosmetics (external D&C). All synthetic colorants approved for use today must meet the specifications, uses, and restrictions as described in Title 21 of the CFR (Parts 74, 81, and 82). Certified synthetic colorants are the primary source of colorants used in the pharmaceutical industry.

FD&C colorants

The present list of FD&C certified colorants consists of both dyes and lakes. Lakes are pigments. They are insoluble materials that color by dispersion and reflected light. FD&C dyes are water-soluble and exhibit their color by transmitted light.

FD&C dyes: Today, FD&C dyes are synthetic organic molecules produced from highly purified intermediates derived from petrochemicals and other sources. The stigma of the association of FD&C dyes with coal tar no longer exists because of these starting materials. They are marketed in a number of physical forms, such as powder, granular, pastes, liquids, and dispersions. Many of these forms are customized for specific uses and are selected by the user for their particular application.

Dyes are relatively unstable because of their chemical structures. They are subject to instability as a result of: 1) light energy, 2) oxidizing and reducing agents, 3) microorganisms, 4) trace metals, 5) pH, and 6) high temperatures.

Table 1 List of permanently listed color additives subject to U.S. certification in 2000^a

Color	Common name	Color index number	21 CFR references			
			CAS number	Food	Drug	Medical devices
FD&C Blue #1	Brilliant Blue FCF	42090	3844-45-9	74.101	74.1101	74.2101
FD&C Blue #2	Indigotine	73015	860-22-0	74.102	74.1102	—
D&C Blue #4	Alphazurine FG	42090	6371-85-3	—	74.1104	74.2104
D&C Blue #6	Indigo	73000	482-89-3	—	—	74.3106
D&C Blue #9	Indanthrene Blue	69825	130-20-1	—	74.1109	—
D&C Brown #1	Resorcin Brown	20170	1320-07-6	—	—	74.2151
FD&C Green #3	Fast Green FCF	42053	2353-45-9	74.203	74.1203	74.2203
D&C Green #5	Alizarin cyanine green F	61570	4403-90-1	—	74.1205	74.2205
D&C Green #6	Quinizarine Green SS	61565	128-80-3	—	74.1206	74.2206
D&C Green #8	Pyranine concentrated	59040	63-58-69-6	—	74.1208	74.2208
Orange B	—	19235	—	74.250	—	—
D&C Orange #4	Orange II	15510	633-96-5	—	74.1254	74.2254
D&C Orange #5	Dibromofluorescein	45370:1	596-03-2	—	74.1255	74.2255
D&C Orange #10	Diiodofluorescein	45425:1	38577-97-8	—	74.1260	74.2260
D&C Orange #11	Erythrosine Yellowish Na	45425	38577-97-8	—	74.1261	74.2261
[Phthalocyaninato (2-)] Copper	Copper Phthalocyanine	74160	147-14-8	—	—	74.3045
FD&C Red #3	Erythrosine	45430	16423-68-0	74.303	74.1303	—
FD&C Red #4	Ponceau SX	14700	4548-53-2	—	74.1304	74.2304
D&C Red #6	Lithol Rubin B	15850	5858-81-1	—	74.1306	74.2306
D&C Red #7	Lithol Rubin B Ca	15850:1	5281-04-9	—	74.1307	74.2307
D&C Red #17	Toney Red	26100	85-86-9	—	74.1317	74.2317
D&C Red #21	Tetrabromo Fluorescein	45380:2	15086-94-9	—	74.1321	74.2321
D&C Red #22	Eosine	45380	17372-87-1	—	74.1322	74.2322
D&C Red #27	Tetrachlorotetra—Bromofluorescein	45410:1	13473-26-2	—	74.1327	74.2327
D&C Red #28	Phloxine B	45410	18472-87-2	—	74.1328	74.2328
D&C Red #30	Helindone Pink CN	73360	2379-74-0	—	74.1330	74.2330
D&C Red #31	Brilliant Lake Red R	15800:1	6371-76-2	—	74.1331	74.2331

(Continued)

Table 1 List of permanently listed color additives subject to U.S. certification in 2000^a (Continued)

Color	Common name	Color index number	CAS number	21 CFR references		Medical devices
				Food	Drug	
D&C Red #33	Acid Fuchsin	17200	3567-66-6	—	74.1333	74.2333
D&C Red #34	Lake Bordeaux B	15880:1	6417-83-0	—	74.1334	74.2334
D&C Red #36	Flaming Red	12085	2814-77-9	—	74.1336	74.2336
D&C Red #39	Alba Red	13058	6371-55-7	—	74.1339	—
FD&C Red #40	Allura Red AC	16035	25956-17-6	74.340	74.1340	74.2340
FD&C Red #40 Lake	Allura Red AC	16035:1	68583-95-9	74.340	74.1340	—
Citrus Red #2	—	12156	6358-53-8	74.302	—	—
D&C Violet #2	Alizarin Violet	60725	81-48-1	—	74.1602	74.2602
Ext. D&C Violet #2	Alizarin Violet	60730	4430-18-6	—	—	74.2602a
FD&C Yellow #5	Tartrazine	19140	1934-21-0	74.705	74.1705	—
FD&C Yellow #6	Sunset Yellow FCF	15985	2783-94-0	74.706	74.1706	—
D&C Yellow #7	Fluorescein	45350:1	2321-07-5	—	74.1707	74.2707
Ext. D&C Yellow #7	Naphthol Yellow S	10316	846-70-8	—	74.1707a	74.2707a
D&C Yellow #8	Uranine	45350	518-47-8	—	74.1708	74.2708
D&C Yellow #10	Quinoline Yellow WS	47005	8004-92-0	—	74.1710	74.2710
D&C Yellow #11	Quinoline Yellow SS	47000	8003-22-3	—	74.1711	74.2711

^aBased on 21 CFR 2000. Restrictions may exist limiting the use of some of these colors to specific applications (i.e., external drug use only, etc.). Additionally, there may be quantitative limits for the use of some colors. The specific 21 CFR reference for each color should be reviewed to determine potential restriction status.

Table 2 List of provisionally listed color additives subject to U.S. certification in 2000^a

Color	Common name	CI number	CAS number	21 CFR references		
				Food	Drug	Cosmetic
FD&C lakes	Lakes	See individual color	See individual color	82.51	82.51	82.51
D&C lakes	Lakes	See individual color	See individual color	—	82.1051	82.1051
Ext. D&C lakes	Lakes	See individual color	See individual color	—	82.1051	82.1051
FD&C blue #1 lake	Brilliant blue FCF	42090:2	53026-57-6	82.101	82.101	82.101
FD&C blue #2 lake	Indigotine	73015:1	16521-38-3	82.102	82.102	82.102
D&C blue #4 lake	Alphazurine FG	42090	6371-85-3	—	82.1104	82.1104
FD&C green #3 lake	Fast green FCF	42053	2353-45-9	82.203	82.203	82.203
D&C green #5 lake	Alizarin cyanine green F	61575	4403-90-1	—	82.1205	82.1205
D&C Green #6 Lake	Quinizarine Green SS	61565	128-80-3	—	82.1206	82.1206
D&C orange #4 lake	Orange II	15510:2	633-96-5	—	82.1254	82.1254
D&C orange #5 lake	Dibromo- fluorescein	45370:2	596-03-2	—	82.1255	82.1255
D&C orange #10 lake	Diiodofluorescein	45425:2	38577-97-8	—	82.1260	82.1260
D&C orange #11 lake	Erythrosine yellowish Na	45425:2	38577-97-8	—	82.1261	81.1261
FD&C red #4 lake	Ponceau SX	14700	4548-53-2	82.304	82.304	82.304
D&C red #6 lake	Lithol rubin B	15850:2	17852-98-1	—	82.1306	82.1306
D&C red #7 lake	Lithol rubin B Ca	15850:1	5281-04-9	—	82.1307	82.1307
D&C red #17 lake	Toney lake	26100	85-86-9	—	82.1317	82.1317
D&C red #21 lake	Tetrabromo- fluorescein	45380:3	15086-94-9	—	82.1321	82.1321
D&C red #22 lake	Eosine	45380:3	17372-87-1	—	82.1322	82.1322
D&C red #27 lake	Tetrachlorotetra—bromofluorescein	45410:2	13473-26-2	—	82.1327	82.1327
D&C red #28 lake	Phloxine B	45410:2	18472-87-02	—	82.1328	82.1328
D&C red #30 lake	Helindone pink CN	73360	2379-74-0	—	82.1330	82.1330
D&C red #31 lake	Brilliant lake red R	15800:1	6371-76-2	—	82.1331	82.1331
D&C red #33 lake	Acid fuchsin	17200	3567-66-6	—	82.1333	82.1333
D&C red #34 lake	Lake bordeaux B	15880:1	6417-83-0	—	82.1334	82.1334
D&C red #36	Flaming red	12085	2814-77-9	—	82.1336	82.1336
D&C violet #2 lake	Alizuroil purple SS	60725	81-48-1	—	82.1602	82.1602
FD&C yellow #5 lake	Tartrazine	19140:1	12225-21-7	82.705	82.705	82.705
FD&C yellow #6 lake	Sunset yellow FCF	15985:1	15790-07-5	82.706	82.706	82.706
D&C yellow #7 lake	Fluorescein	45350:1	2321-07-5	—	82.1707	82.1707
Ext. D&C Yellow #7 lake	Naphthol yellow S	10316	846-70-8	—	82.2707a	82.2707a
D&C yellow #8 lake	Uranine	45350	518-47-8	—	82.1708	82.1708
D&C yellow #10 lake	Quinoline yellow WS	47005:1	68814-04-0	—	82.1710	82.1710

^aBased on 21 CFR 2000. Restrictions may exist limiting the use of some of these colors to specific applications (i.e., external drug use only, etc.). Additionally, there may be quantitative limits for the use of some colors. The specific 21 CFR reference for each color should be reviewed to determine potential restriction status.

Table 3 List of color additives exempt from certification permitted for use in the U.S. in 2000^a

Color	Color index number	CAS number	21 CFR references			Medical devices
			Food	Drug	Cosmetic	
Algae Meal (Dried)	—	—	73.275	—	—	—
Alumina	77002	1332-73-6	—	73.1010	—	—
Aluminum Powder	77000	7429-90-5	—	73.1645	73.2645	—
Anatto Extract	75120	8015-67-6	73.30	73.1030	73.2030	—
Astaxanthin	—	—	73.35	—	—	—
β-APO-8-Carotenal	40820	1107-26-2	73.90	—	—	—
β-carotene	40800	7235-40-7	73.95	73.1095	73.2095	—
Beet Powder	—	57917-55-2	73.40	—	—	—
1, 4-BIS[2-(methylphenyl) Amino] 9,10-Anthracenedione	—	6737-68-4	—	—	—	73.3105
Bismuth Citrate	—	—	—	—	73.2110	—
Bismuth Oxchloride	77163	7787-59-9	—	73.1162	73.2162	—
Bronze Powder	77440	7440-50-8	—	73.1646	73.2646	—
Calcium Carbonate	77220	7740-66-6	—	73.1070	—	—
Canthaxanthin	40850	471-34-1	73.75	73.1075	—	—
Caramel	—	514-78-3	73.85	73.1085	73.2085	—
Carbazole Violet	51319	6358-30-1	—	—	—	73.3107
Carmine	75470	1390-65-4	73.100	73.1100	73.2087	—
Carrot Oil	—	—	73.300	—	—	—
Chlorophyllin Copper Complex	75810	—	—	—	—	73.3110
Chromium—Cobalt—Aluminum Oxide	77343	68187-11-1	—	73.1015	—	73.3110a
Chromium Hydroxide Green	77289	12182-82-0	—	73.1326	73.2326	—
Chromium Oxide Green	77288	1308-38-9	—	73.1327	73.2327	73.3111
C.I. Vat Orange 1	59105	—	—	—	—	73.3112
Cochineal Extract	75470	1260-17-9	73.100	73.1100	—	—
Corn Endosperm Oil	—	—	73.315	—	—	—
Copper Powder	77400	7440-50-6	—	73.1647	73.2647	—
1,4-Bis [(2-hydroxyethyl) amino]	—	10956-07-1	—	—	—	73.3100
-9,10-anthracenedione bis(2-propenoic) estercopolymers	—	—	—	—	—	—
1,4-Bis [4-(2-methylphenyl)amino]-9,10-anthracenedione	—	6737-68-4	—	—	—	73.3105
1,4-Bis[4-(2-methacryloxyethyl) phenylamino] anthraquinone Copolymers	—	121888-69-5	—	—	—	73.3106
2-[[2,5-Diethoxy-4-[(4-methylphenyl)thio]] phenyl]azo]-1,3,5-benzenetriol	—	—	—	—	—	73.3115

(Continued)

Table 3 (Continued)

Color	Color index number	CAS number	21 CFR references			Medical devices
			Food	Drug	Cosmetic	
16,23-Dihydrodinaphtho[2,3-a:2',3'-i]naphth[2',3':6,7]indolo[2, 3-c]carbazole-5,10,15,17,22,24-hexone	70800	2475-33-4	—	—	—	73.3117
<i>N,N'</i> -(9,10-Dihydro-9,10-dioxo-1,5-anthracenediyl) Bisbenzamide	61725	82-18-8	—	—	—	73.3118
7,16-Dichloro-6,15-dihydro-5,9,14,18-anthrazinetetrone	69825	130-20-1	—	—	—	73.3119
16,17-Dimethoxydinaphtho [1,2,3-cd:3',2',1'-lm] Perylene-5,10 dione	59825	128-58-5	—	—	—	73.3120
4-[2,4-(Dimethylphenyl)azo]-2,4-dihydro-5-methyl-2-phenyl-3H-pyrazol-3-one	—	6407-78-9	—	—	—	73.3122
Dihydroxy acetone	—	62147-49-3	—	73.1150	73.2150	—
Disodium Edta copper	—	—	—	—	73.2120	—
6-Ethoxy-2-(6-ethoxy-3-Oxobenzol[b]Thien 2(3H)-ylidene) Benzo[b]Thiophen-3(2H)-one	73335	3263-31-8	—	—	—	73.3123
Ferric Ammonium Citrate	—	1185-57-5	—	73.1025	—	—
Ferric Ammonium Ferrocyamide	77510	25869-00-5	—	73.1298	73.2298	—
Ferric Ferrocyamide	77510	14038-43-8	—	73.1299	73.2299	—
Ferrous Gluconate	—	299-29-6	73.160	—	—	—
Ferrous Lactate	—	5905-52-2	73.165	—	—	—
Fruit Juice	—	—	73.250	—	—	—
Grape Color Extract	—	—	73.169	—	—	—
Grape Skin Extract	—	—	73.170	—	—	—
Guaiaculene	—	489-84-9	—	—	73.2180	—
Guanine	75170	68-94-0, 73-40-5	—	—	73.2329	—
Henna	75480	83-72-7	—	73.1329	73.2190	—
Iron Oxides, Synthetic	77491 (Red) 77492 (Yellow) 77499 (Black)	1309-37-1 51274-00-1 12227-89-3	73.200	73.1200	73.2250	73.3125
Lead Acetate	—	6080-56-4	—	—	73.2396	—
Logwood Extract	75290	8005-33-2	—	73.1410	—	—
Manganese Violet	77742	10101-66-3	—	—	73.2775	—
Mica	77019	12001-26-2	—	73.1496	73.2496	—

(Continued)

Table 3 List of color additives exempt from certification permitted for use in the U.S. in 2000^a (Continued)

Color	Color index number	CAS number	21 CFR references		Medical devices
			Food	Drug	Cosmetic
Paprika	—	—	73.340	—	—
Paprika Oleoresin	—	8023-77-6	73.345	—	—
Potassium Sodium Copper Chlorophyllin	75180	—	—	73.1125	73.2125
Phthalocyanine Green	74260	1328-53-6	—	—	73.3124
Poly(Hydroxyethyl methacrylate)-Dye Copolymers	—	—	—	—	73.3121
Pyrogallol	76515	87-66-1	—	73.1375	—
Pyrophyllite	77004	8047-76-5	—	73.1400	—
Riboflavin	—	83-88-5	73.450	—	—
Saffron	75100	42553-65-1	73.500	—	—
		27876-94-4	—	—	—
Silver	77820	7440-22-4	—	—	75.2500
Tagetes Meal and Extract	75125	—	73.295	—	—
Talc	77019	14807-96-6	—	73.1550	—
Toasted Cotton Seed Meal	—	—	73.140	—	—
Titanium Dioxide	77891	13463-67-7	73.575	73.1575	73.3126
Turneric	75300	458-37-7	73.600	—	—
Turneric Oleoresin	75300	458-37-7	73.615	—	—
Ultramarine Blue	77007	57455-37-5	73.50	—	73.2725
Ultramarine Green	77013	—	—	—	73.2725
Ultramarine Pink	77007	12769-96-9	—	—	73.2725
Ultramarine Red	77007	12769-96-9	—	—	73.2725
Ultramarine Violet	77007	12769-96-9	—	—	73.2725
Vegetable Juice	—	—	73.260	—	—
Vinyl alcohol/Methyl Methacrylate-Dye Reaction Products	—	—	—	—	73.3127
Zinc Oxide	77947	1314-13-2	—	73.1991	73.2991

^aBased on 21 CFR 2000. Restrictions may exist limiting the use of some of these colors to specific applications (i.e., external drug use only, etc.). Additionally, there may be quantitative limits for the use of some colors. The specific 21 CFR reference for each color should be reviewed to determine potential restriction status.

The instability to these parameters and the solubility limits differ from color to color and should be considered when selecting dyes for use in various applications.

FD&C lakes: The only lakes permitted for use in all three categories—foods, drugs, and cosmetics—are the aluminum lakes. These are manufactured through the adsorption of an aluminum salt of an FD&C dye on a base of alumina hydrate.

The properties of these lakes can be controlled by variations in process conditions during manufacturing (for example, starting materials, order of additions, pH, and temperature). The most important attributes of aluminum lakes are the shade and particle size. The shade of the lake may be influenced by the quantity of dye adsorbed onto the alumina hydrate and the particle size distribution. The particle size also affects the tinting strength (that is, coloring power) of the pigment. Smaller particles result in increased surface area, which allows for an increase in reflected light and hence more color.

For additional information on the chemistry and properties of certified colors see Ref. (12).

D&C and external D&C colorants:

D&C and external D&C colorants may be used to color drugs and cosmetics with certain restrictions. A basic regulatory difference between FD&C, D&C, and external D&C colorants is that D&C and external D&C colorants have specific uses and restrictions. The classifications of D&C and external D&C mean little today, because many of the colorants listed as D&C are restricted to external uses.

D&C and external D&C colorants may also exist in either the dye or lake form; however, the majority of the commercially significant D&C and external D&C colorants are lakes.

D&C and external D&C dyes: The starting materials used in the manufacture of this class of colors are similar to those used for FD&C colors. D&C and external D&C dyes may or may not be soluble in water. Some are insoluble metal salts, and others are insoluble because they contain no water solubilizing groups. Several, however, are soluble in organic solvents.

Analogous stability problems exist between D&C, external D&C, and FD&C dyes, although there are a few D&C colors that are considerably more stable than the FD&C colorants.

D&C and external D&C lakes: These lakes are usually manufactured by precipitating a soluble dye onto an approved substrate. In the case of D&C colors, the substrate may be alumina, blanc fixe, gloss white, titanium dioxide, zinc oxide, talc, rosin, aluminum benzoate, calcium carbonate, or any combination of these materials.

A notable difference between FD&C and D&C lakes is that FD&C lakes must be manufactured using previously certified dyes, whereas D&C lakes are not restricted by this requirement. However, the regulations may change in the future requiring all lakes to be made from previously certified color.

The lakes made from D&C dyes are subject to the same uses and restrictions as the color from which they are derived.

The important physical properties of the FD&C lakes, such as particle size and shade, are equally important characteristics of the D&C and external D&C lakes.

Colorants Exempt from U.S. Certification

Certain colorants are not required by the FD&C Act to be certified by the FDA before use (Table 3). These include natural organic and inorganic colorants and certain synthetically produced so-called, nature-identical colors such as β -carotene. Although these colors are exempt from certification, they are still regulated by the FDA regarding their specifications, uses, and restrictions.

Most colors exempt from certification are obtained from animal, vegetable, or mineral sources. Most of the inorganic pigments are synthetically produced. Some examples would be titanium dioxide, synthetic iron oxide, and zinc oxide.

The inorganic colorants generally have good stability in a variety of applications. They tend to be less sensitive to heat and light than most of the other types of colorants used in pharmaceutical products. Titanium dioxide is a widely used white pigment. The other inorganic pigments are mainly earth tones.

Many of the natural organic colors in this category are extracts of a specific plant or insect. These extracts are then processed into a usable form. Because of the different sources of these materials, the properties vary greatly for each colorant and also exhibit batch-to-batch variation. Because of the types of raw materials used and the difficulties in manufacturing, these materials tend to be very costly.

Most of the natural organic colorants exhibit poor stability. They are very sensitive to heat, light, and environment. At present, their use is limited because of difficulties in incorporating them into specific products. In general, they exhibit far less tinting strength and lack the range of colors and color reproducibility that can be obtained when using synthetic colorants. In some cases, they may also flavor the product being colored.

Many schemes have been contrived to improve the stability of the natural organic colors. There are also many novel approaches for incorporating these colors into

Table 4 Approved drug colorants listed by the European union in 2000^a

Color	EEC number ^b	Color index number	Alternate names
Allura Red AC	E129	16035	FD&C Red #40
Aluminum	E173	77000	—
Amaranth ^c	E123	16185	Delisted FD&C Red #2
Annatto	E160b	75120	Bixin, Norbixin
Anthocyanins	E163	—	—
Beet Root Red	E162	—	Betainin
β-APO-8'-Carotenal	E160e	40820	—
β-APO-8'-Carotenoid Acidethyl Ester	E160f	40825	—
Brilliant Black BN	E151	28440	Black PN
Brilliant Blue FCF	E133	42090	FD&C Blue #1
Brown FK	E154	—	—
Brown HT	E155	20285	—
Calcium Carbonate	E170	77220	—
Canthaxanthin ^c	E161g	40850	—
Caramel	E150a	—	—
Caramel,-Caustic Sulfite	E150b	—	—
Caramel,-Ammonia	E150c	—	—
Caramel, Sulfite Ammonia	E150d	—	—
Carbon Vegetable Black	E153	77268:1	Carbo Medicinalis Vegetalis
Carmin	E120	75470	Carmin 40, Cochineal Lake, Carminic Acid Lake
Carmoisine	E122	14720	Azorubine
Carotene	E160a	75130	α-, β-, and γ-Carotene
Mixed Carotenes	—	75130	—
β-Carotene	—	40800	—
Chlorophylls/Chlorophyllins	E140	—	—
Chlorophylls	—	75810	—
Chlorophyllins	—	75815	—
Chlorophylls/Chlorophyllins	E141	75815	—
Copper Complexes	—	—	—
Copper Complexes of Chlorophylls	—	—	—
Copper Complexes of Chlorophyllins	—	—	—

(Continued)

Table 4 (Continued)

Color	EEC number ^b	Color index number	Alternate names
Cochineal	E120	75470	Carminic Acid
Erythrosine ^c	E127	45430	FD&C Red #3
Gold	E175	77480	—
Green S	E142	44090	Acid Brilliant Green BS
Indigotine	E132	73015	FD&C Blue #2, Indigo Carmine
Iron Oxides and Hydroxides	E172	77491	Iron Oxide Red
		77492	Iron Oxide Yellow
		77499	Iron Oxide Black
Lutein	E161b	—	—
Lycopene	E160d	—	—
Paprika Extract	E160c	—	Capsanthin, Capsorubin
Patent Blue V	E131	42051	Acid Blue 3
Ponceau 4R	E124	16255	Cochineal Red A
Quinoline Yellow ^d	E104	47005	China Yellow
Riboflavin	E101	—	—
Riboflavin	—	—	—
Riboflavin-5'-phosphate	—	—	—
Sunset Yellow FCF	E110	15985	FD&C Yellow #6, Orange Yellow S
Tartrazine	E102	19140	FD&C Yellow #5
Titanium Dioxide	E171	77891	—
Turmeric	E100	75300	Curcumin

Note: Aluminum lakes prepared from colors mentioned in this list are also authorized.

^aThese colors are approved for specific types of drug use based on an EU legal opinion that cross-references the list of approved drug colors to those listed in EC Directive 94/36/EC that regulates food colors (15). Restrictions may apply to some of the colors concerning the types of applications where they can be used and the maximum amounts which can be present. A new EC Directive, which specifically covers drug colors, is expected in the future to help clarify the situation.

^bSome colors and their derivatives share the same EEC Number. The number and lowercase letter identify the derivatives.

^cThese colors are currently listed for use in previously approved drugs. However, these colors should be avoided in new drug applications.

^dThis is not D&C yellow #10. Although the C.I. numbers are the same, the dyes differ in composition. Quinoline Yellow is primarily the disulfonated quinoline dye, whereas D&C Yellow #10 is the monosulfonated color. Quinoline Yellow is not accepted for use in the U. S.; conversely, D&C yellow #10 cannot be used in the EU.

Table 5 Approved drug colorants for use in Canada in 2000^a

Color	Alternate name	Color index number	CAS number
A. Colorants approved for internal and external drug use			
Acid Fuchsin D	D&C Red #33	17200	3567-66-6
Alizarin Cyanine Green F	D&C Green #5	61570	4403-90-1
Allura Red AC	FD&C Red #40	16035	25956-17-6
Amaranth	Delisted FD&C Red #2	16185	915-67-3
β-APO-8' Carotenal	—	40820	1107-26-2
Brilliant Blue FCF Sodium Salt	FD&C Blue #1	42090	3844-45-9
Brilliant Blue FCF Ammonium Salt	D&C Blue #4	42090	6371-85-3
Canthaxanthin	—	40850	514-78-3
Caramel	—	—	—
Carbon Black	—	77266	1333-86-4
Carmine	—	75470	1260-17-9
Carmoisine	Azorubine	14720	3567-69-9
β-carotene	—	40800	7235-40-7
Chlorophyll	—	75810	479-61-8
Eosin YS Acid Form	D&C Red #21	45380:2	15086-94-9
Eosin YS Sodium Salt	D&C Red #22	45380	17372-87-1
Erythrosine	FD&C Red #3	45430	16423-68-0
Fast Green FCF	FD&C Green #3	42053	2353-45-9
Flaming Red	D&C Red #36	12085	2814-77-9
Helindone Pink CN	D&C Red #30	73360	2379-74-0
Indigo	D&C Blue #6	73000	482-89-3
Indigotine	FD&C Blue #2	73015	860-22-0
Iron Oxides	Iron Oxide Red	77491	1309-37-1
	Iron Oxide Yellow	77492	51274-00-1
	Iron Oxide Black	77499	12227-89-3
Lithol Rubin B Sodium Salt	D&C Red #6	15850	5858-81-1
Lithol Rubin B Calcium Salt	D&C Red #7	15850:1	5281-04-9
Phloxine B Sodium Salt	D&C Red #28	45410	18472-87-2
Phloxine O Acid Form	D&C Red #27	45410:1	13473-26-2
Ponceau 4R	—	16255	2611-82-7
Ponceau SX	FD&C Red #4	14700	4548-53-2
Quinoline Yellow WS	D&C Yellow #10	47005	8004-92-0
Riboflavin	—	—	83-88-5
Sunset Yellow FCF	FD&C Yellow #6	15985	2783-94-0
Tartrazine	FD&C Yellow #5	19140	1934-21-0
Titanium Dioxide	—	77891	13463-67-7
B.	Preparations made by extending any of the above colors on a substratum of alumina, blanc fixe, gloss white, clay, zinc oxide, talc, rosin, aluminum benzoate, calcium carbonate, or any combination of these listed substances.		
C.	Preparations made by extending any sodium, potassium, aluminum, barium, calcium, strontium, or zirconium salt of any of the colors listed in paragraph A on a substratum of alumina, blanc fixe, gloss white, clay, zinc oxide, talc, rosin, aluminum benzoate, calcium carbonate, or any combination of these listed substances.		
Color	Alternate name	Color index number	CAS number
A. Colorants approved for external drug use			
Alizuro Purple SS	D&C Violet #2	60725	81-48-1
Annatto	—	75120	—
Bismuth Oxychloride	—	77163	—
Chromium Hydroxide Green	Pigment Green 18	77289	—
Dibromofluorescein (Solvent Red 72)	D&C Orange #5	45370:1	—

(Continued)

Table 5 (Continued)

	Color	Alternate name	Color index number	CAS number
	Deep Maroon	D&C Red #34	15880:1	6417-83-0
	Guanine	—	75170	—
	Orange II	D&C Orange #4	15510	633-96-5
	Manganese Violet	—	77742	—
	Mica	—	77019	—
	Pyranine Concentrated	D&C Green #8	59040	6358-69-6
	Quinizarin Green SS	D&C Green #6	61565	128-80-3
	Toney red	D&C Red #17	26100	85-86-9
	Uranine Acid Form	D&C Yellow #7	45350:1	2321-07-5
	Uranine Sodium Salt	D&C Yellow #8	45350	518-47-8
	Zinc Oxide	—	77947	—
B.	Preparations made by extending any of the above colors on a substratum of alumina, blanc fixe, gloss white, clay, zinc oxide, talc, rosin, aluminum benzoate, calcium carbonate, or any combination of these listed substances.			
C.	Preparations made by extending any sodium, potassium, aluminum, barium, calcium, strontium, or zirconium salt of any of the colors listed in paragraph A on a substratum of alumina, blanc fixe, gloss white, clay, zinc oxide, talc, rosin, aluminum benzoate, calcium carbonate, or any combination of these listed substances.			

^aBased on the Canadian Department of Health's Food and Drug Regulations on Coloring Agents; Part C.01.040.2; Aug 30, 1995 (16).

products and processes. However, they have met with little success. Although the use of natural organic colors may be desired for a number of reasons, in a majority of these cases it proves impractical.

Much research is currently underway to support the use of additional colors exempt from certification. Color Additive Petitions are being submitted and reviewed by the FDA for a number of new colors and extended uses of existing colors. During the next several years, it is expected that many of these colorants may be approved for these uses in the U.S. pharmaceutical industry.

Major Colorants Used in Multinational Pharmaceutical Applications

As mentioned earlier, when using colorants in pharmaceutical products to be produced or marketed in countries other than the United States, it is extremely important that the target market be established during the product development stage, because the colorant regulations vary significantly from country to country. Various official references are available that can be used to assess the regulations and requirements regarding approved colorants in specific countries (4, 5, 11, 5, 13–17). Additionally, some color usage guides are available from several color manufacturers, which can be helpful.

Most European countries follow the European Directives that list the colorants and specifications for use in foods and drugs in the European Union (EU). The directive that has previously controlled the approved colorants for

use in pharmaceuticals in Europe is 78/25/EEC, which refers to a list of colorants from a 1962 directive (13, 14). The EC published European Directive 94/36/EC in 1994, which significantly changed the list of approved colorants for use in foods. For example, Allura red AC (FD&C red #40) and Brilliant blue FCF (FD&C blue #1) were now approved for use in foods, however, these materials did not exist on the list of approved drug colorants because of the cross-referencing of the pharmaceutical directives back to the 1962 list. This has created much confusion throughout industry and the regulatory community.

Recent discussions within the EU have indicated that the colorant list from 94/36/EC will be rationalized into a new pharmaceutical color directive in the future so that the colorant lists will be aligned. It is now considered acceptable to use these colors in new drug applications even though the formal directives have not yet been finalized. However, each country may have some specific regulations for the use of these colorants in pharmaceuticals that must also be considered. Some nonmember European countries, such as some Scandinavian countries, severely restrict the use of synthetic colors and in some cases oppose their use entirely.

The purity criteria specifications for colorants approved for use in Europe are laid down in European Directive 95/45/EC (8). Each colorant defined by an EEC number has a monograph listing the required tests and limits. The regulation does not specify the test methods to be used for these tests; however, the Joint Expert Committee on Food Additives (JECFA) methods are generally acceptable. Any

Table 6 Approved drug colorants for internal use in Japan in 2000^a

Color	Alternate name	Color index number	CAS number
Acid Red	Acid Red #52	45100	—
Amaranth ^{b, c}	Delisted FD&C Red #2	16185	915-67-3
Brilliant Blue FCF Sodium Salt ^b	FD&C Blue #1	42090	3844-45-9
Caramel	—	—	—
Carmine	—	75470	1260-17-9
β -Carotene ^{R1}	—	40800	7235-40-7
Copper Chlorophyll ^d	—	—	—
Erythrosine ^b	FD&C Red #3	45430	16423-68-0
Fast Green FCF ^b	FD&C Green #3	42053	2353-45-9
Indigotine (Indigo Carmine) ^b	FD&C Blue #2	73015	860-22-0
Iron Oxide Red ^e	Red Ferric Oxide JPE	77491	1309-37-1
Iron Oxide Yellow ^f	Yellow Ferric Oxide JPE	77492	51274-00-1
Iron Oxide Black ^g	Black Iron Oxide JPE	77499	12227-89-3
Phloxine B Sodium Salt ^b	D&C Red #28	45410	18472-87-2
Ponceau 4R ^b	—	16255	2611-82-7
Riboflavin	—	—	83-88-5
Riboflavin Butyrate ^h	—	—	—
Riboflavin Sodium Phosphate ^h	—	—	—
Rose Bengal	Acid Red #94	45440	—
Sodium Copper Chlorophyllin ^j	—	—	—
Sunset Yellow FCF ^b	FD&C Yellow #6	15985	2783-94-0
Tartrazine ^{b, c}	FD&C Yellow #5	19140	1934-21-0
Titanium Dioxide	—	77891	13463-67-7
Turmeric Extract	Curcumin	75300	458-37-7
Vegetable Carbon Black ^k	—	77268:1	1333-86-4

^aBased on colors approved by the MHW's "Ministerial Ordinance to establish Tar colors which can be used in Pharmaceuticals"; No. 30; August 31, 1966. Aluminum lakes of these colors are also authorized.

^bNot more than 0.1% by weight of color (lake or dye) can be used in a dosage form (BIRYO limit on trace amounts).

Note: ^{d, e, f, g, h, i, and j}—Precedent limit listed in the Japanese Pharmaceutical Excipients Directory (JPED) 2000, Japanese Version (18). This limit represents the maximum daily dosage that a patient should consume from the use of a particular dosage form.

^cColor is permitted but use is generally discouraged for new drug applications.

^dNot more than 1.8 mg/day.

^eNot more than 95.4 mg/day.

^fNot more than 5.67 mg/day.

^gNot more than 1.539 mg/day.

^hNot more than 0.4 mg/day.

ⁱNot more than 2.0 mg/day.

^jNot more than 75 mg/day.

^kMust comply with Medicinal Carbon JP specifications if used for manufacture in Japan.

colorant that is to be used in Europe must meet these specifications, which are somewhat different than those used in other countries.

See Table 4 for a list of commonly used European colorants. This list is based on the colorants approved in 94/36/EC (15).

Regulations for colorant use in Canada have many similarities to the U.S. regulations; however, differences do exist. Some colors approved in Canada are not approved in the United States and vice versa. Canada's color regulations for drug color generally list most colors

acceptable for use in Europe as well as those listed for use in the United States. However, some differences do exist. See Table 5 for colorants that can be used in Canadian drug applications (16).

In Japan, pharmaceutical colorants are regulated by the MHWs "Ministerial Ordinance to Establish Tar Colors which can be Used in Pharmaceuticals": No. 30, August 31, 1966 (17). This regulation provides a list of approved synthetic colorants and their specifications. In 1993, MHW reviewed these regulations, and proposals were made to update the regulation. However, no action has been

officially taken regarding these proposals at this time. Therefore, the 1966 regulation is still in effect. Quantity restrictions exist for most synthetic colorants in Japan based on a requirement that these colorants only be present in trace quantities. These trace requirements in Japan are commonly referred to as BIRYO limits. The BIRYO limit for the use of synthetic colorants in oral drug applications is not $>0.1\%$ by weight of color (lake or dye) in the dosage form.

Additional nonsynthetic colors may be acceptable for pharmaceutical applications based on previous precedence of use. These colors are typically listed in the Japanese Pharmaceutical Excipients (JPE) Directory along with their Japanese Pharmacopoeia (JP) or JPE specification (18).

See Table 6 for a list of commonly used pharmaceutical colorants in Japan. This list does not contain all of the acceptable colorants that may be used, but does highlight the major colorants with significant commercial implications along with some of the quantity restriction information.

Most countries have their own specific set of regulations concerning food and drug colorants based on local systems of additive approval. Because of these local evaluations, these regulations vary quite significantly and must be assessed when selecting colorants for target markets that include these countries.

Some countries throughout the world use colorants and specifications recommended by the JECFA, which is organized through the Food and Agricultural Organization (FAO) and the WHO (4, 5). In many cases, approved drug colors are considered by these countries to be the same as the approved food colors.

COLORING SYSTEMS FOR VARIOUS DOSAGE FORMS

Once the colorants have been identified that are approved for use in the target market, other performance-oriented criteria should be assessed. In selecting a colorant for a given application, prime consideration should be given to the type of formulation in which the colorant is to be incorporated.

Whatever the form of colorant chosen, it should meet certain criteria, namely:

- It must be nonreactive;
- It should be the most stable form for the application;
- It should be easily incorporated into the system in which it is being used;
- It should meet the aesthetic criteria required;
- It should not impart any offensive odor or taste to the product.

Tablets

Sugar-coating

Sugar-coating is a process involving several steps, and success is measured in terms of the elegance of the final product. To achieve this success, highly skilled manpower must be used, and care must be taken throughout the process. In the past, because of these factors, sugar-coating was a very time-consuming part of the manufacture of coated tablet dosage forms. The sugar-coating process is divided into several operations. These include sealing, subcoating, smoothing, coloring, and polishing.

For the purposes of this article, only the coloring stage will be discussed. This is probably one of the most critical parts of the operation. It gives the tablet its color and, in some cases, its finished size.

Before the 1950s, traditional color coating for solid dosage forms was usually performed using soluble dyes as the prime colorant. This system can produce the most elegant tablet. However, many difficulties can arise during the processing. These usually relate to the dye being soluble. Color migration readily occurs if the drying stage after each application of color is not handled properly. This results in nonuniform color or mottling. Small depressions or irregularities in the surface of the tablet may also cause nonuniform color. This results from the concentration of dye in these imperfections. Many smoothing coats are needed before any color can be applied.

Careful control must be exercised during the coating process to achieve a good color match. Care also must be taken to ensure that the tablets do not become overcolored. Syrups of increasing dye concentrations usually are used to achieve a color match and to control mottling. This operation may take from 20 to 60 applications for the color to develop fully.

All of the aforementioned problems combine to make color reproducibility from batch to batch difficult to achieve. It becomes obvious that the services of a skilled coater are needed to obtain the best possible results. As can be seen, dye sugar-coating is a very time-consuming and delicate operation.

Late in the 1950s, the pigment sugar-coating process was developed and subsequently patented by Arnold Nicholson and Stanley Tucker (19). The coloring composition of this invention was essentially an aluminum lake and an opacifier dispersed in a syrup solution. This system produced brightly colored, elegant tablets and eliminated many of the problems associated with the standard sugar-coating technique.

The use of insoluble certified lakes have several advantages; namely:

- The fact that they are insoluble enables the drying stages to be performed more quickly, because color migration is unlikely.
- Mottling is reduced because the opacity of the system minimizes the effect of tablet surface depressions.
- Overcoloring is not a problem because the system is opaque; hence, only one shade of color will result.
- Full color development can be achieved with a fewer number of application stages. This results in significant time savings, yet produces a final product that has all the properties of the standard sugar-coated tablet plus increased light stability. Raw material costs are also improved.
- Many of the problems associated with color reproducibility have been eliminated entirely. The finished tablet color has been predetermined by color matching the pigment dispersion before coating.

All of these advantages have resulted in pigment systems virtually replacing the soluble dye sugar-coating process.

Effective use of these pigments can be best made if they are adequately dispersed in the syrup. Obviously, it is more difficult to disperse pigment than it is to dissolve dye. The dusty nature of pigments sometimes requires the use of air filtration and dust collection systems to avoid contamination of other areas of the plant.

Today, there are a number of manufacturers who offer color-matched, predispersed, pigment sugar-coating concentrates. These concentrates are easily incorporated into the bulk of the syrup-coating solution. The convenience of these concentrates and the ability of the manufacturers to reproduce color batch after batch make these products an attractive alternative to self-preparation of dispersions.

Film-coating

Film-coating was developed early in the 1950s to help resolve many of the problems associated with sugar-coating. It involves the application of a film-forming polymer onto the surface of a substrate (such as tablets, granules, and capsules). In addition to the polymer, the film contains plasticizers and colorants, which are needed to achieve the desired properties in the final dosage form.

The polymer and the plasticizer are usually dissolved in a solvent to form a coating solution in which the colorants can be dissolved or dispersed. The original film-coating systems used organic solvents for polymer solution. Today, aqueous systems have largely replaced the organic solvents for environmental reasons.

When using organic solvents, water-soluble dyes are not used because of solubility constraints. In aqueous film-coating systems, water-soluble dyes can be used as colorants. However, many of the same problems observed

in sugar-coating may exist relating to color migration on drying of the films. Additionally, because film coatings are relatively thin, small differences in film thickness on tablets may result in significant color variations. There has been some success in using opacified dye systems; however, these systems have been shown to have poorer light stability than pigmented coatings.

The colorants of choice for these applications are lakes and inorganic pigments. In addition to providing color, pigments have been reported to reduce moisture diffusion through the film and improve light stability as compared with dye (20).

Again, it is important that these pigments be well dispersed in the system to maximize color and provide an elegant appearance. Because dispersing these pigments into the coating system is a difficult task, the use of commercially available predispersed color-matched pigment concentrates are recommended. These concentrates are available both in liquid and in dry forms. The dry forms may contain all of the components needed for a total film-coating system that can be dispersed in water or other solvents directly at the coating pan.

Colored wet granulations

Dissolving water-soluble dyes in a binding solution for the granulating process is the most common approach to coloring a tablet formulation. However, during drying of the granulation, the soluble colors may migrate, and if more than one color is used, the dyes may migrate at different rates. This results in an uneven coloring of the granulation, which will have a mottled appearance after compression. Some additives, such as starches, clays, and talc, have been used to adsorb the dye, thereby reducing but not completely eliminating the migration.

This entire problem can be avoided by using lakes or other pigments. The colors will not migrate because they are insoluble. In addition, the light stability of the product will be improved.

Direct compression

A growing interest in direct compression formulas has developed, mainly for economic reasons. The number of processing steps has been reduced, and the availability of many direct compression materials allows for a simplified tablet formulation. Direct compression formulations require blending only; therefore, lakes and other pigments are used because the elimination of the wetting step prevents the effective use of soluble colors.

The dry-coloring of tablets with pigments is not without problems. Although there is little chance of color migration, poor blending of the pigments into the powder can result in color specking and "hot spots." This problem

can be minimized by preblending the pigment with a small part of one of the other ingredients before addition to the entire mixture to reduce pigment particle agglomeration. The ease with which the pigment can be incorporated into the formulation may depend on the components in the mixture. In extreme cases, some type of milling may be required to achieve an adequate dispersion.

Hard Gelatin Capsules

Hard gelatin capsules are manufactured by melting gelatin, coloring the gelatin, and casting onto pins to form half capsules. They are then dried, stripped off, trimmed, and mated. The capsules are colored primarily using FD&C or D&C colorants and sometimes an opacifying agent such as titanium dioxide.

The clear type of capsule is colored using water-soluble dyes. Solutions of these colors are simply added to the gelatin melt. The pH of the gelatin is important because it can alter the shade of the color. It is also important to control the thickness of the capsule wall because variations can change color intensity. If the active ingredient is photosensitive, it is advisable to use an opaque capsule. Opaque capsules can contain pigments or dyes and an opacifier. The colorants are usually dissolved or dispersed in water, glycerine, or combinations of these vehicles before addition to the gelatin mixture. Wall thickness is rarely a factor in determining the shade of an opaque capsule.

Recent technological advances in the area of spin printing have allowed some manufacturers to color-identify capsules by printing bands of varying widths and colors on the capsule bodies through the use of colored imprinting inks. The inks are generally colored using pigments because they offer greater light stability. Dyes can be used, however, if a transparent color band is desired.

Liquid Products

The appearance of clear liquid products depends primarily on the color and clarity of the solution. Dyes should be used that are completely soluble in the particular solvent and at the required concentration. Many times dyes that correspond to the flavor of the product (for example, red for cherry or yellow for lemon) will be chosen.

Factors influencing the shade and stability of dyes in the liquid system must be carefully considered as well. Among these properties are pH, microbiologic activity, light exposure in the final product package, and the compatibility of the dye with other ingredients. The color shade of most dyes varies greatly at different pHs, making control of product pH extremely important.

All soluble dyes contain reactive sites, and some may be incompatible with drug substances in the formulation.

The anionic colors can also react with compounds containing polyvalent cations (such as calcium, magnesium, or aluminum) and precipitate. Certain dyes, such as FD&C blue #2 and FD&C red #3, exhibit such poor stability in aqueous solutions that they should never be used for coloring aqueous liquid products.

When formulating liquid products with dyes, the lowest possible concentration of dye needed to give the desired color should be used, because higher concentrations can result in a dull color. Most liquid products have dye concentrations of <0.001%. The actual concentration depends on how the product will normally be viewed and the depth of color required.

Because dyes are usually present in small concentrations, they should be predissolved before mixing with the bulk of the formulation. This will ensure that all the dye has been dissolved before additional processing. If the dye was added directly to the bulk mixing tank, the presence of small amounts of undissolved material would be difficult to determine and could cause additional problems later during the compounding procedure.

Pigments or dyes can be used for coloring opaque liquids such as suspensions, emulsions, or imprinting inks. In aqueous systems, pigments offer improved light stability and are less affected by pH and vehicle compatibility. This results in increased shelf life. In nonaqueous systems, because of solubility restrictions, the use of pigments is necessary. The liquid media may cause certain organic pigments to bleed. The extent of bleed will depend on the solvent composition and pH.

If pigments are chosen as the colorants, it may be necessary to predisperse them before adding them to the final product. Concentrated dispersions in a wide variety of vehicles are commercially available.

Ointments and Salves

Both dyes and pigments can be used for coloring ointments and salves, depending on the vehicle. Pigments are preferred because they will not migrate to the surface. To incorporate the pigments into the system, it may be necessary to blend the pigment and the product on a roll or ointment mill.

COLORANT PROPERTIES AND APPLICATIONS

Having discussed the different methods of coloring pharmaceutical products and the various forms of colorants available, the next section can be used to assist in product formulation.

Table 7 Approximate solubilities of some FD&C colors at 25°C in grams per 100 ml of solvent^a

Color	Distilled water	Glycerine	Propylene glycol	95% Ethanol	50% Ethanol
FD&C Blue #1	18	20	20	1.5	20
FD&C Blue #2	1.5	1	0.1	Trace	0.2
FD&C Green #3	17	15	15	0.2	7
FD&C Red #3	12	22	22	2	4
FD&C Red #40	20	3	1.5	Trace	1
FD&C Yellow #5	15	18	8	Trace	4
FD&C Yellow #6	18	15	2	Trace	2

^aThe solubility of commercial lots of FD&C colors differs widely depending on the amounts of salt, pure dye, moisture, and subsidiary dyes present.

Soluble Dyes

If the choice of colorant is a soluble dye, many are offered in a powder, granular, or liquid form. The liquid form offers ease of handling, is dust-free, and is ready to use. However, it is probably the most expensive form of coloring and is subject to both light and microbial stability problems. The powder form is least expensive but poses potential dust problems and can result in plant contamination. The granular product is recommended because dust problems are reduced, although, in some cases, it may have a slower rate of dissolution than the powder form, depending on the color.

Tables 7–9 show typical solubility and stability properties of some of the dyes.

Lake Pigments

FD&C lakes are available in dye strengths of 3–60%. The actual shade and tint will vary for each color and dye strength. There also may be slight differences from one manufacturer to another.

Properties that have made lakes more suitable (for coloring dosage forms) are their relative opacity, stability with regard to heat and light, and ability to be used dry when coloring products (such as tablets) made by direct

compression. They are the only choice for coloring fats and oils, because there are no oil-soluble colorants approved for this application.

Some of the physical characteristics listed for lakes in the literature include particle size distribution, moisture content, oil absorption, bulk density, specific gravity, and pH stability. These properties may or may not be of interest, depending on the user's application.

The particle size distribution determines the tinting properties of a particular pigment. A good distribution of particle size contributes to the relative ease of dispersion of the pigment into a given system.

If the lakes used are aluminum lakes, the alumina substrate contains water of hydration, and the lake itself may have total moisture content of between 12 and 25%. High processing temperatures can release some of the hydrated water; however, in most pharmaceutical applications this is not a problem.

Oil absorption is a measure of the amount of oil necessary to wet out a pigment particle such that each particle becomes enclosed in a shell of oil. The amount of oil required depends on the surface area of the pigment and the porosity of the particle. This property probably has more application in the paint industry than in coloring pharmaceuticals. It may be of some limited value in producing ointments and salves.

Table 8 FD&C color stability data

Dye	Heat	Light	Acid	Base	Oxidizing agents	Reducing agents
FD&C Blue #1	Good	Moderate	Very good	Moderate	Moderate	Poor
FD&C Blue #2	Good	Very poor	Moderate	Poor	Poor	Good
FD&C Green #3	Good	Fair	Good	Poor	Poor	Very poor
FD&C Red #3	Good	Poor	Insoluble	Good	Fair	Very poor
FD&C Red #40	Good	Moderate	Good	Moderate	Fair	Fair
FD&C Yellow #5	Good	Good	Good	Moderate	Fair	Fair
FD&C Yellow #6	Good	Moderate	Good	Moderate	Fair	Fair

Table 9 Solubilities and fastness properties of some D&C colors^a

	Solubilities			Stability				
	Water	Glycerol	EtOH	Light	Acid	Alkalis	Reducing agents	Oxidizing agents
D&C Blue #4	S	S	S	F	G	F	P	P
D&C Blue #9	I	I	I	G	G	G	P	G
D&C Green #5	S	S	SS	G	G	G	P	F
D&C Green #6	I	I	SS	G	G	G	P	F
D&C Green #8	S	SS	SS	P		G	F	F
D&C Orange #4	S	S	SS	G	G	P	F	F
D&C Orange #5	I	SS	S	P	F	G	F	F
D&C Orange #10	I	SS	S	P	F	G	F	F
D&C Orange #11	S	S	SS	P	F	G	F	F
D&C Red #6	S	S	I	G	F	F	P	F
D&C Red #7	I	I	I	G	F	G	P	F
D&C Red #17	I	SS	SS	F	F	G	P	F
D&C Red #21	I	I	SS	P	F	G	F	F
D&C Red #22	S	S	SS	P	P	G	F	F
D&C Red #27	I	I	SS	P	F	G	F	F
D&C Red #28	S	S	S	F	F	G	F	F
D&C Red #30	I	I	I	G	G	G	P	G
D&C Red #31	S	SS	SS	G	F	G	P	F
D&C Red #33	S	S	SS	G	F	G	P	F
D&C Red #34	I	I	I	F	F	F	P	F
D&C Red #36	I	I	I	G	F	F	P	F
D&C Red #39	I	S	S	P	P	G	F	F
D&C Violet #2	I	I	SS	F	G	G	P	P
Ext. D&C Violet #2	S	S	SS	G	G	G	P	F
D&C Yellow #7	I	SS	SS	P	G	G	F	F
Ext. D&C Yellow #7	S	S	SS	F	G	G	F	F
D&C Yellow #8	S	S	S	F	F	G	F	F
D&C Yellow #10	S	S	SS	F	G	F	G	P
D&C Yellow #11	I	SS	S	P	G	P	G	P

^aI—insoluble; SS—slightly soluble; F—fair; G—good; P—poor.

The bulk density is the mass of a unit volume of loose dry pigment, including voids. Sometimes this is called the apparent density. Bulk density may vary from one manufacturer to another, depending on the particular milling process.

Specific gravity is defined as the weight of pigment per unit volume relative to a standard, usually water. This property is important, because it can sometimes have an effect on dispersion stability. Pigments that are heavier than the vehicle in which they are dispersed are prone to settling.

Most aluminum lakes are within a pH range of 4.0–8.0. Outside this range, the dye and substrate may be separated and the lake dissolved. Other lakes may exhibit some stability outside of this range. The color shade of a lake may also vary with pH.

Table 10 gives typical physical properties for most FD&C lakes. Table 11 lists relative light stability information for some lake pigments. These values can

vary, depending on the environment in which the pigment is dispersed. The data shown relate fading properties of one pigment to another in a hydroxypropyl methylcellulose (HPMC) film system. This information can be used as a general guideline. However, actual light stability

Table 10 FD&C aluminum lakes: typical physical data

Average particle size	5.0–10.0 μm
Moisture content	12%–25%
Oil absorption	40–45 ^a
Bulk density	15–20 lb/cu ft
Specific gravity	1.7–2.0 g/cc
pH Stability range	4.0–8.0

^aASTM method D281-31, expressed in grams of oil per 100 g of pigment.

Table 11 Fading characteristics of some FD&C and other lakes^a

Pigment	Rating
D&C Red #30, 30% AL	0.5
D&C Red #7, 50% toner	0.5
FD&C Blue #2, 34% AL	1
FD&C Yellow #5, 38% AL	1.5
FD&C Yellow #5, 15% AL	1.5
FD&C Blue #2, 12% AL	1.5
Quinoline yellow, 23% AL	2
FD&C Red #40, 40% AL	3
Carmoisine, 25% AL	4.5
FD&C Red #40, 15% AL	5
D&C Yellow #10, 17% AL	5
FD&C Yellow #6, 40% AL	5.5
FD&C Blue #1, 30% AL	6
FD&C Yellow #6, 15% AL	7
Ponceau 4R, 25% AL	7
FD&C Red #3, 40% AL	7.5
FD&C Blue #1, 12% AL	7.5
FD&C Red #3, 15% AL	8.5
Patent blue V, 25% AL	9

^aA hydroxypropyl methylcellulose film of pigment was exposed in a fadeometer for 3, 6, and 12 h. The rate at which the colorant degraded was noted from tint strength calculations and placed on a scale of 0–10, with the higher numbers indicating an increased tendency to fade. AL—aluminum lake.

information on a given dosage form must be evaluated when assessing shelf life and packaging concerns.

SUMMARY

As you can see, selection of colorants for use in global drug development can be very complicated and time-consuming. With the differences in colorant regulations worldwide and the need for various performance attributes based on the dosage form, there are numerous considerations that must be assessed. Therefore, it is important that appropriate expertise be consulted before finalizing colorant selection to help prevent future development or registration problems.

REFERENCES

- Newsome, R.L. Food Colors—A Scientific Status Summary by the Institute of Food Technologists Expert Panel on Food Safety and Nutrition. *Food Technol.* **1986**, 49–56.
- Catalog of Food Colors*; ILSI Color Committee of the International Life Sciences Institute: Washington, DC, 1983; 1–3.
- Food and Drug Administration. Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Foods. Red Book **1982**.
- Joint FAO/WHO Expert Committee on Food Additives. Specification for the Identity and Purity and Toxicological Evaluation of Food Colors. WHO Food Add. **1966**, 66 (25), FAO Nutr. Mtgs. Rept. Ser. No. 38B.
- Joint FAO/WHO Expert Committee on Food Additives. Toxicological Evaluation of Some Food Colors, Emulsifiers, Stabilizers, Anti-Caking Agents and Certain Other Substances. WHO Food Add. **1969**, 70 (36), FAO Nutr. Mtgs. Rept. Ser. No. 46A.
- Evaluation of Certain Food Additives. 21st Report of the Joint FAO/WHO Expert Committee on Food Additives, **1978**, WHO Tech. Rep. Ser. 617.
- Code of Federal Regulations. **2000**, Title 21, Part 201.20.
- Official Journal of the European Communities **22.9.95**, L226.
- Code of Federal Regulations **2000**, Title 21, Part 74.1710.
- Code of Federal Regulations **2000**, Title 21, Part 70.3f.
- Code of Federal Regulations **2000**, Title 21, Parts 74, 81, 82.
- Marmion, D.M. *Handbook of U.S. Colorants for Foods, Drugs, and Cosmetics*, 2nd Ed.; Wiley: New York, 1984.
- Official Journal of the European Communities **1962**, 5, 82–85.
- Official Journal of the European Communities **1979**, 22 (L33), 1–4.
- Official Journal of the European Communities **10.9.94**, L237.
- Food and Drug Regulations—Coloring Agents*, Part C. 01.040.2, Canadian Department of Health, August 30, 1995.
- Ministerial Ordinance to Establish Tar Colors Which Can Be Used in Pharmaceuticals*, No. 30, Japan Ministry of Health, August 31, 1966.
- Japanese Pharmaceutical Excipients Directory, *The Japan Pharmaceutical Excipients Council*; Yakuji Nippo Ltd: Tokyo, 2000.
- Nicholson, A.E.; Tucker, S.J. Coloring Solid Pharmaceutical Forms and Compositions. US Patent 2,925,365 February 16, 1960.
- Porter, S.C. Effect of Additives on Properties of an Aqueous Film Coating. *Pharm. Technol.* **1980**, 4 (3), 66.
- Blumenthal, W.B. A Review of the Literature on Color Lakes. *Ann. Dyestuff Rep.* **1946**, 35 (23), 529.
- Food Colors. *Food and Nutrition Board*; National Academy of Sciences: Washington, DC, 1971.
- Francis, F.J. *Handbook of Food Colorant Patents*; Food & Nutrition Press, Inc.: Westport Connecticut, 1986.
- Henkel, J. From Shampoo to Cereal—Seeing to the Safety of Color Additives, U.S. Food and Drug Administration, FDA Website <http://vm.cfsan.fda.gov/~dms/cos-221.html>, FDA Consumer, December 1993.
- Lachman, L.; Lieberman, H.; Kanig, J. *The Theory and Practice of Industrial Pharmacy*, 3rd Ed.; Lea & Febiger: Philadelphia, 1986.
- Lieberman, H.A., Lachman, L., Eds. *Pharmaceutical Dosage Forms*; Marcel Dekker, Inc.: New York, 1982; 1–3.
- From Color Facts—FDA/IFIC Brochure, U.S. Food and Drug Administration, FDA Website <http://vm.cfsan.fda.gov/~lrd/colorfac.html>, (accessed January 1993).
- Walford, J., Ed. *Developments in Food Colours*; Elsevier: London and New York, 1980; 1 and 2.

COLLOIDS AND COLLOID DRUG-DELIVERY SYSTEM

Diane J. Burgess

University of Connecticut, Storrs, Connecticut, U.S.A.

INTRODUCTION

The term *colloid* applies broadly to systems containing at least two components, in any state of matter, one dispersed in the other, in which the dispersed component consists of large molecules or small particles. These systems possess certain characteristic properties that are related mainly to the dimensions of the dispersed units. The colloidal size range is determined by two limits: the particles or molecules must be large relative to the molecular dimensions of the fluid in which they are dispersed so that the fluid can be assigned continuous properties; and they must be sufficiently small so that thermal forces dominate gravitational forces and they remain suspended. This sets the lower size limit at approximately 1 nm and the upper limit at approximately 1 μm . To qualify as a colloid, only one of the dimensions of the particle must be within this size range. For example, colloidal behavior is observed in fibers in which only two dimensions are in the colloidal size range. There are no sharp boundaries between colloidal and noncolloidal systems, especially at the upper size range. An emulsion system may display colloidal properties, yet the average droplet size may be larger than 1 μm . The word colloid was coined by Graham in 1861 from the Greek word for glue (*Rolla*) (1).

BIOLOGICAL AND PHARMACEUTICAL SIGNIFICANCE OF COLLOIDS

Many biological structures are colloidal in nature, such as blood (a dispersion of corpuscles in serum) and bone (a dispersion of calcium phosphate in collagen). There are also many macromolecular dispersions in the body, including enzymes and other proteins. Colloids are used medically for diagnostic imaging (radiolabeled, parenterally administered colloids), as adjuvants to improve therapy (to enhance the immune effect of various agents, such as toxins that are adsorbed onto a colloidal carrier), as a means of drug preparation (such as colloidal silver protein an effective germicide), in the preparation of dosage forms (e.g., the macromolecular colloid acacia is used as an emulsifying and suspending agent in creams, emulsions,

and suspensions), and as colloid drug-delivery systems (2–4) (including nanoparticle, microspheres, liposomes, and macromolecular drug complexes) for the purposes of drug targeting, controlled release, and protection of the drug substance. Colloid drug-delivery systems are used topically, orally, parenterally, and by inhalation. The reader is referred to the colloid drug-delivery system section of this article for further details on these systems.

CLASSIFICATION OF COLLOIDS

On the basis of interaction between the particles or macromolecules of the dispersed phase with the molecules of the dispersion medium, colloidal systems are classified into three groups: 1) *lyophilic*, solvent “loving” colloids, in which the disperse phase is dissolved in the continuous phase; 2) *lyophobic*, solvent “hating” colloids, in which the disperse phase is insoluble in the continuous phase; and 3) *association* colloids, in which the dispersed phase molecules are soluble in the continuous phase and spontaneously “self-assemble” or “associate” to form aggregates in the colloidal size range.

Lyophilic Colloids

The dispersed phase generally consists of soluble macromolecules, such as proteins and carbohydrates. These are thermodynamically true solutions; that is, they are best treated as a single phase system. The disperse phase has a significant effect on the properties of the dispersion medium and introduces an extra degree of freedom to the system. Lyophilic colloidal solutions are thermodynamically stable and form spontaneously when a solute and solvent are brought together. There is a reduction in Gibbs free energy (ΔG) on dispersion of a lyophilic colloid. ΔG is related to the interfacial area (ΔA), the interfacial tension (γ), and the entropy of the system (ΔS):

$$G = \gamma A - TS$$

where T is the absolute temperature.

The strong interaction between the solute and solvent usually supplies sufficient energy to break up the disperse

phase. In addition, there is an increase in the entropy of the solute on dispersion, which is generally greater than any decrease in solvent entropy. The interfacial tension (γ) is negligible if the solute has a high affinity for the solvent; thus, the $\gamma\Delta A$ term will approximate to zero. The shape of macromolecular colloids will vary with affinity for the solvent. Macromolecules will take on elongated configurations in a solvent for which they have a high affinity and will tend to decrease their total area of contact with a solvent for which they have little affinity by forming compact coils.

Lyophobic Colloids

The disperse phase is broken down into very small particles, which are distributed more or less uniformly throughout the solvent. The disperse phase and the dispersion medium may consist of solids, liquids, or gases and are two-phase or multiphase systems with a distinct interfacial region. As a consequence of the poor dispersed phase–dispersion interactions, lyophobic colloids are thermodynamically unstable and have a tendency to aggregate. The Gibbs free energy (ΔG) increases when a lyophobic material is dispersed throughout a medium. The greater the extent of dispersion, the greater the total surface area exposed and hence the greater the increase in the free energy of the system. When a particle is broken down into smaller particles, work is needed to separate the pieces against the forces of attraction between them (W). The resultant increase in free energy is proportional to the area of new surface created (A):

$$G = W = 2\gamma A$$

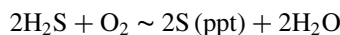
Molecules that were originally bulk molecules become surface molecules. In the surface environment the molecules have different configurations and energies than those in the bulk. An increase in free energy arises from the difference between the intermolecular forces experienced by surface and bulk molecules. Lyophobic colloids are aggregatively unstable and can remain dispersed in a medium only if the surface is treated to cause a strong repulsion between the particles. Such treated colloids are thermodynamically unstable yet are kinetically stable since aggregation can be prevented for long periods.

Preparation of lyophobic colloids

As a direct consequence of the thermodynamics of lyophobic colloidal systems, their preparation requires an energy input. If the dispersed phase is a solid, then particles may be produced by crushing, grinding, or controlled crystallization. The most direct method is by

grinding in a colloid mill (5–7). Coarse suspension particles are subjected to high shear by forcing them through a narrow gap between two rapidly rotating surfaces. The particles are torn apart by the shearing process. Ultrasonic energy may also be used to break up the disperse phase. Dispersing agents such as surfactants are often employed to prepare stable colloids. Surfactants reduce the interfacial tension and hence the free energy of the system.

Chemical methods of preparing lyophobic colloids include dissolution and reprecipitation, condensation from a vapor, and chemical reactions such as reduction, oxidation, and double decomposition. For example, mists or fogs may form spontaneously from a supersaturated vapor, provided that the degree of supersaturation is sufficiently high. Gold colloids can be produced by reducing chlorauric acid (HAuCl_4) with hydrogen peroxide or red phosphorous. Sulfur colloids can be produced by oxidation of hydrogen sulfide as follows:



The sulfur precipitate forms as a colloidal dispersion.

Emulsions

Emulsion systems can be considered a subcategory of lyophobic colloids. Like solid–liquid dispersions, their preparation requires an energy input, such as ultrasonication, homogenization, or high-speed stirring. The droplets formed are spherical, provided that the interfacial tension is positive and sufficiently large. Spontaneous emulsification may occur if a surfactant or surfactant system is present at a sufficient concentration to lower the interfacial tension almost to zero.

Clay minerals

Clay minerals such as silicates and kaolin form another subcategory of lyophobic colloids (5). Kaolin is used as a suspending agent in the preparation of dosage forms. Clays have a plate-like crystal habit. A kaolinite crystal is composed of alternate sheets of silica tetrahedra and alumina octahedra, which are bound by van der Waals forces and hydrogen bonding. The silicon–oxygen tetrahedra are linked to form hexagonal rings, and, in turn, these rings are linked to form a 2D sheet. Refer to the text by van Olphen (8) for further information on clay colloid chemistry.

Association Colloids

Association colloids are aggregates or “associations” of amphipathic surface active molecules. These molecules

are soluble in the solvent, and their molecular dimensions are below the colloidal size range. When present in solution at concentrations above a certain critical value (the critical micelle concentration), these molecules tend to form association colloids (micelles) (Fig. 1).

Amphipathic molecules consist of two parts: one of which has a high affinity for the medium (lyophilic), and the other has a low affinity for the medium (lyophobic). They tend to adsorb at interfaces to reduce the interfacial energy between the lyophobic portion of the molecule and the medium. On micellization the lyophobic portions of surfactant molecules associate to form regions from which the solvent is excluded, while the lyophilic portions of the molecules remain on the outer surface. The reader is referred to (9, 10) for the pioneering work in this area. Not all surfactants form micelles since a subtle balance between the lyophilic and lyophobic portions of the surfactant molecule is required. It appears that a charged, a zwitterionic, or a bulky oxygen-containing hydrophilic group is required to form micelles in an aqueous medium (11). These moieties are able to undergo significant hydrogen bonding and dipole interactions with water to stabilize the micelles. Micelle formation in strongly hydrogen-bonded solvents is very similar to that in water (11). Micelle formation is spontaneous, depending on the

lyophilic–lyophobic balance of the surfactant, surfactant concentration, and temperature.

At room temperature, micellization of amphipathic surfactants dissolved in water is driven by a significant increase in entropy. The hydrophobic part of the molecules induces a degree of structuring of water in the immediate area as a result of unfavorable interactions. This structuring disrupts the hydrogen-bond pattern and causes a significant decrease in the entropy of the water. This is known as the “hydrophobic” effect. The effect of temperature and pressure on micellization is dependent on surfactant properties (11, 12). Ionic surfactants generally exhibit a “Krafft” point. In these systems, micelles form only at temperatures above a certain critical temperature, the Krafft temperature. This is a consequence of a marked increase in the solubility of the surfactant at this temperature. Nonionic surfactants tend to behave in an opposite manner. Above a certain temperature (the cloud point), nonionic surfactants tend to aggregate and separate out as a distinct phase (13).

Micellar shape (Fig. 1) is dependent on surfactant concentration. At low concentration, the micelles are spherical, have well-defined aggregation numbers, and therefore are monodisperse. As the value is increased above the critical micelle concentration, micellar shape becomes distorted, forming cylindrical rods or flattened discs (14). At very high concentrations, liquid crystals form (15–17). Under specific conditions surfactants can form 2D membranes or bilayers separating two aqueous regions (similar to biological membranes). If this bilayer is continuous and encloses an aqueous region, then a vesicle results (liposomes). Insoluble drug substances can be solubilized within the interior of a micelle (Fig. 1). Micellar solubilization allows the preparation of water-insoluble drugs within aqueous vehicles (18). This is advantageous, particularly for intravenous delivery of water-insoluble drugs. Entrapment within a micellar system may increase the stability of poorly stable drug substances and can enhance drug bioavailability (19).

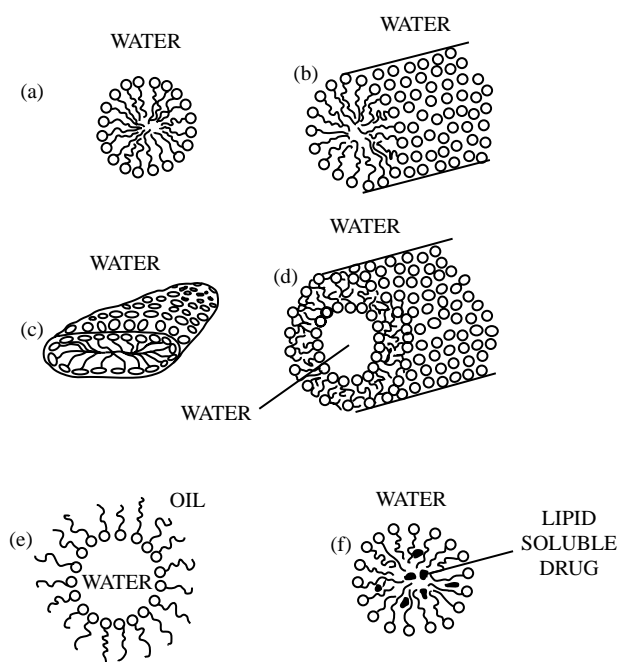


Fig. 1 Association colloids: (a) spherical micelle, (b) cylindrical micelle, (c) flattened disc-shaped micelle, (d) microtubular micelle, (e) inverted micelle, (f) micelle swollen by the presence of solubilized lipid soluble drug.

PROPERTIES OF COLLOIDS

The significant characteristics of colloids are particle size and shape, scattering of radiation, and kinetic properties.

Particle Size and Shape

The colloidal size range is approximately 1 nm to 1 μm and most colloidal systems are heterodisperse. Solid

dispersions usually consist of particles of very irregular shape. Particles produced by dispersion methods have shapes that depend partly on the natural cleavage planes of the crystals and partly on any points of weakness (imperfections) within the crystals. The shape of solid dispersions produced by condensation methods depends on the rate of growth of the different crystal faces. Treatments of particle shape are given by Beddow (20), Allen (21), and Shutton (22). The method of particle size measurement should reflect the aspect of the particle that is of most interest. This may be the surface area of the particle or its settling radius. To characterize hetero-disperse systems, it is necessary to determine the particle size distribution. Various theoretical distribution functions have been proposed such as the normal or Gaussian distribution and the log-normal distribution.

Microscopy

Most colloids are below the limit of resolution of the optical microscope but can be visualized by transmission electron microscopy (TEM) and scanning electron microscopy (SEM) (23, 24). TEM produces 2D images from which particle shape interpretation is difficult; however, size may be determined because of the high resolution. SEM produces 3D images. To obtain a value for the dimensions of an irregular particle, several possible measurements can be made (21). These include

1. Martin's diameter (d_m). This is defined as the length of a line that dissects the image of the particle.
2. Feret's diameter (d_f). This is defined as the distance between two tangents on opposite sides of the particle, which are parallel to some fixed direction.
3. Projected area diameter (d_a). This is defined as the diameter of a circle having the same area as the particle.

Ultracentrifugation

The Stokes settling radius of colloidal particles can be obtained from their sedimentation rate. An ultracentrifuge is used to increase the sedimentation rate (21), since colloidal particles settle very slowly under the influence of gravity alone.

Electronic pulse counters

Counters such as the Coulter Counter determine the number of particles in a known volume of an electrolyte solution. The suspension is drawn through a small orifice that has an electrode on either side. The dispersed phase particles interfere with current flow, causing the resistance to change. Resistance changes are related to the particle volume.

Light scattering

Both intensity and dynamic light scattering can be used as methods of particle size analysis of colloids. The simplest method is to measure the turbidity. Turbidity is defined as the reduction in the intensity of light as it passes through a colloidal sample. The loss in intensity is due to scattering of light and can be used to determine the average molar mass of a lyophilic colloid (25, 26). Dynamic light scattering, which is known as quasielastic light scattering or photon correlation spectroscopy, is used to obtain an estimate of the diffusion coefficient of a colloidal system, from which a particle radius can be determined by applying the Einstein and the Stokes equations (27, 28).

Hydrodynamic chromatography

To determine the particle size, a colloidal dispersion is forced through a long column packed with nonporous beads with an approximate radius of 10 μm . Particles of different size travel with different speeds around the beads and thus are collected in size fractions (29).

Scattering of Radiation

The scattering of a narrow beam of light by a colloidal system, such as a fog or a mist, to form a visible cone of scattered light is known as the Faraday–Tyndall effect (30, 31). The first detailed theory of light scattering by small particles was developed by Rayleigh in 1871. Light scattering theory was further developed by Mie (1908), Debye (1915), and Gans (1925). A full account of light scattering by colloidal systems is given in the texts of Hiemenz (26) and van de Hulst (25). When electromagnetic radiation falls on a material, oscillating dipoles are induced in the material. These serve as a secondary source of emission of scattered radiation with the same wavelength (λ) as the incident light. The intensity of the scattered light depends on the intensity of the original light, the polarizability of the material, the size and shape of the material, and the angle of observation. The scattering intensity increases with increases in the particle radius, reaching a maximum and then decreasing. Small particles ($\lambda/20 \geq r$) act as point sources of scattered light. In large particles, different regions of the same particle may behave as scattering centers. These scattering centers interfere with one another, either constructively or destructively. As particles become larger, the number of scattering centers increases, and the resultant destructive interference causes a reduction in the intensity of light scattering. As a consequence of these conflicting effects, visible light is scattered most intensely by particles within the colloidal size range.

Kinetic Properties

The kinetic properties of colloids are characterized by slow diffusion and slow, often negligible, sedimentation under gravity.

Thermal motion

The thermal motion of particles in the colloidal size range is known as Brownian motion (32, 33), after the English botanist Robert Brown who first observed the random directional movement of colloidal particles in 1827. The particles display a zigzag-type movement, which is a result of random collisions with the molecules of the suspending medium, other particles, and the walls of the containing vessel. The distance moved by a particle in a given period of time is related to the kinetic energy of the particle and the viscous friction of the medium. As a result of Brownian motion, colloidal particles diffuse from regions of high concentration to regions of lower concentration until the concentration is uniform throughout. Since diffusion is inversely related to particle radius, the diffusion of colloidal particles is relatively slow compared with that of small molecules or ions. Gravitational forces, which cause particles to sediment, and Brownian motion (diffusion forces) oppose one another. Colloids are of the size range at which the Brownian forces are stronger than gravitational forces, so they tend to remain suspended.

Osmosis

Osmosis is similar to diffusion in that the molecules move from a location of high chemical potential to one of low chemical potential. An osmotic pressure is generated in a colloidal solution when it is separated from its solvent by a barrier that is impermeable to the solute but is permeable to the solvent. The pure solvent will flow across the membrane, diluting the colloidal dispersion and, as the colloidal material cannot flow in the opposite direction, a pressure difference (osmotic pressure) will be created between the two compartments. Osmotic pressure is a colligative property and therefore can be related to the relative molecular mass of the colloidal material.

Viscosity

The viscosity of a system is a measure of its resistance to flow under an applied stress. Colloids, especially lyophilic colloids, tend to increase the viscosity of a system. The large molecular chains become entangled in one another, increasing the resistance to flow. The viscosity of a colloidal system is related to the shape, molecular weight, and concentration of the colloid. Viscosity measurement can be used to obtain an approximate molecular weight value.

LYOPHOBIC COLLOID STABILITY

There is a well-developed theory—the Derjaguin–Landau and Verwey–Overbeek (DLVO) theory—to describe the interaction between particles of a lyophobic colloid. This is reviewed in the texts of Hunter (1987) and Hiemenz (1986) and is based on the assumption that the van der Waals interactions (attractive forces) and the electrostatic interactions (repulsive forces) can be treated separately and then combined to obtain the overall effect of both of these forces on the particles.

Attractive Forces

van der Waals attractive forces between two particles are considered to arise from dipole–dipole interactions and are proportional to $1/H^6$, where H is the separation distance between the particles (34). These attractive forces were used to explain the phenomenon of colloid aggregation employing the Hamaker summation method to calculate the van der Waals interaction energy (V_A). This is based on the assumptions that the interactions between individual molecules in two colloidal particles can be added together to obtain the total interaction and that these interactions are not affected by the presence of all the other molecules. The modern dispersion theory of Lifshitz (35) overcomes the assumptions made in the London theory. The Lifshitz theory is based on the idea that the attractive interaction between particles is propagated as an electromagnetic wave over distances that are large compared with atomic dimensions. The colloidal particle is considered to be made up of many local oscillating dipoles that continuously radiate energy. These dipoles are also continuously absorbing energy from the electromagnetic fields generated by all the surrounding particles. Refer to Hunter (1987) for a full treatment of this theory.

Repulsive Forces

The electrostatic interaction between two colloidal particles is a consequence of the charge carried by the particles. All colloidal particles acquire a charge when dispersed in an electrolyte solution, by adsorption of ions from the solution onto the particle surface, by ionization of ionizable groups on the particle surface, or by selective ion dissolution from the particle surface (36). The charged particle surface attracts ions of the opposite sign (counter ions). This attraction is so strong that some of the counter ions are tightly bound to the particle surface. The counter ions also experience an attractive force, which

is exerted by the bulk solution and is caused by thermal motion. As a result of these two opposing effects, electrostatic attraction and thermal motion, a diffuse layer of ions builds up so that at a distance from the particle surface it appears to be electrically neutral. The tightly bound and the diffuse layers are known as the electrical double layer (30, 31). The repulsive interaction between the electrical double layers of two particles is caused by the free energy change involved when overlap occurs and the osmotic pressure generated by the accumulation of ions between the particles. The repulsive interaction (V_R) decreases exponentially with increase in the distance between the particles (30).

The DLVO theory is based on a combination of these two effects to explain the aggregative instability of two particles at any given separation distance. The resultant effect on the particles is calculated by summing these two opposing forces. This is shown diagrammatically in Fig. 2, where V_T is the summation of V_A and V_R ($V_T = V_A + V_R$). van der Waals attraction dominates at both large and small separation distances. At very small distances, van der Waals attraction increases markedly, resulting in a deep attractive well (known as the primary minimum). This well is not infinitely deep due to the very steep short range

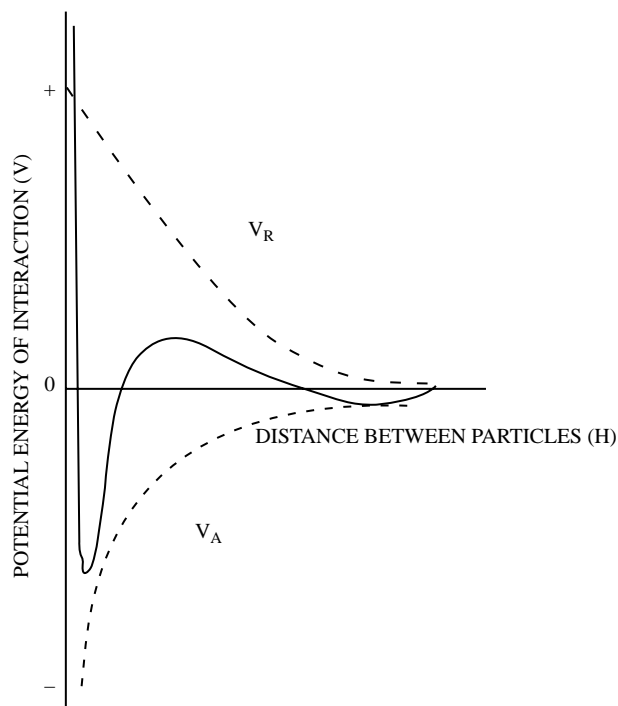


Fig. 2 Potential energy of interaction (V_T) of two spheres of equal radius, obtained by the summation of van der Waals interaction potential energy (V_A) and the repulsive interaction potential energy (V_R) ($V_T = V_R + V_A$).

repulsion between the atoms on each surface (37). A complication to the DLVO theory is the influence of the surface on adjacent solvent layers (38). The DLVO theory can be applied in a broad sense to most systems, although the kinetics of aggregation may be uncertain.

Stabilization of Lyophobic Colloids

There are two methods of stabilization of lyophobic colloids: electrostatic and polymeric. Electrostatic stabilization results from charge–charge repulsion, as discussed previously. Polymeric stabilization is achieved by the adsorption of macromolecules (lyophilic colloids) at the surface of a lyophobic colloid (39, 40). Macromolecules of at least a few thousand molecular weight are required, as they must extend in space over a distance comparable to, or greater than, the distance over which van der Waals attraction is effective. The macromolecules must be mutually repulsive to impart colloid stability. This is achieved when they are present at a sufficiently high concentration that they saturate the surfaces of the particles (39). The colloidal particles will then repel one another as a result of volume restriction and osmotic pressure effects as a result of the high concentration of polymer chains in a confined area (Fig. 3). Hunter (37) gives a thermodynamic account of polymeric stabilization. At low polymer concentrations (and hence low particle surface coverage), bridging flocculation may occur (Fig. 4). Bridging is a consequence of the adsorption of segments of an individual polymeric flocculant molecule onto the surface of more than one particle. A polyelectrolyte may stabilize a lyophobic colloid by a combination of steric and electrostatic stabilization, “electro-steric” stabilization.

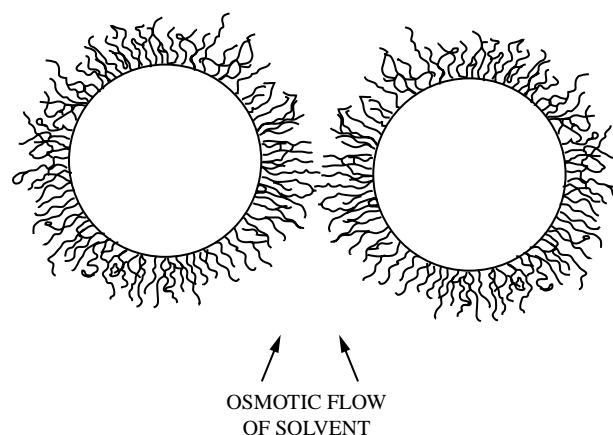


Fig. 3 Steric stabilization of lyophobic colloidal particles. The particles repel one another because of volume restriction and osmotic pressure effects.

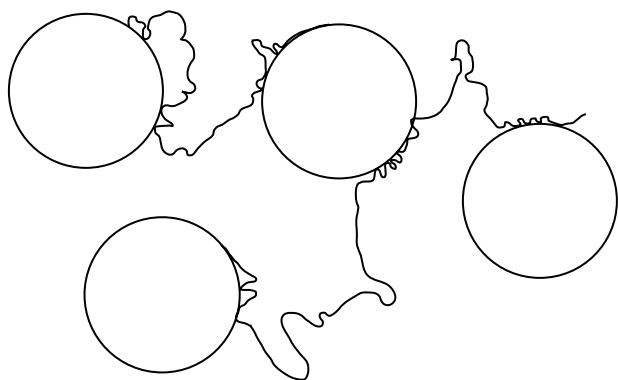


Fig. 4 Bridging flocculation. Lyophobic particles are bound together by the adsorption of polymer chains.

COLLOID DRUG-DELIVERY SYSTEMS

Colloid drug-delivery systems are used to increase the bioavailability of drug substances, to improve drug stability, to sustain and control drug-release rates, to target drugs to specific sites in the body, and to stimulate the immune system. Rate-controlled delivery can be achieved by delayed diffusion through a polymer matrix. The reader is referred to the article in this Encyclopedia on Microsphere Technology and Applications for details on controlled release from polymer microspheres (41). Rate-controlled delivery allows drugs with short biological half-lives and narrow therapeutic indices to be utilized. Drug targeting can improve the therapeutic index of a drug by optimizing the access, amplitude, and nature of interactions with the pharmacological receptor. Drug targeting can also protect the drug and the body from any unwanted and deleterious disposition. Biotechnological therapeutics are ideal candidates for site-specific delivery as these are very potent substances and are rapidly proteolytically degraded after administration. Encapsulation within a colloidal system can protect these therapeutic agents from degradation and deliver them to their sites of action. Drug targeting may be simply targeting a particular body compartment, such as the knee joint. This can be achieved by passive targeting via direct injection into the joint cavity (intra-articular injection). Greater specificity may be required, such as delivery to a particular organ, to a set of cells within an organ, or even to an intracellular structure. This usually requires active targeting, meaning that the natural distribution of the carrier is altered. In some cases it is possible to target to specific cells or intracellular structures by exploiting a natural physiological process such as macrophage uptake of foreign materials.

Colloid drug-carrier systems are of two types: particulate carriers (capsular, monolithic, or cellular) and soluble carriers (macromolecular drug conjugates). The pioneering work on soluble carrier systems was carried out by Ghose and Cerini (42), on radioimmunolocalization of tumors through the use of antibodies, and by Ghose and Nigam (43) and Rowland et al. (44) on drug-antibody conjugates. Particulate-carrier systems, including liposomes, microspheres, nanoparticles, microemulsions, erythrocytes, and vaccines, have been used as a means of drug targeting. The use of liposomes (phospholipid vesicles) for drug targeting was proposed by Sessa and Weissmann (45) and Gregoriadis et al. (46).

The fate of intravenously administered colloidal particles is controlled by particle size and particle surface properties. Particles larger than 7 μm are rapidly filtered by the fine capillary beds of the lung (2, 46, 47). Particles (100 nm to 7 μm) are rapidly taken up by the cells of the reticuloendothelial system (RES) if they are recognized as foreign (48, 49). Particle size can thus be utilized as a means of passive targeting. To target specific cells in other organs, colloidal carriers must escape from the vascular compartment (extravasate). This is generally difficult to achieve, as the plasma membrane is usually continuous and impermeable to large molecules and small particles. Molecules (5000–10,000 Da) may pass into the interstitial space to be collected by the lymphatic system. Some capillary membranes are fenestrated, containing small holes (fenestrae) 50–60 nm in size, which allow small particles to pass through. Sinusoidal or discontinuous membranes are present in the parenchymal cells of the liver, spleen, and bone marrow. These membranes have gaps of the order of 150 nm. Fenestrations may also occur in areas of tissue inflammation and infection (50); hence, the uptake of colloid carrier systems may be enhanced in these areas. The intracellular transport of certain drug-carrier complexes may occur via endocytosis (51). Endocytotic vesicles form at cell surfaces for the purpose of transporting fluids (pinocytic vesicles) and solids (phagocytic vesicles) into the cells and through the cells to the opposite membrane. Receptor-mediated transcytotic routes exist for certain macromolecules, and these can be exploited for the purpose of drug delivery.

Targeting the RES

Passive targeting to the RES can be utilized to treat diseases of the RES (neoplasms and parasitic, viral, and bacterial infections) and for macrophage stimulation to treat pulmonary metastases (52). Liposomes have been used to target antimony compounds to the liver and to treat leishmaniasis (53), and nanoparticles have been used to

target the flukicide nitroxylin (54) to the liver. Various carrier systems have been used to target cytostatic agents to the RES to treat monocytic leukemia, histiocytic medullary reticulosis, and certain forms of Hodgkins disease (55). Macrophages are the primary barrier against the growth and metastatic spread of neoplastic cells (56–58). Macrophage-activating factors (such as, muranyl di- and tripeptides, modified lymphokines, and interferon) that activate the macrophage to the tumoricidal state can be incorporated into colloidal systems (such as liposomes), and thus these agents will be taken up at their active site (the macrophage). Fidler (59) has successfully delivered lymphokines to activate macrophages to abrogate the development of micrometastases after surgical removal of a primary tumor.

Avoiding the RES

Colloidal properties that influence RES uptake are particle size, surface charge, surface hydrophobicity, and the adsorption of macromolecules onto the particle surface. The surface of colloidal particles can be altered to avoid RES uptake by adsorption or grafting of a hydrophilic polymer onto the surface of a particle and thereby creating an energy barrier to particle interaction (e.g., the nonionic surfactant Tween 20 can be adsorbed) (60). Both biological and synthetic polymers have been used for RES masking of colloidal particles, for example, albumin (61), immunoglobulin G (62), carboxymethylcellulose (63), poloxamers (64, 65), and poloxamines. RES uptake can also be avoided by saturating the RES with placebo colloidal particles prior to the administration of the colloidal carrier system (66). However, blockade of the RES is not clinically acceptable.

Types of Colloidal Carrier Systems

Soluble-carrier systems

At least three types of polymer–drug conjugate systems exist: 1) those active at the cell surface, 2) those active following endocytic capture by cells, and 3) those capable of releasing drug extracellularly in a controlled manner (67). Targeting to the cell surface can be achieved by utilizing the specific properties of the target cell. Antibodies specific for a variety of cell surface determinants on neoplastic B cells, such as sIgD or the sIg idiotype, can be complexed to drugs, toxins, or isotopes. Although drug–antibody conjugates have the ability to target tumor sites, the linking procedure may result in inactivation of the drug or of the antibody (68). Monoclonal antibodies can react with normal tissue, and

immune responses may occur. Tumor heterogeneity also causes problems with this approach to drug targeting. Specific intracellular drug action can be achieved through endocytosis of drug–carrier conjugates. Specific receptors exist for various macromolecules, and these can be exploited to achieve intracellular transport of drugs. Dextran have been used to deliver drugs extracellularly (69).

Particulate-carrier systems

Liposomes: These are artificial lipid vesicles consisting of one or more lipid bilayers enclosing a similar number of aqueous compartments (70). Liposomes can be subcategorized into 1) small unilamellar vesicles (SUV), 25–70 nm in size, that consist of a single lipid bilayer, 2) large unilamellar vesicles (LUV), 100–400 nm in size, that consist of a single lipid bilayer, and 3) multilamellar vesicles (MLV), 200 nm to several microns, that consist of two or more concentric bilayers. Unmodified liposomes passively target the Kupffer cells in the liver (71) and have been used to target drugs to the liver. SUVs may target the liver parenchymal cells, as they pass through the “sieve plates.” Liposomes are relatively unstable, have low carrying capacities, and tend to be “leaky” to entrapped drug substances. The properties of liposomes can be altered by the incorporation of various molecules such as cholesterol, cetylphosphate, and stearylamine into the phospholipid bilayer. The presence of cholesterol results in a more stable, less “leaky” membrane. Polymerized liposomes have been developed that are more stable and less “leaky.” These are composed of phospholipids with polymerizable moieties. Immunoglobulins have been attached to the surface of liposomes to actively target them to specific sites. Polyethylene glycol-coated (pegylated) liposomes, also known as “stealth” liposomes, have been prepared, which have extended biological half-lives (72–74). As a consequence of the polyethylene glycol coating these liposomes are able to avoid RES uptake. Liposomes prepared with cationic and fusogenic lipids are currently being utilized in gene therapy to deliver DNA into target cells (75–77). The cationic lipids bind with the DNA and the fusogenic lipids are able to fuse with cell membranes, allowing the DNA to be introduced into cells.

Emulsions (lipid microspheres): Colloid-sized emulsion droplets are used in drug delivery to solubilize water-insoluble drug substances and for the purposes of controlled and sustained release and drug targeting. Fat emulsions such as intralipid are used for parenteral nutrition (78). The stability of parenterally administered emulsion droplets is dependent on the nature of the emulsifying agent used and on the presence of minor

components. Emulsion droplets are usually taken up by the cells of the RES. Lecithin is most frequently used to stabilize parenteral emulsions. The coalescence and behavior of oil droplets stabilized by phospholipid emulsifiers have extensively studied by Davis and Hansrani (79). The deposition of emulsion droplets is dependent on the charge carried by the droplets and the nature of the oil phase. Emulsions of vegetable oils are preferentially deposited in adipose tissue, lactating mammary glands, and myocardium muscle (80). Interaction of emulsions with macrophage cells can be reduced by coating with hydrophilic polymeric materials. Drug-release rates from emulsions are controlled by the particle size of the dispersed phase and emulsion viscosity. Lipid emulsions are very stable, can be stored for 2 years at room temperature (81), and have a low incidence of side effects, even at doses of 500 ml (which are typically given for parenteral nutrition purposes). Emulsion systems for IV nutrition and drug delivery have been reviewed recently by Lyons and Carter (82).

Lipoproteins: These are natural body transporters for cholesterol, triacycloglycerols, and phospholipids. These include high-density lipoproteins (HDL), 10 nm in size; low-density lipoproteins (LDL), 23 nm in size; and very low density lipoproteins (VLDL), 30–100 nm in size (83). LDL has been targeted into endothelial cells (84), and trigal-chol LDL has been targeted into Kupffer cells (85). The use of lipoproteins for site-specific delivery has been reviewed by Vitols et al. (86).

Microspheres: Microspheres are usually solid, approximately spherical particles containing dispersed drug in either solution or microcrystalline form. Incorporation of drugs can be achieved by entrapment during production (polymerization, gelation, or encapsulation techniques, such as coacervation and phase separation) and by covalent and ionic attachment. The drug-loading capacities are usually fairly high, up to 50%. Microspheres may be prepared from natural polymers such as gelatin and albumin (87, 88) and from synthetic polymers such as polylactic and polyglycolic acid. The drug is either totally encapsulated within a distinct capsule wall or is dispersed throughout the microsphere. Drug release is controlled by dissolution and diffusion of the drug through the microsphere matrix or the microcapsule wall, or by polymer degradation. Microspheres range in size between approximately 1 and 1000 μm . Consequently, these systems are outside the conventional colloidal size range. In the pharmaceutical literature, however, microcapsules with sizes up to approximately 15 μm are considered as colloidal-delivery systems. Magnetic microspheres composed of albumin and iron oxide have been used as a means of targeting specific areas in the body (89). Polymeric

microsphere systems have been reviewed by Davis et al. (2) and Benita (90).^a

Nanoparticles: Nanoparticles are similar to microspheres but have particle sizes in the nanometer range (10–1000 nm). It is unlikely that nanometer-sized polymeric systems are “capsule-like” and have discrete shells. Nanoparticles can be made from nonbiodegradable materials, such as methylmethacrylate (91), or from biodegradable materials, such as alkylcyanoacrylate (92) and albumin. Widder (89) prepared magnetic albumin microspheres containing Fe_3O_4 in the size range 10–20 nm. Nanoparticles are used for drug targeting, both active and passive. The relatively small size of these systems limits their use, as only small quantities of material can be encapsulated. Other types of (nonbiodegradable) nanoparticle systems include colloidal sulfur and colloidal gold. Colloidal sulfur is used as a diagnostic agent (labeled with $^{99\text{m}}\text{Tc}$). It is usually protected from aggregation by the addition of gelatin as a polymeric stabilizer. Colloidal gold is also used as a diagnostic (^{198}Au) and as a therapeutic agent (93). Again, gelatin is used as a stabilizer. Heat denatured macroaggregated $^{99\text{m}}\text{Tc}$ albumin spheres are used as diagnostic agents. Polyacrylamide nanoparticles containing fluorescein have been used to monitor cellular processes (92). Diagnostics can be passively or actively targeted. Particle coating to avoid RES uptake has been utilized to achieve targeting of diagnostic agents to the bone marrow (90). Nanoparticles are also used as immune adjuvants. Polymethylmethacrylate nanoparticles have been shown to possess adjuvant activity for tetanus toxoid (92) and for influenza virus (94). The adjuvant effect was equal to or better than that of colloidal aluminum hydroxide particles, the traditional adjuvant.

Viruses: Viruses provide a unique and extremely efficient method of insertion of genes into mammalian cells, such as human bone marrow stem cells (95). SV 40, adenovirus, vaccine virus, and polyomavirus have been used as gene transfer vectors. A nonreplicating adenoviral vector that efficiently infects human cells has recently been developed (96). The essential feature of a retroviral gene delivery system is the presence of an RNA copy of the replacement gene that is packaged in a viral particle, capable of specific and efficient entry into the cytoplasm of a cell (97). A retrovirus consists of an RNA-protein core encapsulated in a lipid envelope. The viral glycoproteins bind with specific receptors on the target cells, and the virus envelope then fuses with the cell membrane, resulting in the introduction of the RNA genome into the cytoplasm. This method of gene insertion

^aSee *Microsphere Technology and Applications*, page 1793.

is not normally harmful to cells. Potential problems include deletion of sequences during replication, recombination with endogenous viral sequences to produce infectious recombinant viruses, activation of cellular oncogenes, introduction of viral oncogenes, and inactivation of genes.

Ideal Properties of Drug-Carrier Systems

The drug carrier should accumulate selectively at the required site, achieve sufficient drug loading, be able to release the drug at the appropriate rate at the site of action, be stable in vitro and in transit to the target site in vivo, be biodegradable, be nontoxic and nonimmunogenic, be easy and inexpensive to prepare, and be sterile for parenteral use.

REFERENCES

- Bungenberg de Jong, H.G. A Survey of the Study Objects of the Volume. *Colloid Science: Reversible Systems*; Kruyt, H.R., Ed.; Elsevier: New York, 1949; 2, 1–17.
- Davis, S.S., Illum, L., McVie, J.G., Tomlinson, E., Eds. *Microspheres and Drug Therapy*; Elsevier: Amsterdam, 1984.
- Morimoto, Y.; Fujimoto, S. *CRC Crit. Rev. Ther. Carriers*. **1985**, 2, 19–63.
- Davis, S.S.; Illum, L.; Burgess, D.J.; Ratcliffe, J.; Mills, S.N. Microspheres a Controlled-Release Systems for Parenteral and Nasal Administration. *Controlled Release Technology*; Lee, P.L., Good, W.R., Eds.; ACS Symposium Series 348; American Chemical Society: Washington, 1987; 201–213.
- Hunter, R.J. *Foundations of Colloid Science*; Clarendon Press: Oxford, 1989; I, 1–48.
- Shaw, D.J. *Introduction to Colloid and Surface Chemistry*, 4th Ed.; Butterworths: London, 1992; 1–8.
- Alexander, A.E.; Johnson, P. *Colloid Science*; Oxford University Press: Oxford, 1949.
- van Olphen, H. *Introduction to Clay Colloid Chemistry*, 2nd Ed.; John Waters: New York, 1977.
- Hartley, G.S. *Aqueous Solutions of Paraffin Chain Salts*; Herman et Cie: Paris, 1936.
- McBain, J.W. *Colloid Science*; D.C. Heath Boston, 1950.
- Kresheck, G.C. Surfactants. *Water A Comprehensive Treatise*; Franks, F., Ed.; Plenum Press: New York, 1975; 95–167.
- Lindman, B.; Wennerstrom, H. *Topics in Current Chemistry*; Springer: Berlin, 1980; 87, 1–83.
- Leja, J. *Surface Chemistry of Froth Flotation*; Plenum Press: New York, 1982; 284–286.
- Tanford, C. *The Hydrophobic Effect Formation of Micelles and Biological Membranes*, 2nd Ed.; John Wiley & Sons: New York, 1980.
- Mittal, K.L., Ed. *Micellization, Solubilization and Microemulsions*; Plenum Press: New York, 1977; 1 and 2.
- Everett, D.H. *Basic Principles of Colloid Science*; Royal Society of Chemistry, Paperbacks: London, 1987; 153–166.
- Tadros, Th.F., Ed. *Surfactants*; Academic Press: London, 1984.
- Mukerjee, P.; Mysels, K.J. *Critical Micelle Concentrations of Aqueous Surfactant Systems*, NSRDS-NBS 36, National Bureau of Standards, US Government Printing Office: Washington, DC, 1971.
- Elworthy, P.H.; Florence, A.T.; MacFarlane, C.B. *Solubilization by Surface Active Agents*; Chapman and Hall: London, 1968.
- Beddow, J.K. *Particulate Science and Technology*; Chemical Publishing: New York, 1980.
- Allen, T. Particle Size Measurement. *The Powder Technology Series*; Williams, J.C., Ed.; Chapman and Hall: London, 1975.
- Shutton, H.M. *Characterization of Powder Surfaces*; Parfitt, G.D., Sing, K.S., Eds.; Academic Press: London, 1976; Ch. 3.
- Hall, C.E. *Introduction to Electron Microscopy*, 2nd Ed.; McGraw-Hill: New York, 1966.
- Silverman, L.; Billings, C.E.; First, M.W. *Particle Size Analysis in Industrial Hygiene*; Academic Press: New York, 1971.
- van de Hulst, H.C. *Light Scattering by Small Particles*; Dover Publications, Inc.: New York, 1981.
- Hiemenz, P.C. *Principles of Colloid and Surface Chemistry*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1986; 160–208.
- Chabay, I.; Bright, D.S. *J. Colloid Interface Sci.*; **1978**, 63, 304–309.
- Spragg, S.P. *The Physical Behaviour of Macromolecules with Biological Functions*; Wiley: New York, 1980; 126–140.
- Hunter, R.J. *Foundations of Colloid Science*; Clarendon Press: Oxford, 1989; I, 104–167.
- Verwey, E.J.W.; Overbeek, J.Th.G. *Theory of Stability of Lyophobic Colloids*; Elsevier: Amsterdam, 1948.
- Hiemenz, P.C. *Principles of Colloid and Surface Chemistry*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1986; 661–731.
- Shaw, D.J. *Introduction to Colloid and Surface Chemistry*, 2nd Ed.; Butterworths: London, 1992; 41–48.
- Everett, D.H. *Basic Principles of Colloid Science*; Royal Society of Chemistry; Paperbacks: London, 1987; 76–108.
- London, F. *Z. Phys.* **1930**, 63, 245.
- Lifshitz, E.M. *Sov. Phys.; JETP* **1956**, 2, 73.
- Shaw, D.J. *Introduction to Colloid and Surface Chemistry*, 4th Ed.; Butterworths: London, 1992; 133–180.
- Hunter, R.J. *Foundations of Colloid Science*; Clarendon Press: Oxford, 1989; I, 395–493.
- Israelachvili, J.N.; Pashley, R.M. *Nature* **1983**, 306, 249.
- Napper, D.H. *Polymeric Stabilization of Colloidal Dispersions*; Academic Press: New York, 1983.
- Sato, T.; Ruch, R. *Stabilization of Colloidal Dispersions by Polymeric Adsorption*; Marcel Dekker, Inc.: New York, 1980; IV, Surfactant Science Series.
- Burgess, D.J.; Hickey, A.J. Microsphere Technology and Application. *Encyclopedia of Pharmaceutical Technology*, 2nd Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 2000.

42. Ghose, T.; Cerini, M. *Nature* **1969**, 222, 993–995.
43. Ghose, T.; Nigam, S.P. *Cancer* **1972**, 29, 1398–1400.
44. Rowland, G.F.; O'Neil, G.J.; Davies, D.A.L. *Nature* **1972**, 255, 487–488.
45. Sessa, G.; Weissmann, G.J. *J. Clin. Invest* **1969**, 48, 76a–77a.
46. Gregoriadis, G.; Leathwood, P.D.; Ryman, B.E. *FEBS Lett.* **1971**, 14, 95–99.
47. Illum, L.; Davis, S.S. *J. Parent. Sci. Technol* **1982**, 36, 242–248.
48. Bellanti, J.A. *Immunology III*. W.B. Saunders: Philadelphia, 1985; 598.
49. Illum, L.; Jones, P.D.E.; Davis, S.S. *Microspheres and Drug Therapy Pharmaceutical, Immunological and Medical Aspects*; Davis, S.S., Illum, L., McVie, J.C., Tomlinson, E., Eds.; Elsevier: Amsterdam, 1984; 353–363.
50. Mizushima, Y.; Wada, Y.; Etoh, Y.; Watanabe, K. *J. Pharm. Pharmacol.* **1982**, 34, 49–50.
51. Trouet, A.; Deprez-De Campeneere, D.; DeDuve, C. *Nature (New Biol.)* **1972**, 235, 110–112.
52. Poste, G.; Fidler, I.J. *Design of Models for Testing Cancer Therapeutic Agents*; Fidler, I.J., White, R.J., Eds.; Van Nostrand Reinhold: New York, 1982; 225–238.
53. New, R.R.C.; Chance, M.L.; Thomas, S.C.; Peters, W. *Nature* **1978**, 272, 55–56.
54. Marly, J.J.; Oppenheim, R.C. *Austr. J. Pharm. Sci.* **1977**, 6, 65–76.
55. Poste, G. *Receptor-Mediated Targeting of Drugs*; Gregoriadis, G., Poste Senior, J.G., Trouet, J., Eds.; Plenum Press: New York, 1985; 427–474.
56. Kleinerman, E.S.; Erickson, K.L.; Schroit, A.J.; Folger, W.E.; Fidler, J. *Cancer Res.* **1983**, 43, 2010–2014.
57. Kleinerman, E.S.; Fidler, L.S. *Novel Approaches to Cancer Chemotherapy*; Sunkura, P.S., Ed.; Academic Press: New York, 1984; 231–249.
58. Phillips, N.C.; Moras, M.L.; Chedid, L.; Lefrancier, P.; Bernard, J.M. *Cancer Res.* **1985**, 45, 128–134.
59. Fidler, I.J. *Science* **1985**, 208, 1460–1471.
60. Davis, S.S.; Illum, L. *Site-Specific Drug Delivery*; Tomlinson, E., Davis, S.S., Eds.; John Wiley & Sons: Chichester U.K., 1986; 93–110.
61. Poznansky, M.J. *Pharmacol. Ther.* **1983**, 51, 53–76.
62. Borel, Y. *Immunol. Rev.* **1980**, 50, 71–104.
63. Diner, U.E.; Kunimoto, D.; Diener, E.L. *J. Immunol* **1979**, 122, 1886–1891.
64. Illum, L.; Davis, S.S. *FEBS Lett* **1984**, 167, 79–82.
65. Leu, D.; Manthey, B.; Kreuter, J.; Speiser, P.; DeLuca, P.P. *J. Pharm. Sci.* **1984**, 73, 1433–1437.
66. Proffitt, R.T.; Williams, L.E.; Presant, C.A.; Tin, G.W.; Uliana, J.A.; Gamble, R.C.; Baldeschweiler, J.D. *Science* **1983**, 220, 502–505.
67. Kopecek, J.; Duncan, R. *Poly [N-(2-Hydroxy Propyl) methacrylamide] Macromolecules as Drug Carrier Systems. Polymers in Controlled Drug Delivery*; Illum, L., Davis, S.S., Eds.; Wright: Bristol, UK, 1987; 152–187.
68. Vitetta, E.S.; Fulton, R.J.; Uhr, J.W. *Site-Specific Drug Delivery*; Tomlinson, E., Davis, S.S., Eds.; John Wiley & Sons: Chichester, UK, 1986; 69–91.
69. Schacht, E. *Polymers in Controlled Drug Delivery*; Illum, L., Davis, S.S., Eds.; Wright: Bristol, UK, 1987; 131–151.
70. Szoka, F.J.; Papahadjopoulos, D. *Ann. Rev. Biophys. Bioeng* **1980**, 9, 467–508.
71. Godfredsen, C.F.; Van Berkel, Th.J.C.; Kruijt, J.K.; Goethals, A. *Biochem. Pharmacol* **1983**, 32, 3389–3396.
72. Gabizon, A.; Martin, F. *Polyethylene Glycol-Coated (Pegylated) Liposomal Doxorubicin. Rationale for Use in Solid Tumours. Drugs* **1997**, 54, 15–21.
73. Torchilin, V.P. *Polymer-Coated Long-Circulating Micro-particulate Pharmaceuticals. J. Microencapsul.* **1998**, 15 (1), 1–19.
74. Lasic, D.D., Martin, F., Eds. *Stealth Liposome*; CRC Press: Boca Raton, FL, 1995.
75. Felgner, P.L. *Nonviral Strategies for Gene Therapy. Sci. Am.* **1997**, 276 (6), 102–106.
76. Felgner, P.L.; Ringold, G.M. *Cationic Liposome-Mediated Transfection. Nature* **1989**, 337, 387–388.
77. Templeton, N.S.; Lasic, D.D. *New Directions in Liposome Gene Delivery. Mol. Biotechnol* **1999**, 11 (2), 175–80.
78. Davis, S.S. *The Stability of Fat Emulsion for Intravenous Administration. Advances in Clinical Nutrition: Selected Proceedings of the 2nd International Symposium*; Johnston, I.D., Ed.; Lancaster, England, 1983; 213–239.
79. Davis, S.S.; Hansrani, P. *J. Colloid Interface Sci.* **1985**, 108, 285–287.
80. Schoeft, G.I.; French, J.E. *Proc. Roy. Soc. Lond. [Biol.]* **1968**, 109, 153–165.
81. Mizushima, Y. *Drugs Expt. Clin. Res.* **1985**, XI (9), 595–600.
82. Lyons, R.T.; Carter, E.G. *Lipid Emulsions for Intravenous Nutrition and Drug Delivery. Lipid Technologies and Applications*; Gunstone, F.D.; Padley, F.B., Eds.; Marcel Dekker, Inc.: New York, 1997; 535–556.
83. Mahley, R.W.; Innerarity, T.L. *Biochim. Biophys. Acta* **1982**, 737, 197–222.
84. Nagelkerke, J.E.; Barto, K.P.; Van Berkel, Th.J.C. *J. Biol. Chem.* **1983**, 258, 12221–12227.
85. Van Berkel, Th.J.C.; Kruijt, J.K.; Spanjer, H.H.; Nagelkerke, J.F.; Harkes, L.; Kempen, H.J.M. *J. Biol. Chem.* **1985**, 260, 2694–2699.
86. Vitols, S.G.; Masquelier, M.; Peterson, C.O. *J. Med. Chem.* **1985**, 28, 451–454.
87. Burgess, D.J.; Carless, J.E. *Int. J. Pharm.* **1985**, 27, 61–70.
88. Burgess, D.J.; Davis, S.S.; Tomlinson, E. *Int. J. Pharm.* **1987**, 39, 129–136.
89. Widder, K.; Bouret, G.; Senyei, A. *J. Pharm. Sci.* **1979**, 68, 79–82.
90. Benita, S. *Microencapsulation*; Marcel Dekker, Inc.: New York, 1996.
91. Birrenback, G.; Speiser, F. *J. Pharm. Sci.* **1976**, 65, 1763–1766.
92. Couvreur, P.; Kante, B.; Roland, M.; Guiot, P.; Bauduin, P.; Speiser, P. *J. Pharm. Pharmacol* **1979**, 31, 331–332.
93. Wade, A., Ed. *Martindale: The Extra Pharmacopoeia*; The Pharmaceutical Press: London, 1997; 1357–1358.
94. Kreuter, J.; Liehl, E. *Microbiol. Immunol.* **1978**, 165, 111–117.
95. Wilson, G. *Gene Therapy: Rationale and Realisation. Site-Specific Drug Delivery*; Tomlinson, E., Davis, S.S., Eds.; John Wiley & Sons: Chichester, UK, 1986; 149–164.
96. Karlson, S.; Humphries, R.K.; Gluzman, Y.; Nienhuis, A.W. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 82, 158–162.
97. Varmus, H.E. *Science* **1982**, 216, 812–820.

FURTHER READING

Colloids

Modern Advanced Treatises

- Hiemenz, P.C.; Rajagopalan, R. *Principles of Colloid and Surface Chemistry*, 3rd Ed.; Marcel Dekker, Inc.: New York, 1999.
- Hunter, R.J. *Foundations of Colloid Science*; Clarendon Press: Oxford, 1992; I.

Basic Treatises

- Everett, D.H. *Basic Principles of Colloid Science*; 1989.
- Shaw, D.J. *Introduction to Colloid and Surface Chemistry*, 4th Ed.; Butterworths: London, 1992.

GENERAL READING

- Alexander, A.E.; Johnson, P. *Colloid Science*; Cambridge University Press: Cambridge, 1949; 2 vols.
- Baldwin, R.W.; Byers, V.S. *Monoclonal Antibodies for Cancer Detection and Therapy*; Academic Press: London, 1985.
- Benita, S. *Microencapsulation*; Marcel Dekker, Inc.: New York, 1996.
- Davis, S.S.; Illum, L.; McVie, J.G.; Tomlinson, E. *Microspheres and Drug Therapy Pharmaceutical and Medical Aspects*; Elsevier: Amsterdam, 1984.
- Diederichs, J.E.; Muller, R.E. *Future Strategies for Drug Delivery with Particulate Systems*; CRC Press: Boca Raton, FL, 1998.
- Donbrow, M. *Microcapsules and Nanoparticles in Medicine and Pharmacy*; CRC Press: Boca Raton, FL, 1992.
- Everett, D.H. *Colloid Science (Specialist Periodical Reports) vols. 1–3*; The Chemical Society: London, 1973–1979; 4, The Royal Chemical Society, London, 1983. (Reviews of advances in the period 1970–1981.).
- Freundlich, H. *Colloid and Capillary Chemistry*, 1st Engl. Ed.; Methuen: London, 1926.
- Gregoriadis, G.; Poste, G.; Senior, J.; Trouet, A. *Receptor Mediated Targeting of Drugs*; Plenum: New York, 1985.

- Guiot, P.; Couvreur, P. *Polymeric Nanoparticles and Microspheres*; CRC Press: Boca Raton, FL, 1986.
- Hatschek, E., Ed. *The Foundations of Colloid Chemistry*; Ernest Benn: London, 1925, This book contains reprints and translations of papers by Ascherson, 1840; van Bemmelen, 1888; Faraday, 1857; Graham, 1864; Carey-Lea, 1889; Muttman, 1887; Selmi, 1845.
- Illum, L.; Davis, S.S. *Polymers in Controlled Drug Delivery*; Wright: Bristol, UK, 1987.
- Kreuter, J. *Colloidal Drug Delivery Systems*; Marcel Dekker, Inc.: New York, 1994.
- Muller, R.H. *Colloidal Carriers for Controlled Drug Delivery and Targeting: Modification, Characterization, and In Vivo Distribution*; CRC Press: Boca Raton, FL, 1991.
- Ostro, M.J. *Liposomes*; Marcel Dekker, Inc.: New York, 1983.
- Parfitt, G.D. *Principles of the Colloid State; Monographs for Teachers*; Royal Institute of Chemistry: London, 1967; No. 14.
- Papiel, W.J. *Introduction to Colloid and Surface Chemistry*; Exposition Press: New York, 1978.
- Spragg, S.P. *The Physical Behaviour of Macromolecules with Biological Functions*; John Wiley and Sons: New York, 1980.
- Tomlinson, E.; Davis, S.S. *Site-Specific Drug Delivery*; John Wiley & Sons: Chichester, UK, 1986.
- van Olphen, H.; Mysels, K.J. *Physical Chemistry: Enriching Topics from Colloid and Surface Science*; IUPAC Commission: La Jolla CA, 1975; 1.6, Theorex.
- Void, M.J.; Void, R.D. *Colloid Chemistry*; Chapman and Hall: London, 1965.
- Von Buzagh *Colloid Systems*; Technical Press: London, 1937.
- Ward, A.G. *Colloids*; Blackie and Son: London, Glasgow, 1945.

REVIEWS

- Davis, S.S. Pharm. Techno. Colloids and Drug Delivery Systems **1987**, 11, 110–117.
- Morimoto, Y.; Fujimoto, S. CRC Crit. Rev. Ther Drug Carrier Systems **1985**, 2, 119–116.
- Oppenheim, R.C. Solid Colloidal Drug Delivery Systems: Nanoparticles. Int. J. Pharm. **1981**, 8, 217–234.
- Tomlinson, E. Theory and Practice of Site-Specific Drug Delivery. Adv. Drug Deliver Rev. **1987**, 1, 87–198.
- Torchilin, V.P. CRC Crit. Rev. Ther. Drug Carrier Systems **1985**, 2, 65–115.

COMPLEXATION AND CYCLODEXTRINS

Gerold Mosher

Diane O. Thompson

CyDex, Inc., Overland Park, Kansas, U.S.A.

INTRODUCTION

Complexation is one of several ways to favorably enhance the physicochemical properties of pharmaceutical compounds. It may loosely be defined as the reversible association of a substrate and ligand to form a new species. Although the classification of complexes is somewhat arbitrary, the differentiation is usually based on the types of interactions and species involved, e.g., metal complexes, molecular complexes, inclusion complexes, and ion-exchange compounds. Cyclodextrins (CDs) are classic examples of compounds that form inclusion complexes. These complexes are formed when a "guest" molecule is partially or fully included inside a "host" molecule e.g. CD with no covalent bonding. When inclusion complexes are formed, the physicochemical parameters of the guest molecule are disguised or altered and improvements in the molecule's solubility, stability, taste, safety, bioavailability, etc., are commonly seen.

CDs form inclusion complexes with many different types of compounds, thus their potential as formulation additives has been investigated for over 40 years. They were discovered in 1891 when Villiers (1) observed crystallization occurring in a bacterial digest of starch. Schardinger's (2) evaluation of the unusual crystalline dextrins in 1903 suggested their cyclic nature but their complete structural definition did not occur until the 1940s (3, 4). This coincided with the identification of the enzyme responsible for their production (*Bacillus macerans* amylase, now referred to as cyclodextrin glucosyltransferase: CGTase: EC 2.4.1.19), and the recognition of the complexing properties of the CD cavity. In the next 30–40 years, extensive work resulted in the ability to produce each of the parent CDs in bulk quantities.

With improvements in the cost and availability of the parent CDs came increases in the volume of scientific investigation. Limitations in the pharmaceutical utility of the CDs were becoming known and derivatives were prepared with the goal of improving characteristics such as complexing ability, solubility, and safety.

Biennial international conferences (5) and reviews (6, 7) have presented the latest research in the production, characterization, and utilization of CDs in biomedical

products, foods, and cosmetics. A literature search of the 19-year period from 1967 to 1985 yields approximately 400 journal references describing CDs in pharmaceutical applications. Uekama and Otagiri (8) reviewed over half of this published literature in 1986. A search of the literature since 1986 shows that journal references have more than tripled and the patent literature has continued to grow rapidly (Fig. 1). Scientific articles have established the research applications of CDs but it is the patents that have shown the increasing interest in the commercial protection of CDs in pharmaceutical products.

The commercial viability of a CD formulation has been established with the marketing of 20 products (Table 1). Two of these products are currently on the market in the United States; one product (oral and parenteral formulations) containing a derivatized CD and the other containing α -CD. Numerous clinical trials using CD formulations have, however, been conducted or are in progress in the United States, with at least one other NDA under review.

Increasing numbers of pharmaceutical products are reaching the market place as CD formulations and research studies exploring their applications are growing exponentially. Nevertheless, the routine use of CDs in formulations is still questioned. The reluctance to develop a CD formulation is mainly due to the uncertain regulatory acceptance of a formulation containing a "nonstandard" inactive ingredient.

INCLUSION COMPLEXATION AND CDs

CDs are cyclic oligosaccharides containing 6, 7, or 8 glucopyranose units, referred to as α -, β -, or γ -CD, respectively. Each glucose unit contains two secondary alcohols at C-2 and C-3 and a primary alcohol at the C-6 position, providing 18–24 sites for chemical modification and derivatization (Fig. 2). Numerous derivatives have been prepared and described in the literature, but because of all the possible derivatives and positional and regioisomers, appropriate nomenclature must be used. The nomenclature should include at a minimum, the parent CD (α , β , or γ -CD) and the type and number of

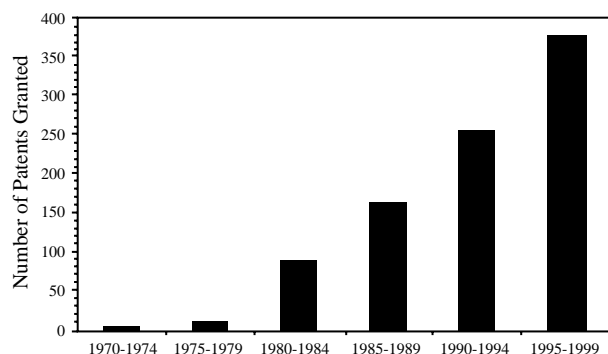


Fig. 1 Patents and patent applications in the use of CDs in pharmaceuticals. (Data collected from *Chemical Abstracts* 1967–1999.)

substituents. The substituents are usually noted by an abbreviation placed before the parent CD. Further description of the substituent group can be included with the abbreviation, if needed. For example, the hydroxyl group on the hydroxypropyl substituent can exist at one of three carbons in the propyl chain. This isomeric position is noted by a number preceding the HP notation and enclosed in parentheses. The most commonly occurring HP derivative is the (2HP)- β -CD, which is often referred to as HP- β -CD.

The number following the abbreviation of the substituent indicates the average number of substituents on the molecule, rounded to the nearest whole number. For example, HP4- β -CD indicates a β -CD with an average of four hydroxypropyl substituents. This is referred to as the molar degree of substitution (MS). It is simply a statement of the molar ratio of substituent to CD. The four substituents might be located on four different hydroxyl functionalities around the CD ring or, at the other extreme, one hydroxyl might be derivatized with a chain of four HP units. Adding further confusion is the fact that the MS can refer to the average molar degree of substitution around the CD ring, or only around each glucose ring. The latter interpretation is a carry-over from traditional carbohydrate chemistry where the lengths of the glucose chains (i.e., the molecular weights) were often not well characterized.

A similar term, the degree of substitution (DS) describes the average number of hydroxyls derivatized. Unfortunately, it has also been used to describe both the average number of hydroxyls around the CD ring and the average number around each glucose ring. A clarifying proposal to use RS (ring substituents) to designate the average number of hydroxyls derivatized in the entire CD ring and DS, the number around each glucose ring, has not been broadly embraced. The term MS will be used here

and will denote the average molar degree of substitution around the CD ring.

Regardless of which convention is used for degree of substitution, none provide any indication of the position of these substituents on the glucopyranose units. If it is known, the position of the substituent on the glucopyranose unit is indicated by a number preceding the substituent abbreviation. 6-SBE1- β -CD describes the monosubstituted sulfobutyl ether derivative with the substituent attached at one of the C-6 positions. More often than not, the substituent is introduced in a random reaction process such that introduction occurs with some defined distribution at the 2-, 3-, and/or 6-positions. For these preparations, no number precedes the substituent abbreviation. HP4- β -CD implies a tetra-substituted hydroxypropyl preparation with substituents randomly distributed over all three positions of the seven glucopyranose units. A number of common CDs and their nomenclature are given in Table 2.

The 3D structure of the parent CD provides a cavity (Fig. 3) that is hydrophobic relative to an aqueous environment. The sequestration of hydrophobic drugs inside the cavity of the CD can improve the drug's solubility and stability in water, the rate and extent of dissolution of the drug:CD complex, and the bioavailability of the drug when dissolution and solubility are limiting the delivery. These properties of the CDs enable the creation of formulations for insoluble drugs typically difficult to formulate and deliver with more traditional excipients. Numerous CDs that have different complexing abilities are available. A quantitative determination of their complexing properties is necessary for proper evaluation.

CDs form inclusion complexes with hydrophobic drugs through an equilibrium process (Fig. 4), quantitatively described in Equation (1) by an association or stability constant ($K_{a:b}$),

$$K_{a:b} = \frac{[Drug_a CD_b]}{[Drug]^a [CD]^b} \quad (1)$$

where a and b represent the molar ratio of the sequestered drug molecule to the CD. The magnitude of this associate constant can be used to compare the binding effectiveness of different CDs. Various complexes with different ratios of drug-to-CD molecules can be formed, depending on the type of CD used and the size and physicochemical characteristics of the drug molecule. In dilute solutions and/or if the drug fits entirely into the CD cavity, a 1:1 complex results. However, if the cavity is large enough, two drug molecules may be accommodated, resulting in the formation of a 2:1 complex. Conversely, if the drug is

Table 1 Commercial pharmaceuticals with CD-based formulations

Drug product	Trade name	Company	Country
PGE ₁ /α-CD	Prostandin	Ono	Japan
Intra-arterial infusion	Prostavasin	SchwarzPharma	Germany
			Italy
Intracavernous injection	Edex	SchwarzPharma	USA
Cefotiam Hexetil HCl/α-CD	Pansporin T	Takeda	Japan
Tablet			
Piroxicam/β-CD	Various	Various	Belgium
Tablet			Brazil
Suppository			France
Oral liquid			Germany
			Italy
			The Netherlands
			Scandinavia
			Switzerland
Dextromethorphan/β-CD	Rynathisol	Synthelabo	Italy
PGE ₂ /β-CD	Prostarmon E	Ono	Japan
Sublingual tablet			
Benexate/β-CD	Ulgut	Teikoku	Japan
Capsule	Lonmiel	Shionogi	
Iodine/β-CD	Mena-Gargle	Kyushin	Japan
Gargling solution			
Dexamethasone Glyteer/β-CD	Glymesason	Fujinaga	Japan
Ointment			
Nitroglycerin/β-CD	Nitropen	Nippon Kayaku	Japan
Sublingual tablet			
Cephalosporin ME 1207/β-CD	Meiact	Meiji Seika	Japan
Tablet			
Nimesulide/β-CD	Nimedex	Italfarmaco	Italy
Tablet	Mesulid Fast	Novartis	Switzerland
		Boehringer	
		Mannheim	Germany
Tiaprofenic acid/β-CD	Surgamyl	Roussel-Maestrelli	Italy
Tablet			
Chlordiazepoxide/β-CD	Transillium	Gador	Argentina
Tablet			
Omeprazol/β-CD		Hexal	Germany
Capsule			
OP-1206/γ-CD	Opalmon	Ono	Japan
Tablet			
Chloramphenicol/Me-β-CD	Clorocil	Oftalder	Portugal
Eye drop			
Cisapride/HP3-β-CD	Prepulsid	Janssen-Cilag	Belgium
Suppository			
Diclofenac/HP3-β-CD		Ciba Vision	Switzerland
Eye drop			
Ziprasidone/SBE7-β-CD	Zeldox	Pfizer	Sweden
Intramuscular injection			USA (NDA pending)
Itraconazole/HP3-β-CD	Sporanox	Janssen	USA
Oral/i.v. solution			Belgium

(Adapted from Ref. 6.)

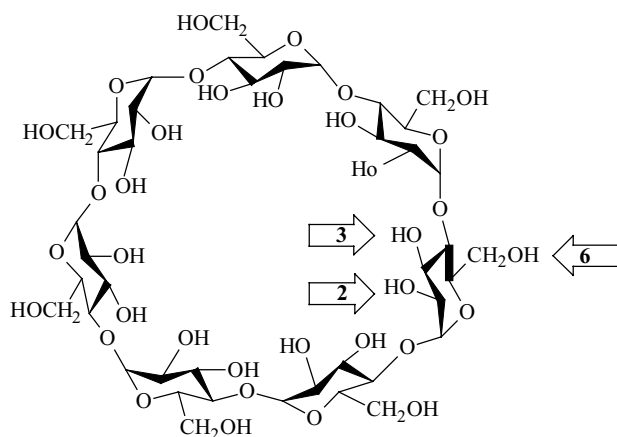


Fig. 2 Chemical structure of β -CD. Arrows indicate the 2-, 3-, and 6-hydroxyls of a glucopyranose unit. (Adapted from Ref. 6.)

very large, then several CD molecules might enclose the drug for the formation of 1:2 or higher order complexes. Although each complex has a finite stoichiometry, more than one complex may be formed in a given system. Depending on the method used to determine the association constant, it is possible to obtain a description of the stoichiometry of the complex ($a:b$).

Evaluating Inclusion Complexation

One of the most common methods of determining association constants and stoichiometry is the phase solubility technique (9). The technique involves adding an equal weight (in considerable excess of its normal solubility) of the compound to be complexed into each of several vials or ampoules. A constant volume of solvent is added to each container. Successively increasing portions of the complexing agent are then added to the vessels. The vessels are then closed and the contents brought to solubility equilibrium by prolonged agitation at constant temperature. The solution phases are then analyzed for total solute content. A phase diagram is constructed by plotting the molar concentration of dissolved solute, found on the vertical axis, against the concentration of complexing agent added on the horizontal axis. Two general types of phase solubility profiles are generated; Type A where soluble complexes are formed, and Type B where complexes of limited solubility are formed.

In Type A diagrams, an increase in solubility of the compound occurs as the amount of complexing agent increases. Soluble complexes are formed between the compound and the complexing agent, thereby increasing the total amount of compound in solution. Depending on the nature of the complexes formed, the diagram can be

linear, A_L , or show curvature in a positive, A_P , or negative, A_N , fashion (Fig. 5). Linear diagrams are formed when each complex contains only one molecule of complexing agent. When more than one molecule of complexing agent is found in the complex, an A_P -type diagram is formed. A_N diagrams are uncommon but may result if self-association is present or high concentrations of complexing agent cause alterations in the nature of the solvent.

Type B diagrams are observed when complexes of limited solubility are formed. In Fig. 5, the segment xy in curve B_S shows the formation of a complex that increases the total solubility of the compound. This is similar to a Type A diagram. At point y , however, the solubility of the complex is reached and as additional compound goes into solution, some solid complex precipitates. At point z , all of the excess solid compound added to the vials has been consumed by this process. Further addition of complexing agent beyond point z results in depletion of the compound from solution by complex formation. Curve B_I is interpreted in a similar manner except that the complex formed is so insoluble that no increase in solubility is observed.

The stoichiometry of the complexes can often be determined from the ascending and descending portions of these diagrams if certain assumptions can be made (9). If a 1:1 complex is formed, the association constant $K_{a:b}$ can be determined from the slope of the initial linear portion of the phase solubility curve, and the intrinsic solubility of the compound, S_0 , using Equation (2):

$$K_{1:1} = \frac{\text{Slope}}{S_0(1 - \text{Slope})} \quad (2)$$

Additional methods are available to determine these association or stability constants including spectroscopy [UV (10, 11), fluorescence (12, 13), NMR (14), and ORD-CD (15, 16)], potentiometry (17), microcalorimetry (18, 19), surface tension (20), membrane permeation (21), electrophoresis (22, 23), and freezing point depression (24). Chromatographic methods include HPLC (25), paper (26, 27), and TLC (28, 29) techniques.

The binding constants obtained by different methods often correlate. For example, diazepam forms a complex with β -CD with an association constant of 220 or 208 M^{-1} as determined by phase solubility (30) vs. circular dichroism (31). There is a close correlation of the binding constants (32) for bendroflumazide and cyclopentiazide as determined by the phase solubility method (56 and 165 M^{-1}) and UV method (60 and 178 M^{-1}).

However, while the above methods gave similar results, the association of β -CD and FCE24578 (17), a synthetic immunomodulator, exhibits a binding constant of 690 M^{-1}

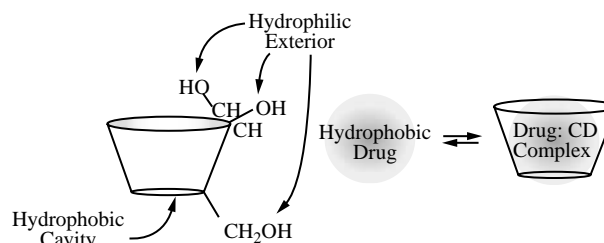
Table 2 Nomenclature and substituent structures for modified CDs

	Position of substituent	Substituent structure ^a	Nomenclature # ^b -XYZ ^{c,d} -CD ^e
<i>Parent cyclodextrins</i>			
Alpha-CD		-OH	α -CD
Beta-CD		-OH	β -CD
Gamma-CD		-OH	γ -CD
<i>Modified cyclodextrins neutral</i>			
<i>Methyl derivatives</i>			
Dimethyl	2, 6-	-O-CH ₃	2,6-DM14-CD
Methyl	Random	-O-CH ₃	M#-CD
Trimethyl	2, 3, 6-	-O-CH ₃	2,3,6-TM-CD
Ethyl derivatives	Random	-O-CH ₂ -CH ₃	E#-CD
<i>Hydroxyalkyl derivatives</i>			
2-hydroxyethyl	Random	-O-CH ₂ -CH ₂ OH	(2HE)#-CD
2-hydroxypropyl	Random	-O-CH ₂ -CHOH-CH ₃	(2HP)#-CD or HP#-CD
3-hydroxypropyl	Random	-O-CH ₂ -CH ₂ -CH ₂ OH	(3HP)#-CD
2,3-dihydroxypropyl	Random	-O-CH ₂ -CHOH-CH ₂ OH	(2,3-DHP)#-CD
<i>Modified cyclodextrins anionic</i>			
<i>Carbon Based Derivatives</i>			
Carboxy	6-	-CO ₂ M	6-C#-CD
<i>Carboxyalkyl</i>			
Carboxymethyl	Random	-O-CH ₂ -CO ₂ M	CM#-CD
Carboxyethyl	Random	-O-CH ₂ -CH ₂ -CO ₂ M	CE#-CD
Carboxypropyl	Random	-O-CH ₂ -CH ₂ -CH ₂ -CO ₂ M	CP#-CD
Carboxymethyl ethyl	2,6-; 3-	-O-CH ₂ -CO ₂ M; -O-CH ₂ -CH ₃	CME#-CD
<i>Sulfur Based Derivatives</i>			
Sulfates	2,6-random	-O-SO ₃ M	S#-CD
Alkylsulfates	6-	-O-(CH ₂) ₁₁ -O-SO ₃ M	SU#-CD
Sulfonates	6-	-SO ₃ M	6-SA#-CD
<i>Alkylsulfonates</i>			
Sulfoethyl ether	Random	-O-(CH ₂) ₂ -SO ₃ M	SEE#-CD
Sulfopropyl ether	Random	-O-(CH ₂) ₃ -SO ₃ M	SPE#-CD
Sulfobutyl ether	Random	-O-(CH ₂) ₄ -SO ₃ M	SBE#-CD

^aM: Cation^bNumbers represent position of substituents if known; if the preparation is a random distribution, then no notation implies an undefined distribution at the 2-, 3-, and 6-positions.^cLetters represent abbreviated notation of substituent.^dNumbers represent the average MS rounded to the closest whole number.^eIndication of parent CD structure, i.e., α -CD.

(Adapted from Ref. 6.)

by a phase solubility determination but a binding constant over four times higher with a UV method. This discrepancy is due to the fact that higher order complexes contribute to spectral changes and these have not been accounted for in the calculation of the UV association constant. Inconsistencies are also observed between binding constants determined by fluorescence intensity and HPLC methods in the binding of estradiol, ethinylestradiol, and estriol, to β - and γ -CD (12). The fluorometric association constants with β -CD were lower than the corresponding HPLC values, whereas for the

**Fig. 3** Complexation of drugs inside the hydrophobic cavity of CDs.

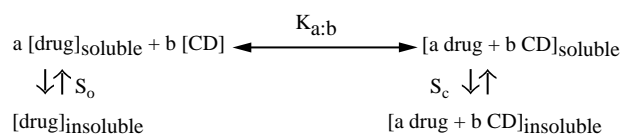


Fig. 4 Equilibrium process describing the interaction between a CD and an insoluble drug molecule to form a soluble or insoluble complex. (Modified from Ref. 6.)

complexes with γ -CD the results of both methods coincide.

pH conditions have also been observed to exert a unique effect on one method and not the other. Doxorubicin (33) and γ -CD form a complex with a $K_{1:1}$ of 617 and 718 M^{-1} as measured at pH 10 by UV and circular dichroism. This close correlation was not observed when the measurements on doxorubicin were conducted at pH 7. At pH 7, the binding constant for doxorubicin (34) was 235 M^{-1} as measured by UV but was 977 M^{-1} as measured by circular dichroism. Under a given set of conditions, a drug has only one binding constant. Therefore, this difference is reflecting how the ionization state of the drug affects the analytical measurements.

In general, binding constants can be used as an indicator of differences in binding only if the methods or conditions for determining the constants are equivalent or unaffected by the conditions. And although determining values by phase solubility might be appropriate for formulation studies, it is probably not appropriate for determination of true thermodynamic values due to the concentrations involved.

Factors Affecting Complexation

Steric effects

Cyclodextrins are capable of forming inclusion complexes with compounds having a size compatible with the dimensions of the cavity. Complex formation with molecules significantly larger than the cavity may also be possible in such a way that only certain groups or side chains penetrate into the carbohydrate channel. The three natural CDs, α , β , and γ , have different internal diameters and are able to accommodate molecules of different size. Cyclohexane is able to complex with all three CDs, but because of size naphthalene does not complex with α -CD, which has the smallest cavity. Anthracene fits only into γ -CD, which has the largest cavity.

Derivatization of the hydroxyls on one or both faces of the natural CD can impact the steric requirements for an acceptable guest molecule. The presence of bulky groups can sterically block entrance to the CD cavity. However,

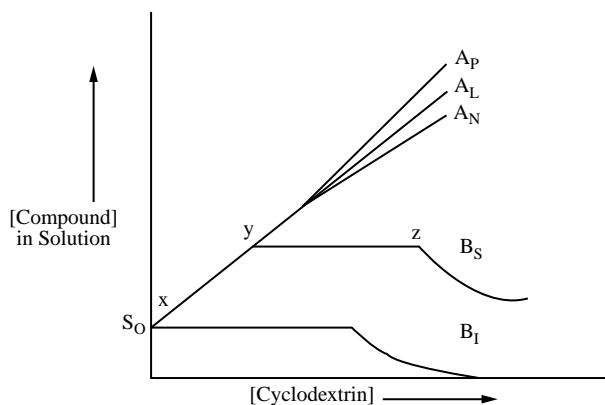


Fig. 5 Possible types of phase solubility diagrams.

some groups, depending on their number, flexibility, and position of attachment, may actually act to extend the cavity and provide for better complexation. Substitution at the 3- and 6-positions will be more likely to narrow the cavity opening while substitution at the 2- and 6-positions may allow for extension of the opening. The binding constants of flurbiprofen (35), bromazepam (36), and nitrazepam (37) to 2,6-DM- β -CD (methyl substituents in the 2- and 6-positions) are 2.3-, 2.9-, and 3.8- fold higher, respectively, than their binding to β -CD. Binding to 2,3,6-TM- β -CD (methyl substituents in the 2-, 3-, and 6-positions), however, shows constants that are less than half of that observed with β -CD.

The numbers of substituents added to the ring (MS) can also affect binding in both positive and negative manners. Müller and Brauns (38) showed that increasing the MS from 3 to 11 decreased the solubilization of hydrocortisone from 10.98 to 5.76 mg/ml for a 0.04 M HE- β -CD (hydroxyethyl) solution ($\sim 5\%$ w/v). A similar effect was observed for digitoxin, diazepam, and indomethacin. The decrease in solubilization was thought to be due to steric hindrance of the increased number of HE substituents. An additional explanation may be that some polymerization of the HE groups may have occurred during preparation of the higher MS products, thereby creating bulkier side chains that may have crowded the cavity entrance.

The hydroxypropyl substituent, being larger yet, appears to require a lower DS to improve binding without sterically obscuring the cavity entrance. Müller and Brauns (38) have studied the effect of the DS on complexing ability (Table 3) and have observed that lower degrees of hydroxypropyl substitution (2 to 5) are more conducive to complexation. As the DS increases, the solubilization of six different drugs decreases but when the DS is from 4 to 8 the solubilization is fairly consistent.

Table 3 Effect of degree of substitution on complexation of drugs by HP- β -CD

Drug	Solubility of drug in HP- β -CD solutions ^{a,b} at 25°C, pH 7.4			
	MS = 2.03	MS = 4.83	MS = 7.84	MS = 8.47
Digoxin ^a	13.12	6.39	3.76	3.70
Digitoxin ^a	8.06	4.51	1.96	2.36
Levocabastin ^b	2.20	0.45	0.31	0.09
Indomethacin ^{b, c}	6.93	8.12	6.63	8.57
Hydrocortisone ^b	19.03	13.43	10.46	10.38
Diazepam ^b	0.72	0.67	0.44	0.46

^a5% HP- β -CD solution.^b10% HP- β -CD solution.^cpH 7.4: Indomethacin in ionized state.

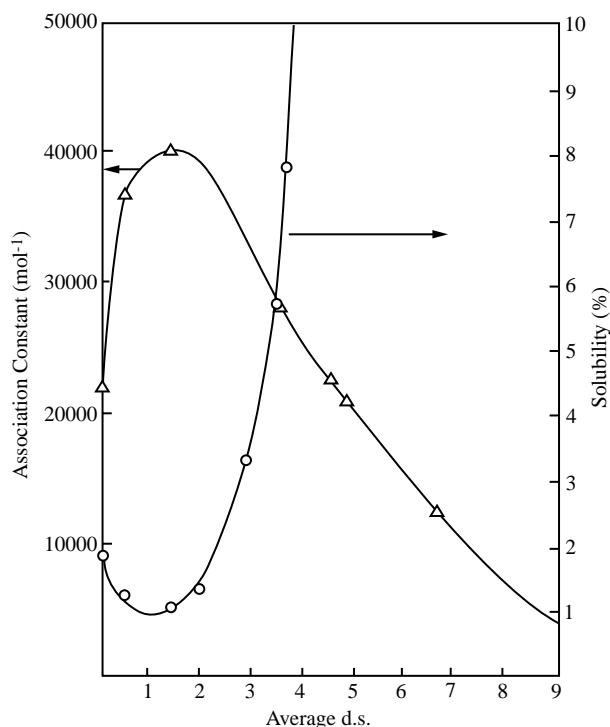
(Adapted from Ref. 39.)

There is a compromise between the steric hindrance of a substituent and its ability to extend the hydrophobic cavity. Yoshida et al. (40) have shown that introduction of the 3-hydroxypropyl (3HP) substituent ($-\text{O}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{OH}$) at an MS of ~ 6 results in higher binding constants than those observed with β -CD, apparently due to the extension of the hydrophobic cavity. The introduction of an equivalent number of 2,3-dihydroxypropyl (2,3-DHP) substituents ($-\text{O}-\text{CH}_2-\text{CH}(\text{OH})-\text{CH}_2-\text{OH}$), however, results in a decrease in the binding constants. This was speculated to be due to steric hindrance of the larger 2,3-DHP substituent, though this group, being more hydrophilic than 3HP, may not serve to extend the hydrophobicity of the cavity.

There is also a compromise between the ability to form complexes and the intrinsic water solubility. Rao et al. (41) have shown that increasing the DS of (2HP)- β -CD improves the aqueous solubility but impairs the complexation capability. Fig. 6 shows this affect for the complexation of phenolphthalein.

Greater steric interferences would be anticipated for the bulkier charged sulfoalkyl ether groups, but have not been observed. Kano et al. (42) evaluated the use of sulfopropyl ether (SPE) derivative of β -CD to interact with naphthalene. Higher association constants were observed for the SPE3- β -CD ($K = 2100 \text{ M}^{-1}$) and SPE5- β -CD ($K = 1800 \text{ M}^{-1}$) than were observed for β -CD ($K = 730 \text{ M}^{-1}$). Similar results were observed for the undecylsulfated methyl CD described by Menger and Williams (43). The lack of a steric hindrance by this highly substituted ionic derivative was explained through a "micellar" arrangement of the ionic substituents. The derivative was described as a "micellar" CD because the long hydrophobic alkyl groups in the substituent are expected to align themselves to reduce interactions with the

aqueous environment similar to micelle formation. The anionic charge at the end of the alkyl chain is expected to repel adjacent substituents effectively maintaining an opening to the CD cavity. Although the substituents are long enough to bend into the cavity, this is not expected due to the hydrophilic character of the ionic sulfate, which would prefer to interact with the aqueous solvent. The

**Fig. 6** Effect of changing the DS of 2HP- β -CD on its solubility (O) and the association constant (Δ) with phenolphthalein. (From Ref. 41.)

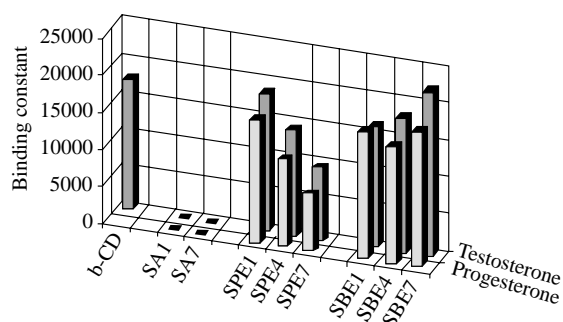


Fig. 7 Comparison of the binding constants of hydrophobic steroids, testosterone, and progesterone with β -CD and anionic β -CD derivatives. SA: Sulfonate anion at the 6-position, SPE: Anionic sulfopropyl ether substituent; and SBE: Anionic sulfobutyl ether substituent. (Adapted from Ref. 6.)

authors suggested the interaction of naphthalene occurred with the hydrophobic “arms” of the side chain and not the CD cavity although interaction with the cavity was not ruled out.

The sulfopropyl and sulfobutyl ether derivatives (44) have been further evaluated for their complexation with testosterone and progesterone (Fig. 7). Even though increasing the DS should produce more steric hindrance to complexation, the mono, tetra, and hepta substituted sulfobutyl ether (SBE1, 4, 7) derivatives all displayed comparable binding abilities for the steroids and the strength of binding was similar to that observed for β -CD. The SBE substituent behaves similar to that proposed for the undecyl sulfate CD, however, complexation with the SBE- β -CDs involves the CD cavity, as well as the hydrophobic butyl side arms.

Electronic effects

Effect of proximity of charge to CD cavity: The ionic derivatives that have charges closest to the CD cavity are the carboxylate, sulfate, and sulfonate derivatives. The complexation characteristics of the directly carboxylated CDs, C- β -CDs have not been reported but the highly anionic sulfated CD derivative (S14- β -CD) does not appear to form inclusion complexes (45). This may be either due to steric effects from the 14 sulfate substituents or due to the ionic state of the CD.

The effects of charge proximity on CD complexation behavior were evaluated (Fig. 7) by studying the complexation of two steroids by the sulfonate, sulfopropyl ether (SPE), and sulfobutyl ether (SBE) derivative (44). Electronic effects seem to be more of a factor than steric effects because even when only one sulfonate substituent is attached at the 6-position, (6-SA1- β -CD) the derivative

loses its complexation capability. The binding constant for testosterone is only 64 M^{-1} for 6-SA1- β -CD versus $17,800\text{ M}^{-1}$ for the neutral β -CD. The attachment of a single negative charge close to the CD cavity appears to disrupt the thermodynamics driving the complexation.

When one sulfonate ion (SA1) is directly attached to the CD, there is a minimal binding of the steroids but as the charge is spaced away by the three carbon propyl (SPE1) or a four carbon butyl group (SBE1), the derivatives regain the binding capability of the β -CD molecule. The monosubstituted sulfopropyl and sulfobutyl derivatives (SPE1 and SBE1) are able to bind progesterone and testosterone as well as β -CD. This suggests that ionic substituents too close to the CD cavity adversely disrupt the thermodynamics driving the inclusion complexation. Moving the charge away from the cavity re-establishes the complexation characteristics but this is dependent on the charge density in the structure.

Effect of charge density: As the charge density increases in the sulfopropyl family from a mono to a tetra and hepta anion, the binding of the steroids decreases. However, when the sulfonate anion was spaced four methylene units away, the charge density did not adversely affect the binding of the steroids. The mono, tetra, and hepta substituted sulfobutyl ether derivatives all displayed comparable binding abilities for the steroids and the strength of binding was similar to that observed for β -CD.

Effect of charge state of CD and drug: Ionic CDs are capable of complexing neutral hydrophobic drugs, if the ionic charge is not directly attached to the carbohydrate backbone of the CD. The trianion of CM3- β -CD (25) is able to complex a neutral drug, hydrocortisone with an association constant that is 74% of that observed for neutral β -CD. Although this anionic derivative is less effective than the neutral β -CD, a more favorable situation has been observed for the interaction of anionic SBE- β -CDs and neutral drugs. Okimoto et al. (46) reported that the anionic SBE- β -CD (Table 4) often exhibits 1:1 binding constants with neutral drugs that are comparable to or better than those observed for the neutral HP- β -CD. The better binding may be due to the butyl “micellar” arms extending the hydrophobic cavity of the CD.

When the drug and the CD are both charged, electrostatic effects may be observed. Adverse electronic effects have been observed for the complexation between the anionic form of indomethacin and the dianion of carboxymethyl- β -CD, CM2- β -CD (38). At pH 6.6, indomethacin exists as an anion and under these conditions, the anionic carboxymethyl CD did not complex the drug at all, probably due to electrostatic

Table 4 Effect of charge state of drug on 1:1 binding to neutral HP- β -CD and anionic SBE- β -CD

Drug	Neutral drug <i>K_a</i> (<i>M</i> ⁻¹)		Anionic drug <i>K_a</i> (<i>M</i> ⁻¹)		Cationic drug <i>K_a</i> (<i>M</i> ⁻¹)	
	HP- β -CD	SBE- β -CD	HP- β -CD	SBE- β -CD	HP- β -CD	SBE- β -CD
Cinnarizine ^b (46)	22,500	69,700			4,000	17,500
Cinnarizine(1:2) ^b (46)	494	—			6	—
Danazol ^c (47)	76,600	94,900				
Digoxin ^d (48)	4,900	6,880				
Hydrocortisone ^d (48)	1,340	2,150				
Indomethacin ^b (46)	1,590	4,710	955	819		
Kynostatin ^c (49)	95	292			20	96
Kynostatin(1:2) ^c (49)	26	4			3	0
Miconazole ^b (46)	104,000	417,000			42,300	410,000
Miconazole(1:2) ^b (46)	45	12			11	<1
Naproxen ^b (46)	1,670	3,600	331	432		
Papaverine ^b (46)	337	1000			17	94
Phenytoin ^b (48)	1,070	756				
Progesterone ^d (48)	11,200	18,300				
Testosterone ^d (48)	11,600	22,500				
Thiabendazole ^b (46)	136	443			7	56
Warfarin ^b (46)	2,540	10,100	509	262		

^aBinding constants for 1:1 complexation unless noted.^bHP = EncapsinTM MS = 3.5; SBE- β -CD MS = 7.^cHP = Roquette MS = Not reported; SBE- β -CD MS = 7.^dHP = Molecusol[®] MS = 7–8; SBE- β -CD MS = 7.^eHP = Molecusol[®] MS = 7–8; SBE- β -CD MS = 4.

(Adapted from Ref. 6.)

repulsions. However, the tri-anion, CM3- β -CD (25) has been reported to complex the anionic forms of warfarin and indomethacin (Table 5) although only at 71 and 60% of the binding observed for the neutral β -CD.

Experience with the carboxymethyl derivatives suggested the position of the charge in the drug structure may affect the interaction with an anionic CD. The spacing of the charge by the butyl group in the SBE substituent appears to lessen these repulsive effects observed for the

shorter carboxymethyl substituent. The binding constants between the anionic forms of indomethacin, and naproxen and the anionic SBE- β -CD (Table 4) are almost equivalent to those observed for the neutral HP- β -CD. The binding constant between SBE- β -CD and the anionic warfarin molecule, however, is much lower than that with HP- β -CD, suggesting that the position of the charge in the drug and how this interacts with the charge in the CD may be important.

Table 5 Effect of charge state of drug on binding to neutral β -CD and anionic carboxymethyl- β -CD

Drug	Charge state of drug	β -CD (neutral) binding constant (<i>M</i> ⁻¹)	CM3- β -CD (anionic) binding constant (<i>M</i> ⁻¹)
Hydrocortisone	Neutral	6200	4600
Indomethacin	Anionic	620	250
Warfarin	Anionic	520	150
Propranolol	Cationic	220	400

(Adapted from Ref. 25.)

Cooperative electrostatic interaction between the cationic drugs and the anionic CDs have been observed. Enhanced complexation is observed for the complexation of the cationic form of propranolol with the anionic CM3- β -CD (Table 5) and is probably due to cooperative electrostatic interactions. Similar positive interactions are observed with the SBE- β -CD and the cationic forms of cinnarizine, miconazole, papaverine, and thiabendazole (Table 4).

One difference in complexation performance of ionic versus neutral CDs is in their inability to participate in 1:2 or 1:3 complexations. The ionically charged CDs do not effectively form higher order complexes probably due to electrostatic repulsions between the first CD to sequester the drug and the incoming ionic CD. As the charge density increases, this repulsive effect is magnified. Rajewski et al. (50) demonstrated that as the charge density of the SBE- β -CD increases from one to four to seven, the solubilization of cholesterol decreases. Fortunately, the SBE-CDs are able to complex drugs effectively with the 1:1 complexation so the inability to effectively participate in 1:2 complexes does not impose any practical disadvantages.

Temperature, additives, and co-solvent effects

Inclusion complexation is an equilibrium process and the strength of association is affected by the temperature of the system. In most cases, as the temperature increases, the binding constant will decrease. For example, the binding constant for the neutral naproxen molecule (51) and β -CD decreased from 1379 to 975 to 778 M^{-1} as the temperature increased from 25°C to 35°C and 45°C, respectively. The solubility of a drug in the CD solution may increase with an increase in temperature even though the binding constant is decreasing because the increased

temperature improves the intrinsic solubility of the free drug (S_o in Fig. 5) (52, 53).

Organic solvents (54–56) typically reduce the complexation of a drug with CD by competing for the hydrophobic cavity. They also reduce the solubility of most CDs and their complexes. Recently, Loftsson et al. (57) and Redenti et al. (58) have reported on the use of water soluble polymers and hydroxy acids, respectively, to increase CD:drug complexation and improve the solubilizing effect.

Release from the Complex

Complexation of drugs by CDs improves their delivery characteristics and does not interfere with their activity because complexation is a rapidly reversible process. In aqueous solution, drug:CD complexes are continually forming and dissociating with lifetimes in the range of milliseconds or less (59, 60). Although slower kinetics of dissociation are seen with stronger binding, the rates are still fast and essentially instantaneous. After administration, the drug is released from the complex upon dilution, and in some cases with contributions from competitive displacement with endogenous lipophiles, binding to plasma and tissue components, drug uptake into tissues is not available to the complex, and rapid elimination of the CD (61).

The effects of dilution are demonstrated in Fig. 8 (62) for complexes with various binding constants. Most drug:CD complexes exhibit binding constants in the range of 100–20,000 M^{-1} and Fig. 8 demonstrates that even for the more tightly bound drugs, a 1:100 dilution will reduce the percentage of drug complexed from 100% to 30%. A 1:100 dilution is readily attained for

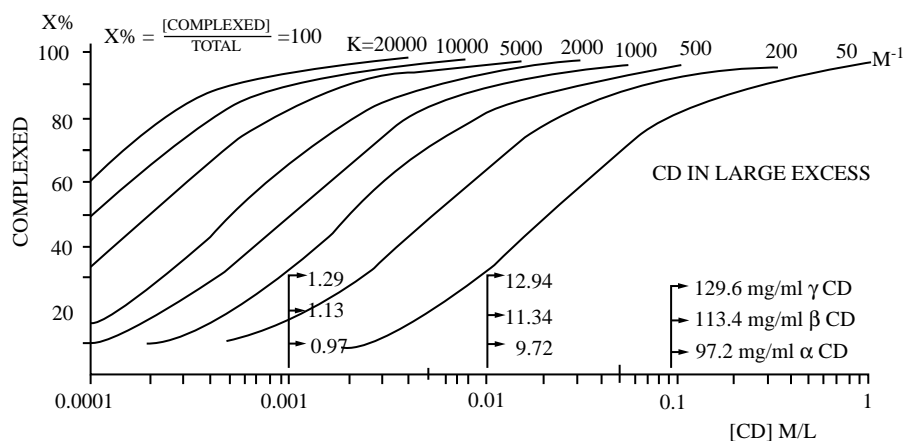


Fig. 8 Correlation between percentage of complexed drug and CD concentration at various K values. (Adapted from Ref. 62.)

intravenous products and upon dilution in the stomach and intestinal contents.

Dilution is minimal, however, when drugs are administered via routes such as ophthalmic, transmucosal, and transdermal. Under these conditions, the drug can still be displaced from the CD cavity by competing lipophiles such as triglycerides, cholesterol, bile salts, and other hydrophobic compounds often found in high concentrations at the site of delivery.

BENEFITS OF COMPLEXATION

Improvement in Solubility, Dissolution, and Bioavailability

CD formulations provide improved aqueous solubility to poorly soluble drugs, and the drug:CD complex often exhibits improved dissolution characteristics compared to other formulations of the drug. These two features can provide for an improvement in oral bioavailability when solubility and the rate of dissolution are limiting the availability of the drug for absorption. For example, the drug cefotiam hexetil hydrochloride forms a gel under the acidic conditions of the gastric contents and shows poor dissolution. A variety of excipients were screened to prevent gelation and α -CD complexation afforded the best formulation for the dissolution and solubilization of the drug (63).

Another example is the calcium channel blocker, cinnarizine. This drug exhibits a very low and erratic bioavailability after oral administration as a suspension ($F = 8 \pm 4\%$) or capsule ($F = 0.8 \pm 0.4\%$). When it was administered as a complex with SBE4- β -CD or HP- β -CD (64), either as a solution ($F = 55\text{--}60\%$) or in a capsule ($F = 38 \pm 12\%$), the bioavailability was significantly enhanced. The improvement in bioavailability was attributed to enhanced dissolution and solubilization via the complexation.

A review of the literature reveals several hundred citations and reviews that describe the effects of complexation on dissolution and bioavailability of drugs. A broad range of CD and CD derivatives have been investigated as well as many different drugs. Some other representative examples are spironolactone (65), meclizine (66), ketoprofen (67), oxazepam (68), danazol (69), phenytoin (70), and tolbutamide (71). Although these studies demonstrate the general application of complexation for improvements in dissolution and bioavailability, the use of complexation may not be practical for some dosage forms due to the amount of CDs required. β -CD for example has a molecular weight of 1135. If one uses a

mole ratio of 5:1 to promote solubility, then over 350 mg of CD will be required for a 25 mg dose of a drug having a molecular weight of 400. This can limit the type and dose of drug that can realistically be used with complexing agents for solid oral dosage forms.

Solution formulations, however, do not typically have these same constraints, and complexation provides an alternative to the use of non-aqueous solvents or large volumes. A few derivatized CDs (e.g. hydroxypropyl and sulfobutyl ether) can be safely administered by parenteral routes. This is often where complexation and its improvements in aqueous solubility can be most readily utilized. The derivatized CDs often can be used to replace cosolvents such as ethanol, polyethylene glycol, and lipids, as well as provide an alternative to the use of emulsions and liposomes. The hydroxypropyl and sulfobutyl ether derivatives are stable in solution and can be readily autoclaved, often improving the heat stability of drugs. There are however, reports of complexation of CDs with anti-oxidants (72) and preservatives (73, 74) with both decreased and increased efficacy (75).

Reduction of Unpleasant Side Effects and Bitter Taste

Improvements in the rate and extent of dissolution of a drug can improve the rate of absorption of the drug. Reducing the contact time between the drug and the tissue mucosa can help minimize tissue irritation produced by drugs. Nonsteroidal anti-inflammatory drugs cause a high incidence of gastrointestinal ulcerative lesions that are a result of both local irritation from the drug and systemic inhibition of prostaglandin synthesis by the drug. CD formulations of naproxen (76), diclofenac (77), and piroxicam (78) cause fewer gastric lesions associated with the acute local tissue irritation than produced by the drug alone. Formulations containing CDs have also shown less irritation than nonCD containing formulations for ophthalmic (79), intravenous (80), and intramuscular (81) administration, and in cellular injury screening tests (82).

Complexation with CDs can also have the effect of reducing the amount of contact with taste receptors. This can be of great benefit in the preparation of oral solutions. Not only are the drugs "masked" from the receptors by inclusion in the CD cavity, but the increased hydrophilicity enables the easier removal of the bitter substance from the receptor surface as well. The apparent concentration of the uncomplexed bitter drug is a function of the complexation constant, the amount of free CD, and the water solubility of the drug (83). Complexation has been used to mask the unpleasant bitter taste of a number of drugs such as oxyphenonium

bromide (84), propantheline bromide (85), clofibrate (86), and acetaminophen (83).

Improvements in Drug Stability

CDs are normally thought of as stabilizing agents in pharmaceutical formulations (87, 88). They have been shown to stabilize drugs to hydrolysis (89) and hydrolytic dehalogenation (90), oxidation (91), decarboxylation, and isomerization (92), both in solution and in the solid state. They can, however, accelerate these same reactions (93, 94). The nature of the stabilization or destabilization depends on the CD used (parent and functional groups of any derivative) and on the position of the guest molecule inside the CD. If the molecule is positioned such that the area of instability is located outside the CD, no effect on stability may be observed. When the position allows interaction of the CD hydroxyls (or derivative functional groups) with a hydrolytically prone site, decreased stability may be observed but if the site is located fully within the CD, enhanced stability usually results.

In the solid state, stabilization of drugs to degradation has been reported for numerous drug including nicardipine (95), colchicine (96), prostaglandin E₁ (97), diclofenac (98), and sulfamethoxazole (99).

Stabilization is not limited to small compounds, as larger molecules such as peptides and proteins can also form complexes that result in enhanced chemical and physical stability (100). The CDs will typically interact with functional groups present on exposed surfaces of the macromolecules and often form multiple complexes (several CDs per molecule). Stabilization against aggregation has been observed for CD complexes in solution with proteins such as ovalbumin and lysozyme (101), carbonic anhydrase (102), and insulin (103), and in the solid state with albumin and gamma-globulin (104). CD complexes have also been investigated as chaperone-mimics (105) in the refolding of denatured proteins (106).

The degree of stabilization/destabilization of a drug complexing with a CD depends not only on the rate of degradation within the complex, but also on the fraction of drug that is complexed (88), and the stoichiometry (107). Increased stability is often observed for compounds having high association constants and those that tend to form higher order complexes.

Reduction in Volatility

Inclusion complexes have been prepared with a number of volatile substances (108, 109) including spices, flavors,

essential oils, and several drugs. CD complexation has been shown to reduce the volatility and improve the stability of many compounds. Examples include lemon oil (110) and other flavoring agents (109), clofibrate (86), isosorbide 5-mononitrate (111), and nitroglycerine (112). In addition, complexation facilitates the handling of products, particularly because they transform liquids to solids. The solid form can also provide certain formulation advantages over liquids such as eliminating the melting point and hardness reduction of suppositories commonly observed when liquids are added (113).

CYCLODEXTRINS

α -, β -, and γ -CDs

CDs, also called Schardinger dextrins, cycloglucans, or cycloamyloses, are α -1,4 linked cyclic oligosaccharides obtained from enzymatic conversion of starch. The parent or natural CDs contain 6, 7 or 8 glucopyranose units and are referred to as alpha-(α -CD), beta-(β -CD), and gamma-(γ -CD) CD, respectively. The chemical structure of β -CD (Fig. 2) shows the cyclic nature of the molecule,

Table 6 Characteristics of α -, β -, and γ -CDs

	α	β	γ
No. of glucose units	6	7	8
Molecular weight	972	1135	1297
Cavity diameter, Å	4.7–5.3	6.0–6.5	7.5–8.3
Solubility @25°C (g/100 mL)			
Water	14.5	1.85	23.2
Methanol	i	i	>0.1
(Aqueous) 50%	0.3	0.3	208
Ethanol	i	i	>0.1
(Aqueous) 50%	>0.1	1.3	2.1
2-propanol	i	i	>0.1
Dimethylsulfoxide	2	35	
Propylene glycol	1	2	
Glycerin	i	4.3	
Solubility in water (g/100 g)			
20 °C	0.90	1.64	1.85
25 °C	1.27	1.88	2.56
30 °C	1.65	2.28	3.20
35 °C	2.04	2.83	3.90
40 °C	2.42	3.49	4.60
45 °C	2.85	4.40	5.85
50 °C	3.47	5.27	
55 °C		6.05	

(Adapted from Ref. 7.)

and the presence of three hydroxyl groups on each glucopyranose unit. Two of the hydroxyls are secondary alcohols and are located at the C-2 and C-3 positions of the glucopyranose unit. The third hydroxyl is a primary alcohol at the C-6 position. The hydroxyls provide the hydrophilic exterior responsible for the aqueous solubility (Table 6) of the CDs.

In three dimensions, this structure forms a truncated cone where the primary hydroxyl groups are located on one face and the secondary hydroxyl groups on the other. The interior of the cone is hydrophobic due to the presence of the glycosidic ether oxygens at O-4 and the hydrogens attached to C-3 and C-5, and thereby provides a cavity for the inclusion of hydrophobic compounds. The cavity varies in size with α -CD being the smallest at about 5.3 Å across and γ -CD the largest at 8.3 Å diameter (Table 6).

Properties in solution

The solubilities of the natural CDs in water varies and is quite dependent on temperature (Table 6). The unusually low water solubility of β -CD is due to the very rigid structure that results from the H-bonding of the C-2 hydroxyl of one glucopyranose unit with the C-3 hydroxyl of an adjacent unit (114). In the β -CD molecule, a complete set of seven intramolecular H-bonds can form, effectively limiting interactions with the solvent. This "belt of H-bonds" is incomplete in the other parent CDs thus allowing more favorable interactions between α - and γ -CD and water molecules. This is consistent with the observation of a less favorable enthalpy and entropy of dissolution (115) for β -CD versus α - and γ -CD. Recent studies have suggested that the abnormally low water solubility of β -CD may be exacerbated by aggregation of these rigid β -CD molecules (116). The solubility of β -CD can be increased by disrupting this aggregation through the addition of solvent structure-altering substances such as urea (117), inorganic salts (118), and hydrophilic polymers (119).

Solubility of the CDs is low in most organic solvents (Table 6). In aqueous/organic cosolvent systems, the solubility decreases as the organic concentration increases, with the exceptions of ethyl and propyl alcohol where a maximum is observed at around 30% alcohol (120).

In solution, the CDs are fairly stable to hydrolysis in alkaline medium. Under acidic conditions, the α -1, 4 glycosidic bonds are slowly broken to open the ring, and then to give glucose and a series of linear maltosaccharides. The initial opening of the ring is a slower process by about 2–5-fold than the subsequent hydrolysis of the linear dextrins. The initial ring-opening kinetics are the most important for pharmaceutical

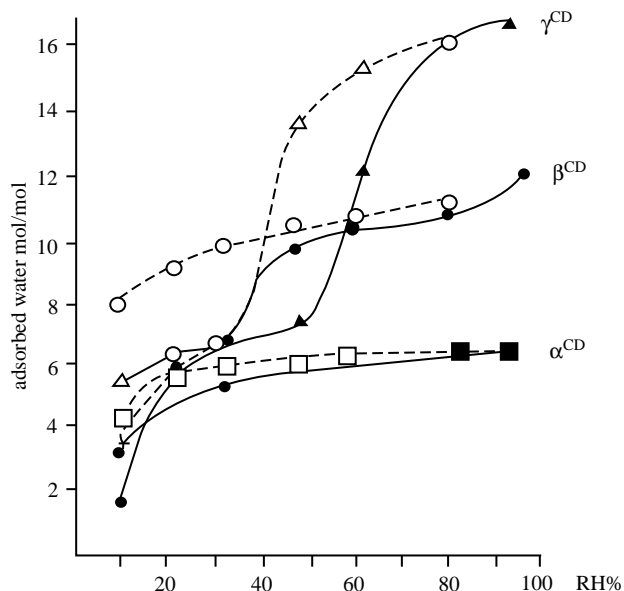


Fig. 9 Water vapor sorption isotherms for α -, β -, and γ -CD at 40 °C. (Dashed line: adsorption, solid line: desorption). (Adapted from Ref. 7.)

preparations as complexation requires the intact cyclic structure. Half-lives for the ring opening reaction step at 70°C and 0.2 M HCl are 25.2, 14.5, and 7.1 h for α -, β -, and γ -CD, respectively (121). Additional pH/rate data are available in the literature (122).

Similar reaction products are observed with gamma-irradiation in solution (123), but in the crystalline solid state, the mechanism appears to be different and no glucose is formed.

Solid state properties

The three natural CDs form crystalline structures in the solid state that decompose above 200°C with no definite melting points. They are not considered hygroscopic, but they do form various stable hydrates. The water vapor sorption isotherms (Fig. 9) show two phases for the β - and γ -CDs, and one phase for α -CD. At 11% RH, α -CD absorbs 4 water molecules and upon long-term storage, forms a stable hydrate with 6 water molecules. The water content gradually increases with increasing humidity to a constant value of 6.6 water molecules per CD molecule at and above 79% RH (124). Four different crystalline forms of α -CD have been reported; two forms containing approximately 6 water molecules, one form containing 7.6 water molecules and a dehydrated form.

The water content of β -CD increases with increasing humidity and passes through a plateau region at about 23–31% RH where the water content is about 5–6

water molecules. Another leveling of the plot occurs at humidities of 60–79%. Both 12-water (125) and 11-water (126) hydrated crystalline forms have been reported along with a dehydrated form. Upon standing for several weeks, the 11-water form will convert to the 12-water form, which is stable over a large range of humidity conditions.

X-ray diffraction patterns of γ -CD stored under various humidity conditions also show the existence of three distinct crystalline forms. A dehydrated form is observed at low humidities and a hydrated form containing almost 17 water molecules occurs at 93.6% RH. An intermediate crystalline form containing 7 water molecules is found at intermediate RH values which corresponds to the plateau region at 20–30% RH in the sorption isotherm. The hydrate and dehydrate forms pass through the intermediate form during dehydration, and hydration respectively (124).

CD Derivatives

Hundreds of modified CDs have been prepared and shown to have research applications. However, only the derivatives containing the hydroxypropyl (HP), methyl (M), and sulfobutyl ether (SBE) substituents are in a position to be used commercially as new pharmaceutical excipients. These substituents vary in size and electronic character and are attached to the CD structure through reaction with one or more of the three hydroxyl groups of the glucopyranose units. The parent CDs contain 18 (α -CD), 21 (β -CD), or 24 (γ -CD) hydroxyl groups that are available for modification. The most reactive hydroxyls are in the C-6 position and the C-3 hydroxyls are the least reactive. However, the difference in reactivity is not great, and changing reaction conditions can often alter the position of substitution. The preparation of homogenous, selectively derivatized CDs is, therefore, not an easy task. With all the options available for positional and regioisomers to be formed, one must be careful in describing the various derivatives. A discussion of nomenclature is provided earlier.

The main derivatives under development as excipients are all derivatives of β -CD: 1) A randomly methylated derivative with an average MS of 14 (M14- β -CD), 2) Two different 2-hydroxypropyl derivatives, one with an average MS of approximately 3 ((2HP)3- β -CD) and the other with an average MS of 7 ((2HP)7- β -CD), and 3) A sulfobutyl ether derivative with an average MS of 7 (SBE7- β -CD). Glucosyl and maltosyl CDs (127, 128) which contain a mono- (G_1 - β -CD) or disaccharide (G_2 - β -CD) substituent, have also been reported and show promise for the future.

Methylated

Methylation can be controlled to produce mono- to fully derivatized CDs. The introduction of the methyl substituent dramatically improves the water solubility of the derivative versus the parent CD. Aqueous solubility increases as the number of methyl groups reaches 14 and then decreases as substitution approaches 21. The 2,6-DM14- β -CD and the 2,3,6-TM21- β -CD have solubilities of 57 and 31 g/100 ml, respectively, versus 1.8 g/100 ml for the parent β -CD. The introduction of the methyl groups disrupts the “belt of H-bonds” effectively increasing the polarity of the derivative.

The aqueous solubility of these derivatives is adversely affected by temperature, however, and precipitation occurs during heat sterilization. The mixture of randomly methylated β -CD (M14- β -CD) (129), however, exhibits a favorable water solubility (>50 g/100 ml) that increases as temperature increases (130).

The extent of methylation is also important in optimizing complexation. The introduction of the methyl substituent at the 2- and 6- positions appears to improve complexation. Binding constants for 2,6-DM14- β -CD with many drugs is an average five times of that observed with β -CD. The methyl groups seem to increase the hydrophobicity of the CD cavity possibly by providing an “extension” of the cavity. Derivatization of the remaining C3 hydroxyls, however, results in a dramatic decrease in complexation ability. This may result from the distorted cyclic structure formed when the CDs are permethylated (131). The altered conformation also impacts the stability of the derivative in acidic solutions. Degradation half-lives of 2.1 and 12.0 h have been reported for a randomly methylated (2, 3, 6) M14- β -CD and a 2,6-DM- β -CD, respectively, in 1 M HCl at 60°C (129). Under similar conditions, β -CD has a half-life of 5.4 h.

The mixture of randomly methylated β -CD, although partially derivatized at the 3-position, still maintains the favorable binding characteristics of 2,6-DM14- β -CD. One report (129) demonstrated that M14- β -CD solubilized 26 drugs more effectively than β -CD and the extent of solubilization was on average 80% of that observed for the purified 2,6-DM- β -CD preparation.

Studies suggest that an optimal definition for a commercially viable methylated CD is the partially methylated β -CD (M14- β -CD) containing an average MS of approximately 14 with the substituents at the 2-, 3-, and 6-positions. This material is produced economically, has an aqueous solubility that increases with temperature, and has binding constants higher than those observed with the unsubstituted β -CD and close to those observed with the 2,6-DM- β -CD.

Hydroxypropyl

Hydroxy alkylation of β -CD requires treating base-solubilized β -CD with the appropriate epoxide or haloalcohol (132, 133). Propylene oxide or propylene carbonate are used in the preparation of 2-hydroxypropyl β -CD ((2HP)- β -CD), the derivative being commercialized. The reaction occurs at both primary and secondary alcohols on the β -CD generating a mixture of numerous isomeric forms (134, 135). This results in a heterogeneous product that is amorphous and highly water soluble.

The 2-hydroxypropyl derivative has been the subject of numerous clinical trials and is commercially available from several suppliers. Brandt (136), Müller (137, 138), and Pitha (139) have described its preparation and use. The DS can affect the ability of the hydroxypropyl derivatives to form complexes. It can also affect the solubility of the derivatives. The mono substituted derivative, (2HP)1- β -CD is actually less soluble than β -CD (140). However, at degrees of substitution of ~ 2.7 and higher, the solids are amorphous and exhibit solubilities in excess of 50% w/v (135). Water uptake by the solid forms is low. At 75% RH and 25°C, the (2HP)- β -CDs show less water uptake than the parent β -CD (135), and the water uptake decreases with increasing MS.

As discussed earlier, the need to control the DS becomes important to balance water solubility and complexation capability. Two commercial preparations of (2HP)- β -CD, EncapsinTM and Molecusol[®], have recognized the need for this compromise and have substitution levels that provide a balance between solubility and complexation. Encapsin and Molecusol have MS values of approximately 3 and 7, respectively. Although both (2HP)- β -CD commercial preparations are unique, each manufacture can reproducibly generate materials to meet defined specifications. These (2HP)- β -CD derivatives appear to be equally effective in complexation and have water solubilities exceeding 50% w/v. Both have been administered parenterally.

Sulfobutyl ether

Rajewski (48) prepared the directly sulfonated CDs through the introduction of the sulfonic acid moiety at the C-6 position. These anionically charged sulfonic acid substituents were spaced away from the CD with alkyl groups by Parmeter (141) and Lammers (142) in the preparation of sulfopropyl derivatives of CDs.

Stella and Rajewski (44) later described the preparation of sulfoethyl through sulfohexyl derivatives of the CDs. The sulfonate and sulfoalkyl ether derivatives can be prepared with different average

degrees of substitution (143), are isolated as the sodium salts, and demonstrate water solubilities independent of the MS. Likewise, no effect on complexation is seen with changes in MS when the alkyl spacer is butyl (Fig. 7). The SBE- β -CD derivatives are amorphous and similar to HP- β -CDs, tend to form amorphous complexes. They are highly water soluble (>50 mg/ml), and somewhat hygroscopic, reversibly picking up water at humidities below RH 60%.

The SBE7- β -CD derivative has been used in clinical trials and is being developed commercially as Captisol. It is well characterized and suitable for parenteral administration.

ABSORPTION, DISTRIBUTION, METABOLISM, AND EXCRETION

Oral Pharmacokinetics

The parent CDs are poorly absorbed from the gastrointestinal tract. Reported values for absorption range from 0.1 to 0.3% (144) for rats fed a diet containing 5–10% β -CD, to $\sim 2\%$ (145) when the doses were administered in an isolated rat ileum closed-loop experiment. When ^{14}C β -CD was administered orally, values as high as 4.8% have been reported for appearance of the label in the urine (146). This higher value was attributed to absorption of the metabolites of β -CD. The small amount of intact CDs absorbed orally probably does so by passive means (147), via the paracellular route (148). Oral absorption studies with α - and γ -CD have shown $\leq 2\%$ and $\leq 0.1\%$ absorption, respectively (149, 150).

The majority of an orally administered dose of α - and β -CD will be metabolized in the colon. This has been demonstrated both in rats (146) and man (151) with very little hydrolysis occurring in the upper gastrointestinal tract (GIT). Microbiological studies (152) have shown that most of the human colonic bacterial strains can degrade α - and β -CD and this activity can be stimulated by as little as 2–4 h of exposure to the CDs. The typical 40 h transit time through the human colon provides adequate time to induce the bacterial enzymes to provide for complete hydrolysis of the CDs in the colon. Likewise, most of an oral dose of γ -CD is metabolized in the GIT. However, studies with radiolabelled γ -CD suggest that most of its metabolism occurs in the upper GIT (153).

The derivatized CDs are generally more resistant to hydrolysis in the GIT than the parent CDs. Oral bioavailability of HP- β -CD in dogs is estimated at 3.3% and is less in rats, and about 60% of the dose is excreted

Table 7 Pharmacokinetic parameters for several CDs

CD	Species	$t_{1/2,\alpha}$ (min)	$t_{1/2,\beta}$ (min)	Vd_{ss} (mL/kg)	CL_T (mL/h/kg)	Ref.
β -CD	Rat	1.5–2.9	23.9–50.2	152–176	204–372	(158)
γ -CD	Rat		20			(153)
(G ₁ - β -CD)	Rabbit			191	283	(159)
(G ₂ - β -CD)	Rat	4.3	31.1	534.6	979.4	(127)
HP β -CD	Rat		24		512	(154)
(MS = 2.7)	Dog		48		188	(154)
HP β -CD	Human		72–108	164–240	96–126	(160)
DM- β -CD	Rat		22.7–42.3			(156, 161)
S- β -CD	Rabbit			144	32	(159)
(MS = 9.6)						
S- β -CD	Rabbit			172	47	(159)
(MS = 17.6)						
S- β -CD	Rat			113	52	(159)
(MS = 13.3)						
SBE- β -CD	Rat		18	300	588	CyDex unpublished
(MS = 7)	Dog		66	400	282	CyDex unpublished
	Man		84	185	114	CyDex unpublished

unchanged (154). Oral absorption in humans has not been observed (150). Oral administration of ^{14}C HP- β -CD to rats results in approximately 3% of the radiolabel appearing in the urine, 71% in the feces, and 3% being exhaled (155).

The methylated derivatives have shown somewhat greater oral absorption. The absorption of DM- β -CD has been reported as 6.3–9.6% in the rat (156), and M- β -CD as 0.5–11.5% (153).

Parenteral Pharmacokinetics

Intravenously administered CDs disappear rapidly from the systemic circulation and are excreted mainly through the kidney. α - and β -CD are excreted almost completely in their intact form, but some metabolism is observed with γ -CD. Reports vary with regard to the amount of metabolism from “substantial” (157) to 10% or less (153). The hydrophilic CD derivatives are likewise rapidly cleared following intravenous administration and most are excreted unchanged in the urine. Linear, two compartment pharmacokinetics are usually observed although the initial distribution kinetics are very rapid and may not always be captured.

The disposition parameters for several CDs are given in Table 7. The steady state volumes of distribution (Vd_{ss}) correspond well with extracellular fluid volume in each species evaluated, suggesting little or no distribution of most CDs into other tissues or storage compartments.

Studies with ^{14}C HP- β -CD have shown that the small amount that does distribute, has been found mainly in the kidney and lungs of rats following single intravenous doses, and in the kidney and liver of dogs after chronic (1 month) intravenous dosing (154). The total plasma clearances (CL_T) are dose independent and are indicative of clearance at a rate comparable to glomerular filtration (158, 160). Thus, as with any compound whose elimination is closely tied to kidney function, linear pharmacokinetics may not always be observed in the presence of poor renal function.

SAFETY OF CD

Oral Safety

The oral safety of the parent β -CD was first reported in 1957, and it was erroneously suggested that the material was unsafe (162). Subsequent studies by Anderson et al. (163) and Gerlóczy (164) demonstrated that α - and β -CD produced no toxic effects when fed to rats for 30–90 days at 1% of the diet or at 1 and 2 g/kg daily doses. The odd, irreproducible results of the first report were probably due to the inconsistent purity of early CD materials and the possible presence of residual organic solvents.

Both rodent and nonrodent studies have been conducted on the parent CDs. Szejtli and Sebestyén (165) reported the parent CDs to be nontoxic at very high oral doses. Mortality was not observed, even in animals treated with

the highest possible oral doses. Therefore, the LD₅₀ in rats is reported to be greater than 12.5, 18.8, and 8 g/kg body weight for α -, β -, and γ -CD, respectively.

In general, the oral administration of α -, β -, and γ -CD appears to cause several changes reflective of an adaptation to a diet containing a poorly digestible carbohydrate. The changes are species dependent, with rats being more susceptible than dogs. In both cases, the effects are reversible upon cessation of treatment.

α -, β -, and γ -CD

The safety of orally administered β -CD has been investigated in numerous studies (120, 144, 166, 167) with extensive evaluation of hematology, blood chemistry, urinalysis, and necropsy (macro and microscopic). No significant toxic effects were observed in any of these studies after oral administration of β -CD to mice, rats, or dogs.

Although no macroscopic pathologies were observed, microscopic evaluation of the tissues revealed several treatment-related changes from the 1-year exposure of rats to β -CD (167). The organs affected by the treatment were the kidneys and the liver. The kidney effects were not thought to be of any toxicological importance.

Some cellular necrosis was observed in the liver of male rats receiving a 5% β -CD diet and in female rats receiving a 2.5 and 5.0% β -CD diet. An increase in portal inflammatory cell infiltration was also observed in male rats receiving the 2.5% β -CD diet and male and female rats receiving the 5.0% β -CD diet. These observations were considered to represent a mild hepatotoxicity (mechanism unknown), which was further evidenced by increases in serum liver enzymes.

The 1-year exposure (167) of dogs to β -CD diets did not result in the kidney or liver pathologies observed in the rats. Therefore, the mild hepatotoxicity may be species related and not reflective of a general hepatotoxicity. Dogs fed 5% β -CD for 1 year exhibited increased urinary protein levels and the urinary excretion of calcium. However, these changes were not noted in the rat study.

From the results of the 1-year studies, the nontoxic effect levels for oral use of β -CD are considered to be 1.25% of the diet for rats and 5% for dogs. Considering the quantity of food that was consumed under these conditions, this is equivalent to approximately 760 and 1899 mg/kg/day for rats and dogs, respectively.

α - and γ -CD show similar oral safety profiles to those observed for β -CD. Ninety-day feeding studies in rats and dogs (153) consuming diets containing α -CD or γ -CD have shown effects that are consistent with the consumption of a poorly digestible carbohydrate such as β -CD or lactose. Some increases in organ weights (spleen and male

Table 8 Reported oral safety studies with HP3- β -CD(198) and SBE7- β -CD

Species	Dosing duration (days)	Doses (mg/kg/day)
SBE7- β -CD (Captisol: CyDex)		
Rats	1	600
HP3- β -CD (Encapsin: Janssen)		
Mice	1	5000
	90	500, 2000, 5000
	90	500, 2000, 5000
	730	500, 2000, 5000
Rats	1	5000
	14	5000
	365	500, 2000, 5000
	730	500, 2000, 5000
Dogs	1	5000
	365	500, 1000, 2000

adrenals) have been observed but were reversible. The ingestion of γ -CD for 13 weeks at dietary levels of up to 20% (corresponding to intakes of 11.4 and 12.7 g/kg body weight/day for male and female rats, respectively) has been shown to be well tolerated (168).

The treatment of dogs with 0, 5, 10, and 20% α -CD and γ -CD diets resulted in minimal effects as compared to those observed in rats (153). A subsequent study concluded that daily γ -CD consumption of up to 20% in the diet (approximately 7.7 g/kg body weight in male and 8.3 g/kg body weight in female dogs) is tolerated without any toxic effects (169).

CD derivatives

Oral safety studies have been conducted with at least two derivatives, the HP3- β -CD and SBE7- β -CD. The reported studies for these derivatives are listed in Table 8. The oral safety of HP3- β -CD has been assessed in mice, rats, and dogs for dosing periods up to 2, 2, and 1 year, respectively. Doses reached as high as 5000 mg/kg/day. No adverse effects were noted except for an increase in diarrhea in dogs treated with 5000 mg/kg. The 2-year carcinogenicity studies are discussed separately later.

The oral safety of SBE7- β -CD derivative is currently under evaluation.

Parenteral Safety

The most encompassing test of an excipient's safety is the systemic safety of the material because many of the routes of administration ultimately result in at least some minor systemic exposure. Numerous studies with the parent CDs have shown that their parenteral toxicity is observed primarily as renal and cytotoxicity (hemolysis and tissue

irritation). These toxicities were the driving force for the preparation of new CD derivatives, many of which exhibit improved parenteral safety.

α -, β -, and γ -CD

Renal issues: The parent CDs can all show a toxic effect on the kidney when given parenterally. The nephrotoxicity of α - and β -CD manifests itself as a series of alterations in the organelles of the proximal tubule cells (170). The toxicity is initially expressed as an increase in apical vacuoles, which is typical of an adaptive response to the excretion of osmotic agents at extremely high concentrations. This effect reverses upon discontinuation of CD administration. However, there are also other cellular changes not typical of osmotic agents that are not reversible. Giant lysosomes appear and prominent acicular (needle-like) microcrystals are observed in the epithelial cells of the renal proximal tubules. Both the occurrence and abundance of the microcrystals are dose dependent. The content of the crystals has not been confirmed but suggestions include precipitated parent CD (unlikely for α -CD with a solubility of 145 mg/ml), and complexes of CDs with cholesterol (171) or lipoproteins (172). The proximal tubules progressively show dramatic alterations in other organelles. The mitochondria are observed to swell and become distorted. The Golgi apparatus is affected along with the smooth endoplasmic reticulum. The interstitial membrane on the basal lateral side of the cell is disrupted. All of these events are irreversible, and as the toxic condition progresses, kidney function is lost and death ensues.

Parenterally administered γ -CD appears to be less nephrotoxic than α - or β -CD. Subcutaneous and intravenous doses as high as 4000 mg/kg in mice and 2400 mg/kg in rats have shown no toxic effects (173). Schmid (174) reported that the intravenous LD₅₀ for γ -CD in mice was 10,000 mg/kg and >3750 mg/kg for rats. For acute intravenous administration, γ -CD has been shown to be safer than α - and β -CD, which exhibit LD₅₀ values of 1000 and 788 mg/kg, respectively, in the rat (170, 174). Antlisperger (153), and more recently Donaubaue (175), evaluated the intravenous administration of γ -CD to rats for 30 and 90 days. A no adverse effect level (NOAEL) of 200 mg/kg was reported for the 30-day studies and 120 mg/kg was suggested for the 90-day studies.

Cytotoxicity issues: In-vivo hemolysis has been observed with parenteral administration of all of the parent CDs. In-vitro studies with human erythrocytes have demonstrated that the damaging effect of the CDs is in the order β -CD > α -CD > γ -CD (176). This cellular destruction has also been observed in studies with human skin fibroblasts and intestinal cells (177), P388

murine leukaemic cells (178), *E. coli* bacterial cells (179), and immortalized human corneal epithelial cells (180). Mechanistic studies suggest that CDs extract either cholesterol (β -CD and γ -CD) or phospholipids (α -CD) from the cell membrane causing small pores which allow leakage and eventually lead to cell lysis.

These in vitro cytotoxicity studies are not indicative of in vivo toxicity but rather provide a method to classify the CDs for their potential to destabilize or disrupt cellular membranes. In fact, when whole blood is used instead of erythrocytes for the hemolysis tests, the cytotoxicity of the CDs is diminished 10-fold by the presence of hydrophobic serum components. Thus, the membrane damaging effects of the CDs are observed in vivo only under situations of high concentrations.

CD derivatives

Renal issues: The derivatized CDs vary widely in their potential for renal safety. Renal nephrosis was observed for methylated β -CDs following intramuscular injections of as little as 50 mg/kg/day over 12 days (181). The damaging effect of the methylated CDs was in the order of TM- β -CD > M- β -CD > DM- β -CD > β -CD. An LD₅₀ value of 220 mg/kg has been reported for DM- β -CD (159). Administration of the maltosyl/dimaltosyl derivatives G₂- β -CD/(G₂)₂- β -CD on the other hand, showed no toxic effects on the kidney of rats at intravenous doses of 200 mg/kg for 14 days (182). There was however, a flushing of the eyes, nose, mouth and extremities observed, prompting additional investigation.

Safety of the hydroxypropyl and sulfobutyl ether derivatives has been studied in considerable detail and little or no renal toxicity has been demonstrated at moderate doses. Summaries of available intravenous safety studies are given in Table 9. Parenteral exposure with either derivative results in the osmotic adaptive response seen with β -CD but further progression to the irreversible damage does not occur.

Ninety-day intravenous dosing of (2HP)3- β -CD at 400 mg/kg resulted in moderate toxicity as evidenced by decreases in body weight gains, changes in blood and serum parameters, increased activity of mononuclear phagocytosing cells of the lungs and liver, and an increase in the red pulp hyperplasia in the spleen (183).

The evaluation of the sulfobutyl ether derivative, SBE7- β -CD for 6 months with daily intravenous dosing up to 600 mg/kg did not present evidence of the effects noted with (2HP)3- β -CD and demonstrates the extensive systemic safety of this CD.

Cytotoxicity issues: As with renal toxicity, the various derivatives show dramatically different hemolytic behaviors. The dimethyl derivative shows substantial

Table 9 Reported intravenous safety studies with HP3- β -CD(198) and SBE7- β -CD

Species	Dosing duration (days)	Doses (mg/kg/day)
SBE7- β -CD (Captisol: CyDex)		
Mice	1	2000
Rats	1	600
	1	2000
	14	160, 240, 600, 1500, 15000
	30	40, 80, 160
	30	160, 240, 320
	30	300, 1000, 3000
	180	200, 320, 600
Dogs	1	240
	14	160, 240, 750
	30	30, 60, 120
	30	100, 200, 300
	30	300, 1000, 3000
	180	150, 300, 600
HP3- β -CD (Encapsin: Janssen)		
Mice	1	5000, 7000, 10000, 14000, 20000
Rats	1	2000, 4000
	4	1600, 3200
	10	400
	90	25, 50, 100, 400
	90 ^a	50, 100, 400
Dogs	1	5000
	4	3200
	90	25, 50, 100, 400
		50, 100, 400

^aTwo 90-day studies were conducted at these levels.
(Adapted from Ref. 198.)

hemolysis; more than even the parent β -CD. This is well illustrated in Fig. 10 where the percentage of cells undergoing hemolysis is shown as a function of CD concentration. Hemolysis started at concentrations below 0.1% for the DM- β -CD. Four to five times higher concentrations of β -CD are required to give the same hemolysis. This behavior is in agreement with the demonstration of DM- β -CD as a penetration enhancer in skin (184) and nasal tissue (185).

The hydroxypropyl and sulfobutyl ether derivatives, on the contrary, show much less hemolysis than β -CD. This is shown in Fig. 11 where the hemolysis caused by β -CD is compared to two (2HP)- β -CDs and three SBE- β -CDs (186). The hemolysis profiles show a dependence on the MS for the derivatives, showing less effect with higher MS. The two hydroxypropyl derivatives, (2HP)3- β -CD and (2HP)7- β -CD, are almost equivalent in their hemolytic behavior and are both less hemolytic than

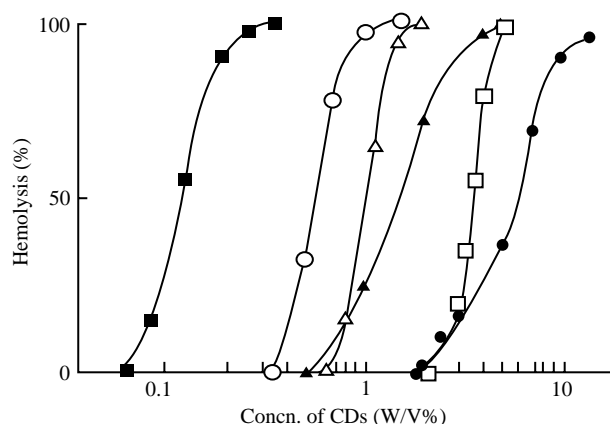


Fig. 10 Hemolytic effects of CD derivatives on human erythrocytes in isotonic phosphate buffer (pH 7.4) at 37 °C for 30 minutes. (Δ , α -CD; \circ , β -CD; \square , γ -CD; \blacksquare , DM- β -CD; \blacktriangle , HP- β -CD; \bullet , HE- β -CD.) (Reprinted from Ref. 135.)

β -CD. Likewise, the sulfobutyl ether derivatives are less hemolytic than β -CD, but the effect of MS is quite dramatic. As the MS increases from one to four to seven, the hemolytic activity drops precipitously to where essentially no hemolysis is observed with SBE7- β -CD.

Mutagenicity and Carcinogenicity

The potential for interaction with genetic material (and therefore risk of carcinogenicity) can be investigated using bacterial and mammalian gene mutation assays and

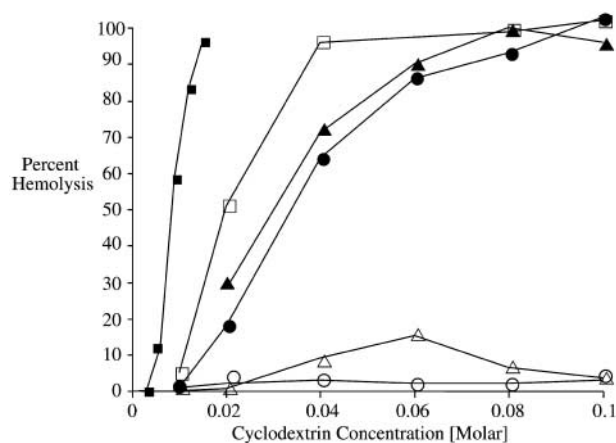


Fig. 11 Hemolytic effects of CD derivatives on human erythrocytes in isotonic phosphate buffer (pH 7.4) at 37 °C for 5 minutes. \blacksquare , β -CD; \square , SBE1- β -CD; \bullet , (2HP)3- β -CD; \blacktriangle , (2HP)7- β -CD; \triangle , SBE4- β -CD; \circ , SBE7- β -CD. (Adapted from Ref. 186.)

chromosomal aberration assays. The parent CDs do not exhibit mutagenic behavior in any of these assays (153, 165), and there have been no reports of tumors in oral feeding studies or in the parenteral administration of any of the parent CDs.

Several of the CD derivatives have also been evaluated for mutagenicity and carcinogenicity. Both HP- β -CD (187, 188) and SBE7- β -CD (189) show negative results for the mutagenicity tests. However, a 2-year carcinogenicity study of HP- β -CD in rats demonstrated hyperplastic and neoplastic changes in the acinar cells of the exocrine pancreas (190). The neoplasia in the rat study is inconsistent with the mutagenicity assay results and with the lack of carcinogenicity of the parent CDs. In separate and shorter studies with mice and dogs, no adverse effects were observed for the pancreas.

The rat pancreatic hyperplasia may be due to the ability of high concentration of HP- β -CD to increase the fecal elimination of bile salts indirectly causing a stimulation of the production of cholecystokinin (CCK). In the rats, CCK functions as a mitogen causing an increase in the cellular hyperplasia in the acinar cells. Sensitivity to this effect is species dependent (191), the rat is most sensitive and dog show no effects. The carcinogenicity study for HP- β -CD may have been conducted at levels that are affecting an important nutritional balance. The FDA guidelines for carcinogenicity studies suggest that safety studies be conducted with the highest levels possible to determine maximum tolerated doses but care should be taken to minimize possible nutritional deficiencies (192). The observation of pancreatic neoplasms observed with the 5 g/kg/day oral doses of HP- β -CD may have been the consequence of a nutritional deficiency not a carcinogenic effect of HP- β -CD itself.

Reproductive Safety

In oral-safety studies involving both male and female animals, some minor differences were observed between the sexes. The parent CDs, however, do not adversely affect either gender and the effect of CDs on reproduction is minimal (166). Embryotoxicity and teratogenicity studies have been reported for α - and γ -CD (193). Several 90-day feeding studies in rats and rabbits have been conducted with no effects being observed for maternal health or reproduction (153).

A more extensive evaluation of reproductive and developmental safety of β -CD was reported in a three-generation study by Barrow et al. (194). The only adverse effect observed during the study was a dose related decrease in pup weight gain from birth until weaning but

this was statistically significant only for the 5% β -CD diet during days 7–14 postpartum. This preweaning growth retardation did not result in any permanent defects and the affected pups returned to normal weights upon weaning. The NOAEL for oral β -CD, under the conditions of the study, was suggested to be at 1.25% dietary β -CD.

Reproductive safety studies have been conducted for both SBE7- β -CD and HP3- β -CD in rats and rabbits. A listing of the reported studies is given in Table 10. Oral administration of up to 5000 mg/kg HP- β -CD to pregnant rats produced no maternal toxicity, embryotoxicity, or teratogenicity. Oral administration of 1000 mg/kg HP- β -CD to pregnant rabbits caused a slight maternal and embryotoxicity but no teratogenicity.

Intravenous administration of HP- β -CD at 400 mg/kg to the dams from day 18 of the pregnancy to 3 weeks of lactation produced no adverse effects on the rat pups. However, when the dosing occurred from day 16 of gestation to week 3 of lactation, the low dose (50 mg/kg) and the high dose (400 mg/kg) presented significantly lower pup survival than the vehicle control groups.

These effects were not observed with the intravenous administration of SBE7- β -CD at doses of 100, 600, and 3000 mg/kg to pregnant rats. There were no effects of intravenous treatment with SBE7- β -CD on fertility or early embryonic development, nor was the material observed to be teratogenic. The only effect of treatment was a decrease in maternal body weight gains and food consumption at the highest doses administered.

REGULATORY STATUS

Regulatory Process for New Excipients

CDs are not “standard” inactive ingredients, and their uncertain regulatory status causes hesitancy in their use in formulations. A common perception exists that an approval process is in place for the evaluation of new excipients, such as the CDs. In fact, there is no mechanism for submission and review of data on a new excipient that would lead to approval of that excipient. In the United States, the FDA reviews a new excipient only in relationship to the review of a drug formulation. Only the final drug product is approved by the FDA. By this method the excipient data are reviewed with each drug application. The dossier on a new excipient is filed by the excipient manufacturer as a Drug Master File (DMF)-Type 4 (195). These data are then referenced when an Investigational New Drug application (IND) or New Drug Application (NDA) is filed for a drug dosage form using the excipient.

Table 10 Reported reproductive safety studies with HP3- β -CD(198) and SBE7- β -CD

Species	Route	Doses (mg/kg/day)
<i>SBE7-β-CD (Captisol: CyDex)</i>		
Maternal range finding toxicity		
Rats	i.v.	300, 1000, 3000
Rabbits	i.v.	250, 600, 1500
Segment I: Fertility & early embryonic development		
Rats	i.v.	100, 400, 1500
Segment II: Embryotoxicity & teratology		
Rats	i.v.	100, 600, 3000
Rabbits	i.v.	100, 400, 1500
Segment III: Pre- & post-natal development		
Rats	i.v.	100, 600, 3000
<i>HP3-β-CD (Encapsin: Janssen)</i>		
Segment I: Fertility & early embryonic development		
Rats	i.v.	50, 100, 400
	Oral	500, 2000, 5000
Segment II: Embryotoxicity & Teratology		
Rats	i.v.	50, 100, 400
	Oral	500, 2000, 5000
Rabbits	Oral	50, 100, 400
	Oral	250, 500, 1000
Segment III: Pre- & post-natal development		
Rats	i.v.	50, 100, 400
	i.v.	50, 100, 400
	Oral	500, 2000, 5000

(From Ref. 198.)

A petition can also be made for approval as a food additive and to be placed on the GRAS (generally regarded as safe) list. The GRAS list (21 Code of Federal Regulations 182.1–184.1) actually applies only to food additives that are reviewed by the FDA and determined to be generally recognized as safe for the purpose and use conditions described in the statute. The use of GRAS excipients is often but not always transferable to oral pharmaceutical formulations. Once the material is approved for use in foods, the material may be considered suitable for use in an oral formulation if the dose fits within the quantities consumed as a food additive. This suitability does not, however, transfer to non-oral routes.

The process is similar in Japan, in that a new excipient's dossier is evaluated in reference to an application for a drug dosage form containing the excipient. The data is evaluated both in terms of the excipient and the active, but only the drug product is approved. After the excipient has seen extensive utilization in multiple marketed products, the regulatory system has a process for review of the data

resulting in possible inclusion of the monograph in the *Japanese Pharmacopoeia* (JP). The JP defines the mandatory standards for substances used in a pharmaceutical product. Inclusion in the JP establishes “precedent” status for the excipient and this notation permits use of the material in new drug products under defined conditions without the need to submit extensive supporting data.

In Japan however, new is new. Even with precedent status, if a new higher dose or a new route of administration is pursued, the examination will treat the excipient as new. This also applies to an approved food additive or cosmetic ingredient. The first use in a pharmaceutical formulation is considered a new use and the application will be examined as such.

Current Regulatory Status of CDs

The parent CDs in Japan are classified as natural starches that have received approval by the Ministries of Health for use in foods. Relative to pharmaceutical applications,

monographs for α - and β -CD have been included in the Japanese Pharmaceutical Excipients (196) compendium (JPE). Even though nine pharmaceutical products with CD formulations have been marketed in Japan, the use of CDs has not been extensive enough in approved formulations to receive "precedent status."

In the United States, two drug products are approved containing CDs (one each containing α -CD and (2HP)- β -CD) and at least one additional NDA has been filed (product containing SBE7- β -CD). In addition, Drug Master Files have been submitted for β - and γ -CD and the (2HP)- β -CD and SBE7-7 β -CD derivatives. These DMFs are available for referencing in IND and NDA applications through agreements with the individual manufacturers. A food additive petition is also under review for β -CD, and a monograph on β -CD has been included in volume 19 of the NF under the name Betadex (197). The United States Pharmacopeial Convention is reviewing proposal(s) to include monographs on additional derivative(s).

An expert panel concluded in 1997 that β -CD is GRAS for its intended use as a flavor carrier and protectant at a level of 2% in numerous food products. The products included chewing gum, graham and puddings, soups prepared from dry mixes, coffee and tea products with added flavors, savory snacks and crackers with added flavorings, baked goods prepared from dry mixes, beverages prepared from dry mixes, and breakfast cereals. A petition has been filed with the FDA requesting their affirmation of β -CD as GRAS for these products. The Joint Expert Committee on Food Additives (JECFA) of the World Health Organization and the Food and Agriculture Organization has reviewed β -CD and established an acceptable daily intake (ADI) of 0–5 mg/kg body weight. The Scientific Committee for Foods of the European Union has also assigned β -CD an ADI of 5 mg/kg body weight/day. The FDA granted GRAS status for γ -CD in 2000.

REFERENCES

- Villiers, A. Sur La Fermentation De La Féculé Par L'action Du Ferment Butyrique. *C R Hebd. Seances Acad. Sci.* **1891**, *112*, 536–538.
- Schardinger, F. Thermophile Bakterien Aus Verschiedenen Nahrungsmitteln Und Milch Und Die Gebildeten Produkte, Wenn Diese Bakterien in Nahrlosungen Kultiviert Werden, Die Kohlenhydrate Enthalen. *Z. Unters. Nahr. Genussm.* **1903**, *6*, 865.
- French, D.; Rundle, R.E. The Molecular Weights of the Schardinger Alpha and Beta Dextrins. *J. Am. Chem. Soc.* **1942**, *64*, 1651–1653.
- Freudenberger, K.; Cramer, F. Die Konstitution Der Schardinger Dextrine Alpha, Beta, Und Gamma. *Z. Naturforsch.* **1948**, *3b*, 464.
- Proceedings of 10th International Symposium on Cyclodextrins Ann Arbor MI May 2000, Schmid, G., Eds.; MIRA Digital Publishing: St. Louis, MO, 2001.
- Thompson, D.O. Cyclodextrins-Enabling Excipients: their Present and Future Use in Pharmaceuticals. *Crit. Rev. Ther. Drug Carrier Syst.* **1997**, *14* (1), 1–104.
- Frömming, K.-H.; Szejtli, J. Cyclodextrins in Pharmacy. Kluwer Academic Publishers: Dordrecht, The Netherlands, 1994; 224.
- Uekama, K.; Otagiri, M. Cyclodextrins in Drug Carrier Systems. *CRC Crit. Rev. Ther. Drug Carrier Syst.* **1987**, *3* (1), 1–40.
- Higuchi, T.; Connors, K.A. Phase-Solubility Techniques. *Advances in Analytical Chemistry and Instrumentation*, Reilly, C.N., Ed.; John Wiley & Sons, Inc.: New York, 1965; 4, 117–212.
- Merino, C.; Junquera, E.; Jimenez-Barbero, J.; Aicart, E. Effect of the Presence of β -Cyclodextrin on the Solution Behavior of Procaine Hydrochloride. Spectroscopic and Thermodynamic Studies. *Langmuir* **2000**, *16* (4), 1557–1565.
- Maupas, B.; Letellier, S.; Guyon, F. Determination of the Formation Constant for the Inclusion Complex of Methyl- β -Cyclodextrin with Anticoagulant Drugs Warfarin and 8-Chlorowarfarin in Aqueous Solution. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1996**, *23* (4), 259–267.
- Sadlej-Sosnowska, N. Fluorometric Determination of Association Constants of Three Estrogens with Cyclodextrins. *J. Fluoresc.* **1997**, *7* (3), 195–200.
- Escandar, G.M. Spectrofluorimetric Determination of Piroxicam in the Presence and Absence of β -Cyclodextrin. *Analyst* **1999**, *124* (4), 587–591.
- Nishijo, J.; Ushiroda, Y. Interaction of 2-Naphthalene-sulfonate With β -Cyclodextrin: Studies with Calorimetry and Proton Nuclear Magnetic Resonance Spectroscopy. *Chem. Pharm. Bull.* **1998**, *46* (11), 1790–1796.
- Li, S.; Purdy, W.C. Circular Dichroism, Ultraviolet, and Proton Nuclear Magnetic Resonance Spectroscopic Studies of the Chiral Recognition Mechanism of β -Cyclodextrin. *Anal. Chem.* **1992**, *64* (13), 1405–1412.
- Patonay, G.; Warner, I.M. Investigation of Induced Circular Dichroism of Benzo(a)pyrene Cyclodextrin Complexes. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1991**, *11* (4), 313–322.
- DePonti, R.; Torricelli, C.; Motta, A.; Crivellente, M. Use of a Polarographic Method, a UV Method, and the Phase-Solubility Technique to Determine the Stability Constant in Aqueous Solution of a β -Cyclodextrin Complex with a New Immunomodulating Agent. *Eur. J. Pharm. Biopharm.* **1991**, *37* (2), 106–109.
- Koenigbauer, M.J. Pharmaceutical Applications of Microcalorimetry. *Pharm. Res.* **1994**, *11* (6), 777–783.
- Mwakibete, H.; Cristantino, R.; Bloor, D.M.; Wyn-Jones, E.; Holzwarth, J.F. Reliability of the Experimental Methods to Determine Equilibrium Constants for Surfactant/ β -Cyclodextrin Inclusion Complexes. *Langmuir* **1995**, *11* (1), 57–60.
- Dharmawardana, U.R.; Christian, S.D.; Tucker, E.E.; Taylor, R.W.; Scamehorn, J.F. A Surface Tension Method

- for Determining Binding Constants for Cyclodextrin Inclusion Complexes of Ionic Surfactants. *Langmuir* **1993**, 9 (9), 2258–2263.
21. Ono, N.; Hirayama, F.; Arima, H.; Uekama, K. Determination of Stability Constant of β -Cyclodextrin Complexes Using the Membrane Permeation Technique and the Permeation Behavior of Drug-Competing Agent- β -Cyclodextrin Ternary Systems. *Eur. J. Pharm. Sci.* **1999**, 8 (2), 133–139.
22. Bartak, P.; Bednar, P.; Kubacek, L.; Stransky, Z. Advanced Statistical Evaluation of Complex Formation Constant from Electrophoretic Data. *Anal. Chim. Acta* **2000**, 407 (1/2), 327–336.
23. Neubert, R.H.H.; Schwarz, M.A.; Mrestani, Y.; Platzer, M.; Raith, K. Affinity Capillary Electrophoresis in Pharmaceutics. *Pharm. Res.* **1999**, 16 (11), 1663–1673.
24. Suzuki, M.; Ito, K.; Fushimi, C.; Kondo, T. Application of Freezing Point Depression to Drug Interaction Studies. I. a Study of 1:1 Plus 1:2 Complexes Between Barbiturate and α -Cyclodextrin Using the Freezing Point Depression Method. *Chem. Pharm. Bull.* **1993**, 41 (8), 1444–1447.
25. Thuaud, N.; Seville, B.; Deratani, A.; Lelievre, G. Determination by High-Performance Liquid Chromatography of the Binding Properties of Charged β -Cyclodextrin Derivatives with Drugs. *J. Chromatogr.* **1990**, 503 (2), 453–458.
26. Hamai, S.; Handa, M. Paper Chromatographic Study on the Inclusion Complexes of Cyclodextrins and Methyl Orange. *Anal. Lett.* **1999**, 32 (5), 1037–1047.
27. Wang, M.; Ueda, H.; Nagai, T. Simple Prediction of Stability Constants for Inclusion Complexes of β -Cyclodextrin with Various Drug Molecules Using β -Cyclodextrin Bonded Phases (Cyclobond I Column). *Drug Dev. Ind. Pharm.* **1990**, 16 (4), 571–579.
28. Cserhati, T.; Forgacs, E. Charge-Transfer Chromatographic Study of the Complex Formation of Some Steroidal Drugs with Carboxymethyl γ -Cyclodextrin. *Anal. Biochem.* **1997**, 246 (2), 205–210.
29. Lederer, M.; Leipzig-Pagani, E. A Simple Alternative Determination of the Formation Constant for the Inclusion Complex Between Rutin and β -Cyclodextrin. *Anal. Chim. Acta* **1996**, 329 (3), 311–314.
30. Kraus, C.; Mehnert, W.; Froemming, K.H. Interactions of β -Cyclodextrin With Solutol HS 15 and their Influence on Diazepam Solubilization. *PZ Wiss.* **1991**, 4 (1), 11–15.
31. Lin, S.Y.; Kao, Y.H.; Yang, J.C. Grinding Effect on Some Pharmaceutical Properties of Drugs by Adding β -Cyclodextrin. *Drug Dev. Ind. Pharm.* **1988**, 14 (1), 99–118.
32. Abdel-Rahman, S.I.; El-Sayed, A.M. Interaction of Some Benzothiadiazine Diuretics with β -Cyclodextrin. *Acta Pharm. Hung.* **1990**, 60 (2/3), 69–75.
33. Bekers, O.; Beijnen, J.H.; Otagiri, M.; Bult, A.; Underberg, W.J.M. Inclusion Complexation of Doxorubicin and Daunorubicin with Cyclodextrins. *J. Pharm. Biomed. Anal.* **1990**, 8 (12), 671–674.
34. Husain, N.; Ndou, T.T.; Munoz de la Pena, A.; Warner, I.M. Complexation of Doxorubicin with β - and γ -Cyclodextrins. *Appl. Spectrosc.* **1992**, 46 (4), 652–658.
35. Imai, T.; Irie, T.; Otagiri, M.; Uekama, K.; Yamasaki, M. Comparative Study on Inclusion Complexations of Antiinflammatory Drug Flurbiprofen with β -Cyclodextrin and Methylated Beta-Cyclodextrins. *J. Inclusion Phenom.* **1984**, 2 (3/4), 597–604.
36. Abdel-Rahman, A.A.; Saleh, S.I.; Nakai, Y.; Aboutaleb, A.E.; Ahmed, M.O. Investigation of the Interaction of Bromazepam with Cyclodextrins in Solutions and in Ground Mixtures. *J. Pharm. Belg.* **1994**, 49 (1), 23–32.
37. Ahmed, M.O.; Nakai, Y.; Aboutaleb, A.E.S.; Yamamoto, K.; Abdel Rahman, A.A.Z.; Saleh, S.I. Complex Formation of Nitrazepam in Coprecipitating and in Co-grinding with Methylated Beta-Cyclodextrins. *Chem. Pharm. Bull.* **1990**, 38 (12), 2423–2427.
38. Müller, B.W.; Brauns, U. Solubilization of Drugs by Modified Beta-Cyclodextrins. *Int. J. Pharm.* **1985**, 26 (1/2), 77–88.
39. Müller, B.W.; Brauns, U. Hydroxypropyl-Beta-Cyclodextrins Derivatives: Influence of Average Degree of Substitution on Complexing Ability and Surface Activity. *J. Pharm. Sci.* **1986**, 75 (6), 571–572.
40. Yoshida, A.; Yamamoto, M.; Irie, T.; Hirayama, F.; Uekama, K. Some Pharmaceutical Properties of 3-Hydroxypropyl- and 2,3-Dihydroxypropyl- β -Cyclodextrins and their Solubilizing and Stabilizing Abilities. *Chem. Pharm. Bull.* **1989**, 37 (4), 1059–1063.
41. Rao, C.T.; Pitha, J.; Lindberg, B.; Lindberg, J. Distribution of Substituents In *O*-(2-Hydroxypropyl) Derivatives of Cyclomalto-Oligosaccharides (Cyclodextrins): Influence of Increasing Substitution, of the Base Used in the Preparation, and of Macrocyclic Size. *Carbohydr. Res.* **1992**, 223, 99–107.
42. Kano, K.; Zhou, B.; Sakaguchi, M.; Matsumoto, H.; Hashimoto, S. Effects of Aeration, Freezing and Thawing on Self-Aggregation of Naphthalene in Aqueous β -Cyclodextrin Solutions. *Sci. Eng. Rev. Doshisha Univ.* **1985**, 25 (4), 253–259.
43. Menger, F.M.; Williams, D.Y. Synthesis and Properties of a Surfactant-Cyclodextrin Conjugate. *Tetrahedron Lett.* **1986**, 27 (23), 2579–2582.
44. Stella, V.; Rajewski, R.A. Derivatives of Cyclodextrins Exhibiting Enhanced Aqueous Solubility and the Use Thereof. US Patent 5,134,127, 1992.
45. Gerloczy, A.; Hoshino, T.; Pitha, J. Safety of Oral Cyclodextrins: Effects of (Hydroxypropyl) Cyclodextrins, Cyclodextrin Sulfates and Cationic Cyclodextrins on Steroid Balance in Rats. *J. Pharm. Sci.* **1994**, 83 (2), 193–196.
46. Okimoto, K.; Rajewski, R.A.; Jona, J.A.; Stella, V.J. The Interaction of Charged and Uncharged Drugs with a Neutral (HP- β -Cyclodextrin) and Anionically Charged (SBE7-b-Cyclodextrin) β -Cyclodextrin. *Pharm. Res.* **1996**, 13 (2), 256–264.
47. Badawy, S.I.F.; Marshall, A.L.; Ghorab, M.M.; Adeyeye, C.M. A Study of the Complexation Between Danazol and Hydrophilic Cyclodextrins Derivatives. *Drug Devel. Indust. Pharm.* **1996**, 22 (9/10), 959–966.
48. Rajewski, R.A. Ph.D. Dissertation. *Development and Evaluation of the Usefulness and Parenteral Safety of Modified Cyclodextrins*; University of Kansas: Lawrence, KS USA, 1991.
49. Johnson, M.D.; Hoesterey, B.L.; Anderson, B.D. Solubilization of a Tripeptide HIV Protease Inhibitor Using a Combination of Ionization and Complexation with

- Chemically Modified Cyclodextrins. *J. Pharm. Sci.* **1994**, *83* (8), 1142–1146.
50. Rajewski, R.A.; Traiger, G.; Bresnahan, J.; Jaberaboansari, P.; Stella, V.J.; Thompson, D.O. Preliminary Safety Evaluation of Parenterally Administered Sulfoalkyl Ether β -Cyclodextrin Derivatives. *J. Pharm. Sci.* **1995**, *84* (8), 927–932.
 51. Otero-Espinar, F.J.; Anguiano-Igea, S.; Garcia-Gonzalez, N.; Vila-Jato, J.L.; Blanco-Mendez, J. Interaction of Naproxen with β -Cyclodextrin in Solution and in the Solid State. *Int. J. Pharm.* **1992**, *79* (2/3), 149–157.
 52. Hoshino, T.; Uekama, K.; Pitha, J. Increase in Temperature Enhances Solubility of Drugs in Aqueous Solutions of Hydroxypropyl Cyclodextrins. *Int. J. Pharm.* **1993**, *98* (1/3), 239–242.
 53. Menard, F.A.; Dedhiya, M.G.; Rhodes, C.T. Physicochemical Aspects of the Complexation of Some Drugs with Cyclodextrins. *Drug Dev. Ind. Pharm.* **1990**, *16* (1), 91–113.
 54. Pitha, J.; Hoshino, T. Effects of Ethanol on Formation of Inclusion Complexes of Hydroxypropyl Cyclodextrins with Testosterone or with Methyl Orange. *Int. J. Pharm.* **1992**, *80* (2/3), 243–251.
 55. Reer, O.; Müller, B.W. The Influence of Cosolvents and Surfactants on the Complexation of Dexamethasone with Hydroxypropyl- β -Cyclodextrin Investigated with a Simplex Lattice Design. *Eur. J. Pharm. Biopharm.* **1993**, *39* (3), 105–111.
 56. Mulski, M.J.; Connors, K.A. Solvent Effects on Chemical Processes. 9. Energetic Contributions to the Complexation of 4-Nitroaniline with Alpha-Cyclodextrin in Water and in Binary Aqueous-Organic Solvents. *Supramol. Chem.* **1995**, *4* (4), 271–278.
 57. Loftsson, T.; Masson, M.; Sigurjonsdottir, J.F. Methods to Enhance the Complexation Efficiency of Cyclodextrins. *Stp Pharma Sciences* **1999**, *9* (3), 237–242.
 58. Redenti, E.; Szente, L.; Szejtli, J. Drug/Cyclodextrin/Hydroxy Acid Multicomponent Systems. Properties and Pharmaceutical Applications. *J. Pharm. Sci.* **2000**, *89* (1), 1–8.
 59. Turro, N.J.; Okubo, T.; Chung, C.J. Analysis of Static and Dynamic Host-Guest Associations of Detergents with Cyclodextrins via Photoluminescence Methods. *J. Am. Chem. Soc.* **1982**, *104* (7), 1789–1794.
 60. Hashimoto, S.; Thomas, J.K. Fluorescence Study of Pyrene and Naphthalene in Cyclodextrin-Amphiphile Complex Systems. *J. Am. Chem. Soc.* **1985**, *107* (16), 4655–4662.
 61. Stella, V.J.; Rao, V.M.; Zannou, E.A.; Zia, V. Mechanisms of Drug Release from Cyclodextrin Complexes. *Adv. Drug Del. Rev.* **1999**, *36* (1), 3–16.
 62. Szejtli, J. Medicinal Applications of Cyclodextrins. *Med. Res. Rev.* **1994**, *14* (3), 353–386.
 63. Hirai, S. Formulation Studies of Cefotiam Hexetil Hydrochloride: Effect of α -Cyclodextrin as Dissolution Enhancer. *Proceedings of 7th International Symposium on Cyclodextrins*; Osa, T., Ed.; Komiyama Printing Co., Ltd.: Tokyo, 1994; 39–44.
 64. Järvinen, T.; Järvinen, K.; Schwarting, N.; Stella, V.J. β -Cyclodextrin Derivatives, SBE4- β -Cyclodextrin and HP- β -Cyclodextrin, Increase the Oral Bioavailability of Cinnarizine in Beagle Dogs. *J. Pharm. Sci.* **1995**, *84* (3), 295–299.
 65. Soliman, O.A.; Kimura, K.; Hirayama, F.; Uekama, K.; El-Sabbagh, H.M.; El-Gawad, A.E.; Hashim, F.M. Amorphous Spironolactone-Hydroxypropylated Cyclodextrin Complexes with Superior Dissolution and Oral Bioavailability. *Int. J. Pharm.* **1997**, *149* (1), 73–83.
 66. Saket, M. Improvement of Solubility and Dissolution Rate of Meclizine Hydrochloride Utilizing Cyclodextrins and Non-Ionic Surfactant Solutions Containing Cosolvents and Additives. *Acta Technologiae et Legis Medicamenti* **1997**, *8* (1), 33–48.
 67. Nagarsenker, M.S.; Tantry, J.S.; Shenai, H. Influence of Hydroxypropyl- β -Cyclodextrin on the Dissolution of Ketoprofen and Irritation to Gastric Mucosa after Oral Administration in Rats. *Pharm. Sci.* **1997**, *3* (9), 443–445.
 68. Moyano, J.R.; Arias, M.J.; Gines, J.M.; Perez, J.I.; Rabasco, A.M. Dissolution Behavior of Oxazepam in Presence of Cyclodextrins: Evaluation of Oxazepam-DIMEB Binary System. *Drug Dev. Ind. Pharm.* **1997**, *23* (4), 379–385.
 69. Badawy, S.I.F.; Ghorab, M.M.; Adeyeye, C. Study of the Complexation Between Danazol and Hydrophilic Cyclodextrin Derivatives. *Pharm. Res.* **1995**, *12* (9), S204.
 70. Savolainen, J.; Jarvinen, K.; Matilainen, L.; Jarvinen, T. Improved Dissolution and Bioavailability of Phenytoin by Sulfobutylether- β -Cyclodextrin ((Sbe)(7m)- β -Cd) and Hydroxypropyl-B-Cyclodextrin (Hp- β -Cd) Complexation. *Int. J. Pharm.* **1998**, *165* (1), 69–78.
 71. Kimura, K.; Hirayama, F.; Arima, H.; Uekama, K. Effects of Aging on Crystallization, Dissolution and Absorption Characteristics of Amorphous Tolbutamide-2-Hydroxypropyl- β -Cyclodextrin Complex. *Chem. Pharm. Bull.* **2000**, *48* (5), 646–650.
 72. Vora, J.; Boroujerdi, M. Enhanced Aqueous Solubility of Phenolic Antioxidants Using Modified Beta-Cyclodextrins. *Drug Dev. Ind. Pharm.* **1995**, *21* (4), 495–502.
 73. Lehner, S.J.; Müller, B.W.; Seydel, J.K. Effect of Hydroxypropyl- β -Cyclodextrin on the Antimicrobial Action of Preservatives. *J. Pharm. Pharmacol.* **1994**, *46* (3), 186–191.
 74. Loftsson, T.; Stefansdottir, O.; Friordottir, H.; Guomundsson, O. Interactions Between Preservatives and 2-Hydroxypropyl β -Cyclodextrin. *Drug Dev. Ind. Pharm.* **1992**, *18* (13), 1477–1484.
 75. Nunuz-Delgado, E.; Sanchez-Ferrer, A.; Garcia-Carmona, F. Cyclodextrins As Secondary Antioxidants: Synergism With Ascorbic Acid. *J. Agric. Food Chem.* **1997**, *45* (8), 2830–2835.
 76. Otero Espinar, F.J.; Anguiano Igea, S.; Blanco Mendez, J.; Vila Jato, J.L. Reduction in the Ulcerogenicity of Naproxen by Complexation with β -Cyclodextrin. *Int. J. Pharm.* **1991**, *70* (1/2), 35–41.
 77. Park, J.H.; Kim, J.H.; Kim, J.I.; Kim, S.J.; Seo, S.H.; Lee, K. Comparison of Diclofenac Sodium and Diclofenac Sodium- β -Cyclodextrin Complexation on Gastric Mucosal Injury in Rats. *Yakche Hakhoechi* **1997**, *27* (1), 11–14.
 78. Santucci, L.; Fiorucci, S.; Chiucchiu, S.; Sicilia, A.; Bufalino, L.; Morelli, A. Placebo-Controlled Comparison of Piroxicam- β -Cyclodextrin, Piroxicam, and Indomethacin on Gastric Potential Difference and Mucosal Injury in Humans. *Dig. Dis. Sci.* **1992**, *37* (12), 1825–1832.

79. Loftsson, T.; Jarvinen, T. Cyclodextrins in Ophthalmic Drug Delivery. *Adv. Drug Del. Rev.* **1999**, *36* (1), 59–79.
80. Doenicke, A.; Roizen, M.F.; Nebauer, A.E.; Kugler, A.; Hoernecke, R.; Beger-Hintzen, H. A Comparison of Two Formulations for Etomidate, 2-Hydroxypropyl- β -Cyclodextrin (HPCyclodextrin) and Propylene Glycol. *Anesth. Analg. (NY)* **1994**, *79* (5), 933–939.
81. Stella, V.J.; Lee, H.K.; Thompson, D.O. The Effect of SBE4- β -Cyclodextrin on I.m. Prednisolone Pharmacokinetics and Tissue Damage in Rabbits: Comparison to a Co-Solvent Solution and a Water-Soluble Prodrug. *Int. J. Pharm.* **1995**, *120* (2), 197–204.
82. Howard, D.; Morris, R.M.; Rinkema, L.; Morrow, T.E.; Laska, D.A. Moderation of Untoward Side Effects during In Vivo and In Vitro Screening of Potential I.V.z. Preparations. *Min. 6th Int. Symp. Cyclodextrins*; Hedges, A.R., Ed.; Editions de Santé: Paris, 1992; 474–477.
83. Weiszfeiler, V.; Szejtli, J. Bitterness Reduction with β -Cyclodextrin. *Proceedings of 4th International Symposium of Cyclodextrins*; Huber, O.; Szejtli, J., Eds.; Kluwer Academic Publishers: Dordrecht The Netherlands, 1988; 359–366.
84. Funasaki, N.; Kawaguchi, R.; Ishikawa, S.; Hada, S.; Neya, S.; Katsu, T. Quantitative Estimation of the Bitter Taste Intensity of Oxyphenonium Bromide Reduced by Cyclodextrins from Electromotive Force Measurements. *Anal. Chem.* **1999**, *71* (9), 1733–1736.
85. Funasaki, N.; Uemura, Y.; Hada, S.; Neya, S. Reduction of the Bitter Taste Intensity of Propantheline Bromide by Cyclodextrins as Predicted by Surface Tension Measurements. *J. Phys. Chem.* **1996**, *100* (40), 16298–16301.
86. Uekama, K.; Oh, K.; Otogiri, M.; Seo, H.; Tsuruoka, M. Improvement of Some Pharmaceutical Properties of Clofibrate by Cyclodextrin Complexation. *Pharm. Acta Helv.* **1983**, *58* (12), 338–342.
87. Loftsson, T. Effects of Cyclodextrins on the Chemical Stability of Drugs in Aqueous Solutions. *Drug Stab.* **1995**, *1* (1), 22–33.
88. Loftsson, T.; Brewster, M.E. Pharmaceutical Applications of Cyclodextrins. 1. Drug Solubilization and Stabilization. *J. Pharm. Sci.* **1996**, *85* (10), 1017–1025.
89. Ueda, H.; Ou, D.; Endo, T.; Nagase, H.; Nagai, T. Evaluation of a Sulfobutyl Ether β -Cyclodextrin as a Solubilizing/Stabilizing Agent for Several Drugs. *Symposium on Pharmaceutical Applications of Cyclodextrins*; Lawrence, KS, June 29 – July 2, 1997.
90. Ma, D.Q.; Rajewski, R.A.; Vander Velde, D.; Stella, V.J. Comparative Effects of (Sbe)(7m)- β -Cd and Hp- β -Cd on the Stability of Two Anti-Neoplastic Agents, Melphalan and Carmustine. *J. Pharm. Sci.* **2000**, *89* (2), 275–287.
91. Schlenk, H.; Sand, D.M.; Tillotson, J.A. Stabilization of Autoxidizable Materials by Means of Inclusion. *J. Am. Chem. Soc.* **1955**, *77*, 3587–3590.
92. Hirayama, F.; Kurihara, M.; Uekama, K. Improving the Aqueous Stability of Prostaglandin E2 and Prostaglandin A2 by Inclusion Complexation With Methylated- β -Cyclodextrins. *Chem. Pharm. Bull.* **1984**, *32* (10), 4237–4240.
93. Jarho, P.; Vander Velde, D.; Stella, V.J. Cyclodextrin-Catalyzed Deacetylation of Spironolactone is pH And Cyclodextrin Dependent. *J. Pharm. Sci.* **2000**, *89* (2), 241–249.
94. Oguchi, T.; Yonemochi, E.; Yamamoto, K.; Nakai, Y. Freeze-Drying of Drug-Additive Binary Systems. II. Relationship Between Decarboxylation Behavior and Molecular States of p-Aminosalicylic Acid. *Chem. Pharm. Bull.* **1989**, *37* (11), 3088–3091.
95. Mielcarek, J. Photochemical Stability of the Inclusion Complexes of Nicardipine with α -, γ -Cyclodextrin, Methyl- β -Cyclodextrin, and Hydroxypropyl- β -Cyclodextrin in the Solid State and in Solution. *Pharmazie* **1996**, *51* (7), 477–479.
96. Ammar, H.O.; El-Nahhas, S.A. Improvement of Some Pharmaceutical Properties of Drugs by Cyclodextrin Complexation. 2. Colchicine. *Pharmazie* **1995**, *50* (4), 269–272.
97. Yamamoto, M.; Hirayama, F.; Uekama, K. Improvement of Stability and Dissolution of Prostaglandin E1 by Maltosyl Beta-Cyclodextrin in Lyophilized Formulation. *Chem. Pharm. Bull.* **1992**, *40* (3), 747–751.
98. Cwiernia, B.; Hladon, T.; Stobiecki, M. Stability of Diclofenac Sodium in the Inclusion Complex with β -Cyclodextrin in the Solid State. *J. Pharm. Pharmacol.* **1999**, *51* (11), 1213–1218.
99. Lee, H.J.; Kim, S.U.; Seo, S.H. Pharmaceutical Studies on the Drug (III). Enhancement of Dissolution and Stability of Sulfamethoxazole by Beta-Cyclodextrin Complexation. *Yakche Hakhoechi* **1988**, *18* (4), 181–186.
100. Sigurjonsdottir, J.F.; Loftsson, T.; Masson, M. Influence of Cyclodextrins on the Stability of the Peptide Salmon Calcitonin in Aqueous Solution. *Int. J. Pharm.* **1999**, *186* (2), 205–213.
101. Sah, H. Stabilization of Proteins Against Methylene Chloride Water Interface-Induced Denaturation and Aggregation. *J. Controlled Release* **1999**, *58* (2), 143–151.
102. Karuppiiah, N.; Sharma, A. Cyclodextrins As Protein Folding Aids. *Biochem. Biophys. Res. Commun.* **1995**, *211* (1), 60–66.
103. Tokihiro, K.; Irie, T.; Uekama, K. Varying Effects of Cyclodextrin Derivatives on Aggregation and Thermal Behavior of Insulin in Aqueous Solution. *Chem. Pharm. Bull.* **1997**, *45* (3), 525–531.
104. Katakam, M.; Banga, A.K. Aggregation of Proteins and Its Prevention by Carbohydrate Excipients: Albumins and γ -Globulin. *J. Pharm. Pharmacol.* **1995**, *47* (2), 103–107.
105. Cooper, A.; Lovatt, M.; Nutley, M.A. Energetics of Protein-Cyclodextrin Interactions. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1996**, *25* (1–3), 85–88.
106. Sharma, A. Cyclodextrin-Assisted Protein Folding. *Symposium on Pharmaceutical Applications of Cyclodextrins*; Lawrence, KS, June 29 – July 2, 1997.
107. Utsuki, T.; Hirayama, F.; Uekama, K. Different Photo-dimerization Behavior of Tranilast in α -, β - and Gamma-Cyclodextrin Complexes: Cavity-Size and Stoichiometry Dependence. *J. Chem. Soc., Perkin Trans* **1993**, *2* (1), 109–114.
108. Lin, G.; Xu, S. A New Method of Encapsulating Volatile Oils in Chinese Medicinal Herbs with Beta-Cyclodextrin. *Zhongguo Zhongyao Zazhi* **1992**, *17* (5), 283–285.
109. Qi, Z.H.; Hedges, A.R. Use of Cyclodextrins for Flavors. *ACS Symp. Ser., Flavor Technology* **1995**, *610*, 231–243.
110. Bhandari, B.R.; D'arcy, B.R.; Bich, L.L.T. Lemon Oil to β -Cyclodextrin Ratio Effect on the Inclusion Efficiency of

- β -Cyclodextrin and the Retention of Oil Volatiles in the Complex. *J. Agric. Food Chem.* **1998**, *46* (4), 1494–1499.
111. Uekama, K.; Oh, K.; Irie, T.; Otagiri, M.; Nishimiya, Y.; Nara, T. Stabilization of Isosorbide 5-Mononitrate in Solid State by β -Cyclodextrin Complexation. *Int. J. Pharm.* **1985**, *25* (3), 339–346.
112. Umemura, M.; Ueda, H.; Tomono, K.; Nagai, T. Effect of Diethyl Beta-Cyclodextrin on the Release of Nitroglycerin from Formulations. *Drug Des. Delivery* **1990**, *6* (4), 297–310.
113. Szente, L.; Apostol, I.; Szejtli, J. Suppositories Containing Beta-Cyclodextrin Complexes. Part I: Stability Studies. *Pharmazie* **1984**, *39* (10), 697–699.
114. Saenger, W.; Betzel, C.; Hingerty, B.; Brown, G.M. Flip-Flop Hydrogen Bridging Bonds in Beta-Cyclodextrin—A General Principle in Polysaccharides. *Angew. Chem.* **1983**, *95* (11), 883–884.
115. Jozwiakowski, M.J.; Connors, K.A. Aqueous Solubility Behavior of Three Cyclodextrins. *Carb. Res.* **1985**, *143*, 51–59.
116. Coleman, A.W.; Nicolis, I.; Keller, N.; Dalbiez, J.P. Aggregation of Cyclodextrins: An Explanation of the Abnormal Solubility of Beta-Cyclodextrin. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1992**, *13* (2), 139–143.
117. Pedersen, M. Effect of Hydrotropic Substances on the Complexation of Clotrimazole with Beta-Cyclodextrin. *Drug Dev. Ind. Pharm.* **1993**, *19* (4), 439–448.
118. Coleman, A.W.; Nicolis, I. Inorganic Salt Modulation of the Aqueous Solubility of Beta-Cyclodextrin. *Supramol. Chem.* **1993**, *2* (2/3), 93–97.
119. Loftsson, T.; Frioriksdottir, H. The Effect of Water-Soluble Polymers on the Aqueous Solubility and Complexing Abilities of β -Cyclodextrin. *Int. J. Pharm.* **1998**, *163* (1/2), 115–121.
120. Szejtli, J. Cyclodextrins and their Inclusion Complexes. *Akadémiai Kiadó: Budapest Hungary*, 1982; 296.
121. Schonberger, B.P.; Jansen, A.C.A.; Janssen, L.H.M. The Acid Hydrolysis of Cyclodextrins and Linear Oligosaccharides: A Comparative Study. *Proceedings of 4th International Symposium on Cyclodextrins*; Huber, O.; Szejtli, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1988; 61–63.
122. Szejtli, J.; Budai, Z. Acid Hydrolysis of Beta-Cyclodextrin. *Acta Chim. Acad. Sci. Hung.* **1976**, *91* (1), 73–80.
123. Al-Rawi, A.M. The Susceptibility of Gamma-Cyclodextrin (Schardinger Dextrin) to Ionizing Radiation. *Food Preserv. Irradiat., Proceedings of the International Symposium*; IAEA: Vienna Austria, 1978; 487–500.
124. Nakai, Y.; Yamamoto, K.; Terada, K.; Kajiyama, A.; Sasaki, I. Properties of Crystal Water of α -, β -, and Gamma-Cyclodextrin. *Chem. Pharm. Bull.* **1986**, *34* (5), 2178–2182.
125. Lindner, K.; Saenger, W. Topography of Cyclodextrin Complexes. Part XVII. Crystal and Molecular Structure of Cycloheptaamylose Dodecahydrate. *Carbohydr. Res.* **1982**, *99* (2), 103–115.
126. Fujiwara, T.; Yamazaki, M.; Tomizu, Y.; Tokuoaka, R.; Tomita, K.; Matsuo, T.; Suga, H.; Saenger, W. The Crystal Structure of a New Form of Beta-Cyclodextrin Water Inclusion Compound and Thermal Properties of Beta-Cyclodextrin Inclusion Complexes. *Nippon Kagaku Kaishi* **1983**, *2*, 181–187.
127. Yamamoto, M.; Aritomi, H.; Irie, T.; Hirayama, F.; Uekama, K. Biopharmaceutical Evaluation of Maltosyl Beta-Cyclodextrin as a Parenteral Drug Carrier. *S.T.P. Pharma Sci.* **1991**, *1* (6), 397–402.
128. Yamamoto, M.; Yoshida, A.; Hirayama, F.; Uekama, K. Some Physicochemical Properties of Branched Beta-Cyclodextrins and their Inclusion Characteristics. *Int. J. Pharm.* **1989**, *49* (2), 163–171.
129. Ou, D.; Ueda, H.; Nagase, H.; Endo, T.; Nagai, T. Some Pharmaceutical Properties of 2,3,6-Partially Methylated- β -Cyclodextrin and its Solubilizing and Stabilizing Abilities. *Drug Dev. Ind. Pharm.* **1994**, *20* (12), 2005–2016.
130. Tsuchiyama Y.; Sato M.; Yagi Y.; Ishikura, T. Partially Methylated Cyclodextrins and Process for the Producing the Same. US Patent 4,746,734, 1988.
131. Harata K. Macrocyclic Conformation of Methylated Cyclodextrins Minutes of the 5th International Symposium on Cyclodextrins, Duchene, D. Editions de Santé; Paris, 1990; 77–81.
132. Irie, T.; Fukunaga, K.; Yoshida, A.; Uekama, K.; Fales, H.M.; Pitha, J. Amorphous Water-Soluble Cyclodextrin Derivatives: 2-Hydroxyethyl, 3-Hydroxypropyl, 2-Hydroxyisobutyl, and Carboxamidomethyl Derivatives of Beta-Cyclodextrin. *Pharm. Res.* **1988**, *5* (11), 713–717.
133. Pitha, J.; Milecki, J.; Fales, H.; Pannell, L.; Uekama, K. Hydroxypropyl- β -Cyclodextrin: Preparation and Characterization; Effects on Solubility of Drugs. *Int. J. Pharm.* **1986**, *29* (1), 73–82.
134. Pitha, J.; Rao, C.T.; Lindberg, B.; Seffers, P. Distribution of Substituents in 2-Hydroxypropyl Ethers of Cyclomaltoheptaose. *Carbohydr. Res.* **1990**, *200*, 429–435.
135. Yoshida, A.; Arima, H.; Uekama, K.; Pitha, J. Pharmaceutical Evaluation of Hydroxyalkyl Ethers of Beta-Cyclodextrins. *Int. J. Pharm.* **1988**, *46* (3), 217–222.
136. Brandt, L.; Felcht, U.H. Water-Soluble Mixed Ether of Beta-Cyclodextrin. EP Patent 146, 841, 1985.
137. Brauns, U.; Mueller, B.W.W. Beta-Cyclodextrins Inclusion Compounds for Pharmaceuticals. DE Patent 3, 346, 123, 1985.
138. Muller, B.W.W. Preparation of Gamma-Cyclodextrins Ethers and their Use as Complexants for Chemicals and Pharmaceuticals. EP Patent 197, 571, 1986.
139. Pitha, J. Preparation of Inclusion Compounds of Cyclodextrins Ethers with Lipophilic Drugs. US Patent 4, 727, 064, 1988.
140. Lindberg, B.; Lindberg, J.; Pitha, J.; Rao, C.T.; Harata, K. Synthesis of Some 2-O-(2-Hydroxyalkyl) and 2-O-(2,3-Dihydroxyalkyl) Derivatives of Cyclomaltoheptaose. *Carbohydr. Res.* **1991**, *22*, 113–119.
141. Parmerter, S.M.; Allen, E.E., Jr.; Hull, G.A. Cyclodextrins With Anionic Properties. US Patent 3, 426, 011 1969.
142. Lammers, J.N.J.J.; Koole, J.L.; Hurkmans, J. Properties of Cyclodextrins. VI. Water Soluble Cyclodextrins Derivatives. Preparation and Analysis. *Staerke* **1971**, *23* (5), 167–171.
143. Luna, E.; Tait, R.J.; Thompson, D.O.; Stella, V.J.; Stobaugh, J.F. Evaluation of the Utility of Capillary Electrophoresis (CE) for the Analysis of Modified Cyclodextrins Mixtures Proceedings of 7th International Symposium on Cyclodextrins. Osa, T., Komiyama Printing Co. Ltd.: Tokyo 1994; 198–201.

144. Olivier, P.; Verwaerde, F.; Hedges, A.R. Subchronic Toxicity of Orally Administered Beta-Cyclodextrins in Rats. *J. Am. Coll. Toxicol.* **1991**, *10* (4), 407–419.
145. Koizumi, K.; Kubota, Y.; Okada, Y.; Utamura, T. Microanalyses of Beta-Cyclodextrins in Plasma by High-Performance Liquid Chromatography. *J. Chromatog.* **1985**, *341* (1), 31–41.
146. Gerloczy, A.; Fonagy, A.; Keresztes, P.; Perlaky, L.; Szejtli, J. Absorption, Distribution, Excretion and Metabolism of Orally Administered 14C- β -Cyclodextrin in Rat. *Arzneim.-Forsch.* **1985**, *35* (7), 1042–1047.
147. Szabo, P.; Ferenczy, T.; Serfözy, J.; Szejtli, J.; Liptak, A. Absorption and Elimination of Cyclodextrin Derivatives by Rabbits and Rats Proceedings of 1st International Symposium on Cyclodextrins, Reidel: Dordrecht, The Netherlands 1982; 115–122.
148. Irie, T.; Tsunenari, Y.; Uekama, K.; Pitha, J. Effect of Bile on the Intestinal Absorption of Alpha-Cyclodextrins in Rats. *Int. J. Pharm.* **1988**, *43* (1/2), 41–44.
149. De Bie, A.; Van Ommen, B.; Bar, A. Disposition of [C-14]Gamma-Cyclodextrin in Germ-Free and Conventional Rats. *Regul. Toxicol. Pharmacol.* **1998**, *27* (2), 150–158.
150. Szathmary, S.C.; Seiler, K.U.; Luhmann, I.; Huss, H.J. Pharmacokinetic Behavior and Absolute Bioavailability of Hydroxypropyl Beta-Cyclodextrin after Increasing Doses in Volunteers Minutes of the 5th International Symposium Cyclodextrin. Duchene, D., Editions de Santé: Paris, 1990; 535–540.
151. Flourie, B.; Molis, C.; Achour, L.; Dupas, H.; Hatat, C.; Rambaud, J.C. Fate of Beta-Cyclodextrin in the Human Intestine. *J. Nutr.* **1993**, *123* (4), 676–680.
152. Antenucci, R.N.; Palmer, J.K. Enzymic Degradation of α And Beta-Cyclodextrins by Bacteroides of the Human Colon. *J. Agric. Food Chem.* **1984**, *32* (6), 1316–1321.
153. Antlisperger, G. New Aspects in Cyclodextrin Toxicology Minutes 6th International Symposium on Cyclodextrins Hedges, A.R., Ed.; Editions de Santé: Paris, 1992; 277–283.
154. Monbaliu, J.; Van Beijsterveldt, L.; Meuldermans, W.; Szathmary, S.; Heykants, J. Disposition of Hydroxypropyl Beta-Cyclodextrin in Experimental Animals. *Minutes 5th International Symposium on Cyclodextrins*; Duchene, D., Ed.; Editions de Santé: Paris, 1990; 514–517.
155. Gerloczy, A.; Antal, S.; Szatmari, I.; Muller-Horvath, R.; Szejtli, J. Absorption, Distribution and Excretion of Carbon-14 Labeled Hydroxypropyl Beta-Cyclodextrin in Rats Following Oral Administration. *Minutes 5th International Symposium on Cyclodextrins*; Duchene, D., Ed.; Editions de Santé: Paris, 1990; 507–513.
156. Szatmari, I.; Vargay, Z. Pharmacokinetics of Dimethyl- β -Cyclodextrin in Rats. *Proceedings of the 4th International Symposium on Cyclodextrins*; Huber, O., Szejtli, J., Eds.; Kluwer Academic Publishers: Dordrecht Netherlands, 1988; 407–413.
157. Uekama, K.; Hirayama, F.; Irie, T. Pharmaceutical Uses of Cyclodextrin Derivatives. *High Performance Biomaterials, A Comprehensive Guide to Medical and Pharmaceutical Applications*; Szycher, M., Ed.; Technomic: Lancaster, PA, 1991; 789–806.
158. Frijlink, H.W.; Visser, J.; Hefting, N.R.; Oosting, R.; Meijer, D.K.F.; Lerk, C.F. The Pharmacokinetics of Beta-Cyclodextrins and Hydroxypropyl- β -Cyclodextrins in the Rat. *Pharm. Res.* **1990**, *7* (12), 1248–1252.
159. Irie, T.; Uekama, K. Pharmaceutical Applications of Cyclodextrins. III. Toxicological Issues and Safety Evaluation. *J. Pharm. Sci.* **1997**, *86* (2), 147–162.
160. Mesens, J.L.; Putteman, P.; Verheyen, P. Pharmaceutical Applications of 2-Hydroxypropyl Beta-Cyclodextrin. *New Trends in Cyclodextrins and Derivatives*; Duchêne, D., Ed.; Editions de Santé: Paris, 1991; 369–407.
161. Yamamoto, M.; Aritomi, H.; Irie, T.; Hirayama, F.; Uekama, K. Pharmaceutical Evaluation of Branched Beta-Cyclodextrins As Parenteral Drug Carriers. *Minutes 5th International Symposium on Cyclodextrins*; Duchene, D., Ed.; Editions de Santé: Paris, 1990; 541–544.
162. French, D. The Schardinger Dextrins. *Adv. Carbohydr. Chem.* **1957**, *12*, 189–260.
163. Anderson, G.H.; Robbins, F.M.; Domingues, F.J.; Moores, R.G.; Long, C.L. The Utilization of Schardinger Dextrins by the Rat. *Toxicol. Appl. Pharm.* **1963**, *5*, 257.
164. Szejtli, J. Chemistry and Preparation of Cyclodextrins. *Cyclodextrins and Their Inclusion Complexes*; Akadémiai Kiadó: Budapest Hungary, 1982; 41.
165. Szejtli, J.; Sebestyen, G. Resorption, Metabolism and Toxicity Studies on the Peroral Application of Beta-Cyclodextrin. *Starch/Stärke* **1979**, *31* (11), 385–389.
166. Gergely, V.; Sebestyen, G.; Virag, S. Toxicity Studies of β -Cyclodextrin. *Proceedings of 1st International Symposium on Cyclodextrins*; Szejtli, J., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1982; 109–113.
167. Bellringer, M.E.; Smith, T.G.; Read, R.; Gopinath, C. Olivier, P. H. β -Cyclodextrin: 52-Week Toxicity Studies in the Rat and Dog. *Food Chem. Toxicol.* **1995**, *33* (5), 367–376.
168. Lina, B.A.R.; Bar, A. Subchronic Oral Toxicity Studies With Gamma-Cyclodextrin in Rats. *Regul. Toxicol. Pharmacol.* **1998**, *27* (2), 178–188.
169. Til, H.P.; Bar, A. Subchronic (13-Week) Oral Toxicity Study of Gamma-Cyclodextrin in Dogs. *Regul. Toxicol. Pharmacol.* **1998**, *27* (2), 159–165.
170. Frank, D.W.; Gray, J.E.; Weaver, R.N. Cyclodextrin Nephrosis in the Rat. *Am. J. Pathol.* **1976**, *83* (2), 367–382.
171. Frijlink, H.W.; Eissens, A.C.; Hefting, N.R.; Poelstra, K.; Lerk, C.F.; Meijer, D.K.F. The Effect of Parenterally Administered Cyclodextrins on Cholesterol Levels in the Rat. *Pharm. Res.* **1991**, *8* (1), 9–16.
172. Sharma, A.; Janis, L.S. Lipoprotein-Cyclodextrins Interaction. *Clin. Chim. Acta* **1991**, *199* (2), 129–137.
173. Matsuda, K.; Mera, Y.; Segawa, Y.; Uchida, I.; Yokomine, A.; Takagi, K. Acute Toxicity Study of Gamma-Cyclodextrin (γ -Cyclodextrin) in Mice and Rats. *Oyo Yakuri* **1983**, *26* (2), 287–291.
174. Schmid, G. Preparation and Application of Gamma-Cyclodextrin. *New Trends in Cyclodextrins and Derivatives*; Duchêne, D., Ed.; Editions de Santé: Paris, 1991; 27–54.
175. Donaubauer, H.H.; Fuchs, H.; Langer, K.H.; Bar, A. Subchronic Intravenous Toxicity Studies with Gamma-Cyclodextrin in Rats. *Regul. Toxicol. Pharmacol.* **1998**, *27* (2), 189–198.

176. Irie, T.; Otagiri, M.; Sunada, M.; Uekama, K.; Ohtani, Y.; Yamada, Y.; Sugiyama, Y. Cyclodextrin-Induced Hemolysis and Shape Changes of Human Erythrocytes in Vitro. *J. Pharmacobio-Dyn.* **1982**, *5* (9), 741–744.
177. Garay, R.P.; Feray, J.C.; Fanous, K.; Nazaret, C.; Villegas, M.J.; Letavernier, J.F. A New Approach for the in Vitro Evaluation of Cyclodextrin Effects on Cellular Membranes of Human Cells. *Minutes 7th International Symposium on Cyclodextrins*; Tetsuo, O., Ed.; Komiyama Printing Co., Ltd.: Tokyo, 1994; 373–376.
178. Leroy-Lechat, F.; Wouessidjewe, D.; Andreux, J.P.; Puisieux, F.; Duchene, D. Evaluation of the Cytotoxicity of Cyclodextrins and Hydroxypropylated Derivatives. *Int. J. Pharm.* **1994**, *101* (1/2), 97–103.
179. Bar, R.; Ulitzur, S. Bacterial Toxicity of Cyclodextrins: Luminous *Escherichia Coli* as a Model. *Appl. Microbiol. Biotechnol.* **1994**, *41* (5), 574–577.
180. Saarinen-Savolainen, P.; Jarvinen, T.; Araki-Sasaki, K.; Watanabe, H.; Urtti, A. Evaluation of Cytotoxicity of Various Ophthalmic Drugs, Eye Drop Excipients and Cyclodextrins in an Immortalized Human Corneal Epithelial Cell Line. *Pharm. Res.* **1998**, *15* (8), 1275–1280.
181. Serfozo, J.; Szabo, P.; Ferenczy, T.; Toth-Jakab, A. Renal Effects of Parenterally Administered Methylated Cyclodextrins on Rabbits. *Proceedings of 1st International Symposium on Cyclodextrins*. Szejtli, J.D., Ed.; Reidel Publishing Co.: Dordrecht, The Netherlands: 1982; 123–132.
182. Anderson, W.R.; Calderwood-Mays, M.; Brewster, M.E.; Bodor, N. The Effects of Chemically-Modified Cyclodextrins on the Renal Histology of the Sprague-Dawley Rat. *Minutes 6th International Symposium on Cyclodextrins*; Hedges, A.R., Ed.; Editions de Santé: Paris, 1992, 288–291.
183. Encapsin, HPB: R81216 Hydroxypropyl- β -Cyclodextrin—A Real Solution for Real Drug Delivery Problems. Product Literature from Janssen Biotech N. V. Drug Delivery Systems. Lammerdries 55, B-2250 Olen Belgium, 1992.
184. Vollmer, U.; Müller, B.W.; Peeters, J.; Mesens, J.; Wilffert, B.; Peters, T. A Study of the Percutaneous Absorption-Enhancing Effects of Cyclodextrin Derivatives in Rats. *J. Pharm. Pharmacol.* **1994**, *46* (1), 19–22.
185. Park, G.-B.; Seo, B.; Ann, H.-J.; Rho, H.-G.; Onn, Y.-S.; Lee, K.-P. Enhanced Nasal Absorption of Ketoconazole by Inclusion With Cyclodextrin. *Yakche Hakhoechi.* **1994**, *24* (2), 95–104.
186. Captisol®. *Sulfobutyl Ether β -Cyclodextrin*; Product Literature from CyDex, Inc.: 12980 Metcalf Ave., Suite 470, Overland Park KS 66213, 2000, CDinfo@cydexinc.com..
187. Coussement, W.; Van Cauteren, H.; Vandenberghe, J.; Vanparys, P.; Teuns, G.; Lampo, A.; Marsboom, R. Toxicological Profile of Hydroxypropyl Beta-Cyclodextrin (HP Beta-Cyclodextrins) in Laboratory Animals. *Minutes 5th International Symposium on Cyclodextrins*; Duchene, D., Ed.; Editions de Santé: Paris, 1990; 522–524.
188. Brewster, M.E.; Bodor, N. Parenteral Safety and Uses of 2-Hydroxypropyl Beta-Cyclodextrins. *Minutes 5th International Symposium on Cyclodextrins*; Duchene, D., Ed.; Editions de Santé: Paris, 1990; 525–534.
189. Ladola, A.; Abbott, D.; George, C.; Guzzie, P.; Nahas, K.; Provost, J.-P. Safety Studies With Captisol. Symposium on Pharmaceutical Applications of Cyclodextrins, Lawrence: KS, June-July 29-2 1997.
190. Van Cauleren, H.; Lampo, A.; Lammens, L.; Benze, H.; Coussement, W.; Vandenberghe, J. Monitoring Pharmacodynamic Effects in Toxicology Studies. *Excerpts From a Workshop on The Use of Pharmacology Studies in Drug Safety Assessment—Present Situation and Future Perspectives*; Sundwall, A.; Johansson, B.; Lindbom, L.-O.; Lindgren, E.; Sjöberg, P., Eds.; 1994; 101–106.
191. Sundaram, S.; Dayan, A.D. Effects of a Cholecystokinin Receptor Antagonist on Rat Exocrine Pancreatic Response to Raw Soya Flour. *Human Exp. Toxicol.* **1991**, *10*, 179–182.
192. Expert Working Group (Safety) of ICH. FDA Guidance for Industry ICH-S1C: Dose Selection for Carcinogenicity Studies of Pharmaceuticals, Drug Information Branch, HFD-210, CDER, 5600 Fishers Lane, Rockville, MD 20857 accessed Sept. 22, 1998, <http://www.fda.gov/cder/guidance/index.htm>.
193. Waalkens-Berendsen, D.H.; Smits-Van Prooijje, A.E.; Bar, A. Embryotoxicity and Teratogenicity Study with Gamma-Cyclodextrin in Rabbits. *Regul. Toxicol. Pharmacol.* **1998**, *27* (2), 172–177.
194. Barrow, P.C.; Olivier, P.; Marzin, D. The Reproductive and Developmental Toxicity Profile of Beta-Cyclodextrin in Rodents. *Rep. Toxicol.* **1995**, *9* (4), 389–398.
195. Möller, H.; Oeser, W.H. Drug Master Files, Global Harmonization of Quality Standards, Wissenschaftlich Verlagsgesellschaft MbH: Stuttgart, 1992; 213.
196. *The Japanese Pharmaceutical Excipients Council. Japanese Pharmaceutical Excipients 1993*. Yakuji Nippo, Ltd.: Tokyo, 1994; 415.
197. Ed. *United States Pharmacopeia 24/National Formulary 19*, United States Pharmacopeial Convention, Inc.: Rockville, MD, 1999; 24/19, 2569.
198. Janssen Research Foundation, Itraconazole Oral Solution, NDA 020657, Food and Drug Administration, Freedom of Information Staff (HFI-35), 5600 Fishers Lane, Rockville MD 20857, <http://www.fda.gov/>.

COMPLEXATION: NON-CYCLODEXTRINS

Galina N. Kalinkova

Medical University, Sofia, Bulgaria

INTRODUCTION

Complexation processes, also known as complexation, are based on the ability of many well-known drugs to interact and to form new complex drugs with altered properties in comparison with a drug alone. The pharmaceutical technology and the pharmaceutical industry have long considered research and development in the area of complexation a priority. The complexation process offers new possibilities for the improvement of existing drugs (side effects, therapeutical activity, and solubility). Such drug complexes with optimized characteristics can be prepared by complexation as a result of various interactions as drug–metal ion, drug–drug, drug–excipient(s), etc.

Today, two of the greatest advances in pharmaceutical technology can be found in the field of complexation processes. These are chelathotherapy and biotechnology. The significance of chelathotherapy is evident in relation to recent increased problems connected with the pollution of the environment.

Biotechnology in relation to the pharmaceutical industry ensures new special drugs, some of which include proteins, antibodies, and peptides. Others include insulin, interferons, growth factors (GFs), and sensitive diagnostics for diseases such as hepatitis, AIDS, and herpes.

It is interesting to note that the history of medical treatment with metal ions and their compounds has been known for thousands of years. In 2500 B.C, the Chinese used gold (Aurum, Au) for medical treatment, whereas in the middle of this past millennium, the gold's compounds were considered an effective treatment against leprosy.

COORDINATION THEORY AND SOME RECENT THEORETICAL CONSIDERATIONS

The coordination theory was promoted by the Swiss chemist Alfred Werner (1866–1919) in 1891. He became a professor in Zurich in 1893, and in 1913, he received the Nobel Prize in Chemistry for his investigations of complex compounds (1). Werner found many nonorganic

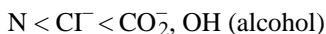
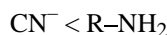
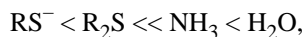
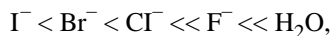
compounds with asymmetrical molecules that were also optically active in solutions. Such complex compounds include Co, Cr, and Fe.

The development of the chemistry of complex compounds was also promoted by the Scandinavian scientists K. Blomstrand and S. Jorgensen (1837–1914).

Gilbert Newton Lewis (1875–1946), University of California, contributed to the development of the electronic theory of the valence. He also carried out investigations on absorption spectra of organic compounds.


























The atom of the complex-constitutor is the so-called central atom or central ion. In order to emphasize the difference between the central ion of the same metal in a free state, the central ion is marked with the symbol of the chemical element followed by its valency in roman numerals in brackets. Molecules and ions directly bound with the central atom are called coordinated groups (Cl, NH₃, H₂O, etc.) or intraspherical substituents (ligands or addenda) (2–4).

Ligands are “hard” or “soft” (5). The former are electronegative with electrostatic interactions, low delocalization of electron density, and formation of covalent bonds with cations. They include F[−] ions and H₂O molecules. “Soft” ligands are polarizable, covalent bonds, such as chloride, bromide, iodide, sulfur-containing ligands, imidazole, etc. This division into “hard” and “soft” ligands is conventional. It is more correct to consider the series of ligands in their increasing hardness:



It should be noted that complexing ions are characterized with definite steric structures. The most widespread drug complexes with coordination number 6 are built according to the type of the octahedron, with the

Table 1 Drug configurations

Metal ion	Coordination number	Polyhedron type	Schematic presentation
Cu(I)	4	tetrahedron	
	(2)	(chain)	
	(3)	(triangle)	
Cu(II)	4	square	
	6	distorted octahedron	
	(4)	(distorted tetrahedron)	
	(5)	(square-pyramid or trigonal-bipyramid)	 
Ni(II)	4	square	
	4	tetrahedron	
	6	octahedron	
	(5)	(trigonal-bipyramid)	
Co(II)	4	tetrahedron	
	(4)	(square)	
	(5)	(trigonal-bipyramid)	
Co(III)	6	octahedron	
	(4)	(tetrahedron)	
	(5)	(square-pyramid)	
Fe(II)	6	octahedron	
	(4)	(tetrahedron)	
Fe(III)	6	octahedron	
	(4)	(tetrahedron)	
Mn(II)	6	octahedron	
	(4)	(tetrahedron)	
	(4)	(square)	

The parentheses used in the second column mark the addition of nonordinary states.

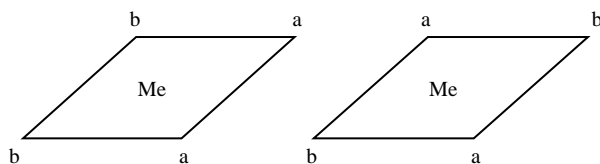


Fig. 1 Steric structures of *cis*- and *trans*-isomers.

coordinated groups oriented to the octahedron's peaks (Table 1). The drug complexes with coordination number 4 also can be built on the type of tetrahedron or on the type of plane (Table 1).

The steric structures presented explain the early detection of isomery (isomerism) of the complex compounds of Co, Pt, etc., as well as predict the number of geometric(al) isomers of the complex ion. For example, complex ions Mea2b2 may exist as two geometric isomers (*cis*- and *trans*-). The content's complication of the complex ions increases the number of geometric isomers (Fig. 1).

Consideration of the octahedral model in accordance with symmetry knowledge also has been used to predict the presence of the mirror isomerism in complex compounds with definite content and structure. Its discovery was made by Werner in 1911, and is a confirmation of the coordination theory (2, 3).

The first steric ideas were based on the pure chemical ways. The number of the isomers in the substitution reactions was determined by their comparison. These models were confirmed by X-ray diffraction analysis.

COMPLEXES FORMED BY INTERACTIONS WITH METAL IONS

Azathioprine, an immunosuppressant and cytostatic drug, forms complexes with various metal ions (6). The complexes have been investigated by the potentiometric titration method, infrared (IR) spectroscopy, etc. Azathioprine is found to form 2:1 complexes with Co(II), Cu(II), and Ni(II). The order of their stability is established as Cu(II) > Ni(II) > Co(II). Azathioprine-metal complexes have been proven by their IR spectra. The evidence of the effect of complexation was established by the absence of an absorption band at 3191 cm^{-1} (N-H stretching characteristic of a purine function in Azathioprine); e.g., the aminogroup is involved in the formation of the metal complex. The anticancer action of purine derivatives was discovered in 1949. This cytostatic drug biotransforms to 6-mercaptopurine in the body (6).

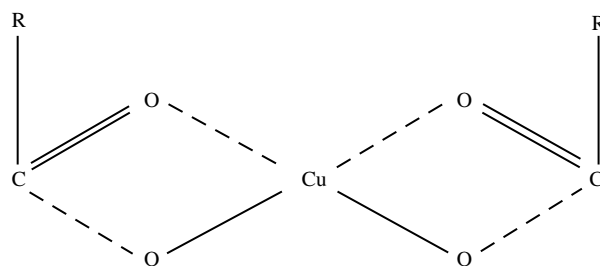


Fig. 2 The complex formation between Cu(II) ions and carboxylic groups in penicillins.

Drug interactions often constitute major problems in drug therapy. This occurs when metallic therapeutic agents and drugs (able to form complexes) are administered simultaneously (7). The metallic compounds used in pharmaceutical preparations and included in pharmacopoeias are as follows: magnesium trisilicate, aluminum hydroxide, ferrous sulfate, calcium carbonate, sodium bicarbonate, and potassium citrate. Norfloxacin, a 4-Quinolone carboxylic acid derivative with antimicrobial activity, has shown significant interaction with compounds containing Fe^{2+} , Mg^{2+} , and Al^{3+} which have led to less active complexes (unabsorbable and/or antibacterially inactive). Such complex formation is proposed as a factor responsible for the alteration of drug activity by simultaneous co-administration with metallic medicinal agents (7).

Platinum(II) complexes have antitumor activity, and have been tested against P-388 leukemia (derivatives of the substituted *o*-phenylenediamine). Their antitumor activity has been connected with many factors. These include the formation of chelate rings and their strength, the nature and the influence of different substituting groups, and the relative stability of the Pt(II) complexes (8).

Copper(II) complexes of penicillins (benzylpenicillin, phenoxymethylenepenicillin, ampicillin, amoxicillin, and carbanicillin) have been prepared. They have shown stoichiometries of the type CuL^* , CuL_2^* , and CuL_3^* (where L^* penicilloic acids). Structures of coordination penicillin compounds have been suggested by IR spectroscopy as monoatomic bidentate ligands coordinating the Cu(II) ion through the carboxylic groups (Fig. 2).

Copper(II) ions have promoted the hydrolysis of penicillins to corresponding penicilloic acids owing to β -lactam group. The same process has been confirmed between Cu(II) ions and carboxylic groups of cephalosporins, which also belongs to β -lactam antibiotics (9).

Advances include radioprotective drugs applied during radiotherapy of neoplastic diseases (10). Effects similar to

the enzyme superoxide dismutase have been found in copper complexes of Schiff's bases (derived from different amino acids and salicylaldehyde). Activity of complexes is dependent on their structures. The structural changes of their chelate rings are responsible for their effects on free radicals produced in organisms during radiation. Complexes have square pyramidal pentacoordination, which is similar to the coordination polyhedron in the active center of the Cu-dependent superoxide dismutase. The complexes used as radioprotective drugs play an important role in radiotherapy. They protect healthy tissues and cells from injurious radiation (10).

The first natural product in which the presence of the metal cobalt (Co) was shown is Vitamin B₁₂ (cobalaminum). It was isolated from the liver simultaneously by K. Fokers (United States) and Lester Smith (England). The Vitamin B₁₂ structure was established by Todd in 1955. The center of the molecule contains the metal Co(III) with a coordination number 6. It is connected with four mutually bound pyrrol cycles. Vitamin B₁₂ is used in the treatment of pernicious anemia.

COMPLEXES FORMED BY INTERACTIONS WITH EXCIPIENTS

Naltrexone (11) is a potent narcotic antagonist approximately 30–40 times more active than nalorphine, and 2–3 times more active than naloxone. Polymeric complexes as a result of hydrogen bond interactions between Naltrexone and Eudragit[®] have been studied. High performance liquid chromatography (HPLC), ultraviolet (UV) spectrophotometry, scanning electron microscopy (SEM), ¹H- and ¹³C-nuclear magnetic resonance (NMR), differential scanning calorimetry (DSC), and hot stage microscopy (HSM) have been used to characterize naltrexone polymeric complexes. A very high efficiency in the dissolution process, as well as a significant reduction in the drug release rate from the complex, has been observed (11).

Eudragit[®] L (12) is a polymer with anionic character that is based on the methacrylic acid and its methyl ester (in ratio 1:1 approximately; m.w = 135000). Eudragit[®] L is widely used in drug preparations. A proposed complexation process allows the obtainment of Eudragit[®] L-Cardiol complexes. The nature of the latter was established through the use of spectroscopic techniques (IR, ¹H, and ¹³C-NMR). The complex was characterized as intermolecular association (polymer–drug), and the drug interacted with ammonium (maximum cardiolol salt content in complex—22%). Comparative studies of cardiolol and

morphine complexes have shown differences. These differences are explained by the different chemical structures related to the amino group of the two drugs (12).

The excipients ethylcellulose (13) and pectin (14) show the possibilities for complex interactions of the type intermolecular H-bonds. The latter result in increased therapeutic activity of the amoxicillin trihydrate (granules) and nystatin (plaque). The nature and complex character of these interactions have been investigated by means of IR spectroscopy and X-ray diffraction (13, 14).

Indomethacin was introduced into medical practice in 1963. Its anti-inflammatory activity is much greater than aminophenazon and hydrocortisone, and it is a much stronger antipyretic than aminophenazon. The properties of indomethacin in complex formations have been studied. The drug interacts with various agents, such as zinc (15), calcium glycerophosphate (CGP) (16), polycarbophil (17), chitosans (18), 2-(*N,N*-dimethylamino) propionate (19), and eprizole (20). The indomethacin complexes show better characteristics when compared to the drug alone. Improvements in water solubility, increased dissolution and absorption rates, and increased bioavailability have been proven, as well as decreased side effects (gastrointestinal ulceration and hemorrhage). Some of the zinc–indomethacin complex characteristics are as follows: m.p.—232°C (decomposition), IR spectrum—1586 cm⁻¹ (asymmetrical stretching of carboxylate anion), ligand and metal ratio 2:1, and two molecules of crystal water (15).

The preparation methods of Indomethacin complexes and their advantages have been described in detail. These include: 1) Zn–indomethacin (15)—economical and less time consuming; 2) indomethacin–CGP—simplicity, facility of design, and nearly cost-free production (of the complex) on a large scale (industrially (16); 3) the concentration of polycarbophil can be used as a key step in suppositories (17); and 4) indomethacin–epirizole complex (20)—new spherical crystallization technique applied without further processing as granulation.

Propranolol–methacrylic acid copolymer complex (21) has been evaluated as a potential prolonged releasing drug. Propranolol content is found to be 68% in the complex. The complexation process has been defined as positive interaction of a high degree between propranolol and the polymer (specific ion–ion interactions and hydrophobic binding to the overall complexation process). The slow release of propranolol from the complex may be due to hydrogel formation when the drug (the tablet) is exposed to the dissolution medium. The complex has been investigated by differential thermal analysis (DTA) and IR- and UV-spectroscopic methods.

Chitosan is a cationic polymer used for controlled drug delivery. It forms polyion complexes (interpolymer) as a result of its interactions with anionic polymers. The polyion complexes and their basic properties have been investigated for their pharmaceutical application (22). The specific properties of the complexes (chitosan–sodium alginate and chitosan–sodium acrylate) are due mainly to rigidity or flexibility of the polymer chains. The former is stable to pH change, and the latter is quite sensitive to pH change, which makes them applicable to the design of more precisely controlled drug delivery systems. Fourier transform infrared spectroscopy (FTIR), elementary analysis, and viscosity measurements have been used to explain the nature of these complexes (22).

Recently, the tranquilizer action of phenothiazine derivatives has been connected with the flexibility of their molecules (23). They form complexes with charge transition. These complexes have been obtained as a result of the interactions between phenothiazine derivatives, dextrans, and pectins. IR spectroscopy, X-ray diffraction, UV spectroscopy, and Dreiding models (a 3-D research model) have been applied (24). Hypochromic effects (changes in the band's intensity) in UV spectra have been observed. The degree of complex binding correlates with the concentrations of dextrans.

COMPLEXES FORMED BY DRUG–DRUG INTERACTIONS

A new complex of Ind and epirozone (molecular ratio equal to 2:1) has been prepared by the spherical crystallization technique and proven by means of IR spectroscopy and X-ray diffraction (20).

ANALYSIS

Drug complexes require application of numerous methods of analysis (physico-chemical, biological, etc.) for their complete characterization (25). The choice and combination of these methods are connected mainly with their effectiveness in establishing the type of the complexes as well as proving their mechanism. Today, the commercial availability and economical convenience (26, 27) of the drug complexes (their preparation, analysis, and manufacturing) propose a new strategy and balance for pharmaceutical technology (the pharmaceutical industry, respectively). In this context, the combination of IR spectroscopy and X-ray diffraction are methods of choice (28).

Complexes obtained by interactions between aliphatic amines and carboxylic acids (23) have a structure type of the ion pair and complex composition 1:1. Their IR spectra do not have the characteristic absorption band for free $\nu\text{C=O}$ in the region from 1780 to 1700 cm^{-1} . However, new characteristic absorption bands appear for the carboxylate anion in 1680–1560 cm^{-1} ($\nu_{\text{as}} \text{COO}^-$) and 1400–1300 cm^{-1} ($\nu_{\text{s}} \text{COO}^-$). Some additive bands also appear: The vibration NH^+HO^- (in the region 2800–2200 cm^{-1}), as in the salts in solid state, and the bands δNH_2^+ or δNH_3^+ at 1620–1600 cm^{-1} , and δCO_2 at 670 cm^{-1} (where ν is stretching or valence vibration and δ is deformation vibration). The complex with composition 2:1 can be formed as a result from the addition of a second acid molecule to an ion pair (in increased amine concentration). Characteristic absorption band of the carboxylate anion shifts to the lower values of the wave numbers (as compared to the complex with 1:1 composition) (23).

Electron Paramagnetic Resonance

Electron paramagnetic resonance (EPR) gives basic information on the complexes containing Cu and Fe ions (in drugs or proteins) mainly for the character of their bonds and ligands. Free organic radicals also give EPR signals, which makes interpretation of spectra very complicated (29).

Electron Spectra (UV–Visible Spectra)

Usually, the intensive bands appear in UV–visible regions that are connected with the charge transition (between ligand and metal ion or between ligand atoms). The band positions (in charge transition) are dependent on the nature of metal and ligand as well as on their relative ability to oxidation and restoration) (23).

Nuclear Magnetic Resonance (NMR)¹H

These spectra are applicable in conformation analysis, e.g., stereochemistry. They are used to provide very necessary information on the role of metal and on the nature of metal complexes.

Complexometry

This is a quantitative (titrimetric) analysis for drugs containing bismuth (Bi), calcium (Ca), magnesium (Mg), plumbum (Pb), and zinc (Zn), among others. The titration is performed with solutions of polyaminocarboxylic acids and their salts, the so-called complexons (I, II, and III or Trilon B, respectively). They are able to form stable,

water-soluble complexes in stoichiometry (1:1) with the drug metals. The titration is carried out in the presence of one of the metal-indicators used (eriochrome black T, murexide, or xylenolorange). The requirements for the metal indicators include their reversible interactions with the metals of drugs analyzed, unstable complex formation, and color differences in free or complex states (2). Therefore, complexometry is an analysis in principle based on the complex formation ability of some drugs.

CHELATOTHERAPY: COMPLEXATION PROCESSES IN THE HUMAN ORGANISM

The treatment of poisoning with metals includes the use of drug chelating agents. The latter form stable and nontoxic complexes with metals and are quickly eliminated from the body.

Criteria of the Drug Choice as Chelating Agents

Chelathotherapy requires comparative evaluation of the chelating drugs. The following must be taken into consideration when choosing chelating drug agents for a given metal: 1) stable binding with the metal so that it may be concurrent to the biological ligands in the organism; 2) selectivity; and 3) nontoxicity.

The selectivity of the drug ligand to a definite metal can be evaluated by analysis of the formation constant. In cases where the selectivity is decreased or absent, unwanted side effects have been observed as a result of the complexation with other metals (such as calcium and zinc). The chelating drug agents attack heavy metals. In the organism, the metal will be in a lipophilic medium. When the ligand is lipophilic, the lipophilic complex cannot have toxic action. If the ligand is adequately lipophilic, it can penetrate the membrane of the depo metal forming a lipophilic complex (29).

A discussion of some chelating drugs follows.

Dimercaprol (BAL): This drug preparation appeared during the World War I and was used for treatment of battle poisoning along with luisit ($\text{ClHC}=\text{CHAsCl}_2$). The luisit action is based on the binding with $-\text{SH}$ groups.

BAL as a drug ligand makes stable bonds with arsen (As) and, thus, is able to remove it from the luisit molecules. BAL also has been used in the treatment of poisoning with other metals (Hg, Ca, Au, Ti, Tl, and Bi). BAL forms solid (strong) bonds with a series of metals; consequently, it must be used in low concentrations. The use of BAL requires more attention since the organic

compounds of Hg increase their concentrations in the brain in the early stages of the poisoning (30).

EDTA (30): In this drug preparation, calcium is bound as complex and can be displaced from the ions of heavy metals. These metals are bound loosely in the biochemical systems of the tissues and are liberated easily from them.

Water-soluble stable complexes with low toxicity are formed with EDTA and are very rapidly eliminated with urea. EDTA is given by injection in acute and chronic poisoning with Pb, Cd, Co, Hg, U, It, and Ce.

Penicillamine: Penicillamine is used in the treatment of cronic copper accumulation (Wilson's disease).

X-ray diffraction investigations explain the structure of this complex. The central ion is a halogen surrounded with eight Cu(I) atoms with sulfur donors. The coordination sphere of the Cu(II) is completed with amino groups of the penicillamine (29).

Anticancer drugs: Anticancer drugs and their metal complexes appear as *cis*- and *trans*-isomers. The *cis*-isomers of the dichlordiaminoplatin(II) (*cis*-DDP) and the tetrachlordiaminoplatin(IV) are known as the most effective complexes. The former *cis*-isomer of platinum(Pt;II) has a large application in the treatment of cancer of the ovaries and testicles. *Trans*-complexes have shown toxicity and do not possess non-anticancer action.

All of the anticancer drugs (*cis*-DDP) have side effects, such as nausea, diarrhea, decrease of the hemoglobin levels, and destruction of the kidneys. The latter is connected with the difficulty of creating and supporting the high liquid content in patient organisms during treatment.

Thus, it is extremely important to determine the most effective methods that allow in vivo low drug concentrations. Some basic protocols have been established on restoration of in vivo complex Pt(IV), disintegration polymeric platinum compounds, and slow evacuation of blocking inert ligands. Other platinum complexes also have been examined (29).

The method of circular dichroism has been applied in investigations that have determined the reactions between Pt(II) compound and nucleosides, nucleotides, and DNA. Many antimicrobial preparations are excellent ligands. The activity of some of the antimicrobial preparations has been based on their complexing with metal ions. For example, increased effectiveness of tetracycline has been observed after its coordination with Ca^{2+} . The complex is more lipophilic and very oil-soluble, which explains its transportation through the cell membrane. The opposite situation occurs when the ion-metal is toxic and the coordinated antibiotic serves as carrier through the membrane.

Diuretic drug preparations have promoted urine formation. They are derivatives of mercury propanol $\text{RCH}_2\text{CH}(\text{OH})\text{CH}_2\text{HgX}$, where R is a polar hydrophilic group. The mercury diuretic preparations act as ferment inhibitors (latter containing—). They also inhibit adenosine triphosphate (ATP). These properties led to the use of mercury drug compounds in the treatment of bacterial infections. In these cases, they interacted with —SH groups of the bacteria proteins.

Interactions between Drug Preparations and Metal Ions in Humans

Many drug preparations can interact with metal ions. Therefore, they function in humans as chelating agents and form complexes. In some cases the therapeutic activity of a drug compound can be explained by inhibition through a chelated metal ion. For example, disulfiram (tetraethylthiourea disulfide) is used in the treatment of chronic alcoholism. It inhibits aldehyde oxidase, and the ethanol metabolism is interrupted in the stage of formation of the acetaldehyde. The latter leads to an unpleasant feeling. It makes the alcohol (ethanol) nontransportable in its later use. The complex with metal ion can lead to formation of the neutral oil-soluble particles able to penetrate the cell membranes. The metal ion acts as a carrier of the drug substance preparation (29).

Drug preparations able to interact with metal ions include acetazolamide, amphetamine, aspirin, ethambutol, phenacetin, nialamide, disulfiram, and thioacetazone. The evaluation of new drugs includes the emphasis of the following important requirements: 1) the drug's strength and the ways (or mechanisms) of its action; 2) the side effects of the drug and its metabolites; 3) its stability in vivo; and 4) its possible interactions with other drugs.

Consideration of these requirements is extremely important. The possible interactions between two or more drugs, taken simultaneously, make their use dangerous or change their therapeutic activity (either increase or decrease). Only the nonionized drug compounds are oil-soluble. Sometimes, drug preparations do not appear active since some of their metabolites are in pharmacologic active forms. Drug characteristics are connected with the whole drug molecule or with receptors. The degree of ionization of the drug in vivo has been determined by the values of pK_a and pH of the organism liquids. The effectiveness of the sulfonamides against infections is stimulated by the presence of the NH₂ groups. The latter prevents the growth and the multiplication of bacteria. The applications of the metal complexes against numerous microorganisms and against strains highly resistant to the traditional antibiotics have

great importance (29). The form and the distribution of the charges in the complex give information concerning the necessary pharmacological properties.

Metals as Drugs

Lithium (Li) has been used in the treatment of bipolar disorder (manic depression) for approximately 50 years. Over 2000 British patients use Li for this disease. The illness is characterized by alternative states of depression and mania, or overexcitement. At times, the cycle of mood changes continue for several weeks or for one year with good results (intervals) between the phases. In Texas, a correlation has been noted between the influence of Li-salts and the drinking water as well as the increase in visits to psychiatric (mental) public health hospitals.

The compounds of gold were first applied in the treatment of rheumatoid arthritis in 1927, and are used even today in the treatment of severe cases.

Thiomalate (salt of ester of the malic acid) also has been used very successfully to treat severe cases of rheumatoid arthritis. A complete cure or a significant improvement has been noted in approximately 50% of patients using this treatment modality, with 40% of patients exhibiting side effects. Thiomalate is also toxic and possesses cumulative action. Minimal concentrations of the gold compounds reach many cells and remain in the body for years (29).

Complexing and Nonmetals as Microelements

Many nonmetals are important microelements necessary for the normal function of organisms. Some polyvitamin preparations contain microelements as microadding. Their therapeutic and toxic action depends on the concentration. Therefore, the presence of these microelements in the organism in proper amounts is extremely important. Examples of nonmetals include boron (B), silicon (Si), selenium (Se), arsenic (As), chlorine (Cl), fluorine (F), and iodine (I).

Studies on the influence of boron have shown that its compounds as borate have inhibited some fermentative reactions. In addition, the influence of boron results in membrane destruction (29).

The lack of silicon (Si) leads to destruction of bone structure and connective tissues (29).

Deficiency of selenium (Se) influences the concentration of vitamin E. The latter is an antioxidant and protects membranes from oxidation. The ability of Se to protect organisms from the poisoning with Hg or Cd are well known. The normal concentration of this microelement is 0.09–0.2 mg/cm³ (29). Interestingly, the

concentration of Se in the blood of New Zealanders shows a decrease (0.068 mkg/cm^3).

Fluorine (F) and its metabolites are of importance in protecting teeth from caries. Fluorine is included in calcium hydroxyapatite, and it promotes the precipitation of calcium phosphate $\text{Ca}(\text{PO}_3)_2$ and accelerates the remineralization. The necessary concentration of Fluorine added to drinking water to prevent caries is approximately 1 mg/L . Application of higher Fluorine concentrations (above 8 mg/L) leads to fluorosis. This is a disease that is characterized by a disturbance in the function of the thyroid gland. A long-term application of fluorine leads to intensive mineralization (possible precipitation of calcium sulfate), deformation of bones with possible accretion, and calcification of the connections (29).

Iodine (I) has a great role in the function of the thyroid gland. It takes part in a complex biochemical synthesis schema and interacts with hormones. Deficiency of iodine is characterized weakness, feelings of cold and dryness, and yellow-colored skin. The treatment of these symptoms is achieved with iodine or thyroid hormones (29).

BIOTECHNOLOGY

Drugs as special products of biotechnology have an important role in the pharmaceutical industry. Biotechnology is based on the progress of genetic engineering and fermentation technologies. These advances make possible a host of new products, some of which include peptides, proteins, and antibodies, and are used as therapeutics or diagnostics.

Potential drugs, such as interferons, interleukins (as well as other lymphokines), growth factors, and plasminogen activators are also being used as diagnostics. They are very sensitive to many diseases, such as hepatitis, herpes, and AIDS (31).

The beginning of biotechnology stems from 300 small biotechnology companies (often with roots from academic laboratories) that operated in the late 1970s. Today, the list of large American biotechnology companies is long. In addition, the governments of the United Kingdom, France, Germany, and Japan have launched major efforts to assist their countries' development of biotechnology. Some of the large European biotechnology companies include Novo Industri in Denmark and Celltech in the United Kingdom, but the majority of biotechnology companies are in the United States (approximately 75%).

The properties of insulin (5) have been investigated in many ways—as a drug, a protein, and a hormone. Its complex chemical structure has been proven (two

polypeptide chains bound by means of two disulfide bridges). Insulin appears as monomers, dimers, tetramers, or hexamers, depending on the number of its associated molecules. Thus, insulin monomers form dimers, which aggregate in hexamers in the presence of zinc (Zn) in neutral pH. Insulin in concentrations 10^{-5} M (molar) exists as dimers. Many of the latter have been formed as a result of van Der Waal's interactions. Insulin molecules are also bound with hydrophobic forces and hydrogen bonds. The complex chemical structure of insulin explains its ability to take part in complexation processes. Heavy atomic reagents, together with insulin, form various complexes (e.g., with cations of Lantanides, lead [Pb], thallium [Tl], and uranium [divalent group: UO_2]). The latter possesses a linear geometry ($\text{O}=\text{U}=\text{O}$)²⁺. Insulin complexes have been formed with different mercury (Hg) reactants. T.B. Blundell is well known as the British scientist who deciphered the structure of insulin. (He is also known for his work on the building of proteins.) Crystallography of insulin and its complexes have been studied by means of X-ray diffraction and electron microscopy.

The presence of Calcium (Ca^{2+}) is necessary for blood coagulation and subsequently prevents hemorrhaging from tissue injuries. The mechanism consists of cascade-type process in which stages are connected with the presence of Ca^{2+} . Many of the so-called factors of blood coagulation are well known. Vitamin K is necessary in the biosynthesis of the factors IX, X, and VII and of prothrombin. The 1,25-dioxiform of vitamin D_3 facilitates the process of Ca^{2+} reception from the intestines. Accumulation of Ca^{2+} and its release from humans is a complex system, and also includes vitamin D.

Almost all proteins (ferments; redox-, transport- and spare-proteins; hormones; and antibodies) have been studied by X-ray diffraction analysis (5).

Ferments are catalysts. They change the rapidity of the biochemical reactions without displacing the equilibrium ratio.

Redox-proteins are connected with cell organelle as mitochondria, chloroplasts, membranes with specific functions, and high oxidation–reduction potential.

Transport- and spare-proteins (the hemoglobin with its four subunits) ensures the rapid transport of oxygen to a definite place. They have been studied by X-ray diffraction analysis since 1937.

Hormones regulate intracellular metabolic processes. Insulin intensifies transfer (transport) of the glucose in the cell and decreases its level in the circulating blood. Glucagon increases the sugar level in the blood. The two hormones are synthesized by the pancreas. Antibodies discern foreign bodies in an organism.

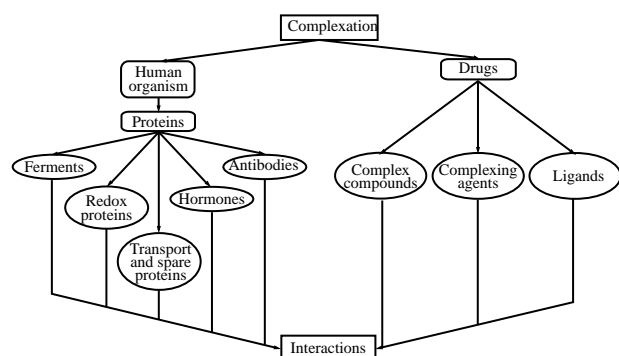


Fig. 3 Schematic presentation of the complexation processes—drugs and human organism interactions.

DISCUSSION

The preparation of drug complex compounds and the use of their qualitative properties have enormous importance in the pharmaceutical industry. Complex compounds have also taken part in the processes of the vital activity of humans as inner-complex compounds in hemoglobin, tissues, etc. (Fig. 3). Complex bound metals are important components of the ferments (in particular, the oxidative ferments).

A systematic view on the topic of “complexation” and complexes in the literature shows a great number of important research results. This expanding field covers studies of various systems, including antibody–antigen complexes, proteinase–inhibitor complexes (32), lipid-based delivery vehicles, metal complexes with DNA base pair, and aqua ligand in the coordination sphere of metal (33). The metal–modified structures are dominated by these metal-base interactions. However, the structural role of the water molecules in the complexes is quite apparent, as suggested by crystallographic studies (33).

A possibility for the next generation of therapeutics is connected with the complexation processes of proteins with virus membranes, which destabilizes them (34). This knowledge is of interest for future use in genetic medicines. An area of much promise is complexes of polycations with DNA. They result in major improvements. Cationic liposome-based gene delivery accounts for 9–12% of ongoing gene therapy clinical trials in the United States and Europe (35).

Today, many computer programs assist in the investigation of new drug ligand designs, geometry of interactions, and conformational flexibility of the potential ligands (26).

CONCLUSION

Hughes (29) is quite right when he concludes “It is evident today that the metal ions are controlling (are checking) the various biological processes and that life is based on the organic and inorganic chemistry.” Finally, modern pharmaceutical technology has successfully applied the advanced knowledge of many sciences (chemistry, biochemistry, molecular biology, pharmacology, etc.) to solve the various complex problems of drug manufacturing (the pharmaceutical industry) and of drug development. These scientific approaches, to integrate and use all the best of human thought for human health, confirm the pharmaceutical industry as an important interdisciplinary science today and in the future.

Recent research on drug complexation can be described as a tunnel that is being dug from two opposite ends with the efforts of many scientists. On one side are those who solve the problems with drug complexation, while on the other side are those who solve them with protein complexation. When they meet in the middle of the tunnel, the mechanisms of the interactions between drugs and human organisms will be elucidated.

REFERENCES

1. *Meyers Neues Lexikon*; VEB Bibliograph Institute: Leipzig Germany, 1977; 15, 155–155.
2. *Meyers Neues Lexikon*; VEB Bibliograph Institute: Leipzig Germany, 1973; 7, 679–682.
3. *Meyers Neues Lexikon*; VEB Bibliograph Institute: Leipzig Germany, 1974; 8, 43–44.
4. *Short Chemical Encyclopedia*; State Scientific Publishing House “Sovietskaia Encyclopedia”: Moscow, 1963; 2, 655–674.
5. Blundell, T.L.; Johnson, L.N. *Protein Crystallography*; Mir: Moscow, 1979; 9–262.
6. Singh, S.; Gulati, M.; Gupta, R.L.; Complexation Behavior of Azathioprine with Metal Ions. *Int. J. Pharm.* **1991**, 68, 105–110.
7. Okhamafe, A.O.; Akerele, J.O.; Chukuka, C.S. Pharmacokinetic Interactions of Norfloxacin with Some Metallic Medicinal Agents. *IJP* **1991**, 68, 11–18.
8. Kidani, Y.; Asano, Y.; Noji, M. Synthesis of Platinum(II) Complexes of 4-Substituted *o*-Phenylenediamine Derivatives and Determination of their Antitumor Activity. *Chem. Pharm. Bull.* **1979**, 11, 577–2588.
9. Sher, A.; Veber, M.; Marolt-Gomiscek, M. Spectroscopic and Polarographic Investigations: Copper(II)–Penicillin Derivatives. *Int. J. Pharm.* **1997**, 148, 191–199.
10. Valentova, J.; Gombosova, I.; Zemlicka, M.; Laginova, V.; Svec, P. Radioprotective Activity of *N*-Salicylideneaminoalkanoatocopper(II) Complexes. *Pharmazie* **1995**, 50 (6), 442–443.

11. Alvarez-Fuentes, J.; Caraballo, I.; Boza, A.; Llera, J.M.; Holgado, M.A.; Fernández-Arévalo, M. Study of a Complexation Process between Naltrexone and Eudragit® L as an Oral Controlled Release System. *Int. J. Pharm.* **1997**, *148*, 219–230.
12. Holgado, M.A.; Fernandez-Arevalo, M.; Alvarez-Fuentes, J.; Caraballo, I.; Llera, J.M.; Rabasco, A.M. Physical Characterization of Carteolol: Eudragit L Binding Interaction. *Int. J. Pharm.* **1995**, *114*, 13–21.
13. Gasheva, L.M.; Kalinkova, G.; Minkov, E.; Kretev, V. IR Spectroscopic Investigations of Amoxicillin Trihydrate Included in the Technological Models Sirup Granules in Ethylcellulose. *J. Mol. Struct.* **1984**, *115*, 323–326.
14. Kalinkova, G.; Krasteva, S. IR-Spectroscopic Investigation of Interactions in the Drug Preparation Nystatin-Plaque. *Recent Developments in Molecular Spectroscopy*; Jordanov, B., Kirov, N., Simova, P., Eds.; World Scientific: Singapore, 1989; 705–709.
15. Singla, A.K.; Wadhwa, H. Zinc–Indomethacin Complex: Synthesis, Physicochemical and Biological Evaluation. *Rat. Int. J. Pharm.* **1995**, (120), 145–155.
16. El-Monem, A.; Foda, Said, S. Effect of Complexation of Indometacin with Calcium Glycerophosphate on its Bioavailability, Ulcerogenicity and Antiinflammatory Activity in Rats. *Pharm. Ind.* **1991**, *53* (1), 94–97.
17. Hosny, E.A.; Al-Angary, A.A. Bioavailability of Sustained Release Indomethacin Suppositories Containing Polycarbophil. *Int. J. Pharm.* **1995**, *113*, 209–213.
18. Imai, T.; Shiraishi, S.; Saito, H.; Otagiri, M. Interaction of Indomethacin with Low Molecular Weight Chitosan, and Improvements of Some Pharmaceutical Properties of Indomethacin by Low Molecular Weight Chitosans. *Int. J. Pharm.* **1991**, *67*, 11–20.
19. Büyüktimkin, S.; Büyüktimkin, N.; Rytting, J.H. Interaction of Indomethacin with a New Penetration Enhancer, Dodecyl 2-(*N,N*-Dimethylamino) Propionate (DDAIP): Its Effect on Transdermal Delivery. *Int. J. Pharm.* **1996**, *127*, 245–253.
20. Kawashima, Y.; Lin, S.Y.; Ogawa, M.; Handa, T.; Takenaka, H. Preparations of Agglomerated Crystals of Polymorphic Mixtures and a New Complex of Indomethacin–Epirizole by the Spherical Crystallization Technique. *J. Pharm. Sci.* **1985**, *74* (11), 1152–1156.
21. Lee, H.-K.; Hajdu, J.; McGoff, P. Propranolol: Methacrylic Acid Copolymer Binding Interaction. *J. Pharm. Sci.* **1991**, *80* (2), 178–180.
22. Takahashi, T.; Takayama, K.; Machida, Y.; Nagai, T. Characteristics of Polyion Complexes of Chitosan with Sodium Alginate and Sodium Polyacrylate. *Int. J. Pharm.* **1990**, *61*, 35–41.
23. Ratajczak, H.; Orville-Thomas, W.S. *Molecular Interactions*; Mir: Moskow, 1984; 11–150.
24. Rachnev, N.; Kalinkova, G.; Toneva-Balucheva, R.; Kretev, V. Some Investigations of Interactions Between Dextrans Promethazine Hydrochloride and Pectins-Promethazine Hydrochloride. *Pharmazie* **1986**, *36* (5), 20–25.
25. Kalinkova, G.N. Infrared Spectroscopy in Pharmacy. *Vibrational Spectroscopy* **1999**, (19), 307–320.
26. Kubinyi, H. Review: Strategies and Recent Technologies in Drug Discovery. *Pharmazie* **1995**, *50* (29), 647–662.
27. Arlington, S.A. How to Survive in Tomorrow's Healthcare Market. *Pharm. Sci. Technol. Today* **2000**, *3* (1), 1–3.
28. Kalinkova, G.N. Review: Studies of Beneficial Interactions Between Active Medicaments and Excipients in Pharmaceutical Formulations. *Int. J. Pharm.* **1999**, *187*, 1–15.
29. Hughes, M.N. *The Inorganic Chemistry of Biological Processes*, 2nd Ed.; Mir: Moscow, 1983; 10–407.
30. *USP 24/NF19 U.S. Pharmacopeia E National Formulary 2000*; United States Pharmacopoeal Convention, Inc.: Rockville, MD, 2000.
31. Dibner, M.D. The Pharmaceutical Industry: Impacts of Biotechnology. *Trends in Pharmacol. Sci.* **1985**, *6* (9), 343–346.
32. Taylor, P.; Mikol, V.; Kallen, J.; Burkhard, P.; Walkinshaw, D.M. Conformational Polymorphism in Peptidic and Nonpeptidic Drug Molecules. *Biopolymers* **1996**, *5*, 585–592.
33. Sponer, J.; Sabat, M.; Burda, J.V.; Leszczynski, J.; Hobba, P.; Lippert, B. Metal Ions in Non-complementary DNA Base Pairs: An Ab Initio Study of Cu(I), Ag(I) and Au(I) Complexes With Cytosine–Adenine Base Pair. *J. Biol. Inorg. Chem.* **1999**, *4* (5), 537–545.
34. Fujii, G. To Fuse or not to Fuse: The Effect of Electrostatic Interactions, Hydrophobic Forces and Structural Amphiphilicity on Protein-Mediated Membrane Destabilization. *Adv. Drug Deliv. Rev.* **1999**, *38*, 257–277.
35. Kabanov, A.E. Taking Polycation Gene Delivery Systems From *in vitro* to *in vivo*. *Pharm. Sci. Technol. Today* **1999**, *2* (9), 365–372.

FURTHER READING

- Blundell, T.L.; Johnson, L.N. *Protein Crystallography*; Mir: Moscow, 1979, 9–262.
- Hughes, M.N. *The Inorganic Chemistry of Biological Processes*, 2nd Ed.; Mir: Moscow, 1983; 10–407.
- Kalinkova, G.N. Review: Studies of Beneficial Interactions Between Active Medicaments and Excipients in Pharmaceutical Formulations. *Int. J. Pharm.* **1999**, *187*, 1–15.
- Kubinyi, H. Review: Strategies and Recent Technologies in Drug Discovery. *Pharmazie* **1995**, *50* (29), 647–662.
- Ratajczak, H.; Orville-Thomas, W.S. *Molecular Interactions*; Mir: Moscow, 1984; 11–150.

COMPOUNDING, CONTEMPORARY

Lloyd V. Allen, Jr.

International Journal of Pharmaceutical Compounding, Edmond, Oklahoma, U.S.A.

INTRODUCTION

Pharmacists are unique professionals—they are well trained in the natural, physical, and medical sciences and sensitized to the potential tragedy that may result from a single mistake that can occur in the daily practice of their profession. The demonstrated expertise, demeanor, and manner in which pharmacists have practiced over the years has resulted in a continued rating of pharmacists as the most respected and trusted individuals in our society. Pharmacists have the reputation of being available (in the local community) by interacting with patients, providing needed medications, and working with patients to regain or maintain a certain standard or quality of health, as well as just being there in time of need.

Pharmacy is a complex mixture of different practices and practice sites. No longer is pharmacy simply community pharmacy or hospital pharmacy. Pharmacy is diverse and offers many opportunities for those willing to look around, find their niche, and practice pharmacy to meet the needs of their own community of patients. Pharmaceutical compounding is an area that is rapidly growing and providing needed products and services to patients and healthcare practitioners. Most compounding pharmacists appear to be very interested and excited about their practices. In fact, many pharmacists intimately involved in pharmaceutical care have now realized the importance of providing individualized patient care through the preparation of patient specific products. Compounding is a professional prerogative that pharmacists have performed since the beginning of the profession. Even today, definitions of pharmacy include the "preparation of drugs" (1, 2).

Pharmacy is the art or practice of preparing and preserving drugs, and of compounding and dispensing medicines according to the prescriptions of physicians (3).

Compounding has always been a basic part of pharmacy practice; the drugs, dosage forms, and equipment/techniques used are the variables. Pharmacists possess knowledge and skills that are unique and not duplicated by any other profession. Pharmacy activities

that individualize patient therapy include compounding and clinical functions; either activity in the absence of the other results in placing pharmacy in a disadvantaged position. It is important to utilize a pharmacist's expertise to adjust dosage quantities, frequencies and even dosage forms for enhanced compliance. All pharmacists should understand the options presented by compounding.

Pharmaceutical compounding is increasing dramatically due to the impact of home health care, nonavailable drug products (especially pediatric formulations), orphan drugs, veterinary compounding, and biotechnology-derived drug products. Newly evolving dosage forms and therapeutic approaches suggest that compounding of pharmaceuticals and related products specifically for individual patients will become even more common in pharmacy practice.

A pharmacy compounder may be defined as a pharmacist who actively promotes and provides prescription-compounding services for the express purpose of attracting this type of prescription to his or her practice. One of the responsibilities of a compounder requires that the pharmacist become actively engaged in the clinical assessment of a patient in order to assist the prescriber in determining the customized patient-specific formula to be extemporaneously compounded. In addition, this responsibility requires the pharmacist to interact with prescribers and the patient when the customized formulation and dosage form are determined. Previous studies have identified the use of clinical skills and physician/patient interaction as intrinsic factors that enhance a pharmacist's job satisfaction (4–7). Therefore, a compounder who utilizes clinical skills and interacts with prescribers and patients is usually predisposed to a higher job satisfaction than would be a noncompounder whose responsibilities may not require such activities (8).

Pharmaceutical compounding requires the use of one's training in mathematics, science, and technology more than many other practices of pharmacy. It has been stated:

The sciences are what support pharmacy's expertise in drug distribution and drug use. Recent history leads one to question whether we in the profession, and some in pharmaceutical education, recognize

and appreciate the contribution that the pharmaceutical sciences have made and continue to make to the pharmacy profession and health care. The pharmaceutical sciences are what make us unique. They provide us the special value that we bring to the bedside. No other health professional is capable of bringing to the pharmacotherapeutic decision-making table such concepts as pH, particle size, partition coefficient, protein binding, structure–activity relationships, economics, and epidemiology. The pharmaceutical sciences, combined with pharmacy's infrastructure including pharmaceutical education, are what make the pharmacist an indispensable participant on the health care team (9).

Furthermore, what area of pharmacy practice has the opportunity of using the scientific education and training as much as pharmacists involved in individualizing patient care through extemporaneous compounding? The pharmaceutical sciences, especially chemistry and pharmaceuticals, serve as the foundation for pharmacists' ability to formulate specific dosage forms to meet patients' needs.

DEFINITION OF COMPOUNDING

The National Association of Boards of Pharmacy has defined compounding as the following:

Compounding means the preparation, mixing, assembling, packaging, or labeling of a drug or device (i) as the result of a practitioner's prescription drug order or—initiative based on the pharmacist/patient/prescriber relationship in the course of professional practice, or (ii) for the purpose of, as an incident to research, teaching, or chemical analysis and not for sale or dispensing. Compounding also includes the preparation of drugs and devices in anticipation of prescription drug orders based on routine, regularly observed patterns (10).

Compounding may hold different meanings to different pharmacists. It may mean the preparation of oral liquids, topicals, and suppositories. It may include the conversion of one dose or dosage form into another, the preparation of select dosage forms from bulk chemicals, the preparation of intravenous (IV) admixtures, parenteral nutrition solutions, pediatric dosage forms from adult dosage forms, the preparation of radioactive isotopes, or the preparation of cassettes,

syringes, and other devices with drugs for administration in the home setting.

TYPES OF COMPOUNDING

Ambulatory Care Compounding

If one is able to walk, the person is considered mobile or ambulatory, i.e., the individual is not bedridden. Consequently, most pharmacists are involved in ambulatory care and most ambulatory patients are outpatients. The term actually can also be applied to home care patients and even institutionalized patients that are mobile. Ambulatory patients are generally responsible for obtaining their own medication, storing it, preparing it (if necessary), and taking it (11). It seems almost incongruous that in healthcare today, as we become more aware that patients are "individuals," respond as "individuals," and must be treated as "individuals," some healthcare providers appear to be grouping patients into "categories" for treatment and "categories" for reimbursement from third parties. Furthermore, they determine "categories" for levels of care in managed care organizations and use "categorized" fixed-dose products provided by pharmaceutical manufacturers that are available because the marketing demand is sufficiently high to justify their manufacture and production. Since when does the availability, or lack of availability, of a specific commercially available product dictate the therapy of a patient?

Pharmacists have an opportunity to extend their activities in patient care as the emphasis continues to shift from inpatient care to ambulatory care. Ambulatory care can generally encourage a team approach to health improvement, prevention, health maintenance, risk assessment, early detection, management, curative therapy, and rehabilitation (12). Ambulatory care offers a variety of opportunities for individualizing patient care through pharmaceutical compounding. In fact, it is the area where most compounding pharmacists practice.

The pharmacist's role in ambulatory care patients can include the following: 1) dispensing; 2) compounding; 3) counseling; 4) minimizing medication errors; 5) compliance enhancement; 6) therapeutic drug monitoring; and 7) minimizing expenditures (11–13). Most reimbursement for ambulatory patients comes from the dispensing or compounding process. Little financial consideration is given to counseling, minimizing medication errors, compliance enhancement, therapeutic monitoring and minimizing expenditures; however, these

activities are important and should be done. Due to the unique nature of compounded medications, counseling is an absolute must for these patients. From the previous discussion of the activities of an ambulatory care pharmacist, it should be evident that extemporaneous compounding can be vitally important in ambulatory patient care.

Institutional Pharmacy Compounding

The ever-present responsibility of the health care industry is to provide the best available care for the patient, using the best means to do so, and to provide that care in a conducive environment. This requires cooperation on the part of the institutional administration, the medical staff, and the employees (nurses and pharmacists in particular as regards to medication usage). It also must involve the patient. One of the effective means by which institutions, and therefore institutional pharmacies, can meet these challenges is to consider expanding extemporaneous compounding services within the institutional pharmacy. Pharmaceutical care and pharmaceutical compounding can provide cost savings to the institution while providing needed options to the physician through problem-solving approaches, and stimulating the institutional pharmacist through new challenges that allow the expression of both their skills and their art.

Institutional pharmacists have always been actively involved in compounding, or producing medications for the patient. Daily IV therapy is provided through compounding of medications. Antibiotic piggybacks, total parenteral nutrition (TPN) solutions, IV additives, and many others are daily calculated, compounded, dispensed, and generally administered by the nursing staff. The preparation of pediatric dosage forms also has been an area of extensive activity in some institutions.

Members of the institutional staff are constantly reading journal articles and are generally aware of innovative thought and practice by their peers. When physicians become aware of the skill, availability, and awareness of pharmaceutical compounding and that they can literally have almost any medication they need, in the form and strength needed for a specific situation, they generally request it more often. As the institutional pharmacy staff demonstrates their expertise and problem-solving skills, the medical staff consistently calls upon them (14).

In the consideration of meeting patient-specific needs, the institutional pharmacist must look at various modalities as potential solutions. When traditional institutional processes and procedures are not meeting

patients' needs, extemporaneous compounding should be a consideration. Improving outcomes and getting patients well and out of the institution as quickly as possible should be the end goal. Individualized dosage forms, dosage strengths, and alternative routes of administration can often help attain these goals. Improving outcomes will assist the medical staff by allowing them to spend their time dealing with new problems as the hospital pharmacy meets the challenge of past problems. Nursing and pharmacy will have an enhanced opportunity to interest and use the skills they have developed, and provide opportunities for pharmacy to have more patient involvement and job satisfaction.

Veterinary Compounding

The first symposium on veterinary compounding, which occurred in September 1993, was a significant forum for discussion by experts and was a pivotal point in the history of veterinary compounding (15). The meeting was important since it assembled an impressive group of experts on veterinary compounding, who then set about explaining and defining the roles of the veterinarian and the pharmacist.

The Food and Drug Administration's (FDA's) interest in compounding by veterinarians dates back to the beginning of the 1990s. The avowed purpose of the symposium was to provide a forum for a comprehensive, public debate in response to the American Veterinary Medical Association (AVMA) position on compounding prior to the issuance of the FDA Compliance Policy Guide on veterinary compounding. Numerous speakers presented views on compounding by veterinarians and compounding for veterinarians by pharmacists. Topics such as conflicts of interest, lack of compounding training by veterinarians, the "new drug" issue, and bioequivalency standards were discussed in detail.

There are many reasons why veterinary compounding is necessary. For example, with multiple species ranging from very small to extremely large, it would be impossible to practice effective medicine without compounded products! Do we simply refuse to treat exotic species or very small animals? Do we abandon oncology in veterinary medicine?

Also, a more specific area of need is the lack of an ideal anesthetic drug, which has led veterinarians to devise anesthetic combinations that induce good-quality anesthesia with minimal risk to the patient. Compounding is essential for safe and effective veterinary anesthetic practice. Veterinarians need to administer anesthetic drugs to a wide variety of patients with a wide variety of temperaments, in settings that are less than ideal.

Veterinarians are called upon to anesthetize elephants, gorillas, tigers, ostriches, sharks, horses, cows, and poisonous snakes, among others. Other reasons why veterinary compounding is necessary includes: 1) the necessity for multiple injections in the absence of a compounded product; 2) rapid changes in management and disease problems in veterinary medicine; 3) problems associated with the treatment of large numbers of animals with several drugs within a short period of time; 4) cost-prohibitive factors associated with the very large volume of some large-volume parenterals required for animals; and 5) the need for previously prepared antidotes for use in cases of animal poisoning.

Unique aspects of veterinary compounding, as compared to compounding for human patients, include the potential impact on human health of compounded veterinary products in food animals, and variability in animal response and size.

The 1993 symposium expressed the following ideas: 1) veterinarians have a definite need for drug compounding; 2) drug compounding was reported to be necessary in all areas of veterinary medicine; 3) the necessity of compounding poisoning antidotes (e.g., sodium nitrite, sodium thiosulfate, methylene blue, and calcium EDTA).

Veterinary compounding will continue to exist in the future for the same reason as it does now—to fill therapeutic needs in veterinary medicine, as well as in medicine for human patients. Difficulties and costs associated with the veterinary drug-approval process will make compounding necessary to fill therapeutic needs not being met by the introduction of therapeutic agents. An increasing interdependence between the veterinarian and the pharmacist is developing, which will result in higher standards of veterinary care (15).

Nuclear Pharmacy Compounding

Nuclear pharmacy is a specialty practice of pharmacy defined as a patient-oriented service that embodies the scientific knowledge and professional judgment required to improve and promote health through assurance of the safe and efficacious use of radioactive drugs for diagnosis and therapy (16). Radioactive drugs, commonly referred to as radiopharmaceuticals, are a special class of drugs regulated by the FDA. They are unique in that they contain an unstable nuclide (radioactive nuclide) as a part of the compound designed to localize in an organ or tissue. Since radiopharmaceuticals are radioactive, the Nuclear Regulatory Commission, or a similar state agency, is involved in regulatory matters relevant to radiopharmaceuticals.

Most radioactive nuclides employed in radiopharmaceuticals have a short half-life. This is beneficial to the patient, as the total number of radioactive atoms given to the patient to produce an image will be small when the half-life of the radioactive nuclide is short, as compared to longer half-life radioactive nuclides. Fewer total atoms reduce the radiation dose to the patient and thus, the risk from a nuclear medicine procedure. However, the short half-life of the radioactive nuclide results in a short shelf-life for the radiopharmaceutical. As a result, most radiopharmaceuticals are compounded on a daily basis. The most common radioactive nuclide used in the preparation of radiopharmaceuticals is technetium-99m (Tc-99m). Tc-99m has a half-life of 6 h and emits only gamma radiation with an energy almost ideal for detection.

A nuclear pharmacist is expert at preparing (compounding) radiopharmaceuticals with Tc-99m sodium pertechnetate and reagent kits. The kits are multidose vials that contain the compound to be “labeled” with the radioactive nuclide Tc-99m in order to create the radiopharmaceutical. The contents within the vial are sterile and pyrogen free, as is the Tc-99m sodium pertechnetate. Most radiopharmaceuticals are administered intravenously so a nuclear pharmacist is expert at maintaining aseptic conditions during compounding.

The most common setting for the provision of radiopharmaceuticals by nuclear pharmacists is a commercially centralized nuclear pharmacy. Radiopharmaceuticals are prepared early in the morning (2:00–3:00 A.M.) and unit doses are delivered by automobile to hospitals near the nuclear pharmacy.

Today, there are several hundred commercial centralized nuclear pharmacies that provide a large percentage of radiopharmaceuticals used in nuclear medicine procedures. What started as limited service in large medical centers and universities by a few pharmacists with education beyond the Bachelor of Pharmacy degree has grown to extensive services provided by several hundred first professional degree pharmacists. This is truly a remarkable change in a time period of 20–25 years, and has resulted from dedicated entrepreneurs who work to make a difference in patient care through quality products and pharmaceutical care.

HISTORY OF COMPOUNDING

The heritage of pharmacy, spanning some 5000 years, has centered around the provision of pharmaceutical products for patients. Pharmacists are the only health professional who possess the knowledge and skill required to compound and prepare medications to meet the unique

needs of patients. The responsibility to extemporaneously compound safe, effective prescription products for patients who require special care is fundamental to the profession.

The 19th century did not see an end to the art of compounding, but the art did give way, however grudgingly, to new technology. According to estimates, a “broad knowledge of compounding” was still essential for 80% of the prescriptions dispensed in the 1920s. Although pharmacists increasingly relied on chemicals purchased from the manufacturer to make up prescriptions, much remained to be done *secundum artem* (1).

The pharmaceutical industry began to take over the production of most medications used by the medical profession in the mid-1990s. In many ways, this has provided superior service, new methods, and a vast array of innovative products that could not have been provided on a one-on-one basis. Research and development (R&D) have been the hallmarks of the pharmaceutical manufacturers. However, the very nature of providing millions of doses of a product requires that the dosage forms (capsule, tablet, suppository) and doses (individual strengths of each dose) be limited and results in a very one-sided approach to therapy. It is simply not economical for a pharmaceutical company to produce a product in 50 different conceivable doses or 15 different dosage forms to meet the needs of the entire range of persons receiving therapy. Windows of activity that meet the majority of patient needs are determined, but the very nature of the process will not be able to meet all patient needs.

We also must recognize that some individuals and their health care needs do not fall in the “windows” or “categories” of theoretical dosage strength and forms. Furthermore, large-scale manufacturers cannot tailor-make a medication for a handful of patients and do so cost effectively and meet the ever-changing needs of a given patient or institution. The pharmacist’s compounding skill fills in this gap in order to meet individualized needs. By this assessment, the pharmacist, through understanding the principals of compounding and recognition of one’s skill level in working *secundum artem*, may recommend a therapy that is not provided by the pharmaceutical industry, but one that is individualized for a specific patient’s needs at a specific time.

Pharmaceutical compounding is increasing due to the impact of home health care, nonavailable drug products, orphan drugs, veterinary compounding, and biotechnology-derived drug products. Newly evolving dosage forms and therapeutic approaches suggest that compounding of pharmaceuticals and related products specifically for individual patients will become more common in

pharmacy practice. Compounding pharmacy is unique as it allows one to use much of one’s scientific, math, and technology background to a fuller extent than many other types of practices. Compounding pharmacists develop a special and unique relationship with their patients. They work hand-in-hand with physicians to solve problems not addressed by commercially available dosage forms.

TECHNICAL AND OTHER CONSIDERATIONS FOR COMPOUNDING

Some considerations related to compounding a prescription are as follows:

1. Is the product commercially available in the exact dosage form, strength, and packaging?
2. Is the prescription rational, i.e., ingredients, use, dose, and method of administration?
3. Are the physical, chemical, and therapeutic properties appropriate?
4. Will the compounded preparation satisfy the intent of the prescribing physician for the needs of the patient?
5. Is there an alternative dosage form, route of administration, etc. by which the patient could benefit?
6. Can the ingredient identity, quality, and purity be assured?
7. Does the pharmacist have the required training and expertise to prepare the prescription?
8. Are the proper equipment, supplies, and chemicals/drugs available?
9. Does documentation on the use, preparation, stability, administration, and storage of the preparation exist?
10. Can the pharmacist perform the necessary calculations to prepare the product?
11. Can the pharmacist project a reasonable and rational “beyond-use” date for the prescription?
12. Can the pharmacist do some basic quality control to check the preparation prior to dispensing (capsule weight variation, pH, visual observations)?

GUIDELINES AND REGULATIONS FOR COMPOUNDING

Two documents are of special importance in providing regulations and guidelines for pharmaceutical compounding. These include the *United States Pharmacopeia/National Formulary* (Chapters $\leq 795 \geq$ “Pharmacy

Compounding Practices” and ≤1206≥ “Sterile Drug Products for Home Use” and other portions of the USP/NF) and the National Association of Boards of Pharmacy *Good Compounding Practices Applicable to State Licensed Pharmacies*.

United States Pharmacopeia

The following are summaries of the lengthy chapters ≤1161≥ and ≤1206≥ in the USP/NF.

Chapter ≤795≥ “Pharmacy Compounding Practices” (17) is divided into: 1) Compounding Environment, 2) Stability of Compounded Preparations, 3) Ingredient Selection and Calculations, 4) Checklist for Acceptable Strength, Quality, and Purity, 5) Compounded Dosage Forms, 6) Compounding Process, 7) Compounding Records and Documents, 8) Quality Control, and 9) Patient Counseling. The compounding environment section describes guidelines for the facilities and equipment that should be available, calibrated, maintained, and used in a compounding pharmacy. The stability section has been previously cited in this chapter, in part to provide guidelines for “beyond-use” dates to be placed on compounded preparations. The selection of ingredients has been previously discussed in this chapter and sample calculations are presented. The checklist for the USP/NF hallmarks of standards of acceptable strength, quality, and purity is presented in a series of questions to be answered. Examples of compounded dosage forms are discussed along with some precautionary statements, as appropriate. A step-by-step presentation on the compounding process is outlined to ensure uniformity of activities in preparing each preparation. Documentation is described for the Formulation Record, the Compounding Record, and the Material Safety Data Sheets (MSDS) files that should be maintained. The section ends with various aspects for patient counseling involving the proper use, storage, and evidence of instability of the compounded preparation.

Chapter ≤1206≥ “Sterile Drug Products for Home Use” covers the areas of: 1) Responsibility of the Dispensing Pharmacist, 2) Risk Levels, 3) Validation, 4) Low-Risk Operations, 5) High-Risk Operations, 6) Environmental Quality and Control, Finished Product Release Checks and Tests, 7) Storage and Expiration Dating, 8) Maintaining Product Quality and Control After it Leaves the Pharmacy, 9) Patient or Caregiver Training, and 10) Patient Monitoring and Complaint System. The compounding pharmacist dispensing any home sterile drug product is responsible for ensuring that the product has been prepared, labeled, controlled, stored, dispensed, and distributed properly. Low-risk and high-risk levels of

sterile products compounding are defined with examples of each. Validation of the sterilization and aseptic processing procedures are described as related to personnel, facilities/equipment, and processes. Low-risk and high-risk operations are described, along with the validations required for each. Environmental quality and control procedures for the work area and personnel, along with suggested standard operating procedures (SOPs) and an example of an environmental monitoring program, are described. Tests and procedures for the finished product are described with the guidelines that only those products that are free from defects and meeting all quality specifications will be distributed. Guidelines are discussed for preparation, storage, and beyond-use dating after the preparation leaves the pharmacy. After the preparation leaves the pharmacy, the caregiver or patient should receive training to ensure understanding and compliance with the storage, handling, and administration of the preparations. The various aspects of the training program are outlined in this chapter. Also included are the recommendation for written policies and procedures for monitoring of patients using home-use sterile drug products and the handling and reporting of adverse events. It is evident in this chapter that the responsibility of the compounding pharmacist ranges from the activities involved in the compounding of the product through its proper storage, distribution, use, and disposal.

National Association of Boards of Pharmacy

The following is a summary of the Good Compounding Practices (GCPs) of the National Association of Boards of Pharmacy (NABP).

Subpart A contains general provisions and definitions, and sets forth the minimum current good compounding practices for the preparation of drug products by state-licensed pharmacies for dispensing and/or administration to humans or animals and includes definitions and requirements, as follows:

Based on the existence of a pharmacist patient prescriber relationship and the presentation of a valid prescription, pharmacists may compound, in reasonable quantities, drug products that are commercially available in the marketplace. Pharmacists shall receive, store, or use drug substances for compounding that have been made in a FDA-approved facility and/or drug components that meet official compendial requirements. If neither of these requirements can be met, pharmacists shall use their professional judgment to procure alternatives.

Pharmacists may compound drugs in very limited quantities prior to receiving a valid prescription based on a history of receiving valid prescriptions. Pharmacists shall not offer compounded drug products to other state-licensed persons or commercial entities for subsequent resale, except in the course of professional practice for a prescriber to administer to an individual patient.

Subpart B discusses the organization and personnel and the responsibilities and authority of the compounding pharmacist and other individuals involved in the compounding process.

In Subpart C, drug compounding facilities are discussed further in this section. It also discusses bulk drugs and materials as well as the compounding of sterile products, radiopharmaceuticals, and special precaution products, such as handling penicillins.

Subpart D discusses equipment and states that equipment used in the compounding of drug products shall be of appropriate design, adequate size, and suitably located to facilitate operations for its intended use and for its cleaning and maintenance.

In Subpart E control of components and drug product containers and closures are discussed. Components, drug product containers, and closures used in the compounding of drugs shall be handled and stored in a manner to prevent contamination. Their required characteristics are mentioned as well.

Subpart F, Drug Compounding Controls, explains the written procedures for the compounding of drug products in order to assure that the finished products have the identity, strength, quality, and purity they purport or are represented to possess. It also discusses some of the final quality control procedures that can be done, including, but not limited to, the following (as appropriate): 1) capsule weight variation; 2) adequacy of mixing to assure uniformity and homogeneity; and 3) clarity, completeness, or pH of solutions.

Appropriate written procedures designed to prevent microbiological contamination of compounded drug products purporting to be sterile shall be established and followed. Such procedures shall include validation of any sterilization process.

Subpart G explains labeling control of excess products. In the case where a quantity of a compounded drug product in excess of that to be initially dispensed in accordance with Subpart A is prepared, the excess product shall be labeled or documentation referenced with the complete list of ingredients (components), the preparation date, and the assigned expiration date based upon professional judgment, appropriate testing, or published data.

Subpart H discusses records and reports. Any procedures or other records required to be maintained in compliance with these good compounding practices shall be retained for the same period of time as each State requires for the retention of prescription files.

FACILITIES, EQUIPMENT, AND SUPPLIES

Compounding Facility

A separate area for traditional compounding is recommended, rather than simply cleaning off a small area of the dispensing counter. The compounding pharmacist needs a clean, neat, well-lit, and quiet working area. If aseptic compounding is considered, a "clean room" with a laminar airflow hood should be used, dependent also upon the individual state board of pharmacy requirements. The actual facility to be used will depend on the level of compounding to be done.

Adequate lighting and ventilation shall be provided in all drug-compounding areas. Potable water shall be supplied under continuous positive pressure in a plumbing system free of defects that could contribute contamination to any compounded drug product. Adequate washing facilities, easily accessible to the compounding area(s) of the pharmacy, shall be provided. These facilities shall include, but not be limited to, hot and cold water, soap or detergent, and air-driers or single-use towels.

The area(s) used for the compounding of drugs shall be maintained in a clean and sanitary condition. Trash shall be held and disposed of in a timely and sanitary manner.

Compounding Equipment

Much of the equipment used today in compounding has changed. Today, electronic balances are more often used than torsion balances. Micropipets are commonplace and ultrafreezers are sometimes required in addition to standard refrigerator-freezers. This area is constantly changing and the compounding pharmacist should be aware of the available technology necessary to prepare accurate and effective prescriptions. It is very helpful to become acquainted with the local representative for a laboratory supply company.

The equipment needed will be determined by the type and extent of the services one chooses to provide. Hospitals already utilize laminar flow hoods where aseptic compounding of sterile solutions is performed. These

same hoods can be used to compound other sterile products, such as eye drops. A balance, preferably electronic, is essential. Ointment slabs (pill tiles), along with spatulas of different types and materials, should be purchased. A few mortars and pestles (glass, ceramic, and/or plastic) should be obtained, as well as some glassware. It may not be necessary to buy a roomful of equipment, but one should purchase what is needed to start the service and build it as the service grows and expands to different arenas.

Equipment used in the compounding of drug products shall be of suitable composition so that surfaces that contact components, in-process materials, or drug products shall not be reactive, additive, or absorptive so as to alter the safety, identity, strength, quality, or purity of the drug product beyond that desired.

Equipment and utensils used for compounding shall be cleaned and sanitized immediately prior to use to prevent contamination that would alter the safety, identity, strength, quality, or purity of the drug product beyond that desired. In the case of equipment, utensils, and containers/closures used in the compounding of sterile drug products, cleaning, sterilization, and maintenance procedures as set forth in the NABP Model Rules for Sterile Pharmaceuticals must be followed.

Previously cleaned equipment and utensils used for compounding drugs must be protected from contamination prior to use. Immediately prior to the initiation of compounding operations, they must be inspected by the pharmacist and determined to be suitable for use.

Automatic, mechanical, or electronic equipment, or other types of equipment or related systems that will perform a function satisfactorily, may be used in the compounding of drug products. If such equipment is used, it shall be routinely inspected, calibrated (if necessary), or checked to assure proper performance.

Compounding Supplies

Throughout history, pharmacists have been using chemicals and other materials for prescription compounding. In the past, these chemicals and materials have been obtained from natural products, raw materials, and household ingredients. Today, compounding pharmacists use chemicals from various sources, depending upon their availability.

The <795> “Pharmacy Compounding Practices” monograph in the USP 24/NF 19 states as follows (17):

A USP or NF grade drug substance is the preferred source of ingredients for compounding all drug preparations. If that is not available, the use of another high-quality source, such as analytical reagent (AR) or certified American Chemical Society (ACS) grade, is an option for professional judgment. If the substance is not an official preparation or substance, additional information, such as a certificate of analysis, needs to be obtained by the pharmacist to ensure its suitability.

A manufactured drug product may be a source of an active ingredient. Only manufactured drugs from containers labeled with a batch control number and a future expiration date are acceptable as a potential source of active ingredients. When compounding with manufactured drug products, the pharmacist must consider all ingredients present in the drug product relative to the intended use of the compounded preparation.

In summary, it is the responsibility of the pharmacist to select the “most appropriate” quality of chemical for compounding, beginning with the USP/NF as the first choice and, if this is not available, then descend

Table 1 Description of chemical grades

Grade	Description
Technical or commercial	Indeterminate quality
Chemically pure (CP)	More refined but still of unknown quality
USP/NF	Meets minimum purity standards; conforms to tolerances set by the <i>USP 23/NF 18</i> for contaminants dangerous to health
ACS reagent	High purity; conforms to minimum specifications set by the Reagent Chemicals Committee of the American Chemical Society
AR	Very high purity
HPLC	Solvents purified for use in HPLC; very high purity
Spectroscopic grade	Very high purity
Primary standard	Highest purity; required for accurate volumetric analysis (for standard solutions)

the list of purity grades (Table 1), using professional judgment and discretion. A certificate of analysis for the chemicals should be obtained and kept on file in the pharmacy for these selected chemicals (17, 18).

STABILITY OF COMPOUNDED PREPARATIONS

Stability is the extent to which a product retains, within specified limits and throughout its period of storage and use, the same properties and characteristics that it possessed at the time of its preparation. Chemical stability is important when selecting storage conditions (temperature, light, humidity), selecting the proper container for dispensing (glass versus plastic, clear versus amber or opaque, cap liners), and for anticipating interactions when mixing drugs and dosage forms.

Factors Affecting Stability

An important component of compounding is the consideration of factors that affect the stability of the final preparation. These factors include pH, temperature, solvent, light, air (oxygen, carbon dioxide, moisture), humidity, particle size, ionic strength, dielectric constant, polymorphism, crystallization, vaporization, and adsorption.

Types and Examples of Stability

Five types of stability are defined by the USP/NF, three of which will be described. Physical stability is concerned with the original physical properties of the preparation, and include appearance, palatability, uniformity, dissolution and suspendability. Chemical stability states that each active ingredient retains its chemical integrity and labeled potency within specified limits.

Microbiological stability states that sterility or resistance to microbial growth is retained according to the specified requirements. Antimicrobial agents that are present will retain their effectiveness within specified limits.

Beyond-Use Dating

Whereas commercially manufactured products are required to possess an "expiration date," compounded products are assigned a "beyond-use" date. Numerous sources of information can be utilized in determining an appropriate "beyond-use" date, such as chemical companies, manufacturers literature, laboratory data, journals,

and published books on the subject. Generally, most pharmacists prepare/dispense small quantities of compounded products, recommend storage at room, cool, or cold temperatures, and use a conservative "beyond-use" date.

The guidelines published in the USP 24/NF 19 Chapter <795> entitled "Pharmacy Compounding Practices" state that:

In the absence of stability information that is applicable to a specific drug and preparation, the following maximum beyond-use dates are recommended for nonsterile compounded drug preparations that are packaged in tight, light-resistant containers and stored at controlled room temperature unless otherwise indicated.

For nonaqueous liquids and solid formulations (where the manufactured drug product is the source of active ingredient), the beyond-use date is not later than 25% of the time remaining until the product's expiration date or 6 months, whichever is earlier.

A USP or NF substance is the source of active ingredient-the beyond-use date is not later than 6 months. For water-containing formulations (prepared from ingredients in solid form)-the beyond-use date is not later than 14 days when stored at cold temperatures.

For all other formulations-the beyond-use date is not later than the intended duration of therapy or 30 days, whichever is earlier. These beyond-use date limits may be exceeded when there is supporting valid scientific stability information that is directly applicable to the specific preparation (i.e., the *same* drug concentration range, pH, excipients, vehicle, water content, etc.) (17).

COMPOUNDING QUALITY CONTROL

Physical Tests

Pharmacists can perform a number of physical quality control tests to ensure the uniformity and accuracy of many small-scale compounded preparations. These quality control tests include individual dosage unit weights, average individual dosage unit weights, total product weight, pH, and physical observations such as appearance, taste, and smell.

Physical Observations

Physical observations can include color, clarity, uniform distribution, hardening, brittleness, softening, discoloration, expansion/distortion, caking, odor, precipitation, discoloration, haziness, gas formation, clarity, breaking, creaming, difficulty in resuspending, consistency, grittiness, dryness, shrinkage, water evaporation, shriveling, and the presence of oil stains on packaging.

Sterile Products Testing

If appropriate, due to the number of sterile preparations compounded, products may be tested for sterility and the effectiveness of incorporated preservatives, if present. This may be done either at the pharmacy or in a contract laboratory.

Pyrogen Testing

Pyrogen testing should be done especially if sterile preparations are made from nonsterile bulk materials. This testing can be done either at the pharmacy or in a contract laboratory.

Contract Analytical Laboratories

Contract analytical testing can be utilized for purity, potency, sterility, and pyrogenicity testing. The frequency of testing is generally related to the volume and frequency of product preparation.

FLAVORS, SWEETENERS, AND COLORS

Preparation of an Aesthetic Product

Flavoring, sweetening, and coloring are important to enhance patient compliance when medications are administered orally. Oftentimes, these can be adjusted to meet the preferences of the patient.

Basics and Examples of Flavoring

Pharmacists must be familiar with the four primary tastes (sweet, sour, salty, and bitter) and be aware of the correlations between select chemical properties and taste and odor. Using this information, they can use a number of approaches to prepare an acceptable product to minimize the bad taste of drugs. These approaches include blending,

overshadowing, physical methods (insoluble compounds, emulsification, effervescence, viscosity), chemical methods (adsorption, complexation), and physiological techniques (cooling due to negative heat of solution anesthetic action of some ingredients). Also, flavor intensifiers can be used, including citrus enhancers, such as citric, maleic, or tartaric acids.

Basics and Examples of Sweetening

A number of different types of sweeteners, caloric and noncaloric, are available, depending upon the specific prescription and patient. Some sweeteners have aftertastes that must also be considered. Today, patients also can have their choice of natural or synthetic sweeteners.

Basics and Examples of Coloring

Coloring is not always necessary but may be of value in certain medications. For example, oral liquids are generally colored with a dye that matches the flavor of the medication. Pharmacists must be aware of the different oil, alcohol, and water solubilities of the dyes they use.

PRESERVATION, STERILIZATION, AND DEPYROGENATION

Methods of Preservation

Some compounded preparations are naturally preserved, as in the case of some syrups and elixirs. Others require the addition of a preservative. Preservatives are commonly added to products to minimize microbial growth, as in the case of oral liquids, topicals, and multi-dose parenterals.

Physicochemical Considerations in Preservation

A preservative is selected based upon its characteristics, including concentration, pH, taste, odor, and solubility.

Preservative Effectiveness Testing

In some situations, it is advisable to have a preservative effectiveness test conducted on a preparation that may

require routine compounding or may require an extended storage period.

Methods of Sterilization

Compounding pharmacists routinely use in-process sterilization, such as sterile filtration, or terminal sterilization, such as autoclaving or dry-heat sterilization. In some situations, combinations of these methods may be used along with chemical sterilization.

Methods of Depyrogenation

Depyrogenation methods used by compounding pharmacists include dry heat and rinsing with Sterile Water for Injection, USP.

COMPOUNDING PHARMACEUTICAL SOLIDS

Powders and Granules

Powders are thorough mixtures of dry, finely divided drugs and excipients that are intended for internal or external use. Granules are dosage forms that consist of particles ranging in size from about 4–10 mesh. Both powders and granules are easy to use and are easy to compound. An example formula of a currently compounded powder includes the following:

Rx Misoprostol	600 g
Lidocaine hydrochloride	500 mg
Polyox WSR-301	2.45 g
Methocel E4M Premium	22.05 g

Capsules

Capsules are dosage forms in which unit doses of powder, semisolid, or liquid drugs are enclosed within either a hard or a soft envelope or shell. Examples of currently compounded capsules include the following with the contents either as powders, in oil or in a semisolid-fill capsule, as well as modified strength capsules:

Rx Dextromethorphan Hydrobromide 30 mg and Morphine Sulfate 10 mg Capsules (#100)	
Dextromethorphan hydrobromide	3 g
Morphine sulfate	1 g
Lactose	35.5 g
Capsule size #1	#100

Rx Triple Estrogen 2.5 mg Slow-Release Capsules (#100)	
Estriol	200 mg
Estrone	25 mg
Estradiol	25 mg
Methocel E4M Premium	10 g
Lactose	23.75
Capsule size #1	#100

Rx Triple Estrogen 2.5 mg in Oil Capsules (#100)	
Estriol	200 mg
Estrone	25 mg
Estradiol	25 mg
Peanut oil	20 ml (18.38 g)
Capsules size #1	#100

Rx Triple Estrogen 2.5 mg Semisolid-Filled, Hard-Gelatin Capsules (#100)	
Estriol	200 mg
Estrone	25 mg
Estradiol	25 mg
Polyethylene glycol 1450	20 g
Polyethylene glycol 3350	20 g
Capsules size #1	#100

Tablets

Tablets are solid dosage forms that are generally either compressed or prepared by a sintering process, which includes sublingual, buccal, chewable, effervescent, and compressed tablets. Some of these can be easily compounded. An example of a tablet triturate is as follows:

Rx Sodium Fluoride 2.2 mg Tablet Triturates (#100)	
Sodium fluoride	220 mg
Sucrose, powdered	1.15 g
Lactose, hydrous	4.63 g

Lozenges/Troches

Lozenges/troches are solid preparations designed to dissolve or disintegrate slowly in the mouth. Their base is usually flavored and sweetened. Examples of compounded troches include anesthetic (lidocaine), hormonal (testosterone), analgesic (ketamine), and anti-fungal (nystatin) preparations.

Rx Testosterone 10 mg Troches (#24)	
Testosterone	240 mg
Aspartame	500 mg
Silica gel	480 mg
Acacia	360 mg
Flavor	qs
Polyethylene glycol 1450	23 g

Rx Pediatric Chewable Gummy Gels

Active drug	qs
Bentonite	500 mg
Aspartame	550 mg
Acacia powder	500 mg
Citric acid monohydrate	700 mg
Flavor	qs
Gummy gel base	26.6 g

Rx Gummy Gel Base

Gelatin	43.4 g
Glycerin	155 ml
Purified water	21.6 ml

Rx Nystatin Popsicles (#10)

Nystatin powder	2,500,000 units
Sorbitol 70% solution	20 ml
Simple syrup	50 ml
Flavor	qs
Purified water qs	300 ml

Rx Tetracaine 20 mg Lollipops (#30)

Tetracaine HCl	600 mg
Lemon essence	0.5 ml
Green food color	2 ml
Sucrose	150 g
Potassium bitartrate	500 mg
Purified water	55 ml

Suppositories

Suppositories are solid dosage forms that are used to administer medicine through the rectum, vagina, or urethra. They are of different sizes or shapes, depending upon the body orifice for their administration. Examples of compounded suppositories include the antinauseant combinations (lorazepam, diphenhydramine, haloperidol, and metoclopramide), analgesic (morphine), antifungal, and hemorrhoidal (lidocaine, tannic acid) preparations.

Rx Fluconazole 200 mg Vaginal Suppositories

Fluconazole	200 mg
Polyethylene glycol base	qs

Rx Morphine Sulfate Slow-Release Suppositories (#1)

Morphine sulfate	50 mg
Alginic acid	25%
Witepsol H-15	qs

Sticks

Sticks are a convenient form of administering topical medications and come in different sizes and shapes. They are readily transportable and can be easily compounded.

Rx Gralla Type Antiemetic Suppositories

<i>Ingredient</i>	#1	#2	#3	#4	#5	#6	#7
Metoclopramide HCl	10	—	20	10	40	20	20
Haloperidol	0.5	5	—	—	1	1	—
Diphenhydramine HCl	—	25	25	—	25	25	25
Dexamethasone	—	—	10	—	10	10	5
Lorazepam	0.5	2	—	—	1	—	—
Diazepam	—	—	—	5	—	—	—
Hydroxyzine HCl	25	—	—	—	—	—	—
Promethazine HCl	—	—	—	25	—	—	—
Benzotropine mesylate	—	—	—	—	1	—	—
Silicon dioxide	20	30	20	15	—	—	—
Fatty acid base qs	2g	—	2g	2g	2g	2g	2g
Polyethylene glycol base qs	—	2g	—	—	—	—	—

Examples of compounded sticks include those containing antivirals for herpes and emollients/sunscreens for environmental exposure.

Rx Fluorouracil 5% Topical Stick (25-g tubes)

Fluorouracil	5 g
Polyethylene glycol 3350	27 g
Polyethylene glycol 300	68 g

Rx Acyclovir Stick with Sunscreen (five 5-g tubes)

Acyclovir 200 mg capsules	#5
<i>para</i> -Aminobenzoic acid	150 mg
Silica gel, micronized	120 mg
Polyethylene glycol 3350	6.5 g
Polyethylene glycol 300	15 ml

COMPOUNDING PHARMACEUTICAL LIQUIDS

Solutions

Solutions are liquid preparations that contain one or more drug substances molecularly dispersed in a suitable solvent or a mixture of mutually miscible solvents. Solutions include those for use topically as well as internally.

Rx Dexamethasone and Lidocaine Solution for Iontophoresis (100 ml)

Dexamethasone sodium phosphate	200 mg
Lidocaine hydrochloride	1 g
Sterile water for injection	100 ml

Rx Buprenorphine Hydrochloride 150 mg/100 ml Nasal Spray (100 ml)

Buprenorphine hydrochloride	150 mg
Glycerin	5 ml
Methylparaben	200 mg
0.9% Sodium chloride injection	95 ml

Examples of compounded liquids include topicals (wart solutions), oral syrups and elixirs, nasal solutions, otic solutions, iontophoretic solutions (dexamethasone sodium phosphate), and many others.

Suspensions

Suspensions are two-phase systems that consist of a finely divided solid dispersed in a liquid, solid, or gas. They are appropriate when the drug to be incorporated is not sufficiently soluble in an ordinary solvent or cosolvent system. They are used orally and topically. Examples of compounded suspensions include many pediatric oral liquids where a commercial pediatric dosage form is not available. Commercial tablets and capsules are formulated into a vehicle for the patient and can be individually flavored to the patient's preferences.

Rx Indomethacin 4% Topical Spray (100 ml)

Indomethacin	4 g
Hydroxypropyl cellulose	200 mg
Sodium lauryl sulfate	100 mg
Purified water	10 ml
Alcohol, 95% qs	100 ml

Rx Testosterone 10 mg/0.1 ml Sublingual Drops (10 ml)

Testosterone	1 g
Saccharin	100 mg
Silica gel	200 gm
Flavor	qs
Almond oil qs	10 ml

Emulsions

Emulsions are heterogeneous systems consisting of at least one immiscible liquid that is intimately dispersed in another liquid in the form of droplets or globules, whose diameters generally exceed 0.1 micron. They are also thermodynamically unstable mixtures of two essentially immiscible liquids and an emulsifying agent that helps hold them together. Examples of compounded emulsions include those for both topical and oral used. Topical emulsions include creams and even liposomal

Rx Emulsion Base

Mineral oil, heavy	25 ml
Isopropyl myristate	25 ml
Polysorbate 80	7 ml
Methylparaben	200 mg
Propylparaben	100 mg
Purified water qs	100 ml

preparations. Compounded emulsions include preparations for both oral and topical use. This category also includes the pluronic-lecithin-organogels, which are penetration-enhancing gels.

Rx Ketamine 10% in Pluronic Lecithin Organogel

Ketamine hydrochloride	10 g
Isopropyl palmitate:Soy lecithin 1:1	20 g
Pluronic F127 20% gel qs	100 ml

COMPOUNDING PHARMACEUTICAL SEMISOLIDS

Ointments and Pastes

Ointments are semisolid preparations that are intended to be applied externally to the skin or mucous membranes. They soften or melt at room temperature. Pastes are thick, stiff ointments that ordinarily do not flow at body temperature and thus protect and coat the areas to which they are applied. Examples of compounded ointments and pastes include the following:

Note: The testosterone-menthol eutectic mixture can be prepared by mixing 31.6 g of testosterone with 68.4 g menthol, using sufficient methyl alcohol to dissolve them both and allowing the alcohol to evaporate to dryness.

Rx Testosterone–Menthol Eutectic Ointment (2% Testosterone)

Testosterone–menthol eutectic mixture	6.33 g
Hydrophilic petrolatum	93.67 g

Rx Aluminum Acetate Paste (100 g)

Aluminum acetate solution	17 ml
Anhydrous lanolin	34 g
Lassar's Plain Zinc Paste	49 g

Creams

Creams are opaque, soft solids or thick liquids intended for external application. Creams may contain medications dissolved or suspended in water-soluble or vanishing cream bases and can be either water-in-oil or oil-in-water. Examples of creams are as follows:

Rx Progesterone 5% Cream

Micronized progesterone	5 g
Glycerin	qs
Hydrophilic ointment	95 g

Note: Commercial oil-in-water vehicles can be used and the quantity of progesterone is variable.

Gels

Gels are semisolid systems that consist of suspensions made up of either small inorganic particles or large organic molecules interpenetrated by a liquid. Some gels are clear and others are turbid since their ingredients may or not be completely molecularly dispersed or they may form aggregates, which disperse light. Examples of compounded gels include many topical, oral cavity, and even rectally administered preparations, such as the following:

Rx Piroxicam Topical Gel	
Hydroxypropylcellulose	1.75 g
70% isopropyl alcohol	98.25 ml
Propylene glycol	4.1 ml
Polysorbate 80	1.7 ml
Piroxicam 20-mg capsules	25 capsules
(Piroxicam powder can be used, if available.)	

Rx Ketoprofen, Topical Gel Cyclobenzaprine and Lidocaine	
Ketoprofen	10 g
Cyclobenzaprine hydrochloride	1 g
Lidocaine	5 g
Propylene glycol	10 ml
Sorbic acid	200 mg
Lecithin:isopropyl palmitate solution	20 g

COMPOUNDING STERILE PREPARATIONS

Ophthalmics

Ophthalmic preparations are sterile, free from foreign particles, and prepared especially for instillation into the eye. They include solutions, suspensions, and ointments. Examples of compounded ophthalmics include drugs such as acetylcysteine, acyclovir, alteplase, amikacin, amphotericin B, ascorbic acid, bacitracin, calcium gluconate, penicillins, cephalosporins, aminoglycosides, cromolyn, cyclosporin, deferoxamine, anti-inflammatory corticosteroids, chelating agent, oncology agents, and numerous others prepared as both solutions and ointments.

Rx Acetylcysteine 15% Ophthalmic Solution	
Acetylcysteine	15 g
Sterile water for injection qs	100 ml
Sodium hydroxide to pH of 6 to 7.5	

Inhalation Solutions

Inhalation solutions are designed to deliver a drug into the respiratory tree of a patient for either a local or systemic effect. Examples of compounded inhalation solutions include individual and combinations of albuterol, cromolyn, morphine sulfate, corticosteroids, ipratropium, metaproterenol, terbutaline, and others.

Rx Morphine Sulfate 0.25% Inhalation Solution	
Morphine sulfate	250 mg
Citric acid, hydrous	100 mg
Sterile water for injection, qs	100 ml

Rx Albuterol Sulfate and Ipratropium Bromide Inhalation Solution	
Albuterol sulfate	100 mg
Ipratropium bromide	16.7 mg
Benzalkonium chloride 50% solution	0.003 ml
0.9% Sodium Chloride Injection qs	100 ml

Parenterals

Parenterals are products that are administered to the body by injection. They must be sterile, nonpyrogenic, and particulate-free. Examples of compounded parenterals include high-dose analgesics for patient-controlled analgesia (morphine sulfate 50 mg/ml), antiemetic injections, fentanyl and bupivacaine injections for ambulatory pump reservoirs, oncology combinations, and others.

Rx Reglan, Ativan and Compazine Injection		
Reglan (5 mg/ml)	30 ml	150 mg
Ativan (2 mg/ml)	0.5 ml	1 mg
Mannitol 25%	50 ml	12.5 mg
Compazine (5 mg/ml)	2 ml	10 mg
5% Dextrose injection		50 ml

Rx Fentanyl Citrate 1.25 g/ml, Bupivacaine Hydrochloride 0.4375 mg/ml, Epinephrine Hydrochloride 0.69 g/ml		
Fentanyl citrate 50 µg/ml		2.5 ml
Bupivacaine hydrochloride, 0.5%		8.75 ml
Epinephrine hydrochloride, 1:100,000		6.9 ml
0.9% sodium chloride injection		81.85 ml

Rx TPN Solution	
50% Dextrose injection	500 ml
Amino acids 8.5% with Lytes	500 ml
Lipids 10%	200 ml
Calcium gluconate	1 g
Magnesium sulfate	ss2 g
Trace elements	1 unit
M.V.I.-12	1 unit

COMPOUNDING WITH BIOTECHNOLOGY PREPARATIONS

Definitions

Biotechnology preparations are those that are developed using the techniques of engineering and technology with living organisms. Biotechnology presents compounding pharmacists with a unique and new source of therapeutic agents that may require their special expertise.

Types of Preparations

Biotechnology products differ in their method of preparation and potential problems presented in their formulation. Pharmacists involved in compounding with biologically active proteins will be interested in their stabilization, formulation, and delivery. Most of the biotechnology products currently are proteins but some may soon be smaller peptide-like molecules.

Physicochemical Considerations

In working with biotechnology-derived drugs, the compounding pharmacist must be cognizant of both the active drug constituent and the total drug delivery system or carrier. Proteins are generally very potent and used in very low concentrations. Pharmacists must be aware of the vehicle, buffer, and stabilizer requirements for these preparations, including the use of surfactants, amino acids, polyhydric alcohols, fatty acids, proteins, antioxidants, reducing agents, and metal ions. Proper pH, chelating agents, preservatives and tonicity-adjusting agents must be considered.

Quality Control and Stability

One must be aware of the factors involved in handling proteins in order to retain a drug's biologic activity up to the time when it is administered to the patient. Proteins are inherently unstable molecules and their degradation profiles can be very complex. Compounding pharmacists may be involved in the selection of an appropriate vehicle for drug delivery, individualizing dosages, administering drugs through novel drug delivery systems, preparing drugs for delivery through these systems, monitoring their efficacy, and counseling patients on their use.

An example of a compounded prescription using a biotechnology-derived product is as follows:

Rx Tissue Plasminogen Activator 25 mg/100 L	
Ophthalmic Solution	
Tissue plasminogen activator	20 mg vial
0.9% Sodium chloride injection	60 ml
Sterile water for injection	20 ml

COMPOUNDING VETERINARY PREPARATIONS

Guidelines for Veterinary Compounding

Veterinary compounding can be considered when there are no effective FDA-approved products available, when available dosage forms are inappropriate, when multiple and concurrent disease states are present, when an additive therapeutic effect could be obtained from simultaneous administration of two or more products or to minimize side effects, when economic realities would preclude treatment with the approved product, and when compounding would encourage compliance of dosage/therapeutic regimens.

Considerations in Veterinary Compounding

Most drugs that are FDA-approved are specified for certain species, either food producing or for a large target population. If veterinarian pharmaceutical companies do not perceive a sufficiently large market for a product, they will not seek approval for products. This has left a large vacuum, or a potential market, to be filled by compounding pharmacists. There are no FDA-approved products for exotic species due to the limited market and there are only limited FDA-approved products for some of the more common species. Veterinarians need patient-specific products and pharmacists know how to prepare these products. Consequently, a team approach has developed to the benefit of the veterinarian, pharmacist, and animal patient.

Questions that often arise include the following. What is the overall goal of the treatment of this animal? Are there any commercially available products that can be used? What are the regulatory concerns? Is this a food or a milk-producing animal? Will there be a residue problem? What do we know about the physical and chemical compatibility of these drugs? What do we know about the stability of these drugs before, during, and after the compounding process? What do we know about the pharmacokinetics of the active ingredients? Are personnel going to be at risk from handling the drug during compounding or while using the compounding form?

Examples of some veterinary preparations are as follows:

Rx Acetylcysteine, Gentamicin and Atropine Solution

Acetylcysteine	720 mg
Gentamicin (as gentamicin sulfate)	36 mg
Atropine sulfate	36 mg
Sterile water qs	15 ml

Rx 4-Methyl-pyrazole Solution

4-Methyl-pyrazole	1 g
Polyethylene glycol 400	9 ml
Bacteriostatic water for injection qs	20 ml

SUMMARY

Pharmacy compounding provides pharmacists with a unique opportunity to practice their time-honored profession. It will become an even more important part of pharmacy practice in the future, and include pharmacists involved in community, hospital, nursing home, home health care, veterinary, and specialty practices. Pharmaceutical compounding is a practice where the clinical expertise of pharmacists can be merged with the scientific expertise of pharmacists to make pharmaceutical care a reality.

REFERENCES

1. Allen, Loyd V., Ed. *The Compact Oxford English Dictionary*, 2nd Ed.; Oxford University Press, Inc.: New York, 1991.
2. Allen, Loyd V., Ed. *American Heritage Dictionary of the English Language*, 3rd Ed.; 1992; In Electronic Form in Microsoft Bookshelf'95, Microsoft Corporation, 1995.
3. Allen, Loyd V., Ed. *Webster's Revised Unabridged Dictionary of the English Language*, G & C Merriam Company: Springfield, MA, 1913; 1075.
4. Williamson, R.E.; Kabat, H.F. Job Satisfaction among Minnesota Hospital Pharmacists. *Am. J. Hosp. Pharm.* **1972**, 29, 942–946.
5. Humphrys, P.; O'Brien, G.E. The Relationship between Skill Utilization, Professional Orientation and Job Satisfaction for Pharmacists. *J. Occupational Behav.* **1986**, 59, 315–326.
6. Johnson, C.A.; Hammel, R.J.; Heinen, J.S. Levels of Satisfaction among Hospital Pharmacists. *Am. J. Hosp. Pharm.* **1977**, 34, 341–247.
7. Olson, D.S.; Lawson, K.A. Relationship between Hospital Pharmacists' Job Satisfaction and Involvement in Clinical Activities. *Am. J. Health-Syst. Pharm.* **1996**, 53, 281–284.
8. Letendre, W.R.; Shepherd, M.D.; Brown, C.M. Comparison of Job Satisfaction for Compounders and Concompounders. *Int. J. Pharmaceutical Compounding* **1998**, 2 (6), 455–462.
9. Penna, R. What Makes Pharmacists Different?. *Am. J. Pharmaceutical Ed.* **1997**, 61 (Spring), 103.
10. *Good Compounding Practices Applicable to State Licensed Pharmacies* NABP: Park Ridge, IL, .
11. Popovich, N.G. Ambulatory Patient Care. *Remington: The Science and Practice of Pharmacy*, 19th Ed.; Gennaro, A.R., Ed.; Mack Publishing Co.: Easton, PA, 1995; 1695–1719.
12. Raehl, C.L.; Bond, C.A.; Pitterle, M.E. Ambulatory Pharmacy Services Affiliated with Acute Care Hospitals. *Pharmacotherapy* **1993**, 13 (6), 618–625.
13. Goode, M.A.; Gums, J.G. Therapeutic Drug Monitoring in Ambulatory Care. *Ann. Pharmacother.* **1993**, 27, 502–505.
14. Sundberg, J.A. Extemporaneous Compounding in the Hospital Pharmacy. *Int. J. Pharmaceutical Compounding* **1997**, 1 (5), 314–317.
15. Proc. J. Amer. Vet. Med. Assoc. **1994**, 204 (2), 189–305.
16. Shaw, S.M. Introduction to Nuclear Pharmacy. *Int. J. Pharmaceutical Compounding* **1998**, 2 (6), 424–425.
17. *United States Pharmacopeia XXIV/National Formulary 19*; US Pharmacopeial Convention, Inc.: Rockville, MD, 2000; 2128–2143, 2698–2704.
18. Food and Drug Modernization Act of 1997.

COPRECIPITATES AND MELTS

Madhu K. Vadnere

NJ Pharma LLC, Chatham, New Jersey, U.S.A.

INTRODUCTION

The bioavailability of a poorly water-soluble drug is often limited by its dissolution rate, which in turn is controlled by the surface area available for dissolution. The effect of the particle size of a drug on its dissolution rate and its biological activity is well known. For example, Atkinson et al. (1) reported that micronization of griseofulvin resulted in reduction of the therapeutic dose by half.

The conventional methods for reducing particle size and increasing surface area include trituration, grinding, ball milling, fluid energy micronization, and controlled precipitation (2). Coprecipitates and melts are solid dispersions that provide a means of reducing particle size to the molecular level. Sekiguchi and Obi (3) first introduced the concept of using solid dispersions to improve bioavailability of poorly water-soluble drugs in 1961. They demonstrated that the eutectic of sulfathiazole and the physiologically inert water-soluble carrier urea exhibited higher absorption and excretion after oral administration than sulfathiazole alone. Recent work on solid dispersions has been extended to the development of sustained-release preparations.

DEFINITIONS AND METHODS OF PREPARATION

Definitions

Chiou and Riegelman (2) defined the term solid dispersion as “a dispersion of one or more active ingredients in an inert carrier or matrix at solid state prepared by the melting (fusion), solvent, or melting-solvent method.” Dispersions obtained through the fusion process are often called melts, and those obtained by the solvent method are frequently referred to as coprecipitates or coevaporates. Examples include sulfathiazole-providone (PVP; 4) and reserpine PVP (5).

Methods of Preparation

The two basic procedures used to prepare solid dispersions are the fusion and cosolvent techniques. Modifications of

these methods and combinations of them have also been used (2). Recently, application of supercritical fluid process has been explored to form pharmaceutical solid dispersions (6).

Melting or Fusion Method

This method was first reported by Sekiguchi and Obi (3). A physical mixture of an active agent and a water-soluble carrier is heated until it is melted. The melt is solidified rapidly in an ice bath under vigorous stirring, pulverizing and then sieving. Rapid congealing is desirable because it results in supersaturation of the drug as a result of entrapment of solute molecules in the solvent matrix by instantaneous solidification. The solidification process can be achieved on stainless steel plates attached to a cooling system to favor rapid heat loss. Spray congealing from a modified spray drier onto a cold metal surface has also been used. Products from this process can be obtained in pellet form without the necessity of a grinding step that may alter crystalline modification.

Two advantages of the melt method are its simplicity and its economy, as no solvents are involved. However, the method may not be suitable if the drug or the carrier is unstable at the fusion temperature or evaporates at high temperatures. Succinic acid, for example, used as a carrier for griseofulvin, is quite volatile and partially decomposes by dehydration near its melting point. Such problems can be avoided by melting in a sealed container under vacuum or under an inert gas, such as nitrogen. By proper selection of carrier system and composition, the melting point of a binary system can be much lower than the melting point of either of the components.

Other disadvantages of this method may include the tacky and intractable nature of the resulting solidified melt and irregular crystallization owing to the presence of a miscibility gap on the phase diagram for a given drug-carrier system.

Solvent Method

Tachibana and Nakamura (7) first used this method to prepare a solid dispersion of β -carotene in PVP by using

chloroform as a cosolvent. The solvent is usually removed by evaporation under reduced pressure at varying temperatures. The choice of solvent and its removal rate are critical to the quality of the dispersion. A mixed solvent system may be used. Some examples of solid dispersions prepared by this method include sulfathiazole-PVP (4), reserpine-PVP (5), reserpine-deoxycholic acid (8) and griseofulvin-PVP (9).

The freeze-drying process has been used to prepare dispersions of ketoprofen (10) and dicumarol (11) in PVP from their ammoniacal solutions. Similarly, the spray-drying process has been used to prepare dispersions of acetohexamide in PVP (12) and chlorthalidone in pentaerythritol (13).

The major advantage of the solvent method is that thermal decomposition of drugs and carriers associated with the fusion method can be avoided. The disadvantages include the higher cost of preparation, the use of large quantities of solvent and the difficulty in complete removal of solvent, the possible adverse effect of residual solvent, the selection of a common volatile solvent, the difficulty of reproducing crystal forms, and the inability to attain a supersaturation of the solute in the solid system unless the system goes through a highly viscous phase.

Supercritical Fluid Process

Supercritical CO₂ is a good solvent for water-insoluble as well as water-soluble compounds under suitable conditions of temperature and pressure. Therefore, supercritical CO₂ has potential as an alternative for conventional organic solvents used in solvent-based processes for forming solid dispersions due to its favorable properties of being nontoxic and inexpensive. The process developed by Ferro Corporation (14) consists of the following steps: 1) charging the bioactive material and suitable polymer into the autoclave; 2) addition of supercritical CO₂ under precise conditions of temperature and pressure, that causes polymer to swell; 3) mechanical stirring in the autoclave; and 4), rapid depressurization of the autoclave vessel through a computer-controlled orifice to obtain desired particle size. The temperature conditions used in this process are fairly mild (35–75°C), which allows handling of heat sensitive biomolecules, such as enzymes and proteins.

CLASSIFICATION OF SOLID DISPERSIONS

Chiou and Riegelman (2) classified solid dispersions into the following six representative types: 1) simple eutectic mixtures; 2) solid solutions; 3) glass solutions and

glass suspensions; 4) amorphous precipitations in a crystalline carrier; 5) compound or complex formation; and 6) combinations of the previous five types. Many techniques have been used to characterize the physical nature of solid dispersions. These include thermal analysis (e.g., cooling-curve, thaw-melt, differential scanning calorimetry and x-ray diffraction, microscopic, spectroscopic, dissolution rate, and thermodynamic methods) Usually, a combination of two or more methods is required to obtain a complete picture of the solid dispersion system.

Simple Eutectic Mixtures

These are prepared by rapid solidification of the fused melt of two components that show complete liquid miscibility but negligible solid–solid solubility. Thermodynamically, such a system is an intimately blended physical mixture of its two crystalline components. Thus, the x-ray diffraction pattern of a eutectic constitutes an additive composite of the two components. A phase diagram of a two-component system is shown in Fig. 1. Examples of this type include phenacetin-phenobarbital (15), chloramphenicol-urea (2), griseofulvin-succinic acid (16), paracetamol-urea, and the dispersions of griseofulvin and tolbutamide in polyethylene glycol-(PEG-2000; 17).

Solid Solutions

In a solid solution, the two components crystallize together in a homogeneous one-phase system. The particle size of

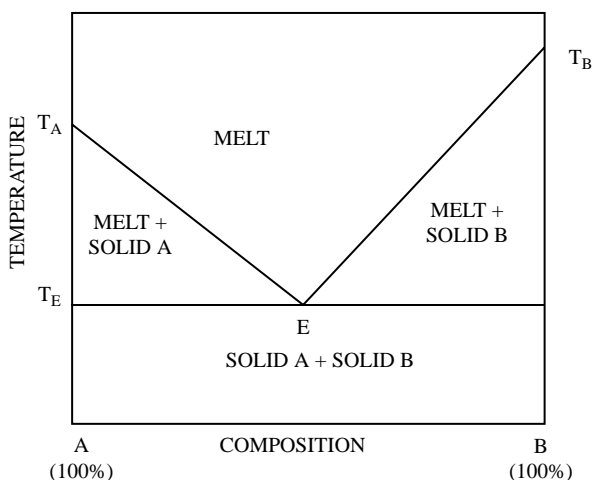


Fig. 1 Simple binary-phase diagram with eutectic formation. T_A is melting point of pure A; T_B is melting point of pure B; and E is eutectic point.

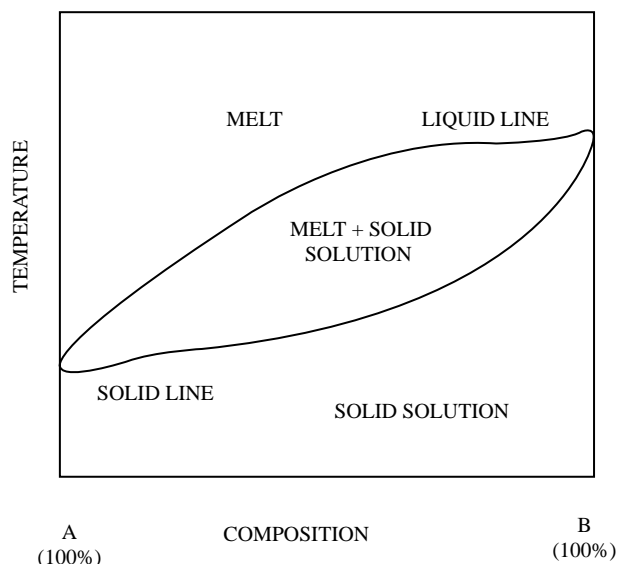


Fig. 2 A phase diagram of continuous solid solution for a binary system A and B.

the drug in the solid solution is reduced to its molecular size. Thus, a solid solution can achieve a faster dissolution rate than the corresponding eutectic mixture.

Solid solutions can be classified by two methods. According to the extent of miscibility of the two components, they may be classified as continuous or discontinuous. In continuous solid solutions, the two components are miscible in the solid state in all proportions. Typical phase diagrams of continuous and discontinuous solid solutions are shown in Figs. 2 and 3, respectively. Discontinuous solid solutions exist at extremes of composition. In general, some solid-state solubility can be expected for all two-component systems.

According to the criterion of molecular size of the two components, the solid solutions are classified as substitutional or interstitial. In the substitutional type, the solute molecule substitutes for the solvent molecule in the crystal lattice (Fig. 4). The molecular size of the two components should not differ by more than 15%. This class is represented by solid solutions of p-dibromobenzene-p-chlorobromobenzene, anthracene-acenaphthene, and ammonium and potassium thiocyanate.

An interstitial solid solution is obtained when the solute (guest) molecule occupies the interstitial space (Fig. 4) in the solvent (host) lattice. For this to occur, the solute molecule diameter should be less than 0.59 times that of the solvent molecule; therefore, the volume of the solute molecule should be less than 20% of the solvent molecule. Owing to their large molecular size, polymers favor the

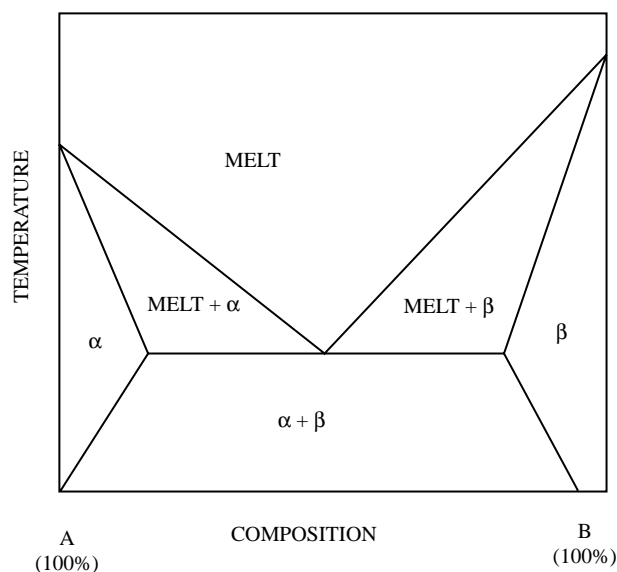


Fig. 3 A typical phase diagram of a discontinuous solid solution for a binary system A and B; α and β are regions of solid solution formation.

formation of interstitial solid solutions. Examples of this type include solid solutions of digitoxin, methyltestosterone, prednisolone acetate, and hydrocortisone acetate in the matrix of PEG-6000. They all exhibit a fast rate of dissolution.

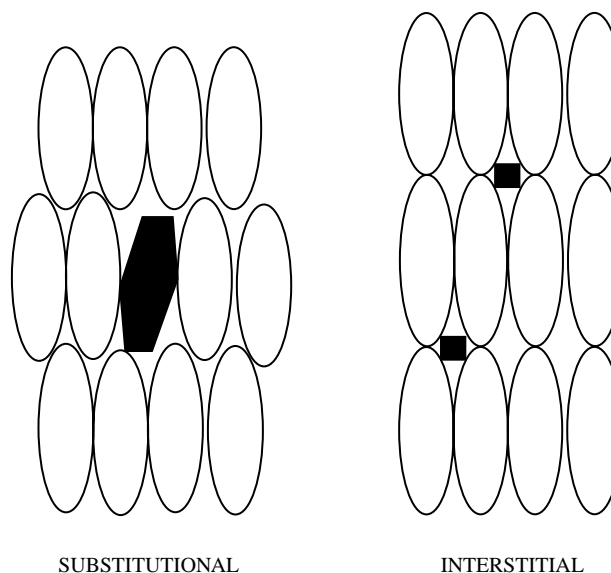


Fig. 4 Schematic representation of substitutional and interstitial solid solutions. Dark symbols represent solute atoms or molecules; open symbols indicate solvent atoms or molecules.

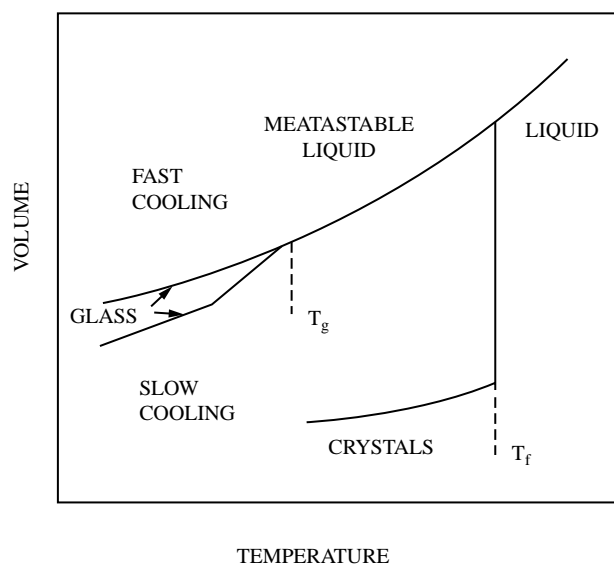


Fig. 5 Volume changes associated with cooling of a melt: T_g is the glass transition temperature and T_f is the melting point of the material.

Glass Solutions and Suspensions

A glass solution is a homogeneous glassy system in which a solute dissolves in the glassy carrier. A glass suspension refers to a mixture in which precipitated particles are suspended in a glassy solvent. The glassy state is characterized by transparency and brittleness below the glass transition temperature. Glasses do not have sharp melting points. Instead, they soften progressively on heating. The lattice energy, which represents a barrier to rapid dissolution, is much lower in glass solutions than in solid solutions. Fig. 5 shows the volume changes associated with glass formation when a melt is cooled down. Examples of carriers that form glass solutions and suspensions include citric acid, sugars such as dextrose, sucrose, and galactose, PVP, urea, and PEG.

Amorphous Precipitations in a Crystalline Carrier

This type of solid dispersion is distinguished from a simple eutectic mixture by the fact that the drug is precipitated out in an amorphous form. In a simple eutectic mixture, the drug is precipitated out in a crystalline form. An example of this is the precipitation of sulfathiazole in the amorphous form in crystalline urea (3). It is postulated that a drug with a propensity to supercooling has more tendency to solidify as an amorphous form in the presence of a carrier.

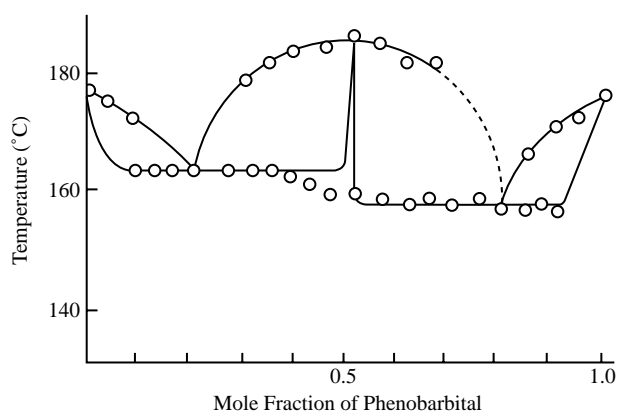


Fig. 6 Quinine-phenobarbital system, showing molecular compound formation. (Reproduced from Ref. 15. Copyright the American Pharmaceutical Association.)

Compound or Complex Formation

When two substances form a molecular compound, it usually gives rise to a maximum in the phase diagram. An example of this is the quinine-phenobarbital system (15) shown in Fig. 6. It is difficult to generalize the influence of complex formation on dissolution. A complex between digoxin and hydroquinone exhibited a high dissolution rate (16), whereas the insoluble complex between phenobarbital and PEG was shown to reduce both the rates of dissolution and the permeation of phenobarbital through everted rat gut (17).

MECHANISM OF INCREASED DISSOLUTION RATE

The enhancement in dissolution rate as a result of solid dispersion formation, relative to pure drug, varies from as high as 400-fold (18) to less than twofold. Corrigan (19) reviewed the current understanding of the mechanism of release from solid dispersions. The increase in dissolution rate for solid dispersions can be attributed to a number of factors. It is very difficult to show experimentally that any one particular factor is more important than another. The main reasons postulated for the observed improvements in dissolution of these systems are as follows:

1. Reduction of particle size. In the case of glass, solid solutions, and amorphous dispersions, particle size is reduced to a minimum level. This can result in an enhanced dissolution rate due to an increase in both the surface area solubilization.

2. Solubilization effect. The carrier material, as it dissolves, may have a solubilization effect on the drug. This was shown to be the case for acetaminophen and chlorpropamide in urea, as well as for numerous other drugs (20).
3. Wetability and dispersibility. The carrier material may also have an enhancing effect on the wetability and dispersibility of the drug in the dissolution media. This should retard any agglomeration or aggregation of the particles, which can slow the dissolution process.
4. Metastable forms. Formation of metastable dispersions with reduced lattice energy would result in faster dissolution rates. It was found that the activation energies for dissolution for furosemide was 17 kcal per mol, whereas that for 1:2 furosemide: PVP coprecipitate was only 7.3 kcal per mol (21).

RECENT ADVANCES

Serajuddin (22) reviewed recent advances in solid dispersion technology with particular emphasis on self-emulsifying systems. The advances in filling solid dispersion directly into hard gelatin capsules and availability of surface active and self-emulsifying carriers have renewed interest in commercial development of drug products based on solid dispersion. For ease of manufacturing, the carrier must be amenable to liquid filling into hard gelatin capsules as melts. The melting temperatures of carriers should be such that the solutions do not exceed approximately 70°C, which is the maximum acceptable temperature for hard gelatin capsule melts. Two such surface-active carriers that are being explored in solid dispersions for the bioavailability enhancement are Gelucire 44/14 (Gattfosse Corp, France) and Vitamin E TPGS NF (Eastman, Kingsport, TN). Gelucire 44/14 is a mixture of glyceryl and PEG-1500 esters of long chain fatty acids (lauryl macroglycerides). The suffixes 44 and 14 refer to its melting point and HLB value respectively. Serajuddin et al. also has shown that appropriate combinations of polysorbate 80 and PEG yield self-emulsifying systems.

SUSTAINED-RELEASE SOLID DISPERSIONS

More recently, the concept of solid dispersions has been explored using insoluble carrier materials. These systems are suitable for formulating sustained-release dosage forms. Hasegawa et al. (23, 24) prepared sustained-release

dosage forms of nifedipine by forming solid dispersions with anionic polymers, such as hydroxypropylmethyl cellulose phthalate and methacrylic acid-methacrylic acid methyl ester copolymers. Nifedipine in these solid dispersions was amorphous and was practically insoluble in gastric fluid (pH 1.2). However, it dissolved rapidly in intestinal fluid (pH 6.8) and showed a supersaturation phenomenon. These solid dispersions provided sustained absorption of nifedipine in beagle dogs with good availability after oral administration. Stability studies indicated that these dispersions were stable for at least 6 months under accelerated conditions. Similar dispersions of digoxin and dipyridamole (25) also showed delayed absorption with good bioavailability. Moreover, the chemical stability of digoxin in acidic medium was improved.

Fassihi et al. (26) used a combination of hydrophilic and lipophilic polymers to control the release rate. Thus, solid dispersions of theophylline were prepared by the fusion method using various ratios of PEG-6000, ethyl cellulose, and acrylic/methacrylic esters.

Takahashi et al. (27) showed that the coprecipitates of cationic water-soluble drugs (e.g., thioridazine hydrochloride), with pectin can be used as sustained-release preparations.

Controlled-release formulations of acetaminophen, aminopyrine, chlorpheniramine maleate, and salicylic acid that use Eudragit RSPM as a water-insoluble carrier, prepared by the solvent method, have been reported (28). A novel approach that uses a less soluble derivative of the drug as a carrier was used by Yang and Swarbrick (29) to prepare sustained-release solid dispersions of dapsone.

SELECTION OF A CARRIER

The properties of the carrier have a major influence on the dissolution characteristics of the dispersed drug. A carrier should meet the following criteria to be suitable for increasing the dissolution rate of a drug: 1) be freely water-soluble with intrinsic rapid dissolution properties; 2) be nontoxic and pharmacologically inert; 3) be heat stable with a low melting point for the melt method; 4) be soluble in a variety of solvents and pass through a vitreous state upon solvent evaporation for the solvent method; 5) be able to, preferably, increase the aqueous solubility of the drug; and 6) be chemically compatible with the drug and not form a strongly bonded complex with the drug.

Table 1 Materials used as carriers for solid dispersions

Sugars: Dextrose, sucrose, galactose, sorbitol, maltose, xylitol, mannitol, lactose
Acids: Citric acid, succinic acid
Polymeric materials: Povidone (PVP), polyethylene glycols (PEG), hydroxypropyl-methylcellulose, methylcellulose, hydroxyethylcellulose, cyclodextrins, hydroxypropylcellulose, pectin, galactomannan
Insoluble or enteric polymers: Hydroxypropylmethylcellulose phthalate, Eudragit L-100, Eudragit S-100, Eudragit RL, Eudragit RS
Surfactants: Polyoxyethylene stearate, Renex, Poloxamer 188, Texafor AIP, deoxycholic acid, Tweens, Spans
Miscellaneous: Pentaerythritol, pentaerythrityltetracetate, urea, urethane, hydroxyalkylxanthins.

(Adapted from Ref. 19).

Table 1 shows a list of materials used as carriers for solid dispersion formation. An excellent review of many of these carriers is included in an article by Ford (30). Enteric polymers are useful in the formation of solid dispersions of acid labile drugs. In some cases, a combination of carriers is more useful.

ADVANTAGES AND DISADVANTAGES OF SOLID DISPERSIONS

Among the advantages of solid dispersions are the rapid dissolution rates that result in an increase in the rate and extent of the absorption of the drug, and a reduction in presystemic metabolism. This latter advantage may occur due to saturation of the enzyme responsible for biotransformation of the drug, as in the case of 17- β -estradiol (31); or inhibition of the enzyme by the carrier, as in the case of morphine-tristearin dispersion (32). Both can lead to the need for lower doses of the drug. Other advantages include transformation of the liquid form of the drug into a solid form (e.g., clofibrate and benzoyl benzoate can be incorporated into PEG-6000 to give a solid (33), avoidance of polymorphic changes and thereby bioavailability problems, (as in the case of nabilone and PVP dispersion (34), and protection of certain drugs by PEGs (e.g., cardiac glycosides) against decomposition by saliva to allow buccal absorption (35).

The major disadvantages of solid dispersions are related to their instability. Several systems have shown changes in crystallinity and a decrease in dissolution rate with aging. The crystallization of ritonavir from the

supersaturated solution in a solid dispersion system was responsible for the withdrawal of the ritonavir capsule (Norvir, Abbott) from the market (36). Moisture and temperature have more of a deteriorating effect on solid dispersions than on physical mixtures. Some solid dispersions may not lend themselves to easy handling because of tackiness.

REVIEW OF IN VIVO STUDIES

Several examples of in vivo performance of solid dispersions have been published. The in vivo performance of solid dispersion systems containing sulfathiazole-urea, chloramphenicol-urea, reserpine-bile acids, and griseofulvin-PEGs were reviewed by Chiou and Riegelman (2). Some other representative examples are summarized below:

1. Nifedipine. Solid dispersions of nifedipine with PVP (37) and enteric polymers (23, 24) were evaluated in dogs and humans. PVP dispersions showed rapid absorption with a threefold increase in bioavailability when compared to physical mixtures. Sustained blood levels were obtained from solid dispersions with enteric polymers.
2. Acetaminophen. The effect of aging of the solid dispersion of acetaminophen in PEG-20,000 on human bioavailability was studied (38). It was found that the bioavailability of the sample stored at room temperature for 9 months decreased. This was due to an increase in the crystallinity of the drug.
3. Phenytoin. Solid dispersions of phenytoin in PEG-4000 (39), PEG-6000 (40), and PVP (41) were evaluated in vivo. The total areas under blood concentration curve (AUC) after oral administration to human volunteers were fourfold greater from a 1:10 PEG-4000 dispersion and 2.7-fold greater from a physical mixture than from phenytoin crystals. PEG-6000 dispersions (40% drug) were examined in mixed-breed dogs and compared with phenytoin sodium. Although phenytoin sodium dissolved several times faster in vitro than the solid dispersion of phenytoin, the two preparations were found to be bioequivalent. Sekikawa et al. (41) studied absorption of phenytoin in humans from PVP dispersions. The extent of bioavailability of phenytoin in phenytoin-PVP coprecipitate was 1.54 times greater than that of phenytoin alone.
4. Nitrofurantoin. Nitrofurantoin dispersed in PVP, PEG, and mannitol was evaluated in humans by studying cumulative urinary excretion of the drug (42). A linear

correlation was found between the amount of nitrofurantoin dissolved in acidic medium after 30 and 90 minutes and the cumulative amount of unchanged drug excreted after 12 h. The bioavailability relative to pure drug was 239% for a 1:4 PVP dispersion, 190% for a 1:4 PEG-6000 dispersion and 150% for a 1:10 mannitol dispersion. Stoll et al. (43) studied the absorption characteristics of various nitrofurantoin and nitrofurantoin-deoxycholic acid preparations. The 1:5 (w/w) nitrofurantoin-deoxycholic acid dispersion showed significant increases in both the rate and extent of absorption when compared with either the drug alone or the physical mixture.

5. Dicumarol. Sekikawa et al. (11) used the rabbit as a model to evaluate the performance of solid dispersions of dicumarol in PVP and β -cyclodextrin. Peak levels of the drug were observed at 4–6 h postadministration in the cases of the solid dispersion systems. In the case of dicumarol crystal powder, peak levels were observed at 2 to 12 hours postadministration. Average AUC values (0–48 h) of dicumarol following the administration of the dicumarol-PVP solid dispersion systems were 3.31 times (coevaporation method) and 1.54 times (freeze-drying method) that of control. The corresponding numbers for β -cyclodextrin dispersions were 2.18 and 1.72.
6. HIV Protease Inhibitor. The bioavailability of ritonavir (Norvir, Abbott), an HIV protease inhibitor, was enhanced by formulation as a solid dispersion in a mixture of surface active carrier, such as Gelucire 50/13, polysorbate 80, and polyoxyl-35 castor oil.

FUTURE OF SOLID DISPERSIONS

In spite of tremendous research activity in the area of solid dispersion since its introduction to pharmaceutical applications in 1961, only a few systems, such as Gris-PEG (Sandoz), a griseofulvin-PEG solid dispersion, Cesamet (Lilly), a nabilone-PVP solid dispersion, Sporanox capsules (Janssen), an itraconazole-HPMC solid dispersion, and Norvir capsules (Abbott), are identified as such in the market place. Most likely many others are utilized but masked as formulation methods. Perhaps the limited commercial utilization of the concept stems in part from the instability of solid dispersions on aging and the need for a high content of carrier, which might restrict its use to very potent drugs for physical and economic reasons. Nevertheless, solid dispersions have great potential both for increasing the bioavailability of drugs and for developing controlled-release preparations.

In regard to manufacturing considerations, the problem of total solvent removal in dispersions prepared by the solvent method needs to be addressed. The method created by Hasegawa et al. (23), which involves spray coating of nonpareils or any other inert core with drug-carrier solution, provides a one-step process of achieving a multiunit dosage form of solid dispersions. With particle-coating equipment now commercially available (44), this process has a promising future, as exemplified by commercial success of Sporanox capsules manufactured by this technique (45).

The problem of instability of the supersaturated state upon dissolution, which results in a stable form, has been dealt with by addition of a retarding agent. Methyl cellulose was used as a retarding agent in dispersions of indomethacin and flufenamic acid in PVP (46, 47). Computer optimization of such compositions, as described by Takayama, et al. (47), will certainly be of great value. More work is needed in these areas.

Novel formulating methods can provide cleaner manufacturing conditions. For example, Walker et al. (48) demonstrated the feasibility of liquid-filling gelatin capsules with the liquid melt and avoiding grinding-induced changes in crystallinity. This would be a very attractive feature for potent drug candidates. Exploration for new carrier systems will certainly continue. The most commonly used carriers are polymers, such as PVP and PEG. Recently lipids (49), freeze-dried milk (50), and self-emulsifying agents (22) have been used as carriers. Finally, the emergence of considerable research on novel drug-delivery systems may provide the greatest impetus for increased use of solid dispersions. Valuable preliminary studies of the use of solid dispersions to provide sustained- or controlled-release of drugs (23–27) have been reported. A U.S. patent (51) describes a method of preparation for a controlled release preparation of cyclosporine in biodegradable polymer, such as poly-D,L-lactide or a blend of poly-D,L-lactide and poly-D,L-lactide-co-glycolide. More and more commercial applications are expected in this area.

REFERENCES

1. Atkinsons, R.M.; Bedford, C.; Child, K.J.; Tomich, E. *Antibiot. Chemother.* **1962**, *12*, 232–238.
2. Chiou, W.L.; Riegelman, S. *J. Pharm. Sci.* **1971**, *60*, 1281–1302.
3. Sekiguchi, K.; Obi, N. *Chem. Pharm. Bull.* **1961**, *9*, 866–872.
4. Simonelli, A.P.; Mehta, S.C.; Higuchi, W.I. *J. Pharm. Sci.* **1969**, *58*, 538–549.

5. Bates, T.R. *J. Pharm. Pharmacol.* **1969**, *21*, 710–712.
6. Ye, C.; Frank, S. AAPS Annual Meeting Abstract. 1988.
7. Tachibana, T.; Nakamura, A. *Kolloid-Z. Polym.* **1965**, *203*, 130–133.
8. Malone, M.H.; Hochman, H.I.; Nieforth, K.A. *J. Pharm. Sci.* **1966**, *55*, 972–974.
9. Mayersohn, M.; Gibaldi, M. *J. Pharm. Sci.* **1966**, *55*, 1323–1324.
10. Takayama, K.; Nambu, N.; Nagai, T. *Chem. Pharm. Bull.* **1982**, *30*, 3013–3016.
11. Sekikawa, H.; Fukuya, W.; Takada, M.; Ohtani, K.; Arita, T.; Nakano, M. *Chem. Pharm. Bull.* **1983**, *31*, 1350–1356.
12. Graf, E.; Beyer, C.; Abdallah, O. *Acta Pharm. Technol.* **1982**, *8*, 225.
13. Bloch, D.W.; Elegakey, M.A.; Speiser, P. *Pharm. Acta Helv.* **1983**, *58*, 14–22.
14. Ferro Corporation, Independence, Ohio. Company Literature on Ferro ProcessTM Using Supercritical Fluid Technology.
15. Guillory, J.K.; Hwang, S.C.; Lach, J.L. *J. Pharm. Sci.* **1969**, *58*, 301–308.
16. Bochner, F.; Huffman, D.H.; Shen, D.D.; Azarnoff, D.L. *J. Pharm. Sci.* **1977**, *66*, 644–647.
17. Singh, P.; Guillory, J.K.; Sokolosky, T.D.; Benet, L.Z.; Bhatia, V.N. *J. Pharm. Sci.* **1966**, *55*, 63–68.
18. Said, S.A.; Fatatry, H.M.; Geneidi, A.S. *Aust. J. Pharm. Sci.* **1974**, *53*, 42–45.
19. Corrigan, O.I. *Drug Dev. Ind. Pharm.* **1985**, *11*, 697–724.
20. Roseman, M.; Jencks, W.P. *J. Amer. Chem. Soc.* **1975**, *97*, 631–640.
21. Shin, S.C. *Arch., Pharm. Res.* **1979**, *2*, 35.
22. Serajuddin, A.T.M. *J. Pharm. Sci.* **1999**, *88*, 1058–1066.
23. Hasegawa, A.; Nakagawa, H.; Sugimoto, I. *Chem. Pharm. Bull.* **1985**, *33*, 1616–1619.
24. Hasegawa, A.; Nakagawa, H.; Sugimoto, I. *Chem. Pharm. Bull.* **1985**, *33*, 388–391.
25. Hasegawa, A.; Kawamura, R.; Nakagawa, H.; Sugimoto, I. *Chem. Pharm. Bull.* **1986**, *34*, 2183–2190.
26. Fassihi, A.R.; Parker, M.S.; Pourkavoos, N. *Drug Dev. Ind. Pharm.* **1985**, *11*, 523–535.
27. Takahashi, Y.; Nambu, N.; Nagai, T. *Chem. Pharm. Bull.* **1986**, *34*, 327–332.
28. Athanikar, N. *Diss. Abstr. Int. B* 1978 **1977**, *38* (10), 4740–4741.
29. Yang, T.T.; Swarbrick, J. *J. Pharm. Sci.* **1986**, *75*, 53–56, 264–270.
30. Ford, J.L. *Pharm. Acta Helv.* **1986**, *61*, 69–88.
31. Resetarits, D.E.; Cheng, K.C.; Boltan, B.A.; Prasad, P.A.; Shefter, E.; Bates, T.R. *Int. J. Pharm.* **1973**, *2*, 113–123.
32. Chang, K.S.; Jarowski, C.I. *J. Pharm. Sci.* **1980**, *69*, 466–469.
33. Chiou, W.L.; Smith, L.D. *J. Pharm. Sci.* **1971**, *60*, 125–127.
34. Thakkar, A.; Hirsh, C.A.; Page, J.G. *J. Pharm. Pharmacol.* **1977**, *29*, 783–784.
35. Halpern, A.; Bradney, C.H. US Patent 2,698,822, 1955.
36. *Abbott to Substitute Norvir Capsules with Liquid Formulation*; Abbott Laboratories, 1998; press release, July 29.
37. Sugimoto, I.; Kuchiki, A.; Nakagawa, H.; Tohgo, K.; Kondo, S.; Iwane, I.; Takahashi, K. *Drug Dev. Ind. Pharm.* **1980**, *6*, 137–160.
38. Vila, Jato, J.L.; Vlanco, J.; Martinez, R.; Concheiro, A. *Biopharmaceutics Pharmacokinetics, Second European Congress*; Aiache, J.M., Hirtz, J., Eds.; Lavoisier: Paris France, 1984; 353–359.
39. Yakou, S.; Umehara, K.; Sonobe, T.; Nagai, T.; Sugihara, M.; Fukuyama, Y. *Chem. Pharm. Bull.* **1984**, *32*, 4130–4136.
40. Stavchansky, S.; Gowan, W.G. *J. Pharm. Sci.* **1984**, *73*, 733–736.
41. Sekikawa, H.; Fujiwara, J.; Naganuma, T.; Nakano, M.; Arita, T. *Chem. Pharm. Bull.* **1978**, *26*, 3033–3039.
42. Ali, A.A.; Gorashi, A.S. *Int. J. Pharm.* **1984**, *19*, 297–306.
43. Stoll, R.G.; Bates, T.R.; Swarbrick, J. *J. Pharm. Sci.* **1973**, *62*, 65–68.
44. Mehta, A.M.; Valazza, M.J.; Abele, S.E. *Pharm. Technol.* **1986**, *10* (4), 46–56.
45. Gilis, P.; De Conde, M.V.; Valentin, F.V.; Vandecruys, R.P.G. US Patent 5,633,015, May 27, 1997.
46. Takayama, K.; Nambu, N.; Nagai, T. *Chem. Pharm. Bull.* **1983**, *31*, 4496–4507.
47. Takayama, K.; Imaizumi, H.; Nambu, N.; Nagai, T. *Chem. Pharm. Bull.* **1985**, *33*, 292–300.
48. Walker, S.E.; Ganley, J.A.; Bedford, K.; Eaves, T. *J. Pharm. Pharmacol.* **1980**, *32*, 389–393.
49. Venkatram, S.; Rogers, J.A. *Drug Dev. Ind. Pharm.* **1985**, *11*, 223–238.
50. Macheras, P.E.; Reppas, C.I. *J. Pharm. Sci.* **1986**, *75*, 692–696.
51. Zeibun, R. US Patent 5,641,745, June 24, 1997.

COOLING PROCESSES AND CONGEALING

Richard Turton

West Virginia University, Morgantown, West Virginia, U.S.A.

Xiu Xiu Cheng

Andrx Pharmaceuticals, Inc., Fort Lauderdale, Florida, U.S.A.

INTRODUCTION

The purpose of this article is to introduce the major concepts and applications of cooling and congealing processes in the pharmaceutical industry. A cooling and congealing process must follow any process involving a melt. The melt-congealing technique has been practiced in the manufacture of wax-based suppositories for over 100 years. New applications have evolved quickly in the last 40 years to meet the increasing need for improving the solubility and consequent bioavailability of poorly water-soluble drugs and making modified-release drug products. Solid dispersions, matrix tablets, and coated drug-laden beads are some common examples. One of the methods of preparing solid dispersions is melt-congealing. The basic concept and process required to form solid dispersions of poorly water-soluble drugs in solid matrices is attributable to the work of Sekiguchi and Obi in 1961 (1). They melted a sulphathiazole-urea mixture of eutectic composition at above its eutectic temperature, solidified the dispersion in an ice bath, and pulverized it into a powder. A modification of the process involves spray-congealing from a modified spray-dryer onto coated metal surfaces and has been used for dispersions containing mannitol (2). Hot-melt coating and granulation processes have also been used in the development of wax-coated or wax-based matrix sustained-release products.

In this article, typical systems that make use of melt-congealing processes, the equipment and important operating variables used, typical matrix substrate materials, important processing considerations, and the advantages and disadvantages of using these systems are addressed.

PRINCIPAL CONCEPTS

Many review articles and many more research studies have been written on the use of systems involving the congealing and cooling of solid mixtures or solutions to produce both fast-release and sustained-release dosage

forms. There are many reasons for using such systems. An overview of the uses of congealed solids in the production of pharmaceutical dosage forms is shown in Fig. 1. From this figure we can see that the congealable solid may play an active role in controlling the rate of release of the active ingredient or may just be a cosmetic agent.

When the role of the congealable solid is to retard the rate at which drug is released, the congealed solid may be used as an overcoat on a substrate containing the active (e.g., a drug-laden nonpareil). In this case, typical coating equipment such as fluidized beds and perforated pans can be used. Alternatively, the congealable solid may be part of a drug-containing matrix that is permeable to gastric fluids or is erodable. In this case, the dosage form may be produced using a variety of methods, e.g., spray-congealing of a liquid (melt), granulation using a high-shear granulator, extrusion using a twin screw extruder, etc.

When the object is to increase the rate of drug dissolution, the drug is usually dispersed in a matrix composed of the congealable solid. The first step in this process is to form a solution or well-mixed dispersion of drug in the molten carrier. Subsequently, this mixture is cooled and solidified, and the resulting solid forms a solid matrix in which drug is very finely dispersed. This method can yield a solid dispersion in which the drug is dispersed at a molecular level, and the rate of dissolution of even very poorly soluble drugs can be enhanced significantly owing to the increase in exposed drug surface area. Spray-congealing and melt-granulation processes are often used to produce this type of product.

The use of a congealable solid as a taste-masking agent or for cosmetic purposes is also reported in the literature. Cosmetic uses usually involve a coating process, whereas taste-masking can be implemented using both overcoating and drug-carrier matrix formation.

Formation of Dispersions of Drugs in Carriers

One important reason for forming a drug dispersion in a carrier is to increase solubility of poorly water-soluble

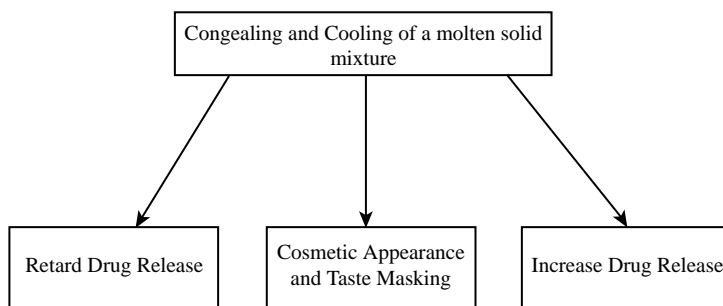


Fig. 1 General overview of uses of congealed solids in pharmaceutical dosage forms.

drugs. The dispersion of the active ingredient in a suitable matrix or carrier provides a large, exposed, drug surface area that compensates for the poor water solubility of the drug. In principle, a large surface area can be provided by fine milling (micronization) of the drug and subsequent dispersion of the drug powder in a carrier. However, owing to the cohesive nature of very fine particles, the effective dispersion of the micronized drug into the carrier can be problematic, resulting in maldistribution of the drug and consequent poor release characteristics. Despite this disadvantage, the rapid cooling and “freezing” of liquid carriers containing dispersions or slurries of fine drug powder are practiced, as is the melt granulation of a drug with a molten substrate. Another approach is to dissolve the drug into a liquid carrier and then to rapidly cool and congeal the mixture. The resulting solid will contain drug particles dispersed at the molecular or near molecular level within the solidified matrix carrier. Clearly, for this process to be effective, it is preferable that the drug is soluble in the molten carrier. A significant increase in drug-dissolution rate can then be achieved. It should be pointed out that the properties of both the drug and carrier, their interactions, the rate of cooling and congealing, and the conditions of storing the final dosage form all play an important role in the performance of the final product. All these factors are addressed in the sections below. The phase diagrams describing the behavior of different binary drug-carrier systems are reviewed in the following section.

Phase Diagrams

As an example of a liquid-solid phase diagram, we consider the salol-thymol system shown in Fig. 2 (3). This system represents a simple eutectic in which the two components are totally miscible in the liquid phase and totally immiscible in the solid phase. From Fig. 2, we see that the diagram is divided into four zones. In zone 1, both

components exist as a liquid solution; in zones 2 and 3, both components exist as a liquid solution in equilibrium with the other component in the solid form. In zone 4, both components exist as solids. Each zone boundary represents an equilibrium line. For example, consider a salol-thymol system containing 64 wt% thymol at a temperature of 48°C. This is denoted by point A on the diagram. If we slowly cool this solution (at constant pressure), then at approximately 30°C (point B) we will see that the first crystal of pure, solid thymol forms. If we cool the mixture more, pure thymol will continue to crystallize, and the resulting liquid solution becomes depleted of thymol. At approximately 20°C, we are at point C, and the resulting solution, which is in equilibrium with pure thymol (point D), has a composition of 48 wt% thymol, shown by point E. Continued cooling brings us to point F, where the solution composition is shown by point G. This is the eutectic composition. Further cooling beyond point G to point H, for example, gives rise to the formation of two solid phases, namely, pure salol (point I) and pure thymol

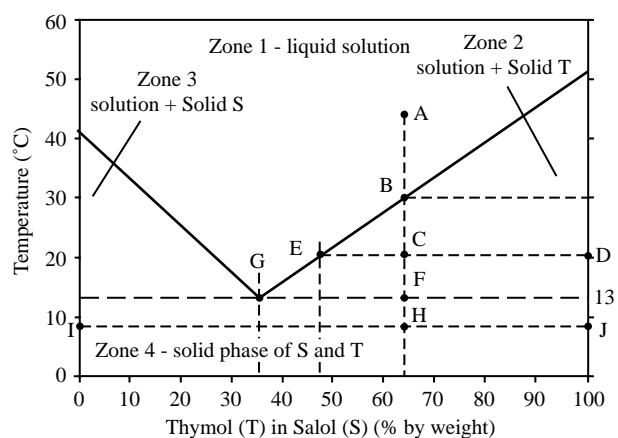


Fig. 2 Phase diagram for the thymol-salol system. (From Ref. 30.)

(point J). If we were to repeat this cooling process but start with a liquid solution with a composition to the left of the eutectic point (G), then the phase transformations would be similar except that pure salol would crystallize, and the remaining solution would become depleted in salol. The composition of this solution would move to the right and would eventually reach the eutectic composition G when the temperature reached 13°C. Further cooling would produce two solid phases of pure salol and pure thymol. The eutectic point (G) is the only condition at which a liquid solution of salol and thymol is in equilibrium with both pure solid thymol and pure solid salol. Recalling the Gibbs phase rule for systems in equilibrium:

$$F = 2 + C - P \quad (1)$$

For our salol–thymol system at the eutectic, we have three phases ($P = 3$) and two components ($C = 2$); therefore, we have only one degree of freedom ($F = 1$). For the system considered in Fig. 2, the pressure of the system is fixed, thus accounting for the single degree of freedom. Therefore, for such a system at a known pressure, there only exists a single temperature at which the eutectic can form, which is 13°C for this case.

The salol–thymol system presented is an example of a system that yields a solid dispersion when a solution is cooled rapidly. This solid dispersion consists of an intimate crystalline dispersion or mixture of one component distributed in the other. The uniformity of the dispersion of the two solid phases in each other is important to the dissolution behavior of the system. With all other factors being equal, the more uniform the dispersion, the faster and more reproducible will be the dissolution. It should be noted that the information given in Fig. 2 is accurate only if the system is in equilibrium. The conditions prevailing in any production process will most often fall short of providing equilibrium. Indeed, the idea of rapidly cooling a solution will inevitably lead to temperature and concentration gradients within the liquid and result in significantly different local congealing rates. The cooling and thickening (increase in viscosity) of the liquid act to trap crystals within the matrix and tend to minimize the diffusion and migration within the matrix. This has an added effect of limiting aggregation of the crystals and maintaining the drug in a well-dispersed form. Therefore, it can be seen that the design of the congealing device or equipment is of significant importance in the production of well-controlled dosage forms utilizing this technology. From Fig. 2, it is clear that rapid cooling of a solution with the eutectic composition can give rise to a well-dispersed solid dispersion and requires the least amount of cooling duty. However, the drug-loading in the

final dosage form, along with other factors, often precludes the congealing of eutectic mixtures in favor of other liquid compositions. Solid dispersions that exhibit negligible mutual solubility in the solid phase giving rise to eutectic mixtures, as previously described, are uncommon. Examples are given by Ford (4) and include paracetamol-urea, by Goldberg et al. (5), griseofulvin–succinic acid, by Goldberg et al. (6), and griseofulvin–PEG 4000 and 6000, by Chiou (7).

Another form of solid dispersion is the solid solution. The phase diagram for two materials (X and Y) that give rise to solid solutions is different from that previously discussed; an example of such a system is shown in Fig. 3 (8). This diagram is an example of a system in which both components (drug and carrier) have a limited solubility in each other in the solid phase. Consider the cooling of a liquid with composition and temperature given by point A. As the liquid is cooled, it will reach a point B on the equilibrium line. At this point, the first crystal of solid phase will be formed. The composition of this solid phase (α) is given by point C, and comprises a solid mixture of both X and Y. The composition of the liquid in equilibrium with this solid is given by point B. As the cooling continues, more solid will crystallize out of the solution, and the compositions of both the liquid solution and the solid change during this process. At point D, the eutectic temperature is reached. Cooling below the eutectic temperature causes two solid phases to crystallize from solution. The compositions of the two phases are given by the endpoints of the constant temperature line (points E and F); these two solid phases are in equilibrium with the remaining eutectic solution. Just as in the simple system illustrated in Fig. 2, there is a single eutectic

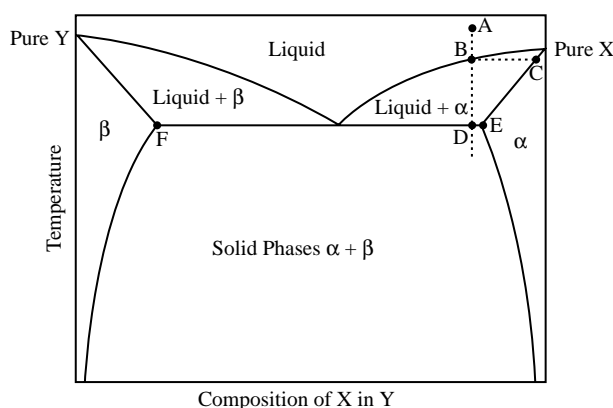


Fig. 3 Phase diagram for two components, X and Y, which are miscible in the liquid phase and mutually soluble in the solid phase. (From Ref. 8.)

temperature at which liquid and both solid phases can coexist. However, unlike the simple system, the composition of the solid phases will change with temperature.

Other structures may exist owing to interactions between the drug and carriers. Ford (4) addresses the following systems: glass solutions and glass suspensions, amorphous precipitations in a crystalline carrier, compound or complex formation, and combinations of these systems.

COMMON CARRIERS/EXCIPIENTS

Desirable Properties of Carriers

The properties of the congealed carriers clearly control the release of the drug from the dosage form. Reviews by Ford (4) and Achanta and associates (9) have been carried out in which the carrier properties are compared and contrasted, and these are summarized here. For hot-melt coatings, Achanta and associates (9) give the following ideal coating material characteristics:

1. It should be stable below 200°C.
2. It should have a melting point in the range of 75–80°C.
3. It should possess a narrow melting-point range and should not undergo softening before melting.
4. Its thermal behavior in the range of 30–200°C should be independent of its thermal history, method of preparation, and method of storage.
5. No crystal modifications should occur when it is exposed to temperatures up to 200°C.
6. It should be stable when subjected to thermal cycling.
7. It should possess a low melt viscosity to facilitate flow and spray formation.

These characteristics essentially ensure that the coating material is stable under typical coating conditions (items 4–6), that unwanted agglomeration is minimized (item 3); that it can be easily sprayed (item 7); and that operating temperatures are moderate so as to minimize drug degradation (items 1 and 2). In addition to these attributes, the coating material should clearly not interact with or adversely affect the drug's bioavailability.

For the formation of solid dispersions by congealing a melt of poorly water-soluble drug and carrier, Ford (4) recommends the following "ideal" characteristics for the substrate/carrier material:

1. It should be freely water soluble with intrinsic rapid dissolution properties
2. It should be nontoxic.

3. It should be chemically, physically, and thermally stable, with a low melting point. On solidification, it should crystallize rapidly and maintain the drug as a fine crystalline dispersion. The carrier and drug should be miscible in the liquid to avoid irregular crystallization and subsequent variability in dissolution rate.
4. The carrier should ideally increase the water solubility of the drug.
5. The carrier should not form stable complexes with the drug that would retard dissolution.
6. The carrier should be pharmacologically inert.

These characteristics ensure that, on solidification, the drug will be well dispersed with a large surface area available for dissolution (item 3); that the carrier will not inhibit the dissolution process or effect the drug (items 1, and 3–5); and that the carrier is inert (items 1 and 6). Above all, all of the carriers/excipients should have necessary regulatory approval for pharmaceutical use.

The carriers used in the production of sustained-release matrix tablets of water-soluble drugs by melt-congealing share some of the characteristics of carriers used for hot-melt coating and solid dispersion preparations. However, most of them have melting points in the range of 50–85°C.

Suitable Commercial Excipients

In general, an ideal carrier or coating material will not be available, and some compromises will have to be made. Perhaps some of the most important and common problems associated with the use of these carriers are their tendency to form crystalline polymorphs during the congealing process. In addition, there are subsequent changes in dissolution characteristics that occur on storage caused by changes in crystalline structure. The problems of changes in crystallinity are addressed later.

Congealed solids commonly used as carriers and coating materials have been summarized by Ford (4) and Achanta (9) and are:

Citric and Succinic Acids: Citric acid was used in solid dispersion preparations because of its high water solubility and its capability of glass formation. The monohydrate melts at approximately 100°C, whereas the anhydrous form melts at approximately 153°C. Drug degradation is quite probable at the higher temperature because the release of the water of hydration from the monohydrate form can affect moisture-sensitive drugs.

Bile Acids, Sterols, and Steryl Esters: Many of these compounds have high melting points, making them unsuitable for melt applications. Exceptions are the salicylic acid–cholic acid system formed by a melt-granulation technique reported by Froemming and

Vetter (10). The system provided a sustained release in acid media (pH 1–3.5).

Sugars: Again, many sugars have excessive melting points to be considered viable candidates for melt-congealing processes. However, the use of a few of the sugars has been reported in the literature. The application of zylitol in the formulation of the diuretic hydrochlorothiazide was demonstrated by Sirenus and coworkers (11). Hirasawa et al. (12) prepared naproxen solid dispersions by melting and rapid cooling with liquid nitrogen, using lactose as a carrier. Danjo and associates (13) prepared ethenzamide solid dispersions using sugars such as sucrose, maltose, galactose, and mannitol as carriers by melting and rapid cooling with liquid nitrogen. It was found that solid dispersions made with amorphous sucrose were more stable than those made with other sugars.

Urea: Urea has been used both to form solid dispersions by the cooling of urea–drug mixtures and to directly coat drug-laden particles in melt-coating operations. It has a melting point of approximately 130°C and is highly soluble in water. A variety of urea–drug dispersions obtained from melt processes have been reported in the literature, and these include aspirin, paracetamol, phenobarbitone, and tolbutamide. One disadvantage of using molten urea is its instability, resulting in the evolution of ammonia.

Polyethylene Glycols (PEG): These represent one of the most common materials used in the formation of solid dispersions from melts. The family of PEG polymers has molecular weights ranging from a few hundred to a few hundred thousand. Typical values for solid dispersions are 2000–20,000. Examples of solid dispersions of griseofulvin, indomethacin, and tolbutamide in PEG formed by the coagulation of PEG–drug melts have been reported extensively in the literature. PEG has the added advantage of increasing the water solubility of many drugs. This, along with its availability in many molecular-weight forms, makes it a popular material for the formation of solid dispersions.

Poly (Ethylene Oxide) (PEO) Polymers: In the same family as PEG, but having higher molecular weights are POLYOX[®] resins, which are water-soluble polymers. Because of their low melting point and unique swelling properties, coupled with the controlled rate of dissolution, POLYOX resins have been used to make sustained-release formulations by hot-melt extrusion (14).

Hydroxypropyl Methylcellulose (HPMC): Suzuki and Sunada (15a, 15b) report the production of solid dispersions prepared with nicotinamide and HPMC as combined carriers using nifedipine and nitrendipine as model drugs. Their solid dispersions were obtained using the fusion method. After both the drug and HPMC were dissolved in

the liquid melt of nicotinamide at 140°C, the fused mixture was cooled to solidify it.

Natural and Synthetic Waxes: The most common natural wax used as a congealing agent is carnauba wax derived from the wax palm or carnauba (*Copernicia cerifera*) found in Brazil. The major constituents of this wax are aliphatic esters, hydroxy esters, and methoxycinnamic and hydroxycinnamic aliphatic diesters. This is one of the hardest and highest-melting-point natural waxes (m.p. 82–85°C). It is used both as a coating agent, (9, 16) and as solid dispersion carrier (4, 17). Other natural waxes used in congealing applications include paraffin waxes and beeswax (18, 19). Synthetic waxes, such as stearic acid and stearyl alcohol, are often used in the formulation of sustained-release wax matrices by melt-congealing processes (20).

Hydrogenated Vegetable Oils: These primarily consist of the family of stearines. Hydrogenated castor oil (Cutina HR[®]) has been used as a wax matrix material both in melt-granulation (21) and spray-congealing processes (22).

Glycerides and Polyglycolized Glycerides: Most common carriers used in the preparation of sustained-release wax matrix formulations are glycerides and polyglycolized glycerides. There are several commercial products on the market, for example, Compritol 888 ATO (glyceryl behenate), Myverol 18092 (distilled monolinoleate), Myverol 18-99 (distilled monooleate), and Myveplex 600 (glycerol monostearate). They have been used as hydrophobic coatings (19), spray-congealing agents (11, 23), and retardant materials to form wax matrices (24).

Other Polymers and Excipients: Some other excipients have been used in the preparations of solid dispersions and sustained-release wax matrices by melt-congealing processes. The most common examples are the family of ethylacrylate and methylmethacrylate copolymers (Eudragits). Although these agents are generally not suitable for melt formation, they have been used successfully to modify the release of products formed by melt processes. For example, Miyagawa et al. (25) used both hydropropylcellulose and Eudragit L-100 to modify the release characteristics of diclofenac sodium from granules formed by the melt extrusion of a mixture of drug, release agents, and carnauba wax. The application of these excipients has also been demonstrated by Emori and colleagues (26).

IMPORTANT PROCESSES AND EQUIPMENT

Coating Processes

When the active ingredient is contained in an inert core material, for example, granulated with sucrose, and a

sustained-release coating of a congealable material is desired, then the equipment of choice is often the fluidized bed using a top-spray configuration (27). However, a bottom-spray configuration using a Wurster column insert can also be used. The fluidized bed offers the unique advantage of very high rates of heat and mass transfer from the molten solid to the particles to be encapsulated. Important processing variables are the temperature of the bed, temperature and flow rate of fluidizing air, temperature of the molten material and atomizing air; spray rate; and atomizing air pressure. According to Jones and Percel (28), the product temperature and droplet size of the molten material being sprayed (this is a function of atomization air pressure and spray rate) are the key variables. Special attention must also be given to the insulation of the spray nozzle within the bed to avoid remelting coated particles that come in contact with the nozzle during processing (9). The heat tracing of the lines containing the molten material is also imperative to avoid the solidification of coating material before reaching the nozzle. Jozwiakowski et al. (27) investigated the coating of a sugar-based granulation with a partially hydrogenated cottonseed oil (Durkee 07 Stearine, Durkee Industrial Foods) using a factorial experimental design. The melting point of this wax was 64°C, and it was found that excessive growth and rough coated surfaces were obtained when the bed temperature was held at 58°C. However, particles coated at bed temperatures of 54 and 50°C had uniform, smooth coated surfaces.

A novel fluidized-bed coating application that avoids the use of a spray nozzle for the delivery of the congealable solid additive was demonstrated by Kennedy and Niebergall (18, 19). They coated chlorpheniramine maleate (CPM)-loaded nonpareils with a variety of different waxes, including beeswax. The method of preparation involved loading a fluidized bed with both the nonpareils and the coating agent, both in powder form. The bed was then fluidized with heated air, the temperature of which was varied. The process consisted of four steps: 1) equipment warm-up, 2) preheating of substrate, 3) melting and spreading of the coating agent, and 4) cooling and subsequent solidification of the coating agent. Dissolution-release profiles for the nonpareils

coated with beeswax showed significant sustained-release characteristics. The coating of nonpareils from 10 to 35 mesh and particles up to 1 g are reported to be feasible.

Spray-Congealing Processes

The formation of drug dispersions within a congealed matrix is conveniently carried out in a spray-congealing apparatus. The process and important operating parameters for this type of equipment are addressed in detail by Killeen (29). The basic steps in the spray-congealing process are described in Fig. 4. First, the molten liquid, containing all the excipients, is fed to an atomizing nozzle. The atomized droplets then leave the nozzle and enter the cooling vessel or chamber, where they mix with chilled air. The droplets solidify and either fall to the bottom of the chamber or become entrained in the upward-moving airflow. The smaller particles that become entrained in the flow of air are separated from the air in a cyclone separator. The larger particles that settle to the bottom of the chamber are discharged. If necessary, the product can be size-separated as needed before further processing.

The atomizing nozzle may be one of four basic types: 1) dual-fluid, 2) single-fluid pressurized, 3) single-fluid disk or rotary, and 4) single-fluid ultrasonic. As with all congealing processes, the feed lines must be heat-traced to avoid unwanted solidification, and the nozzle should be insulated within the equipment to avoid the remelting of product coming in contact with it. For the best control of particle size and uniformity, the dual fluid or atomizing air nozzle is recommended. This nozzle can produce particles well below 50 μm , and because the airflow rate and pressure can be controlled separately, the size of the droplets can be controlled independently of the liquid flow rate. The air atomizing nozzles become problematic when the viscosity of the molten liquid is very high or the amount of solids loading in the melt is high. In these cases, the single-fluid nozzles are often preferred. Single-fluid pressurized (hydraulic) nozzles make use of the pressure of the fluid to cause atomization when flowing through the small orifice holes in the nozzle. These devices are suitable when the solids content of the melt is low and droplet size

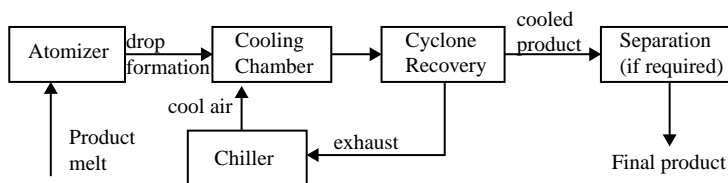


Fig. 4 Subprocesses occurring during spray congealing. (From Ref. 29.)

is not too small (for example, not less than 100–200 μm). Disk or rotary atomizers consist of a disk connected to a rotating shaft onto which the molten liquid flows. The liquid is broken up into drops by the spinning motion of the disk that produces a thin, radial outward-moving droplet spray pattern. This gives rise to a collection chamber that is short and wide, which requires less headroom than for processes using hydraulic and air atomized nozzles. The size distribution of the product is controlled by the liquid feed rate and the speed of rotation of the disk. These types of nozzle can process molten liquids that are corrosive, heat-sensitive, and contain less than 20% solids. In processes that use either hydraulic or rotary nozzles, the rate of cooling of the drops is relatively slow, and the flow of cocurrent and/or countercurrent air in the cooling chamber is significantly greater than for air atomizing nozzle systems. The use of ultrasonic nozzles for the manufacture of congealed solid products is relatively new. This type of nozzle uses the movement of an oscillating plate vibrating at ultrasonic frequencies to break up a liquid stream fed to the surface of the plate. The droplets formed are generally very small (less than 30 μm) and are suitable for low liquid flow rates. Because the droplet formation process imparts little directed kinetic energy to the drops, the spray pattern is quite small and controlled. In a recent study, Rodriguez and associates (22) demonstrated the efficacy of using an ultrasonic nozzle to produce spray-congealed products of theophylline and fenbufen in wax matrices.

The cooling chamber design must allow for the disengagement of the solidified droplets from the cooling air stream while also allowing the droplets sufficient residence time to cool and solidify. For a given liquid feed rate and composition, the amount of cooling is fixed. This is because the heat removal is equal to the latent heat of solidification plus the sensible heat of cooling the product, both of which are fixed by the liquid feed rate. The finer the droplets that are produced, the slower the air velocity must be to allow for disengagement; however, heat transfer tends to be rapid for small drops. Therefore, because the flow of air is essentially fixed owing to the cooling duty, this leads to wider and shorter cooling vessels. For larger droplets, particle disengagement can take place at higher gas velocities. However, heat transfer is slower, and droplet residence time must be longer. This leads to narrower and taller cooling vessels. Countercurrent and cocurrent airflows are possible as are mixed airflow patterns; several other design considerations are discussed by Killeen (29). Other important product parameters such as molten liquid viscosity, surface tension, specific gravity, and solids concentration, and processing parameters such as air flow rate, flow patterns,

spray angle, spray impingement, and nozzle resistance to abrasion are also addressed in this reference. According to Yajima et al. (30), the key product variable for the successful production of a wax matrix containing clarithromycin that masked the bitter taste of this macrolide antibiotic was the congealing speed. The important process variables were the liquid feed rate and atomizer (rotary) wheel speed.

Granulation Processes

The mixing of drug powder into a wax matrix and the subsequent uniform dispersion of the drug within the matrix can be carried out using a variety of granulating techniques. Thies and Kleinebudde (31) describe a process by which a hygroscopic drug, sodium valproate, was dispersed in a meltable binder, glycerol monostearate, using a high-shear mixer. The high-shear mixer was modified by adding a temperature bath that surrounded the mixing bowl. The prescreened drug was first added to the bowl and preheated and mixed for 10 min. The binder (glycerol monostearate) was then added to the mixer, and the granulation process began. Because of the hygroscopic nature of the drug, all formulation-related activities were performed in atmospheres with a relative humidity of less than 40%. The solidification temperature of the drug–binder combination was approximately 35°C. This temperature was approximately 35°C lower than the pure binder owing to the addition of the drug. The resulting granules, in the size range of 4 mm–500 μm , were cooled to room temperature and used for analysis. The effects of several process variables on the properties of the granules were investigated. These variables included the amount of binder, the process temperature, the granulation (massing) time, and the speed of the impeller.

Evrard et al. (21) investigated the influence of the melting and rheological properties of the binder in high-shear mixer granulation processes. These researchers concluded that successful granule formation took place in high-shear mixers using a variety of low-melting-point binders (Compritrol® 888, Cutina® HR, and Precirol® ATO5) at temperatures below their melting points. The binder only needed to be softened to produce granules. The melting range of the binder and the effect of temperature on the binder viscosity influenced the rate at which granules grew during the particle-formation process.

Melt-Extrusion Processes

The hot-melt-extrusion process is a well-established unit operation in the polymer pelleting and compounding

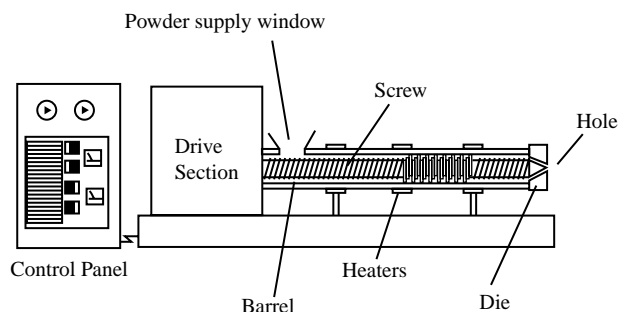


Fig. 5 Schematic diagram of a twin-screw compounding extruder. (From Ref. 25.)

industry. The concept is to feed a polymer powder along with desired additives, such as fillers, dyes, and other polymers, and then subject them to a high temperature and very high-shear environment. The temperature causes the polymers to melt, and the high shearing action of the screw(s) causes the other additives to be intimately mixed and dispersed within the polymer melt. Both single-screw and twin-screw units are common; the latter is capable of producing much higher pressures than is the single-screw unit. A schematic diagram of a twin-screw extruder is shown in Fig. 5. The extruder consists of essentially seven regions: the powder supply window, the barrel, the screws, the die, the drive section, the cooling/heating unit, and the control panel.

Because the process of polymer-compounding is very similar to the dispersion of a drug in a congealable solid matrix, it is not surprising that hot-melt extrusion techniques have been applied to the production of wax matrix formulations. Miyagawa et al. (25) used a twin-screw compounding extruder to prepare granules of diclofenac sodium dispersed in a carnauba wax matrix. They cited the following advantages of using such a device:

High kneading and dispersing ability can be achieved within the unit independently of the physical and chemical properties of the excipients.

The temperature in each zone of the unit can be controlled accurately.

The unit demonstrates superior extruding capability.

Residence time of powders in the barrel is minimized, thus minimizing the product degradation.

The unit is easy to clean.

Zhang and McGinity (14) also demonstrated the use of this type of equipment to formulate a sustained-release tablet of chlorpheniramine maleate (CPM) dispersed in a wax matrix consisting of polyethylene oxide (PEO). The content uniformity of their extruded tablets was in the range of 99–101% of theoretical drug content, which for a low-dose drug such as CPM was considered very good.

For this application, the sustained-release profiles were significantly affected by the addition of polyethylene glycol (PEG). The release rates increased with increasing amounts of PEG. PEG was also found to significantly lower the required processing temperature and reduce the required torque needed to extrude the tablets. These investigators concluded that the hot-melt-extrusion process was suitable for the preparation of these sustained-release tablets and that significant drug degradation, caused by the relatively high processing temperatures, was avoided because of the short residence time of the drug in the barrel (2–3 min).

IMPORTANT APPLICATIONS

Sustained-Release Dosage Forms Made Using Melt-Congealing Processes

Many prolonged- or sustained-release dosage forms are made using melt-congealing processes. The dosage forms can be melt-congealed matrix granules in the form of capsules or compressed into tablets or melt-congealed coatings on drug-loaded substrates (particles or pellets in capsules). Wax matrices are often prepared by melt-granulation or melt-extrusion processes, whereas microspheres or microparticles are usually produced by spray-congealing processes, and coated particles or pellets can be made by hot melt-coating processes.

The formation of drug containing wax-coated microspheres can be considered a coating process, although it is more correct to describe it as an encapsulation process involving a phase-inversion technique. Giannola et al. (16) demonstrated this technique by producing carnauba wax microspheres loaded with valproic acid (VA). These microspheres were between 200 and 425 μm in diameter, with an average drug content of 26% w/w. The process used to manufacture this product consisted of melting carnauba wax in an oil bath at 110°C and then adding a sample of the VA. To this homogeneous melt was added an acidic aqueous solution (pH 4.5), which minimized the solubility of the VA, combined with glycerine, which increased the melting point. A small amount of surfactant (Tween or Spans[®]) was added, and then the mixture was stirred at a predetermined rate. On dispersion in the aqueous medium, the molten mass formed spherical particles. The mixture was then quenched by the addition of iced water, with the result that the carnauba wax solidified and enveloped the drug.

Miyagawa and coworkers (25) used a twin-screw extruder to prepare wax matrix granules (WMG)

consisting of diclofenac sodium (DS) as a model drug; carnauba wax as the matrix material; and hydroxypropylcellulose (HPC-SL), methacrylic acid and copolymer L (Eudragit L-100), and sodium chloride as rate-controlling agents. The dissolution behavior of DS from WMG was strongly influenced by granule formulation. The release rate can be manipulated by the addition of other rate-controlling agents.

Emori et al. (26) formulated wax matrix tablets for prolonged drug release using a melt-congealing process. They studied the addition of acrylic acid polymer on the release of drug from a wax matrix consisting of carnauba wax and stearyl alcohol in a 1:1 ratio.

Perez and colleagues (24) prepared sustained-release phenylpropanolamine HCl tablets and investigated the effects of varying wax levels and methods of matrix formulation on drug release. Two methods were used for the preparation of a drug-wax system: physical mixtures and solid dispersions. In the first method, the drug, wax, and diluent were blended in a Turbula mixer by geometric dilution for a total of 20 min. Then the mixture was compressed into tablets. For the solid-dispersion method, the wax was melted in a water bath at a temperature of 80–85°C. The drug was incorporated into the melted wax using constant stirring over a 5-min period, and then the mixture was allowed to cool until it solidified. The solidified mass was granulated using a Stokes oscillating granulator equipped with a no. 12 screen. Blending of milled material (wax and active ingredient) and diluent was done in the Turbula mixer for 20 min. The tablets that were prepared as a physical mixture gave higher drug release than did tablets prepared by the solid-dispersion process.

The formation of a barrier coat on drug-laden substrates, to protect the drug from gastric fluid or to provide sustained release of the drug, can be achieved by the hot-melt fluid-bed coating process. Processing conditions must be chosen to prevent solidification of the coating material in the feed lines and to obtain smooth, uniform coats without excessive spray-drying and agglomeration. Kennedy and Niebergal (19) coated chlorpheniramine maleate (CPM)-loaded nonpareils with hydrophobic coating agents such as beeswax, paraffin, etc. in a hot-melt fluid-bed coating process. They demonstrated the ability to extend the release profile of a highly water-soluble drug, CPM.

Solid Dispersions for Improving the Solubility of Poorly Water-Soluble Drugs

As previously noted, the use of solid dispersions to improve the solubility of poorly water-soluble drugs dates

back to 1961 (1). Pharmaceutical composition with good dissolution and bioavailability can be formulated from solid dispersions of pharmaceutically active ingredients. Solid dispersions also can be used in controlled-release formulations.

The principle of improving the solubility of poorly water-soluble drugs using solid dispersions has already been presented. One common method of preparing solid dispersions is melt-congealing. The carriers used to make solid dispersions of poorly water-soluble drugs have also been addressed.

Dispersions of drugs in matrices can be formed either by dissolution of the drug in a molten solid and subsequent solidification or by the mechanical dispersion of the drug in a molten liquid followed by solidification. If the drug is poorly water soluble, then the matrix material should be either highly water soluble or highly water permeable. The equipment used for the preparation of drug dispersions is quite varied and includes spray congealers, granulators, and melt extruders.

Nifedipine is a poorly water-soluble drug, and much research has been conducted to improve its solubility. Suzuki and Sunada (15a) prepared solid dispersions of nifedipine in a combined carrier of nicotinamide and hydroxypropyl methylcellulose (HPMC). The nicotinamide was melted at 140°C, and then the drug and HPMC were added into the melt and dissolved. The melt mixtures were cooled and solidified to form the nifedipine solid dispersions. The solubility of the drug in the solid dispersion was enhanced. The drug dissolution of this ternary dispersion system was influenced by the viscosity and weight fraction of HPMC, the solubility and weight fraction of the drug, and the humidity during storage.

Suzuki and Sunada (15b) also prepared solid dispersions of nifedipine with other combined carriers using a fusion (melt) method. The combined carriers were nicotinamide and four different water-soluble polymers: hydroxypropyl methylcellulose (HPMC), polyvinylpyrrolidone (PVP), partially hydrolyzed polyvinyl alcohol (PVA), and pullulan. HPMC, PVP, and PVA dissolved in the melt of nicotinamide and were effective in the amorphous formation of nifedipine in solid dispersions. In dissolution studies, the drug concentration for these dispersions increased to more than twice the intrinsic drug solubility.

Hirasawa and coworkers (12) prepared a naproxen solid dispersion by melting, followed by rapid cooling with liquid nitrogen using lactose as a carrier. The dissolution studies of naproxen indicated that the dissolution rate was markedly increased in solid dispersions compared with physical mixtures and pure drugs.

Doshi et al. (32) prepared solid dispersions of carbamazepine in polyethylene glycols (PEG), PEG 4000

and PEG 6000, by both the melt and solvent methods. A comparison of dissolution profiles of the solid dispersions indicated dramatic increases in the rate and extent of carbamazepine dissolution from the solid dispersions. The melt (congealing) method provided a significantly higher rate and extent of dissolution of carbamazepine than did the solvent method. In addition, the rate and extent of dissolution of carbamazepine were significantly greater when the solid dispersion was cooled (slowly) at room temperature compared with faster cooling with ice.

Van den Mooter et al. (33) prepared solid dispersions of tamazepam in PEG 6000 and polyvinylpyrrolidone K30 (PVP K30) by the fusion (melting) and cooling process. In contrast to the very slow dissolution rate of pure tamazepam, the dispersion of the drug in the polymers enhanced the dissolution rate considerably. This can be attributed to improved wettability and dispersibility and a decrease of the crystalline fraction of the drug.

Taste-Masking and Prevention of Environment Degradation

Yajima et al. (30) applied the spray-congealing technique to mask the bitter taste of clarithromycin, a macrolide antibiotic. An optimum wax matrix formulation was developed consisting of 30% clarithromycin (CAM), 60% glyceryl monostearate (GM), and 10% aminoalkyl methacrylate copolymer E (AMCE). The CAM wax matrix was made by a spray-congealing agglomeration process. AMCE was dissolved in melted GM at 12°C. CAM was added to the melt and homogeneously suspended. Subsequently, the suspension was transferred to a spray-dryer and atomized under various atomizer wheel speeds and liquid feed rates. It was found that a small spherical matrix with a smooth surface could be obtained with a high atomizer wheel speed and optimum liquid feed rate. This matrix also possessed excellent properties for taste-masking with small initial amounts of release and subsequent high rates of release.

The masking of unpleasant taste and the prevention of environmental degradation of the drug can also be achieved by applying a barrier coating in a melt-congealing process.

THE STABILITY AND STRUCTURE OF THE CONGEALED MATERIAL

Analytical Techniques

Differential scanning calorimetry (DSC), x-ray diffraction (XRD), and infrared spectroscopy are the common

techniques used in the characterization of the structure of the congealed solid. Thermal analytic methods, such as DSC and differential microcalorimetric analysis (DMA), are routinely used to determine the effect of solutes, solvents, and other additives on the thermomechanical properties of polymers such as glass transition temperature (T_g) and melting point. The x-ray diffraction method is used to detect the crystalline structure of solids. The infrared technique is powerful in detecting interactions, such as complexation, reaction, and hydrogen bonding, in both the solid and solution states.

Case Studies

Numerous studies have been reported in the literature that investigate the micro- and macrostructure of the congealed solid and its affect on the dissolution of the active ingredient. The structure of the solid phase formed on congealing is affected by the process by which the material is congealed, the formulation, and the conditions at which the solid product is stored.

Zhang et al. (14) produced matrix tablets containing chlorpheniramine maleate (CPM) dispersed in mixtures of polyethylene oxide (PEO) and polyethylene glycol (PEG) using melt-extrusion. Physical characterization of the samples was performed using differential scanning calorimetry (DSC), wide-angle x-ray diffraction (WAXD), infrared spectroscopy, and gel permeation chromatography. Results showed that the PEO and PEG consisted of ordered crystal regions interdispersed in random amorphous regions. WAXD verified that crystals of CPM and PEO existed in the matrix tablets. The effect of PEG was to lower the processing temperature and also the required torque necessary to extrude the melt through the die. This, in turn, had the effect of reducing damage to the polymer (and drug) during tablet formation. Polymer damage was primarily attributed to the intense shear force occurring during extrusion. This caused bond breakage and a reduction in molecular weight. The introduction of PEG caused a significant reduction in the melt viscosity and thus allowed a reduction in the processing temperature. An increase in the ratio of PEG to PEO also caused an increase in the rate of drug release during in vitro dissolution tests.

Emas and Nyquist (17) investigated the aging and stabilization of spray-congealed solid dispersions of carnauba wax. Isothermal microcalorimetry (IM) was used to measure the behavior of samples of freshly congealed wax and samples stored for 2 days at elevated temperatures of 40, 50, and 60°C, respectively. Results for the freshly obtained sample and the annealed samples are illustrated in Fig. 6, where the rate of change of heat content (dq/dt) is plotted as a function of testing time.

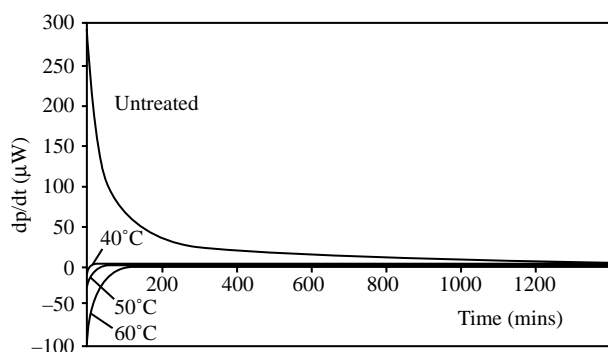


Fig. 6 Results of microcalorimetric measurements of spray-congealed carnauba wax. Untreated sample and samples annealed at different temperatures and stored for 5 days at ambient conditions; all measurements taken at 45°C. (From Ref. 17.)

Exothermic reactions are represented by positive deviations, whereas endothermic reactions are represented by negative deviations. From this figure, it can be clearly seen that the effect of annealing the samples at different temperatures is to reduce the magnitude of the exothermic reaction. The absolute values of the deviations depend on the temperature at which the experiments are run. However, the trend of reduced exotherms with increasing annealing temperature is followed for all temperatures. A series of experiments was also performed to evaluate the effect of long-term storage on the rate of change of heat content with time. Again, the results depend on the testing temperature, but the trends are similar. Fig. 7 illustrates the results for a wax sample annealed at 50°C for 2 days and then sealed and stored at room temperature for up to 12 months. From this figure, it can be seen that the initial

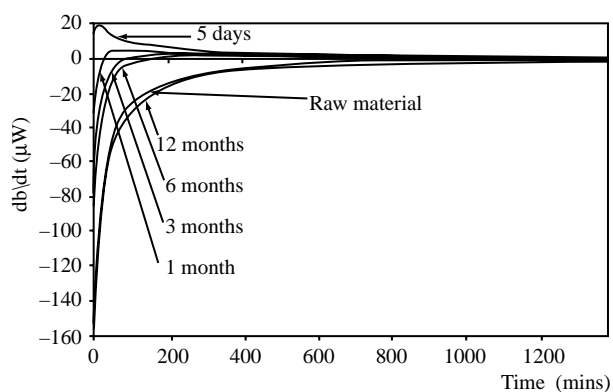


Fig. 7 Results of microcalorimetric measurements of spray-congealed carnauba wax. Samples annealed at 50°C for 2 days and stored for various times at ambient conditions all measurements taken at 45°C. (From Ref. 17.)

exothermic reaction is reduced in magnitude with increased storage temperature and that a stable condition (similar to that obtained for the raw wax) is approached asymptotically with time. From this work, it can be seen that spray-congealed carnauba wax initially exists in an unstable form that slowly changes to a stable form on storage (aging). This aging process can be accelerated by annealing the samples at an elevated temperature (40–60°C) for several days before storage. However, the rate of aging is a complex function of the annealing history because the wax appears to have different stable states at different temperatures.

Eldem et al. (34) investigated the polymorphic behavior of congealed micropellets of two different drugs (estradiol cypionate and medroxyprogesterone acetate) in two lipids, GTS-33 (glycerol tristearate) and Compritol 888 (glycerol behenate). Micropellets were formed from the drug and congealable matrix (tristearate-behenate), and then separate samples were taken, sealed, and stored for 6 months at temperatures of –18, 4, 25, and 37°C, respectively. DSC was used to investigate the stability of the samples after the storage period, and scanning electron microscopy (SEM) was used to measure the surface morphology of the microspheres. It was concluded that all the samples initially possessed an unstable polymorphic structure with a smooth surface morphology corresponding to the α -form, which is the result of rapid crystallization from the melt. According to Garti and Sato (35), the α -form possesses a very small crystal size accounting for the smooth surface morphology. The rate of attainment of the more stable β -form is dependent on the storage temperature. The presence of additives, such as lecithin, can also significantly affect the rate of change of crystalline structure. Higher storage temperatures increase the rate of polymorphic transformation, whereas the addition of lecithin acts as a stabilizer, slowing this transition.

McGinity et al. (36) investigated how the cooling process influenced the properties of solid dispersions prepared by congealing mixtures of drug and wax matrix. The system that was studied was tolbutamide in urea and in PEG (6000). Powder x-ray diffraction was used to determine the extent of crystallinity of these solid dispersions. For the tolbutamide–urea system, rapid cooling of the melt gave rise to distinct crystalline forms of the drug and wax. However, slow cooling of the melt in an oil bath at ambient conditions over a period of several hours yielded a solid that exhibited a complete absence of crystallinity. This amorphous solid did yield a crystalline structure after 5 months of storage, but the crystals were those of urea only. This is in contrast to simple physical mixtures of the two solids that clearly

showed crystalline mixtures of both tolbutamide and urea. In contrast to the tolbutamide–urea system, the tolbutamide–PEG system showed similar degrees of crystallinity for both the rapid-cooled and slow-cooled systems. Dissolution profiles for this system were compared for the rapid-cooled, slow-cooled, and physical mixture samples. The profiles for all three samples were very similar; however, the extent of release from the rapid-cooled sample was approximately 10% higher than that of the other two samples at any given time during the dissolution process.

Coben and Lordi (37) investigated the hardening of a variety of suppositories using a modified Krowczynski (38, 39) apparatus. The suppository bases were commercially available materials that consisted of a mixture of natural and synthetic waxes and fats. Samples were prepared from the different base materials by coagulating melted materials in 50-cavity brass molds. The samples were allowed to set at different storage temperatures (–5, 4, and 22°C), and this resulted in molding times of approximately 5, 15, and 30 min, respectively. After molding, samples from each batch were stored at three different temperatures of 4, 20, and 30°C, respectively. Both DSC and XRD were used to evaluate crystalline and polymorphic changes in the suppository base material. The results clearly indicated that the age-hardening phenomenon was solely attributable to a shift from an amorphous to crystalline structure and that polymorphic changes did not occur. Results also showed that the times required to resoften the samples at a given condition were dependent on the storage temperature but did not depend on the rate of solidification during the molding process.

From the case studies presented here, it is clear that the structural changes in the congealed material vary considerably and that it is difficult to identify and generalize these trends for a family of excipients. Nevertheless, it is fair to say that aging effects are common to many congealed materials and that carefully planned experimental studies must be conducted to evaluate the different phenomena causing these changes.

ADVANTAGES AND DISADVANTAGES

The advantages of using melt-congealing processes are numerous. Generally, no solvent is required in the formulation and manufacturing processes, and the subsequent environmental requirement of solvent capture and recycle is eliminated. Processing times are often much shorter because solvent evaporation is not required. For example, in coating processes, the time to congeal a molten

liquid is much shorter than that required to evaporate a solvent. In addition, undesirable drug–solvent interactions are also eliminated. For example, drugs that are highly water labile can be processed using melt processes.

There are also disadvantages to using melt-congealing processes. The drug must be stable at the temperature required to melt the carrier. For many drugs, the temperature at which degradation takes place is low and may preclude the use of all suitable matrix materials. Processing using hot-melt flows requires careful engineering to avoid feed-line solidification and unwanted agglomeration. Many of the matrix materials used to form dispersions and coatings undergo aging during storage, and this can affect the stability of the drug and/or the release rate of the drug.

SUMMARY

There are many applications of melt-congealing techniques in controlled-release dosage forms, improving the solubility of poorly water-soluble drugs and taste-masking of bitter drugs. Although the incorporation of an active ingredient into solid dispersions is a good way to enhance the solubility of poorly water-soluble drugs, the stability and aging problems associated with the congealed solids still need extensive additional study.

REFERENCES

1. Sekiguchi, K.; Obi, N. Studies on Absorption of Eutectic Mixtures. I. A Comparison of the Behaviour of Eutectic Mixture of Sulfathiazole and that of Ordinary Sulfathiazole in Man. *Chem. Pharm. Bull.* **1961**, *9*, 866–872.
2. Kanig, J.L. Properties of Fused Mannitol in Compressed Tablets. *Pharm. Sci.* **1964**, *53*, 188–192.
3. Martin, A.; Bustamante, P.; Chun, A.H.C. States of Matter. *Physical Pharmacy*, 4th Ed.; Lea and Febiger: Philadelphia, 1993; 22–52.
4. Ford, J.L. The Current Status of Solid Dispersions. *Pharm. Acta Helv.* **1986**, *61* (3), 69–88.
5. Goldberg, A.H.; Gibaldi, M.; Kanig, J.L. Increasing Dissolution Rates and Gastrointestinal Absorption of Drugs via Solid Solutions and Eutectic Mixtures. II. Experimental Evaluation of a Eutectic Mixture: Urea-Acetaminophen System. *J. Pharm. Sci.* **1966**, *55*, 482–487.
6. Goldberg, A.H.; Gibaldi, M.; Kanig, J.L. Increasing Dissolution Rates and Gastrointestinal Absorption of Drugs via Solid Solutions and Eutectic Mixtures. III. Experimental Evaluation of Griseofulvin–Succinic Acid Solid Solution. *J. Pharm. Sci.* **1966**, *55*, 487–491.

7. Chiou, WinL. Pharmaceutical Applications of Solid Dispersion Systems: X-ray Diffraction and Aqueous Solubility Studies on Griseofulvin-Polyethylene Glycol 6000 Systems. *Pharm. Admin.* **1977**, 66 (7), 989–991.
8. McGinity, J.W.; Coffin, M.D. Cooling Processes and Congealing. *Encyclopedia of Pharm. Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1990; 3, 318.
9. Achanta, A.S.; Adusumilli, P.S.; James, K.W.; Rhodes, C.T. Development of Hot Melt Coating Methods. *Drug Dev. Ind. Pharm.* **1997**, 23 (5), 441–449.
10. Froemming, K.-H.; Vetter, G. *Pharm. Ind.* **1975**, 37, 1051.
11. Sirenus, I.; Krogerus, V.E.; Leppanen, T.J. Dissolution Rate of p-Amino Benzoate from Solid Xylitol Dispersions. *Pharm. Sci.* **1979**, 68, 791–792.
12. Hirasawa, N.; Danjo, K.; Haruna, M.; Otsuka, A. Physicochemical Characterization and Drug Release Studies of Naproxen Solid Dispersions Using Lactose as a Carrier. *Chem. Pharm. Bull.* **1998**, 46 (6), 1027–1030.
13. Danjo, K.; Nakata, T.; Otsuka, A. Preparation and Dissolution Behavior of Ethenzamide Solid Dispersions Using Various Sugars as Dispersion Carriers. *Chem. Pharm. Bull.* **1997**, 45 (11), 1840–1844.
14. Zhang, Feng; McGinity, J.W. Properties of Sustained-Release Tablets Prepared by Hot-Melt Extrusion. *Pharm. Dev. Technol.* **1999**, 4 (2), 241–250.
- 15a. Suzuki, H.; Sunada, H. Some Factors Influencing the Dissolution of Solid Dispersions with Nicotinamide and Hydroxypropylmethylcellulose as Combined Carriers. *Chem. Pharm. Bull.* **1998**, 46 (6), 1027–1030.
- 15b. Suzuki, H.; Sunada, H. Influence of Water-soluble Polymers on the Dissolution of Nifedipine Solid Dispersions with Combined Carriers. *Chem. Pharm. Bull.* **1998**, 46 (3), 482–487.
16. Giannola, L.I.; De Caro, V.; Severino, A. Carnauba Wax Microspheres Loaded with Valproic Acid: Preparation and Evaluation of Drug Release. *Drug Dev. Ind. Pharm.* **1995**, 21 (13), 1563–1572.
17. Emas, M.; Nyqvist, H. Methods of Studying Aging and Stabilization of Spray-Congeaed Solid Dispersions with Carnauba Wax. 1. Microcalorimetric Investigation. *Int. J. Pharm.* **2000**, 197, 117–127.
18. Kennedy, J.P.; Niebergall, P.J. Development and Optimization of a Solid Dispersion Hot-Melt Fluid Bed Coating Method. *Pharm. Dev. Technol.* **1996**, 1 (1), 51–62.
19. Kennedy, JohnP.; Niebergall, PaulJ. Evaluation of Extended-Release Applications for Solid Dispersion Hot-Melt Fluid Bed Coatings Utilizing Hydrophobic Coating Agents. *Pharm. Dev. Technol.* **1998**, 3 (1), 95–101.
20. Motycka, S.; Nairn, J.G. Influence of Wax Coatings on Release Rate of Anions from Ion-Exchange Resin Beads. *J. Pharm. Sci.* **1978**, 67 (4), 500–503.
21. Evrard, B.; Amighi, K.; Beten, D.; Delattre, L.; Moes, A.J. Influence of Melting and Rheological Properties of Fatty Binders on the Melt Granulation Process in a High-Shear Mixer. *Drug Dev. Ind. Pharm.* **1999**, 25 (11), 1177–1184.
22. Rodriguez, L.; Passerini, N.; Cavallari, C.; Cini, M.; Sancin, P.; Fini, A. Description and Preliminary Evaluation of a New Ultrasonic Atomizer for Spray-Congealing Processes. *Int. J. Pharm.* **1999**, 183, 133–143.
23. Chang, Chin-Ming; Bodmeier, R. Swelling of and Drug Release from Monoglyceride-Based Drug Delivery Systems. *Pharm. Admin.* **1997**, 86 (6), 747–752.
24. Perez, M.A.; Ghaly, E.S.; Marti, A. Sustained Release Phenylpropanolamine Hydrochloride from ATO 888 Matrix. *Puerto Rico Health Sci. J.* **1993**, 12 (4), 263–267.
25. Miyagawa, Y.; Okabe, T.; Yamaguchi, Y.; Miyajima, M.; Sato, H.; Sunada, H. Controlled-Release of Diclofenac Sodium from Wax Matrix Granule. *Int. J. Pharm.* **1996**, 138, 215–224.
26. Emori, H.; Ishizaka, T.; Koishi, M. Effects of Acrylic Acid Polymer and its Arrangement on Drug Release from a Wax Matrix. *J. Pharm. Sci.* **1984**, 73 (7), 910–915.
27. Jozwiakowski, M.J.; Jones, D.M.; Franz, R.M. Characterization of a Hot-Melt Fluid Bed Coating Process for Fine Granules. *Pharm. Res.* **1990**, 7 (11), 1119–1126.
28. Jones, D.M.; Percel, P.J. Coating of Multiparticulates Using Molten Materials: Formulation and Process Considerations. *Multiparticulate Oral Drug Delivery*; Ghebre-Sellasie, I. Ed.; Marcel Dekker, Inc.: New York, 1994; 113–142.
29. Killeen, M.J. The Process of Spray Drying and Spray Congealing. *Pharm. Engineering* **1993**, 13 (4), 56–64.
30. Yajima, T.; Umeki, N.; Itai, S. Spray Congealing for Masking the Bitter Taste of Clarithromycin in Wax Matrix. *Chem. Pharm. Bull.* **1999**, 47 (2), 220–225.
31. Thies, R.; Kleinebudde, P. Melt Pelletisation of a Hygroscopic Drug in a High Shear Mixer Part 1. Influence of Process Variables. *Int. J. Pharm.* **1999**, 188, 131–143.
32. Doshi, D.H.; Ravis, W.R.; Betageri, G.V. Carbamezepine and Polyethylene Glycol Solid Dispersions: Preparation, In Vitro Dissolution, and Characterization. *Drug Dev. Ind. Pharm.* **1997**, 23 (12), 1167–1176.
33. Van den Mooter, G.; Augustijns, P.; Bleton, N.; Kinget, R. Physico-Chemical Characterization of Solid Dispersions of Temazepam with Polyethylene Glycol 6000 and PVP K30. *Int. J. Pharm.* **1998**, 164, 67–80.
34. Eldem, Turkan; Speiser, Peter; Altorfer, Hans. Polymorphic Behavior of Sprayed Lipid Micropellets and its Evaluation by Differential Scanning Calorimetry and Scanning Electron Microscopy. *Pharm. Res.* **1991**, 8 (2), 178–184.
35. Crystallization and Polymorphism of Fats and Fatty Acids. Garti, N., Sato, K., Eds.; Marcel Dekker, Inc.: New York, 1998; 31.
36. McGinity, James W. Maincent Philippe; Steinfink, Hugo. Crystallinity and Dissolution Rate of Tolbutamide Solid Dispersions Prepared by the Melt Method. *Pharm. Admin.* **1984**, 73 (10), 1441–1444.
37. Coben, Larry J.; Lordi, Nicholas G. Physical Stability of Semisynthetic Suppository Bases. *Pharm. Admin.* **1980**, 69 (8), 955–960.
38. Krowczynski, L. *Diss. Pharm.* **1959**, 11, 269.
39. Krowczynski, L. *Acta Pol. Pharm.* **1965**, 22, 1949–1954.

BIBLIOGRAPHY

- Aitken-Nichol, C.; Zhang, Feng; McGinity, J.W. Hot Melt Extrusion of Acrylic Films. *Pharm. Res.* **1996**, *13* (5), 804–808.
- Ansel, H.C.; Popovich, N.G. *Pharmaceutical Dosage Forms and Drug Delivery Systems*, 5th Ed.; Lea & Febiger: Philadelphia, 1990.
- Evrard, B.; Delattre, L. In Vitro Evaluation of Lipid Matrices for the Development of the Sustained-Release Sulfamethazine Bolus for Lambs. **1996**, *22* (2), 111–118.
- Hincal, A.A.; Kas, H.S. Preparation of Micropellets by Spray Congealing. *Multiparticulate Oral Drug Delivery*; Ghebre-Sellasie, Ist Ed.; Marcel Dekker, Inc.: New York, 1994; 17–34.
- Krape, P.J.; Chang, S.C.; Hein, W.A. II; Teleha, C.A. Process for the Manufacturing Paroxetine Solid Dispersions. US Patent, 5,955,475, 1999.
- Oshlack, B.; Chasin, M.; Minigue, J.J.; Kaiko, R.F. Controlled Release Oxycodone Compositions. US Patent 5,549,912, 1996.
- Oth, M.P. Sustained Release Solid Dispersions of Indomethacin with Eudragit RS and RL. *Int. J. Pharm.* **1989**, *55*, 157–164.
- Qui, Y.; Trivedi, J.S.; Graham, S.L.; Flood, K.M.; Krill, S.L. Tablet for the Controlled Release of Active Agents. US Patent, 6,033,685, 2000.
- Rudnic, E.M.; McCarty, J.A.; Belendui, G.W. Hydrophobic Drug Delivery Systems. US Patent, 5,430,021, 1995.
- Sugao, H.; Yamazaki, S.; Shiozawa, H.; Yano, K. Taste Masking of Bitter Drug Powder without Loss of Bioavailability by Heat Treatment of Wax-Coated Microparticles. *J. Pharm. Sci.* **1998**, *87* (1), 96–100.
- Suzuki, H.; Sunada, H. Influence of Water-Soluble Polymers on the Dissolution of Nifedipine Solid Dispersions with Combined Carriers. *Chem. Pharm. Bull.* **1998**, *46* (3), 482–487.

CONTRACT MANUFACTURING

Duncan E. McVean

Consultant, Solon, Ohio, U.S.A.

INTRODUCTION

Changes in Technology

Exciting changes are occurring in the pharmaceutical industry. Due to new advances in technology the industry is expanding more rapidly than ever before. For example, scientists are now using computers to manipulate molecules instead of having to go into the laboratory. Using computer simulations, they can design molecules that fit into cellular receptor sites. These “designer” molecules then produce desired therapeutic effects. Other scientists are using genetic manipulation to expand the realm of therapeutics. By inserting specific genetic material into mammalian, insect, and bacterial cells, cultures of manipulated cells produce hormones and other naturally occurring, physiologically active substances that are useful in the treatment of many disease states. Genes are also being inserted into embryos of farm animals; in adulthood, these animals are able to secrete therapeutically active substances in their milk and will become very low cost producers of substances that had either been in limited supply or had been very expensive to produce by other methods.

Changes in Core Business of Companies

To accommodate this changing technology, the core business of many companies will require capabilities that they do not presently have. In addition, many businesses have realized that being “prime” from Basic Research to Marketing and Distribution is no longer economically practical. Some companies have even decided to dedicate their facilities solely to research. Others have decided to exploit a particular niche. Companies that continue to be prime are becoming fewer and fewer.

Outsourcing Defined

Previously, the paradigm of the pharmaceutical industry required companies to be vertically integrated, i.e., the company itself performed all the operations required of its business. Now investors are demanding continued

high financial performance. In an effort to reduce time-to-market, companies are more and more focusing their attention on their core competencies. As a result “outsourcing” has become a significant way of doing business. Outsourcing is the system of using a nonrelated company to produce materials or perform tasks for another on a contractual basis.

The New Paradigm of Doing Business

Outsourcing has resulted in the development of a new paradigm for doing business. This paradigm offers companies new opportunities for improving their bottom lines through the conversion of fixed costs to variable costs. They accomplish this by reducing or eliminating in-house production capabilities and replacing them with contract manufacturers. As a result, contract manufacturers that perform custom synthesis and produce intermediates, active pharmaceutical ingredients, and dosage forms are becoming increasingly important to the conduct of today’s business.

Change from Vertical to Horizontal Integration

As a result, companies are becoming more horizontally rather than vertically integrated. Rather than expanding or upgrading their own facilities to accommodate new technology, companies that outsource have realized that a new product can be brought to market with greater speed than if they had performed all of the tasks internally. As an additional benefit they have positively affected their profitability through reduction of their need for investment in bricks and mortar.

The Virtual Company

The outsourcing industry has also spawned a new type of company, the “virtual” company. The concept of a virtual company is based entirely on the use of contract manufacturers, laboratories, and other providers of services to take a product from concept to market. Such a company is truly horizontally integrated. Furthermore, it has the advantage of requiring only a few managers to

coordinate the activities of various contractors as it conducts its business.

Supply Chain Management

A new breed of managers is developing within existing organizations that have decided to outsource. Their job is to effectively link the services of several contract manufacturers to forge a supply chain that performs all of the steps between R&D and Sales/Marketing. The resulting supply chain requires little or no investment of capital and achieves the same results previously obtained through vertical integration.

CONTRACT MANUFACTURING DEFINED

Origin

Contract manufacturing first began when one company asked another to produce for them a product that they chose not to produce themselves. Contract manufacturers can be found throughout all industries. Instead of producing their own “brand name” products, they provide production and other services to all comers. And, because their facilities are designed to be flexible, the same facility can make products for many customers.

Types

Contract manufacturers can be classified into two types, i.e., those who “supply” and those who “toll”. Regardless of type, all contract manufacturers have the common denominator of providing one or more services for a fee.

Suppliers and Tollers

A contract manufacturer is one who supplies manufacturers materials for inventory. This supplier sells products from its inventory to one or more companies for their use or disposition. This type of contractor is sometimes known as an “original equipment manufacturer.” On the other hand, a “toll manufacturer,” or “toller,” is a manufacturer who contracts to

- receive a raw material from another company.
- convert that material into another form, and
- return the converted material to the contracting company for its use or further disposition.

The basic difference between these two types of contractors is that one manufactures for its own inventory,

while the other manufactures according to a custom order. To cloud this picture, either type of contractor may provide both functions.

Types of Tollers

Sellers of Excess Capacity

Companies that engage in toll manufacturing can be further classified into three fundamental types. First are companies that have more capacity than they need to produce an established product line. Rather than let that excess capacity stand idle, the companies use it for toll manufacturing to reduce the burden of salaries and overhead attached to the excess capacity while concurrently producing additional income. Pricing for services from these manufacturers are usually quite favorable, since their product line is absorbing their fixed costs. Only variable costs associated with the excess capacity are incurred.

Providers of Niche Services

The second type of toller is comprised of companies that, rather than marketing an established product line of their own, have concentrated on a niche service. Their facilities, equipment, and personnel have been designed only to accommodate specialized operations such as lyophilization and production of parenterals. Their success is dependent upon being able to lease their facilities and technology for the production of products for one or more customers. Their focus is on making products without having to worry about heavy R&D and marketing. A positive advantage for the use of this type of company is that they are less likely to compete with their customers.

Academic Institutions

The last type is comprised of academic institutions that provide assistance with custom synthesis, dosage form development, pilot-scale production, and production of clinical supplies in order to obtain additional income to support their academic programs. They have no interest in, nor are they equipped for, commercial production. Companies seeking outsourcing should be aware that students and faculty could form part or all of the staff of their operations; this may be a disadvantage when considering using this type of contract manufacturer.

HOW TO DETERMINE WHEN A CONTRACT MANUFACTURER IS NEEDED

The need for the services of a contract manufacturer can occur at any time during the developmental phases and/or

commercial manufacture of a product's life cycle. Such situations occur when

- Specialized manufacturing capabilities are required that are not available in-house.
- Assistance is needed with product and/or process development.
- The need to establish the market potential of a new product is required before investing in specialized capabilities.
- Difficulty is encountered in breaking into the manufacturing schedule in a timely manner to produce small research, clinical, or commercial batches.
- Production requirements cannot be accommodated when sales exceed capacity.
- Capacity is needed for the production of new, growing products, yet a place for the manufacture of products that are at the end of their life cycle still needs to be provided.

Advantages of Working with a Contractor

Developing companies especially can gain many advantages from working with a contract manufacturer. First, they gain access to production facilities without having to invest any of their own limited capital. Second, they instantly add a breadth and depth of expertise in pharmaceutical manufacturing to their operations that would have taken years to develop. Third, they gain the use of the contractor's operating personnel. Depending on the special niche of the contractor, the contractor's personnel will be experienced in one or more pharmaceutical specialties. These include custom synthesis; production of intermediates and active pharmaceutical ingredients; formulation development; lyophilization cycle development; aseptic processing; sterilization process development; regulatory assistance; sourcing of excipients; analytical methods development; validation of processes and methods; package development; and storage and testing of stability samples.

Disadvantages of Working with a Contractor

But, a company seeking to outsource must also be familiar with negative aspects associated with working with a contract manufacturer. A major disadvantage to developing companies is that the client is entrusting to another entity its valuable intellectual property that before was a closely guarded secret. Clients also must consider that they are but one of the contractor's many clients. As a result, due to competition for the contractor's resources, projects may not move forward as fast as a client may desire.

Finally, clients incur the costs of at least a portion of the salaries and overhead of the contract manufacturer.

HOW TO SELECT A CONTRACT MANUFACTURER

Involvement of Purchasing Department

After evaluating the pros and cons of using a contract manufacturer, how is one selected? The selection process is very important to avoid disappointment later. Some companies require that all searches be conducted under the auspices of their purchasing department. In such cases, the purchasing department should be requested to initiate a search. Regardless, whether such a policy exists or not, close collaboration with the purchasing department is advisable during all of the following steps in the selection process.

List of Requirements

Begin any search by developing a list of requirements that a potential contractor must satisfy in order to be considered. Provide this list of requirements to the purchasing department. The most fortunate situation that can occur in the selection process is finding that the Purchasing Department is already employing a contract manufacturer with the capabilities to perform the required task.

Referrals from Colleagues

After developing the list of requirements, begin your search by seeking referrals from colleagues in the industry. There is nothing better than obtaining a recommendation about a particular contract manufacturer from someone that has successfully used its services. You can question your colleagues about any idiosyncrasies they experienced with the operations of recommended contractors. These queries can also explore the degree of satisfaction experienced regarding the manufacturer's compliance with mutually agreed upon requirements. Finally, an opinion can be obtained regarding the qualifications of the contractor's personnel assigned to projects. Be sure to solicit information as to how those personnel interrelate with customers.

Use of Information Sources

Contract manufacturers can be identified by their specialty through the use of commercial information sources (1, 2) and by review of advertisements in trade journals (3).

Commercial information sources solicit profiles of contract manufacturers for their databases. Contract manufacturers usually advertise their services in journals having the widest circulation within an industry. Many of these journals print annual buyer's, guides in which their advertisers are grouped according to the services they provide. The Internet is also a useful tool to identify contract manufacturers. Several sites are listed in the Reference section (4).

Trade Shows

Trade shows provide excellent opportunities to seek out and meet contract manufacturers. Contract manufacturers operate exhibits where literature on their companies can be obtained for later evaluation. Talking with the exhibitor's personnel provides an opportunity to conduct preliminary discussions with contractors about the services they offer. You can determine on the spot whether or not a particular contractor will be suitable for your project.

Trade shows often feature speakers on topics concerned with contract manufacturing. Attend such programs to familiarize yourself with the intricacies of contract manufacturing. You will find that your notes and handouts from the speakers will contain the latest information on this subject.

The Interview Process

Once you have identified a list of potential contractors, begin communicating with them by telephone. Table 1 provides suggested questions that can be used during interviews. Begin the interview by explaining that you are conducting a search for a contractor, and that they have been preliminarily identified as having the capability to meet the requirements of your upcoming project. Then, request them to send the most current information about themselves and their capabilities. To assure due diligence in your search, include at the conclusion of each interview a request for them to give you the names of any other contractors that they feel might be able to meet your requirements. Additional contractors identified in this manner should be contacted and interviewed, especially those who were suggested by more than one other company.

“Very Good” vs. “Excellent” Contractors

A few contractors have reached the point of being identified as being “very good” at what they do. Once this status has been reached, their business usually increases and downtime decreases. However, the nature of a

contractor's business poses problems with advancing further in status, i.e., involvement with many customers, frequent changeovers of equipment, and production of many batches of a multiplicity of products. Advancing to the status of being “excellent” presents a unique challenge, because an “excellent” contractor is one that manages schedules so that time is allocated for concentration upon the nuances of any one process, smoothing out the rough edges, and achieving perfection. Your search should strive to identify the “excellent” contractors.

Confidentiality Agreement

After a list of “finalists” has been assembled, execute a confidentiality agreement with them. Only after such an agreement has been executed between your company and the contractor should the project be discussed in greater detail. Once provided with detailed knowledge of the project, the potential client and contractor can then determine whether a match exists between the requirements of the project and the contractor's capabilities.

Agenda for First Meeting with Selected Contractor

The final step in the process is to select one of the finalists and to request a meeting to discuss the project in greater detail. A sample agenda for such a meeting is contained in Table 2. Any information that the contractor will require in preparation for this meeting should be provided before the meeting. Additional meetings will probably be required before work can be initiated.

Statement of Work

Following the initial meeting, the contractor will evaluate the information provided. A prudent contractor will then provide an itemized proposal describing the work that would be necessary to be completed before the client's product can be introduced into the production schedule. The contractor and client should meet as often as necessary until both sides can agree upon a statement of work to be provided by the contractor that will accomplish the goals of the client.

DEVELOPING A CONTRACT

Basic Elements and Governing Procedures

The parties involved in developing a manufacturing contract generally understand that the contract will include

Table 1 Questions to be used when defining requirements sought in a contract manufacturer

-
1. What services are to be provided by the contractor?
 - Formulation development?
 - Process development?
 - Experience with and the capability to produce a particular dosage form?
 - What manufacturing capacity is required?
 - Testing laboratories?
 - Document preparation and editing?
 - Capability for distribution?
 - Regulatory assistance?
 - Aseptic processing?
 - Lyophilization?
 - Terminal sterilization?
 - Validation?
 - Quality assurance?
 - Inspection?
 - Labeling?
 - Packaging?
 - Stability studies?
 2. Does the contractor have the capability, facilities, and capacity to manufacture both clinical supplies and commercial products? (Otherwise, approval of another manufacturer will have to be obtained before the product can go to market.)
 3. How flexible is the contractor with regard to formulation, manufacturing, quality assurance, labeling, and packaging?
 4. Are “turn key” services such as development, control, and distribution available?
 5. Do the manufacturing facilities conform to current good manufacturing practices ?
 6. Does the contractor allow inspection of the facilities?
 7. Does the contractor allow the client to audit the manufacture of each batch of a client’s product?
 8. Does the contractor have an established process to monitor and improve quality?
 9. Does the contractor supply a list of minimum requirements that are expected of its clients?
 10. Does the contractor provide a product information questionnaire to be completed before a project begins?
 11. Does the contractor insist that both parties review and approve master batch records before their use?
 12. Does the contractor require that a representative of the client be on site during the manufacture of the client’s product, or require that the client have a knowledgeable person on-call at all times to provide immediate response to questions that arise during manufacturing?
 13. Does the contractor offer analytical support? Is this support adequate?
 14. Does the contractor provide a protocol of analysis or a certificate of compliance for each lot of product at the time of shipment as well as copies of completed batch records?
 15. What is the contractor’s policy regarding loss of customer supplied materials when the loss is due to its own negligence, or its failure to perform according to mutually agreed upon standards or obligations?
 16. Does the contractor encourage innovation rather than maintaining the status quo? What examples of new practices, procedures, etc., can it describe?
 17. Will the contractor provide a letter of access to its facilities master file?
 18. What is the contractor’s relationship with regulatory authorities?
 19. What is the contractor’s posture during regulatory inspections?
 20. Does the contractor offer confidentiality agreements?
 21. Is the contractor willing and able to enter into long-term supply agreements?
 22. Is the contractor selling the excess capacity of its parent company?
 23. Can the contractor comply with the established timetable?
 24. How responsive is the contractor’s customer service department?

(Continued)

Table 1 Questions to be used when defining requirements sought in a contract manufacturer (*Continued*)

25.	Is the contractor an equal opportunity employer?
26.	Has the contractor had experience in supplying materials for government contracts?
27.	Does the contractor produce any products for its clients that are shipped or sold internationally?

The answers to the above interrogatories will help in the selection of a contractor who can provide the services desired.

a definition of what is to be manufactured, what the cost will be, and when the product is to be delivered. Beyond these basic elements, the client and contractor are well advised to jointly develop written procedures that detail how the client and contract manufacturer are to interact and thus ensure that appropriate actions are taking place. These procedures can either be included in the contract or take the form of mutually agreed upon internal operating procedures. Further, the contract must specify how the client reviews and approves all processes specific to the manufacture of the drug product. The client is then obligated to generate operating procedures that describe the mechanisms that will be used for review and control and provide these to the contractor. The client has the added responsibility of having written procedures in place to ensure that the contractor complies with all previously agreed upon requirements. A copy of these procedures must be given to the contractor, providing awareness of the client’s standards.

Conformance to Precepts

Basically, the contract must ensure that the client can expect the highest level of professionalism from the contractor, that a high quality product is being provided, and that the product is being produced according to cGMP at a fair price. The contract manufacturer is only selling its

services and its quality process; therefore, conformance to these precepts is the reason for being selected as your contract manufacturer. A list of the clauses usually incorporated into a standard supply agreement for Goods and Services is included in the Appendix to this article.

COSTS

Contractor’s Fee for Service

What does it cost to use a contract manufacturer? Although contract manufacturing may seem expensive, the use of a contractor does enable a client to dedicate its resources to more productive activities than maintaining a variety of production facilities. In return for providing the client with other opportunities for its resources, the contractor expects to receive a fee commensurate with its investment in capital equipment, wages and benefits for personnel, materials to be supplied, overhead, batch size, and a reasonable profit.

Fee Based on Expected Business

However, there is no set fee that can be universally applied to a contractor’s services. Remember, in most cases a contractor, unlike a regular manufacturer, has no product line. Whether or not it has business is entirely at the whim of its customers. The quantity of business available to the contractor is governed by how well a client’s products are doing in the clinic or in the marketplace. The contractor can make a guess as to how much business it will do each year, but the uncertainty is great. Therefore, a contractor usually bases his fees on the costs associated with each batch to be produced, rather than the projected total output of its production each year. To gain discounts, clients must be prepared to offer reasonable guaranteed production requirements.

Fee Related to Size of Batch

The size of the batch to be produced is a very important aspect to be considered when evaluating costs. Regardless

Table 2 A Sample agenda for an initial meeting with a contractor

I.	Welcome and introductions
II.	Tour of contractor’s facilities
III.	Detailed description of project
IV.	Question and answer period with contractor’s staff
V.	Client’s requirements and expectations
VI.	Contractor’s requirements and expectations
VII.	Agreement to proceed with or discontinue further discussions
VIII.	Development of timetable and identification of milestones, if appropriate
IX.	Business issues
X.	Adjournment

of the batch size, certain minimum charges for setup must be recovered. Obviously, the larger the batch, the more units are available over which to spread these charges, and unit costs diminish accordingly. Conversely, for the same reason, the smaller the batch size, the greater is the unit cost.

In-House vs. Using Contractor

When evaluating the appropriateness of a contractor's fee, a client must know its costs for doing the project in-house. Then it must consider whether the expense of adding the contractor's overhead and profit to the cost of goods is more than the in-house costs. Finally, a client must justify the cost of any delays in reaching the market as being more economical than the contractor's fee.

WORKING WITH A CONTRACT MANUFACTURER

Timetable

As a first step before beginning work with a contractor, both the client and the contractor must agree on a mutually acceptable timetable for completion of the project. This timetable should include decision points where reports and/or review meetings are required. Then, a principal contact at the contractor's site and one at the client's site need to be identified so as to assure that a uniform flow of communications can occur between the companies. The person assigned to be the client's principal contact will be responsible for monitoring the contractor's compliance with the timetable. Each party to the timetable must be agreeable to making revisions in the timetable should legitimate situations require.

Contractor's Input

From a contract manufacturer's viewpoint, it often appears that the client considers the contract manufacturer only as a provider of services with no expectation of providing input into the process of bringing the drug product to market. For a successful relationship this concept must be reexamined. Actually, the mutual best interest of both is best served by each acting as a partner of the other, freely sharing information that assures mutual success.

Communications

A contract manufacturer brings more than just a manufacturing facility to the partnership. The contractor is not just another pair of hands. The contract manufacturer

and its client must be able to communicate with one another as equals and with respect for each other's contributions.

Key to Successful Relationship

The client is well advised to listen to advice provided by the contractor and respect any concerns. After all, the contractor, while being careful to protect confidences, can call upon years of experience gained from working with other clients to provide insight into problems, regulatory concerns, etc., which a client may be experiencing for the first time.

Master Production Batch Information

Open channels of communication assure continued cooperation and successful outcomes. Encourage your contractor to have timely discussions with you concerning any problems or concerns experienced with the manufacturing process or with regulatory matters. Likewise, share with the contractor your data and any other information concerning analytical results, results from focus groups, any product complaints, and conclusions from clinical studies. By working together, adjustments can be made to manufacturing processes that will assure that a quality product is being produced. When using a contractor, always remember the following: A successful relationship can only occur when both parties work together to win.

Before the contractor can begin manufacturing activities, master production batch information must be developed. This master information package normally includes specifications for raw materials, components, analytical methods, finished product specifications, packaging instructions, and any other information pertinent to the product. The client is obligated to review and approve the master production information before allowing the contractor to proceed. This approval indicates that the client is satisfied that the contractor will be producing a product that will meet specifications.

Client's Representatives

A representative of the client should be present in the contractor's facilities during the production of the product. If the client elects not to have a representative present, the client should provide the contractor with the name of an authorized person who can be contacted at any time, day and night, to resolve any issues with the production which might occur during processing.

DISADVANTAGES TO WORKING WITH A CONTRACT MANUFACTURER

Use of Commercial-Sized Equipment for Small Batches

Some contract manufacturers are equipped to produce only commercial-sized batches of product. Because of this, the same equipment is also used for the production of smaller research and development scale batches. This is a disadvantage when batches are small, particularly where expensive materials in short supply are to be produced, because a considerable amount of bulk formulated product may not be able to be converted into finished product. Some materials will always be hung up in processing equipment, regardless of size. The materials that are hung up are not available for further processing. The quantity of hung up materials is related to the size of the processing equipment being used, and is relatively constant for that equipment. Unfortunately, when losses are being calculated, the smaller the batch size that is being produced in equipment designed for larger batches, the greater is the percentage loss of product.

Buy-Back Arrangements for Specialized Equipment

Another consideration when working with a contract manufacturer is whether specialized or dedicated processing equipment will be required. If dedicated or special equipment is required, the contract manufacturer may feel justified to ask the client to supply this equipment with or without a later buy-back arrangement. In a buy-back arrangement, the contractor asks the client to initially assume the full risk of loss of the capital investment in the equipment should the product not become a viable commercial entity. However, as the product successfully completes each stage of the steps to marketability, the client's capital investment would be gradually returned. The end result is that the contract manufacturer has all of the equipment needed for the manufacture of commercial-sized batches of the client's product at the time it is finally commercialized.

Liability

Then there is the question of liability. If an adverse event occurs in a client's own operation, the client absorbs the financial loss as a cost of doing business. How much liability for loss of product or materials during the manufacturing process should a client then expect a contractor to assume? A contractor will argue that, even

though a separate entity, during the time the contractor is manufacturing the client's product, the contractor's facility is an extension of the client's shop floor. Therefore, should an adverse event occur during production of the client's product, the contractor will expect the client to absorb any financial losses as if the losses had occurred in the client's own operation. Whether or not a client agrees with this argument, the amount of liability to be assumed by each party must be agreed upon during contract negotiations.

REGULATORY ISSUES

Status of a Contract Manufacturer

From a regulatory standpoint, Regulatory agencies do not differentiate a contract manufacturer from any other manufacturer under their jurisdiction. A contract manufacturer must comply with the same federal, state, and local regulations as would any other manufacturer. In the United States, prior approval must be obtained from the U.S. Food and Drug Administration (FDA) before a contractor can be used as either the primary or alternative manufacturer for a registered pharmaceutical product. Even though a contractor is being used, the contracting company still retains primary responsibility for assuring that the contract manufacturer complies with all of the commitments that were included in the product's registration and with all aspects of current good manufacturing practices (cGMPs).

Inspections

Regulatory agencies inspect contract manufacturers on a regular basis to ensure that these manufacturers are in compliance with cGMPs. In the United States, all reports of inspections by regulatory agencies are available to anyone through freedom of information. Any company contemplating the use of a contract manufacturer should review the regulatory history of all companies under consideration as part of the selection process. Likewise, client companies, as part of their contractual relationship, should have a contractual requirement that the contractor notify them should there be an inspection involving their product, or should there be any adverse findings from any inspection which would affect the continued supply of their product by that manufacturer.

Pre-approval Inspections

Regulatory Agencies also conduct pre-approval inspections as part of the process of review of registration

documentation. Prior to a pre-approval inspection, the client and the contractor should review all documentation that will be reviewed during the inspection. Clients should also assure themselves that the contractor is, indeed, in compliance with cGMPs.

International Considerations

If the client is marketing the product internationally, the client should provide the contract manufacturer with copies of the regulatory requirements that might affect the client's product in all countries in which the product is being marketed. A contractor cannot be held responsible for compliance with regulations with which the contractor is unfamiliar.

Master Files

A contract manufacturer should be willing to provide its clients with authorization to permit the regulatory authorities to reference the facilities master file that the contractor maintains in the archives of the regulatory authority. Or, if a facilities master file is not an appropriate vehicle for a particular type of registration, the contractor should provide sufficient information about its processes and procedures to enable a client to satisfy the requirements for manufacturing information in its application for the registration of its product. Although it is convenient to be able to use contractor-supplied information in the preparation of an application for registration of a product, the client must always completely review that information to ensure that, from a regulatory viewpoint, control of the product has not been inadvertently transferred to the contractor.

Special Requirements of the Biotechnology Industry

The biotechnology industry is one of the biggest users of contract manufacturing. Contract manufacturers are used to produce bulk active pharmaceutical ingredients by cell culture and/or fermentation. Then, other contractors may be used to convert the bulk active pharmaceutical ingredients into finished dosage forms. The regulations governing the use of multiple companies for these production activities are continually being refined by the FDA.

Dedicated Equipment

In some situations, the regulatory agency will require that dedicated equipment be used in manufacturing activities.

Should this requirement be imposed, many contractors require the client to purchase this equipment. This becomes another initial expense that must be added to the total cost of using a contract manufacturer.

Common Regulatory Theme

Regardless if the product is a "traditional" pharmaceutical or is a product of biotechnology, one common regulatory theme is always present. That is, the company employing a contractor retains responsibility for the product being manufactured and for ensuring that the contractor remains in compliance with the terms of the license and with cGMPs.

History of Contractor's Relations with Regulatory Agencies

Finally, before working with any contract manufacturer, potential clients must be assured that the contractor has a satisfactory history of cooperation with regulatory authorities. This type of information can be obtained from a review of several establishment investigation reports (available through freedom of information in the United States) involving that contractor.

Cooperation with Regulatory Agencies

Should a contractor regularly take an adversarial stance when dealing with regulatory authorities, there is more at stake than just defending that contractor's position. That contractor is potentially placing the continued production of the client's products in jeopardy especially if the dispute involves processes or procedures related to those products. For that reason alone, a potential client should not select that particular contractor for their project.

MONITORING THE CONTRACTOR

Quality Assurance Initial Audits

Prior to establishing any relationship with a contractor, the quality assurance staff of a potential client must audit the contractor's facilities. During this audit, the contractor should be willing to provide a comprehensive tour of the facilities. Following the tour, the auditors need to review all of the contractor's standard operating procedures that may apply to their company's proposed project. In addition, the auditors need to review the quality improvement process being used by the contractor and the results of that process.

Any concerns that have been identified during the audit should be discussed with the contractor at that time.

Actions to be Taken Upon Receipt of Audit Report

Upon completion of the audit, the auditors should provide the contractor with a report citing any problems which they have identified with the contractor's operation that might interfere with the contractor being able to successfully work on their company's project. The contractor should respond in writing to the audit report and address how they will handle each observation. Based on the results of their audit and the responses of the contractor, the audit team can then make a recommendation to the responsible personnel within their organization as to the acceptability of the contractor.

Representatives in Contractor's Facilities

Once manufacturing begins, most clients elect to have a representative present. The contractor should be willing to permit that representative to observe the processes in their entirety. Before any visit, however, the client must instruct its representative to comply with all of the standard operating procedures of the contractor while on the contractor's premises. Further, the extent a representative can be involved in the production process needs to be negotiated between the client and the contractor prior to sending a representative to the contractor's site.

Periodic Audits by Quality Assurance Staff of Client

After a client has established a working relationship with a contractor, the quality assurance staff of the client should be allowed to audit the contractor's facilities and operations on a regular basis, but usually not more frequently than once a year. Again, any observations concerning the contractor's operations arising from these audits require a written response from the contractor. If the responses are not satisfactory to the audit team, then the upper management of both companies must be called upon to immediately resolve the differences. Should this be necessary, the management of both parties must ensure that any proposed resolution remains within the terms of the contractual agreement.

Quarterly Meetings

Finally, it is good practice to hold quarterly meetings between the decision makers of both the client and the

contractor. These meetings ensure that both companies are still aligned to achieve mutual goals. These meetings allow the contractor to participate in the client's decision-making process as performance for the previous quarter is reviewed and plans for the following quarter are presented. Topics for discussion should include the contractor's conformance to the client's expectations, regulatory issues concerning the client's product(s), the contractor's recent regulatory experiences, sales forecasts, and any other items appropriate to the occasion.

CONCLUSION

Through skillful management and control, contract manufacturers can be used to take a product from the "test tube miracle" stage through the stage of being a "production masterpiece." Virtual companies have learned to do these tasks well, and have successfully linked multiple contract manufacturers together to form the horizontally integrated company of the future.

Points to Follow When Working with a Contract Manufacturer

Working with a contract manufacturer is quite easy provided that the following points are followed:

- Finalize the decision to use a contract manufacturer early in the timetable of the project.
- Involve the contractor in decisions, e.g., formulation, components, regulatory, etc.
- Do not file specific requirements in regulatory documents without first assuring that the contractor agrees.
- The client and the contractor have mutually agreed upon the requirements of the project *before* beginning the project.
- The client has designated someone from their staff with whom the contractor can make contact *at any hour* to resolve difficulties.
- Remember the Golden Rule—Do unto the contractor, as you would have the contractor do unto you.

APPENDIX: ELEMENTS OF A CONTRACT

The following list of clauses are usually incorporated into supply agreements for goods and services. The author developed this list using his experience of over 15 years reviewing various supply agreements. Potential clients can

use this list to ensure the completeness of their agreements. Depending on the situation, additional elements may be appropriate while others may not. The aim in any contractual relationship is to achieve a win/win situation for both parties. This aim can only be accomplished through good faith and due process during the process of negotiation.

The Checklist

1. Payment terms
2. Forecasting process
3. Quality standards—Test methods
4. Pricing
5. Liabilities—Both parties
6. Specification(s) of item(s) or service(s) to be supplied
7. Transportation and limit of responsibility from the point of ownership
8. Right of audit and inspection
9. Term and renewals
10. Warranties
11. Prior approval over product or process changes
12. Confidentiality
13. Alternate supply in case of failure to supply
14. Limitations to subcontract
15. Acceptance
16. Third-party laboratory for resolution of testing disputes
17. Insurance
18. Supply over-forecast amounts
19. Firm purchase orders
20. Change of control
21. Assignment rights
22. Penalties for failure to perform
23. Stability responsibilities
24. Recall responsibilities
25. Survival
26. Exclusive/nonexclusive total requirements
27. Force majeure
28. Regulatory/state/federal law/cGMP compliance
29. Separability
30. Notices
31. Arbitration and dispute resolution
32. Governing law
33. Amendments
34. Notification of price changes
35. Accuracy
36. Complete agreement
37. Supplier certification
38. Adverse reaction reporting—Product complaints
39. Independent contractor
40. Allocation of capacity in event of shortage
41. Royalties
42. Supply and storage conditions
43. Capital costs and tooling
44. Yields—Over/under shipments quantity
45. Trademark rights
46. Patent rights
47. Rejections
48. Safety stock requirements and liability for obsolescence
49. Ownership of technology or proprietary information
50. Taxes
51. Human resources—Government, EEOC requirements
52. Definitions
53. Index
54. Customs and duties
55. Noncompete

REFERENCES

1. PharmSourceTM Information Services, P.O. Box 8163, Springfield, VA, 22151.
2. TechnomarkTM Consulting Services, Ltd., King House, 5-11 Westbourne Grove, London, W24UA, United Kingdom.
3. Magazines: Pharmaceutical Technology; Contract Pharma; Pharmaceutical Processing; Formulation & Quality; Drug Discovery & Development; Biopharm.
4. Internet Sources: Member/Subscriber Only Access; www.pharmsource.com PDA, AAPS, BIO, etc.; Free Access: Individual Company Sites; www.pharmaportal.com (All Advanstar Publications) (www.pharmaceuticalonline.com) (www.pharmpro.com) (www.contractpharma.com) (www.pharmaquality.com).

COMPUTERS IN PHARMACEUTICAL TECHNOLOGY

Onkaram Basavapathruni

Pharmacia Corporation, Peapack, New Jersey, U.S.A.

INTRODUCTION

The computer has become a very common tool in all areas of science and technology, and there seems to be no end in sight for future applications. With the proliferation of the Internet and the developments in computer technology and manufacturing, the ratio of price-to-performance of computers continues to decrease. This has resulted in the development of a number of computer applications. The field of pharmaceutical technology has also benefited from the use of computers and will continue to benefit as the professionals in the field gain more familiarity with computers. This chapter will discuss some examples of existing computer applications, the fundamentals of computer technology, and issues to be addressed when applying computers in pharmaceutical technology and assessing their future applications.

COMPUTER APPLICATIONS

Computers have been successfully utilized in pharmaceutical technology to improve productivity, collaborate with other professionals, and to provide solutions for time-consuming manual tasks. For purposes of discussion, the various computer applications are classified as follows:

- Data and information management systems
- Interactive voice response systems
- Group collaboration tools
- Document management and publishing systems
- Internet-based applications and tools
- Problem-solving applications
- Communication aids
- Laboratory automation
- Process control
- Computer-based training

Some computer applications are so complex that it is difficult to classify them into any one of the categories. Nevertheless, any complex computer application can be broken down into smaller parts, and each part may then be described under one of the classifications.

Data and Information Management Systems

The first computers were primarily used for computational purposes. As hardware prices dropped and computer storage technology developed, it became cost-effective to use computers for storing vast amounts of data and information in a variety of formats (e.g., text, numbers, audio, image, and video). Along with the advances in computer hardware, software development also advanced rapidly and resulted in the development of database management packages, such as Oracle, Sybase, and Microsoft Access. By using these packages, a number of computer-based systems can be developed in-house in order to organize data and information and then to query the data in a number of ways. An alternative would be to acquire commercially available systems that use these popular database packages. In either case, the challenge with these systems is to determine what type of information has to be stored in the database and how it can be retrieved in a number of ways. These systems, when designed properly and implemented with active user participation, minimize the paperwork and improve the productivity of personnel involved. Data and information management systems are usually built by the in-house data processing department or acquired from a vendor to meet the needs of users in a given organization. More often than not, the same data management system can be implemented differently in different organizations. A number of systems will be described in general terms in order to give the reader an idea of the available data management programs.

Material inventory system

This system maintains a running inventory of raw materials used in pharmaceutical manufacturing. It will be updated at regular intervals when new materials arrive, as well as when materials are drawn out. This system is useful in tracking existing inventory, lot numbers, and quantities of raw materials needed for manufacturing a product, as well as other pertinent information. Some systems also accommodate the need to reserve a certain lot of raw material for use in manufacturing a particular batch of the product to be used for a stability or clinical study. A sample output from a material inventory system is shown in Fig. 1.

Inventory Report for GMP items

Description	Item #	Unit of measure	Lot #	Quantity
MICROCRYSTALLINE CELLULOSE NF/BP/DAB (AVICEL PH101/EMCOCEL)	Z00106	KGS	SP2359	50.00
MICROCRYSTALLINE CELLULOSE NF/BP/DAB (AVICEL PH102/EMCOCEL 90M)	Z00102	KGS	L00616	10.57
MICROCRYSTALLINE CELLULOSE NF/BP/DAB (AVICEL PH103)	Z00103	KGS	SP1029	9.70
MILLER FROCKS 5 SNAPS L CODE 1212	Z02204	CS	SP2059	1.00
MILLER FROCKS 5 SNAPS M CODE 1212	Z02205	CS	SP2060	1.00
MILLER FROCKS 5 SNAPS S CODE 1212	Z02206	CS	L00604	90.00

Fig. 1 Sample output from a material inventory system.

Formulation information system

This system can be used to store the information on raw materials that constitute a batch of a pharmaceutical bulk product. Typically, the lot numbers of the ingredient, the name of the ingredient, amount per dosage form, percent composition, weight of the batch, and the actual amount of each ingredient are stored in these systems. Other information, such as manufacturing summary, shipping history, and relevant storage information, is also stored. The system's greatest benefit is its different retrieval methods. For example, during product recalls, audits, or tracking an excipient problem, one might be interested in determining all the batches of products made using a particular lot of raw material. This system can also be used to archive the formulation information generated during formulation screening studies. Consequently, data collected in this fashion can be utilized for subsequent statistical analysis. This system can also act as a repository of information that might be helpful in serving as a knowledge base to develop new formulations.

Clinical supplies inventory system

In pharmaceutical research and development (R&D), a number of clinical trials are conducted to determine the therapeutic effectiveness of potential new drugs and novel dosage forms of well-established drugs. In order to provide the necessary clinical supplies in a timely manner and to plan for the manufacture of needed supplies, clinical supplies inventory systems are used.

Clinical supplies labeling

The clinical pharmacy is often called upon to supply complex labeling requirements for investigational drugs used in clinical studies. Depending on the number of patients and investigators, several hundreds or even thousands of labels with randomized patient and

investigator numbers are prepared. Well-designed computer programs eliminate the manual labor involved in hand-numbering each label. The computer also sorts the labels by investigators and treatment groups, and prints the labels accordingly. Examples of labels generated by a computer program are shown in Fig. 2.

Stability information systems

The pharmaceutical manufacturer conducts stability studies to develop stable dosage forms in a variety of packages and, in the process, generates vast amounts of data and information. Stability information systems are designed to organize these data and help retrieve the needed information to establish an expiration date for a product. The stability information is also submitted to the regulatory agencies, such as Food and Drug Administration (FDA), from time to time in support of investigational new drug applications (INDAs), new drug applications (NDAs), new dosage forms, abbreviated new drug applications (ANDAs), biological license applications (BLAs), and product license applications. For additional information, the reader is advised to check the following website: <http://www.fda.gov/cber/gdlns/stabdft.pdf>.

Prior to the start of a stability study, the pharmacist designs protocols, such as the one shown in Fig. 3. When it is time to initiate a study, the computer is used to generate the stability calendar (Fig. 4). Using this calendar and the protocol, the list of samples for chemical, physical observations, and microbiology analysis of a given time period is generated.

The chemical analysis results and the physical observations are stored in a database for further retrieval. The cumulative chemical data are retrieved in the form of tables for inclusion in a regulatory report (Fig. 5) or for review (Fig. 6). The data can also be presented in a graphic form (Fig. 7), which makes it easy to review large amounts

KEEP OUT OF REACH OF CHILDREN	SEARLE	Bottle A
	G. D. Searle & Co. Skokie, IL 60077	32 Tablets
	Test Compound	
	XX -XX -XX -XX -X	
	BASELINE	SUBJ # 0000
	Take TWO tablets from Bottle A	
	in the morning with breakfast.	
	Lot RCT XXXX	Expires MAY 2001
	Store between 59° - 77°F (15° - 25°C).	
	CAUTION : New Drug - Limited by Federal (U.S.A.) law to investigational use.	

This content must be returned to depending physician	SEARLE G.D. Searle & Co. Skokie, IL 60077	SEARLE G.D. Searle & Co. Skokie, IL 60077	MODEL PANEL DISCLOSURE
	TEST LABEL	TEST LABEL	
	SUBJ:	SUBJ:	
	STUDY II-II-II-II-I LOT: RCT XXXX	STUDY II-II-II-II-I LOT: RCT XXXX	
	TAKE 1 (ONE) TO 2 (TWO) TABLETS 4 TO 6	TAKE 1 (ONE) TO 2 (TWO) TABLETS 4 TO 6	
	TIMES A DAY AS NEEDED FOR PAIN. DO NOT	TIMES A DAY AS NEEDED FOR PAIN. DO NOT	
	TAKE MORE THEN A TOTAL OF 12	TAKE MORE THEN A TOTAL OF 12	
	TABLETS IN A DAY.	TABLETS IN A DAY.	
	KEEP THIS AND ALL MEDICATION OUT OF	KEEP THIS ALL MEDICATION OUT OF	
	REACH OF CHILDREN.	REACH OF CHILDREN.	
STORE BELOW 86°F AND IN A DRY PLACE.	STORE BELOW 86°F AND IN A DRY PLACE.	II-II-II-II-I SUBJ: BULK LOT RCT XXXXX PLACEBO TABLETS	
EXPIRES MAY 2001	EXPIRES MAY 2001		
KEEP OUT OF REACH OF CHILDREN			
Caution: New Drug - Limited by Federal (U.S.A.) Law to Investigational use.			
DETAC - HCTE			

Fig. 2 Sample labels generated by a computer program.

of data in a short time. Other programs that are used to carry out statistical analysis can access the required data from the stability database. The physical observation data are also presented in the form of tables for reporting and review purposes.

Analytical information systems

The analytical laboratory generates vast amounts of data and information while developing analytical methods, supporting product stability studies, and aiding formulation development. The analytical laboratory needs sample management in addition to the management of data it generates. The analytical systems help by providing lists of samples to be analyzed. These are sorted by project, laboratory location, etc. The system also generates reports of analysis (Fig. 8), cumulative analytical data, and other types of reports needed by an organization.

Quality assurance information system

Several systems are used by the quality assurance unit to help carry out quality assurance functions. One such

system helps the quality assurance function track information on chemical raw materials, package components, intermediate raw materials, and finished products. The information maintained in this system usually includes a lot number assigned in-house, name of manufacturer, date sampled, and type of release and release status. Using other systems, the quality assurance function compares the physical, chemical, and biological test data generated with the specifications established on products or package components and determines whether the product is released or rejected. Production personnel can access these systems online and determine the status of the materials they have produced or determine the status of raw materials to be used in production. A sample output is shown in Fig. 9.

Interactive Voice Response (IVR) Systems

An IVR system is a specially configured personal computer that contains unique voice software that enables a sound-based interface between a user and the system via a telephone. On the most basic level, these systems let

STABILITY PROTOCOL FOR A TABLET DOSAGE FORM PACKAGED IN ALUMINUM FOIL POUCH												
STABILITY NO. : 20000												
***** EVALUATION PERIOD - WEEKS *****												
STORAGE CONDITION	CODE	0	8	13	26	39	52	78	104	156	208	260
+5 °C	C		Ct	Ct	Ct	Ct	Ct	Ct	Ct	Ct	Cn1	Cn1
25°C-60% RH CL	NH	CP		CP	CP	CP	CP	CP	CP	CP	Cn	Cn
30°C-60% RH CL	XH			CP	CP	Cn1	CP	Cn				
40°C-75% RH CL	BH		CP	CP	CP	Cn						
<u>Abbreviations :</u> P - Physical Tests Ct - Control Sample Cn - Contingency Sample S - Special Instructions <u>Analytical Tests:</u> E--EXPLORATORY TEST (For Internal Use Only) C: 3, 10, 11, 17, 22 <u>Compound for Assay:</u> COMPOUND A <u>Impurities or Degradation Products:</u> COMPOUND B <u>Dissolution :</u> IN GASTRIC FLUID <u>Physical Tests:</u> Appearance of tablet and package. Compare to control stored at 5 deg C. Hardness <u>Instructions:</u> ICH Study . Stored locally and testing at Contract labs. Store the following number of samples counted as 1's : c=100 P=10 Ct=10 Cn=300 Cn1=20 Sample Cn (contingency samples). A1 pouches are packaged in strips of 2x4's for a total of 8 pouches. Each pouch =1's. (3). ASSAY (10). DISINTEGRATION (11). DISSOLUTION (17). IMPURITIES OR DEGRADATION PRODUCTS (22). MOISTURE												

Fig. 3 A sample stability protocol generated by using a computer program.

callers exchange information with a computer over the telephone without a human intermediary. Popular applications of IVR systems include banking by phone, flight scheduling, and shipment tracking. In the last five years, IVR systems have gained acceptance in the management of clinical trials. Some of the current uses

include randomization of patients, drug supply inventory management, real-time patient enrollment status, and emergency code breaking in double blind clinical trials. These systems enable the conduct of clinical trials in multiple countries by providing customized voice prompts in various languages.

STABILITY CALENDAR FOR THE PROTOCOL 20000

NH - 0.....	10 - DEC-1999
C - 8 BH - 8.....	4 - FEB-2000
C - 13 NH - 13 XH - 13 BH - 13.....	10 - MAR-2000
C - 26 NH - 26 XH - 26 BH - 26.....	9 - JUN-2000
C - 39 NH - 39 XH - 39 BH - 39.....	8 - SEP-2000
C - 52 NH - 52 XH - 52.....	8 - DEC-2000
C - 78 NH - 78 XH - 78.....	8 - JUN-2001
C - 104 NH - 104.....	7 - DEC-2001
C - 156 NH - 156.....	6 - DEC-2002
C - 208 NH - 208.....	5 - DEC-2003
C - 260 NH - 260.....	3 - DEC-2004

Fig. 4 A sample computer-generated stability calendar.

STABILITY DATA TABLE FOR INCLUSION IN A REPORT							
STABILITY OF ABC PRODUCT, 100 MG CAPSULES							
STABILITY STUDY NUMBER	20000						
PACKAGE DESCRIPTION	ALUMINUM FOIL POUCH, 30 CAPSULES						
IDENTIFICATION CODE	AXY-142-84						
DATE STORED	21-FEB-95						
ASSAY SUBJECT	COMPOUND A		<===== DISSOLUTION =====>				
ASSAY METHOD	HPLC, %		UV				
STORAGE CONDITION	EVALUATION PERIOD (WEEKS)		% DISSOLVED				
			Hours				
			1	2	4	5	8
	0	100.4	6.3	13.7	29.2	36.8	58.0
40°C	4	104.1	5.6	11.7	24.0	30.1	47.3
	13	98.9	6.2	12.9	26.4	33.2	51.7
	26	100.1	--	--	--	--	--
	52	99.0	--	--	--	--	--
	104	101.6	--	--	--	--	--
40°C-75% RH OP	4	105.0	6.0	12.4	25.4	31.8	49.5
	13	99.9	6.2	13.1	27.1	34.0	53.1
30°C-60% RH CL	13	99.7	5.6	13.0	28.7	36.3	57.0
	26	99.7	6.7	14.2	29.3	36.7	56.5
	52	99.6	6.2	13.1	27.3	34.2	53.2
	78	100.0	6.3	13.4	28.0	35.3	55.3
	104	100.0	5.2	11.3	24.4	31.0	49.6
	156	105.2	5.3	11.2	23.9	30.4	49.3

Fig. 5 A sample table generated from a stability database for inclusion in a report.

CUMULATIVE ANALYTICAL DATA										
Name of Product: ABC PRODUCT 100 MG TABLETS									7-Apr-2000	10:37 AM

CUMULATIVE ASSAY DATA BY STABILITY NO: 20000										

COMPOUND	STAB NO	CODE	DATA			AVG	S.D	UNITS	METHOD NO.	ANALYSIS NUMBER
COMPOUND A	20000	A-4	94.8	95.9	96.6	95.8	0.9	#LBL CLM	TAM 30-992306	99-0862
COMPOUND A	20000	A-8	95.4	97.8	95.3	96.2	1.4	#LBL CLM	TAM 30-992306	99-0862
COMPOUND A	20000	A-13	97.9	96.4	97.2	97.2	0.8	#LBL CLM	TAM 30-992306	99-0864
...	
...	

CUMULATIVE IMPURITY DATA BY STABILITY NO: 20000										

COMPOUND	LOT	CODE	DATA			AVG	S.D.	UNITS	METHOD NO.	ANALYSIS NUMBER
COMPOUND B	20000	A-4	<0.1	<0.1	<0.1	<0.1		%	TAM 30-892306	99-0862
COMPOUND B	20000	A-8	0.1	0.1	0.1	0.1		%	TAM 30-892306	99-0863
COMPOUND B	20001	A-13	0.1	0.1	0.1	0.1		%	TAM 30-892306	99-0864
...	
...	
The letter A in CODE field refers to storage condition 55°C										

Fig. 6 Cumulative analytical data for review purposes.

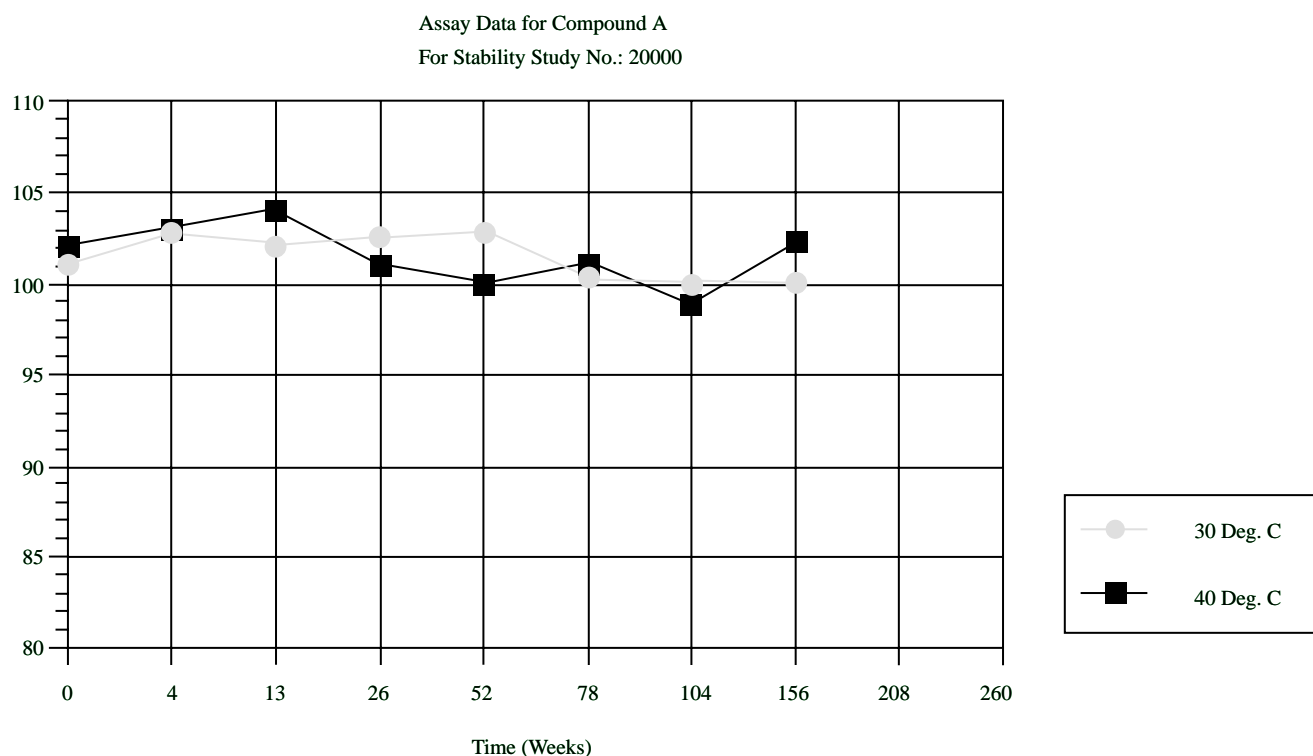


Fig. 7 Graphical representation of cumulative assay data for a stability study.

Group Collaboration Tools

The exponential growth in Internet technologies has allowed companies to set up private Internets, known as Intranets, for effective sharing of information and online meetings where the participants can see each other to collaborate on projects. These tools help to archive project documentation and provide capabilities for searching information across several projects with ease. Some current uses of these tools include collaboration between geographically distributed pharmaceutical R&D and manufacturing plants on process technology data and information.

Document Management and Publishing Systems

In the last 15 years, the size of a typical new drug application has grown from thousands of pages to several hundred thousand pages. To cope with this and to help speed up the submission process, a number of software tools have been developed. These tools provide control and access to documents in a collaborative environment and help publish electronic and paper copies of submissions.

Internet-Based Applications and Tools

Easy navigation of the Internet led to the development of software applications that use web pages for data collection, analysis, browsing, and reporting. As a result, it is becoming easier to deploy software that gathers patient enrollment status information and certain types of clinical trial data. In addition, a number of search tools have been developed that query information across several hundred million pages of information currently available on the Internet. Examples of these tools include Alta Vista, Excite, and HotBot. More information on these tools can be found on www.altavista.com, www.excite.com, and www.hotbox.com.

Project Management Systems

The successful development of a new pharmaceutical product requires careful planning of various activities and resources, as well as tracking the project's progress. A small project is not difficult to monitor manually; however, multiple projects benefit from an automated tool to support the planning process as well as the monitoring of various activities. In addition, these systems will help develop "what if" scenarios for the resources in

REPORT OF ANALYSIS					
PRODUCT DEVELOPMENT ANALYTICAL DEPARTMENT					
REPORT OF ANALYSIS					
Analysis No: 99-2721 Version 1			Group: AN		
Material: PLACEBO 50mg FILM-COATED TAB			Proj Code: 5811		
Lot No. RCT XXXXX Sample :BEGIN, END					
Requested By: C MAH			Dept: CLP		
Objective: R & D Q.C. RELEASE					
Received Date: 17-Dec-1999			Completion Date: 14-Jan-2000		

R & D QUALITY CONTROL		* G.R. Dill		18-JAN-2000	

APPROVED					
*REEVALUATION DATE: 30-APR-2002					

Almedica blister strips					

<u>APPEARANCE:</u>					
Lot	Sample	Result			
RCT XXXXX	BEGIN	WHITE ROUND CONVEX TABLET.			
RCT XXXXX	END	WHITE ROUND CONVEX TABLET.			
<u>IDENTITY:</u>					
Lot	Sample	TEST/COMPOUND	RESULT		
RCT XXXXX	BEGIN	UV/AA-66106	TABLET STRENGTH AND UV SPECTRUM CONFORM TO AA -66106 50 mg TABLET.		
RCT XXXXX	END	UV/AA-66106	TABLET STRENGTH AND UV SPECTRUM CONFORM TO AA -66106 50 mg TABLET.		
COMMENT					
STRIP COLOR CODE OF BEGINNING AND END SAMPLES: MAGENTA.					
Lot. No. RCT XXXXX		Sample : BEGIN, END			
TEST	ANALYST	BOOK	PAGE	SUPERVISOR	METHOD NO.
1	L. K. BROADUS	11511	56	G. MADSEN	N99-044-B1099
16	L. K. BROADUS	11511	56	G. MADSEN	N99-044-B1099

Fig. 8 A sample report of analysis.

QUALITY ASSURANCE INFORMATION								
LOT NUMBER	ITEM NUMBER	MATERIAL	VENDOR INFORMATION	SUBMITTER DEPT.	GROUP	DATE RECEIVED	TYPE OF RELEASE	ANALYSIS NUMBER
SP1890	R00106	MICROCRYSTALLINE CELLULOSE NF/BP/DA	FMC 175	K.DUN MMD	* PDAD *MICRO *RDQC	29-Jan-1998	FULL	98-0268
RELEASED ON 27-Mar-1998 Rev. Date: JAN 2000								
SP1890	R00106	MICROCRYSTALLINE CELLULOSE NF/BP/DA	FMC	K.DUN	* RDQC	29-Jan-1998	FULL	NONE
RELEASED FOR EXPLORATORY USE ONLY ON 29 -Jan-1998 Rev. Date: JAN 2000								
SP6846	R00106	MICROCRYSTALLINE CELLULOSE PH101	FMC 1930	K.DUN MMD	PDAD MICRO RDQC	7-Sep-1999	FULL	99-1848
STATUS : PENDING								
SP3591	R00102	MICROCRYSTALLINE CELLULOSE PH102	FMC CORP 2855	A.SUL PDD	* PDAD *MICRO *RDQC	25-Jan-1999	FULL	99-0189
RELEASED ON 23-Feb-1999 Rev. Date: JAN 2001								
SP3591	R00102	MICROCRYSTALLINE CELLULOSE PH102	FMC CORP	A.SUL	* RDQC	27-Jan-1999	FULL	NONE
RELEASED FOR EXPLORATORY USE ONLY ON 21 -Jan-1999 Rev. Date: JAN 2001								

Fig. 9 A sample output from a quality assurance information system.

Table 1 Commercially available statistical software packages

Name	Source
Statistical Analysis System (SAS)	SAS Institute Inc. SAS Campus Drive Cary, NC 27513-2414 Web address: http://www.sas.com/
MINITAB	Minitab Inc. 3081 Enterprise Drive State College PA 16801-3008 Web address: http://www.minitab.com/
WINNONLIN	Pharsight Corporation 800 W. El Camino Real, Suite 200 Mountain View, CA 94040 Web address: http://www.pharsight.com/

the new projects as well as help to terminate the projects. Several project management systems currently available on the market are designed to fulfill these needs. These systems are available for all types of computers—from personal computers to mainframes.

Problem-Solving Applications

The computer is an excellent tool for statistical analysis of data and for solving mathematical problems. Commercial software packages (Table 1) are generally used for statistical analysis. The scientist often works with the statistician to design an experiment and determine the most appropriate statistical method to analyze the collected data. Custom-designed software or commercial software that allows tailoring is generally used to solve mathematical problems.

Spreadsheet Software

This software is available commercially for almost all types of computers and is becoming a valuable tool to solve mathematical problems. The data are entered in the form of tables, the mathematical formulae are defined, and at the push of a button the answer is obtained. The user can easily modify the formula as well as add rows or columns of data and obtain the results easily and quickly. As such, this software enables the user to determine “what-if” scenarios for the problem at hand. It also is extremely valuable in helping the pharmacist compute percent composition and the amount of each ingredient necessary for preparing different strengths of dosage forms. In addition to providing the computational ability, the

software also enables the user to prepare data and information in the form of tables and plots.

Expiry Date Prediction

Expiration dating, required on the label of a drug product by good manufacturing practices (GMPs), is arrived at by analysis of data collected on samples exposed to storage conditions defined in a stability protocol. The establishment of an expiry date has evolved from “eyeballing” the time–temperature plot on graph paper and drawing an approximate regression line, to the rigorous application of physical–chemical laws and sophisticated statistical analysis using computers (1). Using the speed and accuracy available from the computer and expert advice from a statistician, the pharmaceutical scientist can try various statistical models to fit the data and arrive at an optimal expiration date.

Pharmacokinetics

For a number of years, computers have been successfully utilized in pharmacokinetics (2) to: 1) fit blood-level data to the appropriate model (single, two, or multiple compartments) and to calculate model parameters, such as absorption rate constant, elimination rate constant, half-life, and volume of distribution (3); 2) evaluate bioavailability parameters, such as peak plasma concentration, time of peak concentrations, and area under the concentration time curve obtainable from a blood-level curve (4); and 3) calculate dosage regimens in patients with renal failure (3).

Currently, the growing trend is to make use of physiologically–based pharmacokinetic models to study the behavior of drugs in animals and extrapolate the data to humans (4, 5). In this context, computers will be of immense help in developing predictive models that might assist in the scale-up of animal data to humans and predicting the concentration of drugs in human body fluids.

Microcomputers are also used to systematize, speed up formulation, simplify manufacturing processes, and reduce the number of needed bioavailability studies through simulated models and plasma level predictions. Commercially available software packages, such as WINNONLIN, SAS (Table 1), and other custom-designed programs, are generally used to solve the often complex mathematical formulae encountered in pharmacokinetic research and applications.

Communication Aids

The computer has become an excellent tool for communication. It can store information entered by one user and send a signal to another user who then reads the information. When several computers are connected through a network, users of one computer can send information to other users in the network almost instantaneously. This feature makes the computer a powerful and effective communication medium for organizations with various remote locations. Examples of communication aids are discussed below.

Electronic mail

With the proliferation of the Internet, electronic mail (e-mail) has become an integral part of browsers, such as Netscape Navigator and Microsoft Internet Explorer. At times, e-mail software is also bundled into other office automation software packages, such as word processing. This e-mail software allows users to send messages to other users (with e-mail accounts) anywhere on the Internet. In addition to sending mail messages to other users, the software allows the user to file, forward, reply, delete, and print messages sent by others. Most e-mail software packages allow attachment of files that contain text, graphics, video, or audio information. Electronic mail provides almost instantaneous communication to remote users, helps improve the productivity of firms operating on a worldwide basis, and improves communication with regulatory agencies.

Distributed information management systems

A pharmaceutical company with different geographic locations often has a need to communicate between locations in order to share data and information. Using the concept of distributed data processing, the company sets up a data processing center at each location and connects these centers by networks. In this setup, common data management systems running at each site make it easy for the company to work on the same project at several sites and funnel the information back and forth. For example, a company conducting toxicology or stability studies can consolidate the information at one site and use it for generating regulatory reports or for other purposes. In addition, these systems help transfer technology developed at one site to another site and allow access to data generated from other sites.

Laboratory Automation

Many laboratory instruments available on the market today contain built-in microprocessors that process data

collected on samples and display or send the answer to a computer. In addition, they may have an interface that attaches to an external computer for processing the data generated by the instrument. With regard to experiments or analyses performed frequently, it is often desirable to interface the instrument to a computer to aid in the subsequent analysis of data. Some of the commonly encountered systems following.

Chromatography systems

The computer has become a valuable tool in automating chromatographic techniques, especially high-performance liquid chromatography (HPLC). Prior to automation, strip chart recorders were used to record the analog signals from an HPLC detector, and the calculations were done manually. In automated systems, the output from an HPLC detector is digitized through an analog-to-digital (A/D) converter, the digitized information is stored in the computer, and the data is analyzed to compute the final result. The inherent disadvantages associated with a strip-chart recorder are overcome in automation because the computer enables the analytical chemist to change calculation parameters (e.g., the base line, peak start, and peak end) as needed. As a result, the number of repeat analyses to be performed is minimized. The computer, utilized as a systems controller, is especially useful in applications that require techniques such as column switching and solvent gradients.

Automated dissolution systems

The application of computers to solid dosage form dissolution allows for nearly complete automation. Analyst intervention is limited to the analysis setup. The computer executes all other steps by controlling the system devices, such as the sampling pump and the spectrophotometer generally used for analysis. Typically, a fraction of the sample solution is pumped into a spectrophotometer flow-through cell, where its absorbance is measured. The computer uses the absorbance reading to perform calculations and reports the analysis results in a tabular or graphic format. Sampling, analysis, quantitation, data handling, and reporting are all performed by the computer according to the analysis parameters specified before each run or included in a setup table (6). Automation has played a key role in the development of dissolution systems, which attempt to simulate changes in pH in the gastrointestinal tract or the presence of bile salts (7).

Microprocessor-based balances

Microprocessor-based balances are used in the pharmaceutical industry for automating a number of routine operations. Some organizations use these balances to

automate the United States Pharmacopoeia (USP) weight variation test. In this application, the printer attached to the balance prints a hard copy of the individual weights of the dosage form as well as the average and standard deviation (8). The balances are also used in the toxicology and pathology laboratories to weigh animals or organs and then transmit the information to a central computer for further processing.

Process Control

Computerized process-control systems are used to measure process variables through sensors at predefined intervals, make appropriate decisions, and take appropriate actions to keep the process under control. In an open-loop process-control operation, the computer records sensor readings, compares readings against standards, and notifies human operators of needed actions to regulate devices. In more complex closed-loop process-control operations, the computer records the measurements, makes the comparisons with standards, and transmits signals to the regulating devices to make the necessary changes (9). The automation of process control is often limited by the availability of pharmaceutical processing equipment. To date, a number of companies have succeeded in implementing the process-control applications. Two examples are automated tableting and automated freeze-drying.

Automated tableting

Instrumented tablet presses with computer interfaces allow the pharmaceutical scientist to study the mechanism of compaction and the relationship of the mechanism to tablet-compaction properties and formulations. In addition, automated systems are useful to develop compression profiles for reference purposes, to control weight of tablets during development and production, and to monitor punch wear. This automation reduces the burden on personnel faced with the requirements of quality control (10). Merck Sharp and Dohme's major production facility in the United Kingdom is fully computerized to manufacture a high-volume tablet product as well as multiple-tablet products (11).

Automated freeze-drying

Pharmaceutical or biological products that are unstable in solution form are converted to a stable solid state using the freeze-drying technique and later reconstituted prior to administration to a patient. Freeze-dryer equipment manufacturers are now upgrading their equipment with automated control systems. Typically, the automated control system has a personal computer and a programmable

logic controller. The personal computer is used to initiate the process, monitor the process, maintain recipes, and archive data. The programmable logic controller is used to control the freeze-drying, sterilization, and cleaning process by means of instructions downloaded from the personal computer.

The computer is useful in avoiding long freezing times since it is possible to monitor product temperature with ease, and eliminate the need for human operator intervention during transition from one freeze-drying phase to the other. The computer also helps in process documentation and in developing cost-effective freeze-drying profiles that should be useful in scaling up a product from the pilot plant to the production area.

Regulatory Issues Affecting Use of Computers

As previously described, the use of computers and computerized systems in the pharmaceutical industry is growing at a rapid rate. Some of the systems used in the industry range in complexity from the use of personal computers for performing simple tasks (word processing, e-mail, Internet access) to the use of powerful computers in process-control applications. In addition, to help eliminate or reduce paper usage, the pharmaceutical industry has implemented a number of electronic batch record systems in drug substance and product manufacturing to keep track of process documentation.

As the regulatory authority, FDA wants the computerized systems to be validated and requires manufacturers to comply with regulations that cover electronic records and signatures (see http://www.fda.gov/oralcompliance_ref/part). The FDA's *General Principles of Validation Guideline* (12) defines validation as establishing documented evidence that provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes. This definition applies to computer systems as well as to the processes. To date, the industry has done a good job in validating computerized systems and is beginning to address the implications of electronic records/signature rule. For additional information on computer systems validation and guidance on electronic records and signatures, visit <http://www.pharmaportal.com/articles/pt.fda99.cfm>.

Future Trends

The computer provides solutions to problems that can be defined in a language it understands. Although the computer works faster than humans, it still lacks the

human qualities of intuition, insight, and experience. A great deal of effort has been expended in the development of computer languages and software tools that allow for ease in writing programs and obtaining desired solutions. In spite of this, researchers still need programmers to translate their requirements to the computer; it is probably unrealistic at this time to assume that they can do away with programmers. The challenge facing scientists today is to determine ways in which to help nonexperts accomplish sophisticated tasks with computers. In this regard, Internet technologies have helped a great deal by implementing simple-to-use and standardized navigation techniques for accessing and interacting with information on a personal computer. Any savvy Internet user can now buy airline tickets, shop for electronic items, and search the Internet for information on diseases, product specifications, etc. This trend will continue for the foreseeable future.

Human genomics is another area that is beginning to benefit from the use of computers. Researchers have just begun amazing discoveries in this particular field. As more information on the human genome is discovered, we will begin to understand human diseases at a gene level, thus allowing researchers the opportunity to discover cures for deadly diseases.

REFERENCES

1. Vadas, E.B. Stability of Pharmaceutical Products. *Remington: The Science and Practice of Pharmacy*; Mack Publishing Company: Eaton, PA; 1995; 639–647.
2. Danek, A. *Int. J. Clin. Pharmacol.* **1978**, *16*, 345–350.
3. King, F.G.; Dedrick, R.L.; Farris, F.F. *J. Pharmacokinet. Bonham* **1986**, *14*, 131–155.
4. Tsuji, A.; Nishide, K.; Minami, H.; Nakashima, E.; Terakasi, T.; Yamana, T. *Drug Metab. Dispos.* **1985**, *13*, 729–739.
5. Aiache, J.S.; Turlier, M.; Tanguy, A. *Eur J. Drug Metab. Pharmacokinet.* **1984**, *9*, 89–102.
6. Pagone, F.M. Pittsburgh Conference and 2 Exposition Abstracts, Atlantic City, NJ, 1986; No. 158.
7. Wickman, A.; Pagone F.M. Pittsburgh Conference and 2 Exposition Abstracts, Atlantic City, NJ, 1986; No. 280.
8. Paruta, A.N.; Breon, T.L.; Greene, D.S. Computerized Weight-Variation Testing for Uncoated Tablets and Capsules. *Automation of Pharmaceutical Operations*; Fraad, D.J., Ed.; Pharmaceutical Technology Publications: Springfield, OR, 1983; 59–64.
9. Sanders, D.H. *Computers Today*; McGraw-Hill: New York, 1983; 595–596.
10. Schwartz, J.B. The Instrumented Tablet Press: Uses in Research and Production. *Automation of Pharmaceutical Operations*, Fraad, D.J., Ed.; Pharmaceutical Technology Publications: Springfield, OR, 1983; 229–239.
11. Corbin, R.L. A Computerized Manufacturing Facility. *Automation of Pharmaceutical Operations*; Fraade, D.J., Ed.; Pharmaceutical Technology Publications: Springfield, OR, 1983; 241–243.
12. *U.S. Department of Health and Human Services Guide to Inspection of Computerized Systems in Drug Processing*; 1983.
13. Fraade, D.J. *Automation of Pharmaceutical Operations*; Pharmaceutical Technology Publications: Springfield, OR, 1983.
14. Fraade, D.J. *The Aster Guide to Computer Applications in the Pharmaceutical Industry*; Aster Publishing Corporation: 1984.
15. Harris, J.R. *The Aster Introduction to Microcomputers in Pharmaceutical Operations*; Aster Publishing Corporation: 1984.
16. deSpautz, J.F. *Automation and Validation of Information in Pharmaceutical Processing*; Marcel Dekker, Inc.: New York, 1998.
17. Rowland, M.; Tozer, T.N. *Clinical Pharmacokinetics: Concepts and Applications*, 3rd Ed.; Williams & Wilkins: Baltimore, 1995.

COMPUTER-ASSISTED DRUG DESIGN

J. Phillip Bowen

University of Georgia, Athens, Georgia, U.S.A.

Michael Cory

GlaxoSmithKline, Research Triangle Park, North Carolina, U.S.A.

INTRODUCTION

Drug discovery and development are extremely time-consuming and costly processes. For every drug that reaches the market, there are more than 10,000 compounds synthesized, characterized, and tested for biological effects. Hundreds of millions of dollars are invested in the basic research and clinical studies which lead to its FDA approval and subsequent marketing. Traditionally, drugs have been "discovered" predominately through random or targeted screening efforts, followed by discrete structural changes in the molecule to optimize the properties responsible for the desired activity. With rapidly increasing costs, diminishing resources, more intense public scrutiny, greater government regulations, and higher expectations, this hit-and-miss approach is neither efficient nor economical.

The development of new medicines has been and continues to be a spectacularly successful scientific endeavor. While the traditional preparation and modification of complex pharmaceuticals has been a laudable achievement of synthetic medicinal chemists, the advent of computers has ushered in a new exciting approach to drug discovery. The new drug discovery activities, including that of computational chemists, have grown out of these technology changes. One of the clear trends of science has been the reduction of chemistry to particles in motion. Molecules undergo rapid translational, rotational, vibrational, and bending motions, which, in most cases, can be quantified using theoretical physics. Chemical reactivity may be attributed to these dynamic features of molecules. There is a direct relationship between molecular structure and biological activity. Through an intimate understanding of such molecular behavior, the medicinal chemist has the ability to gain profound insights into the most probable drug conformations, as well as the structural and energetic factors responsible for favorable drug-receptor interactions. With a combination of experimental data and computer-based information, the medicinal chemist is in a much better position to predict drug action. Unlike hand-held mechanical models,

computer-generated structures add a new dimension to chemical perception by including the energy for a specific conformation defined by the Cartesian coordinates, which define the molecular array. Of course, other factors such as administration routes, absorption, distribution, metabolism, and elimination (the domain of pharmacokinetics) affect the activity of drugs and are increasingly being considered in the early phases of drug design.

The past twenty years have seen the development of a number of new approaches to drug design. The increased processing power of computers and the rapid display of computer-generated structures is simply amazing since the first writing of this article approximately a decade ago. These exciting developments over the last two or three decades have allowed the emergence of sophisticated computer programs specifically designed to assist medicinal chemists in developing new drugs. The combined application of molecular graphics, computational chemistry, as well as chemical and biological information is commonly called computer-assisted drug design (CADD). These computer-based approaches promise to fulfill a long-coveted goal of medicinal chemists: the prediction of biological activity prior to extensive laboratory synthesis and biological testing. One day it may not be unreasonable to expect medicinal chemists to design active molecular structures in a fashion analogous to the way engineers plan buildings, although the problems associated with biological problems are much more complex and frankly less understood. Computer-aided molecular modeling methods are still in their infancy, with exciting new methods and applications being reported at a staggering rate. Nevertheless, CADD approaches have already had a substantial impact on pharmaceutical design, discovery, and development, moving from the realm of after-the-fact rationalizations to a more beneficial predictive role.

This article will provide a brief introduction to computational chemistry, molecular modeling, and CADD theory. Clearly, it is impossible to cover in this short review all of the useful and/or interesting CADD approaches. Although this is relatively a new field of

research, there are many examples from which to choose. Today, there are a number of pharmaceutical agents in various stages of clinical trials and/or on the market in which computer-based methods were used. Research examples have been included to illustrate more completely the conceptual points. Some of the fundamental methodologies are discussed, including the important underlying mathematical formalisms. In addition, historical perspectives have been incorporated. Finally, the ever-changing hardware and software trends are discussed.

METHODS

Molecular Modeling

Molecular modeling of organic compounds has a long history. John Dalton (1) had a set of wooden spheres, representing atoms, drilled to receive rods representing bonds. These original ball-and-stick models are on display at the London Museum of Science and date to 1810. The most widely known recent application of molecular modeling to the solution of a significant organic chemistry structural problem is the Nobel prize-winning work of J. D. Watson and F. H. C. Crick for the development of their model for the structure of DNA. Initially, Watson and Crick used homemade molecular models cut by hand from cardboard. Later, as they improved their understanding of the structures and the conformations of the bases in the nucleic acids, they had more exact models machined from metal. At the time of their work (mid 1950s) controversy existed over the correct tautomeric forms for the heterocyclic bases. Based on old experimental results, the first Watson and Crick models had the incorrect tautomers. After building models with the correct tautomeric forms of the bases, they quickly arrived at a conformation for DNA that not only matched experimental data but also suggested a method by which DNA could be responsible for cellular replication (2, 3).

In the two decades following this landmark work, various types of hand-held models became commercially available, and they have been used extensively by medicinal chemists. The wooden ball-and-stick models frequently used by high school students, as well as the metal Dreiding and Kendrew (4) models (not to mention the plastic varieties), are still quite popular in organic chemistry laboratories. Before the availability of high-resolution computer graphics systems, X-ray crystallographers depended heavily on metal models, including the Kendrew type, for converting computer-plotted electron density maps into physical models that could be used to study proteins.

Linus Pauling, in his early studies of the conformations of the *alpha* helix in proteins, suggested it would be helpful to obtain solid molecular models that would allow modeling of large pieces of a polypeptide structure. This suggestion eventually led to the CPK (Corey, Pauling, Koltun) (5) spacefilling models which are made from plastic and are designed to effectively represent the true size of atoms. These models were designed under the guidance of an IUPAC committee. Their quality and popularity led to their commercial availability (6). Many laboratories used large quantities of these models to assist in drug design and the understanding of molecular shapes. A few laboratories attempted to use such models for the study of proteins and nucleic acids. Due to their tactile properties, these models are excellent teaching tools and led to an easy understanding of problems in stereochemistry and molecular shape. The mechanical stress on the plastic parts, however, resulted in compromising the integrity of models of large molecules. It turned out that the models for large molecules were less useful than computer representations. Further complications arise when attempts were made to simulate ligand-protein (or drug-receptor) interactions. Importantly, there is no possibility of determining accurately the energy differences among various conformations.

The display of molecular representations on a computer screen has been an objective of many computer scientists over the years. Significant decreases in the cost of computing coupled with dramatic increases in the speed and quality of computer displays have led to the increasing use of computer graphics displays for molecular modeling. Graphics displays have grown from the early ball-and-stick ORTEP (7) plots of small X-ray crystal structures to interactive displays that show spacefilling models of entire proteins. Computer-generated molecular displays are not only integral parts of structural biology and structural chemistry research, but they have also moved into the realm of advertising, as seen on the covers of scientific journals and in the many pharmaceutical media advertisements. In other words, computer-generated images not only impart specific and sometimes subtle chemical and biological information not possible to obtain with hand-held models, but they also tend to represent state-of-the-art methods. Most of these developments were driven by the rapidly growing computer-assisted design market, rather than the much smaller chemistry market.

The growth of visualization applications has been boosted by improved computer graphics displays and increasing computational chemistry capabilities, as indicated above. The initial growth of the market was induced through the rapid acceptance of computer-assisted

design in the aircraft manufacturing industry. Until about 1983, sophisticated molecular modeling software was the domain of some academic scientists; most of the advances took place in X-ray crystallography laboratories. The most sophisticated of these displays allowed full three-dimensional transformation of the coordinate display by turning a dial. Custom-designed computer molecular modeling and display software, now available from multiple vendors, allow the interactive manipulation of the display through a joystick, mouse, dial, and/or keyboard commands.

The use of computer-generated molecular models has many advantages over the more traditional manipulation of physical molecular models. Since the molecules are represented in the computer as sets of three-dimensional (3-D) atomic coordinates, a large number of molecules can be handled simultaneously. A feat which is essentially impossible using mechanical models. Any number of computer-generated images can be superimposed and analyzed. These techniques have provided medicinal chemists with information about the relative size and shape of sets of molecules with similar biological properties. The superimposition of different conformations of the same molecule can provide information about the relative space a molecule can occupy. Since the computer maintains a set of coordinates for each of the atoms in each molecule, information about bond lengths, bond angles, and torsion angles of a model can be rapidly obtained. Additionally, nonbonded and direct through-space distances, so important in defining the shape of receptor sites, can also be readily obtained and compared. Moreover, the information can be stored, retrieved, and re-analyzed at later times.

From the preceding discussion, one extremely important aspect of molecular modeling which cannot be overstated: visualization of molecules. When molecular interactions between a drug and its receptor are displayed, manipulated, and rotated on a graphics terminal, researchers can identify important and unimportant molecular interactions. It is through such computer exercises that medicinal chemists can confirm, reject, or develop hypotheses regarding the molecular origins of biological activity. Perhaps one major drawback to computer-generated structures is the fact that one is unable to handle them physically, which could be an artifact of previous experiences with hand-held molecules in sophomore organic chemistry classes. Interestingly, there is a growing trend in organic chemistry education to move toward computer graphic representations of organic structures. On computer graphics terminals, to give the illusion of three-dimensional visualization, a technique

called *depth cueing* is used to shade structures in such a way that objects further away from the viewer are slightly dimmer, while objects that are closer appear brighter. This optical effect works reasonably well. Another stratagem that creates the illusion of three-dimensional perception is to have the molecular structure slowly rotate about its center of mass or slowly rock back and forth. Current technology allows breathtaking stereo viewing of molecular display on desktop computers. An example of this is the crystal-eyes system (<http://www.stereographics.com/>).

There continues to be important developments in the molecular modeling area. The first is the development and use of algorithms that describe the surface area of molecules. Software utilizing these algorithms allows the comparison of computed surface areas. One of the first algorithms to be widely implemented in the study of molecule surfaces is that of Connolly (8). In it, a sphere of a given radius (usually the radius of a water molecule) is "rolled" over a defined surface, usually the van der Waals surface. Wherever the sphere touches the macromolecule, a new contact surface is generated. The display of this surface allows the medicinal chemist to see what should be considered the *solvent accessible* surface. An alternative approach, much more efficient in computational and display time, is to place a dot surface over the molecule and to compute the surface using expanded atomic radii. The more recent implementations of this algorithm display the electrostatic density at the surface. This provides the chemist with a picture of what charge might be exposed to solvent. As applied to drug design, the *solvent excluded* surface allows an indication of the topology of a receptor site. Plotting physicochemical functions, such as the electrostatic potential, on the surface of this model indicates the forces acting upon a molecule that binds to that surface. Today, it is possible to display transparent molecular surfaces rapidly, which gives the impression that the molecular skeleton is suspended in a gelatin-like volume.

The second major change, interrelated to the discussion above, is related to the decreased cost and increased speed of computation. Raster computer displays generated widespread interest in the development of display algorithms for calculating high resolution spacefilling representations of molecules. Many algorithms, motivated in part by the advances in computer technology, were developed to give rapid, high resolution CPK-like displays of molecular surfaces. Most of these systems provide a static display computed from a specific molecular orientation. Real-time, three-dimensional manipulation of these "CPK" models was demonstrated successfully on some systems (9) and is now standard.

Computational Chemistry

Overview

Much of the recent progress in drug design has been based on the fundamental understanding of drug-receptor interactions and the increased availability of high quality structural data. Quantitative correlations have been made between biological activity and molecular properties. Computer experiments offer complimentary means to acquire essential information that may be quite difficult or impossible to obtain empirically. Of course, these theoretical approaches need to be properly calibrated to existing experimental data to insure that if there are any systematic differences they are clearly understood at the outset. No one is seriously suggesting that computer analysis will replace experimental work. The mathematical models presently used are just that—mathematical models approximating complex physical and biological properties. Traditionally, a great deal of attention has been focused on the structural fit between receptors and ligands, like keys within locks. Today, this idea of a lock-and-key analogy is a well-recognized exaggeration since small molecules and receptors are not static structures but constantly changing to adopt new conformations in response to their environment. In considering pharmacodynamic interactions, favorable electrostatic interactions are critically important. Charge density and the complementarity of charges between ligand and receptor are readily accessible to calculation, although there are several different ways to generate charges. Important quantitative structure–activity relationships (QSAR) between nontraditional parameters such as HOMO–LUMO interactions and bond orders have been correlated with biological activity.

Researchers involved in drug design have been criticized for focusing too extensively on the interaction energies governing ligand–receptor binding while neglecting important solvation and desolvation effects and pharmacokinetic effects. These other aspects of drug availability, as well as toxicity, are more complex to model and consequently more difficult to predict. Progress, however, is being made. Until much more research verifies the accuracy of pharmacokinetic and toxicity models, energy-based approaches focused on drug–receptor fitting will dominate. Alternative chemical information based methods will play increasingly important roles in drug design.

Quantum mechanics

History and concepts: In the late nineteenth and early twentieth centuries, the applications of classical

mechanics failed to describe accurately blackbody radiation or atomic spectra (10a, 10b). For nearly two centuries, Newtonian mechanics had been the cornerstone of physics, essentially unassailed and unailing. According to the prevailing attitude among most physicists of the late nineteenth century, all physical theories had been discovered (10a, 10b). The only remaining challenge for future generations of scientists was to refine minor details and extend the accuracy to an additional decimal place. Wein in 1896 and Rayleigh in 1900 had tried to explain blackbody radiation using standard classical methods. Every attempt, however, by theoreticians of the day failed to reproduce all of the available empirical data. This failure of classical physics has been termed the “ultraviolet catastrophe.” It was not until Planck made his revolutionary proposition, that the oscillators themselves needed to be quantized, that a consistent mathematical solution emerged.

Based on the Einstein wave-particle solution for the photoelectric effect, De Broglie was the first to suggest that if particle characteristics were associated with electromagnetic radiation, then matter should also have wavelike characteristics. De Broglie derived a simple relationship (Eq. 1) correlating the wavelike characteristics, λ , with momentum, p . Large objects would have a vanishingly small wavelength. The wavelike characteristics of matter would only manifest themselves significantly at the atomic and subatomic levels since the wavelength is inversely proportional to the mass. The de Broglie equation was used to predict exact diffraction patterns which were later verified experimentally by Davisson and Germer.

$$\lambda = \frac{h}{p} \quad (1)$$

Since matter had wavelike properties, it could argue that there must be an underlying fundamental wave equation governing the behavior of matter. Erwin Schrödinger (11a, 11b, 11c) formulated the celebrated differential equation, which, for chemists, is the basic molecular expression of quantum mechanics. The idea being that if one could solve the Schrödinger equation, then all of the physical properties of a molecular system (or any system of any size for that matter) would be known. Chemistry, and ultimately biology, could be reduced to solving mathematical equations. Another more general formulation of quantum theory, using matrix algebra, was developed independently by Heisenberg and Dirac (12). Both methods have been demonstrated to be equivalent mathematical representations (13a, 13b).

The concept of spin falls naturally out of the matrix mechanics approach, but it is an ad hoc addition to wave mechanics.

To date, quantum theory, despite its many peculiar “nonclassical” microscopic aberrations, has been used effectively to explain experimental observations and to predict accurately physical effects in advance of the experiment. Interestingly, many of the founders of quantum mechanics later rejected it, primarily because it was a nondeterministic theory. The break from their classical mechanics view of the universe appears to have been too severe for them to accept. But a new generation of physicists was prepared to embrace the quantum mechanics and apply the methods to chemical structures. With a series of novel concepts and observations, physics and chemistry were changed forever. From the brilliant work of Schrödinger, de Broglie, Heisenberg, Dirac, Mulliken, Pauling, and many other pioneers, the new quantum chemistry was born.

Methodology: Unquestionably, the application of quantum mechanics to chemical bonding has revolutionized scientific thinking. In fact, the modern theoretical framework of chemistry rests on quantum physics. In principle, the Schrödinger equation may be solved for any chemical system. No prior knowledge of any analogous or related system is necessary. Exactly solvable problems are rare, due to the mathematical complexities; recourse must then be made to approximate methods, and many powerful approaches have been devised. Generally, approximate solutions must suffice for the size of molecules of pharmaceutical interest.

The Schrödinger equation is the starting point for molecular problems (13a, 13b). The symbol \hat{H} is a differential operator called the Hamiltonian operator, which is analogous to the classical Hamiltonian, inasmuch as it is a sum of kinetic and potential energy terms. E is the total energy for the system. The wavefunction Ψ depends on the position of all the particles comprising the system. Born (10a, 10b) proposed that $|\Psi|^2$, and not Ψ , is a measure of the probability distribution of the particles within a molecule. The Schrödinger equation is an eigenfunction, which in its most general form is written as follows (Eq. 2):

$$\hat{H}\psi = E\psi \quad (2)$$

All quantum chemical calculations rely on a very important assumption called the Born–Oppenheimer approximation, which treats nuclear and electronic

motions separately (Eq. 3), and it is discussed in most elementary quantum mechanics textbooks.

$$(\hat{H}_{el} + V_{NN})\psi_{el}(r, R) = E\psi_{el}(r, R) \quad (3)$$

The nuclei, being much heavier than electrons, are considered fixed in space relative to the faster moving electrons. V_{NN} is a function describing the nuclear potential energy and may be factored out of the electronic expression. This uncoupling of electronic and nuclear motion is critical to quantum mechanical formulations. Nevertheless, the Born–Oppenheimer approximation may not be accurate under certain circumstances. Once a nuclear configuration has been determined, the Schrödinger equation is then solved to give the electronic energy for a particular nuclear configuration. The process may be repeated over every conceivable nuclear arrangement. A potential energy surface can then be mapped. The most stable nuclear–electronic configuration has the lowest energy, and it is referred to as the *ground state*. Promotion of electrons to higher energy states, referred to as *excited states*, usually is not of importance in most drug design applications, although there are exceptions. The vast majority of energy computations related to pharmaceutical agents and their design have been carried out on ground state electronic configurations.

In general, quantum chemical calculations may be divided into three broadly defined subdisciplines based on the approaches taken to solve the Schrödinger equation: 1) semiempirical-based methods, 1, 2f) ab initio-based methods, and (c) density functional theory. John Pople, ab initio methods, and Walter Kohn, density functional theory (DFT), received the 1998 Nobel Prize in chemistry for their pioneering work in computational quantum chemistry.

a. *Semiempirical calculations:* Exact solutions to the Schrödinger equation are limited to special cases: a particle in a box, a particle on a ring, a harmonic oscillator, the hydrogen atom, and the hydrogen molecule ion, etc. In an effort to apply quantum mechanics effectively to chemical systems of more practical importance, several simplifying approximations have been made and refined over the years. Early work by Pople and co-workers (14a, 14b, 14c, 14d, 14e, 14f, 14g) led to an approximation called Complete Neglect of Differential Overlap (CNDO). The central idea behind this approximation was to eliminate certain sets or families of two-electron integrals of the following expression (Eq. 4). The solution of these expressions are the most complicated mathematical steps in quantum chemical calculations,

especially when integration is over four different atomic centers a, b, c, and d. The omitted integrals could then be accounted for by adding empirical equations and parameters. Years later, Dewar developed a series of methodologies that approach chemical accuracy (15). These methodologies have been incorporated into a series of computer programs, MINDO (modified intermediate neglect of differential overlap; 16), MINDO/3 (17a, 17b, 17c, 17d, 17e), MNDO (moderate neglect of differential overlap; 18a, 18b, 18c, 18d, 18e, 18f), and AM1 (Austin Method 1) (19). Correlations between drug activity and molecular orbital energies and coefficients have been made (20a, 20b, 20c). These and similar approximations are classified as semiempirical calculations.

$$\Psi_a^*(1)\Psi_b(1)\left(\frac{1}{r_{12}}\right)\Psi_c^*(2)\Psi_d(2) \quad (4)$$

- b. *Ab initio calculations*: The second quantum mechanical approach is based on what is referred to as the *ab initio* method (21). As the name implies, *ab initio* calculations use a nonparameterized method. Strictly speaking, these calculations also have a number of simplifying approximations; but they are much more demanding computationally, as no classes of integrals are eliminated, and consequently are much more expensive. There have been numerous reports and reviews in the literature on *ab initio* and semiempirical calculations, as well as detailed comparisons of the two methodologies (21a, b).

Usually the Ψ function is formulated from Molecular Orbital (MO) theory, where one-electron orbitals are used to approximate the true many-electron wavefunction. It is assumed that each molecular orbital is a truncated linear combination of n one-electron functions, known as the basis set, and are primarily atomic orbitals. This approach is known as the linear combination of atomic orbital (LCAO) method (23a, 23b). The actual functional form of the wavefunction Ψ determines the complexity of the calculation. Boys (24) originally suggested that Gaussian Type (GT) functions may be used to describe Slater Type (ST) functions, which in turn are approximations of the true hydrogenic orbitals. The use of GT functions greatly reduces computational time because they can be solved explicitly. Since GT functions, unlike ST orbitals, do not have a cusp at the origin, they are less accurate. In addition, as distance from the origin increases the GT functions decay too quickly. Nevertheless, despite these limitations, a linear combination of GT functions can be made to fit an ST function.

Pople found that a minimal basis set was composed of a linear combination of three GT functions; they are referred to as STO-3G (25a, 25b, 25c, 25d). Pople also found that more accurate representations may be achieved with more sophisticated GT combinations. The next level of theory splits the outer shell basis functions into two parts, appropriately known as split-level basis sets. In the early 1990s, 3-21G calculations became the new standard (26a, 26b, 26c), replacing the STO-3G formulation. Presently, more complex basis sets have become the new standard. Additional accuracy was achieved by adding polarization functions, which may be denoted by (*). Other conventions are also used: for example, 6-31G* calculations, the new standard, treat the core electrons with six primitive basis functions (27a, 27b, 27c). The outer shell electrons are divided into two parts, which are composed of a set of three GT functions and one GT function, respectively. Finally, a set of six d polarization functions has been added for non-first-row elements (21, 28). The split basis sets and polarization functions add flexibility to the molecule such that the atomic orbitals can more accurately respond to their environment. Polarization of hydrogen atoms is achieved by adding a set of three p orbitals, denoted by a second * (e.g., 6-31G**). The next level of accuracy divides the outer electrons into three sets of Gaussians (29). With the advances in computers, 6-311G* calculations are more economical and becoming more popular.

Quantum mechanical calculations are carried out using the Variational theorem and the Hartree-Fock-Roothaan equations (13a, 13b, 21, 30a, 30b, 30c). Solution of the Hartree-Fock-Roothaan equations must be carried out in an iterative fashion. This procedure has been called self-consistent field (SCF) theory, because each electron is calculated as interacting with a general field of all the other electrons. This process underestimates the electron correlation. In nature, electronic motion is correlated such that electrons avoid one another. There are perturbation procedures whereby one may carry out post-Hartree-Fock calculations to take electron correlation effects into account (31a, 31b, 31c, 31d, 31e). It is generally agreed that electron correlation gives more accurate results, particularly in terms of energy.

- c. *Density functional theory*: Density functional theory (DFT) is the third alternative quantum mechanics method for obtaining chemical structures and their associated energies (32). Unlike the other two

approaches, however, DFT avoids working with the many-electron wavefunction. DFT focuses on the direct use of electron densities $P(r)$, which are included in the fundamental mathematical formulations, the Kohn–Sham equations, which define the basis for this method (33). Unlike Hartree–Fock methods of *ab initio* theory, DFT explicitly takes electron correlation into account. This means that DFT should give results comparable to the standard *ab initio* correlation models, such as second order Møller–Plesset (MP2) theory.

In some ways, DFT may be more easily understood. According to the Kohn–Hohenberg theorem (34), the energy is minimized when the calculated and true electron densities are equal. One important consequence of DFT calculations is that their accuracy corresponds to standard post Hartree–Fock methods. Another consequence of DFT calculations of more practical importance is the reduction in the computation time required to complete a calculation. The time required to complete Hartree–Fock calculations is a function of the number of electrons in the system being examined, and it is proportional to n^4 , where n represents the number of electrons. DFT calculations, also a function of the number of electrons in the system, are proportional to n^3 . (Contrast the time for quantum chemical calculations with those of molecular mechanics calculations, discussed later. In molecular mechanics, the time scales to m^2 , where m is the number of atoms, not the number of electrons. Molecular mechanics is an attractive and accurate alternative for larger molecular systems where it is impossible to use quantum mechanics based methods.) For systems of interest to medicinal chemists, the differences in computer time between DFT or *ab initio* calculations may be the most important factor in selecting which method is used. Although there are more examples of *ab initio* calculations in the literature and consequently more experience with their accuracy and limitations, there is a growing number of DFT calculations being reported.

Molecular mechanics

History and concepts: A complementary approach for molecular structure calculations is available, and it is referred to as the molecular mechanics or force field method; it is also known as the Westheimer method (35). In 1946, twenty years after the impressive development of quantum theory, three papers appeared in the literature which applied *classical* mechanical concepts to problems of chemical interest (36–38). Westheimer investigated the racemization of some optically active biphenyl

derivatives. His work demonstrated the potential usefulness of molecular mechanics. The other two papers were attempts to tackle more complex problems.

The fundamental concepts of molecular mechanics are deeply rooted in traditional chemical thought, which takes the view that molecules may be considered as a series of masses held together by elastic or harmonic forces (somewhat like balls fastened together by weightless springs). The idea central to molecular mechanics theory, concisely stated, is that any deformation of the ball-and-spring model from its *natural* bond lengths (l_0) and angles (θ_0) results in a strain. These deviations from geometric ideality are reflected in a corresponding increase in energy for the model. A set of classical equations are used to describe the motion and corresponding energy of the spring. Hence, with the wealth of chemical information found in the literature, parametrization of these *classical* equations results in the accurate reflection of molecular behavior—that is to say, the potential used to describe the energy surface are generally referred to as the force field. In physics the term *field* traditionally is given to the continuous distribution of some “condition” prevailing through space.

An isolated molecule may presumably adopt various conformations in response to specific intramolecular forces. In general, the total energy of a molecule (E_{total}) may be regarded as the summation of stretching (E_s), bending (E_b), torsional (E_{tor}), van der Waals (E_{vdw}), and electrostatic (E_{ele}) energy components. For more advanced force fields it is crucial for structural accuracy to include various cross-terms ($E_{\text{cross terms}}$), e.g., stretch–bend, torsion–stretch, and torsion–bend, etc., are added to reflect the fact that motions and interactions are coupled to one another. Additional terms may be added, such as hydrogen bonding potentials. One of the fundamental theorems of molecular mechanics is the division of the total energy into separate, readily identifiable parts, as indicated by Equation (5). As discussed previously, deviations from the *natural* geometry is accompanied by an increase in the total energy (E_{total}) of the molecule. The total energy may be viewed as a summation of various energy terms that medicinal scientists feel comfortable in using an visualizing (Eq. 5). A second tenet of molecular mechanics theory concerns the transferability of the parameters used in the individual potential energy terms of Eq. 4. In other words, the parameters used are transferable from one molecule to the next. The continued success of molecular mechanics in treating a wide variety of compounds has justified these assumptions. When several molecular structures are to

be considered together, however, one must obviously think about intermolecular forces as well.

$$E_{total} = E_s + E_b + E_{tor} + E_{vdw} + E_{elec} + E_{cross\ terms} + \dots \quad (5)$$

Some theoretical purists tend to view molecular mechanics calculations as merely a collection of empirical equations or as an interpolative recipe that has very little theoretical justification (39). It should be understood, however, that molecular mechanics is not an ad hoc approach (35). As previously described, the Born–Oppenheimer approximation allows the division of the Schrödinger equation into electronic and nuclear parts, which allows one to study the motions of electrons and nuclei independently. From the molecular mechanics perspective, the positions of the nuclei are solved explicitly via Eq. 2. Whereas in quantum mechanics one solves, Ψ , which describes the electronic behavior, in molecular mechanics one explicitly focuses on the various atomic interactions. The electronic system is implicitly taken into account through judicious parametrization of the carefully selected potential energy functions.

Methodology: The fundamental molecular mechanics formulation may be derived from the expansion of the potential energy function $E(q_1, q_2, q_3, \dots, q_1)$ in a Taylor series about the equilibrium position q_0 , where q represents generalized coordinates (40), yielding Eq. 6:

$$E = E_0 + \sum_{i=1}^n \left(\frac{\partial E}{\partial q_i} \right)_0 \Delta q_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m \left(\frac{\partial^2 E}{\partial q_i \partial q_j} \right)_0 \Delta q_i \Delta q_j + \frac{1}{6} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^o \left(\frac{\partial^3 E}{\partial q_i \partial q_j \partial q_k} \right)_0 \Delta q_i \Delta q_j \Delta q_k + \dots \quad (6)$$

The first term in the expansion is a constant. Since there is the freedom to select a point of zero potential energy E_0 can be defined as zero. The second term is the negative of the restoring force, which is equal to zero at the equilibrium position (Eq. 7):

$$-F_i = \left(\frac{\partial E}{\partial q_i} \right)_0 = 0 \quad (7)$$

Hence, the first nonvanishing term in the Taylor series expansion is the third term, which is a quadratic expression. This third nonvanishing term corresponds to a Hooke's law potential in the limit of small vibrations. The series may be truncated after the third term to provide the following potential energy expressions (Eqs. 8–10):

$$E_s = \frac{1}{2} \sum_{i=1}^n k_q (q_i - q_0)^2 \quad (8)$$

$$E_\theta = \frac{1}{2} \sum_{i=1}^n k_\theta (\theta_i - \theta_0)^2 \quad (9)$$

Additionally, cross terms such as the stretch–bend potential, shown below in Eq. 10, may be included in the force field equation to give better agreement between experiment and calculation.

$$E_{\text{stretch-bend}} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m k_{\theta} (q_i - q_0)(\theta_j - \theta_0) \quad (10)$$

The stretching, bending, and various cross-term potentials [specifically illustrated with a stretch–bend effect in Equation (10) above] naturally arise from the Taylor series expansion. Cross-terms have been successfully added to various force fields to give more structural flexibility. For example, when intramolecular angles are compressed, experimental studies indicate that the bonds stretch to relieve the strain resulting from the angle bending. Mathematically, in a valence bond force field, this may be incorporated into a force field potential through cross terms. In the case of angle compression and bond stretching, a stretch–bend term may be introduced (see Eq. 10). There are other ways to treat these effects. Explicit terms between 1,3 nonbonded atoms may also be incorporated, and are called Urey–Bradley force fields. The MM2, MM3, and MM4 force fields (41a, 41b, 41c, 41d, 41e, 41f, 41g, 41h, 41i, 41j, 41k) uses stretch–bend cross terms to mimic this type of molecular behavior. In the latest formulation of molecular mechanics treatments, Allinger and co-workers have incorporated a torsion stretch term to represent quantitatively the bond stretching that occurs in strained molecules. It has been shown that some cross terms are more important than others, which means that some can be neglected without drastically affecting the calculated structures. Additionally, certain cross terms are known to affect calculated vibrational spectra.

Higher terms from Eq. 6 may also be added to the simple quadratics to furnish better representations of chemical bonding. Nevertheless, these energy terms alone do not adequately describe molecular behavior.

The third energy term of Eq. 5 arises from pairwise nonbonded interactions summed over every pair i and j . Two opposing forces play a role in this energy term. For relatively long distances, two atoms attract each other due to London dispersion interactions, which are proportional to r^{-6} . As the two atoms come into close proximity, they

exert a mutual repulsion known as van der Waals repulsion. the Leonard-Jones 6–12 potential function [Eq. 11] describes this behavior (42), originally derived for the noble gases, and is most often used in simple force fields. Quantum mechanical calculations indicate that the repulsion part of the curve may in fact be somewhat overestimated. Some force fields use a 6–10 potential to reduce the steepness or hardness of the curve as two atoms approach one another, which is perhaps most commonly used for cases of hydrogen bonding. The use of an exponential term instead of the r^{-12} term in force field equations better reproduces experimental data for organic structures and is more consistent with quantum chemical calculations. Potential exponential functions have been known for some time. In fact, MM2, MM3, and MM4 use a modified Hill equation to determine the nonbonded interactions.

$$E_{vdw} = \sum_{ij} \sum \left[\frac{A}{r^{12}} - \frac{B}{r^6} \right] \quad (11)$$

It has long been known that molecules do not rotate freely. In 1891, Bischoff proposed that ethane preferred a staggered conformation and that restricted rotation occurred in substituted ethanes (43). Christie and Kenner first demonstrated restricted rotation in 1922 by resolving 2,2'-dinitrophenyl-6,6'-dicarboxylic acid into optically active isomers (44). Pitzer showed that the calculated and observed entropies for ethane were identical if restricted rotation was considered (45a, 45b). The phenomenon of restricted rotation appears to be ubiquitous and has stimulated intense interest and research.

There has been much speculation concerning the origin of the internal energy barrier (46). Three explanations which have been proposed: 1) repulsion between the bonds not covered by the van der Waals repulsion; 2) correction

for the anisotropy of van der Waals repulsion; and 3) various quantum mechanical suggestions.

The barrier to internal rotation must be included in the force field. Usually, the internal rotational energy is a function of the torsion angle ω . (See Fig. 1.) Klyne and Prelog have defined the torsional angle ω for the connected atoms A—C—C—D, as depicted previously. Rotation of A towards D along the shortest arc, viewing the molecule along the C—C axis, is defined as a positive value when clockwise.

The torsional potential may be expressed as a truncated Fourier series, as shown in Eq. 12, where the summations is over all n unique bond dihedrals (torsions), ω . Usually, the first three terms are sufficient to describe the torsional energy adequately. Higher terms may improve the fit between experimental data and calculation. Inclusion of V_1 and V_2 terms has improved the agreement between experiment and calculations (35a, 47b, 47c). Radom, Hehre, and Pople (48) had given a physical explanation to the torsional terms. The onefold term, V_1 , is a dipole–dipole interaction. The twofold term, V_2 , arises from hyperconjugation. The threefold term, V_3 , has a steric origin.

$$E_{\text{tor}} = \sum_{i=1}^n \left[\frac{1}{2} V_{i,1} (1 + \cos \omega) + \frac{1}{2} V_{i,2} (1 - \cos 2\omega) + \frac{1}{2} V_{i,3} (1 + \cos 3\omega) + \dots \right]$$

Other related Fourier expressions may include phase angles, which may also be used to describe potential energy curves; this is the approach employed in AMBER (49a, 49b, 49c, 49d). See Eq. 13, where ϕ is the dihedral angle, V^ϕ is the force constant, n is the multiplicity, and δ is the phase angle. These Fourier series expressions allow the incorporation of additional experimental or quantum mechanical phenomena into force field calculations.

$$E_{\text{tor}} = \sum_{i=1}^n V_\phi [1 + \cos(n\phi - \delta)] \quad (13)$$

Electrostatic interactions play a crucial role in the binding associated with drug–receptor interactions. The electrostatic part of any molecular mechanics potential energy expression is extremely important for many types of calculations beyond drug–receptor considerations. Interestingly, consensus opinion on the best formulation has not been reached. It appears that two alternative, but ultimately equivalent, representations may be used: 1) point charge–point charge and 2) dipole–dipole

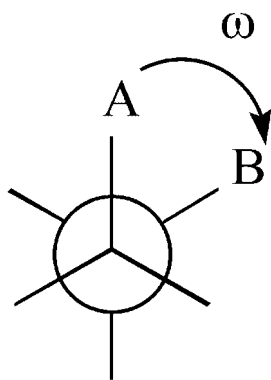


Fig. 1 Dihedral angle representation.

interaction models. Most force fields utilize the former, explicitly assigning point charges to each atom in a molecule and relying on Coulombic interaction terms, where the energy varies between two point charges, q_1 and q_2 , by r^{-1} . Mathematically, this formulation requires the evaluation of square root terms. Originally when applied to large biopolymers, the electrostatic calculations were quite time consuming. An approximation which successfully circumvents this problem sets the dielectric constant equal to r (50a, 50b): With this stratagem, q_1 and q_2 vary as r^{-2} , and this in effect dampens the charge interactions. The Allinger force fields and the many variants treat electrostatic interactions with a dipole–dipole interaction equation, where the dipole energies vary by r^{-3} . For charged structures, there must be dipole–charge and charge–charge terms. The charges or dipoles are not static, and the recipes are influenced by neighboring groups. Ideally, terms or methods which consider the polarizability of groups are increasingly being used (51). Progress over the years continues.

The treatment of charges in molecular mechanics calculations remains a central question. The exact computational scheme used to obtain charges has been under considerable review. One obvious way is to obtain the charges from quantum mechanics calculations. Since quantum mechanically derived point charges are based on various population analyses, the level of accuracy required and how the charges should be distributed between the atoms have not been universally accepted. One final check on the charge distribution in a molecule is the comparison of the calculated dipole moment with experimentally determined dipole moments. If there is a reasonable charge distribution, based on calculated charges, then the calculated and experimental dipole moments should be in close agreement.

As indicated earlier, molecular mechanics has also been described as the empirical method. Most of the parameters and equations are designed to reproduce experimental data (for example, molecular geometry, dipole moments, and heats of formation). Perhaps the term *empirical method* is becoming outdated. In many cases, high level ab initio (or DFT) calculations are being used to augment sketchy or nonexistent experimental data such as charge distribution, conformational energies, and rotational barriers (52a, 52b, 52c, 52d, 52e, 52f). Recent work by several groups has used a combination of experimental and theoretical work to parameterize transition state force fields (53a, 53b, 53c, 53d, 53e).

Geometry optimization: Once the three-dimensional Cartesian coordinates and necessary parameters (such as atom and/or bond types) have been established for a molecular structure, the next step is to determine the

energy-optimized structure. Various molecular mechanics programs have geometry optimization algorithms built into them, and they are capable of locating an atomic arrangement at an energy minimum. The initial structure has an energy associated with it based on the particular force field. Every atom has a net force acting upon it. The objective is to relieve as many of the forces as possible, such that the derivative of the potential energy equation is zero.

A variety of methods have been developed for these purposes. Most of the modern methods involve differentiation of the analytical functions, whereas earlier methods used numerical techniques to reach the lowest point (bottom) of the potential energy well. These methods assume the potential energy function is well behaved and differentiable. On the one hand, the simplest approach is called *steepest descent* (54a, 54b, 54c). In this method, displacement steps k_i are taken in the direction parallel to the net force acting on the atom. Since the displacement vector is parallel to the force, which in turn is equal to the negative of the gradient of the potential, the direction is straight downhill. Steepest descent is relatively efficient when the geometry is poor, i.e., far from the minimum. As the minimum is approached, however, the derivatives become smaller and smaller. Consequently this algorithm has reduced efficiency in these cases.

On the other hand, conjugate gradient methods (55a, 55b) are more effective in locating the minimum energy structure. In this approach previous optimization information is utilized. The second and all subsequent descent directions are linear combinations of the previous direction and the current negative gradient of the potential function. Conjugate gradient methods are generally more efficient than steepest descent. In general, for an N -dimensional quadratic surface, the minimum will be located in N steps.

The Newton–Raphson approach is another minimization method (56a, 56b). It is assumed that the energy surface near the minimum can be described by a quadratic function. In the Newton–Raphson procedure the second derivative or **F** matrix needs to be inverted and is then used to determine the new atomic coordinates. **F** matrix inversion makes the Newton–Raphson method computationally demanding. Simplifying approximations for the **F** matrix inversion have been helpful. In the MM2 program, a modified block diagonal Newton–Raphson procedure is incorporated, whereas a full Newton–Raphson method is available in MM3 and MM4. The use of the full Newton–Raphson method is necessary for the calculation of vibrational spectra. Many commercially available packages offer a variety of methods for geometry optimization.

In many cases, low-energy conformations are sought, although the drugs bound to receptors are required (and usually are not) in their lowest energy state for the isolated molecule. All of the algorithms, described above, have been almost exclusively designed to find minima, but not necessarily the global minimum. When an energy optimization has been carried out for a complex molecule, it is impossible to tell whether or not the absolute lowest energy conformation has been calculated. In general, programs are not designed to go over energy barriers. The energy-minimized structure depends on the starting coordinates initially used. So if the coordinates corresponding to the boat conformation of cyclohexane were used as the starting point, then the force field optimizer would seek out the closest energy well. In this case, the final geometry would correspond to a twist-boat structure, a minimum, but not the chair form, the global minimum.

As indicated, there is no assurance that the *bioactive* conformation is necessarily the *global* minimum or, for that matter, a minimum at all (57a, 57b, 57c). Nevertheless, the simplifying assumption is that minimum energy conformations are at the very least good starting points for examining potential drug candidates. To alleviate the concern of not finding the global minimum, automated conformational search procedures have been devised. In general, one or more bonds may be rotated and evaluated based on some criterion, such as energy differences and/or van der Waals interactions. From these searches, a series of compounds can be saved in a computer file, retrieved later, and then subjected to energy optimization procedures. The problem with conformational searching techniques is that they are computationally demanding; and, if the dihedral angle increment is too large, one could miss some low-energy conformers. The use of conformational searching formed the basis of the active analog approach championed by Marshall (57a, 57b, 57c).

Molecular simulations

Dynamics: Molecular mechanics calculations treat molecules as static, time-averaged structures. Since there is an exponential growth in the number of conformations with increasing degrees of freedom in a molecule, a systematic search of the configuration space for the low-energy conformations of large molecules is virtually impossible. Another powerful computational method which has been applied to conformational problems is molecular dynamics (58a, 58b, 58c). Molecular dynamics (MD) simulations describe the trajectory (configurations as a function of time) of a system by integration of Newton's equations of motion, where the force on particle

i is represented as F_i . The phase space trajectories generated with MD calculations are analyzed to yield equilibrium and dynamical information. See Equations (14a, 14b, 14c, 14d, 14e, 14f, 14g) and (15), where r represents the position and E the potential energy.

$$\frac{d^2 r_i}{dt^2} = m_i^{-1} F_i \quad (14)$$

$$F_i = -\frac{\partial}{\partial r_i} E(r_1, r_2, r_3, \dots, r_N) \quad (15)$$

This technique, provided a sufficiently good force field has been developed, has certain advantages over molecular mechanics calculations by themselves in that solvation interactions may be determined and different energy minima can be located.

Monte Carlo simulations: The use of Monte Carlo (MC) methods (59) which incorporate the metropolis algorithm has been used the basis for conformational searching (60), drug-receptor docking (61), and free energy perturbation calculations (see next section) (62). MC is based on the principles of statistical mechanics, where an ensemble of molecular states are generated and evaluated. The metropolis MC method for finding available conformations are summarized in the following steps. First, the starting molecular geometry is calculated using a suitable molecular mechanics potential. Second, a small random perturbation is applied, i.e., one of the atoms is kicked from its original position to a new position. Third, a new energy is calculated. If the energy is lower, then the move is accepted, and the new configuration or conformation (which depends on whether the solvent or solute is being perturbed) is stored for later use. If the energy is higher there is a probability that it may or may not be rejected. The probability is based on a Boltzmann factor ($e^{-\Delta E/kT}$), where ΔE is the energy difference between the final and original states ($E_2 - E_1$), k is the Boltzmann constant, and T is the temperature. It is necessary to accept some higher-energy-state configurations or conformations to avoid becoming trapped in a local minimum. The procedure described is repeated for many iterations until the lowest free energy is achieved. MC, like MD simulations, can be used to treat solvation.

Free energy perturbations: One of the challenges facing medicinal chemists is to predict (and ultimately prepare and test) compounds with high affinity and efficacy for target receptor sites. Knowing the free energy of binding, $\Delta G_{\text{binding}}^\circ$ for a series of structurally related compounds is important information. Unfortunately, determining the $\Delta\Delta G_{\text{binding}}^\circ$ for many different analogs is experimentally demanding. Free energy perturbation

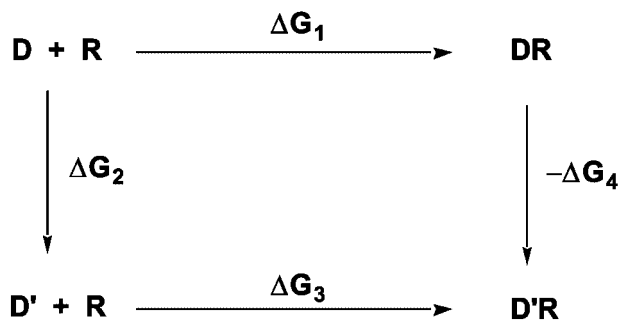


Fig. 2 Free energy perturbation (FEP) diagram, showing drugs D and D' binding to receptor R to give the DR and DR' receptor complex and the free energy changes.

(FEP) calculations (63a, 63b) take advantage of the fact that the change in a complete thermodynamic cycle should be zero (see Fig. 1). By measuring the equilibrium constants K_2 and K_1 , the experimental $\Delta\Delta G_{binding}^0$ can be obtained according to Eq. 16. Inspection of Fig. 2, however, reveals that if D_1 could be mutated to D_2 and DR_1 could be mutated into DR_2 , we would have a relationship where knowing the difference between $\Delta^*G_4^0 - \Delta G_3^0$ would be equivalent to $\Delta G_2^0 - \Delta G_1^0$ (Eq. 17). The former relationship can be simulated using either MD or MC methods.

$$\Delta\Delta G_{binding}^0 = \Delta G_2^0 - \Delta G_1^0 = -RT \ln \frac{K_2}{K_1} \quad (16)$$

$$\Delta G_2^0 - \Delta G_1^0 = \Delta G_4^0 - \Delta G_3^0 \quad (17)$$

FEP is a major conceptual achievement with important ramifications for drug design. Unfortunately, two major problems have prevented FEP from being routinely applied to drug design outside of academic research. First, the mutations must be small and gradual. Calculations need to be carried out during the mutations process. Second, the calculations are computationally demanding even on high speed workstations. Nevertheless, it is an attractive computational approach.

COMPUTER-ASSISTED DRUG DESIGN

Computational chemistry applied to drug design relies on two major experimental sources of detailed three-dimensional (3-D) data: 1) X-ray crystallographic analysis and 2) nuclear magnetic resonance (NMR) spectroscopy.

High precision X-ray crystallographic studies now are reported routinely in the literature for small molecules and for proteins and nucleic acids (64). These data provide

spatial locations (3-D coordinates) for the atoms in a molecular structure. Since X-rays are diffracted by electron density, hydrogen atoms are not generally located with the same accuracy as heavy atoms. In general, the crystallographer is able to provide complete structures. In protein X-ray crystallography the knowledge of the amino acid sequence of the protein assists the crystallographer in developing a model of the atomic positions and bonding (65). Backbone traces based on three-dimensional electron density make the viewing easier for macromolecular structures. Interestingly, computer-based approaches are also used to fit experimentally determined electron density maps to proposed structures.

NMR techniques are also being used to determine 3-D macromolecular conformation. High-resolution NMR combined with pulse experiments can provide connectivity information along a protein chain. Nuclear overhauser effect (NOE) signals are sensitive to the through-space distances between atoms, and they can be used to determine the folding pattern in a protein. The combination of connectivity data, NOE-determined distance constraints, and amino acid sequence information provides a molecular modeler with a substantial set of geometric constraints for building models of macromolecules. These constraints can be combined to build a macromolecular model using the mathematical technique of distance geometry.

Distance geometry provides sets of 3-D structures of a protein or nucleic acid that fulfill the constraints. The combination of distance geometry, for generation of molecular starting points, with molecular dynamics computations can yield 3-D models of small proteins with precision equal to X-ray crystallography (66). This combination of NMR, molecular mechanics, and molecular dynamics (67) can be used to provide a three-dimensional protein structure in a situation where the protein cannot be crystallized or the crystals are not appropriate for X-ray crystallography.

The term CADD has been used to describe two aspects of the recent use of computational tools that aid computational and medicinal chemists in the search for new drug candidates. In the first approach, medicinal chemists attempt to describe the predominant statistical correlation of biological activity with directly measurable physico-chemical parameters or characteristics of drugs and is known as Quantitative Structure-Activity Relationships (QSAR) (68). The central idea is that compounds exhibit biological activity based on structural characteristics. It should then be possible to correlate the associated biological activity with various critical parameters. In general, the biological activity may be considered a function of hydrophobicity, electrostatics, and steric forces (Eq. 18).

$$\begin{aligned}\text{Biological Activity} = & f(\text{hydrophobicity}) \\ & + f(\text{electrostatics}) + f(\text{sterics})\end{aligned}\quad (18)$$

The seminal work of Corwin Hansch initiated the field of QSAR (69). Two-dimensional and three-dimensional QSAR methods (2-D QSAR and 3-D QSAR) have been widely applied to problems of biological interest. The latter approach has increased in popularity with the introduction of the comparative molecular field analysis (CoMFA) (70) and commercial availability of similar methods.

The second aspect of CADD relates more closely to computer modeling. This approach can be characterized as pharmacophore mapping, which uses structural information from receptors or small molecules (in combination with computational chemistry techniques) to develop new candidate drug targets. A pharmacophore, first defined by Paul Erlich, is the set of essential molecular features in a compound responsible for a specific pharmacological action. This approach has also been called receptor-based drug design, using a very general definition of the word receptor (71), which broadens the definition of classical pharmacology. In this technique, a receptor or a model of a receptor is used as the target for the drug design efforts. This receptor model can be an X-ray crystallographic structure or a model built from the X-ray structure of a homologous protein. These can be viewed as recognition sites, since it could be argued that the receptor *recognizes* certain patterns of geometry and charge density of a drug as it approaches and binds to its receptor. As improved computational and X-ray structural tools

become available, modeling of pharmacophores in 2-D as well as in 3-D, also an old technique, has become a more commonly used and much more sophisticated approach.

Some early pharmacophore models contained sets of hexagonal rings (72) presented as a single plane. These rings were used to describe, in a 2-D fashion, the relative 3-D relationship of receptor functional groups. If no receptor surfaces are known to exist in a specific orientation, this area is called a "bulk tolerance area" of the unknown, but complex, 3-D area of an enzyme's active site (73). Another later example of pharmacophoric models is the triangular N O O hypothesis which Cheng (74a, 74b) used to describe, in simple geometric terms, the conformation of active antitumor agents. This hypothesis used early results from X-ray crystal studies on small molecules and measurements from physical molecular models to suggest the required geometry of active antitumor agents. Through-space distances were used to connect a basic nitrogen with two hydrogen-bond-donating oxygen groups. This pharmacophore model was more sophisticated than most previous models because it was developed with three-dimensional data.

More recently CADD software, for example the Catalyst program from MSI (www.msi.com) (75, 75), is available for pharmacophore modeling and searching of 3-D databases. The use of such a program, only one of many with similar power, shows the increasing sophistication of computational search strategies for drug design and discovery. The input to this program is a structure activity set of molecules. The program then analyzes the set of molecules and generates a set of pharmacophoric hypotheses that can be represented in 3-D. The

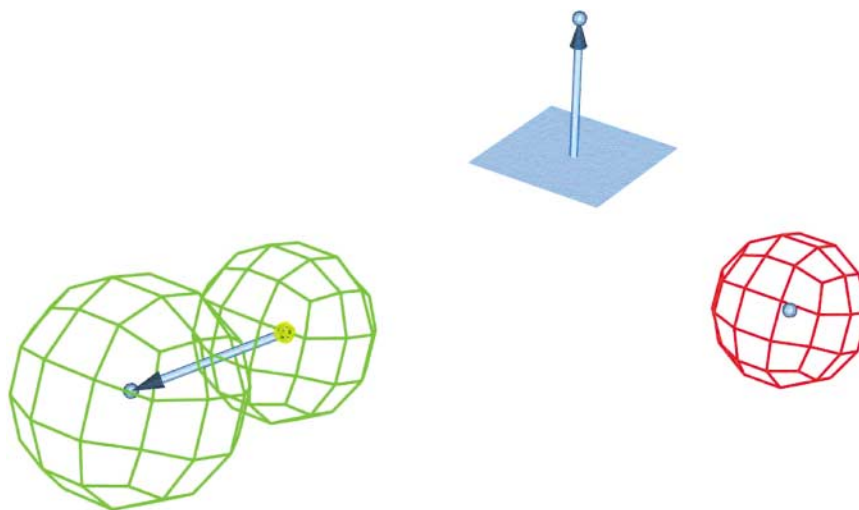


Fig. 3 Three-dimensional pharmacophore generated in catalyst.

pharmacophoric hypotheses can then be used by catalyst to search databases of molecule for pharmacophore matches.

Fig. 3 represents an example of a 3-D pharmacophoric hypothesis as generated in catalyst. This is a 2-D snapshot of the 3-D pharmacophores. They would have inter-pharmacophore distances associated with the computer data object. The blue plane represents the location of a plane of an aromatic ring with the necessary perpendicular vector for orienting aromatic ring systems in new molecules. The smaller of the green spheres represents a hydrogen bond acceptor, sized at 1.7 Å, and the larger sphere representing the projected point and locations for a possible hydrogen atom donating group with a sphere of 2.3 Å. The red sphere represents a positively charged group with a radius of 1.5 Å. This type of pharmacophore hypothesis can then be used as a 3-D search query that, in many cases has proven useful in finding molecules related to the initial input set.

As previously mentioned, X-ray crystallography provides the bulk of the 3-D information used by medicinal and computational chemists in structure based drug design. Two research-based organizations (76, 77) have compiled and provided a subscription service to databases which are the primary source for X-ray structural data. Besides the information contained in the databases, the existence of these organizations and defined computer formats for crystallographic data has the advantage of providing the scientist with some standardization. Commercial molecular modeling software packages provide a transparent interface between graphics display programs and these data. The two databases divide their holdings on the basis of the size of the molecule. Large polynucleotide crystal structures and structures with drugs bound to polynucleotides are stored in both the protein data bank (PDB <http://www.rcsb.org/pdb/>) and Cambridge files, with the division basically being subjective. Generally, the larger structures are in the PDBdatabase.

The Cambridge Crystallographic Data Center (<http://www.ccdc.cam.ac.uk/>) (76) now provides a computerized database of more than 200,000 small molecular structures. In addition to this extensive structural database, the Cambridge group has developed a suite of programs for systematic structure search and retrieval. The most unique aspect of these programs is that database search queries can specify desired structures in three dimensions.

With the rapid advances in macromolecule crystallography, the PDB (<http://www.rcsb.org/pdb/>) (77) currently contains about 12,600 macromolecular NMR or crystal structures. While this may not represent all macromolecular structures, it does represent the large majority. Most of the major scientific journals require simultaneous deposition of macromolecular crystal data into the PDB as the journal article is published. Besides structural proteins, and the classic crystal structures such as hemoglobin and lysozyme, the data base includes many enzymes or representative enzymes from classes that could be good targets for chemotherapy.

Table 1 shows a realistic overview of the successes of CADD in the pharmaceutical industry. Target proteins have been vigorously explored as indicated by the number of structures in the PDB. The only criteria of success that can be used in this situation is to measure compounds put into clinical trials, since about 80% of candidate drugs fail in clinical trials many of these compounds may not end up on the market.

Automated X-ray diffractometers integrated with fast computer workstations are commercially available for rapid acquisition of protein diffraction data. Area detectors for X-ray diffraction allow for rapid collection of much larger quantities of reflection data and provide the ability to find structures for materials that were too unstable to withstand the older, slower techniques. New automated software routines that use direct methods allow companies to provide a commercial service solving crystal structures. Increasing use of synchrotrons around the world has

Table 1 Overview of the success of CADD

Target (# in PDB)	Research group	Method	Outcome
Thrombin (105)	Hoffman-La Roche	X-ray, models (2.2 Å)	Clinical candidate
Thrombin	Biogen Inc.	X-ray, models	Phase III
Neuraminidase (53)	Monash University/GW	GRID search, X-ray (2.4 Å)	Phase II*
Purine nucleoside phosphorylase (16)	Biocryst	Molecular mechanics X-ray (1.5–2.0 Å)	Phase II
Thymidylate synthase (66)	Agouron	Modeling X-ray (1.9–2.5 Å)	Clinical candidate
Carbonic Anhydrase (111)	Merck	Multiple X-ray (1.9–2.5 Å)	Market
Rhinovirus-14 (28)	Sterling Winthrop	Screening X-ray (2.9 Å)	Phase I
Aldose reductase (13)	Ayerst	Molecular orbital QSAR (1.8–2.3 Å)	Market

allowed rapid collection of X-ray data. The technology has expanded widely into the pharmaceutical industry. Consortia such as IMCA provide drug discovery with a significant resource for crystallography. (<http://www.imca.aps.anl.gov/>)

STATE-OF-THE-ART COMPUTER-ASSISTED DRUG DESIGN

Theory

Most recently an interesting model for approaching CADD has been presented (78), where a set of guidelines for the practical application of CADD techniques are provided. It relates the number of compounds that need to be investigated in a drug-discovery project to the amount of information available about the drug target. More information, available about the target, means that fewer compounds must be investigated in the drug design process and also changes the techniques used by the computational chemist. This is illustrated graphically in Fig. 4.

On the one hand, if the protein X-ray crystallographic structure has been characterized, then the techniques of CADD are used directly. This would involve use of the crystal structure as a target of design efforts using the structural information. That initial effort would be followed by solving additional structures with new, more potent, ligands bound to the protein. The bound structures are modified, graphically, and then evaluated using methods previously discussed. Only compounds that show promise as possible ligands are actually prepared. On the other hand, if there is less information about the target available: 1) a protein sequence and no characterized structure or 2) a set of compounds with different degrees of biological activity. Then the pharmacophoric model and hypothesis approach, discussed above, can be used. A 3-D pharmacophoric model building effort, using

catalyst might develop pharmacophore models that can be used to search libraries of structures.

At the other extreme if you have no knowledge of the protein structure, no candidate ligands, and the X-ray structure is not yet known then a high throughput screening of primary large scale collections of compounds or combinatorial libraries would be screened. These techniques would use roboticized and miniaturized assays and be done to search for possible binding compounds.

Methodology

Design of enzyme inhibitors guided by molecular modeling and X-ray crystallography

An excellent, and successful, structure based drug design example in a system where there is significant structural information is work on the development of new inhibitors of HIV protease (79a, 79b). This work was carried out by scientists from Dupont Merck Pharmaceuticals. A substantial amount of structural data has been available for a decade on the structure of HIV protease (80).

HIV protease is one of the proteins coded by the HIV viral genome and expressed as part of the reproductive cycle of the virus. HIV protease contains active aspartic acid residues that classify it as an aspartyl protease. The viral genome once incorporated into the host genome codes for a polyprotein. The protease is responsible for cleaving specific sites in the polyprotein into specific proteins that allow the virus to mature. The HIV protease is a symmetrical dimer of with each monomer containing 99 amino acids. Fig. 5 shows a cartoon of the protease structure with the active site region labeled. This region is responsible for the binding of the protein substrate sequence. It recognizes and binds a hexapeptide and hydrolyzes the labile central peptide bond.

Fig. 6 shows the binding mode of an HIV protease inhibitor from a X-ray study (file 4phv). In this figure two amino acids, glycine 49 (GLY49) and isoleucine 50 (ILE50) of the HIV protease, are shown in stick form. The remainder of the protein-active site is in line form. Since the enzyme is a symmetrical dimer, the residues from each dimer are shown although only one pair of residues is labeled. Also shown is a water molecule bound in the active site of the enzyme. It is believed that this water molecule assists in the peptide bond hydrolysis. The protein NH groups bind to this water molecule. Carbonyl groups of the inhibitor also interact with this water molecule. Additionally this crystal structure shows the position of the peptidomimetic, HIV protease inhibitor, **1**, and the manner in which the four aromatic ring systems fit into the enzyme.

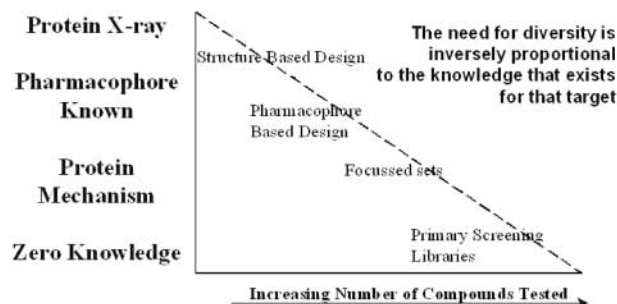


Fig. 4 Plot of relationship between information and number of compounds necessary for synthesis.

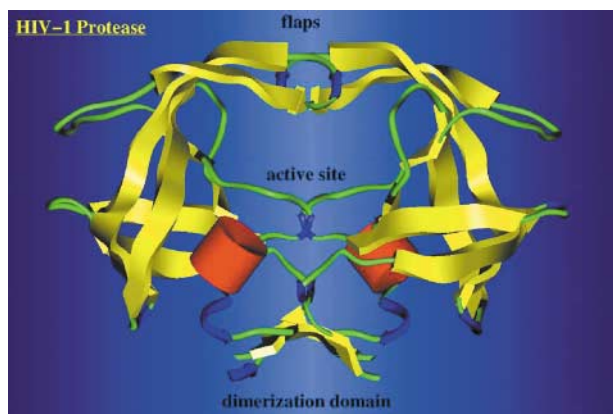
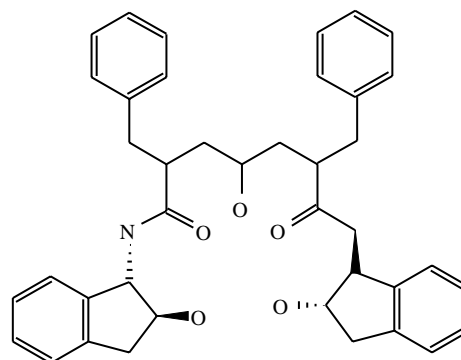


Fig. 5 HIV protease structure with active site region labeled.

The second structure in Fig. 6 represents a similar view of the crystal structure (file lajx). In this structure-based design, a cyclic urea 2 was synthesized in which it was hypothesized that the water molecule could be replaced by a functional group of the inhibitor (81). In the case of compound 2, the oxygen of the water was replaced by the oxygen of a carbonyl group that interacts with the backbone NH groups just as the water molecule does. The four aromatic groups of the cyclic urea structure.

Equipment

Currently, molecular modeling hardware is concentrated on two groups of computer systems. High-end



Structure 1.

workstations with high-resolution graphics display systems for molecular modeling are primarily Silicon Graphics models, running the Silicon Graphics version of the UNIX operating system. These systems have held predominance in the market for about a decade. A rapidly growing competitor is the standard Intel chip powered, PC running Windows or Windows/NT. The vast majority of the pharmaceutical industry has supported Windows/NT desktops for scientists. Marketing efforts by computational chemistry software vendors have been shifting rapidly to that arena.

While there is considerable press about the Linux version of UNIX, it has still not penetrated the computational chemistry market to a large extent. Future high end computing will probably depend upon versions of UNIX with transparent delivery to desktop workstations.

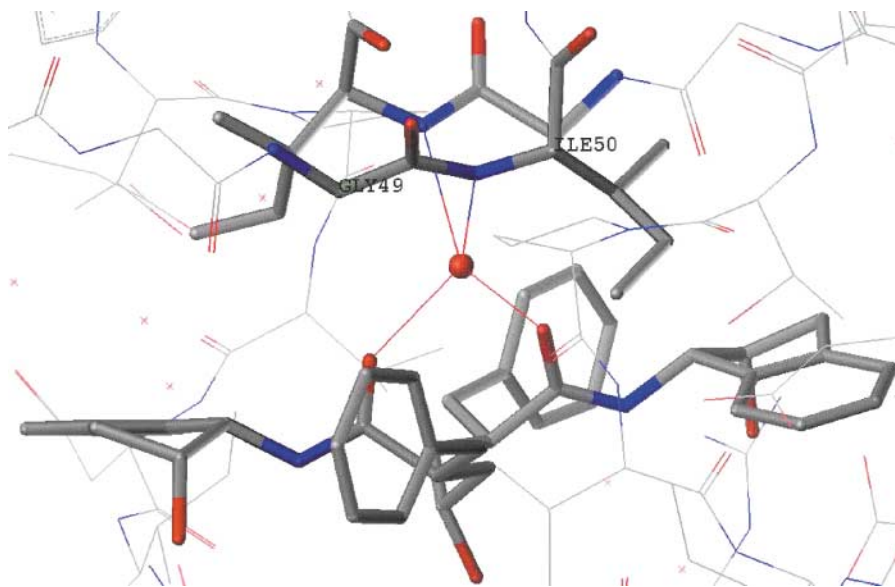


Fig. 6 X-ray structure of HIV protease and protease inhibitor taken from PDB.

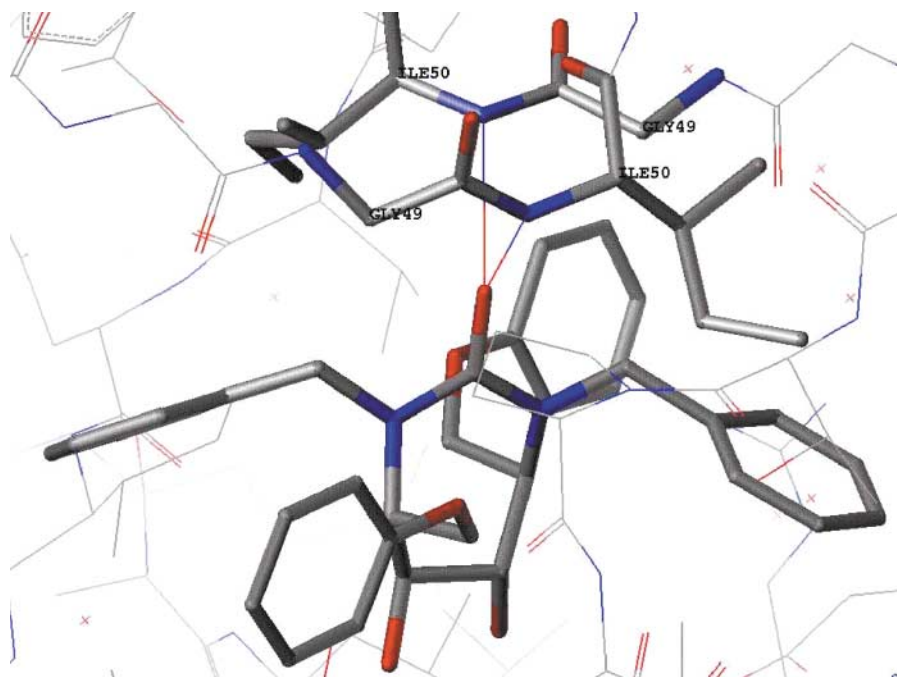


Fig. 7 X-ray structure of cyclic HIV protease inhibitor with carbonyl group mimic of water.

Software

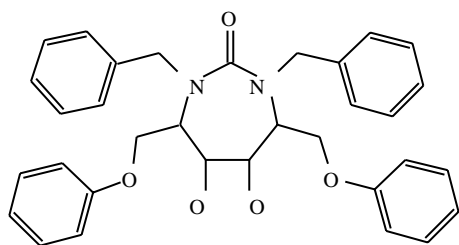
The emergence of molecular modeling and computational into a discipline has been closely associated with advances in computer hardware, computer graphics, and software. The availability of more sophisticated number-crunching power has allowed scientists to calculate the properties of larger and larger molecules. With the advent of supercomputers, dynamics calculations on biomolecules surrounded by an environment of water are becoming common.

The advances in software have also been spectacular. It was not terribly long ago when medicinal scientists were happy to get a color picture that somehow conveyed a three-dimensional perspective. Now software packages allow chemists to conveniently overlay a series of structures and do real-time rotations. Many companies have emerged, most originated by academicians, which

specialize in designing and marketing modeling software. The origins of much of the available software can be traced to university research laboratories. Although many of the packages have similar features and capabilities, they each have a characteristic style, which reflects the philosophies of the developers.

CONCLUSIONS AND FUTURE TRENDS

Chemists have long been associated with ball-and-stick mechanical structures, and the concepts for molecular modeling have been appreciated for a number of years by medicinal chemists. The precomputer modeling endeavors of Pauling and Watson and Crick demonstrated the utility of these methodologies. The mathematical formalisms for quantum and molecular mechanics were worked out years ago; however, the lack of sufficient computational means prevented practical applications. The explosion of modeling endeavors today is directly related to the advent of powerful new computers and algorithms. Now, larger and larger structures can be subjected readily to the rigors of calculations. More realistic simulations, including solvent interactions, can be treated. Molecular structures can be easily displayed and manipulated on computer graphics terminals. The full impact of high-speed desktop workstations, the Internet and the genome initiative.



Structure 2.

Computational chemists, medicinal chemists, and pharmaceutical scientists face an exciting future, as the computer-assisted approaches for drug design are becoming more powerful and more accessible. Every major pharmaceutical company has invested resources into these new tools. Education in the techniques, scope, and limitations are ever increasing. Improvements in the methodologies and computational resources will undoubtedly increase. Nevertheless, biological processes are still poorly understood, for the most part, at the molecular level. Many challenges remain, and theoretical and experimental work will remain tightly correlated in drug design and discovery.

REFERENCES

1. Wooden Ball-and-Stick Models Attributed to John Dalton. The Science Museum, Exhibition Road, South Kensington: London, England.
2. Watson, J.D. *The Double Helix*; New American Library: New York, 1968.
3. Watson, J.D.; Crick, F.H.C. *Nature* **1953**, *171*, 964–967.
4. Gordon, A.J. *J. Chem. Educ.* **1970**, *47*, 30–33.
5. Koltun, W. *Biopolymers* **1965**, *3*, 665–679.
6. Harvard Apparatus Inc.: South Natick, Mass, 01760.
7. Johnson, C.K. Report No. ORNL-3794, 1970 Oak Ridge National Laboratory: Oak Ridge, Tennessee.
8. Connolly, M.L. *Science* **1983**, *221*, 709–713.
9. Fuchs, H.; Goldfeather, J.; Hultquist, J.P.; Spach, S.; Austin, J.D.; Brooks, F.P., Jr.; Eyles, J.G.; Poulton, J. *Computer Graphics* **1985**, *19*, 111–120.
- 10a. D'Abro, A. *The Rise of the New Physics: Its Mathematical and Physical Theories*; Dover: New York, 1951.
- 10b. Pauling, L.; Wilson, E.B., Jr. *Introduction to Quantum Mechanics, With Applications to Chemistry*; McGraw-Hill: New York, 1935.
- 11a. Schrödinger, E. *Ann. Phys.* **1926**, *79*, 361.
- 11b. Schrödinger, E. *Ann. Phys.* **1926**, *80*, 437.
- 11c. Schrödinger, E. *Ann. Phys.* **1926**, *81*, 109.
12. Heisenberg, W. *Z. Phys.* **1925**, *33*, 879–893.
- 13a. Schrödinger, E. *Ann. Phys.* **1926**, *79*, 734.
- 13b. Eckart, C. *Phys. Rev.* **1926**, *28*, 711.
- 14a. Pople, J.A.; Beveridge, D. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.
- 14b. Dewar, M.J.S. *The Molecular Orbital Theory of Organic Chemistry*; McGraw-Hill: New York, 1969.
- 14c. Murrell, J.N.; Harget, A.J. *Semiempirical Self-Consistent-Field Molecular Orbital Theory of Molecules*; Wiley-Interscience: London, 1972.
- 14d. Pople, J.A.; Santry, D.P.; Segal, G.A. *J. Chem. Phys.* **1965**, *43*, S129–S135.
- 14e. Pople, J.A.; Segal, G.A. *J. Chem. Phys.* **1965**, *43*, S136–S149.
- 14f. Pople, J.A.; Segal, G.A. *J. Chem. Phys.* **1966**, *44*, 3289–3296.
- 14g. Pople, J.A.; Beveridge, D.L.; Dobosh, P.A. *J. Chem. Phys.* **1967**, *47*, 2026–2033.
15. Dewar, M.J. *J. Mol. Struct.* **1983**, *100*, 41–50.
16. Dewar, M.J.S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.
- 17a. Bingham, R.C.; Dewar, M.J.S.; Lo, D.H. *J. Am. Chem. Soc.* **1975**, *97*, 1285–1293.
- 17b. Bingham, R.C.; Dewar, M.J.S.; Lo, D.H. *J. Am. Chem. Soc.* **1975**, *97*, 1294–1301.
- 17c. Bingham, R.C.; Dewar, M.J.S.; Lo, D.H. *J. Am. Chem. Soc.* **1975**, *97*, 1302–1306.
- 17d. Bingham, R.C.; Dewar, M.J.S.; Lo, D.H. *J. Am. Chem. Soc.* **1975**, *97*, 1307–1310.
- 17e. Dewar, M.J.S.; Lo, D.H.; Ramsden, C.A. *J. Am. Chem. Soc.* **1975**, *97*, 1311–1318.
- 18a. Dewar, M.J.S.; Thiel, W.J. *J. Am. Chem. Soc.* **1977**, *99*, 4907–4917.
- 18b. Dewar, M.J.S.; McKee, M.L. *J. Am. Chem. Soc.* **1977**, *99*, 5231–5241.
- 18c. Dewar, M.J.S.; Rzepa, H.S. *J. Am. Chem. Soc.* **1978**, *100*, 777–794.
- 18d. Davis, L.P.; Guidry, R.M.; Williams, J.R.; Dewar, M.J.S.; Rzepa, H.S. *J. Comput. Chem.* **1981**, *2*, 433–445.
- 18e. Dewar, M.J.S.; McKee, M.L.; Rzepa, H.S. *J. Am. Chem. Soc.* **1978**, *100*, 3607.
- 18f. Dewar, M.J.S.; Healy, E. *J. Comput. Chem.* **1983**, *4*, 542–551.
19. Dewar, M.J.S.; Zebisch, E.G.; Healy, E.F.; Stewart, J.J.P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- 20a. Richards, W.G. *Quantum Pharmacology*; Butterworths: London, 1984.
21. Kier, L.B. *Molecular Orbital Theory in Drug Research*; Academic Press: New York, 1984.
22. Martin, Y.C. *Quantitative Drug Design*; Marcel Dekker, Inc.: New York, 1978.
23. Hehre, W.J.; Random, L.; Schleyer, P.; Pople, J. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986.
24. Pople, J.A. *J. Am. Chem. Soc.* **1975**, *97*, 5306–5308.
25. Hehre, W.J. *J. Am. Chem. Soc.* **1975**, *97*, 5308–5310.
26. Dewar, M.J.S. *J. Am. Chem. Soc.* **1975**, *97*, 6951.
27. Roberts, J.D. *Molecular Orbital Calculations*; Benjamin/Cummings: Reading, Massachusetts, 1961.
28. Streitwieser, A., Jr. *Molecular Orbital Theory for Organic Chemists*; John Wiley & Sons: New York, 1961.
29. Boys, S.F. *Proc. R. Soc. (London A)* **1950**, *200*, 542–554.
30. Hehre, W.J.; Stewart, R.F.; Pople, J.A. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
31. Hehre, W.J.; Ditchfield, R.; Stewart, R.F.; Pople, J.A. *J. Chem. Phys.* **1970**, *52*, 2769–2773.
32. Pietro, W.J.; Levi, B.A.; Hehre, W.J.; Stewart, R.F. *Inorg. Chem.* **1980**, *19*, 2225–2229.
33. Pietro, W.J.; Blurock, E.S.; Hout, R.F., Jr.; Hehre, W.J.; DeFrees, D.J.; Stewart, R.F. *Inorg. Chem.* **1981**, *20*, 3650–3654.
34. Binkley, J.S.; Pople, J.A.; Hehre, W.J. *J. Am. Chem. Soc.* **1980**, *102*, 939–947.
35. Gordon, M.S.; Binkley, J.S.; Pople, J.A.; Pietro, W.J.; Hehre, W.J. *J. Am. Chem. Soc.* **1982**, *104*, 2797–2803.
36. Dobbs, K.D.; Hehre, W.J. *J. Comput. Chem.* **1986**, *7*, 359–378.
37. Hehre, W.J.; Ditchfield, R.; Pople, J.A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

38. Binkley, J.S.; Pople, J.A. *J. Chem. Phys.* **1977**, *66*, 879–880.
39. Franci, M.M.; Pietro, W.J.; Hehre, W.J.; Binkley, J.S.; Gordon, M.S.; DeFrees, D.J.; Pople, J.A. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
40. Carlsen, N.R. *Chem. Phys. Lett.* **1977**, *51*, 192–196.
41. Krishnan, R.; Frisch, M.J.; Pople, J.A. *J. Chem. Phys.* **1980**, *72*, 4244–4245.
42. Hartree, D.R. *Proc. Cambridge Philos. Soc.* **1928**, *24*, 105.
43. Roothaan, C.C.J. *Rev. Mod. Phys.* **1951**, *23*, 69.
44. Hall, G.G. *Proc. roy. soc. (London A)* **1951**, *205*, 541.
45. Møller, C.; Plesset, M.S. *Phys. Rev.* **1934**, *46*, 618–622.
46. Pople, J.A.; Binkley, J.S.; Seeger, R. *Int. J. Quantum Chem. Simp.* **1976**, *10*, 1–19.
47. Pople, J.A.; Seeger, R.; Krishnan, R. *Int. J. Quantum Chem. Quantum Chem. Simp.* **1977**, *11*, 149–163.
48. Krishnan, R.; Pople, J.A. *Int. J. Quantum Chem.* **1978**, *14*, 91–100.
49. Schlegel, H.B. *J. Chem. Phys.* **1982**, *77*, 3676–3681.
50. Parr, R.G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.
51. Kohn, W.; Sham, L.J. *Phys. Rev.* **1934**, *A140*, 1133–1138.
52. Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *B136*, 864–871.
53. Burkert, U.; Allinger, N.L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
54. Westheimer, F.H.; Mayer, J.E. *J. Chem. Phys.* **1946**, *14*, 733–738.
55. Hill, T.L. *J. Chem. Phys.* **1946**, *14*, 465.
56. Dostrovsky, I.; Hughes, E.D.; Ingold, C.K. *J. Chem. Soc.* **1946**, 173–194.
57. Boyd, D.B. *Drug Inf. J.* **1983**, *17*, 121–131.
58. Goldstein, H. *Classical Mechanics*; Addison-Wesley: Reading, Massachusetts, 1981.
59. Allinger, N.L. *J. Am. Chem. Soc.* **1977**, *99*, 8127, and subsequent papers; MM2.
60. Allinger, N.L.; Yuh, Y.H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551, MM3.
61. Lii, J.-H.; Allinger, N.L. *J. Am. Chem. Soc.* **1989**, *111*, 8566.
62. Lii, J.-H.; Allinger, N.L. *J. Am. Chem. Soc.* **1989**, *111*, 8576.
63. Bowen, J.P.; Allinger, N.L. *Molecular Mechanics: The Art and Science of Parameterization. Reviews Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; VCH: New York, 1991; 2, 81–97.
64. Allinger, N.L.; Chen, K.S.; Lii, J.-H. *J. Comp. Chem.* **1996**, *17*, 642, MM4.
65. Nevins, N.; Chen, K.S.; Allinger, N.L. *J. Comp. Chem.* **1996**, *17*, 669.
66. Nevins, N.; Lii, J.-H.; Allinger, N.L. *J. Comp. Chem.* **1996**, *17*, 695.
67. Nevins, N.; Allinger, N.L. *J. Comp. Chem.* **1996**, *17*, 730.
68. Allinger, N.L.; Chen, K.S.; Katzenellenbogen, J.A.; Wilson, S.R.; Anstead, G.M. *J. Comp. Chem.* **1996**, *17*, 747.
69. Bowen, J.P. *New Vistas in Molecular Mechanics Calculations. Drug Design*; Charifson, P., Ed.; Marcel Dekker, Inc.: New York, 1997; 495–538.
70. Lennard-Jones, J.E. *Proc. Roy. Soc. (London, A)* **1924**, *106*, 463.
71. Bischoff, C.A. *Ber. Dtsch. Chem. Ges.* **1890**, *23*, 620, 1891, 24, 1074, 1086; 1891, 26, 1452.
72. Christie, G.H.; Kenner, J. *J. Chem. Soc.* **1922**, *121*, 614–620.
73. Kemp, J.D.; Pitzer, K.S. *J. Chem. Phys.* **1936**, *4*, 749.
74. Kemp, J.D.; Pitzer, K.S. *J. Am. Chem. Soc.* **1937**, *59*, 276–279.
75. Pitzer, R.M. *Acc. Chem. Res.* **1983**, *16*, 207–210.
76. Bartell, L.S. *J. Am. Chem. Soc.* **1977**, *99*, 3279–3282.
77. Allinger, N.L.; Hindman, D.; Honig, H. *J. Am. Chem. Soc.* **1977**, *99*, 3282–3284.
78. Radom, L.; Hehre, W.J.; Pople, J.A. *J. Am. Chem. Soc.* **1972**, *94*, 2371–2381.
79. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P.J. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
80. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. *J. Comp. Chem.* **1986**, *7*, 230–252.
81. Pearlman, A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham, T.E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.
82. Cornell, D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
83. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
84. Brooks, B.R.; Brucoleri, R.E.; Olafson, B.D.; Stales, D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
85. Buyong, M.; Lii, J.-H.; Allinger, N.L. *J. Comp. Chem.* **2000**, *21*, 814–825.
86. Bowen, J.P.; Pathiaseril, A.; Profeta, S., Jr.; Allinger, N.L. *J. Org. Chem.* **1987**, *52*, 5162–5166.
87. Bowen, J.P.; Allinger, N.L. *J. Org. Chem.* **1986**, *51*, 1513–1516.
88. Goldsmith, D.J.; Bowen, J.P.; Quamhiyeh, E.; Still, W.C. *J. Org. Chem.* **1987**, *52*, 951–953.
89. Bowen, J.P.; Allinger, N.L. *J. Org. Chem.* **1987**, *52*, 1830–1834.
90. Bowen, J.P.; Allinger, N.L. *J. Org. Chem.* **1987**, *52*, 2937–2938.
91. Bowen, J.P.; Reddy, V.; Patterson, D., Jr.; Allinger, N.L. *J. Org. Chem.* **1988**, *53*, 5471–5475.
92. Green, M.M.; Boyle, B.A.; Vairamani, M.; Mukhopadhyay, T.; Saunders, W.H., Jr.; Bowen, P.; Allinger, N.L. *J. Am. Chem. Soc.* **1986**, *108*, 2381–2387.
93. Spellmeyer, D.C.; Houk, K.N. *J. Org. Chem.* **1987**, *52*, 959–974.
94. Dorigo, A.E.; Houk, K.N. *J. Am. Chem. Soc.* **1987**, *109*, 3698–3708, (corrections), 1988, 110, 4874.
95. Rudolf, K.; Hawkins, J.M.; Loncharich, R.J.; Houk, K.N. *J. Org. Chem.* **1988**, *53*, 3879–3882.
96. Wu, Y.-D.; Houk, K.N. *J. Am. Chem. Soc.* **1987**, *109*, 908.
97. Jacoby, S.L.S.; Kowalik, J.S.; Pizzo, J.T. *Iterative Methods for Nonlinear Optimization Problems*; Prentice Hall, Englewood Cliffs: New Jersey, 1972.
98. Williams, J.E.; Stang, P.J.; Schleyer, P.V.R. *Annu. Rev. Phys. Chem.* **1968**, *19*, 531.
99. Wibeg, K.B. *J. Am. Chem. Soc.* **1965**, *87*, 1070–1078.
100. Fletcher, R.; Reeves, C.M. *Comput. J.* **1964**, *7*, 149.

101. Fletcher, R.; Powell, M.J.D. *Comput. J.* **1963**, *6*, 163.
102. Jacoby, S.L.S.; Kowalik, J.S.; Pizzo, J.T. *Iterative Methods for Nonlinear Optimization Problems*; Prentice Hall, Englewood Cliffs: New Jersey, 1972.
103. Ermer, O. *Struct. Bonding (Berlin)* **1976**, *27*, 161.
104. Marshall, G.R. *Medicinal Chemistry VI, Proceeding of the 6th International Symposium on Medical Chemistry*. Sinkin, A., Ed.; Cotswold Press: Oxford, 1979; 225–235.
105. Marshall, G.R.; Motoc, I. *Top. Mol. Pharmacol.* **1986**, *3*, 115–155.
106. Marshall, G.R. *Annu. Rev. Pharmacol. Toxicol.* **1987**, *27*, 193–213.
107. Beveridge, D.L., *Computer Simulation of Chemical and Biochemical Systems*; Jorgensen, W.L., Eds.; New York, 1986.
108. McCammon, J.A.; Harvey, S.C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University: Cambridge, 1987.
109. van Gunsteren, W.F.; Weiner, P.K., Eds.; *Computer Simulation of Biomolecular Systems*; ESCOM: Leiden, 1989.
110. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E.J. *Chem. Phys.* **1953**, *21*, 1087.
111. Saunders, M.; Houk, K.N.; Wu, Y.-D.; Still, W.C.; Lipton, M.; Chang, G.; Guida, C. *J. Am. Chem. Soc.* **1990**, *113*, 1419–1427.
112. Miranker, A.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161–9175.
113. Jorgensen, W.L.; Ravimohan, C.J. *Chem. Phys.* **1985**, *83*, 3050–3054.
114. McCammon, J.A. *Science* **1987**, *238*, 486–491.
115. Jorgensen, W.L. *Acc. Chem. Res.* **1989**, *22*, 184–189.
116. Horn, A.S., de Ranter, C.J., Eds. *X-ray Crystallography and Drug Action*; Clearendon: Oxford, 1984.
117. Stezowski, J.J.; Chandrasekhar, K. *Annu. Rep. Med. Chem.* **1986**, *21*, 293–302.
118. Jefson, M.R. *Annu. Rep. Med. Chem.* **1988**, *23*, 275–283.
119. Clore, G.M.; Gronenborn, A.M.; Brünger, A.T.; Karplus, M. *J. Mol. Biol.* **1985**, *186*, 435–455.
120. Martin, Y.C. *Quantitative Drug Design—A Critical Introduction*; Marcel Dekker, Inc.: New York, 1978.
121. Hansch, C. *Acc. Chem. Res.* **1969**, *2*, 232–239.
122. Cramer, R.D., III; Patterson, D.E.; Bunce, J.D. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
123. Gund, P. *Annu. Rep. Med. Chem.* **1978**, *14*, 299–308.
124. Baker, B.R.; Wood, W.F. *J. Med. Chem.* **1968**, *11*, 644–649.
125. Baker, B.R. *Design of Active-Site-Directed Irreversible Enzyme Inhibitors*; Wiley: New York, 1967.
126. Zee-Cheng, K.-Y.; Cheng, C.C. *J. Pharm. Sci.* **1970**, *59*, 1630–1634.
127. Townsend, L.B.; Schram, K.H.; Beranek, J. *Proc. Natl. Acad. Sci. USA.* **1979**, *76*, 4075–4078.
128. Smellie, A.; Kahn, S.D.; Teig, S.L. *J. Chem. Inf. Comput. Sci.* **1995**, *35* (2), 295–304.
129. Smellie, A.; Kahn, S.D.; Teig, S.J. *J. Chem. Inf. Comput. Sci.* **1995**, *35* (2), 285–294.
130. Allen, F.H.; Bellard, S.; Brice, M.D.; Cartwright, B.A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B.G.; Kennard, O.; Motherwell, W.D.S.; Rogers, J.R.; Watson, D.G. *Acta Crystallogr. Sect. B* **1979**, *35*, 2331–2339.
131. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
132. Hann, M.; Green, R. *Chemoinformatics Chem. Biol.* **1999**, *3* (4), 379–383.
133. Swannstrom, R.; Erona, J. *J. Pharmacol. Ther.* **2000**, *86* (2), 145–170.
134. Leung, D.; Abbenante, G.; Fairlie, D.P. *J. Med. Chem.* **2000**, *43* (3), 305–341.
135. Wlodawer, A.; Erickson, J.W. *Annu. Rev. Biochem.* **1993**, *62*, 543–585.
136. Lam, P.Y.S.; Jadhav, P.K.; Eyermann, C.J.; Hodge, C.N.; Ru, Y.; Bacheler, L.T.; Meek, J.L.; Otto, M.J.; Rayner, M.M.; Wong, Y.N.; Chang, C.-H.; Weber, P.C.; Jackson, D.A.; Sharpe, T.R.; Erickson-Viitanen, S. *Science* **1994**, *263*, 380–384.

Computer Systems Validation

Orlando Lopez

Johnson & Johnson, NCS, Raritan, New Jersey, U.S.A.

INTRODUCTION

Computer systems validation, as established in *21 Code of Federal Regulation* (CFR) Part 11.10(a) and defined in the recent draft United States (US) Food and Drug Administration (FDA) guideline,^[1] is one of the most important requirements applicable to computer systems performing FDA-regulated operations. It involves establishing the conformance to the intended use, user, regulatory, safety, and function allocated to the computer system.

Similar to any FDA-regulated products, quality is built into a computer system during its conceptualization, development, and operational life. The quality of computer systems cannot be tested after being developed. In addition to software and hardware testing, other verification activities include code walkthroughs, dynamic analysis, and trace analysis. The documentation generated during the validation can be subject to examination by FDA field investigators. The results of a high-quality validation program can ensure with a high degree of assurance the trustworthiness of electronic records and computer systems-related functionality.

OVERVIEW

The introduction in 1997 of 21 CFR Part 11, Electronic Records, Electronic Signatures Rule (hereafter referred to as Part 11) provided the formal codification applicable to computer systems performing FDA-regulated operations. One fundamental principle in Part 11 is that it requires organizations to store regulated electronic data in their electronic form once a record is saved to durable media, rather than keep paper-based printouts of the data on file, as had been the long-term practice in organizations performing regulated operations.^[2] If information is not recorded to durable media, the stored data will be lost and they cannot be retrieved for future use. If “retrievability” is an attribute, then procedural and technological controls contained in Part 11 are essential to ensuring integrity. The implementation of procedural and technological controls to achieve compliance with Part 11 shall be monitored during the SLC.

The approach to be used in covering computer systems validation is by presenting key elements applicable to

any development/maintenance methodology. It is not intended to cover everything that computer system validation should encompass, including Part 11. A wide range of information about computer systems validation is listed in the bibliography.

KEY VALIDATION ELEMENTS

The key elements required to successfully execute computer system validation projects are as follows:

- Selection of a development/maintenance methodology that best suits the nature of the system under development.
- Identification of operational functions associated with the users, operational checks, regulatory, company standards, and safety requirements.
- Selection of hardware based on capacity and functionality.
- Inspection and testing of the operational functions.
- Identification and testing of “worst case” operational/production conditions.
- Reproducibility of the testing results based on statistics.
- Documentation of the validation process.^[3]
- Written design specification that describes what the software is intended to do and how it is intended to do it.
- A written validation plan based on the design specification, including both structural and functional analysis.
- Test results and evaluation of how these results demonstrate that the predetermined design specification has been met.
- Availability of procedural controls to maintain the validation state of the computer system and its operating environment.
- Any modification to a component of the system and its operating environment must be evaluated to determine the impact to the system. If required, qualification/validation is to be re-executed totally or partially.

Selection of a Development/Maintenance Methodology

The SLC is the “period of time that begins when a product is conceived and ends when the product is no longer available for use.”^[4] Certain overall discrete work



products are expected when evidencing the development and maintenance work of computer systems compliance to regulatory requirements. Refer to "Documentation of the Validation Process." The selected SLC specifies the overall periods and associated events. Different system acquisition strategies and software development models can be adapted to the SLC depicted in Fig. 1.^[5] The SLC model focuses on software engineering key practices and does not specify or discourage the use of any particular software development method. The acquirer determines which of the activities outlined by the standard will be conducted, and the developer is responsible for selecting the methods that support the achievement of contract requirements. A modifiable framework must be tailored to the unique characteristics of each project. The SLC includes the following periods:

Conceptualization.
Development.
Early operational life.
Maturity.
Aging.

Project Recommendation, Project Initiation, Release for Use, and Retirement are events. These events are considered phase gates or major decision points, which include formal approvals before the development can proceed to the next period.

The development methodology associated with the SLC is a structured process that decomposes the engi-

neering tasks and associated work products in support of the computer system validation effort. It breaks down the systems development process into subperiods, containing specific inspection and testing tasks that are appropriate for the intended use of the computer system. During each subperiod, detailed discrete work products are developed. This approach leads to well-documented systems that are easier to test and maintain, and for which an organization can have confidence that the system's functions will be fulfilled with a minimum of unforeseen problems.

The most common development methodologies are the Waterfall Model, Incremental Development, Evolutionary Model, Object Oriented, and Spiral Model.

A critical component of the validation process is providing assurance that the development/maintenance methodology is being followed. The SLC and associated development/maintenance methodology applicable to computer systems performing regulated operations shall be specified in procedural control(s). A project team should have the authority to select a developmental/maintenance methodology that best suits the nature of the system under development/maintenance and that is different from the one included in the related procedural control. If this is the case, the selected development or maintenance methodology must be explained in the validation plan.

It is the objective of FDA-regulated companies to select the appropriate SLC and associated development/maintenance methodology. Development and mainten-

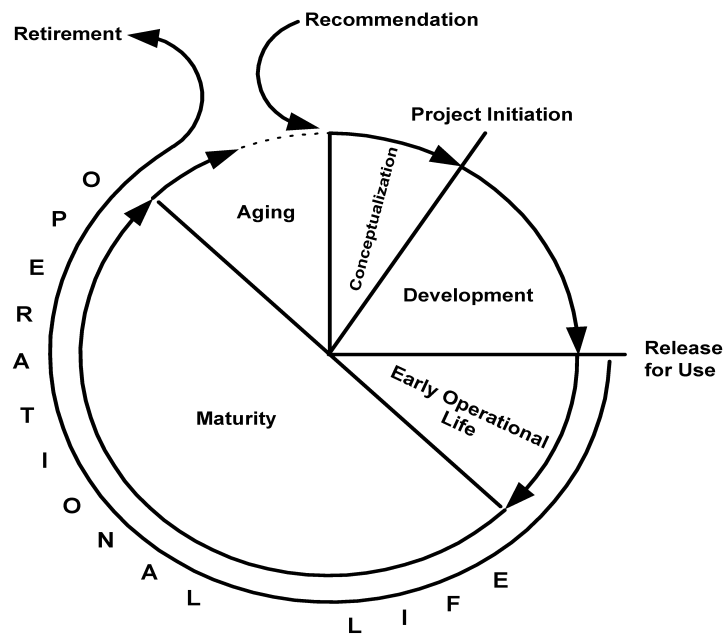


Fig. 1

ance teams shall receive adequate training in the use of the chosen methodology.

Identification of Operational Functions

Using a structured process, the goal of the Development Period is to specify, design, build, test, and install the application to be automated or updated. One of the deliverables of the Development Period is the written and approved operational functions (e.g., user's, functional, design) that describe what the application is intended to do.

A key activity to identify operational functions is by gathering systems requirements. The term "requirement" defines a bounded characterization of the scope of the system. It contains the information essential to support the operation/operators. These requirements include functional capacity, execution capability, safety, operational, installation, system maintenance, and regulatory compliance.

The refined scope is captured in the requirements specification, which describes what the system is supposed to do from the process/user's/compliance perspective. The requirements specification is used as part of the framework to select the computer technology supplier and/or contract developer. The system functionality must be well defined at the outset in order to provide the prospective supplier/integrator with enough information to provide a detailed and meaningful quotation. The requirements specification is used to develop the performance qualification (PQ) protocol.

The requirements specification addresses the following:

- The process to familiarize the developer with the user, process and data acquisition requirements, and special considerations of the project.
- The scope of the system and strategic objectives.
- The problem to be solved.
- Process review and sequencing, as well as where each operation is to be completed.
- The direction to solve the problem (e.g., device driven or may just be the mode of presentation of data, data security, data backup, data and status reporting and trending).
- Redundancy and error-detection protocol.
- Environmental control.
- Interfaces (e.g., to field devices, data acquisition, reports and HMI), input/output (I/O) list, communication protocol, and data link requirements.
- Type of control/process to be performed.
- Operational checks and sequencing.
- Data management.

- Definition of the input and output domains.
- How the data are to be collected, used, and stored.
- How input data influence the operation of the system.
- Retention requirements.
- Data security requirements.
- Audit trails and metadata.
- Timing requirements.
- Regulatory requirements.
- Preliminary evaluation of technology.
- Feasibility study and preliminary risk assessment.
- Safety and security considerations.
- Nonfunctional requirements (e.g., development standards, program-naming convention standards).

Each requirement in the requirements specification must be "testable." A "testable" requirement includes an objective criterion and it is nonambiguous. A "testable" requirement provides the advantage that it can be recorded in quantified terms and allows for a subsequent review and independent evaluation of the test results.

Selection of Hardware Based on Capacity and Functionality

Based on the identification of operational functions and design, computer hardware technologies can be selected. Depending on available technology and cost, automated functions can be assign to the computer hardware or software. Computer hardware can be further decomposed into a number of subelements. Processor, memory, I/O and networks are some examples.

Process, field instruments, control requirements, available technology and cost are some of the factors driving the selection of the hardware. This selection is specified in the requirement specification and the implementation described in the design specification. The design specification needs to be sufficiently detailed in order to familiarize the implementation team (e.g. engineering) and the hardware vendor with the requirements and special considerations of the process, field instrument, and control requirements.

The requirements include, but are not limited to, the following:

- Purpose of the system.
- Regulatory requirements.
- Information and material inputs.
- Preliminary block diagrams.
- Data processing requirements (e.g., supervisory control and data acquisition).
- Number and type of I/O cards.
- Instrumentation and cabling.
- Control and information outputs.



Operating modes.
 Alarms and alerts.
 Safety features.
 Error checking.
 Reporting.
 Redundancy requirements.
 Environmental requirements.
 Network requirements.
 Supporting utilities requirements.
 Hardware/human machine interfaces.
 General plan and acceptance criteria.

The critical field instrumentation must support accuracy and reliability requirements over the entire process range conditions.

Inspection and Testing of the Operational Functions

Software engineering practices may include documented unit testing, code reviews, explicit high-level and low-level design documents, explicit requirements and functional specifications, structure charts and data flow diagrams, function-point analysis, defect and resolution tracking, configuration management, and a documented software development process.

These are the same quality principles that the FDA expects to be used during the development and maintenance of computer systems. These quality principles shall be contained in procedural controls. Inspections and testing are part of these principles.

Testing and inspections are activities performed as part of the development methodology. Numerous inspection steps are undertaken throughout the system development and operational life to determine whether a computer system is validated. These include static analyses such as document and code inspections, walk-through, and technical reviews. These activities and their outcomes help to reduce the amount of system-level functional testing needed in the operational environment in order to confirm that the software meets the requirements and intended uses.

Reproducibility of the Testing Results Based on Statistics

Testing is not just executing a program using a test data file or randomly selected test cases just prior to implementation. It is an on going process using techniques based on Statistical Process Control (SPC) principles and product quality concepts implemented as components of a Statistical Quality Control (SQC) program.^[6]

If software systems are viewed as a manufacturing facility that produces the desired output products, then statistical sampling procedures and statistical inference can be used to predict the reliability of test results. Instead of paying too much attention to the development of the application, another factor that requires attention is the data (raw material) to be converted into information (product). Information flow is a design technique that may be helpful in achieving this task.

Concerning testing, the input and output domains must be strictly defined. This suggests a response to the following: 1) What sampling technique will ensure an adequate subset of possible input values that will provide as complete a test of the software as possible? and 2) What information sampling procedures will allow the developers to determine product reliability?

It is suggested that White and Black box test cases design strategies be used to sample the data (input domain), including Equivalent Partitioning, Boundary Analysis and Error Guessing, Cause-Effect Graphing, and Structural Tableau.

The basics of Software Testing must be understood before the more abstract principles of statistical inference are tackled.

Documentation of the Validation Process

Design specification

System requirements are allocated to the software design. During the technical design it is described how each specification described in the system specification deliverable is to be implemented. This includes also developed subsystems components and interfaces, data structure, design constraints, algorithms and system decomposition. This activity is very critical to medical device companies. Design inputs are contained in the requirements of the computer system, and design output(s) can be included as part of the specification of the design.^[7] This design is the input for developing integration test and operational checks.

Design according to the development methodology and specific procedural controls.

Computer hardware and software architecture.

Data structures.

Flow of information.

Interfaces

Put together the design.

Perform design reviews. Verify whether the risks previously identified were mitigated as part of the solution presented in the design.

Finalized the test planning.

Design Part 11 technical controls.

Approve the design specification deliverable.
 Conduct in-process audit activities associated with the technical design.
 Re-visit the risk analysis.

Begin the planning of development of procedural controls for those Part 11 requirements not covered by technology.

Validation plan

Validation plans are documents that tailor a firm's overall philosophies, intentions, and approaches to be used for establishing performance adequacy to a specific project. They state who is responsible for performing development and validation activities. They identify which systems are subject to validation, define the nature and extent of inspection and testing expected to be done on each system, and outline the protocols to be followed to accomplish the validation.

In summary, validation plans describe the following:

- Organizational structure of the computerization project.
- Responsible departments and/or individuals.
- Resource availability.
- Risk management.
- Time restrictions.
- SLC and development methodology to be followed.
- Deliverable items.
- Overall acceptance criteria.
- Development schedule and timeline.
- System release sign-off process.
- Sample format for key documentation.

Test results and evaluation

Although installation qualification (IQ)/operational qualification (OQ)/PQ terminology has served its purpose well and is one of the many legitimate ways to organize computer system testing tasks in FDA-regulated industries, this terminology may not be well understood among many software professionals. However, organizations performing regulated operations must be aware of these differences in terminology as they ask for and provide information regarding computer systems.

Once the qualification protocols have been completed, test results and data need to be formally evaluated. Written evaluation needs to be presented clearly in a manner that can be readily understood. The report should also address any nonconformance or deviation to the validation plan encountered during the qualification and resolution. The outline of the report parallels the structure of the associated protocol. The qualification testing

should be linked with relevant specification's acceptance criteria, such as PQ vs. system requirements specification deliverable, OQ vs. system specification deliverable, and IQ vs. technical design specification deliverable. If applicable, it is included as part of the summary of the results of inspections and technical review of all technologies that are elements of the systems.

In very large validation efforts, a report references (by title and document reference number) other documents that satisfy the protocol requirements. In smaller validation efforts, actual evidence is incorporated as appendices to the report.

The documentation and results of the qualification efforts are assembled and reviewed by appropriate and qualified personnel. Following the review, the personnel responsible for the criticality of the system, including QA, approve the qualification effort.

The approval of all qualification reports is a confirmation that the computer system as a whole has been proven to fit its purpose and that all essential elements of documentation are available. On computer systems controlling manufacturing equipment (process control systems), the approval of all qualification reports indicates the release of the computerized systems to the Process/Product Performance Qualification. On other computer systems, the approval of all reports indicates the release of the system to the user.

Test results on Part 11 shall be addressed in the associated qualification report.

The Project Report summarizes the outcome of each activity performed to develop or maintain computer systems and the verification of critical checkpoints throughout the entire development process. The end-result is to verify that good quality development procedures were adhered to as established in the project plan.

All verification and testing results completed during the project shall be addressed in the Project Report as well.

The approval of the project report is the event to be considered prior to the release of the system for operation.

Validation maintenance

After the system has been released for operation, computer system maintenance activities take over. The maintenance activities must be governed by the same procedures followed during the Development Period.

The validated status of computer systems performing regulated operations is subject to threat of changes in its operating environment, either known or unknown. Adherence to security, operational management, business continuity, change management, periodic review, and decommissioning provides a high degree of assurance that the system is being maintained in a validated state. It is



the objective of organizations to have procedures in place to minimize the risk of computer systems performing regulated operations out of validated state.

Maintenance in computer systems becomes an essential issue, particularly when a new version of the supplier-provided standard software is updated. A change control procedure must be implemented whereby changes in the software and computer hardware may be evaluated, approved, and installed.

If necessary, additional analysis may be needed to evaluate the changes (e.g., impact analysis) to the computer systems. The procedure should allow for both planned and emergency changes to the system. This procedure must include provision for updating of pertinent documentation on the system, including procedures. Records of changes to the system must be kept for the same period as any other regular production document.

Table 1^[8] summarizes the periods and events applicable to the operational life of computer systems and associated key practices.

Evaluation of Modification to Computer Systems

As required by regulations, all maintenance work must be performed after the evaluation and approval of the work and must be consistent with the selected SLC methodology. Maintenance to a software system includes, among other things, the following:

- Perfective maintenance or correcting the system because of new requirements and/or functions.
- Adaptive maintenance or correcting the system because of a new environment, which could include new hardware, new sensors or controlled devices, new operating systems, new regulations.
- Corrective maintenance or correcting the system because of detection of errors in the design, logic, or programming of the system. It is essential to recognize that the longer a defect is present in the software before it is detected the more expensive it is to fix it.

Preventive maintenance or correcting the system to improve future maintainability or reliability in order to provide a better basis for future enhancements.

Change management procedural control is in place when the validated system is released for use. The change management procedural control provides for the following activities:

- Identifying and specifying the change.
- Assessing risk, criticality, and impact of change.
- Specifying testing requirements and acceptance criteria.
- Implementing change after authorization.
- Performing regression testing.
- Reviewing of change(s) with an independent reviewer.
- Updating system and user documentation to reflect implemented change(s).
- Establishing of provisions for the management of “emergency change,” including expeditious documentation modifications.

SYSTEM DEVELOPMENT FILES

One key element to support the SLC is the availability and maintenance of system development files. The developer shall document the development of each system unit, system component, and configuration items in software development files.

The developer should establish a separate system development file for each unit or a logically related group of units. The developer should document and implement procedures for establishing and maintaining system development files. The developer should maintain the system development files until the retirement of the system. The system development files should be available to the agency review upon request. System development files may be generated, maintained, and controlled by automated means. To reduce duplication, system development files should not contain information provided in other documents or system development files. The set of system

Table 1 Period/events computer systems operational life

Period/event	Representative characteristics	Key practices
Early operational life	Phased roll outs	Problem reporting and maintenance
Maturity	Corrective, adaptive, perfective, and preventive maintenance	Operational audit Performance evaluation
Aging	Maintainability issues of obsolete technology (e.g., access to electronic records)	Periodic review Re-engineering analysis

development files shall include (directly or by reference) the following information:

Design considerations and constraints.
Design documentation and data.
Schedule and status information.
Test requirements and responsibilities.
Verification and test procedures, and results.

CONCLUSION

The validation of computer systems in the U.S. FDA-regulated environment is an ongoing process that is integrated with the entire System Life Cycle (SLC). Quality to a software system is introduced by following the system life cycle and following the key validation elements.

REFERENCES

1. Food and Drug Administration. *Draft "Guidance for Industry: 21 CFR Part 11; Electronic Records; Electronic Signatures Glossary of Terms"*; U.S. Department of Health and Human Services: Washington, DC, August, 2001.
2. Note that Part 820 requires that "results" of acceptance activities be recorded but not necessarily all raw data. "Results" must have audit trails. Be sensitive to the need for raw data during failure investigations under Corrective and Preventive Actions. Refer to Part 820 preamble, pp. 52631 and 52646.
3. FDA. *Guidance for the Industry: Computerized Systems Used in Clinical Trials*; April, 1999.
4. ANSI/IEEE Std 610.12-1990. *Standard Glossary of Software Engineering Terminology*; Institute of Electrical and Electronic Engineers: New York, 1990.
5. Herr, R.R.; Wyrick, M.L. *A Globally Harmonized Glossary of Terms for Communicating Computer Validation Key Practices*; PDA Journal of Pharmaceutical Science and Technology, March/April, 1999.
6. Cho, C.-Q. *Quality Programming: Developing and Testing Software Using Statistical Quality Control*; Wiley: New York, 1987.
7. López, O. Applying design controls to software in the FDA-regulated environment. *J. cGMP Compliance* **July, 1997**, 1 (4).
8. Grigonis, G.J.; Subak, E.J.; Wyrick, M.L. Validation key practices for computer systems used in regulated operations. *Pharm. Technol.* **June, 1997**.

BIBLIOGRAPHY

- Annex 11 of the EU-GMP.
- ANSI/AAMI SW68. *Medical Device Software—Software Life Cycle Processes*; Association for the Advancement of Medical Instrumentation, 2001.
- ANSI/ISO/ASQ Q9000-3. *Quality Management And Quality Assurance Standards—Part 3: Guidelines for the Application of ANSI/ISO/ASQC Q90001-1994 to the Development, Supply, Installation and Maintenance of Computer Software*; 1997.
- FDA Compliance Policy Guides for Computerized Drug Processing. *CGMP Applicability to Hardware and Software (CPG 7132a.11); Vendor Responsibility (CPG 7132a.12); Source Code for Process Control Application Programs (CPG 7132a.15); Input/Output Checking (CPG 7132a.07); Identification of "Persons" on Batch Production and Control Records (CPG 7132a.08); Enforcement Policy: 21 CFR Part 11; Electronic Records; Electronic Signatures (CPG 7153.17)*.
- Grigonis, Subak, Wyrick. Validation key practices for computer systems used in regulated operations. *Pharm. Technol.* **June, 1997**.
- Good Automated Manufacturing Practice Forum (Rev 4). *Supplier Guide for Validation of Automated Systems in Pharmaceutical Manufacture*; ISPE, 2001.
- Herr; Wyrick. A globally harmonized glossary of terms for communicating computer validation key practices. *PDA J. Pharm. Sci. Technol.* **March/April, 1999**.
- IEC 62A/62304/Ed.1 (IEC 62A/338/NP). *Medical Device Software—Software Life Cycle Processes*; IEC, 2001.
- ISO/IEC 1207. *Information Technology—Software Life Cycle Processes*; ISO/IEC, 1995.
- ISO/IEC 12119. *Information Technology—Software Packages—Quality Requirements and Testing*; ISO/IEC, 1994.
- López, O. FDA regulations of computer systems in drug manufacturing—13 years later. *Pharm. Eng.* **May/June, 2001**, 21 (3).
- López, O. Implementing software applications compliant with 21 CFR part 11. *Pharm. Technol.* **March, 2000**.
- López, O. *21 CFR Part 11 A Complete Guide to International Compliance*; Sue Horwood Publishing Limited, 2002, (<http://www.computer-systems-validation.com>), ISBN 09540706 7 4.
- PIC/S Guidance. *Best Practices for Computerized Systems in Regulated 'GxP' Environments*; PIC/S, January, 2000.
- Wyn, S. Regulatory requirements for computerized systems in pharmaceutical manufacture. *Softw. Eng. J.* **1996**, 11 (1).

Dendrimers

Antony D'Emanuele

David Attwood

Ragheb Abu-Rmaileh

University of Manchester, Manchester, U.K.

INTRODUCTION

Dendrimers are highly branched macromolecules that can be subdivided into three architectural components: a central core branched cell, interior branch cells, and branch cells possessing surface groups. The term dendrimers was coined in the early 1980s by Tomalia et al.^[1] from the Greek words dendron (tree) and meros (part) and relates to the symmetrical branch-like structure of these polymers. Dendrimers are synthesized through a stepwise repetitive reaction sequence, which gives rise to different generations of the same molecule and determines the size and surface functionality of the macromolecule. The dendrimer microenvironment possesses some interesting properties. Cavities in the core structure and folding of the branches create cages and channels, which, depending upon how the dendrimer is constructed, may be either hydrophilic or hydrophobic in nature. Specific binding sites may also be incorporated. The surface groups of dendrimers are amenable to modification and can be tailored for specific applications. The dendrimer architecture, therefore, permits control over properties such as shape, size, density, polarity, reactivity, and solubility. The structure of a generation 3 polyamidoamine (PAMAM) dendrimer is shown in Fig. 1. Dendrimers can be very large; PAMAM dendrimers have been synthesized with a diameter in excess of 10 nm. Fig. 2 compares the size of a series of PAMAM dendrimers to a number of biological structures. Table 1 gives the relationship between generation and dendrimer size, molecular weight and number of surface groups.^[2] As can be seen, an increase in dendrimer generation results in a doubling in the number of surface groups.

Although the synthesis of dendrimer-like structures was initially described in 1978,^[3] it is only since the 1990s that there has been an intense interest in these polymers, partly attributable to the availability of a commercial source of PAMAM dendrimers.

SYNTHESIS

Dendrimers may be prepared by either a convergent^[4–6] or a divergent approach.^[1] Much of the work on dendrimers has been based on the commercially available Starburst® PAMAM dendrimers,^[1,7] which may be synthesized from

an ammonia or ethylenediamine core (EDA), and possess an amidoamine branching structure. This family of dendrimers was introduced by Tomalia et al. in 1985^[1] and are prepared by the divergent approach, where the branching dendritic structure is built up from a central core. PAMAM dendrimers are synthesized by an iterative process involving two reactions: Michael addition followed by amidation. The first iteration of these two reactions results in the formation of a zero-generation (G0) dendrimer, the subsequent addition–amidation cycles each result in growth and the formation of a higher generation (Fig. 3), this synthetic cycle yields a full-generation (amine terminated) dendrimer. Cessation of the reaction after the Michael addition results in the eventual formation of a half-generation (carboxyl terminated) dendrimer. The growth of a dendrimer is self-limiting, and governed by steric hindrance arising from the introduction of numerous surface groups.^[8,9] The convergent method of synthesis involves the initial creation of the dendrimer branches, followed by assembly to form the dendrimer. The convergent method produces a system with a low polydispersity, while the divergent method can produce dendritic structures with defects arising from incomplete reactions.^[10] However, the divergent method has the advantage of being able to yield higher generation dendrimers. The methods of synthesis and their relative merits have been reviewed by Dykes.^[11]

Three structural components are common to all dendrimers: a core unit, peripheral groups, and the multiple branching units that span the two. The core unit in dendrimers is usually an important part of the structure as it covalently links the dendritic “wedges” (dendrons). However, cored dendrimers have been synthesized in which surface crosslinking maintains the integrity of the dendritic structure.^[12]

The preparation of structurally perfect dendrimers traditionally can be time-consuming due to the required repetitive coupling and activation steps, and the necessity for extensive purification.^[13] PAMAM dendrimers are just one of several types of dendrimers with pharmaceutical applications that have been synthesized. Poly(aryl ether) dendrimers were developed by Hawker and Fréchet and were the basis of unimolecular micelles and cancer drug conjugates.^[4,14,15] Wang, Zeng, and Zimmerman^[16]

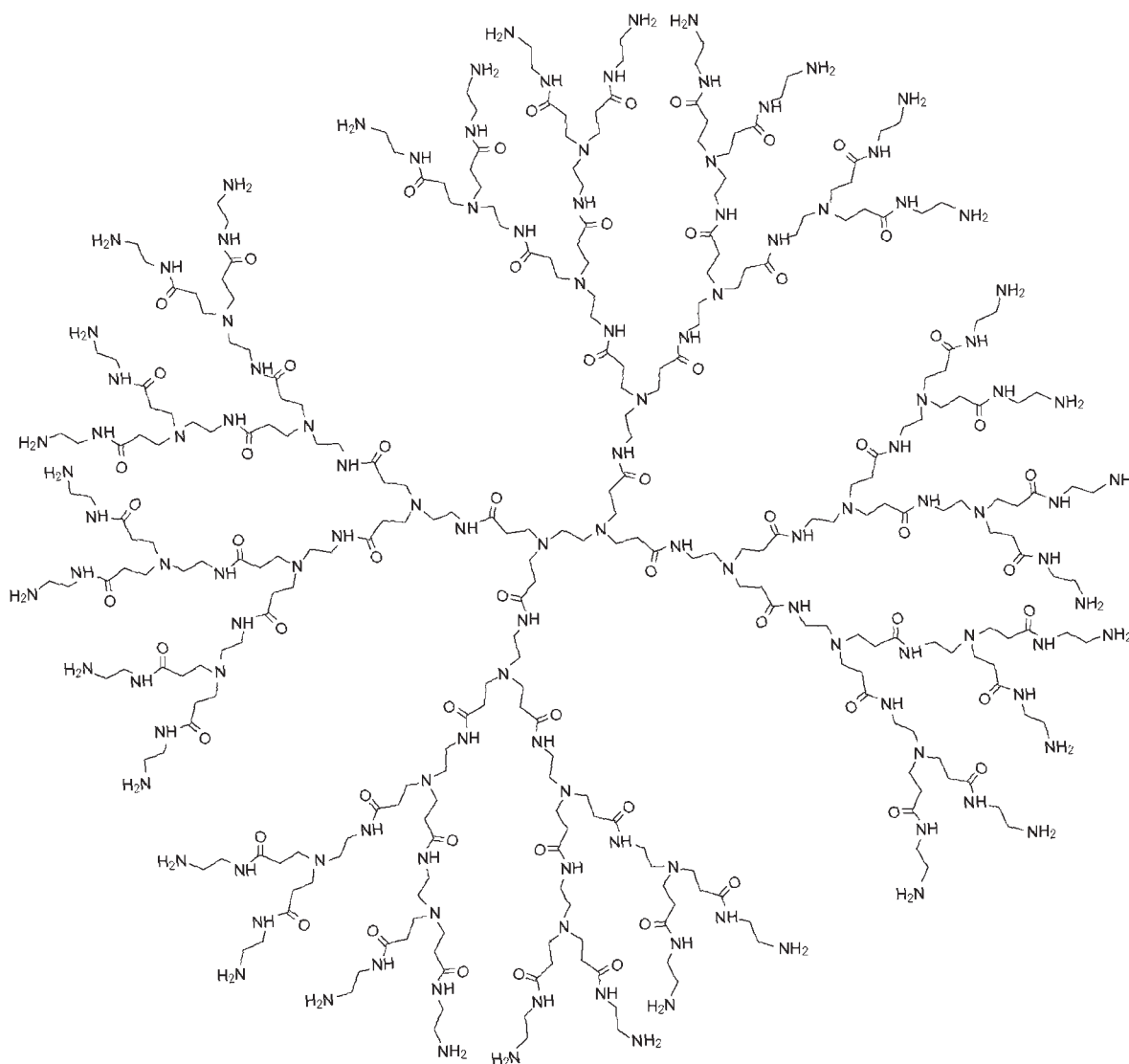


Fig. 1 Structure of a G3 PAMAM dendrimer. (From Ref. [110], © 2000 Elsevier Science.)

reported the synthesis of poly(phenylacetylene) and poly(benzyl ether) (PBE) dendrimers to encapsulate benzamidinium guests by hydrogen bonding. Jansen, de Brabander-van den Berg, and Meijer^[17] described the synthesis of poly(propyleneimine)-based dendrimers and their applications in the encapsulation of guest molecules. Numerous other dendrimer families have been synthesized with a wide range of applications.^[18–20]

Several peptide dendrimers have been reported,^[21–25] a major application of which is for the preparation of multiple antigen peptides (MAPs). For example, Tam^[26] used the poly(lysine) platform to prepare a MAP. These peptides are used to activate the immune system to produce large numbers of anti-peptide antibodies, and their

use avoids the immunogenicity and the other disadvantages associated with conventional antigenic systems.

Dendrimer structure is often confirmed with a variety of techniques, including ¹H- and ¹³C-NMR, mass spectrometry, size-exclusion chromatography, high performance liquid chromatography, electrophoresis, elemental analysis, and thermal analysis.^[14,27–35]

PHYSICOCHEMICAL PROPERTIES

Unlike classical polymers, dendrimers have a high degree of molecular uniformity, narrow molecular weight distribution, specific size and shape characteristics, and

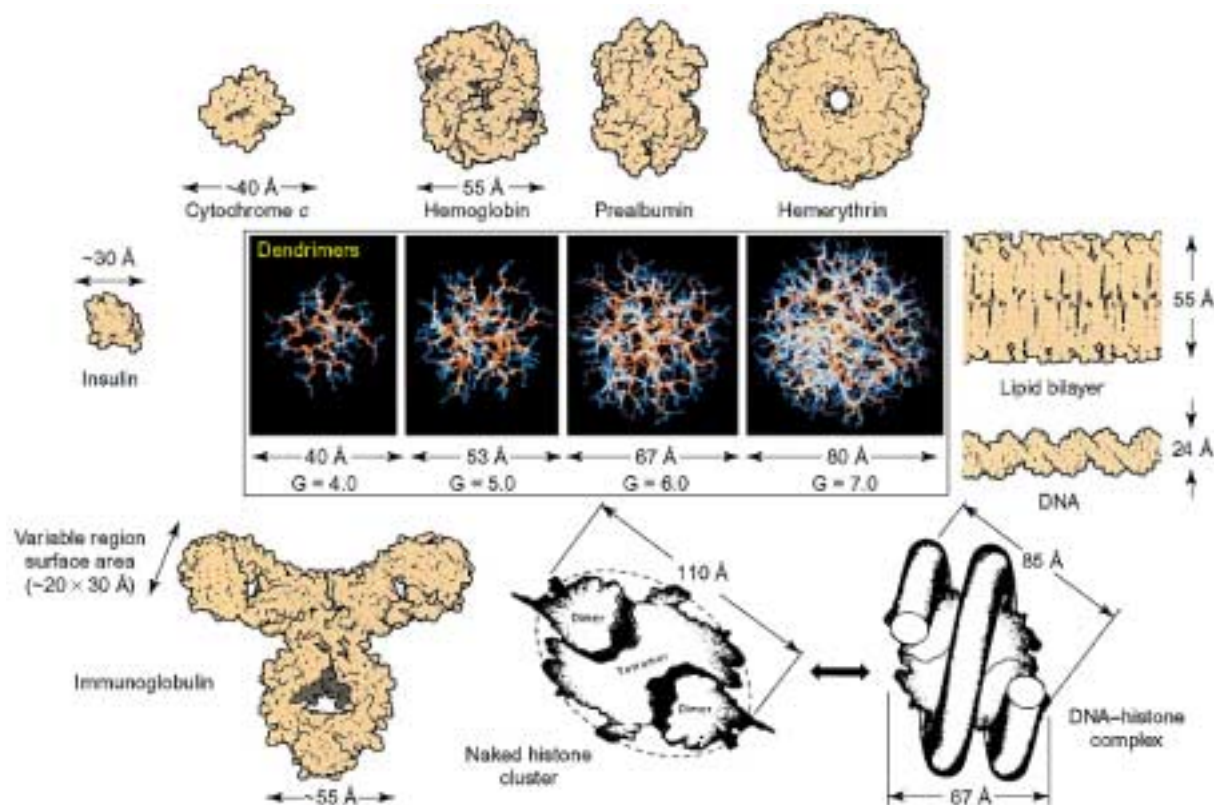


Fig. 2 A dimensionally scaled comparison of a series of PAMAM dendrimers (NH_3 core) with a variety of proteins and bioassemblies. (From Ref. [7], © 2001 Elsevier Science.)

a highly functional terminal surface. The branching nature of the structure can lead to large 3-D globular structures, which at high molecular weights may approximate spheres. These structures are relatively fixed, in marked contrast to linear polymers that are random coils and,

depending on the solvent, may adopt a variety of configurations.

Many of the physical properties of dendrimers may be predicted, at least qualitatively, from molecular modeling of the growth process. With the PAMAM dendrimers,

Table 1 Physical characteristics of PAMAM dendrimers (EDA core)

Generation	Molecular weight	Measured diameter (nm)	Surface groups
0	517	1.5	4
1	1,430	2.2	8
2	3,256	2.9	16
3	6,909	3.6	32
4	14,215	4.5	64
5	28,826	5.4	128
6	58,048	6.7	256
7	116,493	8.1	512
8	233,383	9.7	1024
9	467,162	11.4	2048
10	934,720	13.5	4096

(From Ref. [2].)

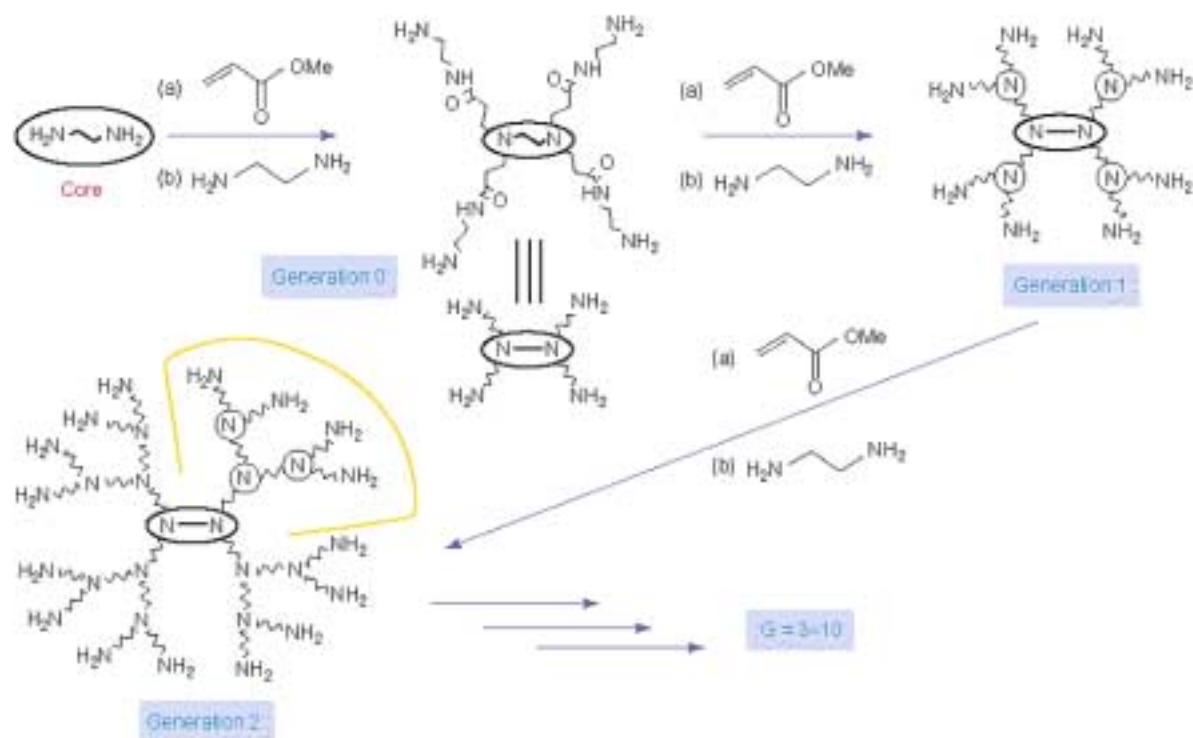


Fig. 3 Synthesis of a PAMAM dendrimer: exhaustive Michael addition of amino groups with methyl acrylate, followed by amidation of the resulting esters with EDA. (From Ref. [7], © 2001 Elsevier Science.)

a fully developed dendrimer structure first appears only after dendrimer growth to at least G1.5, since it is not until the point that the dendrimer contains all three branch cell components, i.e., the core, interior, and surface branch cells. This transition from lightly branched structures to fully developed dendritic structure is referred to as the critical branching stage.^[36] Because of the progressive growth pattern, the dendrimers are constructed in a precise manner and a linear increase of the radii of the fully developed dendrimers with increase of generation is expected. However, because the surface cells amplify according to a geometric progression with increasing generation, it is clear that ideal growth cannot continue indefinitely and there will be a critical generation at which the reacting dendrimer surface will not have sufficient space to accommodate all of the required new units. This stage is referred to as the de Gennes dense-packed state and occurs around G7 in the PAMAM dendrimers.^[37] Dendritic growth beyond this point is of course possible but leads to products of imperfect structure because not all of the surface groups are able to participate in reaction due to steric effects. Such defective generations have been called “hairy dendrimers.”

Similar molecular modeling of the 3-D character of the polyether dendrimers, [37] on the other hand, shows that

steric crowding of the surface branches becomes prohibitively high on approaching G4. As a consequence, it is predicted that the polyether dendrimers will be denser with far fewer internal cavities than the PAMAM series. The PAMAM dendrimers have much higher internal surface area and solvent-filled volume than the polyether dendrimers. Within the PAMAM dendrimer spheres formed at $G > 4$, the solvent accessible surface increases with generations so that the fraction of the internal molecular surface increases from about 29% of the total solvent accessible surface at G4, to 69% for G5, and about 124% for G6.^[37] Thus, at G6 there is more internal surface area than external surface area. For polyether dendrimers on the other hand, the internal surface area reaches a maximum of approximately 20% for G3 and G4. The possibility of binding or entrapping small molecules in these cavities mimics the drug delivery attributes that are offered by liposomes.

Dendrimer Size and Shape

Computer-simulated modeling of the structures of the PAMAM dendrimers^[37–39] highlights changes in external appearance with increase in generation (Fig. 2), through a continuum of molecular shapes ranging from open

amorphous shape for the early generations (G0 to G3) to a more tangled spheroidal network for the fully developed structures at $G > 4$. These shape changes are a consequence of the tethered steric constraints imposed on the developing branches. In Fig. 4, these shape changes are expressed in terms of the aspect ratios of the corresponding longest and shortest principle moments. Experimental determination of the exact shape of dendrimers presents difficulties mainly because most of the commonly used techniques are at their limits of reliability in the size range involved.

Very few experimentally determined transmission electron microscope (TEM) images of dendrimers have been published in the literature, probably because their size range and their fragile organic composition make the resolution of such objects by electron beam techniques very difficult.^[40] Quality images of dendrimer molecules may provide insight into the actual uniformity, sphericity, and hollowness of the macromolecules. TEM studies on PAMAM dendrimers by Tomalia et al.^[1] showed highly monodisperse spheroids. For example, images of the sodium salt of G3.5 dendrimer showed that almost 90% of all particles had diameters ranging within 10% of the average value determined by computer simulations. Jackson et al.^[41,42] examined positively stained PAMAM dendrimers of G5 to G10 using this technique. The shapes of the stained molecules were shown to be spherical for G7 to G10 with some molecules showing "edges." For G5 and G6, the resolution of these molecules was less than with larger dendrimers because smaller molecules took up less amount of stain. The mean diameters of the dendrimers measured by TEM compared well with measurements made with small-angle x-ray scattering (SAXS).

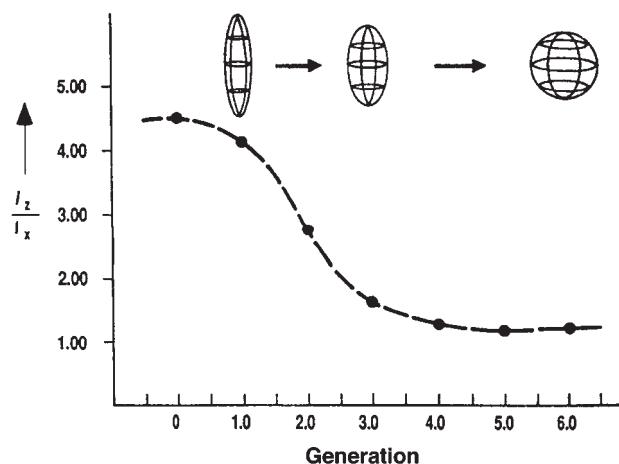


Fig. 4 Comparison of the change in PAMAM dendrimer morphology (aspect ratio I_z/I_x) as a function of generation. (From Ref. [37], © 1990 Wiley-VCH.)

Dendritic structures have been studied on a variety of surfaces including glass,^[43] graphite,^[43,44] and charged solid surfaces^[45,46] using atomic force microscopy (AFM). In a recent study, Zhang et al.^[47] reported the topographical imaging of well-separated, individual polyphenylene dendrimer molecules and their aggregates on mica surfaces using noncontact AFM techniques. The observed height was in good agreement with the size of a single dendrimer molecule as calculated by molecular dynamics simulation. In addition, pulse force mode AFM was used to study the stiffness and adhesion properties of the individual dendrimer molecules to this surface and these were related to the molecular structure and the chemical nature of the outer surface of the dendrimers.

Experimental measurements of the hydrodynamic radius of PAMAM dendrimers with ammonia or EDA cores by size exclusion chromatography and viscometry^[37,39] have shown the expected linear increase of hydrodynamic radius with generation beyond G3. The values were intermediate between the calculated theoretical maximum values that would result from maximum extension of all branches, and the minimum value allowed by acceptable packing of the atomic mass units around the initiator core. The expanded dendrimer would be expected to occur with more ideal solvents; the compressed form would be more likely in theta solvents. A small-angle neutron scattering (SANS) study of the effect of solvent quality on the molecular dimensions of G5 and G8 PAMAM dendrimers^[48,49] showed a decrease of the radius of gyration of the dendrimers by approximately 10% with decreasing solvent quality. The same studies showed that the dimensions of these two dendrimers were constant to within 5% over the temperature range -10 to 50°C .

PAMAM dendrimers have been obtained as very monodisperse samples by careful removal by ultrafiltration of propagating agents and any bridged forms. Measurement by a variety of techniques, notably size exclusion chromatography and low-angle laser light scattering,^[37] SANS and SAXS,^[50,51] and intrinsic viscosity measurement^[39] have shown polydispersity indices for PAMAM dendrimers in the range 1.01–1.08, indicative of a relatively monodisperse distribution of dendrimer sizes in solution. A similar monodispersity of size was noted from SAXS measurements^[52] on dilute methanolic solutions of G3 to G10 PAMAM dendrimers and poly(propyleneimine) dendrimers made with primary amines built around a diaminobutane core (DAB-dendr- $(\text{NH}_2)_x$, $x = 16, 32$, and 64). In contrast, dilute solutions of polyol hyperbranched polymers exhibited scattering indicative of the inherent irregularity of internal segment densities and of overall size, giving more polydisperse solutions. SANS measurements on poly(propyleneimine) dendrimers terminated by primary

amines, DAB-*dendr*-(PA)₃₂ and DAB-*dendr*-(PA)₆₄, in methanol over a wide concentration range (dendrimer mass fraction between 0.01 and 0.80) have been reported by Topp et al.^[53] At concentrations at which the swollen dendrimer volume fraction was below 0.64 (the critical volume fraction for close packing of hard spheres), which corresponds to a dendrimer weight fraction of 0.25, the dendrimers in solution behaved as a dispersion of uniform soft spheres with no significant interpenetration between the segments of different dendrimers. At higher concentrations, however, the dendrimers “collapsed,” their size decreasing with increase of number density so that the volume fraction of the solution was maintained at approximately 0.64.

Surface Properties of Dendrimers

The surface properties of dendrimers are specific to the functional terminal groups that make up the surface of the dendrimers, which can be either reactive or passive moieties or even a combination of both. The type and number of functional groups and their ionization characteristics may affect dendrimer solution properties in various ways, e.g., by changing the solubility, potential for aggregation, and inter-dendrimer charge interaction. A dendrimer surface may contain multiple copies of a particular functional group and so would be an ideal molecule for substrate binding. Under conditions where the functional groups are ionized, the dendrimer becomes a macromolecular polyelectrolyte and as such it will interact strongly with oppositely charged particles. Polyelectrolyte dendrimers have been shown to adsorb strongly at various interfaces such as alumina/water and silica/water as well as associating with proteins or DNA.^[54–57]

Although dendrimers possessing a wide variety of functional groups have been synthesized, the most studied are those with amine or carboxyl groups. A titration curve of a 2% aqueous solution of a G4 PAMAM dendrimer against 0.2 M HCl^[58] shows two distinct inflection points that are indicative of the titratable terminal primary amines (pK_a 10.7) and the interior amines (pK_a 6.5). The titration characteristics have been confirmed by observation of ¹³C chemical shifts.^[1] The surface potential of the G4 PAMAM dendrimer calculated from titration curves is 63 mV.^[58]

Charge Interactions and Aggregation

A pronounced influence of pH and added electrolyte on the charge interactions between dendrimers with ionizable groups is expected in aqueous solution. Briber et al.^[50] investigated the intermolecular interactions between PAMAM dendrimers in concentrated solutions using

SANS and SAXS techniques. They showed that the molecules develop large scale interactions which can be screened by the addition of excess acid or electrolyte. SANS studies of the structural properties of poly(propyleneimine) dendrimers (DAB-*dendr*-(NH₂)₆₄) in D₂O examined the influence of the degree of ionization and solvent pH.^[59,60] Upon addition of HCl, the dendrimers become charged and the scattering patterns exhibited a single correlation peak indicating a spatial arrangement of the molecules due to the electrostatic repulsion. In the uncharged state, these dendrimers behave as “soft” molecules, i.e., they have a very high degree of flexibility with possible interpenetration at high concentration. In conditions where they are charged, however, the electrostatic forces cause a stretching of the branches and reduce their flexibility, i.e., they behave as “hard” particles.

Changes in the G6 PAMAM structure upon pH titration were studied using a polarity-responsive probe by Chen, Tomalia, and Thomas.^[61] Two possible models were proposed to explain the dendrimer behavior; in the first, some inward folding of the dendritic termini is presumed to occur at all basic pHs, resulting in a slightly less polar interior. At high pH (ca. pH 10) any protonation will occur on the more polar surface amines; as the pH is lowered to pH 8.3, the less polar amines will protonate and the average environmental polarity decreases, as was observed from the experimental measurements. Finally, further decrease of pH causes protonation of the internal amines and molecular expansion due to charge repulsion. In the alternative model, increasing protonation of dendritic termini is able to overcome entropy, forcing charged termini to the periphery; the resulting surface crowding results in the observed lower average environmental polarity. In this model, the polarity increase on decreasing pH below 8.3 arises from protonation of the amines that are located in the less dense, more polar dendritic interior. The first model is considered a denser core model of the G6 dendrimer, while the second is a denser shell model.

A detailed study of the physicochemical properties of G5 PAMAM dendrimers in aqueous solution by Nourse, Millar, and Minton^[35] revealed contrasting properties of amino-surface (G5-NH₂) and hydroxyl-surface (G5-OH) dendrimers. The hydrodynamic properties of G5-OH in dilute solution could be described by a model in which the solute was represented by a single species of quasi-spherical particle having a molar mass equal to the theoretically predicted molar mass of the dendrimer and a hydrodynamic radius of about 3.1 nm. In contrast, size exclusion chromatography and sedimentation equilibrium ultracentrifugation studies on G5-NH₂ have indicated the formation of oligomeric aggregates in aqueous solution even in the presence of high salt concentration. Measurement of the concentration dependence of

sedimentation of G5-OH in pH 7.2 phosphate buffer indicated the presence of significant electrostatic repulsion overlaid on weakly attractive interactions leading to the formation of nonspecific aggregates at sufficiently high dendrimer concentration.

Static and dynamic light scattering studies by Milhem^[58] have shown that G4 PAMAM dendrimers do not exhibit intermolecular aggregation in aqueous solution at concentrations up to 7% w/w at pH 10.5, or even when fully unionized at pH 14.

Internal Structure

Many of the potential applications of dendrimers depend directly on the organization and distribution of internal segment densities and the possibility of reduced density at the core. The first theoretical treatment of internal structure of the amine-terminated PAMAM dendrimers by de Gennes and Hervet^[8] predicted a segment distribution function that has the highest density on the periphery and a relatively hollow core. On the other hand, Monte Carlo simulations by Lescanec and Muthukumar^[62] predict that the highest density is at the center with a decaying profile to the edge of the molecule. Both the models have shortcomings; the de Gennes model does not account for backfolding, which could be an incorrect assumption for flexible dendrimers, whereas the Lescanec and Muthukumar model is based on kinetically grown rather than equilibrium structures. The predictions of these models have been tested against the experimental data from a variety of techniques.

The SAXS studies on methanolic solutions of G3 to 10 PAMAM dendrimers by Prosa et al.^[52] indicate that lower generation dendrimers are less dense than their higher generation relatives, but after the first few generations the average density appears to be roughly independent of generation and uniform throughout the structure with no indication of any sizeable minimum in density near the dendrimer core (with the G10 dendrimer) contrary to the predictions of the de Gennes model.

Measurements of molecular density^[39] and intrinsic viscosity^[37,39] of PAMAM dendrimers indicate an unusual variation with dendrimer generation. Minimum density and maximum intrinsic viscosity were observed at around G4, which suggests that the fully developed dendrimers have a high accessible internal surface area in a solvent-filled intramolecular free volume that may consist of internal cavities and channels. Similar findings were reported by Mourey et al.^[63] for PBE monodendrons (based on dihydroxybenzyl alcohol) and tridendrons produced by coupling these dendrons to a trifunctional core, 1,1,1-tris(4'-hydroxyphenyl)ethane prepared by the convergent method. A maximum intrinsic viscosity occurred at G3 for tridendrons and G5 for monodendrons,

consistent with the model developed by Lescanec and Muthukumar that predicts inward folding of branch units, a maximum density in the center, and a distribution of terminal groups throughout the structure. If there are no specific interactions between the groups and if the branch units of the dendrimer are flexible enough to allow some gradual backfolding, then it is reasonable that the equilibrium structure has maximum density in the center because this corresponds to maximum entropy for the molecule and also provides relief of steric crowding of terminal groups. It is also significant that this model describes the behavior of both PAMAM and polyether dendrimers implying that it may be of general application to all flexible dendrimers.

A study of the molecular characteristics in dilute solution (D_2O , 1%) of the first five generations of the poly(propyleneimine) dendrimers DAB-*dendr*-(NH_2)_x and DAB-*dendr*-(CN)_x, ($x = 4, 8, 16, 32$, and 64) by Scherrenberg et al.^[64] has suggested some degree of backfolding of molecules within the dendrimer. Comparison with molecular dynamics calculations indicated that the poly(propyleneimine) dendrimers could be regarded as flexible molecules with a relatively homogeneous density distribution, not in line with either the de Gennes or Lescanec and Muthukumar models.

There is considerable evidence from a variety of techniques for a dendrimer model composed of a relatively soft or spongy interior surrounded by a considerably harder outer molecular surface, the so-called "dendritic box."^[17] ¹³C-NMR measurements of spin lattice relaxation times of specifically tagged PAMAM dendrimers^[65–68] have shown considerably reduced mobility in the outer surface groups relative to the interior segments. Meltzer et al.^[65,66] performed ¹³C-NMR relaxation studies on hydroxyl terminated PAMAM dendrimers, in order to understand the influence of the end groups on the chain dynamics of the molecule (G0.5 to G10.5). The spin relaxation times (T_1) for each terminal carbon decreased continuously with an increase in generation, suggesting that segmental motion was increasingly restricted at the surface of the higher generations. Thus, it was concluded that the iterative branching process that creates the PAMAM architecture must at some point lead to severe steric crowding either at the molecular surface or throughout the volume of the molecule, as indeed is indicated from modeling of dendrimer growth. It was, however, noted that the mobility of the terminal groups on the surface of dendrimer molecules was higher than expected for the available area. Therefore, either the number of terminal groups must be much less than expected, possibly due to growth failure, or the conformation must be such that all the terminal groups do not reside at the surface due to their folding back into the interior of the molecule. Gorman et al.^[69] also showed

that some terminal groups (approximately 3) of a G3 dendrimer were located very close to the dendrimer core. An average of 25 of the 32 terminal groups were found on the geometric periphery which were more mobile than the others found elsewhere in the dendrimer molecule.

Differential scanning calorimetric (DSC) measurements on a variety of dendrimers including PAMAM dendrimers with NH_3 and EDA cores,^[39] and PBEs and phenolic-terminated polyesters^[27] have shown exponentially increasing glass transition temperatures with increasing molecular weight reaching an asymptotic value generally at about G3 or G4. These results show that a pronounced segmental mobility is retained within the internal volume, possibly around the core and over the first two or three generations in support of the conclusions from ^{13}C -NMR relaxation studies. Moreover, at about the fifth branch layer around the core, segmental motions responsible for the glass transition apparently reach their limiting values and cannot be extended any further.

Flow Properties and Inter-dendrimer Interaction

The flow properties of solutions of dendrimers have relevance not only for their possible use in pharmaceutical formulation, but also because of the insight that might be gained on the nature of any intermolecular interactions between dendrimers. It might be expected from computer modeling of the dendrimer shape that lower generations, which have an open "plate-like" or "dome-like" entity, would readily permit a branch from a neighboring dendrimer to penetrate into the interior. This tendency would be further enhanced by intermolecular hydrogen bonds between the interior amide groups of two interpenetrating molecules or between the primary amine units of one molecule and the amide carbonyl oxygens of another. In contrast, the closure of the dendrimer outer surface at higher generations should result in minimal inter-dendrimer interactions.

Measurements by Uppuluri et al.^[36] using a cone and plate rotational rheometer on the first seven generations of EDA-core PAMAM dendrimers in medium and highly concentrated solution (30%–75% in EDA) showed typical Newtonian flow behavior over the entire range of shear stress and strain examined. An absence of any abrupt change in the slope of log zero-shear viscosity against log weight average molecular weight relationship was indicative of a lack of any significant interpenetration of dendrimers to form quasi-networks even at such high concentration where any such inter-dendrimer interaction would be accentuated. Similar measurements on neat linear polyamidoamine illustrated the unique bulk flow properties of the dendrimers compared with a chain polymer of comparable molecular weight, which showed

an inflection in these plots associated with the onset of chain entanglement. There is also a fundamental difference in the flow properties of these dendrimers and colloidal suspensions of spheroidal particles, which typically show shear thinning properties above some critical shear rate, molecular weight, or concentration originating from a tendency of individual particles to aggregate. The viscosities of the dendrimer solutions were at least an order of magnitude lower than those usually found for similar solutions of chain-type macromolecules of comparable molecular weights and concentrations, i.e., they flowed much more easily than their chain-type equivalents. It was proposed that the atypical flow properties of the dendrimers compared to other high molecular weight synthetic polymers were a consequence of the unique dendrimer architecture, which above the critical generation results in globular spheroids whose outer surfaces close upon themselves or at least become dense enough to be impenetrable for other dendrimers or large molecules, giving rise to the Newtonian flow properties of these solutions. The situation is thus the reverse of that of typical long-chain polymers, where the probability of entanglement increases with increase of chain length. A pronounced sensitivity of solution viscosity to temperature and a stability of the dendrimer solutions to repeated loading, observed by these workers, suggested surprising flexibility of the dendrimer interior, in agreement with their reported ability to expand^[70] or shrink^[71] depending on solvent quality, solution concentrations, and temperature. These observations support a model of the dendrimer having a soft or spongy interior and a surrounding denser more congested shell.

A detailed rheological examination of the first eight generations of PAMAM dendrimers^[72] explored rheological behavior in the bulk (rather than solution) state under conditions of steady shear, shear creep, and dynamic oscillatory shear over a wide temperature range. A distinct change from single relaxation mode to a multirelaxation mode, Maxwell-type behavior at G4 was consistent with the soft interior-dense shell model involving the closure of the dendrimer molecular surface upon itself. Furthermore, an analysis of the free volume of PAMAM dendrimers revealed a dependence on generation that was consistent with the qualitative conformational change at G4 predicted by this model.

Further insight into the nature of the dendrimer surface has been provided by measurements of the melt viscosity of PBE monodendrons and tridendrons.^[73] Logarithmic plots of melt viscosity against molecular weight were linear with a slope close to the expected value of unity and no detectable inflection attributable to onset of chain entanglement. These observations confirm that the melt viscosity behavior is not dominated by chain entanglements and is fully consistent with a globular highly

branched structure with an increasingly congested surface lacking intermolecular chain entanglements.

PHARMACEUTICAL APPLICATIONS OF DENDRIMERS

Encapsulation/Solubilization of Drugs by Dendrimers

The concept of encapsulating guest molecules into a special, egg shell-like structure was proposed by Maciejewski in 1982.^[74] There has been a considerable interest in the use of dendrimers to encapsulate or interact with labile or poorly soluble drugs in order to protect the drug and enhance bioavailability. Such guest–host dendrimer systems have also been proposed as controlled delivery systems. Although several terms have been applied to systems in which drugs have been encapsulated within a dendritic structure, these systems are usually only distinguishable by subtle differences in dendrimer architecture. The interaction within the dendrimer may

be simple physical entrapment, or can be more specific involving hydrophobic interactions or hydrogen bonding.

Dendritic structures having a hydrophobic core and hydrophilic surface layer have been described as unimolecular micelles.^[14,15,75–79] However, unlike conventional micelles, the dendritic structure is independent of dendrimer concentration; thus, they do not possess a critical micelle concentration. Newkome et al. described a monomolecular micelle cascade polymer based on a neopentyl core and 36 carboxylic acid surface groups.^[77] A number of lipophilic probes (e.g., phenol blue) were found to be associated within the lipophilic infrastructure of the micellanoates whose size ranged 3 nm–5 nm in diameter. Hawker et al. described the convergent synthesis of dendritic polyether unimolecular micelles based on an 3,5-dihydroxybenzyl alcohol building block with carboxylate surface groups.^[14] The dendritic micelles were capable of specific nonbonding interactions (π – π interactions in this case) through tailor-made molecular inclusion sites. The electron-rich macromolecular structures (Fig. 5) were able to solubilize a variety of polycyclic aromatic compounds in water. The solubilizing power of

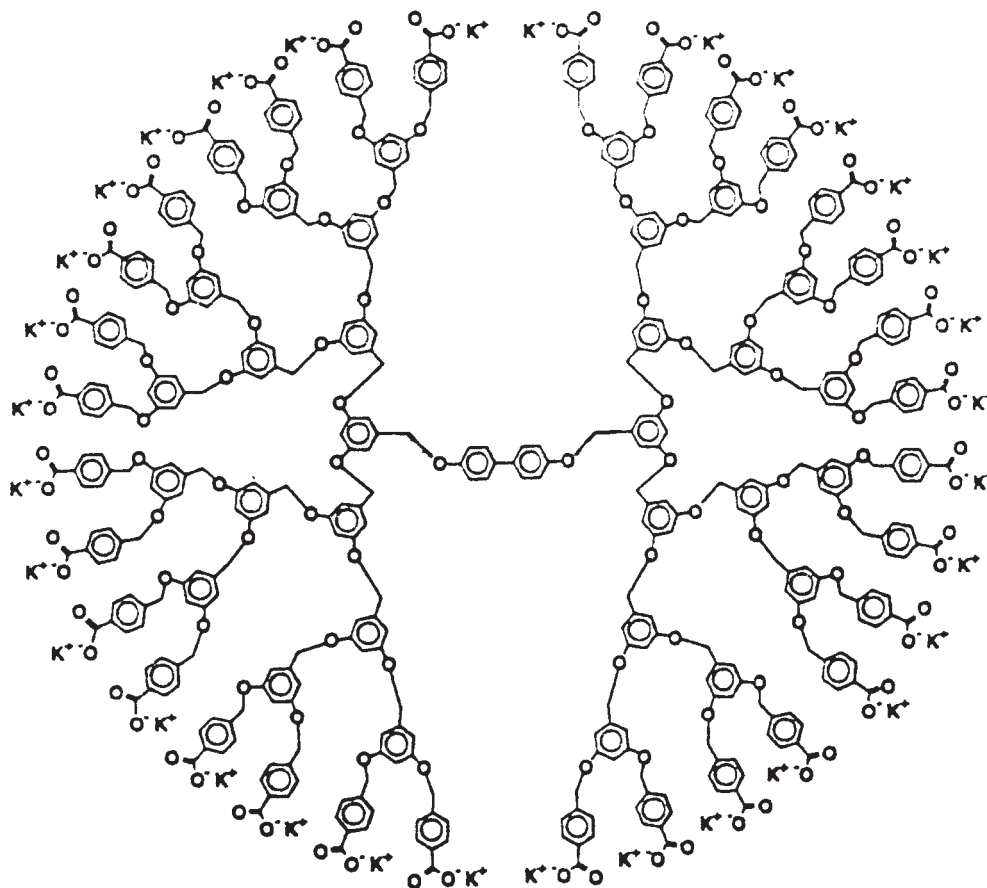


Fig. 5 Water soluble unimolecular dendritic polyether micelle. (From Ref. [14], © 1993 Royal Society of Chemistry.)

the dendrimer was found to increase with dendrimer concentration and also as the electron density of the polycyclic aromatic compounds increased. Hybrid dendritic structures were also described in which hydrophobic and hydrophilic chain ends were segregated at distinct ends of the globular structure, thus allowing for preferential orientation at certain interfaces. Dendrimers based on 3,5-dihydroxybenzyl alcohol have limited solubilization capacity as a result of small internal cavities within the dendrimer. Liu et al. synthesized water-soluble dendritic unimolecular micelles based on 4,4-bis(4'-hydroxyphenyl)pentanol building blocks and a surface shell of polyethylene glycol (PEG) chains.^[15,32] The building block was chosen in order to increase the flexibility and cavity size of the dendritic structure while the PEG was chosen because of its good water solubility and demonstrated biocompatibility (Fig. 6). The solubilization capacity of the unimolecular micelle was demonstrated by the solubilization of pyrene (~365-fold increase in solubility) in aqueous solution and entrapment of a model drug (indomethacin) at a loading of 11% w/w

(G3 micelle), a value that corresponds to approximately nine drug molecules per micelle. The drug-loaded dendrimer provided sustained release of indomethacin over a period of 30 hr.

Jansen et al. describe the entrapment of molecules in a dendritic box,^[17,80–82] based on poly(propyleneimine) dendrimers with a chiral shell of protected amino acids (Fig. 7). The resulting dendritic structure, 5 nm in size, possesses a dense shell (as a result of bulky surface groups) with solid-phase character. Guest molecules were entrapped within the internal cavities of the dendrimer. The dense outer shell prevents diffusion from the dendritic box, even after solvent extraction, prolonged heating, or sonication. The entrapment effect was only seen with higher generation structures, e.g., a dendrimer with 64 amine end groups entrapped up to four Bengal Rose molecules. An extension of this work involved the shape-selective liberation of guests from dendritic boxes.^[81] Hydrolysis of the surface t-BOC groups (with formic acid) of a dendritic box containing entrapped Bengal Rose and 4-nitrobenzoic acid resulted in perforation of the dendrimer

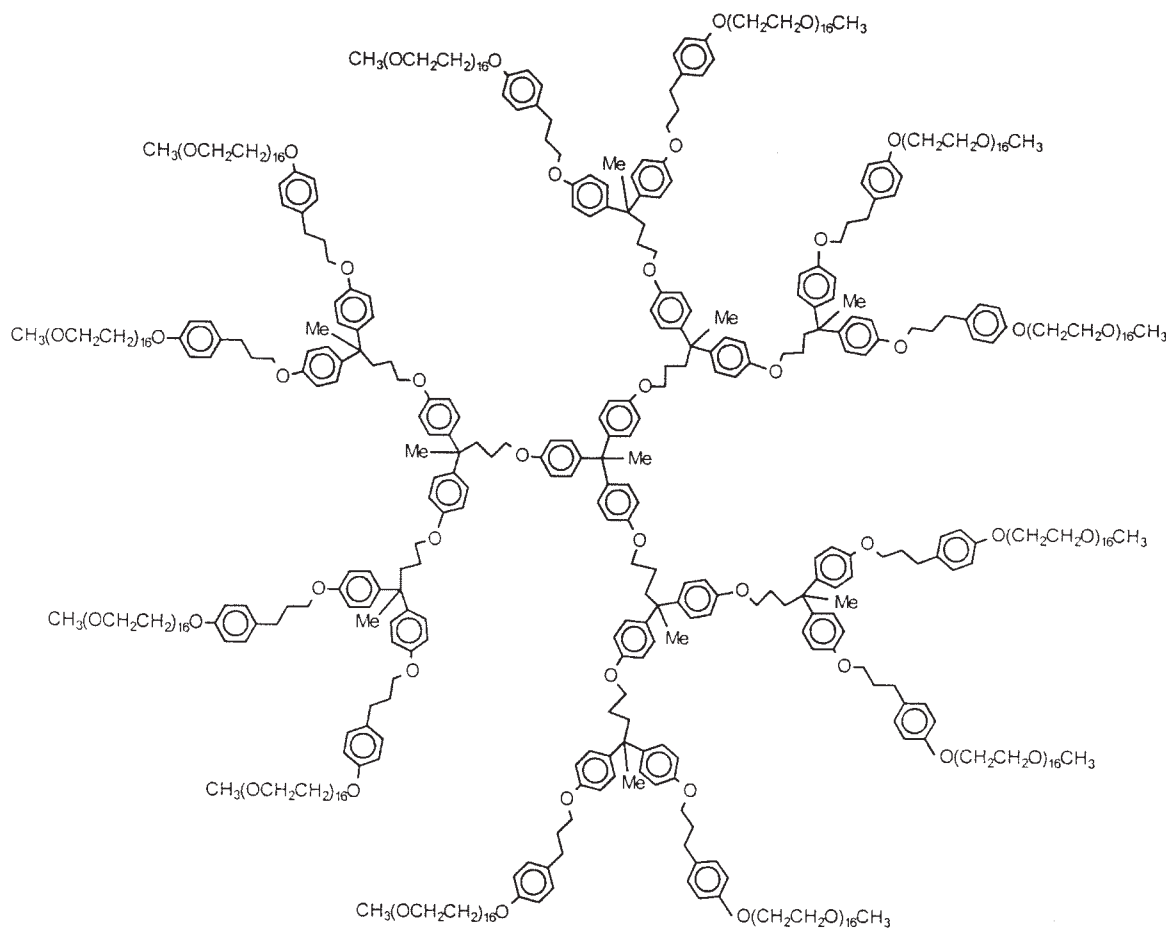


Fig. 6 G2 dendritic unimolecular micelle. (From Ref. [15], © 2000 Elsevier Science.)

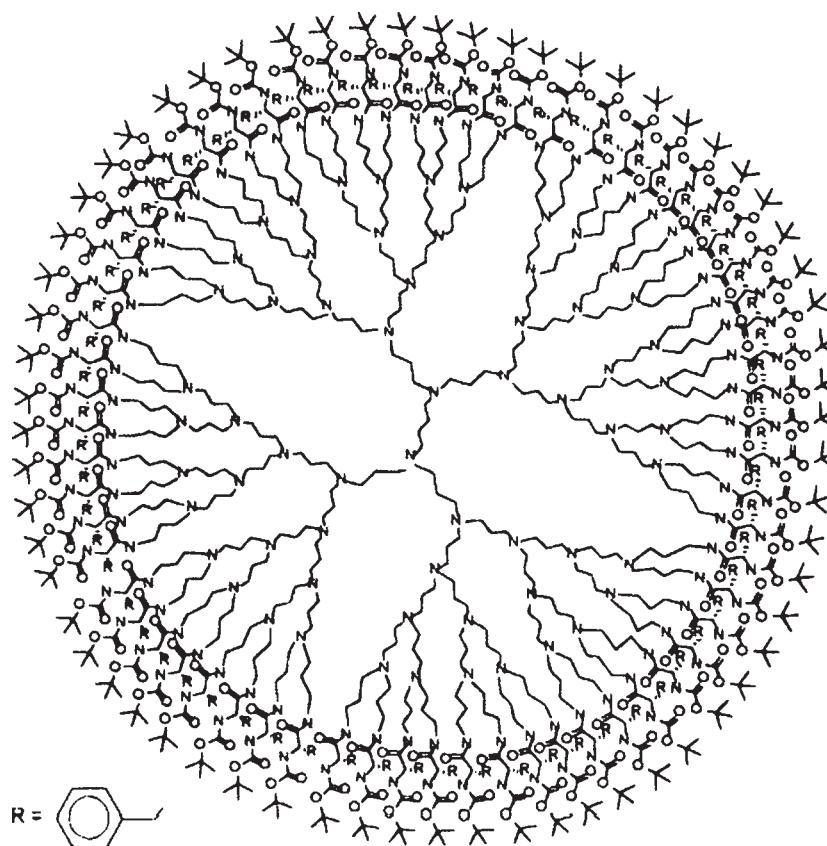


Fig. 7 Dendritic box based on poly(propyleneimine) dendrimer. (From Ref. [81], © 1995 American Chemical Society.)

allowing the release of 4-nitrobenzoic acid but not of Bengal Rose. The larger Bengal Rose molecule could only be liberated following hydrolysis of the outer shell by 12 M HCl under reflux for 2 hr.

An approach to increase the loading of guest molecules within a dendritic structure involves the creation of additional void space by removal of the dendrimer core, producing the so-called cored dendrimers.^[12,83] This approach obviously requires a means of maintaining the structural integrity of a cored dendrimer and thus the connections among the dendritic wedges. The approach taken by Zimmerman's group was to extensively cross-link the dendrimer surface groups by a ring-closing metathesis reaction prior to removal of the core (Fig. 8). Core removal was via three cleavable ester bonds, with the remaining structure being unaffected as a consequence of robust ether linkages. Cored dendrimers have been compared to hollow polymeric nanospheres with the potential to encapsulate substances. It is not clear, however, how guest molecules could be loaded into such structures.

A recent report focusing on the encapsulation of drugs referred to two types of possible systems, such as

the Dendrilock[®] and Dendripore[®] type structures.^[84] The Dendrilock structure is based on a dendrimer with a congested outer shell, the release of guest molecules from which was either immeasurably slow or nonexistent. In contrast, Dendripores are lower generation dendritic host structures with less compact surfaces allowing time dependent release of guest molecules. The proposed mechanism of release is similar to that reported by Meijer's group,^[17,80–82] i.e., dendrimer surface density is used as a means of controlling the release of guest molecules.

Kojima et al. synthesized G3 and G4 PAMAM dendrimers with PEG grafts and examined their ability to encapsulate the hydrophobic drugs adriamycin and methotrexate.^[85] The modified PAMAM dendrimers were reported to have PEG grafted to every surface group and were estimated to have a hydrodynamic diameter of up to 14.7 nm. The purpose of the grafts was to improve biocompatibility and modify biodistribution. It was found that drug loading increased with dendrimer size and increasing chain length of PEG grafts, and up to 6.5 adriamycin or 26 methotrexate molecules could be

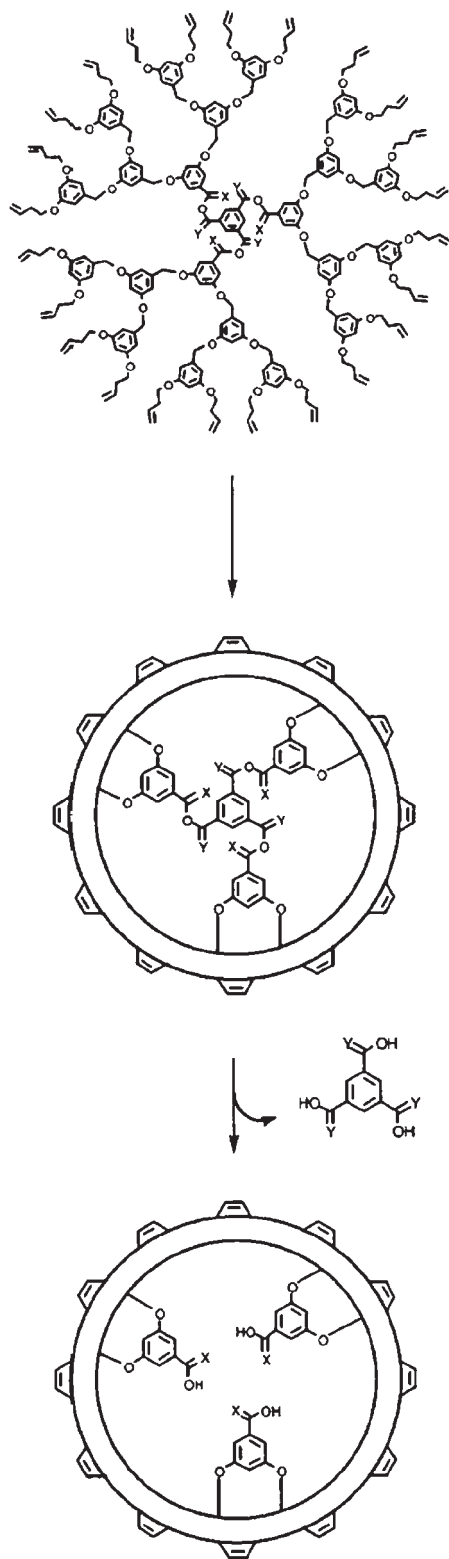


Fig. 8 Synthesis of a cored dendrimer. (From Ref. [12], © 1999 American Chemical Society.)

incorporated into a dendrimer. While there was evidence of the sustained release of methotrexate from a dendrimer carrier in an aqueous solution of low ionic strength, no control could be achieved in isotonic solutions.

Although the number of guest molecules incorporated into a dendrimer may be dependent to a limited extent on the architecture of a dendrimer, the loading capacity may be dramatically increased by the formation of a complex with the large number of groups on the dendrimer surface. The interaction between charged dendrimers and oppositely charged polyelectrolytes may result in the formation of soluble complexes.^[86–91] Studies on the interaction of poorly soluble drugs have shown that solubility may be dramatically enhanced in the presence of dendrimer.^[92–95] The solubility of ibuprofen, for example, was increased by a factor of over 140-fold in the presence of a 2% w/v PAMAM G4 dendrimer solution, over twice that achieved using a micellar solution of 2% w/v sodium dodecyl sulfate. The dendrimer–drug complexes were thought to involve an electrostatic interaction between the carboxyl group of the ibuprofen molecule and the amine groups on the dendrimer surface, though recent studies indicate the additional involvement of internal amine groups.^[58] Although enhancing solubility, complexes formed as a result of electrostatic interactions may not provide sufficient control over drug release.^[93] In order to provide a more controlled release of drugs, covalent conjugates have been synthesized between drugs and the numerous surface groups on a dendrimer. The hydrophobic drug ibuprofen was coupled onto the surface of G4 PAMAM dendrimers.^[58,96,97] It was found that insoluble complexes were formed when more than five ibuprofen molecules were attached to the dendrimer surface. However, much higher drug loadings (up to 32 ibuprofen molecules per dendrimer) could be achieved when PEG chains were attached to the dendrimer. The release of ibuprofen from these dendrimer conjugates was examined. Liu et al. characterized a number of conjugates based on polyether dendrimers and a number of model drugs (cholesterol and two amino acid derivatives).^[32] The solubility of the conjugates was maintained by grafting short PEG chains onto the dendrimer surface. Zhuo et al. synthesized a series of dendrimers with a branch structure similar to PAMAM dendrimers, but with a cyclic core of 1,4,7,10-tetraazacyclododecane.^[34] The resultant dendrimer had four branches emanating from the core. 5-fluorouracil (SFU), conjugated onto the surface of the dendrimers, could be released by hydrolysis of the resulting conjugate.

Uptake of Dendrimers

A polymeric drug carrier such as a dendrimer must traverse through various cell barriers, i.e., epithelial and endothelial cells, before reaching its target. Transepithelial

transport is an important factor when considering movement across the intestinal cell wall, while transendothelial transport is relevant to the movement of carriers across the vasculature to reach their target cells/receptors. Generally, transport across biological barriers is determined by factors pertaining to both the nature of the carrier (e.g., drug molecule, macromolecule, and particle) and the nature of the biological barrier. Important carrier attributes are molecular weight, charge, hydrophobicity, flexibility, and the geometry of the molecule. Biological barrier attributes include the barrier's location and function; the blood-brain barrier is, for example, less permeable to molecules than the intestinal epithelium.

Proposed mechanisms of uptake of structures of dendritic size from the gastrointestinal tract include persorption, endocytosis by enterocytes, paracellular transport, uptake by intestinal macrophages and uptake through the gut-associated lymphoid tissue.^[98] Wiwattanapatapee et al.^[99] investigated the potential of 3 nm–7 nm diameter anionic and cationic PAMAM dendrimers to transverse an everted rat intestinal system. It was found that the uptake of PAMAM dendrimers was most likely to occur across enterocytes by transcytosis. It was also reported that cationic (full generation) PAMAM dendrimers are retained in the tissue. This retention could be a consequence of adsorption of positively charged dendrimers (as well as other polycations) onto negatively charged cell membranes, an effect also reported by Katchalsky.^[100] Anionic PAMAMs were found to have high serosal transfer rates with potential as an oral drug delivery system.

Florence and coworkers studied the uptake, after oral administration, of a G4 poly(lysine) dendrimer (diameter = 5 nm) modified with a lipid surface.^[98,101] In vivo studies indicated that there was a rapid uptake of dendrimers from the gastrointestinal tract, demonstrating their potential as particulate delivery systems. It was found that uptake of the dendrimers was not uniform along the gastrointestinal tract; there was preferential uptake through lymphoid tissue in the small intestine, but no uptake in lymphoid tissue in the large intestine. Peyer's patches in the small intestine preferentially absorb dendrimers over enterocytes, whereas the opposite is true in the large intestine. The level of uptake of the dendrimers was lower than that exhibited by polystyrene particles in the size range 50 nm–3000 nm. It was suggested that there was an optimum size for nanoparticulate uptake by the gut and that a reduction in size did not necessarily correlate with an increase in uptake.

Tajarobi et al.^[102] studied the transport of a series of full generation PAMAM dendrimers (G0 to G4) across Madin-Darby canine kidney (MDCK) cells. It was found that the permeability of dendrimer across these cells was in the order $G4 \gg G1 > G0 > G3 > G2$ and was governed

by the balance between the size of the dendrimer and its interaction with the cells. One of the factors attributing to the high permeability of the G4 dendrimer was the fact that it compromised cell integrity. Zhang and Smith^[103] also showed that higher generation PAMAM dendrimers may disrupt or influence the integrity of the cells.

The factors influencing the extravasation of PAMAM dendrimers across microvascular network endothelium was reported by El-Sayed et al.^[104] It was concluded that the rate of transport is dependent on size and molecular weight, as well as physicochemical properties such as geometry and charge. An increase in size of PAMAM dendrimers resulted in a corresponding exponential increase in extravasation time. Cationic dendrimers interact with the negative cell walls, and their extravasation time is therefore increased. The balance between charge and geometry was illustrated by the fact that at physiological pH, the polycationic PAMAM dendrimers exhibit faster extravasation than neutral PEG molecules of the same radius.

A certain degree of caution is required in the interpretation of uptake data where dendrimers are conjugated to a label. As reported by Yoo and Juliano, a dendrimer labeled with a fluorescent dye (Oregon Green 488) was a much better delivery agent for antisense compounds than unmodified dendrimer.^[105] It is also important to fully characterize dendrimers as there is a concern regarding polydispersity, particularly with higher generation dendrimers.^[103,104]

Delivery of DNA and Oligonucleotides by Dendrimers

Dendrimers have emerged as efficient carriers of DNA for the transfection of cells.^[29,106–115] A study by Haensler and Szoka^[106] demonstrated the high efficiency transfection of a variety of suspension and adherent cultured mammalian cells using plasmid-PAMAM dendrimer complexes. A maximal value of transfection was obtained by varying complex size and plasmid/dendrimer ratio. Expression was found to be unaffected by lysomotrophic agents and transfection efficiency was increased by over two orders of magnitude when a membrane-destabilizing peptide was attached to the dendrimer. The high transfection efficiency of dendrimers was believed to be related to dendrimer size, shape, and ability to buffer pH change in the endosomal compartment. The dendrimers were found to be well tolerated by cells and less toxic than poly(lysine). It is believed that polycations such as dendrimers act as a scaffold to condense DNA (the G6 dendrimer used in the study has a diameter similar to the histone core of chromatin, 7 nm). Efficient internalization is a result of the net positive charge of the DNA-polycation complex interacting with negatively charged cell-surface groups. Kukowska-Latallo et al.^[107] reported much higher levels of

transfection mediated by dendrimer compared to poly(lysine), and noted that different types of cell might be specifically transfected with different types of dendrimers depending on the cellular metabolism of the DNA–dendrimer complex. Lysomotropic agents were found to have a beneficial effect on transfection. The difference in levels of transfection achieved in the two studies was attributed to the possible presence of impurities in the commercial dendrimer preparations used in the earlier study by Haensler and Szoka. A subsequent study from the Szoka group^[29] reported that monodisperse PAMAM dendrimers in fact produce low levels of transfection whereas partially degraded dendrimers produced much higher levels of transfection (> 50-fold). The higher levels of transfection activity were attributed to the high molecular weight degraded component (fractured dendrimer) being less sterically constrained. The increased flexibility of the dendrimer branches enables the fractured dendrimer to be compact when complexed with DNA and to swell when released from DNA in cells. A comparison of the transfection efficiency of dendrimers to a range of polycationic gene transfer agents indicated that degraded dendrimers displayed high transfection activity with relatively low cytotoxicity.^[116] Polyethylene glycol conjugated low generation PAMAM dendrimers have been shown to be more efficient transfection agents than partially degraded (fractured) dendrimers.^[115] It is believed that the presence of PEG chains on PAMAM dendrimers would mimic some of the essential properties of fractured high generation dendrimers (enhanced flexibility of chains). The use of lower generation dendrimers and the addition of PEG chains also enhances biocompatibility.

Dendrimers have been successfully used as a delivery system for antisense oligonucleotides (ODNs).^[105,117–120] Hughes et al.^[117] reported the significant enhancement of the effectiveness of antisense ODNs in the presence of non-toxic concentrations of G5 PAMAM dendrimer. Bielinska et al.^[118] reported the enhanced activity of antisense ODNs and antisense cDNA plasmids in the presence of PAMAM dendrimers in an *in vitro* cell culture system. Dendrimer was believed to enhance the transfer of ODN into cells. The electrostatic binding of the phosphodiester ODNs to dendrimers also extended their intracellular survival. Dendrimers at concentrations required to be effective ODN carriers were reported to be noncytotoxic.

Cancer Therapy Based on a Dendritic Platform

A number of studies have examined the use of a dendrimer drug carrier to treat a variety of tumors. One approach has been based on the exploitation of the enhanced permeability and retention effect (EPR effect) to localize drug conjugates in tumor tissue.^[121] A second approach

has involved the conjugation of a drug-loaded dendrimer to a targeting moiety such as an antibody or a receptor-specific molecule (ligand) to target specific tumor cells. Structurally, dendrimers are multibranched and can therefore carry a relatively large payload of drug, as well as other bioactive macromolecules on their surfaces or interior. They are also monodisperse systems and the size of conjugates may be calculated with relative accuracy.

Cisplatin is an important anticancer agent, but as with other platinate drugs it suffers from low water solubility, severe toxic side effects, and the inherent or acquired resistance seen in a number of tumors. Malik et al. conjugated a G3.5 PAMAM dendrimer with a carboxylate surface to cisplatin to produce a dendrimer–platinate that had a high drug loading (~ 25%), was highly soluble and released platinum slowly *in vitro*.^[122] *In vivo* studies demonstrated that intraperitoneal administered dendrimer–platinate and cisplatin were equally effective in mice bearing L1210 tumors, whereas only the dendrimer–platinate was active in mice bearing B16F10 tumors. When administered intravenously to treat a palpable subcutaneous melanoma, the dendrimer–platinate displayed antitumor activity, whereas the cisplatin was inactive. It was shown that the dendrimer–platinate selectively accumulated in solid tumor tissue by the EPR effect, and that the conjugate was also significantly less toxic than cisplatin.

Kono, Liu, and Fréchet^[123] reported the synthesis of polyether-based dendrimers conjugated to methotrexate or folate residues, which may be used to target the folate receptor that is overexpressed in almost all solid tumors. Wiener et al.^[124] also used folate-conjugated PAMAM dendrimers to target tumors, and showed that these dendrimers were accumulated in cells in a receptor specific manner.

Other examples exist of dendrimers (particularly PAMAM dendrimers) being conjugated to antibodies to target tumors. Barth et al.^[125] synthesized boronated PAMAM dendrimers conjugated to the IB16-6 monoclonal antibody, which is directed against the murine B16 melanoma. The conjugate was designed for boron neutron capture therapy. The immunoconjugates showed a high level of immunoreactivity *in vitro*, but *in vivo* were localized in the liver and spleen. The same group conjugated epidermal growth factor (EGF) onto a boronated G4 PAMAM dendrimer to target brain tumors.^[126] When administered intrathecally the conjugate was selectively delivered to EGF receptor positive gliomas, but when delivered intravenously the conjugates were once again localized in the liver and spleen, and very little crossed the blood–brain barrier. Abu-Rmaileh et al. proposed a drug loaded PAMAM based dendrimer conjugated to an antibody fragment specific to tumor cells bearing the carcinoembryonic antigen.^[127]

The conjugates had PEG chains grafted onto their surface to minimize uptake by the reticuloendothelial system. Kobayashi et al.^[128] demonstrated that monoclonal antibodies may be conjugated to dendrimers with minimal loss of immunoreactivity. There have also been a number of reports on the conjugation of dendrimers to antibodies for use in immunoassay.^[129,130]

Baker et al. discussed the technical requirements of a smart nano-device therapeutic to diagnose and treat cancer and proposed two approaches.^[131] The first involved the attachment of all the active moieties required for functionality to the same dendrimer, while the second involved the construction of a dendrimer cluster of many dendrimers each carrying a different functional group (Fig. 9). A G5-Folate-FITC conjugate was synthesized and used to successfully target cell lines expressing the folate receptor.

Biological Evaluation of Dendrimers

While several pharmaceutical applications for dendrimers have been proposed, there is limited evidence to suggest that they are safe for human use. This may be partly due to difficulties in assessing these materials. An ongoing concern in dendrimer science is sample purity, and it is not a simple procedure to unambiguously establish the purity of high-generation dendrimers.^[103,104] It is likely that new dendrimers will be developed from building blocks known to be biocompatible or degradable in vivo to natural metabolites, e.g., the poly(glycerol-succinic acid) dendrimers as described by Carnahan and Grinstaff.^[132] Information on dendrimer toxicity and biocompatibility has been included in a number of studies,^[99,102,106,107,115,117,125,126] but there are only two

reports that specifically address the safety of dendrimers. Roberts, Bhalgat, and Zera^[133] investigated G3, G5, and G7 PAMAM dendrimers. The dendrimers were tested in V79 cells and Swiss-Webster mice for a number of biological properties including toxicity, immunogenicity, and biodistribution. Biological complications were observed only with G7 dendrimers at high concentrations. The biodistribution of the dendrimers was generation dependent; G3 showed the highest accumulation in kidney tissue, whereas G5 and G7 preferentially localized in the pancreas. Although it was concluded that the dendrimers did not exhibit properties that would preclude their use in biological applications, it was stressed that the biodistribution of dendrimer preparations should be carefully evaluated. The authors also noted that the dendrimers were not pure (yield ranged 62.7%–76.7%). Malik et al.^[134,135] examined the biocompatibility of PAMAM, poly(propyleneimine) with either diaminobutane or diaminoethane as core, and poly(ethylene oxide) (PEO) grafted carbosilane (CSi-PEO) dendrimers. These dendrimers were used to systematically study the effect of dendrimer generation and surface functionality on biological properties in vitro. It was reported that dendrimers bearing amine surface groups showed concentration-dependent hemolysis, and changes in red cell morphology were observed. CSi-PEO dendrimers and those dendrimers with carboxylate terminal groups were neither hemolytic nor cytotoxic towards a panel of cell lines in vitro. In general, cationic dendrimers were found to be cytotoxic. Preliminary studies with polyether dendrimers showed that dendrimers with carboxylate and malonate surfaces were not hemolytic at 1 hr, but unlike anionic PAMAM dendrimers they were lytic after 24 hr. Cationic ¹²⁵I-labeled PAMAM

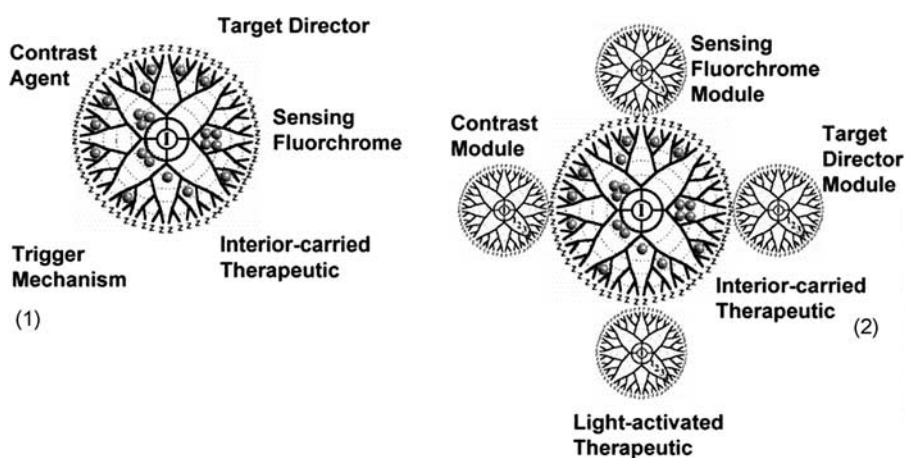


Fig. 9 Two alternative designs for a dendritic nanodevice cancer therapeutic. Configuration 1 is a therapeutic agent built around a single dendrimer molecule (70 Å in diameter). Configuration 2 is a cluster reagent where multiple dendrimers are clustered together, each providing a different functional unit (approx. 200 Å in diameter). The advantage to this latter approach is that functional units can be mixed and matched in a combinatorial method to address almost any type of cancer. (From Ref. [131], © 2001 Kluwer Academic Publishers.)

dendrimers (G3 and G4) administered i.v. to Wistar rats were cleared rapidly from the circulation. Anionic PAMAM dendrimers (G2.5, 3.5, and 5.5) showed longer circulation times with generation-dependent clearance rates. In general, lower generation dendrimers had longer circulation times. The inherent toxicity of the higher generation cationic dendrimers suggest they are unlikely to be suitable for parenteral administration, especially if they are to be used at a high dose. The results indicated that dendrimer structure must also be carefully tailored to avoid rapid hepatic uptake if targeting elsewhere is a primary objective. The study did not consider the issue of dendrimer purity. It is not clear whether the concentrations reported are clinically relevant. In both of the above studies, unmodified dendrimers were evaluated. However, the biological profile of a dendrimer-based delivery system (with surface modifiers and a payload of drug) is likely to be different. It has been shown that DNA-PAMAM dendrimer complexes are less mytotoxic than free dendrimer, possibly due to the fact that the complex reduces the overall positive charge of the dendrimer.^[136] The addition of PEG chains to PAMAM conjugates has been found to increase circulation times and modify biodistribution, with less conjugate being accumulated in the liver and kidney.^[137]

CONCLUSIONS

Dendrimers are a unique class of hyperbranched polymers with well-defined size, shape, and chemical functionality and with properties not found in classical linear and cross-linked polymers. There have been significant developments in the last decade in many areas of dendrimer research, partly due to the commercial availability of dendrimers such as PAMAM, but also the result of the synthesis of several novel dendritic structures. Numerous pharmaceutical applications have been proposed for these hyperbranched polymers, and given the rate of current developments, it is envisaged that dendrimer-based formulations will appear in the next decade.

REFERENCES

1. Tomalia, D.A.; Baker, H.; Dewald, J.; Hall, M.; Kallos, G.; Martin, S.; Roeck, J.; Ryder, J.; Smith, P. A New Class of Polymers: Starburst-Dendritic Macromolecules. *Polym. J. (Tokyo)* **1985**, *17*, 117–132.
2. EDA Core PAMAM Dendrimers in, <http://www.dendritech.com/pamam.html> (accessed April 2002).
3. Buhleier, E.; Wehner, W.; Vögtle, F. "Cascade"- and "Nonskid-Chain-Like" Syntheses of Molecular Cavity Topologies. *Synthesis* **1978**, 155–158.
4. Hawker, C.J.; Fréchet, J.M.J. Preparation of Polymers with Controlled Molecular Architecture. A New Convergent Approach to Dendritic Macromolecules. *J. Am. Chem. Soc.* **1990**, *112*, 7638–7647.
5. Hawker, C.J.; Fréchet, J.M.J. Control of Surface Functionality in the Synthesis of Dendritic Macromolecules Using the Convergent-Growth Approach. *Macromolecules* **1990**, *23*, 4726–4729.
6. Grayson, S.M.; Fréchet, J.M.J. Convergent Dendrons and Dendrimers: From Synthesis to Applications. *Chem. Rev. (Washington, DC)* **2001**, *101*, 3819–3867.
7. Esfand, R.; Tomalia, D.A. Poly(Amidoamine) (PAMAM) Dendrimers: From Biomimicry to Drug Delivery and Biomedical Applications. *Drug Discov. Today* **2001**, *6*, 427–436.
8. de Gennes, P.G.; Hervet, H. Statistics of "Starburst" Polymers. *J. Phys. Lett.* **1983**, *44*, 351–360.
9. Tomalia, D.A.; Hall, M.; Hedstrand, D.M. Starburst Dendrimers. III. The Importance of Branch Junction Symmetry in the Development of Topological Shell Molecules. *J. Am. Chem. Soc.* **1987**, *109*, 1601–1603.
10. Hummelen, J.C.; Van Dongen, J.L.J.; Meijer, E.W. Electrospray Mass Spectrometry of Poly(Propylene Imine) Dendrimers—The Issue of Dendritic Purity or Polydispersity. *Chem. Eur. J.* **1997**, *3*, 1489–1493.
11. Dykes, G.M. Dendrimers: A Review of Their Appeal and Applications. *J. Chem. Technol. Biotechnol.* **2001**, *76*, 903–918.
12. Wendland, M.S.; Zimmerman, S.C. Synthesis of Cored Dendrimers. *J. Am. Chem. Soc.* **1999**, *121*, 1389–1390.
13. Hecht, S.; Vladimirov, N.; Fréchet, J.M. Encapsulation of Functional Moieties Within Branched Star Polymers: Effect of Chain Length and Solvent on Site Isolation. *J. Am. Chem. Soc.* **2001**, *123*, 18–25.
14. Hawker, C.J.; Wooley, K.L.; Fréchet, J.M.J. Unimolecular Micelles and Globular Amphiphiles: Dendritic Macromolecules as Novel Recyclable Solubilization Agents. *J. Chem. Soc., Perkin Trans. 1.* **1993**, 1287–1297.
15. Liu, M.; Kono, K.; Fréchet, J.M.J. Water-Soluble Dendritic Unimolecular Micelles: Their Potential as Drug Delivery Agents. *J. Control. Release* **2000**, *65*, 121–131.
16. Wang, Y.; Zeng, F.; Zimmerman, S.C. Dendrimers with Anthryridine-Based Hydrogen-Bonding Units at Their Cores: Synthesis, Complexation and Self-assembly Studies. *Tetrahedron Lett.* **1997**, *38*, 5459–5462.
17. Jansen, J.F.G.A.; de Brabander-van Den Berg, E.M.M.; Meijer, E.W. Encapsulation of Guest Molecules into a Dendritic Box. *Science* **1994**, *266*, 1226–1229.
18. Bosman, A.W.; Janssen, H.M.; Meijer, E.W. About Dendrimers: Structure, Physical Properties, and Applications. *Chem. Rev. (Washington, DC)* **1999**, *99*, 1665–1688.
19. Majoral, J.P.; Caminade, A.M. Dendrimers Containing Heteroatoms (Si, P, B, Ge, or Bi). *Chem. Rev.* **1999**, *99*, 845–880.
20. Fischer, M.; Vogtle, F. Dendrimers: From Design to Application—A Progress Report. *Angew. Chem. Int. Ed.* **1999**, *38*, 885–905.



21. Denkwalter, R.G.; Kolc, J.; Lukasavage, W.J. Macromolecular Highly Branched Homogeneous Compound Based on Lysine Units. US Patent 4,289,872.
22. Mulders, S.J.E.; Brouwer, A.J.; van der Meer, P.G.J.; Liskamp, R.M.J. Synthesis of a Novel Amino Acid Based Dendrimer. *Tetrahedron Lett.* **1997**, *38*, 631–634.
23. Kim, Y.; Zeng, F.; Zimmerman, S.C. Peptide Dendrimers from Natural Amino Acids. *Chem. Eur. J.* **1999**, *5*, 2133–2138.
24. Veprek, P.; Jezek, J. Peptide and Glycopeptide Dendrimers. Part I. *J. Pept. Sci.* **1999**, *5*, 5–23.
25. Veprek, P.; Jezek, J. Peptide and Glycopeptide Dendrimers. Part II. *J. Pept. Sci.* **1999**, *5*, 203–220.
26. Tam, J.P. Synthetic Peptide Vaccine Design: Synthesis and Properties of a High-Density Multiple Antigenic Peptide System. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 5409–5413.
27. Wooley, K.L.; Hawker, C.J.; Pochan, J.M.; Fréchet, J.M.J. Physical Properties of Dendritic Macromolecules: A Study of Glass Transition Temperature. *Macromolecules* **1993**, *26*, 1514–1519.
28. Mengerink, Y.; Mure, M.; de Brabander-van den Berg, E.M.M.; van der Wal, S. Exclusion Chromatography of Polypropylenamine Dendrimers. *J. Chromatogr. A* **1996**, *730*, 75–81.
29. Tang, M.; Redemann, C.T.; Szoka, F.C., Jr. In Vitro Gene Delivery by Degraded Polyamidoamine Dendrimers. *Bioconjug. Chem.* **1996**, *7*, 703–714.
30. Brothers, H.M., III. Piehler, L.T.; Tomalia, D.A. Slab-Gel and Capillary Electrophoretic Characterization of Polyamidoamine Dendrimers. *J. Chromatogr. A* **1998**, *814*, 233–246.
31. van der Wal, S.; Mengerink, Y.; Brackman, J.C.; de Brabander-van den Berg, E.M.M.; Jeronimus-Stratingh, C.M.; Bruins, A.P. Compositional Analysis of Nitrile Terminated Poly(Propylene Imine) Dendrimers by High-Performance Liquid Chromatography Combined with Electrospray Mass Spectrometry. *J. Chromatogr. A* **1998**, *825*, 135–147.
32. Liu, M.; Kono, K.; Fréchet, J.M.J. Water-Soluble Dendrimer–Poly(Ethylene Glycol) Starlike Conjugates as Potential Drug Carriers. *J. Polym. Sci. Part A: Polym. Chem.* **1999**, *37*, 3492–3503.
33. Yu, D.; Vladimirov, N.; Fréchet, J.M.J. MALDI-TOF in the Characterizations of Dendritic-Linear Block Copolymers and Stars. *Macromolecules* **1999**, *32*, 5186–5192.
34. Zhuo, R.X.; Du, B.; Lu, Z.R. In Vitro Release of 5-Fluorouracil with Cyclic Core Dendritic Polymer. *J. Control. Release* **1999**, *57*, 249–257.
35. Nourse, A.; Millar, D.B.; Minton, A.P. Physicochemical Characterization of Generation 5 Polyamidoamine Dendrimers. *Biopolymers* **2000**, *53*, 316–328.
36. Uppuluri, S.; Keinath, S.E.; Tomalia, D.A.; Dvornic, P.R. Rheology of Dendrimers. I. Newtonian Flow Behavior of Medium and Highly Concentrated Solutions of Polyamidoamine (PAMAM) Dendrimers in Ethylenediamine (EDA) Solvent. *Macromolecules* **1998**, *31*, 4498–4510.
37. Tomalia, D.A.; Naylor, A.M.; Goddard, W.A., III. Starburst Dendrimers: Molecular-Level Control of Size, Shape, Surface Chemistry, Topology and Flexibility in the Conversion of Atoms to Macroscopic Materials. *Angew. Chem. Int. Ed.* **1990**, *29*, 138–175.
38. Naylor, A.M.; Goddard, W.A., III. Kiefer, G.E.; Tomalia, D.A. Starburst Dendrimers. 5. Molecular Shape Control. *J. Am. Chem. Soc.* **1989**, *111*, 2339–2341.
39. Tomalia, D.A.; Dvornic, P.R. Dendrimers. In *Polymeric Materials Encyclopedia*; Salamone, J.C., Ed.; CRC Press: Boca Raton, 1996; 1814–1830.
40. Hudson, S.D.; Jung, H.T.; Percec, V.; Cho, W.D.; Johansson, G.; Ungar, G.; Balagurusamy, V.S.K. Direct Visualization of Individual Cylindrical and Spherical Supramolecular Dendrimers. *Science* **1997**, *278*, 449–452.
41. Jackson, C.L.; Chanzy, H.D.; Booy, F.P.; Tomalia, D.A.; Amis, E.J. Characterization of Dendrimer Molecules by Staining and Cryoelectron Microscopy Techniques. *Polym. Mater. Sci. Eng.* **1997**, *77*, 222–225.
42. Jackson, C.L.; Chanzy, H.D.; Booy, F.P.; Drake, B.J.; Tomalia, D.A.; Bauer, B.J.; Amis, E.J. Visualization of Dendrimer Molecules by Transmission Electron Microscopy (TEM): Staining Methods and Cryo-TEM of Vitrified Solutions. *Macromolecules* **1998**, *31*, 6259–6265.
43. Sheiko, S.S.; Eckert, G.; Ignat'eva, G.; Muzafarov, A.M.; Spickermann, J.; Raedar, H.J.; Moeller, M. Solid-like States of a Dendrimer Liquid Displayed by Scanning Force Microscopy. *Macromol. Rapid Commun.* **1996**, *17*, 283–297.
44. Huck, W.T.S.; Van Veggel, F.C.J.M.; Sheiko, S.S.; Moller, M.; Reinhoudt, D.N. Molecularly Thin Films of Metallo-dendrimers. *J. Phys. Org. Chem.* **1998**, *11*, 540–545.
45. Bliznyuk, V.N.; Rinderspacher, F.; Tsukruk, V.V. On the Structure of Polyamidoamine Dendrimer Monolayers. *Polymer* **1998**, *39*, 5249–5252.
46. Tsukruk, V.V.; Rinderspacher, F.; Bliznyuk, V.N. Self-Assembled Multilayer Films from Dendrimers. *Langmuir* **1997**, *13*, 2171–2176.
47. Zhang, H.; Grim, P.C.M.; Foubert, P.; Vosch, T.; Vanoppen, P.; Wiesler, U.M.; Berresheim, A.J.; Muellen, K.; De Schryver, F.C. Properties of Single Dendrimer Molecules Studied by Atomic Force Microscopy. *Langmuir* **2000**, *16*, 9009–9014.
48. Bauer, B.J.; Topp, A.; Tomalia, D.A.; Amis, E.J. Effect of Solvent Quality on the Molecular Dimensions of PAMAM Dendrimers. *Polym. Mater. Sci. Eng.* **1998**, *79*, 312–313.
49. Topp, A.; Bauer, B.J.; Tomalia, D.A.; Amis, E.J. Effect of Solvent Quality on the Molecular Dimensions of PAMAM Dendrimers. *Macromolecules* **1999**, *32*, 7232–7237.
50. Briber, R.M.; Bauer, B.J.; Hammouda, B.; Tomalia, D.A. Small-Angle Neutron Scattering from Solutions of Dendrimer Molecules: Intermolecular Interactions. *Polym. Mater. Sci. Eng.* **1992**, *67*, 430–431.
51. Bauer, B.J.; Hammouda, B.; Briber, R.M.; Tomalia, D.A. SANS from Solutions of Large Dendrimer Molecules. *Polym. Mater. Sci. Eng.* **1993**, *69*, 341–342.



52. Prosa, T.J.; Bauer, B.J.; Amis, E.J.; Tomalia, D.A.; Scherrenberg, R. A SAXS Study of the Internal Structure of Dendritic Polymer Systems. *J. Polym. Sci. Part B: Polym. Phys.* **1997**, *35*, 2913–2924.
53. Topp, A.; Bauer, B.J.; Prosa, T.J.; Scherrenberg, R.; Amis, E.J. Size Change of Dendrimers in Concentrated Solution. *Macromolecules* **1999**, *32*, 8923–8931.
54. Esumi, K.; Gojino, M. Adsorption of Poly(Amidoamine) Dendrimers on Alumina/Water and Silica/Water Interfaces. *Langmuir* **1998**, *14*, 4466–4470.
55. Kabanov, A.V.; Kabanov, V.A. DNA Complexes with Polycations for the Delivery of Genetic Material into Cells. *Bioconjug. Chem.* **1995**, *6*, 7–20.
56. Ottaviani, M.F.; Sacchi, B.; Turro, N.J.; Chen, W.; Jockusch, S.; Tomalia, D.A. An EPR Study of the Interactions Between Starburst Dendrimers and Polynucleotides. *Macromolecules* **1999**, *32*, 2275–2282.
57. Ottaviani, M.F.; Furini, F.; Casini, A.; Turro, N.J.; Jockusch, S.; Tomalia, D.A.; Messori, L. Formation of Supramolecular Structures Between DNA and Starburst Dendrimers Studied by EPR, CD, UV, and Melting Profiles. *Macromolecules* **2000**, *33*, 7842–7851.
58. Milhem, O.M. An Investigation into the Potential Use of Polyamidoamine Dendrimers for Drug Delivery. Ph.D. Thesis, University of Manchester, Manchester.
59. Ramzi, A.; Bauer, B.J.; Scherrenberg, R.; Joosten, J.; Amis, E.J. SANS Study of Fatty Acid Modified Dendrimers. *Polym. Mater. Sci. Eng.* **1998**, *79*, 382–383.
60. Ramzi, A.; Scherrenberg, R.; Joosten, J.; Lemstra, P.; Mortensen, K. Structure–Property Relations in Dendritic Polyelectrolyte Solutions at Different Ionic Strength. *Macromolecules* **2002**, *35*, 827–833.
61. Chen, W.; Tomalia, D.A.; Thomas, J.L. Unusual pH-Dependent Polarity Changes in PAMAM Dendrimers: Evidence for pH-Responsive Conformational Changes. *Macromolecules* **2000**, *33*, 9169–9172.
62. Lescanec, R.L.; Muthukumar, M. Configurational Characteristics and Scaling Behavior of Starburst Molecules: A Computational Study. *Macromolecules* **1990**, *23*, 2280–2288.
63. Mourey, T.H.; Turner, S.R.; Rubinstein, M.; Fréchet, J.M.J.; Hawker, C.J.; Wooley, K.L. Unique Behavior of Dendritic Macromolecules: Intrinsic Viscosity of Polyether Dendrimers. *Macromolecules* **1992**, *25*, 2401–2406.
64. Scherrenberg, R.; Coussens, B.; van Vliet, P.; Edouard, G.; Brackman, J.; de Brabander-van den Berg, E.M.M.; Mortensen, K. The Molecular Characteristics of Poly(Propyleneimine) Dendrimers as Studied with Small-Angle Neutron Scattering, Viscosimetry, and Molecular Dynamics. *Macromolecules* **1998**, *31*, 456–461.
65. Meltzer, A.D.; Tirrell, D.A.; Jones, A.A.; Inglefield, P.T.; Downing, D.M.; Tomalia, D.A. Carbon-13 NMR Relaxation in Poly(Amido Amine) Starburst Dendrimers. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **1989**, *30*, 121–122.
66. Meltzer, A.D.; Tirrell, D.A.; Jones, A.A.; Inglefield, P.T.; Hedstrand, D.M.; Tomalia, D.A. Chain Dynamics in Poly(Amidoamine) Dendrimers: A Study of Carbon-13 NMR Relaxation Parameters. *Macromolecules* **1992**, *25*, 4541–4548.
67. Ottaviani, M.F.; Bossmann, S.; Turro, N.J.; Tomalia, D.A. Characterization of Starburst Dendrimers by the EPR Technique. 1. Copper Complexes in Water Solution. *J. Am. Chem. Soc.* **1994**, *116*, 661–671.
68. Gopidas, K.R.; Leheny, A.R.; Caminati, G.; Turro, N.J.; Tomalia, D.A. Photophysical Investigation of Similarities Between Starburst Dendrimers and Anionic Micelles. *J. Am. Chem. Soc.* **1991**, *113*, 7335–7342.
69. Gorman, C.B.; Hager, M.W.; Parkhurst, B.L.; Smith, J.C. Use of a Paramagnetic Core to Affect Longitudinal Nuclear Relaxation in Dendrimers—A Tool for Probing Dendrimer Conformation. *Macromolecules* **1998**, *31*, 815–822.
70. Uppuluri, S.; Tomalia, D.A.; Dvornic, P.R. A “Dense-Shell” Model for Intramolecular Morphology of Polyamidoamine Dendrimers Based on Experimental Rheological Data. *Polym. Mater. Sci. Eng.* **1997**, *77*, 116–117.
71. Prosa, T.J.; Bauer, B.J.; Topp, A.; Amis, E.J.; Scherrenberg, R. Size Changes and Interpenetration Within Concentrated Dendrimer Solutions. *Polym. Mater. Sci. Eng.* **1998**, *79*, 307–308.
72. Uppuluri, S.; Morrison, F.A.; Dvornic, P.R. Rheology of Dendrimers. 2. Bulk Polyamidoamine Dendrimers Under Steady Shear, Creep, and Dynamic Oscillatory Shear. *Macromolecules* **2000**, *33*, 2551–2560.
73. Hawker, C.J.; Farrington, P.J.; Mackay, M.E.; Wooley, K.L.; Fréchet, J.M.J. Molecular Ball Bearings: The Unusual Melt Viscosity Behavior of Dendritic Macromolecules. *J. Am. Chem. Soc.* **1995**, *117*, 4409–4410.
74. Maciejewski, M. Concepts of Trapping Topologically by Shell Molecules. *J. Macromol. Sci. Chem.* **1982**, *A17*, 689–703.
75. Newkome, G.R.; Yao, Z.; Baker, G.R.; Gupta, V.K. Micelles. Part 1. Cascade Molecules: A New Approach to Micelles. A [27]-Arborol. *J. Org. Chem.* **1985**, *50*, 2003–2004.
76. Tomalia, D.A.; Berry, V.; Hall, M.; Hedstrand, D.M. Starburst Dendrimers. 4. Covalently Fixed Unimolecular Assemblages Reminiscent of Spheroidal Micelles. *Macromolecules* **1987**, *20*, 1164–1167.
77. Newkome, G.R.; Moorefield, C.N.; Baker, G.R.; Saunders, M.J.; Grossman, S.H. Unimolecular Micelles. *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 1178–1180.
78. Stevelmans, S.; Hest, J.C.M.v.; Jansen, J.F.G.A.; Van Boxtel, D.A.F.J.; de Berg, E.M.M.; Meijer, E.W. Synthesis, Characterization, and Guest–Host Properties of Inverted Unimolecular Dendritic Micelles. *J. Am. Chem. Soc.* **1996**, *118*, 7398–7399.
79. Liu, M.; Fréchet, J.M.J. Preparation of Water-Soluble Dendritic Unimolecular Micelles as Potential Drug Delivery Agents. *Polym. Mater. Sci. Eng.* **1999**, *80*, 167–168.
80. Jansen, J.F.G.A.; de Brabander-van den Berg, E.M.M.; Meijer, E.W. The Dendritic Box and Bengal Rose. *Polym. Mater. Sci. Eng.* **1995**, *73*, 123–124.

81. Jansen, J.F.G.A.; Meijer, E.W.; de Brabander-van den Berg, E.M.M. The Dendritic Box: Shape-Selective Liberation of Encapsulated Guests. *J. Am. Chem. Soc.* **1995**, *117*, 4417–4418.
82. Jansen, J.F.G.A.; Meijer, E.W.; de Brabander-van den Berg, E.M.M. Bengal Rose Dendritic Box. *Macromol. Symp.* **1996**, *102*, 27–33.
83. Schultz, L.G.; Zhao, Y.; Zimmerman, S.C. Synthesis of Cored Dendrimers with Internal Cross-Links. *Angew. Chem. Int. Ed.* **2001**, *40*, 1962–1966.
84. Esfand, R.; Tomalia, D.A.; Beezer, A.E.; Mitchell, J.C.; Hardy, M.; Orford, C. Dendriopore and Dendrilock Concepts New Controlled Delivery Strategies. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2000**, *41*, 1324–1325.
85. Kojima, C.; Kono, K.; Maruyama, K.; Takagishi, T. Synthesis of Polyamidoamine Dendrimers Having Poly(Ethylene Glycol) Grafts and Their Ability to Encapsulate Anticancer Drugs. *Bioconjug. Chem.* **2000**, *11*, 910–917.
86. Li, Y.; Dubin, P.L.; Spindler, R.; Tomalia, D.A. Complex Formation Between Poly(Dimethyldiallylammonium Chloride) and Carboxylated Starburst Dendrimers. *Macromolecules* **1995**, *28*, 8426–8428.
87. Zhang, H.; Dubin, P.L.; Spindler, R.; Tomalia, D. Binding of Carboxylated Starburst Dendrimers to Poly(Diallyldimethylammonium Chloride). *Ber. Bunsen-Ges.* **1996**, *100*, 923–928.
88. Kabanov, V.A.; Zezin, A.B.; Rogacheva, V.B.; Gulyaeva, Z.G.; Zansochova, M.F.; Joosten, J.G.H.; Brackman, J. Interaction of Astramol Poly(Propyleneimine) Dendrimers with Linear Polyanions. *Macromolecules* **1999**, *32*, 1904–1909.
89. Miura, N.; Dubin, P.L.; Moorefield, C.N.; Newkome, G.R. Complex Formation by Electrostatic Interaction Between Carboxyl-Terminated Dendrimers and Oppositely Charged Polyelectrolytes. *Langmuir* **1999**, *15*, 4245–4250.
90. Zhang, H.; Dubin, P.L.; Ray, J.; Manning, G.S.; Moorefield, C.N.; Newkome, G.R. Interaction of a Polycation with Small Oppositely Charged Dendrimers. *J. Phys. Chem. B* **1999**, *103*, 2347–2354.
91. Huang, Q.R.; Dubin, P.L.; Moorefield, C.N.; Newkome, G.R. Counterion Binding on Charged Spheres: Effect of pH and Ionic Strength on the Mobility of Carboxyl-Terminated Dendrimers. *J. Phys. Chem. B* **2000**, *104*, 898–904.
92. Milhem, O.M.; Myles, C.; McKeown, N.B.; Attwood, D.; D'Emanuele, A. Solubility Enhancement of a Hydrophobic Drug Using Polyamidoamine Starburst Dendrimers. *Proc. Int. Symp. Control. Release Bioact. Mater.* **1999**, *26*, 931–932.
93. Wiwattanapatapee, R.; Jee, R.D.; Duncan, R. PAMAM Dendrimers as a Potential Oral Drug Delivery System: Dendrimer Complexes with Piroxicam. *Proc. Int. Symp. Control. Release Bioact. Mater.* **1999**, *26*, 145–146.
94. Twyman, L.J.; Beezer, A.E.; Esfand, R.; Hardy, M.J.; Mitchell, J.C. The Synthesis of Water Soluble Dendrimers, and Their Application as Possible Drug Delivery Systems. *Tetrahedron Lett.* **1999**, *40*, 1743–1746.
95. Milhem, O.M.; Myles, C.; McKeown, N.B.; Attwood, D.; D'Emanuele, A. Polyamidoamine Starburst Dendrimers as Solubility Enhancers. *Int. J. Pharm.* **2000**, *197*, 239–241.
96. Milhem, O.M.; Mobedi, H.; Day, N.; McKeown, N.B.; Attwood, D.; D'Emanuele, A. Development and Characterization of Dendrimer Conjugates Derivatives for Drug Delivery. *Proc. Int. Symp. Control. Release Bioact. Mater.* **2001**, *28*, 329–330.
97. Milhem, O.M.; Mobedi, H.; Day, N.; McKeown, N.B.; Attwood, D.; D'Emanuele, A. Development of Dendrimer Conjugates for Drug Delivery. *Polym. Mater. Sci. Eng.* **2001**, *84*, 721.
98. Sakthivel, T.; Toth, I.; Florence, A.T. Distribution of a Lipidic 2.5 nm Diameter Dendrimer Carrier After Oral Administration. *Int. J. Pharm.* **1999**, *183*, 51–55.
99. Wiwattanapatapee, R.; Carreno-Gomez, B.; Malik, N.; Duncan, R. Anionic PAMAM Dendrimers Rapidly Cross Adult Rat Intestine In Vitro: A Potential Oral Delivery System? *Pharm. Res.* **2000**, *17*, 991–998.
100. Katchalsky, A. Polyelectrolytes and Their Biological Interactions. *Biophys. J.* **1964**, *4*, 9–41.
101. Florence, A.T.; Sakthivel, T.; Toth, I. Oral Uptake and Translocation of a Polylysine Dendrimer with a Lipid Surface. *J. Control. Release* **2000**, *65*, 253–259.
102. Tajarobi, F.; El-Sayed, M.; Rege, B.D.; Polli, J.E.; Ghandehari, H. Transport of Polyamidoamine Dendrimers Across Madin-Darby Canine Kidney Cells. *Int. J. Pharm.* **2001**, *215*, 263–267.
103. Zhang, Z.Y.; Smith, B.D. High-Generation Polycationic Dendrimers Are Unusually Effective at Disrupting Anionic Vesicles: Membrane Bending Model. *Bioconjug. Chem.* **2000**, *11*, 805–814.
104. El-Sayed, M.; Kiani, M.F.; Naimark, M.D.; Hikal, A.H.; Ghandehari, H. Extravasation of Poly(Amidoamine) (PAMAM) Dendrimers Across Microvascular Network Endothelium. *Pharm. Res.* **2001**, *18*, 23–28.
105. Yoo, H.; Juliano, R.L. Enhanced Delivery of Antisense Oligonucleotides with Fluorophore-Conjugated PAMAM Dendrimers. *Nucleic Acids Res.* **2000**, *28*, 4225–4231.
106. Haensler, J.; Szoka, F.C., Jr. Polyamidoamine Cascade Polymers Mediate Efficient Transfection of Cells in Culture. *Bioconjug. Chem.* **1993**, *4*, 372–379.
107. Kukowska-Latallo, J.F.; Bielinska, A.U.; Johnson, J.; Spindler, R.; Tomalia, D.A.; Baker, J.R., Jr. Efficient Transfer of Genetic Material into Mammalian Cells Using Starburst Polyamidoamine Dendrimers. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 4897–4902.
108. Turunen, M.P.; Hiltunen, M.O.; Ruponen, M.; Virkamaki, L.; Szoka, F.C., Jr.; Urtti, A.; Yla-Herttuala, S. Efficient Adventitial Gene Delivery to Rabbit Carotid Artery with Cationic Polymer–Plasmid Complexes. *Gene Ther.* **1999**, *6*, 6–11.
109. Kukowska-Latallo, J.F.; Raczka, E.; Quintana, A.; Chen, C.; Rymaszewski, M.; Baker, J.R., Jr. Intravascular and Endobronchial DNA Delivery to Murine Lung Tissue

- Using a Novel, Nonviral Vector. *Hum. Gene Ther.* **2000**, *11*, 1385–1395.
110. Eichman, J.D.; Bielinska, A.U.; Kukowska-Latallo, J.F.; Baker, J.R., Jr. The Use of PAMAM Dendrimers in the Efficient Transfer of Genetic Material into Cells. *Pharm. Sci. Technol. Today* **2000**, *3*, 232–245.
111. Bielinska, A.U.; Yen, A.; Wu, H.L.; Zahos, K.M.; Sun, R.; Weiner, N.D.; Baker, J.R., Jr.; Roessler, B.J. Application of Membrane-Based Dendrimer/DNA Complexes for Solid Phase Transfection In Vitro and In Vivo. *Biomaterials* **2000**, *21*, 877–887.
112. Cheng, H.; Zhou, R.; Liu, L.; Du, B.; Zhuo, R. Cyclic Core Dendrimer as a New Kind of Vector for Gene Transfer into Mammalian Cells. *Genetica (The Hague)* **2000**, *108*, 53–56.
113. Shah, D.S.; Sakthivel, T.; Toth, I.; Florence, A.T.; Wilderspin, A.F. DNA Transfection and Transfected Cell Viability Using Amphipathic Asymmetric Dendrimers. *Int. J. Pharm.* **2000**, *208*, 41–48.
114. De Jong, G.; Telenius, A.; Vanderbyl, S.; Meitz, A.; Drayer, J. Efficient In-Vitro Transfer of a 60-Mb Mammalian Artificial Chromosome into Murine and Hamster Cells Using Cationic Lipids and Dendrimers. *Chromosome Res.* **2001**, *9*, 475–485.
115. Luo, D.; Haverstick, K.; Belcheva, N.; Han, E.; Saltzman, W.M. Poly(Ethylene Glycol)-Conjugated PAMAM Dendrimer for Biocompatible, High-Efficiency DNA Delivery. *Macromolecules* **2002**, *35*, 3456–3462.
116. Gebhart, C.L.; Kabanov, A.V. Evaluation of Polyplexes as Gene Transfer Agents. *J. Control. Release* **2001**, *73*, 401–416.
117. Hughes, J.A.; Aronsohn, A.I.; Avrutskaya, A.V.; Juliano, R.L. Evaluation of Adjuvants That Enhance the Effectiveness of Antisense Oligodeoxynucleotides. *Pharm. Res.* **1996**, *13*, 404–410.
118. Bielinska, A.; Kukowska-Latallo, J.F.; Johnson, J.; Tomalia, D.A.; Baker, J.R., Jr. Regulation of In Vitro Gene Expression Using Antisense Oligonucleotides or Antisense Expression Plasmids Transfected Using Starburst PAMAM Dendrimers. *Nucleic Acids Res.* **1996**, *24*, 2176–2182.
119. Delong, R.; Stephenson, K.; Loftus, T.; Fisher, M.; Alahari, S.; Nolting, A.; Juliano, R.L. Characterization of Complexes of Oligonucleotides with Polyamidoamine Starburst Dendrimers and Effects on Intracellular Delivery. *J. Pharm. Sci.* **1997**, *86*, 762–764.
120. Yoo, H.; Sazani, P.; Juliano, R.L. PAMAM Dendrimers as Delivery Agents for Antisense Oligonucleotides. *Pharm. Res.* **1999**, *16*, 1799–1804.
121. Maeda, H. The Tumor Blood Vessel as an Ideal Target for Macromolecular Anticancer Agents. *J. Control. Release* **1992**, *19*, 315–324.
122. Malik, N.; Evagorou, E.G.; Duncan, R. Dendrimer – Platinate: A Novel Approach to Cancer Chemotherapy. *Anti-Cancer Drugs* **1999**, *10*, 767–776.
123. Kono, K.; Liu, M.; Fréchet, J.M.J. Design of Dendritic Macromolecules Containing Folate or Methotrexate Residues. *Bioconjug. Chem.* **1999**, *10*, 1115–1121.
124. Wiener, E.C.; Konda, S.; Shadron, A.; Brechbiel, M.; Gansow, O. Targeting Dendrimer-Chelates to Tumors and Tumor Cells Expressing the High-Affinity Folate Receptor. *Investig. Radiol.* **1997**, *32*, 748–754.
125. Barth, R.F.; Adams, D.M.; Soloway, A.H.; Alam, F.; Darby, M.V. Boronated Starburst Dendrimer-Monoclonal Antibody Immunoconjugates: Evaluation as a Potential Delivery System for Neutron Capture Therapy. *Bioconjug. Chem.* **1994**, *5*, 58–66.
126. Yang, W.; Barth, R.F.; Adams, D.M.; Soloway, A.H. Intratumoral Delivery of Boronated Epidermal Growth Factor for Neutron Capture Therapy of Brain Tumors. *Cancer Res.* **1997**, *57*, 4333–4339.
127. Abu-Rmaileh, R.; Ashford, M.; Day, N.; McKeown, N.B.; Attwood, D.; Stratford, I.J.; D'Emanuele, A. Polyamidoamine (PAMAM) Dendrimers as a Platform for Delivery to Tumour Cells Bearing the Carcinoembryonic Antigen (CEA). *Proc. Int. Symp. Control. Release Bioact. Mater.* **2002**, *29*, in press.
128. Kobayashi, H.; Sato, N.; Saga, T.; Nakamoto, Y.; Ishimori, T.; Toyama, S.; Togashi, K.; Konishi, J.; Brechbiel, M.W. Monoclonal Antibody–Dendrimer Conjugates Enable Radiolabeling of Antibody with Markedly High Specific Activity with Minimal Loss of Immunoreactivity. *Eur. J. Nucl. Med.* **2000**, *27*, 1334–1339.
129. Roberts, J.C.; Adams, Y.E.; Tomalia, D.; Mercer-Smith, J.A.; Lavalley, D.K. Using Starburst Dendrimers as Linker Molecules to Radiolabel Antibodies. *Bioconjug. Chem.* **1990**, *1*, 305–308.
130. Singh, P.; Moll, F., III. Lin, S.H.; Ferzli, C.; Yu, K.S.; Koski, R.K.; Saul, R.G.; Cronin, P. Starburst Dendrimers: Enhanced Performance and Flexibility for Immunoassays. *Clin. Chem.* **1994**, *40*, 1845–1849.
131. Baker, J.R., Jr.; Quintana, A.; Piehler, L.; Banazak-Holl, M.; Tomalia, D.; Raczka, E. The Synthesis and Testing of Anti-cancer Therapeutic Nanodevices. *Biomed. Microdevices* **2001**, *3*, 61–69.
132. Carnahan, M.A.; Grinstaff, M.W. Synthesis and Characterization of Poly(Glycerol-succinic acid) Dendrimers. *Macromolecules* **2001**, *34*, 7648–7655.
133. Roberts, J.C.; Bhargat, M.K.; Zera, R.T. Preliminary Biological Evaluation of Polyamidoamine (PAMAM) Starburst Dendrimers. *J. Biomed. Mater. Res.* **1996**, *30*, 53–65.
134. Malik, N.; Wiwattanapatapee, R.; Klopsch, R.; Lorenz, K.; Frey, H.; Weener, J.W.; Meijer, E.W.; Paulus, W.; Duncan, R. Dendrimers: Relationship Between Structure and Biocompatibility In Vitro, and Preliminary Studies on the Biodistribution of ¹²⁵I-Labeled Polyamidoamine Dendrimers In Vivo. *J. Control. Release* **2000**, *65*, 133–148.
135. Malik, N.; Wiwattanapatapee, R.; Klopsch, R.; Lorenz, K.; Frey, H.; Weener, J.W.; Meijer, E.W.; Paulus, W.; Duncan, R. Erratum to “Dendrimers: Relationship Between Structure and Biocompatibility In Vitro, and Preliminary Studies on the Biodistribution of ¹²⁵I-Labelled Polyamidoamine Dendrimers In Vivo” [*Journal of Controlled Release* 65 (2000) 133–148]. *J. Control. Release* **2000**, *68*, 299–302.



Dendrimers

21

136. Brazeau, G.A.; Attia, S.; Poxon, S.; Hughes, J.A. In Vitro Myotoxicity of Selected Cationic Macromolecules Used in Non-viral Gene Delivery. *Pharm. Res.* **1998**, *15*, 680–684.
137. Kobayashi, H.; Kawamoto, S.; Saga, T.; Sato, N.; Hiraga, A.; Ishimori, T.; Konishi, J.; Togashi, K.; Brechbiel, M.W. Positive Effects of Polyethylene Glycol Conjugation to Generation-4 Polyamidoamine Dendrimers as Macromolecular MR Contrast Agents. *Magn. Reson. Med.* **2001**, *46*, 781–788.

CRYSTALLIZATION: SIGNIFICANCE IN PRODUCT DEVELOPMENT, PROCESSING, AND PERFORMANCE

Naír Rodríguez-Hornedo
Brent D. Sinclair

University of Michigan, Ann Arbor, Michigan, U.S.A.

INTRODUCTION

Understanding crystallization processes is important for the rational formulation, process development, and stability of pharmaceutical products. Whereas thermodynamics describes the equilibrium behavior of a system, pharmaceutical products most often encounter far-from-equilibrium conditions during processing, storage, or delivery. These conditions lead to the creation of metastable liquid or solid states. Metastable liquid states may include freeze concentrated solutions, solutions of weak acids or bases exposed to a pH change, solutions prepared by dissolution of salts of weak acids or bases, and solutions prepared by dissolution of a high energy form. Metastable solid states include amorphous solids, polymorphs, and solvates. Because crystallization provides a means to reduce the free energy of the system to the most stable state, crystallization mechanisms and kinetics determine the extent to which metastable states are reached and maintained.

Whereas knowledge of equilibrium phase diagrams is useful in identifying the concentration (or activity) and temperature regions of thermodynamic stability of solid phases, information on the crystallization kinetics is essential in determining solid phase outcomes. Nucleation and growth rates control the isolation of desired solid state modifications as well as the particle size distribution (number of particles, mean diameter, and standard deviation) and particle shape or morphology. The lack of information on crystallization processes leads to unwanted or previously unknown nucleation events that threaten the development of a pharmaceutical product. Dunitz and Bernstein (1) documented cases of “disappearing or elusive polymorphs,” which provide evidence for the consequences of poor process control in crystallization of polymorphic systems. The recent shortage in the supply of capsules of the HIV protease inhibitor Norvir (ritonavir), due to the sudden formation of a crystalline structure different from the one harvested for months (2), illustrates the decisive role that nucleation mechanisms and kinetics have on crystallization. Nichols and Frampton (3) have reported considerable efforts that failed to crystallize the

metastable polymorph of paracetamol as described in the initial publication of the crystal structure (4). The critical role of crystallization kinetics in determining the appearance of crystalline modifications is also recognized by the FDA and described in the guidelines for the manufacturing of drug substances (5): “Appropriate manufacturing and control procedures (including in-process testing when needed) should be established for the production of the desired solid-state form(s). It should be emphasized that the manufacturing process (or storage condition) is responsible for producing particular polymorphs or solvates; the control methods merely determine the outcome.”

The goal of this chapter is to describe ways in which crystal structure, morphology, and crystallization kinetics can be utilized to reproducibly maintain metastable states and control solid-state outcomes. Experimental methods that can be employed to investigate the factors that regulate crystallization from solution will be presented.

CRYSTALLIZATION

A crystalline phase is created as a consequence of molecular aggregation processes in solution. These prenucleation clusters may achieve a certain size for a sufficient time to enable growth into macroscopic crystals. The rate and mechanisms by which crystals form in liquid solutions are determined by numerous factors, including:

- Solubility or solubility product (solubility product is important in solutions of nonstoichiometric composition)
- Supersaturation: concentration of crystallizing solute or ions that participate in crystallization
- Diffusivity or viscosity
- Temperature
- pH
- Solvent
- Soluble additives and impurities
- Reactivity of surfaces toward nucleation
- Volume of solution

- Rate at which supersaturation is created: cooling rate, freezing rate, rate of pH change.

The effect of these factors on crystallization processes is thoroughly explained in various books (6–10).

Questions on the transferability of crystallization microtechniques to larger scale processes that are reproducible require careful consideration and study of crystallization kinetics. Whereas operationally useful variables that describe crystallization methods are often related to crystallization outcomes, this approach lacks meaningful information for developing a process that yields reliable outcomes because the factors that determine the crystallization kinetics and outcomes are not explicitly considered. For instance, compare the following two approaches to describe the processes for the selective crystallization of polymorphs: (1) form I obtained by cooling; form II obtained by evaporation; and (2) form I obtained at a supersaturation x , temperature y , and time z at which crystals were harvested after crystallization onset; form II obtained at supersaturation x' , etc. Whereas the former approach is at best anecdotal, the latter employs the causative factors and leads to relationships between the crystallization kinetic parameters and the outcomes.

Control of the processes of nucleation and crystal growth is possible as long as the required information is available. In a study of crystallization of polymorphs of paracetamol by Nichols and Frampton (3), seeding with crystals of the desired solid phase and the time of harvesting the solid phase were found to have a significant effect on the isolation of the desired polymorph. The orthorhombic polymorph (metastable) could not be obtained by a previously reported method (4) “—slow evaporation from ethanol.” Dunitz and Bernstein (1) have explained the mystery of disappearing polymorphs by considering various examples and the relevant questions that were left unanswered. They state that, “Once a particular polymorph has been obtained, it is always possible to obtain it again; it is only a matter of finding the right experimental conditions.”

SUPERSATURATION

A solid phase is precipitated from solution if the chemical potential of the solid phase is less than that of the dissolved component. A solution in which the chemical potential of the solute is the same as that of the corresponding solid phase is in equilibrium with the solid phase under the given conditions (temperature, pH, and concentration) and called a saturated solution. In order for crystallization to occur, however, this equilibrium concentration or

solubility must be exceeded. This excess concentration or chemical potential is called supersaturation. Supersaturated states may be created by:

1. Methods that regulate the solute activity or activity product (concentration) include:
 - a. solvent removal (evaporation or freezing);
 - b. addition of indifferent salts with ions that participate in precipitation (addition of NaCl to sodium phosphate solution causing the precipitation of Na_2HPO_4); and
 - c. dissolution of a metastable solid phase (transformations of amorphous to crystalline, anhydrous to hydrate, more soluble to less soluble polymorph, and salt to free acid or free base).
2. Methods that regulate the solute solubility include:
 - a. temperature change;
 - b. pH change; and
 - c. addition of solvent or additives that lower the solubility of the solute.

Knowledge of the driving force for crystallization is essential, not only to characterize the kinetics but also to relate the crystallization outcomes to the parameters that regulate crystallization. The number of molecules required to achieve an effective nucleating cluster is inversely proportional to the supersaturation. Therefore, as the supersaturation is increased the probability of nucleation increases. However, nucleation is energetically more demanding than crystal growth, and there are supersaturation regions in which crystal growth proceeds while spontaneous nucleation is suppressed (metastable zone; 6,7,9).

The driving force for nucleation and growth is the difference in chemical potential of the solute in a supersaturated solution, μ_1 , and in a saturated solution, μ_{eq} :

$$\Delta\mu = (\mu_1 - \mu_{eq}) \quad (1)$$

Because $\mu = \mu^\circ + RT \ln a$, then

$$\Delta\mu = RT \ln (a/a_{eq}) = RT \ln (\gamma_1 c_1 / \gamma_{eq} c_{eq}) \quad (2)$$

and the supersaturation is

$$\sigma = \Delta\mu / RT = \ln (a/a_{eq}) = \ln (\gamma_1 c_1 / \gamma_{eq} c_{eq}) \quad (3)$$

If the activity coefficient, γ , is independent of concentration in the given concentration regime, then $\gamma_1 = \gamma_{eq}$, and the supersaturation becomes

$$\sigma = \ln (c/c_{eq}) = \ln (c/s) \quad (4)$$

where c is the concentration of the crystallizing substance in the supersaturated solution and s is the solubility. (This

notation is adopted to avoid the use of subscripts.) For c/s values smaller than 1.15

$$\sigma = \ln(c/s) \approx (c-s)/s. \quad (5)$$

Frequently used expressions for supersaturation are $[(c-s)/s]$ and the ratio (c/s) . If the supersaturation is attained by changing the pH, then the ratio of the concentration of the solution to the solubility at the given pH is the supersaturation ratio. Fig. 1 illustrates an example of the effect that changing the pH can have on the supersaturation of a weakly basic drug ($pK_a = 5.25$). When the pH of the solution is increased along the path represented by line ABC, spontaneous crystallization will occur when the pH corresponding to point C is achieved. When this boundary is exceeded (i.e., pH increased past point C), the rate of nucleation rapidly increases and the crystallization process may become less controlled. The concentration in the solution decreases as nuclei form and grow until the equilibrium concentration at the final pH is reached at point D. If the concentration is maintained at a pH somewhere in between points B and C, the solution is in the metastable region—where the probability of spontaneous nucleation is negligible while crystal growth on seeds occurs.

NUCLEATION

Nucleation is often the decisive step in the crystallization process and is of practical importance in pharmaceutical systems. Crystallization may occur intentionally or unintentionally during pharmaceutical processes in which

supersaturated conditions are achieved. Pharmaceutical processes in which supersaturated states are attained due to solvent removal include: freezing, freeze drying, spray drying, and drying of wet granulations. Other processes achieve supersaturation by changing solubility, during isolation of solid state forms, or during dissolution of weak bases in an environment in which pH is changing, such as in the gastrointestinal tract, as previously described.

The concentration threshold at which crystallization is observed at times shorter than the processing time or desired product shelf-life or GI transit time, is determined by the kinetic stability of supersaturated states and is regulated by the nucleation mechanisms and kinetics. Nucleation phenomena are equally important in the control of micromeritic properties and in the selective crystallization of a particular polymorph.

The crystallization-induced pH changes that occur during the freezing of sodium phosphate buffers illustrate the importance of kinetic events in determining final pH and product stability. The pH changes of sodium phosphate buffers during freezing at close-to-equilibrium conditions have been thoroughly studied (11, 12), and the phase diagram—phases, composition, and pH—of sodium phosphate buffers during freezing is shown in Fig. 2. However, it is recognized that during freeze-drying, freezing does not take place close-to-equilibrium and that large undercoolings and supersaturations are often encountered (13–16). During cooling of hypoeutectic solutions of monosodium and disodium phosphate, ice crystallization increases the total salt concentration, and the ratio of the dissolved salts will change when precipitation of

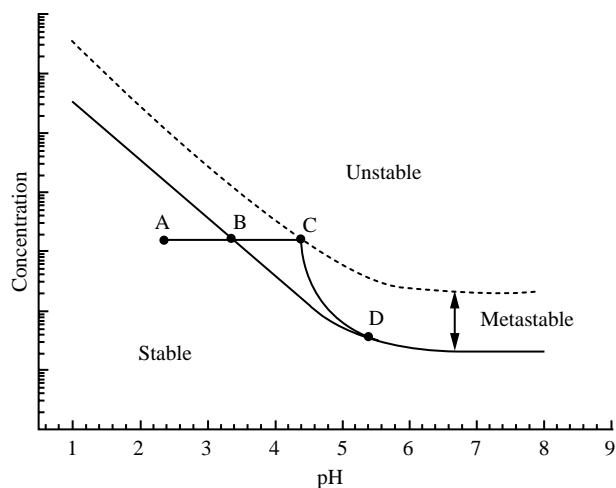


Fig. 1 Regions of varying thermodynamic stability for solutions of a weak base as a function of pH.

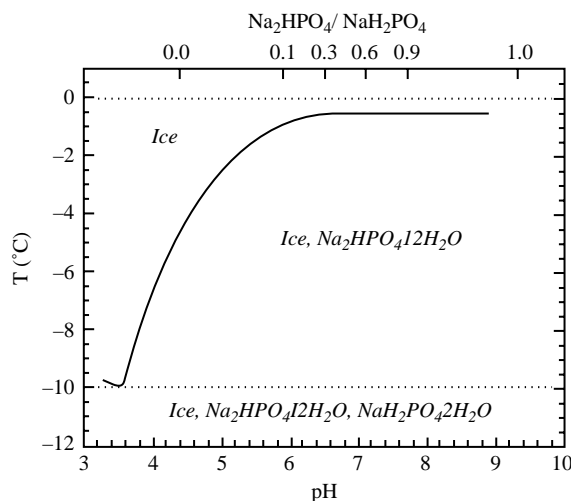


Fig. 2 pH and ratio of mono- to disodium phosphate salts during equilibrium freezing of sodium phosphate buffers, at initial buffer concentration of 20 mM. (Based on Ref. 11.)

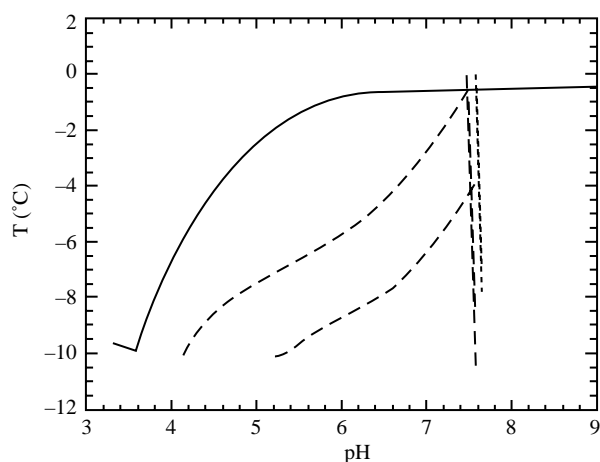


Fig. 3 Comparison of pH during equilibrium and far-from-equilibrium freezing of sodium phosphate buffers. Equilibrium freezing, initial buffer concentration 20 mM, (—). Far-from-equilibrium freezing: Initial buffer concentrations of 100 mM (---) and 8 mM (- - -). (From Refs. 11 and 17).

one of the salts occurs. Disodium phosphate dodecahydrate is the least soluble salt in this system (eutectic temperature = -0.5°C and eutectic concentration = 0.11 M). Under close-to-equilibrium freezing, it precipitates first at initial pH values greater than 3.6, and at temperatures above the ternary eutectic of -9.9°C . This corresponds with the ternary eutectic composition at monosodium to disodium phosphate molar ratios of 3.42:0.06. Because precipitation of $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ consumes HPO_4^{2-} ions, the pH of the unfrozen solution decreases to the equilibrium value of 3.6.

If one wishes to design a formulation to be freeze-dried in which the active product ingredient has a narrow pH range for optimal stability, at what initial pH and buffer concentration should the formulation be prepared before freezing so that the pH remains within the specified range during freezing? In other words, does selective salt crystallization occur at the freeze concentrations achieved during the freeze-drying process, and what is the concentration dependence of salt crystallization in such systems? Based on thermodynamic considerations, sodium phosphate buffer solutions are expected to reach a buffer concentration of 3.5 M and a pH of 3.6 upon freezing to -10°C , if crystallization of ice and $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ proceed to equilibrium. This may be the case in seeded systems under slow cooling rates in which the pH and composition of the residual solution follow very closely the phase diagram (Fig. 3) (11). In contrast to the equilibrium behavior, studies during far-from-equilibrium freezing of sodium phosphate buffers show that the more dilute buffer solutions (lower concentrations of precipitating ions) experience greater departures from equilibrium and maintain pH values as high as 3 units above the equilibrium pH values (17). Thus, crystallization-induced pH changes are determined by the factors that regulate salt crystallization kinetics and depend on the initial salt concentration (pH and concentration).

Whereas buffer salt concentrations are not directly measured during freezing, they can be calculated from in situ pH measurements. Temperature–pH time profiles during freezing reveal the onset of ice crystallization

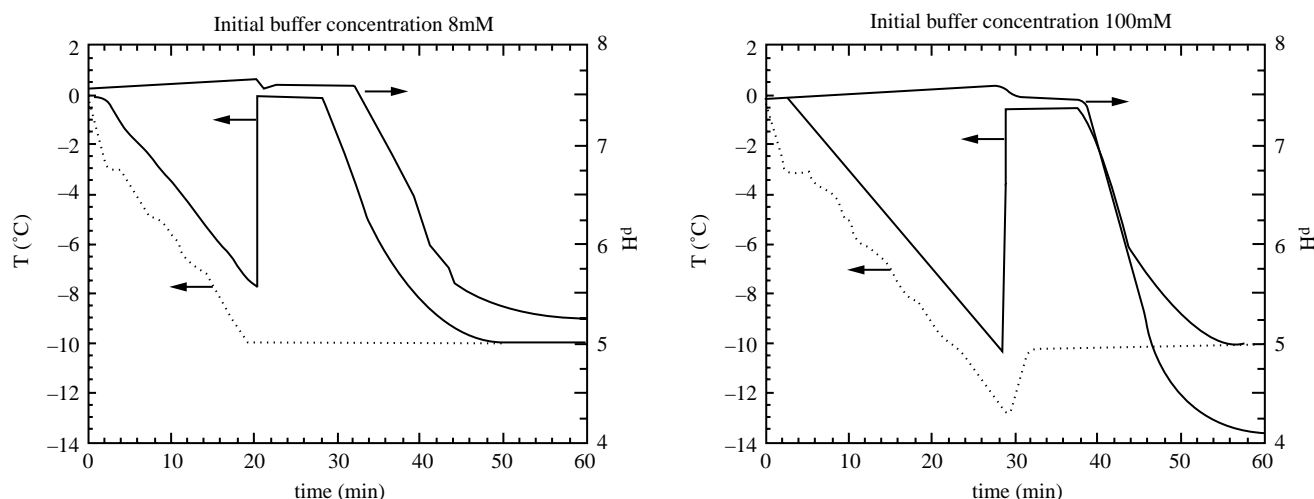


Fig. 4 Effect of initial buffer concentration on pH during far-from-equilibrium freezing of sodium phosphate buffer solutions. Initial pH of 7.4 at 25°C and initial buffer concentration of 100 mM and 8 mM. Solution temperature (—), bath temperature (---), and (—) pH. (From Ref. 17.)

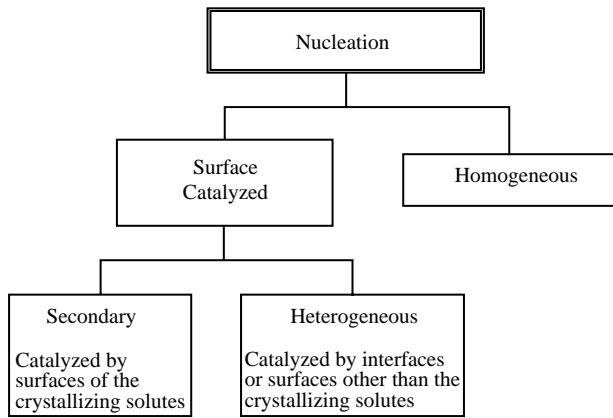


Fig. 5 Mechanisms for crystal nucleation.

(increase in temperature) followed by $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ crystallization (decrease in pH) in sodium phosphate buffer solutions of initial pH 7.4 and initial concentration in the range of 8–100 mM (Fig. 4) (17). This salt precipitates very early in the freezing process and within 15 min. after the onset of ice crystallization.

Nucleation Mechanisms

Nucleation mechanisms can be divided into two main categories: homogeneous and surface- or interface-catalyzed (7–9,18), as shown in Fig. 5. Homogeneous nucleation rarely occurs in large volumes (greater than 100 μl) because solutions contain random impurities that may induce nucleation (19, 20). This type of nucleation is referred to as heterogeneous. A surface or an interface of different composition than the crystallizing solute may serve as a nucleation substrate by decreasing the energy barrier for the formation of a nuclei that can grow into a mature crystal. Nucleation that is promoted by crystals of the crystallizing solute is known as secondary nucleation. These mechanisms are thoroughly discussed by Mullin (7), Myerson (9), and Zettlemoyer (18).

Homogeneous Nucleation

Thermodynamic considerations for nucleation are based on the work of Gibbs (21), Volmer (22), and others, where the free energy change for an aggregate undergoing a phase transition ΔG is given by

$$\Delta G = \Delta G_V + \Delta G_S \quad (6)$$

where ΔG_S is the surface free energy change associated with the formation of the aggregate (a positive quantity),

and ΔG_V is the volume free energy change associated with the phase transition (a negative quantity). For homogeneous or heterogeneous nucleation

$$\Delta G_V = -\alpha l^3 \nu k_B T \ln \left(\frac{c}{s} \right) \quad (7)$$

where α is the volume-shape factor, l is the characteristic length, ν is the molecular volume of the crystallizing solute, k_B is Boltzmann's constant, and T is temperature. For homogeneous nucleation

$$\Delta G_S = \beta l^2 \gamma_{12}, \quad (8)$$

where β is the area shape factor and γ is the interfacial energy per unit area between the crystallization medium, 1, and the nucleating cluster, 2. Consequently, the overall free energy change for nucleation is decreased by a large supersaturation ratio (c/s) and by a low interfacial energy.

The factors that regulate nucleation are best appreciated by considering the equation for the rate of homogeneous nucleation from solutions:

$$\begin{aligned} J &= N_o \nu \exp \left(\frac{-\Delta G^*}{k_B T} \right) \\ &= N_o \nu \exp \left(\frac{-4\beta^3 \nu^2 \gamma_{12}^3}{27\alpha^2 (k_B T)^3 (\ln(cs))^2} \right) \end{aligned} \quad (9)$$

J is the number of nuclei formed per unit time per unit volume, N_o is the number of molecules of the crystallizing phase in a unit volume, ν is the frequency of atomic or molecular transport at the nucleus–liquid interface, and ΔG^* is the maximum in the Gibbs free energy change for the formation of clusters at a certain critical size, l^* . The nucleation rate was initially derived for condensation in vapors (23), where the pre-exponential factor is related to the gas kinetic collision frequency. In the case of nucleation from condensed phases, the frequency factor is related to the diffusion process (24). The value of l^* can be obtained by minimizing the free energy function with respect to the characteristic length:

$$l^* = \frac{2\beta \nu \gamma_{12}}{3\alpha k_B T \ln \left(\frac{c}{s} \right)} \quad (10)$$

For spherical clusters, $\alpha = 4\pi/3$ and $\beta = 4\pi$, based on the radius of the cluster. Therefore,

$$r^* = \frac{2\nu \gamma_{12}}{k_B T \ln \left(\frac{c}{s} \right)} \quad (11)$$

Considering these geometric factors, the rate for homogeneous nucleation of spherical clusters is

$$J = N_o \nu \exp \left(\frac{-16\pi v^2 \gamma_1^3 2}{3(k_B T)^3 (\ln(\frac{C}{S}))^2} \right) \quad (12)$$

Whereas the classical theory of nucleation is limited by the implicit assumptions in its derivation (described in detail in Ref. 6), it successfully predicts the nucleation behavior of a system (7–9, 18, 25). Inspection of the aforementioned equation clearly suggests that the nucleation rate can be experimentally controlled by the following parameters: molecular or ionic transport, viscosity, supersaturation, solubility, solid–liquid interfacial tension, and temperature.

Nucleation kinetics are experimentally determined from measurements of nucleation rates, induction times, and metastability zone widths (the supersaturation or undercooling necessary for spontaneous nucleation) as a function of initial supersaturation (6–8, 26). The nucleation rate will increase by increasing the supersaturation while all other variables are constant. However, at constant supersaturation, the nucleation rate will increase with increasing solubility. Solubility affects the pre-exponential factor and the probability of intermolecular collisions. Furthermore, when changes in solvent or solution composition lead to increases in solubility, the interfacial energy decreases because the affinity between crystallizing medium and crystal increases (6). Consequently, the supersaturation required for spontaneous nucleation decreases with increasing solubility (27), as shown in Fig. 6.

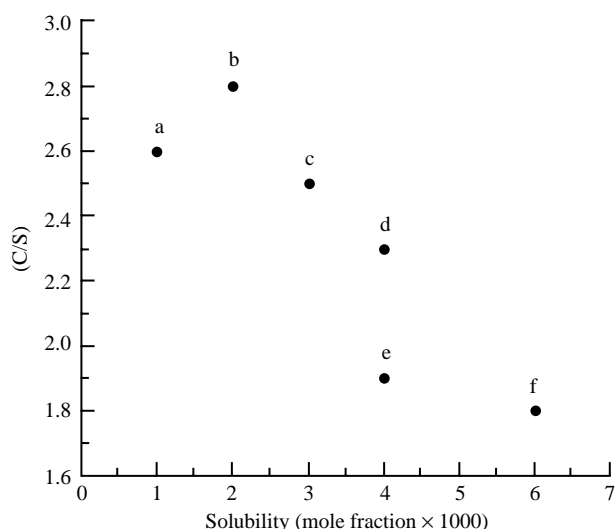


Fig. 6 Dependence of the critical supersaturation for nucleation on solubility of nitrofurantoin in (a) formic acid, (b) formic acid: water (4:1), (c) formic acid: ethanol (2:1), (d) formic acid: dioxane (2:1), (e) formic acid: methanol (2:1), (f) formic acid: water (2:1). (From Ref. 27.)

The dependence of nucleation rate on solubility is also consistent with Ostwald's law of stages (28) regarding the preferential formation of a metastable solid phase. It states the following: "When leaving an unstable state, a system does not seek out the most stable state, rather the nearest metastable state which can be reached with loss of free energy." This indicates that if the unstable solid state modification (the system with highest solubility) precipitates before more thermodynamically stable solid phases, it must have higher nucleation and growth rates than solid states of lower solubility. However, Ostwald's law of stages is not universally valid because the appearance and evolution of solid phases are determined by the kinetics of nucleation and growth under the specific experimental conditions (29–31).

Accounts of nucleation inhibition in the pharmaceutical literature are sometimes confusing because the dependence of the nucleation event (nucleation rate, metastability zone width, or induction time) on supersaturation is not considered. In search of additives that inhibit nucleation, induction times are often measured as a function of additive concentration, whereas the dependence of the nucleation event on supersaturation is neglected. Results from such studies possibly lead to the erroneous conclusion that the additive inhibited nucleation (32, 33) when, indeed, the additive decreased the supersaturation and frequently led to an undersaturated state. Hence, the system is under thermodynamic control instead of kinetic control.

An important factor contributing to the nucleation mechanism and kinetics is the volume of solution in which nucleation occurs. The dispersal of a bulk liquid into a collection of small droplets has been shown to be an effective way of achieving large supersaturations or undercoolings (19, 20). Precipitation and solidification in small volumes (droplets) involving emulsions have been used to study homogeneous nucleation processes (34) and for the control of purity, particle size, and morphology (35, 36). Dispersing a solution into small volumes isolates heterogeneous nucleants within a fraction of the drops and makes nucleation more difficult. Consequently, larger supersaturations need to be reached for nucleation to occur. The boundaries of possible outcomes are represented by the following scenarios: (1) crystals of very small size (even in the nanometer range) are formed as a result of the high nucleation rates (35, 36), or (2) a glass or amorphous solid is formed due to the low diffusion rates of molecules that inhibit the evolution of clusters to crystals within the time scale of the experiment (37).

Nucleation outcomes from solutions with initially the same composition may vary as a consequence of impurities, rates at which supersaturation was created, thermal histories, experimental techniques employed to detect

Table 1 Disodium phosphate crystallization behavior during far-from-equilibrium freezing of buffer solutions

Observation	$C_{\text{Na}_2\text{HPO}_4}^a$ (mM)	Reference
Crystallization begins to decrease	≤ 200	Murase and Franks (16, 38)
Does not crystallize	≤ 10	
Crystallization begins to decrease	≤ 15	Gómez (17)
Crystallization readily occurs	≥ 0.3	
Crystallization is not inhibited by $C_{\text{NaH}_2\text{PO}_4}^b \leq 94$ mM	6	
Readily crystallizes	190	Cavatur and
Crystallization inhibited by $C_{\text{NaH}_2\text{PO}_4} \geq 730$ mM	190	Suryanarayanan (39)

^a $C_{\text{Na}_2\text{HPO}_4}$ = initial disodium phosphate concentration in solution at 25°C.

^b $C_{\text{NaH}_2\text{PO}_4}$ = initial monosodium phosphate concentration in solution at 25°C.

precipitation, and solution volumes in which nucleation occurred. This is illustrated by comparing results of the selective crystallization of buffer components during freezing from various laboratories (16, 17, 38–39). The initial salt concentrations and the crystallization behavior of disodium phosphate during freezing of sodium phosphate buffer solutions are shown in Table 1. Murase et al. (16, 38) report that disodium phosphate precipitation occurs at higher initial buffer and disodium phosphate concentrations, compared to those observed in our laboratory (17). Murase et al. report that salt precipitation begins to decrease at initial buffer concentrations below 500 mM, corresponding to 200 mM disodium phosphate, compared with results from our studies where decreases in salt precipitation were detected at initial buffer concentrations below 8 mM, corresponding to 0.3 mM disodium phosphate. Furthermore, disodium phosphate did not precipitate at initial disodium salt concentrations below 10 mM, whereas our studies show that precipitation occurs at salt concentrations as low as 0.3 mM. Inspection of the experimental conditions under which these studies were done (Table 2) shows a trend towards lower extent of crystallization (and slower crystallization rates) with decreasing volume of solutions from 25 ml to 3 μ l and increasing rates of cooling to lower temperatures.

Neglecting the factors that regulate nucleation leads to misleading generalizations when developing guidelines to control precipitation by the addition of noncrystallizing additives. Consider, for example, the conflicting interpretation of additive effects on nucleation when these are expressed in terms of concentration ratios (additive to crystallizing solute) while ignoring other parameters. Data in Table 1 indicates that disodium phosphate precipitation is inhibited at $(\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}) \leq 4$ (0.73 M/0.19 M) (39) whereas it is not inhibited at $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-} \geq 16$ (0.094 M/0.006M) (17). Compared to the systems we studied, the buffers studied by Cavatur and Suryanarayanan (39) have much higher concentrations of monosodium phosphate (which increases the viscosity of solutions) in addition to smaller solution volumes and faster rates of cooling to lower temperatures. All of these factors contribute to delaying precipitation.

The effect of the viscosity of the crystallization medium on the nucleation rate has been described by Turnbull and Fisher (24). The frequency of atomic or molecular transport at the nucleus-liquid interface, ν , can be related to the bulk viscosity, η , with the Stokes–Einstein relation:

$$\nu \approx \frac{kT}{3\pi a_o^3 \eta(T)} \quad (13)$$

Table 2 Experimental conditions and methods of measuring progress of sodium phosphate crystallization during freezing

Reference	Volume of solution (μ l)	Cooling rate (°C/min)	Temperature (°C)	Method
Murase and Franks (16, 38)	2–5	0.62	–52	DSC, SEM
Gómez (17)	25×10^3	0.3–0.5	–10	pH
Cavatur and Suryanarayanan (39)	300 ^a	15	–40	XRPD

^a Personal communication.

where a_o is the mean effective diameter of the diffusing species. If the viscosity dependence on temperature is described by Arrhenius behavior, then

$$J = \frac{K}{\eta_o} \exp\left(\frac{-\Delta G^* - \Delta G_a}{kT}\right) \quad (14)$$

where ΔG_a is the activation energy for transport across the nucleus–liquid interface. Thus, the nucleation rate may go through a maximum when an increase in undercooling or supersaturation is accompanied by an increase in viscosity. This behavior has been observed in the nucleation of citric acid in aqueous solutions (40) and the crystallization of ice (41).

Heterogeneous Nucleation

Heterogeneous nucleation processes are of fundamental and practical importance in pharmaceutical systems because unintentionally or intentionally added surfaces or interfaces may promote nucleation. The reactivity of crystal surfaces as heterogeneous nucleants has significant consequences in the isolation of the desired solid-state modification and in the control of conversions between these modifications, because the free energy required for the formation of $2 = D$ nuclei is lowered by the presence of an appropriate substrate. Quantitatively, this is described by the following equation (42, 43):

$$\Delta G_s = \gamma_{12}A_{12} + (\gamma_{23} - \gamma_{13})A_{23} \quad (15)$$

where γ is the interaction energy per area, A is the surface area of the interfaces, and the subscript 3 represents the substrate. The total change in surface free energy will be lowered by favorable surface interactions between the aggregate and the substrate and unfavorable interactions between the crystallization medium and the substrate, due to the negative value of the second term in Equation (15). Consequently, increasing the surface area of the substrate will enhance nucleation.

The effectiveness of crystal seeding in controlling crystallization outcomes relies on the potential of crystal surfaces to promote heterogeneous or secondary nucleation (7, 8) while avoiding heterogeneous nucleation mediated by unknown contaminants. A review by Ward (44) on the structure, properties, and reactivity of organic crystal surfaces is recommended to develop strategies for the choice of surfaces that promote nucleation. Various studies (45–49) have demonstrated the influence of substrate topography, lattice parameters, crystallographic symmetry, and intermolecular interactions on surface-directed nucleation.

Heterogeneous nucleation mechanisms can significantly affect dissolution of metastable solid phases

because this form of nucleation can occur at low driving forces. Whereas the choice of a metastable solid phase with a solubility higher than other crystalline modifications is motivated by the expectation of faster dissolution rates, achievement of faster dissolution rates and higher concentrations in solutions is jeopardized by surface-mediated nucleation. We have reported (47) that the surface of the metastable phase of theophylline promoted the nucleation of the stable monohydrate crystals. The observed oriented growth of monohydrate crystals on the anhydrous surface is consistent with a close lattice match between the **b** and **c** crystallographic axes (47). Other studies on the dissolution of metastable solids, such as anhydrous theophylline (50) and anhydrous carbamazepine (51–53), have shown that crystallization of the stable phase occurs during dissolution (Fig. 7). It is unfortunate that, in view of the important influence that nucleation mechanisms have on dissolution of metastable solid phases, very seldom are studies carried out to identify the potential of substrate-mediated nucleation by the metastable modification. The information gained from this

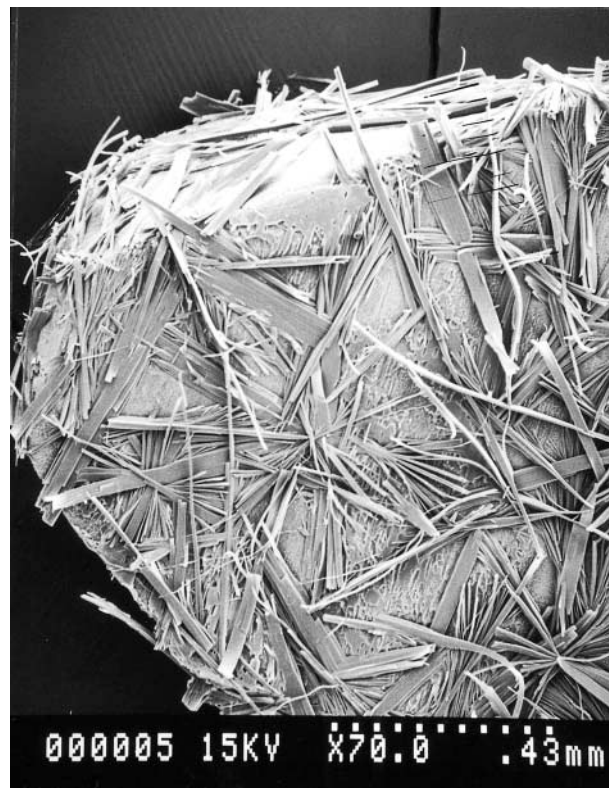


Fig. 7 Nucleation and growth of dihydrate carbamazepine on crystal faces of anhydrous monoclinic carbamazepine in aqueous solutions of sodium lauryl sulfate (17.3 mM, 0.5%), $C/S = 1.15$ at 25°C. (From Ref. 136.)

type of study can be used in the design of methods to regulate crystallization during dissolution as well as during isolation of the desired solid form.

Carter and Ward (43) have identified a surface-mediated nucleation mechanism that involves a geometric shape match between planes of a ledge site on the substrate and planes of prenucleation aggregates. They have applied these concepts to the directed nucleation of polymorphs (45, 48). This work provides us with the attractive possibility that “a library of organic seeds can be used to control polymorphism, or to search for unknown polymorphs” (44). Molecular interpretations based on this approach are experimentally more accessible than those based on solvent-selective polymorph crystallization.

Experimental and Computational Strategies

Whereas nucleation phenomena have their origin at the molecular level, they are often described in terms of macroscopic properties due to the scarcity of experimental techniques that allow for monitoring events at the molecular level. Nevertheless, information about molecular association processes in supersaturated systems obtained by laser Raman spectroscopy and laser light scattering has been used to identify prenucleation clusters and growth units under well-defined experimental conditions. Methods that measure cluster size distributions are more appropriate for studying crystallization of macromolecules (54–58) due to the large sizes of prenucleation clusters, whereas Raman and fluorescence spectroscopic techniques are capable of providing information about the solution structure or the species present in solution (59–63). The implications for crystallization pathways are examined by comparing the solution and crystal structures at a molecular level and by combining information obtained from macroscopic analysis with results from molecular simulations (64–74).

Gavezzotti (75–78) and Desiraju (79–81) have described crystal engineering strategies to understand the molecular aggregation processes involved in crystallization and to elucidate the supramolecular motifs in organic crystals. In this context, crystals are viewed as solid-state supermolecules assembled by intermolecular interactions, with the basic approach of establishing a relation between molecular interactions and supramolecular structure. Recent studies by Gavezzotti (78) show how molecular dynamics calculations allow for simulation of solvent and kinetic effects on molecular aggregation.

The supramolecular assembly process can be controlled so that the precursor nuclei in solution adopt a structure that resembles the structure of the desired crystalline

modification (78, 82, 83). This concept has been used in the design of nucleation inhibitors to prevent growth of the stable polymorph and enhance the growth of the metastable polymorph (29, 84, 85). Davey et al. (64) have explained the solvent-dependent polymorph appearance of sulfathiazole by analyzing the intermolecular interactions in the various polymorphic structures and comparing them with the supramolecular assemblies that could exist in the different solvents. In this case, however, the solvent-dependent selective crystallization of a polymorph was not correlated with solubility (64, 86).

CRYSTAL GROWTH

As stable nuclei (those larger than the critical cluster size) form, they grow into macroscopic crystals. This portion of the crystallization process is known as crystal growth. This process consists of several stages through which the growth units (i.e., the crystallographic basis that is “tacked onto” the space lattice to form the crystal structure) pass. These include: (1) transport of the growth unit from or through the bulk solution to an impingement site, which is not necessarily the final growth site, (2) adsorption of the growth unit at the impingement site, (3) diffusion of the growth units from the site of impingement to a growth site, and (4) incorporation into the lattice.

Desolvation of the growth unit may occur anywhere in steps 2–4, or the solvent may be adsorbed with the growth unit. Because any of these steps can be the rate-limiting step in the crystal growth process and they are dependent on conditions such as supersaturation, temperature, additives or solvent, and the hydrodynamics of the system, crystal growth is generally divided into two main mechanisms: volume-diffusion controlled or surface-integration controlled (Fig. 8) (6–9).

Crystal growth is volume-diffusion controlled when the diffusion of molecules from the bulk to the crystal surface is the rate-limiting step whereas growth is considered surface-integration controlled if the incorporation of a growth unit into the lattice is the slowest process. Many crystallization studies involving proteins, small organic electrolytes, and nonelectrolytes have reported growth controlled by surface-integration (71–73, 87–94). The roughness of the crystal surface determines whether growth occurs by the continuous (relatively rough surfaces) or the layer (relatively smooth surfaces) mechanism.

Crystal growth by the layer growth mechanism describes the formation of steps (i.e., layers) by two different mechanisms—2-D nucleation and screw dislocation. The model for 2-D nucleation was developed by

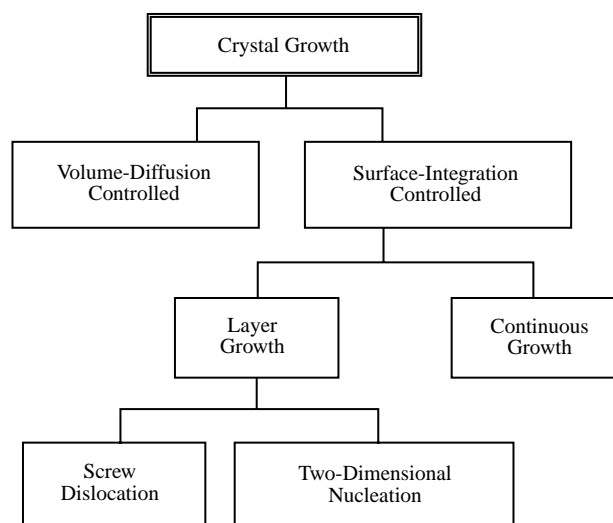


Fig. 8 Mechanisms for crystal growth.

Volmer (95) and Stranski (96). The screw dislocation model was first described by Burton, Cabrera, and Frank (BCF;97). The details of the derivations for these models have been summarized in a number of other references (7, 9, 98, 99).

Two-dimensional nucleation occurs when nuclei at the crystal surfaces act as sources of steps that allow for the further incorporation of growth units. In general, this mechanism accounts for the crystal growth observed at high supersaturations. The screw dislocation mechanism often accounts for growth at lower supersaturations. When the supersaturation is below the threshold for formation of 2-D nuclei, the presence of screw dislocations provides a source of steps for the addition of growth units in an infinite sequence of equidistant and parallel steps.

Experimental Strategies

Depending on the objectives and applications of growth rate measurements, the growth rate may be expressed as:

1. overall linear growth rate, which is the rate of change of the volume equivalent diameter with time;
2. linear growth rate of a face, which is the rate of displacement of a crystal face in a direction perpendicular to the face; and
3. velocity, height, and spacing of growth steps spreading across a crystal surface.

The linear growth rate of a face can be expressed in terms of the step velocity, step height, and step spacing. Techniques used for in situ measurement of crystal growth rates as a function of supersaturation include monitoring:

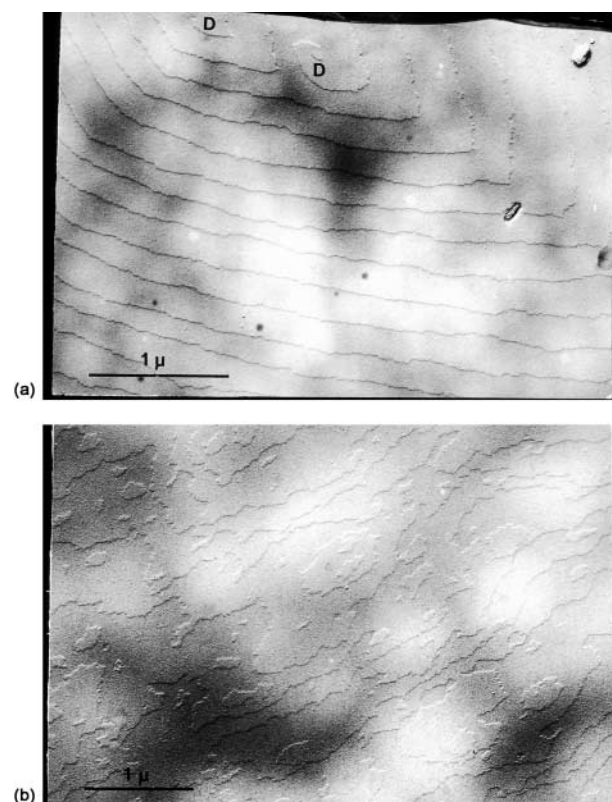


Fig. 9 Surfaces of human insulin crystals showing that steps arise by different mechanisms depending on supersaturation. (a) At low supersaturation, steps are created by screw dislocations, D. (b) At high supersaturations, $c/s > 10$, steps occur at the edges of islands as a consequence of 2-D nucleation. (From Ref. 136.)

1. the crystal population by methods that measure particle size and number (100, 101);
2. the growth rates of individual crystal faces by optical microscopy with the use of a flow cell system (73, 87–90); and
3. the development of surface topography at the molecular level by atomic force microscopy (AFM 89, 91–94) and interferometric microscopy (71, 72).

Although various mathematical models have been used to identify the growth mechanism by fitting various equations to the experimentally measured face-growth rate dependence on supersaturation, it may be difficult to discriminate between mechanisms (73, 89, 102). This approach is often combined with examination of the surface topography to confirm the growth mechanism, as shown for human insulin crystals in Fig. 9. Monitoring the development of surface topography and transport processes during crystal growth will reveal events that are not evident from monitoring growth rates with optical microscopy. For instance, interactions of growth units

and additives with different crystal planes exposed on a surface may be deduced from the shape of 2-D nuclei and the kinetic anisotropy of the growth steps along crystallographic directions. These techniques have been successfully applied for identifying the crystal growth mechanisms and kinetics of small molecules (71, 72, 89) and proteins (91–94, 103).

CRYSTAL MORPHOLOGY

Crystal growth is governed by both internal and external factors. Internal factors, such as the 3-D crystal structure and crystal defects, will determine the nature and strength of the intermolecular interactions between the crystal surface and the solution; whereas external factors, such as temperature, supersaturation, solvent, and the presence of impurities, will affect the type of interactions at the solid–liquid interface. The external shape or morphology of a crystal is a consequence of the relative growth rates of the faces, with the slowest growing faces impacting the crystal morphology to the largest extent. Examination of crystal morphology can reveal the molecular events occurring at the crystal face–liquid interfaces during growth. Consequently, even when morphology does not play a significant role in quality control, studying it is essential to understand the kinetics of crystallization (104).

Crystal Structure and Nomenclature

Crystals are comprised of the long-range, 3-D periodic order of atoms held together through intramolecular covalent and ionic bonds and intermolecular noncovalent bonds. It is this long-range order that differentiates a crystalline solid from a glassy or amorphous solid which only demonstrates periodic order over a short range of atoms. When a chemical entity exists in more than one crystalline state—polymorph or solvate—each of these forms will have a different arrangement of atoms in the 3-D structure. Consequently, in addition to different physical and chemical properties, each of these forms has a unique crystal structure.

Considering the crystal from a purely geometric view (i.e., ignoring the exact arrangement of atoms), each crystal can be described by the smallest repeating translational unit in three dimensions, known as the unit cell. A unit cell is a parallelepiped, which can itself be described by six geometric measurements, three axes (designated “a,” “b,” and “c”) and the angles between these axes (designated “α,” “β,” and “γ”), known as the lattice parameters. The relationship between the six lattice

Table 3 Crystal systems and lattice parameters

System	Relationship between lattice parameters
Cubic	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$
Tetragonal	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$
Orthorhombic	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$
Rhombohedral (or Trigonal)	$a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$
Hexagonal	$a = b \neq c$ $\alpha = \beta = 90^\circ, \gamma = 120^\circ$
Monoclinic	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ, \beta \neq 90^\circ$
Triclinic	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90^\circ$

parameters defines to which of the seven unique crystal systems a particular crystal belongs (Table 3). Whereas the crystal system describes the geometric arrangement of the unit cell, the space group describes the individual translational and symmetrical relationships of the actual atoms within the unit cell. The combination of the lattice parameters, crystal system, space group and, ultimately, the individual atomic coordinates defines each unique crystal structure that exists for a crystalline solid. An example of the unit cell (with lattice parameters) and atomic coordinates for an anhydrous polymorph of the drug carbamazepine is shown in Fig. 10. This polymorph belongs to the monoclinic crystal system and the $P2_1/n$ space group (105).

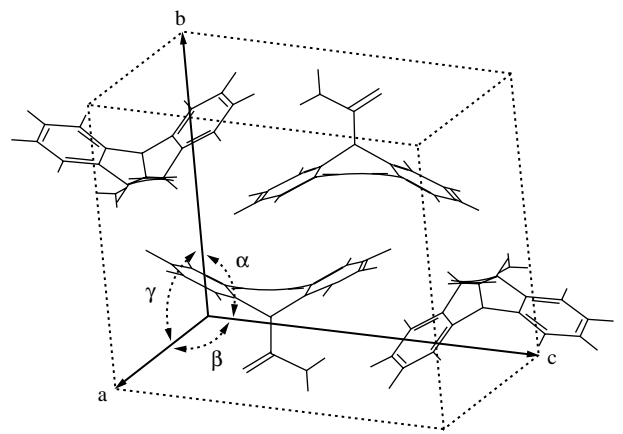


Fig. 10 Unit cell of anhydrous monoclinic carbamazepine. Lattice parameters are marked with “a,” “b,” “c,” “α,” “β,” and “γ.” (From Ref. 105.)

Table 4 Nomenclature for anhydrous carbamazepine polymorphs

	CBZ(AM)	CBZ(ATrg)	CBZ(ATrc)	Reference
Crystal system	Monoclinic	Trigonal	Triclinic	
T_{melt}	175–180°C	192–194°C	192–195°C	
Nomenclature	III		I	106–109
	I	II	III	110–112
	beta	alpha		113
	monoclinic	trigonal		114
			triclinic	115

The convention of Miller indices is used to describe the planes within a unit cell. Miller indices are defined as the reciprocals of the intercept which the plane makes with each of the three crystal axes. Each plane is denoted by three parameters “*h*,” “*k*,” and “*l*.” Planes that are parallel to a crystal axis are given the Miller index of 0, whereas planes formed in the negative direction are written with a bar over the number in the Miller index. The Miller index of a single specific plane is written within parentheses (*hkl*), whereas the Miller indices describing a whole family of faces are written with braces {*hkl*}. The faces that exist and define the crystal morphology are termed morphologically important and are commonly identified by the Miller indices of the planes represented by those faces.

The use of nomenclature indicative of the crystal structure would alleviate much of the confusion that exists in the literature when describing different polymorphs or solvates. Whereas it is often convenient to use terminology such as “Form I and Form II” or “ α -form versus β -form,” this can lead to confusion when different investigators use inconsistent or, in some cases, conflicting nomenclature to describe the same polymorphic forms. For instance, carbamazepine has been shown to exist in at least three anhydrous forms and two solvated forms (a dihydrate and an acetate). Overlapping or conflicting nomenclature has been used when describing the three anhydrous forms and has led to much confusion (Table 4) (106–115). As can be seen, nomenclature arising from crystallographic information (in this case, crystal system) would be more meaningful and less confusing than the commonly used terminology.

The Role of Crystal Morphology in Pharmaceutical Processes

Knowledge of crystal morphology and ways to control and predict the morphology of drugs and excipients could prove invaluable in processing and product development. In a review on solid-state properties of powders, York has

discussed how changes in crystal morphology can affect several formulation parameters, such as particle size distribution, powder flow, mixing (and agglomeration), dissolution, compression, and tablet hardness (116). Crystal morphology could also impact the rheology and filtration of suspensions. Additionally, certain crystal habits may lead to lack of dose uniformity or erratic drug delivery, as is the case with aerosols where particle size is critical for delivery (117–119). It should also be emphasized that the effect of crystal morphology does not impact the active ingredient solely. Because excipients often comprise much larger percentages of the formulation, the crystal morphology of each component may have profound effects on the characteristics previously described.

Often during processing or delivery of a formulation, supersaturated states are created, which may lead to the crystallization of drug or excipient. Consequently, the interactions between drugs and excipient (or excipient–excipient interaction) could have a tremendous impact on the morphology. In the strictest sense, the excipient is an impurity with respect to the active ingredient (or vice versa) and may result in changes in the crystal morphology that could adversely effect both processing and/or delivery.

The presence of additives (intentional) or impurities (unintentional) have long been known to have a significant impact on crystallization (both nucleation and growth; 7, 9). If an additive (or impurity) selectively adsorbs to a crystal face, the growth rates of those faces will be altered resulting in a change in crystal morphology. For example, interaction of sodium taurocholate (STC) with the {111} faces of carbamazepine dihydrate (CBZ(D)) decreases the growth rate along the needle axis and results in more prismatic crystals (Fig. 11) (53, 120). Photomicrographs demonstrating the effects of STC on CBZ(D) morphology are seen in Fig. 12. Lechuga-Ballesteros and Rodríguez-Hornedo have shown how the adsorption of hydrophobic L-amino acids can inhibit the growth rate of the {012} faces of L-alanine resulting in a change in crystal morphology (89, 90, 101, 121). The presence of structurally similar additives has also been shown to significantly change the crystal

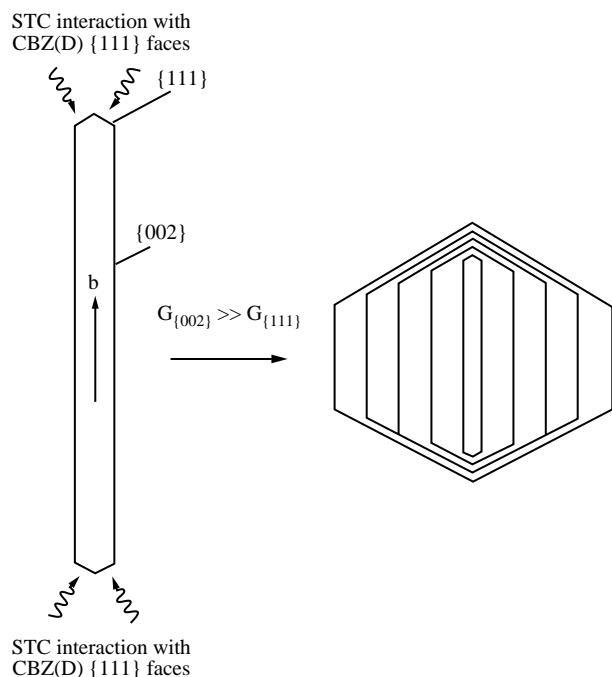


Fig. 11 Two-dimensional representation of the effect of sodium taurocholate (STC) on the crystal morphology of dihydrate carbamazepine (CBZ(D)). STC interacts specifically with the {111} faces of CBZ(D), resulting in inhibition of the growth of those faces and changing the morphology from needles to prisms.

morphology of paracetamol (70). Recently, there has been much work exploiting the selective interaction of additives with crystal faces in order to “tailor” the crystal morphology to a more favorable habit (122).

Monitoring the changes in morphology by visual examination can be a powerful tool when troubleshooting formulation problems. The appearance of crystals with different morphology is indicative of changes in the processing conditions, such as in the degree of supersaturation, rate of heating or cooling, or the presence (or absence) of impurities. Changes in morphology could also be the first suggestion of the appearance of a new crystal form (polymorph or solvate). Characterization techniques, such as x-ray powder diffraction, thermal analysis, or spectroscopy, can be used to identify the appearance of new forms first identified by observing changes in morphology.

Morphology Prediction and Other Computational Strategies

With the objective of predicting the morphology of crystals, several models have been developed. One of the first models for predicting crystal morphology is based on

the general Law of Bravais (123), which states that crystal faces with the most morphological importance have the greatest interplanar distance, d_{hkl} . The Bravais–Friedel–Donnay–Harker (BFDH) model predicts that the morphological importance of a face is proportional to the interplanar spacing corrected for extinction conditions and translational repetitions of the crystal space group (123–125). Consequently, the growth rate of any given face is inversely proportional to its interplanar spacing:

$$G_{hkl} \propto \frac{1}{d_{hkl}} \quad (16)$$

where G_{hkl} is the growth rate of a face and d_{hkl} is the interplanar spacing. This model has been used to predict the morphology for many small molecules including compounds of pharmaceutical interest. (3, 73, 120, 126, 127). The observed (Fig. 13a) and predicted morphology (Fig. 13b) of carbamazepine dihydrate shows that the BFDH model correctly predicts the axis of elongation as well as the appearance of the {002}, {200}, and {111} faces. This model, however, did not correctly predict the aspect ratio of the crystal because it is based on geometric parameters solely. It considers neither molecular interactions nor the effect of anisotropic forces, such as intermolecular hydrogen bonding or π – π interactions.

Early work by Hartman and Perdok (128) described crystal growth in terms of the formation of strong bonds between neighboring crystallizing units. Uninterrupted straight chains of these bonds were classified as periodic bond chains (PBC). This theory led to the classification of three types of crystal faces: F-faces (flat), S-faces (stepped), and K-faces (kinked), based on the number of PBCs in a slice thickness, d_{hkl} . K-faces, which had no PBCs present in a slice were shown to be the fastest growing, whereas the F-faces, which had two or more PBCs, grew the slowest. Consequently, the F-faces were the most morphologically important faces. The PBC theory ultimately led to the development of the attachment energy (AE) model (122, 129) which states that the growth rate of a face, G_{hkl} , is proportional to its attachment energy, E_{att} :

$$G_{hkl} \propto E_{att} \quad (17)$$

where the attachment energy can be defined as the energy released when one growth layer (d_{hkl}) crystallizes on the surface of a crystal. The attachment energy can be calculated as the difference between the energy of crystallization (or lattice energy) and the energy of a growth slice:

$$E_{att} = E_{crystal} - E_{slice} \quad (18)$$

The values for the energy of crystallization can be calculated by summation of the intermolecular interactions

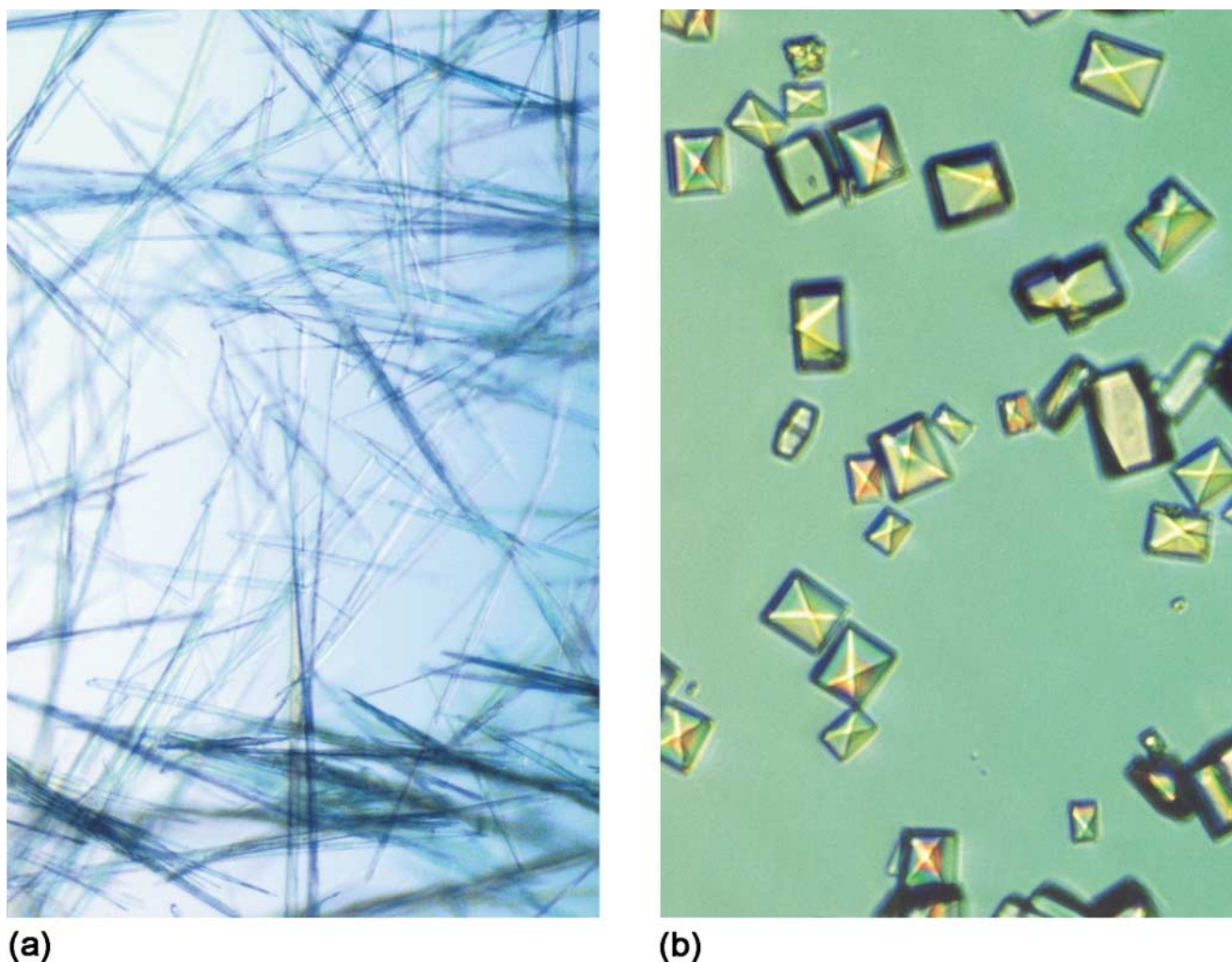


Fig. 12 Photomicrographs of CBZ(D) crystals grown in aqueous solutions $C/S = 2.5$ at 25°C . (a) In the absence of surfactant. (b) In the presence of 9 mM STC. (From Ref. 53.)

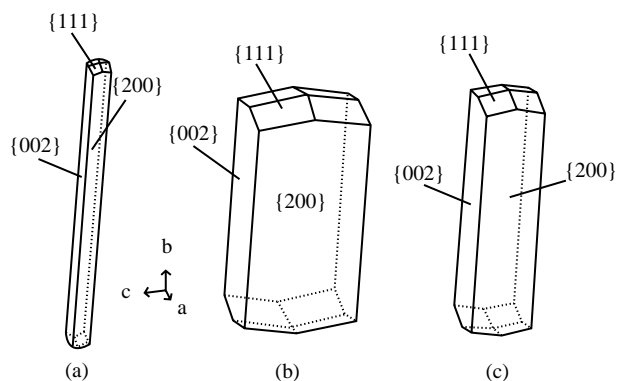


Fig. 13 Dihydrate carbamazepine morphology: (a) Observed morphology when grown from water, (b) Predicted by the BFDH model, (c) Predicted by the AE model. (From Ref. 53.)

from a central reference molecule and all other molecules within the crystal. Similarly, the energy of a growth slice can be calculated for all molecules in that particular slice.

The AE method has been used to successfully predict the morphology for many molecular crystals, especially those that are dominated by strong anisotropic forces (3, 73, 120, 126, 127). In the case of carbamazepine dihydrate, which is dominated by intermolecular hydrogen bonding (especially along the **b**-axis), this model predicts a morphology much closer to the observed morphology (Fig. 13c) (53). Whereas this method can account for anisotropic forces, predictions are very dependent on the force field chosen to calculate the nonbond intermolecular interactions. Much care should be given to the choice of appropriate force fields for the molecules in question. The force field developed by Lifson and coworkers (130),

which is explicitly parameterized for C, H, N, and O atoms, was used in the AE method prediction of carbamazepine dihydrate (53).

Morphology predictions based on the AE method also assume that crystals are grown in vacuo, i.e., in the absence of solvent. Consequently, the solvent effects on morphology are ignored using this method. Roberts et al. have developed a method that calculates a modified attachment energy that can be used to predict the effects of solvents as well as other additives on crystal morphology (131).

With knowledge of the crystal structure, morphology, and intermolecular interactions present, strategies have been developed for the rational choice of additives that will specifically inhibit the growth of the stable polymorph and enhance the growth of a metastable polymorph (29, 64, 65, 74, 84–85, 132–134). These strategies can also be applied to systems for which a change in crystal morphology is desired to aid in product development or delivery (e.g., creating a prismatic habit when a crystal grows preferentially as a needle). The main features of these strategies are:

1. identification of the fastest growing faces of the stable crystalline modification (from experimental data as well as morphology predictions);
2. characterization of intermolecular interactions along the crystallographic direction with fastest growth;
3. selection of additives that can be incorporated into (or adsorbed onto) the crystalline structure along this direction; and
4. development of experimental methods to investigate the effectiveness of the additive to kinetically stabilize the metastable modification.

The selection of additives can be guided by molecular visualizations of the crystal structure based on geometric fits (29, 90) or binding energy calculations (70, 120, 135).

SOLUTION-MEDIATED TRANSFORMATIONS

Knowledge of the propensity of a metastable solid phase to dissolve in a liquid phase from which a stable solid phase nucleates and grows is crucial in many stages of pharmaceutical development. This is because pharmaceutical solids are designed to be dissolved and to come in contact with solvents from the early stages of development (isolated by crystallization from solution) and during processing (wet granulation, spray-drying, freeze-drying, etc.). Given that the sudden disappearance or appearance of a crystalline modification can threaten process development, characterization of the kinetics and mechanisms of

solvent-mediated transformations is of practical importance. It will provide answers to questions such as: What is the relation between processing conditions and the solid-state modification manufactured? Is there a correlation between dissolution conditions, solid phase(s) dissolving, and concentration of drug dissolved?

The importance of phase transition kinetics, molecular interpretations, and process implications are emphasized by several investigators (29–31, 47, 64, 133, 136–138). Cardew and Davey (30) developed a theoretical framework to investigate solvent-mediated transformations in terms of dissolution kinetics of one phase and growth of a second phase. The model represents the time development of the supersaturation with respect to the stable phase, or solute concentration in solution, and to the solid phase composition during the transformation. This experimental approach involves saturating the solution with respect to the metastable phase under consideration and to monitor both solution concentration and solid phase composition in the presence of the metastable phase, under constant external conditions. More useful information is obtained from the concentration or supersaturation profiles than from the solid phase composition profiles with time because the former is related to the driving forces that regulate the transformation rate and can be used to identify the rate-controlling process: dissolution or growth. A calculated supersaturation profile with growth-limited and dissolution-limited regimes is shown in Fig. 14 for the case in which dissolution and growth are linearly dependent on supersaturation (30, 31). Experimental studies of the phase transitions of organic crystals have shown this model to be applicable in explaining the solution-mediated transformation kinetics of polymorphs (30, 31, 138) and solvents (47) and to be applicable to process development (138).

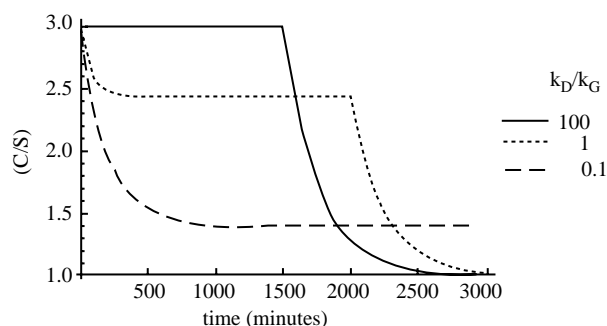


Fig. 14 Simulation of the supersaturation–time profiles as a function of the relative rates of dissolution and crystallization during a solution–mediated transformation. Generated from a kinetic model developed by Cardew and Davey (30), assuming that both dissolution and growth rates are linearly dependent on their respective driving forces. (Adapted from Ref. 136.)

Table 5 Information on crystallization kinetics and mechanisms provided by various experimental techniques

Techniques	Information
Optical microscopy (inverted microscope)	Study crystallization processes in situ Monitor transformations in suspensions Determine transformation times Screen and characterize additive/solvent interactions with specific crystal faces Identify nucleation mechanisms Measure crystal growth rates
Electron microscopy	Characterize additive/solvent interactions with specific crystal faces Identify nucleation and growth mechanisms
Atomic force microscopy	Study crystallization processes in situ
Interferometric microscopy	Examine surface topography Identify nucleation and growth mechanisms Measure crystal growth rates
Raman spectroscopy	Monitor molecular association processes that direct nucleation and crystal growth
Infrared (FT-IR, NIR) spectroscopy, solid-state NMR Spectroscopy	Monitor structure (polymorphs and solvates), conformation, and intermolecular environment of crystals
Spectrophotometry, chromatography	Monitor concentration of solute in solution and supersaturation
Diffraction, Calorimetry, Spectroscopy	Monitor solid phase composition

Some useful experimental techniques for studying small and large scale crystallization and solution-mediated transformations are summarized in Table 5.

SUMMARY

In this chapter, we have discussed the significance of crystallization mechanisms and kinetics in directing crystallization pathways and have presented numerous examples that confirm the importance of controlling the crystallization events (both nucleation and growth) in the pharmaceutical industry. Advances in computational and analytical techniques now provide access to the molecular events that direct nucleation and crystal growth. Consequently, there is no reason for the development of ill-defined crystallization processes even when the desired product is obtained. Understanding the thermodynamic and kinetic behavior of the system is vital for the design of reliable processes; thus, at least a holistic approach that identifies the relevant experimental parameters is essential. Whereas the objectives of formulation or process development may not explicitly include interpretations at the molecular level, the information concealed in the interfaces present during the process, the structure of solutions, and the solids harvested can be significant in the identification of nucleation and growth mechanisms. However, there may be instances, such as the unintentional isolation of a new crystal structure or morphology with

poor formulation characteristics, where knowledge of the molecular level events could prove invaluable in solving formulation or processing problems. Although the significance of crystallization mechanisms and kinetics has been underestimated in the pharmaceutical industry, we hope that the recent developments in computational and analytical techniques will inspire interest in their application to pharmaceutical systems.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the National Science Foundation, the Pharmaceutical Research and Manufacturers of America Foundation, Inc., Fellowship for Advanced Predoctoral Training in Pharmaceutics, the Valteich Research Award, and the Fred W. Lyons, Jr., Fellowship from the University of Michigan College of Pharmacy.

REFERENCES

1. Dunitz, J.D.; Bernstein, J. Disappearing Polymorphs. *Acc. Chem. Res.* **1995**, 28(4), 193–200.
2. Abbott Laboratories. Letter to Health Care Providers <http://www.fda.gov/medwatch/safety/1998/norvir.htm> (accessed Mar 2000).
3. Nichols, G.; Frampton, C.S. Physicochemical Characterization of the Orthorhombic Polymorph of Paracetamol Crystallized from Solution. *J. Pharm. Sci.* **1998**, 87, 684–692.

4. Haisa, M.; Kashino, S.; Maeda, H. The Orthorhombic Form of P-Hydroxyacetanilide. *Acta. Crystallogr.* **1974**, B30, 2510–2512.
5. Food and Drug Administration. *Guideline for Submitting Supporting Documentation in Drug Applications for the Manufacture of Drug Substances*; FDA Center for Drug Evaluation and Research; Office of Drug Evaluation: Rockville, Maryland, 87.
6. Nyvlt, J.; Söhnel, O.; Matuchová, M.; Broul, M. *The Kinetics of Industrial Crystallization*; Elsevier: New York, 1985.
7. Mullin, J.W. *Crystallization*; Butterworth-Heinemann Ltd., Oxford, 1993.
8. Söhnel, O.; Garside, J. *Precipitation: Basic Principles and Industrial Applications*; Butterworth-Heinemann Ltd.: Oxford, 1992.
9. Myerson, A.S. *Handbook of Industrial Crystallization*; Butterworth-Heinemann Ltd.: Oxford, 1993.
10. Mersmann, A. *Crystallization Technology Handbook*; Marcel Dekker, Inc.: New York, 1995.
11. Van den Berg, L. The Effect of Addition of Sodium and Potassium Chloride to the Reciprocal System: KH_2PO_4 – Na_2HPO_4 – H_2O on pH and Composition during Freezing. *Arch. Biochem. Biophys.* **1959**, 84, 305–315.
12. Van den Berg, L. pH Changes in Buffers and Foods During Freezing and Subsequent Storage. *Cryobiology*, **1966**, 3, 236–242.
13. Larsen, S.S. Studies on Stability of Drugs in Frozen Systems VI: The Effect of Freezing upon pH for Buffered Aqueous Solutions. *Arch. Pharm. Chem. Sci. Ed.* **1973**, 1, 41–53.
14. MacKenzie, A.P. Non-equilibrium Freezing Behaviour of Aqueous Systems. *Philos. Trans. R. Soc. London, Ser. B.* **1977**, 278, 167–189.
15. Taylor, M.J. Physico-Chemical Principles in Low Temperature Biology. *The Effects of Low Temperatures on Biological Systems*; Grout, B.W.W., Morris, G.J., Eds.; E. Arnold: London, 1987; 1–71.
16. Murase, N.; Franks, F. Salt Precipitation during the Freeze-Concentration of Phosphate Buffer Solutions. *Biophys. Chem.* **1989**, 34(3), 293–300.
17. Gómez, G. *Crystallization-related pH Changes during Freezing of Sodium Phosphate Buffer Solutions*; University of Michigan: Ann Arbor, MI, 1995.
18. Zettlemoyer, A.C. *Nucleation*; Marcel Dekker, Inc.: New York, 1969.
19. Perepezko, J.H. Nucleation Reactions in Undercooled Liquids. *Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing*. **1994**, 178(1–2), 105–111.
20. Perepezko, J.H. Kinetic Processes in Undercooled Melts. *Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing*. **1977**, 226, 374–382.
21. Gibbs, J.W. *Collected Works*; Vol. I. Thermodynamics Yale University Press: New Haven, 1948.
22. Volmer, M. *Kinetik Der Phasenbildung*; Steinkopff: Leipzig, 1939.
23. Becker, R.; Döring, W. Kinetische Behandlung Der Keimbildung in Übersättigten Dämpfen. *Annalen Der Physik* **1935**, 24, 719–752.
24. Turnbull, D.; Fisher, J.C. Rate of Nucleation in Condensed Systems. *J. Chem. Phys.* **1949**, 17, 71–3.
25. Boistelle, R.; Astier, J.P. Crystallization Mechanisms in Solution. *J. Cryst. Growth* **1988**, 90, 14–30.
26. Hendriksen, B.A.; Grant, D.J.W. The Effect of Structurally Related Substances on the Nucleation Kinetics of Paracetamol (Acetaminophen). *J. Cryst. Growth* **1995**, 156, 252–260.
27. Garti, N.; Tibika, F. Habit Modifications of Nitrofurantoin Crystallized from Formic Acid Mixtures. *Drug Dev. Ind. Pharm.* **1980**, 6, 379–398.
28. Ostwald, W. Studien Über Die Bildung Und Umwandlung Fester Körper. *Zeitschrift Für Physikalische Chemie* **1897**, 22, 289–330.
29. Davey, R.J.; Blagden, N.; Potts, G.D.; Docherty, R. Polymorphism in Molecular Crystals: Stabilization of a Metastable Form by Conformational Mimicry. *J. Am. Chem. Soc.* **1997**, 119(7), 1767–1772.
30. Cardew, P.T.; Davey, R.J. The Kinetics of Solvent-Mediated Phase Transformations. *Proc. R. Soc. Lond. Ser. A* **1985**, 398, 415–428.
31. Davey, R.J.; Cardew, P.T.; McEwan, D.; Sadler, D.E. Rate Controlling Processes in Solvent-Mediated Phase Transformations. *J. Cryst. Growth* **1986**, 79, 648–653.
32. Pellett, M.A.; Castellano, S.; Hadgraft, J.; Davis, A.F. The Penetration of Supersaturated Solutions of Piroxicam Across Silicone Membranes and Human Skin In Vitro. *J. Controlled Release* **1997**, 46(3), 205–214.
33. Katzhendler, I.; Azoury, R.; Friedman, M. Crystalline Properties of Carbamazepine in Sustained Release Hydrophilic Matrix Tablets Based on Hydroxypropylmethylcellulose. *J. Controlled Release* **1998**, 54(1), 69–85.
34. Davey, R.J.; Hilton, A.M.; Garside, J. Crystallization from Oil in Water Emulsions: Particle Synthesis and Purification of Molecular Materials. *Chem. Eng. Res. Des.* **1997**, 75(A2), 245–251.
35. Davey, R.J.; Garside, J.; Hilton, A.M.; McEwan, D.; Morrison, J.W. Purification of Molecular Mixtures below the Eutectic by Emulsion Crystallization. *Nature* **1995**, 375(6533), 664–666.
36. Davey, R.J.; Garside, J.; Hilton, A.M.; McEwan, D.; Morrison, J.W. Emulsion Solidification of Meta-Chloronitrobenzene: Purification and Crystallisation. *J. Cryst. Growth* **1996**, 166(1–4), 971–975.
37. Turnbull, D. Under what Conditions can a Glass be Formed? *Contemp. Phys.* **1969**, 10, 473–488.
38. Murase, N.; Echlin, P.; Franks, F. The Structural States of Freeze-Concentrated and Freeze-Dried Phosphates Studied by Scanning Electron-Microscopy and Differential Scanning Calorimetry. *Cryobiology* **1991**, 28(4), 364–375.
39. Cavatur, R.K.; Suryanarayanan, R. Characterization of Frozen Aqueous Solutions by Low Temperature X-ray Powder Diffractometry. *Pharm. Res.* **1998**, 15(2), 194–199.
40. Mullin, J.W.; Leci, C.L. Some Nucleation Characteristics of Aqueous Citric Acid Solutions. *J. Cryst. Growth* **1969**, 5, 75–76.
41. Franks, F.; Mathias, S.F.; Trafford, K. The Nucleation of Ice in Undercooled Water and Aqueous Polymer Solutions. *Colloids Surf.* **1984**, 11, 275–285.
42. Fletcher, N.H. Nucleation by Crystalline Particles. *J. Chem. Phys.* **1963**, 38, 237.

43. Carter, P.W.; Ward, M.D. Topographically Directed Nucleation of Organic-Crystals on Molecular Single-Crystal Substrates. *J. Am. Chem. Soc.* **1993**, *115*(24), 11521–11535.
44. Ward, M.D. Organic Crystal Surfaces: Structure, Properties and Reactivity. *Curr. Opin. Colloid Interface Sci.* **1997**, *2*(1), 51–64.
45. Carter, P.W.; Ward, M.D. Directing Polymorph Selectivity During Nucleation of Anthranilic Acid on Molecular Substrates. *J. Am. Chem. Soc.* **1994**, *116*(2), 769–770.
46. Black, S.N.; Bromley, L.A.; Cottier, D.; Davey, R.J.; Dobbs, B.; Rout, J.E. Interactions at the Organic Inorganic Interface—Binding Motifs for Phosphonates at the Surface of Barite Crystals. *J. Chem. Soc., Faraday Trans.* **1991**, *87*(20), 3409–3414.
47. Rodríguez-Hornedo, N.; Lechuga-Ballesteros, D.; Wu, H.J. Phase Transition and Heterogeneous Epitaxial Nucleation of Hydrated and Anhydrous Theophylline Crystals. *Int. J. Pharm.* **1992**, *85*(1–3), 149–162.
48. Bonafede, S.J.; Ward, M.D. Selective Nucleation and Growth of an Organic Polymorph by Ledge-Directed Epitaxy on a Molecular-Crystal Substrate. *J. Am. Chem. Soc.* **1995**, *117*(30), 7853–7861.
49. Davey, R.J.; Maginn, S.J.; Andrews, S.J.; Black, S.N.; Buckley, A.M.; Cottier, D.; Dempsey, P.; Plowman, R.; Rout, J.E.; Stanley, D.R.; Taylor, A. Morphology and Polymorphism in Molecular Crystals—Terephthalic Acid. *J. Chem. Soc., Faraday Trans.* **1994**, *90*(7), 1003–1009.
50. De Smidt, J.G.; Fokkens, J.G.; Grijseels, H.; Crommelin, D.A.J. Dissolution of Theophylline Monohydrate and Anhydrous Theophylline in Buffer Solutions. *J. Pharm. Sci.* **1986**, *75*, 497–501.
51. Luhtala, S.; Kahela, P.; Kristofferson, E. Effect of Benzalkonium Chloride on Crystal Growth and Aqueous Solubility of Carbamazepine. *Acta Pharmaceutica Fennica* **1990**, *99*, 59–68.
52. Luhtala, S. Effect of Sodium Lauryl Sulfate and Polysorbate 80 on Crystal Growth and Aqueous Solubility of Carbamazepine. *Acta Pharm. Nord.* **1992**, *4*, 85–90.
53. Murphy, D. *The Solvent-Mediated Phase Transformation of Carbamazepine and the Influence of Surfactants on the Nucleation Mechanism and Crystal Morphology*; University of Michigan: Ann Arbor, MI, 1997.
54. Kadima, W.; McPherson, A.; Dunn, M.F.; Jurnak, F. Precrystallization Aggregation of Insulin by Dynamic Light-Scattering and Comparison With Canavalin. *J. Cryst. Growth* **1991**, *110*(1–2), 188–194.
55. Georgalis, Y.; Umbach, P.; Raptis, J.; Saenger, W. Lysozyme Aggregation Studied by Light Scattering 2. Variations of Protein Concentration. *Acta Crystallogr. Sect. D—Biological Crystallography* **1997**, *53*, 703–712.
56. Georgalis, Y.; Umbach, P.; Zielenkiewicz, A.; Utzig, E.; Zielenkiewicz, W.; Zielenkiewicz, P.; Saenger, W. Microcalorimetric and Small Angle Light Scattering Studies on Nucleating Lysozyme Solutions. *J. Am. Chem. Soc.* **1997**, *119*(49), 11959–11965.
57. Peters, R.; Georgalis, Y.; Saenger, W. Accessing Lysozyme Nucleation with a Novel Dynamic Light Scattering Detector. *Acta Crystallogr. Sect. D—Biological Crystallography* **1998**, *54*, 873–877.
58. Rosenberger, F.; Vekilov, P.G.; Muschol, M.; Thomas, B.R. Nucleation and Crystallization of Globular Proteins—What we Know and What is Missing. *J. Cryst. Growth* **1996**, *168*(1–4), 1–27.
59. Cerreta, M.K.; Berglund, K.A. The Structure of Aqueous—Solutions of Some Dihydrogen Ortho—Phosphates by Laser Raman-Spectroscopy. *J. Cryst. Growth* **1987**, *84*(4), 577–588.
60. Yedur, S.K.; Berglund, K.A. Use of Fluorescence Spectroscopy in Concentration and Supersaturation Measurements in Citric Acid Solutions. *Appl. Spectrosc.* **1996**, *50*(7), 866–870.
61. Rasimas, J.P.; Berglund, K.A.; Blanchard, G.J. Measuring Self Assembly in Solution: Incorporation and Dynamics of a “Tailor-Made Impurity” In Precrystalline Glucose Aggregates. *J. Phys. Chem.* **1996**, *100*(42), 17034–17040.
62. Rasimas, J.P.; Berglund, K.A.; Blanchard, J.G. A Molecular Lock and Key Approach to Detecting Solution Phase Self Assembly. A Fluorescence and Absorption Study of Carminic Acid in Aqueous Glucose Solutions. *J. Phys. Chem.* **1996**, *100*(17), 7220–7229.
63. Dunuwila, D.D.; Berglund, K.A. ATR FTIR Spectroscopy for In Situ Measurement of Supersaturation. *J. Cryst. Growth* **1997**, *179*(1–2), 185–193.
64. Blagden, N.; Davey, R.J.; Lieberman, H.F.; Williams, L.; Payne, R.; Roberts, R.; Rowe, R.; Docherty, R. Crystal Chemistry and Solvent Effects in Polymorphic Systems—Sulfathiazole. *J. Chem. Soc., Faraday Trans.* **1998**, *94*(8), 1035–1044.
65. Weissbuch, I.; Popovitzbiro, R.; Lahav, M.; Leiserowitz, L. Understanding and Control of Nucleation, Growth, Habit, Dissolution and Structure of 2-Dimensional and 3-Dimensional Crystals Using Tailor-Made Auxiliaries. *Acta Crystallogr., Sect. B: Struct. Sci.* **1995**, *51*, 115–148.
66. Leiserowitz, L.; Addadi, L.; Lahav, M. Macroscopic Phenomena in Crystals and Molecular Shape. *J. Mol. Graphics* **1989**, *7*(2), 95–5.
67. Weissbuch, I.; Addadi, L.; Lahav, M.; Leiserowitz, L. Molecular Recognition At Crystal Interfaces. *Science* **1991**, *253*(5020), 637–645.
68. Gallagher, H.G.; Roberts, K.J.; Sherwood, J.N.; Smith, L.A. A Theoretical Examination of the Molecular Packing, Intermolecular Bonding and Crystal Morphology of 2,4,6-Trinitrotoluene in Relation to Polymorphic Structural Stability. *J. Mater. Chem.* **1997**, *7*(2), 229–235.
69. Clydesdale, G.; Roberts, K.J.; Telfer, G.B.; Grant, W.D.J. Modeling the Crystal Morphology of Alpha-Lactose Monohydrate. *J. Pharm. Sci.* **1997**, *86*(1), 135–141.
70. Hendriksen, B.A.; Grant, D.J.W.; Meenan, P.; Green, A.D. Crystallisation of Paracetamol (Acetaminophen) in the Presence of Structurally Related Substances. *J. Cryst. Growth* **1998**, *183*(4), 629–640.
71. Shekunov, B.Y.; Lahtam, R. Growth Anisotropy Of N-Methylurea Crystals in Methanol. *J. Phys. Chem. A* **1996**, *100*, 5464–5469.
72. Shekunov, B.Y.; Grant, D.J.W. In Situ Optical Interferometric Studies of the Growth and Dissolution Behavior of Paracetamol (Acetaminophen) .1. Growth Kinetics. *J. Phys. Chem. B* **1997**, *101*(20), 3973–3979.
73. Zipp, G.L.; Rodríguez-Hornedo, N. Growth Mechanism and Morphology of Phenytoin and their Relationship with Crystallographic Structure. *J. Phys. D: Appl. Phys.* **1993**, *26*(8B), B48–B55.

74. Khoshkhoo, S.; Anwar, J. Study of the Effect of Solvent on the Morphology of Crystals Using Molecular Simulation: Application to Alpha-Resorcinol And *n*-n-Octyl-D-Gluconamide. *J. Chem. Soc., Faraday Trans.* **1996**, 92(6), 1023–25.
75. Gavezzotti, A. Computer Simulations of Organic Solids and Their Liquid-State Precursors. *Faraday Discussions* **1997**, 106, 63–77.
76. Gavezzotti, A.; Filippini, G. Polymorphic Forms of Organic Crystals at Room Conditions —Thermodynamic and Structural Implications. *J. Am. Chem. Soc.* **1995**, 117(49), 12299–12305.
77. Gavezzotti, A.; Filippini, G. Computer Prediction of Organic Crystal Structures Using Partial X-ray Diffraction Data. *J. Am. Chem. Soc.* **1996**, 118(30), 7153–7157.
78. Gavezzotti, A.; Filippini, G. Self-organization of Small Organic Molecules in Liquids, Solutions and Crystals: Static And Dynamic Calculations. *Chem. Commun.* **1998**, 3, 287–294.
79. Desiraju, G.R. Supramolecular Synthons in Crystal Engineering—A New Organic—Synthesis. *Angew. Chem., Int. Ed. Engl.* **1995**, 34(21), 2311–2327.
80. Desiraju, G.R. Crystal Gazing: Structure Prediction and Polymorphism. *Science* **1997**, 278(5337), 404–405.
81. Desiraju, G.R. Designer Crystals: Intermolecular Interactions, Network Structures and Supramolecular Synthons. *Chemical Communications* **1997**, 16, 1475–1482.
82. Boistelle, R.; López-Valero, I. Growth Units and Nucleation: The Case of Calcium Phosphates. *J. Cryst. Growth* **1990**, 102, 609–617.
83. Gidalevitz, D.; Feidenhansl, R.; Matlis, S.; Smilgies, D.M.; Christensen, M.J.; Leiserowitz, L. Monitoring in Situ Growth and Dissolution of Molecular Crystals: Towards Determination of the Growth Units. *Angew. Chem., Int. Ed. Engl.* **1997**, 36(9), 955–959.
84. Weissbuch, I.; Leiserowitz, L.; Lahav, M. Tailor –Made and Charge–Transfer Auxiliaries for the Control of the Crystal Polymorphism of Glycine. *Adv. Mater.* **1994**, 6(12), 952–956.
85. Weissbuch, I.; Zbaida, D.; Addadi, L.; Leiserowitz, L.; Lahav, M. Design of Polymeric Inhibitors for the Control of Crystal Polymorphism—Induced Enantiomeric Resolution of Racemic Histidine by Crystallization At 25 °C. *J. Am. Chem. Soc.* **1987**, 109(6), 1869–1871.
86. Khoshkhoo, S.; Anwar, J. Crystallization of Polymorphs—The Effect of Solvent. *J. Phys. D: Appl. Phys.* **1993**, 26(8B), B90–B93.
87. Li, L.; Rodríguez-Hornedo, N. Growth Kinetics and Mechanisms of Glycine Crystals. *J. Cryst. Growth* **1992**, 121, 33–38.
88. Zipp, G.L.; Rodríguez-Hornedo, N. The Mechanism of Phenytoin Crystal-Growth. *Int. J. Pharm.* **1993**, 98(1–3), 189–201.
89. Lechuga-Ballesteros, D.; Rodríguez-Hornedo, N. The Influence of Additives on the Growth Kinetics and Mechanism of L-Alanine Crystals. *Int. J. Pharm.* **1995**, 115(2), 139–149.
90. Lechuga-Ballesteros, D.; Rodríguez-Hornedo, N. Effects of Molecular Structure and Growth Kinetics on the Morphology of L-Alanine Crystals. *Int. J. Pharm.* **1995**, 115, 151–160.
91. Yip, C.M.; Ward, M.D. Atomic Force Microscopy of Insulin Single Crystals: Direct Visualization of Molecules and Crystal Growth. *Biophys. J.* **1996**, 71(2), 1071–1078.
92. Malkin, A.J.; Kuznetsov, Y.G.; Glantz, W.; McPherson, A. Atomic Force Microscopy Studies of Surface Morphology and Growth Kinetics in Thaumatin Crystallization. *J. Phys. Chem.* **1996**, 100(28), 11736–11743.
93. Yip, C.M.; Brader, M.L.; Defelippis, M.R.; Ward, M.D. Atomic Force Microscopy of Crystalline Insulins: The Influence of Sequence Variation on Crystallization and Interfacial Structure. *Biophys. J.* **1998**, 74(5), 2199–2209.
94. Kuznetsov, Y.G.; Malkin, A.J.; McPherson, A. Atomic Force Microscopy Studies of Phase Separations in Macromolecular Systems. *Phys. Rev. B: Condens. Matter* **1998**, 58(10), 6097–6103.
95. Volmer, M. Crystal Growth. *Z. Physik.* **1922**, 9, 193.
96. Stranski, I.N. Zur Theorie Der Kristallwachstums. *Z. Phys. Chem.* **1928**, 136, 259–278.
97. Burton, W.K.; Cabrera, N.; Frank, F.C. The Growth of Crystals and the Equilibrium Structure of Their Surfaces. *Philosophical Transactions* **1951**, A243, 299–358.
98. Chernov, A.A. Formation of Crystals in Solution. *Contemporary Physics* **1989**, 30, 251–276.
99. Bennema, P. Spiral Growth and Surface Roughening: Development Since Burton, Cabrera and Frank. *J. Cryst. Growth* **1984**, 69, 182–197.
100. Zipp, G.L.; Rodríguez-Hornedo, N. Phenytoin Crystal Growth Rates in the Presence of Phosphate and Chloride Ions. *J. Cryst. Growth* **1992**, 123, 247–254.
101. Lechuga-Ballesteros, D.; Rodríguez-Hornedo, N. The Relation Between Absorption of Additives and Crystal Growth Rate of L-Alanine. *J. Colloid Interface Sci.* **1993**, 157, 147–153.
102. Garside, J.; Janssen-Van Rosmalen, R.; Bennema, P. Verification of Crystal Growth Rate Equations. *J. Cryst. Growth* **1995**, 29, 353–366.
103. Durbin, S.D.; Feher, G. Studies of Crystal Growth Mechanisms of Proteins by Electron Microscopy. *J. Mol. Biol.* **1990**, 212(4), 763–774.
104. Sunagawa, I. *Morphology of Crystals*; Terra Scientific Publishing Company: Tokyo, Japan, 1987.
105. Himes, V.L.; Mighell, A.D.; De Camp, W.H. Structure of Carbamazepine: 5H-Dibenz[b,f]azepine-5-carboxamide. *Acta Crystallogr.* **1981**, B37, 2242–2245.
106. Behme, R.J.; Brooke, D. Heat of Fusion Measurement of a Low Melting Polymorph of Carbamazepine That Undergoes Multiple-phase Changes During Differential Scanning Calorimetry Analysis. *J. Pharm. Sci.* **1991**, 80(10), 986–990.
107. McMahon, L.E.; Timmins, P.; Williams, A.; York, P. Characterization of Dihydrates Prepared from Carbamazepine Polymorphs. *J. Pharm. Sci.* **1996**, 85(10), 1064–1069.
108. Krahn, F.U.; Mielck, J.B. Relations Between Several Polymorphic Forms and the Dihydrate of Carbamazepine. *Pharm. Acta Helv.* **1987**, 62, 247–254.
109. Roberts, R.J.; Rowe, R.C. Influence of Polymorphism on the Young's Modulus and Yield Stress of Carbamazepine, Sulfathiazole, and Sulfanilamide. *Int. J. Pharm.* **1996**, 129, 79–94.
110. Kaneniwa, N.; Ichikawa, J-I.; Yamaguchi, T.; Hayashi, K.; Watari, N.; Sumi, M. Dissolution Behavior of

- Carbamazepine Polymorphs. *Yakugaku Zasshi* **1987**, *107*, 808–813.
111. Matsuda, Y.; Akazawa, R.; Teraoka, R.; Otsuka, M. Pharmaceutical Evaluation of Carbamazepine Modifications: Comparative Study for Photostability of Carbamazepine Polymorphs by Using Fourier-transformed Reflection-absorption Infrared Spectroscopy and Colorimetric Measurement. *J. Pharm. Pharmacol.* **1994**, *46*, 162–167.
 112. Kobayasji, Y.; Ito, S.; Itai, S.; Yamamoto, K. Physicochemical Properties and Bioavailability of Carbamazepine Polymorphs and Dihydrate. *Int. J. Pharm.* **2000**, *193*, 137–146.
 113. Lefebvre, C.; Guyot-Herman, A.M.; Draguet-Brughmans, M.; Bouche, R.; Guyot, J.C. Polymorphic Transitions of Carbamazepine During Grinding and Compression. *Drug Dev. Ind. Pharm.* **1986**, *12*, 1913–1927.
 114. Lowes, M.M.; Cairn, M.R.; Lotter, A.P.; Van Der Watt, J.G. Physicochemical Properties and X-ray Structural Studies of the Trigonal Polymorph of Carbamazepine. *J. Pharm. Sci.* **1987**, *76*(9), 744–752.
 115. Ceoline, R.; Toscani, S.; Gardette, M.; Agafonov, V.N.; Dzyabchenko, A.V.; Bachet, B. X-ray Characterization of the Triclinic Polymorph of Carbamazepine. *J. Pharm. Sci.* **1997**, *86*(9), 1062–1065.
 116. York, P. Solid-state Properties of Powders in the Formulation and Processing of Solid Dosage Forms. *Int. J. Pharm.* **1983**, *14*, 1–28.
 117. Williams, R.O.; Brown, J.; Liu, J. Influence of Micronization Method on the Performance of a Suspension Triamcinolone Acetonide Pressurized Metered-Dose Inhaler Formulation. *Pharm. Dev. Tech.* **1999**, *4*(2), 167–179.
 118. Chew, N.; Chan, H. Influence of Particle Size, Air, Flow, and Inhaler Device on the Dispersion of Mannitol Powders As Aerosols. *Pharm. Res.* **1999**, *16*(7), 1098–1103.
 119. Vernaet, C.; Byron, P.R. Drug-Surfactant-Propellant Interactions in HFA-Formulations. *Int. J. Pharm.* **1999**, *186*(1), 13–30.
 120. Sinclair, B.D.; Rodríguez-Hornedo, N. A Molecular Simulation Approach to Interpret the Effects of Intermolecular Interactions Between Additives and Crystal Surfaces on the Morphology of Carbamazepine Dihydrate, *In Preparation*.
 121. Lechuga-Ballesteros, D.; Rodríguez-Hornedo, N. Growth and Morphology of D-Alanine Crystals: Influence of Additive Adsorption. *Pharm. Res.* **1993**, *10*, 1008–1014.
 122. Berkovitch-Yellin, Z. Toward An Ab Initio Derivation of Crystal Morphology. *J. Am. Chem. Soc.* **1987**, *107*, 8239–8253.
 123. Bravais, A. *Etudes Crystallographiques*; Gauthier-Villars: Paris, 1907.
 124. Friedel, G. *Bull. Soc. Fr. Mineral* **1907**, *30*, 326–455.
 125. Donnay, J.D.H.; Harker, D. A New Law of Crystal Morphology Extending the Law of Bravais. *Am. Mineral.* **1937**, *22*, 446–467.
 126. Docherty, R.; Roberts, K.J. Modelling the Morphology of Molecular Crystals; Application to anthracene, biphenyl, and β -succinic acid. *J. Cryst. Growth* **1988**, *88*, 159–168.
 127. Clydesdale, G.; Docherty, R.; Roberts, K.J. HABIT — A Program for Predicting the Morphology of Molecular Crystals. *Comput. Phys. Commun.* **1991**, *64*, 311–328.
 128. Hartman, P.; Perdok, W.G. On the Relations Between Structure and Morphology of Crystals. *Acta Crystallographica* **1955**, *8*, 49–52.
 129. Hartman, P.; Bennema, P. The Attachment Energy As a Habit Controlling Factor. *J. Cryst. Growth* **1980**, *49*, 145–149.
 130. Lifson, S.; Hagler, A.T.; Dauber, P. Consistent Force Field Studies of Intermolecular Forces in Hydrogen-bonded Crystals. *J. Am. Chem. Soc.* **1979**, *101*(18), 5111–5121.
 131. Clydesdale, G.; Roberts, K.J.; Docherty, R. HABIT95 — A Program for Predicting the Morphology of Molecular Crystals As a Function of the Growth Environment. *J. Cryst. Growth* **1996**, *166*, 78–83.
 132. Black, S.N.; Davey, R.J.; Halcrow, M. The Kinetics of Crystal-Growth in the Presence of Tailor-Made Additives. *J. Cryst. Growth* **1986**, *79*(1–3), 765–774.
 133. Davey, R.J.; Black, S.N.; Goodwin, A.D.; Mackerron, D.; Maginn, S.J.; Miller, E.J. Crystallisation in Polymer Films: Control of Morphology and Kinetics of an Organic Dye in a Polysilicone Matrix. *J. Mater. Chem.* **1997**, *7*(2), 237–241.
 134. Chen, B.D.; Garside, J.; Davey, R.J.; Maginn, S.J.; Matsuoka, M. Growth Of *M*-Chloronitrobenzene Crystals in the Presence of Tailor-Made Additives—Assignment of the Polar Axes from Morphological Calculations. *J. Phys. Chem.* **1994**, *98*(12), 3215–3221.
 135. Etter, M.C. Hydrogen — Bonds As Design Elements in Organic Chemistry. *J. Phys. Chem.* **1991**, *95*(12), 4601–4610.
 136. Rodríguez-Hornedo, N.; Murphy, D. Significance of Controlling Crystallization Mechanisms and Kinetics in Pharmaceutical Systems. *J. Pharm. Sci.* **1999**, *88*(7), 651–660.
 137. Morris, K.; Rodríguez-Hornedo; Hydrates, N. *Encyclopedia of Pharmaceutical Technology*; Boylan, J.C., Swarbrick, J., Eds.; 1993; 393–440.
 138. Nass, K.K. Process Implications of Polymorphism in Organic Compounds. *AIChE* **1991**, *284*, 72–81.

FURTHER READING

- Brittain, H.G. *Polymorphism in Pharmaceutical Solids*; Marcel Dekker, Inc.: New York, 1999.
- Byrn, S.R.; Pfeiffer, R.R.; Stowell, J.G. *Solid-State Chemistry of Drugs*; SSCI, Inc.: West Lafayette, IN, 1999.
- Chernov, A.A. *Modern Crystallography III: Crystal Growth*; Springer-Verlag: Berlin, 1980.
- Cullity, B.D. *Elements of X-ray Diffraction*; Addison-Wesley Publishing Company, Inc.: Reading Massachusetts, 1978.
- Gavezzotti, A. *Theoretical Aspects and Computer Modeling*; John Wiley & Sons Ltd.: New York, 1997.
- Mersmann, A. *Crystallization Technology Handbook*; Marcel Dekker, Inc.: New York, 1995.
- Mullin, J.W. *Crystallization*; Butterworth-Heinemann Ltd.: Oxford, 1993.
- Myerson, A.S. *Handbook of Industrial Crystallization*; Butterworth-Heinemann Ltd.: Oxford, 1993.
- Söhnel, O.; Garside, J. *Precipitation: Basic Principles and Industrial Applications*; Butterworth-Heinemann Ltd.: Oxford, 1992.
- Zettlemoyer, A.C. *Nucleation*; Marcel Dekker, Inc.: New York, 1969.

DOSAGE FORMS: NON-PARENTERAL

Shailesh K. Singh

Wyeth–Ayerst Research, Pearl River, New York, U.S.A.

Venkatesh Naini

Barr Laboratories, Inc., Pomona, New York, U.S.A.

INTRODUCTION

Dosage form is a drug delivery system designed to deliver the active ingredient to the body and, upon administration should deliver the drug at a rate and amount that assures the desired pharmacological effect. Such dosage forms are manufactured under current good manufacturing procedures (cGMP), using equipment and packaging to ensure product stability. The dosage form must produce the same therapeutic response each time it is administered. To maintain this reproducibility between and within batches, manufacturing procedures are validated under a specific quality assurance program. Non-parenteral dosage forms can be categorized based on the route of administration or physical form. Based on physical form they can be classified as solids, liquids (homogenous and heterogeneous systems), semisolids, and aerosols. Dosage forms can also be categorized based on the route of administration. Solid dosage forms include different types of compressed tablets, granules, troches, lozenges, coated dosage forms, and hard and soft gelatin capsules. Liquid dosage forms include solutions, suspensions, emulsions, and buccal and sublingual sprays. Topical dosage forms are applied to the skin and include ointments, pastes, creams, lotions, liniments, and transdermal patches. Some dosage forms are formulated for application to body cavities, viz. rectal and urethral suppositories and vaginal pessaries. Inhalation aerosols, using metered dose inhalers (MDIs), dry powder inhalers (DPIs) and nebulizers, are used to deliver drugs to the respiratory tract. Nasal route uses solution and suspension dosage forms. Occular route is used to administer solutions and suspensions to the eye for local and systemic effects.

SOLID DOSAGE FORMS

Powders and Granules

Powders are intimate mixtures of dry, finely divided drugs and/or chemicals that are intended for oral administration

or external use. Powders may also be formulated as larger particle sized, free-flowing granules to aid in handling and administration. Bulk powders usually are packed into a suitable wide-mouth container and contain relatively nontoxic medicaments in large doses, e.g., compounded magnesium trisilicate oral powder. Insufflations are medicated powders blown into ear, nose, or throat.

Tablets

Flowability and compressibility are two important parameters essential for successful manufacture of tablets. Flowability determines ease of material flow from tablet hopper to the press. Inadequate flow gives rise to arching, bridging, or rat-holing in hoppers. Powder flow can be improved mechanically by use of force feeders. Flowability can also be increased by incorporation of glidants like fumed silica and talc. Another method involves the conversion of powder to spherical particles by spray drying or spheronization. Tablets are manufactured by dry and wet methods. Dry methods consist of direct compression, slugging, and roller compaction of drug-excipient blends. Directly compressible excipients may be disintegrants with poor flow, e.g., microcrystalline cellulose (Avicel PH102); free-flowing materials which do not disintegrate, e.g., dibasic calcium phosphate (DiPac®); or free-flowing powders which disintegrate by dissolution (e.g., spray-dried lactose, anhydrous lactose, dextrose, sucrose, amylose, etc.). The drug is mixed with excipients in a blender and then compressed directly on a tablet press. Dry granulation by compression or slugging is used for moisture or heat sensitive actives. The powder blends are compressed into compacts or slugs. An alternative method is to squeeze the powder blends into solid cake between rollers called roller compacts. These slugs or compacts are milled and screened in order to produce granules with improved flow. Granulation is the process of particle size enlargement of homogeneously mixed powder ingredients and simultaneously increasing bulk density, flowability, and compressibility of the system. Wet granulation process involves the massing of the powder mix, using a binder and

Table 1 List of commonly used tablet diluent/fillers

Diluent	Properties
Dextrose	Hygroscopic and soft granules; available as anhydrous and monohydrate; anhydrous has poor compression
Dicalcium phosphate	Inexpensive, insoluble in water; commercially Ditas [®] is unmilled while Encompress [®] is of specific particle size (better flow)
Alpha lactose monohydrate	Inexpensive, relatively inert; most widely used; often used with Avicel [®] PH MCC to improve disintegration
Mannitol	Freely soluble, used particularly for chewable tablet; powder form has poor flow and compaction; granular form has good flow
Microcrystalline cellulose	Excellent compressibility, some disintegration properties; available in different grades with specific applications
Sodium chloride	Freely soluble, used for solution tablets
Sucrose	Sweet taste but hygroscopic, may be diluted with lactose

solvent. The solvent should be volatile, nontoxic, and removed by drying. This process is not suitable for hydrolysable and thermolabile drugs. The binder is added in the form of a solution, or added dry or its mucilage incorporated with the powder blend. The choice of liquid depends on the properties of the material being granulated. Water is widely used alone or along with a binding agent. Commonly used nonaqueous liquids are isopropanol and ethanol. Massing process is usually performed in a low or high shear granulator where the liquid is poured or sprayed onto a moving powder bed until a moist mass of finely divided material is formed. This is passed through an oscillating granulator with the appropriate screen size to obtain the required granule particle size. Sometimes both intra- and extra-granular portions are divided to prevent incompatibility between excipient material hiding or better distribution, tableting or dissolution. Tablets require different functional excipients for their manufacture. Diluents are inert bulking agents added to actives to make a reasonably sized tablet. Generally, a tablet should weigh about 50–60 mg and therefore very low dose drugs will require these diluents to make at least a 50 mg tablet. Table 1 lists some commonly used excipients. Adsorbents such as fumed silica and kaolin are sometimes used for holding large amounts of fluids in an apparently dry state. Binders are used as adhesives to bind powder in wet granulation and give strength to compacts during compression. Binders may be incorporated into the dry blend or added as a solution to the mixed powder during wet granulation. Table 2 lists some of the binders commonly used in tableting. Disintegrants are usually added to promote rapid breakup of tablets to increase surface area and aid drug dissolution. Disintegrants can act by different mechanisms such as like swelling and capillary action. Table 3 lists some commonly used disintegrants.

Glidants are materials that are added to tablet formulations to improve flow properties of the granulation. They act by reducing inter-particulate friction (e.g., fumed silica). Lubricants are added to prevent the adherence of granules to the punch and die faces of the tablet press. Many lubricants also facilitate flow of granules. Talc and magnesium stearate are more effective as punch lubricants. Stearic acid works better as a die lubricant. Table 4 lists some commonly used lubricants and glidants.

Specific Types of Tablets

Lozenges: These are compressed tablets formulated, without a disintegrant and must be allowed to dissolve in the mouth. They are used for local activity (throat lozenges) or for systemic effect (vitamins).

Effervescent tablets: These tablets undergo quick dissolution of actives in water due to internal liberation of carbon dioxide. By combining alkali metal carbonates or bicarbonates with tartaric or citric acid, carbon dioxide is liberated when placed in water. They are prepared by the heat fusion technique. Usually a water-soluble lubricant is used to prevent scum formation at the water surface. Sweetness is achieved by the addition of saccharin, since sucrose is hygroscopic and increases the bulk of the tablet, e.g., Rochelle Salt[®].

Chewable tablets: These tablets are preferred for pediatric and geriatric patients who have difficulty swallowing whole tablets. Another advantage is that they do not need water for administration. Mannitol is normally used as the base diluent because of its pleasant taste and texture, and because it can effectively mask the taste of objectionable actives. They are usually prepared by wet granulation and are not compressed very hard. High amounts of flavor are added to increase

Table 2 List of commonly used binders in tableting

Binders	Concentration (wt%)	Properties
Natural gums (acacia, tragacanth)	1–5	Form very hard granules; variability in quality
Cellulose derivatives	2–5	HPMC is the most common; used as wet binders
Gelatin (replaced by synthetic polymers)	5–10	Strong adhesive, hence used in lozenges; gels when cold; not very popular in tropical climates
Glucose	Up to 50	Strong adhesive; but hygroscopic
Polyvinyl pyrrolidone (PVP)	2–20	Soluble in water and some organic solvents; (may vary with molecular weight grades)
Starch mucilage	5–10	Commonly used adhesive; insoluble in dry state
Pre-gelatinized starch	10–25	A better alternative to starch paste
Sodium alginate	0.5–3	Forms hard granules; prolongs disintegration time
Sucrose NF	Up to 70	Hygroscopic; tablet hardens on storage

palatability. Antacids are typically formulated as chewable tablets.

Sublingual and buccal tablets: These tablets are placed under the tongue (sublingual) or the cheek (buccal) and can produce immediate systemic effects by enabling the drug to be directly absorbed through the mucosa by preventing the first pass effect (e.g., isoprenaline sulphate and glyceryl trinitrate). Tablets are small, flat, without a disintegrant, and are compressed lightly to produce soft tablets.

Molded tablets: These are prepared from mixtures of medicinal substances and a diluent usually consisting of lactose and powdered sucrose in varying quantities. The powders are dampened with solutions containing high proportions of alcohol depending on the solubility of the active and filler. The dampened powders are pressed under low pressure in die cavities. Solidification depends upon crystal bridging during the subsequent drying process, and not upon the compaction forces.

Multi-layered tablets: A multilayered tablet consists of several different granulations compressed on top of each other to form a single tablet. They may also be bi-layer when incompatible drug substances are used, e.g., phenylephedrine HCl in one layer and ascorbic acid and paracetamol in another.

Modified Release Dosage Forms

Modified release (MR) has been used to describe dosage forms having drug release characteristics based on time, course, and/or location and are designed to accomplish therapeutic or convenience objectives not offered by conventional or immediate release dosage forms. Drugs for chronic conditions with short half-lives, possessing a good therapeutic index and uniform absorption pattern are ideal candidates for such dosage forms. These are either delayed release or extended release (ER) preparations. ER dosage

Table 3 List of commonly used tablet disintegrants

Disintegrants	Concentration (wt%)	Commercial name; property
Alginic acid and alginates	2–10	Created in situ in effervescent tablets
Carbondioxide		Amberlite [®]
Ion exchange resins		Veegum [®] ; often slightly colored
Magnesium aluminium silicate	Up to 10	Avicel; lubricant properties; directly compressible
Microcrystalline cellulose	Up to 20	Starch 1500 [®] (pre-gelatinized starch)
Starch NF	5–20	Primarily a wetting agent but this aids disintegration
Sodium dodecyl sulphate	0.5–5	Nymcel [®]
Sodium carboxymethyl cellulose	1–2	Primojel [®] ; Explotab [®]
Sodium starch glycolate	2–8 (dry)	Corn, potato, maize are most frequently used
Starch	5–10	Ac-Di-Sol [®] ; It may be used both at inter- and intra-granular portions (1–2%)
Croscarmellose sodium	2–4	Polyplasdone XL [®] , Crospovidone NF
Cross linked PVP	2–5	

Table 4 List of commonly used lubricants and glidants in tableting

Excipient	Type	Concentration (wt%)	Property
Calcium and magnesium stearate	Lubricant	0.25–2	Reduces tablet strength, prolongs disintegration, insoluble in water, excellent lubricant
Stearic acid	Lubricant	1–4	More preferred as a die wall lubricant
Polyethylene glycol	Lubricant	2–5	M.W. 4000–6000, soluble in water, moderately effective
Liquid paraffin	Lubricant	Up to 5	Dispersion problems
Sodium lauryl sulphate (SLS)	Lubricant	0.5–5	Moderate lubricant with wetting properties, used with stearates
Sodium stearyl fumarate	Lubricant	0.5–2	Less sensitive to overblending and compressibility
Magnesium lauryl sulphate	Lubricant	1–2	Water soluble
Talc	Lubricant	1–2	Insoluble but not hydrophobic
	Glidant	0.2–0.3	
Colloidal silica (Aerosil [®] , Cabosil [®])	Glidant	0.1–0.2	Excellent glidant
Starch	Glidant	0.2–0.3	Primarily disintegrant
Microcrystalline cellulose	Glidant	0.2–0.5	Primarily used as diluents/filler

forms allow at least a twofold reduction in dosing frequency as compared to the conventional dosage form. Delayed release dosage forms are designed to release all or a portion of drug at times much later than the time of administration. The delay may be time based or environment specific, as in enteric-coated dosage forms. Some other MR dosage forms include repeat action and targeted release dosage forms. Most controlled release products are good examples of ER dosage forms. These dosage forms can be classified by their mechanism of release and/or type of formulation. Coated beads, granules microspheres, and other particulate systems are pellet type controlled release dosage forms, where the drug is usually coated onto nonpareil beads (low dose) or made from granules composed of the drug (high dose). These pellets are further coated with functional coating agents (Eudragits[®], HPMCs, Surelease[®], etc.) to provide various release characteristics. These pellets can be used to fill capsules (e.g., Ornade Spansules[®]) or compressed at low pressure into tablets (e.g., Theo-Dur[®]). In some cases small mini-tablets of about 3–4 mm diameter can be compressed. These tablets function like pellets and can be used to fill capsules. Microencapsulation is a process of encapsulating microscopic drug particles with a thin wall of coating material. Several coating materials have been used including gelatin, ethylcellulose, and polyvinyl alcohol, e.g., Micro-K-Extencaps[®] (Wyeth–Ayerst Research). Matrix systems are dosage forms where drug substance is combined with hydrophilic cellulose polymers (excipient material), which slowly erode in the presence of

body fluids. On hydration, the polymers behave like a gel and prevent the fast disintegration of the tablet. Diffusion from the gel controls the drug release (e.g., Oramorph SR[®] tablets). A multi-layered tablet consists of several different granulations compressed on top of each other to form a single tablet composed of two or more layers. Each layer is fed from a separate feed frame with individual weight control. Precompression tamping helps in good binding of layers. Also, reduced pressures prevent intermixing of granules during compression. They may be bilayer where IR/ER combination are used or mainly when incompatible drug substances is used. Sometimes the release from individual layers is controlled to give a drug delivery system, such as, Geomatrix[®] system (1). In some cases if the bulk density of the tablet is less than one, it floats in the gastric fluids thus extending the residence time in the gastrointestinal tract (GIT). Such dosage forms are called Hydrodynamically Balanced Systems (HBS), an example being Valrelease[®] (Roche). In some cases the drug is embedded inside inert polymeric matrices with materials such as polyethylene, polyvinyl acetate, and polymethacrylates. The granulations are then compressed into tablets. These inert matrices are excreted in the feces unchanged (e.g., Ferro-Gradumet[®] (Abbott)). Some drugs form complexes resulting in slower dissolution and behave as extended release dosage forms, e.g., Rynatan[®] (Carter-Wallace). A slowly eroding tablet may be granulated with hydrophobic excipients (waxy lipophilic material) so that the drug leaches out over an extended period with an outer shell containing the IR dose (buffered aspirin). In some

cases a cationic or anionic drug solution can react with an insoluble resin to form a complex. This complex can be tableted, encapsulated or suspended in a vehicle, e.g., Tussionex Pennkinetic[®] Extended Release Suspension (Medeva). Osmotic pump drug delivery systems consist of a core tablet coated with a semi-permeable membrane with a fine orifice made by laser beams. The core usually forms two layers containing the active and osmotic agents. In the gastric fluids water is imbibed by the osmotic agent (pull) and then exerts pressure on the drug (push) in solution, out through the orifice. Such dosage forms are independent of pH of the gastric fluids and are termed as gastrointestinal therapeutic systems (GITS), e.g., Procardia XL[®] (Pfizer). Repeat action tablets consist of slow release inner core and a immediate release (IR) as in Repetabs[®] (Schering) or as bilayer IR/ER tablets. Delayed release dosage forms are used for drugs that are destroyed in the gastric fluids, or cause gastric irritation, or are absorbed preferentially in the intestine. Such dosage forms are enterically coated using materials such as cellulose acetate phthalate, shellac, and waxes. The coating allows the drug to release at higher pH (pH dependent) or by enzyme catalyzed reactions, e.g., Erythromycin or Aspirin delayed release dosage forms. Several forms of oral controlled release systems are available in the market; however, most of them are dependent on the rate at which the system passes along the gastro-intestinal tract. This can be overcome by regulating the release of the drug by physical chemical means, or by a process related to the environment in which the delivery system is present at the specific time (2).

Capsules

The word “capsule” is derived from the Latin word *capsula* meaning a small box. Gelatin, a substance of natural origin with unique properties, is the major component of capsules. Gelatin is used because it is nontoxic and readily soluble in biological fluids at body temperature. It has good film forming properties and, as a in water and water–glycerol systems, undergoes reversible phase change from a solution to gel at only a few degrees above ambient temperature. There are two forms of gelatin (A and B) based on the method of manufacture from animal bone and skin. The properties important for capsule shell manufacture are viscosity and bloom strength. Hard gelatin capsules are firm and rigid while soft gelatin capsules are soft and flexible. This is because soft capsules contain a larger proportion of plasticizers like glycerol, sorbitol, propylene glycol, acacia, and sucrose. Varying proportions of plasticizers are added depending on the intended use of soft gelatin capsules. The colorants used consist of soluble and insoluble dyes. Titanium

dioxide and iron oxide pigments are common, although recently aluminum lakes are being used. Preservatives are added to capsules to prevent microbial contamination. Moisture levels are also maintained at low levels to prevent bacterial growth on storage.

Hard gelatin capsules (HGCs)

Hard gelatin capsules are available in sizes ranging from size 000, (the largest) to size 5 (the smallest). The fill weight of capsule and tapped bulk density of the powder blend determines the selection of capsule size. Recently, better techniques for capsule sealing, like, self-locking and have made it possible for a range of materials. The filler material should not react with gelatin. Aldehydes lead to gelatin cross-linking affecting the integrity of the shell and water in the formula can act as a plasticizer. On the other hand, hygroscopic agents can make the capsule shell brittle. Powders filled into hard gelatin capsules should have good flow properties to maintain uniform fill weights during filling operations. Granules and pellets of spherical shape making them free-flowing and nonfriable are good candidates for capsule filling using gravitational systems or specialized dosing chambers to maintain uniform fill weight. In some instances, minitabets (filmcoated, nonfriable) can be filled into capsules to produce specialized dosage forms or to separate incompatible ingredients. A recent innovation in hard gelatin capsule filling is a revival of the old practice of filling liquids or semisolids. The main problem encountered is product leakage. This difficulty was overcome by using self-locking capsules and formulation techniques. The use of mixtures of material which are either thermosoftening or thixotropic in nature has become prevalent. These materials are liquefied by heat or shearing force, and revert to solid state within the capsule shell after filling. Filling machines have been developed to handle such formulations with existing powder filling equipment. This system works with solid, liquid, semisolids, and potent drugs. The application of semisolids filling is also getting prevalent. A more recent innovation in HGCs to fill liquid dosage forms with the use of new machines which heat-seal the caps permanently to prevent leakage as observed with liquid filled HGCs in the past. A good example is the introduction of Licaps[®] (Capsugel) for liquid fills. Capsules made of nongelatin ingredients for materials not compatible with gelatin are also available, e.g., cellulose (Vegecaps[®]).

Soft gelatin capsules (SGCs)

SGCs or softgels are continuous gelatin shells surrounding a liquid or semi-solid fill. These capsules are formed, filled, and sealed, all in one operation. These capsules are

available in different shapes and sizes. SGCs are preferred for drugs with poor compressibility, poor powder flow, mixing problem, unstable or poor solubility in gastric pH, and bioavailability problems. Such drugs can be solubilized or dispersed in a liquid, where dosage uniformity is more accurate. Some drugs that are liquid or that melt during compression are good candidates, if other means of tableting are expensive. Gelatin used in SGCs has lower bloom strength than HGCs. The plasticizer type and concentration controls the mechanical strength of the shell. In general, plasticizer amounts are larger, making them more flexible than HGC shells. Preservatives, colorants, and opacifiers are used in the same manner as in HGCs. Sometimes softgel capsules are enteric coated for drugs which are absorbed in the small intestine. Once the capsule shells dissolve *in vivo*, the drug is available in a liquid or semi-solid form that dissolves or disperses into fine particles with enhanced bioavailability. SGCs can be filled with several materials such as aqueous solutions, nonaqueous solutions, suspensions, pastes, oily solutions of drug, self-emulsifying system, and water-miscible liquids. Materials that cause migration of water or plasticizer from the shell cannot be filled. Surfactants and systems with extreme pH should be avoided.

LIQUID DOSAGE FORMS

Solution

A solution is a homogenous single-phase system consisting of two or more components. Solutions are easier to swallow and are acceptable dosage forms for pediatric and geriatric use. The drug in solution is readily available for absorption and therapeutic response is faster. Solutions, however, are bulky and inconvenient to transport. The stability of actives is poorer than in solids and they provide suitable media for microbial growth. Aqueous solutions are preparations made with water as solvent. Purified Water USP is widely used for most preparations. Some drugs are unstable in water or sensitive to the presence of carbon dioxide or oxygen. Not all substances are completely soluble in water and may lead to precipitation. Several other techniques are used to increase solubility of drugs in solution. Cosolvency is a process of increasing solubility of a drug by using a combination of solvents. Some suitable cosolvents are ethanol, isopropyl alcohol, sorbitol, glycerol, and propylene glycol. If a drug is a weak acid or base, then its solubility in water is influenced by pH. The pH for optimum solubility may not give a stable product. Thus, a compromise must be reached to ensure proper formulation and bioavailability. Suitable

buffer systems may be used if necessary. Solubility of insoluble or poorly soluble drugs can also be increased by addition of surface-active agents. Most surfactants are miscible with solvent system and compatible with other ingredients. Hydrophilic surfactants with HLB values >15 are generally preferred. In some cases complexation of a drug with a material may result in formation of soluble molecular complex. However, such complexation needs to be reversible for the active to cross the biological barrier. Chemical modifications of the drug can also result in more water-soluble derivatives. However, these modified drugs are regarded as new chemical entities. Nonaqueous solutions are used when complete solution is not possible in water or if the drug is unstable. Ethyl alcohol is the most widely used water-miscible solvent for external preparations. Ethyl ether is occasionally used as a co-solvent, in combination with alcohol in the preparation of some colloids. Other solvents such as isopropyl myristate and isopropyl palmitate are solvents with low viscosity and are ideally used in cosmetics preparations. Xylene is present in ear drops for human use to dissolve ear wax.

Liquid formulation additives used include buffers, colorants, flavoring agents, and preservatives. Buffers are dissolved in solvents to resist pH changes. The choice of buffers depends on the pH and the buffering capacity. Most pharmaceutically acceptable buffer systems include carbonates, phosphates, citrates, gluconates, and lactates. Colors are added for attractiveness and product identification. Flavors are added to solutions to increase their palatability, particularly for drugs with unpleasant taste. This is especially useful in pediatric formulations. Flavors also help in product identification and are of natural or synthetic sources. Fruit juices, peppermint oil, and menthol are some examples of flavors. Some flavors are preferred for specific products, e.g., mint is associated with antacid formulas. Similarly, flavors are preferred by specific patient groups, e.g., children prefer fruity tastes and smell, while adults prefer flowery and acid flavors. Preservatives help prevent microbial growth. The choice of preservatives should be based on their performance from a microbial challenge test. Care should be taken to ensure there is no adsorption of preservatives onto product containers or packaging material. Antioxidants are added to prevent degradation of the drug in solution; the amount and type can be determined after careful determination of the degradation pathway and stability testing with different agents. Sucrose is widely used as sweetening agent, because it is water soluble, and stable at a wide range of pH. It has a pleasant texture and soothing effect on the throat. There are several other sweeteners that are less widely used. Artificial sweeteners like sugar alcohols and aspartame are used by diabetic patients.

Types of Liquid Preparations

Draught and elixirs: Draught is a mixture by which one or two large doses of about 50 ml are given. Traditionally, elixirs are solutions of potent or nauseating drugs containing alcohol as a cosolvent (60–70%).

Linctuses: A linctus is a viscous preparation usually prescribed for relief of cough. They usually consist of a simple solution of active in a high concentration of sucrose, often with other sweetening agents.

Mouthwashes and gargles: These are aqueous solutions for prevention and treatment of mouth and throat infections. They usually contain antiseptics, analgesics, and/or astringents. These solutions are used directly or diluted with warm water.

Nasal drops: These are small volume aqueous solutions. They are usually buffered to pH of 6.8 and are isotonic solutions. These drops are used locally as antibiotics, anti-inflammators, and decongestants.

Ear drops: These are simple solutions of drugs in water, glycerol, and propylene glycol for local use in the ear and include antibiotics, antiseptics, cleaning solutions, and wax softeners (xylene).

Enemas: These are available as solutions (aqueous or oily) as well as suspensions for rectal administration of drugs for cleaning, diagnostic, or therapeutic effect.

Lotions: These are available as solutions and suspensions to be applied topically without friction. They may either contain humectant, so that moisture is retained on the skin after application, or alcohol, which evaporates quickly imparting a cooling sensation to the skin.

Liniments: These are intended for massaging the skin. They may contain ingredients such as methyl salicylate or camphor as counter-irritants.

Colloidons: These are prepared from volatile solvents that evaporate quickly leaving a tough, flexible film on the skin that seals small cuts and or holds the active in intimate contact with the skin.

Intermediate solutions: Pharmaceutical solutions are used as intermediates for manufacturing other preparations. Aromatic water is used as a flavoring agent and peppermint and anise waters have some carminative properties. These are manufactured as concentrated waters and are diluted before use. Infusions are prepared by extracting the drug using 25% alcohol without heat. Extracts are similar to infusions, but are concentrated by evaporation. Tinctures are alcoholic or hydro-alcoholic solutions prepared from vegetable materials or from chemical substances. They are relatively weak compared to extracts. Spirits are alcoholic or hydro-alcoholic solutions of volatile substances prepared by simple solution or by admixture of ingredients. These are used

as flavoring agents and may have medicinal value. Syrups are concentrated solutions of sucrose or other sugars to which medicaments or flavoring agents are added. These are bacteriostatic by virtue of their osmotic effect, e.g., simple syrup, USP.

Suspensions

These are liquids consisting of insoluble solid particles dispersed throughout a liquid phase. Most suspensions are ready to use while some are prepared as solids to be reconstituted just before use. Ideally, suspension should be homogenous between the time of shaking and dispensing the required dose. The suspended particles should be small, uniformly sized to give a smooth elegant product free from grittiness. Some insoluble solids are not easily wetted by water and thus need wetting agents to be able to disperse readily throughout the medium. Some wetting agents include surfactants, hydrophilic colloids, and solvents. Surface active agents or surfactants possessing HLB value between 7 and 9 are suitable as wetting agents. Most surfactants are used at concentrations of 0.1%. For oral use Tweens and Spans are commonly used, while sodium lauryl sulphate (SLS) is used for external applications. Some wetting agents may cause foaming and formation of deflocculated systems. Hydrophilic colloids like acacia, bentonite, tragacanth, alginates, and cellulose derivatives function as a protective colloid by coating the surface of the particles and thus imparting hydrophilic character to the solid particles. Solvents such as alcohol, glycerol, and glycols are water-miscible and reduce the liquid/air interfacial tension, increasing wetting.

Flocculation and deflocculation

Flocculation comes from the Latin word *flocculate* meaning loose and woolly. Flocculated systems result in rapid rate of settling because each individual unit is composed of many particles and is therefore larger. However, due to the loose packing of flocs they are easily dispersible on shaking. Deflocculated systems on the other hand are made up of smaller particles whose settling rate is slower, but the settled particles tend to form an irreversible compact and are difficult to redisperse. This phenomenon is called *caking*. For coarse suspensions, a deflocculated suspension will have better uniformity of dose but poorer stability due to formation of cake. Thus, a stable suspension is obtained by preparing a partially flocculated suspension with controlled viscosity so that settling is minimal. Controlled flocculation is achieved by a combination of particle size control, electrolytes to control zeta-potential and by the addition of polymers. Inorganic electrolytes, added to an

aqueous suspension alter the zeta-potential of the dispersed particle. Lowering the zeta-potential sufficiently will result in flocculation. Some of the commonly used electrolytes include sodium salts of acetates, phosphates, and citrates. Use of ionic surfactants may also result in flocculation by neutralization of particle charges. Starch, alginates, tragacanth, and cellulose derivatives are sometimes added to control the degree of flocculation so that the suspension is in a flocculated state and the sedimentation volume is large. Suspensions should exhibit high viscosity at low shear rate and vice versa. Also, the viscosity should be low enough to be poured from the container but should spread evenly, if it is intended for external application. Suspensions for injection should be able to pass through hypodermic needles. Acacia gum is used as a thickening agent for extemporaneously prepared suspensions. Tragacanth forms viscous aqueous solutions, and its thixotropic and pseudoplastic properties make it a better thickening agent than acacia. Sodium alginate is used as a suspending agent but is incompatible with cationic materials. Several cellulose derivatives, such as methylcellulose (Celacol[®]), hydroxyethylcellulose (Natrasol[®] 250), sodium carboxymethylcellulose, and microcrystalline cellulose, disperse in water to produce viscous colloidal solutions and are suitable, as suspending agents. Montmorillonite clays or hydrated silicates like bentonite, veegum, and hectorite readily hydrate and absorb up to 12 times their weight of water. The gels formed are thixotropic and therefore have wonderful suspending properties. Carbopols[®] are synthetic polyacrylic acid copolymers that function as thickening agents at higher pH values.

Buffers are included in suspensions to maintain chemical stability and control tonicity. Density modifiers like sucrose and propylene glycol can be added to prevent large differences in densities that could result in sedimentation. Flavors, colors, and perfumes may be added to improve palatability and appearance of the product. Humectants like glycerol and propylene glycol are added in concentrations of 5% for external applications to prevent the product from drying out after application to the skin. Addition of preservatives is important, particularly when using naturally occurring adjuvants. In some situations sweeteners may be added but their effect on final product viscosity and degree of flocculation should be well understood. Suspensions are normally manufactured using colloidal mills with rotor–stator mechanism to ensure free flowing and evenly dispersed particles. Oral suspensions usually have flavoring agents intended for oral administration. Good examples are milk of magnesia, bentonite, magma, and jellies. All these systems swell and form a gel like consistency with non-Newtonian

characteristics. Topical suspensions like Calamine lotion are for external use only.

Emulsions

These are dispersions of one liquid (dispersed phase) in the form of uniformly divided droplets in another liquid (dispersion medium). Depending on which liquid is the dispersed phase oil-in-water (o/w) or water-in-oil (w/o) systems are obtained. To test the identity, emulsion miscibility, staining and conductivity tests are performed. The choice of emulsion depends on the route of administration and the end use. For oral administration, o/w emulsions are used while for external use, both o/w and w/o systems can be employed. Semisolid o/w emulsions are termed as “creams,” and are easily washable after application. Water-in-oil emulsions have an occlusive effect and are therefore preferred as moisturizing lotions and cleansing agents. The choice of oil depends on the application with some oils like castor and cod liver oil having a therapeutic value. Thus, w/o preparations are greasy, with high apparent viscosity while o/w emulsions are less greasy and readily absorbed and washable. Ideally emulsions should exhibit pseudoplasticity and thixotropy, that is, high viscosity at low shear rates and vice versa. They should be dispensable from containers, bottles, and tubes but at the same time should spread on the skin with light pressure. The rheological properties of emulsions are controlled by the concentration, particle size, and viscosity of dispersed phase, concentration of dispersion phase, and the nature and concentration of the emulsifier. The choice of emulsifying system depends on the route of administration, its HLB value and its toxicity. There is no approved list of emulsifiers but pharmaceutical companies employ emulsifiers approved for use in the food industry. Emulsifiers with surface activity reduce the interfacial tension between the phases, thereby decreasing the need for energy to disperse the internal phase. The surfactants used can be anionic like sodium, potassium, and ammonium salts of long chain fatty acids (e.g., sodium stearate), soaps of di- and tri-valent metal ions (e.g., calcium oleate and amine soaps, sulphated and sulphonated compounds (e.g., SLS). Cationic surfactant, amphoteric surfactants like lecithin, and non-ionic surfactants, like glycol and glycerol esters (e.g., glycerol monostearate), sorbitan esters, polysorbates, fatty alcohol polyglycol ethers (e.g., cetyl or cetostearyl alcohol), fatty acid polyglycol esters are also used. Sometimes naturally occurring materials and their derivatives, such as acacia, semi-synthetic polysaccharides (e.g., methylcellulose), sterol containing substances (e.g., beeswax), and wool fat

(anhydrous lanolin), can also be employed as emulsifiers. Other ingredients used in emulsions are finely divided solids, antioxidants, such as butylated hydroxy toluene (BHT) and butylated hydroxyanisole (BHA), humectants and preservatives like benzoic acid, parahydroxybenzoic acid esters, chlorocresol, and phenoxyethanol. Emulsions can be used orally (o/w) with the therapeutic agent included in the internal phase (as for taste masking bad tasting medicaments). Externally, they can be used as lotions either with therapeutic agents or without (as in cosmetics). Emulsions are made using a variety of equipment depending on the stability requirement and the kind of process used. Both batch and continuous processes can be used. Colloid mills have been used traditionally. However, high-pressure homogenizers, microfluidizers, and ultrasonic homogenizers are being used for manufacturing emulsions.

SEMISOLID DOSAGE FORMS

Ointments, creams, and pastes are semisolid dosage forms intended for topical application. They may be applied to the skin, used nasally, rectally, and vaginally. Most of them contain some form of medicament. Medicated ointments are semi-solid preparations intended for application to skin or mucous membranes. Nonmedicated ointments are used as protectants, lubricants, and emollients. Ointment bases used for ointment preparation are of four types, hydrocarbon bases, absorption bases, water-removable bases, and water-soluble bases. Hydrocarbon bases have emollient properties and are effective as occlusive dressings (e.g., Petrolatum, USP). Absorption bases permit the incorporation of aqueous solutions to form w/o emulsions (e.g., hydrophilic petrolatum and lanolin). Water-removable bases are also o/w emulsions and are water washable (e.g., hydrophilic ointment). Water-soluble bases have no oleaginous component and are referred to as *greaseless* water-washable bases (e.g., polyethylene glycol ointment). The potential for absorption depends on the choice of the bases, and intended use of the medicament. Appropriate selection of ointment bases is important for dermal therapy. Ointments for rectal preparation (e.g., Tronolane[®] ointment for hemorrhoidal analgesia) and vaginal preparations (e.g., Mycelex-7[®] ointment as antifungal) are available in the market. Creams are semisolid emulsions with one or more medicinal agents intended for external use. The so called *vanishing creams* are o/w emulsions with stearic acid and cold creams are w/o emulsions with an oily base. Creams spread more easily than ointments and are preferred by

some patients. Gels are semisolid systems with dispersions of small or large molecules in an aqueous vehicle with a gelling agent e.g., high molecular weight Carbopols[®] that are cross-linked polyacrylic acid. Some gels like milk of magnesia or magma has two phases. These behave as thixotropic systems with the viscosity changing due to a *gel-sol* transition on shaking. Pastes are ointments with large amount of powder levigated into the base and are intended for application to the skin. Pastes are more hygroscopic than ointments and are used to absorb serous secretions. (e.g., zinc oxide paste). Plasters are solid or semisolid masses spread on backing paper, plastic, or fabric. Nonmedicated forms are termed as *adhesive plasters* while medicated plasters provide a therapeutic effect at the site of application. Some medicated plasters are termed as *cataplasms*, e.g., ibuprofen (3) and salicylic acid plasters. Transdermal drug delivery systems (TDDS) are used for delivery of actives through the skin into the systemic circulation. Several methods have been used to facilitate the transport of active through the barriers of the skin. Certain absorption enhancers are used to temporarily increase permeability of the skin for increased delivery. More recently, iontophoretic techniques have allowed the delivery of charged chemicals across the skin using an applied electric field (e.g., amino acids and proteins). Sonophoresis or high-frequency ultrasound is also being studied as a means of effectively enhancing transdermal delivery of drugs. Transdermal delivery systems are of two main types the monolithic matrix system that contains the excess drug dispersed in the polymeric matrix and cast into a matrix with a backing layer and frontal membrane, e.g., Estraderm[®] (Novartis). Membrane-controlled transdermal patches contain drug reservoir in the form of a gel or saturated solution of drug with a backing adhesive and a rate controlling membrane e.g., Transderm-Scop[®] (Novartis). Currently, the market is flooded with a variety of transdermal systems for smoke cessation and hormonal delivery patches. More recently efforts are on to develop transmucosal delivery patches.

OTHERS DOSAGE FORMS

Suppositories are solid dosage forms intended for insertion in body cavities like rectum, vagina, and occasionally in the urethra for local or systemic effects. The length, shape, and weight of these depend on the body cavity it is used for. These melt, soften, or dissolve after application depending on the type of suppository base applied. Cocoa butter base suppositories usually melt in contact with the

body temperature. Other bases, such as polyethylene glycol, glycerin, and soap based suppositories solubilize. Cocoa butter suppositories are hydrophobic, and dissolve oil soluble drugs but absorption is poor in the aqueous rectal fluids. Thus, better absorption is obtained with water-soluble bases. Rectal suppositories can be used for local (hemorrhoids) or systemic effects. Systemic bioavailability is poor and the amount of drug required is more than oral administration. Vaginal and urethral suppositories are usually used for their local effects (e.g., anti-infectives). Rectal suppositories weigh about 2 g and have varying shapes like bullets and torpedoes. Urethral suppositories are thin pencil-shaped with tapered ends. Vaginal suppositories are globularized and weigh about 4–5 g. Generally absorption from suppository bases rectally is better when the rectum is empty. Particle size of solids plays an important role in absorption. Most importantly rectal absorption allows to bypass the first pass effect. Most suppositories are manufactured by molding from melts or by compression.

AEROSOL DOSAGE FORMS

Aerosols are pressurized dosage forms containing one or more active drug dissolved, suspended, or emulsified in a propellant or a mixture of solvent and propellant, which is released on actuation of the valve as a fine dispersion of liquid or solid in a gaseous medium. Aerosols are intended for topical administration; for administration into body cavities; for administration orally or nasally as fine solid particles or liquid mists through the pulmonary airways, nasal passages, or oral cavity (buccal or sublingual). Those that provide an airborne mist are called space sprays; those intended for carrying actives to surface are termed as surface sprays and other are termed foam aerosols. Aerosol consists of product concentrate and the liquefied propellant. The pressure depends on the types and amounts of propellants and the nature and amount of active present. Propellant is a liquefied gas or mixture of liquefied gas, which serves as the solvent, or vehicle. In some cases nonliquefied gases like nitrogen, and carbon dioxide are used. Space aerosols (85% propellant) operate at 30–40 psig at 70°F. Surface aerosols (30–70% propellant) operate between 22–55 psig at 70°F. Foam aerosols operate at a slightly higher pressure. Foam aerosols are emulsions of the propellant and product concentrate. Aerosols can be two-phase systems comprising of a solution of drug in liquefied propellant and a vapor phase of propellant/gas. Some of them are present as three-phase systems, comprising of water-immiscible propellant, an aqueous product

concentrate or drug in suspension/emulsion and a vapor phase. Aerosol containers are made from glass (coated and uncoated), tin, aluminum, and stainless steel containers. There are various forms of spray valves and metered valves (for more accurate dosing of potent drug). Oral aerosols are mostly used via the buccal route, e.g., Nitrolingual[®] spray that emits nitroglycerin at a dose of 0.4 mg per metered dose. The respiratory tract offers several advantages for administration of drugs. Inhalation systems should be capable of producing fine particles, usually <10 µm for effective drug delivery. Inhalation drug delivery has been traditionally used to treat respiratory disease, but in recent times the lung has been used as a portal for administering drugs to the systemic circulation. With their large effective surface area, the lungs offer an attractive route for systemic drugs. Three main dosage forms, viz. metered dose inhalers (MDIs), dry powder inhalers (DPIs), and nebulizers have gained prominence (4). Metered dose inhalers are pressurized systems consisting of drug suspension in a propellant and may contain other additives such as surfactants, antioxidants, and solvents, although some solution systems are available. Traditionally MDIs have been formulated using chlorofluorocarbons (CFCs). These propellants are now being phased out under the terms of the “Montreal Protocol” due to their ozone depleting potential (5). This has resulted in an increasing urgency to reformulate existing MDI formulations with alternative hydrofluorocarbon (HFC) propellants such as HFC-143a and HFC-227. Formulations using the new propellants have recently won FDA approval (Airomir[®], 3M Pharmaceuticals). To offer better coordination between actuation and patient inhalation of the emitted dose, several spacer devices are available (Nebuhaler[®], AeroChamber[®], and Breathancer[®]) with marketed MDI formulation. Recent trends in devices include the breath actuated MDI (Autohaler[®], 3M Pharmaceuticals), where the patient inhalation triggers the dose. Due to the phase out of CFCs and extensive difficulties in reformulating using HFCs, dry powder inhalers, and nebulizers are becoming popular. Dry powder inhalers consist of mixtures of micronized drug and a large particle size carrier (usually lactose). These drug-carrier ordered mixtures are packaged in unit doses (capsules or blisters), as in Spinhaler[®] (Fisons Pharmaceuticals) and Rotahaler[®] (GlaxoWellcome Pharmaceuticals) or as pure drug bulk powder, which can be metered into single doses (Turbohaler[®], AstraZeneca Pharmaceuticals). In many instances the patients inhalation maneuver causes the active particles to separate from the carrier in the air stream. More recently, to overcome intra-patient variability in dosing, active DPIs have been designed (Dryhaler[®], Dura Pharmaceuticals). In these devices, the patient’s breath triggers a deaggregation mechanism, which

separates the drug particles from the carrier, which are then inhaled. Newer fine particle generation technologies, such as spray drying and supercritical fluid extraction, are being used to produce fine particles of drug including proteins and peptides for delivery to the deep lung. This can facilitate the use of inhalation delivery systems for systemic drug delivery. Nebulizers generate fine mists from aqueous and nonaqueous drug solutions, using either compressed air or ultra-sonication. Their main disadvantage is that they are bulky and are not portable. Recent developments in this field have focused on the development of battery operated portable devices, which can nebulize aqueous solution containing minimal, or no preservatives.

PEDIATRIC AND GERIATRIC DOSAGE FORMS

Physiology plays an important role in the development and performance of different functions in the body. It is important to determine their consequence on dosage form development. Pediatric dosing is usually determined by weight and age. FDA classifies the pediatric groups into neonates, infant, child, and adolescent. There are several excipients that have been reported to cause adverse reactions, e.g., azo dyes cause bronchoconstriction, lactose may cause prolonged diarrhea and intolerance, and sweeteners such as saccharin are weak carcinogens. Alcohol is a common solvent for most pediatric OTC liquid products. However, limits on the alcohol content of OTC products have been set to minimize toxicity in children. Oral administration is the preferred route for children. However, children younger than 5 years have difficulty swallowing solid tablets. Thus, oral liquid is the most preferred dosage form in pediatric patients. Liquids are often unstable and have short expiration and accurate dosing is difficult. Recently, there has been increased interest in chewable tablets and “sprinkle” powders in capsule formulations as they are well received by children with dentition. Rectal administration is not popular because of wide variability in absorption. Pulmonary administration is emerging to be a popular delivery mode for children, but needs to be studied further for systemic effects. Transdermal route may be another area to explore for children, since the stratum corneum is well developed in children as in adults. This may be beneficial as an alternate route for children. Pediatric drug therapy has very few drug delivery systems. There has been some inroads made with OTC cough and cold products. However, most industries do not have resources to perform separate studies for safety and efficacy in children (smaller consumer than adults) for new chemical entities. Perhaps FDA should take initiatives to provide pharmaceutical

companies with returns like tax break or patent extension and specific market for such developmental work especially for life threatening diseases (6).

Apart from alterations in the pharmacodynamics (PD) and pharmacokinetics (PK), the geriatric population suffers from a number of chronic conditions and physical limitations. Clinical monitoring becomes very important to titer dosing accurately. Most of their PK and PD processes take a down turn. Absorption is slower from the oral cavity. In general, the aged skin is more permeable to water and other chemicals. However, the clearance to the blood stream is lowered thus distribution may not be complete. Physically, impairment or decline in vision may hinder one's ability for self-medication. Also, swallowing and chewing may be a problem in elderly patients. For example, patients suffering from dry mouth may have difficulty swallowing a tablet or capsule. Similarly, elderly patients who are edentulous (i.e., toothless) are incapable of chewing any tablet dosage form. Although sublingual and buccal tablets are used by the elderly population there is very less emphasis on its effect on bioavailability with aging. Patients with dry mouth condition may feel local irritation with such dosage forms. Capsules like tablets may hinder swallowing and are not advisable for elderly patients. Liquids are easier to swallow, but are usually not packaged as unit dosages. Patients with impaired vision and dexterity may not be able to accurately self-administer the required dose. The transdermal route seems to offer better compliance with elderly patients but bioavailability needs to be determined before using this route. Several alternative dosage forms and packaging techniques are emerging to improve compliance of dosage form in elderly patients. Several types of packaging aids like dosett tray, calendar-packs and med packs are used to remind the dosing schedule for elderly patients. Granules of drug may be easy to swallow. They can be mixed with water or food and swallowed easily. Unit dose packs may still be difficult to use for some elderly patients. Effervescent tablets provide an alternative dosage form. These tablets dissolve in water to form a ready-to-use product. The use of an irregular shaped tablet that prevents it from lying flat may be another form that can help patients with impaired dexterity (e.g., Tiltabs®). Similar to granules the drug may also be presented in the form of a small amount of concentrated solution for the entire dose, e.g., 5 ml Rapamune® concentrated oral solution (Wyeth–Ayerst Research). Such products can be mixed with food or drink. This is similar to the use of a dispersible tablet that forms a uniform stable suspension when dispersed in water. Emerging technology has focussed on newer dosage forms like the use of quick or rapid dissolving technology (RDTs), wherein, the dosage form quickly dissolves in the mouth and rapid absorption of

the drug can occur systemically or even from the mouth (7). Furthermore, compliance in elderly patients has also resulted in the availability of sugar and sodium-free products that are beneficial for such age groups.

NEW DRUG-DELIVERY TECHNOLOGIES

Newer technologies are emerging as we move into the new millennium. These technologies promise to have lots of benefits such as simplifying administration regimens, enhancing compliance, improving clinical benefits, and reducing overall healthcare costs. Rapid-dissolving tablets (RDTs) are designed for patients who have difficulty in swallowing standard tablets/capsules, such as pediatric and geriatric patients. These include the lyophilized foam from Zydis[®] (Claritin Reditabs), Flashtab[®] (Prographarm), Orasolv[®] (Cima Labs), Wowtabs[®] (Shaklee), and Flashdose[®] (Fuisz Technology) (8, 9). Hydrogel based technology offered by Professor Neil Graham of British Technology group for development of several systems including morphine suppositories. Nanocrystal[™] technology offered by Elan Corporation where the crystalline drug (<400 nm) is thinly coated with a surface modifier to impart physical stability. Liquitard[®] is a liquid taste masking sustained release granules as a suspension and is offered by Eurand America (10). Inhale Therapeutics offers proprietary technology for pulmonary delivery of proteins and peptides, using innovations in powder processing to develop formulations for deep lung delivery for systemic and local indications (10). Jago Pharma developed the Geomatrix[®] systems which involves the use of multi-layered hydrophilic matrix systems (10). Several newer excipients are now available as matrices for controlled delivery. These include polysaccharides from Galactomannan such as guar gum and locust bean gum (e.g., Timer_x Technology (11). Considerable research efforts have been towards the development of safe and efficient chitosan-based dosage forms (12). The development of solid-lipid nanoparticles (SLN) have made it possible for delivering drugs with less side-effects, better targeting, and protection from enzymes (13). Zambon group in Europe has developed Timeclock[®] technology that involves coating solid drug with hydrophobic surfactant. Chronotropic drug delivery, which targets delivery to a specific absorption window for local as well as systemic effect. Emisphere[®] technology uses a carrier that binds to drug molecules noncovalently to form a complex (14). These complexes easily cross the membranes and then dissociate to release the active. Theratech uses the Theriform[®] microprinting technology to develop oral and implantable dosage form for

making Microdose[®] tablets (15). Delsys Corporation has revolutionized their Accudrop[®] technology of electrostatic deposition of dry powder to any surface with great accuracy (16, 17). Labopharm Inc. has developed Contramid[®] technology obtained by cross-linking of high-amylose starch in three dimensional network which is combined with the active. Once in the stomach, the tablet surface turns into a gel and the active diffuses at an even rate. PORT[™] (Programmable Oral Release Technology) is a technology platform to resolve drugs with biopharmaceutical problems. PORT[™] is a capsule based system with opportunity to provide multiple prolonged release of one or more drugs. This technology is applicable to a wide variety of drug classes and can be used to achieve difficult PK profiles, e.g., zero order with burst system. Quadrant Healthcare has engineered micron-sized particles with enhanced stability and PK profiles (10). Solidose[®] technology is based on chemically modifying oligosaccharides to make them more hydrophobic. In contact with body fluids they undergo phase change and release the drug. Several self-emulsifying and lipid-based systems have been developed that form micro-emulsions of water-insoluble drugs for oral delivery (18). DanBiosystems has developed a proprietary Targit[®] technology, which involves enteric coating with azo polymers, which degrade only at specific sites by bacterial enzymes. Ethypharm is another company involved with coating of nano- and micro-particles, using supercritical fluid (CO₂) technology, suited for fragile water-soluble molecules (peptides and proteins) (10, 19). Protarga[™] is another company which is involved in the development of dosage form by covalently attaching fatty acids (docosahexaenoic acid or DHA) to actives to create new compounds that can be taken up by cells targeted for treatment. A lot of focus has been directed to tissue engineering that is used to design biological substitutes or regenerate natural tissues for defective or lost tissues and organs through the use of cells and their scaffolds (20, 21). Fentanyl Oralet[®] (Abbott labs) is a lollipop that has painkillers used as preoperative sedative. Several forms of intra-vaginal drug delivery systems like Progestasert[®] system (Alza Corp.) and Dinoprostone[®] vaginal insert have been developed. Implants like the levonorgestrel Norplant[®] device (Wyeth–Ayerst Research) that is incorporated as contraceptive on the upper arm. Similarly, Gliadel[®] wafers (Guilford Pharmaceuticals) are implanted in the brain tumor cells.

CONCLUSION

Non-parenteral dosage forms can be administered by different mechanisms. In recent years there has been a

plethora of emerging drug delivery technology companies. Most of these technologies are addressing issues related to unique delivery systems and selectivity of these systems for affected organs or diseases. However, rarely it is recognized that improvements in drug therapy are a consequence of not only the new chemical entity but also the combination of active and the delivery system (dosage form). Currently the trend is to develop a delivery system and then look for suitable drug candidates to apply. This needs to be changed to gain understanding of what unmet medical need is the active aimed for, the rate, time and site for the active to be delivered and then provide a suitable delivery system. More emphasis should be made on the improvement of drug effect profile. Thus, the future of drug delivery will depend on how they can contribute to drug therapy for unmet medical needs (22).

REFERENCES

- Conte, U.; Maggie, L.; Colombo, P.; LaManna, A. Multi-Layered Hydrophilic Matrices as Constant Release Devices (Geomatrix™ Systems). *J. Controlled Release* **1993**, *26*, 39–47.
- Swarbrick, J. Advances in Controlled Drug Delivery. *S.T.P. Pharm. Pratiques* **1996**, *6* (1), 53–60.
- Agarwal, V.; Singh, S.K.; Reddy, I.K.; Durranu, M.J.; Khan, M.A. Catapasm-Based Controlled Drug Delivery: Development and Optimization of a Novel Formulation. *Drug Dev. Ind. Pharm.* **1999**, *25* (5), 659–665.
- Gonda, I. The Ascent of Pulmonary Drug Delivery. *J. Pharm. Sci.* **2000**, *89* (7), 940–945.
- McDonald, K.J.; Martin, G.P. Transition to CFC—Free Metered Dose Inhalers—Into the New Millennium. *Int. J. Pharm.* **2000**, *201* (5), 89–107.
- Nahata, M.C. Lack of Pediatric Drug Formulations. *Pediatrics* **1999**, *104* (3), 607–609.
- Sastry, S.V.; Nyshadham, J.R.; Fix, J.A. Recent Technological Advances in Oral Drug Delivery—A Review. *Pharm. Sci. Tech. Today* **2000**, *3* (4), 138–145.
- Habib, W.; Khankari, R.; Hontz, J. Fast-Dissolve Drug Delivery Systems. *Crit. Rev. Ther. Drug Carrier Sys.* **2000**, *17* (1), 61–72.
- Chang, R.; Guo, X.; Burnside, B.A.; Couch, R.A. Fast-Dissolving Tablets. *Pharm. Technol.* **2000**, June, 52–58.
- Evers, P. *Developments in Drug Delivery: Technology & Markets*; Healthcare and Pharmaceutical Publishing: London, 1995; Financial Times.
- Altaf, S.; Yu, K.; Parasrampur, J.; Friend, D.R. Guar Gum-Based Sustained Release Diltiazem. *Pharm. Res.* **1998**, *15* (8), 1196–1201.
- Illum, L. Chitosan and Its Use as a Pharmaceutical Excipient. *Pharm. Res.* **1998**, *15* (9), 1326–1331.
- Pinto, J.F.; Muller, R.H. Pellets as Carriers of Solid Lipid Nanoparticles (SLN) for Oral Administration of Drugs. *Pharmazie* **1999**, *54* (7), 506–509.
- Leone-Bay, A.; Leipold, H.; Agarwal, R.; Rivera, T.; Baughman, R.A. The Evolution of an Oral Heparin Dosing Solution. *Drugs Future* **1997**, *22* (8), 885–891.
- Kastra, W.E.; Palazzolo, R.D.; Rowe, C.W.; Giritliglu, B.; Teung, P.; Cima, M.J. Oral Dosage Forms Fabricated by Three-Dimensional Printing. *J. Controlled Release* **2000**, *66*, 11–17.
- Chrai, S.S. Electrostatic Dry Deposition Technology. *Pharm. Tech.* **1998**, *4*, 106–112.
- Houston, S. A Revolution in Dose Form Manufacture. *Manuf. Chemist.* **1998**, *6*, 20–21.
- Charman, W.N. Lipids, Lipophilic Drugs, and Oral Drug Delivery—Some Emerging Concepts. *J. Pharm. Sci.* **2000**, *89* (8), 967–978.
- York, P. Strategy for Particle Design Using Supercritical Fluid Technologies. *Pharm. Sci. Tech. Today* **1999**, *2* (11), 430–440.
- Langer, R. Selected Advances in Drug Delivery and Tissue Engineering. *J. Controlled Release* **1999**, *62*, 7–11.
- Tabata, Y. The Importance of Drug Delivery Systems in Tissue Engineering. *Pharm. Sci. Technol. Today* **2000**, *3* (3), 80–89.
- Breimer, D.D. Future Challenges for Drug Delivery. *J. Controlled Release* **1999**, *62*, 3–6.

FURTHER READING

- Ansel, H.C.; Allen, L.V.; Popovich, N.G. *Pharmaceutical Dosage Forms and Drug Delivery Systems*, 7th Ed. Lippincott's Williams and Wilkins: 1997.
- Banker, G.S., Rhodes, C., Eds.; *Modern Pharmaceutics*, 3rd Ed.; Marcel Dekker, Inc.: New York, 1996.
- Block, L.H.; Yu, A.B.C. *Pharmaceutical Principles and Drug Dosage Forms. Comprehensive Pharmacy Review*; Shargel, L., Mutnick, A.L., Souney, P.F., Swanson, L.N., Block, L.H., Eds.; Williams & Wilkins: 1997; 28–76.
- Byron, P.R., Ed.; *Respiratory Drug Delivery*, CRC Press: Boca Raton, FL, 1990.
- Chien, Y.W. *Novel Drug Delivery Systems*; Marcel Dekker, Inc.: New York, 1992.
- General Information/Pharmaceutical Dosage Forms. *USPXXIV*; The United States Pharmacopoeial Convention, Inc. Rockville MD, 2000, 2107–2117, <1151>.
- Hickey, A.J., Ed.; *Pharmaceutical Inhalation Aerosol Technology* Marcel Dekker, Inc.: New York, 1992.
- Hickey, A.J., Ed.; *Inhalation Aerosols: Physical and Biological Basis for Therapy*; Marcel Dekker, Inc.: New York, 1996.
- Lachman, L., Lieberman, H.A., Eds.; *Pharmaceutical Dosage Forms: Tablets*; Marcel Dekker, Inc.: New York, 1980–1982; 1–3.
- Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; *Theory and Practice of Industrial Pharmacy*, 4th Ed., Lea & Febiger: Philadelphia, 1987.
- Martin, A.N., Swarbrick, J., Cammarata, A., Eds. *Physical Pharmacy*, 3rd Ed., Lea & Febiger: Philadelphia, 1983.
- Robinson, J.R.; Lee, V.H.L. *Sustained and Controlled Release Drug Delivery*, 5th Ed., Marcel Dekker, Inc.: New York, 1995.
- Shargel, L.; Yu, A. *Applied Biopharmaceutics and Pharmacokinetics*; Appleton & Lange: 1999.

DOSAGE FORM DESIGN: A PHYSICOCHEMICAL APPROACH

Michael B. Maurin

DuPont Pharmaceuticals Company, Wilmington, Delaware, U.S.A.

Anwar A. Hussain

Lewis W. Dittert

University of Kentucky, Lexington, Kentucky, U.S.A.

INTRODUCTION

Over the past several years, the fraction of new drug products that are new chemical entities has steadily decreased, reflecting the tremendous cost required to bring new chemical entities to the marketplace. Increased understanding of drug metabolic and toxicologic factors, such as the effect of the patient age on drug distribution, the genetic factors that may result in dramatic intersubject variability in metabolism, short-term versus long-term exposure toxicities, and the potential for teratogenic, mutagenic, and embryotoxic effects, has increased the scrutiny under which governmental agencies view the new chemical entity. This careful inspection is intended to minimize the possibility of toxic reaction(s) and to demonstrate the safety and efficacy of new drug products. The regulatory process has also resulted in significantly more costly and time-consuming testing prior to commercialization.

This increased emphasis on safety has placed an additional burden on those who are involved in the development of new drugs, while increasing financial pressures have led to the need for decreased development time. The investigation of approved drugs has resulted in enhanced patient safety and therapeutic efficacy by directing research efforts toward the more efficacious delivery of known pharmacologically active agents to the appropriate physiologic site. This trend has caused pharmaceutical researchers to seek the most suitable methods to deliver both new and existing compounds in the most pharmacologically appropriate manner. The methods may be designed to optimize bioavailability, minimize toxicity and side effects, and improve stability. The objective of this article is to present approaches that have been employed to improve bioavailability and/or minimize the toxicity and side effects of various drugs.

A rational approach to dosage form design requires a complete understanding of the physicochemical and biopharmaceutical properties of the drug substance.

For example, the successful design of an efficacious oral dosage form requires an understanding of the pathways of physiologic disposition of the drug. Some of the physiologic factors associated with drug disposition are illustrated in Fig. 1 (1). On oral administration of an immediate release dosage form (the most common delivery system), the dosage form must disintegrate, the drug must dissolve in the gastrointestinal (GI) fluids, cross the GI mucosa, enter the mesenteric blood system, and pass through the liver prior to reaching the systemic circulation and the site of action. The drug may be metabolized by GI fluids, by enzymes in the gut wall, or by hepatic metabolism prior to reaching the systemic circulation (Fig. 2) (2). The net result is incomplete bioavailability due to first-pass metabolism (inactivation), and/or metabolic formation of a pharmacologically active species.

The physicochemical and biopharmaceutical properties of the drug can have a tremendous impact on its bioavailability and, hence, on its efficacy and toxicity profile. Thus, understanding these parameters is often tantamount to the selection and development of the optimum dosage form.

These properties of the drug are its:

- pH solubility profile and dissolution rate
- Partition coefficient between lipoidal barriers and aqueous physiologic media
- Stability and/or degradation rate in the physiologic fluids
- Susceptibility to metabolic inactivation
- Mechanism of transport through biologic membranes.

The aqueous solubility of a drug in the 2–8 pH range has a direct influence on its oral and parenteral formulations. A drug with poor solubility (i.e., less than 0.1 mg/ml) in acidic media may show poor and erratic oral bioavailability due to the dependency of absorption processes in GI fluids. Intravenous dosing requires that the drug be administered in a soluble form. The adjustment

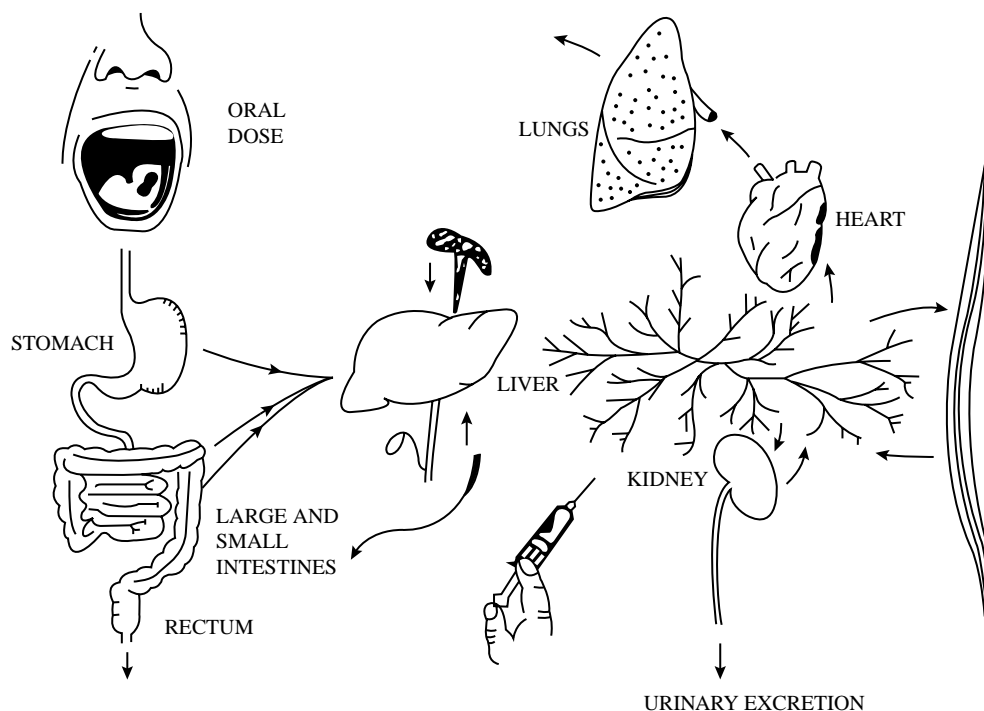


Fig. 1 Physiological factors associated with bioavailability. (From Ref. 1.)

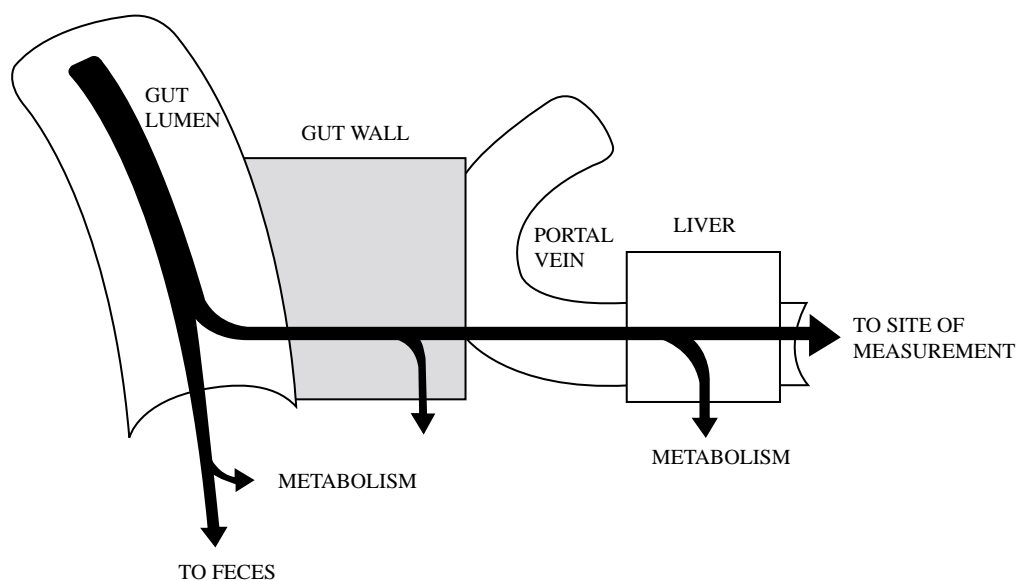


Fig. 2 A drug, given as a solid, encounters several barriers and sites of loss in its sequential movement during gastrointestinal absorption. Dissolution, a prerequisite to movement across the gut wall, is the first step. Incomplete dissolution or metabolism in the gut lumen or by enzymes in the gut wall is a cause of poor absorption. Removal of a drug as it passes through the liver further reduces absorption. (From Ref. 2.)

of pH, the addition of a cosolvent or a ligand for complexation, or the formation of an emulsion may permit solubilization, but each of these techniques has limitations. Rapid intravenous administration of a solubilized drug can result in rapid dilution in an environment in which the drug is insoluble, resulting in incomplete availability and a delayed response due to the formation of particulate matter within the vascular system.

Poor aqueous solubility is not always a limitation; in fact, it may be a desirable feature for a sustained effect after oral or parenteral administration. Oral sustained release may be achieved if the drug combines poor aqueous solubility with the ability to be adsorbed throughout the GI tract. Parenteral sustained release can be achieved after intramuscular administration of a suspension of a drug with low solubility under physiologic conditions or from a drug that precipitates from an aqueous vehicle or that forms a reservoir or depot from an oil-containing dosage form.

The lipid-aqueous partition coefficient of a drug molecule affects its absorption by passive diffusion. In general, octanol/pH 7.4 buffer partition coefficients in the 1–2 pH range are sufficient for absorption across lipoidal membranes. However, the absence of a strict relationship between the partition coefficient of a molecule and its ability to be absorbed is due to the complex nature of the absorption process. Absorption across membranes can be affected by several diverse factors that may include the ionic and/or polar characteristics of the drug and/or membrane as well as the site and capacity of carrier-mediated absorption or efflux systems.

Compounds that are intended for oral administration and can undergo rapid degradation at low pH may require protection from the acidic environment of the stomach. Protection can often be afforded by administering the drug in the form of an acid-insoluble chemical species or in a dosage form with an acid-resistant coating. The insoluble chemical species must remain insoluble and unavailable for solution degradation as it passes through the stomach and must dissolve upon reaching the chemically more stable environment of the intestine at a higher pH. To be effective, an acid-resistant coating must remain intact and protect its contents until it reaches the required pH to dissolve the coating and release the contents in the intestine where the drug may be more stable.

Metabolic inactivation of a compound following oral administration can occur in the GI lumen, the GI mucosa, or the liver. The site of metabolism and the susceptibility of the metabolic processes to saturation are factors that may influence oral bioavailability. Occasionally, some of these factors may be altered to

optimize oral bioavailability. For example, segment specific metabolic sites within the GI tract may be avoided through the use of pH-dependent coating materials that rely on the local pH environment of the GI tract to release the drug. Absorption from the lower colon and rectum can reduce exposure to the portal circulation and the first-pass inactivation that can occur in the liver and, thus, provide the opportunity to improve systemic availability following oral administration. Enzyme systems may be saturated by the rapid release of the contents of the dosage form at a local site or by coadministration within the dosage form of a competitive inhibitor.

Once the physicochemical and biopharmaceutical properties of the drug are determined and the desired plasma concentration profile is defined, the pharmaceutical scientist can select and develop an efficacious dosage form by utilizing a formulation approach, a prodrug approach, a device approach, or an alternative administration route approach.

FORMULATION APPROACH

Use of formulation techniques can improve the bioavailability and/or minimize the toxicity and side effects of drugs. Factors to consider include those that impact on solubility and dissolution rates, chemical and enzymatic stability, and absorption capability.

Several parameters, including particle size, crystalline habit, and salt form, can affect the solubility and dissolution rates of the drug. The effect of particle size on the dissolution rate of relatively insoluble compounds becomes significant when the drug is administered as a suspension or solid dosage form and the material is well dispersed within the GI tract. However, caution must be exercised when compressing the material into a final dosage form because excessive force can result in particle agglomeration and an actual increase in the effective particle size.

Polymorphism is the ability of a chemical species to crystallize in more than one distinct crystal habit. The pharmaceutical applications of polymorphism have been reviewed by several authors (3–5). The differences in dissolution rate and solubility that polymorphs can produce may have a dramatic impact on bioavailability when dissolution is the rate-limiting step in the absorption process.

Tawashii (6) investigated the GI absorption of two polymorphs of aspirin, the stable and metastable forms, forms I and II, respectively. He found that the metastable

form produced a 70% higher total serum salicylate levels than the stable form I.

The selection of a salt form directly influences the physicochemical and biopharmaceutical properties of a compound. The impact of salt selection has been reviewed (7–9). Nelson (10) examined the dissolution of theophylline salts and commented on their impact on oral administration. The dissolution rates of the theophylline salts proceeded independently of the pH of the medium but was governed by the diffusion layer pH. The choline and isopropanolamine salts dissolved three to four times faster than the ethylenediamine salt and produced higher and prolonged blood levels.

For highly insoluble amine bases, such as ergotamine and certain antimalarials, the drug may precipitate in the small intestine as soon as the pH rises following stomach emptying. The authors' experience has been that committing the effort and expense required to develop a new, more soluble salt form of this type of drug may be futile, especially if the new salt is only two to three times more soluble than existing salts. Thus, the selection of the most appropriate salt form should be made early in the development process to optimize bioavailability.

The stability of a drug in the gut is influenced by both chemical and enzymatic factors. Protection from chemical degradation may be accomplished via coating techniques, and enzymatic protection may be achieved with enzymatic inhibitors.

An enteric coating protects the drug during transit through the acidic medium of the stomach. Upon entering the higher pH environment of the duodenum, the coating is dissolved, and the drug becomes available for absorption. Such a coating also provides protection for the gut mucosa when the drug is capable of producing GI irritation. This method has been employed for a number of drugs, including potassium chloride, ammonium chloride, aspirin, diethylstilbestrol, erythromycin, and divalproex. Nishimura et al. (11) employed enteric coating of levodopa to improve its bioavailability. Levodopa is absorbed from the upper portion of the small intestine but undergoes rapid degradation in the intestinal mucosa by levodopa decarboxylase. The authors developed an enteric-coated dosage form with an effervescent core. The dosage form remains intact through the stomach, dissolves upon reaching the small intestine, and bursts open due to the effervescence of the incorporated sodium bicarbonate (Fig. 3). The rapid release in the duodenum results in concentrations of levodopa sufficient to saturate the enzyme at the absorption site.

The limitations of enteric-coated dosage forms include the possibility of duodenal irritation from caustic drugs and an increase in intersubject variability due to

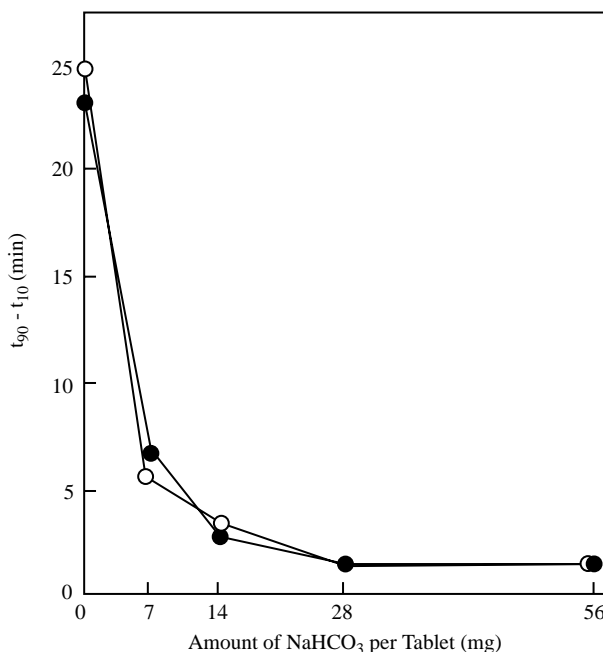


Fig. 3 Relationship between the dissolution parameter t_{90-10} of effervescent enteric tablets of levodopa and the amount of sodium bicarbonate formulated in the tablet. The number of strokes was fixed at 5/min and the pH was 7.5. Key: (●) = uncoated tablet; (○) = enteric tablet. (From Ref. 11.)

the presentation to the small intestine of a dosage form that needs to undergo disintegration and dissolution versus a disintegrated and partially dissolved drug substance.

Enzyme inhibitors compete with the active drug for the enzyme and, thereby, reduce the degradation of the drug and deliver it more efficiently to the systemic circulation. An example is the carbidopa–levodopa combination. Carbidopa competes for levodopa decarboxylase, thereby reducing the levodopa degradation and improving the low bioavailability of levodopa.

Alternatively, the enzyme in the gut can be utilized to control the release of the active drug in the gut. For example, sulfasalazine, which is employed in the treatment of ulcerative colitis, is a combination of sulfapyridine and 5-aminosalicylate chemically linked via an azo bond. It remains absorbed and intact throughout the GI tract until it reaches the large intestine, where bacterial azoreductase enzymes degrade the azo bond and release sulfapyridine and 5-aminosalicylate to act locally on the lesions.

Altering the availability for absorption permits tailoring of the concentration-time profile for a drug. In the case of synthetic contraceptive steroids and theophylline dosage

forms, the approach should be to slow the release rate so that undesirable spikes in the plasma levels are minimized. This effect can be accomplished with a sustained-release formulation. In formulating a drug in a sustained-release dosage form, the following must be taken into consideration: the effective drug plasma level, the rate of absorption of the compound, the rate of elimination of the compound, and the site(s) of absorption. Having established these parameters, calculating the required release rate from the dosage form is a relatively simple matter and formulation techniques are readily available to produce the desired release rate. Predicted plasma levels for a steroid drug administered in sustained and conventional dosage forms are shown in Fig. 4.

The mechanism of absorption must always be evaluated when a sustained-release dosage form is considered. A drug that is passively absorbed throughout the GI tracts is an ideal candidate for sustained release. Drugs such as riboflavin, folic acid, aminopenicillins, amino- β -lactams and nucleoside analogs, which have windows of absorption due to site-specific and/or active transport processes, may have incomplete bioavailability when formulated in oral, sustained-release dosage forms.

PRODRUG APPROACH

An alternative to the formulation approach is the prodrug approach. A prodrug is defined as a drug that is prepared by chemically modifying a pharmacologically active species to form a new chemical entity that undergoes

transformation to the active species within the body. The modification alters the physicochemical and biopharmaceutical properties of the drug in some beneficial manner.

The ideal prodrug should have the following characteristics:

- Possess no pharmacologic activity
- Be eliminated more slowly than its rate of cleavage to the parent
- Be nontoxic
- Be inexpensive to prepare.

Prodrugs can be used to increase or decrease the aqueous solubility, mask bitterness, increase lipophilicity, improve absorption, decrease local side effects, and alter membrane permeability of the parent molecule. For example, chloramphenicol has an aqueous solubility of 2.5 mg/ml, but chloramphenicol sodium succinate, a prodrug, has an aqueous solubility of 100 mg/ml. Hydantoins also possess low aqueous solubilities that result in low and variable availability and precipitation following injection. In an effort to increase the aqueous solubility of phenytoin, Stella et al. (12, 13) prepared the ethyl and triethylamine esters of diphenylhydantoic acid. The esters improved aqueous solubility by adding an amine function to the molecule, and under physiologic conditions, the cyclization to phenytoin is irreversible, rapid, and complete.

The perception of taste requires that some minimum aqueous concentration be exceeded so that the taste can be detected (perceived). As a result, bitterness can be masked by reducing solubility. The technique has been applied to

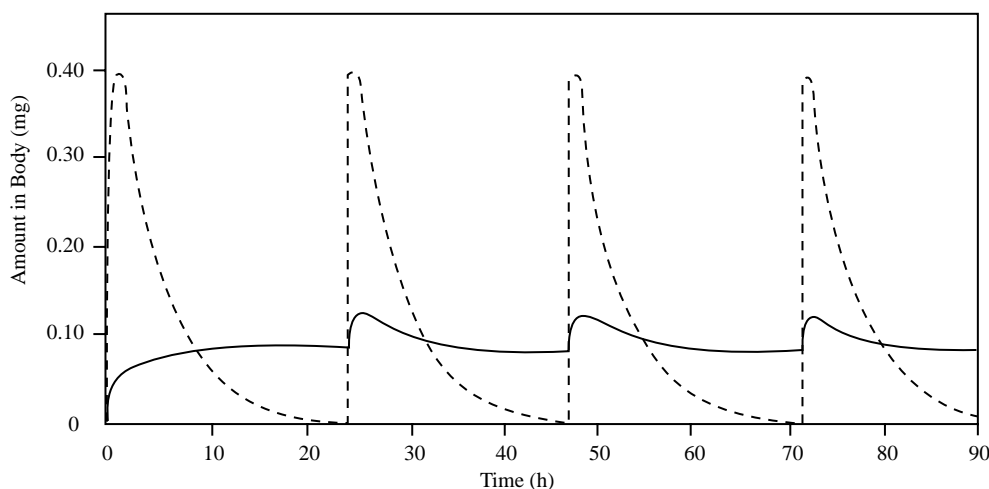


Fig. 4 Computer simulation of a plasma level-time profile of norethindrone from fast-releasing tablet (---) and sustained release tablet (—), from a dose of 0.5 mg/tablet and an elimination half-life of 3 h.

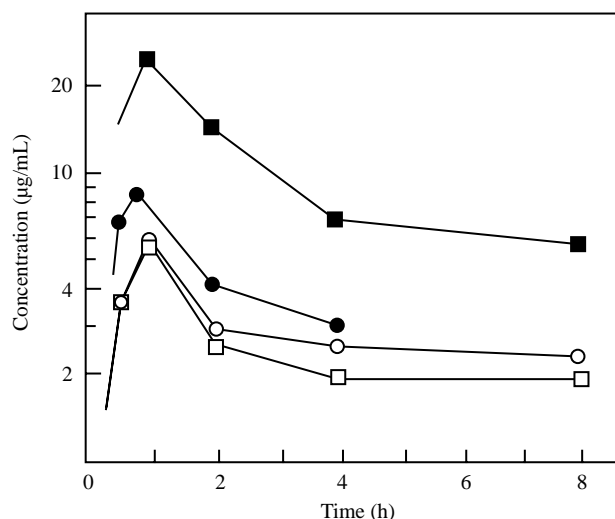


Fig. 5 Plasma acetaminophen concentration following the oral administration of acetaminophen and its ethyl vinyl ether prodrug. Dog 1 received 10 mg/kg doses of acetaminophen (○) and the prodrug (□). Dog 2 received 25 mg/kg doses of acetaminophen (●) and the prodrug (■). The 0.5-h prodrug and 8-h acetaminophen samples for dog 2 could not be obtained. (From Ref. 14.)

acetaminophen by blocking the phenolic substituent with ethyl vinyl ether (14). Chemical hydrolysis to acetaminophen is rapid in the acidic conditions of the stomach. Plasma acetaminophen concentrations in dogs following oral administration of acetaminophen and the prodrug were found to be similar (Fig. 5).

Membrane permeability is governed in part by the lipophilicity of a compound. Highly polar compounds have low lipophilicities and, therefore, low membrane permeability. Epinephrine is a compound of this type. It is very effective in the treatment of glaucoma, but it produces a myriad of side effects. Ocular side effects include hyperemia, mydriasis, corneal edema, and allergic sensitivity. Systemic side effects, such as cardiac arrhythmias, elevated blood pressure, cerebral vascular accidents, dizziness, fear, and restlessness, are observed frequently. In an attempt to enhance the absorption and minimize the side effects of epinephrine, a prodrug, the dipivaloyl ester, was synthesized (15). It was found to be devoid of cardiac symptoms, with no effect on heart rate or blood pressure (Fig. 6) and was effective in lowering intraocular pressure (IOP) (Fig. 7).

On the other hand, highly lipophilic compounds, such as hormones, can be solubilized via the prodrug approach. For example, the rate of transdermal absorption of the highly lipophilic drug, testosterone, was enhanced over

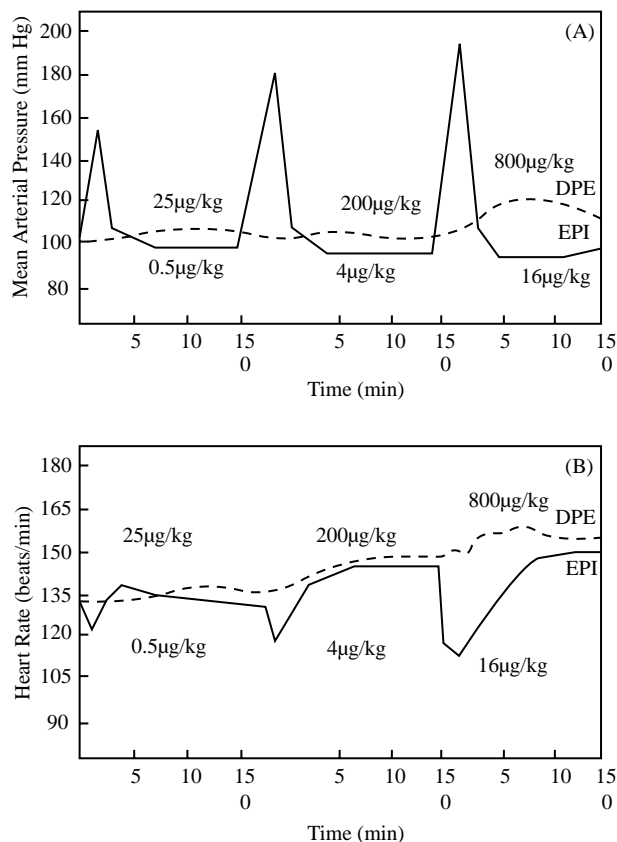


Fig. 6 (A) The effect of intravenous epinephrine (EPI) and its dipivaloyl prodrug (DPE) on blood pressure in dogs. (B) The effect of EPI and its DPE on heart rate in dogs.

50-fold by forming water-soluble, yet lipophilic, prodrug ester (16). The prodrug testosterone-4-dimethylaminobutyrate was found to penetrate human skin tissue, *in vitro*, 54 times faster than testosterone itself (Fig. 8). Furthermore, the prodrug was found to generate testosterone rapidly in biological fluids by enzymatic hydrolysis (Fig. 9).

Oral administration of aspirin can result in gastrointestinal bleeding. The bleeding has been attributed to local irritation due to the acidic nature of the carboxylic acid substituent. In an attempt to reduce the gastric irritation of aspirin, acylal prodrugs were synthesized (17, 18). *In vitro*, the prodrugs generated rapidly aspirin and in a pH-independent fashion. The conversion to aspirin was very sensitive to changes in dielectric constant: decreases caused a corresponding decrease in the reaction rate. This sensitivity is strongly indicative of an S_N1 mechanism. The authors postulated that GI irritation would be reduced due to blockage of both the charge and the acidic nature of the aspirin.

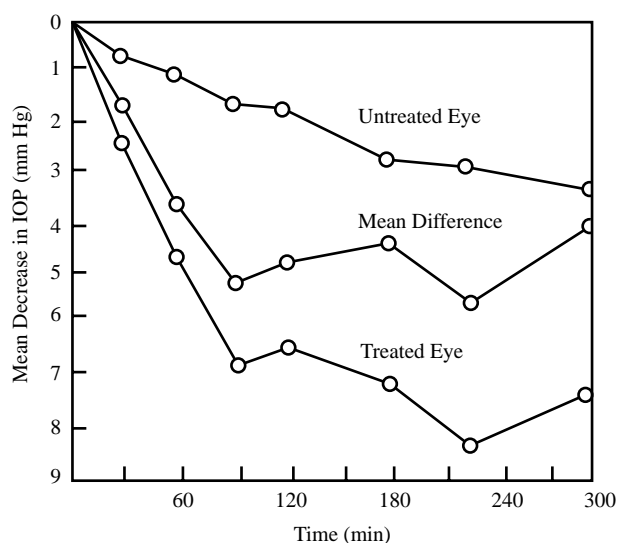


Fig. 7 The mean effect of one drop of 0.025% solution of the DPE of epinephrine on the intraocular pressure of nine glaucomatous individuals.

DEVICE APPROACH

Controlling the release of medication at the site of action is often desirable, especially for compounds that are absorbed rapidly through mucous membranes or are

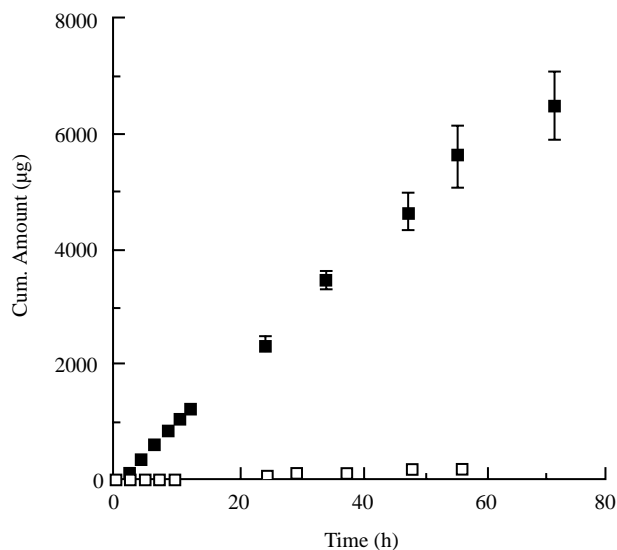


Fig. 8 Cumulative amounts of TS (□) and TSBH (■) crossing isolated human skin tissue in Franz cells ($n = 6$). TS was applied as a 10% suspension. TSBH was applied as a 10% solution. In each case, the vehicle was pH 7.4 phosphate buffer (0.2 M). (From Ref. 16.)

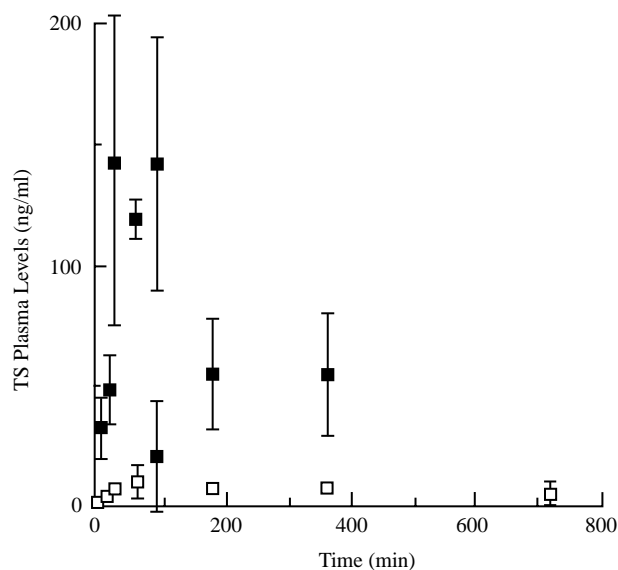


Fig. 9 Total plasma levels of TS (radioimmunoassay) after topical application of 1.25 μmol of TS (□) or TSBH (■) to the backs of HRS/J hairless mice $n = 6$. (From Ref. 16.)

removed rapidly from the site of action. The approach normally reduces the systemic side effects of the agent. The application of this approach can be illustrated with the Progestasert and Pilocarpine Ocusert dosage forms.

Diffusion-controlled devices may be designed for continuous release and usually use either a matrix or reservoir construction. In matrix systems, the drug is dispersed randomly throughout a polymer, whereas reservoir devices surround the drug with an intact rate-controlling membrane. Regardless of the method of construction, the system must be safe and biocompatible for biological application.

The Progestasert intrauterine device (IUD) is a contraceptive IUD marketed in the United States. It is a white, T-shaped unit constructed of ethylene-vinyl acetate copolymer containing titanium dioxide. It releases progesterone at a rate of 65 $\mu\text{g}/\text{day}$ for 1 year, controlled by an outer coat of ethylene-vinyl acetate copolymer. Zaffaroni pointed out several advantages to the uterine progesterone system (19). It permits target-specific delivery of a fertility-control agent for up to 1 year by utilizing a single natural hormone at the lowest effective level of release of hormonal activity. The 1-year duration gives improved patient compliance while eliminating the pulsing seen with repetitive oral or injectable regimens. The efficacy of the Progestasert was investigated by Aznar and Giner (20). Their results indicated that the systems are highly efficacious in avoiding accidental pregnancy and result in decreased menstrual blood loss.

The Pilocarpine Ocusert is a contact lens-shaped device that is inserted into the lower cul-de-sac of the eye. It provides continuous release of pilocarpine at a rate of 10 or 20 $\mu\text{g/h}$, over a 1-week period, for the treatment of open-angle glaucoma. The system is capable of producing significant lowering of ocular pressure and constriction of the pupillary diameter (21). Its use greatly enhances patient compliance and the convenience of the therapeutic regimen while reducing local and systemic side effects.

Armaly and Rao examined the clinical effects of the Pilocarpine Ocusert systems with different release rates (22). The systems exhibited a dose-response relationship such that increases in the release rate above 50 $\mu\text{g/h}$ resulted in increased ocular hypotensive effects with no appreciable change in ocular pressure. The reduction in pressure observed at the 50 $\mu\text{g/h}$ rate was comparable with the changes seen after the administration of 4–8% of pilocarpine solutions.

ALTERNATIVE ADMINISTRATION ROUTES

The administration of drugs by alternative routes avoids absorption and metabolic barriers that may be present in the GI tract. The routes can also provide systematic availability when oral administration is contraindicated due to a physiologic condition, or the route may provide for a concentration-time profile that approaches intravenous dosing profiles. The ophthalmic, nasal, pulmonary, buccal, transdermal, and rectal routes provide one or more of these advantages.

The ophthalmic route has been used traditionally for topical application for local effects. However, with the increasing number of peptide drugs being developed, the ophthalmic route has been considered for systemic drug delivery (23). After topical administration in the eye, peptides can be absorbed from the mucosa during tear turnover as well as via the blood vessels of the conjunctiva. The route suffers from the hesitancy of practitioners to place a drug into the eye for any reason other than to produce ophthalmic effects. Another drawback is the sensitivity of the eye to irritation by foreign substances.

Nasal administration produces rapid blood levels and rapid responses that approach those obtained from intravenous dosing. In addition, the absorbed drug does not pass through the liver before reaching the systemic circulation, and, thus, first-pass metabolism is avoided. In recent years, the nasal route has received a great deal of attention as a convenient and reliable route for the

systemic administration of drugs, especially those that are ineffective orally and must be given by injection. Recently, butorphanol tartrate was introduced commercially in a nasal spray dosage form (Stadol NS[®]) for the relief of pain, such as migraine headache (24). It would appear that the nasal route could be considered for drugs that meet the following criteria: are ineffective orally; are used chronically; are used in small doses; and are desirable to have rapid entry to the general circulation.

Published work carried out in many laboratories has shown that, with the notable exception of the peptides, drugs with a wide variety of chemical structures are well absorbed through the nasal membranes of animals and man. It is believed by many authors that *in vivo* nasal absorption of compounds with molecular weights less than 300 daltons is not significantly influenced by the physicochemical properties of the drug molecule (25). Factors, such as the size of the molecules and, in the case of peptides, their ability to hydrogen bond with the component(s) of the membrane, are more important than their lipophilicity and their ionization state. For example, the *in vivo* rate of absorption of the very lipophilic drug progesterone is similar to that observed for sodium benzoate, and the *in vivo* rate of absorption of benzoic acid is independent of the pH of the medium (26).

Although many drugs are absorbed rapidly and quantitatively following nasal administration, peptides have generally shown low bioavailabilities. Hussain et al. examined the nasal bioavailability of leucine enkephalin (27). The low bioavailability of this pentapeptide was attributed to hydrolysis in the nasal cavity, with dipeptides causing significant inhibition of the hydrolysis. They concluded that polar compounds, such as peptides, can cross the nasal mucosa, and that administration of low concentrations results in extensive hydrolysis in the nasal mucosa and that hydrolysis of leucine enkephalin can be reduced by concomitant administration of peptidase labile peptides.

If the presence of a pharmacologic activity for a peptide is the only criterion for its use in therapy, then nasal administration may be employed, even though bioavailability is not 100%. However, when bioavailability, as determined by plasma level profiles, is considered, the nasal route for peptide administration is not optimal unless enhancers are employed. Enhancers may cause irritation and reduced membrane integrity. Other drawbacks of nasal delivery include the small volume permitted (i.e., 0.2 ml or less), the need for highly potent drugs with low doses, the potential for irritation, and the unknown consequences of long-term nasal administration of drugs, adjuvants, and other formulation components.

The buccal and sublingual routes of administration permit rapid delivery to the systematic circulation. Absorption from the buccal and sublingual vasculature and lymphatics bypasses hepatic circulation and, thereby, reduces first-pass metabolism. The driving force of absorption is the high thermodynamic activity of the compounds. Organic nitrates and testosterone have been administered by these routes to produce rapid plasma concentrations and to minimize hepatic metabolism.

The sublingual administration of methyltestosterone was examined by Alkalay et al. (28) The sublingual tablet produced a 50% higher relative bioavailability when compared with the oral tablet or oral solution. The increased bioavailability was attributed to the avoidance of first-pass hepatic metabolism due to absorption from the sublingual vasculature and lymphatics.

The limitations of the buccal and sublingual routes include the requirement for successful candidates to possess high thermodynamic activities, the restrictive size of the buccal pouch and sublingual area, and the concern over the palatability and local irritation by the compound.

Transdermal administration can avoid first-pass metabolism as well as provide a large surface area for continuous-controlled administration of drugs with short biological half-lives and narrow therapeutic indices. The route has been used for nitroglycerin ointments, and transdermal therapeutical systems (patches) have been developed for scopolamine, nitroglycerin, clonidine, estradiol, and nicotine.

Scopolamine can cause side effects of dry mouth, drowsiness, tachycardia, central nervous system (CNS) disturbances, and amnesia if plasma concentrations exceed the levels required to alleviate motion sickness (Fig. 10). The application of scopolamine to the posterior auricular areas (the most permeable anatomical site) via a microporous patch permits constant delivery of 0.5 mg over 3 days to prevent motion sickness and avoid side effects.

The transdermal route suffers from the inability of the skin to deliver large doses, the relatively slow plasma increase when compared with other routes, and the potential for irritation. However, several products utilizing this route of administration are currently on the market. Absorption can be enhanced with molecules that maintain an appropriate hydrophilic-lipophilic balance for efficient passage through the barrier of the stratum corneum (16).

Rectal administration of systemic effect has traditionally been limited to clinical situations where oral intake is restricted due to a physiologic condition (e.g., vomiting) or to compounds that are irritating to the gastric mucosa. The

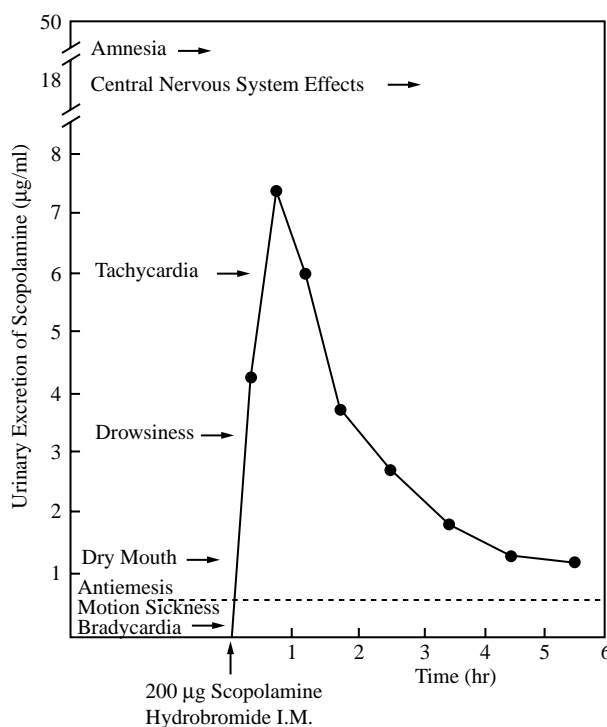


Fig. 10 Relationship between urinary excretion rate of scopolamine and its pharmacological effects; 10% of a parental dose of scopolamine is excreted in the urine unchanged.

route has been utilized for aspirin, acetaminophen, aminophylline, promethazine, perchlorpromazine, chlorpromazine, and indomethacin. The rectal administration of drugs is limited by increased interpatient variability and patient acceptability.

This section has reviewed some of the dosage form design methods available to the pharmaceutical scientist that have been shown to improve the therapeutic efficacy of certain drugs. The key to optimal dosage form design lies in the prerequisite understanding of the physicochemical and biopharmaceutical properties of the drug and the available routes of administration. Once the compound has been characterized and the problem has been defined, the technology is available to optimize bioavailability.

REFERENCES

1. Kaplan, S.A.; Jack, M.L. Utility of Bioavailability Studies in Drug Development. *Drug Dev. Ind. Pharm.* **1977**, *3*, 39–64.
2. Rowland, M.; Tozer, T.N. *Clinical Pharmacokinetics: Concepts and Applications*; Lea & Febiger: Philadelphia, 1980.

3. Haleblan, J.; McCrone, W. Pharmaceutical Applications of Polymorphism. *J. Pharm. Sci.* **1969**, *58*, 911–929.
4. Haleblan, J. Characterization of Habits and Crystalline Modification of Solids and Their Pharmaceutical Applications. *J. Pharm. Sci.* **1975**, *64*, 1269–1288.
5. Wall, G.M. Pharmaceutical Applications of Drug Crystal Studies. *Pharmaceutical Manufacturing* **1986**, *3*, 33–42.
6. Tawashii, R. Gastrointestinal Absorption of Two Polymorphic Forms of Aspirin. *J. Pharm. Pharmacol.* **1969**, *21*, 701–702.
7. Berge, S.M.; Bighley, L.D.; Monkhouse, D.C. Pharmaceutical Salts. *J. Pharm. Sci.* **1977**, *66*, 1–19.
8. Gould, P.L. Salt Selection for Basic Drugs. *Int. J. Pharm* **1986**, *33*, 201–217.
9. Anderson, B.D.; Flora, K.P. Preparation of Water-Soluble Compounds through Salt Formation. *The Practice of Medicinal Chemistry*; Wermuth, C.G., Ed.; Academic Press: London, 1996; 739–754.
10. Nelson, E. Solution Rate of Theophylline Salts and Effects from Oral Administration. *J. Am. Pharm. Assoc., Sci., Ed.* **1957**, *46*, 607–614.
11. Nishimura, K.; Sasahara, K.; Araj, M.; Nitani, T.; Ikegami, Y.; Morioka, T.; Nakajima, E. Dosage Form Design for Improvement of Bioavailability of Levodopa VI: Formulation of Effervescent Enteric-Coated Tablets. *J. Pharm. Sci.* **1984**, *73*, 942–946.
12. Stella, V.; Higuchi, T. Esters of Hydantoic Acids as Prodrugs of Hydantoins. *J. Pharm. Sci.* **1973**, *62*, 962–967.
13. Stella, V.; Higuchi, T.; Hussain, A.; Truelove, J. The Chemistry of a Novel 5, 5-Diphenylhydantoin Prodrug. *Prodrugs as Novel Drug Delivery Systems*; Higuchi, T., Stella, V., Eds.; American Chemical Society: Washington, DC, 1975.
14. Hussain, A.; Kulkarni, P.; Perrier, D. Prodrug Approaches to Enhancement of Physicochemical Properties of Drugs IX: Acetaminophen Prodrug. *J. Pharm. Sci.* **1978**, *67*, 545–546.
15. Hussain, A.; Truelove, J. Prodrug Approaches to Enhancement of Physicochemical Properties of Drugs IV: Novel Epinephrine Prodrug. *J. Pharm. Sci.* **1976**, *65*, 1510–1512.
16. Milosovich, S.; Hussain, A.; Dittert, L.; Aungst, B.; Hussain, M. Testosterone-4-dimethylaminobutyrate •HCl: A Prodrug with Improved Skin Penetration Rate. *J. Pharma. Sci.* **1993**, *82*, 227–228.
17. Hussain, A.; Yamasaki, M.; Truelove, J. Kinetics of Hydrolysis of Acylals of Aspirin: Hydrolysis of (1'-ethoxy) ethyl 2-acetoxybenzoate. *J. Pharm. Sci.* **1974**, *63*, 627–628.
18. Hussain, A.; Truelove, J.; Kostenbauder, H.B. Effect of Hydroxyl Group Substituents of Pyran Ring on Hydrolysis Rate of Benzoates: 2-Tetrahydropyranyl benzoate. *J. Pharm. Sci.* **1979**, *68*, 235–236.
19. Zaffaroni, A. Special Requirements for Hormone Releasing Intrauterine Devices. *Meeting on Pharmacological Models to Assess Toxicity and Side Effects of Fertility Regulating Agents*; Geneva, 1974; 423–434.
20. Aznar, R.; Giner, J. Development of the Intrauterine Progesterone-Releasing System. *Long-Acting Contraceptive Delivery System*; Zatzuchni, G.I., Goldsmith, A., Shelton, J.D., Sciarra, J.J., Eds.; Harper & Row: Philadelphia, 1983; 613–620.
21. Armaly, M.F.; Rao, K.R. The Effect of Pilocarpine Ocusert on Ocular Pressure. *Symposium on Ocular Therapy*; Leopold, I.H., Ed.; The C. V. Mosby Company: St. Louis, 1973; 80–94.
22. Armaly, M.F.; Rao, K.R. The Effect of Pilocarpine Ocusert with Different Release Rates on Ocular Pressure. *Investigative Ophthalmol.* **1973**, *12*, 491–493.
23. Lee, V.H.L. Ophthalmic Delivery of Peptides and Proteins. *Pharm. Technol.* **1987**, *11*, 26–38.
24. Stadol, N.S.®; *Physicians Desk Reference*, 54th Ed., Medical Economics Company: Montvale; NJ, 2000; 853–835.
25. Hussain, A.A. Intranasal Drug Delivery. *Adv. Drug Delivery Rev.* **1998**, *29*, 39–49.
26. Smith, R.; Personal Communication **1998**.
27. Hussain, A.; Faraj, J.; Aramaki, Y.; Truelove, J.E. Hydrolysis of Leucine Enkephalin in the Nasal Cavity of the Rat—A Possible Factor in the Low Bioavailability of Nasally Administered Peptides. *Biochem. Biophys. Res. Communi.* **1985**, *133*, 923–928.
28. Alkalay, D.; Khemani, L.; Wagner, W.E.; Bartlett, M.F. Sublingual and Oral Administration of Methyltestosterone. A Comparison of Drug Bioavailability. *J. Clin. Pharmacol.* **1973**, *13*, 142–151.

Corrosion of Pharmaceutical Equipment

Paul K. Whitcraft

Rolled Alloys, Temperance, Michigan, U.S.A.

INTRODUCTION

Corrosion, the degradation of a material's properties or mass over time because of environmental effects, is a costly reality that effects every industry. A study issued by the Federal Highway Administration (FHWA) in 2002 conservatively estimates the annual direct cost of corrosion in all U.S. industry sectors at US\$276 billion. Costs associated with corrosion include cathodic/anodic protection; coatings; inhibitors; corrosion-resistant alloys and materials; and maintenance, repair, and depreciation of equipment. Indirect costs, such as lost productivity, environmental or product contamination, planning and design, and lost opportunities, can easily outpace direct costs by factors of two or more.

It is relatively easy to recognize the impact that corrosion has on chemical process industries (CPI). The FHWA indicates that annual direct costs for the refinery, chemical, petrochemical, paper, and food processing industries total US\$13.5 billion. In the chemical and pharmaceutical sector alone, the conservative estimate given is that corrosion costs are roughly 8% of capital expenditures. The US\$1.7 billion figure for this sector is very conservative because it does not include operating and maintenance costs. Corrosion within these industries ranges from mild (exposure of structures to atmospherical conditions) to very severe (strong acids, high temperatures, and halogen environments). Additionally, these industries use large quantities of water for not only chemical process, but for heating and cooling. The properties of water, a very mild corrodent, make it conducive to the electrochemical nature of corrosion processes. Acceptable performance of corrosion-resistant materials or systems can range from tens of years to weeks.

The control of corrosion in pharmaceutical product processes is largely managed through the use of stainless steel. Rust-free surfaces and cleanliness issues to prevent product contamination have been the primary corrosion concerns. Resistance to mildly aggressive cleaning solutions and saline solutions and the potential for under deposit or crevice corrosion present the most severe service conditions. The high standards of cleanliness necessary for pharmaceutical processes favor the mitigation of corrosion.

CORROSION BASICS

The broadest definition of corrosion is the degradation of a material's properties or mass over time because of the effect of the environment. We can think of this in simpler terms by recognizing this process as the tendency for a material to return to its most thermodynamically stable state. For most metallic materials, this means the formation of oxides or sulfides, or other basic metallic compounds generally considered to be ores. For polymeric materials, the end result could be a variety of simple organic compounds. Only in vacuums or under inert atmospheres can corrosion processes be expected to halt entirely. In most cases, these processes are slow enough to afford useful and practical equipment life.

Corrosion is an electrochemical process and corrosion processes follow the basic laws of thermodynamics. Under controlled conditions, corrosion can be measured, repeated, and predicted. However, because corrosion takes place on an atomic level, corrosion can take place in an accelerated localized fashion, appear as uniform visible attack, or result in subsurface microscopical damage. Normal service environments can rapidly complicate these processes and mechanisms with such variables as pH, temperature, stress, surface finish, flow rates, etc. With the wide range of variables that can come into play, it should not be surprising that corrosion appears to be unpredictable at times.

FORMS OF CORROSION

Corrosion of metallic materials can take on many forms. Understanding and recognizing the basic forms of corrosion are necessary for developing a strategy for mitigation. These concepts can also be utilized in assessing nonmetallic materials. Fig. 1 illustrates these basic corrosion mechanisms.

Uniform Corrosion

The simplest form of corrosion is "uniform" or "general" corrosion. This mode of corrosion is characterized by a uniform metal loss over the entire exposed surface. It is the most predictable and measurable form of corrosion.



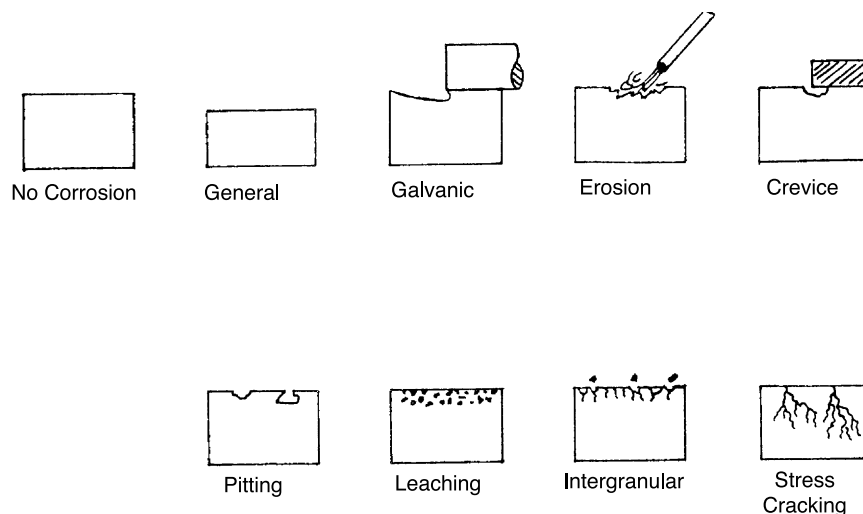


Fig. 1 Forms of corrosion.

There are many sources of general corrosion data that list actual or typical rates of corrosion for many materials in common environments. Where data are not available, simple laboratory tests can be conducted to simulate process environments.

The most common way to report uniform corrosion is in terms of metal thickness loss per unit of time, such as inches per year or millimeters per year. Because uniform corrosion is predictable, even moderately high corrosion rates can be tolerated provided a suitable monitoring and inspection system is utilized. For most chemical process systems, general corrosion rates of less than 2 mils per year (MPY) are acceptable. Rates between 2 and 20 MPY (1 mil = 0.001 in.) are routinely accepted as useful engineering materials. In severe environments, rates between 20 and 50 MPY may be economically justified. Rates exceeding 50 MPY are generally not acceptable.

For pharmaceutical applications in particular, the rate of corrosion must be considered not only for the effect on equipment, but where the lost metal has gone. Product contamination can be of concern for other products as well, and when this is an issue, the selection of more corrosion-resistant materials is of utmost importance.

Galvanic Corrosion

When two different metallic materials are electrically connected and placed in a conductive solution, an electrical potential will exist. This potential difference will provide a stronger driving force for the dissolution of the less noble (more electrically negative) material. It will also reduce the tendency for the more noble material to dissolve.

Although the relative differences in potential will change from one environment to another, they remain fundamentally the same because the potential is related to the energy required to oxidize them to metal ions in the given environment. The precious metals of gold and platinum are at the high potential (more noble or cathodic) end of the series, whereas zinc and magnesium are at the low potential (less noble or anodic) end. A galvanic series for a variety of metals in seawater is shown in Table 1.

Erosion Corrosion

Erosion corrosion results in an increased rate of corrosion attack attributable to the velocity of a corrodent over the exposed surface. The movement of the corrodent can be

Table 1 Galvanic series in seawater

Least Noble	Magnesium	
	Beryllium	
	Aluminum	
	Cadmium	
	Low Alloy Steel	
	Aluminum Bronze	
	Copper	
	Admiralty Brass	
	Martensitic Stainless	
	90-10 Copper-Nickel	
	Ferritic Stainless	
	70-30 Copper-Nickel	
	Nickel Alloy 200	
	Austenitic Stainless	
	Titanium	
	Superaustenitic Stainless	
Most Noble	Platinum	

When Active

associated with mechanical wear. The increased corrosion is usually related to the removal or damage of a protective surface film. The mechanism is usually identified by localized corrosion, which exhibits a pattern that follows the flow of the corrodent.

Fretting corrosion is a specialized form of erosion corrosion where two metal surfaces are in contact and experience very slight relative motion that causes damage to one or both surfaces. Again, in the presence of a corrodent, the movement causes mechanical damage of the protective film, leading to localized corrosion.

A second form of erosion corrosion is the case of cavitation. A type of corrosion familiar to pump impellers, this form of attack is caused by the formation and collapse of tiny vapor bubbles near a metallic surface in the presence of a corrodent. The protective surface film is again damaged, in this case by the high pressures caused by the collapse of the bubbles.

Pitting Corrosion

Pitting corrosion is in itself a corrosion mechanism, but is also a form of corrosion often associated with other types of corrosion mechanisms. It is characterized by a highly localized loss of metal. In the extreme case, a pit can appear as deep, tiny hole in an otherwise unaffected surface.

The initiation of a pit is associated with the breakdown of the protective film on the metal surface. In cases where pit depths increase rapidly, the environment is usually such that no repair or repassivation of the protective layer can be accomplished. In situations where many shallow pits form, the environment is usually one where repassivation of the damaged film can be made, but initiation of new sites occurs on a regular basis.

The localized nature of pitting attack can be associated with component geometry, the mechanics of the corrosion process, compositional inhomogeneity, or imperfection within the material itself. The growth of pits, once initiated, is closely related to another corrosion mechanism, crevice corrosion.

Crevice Corrosion

Crevice corrosion occurs in some environments because the nature of the environment within the crevice becomes more aggressive over time. There is little movement of the corrodent within a crevice. Over time, small changes in chemistry because of minor localized corrosion may become magnified because the solution is not being replenished by the bulk solution.

As a result of a slow initial rate of the corrosion, the pH of the crevice environment may become more acidic, or detrimental ion species may concentrate. As a result of

the low-flow condition, the crevice region may become depleted of oxygen, or preclude the replacement of reacted inhibitors.

Selective Leaching

Selective leaching is the process whereby a specific element is removed from an alloy because of an electrochemical interaction with the environment. Dezincification of brass alloys is the most familiar example of this type of corrosion. It occurs most commonly when there is exposure to soft waters and can be accelerated by high carbon dioxide concentrations and the presence of chloride ions.

The result of this corrosion is the formation of a porous and usually brittle shadow of the original component. Other alloy systems are susceptible to this form of corrosion. Examples include the selective loss of aluminum in aluminum–copper alloys, and the loss of iron in cast iron–carbon steels.

Intergranular Corrosion

As the name suggests, this particular corrosion mechanism attacks those sites where individual grains within a metallic material touch each other. These boundaries are natural regions of higher energy because of the greater frequency of dislocations of atoms from the natural order of the material's structure. In addition, these regions also tend to act as sites for the formation of secondary phases, which are essentially small islands within the matrix that have a chemical composition different from the alloy itself.

Depending on the corrodent and the alloy system, corrosion attack may initiate at these locations because of preferential attack of the secondary phase itself, or attack the surrounding matrix, which was locally dealloyed in forming the secondary phase. Either mechanism will result in the metallic surface being etched along the grain boundaries. As the attack progresses, individual grains are separated from the matrix and the surface layer becomes porous. In severe cases, the surface texture becomes grainy or powdery, leading to more rapid metal loss.

Stress Corrosion Cracking

The mechanism of stress corrosion cracking (SCC) is specific to certain alloys (or alloy systems) in specific environments. It is characterized by one or more crack fronts, which have developed as a result of a combination of the particular corrodent and tensile stresses.

Depending on the alloy system and corrodent combination, the cracking can be either intergranular or transgranular. The rate of crack propagation can vary greatly



and is affected by stress levels, temperature, and the concentration of the corrodent. In some severe combinations, such as type 304 stainless steel in a boiling magnesium chloride solution, extensive cracking can be generated in a matter of hours. In most industrial applications, the progress of SCC is, fortunately, at a much slower pace. However, because of the nature of the cracking, it is difficult to detect until extensive corrosion has already developed, which can lead to unexpected catastrophic failure. Alloy system and corrodent combinations that are known to exhibit SCC are fairly well documented and should be considered in initial design stages.

Apart from the SCC mechanism, stress can assist in other corrosion processes. Because this stress-assisted corrosion is related to tensile stresses, it is logical to expect that it will also accelerate the simple mechanical fatigue process. Corrosion fatigue is often difficult to differentiate from simple mechanical fatigue, but is recognized as a factor when the environment has been judged to have accelerated the normal fatigue process. Such systems can also have the effect of lowering the endurance limit such that fatigue will take place at a stress level wherein, without the environmental effect, fatigue failures would not be expected.

CORROSION MONITORING

The most common method of identifying and monitoring corrosion is visual inspection. Evidence of leakage, staining, or a change in surface appearance can be an indication that some type of corrosion is taking place. Experience with certain types of equipment and processes may help dictate inspection intervals and areas on which to focus the inspection. Records of vessel operation, maintenance, and repair can be helpful in establishing a pattern of performance that will improve predictability and minimize down time.

In areas where general corrosion is the expected form, a simple ultrasonic thickness gage can be utilized to determine the extent of corrosion, based on baseline readings made at installation or previous inspections. The entire unit need not be examined. Attention can be focused on those areas most likely to corrode, such as liquid levels, mixing zones, or areas of high turbulence. Corrosion probes, which can be placed in process equipment or pipelines, can monitor corrosion conditions by measuring an actual corrosion current, or other process parameters known to be related to general corrosion rates. These data can be constantly monitored and recorded to predict equipment wear, or as an alert to upset conditions.

For detection of more localized corrosion, such as crevice corrosion or SCC, other ultrasonic inspection

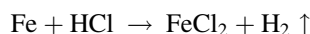
techniques may be useful. Baseline data generated at the time of installation will also be helpful in evaluating results. One benefit derived from this type of inspection technique is that it can often be conducted with little or no interference with production. Periodic planned visual inspection of equipment utilized under conditions likely to cause stress cracking is also an effective technique, especially when combined with nondestructive inspection techniques such as dye penetrant inspection. It may be necessary to remove coatings or insulation from the equipment surface to facilitate inspection.

CORROSION PROCESSES AND MITIGATION

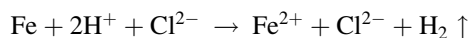
Electrochemical Nature of Corrosion

Corrosion, in its simplest definition, is the process of a material returning to the natural thermodynamic state. For most metallic materials, this means the formation of the oxides or sulfides that existed before being refined into useful engineering materials.

These changes are electrochemical reactions that follow the laws of thermodynamics. This concept aids in understanding why corrosion processes are time-dependent and temperature-dependent, and its application will indicate ways to mitigate corrosion. Corrosion reactions and rates are affected by ion and corrodent concentrations. One of the most basic corrosion reactions involves the oxidation of a pure metal when exposed to a strong acid. A familiar case is that of placing pure iron in hydrochloric acid. The resulting chemical reaction is quite obvious, with the solution beginning to bubble violently. The chemical reaction can be expressed as follows:



The result of this reaction is evidenced by the gradual disappearance of the iron and the hydrogen bubbles rising rapidly to the surface. On an electrochemical level, there is also an exchange of electrons taking place:



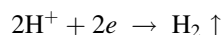
The iron has been converted to an iron ion by giving up two electrons (oxidation), which were picked up by the hydrogen ions. By gaining electrons, the hydrogen ion was *reduced* and formed hydrogen gas. Note that the chlorine atom does not enter into the reaction itself. The transfer of electrons takes place on the metal's surface. Those locations where electrons are being given up are identified as "anodes." The sites where electrons are being absorbed are denoted as "cathodes." A difference in electrical potential exists between these two areas and

a complete electrical circuit is developed. Negatively charged electrons flow in the direction of anode to cathode, and positively charged hydrogen ions in the solution move toward the cathode to complete the circuit. The faster the dissolution of the metal (rate of corrosion) is, the higher is the current flow. On a microscopical level, the sites of the anodes and cathodes can change locations on the surface. In fact, this is exactly what happens when general corrosion takes place, with the anodic areas moving uniformly over the metal's surface.

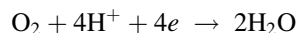
Anodic reactions in metallic corrosion are relatively simple. The reactions are always such that the metal is oxidized to a higher valence state. During general corrosion, this will result in the formation of metallic ions of all the alloying elements. Metals that are capable of exhibiting multiple valence states may go through several stages of oxidation during the corrosion process.

Cathodic reactions are more difficult to predict, but can be categorized into one of five different types of reduction reactions:

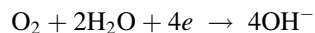
Hydrogen evolution



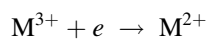
Oxygen reduction in acids



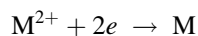
Oxygen reduction–neutral solutions



Metal ion reduction



Metal Deposition



Cell Potentials

Understanding electrochemical behavior and the possible reactions can help in predicting the possibility and extent of corrosion. A reaction will only occur if there is a negative free energy change (ΔG). For electrochemical reactions, the free energy change is calculated from:

$$\Delta G = -nFE$$

where n is the number of electrons, F is Faraday's constant, and E is the cell potential.

Therefore for a given reaction to take place, the cell potential must be positive. The cell potential is taken as the difference between the two half-cell reactions, the

one at the cathode minus the one at the anode. The half-cell potential exists because of the difference in the neutral state compared to the oxidized state, such as Fe/Fe^{2+} ; or, at the cathode, the difference between the neutral state and the reduced state, as in H^+/H_2 . These reduction–oxidation (redox) potentials are measured relative to a standard half-cell potential. The chart shown in Table 2 lists potentials relative to the H^+/H_2 , which is set as zero.

The larger this potential difference is, the greater is the driving force for the reaction. Whether corrosion does occur, and at what rate, is dependent on other factors. For corrosion to occur, there must be a current flow and a completed circuit, which is then governed by Ohm's law: $I = E/R$. The cell potential calculated here represents the peak value for the case of two independent reactions. If the resistance were infinite, the cell potential would remain as calculated but there would be no corrosion at all. The resistance in the circuit is dependent on a number of factors, including the resistivity of the media, surface films, and the metal itself. As current begins to flow, the potentials of both half-cell reactions move slightly toward each other. This change in potential is called polarization.

Once the corrosion current has been determined, the corrosion current density can be calculated by determining the surface area. However, polarization data can be more useful than just estimating corrosion rates. The extent of polarization can help predict the type and the severity of corrosion. As polarization increases, corrosion decreases. Polarization may be preferential to either the cathodic or anodic reactions. Understanding the influence of environmental changes on polarization can offer insights to controlling corrosion. For example, in the iron–hydro-

Table 2 Standard oxidation–reduction potentials

25°C, volts vs. hydrogen electrode	
$\text{Au} \leftrightarrow \text{Au}^{3+} + 3e$	1.498
$\text{O}_2 + 4\text{H}^+ + 4e \leftrightarrow 2\text{H}_2\text{O}$	1.229
$\text{Pt} \leftrightarrow \text{Pt}^{2+} + 2e$	1.2
$\text{Ag} \leftrightarrow \text{Ag}^+ + e$	0.799
$\text{Fe}^{3+} + e \leftrightarrow \text{Fe}^{2+}$	0.771
$\text{O}_2 + 2\text{H}_2\text{O} + 4e \leftrightarrow 4\text{OH}^-$	0.401
$\text{Cu} \leftrightarrow \text{Cu}^{2+} + 2e$	0.337
$2\text{H}^+ + 2e \leftrightarrow \text{H}_2$	0.000
$\text{Ni} \leftrightarrow \text{Ni}^{2+} + 2e$	−0.250
$\text{Fe} \leftrightarrow \text{Fe}^{2+} + 2e$	−0.440
$\text{Cr} \leftrightarrow \text{Cr}^{3+} + 3e$	−0.744
$\text{Zn} \leftrightarrow \text{Zn}^{2+} + 2e$	−0.763
$\text{Al} \leftrightarrow \text{Al}^{3+} + 3e$	−1.662
$\text{Mg} \leftrightarrow \text{Mg}^{2+} + 2e$	−2.363

Source: A. J. de Bethune and N. A. S. Loud, Standard Aqueous Electrode Potentials and Temperature Coefficients at 25°C, Clifford A. Hampel, Skokie, IL, 1964.



chloric acid example, hydrogen gas formation at the cathode can actually slow the reaction (increased circuit resistance) by blocking the access of hydrogen ions to the cathode site. This results in cathodic polarization and lowers the current flow and corrosion rate. If oxygen is bubbled through the solution, the hydrogen will be removed more rapidly by combining to form water and the corrosion rate increases significantly. Although this is an oversimplified view of the effects of oxygen, it does indicate that the degree of polarization can be affected by changes in the environment, either natural or induced.

Polarization

There are three basic causes of polarization. They are termed activation, concentration, and potential drop. Potential drop is the change in voltage associated with effects of the environment and the circuit between the anode and cathode sites. It includes the effects of the resistivity of the media, surface films, corrosion products, etc.

Activation polarization is because of a rate-controlling step within the corrosion reaction(s) at either the cathode or anode sites. An example of this can be seen with the H^+/H_2 conversion reaction. The first step of this process, $2H^+ + 2e^- \rightarrow 2H$, takes place at a rapid pace. The second part of this reaction, $2H \rightarrow H_2$, occurs more slowly and can become a rate-controlling factor.

Concentration polarization is the effect resulting from the excess of a species, which impedes the corrosion process, or from the depletion of a species critical to the progression of the corrosion process. The earlier case with an excess concentration of hydrogen gas impeding the rate of reaction is an example of concentration polarization. Although, in this case, it occurred at the cathode, it can also develop at the anode.

Measuring Polarization

Although polarization always leads to lower rates of corrosion, identifying the effects of the environment on polarization of the corrosion circuit is useful in predicting corrosion behavior. It is possible to measure the corrosion current while the corrosion potential is varied.

Most often, it is the anodic polarization behavior that is useful in understanding alloy systems in various environments. Anodic polarization tests can be conducted with relatively simple equipment and the scans themselves can be done in a short period of time. They are extremely useful in studying the active-passive behavior that many materials exhibit. As the name suggests, these materials can exhibit both a highly corrosion-resistant behavior or that of a material that corrodes actively, while in the same corrodent. Metals that commonly exhibit this type of

behavior include iron, titanium, aluminum, chromium, and nickel. Alloys of these materials are also subject to this type of behavior.

Active-passive behavior is dependent on the material-corrodent combination and is a function of the anodic or cathodic polarization effects, which occur in that specific combination. In most situations where active-passive behavior occurs, there is a thin layer at the metal surface that is more resistant to the environment than the underlying metal. In stainless steels, this layer is composed of various chromium and/or nickel oxides, which exhibit substantially different electrochemical characteristics than the underlying alloy. If this resistant, or passive, layer is damaged while in an aggressive environment, active corrosion of the freshly exposed surface will occur. The damage to this layer can be either mechanical or electrochemical in nature.

The behavior of iron in nitric acid underscores the importance of recognizing the nature of passivity. Iron is resistant to corrosion in nitric acid at concentrations around 70%. Once passivated under these conditions, it can also exhibit low rates of corrosion as the nitric acid is diluted. However, if this passive film is disturbed, rapid corrosion will begin and repassivation is not possible until the nitric acid concentration is raised to a sufficient level.

Anodic Polarization

Active-passive behavior is schematically represented by the anodic polarization curve shown in Fig. 2. Starting at the base of the plot, the curve starts out with a gradually increasing current as expected. However, at point A, there is a dramatic polarizing effect, which drops the current to a point where corrosion is essentially halted. As the

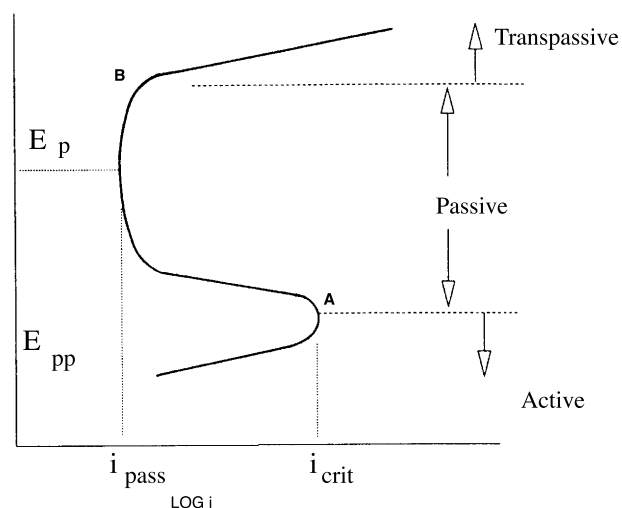


Fig. 2 Anodic polarization curve for a material exhibiting active-passive behavior.

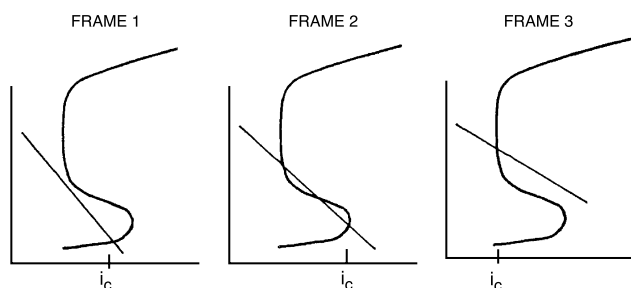


Fig. 3 Conditions within the corrosive environment can alter the cathodic polarization curve to create fluctuations between passive and active behavior.

potential is increased further, there is little change in current flow until the next critical stage, point B, where a breakdown of the passive film occurs, and the corrosion current again begins to rise.

Even with an established anodic polarization behavior, the performance of a material can vary greatly with relatively minor changes in the corrodent. This is also illustrated in Fig. 3. Frame 1 illustrates the case where the anodic and cathodic polarization curves intersect much as in materials with no active-passive behavior. The anode is actively corroding at a high, but predictable, rate.

Frame 2 represents the condition often found perplexing when using materials that exhibit active-passive behavior. With relatively minor changes within the system, the corrosion current could be very low as when the material is in the passive state, or very high when active corrosion begins.

Frame 3 typifies the condition sought after when using materials in the passive state. In this example, the cathodic polarization curve intersects only in the passive region, resulting in a stable and low corrosion current. This type of system can tolerate moderate upset conditions without the onset of accelerated corrosion.

CORROSION MITIGATION

The basic principles outlined here can be applied to identified corrosion problems and can provide solutions or alternatives. Corrosion control in many forms and approaches is founded on these concepts.

The principle of cathodic or “sacrificial” protection is founded in the natural potential differentials between different metals. Zinc anodes are intentionally placed in electrical contact with steel structures so that, as they corrode, the steel is protected. In other systems, a current may be applied to the structure to be protected so as to cause the current to flow to an artificial anode.

For similar reasons, it is desirable to build process systems out of the same materials. In systems where contact of dissimilar metals cannot be avoided, it is helpful to have the less noble material possess the largest surface area. By doing so, the corrosion current that is generated is distributed over a much greater area and slows the overall rate of penetration. In many such systems, it is also possible to electrically insulate one alloy network from the other.

Anodic protection finds its basis in the understanding of active-passive behavior. By increasing the potential of the component to be protected, it moves from an actively corroding situation to one where passivity can be induced. Such techniques can be quite cost-effective, but must be applied under well-controlled operating conditions because slight overprotection or underprotection can lead to accelerated rates of corrosion.

The types and varieties of inhibition systems are quite diverse, but also derive their fundamental logic from the principles reviewed here. Inhibitors slow corrosion by increasing polarization at either the anodic or cathodic reactions, or by increasing the electrical resistance of the media.

As an alternative to controlling the corrosion processes, the environmental conditions under which the system is operating may lend themselves to control. Temperature can have a significant influence on the corrosion process. This is not surprising because it is an electrochemical reaction and reaction rates do increase with increasing temperature.

If reducing the temperature is not possible, then it may be possible to eliminate one or more corrodents present, which are not critical to the product process. Halogen ions can be particularly aggressive on many materials and are often present only as a contaminant. Careful control of the pH can also be employed in many processes. There are additional environmental influences on corrosion other than the corrodent itself.

The relative velocities between the component and the media can have a direct effect on the corrosion rate. In some instances, increasing the velocity of the corrodent over the surface of the metal will increase the corrosion rate. When concentration polarization occurs, the increased velocity of the media will disperse the concentrating species. However, with passive materials, increasing the velocity can actually result in lower corrosion rates.

The surface finish of the component also has an impact on the mode and severity of the corrosion that can occur. Rough surfaces or tight crevices can facilitate the formation of concentration cells. Surface cleanliness can also be an issue with deposits or films acting as initiation sites. Biological growths can behave as deposits, or can change the underlying surface chemistry to promote corrosion.

One of the most common and cost-effective methods of preventing harmful corrosion is the selection of more



corrosion-resistant materials. In addition to selecting the proper material, fabricating and finishing the equipment can have critical impact on performance. Variations within the metal surface on a microscopical level influence the corrosion process. Microstructural differences such as secondary phases or grain orientation will affect the way corrosion manifests itself. For corrosive environments where grain boundaries are attacked, the grain size of the material plays a significant role in how rapidly the material's properties deteriorate. Chemistry variations in the matrix of weld deposits are also factors that must be considered in some corrosive environments.

The corrosion engineer can play a major role in system design, material selection, process or environmental control, and remediation. The focus of these efforts should not necessarily be the complete elimination or avoidance of corrosion, but rather the selection of the most cost-effective means of corrosion control and abatement.

MATERIAL SELECTION

The range and the availability of various types of engineering materials of construction continue to grow. As these materials improve, the benefits appear not only in terms of equipment reliability, but also processes pushed to more extreme conditions. These materials of construction can be generally classified as ferrous metals, nonferrous metals, plastics, elastomers, and other nonmetallics.

Ferrous Metals

This group of materials encompasses some of the most widely used construction materials as it includes steels and stainless steels. In pharmaceutical process applications, low-alloy steels are typically limited to structural uses and see little contact with product. Mild steels have been utilized successfully in stills, provided the oxygen content is maintained at very low levels. Tool and die steels may be used in areas where the product is dry and no corrosion is anticipated.

Steel and Stainless Steel

In areas where steels are in contact with the product or raw materials, stainless steels are often used as the workhorse. The production of stainless steel began in the early 1900s. The original efforts in this area were presumably based on the observation that chromium-plated steels parts were highly corrosion-resistant. The end result was the introduction of the ferritic *family* of stainless steels. The first documentation of the development of this class of steel began to appear in the 1920s. The first American Society

for Testing and Materials (ASTM) Specifications for stainless steels were published in 1935.

Today, there are seven basic families of stainless steels with compositions that contain 11–33% chromium, 0–38% nickel, and 0–7% molybdenum as the major alloying elements. These families are:

- Ferritic
- Austenitic
- Precipitation Hardenable
- Superferritic
- Martensitic
- Duplex (Ferritic-austenitic)
- Superaustenitic

Chromium is a metal that readily forms an oxide, which is transparent and happens to be extremely resistant to further degradation. As a further benefit to alloying with steel, it is less noble than iron and thus tends to form its oxide first. For exposure to mild, wet environments, the addition of about 11% chromium is sufficient to prevent “rusting” of steel components, hence the term “stainless.”

Ferritic stainless steels are magnetic, have body-centered cubic atomic structures, and possess mechanical properties similar to carbon steel, although less ductile. These materials are historically known as “400” series stainless as they were identified with numbers beginning with 400 when the American Institute for Iron and Steel (AISI) had the authority to designate alloy compositions. Alloy identification is now formally handled by the Unified Numbering System (UNS), where stainless alloy identification numbers generally begin with “S” followed by a five-digit number. Ferritic stainless steels find little application in the pharmaceutical industry owing to their low level of corrosion resistance and strength.

The corrosion resistance of the martensitic steels is again dependent solely on chromium, and because the carbon contents are generally higher than the ferritic alloys, they are less corrosion-resistant. Nevertheless, the combination of useful corrosion resistance in mild environments coupled with high strengths makes the martensitic stainless alloys candidates for cutting tools, molds, gears, shafts, and ball bearings.

Austenitic Stainless Steels

This family of stainless accounts for the widest usage of all the stainless steels. These materials are nonmagnetic, have face-centered cubic structures, and possess mechanical properties similar to the mild steels, but with better formability. The AISI designation system identified the most common of these alloys with numbers beginning with 300 and resulted in the term *300 series* stainless. Table 3 lists the chemical analyses of some standard

Table 3 Typical chemical compositions of some common stainless and nickel alloys

Alloy	Family	C	Cr	Ni	Mo	N	Nb	Other
430	Ferritic	0.07	16.5	—	—	—	—	
304 L	Austenitic	0.025	18.2	8.6	—	0.04	—	
316 L	Austenitic	0.025	16.5	10.5	2.2	0.04	—	
410	Martensitic	0.1	12.0	—	—	—	—	
17-4 PH	Precipitation-hardenable	0.04	16.2	4.2	—	—	0.25	3.6 Cu
Alloy 20	Superaustenitic	0.02	19.5	33.0	2.2	—	0.5	3.2 Cu
AL-6XN [®]	Superaustenitic	0.02	20.5	25.0	6.1	0.22	—	
Sea-Cure [®]	Superferritic	0.02	26.0	2.5	3.0	0.025	0.3	0.3 Ti
2205	Duplex	0.02	23.0	5.5	3.0	0.18	—	
C-276	Nickel alloy	0.01	16.0	57.0	16.0	—	—	3.5 W
625	Nickel alloy	0.02	22.0	61.0	8.9	—	3.6	0.25 Ti

AL-6XN is a registered trademark of Allegheny Ludlum.

Sea-Cure is a registered trademark of Colt Industries.

austenitic stainless steels and compares them to a few materials from other families of materials.

Once the corrosion resistance plateau in ferritic alloys of 18% chromium is reached, the addition of about 8% nickel is required to cause a transition from ferritic to austenitic. The primary benefit of this alloy addition is to achieve the austenitic structure, which, relative to the ferritics, is very tough, formable, and weldable. The added benefit, of course, is the improved corrosion resistance to mild corrodents. This includes adequate resistance to most foods, a wide range of organic chemicals, mild inorganic chemicals, and most natural environmental corrosion.

Nickel is used judiciously as an alloying element because its cost is substantially higher than chromium. However, type 304 is balanced near the austenite–ferrite boundary for another reason. Compositions similar to type 304 that are unable to form ferrite when solidifying after welding are prone to cracking during solidification and are more difficult to hot work. As a result, adding more nickel to the 18–8 composition offers little benefit from a corrosion standpoint and would be detrimental in other regards.

The next major step in alloying additions comes from molybdenum. This element also provides excellent corrosion resistance in oxidizing environments, particularly in aqueous corrosion. It participates in strengthening the passive film, which forms on the stainless steel surface along with chromium and nickel. A significant benefit is realized with the addition of only about 2% molybdenum. Added directly to the 18–8 composition, the alloy would contain too much ferrite so it must be rebalanced. The resulting chemistry is roughly 16% chromium, 10% nickel, and 2% molybdenum, and is recognized as type 316 stainless.

Austenitic alloys also make use of the concept of stabilization. Stainless types 321 and 347 are versions of

type 304 stabilized with titanium and niobium, respectively. These elements will preferentially combine with carbon that comes out of solid solution during weld solidification. Rather than a loss of corrosion resistance associated with formation of harmful chromium carbides, the carbides of titanium and niobium are not detrimental to corrosion resistance. The austenitic family of stainless also prompted another approach to avoiding the effects of chromium carbide precipitation. Because the amount of chromium that precipitated was proportional to the carbon content, lowering the carbon could prevent sensitization. Maintaining the carbon content to below about 0.035% vs. the usual 0.08% maximum will avoid the precipitation of harmful levels of chromium carbide. This discovery, along with improvements in melting technology, resulted in the development of the low-carbon version of many of these alloys. When first introduced, Extra Low Carbon (ELC) grades (i.e., 304 L and 316 L) required premiums on pricing because of higher production costs. This differential has essentially disappeared in the face of modern argon–oxygen decarburization (AOD) furnaces.

Argon–oxygen decarburization furnaces, utilized as a final refining stage in melting, are designed to permit the bubbling of the molten steel with oxygen, which facilitates the removal of carbon and sulfur. During this process, the exposed surface of the melt is protected with an inert argon atmosphere. This arrangement also permits bubbling with nitrogen gas, which will dissolve as atomic nitrogen into the steel. Nitrogen acts in a fashion similar to carbon by pinning slip planes, thus leading to higher-strength materials.

Modern melting technology is also responsible for another trend in stainless metallurgy. At one time, the permissible chemistry ranges for alloying elements needed to be broad to accommodate inhomogeneity in electrical furnace melts, chemical analysis variations, and



Table 4 Typical mechanical properties of some common stainless and nickel alloys

Alloy	Condition	Tensile (KSI)	Yield (KSI)	Elongation (%)	Toughness (ft lb at °F)
430	Annealed	74	45	28	75 at RT
304 L	Annealed	75	28	55	80 at -300°
316 L	Annealed	75	28	55	110 at -300°
410	Full hard	188	155	17	49 at RT
17-4 PH	900° aged	198	183	15	16 at RT
Alloy 20	Annealed	90	48	45	145 at -300°
AL-6XN	Annealed	110	55	50	85 at -300°
Sea-Cure	Annealed	95	80	30	DBTT near -10°
2205	Annealed	98	70	25	80 at -40°
C-276	Annealed	110	55	50	240 at -40°
625	Annealed	120	65	50	35 at -320°

raw material quality. For example, the chromium range for type 304 was 8.0–10.0, and still heats were occasionally missed. With current technology, it is possible to maintain $\pm 3\sigma$ limits on chromium to 0.5% or better. The result is that alloys are currently being produced with 0.50–0.75% less of an alloying element than they were just 15 years ago.

Even with alloying additions such as molybdenum to improve localized corrosion resistance to halogens, the workhorse 304 L and 316 L alloys are susceptible to chloride SCC. This cracking mechanism manifests itself as branched, generally transgranular cracks that are so fine as to be virtually undetectable until it has progressed to catastrophic proportions. This mode of failure can occur when the austenitic alloy is under stress in the presence of halogen ions at temperatures above about 120°F. Studies by Copsen underscored the benefit of very low nickel contents, such as the ferritic stainless steels, or nickel levels in excess of about 20%. In fact, the nickel contents in these two alloys are in the range that tends to crack most quickly in chloride-bearing environments.

Duplex Stainless Steels

Stainless alloys that contain roughly equal amounts of austenite and ferrite are termed duplex stainless. This family of alloys grew out of one basic material originally identified as type 329. They are balanced to contain relatively high chromium contents, with only enough nickel and austenitizers to develop about 50% austenite.

These alloys offer several useful advantages. First, their general corrosion resistance is typically slightly above that of 316 L in most media. In addition, because the nickel content is held low, they offer very good resistance to chloride SCC. In combination with good corrosion resistance, duplex stainless alloys offer higher strengths than those typically found with austenitic steels.

Table 4 compares some typical mechanical properties for common stainless and nickel alloys.

Although more formable than the ferritic alloys, they are not as ductile as the austenitic family of alloys. Welding requires more care than with the austenitic alloys because of a greater tendency toward compositional segregation and sensitivity to weld heat input. Improper fabrication techniques can result in equipment that falls short of expectations for corrosion resistance and mechanical properties.

Superaustenitic

During the 1970s and into the 1980s, there was much attention focused on a family of stainless alloys, which came to be identified as superaustenitic. The foundation for the development of this class of materials was in the development of Carpenter no. 20 stainless, introduced in 1951. Consisting of 28% nickel and 19% chromium with additions of molybdenum and copper, this alloy was first produced as a cast material.

20Cb-3 Stainless became popular in the CPI as an intermediate step between type 316 stainless and the more highly alloyed nickel base materials. In particular, it was a cost-effective way to combat chloride SCC. Because of the high nickel content of 20Cb-3 Stainless, it received a nickel base alloy UNS designation as UNS N08020. However, because the major constituent is iron, it is truly a stainless steel. The superaustenitic term is derived from the fact this composition is so far from austenite–ferrite boundary that, unlike the 300 series stainless alloys, there is no chance of developing ferrite in this material.

The main approach to improving the pitting and crevice corrosion resistance of the basic 35% nickel, 19% chromium, and 2% molybdenum alloy was to increase the molybdenum content. Among the first of the newer alloys introduced was 904 L (UNS N08904), which boosted the molybdenum content to 4% and reduced the

nickel content to 25%. The reduction in nickel content was beneficial as a cost-saving factor, with minimal loss of general corrosion resistance and sufficient resistance to chloride SCC.

The next progression was to raise the molybdenum content to a higher level of 6% and to offset the tendency for the formation of sigma phase by the alloying addition of nitrogen. This concept was introduced with two alloys, 254SMO[®] (UNS S31254) and AL-6XN[®] (UNS N08367). The major benefit of the addition of nitrogen was the ability to produce these alloys in heavy product sections such as plate, bar, and forgings. An additional benefit was derived from alloying with nitrogen in terms of increased pitting resistance. A significant amount of work by a large body of researchers has demonstrated a relationship between pitting or crevice corrosion resistance and alloy content, which is approximated by:

$$\text{Cr}\% + 3.3*\text{Mo}\% + 16*\text{N}\%$$

where increasing values indicate increased resistance. A value in excess of approximately 33 is considered necessary for pitting and crevice resistance to ambient seawater.

Other Stainless

The precipitation-hardenable family of stainless alloys utilizes thermal treatment to intentionally precipitate phases, which cause a strengthening of the alloy. The precipitating phase is generated through an alloy addition of one or more of niobium, titanium, copper, molybdenum, or aluminum. The metallurgy is such that the material can be solution-treated (i.e., all alloying elements are in solid solution and the material is in its annealed or softest state). In this condition, the material can be machined, formed, and welded in the desired configuration. After fabrication, the unit is exposed to an elevated temperature cycle (aging), which precipitates the desired phases to cause an increase in mechanical properties.

As a class, these alloys offer high mechanical properties, although not as high as martensitic low-alloy steels, in combination with very useful corrosion resistance. On average, their general corrosion resistance is below that of type 304 stainless. The corrosion resistance of the PH 15-7 Mo and A-286 alloys approaches that of type 316. The martensitic and semiaustenitic pH grades are resistant to chloride cracking. These materials are susceptible to hydrogen embrittlement. In the pharmaceutical industry, these materials might find useful applications in valve components, bolting materials, or wear surfaces.

The ability of the ferritic alloys to resist chloride SCC is one of their most useful features in terms of corrosion resistance. During the 1970s, developmental efforts were directed at producing ferritic materials that could also

exhibit a high level of general and localized pitting resistance as well.

The first commercially significant alloy to meet this expectation was an alloy containing 26% chromium and 1% molybdenum. To obtain the desired corrosion resistance and acceptable fabrication characteristics, the material had to have very low interstitial element contents. To achieve these levels, the material was electron beam-refined under a vacuum, and was introduced as E-BRITE[®] Alloy. Carbon plus nitrogen contents were maintained at levels below 0.020%.

Materials such as SEA-CURE[®] (S43635) and 29-4C[®] Alloy (S44735) represent the most recent developments in superferritic materials. These alloys do exhibit excellent localized corrosion resistance. Although the superferritic materials alloyed with some nickel have improved mechanical toughness and are less sensitive to contamination from interstitial elements, their availability is still limited to heat exchanger tubing with wall thicknesses below about 0.100 in. This is related to the formation of embrittling phase during cooling from annealing temperatures. Section thicknesses over these levels cannot be cooled sufficiently fast to avoid a loss of toughness.

Cast Stainless Steel

The discussion thus far has been devoted to examining the different families of stainless steel metallurgy. The alloys discussed were wrought materials (i.e., materials that are hot-worked following being cast into ingots). The practice of hot working steels improves the uniformity of their chemical, mechanical, and corrosion-resistant properties. These materials are suited for fabrication by bending and welding.

Cast stainless steels can be divided into the same families as the wrought materials, except for the superferritics. Castings offer the particular advantage of being able to obtain complex shapes without extensive fabrication or machining. Cast alloys usually cost less per pound than the wrought counterpart because the hot working operations are avoided. Cast stainless steels can also have chemistry modifications to enhance properties that would otherwise render them unworkable as a wrought product. Heat-resistant cast alloys, such as HK, usually have high carbon and silicon contents, which improve elevated temperature strength considerably, but at the expense of room temperature toughness. The cast structure of such materials is also less resistant to thermal fatigue than the wrought material.

Although the compositions of the basic austenitic cast alloys are very similar to the wrought versions, the cast versions usually contain significant amounts of delta ferrite. As in the solidification of weld metal, ferrite is



beneficial in reducing the tendency for the material to form cracks during solidification. The ferrite content in CF-8M can approach 20% and can readily attract a magnet. Although high ferrite contents are often not of concern, the ferrite can be attacked preferentially in some environments such as urea, nitric acid, and hydrochloric acid.

The existence of significant amounts of ferrite is one form of segregation that can be encountered in cast stainless alloys. Because cooling rates are generally slow for cast components, other secondary phases can form. Chromium carbide precipitation is a particular concern for many of these materials and, under most circumstances, the casting should be solution-annealed prior to being placed in corrosive service.

NONFERROUS METALS

Nickel Alloys

Nickel is very effective in improving the corrosion resistance of stainless steels and is also utilized as the base material for a number of specialty alloys. By definition, nickel alloys contain more nickel than any other constituent. Alloy 625 has been widely used throughout the CPI for its corrosion resistance to strong acids, including hydrochloric acid, and for its resistance to halogen attack. This material also has excellent high-temperature mechanical properties and is used in gas turbine applications. Another nickel base alloy more familiar to the pharmaceutical industry is alloy C-276. This alloy combines the general corrosion resistance advantages of nickel with the benefits of oxidation and localized corrosion resistance derived from chromium and molybdenum.

Nickel alloys are also austenitic, nonmagnetic under all conditions, and possess formability similar to the austenitic stainless steels. The same welding techniques utilized for stainless alloys can be used for joining the nickel alloys.

Recent additions to the wide range of nickel alloys include materials designed to resist even more severe environments, such as hot halogenated acids, which are likely to induce crevice corrosion. Such materials include alloy 686, alloy 59, and alloy C-2000. These materials contain high levels of chromium (over 22%) and molybdenum (over 15%).

Other Nonferrous Materials

Other significant nonferrous materials include copper, aluminum, and titanium, and their alloys. Copper is resistant to most neutral waters, including seawater, and to strong reducing acids. However, it is not resistant to oxidizing acids, amines, metal salts, and sulfur compounds,

and is also sensitive to erosion from high velocities. Aluminum is useful in resisting atmospherical corrosion and neutral waters, but can corrode quickly in acids or bases. As such, copper and aluminum alloys are used sparingly in the CPI.

Titanium offers a much broader range of corrosion resistance and its applications are similar to those of the superaustenitic stainless steels and some nickel alloys. It has good resistance to pitting and crevice corrosion. Titanium alloys offer good general corrosion resistance to a variety of oxidizing and reducing acids. Although titanium alloys can be fabricated by welding, the required procedures are best suited for shops geared for this type of work. Field fabrication and repair are possible, but do require specialized equipment and procedures.

PLASTICS

This is a huge general category of materials, which includes both thermoplastics and thermosetting polymers. Tabular data on the corrosion resistance of these materials in a wide range of environments are available from a variety of sources. Commonly used materials of construction in the CPI include polyvinyl chloride (PVC and CPVC), polyethylene, polypropylene, polystyrene, polycarbonate, polytetrafluoroethylene (PTFE), fiberglass composite materials, and a variety of epoxies used for coatings or adhesives.

This class of materials can be a solution to handling a wide range of very aggressive chemicals, often resulting in a cost savings over more expensive metallic materials. However, some important differences between these materials and metals can limit their application. First, there is a maximum temperature limitation for these materials, which generally ranges from 120°F to 400°F. Many of these materials are limited to useful service temperatures below 200°F. Second, corrosion of these materials rarely occurs by direct material loss through a chemical reaction similar to that of metallic materials. The degradation of properties is usually the result of permeation by, or absorption of, the corrosive media. This attack can result in embrittlement, softening, blistering, crazing, swelling, dissolution, or some other loss of physical properties. Finally, the mechanical strength of this group of materials is generally lower than that of metal alloys. This can usually be accommodated through design, or with suitable exterior support structures.

Within the pharmaceutical industry, these materials are utilized for bulk acid storage at room temperature, storage of dry powders, piping of potable water, sewer lines, and wastewater treatment. The Food and Drug Administration (FDA) recognizes acceptable materials for food contact in the Code of Federal Regulations (CFR) 21 part 177.

ELASTOMERS

Elastomers or rubbers are also available in a wide variety of chemistries starting with natural rubber and including neoprene, urethane, polyester, silicone, and fluoroelastomers. These materials are subject to the same types of degradation as the thermoplastic materials. Temperature limitations are similar, or perhaps slightly lower.

The widest usage of these materials within the chemical process and pharmaceutical industries is for gasket and sealing applications, linings, flexible tubing, electrical insulation, and drive belts.

OTHER NONMETALLICS

Included in this category are materials such as concrete, chemical-resistant grouts, ceramics, and glass. These materials are highly resistant to chemical attack from most media normally encountered in the chemical process and pharmaceutical industries. In general, their usage is restricted as they tend to be brittle, have poor mechanical properties in tension, and are sensitive to thermal shock.

Chemical-resistant grouts and mortars and ceramic tile systems are very effective in providing cleanable surfaces, such as flooring and walls, and for containment. Ceramics are typically silicate-based materials, but there are also types produced from metallic oxides, nitrides, borides, and carbides. Although they generally have good corrosion resistance, they can be attacked by strong, hot alkalis and acidic fluoride media. The chemical-resistant grouts utilize inorganic binders, fillers, and a hardener. The exact formulation utilized must be based on the resistance to the environment to which it will be exposed.

Glassware and glass-lined equipment also offer excellent resistance to corrosion, similar to that of ceramic materials. In addition to the advantages of corrosion resistance, it also affords a very cleanable surface. Glass-lined vessels are used in the pharmaceutical industry for reactors, mixers, storage tanks, transfer piping, and high-purity water systems.

EQUIPMENT AND SERVICE CONSIDERATIONS

Welding

From an engineering standpoint, the ability to weld stainless steels with relative ease is a major advantage to their usefulness. Weld deposits, because they are cast structures, are subject to discussion regarding corrosion resistance similar to the cast materials. The chemistry of a weld deposit is likely to exhibit segregation and, depending on the alloy and the welding technique employed, may

develop deleterious secondary phases in either the weld or heat-affected zone.

Two ways to address this concern have already been discussed. These involve the reduction of carbon content to low levels and the use of stabilizers to prevent chromium depletion. Either of these methods is typically used for components that are assembled in the field, because subjecting the fabricated unit to an annealing treatment is neither practical or desirable in most instances.

In many cases, a small decrease in the corrosion resistance of weldments is tolerable. When the environment is particularly severe for the alloy being used, the weld may be attacked preferentially. This condition can be exaggerated by the area effects of the more noble base metal compared to the small weld zone. An alternate approach for field welding is to select a higher alloy welding consumable so the weld deposit is more noble than the base metal. Preferential attack can also occur in the heat-affected zone of the base metal. This is typical of weldments made in standard type 304 where the carbon content will lead to chromium carbide precipitation. Of course, this condition cannot be avoided by using a different filler metal and the only remedy is a postweld anneal.

Aside from the actual weld deposit chemistry, welding techniques can have an influence on corrosion resistance. First and foremost, the area to be joined should be clean and free of dirt and grease. Carbonaceous materials will contaminate the weld deposit and will deplete chromium from the alloy. For a similar reason, carbon arc gouging or cutting should be avoided. Contamination from other metals should also be avoided. Although free iron will essentially be melted into the weld deposit unnoticed, rust can affect weldability. At best, it can lead to lack of fusion or porosity and, at worst, it may act as a preferential site for the onset of corrosion. Joint preparation should be accomplished using properly sharpened tooling, and wire brushing should be performed using stainless steel brushes.

Low-melting-point metals are of particular concern. Molten copper, zinc, or aluminum will attack the grain boundaries of austenitic alloys preferentially. Copper alloy clamps or fixtures used to hold work, whereas welding has been known to leave smears of metal that have subsequently caused cracking. Zinc from galvanized steel or paint primers has also been known to contaminate weld joints.

Full-penetration weld joints should also be made. This is a good practice from a strength and fatigue resistance standpoint, but is also a factor in avoiding corrosion. Unfused joints are sites likely to trap corrosive products increasing the likelihood that oxygen or metal ion concentration cells are developed. This will generally mean that joints will have to be beveled if the thicknesses to be joined are in excess of 3/16 in. Beveling joints also insures that adequate filler metal will penetrate to the root in those instances where overalloying is desired.

Finally, the surface finish of the weld area should be similar to that of the base metal. Although a slightly higher roughness is usually unavoidable unless welding is followed by grinding to blend in the weld, minimizing roughness in the weld zone can be beneficial. Weld spatter should be removed by grinding. The weld slag from covered electrodes, which prevents oxidation of the metal during solidification, should be completely removed prior to making a second weld pass or placing the weld in service. Slag deposits on the surface will act as crevices in corrosive service. Removal of heat tints on the surface from bare wire or autogenous welding processes is preferred. In severe service, these areas may be attacked preferentially.

Passivation

Stainless steels offer useful resistance because they tend to exhibit passive corrosion behavior as a result of the formation of protective oxide films on the exposed surfaces. Under normal circumstances, stainless steels will readily form this protective layer immediately on exposure to oxygen. When this protective film is violated or fails to form, active corrosion can occur. Some fabrication processes can impede the reformation of this passive layer, and to insure that it is formed, stainless steels are subjected to "passivation" treatments.

The most common passivation treatments involve exposing the metal to an oxidizing acid. Nitric and nitric-hydrofluoric acid mixtures comprise the predominant usage in stainless steel production. The nitric-hydrofluoric acid mixtures are more aggressive and are typically used to remove the oxide scales formed during thermal treatment. This "pickling" process provides two benefits. First, it removes the oxide scale and passivates the underlying metal surface. Second, because of its aggressive nature, the process will remove any chromium-depleted layer that may have formed as a result of the scale formation.

For passivation treatments other than scale removal following thermal treatment, less aggressive acid solutions are usually employed. The primary purpose of these treatments is to remove contaminants that may be on the component's surface and could prevent the formation of the oxide layer locally. The most common contaminant is imbedded or free iron particle from forming or machining tools. Mechanical polishing can be employed to provide a uniform surface finish and to remove these contaminants. The polishing materials should be used for stainless only as they can carry over small particulates from one part to the next. In addition, the work-hardened state of this fine particulate, even from a stainless vessel, can have a lower threshold for corrosion and act as an initiation site if not removed. A dilute (10%) solution of nitric acid is effective at removing free iron, or similar contaminants. For ferritic, martensitic, or precipitation-hardening grades, a nit-

ric acid solution inhibited with sodium dichromate is used so as not to attack the stainless too aggressively. For the more resistant stainless alloys, phosphoric acid at 1% concentration and citric acid at 20% concentrations are also effective. Other commercially available chelating agents can be employed.

Electropolishing

Electropolishing is actually controlled corrosion, resulting in the uniform removal of the surface layer of metal. Electropolishing is not a passivation treatment, although the process does result in a passive surface. Proper electropolishing technique maintains the part in the electrochemically passive range whereas the passivated layer is only allowed to grow several atoms thick, at most. The electrolyte simultaneously promotes dissolution of this layer. The electropolishing process does remove surface impurities as is accomplished with passivation. During the cleaning and rinsing process following electropolishing, the material does passivate naturally on exposure to oxygen containing rinse water or air. No additional passivation procedure is required.

Electrolytes used for electropolishing are usually proprietary mixtures with contents that are not quantitatively revealed. Any electrolyte will typically have the ingredients to facilitate three different actions of the polishing process. These are: 1) a contaminant, which facilitates the breakdown of a passive film; 2) an oxidizer, which helps form a passivating film; and 3) a highly viscous constituent, which promotes the formation of a diffusion layer. The oxidizer and the contaminant assist in maintaining the part in a pseudo-passive state whereas the diffusion layer control is necessary to promote uniform metal loss. This is illustrated in the potentiodynamic graph shown in Fig. 4. For electropolishing 316 L stainless, the electrolyte will often contain perchloric acid, which can provide both oxidizing power and a halide contaminant. Acetic anhydride may be used to control the

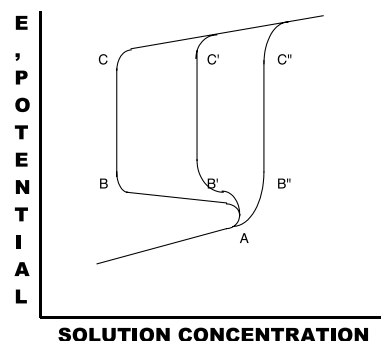


Fig. 4 Anodic polarization curve illustrating the unstable state created and maintained during electropolishing.

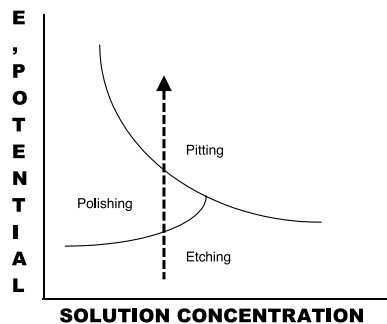


Fig. 5 Schematic diagram illustrating how, as electropolishing potential is increased, conditions may change from etching to polishing and, finally, to pitting.

diffusion layer. Stainless and higher nickel alloys might also use nitric acid, sulfuric acid, phosphoric acid, hydrogen peroxide, and methyl alcohol.

Determining the ideal electrolyte ingredients is an important part of the process. Because this is an electrochemical process, other variables to be controlled include the voltage, current density, and solution temperature. As illustrated in Fig. 5, voltages that are too low can cause etching of the surface because of more rapid general corrosion; too high a voltage will result in pitting. Cathode design is important because it is imperative to distribute the current as uniformly as possible over the part's surface.

Rouging

Rouging is a phenomenon of particular interest to the pharmaceutical industry. It is the presence of a surface layer of oxide found on stainless equipment or piping, typically handling high-purity water at temperatures above ambient. This includes stills, steam systems, purified water, and water for injection (WFI). The oxides can range in composition, degree of oxidation, color, texture, and adherence. Although generally shown to be innocuous, the mere presence of these deposits can raise concern.

The rouge itself is typically composed primarily of iron oxides or iron hydroxides, but because these are developing on stainless surfaces, they also contain oxides of chromium, nickel, and molybdenum as well. There are empirical data indicating that resistance to rouging increases with increasing Cr/Fe ratios in the passive layer, and/or the depth of the passive layer itself. Because electropolishing and passivation both increase the Cr/Fe ratio, application of these processes can increase resistance to rouging. Even with such treatments, the passive layer can break down because of the ionizing effect of high-purity water. The low oxygen content of these waters also slows the rate of repassivation and may cause the layer to linger in intermediate states of oxidation.

Repeated cycles of this process result in the entrapment of various oxides in the passive layer, hence the wide variation in colors.

Removal of rouging can be accomplished mechanically, but is usually addressed by chemical cleaning. Repassivation treatments with nitric, phosphoric, citric, or other oxidizing acid solutions have been effective in removing or fully reoxidizing this layer. As with any chemical reaction, the process is time-dependent and can be influenced by temperature. For more resistant rouge patterns, reducing acids such as hydrofluoric or hydrochloric may be used in combination with a passivation treatment. The use of these acids in strong concentrations may etch the surface.

Potentiodynamic polarization studies have been conducted to measure the efficacy of passivation treatments. It has been shown that the breakdown (pitting) potential is raised by passivation or electropolishing techniques, which result in higher Cr/Fe ratios and increased depth of the passive layer. These potentials can be increased by as much as 50–100 mV over mechanically polished or pickled surfaces, and have been equated to increased resistance to rouging. Additionally, breakdown potentials of the 6% molybdenum alloy N08367 were shown to be 400 mV higher prior to enhancing passivation treatments, and another 50 mV higher following such treatments. Such studies would suggest that higher alloys such as N08367 or the C-276 type alloy are highly resistant to rouging.

SUMMARY

This discussion provides for a fundamental understanding of corrosion and corrosion processes. It also offers an overview of both metallic and nonmetallic materials utilized in the construction of pharmaceutical equipment and some of the special considerations in their application. Suppliers of materials and fabrication services are valuable resources of information and should be included in the design process. Because corrosion processes are often complex, the services of corrosion engineering professionals should be considered in the original design or performance analysis activities as part of an in-house team, or on a consulting basis.

REFERENCES

1. Brasunas, A. *NACE Basic Corrosion Course*; National Association of Corrosion Engineers: Houston, TX, 1970.
2. Dillon, C. *Corrosion Control in the Chemical Process Industries*; Material Technologies Institute for the Chemical Process Industries: St. Louis, MO, 1994.

3. Darken, L.; Gurry, R. *Physical Chemistry of Metals*; McGraw-Hill: New York, NY, 1953.
4. Fontana, M.; Greene, N. *Corrosion Engineering*; McGraw-Hill: New York, NY, 1967.
5. Gonzalez, M. Materials of Construction for Biopharmaceutical Systems. In *ASME Bio Process Seminar, San Juan, PR*; January, 2003.
6. Grant, A. CIP and Passivation. In *ASME Bio Process Seminar, San Juan, PR*; January, 2003.
7. Henthorne, M. *Corrosion, Causes and Control*; Chemical Engineering Magazine: New York, NY, 1971–1972.
8. McGannon, H. *The Making, Shaping and Treating of Steel*; United States Steel Corporation: Pittsburgh, PA, 1971.
9. Sedriks, A. *Corrosion of Stainless Steels*; John Wiley and Sons: New York, NY, 1979.
10. Schweitzer, P. *Corrosion and Corrosion Protection Handbook*; Marcel Dekker: New York, NY, 1996.
11. West, J. *Electrodeposition and Corrosion Processes*; Van Nostrand-Reinhold: London, UK, 1971.
12. *20Cb-3 Stainless Steel*; Carpenter Technology Corporation: Reading, PA, 1987.
13. *AL-6XN Alloy*; Allegheny-Ludlum Corporation: Pittsburgh, PA, 1991.
14. *Corrosion Costs and Preventative Strategies in the United States, A Review by Materials Performance, MP Supplement*; July, 2002.
15. *Mechanical and Physical Properties of the Austenitic Chromium–Nickel Stainless Steels at Elevated Temperatures*; The International Nickel Company: New York, NY, 1963.
16. *Metals Handbook*, 10th Ed.; ASM International: Materials Park, OH, 1990.
17. *Steel Castings Handbook*; Steel Founder's Society of America: Rocky River, OH, 1981.



Crystal Habit Changes and Dosage Form Performance

A. K. Tiwary

Punjabi University, Patiala, India

INTRODUCTION

A crystalline particle is characterized by definite external and internal structures. Habit describes the external shape of a crystal, whereas polymorphic state refers to the definite arrangement of molecules inside the crystal lattice. Crystallization is invariably employed as the final step for purification of a solid. Use of different solvents and processing conditions may alter the habit of recrystallized particles, besides modifying the polymorphic state of the solid. Subtle changes in crystal habit at this stage can lead to significant variation in raw material characteristics. Furthermore, various indices of dosage form performance such as particle orientation, flowability, packing, compaction, suspension stability, and dissolution can be altered even in the absence of significantly altered polymorphic state. These effects are a result of the physical effect of different crystal habits. In addition, changes in crystal habit accompanied with or without polymorphic transformation during processing or storage can lead to serious implications of physical stability in dosage forms. Therefore to minimize variations in raw material characteristics, to ensure reproducibility of results during preformulation, and to correctly judge the cause of instability and poor performance of a dosage form, it is essential to recognize the importance of changes in crystal surface appearance and habit of pharmaceutical powders. Fig. 1 depicts an overview of the critical stages where changes in crystal habit are likely to appear during pharmaceutical processing.

CRYSTAL HABIT CHANGES

The internal structure or polymorphic state represents the molecular arrangement within a crystal and is manifested in the form of a definite heat of fusion (ΔH_f) value. External structure or crystal habit is the outer description of a crystal and is described by its length, width, thickness, and surface appearance (roughness, smoothness, and porosity). Crystal growth may be impeded by adjacent crystals growing simultaneously or contacting container walls. As a result, the development of plane faces may be inhibited, leading to the formation of a tabular (moderate

development of parallel faces), platy (excessive development of parallel faces), prismatic, acicular (inhibited width), or bladed (flattened acicular) crystal habit.^[1,2]

Thus a single internal structure of a compound can have several different habits. In the case of delayed crystallization, an irregularly shaped crystal may be produced because it is constrained to occupy only the spaces left between already crystallized particles. Such irregularly shaped crystals are described as anhedral or allotriomorphic, and those bound by plane faces are known as euhedral or idiomorphic. However, anhedral crystals do have a regular arrangement of building blocks in the crystal lattice.

Crystallization of a solid may be influenced by formulation and process variables of the crystallization process. However, it is difficult to delineate the role of a process variable on crystal habit because an alteration of the variable often leads to change in both deposition and dissolution rate of the material during the crystal growth phase. In addition, the processes are interactive and not independent of each other. For example, a change in temperature simultaneously alters the viscosity of crystallization solvent as well as saturation level of the solute. Hence an increase in temperature affects both deposition and dissolution rate of the solute on crystal nuclei. Similarly, rate of stirring influences the onset of nuclei formation because of its effect on the temperature of the solution. It also influences the growth of crystals because of the uniform distribution of solute throughout the solution. The net result of an altered variable is usually dependent on more dominant factors such as initial supersaturation level, nature of the co-solvent, crystallizing solvent, cooling rate, etc. Few important factors that alter crystal habit and the anticipated influence of altered habit on dosage form performance are summarized in Table 1.

Changes in polymorphic state as a result of the employment of different crystallization conditions and the influence of these polymorphic states on stability and biological attributes of dosage forms have been receiving focused attention of many researchers. However, not much importance seems to have been given toward crystal habit perhaps because of the complexity of the crystallization process and its ability to simultaneously modify



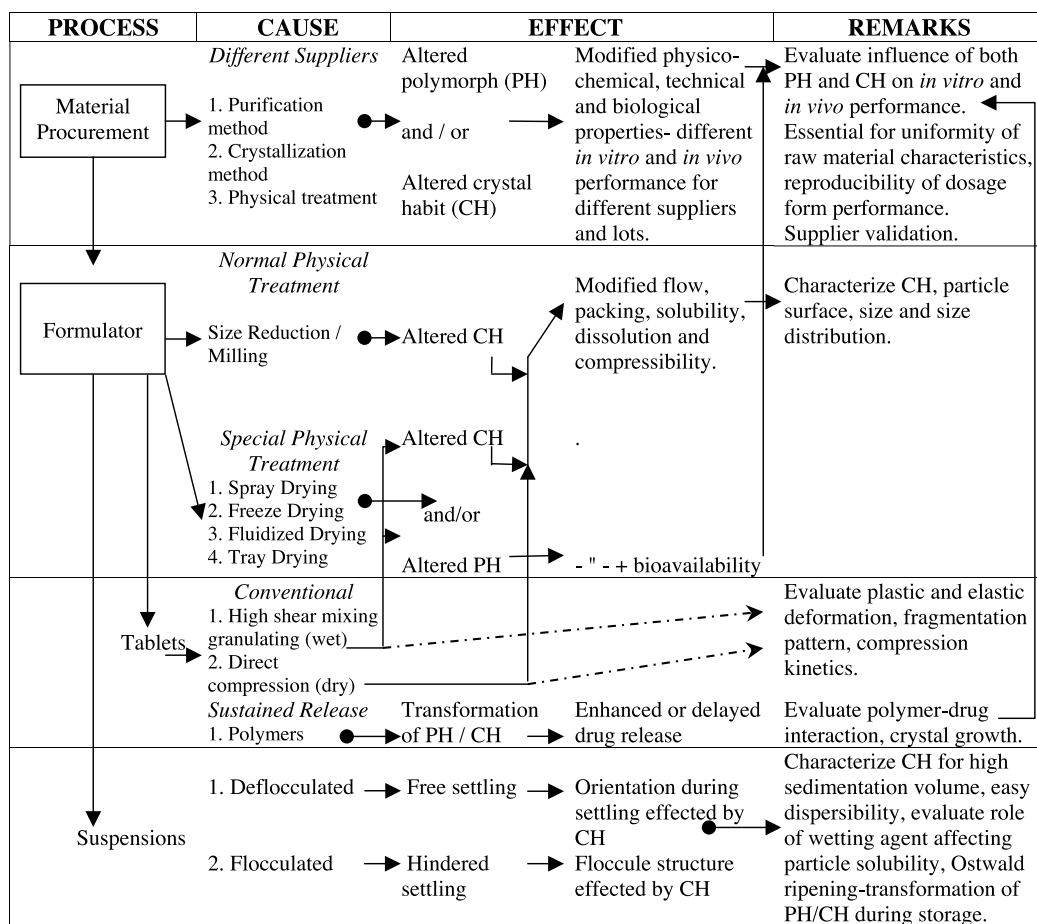


Fig. 1 Flow chart of the cause and effect relationship outlining the importance of crystal habit in dosage form design and performance.

both habit and polymorphic state of compounds. Nevertheless, because of the great impact of this seemingly trivial crystal property on dosage form performance, there is a need to study the important factors that influence crystal habit.

FACTORS INFLUENCING CRYSTAL HABIT

Degree of Supersaturation

The degree of supersaturation of the mother liquor or difference in concentration of solute on opposite sides of a growing crystal influences crystal habit. The effect of supersaturation on the change in habit was described by an equation, $y/x = k\Delta G^n$, where y/x is the ratio of crystal length to breadth, k is a coefficient of proportionality depending on diffusion, ΔG is the degree of supersaturation (moles/1000 mol of solvent at the moment of nuclei for-

mation), and n is a number that depends on the crystallographic classification and the chemical composition of the substance.^[3]

Barium sulfate precipitation has been reported to be controlled initially by nucleation reaction and, finally, by growth reaction. The growth kinetics was represented by the equation:

$$\frac{da}{dt} = -k(C_0 - C)^{2/3}a^q$$

where a is the mean ionic activity of Ba^{2+} and SO_4^{2-} , t is the time, C_0 and C are the molar concentrations of barium sulfate available at t_0 and t , respectively, k incorporates the kinetic constant and the shape factor, and q is proportional to the rate of change of the mean ionic activity and surface of the particle.^[4]

However, Nielsen^[5-7] reported the growth of barium sulfate to be controlled by a chemical reaction in which

Table 1 Influence of process variables of crystallization on crystal habit and dosage form performance

Process variable of crystallization	Possible influence on crystal habit	Possible influence on crystal/dosage form performance ^a
<i>Supersaturation</i>		
1. More saturation or significant solute–solvent interaction	Rate of nuclei formation is greater than crystal growth More growth in one direction producing needle-shaped crystals	Fine particles are produced Needle-shaped crystals exhibit poor flowability and cause bridging in hopper
2. Less saturation or insignificant solute–solvent interaction	Platy crystals are produced	Platy crystals exhibit greater dissolution, but are not preferred for tablet dosage form
<i>Rate of cooling and degree of solution agitation</i>		
1. Rapid cooling	Rapid crystal growth occurs and asymmetric (thin platy crystals) are produced	Platy crystals are not preferred for tablet dosage forms
2. Slow cooling	Rate of crystal growth decreases, and symmetric crystals are produced	Using symmetric, compact crystals gives more predictable and consistent performance
3. High speed of agitation	Even distribution of crystallizing solute on the nuclei produces elongated crystals with small particle size distribution	Desirable particle size range can be obtained; such crystals exhibit good flowability and less sedimentation in suspensions
4. Low speed of agitation or unstirred solutions	Crystallizing molecules deposit on selected crystal face, producing large platy crystals	Crystals with large particle size are unsuitable for formulations
<i>Nature of crystallizing solvent</i>		
1. More affinity for crystallizing solute	Formation of nuclei is delayed and fine, symmetric crystals are produced Interaction of certain functional groups between solvent and solute may impede growth at selected crystal faces Requires a high ratio of crystallizing solute/crystallizing solvent for producing well-defined shaped crystals	Desired crystal size with better dissolution and flowability Elongated crystals may be produced that are suitable for formulations because of their better flowability and sedimentation behavior
2. Less affinity for crystallizing solute	Nuclei are formed immediately and crystal growth is rapid; relatively larger crystals are produced A low ratio of crystallizing solute/crystallizing solvent is required for obtaining well-defined shaped crystals	
<i>Temperature of crystallizing solvent</i>		
1. Low temperature	Rapid nuclei formation because of spontaneous decrease in saturation level produces irregular shaped crystals	Irregular (dendritic) shaped crystals are not suitable for tableting
2. High temperature	Nuclei formation is delayed and fine, symmetric crystals are produced	Desirable shape and size of crystals can be obtained that are suitable for dosage forms
<i>Presence of impurities</i>		
1. Adsorbable ions, solute molecules	Inhibition or excessive growth of certain crystal faces	Desirable crystal morphology can be obtained during purification by recrystallization
2. Structural compatibility between polymer and drug	Interaction of functional groups between polymer and crystallizing solute restricts growth at certain crystal faces	Prevention of habit transformation

^aNo generalization can be made with regard to influence of process variables of crystallization on crystal habit and dosage form performance because the influence of a process variable is interactive and not independent of the other variable. The influences listed here should be used as a guideline only. (From Ref. [72], p. 701, by courtesy of Marcel Dekker, Inc.)



the crystal formation was a fourth-power step. The precipitation was found to be diffusion-controlled when the concentration was greater than 0.4 mM and could be characterized by the equation:

$$K_0 t = \int_0^a a^{-1/3} (1-a)^{-1} da$$

When the concentration was less than 0.4 mM, the kinetics could be described by the equation:

$$K_R t = \int_0^a a^{-2/3} (1-a)^{-4} da$$

where a represents the degree of precipitation. Hence an initial concentration less than 0.5 mM produced small prismatic crystals, and a concentration between 0.5 and 1.5 mM produced distorted prisms (corners grew more than the middle of the faces). However, at concentrations greater than 1.5 mM, the corners grew much more than other parts of the crystal, giving them a star-shaped appearance. It has been suggested that star-shaped crystals should be produced when the growth of crystals is diffusion-controlled because the concentration of crystallizing solute shall be greatest at the corners. On the other hand, rectangular-shaped crystals can be envisaged when the concentration of depositing solute is the same over the entire surface. This takes place when the rate of consumption of the solute is slower than the rate of diffusion.

Crystals of anhydrous cholesterol obtained from ethanol under quiescent conditions with a low supersaturation value were platelike; those obtained from shaken solutions were elongated, and those precipitated from stirred solutions were needlelike. The same trend was observed when acetonitrile or methanol was used as crystallizing solvents.^[8] Formation of needlelike crystals at high supersaturation of less polar solvent can be attributed to the significant solvent-solute interaction. Significant interaction probably results in preferential blocking of some faces, forcing the crystals to grow in one direction, resulting in a needlelike habit. Similarly, platy crystals are formed when the solute-solvent interaction is less.

Solute-solvent interaction is reported to influence the habit of stearic acid.^[9] It is worth noting that the addition of small amounts of surfactants forces stearic acid crystals to grow only in one habit modification regardless of the nature of solvent and the crystallization conditions. This is because of the modification of solute-solvent interaction by added surfactant molecules. Therefore in the absence of significant solute-solvent interaction (using relatively inert solvents), the degree of supersaturation will perhaps predominantly govern crystal habit. Supersaturation also influences the particle size of crystallizing solute. At high supersaturation, nucleation is more rapid than growth.

This results in the precipitation of fine particles. In thermal recrystallization, however, large crystals are produced because growth is faster than nucleation.^[10]

The degree of saturation in diffusional boundary layer next to the predominant face of a growing crystal can alter crystal growth in unstirred systems. Such a stagnant boundary layer that may be as thick as 150 μM can restrict diffusion and deposition of crystallizing molecules on growing crystal faces. Adsorbed impurities/polymers shall further retard crystal growth.^[11] Therefore the state of supersaturation of mother liquor appears to alter the shape of crystallizing particles by influencing the uniform deposition of molecules at different faces. It is important to note that the state of supersaturation depends not only on the choice of solvents used for solubilizing the drug, but also on the characteristics of the selected crystallizing solvent. The combined influence of both these solvents is critical in modifying habit as well as size of the crystallizing particles.

Rate of Cooling and Degree of Solution Agitation

Rate of cooling modifies crystal habit through its influence on the degree of supersaturation in mother liquor. Cooling a supersaturated solution of a drug or pouring it into crystallizing solvent maintained at low temperature immediately decreases the drug's solubility and results in rapid deposition of drug molecules on the nuclei.

Rapid cooling of a solution of naphthalene in ethanol or methanol produces thin plates, whereas slow cooling produces compact crystals. This is because of slow deposition of drug molecules on crystal faces at low rate of cooling.^[12] It has been suggested that rapid cooling usually produces needle-shaped crystals because an elongated shape is most efficient in dissipating heat.^[13] When the rate of cooling is decreased, the solution is maintained below saturation for longer duration. This results in delayed nuclei formation, slow deposition of crystallizing molecules, and, eventually, appearance of symmetric crystals. Slow cooling of a solution of acetazolamide in boiling water produced elongated prisms while faster cooling produced platy crystals.^[14] Similarly, cooling an aqueous saturated solution of paracetamol from 65°C to 25°C produced polyhedral crystals. However, when concentrated solution in hot ethanol is added to water (3°C), platy crystals were obtained.^[14] Hence modulation of the rate of cooling during crystallization could be employed as an effective means to alter crystal habit. As a corollary, habit changes shall be inevitable when a dosage form is repeatedly subjected to changes in temperature during processing or storage where the drug particles shall dissolve and then recrystallize to produce new crystals.

The degree of solution agitation is logically expected to influence saturation level at solid–solvent interface of the nuclei as well as bring about temperature drop in the system. This is probably the reason for the formation of large platy crystals from quiescent solutions and elongated crystals from stirred solutions.^[8] The effect of rate of solution agitation does not seem to have been studied extensively and needs critical evaluation.

Nature of Co-Solvent and Crystallizing Solvent

Crystallizing solvent is the medium into which a saturated drug solution prepared in a co-solvent is added to effect crystallization by precipitation. They may be a buffer (pH change method), water, or an organic solvent. As a prerequisite of this process of crystallization, the crystallizing solvent should be miscible with the liquid in which saturated solution of drug has been prepared. This requirement of miscibility implies that the intensity of solute–solvent interaction can be modified by selecting different co-solvents and/or crystallizing solvents.

Cholesterol is reported to crystallize as needles from ethanol and methanol and as platy crystals from acetonitrile.^[8] Trimethoprim^[16] and sulfamethoxazole^[17] have been crystallized in distinctively different habits from different crystallizing solvents (Fig. 2). These studies also

revealed that it was possible to obtain different habits of either drug belonging to the same polymorphic state using the same crystallizing solvent by just altering the process variables of crystallization such as co-solvent/crystallizing solvent ratio, temperature of co-solvent and crystallizing solvent, and rate of cooling. Co-solvents having high affinity for either drug required a higher ratio of co-solvent to crystallizing solvent for producing a well-defined morphology. A decrease in the initial supersaturation as a result of the elevated temperature of co-solvent or crystallizing solvent delayed the onset of nuclei formation and crystallization, thereby producing small crystals with equidimensional morphology. This was suggested to be a result of the slow and uniform deposition of solute molecules on the nuclei. However, variation of cooling rate did not appreciably affect the habit because of spontaneous precipitation of the drug. Therefore these studies suggest that process variables are as important as crystallizing solvent in modifying crystal habit and should be paid due attention while preparing crystals with specific attributes.

The interaction of a crystallizing solvent at various crystal–solution interfaces may lead to altered roundness of the growing crystal face/edges, change in crystal growth kinetics, and enhancement or inhibition of growth at certain crystal faces thus changing the habit.^[18] In addition, polarity of the solvent and its preferential adsorption at selected crystal faces can significantly alter

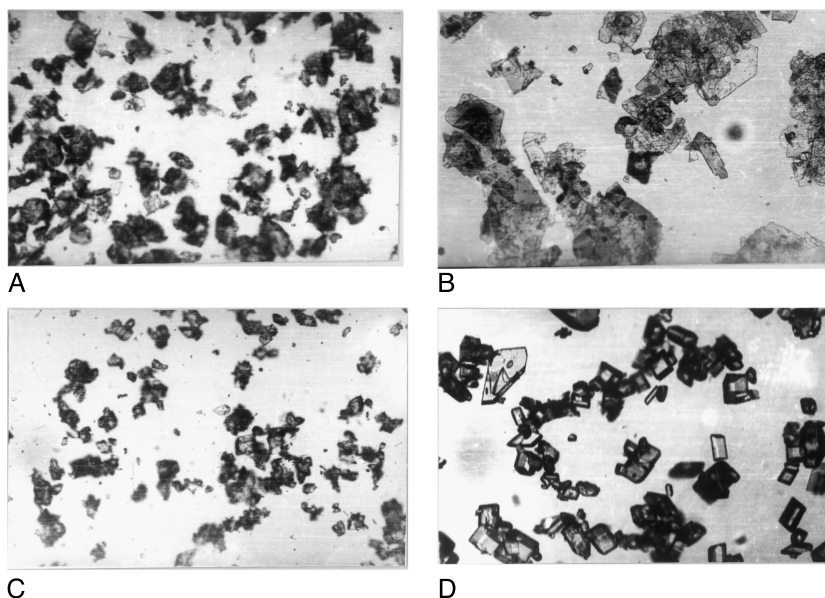


Fig. 2 Photomicrographs (magnification 200 \times) showing different habits of sulfamethoxazole (A and B same polymorph) and trimethoprim (C and D same polymorph): A and C—commercial drug samples; B—obtained by dissolving sulfamethoxazole in PEG 200 at room temperature and adding to water (70°C) in a ratio 1:20 followed by cooling at 4°C; D—obtained by dissolving trimethoprim in dimethyl formamide at room temperature and adding to water (70°C) in ratio 1:20 followed by cooling at room temperature.

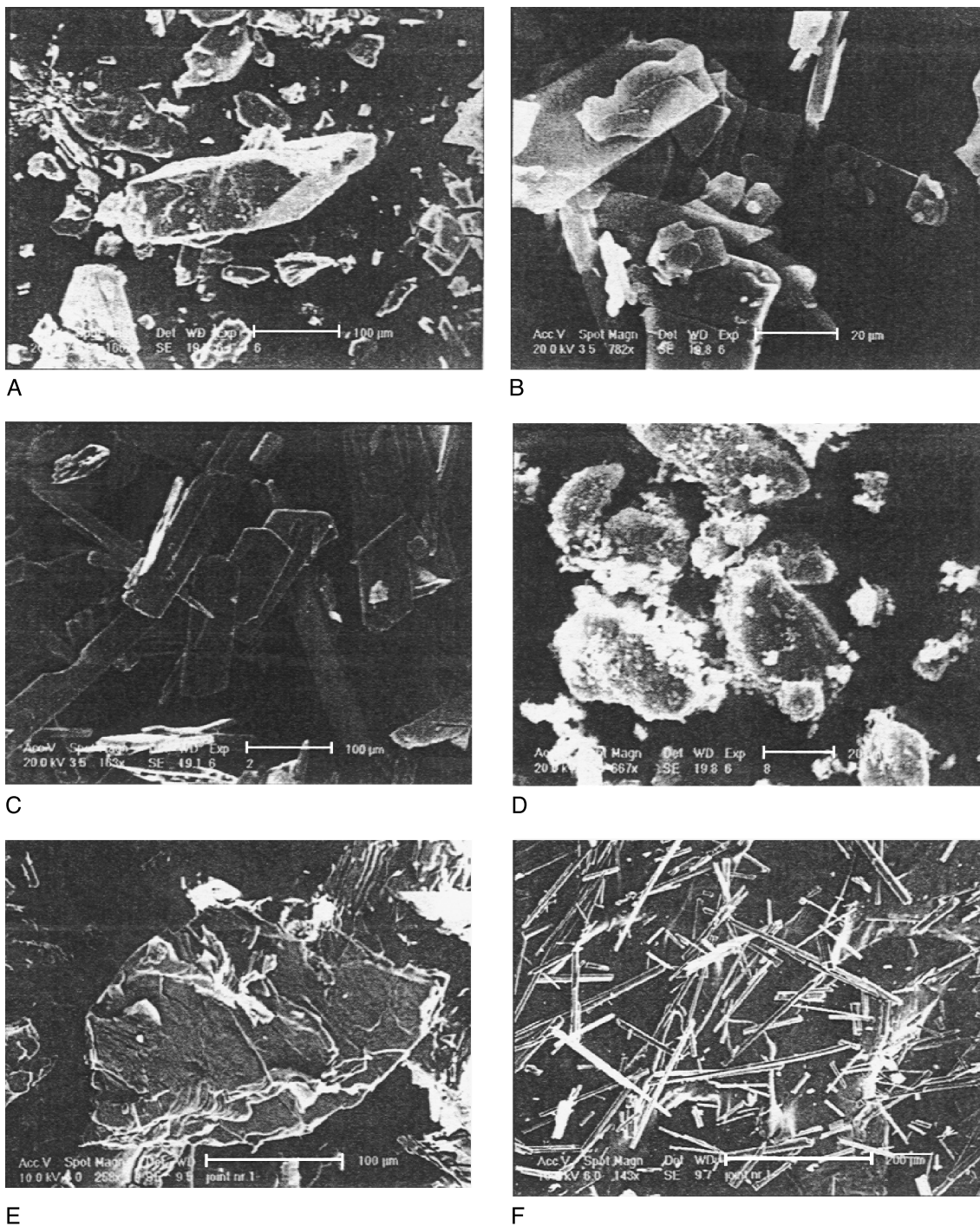


Fig. 3 SEM micrographs of commercial sodium diclofenac (anhydrous) (A); precipitated in tetrahydrate form (B); potassium diclofenac (dihydrate) (C); potassium diclofenac dihydrate (dehydrated at 100°C) (D); sodium diclofenac crystallized from methanol (E); potassium diclofenac crystallized from methanol (F). (From Ref. [20]. Reproduced with permission of John Wiley & Sons, Inc.)

the properties of crystallized solid particles. Nitrofurantoin has been reported to form a monohydrate when crystallized from formic acid/water (2:1) mixture because of the greater interaction of its polar regions with water than with formic acid. These water molecules have been sug-

gested to be retained at the active sites during crystal growth, while desolvation of formic acid occurs more readily thereby producing an altered habit.^[19]

Diclofenac sodium when dissolved in hot water and kept overnight crystallized in tetrahydrate form, whereas

its potassium salt crystallized in dihydrate form. Fig. 3A shows that the commercial sodium diclofenac existed as sturdy opaque crystals. Thin leaflet-shaped crystals were obtained when it was precipitated from water (Fig. 3B). On the other hand, potassium salt was found to exhibit a regular bladed morphology (Fig. 3C) that was lost after dehydration at 100°C (Fig. 3D). Recrystallization of sodium and potassium diclofenac from methanol produced stratified (Fig. 3E) and thin elongated rod-shaped crystals (Fig. 3F), respectively.^[20] Similarly, recrystallization of sulfadiazine from ammonia solution produced long, prismatic, smooth-edged crystals with markedly different physicochemical properties.^[21]

The nature of solvent has been found to have a profound effect on crystal habit of ibuprofen. Ibuprofen crystals precipitated from ethanol and acetone (solvents having high surface tension, dielectric constant, and less specific gravity) were thin, platy, and nearly circular-shaped, whereas those obtained from propylene glycol and 2-propanol were rod-shaped.^[22] Similar observations have been reported by Garekani et al.^[23] where all the habits of crystallized ibuprofen particles belonged to a common polymorphic state. Circular or polyhedral habits of ibu-

profen obtained by precipitation from polar solvents can be ascribed to the insignificant interaction with these solvents. Significant interaction in sodium hydroxide (pH 10) solution resulted in the precipitation of needle-shaped crystals when the pH was decreased by the addition of hydrochloric acid. However, spherical agglomerates were obtained when ibuprofen was dissolved in acetonitrile because of its limited miscibility with water in which crystallization occurred on emulsion droplets (Fig. 4).^[24]

It is important to note that unlike in precipitation method, both polar solvent, such as acetone, and nonpolar solvents, such as diethyl ether and hexane, produced needlelike crystals of ibuprofen when saturated solutions were cooled to 5°C over 120 min. Dichloromethane produced cubic, whereas acetonitrile produced spherical agglomerated crystals by this method. Solvent evaporation of ethanol gave platy crystals and diethyl ether gave needle-shaped crystals.^[24] The ability of a solvent to crystallize ibuprofen into strikingly different habits without changing its polymorphic state can be suggested to be a result of the alteration of solvent-solute interaction brought about by the use of different crystallization methods. Ibuprofen dissolves to a greater extent in semipolar

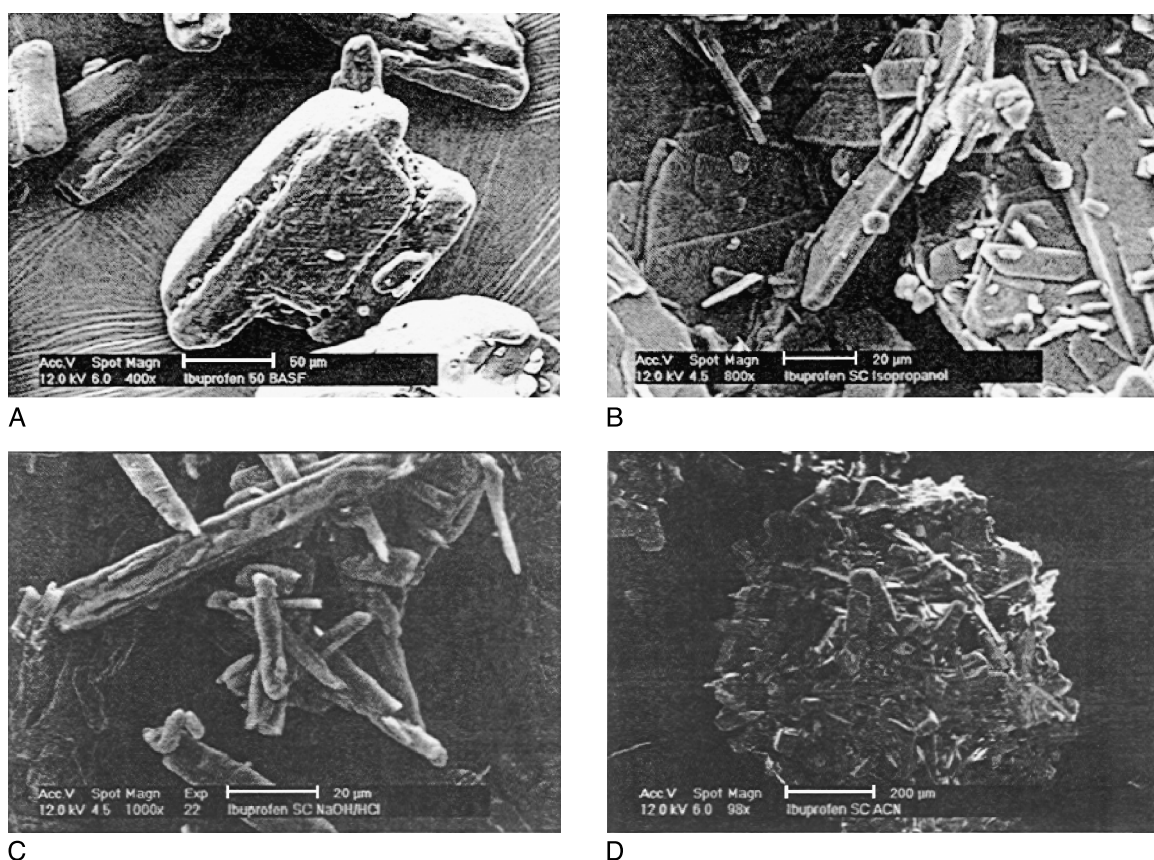


Fig. 4 SEM micrographs of ibuprofen commercial sample (A); crystals prepared by solvent change method using isopropyl alcohol (B); sodium hydroxide (C); acetonitrile (D). (From Ref. [24], p. 1083, by courtesy of Marcel Dekker, Inc.)



and nonpolar solvents, indicating a very high solute-solvent interaction. When crystallization is performed by precipitation in water, this strong interaction is weakened very soon and the crystal growth is not inhibited. This results in the deposition of solute molecules on all faces of the nuclei thus producing platy, polyhedral, or circular-shaped crystals. However, when crystallization is performed by temperature reduction or evaporation, the initially strong interaction is maintained till the end of the process. Therefore nuclei growth is inhibited on certain faces to the extent that both semipolar and nonpolar solvents produce needle-shaped crystals.

Presence of Impurities

Ions, polymeric molecules, or other substances present in solute or solvent can act as impurities for the growing crystals. Surface adsorption of methylated spirit on manitol seed crystals has been reported to produce highly porous surfaced crystals that were more resistant to vibrational segregation.^[25] At the molecular level, impurities may get adsorbed in crystal lattice and disturb the regular and repeating arrangements of a crystal. Such defects give rise to local regions of molecular disorder relative to the original crystal structure and are said to be in an activated state because of the greater molecular mobility and exposure of more reactive chemical groups. Hence such defects in the crystals lead to enhanced reactivity, solubility, and dissolution. Detailed discussion on such crystal lattice defects that lead to alteration in polymorphic state is not within the scope of this topic. However, because of their importance in influencing stability, dissolution, solubility,^[26,27] and thereby biological performance, it is necessary to ensure the absence of polymorphic modifications as a result of the crystal lattice defects to clearly define the role of crystal habit.

It will be more pertinent here to understand the mechanisms by which lattice defects created by added impurities modify crystal habit. Whetstone^[28-30] suggested that habit modification of inorganic salts depended on the anionic and cationic substitution and on the nature of substitution present in dyes. At low saturation level, an increase in the impurity concentration (cationic or anionic surfactants) was found to result in cessation of growth at certain crystal faces, whereas at high saturation, these surfactants produced less effect on the habit of adipic acid. Furthermore, anionic and cationic surfactants as impurities were found to modify the habit to needles and flakes, respectively.^[31,32] When valeric or undecanoic acid were added as impurities, adipic acid crystallized as cigar-shaped spars with rounded edges. Further increase in the concentration of impurities produced fused pairs of spherules. A linear plot of the maxi-

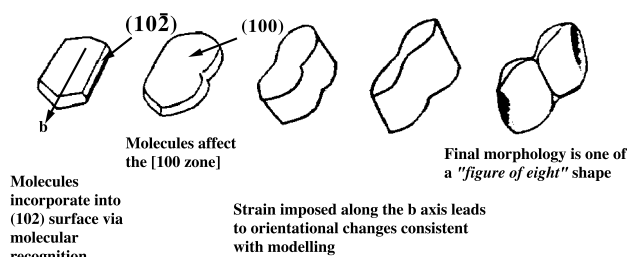


Fig. 5 Schematic illustration of growth process that produces dumbbell-shaped crystals of adipic acid. (From Ref. [36]. Reproduced with permission of John Wiley & Sons, Inc.)

mum growth rate slope of desupersaturation of adipic acid against concentration of caproic acid (impurity) suggested that its selective adsorption on various faces except near the surface was responsible for habit modification of adipic acid.^[33] Earlier, dumbbell-shaped appearance of modified crystals of adipic acid was suggested to arise from crystallographic twinning because of the similarity of the *a*- and *c*-axes of the unit cell.^[34] These crystals have been reported to possess modified density, crystal energy, and dissolution rate.^[35] Recently, it has been found that decanoic acid gets incorporated into adipic acid crystal {102} surface that results in loss of facets in the {010} zone. These lattice distortions result in crystals having misaligned domains, but they are not crystallographically twinned.^[36] Fig. 5 summarizes the steps that lead to the final dumbbell or “figure-of-eight” shaped morphology of adipic acid.

Apart from dyes and surfactants, polymeric molecules also influence the crystal habit of developing crystals. Sulfathiazole has been reported to grow out in finger-like protrusions in the presence of polyvinylpyrrolidone (PVP). It was suggested that PVP formed a “net” over the developing crystals. The effective pore size in the “net” and growth inhibition depended on the relative transport rates of PVP and sulfathiazole to crystal surface.^[37] Carbamazepine (anhydrous) is known to convert to the dihydrate form in water,^[38] and the growth of these crystals occurs by whisker mechanism. Both hydroxypropylmethyl cellulose (HPMC) and egg albumin (EA) have been found to retard the conversion of anhydrous carbamazepine to the dihydrate form in a concentration-dependent manner. In addition, HPMC was found to prevent the crystallization of the α -form. These preventive effects of HPMC at low concentrations could not be explained by simple adsorption because the polymer chain remained entangled and less mobile to enwrap all growing sites effectively. Therefore it was proposed that HPMC served as a template for heterogeneous nucleation, to which carbamazepine dimer attached through hydrogen bonding. More specifically, structural matching between interatomic distances

in the crystal lattice and intra-atomic distances along the polymer chain can be envisaged to be responsible for transformation inhibition.^[39] On the other hand, with EA, the prevention of crystallization of α -form was found to be less pronounced because of its lower ability to form hydrogen bonds with carbamazepine molecules. Egg albumin was found to decrease the degree of carbamazepine conversion to dihydrate form and whisker growth through an increase in aggregation of its dihydrate crystals. An increase in contact angle after addition of EA and reversibility of EA effect on aggregation by sodium dodecyl sulfate suggested that EA prevented the crystal growth of carbamazepine by increasing the dihydrate–liquid interfacial tension. Additionally, EA served as a microsubstitute for nucleation, thereby enhancing the two-dimensional nucleation on whisker sides.^[40]

It is interesting to note that PVP that contained only one hydrogen bonding carbonyl group per monomer unit was ineffective even at 500 times higher molar concentration than the minimum concentration of HPMC required to completely inhibit the transformation of carbamazepine to its dihydrate form.^[40] In contrast, crystal habit of paracetamol was found to be effectively modified by PVP with higher influence being exerted by higher molecular weight derivatives (PVP 10 000 or 50 000). Adsorption of high molecular weight PVP onto paracetamol growing crystals resulted in nearly spherical structure consisting of numerous rod-shaped microcrystals agglomerated together without any change in polymorphic state.^[41]

Discrepancies regarding the influence of polymeric impurities on habit modification seem to be because of the complex nature of nuclei formation and growth during crystallization. Polymers can get incorporated into the crystal only when they are similar to the crystallizing molecule in size or structure, which is often not the case. Hydroxypropylmethyl cellulose has been found to act as both growth inhibitor and habit modifier of hydrocortisone acetate, whereas PVP and polyethylene glycol 400 acted predominantly as growth inhibitors. This was suggested to be a result of the stronger interaction of HPMC (via hydroxyl groups) at the surface of the growing crystal. The influence becomes more pronounced in unstirred systems where a diffusional boundary layer adjacent to the growing crystal surface is formed. The accumulation of polymer molecules in this layer provides resistance for drug molecules to diffuse through and leads to growth inhibition. Additionally, habit gets modified when the polymeric groups get preferentially adsorbed at particular crystal faces thereby preventing growth of these faces, while other faces grow normally. Polymers that do not interact and get just adsorbed therefore lack the ability to modify the habit.^[11]

Polymorphic molecules such as methacrylic co-polymers may modify crystal habit by virtue of their ability to

form micelles because of the presence of quaternary ammonium groups in their molecules. Water-insoluble eudragits (RS and RL) that increased the solubility of ibuprofen produced more reduction in crystal yield than the water-soluble eudragits (L and S). However, all these polymers produced agglomerated spherical crystals with rough surfaces that were exhibited increased porosity. Greater pore diameter in crystals precipitated in the presence of eudragits accompanied with increased intraparticle porosity suggested the absence of polymer deposition in the empty spaces between microcrystals in the agglomerates.^[42]

Tailor-made impurities can be used for specific drug molecules. The use of *p*-acetoxyacetanilide (PAA) during crystallization of paracetamol has been found to produce columnar crystals. The incorporation of PAA into critical nucleus delayed the onset of nucleation and resulted in crystals that exhibited significant mosaic spread, implying the development of significant strain/defect content in the crystals.^[43]

Therefore irrespective of the mechanism under operation, the addition of impurities during crystallization can be advantageously employed to engineer crystal habit if the impurity is known to modify the growing crystal into a morphology that is desirable from the viewpoint of dosage form design and performance.

SIGNIFICANCE OF CRYSTAL HABIT IN DOSAGE FORM PERFORMANCE

Tablet Formulation

It is well known that crystal habits exhibiting symmetric morphological characteristics such as the cubic system present no difficulty during direct compression into tablets. Jaffe and Foss^[44] pointed out that binary compounds are found predominantly in the cubic and hexagonal system and are characterized by higher symmetry than ternary and more complex compounds found in the rhombic, monoclinic, and triclinic systems. Low-symmetry structures such as carbonates of calcium, lead, nickel, potassium, silver, and sodium have been reported not to form tablets on direct compression. The regular spherical particles of spray-dried lactose form stronger tablets compared with angular particles on direct compression.^[45] Similarly, replacement of platy habit (form B) by nonplaty (form A) crystals of tolbutamide has been reported to obliterate capping of tablets.^[46] These findings suggest symmetry of crystals to be a prerequisite for direct compression.

Successful tableting requires uniform flow from the hopper, proper packing, rearrangement, reduction in porosity, and deformation of particles in the die cavity. Flat



needle-shaped crystals of aspirin have been found to align themselves parallel to the punch face, forming a layered structure that exhibited low lateral stress transmission characteristics.^[47] An increased radially transmitted stress with flaky powered material^[48] reinforces this contention. Apart from the mechanical influence of crystal shape, another dominant factor that results from the anisotropy of cohesion and hardness (of low-symmetry crystals) also contributes to the ease of compression. As crystal habit varies, the dominant faces vary in relation to this anisotropy and tend to orient the crystals during compaction process thus exhibiting great differences in packing and compression kinetics.

Cubic sodium chloride crystals that pack in a bricklike fashion have been found to exhibit greater density than the dendritic crystals. In addition, reduced slip between dendritic crystals made rearrangement difficult and resulted in greater loss of compaction force to the die wall.^[49] Ibuprofen is generally crystallized industrially from hexane in the form of elongated needlelike crystals. This shape has been found to be unsuitable for tableting because of the poor flow properties. Equidimensional crystals obtained using methanol have been reported to possess better compaction features and flow properties.^[23,50] Crystal morphology of excipients such as powdered cellulose^[51] and calcium alumina trihydrate^[52] has also been reported to significantly influence strength, content uniformity, and disintegration time of tablets.

The platelike nitrofurantoin crystals are reported to undergo greater densification and plastic deformation than the needlelike crystals. Furthermore, tablets prepared from needlelike crystals exhibited a higher degree of axial recovery after ejection. This was suggested to be a result of the development of larger nonpolar faces in platelike crystals that were crystallized from formic acid than polar faces of needlelike crystals crystallized from formic acid/water mixture. The relative abundance of these faces probably affected the magnitude and the strength of bonding during compression.^[53]

An orthorhombic structure is characterized by high molecular density and weak interplane bonds. During the initial stages of compression cycle, these planes have been suggested to act as slide planes giving rise to substantial interparticle rearrangement, thus resulting in volume reduction. At high pressures, these crystals undergo plastic deformation with low elastic recovery as compared with monoclinic form, thus increasing the points of contact in the compact. Because of these reasons, orthorhombic paracetamol crystals exhibited good tablet forming property with no tendency of capping.^[54] Platy crystals of paracetamol have been reported to exhibit lower correlation coefficient values for Heckel plots and strain rate sensitivity, indicating greater fragmentation compared with polyhedral crystals. Tablets compressed from platy crystals

showed higher elastic recovery, suggesting that these crystals underwent less plastic deformation.^[15] It has been suggested on the basis of molecular geometry of paracetamol crystal lattice that the cleavage along the {010} plane follows a serrated path and the interplanar bonding is of van der Waals type. However, along both {110} and {210} planes, cleavage requires breakage of two hydrogen bonds per unit cell. In addition, the lowest attachment energy and maximum slice energy for the {010} plane strongly suggested that the fracture in paracetamol crystals occurred along the {010} crystal plane.^[55,56]

Compression of particles may alter the internal structure as well as their morphology. These changes are opposed by intermolecular forces that restore the crystal to its original form and results in elastic recovery. If the intermolecular forces are exceeded, then plastic flow occurs. It has been suggested for aspirin that the displacement occurring along the slip planes inside the crystal moved in an orderly manner to a new location, with the molecular packing arrangement remaining unchanged. However, cubic sodium chloride crystals possess numerous potential slip planes for plastic deformation, and microscopic examination did not reveal shearing effects. This indicates that shearing occurred at the molecular level in cubic sodium chloride crystals.^[57] An increase in compression pressure decreased the crystallinity of lactose, producing stronger tablets as a result of the more activated crystals dissipating acquired energy by interparticle bonding.^[58] Compacts made from equidimensional crystals of L-lysine monohydrochloride dihydrate in the absence of excessive pressure showed wider flaws (wider cracks between particles), whereas those from long rod-shaped crystals had longer flaws. These compacts of rod-shaped crystals were found to have poor strength probably because of the higher stress intensity at the crack tip that propagated to longer distances along the interparticle boundary.^[59] In addition, variation in the degree of surface crystallinity has been found to exhibit enormous influence on tablet properties.^[60] Hence an equidimensional habit that undergoes the highest degree of densification and exhibits a tendency to form new bonds at the fragmentation sites during compaction seems to offer great advantage for tablet formulation.

Dissolution

Dissolution of a drug depends on the physicochemical and physicochemical properties of drug particles. These crystal attributes directly affect the absorption kinetics of a drug and thereby bioavailability of dosage forms. This assumes greater importance for drugs exhibiting low solubility that makes absorption to be dissolution rate-limited. It is established that an increase in solubility can be brought by modifying the polymorphic state of a

compound. But the influence on other attributes such as stability, biological efficacy, metabolism, etc. as a result of change in polymorphic state demands a thorough investigation while using this approach. Modifying surface morphology of crystals without altering the polymorphic state seems to offer an attractive alternative approach for enhancing the dissolution of drugs. Furthermore, it is necessary to standardize the surface characteristics for minimizing differences in dissolution behavior of drug particles obtained from different batches/sources.

The size and the number of crystal faces exposed to solvent attack determine the amount of drug dissolved. It was shown that the lifetime of potassium ferricyanide crystal was proportional to the smallest length of the crystal face in contact with liquid paraffin–water interface. Furthermore, nonisometric dissolution of the crystal face indicated a change in crystal shape during dissolution.^[61] The shape factor of a single crystal of potassium dichromate changed significantly after 50% dissolution and was dependent on the degree of nonisometricity of the crystal.^[62] These reports indicate that a continuously changing dissolution profile can be anticipated for certain habits such as rods and needles because of the more solvent attack on predominant faces of the crystals as dissolution progresses.

At the molecular level, dissolution can be visualized to involve interaction of solvent with functional groups in the drug particles. Additives/impurities present in drug particles may alter the intensity of drug-dissolution medium interaction thereby modifying dissolution characteristics. Hydroxypropyl β -cyclodextrin has been found to be more effective than PVP in inhibiting the crystal growth and enhancing the dissolution of nifedipine.^[63] However, additives may sometimes even reduce the dissolution rate. A combination of Eudragit RS and cellulose acetate phthalate has been shown to alter the habit and sustain the release of an iron chelator.^[64] Crystallization of sulfadiazine from ammonia solution significantly decreased the dissolution rate because of the reduced wettability of the outer surface of recrystallized particles that seems to arise from the influence of crystallizing liquid.^[21] Depending on the solvent used for crystallization, internalization of the functional groups that are less attracted to the liquid takes place. Similarly, reduced wettability associated with agglomeration of crystals has been proposed to decrease the dissolution of ibuprofen crystals recrystallized from acetonitrile or methanol.^[24] It is important to note that nonionic impurities such as Tween 80 that do not disturb the arrangement of sulfadiazine molecules in ammonia solution did not appreciably affect the dissolution of recrystallized sulfadiazine. On the other hand, although the addition of sodium chloride as an impurity did not increase wettability, the dissolution of crystals was enhanced. It has

been proposed that probably sodium chloride, as a result of its ionic nature, reduced the interaction of ammonia with sulfadiazine through ion–dipole interaction. This in turn resulted in reducing the effect of ammonia on orientation of functional groups present in sulfadiazine molecules during recrystallization, thereby negating the dissolution retarding influence of ammonia on recrystallized sulfadiazine crystals.^[21]

The literature abounds with reports of investigations on enhancement of dissolution of drugs having low aqueous solubility by change in crystal habit. Ibuprofen crystals prepared by precipitation from ethanol and acetone have been reported to exhibit enhanced dissolution than the rod-like crystals obtained from propylene glycol and 2-propanol.^[22] Preparation of ibuprofen crystals by using phase partition technique and employing Tween 80 in the binding solvent have been found to produce microcrystals which when compressed into tablets exhibited dissolution comparable to that of commercial product with respect to USP 24 requirement ($Q > \text{or} = 80\%$ at 60 min).^[65] Other drugs that are known to have low aqueous solubility such as aspirin,^[66] mefenamic acid,^[67] and etoposide^[68] have also been reported to exhibit better dissolution after modification of crystal habit. However, results of dissolution enhancement should be cautiously interpreted because most studies do not often aim at delineating the role of crystal habit from that of polymorphic state which usually accompanies recrystallization of drug particles.

A pharmaceutical sample usually contains a mixture of habits. This is because of the complex natures of crystallization process in which even a scratch in the container or presence of dust particles is capable of initiating nuclei formation and modifying habit of growing crystals. Because of these mixed habits, the dissolution profile cannot be ascribed to a particular habit unless this shape constitutes a significant proportion of the bulk. Hence to correctly characterize the dissolution behavior and evaluate the contribution of various crystal faces, studies are made on single crystals. Prasad et al.^[69] have reported that the dissolution rate of {001} and {110} faces of paracetamol crystals grown in the presence of molecularly similar additive *p*-acetoxyacetanilide were higher than that of pure paracetamol. This was attributed to the distribution of strain in the crystal that increased its solubility.^[69] Apart from inclusion of impurities, processing variations can also lead to development of such strains. Moisture taken up even in low amounts in these areas of “strain” or “disorder” can plasticize the solid and promote molecular mobility that results in enhanced dissolution, chemical degradation, and solid-state changes such as recrystallization.^[26,27]

It is noteworthy that modified habits of trimethoprim and sulfamethoxazole having different habits but

belonging to the same sieve fraction and polymorphic state exhibited significantly different dissolution profiles. For both drugs, symmetric crystals were found to dissolve faster probably because of the uniform exposure of all the surfaces to dissolution medium. However, symmetric crystals having a higher size factor (length \times breadth) exhibited slow dissolution during the early phase. But during the later phase, as the size factor decreased, the dissolution rate of these crystals was found to increase. Symmetrical crystals that possessed low zeta potential exhibited considerable aggregation and dissolved slowly.^[16,17] It is perhaps for the first time that change in crystal habit has been shown to be associated with alteration of surface charge. The fact that surface charge has an ability to influence not just the aggregation state but also other powder characteristics such as mixing, flowability, dissolution, and solubility, it seems logical to evaluate the electrophoretic motility of recrystallized particles. Nevertheless, altering crystal habit seems to offer an approach for modifying the dissolution behavior of drugs.

Suspensions

The influence of crystal habit on performance of suspension dosage form can be envisaged to be more pronounced than other dosage forms because of greater space available for reorientation and packing of dispersed particles. Furthermore, selection of a stable habit is essential to avoid crystal growth that leads to physical instability during the shelf life of suspensions.

Investigations on trimethoprim^[16] and sulfamethoxazole^[17] suspensions have revealed that crystals with high shape factor ($1/L \times 1/B$) exhibited high sedimentation volume and could be easily redispersed. It was suggested that during settling, these crystals perhaps formed an end-to-face rather than an end-to-end framework because this would result in decreased free energy of the system. Although these anisometric crystals possessed high zeta potential, they tend to induce "self-flocculation," which produced a scaffold-like, porous pack structure that was less susceptible to overhead pressure of settling particles. On the other hand, crystals with irregular shape exhibited low sedimentation volume and were not easily redispersed because of cake formation. This was because of the reorientation of irregular crystals during sedimentation that lead to close-fit packing. It is noteworthy that a massive increase in sedimentation volume (sixfold) was obtained by employing rod-shaped crystals of trimethoprim in deflocculated suspension.^[16] The pharmacokinetics of both trimethoprim and sulfamethoxazole crystals was not significantly altered as compared with the respective pure drug powders probably because of their reported rapid absorption which does not render their bioavailability to be dependent on dissolution. However, studies on drugs that have inherently low solubility and/or low intestinal permeability are advocated to test the validity of this hypothesis.

It is important to consider the influence of interaction between functional groups of drugs that leads to their habit modification when formulated in suspension dosage form. Proton transfer from the N atom of sulfamethoxazole

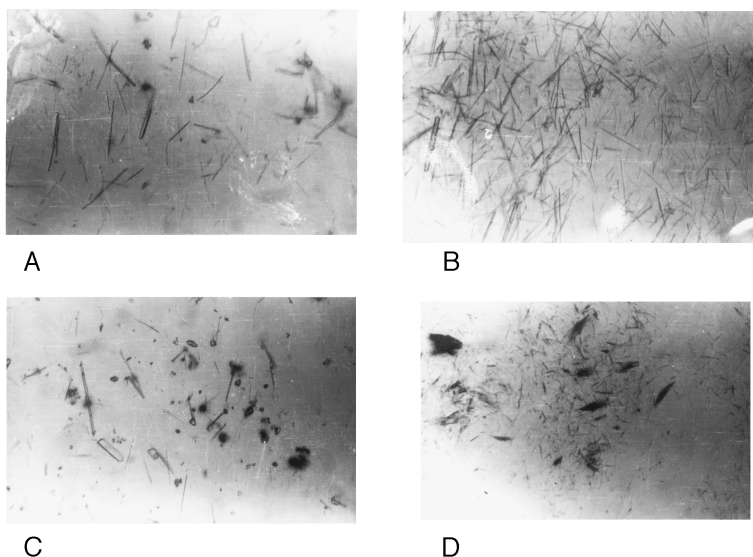


Fig. 6 Photomicrographs (magnification 200 \times) of crystals showing modified habit produced after interaction between aqueous dispersions of sulfamethoxazole and trimethoprim: molar ratio 5.73:1 (A); molar ratio 1:1 (B); HPMC added before mixing aqueous dispersions (C); HPMC added after mixing aqueous dispersions (D).

to the pyrimidine basic N1 atom of trimethoprim has been reported to occur in their equimolar complexes.^[70] Bettinetti et al.^[71] have reported nucleation of the complex of trimethoprim and sulfamethoxypyridazine (1:1) to be accelerated by water or wet granulation. Our studies on cotrimoxazole (unpublished results) revealed immediate formation of fine needle-shaped crystals irrespective of the initial shape of sulfamethoxazole and trimethoprim crystals as a result of the interaction between the two drugs in suspension form. Small needles (Fig. 6A) were produced when the aqueous dispersions of both drugs were mixed in amounts that are normally present in pharmaceutical suspensions (molar ratio sulfamethoxazole/trimethoprim: 5.73:1). Using the same processing conditions, long needle-shaped crystals were produced when the molar ratio was 3:2, 1:1, 2:3, or 1:2 (Fig. 6B), and these dispersions exhibited a sedimentation volume of 1.00 after storage over 2 months at 35°C. Furthermore, addition of aluminum chloride as flocculating agent modified the original crystals to long needle-shaped crystals. This was also evident when aluminum chloride was added to aqueous dispersion of trimethoprim. Addition of HMPc (1% w/v) before mixing dispersions of both drugs did not allow complete interaction as a result of which formation of needle-shaped crystals was found to be incomplete (Fig. 6C) and the resultant dispersions exhibited caking after storage. Perfect needles were formed but possessed shorter length (Fig. 6D) when HPMc was added after mixing the dispersions of both drugs. The results of these investigations strongly indicated that habit modification can result from interaction between functional groups in physical mixtures of drugs and that the intensity of modification can be altered by the formulation and processing variables. This influences the stability of suspension dosage forms, and hence occurrence of such habit modifications should be given due attention during preformulation. As a corollary, this approach could be used to enhance the physical stability of drug suspensions that are prone to caking on storage. This can be achieved by employing interaction between the drug and an inert substance to produce crystals with appropriate habit that can form a porous pack structure that will embed the drug particles. Such suspensions shall exhibit high sedimentation volume and ease of redispersibility. However, feasibility of using this approach is yet to be tested.

CONCLUSION

Crystal morphology is an important attribute of solids. Both crystal shape (habit) and surface appearance can be significantly altered by the process variables of crystal-

lization. Crystallization, which is often used for purifying/recrystallizing a drug powder, may modify either polymorphic state, crystal habit, or both. Although seemingly trivial, crystal habit plays a significant role in influencing packing, flowability, compressibility, dissolution, and sedimentation characteristics of pharmaceutical powders. Therefore selecting a stable polymorph may not alone solve the problem if the polymorph exists in distinctly different habits that can modify the physical stability of the dosage form. In addition, polymorphic transformations or crystal growth during storage may accompany change in crystal habit. Furthermore, development of strain in crystal lattice either because of impurities/additives or during processing can lead to changes in habit that can alter the performance of dosage form. Hence it is imperative to study changes in crystal habit along with changes in polymorphic state at critical processing stages. This shall help in maintaining uniformity in raw material characteristics and batch-to-batch dosage form performance besides formulating a stable dosage form with required performance.

REFERENCES

1. Takubo, H.; Kume, S.; Koizumi, M. Relationship between supersaturation, solution velocity, crystal habit and growth rate in crystallization of ammonium phosphate ($\text{NH}_4\text{-H}_2\text{PO}_4$). *J. Cryst. Growth* **1984**, *67* (2), 217–226.
2. Halebian, J.K. Characterization of habits and crystalline modifications of solids and their pharmaceutical applications. *J. Pharm. Sci.* **1975**, *64* (8), 1269–1288.
3. Grzymek, J. Control of the exterior form of crystals during crystallization. *J. Przemysl. Chem.* **1937**, *21*, 279–280.
4. Johnson, R.A.; O'Rourke, J.D. Kinetics of precipitate formation—Barium sulphate. *J. Am. Chem. Soc.* **1954**, *76*, 2124–2126.
5. Nielsen, A.E. The kinetics of electrolyte precipitation. *J. Colloid Sci.* **1955**, *10*, 576–586.
6. Nielsen, A.E. Nucleation in barium sulphate. *Acta Chem. Scand.* **1957**, *11*, 1512–1515.
7. Nielsen, A.E. Kinetics of crystal growth in barium sulphate precipitation. *Acta. Chem. Scand.* **1958**, *12*, 951–958.
8. Garti, N.; Karpuj, L.; Sarig, S. The effect of solvents and crystallization conditions on crystal habit of cholesterol. *Cryst. Res. Technol.* **1981**, *16* (10), 1111–1115.
9. Garti, N.; Wellner, E.; Sarig, S. Effect of surfactants on crystal structure modification of stearic acid. *J. Cryst. Growth* **1982**, *57* (3), 577–584.
10. Carstensen, J.T.; Ertell, C.; Geoffroy, J.M. Physicochemical properties of particulate matter. *Drug Dev. Ind. Pharm.* **1993**, *19* (1–2), 195–219.
11. Raghavan, S.L.; Trividic, A.; Davis, A.F.; Hadgraft, J. Crystallization of hydrocortisone acetate: Influence of polymers. *Int. J. Pharm.* **2001**, *212*, 213–221.

12. Tipson, R.S. *Separation and Purification Part I*; Weissberger, A., Ed.; Interscience: New York, 1956, Vol. 3, 433.
13. Carstensen, J.T. *Theory of Pharmaceutical Systems: Heterogeneous System*; Academic Press: New York, 1973, Vol. 2, 145, 262.
14. Griesser, U.J.; Burger, A.; Mereiter, K. The polymorphic drug substances of the European pharmacopoeia. Part 9. Physicochemical properties and crystal structure of acetazolamide crystal forms. *J. Pharm. Sci.* **1997**, *86* (3), 352–358.
15. Garekani, H.A.; Ford, J.L.; Rubinstein, M.H.; Riajbi-Siahboomi, A.R. Formation and compression characteristics of prismatic polyhedral and thin plate like crystals of paracetamol. *Int. J. Pharm.* **1999**, *187* (1), 77–89.
16. Tiwary, A.K.; Panpalia, G.M. Influence of crystal habit on trimethoprim suspension formulation. *Pharm. Res.* **1999**, *16* (2), 261–265.
17. Tiwary, A.K.; Panpalia, G.M. Enhanced suspension stability: Choice of crystal habit. *Indian J. Pharm. Sci.* **2000**, *62* (4), 267–272.
18. Davey, R.J. Solvent Effects in Crystallization Processes. In *Current Topics in Material Science*; Kaldis, E., Ed.; Elsevier: Amsterdam, 1982; Vol. 8, 429–479.
19. Marshall, P.V.; York, P. Crystallization solvent induced solid-state and particulate modifications of nitrofurantoin. *Int. J. Pharm.* **1989**, *55*, 257–263.
20. Fini, A.; Garuti, M.; Fazio, G.; Alvarez-Fuentes, J.; Holgado, M.A. Diclofenac salts. I. Fractal and thermal analysis of sodium and potassium diclofenac salts. *J. Pharm. Sci.* **2001**, *90* (12), 2049–2057.
21. Hammouda, Y.E.; El-Khordagui, L.K.; Darwish, I.A.; El-Kamel, A.H. Manipulation of powder characteristics by interactions at the solid–liquid interface: 1-sulphadiazine. *Eur. J. Pharm. Sci.* **1999**, *8*, 280–283.
22. Khan, G.M.; Jiabi, Z. Preparation, characterization and evaluation of physicochemical properties of different crystalline forms of ibuprofen. *Drug Dev. Ind. Pharm.* **1998**, *24* (5), 463–471.
23. Garekani, H.A.; Sadeghi, F.; Badiee, A.; Mostafa, S.A.; Rajabi-Siahboomi, A.R. Crystal habit modifications of ibuprofen and their physicochemical characteristics. *Drug Dev. Ind. Pharm.* **2001**, *27* (8), 803–809.
24. Rasenack, N.; Muller, B.W. Properties of ibuprofen crystallized under various conditions: A comparative study. *Drug Dev. Ind. Pharm.* **2002**, *28* (9), 1077–1089.
25. Stainforth, J.N.; Rees, J.E.; Kayes, J.B.; Priest, R.C.; Cotterill, N.J. The design of a direct compression tablet excipient. *Drug Dev. Ind. Pharm.* **1981**, *7* (1), 179–190.
26. Ahlneck, C.; Zografi, G. The molecular basis of moisture effects on the physical and chemical stability of drugs in the solid state. *Int. J. Pharm.* **1990**, *62*, 87–95.
27. Gerhardt, S.A.; Ahlneck, C.; Zografi, G. Assessment of disorder in crystalline solids. *Int. J. Pharm.* **1994**, *101*, 237–247.
28. Whetstone, J. Crystal habit modification of inorganic salts with dyes. II. Relationship between the structure of crystals and habit-modifying dyes. *Trans. Faraday Soc.* **1955**, *51*, 1142–1153.
29. Whetstone, J. Crystal symmetry and the adsorption of dyes by growing crystals. *J. Chem. Soc.* **1956**, 4841–4847.
30. Whetstone, J. Crystal symmetry and the adsorption of dyes by growing crystal. II. Potassium nitrate. *J. Chem. Soc.* **1957**, 4284–4289.
31. Michaels, A.S.; Colville, A.R., Jr. The effect of surface active agents on crystal growth rate and crystal habit. *J. Phys. Chem.* **1960**, *64*, 13–19.
32. Michaels, A.S.; Tausch, F.W., Jr. Modification of growth rate and habit of adipic acid crystals with surfactants. *J. Phys. Chem.* **1961**, *65*, 1730–1737.
33. Faribrother, J.E.; Grant, D.J.W. Crystal engineering studies with an excipient material (adipic acid). *J. Pharm. Pharmacol.* **1979**, *31*, Suppl., 27 pp.
34. Davey, R.J.; Black, S.N.; Logan, D.; Maginn, S.J.; Fairbrother, J.E.; Grant, D.J.W. Structural and kinetic features of crystal growth inhibition: Adipic acid growing in the presence of *n*-alkanoic acids. *J. Chem. Soc., Faraday Trans.* **1992**, *88* (23), 3461–3466.
35. Chow, K.Y.; Go, J.; Mehhdiazadeh, M.; Grant, D.J.W. Modification of adipic acid crystals: Influence of growth in the presence of fatty acid additives on crystal properties. *Int. J. Pharm.* **1984**, *20*, 3–24.
36. Williams-Seton, L.; Davey, R.J.; Liebermann, H.F.; Pritchard, R.G. Disorder and twinning in molecular crystals: Impurity-induced effects in adipic acid. *J. Pharm. Sci.* **2000**, *89* (3), 346–354.
37. Simonelli, A.P.; Mehta, S.C.; Higichi, W.I. Dissolution rates of high energy sulphathiazole-povidone coprecipitates II: Characterization of form of drug controlling its dissolution rate via solubility studies. *J. Pharm. Sci.* **1976**, *65* (3), 355–361.
38. Laine, E.; Tuonien, V.; Ilvessalo, P.; Kahela, P. Formation of dehydrate from carbamazepine anhydrate in aqueous conditions. *Int. J. Pharm.* **1984**, *20*, 307–314.
39. Katzhendler, I.; Azoury, R.; Friedman, M. Crystalline properties of carbamazepine in sustained release hydrophilic matrix tablets based on hydroxypropylmethyl cellulose. *J. Contr. Rel.* **1998**, *54*, 69–85.
40. Katzhendler, I.; Azoury, R.; Friedman, M. The effect of egg albumin on the crystalline properties of carbamazepine in sustained release hydrophilic matrix tablets and in aqueous solutions. *J. Contr. Rel.* **2000**, *65*, 331–343.
41. Garekani, H.A.; Ford, J.L.; Rubinstein, M.H.; Rajabi-Siahboomi, A.R. Highly compressible paracetamol: I: Crystallization and characterization. *Int. J. Pharm.* **2000**, *208* (1–2), 87–99.
42. Kachrimanis, K.; Nikolakakis, I.; Malamataris, S. Spherical crystal agglomeration of ibuprofen by the solvent-change technique in presence of methacrylic polymers. *J. Pharm. Sci.* **2000**, *89* (2), 250–259.
43. Prasad, K.V.; Ristic, R.I.; Sheen, C.B.; Sherwood, J.N. Crystallization of paracetamol from solution in the presence and absence of impurity. *Int. J. Pharm.* **2001**, *215* (1–2), 29–44.
44. Jaffe, J.; Foss, N.E. Compression of crystalline substances. *J. Am. Pharm. Assoc. (Sci. Ed.)* **1959**, *48*, 29.



45. Alpar, O.; Hersey, J.A.; Shotton, E. Compression properties of lactose. *J. Pharm. Pharmacol.* **1970**, *22*, Suppl., 1S.
46. Simmons, D.L.; Ranz, R.L.; Gyanchandani, N.D. Polymorphism in pharmaceuticals. III. Chlorpropamide. *Can. J. Pharm. Sci.* **1973**, *8*, 125–127.
47. Windheuser, J.J.; Misra, J.; Eriksen, S.P.; Higuchi, T. Physics of tablet compression XIII: Development of the wall pressure during compression of various materials. *J. Pharm. Sci.* **1963**, *52* (7), 767–772.
48. Nakagawa, H.; Fukuoka, E.; Nakai, Y. Preferred orientation of crystalline particles within tablets III—Change in degree of preferred orientation of aspirin crystals during compression. *Yakugaku Zasshi* **1979**, *99*, 936–943.
49. Shotton, E.; Obiorah, B.A. The effect of particle shape and crystal habit on properties of sodium chloride. *J. Pharm. Pharmacol.* **1973**, *25* (Suppl. 37), 43.
50. Gordon, R.E.; Amin, S.I.; Crystalline Ibuprofen. Eur. Patent EP 120,587, October 03, 1984.
51. Szabo-Revesz, P.; Pintye-Hodi, K.; Miseta, M.; Selmeezi, B. Tableting with cellulose products. Powdered cellulose as an excipient for direct compression. *Pharm. Ind.* **1992**, *54* (1), 79–81.
52. Veessler, S.; Boistelle, R.; Delacourte, A.; Guyot, J.-C.; Guyot-Herman, A.-M. Influence of structure and size of crystalline aggregates on their compression ability. *Drug Dev. Ind. Pharm.* **1992**, *18* (5), 539–560.
53. Marshall, P.V.; York, P. The compaction properties of nitrofurantoin samples crystallized from different solvents. *Int. J. Pharm.* **1991**, *67*, 59–65.
54. Joiris, E.; DiMartino, P.; Berneron, C.; Guyot-Herman, A.-M.; Guyot, J.-C. Compression behaviour of orthorhombic paracetamol. *Pharm. Res.* **1998**, *15* (7), 1122–1130.
55. Prasad, K.V.; Sheen, D.B.; Sherwood, J.N. Fracture property studies of paracetamol single crystals using micro-indentation techniques. *Pharm. Res.* **2001**, *18* (6), 867–872.
56. Finnie, S.; Prasad, K.V.; Sheen, D.B.; Sherwood, J.N. Microhardness and dislocation identification studies on paracetamol single crystals. *Pharm. Res.* **2001**, *18* (5), 674–681.
57. Hess, H. Tablets under the microscope. *Pharm. Technol.* **1978**, *2* (9), 38–57.
58. Huettenrauch, R. Mechanism of the particle size effect during tableting. *Pharmazie* **1977**, *32* (2), 130–131.
59. Bandyopadhyay, R.; Grant, D.J. Influence of crystal habit on the surface energy and interparticulate bonding of L-lysine monohydrochloride dihydrate. *Pharm. Dev. Technol.* **2000**, *5* (1), 27–37.
60. Buckton, G. Surface characterization: Understanding sources of variability in the production and use of pharmaceuticals. *J. Pharm. Pharmacol.* **1995**, *47*, 265–275.
61. Schoonen, A.J.; de Vries, N.G.W.; Huizinga, T. Solution rate of crystals at fluid–fluid interface. *J. Pharm. Sci.* **1979**, *68* (2), 163–168.
62. Dali, M.V.; Carstensen, J.T. Effect of change in shape factor of a single crystal on its dissolution behaviour. *Pharm. Res.* **1996**, *13* (1), 155–162.
63. Uekama, K.; Ikegami, K.; Wang, Z.; Horiuchi, Y.; Hirayama, F. Inhibitory effect of 2-hydroxypropyl-beta-cyclodextrin on crystal growth of nifedipine during storage: Superior dissolution and oral bioavailability compared with polyvinylpyrrolidone K-30. *J. Pharm. Pharmacol.* **1992**, *44* (2), 73–78.
64. Venkataram, S.; Khohlokwane, M. Microencapsulation of an iron chelator for sustained release and crystal habit modification. *J. Microencapsulation* **1996**, *13* (5), 519–525.
65. Umprayn, K.; Luengtummueng, A.; Kitiyadisai, C.; Pornpi-putsakul, T. Modification of crystal habit of ibuprofen using the phase partition technique: Effect of aerosol and tween 80 in binding solvent. *Drug Dev. Ind. Pharm.* **2001**, *27* (10), 1047–1056.
66. Goczó, H.; Szabo-Revesz, P.; Farkas, B.; Hasznos-Nezdei, M.; Serwanis, S.F.; Pintye-Hodi, A.K.; Kasa, P., Jr.; Eros, I.; Antal, I.; Marto, S. Development of spherical crystals of acetylsalicylic acid for direct tablet-making. *Chem. Pharm. Bull.* **2000**, *48* (12), 1877–1881.
67. Adam, A.; Schrimpl, L.; Schmidt, P.C. Some physico-chemical properties of mefenamic acid. *Drug Dev. Ind. Pharm.* **2000**, *26* (5), 477–487.
68. Shah, J.C.; Chen, J.R.; Chow, D. Metastable polymorph of etoposide with higher dissolution rate. *Drug Dev. Ind. Pharm.* **1999**, *25* (1), 63–67.
69. Prasad, K.V.; Ristic, R.I.; Sheen, D.B.; Sherwood, J.N. Dissolution kinetics of paracetamol single crystals. *Int. J. Pharm.* **2002**, *238* (1–2), 29–41.
70. Nakai, H.; Takasuka, M.; Shiro, M. X-ray and infrared spectral studies of the ionic structure of trimethoprim-sulphamethoxazole 1:1 molecular complex. *J. Chem. Soc., Perkin Trans.* **1984**, *II*, 1459–1464.
71. Bettinetti, G.; Caira, M.R.; Callegari, A.; Merli, M.; Sorrenti, M.; Tadini, C. Structure and solid-state chemistry of anhydrous and hydrated crystal forms of the trimethoprim-sulphamethoxypyridazine 1:1 molecular complex. *J. Pharm. Sci.* **2000**, *89* (4), 478–489.
72. Tiwary, A.K. Modification of crystal habit and its role in dosage form performance. *Drug Dev. Ind. Pharm.* **2001**, *27* (7), 699–709.



COSOLVENTS AND COSOLVENCY

Joseph T. Rubino

Wyeth-Ayerst Research, Pearl River, New York, U.S.A.

INTRODUCTION

Cosolvents are defined as water-miscible organic solvents that are used in liquid drug formulations to increase the solubility of poorly water-soluble substances or to enhance the chemical stability of a drug. Cosolvency, then, refers to the technique of using cosolvents for the stated purposes; it is also commonly referred to as solvent blending. Cosolvency has been used as an approach for preparing liquid drug preparations throughout the history of drug formulation. Certain drugs of botanic origin were known to be poorly soluble in water and required formulation in water-ethanol mixtures in order to deliver an adequate dose of drug in a small volume of preparation. A common example of a class of formulation containing cosolvents is the elixir, which by definition is a sweetened, hydroalcoholic solution intended for oral use. Tinctures, which generally contain even higher amounts of alcohol, are another classic example of a liquid dosage form containing a cosolvent. The need to employ cosolvents in the formulation of new drugs as solutions for oral, parenteral, and topical use remains high, especially with the increasing structural complexity of new therapeutic agents.

In many cases, cosolvency can increase the solubility of a nonpolar drug up to several orders of magnitude above the aqueous solubility. This would be significant, for example, in a formulation problem where it might be necessary to increase the solubility of a drug 500-fold or more. The use of cosolvents to prepare solution formulations of nonpolar drugs is a simple and potentially effective way to achieve high concentrations of drug.

The primary disadvantages of cosolvency include the potential for biological effects and the potential for drugs that have been solubilized using cosolvents to precipitate upon dilution with aqueous fluids. The biological effects of a cosolvent that may limit or eliminate its use in drug formulations include their general toxicity, target organ toxicity, tissue irritation, or tonicity with respect to biologic membranes. In addition, precipitation of drug upon dilution with aqueous media or during injection or application to mucous membranes must always be considered in deciding if a cosolvent can be used as a vehicle for poorly water-soluble drugs. Other considerations

include the viscosity, tonicity, and taste, as well as the effect of cosolvents on the solubility and stability of formulation components other than the drug.

When used as a method for increasing the chemical stability of a drug, cosolvents may be effective by one or two mechanisms. If a drug is susceptible to hydrolytic degradation, cosolvents may reduce the degradation of the drug by substituting for some or all of the water in the formulation. Alternatively, a cosolvent may enhance the stability of a drug by providing a less suitable environment for the transition state of the reactants, provided the transition state is more polar than the reactants themselves (1).

SOLUBILIZATION BY COSOLVENTS

Methods and Theories

Solubility and solvent polarity

The solubility of a solid nonelectrolyte solute in any solvent can be expressed by Eq. 1:

$$\log X = \frac{-\Delta H_f}{2.303RTT_m}(T_m - T) - \log y \quad (1)$$

where X is the solute mole fractional solubility, ΔH_f is the molar heat of fusion, R is the gas constant, T_m is the solute melting point, T is the temperature, and y is the activity coefficient. The activity coefficient is given by Eq. 2.

$$\log y = C_{11} + C_{22} - 2C_{12} \quad (2)$$

where C_{11} represents the cohesive energy of the solvent, C_{22} is the cohesive energy of the solute (drug), and C_{12} is the adhesive energy. Often C_{12} is calculated as some function of both C_{11} and C_{22} . For example, when dispersion forces are involved:

$$C_{12} = (C_{11}C_{22})^{1/2} \quad (3)$$

The larger the value of $\log y$, the lower the solubility of the drug, with the magnitude of y being controlled by the relative magnitudes of the cohesive and adhesive forces. In the case of hydrophobic drug in water, C_{11} will be large

compared to C_{22} and C_{12} . Thus, the solubility will be low due to the large cohesive energy of water and the low energy of interaction with the drug.

In an ideal solution, all intermolecular forces are equal:

$$C_{11} = C_{22} = C_{12}$$

and $\log \gamma = 0$. Although pharmaceutical solutions are rarely ideal, the goal of cosolvency is to make the solution closer to ideal by modification of the solvent polarity. Upon replacement of some water with an organic cosolvent, C_{11} becomes closer in value to C_{22} , which results in both factors becoming more similar to C_{12} . The value of $\log \gamma$ is reduced and the solubility increases. The maximum solubility for a solute generally occurs when the solvent–solvent intermolecular forces are equal to solute–solute intermolecular forces. Thus, the expression, “like dissolves like,” can be understood by this analysis. In some cases, strong polar or charge transfer interactions between solute and solvent can result in a greater than ideal solubility due to relatively strong C_{12} values, which results in a negative value for $\log \gamma$, but generally this will not be the case.

For the purpose of estimating the solubility of a solute it is necessary to have some measure of the “polarity” of a solute or a solvent. Based on Eqs. 1 and 2, a useful polarity index should be a measure of a material’s intermolecular forces, C_{11} and C_{22} . Table 1 contains a list of solvents that are typically used in liquid pharmaceutical formulations and three measures of solvent polarity. Each measure of solvent polarity, or polarity index, is based upon a different measure of a material’s property. For example, dielectric constant is a measure of the electrical insulating properties of a solvent, solubility parameter is determined from the molar energy of vaporization, and interfacial tension is a measure of the 2-D cohesive forces at the solvent–oil interface. Because solubility parameter and interfacial tension relate more specifically to cohesive

interactions among molecules, they have been used more frequently in various theories to explain and predict solubility in mixed solvents. Despite the differences in their determination, there is a general agreement in the rank order of the polarity among the various solvents. Thus, water and glycerin are relatively polar compared to ethanol or polyethylene glycol (PEG), regardless of which polarity index is considered. Other measurements of polarity or solvent–solvent intermolecular forces have been developed that measure hydrogen bond donating and accepting ability, such as Taft and Kamlet α and β values (2, 3), and 3-D solubility parameters that estimate dispersion, polar, and hydrogen-bonding forces (4, 5), etc.

It is important to realize that a discussion of solvent polarity is only useful when considered in relation to a particular solute. Thus, for a semipolar solute such as theophylline, pure ethanol might be considered nonpolar, but for a relatively nonpolar solute such as hydrocortisone, ethanol might be considered semipolar. This concept is illustrated in Fig. 1. For “nonpolar” and “polar” solutes, solubility will continually decrease or increase, respectively, as solvent polarity increases. For “semipolar” solutes, solubility reaches a maximum at some intermediate solvent polarity. It should be noted that the reverse argument (i.e., solute polarity should be discussed relative to solvent polarity) is also valid.

Due to the complexity of intermolecular interactions in hydrogen-bonded solutions, such as those used in pharmaceutical applications, no single parameter is capable of quantifying all interactions. Several theories have been developed to estimate drug solubility in cosolvent–water mixtures. The relationships that have resulted from these theories range from relatively simple to complex, depending on the desired accuracy of solubility prediction. Some theories and guidelines for selecting an appropriate solvent are presented in the following section.

Table 1 Cosolvents and polarity indices

Cosolvent	Dielectric constant ^a (ϵ)	Solubility parameter (δ) (cal/cm ³)	Interfacial tension ^b (dynes/cm)
Water	78.5	23.4	45.6
Glycerin	42.5	17.7	32.7
<i>N, N</i> -Dimethylacetamide	37.8	10.8	4.6
Propylene glycol	32.0 (30°C)	12.6	12.4
Ethanol	24.3	12.7	0.5
Polyethylene glycol 400	13.6	11.3	11.7
Dimethylisobutide	—	8.63	4.2

^aAll values determined at 25°C unless stated otherwise.

^bDetermined against liquid paraffin.

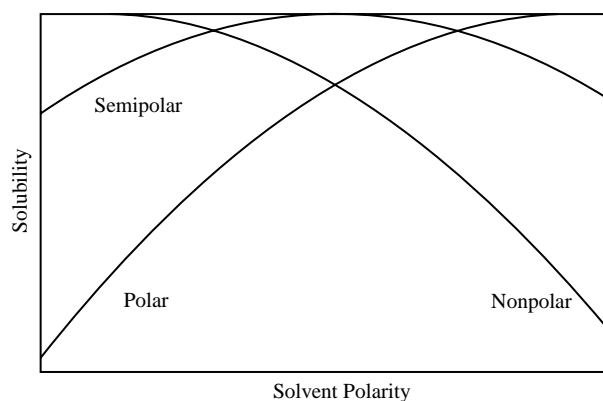


Fig. 1 Solubility vs. solvent polarity for polar, semipolar and nonpolar solutes.

Estimation of solvent composition

The most straightforward method for choosing a solvent composition is via trial and error using a list of pharmaceutically acceptable solvents. In many situations, pharmaceutical formulations can be successfully developed in this way due to the relatively limited number of pharmaceutically acceptable solvents. While this empirical approach is most useful to obtain solubility information in neat solvents, it can become cumbersome when blends of various solvents are desired.

An improvement in the purely empirical approach is the use of aligation methods that can be used to reformulate vehicles based on limited experimental formulation or solubility data. Moore (6) reported a method for the reformulation of liquid vehicles using the approximate dielectric constant ϵ of the pure and mixed solvents. For example, suppose through experimentation that a solvent containing 50% ethanol in water is capable of dissolving the required amount of drug. For some applications, it might be desired to reformulate the solution with propylene glycol–water. It is first necessary to calculate the approximate dielectric requirement (ADR) for the 50% ethanolic solution using the dielectric constants of ethanol and water from Table 1.

$$0.50 (78.5) + 0.50(24.3) = 51.4$$

The basic assumption is that any vehicle with a dielectric constant of 51.4 will solubilize a given drug to the same extent. In general, the ADR can be calculated as follows:

$$\text{ADR} = \sum_{i=1}^n \frac{(\% \text{solvent}_i \epsilon_i)}{100}$$

To calculate the amount of propylene glycol and water required for the formulation, alternate aligation can be used:

$$\begin{array}{r} 78.5 \quad 19.4 \\ 51.4 \\ 32.0 \quad \underline{27.1} \\ 46.5 \end{array}$$

Thus, the required %v/v propylene glycol is $(27.1/46.5) \times 100 = 58.3$. Alternatively, the %v/v of the new cosolvent can be solved using an algebraic method involving the solution of simultaneous equations; however, aligation is a simpler method when more than one cosolvent is to be included in the formulation. When a vehicle is to be formulated for the first time, it is necessary to experimentally determine the concentration of some cosolvent necessary to maintain the required concentration of drug in solution. This value can then be used to calculate the ADR and the final vehicle calculated as illustrated previously.

The use of the ADR method may not always provide accurate vehicle compositions for a given solute since intermolecular forces are dependent on structural characteristics of the solvent and solute that are not expressed by ϵ (7, 8). It is possible, and perhaps desirable, to substitute other measures of cosolvent polarity, such as solubility parameter, surface or interfacial tension, etc., for ϵ when blending solvents, although inaccuracies in vehicle predictions will generally continue to exist.

A second, relatively simple method for estimating the solubility of a drug in cosolvent–water mixtures was developed by Yalkowsky et al. (9). This method is based on the observation that an exponential increase in solubility of a nonpolar solute is observed as the volume fraction of cosolvent, f , increases in a cosolvent–water mixture:

$$\log S_m = S_w + \sigma f \quad (4)$$

S_m represents the solubility of drug in cosolvent–water mixture, S_w represents the solubility of the drug in water, and σ is the slope of a plot of $\log S_m$ versus f . This relationship is illustrated in Fig. 2. Eq. (4) is applicable from 0 to 100% cosolvent when a continuous increase or decrease in solubility occurs, i.e., the equation is not applicable for semipolar solutes that reach a maximum solubility between 0 and 100% cosolvent. Thus, for nonpolar or polar solutes, an estimate of the solubility in cosolvent–water mixtures can be obtained from a knowledge of the drug solubility in water and neat cosolvent, plotting these points on a semilog scale at $f = 0$

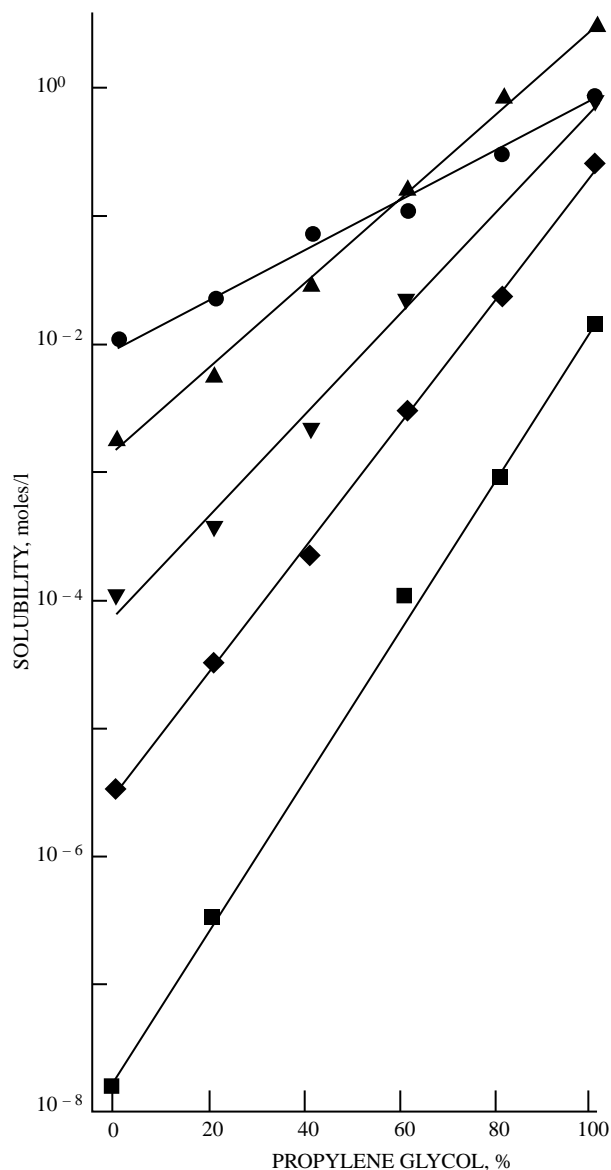


Fig. 2 Log-linear solubility relationship for a series of alkyl *p*-aminobenzoates in propylene glycol–water. (From Ref. 9.)

and $f = 1.0$, respectively, and interpolating along a straight line connecting these points.

Additional studies indicated that σ can be estimated for a solute in specific cosolvent–water combinations using the solute octanol–water partition coefficient, $\log P$, as an index of solute polarity. Thus, for propylene glycol–water mixtures (10, 11):

$$\begin{aligned}\sigma &= 0.714(\log P) + 0.714n = 382 \\ r^2 &= 0.962 \quad s = 0.325\end{aligned}\quad (5)$$

and for ethanol–water mixtures (12)^a:

$$\begin{aligned}\sigma &= 0.903(C \log P) + 0.402n = 107 \\ r^2 &= 0.955, \quad s = 0.567 \\ \delta_1 &= \sum_{i=1}^n \delta_i f_i\end{aligned}\quad (6^*)$$

These equations illustrate the concept that the slope in Eq. 4 will be greater for a given cosolvent–water mixture the more lipophilic the solute. That is, the increase in solubility relative to water will be greater the larger the value of $\log P$ for the solute. In general, solutes with a $\log P \geq 2$ will behave as nonpolar solutes in most pharmaceutical solvents and will demonstrate a continual increase in S_m with increasing f .

Alternatively, Rubino and Yalkowsky (13) found that σ was a linear function of cosolvent polarity for a given solute. This is illustrated in Fig. 3 for the three lipophilic compounds phenytoin, diazepam, and benzocaine. Thus, knowledge of the solubility of a given drug in water and at least two cosolvents would permit σ to be estimated for other cosolvents by interpolation using an index of the desired cosolvent polarity. These studies permit the use of Eq. 4 as a means to rationally choose or eliminate solvents for formulation studies based on limited experimental solubility data and commonly obtained indexes of solute and solvent polarity.

Eqs. (4–6) provide approximate solubilities and in many cases should be used to serve as aids to organize solvent selection and minimize experimental work. Viable formulations using these relationships will most likely require further refinement through experimental effort. As with the ADR method, not all intermolecular forces can be accounted for by the simple relationships presented in Eqs. (4–6). For example, experimental solubilities frequently exhibit characteristic deviations from the linear relationship predicted by Eq. 4, (14). It must also be remembered that Eqs. (4–6) pertain to nonelectrolytes or the unionized form of weak electrolytes. The behavior of salts and ions cannot be reliably estimated using these relationships.

Williams and Amidon (15–17) investigated a method that introduces estimates of solvent–solvent, and solute–solvent interactions into the basic log-linear expression. In their approach, cosolvent–water interactions are estimated from vapor pressure data of the solvent mixtures. The data are obtained from literature sources, if available, or determined experimentally. Solvent–solute interactions are estimated from experimental solubility data. An

^a $C \log P$ is a calculated value. The $\log P$ values in Eq. 5 were obtained experimentally or from a literature source.

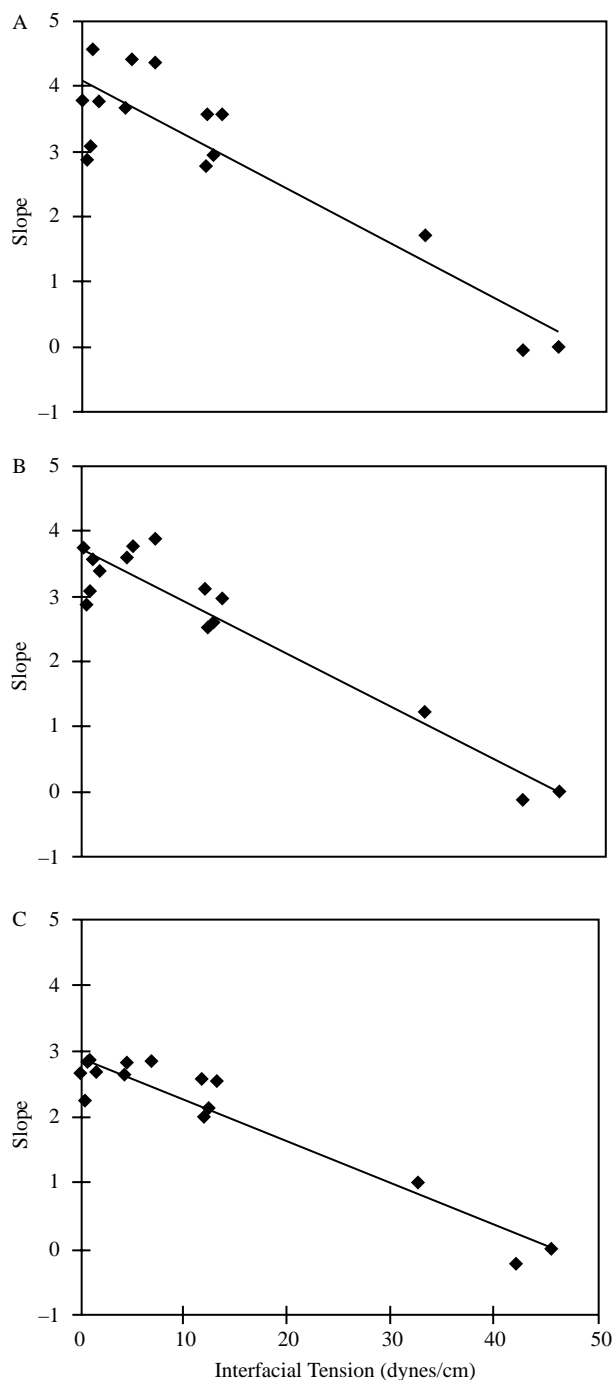


Fig. 3 Relationship between slope σ and solvent polarity, represented by the solvent interfacial tension, for three solutes. A = phenytoin, B = diazepam, C = benzocaine. (From Ref. 13.)

alternate approach, as described by Khossravi and Connors (18), divides the free energy of solubility into crystal, cavity, and solvation components. While the free energy associated with the crystal is estimated from the

solute enthalpy of fusion, the free energy associated with cavity formation is estimated from solvent surface tension data and solute molecular surface area. The solvation component is derived from a model that assumes a series of competitive exchange equilibria between the solute, water, and cosolvent. The equilibrium constants are obtained by a regression analysis of actual solubility data. These methods, previously described, are potentially more accurate in describing solubilities in mixed solvent systems, but require one or more model parameters that are usually specific to a particular solute-solvent system and must be estimated from experimental data.

Additional approaches to understand and predict solubilities in mixed solvents are based on estimation of the activity coefficient, $\log y$, in Eq. 1. Martin and coworkers (19, 20) investigated the use of regular solution theory, as developed by Hildebrand and Scott (21), to predict the solubilities of organic solutes in various solvent mixtures:

$$\log y = \frac{V_2 \phi_1^2}{2.303 RT} (\delta_1 - \delta_2)^2 \quad (7)$$

where V_2 is the molar volume of solute, Φ_1 is the volume fraction of solvent, δ_1 and δ_2 are the solubility parameters of the solvent and solute, respectively. For mixed solvents:

$$\delta_1 = \sum_{i=1}^n \delta_i f_i$$

where f is the volume fraction of solvent i . When $\delta_1 = \delta_2$, $\log y$ will equal 0 and the solution will be ideal. For mixtures of organic solvents, Eq. 7 provided the greatest agreement with experimental solubilities only when the solute and solvent polarities were not very different. In addition, solutions used as pharmaceutical dosage forms are seldom true regular solutions due to the prevalence of strong polar and hydrogen-bonding interactions. Thus, the application of Eq. 7 to pharmaceutical solutions is limited.

In order to account for polar and hydrogen-bonding interactions, the Extended Hildebrand equation was introduced by Martin et al. (22, 23):

$$\log X = \frac{-\Delta H_f}{2.303 RT} \left(\frac{T_m - T}{T_m} \right) - \frac{V_2 \phi_1^2 (\delta_1^2 + \delta_2^2 - 2W)}{2.303 RT} \quad (8)$$

where W is a factor that accounts for stronger solute-solvent interactions, such as hydrogen bonding and dipolar forces. It is determined experimentally for a given solute via regression analysis of the solubility data. Thus, a solubility database must be developed for each solute. In

addition, while solubility parameters can be readily found for various solvents, they must be either determined experimentally or estimated by calculation for a solute.

An alternative method for estimating activity coefficient involves a consideration of interfacial tension between solute and solvent. Amidon et al. (24) reported this method to estimate the solubility of organic solutes in water:

$$\log y = \frac{A_2 \gamma_{12}}{2.303RT} \quad (9)$$

where A_2 is the solute molar surface area and γ_{12} is the solute–solvent interfacial tension.

Yalkowsky et al. (25) developed the approach for mixed solvent systems:

$$\log X_m = \log X_w + \frac{[C(\gamma_{wh} - \gamma_{ch})A_h]f}{2.303 RT} \quad (10)$$

where X_m is the mole fraction solubility in the cosolvent–water mixture, X_w is the mole fraction solubility in water, γ_{wh} and γ_{ch} are the interfacial tensions between solute–water and solute–cosolvent, respectively, A_h is the hydrophobic surface area of the solute, and C is an empirical factor that corrects for the difference between the molecular and macroscopic interfaces (26). The γ terms in this expression are macroscopically determined interfacial tensions between water–tetradecane and cosolvent–tetradecane. C was found to be constant for each solute in different solvent systems and is determined experimentally (26, 27). A_h is determined by a computer program or by group contribution approaches. As predicted by Eq. 10, when the interfacial tension between a solute and solvent is reduced by addition of a cosolvent, the solubility increases relative to pure water. The similarity between Eqs. 4 and 10 can be noted and provides a further theoretical basis for Eq. 4.

The advantage of using interfacial tension as a measure of solute–solvent interactions is that it can be measured for substances whose intermolecular forces are quite different from each other. Thus, it is useful for estimating solubilities for systems that are highly “irregular.” In contrast, regular solution approaches are useful when solute and solvent polarities are similar and the interfacial tensions are immeasurable.

Dilution of Formulations Containing Cosolvents as Solubilizers

Eq. 4 predicts that an exponential increase in solubility of a nonpolar compound occurs as the volume fraction of cosolvent is increased. However, dilution of a cosolvent–solubilized preparation with an aqueous medium, such as

blood or intravenous (IV) infusion fluids, can result in the opposite phenomenon, an exponential decrease in solubility. This must be compared simultaneously to the change in drug concentration upon dilution, which decreases linearly with the degree of dilution. The effect is illustrated graphically in Fig. 4 (11). The straight lines A and B represent the solubility of the drug as a function of cosolvent concentration in two different cosolvent–water systems. The curved lines represent the concentrations of drug at various degrees of dilution. Precipitation would be expected when one of the drug concentration curves crosses a solubility curve. For three concentrations of drug in system A, it can be expected that concentration III will precipitate when it is diluted to the point where approximately 30% cosolvent is present. Formulations I and II, which contain smaller initial concentrations of drug, do not cross the solubility curve A and should be stable upon dilution. Alternatively, formulation of drug in a cosolvent or solvent mixture that produces a larger slope

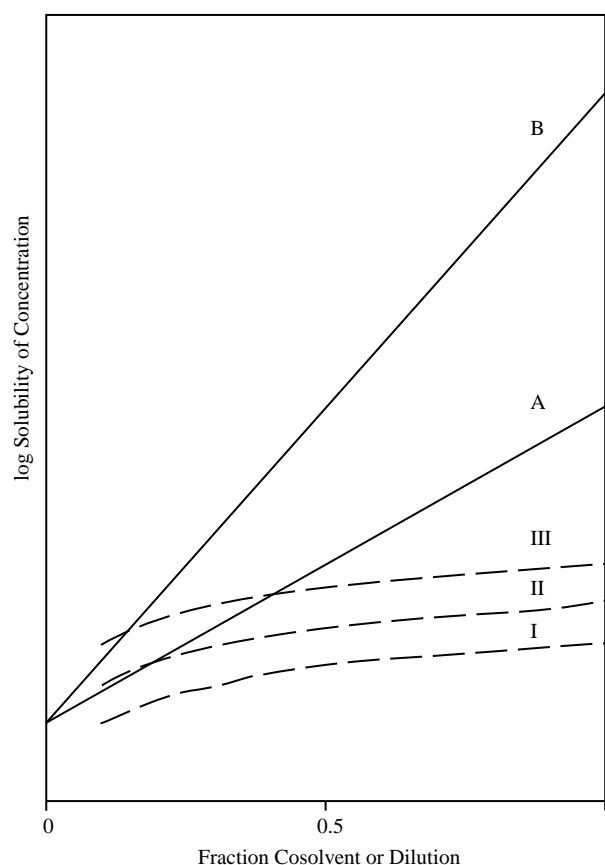


Fig. 4 Illustration of the effect of dilution of cosolvent-solubilized formulations with an aqueous medium. (See text for explanation.)

Table 2 LD 50 data for various cosolvents in mice and rats

Cosolvent	LD 50(mg/kg)					
	Mouse			Cat		
	Oral	IV	IP	Oral	IV	IP
Ethanol	7800	1973	1230	7060	1440	4070
Propylene glycol	24000	8000	9718	20000	6800	6660
Glycerin	4090	6199	8982	126000	5566	8728
PEG 400	28915	8550	9953	—	7312	9708
Dimethylacetamide	4620	3020	2800	5000	2640	2750
Dimethylsulfoxide	16500	5750	2500	17500	5360	8200

(From Ref. 48.)

of the solubility curve (B) may also minimize or reduce precipitation upon dilution. In this case, even concentration III will not precipitate upon dilution. Thus, a careful selection of cosolvent and drug concentration can prevent such occurrences upon IV injection or dilution with aqueous fluids.

BIOLOGIC EFFECTS OF COSOLVENTS

When selecting a cosolvent as a vehicle for a drug formulation, the compatibility of the solvent or solvent mixture with the appropriate tissues must be considered, as well as the potential for systemic effects. Occasional reports of systemic effects on the central nervous system (CNS, 28) or renal system (29) have been reported in humans after relatively large doses of cosolvents have been administered as part of a drug vehicle. In general, systemic effects of cosolvents are of particular interest

when they are used in drug formulations that are administered to animals during pharmacologic or safety evaluation of a drug. Budden (30) demonstrated CNS and muscle relaxant activity in mice for a number of commonly used cosolvents. It must be remembered that undesirable biological effects of a cosolvent depend on several factors that include the dose of cosolvent, route of administration, rate of administration, and concentration in the formulation. For example, a rapid bolus injection of a formulation containing a cosolvent could result in undesirable effects, which could be minimized or eliminated if administration is performed slowly or diluted with an isotonic vehicle prior to administration. When undesirable systemic or local effects of a particular cosolvent limit its use in a formulation, solvent blending that uses alternative solvents is in order.

Table 2 lists LD 50 data for various cosolvents in mice and rats and Table 3 lists the irritation data for various cosolvents when applied topically to the skin or eye. In addition, many product development scientists rely on the

Table 3 Irritation data for selected cosolvents^a

Cosolvent	Skin	Species	Eye	Species
Ethanol	400 mg (unoccluded)	Rabbit	79 mg	Rabbit
Propylene glycol	10% (2 days)	Man	100 mg (mild)	Rabbit
	500 mg (7 days, mild)	Hamster	500 mg (moderate)	Rabbit
Glycerin	500 mg (moderate)	Rabbit	126 (mild)	Rabbit
Dimethylacetamide	10 mg (mild)	Rabbit		
Dimethylsulfoxide	10 mg (unoccluded, mild)	Rabbit	100 mg	Rabbit
	500 mg (moderate)			
Isopropyl alcohol	500 mg (mild)	Rabbit	16 mg; 10 mg (moderate)	Rabbit

^aData expressed as minimum irritant dose using a 24-h (unless otherwise stated) Draize Test. Severity of reaction is listed when available. All tests performed on occluded skin unless stated otherwise.

(From Ref. 48.)

Table 4 List of parenteral products containing cosolvents

Trade name	Generic name	Manufacturer	Cosolvent composition	Route(s) of administration
BICNU	carmustine	Bristol Myers	100% dehydrated alcohol	IV infusion
Librium	chlordiazepoxide	Roche	20% propylene glycol	IM
Sandimmune	cyclosporin	Novartis	cremophor EL 65% ethanol 27.8%	IV infusion
Valium	diazepam	Roche	propylene glycol 40% ethanol 10%	IM, IV
Lanoxin	digoxin	Glaxo Wellcome	propylene glycol 40% ethanol 10%	IV preferred IM
Dimenhydrinate	dimenhydrinate	Steris	propylene glycol 50%	IM IV (req. dilution)
Ergotrate Maleate	ergonovine maleate	Lilly	ethyl lactate 10%	IM, IV
Brevibloc 250 mg	esmolol HCl	Ohmeda	propylene glycol 25% ethanol 25%	IV
Amidate	etomidate	Abbott	propylene glycol 35%	IV
VePesid	etoposide	Bristol Myers Squibb	PEG 300 65% ethanol 30.5%	IV infusion
Hydralazine HCl	hydralazine HCl	Solopak	propylene glycol 10%	IM, IV
Toradol	ketorolac tromethamine	Syntex	ethanol 10%	IM
Ativan	lorazepam	Wyeth-Ayerst	PEG 400 18% propylene glycol 80%	IM IV (req. dilution)
Alkeran	melphalan	Glaxo Wellcome	propylene glycol 60% ethanol 5%	IV
Nitrobid	nitroglycerin	Abbott, Hoechst	propylene glycol 45%	IV infusion
Taxol	paclitaxel	Marion Roussel Bristol Myers Squibb	ethanol 70% cremophor EL 52.7% ethanol 49.7%	IV infusion
Nembutal	pentobarbital	Abbott	propylene glycol 40% ethanol 10%	IM, IV
Luminal	phenobarbital sod.	Sanofi Winthrop Elkins Sinn	propylene glycol 67.8% ethanol 10%	IM, IV
Dilantin	phenytoin sod.	Parke Davis	propylene glycol 40% ethanol 10%	IM, IV
Secobarbital	secobarbital	Wyeth-Ayerst	PEG 400 50%	IM, IV
Prograf	tacrolimus	Fugisawa	polyoxyl 60 hydrogenated castor oil 20%	IV infusion
Vumon	teniposide	Bristol Myers Squibb	dimethylacetamide 6% cremophor EL 50% dehydrated ethanol 43%	IV infusion
Septra, Bactrim	trimethoprim-lfamethoxazole	Glaxo Wellcome, Roche	propylene glycol 40% ethanol 10%	IV infusion

(From Ref. 49.)

composition of past or currently marketed drug formulations as a safety guideline for new drug formulations. Tables 4–6 list the composition of various products that contain cosolvents, including products that are given parenterally, topically, and into the eye or ear.

Considerable interest exists in the local effects that cosolvents exert during IV and intramuscular (IM) injection. Several investigations have studied the hemolytic potential of cosolvents (31–35). Cosolvent-induced hemolysis of red blood cells has been studied under both static and dynamic flow conditions. Table 7 was abstracted from Krzyzaniak et al. (35) and compares the results of in vitro static and dynamic hemolysis experiments with the incidence of hemolysis observed in vivo for various vehicles containing cosolvents. In a separate publication that described the use of a dynamic method of hemolysis detection, both glycerin and propylene glycol were observed to produce significant hemolysis, while ethanol and PEG 400 did not (34). All four cosolvents were tested at a concentration of 50% in water. Propylene glycol was later reported to be nonhemolytic at a concentration less than 20% (35).

Phlebitis has been observed for several poorly soluble drugs that are formulated in cosolvent–water mixtures. Ward et al. tested several formulations of both water-soluble and water-insoluble drugs, as well as various cosolvent–water mixtures after injection in the rabbit ear (36). The cosolvent–water vehicles alone were not found to induce phlebitis. These studies indicate that the irritation was most likely caused by water–immiscible drugs.

A consideration of the local effects of cosolvents on muscle should be considered following IM injection. Brazeau and Fung (37) studied the effects of various cosolvents on muscle damage. The results are illustrated in Fig. 5. In general, glycerin (not shown) and propylene glycol produced more evidence of muscle damage, while PEG 400 was found to produce the least irritation. These results tend to parallel the hemolytic potential of cosolvents. Oshida et al. (38) suggested the use of hemolysis measurements as a way to predict muscle damage of various formulations. Brazeau and Fung also reported that the combination of PEG 400 with other, more irritating cosolvents was found to reduce the muscle damage induced by the more irritating cosolvent.

Other cosolvents, including glycerol formal, solketal, dimethylformamide, and dimethylsulfoxide, have been suggested as potential cosolvents for drug formulations; however, their safety has not been established. For additional information on these as well as other cosolvents, the reader is directed to the review by Spiegel and Noseworthy (39).

EFFECTS OF COSOLVENTS ON THE CHEMICAL STABILITY OF DRUGS

The influence of solvent polarity on the chemical stability of drugs depends on the nature of the reactants, products, and transition state of the reactants. This was discussed by Connors et al. (1) for various combinations of reactants. In general, if a vehicle is less repulsive to the transition state than another, an increase in the reaction rate will occur. Various situations arise, depending on the relative charges on the reactants. For example, when both reactants are ions (40):

$$\log k_{\varepsilon 2} = \log k_{\varepsilon 1} - \frac{Z_1 Z_2 e^2}{2.303 k T r} \left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1} \right) \quad (11)$$

where $k_{\varepsilon 1}$ and $k_{\varepsilon 2}$ are the reaction rate constants in two different solvent media of dielectric constant, ε_1 and ε_2 , respectively, Z is the ionic charge, k is the Boltzmann constant, e is the electronic charge, T is the absolute temperature, and r is the distance between the two ions. For the reaction between an ion and a dipolar molecule (41):

$$\log k_{\varepsilon 2} = \log k_{\varepsilon 1} - \frac{Z \mu e \cos \theta}{2.303 k T r^2} \left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1} \right) \quad (12)$$

where μ is the dipole moment and θ is the angle between the resultant dipole and the axis of collision of the two species. For two dipolar reactants (assuming alignment of the dipoles):

$$\log k_{\varepsilon 2} = \log k_{\varepsilon 1} - \frac{2\mu_1 \mu_2}{2.303 k T r^3} \left(\frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1} \right) \quad (13)$$

Eqs. 11, 12, 13 illustrate the concept that acceleration or deceleration of a reaction rate upon a change in solvent polarity depends on the charge of the reactants. For example, if ε_1 is water and ε_2 is an ethanol–water mixture, two oppositely charged ions will demonstrate a reduced stability in the ethanol–water system, whereas two similarly charged reactants will demonstrate an enhanced stability. This can be understood by realizing that water provides greater insulation to ions and polar species as compared to a cosolvent–water mixture. Thus, addition of cosolvents to water increases the attraction between oppositely charged species and increases the repulsion between similarly charged species.

While Eqs. 11–13 are useful in understanding the general medium effects on reaction rates, their ability to quantitatively predict reaction rates will most likely suffer from inaccuracies similar to those experienced with solubility prediction in mixed solvents. Dielectric constant

Table 5 Examples of ophthalmic and otic products containing cosolvents

Product name (manufacturer)	Active component(s)	Cosolvent composition
Chloroptic (Allergan)	Chloramphenicol	PEG 300
Americaine Otic (Medeva)	Benzocaine	PEG
Debrox Drops (Marion)	Carbamide peroxide	Glycerin, propylene glycol
Auralgan Otic (Wyeth-Ayerst)	Antipyrine, benzocaine, glycerin	Glycerin
VoSol Otic (Wallace)	Acetic acid	propylene glycol

(From Ref. 50.)

as a single parameter is not capable of quantitating all interactions among solvent and solute molecules. LePree and Connors (42) reported on the use of a phenomenological approach to predict reaction rates in mixed solvents. The approach is similar to the solubility studies reported by Khossravi and Connors (18) and uses complex equilibria to characterize solvent–solute interactions.

When water is a reactant in the degradation process, that is, when the reaction is hydrolytic, replacement of all or part of the water with a cosolvent may enhance the stability of the drug. This does not imply that solvolysis will not occur with solvents other than water. Indeed, many pharmaceutical solvents or solvent blends are sufficiently nucleophilic to participate in substitution reactions; however, they are often less reactive than pure water (43).

When cosolvents are used to enhance solubility, the potential effects on drug stability must be remembered. The benefit of the enhanced solubility must be weighed against potentially undesirable effects on drug stability.

Cosolvents may also contain impurities that can initiate or catalyze oxidative degradation. PEGs are well-known initiators due to their ability to form unstable peroxides (44). Likewise, the heavy metal content of a cosolvent can vary from vendor to vendor or lot to lot. The use of such cosolvents in formulations of drugs susceptible to oxidation requires that strict limits be placed on specifications for the heavy metal or peroxide content of the cosolvent.

EXAMPLES OF THE USE OF COSOLVENTS FOR THE FORMULATION OF LIQUID DOSAGE FORMS

Oral Dosage Forms

Ethanol has been used traditionally as a cosolvent for oral solutions. It has been incorporated with sucrose in elixir

Table 6 Examples of topical products containing cosolvents

Product name (manufacturer)	Active component(s)	Cosolvent composition
Dermal Preparations		
Benadryl Itch-Stopping Spray (Warner-Lambert)	diphenhydramine HCl	alcohol, glycerin
Erythromycin topical (various)	Erythromycin	55–77% ethanol propylene glycol
Lotrimin solution (Schering)	Clotrimazole	PEG 400
Diprolene lotion (Schering)	Betamethasone dipropionate	30% isopropanol propylene glycol
Lamisil (Novartis)	terbinafine HCl	28.7% ethanol propylene glycol
Effudex (ICN)	5-fluorouracil	propylene glycol
Lidex (Medicis)	Fluocinonide	35% alcohol propylene glycol
Lotrimin AF Solution (Schering-Plough)	miconazole	PEG
Rogaine (Pharmacia-Upjohn)	minoxidil	60% alcohol propylene glycol
Mouth and Throat Preparations		
Anbesol (Whitehall-Robbins)	Benzocaine, phenol	PEG propylene glycol
Glyoxide (SKB)	Carbamide peroxide	glycerin propylene glycol
Listerine (Warner-Lambert)		21.6% alcohol

(From Ref. 50.)

Table 7 Hemolytic effects of cosolvents

Formulation composition	In vivo	In vitro method (% hemolysis)	
		Reed and Yalkowsky	Krzyzaniak et al ^a
normal saline (NS)	no	0.0	0.0
10% ethanol in NS	no	0.0	0.7
30% ethanol in NS	no	0.0	0.5
40% PG in NS	yes	61.0	5.6
60% PG in water	yes	100.0	9.5
10% PG + 30% ethanol in NS	no	0.0	1.2
10% ethanol + 20% PG in water	no	8.8	2.0
10% ethanol + 40% PG water	yes	69.2	10.3
20% ethanol + 30% PEG 400 in water	no	0.0	0.3

^aVehicles with % hemolysis > 2 considered hemolytic in vivo.
(From Ref. 35.)

formulations, the alcoholic content of which may vary from 3–78%. These formulations resulted in solubilization and/or stabilization of various drugs and have more favorable taste as compared to that of other solvents. Its use is often undesirable, however, in oral preparations intended for pediatric patients or other patients who cannot tolerate the effects of ethanol. Ethanol may also accentuate the saline taste of ionic solutes (45). Propylene glycol has been suggested as an appropriate substitute for ethanol in oral solutions; however, its use in pediatric formulations should be carefully examined in light of previous reports of toxicity (46). In addition, the objectionable taste of propylene glycol may require the addition of flavoring

agents. Liquid PEGs have been used as vehicles for soft gelatin capsules as well as pediatric elixir formulations.

Parenteral Dosage Forms

The use of cosolvents in small-volume parenteral preparations is often critical due to the limited volume of solution that can be administered by a single injection. Thus, the required dose of drug must often be incorporated in 1 or 2 mL of solution. Table 6 lists parenteral products containing cosolvents. The cosolvents most often used include ethanol, propylene glycol, glycerin, PEG 400, and, sometimes, dimethylacetamide. Other cosolvents, such as DMSO, have been used as solvents for parenteral formulations of experimental anticancer agents; however, their use is restricted due to toxicity and potential incompatibilities with plastic administration devices (47).

Irritation, and hemolysis are primary considerations when choosing a cosolvent for parenteral preparations. Concentration and route of administration are important factors that determine the incidence and severity of local reactions.

Ophthalmic and Otic Dosage Forms

Ophthalmic formulations sometimes contain cosolvents, such as propylene glycol or PEG 300, as part of the vehicle. The greatest limitation to the use of cosolvents in ophthalmic preparations is their irritation potential. For example, ethanol is too irritating to be used in the eye (Table 3). Osmotic effects of cosolvents are also important, and the strong osmotic effect of glycerin combined with its poor solubilizing power limit its usefulness in ophthalmic preparations.

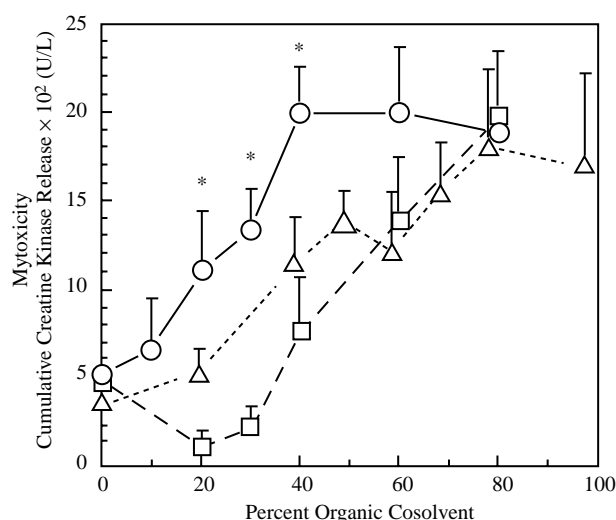


Fig. 5 Myotoxicity of various cosolvents as a function of % cosolvent in water. ○ = propylene glycol, Δ = ethanol, □ = Polyethylene glycol 400. (From Ref. 37.)

Propylene glycol, PEG, glycerin, and isopropyl alcohol have been used in otic formulations. Table 5 contains a list of ophthalmic and otic preparations containing cosolvents.

Topical Dosage Forms

Liquid preparations intended for dermal application contain the largest variety of cosolvents. They most commonly include ethanol, isopropanol, propylene glycol, glycerin, and PEG 400. Irritation and sensitization are important considerations in choosing a cosolvent for dermal use. In addition, it may be necessary to test new compounds for photoirritation or photosensitization reactions.

Other cosolvents, such as DMSO, demonstrate skin penetration enhancement properties for a number of compounds. Although this could be a highly desirable property for many drugs, the use of DMSO as a solvent for dermal application has not been approved.

Preparations that are applied topically to the mouth and throat have contained glycerin, ethanol, propylene glycol, and PEG. Both systemic and local toxic effects would need to be considered in choosing a solvent system for these preparations. Table 6 contains a list of dermal and topical mouth and throat products that contain cosolvents.

REFERENCES

- Connors, K.A.; Amidon, G.L.; Kennon, L. *Chemical Stability of Pharmaceuticals*; Wiley Interscience: New York, 1979; 39–42.
- Taft, R.W.; Kamlet, M.J. The Solvatochromic Comparison Method I. The α Scale of Solvent Hydrogen Bond Donor (HBD) Acidities. *J. Am. Chem. Soc.* **1976**, *98* (10), 2886–2894.
- Taft, R.W.; Kamlet, M.J. The Solvatochromic Comparison Method II. The β Scale of Solvent Hydrogen Bond Acceptor (HBA) Basicities. *J. Am. Chem. Soc.* **1976**, *98* (2), 377–383.
- Beerbower, A.; Hansen, C. Solubility Parameters. *Encyclopedia of Chemical Technology*; Standen, E.A., Ed.; Interscience: New York, 1971; 889–910.
- Martin, A.; Wu, R.L.; Adjei, A.; Beerbower, A.; Prausnitz, J.M. Extended Hansen Solubility Approach: Naphthalene in Individual Solvents. *J. Pharm. Sci.* **1981**, *70* (11), 1260–1264.
- Moore, W.E. The Use of Approximate Dielectric Constant to Blend Solvent Systems. *J. Am. Pharm. Assoc. Sci.* **1958**, *47* (10), 855–857.
- Paruta, A.N.; Sciarrone, B.J.; Lordi, N.G. Solubility of Salicylic Acid as a Function of Dielectric Constant. *J. Pharm. Sci.* **1964**, *53* (11), 1349–1353.
- Sorby, D.L.; Bitter, R.G.; Webb, J.G. Dielectric Constant of Complex Pharmaceutical Solvent Systems I. Water–Ethanol–Glycerin and Water–Ethanol–Propylene Glycol. *J. Pharm. Sci.* **1963**, *52* (12), 1149–1153.
- Yalkowsky, S.H.; Flynn, G.L.; Amidon, G.L. Solubility of Nonelectrolytes in Polar Solvents. *J. Pharm. Sci.* **1972**, *61* (6), 983.
- Yalkowsky, S.H.; Rubino, J.T. Solubilization by Cosolvents I: Organic Solutes in Propylene Glycol–Water Mixtures. *J. Pharm. Sci.* **1985**, *74* (4), 416–421.
- Yalkowsky, S.H.; Roseman, T.J. Solubilization of Drugs by Cosolvents. *Techniques of Solubilization of Drugs*; Yalkowsky, S.H., Ed.; Chap. 3. Marcel Dekker, Inc.: New York, 1981.
- Li, A.; Yalkowsky, S.H. Solubility of Organic Solutes in Ethanol/Water Mixtures. *J. Pharm. Sci.* **1994**, *83* (12), 1735–1740.
- Rubino, J.T.; Yalkowsky, S.H. Cosolvency and Cosolvent Polarity. *Pharm. Res.* **1987**, *4* (3), 220–230.
- Rubino, J.T.; Yalkowsky, S.H. Cosolvency and Deviations from Log-Linear Solubilization. *Pharm. Res.* **1987**, *4* (3), 231–236.
- Williams, N.A.; Amidon, G.L. Excess Free Energy Approach to the Estimation of Solubility in Mixed Solvent Systems I. Theory. *J. Pharm. Sci.* **1984**, *73* (1), 9–13.
- Williams, N.A.; Amidon, G.L. Excess Free Energy Approach to the Estimation of Solubility in Mixed Solvent Systems II: Ethanol–Water Mixtures. *J. Pharm. Sci.* **1984**, *73* (1), 14–18.
- Williams, N.A.; Amidon, G.L. Excess Free Energy Approach to the Estimation of Solubility in Mixed Solvent Systems III: Ethanol–Propylene Glycol–Water Mixtures. *J. Pharm. Sci.* **1984**, *73* (1), 18–23.
- Khossravi, D.; Connors, K.A. Solvent Effects on Chemical Processes I: Solubility of Aromatic and Heterocyclic Compounds in Binary Aqueous–Organic Solvents. *J. Pharm. Sci.* **1992**, *81* (4), 371–379.
- Chertkoff, M.J.; Martin, A.N. The Solubility of Benzoic Acid in Mixed Solvents. *J. Pharm. Sci.* **1960**, *49* (5), 444–447.
- Restaino, F.A.; Martin, A.N. Solubility of Benzoic Acid and Related Compounds in a Series of *N*-Alkanols. *J. Pharm. Sci.* **1964**, *53* (6), 636–639.
- Hildebrand, J.H.; Scott, R.L. *Solubility of Nonelectrolytes*; Chap. 3. Reinhold Publishing Co.: New York, 1950.
- Martin, A.N.; Newberger, J.; Adjei, A. New Solubility Equation. *J. Pharm. Sci.* **1979**, *60* (10), IV.
- Martin, A.N.; Newberger, J.; Adjei, A. New Solubility Equation. *J. Pharm. Sci.* **1980**, *69* (5), 487–491.
- Amidon, G.L.; Yalkowsky, S.H.; Leung, S. Solubility of Nonelectrolytes in Polar Solvents II: Solubility of Aliphatic Alcohols in Water. *J. Pharm. Sci.* **1974**, *63* (12), 1858–1866.
- Yalkowsky, S.H.; Amidon, G.L.; Zografi, G.; Flynn, G.L. Solubility of Nonelectrolytes in Polar Solvents III: Alkyl *p*-Aminobenzoates in Polar and Mixed Solvents. *J. Pharm. Sci.* **1975**, *64* (1), 4852.
- Yalkowsky, S.H.; Valvani, S.C.; Amidon, G.L. Solubility of Nonelectrolytes in Polar Solvents IV. Nonpolar Drugs in Mixed Solvents. *J. Pharm. Sci.* **1976**, *65* (10), 1488–1494.
- Zografi, G.; Yalkowsky, S.H. Interfacial Properties of Polar Liquids Against Nonpolar Phases. *J. Pharm. Sci.* **1974**, *63* (10), 1533–1536.

28. Van de Wiele, B.; Rubinstein, E.; Peacock, W.; Martin, N. Propylene Glycol Toxicity Caused by Prolonged Infusion of Etomidate. *J. Neurosurg. Anesthesiol* **1995**, *7* (4), 259–262.
29. Yorgin, P.D.; Tehodorou, A.A.; Al-Uzri, A.; Davenport, K.; Boyer-Hassen, L.V.; Johson, M.I. Propylene Glycol-Induced Proximal Renal Tubular Cell Injury. *Am. J. Kidney Dis.* **1997**, *30* (1), 134–139.
30. Budden, R.; Kuhl, V.G.; Bahlsen, J. Experiments on the Toxic, Sedative and Muscle Relaxant Potency of Various Drug Solvents in Mice. *Pharm. Ther. Part B* **1979**, *5*(1–3), 467–474.
31. Cadwallader, D.W. Behavior of Erythrocytes in Various Solvent Systems I. Water–Glycerin and Water–Propylene Glycol. *J. Pharm. Sci.* **1963**, *52* (12), 1175–1180.
32. Fort, E.L.; Heyman, I.A.; Kesterson, J.W. Hemolysis Study of Aqueous Polyethylene Glycol 400 Propylene Glycol and Ethanol Combinations In Vivo and In Vitro. *J. Parenter Sci. Technol.* **1984**, *38* (2), 82–86.
33. Reed, K.W.; Yalkowsky, S.H. Lysis of Human Red Blood Cells in the Presence of Various Cosolvents. *J. Parenter Sci. Technol.* **1985**, *39* (2), 64–86.
34. Krzyzaniak, J.F.; Raymond, D.M.; Yalkowsky, S.H. Lysis of Human Red Blood Cells 2: Effect of Contact Time on Cosolvent Induced Hemolysis. *Int. J. Pharm.* **1997**, *152* (2), 193–200.
35. Krzyzaniak, J.F.; Nunez, F.A. Alvarez; Raymond, D.M.; Yalkowsky, S.H. Lysis of Human Red Blood Cells. 4. Comparison of In Vitro and In Vivo Hemolysis Data. *J. Pharm. Sci.* **1997**, *86* (11), 1215–1217.
36. Ward, G.H.; Nolan, P.E.; Chawla, M.; Yalkowsky, S.H. Studies in Phlebitis: Detection and Quantitation Using a Thermographic Camera. *Pharm. Res.* **1991**, *8* (1), 76–79.
37. Brazeau, G.A.; Fung, H.L. Use of an in Vitro Model for the Assessment of Muscle Damage from Intramuscular Injections: In Vitro–In Vivo Correlation and Predictability with Mixed Solvent Systems. *Pharm. Res.* **1989**, *6* (9), 766–771.
38. Oshida, S.; Degawa, K.; Takahoshi, Y.; Akaishi, S. Physicochemical Properties and Local Toxic Effects of Injectables. *Tohoku. J. Exp. Med.* **1979**, *127* (4), 301–316.
39. Spiegel, A.J.; Noseworthy, M.M. Use of Nonaqueous Solvents in Parenteral Products. *J. Pharm. Sci.* **1963**, *52* (10), 971–927.
40. Amis, E.S. Coulomb's Law and the Quantitative Interpretation of Reaction Rates. *J. Chem. Educ.* **1952**, *29* (July), 337–339.
41. Amis, E.S. Coulomb's Law and the Quantitative Interpretation of Reaction Rates II. *J. Chem. Educ.* **1953**, *30* (July), 351–353.
42. LePree, J.M.; Connors, K.A. Solvent Effects on Chemical Processes. 11. Solvent Effects on the Kinetics of Decarboxylative Dechlorination of N-chloro Amino Acids in Binary Aqueous-Organic Solvents. *J. Pharm. Sci.* **1996**, *85* (6), 560–566.
43. Lowry, T.H.; Richardson, K.S. *Mechanism and Theory in Organic Chemistry*, 3rd Ed.; Harper Collins Publishers: New York, 1987; 361–363.
44. McGinity, J.W.; Hill, J.A.; La Via, A.L. Influence of Peroxide Impurities in Polyethylene Glycols on Drug Stability. *J. Pharm. Sci.* **1975**, *64* (2), 356–357.
45. Parnarowski, M. Solutions, Emulsions & Suspensions. *Remington's Pharmaceutical Sciences*, 14th Ed.; Osol, A., Hoover, J.E., Eds.; Mack Publishing Co.: 1970; 1488–1489.
46. Arulanantham, K.; Genel, M. Central Nervous System Toxicity Associated with Ingestion of Propylene Glycol. *J. Pediatr.* **1978**, *93* (3), 515–516.
47. Mojavarion, R.; Repta, A. Development of an Intravenous Formulation for the Unstable Investigational Cytotoxic Nucleosides 5-Azacytosine Arabinoside (NSC 281272) and 5-Azacytidine (NSC 102816). *J. Pharm. Pharmacol.* **1984**, *36* (11), 728–733.
48. Tatken, R.J.; Lewis, R.J. Registry of Toxic Effects of Chemical Substances. *U.S. Dept. Health and Human Services*; U.S. Government Printing Office: Washington, DC, 1986.
49. Trissel, L.A. *Handbook on Injectable Drugs*, 9th Ed.; American Society of Health-System Pharmacists, Inc.: Bethesda, MD, 1996.
50. *Physicians' Desk Reference for Nonprescription Drugs*, 19th Ed.; Medical Economics Co., Inc.: Montvale, NJ, 1998.

COSMETICS AND THEIR RELATION TO DRUGS

Martin M. Rieger

M&A Rieger, Morris Plains, New Jersey, U.S.A.

INTRODUCTION

Under the Food, Drug, and Cosmetics Act (FDCA, 21 U.S. Code 301), the Food and Drug Administration (FDA) of the United States has the authority to regulate foods, prescription (Rx) drugs, over-the-counter (OTC) drugs, and cosmetics. The FDA also administers a second statute, the Fair Packaging and Labeling Act (FPLA, 15 U.S. Code 1400). In order to administer this complex task, the FDA depends on the legalistic and statutory definitions of drug and cosmetics in the FDCA. Nevertheless, the public's interpretation of what constitutes a drug or a cosmetic may differ somewhat from that of regulatory agencies. Philosophically and historically, a cosmetic is a product that helps improve external appearance and has the ability to hide, or at least distract from, unwanted stigmata or skin defects.

A product that changes the color of hair is a cosmetic, as is a product intended to increase the skin's tendency to tan by exposure to sun. This traditional view remains ingrained in the consumer's mind but may not be judicially valid. A change in hair color, for example, can be effected by the following: 1) a wig, which might be viewed as an article of clothing; 2) a variety of dyeing processes, which are properly identified as cosmetic changes; and 3) possibly by a variety of ingested or topically applied substances that gradually alter the hair follicle's ability to synthesize melanin, which should be classified as a drug effect. The common goal of these three approaches is to effect a change in appearance, the key objective of all cosmetics. The method by which this goal is achieved differentiates the three hair "coloring" processes and makes a product a drug or a cosmetic. This can create some confusion, as is demonstrated by a consideration of sunscreen products. Sunburn prevention by topical products was for years considered within the scope of cosmetics, even though ultraviolet-B (UV-B) light absorbers were incorporated into these "cosmetics." The cosmetic industry responded rather calmly when the FDA's review of OTC drugs included suntan preparations and sunburn preventives. What for years had been a cosmetic suddenly became a drug by legislative or administrative fiat. The FDA's rationale is justifiably

based on the concept that sunburn prevention is prevention of disease. Adding an ingredient that enhances the ability of melanocytes to produce melanin in the skin would be viewed as a cosmetic by the user. The FDA is likely to accept cosmetic (color change, appearance) claims for such a product, but a definition of a drug would become mandatory if the melanin is claimed to protect against sunburn. The implication that a parasol intended to prevent exposure to sun is a medical device has not been judicially examined.

One must recognize that the differentiation between cosmetics and drugs is complex and is blurred by the interplay of consumer perception, commercial interest, and statutory interpretation by regulatory agencies, with the ultimate decision in the hands of the judiciary. For these reasons, differences between drugs and cosmetics are discussed in the next section on the basis of existing U.S. laws, the product's composition, and safety and efficacy. Laws and rules covering the distinction between cosmetics and drugs differ from country to country. For this reason, marked divergence from U.S. practices will be noted in this survey.

COMPARISON ON THE BASIS OF U.S. LAW

Definitions

The sharpest distinction between a drug and a cosmetic is based on the statutory definitions in the Federal Food Drug and Cosmetic Act (21 USC 301 et seq.). Cosmetics are clearly defined as:

- 1) articles intended to be rubbed, poured, sprinkled, or sprayed on, introduced into, or otherwise applied to the human body or any part thereof for cleansing, beautifying, promoting attractiveness, or altering the appearance, and 2) articles intended for use as a component of any such articles; except that such term shall not include soap.

On the other hand, drugs are defined as follows:

The term drug means: (A) articles recognized in the official United States Pharmacopoeia, official

Homeopathic Pharmacopoeia of the United States or official National Formulary or any supplement to any of them; and (B) articles intended for the use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or other animal; and (C) articles (other than food) intended to affect the structure or any function of the body of man or other animals; and (D) articles intended for use as a component of any articles specified in clause (A), (B), or (C); but does not include devices or their components, parts, or accessories.

These definitions may differ from the interpretation of the consumer or from generally accepted usage, but courts will adjudicate exclusively on the basis of these statutory definitions. It is clearly the intent of the product, not necessarily its performance, that is used judicially to classify a product as a drug or as a cosmetic. A skin-care product intended to beautify by the removal of wrinkles is both a cosmetic (alters the appearance) and a drug (affects a body structure). Historically and intuitively, the requirements for a drug are more stringent than those for a cosmetic, and the regulatory agency and the courts tend to apply the more stringent requirement to a product that may be perceived to be both a drug and a cosmetic. Table 1 includes a listing of products classified by the FDA as cosmetics. Some of the products are considered drugs or quasi-drugs in other countries.

The Food and Drug Administration's Tasks

The FDA is authorized to enforce the FDCA and the FPLA. The FDA's tools include inspection and seizure, which may be applied equally to drugs or cosmetics.

The FDCA prohibits the use (or presence) of poisonous or deleterious substances. Their presence makes a cosmetic "adulterated" or "misbranded." In this regard, no significant distinction is made between drugs and cosmetics. Similarly, good manufacturing practices (GMPs) are applicable to drugs and with minor changes to cosmetics. Products that are manufactured under conditions that are in violation of the GMPs may become subject to seizure. In recent years, the FDA has not initiated formal cases against violators of cosmetic regulations; instead, the FDA has relied on so-called Warning Letters to obtain compliance without recourse to complicated legal action.

In contrast to Rx drugs, OTC drugs and cosmetics are not subject to preclearance. Preclearance is specifically designed to prevent the introduction of dangerous or undesirable drug entities into the market. The restrictions on ingredients are most severe in the case of OTC drugs and preclude introduction of untested drugs or combinations. In fact, a "new chemical entity," which might be entirely suitable for introduction as an OTC drug, requires workup via the new drug application (NDA) process. The approval of Rx and OTC drugs by the FDA is, in principle, based on the performance of the drug entity. Efficacy against the disease, bioavailability, and lack of adverse side effects are of primary importance. Thus, judicious choice of drug excipients is required for all drug approvals. Nevertheless, in the absence of an FDA-approved list of cosmetic ingredients, the cosmetic manufacturer has the responsibility to provide products that are not injurious to the user under the expected conditions of use. Some ingredients are specifically restricted, and a regulation requiring safety substantiation exists. These will be discussed in the section entitled "Restrictions on the Use of Ingredients."

Table 1 List of products (recognized as cosmetics in the United States)

Baby preparations	Creams, lotions, oil, powders, shampoos
Bath preparations	Bubble baths, capsules, oils, salts, soaps and detergents, tablets
Cleansing preparations	Creams, douches, liquids and pads, lotions, personal cleansing products
Dentifrices	Aerosols, breath fresheners, liquids, mouthwashes, pastes, powders
Fragrance products	Colognes and toilet waters, deodorants, fragrances, perfumes
Hair products	Depilatories, dressings, dyes and colors, grooming aids, lighteners, miscellaneous rinses, permanent wave products, shampoos, sprays, straighteners, tints, tonics, wave sets
Makeup preparations	Blushers, eyebrow pencils, eyeliners, eye makeup preparations, eye makeup removers, eye shadows, face powders, facial makeups, fixatives, foundation makeups, leg and body paints, lip glosses, lipsticks, mascaras, rouges
Miscellaneous products	Paste masks, powders (men's, women's, talcums)
Shaving preparations	Aftershaves, beard softeners, shaving creams (aerosol, brushless lather), preshaves
Skin care preparations	Body and hand preparations (moisturizers), eye creams, face and neck preparations, fresheners and astringents, suntan gels

Color Additives

Color additives are of particular importance to the formulation of cosmetics. Dyes and pigments not only make products more attractive but also are vital to any product that is intended to alter the color of any part of the body. Color additives are regulated meticulously by the FDA, and only some general information on the current regulatory status of colorants in the United States can be provided.

Certified color additives are synthetic organic dyes that are described in an approved color additive petition. Each manufactured lot of a certified dye must be analyzed and certified by the FDA prior to usage.

Color lakes are pigments that generally consist of an insoluble metallic salt of a certified color additive deposited on an inert substrate. These lakes are subject to the color additive regulations of the FDA and must be certified by the agency prior to use.

Color additives that are not classified as certified color or color lakes are identified as *noncertified color additives*. Each of these substances is the subject of an approved color additive petition, but individual batches do not require certification by FDA prior to use.

The fourth major class of color additives is *hair colorants*. These compounds or their mixtures may be used only to color scalp hair and may not be used in the eye area. Use of these colorants is “exempt,” that is, the so-called coal-tar hair dyes may be sold with cautionary labeling, directions for preliminary (patch) testing, and restrictions against use in or near the eye.

Soap Exclusion

Soap is specifically excluded from cosmetics in the FDCA, and no cosmetic or drug regulations are applicable to soap. The FDCA fails to define the term *soap*, but the FDA has ruled that a product is a soap if the bulk of the nonvolatile matter is the alkali salt of a fatty acid and if its deterative properties are due exclusively to the fatty acid salt. In addition, the product must be labeled as a soap. A product is identified as a shampoo when it consists, e.g., only of aqueous potassium oleate. It then must conform to cosmetic regulations.

The term *soap* thus has two meanings. The first is the FDA’s definition, which is used for legal purposes. The second is the generic sense, whereby soaps may refer to cleansing products that may not meet the specifics of FDA’s definition. Such products must, therefore, be labeled as cosmetics. The Federal Trade Commission (FTC) and the Consumer Products Safety Commission (CPSC) handle the regulatory control for soaps.

COMPARISON ON THE BASIS OF COMPOSITION

Drug Ingredients Versus Cosmetic Ingredients

On the basis of the Drug Efficacy Study Implementation (DESI) review, which began in 1962, the FDA ultimately concluded that of about 16,000 claims made for 3400 Rx drugs, only about 2300 drugs were effective for at least one indication. Today, Rx drugs must undergo the NDA process, which, for all practical purposes, is a critical preclearance procedure. As the work on the DESI review neared completion, the FDA initiated the so-called OTC review in 1972. The FDA classified some 250,000 drugs into about 55 therapeutic groups. The panels that reviewed each group had the responsibility to establish the safety and efficacy of each OTC drug and to restrict claims for these drugs to those the panel considered appropriate for a given drug or combination of drugs. It is apparent that the marketability of a drug—Rx or OTC—requires the presence and bioavailability of an identifiable drug entity that can be expected to exert some therapeutic benefit. No such legal requirement for the use of raw materials exists in cosmetics. As a rule, the presence of and claim for any component in a cosmetic that may have a therapeutic effect converts such a cosmetic into a drug.

The FDA has classified the following topically applied products as OTC drugs on the basis of safety and efficacy review of the drug(s) constituents:

- Acne products
- Antidandruff products
- Antimicrobial products
- Antiperspirant products
- Astringent products
- Oral care products
- Skin-protectant products
- Sunscreen products
- External analgesic products

In other countries, some of these products are considered cosmetics.

Some of the actives used in the past in such products have been classified as Category I, i.e., safe and effective. Usage of these agents and the claims made for the finished product make these products OTC drugs, not cosmetics.

The activities of the OTC panels are not yet completed, although most of the tentative final reports have been published. However, no definitive rulings have been made or subjected to judicial review. It appears at this time that the ingredients reviewed by the OTC panels can be used in cosmetics as excipients and the like. To repeat, their use, together with drug or therapeutic claims, transforms the cosmetic into a drug, in which case the labeling and claim

structure must conform to those established for OTC drugs. The designation "cosmetic" places almost no restriction on the use of components. However, claims for therapeutic efficacy convert any cosmetic into a drug, as interpreted by the FDA. The regulations do not restrict the use of a drug substance for purposes unrelated to its drug status.

COSMETICS AND THEIR RELATION TO DRUGS

Restrictions on the Use of Ingredients in Cosmetics

The review of active drugs by the OTC panels was limited to relatively few drug entities, but the cosmetic industry employs thousands of ingredients, including many of plant and animal origin. Many typical cosmetic ingredients are identical to the components used in Rx and OTC drugs, but only very few cosmetic ingredients are subject to restrictions by the FDA. These include mercury compounds, except those used as preservatives in products intended for use in or near the eye. Others are bithionol, vinyl chloride, halogenated salicylanilides, zirconium compounds in aerosol products, chloroform, chlorofluor-carbon propellants, and hexachlorophene. With regard to cosmetic ingredients, the FDA has placed the responsibility for substantiating their safety squarely on the producer. Such safety substantiation also includes finished products. Each ingredient used in a cosmetic product and each finished cosmetic product shall be adequately substantiated for safety prior to marketing. Any such ingredient or product whose safety is not adequately substantiated prior to marketing is misbranded unless it contains the following conspicuous statement on the principal display panel: Warning—The safety of this product has not been substantiated [21 CFR, Par. 740].

These regulatory activities and the need to demonstrate to the public that the cosmetic industry as a whole is prepared to accept responsibility prompted the Cosmetics, Toiletries and Fragrance Association (CTFA) in 1976 to establish the Cosmetic Ingredient Review (CIR) for the purpose of evaluation and review of the safety of the ingredients used in cosmetics. The CIR process established a system for prioritizing ingredients based on frequency of use, concentration used, area of use, frequency of application, use by sensitive subgroups, likelihood of biologic activity, and consumer complaints. This is a scientific review of worldwide data, and nonvoting members from industry and consumer groups participate in the deliberations. In order to speed up the review process and to avoid duplication, the CIR expert panel may defer study of substances that are already under review by other

safety programs. The most important of these are those by the Research Institute for Fragrance Materials (RIFM) and the Flavor and Extract Manufacturers Association (FEMA). The former establishes the safety of fragrance components; its funding and concept permit expenditures for safety testing of substances. The latter addresses issues related to the safety of individual flavor materials. The reviews by the CIR panel are available from the CTFA (1101 17th St., N. W., Washington, D.C., 20036-4702) and have appeared over a period of years in the *Journal of the American College of Toxicology*.

Although the CIR process is sponsored by the CTFA, the latter will conduct safety studies for substances considered crucial to the survival of the cosmetic industry. The CTFA will also establish and pay for research required to confirm the stability, chemical purity, and safety of various cosmetic ingredients. These activities are part of the cosmetic industry's technically oriented self-regulation program. Similar programs exist in the pharmaceutical industry.

RIFM has recommended discontinuance of the use of acetyethyl tetramethyltetralin, 6-methylcoumarin, musk ambrette, and musk ketone. Another self-imposed ingredient restriction concerns the use of potential nitrosating agents in products containing various (secondary) alkanol amines, in light of the hazard of nitrosamine formation. In the European Union, (E.U.), the use of *tri*- and *di*-ethanolamine is restricted.

One of the key self-regulatory procedures in the cosmetic industry is the voluntary reporting process of adverse reactions. The program is intended to provide data on the type and frequency of adverse reactions reported by consumers or by their medical advisors to the industry. It is an important means of detecting problems that are not treated in hospital emergency rooms (i.e., documented in the National Electronic Injury Surveillance System, NEISS) or do not reach poison control centers.

Practical Aspects

For practicing formulators, the border between cosmetics and drugs is not clear. Drug entities for which claims are made on the label differentiate drugs from cosmetics. However, the same or chemically similar excipients and formulation aids are widely used in cosmetics and drugs. Finally, one must recognize that cosmetics may include groups of substances not normally found in drugs. These similarities and differences are illustrated in Table 2. Substances considered drugs by U.S. law have been excluded from the table. This listing is not comprehensive and is presented for illustrative purposes only.

Table 2 Common usage of selected ingredients in cosmetics and drugs and cosmetics

Function	Cosmetics (primarily)	Cosmetics and Drugs
Abrasive	Oatmeal	Dicalcium phosphate
Absorbent		Kaolin, silica
Antifoam		Simethicone
Antioxidant		Ascorbic acid, BHA, BHT, tocopherol
Antistatic	N-Lauroyl- β -Alanine, stearyl/dimethylbenzylammonium chloride [INCI nomenclature: CAS 68650-39-5]	
Binder		Hydrophilic gums, Polyvinyl acetate, starch
Bulking agent		Cellulose, silica
Chelator		Citrates, EDTA and salts, glucuronic acid
Cleanser	Triethanolamine laurylethersulfate, lauryl betaine	Sodium Laurylsulfate, poloxamer 188
Coemulsifier	Beeswax	Fatty alcohols
Emollient	Apricot kernel oil	Cocoa butter, dibutyl sebacate, dioctyl adipate, isopropyl myristate, lanolin, mineral oil, squalene
Emulsifier	Sodium lauroyllactylate [INCI nomenclature: CAS 13557-75-0]	Alkoxylated fatty acids, alkoxylated fatty alcohols, fatty acid salt, lecithin, monoglycerides, poloxamers, polysorbates, sorbitan esters
Folkloric additive	Plant and animal extracts, protein derivatives	
Humectant	N-Acetyl monoethanol-amide, pyrrolidone carboxylic-acid	Glycerin, sorbitol, urea, sodium lactate
Lubricant (hair and skin)	Mink oil, sodium hyaluronate	Coconut oil, lanolin, olive oil, petrolatum
Plasticizer	Ethyltoluenesulfonamide	
Preservative	1,3-Dimethylol-5, 5-dimethyl hydantoin, Quaternium-15 [INCI nomenclature: CAS 4080-31-3]	Camphor, dioctyl sebacate Imidazolidinyl urea, parabens phenoxyethanol
Solvent and viscosity reducer		Acetone, ethanol, ethyl acetate, glycerin, hexylene glycol, <i>i</i> -Propanol, propylene glycol, toluene
Suspending agent	Sodium methylnaphthalene sulfonate [INCI nomenclature: CAS 26264-58-4]	Carbomer, cellulose derivatives, clays
Thickener		Behenyl alcohol, polyethylene, polyisobutene, polyoxyethylene 14,000, tristearin
UV light absorber	Butyl methoxybenzoylmethane	Drometrizole

The overlap on the basis of ingredient usage is apparent from an examination of column 3.

Labeling of Cosmetics

An important ingredient-related topic is cosmetic ingredient labeling. Advocates of consumers' rights have suggested that the public would benefit from full disclosure of the composition of cosmetic products. Pursuant to the regulations by the FDA and the FPLA, all cosmetics are now required to carry the following information on labels: 1) a statement of the identity of the product; 2) a statement of the net quantity of contents; 3) a statement of the name and place of business of the manufacturer, packer, or distributor; 4) a list of the ingredients included in the product in order of predominance; and 5) cautionary or warning language [21 USC 321 (k); 15 USC 1459 (b); 21 CFR 701.10].

Except the aforementioned labeling for item 4, similar requirements exist for Rx and OTC products at the time of this writing. Labeling of all constituents in OTC drug products is still under consideration. The listing of ingredients in cosmetics must be in descending order of predominance, with some exceptions for components present as minor constituent and color additives. In order to achieve uniformity for identification, the CTFA has continuously created shorthand nomenclature for all cosmetic ingredients. The "naming" process is similar to that employed by USAN (U.S. Adopted Name). Many names and chemical descriptions of cosmetic ingredients have been reviewed and accepted by the FDA (for the purpose of using these names on labels). As a result, cosmetics in the U.S. are now labeled in accordance with the nomenclature and the rules of the International Nomenclature of Cosmetic Ingredient Dictionary (INCI Dictionary; 1). This approach has been accepted in the E.U., where the same names (with minor modifications) are used. "Harmonization" of names is a continuing process to make these names linguistically acceptable throughout the world.

COMPARISON ON THE BASIS OF SAFETY AND PERFORMANCE

Safety

Users know that, as a rule, Rx drugs are more likely to cause adverse side effects than OTC drugs. This is an obvious result of the nature and of the distribution system for these products. Prescription drugs are administered under the supervision of a physician, who has the

responsibility and moral obligation to monitor the patient's progress. However, OTC drugs may be used ad lib by the uninformed, who may not always be competent to diagnose the underlying disease or to recognize adverse side effects. In this respect, cosmetics resemble OTC drugs except that cosmetics are used repeatedly and over extended periods of time. Thus, the requirements for the safety of cosmetics should be, in fact, much more stringent than those for many drugs. The level of side effects or adverse effects that can be tolerated by manufacturers of cosmetics is virtually nil. Of particular concern is the sensitizing potential of components during prolonged and repeated use. Photosensitization is another phenomenon that has led to the removal of some cosmetic ingredients from the list of routinely employed substances.

Elegance

Elegance is not a primary concern in the case of Rx drugs but impacts marketing of OTC drugs and helps to ensure patient compliance. By contrast, any feature that detracts from the elegance (appearance, odor, texture, etc.) of a cosmetic interferes with its marketability and acceptance. As a matter of fact, cosmetic elegance is the essential attribute of a successful cosmetic product.

Performance

Performance is the one issue in which cosmetics and drugs are different. Drug efficacy is assessed on the basis of cure or prevention of disease. The FDA has established that cosmetic products that exert therapeutic effects are drugs. As a result, some traditional cosmetics were converted into drugs via the OTC panel process. Any performance claims for these OTC products are limited to the wording approved during the OTC review process.

Claims for (nondrug) cosmetics may be, for example, fashion-oriented (color), beauty-oriented (hiding of blemishes), or texture-oriented (emollient or lubricant). These, and related claims, are easily perceived and can be readily documented. More complex issues arise when cosmetic claims are made for age-related or reparative skin-care preparations. In the past, various regulatory agencies have been permissive with regard to cosmetic puffery claims. More recently, claims made for some cosmetics suggest to consumers that the product may exhibit a druglike effect, as defined by statute. A Commissioner of Food and Drugs has labeled these claims "daring" (2). Advertising copy that implies that a product nourishes the skin, is active, causes tingling, tightens the skin, discourages wrinkle formation, is prepared by a pharmaceutical company, or performs like

a face-lift may be false and misleading if the product does not perform. On the other hand, the product is considered a drug if it performs as claimed. Aside from drug results, claims for skin or hair benefits require documentation for commercial purposes (advertising and promotion). Thus, claims for performance are likely to run afoul of FTC rules. In the E.U., however, claims for efficacy require substantiation by regulation. Guidelines for documenting cosmetic performance exist in Europe but not in the United States (3). On the other hand, the FDA has a powerful tool for stopping unsubstantiated claims. A claim for wrinkle "removal," even on a temporary basis, may make a cosmetic product into a drug. Thus, the cosmetic industry is as tightly controlled as the Rx industry. Whenever a new indication for an existing drug constituent is claimed or whenever a druglike claim is made for *any* ingredient, a new drug application is required.

Shelf Life

Cosmetic preparations need not be labeled for outdating. This does not imply that cosmetic products must (or do) exhibit indefinite stability. The physical and chemical stability of cosmetics is routinely studied by the same procedures as those used for Rx or OTC drugs. Since cosmetic products do not contain active drug entities, the chemical stability of *any* component may be critical to the performance of the product: The components of a fragrance product require monitoring; the performance of a hair-waving preparation may depend on alkalinity and the chemical integrity of the reducing agent.

Physical stability affects cosmetic elegance, for example, by the breaking of an emulsion. Moreover, physical stability may also affect efficacy, as is the case during settling of a pigment in a nail lacquer that might then no longer be readily redispersible. A similar type of instability may occur as a result of the settling of an antiperspirant compound in a suspension aerosol. In these cases, neither the cosmetic (nail lacquer) or the OTC drug (antiperspirant) performs as claimed and may be considered misbranded or mislabeled.

As a rule, therefore, the demands of chemical and physical stability are similar for drugs and cosmetics. Under certain circumstances, the demands on physical stability may be especially critical as, for example, in hand and body lotions that have to perform under tropical conditions after having been exposed to the heat of the sun on the beach or after storage under arctic conditions in a ski hut.

The criteria for microbiologic cleanliness of cosmetic products are especially complex. Like a drug, a cosmetic is deemed adulterated if: 1) it bears or contains any poisonous or deleterious substance which may render it

injurious to users on the conditions of use as are customary or usual; 2) it contains in whole or in part of any filthy, putrid, or decomposed substance; and 3) if it has been prepared, packed, or held under unsanitary conditions where it may have become contaminated with filth, or whereby it may have been rendered injurious to health [Sect. 601, FDCA].

Cosmetic companies adhere to FDA-mandated GMPs or to GMPs promulgated by the CTFA with regard to housekeeping, cleaning, and sanitizing of equipment, and purity of raw materials and process water. Water is a particularly important component of finished cosmetics. Its purity is closely monitored to avoid the inadvertent introduction of contaminating biota into products. The GMPs established by the FDA are available in 21 CFR, Part 211, and the CTFA has published similar recommendations in the form of Quality Assurance Guidelines for Cosmetic Manufacture.

The final check on purity is on the finished product. The high water content and the inclusion of nutrients for unwanted microbiota make cosmetics subject to microbiological contamination. Thus, topical OTC products and cosmetics require preservation and a final check before distribution. Preservative systems and GMPs usually ensure delivery of essentially uncontaminated cosmetics. The microbiologic requirements recommended by the CTFA include the following: 1) baby products: less than 500 microorganisms/g; 2) eye products: less than 500 microorganisms/g; 3) oral products: less than 1000 microorganisms/g; (4) all other products: less than 1000 microorganisms/g; and 5) pathogens should be absent.

The use of preservatives to achieve the desired low levels of contaminating microorganisms is required. Some liquid cosmetic products do not support the growth of microorganisms (e.g., alcohol-based aftershave), while others are excellent growth substrates (e.g., a protein-containing hair conditioner). Because consumers frequently introduce microorganisms during normal product use, some cosmetic manufacturers may require that their products be self-sterilizing. This is a difficult task and not always achievable. A final check for levels of microorganisms is, nevertheless, desirable.

Relation to Health-Care Providers

Like the drug industry, the cosmetic industry requires animal toxicology and human testing to establish the safety of its products. As a rule, most cosmetic products are quite innocuous upon ingestion, even though they may cause laxation or act as emetics. The industry makes a deliberative effort not to market products that might elicit

toxic syndromes when ingested or applied topically. The former type of problem is usually handled by poison control centers and routinely (in 90% of all cases) involves ingestion by children. The number of fatalities was reported as nil between 1971 and 1978.

Dermatologists see many patients who have used cosmetic products properly but still report adverse reactions. The cosmetic industry is conscious of the need to provide products that do not elicit irritation or allergic responses during use. For this reason, the cosmetic industry depends on all types of patch and related testing by dermatologic laboratories to establish the safety of a given product in a predictive fashion. A whole battery of test protocols is available, and hundreds of subjects are tested routinely by dermatologists before product marketing (4).

Topical drugs and cosmetics have the potential of penetrating the skin. In the case of a topical drug (e.g., an antiinflammatory steroid), localized permeation may be a desirable feature. Penetration of cosmetics into and through the skin may elicit undesirable effects, especially since consumers may apply two or more products to the same site. Thus, in contrast to topically used Rx drugs, cosmetics should be retained on the skin with minimal penetration. Dermatologists also recommend cosmetics to patients as, for example, in cases of dry or chapped skin. Clearly, the cosmetic industry–physician–user relationship is not very different from that existing in the drug industry.

COSMETIC OR DRUG?

Formulators and marketers require an answer to the question of whether a given product is a cosmetic or a drug. Some of the answers are almost obvious. In the United States, a product is a drug if a drug claim is made for the preparation. The inclusion of a new chemical entity on which a therapeutic claim is based requires filing of a NDA; a product containing such a substance is automatically viewed as a drug. A typical example is a suntan product that adds an UV light-absorbing substance that was not reviewed by the OTC panel but was previously used as a sunscreen. Although the product in question was labeled as a drug, the FDA ruled that this clearly constituted use of an unapproved *new* drug substance. The addition of an OTC-reviewed sunscreen into a cosmetic makeup preparation—without claims for sun-protective action but with the indication that the product contains a sunscreen for the purpose of reducing UV light-induced skin aging—could be considered a drug use in the United States.

A soap to which an OTC Category I antimicrobial agent has been added is converted into a drug. This raises an interesting secondary issue. Soaps may be tinted with noncertified color additives. This inclusion of the antimicrobial requires not only drug labeling but also reformulation with approved colorants.

As a rule, the status of the product is determined by the claims made for it and its intended purpose. The use of aluminum chloride, a Category I OTC antiperspirant, as an astringent probably does not confer drug status on the product. It was already noted that certain OTC Category I skin protectants, such as petrolatum, can be used freely in cosmetics as long as no drug claims are made for the product. Numerous other issues might arise, and each one might require specific adjudication. An example is the use of an approved sunscreen in a cosmetic product to preserve it against UV light deterioration. Since the intent of the sunscreen's use is clearly not drug related, the product will probably be considered a cosmetic.

SUMMARY

Cosmetics and drugs are distinctly different on the basis of U.S. law. In principle, cosmetics may not contain ingredients that treat or prevent disease or alter the structure or function of the human body. The objective of cosmetics is limited to the enhancement of appearance.

The ingredients used in cosmetics to a large extent are the same as those employed in drugs, with the exception of components that are intended to cure, alleviate, or prevent disease.

The demands on product stability and manufacturing practices are essentially the same for cosmetics and drugs.

Finally, important differences are seen in judging the performance of cosmetics and of drugs. Consumers assess cosmetics based on the products' performance vis-à-vis the demand for better appearance. On the other hand, drugs are assessed on their ability to prevent or improve a disease state.

REFERENCES

1. Wenninger, J.A.; Canterbury, R.C.; McEwen, G.N., Jr. *International Cosmetic Ingredient Dictionary and Handbook*, 8th Ed.; The Cosmetics, Toiletries, and Fragrance Association: Washington, DC, 2000.
2. Hayes A.J. Jr. Annual Meeting of the Cosmetics, Toiletries and Fragrance Association: Boca Raton, FL, March 2, 1983.

3. COLIPA, Guidelines for the Evaluation of the Efficacy of Cosmetic Products; European Cosmetic, Toiletry, and Perfumery Association. Rue du Congr s: 5–7 B-1000: Brussels, Belgium, August 1996.
4. Marzulli, F.; Maibach, H. *Dermatotoxicology*, 5th Ed.; Hemisphere Publishing Corp.: New York, 1997.
5. Bachrach, E.E. The FDA's Flexible Policy for Labeling Over-the-Counter Cosmetic Drugs: A Review and Analysis. *Food Drug Cosm. Law J.* **1987**, *42*, 184–191.
6. Eierman, H.J. Regulatory Requirements for Marketing Cosmetics in the United States. *Drug Inf. J.* **1987**, *21*, 387–391.
7. Gilbertson, W.E. FDA OTC Drug Standards Versus Cosmetic Standards. *Drug Inf. J.* **1987**, *21*, 379–385.
8. Hyman, P.M. Regulatory Risks of Marketing Drugs and Cosmetics. *Drug Inf. J.* **1987**, *21*, 393–401.
9. Rogiers, V. Efficacy Claims of Cosmetics in Europe Must be Scientifically Substantiated from 1997 On. *Skin Res. Technol.* **1995**, *1*, 44–46.

DNA PROBES FOR THE IDENTIFICATION OF MICROBES

Wayne P. Olson

Consultant, Beecher, Illinois, U.S.A.

INTRODUCTION AND SCOPE

In 1997 and the first quarter of 1998, when much of the first version of this article was written (1), what we hereafter refer to as DNA probes were used primarily by those performing genetic-related research. Today (mid-2000), oligonucleotide (DNA, RNA, and related) probes in several guises are a growth industry in biotechnology, and their applications are widespread. In 1995, Theta Reports (*Gene Therapy/DNA Probes/PCR Markets*) predicted a phenomenal growth in polymerase chain reaction (PCR), DNA probes, and the successful use of gene therapy—they were correct about the first two. In 1996, the Frost & Sullivan Market Intelligence indicated that the 1995 market for DNA probes had been over \$145 million, and predicted that it would reach \$1.4 billion by 2003 (*U.S. DNA Probe Markets*). Also in 1996, Business Communications Co., Inc., made similar predictions (*Advances and Opportunities in DNA Testing and Gene Probes*). DNA probes are the fastest growing area of in vitro diagnostics (2), expanding at the rate of 25% per year.

The most important differences between the earlier and the current version of DNA probes (1) relate to the following:

- Literature and applications have boomed; cumulative citations listed by Medline for “DNA probes” are currently 79,270.
- Applications—especially of PCR—now embrace most aspects of life sciences, including agriculture, food science, and nonmicrobial disease states, as well as areas relating to human diseases.
- Use of oligo-on-a-chip (DNA- or RNA-based array, also referred to in some instances as microchip) has burgeoned in pharmaceutical/biotech R&D. This approach, if not yet predominant, is becoming so in many applications.
- Antisense therapeutics, which had lost favor due to a well-publicized Phase-I human death in mid-2000, are already rebounding. Some forms of antisense show very high potential as antimicrobials.

To focus the current version of “DNA probes” on microbiological and related issues, together with

literature not explored in the previous version (1) some background material has been abbreviated. The reader unfamiliar with the fundamentals of DNA probes is urged to consult Vol. 19, Ref. 1 for introductory particulars. Table 1 of that article is repeated herein as Table 1, but without references. This author also recommends that the genomics tyro read one or more of the basic references (3–5), of which the article by Keller and Marak is the most germane. Additional basic texts not cited in (1) are listed in Table 1. Editors of those texts, who are significant contributors to the scientific literature, include (to my knowledge) Clapp, Higgins, Kricka, Marak, and Persing. Presumably, the others also do fundamental research but I have not read their papers.

Rapid growth areas of science, such as oligo probes, are referenced most rapidly on the Internet. Formal print publication may appear 1–3 years after completion of the work. Many, but not all, Internet publications are refereed in a manner similar to that for print publications. Therefore, many of the references herein are from the Internet.

Although I have cited a very small proportion of the patent, regulatory, and market literature herein, the reader should note that

1. Many productive scientists choose to file for letters patent in lieu of, rather than in addition to, conventional publication.
2. Negative findings may not find their way into scientific literature for a long time, if ever.
3. For a new commercial venture, first-rate technology is often insufficient; good business practice and adequate financing are equally important. Therefore, many small operations disappear.

Follow-up on vanished companies and products is difficult, unless they are acquired. For example, Biotech Research Laboratories and their peptide nucleic acid oligos (PNAs) were acquired by Boston Biomedica, Inc. Tracking BRL and its specifically targeted PNAs to BBI is simplified by web-page links, which are easy to follow for the reader with a modest knowledge of computer use.

Table 1 Texts entirely, or in part, about oligonucleotide probes

Year	Editor(s)	Title	Publisher
1996	Persing, D.H. Clapp, J.P.	<i>PCR Protocols for Emerging Infectious Diseases</i> <i>Species Diagnostic Protocols: PCR and Other</i> <i>Nucleic Acid Methods, Vol. 60</i>	ASM Press Humana
1995	Hames, B.D.; Higgins, S.J.	<i>Gene Probes: A Practical Approach, Vols. 1 & 2</i>	Oxford University Press
1994	Svendsen, P.; Hau, J.	<i>Handbook of Lab Animal Science, Vols. 1 & 2</i>	CRC Press
1993	Keller, G.H.; Manack, M.M.	<i>DNA Probes</i>	Stockton Press
1993	Persing, D.H.; White, T.J.; Tenover, F.C.; Smith, T.F.	<i>Diagnostic Molecular Microbiology: Principles</i> <i>and Applications</i>	ASM Press
1991	Stanley, P.E.; Kricka, L.E.	<i>Bioluminescence and Chemiluminescence:</i> <i>Current Status</i>	Wiley
1990	Boulton, A.A.; Baker, G.B.; Campagnoni, A.T.	<i>Molecular Neurobiological Techniques</i>	Humana

DEFINITIONS

Amplification of an oligonucleotide or an entire gene: See polymerase chain reaction.

cDNA: Copy DNA, as by PCR or RT-PCR.

DNA probe: An oligonucleotide sequence that complements a sequence (usually of a gene) in or from an organism. Here, DNA is used generically in the description of probes.

dsDNA: Double-stranded DNA, as in the double helix.

endonuclease: An enzyme that hydrolyzes (breaks) an oligonucleotide chain between particular bases only; consequently, the pattern of pieces of DNA (oligos) generated by the action of endonuclease on DNA of a particular organism is reproducible.

Genome: The entire DNA complement [gene(s)] of an organism; this includes the genetic information in viruses.

Morpholino oligo: An oligo containing the morpholino analog of the conventional backbone (Fig. 1a).

Nucleoside: Molecule containing a purine or a pyrimidine base linked to a pentose sugar.

Nucleotide: A phosphorylated nucleoside.

Oligonucleotide (also termed oligos): Usually a sequence of DNA or RNA with a phosphate backbone (Fig. 1b), but may have a sulfate, peptide, or morpholino backbone in place of a phosphate one, to reduce or eliminate oligo degradation by nucleases.

PCR: See polymerase chain reaction.

Phosphorothioate oligo: An oligo (usually a probe) containing S in place of P (Fig. 1c).

PNA oligomer: Peptide nucleic acid oligo (Fig. 1d); manufacturers include Boston Probes (www.bostonprobes.com); the Danish company; Pantheco (www.pantheco.com), and Research Genetics (www.resgen.com). These, like morpholino analogs, resist nuclease digestion.

Polymerase chain reaction: The method for producing, in vitro and fairly rapidly, millions of copies of a specific segment of DNA or RNA (amplification). The sequence of the ends of the portion to be copied must be known so that the primer can be annealed to the denatured oligo to be copied. An excellent source for PCR applications is www.eppendorfsi.com/application.html.

Primer: The segment of synthetic oligo that is added to one end of the DNA or RNA to be copied. The denatured oligo target is annealed with the primer, and DNA polymerase then adds complementary nucleic acids until the full copy is produced. See Ref. 1 for figures and further explanation.

Reverse transcriptase (RT): The enzyme from RNA viruses that codes (cDNA) from RNA.

Table 2 Commercial firms producing and/or selling oligonucleotide probes and arrays (partial listing, in alphabetical order)

Probes	Affymetrix
Bayer	Aventis
Chiron	BioMerieux
Digene	Nanogen
Gen-Probe	Oncogene Science Diagnostics
Roche	Roche Molecular Systems
Oligos-on-a-chip	

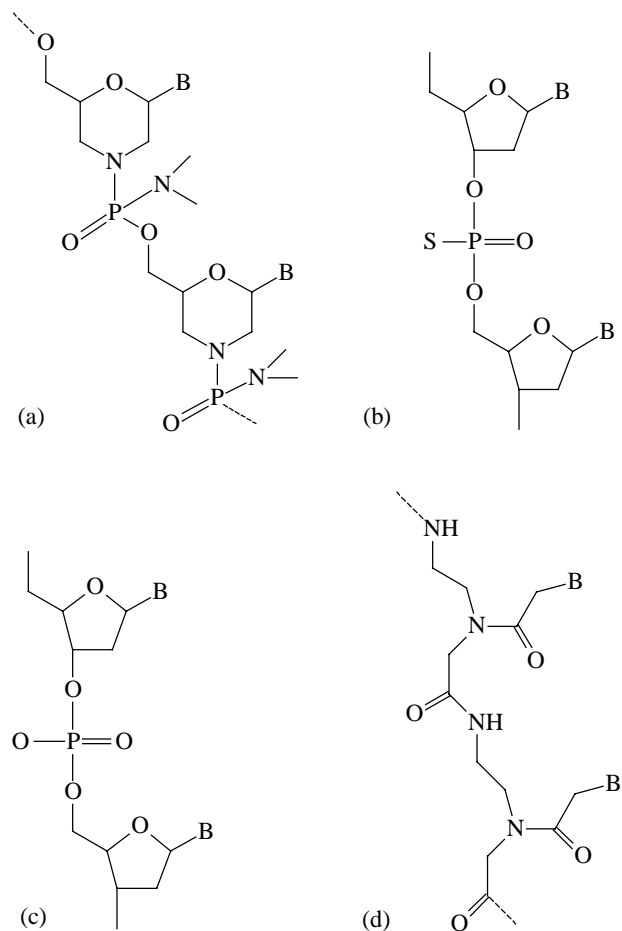


Fig. 1 Dinucleotide structures (as found in respective dinucleotides). (a) morpholino analog; (b) normal DNA or RNA; (c) phosphorothioate analog of DNA; (d) peptide nucleic acid.

Sense strand: In dsDNA, the strand that codes for the protein that is generated from it. The other strand is complementary, and protects the sense strand.

ssDNA: Single-stranded DNA, as contrasted with dsDNA

Target of an oligonucleotide probe: Usually ribosomal RNA (rRNA), which (with uridine in place of thymine) is the complement to the sense strand of the dsDNA in the double helix.

Triplet code: Although the coding for the amino acid sequence of proteins is by sets of three nucleotides in the DNA, the actual blueprint for protein synthesis is in RNA transcribed from the DNA. For example, in the RNA copied from a gene, UAU (uracil–adenine–uracil) or UAC (C is cytidine in the second uracil) are triplet codes for tyrosine.

ABBREVIATED BACKGROUND

Every characteristic of an organism (single-celled or multicelled) is dictated by its genome. For example, all metabolic processes are determined by enzymes that are encoded in the genome (some enzymes are imported and/or exported from a microbial cell in plasmids, which are small closed DNA loops of genes separate from the microbial chromosome). In some viruses, the genome is ssDNA or RNA, but in the majority of organisms, including microbes, it is dsDNA. The double helix of dsDNA contains a sense strand, which carries the sequence of the nucleotide triplet codes that characterize the amino acid sequence of the protein they code for. The other nucleotide strand in dsDNA is complementary to the sense strand and stabilizes it (Fig. 2).

Although some segments of the DNA in a microbial cell provide information that can identify the cell to species, there is only a single copy of DNA per cell. There may be 10^4 or more copies of the 16S rRNA per cell involved in the production of various proteins (usually enzymes) for use by the cell. The rRNA of molecular weight 16S (designated as Svedberg units in the ultracentrifuge) provides regions that are suitable for identification of the microbe to species (Fig. 3).

Therefore, if the 16S rRNA of an organism can be made available to oligo probes, usually by opening the organism in some way, it is possible to see whether the rRNA of the microbe anneals (forms a duplex) with the labeled synthetic oligo probe of known sequence and identity. There are several ways of doing this. However, in my view, unless the genus of the organism is known with reasonably good probability, this method is inordinately expensive in the use of labeled probes.

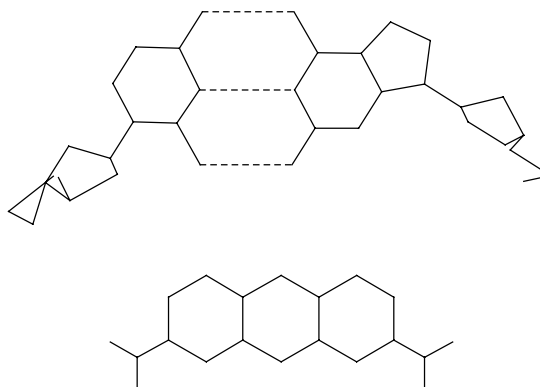


Fig. 2 Schematic of base pair in dsDNA (above) and acridine orange (AO) (below). The AO intercalates between sequential base pairs, extending the phosphate backbone.

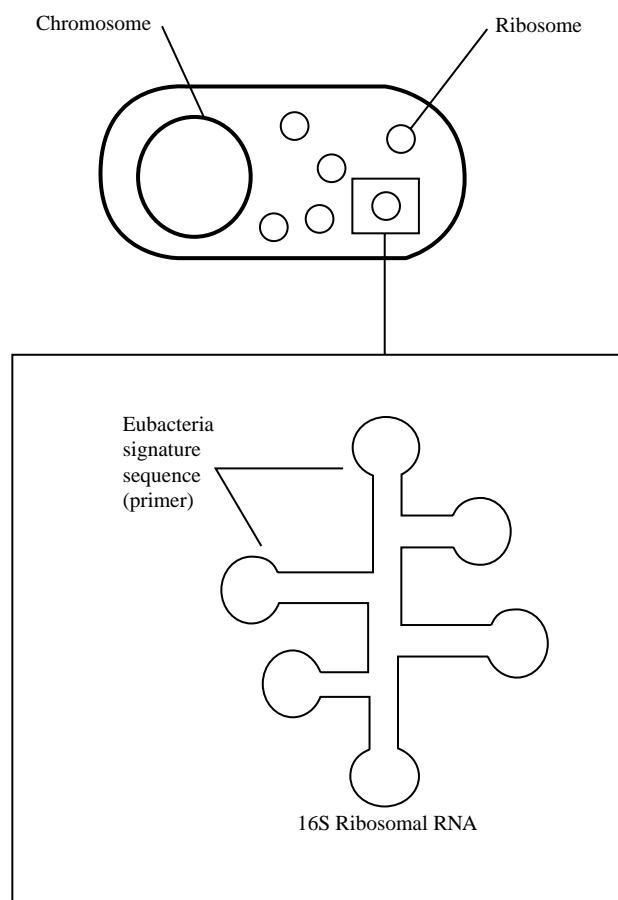


Fig. 3 Extrachromosomal ribosomes in a bacterium or fungus that contain 16S rRNA.

APPROPRIATE USE OF OLIGO PROBES

In my view, before labeled probes are put to use for the identification of an unknown, the organism should be tentatively identified. Conventional (but effective) hand methods derived from *Bergey's Manual* can be painfully slow. The fatty acid methyl ester (FAME), technique reference (6), or one of the methods outlined in (7) are usually quite accurate and sufficiently rapid if the patient is not in extremis. Also, it is wise to remember that conventional bench methods (e.g., *Bergey's*) are dendritic, and for microbiologists lacking considerable experience, an error in any of the sequential steps may result in error in the identification. The earlier a mistake is made in a dendritic pattern, the greater the error in identification (i.e., if a bird's nest is near the end of a limb on the north side of the tree and one chooses to climb on a limb on the south side of the tree in search of the nest, one will be right about the nest being above ground, but remote from the objective).

If preliminary information is available (e.g., indicators of pathogenicity), the most cost-effective and reliable approach may be to go directly to probes. Also, probes make best sense in widespread screening for specific diseases such as tuberculosis. In disease states, whether in humans, animals, or plants, symptoms may indicate the presence of one or a few pathogens. In such instances, the use of probes directly is sensible and economical. Machine identifications, based not on genomic but on chemical analyses, become confirming and often unnecessary.

Probes, depending on the oligo target in the organism, are for practical and definitive purposes. Errors with nongenomic machines (e.g., FAME profile by GC, or mass spectrometry/fuzzy logic profile, etc.) are because the organism sought is not in the machine library, or the concentration of key chemicals (almost always oligonucleotides, oligosaccharides, and/or FAME from the organism) is too low (insufficient cells or incomplete extraction).

For some very slow-growing organisms such as *Mycobacterium tuberculosis* and *M. leprae*, sufficient organisms may be present for analysis by probe in the sample fluid or tissue, in which case the use of penetrant aids such as DMSO facilitates the exposure of chemiluminescent-labeled oligo probes to the genome of the organism. The DNA of the organism then fluoresces under ultraviolet epi-illumination. A marvelous tutorial on probe labeling is available at www.jic.bbsrc.ac.uk/staff/pat-heslop-harrison/methods/probe.html. The method with fluorescent-labeled probes when applied for the detection of cancers or genetic effects, to human or animal genes in tissue, is called fluorescent in situ hybridization (FISH).

DEFINED OLIGOS ON MAGNETIC BEADS

Blotting and related methods have been discussed in a previous review (1). A very rapid technique employs specific oligos immobilized on Dynal[®] magnetic beads. The special advantage is that the beads, covalently coupled to particular oligo sequences (or antigens, antibodies, or other binding entities), can be mixed in a nonmetallic container (e.g., test tube) with a solution or suspension so as to react even with very dilute targets (organisms, especially viruses). The beads are then concentrated simply by placing a magnet against the side of the container. The original solution, less the bead-reacted targets, is then removed and the beads rinsed in situ. An oligo-magnetic bead method is outlined in (7). The recovery of mRNA (terminal polyA sequence) is with poly dT beads (8).

The basic magnetic-bead patent appears to be the property of Dynal, but Clemente Associates (formerly Quantum Magnetics, Madison, CT) produces Ni-bearing magnetic particles of 3- to 5- μ m diameter for performing what they refer to as "magnetic chromatography," of cells, proteins, or both. Qiagen (Valencia, CA, and other locations) markets histidine-labeled Ni-containing agarose magnetic beads. Bangs Laboratories (Fishers, IN) also produces 2- μ m diameter beads with surface carboxyl groups for the coupling of binding ligands. Seradyn (Indianapolis, IN) uses beads for DNA isolation. All of these can bear oligo probes. CPG, Inc. (Lincoln Park, NJ) offers streptavidin magnetic porous glass for the binding of biotinylated molecules for a variety of oligo-probe-related purifications of cells, noncellular particles, and macromolecules, including DNAs and RNAs.

Nonmagnetic beads of a variety of matrices, ready for coupling by various chemistries to oligos, antibodies, antigens, or what-have-you, are available for column or suspension applications from Polymer Laboratories (www.polymerlabs.com).

THE POLYMERASE CHAIN REACTION (PCR) AND THE DETECTION OF MICROBES

If a dozen copies of a pathogenic DNA virus are present per milliliter or gram in body fluid or beef tissue intended for sale, inoculation of the suspect source material into cell culture or in a test animal for analytical testing is a very lengthy, expensive process. It is unrealistic for use in the food industry or in a hospital, as are many other tests. If we seek particular viruses (e.g., rabies), the most rapid, practical, and effective means of detecting the virions is the PCR.

The general sequence of events in performing PCR is shown in Fig. 4. The primer used is an oligo specific to the organism to be detected. Commencing with the DNA recovered from one or several organisms, the multiplication of genomic material can be 10^7 -to 10^8 - fold in 35–90 min. By comparison, growing the virions in cell culture to equivalent numbers (to allow monoclonal antibody type analyses) usually requires 48 h or more. In cases where life hangs in the balance of an accurate analysis, those hours may be critical.

Even with bacterial or fungal pathogens, the time required to generate sufficient genomic material for analysis by PCR is 35–45 min but 24–48 h in conventional culture on an appropriate nutrient medium. Some useful background information on PCR is provided in the Eppendorf web page (www.eppendorphi.com/

applications.html). The reader is also referred to the previous version (1), which provides details in a somewhat different manner. Primers (DNA probes needed for PCR) have been described particularly well by Steve Rozen (Whitehall Institute, MIT Center for Genomic Research, Cambridge, MA). The material he has developed is available, gratis, at www.genome.wi.mit, which appears to be difficult to recover. HIV primers are shown at www.appliedbiosystems.com/pc/catalog/pg25.html.

Reverse transcriptase PCR, (RT-PCR), runs the normal RNA to DNA sequence backward. Reverse transcriptase is an enzyme found in RNA viruses and, when added to a solution of RNA, deoxynucleotides, reverse transcriptase, etc., allows one to produce the DNA equivalent to the RNA recovered from an organism. The cDNA (DNA produced from RNA) recovered in sufficient amounts can then can be "clipped" using a well-characterized endonuclease and a cDNA "fingerprint" generated for comparison with the fingerprints of known RNA sources. Automated systems for doing this can be purchased from Qualicon, a DuPont spinoff (info@qualicon.com) or from Applied Biosystems (www.appliedbiosystems.com/ab/about/mm/microseq/microseq500/).

IMMOBILIZED OLIGOS OR GENES IN ANALYTICAL ARRAYS

As reported in the previous version of this article (1), immobilized oligos in arrays (e.g., in the wells of microtiter plates or on glass slides) are used in the detection of cancer and precursor stages (e.g., mutations), in addition to drug screening of the interactions of candidate drugs (free) with human or animal gene sequences (immobilized). A tutorial on microarray technology is presented by Virginia Tech and the Forest Biotechnology Group at the North Carolina State University (9).

To provide the reader with an appreciation of the flexibility possible with arrays, I consider the possibilities inherent in combinatorial chemistry (10). One possible approach involves the extraction of the gene of each of hundreds of bacterial species, grown as axenic (pure) cultures. In each of the wells of microtiter plates molded from polyethylene-co-acrylic acid, segments of the gene of a single bacterial species (prepared when the DNA from the cells of a particular colony are "clipped" with a known endonuclease) are added and immobilized to the bottom of the plate via the 3'-OH. A good immobilization system is the triazine method for coupling linear microbial DNA strands (Fig. 5). To each of these plates, bearing (for example) ca. 100 wells and the genomes of 100 human

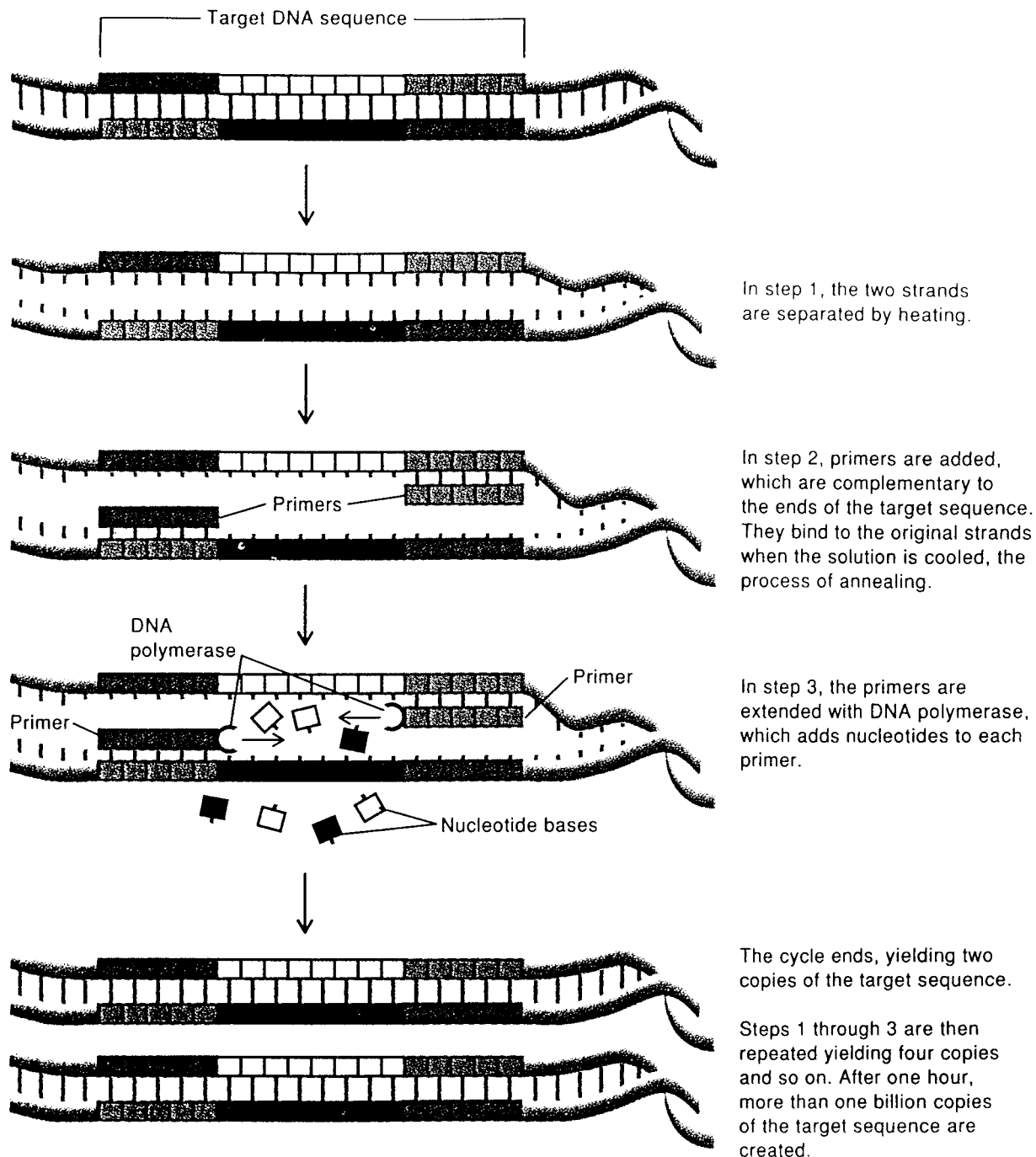


Fig. 4 Sequence of events in the PCR.

microbial pathogens, is added one of the many varieties of possible synthetic oligos intended for test as potential drugs (limited mixtures, each unpurified but reproducible), each with a low toxicity label, preferably chemiluminescent rather than radioactive.

The drugs (e.g., PNAs) and immobilized microbial genomes are allowed a reaction time, rinsed, and then examined (automatically or by eye,) for retention of the synthetic candidate drugs by segments of the genomes of the various pathogens. In this way, it is possible to screen

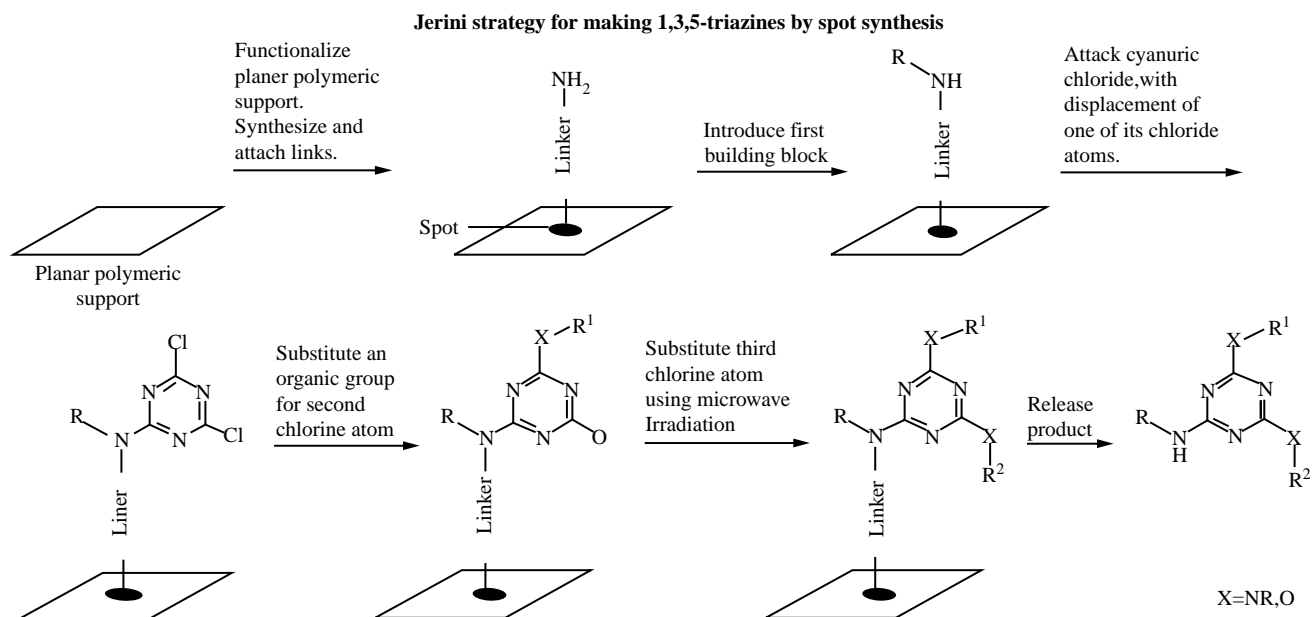


Fig. 5 Coupling microbial ssDNA or RNA or analogs via triazine to wells in a microtiter plate.

candidate oligos for specific binding to immobilized genomes of known organisms, even if gene sequences of the pathogens are unknown, and the precise sequence that binds to a given pathogen's genome is known only with modest accuracy.

In cases where this combinatorial-type approach to new product development appears promising, it can be automated, all or in part. Such systems are in widespread use in a variety of applications in the pharmaceutical industry. The example given is a variation of the method known as combinatorial chemistry.

A variation on this method has been proposed by the U.S. National Institute of Environmental Health Sciences to evaluate chemicals as potential human toxins. Human genes, immobilized on microtiter-type plates, are exposed to known and unknown toxicants, with a view to seeking similarities in gene-binding patterns (11). Origene (Rockville, MD), Qiagen (Valencia, CA, and other sites), CCS Packard (Meriden, CT), and many others may be added to the list of array-type manufacturers/distributors listed in Table 3 in Ref. 1.

SPECIFIC OLIGO PROBES AS ANTIMICROBIALS

The DNA of a cell must be copied if the cell is to reproduce. Also, certain enzymes are required for a cell to metabolize and function (as contrasted with the dormant

state), and these are usually made in the cell as needed. Therefore, an antisense probe that reacts with, for example, a critical RNA in the cell, may effectively render the cell dormant. One can envision this in vitro and in vivo. But oligonucleotides are subject to degradation by nucleases which, depending on their specificity, clip a nucleotide at certain base sequences, usually rendering the oligo less functional or nonfunctional. (This is also how DNA is clipped in the process of generating "fingerprints.")

If a species-specific antisense oligo probe can be made resistant to enzymic degradation, and if it can readily enter a microbe, the probe not only identifies the organism, but also may shut down the metabolic potential of the cell (11). Antisense PNA oligos appear to meet these requirements and bind to sense with a lower dissociation constant (K_d) than does the phosphate-backbone antisense. PNA antisenses have the potential for use as very specific, low-toxicity microbicides for in vivo applications (12). The peptide analogs of DNA probes are protected by appropriate patents (13). The text of choice is from the same authors and others (14).

Boston Bioprobes is producing PNAs for oligo research, development of diagnostics, testing of foods, and the environment. Perkin-Elmer (now largely renamed Analytical Biosystems, Foster City, CA) has invested with BBI and will participate in PNA technology.

ANTISENSE PROBES AS THERAPEUTICS

The primary target of antisense probe studies is the ongoing cell proliferation in all cancers.

Cancerous cells are not functional in, for example, gas transport in the lung. Like viruses, their anthropomorphized *raison d'être* is only to reproduce, not to perform a service for the organism of which they are a part. On the other hand, normal cells have a finite lifetime and must occasionally be replaced.

The toxicants used in conventional chemotherapy and radiation not only kill a substantial proportion of the cells that are reproducing rapidly but also tend to destroy stem cells in the bone marrow and circulating plasma, which gives rise to the panoply of red cells, white cells, and platelets that are essential to a survival free of dialysis machines and other elaborate supports. For this reason, patient-sourced (autologous) stem cells (CD34+),

hopefully cancer-free, are recovered from the patient prior to radiation treatment or chemotherapy (or both) as “heroic” cancer therapy. The recipient of massive “chemotherapy,” or radiation (or both), dies unless those healthy stem cells (removed from the patient prior to treatment) are injected intravenously (they find their way to the marrow) and replace the essential cells that chemotherapy or radiation has destroyed.

In addition, although chemotherapy and radiation may “cure” a patient, the once-healthy cells that are not killed may be mutated in critical portions of the genome. It is fairly common that patients receiving “heroic” therapy develop new tumor sets 20 years downstream of the initial therapy. Ultimately, the exposure to these unselective toxicants that saved the patient may eventually kill the patient. (“The therapy was successful, but the patient died.”)

In theory, the beauty of antisense is that an antisense oligo probe does not damage all types of cells at random,

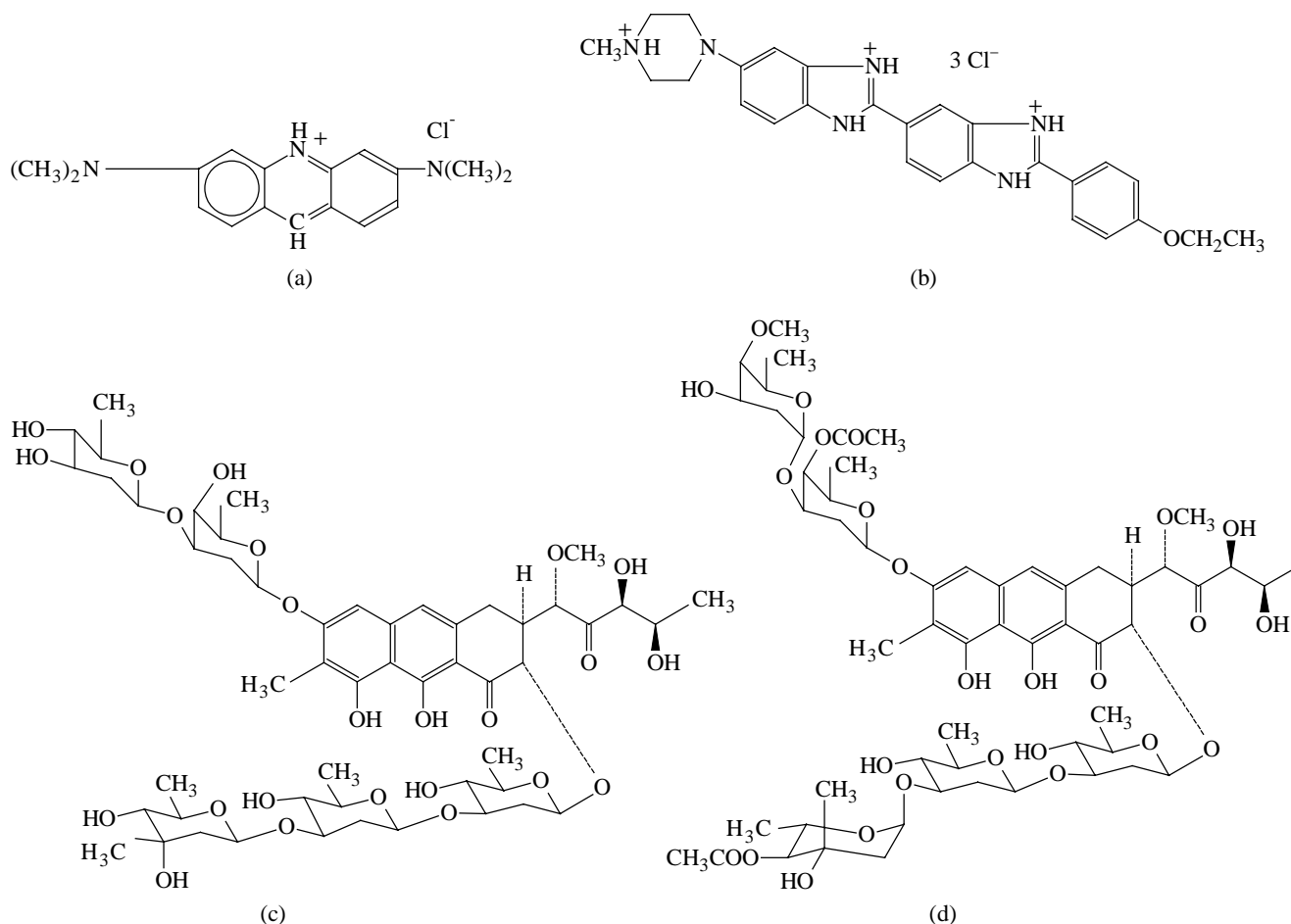


Fig. 6 Small organic molecules that interact with DNA and/or RNA and when excited with one wavelength of light, fluoresce in another. (a) acridine orange; (b) Hoechst 33342; (c) mithramycin; (d) chromamycin A3.

and does not induce mutations. Antisense shuts down the propagation of some cells, and not others. Antisense therapy is distinct from gene therapy, in which a missing or malfunctioning gene (e.g., human clotting factor VIII gene in most hemophiliacs) is placed in autologous cells and returned to the patient.

Professor Eric Wickstrom (Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA) has done some excellent work in mouse models on antisense and Burkitt's lymphoma, rabies virus, and the human *cMYC* oncogene, to name a few systems. His group employs phosphorothioates, PNAs, and other variations on antisense. His web page is at www.kcc.tju.edu/staff/wickstrom and several of his recent papers (15, 16) and his book (17) should be of interest to the reader interested in antisense. His patents on the synthesis (18) and introduction of antisense into cells (19) are germane.

Some of Wickstrom's work points to the issue of antisense entry into a cell. At least one company, Gene Therapy Systems, is devoted to this (www.genetherapy-systems.com).

SMALL, NONSPECIFIC ORGANIC PROBES THAT LOCATE INTACT DNA

Some organic molecules have a high affinity (low K_d) for DNA and are stoichiometric but not specific in the manner of a complementary oligonucleotide. For example, acridine orange (AO; Fig. 6a), is planar and intercalates between the paired sets of H-bonded base pairs (A-T, G-C) in dsDNA (Fig. 2). Where AO penetrates into a cell and "reacts" with the dsDNA, the cell emits green light when excited with blue. AO also reacts with ssDNA or RNA by stacking on the charged phosphate of the backbone; ssDNA fluoresces in the red.

Benzamides such as Hoechst 33342 (Fig. 6b) bind to A-T-rich regions in the small groove of dsDNA. When sufficient dye is excited with blue light, light emission is red. The antibiotics mithramycin and chromamycin A3 (Fig. 6c and Fig. 6d) bind to G-C regions of DNA.

These small molecule probes of DNA (ds and ss) and RNA do not identify organisms, but indicate probable viability because the genome is intact. When cells in suspension or on a slide are stained and examined under a UV-epi-illuminated microscope or in a flow cytometer, with or without cell sorting, then viability can be estimated with fairly high precision and accuracy. This and related topics are developed in (7), and in portions of the molecular probes pages (www.probes.com), which the reader may find particularly useful.

REFERENCES

1. Olson, W.P. *Encyclopedia of Pharmaceutical Technology*; 2000; 19, 103–120.
2. Nelson, N.C. *Molecular tools for building nucleic acid IVDs*; 1998; www.devicelink.com/ivdt/archive/98/03/013.html.
3. Keller, G.F.; Manak, M.M. *DNA Probes: Background, Applications and Procedures*, 2nd Ed.; Macmillan Reference, Ltd.: 1993.
4. Singer, M.; Berg, P. *Genes & Genomes*. University Science Books; Mill Valley: CA, 1991.
5. Davies, J., Reznikoff, W.S., Eds.; *Milestones in Biotechnology*; Butterworth-Heinemann: Boston, MA, 1992.
6. Gilbert, G.L.; James, G.S.; Sintchenko, V. *Med. J. Aust.* **1999**, 171 (10), 536–539.
7. Calomiris, J.J. *Monitoring Waterborne Pathogens Using DNA Probes*; 6th CBW Protection Symposium: Stockholm, May 1998; 10–15.
8. *Miltenyi Biotech*. 1999; www.miltenyibiotec.com.
9. Alscher, R. *Microarray Technology*, www.bsi.vt.edu/ralscher/gridit/intro_ma.htm.
10. Olson, W.P. *Genet. Engineer. News* **1997**, 17 (7), 1–36.
11. *NIEHS Environmental Health Institute to use Gene Chips to Evaluate Chemicals for Potential Harm to Humans*, 2000, www.niehs.nih.gov/oc/news/toxchip.htm.
12. PNA—The Ultimate Antibiotic, Pantheco A/S. Hørsholm, Denmark. www.pantheco.com/pna.html (1999).
13. Nielsen, P.E.; Buchandt, O.; Egholm, M.; Berg, R.H. *Peptide Nucleic Acids*. U.S. Patent 5,539,082, July 23, 1996.
14. Nielsen, P.E.; Egholm, M. *Peptide Nucleic Acids: Protocols and Applications*. Horizon Scientific Press: 1999.
15. Wickstrom, E.; Smith, J.B. *Cancer J.* **4**, S43–S47.
16. Smith, J.B.; Wickstrom, E. *J. Nat. Cancer Inst.* **1998**, 90, 1146–1154.
17. Wickstrom, E., Ed.; *Clinical Trials of Genetic Therapy with Antisense DNA and DNA Vectors*; Marcel Dekker, Inc.: 1998.
18. Wickstrom, E.; LeBec, C. Stereospecific Solid Phase Synthesis of Oligodeoxynucleoside Alkyl Phosphonates by Pentavalent Grignard Coupling, US Patent 5,703,223, 1997.
19. Wickstrom, E.; Cleaver, S.H. Composition and Method for Targeted Integration into Cells. US Patent 5,958,775, 1999.

BIBLIOGRAPHY

- Michotey, V.; Mejean, V.; Bonin, P. *Appl. Environ. Microbiol.* **2000**, 66 (4), 1564–1571.
- Prescott, A.M.; Fricker, C.R. *Lett. Appl. Microbiol.* **1999**, 29 (6), 396–400.
- Patel, R.; Newell, J.O.; Procop, G.W.; Persing, D.H. *Am. J. Clin. Pathol.* **1999**, 112 (1), 36–40.
- Wang, H.; Farber, J.M.; Malik, N.; Sanders, G. *Int. J. Food Microbiol.* **1999**, 52 (1/2), 39–45.
- Burstein, G.; Snyder, M.; Conley, D.; Howell, M.; Bockeloo, B.; Quinn, T.; Zenilman, J. *J. Pediatr. Adolesc. Gynecol.* **2000**, 13 (2), 91.

- Elnifro, E.M.; Cooper, R.J.; Klapper, P.E.; Yeo, A.C.; Tullo, A.B. Investigat. Ophthalmol. Visual Sci. **2000**, *41* (7), 1818–1822.
- Löffler, F.E.; Sun, Q.; Li, J.; Tiedje, J.M. Appl. Environ. Microbiol. **2000**, *66* (4), 1369–1374.
- Monstein, H.J.; Johansson, Y.; Jonasson, J. APMIS **2000**, *108* (1), 67–73.
- Hilali, F.; Ruimy, R.; Saulnier, P.; Barnabe, C.; Lebouguene, C.; Tibayrenc, M.; Andreumont, A. Infect. Immun. **2000**, *68* (7), 3983–3989.
- Anonymous. DNA Probes with PCR Assay for the Detection and Identification of *Haemophilus ducreyi*. www.hrinet.org/techtran/technologies/ducreyi.html.
- O'Connor, L.; Joy, J.; Kane, M.; Smith, T.; Mather, M. J. Food Prot. **2000**, *63* (3), 337–342.
- McFadden, J.; Collins, J.; Beaman, B.; Arthur, M.; Gitnick, G. J. Clin. Microbiol. **1992**, *30* (12), 3070–3073.
- Roring, S.; Hughes, M.S.; Skuce, R.A.; Neil, S.D. Vet. Microbiol. **2000**, *74* (3), 227–236.
- Clement, P.; Springael, D.; Gonzalez, B. Can. J. Microbiol. **2000**, *46* (5), 485–489.
- Rosenberger, C.M.; Scott, M.G.; Gold, M.R.; Hancock, R.E.; Finlay, B.B. J. Immunol. **2000**, *164* (11), 5894–5904.
- Danish, E.P.A., Ed.; *Vibrio vulnificus in Denmark*, www.mst.dk/200002pubs/87-7909-344-2/lelepubl_eng.htm.
- Hardin, B., Ed.; *DNA Probes Foil Spoilage Yeasts*, 1999, www.ars.usda.gov/is/AR/archive/aug99/dna0899.htm.
- Sakai, T.; Ikegami, K.; Yoshinaga, E.; Uesugi-Hayakawa, R.; Wakizaka, A. Tohoku J. Exp. Med. **2000**, *190* (2), 119–128.
- Tamura, M.; Watanabe, K.; Imai, T.; Mikami, Y.; Nishimura, K. Clin. Lab. **2000**, *46* (1/2), 33–40.
- Walsh, S.R.; Bidochka, M.J.; Roberts, D.W.; Humber, R.A.; Tyrrell, D.; Silver, J.C. *Characterization of DNA Probes Specific for Different Pathotypes of the Entomophaga grylli Species Complex of Grasshopper Pathogens. USDA Agricultural Research Service: 'TEKTRAN'*, www.nal.usda.gov/ttic/tektran/data/000006/45/000006454_0.html.
- Coustou, V.; Deleu, C.; Saupe, S.J.; Begueret, J. Genetics **1999**, *153* (4), 1629–1640.
- Clapp, J.P.; Fitter, A.H.; Young, J.P. Mol. Ecol. **1999**, *8* (6), 915–921.
- Ogunseitan, O.A.; Sayler, G.S.; Miller, R.V. Applied & Environ. Microbiol. **1992**, *58* (6), 2046–2052.
- Monpoehoe, S.; Dehee, A.; Mignotte, B.; Schwartzbrod, L.; Marechal, V.; Nicolas, J.-C.; Billadud, S.; Ferre, V. BioTechniques **2000**, *29* (1), 88–93.
- Jardi, R.; Buti, M.; Rodriguez-Frias, F.; Cotrina, M.; Costa, X.; Pascual, C.; Esteban, R.; Guardia, J. J. Virol. Meth. **1999**, *83* (1/2), 181–187.
- Anonymous, Ed.; *Development of DNA Probes for Babesia ovata Detection and Identification*, 1995. <http://ss.niah.affrc.go.jp/topics/1995/95013.html>.
- Selim, A.G.; El-Ayat, G.; Wells, C.A. J. Pathol. **2000**, *191* (2), 138–142.
- Mesarch, M.B.; Nakastu, C.H.; Nies, L. Applied and Environ. Microbiol. **2000**, *66* (2), 678–683.
- Smith, A.H.; Siegel, J.P. In Automated Microbial Identification and Quantitation. Olson, W.P., Ed.; Interpharm Press: 1996.
- Olson, W.P. Rapid Analytical Microbiology. Olson, W.P., Ed.; Serentec Press: Raleigh, NC, 2001.
- Zhang, P.; Fawcett, N.C.; Craven, R.D.; Evans, J.A. *5th Internet World Congress on Biomedical Sciences at McMaster University*, Canada 1998. www.mcmaster.ca/inabis98/cellbio/zhang0791.

DISSOLUTION AND DISSOLUTION TESTING

James L. Ford

Liverpool John Moores University, Liverpool, United Kingdom

Ali R. Rajabi-Siahboomi

Colorcon Limited, Kent, United Kingdom

INTRODUCTION

Dissolution tests are one of the tests most used in the characterization of drugs and in the quality control of dosage forms. During the late 1960s it became recognised that dissolution data should be determined by studying the rate at which dosage forms allow their formulated drug to dissolve. Subsequently, dissolution tests for six products were introduced into the USP 18 (1969). This increased to about 600 tests in the USP 24, which also includes drug-release requirements for modified release products and transdermals.

Although dissolution tests are mainly used as quality control methods to ensure end-product or batch-to-batch consistency and to identify good and bad formulations, dissolution data may also be correlated with in vivo activity. Dissolution tests become especially important if dissolution is the rate-limiting step in drug absorption. Dissolution tests are, therefore, used to confirm compliance with compendial specifications and are needed as part of a product licence application. Additionally they are used during product development and stability testing as part of the specification for the product. No universal dissolution test has been designed that gives the same rank order for in vitro dissolution and in vivo bioavailability from different formulations and batches.

MATHEMATICAL CONCEPTS OF DISSOLUTION

Dissolution of Particles

Eq. 1 is one of the oldest expressions used to describe the dissolution process of a particle:

$$\frac{dW}{dt} = \frac{D}{h} S(C_s - C_t) \quad (1)$$

where dW/dt is the rate at which a material dissolves across a surface S at a time t ; $C_s - C_t$ is the concentration gradient

between the concentration of solute in the stagnant layer (thickness h and immediately adjacent to the dissolving surface) surrounding the dissolving particles, and is assumed to be equal to the difference between the saturated solubility of the drug (C_s) and the concentration of the solute in the surrounding medium at time t (C_t). The parameter D is a function of the diffusion coefficient of the solute molecules. Maximum dissolution rates are predicted when $C_t = 0$. Consequently, as C_t increases, the dissolution rate decreases. The parameter D is also dependent on $C_s - C_t$. Such conditions, where dissolution is followed by absorption of the drug, an in vivo situation, are described as sink conditions (i.e., $C_t < 0.1 C_s$). In vitro systems should ideally maintain a sink condition and the dissolving solid should be tested in fresh solvent.

In Eq. 1 the parameter D is temperature dependent. Consequently, both the temperature of the dissolution fluid and its viscosity (which is also temperature dependent) should be carefully controlled. In addition, the presence of electrolytes and changes in pH may influence the diffusing species by altering their ionization. Such factors imply that dissolution fluids should be as simple as possible.

The constraints imposed by sink conditions may be overcome using various approaches. It may be accepted that non-sink conditions apply and that incomplete dissolution will occur. Alternatively, corrections may be made by increasing the volume of the dissolution fluids, removal of the dissolved drug by partition from the aqueous phase of the dissolution fluid to an organic phase either above or below the dissolution fluid, addition of selective adsorbents to remove the dissolved drug, addition of a water-miscible solvent to the dissolution media to increase drug solubility or removal of dissolved drug, using a flow-through system.

Dissolution of a Mono-Dispersed Powder

Dissolution processes of multiparticulate systems where the specific surface area decrease during the dissolution,

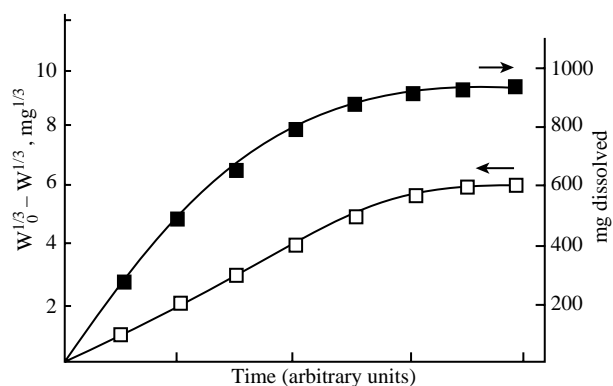


Fig. 1 Dissolution data for a hypothetical solid plotted as cumulative amount released (right axis) and after cube-root-law data treatment (left axis).

may be described by the Hixson and Crowell (1) cube root law in Eq. 2:

$$W_0^{1/3} - W^{1/3} = Kt \quad (2)$$

where W_0 represents the original mass of drug, W is the amount of remaining drug at time t , and K is the dissolution rate constant. Fig. 1 represents a typical plot where a straight-line release exists throughout the substantial period of the dissolution. The cube-root law assumes several factors, including a constant diffusion layer thickness, isotropicity of the sample, an independence of solubility of particle size, smooth surfaces, and sink conditions. Modifications of the equation are required for non-sink conditions and polydisperse systems; for example, Lai and Carstensen (2) derived shape factors to modify cube root behavior. In reality, dissolution from multiparticulate systems requires more complex mathematical approaches to dissolution. Associated problems include correctly assigning the particle size distribution to the powders, the fact that small particles have higher solubilities than larger ones, and accounting for changes in both the size and the number of particles during dissolution.

Dissolution of Disintegrating Tablets and Capsules

The development of theories of dissolution from disintegrating tablets and capsules becomes very difficult because disintegration produces vast changes in surface area. Attempts have been made to develop models to describe dissolution rates from tablets, using complex mathematical approaches (3, 4).

Dissolution of Nondisintegrating Tablets

For systems where drug release involves the dissolution of a soluble drug at high concentrations from an insoluble matrix, the Higuchi (5) equation adequately describes release rates (Eq. 3):

$$\frac{W_r}{t^{1/2}} = 2W_0 \frac{S}{V} (D/\pi\tau)^{1/2} \quad (3)$$

where W_r is the amount of drug dissolved in time t ; W_0 is the dose of the drug; S is the effective diffusional area; V is the volume of the hydrated matrix; D is the diffusion coefficient of the drug in the hydrated matrix; and τ is the tortuosity of the matrix. An analogous equation was developed for drugs of limited water solubility (6).

It is not the aim of this article to evaluate formulation aspects that may influence dissolution of drugs from dosage forms. However, it suffices to mention that dissolution rates from unformulated powders or tablets may be increased or decreased according to many factors, including salt selection, the presence of surfactants, polymorphic modifications, and the use of water-soluble carriers as in solid dispersions.

Having identified some of the important criteria and models for dissolution, let us examine some of the apparatus used to measure dissolution rates.

DESIGN OF APPARATUS

The ideal features of a dissolution apparatus are (7):

1. The fabrication, dimensions, and positioning of all components must be precisely specified and reproducible, run to run.
2. The apparatus must be simply designed, easy to operate, and useable under a variety of conditions.
3. The apparatus must be sensitive enough to reveal process changes and formulation differences but still yield repeatable results under identical conditions.
4. The apparatus, in most cases, should permit controlled variable intensity of mild, uniform, non-turbulent liquid agitation. Uniform flow is essential because changes in hydrodynamic flow will modify dissolution.
5. Nearly perfect sink conditions should be maintained.
6. The apparatus should provide an easy means of introducing the dosage form into the dissolution medium and holding it, once immersed, in a regular reliable fashion.
7. The apparatus should provide minimum mechanical abrasion to the dosage form (with exceptions) during

the test period to avoid disruption of the microenvironment surrounding the dissolving form.

8. Evaporation of the solvent medium must be eliminated, and the medium must be maintained at a fixed temperature within a specified narrow range. Most apparatuses are thermostatically controlled at around 37°C.
9. Samples should be easily withdrawn for automatic or manual analysis without interrupting the flow characteristics of the liquid. In the latter case, efficient filtering should be achieved.
10. The apparatus should be capable of allowing the evaluation of disintegrating, nondisintegrating, dense or floating tablets or capsules, and finely powdered drugs.
11. The apparatus should allow good interlaboratory agreement.

There are two principal types of apparatus design. One is based on limited volume that is constrained to the size of the container used. The second type uses a continuous-flow cell to house the dosage form and permits constant replenishment of the dissolution fluids.

COMPENDIAL APPARATUS WITH FIXED VOLUMES OF DISSOLUTION FLUID

The general principle of dissolution tests is that the powder or solid dosage form is tested under uniform agitation, which is accomplished by either using a stirrer inside the apparatus or rotating the container holding the dosage form (e.g., a basket). Two general methods are currently included in the USP 24 (8) and the BP 2000 (9) to measure dissolution from tablets and capsules. In apparatus 1, the dosage forms are enclosed in a basket; in apparatus 2, agitation is provided by means of a paddle. Each apparatus is a limited volume-type apparatus.

Basket Apparatus

The basket method was first described in 1968 by Pernarowski et al. (10). A container, the basket, constrained the enclosed tablet or capsule, allowed for fluid change and could be used either in continuous flow or in restricted volume modes. This gradually evolved to the USP 24 (8) and BP 2000 (9) apparatus 1—the rotating basket apparatus (Fig. 2). The dimensions are taken from the USP 24, although those given in the BP 2000 are similar. The apparatus consists of a motor, a metallic drive shaft, a cylindrical basket, and a covered vessel made of glass or other inert transparent material.

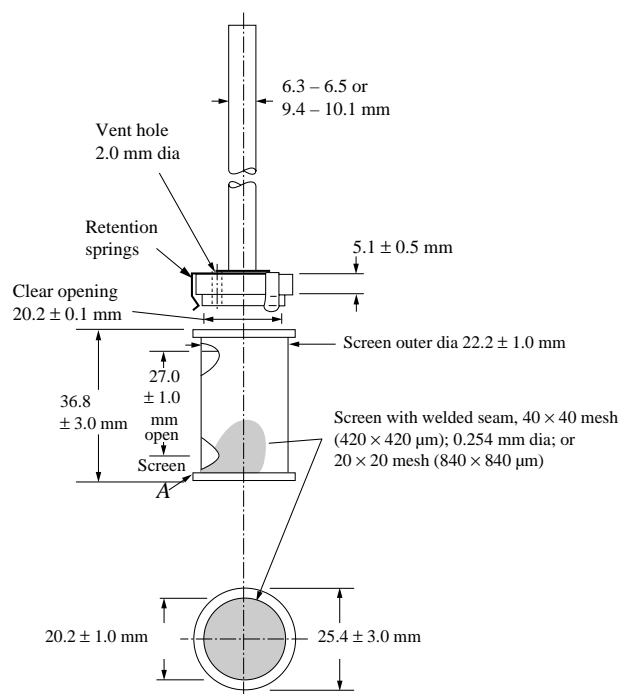


Fig. 2 The basket stirring element of the United States Pharmacopeia 24. (From Ref. 8.)

The latter should be made of materials that do not sorb or react with the specimen being tested. The contents are held at $37 \pm 0.5^\circ\text{C}$. There should be no significant motion, agitation, or vibration caused by anything other than the smoothly rotating stirring element. Ideally, the apparatus should provide observation of the stirring element and specimen. The vessel is cylindrical with a hemispherical bottom and sides that are flanged at the top. It is 160–210 mm high and has an inside diameter of 98–106 mm, and a nominal capacity of 1000 ml. Other sizes are described for 2 and 4 L capacity vessels. A fitted cover may be used to retard evaporation but should provide sufficient openings to allow ready insertion of a thermometer and allow withdrawal of samples for analysis. The shaft is so positioned that its axis is no more than 2 mm at any point from the vertical axis of the vessel and should rotate smoothly, without significant wobble. The shaft rotation speed should be maintained within $\pm 4\%$ of the rate specified in the individual monograph. The shaft has a vent and three spring clips or other suitable means to fit the basket into position. Each should be fabricated of stainless steel, type 316 or equivalent. Welded seam, stainless steel cloth (40 mesh or $425 \mu\text{m}$) is used, unless an alternative is specified. For testing, a dosage unit is placed in a dry basket at the beginning of each test. The distance

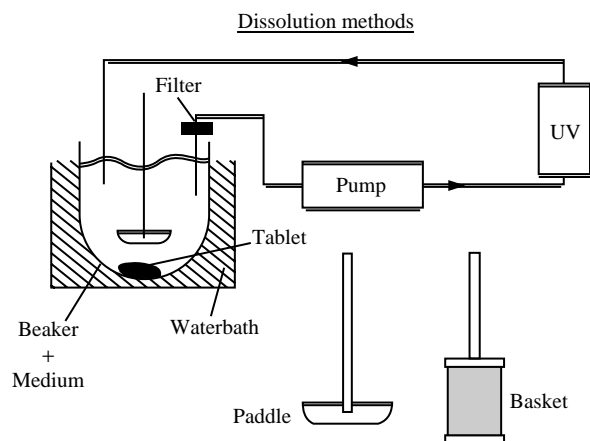


Fig. 3 Typical apparatus set up for basket and paddle apparatus.

between the inside bottom of the vessel and the basket is 25 ± 2 mm. Typical arrangements for the apparatus are shown in Fig. 3.

Although the basket apparatus in the BP 2000 and USP 24 are similar and have a common design, considerable changes have taken place since basket apparatuses were first included in official monographs. The USP 18 (11) described a cylindrical vessel with a slightly concave bottom. No precise specifications were given for the concave bottom and differences in tolerances supplied by different manufacturers were common. Consequently, statistically different dissolution rates could be obtained when determined using containers obtained from two different manufacturers (12). Flask shape had affected the hydrodynamics of systems and consequently it was considered better to have flasks of uniform hemispherical shape. The flat-bottomed flask described in the BP 1980 (13) alleviated the problems of manufacturing tolerances in vessel shape. Irrespective of apparatus design, there are still several potential problems. The wire basket corrodes following exposure to acidic media; the basket method gives poor reproducibility due to inhomogeneity of the agitation conditions produced by the rotating basket; and clogging of the basket can occur due to adhering substances. Additionally, particles can fall from the rotating basket and sink to the bottom of the flask where they will not be subjected to the same agitation as that inside the basket.

Paddle Apparatus

An apparatus described by Levy and Hayes (14) may be considered the forerunner of the beaker method. It consisted of a 400 ml beaker and a three-blade, centrally placed polyethylene stirrer (5 cm diameter) rotated at

59 rpm in 250 ml of dissolution fluid (0.1 N HCl). The tablet was placed down the side of the beaker and samples were removed periodically.

In the pharmacopoeial apparatus 2—the paddle apparatus method—a paddle replaces the basket as the source of agitation. As with the basket apparatus, the shaft should position no more than 2 mm at any point from the vertical axis of the vessel and rotate without significant wobble. The specifications of the shaft are given in Fig. 4. A distance of 25 ± 2 mm between the blade and the inside bottom of the vessel is maintained during the test. The metallic blade and shaft comprise a single entity that may be coated with a suitable inert coating to prevent corrosion. Again typical arrangements for the apparatus are shown in Fig. 3. The dosage form is allowed to sink to the bottom of the flask before rotation of the blade commences. Sinkers are recommended to prevent floating of capsules and other floatable forms. A small, loose piece of nonreactive material (e.g., a few turns of wire helix) may be attached to the dosage form. Soltero et al. (15) thoroughly examined the influence of sinker shapes on dissolution rates obtained from gelatin capsules. Although a stainless steel helix is officially recommended, alternative shapes can greatly affect the dissolution rates.

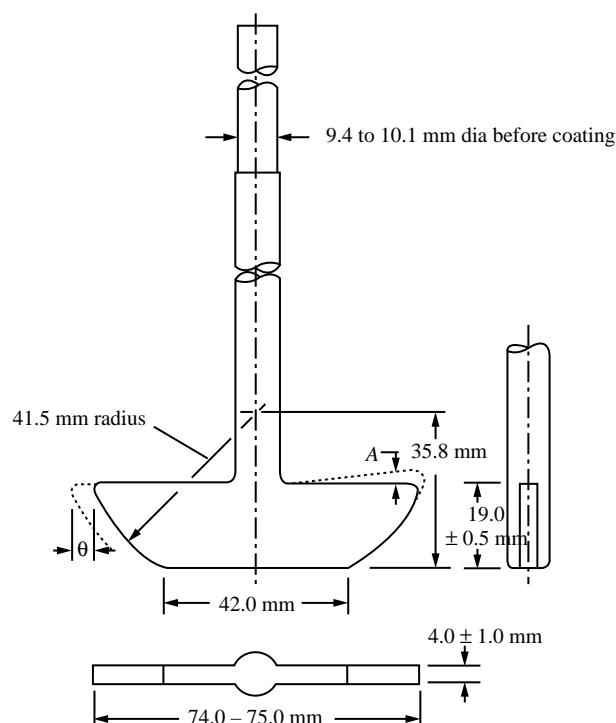


Fig. 4 The paddle-stirring element of the United States Pharmacopeia 24. (From Ref. 8.)

Data Presentation and Interpretation

The data collected during dissolution tests will, especially at the developmental stage, be presented as dissolution profiles (Fig. 5) whereby the amount released is plotted as a function of time. It is common practice to monitor drug release at several points or where possible continuously until 100% of the dose is dissolved and dissolution profiles showing drug release against time can be produced. Values equivalent to the times for 10, 50, 70, or 90% drug release are often cited ($t_{10\%}$, $t_{50\%}$, $t_{70\%}$, or $t_{90\%}$). The dissolution profiles from tablets and capsules are often sigmoidal in shape.

Pharmacopoeias generally do not require a dissolution profile to be determined but do specify that a certain amount of drug must dissolve within a specified time. Should a single time specification be stated, the test may be concluded in a shorter time if the requirement for a minimum amount dissolved is met. If two or more times are specified, then the samples should be withdrawn with a tolerance of $\pm 2\%$ of the stated time. For Digoxin Tablets B.P. (9), when there is more than one tablet per test, all six replicate runs should release at least 75% of the stated amount within 60 min.

The USP 24 (8) assesses dissolution in three stages. Should the specifications be met at either of the initial two stages, there is no requirement to proceed to later stages. In stage 1, six units are tested. To pass the test, each unit, at the specified time, should have $Q + 5\%$ in solution where Q is the amount of dissolved active ingredient expressed as a percentage of the labelled content and 5% is the percentage of labelled content. Failure at stage 1 requires a second stage test on an additional six units. To pass the test at this stage, the average content dissolved, from the combined two

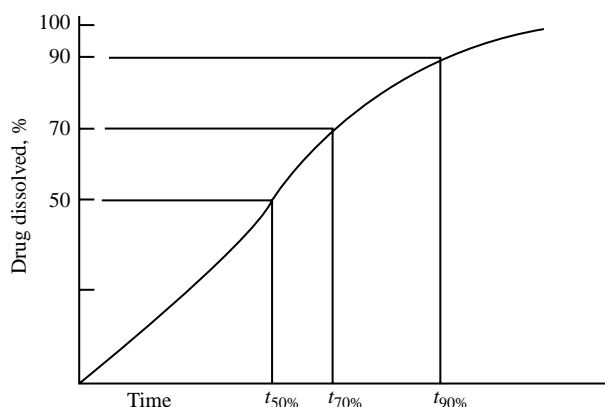


Fig. 5 Dissolution profiles of a hypothetical dosage form indicating the determination of $t_{50\%}$, $t_{70\%}$, and $t_{90\%}$ values.

stages, should be equal to or greater than Q with no unit being less than $Q - 15\%$. Failure leads to stage 3 where a further 12 units are tested. These results are combined with the results from the previous stages. The average of the total of the 24 units thus tested should be equal to or greater than Q . No more than two units should be less than $Q - 15\%$ and no unit should be less than $Q - 25\%$.

Sample Collection Procedures

The USP 24 (8) and the BP 2000 (9) state that samples should be drawn from a zone midway between the surface of the dissolution medium and the top of the rotating basket or blade, not less than 1 cm from the vessel wall. A volume of media equal to the volume of sample withdrawn should be replaced or compensated for by calculation. Samples removed should be filtered. Any substance released from the filter should contribute no significant absorption to the active ingredient in solution or interfere with the assay. No extractable materials interfering with the analytical procedures should be released. The filter pore size should not be greater than 1 μm . When capsule shells interfere with the analysis, the contents of a minimum of six capsules should be removed as completely as possible and the shells dissolved in the dissolution medium. This allows calculation of correction factors, which should be no greater than 25% of the labelled content.

Factors That May Influence Dissolution

Both the BP 2000 (9) and the USP 24 (8) give considerable guidance about external factors that may influence dissolution. This section highlights some of the more important factors.

Vibrations

No part of the assembly, including the environment in which the assembly is placed, should contribute significant motion, agitation, or vibration beyond that caused by a smoothly rotating element. There is little doubt that excessive vibration of the dissolution apparatus considerably increases dissolution rates. Therefore, to minimize the effects of vibration, the distance from the dissolution motor to the stand supporting it and the distance from the rotating basket to the point connecting the shaft to the dissolution motor should be specified (16). Vibrations may arise from various sources, including water bath and pump.

Shaft

Many problems associated with variations in the dissolution rate are caused by misalignment. Limits should be set on shaft eccentricity. Minor changes in physical alignment of the paddle may produce large variations in results.

Vessel

The use of a hemispherical flask allows some variation in construction. Minor changes in vessel shape may considerably alter the dissolution rates determined by the paddle method (17). Glass vessels with an inside bottom flatter than specified gave a high bias in dissolution rates and those with an inside bottom of steeper curvature gave a low bias. Such bias is associated with the different hydrodynamics created by the different shapes. Plastic vessels provide more perfect hemispheres than glass vessels and their shape is more reliable and should be preferred to glass containers provided the drug does not sorb to them and that the dissolution fluid does not interact with the plastic (17). However, plastic containers are not without their problems. During dissolution, some tablets may become more centred in glass vessels than in plastic vessels, which is probably attributable to friction between the tablet and the wetted vessel surface being greater for plastic than for glass (18).

Sampling procedures

Flow through facilities usually allow ultraviolet (UV) analysis of drug dissolved or collection of samples for subsequent analysis (Fig. 3). Filtering must be accomplished before analysis to prevent insoluble excipients or dissolving drug particles from passing through the beam. Filters remove solid particles prior to assay at the stage of sample removal, reduce turbidity problems caused by undissolved drug and excipients, and help eliminate spurious results caused by particles dissolving following removal.

A sampling probe used for continuous sampling must not disturb the hydrodynamics of the system. When continuous monitoring is accomplished by flow-through systems, the flow rate should not be too high or some of the suspended drug will adhere to the filter, resulting in a percolation effect and artificially high estimates of the amount of drug released. Care should be taken with automatic flow methods so that tube length is kept to a minimum to reduce the amount of fluid in the tubing. Excessive amounts will produce long lag periods before dissolved material is transported from the dissolution vessel to the analyzing cell.

Adsorption of the drug onto the filter should be avoided and the release of isooctylphenoxypolyethoxyethanol, which is often included in the membrane filters as a wetting agent, may be a potential problem (19). This surfactant has significant absorption in the 220–240 nm wavelengths. Selection of a suitable pore size is important. Adsorption onto filters of frusemide was increased by using filters of decreasing pore size but of the same diameter (20). Replacement of sampled fluid must be by fluid of the same temperature as that inside the dissolution vessels. Recent developments have focused on the possibility of utilizing a UV-fiber-optic probe to pipe light in the dissolution medium where a spectrophotometric assay *in situ* through fiber-optic light guides may be carried out (21).

Temperature control

The USP directs that the thermometer should be removed before the test and that the temperature should be checked periodically. The dissolution fluids should be maintained at $37 \pm 0.5^\circ\text{C}$. Even slight temperature variations may have a significant effect on tablet dissolution.

Deaeration of dissolution medium

The levels of dissolved gas are not specified in either the BP 2000 (9) or the USP 24 (8); each states that media should be deaerated before use. By testing the dissolution of prednisone tablets in water that had been presaturated with air, Cox et al. (22) showed that the media must remain unsaturated during the test. Higher initial air content resulted in increased percentages of drug dissolved during the paddle method. Air bubbles, in the media that contained higher levels of air, attached to particles of the disintegrating tablets owing to sorption of air. The particles did not sink to a mound at the bottom of the vessel but were suspended by the air bubbles and were, therefore, subject to greater conditions of agitation.

Failure to deaerate media used with the basket apparatus results in the clogging of the pores of the basket by air, causing different flow characteristics (19) and often a decrease in dissolution rates. This has even greater influence when gel-forming extended-release dosage forms are being tested. Bubbles attached to tablets or capsules may cause the dosage form to relocate near to the top of the basket, thereby reducing dissolution rates or amounts dissolved (19). The basket is intended to hold the tablet in a fixed position. Disintegration of the dosage form will occur in the basket and fluid must flow through the basket sufficiently to disperse the tablets and sweep dissolved drug into the bulk of the dissolution fluid.

Variation in speed of agitation

Agitation speed must be uniform throughout the test. Some motor drives result in a satisfactory mean speed but during the test will periodically slow down or speed up. Consequently, speed should be checked at the start and end of each run (19).

Standardization and Calibration

All apparatuses must be calibrated and the variables standardized and known. Experimental design in dissolution testing to examine residual variation between experimental runs and individual dissolution vessels is feasible (23). The objective is to minimize errors in experimental setup, achieved by using mean dissolution times and partial balancing. Lower coefficients of variation occur at higher hydrodynamic intensities. Dissolution tests are critical but difficult to carry out properly. Standardization should remove all variable factors. Similar problems of fluid flow exist with the current USP 24 apparatus, although the hemispherical shape will create different fluid hydrodynamics.

The USP 24 (8) specifies apparatus suitability tests, based on the operating conditions, involving the USP dissolution calibrator, disintegrating type or the USP dissolution calibrator, nondisintegrating type. The apparatus is deemed suitable if the obtained results are within the accepted range stated for that type of calibrator. Calibrators are used to show deficiencies in the equipment, for example, chain looseness, tilting of stirrer motor, excessive vibration, or misalignment of the flasks with the stirrers. Nondisintegrating calibrants are composed of 300 mg salicylic acid, whereas disintegrating calibrants contain 50 mg prednisone. The dissolution rates obtained from the calibrants should fall within established ranges at both 50 and 100 rpm.

Other Variables in Compendial Methods

The BP 2000 (9) generally specifies 1000 mL of dissolution fluid, although for digoxin tablets the volume is 600 mL water. Dissolution fluid volume is varied more in the USP 24 (8). Examples include 500 mL (alprazolam tablets), 750 mL (metyrosine capsules), and 900 mL (ampicillin capsules). The volume may additionally vary as to the strength of the preparation. Thus, for phentermine hydrochloride capsules 500 mL of medium is used if the strength is 15 mg or less, but 900 mL is used for preparations containing in excess of 15 mg. Similar volumes are used for prednisone tablets where the cut-off point is 10 mg and for cinoxacin capsules where 500 mL is

used for capsules containing 250 mg or less but 1000 mL is used for capsules containing in excess of 250 mg.

OTHER METHODS USING LIMITED VOLUMES OF DISSOLUTION FLUIDS

Many other, nonofficial methods that use limited volumes of dissolution fluids have been described. These include the rotating-disc method that measures the dissolution rate from a constant surface area, which is usually in the form of a disc (24). Intrinsic dissolution rates, which relate dissolution rate to the surface area of the dissolving drug, can easily be determined from constant-surface area discs whose areas can be very easily estimated. The rotating filter-stationary basket apparatus (25) uses a large capacity (1.5 L) glass flask containing a stationary sample basket and a rotating-filter basket.

Other apparatuses that have been evaluated for dissolution testing include the magnetic basket apparatus (26) and the commercial Bio-Dis apparatus (27). In the later apparatus, the use of vertically reciprocating tubes, sealed with mesh discs at each end to restrain the dosage form was developed. The USP 24 (8) describes, as drug release apparatus 3, the reciprocating cylinder method. The tubes are raised up and down, 9.9–10.1 cm, in dissolution fluid. The apparatus has been advocated to monitor release from pellet systems although it tended to higher values of amount released than other apparatus (27).

Limited-volume apparatus, with a finite volume of dissolution fluid, suffer from the problem that they operate under non-sink conditions, which results in limitations when poorly soluble drugs are considered. Some apparatus, previously described, may be modified by the use of flow-through systems and reservoirs to provide sink conditions by removing solvent and replacing it with fresh solvent. Continuous circulation may also be used for conditions when sink conditions are not required. The drawbacks of nonflow-through apparatus (28) include (1) lack of flexibility, (2) lack of homogeneity, (3) the establishment of concentration gradients, (4) their semiquantitative agitation, (5) the obscuring of details of the dissolution processes, and (6) their variable shear. Consequently, flow-through apparatus characterized by a dissolution cell of low volume (often less than 30 mL) and a reservoir to provide fresh solvent have been developed.

Flow-Through Apparatus

The USP 24 flow-through cell (24) may be either used in closed mode when the fluid is recirculated and, by

necessity, is of fixed limited volume, or open mode when there is continuous replenishment of the fluids. The basic components are reservoir, pump, heat exchanger, column, tablet support, filter system, and analytical method. The systems enable solvent to be taken from a suitable reservoir and passed straight through the apparatus containing the dosage form to be either assayed and removed (effluent system) or recirculated (recycling system) (Fig. 6).

The design of the pump to remove the solvent from the reservoir is crucial to the results obtained from such systems. The pump used may be either a displacement (oscillating or peristaltic) or a momentum (centrifugal) type. However, peristaltic pumps may create oscillations that result in faster dissolution rates than might otherwise have occurred. Dissolution is affected by factors such as the volumetric flow rate, the cross-sectional area of the cell, the initial drug quantity, liquid velocity, and drug concentration.

Ascending fluid flow is used. Flow-through facilities remove some of the problems associated with non-sink conditions. The material under test (tablet, capsules, or granules) is placed in the vertically mounted dissolution cell, which permits fresh solvent to be pumped (between 240 and 960 mL/h) in from the bottom. Glass beads (1 mm diameter) control the flow across the whole cross-sectional area of the cell. Cells may have diameters of 12 or 22.6 mm.

The maintenance of a controlled flow is crucial to column methods and can be influenced by the inlet system. The design of the apparatus was improved by using a holder constructed of a folded wire cross above the inlet pipe to support the tablet such that the solid is bathed in the outflow contained within a concentric column of relatively stationary liquid (29). It is common to place the tablets on such supports, but attrition (by glass beads) may encourage breakdown of the dosage form thereby increasing dissolution rates. Tablet support and consistent positioning

in the liquid flow are prerequisites for consistent results. Attempts have been made to embed the tablet in glass beads.

DISSOLUTION MEDIA

Ideally, a dissolution medium should be formulated as close as possible to that pH anticipated in *in vivo* fluids; for example, dissolution media based on 0.1 *N* HCl are used to mimic gastric pH. Simulated gastric fluid is similarly used. Food can increase the gastric pH to as high as 3–5. Many compendial dissolution fluids are at a pH near neutral despite the fact that tablets, when swallowed, will meet a lower gastric pH. The use of surfactants and enzymes may also be a coarse approximation of the intestinal fluids, although surfactants may be included to increase drug solubility by solubilization into micelles. Both the USP 24 (9) and the BP 2000 (8) indicate that the pH of dissolution-fluid should be within 0.05 of that specified in the relevant monograph.

The inclusion of alcoholic solvents, for example, isopropanol, has been advocated for insoluble drugs such as norethindrone and was used to provide a rank order of dissolution (30). The dissolution rates increased in hydro-alcoholic medium. Each component of a norethindrone–mestranol combination tablet gave its own dissolution rate in deaerated water but was not discriminated in the hydro-alcoholic medium. Therefore, for quality control purposes, all components of a multiple drug formulation should have their dissolution assessed. Problems are induced by the inclusion of cosolvents. Alcoholic solvents may retard dissolution rates by retarding disintegration as, for example, during the dissolution of chlorthalidone into 40% aqueous methanol (31).

Sink conditions may be accomplished by using dissolution media consisting of two phases to allow partition of the drug into the organic phase; the use of organic solvents such as hexane may provide sink conditions. Supramicellar surfactant concentrations may be used for drugs of low solubility (<0.01% w/v). Consequently, sodium dodecyl sulfate may be added to the media. However, the inclusion of surface agents into the dissolution media may make the media less discriminatory. The use of 0.4% sodium dodecyl sulfate was proposed as an alternative to 0.1 or 0.01 *M* HCl for the dissolution of rifampin (32) because the drug was more stable in surfactant solutions. After 1 and 2 h, in 0.01 *M* HCl, 13 and 29% of rifampicin decomposed, at 37°C. After 2 h, only 2% degraded in water.

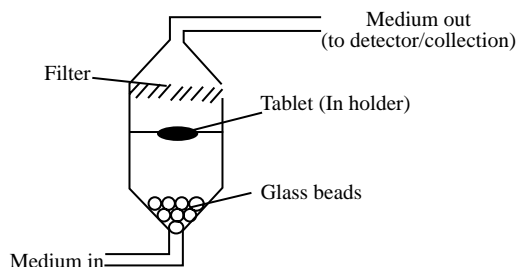


Fig. 6 Schematic diagram of a flow-through dissolution cell.

IN VITRO–IN VIVO CONSIDERATIONS

Ideally, dissolution tests should provide data to distinguish good and bad formulations, changes in production processes that might influence bioavailability, batch to batch reproducibility, site and manufacturer variation, and information on the stability of the product. The test is used in the development of new dosage formulations, for quality control and to give indications for bioequivalence. In addition, one may consider the dissolution testing as a prognostic tool for in vivo performance (33).

Great care must be taken to produce good in vitro–in vivo correlations. This may be accomplished by creating a reproducible dissolution method that provides proper selection of medium, degree of agitation, etc. Ideally, this could be achieved if the gastrointestinal tract conditions were successfully reconstructed in the dissolution apparatus.

Using dissolution parameters, dissolution rate can be related to bioavailability for some drugs. For example, Kingsford et al. (34) showed a linear relationship between the percentage dissolved in 30 min and the bioavailability of frusemide relative to an oral solution when frusemide was examined using the rotating basket apparatus and buffer at pH 5.0. Therefore, in certain cases, it may be appropriate to apply dissolution testing data to evaluate biopharmaceutical implications of a product change rather than automatic bioequivalence study (35). The USP 24 (8) details the requirements for in vitro and in vivo evaluation of dosage forms, including immediate release and modified release dosage forms.

DISSOLUTION TESTING OF MODIFIED-RELEASE ARTICLES

The USP 24 (8) also gives tests (termed “drug-release” tests rather than dissolution tests) for modified release products, which includes extended-release (sustained or slow-release) products and delayed-release (enteric-coated) articles. The former are described as allowing at least a two-fold reduction in the dosing frequency as compared with that drug presented as a conventional dosage form. A delayed-release article is one that releases the drug at a time other than promptly after administration.

The USP 24 testing of both extended-release articles and delayed-release articles proceeds via three stages. For extended-release products three time points are generally

specified, expressed in terms of the labelled dosing interval (D in hours). The amounts dissolved (Q) at each time interval are expressed in terms of the labelled content; the limits embrace each value of Q , the amount dissolved at each specified dosage interval (8). Initially six units are tested. No unit should lie outside the stated range or be less than the stated amount at the final test time. Failure at this stage leads to examination of a further six units whose results are combined with the initial results. The average of the 12 units should not lie outside the stated range and should not be less than the stated amount at the final test time. Additionally none should be more than 10% outside each of the stated ranges and none more than 10% of the labelled content below the stated amount at the final test time. Failure to meet these standards requires the testing of a further 12 units, the data being combined with those from the previous 12 units. The average of the 24 units is within each of the stated ranges and not less than the stated amount at the final test time. No more than two of the 24 units should release more than 10% outside each of the stated ranges; no more than two of the 24 units should release more than 10% below the stated amount at the final test time; and none of the units should be more than 20% of the labelled content outside each of the stated ranges or more than 20% of the labelled content below the stated amount at the final test time.

DISSOLUTION TESTING FOR ENTERIC-COATED TABLETS AND CAPSULES

The USP 24 (8) describes two general methods for delayed-release (enteric coated) articles, each method utilizing a change in pH. In method A, an acid stage consisting of 750 mL of 0.1 M HCl is used. An aliquot is withdrawn after 2 h for assay. The pH is then raised by adding 250 mL of 0.2 M tribasic sodium phosphate and adjusted to pH 6.8 ± 0.05 and the dosage forms are operated for a further 45 min. Method B utilizes 1000 mL of 0.1 M HCl operated for the initial 2 h period and the fluid is completely replaced with 1000 mL pH 6.8 phosphate buffer and run for 45 min. Again, three levels of testing are used for the HCl stage (8).

In addition to the official tolerances for delayed-release products, commercial manufacturers will have their own in-house specifications. Standards for enteric-coated forms ensure the integrity of the coat, but consideration should be given to a pH change method for evaluating extended-release medications. This evaluation should ensure that

dose dumping does not occur at a specific pH, as might be a problem with delayed-release dosage forms based on, for example, hydroxypropyl methylcellulose phthalate.

DRUG RELEASE FROM SUSTAINED-RELEASE ARTICLES

There have been various testing schedules that have been developed to study the influence of pH changes, for example, from 0.1 *N* HCl to pH 7.5 (36) or a progressive rise from pH 1.5, 4.5, 6.9, 7.2–7.5 over 22 h (37). Such changes are designed to follow the pH changes during the passage of a dosage form through the gastrointestinal tract, but the residence time in the tract can be as little as 4–6 h up to in excess of 24 h depending on the food that the patient has ingested. An alternative approach is to develop topographic profiles of the matrix tablets to determine the dissolution profiles at a variety of pHs (38); an apparent three-dimensional profile of amount released versus time versus pH is produced. Such profiles may be used to determine over which pH range dosage forms release their drug at a faster rate than required, thus avoiding the problem of dose dumping at a particular narrow pH range.

In the development stage of new modified release dosage forms, it is vitally important to not only consider dissolution testing as a quality control tool but also as a prognostic tool to discriminate in vivo good and bad formulations. Therefore, one needs to understand the mechanism of drug release from the dosage form in order to utilise the dissolution testing effectively. For extended-release reservoir formulations, where the drug was dissolved and released by diffusion through a barrier coat, factors such as the influence of medium on solubility of the drug need to be established.

In the case of hydrophilic-matrix sustained-release formulations, drug release is via diffusion through and erosion of the hydrated viscous surface polymer. Therefore, factors that may provide additional strength or support to this surface polymer, must be minimised. For example, during the dissolution testing of a large swelling matrix, it may be supported by the basket wall, which combined with slow speed of rotation leads to unrealistically slow drug release. This is in contrast to in vivo where due to agitation of the GIT and also presence of food, the level of erosion is greater and thus faster drug release may occur. A discriminating dissolution testing would assist the formulator to achieve the desired in vivo drug release profiles, through the formulation variables.

Figure 7 illustrates how the type of apparatus and also speed of rotation influences the in vitro drug release profiles of a HPMC matrix formulation (39). In another study, using HPMC matrix formulations, the in vitro data were successfully correlated to in vivo data with an excellent correlation when apparatus 1 at 150 rpm was used (40).

Matrix-sustained release tablets may present additional problems since they tend to stick to the side walls of the vessel, giving a reduction in the partial exposure of surface area and the use of the paddle may destroy the discreteness of the dosage form. Moreover, media containing high concentrations of phosphates may result in the disintegration of sustained-released matrices based on hydroxypropyl methylcellulose.

Pellets, beads, and other particulate systems intended to provide prolonged release are probably best tested by column methods that facilitate changes in pH or dissolution fluids. With compendial apparatuses, problems arise; for instance, it is difficult to capture dispersed particles when the paddle method is used, and small particles may pass through the mesh of the basket in the basket method.

Recently a simple (41) model was proposed using mathematical indices to define a similarity factor f_2 , to compare dissolution profiles as in Equation (4):

$$f_2 = 50 \log\{[1 + 1n_t^{n-1}(R_t - T_t)^2]^{-0.5} \times 100\} \quad (4)$$

where R_t and T_t are percent dissolved at each time point for the reference product and the test product, respectively. Using the f_2 values, dissolution profiles are considered dissimilar if these values were less than 50 with average difference between any dissolution samples not being greater than 50%. The similarity factor and a similarity

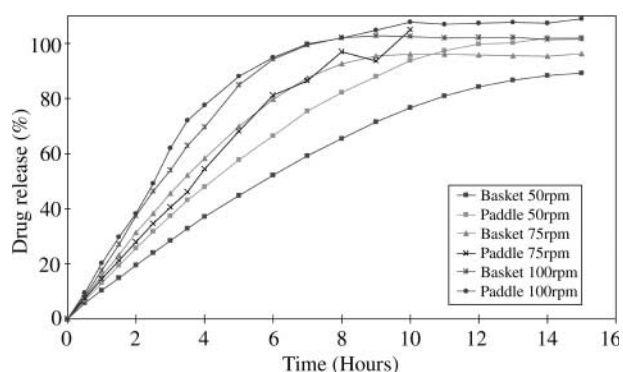


Fig. 7 Dissolution profiles of diltiazem hydrochloride from a swelling matrix system in phosphate buffer pH 6.8 at 37°C, using the basket and paddle methods rotating at 50, 75, and 100 rpm.

testing have been recommended for dissolution profile comparison in the FDA Guidance for Industry (35, 42, 43).

SUPPOSITORIES AND OTHER SEMISOLIDS

Semisolids are intended for percutaneous (ointments, creams, gels, patches), vaginal (pessaries), ophthalmic (ointments, creams, gels), or rectal (ointments, creams, suppositories, gels) release. Some estimate of drug release is desirable to ensure that the drug is not entrapped by preferential partitioning into the oleaginous phase. Unlike tablets and capsules, there must be some constraint on the surface area of such preparations exposed to dissolution fluids to produce a quantitative estimate of release during dissolution studies.

Suppositories create problems because they soften, deform, disintegrate, and change from a solid to an oily state during the test. Glass beads will control the interfacial area during a dissolution test; consequently, the use of a continuous-flow bead-bed column-type apparatus was advocated for the dissolution of suppositories (44). Suppositories based on soluble carriers such as polyethylene glycol create fewer problems in the assessment of drug release since the carrier rapidly dissolves. However, oleaginous-based suppositories (e.g., containing cocoa butter) will melt slowly and spread, giving a constantly changing surface area. Consequently, membranes are frequently employed to entrap the suppository although they may themselves be the factor that limits drug transport. In addition, basket methods are unsuitable since the basket mesh will readily clog although polyurethane baskets of dimension similar to the USP basket have been used. To design a realistic apparatus for both types of suppository probably require two phases.

Transdermal patches have also been examined. The USP 24 (8) includes reference to transdermal drug delivery systems. The temperature of operation is specified as $32 \pm 0.5^\circ\text{C}$. Apparatuses include a modified paddle method, using a disc, 41.2 mm in diameter, to locate the patch at the bottom of the dissolution flask, a cylinder method whereby the patch is attached to the outside of a stainless steel cylinder of radius approximately 2.2 cm and rotated in the dissolution flask, or a reciprocating disc apparatus such that the disc may be reciprocated at 30 cycles/min over an amplitude of 2 cm in volumetrically calibrated solution containers.

ROBOTICS

Dissolution by manual methods is very labor-intensive and time-consuming. Therefore, the use of microprocessors is cost-justified, especially when some tests may take up to 24 h. Robotics speed up assay techniques; rapid insertion and withdrawal of a small sample probe, controlled robotically, produces minimum disturbance of the hydrodynamic conditions. The use of robotics increases the speed of assay, the throughput and the productivity of dissolution testing. However, the speed of assay may then be the rate-limiting step to productivity. Consequently, to speed up the assay process, short high-pressure liquid chromatography (HPLC) columns may be used to increase assay speed.

ASSAY METHODS

Whatever assay technique is chosen, it must be specific to the drug under assay. For single-drug entities, UV or HPLC assays are most suitable and most used; even ion-specific electrodes can be used to monitor drug release in the presence of a suitable anion or cation. However, with any trend to multiple dosage forms, more complicated assay techniques may be required, which makes HPLC useful since more than one drug can be analyzed in the same chromatogram. The use of UV photodiode array spectrophotometry has become popular with multidrug dosage forms.

REFERENCES

1. Hixson, A.; Crowell, J. *Ind. Eng. Chem.* **1931**, 23, 923–931.
2. Lai, T.Y.-F.; Carstensen, J.T. *Int. J. Pharm.* **1978**, 1, 33–40.
3. Laakso, R.; Kristoffersson, E.; Marvola, M. *Int. J. Pharm.* **1984**, 19, 35–42.
4. Kitazawa, S.; Johnno, I.; Minouchi, T.; Ito, Y.; Okada, J. *J. Pharm. Pharmacol.* **1977**, 29, 585–588.
5. Higuchi, W.I. *J. Pharm. Sci.* **1962**, 51, 802–804.
6. Higuchi, T. *J. Pharm. Sci.* **1963**, 52, 1145–1149.
7. Lathia, C.D.; Banakar, U.V. *Drug Dev. Ind. Pharm.* **1986**, 12, 71–105.
8. *United States Pharmacopeia 24* The U.S. Pharmacopeial Convention: Rockville MD, 1999.
9. *British Pharmacopoeia 2000* Her Majesty's Stationery Office: London, 2000; II.
10. Pernarowski, M.; Woo, W.; Searl, R.O. *J. Pharm. Sci.* **1968**, 57, 1419–1421.

11. *United States Pharmacopeia 18* The U.S. Pharmacopeial Convention: Rockville, MD, 1969.
12. Rosolia, A.M.; O'Connell, J.R.; Bavitz, J.F.; Restaino, F.A.; Schwartz, J.B. *J. Pharm. Sci.* **1972**, *61*, 1638–1640.
13. *British Pharmacopoeia 1980* Her Majesty's Stationery Office: London, 1980.
14. Levy, G.; Hayes, B.A. *New Engl. J. Med.* **1960**, *262*, 1053–1058.
15. Soltero, R.A.; Hoover, J.M.; Jones, T.F.; Standish, M. *J. Pharm. Sci.* **1989**, *78*, 35–39.
16. Embil, K.; Torosian, G. *J. Pharm. Sci.* **1979**, *68*, 1336–1338.
17. Cox, D.C.; Wells, C.E.; Furman, W.B.; Savage, T.S.; King, A.C. *J. Pharm. Sci.* **1982**, *71*, 395–399.
18. Cox, D.C.; Furman, W.B. *J. Pharm. Sci.* **1984**, *73*, 1125–1127.
19. Cartwright, A.C. *Drug Dev. Ind. Pharm.* **1979**, *5*, 277–291.
20. Ford, J.L.; Poole, V. *Drug Dev. Ind. Pharm.* **1991**, *17*, 639–647.
21. Bynum, K.C.; Kraft, E. *Pharm. Technol.* **1999**, *23*, 10.
22. Cox, D.C.; Furman, W.B.; Page, D.P. *J. Pharm. Sci.* **1983**, *72*, 1061–1064.
23. Lewis, G.A.; Stevens, H.N.E. *Drug Dev. Ind. Pharm.* **1987**, *13*, 1809–1816.
24. Levy, G.; Sahli, B.A. *J. Pharm. Sci.* **1962**, *51*, 58–62.
25. Shah, A.C.; Peot, C.B.; Ochs, J.F. *J. Pharm. Sci.* **1973**, *62*, 671–677.
26. Needham, T.E.; Shepherd, R.E.; Luzzi, L.A. *J. Pharm. Sci.* **1973**, *62*, 470–472.
27. *Pharmacopeial Forum* **1989**, *15*, 4984–4992.
28. Tingstad, J.E.; Riegelman, S. *J. Pharm. Sci.* **1970**, *59*, 692–696.
29. Groves, M.J.; Alkan, M.H.; Deer, M.A. *J. Pharm. Pharmacol.* **1975**, *27*, 400–407.
30. Goehl, T.J.; Sundaresan, G.M.; Prasad, V.K. *Int. J. Pharm.* **1983**, *15*, 115–123.
31. Poirier, H.; Lewis, G.A.; Shott, M.J.; Stevens, H.N.E. *Drug Dev. Ind. Pharm.* **1983**, *9*, 443–452.
32. Gharbo, S.A.; Cognion, M.M.; Williamson, M.J. *Drug Dev. Ind. Pharm.* **1989**, *15*, 331–335.
33. Dressman, J.B.; Amidon, G.L.; Reppas, C.; Shah, V.P. *Pharm Res.* **1998**, *15*, 11–22.
34. Kingsford, M.; Eggers, N.J.; Soteros, G.; Maling, T.J.B.; Shirkey, R.J. *J. Pharm. Pharmacol.* **1984**, *36*, 536–538.
35. *Immediate Release Solid Oral Dosage Forms, Scale Up and Post Approval Changes, and In Vivo Bioequivalence Documentation* CDER; FDA: USA, 1995; FDA Guidance for Industry.
36. Jalal, I.; Zmaily, E.; Nazib, N. *Int. J. Pharm.* **1989**, *52*, 63–70.
37. Graepel, P.H.; Momberger, H.; Kuhn, D. *Pharm. Zt.* **1986**, *131*, 1820–1823.
38. Skelly, J.P.; Yamamoto, L.A.; Shah, V.P.; Yau, M.K.; Barr, W.H. *Drug Dev. Ind. Pharm.* **1986**, *12*, 1159–1175.
39. Rajabi-Siahboomi, A.R., Ed.; unpublished data.
40. Eddington, N.D.; Marroum, P.; Uppoor, R.; Hussain, A.; Augsburger, L. *Pharm. Res.* **1998**, *15*, 466–473.
41. Moore, J.W.; Flanner, H.H. *Pharm. Tech.* **1996**, *20*, 64–74.
42. *Dissolution Testing of Immediate Release Oral Solid Dosage Forms* : , 1997, FDA Guidance for Industry.
43. *Extended Release Oral Solid Dosage Forms, Scale Up and Post Approval Changes (SUPAC-MR)*, 1997, FDA Guidance for Industry.
44. McElnay, J.C.; Nichol, A.C. *Int. J. Pharm.* **1984**, *19*, 89–96.

DESIGN OF DRUGS: BASIC PRINCIPLES AND APPLICATIONS

Jacques H. Poupaert

Université Catholique de Louvain, Brussels, Belgium

INTRODUCTION

Drug design and discovery techniques have evolved considerably in recent years but they have progressed in two opposite directions. On the one hand, considerable intellectual and financial efforts have been made to implement a rational sequence of events that would ultimately result in the identification of the successful drug candidate. One of these scenarios is known as “structure-based drug design.” At the turn of the millennium, this can be considered as the ultimate stage of development of “rational drug design” (1–4). On the other hand, with the potentialities offered by automated chemical synthesis and robotized biochemistry, the temptation has been ever greater to produce increasingly huge numbers of molecules by combinatorial chemistry techniques and include them in batteries of high throughput screening (HTS). The process of finding novel, active compounds through combinatorial chemistry is akin to finding a needle in a haystack. While per se not irrational, this process can be viewed as using a Monte-Carlo algorithm and as such it can be easily anticipated that convergence in the selection of the successful drug candidate will be inevitably slow. The challenge of the next century will be to reconcile these two approaches to significantly accelerate both discovery and preclinical research. Building a fully integrated, high-throughput drug-discovery platform spanning lead generation through investigational new drugs (INDs) at present remains a critical step to be made in order to efficiently transform proprietary medicinally designed combinatorial libraries into wholly or substantially owned new chemical entities (NCEs).

STRUCTURE-BASED DRUG DESIGN

Historically, the pharmaceutical industry has relied on discovering drug leads by screening-sifting through vast inventories of either naturally occurring or manmade chemicals in search of previously undiscovered substances with the desired biological activity. Traditionally, optimization of a lead compound was achieved by random exploration of a chemical structure (often called lead

compound) through the synthesis of large numbers of chemical derivatives. Over the last decade, this approach has become increasingly unsatisfactory, since it is costly and inefficient, and while the rate of discovery of new therapeutic compounds and the marketing of new drugs has declined, the costs have continued to rise. Most importantly, there remain many important therapeutic needs for which screening-based research has failed to yield acceptably safe and effective drugs.

Almost all biologically active molecules work via an interaction with a “target” or “receptor” molecule that plays key roles in all biological processes. In the most common perception of this interaction, the drug molecule inserts itself into a functionally important cleft of the target protein, such as a key in a lock. The molecule then binds there and either induces or more commonly inhibits the protein’s normal function. This universal drug-target scheme suggests an interesting alternative approach to drug discovery. Indeed, if it were possible to identify in advance the appropriate protein target for a searched therapeutic need and if enough information were known about the structure of that target protein, it ought to be feasible to design the structure of an ideal drug to interact with it. This approach, called structure-based drug design (5–10), offers the promise of eliminating much of the inefficiency of the classical approaches of conventional drug discovery.

Scientists therefore developed an approach for drug discovery that exploits the 3-D structures of molecular targets (protein structure-based drug design; Fig. 1). At the heart of this strategy is protein X-ray crystallography, which enables one to determine the 3-D atomic structures of target proteins and the drugs that bind to them. This novel approach to drug design integrates genetic engineering techniques that allow the identification, purification, and modification of appropriate target proteins with innovations in protein X-ray crystallography (11–14). This approach also uses elaborated softwares that permit chemists to predict molecular structure in its dynamic and thermodynamic extensions. In contrast to the biotechnology industry, several pharmaceutical groups nowadays employ genetic engineering techniques to produce proteins not as products, but as drug targets. Genetic engineering techniques assist scientists in identifying molecular targets for particular therapeutic objectives, produce target

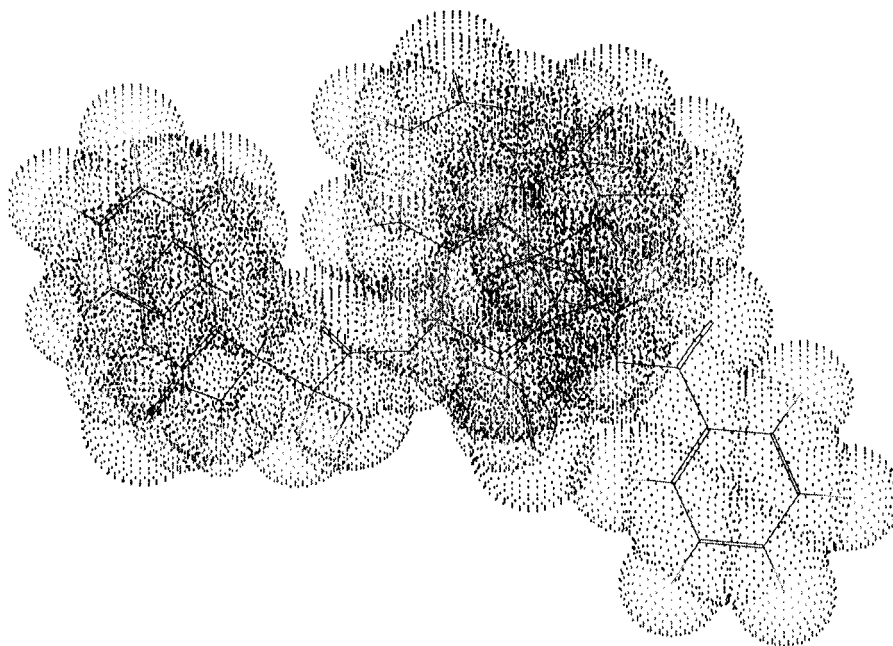


Fig. 1 The 3D vision of pharmacologically active molecules (here taxol) provides the medicinal chemist with a complex information from which a pharmacophore has to be deduced to generate analogs with more adequate pharmacokinetic and toxicological properties.

proteins in sufficient amounts to permit structural studies, and modify these proteins to probe the connections between a target protein's structure and its physiological or pathological functions. So far, the only method that has been successful in determining the precise 3-D atomic structure of large proteins is X-ray crystallography. An important limitation of the method is that X-ray crystallographic studies require a target protein in crystalline form. A powerful X-ray beam bombards a single protein crystal, which diffracts the X-ray beam and generates a definite diffraction pattern. A complex analytical process involving extensive mathematical computations then has to be performed on the X-ray diffraction data. The results of these calculations determine the target protein's exact 3-D structure. It is this information that provides the critical starting point for 3-D drug design. Alternatively, when the target protein cannot be obtained in crystalline form, 2-D nuclear magnetic resonance (2D NMR), or molecular modeling techniques, can provide useful information.

Medicinal chemists and crystallographers begin the process of drug design after determining the 3-D atomic architecture of the target protein and its functionally critical regions. Using a variety of specialized programs on interactive graphics workstations, drug designers generate concepts of drug molecules that complement the unique structure and electronic environment of the target protein. Medicinal chemists then synthesize the most promising candidate structures. As in conventional drug-discovery

strategies, biochemists measure the ability of this newly synthesized drug candidate to produce the intended effects on the target protein. Crystallographers then redetermine the structure of the protein target now in combination with the candidate drug molecule included within the active site of the macromolecule. They see the detailed structural interactions actually achieved by the candidate drug with its biological counterpart. Scientists relate the performance of such a compound measured by conventional biochemical or pharmacological techniques to its structural interactions with the target as revealed by X-ray crystallography. The design team then incorporates the results of this analysis into its next generation of compounds. In this scenario, drug-design methodology consists of iterative cycles of simulation, design, synthesis, structure and biological performance assessments, and redesign. The power of this methodology lies in the ability of the drug designers to see the primary event in drug action, i.e., the interaction of the drug with its target as it actually takes place, and guide the design and optimization of drugs by the intimate details of this interaction.

DRUG DESIGN AND CHEMICAL DIVERSITY

In the 1990s, the average cost for introducing a new drug entity to the marketplace was estimated at greater than

\$300 million dollars. Of this dollar figure, nearly one-third has been estimated to go to the discovery and optimization of a lead chemical structure. Each compound within a company's archives ultimately finds its origin in the labors of several chemists involved in the synthesis or the isolation and identification of natural products. The cost for the preparation of such compounds on a single-compound basis is very important. It has been estimated that the cost of preparing each novel molecule in the traditional pharmaceutical industry paradigm of individually synthesized molecules prepared in serial fashion is between \$5,000 and \$10,000. Clearly, an opportunity exists to reduce costs at this earliest stage of the drug-discovery process, i.e., the identification of a novel lead chemical structure (Fig. 2).

Lead structure compounds are currently identified through rational design and/or mass screening. In the past, routine mass screening has been rather successful in identifying new leads. With the recent introduction of high-throughput, automated screening technologies the evaluation of hundreds of thousands of individual test molecules per year against a large number of targets is now feasible. The source of large chemical libraries still remains a stringent limitation. Compound libraries commonly used in mass screening consist of either an historical collection of synthesized compounds owned by pharmaceutical companies or natural product collections. Each of these libraries has limitations. Historical collections contain a limited number of diverse structures (e.g., thousands of steroids, β -lactams, benzodiazepines, etc.) and, although quite useful, represent only a small fraction of the vast number of possibilities. Natural products are limited by the structural complexity of the leads identified and the (pharmaco) chemical difficulty of modifying them to useful pharmaceutical agents (e.g., taxol) endowed with the pharmacokinetic and toxicological properties required for the medical goal pursued.

During the past decade, a new source of compounds has arisen, i.e., those obtained through the rapid chemical or biological generation of compound libraries. This wealth of new compounds, coupled with the ability to rapidly

carry out their biological evaluation, represents an important shift in the traditional paradigm for generating and optimizing new lead structures not only for the pharmaceutical industry, but also for the agricultural, materials, and chemical industries. From its earliest days over a decade ago to the present, the field of chemical generation of molecular diversity has changed dramatically. Early attempts focused exclusively on the rapid generation of very large numbers of peptides, using solid- and solution-phase techniques. This was primarily due to the ready availability of the natural and unnatural aminoacids and the previous development of well-established coupling methodologies. The application of this approach can now rapidly generate hundreds of thousands to millions of small to medium size peptides for identifying novel leads or to help elucidate the chemical basis of known ligand-receptor interactions by preparing and evaluating a large number of peptide analogs. The new trend, however, is to focus on the generation of nonpeptide, low-molecular-weight compounds (commonly called peptidomimetic compounds). Combinatorial chemistry is one of the important new methodologies developed to reduce the time and costs associated with producing chemical diversity. Scientists use combinatorial chemistry techniques to create large populations of molecules (libraries) that can be screened efficiently en masse. By producing larger, more diverse compound libraries, companies hope to increase the probability that they will find novel compounds of significant therapeutic and commercial value. The field represents a convergence of chemistry and biology, made possible by fundamental advances in miniaturization, robotics, and biotechnology developments (15–17).

As with traditional drug design, combinatorial chemistry relies on the progress of organic synthesis methodologies. The difference, however, is the scope: Instead of synthesizing a single compound, combinatorial chemistry exploits automation and miniaturization to synthesize large libraries of compounds (Fig. 3). Scientists then need a straightforward way to find the active ingredients within these enormous populations. Thus, combinatorial organic synthesis (COS) is not random, but systematic and repetitive, using sets of chemical “building blocks” to form a diverse set of molecular entities. Scientists have developed several different COS strategies, each with the same basic philosophy; i.e., to find ways to determine active compounds within populations, either spatially, through chemical encoding, or by systematic, successive synthesis and biological evaluation (deconvolution).

Three common approaches to COS could be envisaged. During arrayed, spatially addressable synthesis, building

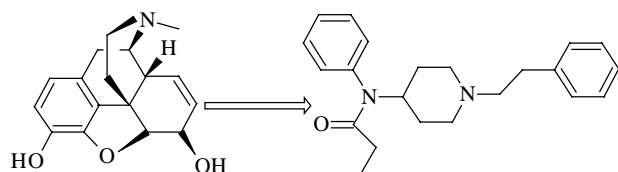


Fig. 2 It took more than 150 years to modify the complex morphine structure (left) to the simpler achiral morphinomimetic fentanyl, using classical drug-design methods.

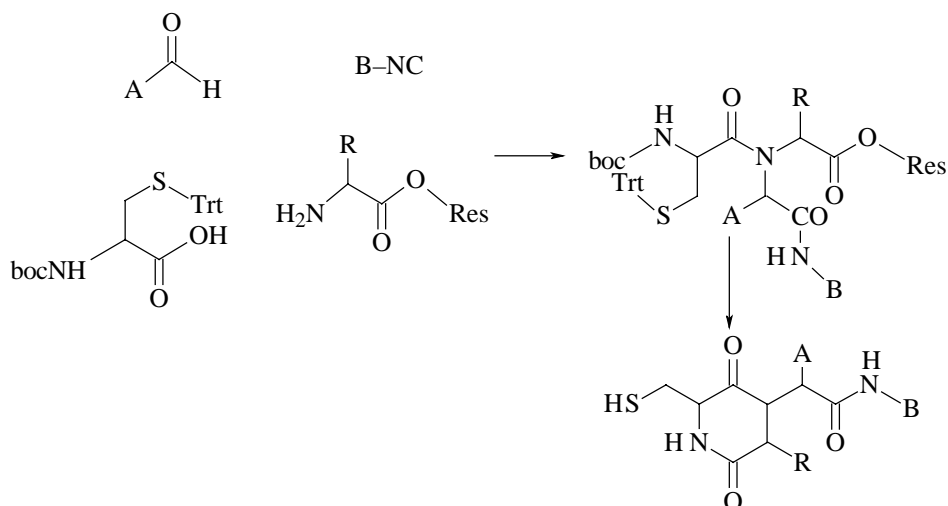


Fig. 3 Example of modular construction of a multifunctional assembly, using solid-phase synthesis technology (A and B are variable alkyl or aryl groups; boc = benzyloxycarbonyl; Trt = trityl; Res = resin).

blocks are reacted systematically in individual reaction wells or positions to form separated “discrete molecules.” Active compounds are identified by their location on the grid. The second technique, known as encoded mixture synthesis, uses nucleotide, peptide, or other types of more inert chemical tags to identify each compound. During deconvolution, the third approach, a series of compound mixtures is synthesized in a combinatorial manner, each time fixing some specific structural feature. Each mixture is assayed as a mixture and the most active combination is pursued. Further rounds systematically fix other structural features until a manageable number of discrete structures can be synthesized and screened. Scientists working with peptides, for example, can use deconvolution to optimize, or locate, the most active peptide sequence from millions of possibilities.

As with traditional drug design, the ability to integrate different types of chemical, biological, and corporate information is crucial to combinatorial chemistry techniques. Combinatorial chemistry also generates an enormous amount of information, which present-day information systems still have a hard time managing. Combinatorial chemists also ask different questions in different ways, and their information systems need to adapt to find these answers quickly. For example, chemists planning a traditional synthesis typically conduct a retrosynthetic analysis to determine the best, and perhaps cheapest, way to obtain the target. In the same way, combinatorial chemists also look at retrosynthetic trees to build combinatorial libraries. Combinatorial chemists need rapid ways to access reaction information efficiently. One of the largest limitations in the construction of

combinatorial libraries is in obtaining the basic building blocks necessary to run each reaction. Chemical information systems that can quickly retrieve commercially available reagents are invaluable tools in reagent acquisition.

Once built, combinatorial libraries produce unprecedented amounts of useful information, provided a consistent quality of the pharmacological evaluation can be ensured throughout the whole process. Reaction histories for each compound must be archived. Robots and other laboratory instruments need permanent monitoring, and the data they acquire have to be archived for future reference. Scientists need to integrate screening results and biological data with structural information. As in single-molecule archival systems, the archival of combinatorial libraries and their corresponding data is essential to cost-effective research and development.

Combinatorial chemistry is a promising new field that stands to revolutionize the chemical industry, and demands completely new scientific information management solutions. Combinatorial chemists will be able to meet their goals if they can find ways to plan libraries quickly, produce libraries that better interrogate biological assays, and learn from past screening results. Libraries have to be designed to systematically order and explore the wide-ranging molecular themes represented in its building block collection. Within each thematic library, subclass chemical properties are to be varied systematically in order to aid medicinal chemists in fine-tuning lead profiles. Using software that can orchestrate the planning, building, screening, and interpretation of synthesized libraries, combinatorial chemistry programs will begin to realize

their promise of minimizing the time and cost associated with bringing new molecular entities to market. Proper management of combinatorial chemistry libraries requires software applications that understand the science behind combinatorial chemistry while managing the chemical and biological data generated by combinatorial chemistry programs (18–22).

HIGH-THROUGHPUT ORGANIC SYNTHESIS (HTOS) AND HIGH-THROUGHPUT SCREENING (HTS)

The pharmaceutical world has passed through a remarkable transition in the past decade in its efforts to identify novel compounds that interact with new molecular targets and to pave the way to new therapeutic agents and product lines. The impact was first apparent in the biological sphere as assays reached the micro level and automation permitted sample throughput in multiple screens, which was unimaginable a few years earlier. HTS was born and quickly adopted as an important, if not the only, source of the initial “hits” required for the generation of leads for commercial development (23–26). The importance and permanence of the HTS focus has been manifested within the organizational structure of most corporations. Such structures tend to be immovable and only bend when economic considerations surface at both the operational and future levels. HTS is such an event. Discovery has shifted from a closed chemist–biologist relationship into a complex multidisciplinary cell that has generated a new level of professionalism within the industry. Compound management alone constitutes new positions and occasionally departments in larger organizations. This function controls the fodder that HTS requires and essentially holds the key to the success of the discovery program.

The success of HTS in uncovering the unknown parameters of a new receptor site, however, relies on access to a vast collection of compounds that possesses a very broad range of chemical functional groups dispersed in 3-D space in diverse structural frameworks. This level of diversity is rarely found in a company’s archives. The development of most corporate collections has relied on acquisition from external sources. Initially, such access was primarily from universities and research institutes worldwide. The fragmentary nature of these resources spawned the creation of small industries, the compound brokers, who have served as “gofers” and data processors in the collection and marketing of such compounds. It is economically preferable for industrial groups to use these

brokers and avoid the need and expense of scouring the multitude of academic archives. A significant savings over that of in-house synthesis is realized; however, the number required to come close to reasonable diversity brings the required investment again to a substantial level. It has been estimated that the average cost of \$50 per sample places the overall expenditure for 100,000 compounds at ~\$5,000,000 plus handling, storage, and maintenance costs. While this value can be reduced with rigorous negotiation when large numbers are purchased, the better bargain is to purchase compounds preplated on microplates. The micro requirements of most HTS assays (<10 μ g) makes this latter approach especially attractive to small companies wherein acquisition and management costs of a large compound collection are prohibitive.

All of these external resources have relied heavily on the university as a primary source. An unfortunate consequence of these activities over the past decade has been a depletion of the academic resources that have accumulated over the past 50 years. Furthermore, the availability of useful quantities from current academic research has become less likely as modern chemical techniques and instrumentation permit studies with only a few milligrams. This is especially apparent in the natural product field, as the unique diversity inherent in that group has placed such acquisition at a premium. Market demands, nevertheless, persist and have led to the emergence of synthesis factories that prepare compounds specifically for the screening market, using conventional chemical techniques. Output has been greatly enhanced at some of these centers by adoption of modular approaches in synthesis. These sources have helped to fill the void in numbers, but have done little to enhance the level of structural diversity or the cost factor.

Chemists have addressed the sample resource question directly. The first approach was in the use of solid- and solution-phase techniques in the synthesis of large libraries of peptides, through combinatorial chemistry. Many discovery-based industries implemented in-house programs in combinatorial chemistry initially to explore the peptide as a supplement to their screening resources, but more recently in the use of the same technologies for the automation of general organic synthesis, as the limits and flexibility of the peptide backbone failed to satisfy the more general structural needs of discovery (Fig. 4). The success of this approach has led to the generation of a new chemical discipline, HTOS, the implementation of such programs as core technologies in many industries, and the emergence, once more, of new industries. Programs in academia also address this need, the tremendous potential it offers in the enhancement of structural diversity and the intellectual challenge of devising techniques for the

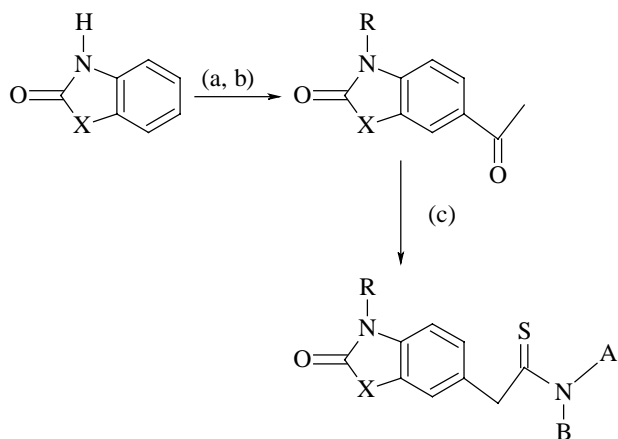


Fig. 4 Example of a three-step synthesis of a series of rather elaborated peptidomimetic compounds synthesized using straightforward reactions that can be easily automated in a fast parallel synthesis process. Methods: (a) acetic acid, polyphosphoric acid, 85°C; (b) RI, DMF, anhydrous potassium carbonate, room temperature; (c) Willgerdt-Kindler reaction (sulfur, secondary amine HNAB, DMF, 110°C).

tracking of the thousands of compounds generated as they proceed through a bioassay network. Several new industries have emerged within this specialized arena with chemical innovation and custom synthetic schemes that provide access to large chemical libraries which include features of specific interest to their client. The market and discovery impact has been dramatic.

Despite the excitement over the recent advances in HTOS development and the promises of prominent researchers in this area, HTOS cannot alone provide a diverse and sustained compound supply for screening. Today's technologies can achieve in parallel synthesis only the skeletal and functional complexity that is economically available through classical bench techniques. This aspect alone places full diversity outside the scope of HTOS because so far only a functional rather than a skeletal chemistry has been developed. Therefore, natural products still retain a commanding lead. While contributions to the enhancement of the structural diversity of a company's compound resources has, at best, been modest, the impact of HTOS on the economics of new drug development through rapid analog synthesis and the subsequent transposition of shorter development time into extended useful patent life has been dramatic. It is here that the strength of HTOS and its most important impact is felt.

The development of a lead into a product, in the traditional framework, is a stepwise process. The first stage is based on the assessment of synthetic possibilities within the limitations of organic synthesis and available starting materials, the extrapolation of preferred features defined

by precedence, any structure–activity relationships (SAR) information, and guides from computational analyses. This has not changed. The next step is the orderly planning of synthesis programs within the structural goals set for the target series, the development of timelines, which integrate synthesis with information feedback from primary and secondary bioassay, and further development of the synthetic targets. Analogs of an initial lead can be prepared singly at an average rate of 100 compounds per chemist per year depending on the complexity of the chemistry involved. In the traditional mode, the number of chemists employed on a project is defined by the management according to the significance of the discovery, the complexity of the syntheses, and the intended patent scope, the primary goal being a critical assessment of market potential and, if justified, the identification of preferred candidates for further development to a marketable product. The subsequent steps involve major capital investment in toxicology and clinical study. Thus, this selection stage is critical to corporate success and, in smaller companies, may mean eventual economic survival. A reasonable time for candidate selection in the traditional mode of drug discovery has been 1–2 years. The limiting factor has generally been the initial synthesis stage and the exploration of the numerous parameters associated with SAR development. With the application of HTOS design principles and robotics in organic synthesis, we see this time shortened dramatically. The 1- to 2-year period for candidate selection can be reduced by as much as 75%. This can constitute a significant extension in useful patent life by many months.

HTOS has clearly become a new chapter in chemical technology. It brings the excitement of further growth with the resolution of the obvious challenges of conducting chemical synthesis in arrays at semimicro and submicro levels. It also significantly impacts product development in the support industries as new goals are defined in liquid handling, automation, and data management. The corollary of this restructuring is the creation of new challenges for the organic chemist, a greater dependence on multidisciplinary interactions, and the internal structure of discovery departments. It has also created new scientific and engineering challenges and product opportunities that permeate support industries. Combinatorial chemistry has become a widely used tool both for the discovery and the optimization of lead structures. The demand for quality combinatorial libraries—characterized by diversity, novelty, purity, medicinal relevance, and facile synthesis—is clearly increasing. However, most of today's libraries provide only limited novelty, owing to the use of widely available commercial reagents and standard chemistries. In a near future, to achieve this goal of

chemical diversity it will be necessary to produce truly novel high-quality, value-added libraries which comprise a wide array of pharmacophore building blocks and scaffolds, featuring a broad range of molecular properties and chemical functionality.

DNA, FUNCTIONAL GENOMICS AND PHARMACO- AND TOXICOGENOMICS

In the past few years, we have witnessed a dramatic increase in the availability of genome-scale DNA sequence information from humans and several model organisms. New technologies incorporating this information are radically altering biological research. Toxicogenomics is a new scientific subdiscipline that combines the emerging technologies of genomics and bioinformatics to identify and characterize mechanisms of action of known and suspected toxicants. Currently, the premier toxico-genomic tools are the DNA microarray and the DNA chip, which are used for the simultaneous monitoring of expression levels of hundreds to thousands of genes (27–30).

An array is an orderly arrangement of samples. It provides a mean for matching known and unknown DNA samples based on base-pairing rules and automating the process of identifying the unknowns. An array experiment can make use of common assay systems such as microplates or standard blotting membranes, and can be created by hand or make use of robotics to deposit the sample. In general, arrays are described as macro-arrays or micro-arrays, the difference being the size of the sample spots. Macro-arrays contain sample spot sizes of about 300 μm or larger and can be easily imaged by existing gel and blot scanners. The sample spot sizes in micro-arrays are typically less than 200 μm in diameter and these arrays usually contains thousands of spots. Micro-arrays require specialized robotics and imaging equipment which generally are not commercially available as a complete system.

DNA micro-arrays, or DNA (gene) chips, are fabricated by high-speed robotics on glass or nylon substrates, for which probes with known identity are used to determine complementary binding, allowing massively parallel gene expression and gene discovery studies. An experiment with a single DNA chip can provide researchers with information on thousands of genes simultaneously—a dramatic increase in throughput. Why do some drugs work better in some patients than in others? And why may some drugs even be highly toxic to certain patients? Pharmacogenomics can be regarded as the hybridization of functional genomics and molecular pharmacology. The

goal of pharmacogenomics is to find correlations between therapeutic responses to drugs and the genetic profiles of patients. In the same way, toxicogenomics is the hybridization of functional genomics and molecular toxicology.

Pharmacogenomics uses genetic and genomic information to predict the response of individual patients and patient populations to drugs. Pharmacogenomics will have an impact on medicine by allowing the use of newly created genomic diagnostic tools to predict which drugs will have the greatest chance of success in treating individual patients safely and effectively. Eventually, drug prescriptions may be tailored or customized to the individual patient's genetic make-up, using this technology. Pharmacogenomics also has the potential to revolutionize the planning and design of clinical trials. By identifying which patients will respond to compounds being tested, as well as by eliminating those who may be at risk for adverse reactions, pharmacogenomics can improve the success rate of clinical development and reduce development time and cost.

The development of quantitative structure–toxicity relationships (an extended form of QSAR) to predict and to help understand the toxicity and metabolism of drugs now becomes an additional challenge for the efficient discovery of new drug candidates. Indeed, while HOTS and HTS techniques become more and more efficient to identify new lead structures and convey adequate information about the description of new pharmacophores, as such these techniques do not give direct access to compounds devoid of toxicity and endowed with the expected pharmacokinetic properties required for the therapeutic goal pursued. At this stage of the research, the professional skills and talents of experienced “traditional” medicinal chemists are again of paramount importance for the success of the whole process. A criticism that can be addressed to HOTS and HTS techniques so far developed is that, to increase potency and selectivity of the future drug candidate, relatively high molecular complexity is employed. These techniques of selection tend to unnecessarily increase molecular weight and lipophilicity and, as a result, the oral bioavailability (a major pharmacokinetic parameter) of such compounds will be inevitably very low or virtually zero. Consequently, after the discovery of initial leads, complementary programs of synthesis by traditional means are still necessary to produce grams of material necessary to assess the toxicological and pharmacokinetic properties. In this connection, while remaining somewhat empirical, the fundamental knowledge as well as the practice do exist. For this purpose, prodrugs have been intensively studied. However, this approach again tends to increase the

molecular complexity and, therefore, the molecular weight. It is, therefore, more and more envisaged to develop approaches with an integrated view of this enormous problem involving, at a very early stage, considerations of structure–toxicity and structure–metabolism relationships with the help of computerized expert systems.

The rapid evolution of the field with many techniques and disciplines involved poses the problem of the basic education and training of the personnel. While the relationship between academia and industry is frequently examined from the perspective of research and technical collaborations, it is also important to view it in terms of supply of human resources by academia to meet the demands of industry. At this level, the educational system based on the Ph.D. programs offered by universities has taken some delay with respect to the culture of the companies. A significant effort will have to be made to fill an undoubted gap in current professional education. There is also an increasing demand from the industry for basic research in organic chemistry to develop new reactions that can be employed in HOTS.

CONCLUSION

In front of the ever-increasing demand by the society and the authority for better drugs at lower cost, the pharmaceutical industry has developed sophisticated techniques of drug design, encompassing HOTS and HTS techniques to speed up the duration and the cost of drug discovery. While these techniques have not reached the level of superb maturity, they have considerably influenced the culture of the companies. However, per se these new methodologies do not directly deliver the new drug candidate, and, therefore, complementary programs of traditional medicinal chemistry are still necessary for the fine tune-up of the drug candidate. Consequently, drug design at present is made of a blend of modern technology and more classical science. The making of a blend has always been an art. It can be anticipated that in the future century drug design will evolve between these two trends, a blend of science and art in which the human being is central.

REFERENCES

1. Parrill, A.L., Rami, Reddy, M., Eds. *Rational Drug Design. Novel Methodology and Practical Applications*. Oxford University Press: New York, 1999.

2. Bugg, C.E.; Carson, W.M.; Montgomery, J.A. *Drugs by Design*. *Sci. Am.* **1993**, 269, 92–98.
3. Truhlar, D.G., Howe, W.J., Hopfinger, A.J., Blaney, J., Eds. *Rational Drug Design*, Springer: New York, 1999.
4. Wupong, S., Ed. *Biopharmaceutical Drug Design and Development*; Humana Press, Inc.: Totowa, New York, USA, 1999.
5. Salemme, F.R.; Spurlino, J.; Bone, R. Serendipity Meets Precision: The Integration of Structure-Based Drug Design and Combinatorial Chemistry for Efficient Drug Discovery. *Structure* London, **1997**, 5 (3), 319–324.
6. Bone, R.; Salemme, F.R. The Integration of Structure-Based Design and Directed Combinatorial Chemistry for New Pharmaceutical Discovery. *Structure-Based Drug Design*; Veerapandian, P., Ed.; Marcel Dekker, Inc.: New York, 1997; 525–539.
7. Babu, Y.S.; Ealick, S.E.; Guida, W.C.; Erion, M.D.; Montgomery, J.A.; Secrist, J.A.; Carson, M.; Narayana, S.V.L.; Cook, W.J.; Bugg, C.E. (1997) Design of Purine Nucleoside Phosphorylase Inhibitors. *Structure Based Drug Design*; Veerapandian, P., Ed.; Marcel Dekker, Inc.: New York, 1997; 151–169, Ch. 5.
8. Graybill, T.L.; Agrafiotis, D.K.; Bone, R.; Illig, C.R.; Jaeger, E.P.; Locke, K.T.; Lu, T.; Salvino, J.M.; Soll, R.M.; et al. Enhancing the Drug Discovery Process by Integration of High-Throughput Chemistry and Structure-Based Drug Design. *Mol. Diversity Comb. Chem.: Libr. Drug Discovery Conf.* **1996**, 16–27.
9. Illig, C.; Eisennagel, S.; Bone, R.; Radzicka, A.; Murphy, L.; Randle, T.; Spurlino, J.; Jaeger, E.; Salemme, F.R.; Soll, R.M. Expanding the Envelope of Structure-Based Drug Design Using Chemical Libraries: Application to Small-Molecule Inhibitors of Thrombin. *Med. Chem. Res.* **1998**, 8 (4/5), 244–260.
10. Spurlino, J.C. Structural Implications in the Design of Matrix-Metalloproteinase Inhibitors. *Structure-Based Drug Design*; Veerapandian, P., Ed.; Marcel Dekker, Inc.: New York, 1997; 171–189.
11. Bossart-Whitaker, P.; Carson, M.; Babu, Y.S.; Smith, C.D.; Laver, W.G.; Air, G.M. Three-Dimensional Structure of Influenza A N9 Neuraminidase and its Complex with the Inhibitor 2-Deoxy 2,3-Dehydro-N-Acetyl Neuraminic Acid. *J. Mol. Biol.* **1993**, 232, 1069–1083.
12. Carson, M.; Ribbons Macromolecular Crystallography. *Methods in Enzymology*; Sweet, R.M., Carter, C.W., Eds.; Academic Press: New York, 1997; 277, 493–505.
13. Bone, R.; Lu, T.; Illig, C.R.; Soll, R.M.; Spurlino, J.C. Structural Analysis of Thrombin Complexed with Potent Inhibitors Incorporating a Phenyl Group as a Peptide Mimetic and Aminopyridines as Guanidine Substitutes. *J. Med. Chem.* **1998**, 41 (12), 2068–2075.
14. Lu, T.; Tomczuk, B.; Illig, C.R.; Bone, R.; Murphy, L.; Spurlino, J.; Salemme, F.R.; Soll, R.M. In Vitro Evaluation and Crystallographic Analysis of a New Class of Selective, Non-Amide-Based Thrombin Inhibitors. *Bioorg. Med. Chem. Lett.* **1998**, 8 (13), 1595–1600.
15. Houghten, R.A.; Pinilla, C.; Appel, J.R.; Blondelle, S.E.; Dooley, C.T.; Eichler, J.; Nefzi, A.; Ostresh, J.M. Mixture-Based Synthetic Combinatorial Libraries. *J. Med. Chem.* **1999**, 42, 3743–3778.
16. Sofia, M.J.; Allanson, N.; Hatzenbuehler, N.T.; Rakesh, J.; Kakarla, R.; Kogan, N.; Liang, R.; Liu, D.; Silva, D.J.;

- Wang, H.; Gange, D.; Anderson, J.; Chen, A.; Chi, F.; Dulina, R.; Huang, B.; Kamau, M.; Wang, C.; Baizman, E.; Branstrom, A.; Bristol, N.; Goldman, R.; Kiho, H.; Longley, C.; Midha, S.; Axelrod, H.R. Discovery of Novel Disaccharide Antibacterial Agents Using a Combinatorial Library Approach. *J. Med. Chem.* **1999**, *42*, 3193–3198.
17. Dolle, R.E.; Nelson, K.H., Jr. Comprehensive Survey of Combinatorial Library Synthesis. *J. Comb. Chem.* **1999**, *1*, 235–282.
18. Agrafiotis, D.K.; Yslik, J.C.; Alemme, F.R. Advances in Diversity Profiling and Combinatorial Series Design. *Mol. Diversity* **1999**, *4* (1), 1–22.
19. Agrafiotis, D.K. On the Use of Information Theory for Assessing Molecular Diversity. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (3), 576–580.
20. Agrafiotis, D.K. Stochastic Algorithms for Maximizing Molecular Diversity. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (5), 841–851.
21. Dhanoa, D.S.; Soll, R.M.; Subasinghe, N.; Wu, Z.; Rinker, J.; Hoffman, J.; Eisennagel, S.; Graybill, T.; Bone, R.; Radzicka, A.; Murphy, L.; Salemme, F.R. Serine Proteases-Directed Small Molecule Probe Libraries. *Med. Chem. Res.* **1998**, *8* (4/5), 187–205.
22. Hajduk, P.J.; Gerfin, T.; Boehlen, J.-M.; Häberli, M.; Fesik, S.W. High-Throughput Nuclear Magnetic Resonance-Based Screening. *J. Med. Chem.* **1999**, *42*, 2315–2317.
23. Andres, C.J.; Denhart, D.J.; Deshpande, M.S.; Gillman, K.W. Recent Advances in the Solid Phase Synthesis of Drug-Like Heterocyclic Small Molecules. *Combinatorial Chem. & High Throughput Screening* **1999**, *2*, 191–210.
24. Kuzniar, E. Inventory Management and Reagent Supply for Automated Chemistry. *Combinatorial Chem. & High Throughput Screening* **1999**, *2*, 239–247.
25. Glaser, V. Trends in Drug Discovery. *Genetic Eng. News* **1998**, *18*, 12.
26. Brush, M. High-Throughput Technology Picks Up Steam. *The Scientist* **1999**, *13*, 11.
27. Cho, R.J.; Campbell, M.J.; Winzeler, E.A.; Steinmetz, L.; Conway, W.; Wodicka, L.; Wolfsberg, T.G.; Gabrielian, A.E.; Landsman, D.; Lockhart, D.J.; Davis, R.W. A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle. *Mol. Cell* **1998**, *2*, 65–73.
28. Wolfsberg, T.; Gabrielian, A.; Campbell, M.J.; Cho, R.; Spouge, J.; Landsman, D. Candidate Regulatory Sequence Elements for Cell Cycle-Dependent Transcription in *Saccharomyces Cerevisiae*. *Genome Research* **1999**, *9*, 775–792.
29. Tavazoie, S.; Hughes, J.D.; Campbell, M.J.; Cho, R.J.; Church, G.M. Systematic Determination of Genetic Network Architecture. *Nat. Genet.* **1999**, *22*, 281–285.
30. Madej, T.; Boguski, M.S.; Bryant, S.H. Threading Analysis Suggests that the Obese Gene Product may be a Helical Cytokine. *FEBS Lett.* **1995**, *373*, 13–18.

DENTAL PRODUCTS

Sebastian G. Ciancio

State University of New York at Buffalo, Buffalo, New York, U.S.A.

INTRODUCTION

A considerable number of products are now recommended for use in the oral cavity. Products for caries control are widely used and include fluorides in dentifrices and mouthwashes. Plaque control is achieved through the use of chemical agents such as chlorhexidine and quaternary ammonium compounds. Mechanical products for plaque control include dental floss, toothpastes, and mouthwashes. Products also exist to combat halitosis, act as topical anesthetics, desensitize sensitive teeth, act as tooth-bleaching agents, and assist in reducing xerostomia. More recently, products designed for the local delivery of antimicrobial agents to the oral cavity have also been made available. These, and other products, are reviewed in this article.

CARIES CONTROL USING FLUORIDES

Over 300 million people worldwide now consume optimally fluoridated water. The U.S. Public Health Service has established recommended levels for fluoride concentrations in water supplies in accordance with mean annual temperatures (1). The daily intake of fluoride not only comes from drinking water but also from food consumed or prepared with fluoridated water. Also, crops are frequently fertilized with phosphate fertilizers of high soluble-fluoride content, and food products including bone in animal feeds contain fluoride.

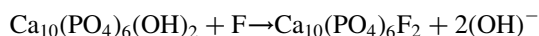
Naturally fluoridated foods and water have been ingested for decades with no serious side effects. In addition, public fluoridation has been widespread in this country for over 30 years without serious adverse effects. The incidence of mottled enamel, one of the earliest and most sensitive signs of fluoride toxicity, has not increased significantly in the past 15 years of water fluoridation. The safety and efficacy of fluoride has definitely been established, with no scientific evidence against fluoridation.

Mechanism of Action

The mechanism by which fluoride prevents caries is not clearly understood. It is known that the fluoride ion (F^-)

can replace the hydroxyl ion (OH^-) in hydroxyapatite, the major crystalline structure of enamel. The substituted crystal, called fluorapatite, is more resistant to acids, such as those produced by plaque bacteria, than the original hydroxyapatite.

As the tooth develops and enamel is formed, ingested fluoride is incorporated into the enamel. Therefore, because enamel develops its outer layer first, more fluoride can be expected to be deposited on the outer layers as compared to the inner layers. It is this surface enamel layer containing fluoride that imports, in part, caries resistance to a tooth (2). Topical fluorides also become incorporated into enamel and provide protection against acid. A number of studies have now shown that topical fluorides may be most beneficial in early enamel caries and that there is an increased uptake of fluoride in early lesions, with some tooth remineralization occurring. This finding has caused some investigators to label fluoride as a remineralizing agent as well as a caries inhibitor. The incorporation of fluoride into enamel can be represented as a chemical reaction:



Dental plaque also tends to concentrate fluoride. This could increase possible antienzymatic activity. Some caries protection from this may be expected. Additionally, studies have suggested that topical application of fluoride may also reduce smooth surface plaque, with a resulting beneficial effect on the periodontal tissues (3).

In areas where there is no fluoridation of the community water supply, fluoride may be added to school water. This is not a substitute for community water fluoridation because fluoride intake from birth is important.

In some areas where fluoride is absent from drinking water, the school water supply has been fluoridated as much as 4.5 times the level usually recommended for community water levels. These studies, conducted for 12 years in fluoride-deficient areas, revealed a 40% reduction in DMF surfaces. Essentially, no undesirable fluorosis resulted from this procedure. However, it should be noted that ingestion occurred only during part of each day and only on school days. This level (4.5 ppm) from birth (during major formation of permanent teeth) would cause undesirable fluorosis with continued intake (4).

Therapeutic Effects

Fluoridated water

The administration of fluoride in drinking water at concentrations of approximately 1 ppm significantly reduces dental caries. The anticaries benefits are similar to those due to natural fluoride in drinking water. Fluoridated drinking water produces the following: a 60% lower dental caries rate, a 75% decrease in the loss of 6-year molars, and a 90% reduction in the incidence of proximal caries of the four upper anterior teeth.

Evidence suggests that greater inhibition of caries occurs when teeth receive fluoride throughout the calcification period. Therefore, maximum benefit may be expected from the continued use of fluoridated drinking water.

Dietary supplements/tablets

Maximum benefits to both deciduous and permanent teeth may result from daily fluoride supplements from infancy until approximately 13 years of age, at which time all permanent teeth except the third molars should have erupted. Because cariostatic benefits may tend to diminish gradually after fluorides are discontinued, periodic applications of topical fluorides may then be necessary.

The use of dietary fluoride or topical applications of fluoride depends partly on the age of the child. Dietary supplements of fluoride are best for very young children, whereas topical fluoride applications are preferred for older children with permanent teeth. Younger children who are highly susceptible to caries may benefit from both measures.

The natural level of fluoride in drinking water where the child lives should be known before dietary fluoride is prescribed. At present, it is suggested that fluoride supplements be limited to where drinking water contains 60% or less of the optimal level of fluorides recommended for community water in the geographic area.

As a precaution, no large quantities of sodium fluoride should be stored in the home. It is recommended that no more than 264 mg of sodium fluoride be dispensed at any one time, which is enough for at least a 4-month period (5). Each package dispensed should also bear the statement: Caution—store out of the reach of children.

Although the optimal level of fluoride in drinking water is well documented, there is no established allowance for fluoride administered once a day. The standard allowance is 1 mg/day for a child over 3 years of age and one-half this amount for a child between the ages of 2 and 3. A more accurate method to use in calculating the daily fluoride needs of a child is to administer a dose based on the child's weight, for example, .025 mg/lb of body weight. However,

Table 1 Adjustment of prescribed fluoride relative to natural content of drinking water for children over 6 years of age

Water fluoride (ppm ^a)	Adjusted allowance sodium fluoride (mg/day)	Provides fluoride ion (mg/day)
0.0	2.2	1.0
0.2	1.8	0.8
0.4	1.3	0.5
0.6	0.0	0.0

^a1 ppm = 1 mg/L.

for all children over 6 months of age, the dose should not exceed 1 mg/day of fluoride ion regardless of weight. In order to avoid the possibility of unesthetic dental fluorosis, the prescribed dietary allowance should be reduced in proportion to the fluoride levels in the drinking water. Table 1 illustrates empirical adjustments for varying amounts of natural fluorides in communities where the recommended optimal level of fluoride is 1 ppm. These allowances are for children over 6 years of age. The allowances should be reduced for children 6 years of age and under, as shown in Table 2 (6, 7).

For children under 6 months of age, experts question the value of prescribing fluoride supplements.

Dental caries have been prevented when fluoride tablets were administered in a school-based program. After two or more years of fluoride ingestion, protection against dental caries ranged from 20–40% (8). In an extended trial of fluoride tablets reported in the literature, there was a 36% reduction in dental caries after 8 years (9).

The use of fluoride tablets can provide both a preeruptive (endogenous) effect and a posteruptive (topical) effect. Therefore, tablets should be chewed or dissolved in the mouth and the teeth rinsed with the resultant solution before swallowing.

One advantage of fluoride tablets compared to water fluoridation is that a specific dosage of fluoride is delivered.

Table 2 Adjustment of prescribed fluoride ion relative to natural content of drinking water for children under 6 yr of age

Patient's age	Content of water fluoride ^a		
	0–0.3 ppm	0.3–0.6 ppm	>0.6 ppm
0–6 mon	0.0	0.0	0.0
6 mon–3 yr	0.25	0.0	0.0
3–6 yr	0.50	0.25	0.0

^a2.2 mg prescribed sodium fluoride = 1 mg fluoride ion.

One disadvantage is that dietary supplements of fluoride taken once a day are rapidly cleared from the body.

Breast-feeding

Fluoride levels in human breast milk have been found to be less than 0.05 ppm. This concentration remains constant, regardless of drinking water and maternal plasma levels. Thus, breast-fed infants who receive no formula bottle feedings ingest considerably less fluoride than infants receiving formula mixed in 1 ppm fluoridated water. Fluoride supplementation for breast-fed infants should be considered.

The administration of fluoride supplements to expectant mothers in an effort to benefit the teeth of the offspring has been evaluated in several studies (10, 11). The evidence, however, is not sufficiently conclusive to warrant recommendation.

Vitamins

Vitamin preparations containing sodium fluoride are also available as drops, tablets, and chewable tablets. These forms of supplementation are useful in areas where the water supply contains less than 0.6 ppm, and they offer a way to provide fluoride to the child, if the parents are conscientious in dispensing the required amount daily and if the child does not object to taking oral medications. In recommending vitamins with fluoride, it is mandatory that one know the fluoride content of the child's water supply, as well as the fluoride content of the vitamin being recommended.

Topicals

Fluoride can be applied topically in various forms, offering the practitioner a number of options to choose for his or her patients.

Topical agents are of lesser value in a caries reduction program when they are used in fluoridated communities. However, when used in nonfluoridated areas, they are more effective (12, 13). In such areas, they often are the only form of fluoride therapy available.

The dosage forms of topical fluoride currently available include varnishes, dentifrices, solutions, gels, mouthwashes, and prophylaxis pastes. These various forms are outlined in Table 3.

Dentifrices

The earliest fluoride dentifrices contained sodium fluoride. However, the fluoride was biologically unavailable because the calcium in the dentifrice abrasive bound the fluoride and thus inactivated it.

Although a number of dentifrices containing fluoride are on the market, not all provide available fluoride because the abrasive systems that some dentifrices contain inactivate the fluoride. Therefore, the product may contain as much fluoride as any other dentifrice but it is not available. Also, if the product has a short shelf life, it will be ineffective if poor marketing gets it to the consumer too late.

For these reasons, only dentifrices approved by the Council on Scientific Affairs of the American Dental Association (ADA) should be recommended. These

Table 3 Various topical fluoride preparations

Preparation	Form	Formulation
Acidulated phosphate fluoride	Topical solution	1.23% in 1% phosphoric acid
	Topical gel, foam	1.23% in 1% phosphoric acid
	Mouthrinse	0.02–0.04%
	Prophylaxis paste	1.2%
Amine fluoride	Dentifrice	1.6%
	Mouthrinse	2.5%
	Topical solution	2%
Sodium fluoride	Mouthrinse	2.5%
	Foam	0.2%
	Varnish	5% (every 3–6 months)
	Dentifrice	0.76–0.8%
Sodium monofluorophosphate	Topical solution	8%
Stannous fluoride	Mouthrinse	0.1%
	Prophylaxis paste	8%
	Dentifrice	0.4%
	Gel	0.4%

Table 4 Representative dentifrice ingredients

Brand name	Fluoride	Abrasive	Sweetener
Aim	0.8% Sodium MFP	10% hydrated silica xerogel 19% hydrated silica	67% Sorbitol
Aim extra strength	1.2% Sodium MFP		
Aquafresh	0.76% Sodium MFP	12.6% Calcium carbonate 12% Silica	52.8%
Colgate®	0.76% Sodium MFP	48.76 Dicalcium phosphate	22% Glycerin
Crest®	0.24% Sodium fluoride	20% Hydrated silica	50% Sorbitol 18% Glycerin
Macleans	0.76% Sodium MFP	38% Calcium carbonate	29.55% Glycerin

products are listed in that association's publications and carry the ADA seal on their packaging.

Currently accepted dentifrices contain sodium monofluorophosphate, sodium fluoride, or, less frequently, stannous fluoride, all of which reduce caries by approximately 25% when used daily. In some clinical studies, stannous fluoride dentifrices stained teeth, particularly in pits and fissures (14). This stain is related to the tin in this compound, which adheres to plaque. The significance of this staining and its esthetic problems have resulted in a decreased usage in dentifrices. Stannous fluoride dentifrices are marketed in a plastic container because a reaction of stannous ions at an acid pH occurs when conventional soft metal tubes are used.

The composition of some popular toothpastes is important for a proper understanding of this topic. With the exception of extra strength products, the various dentifrices are formulated to provide 1000 ppm of fluoride. Comparative compositions of some fluoride dentifrices are shown in Table 4.

Since children ingest most of the fluoride toothpaste when they brush their teeth, only a pea-sized amount should be placed on the brush.

Stannous fluoride

Dentifrices containing stannous fluoride as an active ingredient are no longer widely marketed; however, these formulations were the first to be evaluated for caries-reducing properties. Effectiveness in caries reduction varied from 23 to 34% (15, 16). One stannous fluoride dentifrice containing a patented stabilized form of stannous fluoride is marketed with a claim of both caries and gingivitis reduction. However, this product is not ADA accepted.

Currently, there are no ADA-approved, over-the-counter dentifrices containing stannous fluoride. However, there are a number of ADA accepted stannous fluoride

prescription products approved for application by the dentist or by the patient.

Amine fluoride

Clinical data from several long-term studies in Europe have demonstrated the effectiveness of the use of a dentifrice containing organic amine fluorides (17, 18). The amine fluorides also have strong plaque-reducing properties. However, although the amine fluorides may be more effective for caries reduction than other forms of fluoride, the FDA has not allowed these products to be extensively tested in this country.

Sodium fluoride

Sodium fluoride as an ingredient in dentifrices has been the subject of a number of clinical investigations. Recent studies of sodium fluoride dentifrices formulated to ensure ready availability of fluoride ions have shown anticaries benefits similar to those obtained in clinical caries trials with dentifrices containing stannous fluoride and sodium monofluorophosphate.

Clinical caries trials conducted under well-controlled, daily supervised brushing conditions have reported reductions in dental caries of approximately 25–48% (19, 20).

Sodium monofluorophosphate

A number of clinical studies have been conducted with dentifrices containing 0.76% monofluorophosphate (MFP). The data from these controlled clinical studies of sodium MFP dentifrices have indicated reductions in dental caries ranging from approximately 17–42% (21–23). Two studies indicating effectiveness were conducted in fluoridated communities. In clinical studies comparing this form of fluoride with sodium fluoride, the findings for caries reduction have been similar. Unlike dentifrices containing sodium fluoride, dentifrices

containing MFP are compatible with a number of abrasive systems; this is one of the reasons why there are more ADA-accepted products in this category.

Solutions and Gels

In children, a reduction of 30–40% in dental caries is seen with the following: 2% sodium fluoride, 8% stannous fluoride, and acidulated phosphate-fluoride products. No one agent appears to be superior to any other when used as directed.

Solutions of 8% stannous fluoride have been used to reduce caries (24). As with the other agents, the teeth are polished, dried, and isolated, and a 4-min application follows. The disadvantages of this solution are that it must be freshly prepared, some tooth discoloration (as discussed under dentifrices) has occurred, and it has an unpleasant taste that is difficult to mask.

Topical concentrated fluoride solutions are useful in children with high caries activity because they may have both a caries-arresting property and one of caries prevention. The frequency of application varies with the caries activity of the child. For children with an average incidence of caries, it can be applied annually between the ages of 3 and 13.

When gels are used, they are placed into a tray that is placed against the teeth so that the gel flows around all surfaces. Best results have been reported with custom-fitted trays. It has been estimated that about one-third of the total fluoride placed in a tray is actually available for uptake by teeth. The remainder is simply a filler for the tray, some of which is swallowed (25).

The value of topical fluorides on adults has not been established. However, some studies have suggested that the acidulated phosphate fluoride types may offer some caries protection. Recent reports have suggested that polishing of teeth is not necessary prior to the topical application of fluorides (26, 27). These studies have stated that deplaquing of teeth with a toothbrush is adequate and offers the advantage of not removing surface fluoride from tooth structure. This concept may be valid, and future investigations in this area should be encouraged.

Fluoride mouthwashes

Substantial research has been performed on the caries-inhibiting effect of fluoride mouthrinse products. The effectiveness of these topically applied fluorides varies with patient compliance.

The daily use of fluoride rinses by young children should be carefully monitored, and it should be noted that

the ADA does not recommend the use of fluoride mouthrinse for children under 6 years of age.

Moreover, because these products, when brought into the home, present a potential danger, the ADA Council on Scientific Affairs has recommended that these rinses should not exceed 300 mg of sodium fluoride. Fluoride mouthrinse solutions for use in school or community programs, however, are available in larger volumes, based on the assumption that storing and dispensing of the products in public settings will be closely monitored. The potential dangers involved with unsupervised ingestion of these products should be made known to both parents and children.

Studies evaluating the effectiveness of mouthrinses containing sodium fluoride have shown the usefulness of these agents for children living in nonfluoridated areas. Most studies have been conducted using a mouthrinse containing approximately 0.05% sodium fluoride used daily or a 0.2% sodium fluoride used weekly (28, 29).

The Council on Scientific Affairs has accepted a number of fluoride mouthrinses. All products discussed below have ADA Council on Scientific Affairs acceptance.

Some mouthrinses are marketed as a concentrate for dilution to recommended levels. If concentrates are used, special care should be exercised to keep them out of the reach of children. The products are not packaged in glass containers, because the pH becomes more alkaline in glass. Plastic is the container of choice. Examples of prescription and nonprescription products follow.

ACT

This product is an aqueous solution of 0.05% sodium fluoride. It also contains 8% glycerin, 7% alcohol, a detergent, a preservative, saccharin, coloring, and flavoring agents. It is intended to be used on a daily basis and is available without prescription.

Fluorigard

This product is available as a rinse containing an aqueous solution of a 0.05% sodium fluoride, 15% glycerin, 5% alcohol, a detergent, a preservative, saccharin, coloring, and flavoring agents. It is intended to be used on a daily basis and is available without a prescription.

Fluorinse

This product contains 0.2% sodium fluoride as the active ingredient. It also contains a detergent, a preservative, flavoring, and color agents. It is intended to be used on a daily basis and is available without a prescription.

Phos-Flur oral rinse supplement

This product is an aqueous solution containing 0.044% sodium fluoride, 0.055% phosphoric acid, 1.35% sodium biphosphate, and flavoring and coloring agents. It is intended to be used once daily and is available by prescription.

Desensitizing Agents

A number of studies have reported that topical fluoride application may reduce dental hypersensitivity (30–32). These results have been found when concentrated dosage forms have been applied ranging from 8% stannous fluoride gels to 33.3% sodium fluoride paste. It has been shown that commercial dentifrices containing stannous fluoride may also decrease dental hypersensitivity. Also, a combination of stannous fluoride and potassium nitrate is marketed by one manufacturer to reduce sensitivity. Varnishes containing sodium fluoride have also been shown to reduce dental hypersensitivity.

CHEMICAL AGENTS FOR PLAQUE CONTROL

A number of chemical agents have been evaluated over the years in terms of their antimicrobial effects in the oral cavity and the importance of these effects on oral health.

In 1986, the establishment by the ADA of guidelines for acceptance of these products has served to stimulate properly designed clinical studies for evaluating potential therapeutic agents. Products that have earned the ADA's seal of acceptance are Peridex, Listerine[®], some generic copies of Listerine, and Colgate Total[®].

In this section, data on various available agents are presented according to chemical agent category. When the term *substantivity* is used, it refers to the ability of an agent to be retained in the area cavity and to be released over an extended time period with a continued antimicrobial effect.

Chlorhexidine

Of the products included in this report, chlorhexidine appears to be the most effective agent. Long-term studies in over 700 subjects showed reductions in plaque averaging 55% and in gingivitis 45% (33–35).

The mechanism of action of chlorhexidine is related to a reduction in pellicle formation, alteration of bacterial absorption and/or attachment to teeth, and an alteration of the bacterial cell wall so that lysis occurs. Chemically, it is classified as chlorhexidine digluconate and the U.S.

Adopted Name (USAN) designation is chlorhexidine gluconate. It has high substantivity. Adverse effects reported included staining of teeth, reversible desquamation in young children, alteration of taste, and an increase in supragingival calcified deposits. Long-term and microbiologic studies do not demonstrate the development of resistant strains. It is sold in the United States in a 0.12% concentration as a prescription mouthrinse (Peridex, PerioGard, and generically), which contain 11.6% alcohol with a pH of 5.5 and is approved by the ADA for control of plaque and gingivitis. Recommended usage is twice daily.

Fluorides

Fluorides are purported to have some antiplaque properties. The most widely used topical fluorides are stannous fluoride, acidulated phosphate fluoride, and sodium fluoride. Of the fluorides, short-term studies of stannous fluoride have been promising. However, long-term published studies showed lower plaque scores, but the differences were not significant (36, 37). No effect on gingival health was noted with the exception of one study.

With stannous fluoride, the mechanism of action appears to be related to an alteration of bacterial aggregation and metabolism. In summarizing the properties of this agent, it can be stated that it has moderate substantivity, that the antibacterial activity may be related to the tin ion, and that a 0.4% concentration may be the most effective. Stannous fluoride is the most toxic of the products considered and has the shortest shelf life. Adverse effects have been taste and black stain lines on teeth. Usage of once or twice daily favors compliance. Stannous fluoride is most often available as an aqueous gel.

Stannous fluoride products are accepted by the ADA for their ability to deliver fluoride but have not been approved for their plaque-reducing properties. Examples of accepted products are Activux Basic Control, Gel-Kam, Gel-Tin, Perfect Choice, Pro-Dentx, Schein Home Care, and Super-Dent.

Oxygenating Agents

In evaluating the efficacy of oxygenating agents, one must evaluate the endpoints selected for efficacy and their measurement. Oxygenating agents have anti-inflammatory properties. Therefore, less bleeding on probing, a major sign of inflammation, would be expected following their use, but the bacteria producing the disease process would not necessarily have been reduced. Peroxides are found in dentifrices in combinations with sodium bicarbonate in

concentrations of 1.5% or less. Also, they are found in bleaching agents discussed later in this chapter. As long-term studies of the effect of oxygenating agents are unavailable and short-term studies offer contradictory findings, questions of safety have been raised with regard to chronic use (38, 39).

Phenolic Compounds

Listerine

Short-term studies of phenolic compounds have shown plaque and gingivitis reductions averaging 35%, and long-term studies have shown plaque reduction averaging 35% and gingivitis reduction averaging 30% (40–42).

The only product in this category that has been adequately studied is Listerine. Listerine is a mixture of essential oils—thymol, menthol, a eucalyptol, and methylsalicylate. The mechanism of action appears to be related to alteration of the bacterial cell wall. This product is uncharged and has a low substantivity. Adverse effects reported have been a burning sensation and bitter taste. It is available in a 21.6–26.9% alcohol vehicle with a pH of 4.2. Recommended usage is twice daily, and the ADA accepts the product and some of its generic copies for the control of plaque and gingivitis.

Plax

Only short-term, clinical studies with small numbers of patients have been published. These pilot studies suggested some reduction in plaque when this product was used as a prebrushing rinse (43). Effects on gingivitis have not been reported. Additional long-term studies have questioned the efficacy of this product to reduce plaque and gingivitis.

The active ingredient is stated to be sodium benzoate. However, the product also has about 30% of some of the ingredients (oils) found in Listerine. It contains 7.5% alcohol. Usage is as a prebrushing rinse. It is not ADA accepted.

Quaternary Ammonium Compounds

Quaternary ammonium compounds have been evaluated in a number of short-term studies relative to their effect on plaque and gingivitis. In these studies, an average plaque reduction of 35% has been reported, with mixed effects on gingival health (44, 45). A 6-month study has been reported showing a 14% reduction in plaque and a 24% reduction in gingivitis (46). Cepacol®, Scope®, and Advanced Care Viadent™ are well-known representatives of this group, each with concentrations of approximately

0.05% cetylpyridium chloride. In addition, Scope contains 0.005% domiphen bromide. The mechanism of action is related to increased bacterial cell wall permeability, which favors lysis, decreased cell metabolism, and a decreased ability for bacteria to attach to tooth surfaces. These agents are categorized as being cationic, which favors their attraction to anionic surfaces of teeth and plaque. They are surface-active agents that alter surface tension and have some substantivity.

Adverse effects have been some tooth staining and a burning sensation in the oral cavity. These agents are available in a 14–18% alcoholic vehicle with a pH range of 5.5–6.5. Recommended usage is twice daily, and they are not ADA accepted.

Sanguinarine

Short-term studies of sanguinarine have shown some plaque and gingivitis reduction (47, 48). In the long-term studies of the product in a dentifrice form, no significant reduction in plaque or gingivitis occurred, with the exception of one study in which the product was used as a dentifrice and as a mouthrinse (49, 50).

The proposed mechanism of action is by alteration of bacterial cell surfaces so that aggregation and attachment is reduced.

Sanguinarine (benzophenathradine) is derived from the bloodroot plant (*Sanguinaria canadensis*). The extract concentration in the product is 0.03%, which equals 0.10% sanguinarine. It also contains 0.2% zinc chloride. The product may be cationic, and the degree of substantivity is unclear. Adverse effects have been a burning sensation and a question of epithelial cell dysplasia. It is available as Viadent toothpaste and Viadent mouthrinse. The mouthrinse pH is 4.5, the dentifrice pH is 4.8, and the alcohol content of the rinse is 11.5%. It is not ADA accepted.

Triclosan

Triclosan (2,4,4'-Trichloro-2'-hydroxydiphenyl ether) is a new antiplaque/antigingivitis agent available in dentifrices. The addition of a copolymer, vinylmethyl-ether maleic acid (Gantrez), has been shown to improve the effectiveness of triclosan by enhancing its retention (substantivity) by hard and soft surfaces. This formula (Colgate Total) has been approved by the FDA for sale in the United States, and is ADA accepted. Claims allowed are for the reduction of plaque, gingivitis, calculus, and caries. Studies as early as 1973 showed that this chemical agent had a broad spectrum, antimicrobial effect against a wide range of gram-positive and gram-negative bacteria found in the mouth. The minimal concentration of triclosan for oral pathogens is 0.3 mg/ml. Triclosan's

antibacterial activity is not affected by anionic agents, such as lauryl sulfate, which are essential to dentifrice and mouthwash formulations—a fact that broadens its range of use. In doses lower than 0.5%, taste perception is minimally affected; however, at concentrations of greater than 0.5%, undesirable effects on taste occur.

Zinc citrate

This agent is found in some tartar control dentifrices and also has some plaque- and gingivitis-reducing properties as found in Mentadent® and Advanced Care Viadent®. ADA does not accept dentifrices with this ingredient for reducing plaque and gingivitis

MECHANICAL PRODUCTS FOR PLAQUE CONTROL

Good control of plaque is accomplished by mechanical procedures, which include brushing, flossing, and professional prophylaxis. A professional cleaning is recommended at least twice a year to remove plaque and tartar (calculus), both supragingivally and subgingivally.

Brushing

A recent survey of brushing habits in the United States showed that only 60% of the public follow a strict brushing regimen. Clearly, motivation and education are needed in this area.

For the average adult, a soft brush with rounded bristles is most efficient in removing plaque from supragingival tooth surfaces (with the exception of the surfaces between the teeth). Most bristles are made of nylon, and the bristle ends are rounded. Subgingival plaque can be removed only to a depth of a few millimeters. For patients with a highly developed gagging reflex, a child's toothbrush is recommended. In addition, these patients sometimes find that placing a small amount of salt on the tongue is helpful in checking the desire to gag.

Studies of toothbrushing methods indicate that thoroughness is more important than technique. In the United States, the most widely used technique is one in which the bristles are directed into the gingival crevice at a 45° angle, and a gentle, jab-jiggle action is used. The motion is elliptical, rather than a back-and-forth scrubbing. Power toothbrushes are of special value for people who have motor coordination problems or difficulty in properly removing plaque by manual brushing. For children, the novelty effect of the powered brush is sometimes of motivational value. Also, a number of

studies have shown advantages of 10–15% better plaque removal than manual brushing.

Flossing

Surveys find that only 25% of the population questioned use dental floss regularly. Flossing is essential for removal of plaque from the surfaces between teeth and under the gumline, where the toothbrush does not reach. Because plaque has a propensity to build up in these areas, some dentists feel that flossing is actually more important than brushing.

Patients who do not have the manual dexterity to use dental floss can use the various types of dental floss holders or powered interdental cleaners (e.g., Braun InterClean and WaterPik Interdental Cleaner). Dental floss is available in waxed, lightly waxed, and unwaxed varieties. Most dentists feel that lightly waxed and unwaxed types are the most efficient in plaque removal. If the floss shreds or splits during use, this may be a sign of decay between the teeth or a defective filling margin. However, new flosses have been introduced that do not shred and are easier to use. The first of these new flosses was Glide Floss, followed by similar products from Colgate and Oral B®.

Disclosing Agents

Disclosing agents are dyes similar to those in food colorings that, when introduced into the oral cavity, color the supragingival plaque and make it easily visible. Various dyes are available in both liquid and tablet form. They are used in the dental office and at home both to increase the patient's awareness of plaque and to demonstrate where self-care has been ineffective in removing plaque.

Toothpastes

The majority of toothpastes advertised as specially formulated to control plaque contain (in addition to fluoride) a foaming agent and a mild abrasive, both of which facilitate plaque removal. However, the only toothpaste accepted by the ADA as possessing an active ingredient with proven ability to prevent or control plaque formation and reduce gingivitis is Colgate Total, with Triclosan as the active ingredient. Toothpastes claiming to be effective against plaque are simply more effective than brushing without any toothpaste because the use of toothpaste motivates people to brush longer and more thoroughly. In fact, it is mainly the mechanical action of brushing that removes plaque.

Toothpastes are effective as vehicles to deliver fluoride and Triclosan to the tooth surface, and although fluoride may have some effect against plaque bacteria and their enzymes, its major effect is to make the tooth surface more resistant to destruction by plaque bacteria. With the exception of Colgate Total, other toothpastes are accepted by the ADA for their fluoride content and effectiveness against tooth decay but not for their plaque- and gingivitis-reducing properties.

Other Oral Hygiene Aids

A number of devices aid in the removal of plaque from surfaces between teeth, around bridgework, and in other areas that are difficult to reach. The limitation of many of these devices is that they are effective for control of supragingival plaque but, at best, can remove subgingival plaque only to a depth of few millimeters. Therefore, they are of minimal value against subgingival plaque located deeper within the gingival crevice, as is the case in periodontal disease.

Various oral irrigators on the market remove some loosely attached plaque and particles of debris present around teeth and dental appliances, including braces (51). Because they are not effective in removing all attached plaque, they are not substitutes for brushing and flossing; rather, they should be used as adjuncts to these procedures (52). In addition, oral irrigators are limited in their ability to reach subgingival plaque. However, the introduction of subgingival applicator tips allows solutions to be delivered 6–7 mm apically.

Some studies have suggested that irrigators may alter plaque composition by eluting bacterial endotoxins (53). Patients with severely inflamed gum tissues should be cautioned to use irrigators at low pressures to guard against tissue laceration (if the tissue is severely inflamed), which may aggravate the existing problem.

Tartar (Calculus)-Reducing Products

A number of products, both dentifrices and mouthrinses, are available for reduction of supragingival calculus (tartar) in dental patients (54, 55). Calculus reduction has been shown with dentifrices containing pyrophosphates, zinc salts, triclosan, and papain.

The incidence of calculus formation ranges from 45 to 66%, with some variation between males and females and different age groups. Although supragingival calculus is not a major etiologic agent for gingivitis or periodontitis, its surface porosity provides an environment for plaque

formation. In addition, it serves as a plaque-delivery system by holding plaque against gingival tissues. Although plaque formation has been well correlated with gingivitis and periodontitis, a similar correlation for calculus has not been reported. For this reason, the ADA does not offer an acceptance program for products that reduce calculus formation because this is considered to be a cosmetic issue, rather than an issue of disease.

The mechanism of action of the calculus-reducing chemicals is related to the latter's ability to inhibit crystal growth and interrupt the transformation of calcium phosphate (found in foods and saliva) into dental calculus. This effect may occur as follows:

1. The agents complex on the tooth surface to block receptor sites for calcium phosphate that precipitates from saliva and chemically absorbs to initiate calculus formation.
2. This same receptor site blockage also occurs in the calculus matrix as it begins to form.
3. The pyrophosphate complexes combine with free calcium in saliva to inhibit the attachment at the tooth surface (probably a secondary mechanism).

Because these products offset mineralization, there has been concern over demineralization of teeth. All manufacturers have addressed this issue and have reported that this has not been a problem, probably because of the positive effect of fluoride on remineralization of dentin and enamel.

Some patients cannot use tartar control toothpastes containing pyrophosphates because they develop tooth sensitivity and sloughing of tissue (56). These adverse effects have not been reported with nonpyrophosphate-containing tartar control products such as those made by Den-Mat Corporation (Rembrandt[®] Toothpaste).

Crest tartar control dentifrice

This dentifrice contains 3.4% tetrasodium pyrophosphate and 1.37% disodium dihydrogen pyrophosphate to reduce calculus. Also, 0.243% sodium fluoride is included for caries reduction and prevention. It was the first calculus-reducing dentifrice introduced into the United States. On the basis of various clinical studies, a reduction of 30–40% can be expected. It has also been shown to significantly reduce the tooth staining seen in some patients who use chlorhexidine.

Crest tartar control mouthrinse

This mouthrinse provides 1.6% ionic pyrophosphate from disodium and tetrasodium pyrophosphate to act against calculus formation and 0.05% sodium fluoride as a

caries-reducing agent. Data on the extent of calculus and caries reduction were not available when this article was written because the product was in test market.

Colgate tartar control dentifrice

This product contains 5% tetrasodium pyrophosphate and a polymeric fatty acid with the company-patented name of Gantrez as the calculus-reducing agents. Also, 0.243% sodium fluoride is included for caries reduction and prevention. On the basis of various clinical studies, a calculus reduction of 35–50% can be expected. It is equal to Crest in terms of calculus reduction, with some clinical studies even suggesting a superiority to Crest.

Colgate tartar control mouthrinse

This mouthrinse contains tetrasodium pyrophosphate and tetrapotassium pyrophosphate, which provide 1% ionic pyrophosphate. It also contains 0.02% fluoride. Calculus reduction has been reported to be 35–40% with twice-a-day rinsing, with no claim made for caries reduction.

Listerine tartar control mouthrinse

This mouthrinse contains 0.09% zinc chloride to reduce calculus and also contains the same ingredients as Listerine mouthrinse.

Rembrandt mouth-refreshing rinse

This product has been shown to reduce tartar due to a formulation of surface-active agents and citroxain, a form of papain.

Targon[®]

This mouthrinse reduces staining due to a formulation of surface-active agents. With the introduction of these products, the practitioner is offered a variety of dosage forms and flavors. If one decides to recommend a calculus-reducing agent, the product selection should be based on the product the patient likes to use best and one that will not diminish his awareness of the importance of plaque control. For example, if he is not already using a mouthrinse, would its introduction de-emphasize the mechanical methods of brushing and flossing as a means of plaque and/or calculus reduction?

HALITOSIS

Local factors, systemic factors, or a combination of both can cause halitosis. It is estimated that 80% of all mouth odors are caused by local factors within the oral cavity,

and these odors are most often associated with caries, gingivitis, and periodontitis. Oral malodors occur because of the action of various microorganisms on proteinaceous substances, such as, exfoliated oral epithelium, salivary proteins, food debris, and blood (57, 58).

Studies have shown that saliva from individuals who are free of dental disease produces malodor less rapidly than saliva from patients with dental disease. It has also been observed that after prolonged periods of decreased salivary flow and abstinence from food and liquid malodors tend to be most severe.

Various oral bacteria produce products that are degraded to a number of compounds, foremost of which are sulfides and mucoproteins (59). These compounds have been most often associated with oral malodor. Specifically, it appears that oral malodor usually results from the bacterial-mediated degradative processes of methyl mercaptan and hydrogen sulfide in oral air. Ammonia is also produced but does not appear to contribute significantly to halitosis. It has even been suggested that ammonia production may improve the odor of mouth air.

Control of halitosis is directed at its etiology. If systemic factors are the problem, a medical consultation is indicated. If local factors are responsible, efforts should be directed toward their elimination. However, for many patients, systemic or local factors cannot be identified. Tongue scraping has been shown to reduce malodor in some patients.

Mouthwashes and dentifrices can serve an esthetic function by reducing halitosis. They can accomplish this by masking malodors, acting as antimicrobial agents, or both. There are no ADA-accepted products to reduce halitosis at this time.

TOPICAL ANESTHETICS

Topical anesthetic agents are selected for their ability to diffuse into the oral mucosa. Because many anesthetics used effectively for nerve block or infiltration do not adequately cross the mucosa, they cannot be used for topical anesthesia.

The concentration of anesthetic used for surface application is 2–5%. The rate of onset of topical anesthesia ranges from 2 to 5 min, is of relatively short duration, and has minimal effects deep to the area of application. One exception is a mucosal patch containing 10.4 mg of Lidocaine that gives anesthesia deep into the gingiva after a 5-min application period. Systemic absorption of topical anesthetics applied to the oral

mucosa is rapid, and blood levels may approach those seen following injection (60).

Several drugs used as topical anesthetics are not readily soluble in water but are soluble in organic solvents. They are, thus, prepared in alcohol, propylene glycol, polyethylene glycol, volatile oils, and other vehicles suitable for surface application. Their slower absorption rates make them safer for topical use on abraded or lacerated tissue. They produce anesthesia for short periods.

Topical anesthetics are useful to temporarily relieve the pain of ulcers, wounds, and other injured areas. The topical use of anesthetic agents before injection may produce superficial anesthesia. They are also of value in taking impressions or intraoral radiographs in patients with an excessive gag reflex.

Patients who are allergic to parenterally administered local anesthetics will also be allergic to topical application of these agents. In addition, as the agents may be absorbed into the systemic circulation, careless application of excessive amounts can result in signs of systemic toxicity. Symptoms of toxicity should be treated, as they would be for injectable agents. One can minimize the absorption of these agents by limiting the concentration of the drug, the area of application, and the total amount applied. In general, for topical anesthesia, one should use no more than one-fourth to one-half the maximum recommended dosage for injection of the agent.

Some topical anesthetic preparations are marketed in spray containers. These containers make it difficult to control the amount of material expelled and to confirm the agent was applied to the desired site. If these agents are applied to the posterior part of the mouth, a patient may inhale enough of the aerosol spray to provoke a toxic reaction (61). Use of topical anesthetics on the posterior pharynx may alter the swallowing reflex.

Benzodent

This ester-type anesthetic is poorly absorbed. Because it contains benzocaine, which has a low water solubility, it is prepared in a base containing petrolatum and sodium carboxymethylcellulose. Eugenol is included for its antiseptic and anodyne properties. Hydroxyquinoline sulfate is a preservative. This ointment can be directly applied to abraded or ulcerated lesions with minimal systemic effects. It is sometimes used to temporarily relieve denture sores and painful lesions.

Hurricane

This ester-type anesthetic also contains benzocaine and is prepared in a polyethylene glycol base with flavoring

agents added. It is available as a liquid, gel, or spray. The propellant for the spray is A 70.

Butyn

This ester-type anesthetic contains butacaine, with benzyl alcohol as a preservative. It is available as an ointment. The maximum dose is 5 ml of a 4% solution or 200 mg.

Cetacaine

This ester-type anesthetic is a combination of tetracaine HCl (2%), butyl aminobenzoate (2%), and benzocaine (14%). Benzalkonium chloride and cetyl dimethylammonium bromide are included as surface-active agents to facilitate the passage of benzocaine into the mucosal tissues.

Tetracaine is rapidly absorbed through biologic membranes and requires no facilitating agents. Because of its high toxicity and absorption, agents containing tetracaine should be used with caution and should not be placed under dentures. Spraying this agent is dangerous because the patient may inhale the aerosol. The maximum amount to be applied is 20 mg or 1 ml of a 2% solution. Cetacaine is available as a liquid, ointment, spray, or gel.

Xylocaine

This amide-type anesthetic contains a lidocaine base in a monoaqueous vehicle. It is available as a 2% viscous product containing lidocaine 2%, sodium carboxymethylcellulose, sodium saccharin, methylparaben, propylparaben, flavors, and purified water. It is also available as a liquid containing 4% methylparaben, sodium hydroxide, and flavoring agents. Another dosage form is a 5% ointment containing lidocaine 5%, polyethylene glycol, propylene glycol, and flavoring. The maximum dose is 300 ml of a 5% liquid form or 15 ml of the 2% viscous preparation.

DENTIFRICES AND SENSITIVE TEETH

Sometimes a patient will complain of teeth that are hypersensitive to heat and cold. These teeth usually have exposed root surfaces, sometimes with a loss of cementum. Most teeth, when in an ideal position in the mouth, have only the enamel surface exposed to the oral cavity. On occasion, such teeth may even respond with pain to extreme heat or cold.

However, in true dentinal hypersensitivity, the response to thermal and tactile changes is more pronounced, sometimes eliciting severe pain. Root surface exposure that allows contact with stimuli may occur because of gingival recession or following periodontal therapy.

Several theories have been advanced to explain the mechanism of dentinal hypersensitivity (62): innervation of the dentinal tubules, permitting transmission of impulses to the pulp, or the presence of lymph fluid in the dentinal tubules. In the latter case, exposure of dentin results in increased colloidal pressure on the tubules (thereby increasing pressure on the odontoblastic cells). Also proposed is a hydrodynamic mechanism involving the movements of tubular fluid in either direction, which elicits pain in the nerves of the pulp. Although no one theory has been proved, occlusion of the dentinal tubules by various methods brings relief. Various dentifrices are recommended for the treatment of sensitivity, with some success.

The greatest success occurs with dentifrices containing 5% potassium nitrate (e.g., Sensodyne[®], Rembrandt, Crest, AquaFresh[®], and Protect[®]), and some fluoride-containing dentifrices. Recently, a dentifrice containing potassium nitrate and stannous fluoride has been introduced to treat this problem (Colgate).

The primary mechanisms postulated for these dentifrices are that they occlude dentinal tubules, preventing stimuli from the oral cavity from irritating the dental nerve via these tubules. Also, those containing potassium may depolarize nerve fibers resulting in decreased impulse conduction and an associated decrease in pain.

For maximum effect, a patient must use only one of these dentifrices for at least a month. If no benefit occurs after a month, a different dentifrice should be recommended or other methods employed.

Topical varnishes containing sodium fluoride have also been shown to reduce dentinal hypersensitivity (e.g., Duraphat and Fluor-Protect) as well as reduce root surface caries.

BLEACHING AGENTS

Tooth-bleaching agents can be classified as to whether they are used for external or internal bleaching and whether the procedure is performed in the office by a dentist or at home by a patient. For tooth bleaching, hydrogen peroxide (H_2O_2) is used alone at levels of 30% or at 10–22% levels in a stable gel of carbamide peroxide (urea peroxide) that breaks down to form hydrogen peroxide (3.35% H_2O_2 from 10% carbamide peroxide),

urea, ammonia, and carbon dioxide. The FDA has not approved peroxide solutions for use as a home bleach, however.

Internal Bleaching

Internal bleaching produces reliable results when used to eliminate intrinsic stains in dentin caused by blood breakdown products or endodontics or for stains in receded pulp chambers. Internal bleaching is always an in-office procedure.

External Bleaching

External bleaching is indicated for teeth that are disclosed from aging, fluorosis, or staining due to the effects of tetracycline (63). External bleaching can be applied by the dentist or staff or can be applied by the patient in home-use bleaching. When dentist-administered and home-use bleaching are both used, it is called “dual bleaching.”

Dentist-applied external bleaching can be done with periodic repetitions of an office-bleaching agent using Superoxol or 30% H_2O_2 . An etching gel containing phosphoric acid applied to selected dark areas increases the penetration of the bleach. Light is used to produce heat, which accelerates the bleaching process. External bleaching may need additional treatment every 1–2 years to touch up relapses. Severely stained teeth may require more frequent retreatment.

Home bleaching, supervised by the dentist, is done by the patient at home using a custom-made carrier that holds the bleach against the patient's teeth (64). After the desired result is achieved, overnight use on a periodic basis (1–4× month) can maintain the lightening that has been achieved.

External bleaching is seldom permanent, lasting approximately 1–4 years, after which teeth gradually return to their original color. Usually the younger the patient, the longer the bleaching will last. The more difficult it is to bleach a tooth, the more likely it is to discolor again. Bluish–gray stains seem to reappear more quickly than yellow stains. Because reoccurrence of staining is unpredictable, promises about longevity should not be made. Internal bleaching usually lasts longer than external bleaching.

Whitening Formulas

Whitening of teeth can occur by two mechanisms. One method is mechanical, in which an abrasive is used to

remove debris from the tooth. The other method involves either the use of peroxides, which react with water to form free oxygen radicals that help to whiten the teeth or a combination of mechanical and chemical actions. This latter mechanism is found with bleaching agents and is longer lasting than whitening procedures.

XEROSTOMIA

The widely held belief that saliva production significantly decreases with age is not well supported in the literature dealing with this subject. Aging does not appear to play a major role as a single contributing factor in causing xerostomia (65). However, senior citizens may receive medication that produces the side effect of xerostomia. The aged also develop medical problems that can diminish salivary production.

For these reasons, most of the studies of xerostomia have focused on older patients. One study reported a direct correlation between the intake of anticholinergic drugs, sedatives, and hypnotics and xerostomia (66). Over 400 drugs have been identified as potential reducers of salivary flow by acting on the cholinergic (parasympathetic) system either directly or indirectly (67). Another study found that the use of drugs producing xerostomia increases with age and, as expected, is highest in institutionalized patients. There are over 30 classifications of prescription and nonprescription medications that can reduce salivary flow.

Other factors that can cause xerostomia are systemic disorders and radiation (67). These factors must be considered in the differential diagnosis of xerostomia.

Clinical Problems with Xerostomia

Clinical problems associated with reduced salivary flow include difficulty chewing foods, reduced denture retention, recurrent caries, root surface caries, and oral candidiasis (low grade). When any of these conditions are found in a patient, regardless of age, reduced salivary flow should be considered in a differential diagnosis of the problem.

Treatment

Treatment of patients with reduced salivary flow should include the following: (1) drug and dosage changes by the patient's physician in consultation with the patient's dentist; (2) use of artificial saliva in a spray form; (3) use of mouth moisturizers and lip balms; (4) use of sugarless hard

candy; (5) frequent sipping of water; (6) use of decaffeinated products; (7) use of pilocarpine (Salagen) 3–5× daily; and (8) inclusion of citrus and pineapple flavors in the diet.

LOCAL DELIVERY OF ANTIMICROBIAL AGENTS

In the past decade, significant research and product innovations have focused the attention of dental practitioners on the concept of the local application of antimicrobials to treat periodontal diseases. Three local delivery agents are now available in the United States, and two additional products are available in other parts of the world.

Though the rationale of antimicrobial approaches to treatment is evident, their limitations have also been evident. With systemic therapy, it may be difficult to achieve bacteriostatic or bactericidal antibiotic concentrations in pockets without using doses that evoke systemic side effects. The development of bacterial resistance is also an issue. The rise of antibiotic resistant, disease-producing bacterial strains is currently a major public health concern. Prolonged, repetitive courses of antibiotics for recurring dental infections is discouraged, because such practice can more readily lead to the development of resistance (68). Preferably, the cause of the infection should be eliminated rather than merely “managed” with antibiotics. Once antibiotic therapy is initiated, however, the importance of compliance with dosage and duration of treatment must be stressed with the patient. With poor patient compliance, under-dosing may occur, which, in turn, can favor the emergence of resistant bacteria.

Controlled Medication Delivery

The limitations of systemic therapy have prompted extensive research for the development of alternative delivery systems. Local, controlled delivery systems are available to release pilocarpine to the eye for a week after single placement for treatment of glaucoma. The oral cavity offers another relatively accessible disease site for localized therapy. In localized therapy for periodontal disease, the concern is the difficulty in reaching deep pockets and sustaining bacteriostatic or bactericidal levels long enough to be effective but not causing the development of resistance. The following is summary information about the various products available in the United States.

Tetracycline-containing fibers (Actisite®)

The first local delivery product available in the United States, one which has been extensively studied, is an ethylene–vinyl acetate copolymer fiber, diameter 0.5 mm, containing tetracycline, 12.7 mg/9 in. (Actisite tetracycline fiber; manufactured by Alza Corporation, Piscataway, NJ; distributed by Procter & Gamble Co., Cincinnati, OH; Fig. 1). When packed into a periodontal pocket, it is well tolerated by oral tissues, and for 10 days, it sustains tetracycline concentrations exceeding 1300 µg/ml, well beyond the 32–64 µg/ml required to inhibit the growth of pathogens isolated from periodontal pockets (69, 70). In contrast, crevicular fluid concentrations of only 4–8 µg/ml are reported following systemic tetracycline administration, 250 mg, 4× daily for 10 days (total oral dose, 10 g; 68). Thus, controlled site-specific tetracycline delivery can achieve a bactericidal effect at approximately 1/1000th of the dose administered systemically.

Studies demonstrate that the tetracycline fibers, applied with or without scaling and root planing, reduce probing depth, bleeding on probing, and periodontal pathogens and provide gains in clinical attachment level. Such effects are significantly better than those attained with scaling and root planing alone or with placebo fibers. The fibers used in conjunction with scaling and root planing have also provided a statistically significant improvement in probing depth reduction and clinical attachment level gains of over 60% and in bleeding on probing reductions over scaling and root planing alone at 6 months after therapy (71).

Actisite was the first local delivery system cleared by the FDA for the adjunctive treatment of recurrent periodontal disease. Although 6-month studies have demonstrated their value, longer-term studies are needed.

Chlorhexidine delivery system (PerioChip®)

A newer development in controlled local delivery, one that utilizes the antiseptic chlorhexidine as the antimicrobial agent, has been introduced in a number of countries and was recently cleared by the FDA for use in the United States. This delivery system, PerioChip (manufactured by Perio Products Ltd., Jerusalem, Israel; distributed by Dexell Pharmaceuticals, Edison, NJ), was developed in Israel and has been tested in the United States as well as in Europe.

The PerioChip is a small chip (4.0 × 5.0 × 0.35 mm) composed of a biodegradable hydrolyzed gelatin matrix into which has been incorporated 2.5 mg chlorhexidine gluconate per chip. It is rounded on one end and inserts easily and in less than a minute into periodontal pockets that are 5 mm or greater in depth. The PerioChip releases

chlorhexidine and maintains drug concentrations in the gingival crevicular fluid greater than 100 µg/ml for at least 7 days, concentrations well above the tolerance of most oral bacteria (72). Because the PerioChip biodegrades in 7–10 days, a second appointment for removal is not needed. Studies with the PerioChip were as long as 9 months (73). At 9 months, significant decreases were observed in probing depth from baseline favoring the active chip plus scaling and root planing compared with controls (scaling and root planing only): chlorhexidine chip plus scaling and root planing, -0.95 ± 0.05 mm; placebo chip plus scaling and root planing, -0.69 ± 0.05 mm ($p = 0.00056$); scaling and root planing alone -0.65 ± 0.05 mm ($p = 0.00001$). The proportion of pocket sites with a probing depth reduction of 2 mm or more was increased in the chlorhexidine chip group compared with scaling and root planing alone, a difference which was statistically significant on a per patient basis ($p < 0.0001$). Improvements favoring the chlorhexidine chip compared with controls were also observed for clinical attachment levels at 9 months, improvements that were significant when the data were pooled ($p < 0.05$). Bleeding on probing was reduced in the active chip group compared with both controls, differences which were significant in one of the two studies ($p < 0.05$) and when the data were pooled ($p = 0.012$).

The results of these studies suggest that the Perio Chip may be a valuable adjunct to scaling and root planing in the treatment of periodontal disease. This product is easily placed, requires no appointment for removal, and provides reductions in probing depths similar to subgingivally placed antibiotics now available in various countries across the world. In addition, a major advantage of this system is that its active agent is an antiseptic instead of an antibiotic.

Subgingival Delivery of Doxycycline (Atridox®)

Atridox (manufactured by Atrix Laboratories, Fort Collins, CO) is a recently developed gel system that incorporates the antibiotic doxycycline (10%) in a syringeable gel system. An animal study in beagle dogs initially suggested some benefit to locally delivered doxycycline (74), and it is used in the veterinary population.

A recent 9-month, multicenter study was designed to study the effects of subgingivally placed doxycycline compared to subgingival placement of the vehicle and an herbal agent (Sanguinaria; 75). No scaling or root planing was performed in any of the groups, and there was no untreated group. The patients were instructed in oral

hygiene and randomly assigned to one of three groups: vehicle control, 5% sanguinarine in the vehicle control, and 10% doxycycline in the vehicle control.

Treatment with doxycycline was more effective than the other treatments at all time periods, with the exception of the 3-month clinical attachment level value. Also, when the authors evaluated the effect based on initial probing depth, the differential effect in the doxycycline group in comparison with the other two groups was greater as pretreatment probing depth increased. For the doxycycline group, the reduction in clinical attachment level at 9 months showed a gain of 0.4 mm compared to vehicle control, the reduction in probing depth was 0.6 mm greater than vehicle control, and the reduction of bleeding on probing was 0.2 units greater than vehicle control. Although the differences were small, they were statistically significant. The patient's oral hygiene scores (plaque index) averaged between 0.7 and 1.1 in all three groups throughout the study. Although resistance was not evaluated in this study, the local application of doxycycline has previously been reported to show transient increases in resistance in oral microbes and no overgrowth of foreign pathogens (76).

Data have recently been presented from two multicenter clinical trials (77). All treatment groups showed clinical improvements from baseline over the 9-month period. The results for all parameters measured were significantly better in the doxycycline group compared with vehicle control and oral hygiene only. Compared with scaling and root planing, the effects of doxycycline on clinical attachment level gain and probing depth reduction were equivalent.

Clinical study results suggest a potential periodontal benefit from the subgingival application of doxycycline. However, the value of this agent as an adjunct to scaling and root planing is untested. This product is cleared by the FDA for use in animals and humans.

HOST MODULATION

In 1998, the first nonantimicrobial drug to treat periodontal disease, Periostat, was approved by the FDA. Nonetheless, in the antibiotic family, the dose of the drug used is too low to kill bacteria. Periostat acts through its effects on inhibition of metalloproteinases, such as, collagenase and gelatinase. The therapeutic objective is to modulate the inflammatory host response. Periostat, available as a 20 mg capsule of doxycycline hyclate, is prescribed for use by patients twice daily. The mechanism of action is by suppression of the activity of collagenase, particularly that produced by polymorphonuclear leukocytes. Although this

drug is in the antibiotic family, it does not produce any antimicrobial effects because the dose of 20 mg twice daily is too low to affect bacteria. As a result, resistance to this medication cannot develop.

Four double-blind, clinical, multicenter studies in over 650 patients have demonstrated that Periostat improves the effectiveness of professional periodontal care and slows the progression of the disease process (78).

REFERENCES

1. Maier, F.J. *Manual of Water Fluoridation Practice*; McGraw-Hill: New York, 1963; 48–59.
2. Brudevold, F.; Gardner, D.E.; Smith, F.A. Distribution of Fluoride in Human Enamel. *J. Dent. Res.* **1956**, *35*, 420–429.
3. Tinanoff, N.; Brady, J.M.; Gross, A. The Effect of NaF and SnF₂ Mouthrinses on Bacterial Colonization of Tooth Enamel: TEM and SEM Studies. *Caries Res.* **1976**, *10*, 415–426.
4. Horowitz, H.S.; Heifetz, S.B.; Meyers, R.J.; Driscoll, W.S.; Korts, D.C. Evaluation of a Combination of Self-Administered Fluoride Procedures for the Control of Dental Caries in a Nonfluoride Area: Findings After Four Years. *JADA* **1979**, *98* (2), 219–223.
5. Council on Dental Therapeutics, American Dental Association. *JADA* **1975**, *91*, 1250–1251.
6. Newbrun, E. Current Regulations and Recommendations Concerning Water Fluoridation, Fluoride Supplements, and Topical Fluoride Agents. *J. Dent. Res.* **1992**, *71* (5), 1255–1265.
7. Levy, S.M. Review of Fluoride Exposures and Ingestion. *Commun. Dent. Oral. Epidemiol.* **1994**, *22* (3), 173–180.
8. DePaola, P.F.; Lax, M. The Caries-Inhibiting Effect of Acidulated Phosphate-Fluoride Chewable Tablets: A Two-Year Double-Blind Study. *JADA* **1968**, *76*, 554–557.
9. Driscoll, W.S.; Heifetz, S.B.; Brunelle, J.A. Treatment and Posttreatment Effects of Chewable Fluoride Tablets on Dental Caries: Findings after 7½ Years. *JADA* **1979**, *99* (5), 817–821.
10. Glenn, F.B. Immunity Conveyed by Sodium-Fluoride Supplement During Pregnancy: Part II. *ASDC J. Dent. Child.* **1979**, *46* (1), 17–24.
11. Glenn, F.B. Immunity Conveyed by a Fluoride Supplement during Pregnancy. *ASDC J. Dent. Child.* **1977**, *44* (5), 391–395.
12. Ripa, L.W. A Critique of Topical Fluoride Methods (Dentifrices, Mouthrinses, Operator- and Self-Applied Gels) in an Era of Decreased Caries and Increased Fluorosis Prevalence. *J. Public Health Dent.* **1991**, *51* (1), 23–41.
13. Ripa, L.W. Review of the Anticaries Effectiveness of Professionally Applied and Self-applied Topical Fluoride Gels. *J. Public Health Dent.* **1989**, *49* (5), 297–309.
14. Segal, A.H.; Stiff, R.H.; George, W.A.; Picozzi, A. Cariostatic Effect of a Stannous Fluoride-Containing Dentifrice on Children: Two-Year Report of a Supervised

- Toothbrushing Study. *J. Oral Ther. Pharmacol.* **1967**, 4 (3), 175–180.
15. Keyes, F.M.; Overton, J.J.; McKean, T.W. Clinical Trials of Caries Inhibiting Dentifrices. *JADA* **1961**, 63, 189–193.
 16. DePaola, P.F.; Soparkar, P.; Foley, S.; Bookstein, F.; Bakhos, Y. Effect of High-Concentration Ammonium and Sodium Fluoride Rinses on Dental Caries in School-Children. *Commun. Dent. Oral Epidemiol.* **1977**, 5 (1), 7–14.
 17. Reed, A.J.; Bibby, B.G. Preliminary Report on Effect of Topical Applications of Titanium Tetrafluoride on Dental Caries. *J. Dent. Res.* **1976**, 55 (3), 357–358.
 18. Koch, G. Long-term Study of Effect of Supervised Toothbrushing with a Sodium Fluoride Dentifrice. *Caries Res.* **1970**, 4 (2), 149–157.
 19. Koch, G. Comparison and Estimation of Effect on Caries of Daily Supervised Toothbrushing with a Dentifrice Containing Sodium Fluoride and a Dentifrice Containing Potassium Fluoride and Manganese Chloride. A Three-Year Clinical Test. *Odontol. Rev.* **1972**, 23 (3), 341–354.
 20. Peterson, J.K. A Supervised Brushing Trial of Sodium Monofluorophosphate Dentifrices in a Fluoridated Area. *Caries Res.* **1979**, 13 (2), 68–72.
 21. James, P.M.; Anderson, R.J.; Beal, J.F.; Bradnock, G. A 3-Year Clinical Trial of the Effect on Dental Caries of a Dentifrice Containing 2% Sodium Monofluorophosphate. *Commun. Dent. Oral Epidemiol.* **1977**, 5 (2), 67–72.
 22. Moller, I.J.; Holst, J.J.; Sörensen, E. Caries Reducing Effect of a Sodium Monofluorophosphate Dentifrice. *Br. Dent. J.* **1968**, 124 (5), 209–213.
 23. Gish, C.W.; Muhler, J.C.; Howell, C.L. Self Application of Fluoride as a Community Measure. *J. Dent. Child.* **1962**, 29, 65–71.
 24. Ripa, L.W.; Leske, G.S.; Sposato, A.; Varma, A. Effect of Prior Toothcleaning on Biannual Professional APF Topical Fluoride Gel-Tray Treatments. Results after Two Years. *Clin. Prevent. Dent.* **1983**, 5 (4), 3–7.
 25. Houpt, M.; Koenigsberg, S.; Shey, Z. The Effect of Prior Toothcleaning on the Efficacy of Topical Fluoride Treatment. Two-Year Results. *Clin. Prevent. Dent.* **1983**, 5 (4), 8–10.
 26. Brandt, R.S.; Slack, G.L.; Waller, D.F. The Use of a Sodium Fluoride Mouthwash in Reducing the Dental Caries Increscent in Eleven-Year-Old English School Children. *Proc. Br. Paedoc. Soc.* **1972**, 2, 23–25.
 27. Rugg-Gunn, A.J.; Holloway, P.J.; Davies, T.G. Caries Prevention by Daily Fluoride Mouthrinsing. Report of a Three-Year Clinical Trial. *Caries Br. Dent. J.* **1973**, 135 (8), 353–360.
 28. Ripa, L.W.; Leske, G.S.; Sposato, A.L.; Rebich, T., Jr. Supervised Weekly Rinsing with a 0.2% Neutral NaF Solution: Results of a Demonstration Program after Four School Years. *JADA* **1981**, 102 (4), 482–486.
 29. Stephen, K.W. Fluoride Toothpastes, Rinses, and tablets. *Adv. Dent. Res.* **1994**, 8 (2), 185–189.
 30. Hunter, G.C., Jr.; Barringer, M.; Spooner, G. Analysis of Desensitization of Dentin by Sodium Silico-Fluoride and Gottlieb's Solution. *J. Periodontol.* **1961**, 32, 333–337.
 31. Kanouse, M.C.; Ash, M.M., Jr.; Löe, H.; Schiott, C.R. The Effectiveness of a Sodium Monofluorophosphate Dentifrice on Dental Hypersensitivity. *J. Periodontol.* **1969**, 40 (1), 38–40.
 32. Löe, H.; Schiott, C.R. The Effect of Mouthrinses and Topical Application of Chlorhexidine on the Development of Dental Plaque and Gingivitis in Man. *J. Periodontal. Res.* **1970**, 5 (2), 79–83.
 33. Grossman, E.; Reiter, D.; Sturzenberger, O.P.; De la Rosa, M.; Dickinson, T.D.; Fenetti, G.A. Six Months Study of the Effects of a Chlorhexidine Mouthrinse on Gingivitis in Adults. *J. Periodontal. Res.* **1986**, 16 (Suppl.), 33, 34.
 34. Segreto, V.A.; Collins, E.M.; Beiswanger, B.B.; De la Rosa, M.; Isaacs, R.L.; Lang, N.P. A Comparison of Mouthrinses Containing Two Concentrations of Chlorhexidine. *J. Periodontal. Res.* **1986**, 21 (Suppl.), 23–32.
 35. Tinanoff, N.; Hock, J.; Camosci, D.; Helldén, L. Effect of Stannous Fluoride Mouthrinse on Dental Plaque Formation. *J. Clin. Periodontol.* **1980**, 7 (3), 232–241.
 36. Tolle, S.L.; Bauman, D.B.; Allen, D.S. Effects of Fluoride Gels on Plaque and Gingival Health. *Dent. Hyg. (Chic.)* **1987**, 61 (6), 280–284.
 37. Wolff, L.F.; Pihlstrom, B.L.; Bakdash, D.M.; et al. *Program and Abstracts, International Association for Dental Research*; Montreal Canada, March 1988, (abstract).
 38. Weitzman, S.A.; Weitberg, A.B.; Stossel, T.P.; Schwartz, J.; Shklar, G. Effects of Hydrogen Peroxide on Oral Carcinogenesis in Hamsters. *J. Periodontol.* **1986**, 57 (11), 685–688.
 39. Rees, T.D.; Orth, C.F. Oral Ulcerations With Use of Hydrogen Peroxide. *J. Periodontol.* **1986**, 57 (11), 689–692.
 40. Fornell, J.; Sundin, Y.; Lindhe, J. Effect of Listerine® on Dental Plaque and Gingivitis. *Scand. J. Dent. Res.* **1975**, 83 (1), 18–25.
 41. Lamster, I.B.; Alfano, M.C.; Seiger, M.C. The Effect of Listerine® Antiseptic on Reduction of Existing Plaque and Gingivitis. *Clin. Prev. Dent.* **1983**, 5, 12–16.
 42. Gordon, J.M.; Lamster, I.B.; Seiger, M.C. Efficacy of Listerine® Antiseptic in Inhibiting the Development of Plaque and Gingivitis. *J. Clin. Periodontol.* **1985**, 12 (8), 697–704.
 43. Emling, R.C.; Yankell, S.L. First Clinical Studies of a New Prebrushing Mouthrinse. *Compend. Contin. Educ. Dent.* **1985**, 6 (9), 636, 638–640, 644, 645.
 44. Ciancio, S.G.; Mather, M.L.; Bunnell, H.L. Clinical Evaluation of a Quaternary Ammonium-Containing Mouthrinse. *J. Periodontol.* **1975**, 46 (7), 397–401.
 45. Sturzenberger, O.P.; Leonard, G.J. The Effect of a Mouthwash as Adjunct in Tooth Cleaning. *J. Periodontol.* **1969**, 40 (5), 299–301.
 46. Lobene, R.R.; Lobene, S.; Soparkar, P.M. The Effect of a Cetylpyridinium Chloride Mouthwash on Human Plaque and Gingivitis. *J. Dent. Res. Prog. Abst.* **1977**, 56, 595.
 47. Wennström, J.; Lindhe, J. Some Effects of a Sanguinarine-Containing Mouthrinse on Developing Plaque and Gingivitis. *J. Clin. Periodontol.* **1985**, 12 (10), 867–872.
 48. Klewansky, P.; Vernier, D. Sanguinarine and the Control of Plaque in Dental Practice. *Compend. Contin. Educ. Dent.*, S94–S97.
 49. Lobene, R.R.; Soparkar, P.M.; Newman, M.B. The Effects of a Sanguinaria Dentifrice on Plaque and Gingivitis. *Compend. Contin. Educ. Dent.* **1986**, (Suppl. 7), S185–S188.

50. Hanna, J.J.; Johnson, J.D.; Kuftinec, M.M. Six Month Evaluation of Sanguinarine Products in Orthodontic Patients. *Program and Abstracts, American Association of Dental Research*; Montreal Canada, March 1988, (abstract).
51. Elliott, J.R.; Bowers, G.M.; Clemmer, B.A.; Rovelstad, G.H. A Comparison of Selected Oral Hygiene Devices in Dental Plaque Removal. *J. Periodontol.* **1972**, *43* (4), 217–222.
52. Brady, J.M.; Gray, W.A.; Bhaskar, S.N. Electron Microscopic Study of the Effect of Water Jet Lavage Devices on Dental Plaque. *J. Dent. Res.* **1973**, *52* (6), 1310–1313.
53. Ciancio, S.G. New Evidence that Oral Irrigation Reduces Proinflammatory Proteins—Cytokines. *Biological Therapies in Dentistry* **1999**, *15*, 15–16.
54. Zacheral, W.A.; Pfeiffer, H.J.; Swancar, J.R. The Effect of Soluble Pyrophosphates on Dental Calculus in Adults. *JADA* **1985**, *110* (5), 737, 738.
55. Ciancio, S.G., Ed.; *ADA Guide to Dental Therapeutics*, 1st Ed.; American Dental Association: Chicago, 1999; 210–211.
56. Ciancio, S.G. Tartar Reduction without Pyrophosphates. *Biological Therapies in Dentistry* **1998**, (13).
57. Pitts, G.; Brogdon, C.; Hu, L.; Masurat, T.; Painotti, R.; Schumann, P. Mechanism of Action of An Antiseptic, Antiodor Mouthwash. *J. Dent. Res.* **1983**, *62* (6), 738–742.
58. Tonzetich, J.; McBride, B.C. Characterization of Volatile Sulphur Production by Pathogenic and Non-Pathogenic Strains of Oral Bacteroides. *Arch. Oral Biol.* **1981**, *26*, 963–969.
59. Yaegaki, K.; Sanada, K. Volatile Sulphur Compounds in Mouth Air from Clinically Healthy Subjects and Patients with Periodontal. Disease. *J. Periodontal. Res.* **1992**, *27*, 233–238.
60. Gilman, A.G.; Goodman, L.S.; Gilman, A., Eds. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 9th Ed.; Macmillan: New York, 1996.
61. Ciancio, S.G., Ed. *ADA Guide to Dental Therapeutics*, 1st Ed.; American Dental Association: Chicago, 1999; 4–15.
62. Ciancio, S.G. Dental Hypersensitivity. *Endodont. Dental Traumatol.* **1986**, *12*, 150–152.
63. Albers, H.F. Lightening Natural Teeth. *ADEPT Report* **1991**, *2* (2), 1–24.
64. Haywood, V.B.; Leonard, R.H.; Nelson, C.F.; Brunson, W.D. Effectiveness, Side Effects and Long-Term Status of Nightguard Vital Bleaching. *JADA* **1994**, *125* (9), 1219–1226.
65. Sreebny, L.M.; Valadini, A. Xerostomia: A Neglected Symptom. *Arch. Intern. Med.* **1987**, *147*, 133–137.
66. Sreebny, L.M.; Schwartz, S.S. A Reference Guide to Drugs and Dry Mouth. *Gerontology* **1986**, *5*, 189–192.
67. Butt, G.M. Drug-Induced Xerostomia. *J. Can. Dent. Assoc.* **1991**, *57*, 391–393.
68. Gordon, J.M.; Walker, C.B.; Murphy, C.J.; Goodson, J.M.; Socransky, S.S. Tetracycline: Levels Achievable in Gingival Crevice Fluid and In Vitro Effect on Subgingival Organisms. Part I. Concentrations in Crevicular Fluid after Repeated Doses. *J. Periodontol.* **1981**, *52*, 609–612.
69. Tonetti, M.; Cugini, A.M.; Goodson, J.M. Zero Order Delivery with Periodontol. Placement of Tetracycline Loaded Ethylene Vinyl Acetate Fibers. *J. Periodontal. Res.* **1990**, *25*, 243–247.
70. Walker, C.B.; Gordon, J.M.; McQuilkin, S.J.; Niebloom, T.A.; Socransky, S.S. Tetracycline: Levels Achievable in Gingival Crevice Fluid and In Vitro Effect on Subgingival Organisms. Part II. Susceptibilities of Periodontal. Bacteria. *J. Periodontol.* **1981**, *52*, 613–616.
71. Newman, M.G.; Kornman, K.S.; Doherty, F.M. A 6-Month Multi-Center Evaluation of Adjunctive Tetracycline Fiber Therapy Used in Conjunction with Scaling and Root Planing in Maintenance Patients: Clinical Results. *J. Periodontol.* **1994**, *65*, 685–691.
72. Briner, W.W.; Kayrouz, G.A.; Chanak, M.X. Comparative Antimicrobial Effectiveness of a Substantive (0.12% Chlorhexidine) and a Nonsubstantive (Phenolic) Mouthrinse In Vivo and In Vitro. *Compend. Contin. Educ. Dent.* **1994**, *15*, 1158–1168.
73. Jeffcoat, M.K.; Bray, K.S.; Ciancio, S.G.; Dentino, A.R.; Fine, D.H.; Gordon, J.M.; Gunsolley, J.C.; Killooy, W.J.; Lowenguth, R.A.; Magnusson, N.I.; Offenbacher, S.; Palcanis, K.G.; Proskin, H.M.; Finkelman, R.D.; Flashner, M. Adjunctive Use of a Subgingival Controlled-Release Chlorhexidine Chip Reduces Probing Depth and Improves Attachment Level Compared With Scaling and Root Planing Alone. *J. Periodontol.* **1998**, *69*, 989–997.
74. Polson, A.M.; Southard, G.L.; Dunn, R.L.; Yewey, G.L.; Godowski, K.C.; Polson, A.P.; Fulfs, J.C.; Laster, L. Periodontal. Pocket Treatment in Beagle Dogs Using Subgingival Doxycycline from a Biodegradable System. I. Initial Clinical Responses. *J. Periodontol.* **1996**, *67*, 1176–1184.
75. Polson, A.M.; Garrett, S.; Stoller, N.H.; Bandt, C.L.; Hanes, P.J.; Killooy, W.J.; Southard, G.L.; Duke, S.P.; Bogle, G.D.; Drisko, C.H.; Friesen, L.R. Multi-Center Comparative Evaluation of Subgingivally Delivered Sanguinarine and Doxycycline in the Treatment of Periodontitis. II. Clinical Results. *J. Periodontol.* **1997**, *68*, 119–126.
76. Larsen, T. Occurrence of Doxycycline-Resistant Bacteria in the Oral Cavity After Local Administration of Doxycycline in Patients with Periodontal. Disease. *Scand. J. Infect. Dis.* **1991**, *23*, 89–95.
77. Garrett, S.; Johnson, L.; Drisko, C.H.; Adams, D.F.; Bandt, C.; Beiswanger, B.; Bogle, G.; Donly, K.; Hallmon, W.W.; Hancock, E.B.; Hanes, P.; Hawley, C.E.; Kiger, R.; Killooy, W.; Mellonig, J.T.; Polson, A.; Raab, F.J.; Ryder, M.; Stoller, N.H.; Wang, H.L.; Wolinsky, L.E.; Evans, G.H.; Harrold, C.Q.; Arnold, R.M.; Southard, G.L.; et al. Two Multi-Center Studies Evaluating Locally Delivered Doxycycline Hyclate, Placebo Control, Oral Hygiene, and Scaling and Root Planing in the Treatment of Periodontitis. *J. Periodontol.* **1999**, *70*, 490–503.
78. Ciancio, S.; Ashley, R. Safety and Efficacy of Sub-Antimicrobial-dose Doxycycline Therapy in Patients with Adult Periodontitis. *Adv. Dent. Res.* **1998**, *12*, 27–31.

DRUG DELIVERY — LIQUID CRYSTALS IN

Christel C. Mueller-Goymann

*Universität Braunschweig Mendelssohnstr, Braunschweig,
Germany*

DEFINITION AND FORMATION OF LIQUID CRYSTALS

Definition

The liquid crystalline state combines properties of both liquid and solid states. The liquid state is associated with the ability to flow, whereas the solid state is characterized by an ordered, crystalline structure (1). Crystalline solids exhibit short as well as long-range order with regard to both position and orientation of the molecules (Fig. 1a). Liquids are amorphous in general but may show short-range order with regard to position and/or orientation (Fig. 1b). Liquid crystals show at least orientational long-range order and may show short-range order, whereas positional long-range order disappears (Fig. 1c) (2). Accordingly, liquid crystalline phases represent intermediate states and are also called mesophases.

A pre-requirement for the formation of liquid crystalline phases is an anisometric molecular shape that is generally associated with a marked anisotropy of the polarizability. Molecules that can form mesophases are called mesogens. Depending on the molecular shape, rod-like mesogens form calamitic mesophases and disc-like mesogens form discotic mesophases. Rod-shaped molecules are often excipients of drugs (e.g., surfactants). Even drug compounds themselves (e.g., the salts of organic acids or bases with anisometric molecular shape) fulfill the requirements for the formation of calamitic mesophases.

Formation

Starting with the crystalline state, the mesophase is reached either by increasing the temperature or by adding a solvent; accordingly, a differentiation can be made between thermotropic and lyotropic liquid crystals, respectively. As with thermotropic liquid crystals, a variation of temperature can also cause a phase transformation between different mesophases with lyotropic liquid crystals.

Thermotropic liquid crystals

Calamitic mesophases were the first liquid crystals to be found more than 100 years ago. In 1888, the botanist

Friedrich Reinitzer observed birefringence of cholesteryl esters after melting (3) and contacted the physicist Otto Lehmann, a specialist in crystallization microscopy, who interpreted the birefringence of the molten cholesteryl esters as a parallel orientation of molecules within a liquid crystal, a new kind of state (4). However, these cholesteric liquid crystals exhibit not only a parallel orientation of the anisometric molecules, but the director of the orientation rotates layer by layer in a right- or left-handed helix (Fig. 2b). The layer distance where a 360° rotation has been performed is called pitch, which is often in the magnitude of the visible light. This phenomenon, as well as the variation of pitch with temperature, is responsible for the characteristic color play of cholesteric liquid crystals. Cholesterics require chirality either of the mesogen itself or on addition of a mesogen.

The nematics are similar to cholesteric liquid crystals in having just orientational long-range order, with the deviation that the director of the preferred orientation does not rotate (Fig. 2a). If, however, a chiral mesogen is dissolved in a nematic liquid crystal, the latter will transform into a cholesteric liquid crystal.

Calamitic mesophases with parallel orientation of the molecules, which are additionally arranged in layers, are called smectic liquid crystals (Fig. 2c-e). The layer plane may be oriented either perpendicular or tilted to the long axes of the molecules. Furthermore, the molecules may be arranged regularly within the layer (e.g., in a hexagonal arrangement), thus forming a three-dimensional lattice. As opposed to crystals, the smectic liquid crystalline state enables rotation of the molecules around their long axes. Different smectics may be distinguished in the basis of a variety of arrangements.

Phase transitions occur with increasing temperature, for example, crystalline to smectic C to smectic A to nematic to isotropic, or crystalline to nematic to isotropic. These examples demonstrate that not all possible transitions necessarily occur. Depending on the number of mesophases occurring, thermotropic mono-, di-, tri-, or tetra-morphism may be distinguished.

Discotic liquid crystals arise from disc-shaped molecules as nematic or cholesteric mesophases. Their

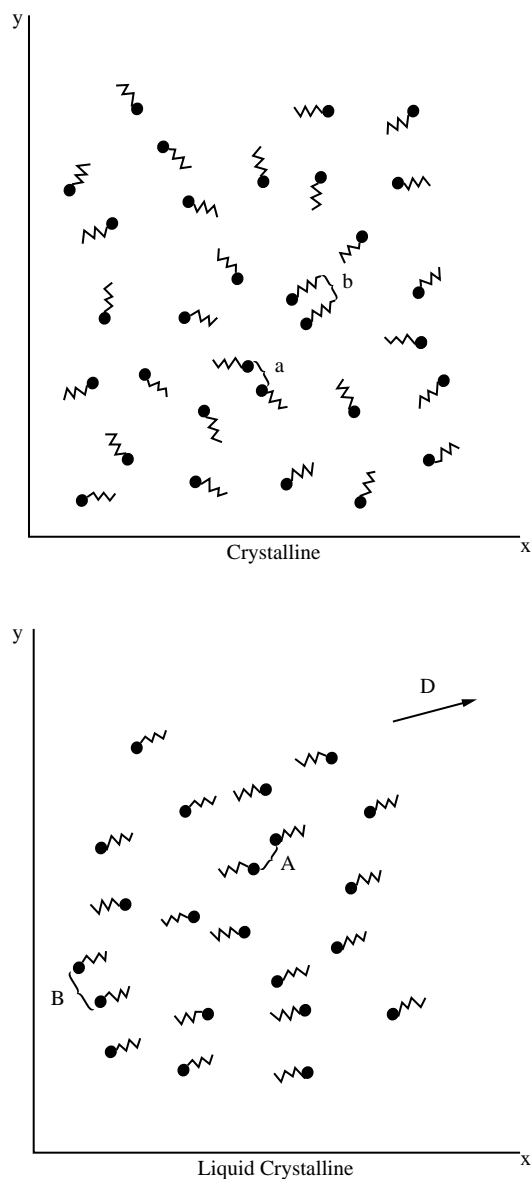


Fig. 1 Two-dimensional representation of short-range order (a, b) and long-range order (A, B) in the crystalline and liquid crystalline states. (From Ref. 2.)

structural characteristics are similar to that of their respective calamitic mesophases, that is, the normals of the discs are oriented parallel. Instead of the smectic mesophases, discotic columnar liquid crystals arise from stapling the discs one on the other. The columns of the discotic columnar mesophase form a two-dimensional lattice that is in either a hexagonal or a rectangular modification. In addition, the columns may or may not be tilted (Fig. 2f and 2g).

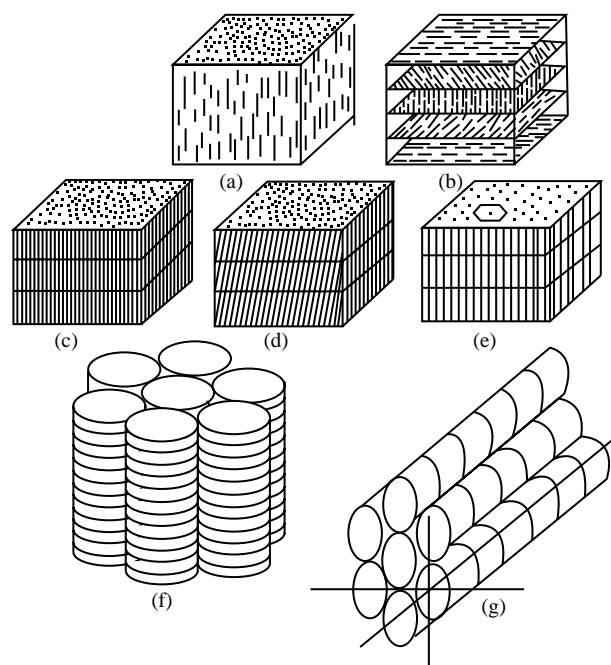


Fig. 2 Schematic representation of different calamitic and discotic thermotropic liquid crystals. (a) nematic; (b) cholesteric; (c–e) smectic; (f) columnar hexagonal; (g) columnar hexagonal tilted. (a–e: Adapted from Ref. 5; f and g: Adapted from Ref. 6.)

Lyotropic liquid crystals

Lyotropic liquid crystals differ from thermotropic liquid crystals. They are formed by mesogens that are not the molecules themselves but their hydrates or solvates as well as by associates of hydrated or solvated molecules. In presence of water or a mixture of water and an organic solvent as the most important solvents for drug molecules, the degree of hydration or solvation depends on the amphiphilic properties of a drug molecule. Hydration—and solvation—of the mostly rod-shaped molecule results in different geometries such as cone and cylinder (Fig. 3) (7).

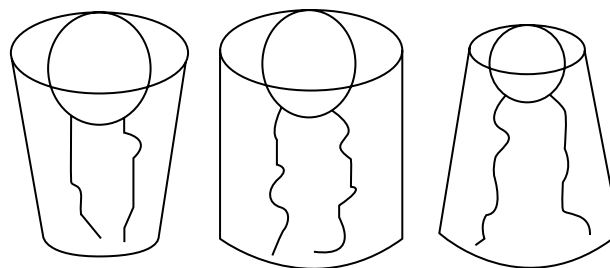


Fig. 3 Geometry of hydrated molecules—cylinders associate to a lamellar liquid crystal, cones to a hexagonal and an inverse hexagonal one. (Adapted from Ref. 7.)

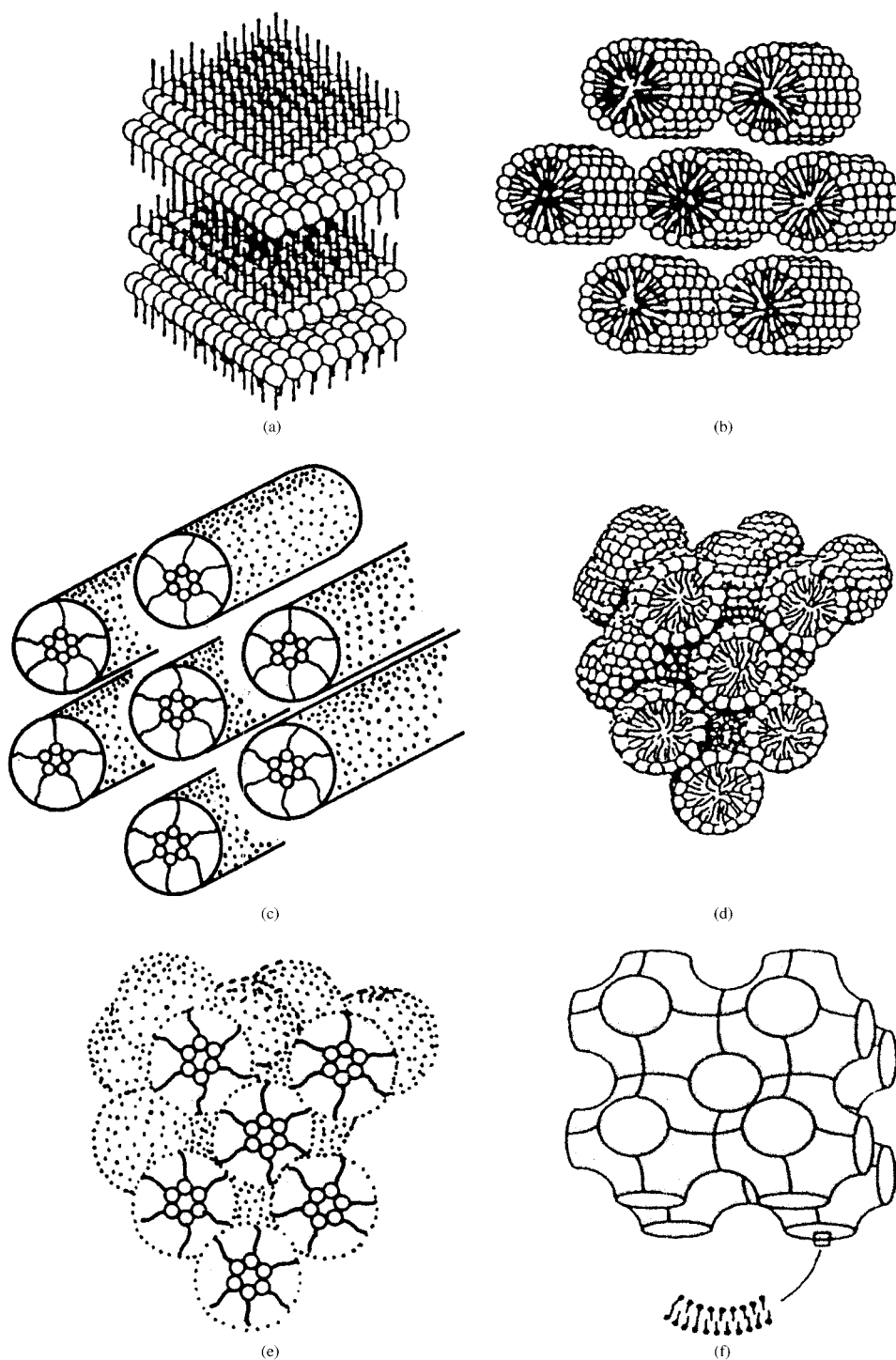


Fig. 4 Molecular structure of lyotropic liquid crystals. (a) lamellar; (b) hexagonal; (c) inverse hexagonal; (d) cubic type I; (e) inverse cubic type IV; (f) cubic type II. (a, b, and d: Adapted from Ref. 8; c: Adapted from Ref. 9; e: Adapted from Ref. 10; f: Adapted from Ref. 11.)

Cylinders arrange in layers; this results in a lamellar phase with alternating polar and nonpolar layers (Fig. 4a). Water and aqueous solutions can be included in the polar layers, resulting in an increase of layer thickness. Analogously, affinic molecules can be included in the nonpolar layers. In addition to the increased layer thickness of the lamellar phase, lateral inclusion between molecules is also possible with an increase in the solvent concentration, which transforms the rod shape of the solvated molecules to a cone shape (Fig. 3), thereby leading to a phase change. Depending on the polar or nonpolar character of the solvating agent and the molecule itself, the transition results in a hexagonal or an inverse hexagonal phase (Figs. 4b and 4c).

The hexagonal phase is named after the hexagonally packed rod micelles of solvated molecules, whereby their polar functional groups point either to the outside (Fig. 4c) or to the inside of the structure (Fig. 4, inverse hexagonal phase). In the hexagonal phase, the additional amount of water or unpolar solvent that can be included is limited. As the molecular geometry changes further during solvation, another phase transformation to a cubic form (type I) or inverse cubic form (type IV) takes place, consisting of spherical or ellipsoidal micelles and/or inverse micelles (Figs. 4d and 4e).

In addition to the cubic and/or inverse cubic forms described previously, further transitional forms exist between the lamellar phase and the hexagonal mesophase (cubic, type II) or inverse hexagonal mesophase (cubic, type III) (12). In contrast to the discontinuous phases of types I and IV, cubic mesophases of type II and type III belong to the bicontinuous phases (Fig. 4f). A range of lyotropic mesophases are possible, depending on the mesogen concentration, the lipophilic or hydrophilic characteristics of the solvent and the molecule itself (Table 1). However, not all theoretically possible mesophases may occur in practice.

Liposomes

With some molecules, a high concentration results in a lamellar phase, but no additional mesophases are formed if the concentration is reduced. The lamellar phase is dispersed in the form of concentric-layered particles in an excess of solvent (water or aqueous solution). This results in a vesicular dispersion. If the mesogenic material consists of phospholipids, the vesicular dispersion is called a liposomal dispersion (13). In principle, liposomes may be dispersed in oily continuous media too. However, the latter systems are of minor interest in drug formulation.

Liposomes consist of many or only few phospholipid bilayers, or just one bilayer (Fig. 5). Based on this, multilamellar vesicles (MLV), oligolamellar vesicles (OLV), small unilamellar (SUV), and large unilamellar vesicles (LUV) can be distinguished. Furthermore, multi-vesicular liposomes (MVL) may be formed.

The polar character of the liposomal core enables the encapsulation of polar drug molecules. Amphiphilic and lipophilic molecules are solubilized within the phospholipid bilayer depending on their affinity for the phospholipids. Participation of nonionic surfactants instead of phospholipids in the bilayer formation results in NiosomesTM. The term sphingosomes is suggested for vesicles from sphingolipids. However, nomenclature is not consistent; the term liposomes is used as a general term, although vesicles would be a better term.

A standard manufacturing procedure of liposomes is the film-forming method. Prior to film formation, the phospholipids are dissolved in an organic solvent. By rotational evaporation of the solvent, a thin multilayered film of phospholipids develops at the inner wall of the vessel. Redispersion of this film in water or aqueous buffer results in the formation of vesicles. The size of the vesicles and the number of bilayers vary. Hence, further manufacturing steps are needed to obtain defined vesicular dispersions with a sufficient shelf-life.

To reduce vesicle size and the number of bilayers, high-pressure filtration via polycarbonate membranes or high-pressure homogenization in a French press or in a microfluidizer are appropriate manufacturing procedures. Sonication may also be applied, although the obtained dispersion does not have the same particle sizes.

Alternatively, injection method and reverse-phase dialysis are appropriate procedures for the formation of SUV and LUV. Freeze-thaw procedures enable drug loading of the liposomes and also offer an evaluation of the stability of the vesicular dispersion. For interested readers, general reading is recommended (13, 14).

Liquid Crystal Polymers (LCP)

Both thermotropic and lyotropic liquid crystal polymers exhibit characteristic features with regard to their microstructure (15, 16). Anisometrical monomers such as rods or discs are connected to chains in an appropriate manner. These anisometrical monomers are considered to be the mesogens and may be part of main chain LCP, side chain LCP, or of both types together (Fig. 6). Between the mesogens are located flexible spacers of nonmesogenic character. Sufficient flexibility is a prerequisite for liquid crystal formation, with an increase in either temperature or solvent concentration.

Table 1 Possible transitions of lyotropic liquid crystals

Micellar	↔	Hexagonal	↔	Lamellar	↔	Inverse hexagonal	↔	Inverse micellar
Cubic I		Cubic II		Cubic III		Cubic IV		
Lipophilicity of the solvent and/or the amphiphilic compound (AC)				AC concentration				

(Modified from Ref. 12.)

METHODS FOR CHARACTERIZATION OF LIQUID CRYSTALS

Methods appropriate for the investigation and characterization of lyotropic liquid crystals are frequently used in drug development and may thus be employed in pharmaceutical laboratories. These methods are both macroscopic and microscopic.

Polarized Light Microscopy

Lyotropic liquid crystals except for cubic mesophases show birefringence just like real crystals do. Birefringence can be observed in a polarization microscope. Two polarizers in cross position are mounted below and above the birefringent object being examined. The cross position of the polarizers provides plane polarized waves perpendicular to each other. Therefore, the light passing the polarizer below an isotropic object cannot pass the polarizer in across position above the object. In an anisotropic material, some parts of the light are able to pass the second polarizer because the plane polarized beam has been rotated by an angle relative to the plane of the incoming beam.

Each liquid crystal shows typical black and white textures. The addition of an λ-plate with strong birefringent properties enables the observation of color effects of the textures in yellow, turquoise, and pink. Color effects arise because rotation of the plane polarized light depends on wavelength. The thickness of the λ-plate is suited for a wavelength of 550 nm. After leaving the plate, this wavelength swings in the same plane as the incoming polarized white light does. Therefore, it is totally absorbed by the second polarizer in the cross position. All the other wavelengths of the white light except for the 550-nm one are more or less rotated with respect to the polarization plane. Hence they pass the polarizer with various intensities. White light minus the 550-nm wavelength (green yellow) gives the impression of a pink color. With an additional

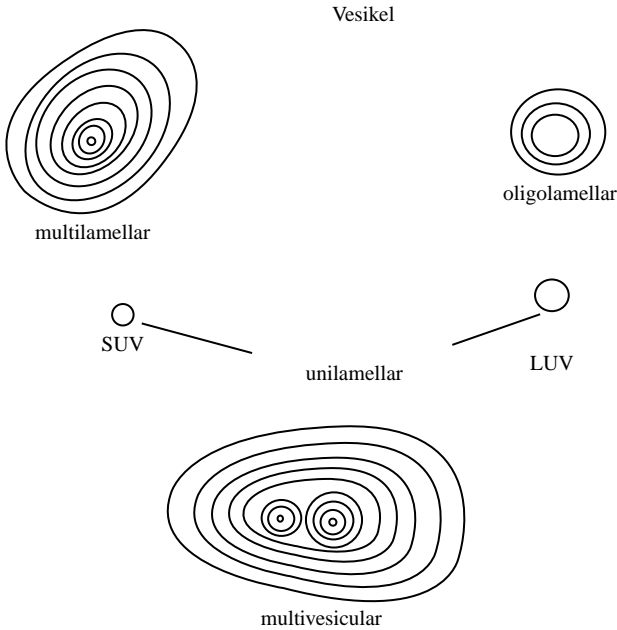


Fig. 5 Schematic cross sections of vesicles, each line represents a bilayer of hydrated molecules. (From Ref. 2.)

birefringent liquid crystalline material in the microscope, small deviations of the wavelengths being absorbed occur; thereby, turquoise and yellow textures can be observed.

Hexagonal mesophases can be recognized by their typical fan-shaped texture (Fig. 7a). Lamellar mesophases typically show oily streaks with inserted maltese crosses (Fig. 7b). The latter result from defects, so-called confocal domains, that arise from concentric rearrangement of plane layers. These defects prevail in some lamellar mesophases. Hence, no oily streaks occur but maltese crosses are the dominant texture (Fig. 7c).

The smectic mesophases of the thermotropic liquid crystals show a variety of textures but resemble the fan-shaped texture of the lyotropic hexagonal mesophase. More comprehensive literature is recommended for further reading (5).

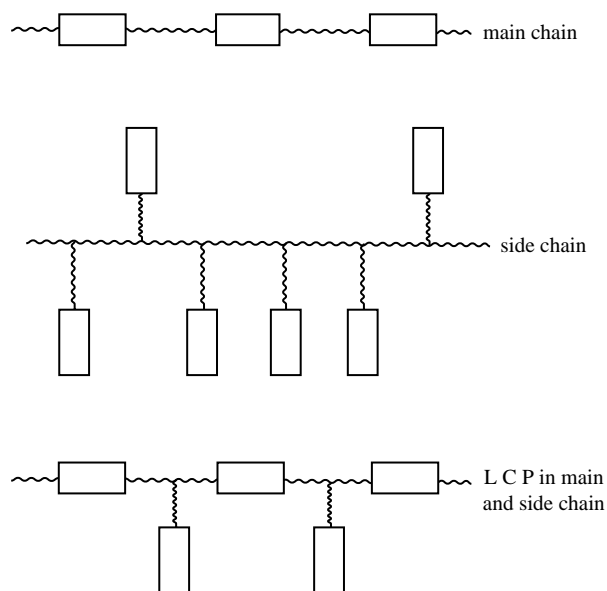
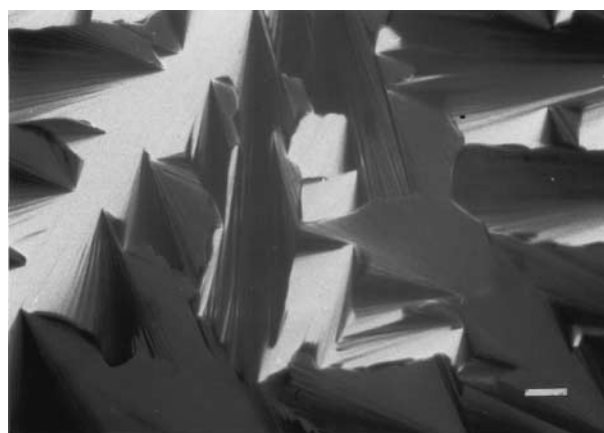


Fig. 6 Liquid crystal polymers with the mesogen within the main chain, the side chain and both the main and side chain, respectively. (From Ref. 2.)

Transmission Electron Microscopy (TEM)

Due to the high magnification power of the electron microscope, the microstructure of liquid crystals can be visualized. However, aqueous samples do not survive the high vacuum of an electron microscope without loss of water and thus their microstructure changes. Therefore, special techniques of sample preparation are necessary prior to electron microscopy. The freeze fracture technique has proven to be successful in this regard (Fig. 8). For this purpose, a replicum of the sample is produced and viewed in the electron microscope. To preserve the original microstructure of the sample during the replication, the first step is shock freeze the sample. For high freezing rates to ($10^5 - 10^6$ K/s), the sample is sandwiched as a thin layer between two gold plates and then shock frozen with either nitrogen-cooled liquid propane at -196°C or slush nitrogen at -210°C . If the temperature of the cooling medium is far below its boiling temperature, an efficient freezing rate can be obtained.

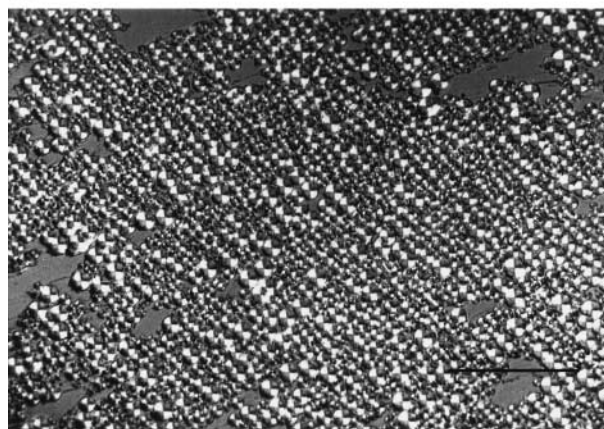
The frozen sample within the sample holder is transduced into the recipient of a freeze fracture apparatus, in which the fracture is performed at a temperature of -100°C and a vacuum between 10^{-6} and 5×10^{-7} bar. Within a homogeneous material, the fracture occurs by randomly because all structural elements have equal probabilities for fracturization. However, even a



(a)



(b)



(c)

Fig. 7 Polarized light micrographs of (a) hexagonal; (b) and (c) lamellar liquid crystals. Bar $50 \mu\text{m}$. (From Ref. 17.)

homogeneous material often consists of more or less polar areas. Within polar areas, stronger interactions via hydrogen bonds prevent the fracture; thus, fracture within polar areas is less probable than is fracture within apolar areas. Therefore, the sample profile obtained after

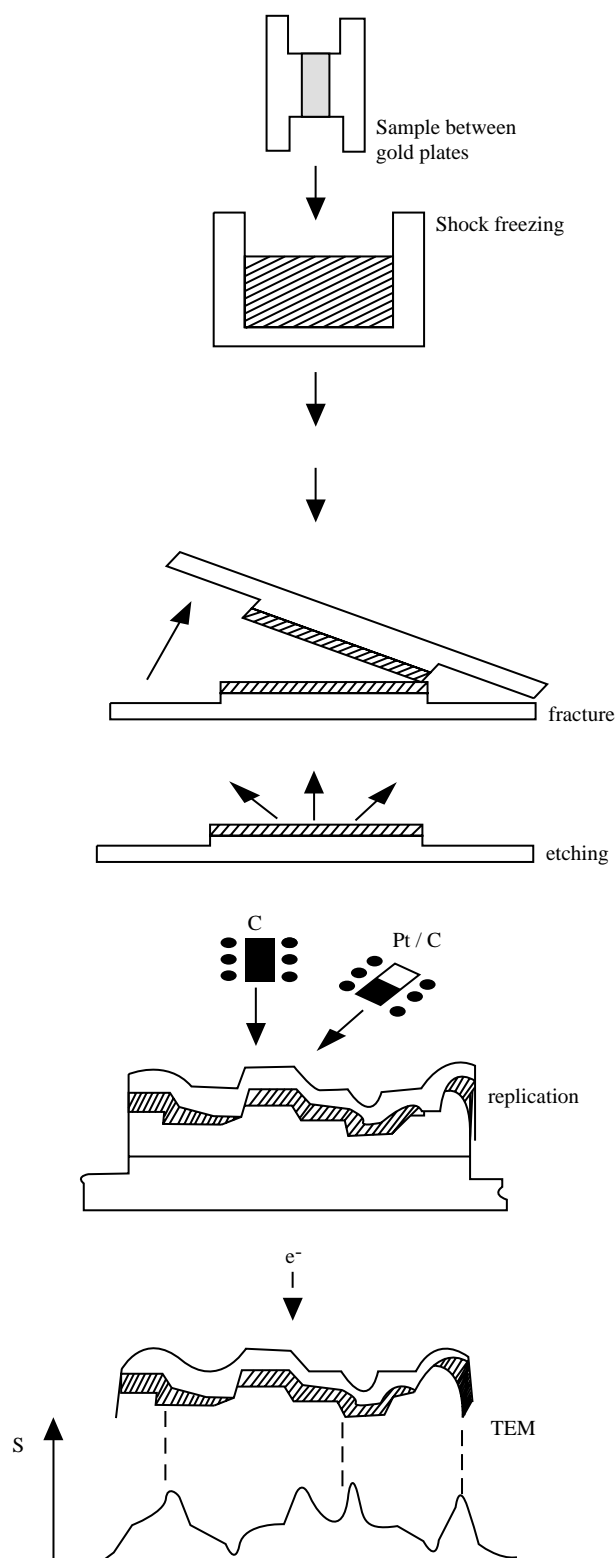


Fig. 8 Freeze fracture replication technique for transmission electron microscopy (TEM). (From Ref. 18.)

fracturization represents the microstructure of the sample just qualitatively and not quantitatively.

Immediate etching after freeze fracture provides sublimation of nonpermanent constituents (commonly ice), with the effect that level differences in the sample surface appear more pronounced.

Following this, the sample surface is shadowed with 2-nm thick platinum under an angle of 45° . Additional vertical shadowing with a 10-times thicker carbon layer of 20-nm platinum provides a high mechanical stability of the replicum, which means easier handling as regards removing, cleaning, drying, and finally observing in the transmission electron microscope (TEM).

The shadowing with platinum under an angle of 45° provides differences in contrast because platinum precipitation takes place preferably at sample positions that face the platinum source in luff whereas sample positions in lee are less or not shadowed. In the TEM, these different thicknesses of platinum absorb the electron beam to different extents, thus forming shadows. This phenomenon results in the formation of a plastic impression of the transmission electron micrographs of the replicum.

Fig. 9a–c represents transmission electron micrographs of different lyotropic liquid crystals after freeze fracture without etching. The layer structure of the lamellar mesophase, including confocal domains, hexagonal arrangement of the rod-like micelles within the hexagonal mesophase, and close by packed spherical micelles within the cubic liquid crystal can be clearly seen.

Fig. 9d and e shows aqueous dispersions of vesicles. The smaller the vesicle, the less probable is an upcoming cross fracture. Thus the question of whether the vesicle is uni- or multilamellar can probably not be answered. At least for the fluid vesicle dispersions, it is possible to solve the problem using cryo-TEM.

For this purpose, it is necessary to give sufficient contrast to a thin film of the frozen sample by using, for example, osmium tetroxide. Then the sample can directly be viewed in the TEM (at a temperature of -196°C). The adjustment of the temperature to -196°C provokes a very low vapor pressure, especially of water, so that the examination of the probe is possible by preservation of the microstructure despite the high vacuum. A disadvantage of cryo-TEM is the classification of vesicles according to their size. Due to the fluid property of vesicle dispersion prior to freezing, the thickness of the sample film varies from the center to the outside. Hence, smaller vesicles stay in the center, where the film is thin, whereas the larger ones linger at the outside margin in the thicker part of the film. In this outer part the vesicles escape detection. Hence the resulting distribution does not represent the actual size distribution.

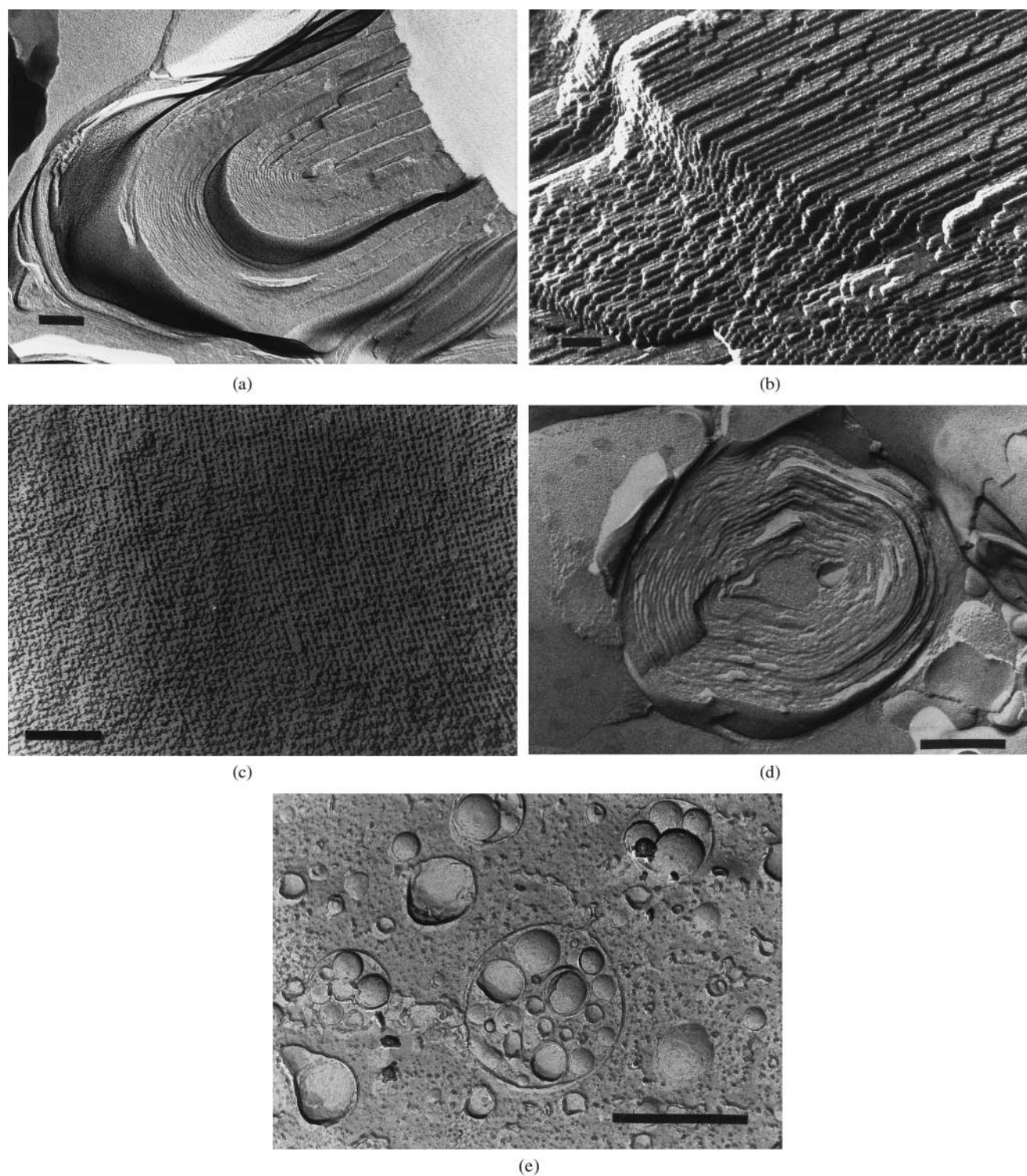


Fig. 9 Transmission electron micrographs of freeze fractured liquid crystals. (a) lamellar with confocal defects, bar 100 nm; (b) hexagonal, bar 100 nm; (c) cubic of type I, bar 100 nm; (d) multilamellar vesicle consisting of dodecyl-PEG-23-ether, cholesterol and water, bar 200 nm; (e) multivesicular vesicle, bar 1 μm . (a and b: Adapted from Ref. 19; c: Adapted from Ref. 20; d: Adapted from Refs. 21 and 22.)

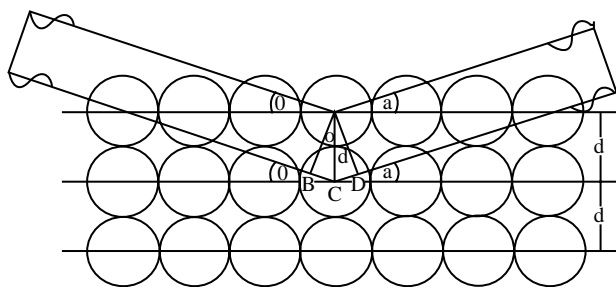


Fig. 10 Schematic representation of the reflection conditions according to Bragg's equation.

X-ray Scattering

With X-ray scattering experiments, characteristic interferences are generated from an ordered microstructure (23). A typical interference pattern arises due to specific repeat distances of the associated interlayer spacing d . By Bragg's equation, d can be calculated as $d = n(\lambda/2)\sin \vartheta$ where λ is the wavelength of the X-ray (e.g., 0.145 nm by using a copper anode or 0.229 nm by using a chromium anode), n is an integer and denotes the order of the interference and, ϑ is the angle under which interference occurs (i.e., reflection conditions are fulfilled). (See Fig. 10 for a further illustration).

Bragg's equation points at the inverse proportionality between d and ϑ . Large terms for d in the region of long-range order are registered by the small-angle X-ray diffraction technique (SAXD), whereas small terms for d in the region of short-range order are registered by the wide-angle X-ray diffraction technique (WAXD). SAXD is important for the exact determination of the distances of d of liquid crystalline systems. With WAXD, the loss of short-range order of liquid crystalline systems can be recognized in terms of the absence of interferences, which are characteristic of the crystalline state.

Interferences can be detected in two ways: (i) the film detection; and (ii) the registration of X-ray counts with scintillation counters or position-sensitive detectors.

However, SAXD does not only detect interferences from which the interlayer spacings can be calculated, but also enables to decide from the sequence of the interferences the type of liquid crystal (24, 25).

The sequence of the interferences for different liquid crystals is as follows:

Lamellar: $1 : 1/2 : 1/3 : 1/4 \dots$

Hexagonal: $1 : 1/\sqrt{3} : 1/\sqrt{4} : 1/\sqrt{7} \dots$

Cubic: $1 : 1/\sqrt{2} : 1/\sqrt{3} : 1/\sqrt{4} \dots$

Cubic: $1 : 1/\sqrt{4} : 1/\sqrt{5} : 1/\sqrt{6} \dots$

Differential Scanning Calorimetry (DSC)

Phase transitions go along with changes in energy content of the respective system. This phenomenon is caused by changing either the enthalpy ΔH or the entropy ΔS . Enthalpy changes cause endothermic or exothermic signals depending on whether the transition is due to consumption of energy (e.g., melting of a solid) or release of energy (e.g., recrystallization of an isotropic melt).

It should be mentioned that the transition from crystalline to amorphous requires much energy, whereas the transition from crystalline to liquid crystalline, from liquid crystalline to amorphous, and particularly the transition between different liquid crystals consume low amounts of energy. Therefore, care has to be taken about the appropriate sensitivity of the measuring device as well as on a sufficiently low detection limit (26).

Entropically caused phase transitions may be recognized by a change in baseline slope according to a change in the specific heat capacity. In particular, the phase transitions of liquid crystalline polymers result from entropic reasons, thus being considered transitions of the second order. These are usually called glass transitions. They can be overlaid from an enthalpic effect so that their detection might be complicated.

Rheology

Different types of liquid crystals exhibit different rheological properties (27, 28). With an increase in the microstructural organization of the liquid crystal, its consistency increases and the flow behavior becomes more viscous. The coefficient of dynamic viscosity η , although a criterion for the viscosity of just ideal viscous flow behavior (Newtonian systems), is rather high for cubic and hexagonal liquid crystals but fairly low for lamellar ones; however, the flow characteristics are not Newtonian but plastic for cubic and hexagonal crystals or pseudoplastic for lamellar ones.

For thermotropic liquid crystals, the viscosity increases in the following sequence:

nematic < smectic A < smectic C.

The low flowability of lyotropic liquid crystals such as cubic and hexagonal mesophases is due to their three-dimensional and two-dimensional order, respectively. Lamellar mesophases with one-dimensional long-range order have a fairly high flowability. Due to their gel character, cubic and hexagonal mesophases even exhibit a yield stress until flow occurs. Unlike the corresponding

inverse liquid crystals, the gel character is much more pronounced because of the interactions between polar functional groups located at the surface of the associates. Via polar interactions, for example hydrogen bonds, the associates may form strong networks with each other. On the other hand, the surface of the associates of inverse mesophases consists of apolar groups of the associated molecules. Thus the resulting interactions are less strong and the gel can get deformed more easily.

A mechanical oscillation measurement is the method of choice for determining the elasticity of liquid crystalline gels. Without applying a superposition of shear strain, the viscoelastic properties of liquid crystals may be studied without a change in network microstructure, which usually occurs in terms of mechanical deformation with rheological investigations. With the oscillation experiments, the viscoelastic character of cubic and hexagonal mesophases as well as that of lamellar mesophases and highly concentrated dispersions of vesicles (which also show viscoelastic behavior) can be quantified. A vesicle dispersion of low content of the inner phase, however, exhibits an ideal viscous flow property. According to the Einstein equation, η is larger than η_0 of the continuous phase, which is usually pure water or solvent, by the multi factor $2.5 \times$ volume ratio of the dispersed phase ϕ .

$$\eta = \eta_0(1 + 2.5\phi)$$

where η_0 = viscosity of pure solvent (i.e., the continuous phase) and, ϕ = volume ratio of the inner phase.

Determination of Vesicle Size by Laser Light Scattering

Vesicle size is an important parameter in not only in-process control but particularly quality assurance because the physical stability of the vesicle dispersion depends on particle size and particle size distribution. An appropriate and particularly quick method is the laser light scattering (for particle size) or diffraction (for particle size distribution). Laser light diffraction can be applied for particles $>1 \mu\text{m}$ and according to the diffraction theory of Fraunhofer, refers to the proportionality between intensity of diffraction and the square of particle diameter.

Rayleigh's theory holds for particles $<200 \text{ nm}$, which considers scattering intensity to be proportional to the sixth potency of the particle diameter. Both Fraunhofer's and Rayleigh's theories are only approximations of Mie's theory, which claims that scattering intensity depends on the scattering angle, absorption, and size of the particles

as well as on the refractive indices of both the particles and the dispersion medium. Unfortunately, the latter parameters are difficult to be determined. Furthermore, most vesicle dispersions consist of a dispersed mesophase with particle sizes $<200 \text{ nm}$ up to $1 \mu\text{m}$. Therefore, photon correlation spectroscopy (PCS) that is based on laser light scattering provides an appropriate method of investigation (29).

Dynamically raised processes in the dispersion, such as Brownian molecular motion, cause variations in the intensities of the scattered light with time, which is measured by PCS. Smaller the particle, higher the fluctuations by Brownian motion. Thus, a correlation between the different intensities measured is only possible for short time intervals. In a monodisperse system following first-order kinetics, the autocorrelation function decreases rather fast. In a half logarithmic plot of the autocorrelation function, the slope of the graph enables the calculation of the hydrodynamic radius by the Stokes–Einstein equation. With the commercial PCS devices the z -average is determined, which corresponds to the hydrodynamic radius.

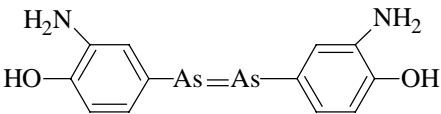
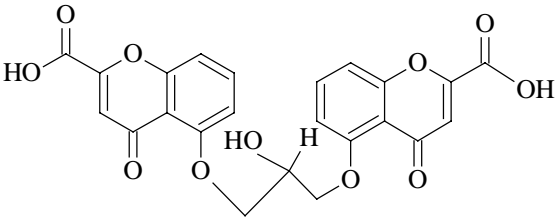
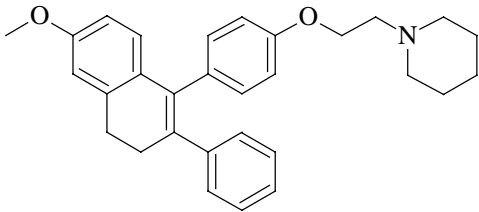
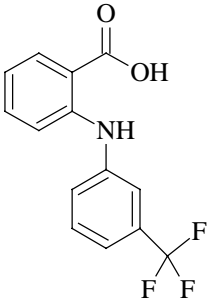
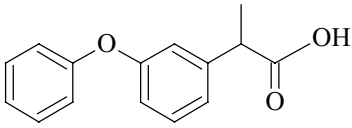
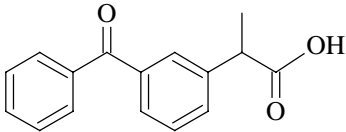
In a polydisperse system, the calculation of particle size distribution is possible, in addition, by using special transformation algorithms. For this, certain requirements need to be fulfilled: a spherical particle shape, sufficient dilution, and a large difference between the refractive indices of the inner and outer phase. As not all requirements can be usually fulfilled, the z -average as a directly accessible parameter is preferred to the distribution function depending on models.

APPLICATIONS OF LIQUID CRYSTALS IN DRUG DELIVERY

Liquid Crystalline Drug Substances

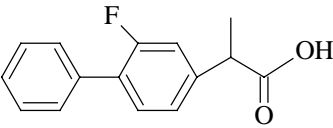
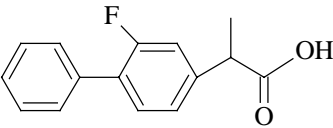
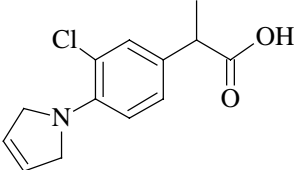
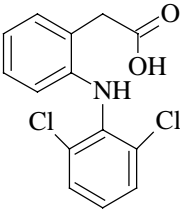
Some drug substances are able to form mesophases either with solvent or alone (30–37). In the latter case, an increase in temperature causes transition from the solid state to the liquid crystalline one. This is called thermotropic mesomorphism. Lyotropic mesomorphism occurs in combination with a solvent, usually water. Furthermore, a change in temperature may cause additional transitions. Thermotropic and/or lyotropic liquid crystalline mesophases of drug substances may interact with mesomorphous vehicles as well as with liquid crystalline structures in humans. Table 2 presents drug substances for which either thermotropic or lyotropic mesomorphism has been proven.

Table 2 Liquid crystalline drug substances

Drug	Type of liquid crystal	Formula	Reference
Arsphenamine	Nematic		30
Disodium cromoglicate	Nematic, hexagonal		31
Nafoxidin-HCl	Hexagonal, cubic, lamellar		32
Diethylammonium flufenamate	Lamellar		33
NSAID salts Fenoprofen	Lamellar		34
Ketoprofen	Lamellar		34

(Continued)

Table 2 Liquid crystalline drug substances (*Continued*)

Drug	Type of liquid crystal	Formula	Reference
Ibuprofen	Lamellar		34
Flurbiprofen	Lamellar		34
Pirprofen	Lamellar		34
Diclofenac	Lamellar		36
Peptide hormone	LH-RH analogue		35

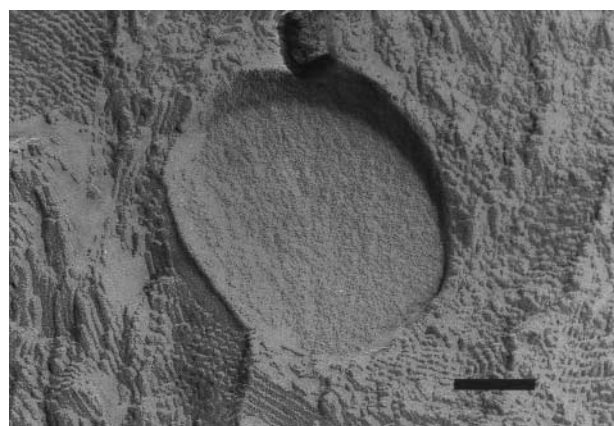
Arsphenamin was the first drug substance with thermotropic mesomorphism (30) to be therapeutically used as Salvarsan during the first half of the nineteenth century. The drug is effective against microorganisms and thus offered for the first time an efficient therapy of venereal diseases such as syphilis. Nowadays, it has been replaced by antibiotics with less serious side effects.

The molecular structure of arspenamin is typical of a thermotropic mesogen. With its symmetrical arrangement of atoms, the same holds for disodium cromoglicinate (DNCG) (30), which forms thermotropic liquid crystals and additionally lyotropic mesophases with water. If micronized DNCG powder is applied to the mucosa of the nose or the bronchi, the powder will absorb water from the high relative humidity of the respiratory tract, and first transform into a lyotropic mesophase and then into a solution.

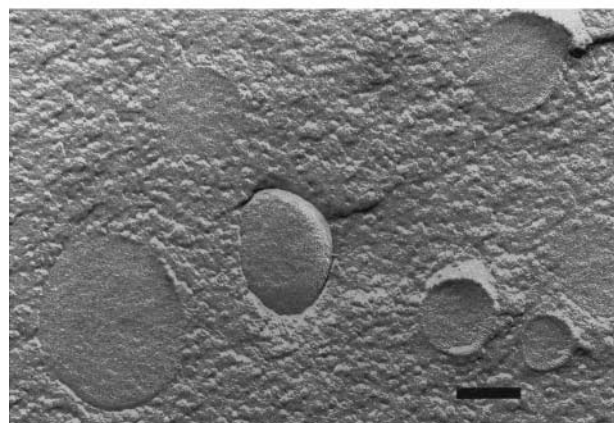
DNCG serves as a mast-cell stabilizer. Mast cells are located on the mucosa of the respiratory tract and act by

releasing the mediator substance histamine on contact with an allergen, provided the patient had been sensitized previously. Due to its mast-cell stabilizing effect, DNCG acts as a prophylactic against allergic reactions associated with asthma and hay fever. In addition to this prophylactic effect, DNCG exhibits a second mode of action in causal therapy of asthma, which has not yet been fully clarified. According to recent findings, DNCG has a positive effect on the inflammation of the mucosa of bronchi.

For therapeutic purposes, a similar frequently used group of drug compounds is the nonsteroidal anti-inflammatory drugs (NSAIDs). One of the best known representatives of the aryl acetic acid derivatives is diclofenac and that of aryl propionic acid derivatives is ibuprofen. As both have acidic properties, they dissociate while being dissolved and may form salts with amphiphilic properties. Together with appropriate



(a)



(b)

Fig. 11 Transmission electron micrographs of freeze fractured oily droplets dispersed in (a) a hexagonal and (b) a cubic liquid crystalline phase, bar represents 100 nm. (From Ref. 38.)

counter ions, these amphiphilic organic acids may form lyotropic mesophases with water at even room or body temperature e.g., diclofenac diethylamine or ibuprofen lysinate (34, 36). Furthermore, some anhydrides of NSAID, e.g., fenoprofen calcium (37), exhibit thermotropic mesomorphism after thermal dehydration of the crystalline salt.

All the other drugs substance listed in Table 2 have not yet been used for therapeutical purposes.

Liquid Crystalline Formulations for Dermal Application

As long as drug molecules with amphiphilic character form lyotropic mesophases, amphiphilic excipients in drug formulations form lyotropic liquid crystals. Especially surfactants, which are commonly used as emulsifiers in dermal formulations, associate to micelles after dissolution

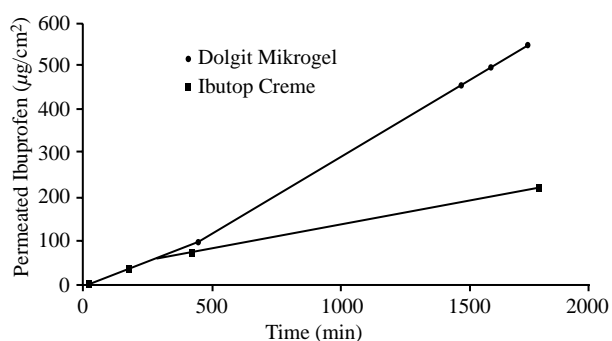


Fig. 12 Permeability of ibuprofen from different formulations via excised human stratum corneum. (Redrawn from Ref. 39.)

in a solvent. With increasing concentration of these micelles, the probability of interaction between these micelles increases, forming liquid crystals.

Surfactant gels

The use of monophasic systems of lyotropic liquid crystals is relatively seldom and is limited to gels. A variety of polar surfactants (e.g., ethoxylated fatty alcohols) are hydrated in presence of water and form spherical or ellipsoidal micelles. At high surfactant concentrations, these associates are densely packed and are thus identified as cubic liquid crystals (20).

Fig. 9c represents a transmission electron micrograph of a liquid crystalline surfactant gel of this type. Such gels are optically transparent. If agitated mechanically, their elastic properties become evident. Due to resonance effects in the audible range, they are also called ringing gels. The lipophilic components are solubilized together with the active ingredients in hydrated associates of the surfactants. However, the solubilization capacity for lipophilic components is generally limited. By exceeding this capacity, the excess of the lipophilic component will be dispersed dropwise in the liquid crystalline phase (Fig. 11b). Such systems exhibit a white appearance according to the change in refractive index at the interface between continuous liquid crystalline and dispersed oil phase. Besides, the dispersed drops are mechanically stabilized because the liquid crystalline phase of either hexagonal (Fig. 11a) or cubic character (Fig. 11c) has a high yield stress.

Ringing gels with cubic liquid crystalline microstructure are used as commercial drug formulations especially for topical NSAID formulations. Examples in the German market include Contrheuma Gel Forte N, Trauma-Dolgit Gel, and Dolgit Mikrogel. Dolgit Mikrogel was introduced in 1996 and contains ibuprofen as an active ingredient. On the one hand the high surfactant concentration of such gels

is necessary to verify the liquid crystalline microstructure; on the other hand, this concentration influences the microstructure of the stratum corneum lipids via increase in permeability. This effect is also achieved by alcohol, which is also solubilized in the formulation. Fig. 12 shows the result of permeation tests with excised human stratum corneum. The amount of ibuprofen permeating per unit time and surface area is much higher for Dolgit Mikrogel than for an aqueous mixed micellar solution of the drug. Although relatively high permeation rates are possible for the liquid preparation, the commercial formulation is significantly more effective because the high surfactant content and the alcohol favor high permeability.

A ringed surfactant gel of liquid crystalline microstructure containing the antimycotic bifonazole (Bifomyk gel) was introduced in 1995 into the German market. Similar to surfactant gels containing NSAID, an improved penetration of the active ingredient is desired in antifungal therapy of the dermis as well. However, because the liquid crystal structure only forms with a relatively high surfactant concentration, the positive effect of improved penetration must be considered together with the potential of irritation. The objective is to achieve improved penetration with minimum irritation, via a change in skin structure. Because hyphae fungal (mycelium) can penetrate deep into the epidermal layers by sliding past corneocytes of the horny layer, improvement of antimycotic therapy is of particular importance. The same holds true for penetration of NSAIDs through several epidermal layers because they have to arrive at the deeply located muscle and joint tissue.



Fig. 13 Transmission electron micrograph of a freeze fractured w/o cream. The aqueous phase is dispersed as droplets within the continuous lipophilic phase and the interface consists of multiple bilayers of hydrated surfactant molecules; bar 500 nm. (From Ref. 41.)

Ointments and creams

Usually the surfactant concentration in ointments and creams is significantly lower than in surfactant gels. Ointments are nonaqueous preparations, whereas creams result from ointments by adding water. The microstructure of both ointments and creams may consist of liquid crystals, as long as a liquid crystalline network or matrix is formed by amphiphilic molecules. In a liquid crystalline matrix, it is easier to deform the system by shear; such formulations show plastic and thixotropic flow behavior on shear. In comparison to systems with a crystalline matrix which are usually destroyed irreversibly by shear, those with a liquid crystalline matrix exhibit a short regeneration time of the sheared matrix. To obtain a liquid crystalline matrix, amphiphilic surfactants, that form lyotropic liquid crystals at room temperature have to be selected. Preferably, lamellar liquid crystals that are able to solubilize high amounts of further ingredients and spread through the whole formulation as a network-forming crosslinked matrix should be formed. In contrast, ointments that contain long-chain fatty alcohols such as cetyl and/or stearyl alcohol have a crystalline structure at room temperature (40).

Although the so-called α -Phase of the fatty alcohols—a thermotropic type smectic B liquid crystal with hexagonal arrangement of molecules within the double layers—is initially formed from the melt during the manufacturing process, it normally transforms into a crystalline modification as it cools. However, the crystallization of the gel matrix can be avoided if the α -Phase can be kept stable as it cools to room temperature. This can be achieved by combining appropriate surfactants such as myristyl or lauryl alcohol and cholesterol, a mixture of which forms a lamellar liquid crystal at room temperature (41). Due to depression of the melting point, the phase transition temperature of crystalline to liquid crystalline as well as liquid crystalline to isotropic decreases. Therefore, a liquid crystalline microstructure is obtained at room temperature.

The polar character of a surfactant molecule enables the addition of water to form creams. Depending on whether the surfactant or the surfactant mixture has a strong or weak polar character, creams of type o/w or w/o are formed. Creams of w/o type are produced from systems that are stabilized solely with weakly polar surfactants such as fatty alcohols, cholesterol, glycerol monostearate, or sorbitan fatty acid esters. The surfactant or surfactant mixtures are adsorbed at the interface of the dispersed aqueous and the continuous lipophilic phase. Even multiple layers of the surfactant will be adsorbed if the concentration of mesogenic molecules is

high enough to form their own liquid crystalline phase (Fig. 13). Apart from the reduction of surface tension and/or surface energy, the liquid crystalline interface also has a mechanically stabilizing effect on the emulsion drops.

Surfactants such as sulfated fatty alcohols may be hydrated to a higher extent than the fatty alcohols alone, and thus stabilize o/w emulsions. The combination of an anionic and a nonionic surfactant has proven to be particularly effective, as the electrostatic repulsion forces among the ionic surfactant molecules at the interface are reduced by the incorporation of nonionic molecules, thereby improving emulsion stability. The combination of cetyl/stearyl sulfate (Lanette E) and cetyl/stearyl alcohol (Lanette O) to yield an emulsifying cetyl/stearyl alcohol (Lanette N) is an example of this approach. The polar properties of this surfactant mixture are dominant; therefore, o/w creams are formed. In contrast to w/o systems, the stabilizing effect of the surfactant mixture is not mainly due to adsorption at the interface. Instead, the mixed surfactants are highly hydrated and form a lamellar network, which is dispersed throughout the continuous aqueous phase, whereas the dispersed lipophilic components are immobilized within the gel network. However, this hydrated gel matrix is not crystalline at room temperature as are the corresponding w/o creams with cetyl/stearyl alcohol, but is in its α -phase, which belongs to the thermotropic smectic liquid crystals and exhibits a strong similarity to lyotropic lamellar liquid crystals.

Analogous gel matrices of liquid crystalline lamellar phases can also be formed with nonionic mesogens, for example, with the combination of cetyl/stearyl alcohol and

ethoxylated fatty alcohol, provided the hydrophilic and lipophilic properties of the surfactant molecules are more or less balanced to favor the formation of lamellar structures.

Liposome Dispersions

Although liposomes have been extensively studied since 1970, only a few commercial drug formulations contain liposomes as drug carriers (42, 43). The first commercial drug formulation with liposomes for topical administration was registered in Italy. The antimycotic econazol was encapsulated in liposomes being dispersed in a hydrogel (Ecosom Liposomengel, formerly Pevaryl Lipogel). Due to the formation of a highly hydrated gel network of the hydrophilic polymers, liposomes are immobilized within the gel network and thus mechanically stabilized. This stabilization via gelation of the continuous aqueous phase can also be applied to other dispersion systems (e.g., suspensions or emulsions). An example of such an emulsion/hydrogel combination that contains heparin sodium as an active ingredient and liposomes as the additional dispersed phase (the latter only since 1995) is Heparin Liposom. A formulation with an analogous emulsion/hydrogel combination but without additional liposomes is Voltaren Emulgel. The transmission electron micrograph (Fig. 14) reveals an adsorption of lamellar liquid crystals at the interface of dispersed oil drops and the aqueous continuous phase. The aqueous continuous phase is again a hydrogel based on polyacrylate in which the lipophilic phase is immobilized. The interface consists of multilamellar layers consisting of both surfactant and drug molecules. Thus the hydrogel is not only stabilized by the hydrogel network itself but also by the liquid crystalline interface, which provides an additional stabilization. The active ingredient diclofenac diethylamine diffuses slowly from the dispersed phase via the multilamellar interface into the continuous phase, from where it penetrates into the epidermis.

Similar to Voltaren Emulgel, oily droplets of an eutectic mixture of lidocaine and prilocaine are dispersed in a hydrogel to provide local anaesthesia of the skin for injections and surgical treatment (Emla cream). A further possibility is the dermal administration of a liposome dispersion as a spray (Heparin PUR ratiopharm Sprühgel). After administration, water and isopropyl alcohol evaporate partially to result in an increase of concentration and thereby in a transition from the initial liposome dispersion to a lamellar liquid crystal (45). The therapeutic effect thus appears to be influenced favorably by the presence of lecithins alone, rather than by the degree of dispersion of liposomes.

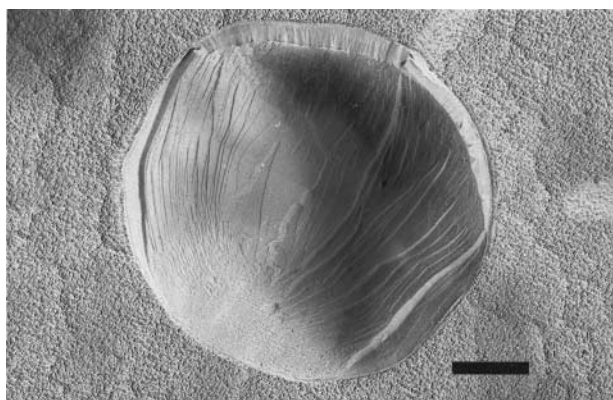


Fig. 14 Transmission electron micrograph of Voltaren Emulgel. The interface between the continuous hydrogel and the dispersed emulsion droplets consists of multiple bilayers of hydrated surfactant molecules bar 500 nm. (From Ref. 44.)

Liposome dispersions for parenteral administration

Depending on their size and surface charge, parenterally administered liposomes interact with the reticulo endothelial system (RES) and provoke an immunological response. After being marked by the adsorption of certain serum proteins, called opsonins, they are identified as an invader and destroyed by specific immune cells mainly in the liver, spleen, and bone marrow.

This passive drug targeting enables an efficient therapy of diseases of these organs or their affected cells. Clinical tests in the therapy of parasitic diseases of the liver and spleen have proven most efficient by having encapsulated the drug substance in liposomes. Apart from passive drug targeting, drug encapsulation within liposomes offers a modification of the therapeutic effect in terms of intensity and duration, together with a minimization of undesired side effects. For this purpose, liposomes have to circulate for as long as possible in the vascular system and remain unrecognized by phagocytic cells.

The antimycotic amphotericine is encapsulated in liposomes and marketed as AmBisome to treat severe systemic mycosis. The liposomal encapsulation reduces the toxicity of amphotericine while increasing half-life of the drug and plasma level peaks (43). Due to stability reasons, the parenteral formulation is a lyophilized powder, which has to be reconstituted by adding the solvent just before administration.

The cytostatic daunorubine, which is administered in the later state of Kaposi sarcoma of AIDS patients, is encapsulated in liposomes of about 45 nm size (43). The liposome dispersion is marketed as a sterile, pyrogen-free concentrate (DaunoXome) and has to be diluted with a 5% glucose solution just before being administered as an infusion. Although daunorubicine by itself is cardiotoxic, the liposomal formulation attacks cardiac tissue only insignificantly, but strongly affects the tumor cells by being taken up preferably. It is postulated that small unilamellar vesicles (SUV) may pass through endothelial gaps in recently formed capillaries of the tumor, thereby entering the tumor tissue. At this site, the drug is released from the liposomal carrier and inhibits the proliferation of the tumor cells.

Liposome dispersions for an installation into the lung

A liposomal formulation consisting of a surfactant, which usually coats the mucosa of the bronchi and prevents a collapse of the alveolar vesicles of the lung, has been developed for patients who suffer either from infant

respiratory distress syndrome (IRDS) or adult/acquired respiratory distress syndrome (ARDS). IRDS often affects prematurely born babies who have not yet developed a functional lung surfactant and therefore develop a failure in pulmonary gas exchange. ARDS is also a life-threatening failure/loss of the lung function, usually acquired by illness or accident. Clinical trials with liposomal surfactant have proven to be efficient in prophylactic treatment of IRDS and ARDS.

The surfactant is obtained by extraction from the lungs of cattle, by washing and centrifuging several times. This raw extract is treated with appropriate organic solvents, sterilized by filtration, dried by solvent evaporation under aseptic conditions, resuspended in water, and finally homogenized in a French press under cooling. Care has to be taken to maintain sterility of the extract during all procedures. Special attention has to be paid to transmissible spongiform encephalopathies (TSE). The whole manufacturing process has been validated in terms of a decrease of infectious material by a factor of 10^{21} , although a factor of 10^8 is sufficient. The result is a formulation (Alvefact) that is considered safe in context of TSE and viruses, and it also contains all relevant components of a lung surfactant in terms of pulmonary exchange of gas (43).

Transdermal Patches

To obtain a systemic effect via percutaneous penetration of a drug compound, a high permeability through the stratum corneum and the living tissue beneath is required as well as a high potency of the drug for a low dose to be administered. For an additional short biological half-life, the development of controlled release transdermal systems is a good choice.

Transdermal patches are high-tech devices, which contain the drug substance in a reservoir from which the drug is released in a controlled manner (i.e. zero-order kinetics). The control element is either a membrane or a matrix. Membrane-controlled patches were the first to be marketed. A major disadvantage of these is the so-called dose dumping that occurs in case of a membrane damage during handling. To ensure the desired drug control, even liquid crystalline polymers have been examined with regard to their usefulness in membrane-controlled transdermal patches (46). The matrix-controlled transdermal patch consists of only one functional element, the porous polymer matrix, which not only controls drug release but simultaneously acts as a drug reservoir and adhesive element.

Transdermal patches are marketed worldwide with the drug substances glycerole trinitrate, estradiol,

testosterone, clonidine, scopolamine, fentanyl and nicotine, respectively. The patch has to remain for up to one week at the appropriate body site. In this case the drug amount in the reservoir is rather high. As liquid crystalline vehicles with lamellar microstructure have high solubilization capacities, they are recommended as reservoirs for transdermal patches (47), although the high surfactant concentration of the lamellar liquid crystal might have a irritate the skin. Especially in terms of the membrane-controlled patch, the liquid crystalline vehicle is not in direct contact with the skin and thus will not exhibit an irritating effect on the skin.

Sustained Drug Release from Solid, Semisolid, and Liquid Formulations

The therapy of a chronic disease requires a repeated dosing of a drug. Drugs having a short biological half-life have to be administered up to several times daily within short intervals. To reduce the application frequency, sustained formulations have been developed. Liquid crystalline excipients are appropriate candidates for this because in a liquid crystalline vehicle the drug diffusion is reduced by a factor of 10 to 1000 in comparison with a liquid vehicle such as a solution (48–50). The factor depends on the kind of liquid crystal being employed.

Solid formulations

Solid formulations for sustained drug release may contain mesogenic polymers as excipients. The mesogenic

polymers form a matrix, which is usually compressed into tablets. Some of the most frequently used excipients for sustained release matrices include cellulose derivatives, which behave like lyotropic liquid crystals when they are gradually dissolved in aqueous media. Cellulose derivatives such as hydroxy-propyl cellulose or hydroxy-propylmethyl cellulose form gel-like lyotropic mesophases in contact with water (51), through which diffusion takes place relatively slowly. Increasing dilution of the mesophase with water transforms the mesophase to a highly viscous slime and then to a colloidal polymer solution.

Semisolid formulations

The solubilization of a drug substance in monophasic liquid crystalline vehicles results in semisolid formulations, which are preferably used for topical application. (See the sections on Surfactant Gels and Transdermal Patches.)

Liquid formulations

Sustained release from disperse systems such as emulsions and suspensions can be achieved by the adsorption of appropriate mesogenic molecules at the interface. The drug substance which either forms the inner phase or is included in the dispersed phase cannot pass the liquid crystals at the interface easily and thus diffuses slowly into the continuous phase and from there further into the organism via the site of application. Such as sustained drug release is especially pronounced in the case of multilamellar liquid crystals at the interface.

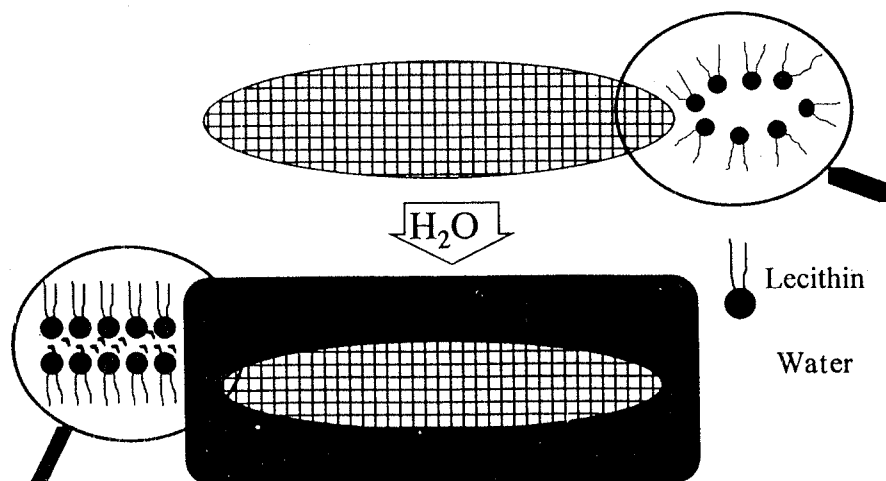


Fig. 15 Application induced transformation of a reverse micellar solution into a liquid crystal on contact with aqueous media. (From Ref. 10.)

A further possibility is the formation of liquid crystals on contact with body fluids at the site of application. The applied drug solution interacts with body fluids such as plasma, tear, fluid, or skin lipids and undergoes a phase transition to a mono- or multiphasic system of liquid crystals (Fig. 15). For example oily solutions of reverse micellar solutions of phospholipids, which solubilize any additional drug, transform into liquid crystalline lamellar phases by the absorption of water when applied to the mucosa. Drug release is controlled by the liquid crystals because diffusion within the liquid crystalline phase is slowest and thus rate-controlling (50). This principle can be used for ophthalmological administration as well as for nasal, buccal, rectal, vaginal, or even parenteral subcutaneous application (52). However, the peroral administration of such reverse micellar solutions either directly or encapsulated within soft gelatin capsules is not recommended (53) because the sustained release effect is limited by interindividual variations in digestion, such as the amount and composition of the gastric fluid as well as its ability to emulsify and solubilize in terms of enteral absorption.

For the treatment of paradontitis of infected gum pockets, the chemotherapeutic metronidazol has proven to be effective. The crystalline prodrug metronidazolbenzoate, which has to provide the active metronidazol through dissolution and hydrolysis, is suspended in an oleogel (Elyzol Dentalgel). The oleogel consists of glycerol monooleate and sesame oil, which are immobilized within the matrix structure of the surfactant. The base melts at body temperature and spreads evenly over the inner surface of the gum pockets. The molten system absorbs water and transforms into a reverse hexagonal phase. This liquid crystalline structure has a high viscosity. The resulting system adheres well to the surface of the mucosa and releases the active ingredient slowly (10).

Liquid Crystals in Cosmetics

Liquid crystals are mainly used for decorative purposes in cosmetics. Cholesteric liquid crystals are particularly suitable due to their iridescent color effects, and find applications as coloring for nail varnishes, eye shadows, and lipsticks. The structure of these thermotropic liquid crystals changes with body temperature, resulting in the required color effect. In recent times, such thermotropic cholesteric liquid crystals have also been included in body care cosmetics, where they are dispersed in a hydrogel. Depending on whether this dispersion in the hydrogel involves stirring or a special spraying process, the

iridescent liquid crystalline particles are distributed statistically in the gel (Estée Lauder Time Zone Moisture Recharging Complex) or concentrated locally (Vichy Restructure Contour des Yeux) to give the formulation the required appearance. Tests of the cosmetic efficiency of the liquid crystalline constituents have not yet been published.

REFERENCES

1. Kelker, H.; Hatz, R. *Handbook of Liquid Crystals*. Verlag Chemie; Weinheim, Germany, 1980.
2. Müller-Goymann, C.C. Flüssigkristalline Systeme in der Pharmazeutischen Technologie. *PZ Prisma* **1998**, *5*, 129–140.
3. Reinitzer, F. Beiträge zur Kenntnis des Cholesterins. *Monatsh. Chem.* **1888**, *9*, 412–441.
4. Lehmann, O. Über fließende Kristalle. *Z. Physikal. Chem.* **1889**, *4*, 462–471.
5. Demus, D.; Richter, L. *Textures of Liquid Crystals*; Chemie: Weinheim, Germany, 1978.
6. Eidenschenk, R. Flüssige Kristalle. *Chem. Unserer Zeit* **1978**, *18*, 168–176.
7. *The Physical Chemistry of Membranes*; Silver, B., Ed.; Allen Unwin Inc. & Solomon Press: Winchester, USA, 1985.
8. Brown, G.H.; Wolker, J.J. *Liquid Crystals and Biological Structures*; Academic Press: New York, 1979.
9. Friberg, S.E. *Food Emulsions*; Marcel Dekker, Inc.: New York, 1976.
10. Müller-Goymann, C.C. Anwendung lyotroper Flüssigkristalle in Pharmazie und Medizin. *Lyotrope Flüssigkristalle: Grundlagen, Entwicklung, Anwendung*; Stegemeyer, H., Ed.; Steinkopff Verlag: Darmstadt, Germany, 1999; 141–167.
11. Larsson, K. Structure of Isotropic Phases in Lipid-Water Systems. *Chem. Phys. Lipids* **1972**, *9*, 181–195.
12. Tiddy, G.J.T. Surfactant-Water Liquid Crystal Phases. *Phys. Rep.* **1980**, *57*, 1–46.
13. Gregoriadis, G. *Liposome Technology*; CRC Press Inc.: Boca Raton, USA, 1993; I–III.
14. Lill, N.; Krempel, H. Liposomen in Pharmazie und Kosmetik Teil 1. Struktur und Herstellungsverfahren. *PZ Prisma* **1996**, *3*, 262–267.
15. Blumstein, A., Ed. *Liquid Crystalline Order in Polymers*; Academic Press: New York, 1978.
16. Ciferri, A.; Kriegbaum, W.R.; Meyer, R.B., Eds. *Polymer Liquid Crystals*; Academic Press: New York, 1982.
17. Müller-Goymann, C.C. Flüssigkristalline Arzneimittel. *Pharmazeutische Technologie: Moderne Arzneiformen*, 2. erw. Auflage; Müller, R.H., Hildebrand, G.E., Eds.; Wissenschaftliche Verlagsges.: mbH: Stuttgart, Germany, 1998; 219–242.

18. Heering, W. *Die Struktur des Gelgerüsts der Wasserhaltigen Hydrophilen Salbe DAB 8- Anwendung der Gefrierbruchätztechnik und TEM auf kolloide pharmazeutische Zubereitungen*; Thesis Technische Universität: Braunschweig Germany, 1985.
19. Mueller-Goymann, C. Liquid Crystals in Emulsions, Creams and Gels, Containing Ethoxylated Sterols as Surfactant. *Pharm. Res.* **1984**, *1*, 154–158.
20. Schuetze, W.; Mueller-Goymann, C.C. Mutual Interactions between Nonionic Surfactants and Gelatin—Investigations in Cubic Liquid Crystalline Systems and Micellar Systems. *Colloid Polym. Sci.* **1992**, *269*, 85–90.
21. Usselman, B. *Beitrag zur Strukturaufklärung topischer Zubereitungen mit Fettalkoholpolyethylenglykolethern und Cholesterol als Tensiden*; Thesis Technische Universität: Braunschweig, Germany, 1987.
22. Schütze, W. *Diffuse Röntgenkleinwinkelstreuung an kolloidalen Drug Delivery Systemen*; Thesis Technische Universität: Braunschweig Germany, 1998.
23. Fontell, K. *X-ray Diffraction by Liquid Crystals-Amphiphilic Systems. Liquid Crystals and Plastic Crystals*; Gray, G., Winsor, P., Eds.; Ellis Horwood: Chichester, 1974; 2.
24. Luzzati, V.; Mustacchi, H.; Skoulios, A.; Husson, F. La Structure des Colloides d'association. I. Les phases liquide-Cristalline des Systemes Amphiphile-eau. *Acta Cryst.* **1960**, *13*, 660–677.
25. Fontell, K.; Mandell, I.; Ekwall, P. Some Isotropic Mesophases in Systems Containing Amphiphilic Compounds. *Acta Chem. Scand.* **1968**, *22*, 3209.
26. Shin, S.; Kumar, S.; Finotello, D.; Keast, S.; Neubert, M. High-Precision Heat Capacity Study of Phase Transitions in a Lyotropic Liquid Crystal. *J. Am. Phys. Soc.* **1992**, *45*, 8683–8692.
27. Roux, D.; Nallet, F.; Diat, O. Rheology of Lyotropic Lamellar Phases. *Europhys. Lett.* **1993**, *24*, 53–58.
28. Kohler, H.; Strnad, J. Evaluation of Viscosity Measurements of Dilute Solutions of Ionic Surfactants Forming Rod-shaped Micelles. *J. Phys. Chem.* **1990**, *94*, 7628–7634.
29. Müller, B.W.; Müller, R.H. Bestimmung von mittleren Durchmessern und Größenverteilungen an Teilchen im submikroskopischen Bereich mit der Photonenkorrelations-spektroskopie. *Pharm. Ind* **1983**, *45*, 1150–1153.
30. Freundlich, H.; Stern, R.; Zocher, H. The Colloidal Chemistry of Arsphenamine and Neoarsphenamine. *Biochem. Z.* **1923**, *138*, 307–317.
31. Hartshorne, N.H.; Woodart, G.D. Mesomorphism in the System Disodium Chromoglycate–Water. *Mol. Cryst. Liq. Cryst.* **1973**, *23*, 343–368.
32. Mlodozieniec, A.R. Thermodynamics and Physical Properties of a Lyotropic Mesophase (Liquid Crystal) and Micellar Solution of an Ionic Amphiphile. *J. Soc. Cosmet. Chem.* **1978**, *28*, 659–683.
33. Eckert, T.; Fischer, W. Organic Salts of Flufenaminic Acid: A New Class of Materials Forming Lyotropic Mesophases in Aqueous Systems. 1. Electron Microscopic Study. *Colloid Polym. Sci.* **1981**, *259*, 553–560.
34. Hamann, H.-J.; Mueller-Goymann, C.C. Lyotroper Mesomorphismus von Arzneistoffmolekülen Unter Besonderer Berücksichtigung der Profene. *Acta Pharm. Technol.* **1987**, *33*, 67–73.
35. Powell, M.F.; Sanders, L.M.; Rogerson, A.; Si, V. Parenteral Peptide Formulations: Chemical and Physical Properties of Native Luteinizing Hormone-Releasing Hormone (LH-RH) and Hydrophobic Analogs in Aqueous Solution. *Pharm. Res.* **1991**, *8*, 1258–1263.
36. Kriwet, K.; Mueller-Goymann, C.C. Binary Diclofenac Diethylamine Water Systems: Micelles, Vesicles and Lyotropic Liquid Crystals. *Eur. J. Pharm. Biopharm.* **1993**, *39*, 234–238.
37. Rades, T.; Padmadisastra, Y.; Mueller-Goymann, C.C. Thermal Behaviour and Solubility of Fenoprofen Calcium. *Pharmazie* **1996**, *51*, 846–851.
38. Mueller-Goymann, C. Liquid Crystals in Emulsions, Creams and Gels, Containing Ethoxylated Sterols as Surfactant. *Pharm. Res.* **1984**, *1*, 154–158.
39. Stoye, I. *Permeabilitätsveränderung von humanem Stratum corneum nach Applikation nicht-steroidaler Antirheumatika in verschiedenen kolloidalen Trägersystemen*; Thesis TU: Braunschweig, 1997.
40. Führer, C.; Junginger, H.; Friberg, S. Structural Studies of Ointments. Part 1: X-ray Structure Studies on the Hydrophilic Ointment DAB 7. *J. Soc. Cosmet. Chem.* **1978**, *29*, 703–716.
41. Müller-Goymann, C. Halbfeste Emulsionsähnliche Zustände. Seifen, Öle, Fette, Wachse. **1984**, *110*, 395–400.
42. Krempel, H.; Lill, N. Liposomen in Pharmazie und Kosmetik, Teil 2: Anwendungen und Wirkungen. *PZ Prisma*. **1997**, *4*, 46–54.
43. Schubert, R. Liposomen in Arzneimitteln. *Pharmazeutische Technologie: Moderne Arzneiformen*, 2. erw. Auflage; Müller, R.H., Hildebrand, G.E., Eds.; Wissenschaftliche Verlagsges: mbH: Stuttgart, Germany, 1998; 219–242.
44. Müller-Goymann, C.; Schütze, W. Mehrschichtige Phasengrenzen in Emulsionen. *Dtsch. Apoth. Ztg.* **1990**, *130*, 561–562.
45. Rades, T.; Gerke, A.; Schütze, W.; Müller-Goymann, C.C. Characterization of a Commercial Liposome Spray. *Pharmazie* **1997**, *52*, 44–50.
46. Euschen, A. *Diffusion in flüssigkristallinen Silastomeren - ein Beitrag zur Kontrolle der Arzneistofffreisetzung durch Diffusion*; Thesis Universität des Saarlandes: Saarbrücken, Germany, 1986.
47. Tiemessen, H.L.G.M. *Nonionic Surfactant Systems for Transdermal Drug Delivery*; Thesis Leiden University: The Netherlands, 1989.
48. Wahlgren, S.; Lindstrom, A.L.; Friberg, S. Liquid Crystals as a Potential Ointment Vehicle. *J. Pharm. Sci.* **1984**, *73*, 1484–1486.
49. Mueller-Goymann, C.C.; Frank, S.G. Interaction of Lidocaine and Lidocaine-Hydrochloride with the Liquid

- Crystal Structure of Topical Preparations. *Int. J. Pharm.* **1986**, 29, 147–159.
50. Mueller-Goymann, C.C.; Hamann, H.-J. Sustained Release from Reverse Micellar Solutions by Phase Transformation into Lamellar Liquid Crystals. *J. Contr. Release* **1993**, 23, 165–174.
51. Gray, D.G. Liquid Crystalline Cellulose Derivatives. *J. Appl. Polym. Sci.: Appl. Polym. Symp.* **1983**, 37, 179–192.
52. Schneeweis, A.; Mueller-Goymann, C.C. In Vivo and In Vitro Diclofenac Sodium Evaluation after Rectal Application of Soft Gelatine Capsules Enabling Application Induced Transformation (AIT) into a Semisolid System of Liquid Crystals (SSLC) for Controlled Release. *Pharm. Res.* **1997**, 14, 1726–1729.
53. Schneeweis, A.; Papantoniou, I.; Mueller-Goymann, C.C. Diclofenac Sodium Plasma Concentrations in Dogs after Peroral Application of Soft Gelatine Capsules Enabling Application Induced Transformation (AIT) into a Semisolid System of Liquid Crystals (SSLC) Compared to In Vitro Drug Release. *Pharm. Pharmacol. Lett.* **1997**, 7, 42–44.

DOSING OF DRUGS: DOSAGE REGIMENS AND DOSE-RESPONSE

Chyung S. Cook

Aziz S. Karim

Pharmacia Corporation, Skokie, Illinois, U.S.A.

INTRODUCTION

Drugs are administered for their pharmacological effects. In some cases, however, drug therapy includes the risk of undesirable side effects, with each drug having inherently different risks associated with its use. Therefore, it is the objective of physicians to administer a drug with an optimal dosing regimen by selecting the appropriate dosage, route, and frequency of administration to achieve maximal pharmacological efficacy with minimal side effects.

In most clinical situations, drugs are administered either repetitively, at time intervals, or as a continuous infusion to maintain a certain blood concentration at steady state within the known therapeutic concentration range for a given drug. A loading dose may be desirable in order to achieve the target concentration rapidly at the onset of therapy for a drug with a long half-life. These maintenance and loading doses as well as dosing frequency can be determined using pharmacokinetic principles.

DETERMINATION OF DOSE

Maintenance Dose

The general principle that is used to select the appropriate maintenance dose and dosing interval for the average patient is as follows: to maintain a target concentration at the steady state, the rate of drug administration should be equal to the rate of elimination. This can be expressed by the following equation (Eq. 1) using the clearance concept:

$$\text{Dosing rate} = (CLF) \times C_{ss} \quad (1)$$

where CL is clearance, C_{ss} is the steady state concentration of drug, and F is the fraction of the dose that is systemically available. Therefore, if CL , F , and C_{ss} are known, then the appropriate dose and dosing interval can be calculated.

For example, for theophylline, the clearance value is 0.65 mL/min/kg (2.34 L/h for a 60 kg man), the half-life in

healthy nonsmoking asthmatics is approximately 9 h and the effective plasma concentration range is 5–15 mg/L. Because the dose is an IV infusion, the systemic availability, F is 1. If the target plasma concentration at the steady state is 10 mg/L, then the

$$\begin{aligned} \text{IV infusion rate for a 60 Kg man} \\ &= 2.34 \text{ L/h} \times 10 \text{ mg/L} \\ &= 23.4 \text{ mgh.} \end{aligned}$$

If the dose is administered not as an infusion but as a bolus dose (e.g., IV injection or an oral dose), high fluctuations in plasma concentrations between peak (C_{max}) and trough values (C_{min}) will occur. If absorption and distribution are very rapid, fluctuation of plasma concentrations between C_{max} and C_{min} values will be controlled almost entirely by the elimination of the half-life. When the dosing interval is the same as the half-life, the ratio of C_{max}/C_{min} is 2, which is usually tolerable as long as pharmacological activity is directly correlated with plasma concentrations.

Using theophylline as an example again, if the targeted average plasma concentration is 10 mg/L, the dose calculated in the above example is 23.4 mg/h. Therefore, to maintain this concentration as an average concentration, IV bolus doses can be given at 140 mg every 6 h (qid), 187 mg every 8 h (tid), 280 mg every 12 h (bid), or 560 mg once a day (qd). The average concentration for a noninfusion dose regimen is defined as

$$\text{Average plasma concentration} = \text{AUC}/\tau \quad (2)$$

where AUC is the area under the plasma concentration–time curve and τ is dosing interval.

Although all these dose regimens will give the same target average plasma concentration, both $C_{ss,max}$ and $C_{ss,min}$ values will be markedly different. For a given dose, $C_{ss,max}$ and $C_{ss,min}$ values can be estimated using the following equations (Eqs. 3 and 4). Alternatively, the dose that will give the targeted maximal and minimal concentrations can be estimated using these equations:

$$C_{ss,max} = \frac{F \cdot \text{Dose} \cdot e^{-KT_{max}}}{V_{ss}(1 - e^{-K\tau})} \quad (3)$$

$$C_{ss,max} = \frac{K_a \cdot F \cdot \text{Dose} \cdot e^{-K\tau}}{V_{ss}(K_a - K)(1 - e^{-K\tau})} \quad (4)$$

K is the elimination rate constant (equal to 0.693 divided by the clinically relevant half-life). K_a is the absorption rate constant. T_{max} is the time to reach C_{max} . V_{ss} is the volume of distribution at steady state.

For an IV dose, T_{max} is 0, and the term $e^{-KT_{max}}$ becomes 1 in Eq. 3. The term $K_a/K_a - K$ in Eq. 4 also becomes 1. The term $e^{-K\tau}$ is the fraction of the last dose that remains in the body at the end of a dosing interval.

For theophylline (V_{ss} for 60 kg man = 30 L, $K = 0.077 \text{ h}^{-1}$, and $\tau = 6, 8, 12$, or 24 h), the calculated C_{max} , C_{min} and C_{max}/C_{min} values for IV doses are as follows:

Dose regimen	C_{max} (mg/L)	C_{min} (mg/L)	C_{max}/C_{min}
140 mg qid	12.6	8.0	1.6
187 mg tid	13.6	7.3	1.9
280 mg bid	15.5	6.1	2.5
560 mg qd	22.2	3.5	6.3

Because theophylline has a relatively narrow therapeutic index, it would be desirable to determine a dose at which the maximal and minimal concentrations at steady state will be within the therapeutic concentration range (5–15 mg/L). Therefore, the 280 mg bid dose regimen is convenient and adequate to achieve C_{max} and C_{min} within the therapeutic range.

As can be seen in the table, the more frequently a drug is given, the smaller the ratio of peak-to-trough plasma concentrations will be. These phenomena are also demonstrated by the computer-simulated plasma concentration-time curves (Fig. 1). If the dosing interval is equal to or smaller than the half-life, then the fluctuation between C_{max} and C_{min} is usually acceptable even for a drug with a narrow therapeutic window. If the dosing interval is greater than the half-life, a large fluctuation is expected, which may not be desirable for drugs with a narrow therapeutic window. However, if the half-life is very short, dosing at every half-life is not practical. Therefore, different formulations (e.g., extended-release or controlled release) are often used. For drugs with a long half-life (greater than 12 h), the dose can be given once or twice a day to maintain appropriate therapeutic levels.

Following multiple oral doses of a drug that obeys bi-exponential kinetics after IV administration (two-compartment model), the estimation of the $C_{ss,max}$ and $C_{ss,min}$

involves a complicated set of exponential constants for absorption and distribution, as shown below.

$$C_{ss,max} = \frac{F \cdot \text{Dose}(K_{21} - \alpha)e^{-\alpha T_{max}}}{V_{ss}(\beta - \alpha)(1 - e^{\alpha\tau})} + \frac{F \cdot \text{Dose}(K_{21} - \beta)e^{-\beta T_{max}}}{V_{ss}(\alpha - \beta)(1 - e^{\beta\tau})} \quad (5)$$

$$C_{ss,min} = \frac{K_a \cdot F \cdot \text{Dose}(K_{21} - \alpha)e^{-\alpha\tau}}{V_{ss}(K_a - \alpha)(\beta - \alpha)(1 - e^{\alpha\tau})} + \frac{K_a \cdot F \cdot \text{Dose}(K_{21} - \beta)e^{-\beta\tau}}{V_{ss}(K_a - \beta)(\alpha - \beta)(1 - e^{\beta\tau})} \quad (6)$$

where α and β are rate constants for distribution and elimination phases, respectively, and K_{21} is the rate constant from tissue to plasma compartment.

As with the one compartment model, T_{max} is 0 for an IV bolus dose or rapid oral absorption, and both terms $e^{-\alpha T_{max}}$ and $e^{-\beta T_{max}}$ become 1 in Eq. 5. Also in Eq. 6, both terms $K_a/(K_a - \alpha)$ and $K_a/(K_a - \beta)$ become 1.

If both the absorption and distribution are very rapid, these terms can be ignored for simplicity, and a maximal steady state concentration can be easily predicted by omitting the $e^{-KT_{max}}$ term in the numerator of the above Eq. 3 even for the two compartment model. Because of this approximation, the predicted maximal concentration from this equation will be greater than that actually observed.

As discussed above, it is possible to design a dose with pharmacokinetic parameters alone. However, in some cases, pharmacodynamic considerations make the selection of the dose regimen deviate from this principle. If pharmacological half-life is substantially different from pharmacokinetic half-life, then pharmacodynamic half-life (instead of pharmacokinetic half-life) should be used. For example, aspirin is rapidly hydrolyzed to salicylic acid, but the half-life for antiplatelet activity is in days. Therefore, a small amount of aspirin once a day is a good dose regimen in order to achieve the desired pharmacological response. Another exceptional case is for the drugs that are relatively nontoxic even though the pharmacological activity is directly related to plasma concentrations (i.e., there is a large therapeutic index). In this case, a high dose can be given so that the dosing interval can be much longer than the elimination half-life. The half-life of penicillin G is less than 1 h, but it is given at very large doses every 6 or 12 h.

If absorption is slow, and the apparent absorption half-life is much longer than the elimination half-life, then the dose regimen can be based on the apparent half-life of the absorption phase. Changing the formulation is a relatively

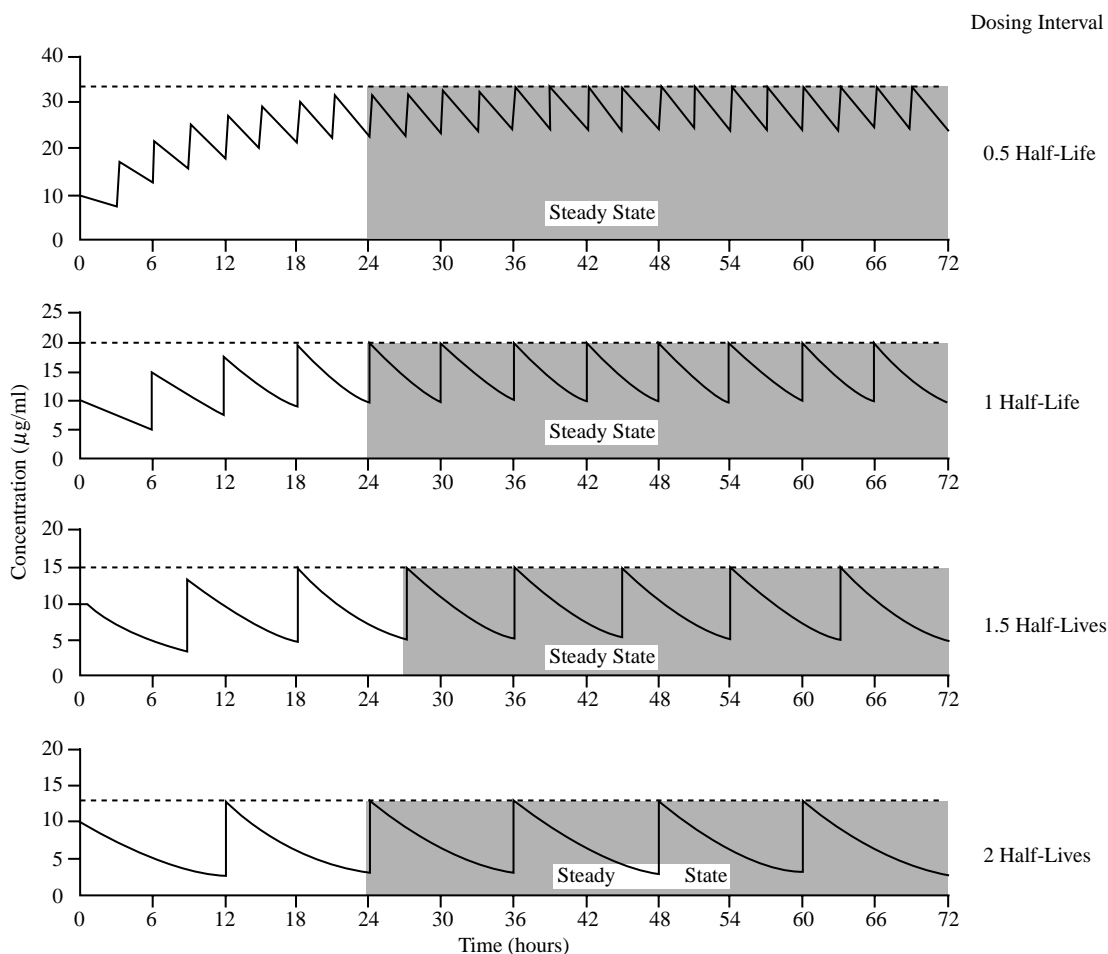


Fig. 1 Simulated plasma concentration-time curves obtained after repetitive doses were given at the intervals of every 1, 1.5, 2, and 3 half-lives. Steady-state levels were obtained between 4 (24 h) and 5 (30 h) half-lives, regardless of dosing interval. The accumulation factors (R) are 3.4, 2, 1.6, and 1.3 for dosing intervals of 0.5, 1, 1.5, and 2 half-lives, respectively.

common approach that is taken for short-acting drugs to extend the duration of absorption. The elimination half-life of nitroglycerin is approximately 2 min. However, nitroglycerin from a transdermal formulation is slowly released, and therapeutic plasma concentrations can be maintained for 24 h.

Another question regarding dose regimen is how often the dosage has to be changed and by how much. This can be usually determined using simple pharmacokinetic principles. The dose will not change by more than 50% and no more often than every 3 to 4 half-lives. Regardless of the half-life and frequency of dosing, it will take 3 half-lives to reach 87.5% of the steady state plasma concentration if the elimination rate constant is smaller than the absorption rate constant (Fig. 2). However, if the elimination rate constant is much greater than the absorption rate constant, the time to reach steady state

will depend on the absorption rate constant and not on the elimination constant.

Loading Dose

The appropriate magnitude for the loading dose can be calculated as follows:

$$\text{Loading dose} = \text{Target plasma concentration} \times V_{ss}/F \quad (7)$$

To achieve steady-state plasma concentrations, it takes approximately 4 elimination half-lives. If the half-life is long and the drug effects are immediately required for treatment of a life-threatening condition, then single or multiple loading doses are sometimes inevitable. For example, the half-life of lidocaine is greater than 1 h.

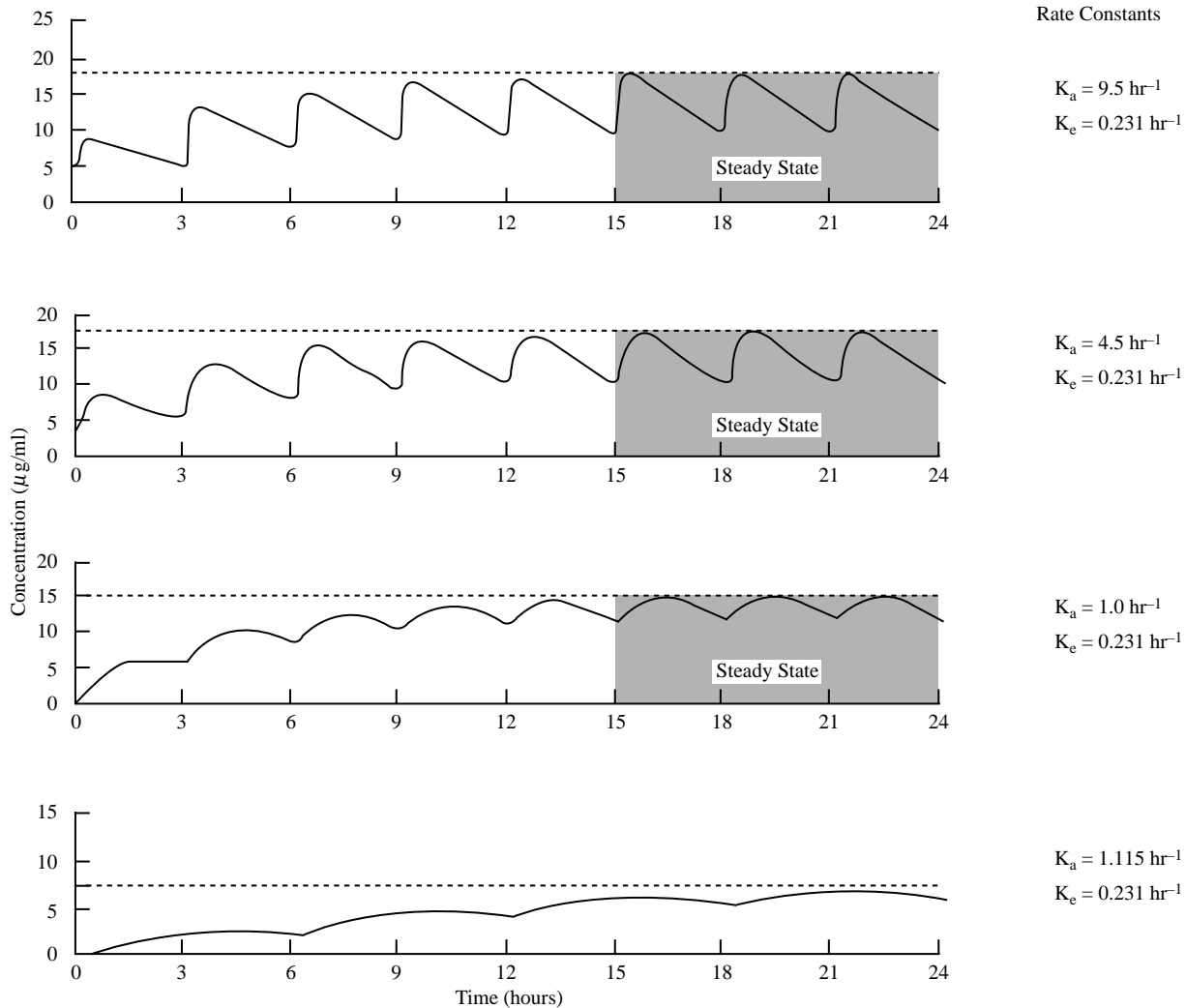


Fig. 2 Computer-simulated plasma concentration-time curves that show effects of absorption rate on the time to reach steady-state plasma concentrations. When absorption rate is higher than the elimination rate, the time to reach steady-state concentration is dependent on the elimination rate (top three figures). When absorption rate is lower than elimination rate, the time to reach steady-state concentration is dependent on the absorption rate (bottom figure).

However, a patient with arrhythmia after a myocardial infarction cannot wait 4–6 h to achieve therapeutic concentrations of the drug by IV infusion. Thus, the use of a loading dose of lidocaine in the coronary unit is a standard therapy.

For oral doses, the loading dose is usually twice that of the single dose. This is based on the pharmacokinetic principle that the accumulation factor is two when a drug is administered at every half-life. However, a loading dose should be cautiously used, particularly for those drugs with a narrow therapeutic window, because the sensitive individual may be exposed to toxic concentrations. Furthermore, for drugs with long half-lives, it

will take a long time to eliminate a large loading dose from the body.

DOSE- (OR CONCENTRATION-)RESPONSE RELATIONSHIP

A drug is primarily carried by the blood from the absorption site to the target organ/tissue or blood component where the drug interacts with a receptor to produce its effect. Consequently, for many drugs, various types of relationships exist between the plasma or serum

concentration of a drug and its clinical efficacy and toxicity. Much effort has been directed toward establishing reliable mathematical relationships between drug bioavailability input (plasma concentrations) and pharmacological output (response) (1–3). Although the relationship is far from being completely understood, this dose-response relationship is very important in therapeutic decisions, and optimal dosage regimens can be deliberately planned for many drug therapies using the plasma concentration as a reasonable marker, particularly when pharmacological effects are easily measurable (e.g., blood pressure). This dose-response relationship generally depends on whether the plasma concentration is directly or indirectly related to response and whether a drug interacts with its receptor in a reversible or irreversible manner. If there is no apparent correlation between plasma/serum concentrations and the pharmacological effects and/or the effects are not easily measurable, the trial-and-error approach may be more practical.

Directly Reversible Pharmacological Response

The concept of a direct and rapidly reversible pharmacological response implies that the intensity of response is directly associated with the drug concentration at the site of action. In this category, two models (pharmacodynamic and pharmacokinetic–pharmacodynamic) are discussed.

In the pharmacodynamic model, the drug concentration at the receptor site is proportional to the drug concentration in the plasma, regardless of the pharmacokinetic model (one compartment or multicompartment), and the interaction between the drug and receptor is directly and rapidly reversible after drug administration.

Plasma concentration (C) and intensity (I) of the pharmacological response often follow an empirical relationship, known as the Hill equation:

$$I = \frac{I_{\max} C^n}{(C_{50}^n + C^n)} \quad (8)$$

where I_{\max} is the maximal effect attributable to the drug and C_{50} is the concentration producing 50% of the maximum effect. Two examples of this are the *in vivo* effect of *d*-tubocurarine on muscle strength in patients and the plasma concentration–antiarrhythmic effect of tocainide in humans.

Another common and empirical relationship between concentration and response is the sigmoidal plot of response vs. logarithm of dose, plasma concentration, or drug concentration in the body. Very often, this curve shows excellent linearity from at least 20–80% of the maximum attainable intensity of response, which is a

region of particular interest in drug therapy. Such apparent linearity between response and log plasma concentration has been demonstrated for a number of drugs, such as theophylline and propranolol. Relating response to the logarithm of plasma concentration rather than to the logarithm of dose reduces the variability in response due to differences in absorption, metabolism, and excretion among patients.

The pharmacokinetic–pharmacodynamic model is one in which drug concentrations at the effect site are not known or cannot be estimated without knowledge of the drug effect. If the pharmacological response is associated with a peripheral compartment (not with a central compartment) that receives a substantial mass of drug, a linear, log-linear, or sigmoidal relationship may also be obtained when the response is plotted against the calculated concentrations of the drug in the peripheral compartment (not plasma concentrations). A linear relationship was obtained between behavior response and the predicted concentration of the drug in the tissue compartment after IV administration of *d*-lysergic acid diethylamide (LSD) to humans.

Indirect Pharmacological Response

In contrast to the direct pharmacological responses discussed above, some pharmacological responses are indirect and represent the net result of several processes of which only one is influenced by the drug. In this case, a direct relationship between plasma concentration and pharmacological response may not be obtained. However, if the process influenced by the drug can be identified, then the drug concentration may relate to changes in this process. This concept is illustrated by the effects of oral anticoagulants such as warfarin.

Warfarin inhibits the synthesis of certain vitamin K-dependent clotting factors. However, warfarin has no effect on the physiological degradation of these factors. Therefore, the pharmacological response of warfarin should be based on the inhibitory effect on the synthesis of the clotting factor (prothrombin) rather than on the change in clotting time. These mechanism-based models were found to be relevant for the clinical effects of numerous drugs (4). Many metabolic and endocrine systems provide similar modeling strategies (5).

Irreversible Pharmacological Response

Most drugs produce a reversible pharmacological response. However, some antibiotics, irreversible enzyme inhibitors, and anticancer agents incorporate irreversibly or

covalently into a cell's metabolic pathway. This results in an irreversible effect—cell death. Complex kinetic models are used to explain the relationship of dose-chemotherapeutic effects for some drugs, such as methotrexate, cyclophosphamide, and arabinosylcytosine (2).

FACTORS AFFECTING DRUG RESPONSE

Dosage regimens are generally derived using average pharmacokinetic parameters to maintain plasma concentrations of the drug within the therapeutic window. For many drugs, both therapeutic range and toxic concentrations have been established (6). However, for a fixed dose, the plasma concentration of a drug within a patient can be influenced by many pharmacodynamic and pharmacokinetic factors. Therefore, these factors must also be taken into consideration for maintaining desired therapeutic concentrations of a drug.

Pharmacokinetic Factors

Factors affecting absorption

Factors affecting drug absorption include formulation, disease state, food effect, and drug–drug interaction. Formulations used for oral administration include solutions, suspensions, capsules, and uncoated and coated tablets. Depending on the formulation of a drug, the absorption characteristics may differ substantially. Slow-release oral formulations are often used to prolong the drug's activity and to reduce the fluctuation between C_{\max} and C_{\min} values.

Intestinal surgery and disease states have been shown to alter the absorption of some drugs, although information on this subject is limited. For a given disease state or surgery, drug absorption may be increased, unchanged, or decreased, depending on the drug. Therefore, the effect of a particular disease condition on drug absorption cannot usually be predicted from the existing information.

The effect of food on the absorption of a drug from the gastrointestinal (GI) tract is quite variable and depends on the physicochemical properties of the drug substance and the mechanism by which it is absorbed. The presence of solid food in the stomach will tend to decrease the rate of stomach emptying and thus delay the absorption of the drug, which often results in decreased systemic availability of the drug. The relative bioavailability of lincomycin is reduced to about 60% when given 1 h before breakfast and to about 20% when given immediately after breakfast, compared with that observed

after oral administration to fasting subjects. A potentially dangerous situation may arise owing to delayed absorption of hypnotic agents in nonfasting patients. With the hypnotic capuride, a 42-min difference in onset of absorption has been observed between fasting and nonfasting subjects.

The presence of food is also found to increase gastrointestinal motility and splanchnic blood flow up to 30%. Consequently, the absorption of some drugs is either unaffected or increased by food. Among the beta-blocking agents, absorption of bevantolol and oxprenolol was unchanged with food, whereas absorption of metoprolol, labetalol, and propranolol was increased, but the absorption of atenolol and sotalol was decreased. The oral absorption in humans of the antifungal antibiotic griseofulvin is substantially greater with food of high fat content than without food. Because a high fat concentration in the small intestine stimulates bile secretion, absorption of the relatively lipophilic drug may be increased by enhancing its dissolution in the GI tract. Furthermore, the increase in splanchnic blood flow resulting from food consumption may contribute to the enhanced absorption. However, relatively polar and poorly permeable drugs show a tendency toward reduced absorption in the presence of food (7). Food decreases absorption of alendronate, astemizole, captopril, didanosine, and penicillamine.

Food effects on drug absorption will depend not only on the physical and chemical properties of the drug substance but also on the formulation. Following oral administration of theophylline, the absorption rate of the drug can be either increased or decreased with food, depending on the formulation (8). Generally, the greatest food effect is observed when the drug is taken immediately after a meal; the degree of interaction decreases as the time between eating and drug-dosing increases (9). Extensive reviews of the effect of food on drug absorption can be found in the literature (10–12).

Drug interaction is another important factor affecting absorption. In general, interactions that interfere with absorption either involve binding or chelation of drugs in the GI tract, rendering them nonabsorbable, or involve effects on gastric emptying or gastrointestinal motility. In addition, gastric pH changes may affect on the absorption of some drugs.

Antacids, particularly aluminum hydroxide gel, reduce the absorption of most tetracycline drugs by forming an insoluble complex in the gut. Simultaneous administration of iron (ferrous sulfate) with tetracycline, oxytetracycline, methacycline, or doxycycline seriously impairs the absorption of these antibiotics. Adsorbants such as kaolin (an antidiarrheal agent) substantially reduce the absorption of lincomycin and promazine. Ion-exchange resins, such as

cholestyramine, strongly bind many anionic and neutral drugs in the GI tract and interfere with the absorption of anticoagulants and thyroxine. Imipramine significantly reduces bioavailability of levodopa in humans, presumably because delayed absorption by reducing gastric emptying time enhances metabolism of levodopa in the gut. Orally administered ketoconazole requires an acidic medium to dissolve adequately and, therefore, antacids, anticholinergic drugs, H₂ blockers, or acid pump inhibitors (e.g., omeprazole) will reduce the bioavailability of this drug. Metoclopramide, cisapride, and cathartics increase GI motility and may decrease the absorption of drugs that require prolonged contact with the absorbing surface and those that are absorbed only at a particular site along the GI tract. Anticholinergics decrease GI motility and may increase absorption by prolonging contact with the area of optimal absorption or may reduce absorption by slowing dissolution and gastric emptying. The interactions between drugs have been reviewed in detail (13–15).

Factors affecting distribution

After entering the general circulation, a drug is carried throughout the body and is distributed to various tissues. The drug distribution depends on the relative affinity of binding to plasma protein, red blood cells, and tissues. However, it can be altered in the presence of other drugs, in some disease states, and due to the age of the patient.

The risk of interactions resulting from the displacement of the drug from proteins is significant, primarily with drugs that are highly protein-bound (>90%) and that have a small apparent volume of distribution and a relatively narrow therapeutic window. This can result in a high concentration of unbound drug temporally (the first few days), which may have clinically important effects. Trichloroacetic acid, a major metabolite of chloral hydrate, displaces warfarin from its binding sites on plasma albumin. This displacement temporarily elevates plasma levels of warfarin and, thereby, increases the pharmacological effect per unit dose. Administration of 1 g of chloral hydrate daily for 1 week to subjects on warfarin increases the hypothermic effect by 40–80%. When valproic acid and phenytoin are coadministered, unbound phenytoin concentrations increase significantly in some patients causing more adverse reactions, even when total phenytoin serum concentrations are within the usual therapeutic range.

Disease states are another factor that may affect the binding of drugs to proteins in plasma and other tissues. Unusually low binding of drugs to plasma proteins has been observed in various diseases. For example, the percentage of unbound phenytoin in plasma was 5.8–7.3% in normal subjects, whereas the percentages of unbound drug in patients with renal diseases (azotemia

or uremia) were 8–25%. In addition to renal diseases, liver diseases, hyperbilirubinemia, and hyperlipidemia are reported to lower plasma protein binding of some drugs. However, during physiological stress (e.g., myocardial infarction, surgery, ulcerative colitis, and Crohn's disease), the plasma protein binding of basic drugs (e.g., propranolol, quinidine, and disopyramide) increases, and volume of distribution decreases accordingly due to the increase in concentrations of the α 1-acid glycoprotein.

Age-related changes in drug distribution have been reported. The apparent volume of distribution is somewhat larger in newborns and infants than in adults. The estimated volume of distribution of sulfamethoxypyridazine in newborns and infants is 0.47 and 0.36 L/kg, respectively, whereas the values are 0.20–0.26 L/kg in children, adults, and elderly subjects. The volume of distribution of chlorthalidone is substantially larger in the elderly (0.52 L/kg) than in the young (0.42 L/kg). The age-related difference in the volume of distribution may be due to a difference in plasma protein binding and/or in the relative size of body compartments.

Factors affecting metabolism

Drug metabolism is a primary mechanism for removal of drugs from the body. Drug metabolism may be affected by age, disease state, polymorphism, and inhibition or induction of drug-metabolizing enzymes. Sex differences in drug metabolism are minor in humans and usually are not clinically important.

Although hepatic drug metabolism is generally slower in neonates as compared with adults, children actually metabolize certain drugs more rapidly than adults. The half-life of theophylline averaged 3.7 h in children but ranged from 5 to 9 h in nonsmoking adults. Children also metabolize antipyrine, clindamycin, diazoxide, and phenobarbital more rapidly than adults. In elderly subjects, metabolism of some drugs (e.g., amobarbital, antipyrine, phenylbutazone, and chlorthalidone) has been reported to be less efficient than in younger subjects. Thus, the aging process may reduce the total clearance of a drug and extend its elimination half-life. After oral administration of amobarbital, the mean percentages of the dose excreted as the hydroxy metabolite in 0–24 h urine specimens was reduced from 14% in the younger subjects (20–40 years of age) to only 4% in the elderly subjects (>65 years of age). The apparent elimination half-life of chlorthalidone showed a strong correlation with age, increasing from a mean of 9.5 h in the young to 37 h in the elderly. Consequently, the plasma concentrations of these drugs were greater in the elderly subjects than in the younger subjects.

Genetic polymorphism in metabolism is a very important factor in the variability in drug elimination. The metabolism of isoniazid (and other drugs that have a hydrazine moiety such as sulfamethazine, sulfapyridine, hydralazine, procainamide, and dapsone) is more rapid in some individuals than in others. The major route of elimination of isoniazid in humans is via metabolism to acetylisoniazid by hepatic *N*-Acetyltransferase; thus, the half-life of the drug depends on how rapidly the drug can be acetylated. In about 50% of the U.S. population and in 5–10% of the Asian population, hepatic *N*-acetylation is slow. In those with extensive (rapid) metabolism, the half-life of the drug is 45–80 min, whereas in those poor (slow) metabolism, the half-life is 140–200 min. Slow acetylators require a longer time to acetylate these drugs and, therefore, are more susceptible to adverse effects caused by the parent compound (e.g., peripheral neuritis with isoniazid, lupus erythematosus with hydralazine or procainamide, and sedation and nausea with phenelzine). Compared with slow acetylators, the rapid acetylators require larger and more frequent doses of such a drug in order to obtain the desired therapeutic response.

Drugs that are metabolized by the cytochrome P-450 (CYP) isoenzymes CYP2D6, CYP2C9, and CYP2C19 also exhibit genetic polymorphisms. An example of CYP2D6 metabolism is debrisoquine. In about 5–10% of Caucasians in North America and Europe and about 1% of Asians, 4-hydroxylation of debrisoquine is reduced, and such individuals are at increased risk for toxicity (orthostatic hypotension). Beta blockers (metoprolol and timolol), antiarrhythmic drugs (encainide and flecainide), tricyclic antidepressants (amitriptyline, nortriptyline, and desipramine) and an antitussive drug (dextromethorphan) are also metabolized by CYP2D6. Therefore, in the CYP2D6 poor metabolizers, unusually large therapeutic responses and/or more adverse effects may be observed compared to extensive metabolizers.

Phenytoin is metabolized by CYP2C9. In CYP2C9 poor metabolizers, increased toxicity (e.g., excessive CNS depression) was observed. *S*-mephenytoin is stereoselectively metabolized by CYP2C19. About 3–5% of Caucasians in North America and 20% of Asians metabolize mephenytoin slowly, which increases their risk of a major side effect (transient sedation). In such persons, other drugs metabolized by the same isoenzyme as mephenytoin will also be poorly metabolized, and pharmacological activity and/or side effects of these drugs may be increased. These drugs include mephobarbital (an anticonvulsant), proguanil (antimalarial), and possibly diazepam (an anxiolytic). These drugs have a substantially longer half-life in poor metabolizers than in extensive

metabolizers. Thus, caution must be used for the patient with limited capacity of metabolism.

Other genetic polymorphisms include deficiencies in aldehyde dehydrogenase-2 (an enzyme involved in ethanol metabolism), alcohol dehydrogenase (another enzyme involved in ethanol metabolism), glucose-6-phosphate dehydrogenase (G6PD), or glutathione synthetase. Patients with G6PD deficiency (about 10% of Black males) are at increased risk of developing hemolytic anemia when given oxidant drugs, such as antimalarials (e.g., chloroquine, pamaquine, and primaquine), aspirin, probenecid, and vitamin K. Patients with low levels of glutathione synthetase in hepatocytes are at increased risk of liver damage if given drugs metabolized by glutathione conjugation, such as acetaminophen and nitrofurantoin.

Some disease states may have clinically significant effects on drug metabolism. If hepatic metabolism is an important route of drug elimination, any dysfunction of the liver may lead to a change in the pharmacokinetics of that drug, and subsequently plasma concentrations of the drug will be different from those in a patient with normal liver function. The clinical significance of the changes in drug elimination depends on the type or severity of the disease. Antipyrine has often been used as a model drug for the investigation of the effects of liver disease on oxidative drug metabolism in humans. In healthy subjects, the average half-life of antipyrine is approximately 12 h. In contrast, the half-life of the drug in patients with liver cirrhosis or chronic active hepatitis averages 34 and 26 h, respectively. However, the half-life of antipyrine does not provide a quantitative index of the degree of impairment of drug elimination in a patient for a specific drug. Some other examples of drugs whose half-lives in patients change due to impairment of metabolism in particular disease states include hexobarbital in hepatitis, meperidine in liver cirrhosis, propranolol in chronic liver diseases, and rifampin in obstructive jaundice.

Enzyme induction or inhibition is another factor that alters the rate of metabolism of many drugs. Many drugs, polycyclic hydrocarbons, insecticides, and other environmental chemicals (e.g., cigarette smoke) stimulate activity of microsomal enzymes in the liver. Consequently, repetitive administration of certain drugs or continuous systemic exposures to certain chemicals may facilitate elimination of other drugs by increasing their metabolism. Phenobarbital and other barbiturates are among the most potent and widely studied inducing agents in humans. During repetitive administration of phenobarbital (2 mg/kg/day) for a period of 4 weeks, a substantial decrease in the efficacy of warfarin was observed after the first week, as shown by a substantial decrease in mean prothrombin time. Rifampin also induces metabolism of

warfarin. After administration of 600 mg of rifampin a day to subjects receiving repetitive daily doses of warfarin, plasma concentrations of warfarin were reduced to near nondetectable levels.

Some drugs, such as rifampin, induce their own metabolism. During repetitive administration of rifampin at a daily dose of 600 mg for 30 days, the elimination half-lives of rifampin were 4.2 and 1.9 h on the first and last days, respectively. Carbamazepine also shows pronounced self-induction. The plasma concentrations at steady state predicted from single-dose data were two to three times higher than the actual concentrations observed during chronic treatment.

Whereas the clinical consequence of enzyme induction is usually a decrease in the efficacy of the drug, inhibition of drug metabolism can lead to serious adverse effects from the accumulation of drugs to toxic concentrations. However, by identifying the CYP isozymes responsible for metabolism of substrates and inhibitors and determining their *in vitro* inhibition constants, it is possible to predict clinically significant drug interaction in many cases. Chloramphenicol is a potent inhibitor of the metabolism of CYP2C9 substrates (e.g., tolbutamide, phenytoin, and dicumarol) in humans. Treatment with 2 g of chloramphenicol for several days resulted in a marked rise in the steady state serum concentrations of tolbutamide and phenytoin. The area under the plasma concentration-time curves of tolbutamide was elevated five-fold by coadministration of sulfaphenazole.

Among the CYP isozymes, CYP3A4 is responsible for the metabolism of the widest range of drugs and endogenous compounds in humans. It accounts for 60% of CYP isoforms in human liver. Therefore, if a drug is metabolized by CYP3A4, concomitant administration of other drug that is metabolized also by CYP3A4 may result in a drug interaction. For example, ketoconazole inhibited metabolism of the nonsedating antihistamines terfenadine and astemizole, and cisapride. As a result of CYP3A4 inhibition, life-threatening arrhythmic side effects of these drugs occurred. An extensive review of clinically significant drug interactions can be found in the literature (16).

Factors affecting excretion

Renal excretion of drugs is accomplished by both a pH-dependent diffusion and active secretion. Changes in urine pH may occur due to disease states, dietary factors, or simultaneous administration of drugs. The effect of urinary pH change on the elimination rate of acidic and basic drugs can be pronounced and may be predicted according to the pH partition hypothesis as long as excretion of the drug in urine is an important route of elimination.

For an acidic drug, a more alkaline urine increases ionized drug concentration and decreases reabsorption at the distal tubule. Consequently, elimination of the drug will be faster. On the other hand, more acidic urine leads to a decreased concentration of the ionized acidic drug, resulting in increased reabsorption and slower elimination. After sulfaethidole was administered to humans, the half-life decreased from 11.4 to 4.2 h when urinary pH increased from 5 to 8, respectively. The elimination rates of sulfalene and sulfasymazine were doubled by the administration of sodium carbonate, which increased urine pH from 6 to 8. Elimination of salicylate was slower in humans with low urine pH, and higher plasma concentrations were achieved.

For basic drugs, urinary excretion increases with low urinary pH and decreases with high pH. As a result, patients receiving amphetamines or closely related drugs may be placed at risk by relatively small changes in urine pH. Patients with alkaline urine had intense psychoses lasting more than 3 days after the last dose of amphetamine.

Age-related alteration of renal function is a very important factor in selecting the dose regimen. Renal function in newborns is incompletely developed. Neonatal renal plasma flow and glomerular filtration rates (normalized for body surface) are only 30–40% of those of adults. The half-life of penicillin G is 3.2 h in newborns (up to 6 days of age) and 1.4 h in infants (14 days of age or older), whereas in older children and adults, it is about 0.5 h. The mean half-life of gentamicin is about 5 h in newborns under 1 week of age and about 3 h in infants 1–4 weeks of age. The half-life of gentamicin in older infants and adults is approximately 2 h. Thus, drugs that depend on renal excretion as the principal mode of elimination would be expected to have a longer residence time in infants.

Impaired renal function may affect drug elimination to varying degrees, depending on the extent of the contribution of renal clearance to the total clearance. The greater the contribution of the renal clearance to the total clearance, the greater the influence of renal disease will be on drug elimination and steady state concentration in the plasma during repetitive dose administration.

Pharmacodynamic Factors

There are many examples of drugs that interact at a common receptor site or that have additive or inhibitory effects. Also, there are some other interactions of an apparently pharmacodynamic nature whose mechanism is poorly understood.

Monoamine oxidase (MAO) inhibitors, such as, phenelzine or tranylcypromine, cause norepinephrine to accumulate within adrenergic neurons. Thus, drugs that release norepinephrine (e.g., indirectly acting sympathomimetic amines, amphetamine, and phenylpropanolamine) can produce an exaggerated response including severe headache, hypertension, and cardiac arrhythmias. MAO inhibitors are present in many over-the-counter drugs for cold, allergy, and diet products, and patients taking MAO inhibitors should avoid such products.

OPTIMIZATION OF DOSAGE REGIMENS FOR INDIVIDUAL PATIENT

Dose regimen is determined using the average pharmacokinetic parameters. However, unpredictable variation exists even among normal individuals. For many drugs, one standard deviation in the values observed for F , CL , and V_{ss} is about 20%, 50%, and 30%, respectively. Therefore, 95% of the time, the C_{ss} that is achieved based on average pharmacokinetic parameters will be between 35 and 270% of the target concentration. This is not an acceptable range, especially for the drugs with narrow therapeutic index. If F , CL , and V_{ss} can be directly measured and the dose is further adjusted for body weight or surface area, drug–drug interactions, and disease state, such as renal impairment (for renally excreted drugs), then a more precise dose regimen can be made for a particular patient.

Dosage Adjusted for Body Weight

The same amount of drug in different individuals can result in different drug concentrations in plasma, because the apparent volume of distribution of a drug depends on the size of the body compartment as well as on the relative binding of the drug in the vascular and extravascular spaces. The volume of total body water and extracellular fluid in subjects with a normal lean to fat ratio is directly proportional to body weight. A relationship exists between the apparent volume of distribution and body weight, as well as between steady state plasma concentration and body weight. To maintain the therapeutic plasma concentration at the steady state in nonobese adults, it is better to determine the dose on a mg/kg body weight basis than to administer a fixed dosage per patient.

The distribution of relatively polar drugs (e.g., digoxin and gentamicin) is limited to body water, whereas highly lipid-soluble drugs (e.g., thiopental) distribute into fat. Therefore, for polar drugs, the use of lean body mass instead of total body weight allows for a more predictable

dose–response relationship in obese patients. In fact, if digoxin dosage for obese individuals is calculated on the basis of total body weight, plasma concentrations of the drug may be dangerously high.

The percentages of fat and lean body mass in an individual may be estimated with the help of (Eqs. 9 and 10).

$$\% \text{ Fat} = 90 - 2(H - G) \quad (9)$$

where H is height and G is girth, both in centimeters, using umbilical level at expiration.

$$\text{Lean bodymass} = (100 - \% \text{ fat}) \times \text{weight(in kg)} \quad (10)$$

Dosage adjusted for body surface area

Children require and tolerate a larger mg/kg dose than adults and cannot be considered as small adults. For example, the doses of digoxin required to maintain mean plasma concentrations of about 1–1.5 ng/ml are 15–20 $\mu\text{g/kg}$ per day for children 4 weeks to 2 years of age, 10–15 $\mu\text{g/kg}$ per day for children 2–12 years of age, and 4–5 $\mu\text{g/kg}$ per day for adults. Reasonable estimates of the doses required in children are often obtained by calculating the child's dose on the basis of body surface relative to that of an adult. An estimate of body surface area may be made from Eq. 11.

$$S = (W^{0.425} \times H^{0.725} \times 71.84) / 10^4 \quad (11)$$

where S is surface area in m^2 , W is body weight in kg, and H is height in cm. The approximation of the body surface area for children can be obtained using the simplified (Eq. 12).

$$S = 0.088 W^{0.728} \quad (12)$$

The body surface area of the average adult is usually taken to be 1.73 m^2 .

In general, for the average 3-month-old child weighing 6 kg, the body surface-adjusted dose is 2 times greater than the mg/kg dose given to an average adult, whereas for the average 5-year-old weighing 20 kg, the body surface-adjusted dose is 1.5 times greater than the mg/kg dose. The requirement of larger mg/kg doses in children than in adults may be related in part to the fact that total body water and extracellular fluid make up a higher percentage of the total body weight in children than in adults.

Dosage adjusted for drug interactions

Epidemiological studies demonstrate that if 5 or fewer drugs are coadministered, the rate of drug-related adverse

reactions is approximately 4%. However, when 20 or more drugs are prescribed, the rate of adverse reactions increases to 45%. Therefore, patients should take as few drugs as possible for concurrent use. If more than one drug therapy is required and interactions between the coadministered drugs are known to be clinically significant, the dosage of one or more drugs may be increased or decreased based on the clinical response and severity of side effects in the individual patient, or each drug may be administered with a certain time interval, depending on the situation during the concurrent therapy. After one or more drugs is discontinued, the dosage regimen of the other drug(s) may be readjusted for continuing therapy. If a drug or a particular formulation of a drug is known to have a food effect, the drug or formulation must be given at least 1 h prior to or after food intake for reliable therapy.

Dosage adjusted for renal impairment

The renal clearance of the drug is proportional to endogenous creatinine clearance, irrespective of the mechanisms (i.e., filtration, reabsorption, or secretion), and the creatinine clearance is used as an indicator of the severity of the disease. Furthermore, it is possible to

predict the half-life of a drug in a patient with renal disease based on the creatinine clearance and on a knowledge of the pharmacokinetics of the drug in normal subjects, as illustrated with cefazolin in Fig. 3.

The half-life of some drugs is changed sufficiently in patients with impaired renal function to warrant consideration of a change in the usual dosage regimen, thus preventing toxic accumulation of the drug in the body. Changes in regimen usually take the form of reducing the dose per dosing interval or increasing the length of the dosing interval. Either change is usually roughly proportional to the relative difference in half-life between the patients with and without renal disease. Cephalexin is administered as a 250-mg to 1-g dose every 4–6 h. Its average half-life in patients with normal renal function is about 0.5–1 h. In a patient with a creatinine clearance of 10–15 ml/min, the half-life of the drug increased about eightfold because cephalexin is eliminated almost entirely by urinary excretion of the parent drug. The dosing frequency suggested for this patient is the usual dose every 24 h, which is 3–6 times longer than the usual dosing interval. A comprehensive guide to drug usage in adult patients with impaired renal function is available (17).

Dosage adjustment for other diseases

If the major route of elimination of a drug is by metabolism in the liver and the patient has liver dysfunction, then dosage of the drug or frequency of dosing may be reduced. However, unlike patients with renal dysfunction, the degree of impairment of drug elimination in an individual with liver dysfunction cannot be predicted. In general, little change is necessary in dosage regimens when liver disease is inactive, although doses should be kept low, and particular care should be taken with sedative and antidepressant drugs. When liver disease is present, the dosage regimens of some drugs (e.g., sedatives and analgesics) in each patient should be titrated to his or her clinical response. In selected instances, plasma concentrations of the drug should be monitored and dosage regimens adjusted based on plasma concentration measurements.

It is generally recommended that altered plasma protein binding in the disease state does not require dosage adjustment. However, in patients with hypoalbuminemia, the incidence of drug toxicity appears to be increased, and, thus, dosages should be reduced. Patients showing toxicity may have impaired elimination in addition to altered plasma protein binding. In fact, it has been suggested that low serum albumin concentrations may be a marker for the integrity of liver function.

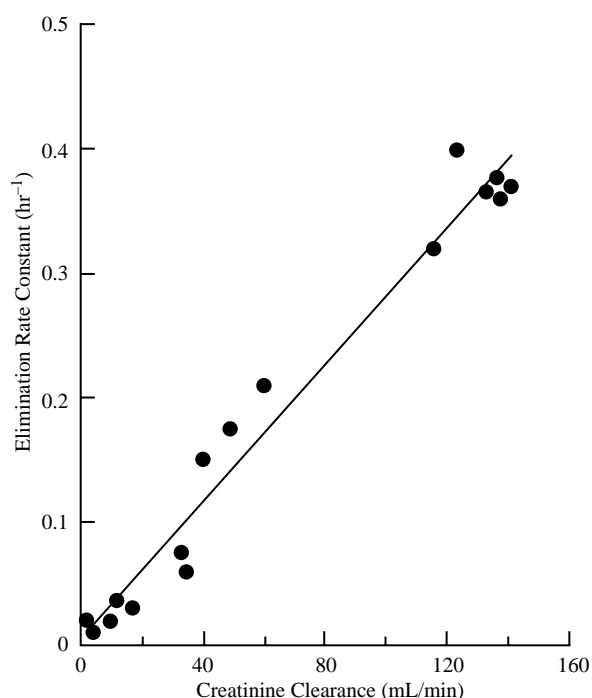


Fig. 3 Relationship between renal function (expressed by creatinine clearance) and overall elimination rate constant for cefazolin after a single intramuscular injection. (From Ref. 17.)

REFERENCES

- Holford, N.H.G.; Sheiner, L.B. *Clin. Pharmacokinetics* **1981**, *6*, 429–453.
- Gibaldi, M.; Perrier, D. *Pharmacokinetics*; Marcel Dekker, Inc: New York, 1982.
- Sheiner, L.B.; Stanski, D.R.; Vozeh, S.; Miller, R.D.; Ham, J. *Clin. Pharmacol. Ther.* **1979**, *25*, 358.
- Jusko, W.J.; Ko, K.C. *Clin. Pharmacol. Ther.* **1994**, *56*, 356–358.
- Carson, E.R.; Cobelli, C.; Finkelstein, L. *The Mathematical Modeling of Metabolic and Endocrine Systems*; John Wiley & Sons: New York, 1983.
- Benet, L.Z.; Sheiner, L.B. Design and Optimization of Dosage Regimens: Pharmacokinetic Data. *The Pharmacological Basis of Therapeutics*, 7th Ed.; Gillman, A.G., Ed.; Macmillan Publishing Co.: New York, 1985; 1663–1733.
- Lee, K.H.; Xu, G.X.; Schoenhard, G.L.; Cook, C.S. *Pharm. Res.* **1997**, *14*, 1030–1038.
- Cook, C.S.; Hauswald, C.L.; Grahn, A.Y.; Kowalski, K.; Karim, A.; Koch, R.; Schoenhard, G.L.; Oppermann, J.A. *Inter. J. Pharm.* **1990**, *60*, 125–132.
- Karim, A.; Burns, T.; Janky, D.; Hurwitz, A. *Clin. Pharmacol. Ther.* **1985**, *38*, 642–647.
- Welling, P.G. *J. Pharmacokinet. Biopharm.* **1997**, *5*, 291–334.
- Toothaker, R.D.; Welling, P.G. *Ann Rev. Pharmacol.* **1980**, *20*, 173–199.
- Fleusher, D.; Zhou, Y.; Pao, L.H.; Karim, A. *Clin. Pharmacokinet.* **1999**, *36*, 233–254.
- Griffin, J.P.; D'Arcy, P.F.; Speirs, C.J. *A Manual of Adverse Drug Interactions*, 4th Ed.; Butterworth & Co.: Oxford, UK, 1988.
- Grahame-Smith, D.G. *Drug Interactions*; The Macmillan Publishing Co: New York, 1977.
- Prescott, L.F. *Drugs* **1973**, *5*, 161–186.
- Michalets, E.L. *Pharmacotherapy* **1998**, *18*, 84–112.
- Bennett, W.M.; Singer, I.; Coggins, C.H. *JAMA* **1973**, *223*, 991–996.
- DeVane, C.L.; Jusko, W.J. Dosage Regimen Design. *Pharmacokinetics: Theory and Methodology*; Rowland, M., Tucker, G., Eds.; Pergamon Press: Oxford, UK, 1986; 213–233.
- Holford, N.H.G.; Sheiner, B.L. Kinetics of Pharmacologic Response. *Pharmacokinetics: Theory and Methodology*; Rowland, M., Tucker, G., Eds.; Pergamon Press: London, UK, 1986; 189–212.
- Welling, P.G. *Pharmacokinetics*; American Chemical Society: Washington, DC, 1986; 7–76.
- Rowland, M., Tozer, T.N., Eds. *Clinical Pharmacokinetics Concepts and Applications*; Williams and Wilkins: Media, PA, 1995.

DRESSINGS IN WOUND MANAGEMENT

Sarah M.E. Cockbill

University of Wales College, Cardiff, United Kingdom

INTRODUCTION

Throughout history, many diverse materials of animal, vegetable, and mineral origin have been used to treat wounds. They range from the hot oils and waxes reported in the Ebers papyrus (1) to the animal membranes and faeces of the Middle Ages to the picked oakum of the 19th century. Moist poultices are described on Sumerian tablets inscribed as early as 2100 B.C. (1, 2). Medieval manuscripts illustrate examples of both *moist environment* and *biological healing*, such as the practice of using dogs to lick wounds, particularly post surgery, to moisten, cleanse, and stimulate healing, thus relying on the now recognised, but at that time unknown, presence of antibacterial agents and growth factors in the saliva (3). Ambroise Paré (1510–1590), the *father* of wound management, used a semi-occlusive, oil-impregnated dressing in the 16th century to obtain “a softness of the tissues” (4). A Diachylon plaster consisting of a mucilagenous, moist mass made from linseed and marshmallow was originally described by Galen (130–201). This could be considered a precursor to our modern hydrogels (4). In 1819, Abraham Rees, in his treatise on the use of lynette, emphasized the need to prevent “scab” formation by keeping the wound edges apart with oil-soaked dossils, which were nonadherent and moisture retaining (5). In the 19th century, wet compresses or cataplasma kaolin were applied to wounds and covered in waterproof fabrics such as jaconet, battiste, or oiled silk to maintain humidity. The 20th century saw the production of leno gauze impregnated with soft paraffin, Tulle Gras.

The literature references to early wound management give detailed descriptions of the wound types but only a cursory description of wound dressing materials. The first authoritative monographs appeared in early London and Edinburgh hospital dispensaries and later in the British Pharmaceutical Codices. The development of wound management products can be traced by examining these Codices together with the *British Pharmacopoeia* (6). The information is reflected in similar publications in the *United States Pharmacopoeia* and other national standards.

Advances in the design and efficacy of wound management products were spasmodic and limited to the adaptation of available materials until 1960. Up to that date, the products were primarily of the *plug and conceal* variety, exemplified by lint, gauze, cotton, wool, and tow, which were considered to be *passive* products that took no part in the healing process. Little or no attention was paid to the functional performance of a product, and minimal consideration was given to the different healing environments required for different wound types.

The new generation of products was a rejection of the traditional passive *cover all* dressing philosophy and was potentiated by advances in knowledge of the humoral and cellular factors associated with the healing process and the realization that a controlled microenvironment is necessary if wound healing is to progress at the optimum level. Such environmental control dressings could be classified as *interactive*.

Both the acellular and cellular activities involved in the healing cascade are optimized by a wound microenvironment that will allow the free movement of cells and effective response to *bioactive* compounds. This optimal response can be expected where environmental factors such as temperature and humidity are at subdermal levels.

PERFORMANCE CRITERIA

The concept of moist wound healing is generally attributed to George Winter after his much cited publication in *Nature* (7), although Bull and his co-workers published results in 1948 (8) showing enhancement of healing under a ‘film’ dressing, but no explanation was given. In 1975, Winter expressed his initial concern with the secondary trauma produced at the wound surface by removing the strongly adhered fibrous and fabric dressings of the 1960s and described his attempts to produce a moist wound environment and, thereby, reduce dressing adherence by using polythene sheets as wound covers (9). This gave an unacceptable *wet* environment that resulted in tissue maceration. An alternative microporous polyvinyl chloride material that Winter reported as having a permeability

to water vapor of “about $2500 \text{ g m}^{-2} 24 \text{ h}$ ” proved more successful.

In 1979, Turner (10, 11) identified the performance criteria for a wound dressing product that would successfully contribute to an acceptable microenvironment. These were:

- to maintain a high humidity at wound/dressing interface
- to remove excess exudate and toxic components
- to allow gaseous exchange
- to provide thermal insulation
- to afford protection from secondary infection
- to be free from particulate or toxic contaminants
- to allow removal without trauma at dressing change

These criteria are still valid and have not been greatly modified. An additional requirement that results from our greater knowledge of the growth factors involved in the healing process is to be compatible with the humoral and cellular factors involved in healing.

Humidity Levels and Removal of Exudate

Partial or full thickness wounds exposed to the air will demonstrate a lower temperature than ambient due to the latent heat lost through tissue fluid evaporation. The clotting process and fibril development produces an occlusive dry eschar or scab that effectively seals and insulates the wound, thereby limiting moisture and gaseous transmission and restricting the migration of epithelial cells to the moist subscab tissue. The result is slow healing with a high contamination and infection risk, together with possible excessive scarring in an excised wound from closure without full cavity granulation.

The maintenance of a high humidity between the wound and the dressing is, therefore, a requirement for rapid epidermal healing. A drying wound will result in a gas-impermeable eschar and will require epithelial cell penetration to a moist lower level, resulting in prolonged healing time.

The absorption of excess exudate not only avoids tissue maceration but also removes exotoxins or cell debris that may retard growth or extend the inflammatory phase of the healing process. The balance between humidity and absorption is critical, and excessive wicking must be avoided to prevent drying and necrosis.

Gaseous Exchange

Gaseous permeability will allow water vapor transmission, which may be particularly important in a high exudate wound such as a burn, sacral, or high exudate leg ulcer. Of equal significance will be the effect of gaseous exchange on oxygen (pO_2) and hydrogen ion (pH) levels; movement

across the dressing of oxygen and carbon dioxide will directly relate to cellular and humoral factor activity. Epithelization of the wound is greatly accelerated by the availability of atmospheric oxygen, which dissolves in the serous exudate to supplement the oxygen transported to the wound area by haemoglobin and which is subsequently directly utilized by the migrating epidermal cells.

Thermal Insulation

Thermal insulation will assist in maintaining the wound temperature at a level as close to body core temperature as possible. Phagocytic and mitotic activity are particularly susceptible to temperatures below 28°C . Thermal insulation and *warm* dressing change conditions are very important if the optimum healing rate is to be maintained. A long exposure of wet wounds may reduce the surface temperature to the point where mitotic activity ceases and the recovery of that tissue may take up to 3 h. Temperatures of 30°C and above may be found beneath a good insulating dressing, and this will result in high mitotic activity with rapid epithelization and improved granulation (12).

Impermeability to Microorganisms

Bacterial impermeability has a dual role. One, the wound will not heal if it is heavily infected, and the inflammatory phase will be extended, and two, unless topical or systemic antibacterial agents are used, a more general infection could result. However, a limited number of microorganisms are tolerated by most wounds, and the destructive or cleansing phase produced by phagocytic activity should result in a self-sterilized environment. The wound should then be protected from secondary infection or, if still contaminated, be prevented from transmitting the infective organisms (Fig. 1).

A dressing should, therefore, be impermeable to airborne microorganisms that may fall on its surface and penetrate to and infect the wound. It should also act as a barrier to any wound organism that may be transmitted to the dressing surface and become airborne, thus causing cross-infection. Organism transmission occurs most frequently in dressings that exhibit *strike through* of the exudate to the wound surface, providing a wet pathway to or from the wound surface. The passage of organisms can take as little as 6 h from the time of *strike through*.

Freedom from Particulate and Toxic Wound Contaminants

Both particles and toxic compounds that may contaminate a wound will be responsible for disrupting the healing

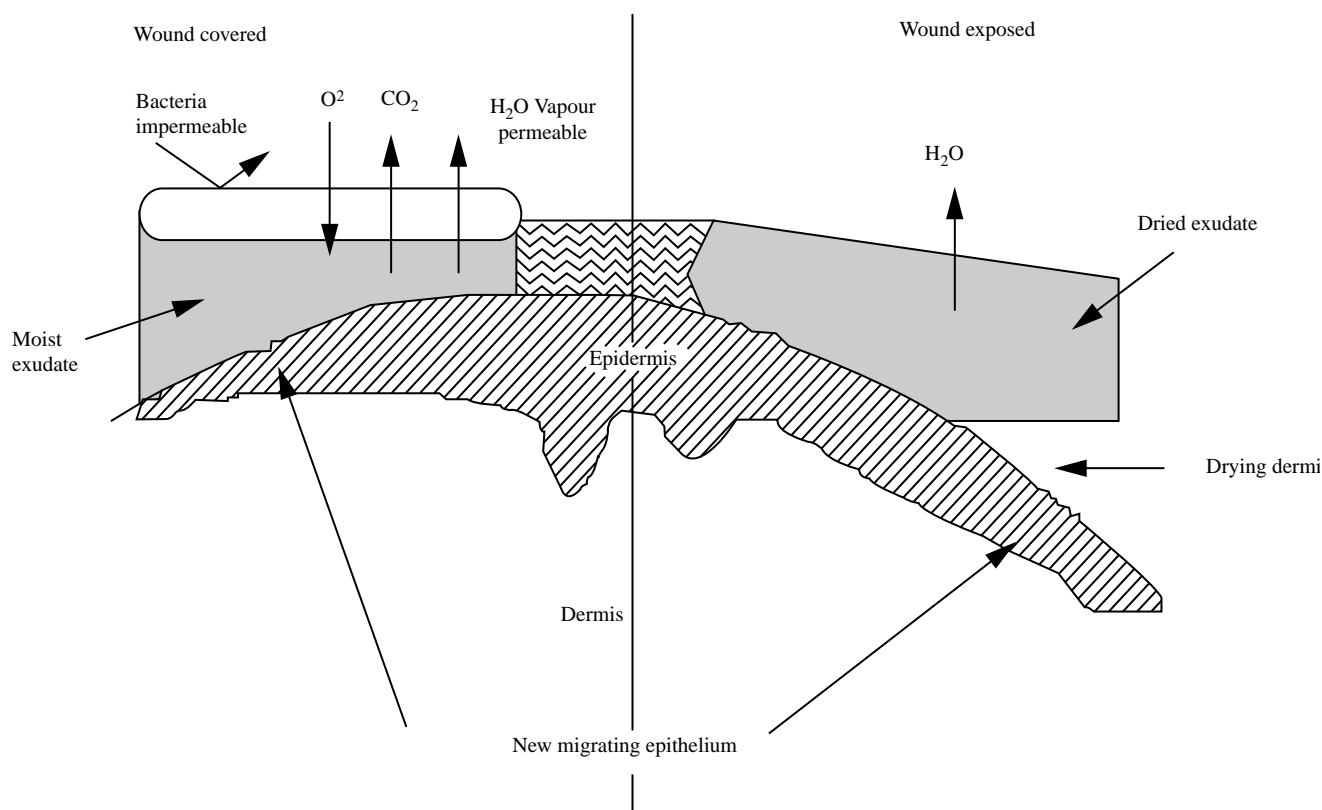


Fig. 1 Wound healing response to a controlled microenvironment.

pattern. The incorporation of fibrous particles into a wound may result in a granuloma, which could subsequently reduce the wound strength and induce keloid scarring. It is well documented that particulate contamination can also reduce the infection resistance levels by a factor of 1.0×10^{-6} . Thus, organism levels that previously would have been self-sterilizable become infective and colonize the wound to produce a gross infection.

Trauma during Dressing Change

The wound environment may be optimally maintained with a product that has the preferred performance parameters but, nevertheless, is disrupted during the dressing change. The hazards of temperature change and secondary infection may be accompanied by a secondary trauma caused by the dressing adhering to the wound and, on removal, stripping newly formed tissue. This adhesion is normally caused by the adhesiveness of the drying exudate, and the trauma can be exaggerated on removal by the destruction of capillary loops that have penetrated the dressing material. This produces a bleeding surface which,

at most, may revert the wound to the primary inflammatory phase and, at least, delay epithelization.

Although not associated with the production of an acceptable microenvironment, there are certain physical characteristics that are required to assist in the overall dressing procedure. The dressing should have (11): 1) a size range to match the wounds; 2) an absorption range for dry and heavy exudate wounds; 3) good conformability and good handle when both dry and wet; 4) sterility and be stable in storage; 5) ease in disposability.

These parameters stimulated the development of functional and well-designed products using the advances that had accrued in the technology of materials. As an example, two-layered systems were recommended, consisting of primary and secondary materials. The primary dressing should meet the requirements of permeability, nonadherence, and bacterial impermeability, and the secondary, the need for absorption, protection, and insulation. This development was closely followed by products that were derived from the advances in the development of synthetic polymers. A statement of required performance could now be considered as a possible specification for a polymeric product.

The range of polymeric products produced as surgical dressings have included vapor-permeable films, polymeric foams, particulate and fibrous polymers, xerogels and hydrogels, and hydrocolloids.

These materials mark the progression towards the production of an *ideal wound dressing*. It should, however, be emphasized that no *single* dressing will produce the optimum microenvironment for *all* wounds or for all of the healing stages of *one* wound. The spectrum of performance requires that the wound is diagnosed and the treatment progressed by prescribing the most suitable dressing at each stage of the healing process.

This article will now examine the range of dressings currently available, identify their principal chemical and physical characteristics, and indicate their recommended clinical usage.

ABSORBENTS

The overall function of surgical absorbents is self-explanatory. They are required to absorb and retain a wide range of fluids from the blood and serous exudate of damaged tissue to the variable gut content met during surgical intervention (10, 13). They are available in a number of forms: 1) Fibrous (staple) absorbents, 2) Fabric absorbents, 3) Fiber plus fabric absorbents, and 4) Wound dressing pads.

Fibrous Absorbents

These are made from cotton staple or from the fibers of viscose or cellulose; viscose and cotton may be admixed.

Absorbent cotton is available in different qualities varying with the length and diameter of the cotton staple. It is available in the form of rolls and balls and is used for cleansing and swabbing wounds, for preparing preoperative skin, and for applying topical medicaments to the skin.

The absorption performance and physical character of *absorbent viscose* varies markedly with the manufacturing process. It is available in the bright or dull form, the latter containing a particulate material, such as titanium dioxide, within the fiber. The fibers are, in general, a continuous staple with a crenate transsectional profile, but smooth and luminated forms are available that show different degrees of absorptive capacity and wet tensile strength.

Some fibrous absorbents contain a proportion of acrylamide or other synthetic polymeric fiber. They frequently enhance the absorptive performance and give body to the fleece, thus improving fluid retention and

avoiding squeeze out, which is caused by fleece collapse after wetting.

Cellulose wadding is produced from delignified wood pulp and manufactured in a multiple laminate material form. It is used in large pieces to absorb large volumes of fluid in incontinence but is not used in contact with a wound unless enclosed in an outer fabric sleeve to prevent fiber loss to the wound.

Fabric Absorbents

Absorbent lint is a close weave cotton cloth with a raised nap on one side that offers a large surface area for evaporation when placed with the nap upwards on an exuding wound. It is generally unacceptable in modern wound management.

Absorbent gauze is the most widely used absorbent and consists of a cotton cloth of plain weave that is bleached white, is clean, and is reasonably free from weaving defects and contains not more than traces of leaf residue, seed coat, and other impurities (6). It may be slightly off white if sterilized. It absorbs water readily, but its performance may be reduced by prolonged storage or exposure to heat.

Gauze products are primarily absorbents when used preoperatively, perioperatively, and postoperatively, but perioperatively, they are also required to perform other functions, including the protection of tissue and organs by occluding areas not involved in the procedure, to assist the application of wet heat, which may establish the viability of doubtful tissue, and to assist in blunt dissection where fascias are separated along the lines of cleavage, thus avoiding unnecessary cutting (14).

The gauze fabric may contain a proportion of viscose incorporated with the cotton either in the warp and the weft or exclusively in the weft. A maximum level of 45% of viscose is widely accepted. A range of gauze fabrics exist that are graded according to the number of threads per 10 cm width of gauze.

Gauze products fall into two broad categories—the “swab” or “sponge” type produced by folding and stitching the cloth and the “plain cloth” type. The “swab” type includes swabs, strips, pads, and pledgets. The plain type includes packs and ribbon. Gauze swabs or sponges are commercially available as gauze folded into rectangles or squares to give various sizes or ply. They are folded in such a way that no cut edges are visible and the edges may be stitched. For use in an operating theatre, they are available with and without a radio opaque (X-ray detectable) mono- or multi-filament thread containing barium sulphate woven into or heat bonded to the fabric.

Table 1 Absorbent gauze

Type	Threads per 10 cm		
	Warp	Weft	Weight (g/m ²)
13 Light	73	57	14
13 Heavy	70	60	17
17	100	70	23
18	100	80	24
20	120	80	27
22	120	100	30
24a	120	120	32
24b	140	100	32

The commercial product can vary in size from 5 × 5 cm to 10 × 15 cm with a variation in ply from 4 to 32. Some are colored with a suitable fast, nontoxic dye.

Nonwoven fabrics include a wide range of products manufactured from synthetic and semisynthetic fibers.

Nonwoven swabs consist of a nonwoven viscose fabric and are available in folded pieces of various dimensions. These are occasionally used in error as a single wound dressing. They have a lower total absorbent capacity than gauze but absorb more quickly because of the random orientation of the viscose fibers. As fabrics, they constitute the outer layer on a number of wound dressing pads and are sometimes suitably coated with a polymer to reduce adherence at dressing change.

The types as defined in Table 1 are derived from the *European Pharmacopoeia* (15) and the designation of type numbers is one-tenth of the sum of the threads in warp and weft.

Absorbent muslin is a bleached cotton cloth of open weave used infrequently for the treatment of extensive burns and as a wet dressing.

Cellulose sponge is a cavity foam cellulose-based sponge available in sheets and thin bands. It is used to absorb at small sites in surgery. The material is not radio opaque and has a tendency to lose particles. Additional precautions must, therefore, be taken if such surgical use is contemplated.

Neuropatties are small squares or strips of nonwoven absorbent viscose with thread stitched through the nonwoven fabric and left long. These are used as spot absorbents, particularly in neurosurgery. They are frequently moistened in saline before application. The threads are left outside the surgical area, and on completion of surgery, the recovery of each pattie is facilitated by lifting each thread. Products vary in size and shape, and there may also be a device for attaching the ends of all the threads, thus producing a mini count rack.

Fibrous and Fabric Absorbents

Gauze and cellulose wadding consists of a thick layer of cellulose wadding enclosed in a tubular form gauze. The properties of the two separate materials have already been described. The gauze and cellulose wadding tissue is used as an absorbent and protective pad. It should only be used as a wound dressing with a nonadherent layer placed between the pad and the wound. It has a high absorbency, and because of its thickness, the additional property of insulation, which results in raising the temperature at skin surface, has been shown to accelerate the wound healing rate. On a highly exuding surface, there is a tendency for the cellulose wadding element to collapse when wet and become a semisolid wet mass. This may cause difficulty in practice. In such high fluid loss situations, the gauze and cotton tissue is preferred.

Gauze and cotton tissue (Gamgee tissue) is a thick layer of absorbent cotton enclosed in a tubular form gauze. It has the same uses as gauze and cellulose wadding tissue but has the advantage of a higher absorbent capacity and less wet collapse. It is also softer in use and, thus, conforms more readily to the wound surface. It should be used in place of gauze and cellulose wadding tissue on high exuding surfaces, such as burns, but as previously stated for gauze and cellulose wadding tissue, it should not be used in direct contact with the wound surface but placed upon a nonadherent primary dressing.

Wound Dressing Pads

Wound dressing pads are widely available in a number of formulations, which include the fibrous and fabric absorbents previously described plus other materials combined to meet some aspects of the acceptable performance profile. The pads can be subdivided into the following categories:

- absorbents and filmated products
- sleeved pads with a single layer core
- sleeved pads with a multiple layer core
- low adherence pads.

The absorbents and filmated products have been described. The simplest sleeved pads contain cotton, viscose, or cellulose fiber with an outer sleeve of gauze or nonwoven material. Those with a multilayer core have an outer sleeve of cotton, viscose, or nonwoven fabric that may have been treated with a polymer, such as polypropylene, to reduce adherence. The multilayer core is designed to increase absorptive capacity and to prolong usage by delaying *strike-through* to the outer surface. This is facilitated by using a fluid-retardant layer within the

upper and outer sleeve, which encourages lateral movement of fluid within the pad.

Low-adherence pads have wound contact faces designed to be of low adherence. They vary from aluminum-coated fabrics to perforated polymeric films or heat-bonded polyethylene films. The wound contact film may be attached to an absorbent fibrous mat and an outer woven or nonwoven fabric. In some products, the polymeric film forms a continuous sleeve on both dressing surfaces. They are dressings for low exudate and drying wounds where high adherence can be expected. These low-adherence, low-absorptive capacity dressings are sometimes centered on an adhesive backing to produce an *island* dressing, which is presented as a postoperative adhesive dressing or in the more familiar form of a first aid island or strip dressing for superficial injuries.

Low-adherence primary dressings consist of a partially open cellstructured nylon or viscose fabric that may be finished with a silicone coating. The open cell structure allows fluid transmission to a superimposed absorbent dressing pad. The pad is changed when required without disturbing the primary contact layer.

Deodorizing Dressings

Deodorizing dressings have been formulated from high gaseous sorbative activated-charcoal presented either as a woven fabric or a fibrous mat backed by a nylon sleeve, as a vapor permeable film, or as a polyurethane foam. In each formulation, the objective is to reduce odor, and the dressings must, therefore, be large enough to cover the entire malodorous area. One product encourages direct contact of the carbon layer with the wound exudate and whereas this will limit gaseous absorption, it is claimed that the incorporation of bound silver into the charcoal cloth inactivates bacteria that are absorbed onto the fabric surface, thus reducing the infective level and resulting in a reduction of odor.

POLYMERIC DRESSINGS

Vapor-Permeable Films

Vapor-permeable films are of use in those wounds in which granulation tissue is established and wound exudate is declining (16). These products have been developed as materials that would, in part, mimic the performance of skin. The resultant products were transparent, synthetic, adhesive films generically described as vapor-permeable adhesive membranes and consisting of transparent polyurethane or other synthetic film of low reflectance

evenly coated on one side with a synthetic adhesive mass. The adhesive is cohesive and inactivated by contact with moisture and will not, therefore, stick to moist skin or the wound bed. The films are permeable to water vapor, oxygen, and carbon dioxide but occlusive to water and bacteria. They have highly elastomeric and extensible properties that contribute to both their conformability and their resistance to shear and tear. The products are sterile and particle free.

The films also possess permeability functions that are essential to their efficacy as wound management material. It should be noted that the removal of the stratum corneum results in a water vapor loss from tissues of between 3000 g m² and 5000 g m² over a period of 24 h. This loss will result in progressive dehydration that could be of great significance, particularly in a full thickness burn. The loss through a positioned vapor-permeable membrane is reduced to 2500 g m² over 24 h or less, depending upon the structure of the membrane. This allows excess fluid to be lost by water vapor transmission through the membrane but prevents dehydration and maintains a moist wound interface. Where the volume of exudate produced is significantly greater than the volume removed as vapor, the water impermeability will result in serous effusion accumulating below the film. Impermeability to water prevents wetting from external sources.

The importance of a moist interface to wound healing is now well recognized. It allows the rapid migration of new epithelium across the wound surface, precludes trauma due to adherence at dressing change, and contributes to gaseous diffusion in the damaged tissue. Oxygen and carbon dioxide transfer are accomplished by intramolecular diffusion through the membrane and by solution in the wound surface moisture. The oxygen permeability of the films is variously described as 4000–10,000 cm³ m² 24 h at ambient temperature. The pO₂ and pH levels of the wound surface are directly related to the gaseous permeability and contribute to cellular activity. The wound is protected against secondary infection by the bacterial impermeability of the film to such organisms as *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Eschericia coli*.

The physical performance is applicable to the management of superficial tangential wounds, such as dermabrasions, split skin graft donor sites, and burns. In a dermabrasion, haemostasis must first be obtained and the margin of the wound dried before the film is applied. Providing the film is correctly positioned, it may be left in situ until epithelization is complete. In its application for the treatment of burns, careful disinfection must precede the positioning of the film. It is only recommended for superficial and clinically clean burns and contraindicated

for deep burns where it retards the separation of necrotic tissue.

Decubitus ulcers and pressure sores can be covered with a vapor-permeable film, with the added advantage that the films' resistance to shear and low frictional surface properties protect the dermal layers from additional physical abrasion while producing the minimal barrier to normal skin function. This performance allows the film to be used as a prophylactic in areas that are traumatized by pressure but not ulcerated.

The film dressings can also be used for the retention of canulae and tubes in both ward and theatre. Specific products have now been produced with a variable water vapor permeability to reduce the build-up of moisture beneath the film and the resultant infective hazard.

In theatre, the large, synthetic, adhesive vapor-permeable films were originally used as incise drapes to protect tissue against additional trauma and were used postsurgery as a direct dressing over suture lines, allowing postoperative inspection without disturbing the area.

Polymeric Foams

These materials demonstrate the greatest variability between products. At their simplest, they are foamed polymers that have been made into sheets. The wound contact layer is often heat treated to give a smooth surface that absorbs fluids by capillarity. Foaming the polymer creates small, open cells that are able to hold fluids, and the cell size may be controlled during the foaming process. The most common polymer used is polyurethane. Their structure and softness also provide a cushion that protects and contributes to thermal insulation of the wound. They also may be tailored for particular applications, such as tracheostomy dressing, without particle loss to the wound and with the retention of their conformable characteristics. The nonadhesive foams will require a secondary dressing.

The dressings are available as sheet dressings and as *in situ* formed foams. They have a nonadherent wound contact surface and are also available as adhesive island dressings and cavity fillers (16). They are absorbent dressings used for wounds with moderate-to-heavy exudation.

The sheet dressings are, in the main, polyurethane foams where the absorbency and water vapor permeability are varied either by a physical modification to the foam or by combining the foam with an additional sheet component.

An early foam dressing consisted of a hydrophobic polyurethane foam, where one surface had been heat- and pressure-modified to produce a hydrophilic porous membrane about 0.5 mm thick. The outer surface of the

dressing consists of a layer of relatively large cells of approximately 5 mm thick and remains hydrophobic. Absorption of serous exudate is limited to the wound/dressing interface. In use, the absorptive capacity of the hydrophobic portion will be exceeded in a high exudate wound, and although moisture vapor transmission occurs through the dressing, frequent changes may be required until the exudate level diminishes. The superpositioning of a secondary absorbent pad will solve the problem of excess exudate while retaining the nonadherence, bacterial impermeability, gaseous and water vapor permeability functions and the additional cushioning and thermal insulation properties of the foam. The dressing combines the function of absorbency with that of producing an acceptable microenvironment that will allow healing to take place at the fastest rate concomitant with the total clinical condition of the patient.

Polyurethane foam dressings of this type are recommended for the management of dry sutured wounds, minor lacerations, early pressure ulcers, and venous ulcers (17, 18).

A partially expanded, modified polyurethane foam was developed by Lock (12). It comprised a lower layer of open cells and an upper hydrophilic surface with closed impermeable layers that reduce the loss of water vapor and prevent strike through of absorbed fluid. This primary dressing expands when it becomes wet and conforms to the contours of the wound, producing an environmental chamber with entrapped solutes and cell debris. It is claimed that this function enhances the inflammatory response of the wound and, subsequently, stimulates the production of granulation tissue and revascularization. These polyurethane membranes are recommended specifically for the management of stasis ulcers, with a superimposed absorbent pad and graduated pressure applied either by stretch bandages or elasticated stockings.

A foam dressing with the prime function of absorbency has been designed for the management of burns. It consists of a highly absorbent hydrophilic polyurethane foam backed with a moisture-permeable polyurethane membrane and bonded to an apertured polyurethane net on the wound contact face. It is capable of absorbing and retaining large volumes of fluid even under pressure. The backing, while permeable to water vapor, is impermeable to water, thus avoiding strike through. As the exudate level decreases, the membrane retains moisture and prevents the drying of the wound. The apertured polyurethane net interface reduces adherence to the wound surface. Whilst recommended for burns, these dressings have been used successfully on other exuding lesions (19).

Low-absorptive capacity, primary foam dressings have been produced from a carboxylated styrene

butadiene rubber latex foam. The foam is bonded to a nonwoven fabric coated with a polyethylene film that has been vacuum ruptured. The basic foam is naturally hydrophobic, and a surface active agent is incorporated to facilitate the uptake of wound exudate. The polyethylene foam layer is particularly effective in preventing adherence, and the dressing is recommended for minor wounds and abrasions where exudate levels are low and adherence a prominent hazard at dressing change.

In Situ Foam

One of the major problems in wound management is the treatment of large cavity wounds produced either postoperatively, such as a pilonidal sinus, or by trauma, such as pressure ulceration. It is necessary to occlude the cavity by packing to absorb excess exudate and to stimulate the production of granulation tissue, neovascularization, and collagen deposition.

The traditional procedure is to pack the cavity with ribbon gauze (see Fibrous Absorbents) variously impregnated. The subsequent removal of such a dressing is difficult, and the pain and stress associated with the dressing change may require low-level anesthesia and theatre management.

An in situ-formed foam was originally designed by Dow Corning and found to be clinically superior to ribbon gauze (20). Its status in cytotoxic terms was in dispute, but it is now available as Cavi-Care™ (Smith & Nephew), and its exclusion would be to the detriment of this chapter.

Cavi-Care foam dressing is a two-part foam composed of a filled poly(dimethylsiloxane) base and a stannous octonate catalyst. The two components are mixed together immediately prior to use. The reaction is slightly exothermic and over a period of 2–3 min, the dressing expands to approximately 4 times its original volume and sets to a soft, spongy foam accurately conforming to the contours of the wound cavity. The stent is normally removed twice daily, soaked in a mild antiseptic (0.5% aqueous chlorhexidine), rinsed in cold running water, squeezed dry, and replaced. A new dressing is formed after a week or more to match the reduction in size of the cavity. The product does not adhere to granulation tissue. It maintains free drainage around the wound and has a low, but significant, absorptive capacity at the dressing surface. It has been indicated for the management of pilonidal sinus, hydradenitis suppurativa, perianal and perineal wounds, and in the management of dehiscent abdominal wounds.

Hydropolymer

Hydropolymer appears visually as a foam but is described as a foamed gel. The material expands into the contours of the wound as it absorbs fluid, and is used in an island configuration with an adhesive portion, which is unique. It has the ability to re-adhere once lifted, enabling manipulation of the product for fit or assessment of the wound without dressing change. The hydropolymer wicks fluid into the upper layers of the dressing where it escapes through the backing (21).

Its recommended use is for dynamic fluid management for heavily exuding wounds or where extended periods between dressing changes are desirable.

Hydrogels

Hydrogels, or water polymer gels, are modified, cross-linked polymeric formulations that form 3D networks of hydrophilic polymers prepared from materials such as gelatin, polysaccharides, cross-linked polyacrylamide polymers, polyelectrolyte complexes, and polymers or copolymers derived from methacrylate esters. These interact with aqueous solutions by swelling to an equilibrium value and retain a significant proportion of water within their structure. They are insoluble in water. Hydrogels are available in dry or hydrated sheets or as a hydrated gels in applicators. When hydrated, they contain up to 96% water and have additional, high, absorption properties.

The physical properties of bulk polymers are directly influenced by a number of factors including the nature of monomers, copolymers, the cross-linkers and the degree of cross-linking and the polymerization initiators and processing parameters. The availability of hydrophilic and hydrophobic sites is influenced by the chain of configuration and conformation and could determine the degree of hydrophilicity and oxygen permeability. By varying the nature of the polymer backbone, a range of water-binding behavior and, thus, mechanical, surface, and permeability properties can be obtained. The expanded nature of the hydrogel structure and its permeability allows the extraction and polymerization of initiator molecules, initiator decomposition products, and other extraneous materials from the gel network before the hydrogel is placed in contact with the living system. The tissue-like structure of most hydrogels will contribute to their biocompatibility by minimizing mechanical irritation to surrounding cells and tissues. They possess low interfacial free energies with aqueous solutions and only a weak tendency to absorb biological species, such as proteins or cells. Their high moisture

content maintains a desirable moist interface that facilitates cell migration and prevents dressing adherence. Water can be transmitted through the saturated gel; whereas the unsaturated gel will have a water vapor permeability comparable with the water vapor permeability of vapor-permeable membranes.

The absorption, transmission, and permeability performance result in the maintenance of a moist wound with a continuous moisture flux across the dressing and a sorption gradient that assists in the removal of toxic components from the wound area. The high moisture content allows dissolved oxygen permeability, which varies between products but will allow the continuation of aerobic function at the wound/dressing interface and have an effect upon both epithelization and bacterial growth. It has been observed that the positioning of a hydrogel frequently results in a marked reduction in pain response in patients. It is suggested that the high humidity protects the exposed neurones from dehydration and also produces acceptable changes in pH. A secondary effect that may contribute to this response is the property of the gels to immediately cool the wound surface and maintain a lower temperature for up to 6 h (22). This lowering of temperature could result in a reduction of the inflammatory response.

Sheet hydrogels

Sheet hydrogels are sheets of 3D networks of cross-linked hydrophilic polymers (polyethylene oxide, polyacrylamides, polyvinylpyrrolidone, carboxymethylcellulose, modified corn starch). Their formulation may incorporate up to 96% bound water, but they are insoluble in water and interact by 3D swelling with aqueous solutions. The polymer physically entraps water to form a solid sheet, which may make them feel moist but compression of the sheet will not release any water. They have a thermal capacity that provides initial cooling to the wound surface. A secondary dressing is required.

The recommendation for use of these products includes the management of donor sites and superficial operation sites and also the treatment of fresh, chronic damaged epithelium, such as thermal and other painful wounds, and dermatitic skin where the avoidance of topical agents is indicated. In chronic ulcers, they are used to encourage granulation and formation of cellular tissue.

Amorphous hydrogels

Amorphous hydrogels have additional ingredients, such as alginate, collagen, or complex carbohydrates, besides water and a polymer. They are similar in composition to sheet hydrogels, but the polymer has not been cross-

linked. These amorphous preparations do not have the cooling properties of the sheet dressings, and a secondary dressing is required. Their recommended use includes hydration of dry, sloughy, or necrotic wounds and autolytic debridement.

One of the earliest hydrogel dressings came from Hydro Med Sciences, Inc., NJ, USA. It is formed directly on the wound by using a two-component spray system, consisting of p(HEMA) and polyethylene glycol (PEG). Alternate layers are applied to the wound. The PEG dissolves the p(HEMA), forming a saturated solution that solidifies after 30 min (23).

A primary hydrogel dressing has been produced from a colloidal suspension of radiation cross-linked polyethylene oxide and water with an equilibrium water content of 96%. The gel is sandwiched between polyethylene films. The wound contact film is routinely removed, and the outer film is used to control evaporation from the gel surface (24, 25). A novel hydrogel was developed by the Max Plank Institute for Immunobiology and Dermatology consisting of an insoluble cross-linked polyacrylamide agarose polymer containing 95% water as the disperse phase (26). The gel has a high elasticity and tensile strength and a high capacity to absorb exudate from wounds. It is available in hydrated and dehydrated forms and as granules which, with their increased surface area, absorb larger amounts of exudate. The granules can be used to fill a cavity wound with the gel sheet superimposed to produce a continuous hydrogel dressing.

An alternative to the acrylamide-based composite hydrogel dressing is an acrylamide grafted to a polyurethane film to give a transparent, flexible gel with an equilibrium water content of approximately 50%. The hydrated gel has a low modulus of elasticity and a high water permeability, which facilitates the adherence of the gel to the wound surface. The permeability characteristics allow penetration of antimicrobial agents, which can be applied topically to the dressing surface in situ.

Particulate and Fibrous Polymers

This group of dressings includes synthetic, semisynthetic, and naturally occurring products embracing a range of polysaccharide materials.

Xerogels

The xerogel dressings may be regarded as a subgroup of products within the larger group of polysaccharide dressings. The latter contains the well-known cellulosic

dressing products, such as gauze and absorbent cotton (see Absorbents section), but the products that consist of dextranomer beads, dehydrated hydrogels of the agar/acrylamide group, calcium alginate fibers, and dehydrated granulated Graft T starch polymers are identified specifically as xerogels (27), the material remaining after the removal of most or all of the water from a hydrogel (or the disperse phase from any type of simple gel). These materials, therefore, have no water in their formulation but swell to form a gel when in contact with aqueous solutions.

Particulate Polymers

Dextranomer

This xerogel is a polymer of the polysaccharide dextran, a naturally derived polymer of glucose produced by cultures of a microorganism, *Leuconostoc mesenteroides*. The gel is formed when the dextran molecules comprising the disperse phase of the hydrocolloid are cross-linked by a chemical process utilizing epichlorhydrin and sodium hydroxide (27). Dextranomer is available as beads or paste. The material requires a secondary dressing.

The dextranomer is supplied in beads of 100–300 μm diameter containing poloxamer 187, polyethylene glycol 300, and some water. A paste formulation is also available, which is the dextranomer in polyethylene glycol 600 (PEG 600). The beads are offered as a discrete particle or enclosed in a low adherence pouch for insertion into a cavity wound. One company (Pharmacia & Upjohn) offers a polymeric net that can be placed into a cavity wound before the addition of either granules or paste and facilitates removal and also a vapor-permeable film that is superimposed on the dextranomer dressing to control evaporation and retard the drying of the dextranomer in a low exudating wound. The hydrophilic beads will absorb the aqueous component of wound exudate and dissolved materials ranging from inorganic salts to low molecular weight proteins. Dextranomer has a pore size that produces an exclusion limit of 1,000–8,000 Da, which precludes the sorption of viruses and bacteria. Microorganisms are removed from the wound by the capillary action between the beads, a function that is absent from the paste formulation which, however, demonstrates a marked increase in absorbing capacity for malodorous elements and pain producing compounds released during the inflammatory response.

Dextranomer is used primarily as a debriding agent on sloughy and exuding wounds, whether clean or infected, and on small area burns, where the objective is to produce

a clean tissue bed for the production of a granulating tissue. It is not a product that should be used beyond this phase as its continued application will impair epithelization. Dextranomer is *not* biodegradable, and both granules and paste must be carefully removed with Normal saline before drying to avoid particulate residues and the subsequent development of granulomas.

The use of dextranomer on exuding necrotic injuries, such as leg ulcers, has proved to be beneficial, and it is an acceptable substitute for surgical or enzymatic debridement. The granules are contraindicated in drying wounds.

Fibrous Polymers

Alginate dressings

Alginic acid is a polyuronic acid composed of residues of D-mannuronic acid and L-guluronic acid and is obtained chiefly from algae belonging to the Phaeophyceae, mainly species of *Laminaria*.

Calcium alginate dressings are flat, nonwoven pads of either calcium sodium alginate fiber or pure calcium alginate fiber. The alginate wound contact layer may be bonded to a secondary absorbent viscose pad. Alginate hanks are also available, as are packing and ribbon for deeper cavity wounds and sinuses. Alginates have been shown to be effective in the management of injuries where there has been substantial tissue loss. The nonadhesive formulations require a secondary dressing.

Gel formation is via ion exchange. A biodegradable gel is formed when the fiber is in contact with exudate and the calcium contributes to the clotting mechanism. The gel may be firm or soft, depending upon the proportions of calcium and sodium in the fiber. The gel is removed with saline.

The isomeric acids are present in varying proportions, depending upon the seaweed source. Calcium alginate is capable of gel formation. The guluronic acid forms an association with calcium, providing the stimulus to produce the continuous disperse phase of a hydrogel. Ca^{2+} ion and a phospholipid surface promote the activation of prothrombin in the clotting cascade, and calcium alginate products are used as the source of these ions to arrest bleeding, both in superficial injuries and as an absorbable haemostat in surgery. The rate of biodegradation is related to the sodium/calcium balance in the preparation.

The alginates are produced in fiber form and presented as a fleece or layered needled fabric. When applied to a bleeding surface, both the availability of the Ca^{2+} ions and the fibrous matrix contribute to coagulation and serum absorption produces a gel-like mass. The dressings may be

removed with a sterile 3% sodium citrate solution, followed by washing with sterile water, or they may be removed with sterile Normal saline.

The *wet* integrity of the dressing, which facilitates removal from the wound, may be improved by incorporating fibers of greater strength, such as viscose (rayon) staple fiber, or fibers that interact with the alginate fibers when wet, such as chitosan staple fibers (28).

The primary haemostatic usage of calcium alginate is in the packing of sinuses, fistulae, and bleeding tooth sockets. The alginate dressings have recently become widely used as a soluble wound packing for a number of additional wound types. They have been used as a useful nonadherent for lacerations and abrasions and are effective in the management of hypergranulation tissue (proud flesh), interdigital maceration, and heloma molle. The hospital and community use of calcium alginate includes intractable skin ulcers and pressure sores, where they would appear to accelerate healing, and in the successful management of diabetic ulcers, venous ulcers, burns, and infected surgical wounds.

Alginates have proved to be useful debriding agents. When applied to these injury types, the alginate must be covered by a secondary dressing of foam or film. Some proprietary products bond calcium alginate to a secondary backing, such as absorbent viscose pad or semipermeable adhesive foam, to produce an island dressing.

Hydrocolloids

Hydrocolloid dressings consist of composite products based on naturally occurring hydrophilic polymers. Generally, these dressings are flexible, highly absorbent, occlusive or semiocclusive adhesive pads formulated from biocompatible hydrophilic polymers, such as sodium carboxymethylcellulose, hydroxyethylcellulose, pectins and gelatin incorporated into a hydrophobic adhesive. The dressings may be backed by a polymeric film and may be contoured to fit difficult areas. Hydrocolloids are also available as pastes or powders or gels. The pads do not require a secondary dressing.

In general, hydrocolloids consist of first, a pressure-sensitive, adhesive layer, composed of a so-called *hydrocolloid* dispersed with the aid of a tackifier in an elastomer and secondly, a film coating, composed of a gas-permeable but water-impermeable, flexible, elastomeric material. One of the first hydrocolloid dressings described had a pressure-sensitive, adhesive hydrocolloid layer which consisted of 40–50% by weight of a mixture of gelatin and sodium carboxymethylcellulose dispersed in polyisobutylene with an antioxidant and a tackifier (mineral oil and terpene resin). This was then laminated

with a semi open cell, flexible, polyurethane foam that had previously been laminated with a flexible, polyurethane film. A currently available hydrocolloid dressing is a flexible mass with an adherent inner face and an outer semipermeable polyurethane foam.

The modified formulation is as follows:

Sodium carboxymethylcellulose	20%
Polyisobutylene	40%
Gelatin	20%
Pectin	20%

The product is also available as granules of similar formulation, allowing a continuous *fill* for cavity wounds.

Other hydrocolloid dressings, with formulations consisting of sodium carboxymethylcellulose combined with karaya gum or sodium carboxymethylcellulose on its own, are also available.

The adhesive formulation of hydrocolloids gives an initial adhesion higher than some surgical adhesive tapes. After application, the absorption of transepidermal water vapor modifies the adhesive flow to maintain a high tack performance throughout the period of use. In situ, the dressings provide a gaseous and moisture-proof environmental chamber that strongly adheres to the area surrounding the wound, offering protection against contamination from incontinence or other sources. In the wound contact area, the exudate is absorbed to form a gel that swells in a linear fashion with a higher moisture retention at the contact surface. This results in an expansion of the gel into the wound cavity with the continued support and increasing pressure from the remainder of the elastomeric dressing. The larger the volume of exudate, the greater the expansion into the cavity, up to the limitation imposed by the availability of the gel. The advantage of this system is that it applies a firm pressure to the floor of a deep ulcer, a basic surgical maxim for the production of healthy granulating tissue. It is this function that contributes to the recommended usage for venous ulcers (29).

The formed *colloidal* gel also produces a sorption gradient for soluble components within the serous exudate and allows the removal of toxic compounds arising from bacterial or cellular destruction. The moist gel is soft and conforms to the wound contours. When the dressing is removed, the gel remains in the wound and can be washed away with Normal saline. During use, the dressing in contact with the wound liquefies to produce a pus-like liquid with a somewhat strong odor. The hydrocolloids are suitable for desloughing and for light-to-medium exuding wounds, but are contraindicated if an anaerobic infection is present. They have been used successfully in the

treatment of chronic leg ulcers, pressure ulcers, and as skin barriers in the management of stoma.

As with the hydrogels, the products can be obtained in both powder and paste form, where the powders and pastes have similar formulations to that of the hydrocolloid mass in the sheet dressing. This will allow larger cavity wounds and heavily exuding wounds to be treated with a continuous hydrocolloid system. The recommended use for these dressings is for the treatment of pressure ulcers, minor burns, granulating wounds, and wounds exhibiting slough or necrotic tissue, or wounds with moderate exudate.

Superabsorbents

Superabsorbents are hydrocolloidal compounds that have a high absorbent capacity and entrap exudate so that it cannot be squeezed out once absorbed. One product incorporates this material into an island pad that is covered by a nonwoven absorbent and surrounded by an extra-thin hydrocolloid as the adhesive portion. The covering acts as a transfer layer while its surface stays dry. Superabsorbents are recommended for use with heavily exuding ulcers (16).

Hydrofibers

Hydrofibers are fibers of carboxymethylcellulose formed into flat, nonwoven pads with the facility to form a gel in contact with fluid. The absorbent rate and capacity is approximately 3× that of calcium alginate. The resultant gel is similar to a sheet hydrogel, which does not dry out or wick laterally. Therefore, there is no maceration of the skin surrounding the wound, but moisture is maintained in contact with the wound bed. The high absorbent capacity reduces the frequency of dressing changes. Hydrofibers are recommended for use with heavily exuding wounds or wounds where an extended wear time is desired (16).

Impregnated Dressings

Impregnated dressings coat the area of the wound with hydrophobic paraffin spread on an open mesh gauze, thereby providing both insulation and partial occlusion but allowing excess exudate to be absorbed by a superimposed absorbent pad. Impregnated dressings have the advantage of low adherence and allow gaseous diffusion, but the disadvantages include the incorporation of the soft paraffin or loose cotton fibers into the healing wound, thereby producing a *foreign body* response within the wound, leading to an extended inflammatory phase with

consequent delayed wound healing. These products also, when used excessively, retain wound exudate causing maceration to the surrounding, otherwise healthy, tissue.

Close weave gauze and open weave tulles are used as carriers of medicated and unmedicated ointments to the wound surface.

Paraffin gauze (Tulle) dressing

Paraffin gauze is bleached cotton or a combination cotton and rayon cloth impregnated with yellow or white soft paraffin. It is available as sterile single pieces or multipacks. The paraffin is present to prevent the dressing adhering to a wound. The gauze, which may be leno in nature, is coated so that all the threads of the fabric are impregnated but the spaces between the threads are free of paraffin. The material is used primarily in the treatment of wounds such as burns and scalds where the protective function of the stratum corneum is lost and water vapor can escape from the body surface. Paraffin gauze dressing functions by reducing the fluid loss while the water barrier layer is reforming. The two properties of the paraffin gauze that are most useful are those of nonadherence and semi-occlusiveness.

In addition to its application to burns and scalds, paraffin gauze dressing is used as a wound contact layer in lacerations, abrasions, and in ulcers, where it is used as a packing material to promote granulation. Postoperatively, it is used as a vaginal or penial dressing and for sinus packing.

Povidone iodine 10% w/w, chlorhexidine 0.5% w/w, sodium fusidate 2% w/w, and framycetin sulfate 1% w/w are some examples of available impregnations that are recommended for the reduction of infection. Diffusion of the antibacterial agent into or onto an infected and exuding wound has been shown to be minimal. The possibility of development of resistant strains of infective organisms has reduced the usage of these products.

Biodressings

Biodressings are composed of materials that are almost exclusively originated from living tissue and are said to “participate actively and beneficially in the biochemistry and cellular activity of wound healing” (30). They can be identified as biological dressings and biosynthetic dressings and the biological dressings can be further subdivided into natural or cultured, depending on their origin.

Collagen dressings (biosynthetic dressings)

Collagen dressings are formed from denatured collagen peptides, derived from pig skin, and purified bovine hide collagen, cross-linked with the glycosaminoglycan,

chondroitin-6-sulfate. When applied to damaged tissue, the dressings stimulate the production of fibroblasts and endothelial cells while accelerating the migration of epithelial cells. These biological components are combined with silicone elastomers to control moisture vapor loss during the accelerated growth period (31).

The bovine material used for the extraction of the Type-1 collagen used in these dressings is nonantigenic due to enzymatic purification. It is available as sheets, particles, pastes, or gels. The dry materials absorb exudate to form a gel, and it is believed that the addition of collagen to a wound bed accelerates wound repair by the provision of a matrix for cellular migration (16). One manufacturer has incorporated 10% alginate in the dressing formulation. The materials all require a secondary dressing and are recommended for use on any recalcitrant wound.

Natural biological dressings

Autologous (self) skin is harvested from a healthy area of the patient's body and transferred to the wound, generally a burn. The procedure leaves a second injury or donor site. The advantages of nonallergenicity, nontoxicity, nonpyrogenicity, and direct incorporation into the healed area with the overall performance parameters of intact skin make this the dressing of choice. Skin harvested from fresh cadavers and used as an allograft is a possible alternative to autologous skin. This skin undergoes processing to remove fibroblasts, endothelial cells, and epidermis, which would stimulate an immune response. The resulting product is an acellular, dermal collagen matrix that is immunologically inert but retains elastin, proteoglycans, and the basement membrane complex (31).

Porcine skin: Sterile, denatured lyophilized skin of porcine origin is produced and consists of the dermal and/or epidermal layers. The material is reconstituted by immersion in sterile water or saline and applied dermal side to the wound. A bovine preparation is also available. It is used as a temporary dressing in burns and ulcers, particularly where a site is being prepared for grafting.

Cultured biological dressings

Cultured biological dressings are derived from mammalian cell culture procedures that facilitate the development of sheets of cells derived from a few of the patients' own epidermal cells. Cultured human keratinocytes form cultured epithelial grafts which are not bioengineered tissue but are cultured cell products with limited use. These cultures may be regarded as a precursor, which has led to the development of other products through bioengineering (31).

Bilayered skin equivalent: A bilayered composite skin equivalent has been developed with a viable dermis and epidermis. The epidermis is composed from cornified differentiated keratinocytes and a dermal matrix composed of a collagen lattice containing viable fibroblasts. Its cellular components assist with wound closure through stimulation of the wound bed. The outer layer of the differentiated bilayered skin equivalent, the stratum corneum, acts as a specialized vapor-permeable membrane and protective outer barrier (31).

Human dermal replacement: Human dermal replacement requires cultivation of human diploid fibroblasts on a 3D polymer scaffold. The cells are derived from newborn foreskin and are living cells that are metabolically active following implantation into the wound bed (31).

Bioactive Dressings

The Introduction described the performance parameters of *interactive* dressings which distinguished them from the *passive* products. Current developments have confirmed that the next generation of products will participate actively in the wound healing process by contributing growth hormones, chemotactic agents, angiogenic agents, and other growth factors, either in a depot release mode or as sequentially released compounds (32), as well as controlling the microenvironment surrounding the wound.

Each wound management product will eventually be designed to meet the environmental, nutritional, and growth requirements of particular wound types and will probably be based on those biodressings described earlier. The group will be designated *bioactive* wound management products (33). These biodressings must be the precursors for the development of many more exciting products, which will improve the morbidity of wound healing to the advantage of both patient and clinician.

CONCLUSION

The content of this chapter has been restricted to those products that are in direct contact with the wound. There are many other products outside this limitation that are used successfully for wound management. These include surgical adhesive tapes and nonextensible, conforming, and elastic net bandages, used to retain dressings in position, and the extensible bandages, which may vary from the light support and compression products for the management of sprains and strains and the prevention of

edema to the high compression bandages used either alone or superimposed on various dressings to apply pressure to a limb.

REFERENCES

- Breasted, I.H. *Edwin Smith Surgical Papyrus Facsimile*; University of Chicago Press: Chicago, 1906.
- Najano, G. *The Healing Hand*; Harvard University Press: Cambridge, MA, 1975.
- Mackinney, L. *Medical Illustrations in Medical Manuscripts*; Clowes and Sons: London, 1965.
- Packard, F.R. *Life and Times of Ambroise Paré*; Milford: London, 1922.
- Elliot, J.R. Surgical Materials. *St. Barts Hosp. J.* **1954**, 58, 11–14.
- British Pharmacopoeia*; HMSO: London, 1998.
- Winter, G.D. Formation of the Scab and the Rate of Epithelialisation of Superficial Wounds in the Skin of the Young Domestic Pig. *Nature* **1962**, 193, 293–294.
- Bull, J.P.; Squire, J.R.; Tophey, E. Experiments with Occlusive Dressings of a New Plastic. *Lancet* **1948**, 2, 213–214.
- Winter G. D. In *Epidermal Wound Healing Proceedings Conference on Surgical Dressings in the Hospital Environment* Turner T. D. Brain K. R. SDRU, UWIST Cardiff 1975; 47–83
- Turner, T.D. Hospital Usage of Absorbent Dressings. *Pharm. J.* **1979**, 22, 421–426.
- Turner, T.D. *Proc. Symp. Wound Healing Helsinki*; Sundell, B., Ed.; 1979, 75–84.
- Lock, D.M. *Proc. Symp. Wound Healing Helsinki*; Sundell, B. Ed.; 1979, 103–108.
- Turner, T.D. Surgical Absorbents. *Africa Health* **1979**, 1 (4), 22–24.
- Turner, T.D. Absorbents in Surgery. *Proc. Guild Public Pharm. Conf. York* **1977**, 56–59.
- European Pharmacopoeia*, 2nd Ed.; Maisonneuve SA: France, 1980.
- Ovington, L.G. The Well-Dressed Wound: An Overview of Dressing Types. *Wounds* **1998**, 10 (Suppl. A), 1A–11A.
- Dahle, J.S. *Proc. Symp. Wound Healing Helsinki*; Sundell, B. Ed.; 1979; 143–154.
- Banks, V.; Bale S.; Harding, K.G.; Harding, E.F. A Randomised, Stratified, Controlled, Parallel-group Clinical Trial of a New Polypurethane Foam Dressing (Lyof foam Extra) Versus a Hydrocellular Dressing (Allevyn) in the Treatment of Moderate to Heavily Exuding Wounds Proceedings of 6th European Conference on Advances in Wound Management Leaper, D. J. Cherry, G. W. Dealey, C. Lawrence, J.C. Turner, T. D. Macmillan London: 1997; 235–238
- Thomas, S. The Role of Foam Dressings in Wound Management. *Advances in Wound Management*; Turner, T.D., Schmidt, R.J., Harding, K.G., Eds.; John Wiley and Sons, Inc.: New York, 1986; 23–29.
- Harding, K.G. Silastic Foam. *Advances in Wound Management*; Turner, T.D., Schmidt, R.J., Harding, K.G., Eds.; John Wiley and Sons, Inc.: New York, 1986; 41–52.
- Banks, V.; Fear, M.; Orpine, J.; Hogelstrom, S.; Colgate, G.; Humphries J.; Disley, L.; Bale, S.; Thomas, S.; Harding, K.G. A Comparative Open Multicentre Trial of Tielle Hydropolymer Dressing and Granuflex Improved Formulation Proceedings of 5th European Conference on Advances in Wound Management Cherry, G. W. Gottrup, F. Lawrence, J. C. Moffat, C. J. Turner, T. D. Macmillan London 1995; 163–167
- Turner, T.D. *Pharm. Int.* **1985**, 6 (6), 131.
- Nathan, P.; Law, E.J.; MacMillan, B.G.; Murphy, D.F.; Rouel, S.H.; D'Andrea, M.I.; Abrahams, R.A. *Trans. Am. Soc. Artif. Int. Org.* **1976**, 22, 30–41.
- Eaglestein, W.H. *J. Amer. Acad. Dermatol.* **1985**, 12, 434–440.
- Yates, D.W.; Hadfield, J. *Med. Injury* **1984**, 16, 230–232.
- Myers, J.A. *Pharm. J.* **1983**, 230, 263–264.
- Schmidt, R. Xerogel Dressings. Turner, T.D., Schmidt, R.J., Harding, K.G. Eds.; *Advances in Wound Management*; John Wiley and Sons, Inc.: New York, 1986, 65–72.
- Gilchrist, T.; Mitchell, D.C.; Burrows, T.R. Sorbsan. *Advances in Wound Management*; Turner, T.D., Schmidt, R.J., Harding, K.G., Eds.; John Wiley and Sons, Inc.: New York, 1986; 73–82.
- Turner, T.D. *Royal Soc. Med. Symp.* **1985**, (Series No. 38), 5–14.
- Lofts, P.M. Biodressings. Turner, T.D., Schmidt, R.J., Harding, K.G., Eds.; *Advances in Wound Management*; John Wiley and Sons, Inc.: New York, 1986; 127–131.
- Mulder, G.T. The Role of Tissue Engineering in Wound Care. *J. Wound Care* **1999**, 8 (1), 21–24.
- Turner, T.D. *Wounds* **1989**, 1 (3), 155–179.
- Turner, T.D. *Surgical Dressings and Their Evolution Proceedings of 1st European Conference on Advances in Leaper, D.J. Harding, K.G. Turner, T.D. Wound Management Macmillan: London, 1992; 181–187*

BIBLIOGRAPHY

- An Environment for Healing. *Royal Soc. Med. Int. Cong. Series* **1984**, (88).
- Bucknell, T.E., Ellis, H., Eds. *Wound Healing for Surgeons*, Balliere Tindall: London, 1984.
- Goldsmith, L.A.; Ed. *Biochemistry and Physiology of the Skin* Oxford University Press: Oxford, 1983; 1, 11.
- Leaper, D.J., Harding, K.G.; Eds. *Wounds: Biology and Management*; Oxford University Press: Oxford, 1998.
- Najano, G. *The Healing Hand*; Harvard University Press: Cambridge MA, 1975.
- Peacock, E.E. *Wound Repair*; WB Saunders: London, 1984.
- Turner, T.D., Schmidt, R.J., Harding, K.G., Eds. *Advances in Wound Management* John Wiley and Sons, Inc.: New York, 1986.
- Westaby, E.S., Ed. *Wound Care*; William Heinemann Medical Books: London, 1985.

DOSAGE FORMS: PARENTERALS

Gayle A. Brazeau

Adam Persky

Jintana M. Napaporn

University of Florida, Gainesville, Florida, U.S.A.

INTRODUCTION

Parenteral is derived from the two words “para” and “enteron” meaning to avoid the intestine. Parenteral articles are defined according to the USP 24/NF19 “as those preparations intended for injection through the skin or other external boundary tissue, rather than through the alimentary canal, so that the active substances they contain are administered using gravity or force directly into a blood vessel, organ, tissue, or lesion. Parenteral products are prepared scrupulously by methods designed to ensure that they meet pharmacopeial requirements for sterility, pyrogens, particulate matter, and other contaminants, and, where appropriate, contain inhibitors of growth of microorganisms. An injection is a preparation intended for parenteral administration and/or for constituting or diluting a parenteral article prior to administration” (1). Parenteral drug administration is an attractive route of administration when oral administration is contraindicated, and it has been traditionally used in institutional settings. With an increasing interest in reducing overall health care costs and with the development of new biotechnologically derived compounds and improved and novel infusion-related technologies, parenteral products have become an important component in the care of patients in hospitals and the home health care setting. In the present article, information will be presented on history and the following: the use of parenterals in health care, the advantages and disadvantages of using parenterals, routes of administration, vascular access devices and infusion sets, types of parenteral products, components of parenteral products, parenteral packaging, convenience and needleless systems, needleless injection, extemporaneous compounding of parenteral products, infusion pumps and devices, and future parenteral dosage forms.

HISTORY AND USE OF PARENTERALS

A detailed history of early parenteral medications can be found in the first edition of this encyclopedia (2). One of

the first historical references to the parenteral administration of a compound was in the late 15th century when a blood transfusion from three young boys was given to Pope Innocent VIII, resulting in the death of all four individuals. These deaths led to a ban in the use of this type of medical treatment, namely an infusion, for several centuries. It was not until the 17th century that studies on the parenteral administration of compounds was first studied in animals. The development of the hypodermic needle and the use of parenterally injected drugs in humans is first reported in the mid-19th century. By the end of this century, there was an increased interest in the use of intravenous administration of glucose and normal saline solutions. Baxter produced the first commercially prepared intravenous solutions in 1931. However, parenteral products and their administration became acceptable and a mainstay in the treatment of patients in the mid 20th century. This could be attributed, in part, to our increased understanding of microbial and viral agents, increased recognition of the dangers associated with parenteral therapy, the development of antibiotics and other drug classes, and the availability of new systems or infusion technologies. In the mid 1960s, many hospitals introduced intravenous admixture services. In the last 20 years, the area of infusion pumps and systems and improved vascular access devices has enabled parenteral therapy to extend beyond the institutional setting to clinics, ambulatory infusion centers, and home health care. In addition, the administration of parenteral drugs is also frequently utilized in basic and clinical research in animals and humans.

In today's health care environment, parenteral products are a key component of therapy for hospitalized patients. Vascular access for parenteral infusion therapy is obtained in the vast majority of hospitalized patients at some point in their therapy. There are very little data as to the use of parenteral products in today's health care environment. It was suggested that 40% of all pharmaceutical dosage forms are administered as a type of injection and that over 350 million units of large volume parenterals and 100 million units of IV admixtures (piggybacks) were used annually in the late 1980s (3). These numbers have

certainly increased since that time with the advent of new drugs and infusion methodologies. Infusion therapy in home health care continues to have a strong market in the United States. While the annual growth rate of home infusion therapy decreased between 1982 and 1993, 29% of acute care hospitals provided or were developing a home health care program. In the current market, home infusion therapy is being integrated into alternative sites, such as ambulatory infusion centers (4). Due to the rapid and increasing use of parenteral drugs, it is critical for health care providers and scientists to understand the various available dosage forms, specific products, routes of administration, catheter types, and various infusion devices. Detailed information on these devices can be found in the nursing and pharmaceutical literature (3, 5–10). In addition, company websites are a valuable source of current information, including available parenteral products, infusion sets and ports, devices and infusion pumps. A list of useful worldwide websites is shown in Table 1.

ADVANTAGES AND DISADVANTAGES OF PARENTERAL PRODUCTS

The advantages and disadvantages of parenteral drugs and administration are shown in Table 2. Generally, parenterally administered drugs are advantageous because they can provide rapid and reliable drug systemic effects, long-term drug delivery, and targeted drug delivery. The disadvantages of parenteral products are predominately associated with safety issues related to infection and thrombosis, tissue damage and/or pain upon injection, and the use of requirements for specific equipment, devices, and techniques.

ROUTES OF ADMINISTRATION

The routes of administration for parenteral products are shown in Fig. 1. The most commonly used routes are intravenous, intramuscular, subcutaneous, and intradermal. Formulations intended for administration into the central nervous system should not include preservatives. For intramuscular, intradermal, and subcutaneous, a single needle and syringe is generally used to administer the drugs. For intravenous and intra-arterial, drugs are administered using vascular access or port devices. Other parenteral routes of administration (e.g., epidural, intra-articular, and intrathecal) usually require specialized delivery sets. In some cases, the specific drug requires a

specialized delivery set be utilized due to the dose of the drug or the physicochemical properties of the drug (e.g., nitroglycerin).

VASCULAR ACCESS DEVICES AND INFUSION SETS

Vascular access can be achieved through short peripheral, long dwell peripheral, central lines, and ports (8, 11). These various devices differ in their insertion, characteristics, dwell time or time they should be in place, and usage and safety features. In peripheral access, the distal veins on the hand and arm are often used, with the basilic and cephalic veins in the arms being the most common site for peripheral infusions. Alternatively, the metacarpal veins can also be utilized. The basic devices are a winged infusion device or the over-the-catheter needle, with needle sizes ranging from 16 to 26 G (20- or 22-gauge being the most common size). These catheters are usually utilized only for 48 h. A midline catheter is usually inserted into a large vein and is intended to be used over a 2- to 4- week period (Richardson). Peripherally inserted central venous catheters (PICC) are designed for long-term infusion therapy up to a year. The devices are inserted into a peripheral vein and threaded so that the tip of the catheter is within the central vascular system. Central venous catheters (CVC) are inserted into the subclavian or jugular vein and threaded until the tip is located in the superior vena cava. The types of central venous catheter are a nontunneled, Groshong, Hickman, and Brovaic. A nontunneled CVC can be inserted at the bedside, whereas the Groshong, Hickman, and Brovaic require surgical insertion. Vascular access ports (VAP) are an alternative to central venous access. These devices are surgically implanted usually into the chest wall or arm subcutaneous tissue and are composed of a rigid reservoir with a self-sealing rubber septum, and the tip of the catheter is placed into a central vein. The drug is placed into the reservoir via an injection. A VAP allows repeated, intermittent access and drug delivery (in some cases up to 2000 times), depending upon the size of the needle. Examples of these types of ports include a P.A.S Port® (SIMS Deltec), Vital-Port® (Cook Incorporated), and BardPort® and CathLink® (Bard Medical Division).

There are three basic types of intravenous administration: 1) primary set, 2) secondary set, and 3) a volume control set (Fig. 2) (8). The basic components of all these sets include a piercing spike (to insert into the bag or bottle), drip chamber and drip orifice, tubing ranging in length from 160 to 250 cm (63–98.58 in.), a roller clamp

Table 1 Representative parenteral products and devices websites

Company	Website	Products available
Abbott Laboratories	http://www.abbott.com	Premixes, IV solutions—Lifecare, injectable drug delivery systems, needleless systems, infusion pumps, Add-Vantage
Alaris Medical Systems	http://www.alarismed.com	Infusion pump systems, ambulatory pumps, needlefree systems, intravenous administration sets
Arrow International Critical Care Products	http://www.arrowintl.com	Vascular access, catheters, infusion ports
Bard Access Systems	http://www.bardaccess.com	Access systems and ports
B. Braun Medical	http://www.bbraunusa.com	IV Solutions—Excel or PAB (Partial Additive Bag), IV administration, needlefree administration, infusion pump systems, vascular access
Baxa	http://www.baxa.com	IV admixture tools, syringe infusion systems
Baxter Healthcare	http://www.baxter.com	Large and small volume parenterals, premix medications, reconstitution products and accessories, syringe pumps and sets, interlink, administration sets, infusion pumps
Becton, Dickinson and Company	http://www.bd.com	Needles, vascular access, needleless systems, catheters
Cook Endovascular	http://www.cook-inc.com	Vascular access, infusion catheters and sets
Horizon Medical Products	http://www.hmpvascular.com	Vascular access ports
I-Flow Corporation	http://www.i-flowcorp.com/home.html	Homepump infusion systems, electronic infusion pumps
SIMS Deltec	http://www.deltec.com	Access systems and ports, infusion pumps
VYGON USA	http://www.vygonusa.com	Vascular access, needleless access

Table 2 Advantages and disadvantages of parenteral drugs and administration

Advantages	Disadvantages
Useful for patients who cannot take drugs orally	More expensive and costly to produce
Useful for drugs that require a rapid onset of action (primarily intravenous administration)	Potential for infection at site of injection
Useful for emergency situations	Potential for sepsis
Useful for providing sustained drug delivery (implants, intramuscular depot injections)	Potential for thrombophlebitis
Can be used for self-delivery of drugs (subcutaneous)	Potential for fluid overload
Useful for drugs that are inactivated in the gastrointestinal tract or susceptible to first-pass metabolism by the liver	Potential for air embolism
Useful for injection of drugs directly into a tissue (targeted drug delivery)	Potential for extravasation
Useful for delivering fluids, electrolytes, or nutrients (total parenteral nutrition to patients)	Psychological distress by the patient
Useful for providing precise drug delivery by intravenous injection or infusion utilizing pharmacokinetic techniques	Require specialized equipment, devices, and techniques to prepare and administer drugs
Can be done in hospitals, ambulatory infusion centers, and in home health care	Potential for pain upon injection
	Potential for tissue damage upon injection
	Risk of needlestick injuries and exposure to blood-borne pathogens by health care worker
	Increased morbidity associated with long-term vascular access devices
	Disposal of needles, syringes, and other infusion devices requires special consideration

or other flow control device on the tubing, a Y-site for infusion of other components, an in-line filter (ranging from 0.2 to 170 μm) and leuk-lok adapter to attach to the vascular access device. A primary infusion set is designed to deliver solutions from the parenteral container via gravity. If an additional infusion is needed, the secondary set can be connected to the primary set via one of the Y-sites. A volume control administration set is used to deliver a small amount of solution through the use of the volume chamber. In general, administration sets are classified into macrodrip sets that can deliver 10–20 drops/ml and a microdrip set that delivers drugs at a slower rate of 60 drops/ml.

TYPES OF PARENTERAL PRODUCTS

Parenteral products can be divided into two general classes according to the volume of the product. All parenteral products are sterilized and must meet all the requirements for sterility and particulate matter and must be pyrogen-free (36). They must be prepared using strict sanitation standards in environmentally controlled areas by individuals trained to meet these standards. The injections are overfilled with a small excess over the labeled volume to ensure that the required volume can be obtained from the product. *Small-volume parenterals (SVP)* or injections are 100 ml or less and can be provided as a single- or multidose

product. In contrast, *large-volume parenterals (LVP)* are intended to be used intravenously as a single-dose injection and contain more than 100 ml of solution. SVPs and LVPs are often combined during the extemporaneous preparation of intravenous admixtures, to be discussed later in this article.

The U.S. Pharmacopoeia (USP) classifies injections into five different types. The dosage form selected for a particular drug product is dependent upon the characteristics of the drug molecule (e.g., stability in solution, solubility, and injectability), the desired therapeutic effect of the product (e.g., immediate vs. sustained release), and the desired route of administration. Solutions and some emulsions (e.g., miscible with blood) can be injected via most parenteral routes of administration. Suspensions and solutions that are not miscible with blood (e.g., injections employing oleaginous vehicles) can be administered via intramuscular or subcutaneous injection but should not be given intravenously.

Parenteral products contain excipients such as buffers, solvents, nonaqueous solvents, antimicrobial preservatives, antioxidants, and chelating agents. Coloring agents are prohibited in parenteral products. All excipients must meet compendial standards, and the excipients must not interfere with the efficacy of the product (to be discussed more in detail later in this article). Parenterals are packaged in airtight containers using specific, high quality materials so that they do not interact with the product and to maintain the sterility of the product. For example, the

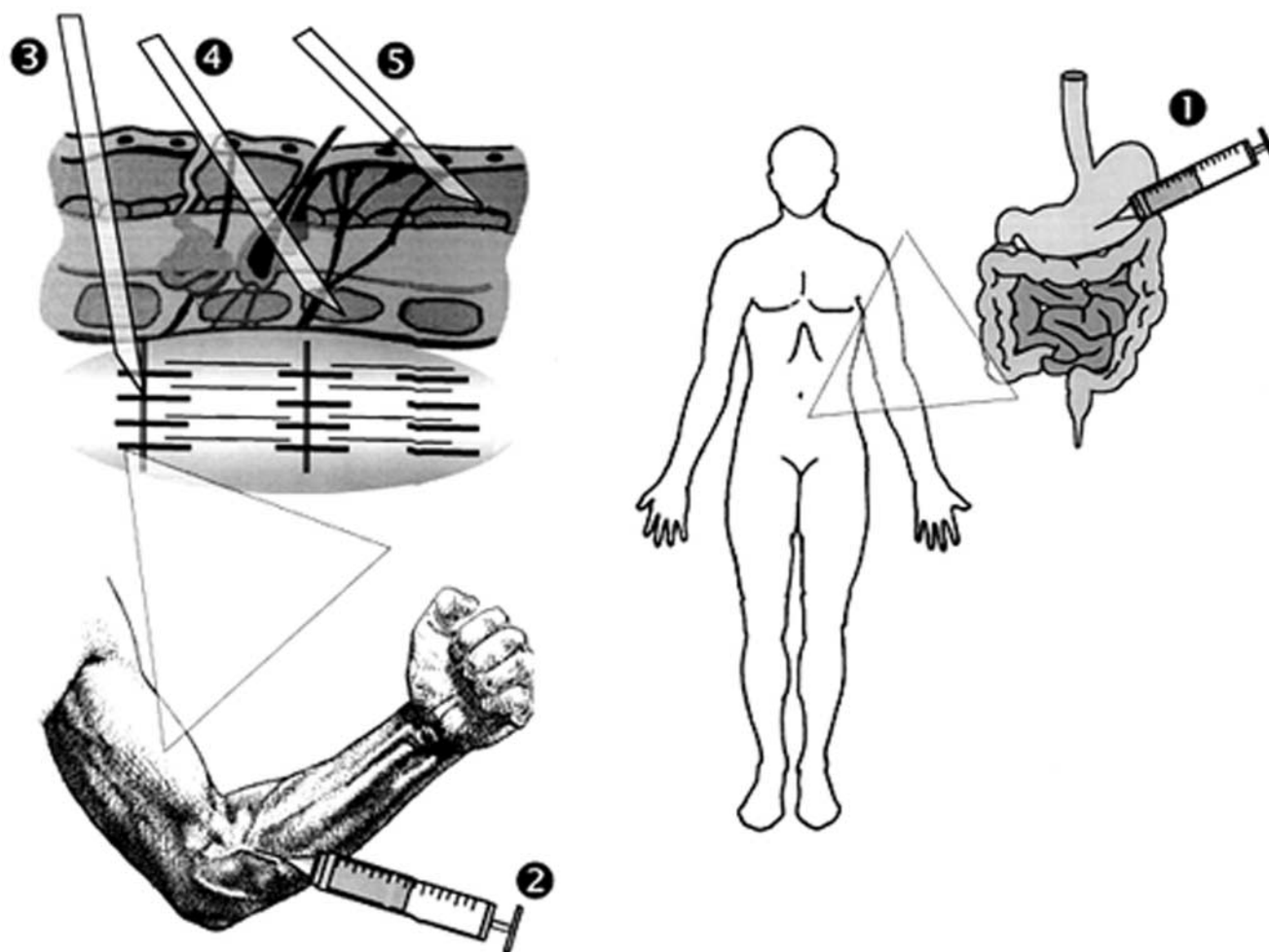


Fig. 1 Routes of parenteral administration. The following represent the most commonly used routes of administration for parenteral products: 1) intraperitoneal, 2) intravenous, 3) intramuscular, 4) subcutaneous, and 5) intradermal.

type of glass to be used in a specific parenteral drug product is indicated in the monograph. The types of packaging and containers for SVPs and LVPs will be discussed later in this article.

A SVP product is available for most of the major therapeutic classes of drug. It is often desirable for a manufacturer to provide both an oral and parenteral dosage form for a specific drug product. A “drug injection” is a liquid preparation that is composed of drug substances and or solutions. A “drug for injection” is a dry solid that upon the addition of a suitable vehicle (usually a vehicle in which the drug is stable and soluble) provides a solution that conforms to the requirements for an injection. Drugs for injection are often lyophilized or freeze-dried to assist in the reconstitution of the solid. A “drug injectable emulsion” is a liquid preparation of a drug or drug substances dissolved in a suitable emulsion vehicle. A

“drug injectable suspension” is a liquid preparation of solids suspended in a suitable vehicle. A “drug for injectable suspension” is a dry solid (often lyophilized) that is intended, upon the addition of a suitable vehicle, to yield a preparation that in all aspects meets the requirements for an injectable suspension.

LVPs are often administered via intravenous infusion in a large single-dose container. The therapeutic goal of these products is to provide electrolytes, body fluids, and nutrition. These solutions may or may not be isotonic with blood depending upon the concentration of the components, which include sodium chloride, dextrose, mannitol, Ringers (sodium, potassium, calcium, and chloride) and Lactated Ringers (calcium, potassium, sodium, and lactate), sodium bicarbonate, ammonium chloride, sodium lactate, fructose, alcohol, dextran, and amino acids. Other drugs (small volume injectables) are

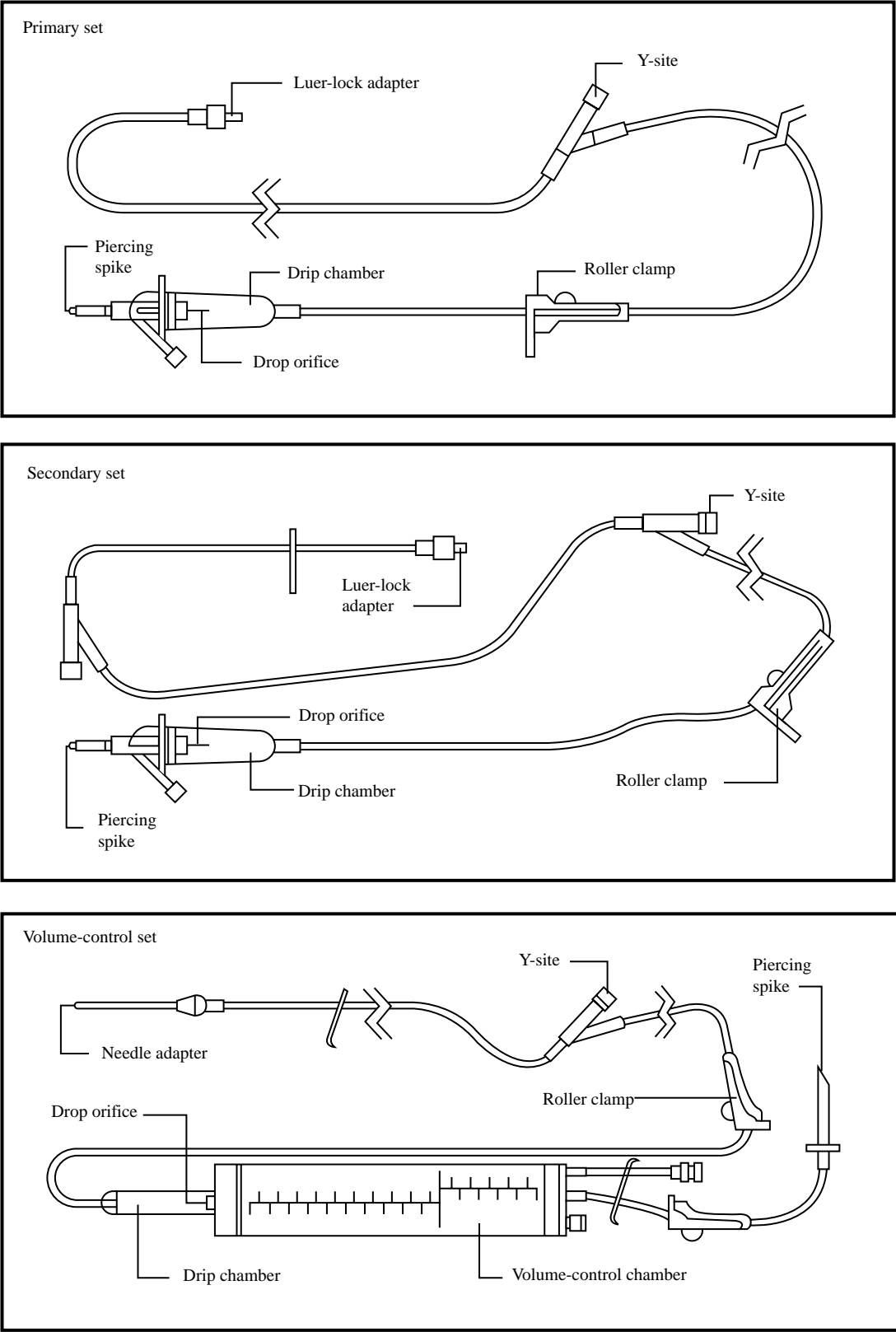


Fig. 2 Types of infusion sets used for parenteral therapy: primary, secondary, and volume-controlled infusion sets.

often combined to these LVPs, provided that these two products are compatible during the extemporaneous preparation of intravenous admixtures (discussed later).

COMPONENTS OF PARENTERAL PRODUCTS

Parenteral products are optimized during their development to provide the requisite solubility (per the required dose), stability, and syringeability. In addition, these products must meet the desired requirements for the rate of drug release based upon the dosage form and biopharmaceutical properties. Finally, it is important that parenteral products also be evaluated for their potential to cause tissue damage and/or pain associated with the injection of the formulation. The adjuvants in parenteral products can include solvents, vehicles, cosolvents, buffers, preservatives, antioxidants, inert gases, surfactants, complexation agents, and chelating agents.

It is important to understand the various types of waters used in parenteral products. The most frequently used solvent in parenteral products is Water for Injection, USP, which is not required to be sterile but must be pyrogen-free. In contrast, Sterile Water for Injection, USP is water that has been sterilized, does not include a preservative or antimicrobial agent, is pyrogen-free, and is provided in single containers no larger than 1000 ml. The use of this product is for the reconstitution of other parenteral products, in most cases antibiotics. This product must not be given alone. Bacteriostatic Water for Injection, USP is sterile water that can contain one or more preservative or antimicrobial agent (specified on the label) and is packaged in prefilled syringes or vials that are no larger than 30 ml. It is also used in the reconstitution of SVPs. The limitation with Bacteriostatic Water for Injection, USP is the presence of the antimicrobial agent that is contraindicated in newborns. Other solvents used for parenteral formulations are Sodium Chloride Injection, USP and Bacteriostatic Sodium Chloride Injection, USP, Ringers Injection, USP, and Lactated Ringer's, USP.

Other vehicles may be added to parenteral products if the aqueous solubility is limited. However, these vehicles must be nontoxic, nonsensitizing, and nonirritating (3, 6). In addition, these solvents must be compatible with the drug and other components in the formulation. Cosolvents often used in parenteral formulations include propylene glycol, ethanol, polyethylene glycols, glycerin, and dimethylacetamide. In addition, fixed vegetable oils, such as peanut, cottonseed, sesame and castor oil, can be used; however the USP provides clear restrictions on their use in parenteral products.

Buffers can also be provided in parenteral formulations to ensure the required pH needed for solubility and/or stability considerations. Other excipients included in parenteral products are preservatives (e.g., benzyl alcohol, *p*-hydroxybenzoate esters, and phenol), antioxidants (e.g., ascorbic acid, sodium bisulfite, sodium metabisulfite, cysteine, and butyl hydroxy anisole), surfactants (e.g., polyoxyethylene sorbitan monooleate), and emulsifying agents (e.g., polysorbates). An inert gas (such as nitrogen) can also be used to enhance drug stability. Stability and solubility can also be enhanced by the addition of complexation and chelating agents such as the ethylenediaminetetraacetic acid salts. For a more detailed list of approved excipients in parenteral products, the reader should consult the monographs within the USP.

PARENTERAL PACKAGING

In general, all parenteral products must be manufactured under strict, current good manufacturing processes (cGMP) to ensure the final product is sterile and pyrogen-free. Sterilization is defined as the complete destruction of all living organisms or their spores or the complete removal from the product (6). Pharmaceutical products can be sterilized by steam sterilization, dry-heat sterilization, filtration sterilization, gas sterilization, and ionizing-radiation sterilization. The USP provides monographs and standards for biological indicators required to test the validity of the sterilization process. These products must also be tested for pyrogens—fever-producing substances that arise from microbial contamination most likely thought to be endotoxins or lipopolysaccharide in the bacterial outer cell membrane.

Injections are provided in either multiple-dose containers or single-dose containers. A multiple-dose container is often a vial that will allow the withdrawal of successive portions of the contents without a change in the strength of the product and while maintaining the sterility. A single-dose product is intended for a single parenteral administration. These products can be an ampul, vial, or a syringe. For some drugs, there are specific double-chambered vials that contain the reconstitution solvent and the powdered drug (e.g., Mix-O-Vial—to be discussed later). Types I, II, and III glass are required for parenteral products and are specified in the individual monograph for a given drug.

Ampuls are utilized for a single dose and, as such, do not require a preservative. However, in many cases, the manufacturer will include a preservative, as the drug formulation is the same for both the ampul and

multiple-dose vial. The disadvantages of ampuls are that these containers become contaminated with glass particles when opened and require the use of a syringe to remove the drug solution. A filter needle must be used sometime during the withdrawal of the solution or delivery of the drug solution to a flexible bag or other intravenous solution to ensure the glass is removed from the solution. Ampuls are opened via breaking the neck at a prescored position.

Vials can be used for single or multiple doses. The glass containers are sealed with rubber closures that permit the withdrawal of the drug solution via a syringe. The disadvantage of these systems is associated with ensuring that the drug solution is compatible with the rubber closure. Furthermore, when utilizing vials in the extemporaneous preparation of sterile intravenous admixtures, the health care practitioner must minimize the potential of coring during the introduction of the needle through the rubber seal. Furthermore, there is always the concern of contamination of the solution with repeated withdrawals. The potential for contamination can be minimized by the use of single-dose vials.

Parenteral solutions can also be packaged in syringe dosage forms for a single-dose use. As such, they can be considered a type of convenience container (to be discussed later). The syringe and needle are sterile until opened. They are ideal for emergency situations or the home health care environment.

LVPs are usually provided in glass containers, flexible plastic bags, or semirigid containers. These systems are also classified as open systems (nonvacuum) and closed systems (vacuum). The largest manufacturers of LVPs are Abbott Laboratories, Baxter Healthcare Corporation, and B. Braun.

Glass containers are sealed with a thick rubber disk and a target in the center for the piercing spike. Glass bottles can be either vented with a plastic venting tube or nonvented, thereby requiring either a nonvented administration set or a vented administration set, respectively. The advantages of glass containers for parenterals are that they are easy to sterilize, can be accurately read, and are generally inert and less susceptible to incompatibilities with drugs or leaching of components compared to the plastic flexible intravenous bags. The disadvantage of glass is associated with handling the glass bottles and the potential for breakage (10).

Plastic intravenous fluid containers were first introduced due to the need to start intravenous therapy while transporting soldiers from the battlefield or triage area to the hospitals. These containers are flexible due to the presence of plasticizers, with the bags being composed of polyvinyl chloride. In contrast, semirigid containers are often composed of polyolefin. Some representative shapes

of these various types of containers are shown in Fig. 3. The most common manufacturers of intravenous solutions are Abbott Laboratories—Lifecare, Baxter Healthcare—Viaflex, and B. Braun Medical—Excel. The major advantages of plastic flexible bag systems for parenterals are that they do not require the use of a vented administration set as they collapse when empty, and they are less susceptible to breakage. It is also easier to store and transport these bags. The difficulties with these flexible plastic bags for infusions solutions are the potential for incompatibilities of the drug substance with the components in the bag (see later), the potential for the bag to get perforated during its use, thus compromising the sterility of the solution, and the difficulty in reading the volume remaining in the bag. One major concern with the use of flexible plastic bags is the potential for the drug compound to leach out the plasticizers from the systems. Semirigid containers are similar to flexible plastic containers in that they are lightweight and nonbreakable and can be easily transported and stored. These containers are less likely to be perforated during their use. More importantly, these containers do not contain plasticizers and, as such, may be more compatible with drug substances. The disadvantages are related to their similar properties to glass containers, in that they require venting, can be more susceptible to cracking upon extreme changes in temperature, should not be frozen, and do not adapt well for ambulatory care.

CONVENIENCE AND NEEDLELESS SYSTEMS

Whereas the compounding and administration of parenteral products and intravenous admixtures continues to be a vital and important component in the care of hospitalized and home health care patients, there is continued interest in easing the preparation, storage, and administration of these products with respect to controlling contamination of the finished product and protecting the health care providers from needlestick injuries. It is estimated that more than 750,000 needlestick injuries occur every year. Commonly used convenience and/or needleless systems include premixes, bags, and vial systems (e.g., Add-Vantage and Mini-Bag Plus[®]), prefilled syringe systems (Carpject[®]—Abbott and Tubex[®]—Wyeth-Ayerst) and double-chambered vial systems (Mix-O-Vial—Pharmacia-Upjohn and Redi-Vial[®]—Lilly).

For drugs with suitable stability in intravenous solutions, premixes provide an alternative to the

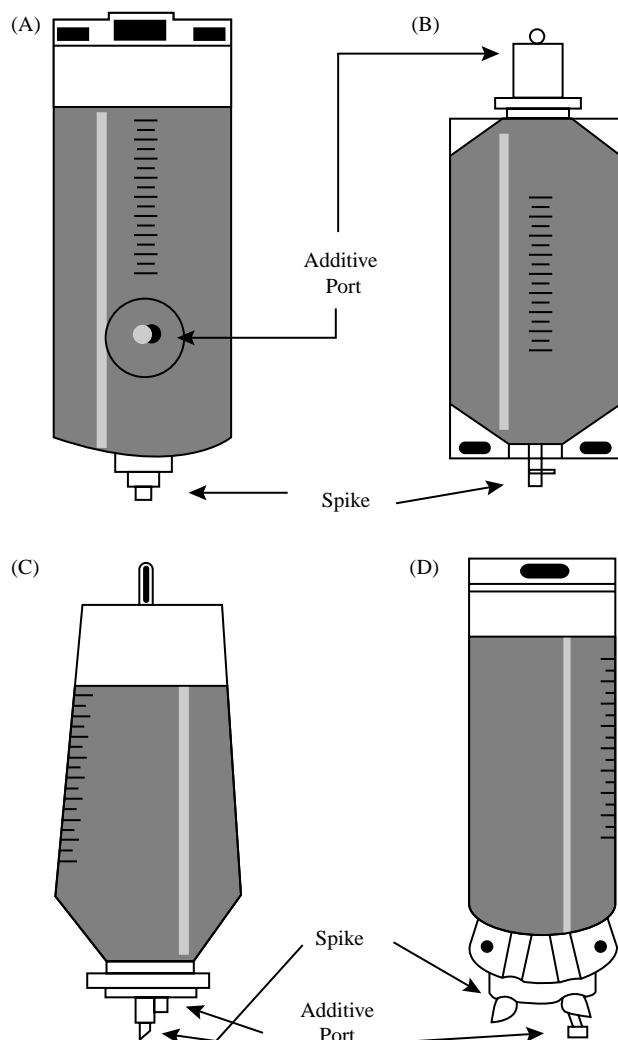


Fig. 3 Types of large volume parenterals. These represent the various types of flexible intravenous bag systems: (A) Lifecare[®] by Abbott Laboratories, (B) Representative vial and flexible IV bag system, such as the Add-Vantage[®]—Abbott Laboratories or Mini-Bag Plus System[®]—Baxter Healthcare, (C) Excel[®]—B. Braun Medical, and (D) Viaflex[®]—Baxter Healthcare. These bags contain either polyvinyl chloride (Lifecare or Viaflex) or polyolefin (Excel).

extemporaneous compounding of admixtures. These products are ready to administer, reduce the chance for a medication error, reduce the potential for infection, and decrease the chance for needlestick injuries. In addition, there is an advantage in using these products with respect to the shelf-life of the product. The stability and storage requirements for each product is provided by the manufacturer. For example, the FirstChoice[®] Premix products (Abbott) in the overwrap have a shelf-life of, typically, 18 months, whereas those products in which the

overwrap has been removed and which are stored at room temperature can be stable for up to 30 days, provided no additional drugs or additives have been added to the product. The diluents in these premix products include 0.45 or 0.9% sodium chloride, 5% dextrose, water for injection, Lactated Ringers, and combinations of these diluents in volumes ranging from 50 to 1000 ml in plastic or glass containers. Drug classes that are currently formulated as premixes include amino acids, dextrans, electrolytes, cardiovascular, anti-infectives, analgesics, and gastrointestinal and respiratory compounds.

The Add-Vantage[®] system (Abbott Laboratories) and the Mini-Bag[®] Plus system (Baxter Healthcare) are needleless drug delivery systems composed of a drug-containing vial and a diluent in a flexible plastic intravenous bag. A schematic of these type of systems is shown in Fig. 3B. The drug in the vial comes in contact with the diluent, followed by the drug being transferred back into the bag with the vial still attached; this system can then be attached directly to the infusion equipment. In the Add-Vantage system[®], there are specialized vials and intravenous bags and it is necessary for the health care professional to activate the system by removing the vial stopper, thus allowing the diluent to enter the vial. The Mini-Bag Plus[®] system is designed to allow the simple reconstitution of standard 20 mm powdered drug vials. The Monovial Safety Guard[®] (Becton, Dickinson and Company) is an integrated, self-contained system that allows the transfer of a reconstitution solution from a flexible intravenous bag or vial into a drug-containing vial, and it is only available for a limited number of drugs.

As such, the advantages of these systems are that the product can be easily stored and quickly prepared without the need for calculations, specialized equipment (such as laminar airflow hoods) or needles and syringes. They also allow for a quicker turnaround time for the first dose. These products enhance safety for both the patient, by reducing the chance for medication errors (e.g., the wrong drug added to the vial, the incorrect amount of drug to be added to the bag, or a product being incorrectly labeled), and for the health care practitioner, by minimizing the chance for a needlestick injury. In addition, these products can help to reduce costs associated with unused doses because the unwrapped products can be redistributed for short periods of time. At present, a variety of therapeutic classes of drugs from anti-infectives to cardiovascular agents to pain management at various doses are available for reconstitution in bags containing 0.9 or 0.45% sodium chloride or 5% dextrose in 50- to 250- ml bags.

Prefilled syringes (Carpject[®]—Abbott and Tubex[®]—Wyeth-Ayerst) are composed of drug solutions placed in a

syringe with a needle and needleless systems. The advantages of these systems are the convenience associated with a standard dose, less chance for medication error associated with extemporaneous compounding of these syringes, usefulness in emergency situations, and ease of storage. The needle and syringe are sterile until opened. Prefilled syringes are available for drugs ranging from anti-infectives, analgesics, and antipsychotics to antiemetics.

Doubled-chambered vials are advantageous in that the reconstitution solution is separated from the drug until desired by the health care practitioner. The Mix-O-Vial[®] (Pharmacia-Upjohn) system is a combination of a powdered or lyophilized drug in a lower container and an appropriate diluent with a preservative and other active ingredients in an upper container. Following removal of the dust cover and upon pressure on the top plunger, the solution comes in contact with the drug and the vial is shaken until a solution is obtained. The upper plunger can then be swabbed with a disinfectant and the appropriate volume of drug removed with a needle as in the standard preparation of an intravenous admixture using a vial.

With an increased interest in eliminating needlestick injuries associated with parenteral drug administration in health care workers, needleless systems are becoming more common in patient care. For example, the Interlink[®] System (Baxter) is designed for needleless access during intravenous therapy. These types of products are available for injection sites, Y-sites, vial adapters, infusion and vein access, syringe products and catheter extension sets. Other needleless catheters and infusion sites include the Introcan[®] Safety IV catheter and the Sifesite[®] injection caps (B. Braun).

NEEDLELESS INJECTION

The concept of needleless injection is not a new one and has been thought about since the 1940s. Current products utilize either spring action or compressed gas (e.g., helium or carbon dioxide) as a propellant to deliver a drug through the skin. These needleless systems offer several advantages. The first potential advantage is reduced pain and anxiety, an advantage for use in children. The second advantage is that needleless injection causes less tissue damage than conventional needles (12). Finally, a needleless system results in a diffuse pattern of exposure and, therefore, increases surface area and absorption rate (12). The main disadvantage of this system is unreliability in reference to pain and discomfort and skin characteristics

that can influence the amount of drug entering the body (12). This system has been used to administer vaccines, insulin (12), and drugs for topical applications (e.g., penile erectile dysfunction) (13) and as a means to deliver DNA for gene therapy (14).

EXTEMPORANEOUS COMPOUNDING OF PARENTERAL PRODUCTS

Whereas the presence of the various convenience parenteral products has assisted health care practitioners in safely and accurately delivering drugs to the patients, the extemporaneous compounding of parenteral products continues to be an important component in institutional settings and home health care (3, 5, 15). Parenteral intravenous admixtures include the withdrawing of the drug solution from an ampul(s) or vial(s) and placing it into various large volume solutions, syringe dosage forms for patients, total parenteral nutrition solutions and cassettes, or other delivery systems for home health care patients. The American Society of Health-System Pharmacists, the National Association of Boards of Pharmacy, and the USP provide practitioners with useful technical assistance bulletins, rules, and standards for the preparation of parenteral products (13). The concerns associated with the extemporaneous preparation of parenteral products are maintaining sterility of the products through proper aseptic techniques, calculating and providing the correct dosage, preventing or reducing drug–drug, drug–solution, or drug–container incompatibilities during the preparation or administration of the product, and maintaining and providing drug stability and quality control.

The majority of extemporaneous parenteral products are prepared by pharmacists working in hospitals, home health care, or long-term care facilities. These products must be prepared using aseptic technique and using the appropriate supplies (e.g., syringes, needles, and filter needles, and caps) and equipment (Class 100 laminar airflow hoods enclosed within a class 10,000 clean room). Aseptic technique which differs from sterilization, is a process by which an individual can manipulate sterile products and containers to prevent microbial contamination. Pharmacists and other personnel involved in the extemporaneous preparation of parenteral products require special knowledge and training and should receive additional training and education on a routine basis to ensure proper aseptic techniques are being followed consistently. In addition, these individuals must be able to perform the required calculations (e.g., dosing,

milliequivalents, milliosmoles, and powder volume) needed to prepare intravenous admixtures. Equally important to the safe preparation of intravenous admixtures is an understanding of general principles and concepts related to drug and solution or container incompatibilities and to drug stability and also specific knowledge as to whether a specific drug is compatible or stable with another drug, solution, or container. For example, general and specific information on intravenous admixture incompatibilities can be found in books such as *Handbook on Injectable Drugs* (16, 17) and *King Guide to Parenteral Admixtures* (18) and in primary literature sources such as the *American Journal of Health-System Pharmacy* and the *International Journal of Pharmaceutical Compounding*.

INFUSION PUMPS AND DEVICES

Infusion pumps and devices are an essential component to the delivery of parenteral drugs, particularly those given by the intravenous route. For drugs that are administered via intravenous infusion, there are two forces that control fluid flow: 1) the pressure of an active force of the liquid that can be generated via gravity flow (*viz.*, hydrostatic pressure) or mechanically via a positive pressure pump and 2) resistance, or an opposing force, that is generated via the infusion sets, a vascular access device and/or blood vessels. The maximum flow rate will depend upon the ratio of the change in pressure exerted by the liquid to that of the change in resistance.

An infusion control device (ICD) is a device that maintains a constant infusion rate in a gravity flow system (controller) or via a positive pressure pump. A positive pressure pump is a device that provides mechanical pressure (2–12 psi) to overcome the resistance to flow in the vessels. The types of positive pressure pumps are categorized according to how they deliver the solution and their degree of precision in the flow rate. Positive pressure pumps include peristaltic pumps, cassette pumps, syringe pumps, nonelectric or disposable pumps, and patient-controlled analgesic pumps (PCA). Syringe pumps are usually the most accurate pumps, with flow variances at 2% or less. Nonelectric or disposable syringe and PCA pumps are useful for ambulatory care. PCA pumps are very useful for the parenteral administration of analgesics (*viz.*, morphine) and can be easily programmed to deliver bolus doses and provide a dosing history. Nonelectric or disposable pumps (*e.g.*, Homepump®—I-Flow Corporation, Readymed®—Alaris Medical Systems, and Smart-Dose®—ProMed) are lightweight, and the solution is

delivered based upon a vacuum or through the generation of a gas in the system. A recent consensus development conference on the safety, cost, simplicity of use, and training of intravenous drug delivery systems, focusing on acute care and nonelectronic devices, reviewed the use of manufacturer-prepared (*e.g.*, premixed or frozen products), point-of-care activated systems (manufacturer-prepared products that require the drug and diluent to be mixed at the point of care), pharmacy-based intravenous admixture, intravenous push medications in prepared or premade syringes, augmented iv push systems (syringe pumps), and volume control chambers (19). Manufacturer-prepared products, point-of-care activated products, and pharmacy-based intravenous admixture programs were recommended as being superior intravenous drug delivery systems, with the manufactured products being considered the safest systems due to the quality assurance in the preparation of these products.

FUTURE PARENTERAL DOSAGE FORMS

Current research is leading to newer types of parenteral dosage forms that will be useful for both immediate and sustained drug delivery, for systemic and targeted drug delivery, and for the delivery of small molecules and macromolecules (*e.g.*, proteins, peptides, and DNA). There has been an increased interest in developing a wide variety of particulate drug delivery systems for parenteral products, which have included liposomes or other phospholipid vesicles, microspheres, microcapsules, nanoparticles, or microemulsions (20–23). The development of new biomaterials, such as the linear and branched biodegradable polyesters, has increased the interest in the development of these systems for microsphere formulations for parenteral drugs (24, 25). An advantage is that drug molecules can be incorporated into these particulate carriers, and, as such, the rate of drug release can be modified, cellular uptake can be facilitated, or the degree of tissue damage or pain can be reduced.

Microemulsions, defined as clear solutions obtained by triturating normal coarse oil-in-water emulsions with a medium chain alcohol and composed of the nonpolar phase, surfactant and cosurfactant, appear to be potentially useful for parenteral administration, as these are clear and stable formulations that are able to be filtered and might be suitable for intravenous administration (21, 22). In addition, other researchers are investigating *in situ* forming gel or implants that can be easily injected intramuscularly or subcutaneously and that result in the formation of a depot at the site of injection with the

potential to modify or extend the release of the drug or macromolecule (25–27).

CONCLUSIONS

Parenteral products will continue to play a vital role in the treatment of patients when the oral route is contraindicated, when it is necessary to carefully control drug blood levels in response to therapeutic effects, when a prolonged therapeutic effect is needed through a long-acting injectable, or when a drug effect is to be targeted to a specific tissue or organ, to name a few instances. Advances in the technology required for the administration of parenteral dosage forms in the last 100 years have expanded their clinical uses for in-patient and out-patient settings. In addition, there is improved convenience and safety for the health care providers who prepare and administer these products. It seems likely that more parenteral dosage forms will become available in the marketplace in response to the compounds being developed through biotechnology.

REFERENCES

- General Requirements for Tests and Assays. *The United States Pharmacopeia, The National Formulary*; US Pharmacopeial Convention, Inc.: Rockville, MD, 1995; 1775–1777.
- Swarbrick, J., Boylan, J.C., Eds. *Dosage Forms: Parenteral. Encyclopedia of Pharmaceutical Technology*; Marcel Dekker, Inc.: New York, 1988; 231–247.
- Turco, S. *Sterile Dosage Forms: Their Preparation and Clinical Application*; Lea & Febiger: Philadelphia, PA, 1994.
- Monk-Tutor, Ma.R. The U.S. Home Infusion Market. *Am. J. Health-Syst. Pharm.* **1998**, 55, 2019–2025.
- Thompson, J.E. *A Practical Guide to Contemporary Pharmacy Practice*; Williams & Wilkins: Baltimore, MD, 1998; 32.1–32.21.
- Ansel, H.C.; Allen, L.V.; Popovich, N.G. *Pharmaceutical Dosage Forms and Drug Delivery System*, 7th Ed.; Lippincott, Williams & Wilkins: Philadelphia, PA, 1999; 397–449.
- Avis, K.E.; Miller, W.A.; Phillips, G.B. *Dispensing of Medication*, 8th Ed.; Hoover, J.E., Ed.; Mack Publishing Co.: Easton, PA, 1976; 255–295.
- Doyle, R.M., Nale, P. Eds.; *Handbook of Infusion Therapy* Springhouse Corporation: Springhouse, PA, 1999.
- Andrews, M.; Kaplan, H.; Weinstock, D.L.; Wittig, P.A. *Nurses Handbook of Home Infusion Therapy*; Springhouse HomeCare: Springhouse, PA, 1997.
- Terry, J.; Baranowski, L.; Lonsway, R.A.; Hedrick, C.E. *Intravenous Therapy Clinical Principles and Practice*; W.B. Saunders Co.: Philadelphia, PA, 1995.
- Moureau, N. Vascular Access with a Focus on Safety. *Infusion* **1999**, 6, 16–26.
- Baker, A.B.; Sanders, J.E. Fluid Mechanics Analysis of a Spring-Loaded Jet Injector. *IEEE Trans. Biomed. Engineer* **1999**, 46, 235–242.
- Seyam, R.M.; Bégin, L.R.; Tu, L.M.; Diona, S.B.; Merlin, S.L.; Brock, G.B. Evaluation of a No-Needle Penile Injector: A Preliminary Study of Evaluation Tissue Penetration and its Hemodynamic Consequences in the Rat. *Urology* **1997**, 50, 994–998.
- Furth, P.A.; Shamay, A.; Hennighausen, L. Gene Transfer into Mammalian Cells by Jet Injection. **1995**, 14, 149–152.
- Allen, L. *The Art, Science and Technology of Pharmaceutical Compounding*; American Pharmaceutical Association: Washington, DC, 1998; 251–264.
- Trissel, L.A. *Handbook on Injectable Drugs*, 10th Ed.; American Society of Health System Pharmacists: Bethesda, MD, 1998.
- Trissel, L.A. *Handbook on Injectable Drugs*; American Society of Health System Pharmacists: Bethesda, MD, 1999; Supplement.
- Catania, P.N. *King Guide to Parenteral Admixtures*; King Guide Publications: Napa, CA.
- Bates, D.W.; Dugger, B.H.; Gosbee, J.W.; Schneider, P.J.; Cousins, D.D.; Flynn, E.; Richason, L. *Safety of Intravenous Drug Delivery Systems*, Consensus Development Conference on the Safety of Intravenous Drug Delivery Systems, Phoenix, AZ, Sept. 17–18, 1999.
- Morris, R.M.; Cuff, G.W. Particulate Drug Delivery Using Microparticulates. *Sterile Pharmaceutical Manufacturing. Applications for the 1990's*; Groves, M.J., Olsen, W.P., Anisfeld, M.H., Eds.; Interpharm Press: Buffalo Grove, IL, 1991; 2, 1–42.
- Lawrence, M.J. Surfactant Microemulsions and Vesicles as Vehicles for Drug Delivery. *Eur. J. Drug Met. Pharmacokin* **1994**, 3, 257–269.
- Tenjarla, S. Microemulsions: An Overview and Pharmaceutical Applications. *Crit. Rev. in Therapeutic Drug Carrier Systems* **1999**, 16, 461–521.
- Gupta, P.K.; Brazeau, G.A. *Injectable Drug Development: Techniques to Reduce Pain and Irritation*; Interpharm Press: Denver, CO, 1999.
- Breitenbach, A.; Li, Y.X.; Kissel, T. Branched Biodegradable Polyesters for Parenteral Drug Delivery Systems. *J. Controlled. Release* **2000**, 64, 167–178.
- Jain, R.A.; Rhodes, C.T.; Raikar, A.M.; Malick, A.W.; Shah, N.H. Comparison of Various Injectable Protein-Loaded Biodegradable Poly(Lactide-co-glycolide) (PGLA) Devices: *In-Situ* Formed Implant Versus *In-Situ* Formed Microspheres. *Pharm. Devel. Technol.* **2000**, 5, 201–227.
- Joshi, R.; Robinson, D.H.; Himmelstein, K.J. *In Vitro* Properties of an *In Situ* Forming Gel for the Parenteral Delivery of Macromolecular Drugs. *Pharm. Devel. Technol.* **1999**, 4, 515–522.
- Joshi, R.; Arora, V.; Desjardins, J.P.; Robinson, D.; Himmelstein, K.J.; Iversen, P.L. *In Vivo* Properties of an *In Situ* Forming Gel for Parenteral Delivery of Macromolecules. *Pharm. Res.* **1998**, 15, 1189–1195.

DRUG DELIVERY—BUCCAL ROUTE

James C. McElnay
Carmel M. Hughes

The Queen's University of Belfast, Belfast, United Kingdom

INTRODUCTION

A drug can be administered via many different routes to produce a systemic pharmacologic effect. The most common method of drug administration is via the *peroral* route, in which the drug is swallowed and enters the systemic circulation primarily through the membranes of the small intestine. Although this type of drug administration is commonly termed *oral*, *peroral* is a better term because *oral* administration more accurately describes drug absorption from the mouth itself. The mouth is lined with a mucous membrane and among the least known of its functions is its capability of serving as a site for the absorption of drugs (1). In general, drugs penetrate the mucous membrane by simple diffusion and are carried in the blood, which richly supplies the salivary glands and their ducts, into the systemic circulation via the jugular vein. Active transport, pinocytosis, and passage through aqueous pores usually play only insignificant roles in moving drugs across the oral mucosa (2).

The administration of drugs by the buccal route has several main advantages over peroral administration, including the following:

1. The drug is not subjected to the destructive acidic environment of the stomach.
2. Therapeutic serum concentrations of the drug can be achieved more rapidly.
3. The drug enters the general circulation without first passing through the liver.

This last phenomenon is important for drugs that are highly metabolized during their first passage through the liver. This metabolism (governed by the hepatic extraction ratio) can lead to a dramatic reduction in the amount of drug available systemically from a given peroral dose but is avoided by buccal absorption.

Two sites within the buccal cavity have been used for drug administration. Using the sublingual route, as for glyceryl trinitrate (GTN), the medicament is placed under the tongue, usually in the form of a rapidly dissolving tablet. The second anatomic site for drug administration is between the cheek and gingiva. Although this second application site is itself known as *buccal absorption*, the

absorption from all areas within the buccal or oral cavity are considered in this article.

Of the range of pharmaceutical preparations available for administration into the oral cavity, the most popular form is that of a rapidly dissolving tablet that releases its drug contents for absorption across the oral mucosa. Alternatively, a tablet or capsule can be chewed to release its contents. This latter method is less successful because mastication tends to produce a large volume of saliva that increases the probability of premature swallowing. The same problem occurs in the administration of drug in the form of a chewing gum.

The aim of the present article is to review the published literature on the absorption of drugs through the oral mucosa. Special attention is given to the prevention of presystemic metabolism via drug administration by the buccal and sublingual routes. Consideration is also given to the types of pharmaceutical preparations that are commercially available for drug administration into the mouth. Before progressing to drug absorption, however, the structure and blood supply of the oral mucosa are discussed because of the important role they play in the transfer of drugs from the mouth into the systemic circulation.

STRUCTURE AND SECRETIONS OF THE ORAL MUCOSA

Epithelial Lining

The major function of the oral epithelium is to provide a protective surface layer between the oral environment and the deeper tissues. The oral epithelium has a squamous epithelium of tightly packed cells that form distinct layers by a process of maturation from the deeper layers to the surface (3). The pattern of maturation differs in different regions of the oral mucosa due to the variation in the specific function of the tissues. The surface layer of the hard palate and tongue forms keratin to yield a tough, nonflexible epithelial surface resistant to abrasion, but the epithelium of the cheek, floor of the mouth, and soft palate is nonkeratinized and facilitates distensibility. The major features of the keratinized and non-keratinized oral

epithelium have been extensively investigated by Squier and Rooney (4). Together with the presence or absence of keratin, the second main feature likely to influence regional differences in drug absorption is the epithelial thickness. This varies in different regions of the mouth: the hard palate, buccal mucosa, lip mucosa, and floor of the mouth have been found to have thicknesses of 100–120 μm , 500–600 μm , 500–600 μm , and 100–200 μm , respectively (5, 6).

Secretion of Saliva

In addition to the protective function afforded by the oral mucosa, it also has the ability to maintain a moist surface, which enhances permeability of the membrane to drugs (3). Although the mucous membrane lining in the mouth contains many minute glands called *buccal glands*, which pour their secretions into the mouth, the chief secretion is supplied by three pairs of glands, namely, the *parotid* (under and in front of the ear), the *submaxillary* (below the jaw), and the *sublingual* (under the tongue) glands. Blood is richly supplied to the salivary glands and their ducts by branches of the external carotid artery and afterwards, travelling through the many branch arteries and capillaries, returns to the systemic circulation via the jugular veins (1). The presence of saliva in the mouth is important to drug absorption for two main reasons:

1. Drug permeation across moist (mucous) membranes occurs much more readily than across nonmucous membranes.
2. Drugs are commonly administered to the mouth in the clinical setting in a solid form. The drug must, therefore, first dissolve in saliva before it can be absorbed across the oral mucosa; that is, the drug cannot be absorbed directly from a tablet.

VASCULAR SYSTEM OF THE ORAL MUCOSA

The vascular system and blood supply to the oral mucosa have been clearly described by Stablein and Meyer (7). Netter's excellent drawings of the blood supply to the mouth and pharynx, venous drainage of the mouth and pharynx, and lymphatic drainage of the mouth and pharynx have been published by Ciba (8). This latter publication also includes definitive documentation of the blood supply and drainage from the mouth.

The blood supply to the mouth is delivered principally via the external carotid artery. The maxillary artery is the major branch, and the two minor branches are the lingual

and facial arteries. The lingual artery and its branch, the sublingual artery, supply the tongue, the floor of the mouth, and the gingiva, and the facial artery supplies blood to the lips and soft palate. The maxillary artery supplies the main cheek, hard palate, and the maxillary and mandibular gingiva (7, 9). The internal jugular vein eventually receives almost all of the blood derived from the mouth and pharynx (8). Drugs diffusing across the membranes have easy access to the systemic circulation via the internal jugular vein.

FACTORS INFLUENCING DRUG ABSORPTION FROM THE ORAL CAVITY

Because the oral mucosa is a highly vascular tissue, the two main factors that influence drug absorption from the mouth are the permeability of the oral mucosa to the drug and the physicochemical characteristics of the drug that is presented at the site of absorption.

Permeability of the Oral Mucosa to Drugs

The lipid membranes of the oral mucosa are resistant to the passage of large macromolecules; however, small un-ionized molecules tend to cross the membrane with relative ease. This passage is in either direction, and indeed passage of drugs from the circulation into the mouth can be used in therapeutic drug monitoring by measuring drug concentrations in saliva. The permeability of the oral mucosa has been comprehensively reviewed by Siegel (10).

Mechanisms involved in drug absorption across the oral mucosa

The mechanisms by which drugs cross biologic lipid membranes are passive diffusion, facilitated diffusion, active transport, and pinocytosis. Small, water-soluble molecules may pass through small, water-filled pores. The main mechanism involved in drug transfer across the oral mucosa, common with all regions of the gastrointestinal tract, is passive diffusion, although facilitated diffusion has also been shown to take place, primarily with nutrients. *Passive diffusion* involves the movement of a solute from a region of high concentration in the mouth to a region of low concentration within the buccal tissues. Further diffusion then takes place into the venous capillary system, with the drug eventually reaching the systemic circulation via the jugular vein. The physicochemical characteristics of a drug are very important for this diffusion process. Although passive diffusion is undoubtedly the major transport mechanism for drugs, the

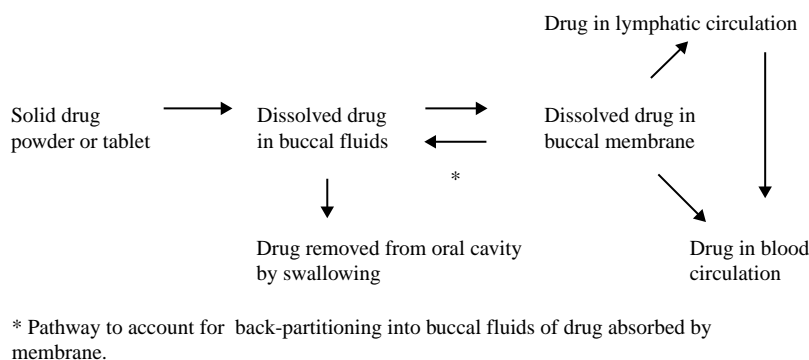


Fig. 1 Schematic representation of the absorption kinetics of buccally presented drugs. (From Ref. 10)

absorption of nutrients from the mouth has been shown to involve carrier systems (*facilitated diffusion*), which lead to a more rapid absorption than the concentration gradient would promote. Such a carrier system, unlike passive diffusion, exhibits stereospecificity, and indeed the absorption of D-glucose and L-arabinose across the buccal mucosa has been shown to be stereospecific (11). The same authors also showed that the absorption of D-glucose, galactose, and 3-O-methyl-D-glucose was at least partially dependent on the presence of sodium ions in the luminal fluids. Furthermore, the transport of D-glucose was inhibited by galactose and 3-O-methyl-D-glucose, suggesting at least one common carrier system. Similarly, Kurosaki et al. (12) demonstrated that the absorption of cefadroxil (a cephalosporin antibiotic) from the human oral cavity occurs through a carrier-mediated mechanism; this absorption was inhibited by the presence of cephalexin, which shares a common carrier-mediated process with cefadroxil in the small intestine of rat.

Membrane storage during buccal absorption of drugs

The absorption of a drug from the mouth is not synonymous with drug entry into the systemic circulation. Instead, the drug appears to be stored in the buccal membranes, sometimes known as the membrane reservoir effect (13). Due to this phenomenon, buccal partitioning has been suggested as a more accurate term to describe the diffusion of drugs across the oral mucosa (14). Although several authors have devised schematic representations of the kinetics of oral drug absorption (Fig. 1) and (1, 14) the mucosal constituents responsible for drug binding have not been identified.

Regional differences in mucosal permeability

The epithelial lining of the mouth differs in both composition (keratinized and nonkeratinized) and

thickness in different regions of the mouth. Therefore, drug absorption may vary from different oral sites. This site-dependent absorption has been shown to take place by Pimlott and Addy (15), who measured the absorption of isosorbide dinitrate into the systemic circulation after applying tablets to the buccal, palatal, or sublingual mucosa in six healthy volunteer subjects. Serum levels of drug were detected from the buccal and sublingual sites after 1 min. The drug concentration progressively increased, peaking at 5 min, and then decreased during the 30-min sampling period. At most of the time periods, serum concentrations were higher from sublingual sites than from buccal sites (Fig. 2). The drug was not detected in the serum of any subject after application to the palatal mucosa. These authors concluded that the keratinized layer of the oral mucosa may be an important barrier to drug absorption because the palatal epithelium is keratinized, but the buccal and sublingual mucosa are not (16). Absorption across the sublingual epithelium is likely to be greater than across the buccal epithelium because the former is thinner and is immersed in a larger volume of saliva.

Rapid absorption from the sublingual mucosa was also demonstrated through work by Al-Furaih et al. (17), who reported that sublingual administration of captopril led to a more rapid attainment of plasma captopril concentrations and had a more rapid pharmacological effect (i.e., lower systolic blood pressure) compared to peroral administration of the drug.

Physicochemical Characteristics of the Drug

Various experimental techniques have demonstrated that cell membranes have a large lipid component (18), and most drugs cross such membranes by simple passive diffusion. In order to cross these lipid membranes, a drug should be in the lipid-soluble or un-ionized form and also

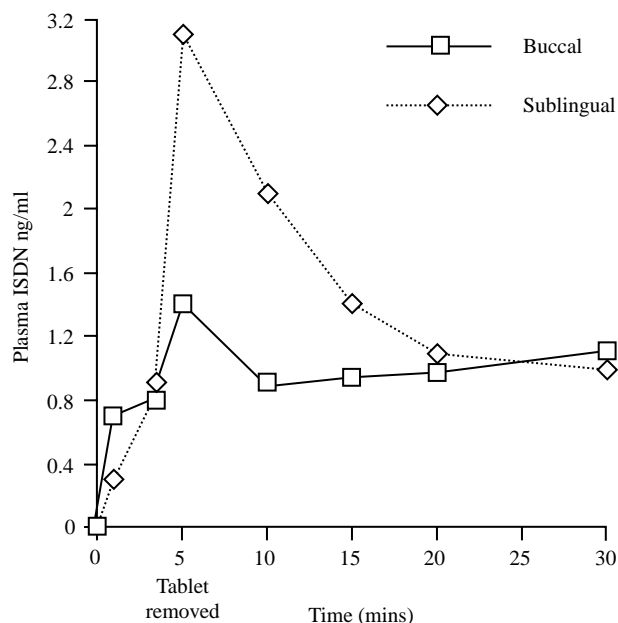


Fig. 2 Mean plasma isosorbide dinitrate concentrations after application of isosorbide dinitrate (5 mg) to the buccal and sublingual mucosa in six healthy male volunteer subjects. (Redrawn from Ref. 15.)

be in solution. The various physicochemical characteristics of the drug are, therefore, of paramount importance as far as drug penetration across the oral mucosa is concerned.

Molecular weight

In general, molecules penetrate the oral mucosa more rapidly than ions, and smaller molecules penetrate more rapidly than larger molecules. However, this rule is not absolute because dextrans with a molecular weight of up to 70,000 cross keratinized rabbit oral mucosa (19), but horseradish peroxidase (molecular weight 40,000) does not (20). High-molecular-weight mucopolysaccharides such as heparin are not well absorbed (21), although inclusion of penetration enhancers in some insulin formulations have improved bioavailability (22).

Degree of ionization

The average pH of saliva is 6.4. Because the un-ionized form of a drug is the lipid-soluble-diffusible form, the pK_a of the drug plays an important role in its absorption across the lipid membranes of the oral mucosa. The degree of ionization of a drug at a specified pH can be calculated using the Henderson–Hasselbalch equation as follows:

For an acid :

$$pH = pK_a + \log_{10} \frac{[\text{un-ionized species}]}{[\text{ionized species}]}$$

For a base :

$$pH = pK_a + \log$$

The importance of pH on drug absorption from the mouth has been extensively studied using the buccal absorption model, in which loss of drug from buffered drug solutions placed in the mouth is monitored (23). The influence of pH on the absorption of the weak base chloroquine and of the weak acid phenobarbitone is shown in Fig. 3 (24).

However, pH does not always influence the rate or extent of absorption. For example, McElnay et al. (25) found that captopril pharmacodynamic parameters (blood pressure, heart rate, and plasma renin activity) did not differ significantly between buffered and unbuffered sublingual administration, suggesting that manipulation of pH had little effect. It was, therefore, proposed that a mechanism other than passive diffusion was involved in the buccal absorption of this drug.

Although many studies illustrate the importance of ionization on drug absorption, the pH of saliva is relatively constant, and in the absence of a buffer, the pK_a of the drug plays the deciding role as to the state of drug ionization.

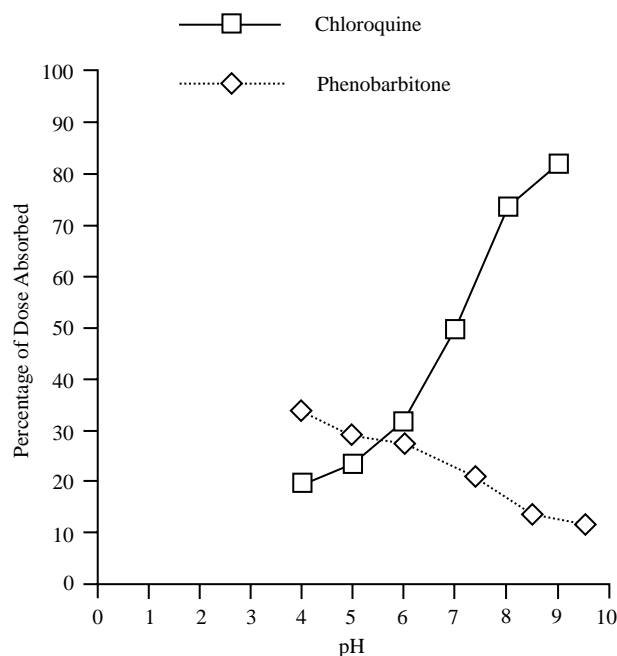


Fig. 3 The influence of pH on the absorption of the weak acid phenobarbitone and the weak base chloroquine from the buccal cavity in three healthy volunteer subjects. (Redrawn from Ref. 24.)

Also, due to the relatively large surface area available for absorption and to the maintenance of an equilibrium between ionized and un-ionized drug, only a small percentage of drug has to be present in the un-ionized form before significant absorption can take place.

Lipid solubility

Although the undissociated (un-ionized) form of a drug has the higher lipid solubility, the un-ionized moieties themselves have differing lipid solubilities. A common way of assessing the lipid solubility of a drug is to measure its oil–water partition coefficient. As with pH, buccal absorption has been shown to be positively correlated with a drug's oil–water partition coefficient. Beckett and Moffat (26), for example, found a correlation of partition coefficients in *n*-heptane/aqueous systems with buccal absorption data for a series of amines and acids when the degree of ionization was held constant.

In conclusion, to penetrate the oral mucosa to a significant degree, a drug should have a relatively low molecular weight and exhibit biphasic solubility patterns, that is, be soluble in both the aqueous salivary fluid and the lipid membrane barrier to penetration. A significant amount of the drug should be un-ionized at salivary pH, and the drug should also not bind strongly to the oral mucosa.

BUCCAL ADMINISTRATION AS A METHOD OF PREVENTING PRESYSTEMIC METABOLISM

The systemic availability of a drug is a measure of the fraction of the administered amount of drug that is absorbed into the general circulation in an unchanged form from its site of administration. Disregarding pharmaceutical reasons (e.g., poor tablet disintegration) and inappropriate physicochemical properties of the drug, the two main reasons for poor bioavailability after peroral administration are drug destruction by stomach acid and drug modification by metabolic enzyme systems prior to its entry into the systemic circulation. The principal organs involved in presystemic elimination are the gut wall, the liver, and the lung (27). Drug metabolism of this type is known as *first-pass* metabolism. A number of drugs have high affinities for the enzyme systems in these organs and are, therefore, highly extracted during their flow through the organs. These drugs, which are said to have a *high extraction ratio* (Fig. 4), include propranolol, terbutaline, levodopa, imipramine, aspirin, morphine, pentazocine, nitroglycerin, lignocaine, hydralazine, verapamil, and methyldopa. The main metabolizing organ in the body is the liver. Because blood draining from the gut via the portal vein must pass

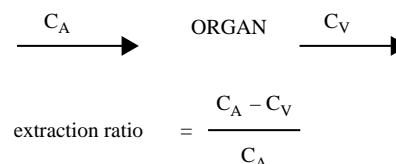


Fig. 4 Diagrammatic representation of the extraction ratio of a drug. The extraction ratio is a measure of the tendency of a drug to be removed from the blood during its passage through an organ such as the liver. In the diagram, C_A is the arterial drug concentration and C_V is the venous drug concentration.

through the liver prior to entry into the general circulation, the total drug absorbed from the gut must pass through the liver before it can reach its site of action. Once the drug has entered the systemic circulation, it is distributed to other areas of the body (depending on its volume distribution); although the extraction ratio remains constant, the proportion of the total drug in the body that is metabolized on subsequent passes through the liver is reduced due to a lowered drug concentration in the plasma after distribution has taken place. The liver receives only 20% of the cardiac output (as compared with 100% from the portal vein), which also protects the drug that has already been absorbed from the metabolic systems of this organ. Presystemic elimination can, therefore, be avoided by choosing a site of administration from which the drug enters the systemic circulation directly, without first passing through the liver, lung, or gut wall. Because blood draining from the oral cavity enters the general circulation via the internal jugular vein, oral administration by the buccal or sublingual routes provides a useful strategy for improving bioavailability of drugs that are susceptible to extensive first-pass metabolism. A high first-pass effect does not, however, mean that drugs with a high extraction ratio cannot be given perorally. If a sufficient dose of the drug is given, an adequate amount of drug (to produce the required therapeutic effect) often remains intact during its first passage through the liver. Also, a high peroral dose of drug or, indeed, serum levels of the drug from previous doses may saturate the high-affinity metabolizing systems in the liver and, thereby, decrease the first-pass effect and increase bioavailability. With some drugs, moreover, the metabolites themselves may have good pharmacologic activity.

DRUGS AND PHARMACEUTICAL FORMULATIONS FOR ADMINISTRATION BY THE BUCCAL AND SUBLINGUAL ROUTES

Although the data produced using the buccal partitioning model of drug absorption (23) have shown that numerous

drugs are absorbed efficiently from the oral cavity, few drugs have been assessed clinically after administration by this route, and not all drugs that have given encouraging clinical data have specific formulations available for intraoral administration. Drugs within the cardiovascular and strong analgesic pharmacologic classes have received the most attention.

Cardiovascular Drugs

Glyceryl trinitrate (GTN)

This vasodilator has been used for over 100 years in the treatment of angina pectoris, and today, many clinicians consider it the most effective drug despite exhaustive efforts to find alternatives. It is also used in the treatment of congestive heart failure. This drug is rapidly absorbed from the mouth, with much of the drug bypassing the liver. The liver has a high metabolic capacity for organic nitrates by virtue of the enzyme glutathione reductase (28). Sublingual administration of GTN is the most appropriate action to alleviate the pain of an acute angina attack because of its rapid action, its long-established efficacy, and its low cost (29). The traditional pharmaceutical formulation of the drug is a rapidly dissolving tablet for administration under the tongue. This approach, however, has two main disadvantages:

1. The time taken for the tablet to disintegrate and dissolve may vary from person to person. A delayed and varied onset of action may result.
2. The tablets of GTN lose significant potency after only 8 weeks of the initial opening of the manufacturer's bottle and should be discarded after that period because exposure to moisture and to the atmosphere accelerates nitrate breakdown. Heat also accelerates drug deterioration.

In an attempt to overcome the previously noted problems with the sublingual tablet formulations, GTN is now widely available in metered-dose aerosol preparations. The sprays usually contain 0.4 mg GTN per unit dose. The manufacturers suggest that 1 or 2 metered doses be sprayed on the oral mucosa (preferably under the tongue) and then the mouth should be closed.

A slightly different approach has been taken by Pharmax, the manufacturer of Suscard Buccal tablets. Instead of the traditional 300-, 500-, and 600- μ g sublingual tablets, the Pharmax tablets contain 1, 2, 3, or 5 mg of GTN and are placed between the upper lip and the gum on either side of the front teeth. During the dissolution phase, the tablet softens and adheres to the

gum, after which dissolution continues in a uniform and gradual manner. Because this is a prolonged-release dosage form, the patient should not increase the tablet's dissolution rate by moving it around the mouth. The tablet should be replaced if accidentally swallowed, and the placement of successive tablets should be alternated on either side of the mouth. As well as an effective prophylactic in angina, this formulation has been shown to be effective in congestive heart failure (30).

Isosorbide dinitrate

This nitrate is also active sublingually and is a more chemically stable drug for those who require nitrates only infrequently. It is a longer-acting drug than GTN. The activity of isosorbide dinitrate may depend on the production of active metabolites, the most important of which is isosorbide 5-mononitrate. Isosorbide mononitrate is also available for angina prophylaxis, though the advantages over isosorbide dinitrate have not yet been firmly established (31). The general consensus is that the activity of the dinitrate is also longer than that of GTN Kattus et al. (32), for example, found that sublingual isosorbide dinitrate offered protection against angina for 2.5–3 h compared to 1 h relief with GTN. The finding of equal bioavailability of chewable (buccal absorption) and slow-release capsules (intestinal absorption) “infers that buccal or sublingual absorption does not circumvent the first pass effect, that presystemic metabolism occurs in the buccal mucosa, that buccal absorption is not as effective as believed or that isosorbide dinitrate is swallowed and not absorbed by the buccal mucosa. The identical pattern of metabolites after buccal and intestinal administration favours the theory that buccal absorption is slow and that isosorbide dinitrate is swallowed with the saliva in which it is dissolved” (33). Current knowledge concerning the buccal absorption route supports this theory. The main advantage of sublingual and buccal dosing may be the rapid disintegration and dissolution of the tablet in saliva. Present knowledge suggests using the drug buccally for the treatment of acute attacks of angina and using a sustained-release formulation for prophylactic purposes. Iga and Ogawa (34) demonstrated that a sustained release buccal formulation of both GTN and isosorbide dinitrate increased the bioavailability of both drugs when administered to dogs, compared to oral administration. A number of isosorbide dinitrate preparations are available for administration by the buccal or sublingual routes, the usual strengths being 5 or 10 mg. Although chewable preparations are available, the more traditional quick-disintegrating tablets predominate. Because mastication tends to increase saliva production, in order to prevent premature swallowing of the drug, the

traditional tablet type may also be preferable. Sublingual rather than buccal administration may also be preferable because higher plasma concentrations have been found in healthy volunteers when the former route was used (15). (Fig. 2)

Nifedipine

In the past, the difficulties presented in the administration of drugs in the treatment of hypertensive emergencies were largely overcome with the use of nifedipine administered sublingually (35). The onset of action was rapid, and the drug was also used sublingually for the treatment of acute attacks of angina pectoris. Presently, two types of formulation of nifedipine are available, both intended primarily for peroral administration. The sustained-release formulation is solely used perorally; however, the rapid-release capsule, which contains nifedipine in solution form, was formerly administered to the buccal cavity. However, the manufacturers now state in their literature that “nifedipine should not be used for the treatment of acute attacks of angina” as it has been associated with large variations in blood pressure and reflex tachycardia (36).

Captopril

Two studies (37, 38) have indicated the usefulness of sublingual captopril in the treatment of severe hypertension. The hypertensive patients thus treated showed a marked decrease in systolic and diastolic blood pressure, with the onset of action being 2–5 min and the peak effect at 10 min (37). Perorally administered captopril takes 1–2 h to achieve a maximal therapeutic effect (39) and, therefore, is unsuitable for the treatment of hypertensive crisis. Al-Furaih et al. (17) reported that sublingual administration of captopril (followed by plasma monitoring of drug levels) led to a more rapid attainment of plasma captopril concentrations and had a more rapid pharmacological effect (i.e., lower systolic blood pressure) compared to peroral administration of the drug as shown in Fig. 5.

Iscan et al. (40) compared a number of parameters of a specially formulated buccal bioadhesive captopril tablet with that of a conventional tablet. The buccal formulation provided controlled release of captopril with a smooth plasma level profile and a long duration of action; however, its bioavailability was 40% via the buccal route as compared to 65% following an oral dose. This was attributed to the intestinal mucosa being more permeable than the buccal mucosa, and it was concluded that further work was required to improve its bioavailability.

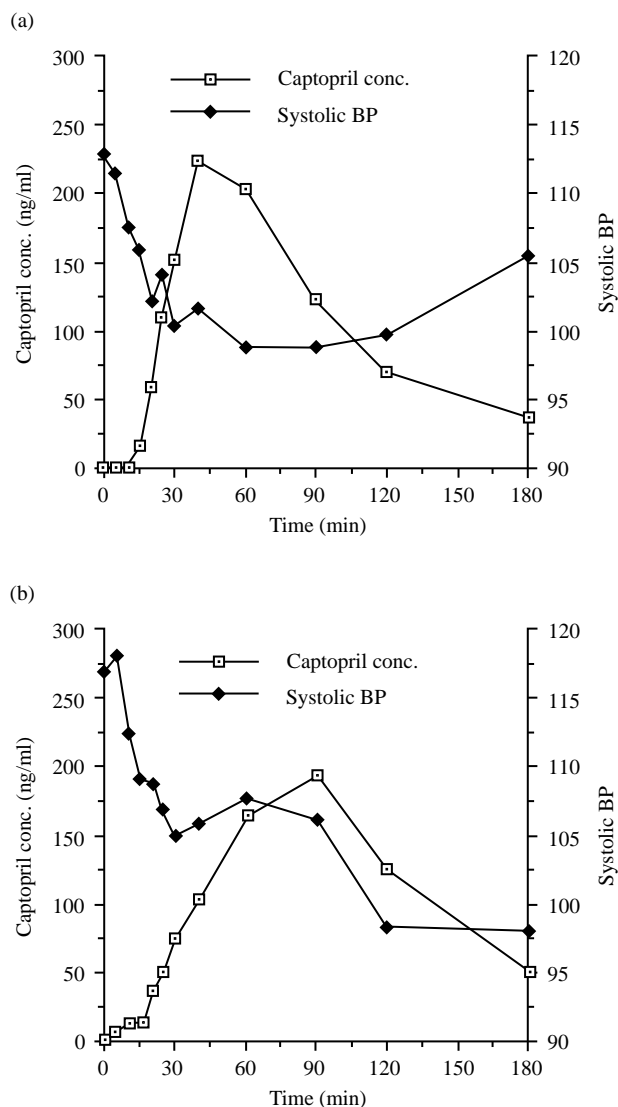


Fig. 5 Correlation over time of systolic blood pressure (◆) with plasma unchanged captopril concentration (□) after sublingual (a) and peroral (b) administration of 25 mg of captopril. (From Ref. 17.)

Analgesics

Buprenorphine

In common with other phenolic opiate analgesics, buprenorphine shows low peroral potency, suggesting a high first-pass metabolism effect; indeed, work in rats has shown this to be the case (27). Intravenous studies have estimated that the extraction ratio of buprenorphine is 85% and that peroral systemic availability is consequently expected to be 15% or less (41). Although absorption from the mouth is slow and, therefore, not as useful as parenteral

administration in the treatment of acute pain, it offers a major bioavailability advantage over the peroral route for this drug (42). If required, the patient can be given a parenteral dose of buprenorphine to achieve rapid pain relief and thereafter be maintained on sublingual drug. The drug is available as a sublingual tablet containing 200 or 400 μg of buprenorphine hydrochloride for the treatment of moderate to severe pain.

Morphine

Although not routinely given by the buccal or sublingual routes, several research studies have shown that absorption of morphine from the mouth gives rise to effective analgesia and that these routes may provide suitable alternatives to parenteral administration (43). Clinical studies have suggested that the bioavailability of morphine is 40–50% greater after buccal than intramuscular administration; as plasma morphine concentrations decline more slowly after buccal administration, buccal morphine may be associated with enhanced analgesia (44). Anlar et al. (45) administered buccoadhesive morphine sulphate tablets to six healthy volunteers, resulting in up to 30% of active drug being absorbed; this is in contrast to absolute bioavailability of a morphine sulphate solution of 23% (46). Christrup et al. (47) found that buccal delivery of morphine sulphate could be enhanced further by using ester prodrugs with higher lipophilicity than the parent drug itself.

Ketobemidone

Ketobemidone is a narcotic analgesic that has been used clinically in Scandinavia and other European countries. The mean bioavailability in humans has been reported to be approximately 35% following oral administration, but this can be substantially improved when administered by the sublingual or buccal route (48). To date, *in vitro* work has focussed on the use of the ketobemidone prodrugs (largely esters), and published results suggest that, as with morphine sulfate, buccal mucosa permeation is greatly improved (49, 50).

Flurbiprofen

Studies have suggested that this nonsteroidal anti-inflammatory drug may be useful in the treatment of periodontal disease. Manipulation of pH was shown to influence the amount of drug absorbed; an increase in pH resulted in a reduction of drug absorbed, thus resulting in a poor local effect (51). Gonzalez-Younes et al. (52) reported that the drug was tightly bound to the membrane of the tissues in the mouth.

Peptide Drugs

The oral mucosa has been cited as a route of administration for peptide drugs as a way of avoiding parenteral delivery, although permeability is low, which reduces its value as a viable option (22). However, modifications to drug formulation may offer greater success. The addition of penetration enhancers to dosage forms appears to improve bioavailability to the greatest extent, by improving the permeability of the epithelium and/or affecting the nature of the drug (53). To date, much of the experimental work has been conducted in animals.

Insulin

To achieve hypoglycemia with insulin, the traditional route of administration has been via subcutaneous injection. Peroral preparations are not feasible due to the degradation of insulin by gastric acid and enzymes. However, studies carried out in animals utilising buccal formulations have been more successful. Ritschel et al. (54) administered insulin to beagle dogs using solutions of different pHs. Bioavailability (22.3%) was maximized at pH 7.5, and the addition of penetration enhancers (bile salts) did not increase this further. Experimental work in rabbits found that in the absence of penetration enhancers, insulin solutions over a range of pHs did not show any significant hypoglycaemic response, indicating that insulin was not absorbed to a significant degree through the buccal mucosa (55).

Buserelin

This luteinizing hormone-releasing hormone has been used in the treatment of endometriosis and hormone-dependent tumors. Modes of administration have included injections, nasal sprays and subcutaneous implantations. One study, conducted in pigs, demonstrated the value of glycodeoxycholate (a penetration enhancer) in improving the bioavailability of buserelin by up to five-fold after buccal delivery (56).

α -Interferon

α -Interferon has broad antiviral and antiproliferative activity and has been used in HIV and certain forms of hepatitis. As with other peptide drugs, ways have been sought to improve the buccal delivery of α -interferon to avoid gastro-intestinal degradation and first-pass metabolism. Using a range of penetration enhancers, Stewart et al. (57) noted improved bioavailability, particularly with sodium taurocholate, in rats. As with all of the studies reviewed under the category of peptide drugs, extrapolation to the human situation should be done with caution.

Miscellaneous Drugs

Nicotine

The absorption of nicotine from chewing tobacco has been widely used for many years. Buccal absorption of nicotine is also the route of absorption for pipe and cigar smokers if the smoke is not inhaled. Nicotine replacement therapy has been used in smoking cessation strategies. Nicotine in the form of chewing gum carries no cancer risk and is a useful part of a smoking cessation strategy (58). A recent innovation has been the development of a sublingual tablet (available in 2 mg); a clinical trial has shown that this formulation is a safe form of administration (59), and patients may use one tablet every 1–2 h.

Zinc

Zinc gluconate in the form of a lozenge has been marketed for the treatment of the common cold; there has been no definitive conclusion as to whether it is effective in treating cold symptoms, although differences in study methodology may partially explain the conflicting results that have been reported (60, 61).

Midazolam

Midazolam administered buccally in solution has been shown to be rapidly absorbed and produces changes in EEG readings (62). The authors suggested that this may offer an alternative to rectal administration of diazepam in the emergency treatment of seizures.

CONCLUSIONS

The buccal cavity provides a highly vascular mucous membrane site for the administration of drugs. The epithelial lining of the oral cavity differs both in type (keratinized and nonkeratinized) and in thickness in different areas, and the differences give rise to regional variations in permeability to drugs. Although some macromolecules have been shown to cross the absorption barrier (lipid membrane), the absorption of smaller drug molecules occurs more reproducibly and rapidly. The main absorption mechanism is passive diffusion of the un-ionized (lipid-soluble) form of the drug. Facilitated diffusion has also been shown to take place with nutrients. Drugs are often stored or bound to the buccal mucosa prior to entry into the bloodstream. The blood drainage from the mouth enters the general circulation directly without first passing through the liver. This feature enhances the bioavailability of certain drugs as compared with peroral administration because first-pass metabolism is avoided.

To ensure adequate absorption from the mouth, a drug administered as a solid dosage form must exhibit *biphasic solubility*, that is, be soluble in saliva and in the lipid membranes of the buccal cavity. The major drugs currently available for buccal administration fall within the vasodilator and strong analgesic pharmacologic classes. Although the main type of formulation available for buccal absorption is rapidly disintegrating tablets, new approaches include mucoadhesive tablets and spray formulations. Much of the current knowledge on the mechanism and characteristics of drug absorption from the buccal cavity has been gained from volunteer studies in which buffered drug solutions are placed in the mouth (buccal absorption model) rather than from clinical pharmacologic studies in patients. This former method provides useful information on the bioavailability of new and existing drugs. The main advantages of the buccal route of administration over the traditional peroral route are that drug degradation in the stomach is avoided, first-pass metabolism is avoided, and therapeutic blood levels of drug can be achieved rapidly. Clearly these advantages are presently clinically relevant for only a limited number of drugs. However, with the recent developments of formulation types, such as mucoadhesive preparations and the use of peptides as drugs, this number may increase in the future. The main disadvantage of the buccal route is that the drug may have an unwanted local effect in the mouth, such as bad taste, or may be absorbed slowly and, therefore, be swallowed prior to sufficient absorption taking place. The buccal route, like the rectal and intranasal routes, has been largely neglected by clinicians and manufacturing companies in the past and clearly merits further intensive research.

REFERENCES

1. Harris, D.; Robinson, J.R. Drug Delivery Via the Mucous Membranes of the Oral Cavity. *J. Pharm. Sci.* **1992**, *81*, 1–10.
2. Beckett, A.H.; Hossie, R.D. Buccal Absorption of Drugs. *Handbuch der Experimentellen Pharmakologie*; Gillette, T.R., Ackermann, H.S., Eds.; Springer-Verlag: Berlin, 1971; 28, 1–60.
3. Feldman, R.S.; Szabó, G. Comparative Microanatomy of Skin and Oral Mucosa. *Textbook of Oral Biology*, 1st Ed.; Shaw, J.H., Sweeney, E.A., Cappuccino, C.C., Meller, S.M., Eds.; W.B. Saunders: Philadelphia, 1978; 166–167.
4. Squier, C.A.; Rooney, L. The Permeability of Keratinized and Nonkeratinized Oral Epithelium to Lanthanum In Vivo. *J. Ultrastruct. Res.* **1976**, *54*, 286–295.
5. Schroeder, H.E. Classification of Stratified Epithelia. *Differentiation of Human Oral Stratified Epithelia*; Karger: Basel, 1981; 33.

6. Chen, S.Y.; Squier, C.A. The Ultrastructure of the Oral Epithelium. *The Structure and Function of Oral Mucosa*; Meyer, J., Squier, C.A., Gerson, S.J., Eds.; Pergamon Press: Oxford, 1984; 7–30.
7. Stablein, M.J.; Meyer, J. The Vascular System and Blood Supply. *The Structure and Function of Oral Mucosa*; Meyer, J., Squier, C.A., Gerson, S.J., Eds.; Pergamon Press: Oxford, 1984; 237–256.
8. Netter, F.H. Upper Digestive Tract. *The Digestive System, Part 1: The Ciba Collection of Medical Illustrations*; Oppenheimer, E., Ed.; Ciba: New York, 1959; 3, 24–28.
9. Boyer, C.C.; Neptune, C.M. Patterns of Blood Supply to Teeth and Adjacent Tissue. *J. Dent. Res.* **1962**, *41*, 158–171.
10. Siegal, I.A. Permeability of the Oral Mucosa. *The Structure and Function of Oral Mucosa*; Meyer, J., Squier, C.A., Gerson, S.J., Eds.; Pergamon Press: Oxford, 1984; 95–108.
11. Manning, A.S.; Evered, D.F. The Absorption of Sugars from the Human Buccal Cavity. *Clin. Sci. Mol. Med.* **1976**, *51*, 127–132.
12. Kurosaki, Y.; Nishimura, H.; Terao, K.; Nakayama, T.; Kimura, T. Existence of a Specialised Absorption Mechanism for Cefadroxil, an Aminocapthosporin Antibiotic, in the Human Oral Cavity. *Int. J. Pharm.* **1992**, *82*, 165–169.
13. Ho, H.F.H. Biophysical Kinetic Modeling of Buccal Absorption. *Adv. Drug Del. Rev.* **1993**, *12*, 61–97.
14. Henry, J.A.; Ohashi, K.; Wadsworth, J.; Turner, P. Drug Recovery Following Buccal Absorption of Propranolol. *Br. J. Clin. Pharmacol.* **1980**, *10*, 61–65.
15. Pimlott, S.J.; Addy, M. A Study into the Mucosal Absorption of Isosorbide Dinitrate at Different Intra-Oral Sites. *Oral Surg. Oral Med. Oral Pathol.* **1985**, *59*, 145–148.
16. Squier, C.A.; Johnson, N.W.; Hopps, R.M. The Human Oral Mucosa: Innervation of the Oral Mucosa. *Human Oral Mucosa*; Blackwell Scientific Publications: Oxford, 1975; 63.
17. Al-Furaih, T.A.; McElnay, J.C.; Elborn, J.S.; Rusk, R.; Scott, M.G.; McMahon, J.; Nicholls, D.P. Sublingual Captopril—A Pharmacokinetic and Pharmacodynamic Evaluation. *Eur. J. Clin. Pharmacol.* **1991**, *40*, 393–398.
18. Gurr, M.I.; Harwood, J.L. Lipids in Cellular Structure: Membrane Structure. *Lipid Biochemistry*; Chapman and Hall: London, 1991; 265–281.
19. Squier, C.A.; Johnston, N.W. Permeability of Oral Mucosa. *Br. Med. Bull.* **1975**, *31*, 169–175.
20. Squier, C.A. The Permeability of Keratinized and Nonkeratinized Oral Epithelium to Horseradish Peroxidase. *J. Ultrastruct. Res.* **1973**, *43*, 160–177.
21. Fuller, H.L. Effect of Sublingual Heparin on Lipemia Clearing and on Recurrence of Myocardial Infarction. *Angiology* **1960**, *11*, 200–206.
22. Merkle, H.P.; Wolany, G. Buccal Delivery for Peptide Drugs. *J. Controlled Release* **1992**, *21*, 155–164.
23. Beckett, A.H.; Triggs, E.F. Buccal Absorption of Basic Drugs and its Application as an In Vivo Model of Passive Drug Transfer through Lipid Membranes. *J. Pharm. Pharmacol.* **1967**, *19*, 31S–41S.
24. McElnay, J.C.; Sidahmed, A.M.; D'Arcy, P.F. Experimental Modeling of Drug Absorption and Drug Absorption Interactions. *Int. J. Pharm.* **1986**, *31*, 107–117.
25. McElnay, J.C.; Al-Furaih, T.A.; Hughes, C.M.; Scott, M.G.; Elborn, J.S.; Nicholls, D.P. The Effect of pH on the Buccal and Sublingual Absorption of Captopril. *Eur. J. Clin. Pharmacol.* **1995**, *48*, 373–379.
26. Beckett, A.H.; Moffat, A.C. Correlation of Partition Coefficients in *n*-Heptane-Aqueous Systems with Buccal Absorption Data for a Series of Amines and Acids. *J. Pharm. Pharmacol.* **1969**, *21*, 144S–150S.
27. Brewster, D.; Humphrey, M.J.; McLeavy, M.A. The Systemic Bioavailability of Buprenorphine by Various Routes of Administration. *J. Pharm. Pharmacol.* **1981**, *33*, 500–506.
28. Needleman, P.; Hunter, F.E. The Transformation of Glyceryl Trinitrate and other Nitrates by Glutathione-Organic Nitrate Reductase. *Mol. Pharmacol.* **1965**, *1*, 77–86.
29. Goldstein, R.E.; Rosing, D.R.; Redwood, D.R.; Beiser, F.D.; Epstein, S.E. Clinical and Circulatory Effects of Isosorbide Dinitrate Comparison with Nitroglycerin. *Circulation* **1971**, *43*, 629–639.
30. Sanghera, S.S.; Goldberg, A.A.J.; Parson, D.F. Buccal Nitroglycerin in Elderly Patients with Congestive Heart Failure. *Practitioner* **1985**, *229*, 1054–1055.
31. British National Formulary No. 40, British Medical Association and the Pharmaceutical Society of Great Britain. The Pharmaceutical Press: London, 2000.
32. Kattus, A.A.; Alvaro, A.B.; Zohman, L.R.; Coulson, A. Comparison of Placebo, Nitroglycerin, and Isosorbide Dinitrate for Effectiveness of Relief of Angina and Duration of Action. *Chest* **1979**, *75*, 17–21.
33. Galeazzi, R.L.; Platzer, R.; Reuteman, G. Plasma Isosorbide Dinitrate Concentrations and Effect after Chewable and High-Dose Sustained-Release Formulations. *Int. J. Clin. Pharmacol. Ther.* **1983**, *21*, 393–397.
34. Iga, K.; Ogawa, Y. Sustained-Release Buccal Dosage Forms for Nitroglycerin and Isosorbide Dinitrate: Increased Bioavailability and Extended Time of Absorption when Administered to Dogs. *J. Controlled Release* **1997**, *49*, 105–113.
35. Mathur, M.M.; Chaturvedi, M.K.; Rai, R.R.; Das, Khatri T. Role of Nifedipine Sublingually in Accelerated Hypertension. *J. Assoc. Phys. India* **1986**, *34*, 137–138.
36. A., B.P.I. *Compendium of Data Sheets and Summaries of Product Characteristics 1999–2000*; Datapharm Publications: London, 1999.
37. Tschollar, W.; Belz, G.G. Sublingual Captopril in Hypertensive Crisis. *Lancet* **1985**, *2*, 34–35.
38. Hauger-Kleven, J.H. Comparison of Sublingual Captopril and Nifedipine. *Lancet* **1986**, *1*, 219.
39. Ohman, K.P.; Kagedal, B.; Larsson, R.; Karlberg, B.E. Pharmacokinetics of Captopril and its Effects on Blood Pressure during Acute and Chronic Administration and in Relation to Food Intake. *J. Cardiovasc. Pharmacol.* **1985**, *7* (Suppl. 1), 20–24.
40. Iscan, Y.Y.; Capan, Y.; Senel, S.; Sahin, M.F.; Kes, S.; Duchene, D.; Hincal, A.A. Formulation and In Vitro/In Vivo Evaluation of Buccal Bioadhesive Captopril Tablets. *STP Pharma. Sci.* **1998**, *8*, 357–363.
41. Bullingham, R.E.S.; McQuay, H.J.; Moore, R.A.; Bennett, M.R.D. Buprenorphine Kinetics. *Clin. Pharmacol. Ther.* **1980**, *28*, 667–672.

42. Cassidy, J.P.; Landzert, N.M.; Quadros, E. Controlled Buccal Delivery of Buprenorphine. *J. Controlled Release* **1993**, *25*, 21–29.
43. Whitman, H.H. Sublingual Morphine: A Novel Route of Narcotic Administration. *Am. J. Nursing* **1984**, *84*, 939.
44. Bell, M.D.D.; Murray, G.R.; Mishra, P.; Calvey, T.N.; Weldon, B.D.; Williams, N.E. Buccal Morphine—A New Route for Analgesia. *Lancet* **1985**, *1*, 71–73.
45. Anlar, S.; Capan, Y.; Guven, O.; Gogus, A.; Dalkara, T.; Hincal, A.A. Formulation and In Vitro–In Vivo Evaluation of Buccoadhesive Morphine Sulphate Tablets. *Pharm. Res.* **1994**, *11*, 231–236.
46. Hoskin, P.J.; Hanks, G.W.; Aheme, G.W.; Chapman, D.; Littleton, P.; Filshie, J. The Bioavailability and Pharmacokinetics of Morphine after Intravenous, Oral and Buccal Administration in Healthy Volunteers. *Br. J. Clin. Pharmacol.* **1989**, *27*, 499–505.
47. Christrup, L.L.; Christensen, C.B.; Friis, G.J.; Jorgensen, A. Improvement of Buccal Delivery of Morphine Using the Prodrug Approach. *Int. J. Pharm.* **1997**, *154*, 157–165.
48. Hansen, L.B.; Christrup, L.L.; Bundgaard, H. Ketobemidone Prodrugs for Buccal Delivery. *Acta Pharma. Nord.* **1991**, *3*, 77–82.
49. Hansen, L.B.; Christrup, L.L.; Bundgaard, H. Enhanced Delivery of Ketobemidone through Porcine Buccal Mucosa In Vitro via more Lipophilic Ester Prodrugs. *Int. J. Pharm.* **1992**, *88*, 237–242.
50. Hansen, L.B.; Jorgensen, A.; Rasmussen, S.N.; Christrup, L.L.; Bundgaard, H. Buccal Absorption of Ketobemidone and Various Ester Prodrugs in the Rat. *Int. J. Pharm.* **1992**, *88*, 243–250.
51. Barsuhn, C.L.; Olanoff, L.S.; Gleason, D.D.; Adkins, E.L.; Ho, N.F.H. Human Buccal Absorption of Flurbiprofen. *Clin. Pharmacol. Ther.* **1988**, *44*, 225–231.
52. Gonzalez-Younes, I.; Wagner, J.G.; Gaines, D.A.; Ferry, J.J.; Hageman, J.M. Absorption of Flurbiprofen through Human Buccal Mucosa. *J. Pharm. Sci.* **1991**, *80*, 820–823.
53. Hoogstraate, A.J.; Bodde, H.E. Methods for Assessing the Buccal Mucosa as a Route of Drug Delivery. *Adv. Drug Del. Rev.* **1993**, *12*, 99–125.
54. Ritschel, W.A.; Ritschel, G.B.; Forusz, H.; Kraeling, M. Buccal Absorption of Insulin in the Dog. *Res. Comm. Chem. Path. Pharmacol.* **1989**, *63*, 53–67.
55. Oh, C.K.; Ritschel, W.A. Biopharmaceutic Aspects of Buccal Absorption of Insulin. *Meth. Find. Exp. Clin. Pharmacol.* **1990**, *12*, 205–212.
56. Hoogstraate, A.J.; Verhoef, J.C.; Pijpers, A.; van Leengoed, L.A.M.G.; Verheijden, J.H.M.; Junginger, H.E.; Bodde, H.E. In Vivo Buccal Delivery of the Peptide Drug Buserelin with Glycodeoxycholate as an Absorption Enhancer in Pigs. *Pharm. Res.* **1996**, *13*, 1233–1237.
57. Steward, A.; Bayley, D.L.; Howes, C. The Effect of Enhancers on the Buccal Absorption of Hybrid (BDBB) α -Interferon. *Int. J. Pharm.* **1994**, *104*, 145–149.
58. Henningfield, J.E. Nicotine Medications for Smoking Cessation. *N. Engl. J. Med.* **1995**, *333*, 1196–1203.
59. Wallstrom, M.; Sand, L.; Nilsson, F.; Hirsch, J.M. The Long-Term Effect of Nicotine on the Oral Mucosa. *Addiction* **1999**, *94*, 417–423.
60. Mossad, S.B.; Macknin, M.L.; Medendorp, S.V.; Mason, P. Zinc Gluconate Lozenges for Treating the Common Cold—A Randomized Double-blind, Placebo-Controlled Study. *Ann. Intern. Med.* **1996**, *125*, 81–89.
61. Macknin, M.L.; Piedmonte, M.; Calendine, C.; Janosky, J.; Wald, E. Zinc Gluconate Lozenges for Treating the Common Cold in Children—A Randomized Controlled Trial. *JAMA* **1998**, *279*, 1962–1967.
62. Scott, R.C.; Besag, F.M.C.; Boyd, S.G.; Berry, D.; Neville, B.G.R. Buccal Absorption of Midazolam: Pharmacokinetics and EEG Pharmacodynamics. *Epilepsia* **1998**, *39*, 290–294.

FURTHER READING

- Chidambaran, N.; Srivatsava, A.K. Buccal Drug Delivery Systems. *Drug Dev. Indust. Pharm.* **1995**, *21*, 1009–1036.
- Gandhi, R.B.; Robinson, J.R. Oral Cavity as a Site for Bioadhesive Drug Delivery. *Adv. Drug Del. Rev.* **1994**, *13*, 43–74.
- Harris, D.; Robinson, J.R. Drug Delivery via the Mucous Membranes of the Oral Cavity. *J. Pharm. Sci.* **1992**, *81*, 1–10.
- Meyer, J.; Squier, C.A.; Gerson, S.J. *The Structure and Function of the Oral Mucosa*; Pergamon Press: Oxford, 1984.
- Rathbone, M.J.; Tucker, I.G. Mechanisms, Barriers and Pathways of Oral Mucosal Drug Permeation. *Adv. Drug Del. Rev.* **1993**, *12*, 41–60.
- Rathbone, M.J.; Drummond, B.K.; Tucker, I.G. The Oral Cavity as a Site for Systemic Drug Delivery. *Adv. Drug Del. Rev.* **1994**, *13*, 1–22.

DRUG DELIVERY—CONTROLLED RELEASE

Yie W. Chien

Kaohsiung Medical University, Kaohsiung, Taiwan

Senshang Lin

St. John's University, Jamaica, New York, U.S.A.

INTRODUCTION

Over the past decades, the treatment of illness has been accomplished by administering drugs to the human body via various pharmaceutical dosage forms, like tablets. These traditional pharmaceutical products are still commonly seen today in the prescription and over-the-counter drug marketplace. To achieve and maintain the drug concentration in the body within the therapeutic range required for a medication, it is often necessary to take this type of drug delivery system several times a day. This yields an undesirable “seesaw” drug level in the body (Fig. 1).

A number of advancements have been made recently in the development of new techniques for drug delivery. These techniques are capable of regulating the rate of drug delivery, sustaining the duration of therapeutic action, and/or targeting the delivery of drug to a specific tissue (1–6). These advancements have already led to the development of several novel drug delivery systems that could provide one or more of the following benefits:

1. Controlled administration of a therapeutic dose at a desirable rate of delivery
2. Maintenance of drug concentration within an optimal therapeutic range for prolonged duration of treatment
3. Maximization of efficacy-dose relationship
4. Reduction of adverse side effects
5. Minimization of the needs for frequent dose intake
6. Enhancement of patient compliance

Based on the technical sophistication of the controlled-release drug delivery systems (CrDDSs) that have been marketed so far, or that are under active development, the CrDDSs can be classified (Fig. 2) as follows:

1. Rate-preprogrammed drug delivery systems
2. Activation-modulated drug delivery systems
3. Feedback-regulated drug delivery systems
4. Site-targeting drug delivery systems

In this article, the scientific concepts and technical principles behind the development of this new generation of drug-delivery systems are outlined and discussed.

RATE-PREPROGRAMMED DRUG DELIVERY SYSTEMS

In this group of CrDDSs, the release of drug molecules from the delivery systems has been preprogrammed at a specific rate profile. This is accomplished by system design, which controls the molecular diffusion of drug molecules in and/or across the barrier medium within or surrounding the delivery system. Fick's laws of diffusion are often followed. These CrDDSs can further be classified as follows:

1. Polymer membrane permeation-controlled drug delivery systems
2. Polymer matrix diffusion-controlled drug delivery systems
3. Polymer (membrane/matrix) hybrid-type drug delivery systems
4. Microreservoir partition-controlled drug delivery systems

Polymer Membrane Permeation-Controlled Drug Delivery Systems

In this type of CrDDS, a drug formulation is either totally or partially encapsulated in a drug reservoir compartment whose drug-releasing surface is covered by a rate-controlling polymeric membrane. The drug reservoir can be drug solid particles, a dispersion of drug solid particles, or a concentrated drug solution in a liquid- or solid-type dispersing medium. The polymeric membrane can be fabricated from a homogeneous or a heterogeneous nonporous polymeric material or a microporous or semipermeable membrane. The encapsulation of drug formulation inside the reservoir compartment can be accomplished by molding, capsulation,

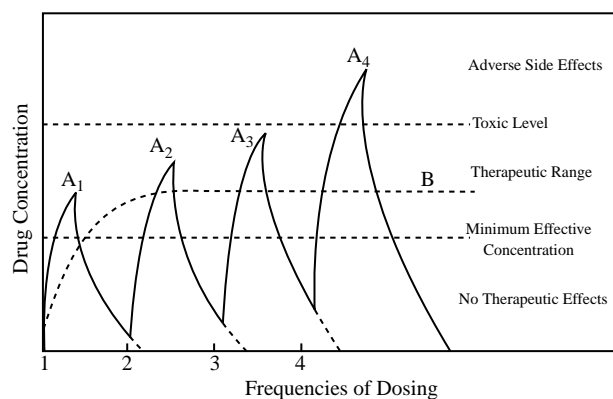


Fig. 1 Drug concentration profiles in the systemic circulation as a result of taking a series of multiple doses of a conventional drug-delivery system (A_1, A_2, \dots) in comparison with the ideal drug concentration profile (B). (Adapted from Ref. 6.)

microencapsulation, or other techniques. Different shapes and sizes of drug delivery systems can be fabricated (Fig. 3).

The release of drug from this type of CrDDSs should be at a constant rate (Q/t), which is defined by the following general equation:

$$\frac{Q}{t} = \frac{K_{m/r} K_{a/m} D_d D_m}{K_{m/r} D_m h_d + K_{a/m} D_d h_m} C_R \quad (1)$$

where $K_{m/r}$ and $K_{a/m}$ are, respectively, the partition coefficients for the interfacial partitioning of drug molecules from the reservoir to the membrane and from the membrane to the aqueous diffusion layer; D_m and D_d are, respectively, the diffusion coefficients in the rate-controlling membrane with a thickness of h_m , and in the aqueous diffusion layer with a thickness of h_d . For microporous membrane, the porosity, and tortuosity of the pores in the membrane should be included in the estimation of D_m and h_m . C_R is the drug concentration in the reservoir compartment.

The release of drug molecules from this type of CrDDS is controlled at a preprogrammed rate by modulating the partition coefficient and the diffusivity of drug molecule and the rate-controlling membrane and the thickness of the membrane. Several CrDDSs of this type have been successfully marketed for therapeutical uses and some representatives are outlined later for illustration.

Progestasert® IUD

In this controlled-release intrauterine device, the drug reservoir exists as a dispersion of progesterone crystals in silicone (medical grade) fluid encapsulated in the vertical limb of a T-shaped device walled by a nonporous membrane of ethylene–vinyl acetate copolymer (Fig. 4). It is engineered to release continuously a daily dose of 65 μg progesterone inside the uterine cavity to achieve

contraception for one year (6). The same technology has been utilized in the development of the Mirena® system, a plastic T-shaped frame with a steroid reservoir containing 52 mg levonorgestrel, which is designed to release a daily dose of levonorgestrel at $\sim 20 \mu\text{g}/\text{day}$ for achieving effective contraception for five years (7–9).

Ocusert® system

In this controlled-release ocular insert, the drug reservoir is a thin disc of pilocarpine–alginate complex sandwiched between two transparent discs of microporous membrane fabricated from ethylene–vinyl acetate copolymer (Fig. 5). The microporous membranes permit the tear fluid to penetrate into the drug reservoir compartment to dissolve pilocarpine from the complex. Pilocarpine molecules are then released at a constant rate of 20 or 40 $\mu\text{g}/\text{h}$ for a 4- to 7-day management of glaucoma (1, 6, 10, 11).

Transderm-Nitro® system

In this controlled-release transdermal therapeutic system, the drug reservoir, which is a dispersion of nitroglycerin–lactose triturate in a silicone (medical grade) fluid, is encapsulated in an ellipsoid-shaped thin patch. The drug reservoir is sandwiched between a drug-impermeable metallic plastic laminate, as the backing membrane, and a constant surface of drug-permeable, rate-controlling membrane of ethylene–vinyl acetate copolymer (Fig. 6). This device is fabricated by an injection-molding process. A thin layer of silicone adhesive is further coated on the drug-permeable membrane in order that an intimate contact of the drug-releasing surface with the skin surface is achieved and maintained. It is engineered to have nitroglycerin delivered transdermally at a rate of 0.5 (mg/cm^2)/day for a daily relief of angina (2, 3).

The same technology has been utilized in the development of the following: 1) the Estraderm® system, which administers a controlled dose of estradiol transdermally over 3–4 days for the relief of postmenopausal syndrome and osteoporosis (12–14); 2) the Duragesic® system, which provides a transdermal-controlled administration of fentanyl, a potent narcotic analgesic, for 72-h relief of chronic pain (14); and 3) the Androderm® system, which provides a transdermal-controlled delivery of testosterone, through nonscrotal skin, for the 24-h replacement therapy of testosterone-deficient patients (14).

Norplant® Subdermal Implant

The controlled-release subdermal implant is fabricated from a nonporous silicone (medical-grade) tubing, by sealing both ends with silicone (medical-grade) adhesive to encapsulate either levonorgestrel crystals alone (generation I) or a solid dispersion of levonorgestrel in

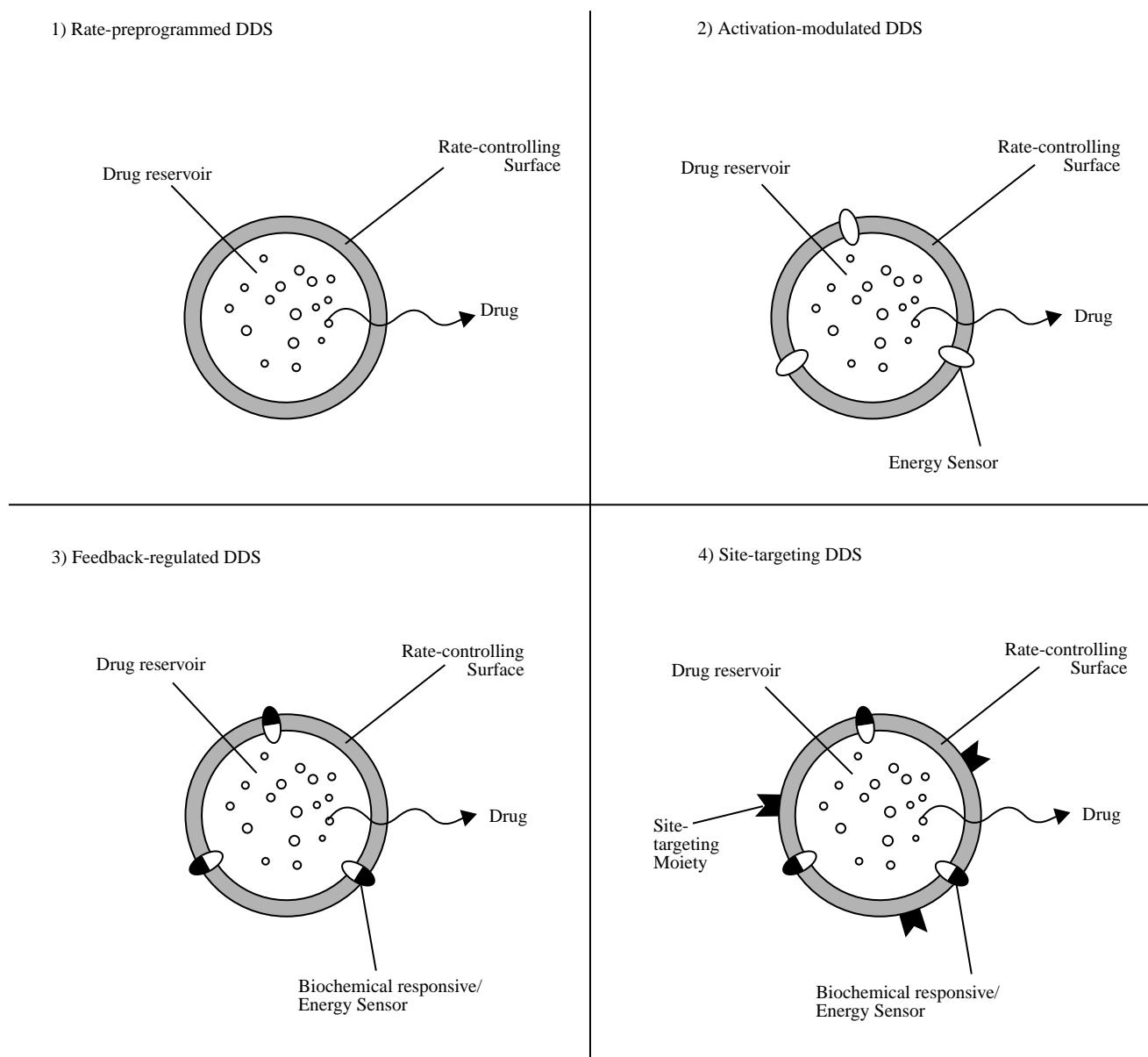


Fig. 2 The four major classes of controlled-release drug delivery systems.

silicone elastomer matrix (generation II). It is designed to attain a continuous release of levonorgestrel, at a daily dosage rate of 30 μg , to each subject (following the subcutaneous implantation of either 6 units of I or 2 units of II); (Fig. 7) for up to 7 years (15–18).

Polymer Matrix Diffusion-Controlled Drug Delivery Systems

In this type of CrDDS, the drug reservoir is produced from the homogeneous dispersion of drug particles in either a lipophilic or a hydrophilic polymer matrix.

The drug dispersion in the polymer matrix is accomplished by either 1) blending a dose of finely ground drug particles with a viscous liquid (or a semisolid) polymer, followed by a crosslinking of polymer chains or 2) mixing drug solids with a melted polymer at an elevated temperature. The resultant drug-polymer dispersion is then molded or extruded to form drug-delivery devices of various shapes and sizes designed for a specific application (Fig. 8). It can also be fabricated by dissolving the drug and the polymer in a common solvent, followed by solvent evaporation, at an elevated temperature and/or under a vacuum, in a mold.

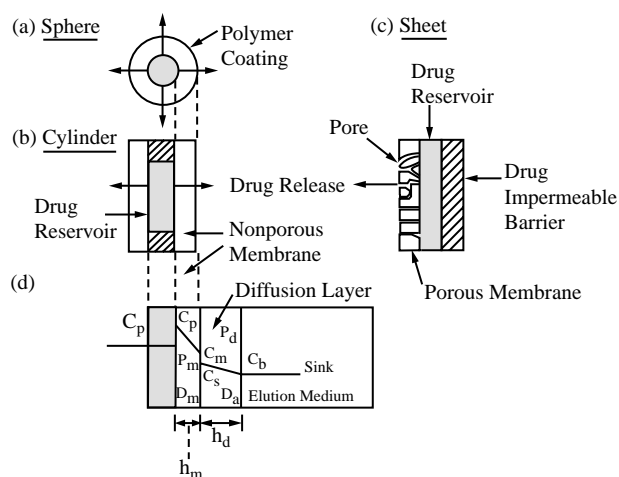


Fig. 3 Release of drug from various shapes of polymer membrane permeation-controlled drug-delivery systems: (a) sphere-type, (b) cylinder-type, and (c) sheet-type. In (d), the drug concentration gradients across the rate-controlling polymeric membrane and hydrodynamic diffusion layer exist in series. Both the polymer membrane, which is either porous or nonporous, and the diffusion layer have a controlled thickness (h_m and h_d , respectively).

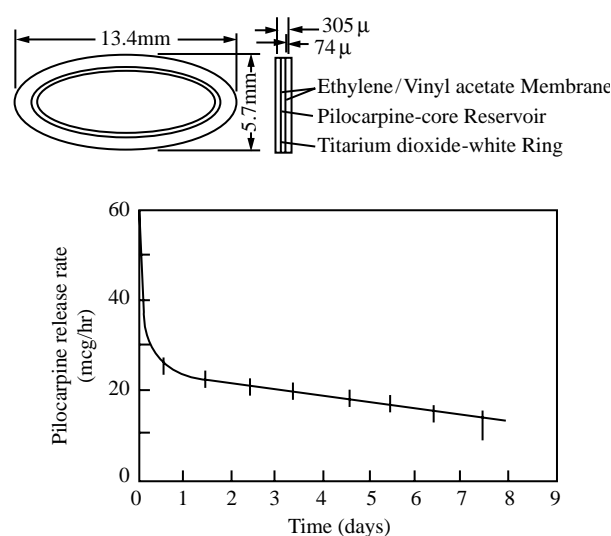


Fig. 5 Diagrammatic illustration of a unit of Ocusert® system, showing various structural components, and the ocular release rate profile of pilocarpine from the Ocusert pilo-20 system. (From Ref. 11.)

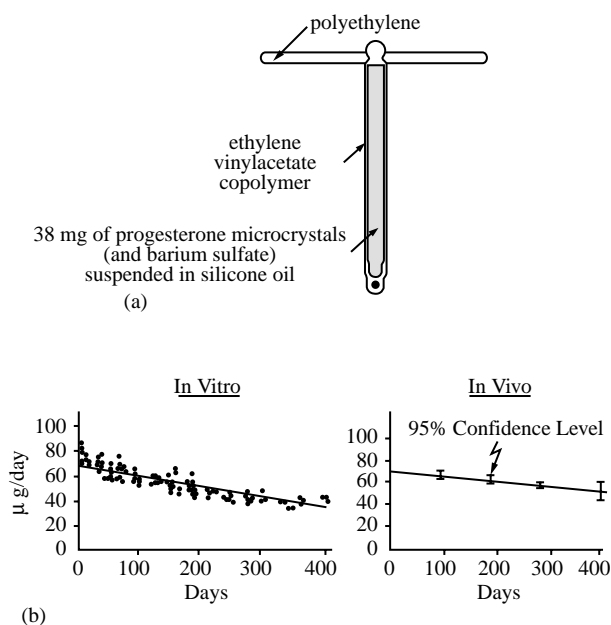


Fig. 4 Diagrammatic illustration of a unit of Progestasert IUD, showing various structural components (a) and the in vitro and in vivo delivery rate profiles of progesterone for up to 400 days (b).

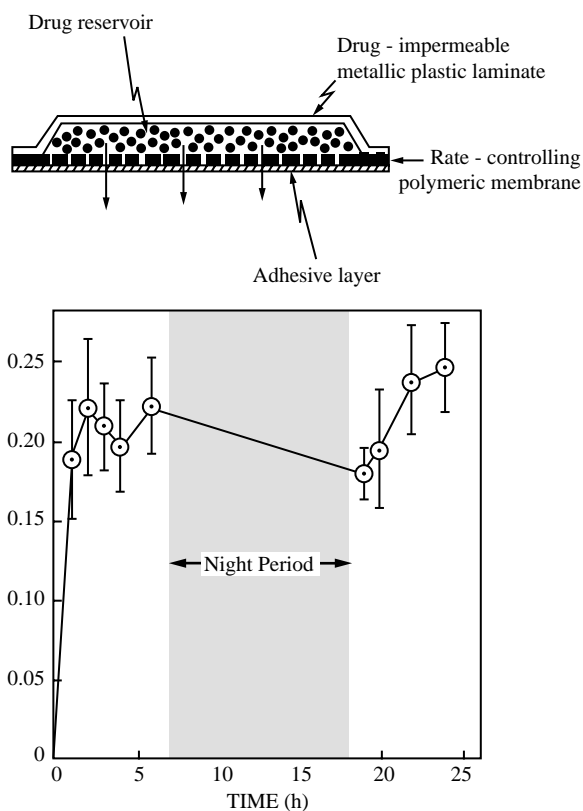


Fig. 6 Cross-sectional view of a unit of Transderm-Nitro® system, showing various structural components, and plasma concentration profiles of nitroglycerin in 14 human volunteers, each receiving one unit of Transderm-Nitro system (20 cm², with a delivery rate of 10 mg/day) for 24 h. (From Refs. 11, 55.)

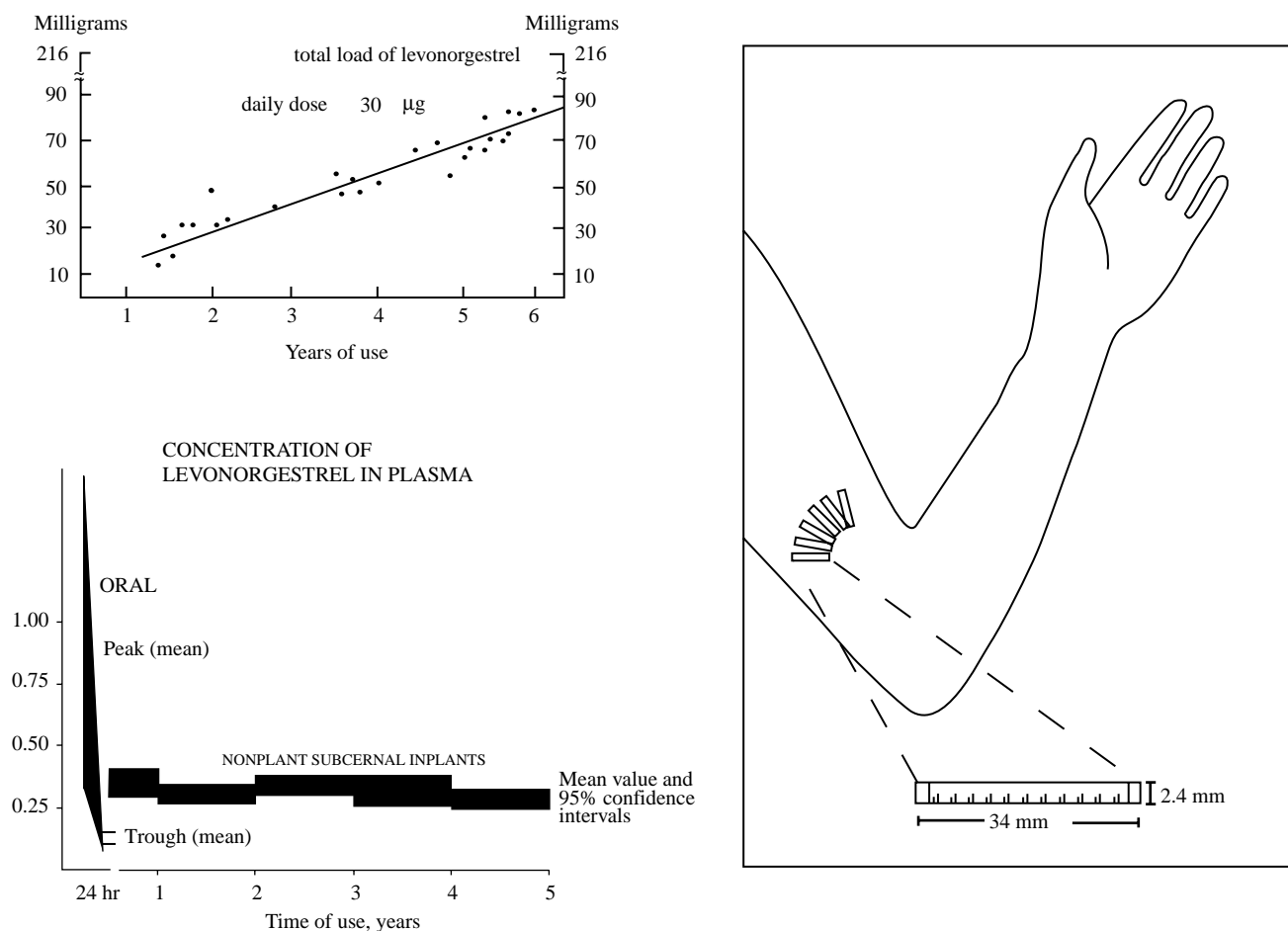


Fig. 7 Diagrammatic illustration of the subcutaneous implantation of Norplant® implants. The subcutaneous release profile of levonorgestrel in female volunteers for up to 6 years and the resultant plasma profile as compared to those obtained by oral administration. (Adapted from Refs. 15–18.)

The release profile of drug from this matrix diffusion-controlled CrDDS is not constant, because the rate of drug release is time dependent as defined by:

$$\frac{Q}{t^{1/2}} = (2AC_R D_p)^{1/2} \quad (2)$$

where A is the initial loading dose of drug dispersed in the polymer matrix; C_R is the drug solubility in the polymer, which is also the drug reservoir concentration in the polymer matrix; and D_p is the diffusivity of the drug molecules in the polymer matrix.

The release of drug molecules from this type of CrDDSs may be controlled at a preprogrammed rate by controlling the loading level and the polymer solubility of the drug and its diffusivity in the polymer matrix. Several CrDDSs of this type have been successfully marketed for therapeutical uses, and some representatives are outlined later for illustration.

Nitro-Dur® System

This controlled-release transdermal therapeutic system is fabricated by first heating an aqueous solution of water-soluble polymer, glycerol, and polyvinyl alcohol and then lowering the temperature of the mixture to form a polymer gel. Nitroglycerin/lactose triturate is dispersed in the gel, and the mixture is then solidified at room temperature to form a medicated polymer disc by a molding and slicing technique. After assembly onto a drug-impermeable metallic plastic laminate, a patch-type transdermal therapeutic system is produced with an adhesive rim surrounding the medicated disc (Fig. 9). It is designed for application onto an intact skin to provide a continuous transdermal infusion of nitroglycerin, at a daily dose of 0.5 mg/cm^2 , for the prevention of angina pectoris (2, 19).

The drug reservoir can also be formulated by directly dispersing the drug in an adhesive polymer, such as poly(isobutylene) or poly(acrylate) adhesive, and then

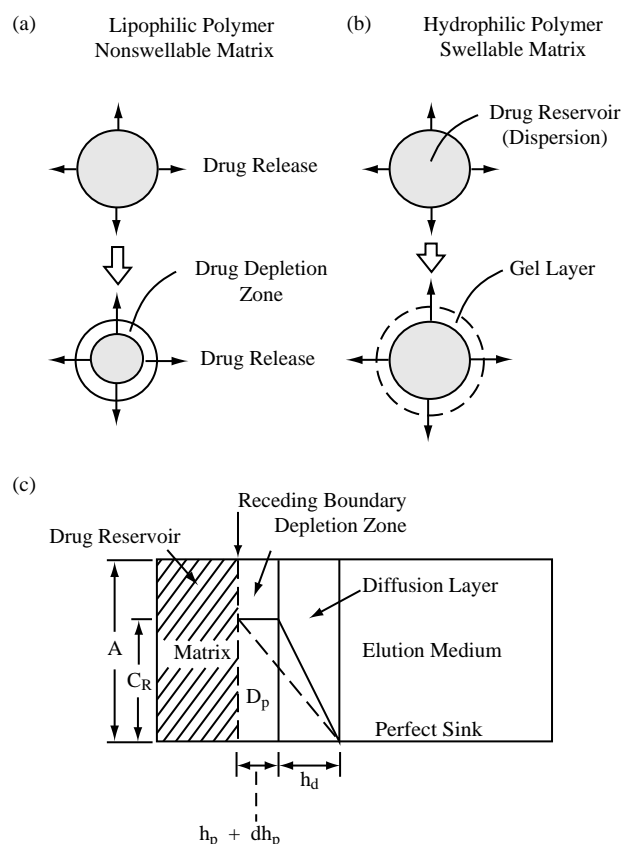


Fig. 8 Release of drug from the polymer matrix diffusion-controlled drug delivery systems with drug reservoir exists as a homogeneous dispersion in (a) lipophilic, nonswellable polymer matrix, with a growing thickness of drug depletion zone, or (b) a hydrophilic, swellable polymer matrix, with a growing thickness of drug-depleted gel layer. In (c), the drug concentration gradients across the time-dependent drug depletion zone, with a growing thickness ($h_p + dh_p$), and the hydrodynamic diffusion layer, with a controlled thickness (h_d), are shown in series.

spreading the medicated adhesive by solvent casting or hot melt, onto a flat sheet of drug-impermeable backing support to form a single- or multiple-layer drug reservoir. This type of transdermal CrDDS (TDD) is best illustrated by the development and marketing of an isosorbide dinitrate-releasing TDD system, named Frandol[®] tape, by Toaeiyo/Yamanouchi in Japan, and of a nitroglycerin-releasing TDD system, named Nitro-Dur[®] II system by Key in the United States, for once-a-day medication for angina pectoris. This second generation of TDD system (NitroDur II) has also received FDA approval for marketing. Nitro-Dur II compares favorably with Nitro-Dur (Fig. 10) and has gradually replaced the first-generation Nitro-Dur from the marketplace. The same technical basis has been also utilized in the development of the following: 1) Habitrol[®] and Nicotrol[®] systems, which

provide a controlled dose of nicotine transdermally over 24 h for smoking cessation (14); 2) Minitran[®] system, which administers a controlled dose of nitroglycerin transdermally over 24 h for the relief of anginal attacks (14); 3) Testoderm[®] system, which administers a controlled delivery of testosterone for transdermal permeation through a scrotal skin (14) for the replacement therapy of testosterone-deficient patients for 24 h; and 4) Climara[®] system, which provides a controlled delivery of 17 β -estradiol for transdermal permeation for once-weekly treatment of vasomotor systems (14) associated with menopause.

Compodose[®] Implant

This controlled-release subdermal implant is fabricated by dispersing micronized estradiol crystals in a viscous mixture of silicone elastomer and catalyst and then coating the estradiol-polymer dispersion around a rigid (drug-free) silicone rod by an extrusion technique to form a cylinder-shaped implant (Fig. 11). This implant is designed for subcutaneous implantation in the steer's ear flap for a duration of 200 or 400 days, during which a controlled quantity of estradiol is released daily for growth promotion (20).

To improve the Q versus $t^{1/2}$ drug release profiles [Eq. (2)], this polymer matrix diffusion-controlled CrDDS can be modified to have the drug-loading level varied, in an incremental manner, to form a gradient of drug reservoir along the diffusional path in the polymer matrix. A constant drug release profile is thus achieved, and the rate of drug release from this drug reservoir gradient-controlled drug delivery system is defined by:

$$\frac{dQ}{dt} = \frac{K_{a/r} D_a}{h_a(t)} C_p(h_a) \quad (3)$$

in which the time-dependent thickness [$h_a(t)$] of the diffusional path for drug molecules to diffuse through, which is increasing with time, is compensated by the proportional increase in the drug-loading level [$C_p(h_a)$], and a constant drug release profile is thus obtained. This type of CrDDS is best illustrated by the nitroglycerin-releasing Deponit[®] system (Fig. 12), first marketed by Pharma-Schwartz/Lohmann in Europe (21). Wyeth-Ayerst has received FDA approval for marketing this system in the United States.

Furthermore, it was recently demonstrated that the release of a drug, such as propranolol, from the multilaminate adhesive-based TDD system can be maintained at zero-order kinetics by controlling the particle size distribution of drug crystals in the various laminates of adhesive matrix (22).

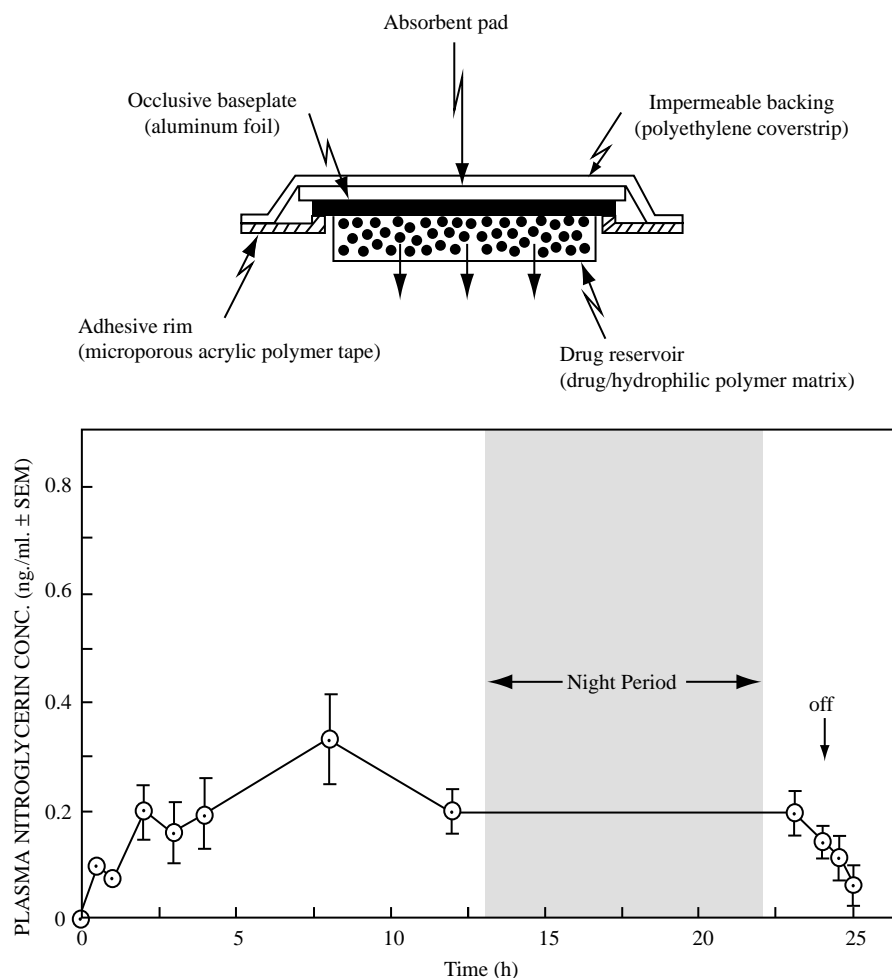


Fig. 9 Cross-sectional view of a unit of Nitro-Dur[®] system, showing various structural components, and the plasma nitroglycerin concentration profiles in six human volunteers, each receiving 1 unit of Nitro-Dur[®] system (20 cm², with a delivery rate of 10 mg/day) for 24 h. (From Refs. 55 and 56.)

Polymer (Membrane/Matrix) Hybrid-Type Drug Delivery Systems

This type of CrDDS is developed with the objective of combining the constant drug release kinetics of polymer membrane permeation-controlled drug delivery systems with the mechanical superiority of polymer matrix diffusion-controlled drug delivery systems. The release profile of drug from a sandwich-type drug delivery system (Fig. 13) is constant, and the instantaneous rate of drug release is defined by:

$$\frac{dQ}{dt} = \frac{AC_p D_p}{[D_p K_m (1/P_m + 1/P_d)]^2 + 4AC_p D_p t^{1/2}} \quad (4)$$

where A is the initial amount of drug solid impregnated in a unit volume of polymer matrix with solubility C_p and

diffusivity D_p ; K_m is the partition coefficient for the interfacial partitioning of drug molecules from polymer matrix toward polymer coating membrane; P_m is the permeability coefficient of the polymer coating membrane with thickness h_m ; and P_d is the permeability coefficient of the hydrodynamic diffusion layer with thickness h_d .

The hybrid system is exemplified by the development of clonidine-releasing and scopolamine-releasing transdermal therapeutic systems (Catapres-TTS[®] and Transderm-Scop[®]) (Fig. 14), in which a rate-controlling nonmedicated polymeric membrane is added to coat the surface of the drug-dispersing polymer matrix, and the release of drug molecules thus becomes controlled by membrane permeation instead of matrix diffusion. The same technology has been utilized in the development of levonorgestrel-releasing subdermal implants (Norplant[®] II).

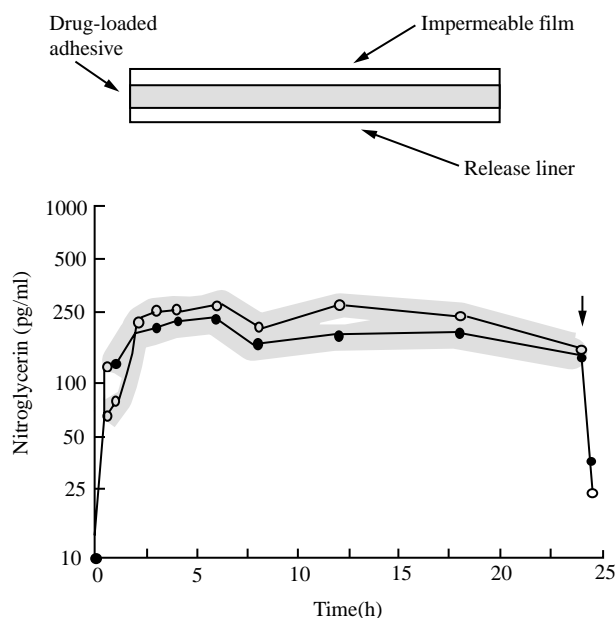


Fig. 10 Cross-sectional view of Nitro-Dur II, showing various structural components, and the comparative 24-h plasma nitroglycerin concentration profiles in 24 healthy male volunteers, each receiving randomly 1 unit of Nitro-Dur II (open circle) or Nitro-Dur (closed circle), 20 cm² each, with a delivery rate of 10 mg/day, over the chest for 24 h (the arrow indicates unit removal). (From Ref. 56.)

Microreservoir Partition-Controlled Drug Delivery Systems

In this type of CrDDS, the drug reservoir is a suspension of drug solid particles in an aqueous solution of a water-miscible polymer, like polyethylene glycols. This forms a homogeneous dispersion of many discrete, unleachable, microscopic drug reservoirs in a biocompatible polymer, like silicone elastomers (Fig. 15). The microdispersion is achieved by applying a high-energy dispersion technique (13, 23). Different shapes and sizes of drug-delivery devices can be fabricated from this microreservoir-type CrDDS by molding or extrusion techniques. Depending upon the physicochemical properties of drugs and the desired rate of drug release, the device can be further coated with a layer of biocompatible polymer to modify the mechanism and the rate of drug release.

The rate of drug release (dQ/dt) from this type of CrDDS is defined by:

$$\frac{dQ}{dt} = \frac{D_p D_a m K_p}{D_p h_d + D_a h_p m K_p} \left[n S_p - \frac{D_l S_l (1-n)}{h_l} \left(\frac{1}{K_l} + \frac{1}{K_m} \right) \right] \quad (5)$$

where $m = a/b$ and n is the ratio of drug concentration at the inner edge of the interfacial barrier over the drug

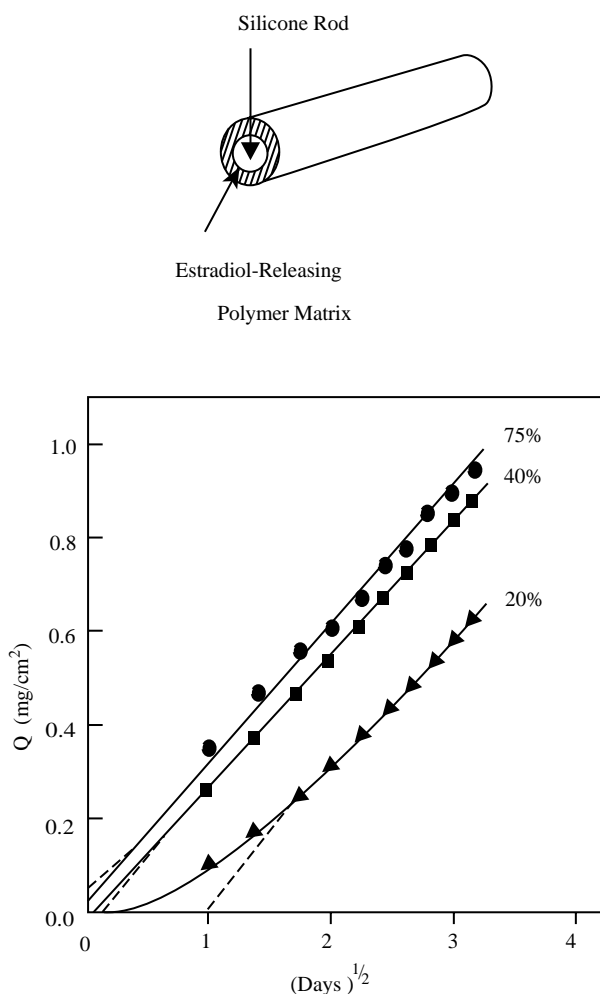


Fig. 11 Diagrammatic illustration of a unit of Compudose® subdermal implant and in vitro release profiles of estradiol from the implants immersed in aqueous solution containing various volume fractions of polyethylene glycol 400.

solubility in the polymer matrix (1, 6), in which a is the ratio of drug concentration in the bulk of elution solution over drug solubility in the same medium and b is the ratio of drug concentration at the outer edge of the polymer coating membrane over drug solubility in the same polymer. K_l , K_m , and K_p are, respectively, the partition coefficients for the interfacial partitioning of drug from the liquid compartments to the polymer matrix, from the polymer matrix to the polymer coating membrane, and from the polymer coating membrane to the elution solution, whereas D_l , D_p , and D_d are, respectively, the diffusivities of the drug in the liquid layer surrounding the drug particles, the polymer coating membrane enveloping the polymer matrix, and the hydrodynamic diffusion layer surrounding the polymer coating

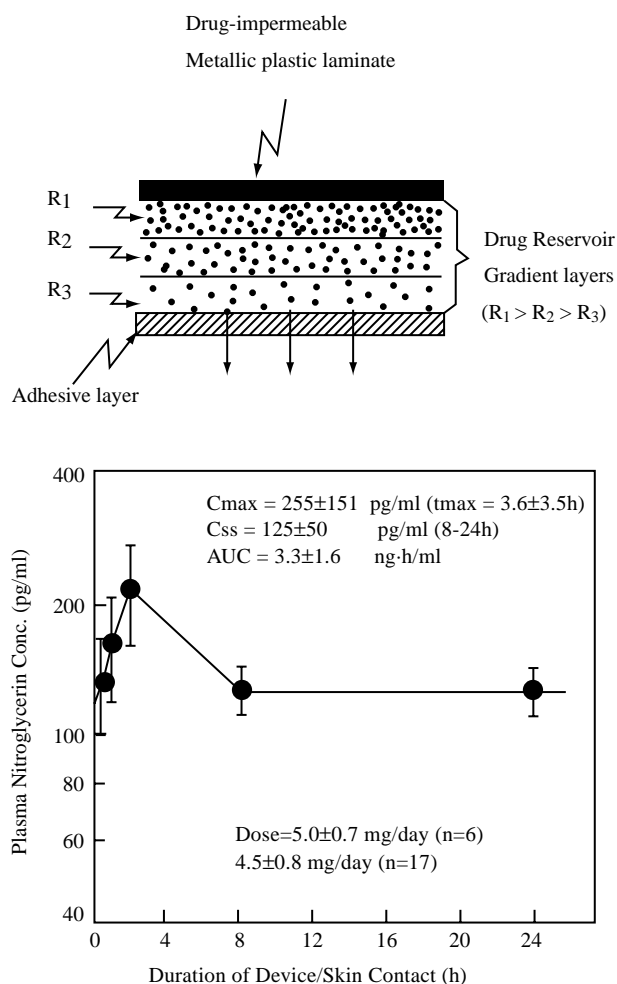


Fig. 12 Cross-sectional view of a unit of Deponit® system, showing various structural components, and the plasma nitroglycerin concentration profiles in six human volunteers, each receiving 1 unit of Deponit system (16 cm², with a delivery rate of 5 mg/day) for 24 h. (Plasma profiles are plotted from data from Ref. 21.)

membrane with respective thicknesses of h_i , h_p , and h_d (Fig.15); and S_l and S_p are the solubilities of the drug in the liquid compartments and in the polymer matrix, respectively.

The release of drug from the microreservoir-type CrDDS can follow either a dissolution- or a matrix diffusion-control process, depending upon the relative magnitude of S_l and S_p (24). Representatives of this type of CrDDS is outlined below.

Nitrodisc® system

In this transdermal CrDDS (Fig. 16), the drug reservoir is a suspension of nitroglycerin/lactose triturate in an

aqueous solution of 40% polyethylene glycol 400. It is dispersed homogeneously by a high-energy mixing technique, with isopropyl palmitate, a skin permeation enhancer, in a mixture of viscous silicone elastomer and catalyst (25). The resultant drug-polymer dispersion is then formed in situ into a solid medicated disc on a drug-impermeable metallic plastic laminate, with an adhesive rim, by an injection-molding technique and application of an instantaneous heating. It is engineered to provide a transdermal administration of nitroglycerin at a daily rate of 0.5 mg/cm² for once-a-day medication of angina pectoris (2, 26). AQ versus $t^{1/2}$ (matrix diffusion-controlled) release profile is obtained.

Syncro-Mate-C implant

This subdermal controlled-release implant is fabricated by dispersing the drug reservoir, which is a suspension of norgestomet in an aqueous solution of PEG 400, in a viscous mixture of silicone elastomers by a high-energy dispersion technique (24). After adding catalyst, the suspension is delivered into a silicone medical-grade tubing, which serves as the mold as well as the coating membrane, and then polymerized in situ. The polymerized drug-polymer composition is then cut into a cylinder-shaped implant with its ends staying open (Fig. 17). This tiny cylindrical implant is designed to be inserted into the subcutaneous tissue of the livestock's ear flap; norgestomet is released continuously into the subcutaneous tissue for up to 20 days for the control and synchronization of estrus and ovulation and up to 160 days for growth promotion. A constant Q versus t (dissolution-controlled) release profile has been achieved, as compared to the Q versus $t^{1/2}$ release profile (matrix diffusion-controlled drug release) for the Syncro-Mate-B implant and the Nitrodisc system discussed above.

Transdermal contraceptive device

The transdermal contraceptive device is based on a patentable micro-drug-reservoir technique (26) to achieve a dual-controlled release of levonorgestrel, a potent synthetic progestin, and estradiol, a natural estrogen, at constant and enhanced rates, continuously, for a period of 7 days (5). By applying 1 unit (10 or 20 cm²) of transdermal contraceptive device per week, beginning on day 5 of an individual's menstrual cycle, for 3 consecutive weeks (3 weeks on and 1 week off), steady-state serum levels of levonorgestrel have been obtained, and the secretion of gonadotropins and progesterone have been effectively suppressed.

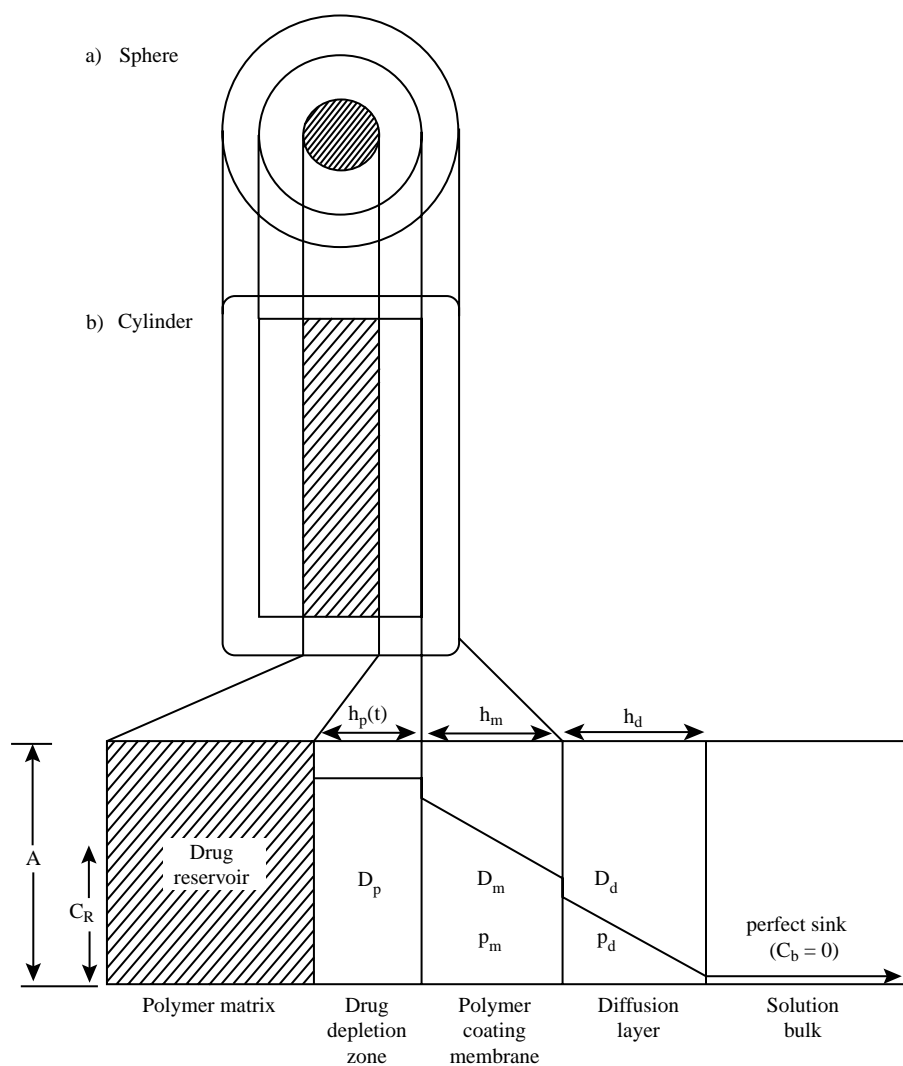


Fig. 13 The controlled release of drug molecules from a (membrane-matrix) hybrid-type drug delivery system in which solid drug is homogeneously dispersed in a polymer matrix, which is then encapsulated inside a polymeric membrane, where D , P , and h are the diffusivity, permeability, and thickness, respectively, and the subscripts p, m, and d denote the drug depletion zone in the polymer matrix, polymer coating membrane, and diffusion layer, respectively.

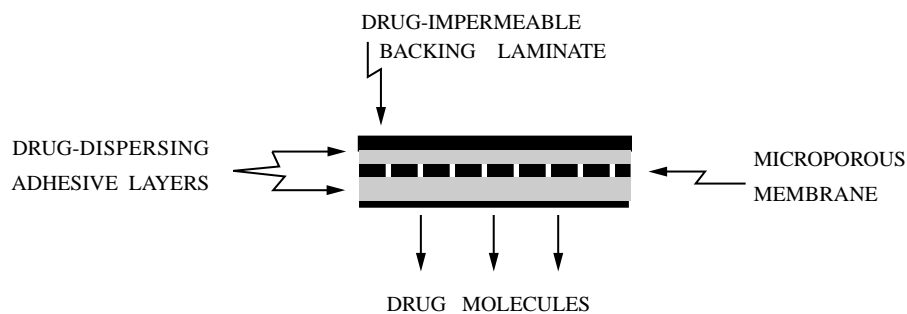


Fig. 14 Cross-section view of various structural components in the Transderm-Scop® and Catapres-TTS® systems.

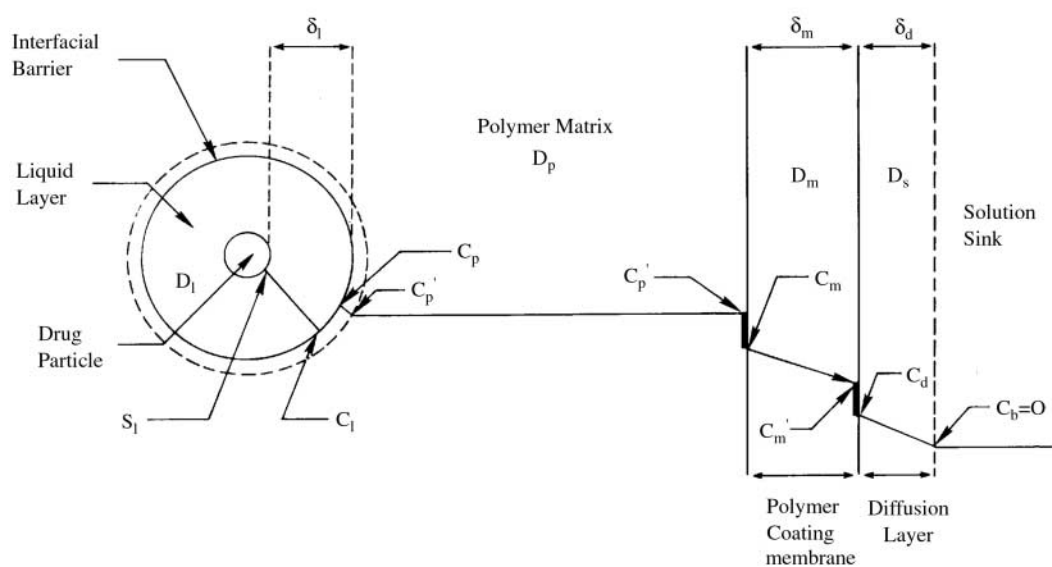
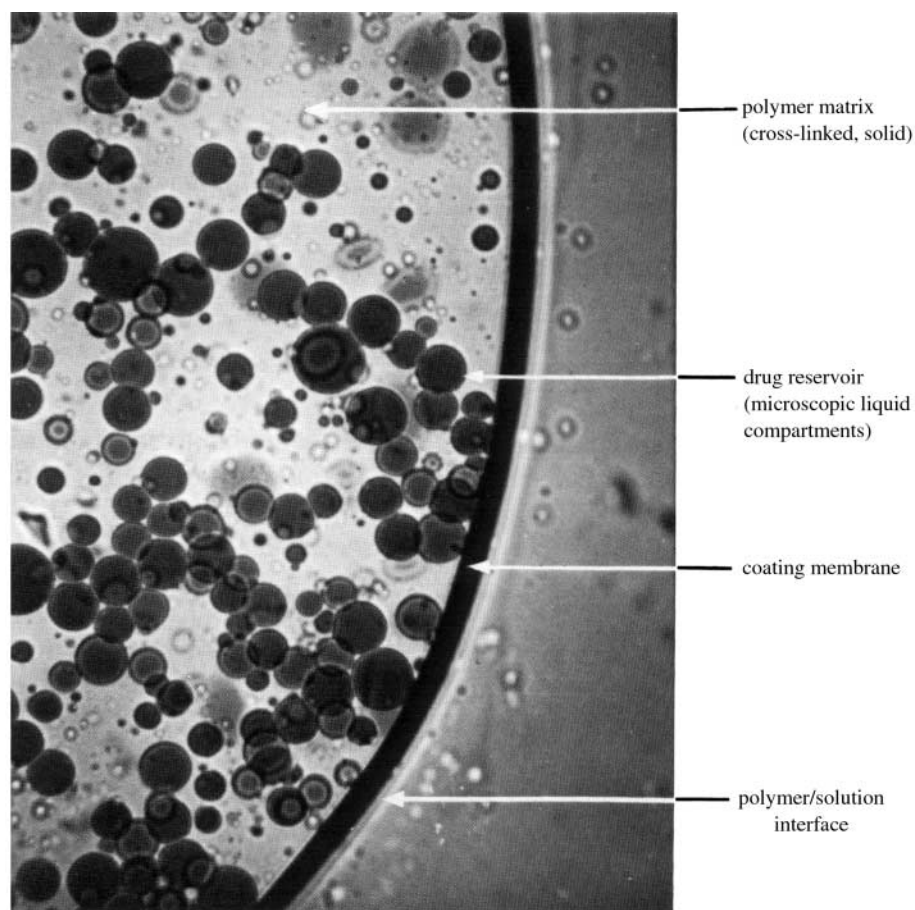


Fig. 15 Microscopic view of a microreservoir-type drug-delivery system, which shows the microscopic structure of various components, and the physical model developed for the mechanistic analysis of the controlled release of drug. (Adapted from Refs. 1 and 57.)

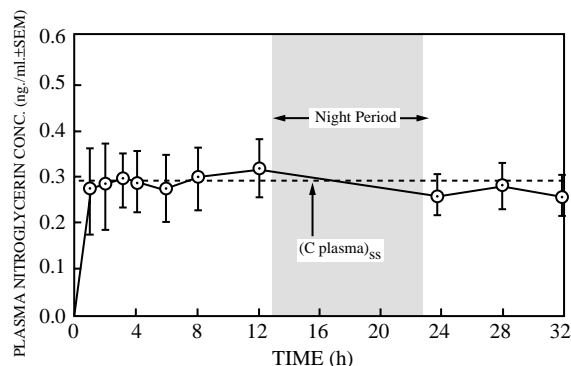
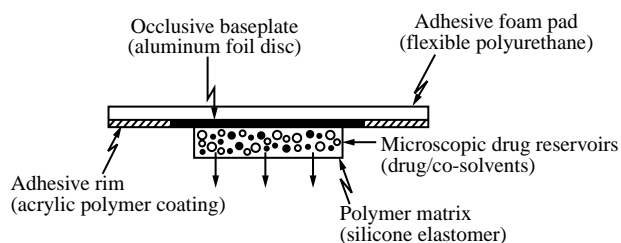


Fig. 16 Cross-sectional view of a unit of Nitrodisc® system, showing various structural components, and the plasma nitroglycerin concentration profiles in 12 human volunteers, each receiving 1 unit of Nitrodisc system (16 cm², with a delivery rate of 10 mg/day) for 32 h. (From Ref. 23.)

ACTIVATION-MODULATED DRUG DELIVERY SYSTEMS

In this group of CrDDSs, the release of drug molecules from the delivery systems is activated by some physical, chemical, or biochemical processes and/or facilitated by an energy supplied externally (Fig. 2). The rate of drug release is then controlled by regulating the process applied or energy input. Based on the nature of the process applied or the type of energy used, these activation-modulated CrDDSs can be classified into the following categories:

1. Physical means
 - a. Osmotic pressure-activated drug delivery systems
 - b. Hydrodynamic pressure-activated drug delivery systems
 - c. Vapor pressure-activated drug delivery systems
 - d. Mechanical force-activated drug delivery systems
 - e. Magnetics-activated drug delivery systems
 - f. Sonophoresis-activated drug delivery systems
 - g. Iontophoresis-activated drug delivery systems
 - h. Hydration-activated drug delivery systems
2. Chemical means
 - a. pH-activated drug delivery systems
 - b. pH-activated drug delivery systems

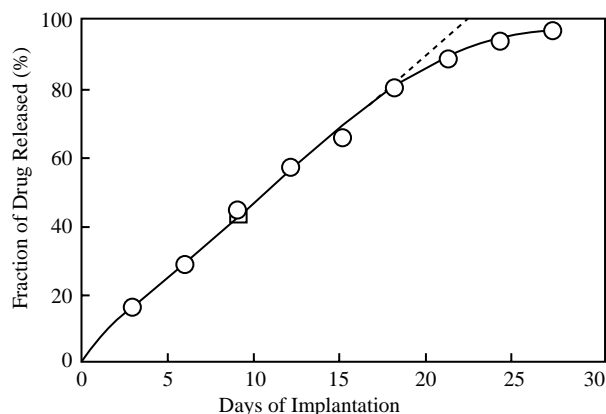
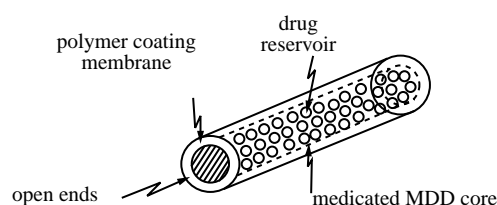


Fig. 17 Syncro-Mate-C implant, a subdermal implant fabricated from the microreservoir dissolution-controlled drug-delivery system, and subcutaneous controlled release of norgestomet, a potent synthetic progestin, at constant rate for 20 days. The open ends on the implant do not affect the zero-order in vivo drug release profile. (Adapted from Ref. 57.)

- c. Ion-activated drug delivery systems
- d. Hydrolysis-activated drug delivery systems

3. Biochemical means

- a. Enzyme-activated drug delivery systems
- b. Biochemical-activated drug delivery systems

Several CrDDSs have been successfully developed and applied clinically to the controlled delivery of pharmaceuticals and biopharmaceuticals. These are outlined and discussed below.

Osmotic Pressure-Activated Drug Delivery Systems

In this type of CrDDSs, the drug reservoir, which can be either a solution or a solid formulation, is contained within a semipermeable housing with a controlled water permeability. The drug in solution is released through a special laser-drilled delivery orifice at a constant rate under a controlled gradient of osmotic pressure.

For a solution-type osmotic pressure-activated CrDDS, the intrinsic rate of drug delivery (Q/t) is defined by:

$$\frac{Q}{t} = \frac{P_w A_m}{h_m} (\pi_s - \pi_e) \quad (6)$$

For a solid-type osmotic pressure-activated CrDDS, the intrinsic rate of drug delivery should also be a constant and is defined by:

$$\frac{Q}{t} = \frac{P_w A_m}{h_m} (\pi_s - \pi_e) S_d \quad (7)$$

where P_w , A_m , and h_m are, respectively, the water permeability, the effective surface area, and the thickness of the semipermeable housing; $(\pi_s - \pi_e)$ is the differential

osmotic pressure between the drug-delivery system with an osmotic pressure of π_s and the environment with an osmotic pressure of π_e ; and S_d is the aqueous solubility of the drug component in the solid reservoir.

The release of drug molecules from this type of CrDDS is activated by osmotic pressure and controlled at a rate determined by the water permeability and the effective surface area of the semipermeable housing as well as the osmotic pressure gradient. Several CrDDSs of this type have been successfully marketed for therapeutical uses and some representatives are outlined later.

Alzet osmotic pump

In this implantable or insertable CrDDS, the drug reservoir, which is normally a solution formulation, is contained within a collapsible, impermeable polyester bag whose external surface is coated with a layer of osmotically active salt, for example, sodium chloride. This reservoir compartment is then totally sealed inside a rigid housing walled with a semipermeable membrane (Fig. 19). At an implantation site, the water content in the tissue fluid will penetrate through the semipermeable membrane at a controlled rate and dissolve the osmotically active salt. This creates an osmotic pressure in the narrow spacing between the flexible reservoir wall and the rigid semipermeable housing. Under the osmotic pressure created [Eq. (6)], the reservoir compartment is thus reduced in volume and the drug solution is forced to release through the flow moderator at a controlled rate (27, 28). By varying the drug concentration in the solution, different doses of drug can be delivered at a constant rate for a period of 1–4 weeks.

In addition to its application in the subcutaneous controlled administration of drugs for pharmacological studies, this technology has recently been extended to the controlled administration of drugs in the rectum by zero-order kinetics. The hepatic first-pass metabolism of drugs is thus bypassed (29).

Acutrim[®] tablet

In this oral CrDDS, the drug reservoir, which is a solid tablet of water-soluble and osmotically-active phenylpropanolamine (PPA) HCl, is enclosed within a semipermeable membrane of cellulose triacetate (2, 30). The surface of the semipermeable membrane is further coated with a thin layer of immediately releasable PPA dose (Fig. 20). In the alimentary tract, the gastrointestinal fluid will dissolve away the immediate release layer of PPA to provide an initial dose of PPA and then penetrate through the semipermeable membrane to dissolve the sustained-release dose of PPA. Under the osmotic pressure created [Eq. (7)], the PPA solution is released continuously at a

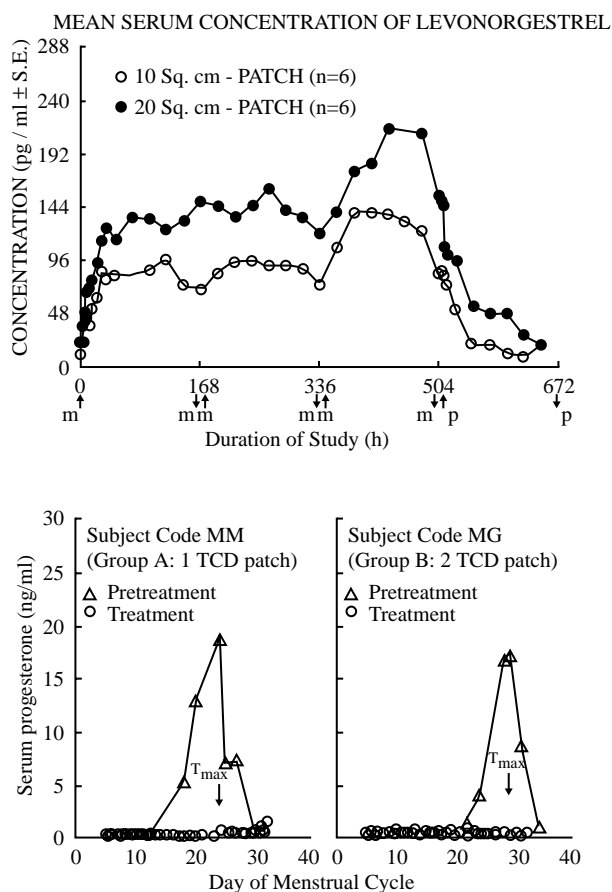


Fig. 18 (Upper panel) The 4-week serum levonorgestrel profiles in 12 human volunteers, each receiving 1 or 2 units of a transdermal contraceptive system (10 cm², with daily dosage of 28.3 μg/day) once a week, consecutively for 3 weeks, and the same size of placebo on week 4. (Lower panel) Comparative serum concentration profiles of progesterone during the pretreatment and treatment cycles in two subjects, each as the representative for group A (receiving 10 cm²) and group B (receiving 20 cm²), respectively. The suppression of progesterone peak during the treatment cycle is an indication of effective fertility control.

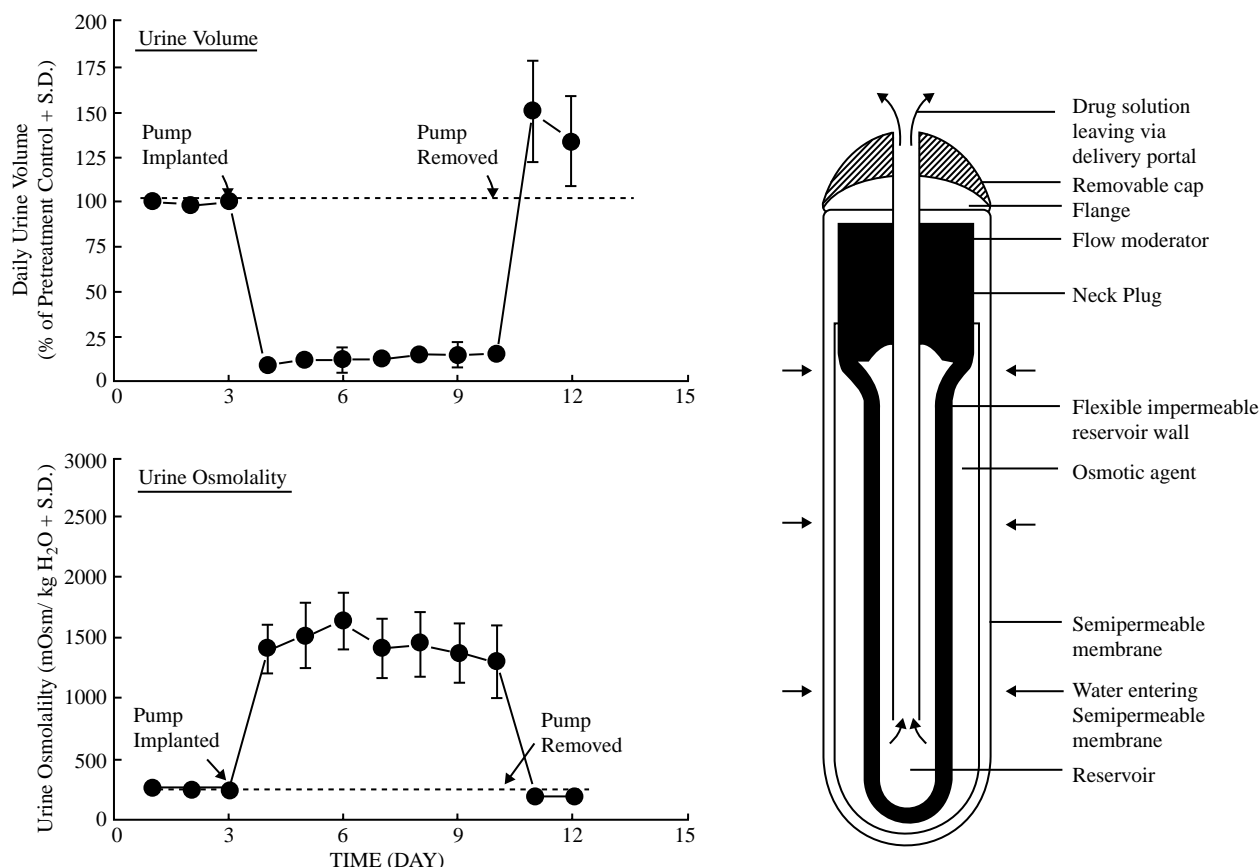


Fig. 19 (Left panel) Cross-sectional view of the Alzet[®] osmotic pump, an osmotic pressure-activated drug-delivery system. (Right panel) The effect of 7 days of subcutaneous delivery of antidiuretic hormone (vasopressin) on the daily volume of urinary excretion and urine osmolality in the Brattleboro rats with diabetes insipidus.

controlled rate, through an orifice pre-drilled by a laser beam (2, 30, 31). It is designed to provide a controlled delivery of PPA over a duration of 16 h for appetite suppression in a weight-control program (31). The same delivery system has also been utilized for the oral controlled delivery of indomethacin. An extension of this technology is the development of a push-pull type osmotic pressure-activated CrDDS for the oral controlled delivery of nifedipine and metoprolol (27). It has been further extended to the delayed-onset and controlled oral delivery of verapamil (14) to produce a maximum plasma concentration in the morning hours.

Hydrodynamic Pressure-Activated Drug Delivery Systems

In addition to the osmotic pressure systems discussed above, hydrodynamic pressure has also been explored as the potential source of energy to modulate the delivery of therapeutic agents (2).

A hydrodynamic pressure-activated drug-delivery system can be fabricated by placing a liquid drug formulation inside a collapsible, impermeable container to form a drug reservoir compartment. This is then contained inside a rigid, shape-retaining housing. A laminate of an absorbent layer and a swellable, hydrophilic polymer layer is sandwiched between the drug reservoir compartment and the housing. In the gastrointestinal tract, the laminate will imbibe the gastrointestinal fluid through the annular openings at the lower end of the housing and become swollen. This generates a hydrodynamic pressure in the system. The hydrodynamic pressure, thus created, forces the drug reservoir compartment to reduce in volume and causes the liquid drug formulation to release through the delivery orifice (32). The drug release rate is defined by

$$\frac{Q}{t} = \frac{P_f A_m}{h_m} (\theta_s - \theta_e) \quad (8)$$

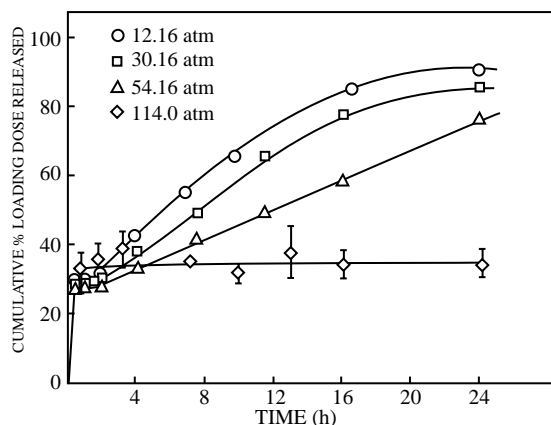
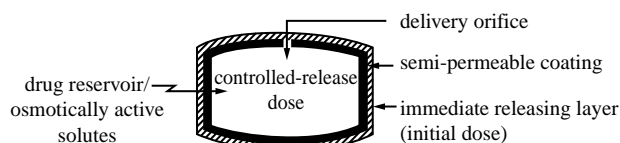


Fig. 20 Cross-sectional view of a unit of Acutrim[®] tablet, a solid-type osmotic pressure-activated drug delivery system, and the effect of increased osmotic pressure in the dissolution medium on the release profiles of phenylpropanolamine HCl from the Acutrim tablet at intestinal condition. (Adapted from Refs. 31 and 58.)

where P_f , A_m , and h_m are the fluid permeability, the effective surface area, and the thickness of the wall with annular openings, respectively; and $\theta_s - \theta_e$ is the difference in hydrodynamic pressure between the drug delivery system (θ_s) and the environment (θ_e).

The release of drug molecules from this type of CrDDS is activated by hydrodynamic pressure and controlled at a rate determined by the fluid permeability and effective surface area of the wall with annular openings as well as by the hydrodynamic pressure gradient.

Vapor Pressure-Activated Drug Delivery Systems

In this type of CrDDS, the drug reservoir, which is a solution formulation, is contained inside the infusion compartment. It is physically separated from the pumping compartment by a freely movable partition (Fig. 21). The pumping compartment contains a vaporizable fluid, such as fluorocarbon, which vaporizes at body temperature and creates a vapor pressure. Under the vapor pressure created, the partition moves upward and forces the drug solution in the infusion compartment to be delivered, through a series of flow regulator and delivery cannula, into the blood circulation at a constant flow rate (1, 6, 33). The process is defined by:

$$\frac{Q}{t} = \frac{d^4 \delta P}{40.74 \mu l} \quad (9)$$

where d and l are, respectively, the inner diameter and the length of the delivery cannula; Δp is the pressure difference between the vapor pressure in the pumping compartment and the pressure at the implantation site; and μ is the viscosity of the drug formulation.

The delivery of drug from this type of CrDDS is activated by vapor pressure and controlled at a rate determined by the differential vapor pressure, the formulation viscosity, and the size of the delivery cannula.

A typical example is the development of Infusaid[®], an implantable infusion pump by Metal Bellows, for the constant infusion of heparin in anticoagulation treatment (34), of insulin in the normoglycemic control of diabetics (33), and of morphine for patients suffering from the intensive pain of a terminal cancer (35).

Mechanical Force-Activated Drug Delivery Systems

In this type of CrDDS, the drug reservoir is a solution formulation in a container equipped with a mechanically activated pumping system. A metered dose of drug formulation can be reproducibly delivered into a body cavity, such as the nose, through the spray head upon manual activation of the drug-delivery pumping system. The volume of solution delivered is fixed and is independent of the force and duration of activation.

A typical example of this type of drug-delivery system is the development of a metered-dose nebulizer for the intranasal administration of a precision dose of luteinizing hormone-releasing hormone (LHRH) and its synthetic analogs, such as buserelin. Through nasal absorption, the hepatic first-pass elimination of these peptide drugs is thus avoided (24).

Magnetic-Activated Drug Delivery Systems

Macromolecular drugs, such as peptides, have been known to release only at a relatively low rate from a polymer-controlled drug-delivery system. This low rate of release can be improved by incorporating an electromagnetism-triggering vibration mechanism into the polymeric delivery device. With a hemispheric-shaped design, a zero-order drug-release profile is achieved (36). By combining these two approaches, a subdermally implantable, magnetic-activated hemispheric drug-delivery device is developed. It is fabricated by first positioning a tiny doughnut-shaped magnet at the center of a drug-dispersing biocompatible polymer matrix and

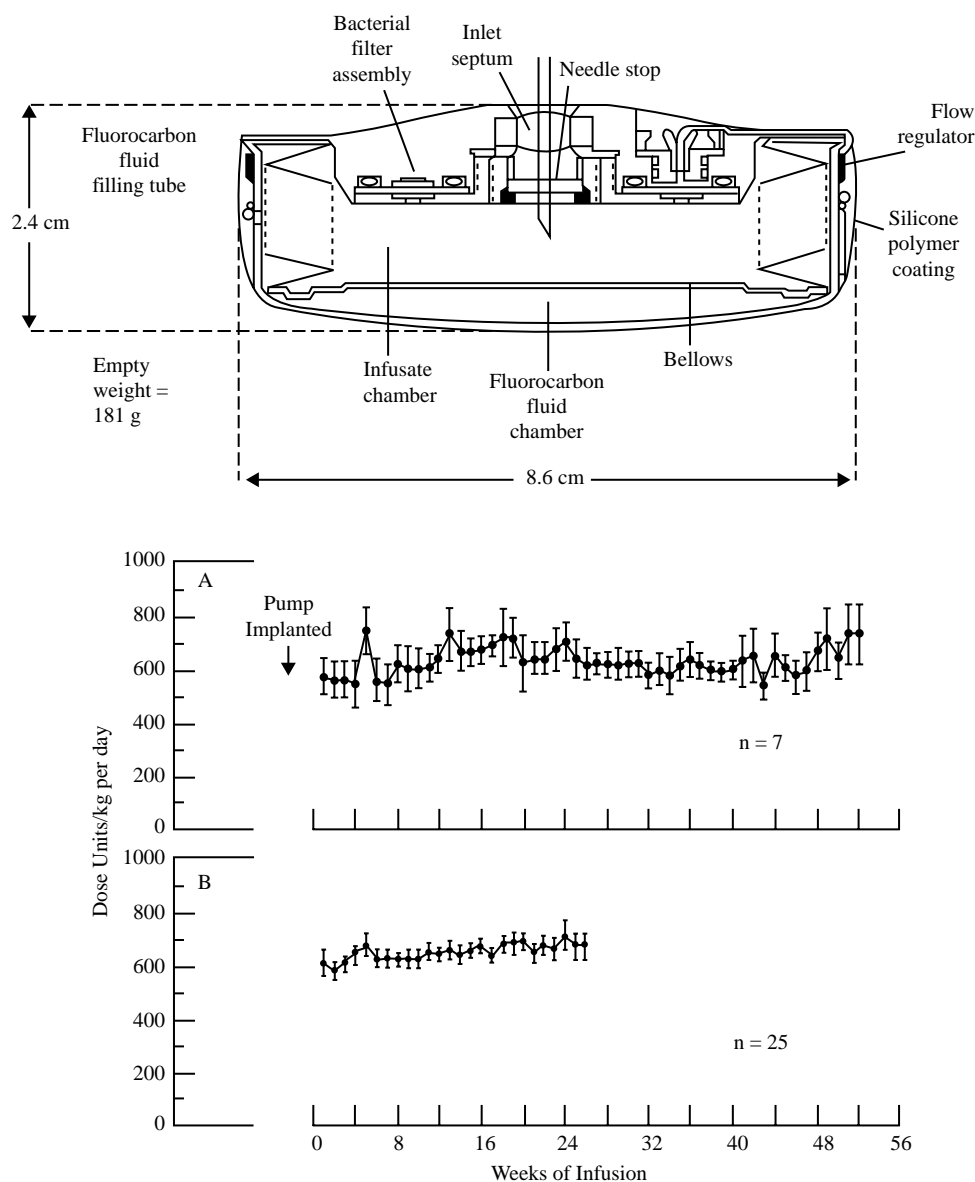


Fig. 21 Cross-sectional view of a unit of Infusaid® system, a vapor pressure-activated drug-delivery system, and daily heparin dose (mean \pm S.E.) delivered to 25 dogs for 6 months and to 7 dogs for 12 months. (Adapted from Ref. 34.)

then coating the external surface of the medicated polymer matrix, with the exception of one cavity at the center of the flat surface, with a pure polymer, for instance, ethylene–vinyl acetate copolymer or silicone elastomers. This uncoated cavity is designed for allowing a peptide drug to release.

The hemispheric magnetic delivery device produced can release macromolecular drugs, like bovine serum albumin, at a low basal rate, by diffusion process, and under a nontriggering condition, or it can release the same

drug at a much higher rate, when the magnet is activated, to vibrate by an external electromagnetic field.

Sonophoresis-Activated Drug Delivery Systems

This type of activation-controlled drug delivery system utilizes ultrasonic energy to activate (or trigger) the delivery of drugs from a polymeric drug delivery device. The system can be fabricated from either a nondegradable polymer, such as ethylene–vinyl acetate copolymer, or a

bioerodible polymer, such as poly[bis(*p*-carboxyphenoxy)alkane anhydride] (37). The potential application of sonophoresis (or phonophoresis) to regulate the delivery of drugs was recently reviewed (38).

Iontophoresis-Activated Drug Delivery Systems

This type of CrDDS use electrical current to activate and to modulate the diffusion of a charged drug molecule across a biological membrane, such as the skin, in a manner similar to passive diffusion under a concentration gradient but at a much facilitated rate. The iontophoresis-facilitated skin permeation rate of a charged molecule *i* consists of three components and is expressed by

$$J_i^{\text{isp}} = J^p + J^e + J^c = \left(K_s D_s \frac{dC}{h_s} \right) + \left(\frac{Z_i D_i F_i}{RT} C_i \frac{dE}{h_s} \right) + (k C_s I_d) \quad (10)$$

where J^p , J^e , and J^c represent, respectively, the flux for the skin permeation by passive diffusion, for the electrical current-driven permeation, and for the convective flow-driven skin permeation; K_s is the partition coefficient for interfacial partitioning from the donor solution to the stratum comeum; D_s and D_i are, respectively, the diffusivity across the skin and the diffusivity of ionic species *i* in the skin; C_i and C_s are, respectively, the donor concentration of ionic species *i* and the concentration in the skin tissue; dE/h_s is the electrical potential gradient across the skin; dC/h_s is the concentration gradient across the skin; Z_i is the electrical valence of ionic species *i*; I_d is the current density applied; F , k , and R are, respectively, the faraday, proportionality, and gas constant; and T is the absolute temperature.

A typical example of this type of activation-controlled CrDDS is the development of an iontophoretic drug delivery system, named Phoresor by Motion Control, to facilitate the percutaneous penetration of antiinflammatory drugs, such as dexamethasone sodium phosphate (39–41), to surface tissues.

Further development of the iontophoresis-activated drug delivery technique has yielded a new design of iontophoretic drug delivery system—the transdermal periodic iontotherapeutic system (TPIS). This new system, which is capable of delivering a physiologically-acceptable pulsed direct current, in a periodic manner, with a special combination of waveform, intensity, frequency, and on/off ratio, for a specific duration, has significantly improved the efficiency of transdermal delivery of peptide and protein drugs (4). A typical example is the iontophoretic transdermal delivery of insulin, a protein drug, in the control of hyperglycemia in diabetic animals.

Hydration-Activated Drug Delivery Systems

In this type of CrDDS, the drug reservoir is homogeneously dispersed in a swellable polymer matrix fabricated from a hydrophilic polymer. The release of drug is activated and modulated by hydration-induced swelling of the polymer matrix. Representatives of this type of CrDDS are outlined below.

Syncro-Male-B implant

This subcutaneous CrDDS is fabricated by dissolving norgestomet, a potent progestin for estrus synchronization, in an alcoholic solution of linear ethylene glycomethacrylate polymer (Hydron S). The drug-polymer mixture is then cross-linked by adding ethylene dimethacrylate, in the presence of an oxidizing catalyst, to form a cylinder-shaped subdermally implantable implant (1, 6). This tiny subdermal implant can be activated by tissue fluid to swell and can be engineered to deliver norgestomet, at a rate of $504 \mu\text{g}/\text{cm}^2/\text{day}^{1/2}$, in the subcutaneous tissue for up to 16 days for the control and synchronization of estrus in livestock (13).

Valrelease[®] tablet

This oral CrDDS is prepared by granulating Valium, an antidepressant drug, with hydrocolloids (20–75 wt%) and pharmaceutical excipients. The granules are then compressed to form an oral tablet. After oral intake, the hydrocolloids absorb the gastric fluid and are activated to form a colloid gel matrix surrounding the tablet surface (Fig. 22). The release of Valium molecules is then controlled by diffusion through the gel barrier, while the tablet remains buoyant in the stomach, due to a density difference between the gastric fluid ($d > 1$) and the gelling tablet ($d < 1$) (2, 3).

pH-Activated Drug Delivery Systems

For a drug labile to gastric fluid or irritating to gastric mucosa, this type of CrDDS has been developed to target the delivery of the drug only in the intestinal tract, not in the stomach (2). It is fabricated by coating a core tablet of the gastric fluid-sensitive drug with a combination of intestinal fluid-insoluble polymer, like ethyl cellulose, and intestinal fluid-soluble polymer, like hydroxymethyl cellulose phthalate (Fig. 23).

In the stomach, the coating membrane resists the degrading action of gastric fluid ($\text{pH} < 3$), and the drug molecules are thus protected from the acidic degradation. After gastric emptying, the CrDDS travels to the small intestine, and the intestinal fluid-soluble component in the coating membrane is dissolved away by the intestinal fluid

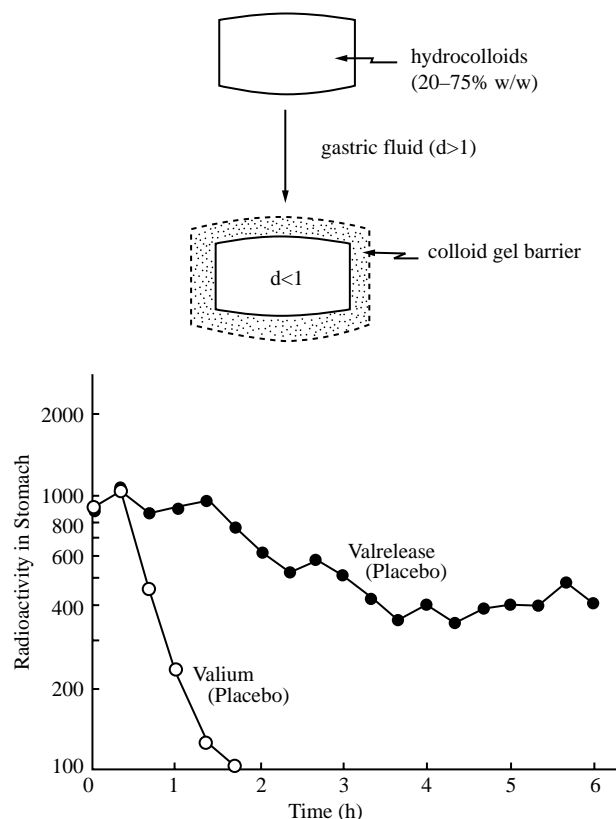


Fig. 22 Upper panel: Schematic illustration of Valrelease® tablet, a swelling-activated drug-delivery system, and the hydration-induced formation of colloid gel barrier. Lower panel: Comparison in the gastric residence profile between Valrelease with the conventional Valium capsule. (From Ref. 58.)

(pH > 7.5). This produces a microporous membrane of intestinal fluid-insoluble polymer to control the release of drug from the core tablet. The drug is thus delivered in a controlled manner in the intestine by a combination of drug dissolution in the core and diffusion through the pore channels. By adjusting the ratio of the intestinal fluid-soluble polymer to the intestinal fluid-insoluble polymer in the membrane, the rate of drug delivery can be regulated. Representative application of this type of CrDDS is in the oral controlled delivery of potassium chloride, which is highly irritating to gastric epithelium.

Ion-Activated Drug Delivery Systems

For controlling the delivery of an ionic or an ionizable drug, this type of CrDDS has been developed (2). Because the gastrointestinal fluid has regularly maintained a relatively constant level of ions, the delivery of drug by this type of CrDDS can be modulated, theoretically, at a constant rate.

Such a CrDDS is prepared by first complexing an ionizable drug with an ion-exchange resin, such as complexing a cationic drug with a resin containing SO_3^- group or an anionic drug with a resin containing $\text{N}(\text{CH}_3)_3^+$ group. The granules of the drug-resin complex are further treated with an impregnating agent, like polyethylene glycol 4000, for reducing the rate of swelling upon contact with an aqueous medium. They are then coated by an air-suspension coating technique with a water-insoluble but water-permeable polymeric membrane, such as ethylcellulose. This membrane serves as a rate-controlling barrier to modulate the release of drug from the CrDDS. In the GI tract, hydronium and chloride ions diffuse into the CrDDS and interact with the drug-resin complex to trigger the dissociation and release of ionic drug (Fig. 24).

This type of CrDDS is exemplified by the development of Pennkinetic® system (by Pennwalt Pharmaceuticals), which permits the formulation of oral liquid-type dosage forms with sustained release of a combination of

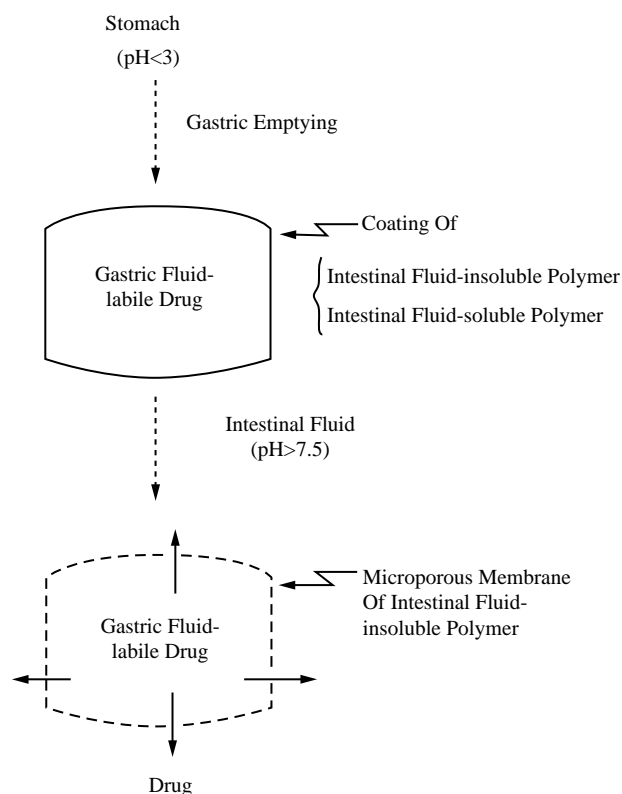


Fig. 23 Schematic illustration of a pH-activated drug-delivery system and the pH-dependent formation of microporous membrane in the intestinal tract.

hydrocodone and chlorpheniramine (Tussionex®) (14, 42–44).

Hydrolysis-Activated Drug Delivery Systems

This type of CrDDS depends on the hydrolysis process to activate the release of drug molecules. In this system, the drug reservoir is either encapsulated in microcapsules or homogeneously dispersed in microspheres or nanoparticles. It can also be fabricated as an implantable device. All these systems are prepared from a bioerodible or biodegradable polymer, such as polylactide, poly(lactide–glycolide) copolymer, poly(orthoester), or poly(anhydride). The release of a drug from the polymer matrix is activated by the hydrolysis-induced degradation of polymer chains, and the rate of drug delivery is controlled by polymer degradation rate (45). A typical example is the development of Lupron Depot®, an injectable microspheres for the subcutaneous controlled delivery of luprolide, a potent biosynthetic analog of gonadotropin-releasing hormone (GnRH) for the treatment of gonadotropin-dependent cancers, such as prostate carcinoma in men and endometriosis in the females, for up to 4 months. Another

example is the development of Zoladex® system, an implantable cylinder for the subcutaneous controlled delivery of goserelin, also a potent biosynthetic analog of GnRH for the treatment of patients with prostate cancer (Fig. 25) for up to 3 months (14).

Enzyme-Activated Drug Delivery Systems

In this type of CrDDS, the drug reservoir is either physically entrapped in microspheres or chemically bound to polymer chains fabricated from biopolymers, such as albumins or polypeptides. The release of drugs is made possible by the enzymatic hydrolysis of biopolymers by a specific enzyme in the target tissue (46–48). A typical example is the development of albumin microspheres, which release 5-fluorouracil, in a controlled manner, by protease-activated biodegradation.

FEEDBACK-REGULATED DRUG DELIVERY SYSTEMS

In this group of CrDDSs, the release of drug molecules is activated by a triggering agent, such as a biochemical

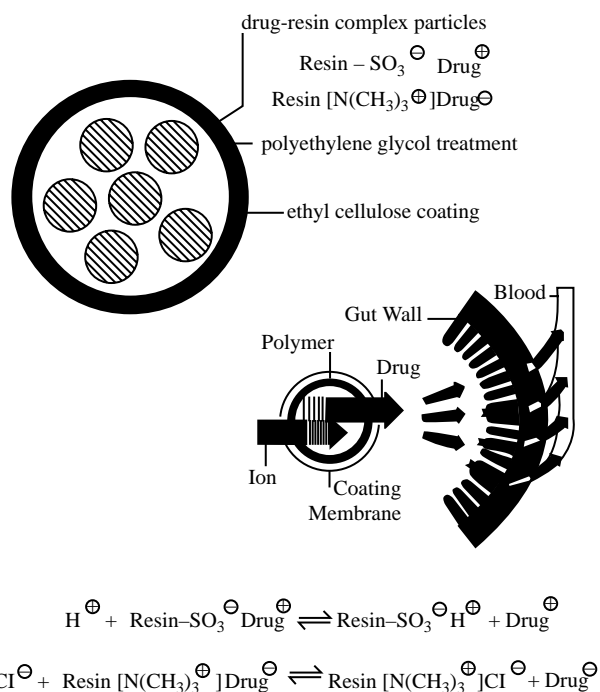


Fig. 24 Cross-sectional view of an ion-activated drug-delivery system, showing various structural components, and diagrammatic illustration of ion-activated drug release. (Adapted from Ref. 58.)

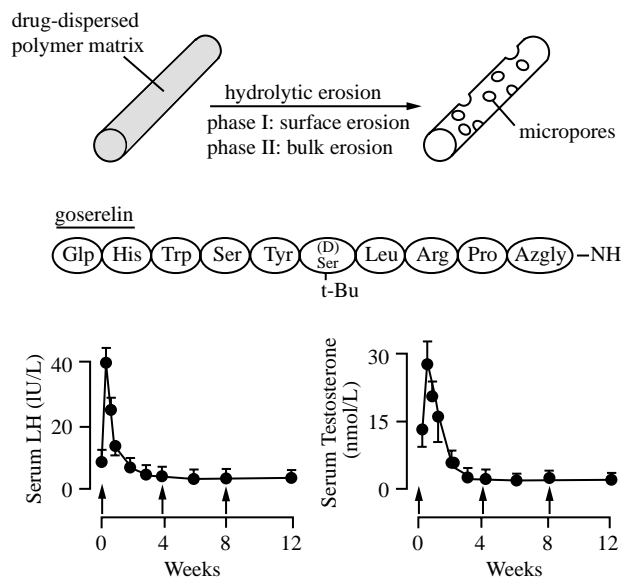


Fig. 25 Amino acid sequence of goserelin, a biosynthetic analog of gonadotropin-releasing hormone, and the effect of subcutaneous controlled release of goserelin from the biodegradable poly(lactide-glycolide) implant on the serum levels of luteinizing hormone and testosterone.

substance, in the body via some feedback mechanisms (Fig. 2). The rate of drug release is regulated by the concentration of a triggering agent detected by a sensor built into the CrDDS.

Bioerosion-Regulated Drug Delivery Systems

The feedback-regulated drug delivery concept has been applied to the development of a bioerosion-regulated CrDDS by Heller and Trescony (49). This CrDDS consists of a drug-dispersed bioerodible matrix fabricated from poly(vinyl methyl ether) half-ester, which was coated with a layer of immobilized urease (Fig. 26). In a solution with near neutral pH, the polymer only erodes very slowly. In the presence of urea, urease at the surface of the drug delivery system metabolizes urea to form ammonia. This causes the pH to increase and activates a rapid degradation of polymer matrix as well as the release of drug molecules.

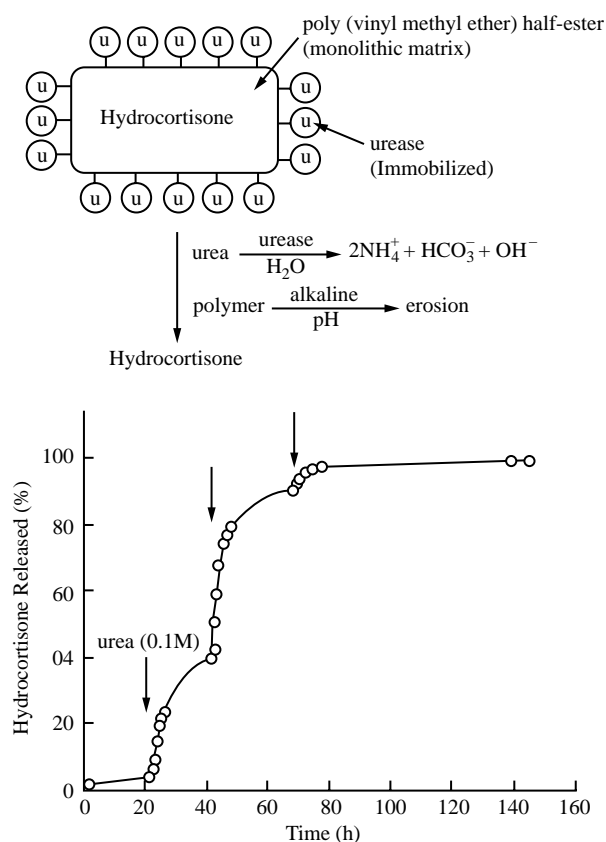


Fig. 26 Cross-sectional view of a bioerosion-regulated hydrocortisone delivery system, a feedback-regulated drug delivery system, showing the drug-dispersed monolithic bioerodible polymer matrix with surface-immobilized ureases. The mechanism of release and time course for the urea-activated release of hydrocortisone are also shown. (From Ref. 49.)

Bioresponsive Drug Delivery Systems

The feedback-regulated drug delivery concept has also been applied to the development of a bioresponsive CrDDS by Horbett et al. (50). In this CrDDS, the drug reservoir is contained in a device enclosed by a bioresponsive polymeric membrane whose permeability to drug molecules is controlled by the concentration of a biochemical agent in the tissue where the CrDDS is located.

A typical example of this bioresponsive CrDDS is the development of a glucose-triggered insulin delivery system, in which the insulin reservoir is encapsulated within a hydrogel membrane containing pendant NR_2 groups (Fig. 27). In an alkaline solution, the NR_2 groups exist at neutral state and the membrane is unswollen and thus impermeable to insulin. As glucose penetrates into the membrane, it is oxidized enzymatically by the glucose oxidase entrapped in the membrane to form gluconic acid. This process triggers the protonation of NR_2 groups to

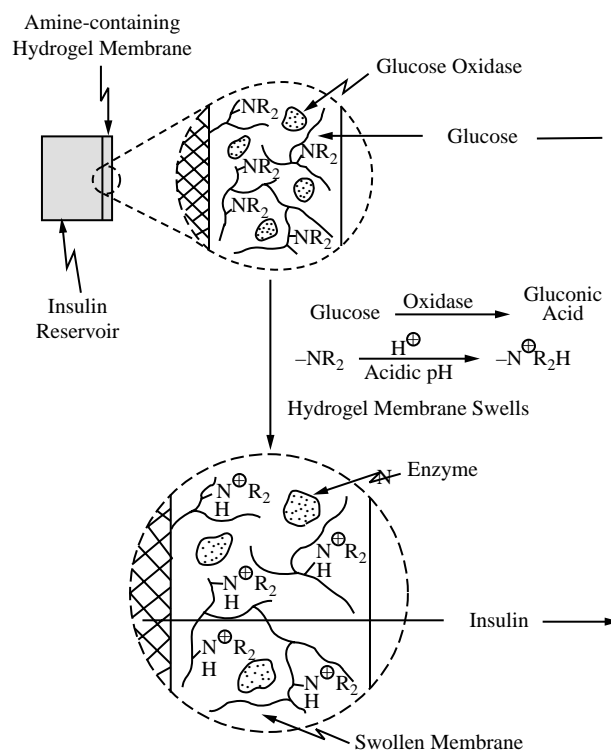


Fig. 27 Cross-sectional view of a bioresponsive insulin delivery system, a feedback-regulated drug delivery system, showing the glucose oxidase-entrapped hydrogel membrane constructed from amine-containing hydrophilic polymer. The mechanism of insulin release, in response to the influx of glucose, is also illustrated. (From Ref. 50.)

form NR_2H^+ , and the hydrogel membrane becomes swollen and is thus permeable to insulin molecules (Fig. 27). The amount of insulin delivered is bioresponsive to the concentration of glucose penetrating into the CrDDS.

Self-Regulating Drug Delivery Systems

This type of feedback-regulated CrDDS depends on a reversible and competitive binding mechanism to activate and to regulate the release of drug. In this CrDDS, the drug reservoir is a drug complex encapsulated within a semipermeable polymeric membrane. The release of drug from the CrDDS is activated by the membrane permeation of a biochemical agent from the tissue where the CrDDS is located.

Kim et al. first applied the mechanism of reversible binding of sugar molecules with lectin into the design of self-regulating CrDDS (51). For this CrDDS, a biologically-active insulin derivative, in which insulin is coupled with a sugar (e.g., maltose), was first prepared and then conjugated with lectin to form an insulin–sugar–lectin complex. The complex is then encapsulated within a semipermeable membrane to produce CrDDS. As blood glucose diffuses into the CrDDS, it binds, competitively, with the binding sites in the lectin molecules and activates the release of the insulin–sugar derivatives from the binding sites. The released insulin–sugar derivatives diffuse out of the CrDDS, and the amount of insulin–sugar derivatives released depends on the concentration of glucose. Thus, a self-regulating drug delivery is achieved. However, a potential problem has remained to be resolved: that is, the release of insulin is nonlinear in response to the changes in glucose level (52).

Further development of the self-regulating insulin delivery system has utilized the complex of glycosylated insulin–concanavalin A, which is encapsulated inside a polymer membrane (53). As glucose penetrates into the system, it activates the release of glycosylated insulin from the complex for a controlled release from the system (Fig. 28). The amount of insulin released is thus self-regulated by the concentration of glucose that has penetrated into the insulin delivery system.

SITE-TARGETING DRUG DELIVERY SYSTEMS

Delivery of a drug to a target tissue that needs medication is known to be a complex process that consists of multiple steps of diffusion and partitioning. The CrDDSs outlined above generally address only the first step of this complex

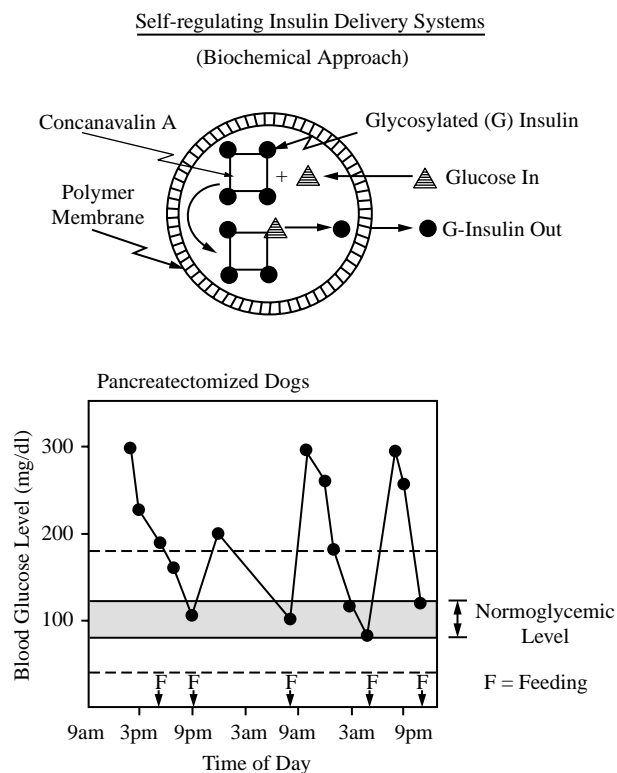


Fig. 28 Various components of a self-regulating insulin delivery system, a feedback-regulated drug delivery system, and its control of blood glucose level in the pancreatectomized dogs. (From Ref. 53.)

process. Essentially, these CrDDSs have been designed to control the rate of drug release from the delivery systems, but the path for the transport of drug molecules from the delivery system to the target tissue remains largely uncontrolled.

Ideally, the path of drug transport should also be under control. Then, the ultimate goal of optimal treatment with maximal safety can be achieved. This can be reasonably accomplished by the development of a CrDDS with a site-targeting specificity (Fig. 2). An ideal site-targeting CrDDS has been proposed by Ringsdorf (54). A model, which is shown in Fig. 29, is constructed from a nonimmunogenic and biodegradable polymer and acts as the backbone to contain three types of attachments: 1) a site-specific targeting moiety, which is capable of leading the drug delivery system to the vicinity of a target tissue (or cell); 2) a solubilizer, which enables the drug delivery system to be transported to and preferentially taken up by a target tissue; and 3) a drug moiety, which is covalently bonded to the backbone, through a spacer, and contains a linkage that is cleavable only by a specific enzyme(s) at the target tissue.

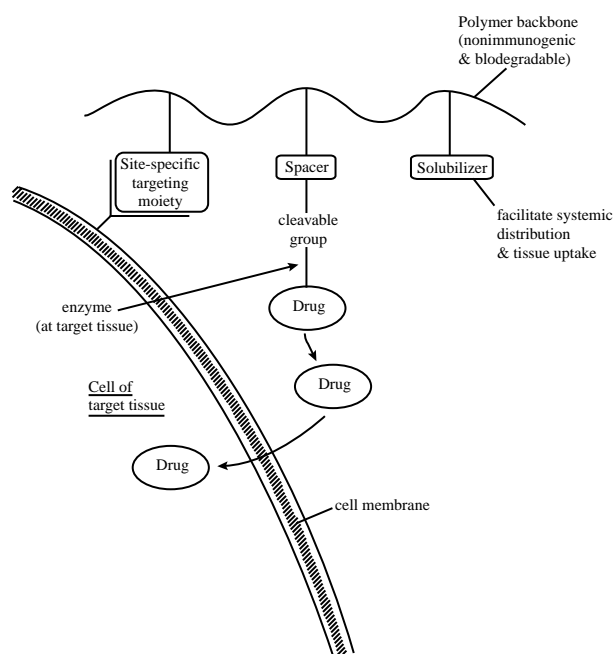


Fig. 29 An ideal site-targeting controlled-release drug delivery. (From Ref. 54.)

Unfortunately, this ideal site-targeting CrDDS is only in the conceptual stage. Its construction remains largely unresolved and is still a challenging task in the biomedical and pharmaceutical sciences.

SUMMARY

The controlled-release drug delivery systems outlined here have been steadily introduced into the biomedical community since the middle of the 1970s. There is a growing belief that many more of the conventional drug delivery systems we have been using for decades will be gradually replaced in the coming years by these CrDDSs.

REFERENCES

- Chien, Y.W. *Novel Drug Delivery Systems: Fundamental, Developmental Concepts and Biomedical Assessments*; Marcel Dekker, Inc.: New York, 1982.
- Chien, Y.W. Industrial Pharmaceutical R&D Symposium on Transdermal Controlled Release Medication. Piscataway, New Jersey, January, 14 & 15 1982 Rutgers University, College of Pharmacy, Proceedings Published in Drug Develop. & Ind. Pharm. 1983; 9, 497–744
- Chien, Y.W. Industrial Pharmaceutical R&D Symposium on Oral Controlled Drug Administrations Piscataway, New Jersey January, 19 & 20 1983; Rutgers University, College of Pharmacy, Proceedings Published in Drug Develop. & Ind. Pharm. 1983; 9, 1077–1396
- Chien, Y.W. International Pharmaceutical R&D Symposium on Advances in Transdermal Controlled Drug Administration for Systemic Medication. Piscataway, New Jersey, June, 20 & 21 1985 Rutgers University, College of Pharmacy Proceedings Published in Transdermal Controlled Systemic Medications; Chien, Y.W., Ed.; Marcel Dekker, Inc.: New York, 1987.
- Chien, Y.W. Rate-control Drug Delivery Systems: Controlled Release vs. Sustained Release. Med. Prog. Technol. 1989, 15 (1–2), 21–46.
- Chien, Y.W. *Novel Drug Delivery Systems: Second Edition, Revised and Expanded*; Marcel Dekker, Inc.: New York, 1992.
- Luukkainen, T.; Allonen, H.; Haukkanen, M.; Lahtenmaki, P.; Nilsson, C.G.; Toivonen, J. Five Years' Experience With Levonorgestrel-Releasing IUDs. Contraception 1986, 33 (2), 139–148.
- Andersson, K.; Odland, V.; Rybo, G. Levonorgestrel-Releasing and Copper-Releasing (Nova T) IUDs During Five Years of Use: A Randomized Comparative Trial. Contraception 1994, 49 (1), 56–72.
- Sivin, I.; Stern, J. Health During Prolonged Use of Levonorgestrel 20 Micrograms/d and the Copper TCu 380Ag Intrauterine Contraceptive Devices: a Multicenter Study. International Committee for Contraception Research (ICCR). Fertil. Steril. 1994, 61 (1), 70–77.
- Robinson, J.R. *Sustained and Controlled Release Drug Delivery Systems*; Marcel Dekker, Inc.: New York, 1978.
- Baker, R.W.; Lonsdale, H.K. Controlled Delivery—an Emerging Use for Membranes. Chemtech. 1975, 5, 668.
- Chien, Y.W. Methods to Achieve Sustained Drug Delivery. The Physical Approach: Implants. *Sustained and Controlled Release Drug Delivery Systems*; Robinson, J.R., Ed.; Marcel Dekker, Inc.: New York, 1978.
- Chien, Y.W. Microsealed Pharmaceutical Delivery Device. US Patent 3,992,518, Nov 16, 1976.
- Physicians' Desk Reference*; Medical Economics Company: Montvale, 1999.
- Segal, S.J. The Development of NORPLANT Implants. Stud. Fam. Plann. 1983, 14 (6–7), 159–163.
- Diaz, S.; Pavez, M.; Miranda, P.; Robertson, D.N.; Sivini, I.; Croxatto, H.B. A Five-year Clinical Trial of Levonorgestrel Silastic Implants (NorplantTM). Contraception 1982, 25 (5), 447–456.
- Weiner, E.; Victor, A.; Johansson, E.D. Plasma Levels of D-norgestrel After Oral Administration. Contraception 1976, 14 (5), 563–570.
- Croxatto, H.B.; Diaz, S.; Miranda, P.; Elamsson, K.; Johansson, E.D. Plasma Levels of Levonorgestrel in Women During Longterm Use of Norplant. Contraception 1981, 23 (2), 197–209.
- Keith, A.D. Polymer Matrix Considerations for Transdermal Devices. Drug Develop. & Ind. Pharm. 1983, 9, 605–625.
- Hsieh, D.S.T. Subcutaneous Controlled Delivery of Estradiol by Compudose Implants: In Vitro and In Vivo Evaluations. Drug Develop. & Ind. Pharm. 1987, 13, 2651–2666.

21. Wolff, M.; Cordes, G.; Luckow, V. In Vitro and In Vivo Release of Nitroglycerin from a New Transdermal Therapeutic System. *Pharm. Res.* **1985**, *1*, 23–29.
22. Corbo, M.; Liu, J.C.; Chien, Y.W. Transdermal Controlled Delivery of Propranolol from a Multilaminate Adhesive Device. *Pharm. Res.* **1989**, *6* (9), 753–758.
23. Microsealed Pharmaceutical Delivery Device. US Patent 4,053,580 Oct 11, 1977.
24. Chien, Y.W. Microsealed Drug Delivery Systems: Fabrication and Performance. *Methods in Enzymology*; Widder, K.J., Green, R., Eds.; Academic Press: New York, 1985; 461–470.
25. Sanvordeker, D.R.; Cooney, J.G.; Wester, R.C. Transdermal Nitroglycerin Pad. US Patent 4,336,243, June 22, 1982.
26. Karim, A. Transdermal Absorption: A Unique Opportunity for Constant Delivery of Nitroglycerin. *Drug Develop. & Ind. Pharm.* **1983**, *9*, 671.
27. Theeuwes, F.; Yum, S.I. Principles of the Design and Operation of Generic Osmotic Pumps for the Delivery of Semisolid or Liquid Drug Formulations. *Ann. Biomed. Eng.* **1976**, *4* (4), 343–353.
28. Theeuwes, F. Elementary Osmotic Pump. *J. Pharm. Sci.* **1975**, *64* (12), 1987–1991.
29. De Leede, L.G.J. *Rate-Controlled and Site-Specified Rectal Drug Delivery*; State University of Leiden: Leiden, The Netherlands, 1983.
30. Theeuwes, F. Oros-Osmotic System Development. *Drug Develop. & Ind. Pharm.* **1983**, *9*, 1331–1357.
31. Liu, J.-C.; Farber, M.; Chien, Y.W. Comparative Release of Phenylpropranolamine HCl for Long-acting Appetite Suppressant Products: Acutrim Vs. Dexatrim. *Drug Develop. & Ind. Pharm.* **1984**, *10*, 1639–1661.
32. Michaels, A.S. Device for Delivering Drug to Biological Environment. US Patent 4,180,073, Dec. 25, 1979.
33. Blackshear, P.J.; Rohde, T.D.; Grotting, J.C.; Dorman, F.D.; Perkins, P.R.; Varco, R.L.; Buchwald, H. Control of Blood Glucose in Experimental Diabetes by Means of a Totally Implantable Insulin Infusion Device. *Diabetes* **1979**, *28* (7), 634–639.
34. Blackshear, P.J.; Rohde, T.D.; Varco, R.L.; Buchwald, H. One Year of Continuous Heparinization in the Dog Using a Totally Implantable Infusion Pump. *Surg. Gynecol. Obstet.* **1975**, *141* (2), 176–186.
35. American Pharmacy, Implantable Pump for Morphine. NS24:20 **1984**.
36. Hsieh, D.S.T.; Langer, R. Zero-order Drug Delivery Systems with Magnetic Control. *Controlled Release Delivery Systems*; Roseman, T.J., Mansdorf, S.Z., Eds.; Marcel Dekker, Inc.: New York, 1983.
37. Kost, J. Ultrasound for Controlled Delivery of Therapeutics. *Clin. Mater.* **1993**, *13* (1–4), 155–161.
38. Tyle, P.; Agrawala, P. Drug Delivery by Phonophoresis. *Pharm. Res.* **1989**, *6* (5), 355–361.
39. Bertolucci, L.E. Introduction of Anti-inflammatory Drugs by Iontophoresis: Double Blind Study. *J. Orthopaed. & Sports Phys. Ther.* **1982**, *4*, 103.
40. Glass, J.M.; Stephen, R.L.; Jacobson, S.C. The Quantity and Distribution of Radiolabeled Dexamethasone Delivered to Tissue by Iontophoresis. *Int. J. Dermatol.* **1980**, *19* (9), 519–525.
41. Harris, P.R. Iontophoresis: Clinical Research in Musculoskeletal Inflammatory Conditions. *J. Orthopaed. & Sports Phys. Ther.* **1982**, *4*, 109.
42. Controlled Release Suspensions, APhA/APS Midwest Reg. Meet.. Chicago Illinois April, 2 1984.
43. Prolonged Release Pharmaceutical Preparations. US Patent 4,221,778 Sept 9, 1980.
44. Raghunathan, Y.; Amsel, L.; Hinsvark, O.; Bryant, W. Sustained-Release Drug Delivery System I: Coated Ion-exchange Resin System for Phenylpropanolamine and Other Drugs. *J. Pharm. Sci.* **1981**, *70* (4), 379–384.
45. Heller, J. Biodegradable Polymers in Controlled Drug Delivery. *Crit. Rev. Ther. Drug Carrier Syst.* **1984**, *1* (1), 39–90.
46. Morimoto, Y.; Fujimoto, S. Albumin Microspheres As Drug Carriers. *Crit. Rev. Ther. Drug Carrier Syst.* **1985**, *2* (1), 19–63.
47. Sezaki, H.; Hashida, M. Macromolecule-drug Conjugates in Targeted Cancer Chemotherapy. *Crit. Rev. Ther. Drug Carrier Syst.* **1984**, *1* (1), 1–38.
48. Heller, J.; Pengburn, S.H. A Triggered Bioerodible Naltrexone Delivery System. *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **1986**, *13*, 35–36.
49. Heller, J.; Trescony, P.V. Controlled Drug Release by Polymer Dissolution. II: Enzyme-mediated Delivery Device. *J. Pharm. Sci.* **1979**, *68* (7), 919–921.
50. Horbett, T.A.; Ratner, B.D.; Kost, J.; Singh, M. A Bioresponsive Membrane for Insulin Delivery. *Recent Advances in Drug Delivery Systems*; Anderson, J.M., Kim, S.W. Eds.; Plenum Press: New York, 1984; 209–220.
51. Kim, S.W.; Jeong, S.Y.; Sato, S.; McRea, J.C.; Feijan, J. Self-Regulating Insulin Delivery Systems—A Chemical Approach. *Recent Advances in Drug Delivery Systems*; Anderson, J.M., Kim, S.W., Eds.; Plenum Press: New York, 1983; 123.
52. Baker, R.W. *Controlled Release of Biologically active Agents*; J. Wiley & Sons: New York, 1987.
53. Jeong, S.Y.; Kim, S.W.; Eenink, M.J.D.; Feijen, J. Self-Regulating Insulin Delivery Systems. I. Synthesis and Characterization of Glycosylated Insulin. *J. Controlled Release* **1984**, *1*, 57–66.
54. Ringsdorf, H. Synthetic Polymeric Drugs. *Polymeric Delivery Systems*; Kostelnik, R.J., Ed.; Gordon and Brech: New York, 1978.
55. Chien, Y.W. Logics of Transdermal Controlled Drug Administration. *Drug Develop. & Ind. Pharm.* **1983**, *9*, 497–520.
56. Noonan, P.K.; Gonzalez, M.A.; Ruggirello, D.; Tomlinson, J.; Babcock-Atkinson, E.; Ray, M.; Golub, A.; Cohen, A. Relative Bioavailability of a New Transdermal Nitroglycerin Delivery System. *J. Pharm. Sci.* **1986**, *75* (7), 688–691.
57. Chien, Y.W. Microsealed Drug Delivery Systems: Theoretical Aspects and Biomedical Assessments. *Recent Advances in Drug Delivery Systems*; Anderson, J.M., Kim, S.W., Eds.; Plenum Press: New York, 1984; 367–387.
58. Chien, Y.W. Potential Developments and New Approaches in Oral Controlled Release Drug Delivery Systems. *Drug Develop. & Ind. Pharm.* **1983**, *9*, 1291–1330.

DRUG DELIVERY—PULMONARY DELIVERY

Michael T. Newhouse

Inhale Therapeutic Systems, Inc., San Carlos, California, U.S.A.

OVERVIEW

New dispersible formulations and drug aerosol delivery devices for inhaleable peptides, proteins and various small molecules have, in the past decade, become of increasing interest for the treatment of systemic and respiratory diseases. These include, but also extend well beyond, the traditional and long available (although still underutilized) therapies for asthma and chronic obstructive pulmonary disease (COPD). Advances in the use of the lungs as portals for delivery of medication to the blood stream have greatly expanded the potential applications of pulmonary delivery. This advanced technology was initially applied to the systemic delivery of large molecules, such as insulin, interferon- β , or α_1 proteinase inhibitor. By facilitating the systemic delivery of large and small molecule drugs through inhalation deep into the lung, this advanced pulmonary technology provides a unique and innovative delivery alternative for therapies that must currently be administered by injection (i.v., i.m., s.c.) or by oral delivery that causes adverse effects or is poorly absorbed. Indeed, a major advantage of therapy via the lungs is the potentially improved therapeutic index, that is, the ratio of therapeutic benefit to adverse effects. This applies mainly to the therapy of pulmonary disease, but may also be applicable to systemic disease due to reduced first-pass metabolism that may be associated with hepatocellular injury. Pulmonary delivery also offers the potential for better and possibly more economical treatment or prophylaxis of respiratory and systemic diseases (e.g. viral vaccines).

ADVANTAGES OF INHALEABLES

Advanced technology for pulmonary delivery is expanding a category of drugs called “inhaleables,” defined as respiratory and systemic therapies administered simply by inhaling. Inhaleables offer several advantages over injectable, transdermal or oral methods of delivery that make them more appealing to both patients and physicians for treating a variety of diseases by means of currently available and future therapies.

First, inhaleables provide a noninvasive method of delivering drugs into the bloodstream for those molecules that currently can only be delivered by injection. These include peptides and proteins, such as insulin for diabetes or interferon beta for multiple sclerosis and most of the drugs developed in recent years by biotechnology companies. Inhale Therapeutic Systems, Inc. (San Carlos, California.) is pioneering advanced pulmonary drug delivery technology to provide a convenient and pain-free alternative to injection for systemic delivery of peptides and proteins. Feedback from patients in the clinical trials and extensive market research support the view that inhaleable drugs will be welcome alternatives to injections.

Second, inhaleables enable effective drug targeting to the lungs for relatively common respiratory tract diseases such as asthma, emphysema, bronchiectasis and chronic bronchitis. This direct delivery most often results in a better treatment outcome while potentially requiring less drug than if given systemically either orally or by injection.

Third, inhaleables provide for very rapid onset of action similar to the i.v. route and quicker than can be achieved with either oral delivery or subcutaneous injections. More rapid delivery could benefit treatments for pain, seizures, panic/anxiety attacks, hypertensive crises, anaphylaxis (severe allergies, food, insect bites), nausea, cardiovascular conditions (arrhythmia, strokes), and Parkinson's “lock-up”—indications where speed is important.

Fourth, inhaling instead of taking pills can help avoid gastrointestinal tract problems such as poor solubility, low bioavailability, gut irritability, unwanted metabolites, food effects and dosing variability.

MACROMOLECULES AND THEIR IMPORTANCE

Macromolecules are polymers composed of three or more amino acids, sugars, nucleotides, etc. While the large protein molecules are usually made by means of recombinant technologies, the smallest peptides are made primarily by chemical synthesis.

After nearly two decades of activity, innovations in biotechnology and recombinant gene techniques have led to an increase in the approved use of many macromolecule drugs. In recent years, at least 30 macromolecule medications have been approved for marketing in the United States alone, and more than 130 are now in human clinical trials, many for the treatment of chronic and subacute diseases that afflict a large percentage of people worldwide. This is particularly significant, as most of the diseases in question require multiple drug doses and, therefore, multiple injections over many years.

For many years, medical science has been looking for an alternative to injections for the delivery of macromolecule drugs. Due principally to their size, these molecules, mostly proteins and peptides, cannot naturally and efficiently pass through the skin or nasal membranes without the use of penetration enhancers, such as detergents or electrical impulses. If administered orally, they are digested or degraded before they reach the bloodstream. Therefore, oral, transdermal and nasal routes of delivery are inefficient for these molecules. In contrast, research has shown that many of those same molecules are absorbed naturally and quickly into the bloodstream if they are delivered to the deep lung through the use of inhaleables.

WHY IS PULMONARY DELIVERY THE BEST ROUTE?

At best, chronic injection is an unpleasant prospect with a host of hygiene issues and potential side effects. At worst, it can create a barrier to patient compliance with the particular drug regimen required to most effectively treat a given disease, since some patients choose irregular treatment or no treatment at all when faced with frequent injections.

While injection has served as the primary means of delivering macromolecules produced by biotechnology, many noninvasive routes have been explored as alternatives. Oral delivery remains the most common method of delivery for most small molecule drugs. However, oral delivery most often does not work for macromolecules because proteins are digested before they have an opportunity to reach the bloodstream. Commercially successful oral delivery of peptides and proteins has not been achievable with the exception of DDAVP (9 amino acids) and cyclosporin (11 amino acids), two digestion resistant small peptides.

The skin offers an even less naturally permeable boundary to macromolecules than the gastrointestinal tract. Thus, passive transdermal delivery of proteins and

peptides using “patch” technology has not succeeded. Peptides and proteins can be shot through the skin into the body using high-pressure “needle-less” injection devices. The devices, which inject proteins like insulin, have been available for years, however they have failed to impress doctors or patients due to the associated discomfort and the potential for “splash back” to transmit blood-borne diseases such as AIDS or hepatitis.

Nasal delivery is inefficient in terms of the amount of drug actually delivered to the body and to increase its efficiency, penetration enhancers must be added that may cause local irritation.

In contrast, research has shown that many molecules are absorbed through the deep lung into the bloodstream naturally with relatively high bioavailability and without the need for enhancers used by other noninvasive routes (1). With regard to the treatment of systemic disease, bioavailability is defined as the amount of drug that actually reaches the bloodstream by any method of delivery, compared to the mass of the agent with which the delivery “device” was charged.

The respiratory tract is accustomed to dealing with chronic exposure to a relatively large load of biological and nonbiological particulates. These are contained in the 20,000 L of air that must be inhaled daily to accomplish gas exchange. It is a tribute to the effectiveness of lung defense mechanisms that in healthy people, for most of their lives, the lungs are sterile below the larynx. According to the American Conference of Governmental Industrial Hygienists, a person can inhale approximately 30 mg/day of inert nuisance dusts into the lung day after day without effect (2), suggesting that the lung is a rather robust organ. Further, there is no evidence that inhaling autologous (self) proteins presents any immune response issues. The high bioavailability provided by the deep lung and the robustness of this organ make it a natural portal of entry for peptides, proteins and other small molecules that could be used to provide systemic therapy.

Riding on Air—Getting to the Deep Lung

The lung provides an enormous surface area through which molecules can be absorbed into the bloodstream. When a breath of air is inhaled, it travels down the trachea and the conducting airways to reach the alveolar epithelium. The conducting airways branch 12–23 times and their surface area measures approximately 0.8 m² in adults. The epithelium of the branching airways of the lungs are lined by a relatively thick, ciliated, pseudo-stratified columnar epithelial layer covered with low viscosity periciliary fluid. Floating above the periciliary

fluid are large “rafts” of thicker gel-like mucus which are propelled towards the pharynx by the rapidly beating cilia.

Once a drug aerosol has made its way through the conducting airways to deposit in the deep lung, the major barriers to entering the body are the 0.15 μm layer of type I alveolar cells that are covered by a very thin layer of epithelial lining fluid consisting mainly of surfactant and the relatively permeable endothelium of the alveolar capillaries. Alveolar cells have so called “tight” junctions that act as a relative barrier to the absorption of large molecules such as proteins and peptides and prevent the development of pulmonary edema.

THE ALVEOLAR EPITHELIUM

The alveolar epithelium measures approximately 100 m^2 in adults—approximately the size of a singles tennis court. It is made up of approximately 500,000,000 tiny airsacs, 300 μm in diameter, called alveoli. These are enveloped by an equally large capillary network and it is across this enormously large and extremely thin (0.1–0.2 μm) membrane that gas exchange and the transcytosis of large and small molecules occurs. The alveolar epithelium is composed of a thin, non-ciliated, nonmucus-covered cell layer consisting mainly of type I and type II fixed alveolar cells. A thin epithelial lining fluid, mainly surfactant, covers the type I and II alveolar epithelial cells.

Type I pneumocytes make up most of the epithelial surface. It is the large, thin, type I pneumocytes that are the primary site of pulmonary protein absorption. The type II pneumocytes, lying in niches between type I cells, are the main source of surfactants and also replace type I cells as they undergo apoptosis (programmed cell death) after about 120 days.

Beneath and extending between these cells are antigen evaluating and presenting dendritic cells. On the epithelial surface are motile phagocytes called macrophages, about 15 to each alveolus. These remove foreign biological and non-biological particles and assist in maintaining the sterility of the alveolar surface. Macrophages are the immune system's first line of defense against inhaled organisms. The key to preventing macrophages from engulfing inhaled drug particles is solubility. Ideally, drugs are rapidly dissolved in the epithelial lining fluid on the surface of the epithelium thus assisting their ingestion by macrophages. This process can be accelerated by means of small, drug-containing, lipid particles.

TRANSCYTOSIS

The body absorbs peptides and proteins into the bloodstream by a natural process known as transcytosis, which occurs deep in the lung. Transcytosis allows drug molecules to move across an impermeable cell membrane without creating holes in the cells and destroying the barrier (Fig. 1).

The process is performed by trillions of tiny membrane bubbles, or transcytotic vesicles, which form from invaginations of the cell membrane on one side of the cell and fuse back into the membrane on the other side of the cell. The result is that small volumes of alveolar fluid, containing dissolved proteins, are rapidly carried by a “bucket brigade” from one side of the alveolar cell to the other.

Small molecules and peptides are also thought to be absorbed through the lung surface by an analogous process called paracellular transport. This is achieved through the tight junctions which connect cells to each other. However, in contrast to transcytosis that is rapid and efficient, paracellular transport is slow and inefficient.

Many dissolved proteins that have been transported through the alveolar cells are rapidly reabsorbed into the blood through the capillary endothelial cells (also by

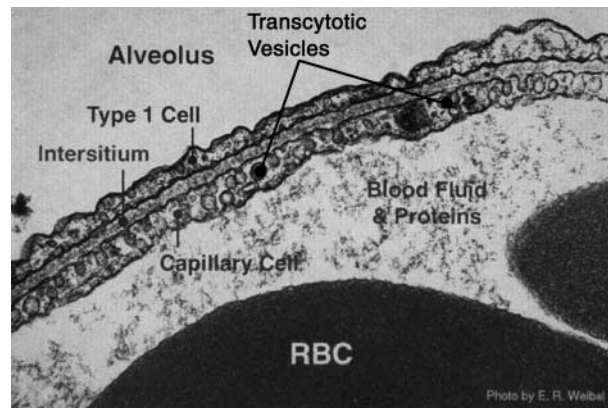


Fig. 1 Natural absorptive mechanism. The body absorbs peptides and proteins into the bloodstream by a natural process known as transcytosis which occurs deep in the lung. Transcytosis is the process by which large molecules move across an impermeable cell membrane without creating holes in the cells and destroying the barrier. It is performed by tiny membrane bubbles, or transcytotic vesicles, which form invaginations of the cell membrane on one side of the cell and dissolve back into the membrane on the other side of the cell. The result is that small volumes of alveolar fluid, including dissolved proteins, are carried by a “bucket brigade” from one side of a cell to the other.

transcytosis). In the case of some large proteins, they are more slowly drained across the interstitial space by means of the pulmonary lymphatics, which also empty into the bloodstream via the thoracic duct and superior vena cava.

The purpose of transcytosis in lung biology is not clearly understood. It is thought to be a natural mechanism for controlling the level of airway fluid and for moving endogenous proteins back and forth across the epithelium during normal physiological “housekeeping.” Both transcytosis and paracellular transport are sophisticated cell processes mediated by complex cell machinery.

The result of these two processes is a noninvasive means of delivering proteins and peptides to the bloodstream with relatively high bioavailability and without the use of penetration enhancers (3).

Because the molecules are delivered rapidly into the bloodstream, there is a much more rapid onset of action than with any other non-i.v. delivery method. This can be particularly useful in indications where speed is important such as pain control, relief of muscle spasm, panic/anxiety attacks, hypertensive crises, cardiac arrhythmias, anaphylaxis (severe allergies to food and insect bites), nausea, cardiovascular conditions (arrhythmia, strokes), Parkinson’s “lock-up,” and epileptic seizures.

Characteristics of Lung Absorption

Large, highly vascularized area available for transcytosis

Conducting airways $\sim 0.8 \text{ m}^2$

Alveoli $\sim 80 \text{ m}^2$

Alveoli highly permeable to many biologics; most small molecules and many macromolecules capable of absorption throughout the respiratory tract

Relatively rapid, first-order absorption

Much less first-pass metabolism and degradation in the gastrointestinal tract and liver

Cytochrome P450 enzyme concentration in human lungs $< 0.7\%$ that of liver

Only 1A1 and 4B1 P450 isozymes reported in the human lung

CLINICAL APPLICATIONS OF PULMONARY DELIVERY

Pulmonary drug delivery in our post-genomics era has opened the door to noninvasive administration of a wide variety of macromolecules. Since drug delivery by the pulmonary route is now practical, drug companies can reconsider the development of new macromolecules and small molecules whose markets are limited as an

injectable or when given by mouth. Furthermore, rather than spending years to develop an oral version (or abandon a project altogether) especially for chronic or subchronic therapies, pharmaceutical companies can save years of development time by using pulmonary delivery for their macromolecule drugs. Pulmonary delivery could also replace some oral drugs due to the much faster onset of action with improved absorption and avoidance of first pass losses with delivery through the GI tract.

Nearly every biotherapeutic product involving chronic or longterm use would benefit from noninvasive delivery, which could provide pharmaceutical companies with a competitive advantage, expand the market for products and/or enable new indications to be considered. Further, proprietary new inhalation delivery systems can extend the patent life of a drug, increase patient compliance and possibly reduce healthcare costs.

Because of the advances in biotechnology that have resulted not only in new macromolecules but also in new devices to deliver them via the lungs, patients and physicians will soon be able to use this route to treat diseases such as diabetes, hepatitis, osteoporosis, multiple sclerosis, genetic emphysema, cystic fibrosis, and other pulmonary infections among others.

Pulmonary delivery could also be used for delivery of vaccines. In the United States alone, children endure as many as 14 vaccine injections by age 16 and that number is climbing as new research makes more vaccines available. The cost of a single injection can quadruple if the costs of equipment and personnel are included. Furthermore, fear of needles reduces compliance, thus the potential for an increased market share in this area. Inhaled vaccines may be used to prevent influenza, pneumonia, tuberculosis, measles, cytomegalovirus, asthma, and mucosal-entry diseases such as sexually transmitted diseases including HIV.

PULMONARY DISEASE AND PULMONARY DELIVERY

Pulmonary drug targeting has long been used to treat lung disease, particularly asthma and COPD, non-CF and CF bronchiectasis, bronchiolitis and recently influenza. Potentially, other pulmonary diseases such as parenchymal fibrosis, acute bronchitis, pneumonia and even carcinoma of the lung in situ might be treated by inhalation. Secondary lung malignancies have been treated by inhalation of IL2 and GM-CSF with some benefit.

With regard to pulmonary infections, much higher concentrations of antibiotics can be achieved in the lungs

by inhalation, which should accomplish greater and more rapid bacterial killing with less likelihood of developing bacterial resistance. Furthermore, topical delivery achieves reduced systemic side effects for equivalent therapeutic benefit.

CHALLENGES AND SOLUTIONS IN PULMONARY DELIVERY

In the mid-1950s the first pressurized metered dose inhaler (MDI) was developed for the administration of bronchodilator drugs locally to the lung. It was a major advance for the treatment of asthma since it made aerosol medications readily available in an inexpensive small multidose device.

The world aerosol market has grown due to the increased incidence of asthma and chronic obstructive pulmonary disease (COPD) as well as due to an increased number of patients receiving aerosol medications as the drug formulation-device combination of choice. Until recently, companies developed pulmonary drug delivery systems primarily to dispense drugs to the airways of the lung for local lung applications. Application systems such as pressurized metered dose inhalers (pMDIs), breath activated dry powdered inhalers (DPIs), liquid jet and ultrasonic nebulizers have proved useful in the management of airway inflammation and bronchoconstriction.

For the systemic delivery of most drugs, however, currently marketed aerosol delivery systems are inadequate due to the following:

1. *Low System Efficiency:* To be commercially feasible for the administration of costly proteins and peptides, the overall efficiency of presently available systems has remained generally too low. Correct aerosol particle size is very important for optimum deep lung delivery. Studies have established that these particles should range from one to three microns in aerodynamic diameter for optimum lung deposition efficiency. If the particles are too large, they impact in the oropharynx and larynx. If they are too small, they will be exhaled. Most existing MDI systems can only deliver a small fraction (about 10–20%) of the dispensed drug in the correct particle size for deep lung deposition (4–6) although, recently developed 1 μm solution aerosols from corticosteroid pMDIs have achieved lung deposition efficiencies of 60% or more.
2. *Low Drug Mass per Puff:* With most existing systems, the total amount of drug per puff delivered to the lower respiratory tract is too low—less than 1000 mcg (6)—to enable practical delivery of many macromolecules which require milligram doses. Device payload

versatility is an important feature with the new macromolecule drugs since they come in a wide variety of potencies—from a few micrograms per dose to tens of milligrams. Traditional inhalation systems have primarily been designed to deliver some of the most potent drugs in use today, namely the inhaled bronchodilators and corticosteroids used for treating asthma. Both of these classes of medication are bioactive in the lung at 5–50 $\mu\text{g}/\text{dose}$. In contrast, many peptide and protein drugs require deep lung doses in the 2–20 mg/dose range (1).

3. *Poor Formulation Stability for Macromolecules:* existing aerosol systems are not designed to protect the formulations of delicate macromolecules. Most traditional small molecule asthma drugs are crystalline and, in the case of corticosteroids, relatively moisture resistant in the dry state. They are also rather stable in liquids as compared to most macromolecules, which are unstable in the liquid state, amorphous, and highly moisture sensitive in the dry state. There are exceptions, including Genentech's Pulmozyme, the first FDA approved aerosol protein, a 33 kDa digestive enzyme (DNAse) used to break up the thick, grossly infected, mucus in cystic fibrosis. It is available as a stable liquid formulation for nebulization. Other proteins such as growth hormone, G-CSF and Interferons aggregate and are partially denatured by nebulization (7).
4. *Poor Dosing Reproducibility:* for a variety of reasons, the dosing reproducibility of many existing systems is too variable for systemic delivery of most macromolecule drugs (6, 8). Physicians and patients alike have tolerated the highly variable dosing of inhaled asthma medications for years because the drugs have a wide therapeutic window and optimizing the drug dose is usually a matter of trial and error. In the case of the bronchodilators, the rapid improvement characterized by easier breathing has enabled patients to know whether or not they have used the proper inhalation technique and dose. So far, no macromolecule drug appears to possess such rapid biofeedback.

Potential solutions to these challenges to ensure effective inhalation drug treatment include active dry powder delivery systems, active liquid blister technology, and hydrofluorocarbon (HFC) propellant nebulization systems.

Dry Powder Inhalation (DPI) Systems

Dry powder aerosols are frequently highly soluble and quickly dissolve in the fluid layer lining the surface of the deep lung before passing through the thin cytoplasm of the

type I alveolar cells the interstitial “space” and capillary endothelium. The main advantages of dry powder systems include product and formulation stability (even at room temperature or above), the potential for delivering a low or high mass of drug per puff, low susceptibility to microbial growth, and applicability to both soluble and insoluble drugs.

Current challenges facing the development of these systems for macromolecules include moisture control, efficient powder manufacturing, reproducible powder filling, unit dose packaging and development of efficient reliable aerosol dispersion and delivery devices.

One of the challenges is that the fine powder particles tend to stick to each other. These clumps can be difficult to break apart into breathable particles for slow inhalation by the patient. Breath-powered powder inhalers for asthmatics attempt to apply the forces generated by a rapid forceful inspiration to break apart the powder clumps. But vigorous rapid inhalation does not efficiently deaggregate and target the very fine powder clumps to the deep lung. This is mainly because particle inertia causes impaction of most of the medication in the oropharynx.

One solution to the clumping problem for macromolecule drug delivery is to use a device with sufficient power to force the deaggregation of even fairly adhesive powders. One company working on this issue, Inhale Therapeutic Systems, of San Carlos, California, has designed a device that uses sonic velocity compressed air (1.5 mL) to aerosolize the powder (Fig. 2). Systemic Delivery. These de-agglomerated particles form a standing cloud in an aerosol holding chamber. The patient then inhales the stationary cloud with one slow, rate controlled deep breath, eliminating the need for patient coordination between the generation of the aerosol powder and inhalation. Furthermore, the slow deep breath encourages efficient alveolar delivery of the drug.

The keys to efficient aerosol targeting to the deep lung are:

1. Formulation of readily dispersible powders by particle engineering
2. A device that is small, inexpensive and user-friendly to generate the drug aerosol, and
3. Inhalation at a low inspiratory flow rate (below 20 L/min) to minimize upper airway and large airway deposition.

Liquid Systems

Liquid systems provide ease of filling and availability (in some cases) of preexisting injectable formulations for macromolecules. These systems, however, also pose



Fig. 2 Advanced inhaler technology for pulmonary systemic delivery. Shown here: a collapsed and extended advanced inhaler used for systemic delivery. A patient would use this device and powdered medicine instead of receiving an injection or taking a pill. The device slides open and closed like a telescope for compact carrying, making it portable for patients who must self-administer drugs regularly. A patient would open the device, insert a blister of powdered medicine and then pump the handle which compresses a small amount of air inside the device. By pushing a button, the patient causes the compressed air to be released at extremely high velocity, breaking apart the powder from the blister and sending it into the 200-mL chamber where it is captured as a stationary cloud. When patients inhale this fine powder formulation of the drug as an aerosolized cloud, they receive the medicine first and then a volume of air that helps push the drug deep into the lung, where it is absorbed into the blood.

hurdles for their use with protein and peptide drug delivery. They currently offer lower drug payload per puff than dry systems (because more than 95% of the mass is

water), difficulty in formulation stabilization (particularly for insoluble drugs), and greater susceptibility to microbial growth within a device. Current challenges facing the development of liquid systems for macromolecules are formulation stability, unit dose packaging, high payload delivery and development of efficient reliable devices.

One of the oldest medical aerosol delivery systems is the air jet nebulizer, which forms a fine mist of liquid droplets from a drug solution that a patient breathes over a period of 10–30 min/dose. The cumbersome electric powered jet nebulizer is often used as a fallback delivery system for asthmatics who do not appear to be getting relief from the small portable MDIs and DPIs. Nebulizers are very hard on macromolecules and cause aggregation in the device before some macromolecules can be inhaled (7).

Some companies are focusing on the development of a new generation of liquid systems that are portable, more efficient, more gentle on the macromolecule and that can deliver the medicine in far fewer puffs than the old fashioned nebulizer. Two companies who are developing advanced pulmonary delivery liquid systems are Aradigm Corporation, Hayward, CA, and AeroGen, Inc., Sunnyvale, CA.

Propellant Systems

The best example of propellant-powered systems is the well-known little canister inhalers (pressurized metered dose inhalers or pMDIs) used by asthmatics since the 1950s. The original, ozone-depleting chlorofluorocarbon (CFC) propellants are being replaced by the more environmentally friendly hydrofluorocarbons (HFCs). HFC propellant systems give patients the convenience of small, inexpensive multidose devices that can be filled easily. However, they appear at this time not to be amenable to stable formulating of macromolecular therapeutic applications. Current designs deliver low payloads per puff, have low lung delivery efficiency, poor stability for water-soluble macromolecules and moderately high dosing variability (6).

OTHER ISSUES REGARDING PULMONARY DELIVERY

Notwithstanding the scientific advances in this field, many still have concerns about acceptable bioavailability and reproducibility with pulmonary delivery while others harbor some fears about the safety of deep lung inhalation of macromolecules.

Dose reproducibility: Several human studies comparing aerosol insulin administration to subcutaneously administered insulin showed that the variability in glucose response from a liquid nebulizer that utilized the standing cloud concept was equivalent or better than that seen with insulin injection (9). Inhale Therapeutics Systems, Inc. has adopted this standing cloud concept for its dry powder inhaler to achieve reproducibility of delivery of macromolecules to the systemic circulation that is equivalent to subcutaneous injections.

Safety: A growing body of data indicates that inhaling proteins can be safe, whether the patient has healthy lungs or a pulmonary disease (see Table 1). With relatively stable lung diseases, unless the damage or deterioration of the lung parenchyma is very severe—that is, if the patient has lost over 60% of lung function—most researchers do not expect systemic variability from pulmonary delivery to differ from that seen with injections. With typical upper respiratory diseases such as colds and flu, lung function is thought to remain within 70–100% of normal. Even with a cold or flu where inspiratory capacity can fall up to 30% below normal, a deep breath that is 70% of a normal deep inspiration can deliver the aerosol drug into the lungs if the standing cloud technique is used to provide an aerosol bolus at the very beginning of a slow deep breath.

Despite considerable clinical experience with aerosolized macromolecules, there have been no serious safety issues to date, nor have there been significant problems with throat irritation or cough. So far, the proteins and peptides under development for human applications have

Table 1 Aerosol safety of peptides and proteins

Molecule	Study ^a
Insulin	89 patients 24 months
DNase	Approximately 10,000 patients Almost four years
Heparin	544 patients 1 day to 1.3 years
Interferon- α	16 patients Up to 67 weeks
Leuprolide acetate	Hundreds of patients 6 months
α_1 -Antitrypsin	12 patients 1 day; 1–7 days
Antibiotics	Hundreds of patients Up to 2 years
Interferon- γ	5 patients 12 days

^aNone of these studies had any adverse effect on lung.
(From Refs. 10–17.)

been virtually tasteless and the incidence of cough, with properly formulated drugs, has been very low. However, large particle sizes (more than 5 μm) can cause cough regardless of chemical composition, so it is important that particle sizes be kept in the so-called fine particle range (i.e., less than 5 μm).

Given the advances in pulmonary delivery technology, the issues for drug companies and patients concerning pulmonary delivery revolve around economic evaluations, approvals, administration and managed health care. As these issues are resolved, pulmonary delivery will doubtless become regarded as one of the leading drug delivery alternatives.

REFERENCES

1. Akwete, A.L., Gupta, P.K., Eds.; Niven, Delivery of Biotherapeutics by Inhalation Aerosol. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Marcel Dekker, Inc.: New York, 1997; 151–231.
2. *American Conference of Governmental Industrial Hygienists* Threshold Limit Values for Chemical Substances and Physical Agents and Biological Exposure Indices (1992), Cincinnati, OH, 29.
3. Patton, J.S. Mechanisms of Macromolecule Absorption by the Lungs. *Adv. Drug Delivery Rev.* **1996**, 3–36.
4. O'Riordan, T.G.; Palmer, L.B.; Smaldone, G.C. Aerosol Deposition in Mechanically Ventilated Patients: Optimizing Nebulizer Delivery. *Amer. J. Resp. Crit. Care Med.* **1994**, 149, 214–219.
5. Hindle, M.; Byron, P.R. Dose Emissions from Marketed Dry Powdered Inhalers. *Int. J. Pharm* **1999**, 116–169.
6. Lalor, C.B.; Hickey, A.J. Generation and Characterization of Aerosols for Drug Delivery to the Lungs. In Ref (1). 235–276.
7. Akwete, A.L., Gupta, P.K., Eds.; Delivery of Biotherapeutics by Inhalation Aerosol. *Inhalation Delivers of Therapeutic Peptides and Proteins*; Marcel Dekker, Inc.: New York, 1993; 151–231.
8. Everard, M.L.; Devadason, S.G.; Summers, Q.A.; Le Souef, P.N. Factors Affecting Total and 'Respirable' Dose Delivered by a Salbutamol Metered Dose Inhaler. *Thorax* **1995**, 746–749.
9. Patton, J.S.; Platz, R.M. Aerosol Insulin—A Brief Review. *Respiratory Drug Delivery IV* **1994**, 65–74.
10. Cefalu, W.T.; Balagtas, C.C.; Landschulz, W.H.; Gelfand, R.A. *Sustained Efficacy and Pulmonary Safety of Inhaled Insulin During 2 Years of Out Patient Therapy*; Poster Presented at American Diabetes Association 60th Scientific Sessions, San Antonio, Texas, June 2000; For the Inhaled Insulin Phase II Study Group, Burlington, VT and Groton, CT.
11. Ogden, Jill, Dr.; Rogerson, Cheryl; Smith, Ian Issues in Pulmonary Delivery. *Scrip Magazine*; June 1996; 56–60.
12. Kohler, D. Aerosolized Heparin. *Journal of Aerosol Medicine* **1994**, 7 (4), 307–314.
13. Kinnula, V.; Cantell, K.; Mattson, K. Effect of Inhaled Natural Interferon-Alpha on Diffuse Bronchioalveolar Carcinoma. *European Journal of Cancer* **1990**, 26, 740–741.
14. Martin, R.J.; Boguniewicz, M.; Henson, J.E.; The Effects of Inhaled Interferon Gamma in Normal Human Airways. *Amer. Rev. of Respiratory Disease* **1993**, 148, 1677–82.
15. Adjei, A.L. Regulatory Issues Concerning Systemic Delivery via the Respiratory Tract: Leuprolide as a Demonstration Model. *Journal of Aerosol Medicine* **1995**, 8, 131.
16. McElvaney, N.G.; Hubbard, R.C.; Birrer, P.; Aerosol Alpha-1 Antitrypsin Treatment for Cystic Fibrosis. *Lancet* **1991**, 337, 392–394.
17. Hall, C.B. Aerosolized Treatment of Acute Pulmonary Infections. *Journal of Aerosol Medicine* **1989**, 2.

FURTHER READING

- American Conference of Governmental Industrial Hygienists* Threshold Limit Values for Chemical Substances and Physical Agents and Biological Exposure Indices, Cincinnati, OH, 1992; CA, 229
- Green, J.D. Pharmacological-Toxicological (Expert Report on Recombinant Human Deoxyribonuclease I (rhDNase: pulmozymeTM). *Human & Experimental Toxicology* **1994**, 13, S1–S42.
- Holt, P.G.; Britten, D.; Sedgwick, J.D. Suppression of IgE Responses by Antigen Inhalation Studies on the Role of Genetic and Environmental Factors. *Immunology* **1987**, 60, 97–102.
- Kinnula, V.; Cantell, K.; Mattson, K. Effect of Inhaled Natural Interferon-Alpha on Diffuse Bronchioalveolar Carcinoma. *European Journal of Cancer* **1990**, 26 (6), 740–741.
- Martin, R.J.; Boguniewicz, M.; Henson, J.E.; Celniker, A.C.; Williams, M.; Giorno, R.C.; Leung, D.Y. The Effects of Inhaled Interferon Gamma in Normal Human Airways. *American Review of Respiratory Disease* **1993**, 148 (6 pt 1), 1677–1682.
- Pauluhn, J., Martins, T. *Effects and Fate of Aerosolized Alpha-1-Antitrypsin (alpha 1-AT) in Rats Following Subacute Inhalation Exposure*; Abstracts of Fourth International Aerosol Conference Los Angeles, CA, Aug. 24–Sept. 2, 1994.
- Van Zandwijk, N.E.; Jassem, R.; Dubbelmann, E.; Braat, M.C.; Rumke, P. Aerosol Application of Interferon-Alpha in the Treatment of Bronchioalveolar Carcinoma. *European Journal of Cancer* **1990**, 26 (6), 738–740.
- Wolff, R.K.; Dorato, M.A. Toxicological Testing of Inhaled Pharmaceutical Aerosols. *Critical Reviews in Toxicology* **1993**, 23 (4), 343–369.

DRUG DELIVERY—PARENTERAL ROUTE

Michael J. Akers

Cook Pharmaceutical Solutions, Bloomington, Indiana, U.S.A.

INTRODUCTION

The *United States Pharmacopoeia* 24 (1) defines a small-volume injectable (SVI) as “an injection that is packaged in containers labeled as containing 100 ml or less.” Therefore, all sterile products packaged in vials, ampuls, syringes, cartridges, bottles, or any other container that is 100 ml or less fall under this classification. Ophthalmic products packaged in squeezable plastic containers, although topically applied to the eye rather than administered by injection, also fall under the classification of small-volume injections as long as the container size is 100 ml or less. (See the article Ocular Drug Formulation and Delivery in this encyclopedia). Large-volume injectables (LVI) have to be terminally sterilized, whereas SVIs can be sterilized terminally or by aseptic filtration and processing. In fact, 80% or greater of all SVIs commercially available are prepared by aseptic processing. LVIs usually involve intravenous infusion, dialysis, or irrigation fluids containing electrolytes, sugar, amino acids, blood, blood products, and fatty lipid emulsions (2). LVIs must be administered by intravenous administration. Small-volume injections may be injected by intravenous, subcutaneous, or intramuscular routes (primary routes of parenteral administration) or by various secondary routes such as intra-abdominal, intra-arterial, intra-articular, intracardiac, intracisternal, intradermal, intraocular, intrapleural, intrathecal, intrauterine, or intraventricular injections.

SVI formulations are relatively simple, composed of the active ingredient, a solvent system (preferably aqueous), a minimal number of excipients present for reasons described later in this chapter, and the appropriate container and closure packaging system. If the active ingredient is unstable in solution or suspension, the product can be a dry powder, processed either by lyophilization or by sterile crystallization.

This chapter introduces the basic aspects of small-volume injectable products—their use, types and primary characteristics of dosage forms, formulation ingredients, and packaging systems. Additional information is available in a variety of reference texts and book chapters (2–7). Only conventional SVI formulations are

addressed in this chapter. Advanced, long-acting (depot) formulations are not covered.

PRIMARY USES OF SMALL-VOLUME INJECTABLES

Small-volume injectables can be therapeutic injections, ophthalmics, diagnostics, radiopharmaceuticals, or allergenic extracts. The active ingredients can be intended for human or animal therapy and can be small molecules, proteins and other large molecules, biologics, vaccines, monoclonal antibodies, antisense oligonucleotides, and, in the future, genes.

Therapeutic Injections

Injections include a wide variety of therapeutic agents, e.g., for the treatment of cancer, infection, cardiovascular disease, arthritis and other inflammatory diseases, diabetes, hormonal deficiencies, central nervous system problems, and many other disease states. There are more than 400 injection products listed in the USP and, because of the huge number of biotechnology molecules in clinical study, this number will continue to grow rapidly over the next several years. Injections are primarily solutions containing the active ingredient and other substances. Product solutions are available either as “ready-to-use” (e.g., amobarbital sodium for injection) or solutions after reconstituting lyophilized (e.g., Gemzar[®]) or crystallized dry powder products (including many injectable cephalosporins). Some solutions may contain only the drug, e.g., vancomycin hydrochloride solution after reconstitution. Some products are suspensions in which the drug is suspended in a suitable medium, again either commercially available as a ready-to-use suspension (e.g., Humulin[®] N) or reconstituted as a suspension rather than as a solution (e.g., amoxicillin for injectable suspension). Injections can also be commercially available as concentrated liquids (e.g., potassium chloride for injection concentrate) that must be diluted before administration. Injectable products are either single dose or multiple dose. Multiple-dose

injections must contain an antimicrobial preservative agent(s), and the volume of injection should not exceed 30 ml (8).

Ophthalmic Products (9)

Ophthalmic drug products include drugs in solution, suspension, gel, or ointment, administered topically to the corneal surface of the eye. Ophthalmic products also include irrigating solutions in LVI sizes. There are many different types of ophthalmic drug products to treat glaucoma, infection, inflammation, and other diseases of the eye. Ophthalmic products must be sterile, but because they are topically applied, they are not required to be pyrogen-free. Ophthalmic solutions and suspensions are usually packaged in squeezeable low-density polyethylene containers for easy administration. Ophthalmic ointments are also sterile and must be free from metallic particles; they are packaged in ointment tubes. Because ophthalmic products are multiple-dose products, they must contain antimicrobial preservative agents. Because of plastic packaging, most ophthalmic products are aseptically processed.

Diagnostic Agents Including Diagnostic Radiopharmaceuticals (10)

There are many SVI diagnostic agents available including solutions containing contrast media and solutions containing radioactive iodine, chromium, technetium, iron, and other radioactive elements. These products are used primarily to evaluate organ functions. Contrast media solutions are stable in solution and, in fact, can be terminally sterilized. Most radioactive agents are produced to be used within hours of preparation because of the very short half-lives of the radioactive element. As with other sterile dosage forms, diagnostic agent products are to be sterile, pyrogen-free, and particulate-free.

Allergenic Extracts (11)

Allergenic extracts are sterile concentrates (solutions or suspensions) of the substances (allergens) responsible for unusual sensitivities in humans. These products can be used for therapeutic or diagnostic purposes. Extracts are aqueous (0.9% sodium chloride used as the diluent) or glycerinated (50% glycerin as the diluent). Most preparations are buffered at pH 8 and contain phenol ($\leq 0.4\%$) as an antimicrobial preservative. They are sterilized by aseptic filtration.

FORMULATIONS

Small-volume injectables are usually considered small-volume solutions in vials or ampuls but are available in a variety of dosage forms and packaging systems.

Liquids

Small-volume injectable liquids are primarily aqueous solutions. However, because many important therapeutic agents are poorly soluble or totally insoluble in water, oily solvents and water-miscible cosolvents are used to produce ready-to-use solutions.

Aqueous solutions

Aqueous ready-to-use SVIs contain the active ingredient, additional substances, if necessary, and water as the solvent. Water-for-injection (WFI), USP, is the solvent of choice for aqueous SVIs. WFI is prepared by distillation or reverse osmosis techniques. Of all the USP types of water (Table 1), WFI is the purest form of water available for sterile products. An essential requirement of WFI is its freedom from pyrogenic contamination. WFI and other USP types of water are now required to pass a certain specification for endotoxin concentration (see Table 1). Endotoxins are pyrogens with pyrogens being metabolic byproducts of microbial growth and death that cannot be destroyed by autoclaving or by sterilizing membrane filters. Aqueous SVI solutions are prepared either by filling the product into containers, sealing, and terminally sterilizing the finished product or, for drugs that cannot physically or chemically withstand high temperatures or radiation doses required for terminal sterilization, the drug product is sterile-filtered and aseptically filled into the final container and the container sealed by aseptic processing.

Nonaqueous solutions

Several SVIs are marketed as oily solutions (Table 2). The oil must be of vegetable origin (sesame, olive, or cottonseed oils are most commonly used) because of safety, purity, and biocompatibility considerations. Oils for injection must meet USP requirements (12):

1. Solid paraffin test (measurement of oil clarity);
2. Saponification value between 185 and 200;
3. Iodine value between 79 and 128; and
4. Test for unsaponifiable matter and free fatty acids.

Oily solutions are prepared by separately sterilizing the solvent, usually using dry heat, and the drug (dry heat or a gas such as ethylene oxide), then combining the solvent and drug aseptically. Terminal sterilization cannot be used for

Table 1 Types of water described in the *United States Pharmacopeia*

Type	Preparation	Pryogen-Free	Comments
Purified water USP	Distillation or ion exchange	No	Pharmaceutical solvent
Water for injection USP (WFI)	Distillation or reverse osmosis	Yes ^a	Not sterile. Must be used within 24 h or stored below 5°C or ≥80°C; used for manuf. of parenteral products to be sterilized
Sterile water for injection USP	Distillation or reverse osmosis	Yes ^a	Same as WFI; single-dose containers; also used to reconstitute sterile solids and dilute sterile solutions
Bacteriostatic water for injection USP	Distillation or reverse osmosis	Yes ^a	Multiple and single dose
Sterile water for irrigation USP	Distillation or reverse osmosis	Yes ^a	1 L or larger, wide mouth, does not meet particulate matter requirements for LVI; labeled “For Irrigation Only”

^a ≤0.25 endotoxin units per mL.

oily solutions because of the lack of moisture in the product necessary to generate saturated steam under pressure required to destroy microbial life. Practical development experiences with oily injection formulations have been reported by Sims and Worthington (13) and Radd et al. (14).

Cosolvent

A fairly large number of SVIs contain cosolvent systems; a partial listing of commercial products is given in Table 3.

Table 2 Small-volume parenteral products containing oil(s) as the solvent system

Product, USP XXII	Oil
Ampicillin (suspension)	Vegetable
Desoxycorticosterone acetate	Sesame
Diethylstilbestrol	Sesame, cottonseed
Dimercaprol (suspension)	Peanut
Epinephrine (suspension)	Sesame
Estradiol benzoate	Sesame
Estradiol cypionate	Cottonseed
Estradiol valerate	Sesame
Estrone	Sesame
Ethiodized iodine	Poppyseed
Fluphenazine enanthate	Sesame
Hydroxyprogesterone caproate	Sesame
Menadione	Sesame
Nandrolone decanoate	Sesame
Penicillin G procaine (suspension)	Vegetable
Propylidone (suspension)	Peanut
Testosterone cypionate	Cottonseed
Testosterone enanthate	Sesame
Testosterone propionate	Sesame

(From Ref. 4.)

Cosolvents are used to increase the solubility of the poorly soluble drug in water. Cosolvents also tend to minimize or even prevent drug chemical degradation by hydrolysis, obviously because of the reduction in the percentage of water in the system. Water-miscible cosolvents operate on the principle of lowering the dielectric constant property of water, thereby increasing the aqueous solubility of poorly water-soluble drugs. Depending on drug stability, products containing cosolvents can be sterilized terminally using saturated steam under pressure. Otherwise, such products are prepared by aseptic processing. A primary concern in using cosolvents in injectable formulations is their potential to cause lysis of red blood cells when administered intravenously (15). Therefore, any addition of a cosolvent to a formulation intended for parenteral administration must be studied for its safety and potential toxicological effects.

Solids

SVIs are available as sterile dry solids that must be reconstituted with a diluent, usually sterile water for injection, USP, before being administered as a solution or suspension. Sterile dry SVIs are prepared using two primary methods.

Freeze-drying

Most commercial sterile dry powders are manufactured by freeze-drying, also called lyophilization. In this process, under strict aseptic conditions, the product is aseptically filtered and filled as a solution. Special slotted rubber closures are inserted partially onto the vials, which are then transferred to a freeze dryer. Freeze-drying involves three primary operations:

Table 3 Small-volume parenteral products containing cosolvents

Trade Name	Manufacturer	Cosolvent composition
Dramamine	Searle	50% propylene glycol
Apresoline	Ciba	10% propylene glycol
MVI	US Vitamins	30% propylene glycol
Nembutal	Abbott	40% propylene glycol, 10% ethanol
Luminal	Winthrop	67.8% propylene glycol
Dilantin	Parke-Davis	40% propylene glycol, 10% ethanol
DHE 45	Sandoz	15% glycerin, 6.1% ethanol
Cedilanid	Sandoz	15% glycerin, 9.8% ethanol
Robaxim	Robbins	50% polyethylene glycol
Serpasil	Ciba	50% polyethylene glycol, 10% dimethylamine
Ativan	Wyeth	20% polyethylene glycol, 80% propylene glycol
Librium	Roche	20% propylene glycol
Valium	Roche	40% propylene glycol, 10% ethanol
Lanoxin	Burroughs Wellcome	40% propylene glycol, 10% ethanol

From Yalkowsky, S.; Roseman, T., *Techniques of Solubilization of Drugs*; Marcel Dekker, Inc.: New York, 1981; 91.

1. Freezing the product below its eutectic temperature (for crystalline materials) or below its glass transition temperature (for amorphous materials);
2. Primary drying in which the frozen solvent is sublimed, a phase transition from a solid directly to a gas; and
3. Secondary drying in which solute bound water is removed to an acceptable product moisture level for long-term stability.

At the completion of the freeze-dry cycle, the partially inserted rubber closures are fully seated in the vials. The finished product contains a white or off-white sterile dry powder.

Freeze-drying operations are used because of limited stability of certain drugs in solution. Most therapeutic proteins are unstable in solution and can only be commercial products if they are freeze-dried. Freeze-dried formulations usually contain bulking agents (e.g., mannitol) that provide an esthetic dry solid matrix and can help in stabilizing the drug in the solid state. Other excipients are added to the freeze-dried formulation for various reasons, primarily to aid in product chemical and/or physical stabilization (e.g., buffers, antioxidants, cryoprotectants). Freeze-dried vials are usually stable for at least 2 years at ambient conditions, except for some protein products that might need to be refrigerated and have a shorter shelf life. Once the freeze-dried product is reconstituted, normal shelf-life storage conditions are 24 h at room temperature and up to 1 week under refrigeration. Excellent freeze-drying science and technology reviews are available (16, 17).

Powder-filled SVIs

Many SVI antibiotics, particularly the injectable cephalosporins, as well as other molecules are manufactured by sterile crystallization of the active ingredient and aseptically filling the sterile powder into the final container. The drug is dissolved in an appropriate solvent, then filtered through a 0.2- μ m membrane filter. Several techniques can be used for sterile crystallization, including adding sterile seed crystals and adjusting the pH level or adding a sterile antisolvent in which the drug is insoluble. The resultant slurry is collected on a filter system (e.g., the Buchner funnel) and dried, then the dried crystals are milled and blended. Obviously, for this approach to work, the drug must be able to be crystallized. Several variables are critical in controlling the purity and quality of the final crystals, including temperature, rate of addition of solvent, adjustment of pH level, mixing rate and time, and the quality of the seed crystals. Sterile crystallization followed by powder-filling is much more economical than freeze-drying. However, sterile powder-filling offers greater challenges with respect to process variability, microbial and particulate contamination, and operator sensitivity.

Suspensions (18)

With sterile suspensions, the active drug ingredient is suspended in a liquid carrier before administration. Commercial suspensions are either ready to use or dry powders reconstituted as suspensions. Drugs

are formulated as suspension dosage forms for one of two reasons:

- poor solubility in aqueous solution but the product does not need to be administered iv, and
- the need for a long-acting depot injection.

Suspension products (Table 4) are prepared by combining sterile vehicle and sterile drug powder aseptically or by combining two sterile solutions, with the drug solution precipitating in the diluent solution.

Major concerns with the suspension dosage form are:

1. Resuspendability of the drug in the vehicle to permit homogeneous filling of the product into the container and to provide homogeneous dosing when withdrawing from the container;
2. Caking or settling of the drug, resulting in a physically unstable product; and
3. Syringeability (the ability to withdraw a homogeneous dose from the vial into a syringe) and injectability (the ability to eject the product through the needle into the patient).

Formulation ingredients include the suspending agent, a wetting agent (if the suspending agent does not also serve this purpose), a buffer, and an antimicrobial preservative for multiple-dose products.

Emulsions (19)

Emulsions are mixtures of oil- and water-based vehicles with an appropriate surface-active agent to facilitate and maintain the miscibility of the oil-in-water phase. Diprivan® (propofol, a local anesthetic agent) is a primary example of an SVI emulsion. The formulation contains soybean oil, glycerol, and egg lecithin.

BASIC CHARACTERISTICS OF SVIs

SVIs must be sterile and free from pyrogens and foreign particulate matter. These three major characteristics distinguish sterile dosage forms from any other pharmaceutical product.

Sterility (20)

Sterility is a state of absolute freedom from microbial contamination. Interestingly, the word *sterile* on the label of a sterile product has had a historic meaning that a sample of the product lot passed the compendial test for sterility (21). Today, to claim that a product is sterile involves much more than passing a sterility test. Achievement of sterility involves the combination and coordination of a wide range of activities and processes such as:

Cleaning and sanitization of all facilities and equipment

Cleaning and sterilization of equipment, packaging, and all other items to be in contact with the sterile product

Installation and certification of laminar air flow areas where sterile air is provided via high-efficiency particulate air (HEPA) filters

Environmental monitoring of the facility, equipment, water, and personnel for strict microbiological and particulate control

Appropriate gowning and training of personnel in aseptic techniques

Table 4 Small-volume parenteral suspensions

Product	Manufacturer	Suspending agent
Aristocort	Lederle	Propylene glycol 4000
Bicillin C-R	Wyeth	Lecithin, carboxymethylcellulose
Decadron-LA	Merck	Carboxymethylcellulose
Depo-medrol	Upjohn	PEG 3350 (polyethylene glycol)
Duracillin	Lilly	Procaine salt
Hydeltra-TBA	Merck	Tebutate salt
Lente insulins	Lilly, Novo	Polymorphic
NPH insulins	Lilly, Novo	Protamine
PZI insulins	Lilly, Novo	Protamine, zinc
Prolixin decanoate	Princeton	Decanoate salt

Validation of sterilization processes

Validation of the filter system

Integrity testing of the filter system before and after filtration

Integrity testing of the container-closure system to maintain sterility of the product

Conductance of the sterility test initially for all lots and at the end of the shelf-life expiration dating period for the product lot under stability testing

The end-product sterility test suffers from at least three serious limitations that minimize its dependability as a sole indicator of the sterility of a lot of product.

1. Concern that the small sample (usually 20 containers per lot) truly represents the entire lot. Probability statistics reveal that with such a small sample size, the extent of contamination must be significant (on the order of at least 1% of the lot) for the sample to fail the sterility test.
2. Concern that the culture test media used for sterility testing can support the growth of low to high levels of any microbial life possibly contaminating the product.
3. Concern that no accidental contamination was introduced during the performance of the sterility test. There is a finite probability that personnel, testing environment, and/or testing materials may introduce contamination, resulting in a false-positive sterility test result. This concern has been alleviated to a great degree by the advent and successful application of barrier isolator technology systems. Such systems remove direct human contact with the sterile product samples and provide a testing environment that is validated as truly sterile.

Freedom from Pyrogens

Pyrogens are metabolic byproducts of microbial growth. Injected in sufficient amounts in humans (in fact, in any mammal), pyrogens can react with the hypothalamus of the brain to raise the body temperature. In addition, they can cause a number of other adverse physiological effects, including death. The serious problems with sepsis are a result of high levels of endotoxins, endotoxins being a major type of pyrogen. Pyrogens are very small, water-soluble, heat-resistant lipopolysaccharides that cannot be destroyed by typical steam-sterilization cycles or removed by 0.2- μ m membrane

filters. Prevention rather than elimination is the key for pyrogen removal. The primary source of pyrogenic contamination in parenteral products is water. Fortunately, pyrogens are destroyed by distillation. Water used to clean containers and closures can also be a source of pyrogens. However, glass is sterilized by dry heat at temperatures hot enough (usually $>250^{\circ}\text{C}$ to destroy pyrogens). Rubber closures are steam-sterilized, which does not destroy pyrogens. Closures are depyrogenated by the cleaning and rinsing process using pyrogen-free water. Chemical raw materials used in parenteral formulations must be crystallized using pyrogen-free water or other solvents. Some raw materials, e.g., sucrose, mannitol, amino acids, etc. must now be tested by incoming quality control for the presence of endotoxins. If the parenteral product is contaminated with pyrogens, there is no practical way to remove or destroy them. Ultrafiltration (nanometer; nominal molecular-weight filters) will depyrogenate and is used in bioprocessing for separating the smallest unit of lipopolysaccharide from therapeutic proteins. However, ultrafiltration is not a practical pyrogen-removal process for commercial processing of parenteral products.

Pyrogenic contamination is detected using two tests. In the older method, rabbits are injected with product samples, and rectal temperature is measured. Compensatory limits are established with respect to how much temperature increase is permitted before the product is judged to be free or contaminated with pyrogens. The newer method involves a relatively simple in vitro technique called the Limulus Amebocyte Lysate (LAL) test. It is based on the high sensitivity of amebocytes of the horseshoe crab (*Limulus*) to the lipopolysaccharide component of endotoxins originating from Gram-negative bacteria. The LAL test is now the USP method of choice with endotoxin limits established for most SVIs (22).

Freedom from Particulate Matter

Particulate matter is viewed as unacceptable contamination in parenteral solutions. It is recognized that subvisible particulate matter will exist in certain amounts, but the USP now has limits for acceptable levels of particulate matter for SVIs (no more than 6000 particles per container $\geq 0.5\ \mu\text{m}$; no more than 600 particles per container $\geq 25\ \mu\text{m}$). The USP is the only compendium in the world that contains limits for subvisible particulates in SVIs. All worldwide compendia have subvisible particle limits (particles per milliliter) for large-volume injections. SVI solutions with

visible particulate matter should not be used. Particulate matter creates problems in product quality and clinical safety. The primary sources of particulate matter are the container-closure systems and personnel.

Stability

Drugs in SVIs are generally unstable. Many drugs are so unstable that they cannot be marketed as ready-to-use solutions. Drugs with sufficient solution stability will still require certain formulation, packaging, and storage conditions to maintain stability during shelf-life storage and use. The primary pathways of drug degradation involve oxidation (reaction with molecular oxygen catalyzed by various factors including high temperature, high pH level, heavy metals, light, and peroxide contaminants) and hydrolysis (reaction with water catalyzed by high temperature and extremes in pH). For protein pharmaceuticals, aggregation of the protein, resulting in a loss of potency, can be a major degradation pathway. Drugs can also react with packaging and formulation components, resulting in physical and chemical degradation.

Oxidation involves the reaction of free radicals with molecular oxygen so the combination of functional groups that can easily form free radicals, e.g., phenolic or sulfhydryl groups, catalysts (see above), and molecular oxygen, will cause a propagation of the self-oxidation process. Many SVI products are packaged in light-protective packaging, require storage at controlled room or lower (refrigeration) temperatures, are formulated at low pH, contain antioxidants and/or metal chelating agents, and are processed in “oxygen-free” conditions where water is saturated with an inert gas, and, before to sealing the container, the product is overlaid with an inert gas to remove oxygen from the headspace of the container.

Many drugs in liquid SVIs will react with water and form hydrolytic degradation products. Hydrolysis and decomposition occur as solution pH may change and are catalyzed by resulting hydrogen and/or hydroxyl ions. Buffers play an important role in certain injectable products to achieve tight control of solution pH. Hydrolysis of solid-state injectables can occur with moisture from the headspace in the container, moisture remaining in the solid product, and/or moisture originating from or through the rubber closure. Control of residual moisture during and after processing and the use of effective container-closure systems to minimize moisture ingress are very important to protect dried powders from hydrolytic degradation.

Isotonicity

SVIs should be isotonic with blood, tears, spinal fluid, and other biological fluids into which the product is injected or instilled. This means that the injected or instilled solution contains the same “number” of solute “particles” in solution as is contained in the biological cell. Isotonicity means that the “tone” of the cell will not be disturbed, either by the ingress of water from the injected solution (if the solution is hypotonic) or egress of water from the cell (if the solution is hypertonic). Solution tonicity can be ascertained by measurement of a colligative property such as osmotic pressure or freezing-point depression. Biological cells are semipermeable membranes, meaning that they allow the passage of water (and some solutes such as boric acid) but do not allow passage of most solutes. Thus, for example, if a hypotonic solution is injected or instilled, there are fewer solute “particles” in the solution than there are in the cell, forcing water from the injected solution to pass through the cell membrane in an attempt to equalize pressure on both sides of the cell membrane. Increasing the water level of the cell may lead to the cell bursting, which, for red blood cells, is a phenomenon called hemolysis. Hypertonic solutions administered cause the opposite effect, whereby water from the cells permeates the membrane to equalize pressure, and the cells shrink (crenation). In either case, cellular damage can occur causing pain and tissue irritation or damage. Blood, muscle, and subcutaneous cells can withstand a fairly wide range of osmotic pressures from injected solutions (e.g., 250–350 mOsm/kg), whereas tear and spinal fluid cells are much more sensitive to slight differences in the osmotic pressure of injected or instilled solutions. In practice, wide osmolality ranges of SVIs can be tolerated when injected except for injections in cerebrospinal fluid (intrathecal, intraspinal, intracisternal injections). However, it is also true that for all injections, achieving isotonicity should be a goal of the product formulation scientist.

FORMULATION INGREDIENTS

SVIs are simple formulations compared with other pharmaceutical dosage forms. Solution SVIs contain water, the active ingredient, and a minimal number of inactive added ingredients. Solid SVIs contain the active ingredient and usually one or two added ingredients. Formulation scientists have severe restrictions in number and choice of added substances because of safety considerations.

Solvent

The most widely used solvent for SVIs is water for injection (WFI), USP. As a solvent, WFI is used in preparing the bulk solution (compounding) and as a final rinse for equipment and packaging preparation. WFI is prepared by distillation or reverse osmosis, although only distillation is permitted for sterile water for injection, USP. Sterile water for injection is used as a vehicle for reconstitution of sterile solid products before administration and is terminally sterilized by autoclaving. Bacteriostatic water for injection, USP, is commercially available as a reconstitution vehicle for solid products intended for multiple-dose use. Benzyl alcohol is a common antimicrobial preservative used in bacteriostatic water for injection.

Sesame oil, cottonseed oil, and other vegetable oils are used as vehicles for water-insoluble drugs such as corticosteroids and oil-soluble vitamins. Oily solutions can be administered only by intramuscular injection.

Solubilizers

Solubilizers are used to enhance and maintain the aqueous solubility of poorly water-soluble drugs (23–27). Examples of solubilizing agents used in sterile products include:

1. *Liquid cosolvents*: glycerin, polyethylene glycol (300, 400, 3350), propylene alcohol, and ethanol, Cremophor EL, sorbitol
2. *Surface active agents*: polysorbate 80 (polyoxyethylene sorbitan monooleate), polysorbate 20, Pluronic 68, lecithin
3. *Complexing agents*: β -Cyclodextrins, Captisol[®], polyvinylpyrrolidone, carboxymethylcellulose sodium

Liquid solubilizers act by reducing the dielectric constant properties of the solvent system, thereby reducing the electrical conductance capabilities of the solvent and increasing the solubility of hydrophobic or nonpolar drugs. Lanoxin[®], Valium[®], and Nembutal[®] are examples of commercially available sterile solutions containing cosolvent solubilizers. A popular combination consists of 40% propylene glycol and 10% ethanol in water.

Surface active agents increase the dispersability and water solubility of poorly soluble drugs owing to their unique chemical properties of possessing both hydrophilic and hydrophobic functional groups in the same molecule (the same is true of β -Cyclodextrins, addressed below). The hydrophobic groups adsorb to surface molecules of the drug, whereas the hydrophilic groups interact with the

water-solvent molecules. Therefore, the drug molecules locate within the hydrophobic core of the surface active agent (sometimes called a micelle) while the polar molecules of the surface active agent are oriented with water, and the drug is solubilized within the surface active agent dissolved in water.

Solid solubilizers such as the β -Cyclodextrins act by forming soluble inclusion complexes in aqueous solution. These molecules, as with surface active agents, are amphiphilic, meaning that they contain hydrophobic interior functional groups and hydrophilic hydroxy exterior functional groups that enable insoluble drugs to remain in the interior core and be solubilized in water. Brewster, et al. (28) reviewed the application of cyclodextrins in parenteral formulations, particularly for the solubilization and stabilization of proteins and peptides. A relatively new cyclodextrin, Captisol[®], has gained prominence as a safe and effective solubilizer and stabilizer (29). It is an anionic β -Cyclodextrin with a sulfobutyl ether substituent.

Antimicrobial Preservative Agents

Antimicrobial preservatives serve to maintain the sterility of the product during its shelf life and use. They are required in preparations intended for multiple dosing from the same container because of the finite probability of accidental contamination during repeated use. They also are included, although this is quite controversial, in some single-dose products that are aseptically manufactured to provide additional assurance of product sterility. The combination of antimicrobial preservative agents and adjunctive heat treatment (usually temperatures below 110°C) also is used to increase assurance of sterility for products that cannot be terminally sterilized. Very few antimicrobial preservative agents are acceptable (Table 5), with this list decreasing as agents such as thimerosal (and other mercury-containing preservatives) and chlorobutanol are no longer being used. Most substances with antimicrobial activity are irritating and toxic at relatively low concentrations and usually have stability limitations (hydrolytic or oxidative degradation). They can be incompatible with the drug and formulation ingredients and can interact adversely with packaging components. Most commonly used parenteral antimicrobial preservatives are alcoholic or phenolic chemicals. These are highly toxic even at low concentrations and easily oxidizable, and their volatility can cause problems with rubber closure permeation. Formulation scientists must also be aware of significant differences comparing USP and EP requirements for preservative efficacy. Basically, the USP

Table 5 Antimicrobial preservative agents in Small-volume parenterals

Agent	Concentration range (%)	Products
Phenol	0.065–0.5	Humulin N, Zantac, Tensilon, Tagamet, Phenergan, Imferon
<i>m</i> -Cresol	0.16–0.3	Humulin N, Humulin R, Humatrope, Demerol
Methylparaben	0.05–0.18	Decadron, Elavil, Prostigmin
Propylparaben	0.011–0.035	Garamycin, Prolixin, Bicillin
Chlorobutanol	0.5–0.55	Epitrate, Bentyt, Dopram
Benzyl alcohol	0.75–2.0	Valium, Protopin, Geopen, Compazine, Pronestyl, Cleocin
Benzalkonium chloride	0.01–0.025	Most ophthalmic products
Thimerosal	0.0075–0.01	Neosporin, Rhogam, Wydase

requires a bacteriostatic preservative system, whereas the EP requires a bacteriocidal preservative system. For example, whereas the USP requires a 1-log reduction 7 days after a bacterial challenge population is added to the product containing the antimicrobial preservative, the EP Criteria A requirement is a 3-log reduction in bacterial population after 1 day.

Buffers

Buffers are used to maintain the pH level of a solution in the range that provides either maximum stability of the drug against hydrolytic degradation or maximum or optimal solubility of the drug in solution. The most common buffer systems used in SVIs are listed in Table 6. Buffers are composed of simple weak acids and their corresponding salt forms. The appropriate choice of buffer depends on the pH range in which the drug in question is most stable (or most soluble) that matches the pK_a (dissociation constant) of the buffer species. For example, if a pH of 4.5 is most desirable, the correct choice of buffer would be an acetate buffer because the pK_a of acetic acid is 4.76. At pH 4.76, acetic acid exists 50% as the acid (unionized form) and 50% as the salt (ionized form). Sufficient acid and salt species exist at this pH level to compensate for any potential drifts in solution pH and maintain the desired pH level. The concentration of buffer

depends on strength of buffer capacity required to maintain the pH level within the desired range. Obviously, the higher the concentration, the greater the buffer capacity. However, high buffer concentrations can lead to other problems such as general acid/base catalysis of drug hydrolytic reactions.

Antioxidants (30)

Antioxidants function by reacting preferentially with molecular oxygen and minimizing or terminating the free radical auto-oxidation reaction. Many drugs are sensitive to the presence of oxygen and will degrade very rapidly in the absence of protection. In addition to the use of antioxidants, other precautions must be taken. These include protection from light, heat, heavy metal and peroxide contamination, and excessive exposure to air. Formulating the product at low pH is preferable if the product is stable and soluble at low pH. Common antioxidants are shown in Table 7. The most widely used agent is sodium bisulfite because its oxidation-reduction potential lies in the range at which it does not preferentially oxidize too slowly or too rapidly. Other sulfurous acid salts also are effective antioxidants, as are ascorbic acid and sodium ascorbate. Sometimes, combinations of antioxidants strengthen oxidative drug protection as well as the combination of an antioxidant and a chelating agent. The most common chelating agent used in parenterals is disodium ethylenediaminetetraacetic acid (DSEDTA).

Table 6 Common buffer systems used in small-volume parenteral products

pH	Buffer system	Concentration (%)
3.5–5.7	Acetic acid–acetate	1–2
2.5–6.0	Citric acid–citrate	1–5
6.0–8.2	Phosphoric acid–phosphate	0.8–2
8.2–10.2	Glutamic acid–glutamate	1–2

Protein Stabilizers

Therapeutic proteins and peptides have exploded on the pharmaceutical scene in recent years. There are at least 30 commercial protein products currently marketed and hundreds more in clinical study. Proteins are very reactive

Table 7 Antioxidants commonly used in small-volume parenterals

Antioxidant	Concentration range (%)
Water soluble	
Sulfurous acid salts	
Sodium bisulfite	0.05–1.0
Sodium sulfite	0.01–0.2
Sodium metabisulfite	0.025–0.1
Sodium thiosulfate	0.1–0.5
Sodium formaldehyde sulfoxylate	0.005–0.15
Ascorbic acid isomers	
L- and D-Ascorbic acid	0.02–1.0
Thiol derivatives	
Acetylcysteine	0.1–0.5
Cysteine	0.1–0.5
Thioglycerol	0.1–0.5
Thioglycolic acid	
Thiolactic acid	
Thiourea	0.001–0.05
Dithiothreitol	
Glutathione	
Oil soluble	
Propyl gallate	0.05–0.1
Butylated hydroxyanisole	0.005–0.02
Butylated hydroxytoluene	0.005–0.02
Ascorbyl palmitate	0.01–0.02
Nordihydroguaiaretic acid	0.01–0.05
α -Tocopherol 9	0.05–0.075

with their environment, with such reactions causing protein degradation. In pharmaceutical dosage forms, proteins are potentially quite reactive with water, formulation components, packaging components, and air. Environmental conditions that promote protein degradation include high temperature, pH excursions, light, oxygen, moisture, and mechanical stress. Degradation reactions are both chemical and physical, with physical stabilization often more challenging than chemical stabilization. Proteins easily aggregate under a variety of conditions, particularly at temperature extremes and with excessive mechanical manipulations. Denaturation in the form of aggregation can occur during the freezing and/or drying and subsequent storage of proteins processed by lyophilization. A number of ingredients have been shown to stabilize proteins both in the solution state and in the dry state (31–34). Serum albumin will compete with therapeutic proteins for binding sites in glass and other surfaces and minimizes loss of the protein caused by surface binding. With concern about viral contamination in natural substances like albumin, other competitive

binding agents are being investigated (e.g., hetastarch). A number of different types of substances are used as cryoprotectants and lyoprotectants to minimize protein denaturation during freeze-drying. Primary examples include amino acids (glycine, lysine, glutamine); polyhydric alcohols (sorbitol, glycerol, polyethylene glycol); nonreducing sugars (sucrose, trehalose); and polymers such as polyvinylpyrrolidone, methylcellulose, and dextran. Surface active agents, such as polysorbate 80, polysorbate 20, and poloxamer 188 (Pluronic 68), are widely used to minimize protein aggregation at air/water and water/solid interfaces. Buffers, antioxidants, and chelating agents also are used to stabilize proteins in solution when necessary.

Tonicity Adjusters

A variety of agents are used in sterile products to adjust tonicity. Most common are simple electrolytes such as sodium chloride or other sodium salts and nonelectrolytes such as glycerin and lactose. Tonicity adjusters are usually the last ingredient added to the formulation after other ingredients in the formulation are established and the osmolality of the formulation measured. If the formulation is still hypotonic (i.e., <280 mOsm/kg as measured by a commonly used osmometer instrument), tonicity adjusting agents are added until the formulation is isotonic. If the formulation is hypertonic, the degree of hypertonicity and the intended route of drug administration need to be considered. For intravenous administration, hypertonicity values up to approximately 360 mOsm/kg are not considered harmful. However, for other routes of administration, efforts should be made to make the final product isotonic before administration. This can be accomplished either by reducing concentrations of ingredients, if acceptable, or by diluting the product before administration.

Other Ingredients

Bulking agents are used in freeze-dried preparations to increase the solid content of the “plug” in the container after the sublimation process during the freeze-drying cycle. Bulking agents not only serve to enhance the elegance of the product but also can serve as stabilizers in adsorbing excess moisture during shelf life. Suspending agents keep the drug suspended in the solvent after shaking and allow homogeneous dosing of the suspended drug from the container. Emulsifying agents lower the interfacial tension of an oil and water interface to allow the two immiscible solvents to mix and form a stable emulsion

dosage form. Semisolid agents aid in the dispersibility of the drug in ophthalmic ointments and provide the ointment base. Examples of these different additives are:

1. *Bulking agents*: mannitol, lactose, sucrose, dextran
2. *Suspending agents*: carboxymethylcellulose, methylcellulose, gelatin, sorbitol
3. *Emulsifying agents*: lecithin, polysorbate 80
4. *Ophthalmic ointment bases*: petrolatum

Two comprehensive references are available that list type and concentration of all excipients used in commercial sterile formulations that should be part of every sterile formulation scientist's library (34, 35).

PACKAGING

The packaging system obviously is an integral part of the parenteral product, providing long-term protection and maintenance of physical and chemical stability of the product formulation. Packaging can also be used as a drug delivery tool by providing more convenient delivery of the drug product (e.g., syringes, dual chamber vials) and offering better control of drug dosing (e.g., cartridges). Packaging is a major source of particulate contamination and can contribute to physical and chemical degradation of the product. Packaging constituents can leach into the product or the product can be adsorbed or absorbed. The primary types of packaging systems are glass, rubber, and plastic. Metal tubes for ophthalmic ointments are not addressed here. Primary SVI packaging systems include glass sealed ampuls, rubber closed vials, prefilled syringes, cartridges, and small- and large-volume bottles made either of glass or plastic.

Glass (36)

Glass used for parenteral products is classified as type I, type II, and type III (Table 8). Type I is the highest quality

grade, composed almost exclusively of borosilicate (silicon dioxide), making it chemically resistant to extreme acidic and alkaline conditions. Type I glass, although more expensive, is preferred for most parenteral products. Often, even type I glass must be surface treated with agents such as ammonium sulfate or silica dioxide to remove surface leachates. Type II glass is made of soda-lime glass but is treated with sodium sulfite or sulfide to neutralize surface alkaline oxides. Type III glass is untreated soda-lime glass. Type II glass generally is used for large-volume injectables and for small-volume products if the solution pH level is less than 7.0. Type III glass can only be used for oily solutions and dry powders. The USP (37) and other compendia provide requirements and tests necessary to qualify the different types of glass.

Formulation scientists must be aware that glass can and will leach out various elements such as boron, sodium, potassium, calcium, iron, and magnesium. Glass leachates can affect solution pH and cause precipitation problems if the drug or other formulation component forms insoluble salts when combined with these leachates. Solutions of high pH level are notorious for causing alkali leachates. Quality control of each lot of glass must be consistent to control these potential difficulties with leachates. Glass particulates can also be a problem owing to delamination of the inner surface of the glass.

Amber glass containers can be used for light-sensitive SVIs. The amber color is produced by the addition of iron and manganese oxides to the glass formulation. Oxide leachates can occur and catalyze oxidation reactions.

Rubber (38)

Rubber formulations are used as rubber closures (vials, cartridges); rubber plungers (syringes, cartridges); and other applications (rubber septum in dual chamber

Table 8 Glass used for small-volume parenterals

Type	General description	USP test	Size (ml)	Limits (ml of 0.02 N acid)
I	Highly resistant, borosilicate glass	Powdered glass	All	1.0
II	Treated soda-lime glass	Water attack	100 or less Over 100	0.07 0.2
III	Soda-lime glass	Powdered glass	All	8.5
NP ^a	General-purpose soda-lime glass	Powdered glass	All	15.0

^aFor nonparenteral articles.

Table 9 Autoclavable rubber compounds used in small-volume parenterals

Type	Additives	Water–vapor permeation	Potential reactivity with product
Butyl	Moderate	Low	Moderate
Natural	High	Moderate	High
Neoprene	High	Moderate	High
Polyisoprene	High	Moderate	Moderate
Silicone	Moderate	Very high	Low

vials, rubber septum for needle introduction in administration set tubing). The formulations can be very complex. Not only do they contain the basic rubber polymer, but also they may contain many additives such as plasticizers, fillers, vulcanizing agents, pigments, activators, accelerants, and antioxidants. Many of these additives are not fully characterized for content or purity and can be sources of physical and chemical degradation problems in parenteral products. The formulation scientist must work as closely with the rubber manufacturer as with the glass manufacturer to choose the appropriate rubber formulation having consistent specifications and characteristics to maintain product stability.

The most common rubber polymers used in SVI closures are natural and butyl rubber (Table 9). Silicone and neoprene also are used but less frequently in sterile products. Butyl rubber has great advantages over natural rubber in that butyl rubber requires fewer additives, has low water vapor permeation properties, and has good characteristics with respect to gaseous (e.g., oxygen) permeation and reactivity with the active ingredient.

Problems with rubber materials include leaching of constituents (e.g., zinc) into the product, adsorption of active ingredients or antimicrobial preservatives, and coring of the rubber by repeated insertion of a needle. Coring produces rubber particulates that affect the quality and, potentially, the safety of the product.

Siliconization of rubber closure is a common practice in manufacturing to facilitate movement of the closure through stainless steel equipment on the filling lines. However, silicone is incompatible with hydrophobic drugs. Excessive silicone on rubber can potentiate protein aggregation and cause precipitation problems with certain drugs. Elastomer manufacturers have developed rubber formulations with specially bonded coatings that provide “slippery” rubber surfaces and, thus, do not require the

need to apply silicone for high-speed processing equipment.

Plastic

Plastic packaging has always been important for ophthalmic drug dosage forms and is gaining in popularity for injectable dosage forms. Plastic bottles are used to enable people to apply droplets of medication into the eye. Plastic bottles for LVIs have been used for many years. Plastic vials for SVIs may be a wave of the future. Plastic packaging offers such advantages of cost savings, elimination of the problems caused by breakage of glass, and increase convenience of use. As with other packaging systems, plastic formulations can interact with the product, causing physical and chemical stability problems. Plastic formulations are less complicated than are rubber formulations and tend to have a lower potential for leachability of its constituents. However, plasticizer leachates are well-known with polymers such as polyvinyl chloride containers used for LVI bags and administration devices. The most commonly used plastic polymer for ophthalmic products is low-density polyethylene. For other SVIs, polyolefin formulations are widely used as well as polyvinyl chloride, polypropylene, polyamide (nylon), polycarbonate, and copolymers such as ethylene vinyl acetate.

STORAGE

Proper storage of SVIs is critical for the safety and potency of the active ingredient(s) contained in the packaging system. Long-term stability studies are necessary for the appropriate storage conditions. Stability studies involve storing the product at various temperatures, exposing to light, exposing to various relative humidities, and assessing the effect of mechanical stress

during transportation and handling. Studies on proper storage conditions and appropriate handling are extremely important in this age of global distribution of drug products, particularly products containing temperature- and stress-sensitive biomolecules. In addition to concerns regarding the maintenance of chemical and physical stability of the drug product during distribution and storage, there is also concern that the container-closure system is adequate to maintain sterility and other microbiological quality attributes of the sterile product. Container-closure integrity studies in recent years have taken on greater prominence as regulatory groups such as the FDA require these data in registration approvals and in the routine conducting of stability studies. Excellent references on container-closure integrity methods are available (39, 40).

REFERENCES

1. *United States Pharmacopeia, National Formulary*, 24th Ed., 19th Ed.; The United States Pharmacopoeial Convention, Inc.: Rockville MD, 2000; 1775.
2. Turco, S. *Sterile Dosage Forms*, 4th Ed.; Lea & Febiger: Philadelphia, 1994.
3. Avis, K.E. Sterile Products. *The Theory and Practice of Industrial Pharmacy*, 3rd Ed.; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea & Febiger: Philadelphia, 1986; 639–677.
4. DeLuca, P.P.; Boylan, J.C. Formulation of Small Volume Parenterals. *Pharmaceutical Dosage Forms: Parenteral Medications*, 2nd Ed.; Avis, K.E., Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1992; 1, 173–248.
5. Boylan, J.C.; Fites, A.L.; Nail, S.L. Parenteral Products. *Modern Pharmaceutics*, 3rd Ed.; Banker, G.S., Rhodes, C.T., Eds.; Marcel Dekker, Inc.: New York, 1995.
6. Avis, K.E. Parenteral Preparations. *Remington's Pharmaceutical Sciences*, 18th Ed.; Gennaro, A.R., Ed.; Mack Publishing Company: Easton, PA, 1990; 1545–1569.
7. Harwood, R.J.; Portnoff, J.B.; Sunbery, E.W. The Processing of Small Volume Parenterals and Related Sterile Products. *Pharmaceutical Dosage Forms: Parenteral Medications*, 2nd Ed.; Avis, K.E., Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1993; 2, 1–92.
8. *United States Pharmacopeia, National Formulary*; 24th Ed., 19th Ed.; The United States Pharmacopoeial Convention, Inc.: Rockville, MD, 2000; 1777.
9. Hecht, G.; Roehrs, R.R.; Cooper, E.R.; Hiddemen, J.W.; Van Duzee, B.F. Design and Evaluation of Ophthalmic Pharmaceutical Products. *Modern Pharmaceutics*, 2nd Ed.; Banker, G.S., Rhodes, C.T., Eds.; Marcel Dekker, Inc.: New York, 1989; 539–603.
10. Olsen, L.E. Sterile Diagnostics. *Pharmaceutical Dosage Forms: Parenteral Medications*, 2nd Ed.; Avis, K.E., Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1992; 2, 321–359.
11. Shough, H.R. Allergenic Extracts. *Remington's Pharmaceutical Sciences*, 18th Ed.; Gennaro, A.R., Ed.; Mack Publishing Company: Easton, PA, 1990; 1405–1415.
12. *United States Pharmacopeia, National Formulary*, 24th Ed., 19th Ed.; The United States Pharmacopoeial Convention, Inc.: Rockville, MD, 2000; 1776.
13. Sims, E.E.; Worthington, H.E.C. Formulation Studies on Certain Oily Injection Products. *Int. J. Pharm.* **1985**, *24*, 287–296.
14. Radd, B.L.; Newman, A.C.; Fegeley, B.J.; Chrzanowski, F.; Lichten, J.L.; Walking, W.J. Development of Haloperidol in Oil Injection Formulations. *J. Parenteral Sci. Technol.* **1985**, *39*, 48–53.
15. Reed, R.W.; Yalkowsky, S.H. Lysis of Human Red Blood Cells in the Presence of Various Cosolvents. *J. Parenteral Sci. Technol.* **1985**, *39*, 64–68.
16. Pikal, M.J. Freeze Drying. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1992; 6, 275–303.
17. Nail, S.L.; Gatlin, L.A. Freeze Drying: Principles and Practices. *Pharmaceutical Dosage Forms: Parenteral Medications*, 2nd Ed.; Avis, K.E., Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1992; 2, 163–223.
18. Akers, M.J.; Fites, A.L.; Robison, R.L. Formulation Development of Parenteral Suspensions. *J. Parenteral Sci. Technol.* **1987**, *41*, 88–96.
19. Floyd, A.G.; Jain, S. Injectable Emulsions and Suspensions. *Pharmaceutical Dosage Forms: Dispersed Systems*, 2nd Ed.; Lieberman, H.A., Rieger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1996; 2.
20. Akers, M.J. *Parenteral Quality Control: Sterility, Pyrogen, Particulate, and Package Integrity Testing*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1994.
21. *United States Pharmacopeia, National Formulary* 24th Ed., 19th Ed.; The United States Pharmacopoeial Convention, Inc.: Rockville, MD, 2000; 1818–1823.
22. *United States Pharmacopeia, National Formulary* 24th Ed., 19th Ed.; The United States Pharmacopoeial Convention, Inc.: Rockville, MD, 2000; 1829–1931.
23. Gibson, M.; Denham, A.J.; Taylor, P.M.; Payne, N.I. Development of a Parenteral Formulation of Trimelamol, A Synthetic S-Triazine Carbinolamine-Containing Cytotoxic Agent. *J. Parenteral Sci. Technol.* **1990**, *44*, 306–313.
24. Chien, Y.W. Solubilization of Metronidazole by Water-Miscible Multi-Cosolvents and Water-Soluble Vitamins. *J. Parenteral Sci. Technol.* **1984**, *38*, 32–36.
25. Howard, J.R.; Gould, P.L. The Use of Cosolvents in Parenteral Formulations of Low Solubility Drugs. *Int. J. Pharm.* **1985**, *25*, 359–362.
26. Tarr, B.D.; Yalkowsky, S.H. A New Parenteral Vehicle for the Administration of Some Poorly Water Soluble Anti-Cancer Drugs. *J. Parenteral Sci. Technol.* **1987**, *41*, 31–33.
27. Rajagopalan, N.; Dicken, C.M.; Ravin, L.J.; Sternson, L.A. A Study of the Solubility of Amphotericin B in Nonaqueous

- Solvent Systems. *J. Parenteral Sci. Technol.* **1988**, 42, 97–102.
28. Loftsson, T.; Brewster, M.E. Pharmaceutical Applications of Cyclodextrins. I. Drug Solubilization and Stabilization. *J. Pharm. Sci.* **1996**, 85, 1017–1025.
29. Thompson, D. *Cyclodextrins—Enabling Excipients: Their Present and Future Use in Pharmaceuticals*, CRC Critical Reviews in Therapeutic Drug Carrier Systems, January, 1977.
30. Johnson, D.M.; Gu, L.C. Autoxidation and Antioxidants. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1988; 1, 415–449.
31. Wang, Y.J., Pearlman, R., Eds.; *Stability and Characterization of Protein and Peptide Drugs*; Plenum Publishing Corp.: New York, 1993.
32. Ahern, T.J., Manning, M.C., Eds.; *Stability of Protein Pharmaceuticals*; Plenum Publishing Corp.: New York, 1992.
33. Carpenter, J.F.; Pikal, M.J.; Chang, B.S.; Randolph, T.W. Rational Design of Stable Lyophilized Protein Formulations: Some Practice Advice. *Pharmaceutical Research* **1997**, 14, 969–975.
34. Nema, S.; Washkuhn, R.J.; Brendel, R.J. Excipients and their Use in Injectable Products. *PDA J Pharm. Sci. Technol.* **1997**, 51, 166–171.
35. Powell, M.F.; Nguyen, T.; Baloian, L. Compendium of Excipients for Parenteral Formulations. *PDA J Pharm. Sci. Technol.* **1998**, 52, 238–311.
36. Glass as a Packaging Material for Pharmaceuticals. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Abendroth, R.P., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1993; 7, 79–99.
37. *United States Pharmacopoeia, National Formulary* 24th Ed., 19th Ed.; The United States Pharmacopeial Convention, Inc.: Rockville, MD, 2000; 1930–1932.
38. Avis, K.E.; Smith, E.J. Elastomeric Parenteral Closures. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1992; 7, 73–88.
39. Guazzo, D.M. Container/Closure Integrity. *Parenteral Quality Control: Sterility, Pyrogen, Particulate, and Package Integrity Testing*, 2nd Ed.; Akers, M.J., Ed.; Marcel Dekker, Inc.: New York, 1994; Chapter 4.
40. Pharmaceutical Package Integrity. Technical Report No. 27. *PDA J. Pharm. Sci. Technol.* **1998**, 52.

DRUG DELIVERY—ORAL ROUTE

Brahma N. Singh

Kwon H. Kim

St. John's University, Jamaica, New York, U.S.A.

INTRODUCTION

Oral drug delivery is the most desirable and preferred method of administering therapeutic agents for their systemic effects. In addition, the oral medication is generally considered as the first avenue investigated in the discovery and development of new drug entities and pharmaceutical formulations, mainly because of patient acceptance, convenience in administration, and cost-effective manufacturing process. For many drug substances, conventional immediate-release formulations provide clinically effective therapy while maintaining the required balance of pharmacokinetic and pharmacodynamic profiles with an acceptable level of safety to the patient.

However, the potential for oral dosage form development is sometimes limited for therapeutic agents that are poorly absorbed in the gastrointestinal (GI) tract and unstable to various enzymes, in particular, to proteolytic enzymes, such as peptide and protein drugs. The overall process of oral delivery is frequently impaired by several physiological and pharmaceutical challenges that are associated with the inherent physicochemical nature of the drugs and/or the variability in GI conditions, such as pH, presence of food, transit times, expression of *P*-Glycoprotein (*P*-Gp) and CYP3A, as well as enzymatic activity in the alimentary canal. Manipulation of these problems and challenges is considered an important strategy for improving oral drug delivery, and requires thorough understanding and appropriate integration of physicochemical principles, GI physiology and biochemistry, polymer science, pharmacokinetics, and pharmacodynamics. Over the last 3 decades, much research effort has been made in this area to address various biological and technological issues. Research has opened many novel avenues for the more effective, sustained, or rate-controlled oral delivery of both existing and new therapeutic agents, including peptide and protein drugs emerging from the biotechnology arena.

Furthermore, the oral route offers an attractive approach of drug targeting at the specific sites within GI tract for the treatment of certain pathological conditions, such as gastroesophageal reflux disorder, gastroduodenal ulcers, inflammatory bowel disease, and stomach and

colon cancers. Oral drug delivery systems (DDS) can be classified into three categories: immediate-release (IR) preparations, controlled-release (CR) preparations, and targeted-release preparations. This chapter describes the recent technological advances in oral drug delivery and various physicochemical and biological barriers, and also provides some insight on future strategies to improve oral drug delivery.

ANATOMICAL AND PHYSIOLOGICAL CHARACTERISTICS OF THE GI TRACT

Mechanisms and Pathways of Drug Absorption

Orally administered drugs are mainly absorbed in the small intestine (duodenum, jejunum, and ileum) and in the large intestine (colon); however, other regions, such as buccal cavity, stomach, and rectum, also can be considered potential sites for drug absorption. The various anatomical and physiological characteristics of each segment are briefly described in Table 1.

Once at the surface of intestinal epithelium, a drug can be absorbed across by one or a combination of the following mechanisms: passive transcellular, passive paracellular, and carrier- and receptor-mediated transport systems. Paracellular transport involves the passage of drug molecules through aqueous pores created by epithelial tight junctions, and is the most likely route for polar, hydrophilic drugs since they exhibit poor membrane partitioning. In the human small intestine, the average size of these water-filled pores is approximately 7–9 Å for jejunum and 3–4 Å for ileum (2). In the colon, the estimated pore radii are about 8–9 Å, albeit this value was obtained from rat colon (3). Nevertheless, the extent of paracellular transport is limited as tight junctions comprise only about 0.01% of the total absorptive surface area of the intestine (villi) (4). Transcellular absorption involves the transport of drugs through the intestinal epithelial cells (enterocytes) and requires partitioning of drugs across both the apical and the basolateral membranes. Obviously, this route is mainly limited to the transport of relatively low molecular weight lipophilic drugs. Furthermore, studies in

Table 1 Anatomical and physiological characteristics of human GI tract

Region	Length (m)	Absorbing surface area (m ²)	Absorption pathways	Transit (or residence) time of solids (h)	pH		Enzymes and others	Microorganism (counts/g content)
					Fasted	Fed		
Stomach	0.2	0.1	P, C, A (?)	1–3	1.5–3	2–5	Pepsin, lipases, rennin, HCl, Cathepsin	10 ²
Small intestine	7	12.0		3–5				
Duodenum	0.3	0.1	P, C, A, F, I, E		5.5	5.0	Bile acids, trypsin, α-chymotrypsin and other peptidases, amylase, maltase, proteases, lipases, nucleases	10 ²
Jejunum	3.0	60	P, C, A, F, I, E		6.1	No change	Erepsin, amylase, maltase, lactase, sucrase, peptidases, lipases	10 ⁵
Ileum	4.0	60	P, C, A, F, I, E		7–8	No change	Enteropeptidase (enterokinase) and other peptidases, lipases, nucleases, nucleotidases	10 ⁷
Large intestine	1.5	0.3		4–16		No change		10 ¹¹
Cecum	0.06–0.07	0.05	P, C, A, E		5.7			
Colon	1.35	0.25	P, C, E		8.0		Reductases, esterases, glycosidases, amidases, sulfatase	10 ¹¹
Rectum	0.12				7.0			

P = Passive transport; C = Convective transport; A = Active transport; F = Facilitated transport; I = Ion pair transport; E = Endocytosis.
(From Ref. 1.)

humans have demonstrated that absorption by the transcellular route decreases significantly in the colon (small intestine > ascending colon > transverse colon), which has implications for delayed or sustained release formulations, whereas no such gradient exists for the paracellular route (5).

Carrier-mediated transport involves interaction of the drug with a specific transporter or carrier, in which drug is transferred across the cell membrane or entire cell and then released from the basal surface of the enterocyte into the circulation (6). The process is saturable and utilized by small hydrophilic molecules (7). Drugs that are shown to be transported by this mechanism include β -Lactam antibiotics, cephalosporins, and ACE inhibitors (7).

Receptor-mediated transport involves internalization of an external substance, which may be a ligand for which there is a surface bound receptor, or a receptor which binds to a surface located ligand (8). During this process, a small region of the cell invaginates and pinches off, forming a vesicle. This process, in general, is known as endocytosis and comprises phagocytosis, pinocytosis, receptor-mediated endocytosis (clathrin-mediated), and potocytosis (nonclathrin mediated) (9).

After a drug is absorbed in the GI tract, it can gain access to the systemic circulation via two separate and functionally distinct absorption pathways—portal blood and the intestinal lymphatics. The relative proportion of drug absorbed via these two pathways is largely dictated by physicochemical and metabolic features of the drug, and the characteristics of the formulation (10).

The portal blood represents the major pathway for the majority of orally administered drugs as it has higher capacity to transport both water soluble and poorly water soluble compounds (10). During this process, hydrophilic molecules are carried to the liver via the hepatic portal vein, and then by the hepatic artery gain across to the systemic circulation for subsequent delivery to their sites of action. On the other hand, highly lipophilic drugs ($\log P > 5$) that cross the same epithelial barrier are transported to the intestinal lymphatics, which directly delivers them to the vena cava, thereby bypassing the hepatic first-pass metabolism (10).

Gastrointestinal Motility

The process of GI motility occurs both during fasted and fed states; however, the pattern differs markedly in the two states. In the fasted (interdigestive) state, the pattern is characterized by a series of motor activities known as interdigestive myoelectric cycle or migrating motor complex (MMC), which usually occurs every 80–120 min. Each cycle consists of four consecutive

phases (11, 12). Phase I (basal state) is a quiescent period of about 45–60 min without any contractions, except for rare occasions. Phase II (preburst state) is a period of similar duration (30–45 min) that consists of intermittent peristaltic contractions that gradually increase in intensity and frequency. Phase III (burst state) is a short period that consists of large, intense peristaltic contractions, lasting about 5–15 min. This phase serves to sweep the undigested materials out of the stomach and for this reason, phase III contractions are known as “housekeeper” waves.

As phase III of one cycle reaches the end of the distal ileum (ileocecal junction), phase I of the next cycle begins in the stomach (proximal) or esophagus (lower esophageal sphincter). However, sometimes MMC may originate in the duodenum or jejunum and some MMC may not have action potentials strong enough to traverse through the entire small intestine (12). Phase IV is a brief transitional phase (0–5 min) that occurs between phase III and phase I of two consecutive cycles.

In the fed state, the onset of MMC is delayed. In other words, feeding results into delayed gastric emptying. The duration of this delay is mainly dependent on the size (light or heavy) and composition (fatty or fibrous) of the meals. Consequently, the fate of pharmaceutical dosage forms is mainly subject to the pattern of GI motility in fasted (or fed) state at the time of dosage administration.

PROBLEMS AND BARRIERS TO ORAL DRUG DELIVERY

The biggest problem in oral drug delivery is low and erratic bioavailability, which mainly results from one or more factors such as poor aqueous solubility, slow dissolution rate, low intestinal permeability, instability in GI milieu, high first-pass metabolism through liver and/or intestine variable GI transit, and *P*-gp mediated efflux. This, in turn, may lead to unreproducible clinical response or a therapeutic failure in some cases due to subtherapeutic plasma drug levels. Indeed, the incomplete and variable oral bioavailability will have its most serious impact for drugs with a narrow “therapeutic window” (13) (e.g., theophylline, carbamazepine, quinidine, etc.) From an economic point of view, low oral bioavailability results in the wasting of a large portion of an oral dose, and adds to the cost of drug therapy, especially when the drug is an expensive one (14). It is, therefore, extremely important that these issues be considered and a suitable technique (or an animal model) be used while estimating the contributions from each factor responsible for low and/or variable bioavailability.

Physicochemical Barriers to Oral Drug Delivery

Aqueous solubility

It has long been recognized that before an orally administered drug becomes available for absorption at specific sites within the GI tract, it must be dissolved in the GI fluid. Since both the dissolution rate and the maximum amount of a drug that can be dissolved are dictated by the solubility of the drug in the medium (15), aqueous solubility of a drug could be regarded as a key factor responsible for low oral bioavailability of poorly water-soluble drugs, thereby limiting their therapeutic potential. Other issues related to low bioavailability for a sparingly soluble drug are lack of dose proportionality, substantial food effect, high intra- and intersubject variability, gastric irritancy, and slow onset of action (16). These problems are further exacerbated when attempts are made to develop CR dosage forms (17). Unfortunately, many potent compounds, including new chemical entities (NCEs), possess very low aqueous solubility at physiological pH, which could be attributed to their high inherent lipophilicity incorporated by drug design in order to ensure good absorption (18). Consequently, various approaches are utilized to improve aqueous solubility, which mainly include chemical modification (19), complexation (17, 20, 21), micronization (22), solubilization using surfactants, solid dispersion (23, 24), and design of drug delivery systems. However, there may be specific practical limitations to each of these approaches. For instance, the salt formation approach is not feasible for neutral compounds, and in many cases also for weakly acidic or weakly basic drugs because of the reconversion of salts into aggregates of their corresponding free acid or base forms (23). This precipitation effect would thus lead to slower dissolution rates and may cause failure of a clinical response or epigastric distress due to change in pH (24). Sometimes, the rate of renal excretion may also increase, particularly when amine drugs are solubilized as acid salts (25).

Similarly, the commercial success of solid dispersion has been limited due to use of higher temperatures ($>100^{\circ}\text{C}$) or harmful organic solvents, such as chloroform or dichloromethane, which may result in chemical decomposition of drugs and carriers or possible toxicity from the residual solvent. These unacceptable problems led scientists to explore novel alternatives, particularly the cogrinding method, which involves cogrinding of a poorly water-soluble drug with a water-soluble polymer, such as polyvinyl pyrrolidone, polyethylene glycol (PEG), hydroxypropyl cellulose, hydroxypropylmethyl cellulose, polyvinyl alcohol (26, 27), or sugars such as D-Mannitol (28). These authors suggest that although the composition of the coground mixture is similar to that of solid

dispersions, supersaturation is maintained for a long period due to better dispersion through particle size reduction.

Lipophilicity

The lipid solubility or lipophilicity of drugs has long been recognized as a prerequisite for transcellular diffusion across the intestinal membrane. Traditionally, the lipophilicity of drug substances is expressed as the apparent partition coefficient or distribution coefficient ($\log P$) between *n*-octanol and an aqueous buffer (pH 7.4), which is pH-dependent in the case of ionizable compounds (29). In general, compounds with low $\log P$ are poorly absorbed, whereas compounds with $\log P > -1$ offer satisfactory absorption (29). It is important, however, that the drug possess an optimum lipophilicity, as too low or too high lipophilicity may result in less than optimum oral bioavailability. For example, Mori et al. (30) did not observe any significant improvement in oral absorption of disodium cromoglycate ($\log P < -3$) despite the fact that lipophilicity was increased by derivatization to a lipophilic (diethyl) ester prodrug ($\log P = 1.78$). This observation was attributed to a marked decrease in the water solubility (from 195.3 mg/mL to 0.0052 mg/mL), which occurred simultaneously with increased lipophilicity. Similarly, Wils et al. (31) reported for the first time that high lipophilicity ($\log P > 3.5$) decreases drug transport across the intestinal epithelial cells and could be accounted for loss of in vivo biological activity. The “cut-off” point of P value, that is, the P value corresponding to an optimal transepithelial passage of drugs, was found to be around 3000. Other consequences of high lipophilicity may include passive exsorption (32), and increased plasma protein binding and biliary excretion; the biliary excretion may limit the oral bioavailability of polypeptides.

Aqueous boundary layer

The aqueous boundary layer or the unstirred water layer (UWL) is a more or less stagnant layer, about 30–100 μm in thickness, composed of water, mucus, and glycocalyx adjacent to the intestinal wall that is created by incomplete mixing of the luminal contents near the intestinal mucosal surface (33). The glycocalyx is made up of sulfated mucopolysaccharides, whereas mucus is composed of glycoproteins (mucin), enzymes, and electrolytes. Until recently, the resistance of the UWL to intestinal absorption was believed to be correlated to the effective intestinal permeability (P_{eff}) values of the solutes; however, considerable evidence suggests instead that the available surface of the apical membrane of the intestinal mucosa is the main barrier for both actively and passively absorbed solutes (33). It is also interesting to note that coadministration of food and prokinetic (motility inducing)

agents such as cisapride tends to decrease the thickness of UWL by increasing segmental and propagative contractions respectively, which may have implications for drug dissolution in the GI tract (34). The reverse is true for some viscous soluble dietary fibers, such as pectin, guar gum and sodium carboxymethylcellulose, which may increase the thickness of UWL by reducing intraluminal mixing (35–37) and could possibly decrease the intestinal exsorption of lipophilic drugs like quinidine and thiopental (38).

Biological Barriers to Oral Drug Delivery

Intestinal epithelial barrier

The intestinal epithelial layer that lines the GI tract represents the major physical barrier to oral drug absorption. Structurally, it is made up of a single layer of columnar epithelial cells, primarily enterocytes and intercalated goblet cells (mucus secreting cells) joined at their apical surfaces by tight junctions or zonula occludens. These tight junctions are formed by the interaction of membrane proteins at the contact surfaces between cells and are responsible for restricting the passage of hydrophilic molecules during the paracellular transport. In fact, electrophysiological studies have suggested that epithelium gets tighter as it progresses distally, which has been implicated in a reduced paracellular absorption in the colon (39).

The epithelium is supported underneath by lamina propria and a layer of smooth muscle called muscularis mucosa (3–10 cells thick). These three layers, i.e., the epithelium, lamina propria, and muscularis mucosa, together constitute the intestinal mucosa (40). On the apical surface, the epithelium along with lamina propria projects to form villi. The cell membranes of epithelial cells that comprise the villi contain uniform microvilli, which give the cells a fuzzy border, collectively called a brush border. These structures, although greatly increase the absorptive surface area of the small intestine, provide an additional enzymatic barrier since the intestinal digestive enzymes are contained in the brush border. In addition, on the top of the epithelial layer lies another layer, the UWL, as previously described. The metabolic and biochemical components of the epithelial barrier will be discussed.

Gastrointestinal transit

Advanced CR dosage forms can provide a precise control over release rates or release patterns for most drugs, which is attractive from a clinical point of view, particularly minimization of peak and trough variations in plasma drug concentration, and chronotherapy. For example, the GI

therapeutic system (GITS[®], Alza Corporation) can provide zero-order drug absorption, in which the rate of drug absorption from the GI tract is constant and not determined by the amount of drug available in the GI tract. However, none of the CR systems can control the extent of drug absorption from the GI tract. This is mainly attributed to the fact that the extent of drug absorption from different regions of the GI tract is different, which in fact partly constitutes a basis for the biopharmaceutic classification scheme for drugs administered as extended-release (ER) products (41). Thus, indirectly the extent of absorption is determined by the GI transit time (GITT) of the dosage form rather than its CR properties or delivery program (42). In general, the GITT of most drug products is relatively short (8–12 h), which in turn impedes the formulation of a once daily dosage form (43).

From the oral delivery standpoint, both gastric emptying time (GET) and small intestinal transit time (SITT) are considered important since the majority of drugs are preferentially absorbed in the upper parts of the GI tract (stomach, duodenum, and jejunum). Moreover, the stomach and intestines have a limited site for drug absorption. This is known as an “absorption window.” The relatively short GET (1–3 h) and SITT (3–5 h) thus provide limited time for drug absorption through the major absorption zone. These problems are further aggravated by the highly variable nature of the gastric emptying process, which can vary depending on several physiological factors, such as food, age, posture, exercise, body mass index, circadian rhythm, etc.; pathological factors, such as stress, diabetes, Crohn’s disease, and motility disorders; and pharmaceutical factors, such as size and density of formulation and coadministration of drugs like anticholinergic and prokinetic agents. It is apparent that GI transit is less likely to be a critical determinant of bioavailability for drugs that are completely absorbed from the GI tract and if their dosage forms reside in the GI tract for a considerable period of time (~14–18 h).

Since SITT has been demonstrated to be constant and practically independent of food as well as physical properties of the releasing units, such as size and density (44), the variability in overall GITT is mainly contributed by gastric emptying. The variability in turn may lead to unpredictable plasma drug levels, and could severely impair the performance of pulsed- or time-release devices, which are basically designed to deliver drug after a predetermined time to a specific site of the GI tract. For instance, if there is a large intrasubject variation in transit times, then the use of time-release systems may be precluded due to the unintended delivery of drug to an inappropriate region of the GI tract (45).

Food effect

The coadministration of drugs with food is known to result in decreased, delayed, increased, or accelerated drug absorption, which may have pharmacokinetic and pharmacodynamic implications (46). Most often, the food effect is nonspecific and manifested as an interplay between physiological effects of food (such as delayed gastric emptying, stimulated bile secretion, increased liver blood flow, and alterations in gastric and duodenal pH), physicochemical characteristics of the drug (e.g., water solubility), or its formulation (e.g., size, structural organization, and dissolution profiles). In general, drugs that are most influenced are those that are primarily absorbed from the upper regions of the GI tract and/or are poorly water-soluble. Apparently, food does not have a clinically significant effect on the absorption of moderately soluble drugs having a pH-independent solubility, and those that are completely absorbed (e.g., glipizide, isosorbide-5-mononitrate, felodipine, and nifedipine) from the GI tract. Furthermore, food may indirectly influence the drug absorption by affecting drug release from both hydrophilic matrix as well as lipid matrix formulations. In the former case, the effect has been attributed to increased hydrodynamic mechanical stress, which is caused by increased gastric motility (47). In the latter, the effect is due to increased pancreatic and biliary secretions, which in turn affect the integrity of matrix (48).

Other potential factors that affect oral delivery are pH of the upper GI tract (49), which varies as a function of age, food and certain disease states, circadian rhythm (50, 51), various diseases of the GI tract (49), and drug–drug interactions (7).

Metabolic and Biochemical Barriers to Oral Drug Delivery

Presystemic metabolism

Orally administered drugs are subject to presystemic metabolism, which is comprised of three subtypes of mechanisms:

1. Luminal metabolism. This may be triggered by digestive enzymes secreted from the pancreas (amylase, lipases, and peptidases including trypsin and α -chymotrypsin), and those derived from the bacterial flora of the gut, especially within the lower part of the GI tract.
2. First-pass intestinal metabolism. This includes brush border metabolism and intracellular metabolism. The former occurs at the surface of the enterocytes by the enzymes present within the brush border membrane.

Furthermore, the brush border activity is generally greater in the proximal small intestine (duodenum \approx jejunum $>$ ileum \gg colon) and involves enzymes such as alkaline phosphatase, sucrase, isomaltase, and a considerable number of peptidases (7).

The intracellular metabolism occurs in the cytoplasm of enterocytes and involves the major class of phase I metabolizing enzymes (i.e., cytochrome P450s, in particular CYP3A4), several phase II conjugating enzymes, and others such as esterases. Important examples of intestinal phase II metabolism include sulfation of isoproterenol and terbutaline, and glucuronidation of morphine and propofol (52). It is obvious that intestinal epithelium as a site of preabsorptive metabolism may significantly contribute to the low bioavailability of therapeutic peptides and ester type drugs like aspirin, although it could serve as a key site for targeted delivery of ester or amide prodrugs (52).

3. First-pass hepatic metabolism. As an absorbed drug reaches the liver through the portal circulation, a fraction of the administered dose is biotransformed before it reaches the systemic circulation. This is known as first-pass hepatic metabolism. In comparison with intestine, the liver dominates the process of first-pass metabolism for most drugs by virtue of its large mass, multiplicity of enzyme families present, and its unique anatomical position (13). However, this is not true for some drugs, such as terbutaline, in which case the sulfation occurs predominantly in the small intestine, as well as for midazolam, in which first-pass extraction by intestinal mucosa and liver appears to be comparable ($44 \pm 14\%$ versus $43 \pm 24\%$) (52). It is also important to emphasize here that efficiency of first-pass metabolism (both hepatic and intestinal) varies considerably among different animal species and human subjects, which should be taken into account when deriving the estimates of oral bioavailability, particularly in the case of poorly soluble drugs.

P-Glycoprotein and other efflux systems

In recent years, it has been recognized that *P*-Gp can significantly contribute to the barrier function of the intestinal mucosa. *P*-Gp is an integral membrane protein, about 170–180 kDa, encoded by the *MDR1* gene in humans and contains 12 putative transmembrane domains and two ATP binding sites (53). In the small intestine and colon, *P*-Gp is expressed almost exclusively within the brush border membrane of mature enterocytes, where it

acts as an energy-dependent drug efflux pump. Although *P*-gp appears to be distributed throughout the GI tract, its levels are higher in more distal regions (stomach < jejunum < colon) (54). Moreover, several studies have shown that *P*-Gp and CYP3A4 have similar substrate specificity [reviewed in Ref. (55)]. The colocalization of *P*-Gp and CYP3A4 in the mature enterocytes and their overlapping substrate specificity reasonably suggest that the function of these two proteins may be synergistic and appear to be coordinately regulated (56). Consequently, a greater proportion of the drug will be metabolized since the repetitive two-way kinetics (drug exsorption from the enterocyte into the lumen via *P*-Gp and reabsorption back into the enterocyte) will simply prolong the drug exposure to CYP3A4 (57). This mechanism not only limits the absorption of a wide variety of drugs, including peptides, but also poses a threat for potential drug interactions when attempts are made to inhibit CYP3A4 or *P*-Gp (58). Other mechanisms of drug efflux process may involve organic cation and anion transporters, which have been described elsewhere (32).

CURRENT TECHNOLOGIES IN ORAL DRUG DELIVERY

Over the last 3 decades, many novel oral drug therapeutic systems have been invented along with the appreciable development of drug delivery technology. Although these advanced DDS are manufactured or fabricated in traditional pharmaceutical formulations, such as tablets, capsules, sachets, suspensions, emulsions, and solutions, they are superior to the conventional oral dosage forms in terms of their therapeutic efficacies, toxicities, and stabilities (59). Based on the desired therapeutic objectives, oral DDS may be assorted into three categories: immediate-release preparations, controlled-release preparations, and targeted-release preparations.

Immediate-Release Preparations

These preparations are primarily intended to achieve faster onset of action for drugs such as analgesics, antipyretics, and coronary vasodilators (17). Other advantages include enhanced oral bioavailability through transmucosal delivery and pregastric absorption (60), convenience in drug administration to dysphagic patients, especially the elderly and bedridden, and new business opportunities (61). Conventional IR formulations include

fast disintegrating tablets and granules that use effervescent mixtures, such as sodium carbonate (or sodium bicarbonate) and citric acid (or tartaric acid), and superdisintegrants, such as sodium starch glycolate, croscarmellose sodium, and crospovidone. Current technologies in fast-dispersing dosage forms include modified tableting systems, floss or Shearform technology, which employs application of centrifugal force and controlled temperature, and freeze drying (61). Some of these technologies are briefly described in Table 2.

Controlled-Release Preparations

The currently employed CR technologies for oral drug delivery are diffusion-controlled systems, solvent-activated systems, and chemically controlled systems (63). Diffusion-controlled systems include monolithic and reservoir devices in which diffusion of the drug is the rate-limiting step, respectively, through a polymer matrix or a polymeric membrane. Solvent-activated systems may be either osmotically controlled or controlled by polymer swelling. Chemically controlled systems release drugs via polymeric degradation (surface or bulk matrix erosion) or cleavage of drug from a polymer chain (63). Recent examples of commercial products that have been developed based on these CR principles are provided in Table 3. It is worth mentioning here that the so-called programmed-release (“tailored-release”) profile of a final CR product is rarely the outcome of a single pharmaceutical principle. Depending on the specific physicochemical properties of the drug in question and desired therapeutic objectives, different formulation and CR principles may be proportionally combined within the same dosage form. This task appears to be simpler when realized in terms of appropriate selection of polymers and excipients that incorporate desired principles.

Targeted-Release Preparations

Site-specific oral drug delivery requires spatial placement of a drug delivery device at a desired site within the GI tract. Although it is virtually possible to localize a device within each part of GI tract, the attainment of site-specific delivery in the oral cavity and the rectum is relatively easier than in the stomach and the small and large intestines. The latter requires consideration of both longitudinal and transverse aspects of GI constraints (66). Some of the potential CR and site-specific DDSs will be described.

Table 2 Immediate release preparations for oral delivery

Technology owner and DDS	Brief description	Unique features	Marketed products (drugs)
<i>Elan Corporation, plc, Dublin, Ireland:</i> EFVDAS [®]	Effervescent drug absorption system. A technology that allows development of user friendly formulations which, in turn allows for ease of administration and fast onset of action.	Ability to incorporate high drug loading within the one dosage form, which also facilitates development of combination dosage forms.	Ibuscent [®] (Ibuprofen) as an effervescent formulation
<i>Prographarm Group, Paris, France:</i> Flashtab [®] (62)	A tablet formulation consisting of micronized drug particles coated with Eudragit [®] E	Ability to offer rapid dissolution in saliva, taste masking, and tailored pharmacokinetics for both OTC and prescription drugs.	Tachipima Flashtab [®] and Febrectol [®] (Acetaminophen or paracetamol)
<i>R.P. Scherer Corporation, MI, USA:</i> Zydis [®] System (60)	A freeze-dried, porous tablet that disintegrates instantaneously in the mouth, releasing the drug, which subsequently dissolves or disperses in the saliva.	Enhanced oral bioavailability and improved patient compliance through transmucosal delivery and pregastric absorption.	Maxalt-MLT [®] (Rizatriptan benzoate); Feldene Melt [®] (Piroxicam); Claritin [®] Reditabs TM (Loratadine); Innovace Melt [®] and Renitec RPD [®] (Enalapril); Zofran [®] (Ondansetron); Zelapar [®] (Selegiline); Zydis Apomorphine under clinical trial

Table 3 Controlled-release preparations for oral delivery

Technology owner and DDS	Brief description	Unique features	Marketed products (drugs)
<i>ALZA Pharmaceuticals, CA, USA:</i>			
Elementary osmotic pump	It combines a water-soluble drug and sometimes an osmotic agent in a monolithic core and delivers drug in solution form.	High, predictable delivery rates can be obtained independent of GI motility and the pH of luminal fluids.	Potassium chloride
OROS® Push-Pull System	It has separate osmotic and drug layers and delivers drug in solution or suspension form.	Ability to deliver drugs ranging from very low to high water solubility.	Procardia XL® and Adalat CR® (Nifedipine) as once daily tablets; Glucotrol XL® (Glipizide)
OROS Delayed Push-Pull System	A controlled onset extended release (COER-24™) system.	Ability to synchronize extended release characteristics with the circadian variation, thereby providing a protective effect during the periods of increased risk as well as throughout the dosing interval.	Covera-HS™ (Verapamil)
OROS Tri-Layer Push-Pull System	Multiple layers, enable patterned, pulsatile or delayed delivery of compounds.		
OROS Push-Stick System	Designed to deliver insoluble drugs that require high loading, with an optional delayed, patterned, or pulsatile release profile.		
OROS-CT	Colon-targeted delivery system for local or systemic therapy.	Ability to maintain a constant drug input in the colon for up to 24 h or to deliver drug over an interval as short as 4 h; Oral delivery of peptides and proteins.	Chronset®
OSMET®-CT	A miniature osmotic pump for use as a research tool in clinical studies.	The system can be used to study colonic absorption in humans prior to development of colon-specific delivery systems.	
L-OROS	An osmotic system designed to deliver lipophilic liquid formulations containing insoluble drugs or polypeptides.		
MOTS®	Mucosal oral therapeutic system. A CR osmotic system based on OROS push-pull technology to deliver drugs to the oral cavity for extended time periods.	Both topical and systemic therapy.	Nystatin

(Continued)

Table 3 Controlled-release preparations for oral delivery (*Continued*)

Technology owner and DDS	Brief description	Unique features	Marketed products (drugs)
<i>Duffar BV, Weesp, The Netherlands:</i> Gradient Matrix System (64)	A multiple-unit matrix system in which a positive concentration of drug is applied in the direction of the core of the matrix using a coating technique. In addition, an inert excipient (xylitol) is added with an inverse concentration gradient, thus avoiding the formation of a barrier layer at the outer surface of the device.	Zero order release for highly soluble drugs.	Acetaminophen, Mebeverine HCl.
<i>Elan Corporation, plc, Dublin, Ireland:</i> PHARMAZOME®	A multiparticulate drug delivery technology for producing CR and taste masked preparations such as liquids, suspensions, effervescent and chewable tablets, reconstitutable powders, and unit dose sachet or sprinkle systems.	Release from the PHARMAZOMES can be controlled to deliver the desired profile, i.e., immediate-release, controlled-release and/or delayed release.	Theo-Dur® (Theophylline) as a twice daily syrup and twice daily sprinkle formulation.
SODAS®	Spheroidal oral drug absorption system. A drug delivery technology for producing multiparticulate dosage forms (capsules or tablets) using CR beads.	Customized release rates to suit the individual drug delivery requirements of a drug; controlled absorption irrespective of the feeding state.	Cardizem® SR (Diltiazem HCl) as a twice daily formulation; Cardizem CD (Diltiazem HCl) as a once daily formulation; Verelan® (Verapamil) as a once daily formulation, Morphelan (Morphine) as a once daily formulation (under clinical trial).
<i>Pennwalt Corporation, NY, USA:</i> Pennkinetic™ System	The system combines CR principles of ion exchange with membrane diffusion. The system is made up of an ion-exchange polymer-drug complex as a core material, which is subsequently coated with ethylcellulose to form a water insoluble but permeable coating.	Release rates are relatively constant and unaffected by variable conditions of the GI tract. The system can be formulated in capsule form as well as suspensions.	Delsym® (Dextromethorphan), Corsym® (Chlorpheniramine) and Cold Factor 12® (Phenylpropanolamine).
<i>Skye Pharma AG, Muttens, Switzerland:</i> Geomatrix® Systems (65)	A multi-layer tablet technology, which consists of a hydrophilic matrix core containing the active ingredient, and one or two impermeable or semipermeable polymeric barrier layers applied on one (two-layered system) or both bases of the core (three-layer system) by film-coating or by compression coating.	Constant drug release with one soluble drug or with drugs of different water solubility for 24 h.	Dilacor® XR (Diltiazem HCl) as a once daily formulation; Diclofenac Ratiopharm® Uno (Diclofenac Na); Nifedipine as a once daily formulation.

Gastroretentive systems

Gastroretentive systems (GRDDS) are designed on the basis of delayed gastric emptying and CR principles, and are intended to restrain and localize the drug delivery device in the stomach or within the upper parts of the small intestine until all the drug is released. Various mechanisms (approaches) of achieving gastric retention include floatation or buoyancy (floating systems) (43), mucoadhesion (bioadhesive systems), sedimentation (high-density systems), expansion (swelling and expanding systems), and geometry (modified-shaped systems). Other approaches include coadministration of drugs or fatty acid salts, or sham feeding of indigestible (e.g., polycarbophil) or enzyme-digestible hydrogels (67) that convert the motility pattern of the stomach to a fed state.

Based on the previously published literature, applications of GRDDS may be summarized for several different drugs with various physicochemical and biopharmaceutical properties:

1. Drugs required to exert local therapeutic action in the stomach: misoprostol, 5-Fluorouracil, antacids and antireflux preparations, anti *Helicobacter pylori* agents, and certain enzymes.
2. Drugs exhibiting site-specific absorption in the stomach or upper parts of the small intestine: atenolol, furosemide, levodopa, *p*-Aminobenzoic acid, piretanide, riboflavin-5'-phosphate, salbutamol (albuterol), sotalol, sulpiride, and thiamine.
3. Drugs unstable in lower part of GI tract: captopril.
4. Drugs insoluble in intestinal fluids (acid soluble basic drugs): chlorthalidone, chlorpheniramine, cinnarizine, diazepam, diltiazem, metoprolol, propranolol, quinidine, salbutamol, and verapamil.
5. Drugs with variable bioavailability: sotalol hydrochloride and levodopa

The common objective in each case is to prolong, in a predictable manner, the overall GITT, which appears to be a major determinant of clinical efficacy of CR dosage forms. This can lead to improved bioavailability of drugs, especially those with limited absorption sites in the small intestine (class 2).

While employing this approach for enhancing oral bioavailability of poorly soluble drugs, in which drug absorption is dissolution-rate limited, the following considerations must be taken into account:

1. The optimization of prolonged gastric residence time (GRT) should be done in the light of dissolution time required to achieve complete dissolution in vivo. In other words, if the inherent physicochemical properties of the drug and biological environment of the

stomach allow for the complete dissolution of the required dose during the normal GRT, the gastroretentive approach is unlikely to offer any advantage.

2. The increase in dissolution rate due to the gastric retention does not automatically translate into an increased and less erratic bioavailability. This feature will rather depend on whether the rate and/or extent of drug absorption is limited primarily by dissolution when the GRDDS is administered, as it has been previously suggested for conventional dosage forms (34).

Additionally, since the stomach is one of the desired absorption sites for weakly acidic drugs, there should be no lag time in drug release from the GRDDS. Ideally, this type of delivery system is required for loop diuretics, such as furosemide and piretanide, in which the release of a certain amount of drug in the stomach is desired to obtain more balanced bioavailability (17). Consequently, this property prompted the development of floating systems for these acidic drugs (68, 69).

It is obvious that drugs such as nifedipine and both isosorbide-5-mononitrate and isosorbide dinitrate, which have nonspecific, wide absorption sites and so are well absorbed along the entire GI tract, may not be suitable candidates for GRDDS. Also, drugs that are irritant to the gastric mucosa (43) and those undergoing significant first-pass metabolism may have some limitations (1). Relevant examples of the latter type are nifedipine, propranolol, levodopa, diltiazem, metoprolol and 5-Fluorouracil.

Mucoadhesive systems (MADDS)

The term mucoadhesion is commonly used to describe an interaction between the mucin layer, which lines the entire GI tract, and a bioadhesive polymer, which could be natural or synthetic in origin (70). From the oral delivery standpoint, these systems are used to immobilize and localize a drug delivery device in the selected regions of the GI tract, which could be an oral cavity (buccal and sublingual routes), the esophagus, stomach, small intestine, or colon (oral route). For the most part, research in this area has focused on the design of polymeric micro- and nanoparticulate systems that use hydrophilic polymers, primarily due to their propensity to interact with the mucosal surface (71).

A review of recent studies indicate that mucoadhesion of polymers, especially hydrogels, involves a combination of surface and diffusional phenomena that contribute to the formation of "semipermanent" inter-chain bridges between the polymer and the mucosal

surface (72). The various proposed mechanisms of mucoadhesion include lowering of interfacial tension, interpenetration of the mucoadhesive polymer and mucin, formation of an electrical double layer at the adhesive/mucin interface, and formation of a hydrogen bond and/or van der Waals' forces of attraction (70).

Important advantages of MADDs include:

1. Intimate and prolonged contact of the drug delivery device with the absorbing membrane, which has the potential to maximize both the rate as well as the extent of drug absorption.
2. Prolonged and controlled GI transit of the dosage forms.
3. Site-specific drug delivery, which has the potential for local (topical) therapy of several conditions, such as aphthous stomatitis, gastric ulcer, esophageal cancer, toothache, and dental sores, as well as the potential for the treatment of systemic diseases such as diabetes mellitus and angina pectoris (73).
4. Potential for delivering drugs that undergo considerable first-pass metabolism through the liver (developed as buccal and sublingual MADDs).

Examples of marketed mucoadhesive formulations are Aftach[®] buccal tablets (triamcinolone acetonide), Susadrin[®] sublingual tablets (nitroglycerin; Forest Laboratories, NY), and Buccastem[®] buccal tablets (prochlorperazine maleate; Reckitt and Colman, England).

Enteric-coated systems

Enteric-coated systems (ECS) utilize polymeric coatings that are insoluble in the gastric media and therefore, prevent or retard drug release in the stomach. Various types of ionizable polymers are commercially available. They dissolve at various pH ranging between 4.8 and 7.2 (74). ECSs are generally applicable to four major types of drugs (75):

1. Drugs that are unstable in the gastric milieu. Examples include erythromycin, penicillin V, pancreatin, and insulin.
2. Drugs that are an irritant to the gastric mucosa and cause unpleasant side effects such as nausea and vomiting. Examples include aspirin and other NSAIDs (e.g., naproxen).
3. Suitable candidates for delayed-release ("time-controlled release") dosage forms, which provide a lag time between 3 and 4 h (74).
4. Drugs with site-specific absorption in the intestines.

Examples of commercial enteric-coated formulations include Ery-Tab[®] (erythromycin; Abbott), Prilosec[®]

(omeprazole; Astra Merck), Pancrease[®] (pancrelipase; McNeil Pharmaceutical), and Ecotrin[®] (aspirin; SmithKline Beecham) (76).

Colon-specific drug delivery systems

Colon-specific drug delivery systems (CSDDS) are designed to permit targeted drug release to the terminal ileum and proximal colon. The objective of delivering drugs to these parts of the GI tract is twofold. First, to provide local delivery for the treatment of colonic diseases, such as inflammatory bowel diseases (ulcerative colitis and Crohn's disease), colon cancer, local spasm of bowel, and constipation. Secondly, to deliver therapeutic peptides and proteins, such as insulin, calcitonin, vasopressin, cyclosporin A, nisin, and cytokine. In addition, drug delivery to the colon is also advantageous from the viewpoint of circadian biorhythms since it can provide a nocturnal release of drugs for diseases that are characterized by night-time or early morning onset (asthma, hypertension, arthritis, cardiac arrhythmias, etc.) (77).

Currently, four strategies are being pursued to achieve colon-specific drug delivery (78).

1. pH-controlled systems. These systems rely on the pH difference between the small and the large intestine. The approach involves coating drugs with pH-sensitive polymers such as Eudragit[®]. Commercial examples include Asacol[®] (mesalazine coated with Eudragit S; for pH ≥ 7), Salofalk[®], Claversal[®], and Mesasal[®] (mesalazine coated with Eudragit L-100; for pH ≥ 6).
2. Enzyme-controlled systems. These systems also are known as microbially controlled systems. They exploit the enzymatic activity of the colonic microflora and therefore, mainly use saccharide-containing polymers or azopolymers, which respectively undergo glycosidic degradation or azoreduction. Alternatively prodrug approach is utilized in which promoieties are resistant to hostile environment of stomach and small intestine, but are susceptible to degradation in the colon. Examples of such prodrug formulations include Dipentum[®] capsules (olsalazine sodium or azodisal sodium; Pharmacia and Upjohn), Colazide[®] tablets (balsalazide disodium; Salix Pharmaceuticals, Ltd., CA), and Intestinol[®] (bensalazine or bensalazide).
3. Time-controlled systems. These systems rely on the relatively consistent small intestinal transit time, which approximates between 3 and 5 h. Obviously, drug release from such systems occurs after a predetermined lag phase (i.e., ≥ 5 h), which can be

precisely programmed by adjusting the thickness and/or composition of the barrier (e.g., coating) formulation. Examples include TIME-CLOCK system ($t_{\text{lag}} = 8.85 \pm 0.90$ h) (79), Pulsincap[®] system ($t_{\text{lag}} = 5.40 \pm 2.53$ h) (80), Geomatrix[®] systems ($t_{\text{lag}} = 17.3$ h in fasted, 12.5 h in fed) (81), and Chronotopic[®] system ($t_{\text{lag}} = 5\text{--}6$ h) (82).

4. Pressure-controlled systems. These systems rely on the strong peristaltic waves in the colon that temporarily increase the luminal pressure.

STRATEGIES TO IMPROVE ORAL DRUG DELIVERY

Chemical Modification Approach

As a general rule, the design of orally active agents must account for the effects of structural modifications on water solubility, lipophilicity, stability (chemical or enzymatic) and therapeutic index. Prodrug design is a commonly used strategy to improve these drug properties. The term prodrug may be defined as a chemical derivative of a drug that is bioconvertible into the active parent drug or an active metabolite responsible for the therapeutic effect. It is important that the prodrug contain a “trigger” site so that its *in vivo* fate can be predicted and the rate of biotransformation can be controlled. Common examples of triggers include ester, amide, carbonate, carbamate, azo and glycosidic (for colon-specific delivery) bonds. In addition, the promoiety (progroup) should be pharmacologically inert, nontoxic, and readily excretable. From the oral delivery standpoint, the design of a prodrug should have following objectives:

1. To enhance the oral bioavailability of a drug. The improved systemic delivery of a drug could be attributed to the fact that a prodrug has the potential to circumvent barriers such as low GI permeability and significant first-pass metabolism (83).
2. To achieve a site-specific drug delivery through a target-specific bioactivation mechanism.

Both of these objectives are achievable provided the delivery (transport) properties of the drug are improved while maintaining its biological activity *in vivo*. In this context, the parent drug properties, such as lipid solubility, aqueous solubility, partition coefficient, and chemical and metabolic stability, are optimized, often in a “balanced” combination.

Based on the definition of a prodrug, the following classifications may be proposed for oral prodrugs:

1. Lipophilic prodrugs. These include slightly soluble salts (e.g., quinidine polygalacturonate, Cardioquin[®]), esters (e.g., enalapril, Vasotec[®]; ramipril, Altace[®]; lisinopril, Prinivil[®]), or complexes.
2. Water-soluble prodrugs. These include water-soluble salts, esters (e.g., fosphenytoin; Cerebyx[®]) or complexes.

The synthesis of lipophilic esters appears to be the most popular approach for improvement of drug lipophilicity, which may be largely attributed to the ubiquitous occurrence of esterases *in vivo* and their potential for enhanced drug absorption via the lymphatic route (10). Briefly, the approach involves either masking (or latentiation) of charged polar groups, such as carboxylic acids, phosphates, and others, or conjugation of lipophilic acyl groups with hydroxyl groups of the parent molecule (84). Once the prodrug is absorbed, it is hydrolyzed by esterases, whereby the parent drug is released.

In contrast to the previous approach, water-soluble prodrugs are formed by esterification of a drug's hydroxyl, amine, or carboxyl group with a promoiety designed to introduce an ionizable function or reduce intermolecular interactions responsible for low solubility (19). However, prior to the selection of an appropriate drug candidate, it is essential to ensure that the required therapeutic dose of the selected poorly water-soluble drug is sufficiently high enough so that the estimated time for its complete (*in vivo*) dissolution far exceeds the typical GI residence times (19).

Due to its higher water solubility, a soluble prodrug provides a greater concentration gradient (driving force) in the intestinal lumen for absorption as compared to its parent compound (19). The further fate of this prodrug would then be determined by the susceptibility of the hydrophilizing promoiety to various enzymes. For instance, if the promoiety were susceptible to cleavage by a membrane-bound brush border enzyme, the lipophilic and permeable parent drug would be released in the vicinity of the mucosal membrane and hence available for subsequent absorption (19). In another example, if the promoiety were susceptible to cleavage by enzymes secreted by the bacterial microflora of the lower GI tract, the lipophilic drug would be available for absorption through the colonic mucosa (85). Glycosidic and glucuronidic prodrugs that exploit bacterial glycosidases and glucuronidases, respectively, are examples of the latter case. Nevertheless, drug conjugates of natural cyclodextrins (ester type) are also known to be susceptible to bacterial enzymes; however, these conjugates are resistant to the gastric and intestinal enzymes. This finding prompted their evaluation as delayed release-type prodrugs for achieving colon-specific delivery (17).

Another interesting approach of improving the oral bioavailability is based on the design of a so-called prodrug of a prodrug (i.e., “pro-prodrugs” or “double prodrugs”). The approach involves protecting the benzylic phenol (or aniline) derivative with a functionality that can be predictably hydrolyzed (25). Obviously, the active drug is generated via an unstable intermediate in a two-step hydrolytic process. Most of the time, the first step is rate-determining and involves an enzymatic cleavage, whereas the second step is faster and involves a molecular decomposition of the intermediate (86). The strategy seems to increase the permeability of poorly soluble drugs, such as steroids, by several orders of magnitude (87).

Greenwald et al. (25) described a combination approach, which combines the concept of double prodrug and permeation enhancement by chemical conjugation with a hydrophilic polymer, PEG. The approach was found to be efficient in delivering an antitumor drug, daunorubicin. While not investigated by the authors of this study, this approach might also possess the potential to reverse multidrug resistance (MDR), in addition to providing high oral bioavailability and tumor selectivity. This statement could be partly supported by a recent finding that PEG-based surfactants can interfere with *P*-Gp substrate binding and thereby reverse MDR (88).

The design of oral prodrugs is not limited to the objective of increasing oral bioavailability and targeted drug delivery. Recently, Kimura et al. (89) described a novel concept of colon targeting for oral treatment of ulcerative colitis, while overcoming the systemic side effects of corticosteroids. The design involves synthesis of glycosidic prodrugs of an antedrug (“pro-antedrug”). Since the resulting pro-antedrugs were hydrophilic due to the polar nature of the promoity, they were not absorbable in the small intestine of guinea pigs. However, they could generate the corresponding antedrugs within the colon by the action of colonic bacteria. The antedrug was thus made available for its local pharmacological action at the colonic mucosa, and when it entered into the systemic circulation, it rapidly metabolized to an inactive metabolite, which results in no systemic effect. To achieve similar objectives, Bodor et al. (90) utilized a “soft drug” concept, in which a lead compound is modified so that the new active drug undergoes a predictable and controllable metabolism *in vivo* to nontoxic moieties after producing its therapeutic effect.

In another recent study, Otagiri et al. (91) described a new concept (“mutual prodrug” or “chimera drug”) in which two drugs with different pharmacological activities were conjugated in order to enhance pharmacological activity or to prevent clinical side effects while

maintaining their individual therapeutic roles and biopharmaceutical profiles. Occasionally, the prodrug approach has been used to develop oral ER dosage forms by preparing a salt or complex of drug that is slightly soluble in the GI fluids (92) and to shift the absorption site of a drug within the GI tract (93).

Use of prodrug strategies for improving oral delivery of peptides and peptidomimetics (e.g., β -Lactam antibiotics) has also received considerable attention in recent years due to the therapeutic potential of these compounds, and ubiquitous peptidases and esterases, respectively. Similar to the prodrugs of nonpeptides, the prodrugs of peptide and protein drugs are designed in light of their metabolic sensitivity to various enzymes (e.g., aminopeptidases, carboxypeptidases, dipeptidases, α -Chymotrypsin, etc.) in order to mimic the absorptive and metabolic characteristics of nutritional substrates. For the most part, this approach has focused on the modification of *N*-Terminal (e.g., amino acid ester approach). However, in order to broaden the applicability of this approach for clinical development of orally bioavailable peptides and peptidomimetics, it is necessary to incorporate structural features that permit the optimization of the pharmaceutical (e.g., solubility), biopharmaceutical (e.g., metabolic stability and membrane permeability), and pharmacological (e.g., receptor affinity) properties of the molecule (94).

Additional examples of commercial formulations of prodrugs for oral administration are Ceftin[®] (cefuroxime axetil; GlaxoWellcome, Inc.); Anzemet[®] (dolasetron mesylate; Hoechst Marion Roussel, Inc.); Trileptal[®] (oxcarbazepine; Novartis Pharmaceuticals Corporation), and Claritin[®] (loratadine; Schering-Plough Corporation).

Formulation Approaches

Use of absorption enhancers

The coadministration of absorption enhancers has long been recognized to enhance the intestinal permeability of several hydrophilic drugs, including peptides and proteins. A number of absorption enhancing agents are commercially available. These include bile salts, surfactants, fatty acids, salicylates, glycerides, chelating agents, and others. Although most of these enhancers have the ability to promote drug uptake, some absorption enhancers may have potential to cause mucosal damage and systemic toxicity, which warrant their safety evaluation prior to selection for formulation development. Ideally speaking, an enhancer of choice for oral delivery will be one that reversibly loosens the tight junctions or transiently increases the membrane per-

meability and that has a relatively higher benefit/risk ratio. To satisfy these interests, a pair of excellent review articles (95, 96) has appeared that are more complete in scope than is possible here.

Based on the routes of permeation, permeation enhancers may act via paracellular or transcellular route or a combination of both. As a common feature, most of the promising absorption enhancers increase drug transport primarily through the modulation of paracellular permeability (96) since transcellular drug transport is not the most likely route for passive diffusion of hydrophilic drugs, including peptides. These are commonly referred to as paracellular promoters, which basically increase the aqueous pore radius through a variety of mechanisms. For instance, EDTA acts as a paracellular promoter by depleting calcium and magnesium in the regions of the tight junction (3). Interestingly, a recent mechanistic study demonstrated that carbonating agents that are integral components of effervescent formulations could be regarded as paracellular promoters since the generated CO₂ might affect the structural integrity of the tight junction (97).

On the contrary, transcellular promoters have been reported to improve oral absorption through membrane perturbation (98). Most of them act on the intracellular lipids, in many cases by fluidizing, solubilizing, or reorganizing the phospholipid domains in the membrane, which causes disruption of the membrane integrity (99). A change in membrane fluidity may also occur primarily through the interaction between these enhancers and membrane protein. Examples of such enhancers are sodium salicylate and sodium caprylate (98). Interestingly enough, there is ample evidence to suggest that nonionic surfactants (e.g., Cremophor EL and Tween 80) may enhance the permeability of drugs through their inhibitory effect on the *P*-Gp-Mediated counter transport process (100, 101). Although the exact mechanism(s) of inhibition is not well understood, these surfactants might directly act on the *P*-Gp protein on the mucosal surface, thereby decreasing the drug efflux from the intestinal membrane (101), or by decreasing the membrane lipid fluidity (102).

Use of metabolism inhibitors

As described previously, the presystemic metabolism of drugs may occur via various mechanisms. It is obvious, therefore, that coadministration of a low bioavailable drug and its metabolism inhibitor, which can selectively inhibit any of the contributing processes, would result in increased fractional absorption and hence a higher bioavailability. In fact, this approach seems to be a

promising alternative to overcome the enzymatic barriers to oral delivery of metabolically labile drugs such as peptides and proteins.

Current novel approaches in this area include: 1) bioadhesive delivery systems that can reduce the drug degradation between the delivery system and absorbing membrane by providing an intimate contact to the mucosa; 2) CR microencapsulated systems that can provide simultaneous delivery of a drug and its specific enzyme inhibitor at the desired site for required period of time; and 3) immobilization of enzyme inhibitors on mucoadhesive delivery systems (103).

Interestingly, the intestinal wall metabolism may also be inhibited by coadministration of certain drugs and diet, which act by selectively inhibiting an enzyme present in enterocytes. An illustrative example is that of cyclosporine, which undergoes extensive intestinal metabolism, resulting in low bioavailability ranging between 30–40% (104). Studies have shown that coadministration of ketoconazole (105) and grapefruit juice (106), which contains the inhibitory components, can significantly decrease the presystemic metabolism (both act via selective inhibition of intestinal, not hepatic, CYP3A4) and consequently increase the oral bioavailability of cyclosporine. In a differential manner, however, ketoconazole moderately inhibits *P*-gp (55), whereas grapefruit juice activates *P*-Gp-mediated efflux of cyclosporine, which is a well characterized substrate of *P*-Gp, thereby partially counteracting the CYP3A4-inhibitory effects of grapefruit juice (107).

Similarly, the coadministration of a drug that can selectively inhibit an enzyme in the liver may lead to increased bioavailability of another drug. For example, coadministration of erythromycin can result in inhibition of hepatic metabolism and thereby significantly increase the oral bioavailability of cyclosporine (108). As a matter of fact, this attribute of erythromycin appears to be selective, which permits a noninvasive measurement of the *in vivo* hepatic CYP3A4 activity, popularly known as erythromycin breath test (109). Many examples also exist related to the effects of diet on hepatic first-pass metabolism that are discussed in detail elsewhere (110, 111). Accordingly, in our view, the dietary approach seems to justify as a prospective alternative to decrease presystemic metabolism, especially in terms of cost effectiveness and clinical acceptance for routine medical practice.

Ion pairing and complexation

The ion pairing approach involves coadministration of a hydrophilic or polar drug with a suitable lipophilic

counterion, which consequently improves the partitioning of the resultant ion pair (relatively more lipophilic) into the intestinal membrane. In fact, the approach seems to increase the oral bioavailability of ionizable drugs, such as atenolol, by approximately 2-fold (112). However, it is important that a counterion possess high lipophilicity, sufficient aqueous solubility, physiological compatibility, and metabolic stability (112). Some of the successful applications of this approach have already been described in previous reviews (14, 112) and therefore, will not be described any further here.

Complexation approach has been frequently used to increase the aqueous solubility of water-insoluble and slightly soluble drugs, and dissolution rate, both in an effort to increase oral bioavailability. However, in certain instances, the approach can be used to increase the drug stability, particularly esters, control drug release rates, improve organoleptic properties, and maximize the GI tolerance by reducing drug irritation after oral administration. Generally speaking, β -Cyclodextrins are potential carriers for achieving such objectives but other complexing agents, such as caffeine, sodium salicylate, sodium benzoate, and nicotinamide, may also be used. The enhancement in solubility by these aromatic and heterocyclic compounds occurs via a π -Electron related donor-acceptor mechanism (21), whereas β -Cyclodextrins are known to form inclusion complexes with nonpolar guest molecules in their central lipophilic cavity during which no covalent bonds are formed or broken (20).

It has been suggested that since β -Cyclodextrins are hydrophilic, they have the ability to keep the hydrophobic drug molecules in solution form, thereby increasing the drug availability at the surface of the GI mucosa (where the drug solution partitions into the membrane). This results in increased drug absorption; however, it is important that cyclodextrin concentration be optimum (20). Furthermore, the commercial availability of a large variety of cyclodextrin derivatives permits the design of advanced oral DDS. For instance, the hydrophilic and ionizable cyclodextrins can serve as potent drug carriers in the formulation of IR and delayed-release (DR) dosage forms, respectively, while hydrophobic cyclodextrins can be used to retard the release rate of water-soluble drugs such as diltiazem and isosorbide dinitrate (17). Some of the commercial formulations that employ complexation with β -Cyclodextrin are Brexin[®] tablets (piroxicam; Chiesi Farmaceutici SpA, Parma, Italy), Surgamyl[®] tablets (tiaprofenic acid; Roussel-Maestrelli, Italy) and Nitro-pen[®] sublingual tablets (nitroglycerin; Nippon Kayaku, Japan) (20).

Lipid-based formulations

Lipid-based formulations of poorly water soluble drugs offer large versatility for oral administration as they can be formulated as solutions, gels, suspensions, emulsions, self-emulsifying systems, multiple emulsions, microemulsions, liposomes, and solid dispersions (113).

Administration of a drug in a lipidic vehicle/formulation can enhance the absorption and oral bioavailability via a combination of various mechanisms (14, 114) that are briefly summarized as follows:

- Physicochemical—Enhanced dissolution and solubility
- Physiological—Slowed gastric emptying (hence increased time for dissolution and absorption), stimulated bile flow and secretion of pancreatic juice, bile salt micellar solubilization, increased membrane permeability of intestinal epithelium, increased uptake from the intestinal lumen into the lymphatic system, reduced intestinal blood flow, and decreased luminal degradation and first-pass metabolism through the GI tract and liver

As a matter of fact, the coadministration of a fatty (“hyperlipidic”) food also has been known to improve the bioavailability of many poorly absorbed drugs by a combination of these mechanisms (115).

Additionally, several recent studies suggest that lipids may affect the oral bioavailability of lipophilic drugs by modulating biochemical processes. For example, Barnwell et al. (116) found that coadministration of oleic acid increases the bioavailability of propranolol, a lipophilic drug. Based on their results, the authors proposed two possible explanations. First, oleic acid may promote lymphatic absorption of propranolol since it is known to activate lymph production. The avoidance of hepatic first-pass metabolism, therefore, led to increased bioavailability. Second, oleic acid may reverse the inhibition of lymph production caused by propranolol. This hypothesis was consistent with the previous findings (listed in Ref. 116) that several lipophilic drugs, including propranolol, inhibit protein kinase C (a key enzyme involved in cellular signal transduction which controls secretory events, including lymph production) and that oleic acid reverses inhibition of protein kinase C by propranolol.

Likewise, several other studies have demonstrated that certain lipidic formulations can act as effective modulators of *P*-Gp counter transport systems (88, 100, 117–119), which offers a qualitative explanation for enhanced drug transport in independently examined cases, for instance, increased bioavailability of cyclosporine when coadministered with water-soluble vitamin E (120, 121). Thus, a lipid-based formulation approach

may be considered as a cost-effective strategy for future oral drug therapy.

For the sake of completeness, it should be noted that certain lipidic excipients, such as hydrogenated castor oil (Cutina[®] HR), polyethoxylated castor oil (Cremophor[®] RH 40), stearic acid, glyceryl behenate (Compritol[®] 888; Gattefossé Corporation, USA), glyceryl monostearate, glyceryl palmitostearate (Precirol[®] ATO 5; Gattefossé Corporation, USA), and saturated polyglycolized glycerides (Gelucire[®]; Gattefossé Corporation, USA), could be used as hydrophobic carriers to develop SR/CR tablets, particularly for water-soluble drugs, such as diltiazem HCl (122) and phenylpropanolamine HCl (123), as well as other drugs, such as theophylline (124). However, fed conditions of the stomach are likely to affect the *in vivo* disintegration of such formulations (125), and therefore, appropriate recommendations with regard to meal timing must be made to assure reproducible drug absorption or clinical response. In fact, the effect seems to be significant with a high-fat meal, which increases the pancreatic and biliary secretions that would affect the lipid matrix of tablets (48, 124). In contrast, however, SR dosage forms that are formulated as an oily “semisolid” matrix system filled in hard gelatin capsules seem to offer prolonged drug response under the fed state, which could be attributed to delayed gastric emptying (both by food and oily vehicles) and resultant prolonged absorption phase (126).

Unfortunately, the number of commercial formulations is very limited, primarily because of stability and manufacturing problems encountered during large-scale production. The list includes Sandimmune Neoral[®] (cyclosporine A; Novartis AG, Switzerland), which contains a microemulsion concentrate and is available as soft gels and solutions, Sandimmune[®] (cyclosporine A; Novartis AG, Switzerland), which contains an emulsion concentrate, and lipid soluble vitamins. Both formulations of cyclosporine have self-emulsifying properties and spontaneously form an o/w microemulsion (particle size <0.15 μm) and an o/w emulsion, respectively in the aqueous fluids of the GI tract (127).

Although the discussion concerning potential of liposomes, niosomes, microemulsions, and solid dispersions for oral delivery is outside the scope of this chapter, the interested reader is referred to recently published reviews on these topics (23, 128–130).

Design of drug delivery systems

Ideally, the ultimate goal of a *de novo* oral DDS should be aimed at providing a desired therapeutic response that comprises a controlled onset, a required duration, and a

sufficient intensity while avoiding or minimizing potential adverse effects. In terms of delivery system design, this objective translates to an advanced DDS that can precisely control the rate, time, and/or site of drug delivery independently of normal physiological variables such as GI pH, digestive state of the GI tract, peristalsis, and circadian rhythm. While it is difficult to collate the information about all the oral DDS available worldwide, Tables 2–4 describe some of the potential and advanced DDS in the alphabetical order of companies that have developed these technologies.

ORAL DELIVERY OF PROTEINS AND MACROMOLECULES

In the past 2 decades, significant progress has been made in the oral delivery of peptides and proteins that has enabled scientists to deliver insulin, calcitonin, heparin, and several vaccines by oral route. As with conventional drugs, the oral delivery of peptides and proteins is subject to similar challenges and biological barriers. However, to a greater extent this is due to their higher metabolic sensitivity and other inherent properties such as hydrophilicity, charge, and large molecular weight (≥ 600 Da). Current approaches to improve oral delivery of peptides and proteins include prodrug design (18, 94), ion-pairing approach (133), microemulsions (134), particulate delivery systems using polymeric carriers (135), liposomes (128), and niosomes (130), and use of metabolism inhibitors (103) and absorption enhancers (136). Depending upon the physicochemical properties of the therapeutic agent and its intended site of delivery, these approaches may be combined in an effort to enhance the oral bioavailability or achieve targeted therapy. The latter objective, in particular, refers to targeting of Peyer’s patch M cells, which offers potential strategy for oral delivery of vaccines.

FUTURE OF ORAL DRUG DELIVERY

In the last 3 decades, the area of oral drug delivery has undergone unprecedented changes. This transition has been mainly driven by the increased understanding of pharmaceutical and biological constraints at the molecular/atomic level, increased generation of drug candidates, and emergence of several advanced technologies. Overall, the area has made significant technological advances not only in drug delivery aspects but also in various processing (fabrication) technologies, as recently

Table 4 Miscellaneous advanced and novel drug delivery systems for oral delivery

Technology owner and DDS	Brief description	Unique features	Marketed products (drugs)
<i>Cortecs International Ltd., Teeside, U.K:</i> Halo™ Drug delivery system (131)	Hepatic avoidance using lymphatic output oral DDS. A liver-bypass drug delivery system that promotes absorption redistribution of lipophilic drug from the hepatic portal blood supply to the lymphatic system, thereby avoiding first-pass liver metabolism.	Improved oral bioavailability of lipophilic drugs.	Propranolol, verapamil, nifedipine, diltiazem, metoprolol, nicardipine, or labetalolol.
<i>Elan Corporation, plc, Dublin, Ireland:</i> DUREDAS®	Dual release drug absorption system. A bilayer tableting technology that combines an immediate-release granulate and a CR hydrophilic matrix complex within the one tablet.	Two different release rates or dual release of one or two different drugs from a single dosage form.	
IPDAS®	Intestinal protective drug absorption system. A multiparticulate tablet technology that is composed of high density CR beads and an IR granulate.	Dosage form is highly flexible in that the release rate from the tablet can be customized to deliver a drug specific absorption profile associated with optimized clinical benefit and tolerability.	Naprelan (naproxen sodium) as a once daily formulation.
<i>Elan Corporation, plc, Dublin, Ireland:</i> INDAS®	Insoluble drug absorption system. A technology for improving the solubility and absorption characteristics of poorly water-soluble drugs, enabling the development of a CR formulation.	A well-maintained absorption of drug is consistently achieved over the 24-h dosing interval.	Nifelan® (nifedipine) as a once daily formulation.
MODAS®	Multiporous oral drug absorption system. A single-unit, immediate-release tablet formulation consisting of an inner core, containing active drug plus excipients, surrounded by a non-disintegrating, timed release coating.	Combination/dual release dosage forms with an immediate release outer coating.	Bron-12® (a 12-h multicomponent OTC cough cold product), potassium chloride (once daily).
BEODAS®	Bioerodible oral drug absorption system. A bioerodible micro/nano particle technology used in the development of oral protein and peptide formulations.	Oral administration of vaccines.	Under development.
PRODAS®	Programmable oral drug absorption system. A technology for producing multiparticulate formulations by encapsulating combinations of IR, delayed release and/or CR minitablets (1.5–4.00 mm in dia).	The hybrid system incorporates the benefits of both multiparticulate and single-unit monolithic systems into one dosage form.	Not available.

(Continued)

Table 4 Miscellaneous advanced and novel drug delivery systems for oral delivery (*Continued*)

Technology owner and DDS	Brief description	Unique features	Marketed products (drugs)
LOCIDAS [®]	Localized drug absorption system. A novel targeted oral drug delivery technology. The system uses targeting ligands, which are attached to coated microparticles of protein and peptide drugs. These are subsequently packaged in enteric-coated capsules that deposit the particles at appropriate sites of absorption within the GI tract.	Increased permeability and ultimate oral bioavailability of macromolecules.	Under development for insulin.
<i>F. Hoffmann-La Roche Ltd., Basel, Switzerland:</i> HBS TM	Hydrodynamically balanced system. A CR oral dosage form (capsule or tablet), which is designed to prolong the residence time within the stomach.	Increased time for dissolution, prolonged absorption phase, and improved pharmacokinetic profile.	Madopar HBS or Prolopa [®] HBS (L-dopa + Benserazide); Valrelease (diazepam).
<i>R.P. Scherer Corporation, MI, USA:</i> Pulsincap [®] System (132)	A unique CR (capsule) device that permits programmed drug release at a predetermined time and/or site within the GI tract (e.g., colon).	Allows sequential dosing of a patient at different predetermined times and/or sites in the GI tract, with same or different drugs.	Under development.
R.P. Scherer sol TM System	Liquid formulations for softgels, which incorporate the drug in a microemulsion preconcentrate or microemulsion form.	Oral delivery of proteins and peptides; enhanced oral bioavailability with reduced inter- and intraindividual variability in pharmacokinetics of certain drugs.	Sandimmune Neoral (cyclosporine).
<i>Yamanouchi Shaklee Pharma, CA, USA:</i> OCAS TM	Oral-controlled absorption system. A hydrophilic gel-forming oral CR system capable of providing predictable drug release and absorption up to 24 h throughout the GI tract, including the colon.	Applicable to wide range of drugs including small molecules and peptides; reduced food effect and intersubject variability in plasma drug levels.	None
CODES TM	A colon-targeted delivery system for proteins/peptides and other drugs. The system contains drug and lactulose as a core formulation, which is subsequently coated first with an enteric coating and then with a cationic polymer coating.		Insulin

reviewed (61, 137). Despite the fact that currently existing technologies have not been exploited in their full capacity, the annual sales of oral products always ranks first in the list of pharmaceutical products since they offer increased patient compliance and have greater therapeutic outcome. In addition, these technologies help pharmaceutical companies manage the life cycles of their products through extended patent protection. It is therefore anticipated that current trends in oral delivery will bring a much-awaited revolution in delivery of therapeutic agents, particularly biotechnology-based molecules.

SUMMARY

The oral route still remains the major and favorite route of delivery for most bioactive agents in many disease states, despite its challenges and limitations. Optimizing drug delivery potential through this route requires careful design of drug and its delivery system while taking into consideration the physicochemical properties of the drug, biological and metabolic barriers to oral delivery, pharmacokinetic principles, and patient-related critical variables. Consequently, combinations of different formulation and CR approaches may play a significant role in the potential development of oral DDS of existing and new lead compounds.

REFERENCES

1. Rouge, N.; Buri, P.; Doelker, E. Drug Absorption Sites in the Gastrointestinal Tract and Dosage Forms for Site-Specific Delivery. *Int. J. Pharm.* **1996**, *136* (1, 2), 117–139.
2. Madara, J.L. Functional Morphology of Epithelium of the Small Intestine. *Intestinal Absorption and Secretion*; Field, M., Raymond, A.F., Eds.; American Physiological Society, 1991; IV, 83–120.
3. Tomita, M.; Shiga, M.; Hayashi, M.; Awazu, S. Enhancement of Colonic Drug Absorption by the Paracellular Permeation Route. *Pharm. Res.* **1988**, *5* (6), 341–346.
4. Pappenheimer, J.R. Physiological Regulation of Transepithelial Impedance in the Intestinal Mucosa of Rats and Hamsters. *J. Membr. Biol.* **1987**, *100* (2), 137–148.
5. Hebden, J.M.; Wilson, C.G.; Spiller, R.C.; Gilchrist, P.J.; Blackshaw, E.; Frier, M.E.; Perkins, A.C. Regional Differences in Quinine Absorption from the Undisturbed Human Colon Assessed Using a Timed Release Delivery System. *Pharm. Res.* **1999**, *16* (7), 1087–1092.
6. Russell-Jones, G.J. Carrier-Mediated Transport, Oral Drug Delivery. *Encyclopedia of Controlled Drug Delivery*; Mathiowitz, E., Ed.; John Wiley & Sons, Inc.: New York, 1999; 1, 173–184.
7. Barthe, L.; Woodley, J.; Houin, G. Gastrointestinal Absorption of Drugs: Methods and Studies. *Fundam. Clin. Pharmacol.* **1999**, *13* (2), 154–168.
8. Russell-Jones, G.J. The Potential Use of Receptor-Mediated Endocytosis for Oral Drug Delivery. *Adv. Drug Del. Rev.* **1996**, *20* (1), 83–97.
9. Swaan, P.W. Recent Advances in Intestinal Macromolecular Drug Delivery Via Receptor-Mediated Transport Pathways. *Pharm. Res.* **1998**, *15* (6), 826–834.
10. Charman, W.N.; Porter, C.J.H. Lipophilic Prodrugs Designed for Intestinal Lymphatic Transport. *Adv. Drug Del. Rev.* **1996**, *19* (2), 149–169.
11. Code, C.F.; Marlett, J.A. The Interdigestive Myo-electric Complex of the Stomach and Small Bowel of Dogs. *J. Physiol. (Lond)* **1975**, *246* (2), 289–309.
12. Sarna, S.K. Cyclic Motor Activity: Migrating Motor Complex. *Gastroenterology* **1985**, *89* (4), 894–913.
13. Thummel, K.E. Preface. *Adv. Drug Del. Rev.* **1997**, *27* (2, 3), 97–98.
14. Aungst, B.J. Novel Formulation Strategies for Improving Oral Bioavailability of Drugs with Poor Membrane Permeation or Presystemic Metabolism. *J. Pharm. Sci.* **1993**, *82* (10), 979–987.
15. Yalkowski, S.H.; Valvani, S.C. Solubility and Partitioning I: Solubility of Nonelectrolytes in Water. *J. Pharm. Sci.* **1980**, *69* (8), 912–922.
16. Robinson, J.R. Recent Advances in Formulation of Poorly Absorbed Drugs. *Capsugel Library* **1994**, 59–63.
17. Hirayama, F.; Uekama, K. Cyclodextrin-based Controlled Drug Release System. *Adv. Drug Del. Rev.* **1999**, *36* (1), 125–141.
18. Oliyai, R. Prodrugs of Peptides and Peptidomimetics for Improved Formulation and Delivery. *Adv. Drug Del. Rev.* **1996**, *19* (2), 275–286.
19. Fleisher, D.; Bong, R.; Stewart, B.H. Improved Oral Drug Delivery: Solubility Limitations Overcome by the Use of Prodrugs. *Adv. Drug Del. Rev.* **1996**, *19* (2), 115–130.
20. Loftsson, T. Pharmaceutical Applications of β -Cyclodextrin. *Pharm. Technol.* **1999**, *23* (12), 40, 42, 44, 46, 48, 50.
21. Higuchi, T.; Kristiansen, H. Binding Specificity Between Small Organic Solutes in Aqueous Solution: Classification of Some Solutes into Two Groups According to Binding Tendencies. *J. Pharm. Sci.* **1970**, *59* (11), 1601–1608.
22. Chaumeil, J.C. Micronization: A Method of Improving the Bioavailability of Poorly Soluble Drugs. *Methods Find. Exp. Clin. Pharmacol.* **1998**, *20* (3), 211–215.
23. Serajuddin, A.T.M. Solid Dispersion of Poorly Water-Soluble Drugs: Early Promises, Subsequent Problems, and Recent Breakthroughs. *J. Pharm. Sci.* **1999**, *88* (10), 1058–1066.
24. Chiou, W.L.; Riegelman, S. Pharmaceutical Applications of Solid Dispersion Systems. *J. Pharm. Sci.* **1971**, *60* (9), 1281–1302.
25. Greenwald, R.B.; Pendri, A.; Conover, C.D.; Zhao, H.; Choe, Y.H.; Martinez, A.; Shum, K.; Guan, S. Drug Delivery Systems Employing 1,4- or 1,6-elimination: Poly (Ethylene Glycol) Prodrugs of Amine-Containing Compounds. *J. Med. Chem.* **1999**, *42* (18), 3657–3667.
26. Sugimoto, M.; Okagaki, T.; Narisawa, S.; Koida, Y.; Nakajima, K. Improvement of Dissolution Characteristics

- and Bioavailability of Poorly Water-Soluble Drugs by Novel Cogrounding Method Using Water-Soluble Polymer. *Int. J. Pharm.* **1998**, *160* (1), 11–19.
27. Yamada, T.; Saito, N.; Imai, T.; Otagiri, M. Effect of Grinding With Hydroxypropyl Cellulose on the Dissolution and Particle Size of a Poorly Water-Soluble Drug. *Chem. Pharm. Bull.* **1999**, *47* (9), 1311–1313.
 28. Kubo, H.; Mizobe, M. Improvement of Dissolution Rate and Oral Bioavailability of a Sparingly Water-Soluble Drug, (+/-)-5-[(2-(2-Naphthalenylmethyl)-5-Benzoxazolyl)-methyl]-2, 4-Thiazolidinedione, in Co-Ground Mixture With D-mannitol. *Biol. Pharm. Bull.* **1997**, *20* (4), 460–463.
 29. Krämer, S.D. Absorption Prediction from Physicochemical Parameters. *Pharm. Sci. Technol. Today* **1999**, *2* (9), 373–380.
 30. Mori, T.; Nishimura, K.; Tamaki, S.; Nakamura, S.; Tsuda, H.; Kakeya, N. Pro-Drugs for the Oral Delivery of Disodium Cromoglycate. *Chem. Pharm. Bull.* **1988**, *36* (1), 338–344.
 31. Wils, P.; Warnery, A.; Phung-Ba, V.; Legrain, S.; Scherman, D. High Lipophilicity Decreases Drug Transport Across Intestinal Epithelial Cells. *J. Pharmacol. Exp. Ther.* **1994**, *269* (2), 654–658.
 32. Arimori, K.; Nakano, M. Drug Exsorption from Blood into the Gastrointestinal Tract. *Pharm. Res.* **1998**, *15* (3), 371–376.
 33. Lennernäs, H. Human Intestinal Permeability. *J. Pharm. Sci.* **1998**, *87* (4), 403–410.
 34. Hörter, D.; Dressman, J.B. Influence of Physicochemical Properties on Dissolution of Drugs in the Gastrointestinal Tract. *Adv. Drug Del. Rev.* **1997**, *25* (1), 3–14.
 35. Fuse, K.; Bamba, T.; Hosoda, S. Effects of Pectin on Fatty Acid and Glucose Absorption and on Thickness of Unstirred Water Layer in Rat and Human Intestine. *Dig. Dis. Sci.* **1989**, *34* (7), 1109–1116.
 36. Johnson, I.T.; Gee, J.M. Effect of Gel-Forming Gums on the Intestinal Unstirred Layer and Sugar Transport in Vitro. *Gut* **1981**, *22* (5), 398–403.
 37. Cerda, J.J.; Robbins, F.L.; Burgin, C.W.; Gerencser, G.A. Unstirred Water Layers in Rabbit Intestine: Effects of Guar Gum. *J. Parenter. Enteral Nutr.* **1987**, *11* (1), 63–66.
 38. Huang, J.D. Role of Unstirred Water Layer in the Exsorption of Quinidine. *J. Pharm. Pharmacol.* **1990**, *42* (6), 435–437.
 39. Davis, G.R.; Santa Ana, C.A.; Morawski, S.G.; Fordtran, J.S. Permeability Characteristics of Human Jejunum, Ileum, Proximal Colon and Distal Colon: Results of Potential Difference Measurements and Unidirectional Fluxes. *Gastroenterology* **1982**, *83* (4), 844–850.
 40. Wang, W. Oral Protein Delivery. *J. Drug Target* **1996**, *4* (4), 195–232.
 41. Corrigan, O.I. The Biopharmaceutic Drug Classification and Drugs Administered in Extended Release (ER) Formulations. *In-Vitro-In-Vivo Correlations*; Young, D., Devane, J.G., Butler, J., Eds.; Plenum Press: New York, 1997; 111–128.
 42. Sakr, F.M. A Programmable Drug Delivery System for Oral Administration. *Int. J. Pharm.* **1999**, *184* (1), 131–139.
 43. Singh, B.N.; Kim, K.H. Floating Drug Delivery Systems: An Approach to Oral Controlled Drug Delivery Via Gastric Retention. *J. Controlled Release* **2000**, *63* (3), 235–259.
 44. Davis, S.S.; Hardy, J.G.; Fara, J.W. Transit of Pharmaceutical Dosage Forms Through the Small Intestine. *Gut* **1986**, *27* (8), 886–892.
 45. Coupe, A.J.; Davis, S.S.; Wilding, I.R. Variation in Gastrointestinal Transit of Pharmaceutical Dosage Forms in Healthy Subjects. *Pharm. Res.* **1991**, *8* (3), 360–364.
 46. Singh, B.N. Effects of Food on Clinical Pharmacokinetics. *Clin. Pharmacokinet.* **1999**, *37* (3), 213–255.
 47. Abrahamsson, B.; Roos, K.; Sjogren, J. Investigation of Prandial Effects on Hydrophilic Matrix Tablets. *Drug Dev. Ind. Pharm.* **1999**, *25* (6), 765–771.
 48. Gai, M.N.; Ferj, S.; Garcia, E.; Seitz, C.; Thielemann, A.M.; Andonaegui, M.T. Evaluation of the In Vitro and In Vivo Performance of Two Sustained-Release Lithium Carbonate Matrix Tablets. Effect of Different Diets on The Bioavailability. *Drug Dev. Ind. Pharm.* **1999**, *25* (2), 131–140.
 49. Dressman, J.B.; Bass, P.; Ritschel, W.A.; Friend, D.R.; Rubinstein, A.; Ziv, E. Gastrointestinal Parameters that Influence Oral Medications. *J. Pharm. Sci.* **1993**, *82* (9), 857–872.
 50. Vyas, S.P.; Sood, A.; Venugopalan, P.; Mysore, N. Circadian Rhythms and Drug Delivery Design. *Pharmazie* **1997**, *52* (11), 815–820.
 51. Lemmer, B. Circadian Rhythms and Drug Delivery. *J. Controlled Release* **1991**, *16* (1,2), 63–74.
 52. Thummel, K.E.; Kunze, K.L.; Shen, D.D. Enzyme-Catalyzed Processes of First-Pass Hepatic and Intestinal Drug Extraction. *Adv. Drug Del. Rev.* **1997**, *27* (2,3), 99–127.
 53. Ambudkar, S.V.; Dey, S.; Hrycyna, C.A.; Ramachandra, M.; Pastan, I.; Gottesman, M.M. Biochemical, Cellular, and Pharmacological Aspects of the Multidrug Transporter. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 361–398.
 54. Fricker, G.; Drewe, J.; Huwyler, J.; Gutmann, H.; Beglinger, C. Relevance Of P-Glycoprotein for the Enteral Absorption of Cyclosporin A: In Vitro–In Vivo Correlation. *Br. J. Pharmacol.* **1996**, *118* (7), 1841–1847.
 55. Benet, L.Z.; Izumi, T.; Zhang, Y.; Silverman, J.A.; Wachter, V.J. Intestinal MDR Transport Proteins and P-450 Enzymes As Barriers to Oral Drug Delivery. *J. Controlled Release* **1999**, *62* (1,2), 25–31.
 56. Watkins, P.B. The Barrier Function of CYP3A4 And P-Glycoprotein in the Small Bowel. *Adv. Drug Del. Rev.* **1997**, *27* (2,3), 161–170.
 57. Hunter, J.; Hirst, B.H. Intestinal Secretion of Drugs: The Role Of P-glycoprotein and Related Drug Efflux Systems in Limiting Oral Drug Absorption. *Adv. Drug Del. Rev.* **1997**, *25* (2, 3), 129–157.
 58. Dresser, G.K.; Spence, J.D.; Bailey, D.G. Pharmacokinetic–Pharmacodynamic Consequences and Clinical Relevance of Cytochrome P450 3A4 Inhibition. *Clin. Pharmacokinet.* **2000**, *38* (1), 41–57.
 59. Takada, K.; Yoshikawa, H. Oral Drug Delivery, Traditional. *Encyclopedia of Controlled Drug Delivery*; Mathiowitz, E., Ed.; John Wiley & Sons, Inc.: New York, 1999; 2, 728–742.

60. Seager, H. Drug-Delivery Products and the Zydis Fast-Dissolving Dosage Form. *J. Pharm. Pharmacol.* **1998**, 50 (4), 375–382.
61. Sastry, S.V.; Nyshadham, J.R.; Fix, J.A. Recent Technological Advances in Oral Drug Delivery: A Review. *Pharm. Sci. Technol. Today* **2000**, 3 (4), 138–145.
62. Bruna, E.; Leneveu, A.; Abou Chacra, M.L.; Delhotal, B.; Chauveau, C.; Ragot, F.; Flouvat, B. Acetaminophen Flashtab Formulation: Fast Disintegration and Optimal Absorption of the Active Ingredient. *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **1998**, 25, 938–939.
63. Gutowska, A.; Bark, J.S.; Kwon, I.C.; Bae, Y.H.; Cha, Y.; Kim, S.W. Squeezing Hydrogels for Controlled Oral Drug Delivery. *J. Controlled Release* **1997**, 48 (2), 141–148.
64. Van Bommel, E.M.G., Ed.; *Gradient Matrix Systems for Controlled Delivery of Drugs*, Company Thesis Duphar B.V., Weesp: The Netherlands, 1990.
65. Conte, U.; Maggi, L. A Flexible Technology for the Linear, Pulsatile and Delayed Release of Drugs, Allowing for Easy Accommodation of Difficult In Vitro Targets. *J. Controlled Release* **2000**, 64 (1–3), 263–268.
66. Rubinstein, A.; Friend, D.R. Specific Delivery to the Gastrointestinal Tract. *Polymeric Site-Specific Pharmacotherapy*; Domb, A.J., Ed.; John Wiley & Sons Ltd. Chichester, England, 1994; 267–313.
67. Hwang, S.J.; Park, H.; Park, K. Gastric Retentive Drug-Delivery Systems. *Crit. Rev. Ther. Drug Carrier Syst.* **1998**, 15 (3), 243–284.
68. Menon, A.; Ritschel, W.A.; Sakr, A. Development and Evaluation of a Monolithic Floating Dosage Form for Furosemide. *J. Pharm. Sci.* **1994**, 83 (2), 239–245.
69. Rouge, N.; Cole, E.T.; Doelker, E.; Buri, P. Buoyancy and Drug Release Patterns of Floating Minitablets Containing Piretanide and Atenolol as Model Drugs. *Pharm. Dev. Technol.* **1998**, 3 (1), 73–84.
70. Gupta, P.K.; Leung, S.H.S.; Robinson, J.R. Bioadhesives/Mucoadhesives in Drug Delivery to the Gastrointestinal Tract. *Bioadhesive Drug Delivery Systems*; Lenaerts, V., Gurny, R., Eds.; CRC Press, Inc.: Boca Raton FL, 1990; 65–92.
71. Ponchel, G.; Irache, J.M. Specific and Non-Specific Bioadhesive Particulate Systems for Oral Delivery to the Gastrointestinal Tract. *Adv. Drug Del. Rev.* **1998**, 34 (2, 3), 191–219.
72. Peppas, N.A.; Sahlin, J.J. Hydrogels As Mucoadhesive and Bioadhesive Materials: A Review. *Biomaterials* **1996**, 17 (16), 1553–1561.
73. Nagai, T.; Machida, Y. Mucosal Adhesive Dosage Forms. *Pharm. Int.* **1985**, 6 (1), 196–200.
74. Healey, J.N.C. Enteric Coatings and Delayed Release. *Drug Delivery to the Gastrointestinal Tract*; Hardy, J.G., Davis, S.S., Wilson, C.G., Eds.; Ellis Horwood Ltd.: Chichester, UK, 1989; 83–96.
75. Chambliss, W.G. Enteric Coatings. *Encyclopedia of Pharmaceutical Technology*, First Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1992; 5, 189–200.
76. *Physician's Desk Reference*, Medical Economics Data Production Company: Montvale, NJ, 1997.
77. Wilson C.G. Design of Formulations for Colonic Delivery Pre-Symposium Workshop on Oral Controlled Drug Delivery Feb 5 & 6 1997, Indian Chapter of the Controlled Release Society, India.
78. Leopold, C.S. Coated Dosage Forms for Colon-Specific Drug Delivery. *Pharm. Sci. Technol. Today* **1999**, 2 (5), 197–204.
79. Steed, K.P.; Hooper, G.; Monti, N.; Benedetti, M.S.; Fornasini, G.; Wilding, I.R. The Use of Pharmacoscintigraphy to Focus the Development Strategy for a Novel 5-ASA Colon Targeting System ("TIME CLOCK[®]" System). *J. Controlled Release* **1997**, 49 (2,3), 115–122.
80. Wilson, C.G.; Bakhshae, M.; Stevens, H.N.E.; Perkins, A.C.; Frier, M.; Blackshaw, E.P.; Binns, J.S. Evaluation of a Gastro-Resistant Pulsed Release Delivery System (Pulsincap) in Humans. *Drug Deliv.* **1997**, 4 (3), 201–206.
81. Wilding, I.R.; Davis, S.S.; Sparrow, R.A.; Ziemniak, J.A.; Heald, D.L. Pharmacoscintigraphic Evaluation of a Modified Release (Geomatrix[®]) Diltiazem Formulation. *J. Controlled Release* **1995**, 33 (1), 89–97.
82. Sangalli, M.E.; Busetti, C.; Maroni, A.; Zema, L.; Olgiati, V.; Giordano, F.; Gazzaniga, A. Chronotopic[®] System for Colon Specific Delivery: Pharmacokinetic and Scintigraphic Evaluation. *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **1998**, 25, 858–859.
83. Kearney, A.S. Prodrugs and Targeted Drug Delivery. *Adv. Drug Del. Rev.* **1996**, 19 (2), 225–239.
84. Taylor, M.D. Improved Passive Oral Drug Delivery Via Prodrugs. *Adv. Drug Del. Rev.* **1996**, 19 (2), 131–148.
85. Tozer, T.N.; Friend, D.R.; McLeod, A.D. Kinetic Perspectives on Colonic Delivery. *S.T.P. Pharm. Sci.* **1995**, 5 (1), 5–12.
86. Bundgaard, H. The Double Prodrug Concept and Its Applications. *Adv. Drug Del. Rev.* **1989**, 3 (1), 39–65.
87. Stewart, B.H.; Amidon, G.L.; Brabec, R.K. Uptake of Prodrugs by Rat Intestinal Mucosal Cells: Mechanism and Pharmaceutical Implications. *J. Pharm. Sci.* **1986**, 75 (10), 940–945.
88. Buckingham, L.E.; Balasubramanian, M.; Safa, A.R.; Shah, H.; Komarov, P.; Emanuele, R.M.; Coon, J.S. Reversal of Multi-Drug Resistance In Vitro by Fatty Acid-PEG-Fatty Acid Diesters. *Int. J. Cancer* **1996**, 65 (1), 74–79.
89. Kimura, T.; Yamaguchi, T.; Usuki, K.; Kurosaki, Y.; Nakayama, T.; Fujiwara, Y.; Matsuda, Y.; Unno, K.; Suzuki, T. Colonic Mucosa-Specific "Pro-antidrug" For Oral Treatment of Ulcerative Colitis: Design, Synthesis and Fate of Methyl 20-Glucopyranosyloxyprednisolones. *J. Controlled Release* **1994**, 30 (2), 125–135.
90. Bodor, N.; Murakami, T.; Wu, W.M. Soft Drugs 18. Oral and Rectal Delivery of Loteprednol Etaborate, a Novel Soft Corticosteroid, in Rats—for Safer Treatment of Gastrointestinal Inflammation. *Pharm. Res.* **1995**, 12 (6), 869–874.
91. Otagiri, M.; Imai, T.; Fukuhara, A. Improving the Pharmacokinetic and Pharmacodynamic Properties of a Drug by Chemical Conversion to a Chimera Drug. *J. Controlled Release* **1999**, 62 (1,2), 223–229.
92. Krówczyński, L. Oral Extended-Release Dosage Forms: Principles of Technology. *Extended Release Dosage Forms*; Krówczyński, L., Ed.; CRC Press, Inc.: Boca Raton, FL, 1987; 97–158.
93. Lee, V.H.L.; Yamamoto, A.; Buur, A.; Bundgaard, H. Use of Prodrugs to Shift the Absorption Site of 5-Fluorouracil

- from the Upper to the Lower Region of the Gastrointestinal Tract of the Albino Rabbit. *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **1989**, *16*, 56–57.
94. Borchardt, R.T. Optimizing Oral Absorption of Peptides Using Prodrug Strategies. *J. Controlled Release* **1999**, *62* (1, 2), 231–238.
 95. Aungst, B.J. Intestinal Permeation Enhancers. *J. Pharm. Sci.* **2000**, *89* (4), 429–442.
 96. LeCluyse, E.L.; Sutton, S.C. In Vitro Models for Selection of Development Candidates: Permeability Studies to Define Mechanisms of Absorption Enhancement. *Adv. Drug Del. Rev.* **1997**, *23* (1–3), 163–183.
 97. Eichman, J.D.; Robinson, J.R. Mechanistic Studies on Effervescent-Induced Permeability Enhancement. *Pharm. Res.* **1998**, *15* (6), 925–930.
 98. Tomita, M.; Hayashi, M.; Horie, T.; Ishizawa, T.; Awazu, S. Enhancement of Colonic Drug Absorption by the Transcellular Permeation Route. *Pharm. Res.* **1988**, *5* (12), 786–789.
 99. Ganem-Quintanar, A.; Falson-Rieg, F.; Buri, P. Contribution of Lipid Components to the Permeability Barrier of Oral Mucosa. *Eur. J. Pharm. Biopharm.* **1997**, *44* (2), 107–120.
 100. Nerurkar, M.M.; Burton, P.S.; Borchardt, R.T. The Use of Surfactants to Enhance the Permeability of Peptides Through Caco-2 Cells by Inhibition of An Apically Polarized Efflux System. *Pharm. Res.* **1996**, *13* (4), 528–534.
 101. Shono, Y.; Fujita, T.; Yamamoto, A. Improvement of Intestinal Absorption of Drugs As *P*-glycoprotein Substrate by Various Surfactants. *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **1998**, *25*, 677–678.
 102. Dudeja, P.K.; Anderson, K.M.; Harris, J.S.; Buckingham, L.; Coon, J.S. Reversal of Multidrug Resistance Phenotype by Surfactants: Relationship to Membrane Lipid Fluidity. *Arch. Biochem. Biophys.* **1995**, *319* (1), 309–315.
 103. Bernkop-Schnürch, A. The Use of Inhibitory Agents to Overcome the Enzymatic Barrier to Perorally Administered Therapeutic Peptides and Proteins. *J. Controlled Release* **1998**, *52* (1), 1–16.
 104. Wu, C.Y.; Benet, L.Z.; Hebert, M.F.; Gupta, S.K.; Rowland, M.; Gomez, D.Y.; Wachter, V.J. Differentiation of Absorption and First-Pass Gut and Hepatic Metabolism in Humans: Studies With Cyclosporine. *Clin. Pharmacol. Ther.* **1995**, *58* (5), 492–497.
 105. Gomez, D.Y.; Wachter, V.J.; Tomlanovich, S.J.; Hebert, M.F.; Benet, L.Z. The Effects of Ketoconazole on the Intestinal Metabolism and Bioavailability of Cyclosporine. *Clin. Pharmacol. Ther.* **1995**, *58* (1), 15–19.
 106. Ku, Y.M.; Min, D.I.; Flanigan, M. Effect of Grapefruit Juice on the Pharmacokinetics of Microemulsion Cyclosporine and Its Metabolite in Healthy Volunteers: Does the Formulation Difference Matter. *J. Clin. Pharmacol.* **1998**, *38* (10), 959–965.
 107. Soldner, A.; Christians, U.; Susanto, M.; Wachter, V.J.; Silverman, J.A.; Benet, L.Z. Grapefruit Juice Activates *P*-Glycoprotein-Mediated Drug Transport. *Pharm. Res.* **1999**, *16* (4), 478–485.
 108. Freeman, D.J.; Grant, D.R.; Carruthers, S.G. The Cyclosporin–Erythromycin Interaction: Impaired First-Pass Metabolism in the Pig. *Br. J. Pharmacol.* **1991**, *103* (3), 1709–1712.
 109. Lown, K.; Kolars, J.; Turgeon, K.; Merion, R.; Wrighton, S.A.; Watkins, P.B. The Erythromycin Breath Test Selectively Measures P450IIIa in Patients With Severe Liver Disease. *Clin. Pharmacol. Ther.* **1992**, *51* (3), 229–238.
 110. Walter-Sack, I.; Klotz, U. Influence of Diet and Nutritional Status on Drug Metabolism. *Clin. Pharmacokinetic.* **1996**, *31* (1), 47–64.
 111. Wilkinson, G.R. The Effects of Diet, Aging and Disease-States on Presystemic Elimination and Oral Drug Bioavailability in Humans. *Adv. Drug Del. Rev.* **1997**, *27* (2, 3), 129–159.
 112. Neubert, R. Ion Pair Transport Across Membranes. *Pharm. Res.* **1989**, *6* (9), 743–747.
 113. Pouton, C.W.; Charman, W.N. Oily Formulations for Administration to the Gastrointestinal Tract. *Adv. Drug Del. Rev.* **1997**, *25* (1), 1–128.
 114. Palin, K.J. Lipids and Oral Drug Delivery. *Pharm. Int.* **1985**, *6* (1), 272–275.
 115. Zhi, J.; Rakhit, A.; Patel, I.H. Effects of Dietary Fat on Drug Absorption. *Clin. Pharmacol. Ther.* **1995**, *58* (5), 487–491.
 116. Barnwell, S.G.; Burns, S.J.; Higginbottom, S.; Whelan, I.; Corness, D.; Hay, G.; Rosenberg, E.; Attwood, D. Demonstration of the Importance of Biphasic Oleic Acid Delivery for Enhancing the Bioavailability of Propranolol in Healthy Volunteers. *Int. J. Pharm.* **1996**, *128* (1, 2), 145–154.
 117. Wachter, V.J.; Silverman, J.A.; Zhang, Y.; Benet, L.Z. Role Of *P*-Glycoprotein and Cytochrome P450 3A in Limiting Oral Absorption of Peptides and Peptidomimetics. *J. Pharm. Sci.* **1998**, *87* (11), 1322–1330.
 118. Buckingham, L.E.; Balasubramanian, M.; Emanuele, R.M.; Clodfelter, K.E.; Coon, J.S. Comparison of Solutol HS 15, Cremophor EL and Novel Ethoxylated Fatty Acid Surfactants As Multidrug Resistance Modification Agents. *Int. J. Cancer* **1995**, *62* (4), 436–442.
 119. Bates, S.E.; Currier, S.J.; Alvarez, M.; Fojo, A.T. Modulation Of *P*-Glycoprotein Phosphorylation and Drug Transport by Sodium Butyrate. *Biochemistry* **1992**, *31* (28), 6366–6372.
 120. Chang, T.; Benet, L.Z.; Hebert, M.F. Effect of Water-Soluble Vitamin E on Cyclosporine Pharmacokinetics in Healthy Volunteers. *Clin. Pharmacol. Ther.* **1996**, *59* (3), 297–303.
 121. Pan, S.H.; Lopez, R.R., Jr.; Sher, L.S.; Hoffman, A.L.; Podesta, L.G.; Makowka, L.; Rosenthal, P. Enhanced Oral Cyclosporine Absorption with Water-Soluble Vitamin E Early After Liver Transplantation. *Pharmacotherapy* **1996**, *16* (1), 59–65.
 122. Khan, M.Z.I. Recent Trends and Progress in Sustained or Controlled Oral Delivery of Some Water Soluble Drugs: Morphine Salts, Diltiazem and Captopril. *Drug Dev. Ind. Pharm.* **1995**, *21* (9), 1037–1070.
 123. Perez, M.A.; Ghaly, E.S.; Marti, A. Sustained Release Phenylpropanolamine Hydrochloride from ATO 888 Matrix. *P. R. Health Sci. J.* **1993**, *12* (4), 263–267.

124. Gai, M.N.; Isla, A.; Andonaegui, M.T.; Thielemann, A.M.; Seitz, C. Evaluation of the Effect of 3 Different Diets on the Bioavailability of 2 Sustained Release Theophylline Matrix Tablets. *Int. J. Clin. Pharmacol. Ther.* **1997**, *35* (12), 565–571.
125. Ishino, R.; Yoshino, H.; Hirakawa, Y.; Noda, K. Absorption of Diltiazem in Beagle Dog from Pulsatile Release Tablet. *Chem. Pharm. Bull.* **1992**, *40* (11), 3094–3096.
126. Seta, Y.; Higuchi, F.; Otsuka, T.; Kawahara, Y.; Nishimura, K.; Okada, R.; Koike, H. Preparation and Pharmacological Evaluation of Captopril Sustained-Release Dosage Forms Using Oily Semisolid Matrix. *Int. J. Pharm.* **1988**, *41* (3), 255–262.
127. Ritschel, W.A. Microemulsion Technology in the Reformulation of Cyclosporine: The Reason Behind the Pharmacokinetic Properties of Neoral. *Clin. Transplant* **1996**, *10* (4), 364–373.
128. Rogers, J.A.; Anderson, K.E. The Potential of Liposomes in Oral Drug Delivery. *Crit. Rev. Ther. Drug Carrier Syst.* **1998**, *15* (5), 421–480.
129. Tenjarla, S. Microemulsions: An Overview and Pharmaceutical Applications. *Crit. Rev. Ther. Drug Carrier Syst.* **1999**, *16* (5), 461–521.
130. Bouwstra, J.A.; Hofland, H.E.J. Niosomes. *Colloidal Drug Delivery Systems*; Kreuter, J., Ed.; Marcel Dekker, Inc.: New York, 1994; 191–217.
131. Barnwell S.G. Biphasic Release Formulations for Lipophilic Acids, US Patent 5,391,377 Feb 21, 1995.
132. Rashid A.; Stevens H.N.E.; Binns J.S.; Mackie J.L. Controlled Release Device. U S Patent 5,750,143 May 12, 1998.
133. Quintanar-Guerrero, D.; Allémann, E.; Fessi, H.; Doelker, E. Applications of the Ion-Pair Concept to Hydrophilic Substances with Special Emphasis on Peptides. *Pharm. Res.* **1997**, *14* (2), 119–127.
134. Sarciaux, J.M.; Acar, L.; Sado, P.A. Using Microemulsion Formulations for Oral Drug Delivery of Therapeutic Peptides. *Int. J. Pharm.* **1995**, *120* (2), 127–136.
135. Allémann, E.; Leroux, J.-C.; Gurny, R. Polymeric Nano- and Microparticles for the Oral Delivery of Peptides and Peptidomimetics. *Adv. Drug Del. Rev.* **1998**, *34* (2, 3), 171–189.
136. Aungst, B.J.; Saitoh, H.; Burcham, D.L.; Huang, S.-M.; Mousa, S.A.; Hussain, M.A. Enhancement of Intestinal Absorption of Peptides and Non-Peptides. *J. Controlled Release* **1996**, *41* (1), 19–31.
137. Rowe, C.W.; Katstra, W.E.; Palazzolo, R.D.; Giritlioglu, B.; Teung, P.; Cima, M.J. Multimechanism Oral Dosage Forms Fabricated by Three Dimensional PrintingTM. *J. Controlled Release* **2000**, *66* (1), 11–17.

FURTHER READING

- Couvreux, P.; Duchene, D.; Kalles, I., Eds. *Formulation of Poorly Available Drugs for Oral Administration*, Editions de Santé: Paris, 1996.
- de Boer, A.G., Ed. *Drug Absorption Enhancement: Concepts, Possibilities, Limitations and Trends* Harwood Academic: Switzerland, 1994.
- Gurny, R., Junginger, H.E., Peppas, N.A., Eds. *Pulsatile Drug Delivery: Current Applications and Future Trends*; Wissenschaftliche Verlagsgesellschaft: Stuttgart Germany, 1993.
- Hardy, J.G., Davis, S.S., Wilson, C.G., Eds. *Drug Delivery to the Gastrointestinal Tract*; Ellis Horwood Ltd.: Chichester UK, 1989.
- Hui, H.W.; Robinson, J.R.; Lee, V.H.L. Design and Fabrication of Oral Controlled Release Drug Delivery Systems. *Controlled Drug Delivery: Fundamentals and Applications*, 2nd Ed.; Robinson, J.R., Lee, V.H.L., Eds.; Marcel Dekker, Inc.: New York, 1987; 373–432.
- Johnson, L.R., Ed. *Physiology of the Gastrointestinal Tract*, 3rd Ed.; Raven Press: New York, 1994.

DRUG DELIVERY—ORAL COLON-SPECIFIC

Vincent H.L. Lee
Suman K. Mukherjee

University of Southern California, Los Angeles, California, U.S.A.

INTRODUCTION

Historically, oral ingestion has been the most convenient and commonly used method of drug delivery. For sustained- as well as controlled-release systems, the oral route of administration has received the most attention. This is because of greater flexibility in dosage form design for the oral rather than the parenteral route. Patient acceptance of oral administration of drugs is quite high. It is a relatively safe route of drug administration compared with most parenteral forms, and the constraints of sterility and potential damage at the site of administration are minimal.

Colon-specific drug-delivery systems offer several potential therapeutic advantages. In a number of colonic diseases such as colorectal cancer, Crohn's disease, and spastic colon, it has been shown that local is more effective than systemic delivery (1). Colonic drug delivery can be achieved by oral or by rectal administration. Rectal delivery forms (suppositories and enemas) are not always effective because a high variability is observed in the distribution of drugs administered by this route. Suppositories are effective in the rectum because of the confined spread and enema solutions can only be applied topically to treat diseases of the sigmoid and the descending colon. Therefore, the oral route is preferred. Absorption and degradation of the active ingredient in the upper part of the gastrointestinal tract is the major obstacle with the delivery of drugs by the oral route and must be overcome for successful colonic drug delivery.

Drugs for which the colon is a potential absorption site (for example, peptides and proteins) can be delivered to this region for subsequent systemic absorption. The digestive enzymes of the gastrointestinal tract generally degrade these agents. However, these enzymes are present in significantly lower amounts in the colon compared with the upper portion of the gastrointestinal tract (2).

Colon-specific drug delivery has been attempted in a number of ways that primarily seek to exploit the changes in the physiological parameters along the gastrointestinal tract (1, 3). These approaches include the use of prodrugs, pH-sensitive polymers, bacterial degradable polymers,

hydrogels and matrices, and multicoating time-dependent delivery systems. The use of these strategies to target drug delivery to the colon are addressed in detail.

ANATOMICAL, PHYSIOLOGICAL, AND BIOCHEMICAL CHARACTERISTICS OF THE COLON

In terms of size and complexity, the human colon falls between that of carnivores such as the ferret, which has no discernable junction between the ileum and colon, and herbivores such as the rat, which has a voluminous cecum. The cecum, colon ascendens, colon transversale, colon descendens, and rectosigmoid colon make up the colon. It is approximately 1.5 m long, with the transverse colon being the largest and the most mobile part (Fig. 1) (4). Unlike the small intestine, the colon does not have any villi. However, because of the presence of plicae semilunares, which are crescentic folds, the intestinal surface of the colon is increased to approximately 1300 cm² (3). The physiology of the proximal and distal colon differs in several respects that can have an effect on drug absorption at each site. The physical properties of the luminal content of the colon also change, from liquid in the cecum to semisolid in the distal colon.

The colon serves four major functions. They are: 1) creation of a suitable environment for the growth of colonic microorganisms such as *Bacteroids*, *Eubacterium*, and *Enterobacteriaceae*; 2) storage reservoir of fecal contents; 3) expulsion of the contents of the colon at a suitable time; and 4) absorption of water and Na⁺ from the lumen, concentrating the fecal content, and secretion of K⁺ and HCO₃⁻ (5). The active secretion of K⁺ is stimulated by mineralocorticoids.

The factors that affect absorption from the colon are given in Table 1. Colon-specific drug delivery is primarily dependent on two physiological factors. These are the pH level and the transit time.

The pH of the gastrointestinal tract is subject to both inter- and intrasubject variations. Diet, diseased state, and food intake influence the pH of the gastrointestinal fluid.

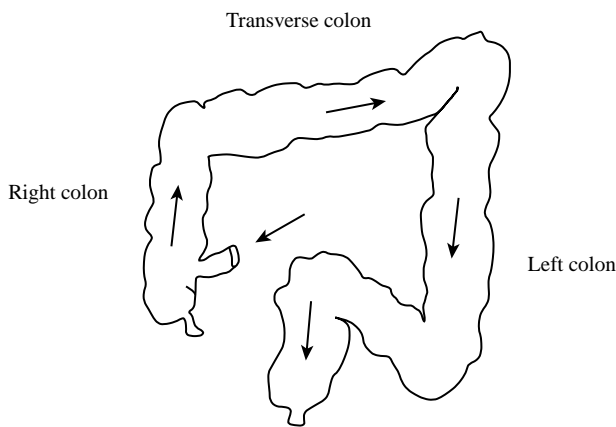


Fig. 1 Right, transverse, and left colon. (Adapted from Ref. 9.)

The change in pH along the gastrointestinal tract has been used as a means for targeted colon drug delivery (6). This can be achieved by using coatings that are intact at the low pH of the stomach but that will dissolve at neutral pH. The right, mid-, and left colon have pH values of approximately 6.4, 6.6, and 7.0, respectively (7). The pH of the colon is often lower than the pH of the small intestine, which can be as high as 8 or 9 (7). In vitro experiments simulating the pH of the different regions of the gastrointestinal tract suggest that drug release from delivery systems that are triggered by a change in pH might already occur in the small intestine (8). However, other studies have concluded the pH can be used for drug release in the colon (6). The contradictions may be a result of nonstandardized in vitro methods as well as attributable to the difference in drug properties and preparation methods of drug-delivery systems. Interspecies variability in pH is a major concern when developing and testing colon-specific delivery systems in animals and applying the information to humans (4).

Drug delivery to the colon via the oral route depends on the gastric emptying and small bowel transit time. Drugs taken before meals usually pass out of the stomach within

1 h, but can take up to 10 h if taken after a meal. The transit time in the small intestine is relatively constant, ranging between 3 and 4 h, regardless of various conditions such as physical state, size of dosage form, and presence of food in the stomach (9).

One of the major determinants of the absorption of a compound from the colon is residence in any particular segment of the colon. The time taken for food to pass through the colon accounts for most of the time the food is present in the gut. In healthy subjects, this is approximately 78 h for expulsion of 50% ingested markers but may vary from 18 to 144 h. Steady-state measurements of mean transit time after ingestion of markers for several days gave a mean transit time of 54.2 h (10).

When the dosage forms reach the colon, transit depends on size. Small particles pass through the colon (236 min for 3-mm tablets and 238 min for 6-mm tablets on study day 1; 226 min for 9-mm tablets and 229 min for 6-mm tablets on study day 2; 218 min for 12-mm tablets and 219 min for 6-mm tablets on study day 3) more slowly than the larger units (11). However, the density and size of larger single units had no real effect on colonic transit. It has been shown that pellets move faster than do tablets through the ascending colon and therefore may be more favorable than tablets with respect to colonic drug absorption (11).

Colonic transit time is only slightly affected by food but is reduced under stress. Studies have shown that drugs that act on the parasympathetic or sympathetic nervous system affect the propulsive motor activity, thereby influencing the colonic transit time (12). Although not significantly affected by most disease (13), the transit time is shorter in patients who complain of diarrhea and longer in patients with constipation.

PERMEABILITY AND METABOLIC CHARACTERISTICS

Drug Permeability in the Colon

Drug absorption in the colon is restricted by a number of barriers. The bacterial flora, which is significantly higher in the distal colon, can affect drug bioavailability (14). The mucus present at the epithelial surface of the colon acts as a physical barrier to uptake attributable to specific and nonspecific drug binding. Drug binding to the colonic lumen may facilitate enzymatic or environmental degradation by increasing the residence time of the drug. Cephalosporins, penicillins, and aminoglycosides are some of the small molecule drugs that bind to the negatively charged mucus (15). This binding can prevent

Table 1 Factors affecting drug absorption from the colon

1. Physical characteristics of drug (pK_a , degree of ionization)
2. Colonic residence time as dictated by gastrointestinal tract motility
3. Degradation by bacterial enzymes and byproducts
4. Selective and nonselective binding to mucus
5. Local physiological action of drug
6. Disease state
7. Use of chemical absorption enhancers, enzyme inhibitors, or bioadhesives

(From Ref. 8.)

drugs from reaching the epithelial surface as well. Some molecules such as carbachol and serotonin that stimulate mucus secretion can, therefore, reduce their own absorption as well as the absorption of any coadministered drug (16). The removal of this mucus barrier with mucolytic agents can result in increased absorption. Schipper et al. showed that chitosans had a pronounced effect on the permeability of mucus-free Caco-2 layers and enhanced the permeation of atenolol 10- and 15-fold (17). The binding of chitosan to the epithelial surface and subsequent absorption-enhancing effects were significantly reduced in mucus-covered HT-29-H cells. Alterations of this mucus layer, however, have been associated with pathological conditions.

The single most formidable barrier to the uptake of drugs occurs at the level of the epithelium (18). The lipid bilayer of the individual colonocytes and the occluding junctional complex (OJC) between these cells provide a physical barrier to drug absorption. OJC structures found at the apical neck of the cells in the gastrointestinal tract limit paracellular transport of water and small molecules (<350 Da) in a cation-selective manner (19). The movement of water and ions via the paracellular route is driven by transcellular ionic gradients, with the proximal colon appearing to be slightly more conducive to such activity than the distal colon (20).

Carrier-mediated absorption in the colon is not extensive and usually relates to the metabolic events in the residual bacteria (21). Drugs that pass from the apical to the basolateral side of this epithelial layer can do so either by the transcellular route (i.e., by passing directly through the colonocytes) or by the paracellular route (i.e., by passing through adjacent colonocytes). Mathematical models describing the transport of small drug molecules across these barriers have been developed (22). The movement of peptides is predominantly dependent on hydrogen bonding (23). Compared with peptides, proteins experience a greater thermodynamic barrier for passing through cell membranes and are more vulnerable to denaturation at the interface.

Drug molecules that have traversed the physical and enzymatic barriers of the colonic mucosa may enter either the blood capillary bed or the lymphatic sinuses. Intact drugs that reach the venous capillaries from the submucosa are transported to the liver via the hepatic-portal system where they may undergo significant metabolism. On the other hand, uptake into the lymphatic sinuses of the colon results in direct delivery into the systemic circulation that causes lesser metabolic breakdown of the absorbed drug (3).

Drug Metabolism in the Colon

Metabolic reactions in the liver and the small intestine are well documented (24). However, only sparse information is available on drug metabolism in the colon. Drug metabolism in the colon can be brought about by the host enzymes in the epithelial cells or by the microbial enzymes in the gut flora. Metabolic activities in the wall of the colon can be attributed to cytochrome P450, esterases, amidases, and various transferases (25). Reductive drug metabolism does not appear to be important at this site.

The colonic mucosa resembles the small intestinal mucosa with respect to the spectrum of metabolizing enzymes (26). However, the total metabolic capacity of the colonic wall is inferior, because the mucosal mass in the lower part of the intestine is several times smaller than that in the upper part. This may be more than offset by the high metabolic capacity found in the gut flora in the large intestine.

The gastrointestinal tract contains a variety of microorganisms that participate in the metabolism of ingested material (Tables 2 and 3) (27). The upper part contains only a small number of primarily gram-positive bacteria, whereas the colon contains a vast amount of primarily anaerobic bacteria, with *Bacteriodes*, *Bifidobacterium* and *Eubacterium* being the most common species found (28). Gram-positive cocci such as *Clostridium enterococci* and various species of *Enterobacteriaceae* are also present (29). The growth of the bacteria is regulated

Table 2 Drug metabolizing enzymes in the human colon that catalyze reductive reactions

Enzymes	Microorganism	Metabolic reaction catalyzed
Nitroreductase	<i>E. coli</i> , <i>Bacterioids</i>	Reduce aromatic and heterocyclic nitro compounds
Azoreductase	<i>Clostridia</i> , <i>Lactobacilli</i> , <i>E. Coli</i>	Reductive cleavage of azo compounds
N-Oxide reductase, sulfoxide reductase	<i>E. coli</i>	Reduce N-Oxides and sulfoxides
Hydrogenase	<i>Clostridia</i> , <i>Lactobacilli</i>	Reduce carbonyl groups and aliphatic double bonds

Table 3 Drug metabolizing enzymes in the colon that catalyze hydrolytic reactions

Enzymes	Microorganism	Metabolic reaction catalyzed
Esterases and amidases	<i>E. coli</i> , <i>P. vulgaris</i> , <i>B. subtilis</i> , <i>B. mycoides</i>	Cleavage of esters or amidases of carboxylic acids
Glucosidase	<i>Clostridia</i> , <i>Eubacteria</i>	Cleavage of β -glycosidases of alcohols and phenols
Glucuronidase	<i>E. coli</i> , <i>A. aerogenes</i>	Cleavage of β -glucuronidases of alcohols and phenols
Sulfatase	<i>Eubacteria</i> , <i>Clostridia</i> , <i>Streptococci</i>	Cleavage of <i>O</i> -sulfates and sulfamates

by gastric acids, peristaltic activity, and microbial interaction including bacterial metabolic byproducts. Administration of antibiotics as well as onset of disease and age can affect the metabolic activity of the intestinal microflora (30).

Redox reactions and hydrolysis are the predominant metabolic conversions triggered by the intestinal microflora. The primary reductive enzymes produced by the intestinal microflora are nitroreductase, deaminase, urea dehydroxylase, and azoreductase. The hydrolytic enzymes are β -glucuronidase, β -xylosidase, β -galactosidase, and α -L-arabinosidase. Studies conducted by Macfarlane and coworkers have shown that proteolysis can also occur in the colon (31). More recent findings by the group indicate that bacterial fermentation of proteins in humans could account for 17% of the short-chain fatty acids in the cecum and for 38% in the sigmoid and the rectum (32).

The unique enzymatic features of the microbial flora present in the colon make it conducive for drug targeting. Active metabolites intended for local therapy can be liberated from poorly absorbed parent molecules by microbial processes. An advantage can also be taken for release of drugs from delivery systems, disintegrating only by the action of microbial enzymes, as a means for targeted delivery. This approach is suitable for drug molecules that are unstable in the upper part of the gastrointestinal tract but can be absorbed from the lower part.

CHARACTERISTICS OF DRUGS THAT FAVOR COLONIC DELIVERY

The colon is a less hostile environment than the stomach and the small intestine (33). It has a longer retention time and is responsive to agents that cause an increase in the absorption of poorly absorbed drugs (34). Drugs that will benefit from colon targeting include those for the treatment of inflammatory bowel disease and irritable bowel syndrome. Drugs metabolized in the upper gastrointestinal

tract also would be candidates for colon targeting. Drugs such as theophylline, nifedipine, ibuprofen, diclofenac, metoprolol, brompheniramine, pseudoephedrine, dinitrate, isosorbide, oxprenolol, and low-molecular-weight peptides and peptide-like drugs have been shown to be effectively absorbed from the colon (35).

The permeability of the colonic epithelium may not be sufficient for achieving a transport rate required for therapeutic activity. This hurdle may be overcome, at least in part, by using penetration enhancers (Table 4). They include chelating agents, nonsteroidal anti-inflammatory drugs, fatty acids, and surfactants (36). Comparison of their rate of onset and recovery of a treated mucosa has also been made (37). Fatty acids have strong and fast reactivity and allow for a fast recovery of the barrier function. Bile salts and salicylates are moderate and fast-acting agents with fast barrier function recovery. Strong surfactants and chelators have strong or moderate reactivity and a slow recovery of barrier function. Solvents such as dimethylsulfoxide and ethanol have moderate reactivity and act primarily as agents to improve drug miscibility in an aqueous environment. The aforementioned enhancers are also effective in the small intestine (22). Some

Table 4 Some common colonic drug absorption enhancers

1. Nonsteroidal antiinflammatory agents (NSAIDs) (e.g., indomethacin and salicylates)
2. Calcium ion-chelating agents (e.g., EDTA and citrate)
3. Surfactants [e.g., polyoxyethylene lauryl ether (BL-9EX) and saponin]
4. Bile salts (e.g., taurocholate and glycocholate)
5. Fatty acids (e.g., sodium caprate, sodium caprylate, sodium laurate, and sodium oleate)
6. Mixed micelles [e.g., monoolein-taurocholate, oleic acid-taurocholate, oleic acid-polyoxyethylene hydrogenated castor oil (HCO 60) and oleic acid-glycocholate]
7. Other agents [e.g., acylcarnitine, azone (1-dodecylazacycloheptan-2-one), dicarboxylic acids and enamine]

enhancers that are more colon-specific include ethyl-acetoacetate, which must be first metabolically transformed to enamine (38).

Several chemical enhancers such as sodium ethylenediaminetetraacetate and sodium taurocholate, oleic acid (37), polyoxyethylated nonionic surfactants, citric acid, and dihydroxy bile salts open the paracellular route presumably by disruption of intracellular OJC function (37). The use of nitric oxide is another approach (39). Moreover, manipulation of the cyclooxygenase pathway activities to trigger the release of compounds like substance P may provide another means for altering colonic permeability (40).

Molecules such as sodium caprate, sodium caprylate, and sodium salicylate can enhance transcellular uptake of poorly permeable compounds through the colonic mucosa in low concentrations (41, 42). Some of these absorption enhancers can cause local irritation as well (43). It has been suggested that these enhancers might function to denature membrane proteins and/or modify their lipid-protein interactions as a means of inducing drug uptake (37). Mixed micelles appear to produce only limited disordering of surface mucosal cells, possibly by reducing the damaging effect of surfactants and somehow augmenting their enhancing activity. Absorption enhancements by surfactants, fatty acids, and mixed micelles may be attributable to improved solubility and/or stability of the drug being transported.

Stable molecules such as dextran and polyethylene glycol have been used in studies to address the size of the transport window that can be opened in the colon. Muranishi and Takada (44) have demonstrated that, on transport enhancement, low-molecular-weight drugs are directed twice as often to a transcellular route than the paracellular route, whereas molecules of approximately 20 kDa are directed almost equally through transcellular and paracellular routes. Enhanced transport with caprate, laurate, and mixed micelles suggest the opening of colonic pores of 14–16 Å, whereas enhancement with taurocholate or caprylate suggests the formation of 11–12 Å pores (37).

Chemical enhancers have been demonstrated to produce transport windows in colonic epithelia large enough for the passage of many bacterial toxins. Patients suffering from inflammatory bowel diseases and colitis typically have increased colonic permeability (45) attributable to bacterial toxins, both enterotoxins and cytotoxins that increase capillary permeability. Increased colon permeability associated with a diseased state may be useful in the treatment, in which improvement of the condition might reduce mucosal permeability and naturally reduce drug transport.

STRATEGIES FOR COLON-SPECIFIC DRUG DELIVERY

Attempts have been made to achieve colon-specific delivery of drugs. These include: 1) prodrugs and 2) polymers (enteric coated, those that are sensitive to degradation by bacterial enzymes, and matrices and hydrogels susceptible to degradation by bacterial enzymes).

Prodrugs

The prodrug sulfasalazine (SAS), used in the treatment of Crohn's disease, ulcerative colitis, and rheumatoid arthritis, is the earliest example of targeted drug delivery in the colon (Figs 2 and 3) (1). Only 12% of the drug is absorbed in the small intestine after oral delivery. When SAS reaches the colon after oral administration, the diazoreductase of the colon bacteria cleaves the azo bond, releasing 5-aminosalicylic acid and sulfapyridine into the colonic lumen (46). Osalazine (dipentum), which consists of two molecules of 5-aminosalicylic acid linked by an azo bond (47), was developed to directly deliver 5-aminosalicylic acid to the colon. Balsalazide is another prodrug that is reputed to benefit from a less toxic carrier molecule than that in SAS. Balsalazide is 5-aminosalicylic acid linked by a diazo moiety to 4-aminobenzoyl-β-alanine (48).

A comparable total release of 5-aminosalicylic acid in rats has been observed from SAS and 5-aminosalicylic acid azo linked to a polymeric prodrug consisting of polysulfonamidoethylene as the carrier molecule (polyasa) (49). Yet, the latter was more effective in reducing inflammation in the guinea pig ulcerative colitis model.

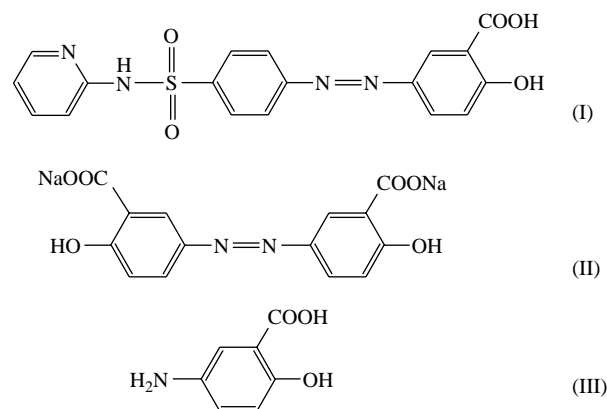


Fig. 2 Chemical structures of sulfasalazine (I), osalazine (II) and 5-ASA. (Adapted from Ref. 1.)

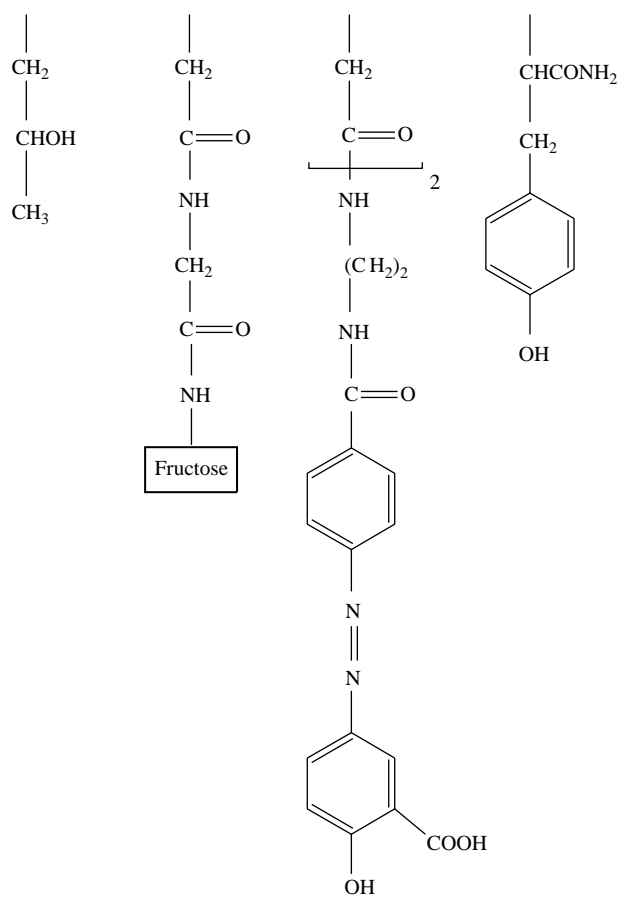


Fig. 3 Polymeric prodrug of 5-ASA. (Adapted from Ref. 1.)

The release of 5-aminosalicylic acid from polymeric prodrugs was dependent on the structure of the polymeric backbone (50). The drawback of coupling 5-aminosalicylic acid to polymeric backbone is the large amount of the drug that has to be taken orally.

A second approach that has been attempted involves prominent bacterial enzymes such as glycosidases and polysaccharides (51). Prodrugs were developed by coupling glucose or galactose to steroids such as dexamethasone, prednisolone, hydrocortisone, and fludrocortisone via a β -glycosidic bond. When these polar prodrugs are taken orally, they undergo minimal absorption in the small intestine. In the colon, the bacterial glycosidases cleave the polar moiety and release the steroid. Corticosteroids have been delivered in this manner to treat disorders of the large intestine. The side effects of corticosteroids can be minimized with this approach because of the minimal absorption of the drug in the small intestine as well as by dose reduction. The oral delivery of the prodrug form of dexamethasone produced the same therapeutic effect as

the parent drug in guinea pig models. However, the dose was lower with the prodrug (52). This approach has been extended to the development of glycoside prodrugs of budesonide and menthol and with dexamethasone poly (L-aspartate) (53).

The prodrug approach has been used where naproxen was conjugated to dextran by an ester linkage (54). The release of naproxen was 17 times higher in the cecum and in the colon homogenates of the pig than in control medium or homogenates of the small intestine. The dextran prodrugs of methylprednisolone and dexamethasone have also been developed (55). In both cases, most of the drugs were released in the contents of the large intestine, whereas only minor chemical hydrolysis occurred in the upper gastrointestinal tract.

Recent developments involve use of cyclodextrins that are absorbed only in minor quantities in the small intestine but are fermented by the colonic bacteria. In recent studies (56, 57), the use of biphenyl acetic acid conjugates of β -cyclodextrins was described. The ester conjugate released the drug preferentially when incubated with rat cecal contents, and almost no release was observed on incubation with the contents of the stomach or the small intestine.

Therefore, it can be seen that prodrug-based approaches are being used extensively to deliver drugs in increased amounts to the colon. Consistency of conversion from the prodrug to the active ingredients in vivo is a concern for these approaches because colonic bacterial populations can be modulated by diet, health, and a variety of antibiotics (29).

Polymers

Enteric coated

Drugs can be delivered locally and selectively to the colon by taking advantage of the difference in the pH of the different regions of the gastrointestinal tract. The pH in the gastrointestinal tract is low in the stomach but increases in the small intestine and the large intestine. For the purpose of targeting drugs to the colon, an outer enteric coating (for example, of cellulose acetate phthalate) can be used to protect the drug in the low pH of the stomach. In the small intestine (pH, 7.5), the enteric coating dissolves, exposing a polymeric coating, typically of ethylcellulose with microcrystalline cellulose and plasticizers. The polymeric coating is designed to stay intact in the small intestine until the dosage form reaches the colon. In the colon, bacteria are expected to digest the microcrystalline cellulose to allow for the disintegration of the polymeric coating around the drug. Such a scenario has been confirmed in a

study by Hirayama et al. (56), in which all the compressed tablets with dual coating remained intact in the small intestine and 85% of which disintegrated in the colon.

The problem with this approach is that the intestinal pH may not be stable because it is affected by diet, disease and presence of fatty acids, carbon dioxide, and other fermentation products. Moreover, there is considerable difference in inter- and intraindividual gastrointestinal tract pH, and this causes a major problem in reproducible drug delivery to the large intestine (58).

Spherical pellets containing 5% triamcinolone acetonide were prepared by Villar-Lopez and coworkers (59) by extrusion/spheronization after formulation with microcrystalline cellulose and/or a hydrophilic excipient such as lactose, sodium carboxymethylcellulose, or β -cyclodextrin. Their suitability for coating, with a view toward colonic drug delivery, was assessed in terms of their size, sphericity, and dissolution test response. The best results were afforded by a 5:90:5 composition of microcrystalline cellulose, β -cyclodextrin, and triamcinolone acetonide, prepared by complexation of triamcinolone acetonide with β -cyclodextrin before the addition of microcrystalline cellulose.

Formulations of 5-aminosalicylic acid that are commercially available use enteric coatings of pH-sensitive methacrylic resins called Eudragit[®] (Fig. 3). Both water-soluble and water-insoluble forms of Eudragit have been tested for colon targeting. Eudragit-L dissolves at a pH level above 5.6 and is used for enteric coating, whereas Eudragit-S, which dissolves at a pH level above 7.0 (attributable to the presence of higher amounts of esterified groups in relation to carboxylic groups) is used for colon targeting. Studies have revealed that Eudragit-S exhibits poor site specificity (60).

In a study performed by Khan et al. (61), lactose placebo tablets were coated using different combinations of Eudragit L and Eudragit S. The Eudragit L–Eudragit S combinations (w/w) studied were 1:0, 4:1, 3:2, 1:1, 2:3, 1:4, 1:5, and 0:1. The disintegration data obtained from the placebo tablets demonstrate that disintegration rate of the tablets is dependent on 1) the polymer combination used to coat the tablets, 2) the pH of the disintegration media, and 3) the coating level of the tablets.

It has been shown that polymers with nonesterified phthalic acid groups dissolve much faster and at a lower pH than those with acrylic or methacrylic groups. The presence of plasticizer and the nature of the salts in the dissolution medium influence the dissolution rate (62).

In a recent study by Peeters and Kinget (63), the free carboxylic groups of Eudragit-S were partially methylated. The product was found to dissolve in water at a slightly higher pH compared with the original polymer. The

effectiveness of this product as a colon-specific coating material had been established with human volunteers using in vivo scintigraphic studies (64).

In a study by Gazzaniga and coworkers (65), an oral dosage form was developed, consisting of a core with two polymeric layers. The outer layer, which was an enteric coating, dissolved at a pH level above 5. The inner layer, made up of hydroxypropylmethylcellulose, acted as a retarding agent to delay drug release for a predetermined period. The thickness of the inner layer determined the lag time. This system was found to release drug in the colon of the rat between the 5th and 10th h.

A pulsed system, called the Time-Clock System, has been developed. The system comprises a solid dosage form coated with a hydrophobic surfactant layer to which a water-soluble polymer is attached to improve adhesion to the core (66). The thickness of the outer layer determines the time required to disperse in an aqueous environment. After the dispersion of the outer layer, the core becomes available for dispersion. An advantage is that common pharmaceutical excipients can be used to manufacture the system. Studies performed in human volunteers showed that the lag time was not affected by gastric residence time. Also, the dispersion of the hydrophobic film was not influenced either by the presence of intestinal digestive enzymes or by the mechanical action of the stomach.

Another system based on the same principle as the Time-Clock System, called the Time-Controlled-Explosion Drug-Delivery System, has also been developed (67). It contains a four-layered spherical structure, with a core containing the drug, a swelling agent, and a water-insoluble polymer membrane made of ethylcellulose. This system is characterized by rapid drug release with a programmed lag time. On contact with water through the polymeric membrane, the swelling agent expands and finally explodes, leading to release of the contained drug. Drug release is not affected by pH, but the lag time is a function of the thickness of the outer polymeric membrane. A similar approach based on ethylene–vinyl acetate polymers was tested for release of the drug isosorbide-5-nitrate (68).

Ishibashi and coworkers have recently developed a Colon-Targeted Delivery Capsule based on pH sensitivity and time-release principles (Fig. 4) (69). The system contains an organic acid that is filled in a hard gelatin capsule as a pH-adjusting agent together with the drug substance. This capsule is then coated with a three-layered film consisting of an acid-soluble layer, a hydrophilic layer, and an enteric layer. After ingestion of the capsule, these layers prevent drug release until the environmental pH inside the capsule decreases by dissolution of the organic acid, upon which the enclosed drug is quickly released.

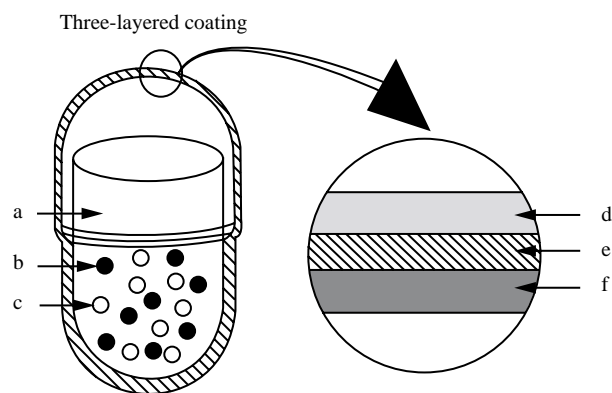


Fig. 4 Design of the colon-targeted delivery capsule: a) gelatin capsule; b) active ingredient; c) organic acid; d) enteric layer; e) hydrophilic layer; and f) acid-soluble layer. (Adapted from Ref. 106.)

Therefore, the onset time of drug release is controlled by the thickness of the acid-soluble layer. In fact, capsule disintegration (and, thus, drug release) does not start until 5 h after gastric emptying regardless of whether the formulation is administered to fasted or fed subjects.

Recently, Yoshikawa et al. (70) reported a new *in vitro* dissolution test called the rotating beads method for drugs formulated in pressure-controlled colon-delivery capsules. This dissolution method was applied to acetaminophen sustained-release tablets and two other drugs having low solubility in the colon, tegafur and 5-ASA. There was good correlation between the *in vitro* dissolution rates and the *in vivo* absorption rates.

In the development of the aforementioned time-dependent systems, care has to be taken to ensure a homogenous coating. If the coat is inhomogenous, there will be a modification of the coating rigidity, possibly leading to undesirable infiltration of the aqueous medium and, in turn, undesired alteration of the lag time before which the drug is supposed to be released.

Polymers sensitive to degradation by bacterial enzymes

Drugs can be administered locally and selectively to the colon if they are enclosed in a dosage form such as a capsule coated with an azo-aromatic cross-linked polymer subject to cleavage by azo-reductases of the colonic microflora. This approach of coating a drug with biodegradable material for colon targeting can be used to administer a large amount of the drug. Moreover, the rate of drug release is dependent on the activity of the bacterial enzymes in the colon rather than on that of the host.

A system was developed by Saffran and coworkers (68) in which insulin or vasopressin was encapsulated in a

gelatin capsule coated with an impermeable polymer. The coat, prepared by using azo functional cross-linking agents based on divinylazobenzene, was resistant to degradation in the stomach and the small intestine. However, problems were encountered attributable to variability in absorption, which may be because of intra- and intersubject differences in microbial degradation of the coating that may not be hydrophilic enough. Indeed, Kimura et al. (72) noted that only polymers with a sufficient degree of hydrophilicity could be degraded within an acceptable period of time. However, there is a possibility of premature drug release if the polymeric coating is too hydrophilic. The impact of the spacer length of the incorporated azo agent appears to be of limited importance.

A popular theory with azo materials is that their degradation products are always aromatic amines such as azo dyes. Ueda and coworkers observed that the azo bonds in segmented polyurethanes were reduced to hydrazo intermediates after incubation with human feces because no decrease in the molecular weight was observed (73). It was then theorized that drug release from pellets coated with these azo polymers was attributable to both a conformational change and a breakdown of the film structure. Other studies also concluded that the polymers were reduced to hydrazo intermediates or were completely degraded to aromatic amines depending on their hydrophilic/hydrophobic nature.

There has been no definitive conclusion regarding the toxicity of azo polymers, although it is known that azo dyes contain several potential carcinogens. To avoid possible azo-related toxicity issues, other biodegradable natural substances capable of forming coatings that degrade in the colon have been studied. Common problems encountered with these natural biodegradable materials are poor film-forming capability and excessive water solubility. Therefore, efforts are currently being made to mix these natural materials with other synthetic polymers to obtain a film-forming mixture or to derivatize them to decrease their water solubility.

Natural polysaccharides such as pectin, xylan, and guar gum are not digested in the human stomach or small intestine, but are degraded in the colon by the resident bacteria. Recent studies conducted with 5-ASA (74) and indomethacin (75) confirmed that selective delivery of these drugs to the colon can be achieved using guar gum as a carrier because guar gum protects the drugs from being released in the physiological environment of the stomach and the small intestine. The polysaccharides under active investigation for colon-specific drug delivery include pectin and its salts, chondroitin sulfate, amylose, and inulin.

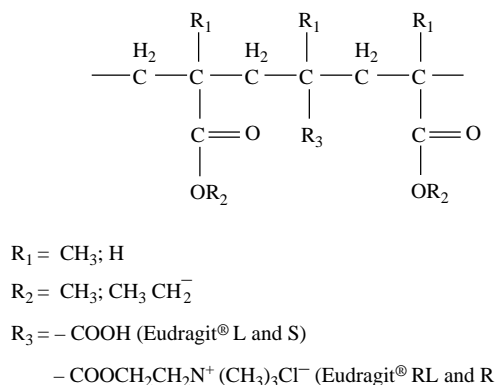


Fig. 5 Chemical structures of Eudragit. (Adapted from Ref. 69.)

Veervort and Kinget (76) demonstrated that the incorporation of inulin in Eudragit films resulted in increased permeability with increase in incubation time in the degradation medium (Fig. 5) (1). After 8, 16, and 24 h of incubation, the permeability coefficient increased with a factor 6, 20, and 70, respectively. However, because of manufacturing problems resulting from the high methoxy pectin content, film coatings were developed consisting of ethylcellulose and pectin. Wakerly and coworkers manufactured (77) film coatings with ethylcellulose and pectin. In vitro degradation studies indicated that release was controlled by the ratio of ethylcellulose and pectin.

Lorenzo-Lamosa and coworkers manufactured a system in which chitosan microcores were entrapped within acrylic microspheres of Eudragit L-100 and Eudragit S-100, forming a multireservoir system (78). This system was designed to combine the specific biodegradability enforced by colonic bacteria with pH-dependent release of the drug sodium diclofenac and tested in in vitro systems. A continuous release for 8–12 h was obtained at the pH in which the Eudragit coats were

soluble. The researchers have proposed a combined mechanism of release, comprising dissolution of the Eudragit coating, swelling of the chitosan microspheres, dissolution of the drug, and its further diffusion through the chitosan gel cores (Fig. 6) (78).

Recent studies conducted by Tozaki et al. (79) with 5-ASA-containing chitosan capsules revealed that the drug concentration in the colon was higher than that afforded by a suspension of the drug. Ramdas et al. (80) used the bioadhesiveness of polyacrylic acid, alginate, and chitosan in formulations with drugs such as 5-fluorouracil and insulin to bypass the acidity of the stomach and to release loaded drug for long periods into the intestine (80). Chitosan succinate and chitosan phthalate have been used successfully as potential matrices for the colon-specific oral delivery of sodium diclofenac as demonstrated by Aiedeh and Taha (81).

Another natural polysaccharide, amylose, when prepared under appropriate conditions, is not only able to produce films, but is also found to be resistant to the action of pancreatic α -amylase while remaining vulnerable to the colonic flora (82). However, incorporation of ethylcellulose was necessary to prevent premature drug release through simple diffusion (83). In vitro release of 5-aminosalicylic acid from pellets coated with a mixture of amylose–ethylcellulose in a ratio of 1:4 was complete after 4 h in a colonic fermenter. By contrast, it took more than 24 h to release only 20% of the drug under conditions that mimic that of the stomach and the small intestine.

A suspension of natural polygalactomannans in polymethacrylate solution applied to a degradable coating around the drug core delayed the drug release in the small intestine but was degraded by bacterial enzymes in the colon (83). This formed the basis for studying the usefulness of guar gum containing polygalactomannans as a carrier for colonic drug delivery. Matrix tablets of dexamethasone were evaluated for colon-specific drug;

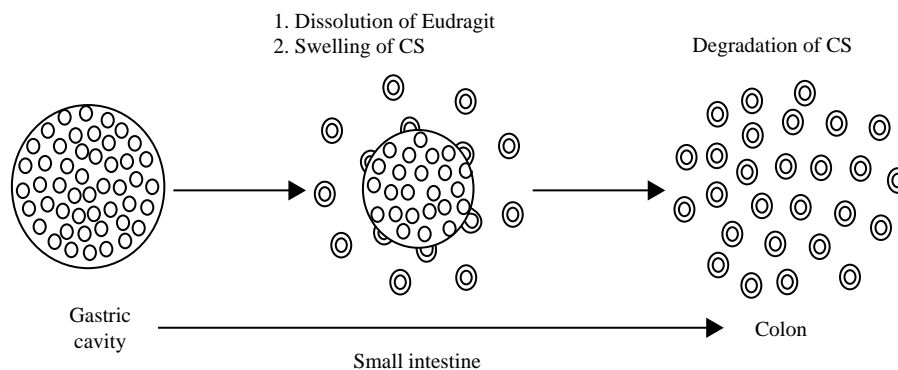


Fig. 6 Scheme of possible drug release from the Eudragit microencapsulated chitosan microspheres. (Adapted from Ref. 83.)

delivery with preparations containing guar gum (84). In a γ -scintigraphic study, guar gum in the form of matrix tablets was evaluated for its performance in healthy human volunteers using technetium-99m-DTPA as tracer (85). It was observed that the matrix tablets entered the colon intact and released the bulk of the tracer in the colon by virtue of the enzymatic action of the colonic bacteria.

Studies conducted by Bauer and Kusselhut have demonstrated that lauryl dextran esters can be film-forming and have been found to release tablets containing theophylline selectively in the colon (86). Theophylline tablets were coated with a dispersion of 4% lauroyl dextran in a study performed by Hirsch and coworkers (87). Theophylline dissolution was monitored for 4 h in a buffer of pH 5.5, after which the passage to the cecum was simulated by the addition of dextranase. Almost linear dissolution was observed during the first 4 h. The rate of release was inversely proportional to the amount of ester applied on the coating. After the addition of dextranase, the coatings were degraded, leading to the complete release of the drug in less than 2 h, after the addition of the enzyme. The results of these studies in which natural biodegradable polymers have been derivatized using acceptable reactants are promising as far as colon-specific drug delivery is concerned.

Matrix and hydrogels susceptible to degradation by bacterial enzymes

In this design, the active ingredient, the degradable polymer, and other additives are compressed to form a monolithic or multiparticulate solid dosage form. The drug is embedded in the matrix core of the degradable polymer. Biodegradable matrix systems of cross-linked chondroitin sulfate with different levels of cross-linking have been tried for the delivery of indomethacin. A direct relationship was found between the degree of cross-linking of the polymer and the amount of drug released in the rat cecal content (88).

Rubinstein and Rudai (89) observed that highly compressed matrices based on pectin in the form of plain tablets or compression-coated tablets were able to retain indomethacin in simulated gastric and intestinal juice before becoming degraded in a medium that contained enzymes for degrading pectin. In vitro experiments showed that methoxyl pectin, when added as a compression coat, was capable of protecting a core tablet under conditions mimicking mouth-to-colon transit and was susceptible to enzymatic attack in the colon. A greater degree of methoxylation of pectin resulted in lower susceptibility degradation to by colonic bacterial enzymes, whereas the presence of calcium increased the vulnerability to enzymatic attack (90).

The problem with the aforementioned monolithic unit system is that it tends to be detained at the ileocecal junction, leading to drug loss before entry in the colon. To circumvent this problem, multiparticulate dosage forms were devised that passed freely through the ileocecal junction. In a recent study, a multiparticulate system, which was based on amidated pectin, was tested (91). Coating of the amidated pectin beads with chitosan significantly reduced the release of sulfamethoxazole and indomethacin in simulated gastric and intestinal juice compared with noncoated beads.

Macleod et al. (92) have studied the potential of pectin: chitosan: hydroxypropylmethylcellulose films for colonic drug delivery (92). The results showed that in all cases, the tablets were able to pass through the stomach and small intestine intact. The tablets started to break up once they were in the colon, as a result of degradation of the coating by colonic bacteria.

Kopecek and coworkers (Fig. 7) (93) have developed novel types of hydrogel capsules, based on acrylic acid, *N,N*-dimethylacrylamide and *N*-tert-butylacrylamide cross-linked with 4,4'-di(methacryloylamino)azobenzene (94). These hydrogels did not swell significantly in the stomach. However, in transit through the small intestine, swelling increased because of increased pH. In the colon, the degree of swelling reached a threshold when the cross-links became accessible to bacterial azoreductases, which in turn, caused the breakdown of the hydrogel and release of the drug. The rate of degradation was found to be directly related to the equilibrium degree of swelling of the hydrogels and inversely proportional to the cross-linking density. Hydrogels prepared by cross-linking polymerization but having the same polymer composition and cross-link structure predominantly followed a bulk degradation-like process. In contrast, hydrogels prepared by cross-linking polymeric precursors or by a polymer-polymer reaction predominantly followed a surface erosion process at a low degree of cross-linking and a bulk degradation-like process when the degree of cross-linking increased (95). In a comparative study to determine the degradation rate of the azo functionality present in a soluble azo polymer and a hydrogel, it was observed that the degradation rate was 125 times lower than that of the soluble azo dye.

Hydrogels of dextran that used diisocyanate as the cross-linking agent were found to be capable of releasing a drug only in the distal part of the colon, where the conditions for absorption are not as conducive as the proximal part (96). pH-sensitive dextran hydrogels were prepared by activation of dextran (T-70) with 4-nitrophenyl chloroformate, followed by conjugation of the activated dextran with 4-aminobutyric acid and

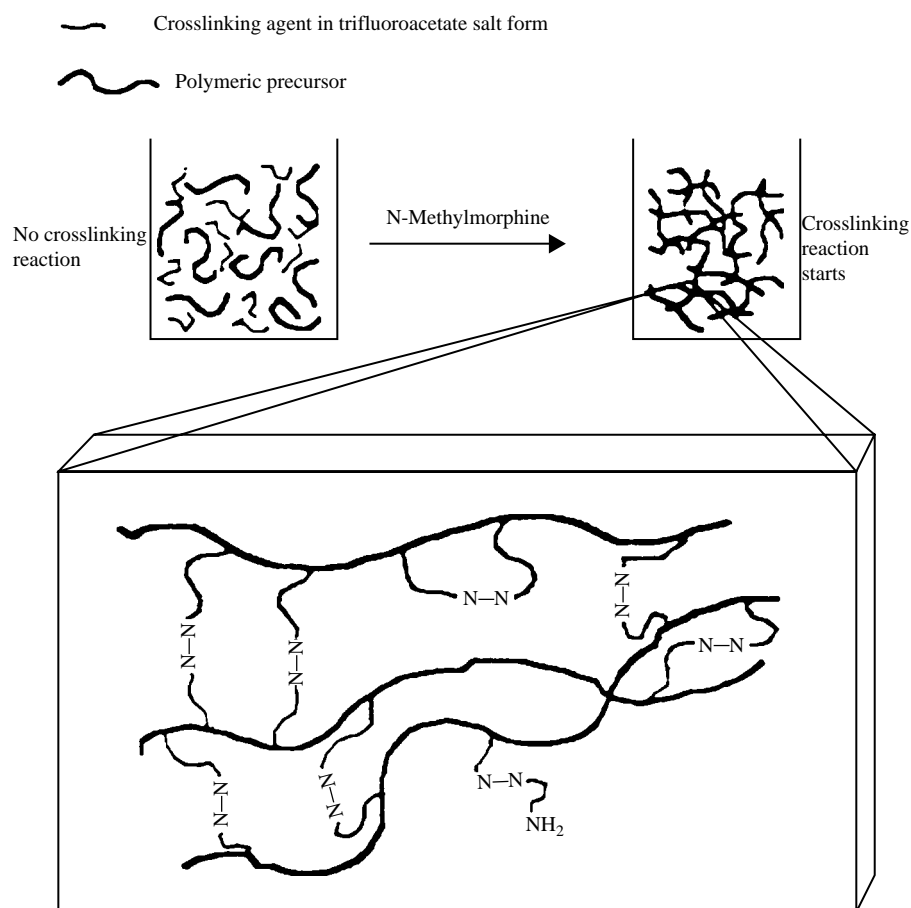


Fig. 7 Schematic representation of the synthesis of hydrogels by cross-linking of polymeric precursors. (Adapted from Ref. 93.)

cross-linking with 1,10-diaminodecane (97). The release rate of bovine serum albumin from this system was further enhanced by the addition of dextranase in buffer solutions. However, in a study in which Hirsh et al. (98) investigated lauroyldextran and cross-linked galactomannan as micro-biologically degradable film-coating materials for site-specific drug delivery to the colon, the ideal zero-order dissolution before and quick degradation after enzyme addition was not realized. On the other hand, hydrogels made by copolymerization of 2-hydroxyethyl methacrylate with 4-methacryloyl-oxyazobenzene led to the release of aniline when the hydrogel was degraded by the colonic bacterial enzymes (99). Another study reported the biodegradable properties of guar gum, which was cross-linked with borax (100). This system was found to be capable of releasing drugs in the proximal part of the colon.

Hydrogels that are based on natural products are more acceptable from the standpoint of toxicity-related issues and are therefore preferable to azo-based polymers. However, chemical derivatization, if performed without

proper understanding, can lead to modifications of the hydrogels to products that will not degrade readily in the colon because it is possible that the new structures will not be recognized by the colonic bacterial enzymes for degradation. Also, bulk degradation is preferred to surface erosion because it leads to a more rapid rate of drug release. The one major drawback with the use of hydrogels and matrix systems is that only a limited amount of drug can be incorporated. Thus, when a large amount of drug is required at the target site in the colon, this may not be the most suitable carrier.

CONCLUSION

A great deal of research has been conducted in the past two decades in delivery systems for targeting drug release in the colon. This is clearly a region of the gastrointestinal tract worthy of attention because it does not have the

hostile environment that is responsible for drug degradation in the stomach and the small intestine. Local therapy of pathologies of the colon and reduced drug availability attributable to degradation of the active ingredients by digestive or mucosal enzymes can benefit from colonic delivery. However, the large intra- and intersubject variation in gastrointestinal pH makes the delivery systems based on pH-dependent polymers less suitable for targeted drug delivery to the colon. Therefore, systems that rely solely on the prevalent conditions of the colon (for example, those that depend on degradation by colonic bacteria for drug release) are promising. However, care should be taken in designing such systems because there are concerns regarding toxicity issues with the use of azo-based polymeric delivery systems. The natural polymers, on the other hand, may not have sufficient film-forming capacity and also may be less prone to degradation in the stomach after chemical derivatization. A major problem in comparing different delivery systems to the colon is that the degradation studies are carried out in different experimental conditions. Moreover, despite promising results in animal studies, none of the polymeric systems have yet been tested clinically. Clearly, much work remains to be done to satisfactorily answer the concerns that have surfaced about colon-specific drug delivery.

REFERENCES

1. Kinget, R.; Kalala, W.; Vervoort, L.; van der Mooter, G. Colonic Drug Targeting. *J. Drug Targeting* **1998**, *6*, 129–149.
2. Lee, V.H.L. Changing Needs in Drug Delivery in the Era of Peptide and Protein Drugs. *Peptide and Protein Drug Delivery*; Lee, V.H.L., Ed.; Marcel Dekker, Inc.: New York, 1991; 1–56.
3. Mersny, R.J. The Colon As a Site for Drug Delivery. *J. Controlled Release* **1992**, *22*, 15–34.
4. Hovgard, L.; Bronsted, H. Current Applications of Polysaccharides in Colon Targeting. *Crit. Rev. Ther. Drug Carr. Syst.* **1996**, *13*, 185–223.
5. Binder, H.J.; Foster, E.S.; Budinger, M.E.; Hayslett, J.E. Mechanism of Electroneutral Sodium Chloride Absorption in Distal Colon of the Rat. *Gastroenterology* **1987**, *93*, 449–455.
6. Thomas, P.; Richards, D.; Richards, A.; Rojers, L.; Evans, B.K.; Drew, M.J.; Rhodes, J. Absorption of Delayed-Release Prednisolone in Ulcerative Colitis and Crohn's Disease. *Int. J. Pharm.* **1985**, *37*, 757.
7. Evans, D.F.; Pye, G.; Bramley, R.; Clark, A.G.; Dyson, T.J.; Hardcastle, J.D. Measurement of Gastrointestinal pH Profiles in Normal Ambulant Human Subjects. *Gut* **1988**, *29*, 1035–1041.
8. Ashford, M.; Fell, J.T.; Attwood, D.; Woodhead, P.J. An in Vitro Investigation into the Suitability of pH Dependent Polymers for Colonic Targeting. *Int. J. Pharm.* **1993**, *91*, 241–245.
9. Davis, S.S.; Hardy, J.G.; Fara, J.W. Transit of Pharmaceutical Dosage Forms Through the Small Intestine. *Gut* **1986**, *27*, 886–892.
10. Read, N.W.; Miles, C.A.; Fischer, D.; Holgate, A.M.; Kime, N.D.; Mitchell, M.A.; Reeve, A.M.; Boche, T.B.; Walker, M. Transit of a Meal Through the Stomach, Small Intestine, and Colon in Normal Subjects and its Role in the Pathogenesis of Diarrhea. *Gastroenterology* **1980**, *79*, 1276–1282.
11. Adkin, D.A.; Davis, S.S.; Sparow, R.A.; Wilding, I.R. Colonic Transit of Different Sized Tablets in Healthy Subjects. *J. Controlled Release* **1993**, *23*, 147–153.
12. Haeblerlin, B.; Friend, D.R. Anatomy and Physiology of the Gastrointestinal Tract. *Oral Colon-Specific Drug Delivery*; Friend, D.R., Ed.; CRC Press: Boca Raton, 1992; 1–43.
13. Hardy, J.G.; Davis, S.S.; Khosla, R.; Robertson, C.S. Gastrointestinal Transit of Small Tablets in Patients with Ulcerative Colitis. *Int. J. Pharm.* **1988**, *48*, 79–82.
14. Rangachari, P.K. Six-Pack Balancing Act: A Conceptual Model for the Intestinal Lining. *Can. J. Gastroenterol.* **1990**, *4*, 201–208.
15. Nubuchi, J.-J.; Aramaki, Y.; Tsuchiya, S. Binding of Antibiotics to Rat Intestinal Mucin. *Int. J. Pharm.* **1986**, *30*, 181–188.
16. Black, J.W.; Bradbury, J.E. Whyllie Stimulation of Colonic Mucus Output in the Rat. *Br. J. Pharmacol.* **1979**, *66*, 456P–457P.
17. Schipper, N.G.; Varum, K.M.; Stenberg, P.; Ocklind, G.; Lennernas, H.; Artursson, P. Chitosans As Absorption Enhancers of Poorly Absorbable Drugs. 3. Influence of Mucus on Absorption Enhancement. *Eur. J. Pharm. Sci.* **1999**, *8*, 335–343.
18. Powell, D.W. Barrier Function of Epithelia. *Am. J. Physiol.* **1981**, *241*, G275–G288.
19. Nellans, H.N. Paracellular Intestinal Transport: Modulation of Absorption. *Adv. Drug Delivery Rev.* **1991**, *7*, 339–364.
20. Luciano, L.; Real, E.; Rechkemmer, G.; von Engelhardt, W. Structure of the Zonula Occludentes and the Permeability of the Epithelium to Short-Chain Fatty Acids in the Proximal and the Distal Colon of Guinea Pig. *J. Membr. Biol.* **1984**, *82*, 145–156.
21. Berne, R.M. *Physiology*; Levy, M.N., Ed.; C.V. Mosby Inc.: Toronto, 1988; 742–743.
22. Jackson, M.J. Drug Transport Across Gastrointestinal Epithelia. *Physiology of the Gastrointestinal Tract*; Johnson, L.R., Ed.; Raven Press: New York, 1987; 1597–1621.
23. Burton, P.S.; Conradi, R.A.; Hilgers, A.R. Transcellular Mechanism of Peptide and Protein Absorption: Passive Aspects. *Adv. Drug Delivery Rev.* **1991**, *7*, 365–386.
24. Armstrong, R.N. Enzyme-Catalyzed Detoxication Reactions: Mechanisms and Stereochemistry. *CRC Crit. Rev. Biochem.* **1987**, *22*, 39–88.
25. Pacifici, G.M.; Franchi, M.; Bencini, C.; Repetti, F.; Di Lascio, N.; Muraro, G.B. Tissue Distribution of Drug-Metabolizing Enzymes in Humans. *Xenobiotica* **1988**, *18*, 849–856.

26. Ritschel, W.A. Targeting in the Gastrointestinal Tract: New Approaches. *Methods Find. Exp. Clin. Pharmacol.* **1991**, *13*, 313–336.
27. Haeberlin, B.; Friend, D.R. Anatomy and Physiology of the Gastrointestinal Tract. *Oral Colon-Specific Drug Delivery*; Friend, D.R., Ed.; CRC Press: Boca Raton, 1992; 1–43.
28. Hill, M.J.; Drasar, B.S. The Normal Colonic Bacterial Flora. *Gut*. **1975**, *16*, 318–323.
29. Moore, W.E.C.; Holdeman, L.V. Discussion of Current Bacteriologic Investigations of the Relationship Between Intestinal Flora, Diet and Colon Cancer. *Cancer Res.* **1975**, *35*, 3418–3420.
30. Lindsey, J.T.; Smith, J.W.; McCluggage, S.G.J.; Nicholas, R.L. Effects of Commonly Used Bowel Preparations on the Large Bowel Mucosa-Associated and Luminal Microflora in the Rat Model. *Dis. Col. Rectum*. **1990**, *33*, 554–560.
31. Macfarlane, G.T.; Hay, S.; Gibson, G.R. Influence of Mucin on Glycosidase, Protease and Arylamidase Activities of Human Gut Bacteria Grown in a 3-Stage Continuous Culture System. *J. Appl. Bacteriol.* **1989**, *66*, 407–417.
32. Macfarlane, G.T.; Gibson, G.R.; Beatty, J.H.; Cummings, J.H. Estimation of Short Chain Fatty Acid Production from Protein by Human Intestinal Bacteria Based on Branched-Chain Fatty Acid Measurements. *FEMS Microbiol. Ecol.* **1992**, *101*, 81–88.
33. Longer, M.A.; Woodley, J.F.; Duncan, R. Comparison of the Activities of Rat Small Intestine and Colon Brush Border Membrane Peptidases. *Proc. Int. Symp. Controlled Release Bioact. Matter.* **1989**, *16*, 225–226.
34. Digenis, G.A.; Sandefer, E. Gamma Scintigraphy and Neutron Activation Techniques in the In Vivo Assessment of Orally Administered Dosage Forms. *Crit. Rev. Ther. Drug Carrier Syst.* **1991**, *7*, 309–345.
35. Kimura, R.E.; Dy, S.A.; Uhing, M.R.; Beno, D.W.; Jiyamapa, V.A.; Lloyd-Still, J.D. The Effects of High Dose Ibuprofen and Pancreatic Enzymes on the Intestine of the Rat. *J. Pediatr. Gastroenterol. Nutr.* **1999**, *29*, 178–183.
36. Tomita, M.; Shiga, M.; Hayashi, M.; Awazu, S. Enhancement of Colonic Drug Absorption by the Transcellular Permeation Route. *Pharm. Res.* **1988**, *5*, 431–435.
37. Muranishi, S. Absorption Enhancers. *Crit. Rev. Ther. Drug Carrier Syst.* **1990**, *7*, 1–33.
38. Nishihata, T.; Kamikawa, K.; Takahata, H.; Kamada, A. Enhanced Intestinal Absorption of Amino Acids and Amino Acid-Like Drug by Possible Formulation of Anamine in Administered Solution by the Presence of Ethylacetoacetate. *J. Pharmacobiodyn.* **1984**, *7*, 143–150.
39. Cheek, J.M.; Kim, K.-J.; Crandall, E.D. NO₂ Decreases Paracellular Resistance to Ion and Solute Flow in Alveolar Epithelial Monolayers. *Exp. Lung Res.* **1990**, *16*, 561–575.
40. Rangachari, P.K.; Prior, T.; McWade, D. Epithelial and Mucosal Preparations from Canine Colon: Responses to Substance. *P. J. Pharmacol. Exp. Ther.* **1990**, *254*, 1076–1083.
41. Tomita, M.; Hayashi, M.; Toshiharu, H.; Awazu, S. Enhancement of Colonic Drug Absorption by the Transcellular Permeation Route. *Pharm. Res.* **1988**, *5*, 786–789.
42. Shimazaki, T.; Tomita, M.; Sadahiro, S.; Hayashi, M.; Awazu, S. Absorption-Enhancing Effects of Sodium Caprate and Palmitoyl Carnitine in Rat and Human Colons. *Dig. Dis. Sci.* **1988**, *43*, 641–645.
43. Lee, V.H.L.; Dodh-Kashi, S.; Grass, G.M.; Bubas, W. Oral Route of Peptide and Protein Drug Delivery. *Peptide and Protein Drug Delivery*; Lee, V.H.L., Ed.; Marcel Dekker, Inc.: New York, 1991; 691–738.
44. Muranishi, S.; Takada, K. Biopharmaceutical Aspects on Enhanced Transmembrane Delivery of Peptides and Proteins. *Therapeutic Peptides and Proteins: Formulation, Delivery, and Targeting*; Marshak, D., Liu, D., Eds.; Cold Spring Harbor Laboratory: New York, 1989; 47–50.
45. Tsuji, S.; Kawano, S.; Tsuji, M.; Masuda, E.; Gunawan, E.S.; Hori, M. Mucosal Microcirculation and Angiogenesis in Gastrointestinal Tract. *Jap. J. Clin. Med.* **1998**, *56*(9), 2247–2252.
46. Khan, A.; Piris, J.; Truelove, S.C. An Experiment to Determine the Active Therapeutic Moiety of Sulphasalazine. *Lancet* **1977**, *2*, 892–895.
47. Van Hozegard, R.A. Pharmacokinetics of Olsalazine and Its Metabolites. *Scan. J. Gastroenterol.* **1998**, *23* (Suppl. 148), 17–20.
48. Chan, R.P.; Pope, D.J.; Gilbert, A.P.; Sarca, P.J.; Baron, J.H.; Lennard-Jones, J.E. Studies of Two Novel Sulphasalazine Analogs, Ipsalazide and Balsazide. *Dig. Dis. Sci.* **1983**, *28*, 609–615.
49. Brown, J.P.; McGarraugh, G.V.; Parkinson, T.M.; Wingerd, R.E.; Oderndonck, A.B. A Polymeric Drug for Treatment of Inflammatory Bowel Disease. *J. Med. Chem.* **1983**, *26*, 1300–1307.
50. Schacht, E.; Gevaert, A.; Kenway, E.R.; Molly, K.; Verstraete, W.; Adriaenssens, P.; Carleer, R.; Gelan, R. Polymers for Colon Specific Drug Delivery. *J. Controlled Release* **1996**, *39*, 327–338.
51. Friend, D.R.; Chang, G.W. A Colon Specific Drug Delivery System Based on Drug Glycosides and Glycosidases of Colonic Bacteria. *J. Med. Chem.* **1984**, *27*, 261–266.
52. Friend, D.R.; Tozer, T.N. Drug Glycosides in Oral Colon-Specific Drug Delivery. *J. Controlled Release* **1992**, *19*, 109–120.
53. Friend, D.R. Glycoside Prodrugs: Novel Pharmacotherapy for Colonic Diseases. *S. T. P. Pharma. Sci.* **1995**, *5*, 70–76.
54. Harboe, E.; Larsen, C.; Johansen, M.; Olesen, H.P. Macromolecular Prodrugs XV. Colon Targeted Delivery Bioavailability of Naproxen from Orally Administered Dextran-Naproxen Prodrugs Varying in Molecular Size in the Pig. *Pharm. Res.* **1989**, *6*, 919–923.
55. Macleod, A.D.; Tolentino, L.; Tozer, T.N. Glucocorticoid-Dextran Conjugates As Potential Prodrugs for Colon-Specific Delivery: Steady-State Pharmacokinetics in the Rat. *Biopharm. Drug Disp.* **1994**, *15*, 151–161.
56. Hirayama, F.; Minami, K.; Uekama, K. In-Vitro Evaluation of Biphenyl Acetic Acid- β -Cyclodextrin Conjugates As Colon-Targeting Prodrugs: Drug Release Behaviour in Rat Biological Media. *J. Pharm. Pharmacol.* **1996**, *48*, 27–31.

57. Minami, K.; Hirayama, F.; Uekama, K. Colon-Specific Drug Delivery Based on a Cyclodextrin Prodrug: Release Behavior of Biphenylacetic Acid from its Cyclodextrin Conjugates in Rat Intestinal Tracts After Oral Administration. *J. Pharm. Sci.* **1998**, *87*, 715–720.
58. Wilson, C.G.; Washington, N. *Physiological Pharmaceutics, Biological Barriers to Drug Absorption*; Ellis Harwood, Ltd: Chichester UK, 1989.
59. Villar-Lopez, M.E.; Nieto-Reyes, L.; Aguiano-Igea, S.; Otero-Espinar, F.J.; Blanco-Mendez, J. Formulation of Triamcinolone Acetonide Pellets Suitable for Coating and Colon Targeting. *Int. J. Pharm.* **1999**, *179*, 229–235.
60. Ashford, M.; Fell, J.T.; Attwood, D.; Sharma, H.; Woodhead, P.J. An In Vivo Investigation into the Suitability of pH Dependent Polymers for Colonic Targeting. *Int. J. Pharm.* **1993**, *95*, 193–199.
61. Khan, M.Z.; Prebeg, Z.; Kurjakovic, N. A pH Dependent Colon Targeted Oral Drug Delivery System Using Methacrylic Acid Copolymers. I. Manipulation of Drug Release Using Eudragit L 100-55 and Eudragit S100 Combinations. *J. Controlled Release* **1999**, *58*, 215–222.
62. Peeters, R. Kinget, R. Film-Forming Polymers for Colonic Drug Delivery. Synthesis and Physical and Chemical Properties of Methyl Derivatives of Eudragit S. *Int. J. Pharm.* **1993**, *94*, 125–134.
63. Peeters, R.; *Studie Over De Ontwikkeling Van Een Colon-Specifieke Artsenijvorm*; Lueven, K.U., Ed.; Doctoral Thesis University of Leuven: Leuven Belgium.
64. Gazzaniga, A.; Bussetti, C.; Moro, L.; Sangali, M.E.; Giordano, F. Time Dependent Oral Delivery Systems for Colon Targeting. *S. T. P. Pharma. Sci.* **1995**, *5*, 70–76.
65. Pozzi, F.; Furlani, P.; Gazzaniga, A.; Davis, S.S.; Wilding, I.R. The Time-Clock System: A New Oral Dosage Form for Fast and Complete Release of Drug After a Predetermined Lag-Time. *J. Controlled Release* **1994**, *31*, 99–108.
66. Ueda, S.; Ibuki, R.; Kawamuza, A.; Murata, S.; Takahashi, T.; Kimuza, S.; Hata, T. Development of a Novel Drug Delivery System, Time-Controlled-Explosion System (TES). *J. Drug Targeting* **1994**, *2*, 133–140.
67. Vandelli, M.A.; Leo, E.; Forni, F.; Bernabei, M.T. In Vitro Evaluation of a Potential Colonic Delivery System that Releases Drug after a Controllable Lag-Time. *Eur. J. Pharm. Biopharm.* **1996**, *43*, 148–151.
68. Saffran, M.; Kumar, G.S.; Neckers, D.C.; Pena, J.; Jones, R.H.; Field, B. Biodegradable Azopolymer Coating for Oral Delivery of Peptide Drugs. *Biochem. Soc. Trans.* **1990**, *18*, 752–754.
69. Ishibashi, T.; Pitcairn, G.R.; Yoshino, H.; Mizobe, M.; Wilding, I.R. Scintigraphic Evaluation of a New Capsule-type Colon Specific Drug Delivery System in Healthy Volunteers. *J. Pharm. Sci.* **1988**, *87*, 531–535.
70. Yoshikawa, Y.; Hu, Z.; Kimura, G.; Muranishi, M.; Yoshikawa, H.; Takada, K. A Dissolution Test for a Pressure-Controlled Colon Delivery Capsule: Rotating Beads Method. *J. Pharm. Pharmacol.* **1999**, *51*, 979–989.
71. Van den Mooter, G.; Samyn, C.; Kinget, R. In Vivo Evaluation of a Colon-Specific Drug Delivery System: An Absorption Study of Theophylline from Capsules Coated With Azo Polymers in Rats. *Pharm. Res.* **1995**, *12*, 244–247.
72. Kimura, Y.; Makita, Y.; Kumagai, T.; Yamane, H.; Kitao, T.; Sasatani, H.; Kim, S.I. Degradation of Azo Containing Polyurethane by the Action of Intestinal Microflora. *Polymer* **1992**, *33*, 5294–5299.
73. Ueda, T.; Yamaoki, T.; Miyamoto, M.; Kimura, Y.; Sasatani, H.; Kim, S.I. Bacterial Reduction of Azo Compounds as a Model Reaction for the Degradation of Azo-Containing Polyurethane by the Action of Intestinal Flora. *Bull. Chem. Soc. Jpn.* **1996**, *69*, 1139–1142.
74. Krishnaiah, Y.S.; Satyanarayana, S.; Prasad, Y.V. Studies of Guar Gum Compression-Coated 5-Aminosalicylic Acid Tablets for Colon-Specific Drug Delivery. *Drug Dev. Ind. Pharm.* **1999**, *25*, 651–657.
75. Prasad, Y.V.; Krishnaiah, Y.S.; Satyanarayana, S. In Vitro Evaluation of Guar Gum as a Carrier for Colon-Specific Drug Delivery. *J. Controlled Release* **1998**, *51*, 281–287.
76. Veervort, L.; Kinget, R. In Vitro Degradation by Colonic Bacteria of Insulin-HP Incorporated in Eudragit RS Films. *Int. J. Pharm.* **1996**, *129*, 185–190.
77. Wakerly, Z.; Fell, J.T.; Attwood, D.; Parkins, D. Pectin/Ethyl Cellulose Film Coating Formulations for Colonic Drug Delivery. *Pharm. Res.* **1996**, *13*, 1210–1212.
78. Lorenzo-Lamosa, M.L.; Remunan-Lopez, C.; Vila-Jato, J.L.; Alonso, M.J. Design of Microencapsulated Chitosan Microspheres for Colonic Drug Delivery. *J. Controlled Release* **1998**, *52*, 109–118.
79. Tozaki, H.; Fujita, T.; Odoriba, T.; Terabe, A.; Okabe, S.; Muranishi, S.; Yamamoto, A. Validation of a Pharmacokinetic Model of Colon-Specific Drug Delivery and the Therapeutic Effects of Chitosan Capsules Containing 5-Aminosalicylic Acid on 2,4,6-Trinitrobenzenesulphonic Acid-Induced Colitis in Rats. *J. Pharm. Pharmacol.* **1999**, *51*, 1107–1112.
80. Ramdas, M.; Dileep, K.J.; Anitha, Y.; Paul, W.; Sharma, C.P. Alginate Encapsulated Bioadhesion Chitosan Microspheres for Intestinal Drug Delivery. *J. Biomater. Appl.* **1999**, *13*, 290–296.
81. Aiedeh, K.; Taha, M.O. Synthesis of Chitosan Succinate and Chitosan Phthalate and Their Evaluation as Suggested Matrices in Orally Administered, Colon-Specific Drug Delivery Systems. *Arch. Pharm (Weinheim)* **1999**, *332*, 103–107.
82. Alwood, M.C.; Archer, D.B.; Ring, S.G. Delayed Release Formulations. Europe Application Patent 0343993A1, 1989.
83. Milojevic, S.; Newton, M.; Cummings, J.M.; Gibson, G.R.; Botham, R.L.; Ring, S.G.; Stockham, M.; Alwood, M.C. Amylose as a Coating for Drug Delivery to the Colon: Preparation and In Vitro Evaluation Using 5-Aminosalicylic Acid Pellets. *J. Controlled Release* **1996**, *38*, 75–84.
84. Kenyon, C.J.; Nardi, R.V.; Wong, D.; Hooper, G.; Wilding, I.R.; Friend, D.R. Colonic Delivery of Dexamethasone: A Pharmacoscintigraphic Evaluation. *Aliment. Pharmacol. Ther.* **1997**, *11*, 205–213.
85. Krishnaiah, Y.S.R.; Satyanarayana, S.; Rama Prasad, Y.V.; Narsimha Rao, S. Gamma Scintigraphic Studies on Guar Gum Matrix Tablets for Colonic Drug Delivery in Healthy Human Volunteers. *J. Controlled Release* **1998**, *55*, 245–252.

86. Bauer, K.H.; Kusselhut, J.F. Novel Pharmaceutical Excipients for Colon Targeting. *S. T. P. Pharma. Sci.* **1995**, *5*, 54–59.
87. Hirsch, S.; Binder, V.; Kolter, K.; Kesselhut, J.F.; Bauer, K.H. Lauroyldextran as a Coating Material for Site-Specific Drug Delivery to the Colon: In Vitro Dissolution of Coated Tablets. *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **1997**, *24*, 379–380.
88. Rubinstein, A.; Nakar, D.; Sintov, A. Chondroitin Sulfate: A Potential Biodegradable Carrier for Colon-Specific Drug Delivery. *Int. J. Pharm.* **1992**, *84*, 141–150.
89. Rubinstein, A.; Rudai, R. In Vitro and In Vivo Analysis of Colon Specificity of Calcium Pectinate Formulations. *Eur. J. Pharm. Biopharm.* **1995**, *41*, 291–295.
90. Ashford, M.; Fell, J.T.; Attwood, D.; Sharma, H.; Woodhead, P.J. Studies on Pectin Formulations for Colonic Drug Delivery. *J. Controlled Release* **1994**, *30*, 225–232.
91. Munjeri, O.; Collet, J.H.; Fell, J.T. Hydrogel Beads Based on Amidated Pectins for Colon-Specific Drug Delivery: Role of Chitosan in Modifying Drug Release. *J. Controlled Release* **1997**, *46*, 273–278.
92. Macleod, G.S.; Fell, J.T.; Collett, J.H.; Sharma, H.L.; Smith, A.M. Selective Drug Delivery to the Colon Using Pectin: Chitosan: Hydroxypropyl Methylcellulose Film Coated Tablets. *Int. J. Pharm.* **1999**, *187*, 251–257.
93. Kopecek, J.; Kopeckova, P.; Bronstead, H.; Rathi, R.; Rihova, B.; Yeh, P.-Y.; Ikesue, K. Polymes for Colon-Specific Drug Delivery. *J. Controlled Release* **1992**, *19*, 121–130.
94. Brondsted, H.; Kopecek, J. Hydrogels for Site-Specific Oral Drug Delivery: Synthesis and Characterization. *Biomaterials* **1991**, *12*, 584–592.
95. Ghandehari, H.; Kopeckova, P.; Kopecek, J. In Vitro Biodegradation of PH-Sensitive Hydrogels Containing Aromatic Azo Bonds. *Biomaterials* **1997**, *18*, 861–872.
96. Simonsen, L.; Hovgaard, L.; Mortensen, P.B.; Brondsted, H. Dextran Hydrogels for Colon-Specific Drug Delivery. Degradation in Human Intestinal Incubation Models. *Eur. J. Pharm. Sci.* **1995**, *3*, 329–337.
97. Chiu, H.C.; Hsiue, G.H.; Lee, Y.P.; Huang, L.W. Synthesis and Characterization of pH-Sensitive Dextran Hydrogels As a Potential Colon-Specific Drug Delivery System. *J. Biomater. Sci. Polym. Ed.* **1999**, *10*, 591–608.
98. Hirsch, S.; Binder, V.; Schehlmann, V.; Kolter, K.; Bauer, K.H. Lauroyldextran and Crosslinked Galactomannan as Coating Materials for Site-Specific Drug Delivery to the Colon. *Eur. J. Pharm. Biopharm.* **1999**, *47*, 61–71.
99. Shanta, K.L.; Ravishandran, P.; Rao, K.P. Azo Polymeric Hydrogels for Colon Targeted Drug Delivery. *Biomaterials* **1995**, *16*, 1313–1318.
100. Rubinstein, A.; Gliko-Kabir, I. Synthesis and Swelling-Dependent Enzymatic Degradation of Borax-Modified Guar Gum for Colonic Delivery Purposes. *S. T. P. Pharma. Sci.* **1995**, *5*, 41–46.

DRUG DELIVERY—OPHTHALMIC ROUTE

Massod Chowhan

Alan L. Weiner

Haresh Bhagat

Alcon Research Ltd., Fort Worth, Texas, U.S.A.

INTRODUCTION

Delivery of medication to the human eye is an integral part of medical treatment. The delivery of drug to the site of action has been practiced since ancient times, which successively advanced in a variety of ophthalmic dosage forms. The writings pertaining to eye medications have been found on Egyptian papyri. Between 20 BC and AD 50, Greeks and Romans practiced the delivery of the necessary components of eye medication by dissolving them in water, milk or egg white (1). They used the term *collyria* for such preparations. The term *belladonna* or “beautiful lady” evolved during the middle ages from such collyria, which contained components to dilate the pupils of a lady’s eyes for cosmetic purpose (2). The collyria gave rise to the birth of modern-day eye drop solutions.

Prior to World War II and well into the 1940s, most solutions for eye use were compounded by the pharmacist in the community pharmacy and were intended for immediate use, perhaps due to an unconfirmed stability of the drug (2). The availability of such solutions in a sterile dosage form marked the most important milestone in this century for the modern day eye drop solution. Alcon Pharmacy (currently known as Alcon Laboratories Inc.), a dispensing pharmacy during the day and a manufacturing pharmacy during the night, was the first provider of sterile ophthalmic solutions in 1947, long before the Food and Drug Administration (FDA) adopted a position in 1953 that a nonsterile ophthalmic solution was considered adulterated. Subsequently the United States Pharmacopoeia (USP) adopted sterility as requirement for ophthalmic solutions in 1955 (3).

The traditional era of the solution-only dosage form for use in the eye ended in the 1950s with the availability of suspension dosage forms. Solid drug particles of cortisone acetate were first suspended and a suspension product commercialized. In an unorthodox approach, for the first time clinical studies revealed that a sufficiently reduced particle sized drug could be instilled on ocular surfaces. This resulted in the availability of water insoluble or

sparingly soluble drugs for the mitigation of ophthalmic disorders in suspension dosage forms.

The successful experimentation of delivering sparingly soluble drugs in a suspended form led the way for products with added values beyond simple presentations. Increasing awareness of basic properties of drug molecules and the wide spread availability of excipients further advanced the discoveries of value added deliveries. Jani et al. (4) reported such delivery of β_1 -adrenergic antagonist as a bound drug to cationic polymer resin beads, suspended in an aqueous vehicle. The authors were able to reduce ocular discomfort of betaxolol yet maintain the therapeutic efficacy of the drug for controlling intra ocular pressure.

Gressel et al. (5) reported viscosity modifiers for enhancing intended effects on an ocular surface. A polyacrylic polymer was used to increase viscosity to a gel consistency and thereby enhancing the treatment of symptoms for dry eyes. Sullivan reported efficacy of carbomer gel in improving the number of subjective and objective symptoms of moderate-to-severe dry eye syndromes (6). Such a gel vehicle offered an advantage of reducing frequency of instillation and has resulted in a commercial product (Pilopine HS Gel).

Recognizing the value of rheological properties of polymers facilitated discovery of gel forming solutions for drugs such as timolol maleate. In these systems, timolol is formulated in an aqueous vehicle which on contact with the ocular surface changes to a gel like consistency, thereby, extending the duration of contact. Extending the contact time improved bioavailability thereby reducing the frequency of the dosage to once daily instillation. Two extended duration timolol maleate products, one containing gellan as a gel forming ingredient (Timoptic XE), and the other with xanthan (Timolol Gel Forming Solution) have been approved for commercial use. Shin reported a single daily application of such dosage in reducing intraocular pressure in human subjects (7). Beyond the use of solutions, suspensions and gels, further developments have been made in formulation of creams and ointments, intra-ocular injections, viscoelastic solutions and newer devices and inserts.

EVOLUTION OF OCULAR DRUG DELIVERY SYSTEMS

Topical ocular application of controlled drug delivery systems is a relatively new science compared to earlier ocular dosage form, with roots beginning in the late 1960s to early 1970s.

Delivery of Drugs from Contact Lens Materials

The earliest attempts to significantly prolong the action of drugs applied to the eye focused on using existing systems known in the ophthalmic field. In particular the most well known and characterized solid systems for ocular use are hydrophilic contact lenses. The primary material initially developed for constructing hydrophilic contact lenses was hydroxyethyl methacrylate (HEMA). Many variations of HEMA polymers have since been produced for contact lenses wherein the HEMA is typically copolymerized with methylmethacrylate, ethoxyethylmethacrylate, 1,3-propanediol trimethacrylate, ethylene dimethacrylate, allyl methacrylate, ethylene glycol dimethacrylate, dimethyl oxybutyl acrylamide, or vinyl pyrrolidone. Depending on copolymer composition and cross-linkage, hydrogels of this type range from 38 to 79% water content. Because of this property, it is possible to load polymers of this type with drug by soaking them in an aqueous solution containing the drug. This application was first reported in the early 1970s in papers examining uptake and release of agents such as fluorescein from Bionite lenses produced by Griffin Laboratories and Soflens lenses from Bausch and Lomb. Studies showed significant differences in uptake and release rates for the two types of lenses. Today these results are not unexpected as it has been shown by Sorenson et al. (8) that elimination of radiotracers from presoaked contact lenses are cleared more slowly in lenses containing higher water content and thicker dimensions. In low water content lenses (minimum of 38%), elimination constants of a technetium label were in the range of $0.278\text{--}0.155\text{ min}^{-1}$. In contrast the elimination constant of a 75% hydrated lens was 0.029 min^{-1} , a factor of 7–10 times slower. These differences have been confirmed in other studies using specific drugs such as tobramycin (9).

From an efficacy point of view, early studies with HEMA based lenses were conducted using pilocarpine, the key available glaucoma drug during that era. Several studies reported improvements in reduction of intraocular pressure and corneal drug flux using presoaked lenses containing lower pilocarpine concentrations than standard

drops (10–14). In later studies following the identification of timolol as a glaucoma agent (15) it was found that polyvinylmethacrylate circular disks of 13-mm diameter and 0.5-mm thickness will release timolol differentially depending on the addition of a basic additive (disodium phosphate). An enhancement in the rate of release and a shifting of peak timolol levels from 4 h to 30 min were observed when formulated with the additive.

Antibiotics have also been examined for uptake and release from contact lens materials. Impregnation of various hydrogels constructed of various methacrylate and polyvinylpyrrolidone copolymers with erythromycin and erythromycin estolate were effective in slow releasing drug over a period sufficient to suppress *Chlamydia* infection in a monkey model (16, 17). As well, modest increases in intraocular gentamicin have been observed upon administration of gentamicin soaked hydrogel lenses in rabbits (18). Clinically, drug penetration of gentamicin, chloramphenicol, or carbenicillin from hydrogel lenses was found to be higher than subconjunctival injection of control solutions over a 2–12-h period posttreatment (19). In prepolymerized poly HEMA minidisks loaded with particles of sulfisoxazole, extended release over a 168-h period was achieved (20). This proved to be successful in treating a *Staphylococcus aureus* infection model in rabbits from one time administration of the minidisk as compared to 3 times daily control solutions.

Absorption and washout characteristics of drugs from poly HEMA based lenses are drug dependent. For example, following a 7-day immersion, increasing levels of uptake are observed for norepinephrine, gentamicin, pilocarpine and dexamethasone, respectively (21). Washout of these drugs from the lens was nearly complete (83–98%). Maximum amounts of these drugs taken up into these lenses represented only one tenth of the dose achievable by application of topical drops. Total uptake of these agents into polymethylmethacrylate was also quite low. These results were supported by earlier studies indicating that intraocular poly HEMA does not act as a significant sponge or long term reservoir for drugs such as dexamethasone, chloramphenicol, epinephrine and pilocarpine (22).

While the majority of reports have examined topical release of drugs from contact lens materials, the implantation of these materials, as is common for intraocular lenses, has been reported. One such example was recorded by Shing et al. (23) who implanted poly HEMA pellets loaded with fibroblast growth factor (FGF)-sucralfate into the cornea as a means to develop a corneal neovascularization model for testing of anti-angiogenic agents. Neovascularization was produced beginning at day 2 and reaching a maximum at day 11 with an elongation rate of 0.41 mm/day until day 13.

Ocusert

At about the same time as investigations were being carried out with contact lens materials for drug release, an ethylene vinyl acetate (EVA) membrane device, Ocusert[®] was developed by the Alza Corp. (Palo Alto, CA) and eventually commercialized in 1974 (20, 24–28). Ocusert is an elliptical shaped device consisting of two outer layers of rate controlling EVA, and an inner layer of pilocarpine in an alginate gel. The device is designed for continuous release of the drug at a 20 or 40 $\mu\text{g/h}$ rate over 7 days. Enhanced release of the pilocarpine in the higher rate device is facilitated by addition of a flux enhancer, di-(ethylhexyl)phthalate. While this device functions effectively in a specific niche of difficult to manage glaucoma patients, it has not been universally adopted for use because of unsatisfactory control of IOP in some patients, ejection of the device from the eye, and irritation or tolerance difficulties (29–31).

Erodible Polymeric Delivery Systems

During the same period that reports appeared on the Ocusert device, research was progressing on the use of erodible polymer systems for ophthalmic drug delivery. This research blossomed in the early 1980s with a particular focus on polymers employed in the manufacture of absorbable sutures. Release of many ophthalmic drugs from polymeric matrices either via dissolution or erosion was investigated. In addition to the development of slower releasing carrier systems, many papers have since appeared on the use of viscosity additives and bioadhesives to extend retention of delivery systems in the eye. For example albumin nanoparticles containing pilocarpine are better retained by the inclusion of methylcellulose, hydroxypropylmethylcellulose, polyvinylalcohol, sodium carboxymethylcellulose, carbopol 941, hyaluronate or mucin in the formulation (32). Similar effects are noted when carbopol is used in coating ophthalmic gentamicin drug delivery inserts comprised of hydroxypropylcellulose, ethylcellulose, and polyacrylate (33).

Polyvinylalcohol

Polyvinylalcohol disks for delivery of drugs to the eye were proposed as early as 1966 for potential use by astronauts (34). Pilocarpine loaded disks exhibited sustained miosis and IOP reduction in human subjects. Maichuk (35–37) has elaborated on these early studies by showing that PVA films containing pilocarpine, antibiotics, or antimetabolites increased drug concentration in the tear film and prolonged the delivery times. Similarly, bioavailability, miotic activity in rabbits, and intraocular

pressure control in human glaucoma subjects were all enhanced over a 24-h period with PVA/pilocarpine–PAA disks of 4-mm diameter and 0.4-mm thickness prepared from cast films (38). A PVA film device termed NODS (“new ophthalmic delivery system”) has also been utilized for studying improvement in delivery of pilocarpine, tropicamide, chloramphenicol, and proparacaine (39). Heat-treated PVA membrane sandwiches have been prepared with 5 mg of ganciclovir, which released at a rate of 0.25 $\mu\text{g/h/mm}^2$ over a 1-week period (40). Similarly, pellets of either thalidomide (41), cyclosporin (42), or leflunomide (43) with either PVA incorporated into the pellet or coating the pellet have exhibited sustained release of those drugs either in vitro or in the vitreous or subconjunctival spaces.

Delivery of drugs from collagen shields

The concept of drug delivery from contact lenses was extended to include contact lens-shaped collagen shields after their approval for use directly on the surface of the eye to treat corneal wounds (44–46). Several types of collagen shields are commercially available under such names of ProShield (Alcon Laboratories, Inc., Fort Worth, TX), Bio-Cor (Bausch & Lomb, Clearwater, FL), and Soft Shield (Oasis, Glendora, CA). The collagen for these products is derived from a bovine or porcine source with dimensions of 14.5 mm in diameter and thickness between 0.012 and 0.071 mm. In the eye these shields dissolve over a 12–72 h period depending on the amount of cross-linkage. Similar to the earlier contact lens studies, the majority of investigations examining release of active agents from collagen matrices first involve loading of active agent into the shield by soaking in the agent over a sufficient period of time (47). Because of the relatively large pore size, diffusion into and out of the shield does not typically surpass 2–3 h.

Many studies have looked into prolongation of aminoglycoside antibiotic release from collagen matrices. Tear film and topical tissue levels (sclera and cornea) of gentamicin delivered from collagen matrices is elevated over controls as determined by pharmacokinetic evaluation of ¹⁴C-gentamicin (48). In models of *Pseudomonas keratitis* in rabbits (49, 50), collagen shields loaded with gentamicin (49, 50) or tobramycin (51) have produced significant reductions in colony-forming units (CFU) over control animals given drug alone. However, changing from the shield design to either a doughnut (52) or disk (53) does not seem to offer an improvement in gentamicin release. Although aqueous humor levels of gentamicin are not achieved when using shields presoaked in 40 mg/ml gentamicin as compared to drops (54), levels of tobramycin delivered from shields

have been detected in aqueous humor (55, 56). Neither repetitive drop treatment nor release from collagen shields results in the establishment of vitreous drug levels. More extensive studies have been reported on collagen shields to deliver tobramycin (57–62) in various other tests of ocular trauma. Enhanced delivery of procaine penicillin, erythromycin, erythromycin estolate (63), silver nitrate 1%, povidone-iodine 5%, chlorhexidine gluconate 1% (64), and amphotericin 5% (65) from collagen shields also has been reported.

Similar to antibiotic studies, delivery of corticosteroids from collagen shields has been demonstrated to produce prolongation of effects, enhanced penetration, and increased efficacy (66–69).

Antimetabolites such as 5-fluorouracil (5-FU) have been used experimentally for retarding the healing of incisions made to improve outflow of aqueous humor and reduction in intraocular pressure in glaucoma patients. Several groups have examined the potential of 5-FU loaded collagen shields for improving current therapy over 5-FU alone (10, 71–73). Increases in duration of action and success rates have been noted. Another antimetabolite, trifluorothymidine, when released from collagen shields was investigated as a potential treatment for ulcerative herpetic keratitis (74). In these studies, cornea and aqueous humor levels were 19–42% higher when eyes were treated with drug loaded shields.

More effective treatment of glaucoma has also been attempted using collagen shield technology. In this regard, shields have been shown to prolong delivery of pilocarpine (75–77) and metipranolol (78).

Collagen disks have had their greatest application in the treatment of wounds. Enhancement of the inherent protective property of the material has been investigated with the subsequent inclusion of growth factors. Collagen disks soaked for 5 min in platelet derived or epidermal growth factors (100 µg/ml) was capable of increasing the wound healing rates of debrided corneas in rabbits (79). Following placement on the cornea, collagen shields containing tissue plasminogen activator shortened fibrin clot lysis time by 50% over controls (80). Another wound modulator, cyclosporin A delivered by topical collagen shields, has been reported to be 10-fold higher in cornea and aqueous humor over an 8-h period as compared to controls (81).

Poly(lactide) (PLA), polyglycolide (PGA) and polycaprolactone (PCL)

A further extension of drug delivery using naturally biodegrading materials other than collagen included the application of known eroding suture materials. Polylactic acid, polyglycolic acid and polycaprolactone are the

primary materials used in dissolvable sutures including Dexon® (Davis and Geck, Danbury, CT) and Vicryl® Polyglactin (Ethicon, Somerville, NJ). The key focus of ocular drug delivery from these polymers has been for sustained release of 5-FU. Sustained release 5-FU from PLA/PGA copolymer microspheres had been utilized for applications in glaucoma filtration surgery and proliferative vitreoretinopathy (82–86). Release of 5-FU is controllable over a 7-day period in vitro and when injected into the vitreous PLA microspheres (with 2 mg drug) degrades over a 48-day period. In a PVR model, tractional retinal detachments are reduced from 60 to 10% of animals when 1.25 mg of drug is delivered from PLA microspheres as compared to control injections (83). In animals, vitreous concentrations of 5-FU delivered from PLA/PGA microspheres (250 µg) or rods (1 mg) can be detected from periods between 11 days (87) and 21 days (88–90). PLA/PGA devices incorporating 5-FU prevented retinal detachments in an experimental model that was not responsive to drug alone. PLA/PGA disks containing 5-FU are also efficacious for periods greater than one month in applications for glaucoma filtration surgery (91, 92). Devices of this type have shown some inflammatory and vascularization reactions depending on site of implantation (90, 93). PLA microspheres have shown sustained efficacy in delivering other antimetabolites such as doxorubicin (94, 95).

PLA, PGA and PCL delivery systems have also proven to be of value in delivering glaucoma agents. Nanoparticles (150-nm diameter) and nanocapsules (300-nm diameter) constructed from polycaprolactone incorporating either 1% carteolol chlorhydrate or 0.5% betaxolol were shown to be more effective than controls in hypertension models (96, 97). Prolongation of miosis has also been observed in rabbits receiving microcapsules constructed from PLA and containing pilocarpine hydrochloride (98).

Anti-inflammatory and anti-infective agents such as indomethacin, fluconazole, and dexamethasone have been incorporated into PLA, PLG, or PCL carriers (99–101). Significant increases in either drug concentration or duration of action were noted.

Polyanhydrides, polyorthoesters, and polyalkylcyanoacrylates

Beyond the use of suture materials, newer erodible polymeric materials were introduced in the 1980s as potential ophthalmic carrier systems for release of drugs.

As with PLA or PLG carriers, the application of either polyanhydride or polyorthoester polymers for 5-FU sustained release in glaucoma treatment or PVR has been investigated (102–107). Using compression

techniques, polyanhydride devices constructed from combinations of (*p*-carboxyphenoxy)alkanes with sebacic acid have been produced. Disk or T-shaped polyanhydride devices containing between 10 and 20% 5-FU prolonged IOP reduction and bleb survival in filtration models or better inhibited tractional retinal detachments in PVR models than standard drug controls. Similar effects on extended filtration bleb survival and IOP reduction have been observed for polyanhydride implants containing daunorubicin (108) or mitomycin C (109). Release of drugs from polyanhydride implants have been examined when administered subconjunctivally (etoposide) (110) or in the vitreous or anterior chamber (gentamicin) (111). In these reports slow release rates were established at least over a 1-week period. In context of vitreous delivery, polyorthoesters have been shown to release ganciclovir for 144 h which is controllable by the pH of the dissolution media (112).

Ocular distribution and elimination of biodegradable polyalkylcyanoacrylate nanoparticles has been examined in a number of reports (113–117). Pilocarpine containing polybutylcyanoacrylate nanoparticles alone or in gel type formulations are capable of enhancing and prolonging the miosis and pressure lowering response in various animal models (118–122).

Non-erodible Systems

Distinct from the development of contact lens materials or the Ocusert in the 1970s, additional promising delivery systems and materials have emerged with the more recent focus on development of intraocular applications for treatment of macular degeneration, cataract, retinopathy, and inflammation.

Vitrasert®

The Vitrasert, approved in 1996 and currently marketed by Bausch and Lomb is an ethylene vinyl acetate cup encasing a cylindrical core of ganciclovir, which is covered on one or two surfaces with a permeable PVA membrane to allow for diffusion (123–134). At the base of the cup is an anchoring strut made of PVA to allow for suturing. The drug core diameter of the device is 2.5 mm. Sustained levels of the drug are maintained in the vitreous over a period of 6–8 months in the treatment for cytomegalovirus (CMV) induced retinitis. This device has also been investigated for the delivery of 5-FU (126–130), flurbiprofen (131) cyclosporin (132), dorzolamide (133), and dexamethasone (134) when implanted at various sites. A variation on this device has been constructed using a core of ganciclovir/poly(lactide-glycolide) encased in a

crosslinked poly[HEMA-co(PVA-AA)] film that releases the drug in the vitreous over 63 days (135).

Silicones

Ophthalmic practice has long employed silicone polymers in surgeries utilizing scleral buckles and sponges, retinal tamponade and foldable intraocular lenses. Silicone elastomers and rubbers are hydrophobic in nature and therefore support long term payout characteristics for many drugs. Silicone polymer disks of 4–5-mm diameter termed minidisks have been studied for the release of gentamicin when placed in the conjunctival cul-de-sac (20, 136). Slow release of the drug is observed over a period of 10–14 days with tear concentrations in the range of 2.5 ppm during this period. Nanoparticles of silicone in the 150–200 μm size range can be made and when containing timolol effectively prolonged release of the drug (137). Sealed silicone tubes 1.46 mm ID by 15 mm length and filled with 2.5–20 mg/ml solutions of timolol slow release the drug over 8 h in vitro (138) or in vivo (139). Recently, development of a solid noodle-like silicone controlled drug delivery rod for the deep cul-de-sac (fornix) known as the Ocufit SR® has been reported (140–143). These rods have been shown to slow release oxytetracycline or diclofenac over several weeks and have shown good wear compliance in humans (144, 145). Liquid silicone can also be used as a surface coat in developing delivery devices. This approach has been reported for polydimethylsiloxane coated solid drug implants that are used to enhance sustained delivery in the eye (41, 42).

CONCLUSIONS

Efficacious delivery of drugs in the eye is dependent on a host of factors including activity at the receptor, absorption and penetration of the drug for the site of application, clearance rates from the biological compartment, delivery rate and duration of the drug from the site of application, toxicology of the total dosage form, and patient compliance with the dosage form. The evolution of acceptable systems as drug carriers in the eye is usually influenced simultaneously by several of these factors. Solution or suspension drops and ointments still remain the first line approach to treatment in standard therapies. However, in circumstances demanding less frequent dosing, or dosing into less accessible compartments of the eye, more unique approaches are indicated. In those cases, use of gelling, eroding or non-eroding polymer systems has proven to be of significant value.

REFERENCES

- Olejnik, O. Conventional Systems in Ophthalmic Drug Delivery. *Ophthalmic Drug Delivery Systems*; Mitra, A.K. Ed.; Marcel Dekker, Inc.: New York, 1993; 177–198.
- King, R.E. *Dispensing of Medication*, 9th Ed. Mack Publishing Co.
- USP Pharmacopeia. *United States Pharmacopeial Convention, Inc.*: Rockville MD, 1995.
- Jani, R.; Gan, O.; Ali, Y.; Rodstrom, R.; Hancock, S. J. Ocul. Pharmacol. **1994**, *10*, 57.
- Gressel, P.D.; Roehrs, R.E., US Patent 5,075,104, 1991.
- Sullivan, K.J.; McCurrach, F.; Lee, S.; Taylor, H.R.; Rolando, M.; Marechal-Courtois, C.; Creuzot, G.C.; East, D.L.; Karabatsas, C.; Bingh-Hoh, M.; Faschinger, C.; Laroche, L. Ophthalmol. **1997**, *104*, 140.
- Hosaka, S.; Ozawa, H.; Tanzawa, H.; Kinitomo, T.; Nichols, R.L. Vivo Evaluation of Ocular Inserts of Hydrogel Impregnated with Antibiotics for Trachoma Therapy Basic Research Laboratories and the Department of Microbiology Harvard School of Public Health, 1982.
- Soronsen, T.B. Acta Ophthalmol. **1984**, *167* (Suppl.), 1.
- Matoba, A.Y.; McCulley, J.P. Ophthalmol., **1985**, *92*, 97.
- North, D.P. Can. Med. Assoc. J. **1971**, *105*, 561.
- Podos, S.M.; Becker, B.; Assoff, C. Am. J. Ophthalmol., **1972**, *73*, 336.
- Kaufman, H.E.; Uotila, M.H.; Gasset, A.R.; Wood, T.O.; Ellison, E.D. Trans. Am. Acad. Ophthalmol. Otolaryngol. **1971**, *75*, 361.
- Krohn, D.L. Trans. Am. Ophthalmol. Soc. **1978**, *76*, 502.
- Krohn, D.L.; Breittfeller, J.M. Invest. Ophthalmol. **1975**, *14*, 152.
- Urtti, A.; Kyyrönen, K.; Finne, U.; Vartianinen, E.; Väisänen, V. Proc. Int. Symp. Contr. Rel. Bioact. Mat. **1990**, *17*, 472.
- Ozawa, H.; Hosaka, S.; Kunitomo, T.; Tanzawa, H. Biomaterials **1983**, *4*, 170.
- Hosaka, S.; Ozawa, H.; Tanzawa, H.; T Kunitomo; Nichols, R.L. Biomaterials **1983**, *4*, 243.
- Mester, U.; Krasemann, C.; Stein, H.J. Klin. Monatsbl. Augenheilkd **1976**, *169*, 492.
- Jain, M.R. Br. J. Ophthalmol. **1988**, *72*, 150.
- Bawa, R. *Ophthalmic Drug Delivery Systems*; Mitra, A.K. Ed.; Marcel Dekker, Inc.: New York, 1993, 223.
- Chapman, J.M.; Cheeks, L.; Green, K. J. Cataract Refract. Surg. **1992**, *18*, 456, 9.
- Heyman, T.P.; McDermott, M.L.; Ubels, J.L.; Edelhauser, H.F. J. Cataract Refract. Surg. **1989**, *15*, 169.
- Shing, Y.; Gonzales, E.M.; Li, W.W.; Casey, R.; Panigrahy, D.; Folkman, J. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1992**, *33*, 1027.
- Baker, R.; Lonsdale, H. Chem. Technol. **1975**, *5*, 668.
- Urquhart, J. *Ophthalmic Drug Delivery System*; Robinson, J.R., Ed.; American Pharmaceutical Association: Washington, DC, 1980; 105.
- Shell, J.W. Ophthalmic Surg. **1974**, *5*, 73.
- Chien, Y.W. *Novel Drug Delivery Systems*; Ch. 2 Marcel Dekker, Inc.: New York, 1982; 13.
- Friederich, R.L. Ann. Ophthalmol. **1974**, *6*, 1279.
- Sihvola, P.; Puustjarvi, T. Acta Ophthalmol. **1980**, *58*, 933.
- Akerblom, T.; Aurell, E.; Christiansson, J.; Kriisa-Kunnos, V.; Wiebert, O. Acta Ophthalmol., **1980**, *58*, 617.
- Pollack, I.P.; Quigley, H.A.; Harbin, T.S. South Med. J. **1976**, *69*, 1296–9.
- Zimmer, A.K.; Chetoni, P.; Saettone, M.F.; Zerbe, H.; Kreute, J. J. Contr. Rel. **1995**, *33*, 31.
- Gurtler, F.; Kaltsatos, V.; Boisrame, B.; Gurny, R. J. Contr. Rel. **1995**, *33*, 231.
- Yakovlev, A.A.; Lenkevich, M.M. Vestn. Ophthalmol. **1966**, *79*, 40.
- Maichuk, Y.F. Antibiotiki **1967**, *5*, 435.
- Maichuk, Y.F. Invest. Ophthalmol. **1975**, *14*, 87.
- Maichuk, Y.F. Lancet **1975**, *173*.
- Saettone, M.F.; Giannaccini, B.; Chetoni, P.; Galli, G.; Chiellini, E. J. Pharm. Pharmacol. **1984**, *36*, 229.
- Richardson, M.C.; Bentley, P.H. *Ophthalmic Drug Delivery Systems*; Mitra, A.K., Ed.; Marcel Dekker, Inc.: New York, 1993; 355.
- Pearson, P.A.; Hollins, J.C.; Baldwin, L.B.; Smith, T.J. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1989**, *30*, 511.
- Velez, G.; Robinson, M.R.; Durbin, T.; Yuan, P.; Sung, C.; Whitcup, S.M. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1999**, *40*, 584.
- Robinson, M.R.; Yuan, P.; Gogolak, L.; Tansey, G.; Smith, J.; Whitcup, S.M. ARVO Abstract. Invest. Ophthalmol., **1999**, *40*, S84.
- Gogolak, L.; Robinson, M.R.; Yuan, P.; Aghera, A.; Sung, C.; Whitcup, S.M. ARVO Abstract. Invest. Ophthalmol. **1999**, *40*, S85.
- Mondino, B.J. Am. J. Ophthalmol., **1991**, *112*, 587.
- Robin, J.B.; Keys, C.L.; Kaminski, L.A.; Viana, M.A. Invest. Ophthalmol. Vis. Sci. **1990**, *31*, 1294.
- Harrison, K.W. J. Ophthalmic Nurs. Technol. **1989**, *8*, 97.
- Fridberg, M.L.; Pleyer, U.; Mondino, B.J. Ophthalmol. **1991**, *98*, 725.
- Bloomfield, S.E.; Miyata, T.; Dunn, M.W.; Bueser, N.; Stenzel, K.H.; Rubin, A.L. Arch. Ophthalmol. **1978**, *96*, 885.
- Silbiger, J.; Stern, G.A. Ophthalmol. **1992**, *99*, 889.
- Brockman, E.B.; Hobden, J.A.; Hill, J.M.; O'Callaghan, J.; Reidy, J.J.; Insler, M.S.; Kaufman, H.E. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1990**, *31*, 558.
- Hobden, J.A.; Reidy, J.J.; O'Callaghan, R.J.; Hill, J.M.; Insler, M.; Rootman, D. Arch. Ophthalmol. **1988**, *106*, 1605.
- Foster, J.A.; Robin, J.B.; Shipplett, C.; Viana, M.; Foster, F.J. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1991**, *32*, 1291.
- Liang, F.-Q.; Viola, R.S.; Nakada, S.; del Cerro, M.; Aquavella, J.V. Invest. Ophthalmol. Vis. Sci. **1992**, *33*, 2194.
- Baziuk, N.; Gremillion, C.M.; Peyman, G.A.; Cho, H.K. Int. Ophthalmol. **1992**, *16*, 101.
- Unterman, S.R.; Rootman, D.S.; Hill, J.M.; Parelman, J.J.; Thompson, H.W.; Kaufman, H.E. J. Cataract Refract. Surg. **1988**, *14*, 500.
- Duzman, E.; Chen, C.C.; Takruri, H. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1991**, *32*, 1291.
- Sawusch, M.R.; O'Brien, T.P.; Dick, J.D. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1988**, *29*, 228.
- O'Brien, T.P.; Sawusch, J.D.; Hamburg, T.R. J. Cataract Refract. Surg. **1988**, *14*, 505.

59. Shofner, R.S.; Kaufman, H.E.; Hill, J.M. *Ophthalmol. Clin. North Amer.* **1989**, *2*, 15.
60. Unterman, S.R.; Rootman, D.S.; Hill, J.M.; Kaufman, H.E. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1988**, *29*, 52.
61. Poland, D.E.; Kaufman, H.E. *J. Cataract Refract. Surg.* **1988**, *14*, 489.
62. Assil, K.K.; Zarnegar, S.R.; Fouraker, B.D.; Schanzlin, D.J. *Amer. J. Ophthalmol.*, **1992**, *113*, 418.
63. Punch, P.I.; Costa, N.D.; Edwards, M.E.; Wilcox, G.E. *J. Vet. Pharmacol. Ther.* **1987**, *10*, 37.
64. Esswein, M.B.; O'Brien, T.P.; Osato, M.S.; Jones, D.B. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1990**, *31*, 452.
65. Schwartz, S.D.; Harrison, S.A.; Engstrom, R.E.; Bowden, R.E.; Lee, D.A.; Mondino, B.J. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1990**, *31*, 558.
66. Hwang, D.G.; Stern, W.H.; Hwang, P.H.; MacGowan-Smith, L.A. *Arch. Ophthalmol.* **1989**, *107*, 1375.
67. Hwang, D.G.; Stern, W.H.; Alyeat, E.B.; MacGowan-Smith, L.A. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1989**, *30*, 248.
68. Aquavella, J.V.; Ruffini, J.J.; LoCascio, J.A. *J. Cataract Refract. Surg.* **1988**, *14*, 492.
69. Sawusch, M.R.; O'Brien, T.P.; Updegraff, S.A.; Gottsch, J.D. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1989**, *30*, 479.
70. Finkelstein, I.; Trope, G.E.; Menon, I.A.; Rootman, D.S.; Basu, P.K. *Curr. Eye Res.* **1990**, *9*, 653.
71. Finkelstein, I.; Trope, G.E.; Menon, I.A.; Spero, L.; Rootman, D.S.; Heathcoat, J.F. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1990**, *31*, 591.
72. Sachdev, S.; Zou, X.; Higgenbotham, E. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1990**, *31*, 3.
73. Hasty, B.; Hever, D.K.; Minckler, D.S. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1989**, *30*, 418.
74. Gussler, J.R.; Ashton, P.; VanMeeter, W.S.; Smith, T.J. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1990**, *31*, 485.
75. Laucuta, S.E. *Int. J. Pharmaceut.* **1989**, *54*, 71.
76. Kitazawa, Y. *Acta Soc. Ophthalmol. Jpn.* **1975**, *79*, 1715.
77. Vasantha, R.; Sehgal, P.K.; Panduranga Rao, K. *Int. J. Pharm.* **1988**, *47*, 95.
78. Ball, S.; Badeaux, J.; Graham, M.; Balazy, M.; Bazan, N.G. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1990**, *31*, 405.
79. Nakada, S.; Viola, R.S.; Aquavella, J.V. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1991**, *32*, 1292.
80. Murray, T.G.; Jaffe, G.J.; McKay, B.S.; Han, D.P.; Burke, J.M.; Abrams, G.W. *Refract. Corneal Surg.* **1992**, *8*, 44.
81. Reidy, J.J.; Gebhardt, B.M.; Kaufman, H.E. *Cornea* **1990**.
82. Moritera, T.; Ogura, Y.; Arai, M.; Honda, Y.; Wada, R.; Hyon, S.-H.; Ikada, Y. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1989**, *30*, 249.
83. Moritera, T.; Ogura, Y.; Yoshimura, N.; Honda, R.; Wada, R.; Hyon, S.-H.; Ikada, Y. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1990**, *31*, 384.
84. Moritera, T.; Ogura, Y.; Yoshimura, N.; Kuriyama, S.; Nakano, T.; Honda, Y.; Tabata, Y.; Ikada, Y. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1991**, *32*, 1294.
85. Moritera, T.; Ogura, Y.; Kimura, H.; Misumi, C.; Takeuchi, A.; Honda, Y.; Ikada, Y. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1992**, *33*, 1012.
86. Moritera, T.; Ogura, Y.; Honda, Y.; Wada, R.; Hyon, S.-H.; Ikada, Y. *Invest. Ophthalmol. Vis. Sci.* **1991**, *32*, 1785.
87. Peyman, G.A.; Conway, M.; Khoobehi, B.; Soike, K. *Int. Ophthalmol.* **1992**, *16*, 109.
88. Davis, P.A.; Kiss, K.; Parrish, R.K.; Villain, F.; Parel, J.-M.; Cousins, S.W. *Proc. Int. Symp. Contr. Rel. Bioact. Mat.* **1992**, *19*, 339.
89. Rubsamen, P.F.; Davis, P.; Hernandez, E.; Parel, J.-M.; Cousins, S.W. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1992**, *33*, 728–737.
90. Portugal, L.; Parel, J.-M.; Parrish, R.; Bichon, D. *Proc. Int. Symp. Contr. Rel. Bioact. Mat.* **1989**, *16*, 105.
91. Sheardown, H.; Chiu, L.-K.; Chiu, W.-J.; Trope, G.E.; Cheng, Y.-L.; Rootman, D.S.; Gould, L.; Liu, G.S. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1992**, *33*, 1390.
92. Gould, L.; Trope, G.E.; Morrison, H.; Cheng, Y.-L.; Rootman, D.S.; Heathcoat, G.; Menon, I.A.; Rootman, D.S.; Liu, G.S.; Chiu, L.-K.; Basu, P.K.; Lyn, W. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1991**, *32*, 1122.
93. Khoobehi, B.; Stradtman, M.O.; Peyman, G.A.; Aly, O.M. *Ophthalm. Surg.* **1990**, *21*, 840.
94. Moritera, T.; Ogura, Y.; Yoshimura, N.; Honda, Y.; Wada, R.; Hyon, S.-H.; Ikada, Y. *Invest. Ophthalmol. Vis. Sci.* **1992**, *33*, 3125.
95. Kimura, H.; Ogura, Y.; Moritera, T.; Honda, Y.; Wada, R.; Hyon, S.-H.; Ikada, Y. *Invest. Ophthalmol. Vis. Sci.* **1992**, *33*, 3436.
96. Maincent, P.; Marchal-Heussler, L.; Sirbat, D.; Thouvenot, P.; Hoffman, M.; Vallet, J.A. *Proc. Int. Symp. Contr. Rel. Bioact. Mat.* **1992**, *19*, 226.
97. Marchal-Heussler, L.; Sirbat, D.; Hoffman, M.; Maincent, P. *Pharm. Res.* **1992**, *10*, 386.
98. Vidmar, V.; Pepeljnjak, S.; Jalsenjak, I. *Microencapsul. J.* **1985**, *2*, 289.
99. Masson, V.; Billardon, P.; Fessi, H.; Devissaguet, J.P.; Puisieux, P. *Proc. Int. Symp. Contr. Rel. Bioact. Mat.* **1992**, *19*, 423.
100. Kimura, H.; Ogura, Y.; Moritera, T.; Honda, Y.; Hyon, S.-H.; Ikada, Y. *Invest. Ophthalmol. Vis. Sci.* **1992**, *33*, 1013.
101. Herrero-Vanrell, R.; Alvarez-Santiago, C.; Barcia, E. *Invest. Ophthalmol.* **1999**, *40*, S84.
102. Lee, D.A.; Leong, K.W.; Panek, W.C.; Eng, C.T.; Glasgow, B.J. *Invest. Ophthalmol. Vis. Sci.* **1988**, *29*, 1692.
103. Lee, D.A.; Flores, R.A.; Anderson, P.J.; Leong, K.W.; Teekhasaene, C.; DeKater, A.W.; Hertzmark, E. *Ophthalmol.* **1987**, *94*, 1523.
104. Jampel, H.D.; Leong, K.W.; Dunkelberger, G.; Quigley, H. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1989**, *30*, 417.
105. Lewis, H.; Schwartz, S.; Lee, D.A.; Leong, K.W. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1991**, *32*, 1047.
106. Alonso, J.I.; Pastor, J.C.; Lopez, I.; Mate, A.; Mateu, C.; Aberola, A.; Sañudo, C. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1991**, *32*, 1293.
107. Tabatabay, C.; Einmahl, S.; Behar-Cohen, F.; Gurny, R. *Invest. Ophthalmol.* **1999**, *40*, S84.
108. Dukes, A.J.; Rabowsky, J.H.; Lee, D.A.; Leong, K.W. *ARVO Abstract. Invest. Ophthalmol. Vis. Sci.* **1992**, *33*, 1391.
109. Charles, J.B.; Ganthier, R. Jr; Wilson, M.R.; Lee, D.A.; Baker, R.S.; Leong, K.W.; Glasgow, B.J. *Ophthalmol.* **1991**, *98*, 503.

110. Koya, P.; Jampel, H.; Quigley, H.; Leong, K.W. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1991**, 32, 745.
111. Merkli, A.; Wüthrich, P.; Heller, J.; Tabatabay, C.; Gurny, R. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1992**, 33, 1392.
112. Gaynon, M.; Heller, J.; Wüthrich, P. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1991**, 32, 883.
113. Wood, R.W.; Li, V.H.K.; Kreuter, J.; Robinson, J.R. Int. J. Pharm. **1985**, 23, 175.
114. Diepold, R.; Kreuter, J.; Guggenbuhl, P.; Robinson, J.R. Int. J. Pharm. **1989**, 54, 149.
115. Li, V.H.; Wood, R.W.; Kreuter, J.; Hamia, T.; Robinson, J.R. Microencapsul. J. **1986**, 3, 213.
116. Zimmer, A.; Kreuter, J.; Robinson, J.R. Microencapsul. J. **1991**, 8, 497.
117. Das, S.K.; Tucker, I.G.; Davies, N.M. Proc. Int. Symp. Contr. Rel. Bioact. Mat. **1992**, 19, 395.
118. Harmia, T.; Speiser, P.; Kreuter, J. J. Microencapsul. **1986**, 3, 3.
119. Harmia, T.; Kreuter, J.; Speiser, P.; Boye, T.; Gruny, R.; Kubis, A. Int. J. Pharm. **1986**, 33, 187.
120. Diepold, R.; Kreuter, J.; Himber, J.; Gurny, R.; Lee, V.H.K.; Robinson, J.R.; Saettone, M.F.; Schnaudigel, O.E. Graefe's Arch. Clin. Exp. Ophthalmol. **1989**, 227, 188.
121. Cheeks, L.; Green, K.; Stone, R.P.; Riedhammer, T. Curr. Eye Res. **1989**, 8, 1251.
122. Desai, S.D.; Blanchard, J. S-200 Abstract PDD 7125. Pharm. Res. **9** **1992**.
123. Goins, K.M.; Blandford, D.L.; Pearson, P.A.; Schmeisser, E.; Hollins, J.C.; Smith, T.J. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1990**, 31, 364.
124. Smith, T.J.; Pearson, P.A.; Blandford, D.L.; Brown, J.D.; Goins, K.A.; Hollins, J.L.; Schmeisser, E.T.; Glavinos, P.; Baldwin, L.B.; Ashton, P.A. Arch. Ophthalmol. **1992**, 110, 255.
125. Sanborn, G.E.; Anand, R.; Torti, R.E.; Nightingale, S.D.; Cal, S.X.; Yates, B.; Ashton, P.; Smith, T.J. Arch. Ophthalmol. **1992**, 110, 188.
126. Baldwin, L.B.; Pearson, P.A.; Smith, T.J. Proc. Int. Symp. Contr. Rel. Bioact. Mat. **1990**, 17, 417.
127. Blandford, D.L.; Smith, T.J.; Brown, J.D.; Pearson, P.A.; Ashton, P. Invest. Ophthalmol. Vis. Sci. **1992**, 33, 3430.
128. Smith, T.J.; Dharma, S.; Rafii, M.; Ashton, P. ARVO, Invest. Abstract. Ophthalmol. Vis. Sci. **1992**, 33, 737.
129. Pearson, P.A.; Solomon, K.D.; Van Meter, W.S.; Smith, T.J.; Ashton, P. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1991**, 32, 797.
130. Amrien-Noll, J.; Brown, J.D.; Ashton, P.; Conklin, J.; Gussler, J.; Smith, T.J.; Hollins, J.C. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1991**, 32, 797.
131. Hainsworth, D.P.; Rafii, M.; Conklin, J.D.; Ashton, P. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1993**, 34, 1491.
132. Pearson, P.A.; Jaffe, G.J.; Martin, D.F.; Ashton, P. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1993**, 34, 1492.
133. Rafii, M.; Blandford, D.L.; Ashton, P. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1993**, 34, 1492.
134. Baker, C.W.; Hainsworth, D.P.; Conklin, J.D.; Ashton, P. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1993**, 34, 1492.
135. Yang, D.C.S.; Wang, Y.J.; Chang, W.C.; Wang, W.W.; Liu, J.H. Invest. Ophthalmol. **1999**, 40, S87.
136. Bawa, R.; Nandu, M.; Downie, W.; Robinson, J.R. Proc. Int. Symp. Contr. Rel. Bioact. Mat. **1989**, 16, 213.
137. Sutinen, R.; Laasanen, V.; Paronen, P.; Urtti, A. J. Contr. Rel. **1995**, 33, 163.
138. Urtti, A.; Pipkin, J.D.; Rork, G.; Repta, A.J. Int. J. Pharm. **1990**, 61, 235.
139. Urtti, A.; Pipkin, J.D.; Rork, G.; Sento, T.; Finne, U.; Repta, A.J. Int. J. Pharm. **1990**, 61, 241.
140. Darougar, S. US Patent 5, 147,647, 1992.
141. Weiner, A.L.; Darougar, S.; Siddque, M.; Raul, V. ARVO Abstract. Invest. Ophthalmol. Vis. Sci. **1993**, 34, 1489.
142. Weiner, A.L.; Darougar, S.; Siddque, M.; Raul, V. Proc. Int. Symp. Contr. Rel. Bioact. Mat. Abstract **1993**, 20, 1222.
143. Darougar, S.; Weiner, A.L. US Patent 5,322,691, 1994.
144. Peiffer, R.L.; Osborne, C.M.; Smith, P.C.; Wang, C.; Weiner, A.L. Invest. Ophthalmol. **1995**, 36, S1021.
145. Hubbell, H.R.; Noonan, J.S.; Lee, S.B.; Swaminathan, A.; Kaminski, E.A.; Edelhauser, H.F. Invest. Ophthalmol. **1999**, 40, S85.

DRUG DELIVERY—NASAL ROUTE

René Bommer

E. Pfeiffer GmbH, Radolfzell, Germany

INTRODUCTION

Today, nasal drug delivery is receiving much attention from the pharmaceutical industry. About 2% of the overall drug delivery is administered via the nasal route. A survey (1) among decision makers in the pharmaceutical industry highlights the importance of this delivery mode. A transmucosal drug delivery route (which includes the nasal route) can target tissue and use active transdermal processes. It is also regarded as a major influence of the market.

In addition, the most popular view of nasal drug delivery is the administration of locally acting products. Topical decongestants or anti-inflammatory drugs used to treat a rhinitis or allergy related indications are well-known drug products. They expand their effect directly at the focus of the disease.

On the other hand, the nasal route is an attractive alternative to invasive administrations, and provides a direct access to the systemic circulation. Certain drugs that are administered intranasally are able to penetrate the nasal mucosa and enter the system. Intranasal application of drugs or pharmacologically active compounds has been a subject of medical knowledge since the beginning of human civilization. In India, ayurvedic medicine uses the advantage of the nasal route. Sniffing tobacco or hallucinogens is used widely throughout all cultures in the world. A rapid onset of action is one advantage of the nasal route for the administration of systemically acting products.

The attractiveness of this noninvasive route is clearly reflected in the increase of approval rates over the past few years. Locally acting products have an average growth rate of approximately 10% per year. The growth rate of systemically acting products is approximately 30%. It is remarkable that more than half of the systemically acting products were approved during the 1990s (2).

The administration of systemically acting products via the nasal route began in the 1980s. The peptide oxytocin, which stimulates uterine contraction and lactation, was one of the first nasally administered peptide hormones. Meanwhile, several peptide-based nasal formulations entered the market. Currently, more attention is being

paid to this delivery system due to the increasing demands of new highly potent drug formulations. In addition, patients' expectations for successful therapy have to be considered. Today's requirements for nasal delivery systems include three key elements: Reliability, safety, and efficacy. These must be taken into account in the development of dispensing systems for locally acting products and systemically acting products.

Nasal drug delivery devices can be divided into multidose and unit-dose/bi-dose systems. Multidose systems consist of a container mounted with a mechanical pump dispenser that is designed to deliver multiple doses from one container (Fig. 1). The various requirements of the customer and of the formulation are respected in the design and performance of the delivery systems. Unit-dose and bi-dose systems are becoming increasingly attractive to the pharmaceutical industry. In particular, therapies that require precise performance in their delivery system employ these single/dual-use disposable systems.

The nose, or more precisely the nasal cavity, is the target of the administration of a drug product. The anatomy and the physiology of the nose play a decisive role in efficient drug administration. The nasal mucosa is much more sensitive to external influences than the digestive mucosa in the stomach. On the other side, nasal administration often requires a smaller amount of drug; therefore, fewer side effects are expected. Nasal administration has several advantages. First, deposition of an active compound in the nasal cavity results in avoidance of its degradation through the "first-pass" metabolism. Second, enzymatic breakdown of the drug in the nose can be neglected. Third, the onset of action of the drug is more rapid and even comparable to an invasive route.

NASAL ANATOMY AND PHYSIOLOGY

The nose actively contributes to two major functions of the human system. The first function is the sense of smell (olfaction) and the second is respiration or breathing.

The nasal septum divides the nasal cavity into left and right halves. The nasal septum is never a straight vertical



Fig. 1 Various multidose systems.

separation of the two cavities. Through normal growth and development, the nasal septum folds together and creates an asymmetry of the two cavities, which results in variations from individual to individual. It is important to realize that the nasal cavities are not round hollow bodies; instead, they are long high slits. The nasal septum is not very accessible for the penetration of drugs into the human system since it consists mostly of cartilage and skin. The most efficient area for drug administration is the lateral walls of the nasal cavity, which consist of highly vascularized tissue, the mucosa.

The surface of the nasal cavity is approximately 150 cm². Inside the cavity, the mucosa is covered with the ciliary epithelium. The ciliary activity is the driving force of the secretory transport in the nose. Approximately 1 L of mucus is transported from the anterior part to the posterior part of the nose per day. The mucus in the rear of the cavity is removed by swallowing. The function of the ciliary activity is to remove any particles that are trapped on the mucus blanket during inhalation. It takes approximately 20–30 min for the whole mucus layer to be renewed.

The nasal mucosa is a natural site for the perception of environmental influences. The nose, with its numerous circulatory and secretory activities, is perfectly qualified for reflex conditioned reactions. One of the most remarkable autonomic reflexes is the so-called nasal cycle. Most people have a cyclic increase in the resistance of one nasal passage. However, the overall resistance remains constant. This variation is due to the swelling of the turbinates. The swelling and subsequent clogging of one nasal passage underlies a cycle time of approximately 4 h. The nose is a very complex organ. For an effective administration of

therapeutic drugs through the nasal mucosa, the following must be taken into consideration (3):

- The method and technique of administration
- The site of drug deposition
- The rate of clearance through the ciliary activity
- The pathological condition of the nose

TRADITIONAL NASAL DISPENSING SYSTEMS

Traditional application systems consist of nasal drops, pipettes, or squeeze bottles. The use of pipettes and nasal drops may work for certain drug formulations but the benefit in terms of reliability, safety, and efficacy is controversial. From an anatomical viewpoint, nasal drops may be suitable for infants only. Since the nasal cavity of an infant is so small, one or two drops can cover the whole mucosa. From the pharmacological viewpoint, the administration of drops to infants and children is controversial. For example, due to insufficient dosing accuracy, scientists do not recommend administering steroids as nasal drops to children (4). In adults, drops into the nasal cavity mostly lead to a rapid clearance of the drug along the floor of the nasal cavity toward the throat. Studies demonstrate a longer duration of sprayed products on the nasal mucosa than formulations administered as drops (5). To achieve a comparable administration mode in adults, the whole nasal cavity has to be flooded with an amount of 20–30 mL. This is pharmacologically not realistic, nor is it acceptable to the patient.

The squeeze bottle was an improvement in the deposition of drugs. The expelled spray-like product is able to reach a larger surface of the nasal mucosa than a drop. One of the major disadvantages of the squeeze bottle is its dose variability. The spray volume and plume geometry, i.e., spray angle and particle size distribution, are highly dependent on the pressure put on the bottle. In addition, the dosage and the plume geometry are influenced by two features present in a squeeze bottle, namely, the liquid formulation and the air. During the period of use, the liquid level in the container decreases and therefore, the ratio of liquid and air changes. The result is a variable spray performance that changes from one actuation to the next. Due to its unreliable performance, squeeze bottles are not recommended for administration of vasoconstrictors to children (6).

A squeeze bottle is regarded as an “open” system. There is no valve or similar mechanism that seals acceptably and prevents contamination of the contents.

The nasal tip, which is usually in contact with the nostril during actuation, naturally becomes contaminated. After releasing the pressure on the bottle, the back flow of the liquid enables bacteria to enter the system. After appropriate incubation time, contamination will occur in the container. Of course, preservatives do oppose the bioburden, but the efficacy of a preservative is limited.

Today's high potency drugs, intended for local and systemic treatments, depend on reliable delivery systems. As an alternative to squeeze bottles and pipettes, propellant-driven or mechanical-dispensing systems are often used. Aerosol systems are well known in inhalation therapy. They are ready to use and easy to handle. While new propellants will replace the old ones, the environmental and pharmacological discussions will continue. The switch to new propellants must be supported by sufficient clinical and toxicological data. In particular, compatibility problems must be addressed. Furthermore, attention must be paid to the surfactants not under dispute.

For the consumer, the sensation of a cold puff into a nostril, with resulting irritation through surfactants and propellants, is not appreciated. A further disadvantage of pressurized systems is the power of the impact of the drug formulation on the nasal mucosa. The deposition occurs on a relatively limited area and therefore, the drug is not distributed evenly over the mucosa. Studies recommend that two doses be administered at different angles into the nostril in order to cover a sufficient area of the mucosa (7). It is recommended that physicians instruct their patients on the proper use of prescribed nasal aerosols.

MECHANICAL DISPENSING SYSTEMS

Mechanical dispensing systems are an alternative to pressurized systems. The devices are not limited to spray solutions. The range of available designed actuators means that suspensions, creams, ointments, and gels also can be dispensed.

The heart of any mechanical dispensing system is its basic pump, or the motor (Fig. 2). At the bottom of the pump is a dip tube, which is the connection to the formulation in the container. An actuator is mounted on the top of the pump. Before using the mechanical dispensing system, a number of priming strokes are required to displace the air in the system and to fill the volume chamber with the targeted nominal volume. A major functional group consists of the preassembly, a combination of the moveable parts inside the pump system. As the preassembly is pushed down during

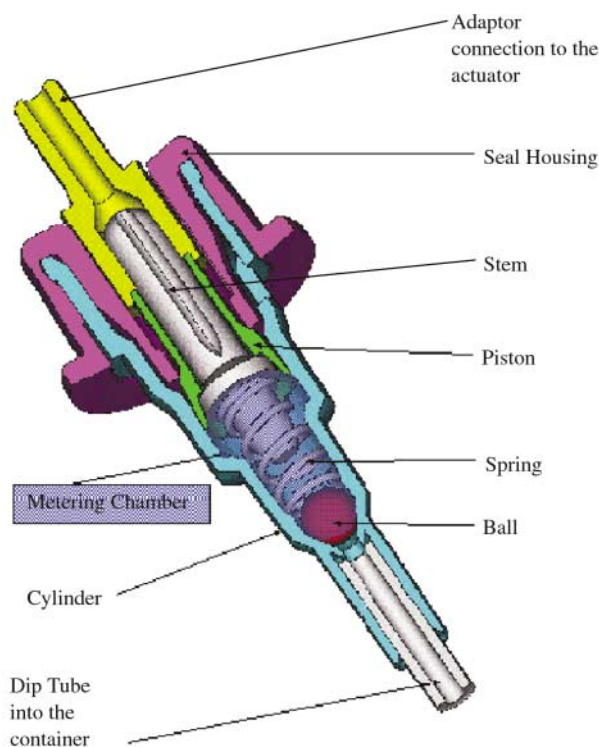


Fig. 2 Mechanical basic pump.

actuation, it compresses the spring and forces the ball into the ball seat. The increasing pressure on the liquid inside the volume chamber opens a valve mechanism and releases the amount of drug formulation through the actuator. The backstroke creates a certain vacuum in the metering chamber. The vacuum lifts the ball out of the ball seat and allows the product to refill the pump system via the dip tube.

Different kinds of pump mechanisms are on the market. Essentially, they follow the same principles. The first step is priming, or filling the volume chamber in the pump through suction. The second step is the displacement of the desired amount in the volume chamber. The third step is the actual dispensing.

Depending on a patient's requirements, the pump dispenser can be fixed to the container via a variety of different closures, such as screw closure, crimp, or snap-on. The most important parameters that define the performance of the dispensing system are the dosage volume, the spray angle, and the particle size distribution. Today, the dispensing industry offers a broad range of dosage volumes, from 25 to 1000 μL . The dimension of the metering chamber of the respective pump (Fig. 2) determines the dosage volume.

Parameters that most influence the behavior of a spray are the viscosity, the thixotropy, the surface tension, and

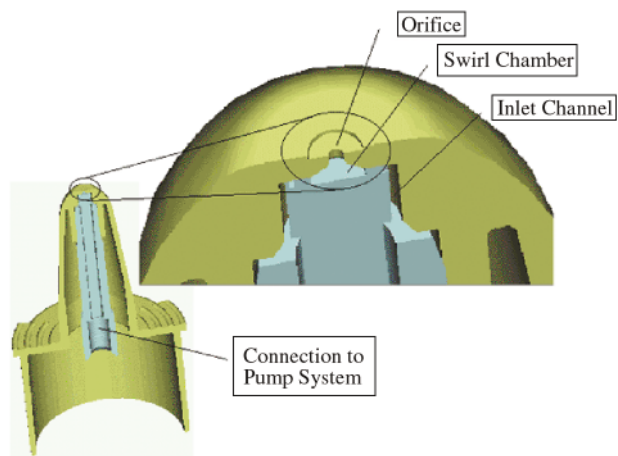


Fig. 3 Cross-section of a nasal actuator.

the density. Modification of the swirl chamber and the inlet channels makes it possible to adjust the spray performance to the patient’s requirements (Fig. 3). The particle size distribution can be altered by varying the dimensions and the geometry of the orifice, as well as the pressure build-up in the volume chamber prior to dispensing. One of the major criteria is to keep the fraction of extremely small particles very low to avoid partial inhalation of the drug formulation.

The increasing demands of the pharmaceutical industry have led to improvements in existing devices and the completion of new concepts.

Different consumer needs, whether for infant, child, or adult, varies the concentrations of the active ingredients in a nasal drug formulation. Fig. 4 shows a typical nasal spray

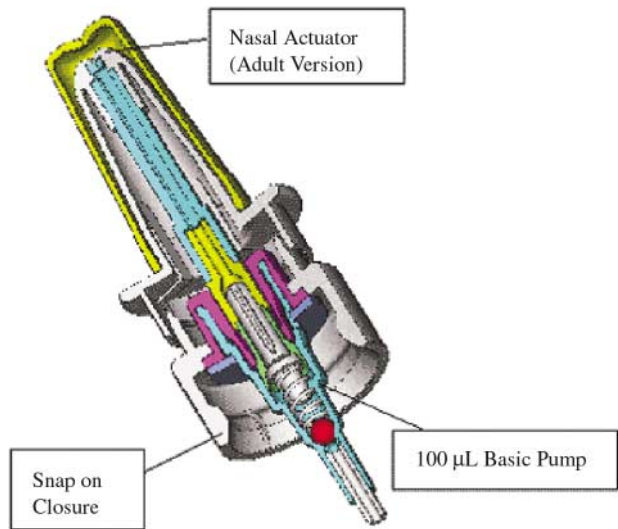


Fig. 4 100 µL nasal spray system with snap-on closure.

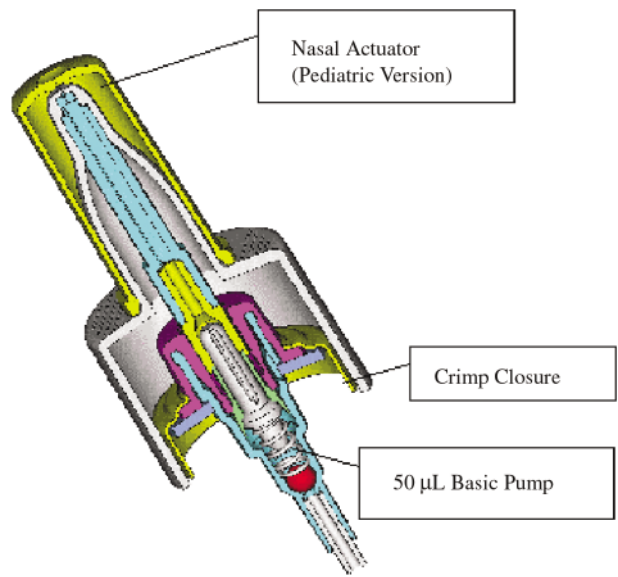


Fig. 5 50 µL Nasal spray system with crimp closure.

system fitted with a snap-on closure. The dispensing system can be adapted to the anatomy of the patient. The actuators for pediatric application are slimmer in their geometry and the dosage volume is reduced (Fig. 5).

Finger flanges have been modified to meet the requirements of elderly people who suffer from arthritis or rheumatism. Actuators can be adapted to deliver drug formulations to animals for veterinary purposes or for preclinical studies. Delivery systems can be fitted out with an upside down pump, which allows actuation in upright and upside down position.

A dispensing system fitted with a counting mechanism also is available (Fig. 6). It allows the consumer to monitor each actuation, beginning from the first priming stroke until delivery of the last dose. A dispensing system for

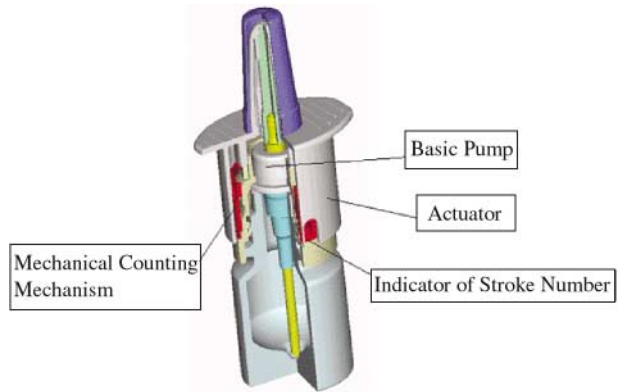


Fig. 6 Nasal spray system with integrated counting mechanism.

controlled substances, such as morphine, can be equipped with a lock-out system. The microchip controlled actuation mode takes into account the patient's specific requirements, such as amount of drug and frequency of administration. Other relevant data are recorded and can be evaluated later by a physician.

PRESERVATIVE-FREE SYSTEMS

A current trend is the development of nasal formulations without preservatives. In long-term treatments, the frequent delivery of preservatives has an irritating effect on the ciliary activity of the nasal mucosa. Economical reasons also play a key role. By reformulating an existing product, a pharmaceutical company can achieve patent extension. The challenge is to create a dispensing system that prevents a preservative-free formulation from being contaminated during the period of usage.

Different systems of preservative-free mechanical pump dispensers are now available. Basically, the prevention of contamination can be achieved in two ways—a chemical and a mechanical way. The implementation of chemical additives in the dispensing system has the advantage of reliable action of the additive. However, most of the additives that are regarded as safe “disinfectants” are not totally effective against certain strains of bacteria or fungi. With a purely mechanical preservative-free system, the contamination source remains independent. A special mechanism, inserted in the nasal actuator, seals at the very end of the tip (Fig. 7). Through actuation of the pump and due to the fact that liquid cannot be compressed, a certain hydraulic pressure is built up in the nasal actuator. As soon as the hydraulic pressure is higher than the spring force, the sealing pin will be forced backward and the liquid is dispensed via the orifice. The combination of high pressure and a small diameter of the orifice allows the liquid to leave the dispensing system at a very high velocity. A possible contamination source has to swim against the flow, which is realistically not possible. This sealing system offers not only protection against contamination but also prevents the evaporation of the remaining drug solution and a subsequent precipitation or crystallization of the solid ingredients in the actuator.

One feature of the preservative-free system is that contamination is prevented from entering the packaging from the actuator. When dispensing from an airtight container, a vacuum will be built up in the system. Two options can be used to keep the pressure balance. Either a form of ventilation is installed or the container adapts itself

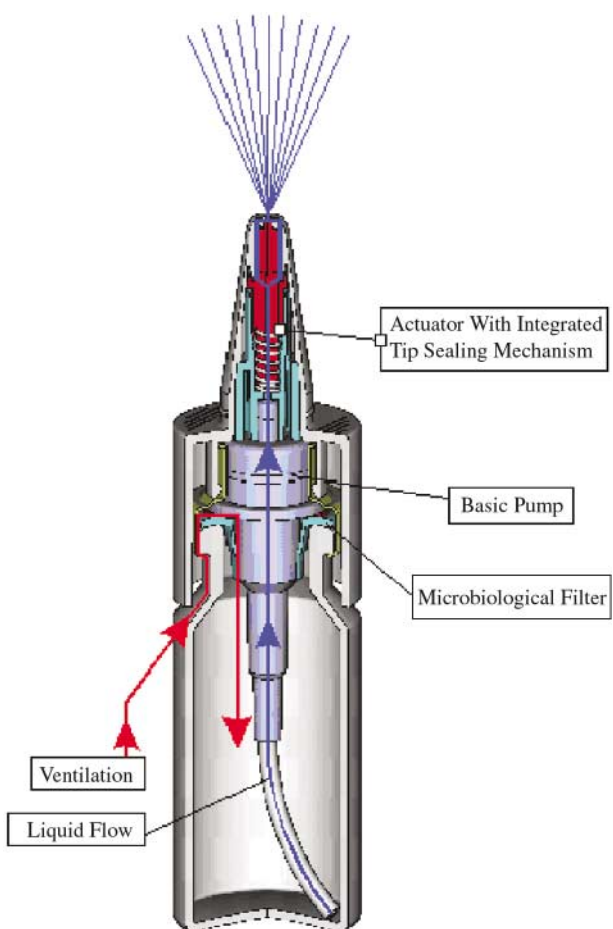


Fig. 7 Preservative-free system.

to the decreasing volume. In the first option, the airflow passes a microbiological filter before entering the container (Fig. 7). In this case, the conventional gasket is replaced with a gasket made of gamma-radiation-resistant material. The filtering gasket prevents any contamination from entering the system (8). The second method employs a collapsible bag or a sliding piston. However, the collapsible bag has to be investigated carefully with regard to compatibility with the formulation and possible permeability of the product through the packaging. Because the rubber formulations and glass that are available for the sliding piston system are already used in syringes, the pharmaceutical industry might prefer this method to the multilayer collapsible bag. The sliding piston glass container, in combination with a tip-sealing mechanism on the actuator, gives the advantage of a preservative-free and airless dispensing system (Fig. 8). From the consumer's point of view, this system offers the advantages of a residual volume, which can be neglected, and in particular, a 360° performance.

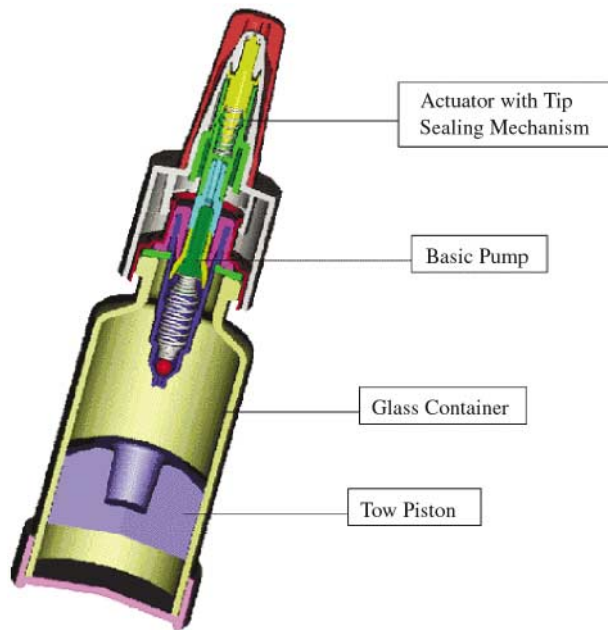


Fig. 8 Preservative-free and airless system.

UNIT-DOSE AND BI-DOSE SYSTEMS

Unit-dose and bi-dose systems are designed to deliver one or two doses into the nostril(s) (Fig. 9). As compared to

multidose pump systems, unit-dose and bi-dose systems are distinguished by a different actuation principle. The dose volume is predetermined by the prefilled glass vial and sealed with a rubber stopper. The glass and rubber are the identical materials used in syringes. The benefit is an optimal protection against environmental influences. Unit-dose and bi-dose systems can be sterilized, and an aseptic filling procedure justifies the omission of preservatives. A pressure-point-controlled actuation mechanism, which builds on the inertia of the actuation finger, guarantees constant performance in terms of delivery volume, plume geometry, and particle size distribution. A manipulation of the spray performance is almost impossible.

Referring to the key words of nasal delivery systems—reliability, safety, and efficacy—the unit-dose and bi-dose dispensing systems meet these demands to the highest degree. In addition, these attributes are a versatile tool for quality control. The pressure point controlled actuation mechanism allows the spray to be independent of the user. Therefore, quality inspection will deliver more reliable results since influence from any individual, in this case the quality inspector, is almost impossible. The patient appreciates the features of a ready, easy-to-use device. The package is small in size, discrete, and works in every position. Another advantage is individual brand recognition due to the many design types on the market.

For delicate drug formulations, the industry offers a modified unit-dose system. The lyophilized ingredient(s)

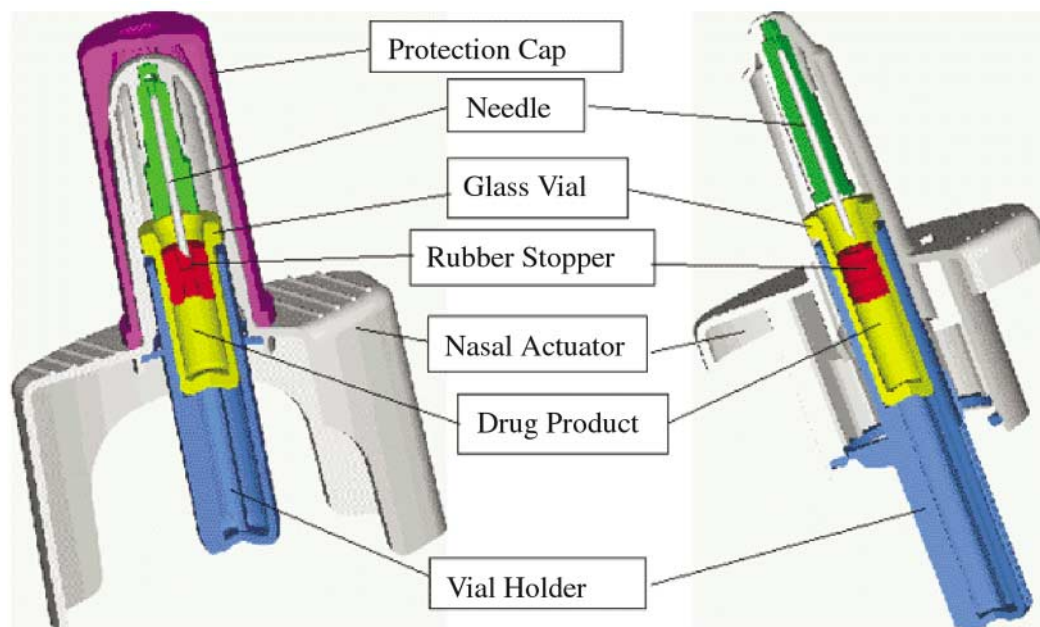


Fig. 9 The unit-dose system and the bi-dose system.

are stored separately from the solvent. Prior to administration, the drug formulation is reconstituted and ready to be dispensed.

POWDER DISPENSING SYSTEMS

Nasal drug delivery is not limited to liquid drug formulations or suspensions. Powder dispensing systems can be more advantageous in dispensing many drug substances. In particular, the biotechnological drugs, mostly represented by proteins and peptides, are more stable in a dry and solid state. Studies demonstrate longer remaining times of powder formulations on the nasal mucosa as compared to liquid formulations (9). The nasal powder processing in terms of the particle size distribution is not nearly as dramatic as it is for dry powders intended for inhalation.

Recently, nasal powder vaccines have been investigated. Vaccines are usually formulated as liquid injectables or as lyophilized powder, which have to be reconstituted prior to administration. A high percentage of the world's vaccines have to be discarded due to the uncertainty of their potency after breaks in the cold chain and possible damage from exposure to high temperatures (10). Regarding physical stability, a dry solid formulation ensures a much higher stability against environmental influences. Influenza vaccines, which are administered via the nasal mucosal route, offer the pharmacological benefit of a mucosal response followed by a seric response.

Nasal dry powder delivery systems can be divided into passive and active systems. Passive systems are powder devices where the act of sniffing delivers the powder. An active system is based mostly on an air pressure driven mechanism. The powder is dispensed through a rapid airflow, which passes through the container and carries the drug into the nasal cavity. A nasal dry powder device takes advantage of the fact that one of the functions of the nose is the filtering of pollutants. A recently developed passive dry powder bi-dose system (Fig. 10) consists of two prefilled blisters, which assures optimal protection against vapor transmission, oxygen, and light. The aerodynamic feature, combined with easy actuation mode, facilitates accurate dosing in terms of expelled amount and inhalation forces. An optimal coverage of the mucosa is facilitated through a very low airflow resistance of the device. It is very convenient for the patient to sniff the powder out of the blister. As soon as the velocity of the inhalation airflow reaches 8 L/min or more, the airflow in the nasal cavity becomes turbulent (11). The turbulence in the nasal cavity

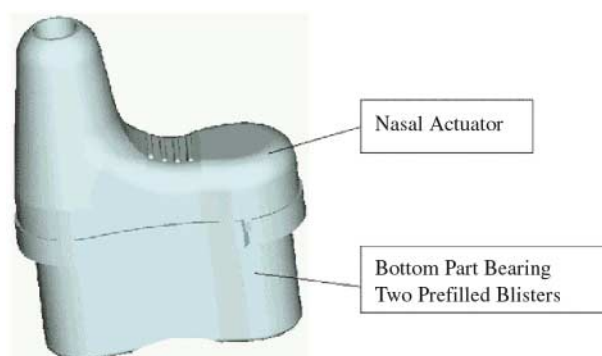


Fig. 10 Powder bi-dose system.

allows the powder to become evenly dispersed, which contributes to an optimal therapeutic effect.

The particle size of nasal powders is not regarded as critical as for inhalation powders. For inhalation, the particle size should be in the range of 1–5 μm . To allow deposition in the nasal cavity, this fine particular fraction must be avoided. The possibility of inhaling particles larger in diameter than 10 μm , or less than 5 μm , is unlikely.

Active systems are fitted with a mechanism that allows pressure to build up and eject the powder into the nostril. These devices are especially suitable for children, where it may be difficult to carry out the required inhalation process.

NASAL FORMULATIONS AND PHARMACOLOGY

During the 1990s, various nasally administered products were approved and entered the market. These included locally acting products, in particular systemically acting products (12). Most of the active ingredients are not newly discovered molecules but are reformulated products taken in the past orally or as injectables.

The benefits of nasally delivered products are their rapid onset of action under avoidance of gastrointestinal breakdown and first-pass metabolism. Furthermore, nasal drugs can be administered at a lower dose, which also means fewer side effects. Other benefits include economical reasons, such as the extensions of patents and intellectual property rights, the economic efficiency of the healthcare systems, and of course, patient compliance.

For a successful nasal product, the drug properties, the delivery system, and the nasal physiology have to act together from the early stage of the development. The development of locally acting products has to focus in

particular on a minimum of absorption into the nasal mucosa and a maximum residence time on the nasal mucosa. Systemically acting products require an efficient absorption into the bloodstream for an optimal bioavailability. It is believed that the possible transport routes of hydrophilic macromolecular compounds are paracellular, and for lipophilic compounds transcellular (13). The delivery system has to be adapted to the formulation in order to enable an optimal deposition in the nasal cavity. As the product is carried in a complex organ, researchers have to be familiar with the nasal anatomy and physiology.

Recently, a comparison study (14) of deposition pattern of aqueous nasal spray pumps and non-portable nebulizers was published. The obtained controversial results showed a relative standard deviation of 35–80%. This demonstrates the variations in nasal anatomy and physiology from individual to individual. Other studies (15) indicate that differences in the spray performance, i.e., spray angle and particle size distribution, of delivery systems do not necessarily result in different in vivo deposition.

Scintigraphic studies (16) demonstrate that the spray angle has almost no influence on the deposition pattern in the nasal cavity. Different nasal characteristics also have to be considered in the interpretation of animal studies. A comparison of interspecies characteristics was summarized by Gizurarson (17).

As mentioned previously, one of the major benefits of nasal drug delivery is the rapid onset of action (t_{\max}). Migraine treatment takes advantage of this fact. Compared to the orally taken tablet, t_{\max} is much shorter. The neurosecretory hormone melatonin, which is used widely against jetlag, shows a very impressive time profile when administered as a nasal spray (18). The peak levels of melatonin after nasal administration appear to be 50 times higher than after oral administration (19).

Lately, the nasal route is receiving attention for the management of postoperative pain. Mucosal administration requires only a 1.1–1.5 times higher dose of fentanyl than the intravenous dose (20). For this new application field, called PCINA (patient-controlled-intra-nasal-analgesia), the pharmaceutical industry demands safety precautions of the delivery device, which can be fulfilled through implementation of intelligent microelectronic features.

The nasal delivery of vaccines is a very attractive route of administration in terms of efficacy and consumer friendliness. A population-wide immunization against influenza has yet to be achieved. The pain of injections discourages many people from receiving a flu shot. The nasal route offers the advantage of a mucosal response followed by a seric response, and has proved to be a very efficient mode of administration (21).

CONCLUSIONS

Worldwide sales of pharmaceuticals in nasal form are approximately \$8 billion. The latest research in drug development for products delivered via the nasal route show very promising results. Not only is the development of “new” drugs taken into consideration, but the reformulation of products already on the market is also taken into account. The development time of a new chemical entity (NCE) is approximately 10–14 years, whereas the nasal reformulation of an existing drug is 4–5 years.

Device technology for nasal pharmaceuticals is becoming more sophisticated, which results in closer communication with regulatory authorities. For a product to be successful, the pharmaceutical industry and the device industry must begin communication at a very early stage of the product’s development. On average, launching a new drug product one year earlier will yield up to 30% more in profit.

The medical device industry has developed many ways to dispense nasal formulations. Device technology is becoming more and more important, not only with regard to pharmacological issues, such as efficacy or safety, but also in terms of responding to healthcare driven trends and marketing aspects, such as patient compliance or brand recognition. After careful evaluation of all aspects, the consumer can be provided with an appropriate nasal drug delivery system for an efficient therapy. The pharmacological issues of the nasal route for systemic drug delivery are widely discussed in the references (3, 6, 22).

REFERENCES

1. Moore, J. *The Drug Delivery Outlook to 2005*; Reuters Business Insight; Datamonitor PLC: London, 1999.
2. www.imshealth.com.
3. Chien, Y.W.; Su, K.S.E.; Chang, S.-F. *Nasal Systemic Drug Delivery*; Marcel Dekker, Inc.: New York, 1989; 1.
4. *BMJ* **1998**, 317, 739.
5. Hardy, J.G.; Lee, S.E.; Wilson, C.G. Intranasal Drug Delivery by Spray and Drops. *J. Pharm. Pharmacol.* **1985**, 37, 294–297.
6. Mygind, N. *Nasal Allergy*; Blackwell Scientific Publications: London, 1979; 260.
7. Mygind, N. *Nasal Allergy*; Blackwell Scientific Publications: London, 1979; 261.
8. Huester, R. Unpublished Results. Fresenius Institute: Taunstein, Germany, 1998.
9. Illum, L. The Nasal Delivery of Peptides and Proteins. *TIBTECH* **1991**, 9, 284–289.
10. *State of the World’s Vaccines and Immunizations*; World Health Organization: Geneva, Switzerland, 1996.

11. Fischer, R. *The Physics of the Airflow in the Nose*; Habilitation Thesis, University Berlin: Berlin, Germany, 1969.
12. Behl, C. Intranasal Drug Delivery Technology for CNS Drugs Nasal Drug Delivery Conference, London, March, 23–24, 2000.
13. McMartin, C.; Hutchinson, L.E.F.; Hyde, R.; Peters, G.E. Analysis of Structural Requirements for the Absorption of Drugs and Macromolecules from the Nasal Cavity. *J. Pharm. Sci.* **1987**, *76*, 535–540.
14. Suman, J.D.; Laube, B.L.; Dalby, R. *Nasal Nebulizers versus Aqueous Nasal Spray Pumps: A Comparison of Deposition in Human Volunteers*. Respiratory Drug Delivery VI, Hilton Head Island, SC, May, 3–7, 1998; School of Pharmacy of Virginia Commonwealth University; Interpharm Press: Buffalo Grove, IL, 1998; 211–218.
15. Suman, J.D.; Laube, B.L.; Ta-chun, L.; Brouet, G.; Dalby, R. *In Vitro Tests of Aqueous Nasal Spray Pumps May Not Predict In Vivo Deposition Pattern*. Postersession Aerosol Delivery, AAPS New Orleans, LA, Nov, 14–18, 1999.
16. Newman, S.P. Deposition Pattern of a Nasal Pump Spray. *Rhinology* **1987**, *26*, 111–120.
17. Gizurarson, S. Interspecies Comparison of Nasal Characteristics. *Acta Pharm. Nord.* **1990**, *2* (2), 108–112.
18. Merkus, F.W.H.M. *Nasal Melatonin*, US Patent 6,007,834, 1999.
19. Merkus, F.W.H.M.; Verhoef, J.C. Cyclodextrins in Nasal Drug Delivery. *Adv. Drug Delivery Rev.* **1999**, *36*, 41–57.
20. Toussaint, S.; Maidl, J.; Schwagmaier, R.; Striebel, H.W. Patient Controlled Intranasal Analgesia (PCINA) with Fentanyl for the Treatment of Postoperative Pain. *Int. Monitor* **1998**, *10*, 63.
21. Kuno-Sakei, H.; Kimura, M.; Ohta, K.; Shimojima, R.; Oh, Y.; Fukumi, H. Developments in Mucosal Influenza Vaccines. *Vaccines* **1999**, *12* (14), 1303–1310.
22. Chien, Y.W. *Transnasal Systemic Medications*; Elsevier: Amsterdam, 1985.

DRUG INTERACTIONS

Daniel A. Hussar

University of the Sciences in Philadelphia, Philadelphia, Pennsylvania, U.S.A.

DRUG INTERACTIONS

The term *drug interaction* is applied most frequently to those situations in which the actions of one drug are altered by the concurrent use of another (drug–drug interactions), and to those situations in which the actions of nutrients affect drugs or vice versa (drug–nutrient–food interactions). The concept of drug interaction has also been extended to include situations in which a drug causes alterations of laboratory test results (drug–laboratory test interactions) and in which a drug causes undesired effects in patients with certain disease states (drug–disease interactions).

CONSEQUENCES OF DRUG INTERACTIONS

One of the most important consequences of a drug interaction is an excessive response to one or more of the agents being used. For example, a significantly enhanced effect of agents such as digoxin (e.g., Lanoxin) and warfarin (e.g., Coumadin) can result in serious complications, and the hazards of using agents having central nervous system depressant properties in combination are also well known. Not as well recognized but also very important are those interactions in which drug activity is decreased, resulting in a loss of efficacy. These interactions are especially difficult to detect because they may be mistaken for therapeutic failure or worsening of the disease.

Although some drug interactions develop unexpectedly and are impossible to predict, others are related to known pharmacodynamic and pharmacokinetic properties of the drugs and can be anticipated. However, in considering the number and diversity of the pharmacological actions of each of the agents utilized in many multiple-drug regimens, it is very difficult to predict the magnitude of a specific action of any given drug. These circumstances point to a need not only for maintenance of complete and current medication records for patients, but also for closer monitoring and supervision of drug therapy so that problems can be prevented or detected at an early stage.

The concerns over drug interactions and the advocacy for appropriate precautionary measures are justifiably

based on the rapidly growing number of observations and studies in which these problems have been described. However, it is also very important to keep in perspective the potential for the occurrence of an interaction. Caution is needed in evaluating and using the available information to avoid overreacting to a potential problem. Although therapeutic alternatives might be considered and/or steps taken to more closely monitor combination therapy, patients must not be deprived of therapy from which they can benefit. The appropriate clinical perspective must be exercised if optimal therapy is to be achieved.

Sometimes a second drug is prescribed intentionally to modify the effects of the first. Such an approach might be used in an effort to enhance the effectiveness or to reduce the adverse effects of the primary agent. In these situations, the efficacy and/or safety of a drug is increased, which indicates that drug interactions are not always harmful but also can be beneficial.

FACTORS CONTRIBUTING TO THE OCCURRENCE OF DRUG INTERACTIONS

A number of factors contribute to the occurrence of drug interactions; the most important factors will be discussed.

Multiple Pharmacological Effects

Most drugs used in current therapy exhibit more than one type of pharmacological action and have the capacity to influence many physiological systems. Therefore, two concomitantly administered drugs will often affect some of the same systems. When considering the potential for interactions between drugs, there is often a tendency to focus on the primary effects of the drugs involved and to overlook their secondary actions. As an example, combined therapy with a phenothiazine antipsychotic [e.g., chlorpromazine (Thorazine)], a tricyclic antidepressant [e.g., amitriptyline (e.g., Elavil)], and an antiparkinsonian agent [e.g., trihexyphenidyl (e.g., Artane)] is employed in some patients. Each of these agents has a considerably different primary effect; however, all three possess anticholinergic activity. Even though the

anticholinergic effect of any one of the drugs may be slight, the additive effects of the three agents may be significant.

Multiple Prescribers

Some individuals go to more than one physician, and it is common for a patient to be treated by one or more specialists in addition to a family physician. Some patients are also seeing other health professionals (e.g., dentists, podiatrists, etc.), who may prescribe medication. It is frequently difficult for one prescriber to become aware of all the medications that have been prescribed by others for a particular patient, and many difficulties arise from such situations. For example, one physician may prescribe an antihistamine having sedative properties for a patient for whom another physician has prescribed an antianxiety agent, with the possible consequence of an excessive depressant effect. Even though the patient is seeing different prescribers, he will often have the prescriptions dispensed at the same pharmacy. Therefore, the pharmacist, by maintaining patient medication records, plays an important role in the detection and prevention of drug-related problems.

Concurrent Use of Nonprescription Drugs

Many reports of drug interactions involve the concurrent use of a prescription drug with a nonprescription drug (e.g., aspirin, antihistamines, antacids, etc.). When a physician questions a patient about medications being taken, the patient will often neglect to mention nonprescription medications. Many patients take preparations such as antacids, laxatives, analgesics, and vitamins for such long periods and in such a routine manner that they do not consider them to be drugs. This information may often be missed when questioning patients, and some physicians and pharmacists prefer to utilize a list of symptoms that might ordinarily be treated with nonprescription drugs when obtaining this information from patients.

Interactions may also result from the concurrent use of two or more products available without a prescription. In some situations, two nonprescription products promoted for different purposes contain the same active ingredient(s), which increases the risk of an excessive response to these agents. Diphenhydramine is included in many products for its antihistamine action but also is included for its sedative effect in many nonprescription sleep aids. Patients are often unaware that the products they purchase for different conditions may contain the same active ingredients. This puts users at increased risk for interaction

difficulties due to the use of products they assume are safe since they do not require a prescription.

The observations made with respect to potential interactions involving nonprescription products also apply to the use of herbal products, dietary supplements, and other related products that are available without a prescription. Although much is still to be learned about the properties of these products, many appear to have a potential to interact with prescription medications. Therefore, patients should be asked whether they are using such products.

Patient Noncompliance

For a variety of reasons, many patients do not take medication in the manner intended by the prescriber. Some patients have not received adequate instructions from the prescriber and pharmacist as to when and how to take their medication. In other situations, particularly those involving patients who are taking several medications, confusion about the instructions may develop even though the patient understood them initially. It is understandable that older patients, who may be taking five or six medications several times a day at different times, can become confused or forget to take their medication, although these occurrences are by no means unique to this population.

At times, patients are noncompliant when they do not take enough of their medication. In other instances, interactions can occur due to excessive dosing. For example, some patients double their dose of medication after realizing they forgot to take the initial dose. Other patients may assume that if the prescribed one tablet dose provides partial but not complete relief of symptoms, a two-tablet dose will be more effective.

MECHANISMS OF DRUG INTERACTION

Understanding the mechanisms by which drug interactions develop will be valuable in anticipating such situations and dealing with problems that do develop. Although the circumstances surrounding the development of some drug interactions are complex and poorly understood, the mechanisms by which many interactions develop are well documented and relate to the basic processes by which drugs act and are acted on in the body.

These mechanisms are often generally categorized as being either pharmacodynamic or pharmacokinetic. Included among the pharmacodynamic interactions are those in which drugs that have similar (or opposing) pharmacological effects are administered concurrently and

situations in which the sensitivity or responsiveness of the receptors/tissues to one drug is altered by the action of another. In these situations, there is a change in drug effect without a change in drug plasma concentration. Pharmacokinetic interactions are those in which one agent (designated by some as the “precipitant drug”) alters the absorption, distribution, metabolism, or excretion (ADME) of a second agent (the “object drug”), with a resultant change in the plasma concentration of the latter agent. Several mechanisms may be involved in the development of certain interactions.

Pharmacodynamic Interactions

Although pharmacokinetic interactions have been studied and publicized to a greater extent, pharmacodynamic interactions are very common and also present challenging clinical problems. Some of the causes of pharmacodynamic interactions will be discussed.

Drugs having opposing pharmacological actions

Interactions resulting from the use of two drugs with opposing effects should be among the easiest to detect. However, opposing pharmacological actions are sometimes caused by the secondary effects of certain drugs, and this and other factors may preclude early identification of such situations.

Diuretics: Thiazides and certain other diuretics may elevate blood glucose concentrations. When the diuretic is prescribed for a diabetic patient being treated with insulin or one of the oral antidiabetic agents, this effect may partially counteract the glucose-lowering action of the antidiabetic drug, necessitating an adjustment in dosage. Similarly, many diuretics may produce a hyperuricemic effect. Therefore, therapy in patients with gout should be closely monitored, as the hyperuricemic action of a diuretic may necessitate an adjustment in dosage of the agent being used in the treatment of gout.

Drugs having similar pharmacological actions

The most common type of pharmacodynamic interaction is an excessive response attributable to the concurrent use of drugs having similar actions. These potential problems warrant particular attention.

Central nervous system (CNS) depressants: An excessive CNS depressant effect, resulting from the concurrent use of two or more drugs exhibiting a depressant action, represents one of the most dangerous drug related problems. Older patients are especially susceptible to this type of response, and patients experiencing effects such as fatigue and dizziness are at increased risk of falls and injuries, such as hip fractures. Patients also must be advised

of the risks of operating motor vehicles or machinery. In considering multiple drug regimens, the prescriber needs to recognize the large number of agents (e.g., antianxiety agents, hypnotics, antipsychotics, tricyclic antidepressants, certain analgesics, and most antihistamines) that can exhibit a depressant effect. Consideration should be given to whether it is necessary to use all the drugs concurrently, and the dosages of the drugs having a depressant effect should be appropriately reduced.

Alcohol and other CNS depressants: The increased CNS depressant effect experienced by most individuals being treated with depressant drugs when they consume alcoholic beverages is among the best known interactions. However, this interaction also illustrates the difficulties in predicting the magnitude of the response experienced by a particular patient, as the response will depend on many variables, including the patient's tolerance of alcohol. Every patient should be alerted to the fact that the depressant effect of the drug prescribed may be enhanced by alcohol. Patients who will not completely avoid alcoholic beverages should be urged to use them only in moderation, particularly when therapy is initiated, and cautioned to observe their own tolerance when such combinations are employed. However, it should be remembered that even though many individuals can take depressant drugs and consume relatively large amounts of alcoholic beverages with no apparent difficulty, this combination can be lethal to some and cause injury to others. Thus, all patients prescribed these drugs need to be warned of the potential interactions.

Drugs having anticholinergic activity: As noted earlier, drugs that differ considerably in their primary pharmacological actions may exhibit the same secondary effects. Some patients treated with antipsychotic agents, such as chlorpromazine, are also given antiparkinsonian agents, such as trihexyphenidyl, to control the extrapyramidal effects of the former. In addition, a number of patients experience depressive symptoms, and a tricyclic antidepressant, such as amitriptyline, might be added to the therapy. These three agents all possess anticholinergic activity, and the additive effect could result in side effects such as dryness of the mouth, blurred vision, urinary retention, constipation, and elevation of intraocular pressure.

While some health professionals might consider side effects such as dryness of the mouth to be minor, these side effects may be especially troublesome in certain patients. For example, persistent dryness of the mouth could make the use of dentures more difficult and cause other dental complications. Increased difficulty in chewing and swallowing may contribute to the problem of malnutrition in some older individuals.

Dryness of the mouth may also result in other problems. For example, the tricyclic antidepressant imipramine (e.g., Tofranil) can cause persistent dryness of the mouth. If nitroglycerin tablets are administered sublingually for the management of exertional angina, the relief of the symptoms may be delayed due to the slower dissolution of the sublingual tablets.

An excessive anticholinergic effect can cause an atropine-like delirium, particularly in older patients. This effect could be misinterpreted as an increase in psychiatric symptoms that might be treated by increasing the dosage of the therapeutic agents that are responsible for the problem. This example points out the difficulty that often exists in distinguishing between the symptoms of the condition(s) being treated and the effects of the drug(s) being employed as therapy. Other potential problems associated with the use of drugs having anticholinergic activity include memory impairment and impairment in self-care capacity.

There have been reports of patients who take a phenothiazine and antiparkinsonian agent concurrently developing severe hyperpyrexia after being exposed to high environmental temperatures and humidity. These combinations may interfere with the thermoregulatory system of the body, and physicians treating patients in hot and humid climates should minimize the outdoor exposure of patients receiving high doses of these agents.

Drugs exhibiting hypotensive effects: Certain antihypertensive drugs, as well as some other classes of medications (e.g., tricyclic antidepressants), can cause orthostatic hypotension, which results in dizziness, lightheadedness, and in more severe cases, syncope. Older patients are more susceptible to this type of response and the associated risks, such as falls and injuries. Appropriate precautions should be exercised whether these agents are given alone or in combination.

Nonsteroidal antiinflammatory drugs (NSAIDs): In some cases, a patient may unknowingly be taking several different products that contain the same NSAID. An arthritic patient whose condition has been managed with ibuprofen obtained via prescription (often at dosage levels at or near the recommended maximum) may purchase a nonprescription ibuprofen product for pain/discomfort not associated with the arthritis. The patient may not know that the two products contain the same drug and that there is an increased risk of adverse effects.

Alteration of electrolyte concentrations

Several important drug interactions occur as a result of therapeutic agents altering the concentrations of electrolytes, such as potassium and sodium. When these drugs are included in a therapeutic regimen, it is important that electrolyte concentrations be periodically monitored.

Digoxin and diuretics: One of the problems associated with the use of most of the commonly employed diuretics [e.g., the thiazides furosemide (e.g., Lasix)], is excessive loss of potassium. Particular caution is necessary in patients also being treated with digoxin, many of whom are also candidates for diuretic therapy. If potassium depletion remains uncorrected, the heart may become more sensitive to the effects of digoxin and arrhythmia may result.

Although potassium supplementation is necessary in some individuals being treated with a potassium-depleting diuretic, the initiation of therapy with such a diuretic must not be viewed as a mandate to provide potassium supplementation. This decision should be based on a consideration of the individual patient's situation and the appropriate parameters should be periodically monitored. It must be recognized that dangers exist if hyperkalemia occurs as a result of excessive supplementation. Although the kidneys are usually able to excrete excessive amounts of potassium rapidly, hyperkalemia may develop, especially in patients with diminished renal function.

In addition to the diuretics, other agents can cause potassium depletion. Prolonged therapy with cathartics and corticosteroids can cause potassium depletion, although this is not likely to occur as quickly or to the same extent as with diuretics.

Lithium and diuretics: Sodium depletion is known to increase lithium toxicity, and it is generally recommended that lithium not be used in patients on diuretic therapy or on a sodium-restricted diet. Even protracted sweating or diarrhea can cause sufficient depletion of sodium to result in decreased lithium tolerance. The sodium depletion caused by diuretics reduces the renal clearance and increases the activity of lithium. However, if preferable therapeutic alternatives are not available, concurrent therapy need not be contraindicated as long as the interaction is recognized and appropriate steps are taken to monitor therapy and adjust the dosage.

Interactions at receptor sites

Examples of interactions occurring at receptor sites include problems involving the use of the monoamine oxidase (MAO) inhibitors [isocarboxazid (Marplan), phenelzine (Nardil), tranylcypromine (Parnate), and procarbazine (Matulane)].

MAO inhibitors and sympathomimetic agents: MAO breaks down catecholamines, such as norepinephrine. When this enzyme is inhibited, the concentrations of norepinephrine within adrenergic neurons increase and drugs that stimulate its release can bring

about an exaggerated response. Interactions between MAO inhibitors and indirectly acting sympathomimetic amines (e.g., amphetamine) develop by this mechanism. If amphetamine is administered to a patient whose stores of norepinephrine have been increased by MAO inhibition, the patient may experience severe headache, hypertension (possibly a hypertensive crisis), and cardiac arrhythmias. The serious consequences associated with these interactions contraindicate the use of these agents in combination.

Most sympathomimetic amines, such as amphetamine, are available only by prescription; others such as phenylephrine and pseudoephedrine, which also are reported to interact similarly with MAO inhibitors, are found in most popular nonprescription cold and allergy preparations. It is important that patients being treated with MAO inhibitors avoid using products containing these agents.

MAO inhibitors and other antidepressants: Product literature and case reports caution against concurrent use of MAO inhibitors with tricyclic antidepressants (e.g., amitriptyline, imipramine) because severe atropine-like reactions, tremors, convulsions, hypothermia, and vascular collapse can occur. The labeling for most of these products warns that therapy with a MAO inhibitor or a tricyclic antidepressant should not be initiated until at least 7–14 days after therapy with the other drug has been discontinued.

Although the labeling for most MAO inhibitors and tricyclic antidepressants warns that concurrent use is contraindicated, there is disagreement as to the degree of risk involved. Several studies show little evidence of interaction. The impression that serious interactions are uncommon, coupled with reports of favorable results with such combinations in selected patients who did not respond to either agent given alone, have led some to conclude that these combinations can be cautiously employed. In patients who are refractory to single antidepressants and who are not candidates for other therapeutic approaches, the potential benefits of combination therapy may outweigh the risks. However, such therapy should be undertaken only by those who are thoroughly familiar with the risks involved and under circumstances in which therapy can be closely monitored.

Very little data regarding the combined use of fluoxetine (Prozac) and a MAO inhibitor are available. It is advised that their combined use be avoided and that at least 14 days elapse between discontinuation of a MAO inhibitor and initiation of treatment with fluoxetine. There are reports of deaths in patients in whom therapy with a MAO inhibitor was initiated shortly after discontinuation

of fluoxetine. Due to the long half-lives of fluoxetine and its active metabolite, it is recommended that at least 5 weeks elapse between discontinuation of fluoxetine and initiation of therapy with a MAO inhibitor.

Pharmacokinetic Interactions

Alteration of gastrointestinal absorption

The interactions that involve a change in the absorption of a drug from the gastrointestinal (GI) tract may develop through different mechanisms and be of varying clinical importance. In some situations, the overall absorption of the drug might be reduced and its therapeutic activity compromised. In other circumstances, absorption may be delayed but the same amount of drug is eventually absorbed. A delay in drug absorption is undesirable when a rapid effect is needed to relieve acute symptoms such as pain. However, in other situations, a delay in drug absorption will not be clinically significant, and this is usually the case when a drug is being used on a chronic basis and therapeutic concentrations in the body have already been achieved. As a general guideline, those drugs not completely absorbed under normal circumstances are most susceptible to alterations of GI absorption.

Ketoconazole and antacids: An acidic medium is required to achieve adequate dissolution of ketoconazole (e.g., Nizoral) following oral administration. Therefore, an antacid, histamine H₂-receptor antagonist [e.g., cimetidine (e.g., Tagamet), ranitidine (e.g., Zantac)], or a proton pump inhibitor [e.g., lansoprazole (Prevacid), omeprazole (Prilosec)] is likely to reduce the dissolution, absorption, and effectiveness of the antifungal agent. An antacid should be administered at least 2 h after ketoconazole. The concurrent use of ketoconazole and a histamine H₂-receptor antagonist or proton pump inhibitor is best avoided, and other agents having a lesser potential for interaction should be considered.

Tetracyclines and metals: Tetracyclines can combine with metal ions, such as calcium, magnesium, aluminum, and iron, in the GI tract to form complexes that are poorly absorbed. Thus, the simultaneous administration of certain drugs (e.g., antacids, iron preparations, products containing calcium salts) by patients on tetracycline therapy could result in a significant decrease in the amount of antibiotic absorbed. When two drugs are recognized as having a potential to interact, there is sometimes a tendency to believe that one of them should be discontinued. In the case of the tetracycline–antacid interactions, problems can be avoided by allowing an interval of at least 1 h to separate the administration of the two drugs.

Fluoroquinolones and metals: Aluminum- and magnesium-containing antacids, as well as other products that contain metals (e.g., iron), markedly reduce the absorption and serum concentrations of the fluoroquinolone derivatives [e.g., ciprofloxacin (Cipro)], probably as a result of the metal ions complexing with the antiinfective agent. Antacids or other metal-containing products must not be administered at the same time as a fluoroquinolone, and the labeling for most of the fluoroquinolones designates a minimum interval of time that should separate the administration of the two drugs. For example, it is recommended that moxifloxacin (Avelox) be taken at least 4 h before or 8 h after taking antacids or other metal-containing products.

Cholestyramine and colestipol: Other interactions involving complexation might be anticipated when cholestyramine (e.g., Questran) and colestipol (Colestid) are used. These resinous materials, which are not absorbed from the GI tract, bind with bile acids and prevent their reabsorption. In addition, cholestyramine and colestipol can bind with drugs (e.g., digoxin and warfarin) that are present in the GI tract. To minimize the possibility of an interaction, the interval between the administration of cholestyramine or colestipol and another drug should be as long as possible.

An interesting application of this interaction is seen with the use of leflunomide (Arava) in the treatment of rheumatoid arthritis. Leflunomide can cause fetal harm if administered during pregnancy, and it has an active metabolite that can persist in the system for at least 2 years. If a woman of childbearing potential discontinues use of leflunomide, it is recommended that cholestyramine (8 g 3 times a day for 11 days) be used to accelerate the elimination of the drug and its active metabolite.

Food: Food can influence the absorption and activity of a number of drugs. In some situations, absorption is delayed but not reduced, whereas in other circumstances, the total amount of drug absorbed is reduced. The effect of food in influencing drug absorption is often due to slower gastric emptying. However, food may also affect absorption by binding with drugs, by decreasing the dissolution rate of solid dosage forms, or by altering the pH of the GI contents.

The presence of food in the GI tract reduces the absorption of many antiinfective agents. Although there are some exceptions (e.g., penicillin V, amoxicillin, and doxycycline), it is generally recommended that penicillin and tetracycline derivatives, as well as certain other antiinfective agents, be given at least 1 h before meals or 2 h after meals to achieve optimum absorption.

Alendronate and food: Food and some beverages (e.g., orange juice, coffee, and mineral water) may markedly reduce the bioavailability of alendronate (Fosamax). Therefore, the drug should be administered at least 30 min before the first food, beverage, or medication of the day, with plain water only.

Acarbose or miglitol and food: Some medications should be administered with food for optimum benefit. Acarbose (Precose) and miglitol (Glyset) are effective in the treatment of diabetes mellitus because they delay the digestion of ingested carbohydrates and reduce the elevation of blood glucose concentrations following meals. Maximum effectiveness is attained when doses are administered at the start (with the first bite) of each main meal.

MAO inhibitors and tyramine: There have been reports of serious reactions (e.g., hypertensive crises) occurring in patients being treated with a MAO inhibitor following ingestion of foods with a high content of pressor substances, such as tyramine.

Tyramine is metabolized by MAO, and normally these enzymes in the intestinal wall and in the liver protect against the pressor actions of amines in foods. However, when these enzymes are inhibited, large quantities of unmetabolized tyramine can accumulate and act to release norepinephrine from adrenergic neurons where greater-than-usual stores of this catecholamine are concentrated as a result of MAO inhibition. Among the foods having the highest tyramine content are aged cheeses (such as cheddar), certain alcoholic beverages (e.g., Chianti), pickled fish (e.g., herring), concentrated yeast extracts, and broad-bean pods (also known as fava beans or Italian green beans).

The pharmaceutical companies that market the MAO inhibitors have developed lists of dietary items to avoid when taking one of these agents. This information should be provided to and discussed with each patient who is prescribed a MAO inhibitor.

Grapefruit juice: Grapefruit juice is reported to increase the serum concentrations and activity of a number of medications, such as certain calcium channel blocking agents [e.g., felodipine (Plendil) and nisoldipine (Sular)], certain HMG-CoA reductase inhibitors [e.g., lovastatin (Mevacor)], and cyclosporine (e.g., Neoral). The bioavailability of most of these agents is generally low, primarily as a result of extensive first-pass metabolism. Components of grapefruit juice possibly reduce the activity of the cytochrome P450 enzymes (primarily CYP3A4) in the gut wall, which are involved in the metabolism of these agents. As a result, larger amounts of unmetabolized drug are absorbed and serum concentrations are increased.

Alteration of distribution

Interactions involving an alteration of distribution may occur when two drugs that are capable of binding to proteins are administered concurrently and one agent displaces the other. Most significant are the situations in which two drugs are capable of binding to the same sites on the protein (competitive displacement). Since protein-binding sites are limited in number, the drug that has the greater affinity for the binding sites will displace the other from plasma or tissue proteins. The protein-bound fraction of a drug in the body is not pharmacologically active. However, an equilibrium exists between bound and unbound fractions, and as the unbound or “free” form of the drug is metabolized and excreted, the bound drug is gradually released to maintain the equilibrium and pharmacological response.

The risk of an interaction occurring is greatest with those drugs that are highly protein bound (more than 90%) and that also have a small apparent volume of distribution. Since only a small fraction of the drug would ordinarily be available in the free form, the displacement of even a small percentage of the amount that is bound to proteins could produce a considerable increase in activity.

Methotrexate: Methotrexate is highly bound to plasma proteins, and agents such as the salicylates may be capable of displacing it from binding sites. Salicylates may also increase the action of methotrexate by inhibiting its renal excretion. The potential for toxicity with methotrexate dictates caution in all situations in which it is used.

Stimulation of metabolism

Drug metabolism occurs primarily in the liver and most commonly involves oxidation, reduction, hydrolysis, and conjugation reactions. Quantitatively, the most important hepatic enzymes are the cytochrome P450 enzymes that have been divided into families and subfamilies (e.g., CYP3A4) based on the similarity of their amino acid sequences. These enzymes are responsible for the metabolism of a large number of drugs.

Many drug interactions result from an effect frequently referred to as enzyme induction: the ability of one drug to stimulate the metabolism of another, most often by increasing the activity of liver enzymes. Enzyme induction usually results in increased metabolism and excretion, and reduces the effect of the agent that is metabolized by the hepatic enzymes. Phenobarbital (and other barbiturates), phenytoin, carbamazepine (e.g., Tegretol), and rifampin (e.g., Rifadin) are among the agents best recognized as causing enzyme induction.

Warfarin and phenobarbital: Phenobarbital, by causing enzyme induction, can increase the rate of metabolism of warfarin. The result of this interaction is a decreased response to the anticoagulant and an increased risk of thrombus formation if the interaction is not recognized.

Smoking: The polycyclic hydrocarbons in cigarette smoke may increase the activity of oxidative enzymes, with the result that certain therapeutic agents [e.g., diazepam (e.g., Valium), propoxyphene (Darvon), theophylline, chlorpromazine, and amitriptyline] are metabolized more rapidly and their effect is decreased. In addition to monitoring therapy carefully with drugs that are metabolized by hepatic enzyme systems in patients who are moderate or heavy smokers, caution must also be exercised if a patient treated with such a medication discontinues smoking. For example, if therapy with a tricyclic antidepressant is initiated in a patient who is a heavy smoker, the maintenance dosage will be determined during the time period in which the enzyme-inducing action of smoking is decreasing the effect of the medication. If the patient stops smoking and is still taking the medication, the dosage that had been appropriate is now likely to be excessive and will have to be reduced.

Inhibition of metabolism

There are numerous situations in which one drug inhibits the metabolism of a second agent, usually resulting in a prolonged and intensified activity of the latter.

Mercaptopurine or azathioprine and allopurinol: Allopurinol (e.g., Zyloprim), by inhibiting the enzyme xanthine oxidase, reduces the production of uric acid, which is the basis for its use in the treatment of gout. Xanthine oxidase also has an important role in the metabolism of such potentially toxic drugs as mercaptopurine (e.g., Purinethol) and azathioprine (e.g., Imuran). When xanthine oxidase is inhibited by allopurinol, the effect of these agents can be markedly increased. When allopurinol is given in doses of 300–600 mg/day concurrently with either of these drugs, it is advised that the dose of mercaptopurine or azathioprine be reduced to about one-third to one-fourth the usual dose.

Cimetidine: Since cimetidine may inhibit certain metabolic pathways, an increased action of concurrently administered drugs that are metabolized via these pathways should be anticipated. For example, cimetidine may inhibit the metabolism of diazepam and certain other benzodiazepines, and the sedative effect of these agents may be enhanced as a result of the interaction. Particular caution is necessary in older patients who may exhibit an increased sensitivity to the depressant effects of the

benzodiazepines, even when one of these agents is given alone.

The metabolism of lorazepam (e.g., Ativan), oxazepam (e.g., Serax), and temazepam (e.g., Restoril) are not likely to be affected, and one of these agents may be preferred when a benzodiazepine is indicated in a patient being treated with cimetidine. The experience with ranitidine (e.g., Zantac), famotidine (Pepcid), and nizatidine (Axid) suggests that these agents are not likely to inhibit hepatic enzyme systems, and these other histamine H₂-receptor antagonists are less likely than cimetidine to interact with other drugs that are metabolized via these pathways.

Macrolide antibiotics: Erythromycin may significantly increase serum concentrations of medications such as theophylline by inhibiting their hepatic metabolism. Clarithromycin (Biaxin) and troleandomycin appear to interact with other medications in a manner similar to erythromycin, whereas azithromycin (Zithromax) is unlikely to interact with these agents.

Fluoroquinolones: Ciprofloxacin and enoxacin (Penetrex) may markedly increase serum concentrations of medications such as theophylline by inhibiting their hepatic metabolism. Certain other fluoroquinolones, such as levofloxacin (Levaquin), are not likely to inhibit hepatic enzyme systems and interact with these medications.

Alteration of excretion

Many drugs and their metabolites are excreted via the kidneys. The most important clinical implications of altering renal excretion involve the use of drugs that are excreted in their unchanged form or in the form of an active metabolite. Thus, substances with pharmacological activity are being reabsorbed or excreted to a greater extent when renal excretion is altered.

Salicylates: A change in urinary pH will influence the ionization of weak acids such as salicylates as well as weak bases (e.g., amphetamine), thereby affecting the extent to which these agents are reabsorbed and excreted. When a drug is in its nonionized form, it will more readily diffuse from the urine back into the blood. Therefore, an acidic drug will have a larger proportion of the nonionized drug in an acid urine than in an alkaline urine, where it will primarily exist as an ionized salt. The result is that acidification of the urine may result in increased salicylate concentrations and a prolonged and perhaps intensified drug action. The risk of a significant interaction is greatest in patients who take large doses of salicylates (e.g., for arthritis).

Penicillins and probenecid: A number of organic acids are actively transported from the blood into the tubular urine and vice versa. In some situations, these agents interfere with each other's excretion. Probenecid

(e.g., Benemid) can increase the serum concentrations and increase and prolong the activity of penicillin derivatives by blocking their tubular secretion. This is an interaction that has been used to therapeutic advantage in the treatment of certain infections.

Methotrexate and NSAIDs: NSAIDs are reported to increase the activity and toxicity of methotrexate, presumably by inhibiting its active renal tubular secretion. Other mechanisms probably also contribute to an increase in serum methotrexate concentrations. Most of the patients in whom these interactions were reported were receiving high-dose methotrexate therapy for neoplastic disorders. However, caution should also be exercised in patients receiving lower doses.

Digoxin and quinidine: Studies have noted significantly greater serum digoxin concentrations when quinidine was administered concurrently than when digoxin was given alone. Digoxin is actively secreted in the renal tubules, and a primary cause of the quinidine-induced increase in serum digoxin concentrations appears to be a reduction in the renal clearance of digoxin. However, quinidine may also reduce the nonrenal clearance of digoxin.

Although other mechanisms may be involved in the development of drug interactions, the ones cited are the most important. More than one mechanism may be responsible for certain interactions and these mechanisms may work in concert or in opposition to determine the resulting effect. Still other drug interactions develop by mechanisms yet to be identified.

Significant limitations often exist in trying to predict the results of combination therapy. In the following discussion, guidelines are provided to reduce the risk of the occurrence of drug interactions.

REDUCING THE RISK OF DRUG INTERACTIONS

Reducing the risk of drug interactions is a challenge that embraces a number of considerations. The following are guidelines to reduce and manage drug interactions.

1. Identify patient risk factors: Factors such as age, the nature of the patient's medical problem (e.g., impaired renal function), dietary habits, smoking, and problems such as alcoholism influence the effect of certain drugs and should be considered during the initial patient interview.
2. Take a thorough drug history and maintain complete patient medication records: An accurate and complete record of both the prescription and nonprescription medications a patient is taking must be obtained prior

- to changing the therapeutic regimen. Numerous interactions have resulted from a lack of awareness of medications prescribed by another physician, or of nonprescription medications the patient did not consider important enough to mention. By maintaining patient medication records, the pharmacist is in an excellent position to detect potential problems and initiate the necessary steps to minimize or avoid them.
3. Be knowledgeable about the actions of the drugs being utilized: The knowledge of the properties and the primary and secondary pharmacological actions of each of the agents used, or being considered for use, is essential if the interaction potential is to be accurately assessed.
 4. Consider therapeutic alternatives: In most cases, two drugs that are known to interact can be administered concurrently as long as adequate precautions are taken (e.g., closer monitoring of therapy or dosage adjustments to compensate for the altered response). However, in those situations in which another agent with similar therapeutic properties and a lesser risk of interaction is available, the other agent should be used.
 5. Avoid complex therapeutic regimens where possible: The number of medications used should be kept to a minimum. In addition, the use of medications or dosage regimens that permit less frequent administration may help avoid interactions that result from an alteration of absorption (e.g., when a drug is administered in close proximity to meals).
 6. Educate the patient: Patients often know little about their illness, let alone the benefits and problems that could result from drug therapy. Individuals who understand this information are more likely to comply with instructions for administering medications and are more attentive to the development of symptoms that could be early indicators of drug-related problems. Patients should be encouraged to ask questions about their therapy and to report any excessive or unexpected responses. There should be no uncertainty on the part of the patient as to how to use medications in the safest, most effective way.
 7. Monitor therapy: The risk of drug-related problems warrants close monitoring, not only for the possible occurrence of drug interactions but also for adverse effects occurring with individual agents and noncompliance. Any change in patient behavior should be suspected as drug-related until that possibility is excluded.
 8. Individualize therapy: Although the development of a therapeutic regimen that meets the specific needs of individual patients is inherent in many of the above guidelines, the importance of this consideration cannot be too strongly emphasized. Wide variations in the response of patients to the same dose of certain drugs are well recognized. It is difficult to predict the response to many therapeutic agents when they are given alone; the challenge and limitations in anticipating the response to a multiple drug regimen are even greater. Therefore, priority should be assigned to the needs and clinical response of the individual patient, rather than to the usual dosage recommendations, standard treatment, and monitoring guidelines.
 9. Involve the patient as a partner in health care. The best efforts of the health professionals involved in the patient's care will fall short of the desired goals unless the patient and/or the family participates in, understands, and complies with decisions regarding the therapeutic regimen. If the optimal benefits of therapy are to be achieved with minimal risk, each participant must be knowledgeable about and diligent in fulfilling his responsibilities.
- The following publications are recommended as comprehensive references in which detailed information regarding specific interactions is provided:

BIBLIOGRAPHY

- Hansten, P.D.; Horn, J.R. *Drug Interactions Analysis and Management*; Facts and Comparisons: St. Louis, 2001.
- Tatro, D.S. *Drug Interaction Facts*; Facts and Comparisons: St. Louis, 2001.



Drug Information Systems

Linda Lisgarten

Michelle Wake

University of London, London, U.K.

INTRODUCTION

All those with an interest in the pharmaceutical sciences need up-to-date and reliable drug information, not only for their own research and studies, but also for communicating with their healthcare colleagues and patients. In this chapter, we will discuss the wide range of modern electronic drug information systems that are currently available, and will give an indication of their main uses and individual strengths.

Since the 1960s, there has been a dramatic increase in the amount of drug information. The number of chemical compounds alone is growing at a tremendous rate. The CAS Registry database adds around 4000 new substances every day. In tandem with this huge growth of information, there have been many developments in data storage and communication technology, which have impacted on the type of medium used to contain and deliver drug information. Along the way we have used, and are still using, paper, microfilm/microfiche, magnetic tape, floppy disks, CD-ROMs, DVDs, and of course Internet servers. There has been a veritable explosion of drug information on the World Wide Web during the past few years.

Also, there are other features and peculiarities about the nature of drug information (apart from the sheer quantity) which can make it difficult to find what you need. Drugs have several names, and those names can differ from country to country: for example, the substance known as acetaminophen in the United States is known as paracetamol in the UK. Then, there is the problem of licensing arrangements which also differ from country to country: a well-known recent example of this being Viagra, which was available in the United States several months before it officially became available in the UK. (For help with nomenclature try the CAS Registry File, the Merck Index, Martindale, or Pharmaprojects. The last resource is very useful for finding out brand names and gives information on when a product was licensed in different countries.)

BACKGROUND

The sources of drug information are extremely diverse and include journals, books, conference proceedings, patents, and trade literature. These sources are generally classified

as primary, secondary, or reference compendia/tertiary information. "Primary" refers to original and new information and "secondary" refers to reviews and commentaries of the primary information, whereas "reference compendia" or "tertiary" usually refers to compilations such as dictionaries and pharmacopoeias. When conducting an information literature search, it is wisest to begin with the reference/tertiary sources as necessary and then move on to the secondary and finally the primary sources of drug information.

REFERENCE COMPENDIA

Pharmacopoeias, Compendia, and Formularies

Many queries can be answered quickly by using a pharmacopoeia, a drug compendium or a formulary. Many of these are now available electronically:

Pharmacopoeias

- United States Pharmacopeia/National Formulary (<http://www.usp.org>)
- British Pharmacopoeia (<http://www.pharmacopoeia.org.uk>)
- The European Pharmacopoeia (<http://www.phEur.org>)

Drug compendia

- Merck Index (<http://www.merck.com/pubs>)
- Physicians' Desk Reference (<http://pdrbookstore.com>)
- Martindale: the extra pharmacopoeia (<http://www.pharmapress.com>)
- DRUGDEX (<http://www.micromedex.com/products/drugdex>)
- Medicines Compendia (<http://emc.vhn.net>)
- Pharmaprojects (<http://www.pharmaprojects.co.uk>)

Formularies

- USP DI, Volume III, Approved Drug Products and Legal Requirements (<http://www.usp.org>)
- BNF (<http://www.bnf.org>)



SECONDARY SOURCES: INDEXING AND ABSTRACTING SERVICES

Journals are usually the most up-to-date form of primary and secondary scientific literature and the number of new biomedical journals has grown rapidly. Computerized abstracting and information databases are often the first port of call to gain access to the published journal literature on a particular drug topic. There has been a plethora of new drug information system products in recent years but perhaps the most useful, with an international outlook, are MEDLINE, EMBASE, IPA, CAPLUS, and BIOSIS Previews. These databases are traditionally classed as indexing and abstracting resources and, traditionally enough, started life as paper-based services, although it should be noted that one of the great advantages of their electronic versions is that it is possible with some of them to link directly to the full text of a journal article on screen.

MEDLINE

Produced by the National Library of Medicine in the United States, MEDLINE is the computerized form of Index Medicus, the Index to Dental Literature, and the International Nursing Index dating back to 1966. Over 4600 biomedical journals are indexed from more than 70 countries, although 52% of its current records are from U.S. journals and 75% are derived from English language publications. Updated weekly, 67% of the entries contain abstracts.

When searching for a drug in MEDLINE, it is best to use a generic name. The U.S. Adopted Names are preferred, as the brand name can only be searched for if it is in the title or abstract of the original article. However, since June 1980, the NLM has added other searchable fields to ease drug searching. It is now possible to search on CAS registry numbers, Enzyme Commission Numbers, and the substances of those numbers.

EMBASE

EMBASE is the electronic form of *Excerpta Medica*, from the publisher Elsevier Science. It contains indexed articles from more than 4000 biomedical journals from 119 countries, although again the English language dominates. EMBASE is stronger on European and Far Eastern literature than MEDLINE. Books were also included from 1975 to 1980. The database goes back to 1974 and is updated weekly.

The generic name, the standard being the WHO International Nonproprietary Name, is the first choice of the EMBASE indexers and for "combination preparations" the brand name is preferred. However, EMBASE

allows drugs to be indexed under numerous names, more so than MEDLINE, and even includes the name of the manufacturer if it is in the original article. EMBASE concentrates more on the pharmacological and toxicological aspects of the biomedical literature than MEDLINE.

IPA

International Pharmaceutical Abstracts, available electronically from 1970, from the American Society of Health-System Pharmacists, may only index around 800 journal titles but, as its name suggests, it specializes in pharmacy literature from around the world. A study by Bonnie Snow found that 12% of IPA's journals were not covered by MEDLINE, EMBASE, BIOSIS Previews, or Chemical Abstracts.^[1] IPA has also indexed ASHP meetings since 1988, meetings of the American Association of College Pharmacists and the American Pharmacy Association. All records have abstracts.

The preferred index name for a drug is its U.S. Adopted Name. If lab codes and chemical names appear in the same article, lab codes are preferred. When looking for an enzyme, try the CAS Registry Number before the EC Number.

For the literature on dosage forms, drug delivery, pharmacy practice, pharmacy law, cosmetics, pharmaceutical technology, and herbal medicines, IPA is one of the premier resources.

CAPLUS File

CAS, a division of the American Chemical Society, produces a large number of databases which cover a range of sciences; not only for chemistry, but also for physics, biology, and the applied sciences. CAS's CAPLUS, developed from Chemical Abstracts, is one of the largest indexing and abstracting databases for chemistry, going back to 1907 and covering over 8000 journals as well as conference proceedings, dissertations, books, and patents from 33 national patent offices. Patents make up 30% of the database. Electronic journals and Web preprints are also included. Updated daily, the literature has worldwide coverage, taking material from 130 countries and from 47 languages.

CAPLUS is particularly useful to do research for the first stages of drug development. However, it is not so useful for clinical practice. Inevitably, the best way to search for a substance is to use its CAS registry number. It is even possible to draw a chemical structure and conduct a search on it for post-1966 data. Generic and brand names are indexed for searching only if they are provided by the author in the original source.



BIOSIS Previews

From the U.S. BioSciences Information Service, BIOSIS indexes life sciences material from around 5500 journals, theses, conference proceedings, books, and organizational reports. Some U.S. patents were included from 1986 to 1989. The literature is from over 100 countries and from 57 languages. Once again, the English language dominates with 86% of the references being derived from English language sources. BIOSIS goes back to 1969 and is updated weekly. Author abstracts have been given for 55% of references since 1976.

The life sciences remit of BIOSIS means that it covers a wider range of scientific subjects than MEDLINE or EMBASE and has a focus on research rather than clinical practice. As a result, it is good for finding literature on early drug development.

When searching for a drug the U.S. Adopted Name is the first choice, but since 1998 it is also possible to search using CAS registry numbers. Note that it is only possible to search on the fields of brand names, Enzyme Commission Numbers and lab codes if they are given in the original literature source. The indexers at BioSciences Information Service will not add them to the database.

Other pharmaceutical and indexing services which have less of an international remit, but nevertheless have an important place in the drug information armoury include Pharm-line and RPS e-PIC.

Pharm-line

Around 100 English-language pharmacy and medical journals are indexed by pharmacists in UK NHS hospitals. The project is based at the Guy's and St. Thomas's Hospital Trust in London and involves the UK Drug Information Pharmacists' Group. The database, which goes back to 1978 and is updated weekly, is good for pharmacy practice and clinical pharmacy.

RPS e-PIC

Produced by the Royal Pharmaceutical Society of Great Britain e-PIC (electronic Pharmacy Information Coverage) consists of five databases which contain references selected from the major UK pharmacy practice and medical journals since 1992. The databases are e-PIC, which looks at pharmacy practice, New Products, Discontinued Products, Pharmaceuticals-Ceuted which covers pharmaceuticals, formulation and technology and Pharmacy History.

Criteria/Issues for Selecting an Indexing and Abstracting Service

Searching the secondary literature can be an expensive and time-consuming business, and the intensive marketing of some of the abstracting databases can make objective choice difficult. We have suggested further criteria that could be useful when choosing a database:

1. Subjects covered and the use of descriptors.
2. The fields that are searchable. What information is indexed? For example, can you specify a brand name search?
3. The type of material covered.
4. Geographical coverage.
5. The criteria for items being included.
6. Designated audience.
7. The standing of the service created—is it from a respected institution?
8. Size of the database—the total number of records and the number of journals considered.
9. The years covered.
10. The currency of the service—how often is it updated and what is the time delay between an item being published and it appearing in the database?
11. The inclusion of abstracts.
12. The user friendliness of the interface and the quality of the "Help" screens.
13. The ease of downloading and/or manipulating results.
14. Cost—not just for the database itself but for also for equipment and online costs, if applicable.

It is worth noting that not all of these criteria are linked to the database itself; they may relate to the supplier of the software operating the database. For example, MEDLINE can be purchased from many software vendors, including Ovid and Compact Cambridge. All vendors will provide you with a different user interface and will have varying prices. They will also vary in the timeliness of the latest material on the database, depending on when they update their products from the National Library of Medicine's MEDLINE source data. The NLM also provides its data for free via PubMed (<http://www.ncbi.nlm.nih.gov/PubMed>) and the NLM Gateway (<http://gateway.nlm.nih.gov>). As several studies have shown, there are overlaps in coverage between the main indexing and abstracting services, but enough differences to mean that searchers need to use more than one service to gain an accurate picture of the relevant literature.^[2,3]



PRIMARY SOURCES

Patent Information

Each country has its own laws for patents, which means it is important to remember that countries differ in the length of protection time given; the name used for the term “patent”; and even about what can be patented. In the United States, patents are sometimes known as “utility patents,” but note that some countries provide “utility models” which are similar to patents, but are not as inventive and which also have a shorter protection period. Utility patents are also called “petty patents” (or “Gebrauchsmustern” in German). In the UK, it is neither possible to patent a new plant or animal, nor a method of treatment of the human or animal body by surgery or therapy or a method of diagnosis. However, this is not the case in all countries; e.g., in the United States, plants can be patented.

Patent searching is a specialized field, but there are several useful options for the novice patent searcher. All patent offices will keep indexes of what they have and many of these will be computerized. Chemical Abstracts include patents. A free service on the Web is esp@cenet (<http://gb.espacenet.com>), which is provided by the European Patent organization. It contains over 30 million worldwide patents going back to 1920, although the past three years are the most comprehensive. For some of these patent applications you can obtain the full text, for others only abstracts or basic bibliographic details. The U.S. Patent and Trade Mark Office Database (<http://www.uspto.gov/patft/index.html>) gives the full text of patents issued since January 1 1976 and the full-page images of every U.S. patent issued since 1790. The UK Patent Status Information Service (<http://webdb4.patent.gov.uk/patents>) provides information on the status of a UK patent, e.g., it is possible to see if it is still in force, if you know the Publication or Application Number.

It is worth noting that there is an inevitable overlap between patent and journal literature, as researchers who are registering patents will often be publishing their work in article form throughout a project.^[4] One study found that 60% of British pharmaceutical patents had been written about in article form. But of course that still leaves 40% which were not, and in any case the information contained in patents often differs from that given in journal papers.

e-Journals

A growing number of printed journals provide their contents online. Sometimes, this is for free—e.g., the *Pharmaceutical Journal* (<http://www.pharmj.com>) of the Royal Pharmaceutical Society of Great Britain. But more often than not a subscription is needed. For details of other

online pharmacy journals try the Virtual Pharmacy Library, PharmWeb (more information on these is given below in the section *Internet Resources*), and the Free Medical Journals site (<http://www.freemedicaljournals.com>). Some journals will provide additional information online to their printed versions. *Nature's* Web site (<http://www.nature.com>) contains auxiliary data from authors, a link to a grant database and the opportunity to take part in online discussions.

Collections of e-journals, online and CD-ROM, can be purchased so that you can access your own personal “virtual” library from your desktop. One of the largest collections is that from the publisher Elsevier Science, which provides access to over 1000 of its journal titles at ScienceDirect.com. Other collections include the Wiley Interscience and the American Chemical Society. Obviously, many of the titles in these collections will not be directly related to the pharmaceutical sciences, so a careful examination of the journals included is recommended before making a purchasing decision. You may find it more economical to use “pay per view” rather than take out a subscription to the full text of a collection/title.

Electronic only journals were pioneered by American physicists, but now encompass all areas of science. To look at free e-only journals, try the Public Library of Science (<http://www.publiclibraryofscience.org>) and BioMedNet (<http://www.bmn.com>). Electronic journals have their advantages: ease of availability in that you no longer have to make a special visit to a Library; they do not take up shelf space, and you no longer need to spend money on binding and binding repairs. For libraries, they have the added advantage of making it easy to gather statistics on usage to help inform acquisition decision-making. However, there are issues/questions that warrant consideration before moving from print to electronic journal subscriptions, some of which are as follows: Identifying and locating electronic resources can take more time than print.^[5] Printed journals are still generally more portable and easier to browse. Continuing access: if you cancel a title, will you still be able to access the issues you paid for? Archiving: will the publisher maintain a permanent archive? (In the UK, at present, there is no legal requirement for electronic/digital material to be placed in a national depository.) Will the publisher be willing to transfer the data to a new medium if the existing technology proves to be transitory? (In the late 1980s, as we all moved away from microfiche, we were told that CDs would last forever, but we have since discovered that their average life-span is from 10 yr to 15 yr.) Will the electronic edition of a printed journal be available before or after the print issue? Also, be sure to read the license conditions to check for any restrictive practices: for example, can you download an article rather than print it out?



INTERNET RESOURCES

The Internet, and in particular the World Wide Web, has brought about a huge change to the communication of drug information. Not only has it dramatically increased the amount of information (it has been estimated that the Web contains approximately 800 million pages of information^[6] and is doubling in size every year^[7]), but it has also enabled the lay person to easily access what in the past would have been “professional only” data. Governments, companies, societies, charities, educational establishments, and individuals are all contributing. It is not just the amount of information that can make the Internet difficult to search, it is also the difficulty in distinguishing between quality material and dubious material that could be downright dangerous. One survey looking at information on Ewing’s sarcoma found that 40% of web pages sampled were from nonreviewed resources and, of these, 6% gave incorrect information.^[8] There are many more such examples.^[9] Of course, it is not just drug information that can be obtained on the Internet, but also the drugs themselves. However, one must keep the dangers of the Internet in perspective.^[10] There is a great deal of extremely useful information available, and one should remember that the medium of print can also be abused.^[11]

If you need to search the Internet for an unknown site, a good place to start is a pharmacy specific “gateway.” This is a Web site which indexes relevant pages. The Argus Clearinghouse (<http://www.clearinghouse.net>) reviews these gateways (which are sometimes referred to as “virtual libraries” or “internet portals”). Started in 1994, PharmWeb (<http://www.pharmweb.net>) was one of the first gateways for pharmacy. It indexes sources from across the globe including governments, companies, societies, and pharmacy schools. The U.S. based Virtual Library: pharmacy section (<http://www.pharmacy.org>) also looks worldwide, but does have a U.S. bias, e.g., the UK’s National Health Service is not listed. PharmiWeb (<http://www.pharmiweb.com>) is designed for UK pharmaceutical industry and provides information on individual companies, employment vacancies, and news. InPharm.com (<http://www.inpharm.com>) is also aimed at industry, in that it creates web sites for companies to encourage commercial exploitation of the Internet, but it also lists web sites by topic (e.g., new drugs for stroke), and includes directories of those involved in medical communications, along with industry news and vacancies. A site designed by the London and South Eastern Drug Information Service, aimed at health professionals, is called DrugInfoZone (<http://www.druginfozone.org>). It contains a new section: the latest issues of current awareness bulletins and monthly bibliographic citations from the PharmLine database. Governments are

increasingly recognizing the need to steer the rapidly expanding number of Internet users towards quality health sites, and to utilize this source of information and communication to the benefit of both health providers and consumers. The U.S. Healthfinder (<http://www.healthfinder.gov>) does not just link to government information; it also lists companies, societies, charities, and academic institutions. The Electronic Library for Health (<http://www.nelh.nhs.uk>) has been designed by the UK government for professionals and the layperson, while NHS Direct (<http://www.nhsdirect.nhs.uk>) provides patients with information on symptoms, conditions, illness, and disease. The U.S. Food and Drug Administration has its own web site (<http://www.fda.gov>) but you can also access FDA information on new drugs and recalls at Intellihealth (<http://www.intelihealth.com>) a site jointly developed by the Harvard Medical School, the University of Pennsylvania and Aetna US Healthcare. Drug information can be searched by generic, brand name, or condition being treated. Although designed for the layperson there is a “professional network” area. Another drug database created for patients and searchable by generic and trade name is MedMaster (<http://www.safemedication.com/about/medmaster.cfm>) from the American Society of Health-System Pharmacists. This has been developed from the ASHP’s *Medication Teaching Manual: The Guide to Patient Drug Information*.

For a list of pharmacy email services, try PharmWeb or the other pharmacy specific gateways discussed earlier.

Once again, much time and energy can be expended on “surfing the net” in the hope of finding relevant, good quality Internet sites providing required drug information. Thus, we have suggested the following guidelines in the hope that they will prove useful.

Criteria/issues for selecting internet resources:

1. Provider/publisher credentials: Are they reputable? A well designed and credible resource will provide an “about us” section, with contact details, and there may also be a site logo. Do you recognize the provider? The URL will also give you a hint as to the origins of an organization, e.g., .ac.uk for a British academic institution and .edu for a U.S. educational organization. If any funding sources are listed, would these detrimentally affect the information included?
2. Geographical origin: The “about us” section or the URL should help. All countries, apart from the United States, have a country code for their URLs, although increasingly the domain .com is being used worldwide without any reference to the country of origin. There is also the problem that anyone is free to register a co.uk domain. However, it is another piece of evidence to take into account.



3. Author credentials: Do you recognize the author and/or the organization from which they originate? If you do not, it is worth doing a search on a relevant quality indexing and abstracting service to discover if they have published on the same topic in a reputable journal.
4. Does the author provide references to back up any statements?
5. Currency: Does the site provide a date for when it was last updated?
6. Criteria for information being included. Check the "about us" and "help" sections for this information.
7. Designated audience: If the site is one you want to suggest to your patients, remember it is helpful and a good practice for the site to state that users should visit a health professional if they have any concerns.
8. Awards: Obviously, if a site has won a reputable award it is a positive indication, but make sure you know the credentials of the award. Reputable awards include those from the Health on the Net Foundation (<http://www.hon.ch>). Their Web site also acts as a gateway to quality medical information.
9. Security: It is worth checking to see if the site has a confidentiality policy, with mechanisms to back it up, if it collects sensitive personal information. Although, of course, one should always remember that no networked computer is totally secure.
10. Speed of access: This can be affected by the providers' and users' hardware, communication technology, and software. Location is also important—Internet sites are always quicker early in the morning providers' local time.
11. Ease of use.
12. Last but not least—price of use, if any.

CONCLUSION

There is little doubt that the already overwhelming quantity of drug information will continue to grow, causing severe problems to those desirous of keeping up-to-date in their field. Additionally, the impact of the Internet explosion has added a further layer of "information overload" to be reckoned with—it is now necessary to search out quality sources of information by the use of gateways and search engines. Indeed (and somewhat ironically) these have developed so quickly and to such an extent that we are now experiencing a difficulty in keeping abreast of the available gateways and portals, let alone the primary information on any topic. New products and packaged services aimed at helping users find drug information are regularly advertised and launched in all professional and scientific fora. Thus, it will become increasingly important for the end user to

develop and hone the evaluative skills of critical appraisal to enable him or her to choose good quality and relevant sources of drug information. Finally, it is hoped that readers will remember that experienced drug information scientists and pharmaceutical librarians are there to help with the quest.

REFERENCES

1. Snow, B. Online Database Coverage of Pharmaceutical Journals. *Database* **1984**, 7, 12–26.
2. Fishman, L.; Stone, V.L.; DiPaula, B.A. Where Should the Pharmacy Researcher Look First? Comparing International Pharmaceutical Abstracts and MEDLINE. *Bull. Med. Lib. Assoc.* **1996**, 84 (3), 402–408.
3. Brown, C.M. The Benefits of Searching EMBASE Versus MEDLINE for Pharmaceutical Information. *Online and CD-ROM Rev.* **1998**, 22 (1), 3–8.
4. Eisenschitz, T. Intellectual Property. In *Information Sources in Pharmaceuticals*; Pickering, W.R., Ed.; Bowker-Saur, 1990; 144–170.
5. Tenopir, C.; King, D.W. The Use and Value of Scientific Journals: Past, Present and Future. *Serials* **2001**, 14 (2), 113–120.
6. Lawrence, S.; Giles, C.L. Accessibility of Information on the Web. *Nature* **1999**, 400 (6740), 107–109.
7. Bradley, P. *The Advanced Internet Searchers Handbook*; Library Association Publishing: London, 1999.
8. Biermann, J.S.; Golladay, G.J.; Greenfield, M.L.V.H.; Baker, L.H. Evaluation of Cancer Information on the Internet. *Cancer* **1999**, 86 (3), 381–390.
9. Purcell, G.P.; Wilson, P.; Delamothe The Quality of Health Information on the Internet: As for Any Other Medium it Varies Widely; Regulation Is Not the Answer. *BMJ* **2002**, 324 (7337), 557–558.
10. Smith, R. Almost No Evidence Exists That the Internet Harms Health. *BMJ* **2001**, 323 (7314), 651.
11. Shepperd, S.; Charnock, D. Against Internet Exceptionalism: There's Nothing Radically Different About Information on the Web. *BMJ* **2002**, 324 (7337), 556–557.

FURTHER READING

- Bowman, L.A.; Adams, M.S.; Christopher, A. Information Resources in Pharmacy and the Pharmaceutical Sciences. In *Remington: The Science and Practice of Pharmacy*, 20th Ed.; Gennaro, A.R., Eds.; Lippincott Williams & Wilkins: Baltimore, 2000; 60–69.
- Judd, A. *UK Drug Information Manual*, 4th Ed.; UKDIPG: Leeds, 1997.
- Newton, D. *How to Find Information: Patents on the Internet*; British Library: London, 2000.
- Snow, Bonnie. *Drug Information: A Guide to Current Resources*, 2nd Ed.; Medical Library Association: Scarecrow, 1999.

DRUG DELIVERY—VAGINAL ROUTE

Yie W. Chien

Kaohsiung Medical University, Kaohsiung, Taiwan

Chi H. Lee

The University of Missouri-Kansas City, Kansas City, Missouri, U.S.A.

INTRODUCTION

Vaginal dosage forms have been developed and used clinically for many years in local therapy and the systemic delivery of systemically effective drugs. Pharmaceutical dosage forms available for intravaginal delivery consist primarily of those used to treat a specific gynecological condition. A small survey conducted to evaluate the preferences of females with regards to the use of intravaginal medications showed a positive market outlook for this mode of delivery (1). Products available include vaginal contraceptives, antifungals, antimicrobials, cleansers, deodorants, and lubricants. These products are formulated as tablets, capsules, creams, suppositories, foams, films, solutions, ointments, and gels. Currently, numerous prescription and over-the-counter (OTC) medications intended only for local activity in the vagina are available.

Recently, the vagina's absorption capacity has been recognized, which suggests that the vagina could provide a potential route for systemic drug delivery with a direct entry into the blood stream (Fig. 1) with the possibility of bypassing the hepatic-gastrointestinal (GI) metabolism. Several pharmacologically active compounds, that are metabolized extensively when taken orally, such as progesterone and estrogen, have been delivered intravaginally for achieving their systemic activity. Use of the vaginal route as a novel site for drug delivery has recently received greater attention, particularly with the new focus on the therapeutic agents that are subject to an extensive hepatic "first-pass" elimination, such as therapeutic proteins and peptides.

This article describes the physiology of the human vagina, its characteristics of absorption, and its permeability. The reader will also be familiarized with current research trends in vaginal delivery and absorption of drugs.

THE HUMAN VAGINA

General Anatomy

The vagina is a canal extending from the vulva to the cervix (Fig. 1). Extensive investigations on its morphology and anatomy have been compiled (2). Physiologically, the vagina serves a few functions, acting primarily as a conduit for the passage of seminal fluid, an excretory duct for menstrual discharge, and as the lower part of the birth canal (3). The anterior portion of the vagina in an adult averages 6–7 cm in length, while the posterior wall is approximately 7.5–8.5 cm.

The vagina is characterized by an exceptional elasticity, having the greatest resiliency at parturition. Along the length of the vagina, a layer of relatively thick connective tissue is located between the anterior vaginal wall and the urinary tract as well as between the posterior vaginal wall and the intestinal canal. The vaginal wall itself consists of three layers: The epithelial layer, the muscular coat, and the tunica adventitia. The epithelial layer is made up of an epithelial lamina and a lamina propria. It is a noncornified, stratified squamous epithelium that is subject to changes with aging. The epithelium atrophies from birth to puberty, at which time hormonal activity increases the thickness and resistance of this layer. In the subepithelial layer, there rests a network of elastic fibers around the lamina propria and collagenous fibers around the tunica adventitia, creating a connection to the muscular coat. Changes in the cytology of the vaginal epithelium occur with the cyclical stages in women. The epithelium is thickest in the proliferative stage, peaking at ovulation, and then diminishing with the secretory phase. The muscular coat of the vagina is composed of smooth muscle and elastic fibers. A spiral arrangement of these fibers provides support to withstand stretching without rupturing the vagina. The tunica adventitia is formed of loose connective tissue that is attached to the muscular coat. Fluctuations in

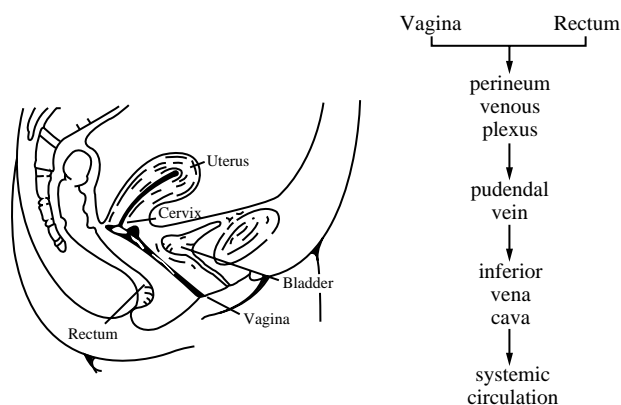


Fig. 1 Lateral view of the female pelvis showing the absorption route to the systemic circulation.

the volume of the vaginal lumen occur due to alterations in the tension of this layer. The vagina is encompassed by a vascular supply of arteries, veins, and lymph capillaries, as well as sensory and autonomous nerves.

Cellular structure

Histological studies of vaginal biopsies from healthy volunteers during follicular and luteal phases,

postmenopause, and following ovariectomy have been conducted to characterize the ultrastructure of the vaginal mucosa by electron microscope (4). The epithelium of the vaginal mucosa is found to have five different layers of cells: The basal, parabasal, intermediate, transitional, and superficial layers (Fig. 2). The cellular types that make up these various layers renew continuously as they are stimulated by hormonal action and intracellular communication. The basal cells are typically columnar or squamous in shape with microvilli present on the surface of the cell membrane. Parabasal cells are similar to the basal cells in size and structure, but have a greater formation of surface microvilli and interdigitations. Their polygonal shapes are formed by adapting to spaces left free from neighboring cells. The cells of the intermediate layer possess microvilli and are of the largest cell type. The transitional cells that follow show noticeable signs of involution and surface characteristics of diminishing and thinning microvilli and intracellular junctions (desmosomes). Superficial cells, as indicated by their nomenclature, are the cells of the outermost layer during the follicular phase of the cycle.

The vaginal epithelium contains a network of intercellular channels that continuously undergo development, reaching a maximum during the ovulatory and

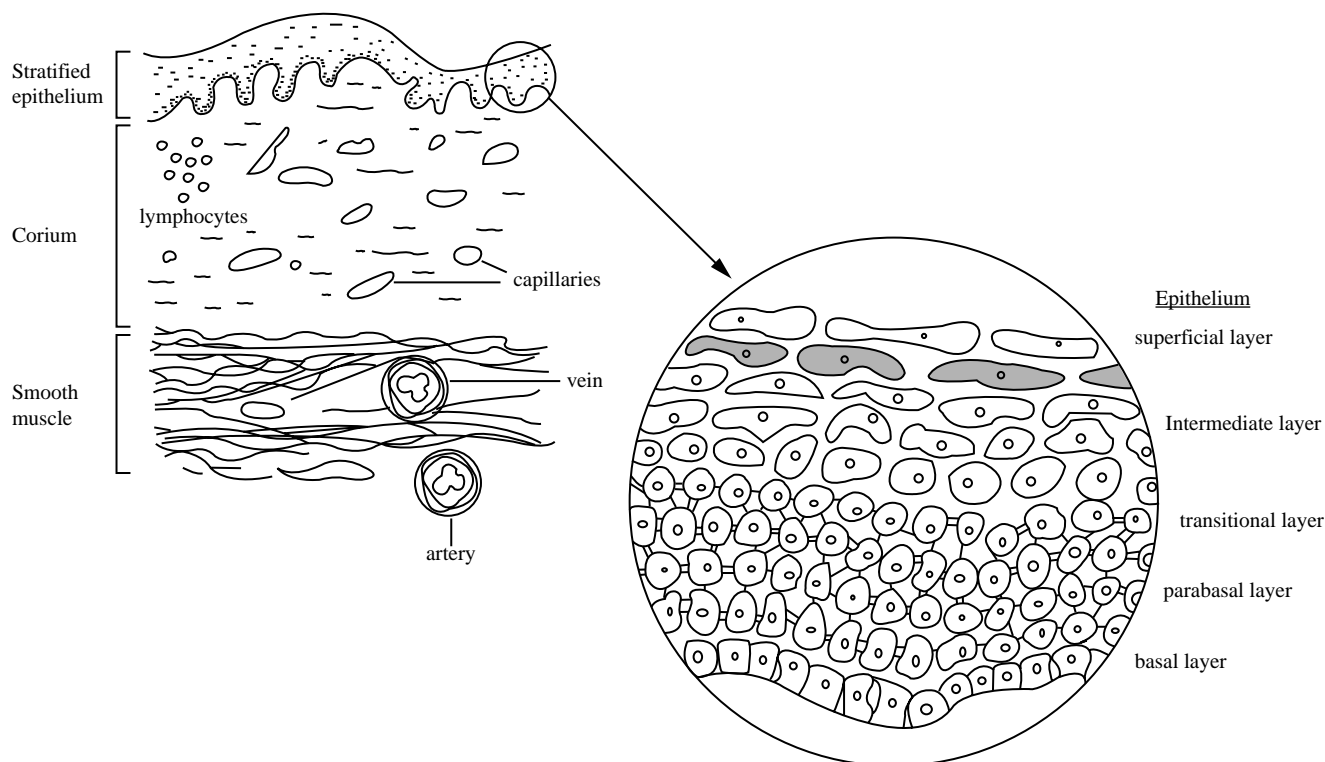


Fig. 2 Cross-sectional view of the vaginal wall with magnification of the stratified epithelial layers. (Based on Ref. 6.)

luteal phases. The channels present in the transitional and superficial layers do not change with the cycle as do those in the basal, parabasal, and intermediate layers. These channels provide a supply of nutrients and transport metabolites to and from the layers of the epithelium. Since the vagina is absent of secretory glands, lubrication is provided via these channels. Interestingly, different substances have been shown to be rapidly absorbed through the vaginal mucosa and will be discussed later. Intercellular junctions, including desmosomes and tight junctions, have also been identified. Desmosomes are most prominent in the intermediate layer and progressively become less toward the superficial layer, which may play a role in the desquamation of vaginal cells.

Since the vaginal epithelium is affected by ovarian hormones, cyclical variations can occur (Fig. 3). Although less intense than the uterine modifications, changes include proliferation, differentiation, and desquamation. During the follicular phase, the time period between the end of menstruation and the day of ovulation, mitosis increasingly occurs in the cells of the basal and parabasal layers, creating an increase in the number of layers and the thickness of the epithelium. The desquamating layers increase until ovulation, after which the layers diminish and are sloughed away through the vaginal lumen. During the luteal phase, the period after ovulation, the transitional cells become superficial due to the absence of the normal superficial layer.

It has been found that the basal cells replicate continuously to provide a self-cleaning mechanism to the epithelial layer. Autoradiographic studies of cell proliferation were performed on normal human cervix and vagina (5). The turnover time, an indication of the time required for the replacement of the cell population, was determined from these studies. Results showed the basal layer to be relatively inactive with a turnover rate of 33 days, while active proliferation occurred in the parabasal layers with a turnover rate of 3 days. The intermediate and superficial layers were found to be inactive differentiating compartments.

Fluids and enzymes

Despite the paucity of glands, the vaginal epithelium is usually kept moist by a surface film. This film, known as vaginal fluid, consists of cervical mucus and exfoliated cells from the vagina itself. Transudation from the blood vessels through the intercellular channels to the lumen can also contribute to the chemical composition (4). The fluid can contain carbohydrates, amino acids, aliphatic acids, protein, and immunoglobulins (Igs) (6). Nonserum proteins in human vaginal secretions have recently been

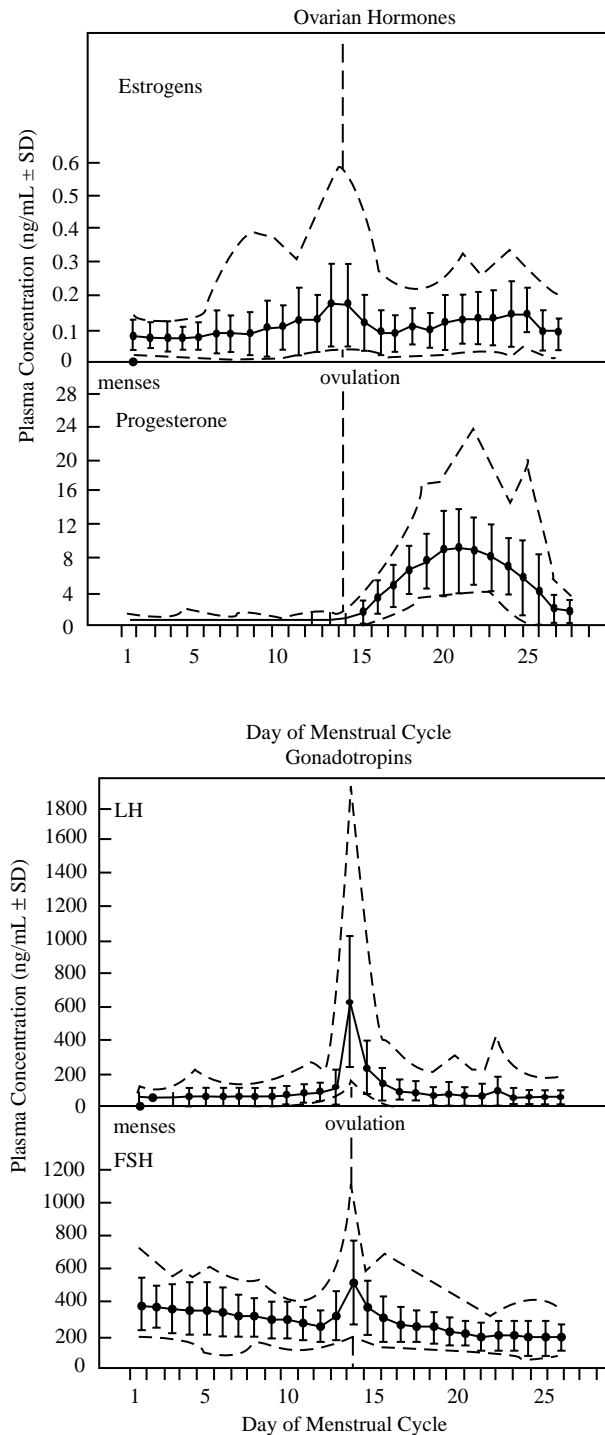


Fig. 3 Profile of the gonadotropin and ovarian hormones during a normal menstrual cycle. The number of cell layers in the vaginal epithelium rises from 22 layers at approximately day 10 of the cycle to 45 layers at ovulation, and drops to 33 layers around day 20. (Based on Refs. 87 and 164.)

localized and analyzed (7, 8). Typically, the vaginal fluid in mature, healthy women has a pH in the range 4–5 (9). This acidic environment is produced by the presence of lactobacilli, which convert carbohydrates to lactic acid. The cervical mucus, a principal component of the vaginal fluid, is produced by glandular units within the cervical canal and has a pH in the range of 6.5–9. The cervical mucus changes in composition and physical characteristics with the menstrual cycle, facilitating sperm migration during ovulation. At the time of ovulation, the vaginal fluid increases in volume. This is due to the augmented amount of cervical secretions. The mucus produced at ovulation has increased spinnbarkeit (fibrosity), ferning (crystallization of the mucus when dried on a slide), pH, and mucin content (10). Additionally, a decrease in the viscosity, cellularity, and albumin concentration is noted.

The variety of enzymes found in the vagina is an important concern for the development of vaginal delivery systems, particularly with proteases and their effect on protein and peptide candidates (11). The outer cell layers of the vagina contain varying amounts of β -Glucuronidase, acid phosphatase, α -Naphthylesterase, diphosphopyridine nucleotide-diaphorase (DPND), phosphoamidase, and succinic dehydrogenase (9). Enzymatic activity has been shown in the basal cell layers as well, and these layers contain β -Glucuronidase, succinic dehydrogenase, DPND, acid phosphatase, and α -Naphthylesterase.

In addition to enzymes, the vaginal lumen is a nonsterile area inhabited by a variety of microorganisms, mainly *Lactobacillus*, *Bacteroides*, and *Staphylococcus epidermidis*, as well as potentially pathogenic aerobes (12). The existence of these microbes and their metabolites may also have a detrimental effect on the intravaginal stability of a vaginal drug delivery device.

Physiology and Dynamics

The unstimulated vagina anatomically consists of a luminal space that exists potentially rather than actually. However, in response to sexual excitement, some tension-induced anatomic variations occur. These anatomic variations may have effects on the long-term intravaginal delivery systems. These variations are reflected in the existence of four phases in a sexual response cycle as outlined by Masters and Johnson (13).

The first sign of physiological response (excitement phase) to stimulation is the production of a vaginal lubricating fluid. This appears on the vaginal mucosal surface within 10–30 sec after an effective stimulation. As sexual tension progresses, individual droplets of

transudation-like mucoid material appear scattered throughout the rugal folds of the vaginal lumen, coalescing to form a smooth coating over the entire vaginal mucosa surface. This transudative mucoid material results from the activation of a massive localized vasocongestive reaction and marked dilation of the venous plexus that encircles the entire vaginal lumen. This sweating phenomenon provides complete lubrication of the vagina. The inner two-thirds of the vaginal lumen lengthen and distend. As sexual tension mounts toward the plateau phase, the vaginal wall in this area expands involuntarily and then partially relaxes in an irregular, tensionless manner. The demand to expand gradually overcomes the tendency to relax. In addition to the expansive effect in the vaginal fornices, the cervix and corpus pull slowly backward and upward into the false pelvis position. This cervical elevation creates a “tenting effect” at the transcervical depth in the midvaginal plane. This phenomenon always occurs in a normal anteriorly positioned uterus (Fig. 4). The vagina of either nulliparous or multiparous women, regardless of prior degree of vaginal expansion or lengthening, increases substantially in length and transcervical width with sexual stimulation.

With attainment of the plateau phase level of sexual tension, a marked localized vasocongestive reaction develops in the outer one-third of the vaginal lumen. The

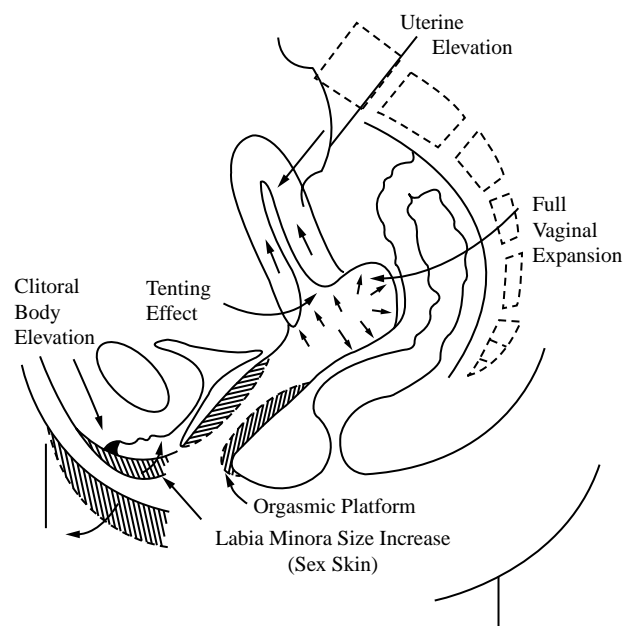


Fig. 4 Lateral view of the female pelvis showing the stimulated vagina. (Based on Ref. 13.)

entire area becomes grossly distended with venous blood, and its central lumen is reduced by at least a third as compared to the distention previously established in the excitement phase. The increase in the width and depth of the vaginal lumen is minimal. The production rate of vaginal lubricating fluid also gradually slows, particularly if this level of sexual tension has been experienced for an extended period of time.

During the orgasmic phase, the basic response of the inner vaginal lumen is essentially expansive rather than constrictive in character. On the other hand, the bulbar vasoconstriction at the orgasmic platform in the outer one-third of the vaginal lumen contracts strongly in a regularly recurring pattern. The intercontractile intervals lengthen in duration, and the intensity of the contractions progressively diminishes.

Along with the onset of the resolution phase, retrogressive changes develop first in the outer one-third of the vaginal lumen. The localized vasocongestion is dispersed rapidly, leading to an increase in the diameter of the central lumen of the outer one-third of the vagina. The previously expanded inner two-thirds of the vaginal lumen also gradually shrinks back to the original collapsed, unstimulated state. This shrinking process is an irregular, zonal-type relaxation of the lateral and posterior walls. The anterior wall and the cervix of the anteriorly positioned uterus descend rapidly toward the vaginal floor, leading to a quick resolution of the tenting effect created earlier during the excitement phase.

Menopause

The natural aging process results in significant changes in the vagina, including a decrease in vaginal size, loss of elasticity, loss of vascularity, and a thinning of the mucosa (14). The cytology of the vagina is variable and many aspects are addressed by Steger and Havez (14). The epithelium becomes markedly thinner and is often invaded with leukocytes. Areas can be completely denuded of an epithelial covering, exposing the subepithelial connective tissue. On the surface of the vagina, the numbers of exfoliating cells and microridges are greatly reduced. Creating the loss of elasticity, collagen replaces many of the elastic fibers in the lamina propria. Glycogen is very low or completely absent, contributing to the change in vaginal microbiology and pH. Vaginal secretions become scant and watery, and the pH increases from 4.5–5.5 to 7.0–7.4. Resistance to bacterial and fungal infections is reduced due to the lower population of acidophilic organisms (15). The enzymes present in the vagina also increase with the onset of menopause, including

β -Glucuronidase, acid phosphatase, and nonspecific esterases (14).

Unlike other tissues, the vagina is greatly affected by steroid replacement therapy. Estrogen replacement therapy is often used to treat menopausal symptoms. The postmenopausal state of the vaginal epithelium, with its thinner epithelium and increased permeability, is an important consideration in drug delivery. The minimization of epithelial fluctuation will result in less fluctuation in absorption, affecting both systemic and local drug deliveries.

VAGINAL ABSORPTION

Much has been written in the literature concerning vaginal absorption. The first experimental studies using animals dates back to 1918. At that time, the histological characteristic of the vaginal wall was known to exist in three simple layers: the connective tissue, muscular, and the mucosa, collectively resembling the skin without the stratum corneum. Originally, the vagina was regarded as an organ impermeable to exogenous agents. Reports began to surface that indicated vaginal absorption of foreign materials as the cause of toxicity and even death in several cases. Using dogs and cats, vaginal absorptions of large varieties of compounds, including alkaloids, inorganic salts, esters, and antiseptics, were demonstrated (16). Later work showed the vaginal absorption of compounds such as hydrocyanic acid, pilocarpine, atropine, and insulin in dogs and cats (17). Using rabbits, cats, and dogs, the vaginal absorption of quinine bisulfate and oxyquinoline sulfate was set forth, intending to emphasize the care and consideration required of physicians with regards to the local applications of medication to the vaginal mucosa (18). A unidirectional transmission of agents was proposed to occur from the vagina to the blood, with no transmission in the reverse direction (19, 20).

An extensive review of the literature documented the absorption of carbohydrates, fats, and proteins (21, 21). Glucose was absorbed and rapidly oxidized. One of the first proteins demonstrated to be absorbed vaginally was peanut protein (23). The vaginal permeation of spermatozoa and bacterial antigens also has been shown (22). Bacterial antigens play an important role in triggering the local immunological mechanisms involved in protecting the area against infection. Other classes of compounds include steroids (e.g., estrogens, progesterone, and testosterone), prostaglandins, antimicrobials, nonoxonyl-9, and methadone. The vaginal delivery of estrogen and

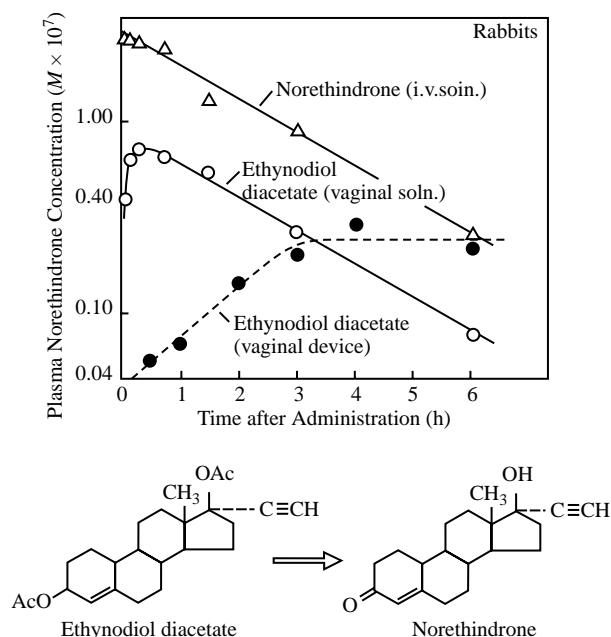


Fig. 5 Rabbit plasma concentration profiles of norethindrone following the intravenous administration of a single dose (solution). Also shown is the intravaginal absorption of ethynodiol diacetate from a solution dose and from a vaginal delivery device. (Based on Ref. 37.)

progesterone has been well documented over the years and used clinically in dosage forms, such as vaginal creams and suppositories.

Vaginal absorption of drugs is dependent upon such physicochemical properties as molecular weight, lipophilicity, ionization, molecular size, chemical nature, and local action, as well as the thickness of the vaginal wall as affected by the ovarian cycle or pregnancy (21). Other factors include changes in the vaginal epithelium and pH with menopause. Prior to absorption, drugs must be in solution. The fluid present in the vagina can

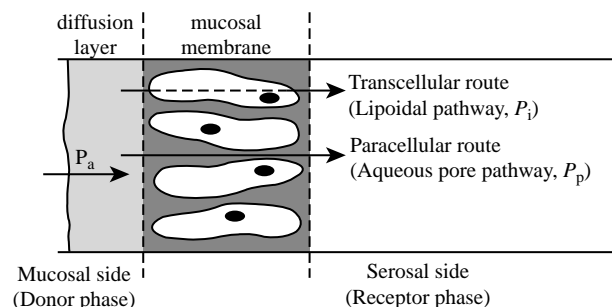


Fig. 6 Schematic of the vaginal membrane as a transport barrier. (Based on Ref. 25.)

help to dissolve drugs, but the cervical mucus secretion also can present a barrier and remove a drug from the site when abundant (11). Dosage forms can undergo different absorption due to the differing dissolution patterns in vaginal fluid. Products such as creams, inserts, and tablets remain for different periods of time in the vagina. A comparison of vaginal inserts versus creams showed that creams have a longer contact time in the vagina (24).

Permeability

Transport across the vaginal membrane can occur by three primary pathways: the transcellular route, by which diffusion occurs through the cell due to a concentration gradient; the intercellular route, where diffusion occurs through spacing between cells; or by a vesicular or receptor-mediated transport (11, 25). Vaginal permeation studies have been conducted using the rabbit as an animal model (26, 27). The female rabbit does not exhibit an estrus cycle, so its vaginal tissues show constancy in the histological, biochemical, and physiological properties not ordinarily seen with most other mammals (28). The lack of a sexual cycle is, therefore, expected to produce a minimal variability in the permeability of the vaginal membrane, making the measurements of vaginal drug permeation more controllable and accessible (26, 27, 29).

The vaginal mucosa permeability of the doe rabbit has been examined by continuous perfusion of straight-chain alkanols and alkanolic acids (26, 27). Similar to the vaginal absorption of ethynodiol diacetate (Fig. 5), the vaginal uptake of both alkanols and alkanolic acids also follows a first-order rate process and is dependent on the drug concentration in the vaginal fluid. The results agree well with a physical model that has a hydrodynamic diffusion layer in series with the mucosal membrane, that consists of two parallel pathways: a lipoidal pathway and an aqueous "pore" pathway (Fig. 6). Immediately behind the mucosa (serosal side) a perfect sink is maintained by hemoperfusion.

The apparent permeability coefficient P_{app} for vaginal membrane permeation is defined by

$$P_{app} = \frac{1}{\frac{1}{P_{aq}} + \frac{1}{P_v}} \quad (1)$$

or

$$P_{app} = \frac{1}{\frac{1}{P_{aq}} + \frac{1}{P_p + P_l}} \quad (2)$$

since

$$P_v = P_p + P_l \quad (3)$$

where P_{aq} , P_v , P_p and P_l are the permeability coefficients of the aqueous diffusion layer, the vaginal membrane, the aqueous pore pathway, and the lipoidal pathway, respectively.

The vaginal permeation kinetics of a series of straight-chain alkanols was investigated (26). Using methanol as a reference permeant, a normalized permeability coefficient: $P_{app}(alc/MeOH)$ was determined for each of the alkanols. The normalized permeability coefficient was observed to increase in value as the alkyl chain length of the alkanols increased (Table 1). The increased permeability can be attributed to the increase in the permeability coefficient for the lipoidal pathway P_l (Eq. 2). It is estimated that for straight-chain aliphatic alcohol the P_l value increases by 2.5 for the addition of each methylene (CH_2) group (26). On the other hand, the P_l value increases by 3.5 for the series of straight-chain alkanolic acids (27).

For the vaginal absorption of ionizable compounds, such as the homologous series of n -alkanoic acids, the apparent permeability coefficient P_{app} becomes pH-dependent and is defined by Eq. 4:

$$P_{app}(n, pH) = \frac{1}{P_{aq}(n)} + \frac{1}{\frac{[H]}{K_a + [H]} P_l^0 10^n \pi + P_p} \quad (4)$$

where n is the number of methylene (CH_2) groups in the alkyl chain; $[H]$ is the concentration of protons; K_a is the dissociation constant of the acid; P_l is the permeability coefficient of the lipoidal pathway for the hypothetical acid with zero carbon atom ($n = 0$); and π is the methylene

Table 1 Effect of alkyl chain length on the normalized permeability coefficients of straight-chain alkanols^a

Alkanol	$CH_3(CH_2)_nOH$	$P_{app}(Alkanol/MeOH)^b$
Methanol	$n = 0$	1.00
Propanol	$n = 2$	1.11
Butanol	$n = 3$	1.13
Pentanol	$n = 4$	1.20
Hexanol	$n = 5$	1.48
Heptanol	$n = 6$	1.91
Octanol	$n = 7$	2.15

^aBased on data from Ref. 26.

^bNormalized permeability coefficient = $P_{app}(alcohol)/P_{app}(methanol)$. Mean value from three rabbits at pH 6.0 and 37°C.

Table 2 Effect of alkyl chain length and pH on the normalized permeability coefficients of straight-chain alkanolic acids

Acids	$P_{app}(Acid/MeOH)^a$		
	pH 3	pH 6	pH 8
Acetic	1.22	0.73	0.25
Butyric	1.62	1.94	0.34
Hexanoic	1.89	2.06	0.81
Octanoic	1.74	2.49	1.24
Decanoic	—	—	1.26

^aNormalized permeability coefficient = $P_{app}(acid)/P_{app}(methanol)$. Mean value from three experiments involving different rabbits at pH 3, 6, and 8.

(Based on data from Ref. 30).

group incremental constant, that is, 3.5 for straight-chain alkanolic acids.

As illustrated earlier in the homologous series of n -alkanols, the normalized permeability coefficient of n -alkanoic acids also shows a dependence on alkyl chain length (Table 2). In addition, the straight-chain alkanolic acids demonstrate a pH-dependence in their normalized permeability coefficients (27). It should be pointed out that the rabbit's vaginal secretion has an effective pK_a value of 6.3 ± 0.1 . However, the rate of vaginal secretion is relatively small, which leads to a surface pH of around 2.0–2.1 (30). This acid surface pH affects the extent of dissociation of n -alkanoic acids and thus the magnitudes of P_l and P_{app} (Eq. 3).

The vaginal uptake of steroids has also been studied and follows a first-order rate process as well (31). The normalized permeability coefficient of steroids appears to be dependent upon steroidal structure (Table 3). The permeability coefficient across the vaginal membrane (P_v) also shows the same trend of structure dependence; the lipophilic steroids (progesterone and estrone) were better absorbed than the more polar steroids

Table 3 Vaginal permeation parameters of representative steroids

Steroids	P_{app}^a	$P_v \times 10^4$ (cm/s)	$P_{aq} \times 10^4$ (cm/s)
Estrone	1.00	7.60	2.81
Progesterone	0.93	6.10	2.80
Testosterone	0.29	0.75	2.76
Hydrocortisone	0.23	0.58	2.79

^aNormalized permeability coefficient = $P_{app}(steroid)/P_{app}(methanol)$. (Based on data from Ref. 31.)

(hydrocortisone and testosterone). However, the P_{aq} values, the permeability coefficient across the hydrodynamic diffusion layer, are very much the same among the four steroids (Table 3). For drugs with high P_v values, such as progesterone and estrone ($6.1\text{--}7.6 \times 10^{-4}$ cm/s), vaginal absorption is mainly controlled by their permeability across the hydrodynamic diffusion layer on the surface of the vaginal mucosa ($P_v > P_{aq}$). Conversely, for drugs with low P_v values, such as testosterone and hydrocortisone ($5.8\text{--}7.5 \times 10^{-5}$ cm/s), vaginal uptake is determined predominantly by their molecular permeation through the vaginal membrane ($P_v \gg P_{aq}$) (32). A linear relationship between the vaginal permeability and a series of progestins and their lipophilicities have been shown in vitro using rabbit epithelium (33).

The apparent permeability coefficient P_{app} is related to the first-order rate constant for the disappearance of drug from vaginal lumen (k_v) as follows:

$$P_{app} = k_v \frac{V_v}{S_v} \quad (5)$$

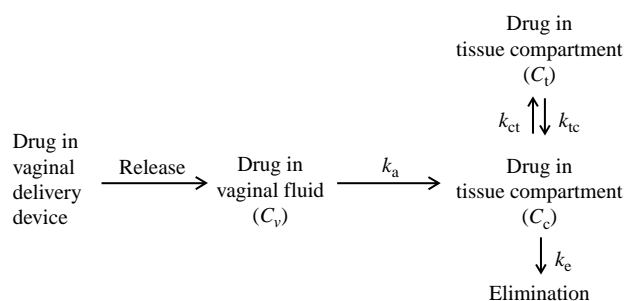
where V_v is the volume of vaginal fluid and S_v is the geometric surface area of the vaginal lumen.

Additional permeability studies have been conducted in rabbits by the measurement of electrical conductance and flux measurements of a hydrophilic fluorescent probe 6-Carboxyfluorescein (34). Membrane permeation selectivity was indicated by KCl diffusion potential or charge-discriminating ability. The fluorescein probe is known to permeate via the paracellular route. The permeation of the 6-carboxyfluorescein was found to be in the order of intestine \approx nasal \geq bronchia \geq tracheal $>$ vaginal \geq rectal $>$ corneal $>$ buccal $>$ skin. The permeation selectivity was found to be negative; in other words, the K^+ ion was more permeable than the Cl^- . With all the tissue types evaluated, the permeation selectivity was found to be similar. Using low-molecular-weight polyvinyl alcohol, the molecular weight cutoff of the vaginal epithelium of rats was found to be greater than other assessable surfaces, such as the GI tract (35). Permeability trends have been evaluated using the spermicide, nonoxynol-9 (36). A linear correlation was observed between permeability and partitioning of nonoxynol-9 oligomers using lamb vaginal mucosa, suggesting the importance of the lipoidal pathway for this particular agent.

Pharmacokinetics

If the absorption, distribution, and elimination of a drug molecule after its release from a vaginal drug delivery device in the vaginal lumen follow the pharmacokinetic

sequences,



then the instantaneous rate of change in drug concentration in the central compartment can be expressed by Eq. 6:

$$\frac{d(C_c)}{dt} = k_a C_v + k_{tc} C_t - (k_{ct} + k_e) C_c \quad (6)$$

where k_a , k_e , k_{cy} , and k_{tc} are the rate constants for absorption, elimination, and central compartment/tissue compartment exchange, respectively, and C_a , C_c , and C_t are the drug concentrations in the tissue fluid surrounding the vaginal device, in the central compartment and in the tissue compartment, respectively.

The vaginal absorption of drug following its release from vaginal drug delivery devices may alternatively be described by a simplified one-compartment open model with first-order drug absorption (Fig. 5) (37). Using this simplified model, Eq. 6 is reduced to Eq. 7:

$$\frac{d(C_c)}{dt} = k_a C_v - k_e C_B \quad (7)$$

where C_v and C_B are the drug concentrations in the vagina and in the body (including blood, tissues, and related compartments with fast drug exchange rates), respectively.

At steady state (Fig. 7), the change in the body concentration of the drug is relatively small, $d(C_B)/dt \approx 0$. For example, the body concentration of norethindrone (C_B), the major metabolite of ethynodiol diacetate (Q), is then related to the amount of (Q) released at time t as shown by Eq. 8:

$$C_B = \frac{k_a \Sigma R_{2v}}{2k_e} \left(\frac{Q}{t} \right)_v \quad (8)$$

where ΣR_v is the total diffusional resistance across the vaginal wall. Eq. 7 suggests that the norethindrone concentration (C_B) in the body of each test animal should be directly proportional to the amount of Q released from the vaginal device, $(Q/t)_v$, for a given duration of intravaginal residence. From the slope of the C_B versus $(Q/t)_v$ plots and the values of k_a and k_e from the vaginal absorption studies of the same drug from a solution, the

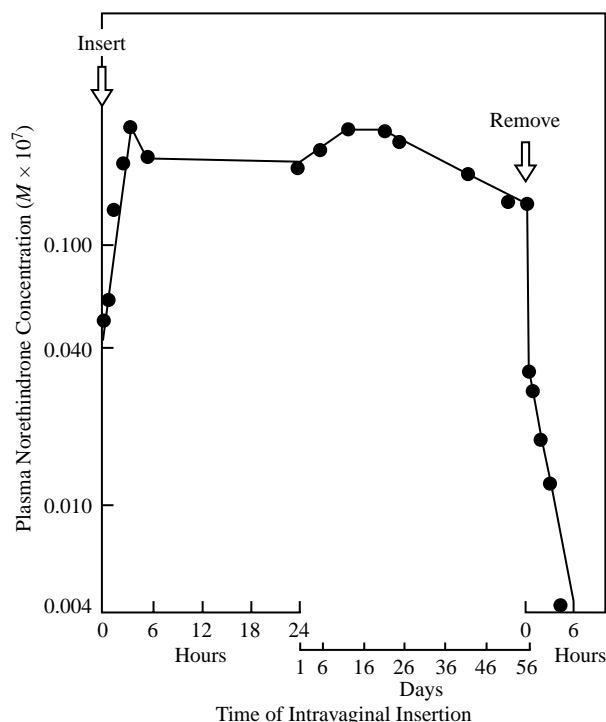


Fig. 7 Plasma profile of norethindrone following the intravaginal insertion of ethynodiol diacetate-releasing vaginal devices in rabbits for 56 days and after device removal. (Based on Ref. 37.)

magnitude of R_v , the total diffusional resistance for the vaginal permeation of the drug, can be estimated.

Physiological Factors Affecting Drug Delivery through Vaginal Mucosa

Mucus: Compositions, characteristics, and roles

The vaginal mucosa, to which a bioadhesive drug delivery system is expected to come into contact when inserted intravaginally, consists of an epithelium having its surface coated with a layer of mucus. The mucus is a heterogeneous secretion and provides a protective and lubricating functions to the epithelium. In a normal woman, the mucus is produced by the cervix at the rate of 20–60 mg/day. During the midcycle, the rate increases to 700 mg/day (38) and the mucus becomes a less viscous and microstructurally more expanded in texture, which facilitates the penetration of sperm (39).

To understand the nature of mucus, it is necessary to understand the structure of glycoprotein, the principal biochemical component of the mucus subunit (40, 41). Glycoprotein, which has a molecular weight of 320–4500 Da, is bound to other subunits through disulfide

bonds and probably interacts with other subunits through ionic bonds and entanglements, especially with sialic acids located at the terminal ends of the oligosaccharide chain (42). In addition to glycoproteins, the cervical mucus also contains a wide range of substances, including plasma proteins, enzymes, amino acids, cholesterol, lipids, and inorganic ions, with concentrations known to fluctuate during the cycle (43). It has been proposed that entanglement of the macromolecules with the specific lectin-like regions contributes to the properties of the cervical mucus (44, 45). The relationship between the intrinsic viscosity and the molecular weight of the whole mucins and their subunits and T-Domains suggests that they are flexible linear macromolecules behaving like a stiff random coil (46).

The hostility of the thickened cervical mucus to sperm penetration has been used as a means to achieve contraception, and low-dose oral contraceptives depend largely on this condition for their effectiveness in fertility control (47, 48). Sequential oral contraceptives do not affect mucus and like estrogen, may even increase sperm penetration.

The change of viscosity in the cervical mucus

Alterations in the biochemical properties of mucus are known to be responsible for changes in rheological behavior and receptivity of mucus to various exogenous compounds (49). The liquefaction can be regulated by mechanical disruption, systemic carrier dilution, or chemicals, such as mucolytic agents. The effect of the alterations in the physical and physiological properties of cervical mucus on the change in drug permeability through the cervical mucosa was studied (50). Since marked changes occur in the plasma membrane during epididymal maturation and capacitation, analysis of the nature of compounds at biochemical level was important to understand the changes in permeation rate in response to the viscosity of mucus (51). The increased ionic strength and consistency of the periovulatory mucus offers better permeability to exogenous compound, and increased charge favors a higher degree of hydration (52).

The viscosity of mucus is affected by binding between calcium and the mucus, which probably arises from an ionic interaction with the sialic acid in the mucin (44). These variations are indicated by a change in the pH, viscoelastic properties, water, and protein content of cervical mucus (53). Calcium is needed to establish an intercellular contact and the assembly of tight junction in the cervical epithelium (54). Changes in extracellular calcium affect the permeability of tight junctions and play a role in regulating the production of cervical mucus (55). Prostaglandin concentrations in the cervix affect the

viscosity of cervical mucus, which in turn affects the cervical softening (56). The effect of dithiothreitol (DTT) on the viscosity was also reported (57). Thus, the physical and physiological properties of cervical mucus may reflect alterations in the macromolecular composition or concentration of its components on exposure to exogenous substrates or hormone replacement therapy (58). As an example, administration of mestranol (or ethinyl estradiol), in combination with norethisterone or its acetate was known to strongly affect the biophysical properties of the cervical mucus and sperm migration (59).

The change in protein composition in cervical mucus

Human cervical mucus contains 1–3% proteins, in which the major soluble proteins are albumin and gamma globulin (60). The effects of various agents on the protein composition were studied (61). When a purified glycoprotein was treated with various proteolytic enzymes, which degraded (the purified) glycoprotein, the charge of cervical mucus increased to confer a rigidity by the mutual repulsion of negative charges and strengthen the coherence and consistency of the secretion (62). It appears that the menstrual cyclic changes in mucus viscoelasticity in an individual can be accounted for by the changes in mucin concentration (63). This appears to result from a decrease in the amount of proteins in both the follicular and the luteal phases, but an increase in the ovulatory phase (64). In addition, mucus compositional differences may occur among individuals, as indicated by the different correlations seen in the viscoelasticity and the mucin concentration (65).

The effect of the combination of nonoxynol-9 and EDTA was studied and showed that spermicidal activity was significantly enhanced (48). The observations may be due to a partial removal of calcium, on addition of chelating EDTA, which may have caused a decrease in the dephosphorylation of regulatory proteins mediated by calcium, and further caused a change in the total amount of protein and its composition in the cervical mucus. The nature of mucus components, especially proteins and enzymes, need to be studied further to fully elucidate the mechanism by which cervical mucus regulates the permeation rate of exogenous compounds. A change in the protein composition is expected to affect enzyme activity in the cervical mucus and further affect the stability and permeation rate of a drug delivered by a vaginal drug delivery device (66).

Cyclic variability

As mentioned previously, the rabbit appears to be an ideal animal model for studying vaginal mucosa permeation due to the absence of an estrus cycle.

Unfortunately, the doe rabbit may not be a suitable animal model for studying long-term vaginal absorption because it lacks the typical cyclic variations observed in the human vaginal tract associated with the rhythmic pattern of hormones during a menstrual cycle. In the human female, the cyclic secretion of estrogenic hormones in the ovarian cycle induces variation in the histology, biochemistry, and physiology of vaginal tissues. It is, therefore, reasonable to expect that the vaginal mucosa undergoes a corresponding cyclic variation in its membrane permeability.

The macaque rhesus monkey has an ovarian cycle of approximately 28 days, as does the human female, and it also exhibits an estrus pattern very similar to the menstrual pattern of the human female. It is widely believed by researchers in the fertility field that rhesus monkeys and humans have comparable anatomy and physiology, as well as similar reproductive functions (67). Therefore, the female rhesus monkey is a superior animal model for studying the vaginal absorption of various drugs from a drug delivery system designed for use in human females.

The effect of the estrus cycle on the permeability of the vaginal mucosa has been demonstrated in the vaginal absorption of a small molecule, like methanol, which has a vaginal membrane-controlled permeation, and a larger molecule, such as *n*-octanol, with vaginal permeation controlled by the hydrodynamic diffusion layer (Fig. 8). Further studies that used intact and ovariectomized monkeys could not establish any systematic relationship between the menstrual cycle and vaginal membrane permeability (68). Conflicting observations were also reported in the vaginal absorption of penicillin in humans (69, 70) as well as in rats (71, 72).

The vaginal permeability of a cycling monkey during the period immediately following menstruation is lower than that of a noncyclic rabbit (Fig. 9). The difference in vaginal permeability between rhesus monkeys and rabbits is greater for hydrophilic molecules, such as the short-chain alkanols (e.g., methanol), whose vaginal permeability is controlled by vaginal membrane permeation. The difference lessens as the alkyl chain length of alkanols increases, since molecular lipophilicity increases at the expense of hydrophilicity. At ovulation, the monkey's vaginal permeability is several-fold lower than that of the noncyclic rabbit (31).

The cyclic variation in vaginal drug permeability observed in rhesus monkeys in association with the rhythmic pattern of the sexual cycle suggests that the vaginal absorption data generated in the rhesus monkeys may be more reflective of what will occur in humans. The rhesus monkey is, therefore, a good animal model for the research and development of intravaginal delivery devices.

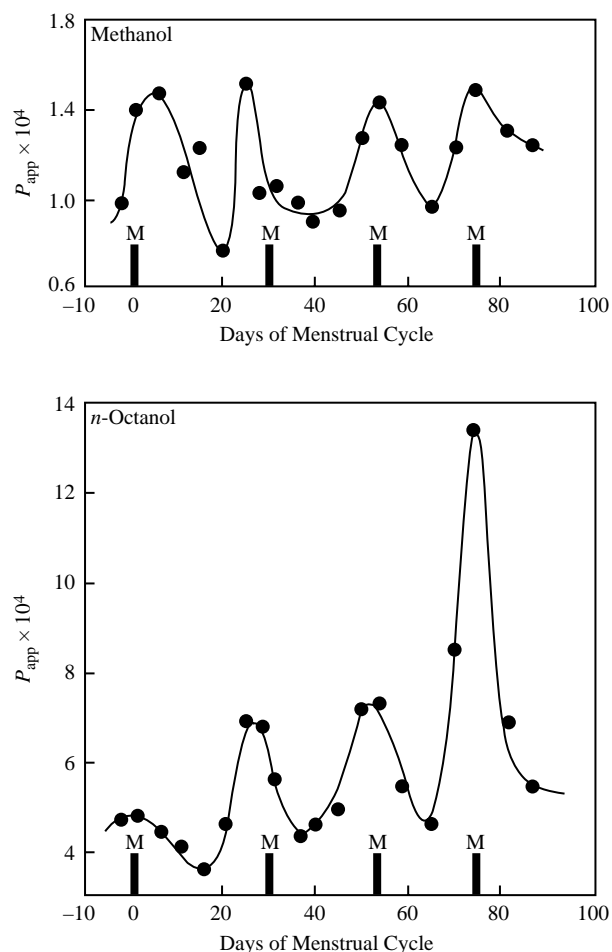


Fig. 8 Cyclic variation in the apparent permeability coefficient, P_{app} , of methanol and *n*-octanol in the female rhesus monkey in response to its estrus cycle. The bars indicate the time of observed menstruation. (Based on Ref. 31.)

TYPES OF VAGINAL DRUG DELIVERY SYSTEMS

Human Applications

Several publications have examined the many types of vaginal delivery systems already marketed or under development (73, 74). In the development of vaginal dosage forms, the following considerations should be addressed (73):

- Maintenance of an optimal pH for vaginal epithelium
- Ease of application
- Even distribution of drug
- Retention in the vagina
- Compatibility with coadministered medicines

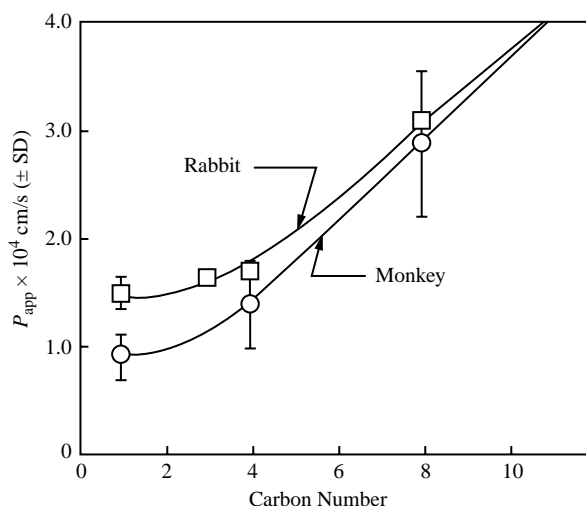


Fig. 9 Comparison of apparent permeability coefficients, P_{app} , for the vaginal absorption of straight-chain alkanols in noncyclic rabbits and cyclic rhesus monkeys. (Based on Refs. 16 and 165.)

In addition, offensive odors, staining, tissue irritation, or pain during sexual intercourse are undesirable. With regard to systemic delivery, one key advantage of intravaginal administration is avoidance of the presystemic elimination associated with oral dosage forms. The perineum venous plexus, which drains the vaginal tissue and rectum, flows into the pudenda vein and ultimately into the vena cava, which circumnavigates the liver on first-pass. This is in marked contrast to GI blood circulation that drains into the portal vein and passes directly through the liver before entering the systemic circulation. The vaginal route is preferable for drug entities associated with GI irritation and for the localized treatment of vaginal disorders that require minimal systemic absorption.

Creams, foams, and jellies

Many OTC vaginal products are available in these types of preparations. Common products include contraceptive creams, foams, gels, suppositories, sponges, and films, that contain an active spermicidal agent, like nonoxynol-9 or octoxynol. Vaginal preps are a recent FDA-approved OTC treatment for yeast infections (75). For treatment of menopausal symptoms, estrogen products are available as vaginal creams. A relatively new product for postmenopausal women, Replens®, is available as a vaginal bioadhesive moisturizer. It has been shown to be as safe and effective as estrogen vaginal creams in increasing vaginal moisture, fluid volume, elasticity, and returning the pH to the premenopausal state (76). Low pH lactate gels have been dispensed for the treatment of bacterial vaginosis. Application of these gels leads to a

disappearance of abnormal discharge and malodor, restores normal acidity, and facilitates recolonization with lactobacilli. Local treatment for vaginosis by the intravaginal delivery of antimicrobials may be preferred over an oral regimen, particularly during pregnancy (77).

When the efficacy of a vaginal ring, which delivers a continuous low dose of estradiol, was compared with a vaginal cream, both formulations were found to be equally effective and safe in the treatment of postmenopausal women with urogenital atrophy symptoms (53).

Vaginal rings

Vaginal rings provide a means of delivering a pharmacologically active agent to the systemic circulation at a controlled rate of release. The vaginal rings developed to date are primarily used for contraception and have been reviewed in the literature (74, 78). Compounds delivered include medroxyprogesterone acetate (MPA) (79–85), estradiol (53, 58, 86, 87), norgestrel (88, 89), levonorgestrel (90–94), combinations of progestins and estradiol (95), and combinations of progesterone and estrogens (96–101). Another area of interest is in the controlled delivery of prostaglandin for cervical ripening and induction of labor or pregnancy termination (102).

An advantage of intravaginal controlled drug administration over conventional oral administration is best illustrated in Fig. 10. After oral ingestion, MPA, reaches a peak plasma drug concentration rapidly within 2 h and declines over the next 22 h. On the other hand, intravaginal controlled delivery of MPA from a vaginal ring attains a steady plasma plateau within 4 h, which is maintained throughout the course of treatment until removal of the ring. The continuous “infusion” of drugs through the vaginal mucosa can prevent the possibility of hepatic GI first-pass metabolism and inefficient therapeutic activity resulting from the alternatively surging and ebbing plasma drug levels that occur with the intermittent use of oral dosage forms (31). Vaginal rings have been shown to be safe and effective for the delivery of estradiol and have been found to be more comfortable than a pessary (103). Local estradiol delivery via vaginal ring, vaginal cream, or suppositories, was reported to alleviate urogenital estrogen deficiency symptoms in postmenopausal women (53, 58). In both cases, patients showed a strong preference for the vaginal ring over other vaginal dosage forms. Danazol, that was administered intravaginally also via the vaginal ring was found to be very effective in the treatment of pelvic endometriosis (104). Danazol was absorbed through the vaginal mucosa and reached the deeply infiltrating endometriosis via diffusion.

Vaginal rings are made of biocompatible silicone elastomers that consist of a drug-free core ring and a

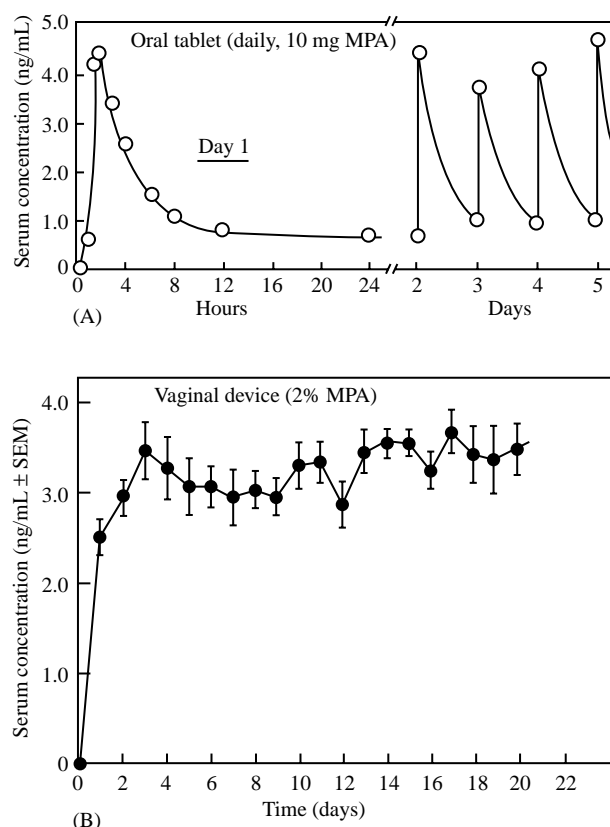


Fig. 10 A) Serum concentrations of medroxyprogesterone acetate after a daily oral administration of a medroxyprogesterone acetate tablet (10 mg) taken by a healthy woman before breakfast for five consecutive days. B) Daily serum concentrations of medroxyprogesterone acetate in women wearing medicated silicone vaginal rings (2% medroxyprogesterone acetate) for 20 days. (Based on Ref. 53.)

drug-containing coat. Vaginal rings are inserted and positioned around the cervix. Those designed for contraception, are kept intravaginally for 21 days and removed for 7 days to allow menstrual flow. The vaginal ring was redesigned due to frequent bleeding irregularities. The new generation of sandwich-type vaginal rings contains a drug-dispersed silicone polymer matrix is coated by a nonmedicated silicone polymeric membrane. The design reduces the initial spike of drug release frequently observed in the first treatment cycle of vaginal rings for contraception. The effect of the overcoat on the release rate profile of d-norgestrel is demonstrated in Fig. 11, which shows that the addition of an overcoat minimizes or eliminates the burst release of drug and shifts the non-zero-order drug release profile to the constant zero-order release rate profile. The concept of intravaginal dual

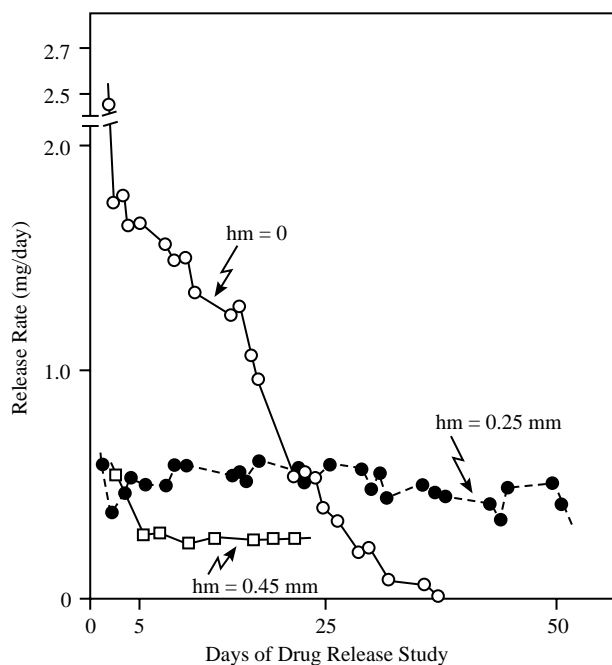


Fig. 11 Comparative in vitro release rate profile of levonorgestrel from a vaginal ring containing a homogeneous dispersion of drug in a silicone-based polymer matrix (open circles) and from one containing an inert overcoat covering the drug reservoir layer (closed circles and squares). The effect of overcoat thickness h_m on the release rate of levonorgestrel is also shown. (Based on Ref. 25.)

administration of progestin and estrogen in combination was recently extended to the development of a combined contraceptive vaginal ring. This new design (Fig. 12) is constructed from two drug reservoir compartments; the major compartment consists of a 3-Keto-desogestrel loaded core, and the other, minor compartment consists of a core loaded with a combination of 3-Keto-desogestrel and ethinylestradiol, a synthetic estrogen. These drug

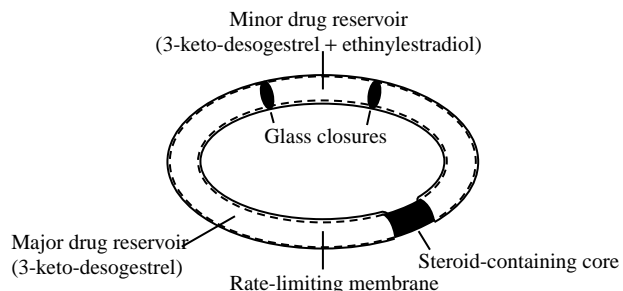


Fig. 12 Structural components of a contraceptive vaginal ring containing the combination of 3-keto-desogestrel and ethinylestradiol. (Based on Ref. 67.)

reservoir compartments are separated by two steroid-impermeable glass closures, as the partitions, and release the steroids at a fixed ratio through a rate limiting silicone membrane (98). Serum profiles of 3-Keto-desogestrel and ethinylestradiol from two prototype vaginal rings are shown in Fig. 13.

Veterinary Applications

Pessaries

Intravaginal drug delivery has been a useful tool in animal husbandry to control the estrus cycle of sheep and cattle. Synchronization of a herd estrus cycle eases the management strains on ranchers and farmers. Fluorogestone acetate, an effective ovulation inhibitor in the ewe, has been formulated into a vaginal pessary for sheep. The pessary is placed in the ewe for about 15 days. The pessary is removed and the sheep regain their estrus and ovulate within 2–4 days. This permits insemination of all sheep within a 2-day period and with a high success rate. Commercial delivery systems include Syncro–Mate vaginal pessary (S.D. Searle & Co.), Chronogest vaginal pessary (Intervet S.A.), and PRID vaginal insert for dairy cattle (Sanofi Animal Health Ltd.), which contains progesterone and estradiol benzoate. A MPA sponge has been used to evaluate the preovulatory gonadotropin surge in ewes (105).

Modified vaginal pessaries

The Syncro–Mate progestin-releasing vaginal pessary is fabricated by dispersing a progestin, such as fluorogestone acetate, in a pessary made of porous polyurethane sponge. The pessary can be readily inserted intravaginally and removed according to a predetermined schedule. The vaginal pessary has been redesigned with the aim of minimizing the loading dose, to overcome the matrix-type release and absorption profiles, and to improve systemic bioavailability. Work evaluating a cylindrical, drug-free polyurethane vaginal sponge coated by a laminate of fluorogestone-containing silicone matrix and a drug-free silicone coating membrane was conducted (106–108).

The system itself was modified and the release rates studied in vitro and in vivo. It was found that the drug-containing silicone layer needed to be in contact with the vaginal wall. A silicone coating did not result in adequate release rate. Also, as the surface area of the drug-containing silicone increased, the drug delivery rate increased. This effort has resulted in two new pessary designs (Fig. 14). Both types make use of the polyurethane sponge in the vaginal pessary as the mechanical support for vaginal insertion and retention, but

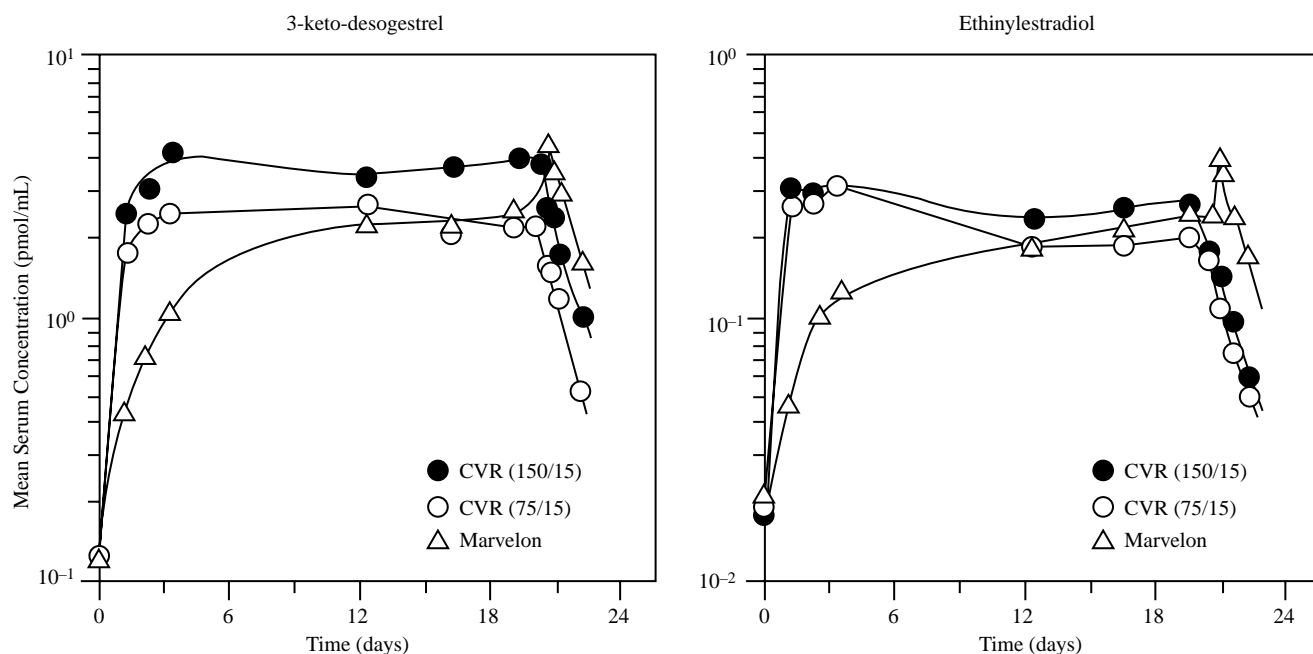


Fig. 13 Mean serum profiles of 3 keto-desogestrel and ethinylestradiol after a 21-day continuous intravaginal administration of two prototype combination contraceptive vaginal rings (CVR). The profiles from an oral combination tablet (Marvelon) is plotted for comparison. (Based on Ref. 67.)

the drug reservoir is relocated from the porous sponge matrix to a sheet-type rate-controlled silicone device that

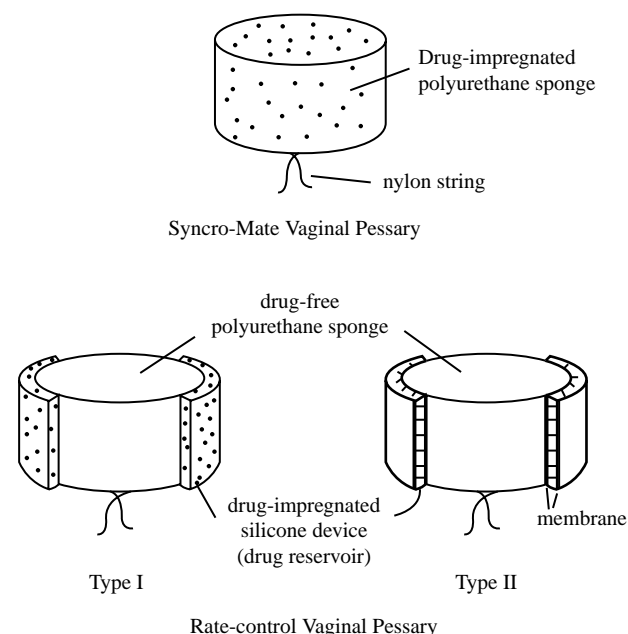


Fig. 14 Comparison of newer progestin-releasing vaginal pessaries (Type I and II) with an older (Syncro-Mate) pessary design. (Based on Ref. 25.)

covers the circumferential surface of the sponge. The type I rate-controlled silicone device consists of a homogeneous dispersion of drug in a silicone polymer matrix. The type II drug-dispersing polymer matrix is sandwiched between two sheets of silicone polymer membrane to form a three layered laminate.

Potential Developments in the Future: Bioadhesives

Novel intravaginal delivery systems include those that employ bioadhesive materials. The bioadhesive properties of compounds such as hydrogels may provide a controlled delivery system with a prolonged residence and intimate contact in the vagina. Many hydrophilic polymers and hydrogels have been used in vaginal products. These include starch, collagen, proteins, gelatin, and cellulose derivatives (hydroxypropyl methylcellulose, hydroxypropyl cellulose, and sodium carboxymethylcellulose) (77). Two synthetic hydrogels that have been reported in applications for vaginal drug delivery are poly(ethylene oxide) and poly(acrylic acid). Bioadhesion is thought to involve an initial interaction of the hydrogel with the mucosal surface, which requires a matching of the polarity between the tissue surface and the polymer surface (109), and subsequently, an interpenetration of the mucosal surface by the polymer chains of the hydrogel.

A review of vaginal bioadhesive formulations indicates that bioadhesive tablets have been used for localized treatment of diseases in the vaginal tissue (77, 109). For example, Bleomycin, an antitumor agent, was incorporated into a flat-faced disk fabricated from a combination of hydroxypropyl cellulose and poly(acrylic acid) (Carbopol 934) (110). The tablet was designed to release Bleomycin at a slow rate to minimize irritation to healthy mucosa. Another vaginal tablet is formulated from the combination of poly(acrylic acid) with hydroxypropyl methylcellulose and ethylcellulose (111). Other polymer combinations evaluated for potential bioadhesive vaginal delivery include poly(acrylic acid) and sodium carboxymethyl cellulose with Avicel PH102 (methylcellulose) as the diluent (112). Insulin has been formulated in a cross-linked poly(acrylic acid) gel and found to be adsorbed onto the vaginal surface of alloxan-induced diabetic rats and rabbits (113).

Emulsion-based formulations with bioadhesive properties also have been designed to deliver antifungal agents, although little has been reported in the literature (78). Bioadhesive microparticles used in nasal and oral delivery have the potential for further development of intravaginal delivery systems. The mucoadhesive benzyl ester of hyaluronic acid has been used in preparing microspheres for the intravaginal delivery of salmon calcitonin to rats (114). Replens[®], which has been marketed as a bioadhesive moisturizer and which remains in the vagina for 2–3 days, consists of a bioadhesive cross-linked polycarboxophil (115). Carbopol 934 polymer could be a good bioadhesive candidate for clinical application in the intravaginal delivery of spermicidal agents (44).

Liposomes, a novel drug delivery system, are widely applied in the topical treatment of diseases, including vaginal diseases. Their applications in contraceptive systems for the intravaginal administration of progesterone (116) and interferon- α (117) (or metronidazole) (118) for the genital papilloma virus infections were previously reported.

DRUG CANDIDATES FOR VAGINAL ADMINISTRATION

Antimicrobials

Antimicrobials, along with antifungals, have been important drug candidates for the treatment of vaginal diseases, such as bacterial vaginosis. Earlier researches on antimicrobials, such as penicillin and sulfanilamide, has been reviewed (22). Evaluations have been made on such drugs as metronidazole (119–121), clindamycin

(122, 123), feticonazole (124), and clotrimazole (124, 125). Metronidazole has been prescribed for the treatment of amoebiasis, trichomoniasis, lambliasis, and anaerobic infection (121). It is administered vaginally as a gel, normally twice a day every 12 h for 7 days for the treatment of bacterial vaginosis. The gel results in lower serum concentrations with negligible adverse systemic effect (119). Metronidazole's low lipid solubility probably contributes to its poor vaginal absorption (120). A comparison of vaginal and oral drug delivery of metronidazole from a film-coated tablet showed that the maximum serum concentrations attained by vaginal delivery are only sufficient to kill the most susceptible anaerobes (121). One recent study confirmed that liposomes could be used as a novel delivery system for metronidazole in the treatment of vaginal infections (118).

Clindamycin also is indicated for the treatment of bacterial vaginosis. Its efficacy as a 2% cream is similar to oral metronidazole treatment (122). Measurable amounts of clindamycin, corresponding to frequency of application, have been detected systemically at levels well below those of intravenous delivery (123). Gentamicin is an additional aminoglycoside that was evaluated for vaginal administration in ovariectomized rats (126). The results indicated the following order of absorption enhancers by their effectiveness: 1% palmitoylcarnitine > 0.5% lysophosphatidylcholine > 1% laureth-9 > 10% citric acid. Severity of desquamation of the vaginal epithelium were scored as lysophosphatidylcholine = laureth-9 > palmitoylcarnitine > citric acid.

The market for vaginal candidiasis treatments has been revamped in recent years due to the availability of OTC products. Recently, oral antifungal agents, i.e., ketoconazole, have been approved for treatment of vaginal candidiasis. Earlier research that compared vaginal feticonazole to oral clotrimazole (125) and vaginal clotrimazole to oral fluconazole (100) have found them equally effective and well tolerated as compared to the oral product used for comparison. An important consideration in the choice of oral vs. vaginal treatment is the contraindication of oral dosage forms during pregnancy. Vaginal treatment for recurrent vaginal conditions may be a good alternative in such a case (127).

Anticancer Agents

Intravaginal investigations into the local treatment of cervicovaginal cancers have made some progress over recent years. [³H]-Feticonazole has been evaluated in normal cervicovaginal mucosa, cervical carcinoma, and relapsing vulvovaginal candidiasis, and was found to be devoid of risk to patients (128). For patients with vaginal

rhabdomyosarcoma and residual vaginal disease following surgery and chemotherapy, the effects of high-dose irradiation from vaginal molds loaded with iridium¹⁹² were examined (129). Individualized molds were made with rapid setting silastic foam and loaded with iridium wires. The patients treated remained well and disease-free for 7 years. The pharmacokinetics properties of ciclopirox olamine have been evaluated after vaginal application to rabbits and patients. The intravaginal absorption was found to be low while the penetration of drug was found in deep tissue. These results led to good local and systemic tolerability (130). Cisplatin suppositories have been tested preoperatively via vaginal administration (131). The Cisplatin penetration of the tumor surface needs further improvement in order to achieve effective local chemotherapy.

Contrary to cancer treatment, cervicovaginal cancers are linked to vaginal pessary use. Review of cancer cases found that tumors occurred at sites of ring insertion. Although chemical carcinogenesis cannot be ruled out, chronic local infection may be the main etiologic factor. The severity of cancer cases has declined in recent years due to the availability of more advanced surgical and radiation treatments with minimal complications (132).

Prostaglandins

For many years, prostaglandins, primarily prostaglandin E₂ (PGE₂), have been studied for cervical ripening and induction of labor as well as an abortifacient. Most of the prostaglandin delivery systems have been in the form of suppositories or pessaries (133–143). More recent work with pessaries and gels has been done due to the use of prostaglandins as an alternative to surgical abortions.

Early work with PGE₂ formulated a sustained-release pessary that used a swelled, cross-linked polymer hydrogel (143). Further work was conducted using a semicrystalline poly(ethylene oxide) and poly(urethane) hydrogel swollen with PGE₂ solution (144). The controlled release of PGE₂ and the cervical score showed a linear release, suggesting the advantage of control over cervical ripening (145). Pessaries provided greater control in labor induction (146), (more favorable than intracervical gels (147)), had high efficacy and low incidence of side effects (148), and showed a reduction in cesarean rates associated with artificial rupture caused by oxytocin infusion (149). Although PGE₂ has been shown to initiate active labor and reduce the need for oxytocin, it has been shown to cause uterine hyperstimulation or fetal heart rate abnormalities, which are reversed with removal of the pessary (150). Tablet forms of PGE₂ are shown to be more chemically stable and cost effective (151, 152). Additionally, the case

of use may allow for shortening of postdate pregnancies in a safe and effective manner on an outpatient basis (153).

In addition, PGE₂ has been formulated as a gel. A comparison of tablets and a triacetin-base gel showed more favorable induction with the gel (154). The use of PGE₂ gel in women with prelabor spontaneous rupture of membranes, significantly improved the time of delivery without influencing the cesarean section rate or fetal-maternal infective morbidity (155). The intravaginally administered gel was more efficacious than that given intracervically (156). In regards to a regimen, either 12 h or 6-hourly, the majority of 12-hourly subjects achieved labor after a single dose, but the induction delivery interval was similar in both groups (157).

Due to its labor-inducing characteristics, prostaglandins also have been used as abortifacients. PGE₁, as a vaginal pessary, has been combined with the oral antiprogesterin, mifepristone (RU486), and has been found to be a safe, efficient nonsurgical outpatient method of termination (158). The PGE₂ product, originally intended for the treatment of GI ulcers caused by nonsteroidal antiinflammatories, was shown to be a safe and low-cost method. Misoprostol alone has been used for the induction of labor with females with late fetal death (159). Attention of the medical profession and the news media has resulted in the use of misoprostol and methotrexate for early abortion. In this application, methotrexate, originally used as a chemotherapeutic agent in cancer and arthritis, is given intramuscularly followed by intravaginal misoprostol 3 days later. This is an effective abortion method for a gestation period of 56 days or less (160).

Spermicides

Spermicidal activity in the vagina is intended for fertility control, by eliminating the motility of sperm and ultimately killing them. Spermicides have become more popular with the rise in social awareness and prevention of sexually transmitted diseases. Many compounds have been evaluated for spermicidal activity. As mentioned, nonoxynol-9 and octoxynol are predominantly available in the United States. A few products that have been evaluated include alkyloxynol-741, which was tested in stump-tailed macaques for spermicidal activity, and gramicidine, which is used as a spermicidal agent in Russia but not in the United States. Alkyloxynol-741 was compared with nonoxynol-9 and chlorhexidine, using a dissolvable polyvinyl alcohol film, and found to be an inexpensive alternative in countries where nonoxynol-9 may not be readily available (161). The combination of nonoxynol-9 and EDTA in a gel formulation, developed from carbopol 934 polymer showed a significant

enhancement of efficacy in fertility control (44). A tablet form drug delivery system fabricated by incorporating nonoxynol-9 into polyvinylpyrrolidone, was reported to provide a short- and long-term release of nonoxynol-9 and produce an immediate and extended enhancement of the contraceptive properties (162). Protectaid, a contraceptive sponge marketed in Canada, contains sodium cholate, nonoxynol-9, and benzalkonium chloride. Sodium cholate itself has been shown to exhibit strong spermicidal and antiviral activity, and offers a new and modern protective method (2828 163). A serine protease from sperm, 44-acetamidophenyl-4-guanidinobenzoate, inhibits acrosin activity and has been found to be more potent and less irritating than nonoxynol-9, as well as providing protection against HIV (164).

Various studies found that the absorption of nonoxynol-9 through the vaginal membrane was very slow and suggested a dependence of molecular weight of the oligomeric components that make up the spermicide (165, 166). Permeation studies have found the hydrophilic–lipophilic balance (HLB) of the oligomers to play a role as well (36). In delivery gel made of calcium chloride cross-linked alginate containing 3% of nonoxynol-9, it was found that pH of delivery gel had a significant effect on spermicidal efficacy of nonoxynol-9 (167). This behavior could be attributed to nonoxynol-9 micelle formation.

Steroids

The vaginal delivery of steroids for urogenital symptoms has been shown to be more appropriate than oral and parenteral administration. Hormone replacement therapy is indicated for peri- and postmenopausal women suffering from vasomotor symptoms, vaginal dryness, and discomfort due to urogenital atrophy and other related symptoms of hormone deficiency. Steroids, progesterone and estrogen, have been used for many years to treat a variety of physiologic conditions, including hormonal replacement and contraception. As an alternative to oral estrogen replacement, vaginal estrogen cream is an effective treatment for vaginal atrophy (168). Use of estradiol cream has been effective for atrophic vaginitis (53) and also for endometrial proliferation and hyperplasia if coadministered with progestins (169). Systemic absorption of progestins could have the risk of endometrial hyperplasia.

Like estrogen, progesterone has been delivered via the vagina as creams, pessaries, and vaginal rings. Vaginal absorption and local redistribution of progesterone was observed in a study using young female pigs (170). Vaginal delivery has shown enhanced progesterone delivery to the uterus when compared with a standard

intramuscular regimen (171). Micronized progesterone in a nonliquefying vaginal cream is promising (172, 173). For systemic delivery of progesterone to the genital organs, doctors have prescribed to their patients a suppository compound of progesterone and cocoa butter. A lactose-based progesterone tablet has been designed to deliver biologically effective amounts of progesterone for up to 48 h (77). These tablets form a milky suspension and stay resident in the vagina for a longer period of time, making them ideal for the treatment of menstrual irregularities, functional uterine bleeding, luteal phase defects, premenstrual tension, infertility, and osteoporosis. A recent study demonstrated that a “first uterine pass effect” occurred when progesterone was delivered intravaginally, thereby providing an explanation for the unexpectedly high uterine concentrations relative to the low serum concentration observed (174). Application of liposomes as the vaginal delivery system for progesterone to achieve contraceptive efficacy was successfully demonstrated (116). The combined use of progesterone delivered from an intravaginal-targeted drug delivery system and estradiol delivered from a transdermal therapeutic system were found to be very effective in producing artificial cycles (95).

Absorption and secretion characteristics of sodium prasterone sulfate have been evaluated in rats after vaginal administration. The absorption was significantly affected by the estrus cycle and the progression of the pregnancy (175).

Proteins and Peptides

Research on the intravaginal delivery of peptides and proteins has focused on insulin and gonadotropin-releasing hormones (GnRH) and their absorption into the systemic circulation. Insulin and thyroid-stimulating hormone have been shown to achieve some absorption from the vagina of rats and rabbits, but with a high dependency on the estrus cycle (113). Hydrophilic molecules, like insulin and GnRH, may be absorbed through intercellular channels, hence absorption would be greater when the epithelium is thinner (77, 176). As commonly observed with protein and peptide research, an enhancer is often needed to assist in absorption. The enhancers are limited in human testing; thus, most of the preliminary research must be done in animal models, such as the rat and rabbit (77). Vaginal administration of insulin was found to increase hypoglycemia in rats when using enhancers such as Na taurodihydrofusidate, polyoxyethylene-9-lauryl ether, lysophosphatidylglycerol, lysophosphatidyl choline, and palmitoylcarnitine chloride (177). Other therapeutic agents that have been investigated include leucine enkephalin (66), salmon calcitonin (115, 178), and

recombinant human relaxin (179, 180). Relaxin, structurally related to insulin, was formulated as a 3% methylcellulose gel for intravaginal delivery. In this form, it had limited permeability through nonpregnant rabbit and rhesus monkey vaginas.

Much work has been done using luteinizing hormone releasing hormones (LHRH) analogs, particularly leuprolide. Initial work with leuprolide found greater potency in rats via vaginal administration over rectal, nasal, and oral administration (181). Enhancement of absorption by organic acids (citric, succinic, tartaric, and glycocholic) increased bioavailability by 20%. The vaginal absorption from jellies was found to be a practical dosage form. Additional investigation of the organic acids used for enhancement was found to work by the acidifying and chelating ability (182). Citric acid also was shown to loosen the blood–vaginal epithelium barrier. The down regulation of the pituitary by chronic intravaginal treatment of leuprolide exhibited regression of hormone-dependent mammary tumors in rats (183–185). The vaginal route appears to be the preferred route according to a recent presystemic metabolism study of first-order LHRH degradation in rabbit homogenates (186). The degradation half-life of vaginal homogenates was 9–12 times longer than that of rectal homogenates and 3–4 times longer than that of nasal homogenates.

Vaccines, Antigens, and Gene Delivery

Factors that have a significant impact on immune responses are the route and method of vaccine or antibody delivery (187). Since the recognition of acquired immunodeficiency syndrome (AIDS) in the early 1980s, researchers have focused on the discovery of pharmacological agents for the treatment and ultimate cure of AIDS, as well as on preventing the spread of the virus. Mucosal infection via the vagina and rectum are reportedly two of the major pathways through which HIV and other sexually transmitting viruses are disseminated (188, 189). Mucosal immunity has been considered as the first line of immunological defense against these pathogens to prevent the systemic infection. The important factors to be considered in the development of vaccines are protection with a minimum number of administrations and a practicability of the approaches in inducing immunity at the mucosal surface (81, 190). Consequently, studies were conducted to elucidate the effects of variations in the routes of immunizations on the type and extent of mucosal response (191). Mucosal immunity in the female reproductive tract was reportedly influenced by immunoglobulins (Igs), cytokines, and reproductive hormones. The types of responses elicited following a DNA

immunization were found to depend on both the identity of the antigen and the route of DNA administration. The epidermal delivery route, including vaginal mucosa, was found to be more efficient in terms of dosage requirements (192, 193). When the effects of exogenous hormones on reproductive tract immunity were evaluated in women on oral contraceptive pills (OCPs), the mean values of IgA in the cervical mucus of women on OCPs were much greater than those in the naturally cycling women (194). The increased levels of IgA in the cervical mucus of women on OCPs may contribute to a lower incidence of sexually transmitted diseases. Using animal models, the transfection of mucosal tissues by gene–gun administered plasmids was demonstrated *in vivo*, and vaginal immunization of human growth hormone was noted to yield a higher titer of cervicovaginal antibodies than other routes of immunization (88).

DNA-based vaccines have been recently developed for immunization to overcome the deficiencies of antigen-based vaccines. The major sites of delivery for DNA-based immunization include the oral, nasal, rectal, and vaginal mucosae (195, 196). The ability of a hepatitis B surface antigen-encoding plasmid to induce responses in mice through the various routes of delivery was compared (197). It was reported that delivery through the vaginal route still produced a cytotoxic T lymphocyte activity, even though it failed to induce antibodies. A human simian virus 1 (HSV) vaccine was tested as an aqueous solution or gel intranasally, vaginally, and subcutaneously in guinea pigs (198). The gel system used was a controlled release carbopol gel. The animals were challenged 3–5 weeks later with only the subcutaneous response producing IgG and IgA. The nasal and vaginal routes showed that the vaccine could take up and elicit antibodies, thereby slightly reducing the severity of the disease, but showed no superiority to the subcutaneous route. In another study, rats were immunized with a synthetic peptide from a HIV envelope glycoprotein and shown to have greater IgG and IgA response with an enhancer, lysophosphatidyl glycerol (199, 200). The serum antibodies from subcutaneous and intravaginal delivery were able to recognize the glycoprotein (HIV 1 gp120), but no neutralizing activity against the virus was seen. An antigen delivery system of lysophosphatidylcholine and degradable starch microspheres demonstrated potential intravaginal delivery to sheep (199, 200). If the vagina is capable of mounting an immune response, antibodies in genital secretions may be able to reduce the transmission of HIV. Intravaginally administered tracers using (fluorescein isothiocyanate) (FITC) bovine albumin, FITC-horseradish peroxidase, and FITC-horse ferritin have shown the vagina and cervix to be major sites of protein uptake (201).

Lactobacilli from the female genital tract have been developed as a vehicle to deliver continued doses of foreign antigen to the vaginal mucosa surface with the aim of stimulating a local immune response (202). The lactobacillus fermentum was chosen for genetic manipulation and delivered intravaginally in guinea pigs and was maintained for 5 days. Despite the short period, the novel vaccination approach shows potential for stimulating mucosa immunization. Horse ferritin was combined with aluminum hydroxide, muramyl dipeptide, monophosphoryl lipid A, dimethyl dioctadecyl ammonium, and cholera toxin (203). The aluminum hydroxide combination was found to be the most effective. However, the doses of antigen used were larger than normal, and consequently the drug combination may be inefficient at more realistic dosages. Additional work with mouse ferritin has shown pelvic immunization at nonmucosal sites to be very effective in stimulating an IgA response in the female reproductive tract, more so than intravaginal delivery because of possible involvement of iliac lymph nodes not associated with the vagina (204).

As a potential means of treatment of gynecological conditions, *Candida albicans* vaccine was developed and an antibody response after intravaginal application of vaccine and effects on recurrent urinary tract infections were evaluated (205, 206). Serum antibody to some of non-*E. coli*, but no antibody to *E. coli* were expressed (207). The vaccine was well tolerated and invites further development. Another evaluated vaccines contained inactivated polio vaccine delivered by intravaginal, intrauterine, mesopharyngeal, and intramuscular routes. The predominant secretory antibodies to polio virus in the vagina were found to be IgA and IgG in the uterus. The genital tract is immunologically reactive and may play a role in protection against other infections such as gonococcus and genital herpes (208). For veterinary use, dairy cattle research has directed attention to bovine herpes virus type 2, which causes ulcerative lesions in teats and udders. Vaccinations subcutaneously and intravaginally have been shown to be beneficial in reducing the severity of the infection and show potential for a vaccine for dairy cattle (209).

Anti-inflammatory Agents and Others

Bromocriptine has been used in the treatment of hyperprolactinaemic women (210). Drug in tablet form was administered intravaginally and found well absorbed from the vagina, with avoidance of side effects. For Benzydamine, a nonsteroidal anti-inflammatory with local anesthetic and analgesic properties, mouthwash, dermal cream, and vaginal douche have been formulated and evaluated for local therapy (211). Another chemical entity

that has been researched for intravaginal delivery is tranexamic acid, an antifibrinolytic drug usually given by mouth for the treatment of menorrhagia associated with intrauterine devices (IUD). The intravaginal delivery of tranexamic acid was found well tolerated with avoidance of GI side effects (212).

REFERENCES

1. Woolfson, A.; Malcolm, R.K.; Gallagher, R. Drug Delivery by the Intravaginal Route. *Crit Rev Ther Drug Carrier Syst* **2000**, *17* (15), 509–555.
2. Platzer, W.; Poisel, W.; Hafez, E.S.E. Functional Anatomy of the Human Vagina. *The Human Vagina*; Hafez, E.S.E., Evans, T.N., Eds.; North Holland Publishing Co.: New York, 1978; 39–53.
3. Kistner, R. Physiology of the Vagina. *The Human Vagina*; North Holland Publishing Co.: New York, 1978; 109–120.
4. Burgos, M.; Roig de Varnas-Linares, C. Ultrastructure of the Vaginal Mucosa. *The Human Vagina*; North Holland Publishing Co.: New York; 1978, 63–93.
5. Averette, H.; Weinstein, G.; Frost, P. Autoradiographic Analysis of Cell Proliferation Kinetics in Human Genital Tissues. I. Normal Cervix and Vagina. *Amer. J. Obstet. Gynec.* **1970**, *108*, 8–17.
6. Wagner, G.; Levin, R. Vaginal Fluid. *The Human Vagina*; North Holland Publishing Co.: New York, 1978; 121–137.
7. Itoh, Y. Analysis of Human Vaginal Secretions Using Rabbit Anti-Human Vaginal Secretions: Detection of Non-serum Proteins. *Nippon Hoigaku Zasshi* **1990**, *44*, 267–271.
8. Itoh, Y.; Furuhashi, A.; Sato, Y. Immunohistochemical Studies on the Localization of Non-Serum Proteins Detectable in Human Vaginal Secretions. *Nippon Hoigaku Zasshi* **1991**, *45*, 26–29.
9. Schmidt, E.; Beller, F. Biochemistry of the Vagina. *The Human Vagina*; North Holland Publishing Co.: New York, 1978; 139–149.
10. Moghissi, K. Composition and Function of Cervical Secretions. *Handbook of Physiology*; American Physiological Society: Washington D.C., 1978; II, 25–48.
11. Richardson, J.; Illum, L. Routes of Delivery: Case Studies, the Vaginal Route of Peptide and Protein Drug Delivery. *Adv. Drug Delivery Rev.* **1992**, *8*, 341–366.
12. Sparks, R.; Purrier, B.; Watt, P.; Elstein, M. The Bacteriology of the Cervix and Uterus. *Br. J. Obstet. Gynecol.* **1977**, *84*, 701–704.
13. Masters, W.; Johnson, V. *Human Sexual Response*; Little Brown, and Co.: Boston, 1966.
14. Steger, R.; Hafez, E. Age-Associated Changes in the Vagina. *The Human Vagina*; North Holland Publishing Co.: New York, 1978; 95–106.
15. Brown, W. Microbial Ecology of the Normal Vagina. *The Human Vagina*; North Holland Publishing Co.: New York, 1978; 407–422.
16. Macht, D. On the Absorption of Drugs and Poisons through the Vagina. *J. Pharmacol. Exp. Ther.* **1918**, *10*, 509–521.

17. Robinson, G. Absorption from the Vagina. *J. Pharmacol. Exp. Ther.* **1927**, 32, 81–88.
18. Macht, D. Concerning the Absorption of Quinin and Oxyquinolin Sulphate through the Vagina. *J. Pharmacol. Exp. Ther.* **1928**, 34, 137–145.
19. Millman, N.; Hartman, C.; Stavorski, J.; Botti, J. Unidirectional Permeability of the Vaginal Wall of the Rabbit. *Fed. Proc.* **1950**, 9, 89.
20. Hartman, C. The Permeability of the Vagina Mucosa. *Ann. N.Y. Acad. Sci.* **1959**, 83, 313–327.
21. El-Sheikha, A.; Hafez, E. Absorption of Drugs and Hormones in the Vagina. *The Human Vagina*; North Holland Publishing Co.: New York, 1978; 179–191.
22. Benzinger, D.; Edelson, J. Absorption from the Vagina. *J. Drug Metab. Rev.* **1983**, 14, 137–168.
23. Rosenzweig, M.; Walzer, M. Absorption of Protein from the Vagina and Uterine Cervix. *Am. J. Obstet. Gynecol.* **1943**, 45, 286–290.
24. Cunningham, F.; Kraus, D.; Brubaker, L.; Fischer, J. The Bacteriology of the Cervix and Uterus. *J. Clin. Pharmacol.* **1994**, 34, 1060–1065.
25. Chien, Y. Mucosal Drug Delivery. *Novel Drug Delivery Systems*; Marcel Dekker, Inc.: New York, 1992; 197–228.
26. Hwang, S.; Owada, E.; Yotsuyanagi, T.; Suhardja, L.; Ho, H.; Flynn, G.; Higuchi, W. Systems Approach to Vaginal Delivery of Drugs II: In Situ Vaginal Absorption of Unbranched Aliphatic Alcohols. *J. Pharm. Sci.* **1976**, 65, 1574–1578.
27. Hwang, S.; Owada, E.; Suhardja, L.; Ho, H.; Flynn, G.; Higuchi, W. Systems Approach to Vaginal Delivery of Drugs V: In Situ Vaginal Absorption of l-Alkanoic Acids. *J. Pharm. Sci.* **1977**, 66, 781–784.
28. Bengtsson, L.; *Acta Endocrinol* **1953**, 13, 11.
29. Yotsuyanagi, T.; Molokhia, A.; Hwang, S.; Ho, N.; Flynn, G.; Higuchi, W. Systems Approach to Vaginal Delivery of Drugs I: Development of In Situ Vaginal Drug Absorption Procedure. *J. Pharm. Sci.* **1975**, 64, 71.
30. Hwang, S.; Owada, E.; Suhardja, L.; Ho, H.; Flynn, G.; Higuchi, W. Systems Approach to Vaginal Delivery of Drugs IV: Methodology for Determination of Membrane Surface pH. *J. Pharm. Sci.* **1977**, 66, 778–781.
31. Flynn, G.; Ho, N.; Hwang, S.; Owada, E.; Molokhia, A.; Behl, C.; Higuchi, W.; Yotsuyanagi, T.; Shah, Y.; Park, J. Interfacing Matrix Release and Membrane Absorption Analysis of Steroid Absorption from a Vaginal Device in the Rabbit Doe. *Controlled Release of Polymeric Formulations*; American Chemical Society: Washington DC, 1976; 87–122.
32. Ho, H.; Suhardja, L.; Hwang, S.; Owada, E.; Molokhia, A.; Flynn, G.; Higuchi, W.; Park, J. Systems Approach to Vaginal Delivery of Drugs III: Simulation Studies Interfacing Steroid Release from Silicone Matrix and Vaginal Absorption in Rabbits. *J. Pharm. Sci.* **1976**, 65, 1578.
33. Corbo, D.; Liu, J.; Chien, Y. Characterization of the Barrier Properties of Mucosal Membranes. *J. Pharm. Sci.* **1990**, 79, 202–206.
34. Rojanasaku, Y.; Wang, L.; Bhat, M.; Glover, D.; Malanga, C.; Ma, J.; *Pharm. Research* **1992**, 9, 1029–1034.
35. Sanders, J.; Matthews, H. Vaginal Absorbance of Polyvinyl Alcohol in Fisher-344 Rats. *Human Exp. Toxicol.* **1990**, 9, 71–77.
36. Yu, K.; Chien, Y. Spermicidal Activity-Structure Relationship of Nonoxynol Oligomers: Physicochemical Basis. *Int. J. Pharm.* **1995**, 125, 81–90.
37. Chien, Y.; Mares, S.; Berg, J.; Huber, S.; Lambert, H.; King, K. Controlled Drug Release from Polymeric Delivery Devices. III: In Vitro–In Vivo Correlation of Intravaginal Release of Ethynodiol Diacetate from Silicone Devices in Rabbits. *J. Pharm. Sci.* **1975**, 64, 1776.
38. Moghissi, K. Composition and Function of Cervical Secretion. *Handbook of Physiology*; American Physiological Society: Washington D.C., 1973; 25–48.
39. Eriksen, G.; Carlstedt, I.; Uldbjerg, N.; Ernst, E. Cervical Mucins Affect the Motility of Human Spermatozoa In Vitro. *Fertil. Steril.* **1998**, 70, 350–354.
40. Morales, P.; Roco, M.; Vigil, P. Human Cervical Mucus: Relationship between Biochemical Characteristics and Ability to Allow Migration of Spermatozoa. *Human Reprod.* **1993**, 8, 78–83.
41. Wolf, D.; Sokoloski, J.; Khan, M.; Litt, M. Human Cervical Mucus. III. Isolation and Characterization of Rheological Active Mucin. *Fertil. Steril.* **1977**, 28, 53–58.
42. Pigman, W. *The Glycoconjugate*; Academic Press: NY, 1977; 1, 1.
43. Murata, L.; Dodson, M. The Herpes Simplex Virus Type 1 Origin-binding Protein. Sequence-specific Activation of Adenosine Triphosphatase Activity by a Double-stranded DNA Containing Box I. *J. Biol. Chem.* **1999**, 274 (52), 37079–37086.
44. Lee, C.; Bagdon, R.; Bhatt, P.; Chien, Y. Development of Silicone-based Barrier Devices for Controlled Release of Spermicidal Agent. *J. Cont. Rel.* **1997**, 43, 283–290.
45. Silberberg, A.; Meyer, F. *Mucus in Health and Disease II*; Chantter, E.; Elder, J.; Eionstein, M., Eds.; 1982; 53–74.
46. Sheehan, J.; Carlstedt, I. Hydrodynamic Properties of Human Cervical Mucus Glycoproteins in 6M-Guanidinium Chloride. *Biochem. J.* **1984**, 217, 93–101.
47. Blasco, L. Clinical Approach to the Evaluation of Sperm-cervical Mucus Interactions. *Fertil. Steril.* **1977**, 28, 1133.
48. Lee, C.; Anderson, M.; Chien, Y. Characterization of In Vitro Spermicidal Effect of Chelating Agent against Human Sperm. *J. Pharm. Sci.* **1996**, 85, 649–654.
49. Vigil, P.; Perez, A.; Neira, J.; Morales, P. Post-partum Cervical Mucus: Biological and Rheological Properties. *Human Reprod.* **1991**, 6, 475–479.
50. Lundholm, P.; Asakura, Y.; Hinkula, J.; Lucht, E.; Wahren, B. Induction of Mucosal IgA by a Novel Jet Delivery Technique for HIV-1 DNA. *Vaccine* **1999**, 17, 2036–2042.
51. Friend, D. Plasma Membrane Diversity in a Highly Polarized Cell. *J. Cell Biol.* **1982**, 93, 243–249.
42. Yanagimachi, R. Sperm-Egg Fusion. *Current Topics in Membranes and Transport*; Academic Press: NY, 1988; 32, 3–43.
53. Barentsen, R.; Peter, H.; Van de Weijer, H.; Schram, J. Continuous Low Dose Estradiol Released from a Vaginal Ring vs Estriol Vaginal Cream for Urogenital Atrophy. *Eur. J. Ob. Gyn. Rep. Biol.* **1997**, 71, 73–80.
54. Jovov, B.; Lewis, S.; Crowe, W.; Berg, J.; Wills, N. Role of Intracellular CA⁺⁺ in Modulation of Tight Junction Resistance in A6 cells. *Am. J. Physiol.* **1994**, 266, F775–784.

55. Gorodeski, G.; Jin, W.; Hopper, U. Extracellular Ca^{++} Directly Regulates Tight Junctional Permeability in the Human Cervical Cell Line CaSki. *Am. J. Physiol.* **1997**, *272*, C511–524.
56. Platz-Christensen, J.; Pernevi, P.; Bokstrom, H.; Wiquist, N. Prostaglandin E and F_{2a} Concentration in the Cervical Mucus and Mechanism of Cervical Ripening. *Prostaglandins* **1997**, *53*, 253–261.
57. Gonzalez-Estrella, J.; Coney, P.; Ostash, K.; Karabinus, D. Dithiothreitol Effects on the Viscosity and Quality of Human Semen. *Fertil. Steril.* **1994**, *62*, 1238–1243.
58. Casper, F.; Petri, E. Local Treatment of Urogenital Atrophy with an Estradiol-Releasing Vaginal Ring: A Comparative and a Placebo-controlled Multicenter Study. *Int. Urogynecol. J.* **1999**, *10*, 171–176.
59. Odeblad, E. Biophysical Composition of Cervical Mucus and Sperm Migration during Treatment with Conluten and Conlunett. *Acta Obstetrica et Gynecologica Scandinavica* **1968**, *47* (suppl. 8), 7–19.
60. Moghissi, K. Cervical Factor in Fertility. *Reproductive Medicine and Surgery*; Mosby Book Co.: Chicago, 1995; 376–397.
61. Katz, D.; Slade, D.; Nakajima, S. Analysis of Pre-ovulatory Changes in Cervical Mucus Hydration and Sperm Penetrability. *Adv. Contraception* **1997**, *13* (2–3), 143–151.
62. Nasir-Ud-Din; Jeanloz, R.; Vercelotti, J.; McArthur, J. Studies on Bonnet Monkey Cervical Mucus: The Effect of Proteases on Mucus Glycoproteins of *Macaca radiata*. *Biochimica. Biophysica. Acta* **1981**, *678*, 483–496.
63. Madsen, F.; Eberth, K.; Smart, J. A Rheological Examination of the Mucoadhesive/Mucus Interaction: The Effect of Mucoadhesive Type and Concentration. *J. Con. Rel.* **1998**, *50*, 167–178.
64. Moriyama, A.; Shimoya, K.; Ogata, I.; Kimura, T.; Nakamura, T.; Wada, H.; Ohashi, K.; Azuma, C.; Saji, F.; Murata, Y. Secretory Leukocyte Protease Inhibitor Concentrations in Cervical Mucus of Women with Normal Menstrual Cycle. *Mol. Human Reprod.* **1999**, *5*, 665–661.
65. Kocevar-Nared, J.; Kristl, J.; Smid-Korbar, J. Comparative Rheological Investigation of Crude Gastric Mucin and Natural Gastric Mucus. *Biomaterials* **1997**, *18* (9), 677–681.
66. Sayani, A.; Chun, I.; Chien, Y. Transmucosal Delivery of Leucine Enkephalin: Stabilizing in Rabbit Enzyme Extracts and Enhancement of Permeation through Mucosa. *J. Pharm. Sci.* **1993**, *82*, 1179–1185.
67. Hartman, C.; *Endocrinology* **1939**, *25*, 670.
68. Behl, C., Ph.D. *Dissertation*; University of Michigan, Ann Arbor.
69. Rock, J.; Barker, R.; Bacon, W.; *Science* **1947**, *105*, 13.
70. Shudmak, M.; Hesseltine, H.; *Am. J. Obstet. Gynecol.* **1951**, *62*, 669.
71. Baker, D.; *Anat. Rec.* **1928**, *39*, 339.
72. Laug, E.; Kunze, F.; *J. Pharm. Exp. Ther.* **1949**, *95*, 460.
73. Despande, A.; Rhodes, C. Intravaginal Drug Delivery. *Drug Dev. Ind. Pharm.* **1992**, *18*, 1225–1279.
74. Chien, Y. Vaginal Drug Delivery and Delivery Systems. *Novel Drug Delivery Systems*; Marcel Dekker, Inc.: New York, 1992; 529–584.
75. Kabongo, M. Problems with Over-the counter Vaginal Preparations. *Amer. Family Phys* **1993**, *48*, 579.
76. Nachtigall, L. Comparative Study: Replens versus Local Estrogen in Menopausal Women. *Fertil. Steril.* **1994**, *61*, 178–180.
77. Holst, E.; Brandberg, A. Treatment of Bacterial Vaginosis in Pregnancy with a Lactate Gel. *Scand. J. Inf. Dis.* **1990**, *22*, 625–626.
78. Brannan-Peppas, L. Novel Vaginal Drug Release Applications. *Adv. Drug Delivery Rev.* **1993**, *11*, 169–177.
79. Mishell, D., Jr.; Talas, M., Jr.; Parlow, A., Jr.; Moyr, D., Jr. Contraception by Means of a Silastic Vaginal Ring Impregnated with Medroxyprogesterone Acetate. *Am. J. Obstet. Gynecol.* **1970**, *107*, 100.
80. Mishell, D., Jr.; Lumkin, M., Jr.; Stone, S., Jr. Inhibition of Ovulation with Cyclic Use of Progestogen-Impregnated Intravaginal Devices. *Am. J. Obstet. Gynecol.* **1972**, *113*, 927–32.
81. Cornette, J.; Kirton, K.; Duncan, G. Measurement of Medroxyprogesterone Acetate (Provera) by Radio-Immunoassay. *J. Clin. Endocrinol. Metab.* **1971**, *33*, 459.
82. Hiroi, M.; Stanczyk, F.; Goebelsmann, U.; Brenner, P.; Lumkin, M.; Mishell, D., Jr. Radioimmunoassay of Serum Medroxyprogesterone Acetate (Provera) in Women Following Oral and Intravaginal Administration. *Steroids* **1975**, *26*, 373.
83. Thiery, M.; Vandekerckhove, D.; Dhondt, M.; Vermeulen, A.; Decoster, J. The Medroxyprogesterone Acetate Intravaginal Silastic Ring as a Contraceptive Device. *Contraception* **1976**, *13*, 605–617.
84. Vermeulen, A.; Dhondt, M.; Thiery, M.; Vandekerckhove, D. Plasma Sex Steroid and Gonadotropin Levels in Control and Silastic Vaginal Medroxyprogesterone Acetate-Impregnated Ring Cycles. *Fertil. Steril.* **1976**, *27*, 773–779.
85. Landgren, B.; Aedo, A.; Johannisson, E.; Cekan, S. Pharmacokinetic Studies with a Vaginal Delivery System Releasing Levonorgestrel at a Near Zero Order Rate for One Year. *Contraception* **1994**, *49*, 139–150.
86. Schmidt, G.; Andersson, S.; Nordle, O.; Johansson, C.; Gunnarsson, P. Release of 17-Beta-Oestradiol from a Vaginal Ring in Postmenopausal Women: Pharmacokinetic Evaluation. *Gynecol. Obstet. Inv.* **1994**, *38*, 253–260.
87. Nash, H.; Brache, V.; Alvarez-Sanchez, F.; Jackanicz, T.; Harmon, T. Estradiol Delivery by Vaginal Rings: Potential for Hormone Replacement Therapy. *Maturitas* **1997**, *26*, 27–33.
88. Victor, A.; Nash, H.; Jackanicz, T.; Johansson, E.; *Contraception* **1975**, *12*, 261.
89. Stanczyk, F.; Hiroi, M.; Goebelsmann, U.; Brenner, P.; Lumkin, M.; Mishell, D., Jr.; *Contraception* **1975**, *23*, 279.
90. World Health Organization. *Contraception* **1990**, *27*, 105–124.
91. Burton, F.; Skiens, W.; Gordon, N.; Veal, J.; Kalkwarf, D.; Duncan, G. Fabrication and Testing of Vaginal Contraceptive Devices Designed for Release of Prespecified Dose Levels of Steroids. *Contraception* **1978**, *17*, 221–230.
92. Landgren, B.; Johannisson, E.; Masironi, B.; Diczfalussy, E. Pharmacokinetic and Pharmacodynamic Investigations with Vaginal Devices Releasing Levonorgestrel at a Constant, Near Zero Order Rate. *Contraception* **1982**, *26*, 567–585.

93. Laundgren, B.; Johannisson, E.; Xing, S.; Aedo, A.; Diczfalusy, E. A Clinical Pharmacological Study of a New Type of Vaginal Delivery System for Levonorgestrel. *Contraception* **1985**, *26*, 581–601.
94. Laundgren, B.; Aedo, A.; Cekan, S.; Diczfalusy, E. Pharmacokinetic Studies with a Vaginal Delivery System Releasing Levonorgestrel at a Near Zero Order Rate for One Year. *Contraception* **1986**, *26*, 473–497.
95. Bals-Pratsch, M.; Al-Hasani, S.; Schopper, B.; Diedrich, C.; Hoepfner, A.; Weiss, J.; Kupker, W.; Felberbaum, R.; Ortmann, O.; Bauer, O.; Diedrich, K.; Hum Reprod. **1999**, *14* (Suppl 1), 222–230.
96. Mishell, D., Jr.; Moore, D., Jr.; Roy, S., Jr.; Brenner, P., Jr.; Page, M., Jr. Clinical Performance and Endocrine Profiles with Contraceptive Vaginal Rings Containing a Combination of Estradiol and D-Norgestrel. *Am. J. Obstet. Gynecol.* **1978**, *130*, 55–62.
97. Nash, H. *Controlled Release Systems for Contraception. Medical Applications of Controlled Release*; CRC Press: Florida, 1984; 35–64.
98. Sam, A. *Minutes of 5th International Pharmaceutical Technology Symposium on New Approaches to Controlled Drug Delivery*; Hincal, A.A. Ed.; Editions de Sante: Paris, 1991; 271–284.
99. Timmer, C.; Apter, D.; Voortman, G. Pharmacokinetics of 3-Keto-desogestrel and Ethinylestradiol Released from Different Types of Contraceptive Vaginal Rings. *Contraception* **1990**, *42* (6), 629–642.
100. Chien, Y. Microsealed Drug Delivery Systems: Fabrications and Performance. *Methods Enzymol.* **1985**, *112*, 461–470.
101. Hermans, R.; Longo, L.; McGivern, R. Decreased Postnatal Testosterone and Corticosterone Concentrations in Rats Following Acute Intermittent Prenatal Hypoxia without Alterations in Adult Male Sex Behavior. *Neurotoxicol Teratol* **1994**, *16* (2), 201–206.
102. Witter, F.; Rocco, L.; Johnson, T. A Randomized Trial of Prostaglandin E2 in a Controlled-release Vaginal Pessary for Cervical Ripening at Term. *Am. J. Obstet. Gyn.* **1992**, *166*, 830–834.
103. Henriksson, L.; Stjernquist, M.; Boquist, L.; Alander, U.; Selinus, I. A Comparative Multicenter Study of the Effects of Continuous Low-Dose Estradiol Released from a New Vaginal Ring versus Estriol Vaginal Pessaries in Postmenopausal Women with Symptoms and Signs of Urogenital Atrophy. *Am. J. Obstet. Gyn.* **1994**, *171*, 624–632.
104. Igarashi, M.; Iizuka, M.; Abe, Y.; Ibuki, Y. Novel Vaginal Danazol Ring Therapy for Pelvic Endometriosis, in Particular Deeply Infiltrating Endometriosis. *Hum. Reprod.* **1998**, *13* (7), 1952–1956.
105. Currie, W.; Joseph, I.; Rawlings, N. Morphine, Naloxone and the Gonadotrophin Surge in Ewes. *Reprod. Fertil.* **1991**, *92*, 407–414.
106. Kabadi, M.; Chien, Y. Intravaginal Controlled Administration of Flurogestone Acetate II. *J. Pharm. Sci.* **1984**, *73*, 1464–1468.
107. Kabadi, M.; Chien, Y. Intravaginal Controlled Administration of Flurogestone Acetate III. *Drug Dev. Ind. Pharm.* **1985**, *11*, 1271–1312.
108. Kabadi, M.; Chien, Y. Intravaginal Controlled Administration of Flurogestone Acetate IV. *Drug Dev. Ind. Pharm.* **1985**, *11*, 1313–1316.
109. Knuth, K.; Amiji, M.; Robinson, J. Hydrogel Delivery Systems for Vaginal and Oral Applications, Formulation and Biological Considerations. *Adv. Drug Delivery Rev.* **1993**, *11*, 137–167.
110. Machida, Y.; Masuda, H.; Fujiyama, N.; Ito, S.; Iwata, M.; Nagai, T. Preparation and Phase II Clinical Examination of Topical Dosage Form for Treatment of Carcinoma Coli Containing Bleomycin with Hydroxypropyl Cellulose. *Chem. Pharm. Bull.* **1979**, *27*, 93–100.
111. Gürsoy, A.; Bayhan, A.; *Drug Dev. Ind. Pharm.* **1992**, *18*, 203–221.
112. Gürsoy, A.; Sohtorik, I.; Uyanik, N.; Peppas, N.; *STP Pharma.* **1989**, *5*, 886–892.
113. Morimoto, K.; Takeeda, T.; Nakamoto, Y.; Morisaka, K.; *Int. J. Pharm.* **1982**, *12*, 107–111.
114. Bonucci, E.; Ballanti, P.; Ramires, R.; Richardson, J.; Benedetti, L. Prevention of Ovariectomy Osteopenia in Rats after Vaginal Administration of Hyaff 11 Microspheres Containing Salmon Calcitonin. *Calcif. Tissue Int.* **1995**, *56*, 274–279.
115. Leung, S.; Robinson, J.; *Polym. News* **1990**, *15*, 333–342.
116. Jain, S.; Singh, R.; Sahu, B.; *Drug Dev. Ind. Pharm.* **1997**, *23*, 827–830.
117. Foldvari, M.; Moreland, A.; *Liposome Res.* **1997**, *7*, 115–126.
118. Pavelic, Z.; Skalko-Basnet, N.; Jalsenjak, I. Liposomes Containing Drugs for Treatment of Vaginal Infections. *Euro. J. Pharmaceutical Sci.* **1999**, *8*, 345–351.
119. Krebs, F.; Miller, S.; Catalone, B.; Welsh, P.; Malamud, D.; Howett, M.; Wigdahl, B. Sodium Dodecyl Sulfate and C31G as Microbicidal Alternatives to N-9: Comparative Sensitivity of Primary Human Vaginal Keratocytes. *Antimicrobial agents Chemotherapy* **2000**, *44*, 1954–1960.
120. Salas-Herrera, I.; Lawson, M.; Johnston, A.; Turner, P.; Gott, D.; Dennis, M. British Plasma Metronidazole Concentrations after Single and Repeated Vaginal Pessary Administration. *J. Clin. Pharmacol.* **1991**, *32*, 621–623.
121. Hoffmann, C.; Focke, N.; Franke, G.; Zschiesche, M.; Siegmund, W. Int. Comparative Bioavailability of Metronidazole Formulations (Vagimid) after Oral and Vaginal Administration. *J. Clin. Pharm. Ther.* **1995**, *33*, 232–239.
122. Fischbach, F.; Petersen, E.; Weissenbacher, E.; Martius, J.; Hosmann, J.; Mayer, H. Efficacy of Clindamycin Vaginal Cream versus Oral Metronidazole in the Treatment of Bacterial Vaginosis. *Obstet. Gynecol.* **1993**, *82*, 405–410.
123. Borin, M. Systemic Absorption of Clindamycin Following Intravaginal Application of Clindamycin Phosphate 1% Cream. *J. Clin. Pharmacol.* **1990**, *30*, 33–38.
124. Lawrence, A.; Houang, E.; Hiscock, E.; Wells, M.; Colli, E.; Scatigna, M. Single Dose Therapy of Vaginal Candidiasis: A Comparative Trial of Fenticonazole Vaginal Ovules versus Clotrimazole Vaginal Tablets. *Curr. Med. Res. Opinion* **1990**, *12*, 114–120.
125. Boag, F.; Houang, E.; Westrom, R.; McCormack, S.; Lawrence, A. Comparison of Vaginal Flora after Treatment with a Clotrimazole 500 mg Vaginal Pessary or a Fluconazole 150 mg Capsule for Vaginal Candidosis. *Genitourinary Med.* **1991**, *67*, 232–234.

126. Richardson, J.; Minhas, P.; Thomas, N. Vaginal Administration of Gentamicin to Rats — Pharmaceutical and Morphological Studies Using Absorption Enhancers. *Int. J. Pharm.* **1989**, *56*, 29–36.
127. Merkus, J.; J. Amer. Acad. Dermatol. **1990**, *23*, 568–572.
128. Novelli, A.; Periti, E.; Massi, G.; Masi, R.; Mazzei, T.; Periti, P. Systemic Absorption of 3H-Fenticonazole after Vaginal Administration of 1 gram in Patients. *J. Chemotherapy* **1993**, *3*, 23–27.
129. O'Connell, M.; Hoskin, P.; Mayles, W.; McElwain, T.; Barrett, A. Intravaginal Iridium-192 in the Management of Embryonal Rhabdomyosarcoma. *Clin. Oncology* **1991**, *3*, 236–239.
130. Coppi, G.; Silingardi, S.; Girardello, R.; De Aloysio, D.; Manzardo, S. Pharmacokinetics of Ciclopirox Olamine after Vaginal Application to Rabbits and Patients. *J. Chemotherapy* **1993**, *5*, 302–306.
131. Fujii, T.; Naito, H.; Kioka, H.; Tanioka, Y.; Murakami, J.; Sanada, M.; Tanimoto, H.; Nakagawa, H.; Tanaka, T.; Furui, J. Effect of Intravaginal Administration of Cisplating (CDDP) Suppositories to Uterine Cervical Cancer—Blood and Tissue Concentrations and the Therapeutic Effects. *Jpn. J. Cancer Chem.* **1995**, *22*, 99–103.
132. Schraub, S.; Sun, X.; Maingon, P.; Horiot, J.; Daly, N.; Keiling, R.; Pigneux, J.; Pourquier, H.; Rozan, R.; Vrousos, C.; *Cancer* **1992**, *69*, 2505–2509.
133. Lauersen, N.; Wlson, K.; Prostaglandins **1976**, *12* (Suppl), 63.
134. Lauersen, N.; Wlson, K.; Prostaglandins **1976**, *13*, 697.
135. Roseman, J.; Spilman, C. *Controlled Release Systems*; AICHE Symp.: New York, 1981; 77.
136. Cohen, L.; Lordi, N.; *J. Pharm. Sci.* **1980**, *69*, 955.
137. Spilman, C.; Roseman, T.; Baker, R.; Tuttle, M.; Lonsdale, H. *Controlled Release Delivery Systems*; Marcel Dekker, Inc.: New York, 1983; 133–140.
138. Duenhoelter, J.; Ramos, R.; Milewich, L.; MacKonald, P. Interruption of Early Pregnancy with a Silastic Device Containing (15S)-15-Methyl Prostaglandin F₂alpha Methyl Ester: Efficacy and Mode of Action. *Contraception* **1978**, *17*, 51.
139. Karim, S.; Ratnam, S. Termination of Pregnancy with Vaginal Administration of 16, 16 Dimethyl Prostaglandin E₂ *p*-Benzaldehyde Semicarbazone Ester. *Br. J. Obstet. Gynecol.* **1977**, *84*, 135–137.
140. Karim, S.; Ratnam, S.; Prasad, R.; Wong, Y. Vaginal Administration of a Single Dose of 16, 16 Dimethyl Prostaglandin E₂ *p*-Benzaldehyde Semicarbazone Ester for Pre-operative Cervical Dilatation in First Trimester Nulliparae. *Br. J. Obstet. Gynecol.* **1977**, *84*, 269–271.
141. Spilman, C.; Roseman, T. Luteolytic and Abortifacient Effects in Rhesus Monkeys of Silicone Vaginal Rings Containing 15(S)-15-Methyl-prostaglandin F₂ Alpha Methyl Ester. *Contraception* **1975**, *11*, 409–18.
142. Spilman, C.; Beuving, D.; Forbes, A.; Roseman, T.; Larion, L. Evaluation of Vaginal Delivery Systems Containing 15[s]-15-Methyl PGF₂ Alpha Methyl Ester. *Prostaglandins. Prostaglandins* **1976**, *12* (Suppl.), 1–16.
143. Graham, N.; McNeill, M. Hydrogels for Controlled Drug Delivery. *Biomaterials* **1984**, *5* (1), 27–36.
144. Embrey, M.; Graham, N.; McNeill, M. Induction of Labour with a Sustained-Release Prostaglandin E₂ Vaginal Pessary. *Br. Med. J.* **1980**, *281*, 901–902.
145. Taylor, A.; Boland, J.; MacKenzie, I. The Concurrent In Vitro and In Vivo Release of PGE₂ from a Controlled-release Hydrogel Polymer Pessary for Cervical Ripening. *Prostaglandins* **1990**, *40*, 89–98.
146. Taylor, A.; Boland, J.; Bernal, A.; MacKenzie, I. Prostaglandin Metabolite Levels during Cervical Ripening with a Controlled-Release Hydrogel Polymer Prostaglandin E₂ Pessary. *Prostaglandins* **1991**, *41*, 585–594.
147. Lyndrup, J.; Nickelsen, C.; Guldback, E.; Weber, T. Induction of Labor by Prostaglandin E₂: Intracervical Gel or Vaginal Pessaries? *Eur. J. Obstet. Gyn. Rep. Biol.* **1992**, *42*, 101–109.
148. Bennett, M.; Horowitz, S.; Wass, D. The Use of 16, 16-Dimethyl-PGE₁-deltamethyl Ester (Cervagem) Vaginal Pessaries for Cervical Dilatation Prior to Evacuation for Second Trimester Termination. *Austr. New Zealand J. Obstet. Gyn.* **1991**, *31*, 44–47.
149. Kurup, A.; Chua, S.; Arulkumaran, S.; Tham, K.; Tay, D.; Ratnam, S. Induction of Labour in Nulliparas with Poor Cervical Score: Oxytocin or Prostaglandin Vaginal Pessaries? *Austr. New Zealand J. Obstet. Gyn.* **1991**, *31*, 223–226.
150. Rayburn, W.; Wapner, R.; Barss, V.; Spitzberg, E.; Molina, R.; Mandsager, N.; Yonekura, M. An Intravaginal Controlled-Release Prostaglandin E₂ Pessary for Cervical Ripening and Initiation of Labor at Term. *Obstet. Gynecol.* **1992**, *79*, 374–379.
151. Stampe Sorensen, S.; Palmgren Colov, N.; Andreasson, B.; Bock, J.; Berget, A.; Schmidt, T. Induction of Labor by Vaginal Prostaglandin E₂. A Randomized Study Comparing Pessaries with Vaginal Tablets. *Acta. Obstet. Gyn. Scand.* **1992**, *71*, 201–206.
152. Bugalho, A.; Bique, C.; Machungo, F.; Bergstrom, S. A Comparative Study of Vaginal Misoprostol and Intravenous Oxytocin for Induction of Labour. *Gyn. Obstet. Invest.* **1995**, *39*, 252–256.
153. Sawai, S.; O'Brien, W.; Mastrogiannis, D.; Krammer, J.; Mastry, M.; Porter, G. Patient-administered Out-patient Intravaginal Prostaglandin E₂ Suppositories in Post-date Pregnancies: A Double-blind, Randomized, Placebo-controlled Study. *Obstet. Gyn.* **1994**, *84*, 807–810.
154. Greer, I.; McLaren, M.; Calder, A. Vaginal Administration of PGE₂ for Induction of Labor Stimulates Endogenous PGF₂ Alpha Production. *Acta. Obstet. Gynecol. Scand.* **1990**, *69*, 621–625.
155. Mahmood, T.; Dick, M. A Randomized Trial of Management of Pre-labor Rupture of Membranes at Term in Multiparous Women Using Vaginal Prostaglandin Gel. *Obstet. Gynecol.* **1995**, *85*, 71–74.
156. Seeras, R. Induction of Labor Utilizing Vaginal vs. Intracervical Prostaglandin E₂. *Int. J. Gyn. Obstet.* **1995**, *48*, 163–167.
157. Seeras, R.; Olatunbosun, O.; Pierson, R.; Turnell, R. Induction of Labor Using Prostaglandin E₂ (PGE₂) Vaginal Gel in Triacetin Base. An Efficacy Study Comparing Two Dosage Regimens. *Clin. Exp. Obstet. Gyn.* **1995**, *22*, 105–110.
158. Hill, N.; Ferguson, J.; MacKenzie, I. The Efficacy of Oral Mifepristone (RU 38,486) with a Prostaglandin E₁ Analog Vaginal Pessary for the Termination of Early

- Pregnancy: Complications and Patient Acceptability. *Am. J. Obstet. Gyn.* **1990**, *162*, 414–417.
159. Bugalho, A.; Bique, C.; Machungo, F.; Bergstrom, S. Vaginal Misoprostol as an Alternative to Oxytocin for Induction of Labor in Women with Late Fetal Death. *Acta. Obstet. Gyn. Scand.* **1995**, *74*, 194–198.
160. Creinin, M.; Vittinghoff, E. Methotrexate and Misoprostol vs Misoprostol Alone for Early Abortion. A Randomized Controlled Trial. *JAMA* **1994**, *272*, 1190–1195.
161. Diao, X.; Zou, S.; Quigg, J.; Kaminski, J.; Zaneveld, L. Comparative Vaginal Spermicidal Studies in the Stump-tailed Macaque with Alkyloxynol-741, Nonoxynol-9 and Chlorhexidine. *Contraception* **1990**, *42*, 677–682.
162. Zavos, P.; Correa, J.; Zarmakoupis-Zavos, P. Assessment of a Tablet Drug Delivery System Incorporation Nonoxynol-9 Coprecipitated with Polyvinylpyrrolidone in Preventing the Onset of Pregnancy in Rabbits. *Fertil Steril.* **1998**, *69*, 768–73.
163. Psychoyos, A.; Creatsas, G.; Hassan, E.; Georgoulis, V.; Gravanis, A. Spermicidal and Antiviral Properties of cholic Acid: Contraceptive Efficacy of a New Vaginal Sponge (Protectaid) Containing Sodium Cholate. *Human Reprod.* **1993**, *8*, 866–869.
164. Bourinbaier, A.; Lee-Huang, S. The Activity of Plant-derived Antiretroviral Proteins MAP30 and GAP31 against Herpes Simplex Virus In Vitro. *Contraception* **1995**, *51*, 319–322.
165. Chapvil, M.; Eskelson, C.; Stiffel, V.; Owen, J.; Droegemuller, W. Studies on Nonoxynol-9. II. Intravaginal Absorption, Distribution, Metabolism and Excretion in Rats and Rabbits. *Contraception* **1980**, *22*, 325–339.
166. Walter, B.; Agha, B.; Digenis, G. Disposition of [14C]nonoxynol-9 after Intravenous or Vaginal Administration to Female Sprague-Dawley Rats. *Toxicol. Appl. Pharmac.* **1988**, *96*, 258–268.
167. Owen, D.; Dunmire, E.; Plenys, A.; Katz, D. Factors Influencing Nonoxynol-9 Permeation and Bioactivity in Cervical Mucus. *J. Cont. Rel.* **1999**, *60*, 23–34.
168. Baker, V.; Obstet. Gyn. Clin. N. Amer. **1994**, *21*, 271–297.
169. Handa, V.; Bachus, K.; Johnston, W.; Robboy, S.; Hammond, CB. Vaginal Administration of Low-Dose Conjugated Estrogens — Systemic Absorption and Effects on the Endometrium. *Obstet. Gynecol.* **1994**, *84*, 215–218.
170. Einer-Jensen, N.; Kotwica, J.; Krzymowski, T.; Stefanczyk-Krzymowska, S.; Kaminski, T.; *Acta Vet. Scand.* **1993**, *34*, 1–7.
171. Miles, R.; Paulson, R.; Lobo, R.; Press, M.; Dahmouh, L.; Sauer, M. Pharmacokinetics and Endometrial Tissue Levels of Progesterone after Administration by Intramuscular and Vaginal Routes: A Comparative Study. *Fertil. Steril.* **1994**, *62*, 485–490.
172. Kimzey, L.; Gumowski, J.; Merriam, G.; Grimes, G., Jr.; Nelson, L., Jr. Absorption of Micronized Progesterone from a Nonliquefying Vaginal Cream. *Fertil. Steril.* **1991**, *56*, 995–996.
173. Norman, T.; Morse, C.; Dennerstein, L. Comparative Bioavailability of Orally and Vaginally Administered Progesterone. *Fertil. Steril.* **1991**, *56*, 1034–1039.
174. Bulletti, C.; de Ziegler, D.; Flamigni, C.; Giacomucci, E.; Polli, V.; Bolelli, G.; Franceschetti, F. Targeted Drug Delivery in Gynaecology: The First Uterine Pass Effect. *Hum. Reprod.* **1997**, *12*, 1073–1079.
175. Sakaguchi, M.; Sakai, T.; Adachi, Y.; Kawashima, T.; Awata, N. The Biological Fate of Sodium Prasterone Sulfate after Vaginal Administration. 1. Absorption and Excretion in Rats. *J. Pharmac. Dyn.* **1992**, *15*, 67–73.
176. Okada, H.; Yashiki, T.; Mima, H. Vaginal Absorption of a Potent Luteinizing Hormone-releasing Hormone Analogue (leuprolide) in Rats III: Effect of Estrous Cycle on Vaginal Absorption of Hydrophilic Model Compounds. *J. Pharm. Sci.* **1983**, *72*, 173–176.
177. Richardson, J.; Illum, L.; Thomas, N. Vaginal Absorption of Insulin in the Rate Effect of Penetration Enhancers on Insulin Uptake and Mucosal Histology. *Pharm. Res.* **1992**, *9*, 878–883.
178. Satoh, T.; Yoshida, G.; Orito, Y.; Koike, T. Drug Delivery System for the Treatment of Osteoporosis. *Nippon Rinsho* **1998**, *56*, 742–747.
179. Chen, S.; Reed, B.; Nguyen, T.; Gaylord, N.; Fuller, G.; Mordenti, J. The Pharmacokinetics and Absorption of Recombinant Human Relaxin in Nonpregnant Rabbits and Rhesus Monkeys after Intravenous and Intravaginal Administration. *Pharm. Res.* **1993**, *10*, 223–227.
180. Chen, S.; Perlman, A.; Spanski, N.; Peterson, C.; Sanders, S.; Jaffe, R.; Martin, M.; Yalcinkaya, T.; Cefalo, R.; Chescheir, N. The Pharmacokinetics of Recombinant Human Relaxin in Nonpregnant Women after Intravenous, Intravaginal, and Intracervical Administration. *Pharm. Res.* **1993**, *10*, 834–838.
181. Okada, H.; Yamazaki, I.; Ogawa, Y.; Hirai, S.; Yashiki, T.; Mima, H. Vaginal Absorption of a Potent Luteinizing Hormone-Releasing Hormone Analog (leuprolide) in Rats I: Absorption by Various Routes and Absorption Enhancement. *J. Pharm. Sci.* **1982**, *71*, 1367–1371.
182. Okada, H.; Yamazaki, I.; Yashiki, T.; Mima, H. Vaginal Absorption of a Potent Luteinizing Hormone-Releasing Hormone Analogue (leuprolide) in Rats II: Mechanism of Absorption Enhancement with Organic Acids. *J. Pharm. Sci.* **1983**, *72*, 75–78.
183. Okada, H.; Sakura, R.; Kawaji, H.; Yashiki, T.; Mima, H. Regression of Rat Mammary Tumors by a Potent Luteinizing Hormone-Releasing Hormone Analogue (leuprolide) Administered Vaginally. *Cancer Res.* **1983**, *43*, 1869–1874.
184. Okada, H.; Yamazaki, I.; Sakura, Y.; Yashiki, T.; Shimamoto, T.; Mima, H. Desensitization of Gonadotropin-Releasing Response Following Vaginal Consecutive Administration of Leuprolide in Rats. *J. Pharm. Dyn.* **1983**, *6*, 512–522.
185. Okada, H.; Yamazaki, I.; Yashiki, T.; Shinamoto, T.; Mima, H. Vaginal Absorption of a Potent Luteinizing Hormone-releasing Hormone Analogue (leuprolide) in Rats. IV: Evaluation of the Vaginal Absorption and Gonadotropin Responses by Radioimmunoassay. *J. Pharm. Sci.* **1984**, *73*, 298–302.
186. Han, K.; Park, J.; Chung, Y.; Lee, M.; Moon, D.; Robinson, J. Identification of Enzymatic Degradation Products of Luteinizing Hormone Releasing Hormone (LHRH)/[D-Ala6]LHRH in Rabbit Mucosal Homogenates. *Pharm. Res.* **1995**, *12*, 1539–1544.

187. Livingston, J.; Lu, S.; Robinson, H.; Anderson, D. Immunization of the Female Genital Tract with a DNA-based Vaccine. *Infect. Immun.* **1998**, *66*, 322–329.
188. D'cruz, O.; Venkatachalam, T.; Uckun, F. Structural Requirements for Potent Humand Spermicidal Activity of Dual Function Aryl Phosphate Derivative of Bromomethoxy Zidovudine. *Biol Reprod* **2000**, *62*, 37–44.
189. O'Hagan, D.; Rafferty, D.; McKeating, J.; Illum, L. Vaginal Immunization of Rats with a Synthetic Peptide from Human Immunodeficiency Virus Envelope Glycoprotein. *J. Gen. Virol.* **1992**, *73*, 2141–2145.
190. Eldridge, J.; Staas, J.; Chen, D.; Marx, P.; Tice, T.; Gilley, R.; Sem. *Hematol.* **1993**, *30* (Suppl), 16–24.
191. Yokoyama, M.; Zhang, J.; Whitton, L. DNA Immunization: Effects of Vehicle and Route of Administration on the Induction of Protective Antiviral Immunity. *FEMS Immunology Med. Microbiology* **1996**, *14*, 221–230.
192. Pertmer, T.; Robersts, T.; Hynes, J. Influenza Virus Nucleoprotein-specific Immunoglobulin G Subclass and Cytokine Responses Elicited by DNA Vaccination are Dependent on the Route of Vector DNA Delivery. *J. Virology* **1996**, *70*, 6119–6125.
193. Gramzinski, R.; Millan, C.; Obaldia, N.; Hoffman, S.; Davis, H. Immune Response to a Hepatitis B DNA Vaccine in Aotus Monkeys: A Comparison of Vaccine Formulation, Route and Method of Administration. *Molecular Medicine* **1998**, *4*, 109–118.
194. Franklin, R.; Kutteh, W. Characterization of Immunoglobulins and Cytokines in Human Cervical Mucus: Influence of Exogenous and Endogenous Hormones. *J. Reprod. Immunology* **1999**, *42*, 93–106.
195. Wang, B.; Dang, K.; Agadjanyan, M. Mucosal Immunization with a DNA Vaccine Induces Immune Responses against HIV-1 at a Mucosal Site. *Vaccine* **1997**, *15*, 821–825.
196. Bagarazzi, M.; Boyer, J.; Javadian, M. Safety and Immunogenicity of Intramuscular and Intravaginal Delivery of HIV-1 DNA Constructs to Infant Chimpanzees. *J. Med. Primatol.* **1997**, *26*, 27–33.
197. McCluskie, M.; Millan, C.; Gramzinski, R.; Robinson, H.; Santoro, J.; Fuller, J.; Widera, G.; Haynes, J.; Purcell, R.; Davis, H. Route and Method of Delivery of DNA Vaccine Influence Immune Responses in Mice and Non-human Primates. *Molecular Medicine* **1999**, *5*, 287–3000.
198. Bowen, J.; Alpar, H.; Phillpotts, R.; Brown, M. Mucosal Delivery of Herpes Simplex Virus Vaccine. *Research Virol.* **1992**, *143*, 269–278.
199. O'Hagan, D.; Rafferty, D.; Wharton, S.; Illum, L. Intravaginal Immunization in Sheep Using a Bioadhesive Microsphere Antigen Delivery System. *Vaccine* **1993**, *11*, 660–664.
200. Parr, M.; Parr, E. Antigen Recognition in the Female Reproductive Tract: I. Uptake of Intraluminal Protein Tracers in the Mouse Vagina. *J. Reprod. Immunol.* **1990**, *17*, 101–114.
201. Rush, C.; Hafner, L.; Timms, P. Genetic Modification of a Vaginal Strain of *Lactobacillus Fermentum* and its Maintenance within the Reproductive Tract after Intravaginal Administration. *J. Med. Microbiol.* **1994**, *41*, 272–278.
202. Thapar, M.; Parr, E.; Parr, M. The Effect of Adjuvants on Antibody Titers in Mouse Vaginal Fluid after Intravaginal Immunization. *J. Reprod. Immunol.* **1990**, *17*, 207–216.
203. Thapar, M.; Parr, E.; Parr, M. Secretory Immune Responses in Mouse Vaginal Fluid After Pelvic, Parenteral or Vaginal Immunization. *Immunology* **1990**, *70*, 121–125.
204. Waldman, R.; Cruz, J.; Rowe, D. Intravaginal Immunization of Humans with *Candida albicans*. *J. Immunology* **1972**, *109*, 662–664.
205. Saltzman, W.; Sherwood, J.; Adams, D.; Haller, P. Long-term Delivery Systems and Biodistribution. *Biotechnol Bioeng* **2000**, *67* (3), 253–264.
206. Uehling, D.; Hopkins, W.; Dahmer, L.; Balish, E. Phase I Clinical Trial of Vaginal Mucosal Immunization for Recurrent Urinary Tract Infection. *Urology* **1994**, *152*, 2308–2311.
207. Ogra, P.; Ogra, S. Local Antibody Response to Poliovaccine in the Human Female Genital Tract. *J. Immunology* **1973**, *110*, 1307–1311.
208. Smee, D.; Leonhardt, J. Vaccination against Bovine Herpes Mammillitis Virus Infections in Guinea Pigs. *Intervirology* **1994**, *37*, 20–24.
209. Ginsburg, J.; Hardiman, P.; Thomas, M. Vaginal Bromocriptine. *Lancet* **1991**, *338*, 1205–1206.
210. Baldock, G.; Brodie, R.; Chasseaud, L.; Taylor, T.; Walmsley, L.; Catanese, B. Pharmacokinetics of Benzydamine after Intravenous, Oral, and Topical Doses to Human Subjects. *Biopharm. Drug. Disp.* **1991**, *12*, 481–492.
211. Moodley, J.; Cohen, M.; Devraj, K.; Dutton, M. Vaginal Absorption of Low-dose Tranexamic Acid from Impregnated Tampons. *S. African Med. J.* **1992**, *81*, 150–152.
212. Banja, A. *Therapeutic Peptides and Proteins: Formulations, Processing, and Delivery Systems*; Technomic Publishing Co.: Lancaster, PA, 1995.

BIBLIOGRAPHY

- Chien, Y. *Novel Drug Delivery Systems*; Marcel Dekker, Inc.: New York, 1992.
- The Human Vagina*; North Holland Publishing Co.: New York, 1978.
- Lee, V. *Peptide and Protein Drug Delivery*; Marcel Dekker, Inc.: New York, 1991.

DRUG DELIVERY—TOPICAL AND TRANSDERMAL ROUTES

Kenneth A. Walters

An-eX Analytical Services Ltd, Cardiff, United Kingdom

INTRODUCTION

Over the past 3 decades there have been significant advances in the science of dermal and transdermal drug delivery. Much of this research has been reviewed and published elsewhere (1–9). This chapter, therefore, will concentrate on more recent innovations and research directions. The author recognizes that the review will not be exhaustive and apologizes unreservedly for any omissions, either deliberate or accidental. Thus, the reader interested in the recent developments in physical methods of skin permeation enhancement, such as iontophoresis, electroporation, and sonophoresis, is directed to the excellent reviews of Banga (5), Lai and Roberts (10), and Kost et al. (11), and the research papers of Guy et al. (12), Ilic et al. (13), and Lombry et al. (14). Likewise, detailed descriptions of the formulations used in dermal and transdermal drug delivery will be omitted because they are fully discussed elsewhere (4). Mathematical aspects of percutaneous absorption and its prediction will be briefly mentioned but readers wishing to delve into the complexities of this subject should consult the detailed analyses of Pugh et al. (15) and Roberts et al. (16). This chapter will concentrate on recent developments in our understanding of the skin barrier and the novel chemical methodologies used to reduce barrier function. Technologies that may be useful in reducing adverse local reactions, often associated with dermal and transdermal drug delivery, will be discussed. There will be a brief review on innovative drug delivery systems. Finally, the recent regulatory initiatives in dermatological therapy will be described.

STRATUM CORNEUM DEVELOPMENT, MICROSTRUCTURE, AND BARRIER FUNCTION

Stratum Corneum Development

The development of the stratum corneum involves several steps of cell differentiation, which has resulted in a structure-based classification of the layers above the basal layer (the stratum basale). Thus, the cells progress from the stratum basale through the stratum spinosum, the

stratum granulosum, and the stratum lucidum to the stratum corneum (17). Cell turnover from stratum basale to stratum corneum is estimated to be on the order of 21 days. The exact mechanism underlying the initiation of keratinocyte differentiation is not fully understood. It is known that protein kinase C and several keratinocyte-derived cytokines may play a regulatory role in the differentiation process (18, 19).

In the outer cell layers of the stratum spinosum, intracellular membrane-coating granules (100–300 nm in diameter) appear within the cells. Within these granules lamellar subunits arranged in parallel stacks are observed. These are believed to be the precursors of the intercellular lipid lamellae of the stratum corneum (20, 21). In the outermost layers of the stratum granulosum the lamellar granules migrate to the apical plasma membrane where they fuse and eventually extrude their contents into the intercellular space (20). The extrusion of the contents of lamellar granules is a fundamental requirement for the formation of the epidermal permeability barrier (22, 23). Thus, the entire process of epidermal terminal differentiation is geared toward the generation of the specific chemical morphology of the stratum corneum. As a result, the end products of this process are the intracellular protein matrix and the intercellular lipid lamellae.

The cornified cell envelope is the outermost layer of a corneocyte, and mainly consists of tightly bundled keratin filaments aligned parallel to the main face of the corneocyte. The envelope consists of both protein and lipid components in that the lipid is attached covalently to the protein envelope. The envelope lies adjacent to the interior surface of the plasma membrane (24). The corneocyte protein envelope appears to play an important role in the structural assembly of the intercellular lipid lamellae of the stratum corneum. The corneocyte possesses a chemically bound lipid envelope comprised of *N*- ω -hydroxyceramides, which are ester linked to the numerous glutamate side chains provided possibly by both the α -helical conformation and β -sheet conformation of involucrin in the envelope protein matrix (25, 26). In the absence of *N*- ω -hydroxyceramides, the stratum corneum intercellular lipid lamellae were abnormal and permeability barrier function was disrupted (27).

The Intercellular Lipids

The stratum corneum intercellular lipids exist as a continuous lipid phase occupying about 20% of the stratum corneum volume and arranged in multiple lamellar structures. They are composed of cholesterol (27%) and ceramides (41%), together with free fatty acids (9%), cholesteryl esters (10%) and cholesteryl sulfate (2%) (28) (Table 1). Phospholipids, which dominate in the basal layer, are converted to glucosylceramides and subsequently to ceramides and free fatty acids, and are virtually absent in the outer layers of the stratum corneum. Eight classes of ceramides have been isolated and identified in human stratum corneum (29, 30) but the functions of the individual ceramide types are not fully understood. Similarly, the exact function of cholesterol esters within the stratum corneum lamellae is also elusive but it is theoretically possible that cholesterol esters may span adjacent bilayers and serve as additional stabilizing moieties.

Overall, the intercellular lipid lamellae are highly structured, very stable, and constitute a highly effective barrier to chemical penetration and permeation. Considerable information on lipid structure within the stratum corneum has been generated by Bouwstra, Poncet, and colleagues using small angle X-ray diffraction and transmission electron microscopic techniques. These and earlier studies have shown that the lipid lamellae of the stratum corneum are orientated parallel to the corneocyte surface and have repeat distances of approximately 6.0–6.4 nm and 13.2–13.4 nm. In more recent studies on lipid packing (31), the Leiden group evaluated lipid

organization of the stratum corneum using electron diffraction and found that although the majority of lipids in the intercellular space were present in the crystalline state, there were some lipids existing in the gel state that had a slightly looser hexagonal packing arrangement in the outer layers of the stratum corneum.

Stratum Corneum Barrier Properties

Systematic evaluation of the skin permeability of many compounds has demonstrated that the intercellular lipids of the stratum corneum are essential for normal skin barrier function. It is clear that the major route of permeation across the stratum corneum is through the continuous intercellular lipid lamellae. Thus, the rate at which permeation occurs is largely dependent on the physicochemical characteristics of the penetrant, the most important being the relative ability to partition into the intercellular lipid lamellae and molecular size. Three major variables account for differences in the rate at which different compounds permeate the skin: the concentration of permeant applied, the partition coefficient of the permeant between the applied vehicle and the stratum corneum, and the diffusivity of the compound within the stratum corneum. A plot of the log of the skin permeability rate versus permeant lipophilicity is usually sigmoidal, and reflects the existence of both lipophilic and hydrophilic barriers. This suggests that compounds with partition coefficients indicating an ability to dissolve in both oil and water (i.e., log P of 1–3) would permeate the skin relatively rapidly. This has been confirmed for a variety of compounds. Data illustrating the skin permeability of various nonsteroidal anti-inflammatory agents are shown in Fig. 1 (32).

Table 1 Lipids of the stratum corneum intercellular space

Lipid	wt%	mol%
Cholesterol esters	10.0	7.5 ^a
Cholesterol	26.9	33.4
Cholesterol sulphate	1.9	2.0
Total cholesterol derivatives	38.8	42.9
Ceramide 1	3.2	1.6
Ceramide 2	8.9	6.6
Ceramide 3	4.9	3.5
Ceramide 4	6.1	4.2
Ceramide 5	5.7	5.0
Ceramide 6	12.3	8.6
Total ceramides	41.1	29.5
Fatty acids	9.1	17.0 ^a
Others	11.1	10.6 ^b

^aBased on C16 alkyl chain.

^bBased on MW of 500.

Mathematical Prediction of Permeation (Limitations)

Mathematical models relate skin permeability of exogenous molecules to physicochemical parameters of the permeant (octanol/water partition coefficients and molecular weight [a surrogate marker of molecular size]). Similar models for animal and human skin relate normalized equations of best-fit regressions based on:

$$\log P_{\text{calc}} = A + B \log K_{o/w} + C \cdot \text{MWt}$$

where P is the permeability coefficient, $K_{o/w}$ is the o/w partition coefficient, and MWt is the molecular weight.

Different values of A , B , and C have been derived depending on species (33) and dataset used (15). The models are significantly limited by the range of partition coefficients (log $K_{o/w}$ range -3 to ~ 6) and molecular weights (MWt range 18–765) of the datasets used, and

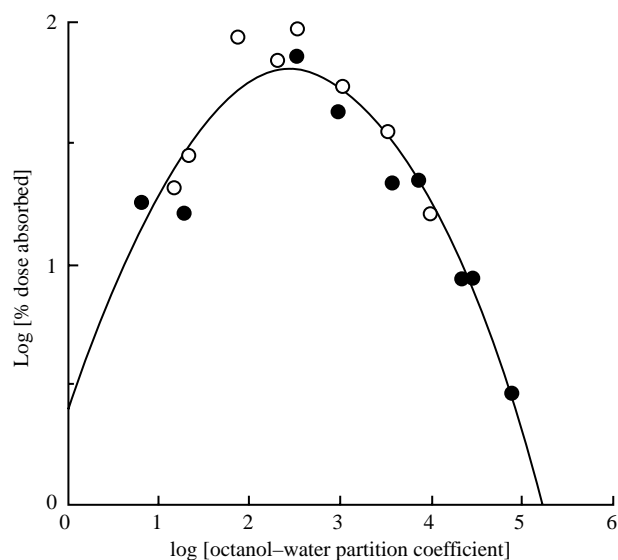


Fig. 1 Relationship between amount of drug absorbed through skin and compound octanol–water partition coefficient (increasing lipophilicity). Open circles: salicylates; closed circles: other nonsteroidal anti-inflammatory agents. (Redrawn from Ref. 32.)

anomalous values can be generated. As the boundaries of the dataset used are approached, the predictions become increasingly unrealistic. For example, the Vecchia equation (33) for human skin predicts that a permeant with molecular weight of 500 and $\log K_{o/w}$ of 6 (not unusual for new drugs and cosmetic ingredients) will permeate skin at a rate of 1.3×10^{-2} cm/h, an order of magnitude faster than water or methanol. In addition, as these models are based on experiments in which permeants were applied in aqueous solution, their value in prediction of permeation from actual formulations applied under in-use conditions is further compromised. For example, while the predicted permeability coefficient for *N*-nitroso-diethanolamine is 1.5×10^{-4} cm/h, the experimentally determined value from isopropylmyristate was 3.5×10^{-3} cm/h (34). The predicted value for octyl salicylate is 1.35×10^{-7} cm/h, while the experimental values from a hydroalcoholic lotion were 4.7×10^{-6} cm/h (infinite dose) and 6.6×10^{-7} cm/h (finite dose), and from an oil-in-water emulsion 1.7×10^{-5} cm/h (infinite dose) and 6.6×10^{-7} cm/h (finite dose) (35).

Datasets continue to be accumulated and predictive methods continue to be refined. However, while mathematical predictions may be useful in the comparison of the behavior of closely related compounds, it is inappropriate to use them for risk assessment or formulation optimization purposes without relevant experimental verification.

CHEMICAL SKIN PERMEATION ENHANCEMENT

The success of dermatological or transdermal drug delivery systems depends on the ability of the drug to penetrate into and/or permeate through skin in sufficient quantities to achieve therapeutic levels. Over the past 2 decades, several chemical skin permeation enhancers have been designed, synthesized, and evaluated (1,2). Many of these enhancers including Azone[®] (1-dodecylazacycloheptan-2-one) and SEPA[®] (2-*n*-nonyl-1,3-dioxolane) have been discussed in full elsewhere (36). Newer enhancers include 1-[2-(decylthio)ethyl]azacyclopentan-2-one (HPE-101), 4-decyloxazolid-2-one (Dermac[™] SR-38), and dodecyl-*N,N*-dimethylamino isopropionate (DDAIP, NexACT 88) (Fig. 2).

HPE-101 is believed to have a similar mechanism of action to Azone and is very sensitive to the vehicle of application (37). Thus, although the enhancer significantly increased the urinary excretion of indomethacin following topical application to hairless mice, it was dependent on the application vehicle (37). When the enhancer was applied in solution in polar solvents, such as dipropylene glycol, triethylene glycol, diethylene glycol, glycerin, water, and triethanolamine, enhancement ratios varied between 1.5 and approximately 67-fold. However, when applied in solution in more lipophilic solvents, such as ethanol, isopropanol, oleyl alcohol, isopropyl myristate, or hexylene glycol, no enhancement was observed. These findings stress the importance of optimization of the delivery vehicle not only for the drug but also for the enhancer. Combinations of HPE-101 with cyclodextrins appear to be useful means to improve drug permeation across the skin (38). All of the available data, however, have been obtained

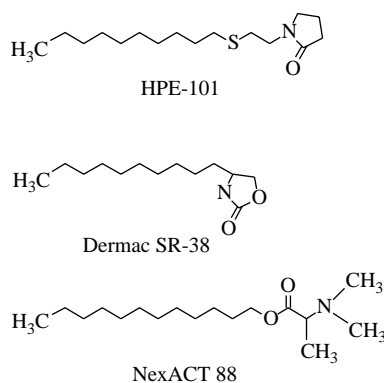


Fig. 2 Structures of the skin permeation enhancers HPE-101 (1-[2-(decylthio)ethyl]aza-cyclopentan-2-one), Dermac SR-38 (4-decyloxazolid-2-one), and NexACT 88 (dodecyl-*N,N*-dimethylamino isopropionate).

using hairless mice or other small laboratory animal skin. Small laboratory animals, especially the hairless mouse, can be uniquely sensitive to skin penetration enhancement and, as yet, the effectiveness of HPE-101 on human skin has not been reported.

Dermac SR-38 is one of a series of oxazolidinones, cyclic urethane compounds, evaluated as transdermal enhancers (39). The compound was designed to mimic natural skin lipids (such as ceramides), to be nonirritating, and to be rapidly cleared from the systemic circulation following absorption (40). In animal and human safety studies, Dermac SR-38 demonstrated a good skin tolerance (no observed irritancy or sensitization at levels of 1–10 wt%; moderate to severe irritation in rabbit at 100%), and a low degree of acute toxicity ($LD_{50}(\text{rat oral}) > 5.0 \text{ g/kg}$). The compound was evaluated for its ability to enhance the human skin permeation of diverse drugs from dermal and transdermal delivery systems. Data for minoxidil indicated an enhancer concentration-dependent effect for permeation enhancement. Dermac SR-38 was also found to enhance the skin retention of both retinoic acid when applied in Retin A cream, and dihydroxyacetone when applied in a hydrophilic cream (41).

NexACT 88 (dodecyl-*N,N*-dimethylamino isopropionate, DDAIP) is one of a series of dimethylamino alkanoates, reported to be biodegradable, which were developed as potential nontoxic skin permeation enhancers (42). Much of the early work was carried out using shed snake skin and it was found, using this model, that most of these compounds were equal to or more active

than Azone. Studies using human skin indicated that dodecyl-*N,N*-dimethylamino acetate (DDAA) was a more effective enhancer of absorption of propranolol hydrochloride and sotalol than was Azone (43). Structural optimization of the compounds led to the identification of the lead candidate DDAIP (44), which appeared to be more effective than DDAA. Mechanism of action studies indicated that the distribution of DDAIP in stratum corneum lipids was somewhat different to that of DDAA, suggesting that other interactions were contributing to the penetration enhancement effect. It is possible that in addition to its effect on stratum corneum lipids, DDAIP may interact with keratin and potentially increase stratum corneum hydration.

Other compounds have been identified and have undergone preliminary evaluation as potential skin penetration enhancers. The data are, however, very limited and these candidate enhancers are mentioned here solely for completeness. The biodegradable fatty acid esters of *N*-(2-hydroxyethyl)-2-pyrrolidone (decyl and oleyl) were synthesized and evaluated for enhancer activity using hairless mouse skin (45). Permeation of hydrocortisone was enhanced two-fold. The activity of *n*-pentyl-*N*-acetylprolinate as a skin permeation enhancer has been determined using human skin (46). The enhancement potential of several sunscreen agents, including padimate O, octyl salicylate, and octyl methoxycinnamate, has been evaluated with some success (47). Table 2 lists some of the more recent patent disclosures concerning skin permeation enhancement.

Table 2 Recent patents containing references to skin permeation enhancers

Patent no.	Enhancer	Assignee
DE 19646050	Neohesperidinedihydrochalcone	Labtec
WO 9818417	Fatty acid esters of lactic acid salts	Theratech
WO 9817315	Polyethyleneglycol monoalkyl ethers	Alza
EP 98480674900	Crotamiton	Lab D'Hygiene et de Dietetique
EP 98440792145	Levulic acid	Lohmann
EP 98270665755	Polyglycolized glyceride	SmithKline Beecham
US 5723114	Proton pump inhibitors	Cellegy
US 5814599	Sonophoresis and liposomes	MIT
US 5885565	Sterols and sterol esters	Cellegy
US 5882676	Acyl lactylates	Alza
US 5879701	Oleic acid dimer; neodecanoic acid	Cygnus
US 5942545	Dioxolanes	MacroChem
US 6001375	Polyoxyethylene cetyl ethers	Gist-brocades
US 6001390	Methyl laurate; glycerol monolaurate	Alza
US 6019988	Dual carrier systems	Bristol-Myers Squibb
WO 9922714	Alkyl-(<i>N,N</i> -disubstituted amino) esters	Nexmed

ADVERSE LOCAL REACTIONS AND STRATEGIES FOR REDUCTION

Human skin is exposed to an environment that contains a variety of natural and synthetic compounds. Inevitably, dermal contact, either accidental or deliberate, will be made with a wide number of these compounds, many of which have the potential for inducing adverse cutaneous reactions, such as irritation and sensitization.

Adverse skin responses to chemical exposure are variable, may be immediate or delayed, and be of long or short duration. They may also be classified as irritant or allergic. Dermatological and transdermal formulations contain a complex mixture of active and inactive ingredients and it is important to appreciate that the cause of any adverse reactions may be a formulation additive (excipient) and not necessarily an active compound. Thus, for example, it is well known to those developing transdermal delivery systems that the pressure sensitive adhesive used to produce intimate contact with the skin is more likely to be the source of any cutaneous reactivity than is the drug. A further complication is that many of the inactive ingredients used in topical pharmaceutical dosage forms may have the ability to alter the barrier function of the skin, which in turn may enhance the percutaneous penetration not only of other formulation ingredients but also of subsequent exposure, either accidentally or deliberately, to chemicals.

Acute toxic contact dermatitis may be induced by a single application of a toxic material. One local inflammatory skin reaction is characterized by erythema and oedema. This type of reaction occurs following contact with materials such as acids, alkalis, solvents, and cleansers and is rarely associated with topical application of medicinal or cosmetic products. In contrast, irritant contact dermatitis (a superficial nonimmunologically based reaction) may occur after repeated exposure to many substances, including topical pharmaceutical agents. The reaction is usually localized to the site of exposure and usually diminishes after the stimulus has been removed. Some materials can stimulate an immune response following an initial topical application. Any future exposure may result in an inflammatory immune reaction, an allergic contact dermatitis, or sensitization.

There are two main sensitization reactions—immediate and delayed hypersensitivity. Immediate type hypersensitivity is the result of antibody–allergen interaction occurring in the skin; the reaction that develops is known as allergic contact urticaria. Delayed type hypersensitivity is the result of cell-mediated immunity and is the most frequently reported side effect of topical

drugs. Both epidermal and dermal cells play pivotal roles in irritation and sensitization. Keratinocytes in the viable epidermis synthesize and secrete proinflammatory mediators and cytokines that activate the biochemical cascade leading to inflammation. Angiogenesis may also play a role in the inflammatory response but although it is possible that inhibitors of neovascularization may modulate the response to irritants, this hypothesis has yet to be fully investigated and is far from commercialization.

Before a compound can induce an adverse skin reaction upon dermal exposure, it has to penetrate into and permeate across the stratum corneum. Many of the strategies for the reduction of adverse reactions are based on reducing or modulating this process.

Retention of Compounds on the Skin (Reducing Penetration)

It is well established that a principle driving force for diffusion across the skin is the concentration, or more accurately the thermodynamic activity, of the permeant in the donor vehicle. Thermodynamic activity is reflected by the concentration of the permeant in the donor vehicle as a function of its saturation solubility within that medium. The closer to saturation concentration, the higher the thermodynamic activity and the greater the escaping tendency of the permeant from the vehicle. This principle has been extensively utilized in pharmaceutical formulation development in attempts to enhance percutaneous absorption of drugs but it is seldom used to reduce absorption. This is not surprising because for the most part, efficacy is improved by greater drug delivery to the therapeutic target. However, in some instances it is preferable to retain the active compound on the skin surface or within the outer layers of the stratum corneum (e.g., insect repellants and sunscreens) and this may be feasible using the vehicle thermodynamic activity approach.

Retardation of penetration can also be facilitated by the use of physical barriers such as protective creams and polymeric materials. There is evidence that emollient creams may be useful as skin barriers, although the results reported to date are somewhat variable and contradictory (48). More recently, however, convincing data have been presented that demonstrated that a barrier lotion containing the organoclay quaternium-18 bentonite (5%) was extremely effective in reducing the occurrence of allergic contact dermatitis to poison ivy and poison oak (49). Other polymeric materials are designed as skin compatible barrier materials. The latter materials include polyolprepolymer [Penederm Inc. (Bertek Pharmaceuticals)] and MacroDermTM (MacroChem Corp.). Both types of polymeric

material have been shown to reduce the skin penetration and permeation of potentially irritating compounds.

The polyolprepolymer products are presently used in several cosmetics, over-the-counter, and prescription topical pharmaceutical products. An extensive toxicology package that illustrates that the polymers are safe in use is available. The polyolprepolymers are polyalkylene glycol-based polyurethanes and are available as three types (PP-2, PP-14, and PP-15) with varying solubility properties and associated formulation characteristics. They are, therefore, amenable to incorporation into a variety of formulation types, including gels, lotions, and creams. Convincing evidence indicates that the polyolprepolymers are capable of reducing the irritant response to retinoids in both laboratory animal and human models. In addition evidence has shown that these polymers are capable of modifying the skin distribution of applied materials, such as alcohols, hydroxy acids and salicylates.

The MacroDerm is low to moderate (2000–25,000 Daltons) in molecular weight, symmetrical polymers that consist mainly of polyoxyethylene and polyoxypropylene linked by alkylene chains and end-capped by long alkyl groups (Fig. 3). The proportionality of the polymer chain mix renders hydrophobicity or hydrophilicity to the resultant block polymer. MacroDerm L, a lipophilic polymer consisting of two units each of 15 polyoxypropylene molecules linked by an alkylene group and end-capped by stearyl chains, has been shown to reduce the skin permeation of various sunscreen agents and pharmaceutically active compounds. A more hydrophilic MacroDerm, in which the polyoxypropylene chains of MacroDerm L

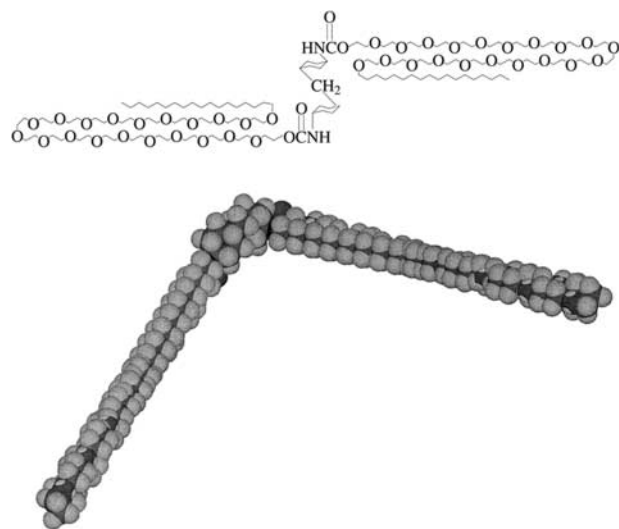


Fig. 3 Structure of MacroDerm SA20C (Courtesy of Dr. S. Krauser, MacroChem Corporation).

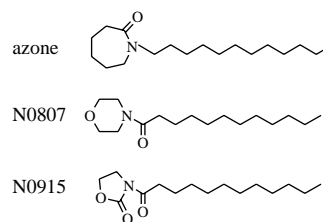


Fig. 4 Structures of Azone and the analogues N0807 and N0915.

have been replaced with polyoxyethylene chains, has been shown to increase the efficiency of a moisturizing formulation. The flexibility in the selection of moieties on the MacroDerm compounds suggests that they may be compatible with and easy to incorporate into a variety of topical formulation types. Given the structure of the MacroDerm and the nature of the constituent moieties (together with their similarities to existing compounds such as poloxamers and other nonionic surfactants), it is unlikely that there will be any safety issues associated with their use.

Reduction of Skin Diffusivity (Reducing Permeation)

The finding that an analogue (N0915) of the skin permeation enhancer Azone (Fig. 4) retarded the penetration of metronidazole through excised human skin, presumably by increasing the order (decreasing fluidity) of the intercellular stratum corneum lipids (50), led to a rationalization of the mechanism of enhancement at the molecular level (51). The structured nature of the stratum corneum intercellular lipid lamellae relies heavily upon lateral cooperative interactions between adjacent ceramide head groups. Enhancers such as Azone intercalate the matrix, but only provide a matching cooperative site on one side of the head group. On the other hand, the Azone analogue N0915, which has cooperative sites on both sides of the head group and retards permeation, while analogue N0807, which has electronegative sites on both sides of the head group, is a more effective enhancer than Azone. In addition, both Azone and N0807 have a preferred bent head group conformation which will disrupt packing of alkyl chains and increase the possibility of permeable defects. This basic concept was refined in the logical design of a series of molecules that were specifically targeted at retardation of skin permeation.

Experiments have been conducted on two of these compounds to determine their penetration retardation effect on two compounds, hydrocortisone and diethyl toluamide (DEET) (52). Fig. 5 shows that the retarders could reduce the permeation of hydrocortisone, while

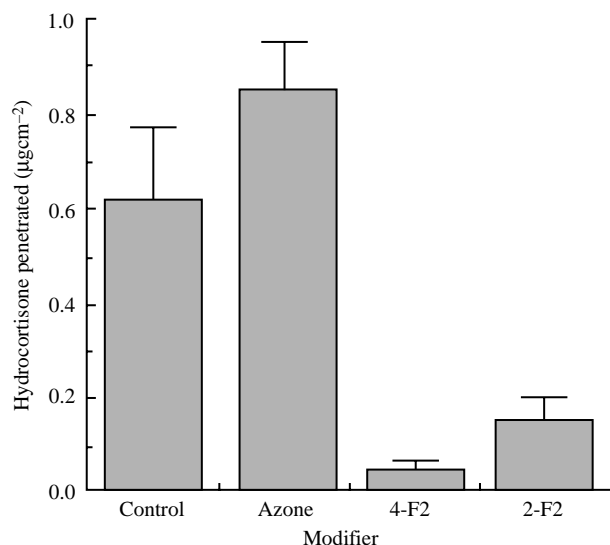


Fig. 5 Effects of Azone and the developmental retarding compounds 4-F2 and 2-F2 on permeation of hydrocortisone across human skin.

Azone, as expected, slightly increased permeation. Similarly, while Azone enhanced permeation of DEET, both retarder compounds reduced permeation. Although these data provide initial validation of the basic hypothesis for the mechanism of skin modulation, much remains to be accomplished before these compounds can be considered as viable strategies for irritation or sensitization reduction.

Miscellaneous Strategies for Modulating Skin Irritation

Many naturally occurring plant extracts are reputed to possess anti-irritant properties and have been recommended for use in cosmetic formulations. These include such diverse mixtures as tea tree oil, borage seed oil, Paraguay tea extract, Kola nut extract, oil of rosemary, and lavender oil. It is, however, difficult to standardize plant extracts and there may be a great deal of lot-to-lot variability in constituents. Understandably, this makes identification and isolation of any specific active constituent complex and laborious. The extracts may be oily or hydrophilic and contain compounds such as α -bisabolol, xanthines, polyphenols, and phytosterols. There is great potential in the use of plant extracts for irritation and sensitization reduction. This has been established within the cosmetic industry, and interest here has stimulated activity into reducing variability by more consistent cultivation techniques and more standardized extraction methods.

Strontium nitrate (Cosmederm-7™) has been shown to dramatically reduce the sensory irritation and erythema produced following application of a 70% free glycolic acid peel. It also has been shown to be effective in suppressing histamine-induced itch. Although the mode of action is unclear, the active principle appears to be elemental strontium in its free ionic form. It is postulated that the mode of action involves the potential ability of strontium to block neurogenic inflammation at type C neurons or nociceptors (53). It is interesting to note that the potential mechanism of action of strontium, involving nociceptor blockade, is also the mechanism by which capsaicin (the compound which renders chilli peppers “hot”) is believed to provide relief in several painful conditions, such as pruritis, postherpetic neuralgia, and diabetic neuropathy. Various analogues of capsaicin have been synthesized and evaluated for anti-inflammatory activity with varying degrees of success (54). However, it is important to appreciate that although the strategy of blocking sensory neurones will inhibit the sensation of pain, it will not inhibit the inflammation associated with an irritation reaction.

A combination of compounds, including oleic acid, a short chain length alcohol, and a glycol, all gelled with a carbomer, has been found to reduce inflammation associated with the topical application of several chemical species (CELLEDIRM, Cellegy Pharmaceuticals, Inc.) (55). This formulation has been patented with the major claim that the use of a combination of oleic acid and glycerol provides skin penetration enhancement while reducing irritation potential (when compared with, for example, oleyl alcohol). CELLEDIRM has been shown to reduce inflammation by up to 40% in animal models challenged with potent irritants or allergens. All the excipient ingredients within CELLEDIRM are either generally accepted as safe (GRAS) status or have been used in various pharmaceutical or cosmetic products. Cellegy Pharmaceuticals plans to utilize CELLEDIRM in the development of several pharmaceutical, dermatological, and transdermal formulations.

NOVEL DELIVERY SYSTEMS IN DERMATOLOGICAL AND TRANSDERMAL THERAPY

In addition to traditional dermal and transdermal delivery formulations, such as creams, ointments, gels, and patches, several other systems have been evaluated. In the pharmaceutical semisolid and liquid formulation area, these include sprays, foams, multiple emulsions, microemulsions, liposomal formulations, transfersomes,

niosomes, ethosomes, cyclodextrins, glycospheres, dermal membrane structures, and microsponges (56–58). Many of these novel systems use vesicles to modulate drug delivery. Novel transdermal formulations include soft patches, microneedles, and powder delivery systems.

Semisolid Vesicular Systems

Mezei first suggested that liposomes may be useful drug carrier systems for the local treatment of skin diseases (59). The suggestion was based on drug disposition data obtained following topical application of the steroid triamcinolone acetonide incorporated in phospholipid liposomes formulated as lotions or gels. Encapsulation of triamcinolone acetonide into liposomes resulted in a vehicle-dependent 4.5- to 4.9-fold increase in the amount of drug recovered from the epidermis. The work of Mezei suggested that application of the dermatological drugs in liposomal form compared to conventional formulations led to increased drug concentration in the skin and subcutaneous tissues and decreased biodisposition in plasma and remote sites. These encouraging early observations were followed by several confirmatory research and clinical investigations, most notably those of Weiner's group at the University of Michigan (60) and Korting's group at Ludwig-Maximilians University in Munich (61). Many other studies have indicated the potential of phospholipid liposomes to increase the skin content of topically applied drugs. Liposomes also have been prepared from lipid mixtures similar in composition to the stratum corneum intercellular lipid (62).

Another vesicle system that has been investigated for potential modification of skin permeation are niosomes (63). Niosomes are composed of nonionic surfactants, such as polyoxyethylene alkyl ethers, and may be prepared as single or multilamellar vesicles. Surfactants of this type are known to enhance skin permeation and this is likely to play a role in any modification of permeation using these vehicles. The effect of nonionic surfactant vesicles on the skin permeation of estradiol was shown to be dependent on the physical state of the niosome. On the other hands niosomes prepared from polyoxyethylene(3)stearyl ether and existing in the gel state did not increase estradiol permeation, and those prepared from polyoxyethylene(3)-lauryl ether and polyoxyethylene(10)oleyl ether, both existing as liquid crystalline vesicles, significantly enhanced transport. Further experiments in which the skin was pretreated with unloaded niosomes indicated that the enhanced transport of estradiol from drug-loaded vesicles was not wholly a result of surfactant-induced

penetration enhancement. The authors postulated that niosomes fused at the surface of the stratum corneum and generated high local concentrations of estradiol which resulted in increased thermodynamic activity of the permeant in the upper layers of the stratum corneum.

Vesicle systems, described as ethosomes composed of phospholipid, ethanol, and water, have been shown to enhance the transdermal delivery of minoxidil and testosterone when compared to more traditional formulations (Fig. 6) (57). The quantities of drug penetrating into and permeating through nude mouse skin *in vitro* were significantly greater from the ethosome systems than from appropriate control vehicles. Furthermore, when evaluated in rabbits *in vivo*, ethosomal transdermal patch systems produced higher testosterone plasma levels than a commercial patch. A tentative synergistic mode of action was proposed in which the ethanol disrupted the stratum corneum intercellular lipid, allowing the flexible ethosome to penetrate and possibly permeate the stratum corneum. The ethosome may subsequently fuse with skin lipids and release its drug content. The authors also point out that there may be a follicular contribution to the enhancement effect. It is interesting and important to note that there was no observed acute or cumulative irritancy (in rabbits) associated with the use of the ethosomal system.

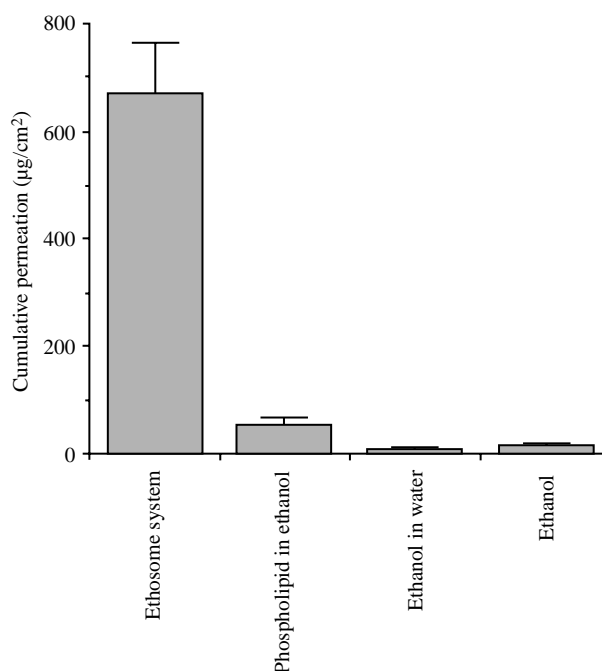


Fig. 6 Minoxidil permeation from ethosome systems and appropriate control vehicles. Data are expressed as the cumulative permeation across male nude mouse skin *in vitro* over 24 h. (Figure plotted from data given in Ref. 57.)

The precise mode of interaction between lipid vesicles and skin remains unclear. There is considerable doubt about the ability of whole vesicles to permeate intact stratum corneum. The majority of evidence suggests that vesicles can penetrate the outer cell layers of the stratum corneum where desmosomal linkages have become disrupted and presumably, the keratinocytes are less tightly bound and surrounded by a mixture of intercellular lipid and sebum. However, continuing diffusion of vesicles through the approximately 60 nm intercellular space of the deeper layers of the stratum corneum seems unlikely. Current thinking suggests that lipid vesicles fuse with endogenous lipid either on the surface or in the outermost layers of the stratum corneum. The fusion is followed by structural changes in the deeper layers of the stratum corneum, as evidenced by freeze-fracture electron microscopy and small angle x-ray scattering techniques. These structural changes are presumed to be the result of intercellular diffusion of vesicle lipid components (not intact vesicles) to the deeper layers, as well as interaction with and disruption of endogenous lipid lamellae. It is simple to postulate that this interaction/disruption of lipid lamellae will lead to an increase in skin permeation rates but this does not explain the observed increase in skin retention of permeants. Apparent increased skin retention may be an artefact from exogenous lipid depot formation on the skin surface. On the other hand, formations of lipid aggregates, possibly comprised of mixtures of endogenous and exogenous lipid, observed in deeper layers of the stratum corneum may provide a reservoir for topically applied drugs.

Cevc and Blume (64), however, suggested that it was possible for whole vesicles to cross intact stratum corneum. The basic premise for this hypothesis was the driving force provided by the osmotic gradient between the outer and inner layers of the stratum corneum and the development of specific mixes of lipids to form modified liposomes termed transfersomes. The requirement for the osmotic gradient to be maintained suggests that transfersomes will not function in occlusive conditions and careful formulation is necessary. Due to their unique structure (a mix of phosphatidyl choline, sodium cholate, and ethanol), transfersomes are reputed to be very flexible vesicles and capable of transporting their contents through the tortuous intercellular route of the stratum corneum. The application of the corticosteroids triamcinolone acetonide, dexamethasone, and hydrocortisone encapsulated in transfersomes resulted in more reliable site specificity for the drug and, therefore, less potential for adverse side effects (65).

Novel Transdermal Systems

In terms of overall composition, traditional transdermal patch systems have changed little in the past few years. The modifications that have been made are, for the most part, refinements of the materials used in their construction (Table 3). This is the case for the “soft” patches that consist of thin flexible films containing a known amount of drug (66). The soft patch is designed to be flexible and to conform to various body flexures.

Table 3 Recent patents containing references to transdermal materials

Patent no.	Material	Assignee
US 5783208	Pressure sensitive adhesive mixture	Theratech
WO 9820869	Electrotransport mechanism	Alza
US 5785688	Electromechanical gas generator	Ceramatec
WO 9737659	Crystallization inhibition	Sano
WO 9813099	Iontophoretic mechanism	Becton Dickinson
US 5753263	Liposomal formulation	Anticancer Inc.
WO 9813024	Hyaluronic acid	Hyal
US 5843979	Irritation/sensitization reduction	Bristol-Myers Squibb
WO 9832488	Irritation/sensitization reduction	Novartis
EP 98160665745	Pressure sensitive adhesive mixture	Lohmann
EP 98140716599	Crystallization inhibition	Lohmann
US 5843114	Skin perforation device	Samsung
US 5820875	Dual delivery rate device	Cygnus
US 5750138	Delayed onset of delivery	Westonbridge International
US 5713845	Laser-assisted drug delivery	ThermoLase
WO 9904838	Electromagnetic injection device	Boehringer Mannheim

Given the limitations imposed on transdermal systemic drug delivery by the barrier properties of the stratum corneum, new technologies have attempted to completely bypass this obstacle by either the creation of a physical conduit (microneedles) or direct powder delivery via compressed gas. The Alza Corporation technology (MacrofluxTM) comprises a patch system that contains a microprojection array designed to create superficial microchannels across the stratum corneum (67). When used in conjunction with their electrotransport system, the Macroflux system provides controlled *in vivo* delivery of therapeutic doses of antisense oligonucleotide, human growth hormone, and insulin. Similarly, the Radeon Inc. system consists of microfabricated microneedles that are 150 μm in length and may be either solid or hollow (68). The Radeon system was effective in enhancing by several orders of magnitude the human epidermal permeability of calcein and bovine serum albumin (Fig. 7) (69).

Transdermal powder delivery uses a supersonic flow of helium to accelerate drug particles to velocities sufficiently high to penetrate the stratum corneum (70). The needle-free injection system is capable of painlessly delivering drugs and vaccines in powder form into the skin. The amount of drug delivered is related to particle size, dose level, and the device operating power. Recent data demonstrated that the powder delivery system was capable of delivering salmon calcitonin across rabbit skin in a dose-dependent (but nonlinear) manner (71). A similar system (the HeliosTM gun system) was used to determine the effect of the dose regimen of a model drug

incorporated into poly-*L*-lactic acid microspheres of varying particle size (72). It was concluded that more frequent applications that contain lower amounts of the model drug generated a superior plasma profile than larger drug loadings at less frequent dosage intervals.

RECENT REGULATORY INITIATIVES

SUPAC-SS—Drug Release from Semisolid Formulations

Determination of the ability of a semisolid formulation to release a drug, the pattern of release and the rate at which this release occurs are important aspects of formulation development and optimization. However, it is also important to appreciate that the data obtained should not be overinterpreted. Release studies normally involve the measurement of drug diffusion out of a mass of formulation into a receiving medium that is separated from the formulation by a synthetic membrane (73). A detailed analysis of the data obtained in this type of experiment can be expected to generate invaluable data concerning the physical state of the drug in the formulation. For example, an examination of the early models and their refined updates derived to describe drug release from semisolids reveals that release patterns are different depending on whether the drug is present as a solution or suspension within the formulation (74). These subtle differences, together with differences in the rate of release, may be used to determine such parameters as drug diffusivity within the matrix of a formulation, the particle size of suspended drug, and the absolute solubility of a drug within a complex formulation (Fig. 8) (75, 76). Although it is generally agreed that drug release rate data cannot be used to predict skin permeation or bioavailability, release rate determinations are important for purposes other than formulation development and characterization.

The Food and Drug Administration (FDA) has issued a guidance document (SUPAC-SS Nonsterile Semisolid Dosage Forms, US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, May 1997) that recommends the use of *in vitro* drug release testing in the scale-up and postapproval changes for semisolids (SUPAC-SS). The FDA intends to promote the use of this test as a quality assurance tool to monitor minor differences in formulation composition or changes in manufacturing sites, but not at present as a routine batch-to-batch quality control test. Thus, the FDA is suggesting *in vitro* release rate data for Level 2 and Level 3 changes in formulation components

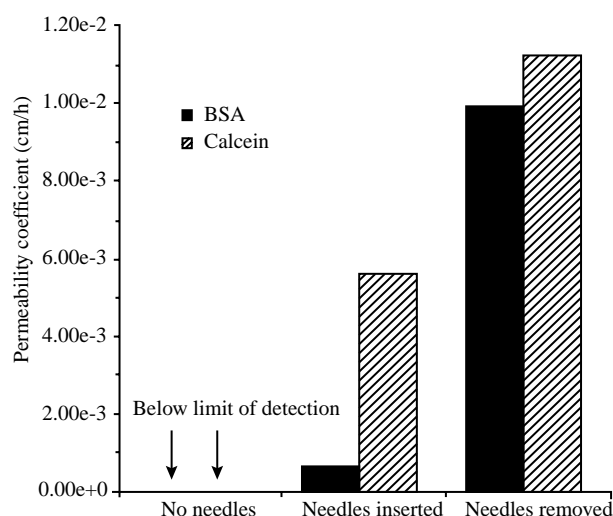


Fig. 7 Permeability of *in vitro* human epidermis to bovine serum albumin (BSA) and calcein. In the absence of microneedles, permeation was below the limit of detection. (Plotted from data given in Ref. 69.)

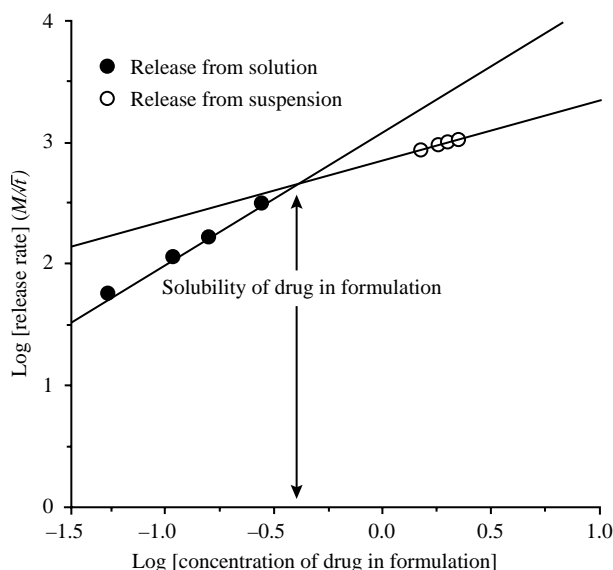


Fig. 8 Illustration of the use of release rates from semisolid preparations to determine drug solubility within a formulation. Data show the release rates of benzocaine from propylene glycol/water gels as a function of drug concentration in the formulation. (Redrawn from Ref. 75.)

and composition but such data are not required for a Level 1 change. In the former, the *in vitro* release rate of the new or modified formulation should be compared to a recent batch of the original formulation, and the 90% confidence limit should fall within the limits of 75–133%. Similarly, *in vitro* release testing is suggested for Level 2 changes in manufacturing equipment, processes, and scale-up, and Level 3 changes in manufacturing site. Recently, the use of *in vitro* testing as a quality assurance tool has been questioned, especially in the case of a hydrophilic formulation that contains the highly water soluble drug ammonium lactate (77). The method was found not to be specific enough to differentiate between small differences in drug loading or minor compositional and processing changes.

Bioequivalence of Dermatological Formulations

In practice, bioequivalence of dermatological dosage forms creates particular difficulties because it is often difficult to determine the very low blood levels of specific drugs following dermal application. The FDA has pioneered the use of alternative methods of evaluation including the investigation of dermatopharmacokinetics using the tape-stripping method. The use of *in vivo* skin stripping in dermatopharmacokinetic evaluation was the subject of an AAPS/FDA workshop concerning the bioequivalence of topical dermatological dosage forms

(Bethesda, MD, September, 1996). Although opinion was somewhat divided, it was concluded that stratum corneum tape stripping “may provide meaningful information for comparative evaluation of topical dosage forms” (78). Furthermore, it was established that a combination of dermatopharmacokinetic and pharmacodynamic data could provide sufficient proof of bioequivalence “in lieu of clinical trials.” However, much remains to be validated in skin stripping protocols. The *in vivo* tape stripping technique is based on the dermal reservoir principle developed by Rogier et al. (79). It is hypothesized that if a compound is applied to the skin for a limited time (for example 0.5 h) and then removed, the amount of drug in the upper layers of the stratum corneum will be predictive of the overall bioavailability of the compound. It follows that determination of the stratum corneum content of a permeating material following a short-term application will predict *in vivo* bioavailability from a corresponding administration protocol. Data obtained in studies of this type have shown reasonable predictability for several compounds.

An outline protocol for skin-stripping bioequivalence studies has been suggested (78). The basic protocol has two phases: uptake and elimination.

Uptake

1. Test and reference drug products are applied concurrently at multiple sites.
2. After exposure for a suitable time (determined by a pilot study), excess drug is removed by wiping three times with tissue or cotton swab.
3. The adhesive tape is applied with uniform pressure. The first strip is discarded (skin surface material). This is repeated if necessary to remove excess surface material.
4. Collect nine successive tape strips from the same site. If necessary collect more than nine strips.
5. Repeat the procedure for each site at designated time intervals.
6. Extract the drug from the combined tape strips for each time point and site and determine the content of drug using an appropriate validated analytical method.
7. Express the data as amount of drug per cm^2 of tape.

Elimination

1. Repeat steps 1, 2, and 3 “Uptake” phase.
2. After a predetermined time interval (e.g., 1, 3, 5, and 21 h postdrug removal) perform steps 4 through 7 of “Uptake” phase.

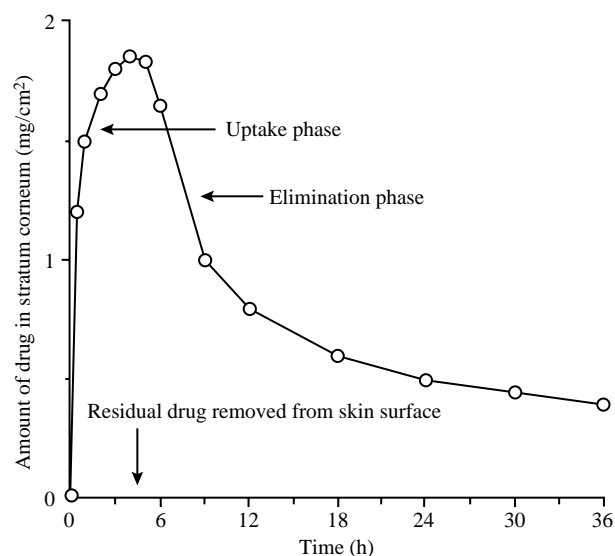


Fig. 9 Idealized dermatopharmacokinetic profile for a topically applied compound illustrating uptake and elimination phases. The amount of drug in the stratum corneum is determined by summing the total amount removed by tape stripping at each time interval.

The results are then expressed as the amount of drug recovered from the tape strips against time. Uptake and elimination phases are observed (Fig. 9) and bioavailability may be predicted from the area under the curve. There are several sources of variability in such studies, all of which must be considered in standard operating procedures. The major causes of concern in variability are:

1. drug application procedure
2. type of tape
3. size of tape
4. pressure applied by investigator
5. duration of application of pressure
6. drug removal procedure
7. drug extraction procedure
8. analytical methodology
9. temperature
10. relative humidity
11. skin type
12. skin surface uniformity

Other concerns have been expressed (80). These include the observation that vehicle components of the products to be evaluated may have different effects on the adhesive properties of the tape. In addition, it is important to appreciate that because the dermatopharmacokinetic bioequivalence studies will most likely be carried out on normal disease-free human volunteers, the generated data may show little resemblance to the actual drug distribution

within the stratum corneum of patients. Nonetheless, following further validation, the technique will have several advantages. For example, basic pharmacokinetic parameters, such as AUC, C_{max} , T_{max} , and half-life, may be approximated from the data obtained. In addition, the approach could be applicable to all types of topical preparation.

Recently, Pershing (81) reviewed much of the extensive validation of typical in vivo skin stripping techniques. Such variables as the test region anatomical site, individual investigator technique, adhesive systems and product dose, application, and removal techniques were discussed. It was concluded that "careful validation of adequate removal of residual applied product, the collection of skin stripping samples and a sensitive analytical assay were critical to the appropriate interpretation of the results" (81). Further details of the dermatological drug product bioequivalence guidelines, including the proposed protocol, may be obtained from the FDA draft guidance document (Topical Dermatological Drug Product NDAs and ANDAs—In Vivo Bioavailability, Bioequivalence, In Vitro Release, and Associated Studies, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, June 1998).

CONCLUDING REMARKS

The title of this article, *Drug Delivery—Topical and Transdermal Routes*, describes a field of pharmaceutical sciences that has expanded rapidly over the past 30 years. For example, it was only a few years ago that noninvasive transcutaneous immunization was only available on the 'Starship Enterprise.' Now the technology is available (82). To include all recent developments would require much more space than is feasible within the scope of this encyclopedia. Quite selfishly, therefore, I have elected to cover those subjects in which *I am* most interested. In partial mitigation, I have provided a reasonably complete bibliography and have attempted to point the reader who wishes to explore in more depth subjects not covered herein toward the most relevant recent references. Topical drug delivery, for either dermatological or transdermal therapy, is a fascinating subject, made more so by the nature of the skin. The more we understand the molecular biology of this unique organ, the more able we will be to fix it when it goes wrong. Similarly, a deeper understanding of skin barrier morphology will allow us to develop strategies to modify its permeation properties.

These are fundamental issues and improvement of our knowledge will increase our capabilities as drug delivery scientists. In this section, I have attempted to outline recent developments in the development of the stratum corneum and some of the methods used to modify the inherent barrier properties of this unique membrane. Some strategies for reducing adverse dermatological events associated with topical therapy have been discussed. Novel delivery systems were outlined and recent regulatory initiatives, which are planned to make life easier for the pharmaceutical industry, were described. I hope that my comments will provide the drug delivery specialist with some insights borne of experience and the experienced topical formulator with some alternative concepts in the field of dermatological and transdermal product development.

REFERENCES

- Walters, K.A., Hadgraft, J., Eds. *Pharmaceutical Skin Penetration Enhancement*; Marcel Dekker, Inc.: New York, 1993.
- Smith, E.W., Maibach, H.I., Eds. *Percutaneous Penetration Enhancers*; CRC Press: Boca Raton FL, 1995.
- Schaefer, H., Redelmeier, T.E., Eds. *Skin Barrier. Principles of Percutaneous Absorption*; Karger: Basel, 1996.
- Ghosh, T.K., Pfister, W.R., Yum, S.I., Eds. *Transdermal and Topical Drug Delivery Systems*; Interpharm Press: Buffalo Grove, 1997.
- Banga, A.K. *Electrically Assisted Transdermal and Topical Drug Delivery*; Taylor & Francis: London, 1998.
- Roberts, M.S., Walters, K.A., Eds. *Dermal Absorption and Toxicity Assessment*; Marcel Dekker, Inc.: New York, 1998.
- Bronaugh, R.L., Maibach, H.I., Eds. *Percutaneous Absorption, Drugs, Cosmetics, Mechanisms, Methodology*, 3rd Ed.; Marcel Dekker, Inc.: New York, 1999.
- Lodén, M., Maibach, H.I., Eds. *Dry Skin and Moisturizers. Chemistry and Function*; CRC Press: Boca Raton, FL, 2000.
- Kydonieus, A.F., Wille, J.J., Eds. *Biochemical Modulation of Skin Reactions, Transdermals, Topicals, Cosmetics*; CRC Press: Boca Raton, FL, 2000.
- Lai, P.M.; Robers, M.S. Iontophoresis. *Dermal Absorption and Toxicity Assessment*; Roberts, M.S., Walters, K.A., Eds.; Marcel Dekker, Inc.: New York, 1998; 371–414.
- Kost, J.; Mitragotri, S.; Langer, R. Phonophoresis. *Percutaneous Absorption, Drugs, Cosmetics, Mechanisms, Methodology*, 3rd Ed.; Bronaugh, R.L., Maibach, H.I., Eds.; Marcel Dekker, Inc.: New York, 1999; 615–631.
- Guy, R.H.; Kalia, Y.N.; Delgado-Charro, M.B.; Merino, V.; López, A.; Marro, D. Iontophoresis Electrorepulsion and Electroosmosis. *J. Controlled Release* **2000**, *64*, 129–132.
- Ilic, L.; Gowrishankar, T.R.; Vaughan, T.E.; Herndon, T.O.; Weaver, J.C. Spatially Constrained Skin Electroporation with Sodium Thiosulfate and Urea Creates Transdermal Microconduits. *J. Controlled Release* **1999**, *61*, 185–202.
- Lombry, C.; Dujardin, N.; Pr  at, V. Transdermal Delivery of Macromolecules Using Skin Electroporation. *Pharm. Res.* **2000**, *17*, 32–37.
- Pugh, W.J.; Hadgraft, J.; Roberts, M.S. Physicochemical Determinants of Stratum Corneum Permeation. *Dermal Absorption and Toxicity Assessment*; Roberts, M.S., Walters, K.A., Eds.; Marcel Dekker, Inc.: New York, 1998; 245–268.
- Roberts, M.S.; Anissimov, Y.G.; Gonsalvez, R.A. Mathematical Models in Percutaneous Absorption. *Percutaneous Absorption, Drugs, Cosmetics, Mechanisms, Methodology*, 3rd Ed.; Bronaugh, R.L., Maibach, H.I., Eds.; Marcel Dekker, Inc.: New York, 1999; 3–55.
- Eckert, E.L. Structure, Function, and Differentiation of the Keratinocyte. *Physiol. Rev.* **1989**, *69*, 1316–1346.
- Dinarello, C.A. Interleukin-1 Receptors and Signal Transduction. *Biochemical Modulation of Skin Reactions*; Kydonieus, A.F., Wille, J.J., Eds.; CRC Press: Boca Raton, FL, 2000; 173–187.
- Kondo, S. The Roles of Keratinocyte-Derived Cytokines in the Epidermis and their Possible Responses to UVA-Irradiation. *J. Invest. Dermatol. Symp. Proc.* **1999**, *4*, 177–183.
- Landmann, L. The Epidermal Permeability Barrier. *Anat. Embryol.* **1988**, *178*, 1–13.
- Wertz, P.W.; Downing, D.T. Glycolipids in Mammalian Epidermis Structure and Function in the Water Barrier. *Science* **1982**, *217*, 1261–1262.
- Wertz, P.W.; Downing, D.T.; Freinkel, R.K.; Traczyk, T.N. Sphingolipids of the Stratum Corneum and Lamellar Granules of Fetal Rat Epidermis. *J. Invest. Dermatol.* **1984**, *83*, 193–195.
- Menon, G.K.; Feingold, K.R.; Elias, P.M. Lamellar Body Secretory Response to Barrier Disruption. *J. Invest. Dermatol.* **1992**, *98*, 279–289.
- Nemes, Z.; Steinert, P.M. Bricks and Mortar of the Epidermal Barrier. *Exp. Mol. Med.* **1999**, *31*, 5–19.
- Lazo, N.D.; Meine, J.G.; Downing, D.T. Lipids are Covalently Attached to Rigid Corneocyte Protein Envelopes Existing Predominantly as β -Sheets a Solid-State Nuclear Magnetic Resonance Study. *J. Invest. Dermatol.* **1995**, *105*, 296–300.
- Downing, D.T.; Lazo, N.D. Lipid and Protein Structures in the Permeability Barrier. *Dry Skin and Moisturizers*; Lod  n, M., Maibach, H.I., Eds.; CRC Press, Boca Raton, FL, 2000; 39–44.
- Behne, M.; Uchida, Y.; Seki, T.; Ortiz de Montellano, P.; Elias, P.M.; Holleran, W.M. Omega-Hydroxyceramides are Required for Corneocyte Lipid Envelope (CLE) Formation and Normal Epidermal Permeability Barrier Function. *J. Invest. Dermatol.* **2000**, *114*, 185–192.
- Suhonen, T.M.; Bouwstra, J.A. Urti, A. Chemical Enhancement of Percutaneous Absorption in Relation to Stratum Corneum Structural Alterations. *J. Controlled Release* **1999**, *59*, 149–161.
- Wertz, W.P.; Miethke, M.C.; Long, S.A.; Strauss, J.S.; Downing, D.T. The Composition of the Ceramides from Human Stratum Corneum and from Comedones. *J. Invest. Dermatol.* **1985**, *84*, 410–412.

30. Stewart, M.E.; Downing, D.T. A New 6-Hydroxy-4-Sphinganine-Containing Ceramide in Human Skin. *J. Lipid. Res.* **1999**, *40*, 1434–1439.
31. Pilgram, G.S.K.; Engelsma-van Pelt, A.M.; Bouwstra, J.A.; Koerten, H.K. Electron Diffraction Provides New Information on Human Stratum Corneum Lipid Organization Studied in Relation to Depth and Temperature. *J. Invest. Dermatol.* **1999**, *113*, 403–409.
32. Yano, T.; Nakagawa, A.; Masayoshi, T.; Noda, K. Skin Permeability of Non-Steroidal Antiinflammatory Drugs in Man. *Life Sci.* **1986**, *39*, 1043–1050.
33. Vecchia, B.E.; Stephens, K.R.; Bunge, A.L. Comparison of Permeability Coefficients for Excised Skin from Humans and Animals. *Pharm. Res.* **1998**, *15*, S-380.
34. Franz, T.J.; Lehman, P.A.; Franz, S.F.; North-Root, H.; Demetruilas, J.L.; Kelling, C.K.; Moloney, S.J.; Gettings, S.D. Percutaneous Penetration of N-Nitroso-Diethanolamine through Human Skin (In Vitro) Comparison of Finite and Infinite Dose Application from Cosmetic Vehicles. *Fundam. Appl. Toxicol.* **1993**, *21*, 213–221.
35. Walters, K.A.; Brain, K.R.; Howes, D.; James, V.J.; Kraus, A.L.; Teetsel, N.M.; Toulon, M.; Watkinson, A.C.; Gettings, S.D. Percutaneous Penetration of Octyl Salicylate from Representative Sunscreen Formulations through Human Skin In Vitro. *Fd. Chem. Toxicol.* **1997**, *35*, 1219–1225.
36. Gauthier, E. *The Dioxolanes A New Class of Percutaneous Absorption Enhancers*; Ph.D. Thesis, University of Paris, XI, 2000.
37. Yano, T.; Higo, N.; Fukuda, K.; Tsuji, M.; Noda, K.; Otagiri, M. Further Evaluation of a New Penetration Enhancer, HPE-101. *J. Pharm. Pharmacol.* **1992**, *45*, 775–778.
38. Adachi, H.; Irie, T.; Uekama, K.; Manako, T.; Yano, T.; Saita, M. Combination Effects of O-Carboxymethyl-O-Ethyl- β -Cyclodextrin and Penetration Enhancer HPE-101 on Transdermal Delivery of Prostaglandin E_1 in Hairless Mice. *Eur. J. Pharm. Sci.* **1993**, *1*, 117–123.
39. Rajadhyaksha, V.J. Oxalodione Penetration Enhancing Compounds, US Patent 4,960,771, 1990
40. Pfister, W.R.; Rajadhyaksha, V.J. Oxazolidinones. A New Class of Cyclic Urethane Transdermal Enhancer (CUTE). *Proc. Intl. Symp. Control. Rel. Bioact. Mater.* **1997**, *24*, 709–710.
41. Pfister, W.R.; Rajadhyaksha, V.J. Oxazolidinones A New Class of Cyclic Urethane Transdermal Enhancer (CUTE). *Pharm. Res.* **1995**, *12*, S280.
42. Wong, O.; Huntington, A.; Nishihata, T.; Rytting, J.H. New Alkyl N,N-Dialkyl-Substituted Amino Acetates as Transdermal Penetration Enhancers. *Pharm. Res.* **1989**, *6*, 286–295.
43. Wongpayapkul, L.; Chow, D. Comparative Evaluation of Various Enhancers on the Transdermal Permeation of Propranolol Hydrochloride. *Pharm. Res.* **1991**, *8*, S140.
44. Büyüktimkin, S.; Büyüktimkin, N.; Rytting, J.H. Synthesis and Enhancing Effect of Dodecyl 2-(N,N-Dimethylamino)-Propionate (DDAIP) on the Transepidermal Delivery of Indomethacin, Clonidine, and Hydrocortisone. *Pharm. Res.* **1993**, *10*, 1632–1637.
45. Lambert, W.J.; Kudla, R.J.; Holland, J.M.; Curry, J.T. A Biodegradable Transdermal Penetration Enhancer Based on N-(2-Hydroxyethyl)-2-Pyrrolidone I. Synthesis and Characterization. *Int. J. Pharm.* **1993**, *95*, 181–192.
46. Harris, W.T.; Tenjarla, S.N.; Holbrook, J.M.; Smith, J.; Mead, C.; Entekin, J. N-Pentyl N-Acetylproline a New Skin Penetration Enhancer. *J. Pharm. Sci.* **1995**, *84*, 640–642.
47. Morgan, T.M.; O'Sullivan, H.M.M.; Reed, B.L.; Finnin, B.C. Transdermal Delivery of Estradiol in Postmenopausal Women with a Novel Topical Aerosol. *J. Pharm. Sci.* **1998**, *87*, 1226–1228.
48. Pigatto, P.D.; Bigardi, A.S.; Legori, A.; Altomare, F.G.; Finzi, A.F. Are Barrier Creams any Use in Contact Dermatitis? *Contact. Dermatitis.* **1992**, *26*, 197–198.
49. Marks, J.G.; Fowler, J.F.; Sherertz, E.F.; Rietschel, R.L. Prevention of Poison Ivy and Poison Oak Allergic Contact Dermatitis by Quaternium-18 Bentonite. *J. Am. Acad. Dermatol.* **1995**, *33*, 212–216.
50. Hadgraft, J.; Peck, J.; Williams, D.G.; Pugh, W.J.; Allan, G. Mechanism of Action of Skin Penetration Enhancers/Retarders Azone and Analogues. *Int. J. Pharm.* **1996**, *141*, 17–25.
51. Brain, K.R.; Walters, K.A. Molecular Modeling of Skin Permeation Enhancement by Chemical Agents. *Pharmaceutical Skin Penetration Enhancement*; Walters, K.A., Hadgraft, J., Eds.; Marcel Dekker, Inc.: New York, 1993; 389–416.
52. Brain, K.R.; Green, D.M.; James, V.J.; Walters, K.A.; Watkinson, A.C.; Allan, G.; Hammond, J. Preliminary Evaluation of Novel Penetration Retarders. *Prediction of Percutaneous Penetration*; Brain, K.R., James, V.J., Walters, K.A., Eds.; STS Publishing: Cardiff, 1996; 4b, 131–132.
53. Hahn, G.S. Strontium is a Selective and Potent Inhibitor of Sensory Irritation (Itch, Burn and Sting) and Neurogenic Inflammation. *Perspectives in Percutaneous Penetration*; Brain, K.R., Walters, K.A., Eds.; STS Publishing: Cardiff, 2000; 7a, 10.
54. Janusz, J.M.; Buckwalter, B.L.; Young, P.A.; LaHann, T.R.; Farmer, R.W.; Kasting, G.B.; Loomans, M.E.; Kerckaert, G.A.; Maddin, C.S.; Bertram, E.F.; Bohne, R.L.; Cupps, T.L.; Milstein, J.R. Vanilloids. 1. Analogs of Capsaicin with Antinociceptive and Antiinflammatory Activity. *J. Med. Chem.* **1993**, *36*, 2595–2604.
55. Thornfeldt, C.R.; Elias, P.M.; Grayson, S. US Patent 5,723,114, 1998.
56. Rogers, K. Controlled Release Technology and Delivery Systems. *Cosmet. Toiletries* **1999**, *114* (5), 53–60.
57. Touitou, E.; Dayan, N.; Bergelson, L.; Godin, B.; Eliaz, M. Ethosomes—Novel Vesicular Carriers for Enhanced Delivery: Characterization and Skin Penetration Properties. *J. Controlled Release* **2000**, *65*, 403–418.
58. Tao, L. Skin Delivery from Lipid Vesicles. *Cosmet. Toiletries* **2000**, *115* (4), 43–50.
59. Mezei, M.; Gulasekharan, V. Liposomes—A Selective Drug Delivery System for the Topical Route of Administration Lotion Dosage Forms. *Life Sci.* **1980**, *26*, 1473–1477.
60. du Plessis, J.; Ramachandran, C.; Weiner, N.; Müller, D.G. The Influence of Particle Size of Liposomes on the Deposition of Drug into Skin. *Int. J. Pharm.* **1994**, *103*, 277–282.
61. Schmid, M.H.; Korting, H.C. Therapeutic Progress with Topical Liposome Drugs for Skin Disease. *Adv. Drug Delivery. Rev.* **1996**, *18*, 335–342.

62. Fresta, M.; Puglisi, G. Corticosteroid Dermal Delivery with Skin-Lipid Liposomes. *J. Controlled Release* **1997**, *44*, 141–151.
63. Schreier, H.; Bouwstra, J. Liposomes and Niosomes as Topical Drug Carriers: Dermal and Transdermal Drug Delivery. *J. Controlled Release* **1994**, *30*, 1–15.
64. Cevc, G.; Blume, G. Lipid Vesicles Penetrate into Intact Skin Owing to the Transdermal Osmotic Gradients and Hydration Force. *Biochim. Biophys. Acta* **1991**, *1104*, 226–232.
65. Cevc, G.; Blume, G.; Schätzlein, A. Transfersomes-Mediated Transepidermal Delivery Improves the Regional Specificity and Biological Activity of Corticosteroids In Vivo. *J. Controlled Release* **1997**, *45*, 211–226.
66. Csóka, G.; Dredán, J.; Marton, S.; Antal, I.; Rácz, I. Evaluation of Different Mathematical Methods Describing Drug Liberation from New, Soft-Patch Type Matrix Systems. *Pharm. Dev. Technol.* **1999**, *4*, 291–294.
67. Daddona, P.E. Minimally Invasive Transdermal Drug Delivery. *Perspectives in Percutaneous Penetration*; Brain, K.R., Walters, K.A., Eds.; STS Publishing: Cardiff, 2000; *7a*, 5.
68. Henry, S.; McAllister, D.V.; Allen, M.G.; Prausnitz, M.R. Microfabricated Microneedles A Novel Approach to Transdermal Drug Delivery. *J. Pharm. Sci.* **1998**, *87*, 922–925.
69. Prausnitz, M.R.; Allen, M.G.; Davis, S.; Kaushik, S.; Jett, E.; Kaye, S.; McAllister, D.V. Microfabricated Microneedles for Transdermal Drug Delivery. *Perspectives in Percutaneous Penetration*; Brain, K.R., Walters, K.A., Eds.; STS Publishing: Cardiff, 2000; *7a*, 4.
70. Sarphie, D.F.; Johnson, B.; Cormier, M.; Burkoth, T.L.; Bellhouse, B.J. Bioavailability Following Transdermal Powdered Delivery (TPD) of Radiolabeled Insulin to Hairless Guinea Pigs. *J. Controlled Release* **1997**, *47*, 61–69.
71. Sweeney, P.A.; Topham, S.J.; Zuurbier, R.J.; McCrossin, L.E.; Muddle, A.G.; Longridge, D.J. Effects of Dose Escalation on Dermal Powderject® Delivery of Salmon Calcitonin to Conscious Rabbits. *Proc. Intl. Symp. Contr. Rel. Bioact. Mater.* **1999**, *26*, 188–189.
72. Uchida, M.; Natsume, H.; Kobayashi, D.; Morimoto, Y. Effect of Dose, Particle Size and Helium Gas Pressure on Transdermal Powder Delivery by Helios™ Gun System. *Proc. Intl. Symp. Contr. Rel. Bioact. Mater.* **1999**, *26*, 483–484.
73. Chattaraj, S.C.; Kanfer, I. Release of Acyclovir from Semi-Solid Dosage Forms A Semi-Automated Procedure Using a Simple Plexiglass Flow-Through Cell. *Int. J. Pharm.* **1995**, *125*, 215–222.
74. Bunge, A.L. Release Rates from Topical Formulations Containing Drugs in Suspension. *J. Controlled Release* **1998**, *52*, 141–148.
75. Caetano, P.A.; Flynn, G.L.; Farinha, A.R.; Toscano, C.F.; Campos, R.C. The In Vitro Release Test as a Means to Obtain the Solubility and Diffusivity of Drugs in Semisolids. *Proc. Intl. Symp. Contr. Rel. Bioact. Mater.* **1999**, *26*, 375–376.
76. Flynn, G.L.; Shah, V.P.; Tenjarla, S.N.; Corbo, M.; DeMagistris, D.; Feldman, T.G.; Franz, T.J.; Miran, D.R.; Pearce, D.M.; Sequeira, J.A.; Swarbrick, J.; Wang, J.C.T.; Yacobi, A.; Zatz, J.L. Assessment of Value and Applications of In Vitro Testing of Dermatological Drug Products. *Pharm. Res.* **1999**, *16*, 1325–1330.
77. Kril, M.B.; Parab, P.V.; Genier, S.E.; DiNunzio, J.E.; Alessi, D. Potential Problems Encountered with SUPAC-SS and the In Vitro Release Testing of Ammonium Lactate Cream. *Pharm. Tech.* **1999 (March)**, 164–174.
78. Shah, V.P.; Flynn, G.L.; Yacobi, A.; Maibach, H.I.; Bon, C.; Fleischer, N.M.; Franz, T.J.; Kaplan, S.A.; Kawamoto, J.; Lesko, L.J.; Marty, J.-P.; Pershing, L.K.; Schaefer, H.; Sequeira, J.A.; Shrivastava, S.P.; Wilkin, J.; Williams, R.L. AAPS/FDA Workshop Report Bioequivalence of Evaluation of Bioequivalence. *Pharm. Res.* **1998**, *15*, 167–171.
79. Rougier, A.; Dupuis, D.; Lotte, C.; Maibach, H.I. Stripping Method for Measuring Percutaneous Absorption in vivo. *Percutaneous Absorption*, 3rd Ed.; Bronaugh, R.L., Maibach, H.I., Eds.; Marcel Dekker, Inc.: New York, 1999; 375–394.
80. Surber, C.; Schwarb, F.P.; Smith, E.W. Tape-Stripping Technique. *Percutaneous Absorption*, 3rd Ed.; Bronaugh, R.L., Maibach, H.I., Eds.; Marcel Dekker, Inc.: New York, 1999; 395–409.
81. Pershing, L.K. Dermatopharmacokinetics for Assessing Bioequivalence of Topically Applied Products in Human Skin. *Cosmet. Toiletries* **2000**, *115 (5)*, 43–51.
82. O'Farrel, C. Transcutaneous Immunization More Effective Non-Invasive Vaccines. *Inn. Pharm. Tech.* **2000**, *(1)*, www.iptonline.com.

BIBLIOGRAPHY

- Banga, A.K. *Electrically Assisted Transdermal and Topical Drug Delivery*; Taylor & Francis: London, 1998.
- Barry, B.W. *Dermatological Formulations: Percutaneous Absorption*; Marcel Dekker, Inc.: New York, 1983.
- Brain, K.R., James, V.J., Walters, K.A., Eds. *Prediction of Percutaneous Penetration*; STS Publishing: Cardiff, 1993; 3b.
- Brain, K.R., James, V.J., Walters, K.A., Eds. *Prediction of Percutaneous Penetration*; STS Publishing: Cardiff, 1996; 4b.
- Brain, K.R., James, V.J., Walters, K.A., Eds. *Perspectives in Percutaneous Penetration*; STS Publishing: Cardiff, 1998; 5b.
- Bronaugh, R.L., Maibach, H.I., Eds. *Vitro Percutaneous Absorption: Principles, Fundamentals, and Applications*; CRC Press: Boca Raton, FL, 1991.
- Bronaugh, R.L., Maibach, H.I., Eds. *Percutaneous Absorption*, 3rd Ed.; Marcel Dekker, Inc.: New York, 1999.
- Chien, Y.W., Ed. *Transdermal Controlled Systemic Medications*; Marcel Dekker, Inc.: New York, 1987.
- de Boer, A.G., Ed. *Drug Absorption Enhancement*; Harwood Academic Publishers: Switzerland, 1994.
- Ghosh, T.K., Pfister, W.R., Yum, S.I., Eds. *Transdermal and Topical Drug Delivery Systems*; Interpharm Press: Buffalo Grove, 1997.
- Hadgraft, J., Guy, R.H., Eds. *Transdermal Drug Delivery Developmental Issues and Research Initiatives*; Marcel Dekker, Inc.: New York, 1989.

- Hsieh, D.S., Ed. *Drug Permeation Enhancement—Theory and Applications*; Marcel Dekker, Inc.: New York, 1994.
- Kemppainen, B.W., Reifenrath, W.G., Eds. *Methods for Skin Absorption*; CRC Press: Boca Raton, FL, 1990.
- Kydonieus, A.F., Wille, J.J., Eds. *Biochemical Modulation of Skin Reactions, Transdermals, Topicals, Cosmetics*; CRC Press: Boca Raton, FL, 2000.
- Lieberman, H.A., Rieger, M.M., Banker, G.S., Eds. *Pharmaceutical Dosage Forms: Disperse Systems*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1998; 1–3.
- Lodén, M., Maibach, H.I., Eds. *Dry Skin and Moisturizers, Chemistry and Function*; CRC Press: Boca Raton, FL, 2000.
- Marzulli, F.N., Maibach, H.I., Eds. *Dermatotoxicology*, 5th Ed.; Taylor & Francis: Washington, DC, 1996.
- Osborne, D.W., Amann, A.H., Eds. *Topical Drug Delivery Formulations*; Marcel Dekker, Inc.: New York, 1990.
- Potts, R.O., Guy, R.H., Eds. *Mechanisms of Transdermal Drug Delivery*; Marcel Dekker, Inc.: New York, 1997.
- Roberts, M.S., Walters, K.A., Eds. *Dermal Absorption and Toxicity Assessment*; Marcel Dekker, Inc.: New York, 1998.
- Schaefer, H.; Redelmeier, T.E. *Skin Barrier—Principles of Percutaneous Absorption*; Karger: Basel, 1996.
- Scott, R.C., Guy, R.H., Hadgraft, J., Eds. *Prediction of Percutaneous Penetration*; IBC Technical Services: London, 1990; 1.
- Scott, R.C., Guy, R.H., Hadgraft, J., Boddé, H.E., Eds. *Prediction of Percutaneous Penetration*; IBC Technical Services: London, 1991; 2.
- Shah, V.P., Maibach, H.I., Eds. *Topical Drug Bioavailability, Bioequivalence, and Penetration*; Plenum Press: New York, 1993.
- Shroot, B., Schaefer, H., Eds. *Skin Pharmacokinetics*; Karger Publishing: Basle, 1987.
- Sjöblom, J., Ed. *Emulsions and Emulsion Stability*; Marcel Dekker, Inc.: New York, 1996.
- Smith, E.W., Maibach, H.I., Eds. *Percutaneous Penetration Enhancers*; CRC Press: Boca Raton, FL, 1995.
- Tyle, P., Ed. *Drug Delivery Devices—Fundamentals and Applications*; Marcel Dekker, Inc.: New York, 1988.
- Walters, K.A., Hadgraft, J., Eds. *Pharmaceutical Skin Penetration Enhancement*; Marcel Dekker, Inc.: New York, 1993.
- Zatz, J.L., Ed. *Skin Permeation, Fundamentals and Application*; Allured Publishing Corp. Wheaton, MD, 1993.

DRUG DELIVERY—RECTAL ROUTE

J. Howard Rytting

The University of Kansas, Lawrence, Kansas, U.S.A.

Joseph A. Fix

Yamanouchi Pharma Technologies, Inc., Palo Alto, California, U.S.A.

INTRODUCTION

Drug Delivery Routes

Although administration via the peroral route is the most commonly targeted goal of new drug and dosage form research and development, oral administration is not always feasible or desirable. The potential for oral dosage form development is severely limited for active agents that are poorly absorbed in the upper gastrointestinal (GI) tract and unstable to proteolytic enzymes. Some agents cause local stomach or upper GI irritation or require doses in excess of 500 mg. Certain patient populations, notably children, the elderly, and those with swallowing problems, are often difficult to treat with oral tablets and capsules. Additionally, treatment of some diseases is best achieved by direct administration near the affected area, particularly with diseases involving ophthalmic, otic, dermal, oral cavity, and anorectal tissues. Although oral administration can be used for drugs targeted for some of these diseased tissues, exposure of the entire body compartment to the administered drug is inefficient and can lead to undesired adverse effects. Rectal drug administration is amenable, however, to both local and systemic drug delivery. It has been effectively utilized to treat local diseases of the anorectal area as well as to deliver drugs systemically as an alternative to oral administration. In this article, factors influencing the utilization of rectal drug administration as well as the advantages and disadvantages of this approach, are discussed. A survey of the pharmaceutical literature for the last 5–7 years suggests that this administration route has not received a great deal of attention in the pharmaceutical community. Relatively few articles have appeared in the literature (1–4) and new products have not captured a significant segment of the pharmaceutical market. Although rectal drug administration is unlikely to ever become a commonly accepted route of administration, the utilization of this technology for particular applications and therapeutic problems offers an alternative delivery route which can be successfully applied in drug therapy.

Rectal Formulations Available

Solid suppositories

Solid suppositories are the most common dosage form used for rectal drug administration and represent greater than 98% of all rectal dosage forms. Typically, these are torpedo-shaped dosage forms composed of fatty bases (low-melting) or water-soluble bases (dissolving) which vary in weight from 1 g (children) to 2.5 g (adult). The composition is largely dictated by the physicochemical properties of the drug and the desired drug release profile. Lipophilic drugs are usually incorporated into water-soluble bases while hydrophilic drugs are formulated into the fatty base suppositories. Theoretically, this method maximizes removal of the drug from the suppository base to the immediate environment of the rectal cavity or lower colon. For suppositories made from fatty bases, melting should occur rapidly near body temperature (37°C). Ideally the resultant melt would readily flow to provide thin, broad coverage of the rectal tissue, thereby minimizing lag time effects due to slow release of the drug from the suppository base. Water-soluble suppositories should likewise readily dissolve at 37°C to facilitate drug release and subsequent absorption. With both fatty-base and water-soluble suppositories, the potential effects of incorporated drug on melting or dissolution properties need to be evaluated. Examples of fatty bases and water-soluble bases suitable for suppository formulations are shown in Table 1. Although not comprehensive, the list shows the variety of bases from which a formulator may choose.

Solutions

Solutions, suspensions, or retention enemas represent rectal dosage forms with very limited application, largely due to inconvenience of use and poor patient compliance. In many cases, these formulations are utilized to administer contrast media and imaging agents for lower GI roentgenography. Although drug absorption from solutions has been shown to exceed that from solid suppositories in some cases (5), this particular administration route is only

Table 1 Suppository bases

Vehicle	Melting range (°C)	Solidification point (°C)
Fatty bases		
Witepsol	32–44	27–38
Cocoa butter	30–35	24
Hard butters	36–45	32–40
Estarinum	29–50	26–40
Suppocire	35–45	30–37
Agrasup A;H	35–40	—
Water soluble		
Myrj 51	39–42	39
PEG ^a	38–49	38–42
Tween 61	35–49	—

^aPolyethylene glycol.

(From *Modern Pharmaceutics*; Banker, G. S., Rhodes, C.T., Eds.; Marcel Dekker, Inc.: New York, Basel 1979.)

infrequently employed and will not be discussed here in detail. Recent studies utilizing liquid formulations that gel at body temperature are discussed in the Gels/foams/ointments section.

Gels/foams/ointments

The use of gels, foams or ointments for rectal administration can afford advantages over liquid formulations because retention of the dosage form in the rectal cavity reduces patient compliance problems. Drug release with semisolid dosage forms is usually limited to local indications such as hemorrhoids and lower bowel inflammation (proctitis). Drug release and subsequent pharmacologic action is usually faster with semisolid formulations than with solid suppositories since a lag time is not required for melting or dissolution.

Miyazaki et al. (1) investigated thermoreversible gels formed by a xyloglucan polysaccharide derived from tamarind seed. Liquids containing 1–2% xyloglucan formed gels over a temperature range of 27–32°C. The gelling temperature decreased with increasing xyloglucan concentration. In vitro release profiles for indomethacin and diltiazem were characterized as a square-root of time function with diffusion coefficients increasing with temperature increases from 10 to 37°C. The slower in vitro indomethacin release from gels was confirmed in vivo where broader absorption peaks and longer residence times were noted. There were, however, no significant differences in bioavailability between the thermal gelling formulations and conventional indomethacin suppositories.

Ryu et al. (2) examined mucoadhesive liquid suppositories that combined bioadhesive properties with

thermal gelling polymers. Hydroxypropylcellulose, polyvinylpyrrolidone, carbopol, polycarbophil and sodium alginate were used as bioadhesive polymers in thermal gelling polymers comprised of poloxamer 407 and poloxamer 188. Gellation temperatures between 30 and 36°C were obtained with mucoadhesive forces ranging from 430 to 5800 dyne/cm². With propranolol as a model compound, bioavailability increased as the mucoadhesive force increased and dosage form migration distance decreased. There was, however, no direct relationship between bioavailability and gellation temperature. Sodium alginate and polycarbophil afforded the greatest mucoadhesive forces and most significant improvements in propranolol bioavailability. Sodium alginate was also free of any quantifiable adverse effects on rectal tissue. The studies suggest that reduced formulation migration as a function of mucoadhesive potential was the primary causative factor for the improved bioavailability of propranolol.

Watanabe et al. (4) have reported improved rectal absorption with reduced mucosal irritation utilizing rectal gels comprising water-soluble dietary fibers, xanthan gum and locust bean gum. Absorption of buprenorphine was more rapid than that observed from polyethylene gels with comparable extents of bioavailability. Mean residence times increased with increasing gum concentrations. Most importantly, the degree of mucosal irritation from these dietary fiber gels was significantly less than that observed with polyethylene based suppositories.

Controlled-release formulations

Controlled-release formulations are designed to release the active agent in a sustained and controlled fashion. They have been the subject of considerable research but have yet to make a significant impact. Hydrogels have been shown in human clinical studies to provide an acceptable polymeric system for rate-controlled delivery of antipyrine and theophylline (6). Rate-controlled osmotic delivery systems have also proven useful in clinical studies in effecting systemic drug delivery comparable to that of intravenous administration for well absorbed drugs (7). Since the total acceptable size of a rectal formulation significantly exceeds the size possible for oral formulations, rectal administration for the purposes of controlled-release offers a significant advantage. A major limiting factor is, however, the need to incorporate controlling agents designed to regulate drug release which would significantly increase the total size of the dosage form. Since adult rectal dosage forms are acceptable up to 2.5 g, the total drug load which can be formulated in a rectal controlled-release formulation can be 2–3 times that possible in an oral formulation. For some therapeutic agents, this higher drug load can offer an advantage which

is not achievable via the oral route. Development and marketing of rectal controlled-release formulations will, however, still be disadvantaged because of the perceived reluctance of patients to employ this route and problems of poor patient compliance.

Marketed drugs and therapeutic classes

Only a limited number of therapeutic agents are currently marketed in the United States as rectal dosage forms. A survey of the 1994 American Hospital Formulary Service shows that 19 agents are sold for systemic therapeutic indications (Table 2) and 15 agents for local GI applications (Table 3), including imaging agents for diagnostic purposes. Many products for local GI application are over-the-counter (OTC) products for treatment of local inflammatory reactions and hemorrhoids. As is readily evident from the limited number of therapeutic agents available for rectal administration, this dosing route does not represent a major share of the physician's choices for drug therapy.

Worldwide market

In certain areas of the world, particularly some European countries and Japan, rectal dosage forms are somewhat more accepted by the patient population and, hence, development of rectal dosage forms has surpassed that in the United States. According to a survey in 1970, approximately 7.5% of all prescriptions in France were formulations intended for rectal administration (8). Even though a few countries may find rectal dosage forms more acceptable, these still represent a small area of the world-wide market share which can be assigned to rectal drug therapy.

PHYSIOLOGIC AND PHARMACEUTICAL ISSUES RELEVANT TO RECTAL DRUG DELIVERY

Physiology and Biochemistry of Rectal Tissues

Anatomical considerations

Unlike the small intestine and upper colon, the vasculature draining the rectal cavity does not totally direct the blood supply to the liver (9). The lower and middle hemorrhoidal veins of the rectum bypass, at least partially, the portal circulation during their first pass into the general circulation, allowing absorbed drug to exert systemic effects prior to possible metabolism and excretion via hepatic mechanisms. Lignocaine, propranolol, and salicylamide have been shown in clinical studies to attain greater systemic bioavailability when administered rectally than

Table 2 Rectal dosage forms marketed in the United States for systemic indications

Therapeutic category and drug	Drug load, solid ^a (mg)
Antihistamine	
Promethazine	12.5–50
Antimigraine	
Ergotamine	2
NSAID	
Aspirin	60–1200
Indomethacin	50
Analgesic	
Hydromorphone	3
Morphine	5–30
Opium	30–60
Oxymorphone	5
Acetaminophen	120–650
Insomnia	
Pentobarbital	30–200
Chloral hydrate	325–650
Promethazine	12.5–50
Tranquilizer	
Chlorpromazine	25–100
Prochlorperazine	2.5–25
Bronchodilator	
Aminophylline	105
Antiemetic	
Thiethylperazine	10
Trimethobenzamide	100–200
Hyperkalemia	
Polystyrene sulfonate	1250 ^b
Portal-systemic encephalopathy	
Lactulose	Variable ^c

^aUnless otherwise indicated.

^bSuspension.

^cSolution.

(From McEvoy, GK., Ed. *American Hospital Formulary Service*; American Society of Hospital Pharmacists, Inc.: Bethesda, MD, 1994.)

when given orally (10). The rectal cavity is also drained by extensive lymphatic circulation which facilitates absorption and systemic exposure of absorbed drugs (11).

Although extensive villi and microvilli are not present in the rectum and colon tissue, sufficient surface area is present to allow absorption of readily permeable drugs. The lack of motility in the rectum and colon, as opposed to extensive motility in the small intestine, provides an additional advantage in terms of maintaining maximum concentration gradients at the absorptive surface. Together with a limited fluid volume in the lower colon, typically 2–3 ml of inert mucous fluid in the absence of fecal material, the static environment of the rectum and lower colon provides an area for maintaining significantly higher

Table 3 Rectal dosage forms marketed in the United States for gastrointestinal indications

Therapeutic category and drug	Drug load	Physical form
Laxatives		
Senna	30 mg	Solid
Glycerin	4 ml	Solid or liquid
Mineral oil	—	Liquid
Potassium bitartrate		Solid
Dibasic sodium phosphate	—	Solution
Docusate sodium	—	Suspension
Hemorrhoids		
Hydrocortisone	100 mg	Suspension or cream
Hydrocortisone acetate or butyrate	—	Aerosol, foam, and suspension
Dibucaine	—	Ointment
Proxamine	1%	Ointment
Proctitis, colitis		
Belladonna	16.2 mg	Solid
Mesalamine	500 mg	Solid
Enteropathogenic diarrhea		
Neomycin	—	Retention enema
Contrast imaging agents		
Barium sulfate	5–70% (w/w)	Suspension
Diatrizoate	60–66%	Solution

(From McEvoy, G.K. Ed., *American Hospital Formulary Service*; American Society of Hospital Pharmacists, Inc.: Bethesda, MD, 1994.)

drug concentrations than is readily achievable in the small intestine. On the negative side regarding potential for drug absorption, the intercellular junctional complexes are tighter in colon and rectum than in small intestine (12) which may reduce opportunities for small, water-soluble drugs to permeate intercellular spaces and gain access to the systemic circulation without passing through cellular membranes.

Biochemistry

Cellular metabolism of drugs, as they pass through the mucosal barrier of the rectum or colon, can be expected to be similar to that seen in the small intestine since the basic intracellular metabolic machinery is common to epithelial cells. The major difference relative to drug metabolism occurs in the enzymes to which drugs are exposed in the intestinal lumen and on the apical membranes of the epithelial cell layer. Since the colon and rectum do not serve digestive functions, the luminal enzymes, which are actively secreted in the upper small intestine, are not present to any significant extent. As such, proteolytically labile drugs such as peptides and proteins should exhibit greater stability if released in the rectum or lower colon. Saffran et al. have shown that vasopressin, a peptide subject to proteolytic hydrolysis, is more active when administered by the rectal route than by oral administration

(13). Significant rectal absorption of growth hormone (14) and insulin (15) have also been demonstrated with the help of absorption enhancing agents. The apical membranes of the small intestine epithelial cell layer express high levels of membrane-associated or membrane-bound enzymes, such as peptidases and saccharidases, which are not present in high amounts on the apical surfaces of epithelial cells in the rectal cavity. This absence of membrane surface metabolic potential affords advantages when delivering drugs susceptible to enzymatic degradation. These membrane-associated enzymes are, however, linked to direct transport carriers for specific nutrients, notably amino acids and sugars. Carriers for many of the vitamins are also present in small intestinal tissue. The relative absence of these transport mechanisms in colon and rectal tissue (16) eliminates this mechanism of absorption as a viable route for rectal drug delivery. Since most of these amino acid, saccharide, and vitamin carriers have relatively specific structure and transport requirements, the absence of these systems in colon or rectal tissue affects only drug candidates with significant structural similarity to natural substrates.

The pH of the rectal compartment is essentially neutral, ranging from 7 to 8, with minimal buffering capacity as compared to the small intestinal milieu. Suppositories or solutions formulated to maintain a specific pH in order to

optimize drug absorption typically will function at that pH following administration. This can be a significant advantage with drugs whose permeation properties are optimized near neutrality. It should be kept in mind, however, that the controlling pH at the epithelial membrane surface, which is under the overlying mucous layer, is still in the range of pH 6.0–6.5 as it is throughout the small and large intestine. Therefore, even though luminal pH may favor drug absorption and diffusion from the lumen to the cell barrier, the effective pH at the cellular barrier is not drastically different from other regions of the intestinal tract. In this sense, perceived pH advantages of rectal drug administration may be obviated by the basic surface pH which is characteristic of mucosal tissue.

Model Systems and Techniques

Whole animal models

Preliminary *in vivo* evaluation of the rectal absorption potential of drug candidates is most easily achieved in a rat model (17, 18). Animals can be maintained at surgical plane anesthesia for 120–180 min, and drug solutions or suspensions can be directly instilled into the rectal cavity. Utilizing a slight elevation above horizontal, instillation of approximately 0.25 ml is readily achieved in rats without loss of fluid from the rectal compartment. Purse-string sutures or moderate pressure clamps can also be used to prevent leakage of administered solutions from the rectal cavity. In most cases, it is advisable to fast animals for 12–16 h prior to study to void the colon and rectal cavity of fecal material which may interfere with administration of the solution or absorption of the drug itself. Microsuppositories can also be evaluated in the rat model by utilizing specially designed molds to prepare small dosage forms, but care must be taken to guarantee drug content uniformity in these small devices. Additionally, since drug absorption can be influenced by the choice of suppository base, this must be taken into consideration in any analysis of the rectal absorption potential of a drug candidate.

Rabbits have been another common animal model utilized for examining rectal drug absorption. Constraints with using this animal model are similar to those encountered with the rat model, although there is more latitude concerning the volume (solid or liquid) of the dosage form which can be tested. A possible drawback to the rabbit model is evidence suggesting that the rabbit colon is somewhat more “leaky” than the colon in other species. In this regard, the rabbit model may overestimate the extent of drug absorption achievable for small, water-soluble compounds which utilize intercellular transport pathways via water flux.

For evaluation of dosage forms designed and formulated for human use, the Beagle dog presents an ideal animal model (18, 19). Solid suppositories (up to 2.5 g total weight) or microenemas (up to 5 ml volume) can be readily tested in Beagles without concern for loss of dosing vehicle from the rectal compartment. Since the physical dimensions of the dog rectum approximates those of the human rectum, variables such as spreading of suppository bases can be reasonably well evaluated. The dog is obviously a convenient model for blood sampling and pharmacokinetics and is, therefore, well suited for the evaluation of rectal dosage forms. Studies from the author’s laboratories and others have shown that dosage-form performance criteria generated in Beagle dog models are accurate predictors of performance in human clinical trials (20).

In vitro models

For assessing the absorption potential of specific drug candidates or conducting studies evaluating correlations between drug structure and transport, *in vitro* models may provide the best approach. Both the tissue-diffusion cell systems and cultured colon cell lines have proved to be particularly useful.

Numerous investigators have employed diffusion chambers of various designs (e.g., Ussing chambers, Sweetana–Grass diffusion cells) to evaluate the permeation properties of drug candidates (21, 22). These systems offer a distinct advantage that drug solutions may be added to the donor compartment under a variety of conditions (e.g., varying concentrations, pH, excipients). Samples are withdrawn periodically from a receiver compartment and analyzed for drug and metabolite content to obtain a true measure of permeation across the cellular barrier. In establishing procedures for these diffusion cell studies, the underlying muscle layer can be teased away and only the intact mucosal cell layer mounted in the diffusion chamber. Since the underlying muscle layer is not a barrier encountered under normal physiologic conditions, this method more accurately reflects the true cellular barrier to drug transport. It should be noted that, in many cases, any mucus layer normally overlying the mucosal cell layer may be disrupted and removed by the physical manipulations and tissue washing procedures normally employed in the experimental design. Therefore, for drugs where diffusion through the mucus layer may represent the rate-limiting step, this model may overestimate drug permeation unless great care is taken to maintain the mucus layer. Conversely, the use of mucolytic agents can be employed to ensure a mucus-free cell layer if true cellular drug permeability is to be determined. Work in the author’s laboratories has demonstrated that the Sweetana–Grass diffusion cell model with muscle-free mucosal layers is

reasonably resistant to experimental conditions. The pH can be varied between 5 and 8 without significant damage to the mucosal layer. Additionally, excipients such as DMSO or ethanol, can be utilized up to 10% (v/v) in order to improve solubility of drugs whose absorption may be limited by poor aqueous solubility. Finally, our experiments have shown that muscle-free mucosal strips can be maintained with good viability for up to 120 min at 37°C, although this should be experimentally determined in each laboratory using this model to ensure that tissue damage is not affecting the measured drug transport values. Typically, standard metabolic viability assays and microscopy are suitable for confirming tissue viability.

Another *in vitro* model that has experienced significant growth in utilization in drug transport studies and which may provide useful information relative to rectal or colonic drug absorption is cultured cell monolayers. HT-29 and Caco-2 cells are two cell lines commonly employed for such studies (20, 25). Both are human colon-cancer cell lines which can be grown on membrane filters and readily form confluent monolayers with intact tight junctional complexes (typically $>250 \Omega \text{ cm}^2$). They can be used in experimental designs similar to the diffusion cell models. These cells are non-mucus secreting cell populations which resemble fetal small intestine in many of their metabolic properties. Although these models are useful for quantifying drug transport across colon cell lines, certain limitations must be kept in mind. First, aqueous pathways in these confluent monolayers are significantly reduced compared to the *in vivo* situation. In other words, the intercellular pathway, as monitored by electrical resistance, is less “leaky” than that encountered in normal colon or rectal tissue and drug transport studies may underestimate the absorption of drugs which utilize the paracellular pathway. Second, there is evidence to suggest that some of the carrier systems present in small intestinal epithelial cells (e.g., vitamins, dipeptides) are also present to some extent in these cell lines (26, 27) which may provide misleading information on the rectal absorption potential for compounds which utilize these carriers. Finally, these are transformed cells which may or may not present metabolic barriers comparable to that seen in normal colon or rectal tissue. Care must be taken in interpreting data from transport studies which show parent drug metabolism during the absorptive phase since this may or may not also occur *in vivo*.

Advantages over Oral Systems

Improved enzymatic drug stability

It is well known that the oral delivery of many drugs, particularly peptides and proteins, is limited due to poor

absorption and/or stability in the stomach and upper intestinal tract. Many proteolytic and other enzymes in the stomach and small intestine result in drug degradation which prevents effective absorption following oral administration. As discussed above, degradative enzymes are present to a much lesser degree in the rectum and therefore many drugs that cannot be administered effectively orally can be administered rectally without as much enzymatically catalyzed degradation. Examples include vasopressin (13), growth hormone (14), and insulin (15). It has been found necessary to include absorption enhancing agents for growth hormone and insulin.

Partial avoidance of hepatic first pass

The rectum is extensively supplied with blood from the various rectal arteries. It is drained by at least three veins (28) and drug absorption primarily occurs through this venous network. Although there is substantial intermingling of these veins due to several anastomoses, it is usually reported that the inferior and middle rectal veins drain into the inferior vena cava. This allows drugs absorbed by this route to partially bypass the portal system and the associated first-pass metabolism in the liver. As an example of this effect, de Boer and Breimer (10) have shown that the systemic bioavailability in humans for lignocaine is about twice that observed following oral administration. It has been suggested that about half of rectally administered drug avoids hepatic presystemic metabolism. Clinical studies have also shown higher systemic bioavailabilities for propranolol and salicylamide when administered rectally compared with oral administration (10). Studies in rats suggest that first pass metabolism is almost totally avoided by rectal administration. For example, the bioavailability of nitroglycerin in rats following unrestricted rectal instillation was about 27% compared to 2% from oral dosing. When the length of rectal tissue exposed to the drug was restricted to 2 cm from the anus, the bioavailability increased to about 90% (29).

The fact that rectal administration in humans only partially avoids delivery of the drug to the portal system has been suggested as an advantage of rectal administration of insulin over parenteral administration (30). Delivery of insulin to the portal blood system is suggested to be more physiological than delivery to the peripheral cells from subcutaneous injections.

Higher drug load

In general, oral administration in a single tablet is limited to about 1 g. Typically, suppositories intended for adults can be as large as 2.5 g. This should allow for two to three

times higher drug loads to be administered, depending on the amounts of other excipients necessary in the formulation of the suppository.

Lymphatic delivery

Some studies have indicated that significant amounts of some drugs are transported through the lymphatic systems following rectal administration. For example, Caldwell et al. (11) have examined the importance of lymphatic transport of water-soluble drugs following rectal administration in the presence of salicylate-type adsorption enhancers. They observed that following i.v. injection to the rat, the concentration of phenol red in the plasma and the lymph (collected from the thoracic duct) were similar. They further found that concentration of phenol red in the lymph after rectal administration to rats via cannulated thoracic ducts was more than 100 times greater than corresponding plasma concentrations when 5-methoxysalicylate was used as an adjuvant. Rectal administration of phenol red in the presence of 5-methoxysalicylate to rats with intact thoracic ducts produced plasma levels of phenol red which were about fivefold greater than those found in the rats in which lymphatic drainage had been diverted. As shown in Fig. 1, insulin was also transported primarily through the lymphatic system to the general circulation. However, theophylline exhibited much lower lymphatic transport.

Constant and static environment

Compared to the oral route of administration, the rectal route provides a much more constant environment for the drug as it is absorbed. An orally administered drug must usually pass through a series of diverse environments prior to absorption. This includes transit through the stomach and the small intestine and perhaps to the large intestine and colon, depending on the drug and the absorption site. The pH and absorptive sites in the GI tract change significantly as the drug traverses it. This can lead to a more complicated set of conditions which influence absorption compared to the rectum which is a short part of the GI tract with relatively constant absorption characteristics.

Patients with swallowing difficulty

For a number of patients, especially children and elderly, problems associated with swallowing and nausea when taking oral medication can be largely obviated by rectal administration. Patients with gastrointestinal disorders or after surgery often have difficulties taking drugs by mouth. Rectal administration may be a good alternative to oral administration in these types of situations.

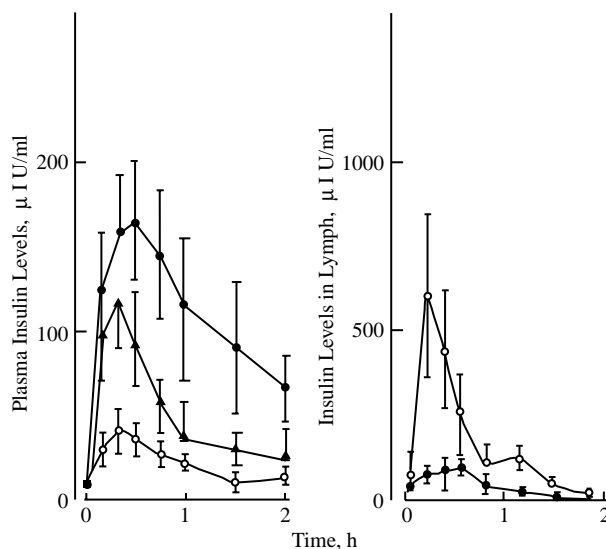


Fig. 1 Concentration of insulin in the plasma and lymph of rats following intramuscular administration of 0.8 IU/body insulin (●), rectal administration of 2.0 IU/body insulin in the presence of 10 mg/body 5-methoxysalicylate and intact thoracic duct (▲), and rectal administration of 2.0 IU/body insulin in the presence of 10 mg/body 5-methoxysalicylate with the thoracic duct cannulated for collection of lymph (○). The error bars represent standard deviations with $n = 6$. (From Ref. 11.)

Avoidance of overdosing

For some patients and with certain drugs, such as certain sedatives, oral administration may raise a concern with respect to the possibility of severe accidental or intentional overdosing. This danger is practically eliminated by rectal administration. Whereas it is relatively easy for a patient to swallow a number of tablets, it is much more difficult to administer numerous suppositories rectally at the same time. Thus rectal administration may be indicated for patients for whom overdosing is a significant concern.

Disadvantages Compared to Oral Systems

Patient acceptance and compliance

In some cultures, such as in the United States, there is a reluctance by many to consider rectal administration. This has resulted in a tendency by pharmaceutical marketing groups to avoid rectal dosage forms, except for the most obvious indications and situations where other dosage forms have substantial disadvantages. Frequently, it is inconvenient to receive or administer a suppository or other rectal dosage form, thereby reducing patient compliance. Thus, the need must be great in most cases for rectal administration to be seriously considered.

Potential for nonspecific drug loss

There are at least two common problems that can lead to drug loss following rectal administration. First, for effective absorption, the dosage form must be retained in the rectum. Thus if the dosage form or parts thereof are lost prematurely from the rectum, drug absorption will be substantially reduced. A study with children showed that when thiopentone suppositories were voided within 40 min, an effective plasma level was maintained for less than 2 h, whereas when the suppositories were retained, an effective level was maintained for about 24 h.

Second, there is the possibility that the drug or some important excipients may interact with constituents of fecal matter or material fluid present in the rectum. This may reduce the drug absorption and diminish effectiveness.

Limited fluid in the rectum

The amount of liquid in the rectum has been reported to be about 3 ml. This is small compared to the volume of fluid available throughout the GI tract when a drug is administered orally. Such a small volume of fluid can limit dissolution of drugs, particularly those with low aqueous solubility. It also may be a constraint to the rapid dissolution and release of compounds from water-soluble vehicles where dissolution of the vehicle is considered to be the rate-determining step in drug release from the vehicle.

Formulation

There are a number of formulation variables and considerations that can lead to difficulties in rectal drug absorption, including the melting and liquefaction characteristics of the vehicle. The solubility of the drug in the vehicle, the particle size of the drug, the vehicle spreading capacity, the viscosity of vehicle and excipients at rectal temperature, and possible retention of the drug by excipients, all can affect the rate of release and consequent drug absorption. Furthermore, the pK_a of the drug, the pH of the rectal fluids, the presence of buffers, and the buffer capacity of the rectal fluid as well as the partition coefficient of the drug influence drug absorption and must be considered in the formulation of a suppository or other rectal dosage form. Storage temperature, time, and conditions can have a profound effect on both the stability and release characteristics of a drug from a rectal dosage form. Each of these considerations lead to potential difficulties in the formulation, manufacture and distribution of rectal dosage forms.

Expense

Suppositories and other rectal dosage forms are more expensive to prepare and dispense than simple tablets. Therefore, unless there is a significant need and advantage by utilizing a rectal dosage form, suppositories are not likely to be used. Recently, techniques have been developed to prepare suppositories more efficiently which could lead to lower manufacturing costs.

Drug Classes Useful for Rectal Administration

Drugs that are currently marketed in the United States for rectal delivery are shown in Tables 2 and 3. In addition to drugs intended for local effects, many other types can be administered rectally for systemic activity, such as drugs with high hepatic first-pass extraction, drugs for which lymphatic absorption is important, and compounds requiring a relatively high therapeutic dose. Generally, drugs requiring significantly more than 500 mg per tablet are not easily taken orally and in some cases may be candidates for rectal delivery.

Although products are not currently marketed containing proteins and peptides, the literature contains several examples of rectal absorption of peptides and proteins. Typically proteins and peptides require some type of penetration-enhancing adjuvant in the formulation to facilitate absorption. Examples of peptides and proteins that have shown significant rectal absorption include: insulin, lysozyme, calcitonin analogs, phenylalanine and di-, tri-, and tetraphenylalanine, as well as gastrin, pentagastrin and tetragastrin (31).

ABSORPTION CHARACTERISTICS AND REGULATION OF DRUG ABSORPTION

Control or Modification of Rectal Drug Absorption

pH partition

The mechanism for absorption from the rectum appears to be similar to that observed for the rest of the GI tract, that is, passive diffusion. Drugs are best absorbed through the rectal mucosa in their un-ionized or neutral form. Drugs with high partition coefficients (more lipophilic) tend to be better absorbed. There are, however, conflicting reports in the literature with some suggesting that simultaneous absorption of ionic species is possible. That is particularly true for relatively small molecules. If the drug can exist in the unionized state at physiological pH, other factors being equal, absorption is improved.

Solubility

As was pointed out earlier, the volume of fluid in the rectum is very small. In most cases, it is believed that the drug should be dissolved in the rectal fluid prior to absorption. This requires that drugs have a reasonably high solubility to be efficiently absorbed from the rectum. Voigt and Falk (32) reported a direct relationship between water solubility and release rate for 35 different compounds. Generally, when a compound can be presented in relatively low solubility form (e.g., neutral acid) or as a more water-soluble form (e.g., the sodium salt), the higher the solubility and consequently the higher the dissolution rate in the rectal fluid, the better is absorption. This factor has to be balanced with the fact that an unionized species tends to pass through the rectal mucosa more readily.

Drug solubility also affects the choice of suppository base or other vehicle. Generally, the drug should have little tendency to remain in the vehicle upon melting or dissolution. Therefore, it is usually suggested that water soluble drugs are best delivered from fatty vehicles and that more lipophilic compounds from water-soluble vehicles.

Molecular size

The smaller the drug molecule, the more readily it can be absorbed. For larger molecules, some type of facilitated transport or the use of penetration enhancers have been found to increase drug absorption from the rectum as well as from other delivery routes.

Charge

Charged molecules have been found to pass through the rectal mucosa less effectively than neutral molecules in most cases. This can sometimes be overcome by modifying the pH or allowing the charged species to interact with another molecule or ion that helps neutralize the charge. Ion pairs, for example, hold some promise for overcoming the drawbacks of charged species. From a solubility point of view, charged species are generally preferred. Often a balance between high solubility and penetration through the rectal mucosa must be obtained, requiring some compromise in properties of choice. It has often been observed that *in vivo* differences are often less pronounced than *in vitro* differences. Furthermore, penetration enhancing agents can often improve the delivery of relatively large water-soluble drugs.

Nonspecific adsorption

The surface properties of a solid may significantly affect the drug when it reaches the interface between the vehicle and the rectal fluid. The amount of wetting, as

demonstrated by contact angle, that occurs or changes in the interfacial tension as caused by interactions with surface-active agents in the vehicle or rectal fluid may have a profound effect on dissolution and consequent absorption of the drug. This can often increase the availability of the drug. On the other hand, adsorption or complex formation with surface-active agents may reduce the availability of the drug and its absorption.

Spreading of the administered formulation

For optimal drug absorption, it is important that the suppository or vehicle melts or dissolves rapidly and spreads over the rectum walls. Thus the rheological behavior of the vehicle can have a significant effect on the release of the drug and the ability of the drug to come into contact with the rectal mucosa. Several studies have suggested that the viscosity of the vehicle is very important for the release of the drug from the vehicle. Studies that compare the spreading behavior of suppository bases and their hydrophilic-lipophilic balance (HLB) values have been inconclusive. Although spreading directly determines the area from which release from the vehicle can occur and thus absorption, there is also the potential difficulty that concentration and thermodynamic activity of the drug may be reduced if it is allowed to spread too much and particularly too high up the rectum. This may result in absorption by the upper hemorrhoidal vein and into the portal blood supply with increased first-pass metabolism.

Yahagi et al. (3) have reported improvements in lidocaine bioavailability utilizing a unique double phase suppository that minimizes spreading of an administered suppository. The front or anchoring phase of the suppository contains Witepsol-H15, Carbopol, 934P, and wax. The terminal layer or drug releasing layer contains Witepsol-H15, Carbopol 934P, and lidocaine. Carbopol provides the bioadhesive properties and the wax confers physical strength to the suppository formulation. This formulation provided greatly prolonged plasma lidocaine levels with improved bioavailability. The authors suggest that a combination of improved rectal retention (decreased formulation migration due to bioadhesive nature) combined with a somewhat slower drug release rate maximizes avoidance of absorption into vascular pathways subject to first-pass hepatic metabolism. This approach may prove useful for drug candidates subject to high first pass effects.

Optimizing Drug Absorption

Enhancing agents

Over the past 20 years, a variety of agents have been identified which significantly increase the permeability of

the GI tract to drug absorption. The enhancing action of a variety of absorption-promoting adjuvants on rectal absorption has been extensively discussed (33). Included are acylcarnitines (19), acylcholines (19), salicylates (34), bile salts (35), phenothiazine derivatives (36), enamines (37), and fatty acids (38). Strong chelating agents and phenothiazines appear to enhance the rectal absorption of both low and high molecular weight compounds with a constant ratio of absorption via a paracellular route. Diethylmaleate enhances rectal absorption of low molecular weight compounds via a transcellular route. Various salicylates, diethyl ethylene malonate, and various fatty acids have been found to enhance both the paracellular and transcellular routes of absorption (33). In some respects, rectal drug administration is optimally suited for coadministration of drug entities with absorption-enhancing agents.

Work in the authors' laboratories has clearly shown that the presentation profile of drug and enhancing agent is critical to optimal activity (19). With agents like the acylcarnitines and acylcholines, the effective increase in permeability due to absorption enhancement is transient with nearly complete loss of activity approximately 30–60 min after dosing. In order to take advantage of the narrow time window of increased permeability, it is essential that the drug and enhancing agent be present at the mucosal barrier at the same time and in sufficient concentration to effect the permeability change. The low motility and limited fluid content of the rectal compartment is ideal for optimizing these requirements. When administered orally, enhancing agents are routinely less effective than when given rectally, presumably due to motility and dilution of the drug and enhancer. The necessary temporal and spatial dosing of a drug and enhancer can be achieved more readily via the rectal than the oral route. Design and performance determination of solid rectal suppositories must, however, address the release profiles of both agents which adds to the complexity of the formulation process. This can be achieved, as demonstrated in a previous study, with sodium cefoxitin and salicylate-type enhancing agents (20). In this study, a rectal suppository of the antibiotic with enhancing agent was shown to be essentially bioequivalent to an intramuscular injection of the antibiotic alone. The rectal absorption of proteins can also be enhanced by salicylates, as shown for insulin in Fig. 2. As an alternative to injectable administration, a rectal suppository offers definite advantages in terms of patient compliance and out-of-clinic administration.

pH control

As discussed earlier, the pH of the rectal fluid can have a marked effect on the absorption of drugs from the rectum.

Since the rectal fluid has a relatively low buffering capacity and the volume of the rectal fluid is small, it might be expected that the contents of the rectal dosage form largely control the pH of the rectum during administration. On this basis, one may be able to utilize the pH characteristics of the drug and incorporate suitable buffers and other excipients in the dosage form to control the pH. It has been reported (40) that a solution having a buffer capacity of 0.1 was sufficient to maintain a pH at about 5.9 during perfusion in humans. In this study, the pH was restored to normal at a rate of about one half a pH unit per minute, following removal of the perfusate. The body appears to try to maintain the pH at the absorbing surface relatively constant by secretion but often requires some time to be able to return to the normal pH after administration of a rectal dosage form having significant buffer capacity and differing pH.

Solubilizing agents

It would be expected that solubilizing agents may increase the rate of drug release from a suppository base by increasing the dissolution rate and perhaps by modifying the viscosity and interfacial tension of the vehicle with the rectal fluid. In addition to the effect that a solubilizing agent or a surfactant can have on the drug and the vehicle, the surfactant may also have an effect on the mucous coating of the rectal membrane. This may increase absorption by reducing the thickness of the layer through which the drug must traverse or it may act as a penetration enhancer by increasing the permeability of the membrane through damaging the rectal mucosa. Nishihata et al. (39) reported that sodium lauryl sulfate appears to interact with the lipoidal fraction of the rectal membrane with irreversible effects in the short term. It was further reported (41) that various metabolic inhibitors had little, if any, effect on the enhancement observed with the surfactant polyoxyethylene 23-lauryl ether. However, definitive studies have not been reported which clearly delineate the complexities that can occur from the simultaneous involvement of several factors that influence drug absorption from the rectum. It has been suggested that for an oleaginous base, solubilization, particularly with a decrease in vehicle viscosity, should improve rectal drug delivery.

Viscosity modifiers

The luminal pressure of the rectal mucosa can act as a shearing stress and influence the rheological behavior of substances showing either plastic or pseudoplastic behavior. It is possible that the viscosity at the shearing stress supplied by the rectum is more important than the yield value of the suppository. It appears that viscosity is very important for drug release from suppositories where

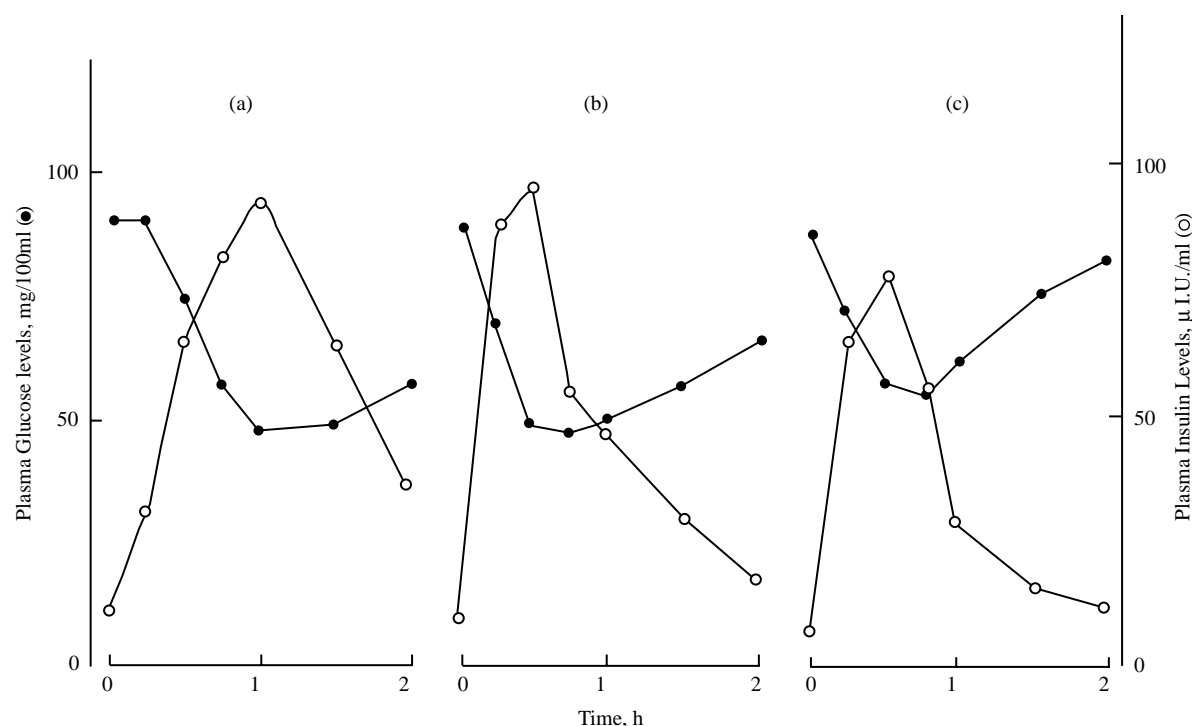


Fig. 2 (a) Concentrations of glucose (mg/100 ml) and insulin (μ IU/ml) in the plasma of dogs following an intramuscular injection of 10 IU of insulin. (b) Concentrations of glucose and insulin in the plasma of dogs after the administration of a 0.5 ml microenema containing 20 IU of insulin with 150 mg of sodium 5-methoxysalicylate in a 0.9% NaCl solution containing 4% gelatin. (c) Concentrations of glucose and insulin in the plasma of dogs following a 0.5 ml microenema containing 20 IU of insulin with 300 mg of sodium salicylate in a 0.9% NaCl solution containing 4% gelatin. (From Ref. 39.)

the melted material behaves like a Newtonian fluid. Generally, the lower the viscosity, the quicker and more complete the release of the drug from the vehicle and the higher the absorption of the drug.

FUTURE OF RECTAL DRUG DELIVERY

Market Potential

Even under the best of conditions and therapeutic needs, rectal dosage forms will remain only an alternative to oral administration. Due to problems associated with patient acceptance and compliance, the need to refrigerate many suppository bases, and the inconvenience of dosing, it is extremely unlikely that rectal drug administration will ever play a significant role in the pharmaceutical market. However, in specific fields of the therapeutic regimen, rectal dosage forms can have a significant impact and can address the needs of several patient populations that are poorly fulfilled by conventional oral formulations.

Examples Where Oral Administration Does Not Satisfy Therapeutic Needs

For certain groups of the patient population, oral dosing is either not desirable or impossible. Several antibiotics which are administered postoperatively via parenteral routes are not available as oral formulations because of the nature of the drug or its absorption limitations, or they are unacceptable due to swallowing difficulties in this patient subset. In these cases, rectal formulations that can contain a significantly higher drug content than oral formulations and which can be administered without difficulty to hospitalized patients, offer an attractive alternative and can be utilized to wean patients from parenteral to non-parenteral drug delivery. Rectal formulations of medications for pain control and sedation fit within this postoperative category.

Both children and the elderly experience swallowing problems with oral formulations. A limited market already exists in pediatric therapeutics which addresses this problem. Although development of rectal formulations for the elderly has not been addressed extensively, with the expanding number of patients in this age group, the need

for acceptable alternatives to oral dosing may increase efforts to develop rectal formulations.

As indicated previously, two general classes of drugs may be most amenable to formulation in rectal dosage forms. Drugs which require relatively high dosing, such as many of the antibiotics, are good candidates for rectal formulations which minimize the need for multiple oral dosings in order to reach the desired drug levels. Drugs that are substrates for proteolytic activity in the upper GI tract, particularly peptides and proteins, may find a useful application in rectal dosage forms if their absorption profile can be improved. Finally, in some very specific cases where extensive first-pass metabolism limits a drug's usefulness, rectal formulations may provide an attractive alternative.

SUMMARY

Rectal administration of therapeutic agents represents a narrow part of the total pharmaceutical approach to disease management but it holds promise for increasing applications. Although rectal administration will not become the first line of drug delivery, the increasing pressure to find exploitable delivery routes for peptides and proteins may well find some answers in rectal administration. From a drug stability perspective, the low level of proteolytic activity in the colon and rectum offers hope that active agents can be delivered to the cellular membrane in effective concentrations. Assuming that techniques can be developed to increase the permeability of the GI tissue to these agents, rectal administration may become a more widely employed route of delivery.

REFERENCES

- Miyazaki, S.; Suisha, F.; Kawasaki, N.; Shirakawa, M.; Yamatoya, K.; Attwood, D. Thermally Reversible Xyloglucan Gels as Vehicles for Rectal Drug Delivery. *J. Controlled Release* **1998**, *56*, 75–83.
- Ryu, J.M.; Chung, S.J.; Lee, M.H.; Kim, C.K.; Shim, C.K. Increased Bioavailability of Propranolol in Rats by Retaining Thermally Gelling Liquid Suppositories in the Rectum. *J. Controlled Release* **1999**, *59*, 163–172.
- Yahagi, R.; Onishih, H.; Machida, Y. Preparation and Evaluation of Double-phased Mucoadhesive Suppositories of Lidocaine Utilizing Carbopol and White Beeswax. *J. Controlled Release* **1999**, *61*, 1–8.
- Watanabe, K.; Yakou, S.; Takayama, K.; Isowa, K.; Nagai, T. Rectal Absorption and Mucosal Irritation of Rectal Gels Containing Buprenorphine Hydrochloride Prepared with Water-soluble Dietary Fibers, Xanthan Gum and Locust Bean Gum. *J. Controlled Release* **1996**, *38*, 29–37.
- Ridolfo, A.S.; Kohlstaedt, K.J. A Simplified Method for the Rectal Installation of Theophylline. *Am. J. Med. Sci.* **1959**, *237*, 585–589.
- De Leede, L.G.-J.; De Boer, A.G.; Portzgen, E.; Feijen, J.; Breimer, D.D. Rate-Controlled Rectal Drug Delivery in Man with a Hydrogel Preparation. *J. Contr. Rel.* **1986**, *4*, 17–24.
- Theeuwes, F. Rate Controlled Delivery by the Oral and Rectal Route. *Topics in Pharmaceutical Sciences*; Breimer, D.D., Speiser, P., Eds.; Elsevier Science Publishers: Amsterdam, 1985.
- Samuels Brusse, F.; Bertens, A.M.; Breimer, D.D. Consumption of Medicines by Outpatients in the Nijmegen Region, 1974–1975. *Pharmaceut. Week.* **1979**, *114*, 385–403.
- Tondury, C. *Angewandte und Topographische Anatomie*; Georg Thieme Verlag: Stuttgart, 1959; 200.
- De Boer, A.G.; Breimer, D.D. Rectal Absorption: Portal or Systemic. *Drug Absorption*; Presscott, L.F., Nimmo, W.S. Eds.; ADIS Press: New York, 1979.
- Caldwell, L.; Nishihata, T.; Rytting, J.; Higuchi, T. Lymphatic Uptake of Water-Soluble Drugs after Rectal Administration. *J. Pharm. Pharmacol.* **1982**, *34*, 520–522.
- Sutton, S.C.; Forbes, A.E.; Cargill, R.; Hochman, J.H.; LeCluyse, E.L. Simultaneous In Vitro Measurement of Intestinal Tissue Permeability and Transepithelial Electrical-Resistance (Teer) Using Sweetana Grass Diffusion Cells. *Pharm. Res.* **1992**, *9*, 316–319.
- Saffran, M.; Bedra, C.; Kumar, G.S.; Neckers, D.C. Vasopressin: Model for the Study of Effects of Additives on the Oral and Rectal Administration of Peptide Drugs. *J. Pharm. Sci.* **1988**, *77*, 33–38.
- Moore, J.A.; Pletcher, S.A.; Ross, M.J. Absorption Enhancement of Growth Hormone from the Gastrointestinal Tract of Rats. *Int. J. Pharmaceut.* **1986**, *34*, 35–43.
- Nishihata, T.; Okamura, Y.; Inagaki, H.; Sudho, M.; Kamada, A.; Yagi, T.; Kawamori, R.; Schichiri, M. Trials of Rectal Insulin Suppositories in Healthy Humans. *Int. J. Pharmaceut.* **1986**, *34*, 157–161.
- Davenport, H.W. *Physiology of the Digestive Tract: An Introductory Text*, 4th Ed.; Year Book Medical Publishers: Chicago, 1982; 229.
- Fix, J.A.; Porter, P.A.; Leppert, P.S. Involvement of Active Sodium Transport in the Rectal Absorption of Gentamicin Sulfate in the Presence and Absence of Absorption Promoting Adjuvants. *J. Pharm. Sci.* **1983**, *72*, 698–700.
- Fix, J.A.; Alexander, J.; Cortese, M.; Engle, K.; Leppert, P.; Repta, A.J. Short Chain Alkyl Esters of L-Dopa as Prodrugs for Rectal Absorption. *Pharm. Res.* **1989**, *6*, 501–505.
- Fix, J.A.; Engle, K.; Porter, P.A.; Leppert, P.S.; Selk, S.J.; Gardner, C.R.; Alexander, J. Acylcarnitines, Drug Absorption-Enhancing Agents in the Gastrointestinal Tract. *Am. J. Physiol.* **1986**, *251*, G332–G340.
- Davis, S.S.; Burnham, W.R.; Wilson, P.; O'Brien, J. Use of Adjuvants for Enhancement of Rectal Absorption of Cefoxitin in Humans. *Antimicrob. Agents Chemother.* **1985**, *28*, 211–215.
- Grass, G.M.; Sweetana, S.A. In Vitro Measurement of Gastrointestinal Tissue Permeability Using a New Diffusion Cell. *Pharm. Res.* **1988**, *5*, 372–374.

22. Grass, G.M.; Sweetana, S.A. A Correlation of Permeabilities for Passively Transported Compounds in Monkey and Rabbit Jejunum. *Pharm. Res.* **1989**, *6*, 857–861.
23. Gueant, J.L.; Masson, C.; Schohn, H.; Girr, M.; Saunier, M.; Nicholas, J.P. Receptro-Mediated Endocytosis of the Intrinsic-Factor Cobalamin Complex in HT-29, a Human Colon-Carcinoma Cell-Line. *FEBS* **1992**, *297*, 229.
24. Rubas, W.; Jezyk, N.; Grass, G.M. Comparison of the Permeability Characteristics of a Human Colonic Epithelial (Caco-2) Cell Line to Colon of Rabbit, Monkey, and Dog Intestine and Human Drug Absorption. *Pharm. Res.* **1993**, *10*, 113–118.
25. Anderberg, E.K.; Artursson, P. Epithelial Transport of Drugs in Cell Culture. Part 8. Effect of Sodium Dodecyl on Cell Membrane and Tight Junction Permeability in Human Intestinal Epithelial (Caco-2) Cells. *J. Pharm. Sci.* **1993**, *82*, 392–398.
26. Ng, K.Y.; Borchardt, R.T. Biotin Transport in a Human Intestinal Epithelial-Cell Line (Caco-2). *Life Sci.* **1993**, *53*, 1121–1127.
27. Inui, K.I.; Yamamoto, M.; Saito, H. Transepithelial Transport of Oral Cephalosprins by Monolayers of Intestinal Epithelial-Cell Line Caco-2 Secific Transport-System in Apical and Basolateral Membranes. *J. Pharmacol. Exp. Ther.* **1992**, *261*, 195–201.
28. de Boer, A.G.; Moolenaar, L.G.; de Leede, L.G.J.; Breimer, D.D. Rectal Drug Administration: Clinical Pharmacokinetic Considerations. *Clin. Pharmacokinet* **1982**, *7*, 285–311.
29. Kamiya, A.; Ogata, H.; Fung, H.-L. Rectal Absorption of Nitroglycerin in the Rat: Avoidance of First Pass Metabolism as a Function of Rectal Length Exposure. *J. Pharm. Sci.* **1982**, *71*, 621–624.
30. Ritschel, W.A.; Ritschel, G.B. Rectal Route of Peptide and Protein Drug Delivery. *Rectal Therapy*; Glas, B., de Blaey, C.J., Eds.; J. R. Prous Publishers: Barcelona, 1984; 67.
31. Rytting, J.H. Rectal Route of Peptide and Protein Drug Delivery. *Peptide and Protein Drug Delivery*; Lee, V.H.L., Ed.; Marcel Dekker, Inc.: New York, 1991; 579–593.
32. Voigt, R.; Falk, G. Die Wasserloslichkeit von Arzneimitteln als Kriterium für die Arzneistoffliberation aus Fettartigen, Galenischen Grundstoffen (cetylum phthallcum, Lasupol G) unter Berücksichtigung Viskosaltserhöhender Hilfsstoffe. *Pharmazie*. **1968**, *23*, 709–714.
33. Nishihata, T.; Rytting, J.H. Absorption-prompting Adjustments: Enhancing Action on Rectal Absorption. *Adv. Drug Delivery Rev.* **1997**, *28*, 205–228.
34. Yoshioka, S.; Caldwell, L.; Higuchi, T. Enhanced Rectal Bioavailability of Polypeptides Using Sodium 5-methoxysalicylate as an Absorption Promoter. *J. Pharm. Sci.* **1982**, *71*, 593–594.
35. Fasano, A.; Budillon, G.; Guandalini, S.; Cuomo, R.; Parrilli, G.; Cangioti, A.M.; Morroni, M.; Rubino, A. Bile-Acids Reversible Effects on Small Intestinal Permeability — An In Vitro Study in the Rabbit. *Dig. Dis. Sci.* **1990**, *35*, 801–808.
36. Fix, J.A.; Leppert, P.A.; Porter, P.A.; Alexander, J. Use of Phenothiazines to Enhance the Rectal Absorption of Water Soluble Compound. *J. Pharm. Pharmacol* **1984**, *36*, 286–288.
37. Murakami, T.; Yata, N.; Tamauchi, H.; Kamada, A. Studies of Absorption Promotols for Rectal Delivery Preparations II. A Possible Mechanism of Promoting Efficiency of Examine Derivatives in Rectal Absorption. *Chem. Pharm. Bull.* **1982**, *30*, 659–665.
38. Tomita, M.; Sawada, T.; Ogawa, T.; Ouchi, H.; Hayashi, M.; Awazu, S. Differences in the Enhancing Effects of Sodium Caprate on Colonic and Jejunal Drug Absorption. *Pharm. Res.* **1992**, *9*, 648–653.
39. Nishihata, T.; Tomida, H.; Frederick, G.; Rytting, J.H.; Higuchi, T. Comparison of the Effects of Sodium Salicylate, Disodium Ethylenediaminetetraacetic Acid and Polyoxyethylene 23 Lauryl Ether as Adjuvants for the Rectal Absorption of Sodium Cefoxitin. *J. Pharm. Pharmacol* **1985**, *37*, 159–163.
40. Bechgaard, E. Perfusion Technique of the Intact Human Rectum and Absorption of Salicylic Acid from Perfused Human Rectum. *Acta Pharmacol. Toxicol.* **1973**, *33*, 123, 129–135.
41. Nishihata, T.; Lee, C.-S.; Rytting, J.H.; Higuchi, T. The Synergistic Effects of Concurrent Administration to Rats of EDTA and Sodium-Salicylate on the Rectal Absorption of Sodium Cefoxitin and the Effects of Inhibitors. *J. Pharm. Pharmacol* **1987**, *39*, 180–184.

Drug Delivery—Pulsatile Systems

Till Bussemer

Aventis Pharma Deutschland GmbH, Frankfurt am Main, Germany

Roland A. Bodmeier

Freie Universität Berlin, Berlin, Germany

INTRODUCTION

Historically, extended release dosage forms were developed, which release the drug continuously over longer periods of time. These dosage forms offer many advantages, such as the nearly constant drug levels at the site of action and therefore minimization of peak-trough-fluctuations, reduced frequency of administration, and an improved patient compliance. In recent years, pulsatile release systems have gained increasing interest. Ideally, with a pulsatile system, the drug is released rapidly and completely after a defined lag time of no drug release (Fig. 1A). Alternative terms used to describe pulsatile release are delayed or sigmoidal release. Besides one-pulse systems, multipulse systems release the drug in subsequent pulses.^[1] The application of pulsatile release systems can be advantageous to adapt a drug therapy to chronopharmacological needs or to target a drug to a specific site in the gastrointestinal tract (GIT), e.g., to the colon.

Pulsatile drug delivery systems (DDS) can be classified in site-specific and time-controlled systems. Drug release from site-specific systems depends on the environment in the GIT, e.g., on pH, presence of enzymes, and the pressure in the GIT. In contrast, time-controlled DDS are independent of the biological environment. The drug release is controlled only by the system. Time-controlled pulsatile delivery has been achieved mainly with drug-containing cores, which are covered with release-controlling layers. The cores serve as reservoirs, which protect the core from the environment, e.g., water, acidic pH, and enzymes, until the drug is released after the lag phase. The coatings can erode/dissolve, rupture, or alter their permeability at the required time. Alternatively, capsular-shaped cores can be combined with release-controlling plugs. These strategies to release drugs in a pulsatile manner are reviewed in detail in this article.

CHRONOPHARMACOLOGY OF DRUG EFFECTS

The dependence of several diseases and body function on circadian rhythms is well known. A genetic control of a

“master clock” located in the nucleus suprachiasmaticus was recently proposed.^[2]

A number of hormones, such as renin, aldosterone, or cortisol, show distinct daily fluctuations.^[3] Circadian rhythms in the onset and extent of disease symptoms were observed, including diseases such as bronchial asthma, myocardial infarction, angina pectoris, rheumatic disease, ulcer disease, and hypertension.^[4]

The incidence of asthmatic attacks increased during the early morning hours with a maximum at 4 A.M.^[5] (Fig. 2). A treatment based on a theophylline controlled release dosage form resulting in a constant drug plasma level would not be optimal. Therefore a therapeutic scheme taking into account diurnal variation should be more effective. This could be realized by a pulsatile dosage form, taken at bedtime with a programmed drug release in the early morning hours.

Circadian effects were also observed for the pH and acid secretion in the stomach. Despite treatment of patients with a continuous infusion of famotidine, a H₂ antagonist, the pH in the stomach showed diurnal fluctuations between pH 7 and pH 2–3.^[6] Therefore a constant blood level, as obtained with a drug infusion and possibly achieved with conventional controlled drug delivery systems, does not always lead to a constant pharmacological effect. Similar effects were observed for the gastric pH in ranitidine-treated patients.^[7] Another example for the need of chronopharmacological adaptation is the treatment of pain in a study with patients suffering from post-operative pain. The analgesia requirements followed a diurnal rhythm with peaks at 9 A.M. and 8 P.M.^[8]

Chronopharmacology can affect the drug therapy in two ways, either in daily variations in pharmacodynamic effects or in pharmacokinetics.^[9] Many drugs were studied with respect to their pharmacokinetics and chronopharmacology, including analgesics, anticancer drugs, antibiotics, psychoactive drugs, local anesthetics, anti-asthmatics, anticonvulsants, and beta-blockers.^[10] Beta-receptor blocking agents reduced ischemic events mainly during the morning hours.^[11]

All body functions involved in absorption, distribution, and elimination of drugs can be dependent on circadian rhythms. For example, the gastric emptying time of solids



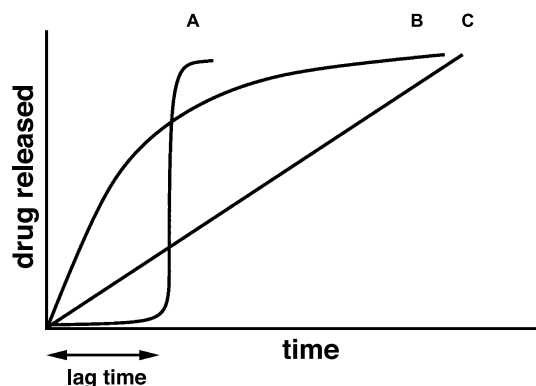


Fig. 1 Drug release profiles: (A) pulsatile, (B) and (C) conventional extended release.

is faster in the morning than in the afternoon.^[12] Blood perfusion of the gastrointestinal tract was also found to be higher during early morning hours, which could affect the absorption via passive diffusion.^[13,14] Especially for lipophilic drugs, the time of maximum plasma concentration, t_{\max} , may decrease and maximum plasma concentration, C_{\max} , may increase when applied in the morning hours.

These findings lead to the requirement of a time-programmed therapeutic scheme, whereby the drug is at the site of action at the right time in the required amount. This can be realized with pulsatile drug delivery systems.

SYSTEMS WITH ERODING OR SOLUBLE COATINGS

Most pulsatile delivery systems are reservoir devices coated with a barrier layer. The barrier dissolves or erodes

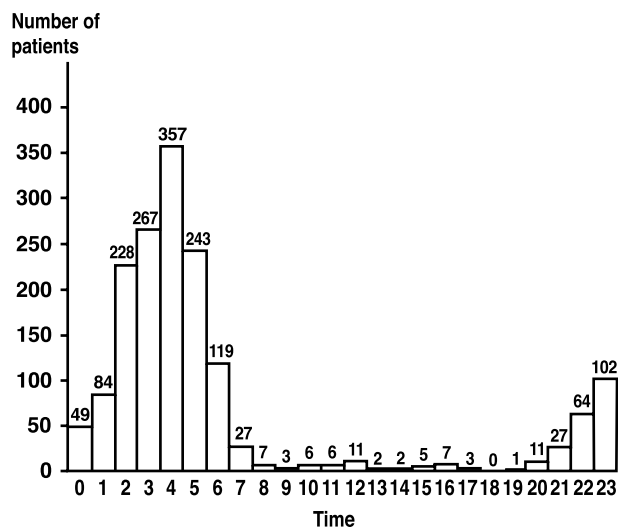


Fig. 2 Incidence of asthmatic attacks in about 1600 patients during a 24-hr cycle. (From Ref. [5].)

after a specified lag time, after which the drug is released rapidly from the reservoir core. In general, the lag time prior to drug release can be controlled by the thickness of the coating layer.

Various lag times have been achieved with press-coated tablets, where the press-coated barrier layer consisted of a mixture of a soluble polymer, hydroxypropylmethyl cellulose (HPMC), and different water-insoluble polymers, such as ethylcellulose, Eudragit[®] RS, or polylactic acid in different ratios.^[15] The release medium permeates through the coating and then results in disintegration of the tablet, whereby the lag time prior to disintegration decreases with increasing proportion of the water-soluble polymer.

The Chronotopic[®] system (Fig. 3) consisted of a core tablet containing the drug and a HPMC layer, optionally coated with an outer enteric coating.^[16–19] The lag time prior to drug release was controlled by the thickness and the viscosity grade of the HPMC layer. After erosion or dissolution of the rubbery HPMC layer, a distinct pulse was observed. To avoid retarding effects in the drug release phase, the thickness as well as the viscosity grade of the HPMC layer should be limited.^[20] The system probably works best for poorly water-soluble drugs. Highly water-soluble drugs could possibly diffuse through the swollen HPMC layer prior to complete erosion. In addition to core tablets, this principle was applied to hard and soft gelatin capsules.^[21] The coating was applied by spraying an aqueous solution of HPMC (Methocel[®] E50) onto the capsules up to a weight gain of 20%. The lag time of coated capsules was longer than that of tablets at the same coating level. This was explained by the lack of disintegration power of the capsules, when compared to the coated tablets. One problem of the coated tablets was the attrition of the swollen polymer from the edges during dissolution test; this would also lead to shortened erosion rates.

The following examples describe delivery systems, whereby an impermeable, insoluble layer protects several

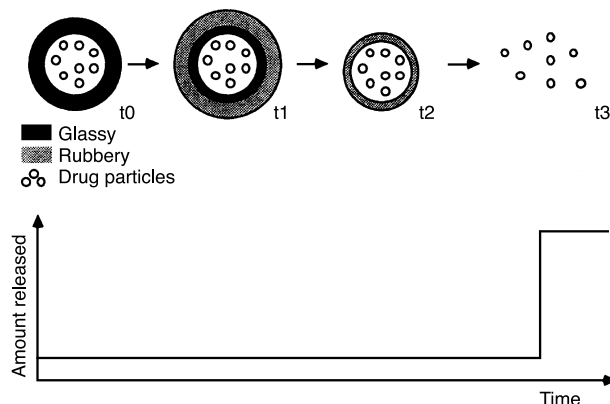


Fig. 3 The Chronotopic[®] system. (From Ref. [18].)

sides of the system from the dissolution medium and the drug release therefore occurs only from one side of the system.

Lippold and Möckel proposed a prototype, which consisted of a three-layer HPMC tablet, containing, from the outside to the inside, an immediate release layer, an intermediate drug-free layer, and an inner drug-containing layer (second pulse dose of drug).^[22] A specially designed holder, a tight-fitting metal mould, was used to allow only one surface to come into contact with the release medium. The hydrophilic matrix eroded continuously, first releasing the drug from the immediate release layer, followed by a period of no drug release (the length of which was controlled by the thickness and therefore the erosion time of the intermediate drug-free HPMC layer) and then the second drug release pulse from the inner layer. The release profile showed a sharp increase in the beginning (first pulse; immediate dose from the top layer released), followed by a lag time, and again a drug release phase (second pulse). The lag time increased from 2.5 to 6.5 hr with increasing thickness of the HPMC layer from 1.44 to 2.87 mm. The HPMC swelled and formed a gel-like layer, thus reducing the release rate of the second pulse. This system was suggested as a prototype for a buccal tablet because of the limited transit time of a solid dosage form through the stomach and the small intestine. A buccal application supposedly would overcome this limitation.

For a final dosage form, a coating for the three sides of the system is still needed in order to overcome the need of the metal mould, which was used during the *in vitro* tests.

A release pattern with two pulses was obtained from a three-layer tablet consisting of two drug-containing layers, separated by a drug-free gellable polymeric barrier layer.^[23,24] The three-layer tablet (Fig. 4) was coated on three sides with an impermeable coating (labeled d), consisting of ethylcellulose. Upon contact with dissolution fluids, the initial dose incorporated into the top layer (a) was released rapidly from the uncoated tablet surface. The second pulse was obtained from the bottom layer (c) after the gelled barrier layer (b) was broken by expanding disintegrants present in the bottom layer. The superdisintegrants were cross-linked polyvinylpyrrolidone

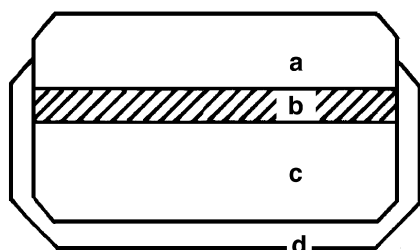


Fig. 4 A three-layer pulsatile system. (From Ref. [25].)

(crospovidone, Polyplasdone[®] XL) and sodium starch glycolate (Primojel[®]). The lag time could be controlled by the choice of the gelling materials (HPMC types with different molecular weight). However, the drug dissolution period after the lag time was quite long due to the loss of swelling efficiency of the disintegrants present in layer (c), which slowly absorbed moisture during the slow gelation of the barrier.^[25] Two distinct plasma peaks obtained *in vivo* (1 hr and 4.5–5 hr after administration) corresponded to the *in vitro* results (immediate release and a second dose released after 40 min). The deviation between *in vitro* and *in vivo* lag time for the second dose was not explained. The main drawback of this dosage form was the complicated manufacturing procedure, including the coating process. The coating step had to avoid the coating of one side (release side) of the tablet, which was obtained by manual coating with a special designed tablet holder. The design of this dosage form has been improved by the application of a press-coating procedure (instead of spray-coating) using a high viscosity HPMC grade.^[26] The obtained lag times ranged from 2 to 3 hr for diclofenac, tested in deionized water, and from 2 to more than 10 hr for ibuprofen, tested in pH 7.5 medium. Although the composition of the system was the same, the ibuprofen-containing tablet showed longer lag times. The authors explained this observation with the use of different dissolution media.

Press-coated tablets were produced on modified rotary tablet machines using special tools with hollow punches, where first an inner tablet was formed and then the outer shell was compressed in the same die (“tablet-within-a-tablet”).^[27]

Another dosage form with an erosion-controlled lag time had a drug-containing core, which was incorporated into a compressed, hollow cylinder consisting of hydroxypropyl cellulose (HPC).^[28] The flat surfaces of the tablet were coated with an impermeable polymer, poly(ethylene vinyl acetate) (Fig. 5). The delivery system was prepared by hand: a hole was drilled into a tablet to obtain the hollow matrix, the inner drug core was placed into this hole, and the system was coated by hand on the two flat base surfaces. This preparation would be quite complicated for large-scale manufacturing, and the use of a nonapproved polymer and benzene as a coating solvent would limit the application of this system. Lag times between 6 and 11 hr were achieved with either a fast drug release after the lag time (using microcrystalline cellulose or lactose in the core) or sustained release (with HPC in the core). The lag time increased with increasing thickness of the matrix cylinder. Lactose-containing cores developed an internal osmotic pressure, leading to a fast separation of the coating from the core. Thus drug was delivered before complete erosion of the outer HPC matrix. On the other hand, microcrystalline cellulose-cores had no osmotic effect, coating separation did not



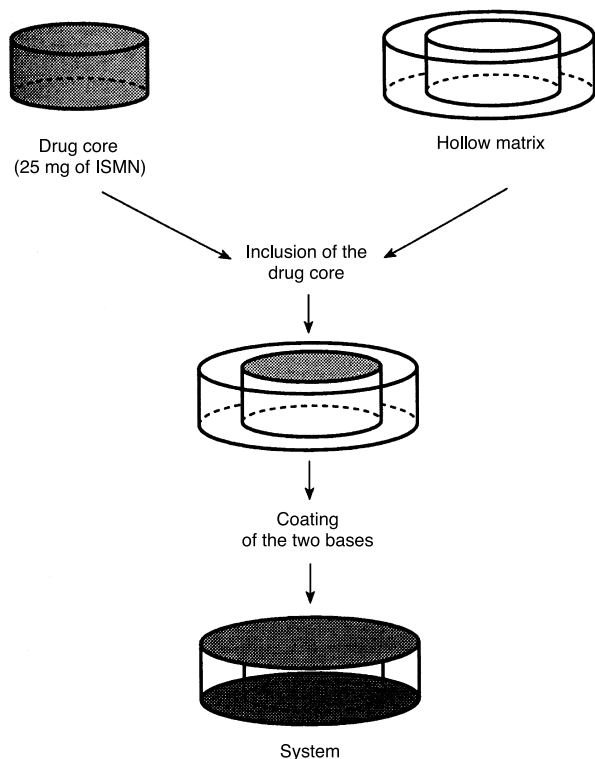


Fig. 5 Eroding system with hollow cylinder and coated surfaces. (From Ref. [28].)

occur, and the lag time was longer. The complete erosion of the matrix was necessary to release the drug.

The lipid barrier of the Time Clock[®] system, containing carnauba wax and beeswax, eroded or was emulsified into aqueous media because of incorporated surfactants (polyoxyethylene sorbitan monooleate).^[29] The lag time increased with increasing coating thickness and was independent of the environmental pH. In vivo assessment with gamma-scintigraphy of an enterically coated Time Clock[®] system was in good agreement with the in vitro predicted data with regard to the lag time.^[30] When the viscosity of the dissolution medium of the in vitro test was raised to 120 cps, the lag time was prolonged to the same value as obtained with in vivo data. The advantage of this system was its ease of manufacturing without the need of special equipment. However, controlled release delivery systems based on lipids may have a high in vivo variability (e.g., food effects).

SYSTEMS WITH CHANGED MEMBRANE PERMEABILITY

The permeability and water uptake of Eudragit[®] RS or RL [chemical name, poly(ethyl acrylate, methyl methacrylate,

trimethylammonioethyl methacrylate chloride)], can be influenced by the presence of different counterions in the release medium.^[31] It was found that theophylline was released faster from Eudragit[®] RS-coated pellets, when succinic, acetic, glutaric, tartaric, malic, or citric acid are present in the release medium.^[32] Increased permeability was explained by the higher hydration of the film, also expressed as the “free volume.”

Several delivery systems with sigmoidal or pulsatile release patterns were derived on this ion exchange. The sigmoidal release system (SRS) consisted of pellet cores, containing drug and succinic acid, coated with Eudragit[®] RS.^[33] The lag time was controlled by the rate of water influx through the coating. The water then dissolved succinic acid and the drug inside the core, and the acid solution increased the drug permeability of the hydrated polymer film by interacting with the quaternary ammonium groups.

In a similar system, theophylline and sodium acetate, acting as the permeability modifying salt, were layered on sugar pellets, followed by coating with Eudragit[®] RS30D.^[34] The lag time increased with increasing thickness of the outer membrane. The slope of the drug release phase was independent of the thickness but was influenced by the amount of the salt in the system.

The release profile of systems based on permeability changes appeared to depend strongly on the physicochemical properties of the drug and its interaction with the membrane. A pulsatile release profile may be obtained for some particular drug molecules in a specific formulation but cannot be generally applied to all drugs.

RUPTURABLE SYSTEMS

The other class of reservoir-type pulsatile systems is based on rupturable coatings in contrast to the swellable/erodible layers of the previous section. The drug is released from a core (tablet or capsule) after rupturing of a surrounding polymer layer, caused by a pressure build-up within the system. The pressure necessary to rupture the coating can be achieved with gas-producing effervescent excipients, an increased inner osmotic pressure or swelling agents, such as cellulose ethers, polysaccharides, or superdisintegrants.

Drug-containing tablets with an effervescent mixture of citric acid and sodium bicarbonate coated with ethylcellulose resulted in a pulsatile release after rupturing of the coating, which was caused by the carbon dioxide development after water penetration into the core.^[35] The rupturing strongly depended on the mechanical properties of the coating layer: highly flexible films, such as Eudragit[®] RS, with high elongation and low elastic moduli, ruptured after a certain lag time as a consequence of the effervescent reaction but left only small fissures

within the film. Therefore the drug release was prolonged and not immediate after the lag time. Using a mechanically weaker and nonflexible film, such as ethylcellulose, plasticized with 20% w/w DBS, the drug release was sigmoidal and reproducible (Fig. 6). The lag time before drug release increased with increasing coating level. Moreover, the disintegrant properties of microcrystalline cellulose, which was primarily used as a filler, supported the film rupturing.

Osmotic pressure was the mechanism of a rupturable dosage form, which was proposed by Baker in 1976.^[36] The core tablet, which contained a drug and a disintegrant, was coated with cellulose derivatives such as ethylcellulose or cellulose acetate. The core protection was defined as the time until the coating ruptured and the drug was released.

Another system was based on a swelling core tablet and a surrounding coating consisting of a combination of hydrophobic and hydrophilic polymers.^[37] The insoluble hydrophilic polymer, such as calcium pectinate or calcium alginate, was dispersed in the coating and served as a channel-former in order to control the water penetration. The core contained a swelling, but water-insoluble polymer, a hardness enhancer (microcrystalline cellulose), and a disintegrant to achieve a fast disintegration after the membrane burst.

Also, soft and hard gelatin capsules could be coated with a swelling layer followed by coating with a rupturable polymeric layer^[38] (Fig. 7). Superdisintegrants, such as Ac-Di-Sol[®] (croscarmellose sodium) or low-substituted hydroxypropyl cellulose (L-HPC), were used as swelling substances, which resulted in a film rupture followed by a rapid drug release. Ethylcellulose or cellulose acetate forms the outer rupturable polymeric membranes. The lag time was controlled by the composition of

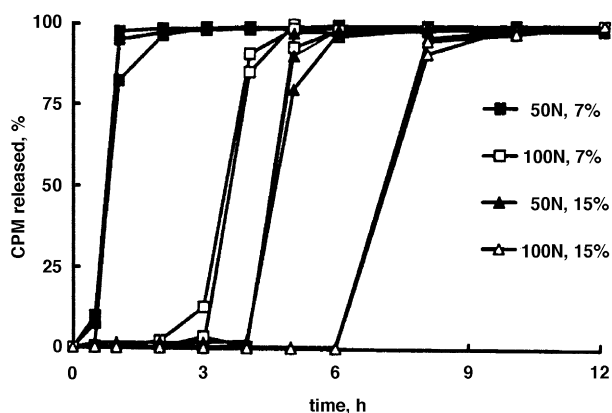


Fig. 6 Pulsatile drug delivery system based on coated effervescent cores—influence of the coating level (7% and 15% w/w) and hardness (50 and 100 N) on the drug release/lag time (coating, EC/DBS; core, 30%, w/w, effervescent agents, microcrystalline cellulose). (From Ref. [35].)

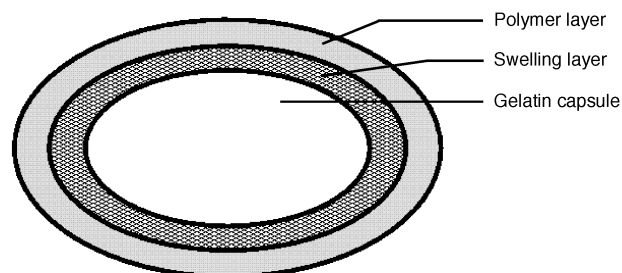


Fig. 7 Capsule-based rupturing pulsatile delivery system. (From Ref. [38].)

the outer polymer layer. Water-soluble polymers such as HPMC increased the permeability and therefore reduced the lag time. The advantage of these capsule-based systems was that both solid and liquid drug formulations could be delivered.

The systems described above are single unit systems. The rupturing of the outer coating of the individual units of a multiple unit dosage form can be induced by the same principles as described for single unit systems.

Multiparticulate drug delivery systems (e.g., pellets) have various advantages, when compared to single unit dosage forms.^[39] These advantages include a reproducible and generally short gastric residence time, no risk of dose dumping, and the flexibility to blend pellets with different compositions or release patterns. However, the potential drug loading of a multiparticulate system is lower because of the proportionally higher need for excipients (e.g., sugar cores).

The time-controlled explosion system (TES) was a multiparticulate system, whereby the drug was layered on an inner core, followed by a swellable layer (e.g., hydroxypropyl cellulose) of optimal thickness (at least 180 μm) and an insoluble polymeric top layer (e.g., ethylcellulose).^[40–43] Upon water ingress, the swellable layer expanded resulting in film rupturing with subsequent rapid drug release. The release was described to be independent of the environmental pH and drug solubility. The lag time increased with increasing coating level and higher amounts of talc or lipophilic plasticizer in the coating, and the release rate increased with increasing concentration of the osmotically active agents. In vivo studies of the time-controlled explosion system, which had an in vitro lag time of 3 hr, showed first drug blood levels after 3 hr and maximal blood levels after 5 hr.^[44]

A combination of osmotic and swelling effects was used in the system developed by Amidon and Leesman.^[45] The permeability-controlled systems consisted of a core, containing an osmotically active substance, a swelling substance, and the drug. NaCl and sorbitol were used as osmotic substances and Na carboxymethyl cellulose as the swelling material. These cores were coated with an



insoluble, semipermeable polymer, such as cellulose acetate. The time T_P , when the insoluble film ruptured, was described by the following equation:

$$T_P = \frac{V_e^* + V_d}{A \cdot a \cdot L_P \cdot \Delta\pi}$$

where T_P represented the time until film rupture (lag time), V_e^* was the volume, to which the tablet has to be expanded for film break, V_d was the displaceable volume inside the core, A was the surface area of the film, a was a constant, L_P was the water permeability of the film, and $\Delta\pi$ was the osmotic pressure difference across the film. The constant a was

$$a = \frac{\Delta r^* / r_0}{e^*}$$

where Δr^* was the critical radius increase, r_0 was the initial radius of the unit, and e^* was the critical strain. This model was based on the following assumptions: the units (tablets, pellets) have a spherical shape, the thickness of the film and the osmotic pressure difference are constant, and the values of the critical stress and strain for film rupture do not change. In practice, the changing mechanical properties of the film, the changing permeability, and weak points (defects) of the film would have to be considered.

Chen has proposed a system with a core, containing an osmotically active substance (NaCl) and the drug, which was coated by an insoluble, permeable membrane.^[46] The coatings included different types of polyacrylate-poly-methacrylate-copolymers (Eudragit[®] NE30D, Eudragit[®] RS30D) and magnesium stearate, which reduced the water permeability of the membrane, thus allowing the use of thinner films for long lag times. Otherwise, thicker films had to be used, which were more difficult to rupture.

The lag time of a similar system was controlled by the addition of an enteric polymer to the surrounding insoluble membrane polymer (e.g., ethyl cellulose). The enteric polymer, poly(methacrylic acid-methylmethacrylate) copolymer (Eudragit[®] S) or hydroxypropyl methylcellulose phthalate (HPMCP), did not dissolve in the acidic pH of the stomach but dissolved in the small intestine, thus weakening the membrane and resulting in rupturing after a predetermined time. Expanding of the core upon water ingress was achieved by the presence of Explotab[®] (starch glycolate), a swelling substance. Again, the inclusion of water-insoluble agent (magnesium stearate) in the outer membrane resulted in thinner, less permeable films with better rupturing properties.^[47]

An osmotically active, drug-containing pellet core has been coated with cellulose acetate, a semipermeable polymer, which is permeable for water but impermeable for drugs.^[48,49] The lag time increased with increasing

coating level and higher amounts of talc or lipophilic plasticizer in the coating, and the release rate increased with increasing concentration of the osmotically active agents. The addition of osmotically active salts, such as sodium chloride, was necessary to achieve a pulsatile release. Otherwise, the release was extended after the lag time because of the lower degree of core swelling, which resulted in only small fissures and not in a complete rupturing of the coating.

Heng et al. investigated multilayered pellets, containing neutral core pellets, an inner drug layer, an intermediate HPMC layer (swelling layer) and an outer insoluble diffusion layer, consisting of Eudragit[®] RS. The addition of sodium chloride to the HPMC layer decreased the rate of swelling and thus delayed the bursting of the pellets.^[50] This phenomenon was first unexpected because of osmotic effects of the salt, which should promote water uptake. The longer time until rupturing was explained by the competition between sodium chloride and HPMC for the imbibed water.

A theoretical approach to calculate the burst time from an osmotically active spherical capsule, which depended on the initial radius, wall thickness, osmotic pressure of the contents, and the material of the capsule, has been presented.^[51] This approach assumed that the spherical core increases in size upon osmotic water influx, leading to an elongation of the membrane until a certain yield stress is reached to rupture it. The rate of volume increase was described by the following equation:

$$\frac{dV}{dt} = D \frac{A}{l} (\Pi_0 - p)$$

where dV/dt was the rate of volume increase, D the dialysis permeability through the wall, A the area of the sphere, l the wall thickness, Π_0 the osmotic pressure difference across the wall, and p the internal pressure of the sphere.

The dimensionless values, $N_1 = \frac{Y}{M}$ and $N_2 = \frac{\Pi_0 r_0}{2Ml_0}$ (whereby Y was the yield stress at which the sphere bursts, M was Young's modulus of elasticity, r_0 the initial radius, and l_0 the initial wall thickness), were derived from the volume of a sphere and Hook's law for elastic materials. N_1 determined the radius, at which the spheres burst, and is the ratio of the material properties, yield stress (Y), and Young's modulus, a value for the stiffness, (M). N_2 determined the initial expansion rate. It was concluded that the dimensionless burst time, t_b^* , increases with increasing N_2 . When $N_2 > 2N_1$, the burst time t_b^* is proportional to l_0 . But if $N_2 < 2N_1$, the bursting time becomes sensitive to the other values.

Increased permeability and decreased wall thickness, concomitant with appropriate mechanical membrane properties, would lead to a faster water influx and shorter lag times of such an expanding system.

CAPSULAR-SHAPED SYSTEMS

Several single unit pulsatile dosage forms with a capsular design have been developed. Most of them consist of an insoluble capsule body, which contains the drug, and a plug, which prevents drug release during the lag phase. Mechanisms of plug removal include dissolution, erosion, or induced pushing-out of the plug by swelling or osmotic pressure.

The Pulsincap[®] system consisted of a water-insoluble body (hard gelatin capsule coated with polyvinyl chloride), filled with the drug formulation.^[52,53] The capsule

half was closed at the open end with a swellable hydrogel plug. Upon contact with dissolution media or gastrointestinal fluids, the plug swelled and pushed itself out of the capsule after a lag time, followed by a rapid release of the capsule content (Fig. 8). The lag time prior to the drug release was controlled by the dimension and the position of the plug. In order to assure a rapid release of the drug content, effervescent agents or disintegrants could be included in the drug formulation, in particular, with water-insoluble drugs. Studies in animals and healthy volunteers proved the tolerability of the formulation (e.g., absence of gastrointestinal irritation).^[54] In order to overcome the

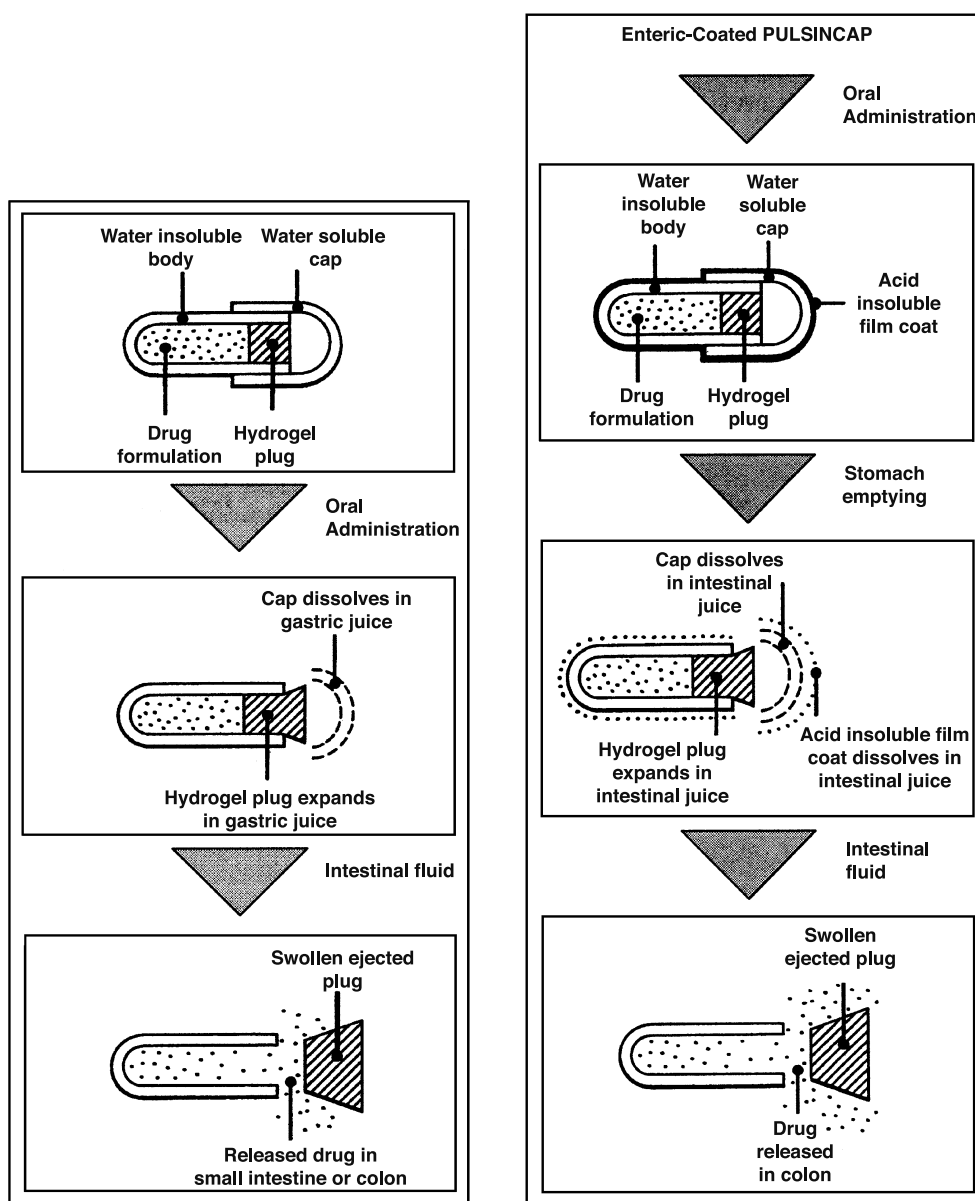


Fig. 8 The Pulsincap[®] system. (From Ref. [51].)



potential problem of variable gastric residence time of a single unit dosage form, the Pulsincap[®] system was coated with an enteric layer, which dissolved upon reaching the higher pH regions of the small intestine. This allowed a more precise control of the drug release after passage of the stomach, because the transit time in the intestinal tract is less variable.^[55,56] The major drawbacks of the Pulsincap[®] system, which led to the withdrawal of commercial activities with this system, were the complicated manufacturing process, reproducibility problems, and the use of a plug material, a cross-linked polyethylene glycol based polymer, which has not been approved in pharmaceutical products.

As an alternative to swellable, cross-linked plugs, erodible plugs have been investigated for Pulsincap-like systems (Fig. 9). The tablet-shaped plugs can be produced by compression of a water-soluble, swellable polymer, such as HPMC, PVA, or polyethylene oxide.^[57] After contact with gastrointestinal fluids, the polymer plugs swelled quickly, forming a gel, followed by a transition into a sol and a subsequent period of erosion. The swelling polymer could also be combined with soluble low-molecular weight excipients, e.g., lactose, to reduce the lag time.^[58] In general, the lag time was adjusted by the choice of the molecular weight of the erodible polymer and by the thickness of the plug.

Plug degradation could also be achieved by enzymes being directly incorporated into the plug.^[59] In an example, plugs containing pectin, a natural polysaccharide, were degraded by pectinolytic enzymes, whereby the lag time of the system was controlled by the ratio of pectin to enzymes (Fig. 10).

Besides compression, erodible plugs were formed by a congealing method with melts of saturated polygly-

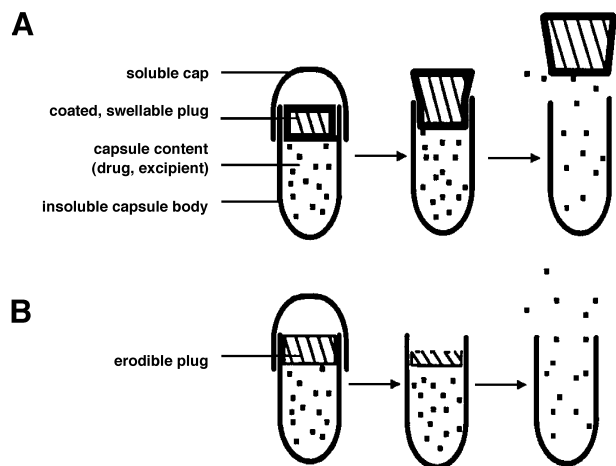


Fig. 9 Capsular-shaped pulsatile drug delivery system, consisting of an impermeable capsule body and (A) a swellable, insoluble coated plug and (B) an erodible plug. (From Ref. [56].)

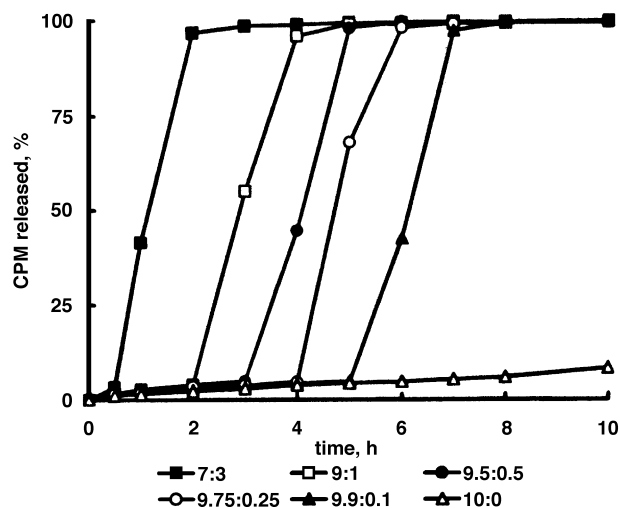


Fig. 10 Chlorpheniramine maleate (CPM) release as a function of the pectin/enzyme ratio within the plug. (From Ref. [58].)

colyted glycerides (Gelucire[®]) or glyceryl monooleate (Myverol[®]).^[57]

Another approach was based on the use of an inner osmotic pressure to push out the plug from the insoluble capsule thus allowing drug release after the lag time. The Port[®] system was composed of a gelatin capsule coated with a semipermeable membrane (e.g., cellulose acetate), which was filled with an osmotic charge and closed with an insoluble plug (e.g., lipids such as Gelucire[®]).^[60] In contact with aqueous media, water diffused across the semipermeable membrane, resulting in a higher inner pressure and ejection of the plug after a certain lag time, which was mainly controlled by the coating thickness of the semipermeable membrane. The system was already tested in a human study, showing good agreement between lag times measured in vivo and in vitro.^[61] An additional immediate or sustained release dose could be placed between the soluble gelatin cap and the insoluble plug.

In the Chronset[®] system (Fig. 11), the driving force for the drug release was an osmotically active layer in the

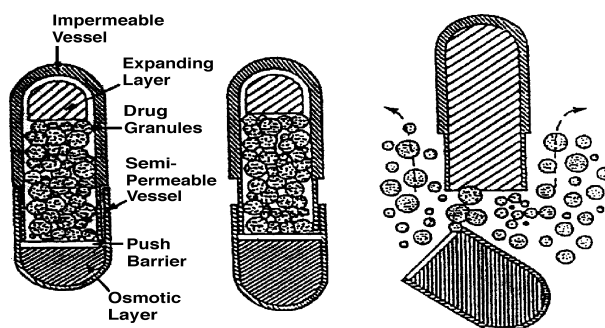


Fig. 11 The Chronset[®] system. (From Ref. [61].)

semipermeable vessel, which pushed the cap out off the impermeable vessel after a predetermined time interval.^[62] The complete release of the drug, often problematic in capsular-shaped dosage forms, was ensured by an expanding layer at the bottom of the capsule body.

Even more sophisticated were insoluble high-frequency (HF) capsules, which released the drug in a pulsed fashion after a high-frequency signal was applied externally to the human body.^[63,64] These HF capsules were used to evaluate the absorption of drugs from distinct regions within the digestive tract. A similar capsule activated by an oscillating magnetic field has been published recently, which ejected an active compound or a radioactive marker to localize the position of the dosage form in the gastrointestinal tract.^[65]

In general, the large-scale manufacturing of the abovementioned capsular-shaped pulsatile drug delivery systems appears to be complicated. Special equipment and several manufacturing steps are necessary to combine all components.

CONCLUSION

The use of pulsatile drug delivery system should be taken into consideration for drugs with a chronopharmacological behavior, a high first-pass effect, the requirement of night time dosing or site-specific absorption in the GI-tract. In recent years, a wide variety of interesting single and multiple unit systems have been developed for oral application. However, many systems are only of academic use because of their complex manufacturing process or non-approved excipients. Most systems perform quite well in vitro, but their performance in vivo has often not been tested. One major challenge will be to obtain a better understanding of the influence of the biological environment on the release performance of pulsatile delivery system in order to develop simple systems based on approved excipients with a good in vitro/in vivo correlation.

REFERENCES

1. Peppas, N.A. Fundamentals on pH- and Temperature-Sensitive Delivery Systems. In *Pulsatile Drug Delivery*, 1st Ed.; Gurny, R., Junginger, H.E., Peppas, N.A., Eds.; Wissenschaftliche Verlagsgesellschaft: Stuttgart, 1993; 41–56.
2. Lemmer, B. Genetic Aspects of Chronobiologic Rhythms in Cardiovascular Disease. In *From Molecule to Man*; Zehender, M., Breithardt, G., Just, H., Eds.; Darmstadt: Germany, 2000; 201–213.
3. Breimer, D.D. The Need for Rate and Time Programming in Future Drug Delivery. In *Pulsatile Drug Delivery*, 1st Ed.; Gurny, R., Junginger, H.E., Peppas, N.A., Eds.; Wissenschaftliche Verlagsgesellschaft: Stuttgart, 1993; 25–40.
4. Lemmer, B. Chronopharmacokinetics: Implications for drug treatment. *J. Pharm. Pharmacol.* **1999**, *51*, 887–890.
5. Dethlefsen, U.; Repges, R. Ein neues Therapieprinzip bei nächtlichem Asthma. *Med. Klin.* **1985**, *80* (2), 44–47.
6. Merki, H.S.; Witzel, L.; Kaufman, D.; Kempf, M.; Neumann, J.; Rohmel, J.; Walt, R.P. Continuous intravenous infusions of famotidine maintain high intragastric pH in duodenal ulcer. *Gut* **1988**, *29* (4), 453–457.
7. Sanders, W.; Buchi, K.N.; Moore, J.G.; Bishop, A.L. Pharmacodynamics of intravenous ranitidine after bolus and continuous infusion in patients with healed duodenal ulcers. *Clin. Pharmacol. Ther.* **1989**, *46* (5), 545–551.
8. Burns, J.W.; Hodsmann, N.B.A.; McLintock, T.T.C.; Gillies, G.W.A.; Kenny, G.N.C.; McArdle, C.S. The influence of patient characteristics on the requirements for postoperative analgesia. A reassessment using patient-controlled analgesia. *Anaesthesia* **1989**, *44* (1), 2–6.
9. Lemmer, B. Why Are So Many Biological Systems Periodic? In *Pulsatile Drug Delivery*, 1st Ed.; Gurny, R., Junginger, H.E., Peppas, N.A., Eds.; Wissenschaftliche Verlagsgesellschaft: Stuttgart, 1993; 11–24.
10. Ritschel, W.A.; Forusz, H. Chronopharmacology: A review of drugs studies. *Methods Find. Exp. Clin. Pharmacol.* **1994**, *16* (1), 57–75.
11. Lemmer, B. Chronopharmacological aspects for the prevention of acute coronary syndromes. *Eur. Heart J.* **1998**, *19* (Suppl. C), C50–C58.
12. Goo, R.H.; Moore, J.G.; Greenberg, E.; Alazraki, N.P. Circadian variation in gastric emptying of meals in humans. *Gastroenterology* **1987**, *93* (3), 515–518.
13. Lemmer, B.; Nold, G. Circadian changes in estimated hepatic blood flow in healthy subjects. *Br. J. Clin. Pharmacol.* **1991**, *32* (5), 627–629.
14. Lemmer, B. Chronopharmacological aspects of PK/PD modeling. *Int. J. Clin. Pharmacol. Ther.* **1997**, *35* (10), 458–464.
15. Korsatko-Wabnegg, B.; Korsatko, W.; Wegleitner, K. Delayed-Release Pharmaceutical Preparation. US Patent 5,151,273, September 29, 1992.
16. Gazzaniga, A.; Busetti, C.; Moro, L.; Crimella, T.; Sangalli, M.E.; Giordano, F. Evaluation of Low Viscosity HPMC as Retarding Coating Material in the Preparation of a Time-Based Oral Colon Specific Delivery System. *Proceed. Int. Symp. Control. Rel. Bioact. Mater.*, 1995; Controlled Release Society: Minneapolis, MN, 1995; 242.
17. Gazzaniga, A.; Iamartino, P.; Maffione, G.; Sangalli, M.E. Oral delayed-release system for colonic specific delivery. *Int. J. Pharm.* **1994**, *108*, 77–83.
18. Gazzaniga, A.; Sangalli, M.E.; Giordano, F. Oral chronotropic drug delivery systems: Achievement of time and/or site specificity. *Eur. J. Pharm. Biopharm.* **1994**, *40* (4), 246–250.
19. Poli, S.; Busetti, C.; Moro, L. Oral Pharmaceutical Composition for Specific Colon Delivery. EP Patent 0 572 942, December 8, 1993.
20. Maroni, A.; Sangalli, M.E.; Busetti, C.; Zema, L.; Giordano, F.

- F.; Gazzaniga, A. Chronotopic System, Influence of Various HPMC Grades as Coating Agents on the Release Behaviour. Proc. 2nd World Meeting APGI/APV, Paris, France, 1998; Arbeitsgemeinschaft für Pharmazeutische Verfahrenstechnik: Mainz, 1998; 337.
21. Maroni, A.; Sangalli, M.E.; Cerea, M.; Buseti, C.; Giordano, F.; Gazzaniga, A. Low Viscosity HPMC Coating of Soft and Hard Gelatin Capsules for Delayed and Colonic Release: Preliminary Investigations on Process Parameters and In Vitro Release Performances. Proceed. Int. Control. Rel. Bioact. Mater., Boston, MA, 1999; Controlled Release Society: Minneapolis, MN, 1999.
 22. Lippold, B.C.; Möckel, J.E. Pulsatile release from laminated methylhydroxypropyl cellulose matrices with KCl as model drug. Acta Pharm. Technol. **1990**, *36* (2), 97–98.
 23. Conte, U.; Manna, A.L.; Colombo, P. Tablet for Pharmaceutical Use Able to Release Active Substances at Successive Times. US Patent 4,865,849, September 12, 1989.
 24. Conte, U.; Colombo, P.; Manna, A.L.; Gazzaniga, A. A new ibuprofen pulsed release oral dosage form. Drug Dev. Ind. Pharm. **1989**, *15* (14–16), 2583–2596.
 25. Conte, U.; Giunchedi, P.; Maggi, L.; Sangalli, M.E.; Gazzaniga, A.; Colombo, P.; Manna, A.L. Ibuprofen delayed release dosage forms: A proposal for the preparation of an in vitro/in vivo pulsatile system. Eur. J. Pharm. Biopharm. **1992**, *38* (6), 209–212.
 26. Maggi, L.; Conte, U.; Burni, R. Delivery Device for the Release of the Active Ingredient in Subsequent Times. Proceed. Int. Control. Rel. Bioact. Mater., Boston, MA, 1999; Controlled Release Society: Minneapolis, MN, 1999.
 27. Hariharan, M.; Gupta, V.K. Encore: A Novel Compression-Coated Tablet Dosage Form. Proceed. Int. Symp. Control. Rel. Bioact. Mater., San Diego, CA, 2001; Controlled Release Society: Minneapolis, MN, 2001; 6102.
 28. Vandelli, M.A.; L., E.; Forni, F.; Bernabei, M.T. In vitro evaluation of a potential colonic delivery system that releases drug after a controllable lag-time. Eur. J. Pharm. **1996**, *43* (2), 148–151.
 29. Wilding, I.R.; Davis, S.S.; Pozzi, F.; Furlani, P.; Gazzaniga, A. Enteric coated timed released systems for colonic targeting. Int. J. Pharm. **1994**, *111*, 99–102.
 30. Pozzi, F.; Furlani, P.; Gazzaniga, A.; Davis, S.S.; Wilding, I.R. The time clock system: A new oral dosage form for fast and complete release of drug after a predetermined lag time. J. Control. Release **1994**, *31*, 99–108.
 31. Bodmeier, R.; Guo, X.; Sarabia, R.E.; Skultety, P. The influence of buffer species and strength on diltiazem release from beads coated with aqueous cationic polymeric dispersions, Eudragit RS, RL30D. Pharm. Res. **1996**, *13* (1), 52–56.
 32. Narisawa, S.; Nagata, M.; Danyoshi, C.; Yoshino, H.; Murata, K.; Hirakawa, Y.; Noda, K. An organic acid-induced sigmoidal release system for oral controlled-release preparations. Pharm. Res. **1994**, *11* (1), 111–116.
 33. Narisawa, S.; Nagata, M.; Hirakawa, Y.; Kobayashi, M.; Yoshino, H. An organic acid-induced sigmoidal release system for oral controlled-release preparations: 2. Permeability enhancement of Eudragit RS coating led by the physicochemical interactions with organic acid. J. Pharm. Sci. **1996**, *85* (2), 184–188.
 34. Beckert, T.E.; Pogarell, K.; Hack, I.; Petereit, H.-U. Pulsed Drug Release with Film Coatings of Eudragit RS30D. Proceed. Int. Symp. Control. Rel. Bioact. Mater., Boston, MA, 1999; Controlled Release Society: Minneapolis, MN, 1999; 5219.
 35. Krögel, I.; Bodmeier, R. Floating or pulsatile drug delivery systems based on coated effervescent cores. Int. J. Pharm. **1999**, *187*, 175–184.
 36. Baker, R.W. Controlled Release Delivery System by an Osmotic Bursting Mechanism. US Patent 3,952,741, April 27, 1976.
 37. Lerner, I.E.; Flashner, M.; Penhasi, A. Delayed Total Release Gastrointestinal Drug Delivery System. WO Patent 99/18938, 1999.
 38. Bussemer, T.; Bodmeier, R. Pulsatile Drug Release from Coated Capsules. AAPS Annual Meeting Abstracts, New Orleans, LA, 1999; American Association of Pharmaceutical Scientists: Alexandria, VA, 1999; 3257.
 39. Bodmeier, R. Tableting of coated pellets. Eur. J. Pharm. Biopharm. **1997**, *43*, 1–8.
 40. Ueda, Y.; Hata, T.; Yamaguchi, H.; Ueda, S.; Kotani, M. Time Controlled Explosion System and Process for Preparation the Same. US Patent 4,871,549, October 3, 1989.
 41. Ueda, S.; Hata, T.; Yamaguchi, H.; Kotani, M.; Ueda, Y. Development of a novel drug release system, time-controlled explosion system (TES): I. Concept and design. J. Drug Target. **1994**, *2*, 35–44.
 42. Ueda, S.; Yamaguchi, H.; Kotani, M.; Kimura, S.; Tokunaga, Y.; Kagayama, A.; Hata, T. Development of a novel drug release system, time-controlled explosion system (TES): II. Design of multiparticulate TES and in vitro drug release properties. Chem. Pharm. Bull. **1994**, *42* (2), 359–363.
 43. Ueda, S.; Yamaguchi, H.; Kotani, M.; Kimura, S.; Tokunaga, Y.; Kagayama, A.; Hata, T. Development of a novel drug release system, time-controlled explosion system (TES): III. Relation between lag time and membrane thickness. Chem. Pharm. Bull. **1994**, *42* (2), 364–367.
 44. Hata, T.; Shimazaki, Y.; Kagayama, A.; Tamura, S.; Ueda, S. Development of a novel drug delivery system (TES): V. Animal pharmacodynamic study and human bioavailability study. Int. J. Pharm. **1994**, *110*, 1–7.
 45. Amidon, G.L.; Leesman, G.D. Pulsatile Drug Delivery System. US Patent 5,229,131, July 26, 1993.
 46. Chen, C.-M. Multiparticulate Pulsatile Drug Delivery System. US Patent 5,260,068, November 9, 1993.
 47. Chen, C.-M. Pulsatile Particles Drug Delivery System. US Patent 5,260,069, November 9, 1993.
 48. Schultz, P.; Kleinebudde, P. A new multiparticulate delayed release system: Part I. Dissolution properties and release mechanism. J. Control. Release **1997**, *47*, 181–189.
 49. Schultz, P.; Tho, I.; Kleinebudde, P. A new multiparticulate delayed release system: Part II. Coating formulation and properties of free films. J. Control. Release **1997**, *47*, 191–199.
 50. Heng, P.W.S.; Chan, L.W.; Chew, S.H. Mechanism of

- pellet coat rupture and its effect on drug release. *Chem. Pharm. Bull.* **1999**, *47* (7), 939–943.
51. Kuethe, D.O.; Augenstein, D.C.; Gresser, J.D.; Wise, D.L. Design of capsules that burst at predetermined times by dialysis. *J. Control. Release* **1992**, *18*, 159–164.
 52. McNeill, M.E.; Rashid, A.; Stevens, H.N.E. Dispensing Device. WO Patent 90/09168, August 29, 1990.
 53. Wilding, I.R.; Davis, S.S.; Bakhshae, M.; Stevens, H.N.E.; Sparrow, R.A.; Brennan, J. Gastrointestinal transit and systemic absorption of captopril from a pulsed-release formulation. *Pharm. Res.* **1992**, *9*, 654–657.
 54. Binns, J.; Stevens, H.N.E.; McEwen, J.; Pritchard, G.; Brewer, F.M.; Clarke, A.; Stewart, J.E.; McMillan, I. The tolerability of multiple oral doses of Pulsincap capsules in healthy volunteers. *J. Control. Release* **1996**, *38*, 151–158.
 55. Davis, S.S. The design and evaluation of controlled release systems for the gastrointestinal tract. *J. Control. Release* **1985**, *2*, 27–38.
 56. Khosla, R.; Davis, S.S. Gastric emptying and small and large bowel transit of non-disintegrating tablets in fasted subjects. *Int. J. Pharm.* **1989**, *52*, 1–10.
 57. Krögel, I.; Bodmeier, R. Pulsatile drug release from an insoluble capsule body controlled by an erodible plug. *Pharm. Res.* **1998**, *15* (3), 474–481.
 58. Ross, A.C.; Macrae, R.J.; Walther, M.; Stevens, H.N.E. Chronopharmaceutical drug delivery from a pulsatile capsule device based on programmable erosion. *J. Pharm. Pharmacol.* **2000**, *52*, 903–909.
 59. Krögel, I.; Bodmeier, R. Evaluation of an enzyme-containing capsular shaped pulsatile drug delivery system. *Pharm. Res.* **1999**, *16* (9), 1424–1429.
 60. Crison, J.R.; Siersma, P.R.; Taylor, M.D.; Amidon, G.L. Programmable Oral Release Technology, Part Systems: A Novel Dosage Form for Time and Site Specific Oral Drug Delivery. *Proceed. Int. Symp. Control. Rel. Bioact. Mater.*, 1995; Controlled Release Society: Minneapolis, MN, 1995; 278.
 61. Crison, J.R.; Siersma, P.R.; Amidon, G.L. A Novel Programmable Oral Release Technology for Delivering Drugs: Human Feasibility Testing Using Gamma Scintigraphy. *Proceed. Int. Symp. Control. Rel. Bioact.*, 1996; Controlled Release Society: Minneapolis, MN, 1996; 51.
 62. Wong, P.S.L.; Theeuwes, F.; Larsen, S.D.; Song, L.C. Osmotic Device for Delayed Delivery of Agent. US Patent 5,443,459, August 22, 1995.
 63. Clear, N.J.; Milton, A.; Humphrey, M.; Henry, B.T.; Wulff, M.; Nichols, D.J.; Anziano, R.J.; Wilding, I. Evaluation of the Intelisite capsule to deliver theophylline and frusemide tablets to small intestine and colon. *Eur. J. Pharm. Sci.* **2001**, *13*, 375–384.
 64. Graul, E.H.; Loew, D.; Schuster, O. Voraussetzung für die Entwicklung einer sinnvollen Retard- und Diuretikakombination. *Therapiewoche* **1985**, *35*, 4277–4291.
 65. Wilding, I.R.; Hirst, P.H.; Connor, A.L. The Enterion Capsule: A Novel Technology for the Assessment of Drug Absorption in Man. *Proceed. Int. Symp. Control. Rel. Bioact. Mater.*, Boston, MA, 2001; Controlled Release Society: Minneapolis, MN, 1999; 6068.



DRUG SAFETY EVALUATION

Farrel L. Fort

TAP Pharmaceutical Products, Inc., Lake Forest, Illinois, U.S.A.

INTRODUCTION

Definition and Goals of Drug Safety Evaluation

Drug safety evaluation is the determination of the safety of drugs for use as therapeutic agents in humans or animals through the conduct of laboratory studies in animals and in vitro systems.

Drug safety evaluation studies are essential to

1. Differentiate between new drug entities that are unacceptably toxic and those that are not
2. Characterize the potential toxic effects of new drugs
3. Determine animal dosage levels that do not cause unwanted side effects and to estimate safe dosages to be used in clinical studies
4. Demonstrate the safety of pharmaceutical compounds or new drug formulations for marketing purposes
5. Support the marketing approval for new drugs by regulatory agencies.

Use of Drug Safety Evaluation Data

Drug safety data are used by scientists and management in the pharmaceutical industry to aid in making decisions about new drugs that can be chosen for further development. Drug safety data are necessary for approval of the Notice of Claimed Investigational Exemption for a New Drug (IND), which is required before clinical trials can begin. Clinicians conducting clinical trials refer to drug safety studies in animals for information on the side effects to be expected from the new drug and for guidance in selecting safe dosages to be used in the clinical trials. Extensive toxicology studies in animals are required for the New Drug Application (NDA) for marketing approval. Even after the marketing launch of the new drug, toxicology studies may be conducted to investigate new side effects reported after widespread clinical use. If new derivatives or new formulations of a marketed drug are discovered, toxicology studies may be conducted to determine whether the new derivative or formulation offers improved safety over the original product.

HISTORICAL PERSPECTIVE

Origins of Drug Safety Evaluation

Drug safety evaluation, as we know it today, is a relatively recent development. It probably began in the late 1930s in the United States as a result of several catastrophic events that illustrated to the public and to the Congress the need for safety testing on pharmaceutical products. These events stimulated legislation leading to regulations requiring animal safety tests before marketing of a new drug would be allowed.

Significant Events Stimulating Development of Drug Safety Evaluation

There have been several famous examples of marketed pharmaceutical products that poisoned hundreds or thousands of people (1). These examples focused the attention of the public, Congress, and the medical community on the need for safety testing of new pharmaceuticals before marketing.

During prohibition, alcoholic extracts of jamaica ginger became popular as an illicit beverage due to their 60 to 80% alcohol content. These extracts usually contained other ingredients, often including castor oil. One manufacturer in Brooklyn, New York, tried to avoid the rising price of castor oil by using triorthocresyl phosphate (TOCP) instead. TOCP is now well known to cause a delayed axonal neuropathy that results in paralysis or impaired use of the legs (known as "jamaica ginger paralysis," "ginger jake paralysis," or "jake leg paralysis"). Approximately 50,000 Americans were affected from 1930 to 1931.

In 1937, an oral formulation of sulfanilamide (sulfanilamide elixir), containing about 72% diethylene glycol, was marketed. Until that time, a suitable vehicle for an oral formulation of sulfanilamide was not available. The manufacturer chose diethylene glycol because the drug exhibited good solubility in that vehicle. Diethylene glycol is now known to cause kidney and liver damage when ingested in relatively large amounts, as occurred with sulfanilamide elixir.

About 350 people were poisoned by sulfanilamide elixir, and 105 died.

During the 1930s, dinitrophenol was touted as a weight-reduction drug. Dinitrophenol acts to uncouple oxidative phosphorylation from electron transport in the mitochondria, thereby stimulating the uncontrolled consumption of metabolic energy. This results in consumption of metabolic energy reserves, for example, fat. It also results in an elevation of body temperature. At least one person died from severe elevation of body temperature after taking an overdose of dinitrophenol. Dinitrophenol also causes cataracts; several hundred people developed cataracts as a result of taking dinitrophenol for weight reduction.

Thalidomide is the most publicized example of a disaster resulting from inadequate testing of a pharmaceutical prior to marketing. Thalidomide was never marketed in the United States but was very popular in other parts of the world, where it was taken as a sleeping aid and for colds, coughs, flu, nervousness, headaches, and asthma. It was used during pregnancy to control nausea. It was eventually reported that thalidomide causes a polyneuritis resulting in sensory and motor disturbances of the hand and thumb. However, it is best known for the induction of birth defects (e.g., phocomelia and amelia). During the late 1950s and early 1960s, over 10,000 babies were born deformed as a result of their mother's use of thalidomide during pregnancy.

Regulatory Events Influencing Development of Drug Safety Evaluation

In the United States, the regulatory environment allowing the development of drug safety evaluation as a distinct discipline began with the Food and Drug Act of 1906. This act provided governmental control over food contaminants, but provided no authority to ensure the safety of new drugs.

About the time of the sulfanilamide elixir episode, a senate bill, which was to become the new Food and Drug Act, was tied up in the House of Representatives. The publicity over the sulfanilamide elixir episode spurred Congress to pass the bill, which was signed into law in June 1938. This act created the Food and Drug Administration (FDA) and required that manufacturers of new drugs prove the product's safety to the FDA prior to marketing.

The thalidomide tragedy and the publicity it generated led to the passage in 1962 of the Kefauver–Harris Drug Amendments establishing the Investigation of New Drugs–New Drug Application (IND/NDA) system in the

United States. These amendments require that before clinical studies are initiated with a new drug, the FDA has to be informed of the intent to conduct clinical studies, the nature of the clinical studies, the qualifications of the clinical investigators, and the preclinical safety studies in animals demonstrating the safety of the new drug. This information is supplied to the FDA in the form of an IND application. The 1962 amendments also require proof of efficacy before a new drug is approved for marketing. This proof is obtained by conducting clinical trials and is submitted to the FDA as part of the NDA for marketing approval (1, 2).

Influence of the International Marketplace on Drug Safety Evaluation

The international pharmaceutical market has a tremendous financial impact on all the major pharmaceutical houses. Therefore, drug safety animal studies must be conducted to satisfy the requirements of regulatory agencies around the world. In practical terms, this means satisfying the requirements of the United States, Japan, the European Economic Community (EEC), and Canada because all the other countries in the world have less stringent requirements. In certain respects, the requirements of Japan and EEC are more stringent than those of the United States. Recently, the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) has developed guidances recommending international standards for nonclinical safety studies. These ICH guidelines serve to reduce differences among regional requirements and to reduce duplication of efforts needed to register a compound internationally. However, most countries not listed above approve new drugs for marketing if those drugs have previously been approved in the United States.

The development of regulations governing drug safety studies around the world has generally followed the development of regulations in the U.S. Such regulations were promulgated in the EEC in 1980, in Canada in 1981, and in Japan in 1984.

GENERAL PRINCIPLES

There are a few general principles followed in conducting drug safety animal studies in the development of new drugs. These principles and the reasoning behind them are sometimes misunderstood by those not involved in drug

safety evaluation, leading to unwarranted criticism of toxicology studies.

First, animal toxicology studies almost always use dosage levels higher than the intended clinical dosage. One of the goals of drug safety toxicology studies is to characterize the toxic side effects of a new drug. To characterize the toxicity of a new drug, high-enough dosages must be used to allow full expression of the toxic effects of the drug. The first goal stated in the introduction was to differentiate between drugs that are acceptably safe and those that are unacceptably toxic. One criterion by which this differentiation is made is the safety margin between the expected clinical dosage, or the pharmacologically active dosage, and the dosage that causes toxicity in animals. To determine this safety margin, dosages higher than the clinical dosage must be used. In addition, dosages in laboratory animals required to elicit pharmacological and toxicological responses are usually higher than those in humans, another reason why high dosages are used in animal toxicology studies.

Margin of safety is not the only criterion used to determine if a drug is safe enough to use in humans. The type of toxicity caused by the drug and the intended clinical indication are also considered. For example, a drug that might cause cancer would not be used for symptomatic relief of a head cold. However, such a drug might be used for the treatment of cancer as the disease to be treated is itself life-threatening. Many anticancer agents fall into this category. In other words, the benefit-to-risk ratio would be too low for the head cold but high enough for the treatment of cancer. This is an extreme comparison. Another example would be the teratogenic anticonvulsants in use today. These drugs can cause malformations in fetuses. Having a seizure during pregnancy also carries serious risk to the fetus and to the mother, however. The benefits of using an anticonvulsant during pregnancy are often considered to outweigh the risk of having a malformed child.

Reversibility of toxic effects is an important consideration in evaluating the safety of a drug. Cancer and fetal malformation are irreversible changes. If, instead of cancer a drug caused reversible enlargement of the liver, such a drug might be acceptable for less life-threatening diseases. In the case of the anticonvulsant used during pregnancy, a reversible retardation of neonatal growth would obviously be a more acceptable side effect than a malformation.

During the development of new drugs, reversibility of toxic effects is often studied to provide information for use in making benefit-to-risk decisions. The information is also helpful to the clinicians conducting clinical trials or using the drug if it eventually reaches the market. For the

clinician it is helpful to know that if a particular side effect appears in study subjects or patients, it will disappear upon cessation of treatment. The margin of safety derived from animal toxicity studies is not always infallible. During the clinical development program, it is often necessary to increase the dosage beyond that used in initial clinical studies. In this situation, where the margin of safety is decreased because the dosage has been increased, side effects are more likely to be encountered. If those side effects have been shown to be reversible in animal studies, the clinical investigators may feel more comfortable about raising the clinical dosage in order to achieve the desired pharmacological effect.

The conduct of animal toxicity studies proceeds concurrently with and in advance of clinical studies. A certain amount of animal toxicity data is required for the IND. This often includes acute toxicity in two rodent species, mutagenicity screening studies, and one-month toxicity studies in a rodent and nonrodent species. After the clinical studies have begun, further animal studies are conducted, usually until the time for the NDA. These would include further mutagenicity studies, reproductive toxicity studies, and long-term toxicity studies, possibly including carcinogenicity studies.

The duration of treatment in animal studies is always at least as long as the duration of treatment in clinical studies and, at the later stages of development, longer than in the intended therapeutic indication. This provides another element of safety. With long-term treatment, toxicity usually occurs at lower dosages than with short-term treatment. Some drugs may be found to accumulate in the body or in certain organs and thereby give rise to toxic changes at lower dosages than with short-term treatment. Occasionally with long-term treatment, different side effects may manifest themselves. If long-term toxicity studies are conducted in advance of long-term clinical trials, any changes in safety margin or toxicity profile will be known before the conduct of those clinical trials. The duration of treatment required by regulatory agencies for marketing approval is usually longer than the intended clinical regimen.

Another principle that is generally followed in drug safety evaluation is the use of two or more species to study the toxicology of a new drug. It is usually not possible to know which species will be the most accurate predictor of the toxicological response in humans. However, if a particular side effect is found in two or more species, we can predict that it will also occur in humans at high-enough dosages. In the case where a side effect is found at a lower dosage in one species than in another, the more sensitive species is generally used to estimate the safe dosage level for humans.

TYPES OF DRUG SAFETY EVALUATION STUDIES

Acute Toxicity

Acute toxicity studies are often misunderstood and criticized for the use of large numbers of animals to generate relatively valueless data, such as LD50 (a dosage that kills 50% of the test animals). It is worth pointing out that there are legitimate uses for such data. For example, in the early stages of new drug discovery, when no toxicity data of any kind are available, it is often very helpful to determine an LD50 for comparison with the ED50 (a dose that produces a pharmacological effect in 50% of the test subjects). This comparison is often used to calculate a therapeutic index (LD50/ED50): the higher the therapeutic index, the greater the implied margin of safety between the pharmacologically active and toxic dosages. These numbers can be generated quickly and inexpensively, helping to provide an early indication of the acceptability of a new drug for further development. Another way in which LD50 data are useful is in the quality testing of new batches of a drug. In mice, LD50 data are often routinely obtained for each new batch of bulk drug for comparison with the LD50 obtained on the original batch(es) of the same drug. This quick and inexpensive test gives a final assurance that the drug is not substantially different from the original drug, or has not been contaminated prior to preparation of the final formulation. To date, there are no good substitutes for these *in vivo* tests. Although they do consume large numbers of animals, the tests are quick, inexpensive, and reliable.

In addition to the applications of acute toxicity studies mentioned previously, such studies are necessary as the initial toxicity studies conducted in the development of a new drug. The information provided is used to select dosages for subacute and subchronic toxicity studies in which animals are treated daily for various periods of time. Studying the acute toxicity of drug combinations is a convenient method for detecting potential drug interactions that may occur clinically. If such interactions are detected, they may be studied further in subacute and subchronic toxicity studies. In these studies, a single dose of test compound is given to animals, which are then observed for a period of 1 or 2 weeks afterward for the development of toxic signs. Surviving animals are usually necropsied at the end of the observation period to determine if grossly observable organ changes have occurred. When organ changes are observed, these are usually examined microscopically to determine the nature of the change. If an LD50 is determined, it is usually calculated using probit analysis (3). Less often, the

up-and-down (4) or moving averages (5) methods, which require fewer animals, will be used. An LD50 value is not necessary for the safety evaluation of a drug or for choosing dosages for subacute toxicity studies. For these purposes, it is only necessary to determine the lowest lethal dosage of the compound and to characterize the toxic manifestations of an acute overdose. This information is useful to clinicians using the drug and to poison-control centers where information regarding the effects of acute overdose must be made available. To select dosages for repeated-dose, subacute toxicity studies, information about the lowest toxic dose is needed.

Often, an LD50 cannot be determined if the compound has low toxicity. Sometimes lethality or toxicity cannot be elicited. In these cases a "limit" dose concept is used. A limit dose, for example, 5 g/kg for rodents or 2 g/kg for dogs, is administered. If no deaths or signs of toxicity occur, the lethal or toxic dose is simply stated to be greater than the respective limit dose. Sometimes the limit dose may be determined by the volume that can be administered in a single bolus. In rodents, for example, 50 mL/kg is the maximum that can be given as a bolus orally or intravenously (IV). For IV acute toxicity, much higher dose volumes can be given if the drug solution is administered as a slow infusion.

The animals used in acute toxicity studies are usually mice, rats, and dogs. The batch tests mentioned above are usually conducted in either male or female mice. Females are generally more sensitive than are males, but either sex is appropriate. Because the objective is simply to compare the toxicity of a new batch of a given drug to that of previous batches, the absolute toxicity is less important than the toxicity relative to previous batches of the same drug. For the safety evaluation of a new drug, acute toxicity studies are usually conducted in both sexes of mice and rats by several routes of administration, including the intended clinical route(s). This is to generate initial information regarding differences in degree or type of toxicity resulting from different routes of administration [e.g., oral vs. IV vs. subcutaneous (SC) vs. intraperitoneal (IP)]. An initial indication of the degree of absorption can be obtained by comparing IV and oral toxicity. Sex differences in acute toxicity can alert the toxicologist to anticipate such differences in subacute/ subchronic studies. If sex differences are marked, this may be taken into consideration in deciding whether to pursue development of the compound. Because subacute and subchronic studies are usually conducted in rats and dogs, it is usually necessary to generate acute toxicity data in these species to aid in dosage selection for those studies.

There is increasing awareness in the industry regarding the need to reduce the numbers of animals used in acute

toxicity studies. This is because of social concerns about the unnecessary use of animals in laboratory studies and because of the financial impact of purchasing and housing large numbers of animals. The use of limit tests as outlined above and of alternative methods for calculating the LD50 offers practical means for reducing the numbers of animals used for acute toxicity studies.

Ocular and Dermal Irritation

If a drug is intended for use by ocular or dermal administration, one of the first requirements is to determine whether it could cause irritation upon contact with those tissues. Drugs or drug formulations that may inadvertently come into contact with the eye should also be tested for ocular irritation. For instance, a topical antibiotic that might be applied to the face could accidentally get into the eyes. Inhalation anesthetics also have the potential for ocular exposure.^a

Muscle and Vein Irritation

Muscle and vein irritation studies are conducted if a drug formulation is to be administered by these routes. These studies may be conducted as a guide during the development of formulations for these routes. They should be conducted prior to any subacute studies so that experimental animals are not subjected to avoidable pain and suffering.^b

Pain on Injection

It seems obvious that drug formulations to be given by injection should be tested for the amount of pain produced upon injection before being tested in humans. In practice, however, this kind of testing is rarely done unless a formulation is first tested in humans and found to be painful upon injection. The muscle and vein irritation studies mentioned previously sometimes give indications of pain if the animals show signs of discomfort during the injection. However, this is usually not a reliable indicator of the painfulness of the formulation. Muscle and vein irritation studies give information on the potential for drug formulations to cause tissue injury but not for pain on injection.

Models that do assess pain on injection are the rat paw lick and mouse scratch models. In the paw lick

model, the drug formulation is injected into the pad of the hindfoot, and the number of times the rat licks the paw is counted (6). The mouse scratch test is similar except that the formulation is injected subcutaneously on the back, and the number of times the mouse scratches the injection site is counted.

Results from the rat paw lick model seem to correlate well with clinical painfulness experience and the muscle irritation test with certain antibiotic formulations (6, 7). However, because rat paw lick results from a wide variety of drug formulations have not been published, these correlations may hold only for a certain subset of painful formulations. A formulation could cause pain but yet not cause muscle damage upon intramuscular injection. Additionally, some formulations are painful upon intravenous injection.

Subacute, Subchronic, and Chronic Toxicity

These studies are designed to characterize the toxic effects of drugs upon repeated daily administration for periods of time ranging from 2 weeks to 1 year and to determine no-toxic-effect dosage levels for short to long-term repeated dosing.

The studies are usually conducted in rats and dogs, unless one of these species is found to be inappropriate. This may be because of a sensitivity or insensitivity to the pharmacological effects of the drug which is not expected to be characteristic of the response in humans. The rat or dog may metabolize the drug substantially differently from humans. These animals may absorb the drug poorly relative to other species, or the pharmacokinetics may be substantially different. Such differences in pharmacological responsiveness, pharmacokinetics, or metabolism are almost never known when the initial toxicology studies are conducted. An exception would be the case of certain monoclonal antibody products with binding sites specific to primate tissues. In this case, primates would be selected as the only species in which to conduct safety evaluation studies. Genetically engineered human protein products where the antigenicity to primates is expected to be less than that to other species less closely related to humans, or where the pharmacological responsiveness in primates is similar to that expected in humans, might also be tested only in nonhuman primates. In most cases, the toxicology program is conducted in rats and dogs until one of these species is found to be inappropriate. The information that might lead to this conclusion would not be available until the Phase I clinical trials are conducted in healthy volunteers.

A certain amount of toxicology data is necessary, usually including 1-month studies in rodents and

^a For typical ocular and dermal irritation protocols, see *Animals in Drug Development*, page 000; 00000.

^b Typical protocols for these studies are described in the article, *Animals in Drug Development*, in this encyclopedia.

nonrodents, prior to the initiation of Phase I clinical trials. The studies are usually conducted in a sequential fashion, using the results from the short-term studies to design subsequent studies of longer duration. Thus, acute toxicity results are used to design 2-week studies, the results of which are used to design 1-month studies, which lead to 3-month studies, and then to studies of 6 month or 1 year. A typical study design would include a period of 2 weeks or more for quarantine and acclimation of the animals to the laboratory environment. During this time, measurements can be made (e.g., body weight, food and water consumption, urine output, serum chemistries, hematology, urinalysis, electrocardiograms (ECG), and general physical and ophthalmological examinations) to determine that the animals are healthy before beginning the treatment phase of the study. Excess animals are ordered and allowed to acclimate so that after unsuitable animals are rejected, a sufficient number of healthy animals remain to begin the study.

During the quarantine period, it is customary to have the animals examined by a veterinarian for signs of disease. In some cases, particularly in preparation for a long-term study where the financial risk is great, some of the animals from the shipment may be sacrificed, blood collected for serology to determine if viral antibodies are present, and the tissues examined for evidence of infectious processes.

During the conduct of the study, all the animals are carefully monitored for changes in physical appearance, behavior, body weight, and food consumption. These variables are often the most sensitive indicators of toxicity. During the study, the other variables mentioned above are periodically measured. The variables to be measured, the intervals between measurements, and the number of such measurements during the study depend on the drug under development, the length of the study, and on the species being used. For example, when dogs are used, blood samples for serum chemistry, hematology, or drug concentration determinations can be taken more frequently because of the larger blood volume relative to rodents and the ease of obtaining blood samples. However, if interim blood samples are required in rodents, extra animals are commonly included in each group to be used for this purpose. Because of the small blood volume in a rodent and the trauma inflicted when blood is collected, there is the possibility that the process of repeated blood collection could alter the response of the test animal to the toxic insult from the drug. This possibility will vary with the frequency of blood collections, the method of blood collection (e.g., tail bleeding, retro-orbital sinus bleeding, or cardiac puncture), and the nature of the drug under study. Many investigators prefer to avoid the possibility of

such confounding effects by assigning extra animals to provide blood samples. In a 1-month study, blood samples might be taken only during acclimation and at the end of the dosing period. However, if the time course of certain expected changes is of interest, then additional blood and/or urine samples may be taken after 2 weeks of treatment. In a 3-month study, samples might be taken at the end of every month or once every 6 weeks. In a 6-month study, samples might be taken every 3 months; in a 1-year study, once every 6 months. If a recovery period is included in the study design, then blood and/or urine samples would be taken at the end of the recovery period. If the recovery period is relatively long in duration, or if the time course of recovery of some function is important, samples may be taken at intervals during the recovery period. A recovery period is a period of time in which some of the animals in a study are held after the cessation of treatment to study the reversibility of any toxic changes that manifested during the treatment period. The length of recovery depends upon the nature of the toxic change under study, the duration of the treatment phase, and, to some extent, the species. For example, if the side effect caused by the drug results from an exaggerated pharmacological effect, for example, central nervous system (CNS) stimulation, cardiac arrhythmia, or hypertension, which might reverse fairly rapidly after the cessation of treatment, then a short recovery period would be appropriate. On the other hand, a change that might take longer to reverse, for example, testicular atrophy or hepatomegaly, would require a longer recovery period. In the absence of such qualifying circumstances (i.e., when the nature of the toxic changes to be expected is not known), typical recovery periods would be 2 weeks after 1 month of treatment, 1 month after 3 months or 6 months of treatment, and 3 months after 1 year of treatment.

Electrocardiograms are commonly recorded in dog studies but only rarely in rodents, although there is no good reason not to record ECGs in rodents. With the dorsal-axial-inferior (DAI) lead system (8), it is actually easier and faster to record ECGs in rats than in dogs. Furthermore, the rat heart is known to respond to pharmacological agents (except the cardiac glycosides) affecting the ECG of other species (8). Because of the more rapid heart rate of rat, the ECG is slightly more difficult to evaluate. However, the main reasons why ECGs are not recorded more commonly in rats in toxicology studies are custom and historical precedent.

It is also fairly easy to perform specialized measurements such as electroretinography in dogs without the anesthesia or surgical preparation required in other species (9, 10). Electroretinography has assumed a status of increased importance in safety evaluation because many

drugs affect retinal function. Examples are the quinolone antibacterials and Viagra (11).

At the end of the study, all the animals are killed under anesthesia, the internal organs are examined for visible changes indicative of pathological responses, the organs are weighed, and samples of all tissues are taken for histological examination. In some studies, particularly long-term studies, animals from each group may be killed at intervals during the study to collect tissues for histological examination. For example, in a 6-month study, some animals may be killed after 3 months of treatment; in a 1-year study, some animals may be killed after 6 months of treatment. Animals would also be killed at the end of any recovery period, or at intervals during the recovery period, for assessment of drug-related tissue changes. In studies with large animals (e.g., dogs or monkeys), protocols are sometimes written so that all the tissues from all the animals are examined histologically. However, in rodent toxicity studies, because of the large number of animals used, protocols are commonly written so that tissues from animals in the control and highest dosage groups killed at the end of the treatment period are examined first. If any changes are found that appear to be drug related, tissues from animals in the next lower dosage group are examined, and so on. Tissues from recovery group animals are usually examined only in groups where drug-related changes are found at the end of treatment.

The determination of the no-toxic-effect dosage level in a toxicity study is not always easy; often the only findings at a particular dosage level are those resulting from the pharmacological action of the drug, for example, hypotension and the resultant behavioral signs with a peripheral vasodilator or sedation with a tranquilizer. Are these changes to be considered due to toxicity? Certainly, they would be unwanted side effects of the drug, but on the other hand, they would not be unexpected based on the pharmacology of the drug. Other examples would be hepatocellular hypertrophy with a drug that also causes induction of liver microsomal enzymes, or renal enlargement at high dosages. These changes might be considered adaptive physiological changes. Should they be considered to be due to toxicity? Some investigators look for a slightly different end point in determining the results of a study—the no-observed-effect dosage level. In their view, all the examples just cited would be lumped along with other findings more clearly labeled toxicity into the category of drug-induced observed effects. They look for a dosage level below which no drug-induced effects are observed.

Dosage selection for subacute, subchronic, and chronic studies is always a sensitive matter, requiring experience and judgment. Three dosage levels plus one or more

vehicle or untreated control groups are usually used. The highest dosage is selected to elicit some degree of toxicity. It should not be so toxic that animals die or are debilitated before the end of the study. The purpose is to characterize the toxicity of the drug on repeated administration. If the dosage is too high, animals may be lost before the end of the treatment period. This reduces the number of animals available for terminal serum chemistry, hematology, histology, and so on, so that the toxicity cannot be characterized adequately. On the other hand, if the dosage is not high enough to cause toxicity, the study may be rejected as inadequate by regulatory agencies. The low dosage is usually selected to provide some small multiple of the expected clinical dosage or systemic plasma exposure level. In cases where the drug is fairly toxic, the low dosage may be selected to match the clinical dosage or plasma exposure level. However, it is usually considered necessary to demonstrate a no-toxic-effect dosage of at least 10 or 6 times the clinical dosage in the case of rats or dogs, respectively. Therefore, the low dosage is usually at least this multiple of the clinical dosage. The mid-dosage is selected to be at equal log intervals between the high and low dosages. This is because the response of biological systems to the vast majority of stimuli are logarithmic. Therefore, to best demonstrate a linear dose-response relationship, the stimulus should be administered on a logarithmic scale. In selecting the dosages for a study, it must be remembered that if the high dosage is toxic as expected, the mid-dosage is likely to be the no-toxic-effect dosage. Therefore, the relationship of the mid-dosage to the clinical dosage should also be considered when selecting the dosage scheme for a study. Often, it is desirable to compare the toxicity in long-term studies with the same dosages in short-term studies. If this is the case, some of the dosage levels in long-term studies may be the same as those in short-term studies. Although three dosage levels are usually considered the minimum, sometimes more dosage levels are used to meet all the considerations discussed above.

The number of animals per group depends on the length of the study, the species, whether a recovery period will be included, and whether extra animals will be included for blood sampling or other special procedures. Generally, for rodent studies to be submitted to a regulatory agency, 10 per sex is the minimum for studies up to 3 months, and 20 per sex is the minimum for longer studies up to 1 year. For dogs, at least 4 per sex is usually required in studies submitted for registration. More animals are preferred for long-term studies because the longer the study duration, the greater the chance of losing animals due to accidents or disease. The study design should compensate for the greater statistical chance of losing animals so that an

adequate number of animals are available at the end of the study for evaluation.

The route of administration is an important consideration in the design of a study. Generally, the route of administration in drug safety evaluation studies is the same route intended clinically. The most common route is oral. Drugs can be administered easily by gavage in rats, dogs, and monkeys and by capsule in dogs. Methods are available for administering capsules to monkeys and rats. In these cases, the capsule is usually affixed to the end of a special gavage needle for rats or gavage tube for monkeys. The tube is inserted into the stomach and the capsule expelled from the end of the tube. In long term (i.e., 1 year) rodent studies, it is common to administer drug mixed in the diet. This assumes that the drug is stable and bioavailable when mixed in the diet and that the drug can be mixed in a homogeneous manner in the diet. All these requirements must be demonstrated before committing to a dietary administration study. When administering drugs via the diet, it is necessary to adjust the drug concentrations in the diet to compensate for variations in food consumption by the animals and for changes in body weight. During the early stages, for example, the first 3 months, of the study when the animals grow most rapidly and when drug effects on food consumption and body weight gain are likely to be most pronounced, this should be done on a weekly basis. Later in the study, say, after 3 months, when the growth curve of rodents approaches a shallow slope, this can be done once per month. This is done by adjusting the concentration of drug in diet based on the food consumption and body weight gain from the previous period. The concentration for each sex group is adjusted separately. If an intravenous study is needed, this could limit the feasible duration of the study and/or the number of animals or groups that can be accommodated. If the drug formulation does not cause vein irritation, it is feasible to conduct 3-month studies in rats, using the lateral tail veins for injection. In dogs, almost any length of study should be possible; however, 3- or 6-month studies seem to be within the usual limits of practicability. If the drug must be given by slow injection rather than by bolus or slow push, there will be limitations on the numbers of groups and animals that can be used due to the equipment and increased technical staff needed to conduct such studies.

Mutagenicity

Mutagenicity studies are required for the marketing approval of new drugs. The role of mutagenicity studies in drug safety evaluation is sometimes misunderstood. The purpose of these studies is not to replace carcinogenicity

bioassays but to predict which drugs are likely to be positive in carcinogenicity bioassays and cause cancer in humans and to detect compounds that might cause hereditary germ-line mutations in humans. Such predictions require an in-depth knowledge of the mutagenicity assays available and scientific judgment. Sound predictions cannot be made strictly on the basis of numbers of positive or negative results in a few tests.

It is currently accepted knowledge that many, but not all, known carcinogens have the ability to cause damage to deoxyribonucleic acid (DNA). Such agents also have the potential to cause heritable germ line mutations. The mutagenicity studies currently available were designed to detect DNA-damaging agents. Positive results in these tests therefore indicate the possibility of carcinogenicity and germ cell mutagenicity resulting from the test compound. However, a negative result in mutagenicity tests designed to detect DNA-damaging agents does not imply that the compound will not be a carcinogen. We now realize that some carcinogens are nongenotoxic. Examples are the rodent liver carcinogens, phenobarbital and chlorophenothane (DDT), which are not mutagenic. These compounds are thought to cause liver tumors through their action as promoters. Bile acids are known to act as promoters in the induction of intestinal tumors. In experimental situations, certain compounds are known to act as cocarcinogens (enhancing the carcinogenic response only when applied concurrently with a DNA-damaging agent) or as promoters (enhancing the carcinogenic response when applied after the DNA-damaging agent). These types of carcinogens are not detected by mutagenicity assays. Currently, there are no short-term assays available for routine testing which reliably detect cocarcinogens or promoters, that is, nongenotoxic carcinogens.

Much publicity has been given to the absence of a high level of concordance among the results of a selected battery of four *in vitro* mutagenicity assays with the results of carcinogenicity bioassays in rats and mice (12). First, the number of compounds available for comparison is hardly an adequate sample size from which to make any general conclusions regarding the predictive capabilities of any series of assays. Second, the compounds selected for comparison of bioassays to genetic toxicity assays were heavily weighted in favor of genotoxic compounds. This makes the estimation of specificity for the mutagenicity assays (the ability to produce a negative result with a true noncarcinogen) difficult. Both sensitivity (the ability to produce a positive result with a true carcinogen) and specificity are used to determine concordance (the ability to produce a correct response with either a true carcinogen or a true noncarcinogen).

Furthermore, the compounds selected surely included compounds that were carcinogenic in the rodent bioassays but are not genotoxic. Finally, the four mutagenicity assays selected represent an artificial restriction on the comparison of mutagenicity assays to bioassay results. In practice, the genetic toxicologist is not restricted to using only a few specific assays.

As stated at the beginning of this section, certain mutagenicity tests are required by regulatory agencies for approval of new drugs. The ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) (13) requires a battery of three categories of tests: (1) bacterial mutagenicity, (2) in vitro cytogenetics or in vitro mammalian cell mutagenicity assay with ability to detect chromosomal damage (i.e., mouse lymphoma tk [thymidine kinase locus] assay), and (3) in vivo test for chromosomal damage using rodent hematopoietic cells. For initial screening of new compounds, I recommend bacterial mutagenicity and in vitro cytogenetics. This is because the bacterial mutagenicity test, usually the Ames test using *Salmonella typhimurium*, is rapid, inexpensive, and sensitive. In vitro cytogenetics should be included because the bacterial test cannot detect agents that act at the level of the eukaryotic chromosome. Therefore, using these two tests, one can determine relatively quickly whether a new drug is a potent mutagen or not. If positive results are obtained in in vitro studies, other tests can be conducted to try to determine the mechanism responsible for the positive results, and to determine the potency and any specificity of the mutagenic response. If the drug is positive for mutagenicity in some tests but has relatively low potency, the drug might still be developed. In this case, in vivo mutagenicity tests may reveal a lack of detectable response. This would aid in defending the benefit/risk ratio for the drug. Some drugs may be positive in in vitro mutagenicity tests only in the presence of activating microsomal preparations. If this is the case, analysis of the metabolic products of the in vitro reaction may reveal that those products are not the same as the metabolites produced in vivo. Thus, the relevance of the in vitro mutagenicity study to in vivo risk assessment can be determined. The point is that one should not just accept in vitro study results at face value. Further study of positive results is usually warranted. Also, some of the mammalian cell in vitro mutagenicity assays are known to be sensitive to alterations in the osmolarity and/or the pH of the culture medium. Examples are the in vitro cytogenetics in Chinese hamster ovary cells and the L5178Y mouse lymphoma assays. The osmolarity in the assays should not be greater than 600 milliosmolar (14) and the pH should be in the range of 6.8–8.0 (15).

Several transgenic animal models are available for mutagenicity testing *in vivo*. The most widely used are the Big BlueTM and MutaTM Mouse models. These animals have portions of the *E. coli* galactosidase operon inserted into their genome. This bacterial DNA is used as the target for mutagenicity testing when the animals are treated with test chemicals. After chemical treatment, the bacterial DNA is removed from the tissues and packaged into bacterial viral (phage) particles. The phages are used to infect bacteria growing on agar plates. The bacteria multiply to form colonies that can be counted. Because of a color reaction that can be catalyzed by the galactosidase protein, a color reaction formed in the bacterial colonies can be used to assess the presence or absence of mutations formed in the target DNA from virtually any tissue within the treated mice.

Guidelines and protocols for specific mutagenicity assays can be found in Naismith, 1987; Brusick, 1980; Kilby et al., 1984; Mirsalis et al., 1994; Provost et al., 1993; and Heddle et al., 2000, listed in the Bibliography.

Reproductive Toxicity

Reproductive toxicology covers the entire process of reproduction from mating to pregnancy, birth, weaning, and, sometimes, the reproductive function of subsequent generations. The species most commonly used in these studies are rats and rabbits, although in special circumstances, other species such as mice, dogs, or monkeys are used. A series of studies is usually conducted so as to cover all the phases reproduction. Within each study, groups of animals corresponding to untreated controls and 2 to 3 drug-treated groups are used.

To assess potential effects on all phases of reproduction, separate studies are usually conducted that focus on distinct aspects of reproduction. The separate studies are usually designed to provide some overlap in the phases tested so as to avoid in gaps in the testing program. Fertility studies (historically called Segment-I studies) are generally intended to assess effects of the drug on the processes of spermatogenesis, oogenesis, mating behavior, copulation, fertilization, early embryonic survival and development, and implantation of the zygote in the uterus. Teratology (Segment II) studies are intended to assess effects of the drug on embryonic/fetal survival and development. Peri-/post-natal (Segment III) studies assess drug effects on late pregnancy, parturition, lactation, and postnatal development. One reason for dividing the reproductive process into phases is that, generally speaking, the shorter the treatment period, the higher the dosage that can be given without causing excessive toxicity. By dividing the reproductive process into shorter

phases instead of using the long phase required to encompass the whole process, higher dosages can be given to maximize characterization of the drug's toxicity in any part of the reproductive process. If a drug effect is found, the stage of the reproductive process affected by the drug can be more readily identified. This facilitates the assessment of risk for clinical use and helps the toxicologist design further studies, if needed, to examine the mechanism of action. A practical reason is that smaller studies can be conducted, sometimes in parallel fashion. Smaller studies allow flexibility in the timing of the evaluation of various parts of the reproductive process in accordance with the development program of a particular drug.

Fertility studies

Segment-I fertility studies are usually conducted in rats, although mice may occasionally be used. Rodents are used because of the convenience of housing large numbers of them, ease of mating, and widespread experience among reproductive toxicologists with these species. In these studies, the drug is usually administered to males for 28–30 days and to females for 14 days prior to mating. The duration of the spermatogenic cycle in rodents is approximately 60 to 80 days. However, a 28- to 30-day exposure period is considered sufficient to detect changes to the male reproduction when used in conjunction with thorough histological evaluation of the male reproductive tract. The treatment duration in females is considered sufficient to allow enough time for the females to be exposed to steady-state blood levels of the drug for at least a week to 10 days before mating. The animals are then placed with individual members of the opposite sex for mating. Treated males may be mated with treated females, or treated members of one sex may be placed with untreated members of the opposite sex. The females are observed for signs of mating, copulatory plug or sperm-positive vaginal smear, to determine mating performance. Generally, mating treated males to treated females is less costly because of the smaller numbers of animals involved. However, if drug-related effects are found, it may be necessary to repeat the study by mating treated animals with untreated animals. Drug treatment of the females is continued through the mating period and through mid-gestation. Treatment of the males continues until it is possible to determine whether there were any effects on fertility. If such effects are found, it may be desirable to remate the males to see if effects on fertility are reversible, or subject the males to histological evaluation of the reproductive tract to try to determine the cause of any fertility effects. Waiting until the end of pregnancy for the intrauterine examination offers the

advantage of an additional assessment for effects on intrauterine development. The fertility study may or may not be initiated prior to the IND application; however, it is usually required for the NDA. This fertility study in male animals is conducted prior to large-scale (phase III) clinical trials in men. This study is generally conducted prior to any clinical trials involving women of child-bearing potential. However, in the U.S. and Europe, women of child-bearing potential confirmed to be non-pregnant and using effective birth control may be enrolled in phase I and phase II (usually small-scale) studies prior to completion of the female fertility study.

Teratology studies

In teratology (historically called segment-II) studies, a drug is administered to pregnant animals during the period of fetal organogenesis. The studies are usually conducted in rats and rabbits, although other species such as mice or monkeys, are also commonly used. Most regulatory guidelines require that teratology studies be conducted on two species, and most suggest rats and rabbits as the preferred species. These species are used because of the ease of housing and handling, the ease of mating or insemination (rabbits), their short gestation period, and their large litter sizes (which facilitates statistical evaluation of the data). The period of drug treatment during gestation can be confusing. Most guidelines recommend gestation Days 6 to 15 for mice and rats and Days 6 to 18 for rabbits. Japanese guidelines recommend Days 7 to 17 for rats. These guidelines assume that the first day of pregnancy (the day when a vaginal plug or sperm in the vagina is found) is counted as gestational day 0. However, the guidelines of the Organization of Economic Cooperation and Development (OECD) recommend that if animals are artificially inseminated or pregnancy is determined by forced mating (rabbits are usually impregnated by one of these methods), the days of treatment be calculated by adding 1 day to the usual number. To accommodate all these guidelines, I recommend administering drugs during gestation Days 6 to 17 for rodents and Days 6 to 19 for rabbits. The extra day or two of dosing will have no adverse effect on the study but will allow compliance with worldwide guidelines.

At the end of the gestation, the dams are killed and the uteri examined internally for the presence of live, dead, resorbed, or degenerating fetuses. At this time the number of implantation sites and the number of corpora lutea are counted, and the internal organs of the dam are examined visually to detect any pathological changes that may have affected the pregnancy. The pups are weighed and examined for any external abnormalities. A number of the pups are fixed and prepared for examination of the

internal organs, and the rest are fixed and prepared for skeletal examination. In the case of rabbits, all the pups are subjected to both internal and skeletal examination. The abnormalities found are classified as either malformations or developmental variations. Malformations are those changes that are irreversible, alter general body conformity, interfere with body function or are possibly life-threatening, and are not commonly found in untreated animals. Variations are those changes that may be reversible, have no effect on body conformity or the health of the animal, represent slight deviations from the normal, and are found occasionally in untreated animals. Some changes, for example, reduced skeletal ossification or enlarged renal pelvis (apparent hydronephrosis), may simply be indicative of delayed intrauterine growth.

Dosages are selected so that some maternal toxicity is demonstrated during the study. This is to ensure a thorough test of teratogenic potential. It also has a practical application when the drug is used clinically. If during drug treatment a pregnant woman develops a toxic response, the physician is alerted to reduce the dosage or withdraw treatment. If it is known that the drug does not cause fetal toxicity or malformations at dosages up to those causing maternal toxicity, the physician can be assured that the fetus remains unharmed. If, on the other hand, a drug caused toxicity or developmental abnormalities in fetuses at dosages below those causing maternal toxicity, the physician no longer has the warning sign of maternal toxicity. With such a drug, the fetus may be harmed in the absence of maternal toxicity. Obviously, this is a more dangerous situation. Thus, the concept of toxicity to the adult mother versus the developing conceptus (or A/D ratio) assumes practical importance. Often, the results of reproductive toxicology studies are reported in terms of no-toxic-effect dosage for the dam and no-toxic-effect dosage for the fetus or offspring.

The special case of teratology studies in rabbits is worth mentioning. Rabbits are very susceptible to changes in the intestinal flora caused by some antibiotics. This usually leads to a marked reduction in food consumption, which leads secondarily to fetal toxicity, abortions, and fetal malformations. In such cases, the rabbit exhibits a species-specific sensitivity to the drug and may be an inappropriate species for a teratology study. However, the standard protocol may be modified so that the dams are treated with a drug for only a few days during the gestation period (e.g., Days 6 to 13 and Days 14 to 19, or smaller intervals). In this manner, excessive toxicity to the dam can be avoided while sufficiently high dosages of the drug can be given to provide an adequate test for teratogenicity.

The teratology study is often the first study of reproductive toxicity because it is less time- and resource consuming than are studies involving other phases of reproduction and because it assesses an end point, teratogenicity, which can have a major impact on further development of a drug.

Perinatal and postnatal toxicity

The perinatal and postnatal study (sometimes called a Segment III study) is usually conducted in rats. In this study, the drug is administered to pregnant rats from implantation (gestation day 6) through parturition and lactation. If desired, some of the fetuses can be taken at the end of gestation for morphological exams to satisfy the requirements for testing effects on organogenesis (teratology or Segment II phase). Effects of the drug on fetal development, parturition, nursing, growth, development, learning/memory, emotionality, and often, reproductive function of the pups are assessed. Postnatal developmental indices such as body weight gain, time of eye opening, pinna detachment, incisor eruption, vaginal opening, and development of righting and startle reflexes are evaluated. Some of the pups are retained after the lactation period for determination of behavioral variables such as learning, memory, and emotionality. Tests such as active or passive avoidance, shuttle box, and water T or M maze assess learning and memory. Test such as activity measurements or open-field behavior give some assessment of emotionality. In addition some of the pups may be retained for mating to determine the fertility and reproductive performance of the F1 generation. In this case, the study is usually terminated at the end of the F1 generation's pregnancy, but may be extended to assess early postnatal survival of the F2 generation. Again, the rationale is to conduct a stringent test for effects of the drug on development of the offspring.

Carcinogenicity Bioassays

Carcinogenicity bioassays are long-term studies conducted in rodents to assess the tumor-producing potential of a compound. These studies are conducted according to guidelines established originally by the U.S. National Cancer Institute. However, other guidelines, for example, OECD and FDA, are also useful in developing protocols. It is required that bioassays be conducted on both rats and mice over a period of time representing a majority of the natural life span for the species. Depending on the strain used, this is usually 2 years for rats and 18 months to 2 years for mice. The animals must be no older than 6 weeks at the initiation of treatment. The strains used are

most commonly Fisher 344 or Sprague Dawley rats and B6C3F1 or CD-1 mice. The animals are dosed daily, usually with orally administered drugs. If the drug is stable in the rodent's diet and can be mixed with it homogeneously, it is commonly administered via the diet because of the reduced staff required. Again, in the case of drugs, the concentrations of test compound in the diet are usually adjusted at regular intervals in order to maintain the target dosage level. The animals are examined frequently for the presence of palpable masses. Blood samples may be taken at 6-month or 1-year intervals for hematological evaluation, with particular attention paid to the development of leukemias. At the end of the study, surviving animals are killed and their tissues subjected to a thorough evaluation for the presence of neoplasia. The study design usually includes three drug-treated groups plus one or more control groups. The number of animals is usually 50 per sex for the drug-treated groups and 100 per sex for the control groups. However, extra animals (e.g., five per sex-group) are often started on treatment for the first 6 to 8 weeks of the study so that if animals die during the early phases of the study, which commonly occurs, there will still be enough animals on study to maintain the intended group sizes. The dosages are selected on the basis of the results of 3-month maximum tolerated dosage (MTD) studies. The object of the MTD studies is to determine the MTD for each sex of each mouse and rat strain to be used in the bioassay. The MTD is defined as a dosage that causes some degree of toxicity but not a decrement in body weight gain greater than 10% compared with that of controls; also, that does not cause toxicity, other than that related to a neoplastic response, which would be expected to shorten the animal's natural life span.

Recently, transgenic strains of mice have become available, which allow detection of a carcinogenic response within 6 months of treatment rather than 2 years. Examples of these are the P53 +/- (P53 knockout), TG.AC and the rasH2 mouse models. These animals contain deletions of genes involved in suppressing the formation of tumors. The use of these animal models is justified based upon the fact that these same genes are known to be involved in many human tumors. Thus, the mechanisms leading to enhanced tumorigenic response in these animals is known to be relevant to human tumorigenesis.

There has been controversy over the use of a maximum tolerated dosage in conducting carcinogenicity bioassays. The argument against the MTD concept is that the dosages usually required to satisfy the MTD criteria are unrealistically high relative to the exposure levels expected for the drug. Thus, a carcinogenic response at

the MTD would not be predictive of the same response in humans. However, it is my opinion that this high dosage is necessary to give a stringent test of the drug's potential to cause cancer. After all, we are trying to predict the response of a heterogeneous population with different susceptibilities to potential cancerogenic agents with relatively small numbers of homogeneous strains of animals. Even though we treat the animals for most of their lives, this is still only a 2-year period. Humans are potentially exposed to some drugs for much longer periods of time. We must, therefore, push the test system in order to understand the potential of the new drug. Nevertheless, there are cases where valid exceptions can be made. For instance, if the metabolism of the drug changes (different metabolites or kinetics are observed) at a particular dosage level, then the capacity of the organism to handle the drug in the usual manner has probably been exceeded. A carcinogenic response, or the lack of one, at dosages higher than that level would not be predictive. Obviously, if a drug is nontoxic, then a MTD may not be possible. If a drug causes alterations in levels of, or target organ responsiveness to, sex or thyroid hormones, high dosages may lead to tumor formation in animals secondary to the endocrine disturbances. If such disturbances do not occur below a certain threshold level, and if they are not expected at the human therapeutic dosage, tumors formed in a bioassay at higher than threshold levels will not be predictive of the risk to humans. The same argument applies if the drug stimulates cell division in the organs such as the liver of experimental animals. Such stimulation of cell division will likely lead to tumor formation in rodents. Again, dosages below those that stimulate cell division must be used for extrapolation of bioassay results to humans (16, 17).

The ICH guidelines allow several alternative ways of selecting the highest dosage for carcinogenicity studies in addition to an MTD. The highest dosage may also be selected based on dose-limiting pharmacodynamic effects, or systemic plasma exposure levels at least 25-fold higher than the human plasma. Certain pharmacodynamic effects if allowed to become exaggerated may affect the viability of animals on a study. Examples would be excessive hypotension with an agent that would be used to treat hypertension. In this case, excessive pharmacological action may lead to the death or poor condition of animals before the potential for carcinogenicity could be assessed. Other agents may cause behavioral changes. Dopamine agonists, for example, cause hyperactivity and stereotypical behaviors such as self-mutilating behaviors which, if allowed to become excessive, represent an issue in terms of humane, ethical treatment of animals. The 25-fold AUC criterion can be used only when there is no evidence for

genotoxicity. Under this condition, a dosage which gives a plasma systemic exposure level in terms of AUC (Area Under the Curve) that is 25 times higher than the maximal anticipated human systemic exposure is allowed even if toxicity is not produced. In this case AUCs of both parent drug and metabolites are considered, both individually and combined. In addition, the plasma exposures must be normalized for any differences in plasma protein binding between animals and humans. The "limit" dose concept can be used under certain conditions. There must be no evidence of genotoxicity, the human dose must be ≤ 500 mg/day, and the AUCs obtained in animals must be ≥ 10 times the human AUCs. If these conditions are met, the high dosage in a carcinogenicity study may be limited to 1500 mg/kg/day. Saturation of absorption means that at some point, if a higher dosage is given, no further increase in systemic exposure (plasma AUC) will be obtained. Under these conditions a dosage that is slightly above the dosage where the AUCs stop increasing with dosage will be accepted as the highest dosage. The maximum feasible dosage is limited by the amount of formulation that can be delivered safely to the animals without trauma (usually 10–20 ml/kg/day by oral gavage) or interfering with nutrition (e.g., 5% mixed in the diet). Other factors such as ability to prepare a homogeneous mixture with vehicle or diet or other physical characteristics of the compound or formulation may also be limiting factors.

The statistical analysis of bioassay results is sometimes controversial. Problems arise when early deaths occur and the distribution of such early deaths is not uniform among study groups. This is because in groups where more animals die early, the number of animals remaining at risk for the development of tumors over the remainder of the study is reduced. Thus, the probability of finding a positive result is artificially reduced for those groups. To overcome this problem, a method is commonly used whereby the entire study period is divided into small time intervals, and animals having tumors discovered at necropsy or tumors considered to have caused the death of the animal are compared among groups within the small time intervals. A comparison of the summation of tumor incidence rates within the time intervals is then used to determine if treated groups exhibit an increased incidence of tumors (18–20).

It is important to understand the situations and criteria that determine the need for a carcinogenicity bioassay of a new drug. If a drug is intended for long-term use, for example, some antihypertensive agents or benzodiazepine tranquilizers where the duration of exposure could be years or the remainder of one's lifetime, or if it will be available for widespread, uncontrolled use, then bioassays will likely be required. If the drug affects rapidly growing

tissues, for example, bone marrow or intestinal mucosa, or is known to affect cell division, bioassays may be required. If on the other hand, the drug is intended for short-term, controlled use, say, a few days, weeks, or months, and does not affect rapidly growing tissues, then bioassays will not be required unless mutagenicity tests indicate the drug represents a risk. If positive results have been obtained in one or more mutagenicity tests, the probability is high that a bioassay will be required. If positive mutagenicity results were obtained in *in vitro* studies only, and appropriate *in vivo* studies are negative, the probability of having to conduct a bioassay is much lower than if the *in vivo* studies are positive. In any case, I recommend holding discussions with the appropriate regulatory authority to determine the case for the particular drug under study. Currently, the cost for a set of bioassay studies in rats and mice is close to two million dollars. Therefore, the cost involved warrants a thorough mutagenicity program and meaningful interaction with regulatory authorities.

Immunological Sensitization

New drugs are sometimes tested for antigenic sensitization if they are suspected of being sensitizers. Such drugs would have a chemical structure similar to that of known sensitizers or would be high-molecular-weight compounds such as peptides. Agents to be administered topically are generally tested for dermal sensitization regardless of chemical structure.

Typical protocols for dermal sensitization, active systemic anaphylaxis, and passive cutaneous anaphylaxis tests are given in the article *Animals in Drug Development*, in this encyclopedia. In a typical gel-diffusion test, solutions of test compound, egg albumin (positive control), or vehicle (usually saline) are mixed with complete Freund's adjuvant. These mixtures are injected into animals (usually guinea pigs) of the respective treatment group once a week for 3 weeks. Two weeks after the last injection, blood is collected by cardiac puncture and serum (antiserum) is prepared. The gel diffusion is carried out on double immunodiffusion (Ouchterlony) discs containing trypan blue. Each disc contains a pattern of wells consisting of a circle of six 5-mm wells around a single 6-mm well. Antiserum is placed in the center well, and test compound or positive control or vehicle is placed in the wells surrounding the center well. The discs are incubated in a humidified 37°C incubator for up to a week to allow development of precipitation bands resulting from interaction between antigens (test compound or positive control) and antibodies (in antiserum). The precipitation bands are visible to the naked eye or with a dissecting microscope (21, 22).

Phototoxicity

Phototoxicity testing is not often included in the safety assessment of new pharmaceuticals; however, such studies may be conducted if phototoxicity is expected on the basis of reports on chemicals of similar structure or pharmacological class. Information on UV absorption spectrum of a compound may prompt phototoxicity testing. Sometimes reports of phototoxicity after a drug has reached the marketplace prompt laboratory studies of phototoxicity to determine if the drug is causing phototoxicity and, if so, by what mechanism.

Phototoxicity is to be distinguished from photosensitization (or photoallergy). Phototoxicity results from interaction of the drug or drug metabolites in the skin with light, usually in the UV region, to produce reactive molecular species that cause cell injury or death in the skin. Phototoxicity is most commonly produced under sunlight where the skin is exposed to the full range of UV light. The phototoxic response generally shows a dose response to both the amount of drug and the intensity or duration of light exposure. Photosensitization, on the other hand, results from an immunological response to by-products formed in the skin when the drug or its metabolites interact with light. The initial exposure to the drug and light may not result in an adverse reaction, but subsequent exposures to drug and light will result in an allergic reaction in sensitized individuals. The models described in this section are confined to the assessment of phototoxicity. Modifications of the models for dermal sensitization referred to in the section on immunological sensitization, incorporating light (i.e., UV light) as well as drug exposure, can be used to test for photosensitization.

Several sensitive and convenient models can be used to test for phototoxicity (23, 24). Guinea pigs, rabbits, and mice are the most commonly used species. Normal, albino guinea pigs and rabbits must be shaved or their hair removed with a depilatory before light exposure. Hairless mice have been used because of the obvious advantage of not having to remove the hair prior to testing. The euthymic, hairless guinea pig has also been used. These animals provide a convenient and sensitive model for phototoxicity testing (25). When conducting a phototoxicity test with normal animals, one finds that the hair can never be completely removed from the test site, and that this interferes to varying degrees with light exposure and scoring the resulting skin reaction. In addition, the ears, a site containing relatively small skin area, are sensitive indicators of phototoxicity. In some cases, the ears may be the only part of the animal where a skin reaction is obtained. With hairless animals, a much larger skin area is

available for accurate observation of the skin reaction. The larger size of the hairless guinea pig compared with that of hairless mice maximizes this advantage. With hairless animals, accurate scoring of skin reactions is facilitated because inevitably there will be an area where exposed and unexposed skin are juxtaposed along the sides of the animal. (This assumes the animal was exposed to light coming from above with the animal in the normal resting position on all four feet.) The comparison of the appearance of exposed and unexposed skin side by side in the same animal makes it easier to discern slight changes in skin appearance due to light exposure.

A test using the mouse tail has also been recommended for determining phototoxicity (26). In this test, mice of the same strain, age, sex, and weight are used. Groups of mice are exposed to the drug and UV light; after a period of 24–48 hours, depending on the drug, the tails are removed and the wet weight of the tails is determined. An increase in tail wet weight indicates an inflammatory response due to phototoxicity. Obviously, an untreated group as well as groups treated with only the drug or the light are included as controls.

In Vitro Toxicity

In vitro toxicity studies are being used more commonly in drug safety evaluation as improved techniques are being developed. Their main uses at present are as a tool for screening compounds for toxicity to particular cell types and in studying cellular mechanisms for target organ toxicity.

The most common assay systems consist of cell cultures that are treated with the drug in the culture medium. Toxicity is commonly assessed by release of intracellular enzyme, for example, lactate dehydrogenase or aspartate transaminase, into the culture medium or by other indicators such as decreases in the rate of radiolabeled amino acid or nucleotide precursor incorporation into macromolecules. A decrease in the intracellular uptake of the vital dye, neutral red, has also been used as a measure of toxicity with a variety of cell types. Fluorescent dyes are also available which allow, for example, estimations of intracellular Ca^{2+} concentrations, intracellular pH, or mitochondrial function. Such measurements can be helpful indicators of cellular toxicity and can help determine mechanism of toxicity.

An example of screening compounds for toxicity to particular cell types is the use of rat hepatocyte primary cultures in screening macrolide antibiotics for potential liver toxicity. The liver is a target organ for this class of compounds. In the development of new macrolides, this in vitro culture system is useful in discriminating those

macrolides that are more toxic to liver cells and those that are more toxic than other well-known and widely used antibiotics such as erythromycin. This information, together with data on antimicrobial potency and spectrum, acute in vivo toxicity, absorption, and so on, can be used to select new drug candidates for further development. The in vitro assay provides a rapid, sensitive means to simultaneously compare several compounds, using only milligram quantities of each drug. Such an in vitro system can predict only the inherent toxicity of the drug to hepatocytes. It cannot predict the modifying influences of the rate of absorption, distribution, and elimination that obtain in vivo. Hepatocytes used in such assays do have the ability to metabolize drugs to a limited extent. However, once the drug is placed into the culture medium, the cells are essentially exposed to that concentration of drug for the duration of the incubation period. Such prolonged exposure at the same concentration may not occur in vivo; therefore, the toxicity observed in vitro may not be predictive of the in vivo situation. The same arguments hold for other cell systems such as renal tubule cell, myocardial cell, or keratinocyte cell cultures. However, cell types such as the latter two have the additional limitation that they do not metabolize drugs as do liver or renal tubule cells. If the metabolite of a drug causes toxicity in these cell types, this would be missed in an in vitro assay unless some provision is made for introducing drug metabolizing enzyme systems into the assay.

Prescreening new compounds in in vitro tests for target organ toxicity can minimize the severity of toxicity and thus suffering and mortality in subsequent animal toxicity studies. Because fewer compounds will be tested in animals with this approach, in vitro tests will reduce the numbers of animals used in drug development.

To use such tests rationally, one needs to know the expected target organ for the type of drug under study. This knowledge may come from experience with similar drugs or from initial pharmacological or toxicological studies with a series of new drugs. Cell types such as the rat hepatocyte can also be used to study the mechanism of drug-induced hepatotoxicity. For instance, in addition to measuring the loss of intracellular enzymes, one could measure the amounts of intracellular glutathione, adenosine triphosphate (ATP), calcium, or enzymes of the various metabolic pathways to gain insight into possible mechanisms of toxicity.

Primary cultures of hepatocyte and tubule cells are prepared by collagenase perfusion and dissociation of the liver or kidney into cell suspensions. Primary rat myocardial cells can be similarly prepared by using trypsin to dissociate the cells. Rhythmic contractions of

myocardial cells can be observed in such cultures. Changes in these contractions serve as another end point for the measurement of toxicity (27). The L6 rat skeletal muscle myoblast cell line has been used to screen compounds for potential muscle irritating and damaging activity (28).

Newer technologies that use the methods of molecular biology are being applied to rapid screening of large numbers of compounds (high-throughput screening) for potential toxicity. These methods make use of knowledge of selected genes or patterns of gene expression that have been associated previously with toxic responses. The methods are known generally as genomics or proteomics. One approach is to expose cells, tissues, or an animal to a test compound and extract either RNA or proteins from the cells or tissue of interest. If RNA is extracted, DNA complementary to this RNA (cDNA) can be applied to a chip containing DNA from thousands of different genes to determine which of those genes are expressed in the drug-exposed tissues. This is determined by labeling the cDNA as it is formed with either radioactivity or a fluorescent molecule. The site on the chip where radioactivity or fluorescence is detected allows determination of genes that had expression levels either higher or lower than in untreated tissues. If protein is extracted, the protein extract is subjected to multidimensional gel chromatography to determine patterns of alteration in gene expression. These patterns are then compared to patterns previously shown to be associated with toxic responses in vivo with other compounds. Efforts are being made in some cases in the pharmaceutical industry to understand toxic responses in humans better by determining genetic susceptibility to adverse effects of specific drugs. Gene expression or genetic polymorphisms are studied in people who have had adverse reactions to the drug of interest. Once the genetic nature of the drug sensitivity is understood, other potential patients can be tested to avoid those who might be sensitive to adverse effects of the drug.

There are a few in vitro models available for rapid testing of ocular irritation potential. These tests were developed to minimize exposure of live animals to potentially distressing procedures. These include the BCOP (Bovine Corneal Opacity) test and the Eyetex test (29, 30). In the BCOP test, bovine corneas obtained from slaughter houses are treated under tissue culture conditions with solutions of the test compound to determine whether corneal opacity occurs. Corneal opacity is considered a type of severe ocular irritation. In the Eyetex test, the test compound is added to a solution of proprietary proteins. The protein solution is then observed for cloudiness or precipitation. Both these tests have a fair record of concordance with in vivo data.

Tests such as these are useful for rapid screening of large numbers of compounds or for preliminary testing prior to testing in animals. Compounds predicted to cause severe irritation presumably would not be tested further in animals.

In vitro systems have also been developed to use as screening tools or for mechanistic studies of teratogenicity. Among the most commonly cited examples are the limb bud assay (31) and the hydra assay (32). In the limb bud assay, cultures of mouse embryonic limb bud mesenchymal cells are prepared and exposed to test compound. Under normal conditions, the mesenchymal cells differentiate into chondrocytes. An agent that disrupts this differentiation process would be expected to be teratogenic. In the hydra assay, two protocols have been recommended. In both cases, the ratio of the test compound's degree of toxicity for the adult form of the organism to that for an undifferentiated form of the organism is the end point used to determine the relative teratogenic potential. In one protocol, the undifferentiated form is a mass culture of dissociated cells. In the other, it is the digestive region dissociated from the adult form. In both cases, the undifferentiated form will differentiate or regenerate the adult form unless inhibited by the test compound. Neither of these systems have found wide usage in new drug development. The hydra assay is probably better suited to screening environmental contaminants because of the economy of the assay and the massive number of environmental chemicals likely to be screened.

Immunotoxicity

Immunotoxicology is an area where refined methods of measuring functional changes in a particular organ system are increasingly being applied to the safety assessment of new compounds. The assays used for functional measurements are taken from methods used for research in immunology and for screening new compounds for in vivo antimicrobial or antitumor efficacy.

In routine subchronic-chronic toxicity studies, several variables are obtained that can be indicators of immunotoxicity. These include differential white blood cell count, spleen and thymus weights, histological assessment of spleen cellularity, and thymus and lymph node morphology. However, the new, more refined methods being applied to immunotoxicity assessment are more specific; they can indicate whether particular cell types or immunological functions are affected. These newer specific assays are not at present being applied on a routine basis to screen new drugs. Instead, they are applied selectively in cases where the particular drug is suspected

of being an immunotoxicant. This suspicion may arise from a structural relationship to previously known immunotoxicants (e.g., antivirals or steroids) or from data obtained in routine toxicity studies.

The National Institute of Environmental Health Sciences National Toxicology Program has recommended that immunotoxicity testing be done using a tier approach (33). Inbred strains of mice are usually used for these assays although methodology has been developed for conducting most of the assays in rats (34). The first tier would include those indicators routinely obtained in subchronic/chronic toxicity studies mentioned above, as well as specific assays for humoral-mediated, cell-mediated, and nonspecific immunity. In these specific assays, the animals are treated with a test compound for 14 days, for example, then their peripheral blood lymphocytes or spleen cells are used in the assay. The length of time for in vivo drug treatment will vary with the type of drug and its pharmacokinetics. The assay for humoral-mediated immunity consists of the plaque-forming cell assay to determine the numbers of splenic cells forming IgM antibody to a T-Cell dependent antigen (e.g., sheep red blood cells). In this assay, spleen cells from the treated animal are incubated in a mixture of complement and sheep red blood cells. The number of antibody-forming cells is determined from the plaques formed by areas of lysed sheep red blood cells on a background of unlysed red blood cells. The recommended assays for cell-mediated immunity are lymphocyte blastogenesis in response to a mitogenic stimulus (concanavalin A or phytohemagglutinin) and the mixed leukocyte response. In these assays, the extent of lymphocyte cell division is determined by incorporation of radiolabeled thymidine into cellular DNA after stimulation by a mitogen or by allogeneic lymphocytes, respectively. Natural killer cell activity is determined by incubating spleen cell suspensions from drug-treated animals with radioactive chromium-labeled target cells such as YAC-1 tumor cells. Release of free chromium into the culture medium is a measure of cell-killing activity.

The second tier includes assays that help determine the specific cell type involved in a response seen in the first tier and assays designed to assess the in vivo relevance of results seen in the first tier. These assays include quantitation of the numbers of splenic B and T lymphocytes, IgG humoral-mediated immunity, cell-mediated and nonspecific immunity, and host-resistance challenge models. Quantitation of B and T cells is done using cell-specific antibody techniques. IgG humoral-mediated immunity is assessed similarly to IgM humoral-mediated immunity (discussed above) except that a longer in vitro incubation time is used to allow the development

of IgG antibodies. Cell-mediated immunity can be assessed by the interleukin 2-dependent expression of cell-killer activity in a manner similar to the natural killer cell assay mentioned above. Nonspecific immunity can be assessed by counting the numbers of peritoneal macrophages in drug-treated animals and by measuring the ability of these macrophages to phagocytize radioactive chromium-labeled chicken red blood cells.

The host resistance models provide a means to directly assess the functional, and clinical significance of any changes found in the preceding tests. Direct assessment is important because of the multifactorial nature of immune responses and because of the well-known reserve capacity of the immune system. These assays consist of infectious agents or tumor cells that will routinely kill 10 to 30% of the animals not treated with drugs. An increase in the lethality of the challenge agent indicates suppression of the immune response. The challenge agent can be selected so as to challenge either humoral or cell-mediated immune systems.

SAFETY EVALUATION OF BIOTECHNOLOGY-DERIVED PRODUCTS

The safety evaluation of biotechnology-derived products is a developing area, which has been the subject of controversy. Safety evaluation has been developed for each product on a case-by-case basis. The ICH and FDA have developed guidelines for the safety evaluation of proteins and peptides produced via biological or cell culture systems (35). Several issues that should be addressed for each product have been identified (36, 37), through discussion among toxicologists concerning the appropriateness of applying the toxicity protocols used for synthetic pharmaceuticals to peptide products of biotechnology, and from examples of the few biotechnology-derived products that have been developed for clinical use.

Much discussion has arisen from the issue of potential contamination of the final dosage form by cellular material from the cell culture system used to produce the product. If a mammalian system has been modified by insertion of a DNA vector into cellular DNA, the concern would be over the potential contamination of the final product by transforming DNA. This might potentially increase the risk for developing cancer in a patient receiving such a contaminated product. For such products, assays for transforming activity of the final product or of DNA extracted from the cell culture system would be appropriate. The target limit for contaminant DNA that has generally been discussed is 10 picograms of DNA per

final dose. Assay for contaminant DNA should be established as a routine quality control procedure for production batches. Tests for transforming activity (e.g., cell transformation in C3H/10T 1/2 or Balb/c 3T3 cells) should be performed initially once the cell culture system is established and at periodic intervals to assure genomic stability of the cell culture system. Another source of problems with potential contamination is the antigenicity of the final product due to foreign cellular materials from the biological system used to produce the product, whether bacterial or mammalian. Also, with bacterial production, systems contamination with pyrogenic cell components must be avoided.

The development of neutralizing antibodies against a human protein that would be foreign to the test species is a major concern. The presence or absence of a neutralizing antibody must be determined in each test species used. This could very well limit the species that are appropriate to use in safety studies, and it could limit the duration of any toxicity studies conducted. Generally, neutralizing antibody titers could develop as early as 10 to 14 days after initiation of treatment. Therefore, for a product inducing such antibodies, toxicity studies of more than approximately 14- to 16-days duration would not be appropriate.

Knowledge of the product's pharmacological activity in the species used for safety testing is requisite for a valid toxicity study; obviously, this should be determined beforehand. The only toxicity induced by the product may be in the form of exaggerated pharmacological effects. Exaggerated pharmacological activity may arise through systemic exposure to a protein normally present in only small quantities at specific sites and/or through stimulation of nontarget receptors. In the latter case, if an inappropriate species is used, the safety of the product may be overestimated. Monoclonal antibody products represent a special case of this requirement. The species selected for testing a monoclonal antibody product should provide antigenic receptors for the antibody on the same target tissues as in humans. If the species selected has antigenic receptors that bind the antibody to unintended tissues, or if there are no antigenic receptors that bind the antibody, the results obtained from that species will be misleading and inappropriate. Generally, the closer in structure the product is to the natural product, the lower the need for extensive long-term toxicity studies; however, the greater the difference in structure between the biotechnology-derived product and the natural product, the more the situation comes to resemble that of a synthetic drug where more extensive toxicity studies are required. Improved predictions of human safety may be obtained through the use of transgenic animals bearing the human receptor. Alternatively, toxicity studies in animals may be

conducted using the animal homologue of the human protein under consideration. In addition, it may be possible to conduct studies in an animal model of the human disease. In these cases it should be demonstrated that the pharmacological response of the test system would be similar to the response in human tissues.

Much information can be gained from in-life physiological measurements such as blood pressure, electrocardiograms, body temperature, and respiratory rate. When the number of studies is reduced due to the lack of appropriate species and the potential for development of neutralizing antibodies, as much information as possible should be obtained from the few studies that are conducted. In-life measurements assume more importance in these cases. When the product is likely to contain proteins that may cause acute "serum sickness" type responses and pyrogens, measurements of blood pressure, body temperature, and respiratory parameters are especially important.

Local injection site irritation studies would be expected with these types of products because they are almost always given by injection.

Mutagenicity studies may be conducted if mutagenic contaminants are suspected. If a biotechnology-derived product is closely related to the natural product, there would be no need to conduct mutagenicity studies to assess mutagenicity of the compound per se. Carcinogenicity studies are not routinely needed for these types of products unless there is a concern about induction of tissue proliferation (e.g., with growth factors) or immunosuppression. Reproductive toxicity studies may or may not be conducted, depending on the nature of the product and its intended use. If the compound is suspected of having abortifacient actions (as with the interferons), a study in pregnant primates would be indicated.

USE OF METABOLISM AND PHARMACOKINETIC DATA IN DRUG SAFETY EVALUATION

There are a host of factors that can affect blood and tissue levels of a drug and thus affect exposure of target organs (38, 39).

Selection of the route of administration can affect the degree of absorption and time course of drug levels. The intravenous route can give a rapid rise and fall in drug levels following bolus administration or a steady level during slow infusion. Oral administration usually gives a slower rise and fall in blood levels than intravenous dosing does; however, the extent of absorption may be the same, that is, 100%. Intramuscular, intraperitoneal,

and subcutaneous administrations generally give rates of rise and fall of blood levels between those of intravenous and oral administration. If a drug is insoluble, or precipitates out of solution after injection, the intramuscular, intraperitoneal, and subcutaneous routes may give rise to a prolonged period of absorption from the injection site until all the drug has become solubilized and absorbed from the site.

For the oral route of administration, various factors such as the use of solutions versus suspensions versus dietary administration, the selection of the vehicle, the particle size, and the particle dissolution rate can affect the rate and extent of absorption. The absorption of a drug is likely to be greater if a solution rather than a suspension can be administered; on the other hand, the time course of rise and fall of drug blood levels is likely to be more prolonged with a suspension. There is some evidence that increasing the concentration of the suspending agent, for example, methylcellulose, can cause slower and less complete absorption. With dietary administration, the rate and extent of absorption are generally expected to be lower than with gavage dosing. Because rodents feed nocturnally, the period of drug consumption will be markedly prolonged compared with gavage dosing. However, because of the degree of dispersion necessary to prepare homogeneous dietary drug mixtures, and because of the long period available for drug administration, absorption from the diet can often approach 100% of that for gavage dosing. If a drug is intended to be administered several times per day, for example, over a 12-h period, dietary administration may provide the most appropriate means of modeling the expected human exposure. Particle size and the rate of particle dissolution can affect the rate and extent of drug absorption. The smaller the particle size, the greater the surface area per mass of the particle, and therefore the greater the rate of dissolution and absorption. The toxicologist has to be careful about using different batches of the same compound because the particle size may differ among batches. Such variation may affect the blood levels of a drug obtained in the test animals, which may in turn affect the toxicity observed.

For many drugs, most absorption probably occurs in the small intestine due to the vast surface area available. The rate of gastric emptying affects the delivery of drug to the small intestine. Therefore, the gastric emptying rate may affect the rate of absorption. The normal gastric emptying times are about 10 min for rodents, 30 min for rabbits, and 1.3 h for dogs. However, these times may be altered by drug administration, either pharmacologically or by the effect of variables such as dose volume or osmotic pressure. This further complicates the effect of gastric emptying on the absorption kinetics of the drug.

Oral absorption may be decreased as a percentage of administered dose at high doses. This can occur if the gastrointestinal transit time is less than the time required for complete absorption. This is sometimes the case with poorly absorbed drugs and drugs with poor solubility.

The selection of the vehicle can make a dramatic difference in the toxicity of a compound. An example would be the use of corn oil instead of methylcellulose to suspend a lipophilic compound. Corn oil may allow a better suspension, and possibly dissolve more of the drug, but it may also delay absorption of the drug due to retention in the lumen of the gastrointestinal tract. Opposite effects can also be obtained. I experienced a case where the toxicity of a drug was markedly greater when suspended in corn oil than when suspended in methylcellulose. In attempting to conduct intravenous studies with drugs having poor aqueous solubility, it is often necessary to use nonaqueous solvents, for example, ethanol, propylene glycol, or polyethylene glycol, to dissolve the drug. All nonaqueous solvents cause some degree of toxicity, whether cardiotoxicity, hemolysis, hepatic toxicity, or renal toxicity, that may complicate the toxicity assessment of the drug. Therefore, nonaqueous vehicles must be carefully selected, and the amounts and concentrations administered must be kept to a minimum.

The rate of intravenous injection can be a critical factor in determining the observed toxicity. Rapid administration of intravenous formulations will give high peak blood levels of drug, which may cause toxicity that could be avoided by infusing the drug at a slower rate. It is important in comparing the results of intravenous toxicity studies to note the injection rate. In designing a series of studies with a given formulation, it is important to use the same rate of injection. When determining upper tolerated limits for volume and rate of injection for a given species, knowledge of the glomerular filtration rate for the species can provide a useful guide. If one knows the rate of injection that will be used clinically, it is prudent to use the same injection rate in the animal toxicity studies. Interspecies differences in the rate and extent of absorption and in the metabolism of drugs are common and have a major effect on observed toxicity. Unfortunately, the toxicologist usually does not have this information when designing the initial toxicity studies. If the rat or dog is found to be an inappropriate species due to major differences in these parameters when compared to humans, some of the toxicity work may have to be repeated in another species. With some drugs, a large percentage may be metabolized to an inactive form by the intestinal mucosa or the liver before all the absorbed drug reaches the systemic circulation. This is called the "first-pass effect" because some of the active drug is removed

during the first pass through the mucosa or liver, which obviously decreases the effective or apparent absorption of the drug. On the other hand, if a large amount of the drug is given (e.g., at high doses in a toxicity study) so that the capacity of the first-pass metabolism is exceeded, the blood levels obtained will be disproportionately high and lead to excessive toxicity. It is important to be aware of any first-pass metabolism when designing oral or intraperitoneal toxicity studies. When a drug is administered intraperitoneally, most of the drug is absorbed via the portal circulation that flows through the liver before reaching general circulation.

When an active drug is excreted in the bile, it may be reabsorbed. This sets up a cycle of biliary excretion, reabsorption, biliary excretion, and reabsorption. Usually only a fraction of the drug is excreted in the bile, and only a fraction thus excreted is reabsorbed. Therefore, the cycle does not continue indefinitely when treatment is stopped. However, such enterohepatic recirculation can lead to higher blood levels of drug than would otherwise be obtained and can prolong the apparent blood or tissue half-life of the drug. Such a cycle can lead to alterations in the blood level-versus-time curve such that after a single dose the initial peak level is followed by a smaller peak caused by the reabsorption of excreted drug from the small intestine. This phenomenon can affect the blood levels and thus the toxicity obtained with a given dosage level, and should be taken into consideration when designing toxicity studies.

With substantial increases in body weight and in pregnancy, the volume of distribution will be increased since the blood and tissue volumes increase. The half-life ($t_{1/2}$) is directly related to the volume of distribution (V_d) as follows:

$$t_{1/2} = \frac{V_d \times AUC \times 0.693}{D}$$

where AUC is the area under the blood concentration-versus-time curve and D is the dose. With substantial increases in body weight, the $t_{1/2}$ for elimination increases also. The time required to reach a steady-state concentration of drug in blood or tissues upon repeated administration, or to eliminate virtually all of a drug after cessation of treatment, is roughly 4 to $5 \times t_{1/2}$. Therefore, the weight gain found in rodents during their phase of rapid growth (the first 3 to 5 months and during pregnancy) will have an effect on the time required to reach steady-state blood-tissue levels, and on the time required to clear a drug from the body.

The concept that four to five half-lives are required to achieve a steady blood level is also of practical importance when determining whether there is any accumulation of

blood levels over the course of a subchronic-chronic toxicity study. It is the usual practice to take blood samples early in the study and again toward the end of the study to determine if blood levels of the drug are higher at the end than at the beginning of treatment. If not enough time is allowed at the beginning of the study for steady-state blood levels to be obtained, the results may mislead one to conclude that accumulation has occurred. Similarly, if a recovery period is included in the study design, it is important to allow enough time for the drug to be eliminated from the test animals and to allow additional time for recovery from drug effects beyond that required to eliminate the drug.

Changes in the organs of elimination (e.g., liver and kidneys) can have dramatic effects on the blood levels of a drug, such that excessive blood levels and concomitant toxicity are obtained due to reduced elimination of drug. However, some drugs induce increased synthesis of metabolizing enzymes, mainly in the liver, and sometimes increased rate of bile flow. In these cases, lower-than-expected blood levels may be obtained. This may occur at higher dosages and after a period of time necessary to stimulate the metabolic machinery.

Many drugs are metabolized to products that have pharmacological and/or toxicological activity. The pharmacokinetics of these metabolites may differ from those of the parent drug, complicating the design and interpretation of toxicity studies.

EXTRAPOLATION OF ANIMAL TOXICITY DATA TO HUMANS

Continually present in drug safety evaluation are questions concerning the predictive value of toxicological findings in animals relative to those expected to occur in humans. Two questions are of importance: Will the same kinds of toxic changes occur in humans that occur in laboratory animals? At what dosage may toxic changes be expected in humans relative to the dosage in animals?

There are differences in the processes of absorption, metabolism, and excretion that have been identified between several laboratory species and humans (40). These differences are useful in the selection of the appropriate species in which to study particular compounds. However, having selected, as far as practical, the appropriate species, it is generally recognized that if toxic changes occur in more than one species, they can be expected to occur in humans as well. Predicting the kind of toxicity that will occur in

humans is usually less of a problem than predicting the dosage at which it will occur.

Generally, laboratory animals are less sensitive on a per body weight basis to the toxic effects of chemicals than are human beings. The sensitivity generally varies inversely with the size of the animals. More specifically, the sensitivity varies inversely with the surface area-to-body weight ratio. The surface area of a mammal is roughly the body weight raised to the two-third power. The ratio of the body surface area to the body weight determines the amount of energy that must be expended to maintain normal body temperature. The energy that is converted to body heat to maintain temperature is derived from the basal metabolic processes of the animal. The smaller the animal, the higher the metabolic rate because the ratio of surface area to body weight is larger. The rate at which drugs and other chemicals are metabolized and excreted generally varies with the basal metabolic rate of the species, which in turn varies with the surface area-to-body weight ratio. Thus, dosages comparable to a 1-mg/kg dosage in humans, after adjustment for differences in surface area, would be about 3 to 5 mg/kg for species such as rabbits, dogs, and monkeys, and about 10 mg/kg for rodents such as rats and mice. These factors are frequently used to establish the safe dosage for humans on the basis of animal toxicity studies. For example, the no-toxic-effect dosage obtained in a rodent toxicity study may be divided by 10 to determine the maximum dosage for clinical use in humans.

Application of the results of carcinogenicity studies to humans continues to be a controversial subject. Some of the issues involved were mentioned in the section on carcinogenicity bioassays. The relevance of tumors found at unrealistically high dosages is questioned as is the extrapolation of tumor incidence-dose response curves to low dosages where, presumably, no additional tumors above spontaneous incidence would be found. In some cases, as mentioned previously, where there is a clear change in metabolism or pharmacokinetics of the drug at high dosages, results found at such dosages may not be expected to occur in humans under normal, presumably much lower, exposure levels. The validity of the extrapolation of tumor incidence-dose response curves to low dosages where, theoretically, the tumor incidence should be zero, depends upon whether the compound acts as a genotoxic initiator or as a nongenotoxic promotor or cocarcinogen. Genotoxic compounds induce initiating, that is, DNA or chromosome damaging, events that can eventually lead to tumor formation. Nongenotoxic promoters or cocarcinogens can lead to tumor formation only in conjunction with the action of a genotoxic initiating compound. Promoters can act when applied

either at the same time as or after the application of the initiating compound. Cocarcinogens can act only when applied at the same time as the initiator. For most genotoxic compounds, theoretically, one molecule is sufficient, if it reaches the target DNA, to cause an initiating lesion. For promoters and cocarcinogens, it is generally understood that one molecule is not sufficient; there is a threshold concentration or dose below which the promoting or cocarcinogenic action is not obtained. Therefore, extrapolation to dosages below the threshold dosage would be appropriate for agents that act via promotion or cocarcinogenesis but not for genotoxic initiating compounds. An example of a class of compounds that cause DNA damage indirectly via interaction with a non-DNA target would be the topoisomerase inhibitors. Through inhibition of topoisomerase action DNA strand breakage normally caused by the enzyme is maintained, whereas without the action of the drug the breaks would be rejoined as part of the normal process of maintaining the DNA. The interaction of drug with the topoisomerase enzyme is subject to a dynamic equilibrium so that a certain concentration of drug is required to interfere with enzyme action. This drug effect, although causing DNA damage should demonstrate a threshold effect; since concentrations of drug below the critical level should be without biological effect (41).

Application of animal data on eye, skin, muscle, and vein irritation and pain on injection to humans merits comment. The rabbit eye is the most commonly used model for predicting eye irritation. This model is generally believed to be slightly more sensitive than the human eye to most irritants (42, 43). The same can be said for the rabbit skin as a model for human skin irritants. The muscle irritation test in rabbits seems to have a good correlation with clinical experience regarding painful intramuscular formulations even though this is not a test for pain (44). If a formulation causes muscle damage in rabbits, because of the similarity of muscle tissue across species, it can be expected to do so in humans. Vein irritation studies in animals are not always predictive of clinical results. If a formulation produces marked vein irritation in animals, it will likely do so in humans. However, a low level or absence of vein irritation in animals does not mean the formulation will not be irritating clinically. Several variables influence the irritating nature of an intravenous formulation. These are concentration of the drug, tonicity, pH, formulation ingredients other than the drug, rate of infusion, and number of repeated injections in the same site. Naturally, any changes in these variables between animal and clinical studies affect the predictability of animal vein irritation studies. Results of animal muscle

and vein irritation studies cannot be used to predict pain on injection in humans because pain is not an end point in these tests. The only available models for pain on injection, by any route, are the rat paw lick and mouse scratch tests mentioned previously. These seem to be predictive for the few compounds that have been tested for a correlation with clinical experience.

The application of data from reproductive toxicity studies to humans involves special considerations in addition to those mentioned above for general toxicity studies. This is especially true of teratology studies. The actual reliability of animal models to predict toxicity in this area is never known until adequate epidemiological studies are conducted in humans. Such epidemiological studies are rarely conducted due to the expense and time required. The ideal animal model would handle the drug (i.e., through absorption, distribution, metabolism, excretion, and maternal fetal distribution) identically to humans and would have the same maternal-placental-fetal relationships as humans. Because the animal models available to us rarely represent the ideal model, the best compromise is to test, for teratogenicity at least, in two or more species (45). If the drug causes fetal toxicity or teratogenicity in two or more species, it would be a good assumption that the drug will do the same in humans. Of the chemicals known to be teratogenic in humans, all but one class, the coumarin anticoagulants, are also teratogenic in at least one species of laboratory animal, (46). However, there are many chemicals known to be teratogenic in animals that are apparently not teratogenic in humans. This may be due to lower levels of exposure in humans or to more careful management of exposure as in the case of pharmaceuticals. Animal models have greater predictive power when combinations of two or more species are used. Among the species commonly used for teratogenicity testing, rats and mice seem to be the best for modeling human responses. However, the use of rabbits in combination with either rats or mice provides greater predictability because rabbits are less likely to yield false positive results. The rabbit was originally selected as a second, nonrodent species for routine teratogenicity testing because of its responsiveness to thalidomide.

Finally, it should be pointed out that although the animal models commonly used can predict the fetal toxicity of a drug in humans, the exact type of fetal toxicity in humans cannot be predicted. If an agent causes fetal toxicity of any kind in animals, it should be suspected as a potential human teratogen or fetal toxin. Valproic acid serves to illustrate this concept: It causes skeletal abnormalities in rats and rabbits. In humans, it has been shown by epidemiological studies to cause a low incidence

of spina bifida, a neural tube defect. Therefore, the type of fetal toxicity caused in the animal models may differ from that caused in humans, but the fetal toxicity in humans could be predicted from the animal studies. It turns out that valproic acid also causes neural tube defects in mice and hamsters. In summary, the animal models used for teratogenicity screening seem to provide reasonable predictability for fetal toxicity in humans. The greater the number of species with positive results, the greater the likelihood that the drug will also cause fetal toxicity in humans.

REFERENCES

- Sherman, M. Drug and Cosmetic Industry. **1988**, 142 (2), 37–78.
- McCutcheon, R.S. Toxicology and the Law. *Casarett and Doull's Toxicology, The Basic Science of Poisons*, 2nd Ed.; J. Doull, J., Klaassen, C.D., Amdur, M.O., Eds.; MacMillan: New York, 1980; 727–733.
- Finney, D.J. Probit Analysis. Cambridge University Press: Cambridge, MA, 1971.
- Brownlee, K.A.; Dodges, J.L.; Rosenblatt, M. J. Am. Statist. Assoc. **1953**, 48, 262–277.
- Thompson, W.R. Bacteriol. Rev. **1947**, 11, 115–145.
- Celozzi, E.; Lotte, J.; Stapley, E.O.; Miller, A.K. J. Pharm. Meth. **1980**, 4, 285–289.
- Comereski, C.R.; Williams, P.D.; Bregman, C.L.; Hottendorf, G.H. Fundam. Appl. Toxicol. **1986**, 6, 335–338.
- Detweiler, D.K. The Use of Electrocardiography in Toxicological Studies with Rats. *The Rat Electrocardiogram in Pharmacology and Toxicology*; Budden, R., Detweiler, D.K., Zbinden, G., Eds.; Pergamon Press: Oxford, 1981; 83–115.
- Schaeppi, U.; Liverani, F. Agents Actions **1977**, 7 (3), 347–351.
- Fort, F.L.; Yamamoto, M.; Ando, T. The Toxicol. **1987**, 7 (1), 96.
- Vobig, M.A.; Klotz, T.; Staak, M.; Bartz-Schmidt, K.U.; Engelmann, U.; Walter, P. Retinal Side-Effects of Sildenafil. The Lancet **1999**, 353–375.
- Tennant, R.W.; Margolin, B.H.; Shelby, M.D.; Seiger, E.; Haseman, J.K.; Spalding, J.; Caspary, W.; Resnick, M.; Staseiwicz, S.; Anderson, B.; Minor, R. Science **1987**, 236, 933–941.
- International Conference on Harmonisation Draft Guideline on Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals; Federal Register 1987; 16026–16029.
- Galloway, S.M.; Deasy, D.A.; Bean, C.L.; Kraynak, A.R.; Armstrong, M.J.; Bradley, M.O. Mutat. Res. **1987**, 189, 15–25.
- Cifone, M.A.; Myhr, B.; Eiche, A.; Blocsfoldi, G. Mutat. Res. **1987**, 189, 39–46.
- Haseman, J.K. Fundam. Appl. Toxicol **1985**, 5, 66–78.
- Clayson, D.B. Mutat. Res. **1989**, 221, 53–67.
- Peto, R.; Br., J. Cancer. **1974**, 29, 101–105.
- Peto, R.; Pike, M.C.; Day, N.E.; Gray, R.G.; Lee, P.N.; Parish, S.; Peto, J.; Richards, S.; Wahrendorf, J. *IARC Monograph on Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal, Suppl. 2*; IARC: Lyon, France, 1980.
- McKnight, B. Fundam. Appl. Toxicol. **1988**, 10, 355–364.
- Arakawa, M.; Hori, T.; Innoue, S.; Kanamura, K. Jap. Pharmacol. Therap. **1985**, 13 (7), 459–469.
- Campbell, D.H.; Garvey, J.S.; Cremer, N.E.; Sussdoff, D.H. Gel Diffusion. *Methods in Immunology*, 2nd Ed.; W.A. Benjamin, Inc.: New York, 1970; 250–260.
- Harber, L.C. J. Invest. Dermatol. **1981**, 77, 65–70.
- Sams, W.M. Arch. Dermatol. **1966**, 94, 773–777.
- Fort, F.L.; Kotz, R.V. The Toxicol. **1988**, 8(1), 129.
- Ljunggren, B. Photodermatol **1984**, 1, 96–100.
- Miletich, D.J.; Albrecht, R.F.; Hoishouser, S.J. In Vitro Toxicol. **1987**, 1 (2), 77–84.
- Williams, P.D.; Masters, B.G.; Evans, L.D.; Laska, D.A.; Hottendorf, G.H. Fundam. Appl. Toxicol. **1987**, 9, 10–17.
- Sina, J.F. Validation of the Bovine Corneal Opacity-Permeability Assay as a Predictor of Ocular Irritation Potential. In Vitro Toxicol. **1994**, 7(3), 283–290.
- Sina, J.F.; Galer, D.M.; Sussman, R.G.; Gautheron, P.D.; Sargent, E.V.; Leong, B.; Shah, P.V.; Curren, R.D.; Miller, K. A Collaborative Evaluation of Seven Alternatives to the Draize Eye Irritation Test Using Pharmaceutical Intermediates. Fundam. Appl. Toxicol. **1995**, 26, 20–31.
- Kistler, A. Arch. Toxicol. **1987**, 60, 403–414.
- Wilby, O.K.; Newall, R.D.; Tesh, J.M. Food Chem. Toxicol. **1986**, 24 (6/7), 651–652.
- Luster, M.I.; Munson, A.E.; Thomas, P.T.; Holsapple, M.P.; Fenters, J.O.; White, K.L.; Dauer, L.D.; Germolec, D.R.; Rosenthal, G.J.; Dean, J.D. Fundam. Appl. Toxicol. **1988**, 10, 2–19.
- Exon, J.H.; Koller, L.D.; Talcott, P.A.; O'Reilly, C.A.; Henninsen, G.M. Fundam. Appl. Toxicol. **1986**, 7, 387–397.
- International Conference on Harmonisation Guidance on Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. Federal Register 1997; 62 222 61515–61519.
- Graham, C.E. *Preclinical Safety of Biotechnology Products Intended for Human Use*; Alan R. Liss, Inc.: New York, 1987.
- Fort, F.L. J. Biomed. Environm. Sci. **1988**, 1 (2), 223–224.
- Smyth, R.D.; Hottendorf, G.H. Toxicol. Appl. Pharmacol. **1980**, 53, 179–195.
- Clark, B.; Smith, D.A. CRC Crit. Rev. Toxicol. **1984**, 12 (4), 343–385.
- Calabrese, E.J. Drug Metab. Rev. **1984**, 15 (3), 505–523.
- Fort, F.L. Mutagenicity of Quinolone Antibacterials. Drug Safety **1992**, 7 (3), 214–222.
- Freeberg, F.E.; Nixon, G.A.; Reer, P.J.; Weaver, J.E.; Bruce, R.D.; Griffith, J.F.; Sanders, L.W. Fundam. Appl. Toxicol. **1986**, 7, 626–634.
- Kay, J.H.; Calandra, J.C. J. Soc. Cosmetic Chem. **1962**, 13, 281–289.
- Gray, J.E. Fundam. Appl. Toxicol. **1981**, 1, 290–292.
- Schardein, J.L.; Schwetz, B.A.; Kenel, M.F. Environ. Health Perspect. **1985**, 61, 55–67.
- Schardein, J.L.; Keller, K.A. CRC Crit. Rev. Toxicol. **1989**, 19 (3), 251–339.

BIBLIOGRAPHY

- Acute Eye Irritation Testing, Recommended Guideline*; Interagency Regulatory Liaison Group: Testing Standards and Guidelines Work Group, January 1981.
- Acute Dermal Toxicity Test, Recommended Guideline*; Interagency Regulatory Liaison Group: Testing Standards and Guidelines Work Group, January 1981.
- Acute Oral Toxicity Testing in Rodents, Recommended Guideline*; Interagency Regulatory Liaison Group: Testing Standards and Guidelines Work Group: January 1981.
- Alder, S.; Zbinden, G. *National and International Drug Safety Guidelines*; M. T. C. Verlag: Zollikon, Switzerland, 1988.
- Brusick, D. *Principles of Genetic Toxicology*; Plenum: New York, 1980.
- Brusick, D.J.; Bradley, M.O. *In Vitro Toxicology* **1988/89**, 2 (3), 151–236.
- Graham, C.E. *Preclinical Safety of Biotechnology Products Intended for Human Use*; Alan R. Liss: New York, 1987.
- Grice, H.C. *The Selection of Doses in Chronic Toxicity/Carcinogen Studies; Age-Associated (Geriatric) Pathology: Its Impact on Long-Term Toxicity Studies; Current Issues in Toxicology*; Springer-Verlag: New York, 1984.
- Grice, H.C. *Interpretation and Extrapolation of Chemical and Biological Carcinogenicity Data to Establish Human Safety Standards; The Use of Short-Term Tests for Mutagenicity and Carcinogenicity in Chemical Hazard Evaluation; Current Issues in Toxicology*; Springer-Verlag: New York, 1984.
- Hayes, A.W. *Principles and Methods of Toxicology*, 3rd Ed.; Raven Press: New York, 1994.
- Heddle, J.A.; Dean, S.; Hohmi, T.; Boerrigter, M.; Casciano, D.; Douglas, G.R.; Glickman, B.W.; Gorelick, N.J.; Mirsalis, J.C.; Martus, H.J.; Skopek, T.R.; Thybaud, V.; Tindall, K.R.; Yajima, N. *In Vivo Transgenic Mutation Assays*. *Environ. Mol. Mutagen.* **2000**, 35, 253–258.
- International Conference on Harmonisation; Guideline on Dose Selection for Carcinogenicity Studies on Pharmaceuticals. *Federal Register* **1995**, 11278–11280.
- International Conference on Harmonisation; Final Guideline on the Need for Long-Term Rodent Carcinogenicity Studies of Pharmaceuticals. *Federal Register* **1996**, 61, 428154–8156.
- International Conference on Harmonisation; Guidance on Dose Selection for Carcinogenicity Studies on Pharmaceuticals: Addendum on a Limit Dose and Related Notes, *Federal Register* **1997**, 64259–64261.
- International Conference on Harmonisation; Guidance on Testing for Carcinogenicity of Pharmaceuticals, *Federal Register* **1988**, 63 35, 8983–8986.
- Kesterson, J.W. *Drug Inform. J.* **1982**, Jan/June, 22–24.
- Kilbey, B.J.; Legator, M.; Nichols, W.; Ramel, C. *Handbook of Mutagenicity Test Procedures*; Elsevier: 1984.
- Klaassen, C.D. *Casarett and Doull's Toxicology*, 5th Ed.; McGraw-Hill: New York, 1996.
- MacGregor, J.T.; Farr, S.; Tucker, J.D.; Heddle, J.A.; Tice, R.R.; Turteltaub, K.W. *New Molecular Endpoints and Methods for Routine Toxicity Testing*. *Fundamen. Appl. Toxicol.* **1995**, 26, 156–173.
- Marzulli, F.N.; Maibach, H.I. *Dermatotoxicology*, 5th Ed.; Taylor & Francis: Washington DC, 1983.
- Marzulli, F.; Maguire, H.C. *Food Chem. Toxicol.* **1987**, 20, 67–74.
- Mirsalis, J.C.; Monforte, J.A.; Winegar, R.A. *Transgenic Animal Models for Measuring Mutations In Vivo*. *Crit. Rev. Toxicol.* **1994**, 24 (3), 255–280.
- Naismith, R. *Mutat. Res.* **1987**, 189 (2), 81–183.
- Nuwaysir, E.F.; Bittner, M.; Trent, J.; Barrett, J.; Afshari, C.A. *Microarrays and Toxicology the Advent of Toxicogenomics*. *Mol. Carcinogen* **1999**, 24, 153–159.
- Principles and Methods for Evaluating the Toxicity of Chemicals, Part I*; Environmental Health Criteria No. 6, World Health Organization: Geneva, 1978.
- Provost, G.S.; Kretz, P.L.; Hamner, R.T.; Matthews, C.D.; Rogers, B.J.; Lundberg, K.S.; Dyaico, M.J.; Short, J.M. *Transgenic Systems for In Vivo Mutation Analysis*. *Mut. Res.* **1993**, 288, 133–149.
- Purchase, I.F.H.; Hard, G.C. *Toxicology In Vitro* **1987**, 1 (1), 3–57.
- Schardein, J.L. *Drugs as Teratogens*; CRC Press: Cleveland, 1976.
- Schardein, J.L. *Chemically Induced Birth Defects*, 3rd Ed.; Marcel Dekker, Inc.: 2000.
- Sontag, J.M.; Page, N.P.; Safflotti, U. *Guidelines for Carcinogen Bioassay in Small Rodents, National Cancer Institute Carcinogenesis, Technical Report Series No. 1*; Publication No. (NIH) 76-801: DHEW, February, 1976.
- Taylor, P. *Practical Teratology*; Academic Press: New York, 1986.
- Teratogenicity Studies in the Rat, Mouse, Hamster, or Rabbit: Recommended Guideline*; Interagency Regulatory Liaison Group: Testing Standards and Guidelines Work Group, January 1981.
- Wilson, J.G. *Environment and Birth Defects*; Academic Press: New York, 1973.

Dry Powder Aerosols: Emerging Technologies

Nora Y.K. Chew
Hak-Kim Chan

University of Sydney, Sydney, New South Wales, Australia

INTRODUCTION

The delivery of pharmaceutical aerosols as a dry powder is no longer perceived as the “second best” method to deliver drugs to the lung. This is because of the mounting technologies, which are capable of making stable powders of respirable size and devices competent to deliver accurate doses and versatile payload.

The identification of the lung to deliver drugs to the bloodstream has expanded the kinds of chemical entities that can be delivered via this route. Therapeutic proteins and polypeptides such as insulin administered as a powder via the lung are examples showing substantial systemic absorption into the bloodstream, resulting in well-controlled blood glucose level for diabetes.^[1,2]

The delivery of therapeutic macromolecules by inhalation, however, presents additional problems and challenges to produce fine powders of particle size 0.5 μm –5 μm that flow well during manufacture, filling, and emptying from the inhaler device. Macromolecules are more “fragile” compared with small molecules, thus imposing certain restrictions during manufacture and storage in order to maintain not only their dispersibility, but also their biochemical and physical stability. Proteins and polypeptides are often substantially more expensive than the small molecules; therefore, the ability to efficiently deliver the dry powder with minimal loss is crucial. Furthermore, adequate dispersion and delivery of the powder to the patient are especially important for reproducible lung distribution and systemic absorption of drugs.

This review captures some of the most recent technologies with examples relating to pharmaceutical dry powder aerosol delivery.

EMERGING TECHNOLOGIES FOR POWDER AEROSOL DELIVERY

Two main areas that have been given most attention to improve inhalation drug delivery of powders are: (I) powder production and formulation and (II) powder inhaler devices.

Powder Production and Formulation

Traditionally, freeze-drying has been the process used to produce dry powders of proteins and peptides from a solution. This is because the solution can be dried without being exposed to elevated temperatures, which may adversely affect the product stability. Unfortunately, powders prepared by this method are too large in size for inhalation. Even though milling of the freeze-dried powders to reduce particle size is feasible, the pressure, temperature, and shear resulting from milling may cause significant loss of bioactivity of proteins.^[3,4]

To date, the two most successful ways for making powders for inhalation are spray drying and solvent precipitation.

Spray drying

Spray drying involves converting the atomized liquid droplets into dry powders by hot air. This one-step process is capable of making particles of size suitable for inhalation.^[5] The particle size and size distribution of the powder can be manipulated by the concentration of the feed solution, the spray temperature, cyclone efficiency, and chemical nature of the feed.^[6]

Therapeutic proteins prepared as dry powders are often found unstable when dried alone. Due to the tremendous surface area of the atomized droplets and the relatively high drying temperatures and mechanical stress during atomization, the integrity of the protein may be adversely affected during spray drying. Protein degradation can be minimized by cospray drying the protein with excipients as has been reported for recombinant human deoxyribose nuclease (rhDNase) for inhalation.^[7,8]

Molecular aggregation of recombinant human growth hormone (rhGH) due to air–liquid interfacial degradation can be prevented by adding polysorbate-20 (with no sugar protectant) or Zn^{2+} into the liquid feed. Polysorbate-20 significantly reduced the formation of insoluble protein aggregates, while Zn^{2+} suppressed the formation of soluble protein aggregates. Combination of polysorbate-20 and Zn^{2+}

resulted in a spray-dried rhGH powder having insignificant protein degradation. The occupancy of polysorbate at the air–liquid interface of spray droplets and the formation of a dimer complex between Zn^{2+} and rhGH are believed to reduce the chance for protein unfolding and aggregates formation.^[9]

Spray drying of recombinant humanized monoclonal antibody, anti-IgE (rhMAbE25) containing trehalose or lactose had $\leq 1\%$ of aggregates formed following spray drying.^[10] Despite that, the powders containing trehalose were too cohesive for aerosol delivery. Mannitol was less capable of stabilizing rhMAbE25, with 1%–3% aggregates found following spray drying. The stabilizing effect of mannitol leveled off at a mannitol concentration of ~ 20 wt. %.^[11] In the case of follicle stimulating hormone (FSH), formulation containing mannitol, sucrose, and raffinose as excipients gave rise to $\leq 2\%$ of the higher order aggregates formed from spray drying. Mannitol-containing formulations had the highest amount of aggregates (1.2%) than those of sucrose and raffinose (0.05%). The production yield of powders from the mannitol-containing formulation (50%) was much less than that of sucrose (79%) and raffinose (74%). Despite these, the mannitol-containing formulation has the highest emitted dose of 66 wt. % vs. 15 wt. % and 55 wt. % obtained from the sucrose- and raffinose-containing formulation, respectively.^[12] These results indicate that each protein/excipient system requires individual characterization to identify an optimal formulation for powder aerosol performance and protein stability.

Modification of the spray drying process

Spray Freeze-Drying. Spray freeze-drying, a method developed for producing protein aerosol powders, involves spraying the feed solution into liquid nitrogen followed by freeze-drying. This process produces large ($\sim 8\text{ }\mu\text{m}$ – $10\text{ }\mu\text{m}$) but porous particles of rhDNase and anti-IgE with high production yields ($> 95\%$). The fine particle fraction (FPF) of the spray freeze-dried powder was significantly higher than that of the spray-dried powder due to improved aerodynamic properties.^[13] The overall process, however, is much more costly, time consuming, and complex as compared with spray drying.

Process for Spray Drying Hydrophobic Drugs. The conventional spray drying process is usually limited to hydrophilic drugs with hydrophilic excipients in aqueous solutions. It is not feasible for systems containing hydrophobic drugs and hydrophilic excipients or vice versa. While spray drying of hydrophobic materials can be accomplished using an organic solvent, a patented invention

has been developed for spray drying pharmaceuticals and other compositions, which comprise both hydrophobic and hydrophilic substances. A basic requirement is that the hydrophilic excipient and hydrophobic drug would be at least partially dissolved in the same organic solvent or cosolvent system.^[14] For example, both budesonide (hydrophobic drug) and povidone (hydrophilic excipient) have high solubility in methanol. Such a drug/excipient composition yielded powders of slightly dimpled spheres, with moisture content of 0.49 wt.%, and particle size of $2.3\text{ }\mu\text{m}$. The delivered dose efficiency was measured as ~ 50 wt.%. The use of a nonaqueous or partially aqueous system has the advantage of preparing powders that are physically or chemically sensitive to water while in solution or to residue moisture in the powder.

Solvent precipitation

The solvent precipitation method utilizes the unique properties of nonsolvent at a critical temperature and pressure to precipitate solid particles of drugs from solutions. Carbon dioxide, which exhibits remarkable solvent power at its critical temperature of 31.1°C and pressure of 70 bar for high molecular weight and low vapor pressure solids, is an ideal nonsolvent choice. CO_2 is also nontoxic, inexpensive, and readily available. Technical details of supercritical fluid technology can be found in Vol. 18 of this encyclopedia series.^[15] The technology has been successfully applied to the production of fine particles for aerosol delivery.^[16–18]

York and Hanna^[17] have developed the SEDS (solution enhanced dispersion by supercritical fluids) process for preparing powders of an anti-asthmatic drugs salmeterol xinafoate. The method is capable of controlling the dispersion of the solution in the system, and thus has the ability to manipulate the characteristics of the particles in the micrometer-sized range. This technique has also been used to prepare powders of therapeutic proteins, with the key experimental component being the three-channeled coaxial nozzle designed to use high velocity supercritical CO_2 to disperse the feed of aqueous and organic nature into the particle formation vessel. Relatively high percent of bioactivity was maintained for lysozyme and a therapeutic peptide (95% and 100%, respectively).

Alternatively, protein particles can be produced using a supercritical or near critical CO_2 -assisted aerosolization and bubble drying process.^[20] This method utilizes the high solubility of CO_2 in water, coupled with expansion of the solution through a nozzle to aerosolize aqueous solutions of drugs. When the microbubbles formed are dried, solid (such as lactose and albuterol sulfate, with size between $0.5\text{ }\mu\text{m}$ and $5\text{ }\mu\text{m}$) or hollow spherical particles (such as sodium

chloride, mannitol, or tobramycin sulfate) are formed depending on the compound. Protein powders such as lysozyme or lactate dehydrogenase can also be produced by this process and can be stabilized through the use of sugars, buffers, and surfactant additives in the formulations. Depending on the solute and conditions of drying, the particles are crystalline in some cases and amorphous in others.^[21]

Recently, Bustami et al.^[22] have investigated the feasibility of the ASES (aerosol solvent extraction system) process to generate microparticles of proteins for inhalation. Protein powders generated were of particle size 100 nm–500 nm. In vitro performance showed 65, 40, and 20 wt. % respirable fraction for lysozyme, albumin, and insulin, respectively. Little or no loss of monomer content was observed for these proteins.

Ways to improve powder flowability and dispersability

Inclusion of Excipient. Powders prepared by any one of the above methods may not be used directly as the powders may be too cohesive for device filling and for administration as an aerosol. This is the case for anti-IgE cospray dried with trehalose^[10] and spray-dried rhDNase powders.^[8] In contrast to the traditional approach of using binary blend systems (i.e., drug plus coarse carrier), the addition of fine carrier particles ($< 10\ \mu\text{m}$) such as lactose and magnesium stearate to form a ternary system has been shown to further enhance the amount of drug particles in the aerosol cloud.^[23,24] The role of the fine carrier particles is believed to reduce the drug–coarse carrier interaction by occupying possible drug binding sites on the larger carrier particles. The formation of multiplets between the fine carrier and drug may hinder the direct contact between the drug and the coarse carrier, thus promoting detachment of drug particles from the carrier surface during powder dispersion. Magnesium stearate at concentration as low as 0.1% was found to be sufficient to increase the fine particles of salbutamol sulphate in the aerosol cloud.^[24]

Recently, Ganderton et al.^[25] have found that the inclusion of low-density amino acid particles produced by spray drying can also enhance the amount of fine particles in the aerosol. Such low-density, flake-like particles have a bulk density of 0.1 g/mL with thickness of $\leq 100\ \mu\text{m}$ and $\text{MMAD} \leq 10\ \mu\text{m}$. Addition of low-density leucine particles as low as 1% enhances the fine particles of salbutamol in the aerosol by $\sim 25\ \text{wt. \%}$. However, the addition of 10% leucine made no difference to the amount of fine particles in the aerosol compared with salbutamol alone.

Another conventional approach to improving powder flow and dispersion is to form agglomerates. However, liquids that are commonly used, e.g., alkanols, may dissolve

and denature the actives such as macromolecules. The recent discovery of fluorocarbon (FC) liquids including perfluorodecalin and perfluorooctyl bromide as alternatives may be useful because they are hydrophobic and do not dissolve proteins. In addition, FC has a low surface tension; relatively weak bonds are expected to form between the fine particles of the loose agglomerates, which can be then redispersed readily as an aerosol. The high vapor pressure of FC renders low solvent residue in the final product.^[26] Fluorocarbon does not contain chlorine atoms, and is therefore ozone friendly.

Tailor-made powders

Besides the use of excipients, powder flow and dispersion can also be improved by manipulation of the physical characteristics of the powder.

AIRTM Particles. AIRTM particles (Alkermes, Inc., Cambridge, Massachusetts, U.S.A., Fig. 1) are prepared by spray drying solutions containing a mixture of drug and biodegradable material (e.g., polyglycolic acid, polylactic acid, or copolymers). AIRTM particles have a low tapped density of less than $\sim 0.4\ \text{g/cm}^3$, with an irregular surface. These low-density particles are physically large but aerodynamically small, which enhances the flow and dispersion of powders, and are ideal for deep lung delivery. Large particles could also minimize phagocytic uptake in alveoli, thereby providing the potential for controlled release of drugs in the lungs.^[27,28]

PulmosphereTM. PulmosphereTM (Inhale Therapeutic Systems, San Carlos, California, U.S.A.) particles are also prepared by spray drying, in which the feed

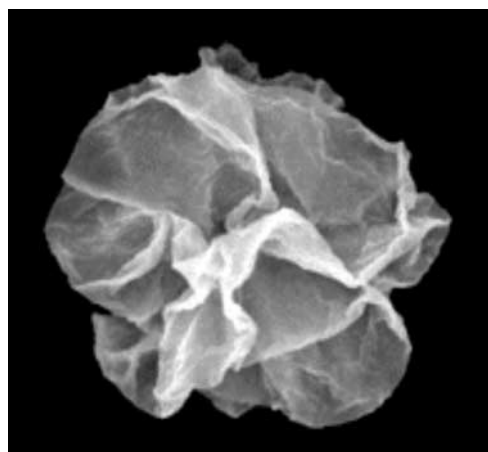


Fig. 1 AIRTM particle. (Courtesy of Alkermes, Inc., Cambridge, Massachusetts, U.S.A.)

comprises an aqueous solution containing dissolved or dispersed active drug, and a fluorocarbon-in-water emulsion, stabilized by a monolayer of phospholipid (such as dipalmitoylphosphatidylcholine).^[29] Being hollow and porous, Pulmosphere™ particles have a bulk density of 0.05 g/cm^3 – 0.2 g/cm^3 , and a size between $2 \mu\text{m}$ and $4 \mu\text{m}$. These particles are dispersed more readily without the need of carriers than the traditional drug–carrier blends with lactose.

Wrinkled Surface Particles. In contrast to the AIR™ particles and Pulmospheres™ which are hollow and porous, *solid* protein particles with wrinkled surfaces have recently been prepared in our laboratory. These wrinkled particles gave a significant improvement in FPF over spherical particles of bovine serum albumin.^[30] The reduction of the surface contacts among the wrinkle particles (Fig. 2) is believed to reduce powder cohesion. Dispersion of the wrinkle particles was shown to be less dependent on the inhaler choice and airflow compared with the smooth spherical particles of the same aerodynamic size.

Powder Inhaler Devices

Dry powder preparation and formulation are only part of the inhalation drug delivery system. Dispersion of these powders is closely linked to the performance of the inhaler device.

Recent inventions of the powder inhaler device are aimed at improving the inhaler's dispersion efficiency and reducing the resistance of the device as well as

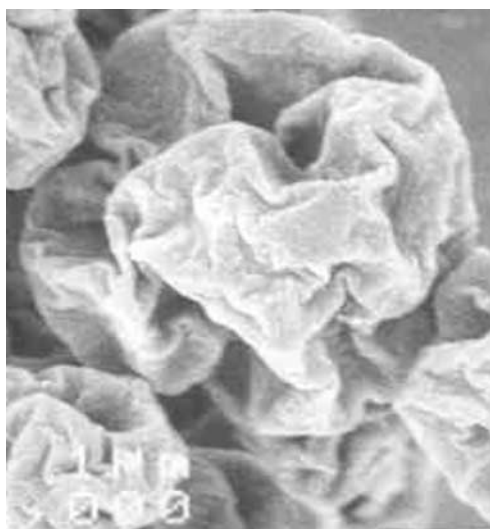


Fig. 2 Wrinkled bovine serum albumin particles, produced by spray drying.

decoupling powder dispersion from the patient's inspiratory effort in order to deliver accurate and flexible dosages for different patient's needs.

Innova™ and Solo™

The Innova™ (Inhale Therapeutic Systems, San Carlos, California, U.S.A., Fig. 3) inhaler device is a unit dose inhaler that has been designed for long-term use. It is powered by a stored bolus of compressed air and is designed to generate aerosol independent of patient's inspiratory effort. A transparent holding chamber enables patients to view

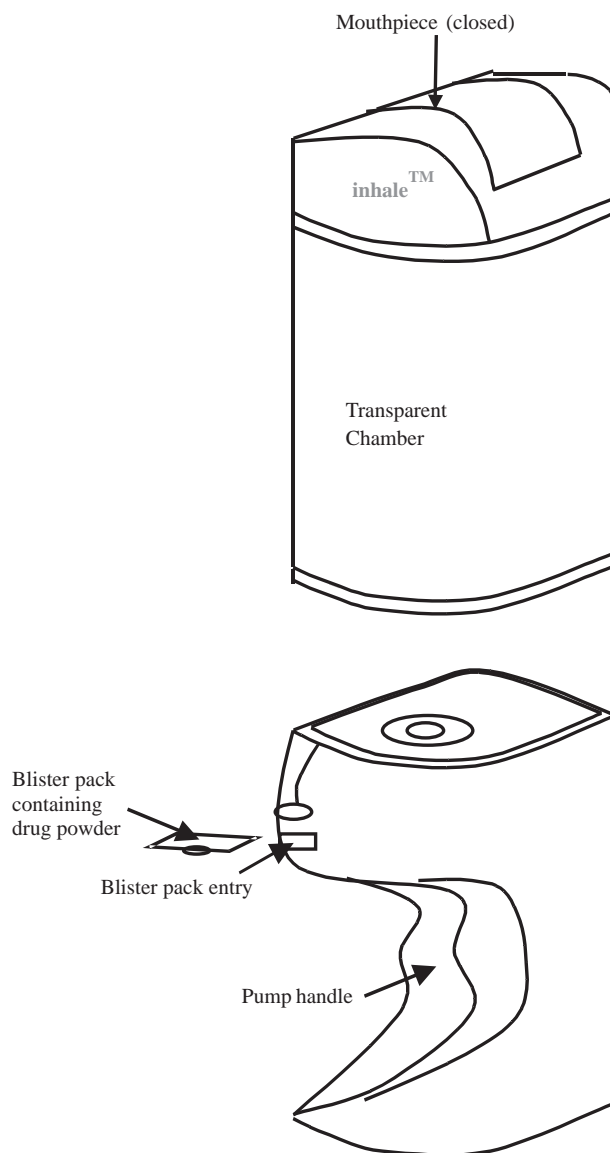


Fig. 3 Schematic diagram of Innova™ from Inhale Therapeutic Systems. (Adapted from Ref. 32.)

the aerosol to assure proper dosing. Further, the device is designed to have the capability to fluidize and extract up to 90% of the dose from the reservoir, thus minimizing waste and enhancing the accuracy and precision of the dosage.^[31]

The Solo™ device from Inhale Therapeutic Systems is a patient-driven unit dose inhaler. It has a built-in flow control to maximize the reproducibility of dose to patient. It is designed for short-term use and when large drug dosages are preferred.^[32]

SkyePharma mDPI

SkyePharma multidose pocket-sized inhaler features an “intelligent” single dose counter to count the dose dispensed and remained in the inhaler. The built-in locking mechanism also allows no “tailoff” effect toward the last doses from the inhaler^[33] (Fig. 4).

Bead inhaler technology

Elan Pharmaceuticals has developed the Spiros® S2 inhaler which features the use of beads to disperse the powders on inhalation. Still under development, the results from testing show that Spiros® S2 is a high dispersion efficiency inhaler capable of delivering drugs at relatively low inspiratory flow rates (30 L/min and 60 L/min).^[34]

Twisthaler®

The Twisthaler® (Schering-Plough, Kenilworth, New Jersey, U.S.A., Fig. 5) features specially designed nozzle



Fig. 4 SkyePharma mDPI. (Courtesy of SkyePharma, PLC, Switzerland.)

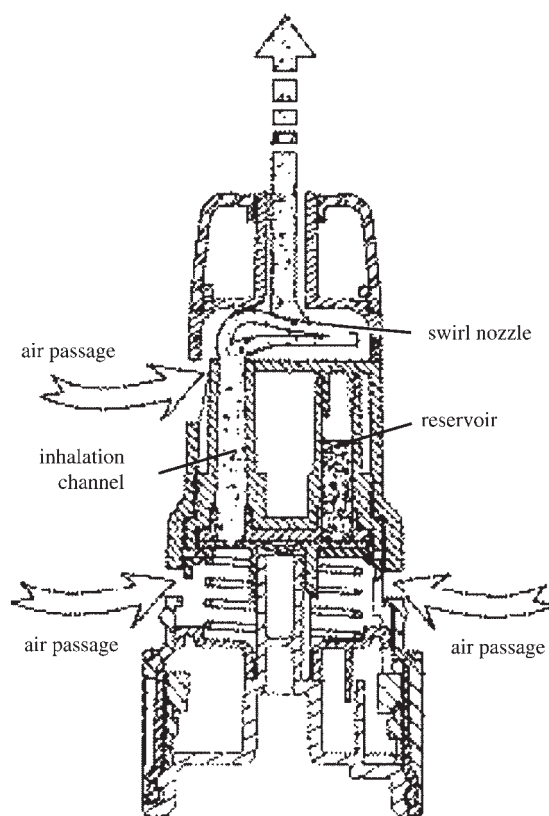


Fig. 5 Inhalation through the ASMANEX™ TWISTHALER™ DPI. (From Serentec Press, Raleigh, North Carolina.)

geometry that creates an airflow pattern which carries the small particles out of the device via the fluted chimney, while the larger particles or agglomerates will be spun into a centrifugal pattern and deagglomerated into fine particles for inhalation. The design optimizes the deagglomeration of powders, but at the same time minimizes drug caught in the inhaler nozzle and mouthpiece.^[35]

Hovine FlowCaps®

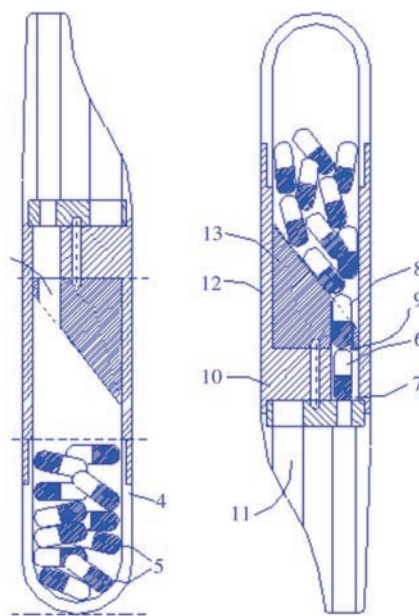
Hovine FlowCaps® (Hovione SA, Loures, Portugal, Fig. 6a and b) is a capsule–powder inhaler. Instead of the traditional needle piercing, the capsule is pierced by two blades, giving rise to a narrow slit across each end of the capsule. The tube-shaped inhaler receives the capsule(s) at one end.^[36,37] Upon inhalation, the air is mainly entrained into the inhaler tube inlets, with only a little air entering the capsules. The powder gets fluidized and experiences turbulence within the capsule before emptying from the device.

Random Loading

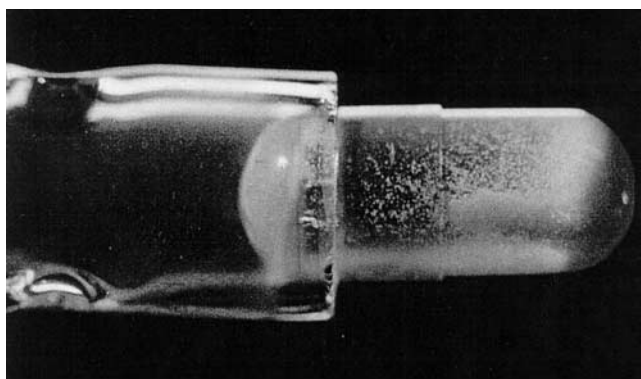
Capsules do not have to be inserted in a proper orientation, as in all other capsule based systems.

A patented ramp will automatically right a capsule for loading into the inhalation chamber. The patient can visually check that this is happening.

Ease of use is fundamental for patient compliance.



(a)



(b)

Fig. 6 (a) Hovione SA FlowCaps[®] dry powder device. (Courtesy of Hovione Produtos Farmaceuticos SA, Portugal.) (b) Photograph showing the dispersion of powders within the capsule in a Hovione SA FlowCaps dry powder inhaler. (Courtesy of Hovione Produtos Farmaceuticos SA, Portugal.)

REFERENCES

1. Patton, J.S.; Bukar, J.; Nagarajan, S. Inhaled Insulin. *Adv. Drug Deliv. Rev.* **1999**, *35*, 235–247.
2. Clark, A.; Foulds, G.H. Aerosolized active agent delivery. World IPO 9,947,196, 1999.
3. Platz, R.M.; Ip, A.; Whitham, C.L. Process for preparing micronized polypeptide drugs. US patent 5,354,562, 1994.
4. Platz, R.M.; Utsumi, J.; Satoh Y.; Naruse, N. Pharmaceutical aerosol formulation of solid formulation of solid polypeptide microparticles and method for the preparation thereof. World IPO 9,116,038, 1991.
5. Vidgren, M.T.; Vidgren, P.A.; Paronen, T.P. Comparison of physical and inhalation properties of spray dried and mechanically micronized disodium cromoglycate. *Int. J. Pharm.* **1997**, *35*, 139–144.
6. Masters, K. *Spray Drying Handbook*; John Wiley & Sons: New York, 1976.
7. Chan, H.-K.; Clark, A.; Gonda, I.; Mumenthaler, M.; Hsu, C. Spray dried powders and powder blends of recombinant human deoxyribonuclease (rhDNase) for aerosol delivery. *Pharm. Res.* **1997**, *14*, 431–497.
8. Clark, A.R.; Dasovich, N.; Gonda, I.; Chan, H.-K. The balance between biochemical and physical stability for inhalation protein powders: RhDNase as an Example. In *Respiratory Drug Delivery V*; Dalby, R.N., Byron, P.R., Farr, S.J., Eds.; Interpharm Press: Buffalo Grove, IL, 1996; 167–174.

9. Maa, Y.-F.; Nuygen, P.A.; Hsu, S.W. Spray-drying of air-liquid interface sensitive recombinant human growth hormone. *J. Pharm. Sci.* **1998**, *87*, 152–159.
10. Andya, J.D.; Maa, Y.F.; Constantino, H.R.; Nguyen, P.A.; Dosovich, N.; Sweeny, T.D.; Hsu, C.C.; Shire, S.J. The effect of formulation excipients on protein stability and aerosol performance of spray-dried powders of recombinant humanized anti-IgE monoclonal antibody. *Pharm. Res.* **1999**, *16*, 350–358.
11. Costantino, H.R.; Andya, J.D.; Nguyen, P.A.; Dasovich, N.; Sweeney, T.D.; Shire, S.J.; Hsu, C.C.; Maa, Y.F. Effect of mannitol crystallization on the stability and aerosol performance of a spray-dried pharmaceutical protein, recombinant humanized anti-IgE monoclonal antibody. *J. Pharm. Sci.* **1998**, *87*, 1406–1411.
12. Nagarajan, S.; Patton, J.S.; Bennet, D.B. Pulmonary administration of dry powder formulations for treating infertility. World IPO 0,061,178, 2000.
13. Maa, Y.F.; Nguyen, P.A.; Sweeney, T.; Shire, S.J.; Hsu, C.C. Protein inhalation powders: Spray drying vs. spray freeze drying. *Pharm. Res.* **1999**, *16*, 249–254.
14. Gordon, M.S.; Lord, J.D. Processes for spray drying solutions of hydrophobic drugs and compositions thereof. US Patent 5,985,248, 1999.
15. Dondeti, P.; Desai, Y. Supercritical fluid technology in pharmaceutical research. In *Encyclopaedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1998; Vol. 18, 219–248.
16. Winters, M.A.; Knutson, B.L.; Debenedetti, P.G.; Sparks, H.G.; Przybycien, T.M.; Stevenson, C.L.; Prestrelski, S.J. Precipitation of proteins in supercritical carbon dioxide. *J. Pharm. Sci.* **1996**, *85*, 586–594.
17. York, P.; Hanna, M. Particle engineering by supercritical fluid technologies for powder inhalation drug delivery. In *Respiratory Drug Delivery V*; Dalby, R.N., Byron, P.R., Farr, S.J., Eds.; Interpharm Press: Buffalo Grove, IL, 1996; 231–239.
18. Sievers, R.E.; Karst, U.; Milewski, P.D.; Sellers, S.P.; Miles, B.A.; Schaefer, J.D.; Stoldt, C.R.; Xu, C.Y. Formation of aqueous small droplet aerosols assisted by supercritical carbon dioxide. *Aerosol Sci. Technol.* **1999**, *30*, 3–15.
19. Cooper, S.M. Aerosol Formulations of Salmeterol Xinafoate. World IPO 9,955,319, 1999.
20. Sievers, R.E.; Sellers, S.P.; Clark, G.S.; Villa, J.A.; Mioskowski, B.; Carpenter, J. Supercritical fluid carbon dioxide technologies for fine particle formation for pulmonary drug delivery. *Book of Abstracts, 219th ACS National Meeting, San Francisco, CA, 2000*; MED-173 American Chemical Society: Washington, DC, 2000.
21. Sievers, R.E.; Milewski, P.D.; Sellers, S.P.; Miles, B.A.; Korte, B.J.; Kusek, K.D.; Clark, G.S.; Mioskowski, B.; Villa, J.A. Supercritical and near-critical carbon dioxide assisted low-temperature bubble drying. *Ind. Eng. Chem. Res.* **2000**, *39*, 4831–4836.
22. Bustami, R.T.; Chan, H.K.; Dehghani, E.; Foster, N.R. Generation of microparticles of proteins for aerosol delivery using high pressure modified carbon dioxide. *Pharm. Res.* **2000**, *17*, 1360–1366.
23. Lucas, P.; Anderson, K.; Staniforth, J.N. Protein deposition from dry powder inhalers: Fine particle multiplets as performance modifiers. *Pharm. Res.* **1998**, *15*, 562–569.
24. Musa R.; Ventura P.; Chiesi, P. Improved powdery pharmaceutical compositions for inhalation. World IPO 0,053,157, 2000.
25. Ganderton, D.; Morton, D.A.; Lucas, P. Improvements in or relating to powders. World IPO 0,033,811, 2000.
26. Johnson, K.A.; Gordon, M.S.; Lyons, S.W. Methods and system for processing dispersible fine powders. US Patent 6,103,270, 2000.
27. Rudt, S.; Muller, R.H. Effect of analytical parameters, particle size and particle concentration. *J. Control. Rel.* **1992**, *22*, 263–271.
28. Kawaguchi, H.; Koiwai, N.; Ohtsuka, Y.; Miyamoto, M.; Sasakawa, S. Phagocytosis of latex particles by leucocytes. I. Dependence of phagocytosis on the size and surface potential of particles. *Biomaterials* **1996**, *7*, 61–66.
29. Tarara, T.; Weers, J.; Dellamary, L. Engineered powders for inhalation. In *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Serentec Press: Raleigh, NC, 2000; Vol. 2, 413–416.
30. Chew, N.Y.K.; Chan, H.K. Use of solid corrugated particles to enhance powder aerosol performance. *Pharm. Res.* **2001**, *18*, 1570–1577.
31. Smith, A.E.; Burr, J.D.; Etter, J.W.; Axford, G.S.; Anthony, J.M. Apparatus and methods for dispersing dry powder medicaments. US Patent 5,740,794, 1998.
32. Inhale, Family of Inhalers. 2001 Inhale Therapeutic Systems Inc. (Access 1 August, 2001). http://www.inhale.com/webpage_templates/tech_frame.php3?page_name=delivery.
33. Skyepharma Technology: Inhalation. Skyepharma PLC [Access 1 August, 2001] <http://www.skyepharma.com/inhalation.html>.
34. Witham, C. In *Bead Inhaler Technology*. Abstracts of Papers, Aerosol in Medicine, Brisbane, Qld, Australia, March 19–20, 2001; S13.
35. Fan, B.J.; Yang, T.T.; Kenyon, D. Application of computer modeling in the design and development of the new mometasone furoate dry powder inhaler (MF-DPI) nozzle. In *Respiratory Drug Delivery VI*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J., Eds.; Serentec Press: Raleigh, NC, 2000; Vol. 2, 491–493.
36. Villax, P.; Peres, R.; Treneman, W.R.; McDerment I.G.; Bunce, M. Medicament inhaler and method. US Patent 5,673,686, 1997.
37. Bunce, M.; Lamb, J.; Greeves, M.; McDerment, I.G.; Villax, P. Apparatus for orienting and positioning an elongate object for dispensing. US Patent 5,881,721, 1999.

Drug Master Files

C. Jeanne Taborsky
Brian James Reamer

SciRegs Consulting, Columbia, Maryland, U.S.A.

INTRODUCTION

In most cases, in order to market a new drug or generic drug in the United States, the manufacturer must file a drug application with the Food and Drug Administration (FDA). The FDA must review and approve the drug application. In support of the application, the manufacturer can reference the information that is filed with the FDA in a Drug Master File (DMF). This information is confidential and the technical contents of a DMF are (usually) only reviewed when authorized by the DMF holder and reviewed only in connection with the review of an Investigational New Drug (IND), New Drug Application (NDA), Abbreviated New Drug Application (ANDA), or an Export Application. A company that intends to conduct a clinical investigation on an approved drug must file an IND. A drug company that wants to market a new drug or a previously approved drug in a new container or a new formula must file a NDA. A company that wants to market a generic drug must file a ANDA. A company that wants to export a drug that is not approved for marketing in the United States must file an Export Application submitted under Section 802 of the Federal Food, Drug, and Cosmetic Act. The DMF is a critical part of the review process and is in increasing use. For this reason, many drug applicants will not use a supplier who does not have a Drug Master File. Drug Master Files are not accepted or rejected, but are rather found to be satisfactory or deficient in support of the drug application in which they are provided. Marketing a pre-1938 drug or a drug for which an Over The Counter (OTC) Monograph status has been granted does not require the filing of an application with the FDA.

FDA—ORGANIZATION

The FDA is divided into different centers, each with their own set of regulations and areas of responsibility.^[1] These centers and their areas of responsibility are as follows:

- CSFAN—The Center for Food Safety and Nutrition: Responsible for regulating foods, dietary supplements, and cosmetics.

- CBER—The Center for Biological Evaluation and Research: Responsible for regulating biologics, blood.
- CDRH—The Center for Devices and Radiological Health: Responsible for regulating medical devices, kits, and diagnostic solutions.
- CVM—The Center for Veterinary Medicine: Responsible for regulating animal medicine.
- CDER—The Center for Drug Evaluation and Research: Responsible for regulating drugs.

Drug applications filed with the CDER can be for INDs, NDAs, and ANDAs, also known as generics.^[2] The CDER, the CBER, and the CVM share access to the same Drug Master Files.

The CDRH does not use DMFs submitted to the CDER, and submissions to the CDRH are not reviewed or used to support drug applications to the CDER. This aspect of the drug review and approval process is important because there is an overlap of responsibility in the use of some materials and components regulated by more than one center. A syringe that is sold on its own merit, without an associated drug, is regulated by the CDRH, while a syringe containing a drug is reviewed by the CDER. There is an internal FDA intercenter agreement that details the divisions of authority. The CDRH would only comment on a component if requested by the CDER reviewer. Therefore filing information to the CDRH on a syringe does not obviate the need to file a DMF to the CDER. The CDER has defined the information that they expect to review in support of a drug application, and the information should appear in the drug submission or a referenced DMF.

DRUG MASTER FILES

A Drug Master File (DMF) is a submission of information to the FDA by a person or a firm.^[3] A DMF is filed with the intent to permit a drug applicant to use (incorporate by reference) the information contained in the DMF in support of a drug application or a supplemental change to an approved application without having to disclose the information to the drug applicant. Drug Master Files were originally used as a way for suppliers to file limited confidential or proprietary information with the FDA in sup-



port of the drug application, while keeping the information confidential from the end user and the competition. In the United States, confidentiality is still an important aspect of the process. In Europe, the relationship between the DMF holder and applicant are somewhat different: It is not uncommon for the DMF holder to provide a copy of the Drug Master File to the drug applicant or customer.

Although firms have the option of filing their own DMF or providing the information to an applicant to file in their drug application, there are many disadvantages to having the drug applicant file the information. The holder of the intellectual property would have to confide the information to the applicant, thereby losing control of it. The term "control" does not refer only to confidentiality of the information, but also to document control: The tracking and update of changes. For multiple drug applications, the information is repeatedly reviewed for each drug submission. Different reviewers may have different questions based on their areas of expertise or the end use of the material. Therefore the information may be rereviewed with questions and responses repeatedly filed to each individual drug application. Filing DMF-type information in the NDA or ANDA significantly increases the size of the drug submission, and increases the workload for the FDA reviewer. The information must be well controlled and updated with all drug applicants whenever a change occurs. With a DMF, the information is reviewed once, and if found satisfactory, may not be reviewed for 2 years, unless the end user is changed or the information is amended. A record of the satisfactory review is maintained in the FDA computer database and filed in the Drug Master File. The DMF holder does not receive a copy of the review and is only contacted by the FDA if the DMF is found to be deficient.

DRUG MASTER FILE PROCESS

Two copies of the Drug Master File are mailed to the Drug Master Files Staff, located at 12229 Wilkins Avenue, Rockville, Maryland 20852.^[4] The Drug Master File staff will audit the nontechnical information for completeness and adequacy for submission. If the key elements are missing, the staff will contact the proposed holder to try to obtain the necessary documents in order to file the DMF. Once the DMFs are determined to be acceptable for filing, the document room staff assigns a DMF number and a letter is sent to the contact person listed in the DMF.

DRUG MASTER FILE FORMAT

The DMF must meet the format requirements.^[5] The DMF is submitted as Original and Duplicate jackets,

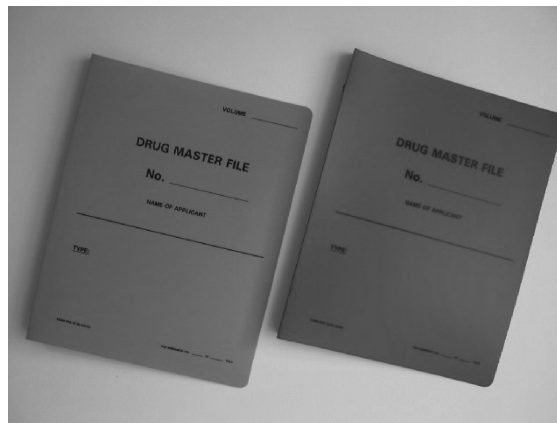


Fig. 1 DMF jackets.

collated, assembled, paginated, and jacketed, using covers obtained from the government printing office. Multiple volumes are numbered, and the paper must be standard U.S. size paper. It is not unusual to receive paper from England or India that is of different size or weight. Because the pages are placed in binders, the left margin should be no smaller than 3/4 in. and the right margin should be no smaller than 1/2 in. The DMF must be submitted in two copies, one with a blue cover and one with a red cover. The jacket covers are purchased from the government printing office and are specifically provided for the DMFs (Fig. 1).

REQUIREMENTS

The regulations for Drug Master Files are listed in the *Code of Federal Regulations* Title 21 § 314.420. In addition, the CDER has issued guidances to better describe the format and provide some suggestions on content. These guidances can be found on the CDER guidances web page www.fda.gov.

A DMF provides (incorporates) information by reference and permits the "holder" to "authorize" other "persons" to disclose information in support of an application. Usually, a DMF may only be accessed (reviewed) by the FDA if the holder of the drug application provides the FDA with authorization (by incorporation), in writing, with two copies, dated, with the DMF number, name of holder, name of material, specific product, reference number/volume, page number, name of authorized persons, and a statement of commitment that the DMF is current. The letters filed to the DMF must have the original signature, typed name, and title of the authorizing person on the letterhead. A copy of each letter is provided to the drug applicant to file in their submission.

The applicant is not required to acquire a new letter for each submission. The DMF holder usually does not know for which drug an applicant has filed their DMF letter. Regulations governing DMFs are very general. For all DMFs, the following must be provided: The names and addresses of the DMF holder, the corporate headquarters, the manufacturing/processing facility, as well as the name, address, phone number, and fax number for the company contact for FDA correspondence, and the contact information for agent(s), if any. The specific responsibilities of each person listed in any of the above categories should be listed. Including an organization chart with job titles helps to communicate this information. A statement of commitment is required, as well as a signed statement by the holder certifying that the DMF is current, and that the DMF holder will comply with the statements made in the DMF. The commitment letter, a letter appointing the U.S. agent, and any debarment or current Good Manufacturing Practices (cGMPs) statements should all be provided on separate pages.

When a U.S. agent is “appointed,” the DMF holder should submit two copies of a letter to the DMF giving the agent’s name, address, and scope of responsibility (administrative and/or scientific). Drug Master File holders are not required to appoint an agent, but foreign DMF holders are encouraged to do so. The holder or agent is required to submit archival and duplicate copies and notify each person authorized to incorporate information of changes, additions, or deletions. It should go without saying that the holder must conform to the procedures listed in the DMF.

The DMF should include the name, address, phone number of holder, a description of items that are the subject of the DMF, a list of materials of construction, and the sources of materials of construction. The standards for testing incoming, in process, and release are sometimes different. There are five types of CDER Drug Master Files listed in the *Code of Federal Regulations* Title 21 §314 Drug Master Files (DMF). They are as follows:

- Type 1—Reserved (no longer used; once used for facilities).
- Type 2—Drug substances, intermediates, and materials used in their preparation of drug products.
- Type 3—Packaging materials (resins, compounds, colorants, inks, components).
- Type 4—Excipients (colorants, flavors, and raw materials).
- Type 5: FDA-accepted reference information.

Type 1 Drug Master Files are no longer used. They were once used to describe facilities. In an effort to better define the relationship between the field investigators and the center reviewers, the Center/Field Agreement was written, transferring the responsibility for review

of facility information to the field investigator. Facility information on companies that manufacture active ingredients, packaging materials, and raw materials was transferred to the respective Drug Master File.

Type 2 Drug Master Files are required for drug substances, also referred to as active ingredients or active pharmaceutical ingredients. A separate DMF is filed for each active ingredient. The DMF must include a brief description of the facility, the address, a contact, phone number, and fax number. The manufacturing facility must be registered and the number should be listed. A U.S. agent is required to list the active ingredient and register the firm if the site is located outside of the United States. The address should be the same as the address used on the FDA registration form. While a U.S. agent is required for registration, an agent is not required to file a DMF at this time.

Type 2 Drug Master Files should contain a flow diagram of the manufacturing or synthetic process (if it is synthetic) including a list of the critical steps, in-processing tests conducted at those steps, and sampling protocol. The testing of all raw materials should be provided as well as the in-process controls, packaging, release, and stability testing. The impurity profile, particle-size distribution, organic volatile impurities, and residual solvent test results have become increasingly important. These tests should reference the current standards of the United States Pharmacopeia (USP) and the International Conference on Harmonization (ICH) Guidelines. These test results are often compared with those obtained by the end user to validate the results when reviewed by the FDA. If the values are not in agreement, it is a signal that the DMF holder may not be in compliance with the DMF. The testing for polymorphism is increasingly important, and many times is an issue for drug absorption and stability and increasingly in patent litigation. The test methods should be validated if they are not USP. CDER reviewers have standardized the information requirements for Type 2 Drug Master Files through their deficiency letters and guidances.

Guidances for residual solvents, methods validation, changes to the synthetic route, and analytical methods should all be cited in the DMF and used to generate the documentation. As these guidances and test methods evolve, the DMF should be amended to provide for changes and be updated to incorporate changes in USP/NF or FDA standards. The establishment of standards for impurities and residual solvents should not be solely based on the ICH/CDER guidances but should also be established at a low-enough level to demonstrate the control of a process.

Changes in the synthetic route, process, test method, and the loosening of specifications can all have a major impact on the drug application, in which the Drug Master File is referenced. These changes must be classified to



minor or major changes, and a determination must be made as to how these changes are to be reported by the drug company. Changes may require the drug company to manufacture a new test batch with the revised procedure and file a supplement to their approved application before releasing the product to the market. It is in the best interest of the DMF holder to discuss the changes with the end user well in advance of the implementation of the change. Failure to file a change in the DMF, or to notify the end user, can result in regulatory action against the DMF holder. The FDA can also place an embargo on a foreign firm and prevent the import of the active ingredient. The FDA can also "cancel" a DMF if the DMF holder does not follow the requirements stated in the Code of Federal Regulations.

The information submitted in Type 3 DMFs is more variable because the articles can be very different. Most Type 3 DMF holders are ISO (International Organization for Standardization) certified, but the rules for ISO and those of CDER are not always the same. There is an increase in the use of Type 3 Drug Master Files for both contract packaging companies and firms that manufacture either components or parts of components. There is a confusion over what is to be filed in packaging DMFs. A Type 3 DMF can contain information on a number of different packaging components. Typically, separate DMFs are filed for closures, plastic bottles, caps, droppers, stoppers, etc. Each packaging material should be identified by the intended use, components, composition, and controls for its release. The names of the suppliers or fabricators of the components used in preparing the packaging material and the acceptance specifications should also be provided. Data supporting the acceptability of the packaging material for its intended use should also be submitted in conformance with the agency's guidances. Toxicological data can be included where appropriate. Components manufactured at alternate sites can still be filed in the same DMF. For example, bottle manufactured at two different plants in the United States can both be in the same DMF. Blister film manufactured in the United States and overseas can be listed in the same DMF. The addresses for both manufacturing facilities would be provided in the facility section of the DMF.

Manufacturers of resins, colorants, inks, compounds, parts, molded components, container systems, and packaging operations all use Type 3 DMFs. Depending on the subject of the DMF, it should contain the following: A short facility description, the formulation providing the trade name, the generic name, *Code of Federal Regulations* Title 21 indirect food additive reference or Food Contact Number, Chemical Abstract number, and references to other supporting DMFs, if applicable. The DMF should list the quality, quantity, and purpose of each

chemical in the component compound. A brief description of the manufacturing process, drawings (if applicable), and Certificate of Conformance, with critical-to-function or fit testing should also be included. Test methods should be based on public standards such as ASTM (American Society for Testing and Materials) or USP^[3,4] to include physiochemical qualification tests, identity tests, and extractions (USP or other). Firms should specify any additional treatments (such as washing). It is also recommended to include a brief description of packaging and labeling procedures and include samples of labels.

Type 3 DMF holders must be aware of the need to meet additional standards. Certain states and organizations such as the Congress of North East Governors (CONEG), and the states of California and New Jersey have local regulations that must be considered. There are limits for heavy metals, lead, hexavalent chromium, mercury, and cadmium. Mad Cow Certification is required if stearates are used. Testing for Organic Volatile Impurities and residual solvents maybe required. In addition, because most firms ship overseas, international rules may apply as well. Drug Master File holders should follow the FDA regulations and guidances, stay informed, maintain documentation, follow applicable cGMPs, follow DMF practices, and keep customers informed.

The needs of the drug applicant must also be considered. The DMF provide all of the required supporting information on the components that is not contained in the drug submission. Standards and requirements are contained in the USP,^[5] FDA, CDER guidances, and ICH/CDER documents. These documents cover most of the testing needed to assure that the components should not adversely impact the drug and render it unacceptable for use. The lower the drug product risk, the more standard the component and its materials, and the easier it is to provide adequate information and meet drug-applicant needs. Standards for high-density polyethylene bottles for a solid oral dosage form are published in the USP <661> and soon to be revised <671> chapters. Solid oral dosage forms are lower-risk products. The standards are understood and have been in place since the 1970s when the standards in the USP chapters were established. Reviewers and manufacturers are well aware of the resins and the properties requirements for this end use.

Where plastic or elastomeric material is used for high-risk products such as inhalation products, injectable drugs, or implants, the risk is greater and the rules are less clear. The FDA provides guidance in this area, but there is a limit as to what help they can provide prior to reviewing the drug product and component information. The drug applicant is placed in a position of trying to test for extractables from the component without knowing what they are, and the DMF holder is trying to

meet the needs of the customer without compromising their product(s).

The end user does not usually inform the DMF holder as to the product for which the component will be used. It is not unusual for the drug applicant to request a sample supply and a letter of authorization. The DMF holder may be unaware of the status of their DMF. The end user would benefit by telling the DMF holder what the end use of the material or components would be. Through this, the DMF holder can better advise the end user on the choice of components and the degree of testing that they themselves have conducted. The applicant can provide a list of criteria to the DMF holder in advance to be certain the article they purchase meets the product needs. An alternative is to use a confidential third party, to work with both the DMF holder and the drug applicant. With a confidentially agreement signed by both parties, there can be consistency in the information in the DMF. A flow of information from the DMF to the container section of the drug application makes it easier for the FDA to review.

Type 4 Drug Master Files are seldom filed and are not required for raw materials. Most raw materials meet USP/NF standards and need only to meet those criteria for acceptance and certain physical tests. Type 4 DMFs are usually used for flavors and cosmetic coatings for tablets for which USP/NF standards do not exist. Coatings, whether cosmetic or controlled release, may be proprietary in nature. The DMF process allows for these articles to be reviewed, again, under confidentiality. Coatings are usually multiple ingredients. The raw material qualifications, formulation, manufacturing process, and release testing should be provided and iron oxide constituents should be quantitated. It is advisable to include the calculation for iron content for these materials in the DMF.

All DMFs require continuing support. The DMF holder is required to file annual reports and inform the drug applicant of changes to the DMF. The DMF holder must notify each firm authorized to reference the DMF if the DMF holder adds, deletes, or changes information in the DMF. If the holder changes any process or document listed in the DMF, then the DMF must also be changed.

As changes occur, both parties must have a clear understanding of the impact of the change, the method of reporting the change, and the time for the FDA to review the change, if necessary. Changes should be classified based on the cGMPs and the CDER guidances. If the DMF holder is uncertain about the classification of a change or the way in which the drug application should report the changes, it is strongly recommended that the DMF holder contact a regulatory consultant or the agency to discuss the change before it is made. Sufficient information should be provided to the agency in order for the FDA to render a meaningful decision. On the anniversary date of the filing, the holder is to update the

DMF to include changes in the authorization list, manufacturing procedures, designs, suppliers, testing, letters of authorization to other DMFs, and provide a complete updated list of people to whom letters of authorization have been issued. Failure to update these items can result in delays in drug approvals. The FDA agencies can also close the DMF based on failure to provide an annual list of persons authorized to incorporate information, failure to provide a list of changes, and failure to provide a statement that the DMF is current.

Based on its experience or newly provided information, the FDA is constantly updating its practices and issuing revised policies and guidances. It is important for both DMF holders and drug applicants to stay vigilant. Documentation and practices should be reviewed on a routine basis and updated to incorporate changes in the regulations, practices, and guidances.

CONCLUSION

The Drug Master File is a critical document used to support a drug application. Deficiencies in the Drug Master File can result in the delay of approval of drug applications. It is important that the DMF be filed in a timely manner and that the standards used to compete it are of the same quality as the actual drug application. The DMF can be considered an extension of the drug application. The drug review process works best when the required information flows from the DMFs to the drug application.

REFERENCES

1. *Food, Drug, and Related Law as Amended Through December 31, 2000, Prepared for the use of the Committee on Energy and Commerce, U.S. House of Representatives*; U.S. Government Printing Office: Washington, DC, 2001.
2. *Code of Federal Regulations Title 21 § Parts 300 to 369, Parts 300 to 499, Office of Federal Register, National Archives and Records Administration, (current version)*; Government Printing Office: Washington, DC, 2001.
3. *Code of Federal Regulations Title 21 § 314. 420, Drug Master Files, Parts 300 to 499, Office of Federal Register, National Archives and Records Administration: (current version)*; Government Printing Office: Washington, DC, 2001.
4. GUIDELINE FOR DRUG MASTER FILES, Center for Drug Evaluation and Research, Food and Drug Administration, Department of Health and Human Services, Office of Drug Evaluation I (HFD-100), 5600 Fishers Lane, Rockville, Maryland 20857, September 1989.
5. *USP, United States Pharmacopeial Convention, Inc. 12601*; Twinbrook Parkway: Rockville, MD, 20852.

DRYING AND DRYERS

Anthony J. Hlinak

Bradley A. Clark

Pharmacia, Skokie, Illinois, U.S.A.

INTRODUCTION

Many pharmaceutical operations, including those used to produce active pharmaceutical ingredients and excipients, use water or organic solvents as essential processing aids. However, the continued presence of these processing aids may hamper downstream manufacturing operations or compromise the safety and stability of the final pharmaceutical product. Drying is a common unit operation used to reduce the levels of water or organic solvent in pharmaceutical materials to acceptable levels. Drying requires the use of a manufacturing firm's scarce resources: facilities, equipment, utilities, human labor, and time. Thus, there is often economic pressure to select aggressive drying conditions in an effort to minimize cycle times and increase throughput. However, the drying rate can impact the properties, functionality, and quality of the material being dried. Thus, the practitioner must balance the economic pressure with the quality and performance needs of the product.

In general terms, drying can be described by three processes operating simultaneously. The first process is energy transfer from an external source to the water or organic solvent in the material. The second process is the phase transformation of the water or organic solvent from a liquid or liquid-like state to a vapor state. The third process is the transfer of the vapor generated away from the pharmaceutical material and out of the drying equipment. Analysis of the drying process is complicated by the fact these three processes are coupled to each other, and all three need to be considered simultaneously.

THEORY

Heat Transfer

For most equipment designs, the energy transferred is in the form of heat. Heat flows naturally in the direction of decreasing temperature and is known to transfer by three distinct mechanisms. The first, conduction, involves purely molecular scale transfers of kinetic energy and

can be modeled in one dimension in the steady state using Fourier's equation of heat conduction (1):

$$q_k = -kA \frac{dT}{dx} \quad (1)$$

where q_k is the rate of energy transferred by the conductive mechanism (calories/s), k the thermal conductivity of the transferring medium (calories/s/cm/°C), A the transfer area (cm²), and

$$\frac{dT}{dx}$$

is the temperature gradient in the direction of transfer (°C/cm).

For the simple case of steady one-dimensional heat flow through a homogeneous medium with parallel, planar boundaries held at fixed temperatures, Eq. 1 can be integrated to yield

$$q_k = kA \frac{(T_s - T_i)}{\ell} \quad (2)$$

where T_s, T_i are the exposed surface and interface temperatures, respectively (°C) and ℓ is the thickness of the material layer (cm).

Although derived assuming the simplest of conditions, Eq. 2 can be used to analyze a variety of drying situations as illustrated in Fig. 1, including cases involving modest curvature, as a reasonable approximation. Solutions for more complex cases, including nonplanar geometries, nonsteady state, and complex boundary conditions can be found in classic texts on the subject (2).

A useful concept, known as thermal resistance, can be derived from Eq. 2 by a simple redefinition of terms:

$$q_k = \frac{(T_s - T_i)}{R_k} \quad (3)$$

Here

$$R_k = \frac{\ell}{kA}$$

is the thermal resistance. Eq. 3 takes the functional form of Ohm's law in electrical circuit theory, with

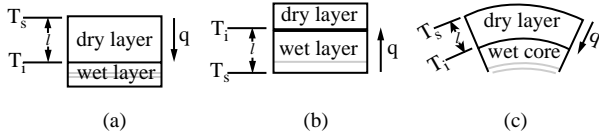


Fig. 1 Application of Eq. 2 to (a) direct heating of a static solids bed, (b) indirect heating of a static solids bed, and (c) fluid bed-drying of a spherical particle.

the temperature difference playing the role of voltage difference, energy flow in the role of current flow, and thermal resistance in the role of electrical resistance. This form emphasizes the importance of the temperature difference as the driving force behind thermal transport and permits the physical factors that make up the thermal resistance to be broken down in detail. Because of this analogy with Ohm's law, thermal "circuits" can be constructed to handle complex cases involving simultaneous transfer with multiple heat transfer mechanisms or sources and the techniques developed for electrical circuits can be used to simplify the analysis.

In the second heat transfer mechanism, convection, molecular scale transfers of kinetic energy are augmented by the macroscopic movement of a fluid transfer medium. Convection is most important as the mechanism of transfer between the solid surface of a static bed or an individual suspended particle and the gaseous medium that surrounds it. Convection has been modeled classically using the following relationship generally attributed to Isaac Newton:

$$q_c = \bar{h}_c A (T_\infty - T_s) \quad (4)$$

where q_c is the rate of energy transferred attributable to convection (calories/s), T_s the exposed solid surface temperature ($^{\circ}\text{C}$), T_∞ the fluid temperature far away from the exposed solid surface ($^{\circ}\text{C}$), A the transfer area (cm^2), and \bar{h}_c is the average convective heat transfer coefficient (calories/s cm^2 $^{\circ}\text{C}$).

The apparent simplicity of Eq. 4 can be misleading, because the convective heat transfer coefficient is actually a very complex function of fluid flow conditions, fluid properties, and system geometry. In addition, the fluid flow patterns are significantly different depending on whether or not the fluid motion is induced by buoyancy forces alone (termed natural or free convection), or generated by external mechanical means using pumps or blowers (termed forced convection). When a heated fluid passes over a solid surface, the regions of significant velocity and temperature change are generally restricted to a small layer in the immediate vicinity of the solid surface. This boundary layer may consist entirely of fluid moving in the

laminar flow regime, where transport of both momentum and energy rely solely on molecular interactions. More generally, the boundary layer consists of both a laminar sublayer immediately adjacent to the solid surface and a turbulent region.

Classical techniques have relied heavily on dimensional analysis (3), the combining of the many variables into physically meaningful nondimensional groups, supported with experiments to quantify heat transfer for various geometries. For most drying applications of pharmaceutical relevance, the most important of these nondimensional groups are the Nusselt number (\overline{Nu}), the Prandtl number (Pr) and the Reynolds number (Re), defined as follows:

$$\overline{Nu} = \frac{hL}{k_f} \quad (5a)$$

$$Pr = \frac{c_p \mu}{k_f} \quad (5b)$$

$$Re = \frac{VL\rho}{\mu} \quad (5c)$$

Here, k_f , c_p , ρ , and μ are, respectively, the thermal conductivity, specific heat at constant pressure, density, and dynamic viscosity of the convective fluid; V is the relative velocity between fluid and solid; and L is a geometry dependent, characteristic length dimension for the system. Note that the Pr is composed exclusively of fluid properties and that the Re will increase in direct proportion to the relative velocity between fluid and solid surface. Example applications are shown in Fig. 2.

Knowledge of \overline{Nu} , fluid phase thermal conductivity, and characteristic length allows computation of the average convective heat transfer coefficient, using Eq. 5a. For flat surfaces, like the surfaces of static beds exposed to air or other gases ($Pr \approx 0.7$), Kreith (4) provides the following for low-velocity, laminar flow conditions

$$\begin{aligned} \overline{Nu} &= 0.664 Re_L^{0.5} Pr^{0.33} \\ \text{for } Pr &> 0.1 \\ \text{and } Re_L &< 5 \times 10^5 \end{aligned} \quad (6a)$$

and for high-velocity, turbulent conditions

$$\begin{aligned} \overline{Nu} &= 0.036 Pr^{0.33} [Re_L^{0.8} - 23, 200] \\ \text{for } Pr &> 0.5 \\ \text{and } Re_L &> 5 \times 10^5 \end{aligned} \quad (6b)$$

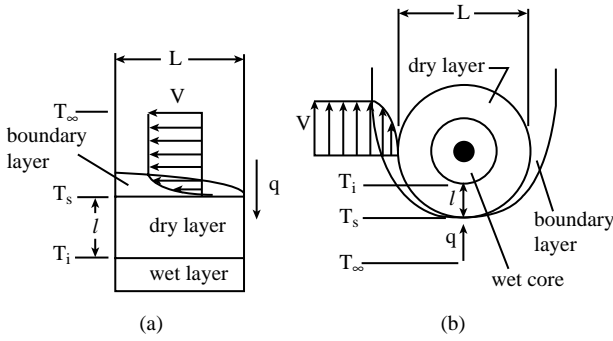


Fig. 2 Convective heat transfer applications in pharmaceutical drying: (a) tray-drying of a static solids bed; and (b) fluid bed-drying of a spherical particle.

For a spherical particle moving in an air or other gaseous stream, Whitaker (5) recommends the following relationship:

$$\overline{Nu} = 2 + (0.4 Re_L^{0.5} + 0.06 Re_L^{0.67}) Pr^{0.4} (\mu_s / \mu_\infty)^{0.25} \quad (6c)$$

for $3.5 < Re_L < 76,000$

where μ_s and μ_∞ are the dynamic viscosities of the gas at the temperature of the particle surface and at the temperature far away from the surface, respectively. In the limiting case of $Re_L \ll 1$, Johnston et al. (6) have shown that the \overline{Nu} approaches the constant value of 2, using assumptions approximating spherical particles in gas streams.

Equipment designs based on indirect conduction usually transfer the heat from the primary heat transfer fluid to the intermediate wall within some kind of internal duct or channel. Transfer coefficients for these cases depend on the nature of the flow (laminar or turbulent) and the geometry of the duct or channel (short or long). Expressions for evaluating the transfer coefficients for these cases are available in standard texts (7).

An expression for the convective thermal resistance can be generated similar to that derived for the conductive resistance:

$$q_c = \frac{(T_\infty - T_s)}{R_c} \quad (7)$$

Here, $R_c = \frac{1}{h_c A}$, is the convective thermal resistance.

The third mechanism of heat transfer is thermal radiation that can be defined as radiant energy emitted by a medium by virtue of its temperature. The wavelengths of thermal radiation produced by emitting bodies fall roughly between 0.1 and 100 μm , which includes portions of the ultraviolet, visible, and infrared spectra. The net exchange of radiant thermal energy between two surfaces can be characterized by the following relationship

$$q_r = \sigma A_1 \mathfrak{J}_{1-2} (T_1^4 - T_2^4) \quad (8)$$

where q_r is the rate of energy transferred attributable to thermal radiation (calories/s), T_1 the absolute temperature of radiating surface 1 (K), T_2 the absolute temperature of radiating surface 2 (K), σ the Stefan-Boltzmann constant (1.35×10^{-12} cal/s $\text{cm}^2 \text{K}^4$), A_1 the transfer area of surface 1 (cm^2), and \mathfrak{J}_{1-2} is a dimensionless factor that corrects for the radiative properties and relative geometries of the surfaces involved in the exchange.

Most of the complexity of radiative heat transfer analysis is thus condensed into evaluation of the dimensionless factor \mathfrak{J}_{1-2} . This factor is a function of both surface properties and the geometric orientation of the surfaces involved in the exchange. For real surfaces the amount of thermal radiation emitted and absorbed depends on the temperature, the wavelength, and the angular direction. These complications are often neglected and the radiative properties of the surface are lumped together into a dimensionless factor that is independent of both wavelength and direction, referred to as emissivity (ϵ). The emissivity expresses the radiative power of a surface as some fraction of an ideal radiator or blackbody. Real surfaces so treated are referred to as greybodies to emphasize this simplification imposed. For exchanges between parallel rectangular surfaces, where the spacing between the surfaces is small compared with the smaller dimension of the rectangles, the factor \mathfrak{J}_{1-2} can be estimated as

$$\mathfrak{J}_{1-2} = \frac{1}{\frac{1}{\epsilon_1} + \frac{1}{\epsilon_2} - 1} \quad (9)$$

where, ϵ_1 , ϵ_2 are the emissivities of the surfaces involved in the exchange. For a small spherical particle inside a large enclosure, the factor \mathfrak{J}_{1-2} can be estimated as

$$\mathfrak{J}_{1-2} = \epsilon_1 \quad (10)$$

where ϵ_1 is the emissivity of the spherical particle. For more rigorous treatments, the reader is advised to consult advanced texts (8). Eq. 8 can be used to generate an expression for the thermal radiative resistance similar to that derived for the conductive and convective resistance:

$$q_r = \frac{(T_\infty - T_s)}{R_r} \quad (11)$$

Here the thermal radiative resistance must assume a more complex form

$$R_r = \frac{(T_\infty - T_s)}{\sigma A_s \mathfrak{J}_{s-2} (T_s^4 - T_2^4)} \quad (12)$$

with the subscripts s and 2 used to denote the product surface and external radiating surface, respectively.

Unfortunately, the resistance defined by Eq. 12 cannot be evaluated without a priori knowledge of temperatures, unlike those defined previously for conduction and convection. However, enough information on temperatures is often available from previous drying experience to permit useful estimates of the radiative resistance to be established.

Application of the Ohm's law analogy allows construction of combined series parallel thermal circuits to describe a specific drying application. The flow of heat energy through the circuit shown in Fig. 3 can be described as

$$q = \frac{(T_\infty - T_i)}{R_T} \quad (13)$$

where

$$R_T = R_k + \frac{R_c R_r}{(R_c + R_r)} \quad (14)$$

is the total resistance to heat transfer for the circuit. If $R_r \gg R_c$ then the radiation transfer mode can be neglected and the total resistance simplifies to

$$R_T = R_k + R_c \quad (15)$$

In the early stages of the drying operation, the thermal resistance attributable to conduction through the dried layer will be negligibly small for the cases illustrated in Figs. 1a and c because the thickness ℓ will approximate zero. For this early stage, the thermal resistance would be

$$R_T = \frac{R_c R_r}{(R_c + R_r)} \quad (16)$$

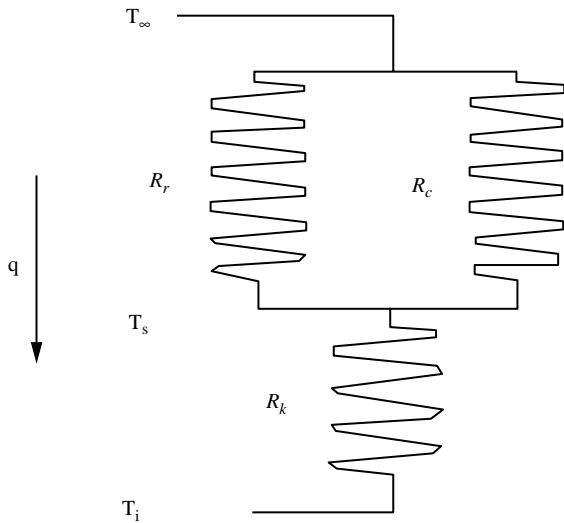


Fig. 3 Construction of a thermal circuit for a drying application.

or

$$R_T = R_c \quad (17)$$

depending on whether or not thermal radiation is appreciable. For a fixed temperature difference and flow rate, we would then expect to generate a constant heat transfer rate during this initial drying period, since the parameters that make up R_T using either Eq. 16 or Eq. 17 are at most dependent on fluid velocity and temperature. As drying proceeds we will expect the thermal resistance attributable to conduction through the growing dried layer to increase and eventually become a significant part of the total resistance. If temperature and flow conditions are fixed, we would therefore expect a decrease in heat transfer rate with time. Heat will continue to flow as long as there is a temperature difference between the energy source and the product.

Mass Transfer

The vapor generated during drying must migrate from the liquid vapor interface through the dried material layer and then be transported out of the drying equipment. For purely diffusional transport, exact solutions to Fick's law are available for a variety of geometric configurations and boundary conditions, usually in the form of infinite series. For a layer of wet material drying off the top surface from an initial uniform concentration of c_0 with the top surface maintained at a constant concentration of c_1 the drying rate for a purely diffusion based transfer mechanism is (9)

$$\dot{m} = \frac{2D(c_0 - c_1)}{\ell} \sum_{n=0}^{\infty} \exp \left[\frac{-(2n+1)^2 \pi^2 D t}{4\ell^2} \right] \quad (18)$$

where \dot{m} is the rate of vapor transferred by the diffusion mechanism (g/s), D the mass diffusivity of the dried layer (cm^2/s), A the transfer area, (cm^2) ℓ the layer thickness (cm), and t the elapsed time (s).

Eq. 18 leads to the following expression for the expected drying curve

$$\frac{M(t)}{M_0} = \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \exp \left[\frac{-(2n+1)^2 \pi^2 D t}{4\ell^2} \right] \quad (19)$$

where $M(t)$ is the amount of solvent in the dried material at time t (g), and M_0 is the initial amount of removable solvent (g).

The corresponding expressions for the drying rate and drying curve of a spherical particle from an initial uniform concentration of c_0 with the exposed surface maintained at a constant concentration of c_1 are (10)

$$\dot{m} = 8Dr_0\pi(c_0 - c_1) \sum_{n=1}^{\infty} \exp\left[\frac{-n^2\pi^2Dt}{r_0^2}\right] \quad (20)$$

and

$$\frac{M(t)}{M_0} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp\left[\frac{-n^2\pi^2Dt}{r_0^2}\right]. \quad (21)$$

where r_0 is the particle radius. Expressions such as Eqs. 18 and 20 illustrate the role of concentration difference as the driving force behind mass transfer and predict a decrease in drying rate with time. However, these expressions tend to overstate the magnitude of the decrease and the dependence on layer thickness and/or particle radius (11).

The total mass transferred will include the combined effect from a number of mechanisms, including molecular diffusion through the solid via vacancies and interstitial defects, migration along dislocations, grain boundaries, and along surfaces of internal pores and fissures, and molecular diffusion through the vapor filled passages defined by the internal pores and fissures (12, 13). In cases where the total pressure inside the material is higher than ambient, the transport mechanism could include convective flow through the pores and fissures.

The two-zone model described above allows for the multiple mechanisms. The rules that govern these mass transfer operations are completely analogous to those governing heat transfer already discussed. The migration of vapor through the dried material layer can be expressed as

$$\dot{m} = -D_{\text{eff}}A \frac{dc}{dx} \quad (22)$$

where $\frac{dc}{dx}$ is the concentration gradient in the direction of transfer ($\text{g/cm}^3/\text{cm}$). The effective mass diffusivity (D_{eff}) will include the combined effect from all the mechanisms outlined above.

Eq. 22 can be recast using vapor phase pressure as the driving force behind the mass transfer, using the ideal gas relationship,

$$\dot{m} = -D_{\text{eff}}A \left(\frac{MW_s}{RT}\right) \frac{dp}{dx} \quad (23)$$

where MW_s is the molecular weight of the solvent (g/mole), R the molar gas constant ($62364.1 \text{ mm Hg cm}^3/\text{mole K}$) and T is the absolute temperature ($^{\circ}\text{K}$).

Using appropriate simplifying assumptions, Eq. 23 can be integrated and placed in a form analogous to Eq. 3

$$\dot{m} = \frac{(p_s - p_i)}{R_D} \quad (24)$$

where p_s, p_i are the vapor pressures at the exposed surface and interface, respectively (mm Hg), and

$$R_D = \frac{\ell}{D_{\text{eff}}A} \left(\frac{RT}{MW_s}\right) \left(\frac{p_m}{P}\right) \quad (25)$$

is the effective mass transfer resistance of the dried layer (mm Hg s/g). Here P is the total pressure and

$$p_m = \frac{(p_s - p_i)}{\ln\left(\frac{P - p_i}{P - p_s}\right)} \quad (26)$$

referred to as the logarithmic mean partial pressure, accounts for the fact that the partial pressures of the individual components in a multicomponent system must equal the system's total pressure. For dilute mixtures of solvent vapor in air, $p_m \cong P$ and the pressure ratio on the right-hand side of Eq. 25 approximates 1.

Solvent transfer from the surface of the dried material can be treated in a manner analogous to Eq. 4 above for heat transfer. An expression for the convective mass resistance can be generated similar to that derived for the thermal resistance:

$$\dot{m} = \frac{(p_{\infty} - p_s)}{R_c} \quad (27)$$

Here, $R_c = \frac{1}{\bar{h}_G A}$, is the convective mass resistance (mm Hg s/g), \dot{m} is the rate of vapor transferred from the exposed surface (g/sec), p_s is the partial pressure of solvent at the exposed solid surface temperature (mm Hg), p_{∞} is the partial pressure of solvent far away from the exposed solid surface (mm Hg), A is the transfer area (cm^2), and \bar{h}_G is the average convective mass transfer coefficient ($\text{g/s-cm}^2\text{-mm Hg}$).

Convective mass transfer coefficients must generally be determined by experiment. Again dimensional analysis can be used to determine physically meaningful nondimensional groups to guide experimental designs. For most drying applications of pharmaceutical relevance, the most important of these nondimensional groups are the Sherwood number (Sh), the Schmidt number (Sc), and the Reynolds number (Re). The Sh and Sc are defined as follows:

$$Sh = \frac{\bar{h}_G L}{D_v} \left(\frac{RT}{MW_s}\right) \left(\frac{p_m}{P}\right) \quad (28a)$$

$$Sc = \frac{\mu}{D_v \rho} \quad (28b)$$

Here, D_v is the mass diffusivity of the solvent through the convective fluid and all other parameters are as defined previously.

The powerful analogy that exists among momentum, heat, and mass transport permits useful values of convective mass transfer coefficients to be calculated from known values of convective heat transfer coefficients. For a particular drying system with a specific geometry and flow characteristics, the following relationship is recommended (14).

$$\bar{h}_G = \left(\frac{\bar{h}_c}{c_p \rho} \right) \left(\frac{MW_s}{RT} \right) \left(\frac{P}{p_m} \right) \left(\frac{Pr}{Sc} \right)^{0.67} \quad (29)$$

Once again, application of the Ohm's law analogy allows construction of mass transfer circuits to describe a specific drying application. The mass flow through the circuits derived from Fig. 2 can be described using

$$\dot{m} = \frac{(p_i - p_\infty)}{R_T} \quad (30)$$

where

$$R_T = R_D + R_c \quad (31)$$

The mass transfer resistance of the dried layer will be negligibly small for some period at the start of drying because the dried layer thickness, starts at zero. During this period the total resistance to mass transfer will equal the convective resistance. For fixed flow, temperature, and solvent concentration far from the exposed product surfaces, the drying rate will be constant during this period. As drying proceeds the resistance of the dried layer becomes a significant portion of the total resistance and continues to increase with time. The drying rate would steadily decrease during this period even if the solvent pressure difference could be held constant.

Phase Transition

The liquid solvent added to a pharmaceutical material generally exists in a variety of states (15). Some will condense or be pulled by capillary forces into macroscopic pores and fissures or into the interstitial spaces between particles. A state of local equilibrium can be assumed to exist at the interface between the liquid and vapor phases of solvent so situated. As a result, the temperature and vapor pressure exerted by the condensed solvent will not be independent of one another. Fig. 4 shows the equilibrium vapor pressure versus temperature relationship for a number of common solvents (16). Heats of vaporization are shown parenthetically (17). Among common solvents, acetone has the highest vapor pressure and water the lowest. Water requires three–five times the energy of the common organic solvents to vaporize.

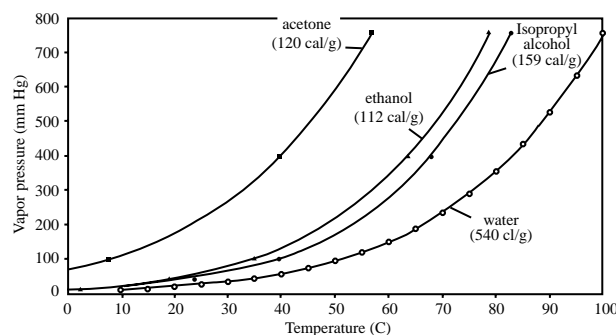


Fig. 4 Vapor pressure curves for common solvents. Heats of vaporization are shown parenthetically. (From Refs. 16 and 17.)

Some of the solvent added will adsorb to the solid surfaces of crystalline solids, particularly at higher energy sites resulting from surface defects and impurities. The amount adsorbed will increase in proportion to the exposed surface area and as the partial pressure of solvent vapor above the surface increases. Solvent can also concentrate in the crystal interior by migrating along high-diffusion paths produced by dislocations and grain boundaries (18). Some polymeric materials of pharmaceutical interest, such as starches and celluloses, often exhibit noncrystalline or amorphous structures. Such materials will typically take up solvent in significantly greater quantities than do crystalline materials with the amount absorbed independent of surface area. As with crystalline solids the amount sorbed will increase as the partial pressure of solvent vapor in contact with the material increases. Sorption data can be experimentally generated and fitted to a variety of available models, including the well-known BET equation, and the more generally applicable 3-state extension developed by deBoer and Guggenheim (19). Data on a number of relevant pharmaceutical materials have been compiled by Callahan and collaborators (20).

In some cases the water or organic solvent added move to regular positions in the crystal lattice and form a stoichiometric relationship with the original molecules resulting in a hydrate or solvate crystalline structure that differs from that of the original crystalline material. Solid state techniques, such as X-ray diffraction, can be used to detect these structural changes. For these materials the impact of solvent addition and removal through drying must be carefully considered as new states with unknown or undesirable properties could be inadvertently generated. In the case of erythromycin, researchers have reported that the method of removing the water of hydration leads to a collapse of the crystalline structure into a metastable amorphous form (21). On the other hand, Schilling and coworkers monitored the formation of a hydrated form of a

5-lipoxaginase inhibitor during wet granulation and subsequent return to the desired anhydrous state after fluid bed-drying (22).

The energy that flows to the water or organic solvent interface is used in two ways. First, and most desirable, it is used to transform the water or organic solvent from a liquid or liquid-like state to a vapor state. The second use, often less desirable, is to raise the temperature of the interface. The distribution can be expressed in terms of an energy balance

$$q_t = \dot{m}\Delta h + Mc_p \frac{(T'_i - T_i)}{\Delta t} \quad (32)$$

where q_t is the total rate of energy transferred to the interface from all sources and mechanisms (calories/s), Δt the time interval under consideration (s), T'_i the temperature of the interface at the end of the time interval ($^{\circ}\text{C}$), T_i the temperature of the interface at the beginning of the time interval ($^{\circ}\text{C}$), Δh the solvent's heat of vaporization (calories/g), M the effective mass of wet product associated with the interface (g), c_p the heat capacity of the wet product (calories/g $^{\circ}\text{C}$), and \dot{m} is the drying rate (g/s).

Eq. 32 can be used to understand the link between drying rate, heat flow, and temperature rise during drying. If the resistance to mass transfer is sufficiently low so solvent vapor molecules generated at the interface can freely escape from the solid, then the bulk of the energy supplied will be absorbed by the first term on the right-hand side of Eq. 32 and the interface will remain cool. This is generally the case near the beginning of the drying cycle because the mass transfer barrier created by a dried product layer has not yet formed. The rate of energy transferred is generally fixed by inlet temperature and flow conditions, leading to a constant drying rate. This portion of the drying cycle is referred to as the constant rate period. As the dried product layer builds the vapor molecules generated cannot readily escape, causing the vapor pressure at the interface to increase. Because temperature and pressure at the interface are related through the equilibrium relationship, the interface temperature increases as the vapor pressure increases. More and more of the energy supplied then shifts from the first to the second term on the right-hand side of Eq. 32 resulting in a drop in the drying rate and a product temperature rise during the time interval. This portion of the drying cycle is referred to as the falling rate period. The higher interface temperature and higher heat transfer resistance created by the dried product barrier serve to reduce the rate of energy transfer in subsequent time intervals as predicted by Eq. 13. The higher interface pressure partially offsets the effect of increasing the mass transfer resistance.

Psychrometrics

The solvent vapor generated during drying must be transported out of the drying equipment. If it isn't, the gas surrounding the material to be dried will soon become saturated with vapor and drying will cease. Various interconvertible terms have evolved over time to express the amount of solvent that is absorbed by the drying gas. Many of the common terms have been defined strictly to apply to the air-water vapor system. However, the concepts involved apply equally well to any solvent-drying gas combination. The most common term is that of relative humidity (ϕ), which expresses the ratio of the actual amount of water vapor present to the maximum amount that could be present at a specified temperature. Amounts can be expressed in any consistent way, including units of mass, moles, or partial pressures. For drying applications, partial pressures are particularly convenient and the relative humidity becomes

$$\phi = \frac{p_{\infty}}{p_{\text{sat}}} \quad (33)$$

where p_{∞} is the partial pressure of solvent vapor present and p_{sat} is the maximum pressure at saturation. The saturation pressures for common solvents have been shown previously (Fig. 4) as a function of temperature. For ease of computation, saturation data can be fit to an equation of the form

$$\ln(p_{\text{sat}}) = A + \frac{B}{T} + C \ln(T) + DT \quad (34)$$

where T is the absolute temperature and the constants A , B , C , and D depend on the solvent (23). Recommended constants for the common solvents, determined through regression, are listed in Table 1 for p_{sat} in mm Hg and temperature expressed in K.

An alternate expression for solvent content is the specific humidity or humidity ratio, which is defined as the ratio of the mass of solvent vapor present to the mass of dry gas

$$\omega = \frac{m_v}{m_g} = \left(\frac{\text{MW}_v}{\text{MW}_g} \right) \left(\frac{P_v}{P_g} \right) \quad (35)$$

Eqs. 33, 34, 35 allow interconversion from one expression for solvent content to another. For example, knowledge of temperature, total pressure, and relative humidity allows the humidity ratio to be determined using

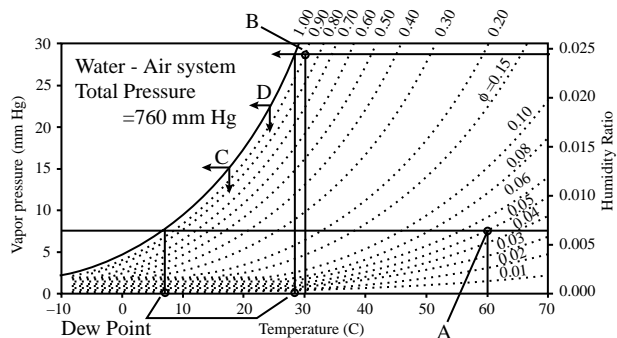
$$\omega = \left(\frac{\text{MW}_v}{\text{MW}_g} \right) \left(\frac{\phi P_{\text{sat}}}{P - \phi P_{\text{sat}}} \right) \quad (36)$$

Solvent content can be portrayed graphically in what is known as a psychrometric chart, such as the one for an air

Table 1 Recommended constants for computing saturation pressure, using Eq. 34 for common solvents (p_{sat} in mm Hg and T in K)

Solvent	A	B	C	D
Water	70.708779	-7175.9470	-7.9064596	0.0053125111
Ethanol	-93.710636	-2458.5969	20.649371	-0.039031369
Isopropyl alcohol	-7.4598754	-5.017.1464	5.7374144	-0.015489516
Acetone	92.141422	-6280.1292	-12.241911	0.013701258

water system at atmospheric pressure shown in Fig. 5. Such a chart is a convenient tool for converting between the different expressions for solvent content and for tracking changes in solvent content during drying. Say, for example, that drying air enters a dryer at 60°C and a relative humidity of 0.05 (point A in Fig. 5) and leaves at 30°C and a relative humidity of 0.90 (point B in Fig. 5). Moving horizontally and to the left from point A shows the inlet condition corresponds to a moisture vapor pressure of approximately 7.5 mm Hg and by moving horizontally and to the right shows a humidity ratio of approximately 0.006. The exit condition (point B in Fig. 5) corresponds to a humidity ratio of 0.024 for a difference of approximately 0.018 g of water vapor carried out of the system per gram of dry air. The intersection of the moisture content (horizontal) lines with the saturation curve ($\phi = 1.0$) uniquely defines the so called dew point temperature, indicated in Fig. 5, which is yet another way of specifying solvent content. Fig. 5 can also be used to illustrate the effect of product temperature on the mass transfer driving force. For example, product at 17.5°C (point C in Fig. 5) would provide a driving force of approximately 7.5 mm Hg (15.0–7.5 mm Hg) between the solvent vapor interface and the inlet drying air. A modest temperature rise to 24°C approximately doubles the driving force with the same inlet air by increasing the vapor pressure at the solvent vapor interface.

**Fig. 5** Psychrometric chart for an air–water system at a total pressure of 760 mm Hg.

PRACTICE

Drying can be carried out successfully using a variety of commercially available equipment designs. Pharmaceutical drying equipment has been classified according to principal mode of operation in a recently published regulatory guidance document (24) as shown in Table 2. Equipment classified as direct heating allows intimate contact between the material being dried and the heat energy source, usually a heated gas. That same gas is used to transport the vapor generated from the equipment. In indirect conduction, the energy is transferred from the source, usually a heated liquid, to the material being dried through a conducting wall. In this case other means must be used to remove the generated vapor from the equipment. Radiant approaches do not rely on temperature to generate or transfer the needed energy to the material being dried. Instead, the material is exposed to electromagnetic energy at frequencies strongly absorbed by the solvent being targeted for removal. Specialized approaches, such as spray drying and lyophilization, are treated in separate articles in this encyclopedia and will not be covered more here.

Tray and Truck-Drying

Historically, the most common method of drying of pharmaceutical powders has been tray-drying. With this method, wet powder or granulation is placed on paper-lined trays, usually solid or perforated metal, which are then placed directly onto racks in a drying chamber (oven) or onto movable racks, or trucks, that are wheeled into an oven. The heat and low relative vapor pressure of solvent provided by the flow of heated, dry air throughout the chamber provide a driving force for solvent transfer to and subsequent removal from the particle surfaces of the powder. This results in the gradual overall loss of solvent from the bulk powder.

The drying process from solids has been characterized by three drying regions, as shown in Fig. 6 (12, 25). The

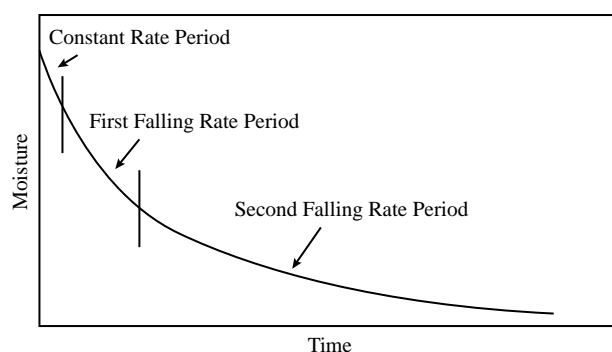
Table 2 Classification of pharmaceutical drying equipment

Class	Subclass	Common names
Direct heating	Static solids bed	Tray and truck dryers
	Moving solids bed	Belt dryer
	Fluidized solids bed	Fluid bed dryer
	Dilute solids bed	Spray dryer
Indirect conduction	Moving solids bed	Tumble dryer
	Gas stripping	Zanchetta
	Static solids bed	Heated shelf tray drier
	Lyophilizers	Freeze dryer
Radiant	Microwave, moving solids bed	Microwave dryer

first, termed the Constant Rate Period, is the initial drying phase in which surface moisture exceeds a critical level and rate is controlled by surface area. When the level of moisture falls below the critical level, it begins to be controlled by mass transfer from inside the solid mass: this is called the First Falling Rate Period. As drying proceeds, mass transfer is not able to supply moisture to the surface of the solid mass at a rate equal to the drying rate, and the free water content at the surface goes to zero. At this time, the surface temperature rises rapidly, and a receding evaporation front may be formed that divides the solid into a wet region and a dry or sorption region. This is the beginning of the Second Falling Rate Period, during which mass transfer of moisture vapor through the sorption region becomes more and more retarded.

The falling rate portion of the drying process can be generally modeled by using a variation of Eq. 19 in which the summation is truncated after one term:

$$\ln\left(\frac{M(t)}{M_0}\right) = -\left(\frac{\pi^2 D}{4\ell^2}\right)t + \ln\left(\frac{8}{\pi^2}\right) \quad (37)$$

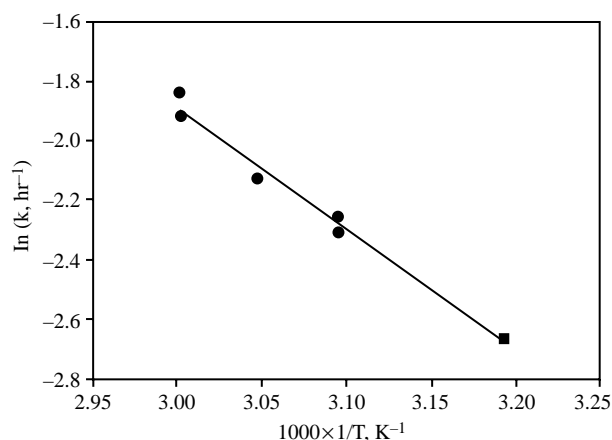
**Fig. 6** The phases of the drying process. (From Ref. 12.)

Eq. 37 can be simplified to

$$\ln\left(\frac{M(t)}{M_0}\right) = -kt - 0.2 \quad (38)$$

where k is a first-order drying rate constant such that when $\ln(M(t)/M_0)$ is plotted versus time a straight-line relationship is obtained with a slope of $-k$ (11). This becomes very useful in trying to model the tray-drying process and evaluating the impact of process variables such as bed thickness and drying temperature changes. An example of this is given in Fig. 7, in which drying rate constants obtained at multiple temperatures are plotted versus inverse temperature, allowing one to predict drying rate at any interpolated temperature.

During the drying process, internal liquid transport occurs via capillary flow, while vapor transport occurs both via diffusion and true mass flow driven by pressure gradients (12, 26). Because the powder bed is static,

**Fig. 7** Temperature dependence of drying rate constant (k from Eq. 38). (From Ref. 11.)

significant resistance to the diffusion of solvent from the bed as a whole reduces the rate of drying, thereby limiting the efficiency of this method of drying. This is demonstrated by the dependence of the first order rate constant k on the depth of the bed being dried. Theoretically it is shown that the drying rate constant is an inverse function of the square of the bed thickness (see Eq. 37), but experimental data shows a relationship that more closely resembles an inverse relationship of k with the first order of bed depth (11).

Tray-drying is also used as a method to remove water from soft elastic gelatin capsules (27), and can be model according to Eq. 39a:

$$\ln(c - c_\infty) = -\frac{t}{\Gamma} + \ln(c_0 - c_\infty) \quad (39a)$$

where

$$\Gamma = \frac{h^2}{5.8D} \quad (39b)$$

Here c is the amount of moisture at time t , c_0 , c_∞ the amounts of moisture at time zero and infinity, respectively, h the thickness of the gelatin film, and D is the diffusion coefficient of moisture through gelatin.

This modeling becomes important as a soft-gel product is being developed and a drying end point needs to be established and reproduced.

Despite the low relative capital investment required for tray-drying, it provides a low rate of drying and the loading and unloading of trays is a labor-intensive process. Although still commonly found in both drug substance and drug product manufacturing procedures, tray-drying has become less popular in comparison to other more efficient, reproducible, and well-defined drying procedures such as fluid bed and vacuum tumble drying.

Fluid Bed Dryers

Fluid bed-drying is a widely used example of the direct heating classification. Drying is accomplished by suspending the particles to be dried directly in a stream of heated air or other gaseous media. The intimate contact and high surface areas available for transfer result in fast, efficient drying, often making fluid bed the approach of choice for high-volume products.

A typical installation is shown in Fig. 8. Ambient air enters an air-handling unit through a coarse filter in the lower right. The air is first passed over a chilled, condensing coil to reduce the moisture content. The air leaving the coils can be assumed to be in equilibrium with the condensed water so the temperature measured at the

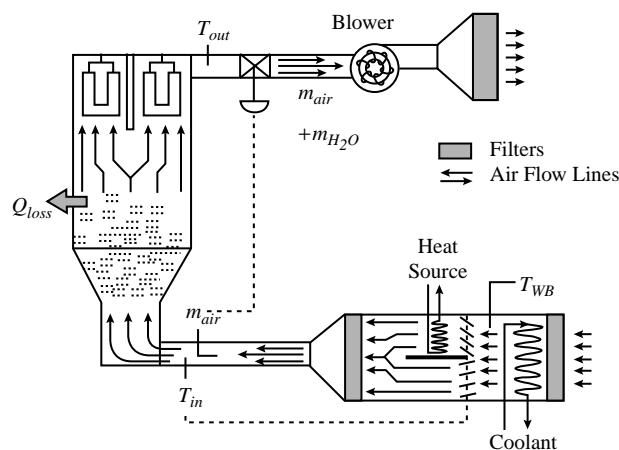


Fig. 8 Schematic of a typical fluid bed dryer installation.

coil outlet represents the wet bulb temperature, a measure of moisture content. The coolant temperature determines the degree of dehumidification achieved. Chilled water and refrigerant are common coolants. A portion of the inlet air is then diverted through louvers past a heat source and then allowed to remix with the portion not diverted. A steam coil is commonly used as the heat source. The louvers are mechanically linked so that one flow path opens as the other closes. A feedback loop can be established between the downstream temperature and the louver position to control drying temperature. If the drying temperature drops below the set point, the louver position is adjusted to divert a larger fraction of the incoming airflow past the heat source, resulting in a higher temperature of the remixed streams. If the drying temperature drifts above the set point, the louvers are repositioned to divert less past the heat source. This type of arrangement is referred to as *face and bypass control* and has the advantage of fast response time and minimal overshoot. The warm, dehumidified air is then passed through a second, finer filter and sent to the dryer.

The product to be dried is placed inside a bowl on top of a retaining screen. The retaining screen can be of the wire mesh, perforated plate, gill plate, or combination design. As the drying air enters the bowl from below, it drags the product particles off the retaining screen and entrains them in the flow stream. The air transfers heat energy to the suspended particles and collects the solvent vapors given off. A small part of the heat energy supplied to the drying air stream is lost through transfer to the surrounding environment. Product filters are provided to prevent the entrained particles from leaving the drying chamber. A split filter design allows for periodic cleaning without disrupting the drying operation. Flow through one filter

segment can be interrupted so it can be mechanically shaken or reverse-pulsed with clean air to remove accumulated particles. Flow then resumes and the cleaning operation is performed on the other segment.

Drying air flow rate control is achieved using a blower that works against a flow control valve. Both are typically located on the downstream side of the drier to maintain the drying chamber at a slight but not excessive negative pressure with respect to ambient. Product particles and organic solvent vapors are thus unable to escape against the negative pressure gradient. The airflow rate is measured, usually on the clean and dry upstream side of the drier, and a feedback loop is established with the flow control valve. Flow control is achieved by adjusting the position of the flow control valve. Before releasing the used air back into the environment, it is filtered once more to remove any pharmacologically active and potentially hazardous product particles that may have leaked past the product filters. For organic vapor applications, the spent air would also be treated to separate and remove the vapors from the air stream before releasing it back to the environment. Grounding, containment, and venting strategies are incorporated into the designs to control explosion hazards.

The dryer bowl is designed in the shape of an inverted frustum of a right circular cone, with the smaller diameter at the bottom of the bowl. As the drying air passes up through the bowl, the increasing area causes the flow velocity to drop in the direction of flow. At the lower velocity the larger, heavier particles can no longer be sustained and they fall back toward the retaining screen. The situation represents a tension between the drag forces exerted on the particle by the moving fluid and the force of gravity trying to pull the particle back down to the retaining screen. For spherical particles moving at low velocity in a fluid stream the expression for the drag force first determined by Stokes (28) can be set equal to the particle weight to yield an expression for the minimum fluidization velocity (29)

$$V = \frac{gd^2}{18\mu}(\rho_p - \rho) \quad (40)$$

where V is the minimum fluidization velocity (cm/s), d the particle diameter (cm), g the acceleration of gravity (980 cm/s²), ρ_p the particle density (g/cm³), and ρ and μ are the density and dynamic viscosity of the fluid, respectively.

Strictly speaking, Eq. 40 is a good approximation only at low Re , that is at particle diameters significantly less than 0.01 cm (100 micrometers). White (30) has provided a formula extending the range to particle diameters as high as 1 cm, using a curve fit of data from many sources. A plot

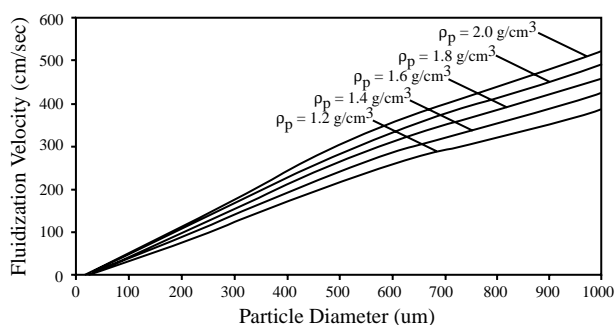


Fig. 9 Fluidization velocity as a function of diameter and density for spherical particles suspended in an air stream at 45°C, using the curve fit of White. (From Ref. 30.)

of fluidization velocity as a function of particle diameter (in microns) and density is shown in Fig. 9 assuming spherical particles in air at 45°C, using White's formula. Typical commercial equipment provides velocities in the range of 150–250 cm/s at the retaining screen that drop after expansion into the range of 60–100 cm/s. Because particle density drops as drying proceeds, flow rates used at the beginning of drying to fluidize the particle bed could be reduced later in the cycle without losing entrainment.

The drying rate at any point in the drying cycle can be derived from information provided from available process instrumentation without resorting to intrusive sampling during the process. An energy balance across a control volume surrounding the drying bed yields:

$$\dot{m} = \frac{\dot{m}_g C_{p,g}(T_{in} - T_{out}) - Q_{loss}}{h_{fg}} \quad (41)$$

where \dot{m} is the drying rate, $C_{p,g}$ the specific heat capacity of drying gas at constant pressure, \dot{m}_g the mass flow of drying gas through dryer, T_{in} the inlet temperature of dryer gas, T_{out} the outlet temperature of dryer gas, Q_{loss} the heat loss to the environment through thermal convection, and h_{fg} is the latent heat of vaporization for solvent.

The heat loss term can be estimated by applying Eq. 41 to conditions near the end of the drying cycle, where the evaporation rate is negligible.

$$Q_{loss} = [\dot{m}_g C_{p,g}(T_{in} - T_{out})]_{\text{end of cycle}} \quad (42)$$

Using these conditions, the $\bar{h}A$ term is then calculated from the following equation and is assumed constant throughout the drying cycle:

$$Q_{loss} = \bar{h}A(\bar{T} - T_{amb}) \quad (43)$$

where \bar{h} is the average convective heat transfer coefficient, A the external dryer surface available for heat transfer, \bar{T}

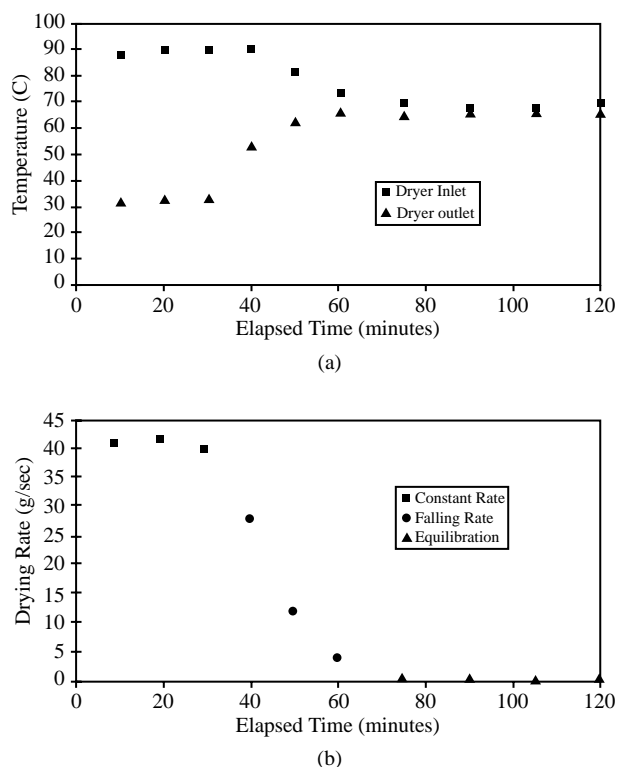


Fig. 10 Temperature and drying rate histories for a water-based drying case in air (a) temperature history; (b) computed drying rate history. (From Ref. 31.)

the average temperature in the drying bed, and T_{amb} is the ambient temperature.

The temperature and computed drying rate histories for a water based drying case in air at a constant flow rate of $2.36 \times 10^6 \text{ cm}^3/\text{s}$ (5000 ft^3/min) is shown in Fig. 10 (31). The heat loss term parameter $\bar{h}A$ is computed to be 17.9 $\text{cal/s}/^\circ\text{C}$ for this case. An early constant rate period is evident that extends out to the first 30 min of drying in which a drying rate of approximately 40 g/s is achieved.

Because different mechanisms limit the drying rate in each of the drying periods, scheduled changes in flow rate and inlet temperature have been used with great success to shorten drying cycles without subjecting the pharmaceutical material to unnecessary stress (31, 32). During the constant rate period, the drying rate is limited by the enthalpy available in the inlet air and its capacity to absorb the vapor that is generated. Increases in flow rate and inlet temperature can be used to reduce the length of the constant rate period. Staged reductions in inlet temperature and flow rate can be scheduled without impacting the rate during the falling rate and equilibration periods because internal moisture transfer limits the overall rate. The lowest flow rates can be used during the equilibration

period because the low-moisture, low-density particles are easiest to fluidize and because dehumidification techniques should become more efficient, resulting in inlet air with lower moisture content.

Vacuum Drying

Vacuum can be used with all of the indirect conduction and microwave approaches to drying. The total pressure surrounding the pharmaceutical material is reduced to levels below the saturation pressure of the solvent at the interface between the wet and dry layers causing generation of vapor. With suitable vacuum levels, drying can be cost-effective at relatively low product temperatures. Vacuum drying is particularly advantageous for heat- or oxygen-sensitive products, for reducing the risk of dust explosions, and for applications requiring solvent recovery or extremely low residual solvent levels.

A typical rotating double-cone vacuum dryer is shown in Fig. 11. Vapor exits the dryer via a tube that passes through a rotary seal along the axis of rotation. A filter prevents particles from leaving the dryer with the exiting vapor. Vacuum can be supplied by conventional pumps, blowers, or steam jets. Heating fluid circulates through a jacket and enters and exits through dynamic seals along the axis of rotation. Typical rotation speeds are 6–8 rpm. Working capacities, generally defined as 50% of total volume, range from 0.1 to 10 m^3 and vacuum levels range from just under ambient to 20 mm Hg (33). Indirect methods rely on contact between the wet material and the jacketed walls of the dryer to supply energy and the drying rate can be heat transfer-limited. Average drying rates range from 1–7 $\text{kg/h}/\text{m}^2$ of heat transfer surface area available. The ratio of jacket area to working volume tends

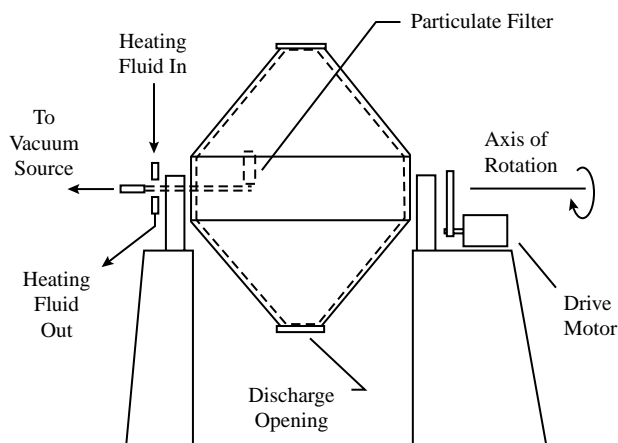


Fig. 11 Rotating double-cone vacuum dryer.

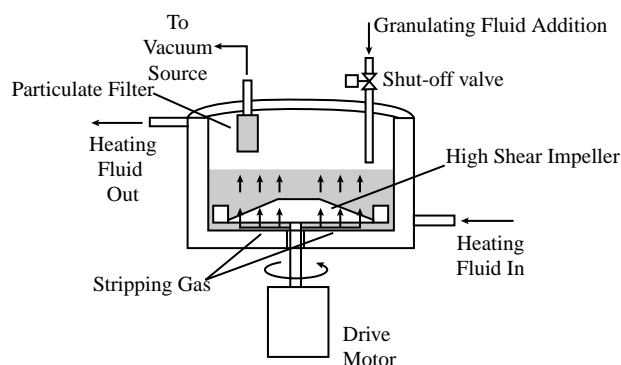


Fig. 12 High shear vacuum processor.

to decrease with increasing size, so larger models often require additional internal plates or pipe coils to increase available area for heat transfer (33).

Vacuum drying can be readily incorporated into high shear granulation designs to permit multiple processing steps to be completed in a single piece of equipment, as shown in Fig. 12. Granulation takes place in an initial processing step by introducing a fluid to the particle bed while mixing it with a high shear impeller. Vacuum drying follows. Typical vacuum conditions are 18–22 mm Hg. The vapor exits through a port in the cover through a tube equipped with a particulate filter. Heating fluid is circulated through the jacket of the bowl with typical operating temperatures of 60–80°C. Inert stripping gas (3–30 m³/h depending of vessel volume) is introduced through the shaft seal to improve the convective transfer of vapor out of the vessel during drying. Gas stripping rates above an optimal level reduce the drying effectiveness by raising the pressure in the vessel. Commercial designs allow for tilting of the unit up through 180° to improve contact between the granules and the heated walls. Microwave and infrared generators can be added to augment the heat transfer rates (34, 35).

Microwave (Dielectric) Drying

By applying microwave energy to pharmaceutical systems to be dried, dielectric materials such as water and solvents with dissolved salts absorb the energy thereby increasing molecular vibration. This movement is in turn converted to friction resulting from interactions with neighboring molecules, solvent temperature increases and ultimately vaporizes, and drying is affected (36). In contrast to previously discussed more conventional means of drying, energy is transferred to the entire volume of solvent in a particle rather than relying on heat transfer from contact

surfaces to the interior of a particle or bed. This mode of energy transfer provides for higher temperatures at the center of the granule or powder bed, generating a temperature gradient directed outward from the center of the material. This facilitates both liquid and vapor mass transfer away from the center of the granule. Vaporization of the solvent inside the granule can occur (36), which allows drying rates to be governed by the diffusion coefficient of the solvent vapor rather than that of the liquid, potentially reducing mass transfer limitations in drying rate.

Microwave dryers can be constructed as stand-alone cabinets, as combination dryers with vacuum, fluid bed, or vibrational capabilities, and as one-pot processors that provide mixing and granulation capabilities in conjunction with microwave drying. Microwaves are generated at typical frequencies of either 915 MHz or 2.45 GHz, and are directed to the powder bed to be dried by way of waveguides. The magnetrons used to generate the microwave output require high-voltage supply and may require water cooling to remove excess heat. The size (output) and number of magnetrons depends on the size of the dryer and mass of wet material to be dried, and in many applications are pulsed on and off by a controller to prevent damage to the product resulting from excessive heat generation.

Some dryers also provide heat energy to the powder mass by a jacketed vessel, thereby increasing overall heat transfer. Moisture can be removed via vacuum or hot air fluidization depending on the design of the dryer allowing for improved evaporative drying and vapor mass transfer. Fig. 13 shows the relationship between power input (W) and first-order drying rate constant in a microwave fluid-bed processor (37).

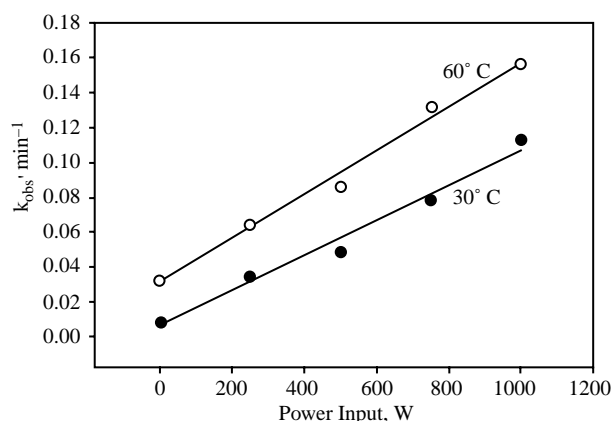


Fig. 13 The influence of microwave power input and inlet air temperature on microwave fluid-bed drying. (From Ref. 37.)

The extent of microwave drying can be correlated to the amount of power absorbed by the product, which is described by Eq. 44 (38):

$$P = 2\pi f V^2 E_0 E_r \tan \delta \quad (44)$$

where P is the power density (W/m^3), f is the frequency (Hz), V the voltage gradient (V/m), E_0 the dielectric constant of vacuum ($8.85 \times 10^{-12} \text{ F/m}$), E_r the dielectric constant of the material being dried (F/m), and δ is the loss angle (a physical property of magnetic waves).

The product of the dielectric constant and the loss tangent ($\tan \delta$) is called the *loss factor* (36), E_r'' , and is a relative measure of how easily a material will be heated by microwave energy.

$$E_r'' = E_r \tan \delta \quad (45)$$

A table of loss factors of some common solvents and excipients are given in Table 3.

Clearly the composition of the powder to be dried plays an integral role in the drying process, using microwaves based on the energy absorption characteristics a formulation possesses. As microwaves penetrate the powder bed the intensity of the electrical field strength is reduced by absorption according to when

$$d = \frac{\lambda_r E_r^{1/2}}{2\pi E_r''} \quad \text{when } E_r'' \ll 1 \quad (46a)$$

or when

$$d = \frac{\lambda_r}{2\pi (E_r'')^{1/2}} \quad \text{when } E_r'' \gg 1 \quad (46b)$$

where d is the depth where the field strength is 37% (or $1/e$) of original value, and λ_r is the wavelength (e.g., 12.3

Table 3 Comparison of loss factors of some common pharmaceutical materials

Material	Loss factor, E_r''
Methanol	13.6
Ethanol	8.6
Water	6.1
Isopropanol	2.9
Acetone	1.25
Corn starch	0.41
Dibasic calcium phosphate	0.06 ^a
Lactose (dry)	0.02, 0.077 ^b
Lactose (15% moisture)	0.50 ^b

^aFrom Ref.(34)

^bFrom Ref.(39)

(From Ref. 38 except where indicated.)

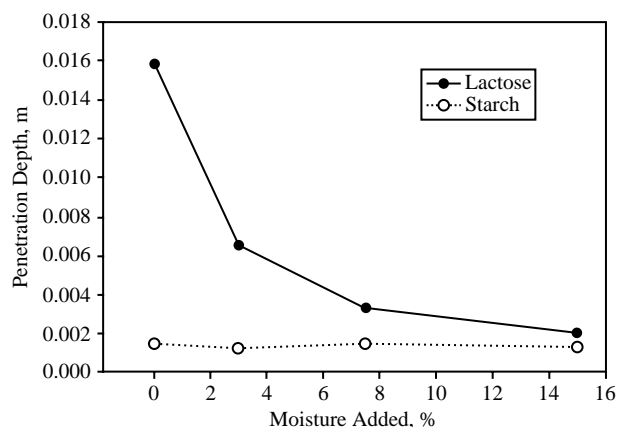


Fig. 14 The effect of moisture on microwave penetration depth. (From Ref. 39.)

cm at a frequency of 2450 MHz). Fig. 14 shows the calculated penetration depth for lactose and starch.

Because the penetration depth is limited, both the speed and the uniformity of drying can be improved by mixing during the drying process. As a material loses moisture during the drying process, both its dielectric constant and its loss tangent change. Because the loss factor is the product of these numbers, an understanding of these property characteristics throughout the drying process may be important. For example, starch with 3% moisture has a higher loss factor than it does with both 7.5% and 15% moisture (39).

Theoretical comparisons have been made between conventional drying techniques and microwave and have shown the superior drying rate of microwave over conductive drying in a jacketed bowl (39) and microwave-aided fluid bed-drying over fluid bed-drying alone (40). Because of the reduced drying time associated with the use of increased microwave energy, the generation of pharmaceutical dust can be reduced in a single-pot drying process (41).

Because of the benefits in drying uniformity and efficiency in energy transfer, microwave drying provides an attractive alternative to more conventional modes of drying. For highly potent pharmaceutical compounds the microwave unit provides a high degree of containment (particularly when coupled with high shear granulation) and is an easily cleanable dryer. However, the initial capital investment to install such a dryer and the significant amount of ancillary equipment is oftentimes prohibitive in conventional applications. Nonetheless, uniformity in drying and reduction in time and manpower may be sufficient to consider microwave drying as a viable alternative.

REFERENCES

1. Rohsenow, W.M.; Choi, H.Y. *Heat, Mass, and Momentum Transfer*, Prentice-Hall: Englewood Cliffs, NJ, 1961; 94–98.
2. Carslaw, H.S.; Jaeger, J.C. *Conduction of Heat in Solids*, Oxford University Press: London, 1959.
3. Langhaar, H.L. *Dimensional Analysis and Theory of Models*, John Wiley & Sons: New York, 1951.
4. Kreith, F. *Principles of Heat Transfer*, 3rd Ed.; Intext: New York, 1973; 327–372.
5. Whitaker, S. Forced Convection Heat Transfer Correlations for Flow in Pipes, Past Flat Plates, Single Cylinders, Single Spheres, and for Flow in Packed Beds and Tube Bundles. *AIChE J.* **1972**, *18*; 361–371.
6. Johnston, H.F.; Pigford, R.L.; Chapin, J.H. Heat Transfer to Clouds of Falling Particles. *Univ. Illinois Bull.* **1941**, *38* (43); 16–20.
7. Kreith, F. *Principles of Heat Transfer*, 3rd Ed.; Intext: New York, 1973; 445–446.
8. Siegel, R.; Howell, J.R. *Thermal Radiation Heat Transfer*, McGraw-Hill: New York, 1972.
9. Crank, J. *Mathematics of Diffusion*, 2nd Ed.; Oxford University Press: Oxford, 1975; 47–48.
10. Crank, J. *Mathematics of Diffusion*, 2nd Ed.; Oxford University Press: Oxford, 1975; 91.
11. Carstensen, J.T.; Zoglio, M.A. Tray Drying of Pharmaceutical Wet Granulations. *J. Pharm. Sci.* **1982**, *7* (1), 35–39.
12. Chen, P.; Pei, D.C.T. A Mathematical Model of Drying Process. *Int. J. Heat Mass Transfer* **1989**, *32* (2), 297–310.
13. Moyers, C.G.; Baldwin, G.W. Psychrometry, Evaporative Cooling, and Solids Drying. *Perry's Chemical Engineering Handbook*, 7th Ed.; McGraw-Hill: New York, 1997; 31–33, Ch. 12.
14. Kreith, F. *Principles of Heat Transfer*, 3rd Ed.; Intext: New York, 1973; 596–600.
15. Zografi, G. States of Water Associated with Solids. *Drug Dev. Ind. Pharm.* **1988**, *14* (14), 1905–1926.
16. Weast, R.C., Ed. *Handbook of Chemistry and Physics*, 70th Ed.; CRC Press: Boca Raton, FL, 1990; D191–D192–D199–D200.
17. Weast, R.C., Ed. *Handbook of Chemistry and Physics*, 70th Ed.; CRC Press: Boca Raton, FL, 1990; C672–C673.
18. Shewmon, P. *Diffusion in Solids*, 2nd Ed.; The Minerals, Metals & Materials Society: Warrendale PA, 1989; 191–215.
19. Kontny, M.J.; Zografi, G. Sorption of Water by Solids. *Physical Characterization of Pharmaceutical Solids*; Brittain, H.G. Ed.; Marcel Dekker, Inc.: New York, 1995; 391–394.
20. Callahan, J.C.; Cleary, G.W.; Elefant, M.; Kaplan, G.; Kensler, T.; Nash, R.A. Equilibrium Moisture Content of Pharmaceutical Excipients. *Drug Dev. Ind. Pharm.* **1982**, *8* (3), 355–369.
21. Bauer, J.; Quick, J.; Oheim, R.J. Alternate Interpretation of the Role of Water in the Erythromycin Structure. *Pharm. Sci.* **1985**, *74*, 899.
22. Schilling, R.J. *Abbott Laboratories*, : North Chicago, IL, unpublished.
23. Van Wylen, G.J.; Sonntag, R.E. *Fundamentals of Classical Thermodynamics*, 2nd Ed.; John Wiley & Sons: New York, 1973; 393.
24. *Guidance for Industry, SUPAC-IR/MR: Immediate Release and Modified Release Solid Oral Dosage Forms, Manufacturing Equipment Addendum CMC 9*, Revision 1 U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research: Washington DC, 1999.
25. Cooper, J.; Swartz, C.J.; Suydam, W., Jr. Drying of Tablet Granulations. *J. Pharm. Sci.* **1961**, *30* (1), 67–75.
26. Wang, Z.H.; Chen, G. Heat and Mass Transfer in Fixed-bed Drying. *Chem. Eng. Sci.* **1999**, *54*; 4233–4243.
27. Carstensen, J. *Solid Pharmaceuticals: Mechanical Properties and Rate Phenomena*, Academic Press: New York, 1980; 149–150.
28. Batchelor, G.K. *An Introduction to Fluid Dynamics*, Cambridge Univ. Press: London, 1970; 229–235.
29. Prandtl, L. *Essentials of Fluid Dynamics*, Blackie & Son: London, 1967; 105–106.
30. White, F. *Viscous Fluid Flow*, McGraw-Hill: New York, 1974; 204–209.
31. Hlinak, A.J.; Saleki-Gerhardt, A. An Evaluation of Fluid Bed Drying of Aqueous Granulations. *Pharm. Dev. Tech.* **2000**, *5* (1), 11–17.
32. Morris, K.R.; Stowell, J.G.; Byrn, S.R.; Placette, A.W.; Davis, T.D.; Peck, G.E. Accelerated Fluid Bed Drying Using NIR Monitoring and Phenomenological Modeling. *Drug Dev. Ind. Pharm.* **2000**, *26* (9), 985–988.
33. Moyers, C.G.; Baldwin, G.W. Psychrometry, Evaporative Cooling, and Solids Drying. *Perry's Chemical Engineering Handbook*, 7th Ed.; McGraw-Hill: New York, 1997; 65–66, Ch. 12.
34. Duschler, G.; Carius, W.; Bauer, K.H. Single-Step Granulation Method with Microwaves: Preliminary Studies and Pilot Scale Results. *Drug Dev. Ind. Pharm.* **1995**, *21* (14), 1599–1610.
35. Duschler, G.; Carius, W.; Bauer, K.H. Single-Step Granulation: Development of a Vacuum-Based IR Drying Method (Pilot Scale Results). *Drug Dev. Ind. Pharm.* **1997**, *23* (2), 119–126.
36. Köblitz, T.; Körblein, G.; Ehrhardt, L. Careful Drying of Temperature-Sensitive Pharmaceutical Granulates in a High-Frequency Field. *Pharm. Technol* **1986**, *April*, *32*.
37. Doelling, M.K.; Nash, R.A. The Development of a Microwave Fluid-Bed Processor. II. Drying Performance and Physical Characteristics of Typical Pharmaceutical Granulations. *Pharm. Res.* **1992**, *9* (11); 1493–1501.
38. Doyle, C.; Cliff, M.J. Microwave Drying for Highly Active Pharmaceutical Granules. *Manuf. Chem.* **1987**, *February*, 23–32.
39. Vromans, H. Microwave Drying of Pharmaceutical Excipients; Comparison with Conventional Conductive Drying. *Eur. J. Pharm. Biopharm.* **1994**, *40* (5), 333–336.
40. Wang, Z.H.; Chen, G. Theoretical Study of Fluidized-Bed Drying with Microwave Heating. *Ind. Eng. Chem. Res.* **2000**, *39*; 775–782.
41. Kiekens, F.; Cordoba-Diaz, M.; Remon, J.P. Influence of Chopper and Mixer Speeds and Microwave Power Level During the High-Shear Granulation Process on the Final Granule Characteristics. *Drug Dev. Ind. Pharm.* **1999**, *25* (12); 1289–1293.

ECONOMIC CHARACTERISTICS OF THE R&D-INTENSIVE PHARMACEUTICAL INDUSTRY

Douglas L. Cocks

Indiana University, Bloomington, Indiana, U.S.A.

INTRODUCTION

This article presents a brief sketch of the economics of the R&D-intensive ethical pharmaceutical industry, highlighting its dynamic characteristics. The approach taken here minimizes the use of static analysis, and thus avoids the use of pure or perfect competition as an analytical tool. In this theoretical discussion, certain empirical studies will be cited as support for aspects of the theory being developed.

The theory discussed here will concentrate on allocative efficiency, but as with all discussions of allocative efficiency, elements of technical efficiency will automatically be involved and at least implicit recognition of these elements will be evident. The allocative efficiency concerns will be placed in a dynamic framework; we will be attempting to establish a notion of "dynamic pure competition" that has analytical and public policy implications. The concept of dynamic pure competition will describe a hybrid form of workable competition as the term is used by industrial organization economists.

AN OUTLINE OF A COMPETITIVE PROCESS

Before we get into an outline of the theory of pharmaceutical economics, we need to establish pure competition as a *competitive process*. Traditional microeconomics has assumed implicitly that the "natural state" is one that is depicted by pure competition. Deviations from the natural state occur as a disequilibrium, by the establishment of monopoly power, or through other often cited market failures. In cases of disequilibrium, the tatonnement will bring us to the equilibrium ideal of pure competition. Interestingly, the model of pure competition never really describes the process of the tatonnement (equilibration) but only the conditions necessary for the process to operate and the final equilibrium to result when the process has worked itself out.

The monopoly power deviation arises because the nature of "economic man" causes him or her to attempt to break out of a pure competitive equilibrium, or the equilibrating tatonnement process, and maximize his or her own economic situation relative to the rest of the world. The economic man will attempt to establish a monopoly power position through "entry barrier" means (1a, 1b).

According to traditional microeconomics, then, the natural economic process is one that proceeds from the natural state of pure competitive equilibrium, or from where the necessary conditions exist for the pure competitive tatonnement process to take place, to conditions of monopoly.

The competitive process that is relevant here is one in which a naturally occurring monopoly is systematically faced with a pressure that erodes this position. It is a process that occurs on a continuum and which must be considered on the basis of changes through time. Reverting to the static sense, the economic concept of deadweight welfare loss is a representation of the social opportunity cost that is associated with having entrepreneurs, singular and corporate, invading previously held monopoly positions by providing new and improved products and services. This in turn represents the economic progress that generates welfare gains, in the technically economic context and not in the sense of providing public funds to needy populations. Through time, economic life is characterized as a continual process of monopoly establishment and systematic erosion via entrepreneurial activity. This entrepreneurial activity constitutes the observation of, and action upon, profit opportunities as evidenced by static monopoly rents.

We can think of dynamic pure competition as a process where naturally occurring monopoly is systematically eroded. It represents a kind of entropy that properly allocates resources in the production of current and future goods and services. The underlying characteristics of the competitive process are that it recognizes that economic imperfections are inherent; that economic man realizes this as a matter of course; and he or she is willing to

compensate economic agents who act to ameliorate these imperfections.

EMPIRICAL EVIDENCE OF COMPETITION IN THE PHARMACEUTICAL INDUSTRY

The issue of competition in the pharmaceutical industry is implicitly addressed in the works of Cocks and Virts (2, 3), who show a significant lack of price rigidity in various drug markets and among individual drug products. But its clearest discussion is given by Brozen (4):

The Cocks data also destroy the common fiction of rigid prices for drugs and the fiction of inelastic demands for each of these patented products. Prices are remarkably flexible, thus producing large effects on market position. Leading products in the anti-infective market, for example, suffered price declines from 1962 to 1971 ranging from 7% (for product number 8) to 67% (for product number 3). The average price decline in this inflationary period for these products was 32%, while the consumer price index rose 34%. The price of leading anti-infectives fell by 51% in constant dollars. This is a remarkable record.

Sales of these products also demonstrate what a complete fiction is the story that the average physician pays no attention to prices in writing prescriptions. Product 11 among the anti-infectives languished at 0.1% of the market for 5 years until it had cut its price by 47%. At that point, its market share rose to 0.7%, a sixfold increase. Another 14% price cut raised its market share another 170%. Still further cuts over the next three years amounting to 12% raised its market share by still another 68%. This would seem to demonstrate a remarkably high price elasticity of demand for a branded patented product; particularly in view of the price cuts of competitive products.

Product number 3 had a fading market position from 1962 through 1969 "despite its price cuts, but then a 16% price cut in 1970 stopped the decline and added 14% to its market share. A further 27% cut in 1971 jumped its market share by another 40%. The market for ethical drugs responds remarkably vigorously to price changes, the myth of the price-insensitive prescribing physician to the contrary notwithstanding.

There appears to be competition among products within each class despite whatever unique features

each possesses. A product only singular enough to win 0.1% of the market over a five-year span won a 310% increase in market share when it cut its price relative to most of the other products in its market. A fading product turned itself around and reclaimed a major portion of its market position as it undertook similar price action."

One area that has been emphasized in economic theory is that price competition does not exist if a firm or group of firms can charge different prices to different segments of the market. In the pharmaceutical industry, this "market failure" has been emphasized relative to the prices charged to the elderly. It has been claimed that the elderly pay higher prices than the rest of the population. A recent study by Berndt et al. provides statistical evidence that the elderly do not.

A study by Reekie provides a more systematic analysis of pricing behavior regarding pharmaceutical products (5). This study provides a statistically strong inference that physicians are indeed sensitive to drug prices. The paper provides statistical evidence on pharmaceutical product price elasticity in which the coefficient of elasticity is determined to be greater than 1. Schwartzman also provides significant evidence on the amount of price competition in the pharmaceutical industry, especially in the area of antibiotics (6). The elderly generally pay higher prices than the rest of the population, and in some drug categories they pay lower prices (7).

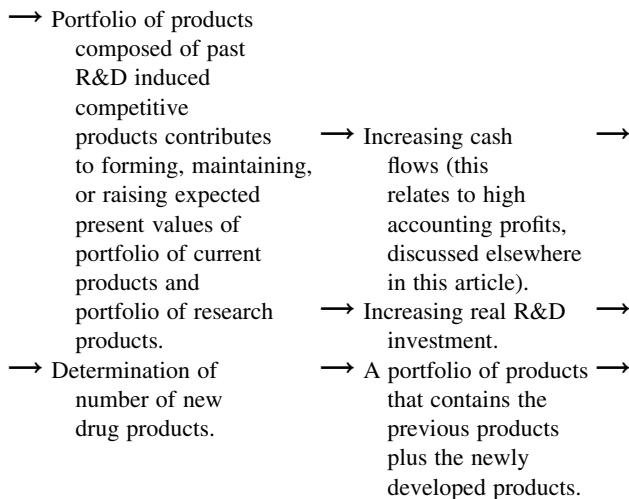
INTERNAL ORGANIZATION CHARACTERISTICS OF THE PHARMACEUTICAL FIRM

The model of the pharmaceutical firm that we have constructed so far leaves us with a fundamental dichotomization of the firm. The dichotomy is between the firm consisting of the production of existing, marketed products, and the production of new products through research and development. The existence of this naturally leads us to question why the firm has to consist of these seemingly different activities. In other words, why could there not be firms who engage in research and development and only produce new products? They could sell these new products on the open market to firms whose specialized function it is to produce and sell the products developed by research firms. In addressing this issue, it is hoped that greater insight into the subtleties of the theory of the pharmaceutical firm can be garnered.

There are two basic elements as to why it is economically efficient to have both characteristics—current product production and new production—present

in one firm. The first element relates to corporate finance and the sources and uses of cash flows in the firm. The second element relates to the efficiencies that can be gained by resorting to markets within the firm (as opposed to external markets that are thought of in conventional microeconomic theory) (8a, 8b, 8c). It should be obvious from the discussion that follows there is an inherent interaction between these elements. This discussion also points out the entrepreneurial function in the firm and its implications for considering the efficiency of the firm.

A characteristic of the pharmaceutical industry, and very likely other R&D-intensive industries, is the interrelationship among new products, the cash flows of the firm, profit expectations, and the utility-enhancing characteristics of new drugs. This flow of economic events can be depicted in the following:



This is a series of events that occurs on a continuum, and the main characteristic is the internally generated cash flows that provide the wherewithal for R&D investment to come from the portfolio of existing products. To provide the necessary cash flows, this portfolio must contain products that have a range of price-marginal manufacturing cost differentials.

The relevance of the two elements, just discussed, can be elaborated on by considering the employment relation that is the primary aspect of the R&D process described above. Pharmaceutical R&D is really an investment in and accumulation of human capital through the employment of scientists and technicians. Like all human capital its “producing” aspects are necessarily embodied in individuals. Unlike normal labor (6) and any associated human capital characteristics that go with it (learning by doing), the human capital associated with pharmaceutical R&D creates

complexities of monitoring and metering work effort. These difficulties exacerbate the contingent claims contracts, bounded rationality, opportunism, and information impactedness problems that would prevail if external markets were used. The use of a hierarchical system clearly presents a less costly alternative. In addition, it is evident from the previous description of the R&D process that the internal market organization allows the combining of the R&D inputs and yields output that is larger than the sum of the products if inputs are used separately.

We can now address the significance of a third element—what can be described as the entrepreneurial, combined element. The six stages of the R&D system process are really the steps that characterize going from invention to innovation, as discussed in the economics of innovation literature. The role of the concept of the entrepreneur is very crucial here. If we view the entrepreneur as the economic visionary, the importance of his or her role is especially apparent in stage 1 of our stage process. At this stage something more than mere “scientific” ability is required. It is also necessary to have the vision to convert “science” or knowledge into a useful product. However, each step of the system process requires entrepreneurial input.

The pharmaceutical firm amalgamates the diverse entrepreneurial activities that make up the complex process from invention through getting a marketable pharmaceutical product. In essence, we are making a distinction between the R&D inputs: scientist and scientist-entrepreneur. In many cases, the scientist does not have the full extent of entrepreneurial ability, and the firm provides the mechanism to achieve this. In addition, when dealing with both the scientist and scientist-entrepreneur, the problems with the Williamson (9) concepts are attenuated; resources are economized because the elements of complexity of contingent claims contracts, bounded rationality, information impactedness, and opportunism are separately prevalent in both the scientists’ and scientist-entrepreneurs’ activities. It is likely that there are distinctive aspects of the Williamson characteristics that are interactive, and this compounds the difficulties and thus makes the internal organization alternative less resource-costly.

In summary, the pharmaceutical R&D process lends itself to the efficiency gains that come from internally organizing these activities. These efficiencies are derived from the existence of the complex technological environment that surrounds the R&D process.

The essence of the theory that we are attempting to apply to the pharmaceutical industry has clearly been outlined by Demsetz. The crucial point is that there are

efficiency gains that are apparent not by comparing them with some ideal, but by comparing them with "real world" alternatives (10).

CONCLUSIONS

The model of the economics of the pharmaceutical industry that is developed here has four basic assumptions:

1. There is price sensitivity on the part of pharmaceutical consumers or, in particular, their agents-physicians, for new products as well as for existing products.
2. Research and development (R&D) serves as the primary catalyst for change among drug firms and is the focal point of entrepreneurial activity that ensures dynamic welfare gains (a continuum of static welfare losses being offset by concomitant higher utility, yielding benefits from new products and systematic erosion of monopoly power through price pressures for older products). As an institutional consideration, there will be a substantial number of firms intensively engaged in R&D activity. In the late 1990s there have been attempts at mega mergers in the industry that would create firms approaching the \$100 billion or more sales amount. These mergers seem to be due to the significant rise in the R&D cost of developing new drugs—possibly exceeding \$500 million.
3. The utility benefits from even small improvements in therapy can theoretically offset substantial differences in the prices of the new improvement relative to existing drug therapies. (This is basically a corollary to assumption 2).
4. The economic profitability of the industry will reflect all dynamic opportunity costs and will through time tend toward normal returns. As such, economic profitability serves as the ultimate guide to the proper allocation of resources as it does with the pure competitive model.

It has been the purpose of this article to apply certain aspects of economic analysis to the pharmaceutical industry. In doing this, we have described a dynamic competitive process that generates new products and

serves as a mechanism that pushes us toward the optimal allocation of resources for the production of existing products. A model of the pharmaceutical firm was also presented. Finally, the welfare implications of the competitive process and the model of the firm were discussed.

REFERENCES

- 1a. Schumpeter, J.A. *Capitalism, Socialism, and Democracy*; Harper & Row, Publishers, Inc.: New York, 1947.
- 1b. Winston, A.P. The Chimera of Monopoly. *The Competitive Economy: Selected Readings*; Brozen, Y., Ed.
2. Cocks, D.L.; Virts, J.R. Pricing Behavior in the Ethical Pharmaceutical Industry. *J. Bus.* **1977**, *47*, 349–362.
3. Cocks, D.L. Product Innovation and the Dynamic Elements of Competition in the Ethical Pharmaceutical Industry. *Drug Development and Marketing*; Helms, R.B., Ed.; American Enterprise Institute: Washington, D.C., 1975; 225–254.
4. Cocks, D.L. Product Innovation and the Dynamic Elements of Competition in the Ethical Pharmaceutical Industry. *Drug Development and Marketing*; Helms, R.B., Ed.; American Enterprise Institute: Washington, D.C., 1975; 225–254.
5. Reekie, W.D. Price and Quality Competition in the United States Drug Industry (Mimeographed). *Pricing New Pharmaceutical Products*; Croom Helm: London, 1977.
6. Schwartzman, D. *Innovation in the Pharmaceutical Industry*; The Johns Hopkins University Press: Baltimore, 1976; 251–299.
7. Berndt is Price Inflation for the Elderly? An Empirical Analysis of Prescription Drugs. *Frontiers of Health Policy Research*; Garber, A., Ed.; National Bureau of Economic Research: Cambridge, MA, 1998.
- 8a. Coase, R.H. The Nature of the Firm. *Economica N. S.* **1937** *November*, *4*, 386–405.
- 8b. Alchian, A.; Demsetz, H. Production, Information Costs, and Economic Organization. *Am. Eco. Rev.* **1972**, *62*, 777–795.
- 8c. Williamson, O.E. *Markets and Hierarchies: Analysis and Antitrust Implications*; The Free Press: New York, 1975; 183–192.
9. Williamson, O.E. *Markets and Hierarchies: Analysis and Antitrust Implications*; The Free Press: New York, 1975; 183–192.
10. Demsetz, H. Information and Efficiency: Another Viewpoint. *J. Law Eco.* **April 1969**, 1–2.

EMULSIONS AND MICROEMULSIONS

Gillian M. Eccleston

University of Strathclyde, Glasgow, United Kingdom

EMULSIONS

An emulsion is a heterogeneous preparation composed of two immiscible liquids (by convention described as oil and water), one of which is dispersed as fine droplets uniformly throughout the other. Emulsions are thermodynamically unstable and revert back to separate oil and water phases by fusion or coalescence of droplets unless kinetically stabilized by a third component, the emulsifying agent. The phase present as small droplets is called the disperse, dispersed, or internal phase and the supporting liquid is known as the continuous or external phase. Droplet diameters vary enormously, but in pharmaceutical emulsions they are typically polydispersed with diameters ranging from approximately 0.1 to 50 μm . Emulsions are conveniently classified as oil-in-water (o/w) or water-in-oil (w/o), depending on whether the continuous phase is aqueous or oily. Fig. 1a shows a photomicrograph of a simple o/w system. Practical pharmaceutical emulsions, however, are rarely simple two-phase oil and water preparations; many are multicomponent systems containing additional solid or liquid crystalline (e.g. lamellar) phases (Fig. 1b). Multiple emulsions, which are prepared from oil and water by the reemulsification of an existing emulsion so as to provide two dispersed phases, are also of pharmaceutical interest. Multiple emulsions of the oil-in-water-in-oil (o/w/o) type are w/o emulsions in which the water globules themselves contain dispersed oil globules; conversely, water-in-oil-in-water (w/o/w) emulsions are those where the internal and external aqueous phases are separated by the oil (Fig. 1c). These more complex emulsions are covered by the broader International Union of Pure and Applied Chemistry (IUPAC) definition of emulsions, which extends the classical definition to include "liquid droplets and/or liquid crystals dispersed in a liquid" (1).

Emulsions are formulated for virtually all the major routes of administration, and there are a number of dermatological, oral and, parenteral preparations available commercially. The internal phase may contain water-soluble drugs, preservatives, and flavoring agents whilst the oil phase may itself be therapeutically active or may act as a carrier for an oil-soluble drug. Such preparations

provide an effective approach to many of the problems in drug delivery, often showing distinct advantages over other dosage forms by way of improved bioavailability and/or reduced side effects. However, despite such advantages, emulsions are not used as extensively as other oral or parenteral dosage forms due to the fundamental problems of emulsion instability that result in unpredictable drug release profiles and possible toxicity. The full potential of emulsions will not be realized until stable systems are developed with predictable *in vitro* and *in vivo* release patterns. Much of the emulsion research over the past decade is based on attempts to understand the relationships between emulsion stability, physicochemical properties, and biological fate. Multiple emulsions are even more difficult to stabilize, and characterize and although there is an increasing interest in their potential applications for drug delivery, at present there are no commercial preparations available (2).

PHARMACEUTICAL APPLICATIONS

The current and potential pharmaceutical applications of emulsions have been the subject of a number of general reviews (3–6). Traditionally the term "emulsion" is restricted to mobile emulsions for internal use; emulsions for external use are described by their pharmaceutical types as liniments, lotions, and creams. This tends to conceal the fact that by far the largest group of emulsions currently used in pharmacy and medicine are dermatological emulsions for external use (7, 8). Both oil-in-water and water-in-oil emulsions are extensively used for their therapeutic properties and/or as vehicles to deliver drugs and cosmetic agents to the skin. The emulsion facilitates drug permeation into and through the skin by its occlusive effects and/or by the incorporation of penetration-enhancing components. Particular attention is paid to patient acceptance of such formulations, which range in consistency from mobile liniments and lotions to semisolid ointments and creams. In the past, the development of dermatological emulsions was essentially empirical with only a limited understanding of the

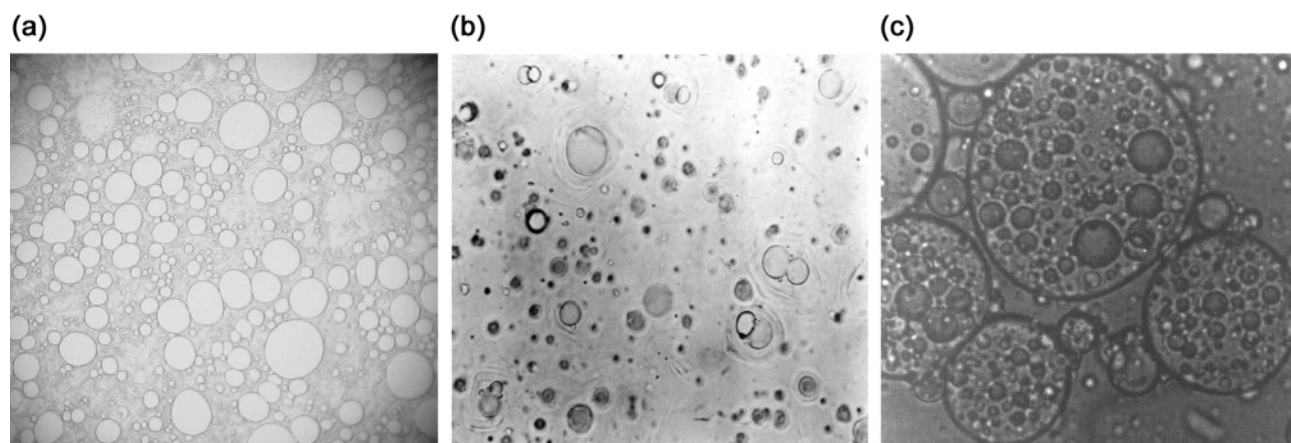


Fig. 1 Photomicrographs of typical emulsions. (a) A liquid paraffin-in-water emulsion stabilized by 1.89% Span 40 and 1.62% Tween 80. The polydispersity of the oil droplets before homogenization is clearly seen. (b) A liquid paraffin–water cream stabilized by a cationic emulsifying wax. Note the lamellar structures surrounding the oil droplets. (c) A multiple w/o/w emulsion. Water droplets can clearly be seen within the larger oil droplets.

underlying principles. Today, although the microstructure of many of these complex formulations is now better understood (9, 10), the mechanisms by which the structure of an emulsion can influence drug bioavailability are far from clear and much of the literature on the role of emulsions in drug release to the skin is contradictory. Confusion arises because the majority of investigations concerning in vitro vehicle effects on drug release are only on bulk formulation. As most emulsions are applied to the skin as a thin film, the drug delivery system is not one of bulk emulsion, but rather a dynamic evaporating system in which phase changes can occur as the preparation is rubbed into the skin and the relative concentrations of volatile ingredients alter. Droplet size appears to influence drug delivery to the skin, with submicron lipid emulsions enhancing the transcutaneous permeation and efficiency of a number of lipophilic drugs (11).

Oral emulsions are almost exclusively of the oil-in-water type. They provide a degree of taste masking as the aqueous external phase effectively isolates the oil from the tongue. Mineral and castor oils have been emulsified in water and administered orally for the local treatment of constipation for many years (cf. Mineral Oil Emulsion USP) as have various nutritional oils from fish liver (generally halibut or cod) or vegetable origin to produce oral liquid food supplements. It has long been established that the use of o/w emulsions as carriers for lipophilic drugs may improve oral bioavailability and efficacy (3–6). For example, griseofulvin formulated as an o/w emulsion has enhanced gastrointestinal absorption when compared with suspensions, tablets, or capsule dosage forms (12). The mechanisms by which emulsions modify and improve

absorption processes are complex and not fully understood, although the oil itself influences gastric motility. Fats and oils are solubilized by the bile salts so that the administration of already emulsified oil droplets containing a high concentration of drug may increase the likelihood of further droplet and drug solubilization and transport across the GI tract by the fat absorption pathways.

The type of emulsion used parenterally depends on the route of injection and the intended use (13–15). Oil-in-water emulsions are administered by all the major parenteral routes whereas water-in-oil emulsions are generally reserved for intramuscular or subcutaneous administration where sustained release is required. Drug action is prolonged in such oily emulsions because the drug has to diffuse from the aqueous dispersed phase through the oil-continuous environment to reach the tissue fluids. Water-in-oil emulsions are used to disperse water-soluble immunizing antigens in mineral oil for injection via subcutaneous or intramuscular routes as adjuvant preparations where they prolong and enhance the antigenic stimulus and increase the antibody titer. Oily emulsion formulations also show promise in cancer chemotherapy as vehicles for prolonging drug release after intramuscular or intratumoral injection, and as a means of enhancing the transport of anticancer agents via the lymphatic system (16). Water-in-oil emulsions for sustained release are often difficult to inject because of the high viscosities of the oily continuous phases. Although these problems can be overcome by reemulsification of the primary w/o emulsion to produce a less viscous multiple w/o/w emulsion, a study using

Table 1 Some commercial lipid emulsions for parenteral nutrition

Trade name	Oil phase (%)	Emulsifier (%)	Other components (%)
Intralipid [®] (Fresenius Kabi)	Soybean (10 and 20)	Egg lecithin (1.2)	Glycerol (2.2), phosphate (15mm/l)
Lipovenos [®] (Fresenius Kabi)	Soybean (10 and 20)	Egg lecithin (1.2)	Glycerol (2.5)
Liposyn [®] (Abbott)	Safflower and soybean, 1:1 (10 and 20)	Egg lecithin (1.2)	Glycerol (2.5)
Lipofundin [®] (Braun)	Cottonseed (15)	Soybean lecithin (0.75)	Sorbitol (5.0), DL- α -Tocopherol
Lipofundin N [®] (Braun)	Soybean (10 and 20)	Egg lecithin (0.75 and 1.2)	Glycerol (2.5)
Lipofundin MCT/LCT [®] (Braun)	Soybean and MCT, 1:1 (10 and 20)	Egg lecithin (0.75 and 1.2)	Glycerol (2.5)

5-fluorouracil implied that sustained release was actually less marked with multiple emulsions (17).

Sterile parenteral oil-in-water emulsions have been used extensively for over 40 years for the intravenous administration of fats, carbohydrates, and vitamins to debilitated patients. Several vegetable oil-in-water emulsions are now available commercially with droplet sizes similar to that of chylomicrons (approximately 0.5–2 μm), the natural fat droplets in the blood that transport ingested fats to the lymphatic and circulatory systems (Table 1). More recently, such emulsions have been employed as intravenous carriers for poorly water-soluble lipophilic drugs such as vitamin K (e.g., Sterile Phytonadione Injection U.S.P.) diazepam (e.g., Diemuls[®]), vitamin A (Vitlipid N[®]), and profenol (Diprovan[®]) as alternatives to the traditionally used cosolvent, surfactant solubilized, or pH controlled parenteral solutions. The drug dissolved in the oil phase of the emulsion is unlikely to precipitate and cause pain when diluted by blood on injection, and if susceptible to hydrolysis or oxidation, it will be protected by the nonaqueous environment. Emulsion formulations of diazepam and more recently clarithromycin have been clinically shown to be less painful than solubilized preparations (18, 19) while emulsions containing amphotericin B are less toxic (20). This emulsion was also shown to be an equally effective, cheaper, and more elegant alternative to a liposomal system. The enormous literature on the potential of lipid emulsions for drug delivery and targeting is discussed in a recent book (21).

Radiopaque emulsions, which have long been used as contrast media in conventional X-ray examinations of body organs, are finding further application with more sophisticated techniques including computed tomography, ultrasound, and nuclear magnetic resonance. Perfluorochemical emulsions are used as artificial blood

substitutes. The potential advantages of such systems over donated blood are enormous with the elimination of major donor associated problems such as blood group incompatibilities and blood disease. The first commercial product, Fluosol-DA[®] (Green Cross Corporation, Osaka, Japan) was licensed several years ago in a number of countries to reduce myocardial ischaemia in patients undergoing angioplasty; however, Fluosol-DA was not a commercial success due to its slow excretion rate and to its marked instability, which meant that it had to be stored in the frozen state. In addition, some patients were sensitive to one of the emulsifiers, pluronic F68[®]. Currently, a second generation of emulsions is being evaluated to resolve the problems encountered with Fluosol (22) and these are discussed in another chapter of this encyclopedia.

There are only a few studies on the ocular and nasal applications of emulsions. Lipid (submicron) emulsions exhibited a long-lasting antidepressant effect on the intraocular pressure of rabbits after a single application when used as carriers for lipophilic antiglaucoma drugs (23). Medium-chain triglyceride emulsions formulated at pH 8 show potential as controlled release formulations for nasal delivery (24, 25) for they give prolonged drug residence in the nasal cavity (Fig. 2). Enhanced nasal delivery of insulin was observed when insulin was incorporated into the continuous phase of an o/w emulsion, but not when incorporated into the aqueous phase of a w/o emulsion (26).

FORMULATION CONSIDERATIONS

The choice of oil, emulsifier, and emulsion type (o/w, w/o, or multiple) is limited by its ultimate use and route of

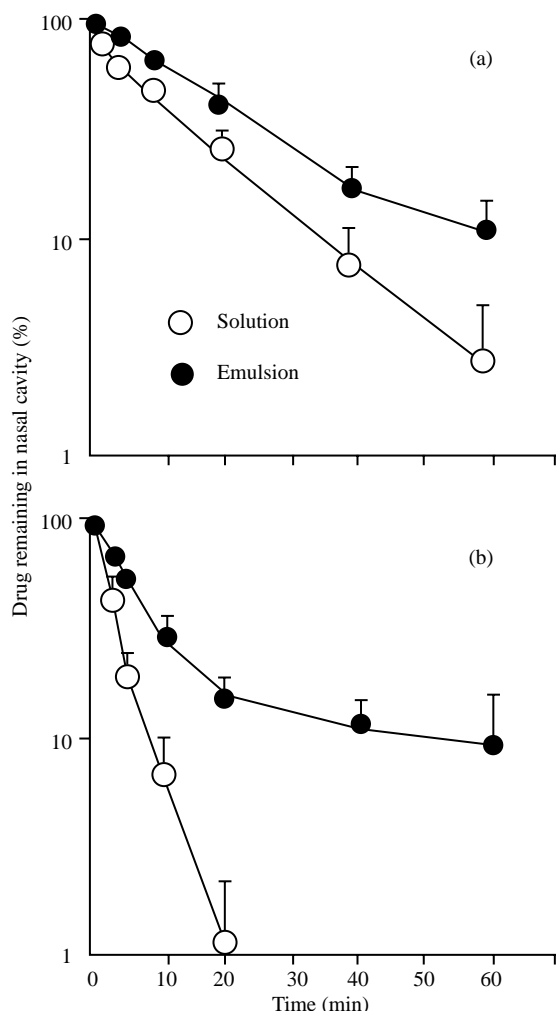


Fig. 2 Disappearance profiles of (a) tetrahydrozoline hydrochloride and (b) chlorpheniramine maleate from rat nasal cavity after nasal administration of an o/w emulsion and an aqueous solution at pH 8. (From Ref. 25.)

administration. Potential toxicity and chemical incompatibilities in the final formulation must be taken into account as must processing details for these also affect the variables that control emulsion stability and therapeutic response such as droplet size distributions and rheology. The design of stable emulsions with the correct pharmacokinetic characteristics and tissue distribution is currently an area of enormous interest, particularly for parenteral IV emulsions. Immediately after injection, the surface of the parenteral emulsion droplets is altered by adsorption of blood components (optosonation) and they are then distributed rapidly through the circulation. Their subsequent fate depends on whether they are treated by the body in the same manner as chylomicrons, or whether they

are recognized as foreign particles and cleared by the RES. Many factors, including droplet size and charge, the type of lipid, and the emulsifier composition influence their fate. A major factor to be considered in the formulation of oral preparations is the low pH and high ionic strength of stomach fluids, which may destabilize the emulsion by its effect on the emulsifier.

Pharmaceutical Oils

Oils used in the preparation of pharmaceutical emulsions are of various chemical types, including simple esters, fixed and volatile oils, hydrocarbons, and turpenoid derivatives. The oil itself may be the medicament, it may function as a carrier for a drug, or even form part of a mixed emulsifier system as in the case of some fixed oils that contain sufficient free fatty acids. Many oils, particularly those of vegetable origin, are liable to autooxidation with subsequent rancidity, and it is frequently necessary to add an antioxidant and/or preservative to inhibit this degradation process. For externally applied emulsions, mineral oils, either alone or combined with soft or hard paraffins, are widely used both as the vehicle for the drug and for their occlusive and sensory characteristics. The most widely used oils in oral preparations are nonbiodegradable mineral and castor oils that provide a local laxative effect, and fish liver oils or various fixed oils of vegetable origin (e.g., arachis, cottonseed, and maize oils) as nutritional supplements.

The choice of oil is severely limited in emulsions for parenteral administration for reasons of toxicity. Purified soybean, sesame, safflower, and cottonseed oils composed mainly of long-chain triglycerides have been used for many years as they are resistant to rancidity and show few clinical side effects. More recently, it has been recognized that the structure of the oil will influence the fate of emulsion droplets after injection. Mixtures containing both long- and medium-chain triglycerides are not only better energy sources for nutritional purposes but they are also cleared more rapidly from the circulation (27); such mixtures are now used in commercial preparations (c.f. Table 1). Structured triglycerides, formed by modifying the oil enzymatically to produce 1,3-specific triglycerides are an area of increasing interest because of their influence on the in vivo circulation time of an emulsion (28). Purified mineral oil is used in some water-in-oil depot preparations where mineral toxicity (e.g., abscess formation at the injection site) must be carefully balanced against efficiency. Emulsified perfluorochemicals are considered acceptable for IV use provided that they are

excreted relatively fast. A major problem in the formulation of the early perfluorocarbon emulsions was that the oils that form the most stable emulsions were not cleared rapidly from the body.

Pharmaceutical Emulsifiers

Emulsifying agents are used both to promote emulsification at the time of manufacture and to control stability during a shelf life that can vary from days for extemporaneously prepared emulsions to months or years for commercial preparations. In practice, combinations of emulsifiers rather than single agents are used. The emulsifier also influences the *in vivo* fate of lipid parenteral emulsions by its influence on the surface properties of the droplets and on the droplet size distributions. For convenience, most pharmacy texts classify emulsifiers into three groups: i) surface active agents, ii) natural (macromolecular) polymers, and iii) finely divided solids.

Surface Active Agents

The range of surfactant emulsifiers used in pharmaceutical preparations is illustrated in Table 2. Surfactants are manufactured from a variety of natural and synthetic sources and consequently they show considerable batch-to-batch variations in their homologue compositions and in trace impurities from the starting material. For example, batch variations in the number of neutral phospholipids occur in lecithin surfactants and nonionic polyethylene surfactants show variations in the number of moles of ethylene oxide. The mechanisms by which such batch variations lead to differences in emulsifying properties are now better understood (29).

Although synthetic and semisynthetic surfactants form by far the largest group of emulsifiers studied in the scientific literature and many of them are available commercially, their use in pharmaceutical emulsions is limited by the fact that the majority are toxic (i.e., haemolytic) and irritant to the skin and mucous

Table 2 Synthetic surface active emulsifying agents

Class	Example	Type	Compatibility
Anionic			Efficient to various degrees above pH 7; incompatible with cationic surfactants and polyvalent cations
Alkali and ammonium soaps	Sodium stearate	o/w	
Divalent and trivalent metallic soaps	Calcium oleate	w/o	
Organic sulfates	Sodium lauryl sulfate	o/w	
Cationic			Efficient below pH 7; incompatible with anionic surfactants and polyvalent anions
Quaternary ammonium compounds	Cetrimonium bromide	o/w	
Pyridium compounds	Hexadecyl pyridinium chloride	o/w	
Nonionic			Efficient to various degrees over pH range 4–8; good tolerance to ionic surfactants and polyvalent ions
Alcohol polyethylene glycol ethers	Ceteth 20	o/w	
Fatty acid polyethylene glycol esters	Polyethylene glycol 40 stearate	o/w	
Ethoxylated fatty acid polyethylene glycol esters	Sorbitan mono-oleate (Span 80)	w/o	
	Polyoxyethylene sorbitan monooleate (Tween 80)	o/w	
Polymeric			
Polyoxyethylene-polyoxypropylene block co-polymers	Poloxomers, Pluronic F-68®	o/w	
Amphiphiles			Generally used combined with a surfactant to form a o/w emulsion
Fatty alcohols	Cetyl alcohol	w/o	
Fatty acids	Stearic acid	w/o	

membranes of the gastrointestinal tract. In general, cationic surfactants are the most toxic and irritant and nonionic surfactants the least. Surfactants are therefore used mainly at relatively low concentrations in topical preparations. The quaternary ammonium compounds constitute an important group of cationic emulsifiers in dermatological preparations because they have antimicrobial properties in addition to their o/w emulsifying action. There are many nonionic surfactants with different oil and water solubilities available commercially because for each fatty starting material the polyoxyethylene chain length can be modified by the systematic addition of ethylene oxide groups. However, a limited number of polysorbate surfactants are used in oral emulsions, and parenteral preparations appear to be based only on the lecithins from plant or animal sources and the nonionic polyoxyethylene oxide/polyoxypropylene oxide block copolymer poloxamer 188 (Poloxamer F68[®]), although some patients using the first generation of perfluorochemical emulsions were sensitive to this poloxamer. The emulsifier influences both emulsion stability and in vivo disposition by its influence on droplet surface properties.

Natural macromolecular materials and finely divided solids

Materials derived from natural sources (Table 3) may originate from animal or vegetable sources and many of these products are susceptible to degradation. For example, depolymerization (the polysaccharides) or hydrolysis (the steroids) usually lead to loss in emulsifying power. Some of these materials, polysaccharides and proteins in particular, provide good culture medium for microorganisms, and therefore preservation of emulsions containing them is imperative. To overcome

these problems, a number of purified and semisynthetic derivatives are available, including various purified wool fat derivatives and semisynthetic celluloses such as methylcellulose and sodium carboxymethylcellulose. These are generally more stable than the unmodified materials. These celluloses are used in oral preparations; they are less suitable for topicals because of their unpleasant feel. Finely divided solids such as clays are used in dermatological preparations as structuring agents.

Preservatives

It is essential that emulsions are formulated to resist microbial attack, as this not only can affect the physicochemical properties of the formulation, causing color, odor, or pH changes and even phase separation, but may also constitute a health hazard. The potential sources of contamination can be from raw materials (especially if these are natural products), water, manufacturing and packaging equipment, or patients themselves. W/o emulsions are less susceptible to attack than o/w emulsions because the aqueous continuous external phase can produce ideal conditions for the growth of bacteria, moulds, and fungi. Preservatives are not used in parenteral emulsions, which are sterilized, generally by autoclaving, but sometimes by using sterile components and aseptically assembling the final emulsion.

There is no simple way of predicting the ideal preservative for a particular emulsion. In addition to requiring a wide spectrum of activity against bacteria, yeasts, and molds, the preservative should be free from toxic, irritant, or sensitizing activity. Some commonly used preservatives in oral and topical preparations include phenoxyethanol, benzoic acid, parabenzates, and chlor-cresol. Emulsions are heterogeneous products, and the

Table 3 Emulsifying agents derived from natural products and finely divided solids

Class	Example	Emulsion type, route of administration	Comments
Polysaccharide	Acacia	o/w; oral	Stable over a wide pH range
	Carageen	o/w; oral	As above
	Methylcellulose	o/w; oral, parenteral	As above, less prone to hydrolysis
Protein	Gelatin	o/w; oral,	Emulsifying properties pH dependent
Glycoside	Saponin	o/w; topical	
Phospholipid	Lecithin	o/w; oral, parenteral	Emulsifying properties dependent on number of negative lipids
Sterol	Wool fat	w/o; topical	Poor emulsifiers alone
	Cholesterol and its esters	w/o; topical	As above
Finely divided solids	Bentonite	o/w and w/o; topical	Gelation dependent on processing conditions
	Veegum	o/w; oral, topical	As above
	Aluminium hydroxide	o/w; oral	

preservative partitions between the oil and aqueous phases. As a sufficient aqueous concentration of the active (usually unionized) form must be present to ensure proper preservation, pH is an additional factor to be considered. Problems often arise because many of the materials used in emulsion formulation, for example hydrocolloids or polyoxyethylene surfactants, can interact with the preservatives, thus depleting their activity. The use of a single preservative is often considered unrealistic, and attention is being increasingly focused on the use of mixtures for a wider spectrum of activity, although this may introduce additional compatibility problems.

Antioxidants and Humectants

Antioxidants are added to many pharmaceutical preparations to prevent oxidative deterioration on storage of the oil, emulsifier, or the drug itself. Such deterioration, as well as destabilizing the formulation, imparts an unpleasant odor or taste. Some oils are supplied containing antioxidants already. Those commonly used in pharmacy include butylated hydroxyanisole (BHA) and butylated hydroxytoluene (BHT) at concentrations up to 0.2%, and the alkyl gallates. Humectants such as propylene glycol, glycerol, and sorbitol (5%) are often added to dermatological preparations to reduce the evaporation of water from the emulsion during storage and use. They are sometimes claimed to prevent evaporation of water from the surface of the skin, although their use at high concentrations would be expected to have the opposite effect (i.e., remove moisture and dehydrate the skin).

EMULSION FORMATION

There are essentially two major considerations in emulsification: first, the formation of emulsions of the correct type, oil-in-water, water-in-oil, or multiple emulsion with the required droplet size distributions and second, the stabilization of the dispersed droplets so formed. When given amounts of two immiscible liquids are mixed or mechanically agitated in the absence of other additives, both phases tend to form droplets of various sizes. The size distributions are related to the forces involved during the agitation process, and the number of droplets of each liquid depends on its relative volume. The surface free energy of the system, which is dependent on both total surface area and interfacial tension is raised by the increase in surface area produced during dispersion, and the system is thermodynamically unstable. To reduce this, high-energy droplets first assume a spherical shape, as this gives the

minimum surface area for a given volume, and then on collision the droplets will coalesce (fuse) to reduce the interfacial area, the interfacial tension remaining constant.

The type of emulsion that forms, either o/w or w/o, depends on the relative rates of coalescence of each type of droplet, with the more rapidly coalescing droplets forming the continuous phase. Generally, this is the liquid present in the larger amount because higher number of droplets increase the probability of collision and coalescence. With phase volumes of oil and water close to 50%, other factors such as the order and rate of addition of each liquid are important. If agitation ceases, coalescence will continue until complete phase separation—the state of minimum free energy—is reached. Thus, emulsification can be considered as the result of two competing processes, namely the disruption of bulk liquids to produce fine droplets and the recombination of the droplets to give back the original bulk liquids. With the inclusion of surfactants or other classes of emulsifiers, the type of emulsion that forms is no longer simply a function of the phase volume and the order of mixing, but also the relative solubilities of the emulsifier in oil and water. In general, the phase in which the emulsifier is most soluble becomes the continuous phase (Bancroft's Rule); thus, hydrophilic polymers and surfactants promote o/w emulsions whereas lipophilic surfactants promote w/o emulsions. Preferential coalescence of the phase in which the emulsifier is most soluble occurs because when droplets collide, the emulsifier is easily displaced from the interface into the droplet, thus providing little protection against coalescence. Theoretically, the disperse phase of an emulsion can occupy up to 74% of the phase volume, and such high internal phase o/w emulsions have been produced with suitable surfactants. It is more difficult to formulate w/o emulsions with greater than 50% disperse phase because of the steric mechanisms involved in their stabilization (discussed later), and the addition of extra water sometimes causes inversion to an o/w emulsion.

Emulsion Characteristics

In general, an emulsion exhibits the characteristics of its external phase. Several methods are available for identifying the emulsion type. Dilution tests are based on the fact that the emulsion is only miscible with the liquid that forms its continuous phase. Conductivity measurements rely on the poor conductivity of oil compared with water, and give low values in w/o emulsions where oil is the continuous phase. Staining tests in which a water-soluble dye is sprinkled onto the surface of the emulsion also indicate the nature of the continuous phase. With an o/w emulsion there is rapid incorporation of the dye into

the system whereas with the w/o emulsion the dye forms microscopically visible clumps. The reverse happens on addition of an oil-soluble dye. These tests essentially identify the continuous phase and do not indicate whether a multiple emulsion has been produced. This can be resolved by microscopy.

Rheology

The rheological properties of emulsions are influenced by a number of interacting factors, including the nature of the continuous phase, the phase volume ratio, and to a lesser extent, particle size distributions. A variety of products ranging from mobile liquids to thick semisolids can be formulated by altering the dispersed phase volume and/or the nature and concentration of the emulsifiers. For low internal phase volume emulsions, the consistency of the emulsion is generally similar to that of the continuous phase; thus, w/o emulsions are generally thicker than o/w emulsions, and the consistency of an o/w system is increased by the addition of gums, clays, and other thickening agents that impart plastic or pseudoplastic flow properties. Some mixed emulsifiers interact in water to form a viscoelastic continuous phase to give a semisolid o/w cream (7).

Droplet Size Distributions

Droplet size distributions in pharmaceutical emulsions are important from both stability and biopharmaceutical considerations. The larger the particle size, the greater the tendency to coalesce and further increase droplet size. Thus, fine particles generally promote better stability. Size distributions are influenced by the characteristics of the emulsifier and the method of manufacture. From a biological point of view, fine emulsification enhances gastrointestinal absorption and whilst this is desirable with oral formulations of nutrient oils alone or with drugs dissolved in them, it may give adverse clinical effects with mineral oils that are used for a local effect and are toxic if absorbed. Droplets in emulsions used as contrast media in computed tomography are approximately 1–3 μm . Parenteral emulsions should be formulated so that the dispersed droplet sizes range from approximately 100 nm to 1 μm . In any event, sizes should never be greater than 5 μm in diameter because of the danger of pulmonary emboli. There is clear evidence that, as with other colloidal preparations, droplet size distributions influence the clearance kinetics of parenteral emulsions. Larger droplets are treated as foreign bodies and rapidly cleared by elements of the RES while smaller droplets may be treated

as natural fat sorting lipoproteins, with a different in vivo fate (30). Drug delivery from dermatological preparations also appears to be improved in lipid emulsions containing submicron droplets (11).

EMULSION STABILITY

A stable emulsion is considered to be one in which the dispersed droplets retain their initial character and remain uniformly distributed throughout the continuous phase for the desired shelf life. There should be no phase changes or microbial contamination on storage, and the emulsion should maintain elegance with respect to odor, color, and consistency. Instabilities of both chemical and physical origins can occur in emulsion formulations. Chemical instabilities, such as the development of rancidity in natural oils due to oxidation by atmospheric oxygen, the depolymerization of macromolecular emulsifiers by hydrolysis, or microbial degradation can be minimized by the addition of suitable antioxidants and preservatives. More general chemical instabilities involving interactions between the drug and emulsion excipients or between the excipients themselves may lead to physical instabilities. If these interactions involve the emulsifying agent, they may destroy its emulsifying properties, causing the emulsion to break. For example, interactions between phenolic preservatives and polyoxyethylene nonionic emulsifiers may lead to loss of emulsifying power as well as poor preservation.

Physical Instability

As emulsions are inherently unstable, they eventually revert to the original state of two separate liquids, that is, will break or crack. In the presence of an emulsifier and other additives, this state is approached via several distinct processes, some of which are reversible such as creaming and flocculation and others irreversible such as coalescence and Ostwald ripening. Phase inversion when an oil-in-water emulsion inverts to form a water-in-oil emulsion or visa versa is a special case of irreversible instability that occurs only under well-defined conditions such as a change in emulsifier solubility due to specific interactions with additives or to a change in temperature (Fig. 3).

Flocculation describes a weak reversible association between emulsion globules separated by thin films of continuous phase. The individual droplets retain their separate identities, but each floccule or cluster of droplets behaves physically as a single unit. The association arises from the interaction of attractive and repulsive forces

between droplets and is reversible in the sense that the original dispersion can generally be regained by mild agitation. Flocculation is generally regarded as a precursor to the irreversible process of coalescence, although sometimes the time scale between flocculation and coalescence can be extended almost indefinitely by the adsorbed emulsifier, giving a kinetically stable emulsion.

Coalescence, where dispersed phase droplets merge to form larger droplets, takes place in two distinct stages. It begins with the drainage of liquid films of continuous phase from between the oil droplets as they approach one another and ends with the rupture of the film when a critical thickness is reached. The approaching droplets may deform as the opposing surfaces distort to either flatten (small droplets) or dimple (larger droplets) under the hydrodynamic pressures generated by viscous flow of the continuous phase.

Coalescence is not the only mechanism by which dispersed phase droplets increase in size. If the emulsion is

polydispersed and there is significant miscibility between the oil and water phases, then Ostwald ripening, where droplet sizes increase due to large droplets growing at the expense of smaller ones, may also occur. This destabilizing process is a result of the Kelvin effect and occurs when small emulsion droplets (less than 1 μm) have higher solubilities (and vapor pressures) than do larger droplets (i.e. the bulk material) and consequently are thermodynamically unstable. To reach the state of equilibrium, molecules from these droplets dissolve and diffuse through the continuous phase to enlarge the larger droplets. As the small droplets lose their oil, they become even smaller, the vapor pressure difference increases, and Ostwald ripening is further enhanced.

Creaming or sedimentation occurs when the dispersed droplets or floccules separate under the influence of gravity to form a layer of more concentrated emulsion, the cream. Generally a creamed emulsion can be restored to its original state by gentle agitation. This process, which

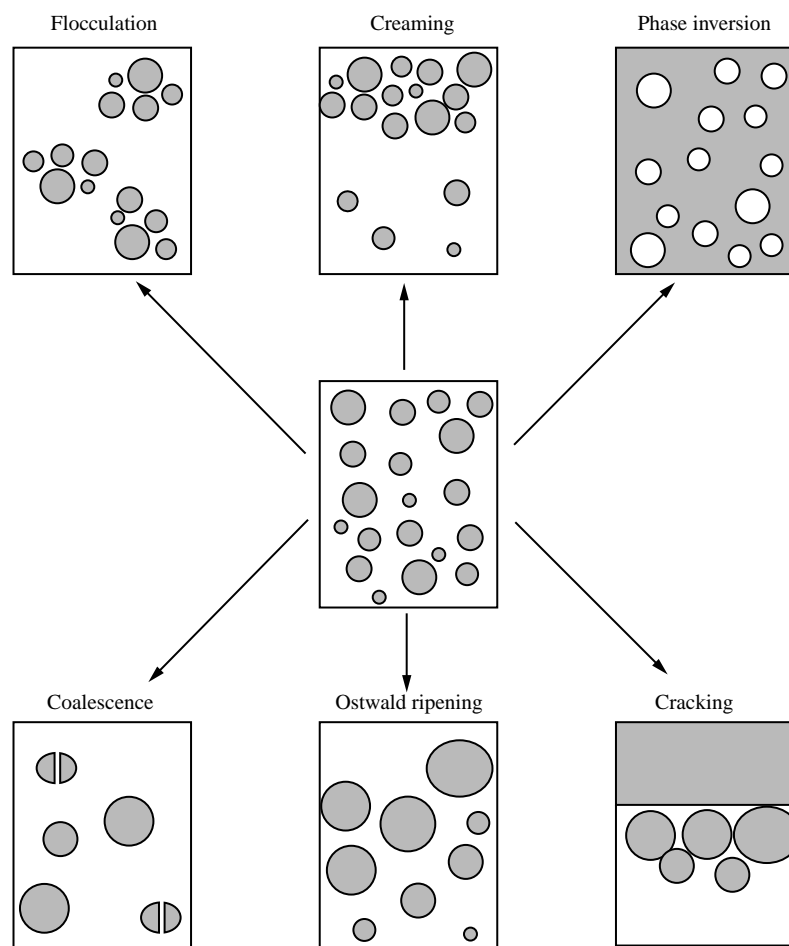


Fig. 3 Schematic representation of the various processes of emulsion breakdown.

inevitably occurs in any dilute emulsion if there is a density difference between the phases as a consequence of Stokes law, should not be confused with flocculation which is due to particle interactions resulting from the balance of attractive and repulsive forces. Most oils are less dense than water so that the oil droplets in o/w emulsions rise to the surface to form an upper layer of cream. In w/o emulsions, the cream results from sedimentation of water droplets and forms the lower layer. According to Stokes Law, the rate of creaming can be minimized by reducing droplet sizes and/or thickening the continuous phase. Adjustment of the densities of the two phases has received little attention.

The destabilization processes are not independent and each may influence or be influenced by the others. For example, the increased droplet sizes after coalescence or Ostwald ripening will enhance the rate of creaming, as will the formation of large floccules which behave as single entities. In practice, creaming, flocculation, and Ostwald ripening may proceed simultaneously or in any order followed by coalescence.

Coalescence and Ostwald ripening are obviously the most serious types of instability as they result in the formation of progressively larger droplets and ultimately lead to phase separation. Creaming and flocculation, on the other hand, are more subtle forms of instability, for although they represent potential steps towards coalescence and breaking due to the close proximity of the droplets, many practical emulsions remain in this state for long periods of time without significant coalescence and can be redispersed simply by shaking the container.

EMULSION STABILIZATION

Emulsifiers stabilize emulsions in a number of different ways, all of which act to prevent or delay the various destabilization processes described previously. The emulsifier may form an interfacial film at the oil–water interface and/or structure (i.e., thicken) the continuous phase. The interfacial film introduces additional repulsive (e.g., electrostatic, steric, or hydrational) forces between droplets to counteract attractive van der Waals forces and inhibit the close approach of droplets. It may also provide a barrier to the coalescence of droplets in close proximity, particularly if the film is close-packed and elastic. Surfactant interfacial films also lower the interfacial tension between oil and water. Although this effect is important during the emulsification process where it facilitates the breakup of droplets, it is not a major factor in maintaining the long-term stability of emulsions. In emulsions that are thickened

by the emulsifier, the interfacial film does not play the dominant role in maintaining stability; rather, it is the structured continuous phase that forms a rheological barrier to prevent the movement and hence the close approach of droplets and also inhibits Ostwald ripening.

Classical (Interfacial) Theories

Classical theories of emulsion stability focus on the manner in which the adsorbed emulsifier film influences the processes of flocculation and coalescence by modifying the forces between dispersed emulsion droplets. They do not consider the possibility of Ostwald ripening or creaming nor the influence that the emulsifier may have on continuous phase rheology. As two droplets approach one another, they experience strong van der Waals forces of attraction, which tend to pull them even closer together. The adsorbed emulsifier stabilizes the system by the introduction of additional repulsive forces (e.g., electrostatic or steric) that counteract the attractive van der Waals forces and prevent the close approach of droplets. Electrostatic effects are particularly important with ionic emulsifiers whereas steric effects dominate with nonionic polymers and surfactants, and in w/o emulsions. The applications of colloid theory to emulsions stabilized by ionic and nonionic surfactants have been reviewed as have more general aspects of the polymeric stabilization of dispersions (4, 31, 32).

The DLVO theory, which was developed independently by Derjaguin and Landau and by Verwey and Overbeek to analyze quantitatively the influence of electrostatic forces on the stability of lyophobic colloidal particles, has been adapted to describe the influence of similar forces on the flocculation and stability of simple model emulsions stabilized by ionic emulsifiers. The charge on the surface of emulsion droplets arises from ionization of the hydrophilic part of the adsorbed surfactant and gives rise to electrical double layers. Theoretical equations, which were originally developed to deal with monodispersed inorganic solids of diameters less than 1 μm , have to be extensively modified when applied to even the simplest of emulsions, because the adsorbed emulsifier is of finite thickness and droplets, unlike solids, can deform and coalesce. Washington (33) has pointed out that in lipid emulsions, an additional repulsive force not considered by the theory due to the solvent at close distances is also important.

The theory states that the forces between droplets can be considered as the sum of an attractive van der Waals part V_A and a repulsive electrostatic part V_R when identical electrical double layers overlap. As the origin of each force is independent of the other, each is evaluated separately, and the total potential of interaction V_T between the two

droplets as a function of their surface-to-surface separation is obtained by summation

$$V_T = V_A + V_R$$

A schematic potential energy of interaction with distance plot is shown in Fig. 4a. It can be seen that a weak attraction occurs at large droplet separations represented by the secondary energy minimum, and a very strong attraction at small droplet separations hence the very deep primary minimum. At intermediate distances, double-layer repulsion dominates and there is

a maximum in the curve. Flocculation occurs in the secondary minimum, where the attractive forces are relatively weak and floccules are easily separated by low energy agitation. Once flocculated, droplets are prevented from approaching closer by the potential energy barrier. If they have sufficient energy to overcome the barrier, the process of coalescence commences as the droplets move closer together. Once in the primary minimum the aggregates formed are separated by only a small distance so that stability against coalescence is determined by the resistance of the interfacial film to rupture.

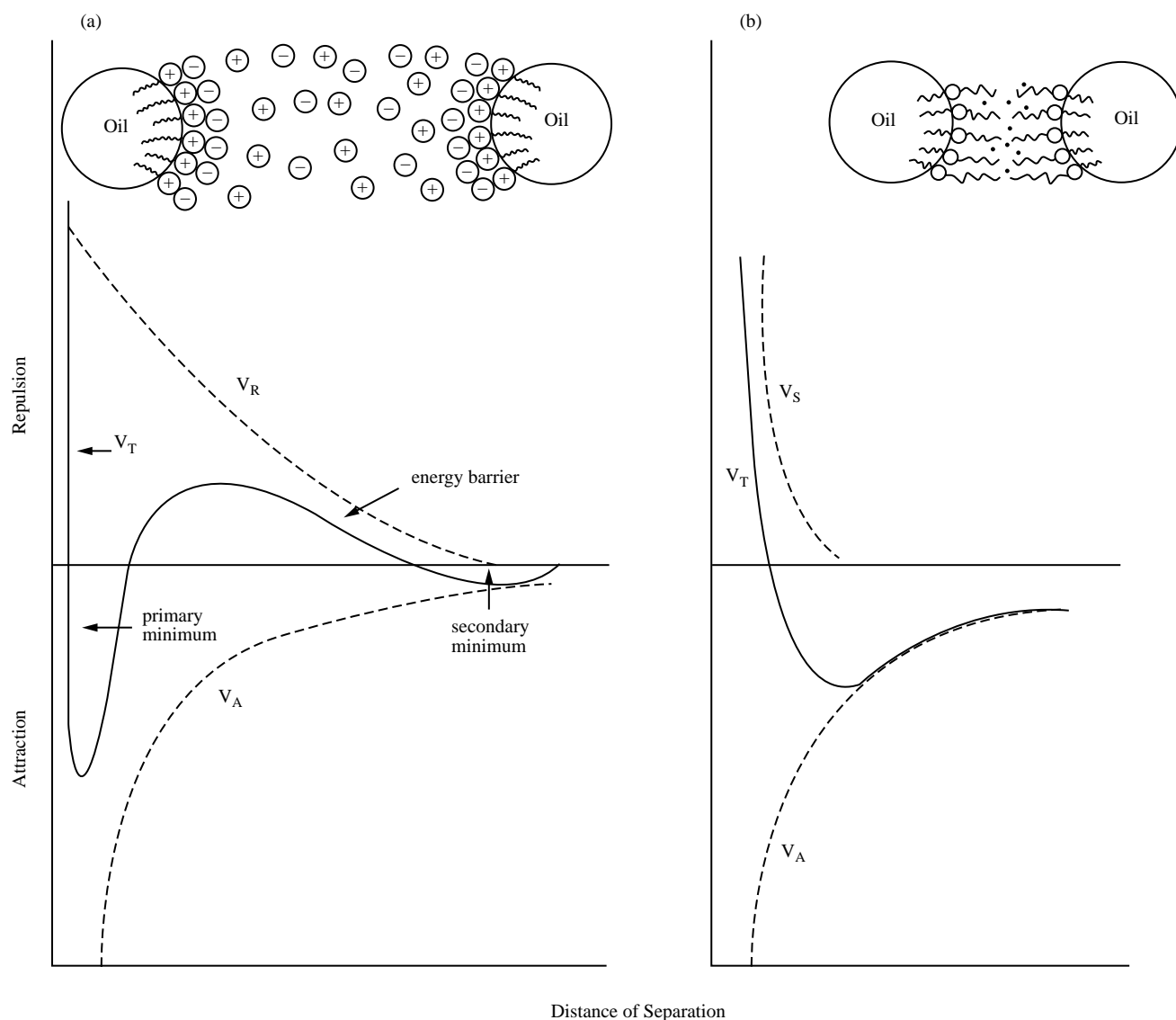


Fig. 4 The total potential energy of interaction V_T as a function of distance of surface separation H for two similar oil droplets in an oil-in-water emulsion. (a) Electrostatic stabilization by a monolayer of ionic surfactant. (b) Steric stabilization by a monolayer of nonionic surfactant. V_A : van der Waals attractive force; V_R : electrostatic repulsive force; V_S : steric repulsive force.

The height of the energy barrier, which is crucial to emulsion stabilization, depends on the state of ionization of the emulsifying agent. Most surfactants are used at pH values where they are totally ionized so that the surface potential is high, giving a correspondingly high energy barrier. The surface potential cannot be measured directly, but can be estimated from the experimentally derived zeta potential. In lipid emulsions for parenteral nutrition, the electrostatic barrier is provided by the ionization of the negatively charged phospholipids in the emulsifier film at the oil droplet–water interface. At physiological pH, a typical fat emulsion carries a negative charge with the zeta potential between 30 and 60 mV. This is sufficient to ensure stability because of the high potential energy barrier. The addition of electrolytes or a change in pH can have a devastating effect on emulsion stability by compressing the double layers, thus reducing the zeta potential and energy barrier and allowing droplets to move into the primary minimum. Thus, great care must be exercised when electrolytes are added nutritional emulsions. With emulsifiers such as proteins and gums, ionization, and hence emulsifying activity, is also pH dependent (c.f. Table 3).

The DLVO theory does not explain either the stability of water-in-oil emulsions or the stability of oil-in-water emulsions stabilized by adsorbed nonionic surfactants and polymers where the electrical contributions are often of secondary importance. In these, steric and hydrational forces, which arise from the loss of entropy when adsorbed polymer layers or hydrated chains of nonionic polyether surfactant intermingle on close approach of two similar droplets, are more important (Fig. 4b). In emulsions stabilized by polyether surfactants, these interactions assume importance at very close distances of approach and are influenced markedly by temperature and degree of hydration of the polyoxyethylene chains. With block copolymers of the ethylene oxide–propylene oxide type, such as the poloxamers, the hydrated polyoxyethylene chains extend into the continuous phase to provide steric stabilization and the hydrophobic propylene oxide portion is anchored onto the droplet surface to form a strong protecting layer against coalescence. Stability is optimized when the droplet surfaces are completely coated by polymer chains so that desorption and lateral movement of the polymer is inhibited. With w/o emulsions, steric hindrance of the adsorbed chains of emulsifier can also result in entropic repulsion effects at small distances of separation.

Some natural polymeric emulsifiers such as the gums, in addition to forming steric and electrostatic barriers form thick multilayered films that are very resistant to film rupture. They may also thicken the continuous phases

of o/w emulsions, thereby reducing the rate of film drainage in the initial stages of coalescence. Small solid particles may stabilize emulsions if they are wetted by both phases and possess sufficient adhesion for one another to form a coherent interfacial film. The film serves as a mechanical barrier to prevent the coalescence of droplets, and if charged, electrostatic mechanisms further assist in the stabilization of the emulsion. Although solids are not generally sufficient to stabilize emulsions on their own, they often reinforce the effectiveness of other emulsifiers.

Stabilization by Mixtures of Emulsifiers

Most pharmaceutical emulsions, whether dilute mobile systems for internal use or thick semisolid creams for application to the skin, contain mixtures of emulsifiers, as these provide more stable preparations. For example, traditional oral preparations are sometimes stabilized by mixtures of gums such as acacia and tragacanth and mixtures of nonionic surfactants of high and low hydrophile–lipophile balance (HLB) generally form more stable emulsions than a single surfactant. The lecithins used to stabilize parenteral emulsions are usually mixtures of neutral and charged lipids as are the partially neutralized glyceryl esters such as self-emulsifying glyceryl monostearate. Combinations of sparingly soluble long-chain acids, alcohols, or glyceryl esters with more soluble ionic and non-ionic surfactants are widely used in dermatological o/w lotions and creams, where they are sometimes added in the form of a preblended emulsifying wax (Table 4). Surfactant/fatty acid combinations are also present in traditional liniment and lotion emulsion formulations prepared by the nascent soap method and in preparations where triethanolamine soaps are formed *in situ* from the interaction of triethanolamine and excess fatty acid.

Equations from the DLVO theory even if modified to allow for the steric repulsive forces cannot cope with mixtures of emulsifiers. Increased stability in model emulsions (c.f. Fig. 1a) is attributed not so much to the control of flocculation (although this does occur), but rather to the prevention or retardation of coalescence by closer packing of the molecules in the adsorbed monolayer to form a more rigid and condensed film. There is now substantial evidence that interactions between emulsifier components to form specific lamellar phases, either liquid crystalline or gel, that are capable of incorporating large volumes of water are important for the stability of many parenteral and dermatological emulsions. Mobile parenteral injections stabilized by

Table 4 Typical emulsifying waxes and their component surfactants

Emulsifying wax	Components
Emulsifying wax USNF	Cetearyl alcohol, polysorbate
Cationic emulsifying wax BP	Cetearyl alcohol, cetrimonium bromide
Cetomacrogol emulsifying wax BP	Cetearyl alcohol, ceteth 20
Glyceryl stearate, SE	Glyceryl stearate, soap

phospholipid mixtures usually contain swollen lamellar liquid crystals (34) whereas a swollen gel phase which generally provides better stability as well as a means of controlling rheological properties dominates in semisolid dermatological emulsions prepared with emulsifying waxes. The relevance of bilayer gel and liquid crystalline phases in dermatological and parenteral emulsions have been discussed in reviews (10, 29). Much of the information about their structures was obtained from investigations of the phase behaviour of emulsifiers and their components in water over the ranges of concentration and temperature relevant to the manufacture, storage, and use of the formulations. It is interesting to note that the same electrostatic, hydration, and steric forces that operate in simple emulsions also dominate the stability and properties of the lamellar phases (35).

Ostwald Ripening

Ostwald ripening has not been studied as extensively in emulsions as has coalescence, although it is a major mechanism for instability in lipid and perfluorochemical emulsions with submicron droplet sizes where a condensed monolayer is not always necessary for emulsion stability (36). Although surfactant interfacial films protect against flocculation and coalescence, Ostwald ripening may in fact be enhanced if the surfactant is above the critical micelle concentration (cmc) because of the diffusion of solubilized oil through the continuous phase. The addition of a third component to the emulsion that has a lower vapor pressure and solubility than the disperse phase will also inhibit Ostwald ripening. The addition of long-chain alkanes to comparatively unstable oil-in-water emulsions prepared with sodium dodecyl sulfate resulted in marked increases in stability even though the alkanes do not effect the composition or mechanical properties of the oil–water interface (37). The stability of pure perfluorodecalin emulsions used as blood substitutes is enhanced by the addition of a small quantity of perfluorotributylamine, and lipid emulsions containing local anaesthetic/analgesic drugs show enhanced stability

in the presence of hydrophobic excipients of lower solubility than the disperse phase (38). Polymeric emulsifiers possibly stabilize emulsions against Ostwald ripening by increasing the viscosity of the continuous phase. The relative lack of Ostwald ripening in emulsions prepared from oils immiscible with water, such as mineral oil, may partly explain why they are easier to emulsify than are more miscible vegetable oils used in parenteral preparations.

Selection of Emulsifier

Over the years there have been many attempts to find systemic methods for screening potential emulsifiers from the enormous number of surfactants available commercially. Although the mechanisms governing the stability of emulsions, including the complex multiple phase systems of pharmacy are becoming clearer, there are still few scientific guidelines to assist in the proper selection of emulsifiers for a particular emulsion. Semiempirical methods based on both interfacial considerations and the phase behavior of the emulsifiers are considered briefly next.

The hydrophile–lipophile balance (HLB) concept

Griffin devised the concept of hydrophile–lipophile balance (HLB) and its additivity many years ago for selection of nonionic emulsifiers and this rather empirical method is still widely used. The enormous literature on the HLB of surfactants has been reviewed by Becher (39). Each surfactant is allocated an HLB number usually on a scale of 0–20, based on the relative proportions of the hydrophilic and hydrophobic part of a molecule. Water-in-oil emulsions are formed generally from oil-soluble surfactants of low HLB number and oil-in-water emulsions from more hydrophilic surfactants of high HLB number. The method of selection is based on the observation that each type of oil will require an emulsifying agent of a specific HLB number to produce a stable emulsion. Thus, oils are often designated two “required” HLB numbers, one low and one high, for their emulsification to form water-in-oil and oil-in-water

emulsions respectively. A series of emulsifiers and their blends with HLB values close to the required HLB of the oil are then examined to see which one forms the most stable emulsion (c.f. Fig. 1a).

Although the HLB concept narrows the range of emulsifiers to select and provides a schematic approach for the formulator, it is limited by its strict relation to molecular structure of the individual surfactants. The concept does not consider the total emulsion and is therefore insensitive to interactions between emulsifier components, the influence of temperature changes, or the presence of additional ingredients in the emulsion. Consequently, not all emulsifier blends of the correct HLB form stable systems. For example, when surfactants of widely different HLB numbers are blended to give the optimum theoretical HLB, the high solubility of the surfactant in the oil and aqueous phases change the balance of the molecules at the interface and unstable emulsions may result. Similarly, if the added surfactants form intermolecular associations at the interface, the association complex is unlikely to have properties that are related in any simple way to the individual properties of the constituent molecules.

The phase inversion temperature (PIT) method (HLB-temperature)

A complementary means of emulsifier selection, the phase inversion temperature (PIT), which employs a characteristic property of the emulsion rather than the properties of the emulsifiers in isolation, was introduced by Shinoda (40). The method uses the fact that the stabilities of oil-in-water emulsions containing nonionic surfactants are closely related to the degree of hydration of the interfacial films. Emulsion stability is reduced by increase in temperature or added salts because these decrease the extent of interfacial film hydration. Phase inversion, due to a change from preferential water solubility of the emulsifier film at low temperature to preferential oil solubility at high temperature, will occur at a specific temperature unique to the particular emulsion and this can be determined experimentally. As a general rule, relatively stable oil-in-water emulsions are obtained when their temperatures during storage and use are between 20 and 65°C below the PIT, presumably because the films are sufficiently hydrated. Mixtures of emulsifiers with identical HLBs produce emulsions with quite different PITs because additives and interactions between the components affect PIT but not HLB.

Microscopic selection for multiple phase emulsions

The better understanding of the mechanisms of stability in complex dermatological emulsions stabilized by

surfactants and amphiphiles has enabled the development of a rapid microscopic method for evaluation of potential emulsifiers. The method is based on the observation that good emulsifier blends that stabilize emulsions by the formation of multilayers of stable gel phase also swell spontaneously in water at ambient temperature and this process can be observed microscopically. Mixtures that do not form gel phase or form metastable gels only after a heating and cooling cycle cannot be observed to swell spontaneously at ambient temperature (4).

Emulsification Techniques

Emulsions are usually prepared by the application of mechanical energy produced by a wide range of agitation techniques. These disrupt droplets by the application of either shear forces in laminar flow or inertial forces in turbulent flow. Emulsifying devices ranging from simple hand mixers and stirrers to the use of propeller or turbine mixers, static mixers, colloid mills, homogenizers, and ultrasonic devices have been used.

Emulsifiers also have an important role in the process of emulsification. Surfactant emulsifiers reduce interfacial tensions during emulsification, making droplets easier to break up as well as reducing the tendency for recombination. Other emulsifiers such as the polymer macromolecules alter the hydrodynamic forces during the agitation process by their influence on rheological properties. Scale-up procedures from the laboratory to manufacture can introduce a number of problems due to the difficulties in matching the exact conditions of mixing, and, because of entrapment-of air, especially in emulsions of high consistency that have a yield value. Along with being inelegant, even traces of atmospheric air can cause decomposition in drugs or excipients susceptible to oxidation.

There are additional constraints when manufacturing parenteral emulsions that must be sterile and of fine particle size. Perfluorochemical and fat emulsions are usually prepared by homogenization at high temperature and pressure, as a large output of energy is required to produce droplet sizes considerably less than 1 μm . Although heat sterilization is widely used, this places a severe test on the stability, and emulsions are sometimes prepared from sterile components under strict aseptic conditions and further sterilized by filtration (15).

Processing variables

Differences in manufacturing techniques such as the rate of the heating and cooling cycle, the extent and order of mixing can cause variations in the consistency and

rheology of the resulting emulsions. The initial particle size of the emulsion depends on the emulsifiers used, the emulsification equipment, the addition speed, and the phase volume. If the surfactant is placed in one of the phases prior to emulsification, it will migrate to the other to establish equilibrium. Thus, emulsification temperatures and cooling rates are important and the time of the mixing should be sufficient to allow the surfactant to migrate to and equilibrate at the interface throughout the process. Oil-in-water emulsions are sometimes prepared by the phase inversion technique, where the aqueous phase is added to the oil phase to form a w/o emulsion that inverts to an o/w emulsion on addition of further amounts of water. This process is claimed to give finer emulsions.

Preparation techniques, in particular cooling rates and mixing procedures, have a marked effect on initial and final consistencies of emulsions prepared with nonionic emulsifying waxes. For example, "shock" cooling and limited mixing initially produces very mobile systems whereas slow cooling with adequate mixing produces semisolid emulsions. Mixing time, when the emulsifiers are in the molten state, influences the distribution of surfactant within the molten masses and bilayers and the relative lamellar order within the system. With ionic emulsifying waxes, different preparation techniques cause comparatively minor variations in the consistency of the final product. It was shown that differences are not due to the gel phase component of cationic ternary systems, but rather due to the variations in size of the crystalline alcohol that precipitates after manufacture. Systems formed by a rapid "shock" cooling method exhibited smaller but greater numbers of cetostearyl alcohol crystals and were thicker than similar ternary systems manufactured by a more lengthy procedure (29).

MICROEMULSIONS

Microemulsions are thermodynamically stable, transparent (or translucent) dispersions of oil and water that are stabilized by an interfacial film of surfactant molecules. The surfactant may be pure, a mixture, or combined with a cosurfactant such as a medium-chain alcohol (e.g., butanol, pentanol). These homogeneous systems, which can be prepared over a wide range of surfactant concentrations and oil to water ratios (20–80%), are all fluids of low viscosity.

The term microemulsion, which implies a close relationship to ordinary emulsions, is misleading because the microemulsion state embraces a number of different microstructures, most of which have little in common

with ordinary emulsions. Although microemulsions may be composed of dispersed droplets of either oil or water, it is now accepted that they are essentially stable, single-phase swollen micellar solutions rather than unstable two-phase dispersions. Microemulsions are readily distinguished from normal emulsions by their transparency, their low viscosity, and more fundamentally their thermodynamic stability and ability to form spontaneously. The dividing line, however, between the size of a swollen micelle (~10–140 nm) and a fine emulsion droplet (~100–600 nm) is not well defined, although microemulsions are very labile systems and a microemulsion droplet may disappear within a fraction of a second whilst another droplet forms spontaneously elsewhere in the system. In contrast, ordinary emulsion droplets, however small, exist as individual entities until coalescence or Ostwald ripening occurs.

Figure 5 shows a hypothetical phase diagram with representation of microemulsion structures. At high water concentrations, microemulsions consist of small oil droplets dispersed in water (o/w microemulsion), while at lower water concentrations the situation is reversed and the system consists of water droplets dispersed in oil (w/o microemulsions). In each phase, the oil and water droplets are separated by a surfactant-rich film. In systems containing comparable amounts of oil and water, equilibrium bicontinuous structures in which the oil and the water domains interpenetrate in a more complicated manner are formed. In this region, infinite curved channels of both the oil and the water domains extend over macroscopic distances and the surfactant forms an interface of rapidly fluctuating curvature, but in which the net curvature is near zero.

Pharmaceutical and Biological Applications of Microemulsions

Microemulsions provide ultralow interfacial tensions and large interfacial areas as well as the ability to concentrate and localize significant amounts of both oil- and water-soluble materials within the same isotropic medium. Over the years, attention has been focused on their potential use as novel reaction media for a wide range of chemical, biochemical, and photochemical reactions, and as carriers for chemicals and small particles, reviewed by Eccleston (41). Inverse microemulsions of the w/o type are the subject of particular interest because of the rapidly emerging range of biotechnological applications based on their ability to solubilize enzymes in the water domains without denaturation or loss of activity. The ability of such solubilized hydrophilic enzymes to transform hydrophobic substrates dissolved in the organic

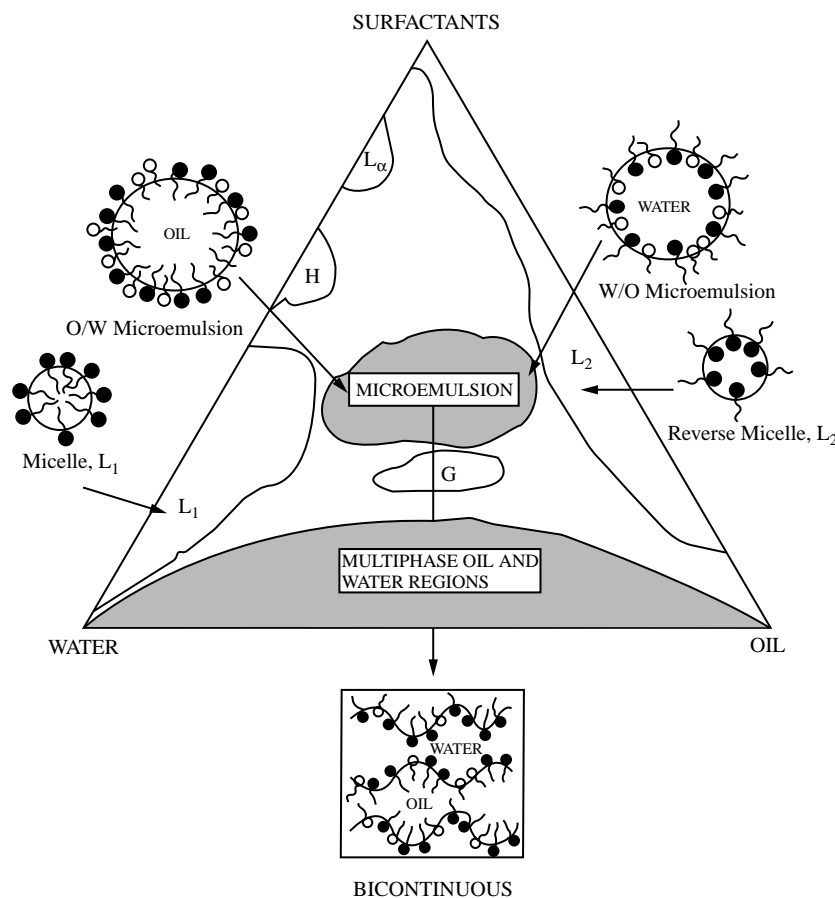


Fig. 5 Ternary phase diagram for oil, water, and surfactant mixtures showing micellar, microemulsion, and multiphase macroemulsion regions with schematic representations of various structures.

phase could lead eventually to the synthesis of new drugs. As with ordinary emulsions, microemulsions show improved gastrointestinal absorption. They also have a number of other advantages over macroemulsions for drug delivery. Microemulsions form spontaneously without the aid of high shear equipment or significant heat input (heat and gentle mixing are required only if it is necessary to melt any of the ingredients) and their microstructures are independent of the order of addition of the excipients. Optical transparency and low viscosity of microemulsions ensure that they are cosmetically elegant and easy to handle and pack, and their indefinite stabilities ensure a long shelf life. Microemulsions have thus attracted much interest in their drug delivery potential. Both o/w and w/o emulsions have been shown to enhance the oral bioavailability of drugs, including various peptides (42). A peroral concentrate of cyclosporine is now available commercially (Sandimmune Neoral[®] Novartis), which forms a microemulsion

in the aqueous fluids of the gastrointestinal tract. In this preparation, the rate of absorption of cyclosporin is more rapid and less variable than it is with the conventional oily dispersion. Calcein administered intraduodenally in the aqueous phase of a w/o microemulsion prepared from medium-chain triglycerides (43) produces significantly higher plasma levels of the drug compared with an aqueous solution.

Microemulsions have also been used for topical delivery where they increase drug absorption. For example, cetyl alcohol, which is commonly used as an emulsifier in lotions and creams, is absorbed faster and deeper into the skin when formulated as a component of a microemulsion (44). Although efficient skin penetration may be desirable for a therapeutic agent, the relatively high concentrations of surfactant (10–25%) and cosurfactant or cosolvent (5–10%) in such formulations could enhance skin absorption of potential irritants or carcinogens. In fact, the main limitations in realizing

the full potential of microemulsions as drug delivery systems are the narrow range of surfactants, cosurfactants, solvents, and other materials acceptable pharmaceutically.

Microemulsion Formation

Many approaches have been used to explore the mechanisms of microemulsion formation and stability [summarized by Eccleston (41) and Attwood (45)]. Early theories considered interfacial aspects of microemulsions and did not distinguish between thermodynamically stable systems and very fine kinetically stable emulsions. For microemulsions to form spontaneously, the free energy involved when the interfacial area is increased, ΔG ($\Delta G = \gamma \Delta A$, where ΔA is the increase in interfacial area) must be negative. An essential requirement is that the interfacial tension between the oil and water phases γ , is reduced to a very low value by the interfacial film, giving a small but positive free-energy value. The dispersion of the droplets in the continuous phase increases the entropy of the system. Microemulsions form because the negative free energy changes due to the entropy of the dispersion of droplets in the continuous phase overcomes the positive product of the small interfacial tension and the large interfacial area A .

The curvature of the oil–water interface in microemulsions varies from highly curved towards oil (o/w) or water (w/o) to zero mean curvature in bicontinuous structures. The type of microemulsion that forms depends on the properties of the surfactant, cosurfactant and the oil. Although there are no strict rules for choosing the appropriate microemulsion components, there are a number of general guidelines based on empirical observations. The surfactant(s) chosen for a particular oil must:

1. lower interfacial tension to a very low value to aid dispersion processes during the preparation of the microemulsion.
2. be of the appropriate hydrophile-lipophile character to provide the correct curvature at the interfacial region for the desired microemulsion type, o/w, w/o or bicontinuous.
3. provide a flexible film that can readily deform round small droplets.

The analysis of film curvature for surfactant associations leading to microemulsion formation has been rationalized by Mitchell and Ninham. They used a packing ratio P defined as V/al , where V is the partial molar volume of the surfactant, a the cross sectional area (i.e. size) of the surfactant head group, and l the maximum length of the surfactant chain (46). The packing ratio provides a direct measure of HLB and is influenced by the same factors. Oil-in-water microemulsions are favored if the effective polar part of the surfactant is more bulky than the hydrophobic part, that is, P varies from 0 to 1, and the interface curves spontaneously towards water (positive curvature). Water-in-oil microemulsions form when the interface curves in the opposite direction, that is, P is greater than 1 (negative curvature). At zero curvature, when the HLB is balanced and P is zero, either bicontinuous or lamellar structures may form according to the rigidity of the film. The critical packing parameter P is based purely on geometric considerations. Hydration of the surfactant head group and penetration of the oil and the cosurfactant into the surfactant film also affect the packing and curvature, as summarized in Table 5, which also illustrates how formulation variables may be manipulated to produce a microemulsion of the desired type (47).

Most single-chain surfactants do not lower the oil–water interfacial tension sufficiently to form microemulsions nor are they of the correct molecular structure, and

Table 5 Factors affecting spontaneous curvature of monolayers

Variable	Curvature effect	Cause
Increase oil chain length	More positive	Less penetration of surfactant tail region
Addition of shorter chain cosurfactant	More positive	Alcohol swells head region more than tail region
Addition of longer chain cosurfactant	More negative	Alcohol swells surfactant chain region more than head region
Addition of salt (ionic surfactant)	More negative	Screened repulsion between polar head groups
Addition of salt (nonionic surfactant)	More negative	Headgroup size reduced by dehydration
Branched or double chained surfactant	More negative	Increased tail group area
Reduced surfactant head group size	More negative	
Increased temperature (nonionic surfactant)	More negative	Headgroup size reduced by dehydration
Increased temperature (ionic surfactant)	More positive	Increased surfactant counter-ion dissociation

(Adapted from Ref. 47.)

short- to medium-chain length alcohols are necessary as cosurfactants. The cosurfactant also ensures that the interfacial film is flexible enough to deform readily around each droplet as their intercalation between the primary surfactant molecules decreases both the polar head group interactions and the hydrocarbon chain interactions. Medium-chain alcohols such as pentanol and hexanol have been used by many investigators as they are particularly effective cosurfactants. They are not, however, suitable for pharmaceuticals due to their high irritant potential. Double-chain surfactants such as anionic Aerosol-OT (bis-2-ethylhexyl sulfosuccinate) or cationic DDAB (didodecyltrimethylammonium bromide), which have relatively small head groups and bulky hydrophobic portions, are already of the required HLB to form w/o microemulsions spontaneously without a cosurfactant. Unfortunately, these widely investigated surfactants are too toxic for general pharmaceutical or biotechnological applications. Double-chain phospholipids such as the phosphatidylcholines of lecithin are an obvious possibility. Although lecithin is too lipophilic to form microemulsions, pharmaceutically acceptable microemulsions have been prepared from double-chain phospholipids by using acceptable cosurfactants such as ethanol, propanol, or *n*-butanol with isopropyl myristate (48, 51). Self-emulsifying drug delivery systems are composed of triglyceride oils and surfactant mixtures that undergo spontaneous emulsification when mixed with water (52). This principle is used in the commercial product Sandimmune Neoral[®], which forms a microemulsion in situ when diluted by gastric fluid.

Formulation and Preparation of Microemulsions

As microemulsions are thermodynamically stable, they can be prepared simply by blending oil, water, surfactant, and cosurfactant with mild agitation. Once the appropriate microemulsion components have been selected, quaternary phase diagrams or ternary pseudo-phase diagrams may be constructed to define the extent and nature of the microemulsion regions and the surrounding two- and three-phase domains. The microemulsion region can be identified and characterized using the range of light, neutron, and X-ray scattering and other techniques such as NMR and microscopy (45). Problems arise in interpretation of data in systems of high droplet volume fraction due to interdroplet interactions. The normal practice of investigating systems at relatively low concentrations and then extrapolating to zero concentration in order to eliminate interparticle interactions cannot be applied to

microemulsions as it is not possible to dilute the systems without affecting their structure. Hard sphere models, such as those adapted from Percus and Yevick, have been successfully used to analyze scattering data from concentrated w/o microemulsions (53).

REFERENCES

1. International Union of Pure and Applied Chemistry, IUPAC. *Manual of Colloid and Surface Chemistry*; Butterworths: London, 1971.
2. Garti, N.; Aserin, A. Double Emulsions Stabilized by Macromolecular Surfactants. *Micelles Microemulsions and Monolayers*; Shah, D.O. Ed.; Marcel Dekker, Inc.: New York, 1998; 333–362.
3. Davis, S.S.; Washington, C.; West, P.; Illum, L.; Liversidge, G.; Sternson, L.; Kirsh, R. Lipid Emulsions as Drug Delivery Systems. *Ann. N. Y. Acad. Sci.* **1987**, *507*, 75–88.
4. Eccleston, G.M.; Emulsions. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C. Eds.; Marcel Dekker, Inc.: New York, 1992; 5, 137–188.
5. Block, L.H. Pharmaceutical Emulsions and Microemulsions. *Pharmaceutical Dosage Forms: Disperse Systems*; Lieberman, H.A., Rieger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1996; 2, 47–109.
6. Rosoff, M.; Specialised Pharmaceutical Emulsions. *Pharmaceutical Dosage Forms: Disperse Systems*; Lieberman, H.A., Rieger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1997; 3, 1–22.
7. Barry, B.W. *Dermatological Formulation. Percutaneous Absorption*; Marcel Dekker, Inc.: New York, 1983; 480.
8. Flynn, G.L.; Cutaneous and Transdermal Delivery: Processes and Systems of Delivery. *Modern Pharmaceutics*, 3rd Ed.; Banker, G.S., Rhodes, C.T. Eds.; Marcel Dekker, Inc.: New York, 1996; 239–298.
9. Eccleston, G.M.; The Microstructure of Semisolid Creams. *Pharm. Int.* **1986**, *7* (3), 63–70.
10. Eccleston, G.M.; Functions of Mixed Emulsifiers and Emulsifying Waxes in Dermatological Lotions and Creams. *Colloids and Surfaces* **1997**, 169–182.
11. Amselem, S.; Friedman, D.; Submicron Emulsions as Drug Carriers for Topical Administration. *Submicron Emulsions in Drug Delivery*; Benita, S. Ed.; Harwood Academic: Amsterdam, 1998; 9, 152–173.
12. Carrigan, P.; Bates, T.; Biopharmaceutics of Drug Administered in Lipid-Containing Dosage Forms 1: GI Absorption of Griseofulvin from an Oil-in-Water Emulsion in Rat. *J. Pharm. Sci.* **1973**, *62*, 1476–1414.
13. Collins Gold, L.C.; Lyons, R.T.; Bartholow, L.C.; Parenteral Emulsions for Drug Delivery. *Adv. Drug Del. Rev.* **1990**, *5*, 189–208.
14. Prankerd, R.J.; Stella, V.J.; The Use of Oil-In-Water Emulsions as a Vehicle for Parenteral Administration. *J. Parent. Sci. Technol.* **1990**, *44*, 139–149.
15. Floyd, A.G.; Jain, S. Injectable Emulsions and Suspensions. *Pharmaceutical Dosage Forms: Disperse Systems*; Lieberman, H.A., Rieger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1996; 2, 261–318.

16. Nomura, T.; Koreeda, N.; Yamashita, F.; Takakura, Y.; Hashida, M. Effect of Particle Size on the Disposition of Lipid Carriers after Intratumoral Injection into Tissue Isolated Tumors. *Pharm. Res.* **1998**, *15* (1), 128–132.
17. Omotosho, J.A.; Whateley, T.L.; Florence, A.T. Release of 5-Fluorouracil from Intramuscular W/O/W Multiple Emulsions. *Biopharm. and Drug Disposit* **1989**, *10*, 257–268.
18. Von Dardel, O.; Mebius, C.; Mossberg, T.; Svensson, B. Fat Emulsion as a Vehicle for Diazepam. A Study of 9492 Patients. *Br. J. Anaesth.* **1983**, *55*, 41–47.
19. Lovell, M.W.; Johnson, H.W.; Hui, H.W.; Cannon, J.B.; Gupta, P.K.; Hsu, C.C. Less Painful Emulsion Formulations for Intravenous Administration of Clarithromycin. *Int. J. Pharm.* **1994**, *109*, 45–57.
20. Kirsh, R.; Goldstein, R.; Tarloff, J.; Parris, D.; Hook, J.; Hanna, N.; Bugelski, P.; Poste, G.J. An Emulsion Formulation of Amphotericin B Improves the Therapeutic Index when Treating Systemic Murine Candidiasis. *Infect. Dis.* **1988**, *158*, 1065–1070.
21. Submicron Emulsions in Drug Targeting and Delivery. *Drug Targeting and Delivery*; Benita, S., Ed.; Harwood Academic: Amsterdam, 1998; 333.
22. Spahn, D.R. Current Status Of Artificial Oxygen Carriers. *Adv. Drug Del. Rev.* **2000**, *40*, 143–151.
23. Naveh, N.; Muchtar, S.; Benita, S. Pilocarpine Incorporated into a Submicron Emulsion Vehicle Causes Unexpectedly Prolonged Ocular Hypotensive Effects in Rabbits. *J. Ocular Pharmacol.* **1994**, *10* (3), 509–520.
24. Kararli, T.T.; Needham, T.E.; Schoenhard, G.; Baron, D.A.; Schmidt, R.E.; Katz, B.; Belonio, B. Enhancement of Nasal Delivery of a Renin Inhibitor in the Rat Using Emulsion Formulations. *Pharm. Res.* **1992**, *9*, 1024–1028.
25. Aikawa, K.; Matsumoto, K.; Uda, H.; Tanaka, S.; Shimamura, H.; Aramaki, Y.; Ysuchiya, S. Prolonged Release of Drug from O/W Emulsion and Residence in Rat Nasal Cavity. *Pharm. Dev. Technol.* **1998**, *3* (4), 461–469.
26. Mitra, R.; Pezron, I.; Chu, W.A.; Mitra, A.K. Lipid Emulsions as Vehicles for Enhanced Nasal Delivery of Insulin. *Int. J. Pharm.* **2000**, *205* (1–2), 127–134.
27. Mascioli, E.A.; Babayan, V.K.; Bistran, B.R.; Blackburn, G.L. Novel Triglycerides for Special Medical Purposes. *J. Parent. Ent. Nut.* **1989**, *12* (6), 127S–132S.
28. Hedeman, H.; Brondsted, H.; Mullertz, A.; Frokjaer, S. Fat Emulsions Based on Structured Lipids (1,3 Specific Triglycerides): An Investigation of the In Vivo Fate. *Pharm. Res.* **1996**, *13* (5), 725–733.
29. Eccleston, G.M. Multiple-Phase Oil-in-Water Emulsions. *J. Soc. Cosmet. Chem.* **1990**, *41*, 1–22.
30. Nishikawa, M.; Yoshinobu, T.; Hashida, M. Biofate of Fat Emulsions. *Submicron Emulsions in Drug Targeting and Delivery*; Benita, S., Ed.; Harwood Academic: Amsterdam, 1998; 99–118.
31. Eccleston, G.M.; Florence, A.T. Application of Emulsion Theory to Complex and Real Systems. *Int. J. Cosmet. Sci.* **1985**, *7*, 195–212.
32. Hough, D.B.; Thompson, L. Effect of Nonionic Surfactants on the Stability of Dispersions. *Nonionic Surfactants. Physical Chemistry*; Schick, M.J., Ed.; Marcel Dekker, Inc.: New York, 1987; 601–676.
33. Washington, C. The Electrokinetic Properties of Phospholipid Stabilized Fat Emulsions. III. Interdroplet Potentials and Stability Ratios in Monovalent Electrolytes. *Int. J. Pharm.* **1990**, *64*, 67–73.
34. Rydhag, L.; Wilton, I. The Function of Phospholipids of Soybean Lecithin in Emulsions. *Am. J. Oil Col. Chem.* **1981**, *58*, 830–837.
35. Eccleston, G.M.; Behan-Martin, M.K.; Jones, G.R.; Townes-Andrews, E. Synchrotron X-Ray Investigations into the Lamellar Gel Phase Formed in Pharmaceutical Creams Prepared with Cetriride and Fatty Alcohols. *Int. J. Pharm.* **2000**, *203*, 127–139.
36. Buscall, R.; Davis, S.S.; Potts, D.C. The Effect of Long-Chain Alkenes on the Stability of Oil-in-Water Emulsions. The Significance of Ostwald Ripening. *Colloid Polym. Sci.* **1979**, *257*, 636–644.
37. Davis, S.S.; Round, H.P.; Purewal, T.S. Ostwald Ripening and the Stability of Emulsion Systems: An Explanation for the Effect of an Added Third Component. *J. Colloid Int. Sci.* **1981**, *80* (2), 508–511.
38. Welin-Berger, K.; Bergenstahl, B. Inhibition of Ostwald Ripening in Local Anaesthetic Emulsions by Using Hydrophobic Excipients in the Disperse Phase. *Int. J. Pharm.* **2000**, *200* (2), 249–260.
39. Becher, P. Hydrophile-Lipophile Balance: An Updated Biography. *Encyclopedia of Emulsion Technology*; Becher, P., Ed.; Marcel Dekker, Inc.: New York, 1985; 2, 425–512.
40. Shinoda, K.; Saito, H. The Stability of O/W Type Emulsions as Functions of Temperature and the HLB of Emulsifiers: The Emulsification by PIT Method. *J. Colloid Int. Sci.* **1969**, *30*, 258–263.
41. Eccleston, G.M. Microemulsions. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1995; 9, 375–421.
42. Ritschel, W. Microemulsions for Improved Peptide Absorption from the Gastrointestinal Tract. *Meth. Find. Exp. Clin. Pharmacol.* **1991**, *13*, 205–220.
43. Constantinides, P.P.; Scalart, J.; Lancaster, C.; Marcello, J.; Marks, G.; Ellens, H.; Smith, P.L. Formulation and Intestinal Absorption Enhancement Evaluation of Water-In-Oil Microemulsions Incorporating Medium Chain Triglycerides. *Pharm. Res.* **1994**, *11* (10), 1385–1390.
44. Linn, E.E.; Pohland, R.C.; Byrd, T.K. Microemulsions for Intradermal Delivery of Cetyl Alcohol and Octyl Dimethyl PABA. *Drug Dev. Ind. Pharm.* **1990**, *16* (9), 899–920.
45. Attwood, D. Microemulsions. *Colloidal Drug Delivery Systems*; Kreuter, J., Ed.; Marcel Dekker, Inc.: New York, 1994; 66, 31–71.
46. Mitchell, D.J.; Ninham, B. Micelles Vesicles and Microemulsions. *J. Chem Soc. Faraday Trans.* **1981**, *2* (677), 601–629.
47. Fletcher, D.I.; Parrott, D. Protein Partitioning between Microemulsion Phases and Conjugate Aqueous Phases. *Structure and Reactivity in Reverse Micelles*; Pileni, M.P., Ed.; Elsevier: New York, 1988; 303–322.
48. Attwood, D.; Mallon, C.; Taylor, C.J. Phase Studies on Oil-in-Water Phospholipid Microemulsions. *Int. J. Pharm.* **1992**, *84*, R5–R8.
49. Aboofazeli, R.; Lawrence, M.J. Investigations into the Formation and Characterisation of Phospholipid Microemulsions. 1. Pseudo-ternary Phase Diagrams of Systems Containing Water-Lecithin-Alcohol-Isopropyl Myristate. *Int. J. Pharm.* **1993**, *93*, 161–175.

50. Aboofazeli, R.; Lawrence, M.J. Investigations into the Formation and Characterisation of Phospholipid Microemulsions. II. Pseudo-Ternary Phase Diagrams of Systems Containing Water-Lecithin-Isopropyl Myristate-Alcohol: Influence of Purity. *Int. J. Pharm.* **1994**, *106*, 51–61.
51. Aboofazeli, R.; Lawrence, M.J. Investigations into the Formation and Characterisation of Phospholipid Microemulsions. II. Pseudo-Ternary Phase Diagrams of Systems Containing Water-Lecithin-Isopropyl Myristate. *Int. J. Pharm.* **1994**, *111*, 63–72.
52. Pouton, C. Lipid Formulations for Oral Administration of Drugs: Non-Emulsifying, Self-Emulsifying and Self-Microemulsifying Drug Delivery Systems. *Eur. J. Pharm. Sci.* **2000**, *1* (11), S93–S98.
53. Percus, J.K.; Yevick, G.J. Analysis of Classical Statistical Mechanics by Means of Collective Co-ordinates. *Phys. Rev.* **1958**, *110*, 1–13.

Electrochemical Detection for Pharmaceutical Analysis

Damon M. Osbourn

Craig E. Lunte

The University of Kansas, Lawrence, Kansas, U.S.A.

INTRODUCTION

In order to bring a drug product from the discovery stage to the commercial market, many analytical methods must be employed. The analytical chemist develops methodology for quality control, stability testing, pharmacokinetics, identification, and clinical studies. Analysts choose from a number of techniques such as NMR, MS, separations, spectroscopy, and electrochemistry, each with its own advantages. Electrochemistry has many advantages, which make it an appealing choice for pharmaceutical analysis. Most electrochemical techniques have excellent limits of detection and a wide dynamic range. Electroanalytical techniques require only very small sample volumes, often in the microliter range, coupled with the low detection limits allowing analysis on subpicogram amounts of analyte. The selectivity of electrochemical detection in complex samples is excellent because fewer electroactive interferents are often encountered than spectroscopic interferents. For these reasons, electrochemistry is uniquely suited for clinical and bioanalysis, where small volumes of blood or urine are analyzed for low concentrations of drug products and metabolites. Additionally, many methods for *in vivo* analysis have been developed. Finally, in comparison to other analytical techniques, the instrumentation required for electrochemical methods is simple and inexpensive, even to the point of having disposable electrochemical cells. With the advent of micromachining technologies, the designs of electrochemical cells promise to become even more reproducible, simple, and inexpensive.

Electrochemical detection is based on the electrical signal arising between two electrodes immersed in a sample solution. Electroanalytical techniques fall into two main categories, potentiometric and Faradaic techniques. Potentiometry is the measurement of a potential difference between two electrodes under equilibrium conditions (i.e., no current flow). The potential is then related to concentration of the analyte species. Faradaic processes are based on the oxidation or reduction of the analyte, where a specific potential waveform is applied and the current is used to extract information about the sample. Many different techniques have been developed to gain

quantitative and qualitative information about analytes of interest. This article covers the basics of many different techniques useful in pharmaceutical analysis and directs the reader to current pharmaceutical applications in the literature.

Potentiometric Techniques

Although electrochemistry has the stigma of being difficult to use, and therefore is often overlooked as an analysis option, potentiometric measurements are probably the most common technique encountered. Many analytical chemists make potentiometric measurements daily, whenever they use a pH meter. Potentiometry is based on the measurement of the potential between two electrodes immersed in a test solution. As the electrical potential of the cell is measured with no current flow between the electrodes, potentiometry is an equilibrium technique. The first electrode, the indicator electrode, is chosen to respond to the activity of a specific species in the test solution. The second electrode is a reference of known and fixed potential. The design of the indicator electrode is fundamental to potentiometric measurements, and should interact selectively with the analyte of interest so that other sample constituents do not interfere with the measurement. Many different strategies have been developed to make indicator electrodes that respond selectively to a number of species including organic ions.

An example of an electrochemical cell for potentiometry is shown schematically in Fig. 1. The cell consists of an external reference electrode and an indicator electrode immersed in a test solution of analyte with some activity, $(a_i)_{\text{sample}}$. The indicator electrode is constructed of a reference electrode contained within a membrane and an internal reference electrolyte of fixed activity, $(a_i)_{\text{internal}}$. The potential (E_{cell}) is measured by a pH/mV meter, and is equal to the sum of the potential between the internal ($E_{\text{ref,int}}$) and external ($E_{\text{ref,ext}}$) reference electrodes, the membrane potential (E_{memb}), plus the liquid junction potential (E_{lj}) that exists at the junction between the external reference electrode and the sample solution.

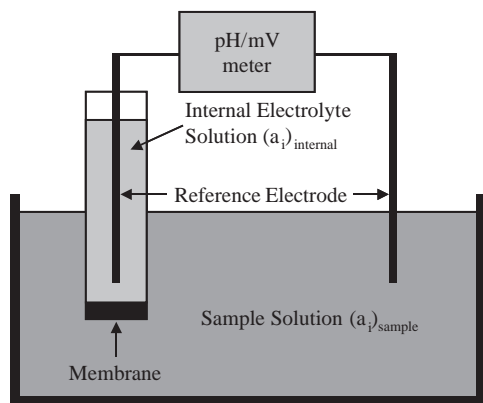


Fig. 1 General schematic of a potentiometric electrochemical cell.

$$E_{\text{cell}} = E_{\text{ref,int}} - E_{\text{ref,ext}} + E_{\text{memb}} + E_{\text{lj}} \quad (1)$$

If the semipermeable membrane is selective for a particular ion (i), a potential gradient will develop across the membrane. The gradient, described by the Nernst equation, depends on the ratio of the activity of the ion on either side of the membrane:

$$E_{\text{memb}} = \frac{RT}{zF} \ln \frac{(a_i)_{\text{sample}}}{(a_i)_{\text{internal}}} \quad (2)$$

where z is the charge on the ion, R is the gas constant ($8.314 \text{ J mol}^{-1} \text{ K}^{-1}$), T is the temperature (K), and F is the Faraday ($96,485 \text{ C mol}^{-1}$). By the substitution of Eq. 2 into Eq. 1, an overall expression for the electrochemical cell can be written as follows:

$$E_{\text{cell}} = E_{\text{ref,int}} - E_{\text{ref,ext}} + \frac{RT}{zF} \ln \frac{1}{(a_i)_{\text{internal}}} + E_{\text{lj}} + \frac{RT}{zF} \ln(a_i)_{\text{sample}} \quad (3)$$

The potentiometric cell is constructed so that the half-cell potentials of the two reference electrodes and the composition of the internal electrolyte are constant. The sample solution is maintained at a high constant ionic strength so that E_{lj} is effectively constant. This allows Eq. 3 to be simplified to yield:

$$E_{\text{cell}} = K + \frac{0.0591}{z} \log(a_i)_{\text{sample}} \quad (4)$$

when the temperature is 25°C , and K represents the terms which are constant in Eq. 3. This relationship between cell potential and analyte activity is the basis of potentiometry. A plot of E_{cell} vs. $\log a_i$ for a series of analyte concentrations

should be linear with a slope of $0.0591/z$ for measurements made at 25°C .

Unfortunately, membranes respond to some extent to ions in solution other than the analyte. A more general expression than Eq. 4 is:

$$E_{\text{cell}} = K + \frac{0.0591}{z} \log(a_i + k_{ij}a_j^{z/x}) \quad (5)$$

where a_j is the activity of the interfering ion j , x is the charge of ion j , and k_{ij} is the selectivity constant of the measurement. Small values of k_{ij} are characteristic of electrodes with good selectivity for the analyte i .

Faradaic Techniques

Faradaic techniques are those in which oxidation or reduction of analyte species occurs at the electrodes and therefore a measurable current is passed through the electrochemical cell. This discussion will be limited to controlled-potential techniques, primarily voltammetry and amperometry, coupled to liquid chromatography. While other Faradaic electrochemical techniques have been developed and electrochemical techniques in bulk solution are common, the use of liquid chromatography employing these two detection strategies is by far the most common electroanalytical technique in pharmaceutical studies.

Application of some potential waveform to an electrode in contact with the sample solution is the basis of all controlled-potential electrochemical techniques. The different techniques simply manipulate the monitored output by either varying the applied potential waveform or by changing the response domain. This response can be current, charge, or in some cases photons. An advantage of electrochemical techniques is that the instrumental requirements are rather simple and essentially the same for all techniques. A basic system consists of a potentiostat for potential control of an electrode and a means of detecting a response (usually the current through the electrode). Indeed, single instruments capable of performing many electrochemical techniques are available from several sources.

For successful application of Faradaic electrochemical techniques, it is necessary to understand the fundamental processes that occur at the surface of the electrode in electrolyte solution. When a potential is applied between two electrodes in solution, a narrow interphase region, the electrical double-layer, develops at the surface of the electrodes. All oxidative or reductive electrochemical reactions between the electrode and the analyte occur in this interphase region between the electrode surface and bulk solution. The bulk solution will remain at electro-neutrality because the potential drop between the electrodes only exists in the interphase region. Therefore,

molecules in the bulk solution cannot feel the presence of the electrodes or the potential gradient. Only molecules within the electrical double-layer (typically less than 10^{-6} m from the electrode) are probed during an electrochemical experiment.

It is equally important to understand how the electrode is experimentally set at some potential relative to the bulk solution. The potential difference between the electrode and the bulk solution cannot be measured; therefore, it is necessary to use a second electrode. Because the potential difference between two electrodes in solution can be determined, if the solution electrode potential difference of one of the electrodes is held constant, then the potential difference between the solution and the second electrode can be directly related to the potential difference between the two electrodes. The potential between one electrode and the solution is held constant by employing a reference electrode. The reference electrode uses a reversible redox couple such as silver–silver chloride to maintain a constant potential of fixed reference value. The potential between the bulk solution and the electrode is established relative to this reference point, even though it is not known in an absolute sense.

In the two-electrode system described before, the electrode of reference potential is known as the counter electrode while the electrode at which the potential difference is applied relative to the counter electrode is termed the working electrode. It is at the working electrode where the redox process of interest occurs. In addition to maintaining a reference potential, the counter electrode serves the second purpose of completing the electrical circuit and allowing charge to pass through the cell. Unfortunately, the two functions of the counter electrode are not independent of one another. When a current is passed through the counter electrode, a change in its potential difference with the solution can arise. This effect appears at even low currents and in most designs the problem is circumvented by splitting the counter electrode into two electrodes to give an overall three-electrode system. In the three-electrode system, the reference electrode maintains the constant reference potential, the electrode used to complete the electrical circuit is termed the auxiliary electrode, and the electrode at which the potential is applied to drive redox reactions is known as the working electrode (Fig. 2). In this arrangement there is negligible current through the reference electrode so its potential relative to the solution is constant regardless of the current through the electrochemical cell (working to auxiliary electrode).

In a high dielectric medium like water, when a potential is applied between an electrode and the solution, charge excess will develop on the electrode surface and the

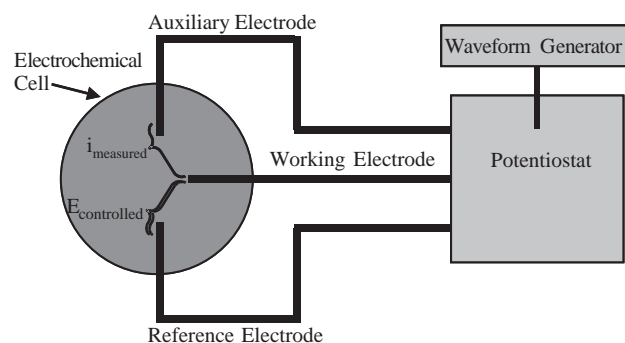


Fig. 2 General schematic of a Faradaic electrochemical cell.

amount of charge will be proportional to the applied potential. Solvent molecules and electrolyte species orient themselves at the electrode surface to counter the electrode charge. This is the interphase region, and is known as the double-layer because of the theoretical arrangement of molecules. This electrode–solution interface behaves in a manner similar to a capacitor. “Double-layer charging currents” are thus produced when the potential difference between the electrode surface and the solution is changed. These charging currents have little consequence in constant potential techniques such as amperometry; however, they can have a significant effect on voltammetric techniques, which will be discussed later.

Faraday’s law is the basis of all electroanalytical chemistry:

$$Q = nFN \quad (6)$$

where Q is the charge (C), n is the number of electrons transferred (equiv. mol^{-1}), F is the Faraday constant ($96,485 \text{ C equiv.}^{-1}$), and N is the number of moles of reactant. Faraday’s law shows the direct relationship between the charge passed through an electrochemical cell and the quantity of chemical reaction that has taken place. For analytical purposes, it is often more useful to take the derivative of Faraday’s law with respect to time:

$$i(t) = (dQ/dt) = nF(dN/dt) \quad (7)$$

where $i(t)$ is the current (A) as a function of time and dN/dt is the rate of conversion of reactant at the electrode surface. In this form, Faraday’s law becomes a rate equation relating the rate of a chemical reaction to an electrical response. Electrochemistry is one of the few techniques that make instantaneous, single point rate measurements. The equations describing the responses for the various electrochemical techniques are all based upon this equation.

Consideration of Faraday's law shows the advantage of electrochemistry for analytical purposes. With currents as small as 10^{-9} A, which are easily measured, conversion rates of as little as 10^{-14} equiv. sec^{-1} are readily detectable. This means detection limits in the range of 10^{-9} M can be achieved by electroanalytical techniques.

While electroanalytical techniques are inherently quite sensitive, the resolution of a mixture of electroactive compounds is quite difficult. Practical considerations limit the usable "potential window" to not more than 3 V and typically around 1.5 V. This is because at more extreme potentials, the medium or the electrode itself begins to oxidize or reduce. In addition, the electrochemical response of compounds as a function of applied potential is fairly broad so that at least a 200 mV–400 mV difference in half-wave potentials is required for adequate resolution. This typically limits electrochemical resolution of mixtures to not more than three or four electroactive compounds.

Because of this lack of resolving power, much electroanalytical research is aimed at providing increased selectivity. This can be accomplished in two ways. First, electrochemistry can be combined with another technique, which provides the selectivity. Examples of this approach are liquid chromatography with electrochemical detection and electrochemical enzyme immunoassay. The second approach is to modify the electrochemical reaction at the electrode to enhance selectivity. This approach is exemplified by modified electrode methods where reaction at the electrode surface is limited beyond mere electrochemical considerations to include physical and chemical properties. The following discussion will illustrate in detail how these approaches can provide analytical techniques with both high selectivity and low detection limits.

ION-SELECTIVE ELECTRODES (ISEs)

The ISE has become key in the application of potentiometric applications. In essence, the ISE measures the potential difference across a membrane, and follows the form shown in Fig. 1. Therefore, the measured cell potential, i.e., the electrochemical response, is based on the interaction between the membrane and the analyte that alters the potential across the membrane. The selectivity of the ISE to the analyte depends on the specificity of the membrane interaction with the analyte.

The development of membranes that exhibit both sensitivity and selectivity for species of interest is paramount in the application of ISEs. Between these two properties, selectivity is by far the more difficult one to

achieve. Three basic types of membranes have been developed, because of their selectivity, for ISEs. These are liquid and polymer, solid state, and glass membranes. Electrodes are commercially available for numerous inorganic ions such as H^+ , Li^+ , Na^+ , K^+ , Ag^+ , Ca^{2+} , Cu^{2+} , Pb^{2+} , Cd^{2+} , F^- , Cl^- , Br^- , I^- , SCN^- , NO_3^- , ClO_4^- , and BF_4^- . Additionally, electrodes have been reported for numerous organic anions and cations.^[1–5]

A number of ISEs for organic ions have found application in pharmaceutical analysis.^[6–8] The use of ISEs ranges from determination of excipients^[9] and formulary impurities^[10,11] in finished products to assay of drug product in pharmacokinetic studies.^[12–14]

Ion-Sensitive Field Effect Transistors (ISFETs)

The ISFET is an electrochemical sensor based on a modification of the metal oxide semiconductor field effect transistor (MOSFET). The metal gate of the MOSFET is replaced by a reference electrode and the gate insulator is exposed to the analyte solution or is coated with an ion-selective membrane^[15] as shown in Fig. 3.

The rest of the device is protected by a suitable encapsulant. Insulators such as SiO_2 , Al_2O_3 , and Ta_2O_5 have surface hydroxyl groups that act as sites for chemical reactions when exposed to an electrolyte solution. The alteration in surface charge resulting from protonation/deprotonation of these surface sites affects the surface potential and results in a relation between the drain current of the ISFET and solution pH. For example, an Al_2O_3 -ISFET exhibits a linear response between amplifier voltage and pH over a range of ca. 2.5–10.5 with a slope of 53 mV/pH.^[15] Additionally, ISFET devices for the determination of lidocaine,^[16] cocaine,^[16] benzalkonium,^[16] salicylic and acetylsalicylic acid,^[17] as well as cholanolic acids^[18] have been reported. The main interest in ISFETs stems from the ease with which inexpensive microelectrodes can be produced. However, the commercial success of these devices has been hindered by problems with encapsulation and stability.

LIQUID CHROMATOGRAPHY/ELECTROCHEMISTRY

Liquid chromatography/electrochemistry (LCEC) has become recognized as a powerful tool for the trace determination of easily oxidizable and reducible compounds. This is because detection of as little as 0.1 pmol of material is readily accomplished with relatively simple

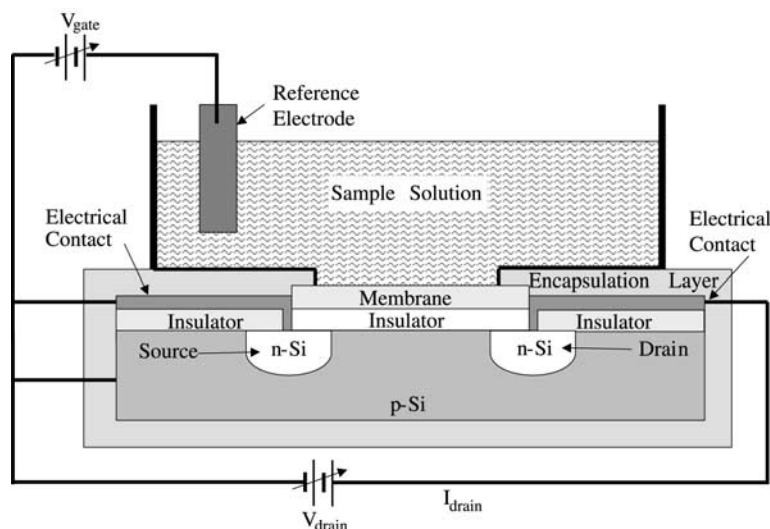


Fig. 3 General schematic of an ISFET.

and inexpensive equipment. Initial interest in LCEC was generated by the determination of several aromatic metabolites of tyrosine in the central nervous system. However, the application of LCEC into other areas of analysis including pharmaceutical analysis and especially pharmacokinetic and pharmacodynamic studies has become common.^[19–22]

Hydrodynamic Voltammetry

The basic characterization of an electrochemical detector for LCEC is the hydrodynamic voltammogram (HDV). Hydrodynamic voltammetry is a steady-state technique in which the electrode potential is altered with successive analyte injections and the current is plotted as a function of the potential. Idealized HDVs for the case of electrolyte solution (mobile phase) alone and with an oxidizable species added are shown in Fig. 4. The HDV of a compound begins at a potential where the compound is not electroactive and therefore no Faradaic current occurs, goes through a region of increasing current, and finally reaches a limiting current plateau, where essentially the entire compound reaching the electrode surface is oxidized. The HDV is characterized by the half-wave potential, $E_{1/2}$, which is defined as the potential at which the current is one-half its limiting value. The slope of the rising portion of the HDV is determined by the Nernst equation and the kinetics of electron transfer at the electrode surface.

The HDV is used to select the operating potential for an LCEC experiment. Two considerations are of major importance. First, quantitative determinations are based on

the extent to which the redox current from the analyte can be distinguished from the background current. Hence, operating at the smallest reasonable applied potential means that the background current typically increases with increasing potential. Second, operating on the limiting current plateau means that the redox current from the analyte is independent of applied potential. Because changes to the electrode surface during the course of an LCEC experiment can cause small changes in the interphase potential difference, operating on the limiting current plateau assures a constant response even if these

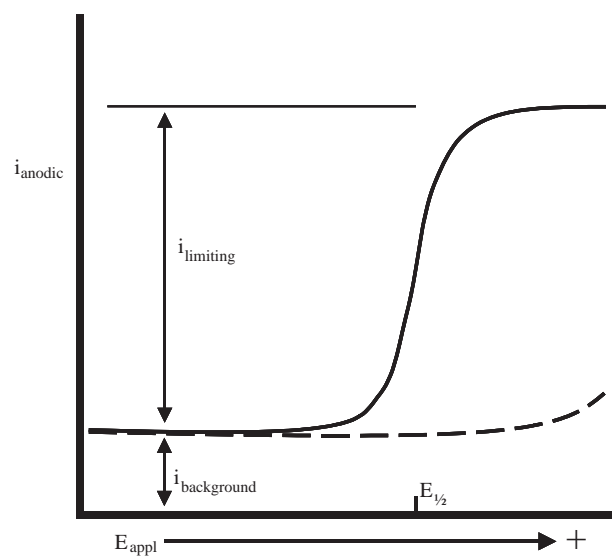


Fig. 4 General depiction of an HDV.

small variations occur. Overall, these considerations dictate that a potential just on the limiting current plateau is normally chosen for LCEC experiments.

Mobile Phase Considerations

While all detectors place some limitations on the mobile phase composition, in electrochemical detection, it is essential to recognize that a complex surface reaction is involved, which depends on both the physical and chemical properties of the medium. To optimize an LCEC determination, it is necessary to consider both chromatographic and electrochemical requirements simultaneously. Fortunately, most commonly applied chromatographic techniques fall into the category of reverse phase separations, the mobile phase requirements of which are consistent with the requirements for electrochemistry. The primary requirement for electrochemical detection is that the mobile phase has a relatively high conductivity. To this end, buffered mobile phases of moderate ionic strength (0.01 M–0.1 M) are typically used. Obviously, totally nonpolar organic mobile phases cannot be used because of their inability to support a significant ionic strength. This rules out classical normal phase separations. The vast majority of LCEC applications have employed reverse phase, ion-exchange, or ion-pair separations. While aqueous in nature, significant amounts (up to 90%, v/v) of organic modifier can be added to the mobile phase. Totally nonaqueous mobile phases can be used if the solvent can support an electrolyte. Organic solvents, which have been used with LCEC, include methanol, acetonitrile, DMF, and DMSO with salts such as tetrabutylammonium hexafluorophosphate or tetrafluoroborate as the electrolyte.

Electrode Materials

The choice of the electrode material can be critical to the successful use of LCEC. Historically, the material of choice is carbon paste, a mixture of graphite powder with a dielectric binder such as mineral oil. Carbon paste exhibits low background current at positive potentials and good mechanical stability. However, carbon paste electrodes are incompatible with mobile phases containing more than about 20% organic solvent. As a result, glassy (or vitreous) carbon electrodes have become more popular than carbon paste electrodes. This is primarily because glassy carbon is compatible with organic solvents and is easier to use. In addition, glassy carbon typically exhibits lower background currents at negative potentials than does carbon paste. More recently, diamond electrodes have been

investigated, although they have yet to find widespread appeal. Mercury electrodes have also been employed with LCEC. Because of the poor mechanical stability of mercury alone, it is often amalgamated with gold to form an electrode. Mercury has the best characteristics when operating at negative potentials, but it is generally not applicable to oxidative detection because of the low oxidation potential of mercury itself. Platinum electrodes have been employed to a limited extent, but have found little acceptance due to the problem of surface oxide layer formation. It is impossible to say that one electrode material will be superior to another in all situations. It is best to evaluate each redox system of interest with several types of electrodes to make the optimal choice.

Cell Design

The design of the electrochemical flow cell can dramatically affect the performance of the detector. For this reason, several different cell designs have been devised for chromatographic detection. The most popular design is the thin-layer cell in which the working electrode is part of one wall of a thin-layer channel through which the chromatographic eluent flows. Thin-layer cells can be constructed such that flow is parallel to the electrode^[23] or directed perpendicular to the surface followed by radial dispersion (known as the wall-jet cell)^[24] as shown in Fig. 5. The thin-layer cell design provides a high ratio of electrode surface area to solution volume providing high sensitivity and the excellent flow characteristics of well-developed laminar flow for low flow-related noise. In addition, the thin-layer cell design is compatible with essentially all electrode materials and offers convenient access to the electrode for resurfacing. Most commercially available electrochemical detectors use thin-layer cells.

Flow-through electrochemical detectors based on a cylindrical geometry, as opposed to a planar geometry, have also been developed.^[25–28] Three cell designs using cylindrical geometry have been used with liquid chromatography (Fig. 5). From a practical perspective, open-tubular electrodes are less attractive because many common electrode materials cannot be easily used (e.g., mercury and carbon paste) and resurfacing of electrodes is difficult. The surface-to-volume ratio is also not very large. The wire in a capillary cell design has become more attractive with the advent of open-tubular liquid chromatography columns. Knecht and coworkers^[29] have described the fabrication of such an electrochemical cell directly at the end of an open-tubular chromatographic column.

Packed-bed electrodes (Fig. 5) are useful for constructing electrochemical detectors of high conversion efficiency.

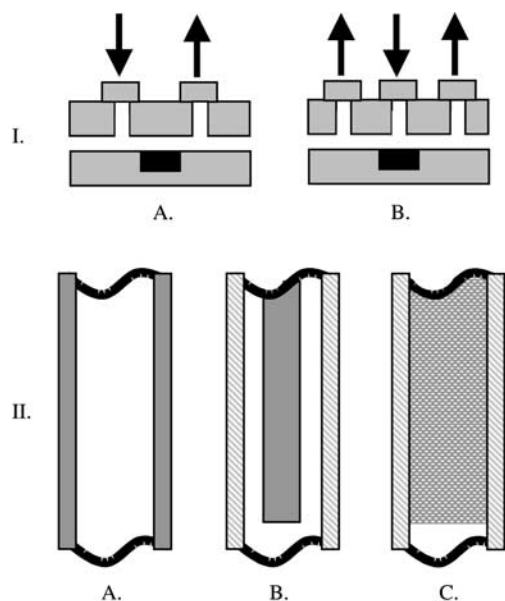


Fig. 5 Flow cells for LCEC: IA) thin-layer cell; IB) wall-jet cell; IIA) open-tubular cell; IIB) wire in a capillary cell; IIC) packed-bed cell.

Because typical packed-bed cells have very large electrode surface areas, high background currents are a problem. Another disadvantage of packed-bed cells is the difficulty of resurfacing the electrode. It is not possible to physically resurface the electrode by polishing and less vigorous chemical methods must suffice. Tubular packed-bed electrode cells are best applicable for electrosynthetic applications where 100% conversion is important.

Most LCEC applications use a single working electrode, although it is possible to monitor the current at several electrodes simultaneously. In practice, this becomes more difficult as more working electrodes are employed so that practical applications have been limited to two working electrodes. Several advantages can be achieved using two working electrodes relative to single-electrode detection.^[30] This configuration, known as a dual-electrode detector, is commercially available in both thin-layer and packed-bed designs.

Both cell designs permit the positioning of the second electrode downstream of the first working electrode (Fig. 6), which is known as the series configuration. This electrochemical transducer is used in the same manner as the classical ring-disk electrode. Products generated at the upstream electrode are detected (or collected) at the downstream electrode.^[31] Selectivity is enhanced when the products of the upstream electrolysis can be detected at a more favorable potential than is necessary for the upstream reaction. In addition, chemically irreversible redox couples

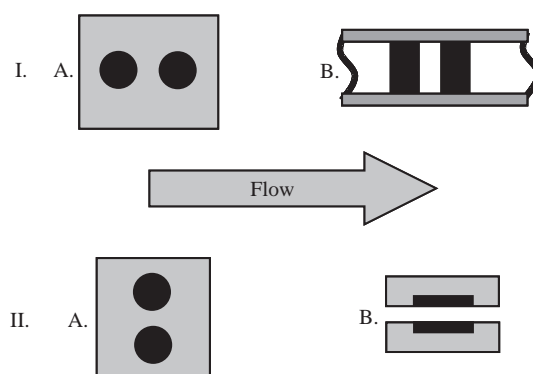


Fig. 6 Dual-electrode cells for LCEC: I) dual-series cells; A) planar; B) cylindrical; IIA) dual-parallel cell; IIB) dual-opposed cell.

are not detected downstream in the series configuration further enhancing selectivity and improving detection limits. This cell design has been applied to the determination of terbutaline in pharmaceutical formulations,^[32] acetaminophen in blood^[33] and urine,^[34] metanephrine and normetanephrine in plasma,^[35] physostigmine and metabolite in blood,^[36] samodil fumarate and metabolite in blood,^[37] oxodipine in plasma,^[38] and roxithromycin in plasma and urine.^[39] The series configuration has also been used to study the electrochemical reactions occurring at the upstream electrode.^[40]

The thin-layer cell design also permits two electrodes to be placed adjacent to each other and normal to the flow path (Fig. 6), which is known as the parallel adjacent configuration. With this configuration, the current at each electrode (typically at different applied potentials) is monitored individually and is analogous to dual-wavelength UV absorption detectors. The parallel configuration can be used in three ways. First, both oxidizable and reducible compounds can be detected simultaneously.^[41] Second, easily oxidizable (or reduced) analytes can be selectively detected in the presence of hard to oxidize (or reduce) analytes by using one electrode at a relatively low potential to detect the easily oxidized analytes and using the other electrode to detect the hard to oxidize analytes at a less selective potential.^[42] This technique was applied to the determination of water-soluble vitamins in pharmaceutical formulations.^[43] Finally, by operating both electrodes at different potentials on the HDV of a compound, voltammetric data can be obtained from a single chromatographic analysis providing qualitative information^[42] as reported for the analysis of Pt-based antineoplastics.^[44]

A third configuration of the dual-electrode thin-layer cell is also shown in Fig. 6. This geometry is known as

the parallel-opposed configuration and provides enhanced response for reversible compounds by redox cycling. As the analyte is cycled from its oxidized to reduced form, many more electrons are transferred than would be with a single working electrode and the current response is thus greatly amplified. Unfortunately, with conventional LC systems, the flow rate through the electrochemical cell is too high to permit a significant number of redox cycles to occur. Goto et al.^[45] have been able to achieve up to a 20-fold increase in response using 1 cm long electrodes spaced 30 μm apart with a flow rate of 20 $\mu\text{L min}^{-1}$ with a microbore chromatography column. With the growing interest in microbore LC and the associated lower volume flow rates, this concept should find more practical application.

While most LCEC experiments directly use the current response at the electrode for quantitation, it is possible to combine the response from two electrodes into a difference response. Many advantages have been described^[46] for LCEC using the series and parallel configurations of the dual-electrode detector in the difference mode. With the parallel configuration, the electrodes are operated at potentials bracketing the region where the current response is most dependent on potential (i.e., $E_{1/2}$). The difference response from the analyte is not significantly different from the response at a single electrode operated at the higher potential. However, the difference response for compounds both harder and easier to oxidize is greatly attenuated. In essence, difference mode detection with parallel configurations establishes a “potential window” through which only compounds whose current response changes can be seen. The major sources of baseline interference in LCEC, such as flow rate and temperature fluctuations, static discharge, and changes in mobile phase composition are common modes (i.e., occur at both electrodes simultaneously). Difference mode detection can be an effective method to improve signal quality because these types of interference signals are greatly attenuated in the difference signal.

With the series configuration, both electrodes are operated at the same potential and on the limiting current plateau of the analyte. Because the diffusion layer is depleted of the analyte at the upstream electrode, the downstream response is greatly attenuated and the difference response is not greatly different than the upstream response alone. However, signal from baseline interferences is greatly attenuated for a large improvement in the signal-to-noise ratio. Because this mode is based on mass transport instead of electrochemistry, no change in selectivity relative to single electrode operation is achieved.

Conversion Efficiency

Depending upon the cell design and flow rate, electrochemical detectors can electrolyze from ca. 5% up to nearly 100% of the material entering the cell. In practice, high efficiency is obtained by using a large surface area electrode. As the electrode surface area is increased, each added increment contributes proportionately less to the total signal, but approximately equal amounts to the background. As background noise is roughly proportional to background current, the signal-to-noise ratio tends to decrease as the electrode area becomes very large. Therefore, although the absolute response of an analyte is larger for high efficiency detectors, limits of detection are typically not better and often worse than for detectors with a conversion efficiency of 10%–20%. A detector of 100% conversion efficiency can be useful for the quantitation of an analyte for which no standard is available. If the number of electrons involved in the electrochemical reaction is known and the chromatographic peak is baseline resolved so that an accurate peak area can be determined, then quantities can be calculated directly from Faraday's law. This is not a very common occurrence in practice.

Voltammetric Detection

To obtain more complete voltammetric information than is readily found with amperometric detection, various potential scanning detectors have been developed for liquid chromatography. Simple use of a linear ramp waveform, such as used in linear sweep and cyclic voltammetry, is not very useful for chromatographic detection. Because the signal current of an analyte is relatively small in LC detectors, double-layer charging currents can easily obscure the voltammetric information. In addition, the cell resistance is typically large in cells used for chromatographic detection causing a significant iR drop that distorts the voltammogram.

To overcome these problems, most voltammetric detectors have used pulsed waveforms such as staircase,^[47] squarewave,^[48] and differential pulse.^[49] The current is sampled at the end of the pulse after the charging current has decayed. In addition, because the charging current is typically the major current source, iR problems are not as severe. Last^[50] has described a coulometric detector based on charge pulses instead of potential pulses, which eliminates iR and charging current problems. Even with these pulse techniques, voltammetric detection methods have not been able to achieve the detection limits readily obtained by simple amperometric detectors. White et al.^[51] have described a voltammetric detector using a fiber

microelectrode that exhibited greatly reduced charging currents. Unfortunately, this detector was limited to use with open-tubular chromatographic systems.

A unique voltammetric detector^[52] has been developed using a series dual-electrode cell. In the detector, the upstream electrode's potential is scanned while the downstream electrode is used to monitor the redox reaction occurring at the upstream electrode without the charging current contributions. In essence, the upstream electrode is operated voltammetrically and the downstream electrode operated amperometrically so that the detector has been named a voltammetric–amperometric detector. Detection limits of 10^{-7} M have been reported using this detection scheme.

ENZYME-LINKED ELECTROCHEMICAL TECHNIQUES

Monitoring enzyme-catalyzed reactions by voltammetry and amperometry is an extremely active area of bioelectrochemical interest. Whereas liquid chromatography provides selectivity, the use of enzymes to generate electroactive products provides specificity to electroanalytical techniques. In essence, enzymes are used as a derivatizing agent to convert a nonelectroactive species into an electroactive species. Alternatively, electrochemistry has been used as a sensitive method to follow enzymatic reactions and to determine enzyme activity. Enzyme-linked immunoassays with electrochemical detection (discussed in “Electrochemical Immunoassay”) have been reported to provide even greater specificity and sensitivity than other enzyme-linked electrochemical techniques.

Enzyme-linked electrochemical techniques can be carried out in two basic ways. The first approach is to use a hydrodynamic technique, such as flow injection analysis (FIAEC) or liquid chromatography (LCEC), with the enzyme reaction being either off-line or on-line in a reactor prior to the amperometric detector. In the second approach, the enzyme is immobilized at the electrode. Hydrodynamic techniques provide a convenient and efficient method for transporting and mixing the substrate and enzyme, subsequent transport of the substrate to the electrode, and rapid sample turnaround. The kinetics of the enzyme system can also be readily studied using hydrodynamic techniques. Immobilizing the enzyme at the electrode provides a simple system that is amenable to *in vivo* analysis. Alternatively, the transport of enzyme product from the enzyme active site to the electrode surface is greatly enhanced when the enzyme is very near the electrode. Enzyme electrodes are an extremely

important area of bioelectrochemical analysis, and many reviews are available in the literature.^[53–55]

A wide variety of enzymes have been used in conjunction with electrochemical techniques. The only requirement is that an electroactive product is formed during the reaction, either from the substrate or as a cofactor (i.e., NADH). In most cases, the electroactive products detected have been oxygen, hydrogen peroxide, NADH, or ferri/ferrocyanide. Some workers have used the dye intermediate used in classical colorimetric methods because these dyes are also typically electroactive. Although an electroactive product must be formed, it does not necessarily have to arise directly from the enzyme reaction of interest. The ability to use several coupled enzyme reactions extends the possible use of electrochemical techniques to essentially any enzyme system.

Off-Line Techniques

The simplest method of coupling enzymatic reactions to electrochemical detection is to monitor an off-line reaction using FIAEC or LCEC. The enzymatic reaction is carried out in a test tube under controlled conditions with aliquots being taken at timed intervals. These aliquots are then analyzed for the electroactive product and the enzyme activity in the sample is calculated from the generated kinetic information. Because LCEC had its initial impact in neurochemical analysis, it is not surprising that many of the early enzyme-linked electrochemical methods are of neurologically important enzymes. Additionally, the activity of lipase in pharmaceutical formulations^[56] has been determined by electrochemical means.

In addition to enzyme activity, the concentration of a nonelectroactive substrate can be determined electrochemically by this technique. By keeping the substrate (analyte) as the limiting reagent, the amount of product produced is directly related to the initial concentration of substrate. Either kinetic or equilibrium measurements can be used. Typically, an enzyme that produces NADH is used because NADH is readily detected electrochemically. Glucuronide and sulfate metabolites of pharmaceuticals^[57] and lecithin in crude drug formulations^[58] have been determined by LCEC through off-line enzymatic reaction.

Enzyme Reactors

By incorporating the entire analytical scheme (enzyme reaction and electrochemical detection) into the flow system, a great improvement in precision can be realized. Sample manipulation is minimized because only a single

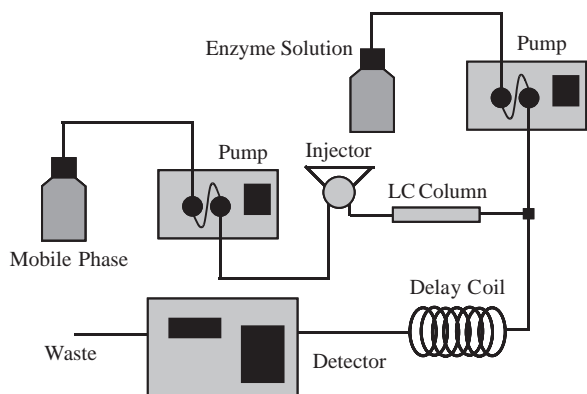


Fig. 7 General schematic of an LCEC system with an on-line enzyme reactor.

injection into the flow system is required vs. sampling of aliquots for the off-line method. Precision is also improved because the timing of the enzyme reaction and detection are much better controlled in the flow system. Finally, less of both enzyme and sample are needed with on-line enzyme reactor methods.

The simplest design for an enzyme reactor is to merely have the substrate and enzyme in two merging buffer streams followed by a reaction delay coil (Fig. 7). The delay coil can be cut to the optimum length for complete substrate conversion to product or for any desired percentage of conversion. The electrochemical detector is placed after the delay coil. Throughput is high because, even though a reaction time of several minutes may be required, several samples can be in the flow stream at the same time. Potter et al.^[59] have determined choline and acetocholine using acetocholinesterase and cholin oxidase to produce hydrogen peroxide. The hydrogen peroxide was detected amperometrically at a platinum electrode. The choline and acetocholine were separated by liquid chromatography before mixing with the enzymes. Other reports include the analysis of salicylic acid by salicylate hydroxylase,^[60] choline derived from L- α -glycerophosphorylcholine in pharmaceutical formulations incubated with glycerophosphorylcholine phosphodiesterase and subsequent post chromatographic enzyme reaction,^[61] glucuronides of fenoldopam with β -glucuronidase,^[62] sulfate conjugates of fenoldopam with sulfatase,^[62] and choline with choline oxidase.^[58]

A great savings in enzyme consumption can be achieved by immobilizing the enzyme in the reactor. In addition to the smaller amount of enzyme required, immobilization often increases the stability of the enzyme. Several designs of immobilized-enzyme reactors (IERs) have been reported, with open tubular and packed bed being the most popular. Open-tubular reactors offer low dispersion, but have a

relatively small surface area for enzyme attachment. Packed-bed reactors provide extremely high surface areas and improved mass transport at the cost of more dispersion.

Enzyme Electrodes (Biosensors)

The final method of coupling enzyme reactions to electrochemistry is to immobilize a biocatalytic material directly at the electrode surface. This biocatalytic material can be an immobilized enzyme, bacterial particles, or a tissue slice, as shown in Fig. 8. The biocatalyst converts substrate (analyte) into product, which is measured by the electrode. Electrodes of this type can be potentiometric or Faradaic, and are often referred to as "biosensors."

Biocatalytic membrane electrodes significantly expand the scope of direct potentiometry by enabling biosensors that respond to a whole host of organic substrates to be made. The selectivity of these sensors is a combination of the selectivity of the biocatalyst for the substrate and the ISE for other constituents in the sample that might reach the ISE surface membrane. Thus, the selectivity with respect to other organic constituents in the sample is determined by biocatalyst specificity, whereas the selectivity with respect to the measured analyte in the sample is determined by the underlying ISE.

A representative example of a potentiometric biosensor is an electrode for L-arginine.^[63] The bacterium *Streptococcus faecium* is immobilized on the gas permeable membrane of an ammonia electrode. Electrodes such as that for L-arginine exhibit excellent selectivity due to the high specificity of the enzyme arginine deiminase for L-arginine and the excellent selectivity of the ammonia electrode for

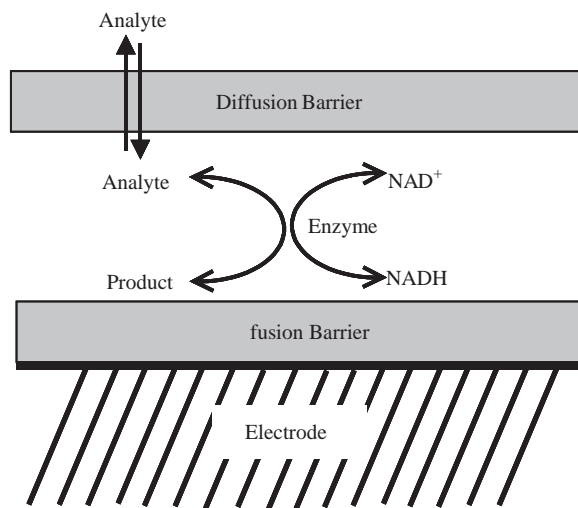


Fig. 8 General diagram of an electrochemical biosensor.

NH₃. The electrode responds to the concentration of NH₃ at its surface, which is proportional to the concentration of L-arginine in the sample. Nernstian plots give slopes of 40 mV–45 mV with a linear range of $1 \times 10^{-4} M$ – $6.5 \times 10^{-3} M$. The electrode is highly selective for L-arginine until contaminants grow in the biocatalyst layer, which initiates response to glutamine and asparagines. The growth of such contaminants can be discouraged under certain storage conditions.^[64] The working range of biosensors of this type is typically only 2–3 orders of magnitude with a detection limit of $10^{-5} M$ – $10^{-4} M$.

Cell-based biosensors have comparable response characteristics to enzyme electrodes, but offer several advantageous features.^[65]

- Strains of bacteria are generally less expensive than isolated enzymes. A specific sterile bacterial culture is not needed to prepare an electrode. For example, human dental plaque suffices to prepare electrodes for D(+)-glucose, D(+)-mannose, and D(–)-fructose.^[66]
- Enzyme activity is often enhanced in bacterial cells and remains longer due to the optimal environment. Consequently, lifetimes of bacterial electrodes average around 20 days compared to 14 days for enzyme electrodes. Cells can be regrown if the catalytic activity is lost.
- Bacteria can readily affect complex reaction sequences requiring cofactors. For example, an electrode for nitrilotriacetic acid (NTA) is practical only because pseudomonas bacterial cells contain all of the enzymes and cofactors necessary to execute the sequence of reactions on which the electrode response is based.^[67] The underlying gas electrode for ammonia responds to NH₃ formed in the last reaction, which is proportional to the concentration of NTA in the sample.

The concept of a biocatalytic membrane electrode has been extended to the use of a tissue slice as the catalytic layer.^[68] An example of this approach is an electrode for AMP, which consists of a slice of rabbit muscle adjacent to an ammonia gas electrode.^[69] NH₃ is produced by enzymatic action of rabbit muscle constituents on AMP. The electrode exhibits a linear range of $1.4 \times 10^{-4} M$ – $1.0 \times 10^{-2} M$ with a response time varying from 2.5 min to 8.5 min, depending on the concentration. Electrode lifetime is about 28 days when stored between use in buffer with sodium azide to prevent bacterial growth. Excellent selectivity enables AMP to be determined in serum.

Cell- and tissue-based electrodes are generally less selective than enzyme-based electrodes because many constituents of cells result in response to many possible

substrates. This problem may be solved by the use of specific inhibitors to block interfering reactions. These electrodes also suffer from long recovery times (3 hr–4 hr) after a measurement has been made. However, such electrodes may be sufficiently cheap to be disposable. Most biocatalytic membrane electrodes are susceptible to irreversible loss of activity when exposed to samples containing denaturing or toxic agents. The difficulties associated with electrode longevity, both during use and storage, have stymied the commercial development of biocatalytic membrane electrodes.

Electrodes based on the immobilization of enzymes at the electrode surface have been reported for a few pharmaceutical applications. Seegopaul and Rechnitz^[70] developed a CO₂ electrode that responds to methotrexate due to the inhibition of dihydrofolate reductase by the analyte. Additional applications have included the detection of penicillin consumption by penicillinase^[71] as well as urea by urease^[72] with a pH electrode and analysis of L-ascorbic acid by ascorbate oxidase.^[73]

Coatings of immobilized enzymes have been used with ISFET devices in a manner that is analogous to their use with the ISE. For example, an ISFET coated with penicillinase immobilized in a membrane responds to penicillin similarly to the ISE mentioned.^[74] Enzymatically coupled pH ISFETs have also been used for the measurement of urea,^[75] acetylcholine,^[75] and glucose.^[76]

The concept of combining an enzyme reaction with an amperometric probe offers all the advantages for potentiometric electrodes with a much higher sensitivity. For this reason, much effort has been applied to the development of new amperometric biosensors including significant work on pharmaceutical applications.^[77–84] Reports have included a dual-enzyme electrode for monitoring quality control of peptides in pharmaceutical formulations.^[85] This system is based on the oxidation of tyrosine by tyrosinase and reduction back by quinoprotein glucose dehydrogenase. The redox cycling of the enzyme sensor lowers the limit of detection of this sensor to $2 \times 10^{-7} M$. Additionally, the natural selectivity of lactate oxidase for the L-enantiomer was used to develop a specific biosensor capable of identifying racemic and enantiomerically pure samples.^[86]

ELECTROCHEMICAL IMMUNOASSAY

The remarkable selectivity that is inherent in the reaction of an antibody with the antigen or hapten against which it was raised is the basis for the extensive use of immunoassays. An electrode in which an antibody or an antigen/hapten is incorporated in the sensing element is

termed an “immuno-electrode.” The response of the immuno-electrode is based on an immunochemical reaction between the sensing element of the electrode and antibody or antigen/hapten in the sample solution. Immunoassay is based on the use of an antibody as a selective chemical reagent for an antigen or hapten analyte. Immunoassays are commonly categorized as heterogeneous, in which antibody-bound antigen is separated from free antigen at some point in the procedure, or homogeneous, in which there is no separation step.

Historically, most immunoassays have relied on radioisotope labeling for detection. The desire to avoid the use of radioisotopes has led to the development of alternative labels. One of the most successful alternatives has been to use enzyme labels, thus enzyme immunoassays. In enzyme immunoassays, substrate is added following the antigenic binding. Product is then detected and its concentration or rate of formation is a measure of antigen (analyte) concentration in the sample. Enzyme immunoassays can be extremely sensitive because of the chemical amplification provided by the enzymatic step. Chemical amplification is the passing of a substance through a cycling or multiplication mechanism to generate a large amount of product.^[87] In this way, a trace concentration of analyte results in a significantly higher concentration of product, which is actually detected, but is still related to the original concentration of analyte. Electrochemical enzyme immunoassay is the use of electrochemical techniques to detect the products from an enzyme immunoassay.^[88]

Potentiometric Immunoassays

Conventional ISEs have been used as detectors for immunoassays. Antibody-binding measurements can be made with hapten-selective electrodes such as the trimethylphenylammonium ion electrode.^[89] Enzyme immunoassays, in which the enzyme label catalyzes the production of a product that is detected by an ISE, take advantage of the amplification effect of enzyme catalysis in order to reach lower detection limits. Systems for hepatitis B surface antigen^[90] and estradiol^[91] use horseradish peroxidase as the enzyme label and an iodide electrode as the detector. The horseradish peroxidase catalyzes the oxidation of *p*-fluoroaniline with the fluoride detected by the ISE. Biotin^[92] and cyclic AMP^[93] have been determined using lysozyme and urease as labels. The immunoreaction between human antibody (IgG) and peroxidase-labeled antihuman IgG antibody can be detected with a fluoride electrode.^[94] Adenosine deaminase, asparaginase, and urease have been examined

as possible enzyme labels for immunoassays using potentiometric detection with the ammonia gas-sensing membrane electrode^[95] and adenosine deaminase with an NH_4^+ electrode.^[96] CO_2 electrodes have been used for the determination of human IgG^[97] and digoxin.^[98]

Heterogeneous Amperometric Immunoassays

Most heterogeneous immunoassays are run in a competitive format in which a standard labeled antigen is allowed to compete with sample antigen for a limited number of antibody sites. The ratio of standard-to-sample antigen that is bound then reflects the concentration of analyte in the sample. Another widely used assay format is the sandwich immunoassay. In this format, the antigen is “sandwiched” between two different antibodies, one of which is labeled for detection. Sandwich assays are more specific than competitive assays because two highly selective antibody reagents are used rather than one; however, the sandwich format is only applicable to antigens large enough to bind two antibodies simultaneously.

Several heterogeneous electrochemical enzyme immunoassays have been demonstrated.^[99–101] These are based on the enzyme-linked immunosorbent assay (ELISA) technique in which antibody is immobilized on the walls of a small volume plastic vessel. The ELISA technique can follow either a competitive equilibrium or a sandwich format. Both formats have been used with electrochemical detection. The general protocol for these two formats is shown in Fig. 9.

Heterogeneous electrochemical enzyme immunoassays have several advantageous features. The detection limit is typically in the low pg mL^{-1} range, and is a function of the antigen–antibody binding constant, rather than the ability to detect the enzyme product. Because the sample is rinsed from the reagent tubes before adding substrate, problems with interferences by electroactive constituents of the sample or possible electrode fouling by protein adsorption are eliminated. Finally, this technique should readily lend itself to automation for the routine analysis of many samples.

Homogeneous Amperometric Immunoassays

Homogeneous immunoassays rely on a change in the intensity of a label signal that occurs when labeled antigen binds with antibody. When the label is an enzyme, a reduction in the rate of enzyme catalysis forms the basis for the assay. This technique is commonly known as an EMIT (enzyme multiplied immunoassay technique) assay.

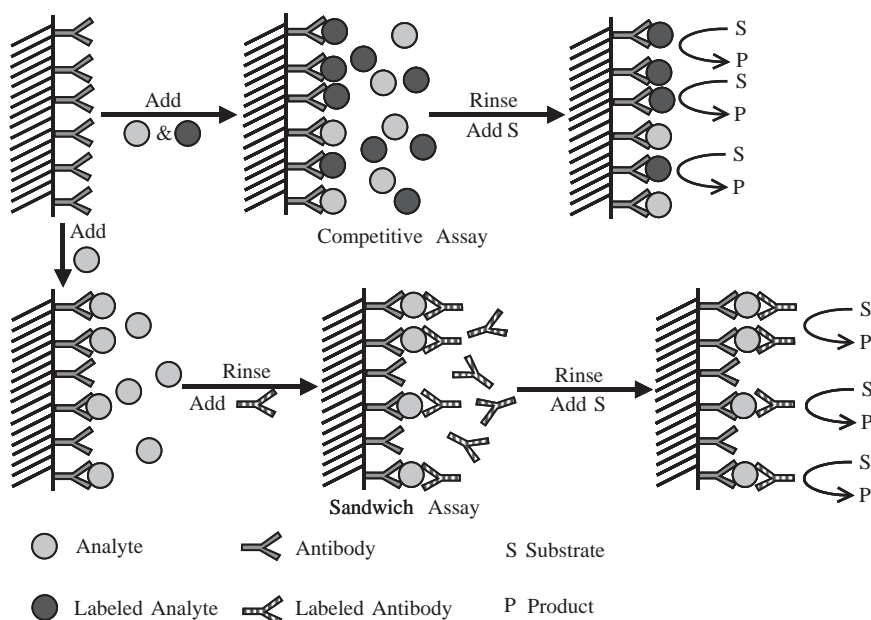


Fig. 9 General schematic of ELISA in competitive and sandwich formats.

This enables free, labeled antigen to be distinguished from bound antigen with no separation step necessary.

Homogeneous electrochemical enzyme immunoassays for both phenytoin^[102] and digoxin^[103] have been developed. In both cases, the label was glucose-6-phosphate dehydrogenase, which catalyzes the reduction of NAD^+ to NADH. The NADH produced was detected by LCEC. Many other homogeneous electrochemical immunoassays have also been reported.^[101,104]

The main advantage of a homogeneous immunoassay, compared to a heterogeneous immunoassay, is the absence of a separation step. This translates into a simpler procedure and easier automation. However, homogeneous assays are typically less sensitive and more easily susceptible to sample interferences which are removed in a separation step.

APPLICATIONS

Oxidative Applications

Many drugs are electroactive, and as such, have been determined using amperometry. Table 1 lists several compounds of pharmaceutical interest, which have been determined by electrochemistry. Alternatively, Jane et al. have reported the electrochemical response of some 462 basic drugs.^[105]

Phenols

Most phenolic compounds are readily oxidized at carbon electrodes. The oxidation potentials vary widely depending upon the number of ring hydroxyl groups and their positions on the ring. Many compounds of biomedical and industrial interest are phenolic and LCEC-based trace analysis determinations are quite popular.

Phenols constitute a major source of xenobiotic exposure to the body in the form of drugs and environmental pollutants. Oxidative metabolism of these compounds can lead to physiological damage; therefore, the metabolism of these compounds is of great interest. LCEC has been a powerful tool for investigating the metabolism of aromatic compounds by the cytochrome p-450 system.^[106]

Aromatic amines

Many pharmaceutical and environmental pollutants are aromatic amines. Like phenols, this class of compounds is generally oxidized at carbon electrodes. LCEC has been used to study the metabolism of aromatic amines of pharmaceutical origin.^[107]

Heterocyclic compounds

Many compounds of biomedical interest, both of endogenous and exogenous origin, are heterocyclic in structure. Many of these compounds are electroactive at potentials

Table 1 List of pharmaceutical products with electroactive structural components and reference to reported electrochemical determination

Class	Compound	Electrophore	Reference
Analgesics	Salicylic acid	Phenolic	[129]
	Acetaminophen	Phenolic	[129,132–137]
	Codeine	Phenolic	[139]
	Morphine	Phenolic	[141]
	Ciramideol	Phenolic	[143]
	Dezocine	Phenolic	[143]
	Nalbuphine	Phenolic	[145]
	Ketobemidone	Phenolic	[146]
Antibacterial	Sulfadimidine	Aromatic amine, heterocyclic	[147]
	Sulfamethoxydiazine	Aromatic amine, phenolic	[147]
	Sulfisoxazole	Aromatic amine	[147]
	Sulfamethoxazole	Aromatic amine	[147]
Antibiotics	Lincomycin	Heterocyclic	[150]
	Cephalosporins	Heterocyclic	[152]
	Penicillins	Heterocyclic	[150]
	Clarithromycin	Heterocyclic	[154]
	Erythromycin	Heterocyclic	[154]
	Gentamicin	Heterocyclic	[157]
Antidepressants	Benperidol	Indolic	[158]
	Haloperidol	Aromatic halide	[158,160]
	Desipramine	Heterocyclic, hydrazine	[130,131]
	Amitriptyline	Amine	[130,131]
Antidepressants	Imipramine	Heterocyclic, amine	[130,131]
Antihypertensives	Hydralazine	Heterocyclic	[138]
	Indoramin	Indolic	[140]
	Isradipine	Heterocyclic	[142]
Anti-inflammatory	Diflunisal	Phenolic	[144]
	Indomethacin	Phenolic, Indolic	[144]
	Naproxen	Phenolic	[144]
	Piroxicam	Heterocyclic	[144]
	Sulindac	Phenyl sulfoxide	[144]
Antineoplastics	Epirubicin	Phenolic	[148]
	Doxorubicin	Phenolic	[148]
	Flavopiridol	Phenolic	[149]
	Cisplatin	Metal center	[151]
Antipsychotics	Chlorpromazine	Heterocyclic	[153]
	Trimeprazine	Heterocyclic	[153]
	Risperidone	Heterocyclic	[155]
Antitubercular	Aconiazide	Hydrazine, phenolic	[156]
	Isoniazid	Hydrazine	[156]
Muscle Relaxants	Tizanidine	Heterocyclic	[159]
Psychotropic	Fluphenazine	Heterocyclic	[161]
	Perphenazine	Heterocyclic	[161]

useful for LCEC analysis. Methods for the determination of both ascorbic acid^[108] and uric acid^[109] were developed in the early days of LCEC. The important enzyme cofactors, the folates,^[110] the pterins^[111] biotin,^[112] and NADH,^[113] are all electroactive heterocycles that have been determined by LCEC.

Amino acids and peptides

While most amino acids are not electroactive at analytically usable potentials, much work is currently directed at general methods of amino acid detection by electrode surface modification or derivatization of the amino acid. Kok et al.^[114] have directly detected amino acids at a copper electrode. Several derivatization methods for amino acids have been reported.^[115,116]

The pentapeptides, met- and leu-enkephalin, have been detected in rat striatum tissue by LCEC at a glassy carbon electrode.^[117] These peptides can be detected directly because they contain an electroactive tyrosine residue. A series of endorphins, also containing tyrosine, have been detected by LCEC.^[118]

Thiols

Thiols are easily oxidized to disulfide in solution, but this reaction occurs only very slowly at most electrode surfaces. However, use can be made of the unique reaction between thiols and mercury to detect these compounds at very favorable potentials. The thiol and mercury form a stable complex that is easily oxidized. In a formal sense, it is the mercury and not the thiol which is actually oxidized in these reactions. For the LCEC determination of thiols, a Au/Hg amalgam electrode is used.^[119] Using a series dual electrode, both thiols and disulfides can be determined in a single chromatographic experiment.^[120,121] Additionally, carbon electrodes can be easily modified, for the determination of thiols, with many compounds including cobalt phthalocyanine^[122] and nickel hexacyanoferrate.^[123]

Reductive Applications

The majority of LCEC applications have used oxidative detection. This is likely because of the perceived difficulties encountered with reductive detection. In particular, dissolved oxygen and trace metal ions must be removed to prevent high background currents. These problems are not difficult to overcome^[124] and more applications of reductive detection should appear as this is more generally realized.

Nitro compounds

Aromatic nitro and nitroso compounds are easily reduced at carbon and mercury electrodes. Other nitro compounds such as nitrate esters, nitramines, and nitrosamines are also typically easily reduced. The complete reduction of a nitro compound consists of three two-electron steps (nitro to nitroso to hydroxylamine to amine). Since most organic oxidations are only two-electron processes, higher sensitivity is typically found for nitro compounds. Several LCEC-based determinations of nitro compounds have been reported.^[125]

Heterocycles

Several heterocycles of biomedical interest are reducible. Among these, the K vitamins^[126] and the pterins^[126] have been determined by LCEC. Some heterocyclic pharmaceuticals have also been determined by reductive LCEC.^[127,128]

FUTURE TRENDS

The application of microchip manufacturing technology to analytical chemistry has encouraged the development of even smaller analytical devices. Because the electrochemical detection does not suffer from a loss in performance when miniaturized, the incorporation of electrodes into microfabricated devices has received much attention. One of the greatest potentials of microfabricated analytical systems is the ability to perform a number of simultaneous parallel analyses. Because combinatorial synthesis provides a means to produce analyte libraries of a million or more compounds, the need for improved high throughput screening methodology will likely turn to microfabricated electrochemical instruments capable of simultaneously assaying many analytes. Developments in this area are hoped to provide more rugged electrochemical detectors that will further reduce to analyte and solvent consumption and as well as reduce analysis time, and provide simple, disposable instrumentation. Another potentially useful application of microfabricated separation-coupled electrochemical instrumentation would be in routine analysis. Often analytical methodology for quality control is limited by the ease with which the method can be applied at the manufacturing site. Reproducibly manufactured, disposable, integrated electrochemical instruments would simplify calibration and eliminate the need for analysts to remove and polish electrodes.

With the availability of less expensive and more dependable commercial instruments, liquid chromatog-

raphy coupled to mass spectrometry is quickly becoming the industry standard. However, the role of electrochemistry in pharmaceutical analysis has been well defined, and will likely continue to be preferentially employed in applications where low analyte concentrations, small sample volumes, or complex sample matrices requiring high specificity challenge the analytical method.

REFERENCES

- Hassan, S.S.M.; El-Bahnasawy, R.M.; Rizk, N.M. Membrane sensors for batch and flow injection potentiometric determination of ethamsylate (cyclonamine) in pharmaceutical preparations. *Mikrochim. Acta* **1997**, *126*, 217–222.
- Elnemma, E.M.; Hamada, M.A. Plastic membrane electrodes for the potentiometric determination of codeine in pharmaceutical preparations. *Mikrochim. Acta* **1997**, *126*, 147–151.
- Badawy, S.S.; Issa, Y.M.; Tag-Eldin, A.S. Potentiometric determination of L-dopa, carbidopa, methyl dopa, and aspartame using a new trinitrobenzenesulfonate selective electrode. *Electroanalysis* **1996**, *8*, 1060–1064.
- Panagos, D.S.; Koupparis, M.A. Construction of an iopanoic acid ion-selective electrode and its application to pharmaceutical analysis. *Analyst* **1995**, *120*, 1643–1646.
- Solich, P.; Macheras, P.E.; Koupparis, M.A. Construction of a diflunisal ion sensor and its use in automated flow-injection methods for assay content uniformity, and dissolution studies of formulations. *J. Pharm. Sci.* **1995**, *84*, 889–894.
- Menon, S.K.; Sathyapalan, A.; Agrawal, Y.K. Ion selective electrodes in pharmaceutical analysis—A review. *Rev. Anal. Chem.* **1997**, *16*, 333–353.
- Stefan, R.-I.; Baiulescu, G.-E.; Aboul-Enein, H.Y. Ion-selective membrane electrodes in pharmaceutical analysis. *Crit. Rev. Anal. Chem.* **1997**, *27*, 307–321.
- Kauffmann, J.-M.; Pekli-Novak, M.; Nagy, A. The potential of electroanalytical techniques in pharmaceutical analysis. *Acta Pharm. Hung.* **1996**, *66*, 57–64.
- Khalil, S.; Borham, N. Phenothiazine drug poly(vinyl chloride) matrix membrane electrodes and their use in pharmaceutical analysis. *Microchem. J.* **1999**, *63*, 389–397.
- Yuwono, M.; Ebel, S. Determination of fluoride impurities in calcium ascorbate. Comparison of gas chromatography and ion selective electrode potentiometry. *Arch. Pharm.* **1997**, *330*, 348–352.
- Hartofylax, V.H.; Efstathiou, C.E.; Hadjiioannou, T.P. Kinetic study of the tartaric acid–periodate reaction: determination of tartaric acid using an improved periodate-selective electrode. *Microchem. J.* **1986**, *33*, 9–17.
- Atta-Politou, J.; Koupparis, M.; Macheras, P. Development of a potentiometric kinetic method for drug adsorption studies: The chlorpromazine–charcoal model case. *Eur. J. Pharm. Sci.* **1996**, *4*, 223–230.
- Mitsana-Papazoglou, A.; Diamandis, E.P.; Hadjiioannou, T.P. Ion-selective electrodes for the H₂-receptor antagonists cimetidine and ranitidine. *J. Pharm. Sci.* **1987**, *76*, 485–491.
- Cunningham, L.; Freiser, H. Ion-selective electrodes for some β -adrenergic and calcium blockers. *Anal. Chim. Acta* **1984**, *157*, 157–162.
- Bergveld, P. The development and application of FET-based biosensors. *Biosensors* **1986**, *2*, 15–33.
- Campanella, L.; Aiello, L.; Colapicchioni, C.; Tomassetti, M. Lidocaine and benzalkonium analysis and titration in drugs using new ISFET devices. *J. Pharm. Biomed. Anal.* **1998**, *18*, 117–125.
- Su, Y.; Tomassetti, M.; Sammartino, M.P.; Crescentini, G.; Campanella, L. A new salicylate ISFET for the determination of salicylic and acetylsalicylic acid in drugs. *J. Pharm. Biomed. Anal.* **1995**, *13*, 449–457.
- Campanella, L.; Battilotti, M.; Borraccino, A.; Colapicchioni, C.; Sammartino, M.P.; Tomassetti, M. Cholanic acids determined in commercial drugs by means of a new ISFET device. *J. Pharm. Biomed. Anal.* **1993**, *11*, 1207–1214.
- LaCourse, W.R.; Dasenbrock, C.O. High-performance liquid chromatography–pulsed electrochemical detection for the analysis of antibiotics. *Adv. Chromatogr.* **1998**, *38*, 189–232.
- Eap, C.B.; Baumann, P. Analytical methods for the quantitative determination of selective serotonin reuptake inhibitors for therapeutic drug monitoring purposes in patients. *J. Chromatogr. B: Biomed. Appl.* **1996**, *686*, 51–63.
- Bersier, P.M.; Bersier, J. Polarographic, voltammetric, and HPLC–EC of pharmaceutically relevant cyclic compounds. *Electroanalysis* **1994**, *6*, 171–191.
- Kissinger, P.T.; Radzik, D.M. Liquid chromatography/electrochemistry in pharmaceutical analysis. *Drugs Pharm. Sci.* **1991**, *47*, 65–100.
- Kissinger, P.T.; Refshauge, C.; Dreiling, R.; Adams, R.N. Electrochemical detector for liquid chromatography with picogram sensitivity. *Anal. Lett.* **1973**, *6*, 465–477.
- Fleet, B.; Little, C.J. Design and evaluation of electrochemical detectors for HPLC [high-pressure liquid chromatography]. *J. Chromatogr. Sci.* **1974**, *12*, 747–752.
- Blaedel, W.J.; Schieffer, G.W. A hydrodynamic voltammetric study of the ferricyanide/ferrocyanide system with convective electrodes of platinum, gold, glassy carbon, carbon film, and boron carbide. *J. Electroanal. Chem. Inter. Electrochem.* **1977**, *80*, 259–271.
- Armentrout, D.N.; McLean, J.D.; Long, M.W. Trace determination of phenolic compounds in water by reversed phase liquid chromatography with electrochemical detection using a carbon–polyethylene tubular anode. *Anal. Chem.* **1979**, *51*, 1039–1045.
- Koile, R.C.; Johnson, D.C. Electrochemical removal of phenolic films from a platinum anode. *Anal. Chem.* **1979**, *51*, 741–744.

28. Messner, J.L.; Engstrom, R.C. Capillary packed-bed electrodes for microanalysis. *Anal. Chem.* **1981**, *53*, 128–130.
29. Knecht, L.A.; Guthrie, E.J.; Jorgenson, J.W. On-column electrochemical detector with a single graphite fiber electrode for open-tubular liquid chromatography. *Anal. Chem.* **1984**, *56*, 479–482.
30. Andrews, R.W.; Schubert, C.; Morrison, J.; Zink, E.W.; Matson, W.R. Dual electrode cells for LCEC: recent developments. *Am. Lab. (Fairfield, CT)* **1982**, *14*, 140–151.
31. Roston, D.A.; Kissinger, P.T. Series dual-electrode detector for liquid chromatography/electrochemistry. *Anal. Chem.* **1982**, *54*, 429–434.
32. Yuan, Y.-S.; Shen, M. Study on pharmacokinetics and bioavailability of terbutaline by high-performance liquid chromatography with electrochemical detection. *J. Chin. Pharm. Sci.* **1994**, *3*, 152–156.
33. Whelpton, R.; Fernandes, K.; Wilkinson, K.A.; Goldhill, D.R. Determination of paracetamol (acetaminophen) in blood and plasma using high performance liquid chromatography with dual electrode coulometric quantification in the redox mode. *Biomed. Chromatogr.* **1993**, *7*, 90–93.
34. Zhou, J.; Wang, E. Direct detection of acetaminophen in urine by liquid chromatography with electrochemical detection using dual electrodes. Preliminary application to a single-dose pharmacokinetic study. *Anal. Chim. Acta* **1990**, *236*, 293–298.
35. Pagliari, R.; Cottet-Emard, J.M.; Peyrin, L. Determination of free and conjugated normetanephrine and metanephrine in human plasma by high-performance liquid chromatography with electrochemical detection. *J. Chromatogr.* **1991**, *563*, 23–36.
36. Lawrence, G.D.; Yatim, N. Extraction of physostigmine from biologic fluids and analysis by liquid chromatography with electrochemical detection. *J. Pharmacol. Meth.* **1990**, *24*, 137–143.
37. Morishima, K.; Tahara, R.; Horiuchi, M. Determination of a new calcium antagonist, sesamol fumarate (SD-3211), and its metabolite in plasma by liquid chromatography with electrochemical detection. *J. Chromatogr.* **1990**, *527*, 381–388.
38. Egros, F.; Moulin, A.; Le Dily, J.; Dufour, A. Routine determination of a new 1,4-dihydropyridine, oxodipine, in human plasma by high performance liquid chromatography with electrochemical detection. *J. Liq. Chromatogr.* **1990**, *13*, 1001–1011.
39. Demotes-Mainaird, F.M.; Vincon, G.A.; Jarry, C.H.; Albin, H.C. Micromethod for the determination of roxithromycin in human plasma and urine by high-performance liquid chromatography using electrochemical detection. *J. Chromatogr.* **1989**, *490*, 115–123.
40. Lunte, C.E.; Kissinger, P.T. Investigation of 6-methylpterin electrochemistry by dual-electrode liquid chromatography/electrochemistry. *Anal. Chem.* **1984**, *56*, 658–663.
41. Lunte, C.E.; Kissinger, P.T. Determination of the oxidized and reduced forms of biopterin in tissue samples. *J. Liq. Chromatogr.* **1983**, *6*, 1863–1872.
42. Lunte, C.E.; Kissinger, P.T. Determination of pterins in biological samples by liquid chromatography/electrochemistry with a dual-electrode detector. *Anal. Chem.* **1983**, *55*, 1458–1462.
43. Hou, W.; Wang, E. Determination of water-soluble vitamins by liquid chromatography with a parallel dual-electrode electrochemical detector. *Talanta* **1990**, *37*, 841–844.
44. Ding, X.D.; Krull, I.S. Dual electrode liquid chromatography–electrochemical detection (LCEC) for platinum-derived cancer chemotherapy agents. *J. Liq. Chromatogr.* **1983**, *6*, 2173–2194.
45. Goto, M.; Zou, G.; Ishii, D. Current amplification in dual electrochemical detection for micro high-performance liquid chromatography. *J. Chromatogr.* **1983**, *268*, 157–167.
46. Lunte, C.E.; Kissinger, P.T.; Shoup, R.E. Difference mode detection with thin-layer dual-electrode liquid chromatography/electrochemistry. *Anal. Chem.* **1985**, *57*, 1541–1546.
47. Caudill, W.L.; Ewing, A.G.; Jones, S.; Wightman, R.M. Liquid chromatography with rapid scanning electrochemical detection at carbon electrodes. *Anal. Chem.* **1983**, *55*, 1877–1881.
48. Samuelsson, R.; O'Dea, J.; Osteryoung, J. Rapid scan square wave voltammetric detector for high-performance liquid chromatography. *Anal. Chem.* **1980**, *52*, 2215–2216.
49. Wang, J.; Dewald, H.D. Potential scanning voltammetric detection for flow injection systems. *Anal. Chim. Acta* **1983**, *153*, 325–330.
50. Last, T.A. Coulostatic voltammetric liquid chromatography detector. *Anal. Chem.* **1983**, *55*, 1509–1512.
51. White, J.G.; St. Claire, R.L., III; Jorgenson, J.W. Scanning on-column voltammetric detector for open-tubular liquid chromatography. *Anal. Chem.* **1986**, *58*, 293–298.
52. Lunte, C.E.; Wong, S.; Chan, K.W.; Ridgway, T.H.; Heineman, W.R. Voltammetric/ampereometric detection for flow-injection systems. *Anal. Chim. Acta* **1986**, *188*, 263–267.
53. De Oliveira Neto, G.; Junior, L.R.; Kubota, L.T. Electrochemical biosensors for salicylate and its derivatives. *Electroanalysis* **1999**, *11*, 527–533.
54. Wang, J. Amperometric biosensors for clinical and therapeutic drug monitoring: a review. *J. Pharm. Biomed. Anal.* **1999**, *19*, 47–53.
55. Harwood, G.W.J.; Pouton, C.W. Amperometric enzyme biosensors for the analysis of drugs and metabolites. *Adv. Drug Deliv. Rev.* **1996**, *18*, 163–191.
56. Fuse, T.; Oda, G.; Arai, K.; Takamura, K.; Kusu, F. Electrochemical determination of lipase activity in pharmaceutical preparations. *Electrochim. Acta* **2000**, *45*, 2933–2937.
57. van der Geest, R.; Kruger, P.; Gubbens-Stibbe, J.M.; van Laar, T.; Bodde, H.E.; Danhof, M. Assay of *R*-apomorphine, *S*-apomorphine, apocodeine, isoapocodeine and their glucuronide and sulfate conjugates in plasma and urine of patients with Parkinson's disease. *J. Chromatogr. B: Biomed. Sci. Appl.* **1997**, *702*, 131–141.

58. Yamasaki, K.; Kikuoka, M.; Nishi, H.; Kokusenya, Y.; Miyamoto, T.; Matsuo, M.; Sato, T. Contents of lecithin and choline in crude drugs. *Chem. Pharm. Bull.* **1994**, *42*, 105–107.
59. Potter, P.E.; Meek, J.L.; Neff, N.H. Acetylcholine and choline in neuronal tissue measured by HPLC with electrochemical detection. *J. Neurochem.* **1983**, *41*, 188–194.
60. Neumayr, M.; Friedrich, O.; Sontag, G.; Pittner, F. Flow-injection analysis with electrochemical detection for determination of salicylic acid in pharmaceutical preparations. *Anal. Chim. Acta* **1993**, *273*, 469–475.
61. Abbiati, G.; Fossati, T.; Arrigoni, M.; Rolle, P.; Dognini, G.L.; Castiglioni, C. High-performance liquid chromatographic assay of L- α -glycerophosphorylcholine using a two-step enzymic conversion. *J. Chromatogr.* **1991**, *566*, 445–451.
62. Boppana, V.K.; Lynn, R.K.; Ziemniak, J.A. Immobilized sulfatase: β -glucuronidase enzymes for the qualitative and quantitative analysis of drug conjugates. *J. Pharm. Sci.* **1989**, *78*, 127–131.
63. Rechnitz, G.A.; Kobos, R.K.; Riechel, S.J.; Gebauer, C.R. A bio-selective membrane electrode prepared with living bacterial cells. *Anal. Chim. Acta* **1977**, *94*, 357–365.
64. Corcoran, C.A.; Kobos, R.K. Selectivity enhancement of a bacterial arginine electrode. *Anal. Lett.* **1983**, *16*, 1291–1302.
65. Corcoran, C.A.; Rechnitz, G.A. Cell-based biosensors. *Trends Biotechnol.* **1985**, *3*, 92–96.
66. Grobler, S.R.; Rechnitz, G.A. Determination of D-(+)-glucose D-(+)-mannose, D-(+)-galactose, or D-(–)-fructose in a mixture of hexoses and pentoses by use of dental plaque coupled with a glass electrode. *Talanta* **1980**, *27*, 283–285.
67. Kobos, R.K.; Pyon, H.Y. Application of microbial cells as multistep catalysts in potentiometric biosensing electrodes. *Biotechnol. Bioeng.* **1981**, *23*, 627–633.
68. Rechnitz, G.A. Bioselective membrane electrode probes. *Science* **1981**, *214*, 287–291.
69. Arnold, M.A.; Rechnitz, G.A. Tissue-based membrane electrode with high biocatalytic activity for measurement of adenosine 5'-monophosphate. *Anal. Chem.* **1981**, *53*, 1837–1842.
70. Seegopaul, P.; Rechnitz, G.A. Enzyme-amplified determination of methotrexate with a $p\text{CO}_2$ membrane electrode. *Anal. Chem.* **1984**, *56*, 852–854.
71. Meier, H.; Tran Minh, C. Determination of penicillin-V in standard solution and in fermentation broth by flow-injection analysis using fast-responding enzyme glass electrodes in different detection cells. *Anal. Chim. Acta* **1992**, *264*, 13–22.
72. Koncki, R.; Chudzik, A.; Walcerz, I. Urea determination using pH-enzyme electrode. *J. Pharm. Biomed. Anal.* **1999**, *21*, 51–57.
73. Fernandes, J.C.B.; Kubota, L.T.; Neto, G.d.O. Potentiometric biosensor for L-ascorbic acid based on ascorbate oxidase of natural source immobilized on ethylene-vinylacetate membrane. *Anal. Chim. Acta* **1999**, *385*, 3–12.
74. Caras, S.; Janata, J. Field effect transistor sensitive to penicillin. *Anal. Chem.* **1980**, *52*, 1935–1937.
75. Miyahara, Y.; Matsu, F.; Moriizumi, T.; Matsuoka, H.; Karube, I.; Suzuki, S. Micro enzyme sensors using semiconductor and enzyme-immobilization techniques. *Anal. Chem. Symp. Ser.* **1983**, *17*, 501–506.
76. Hanazato, Y.; Shiono, S. Bioelectrode using two hydrogen ion-sensitive field-effect transistors and a platinum wire pseudo reference electrode. *Anal. Chem. Symp. Ser.* **1983**, *513*–518.
77. Aboul-Enein, H.Y.; Stefan, R.-I.; Radu, G.-L. Biosensor for enantioselective analysis of S-cilazapril, S-trandolapril, and S-pentopril. *Pharm. Dev. Technol.* **1999**, *4*, 251–255.
78. Akyilmaz, E.; Dinckaya, E. A new enzyme electrode based on ascorbate oxidase immobilized in gelatin for specific determination of L-ascorbic acid. *Talanta* **1999**, *50*, 87–93.
79. Campanella, L.; Favero, G.; Persi, L.; Tomassetti, M. New biosensor for superoxide radical used to evidence molecules of biomedical and pharmaceutical interest having radical scavenging properties. *J. Pharm. Biomed. Anal.* **2000**, *23*, 69–76.
80. Campanella, L.; Favero, G.; Tomassetti, M. Direct determination of nicotine in antismoking pharmaceutical products and in tobacco using an inhibition biosensor. *Anal. Lett.* **2001**, *34*, 855–866.
81. De Carvalho, R.M.; Neto, G.d.O.; Kubota, L.T. Micro-biosensor for salicylate based on modified carbon fiber. *Anal. Lett.* **2000**, *33*, 425–442.
82. Lenigk, R.; Lam, E.; Lai, A.; Wang, H.; Han, Y.; Carlier, P.; Renneberg, R. Enzyme biosensor for studying therapeutics of Alzheimer's disease. *Biosens. Bioelectron.* **2000**, *15*, 541–547.
83. Martin, C.; Dominguez, E. A new enzyme electrode for quantification of salicylic acid in an FIA system. *J. Pharm. Biomed. Anal.* **1999**, *19*, 107–113.
84. Saurina, J.; Hernandez-Cassou, S.; Alegret, S.; Fabregas, E. Amperometric determination of lysine using a lysine oxidase biosensor based on rigid-conducting composites. *Biosens. Bioelectron.* **1999**, *14*, 211–220.
85. Eremenko, A.V.; Makower, A.; Bauer, C.G.; Kurochkin, I.N.; Scheller, F.W. A bienzyme electrode for tyrosine-containing peptides determination. *Electroanalysis* **1997**, *9*, 288–292.
86. Cunningham, D.D.; Stephens, D. Analysis of the enantiomeric purity of lactate with an enzyme electrode. *Anal. Lett.* **1992**, *25*, 1813–1821.
87. Blaedel, W.J.; Boguslaski, R.C. Chemical amplification in analysis: a review. *Anal. Chem.* **1978**, *50*, 1026–1032.
88. Halsall, H.B.; Heineman, W.R. Electrochemical immunoassay: an ultrasensitive method. *J. Int. Fed. Clin. Chem.* **1990**, *2*, 179–187.
89. Meyerhoff, M.; Rechnitz, G.A. Antibody binding measurements with hapten-selective membrane electrodes. *Science* **1977**, *195*, 494–495.
90. Boitieux, J.L.; Desmet, G.; Thomas, D. Potentiometric determination of hepatitis B surface antigen in biological fluids. *Clin. Chem. Acta* **1978**, *88*, 329–336.

91. Boitieux, J.L.; Lemay, D.; Desmet, G.; Thomas, D. Use of solid phase biochemistry for potentiometric enzyme immunoassay of 17β -estradiol—preliminary report. *Clin. Chem. Acta* **1981**, *113*, 175–182.
92. Gebauer, C.R.; Rechnitz, G.A. Ion selective electrode estimation of avidin and biotin using a lysozyme label. *Anal. Biochem.* **1980**, *103*, 280–284.
93. Meyerhoff, M.E.; Rechnitz, G.A. Electrode-based enzyme immunoassays using urease conjugates. *Anal. Biochem.* **1979**, *95*, 483–493.
94. Alexander, P.W.; Maltra, C. Enzyme-linked immunoassay of human immunoglobulin G with the fluoride ion selective electrode. *Anal. Chem.* **1982**, *54*, 68–71.
95. Gebauer, C.R.; Rechnitz, G.A. Deaminating enzyme labels for potentiometric enzyme immunoassay. *Anal. Biochem.* **1982**, *124*, 338–348.
96. Brontman, S.B.; Meyerhoff, M.E. Homogeneous enzyme-linked assays mediated by enzyme antibodies; A new approach to electrode-based immunoassays. *Anal. Chim. Acta* **1984**, *162*, 363–367.
97. Fonong, T.; Rechnitz, G.A. Homogeneous potentiometric enzyme immunoassay for human immunoglobulin G. *Anal. Chem.* **1984**, *56*, 2586–2590.
98. Keating, M.Y.; Rechnitz, G.A. Potentiometric enzyme immunoassay for digoxin using polystyrene beads. *Anal. Lett.* **1985**, *18*, 1–10.
99. Frankfurt, O.S.; Krishan, A. Enzyme-linked immunosorbent assay (ELISA) for the specific detection of apoptotic cells and its application to rapid drug screening. *J. Immunol. Meth.* **2001**, *253*, 133–144.
100. Walshe, K.; Barrett, A.M.; Kavanagh, P.V.; McNamara, S.M.; Moran, C.; Shattock, A.G. A sensitive immunoassay for flunitrazepam and metabolites. *J. Anal. Toxicol.* **2000**, *24*, 296–299.
101. Le Gal La Salle, A.; Limoges, B.; Rapicault, S.; Degrand, C.; Brossier, P. New immunoassay techniques using Nafion-modified electrodes and cationic redox labels or enzyme labels. *Anal. Chim. Acta* **1995**, *311*, 301–308.
102. Wehmeyer, K.R.; Doyle, M.J.; Wright, D.S.; Eggers, H.M.; Halsall, H.B.; Heineman, W.R. Liquid chromatography with electrochemical detection of phenol and NADH for enzyme immunoassay. *J. Liq. Chromatogr.* **1983**, *6*, 2141–2156.
103. Wright, D.S.; Halsall, H.B.; Heineman, W.R. Digoxin homogeneous enzyme immunoassay using high-performance liquid chromatographic column switching with amperometric detection. *Anal. Chem.* **1986**, *58*, 2995–2998.
104. Limoges, B.; Degrand, C.; Brossier, P.; Blankespoor, R.L. Homogeneous electrochemical immunoassay using a perfluorosulfonated ionomer-modified electrode as detector for a cationic-labeled hapten. *Anal. Chem.* **1993**, *65*, 1054–1060.
105. Jane, I.; McKinnon, A.; Flanagan, R.J. High-performance liquid chromatographic analysis of basic drugs on silica columns using non-aqueous ionic eluents II. Application of UV, fluorescence and electrochemical oxidation detection. *J. Chromatogr.* **1985**, *323*, 191–225.
106. Lunte, S.M.; Kissinger, P.T. Detection and identification of sulfhydryl conjugates of *p*-benzoquinone in microsomal incubations of benzene and phenol. *Chem. Biol. Interact.* **1983**, *47*, 195–212.
107. Radzik, D.M.; Brodbelt, J.S.; Kissinger, P.T. Determination of toxic azo dye metabolites in vitro by liquid chromatography/electrochemistry with a dual-electrode detector. *Anal. Chem.* **1984**, *56*, 2927–2931.
108. Pachla, L.A.; Kissinger, P.T. Determination of ascorbic acid in foodstuffs, pharmaceuticals, and body fluids by liquid chromatography with electrochemical detection. *Anal. Chem.* **1976**, *48*, 364–367.
109. Pachla, L.A.; Kissinger, P.T. Estimation of serum uric acid by high performance liquid chromatography with electrochemical detection. *Clin. Chim. Acta* **1975**, *59*, 309–312.
110. Lankelma, J.; VanDerKleijn, E.; Jansen, M.J.T. Determination of 5-methyltetrahydrofolic acid in plasma and spinal fluid by high-performance liquid chromatography, using on-column concentration and electrochemical detection. *J. Chromatogr.* **1980**, *182*, 35–45.
111. Lunte, C.E.; Kissinger, P.T. Study of the tautomerization of quinonoid dihydropterins by liquid chromatography/electrochemistry. *Anal. Chim. Acta* **1984**, *158*, 33–41.
112. Kamata, K.; Hagiwara, T.; Takahashi, M.; Uehara, S.; Nakayama, K.; Akiyama, K. Determination of biotin in multivitamin pharmaceutical preparations by high-performance liquid chromatography with electrochemical detection. *J. Chromatogr.* **1986**, *356*, 326–330.
113. Davis, G.C.; Holland, K.L.; Kissinger, P.T. Amperometric methods for oxidoreductase enzymes based on liquid chromatography with electrochemical detection. *Alcohol dehydrogenase. J. Liq. Chromatogr.* **1979**, *2*, 663–675.
114. Kok, W.T.; Brinkman, U.A.T.; Frei, R.W. Amperometric detection of amino acids in high-performance liquid chromatography with a copper electrode. *J. Chromatogr.* **1983**, *256*, 17–26.
115. Jacobs, W.A.; Kissinger, P.T. Nitroaromatic reagents for determination of amines and amino acids by liquid chromatography/electrochemistry. *J. Liq. Chromatogr.* **1982**, *5*, 881–895.
116. Allison, L.A.; Myer, G.S.; Shoup, R.E. The *o*-phthalaldehyde derivatives of amines for high-speed liquid chromatography/electrochemistry. *Anal. Chem.* **1984**, *56*, 1089–1096.
117. Carlsson, A.; Lundstrom, K. An amperometric detector for use with small-bore liquid chromatographic columns. *J. Chromatogr.* **1985**, *350*, 169–178.
118. Fleming, L.H.; Reynolds, J.N.C. Separation and detection of closely related endorphins by liquid chromatography–electrochemistry. *J. Chromatogr.* **1986**, *375*, 65–73.
119. Allison, L.A.; Shoup, R.E. Dual electrode liquid chromatography detector for thiols and disulfides. *Anal. Chem.* **1983**, *55*, 8–12.

120. Lunte, S.M.; Kissinger, P.T. The use of liquid chromatography with dual-electrode electrochemical detection in the investigation of glutathione oxidation during benzene metabolism. *J. Chromatogr.* **1984**, *317*, 579–588.
121. Lunte, S.M.; Kissinger, P.T. Detection of thiols and disulfides in liver samples using liquid chromatography/electrochemistry. *J. Liq. Chromatogr.* **1985**, *8*, 691–706.
122. Halbert, M.K.; Baldwin, R.P. Determination of cysteine and glutathione in plasma and blood by liquid chromatography with electrochemical detection using a chemically modified electrode containing cobalt phthalocyanine. *J. Chromatogr.* **1985**, *345*, 43–49.
123. Shankaran, D.R.; Narayanan, S.S. Amperometric sensor for thiols based on mechanically immobilised nickel hexacyanoferrate modified electrode. *Bull. Electrochem.* **2001**, *17*, 277–280.
124. Bratin, K.; Kissinger, P.T. Reductive mode thin-layer amperometric detector for liquid chromatography. *J. Liq. Chromatogr.* **1981**, *4*, 1777–1795.
125. Rappaport, S.M.; Jin, Z.L.; Xu, X.B. High-performance liquid chromatography with reductive electrochemical detection of mutagenic nitro-substituted polynuclear aromatic hydrocarbons in diesel exhausts. *J. Chromatogr.* **1982**, *240*, 145–154.
126. Hart, J.P.; Sheares, M.J.; McCarthy, P.T. Enhanced sensitivity for the determination of endogenous phyloquinone (vitamin K1) in plasma using high-performance liquid chromatography with dual-electrode electrochemical detection. *Analyst* **1985**, *110*, 1181–1184.
127. Lund, W.; Hannisdal, M.; Greibrokk, T. Evaluation of amperometric detectors for high-performance liquid chromatography: Analysis of benzodiazepines. *J. Chromatogr.* **1979**, *173*, 249–261.
128. Hanekamp, H.B.; Voogt, W.H.; Frei, R.W.; Bos, P. Continuous flow alternating current polarographic detection of nitrazepam in liquid chromatography. *Anal. Chem.* **1981**, *53*, 1362–1365.
129. Khalil, M.; Gonnet, C. Simultaneous determination of aspirin, its major metabolites, acetaminophen and phenacetin by liquid chromatography. Comparative study of electrochemical and UV detection. *Anal. Chem. Symp. Ser.* **1983**, *14*, 85–95.
130. Smotkin, E.S.; Martin, G.B.; Rechnitz, G.A. Electrochemical determination of tricyclic antidepressants. *Anal. Lett.* **1991**, *24*, 797–808.
131. Ivandini, T.A.; Sarada, B.V.; Terashima, C.; Rao, T.N.; Tryk, D.A.; Fujishima, A. Electrochemical detection of tricyclic antidepressants by HPLC using highly boron-doped diamond electrode. *Chem. Sens.* **2001**, *17*, 163–165.
132. Wilson, J.M.; Slattery, J.T.; Forte, A.J.; Nelson, S.D. Analysis of acetaminophen metabolites in urine by high-performance liquid chromatography with UV and amperometric detection. *J. Chromatogr.* **1982**, *227*, 453–462.
133. Walsh, F.X.; Langlais, P.J.; Bird, E.D. Liquid-chromatographic identification of acetaminophen in cerebrospinal fluid with use of electrochemical detection. *Clin. Chem.* **1982**, *28*, 382–383.
134. Miner, D.J.; Kissinger, P.T. Trace determination of acetaminophen in serum. *J. Pharm. Sci.* **1979**, *68*, 96–97.
135. Munson, J.W.; Weierstall, R.; Kostenbauder, H.B. Determination of acetaminophen in plasma by high-performance liquid chromatography with electrochemical detection. *J. Chromatogr.* **1978**, *145*, 328–331.
136. Riggan, R.M.; Schmidt, A.L.; Kissinger, P.T. Determination of acetaminophen in pharmaceutical preparations and body fluids by high-performance liquid chromatography with electrochemical detection. *J. Pharm. Sci.* **1975**, *64*, 680–683.
137. Meinsma, D.A.; Radzik, D.M.; Kissinger, P.T. Determination of common analgesics in serum and urine by liquid chromatography/electrochemistry. *J. Liq. Chromatogr.* **1983**, *6*, 2311–2335.
138. Wong, J.K.; Joyce, T.H., III; Morrow, D.H. Determination of hydralazine in human plasma by high-performance liquid chromatography with electrochemical detection. *J. Chromatogr.* **1987**, *385*, 261–266.
139. Shah, J.; Mason, W.D. A dose-ranging study of the pharmacokinetics of codeine phosphate following intravenous administration to rats. *J. Pharm. Sci.* **1991**, *80*, 229–231.
140. Leelavathi, D.E.; Soffer, E.F.; Dressler, D.E.; Knowles, J. Liquid chromatographic assay using electrochemical detection for the quantitation of indoramin in human plasma. *J. Pharm. Sci.* **1986**, *75*, 421–423.
141. Rotshteyn, Y.; Weingarten, B. A highly sensitive assay for the simultaneous determination of morphine, morphine-3-glucuronide, and morphine-6-glucuronide in human plasma by high-performance liquid chromatography with electrochemical and fluorescence detection. *Ther. Drug Monit.* **1996**, *18*, 179–188.
142. Takamura, K.; Kusu, F.; Abdel-Wadood, H.; El-Rabbat, N.; Saleh, G.; Refaat, I. Redox properties of isradipine and its electrochemical detection in the HPLC determination of the compound in human serum. *Biomed. Chromatogr.* **2000**, *14*, 453–458.
143. Locniskar, A.; Greenblatt, D.J. Determination of ciramadol and dezocine, two new analgesics, by high-performance liquid chromatography using electrochemical detection. *J. Chromatogr.* **1986**, *374*, 215–220.
144. Kazemifard, A.G.; Moore, D.E. Liquid chromatography with amperometric detection for the determination of nonsteroidal anti-inflammatory drugs in plasma. *J. Chromatogr.* **1990**, *533*, 125–132.
145. Lo, M.W.; Juergens, G.P.; Whitney, C.C., Jr. Determination of nalbuphine in human plasma by automated high-performance liquid chromatography with electrochemical detection. *Res. Commun. Chem. Pathol. Pharmacol.* **1984**, *43*, 159–168.
146. Hynning, P.A.; Anderson, P.; Bondesson, U.; Boreus, L.O. Determination of ketobemidone in plasma by high-performance liquid chromatography with electrochemical detection. *J. Chromatogr.* **1986**, *375*, 207–211.

147. Mokry, M. Use of high-performance liquid chromatography with electrochemical detection for the evaluation of hydrolysis of antibacterial sulfonamides. *Ceska Slov. Farm.* **1996**, *45*, 199–203.
148. Ricciarello, R.; Pichini, S.; Pacifici, R.; Altieri, I.; Pellegrini, M.; Fattorossi, A.; Zuccaro, P. Simultaneous determination of epirubicin, doxorubicin and their principal metabolites in human plasma by high-performance liquid chromatography and electrochemical detection. *J. Chromatogr. B: Biomed. Sci. Appl.* **1998**, *707*, 219–225.
149. Stinson, S.F.; Hill, K.; Siford, T.J.; Phillips, L.R.; Daw, T.W. Determination of flavopiridol (L86 8275; NSC 649890) in human plasma by reversed-phase liquid chromatography with electrochemical detection. *Cancer Chemother. Pharmacol.* **1998**, *42*, 261–265.
150. La Course, W.R.; Dasenbrock, C.O. Pulsed electrochemical detection of sulfur-containing antibiotics following high performance liquid chromatography. *J. Pharm. Biomed. Anal.* **1999**, *19*, 239–252.
151. Treskes, M.; De Jong, J.; Leeuwenkamp, O.R.; Van der Vijgh, W.J.F. Sensitive determination of cisplatin in body fluids with HPLC and on-line reductive electrochemical detection. *J. Liq. Chromatogr.* **1990**, *13*, 1321–1338.
152. Guitton, J.; Laffont, A.; Bruzeau, J.; Rochet-Mingret, L.; Bonnefoy, M.; Bureau, J. Determination of ceftazidime in plasma using high-performance liquid chromatography and electrochemical detection application for individualizing dosage regimens in elderly patients. *J. Chromatogr. B: Biomed. Sci. Appl.* **1998**, *719*, 151–157.
153. McKay, G.; Geddes, J.; Cowper, M.J.; Gurnsey, T.S. Recent advances in the analysis of phenothiazine drugs and their metabolites using high performance liquid chromatography. *Prog. Neuro-psychopharmacol. Biol. Psychiatry* **1983**, *7*, 703–707.
154. Pappa-Louisi, A.; Papageorgiou, A.; Zitrou, A.; Sotiropoulos, S.; Georgarakis, E.; Zougrou, F. Study on the electrochemical detection of the macrolide antibiotics clarithromycin and roxithromycin in reversed-phase high-performance liquid chromatography. *J. Chromatogr. B: Biomed. Sci. Appl.* **2001**, *755*, 57–64.
155. Aravagiri, M.; Marder, S.R.; Wirshing, D.; Wirshing, W.C. Plasma concentrations of risperidone and its 9-hydroxy metabolite and their relationship to dose in schizophrenic patients: simultaneous determination by a high performance liquid chromatography with electrochemical detection. *Pharmacopsychiatry* **1998**, *31*, 102–109.
156. Hansen, E.B., Jr; Dooley, K.L.; Thompson, H.C., Jr High-performance liquid chromatographic analysis of the anti-tuberculosis drugs aconiazide and isoniazid. *J. Chromatogr. B: Biomed. Appl.* **1995**, *670*, 259–266.
157. Adams, E.; Roelants, W.; De Paepe, R.; Roets, E.; Hoogmartens, J. Analysis of gentamicin by liquid chromatography with pulsed electrochemical detection. *J. Pharm. Biomed. Anal.* **1998**, *18*, 689–698.
158. Furlanut, M.; Perosa, A.; Benetello, P.; Colombo, G. Electrochemical detection of benperidol in serum for drug monitoring in humans. *Ther. Drug Monit.* **1987**, *9*, 343–346.
159. Kauffmann, J.M.; Lopez Ruiz, B.; Gotor, M.F.; Patriarche, G.J. Electrochemical behaviour of tizanidine at solid electrodes. *J. Pharm. Biomed. Anal.* **1992**, *10*, 763–767.
160. Korpi, E.R.; Kleinman, J.E.; Costakos, D.T.; Linnoila, M.; Wyatt, R.J. Reduced haloperidol in the post-mortem brains of haloperidol-treated patients. *Psych. Res.* **1984**, *11*, 259–269.
161. Tjaden, U.R.; Lankelma, J.; Poppe, H.; Muusze, R.G. Anodic coulometric detection with a glassy carbon electrode in combination with reversed-phase high-performance liquid chromatography. Determination of blood levels of perphenazine and fluphenazine. *J. Chromatogr.* **1976**, *125*, 275–286.

Electrical Power Systems for Pharmaceutical Equipment

Joseph Maida

Maida Engineering, Inc., Fort Washington, Pennsylvania, U.S.A.

INTRODUCTION

Electrical power systems that serve pharmaceutical equipment must be safe, reliable, functional, predictable, flexible, clean, and sometimes validated. The electrical power systems have voltages ranging from 120 to 69,000 V. Pharmaceutical plants in the United States and Canada can purchase 3 phase, 60 Hz electric power from utility companies that is more reliable than the power they could generate in house. Whereas in other countries, electric power is purchased at 3 phase, 50 Hz, and may not be as reliable as power that is generated in the plant.

This article is meant to convey a basic understanding of various power distribution system configurations and familiarize the reader with the electrical distribution equipment that are commonly installed within pharmaceutical plants in the United States. It will help in the evaluation of both new and existing electrical power distribution systems that serve pharmaceutical equipment.

ELECTRICAL POWER SOURCES

Electricity can be purchased from an energy provider or can be generated in the pharmaceutical plant. Usually, purchased electricity is cheaper than on-site generated electricity. Large plants have many options when purchasing electricity. A thorough understanding of these options can significantly reduce the electrical operating cost and will increase the reliability of the electrical power system. All plants and especially small- and medium-size plants must accurately determine their current and future normal electrical power loads and normal/emergency power loads when deciding how to purchase electricity and how to install a reliable electrical power system.

The pharmaceutical plant's electrical loads can be divided into four categories:

1. Normal loads—Loads that can be turned off for a period of time without creating a hazardous condition or causing a substantial loss of product or research.
2. Standby loads—Loads that if they lose electricity will cause a hazardous condition or cause a substantial loss of product or research but which can sustain a power interruption of 60 sec.

3. Standby noninterruptible load—Loads that cannot sustain any interruption in power.
4. Emergency loads—Legally required emergency and egress lighting and other loads classified as such by governmental agencies and locally adopted building codes.

Loads can be further divided into groups based on their utilization voltage, category, type, and location within the plant. Typical utilization voltages within pharmaceutical plants include 480/277 and 208/120 V. Categories include linear and nonlinear loads. Linear loads typically include incandescent lamps and induction motors that are not controlled by a variable frequency drive (VFD). Nonlinear loads include all loads that have switching power supplies or silicon control rectifiers such as VFDs, uninterruptible power supplies, and electronic ballasts on 480/277 V systems and personal computers, copy machines, faxes, electronic devices, and instruments on 208/120 V systems. Types include continuous loads such as lighting that stay on all day and noncontinuous loads that will not operate for more than 3 hr at a time.^[1] The areas in the plant should be divided into locations based on the function or process performed in the area and based on whether or not the area being classified is a current good manufacturing practices (cGMP) area.

The load projections should be based on a minimum of 7 years. To obtain the group demand for each group of loads, multiply the connected load for the group by a diversity factor that reflects the proportion of load that will operate at any one time to the total connected load. The sum of the group demands will equal the plant's maximum demand. After determining the maximum demand, consult with the local public utility company for assistance in determining the optimum voltage and configuration for the electrical service entrance for the plant.

The ampacity or rated current carrying capacity of the electrical service entrance conductors that connect the utility company's lines to the plant's service entrance equipment must be a minimum of 125% of the calculated maximum demand for continuous loads plus 100% of the maximum calculated demand for noncontinuous loads. Service entrance conductors and equipment with higher ratings or provisions to increase the rating of the service entrance conductors are recommended.



Large plants, which have a central utility plant (CUP) with chiller motors rated above 200 hp, should consider purchasing and in some instances distributing electricity at primary voltage levels such as 4160 V and 13,800 V.

An emergency generator should have a rating that is 25% and possibly 100% higher than the sum of the total calculated standby and emergency demand for the plant. Alternately, provisions for a future emergency generator should be provided as part of the initial installation. Discuss the type of fuel required for the generator's engine with the local authority having jurisdiction. The National Electrical Code, NFPA 70 (NEC), requires that the fuel be on site if the generator provides power for emergency loads. However, exceptions to this requirement have been permitted where reliable natural gas service is available.

ELECTRICAL POWER SYSTEM UTILITY SERVICES

Primary services are typically installed for medium and large pharmaceutical plants and can be purchased as a single service, a dual service, or a regular-reserve service (Fig. 1). The utility company will install two (2) separate power lines to the plant for a dual service and for a regular-reserve service. Each dual service power line will normally serve half of the loads in the plant but each will be capable of serving all of the loads in the plant. Regular-reserve service power lines are connected to a common

bus with only the regular power line normally serving all of the plant's loads.^[2]

Secondary services are typically installed for small- to medium-size pharmaceutical plants. They are available at the utilization voltages of 277/480 or 120/208 V. Secondary services are less expensive to install and maintain than primary services. However, the kilowatt-hour cost for electrical power supplied by a secondary service is higher.

SECONDARY CONFIGURATIONS

The following items should be considered when planning an electrical power system's secondary, less than 1000 V, configuration.

1. The utility company will provide and own the transformer that will supply secondary voltage for the service to the plant. This transformer should be dedicated to the pharmaceutical plant and should have a grounded wye secondary. Sharing a utility-owned transformer with other utility customers will reduce the quality and the reliability of the electric service to the plant.
2. The size of the utility company's transformer will affect the rating of the service entrance equipment and the starting of large motors. If the transformer rating is low, reduced voltage starting will be required for starting large motors. If the transformer rating is high, service entrance equipment will require a high symmetrical short-circuit (IAC) rating. The utility

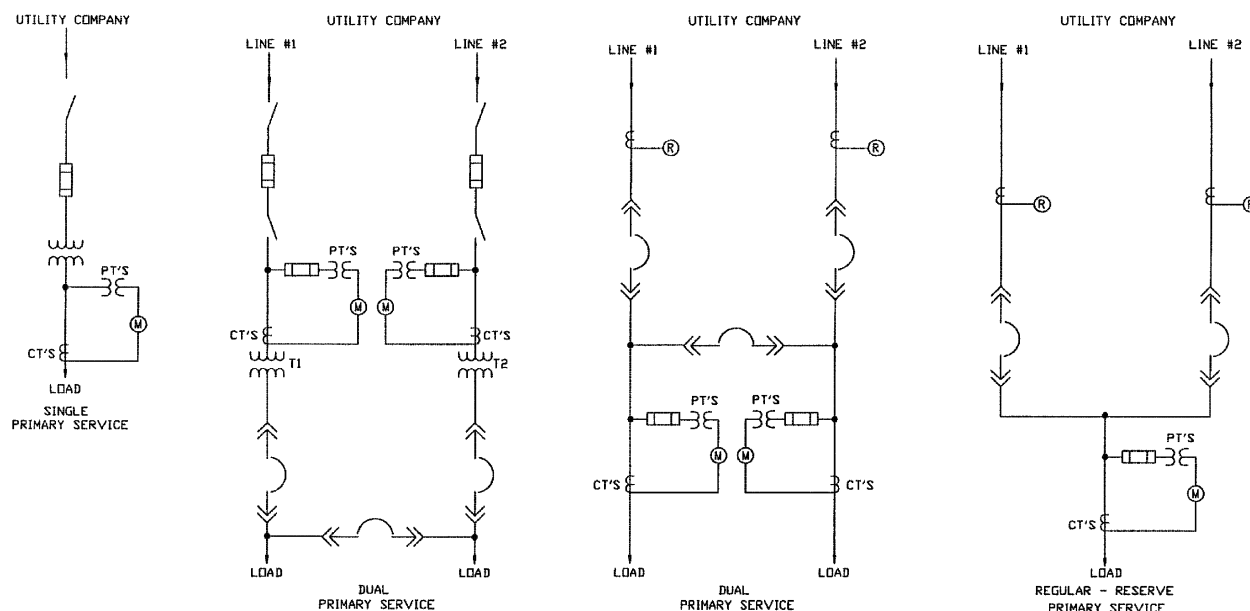


Fig. 1 Typical primary electric service arrangements.

company will provide the maximum available symmetrical short-circuit current available at the secondary of their transformer.

3. Service entrance equipment must have a current rating equal to or greater than the present and future service entrance conductors' ampacity. Moreover, this equipment must have an IAC current rating that is higher than the available short circuit current at the utility company's transformer. Spare service entrance conduits should be installed for future service entrance conductors. The service entrance equipment should include spaces for future feeder circuit breakers to supply future loads. All incoming service entrance conductors must be 3 phase, 4 wire and should be installed in underground metal conduit.
4. Service entrance equipment should be installed in an indoor electrical room whenever possible. The equipment should include one to six main service entrance power circuit breakers or load break disconnect switches and feeder circuit breakers for loads within the plant. Circuit breakers and disconnect switches for 480/277 V systems that are rated at 1000 A or more must include ground fault protection. Feeder circuit breakers will typically be molded case. Integrally fused circuit breakers are available with very high short circuit IAC ratings for both power and molded case circuit breakers. Fuses and current limiters for integrally fused circuit breakers must be stored within the electrical room.
5. Fuses should not be used to protect secondary voltage feeders. The time current characteristics of fuses above 100 A will not coordinate with the ground fault pickup currents and time delays of the main overcurrent protection (circuit breaker or fused disconnect switch) ground fault protection. A main load break disconnect switch can be equipped with current-limiting fuses to reduce the available short-circuit current from the utility and should have a three-phase voltage relay for single-phase protection.
6. Ground fault protection, including circuit breakers or relays, should be set as low as possible without causing nuisance tripping. The ground fault pickup and time delay for the main overcurrent protection should coordinate with the trip characteristics of the largest feeder circuit breaker that does not have ground fault protection and must be less than the maximum current permitted by the NEC.
7. If fuses are used in the power distribution system, three-phase motor starters should have solid-state overload devices that include single-phase protection.
8. Two (2) to six (6) separate service entrance cables connected to the secondary side of one utility owned transformer can connect to separate main circuit breakers or main load break fused disconnect switches that are mounted next to each other. The sum of the ampacity of all of the service entrance conductors must have a current rating equal to or greater than the present and future maximum demand for the plant. This approach can reduce the cost of the service entrance installation but will not provide the additional level of protection provided by having one main circuit breaker or one main fused disconnect switch. Therefore it is not as reliable.
9. A maintenance program should document all tests, inspections, and faults cleared by a main and feeder circuit breaker. Lack of such a program will reduce the reliability of the electrical equipment.^[3] Molded case circuit breakers require no internal maintenance. They should be inspected for broken casing and loose connections after they operate to clear a fault and should be replaced after they have operated to clear two faults. Feeder circuit breakers, especially older circuit breakers, should be exercised (repeatedly opened and closed) whenever possible.^[4] Part of the initial commissioning of an electrical power system must include testing of the ground fault protection.^[1]
10. The secondary selective circuit arrangement is recommended for all critical loads (Fig. 2). It requires the installation of double-ended switchgear or a variation thereof. Double-ended switchgear includes two main circuit breakers that serve separate feeder buses with a tie circuit breaker in the middle. Each feeder bus has its own set of feeder or branch circuit breakers. The load should be evenly split between each of the feeder buses.
11. Important loads can be connected to redundant feeder circuit breakers, one on each side of the double-ended switchgear, by connecting the load sides of the circuit breakers together with a feeder cable. This approach should be used with caution. Another approach is to connect the redundant feeder circuit breakers to a time-delayed automatic transfer switch.

PRIMARY CONFIGURATIONS

The following items should be considered when planning an electrical power distribution system's primary configuration.

1. A single primary service can have one to six primary service entrance circuit breakers or fused disconnect switches. These overcurrent devices will protect primary feeders that distribute power to distribution unit substations located in electrical rooms throughout the plant, to primary voltage motor starters in the

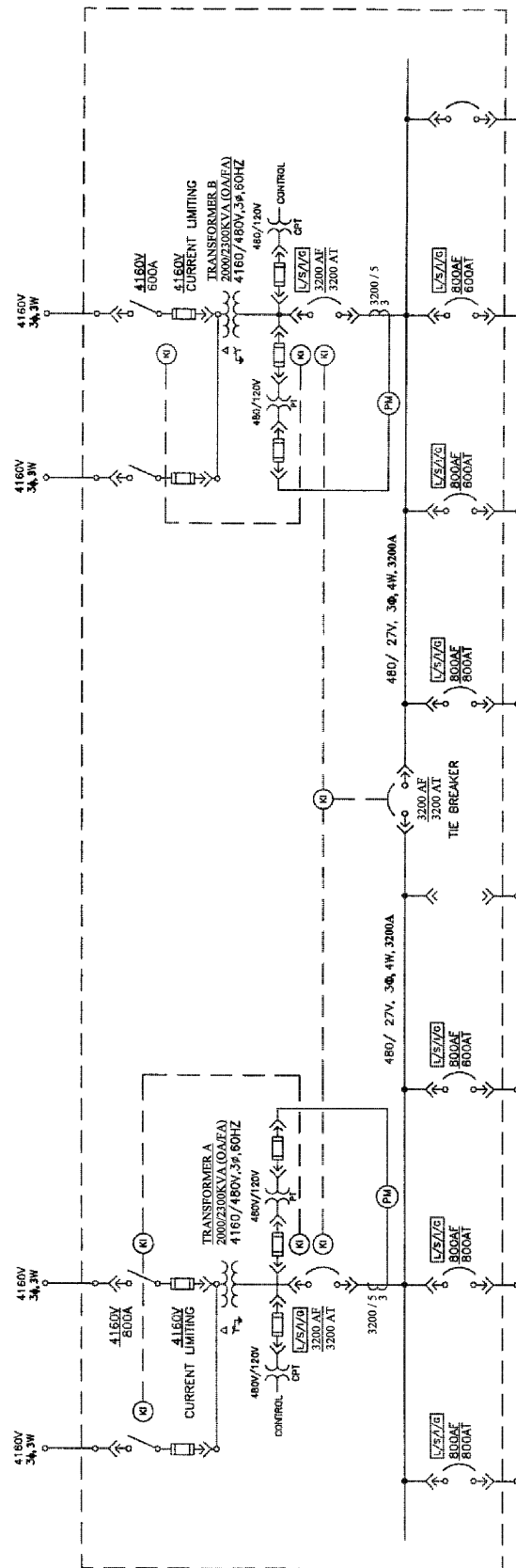


Fig. 2 Secondary selective circuit arrangement.

- CUP, and sometimes to generator switchgear when the generator system is designed for momentary or continuous operation in parallel with the utility line.
2. A dual service will have two main service entrance circuit breakers and a service entrance tie circuit breaker and may have additional primary feeder circuit breakers. A regular-reserve service will have two main service entrance circuit breakers and may have additional primary feeder circuit breakers (Fig. 1). In very large pharmaceutical plants, the primary feeder circuit breakers will connect to outdoor substation transformers that will provide 4160 V for the CUP and 13,800 V for distribution unit substations.
 3. Unit substations, each consisting of a primary fused disconnect switch or circuit breaker, a dry transformer, and secondary switchgear should be in a central location near their loads. Unit substation transformers should be sized to carry at least 125% the maximum present and future connected load.
 4. Double-ended unit substations provide economical, reliable, preengineered, and factory-tested configuration for obtaining secondary selective circuit arrangements at utilization voltage of 480/277 or 120/208 V and for installing a primary selective circuit arrangements (Fig. 2). Double-ended unit substations include one or two main primary circuit breakers or fused disconnect switches on both sides that are connected to transformers that are connected to the feeder circuit breaker sections. Each feeder circuit breaker section has a main secondary circuit breaker that protects the internal bus, which connects to the feeder circuit breaker, and is connected to the other feeder circuit breaker section through a secondary tie circuit breaker. Each transformer should be rated to carry at least 125% of the maximum present and future connected load for the entire double-ended substation.
 5. Primary selective circuit arrangements utilize two primary selectable feeders for each transformer. These feeders are either connected to the line side of two interlocked primary nonfused disconnect switches that have a common load side fuse or to the line side of two interlocked primary fused disconnect switches. The latter configuration is shown in Fig. 2. Only one primary disconnect switch can be closed at any time.
 6. Dry-type transformers are typically installed indoors and include ventilated, sealed, or gas filled, totally enclosed nonventilated, and cast coil types. Cast coil transformers are the most expensive and they have the highest overload capability. Transformers should be sized based on their air to air (AA) rating. Provisions for future automatic fans that will provide a higher force air (FA) rating and an alarm that indicates when the fans are operating due to the temperature within the transformer should be provided.
 7. If installed outdoors, pad mounted liquid filled transformers up to 2500 KVA are available with mineral oil or less flammable silicon liquids. Outdoor liquid-filled transformers sometimes have an advantage over outdoor dry-type transformers, which present a relatively difficult lightning protection problem because of their lower basic impulse levels (insulation level).^[5] Substation transformers with ratings above 2500 KVA are utilized for higher secondary voltages, including 4160 V for large motors and 13,800 V for medium voltage feeders.
 8. All configurations described herein are based on radial feeders. Radial feeders provide power to feeders for panelboards and other electrical distribution equipment within the pharmaceutical plant. Branch circuits connect to feeders and provide power to the electrical loads.

FEEDERS AND BRANCH CIRCUIT ARRANGEMENTS

Feeders originating in switchgear, panelboards, and motor control centers provide power to electrical distribution equipment within the pharmaceutical plant. Branch circuits originating in panelboards and motor control centers provide power to pharmaceutical equipment, lighting, motors, and other equipment.

1. Three phase, 3 wire or 3 phase, 4 wire power and lighting panelboards having current ratings from 100 to 1600 A may have a main circuit breaker or may have main lugs only. They will have feeder circuit breakers for other electrical equipment and/or branch circuit breakers for pharmaceutical equipment, lighting fixtures, etc. The panelboards should have an IAC rating greater than the available short-circuit current calculated at the transformer serving the panelboard. Four-wire switchgear, panelboards, and motor control centers should have a separate neutral bus and a separate ground bus. The secondary neutral of a separately derived 4 wire system should be bonded to ground at one point ahead of the main secondary circuit breaker, ideally in the transformer and not in the panelboard. Panelboard schedules and nameplates must be installed and maintained for reliability and to comply with the NEC.
2. Separately derived systems that originate at the secondary of two winding transformers including single-phase and three-phase delta-wye transformers and at the output of a wye connected generators should be connected to the power distribution system with a 4 pole automatic transfer switch (ATS) when the neutral conductor is installed in the feeder

to the ATS. This will avoid circulating ground currents that can trip ground fault relays and effect sensitive electronic equipment.

3. Secondary unit substations (120/208 V) and transformers need to be distributed throughout the pharmaceutical plant to provide 120 and 208 V branch circuits to pharmaceutical equipment. Because of their lower voltage, the length of these branch circuits should be short.
4. AC motors 1 hp and larger should be 3 phase, 460 V.^[6] Smaller motors can be 120 V, single phase. If the plant purchases power at 208/120 V, 3 phase or 240 V, 3 phase, all motors 1 hp and larger should be 200 or 230 V, 3 phase motors.
5. Motor control centers (MCCs) are typically 480 V, 3 phase, 3 wire or 277/480 V, 3 phase, 4 wire with current ratings from 600 to 2500 A. Motor control centers should be a type that have been designed for industrial applications rather than commercial applications and must have IAC ratings greater than the available short-circuit current calculated at the transformers serving the MCCs. New MCCs that are designed to incorporate intelligent starters, protective equipment, network communication, and 24 V control circuits should be considered for future pharmaceutical process applications.
6. Individual motor controllers are sometimes connected to the main switchgear when they serve large motors but are more often connected to motor branch circuits originating in the distribution panelboards or MCCs. Individual motor controllers can be full voltage nonreversing or full voltage reversing, reduced voltage, and solid state. Reduced voltage and solid state motor controllers are used when the load requires a soft start or when there is insufficient let through (short circuit) current to start the motor without causing a severe voltage drop. When controlling a piece of critical pharmaceutical equipment within a cGMP area or process, the operation of the motor controller and all of its control circuits should be part of the qualification and validation process.
7. Variable frequency drives are special types of motor controllers. They are almost always used with 480 V motors and sometimes for motors greater than 200 hp because they require less line current to deliver the same amount of torque as compared to reduced voltage starters. They are complex and require a clean, cool, ventilated environment to properly operate. Variable frequency drives vary the speed of motors, which if continuously run at a slow speed may need supplemental cooling. Motors, designed to be controlled by VFDs, should be specified. Variable frequency drives generate harmonics on their input (line side) and output (load side) motor branch circuits. Capacitors cannot be installed on the load side of VFDs. Capacitors installed on the line side of VFDs at the switchgear or MCC must have a KVAR rating that will not cause a resonant circuit at any of the harmonic current frequencies generated by the VFDs. Variable frequency drives require many parameter settings, most of which can remain at the factory default settings and all of which should be recorded. When controlling a piece of critical pharmaceutical equipment within a cGMP area or process, the operation of the VFD, its parameter settings, and all of its control circuits should be part of the qualification and validation process.
8. Custom-built control cabinets are provided for various types of pharmaceutical equipment, usually by the equipment manufacturer. These control cabinets should be UL listed or labeled, otherwise all of the wire and electrical equipment in them are subject to the requirements of the NEC. The functions controlled by these control cabinets and sometimes the equipment within the control cabinets and all of its control circuits should be part of the qualification and validation process.
9. Uninterruptible power supplies (UPSs) must be installed for all standby noninterruptible loads. Smaller UPSs that are dedicated to individual pieces of pharmaceutical equipment are usually better than a single large UPS. Because of the amount of nonlinear loads within pharmaceutical distribution systems, UPSs are the only source of clean electricity that does not contain harmonics generated by other nonlinear loads. Uninterruptible power supplies usually have parameter settings and alarms. When providing power to a piece of critical pharmaceutical equipment within a cGMP area or process, the operation of the UPS, its parameter settings, and alarms should be part of the qualification and validation process for the piece of equipment.
10. Automatic transfer switches (ATSs) are commonly used to transfer emergency and standby loads from a normal source of power to an emergency source of power. Automatic transfer switches are available at different voltages, with three or four poles, with contacts to start an emergency generator, with time delays when part of a priority load shedding circuit, with controls that will permit bumpless transfer between the normal and emergency power system, and with integral battery chargers for the emergency generator starting batteries. They can be simple or complex and are critical parts of a reliable power distribution system. Automatic transfer switches are

a latent single point of failure within normal/emergency circuits and therefore must be inspected and tested as often as possible.

GENERATORS AND EMERGENCY POWER SOURCES

A pharmaceutical plant can obtain emergency or standby electrical power from generators, rechargeable batteries, and in rare cases from a separate utility service. Uninterruptible power supplies, central storage battery system, and unit equipment all use rechargeable batteries for their emergency source of power.

1. Unit equipment (self-contained battery packs with integral or remote lamps) provide the lowest initial cost for emergency egress lighting and exit signs but require considerable maintenance.
2. Emergency generators can provide emergency power for egress lighting and exit signs and standby loads. The 1996 NEC introduced a requirement that all legally required emergency loads must be supplied by a dedicated ATS with dedicated feeders and branch circuits. Therefore installations after 1996 must have two or more ATSs when serving both emergency and standby loads. Emergency generators typically have a standby rating and sometimes have a lower prime rating. The standby rating should be used unless the generator is intended to run for weeks at a time. Standby generators can deliver 100% of their rated KW output for short periods of time to linear loads. Non-linear loads will reduce the rating of the emergency generator. Outdoor emergency generators can be purchased with subbase double-wall fuel tanks and can be installed in walk-in enclosures in sizes up to 1600 KW.

When the entire pharmaceutical plant or a large portion thereof requires emergency power, two or more generators are recommended. These generators should have a priority load shedding circuit that turns off the noncritical loads before critical and legally required loads. Emergency generators that can be synchronized with each other can also synchronize with one of the utility lines if approved and designed in accordance with the requirements of the utility company.

HARMONICS

Nonlinear loads have a distorted periodic wave form. Any periodic wave shape can be broken into or analyzed as a

fundamental wave and a set of harmonics. Fundamental wave, 60 Hz, and harmonics currents exist within the branch circuits and feeders that serve nonlinear equipment. The highest magnitudes of currents exist at the 3rd, 5th, 7th, 9th, 11th, and 13th harmonics.

1. The triplet harmonics, including the 3rd, 9th, and 15th, exists only in circuits that have a neutral conductor such as computer circuits and lighting circuits. The triplet harmonics that exist on the secondary side of a Δ -Y transformer will not be transformed to the primary side of the transformer. Transformers with K ratings of 4-50 are specifically designed to handle these harmonic currents.
2. Variable frequency drives generate harmonic other than the triplets, the highest of which are the 5th, 7th, 11th, and 13th harmonic currents, on the three-phase circuits that feed them. Drive transformers have been specifically designed for VFD loads.

A harmonic analysis should be performed before installing capacitors on a power distribution system that has a significant nonlinear load.^[5]

GROUNDING

Grounding for any electrical distribution system can be divided into two areas: system grounding and equipment grounding. These two areas are kept separate from each other except at the point where the system receives its source of power, namely the service equipment or a separately derived system (Δ -Y transformer or generator). Grounding the neutral conductor at more than one point will cause circulating ground currents that could trip ground fault relays and circuit breakers with ground fault detection and can cause noise in the electronic equipment whose enclosures are required to be solidly grounded to the electrical systems equipment ground.^[7]

CONCLUSION

Pharmaceutical plants will become more dependent on reliable electrical power because of technological advances in pharmaceutical equipment and other loads associated with cGMP areas within plant. A reliable electrical power system is not something that just happens. It requires adequate and knowledgeable planning, engineering, design, installation, testing, and maintenance. A failure to provide any one of these will lessen the reliability of the electrical power system and could



leave the plant without electricity. A pharmaceutical plant without electricity is very dark, very quiet, and very unsafe.

REFERENCES

1. *NFPA70, National Electrical Code*; National Fire Protection Association: Quincy, MA. 2002 Ed.
2. Helwig, D.R. *PECO Energy Services Requirements*; Power Delivery Consumer Energy Services Group: Philadelphia, PA, Revised April 1996.
3. *IEEE Standard 493-1997, IEEE Recommended Practice for Design of Reliable Industrial and Commercial Power Systems*; The Institute for Electrical and Electronic Engineers, Inc.: New York, NY, 1st Printing, August 1998.
4. *IEEE Standard 1015-1997, IEEE Recommended Practice for Applying Low Voltage Circuit Breakers Used in Industrial and Commercial Power Systems*; The Institute for Electrical and Electronic Engineers, Inc.: New York, NY. 1st Printing, October 1997.
5. *IEEE Standard 141-1997, IEEE Recommended Practice for Electrical Power Distribution for Industrial Plants*; The Institute for Electrical and Electronic Engineers, Inc.: New York, NY. 2nd Printing, May 1996.
6. *IEEE Standard 241-1990, IEEE Recommended Practice for Electrical Power Systems in Commercial Buildings*; The Institute for Electrical and Electronic Engineers, Inc.: New York, NY. 3rd Printing, January 1998.
7. *IEEE Standard 446-1995, IEEE Recommended Practice for Emergency and Standby Power Systems for Industrial and Commercial Applications*; The Institute for Electrical and Electronic Engineers, Inc.: New York, NY. 1st Printing, July 1996.



ELASTOMERIC COMPONENTS FOR THE PHARMACEUTICAL INDUSTRY

Edward J. Smith

Packaging Science Resources, King of Prussia, Pennsylvania, U.S.A.

INTRODUCTION

The primary function of elastomeric components used by the pharmaceutical industry, which includes both drugs and medical devices, is to protect and deliver. Elastomeric components are typically primary packaging components; that is, they are or may be in direct contact with the dosage form. They must neither interact with the dosage form nor allow the ingress or egress of materials. Elastomeric or rubber components typically provide the means for sealing parenteral containers of varying size and shape because of their unique physical properties. Elasticity particularly permits intimate contact between the closure and the relatively rigid surfaces of container openings. No other material known today has this same unique property. This property is primarily responsible for the ability of the closure to be pierced with a sharp device, such as a hypodermic needle, and then to reseal. This is an example of the delivery function of elastomeric components. Rubber closures are an essential component of the primary package for most parenteral or injectable products. Typical rubber components are pictured in Figs. 1 and 2.

In order to understand how rubber performs its unique “protect and deliver” function, it is necessary to know how rubber is compounded and manufactured.

RUBBER COMPOUNDS

Rubber, like all other primary packaging materials, is not inert. Any material used in the compounding of the rubber component may be leached into and/or chemically react with the dosage form. The basic materials used in the compounding of rubber components used by the drug and medical device industries are listed in Table 1. These materials, and the amounts used, are significantly different from those that may be used in industrial rubber belts, tires, and hoses where heat, abrasion, and solvent resistance are typically the main concern. More than one type of each material may be used in a rubber compound. For example, compounds containing two elastomers, such

as isoprene and chlorobutyl rubber, and two pigments, such as titanium dioxide and carbon black, are very common. Compounds used for parenteral closures consist of an elastomer as the base material combined chemically and physically with other necessary rubber chemicals.

The elastomer determines most of the physical and chemical characteristics of a rubber compound. Typical elastomers are natural elastomers such as natural rubber (NR), sometimes called crepe, and synthetic elastomers such as butyl (including chlorobutyl and bromobutyl), ethylene propylene diene monomer (EPDM), and styrene butadiene rubber (SBR). A list of commonly used elastomers is shown in Table 2.

In the pharmaceutical industry, natural rubber and its synthetic analog, isoprene rubber, are normally used in products that require good physical strength or that must be able to withstand multiple punctures while maintaining seal integrity. Due to the “latex sensitivity” issue, the use of natural rubber is declining. This issue is discussed in a later section. Neoprene, a halogenated form of polyisoprene, is typically used for oil-based pharmaceutical products, such as mineral oil or vegetable oil. SBR and nitrile rubber (NBR), which is used primarily for oil-based products, are specialty elastomers that are not as commonly used. Butyls and halobutyls comprise the largest segment of pharmaceutical rubber compounds, having properties that make them applicable for the packaging of many products, especially those requiring protection from moisture vapor or oxygen. Most lyophilized and powdered products require this protection, and some liquid products require protection from oxygen. EPDMs are less commonly used, though they have been selected for large intravenous (IV) stoppers. Silicones also are not often used due to their permeability to moisture vapor and oxygen as well as their relatively high cost. A list of common elastomers and their chemical structures is found in Table 3.

Curing (or vulcanizing/cross-linking) agents are chemicals used to cross-link elastomer chains into the three-dimensional (3D) network required to give a rubber component the desired elasticity. The term “vulcanization” indicates that heat is employed in the manufacturing or molding process. Common curing agents are sulfur,



Fig. 1 Typical rubber components. A) Syringe plunger. B,C,D) Sleeve stoppers. E) Abbott ADDVantage vial stopper. F) IV bag injection site. G) Serum vial stopper. H) Lyophilization vial stopper. (Photograph provided by Abbott Laboratories.)

thiurams, zinc oxide, peroxides, resins, and amines. A desirable property of pharmaceutical rubber formulations is “cleanliness,” that is, that they contain materials that neither leach nor volatilize into the packaged pharmaceutical. Sulfur-cured rubber, because it requires other chemicals to effect an efficient cure, is not usually as clean as resin-, metal oxide-, or peroxide-cured formulations. The demands of the pharmaceutical industry and regulatory agencies are such that these relatively clean cure systems are becoming more common.

Accelerators reduce the cure time considerably by increasing the cure rate. They are not catalysts because they are chemically altered and, in many cases, also react as curing agents. Common sulfur-cure accelerators are amines, dithiocarbamates, sulfenamides, thiazoles, and thiurams. Some accelerators, because of their reactivity, may form toxic compounds, such as 2-(2-hydroxyethylmercapto)benzothiazole from mercaptobenzothiazole (2-MCBT), residues of which may be extractable. Accelerators that are secondary amines may form toxic nitrosamines.

Activators, which affect the efficiency of accelerators, are commonly added. Normally these are metal oxides, such as zinc oxide or stearic acid.



Fig. 2 Typical rubber syringe plungers. A) Plunger for 10-mL sterile-empty syringe. B) Plunger for sterile-prefilled dental cartridge. C) Plunger for 1mL sterile-empty syringe. D) Tip cap for Luer Tip syringe. (Photograph provided by West Pharmaceutical Services, Inc.)

Antioxidants are classified as antidegradants or age resistors. Chemically, antioxidants protect the reactive (sensitive) sites of the rubber chains against oxygen attack. Typical antioxidants are chemicals such as hindered phenols and amines. Unsaturated elastomers, such as natural rubber, require antioxidants for protection against oxidation, which causes surface cracking and loss of elasticity. Saturated elastomers, such as silicones and fluoroelastomers, are resistant to oxidation and usually require no added antioxidants. Some antioxidants are classified as antiozonants, which are designed to provide protection when high levels of reactive ozone are likely to be in the environment. A list of saturated and unsaturated elastomers is found in Table 4.

Table 1 Rubber compounding materials and their function

Material	Function
Elastomer	Base material
Curing agent	Forms cross-links
Accelerator	Affects type and rate of cross-links
Activator	Alters efficiency of accelerator
Antioxidant	Antidegradant
Plasticizer	Processing aid
Filler	Affects physical properties
Pigment	Color

Table 2 Common elastomers used in parenteral packaging components

Elastomer	% Use	Reason
Butyl/halobutyls	~80	Excellent O ₂ and moisture barrier
Natural/isoprene	~10	Excellent physical properties such as reseal
EPDM	~5	Good heat resistance and surface lubricity
Nitrile and neoprene	~3	Resistance to vegetable and mineral oils
SBR	~1	Blended with isoprene/NR or halobutyls to improve physical properties
Silicone	≤1	Excellent heat resistance; poor O ₂ and moisture barrier; high cost

Table 3 Chemical structures of common elastomers

Common name	Chemical name	Structure
Butyl rubber	Poly(isobutylene-isoprene)	$\left[\text{CH}_2 - \underset{\text{CH}_3}{\overset{\text{CH}_3}{\text{C}}} \right]_{50} \left[\text{CH}_2 - \underset{\text{CH}_3}{\text{C}} = \text{CH} - \text{CH}_2 \right]_n$
Halobutyl rubber ^a	Halogenated poly(isobutylene-isoprene)	$\left[\text{CH}_2 - \underset{\text{CH}_3}{\overset{\text{CH}_3}{\text{C}}} \right]_{65} \left[\text{CH} = \underset{\text{CH}_3}{\text{C}} - \overset{\text{X}}{\text{CH}} - \text{CH}_2 \right]_n$
Ethylene-propylene rubber	Poly(ethylene-propylene)	$\left[\text{CH}_2 - \text{CH}_2 \right]_3 \left[\text{CH}_2 - \underset{\text{CH}_3}{\text{CH}} \right]_n$
Ethylene-propylene-diene rubber	Poly(ethylene-propylene-diene)	$\left[\text{CH}_2 - \text{CH}_2 \right]_{15} \left[\text{CH}_2 - \underset{\text{CH}_3}{\text{CH}} \right]_5 \left[\text{diene} \right]_n$
Silicone rubber	Polydimethylsiloxane	$\left[\underset{\text{CH}_3}{\overset{\text{CH}_3}{\text{Si}}} - \text{O} \right]_n$
Urethane rubber	Adipic acid-ethylene glycol polyester	$\text{HO} - (\text{CH}_2)_2 - \text{O} - \overset{\text{O}}{\parallel} \text{C} - (\text{CH}_2)_4 - \overset{\text{O}}{\parallel} \text{C} - \text{O} - (\text{CH}_2)_2 - \text{OH}$
Fluoroelastomers	Polytetrafluorethylene	$\left[\underset{\text{F}}{\overset{\text{F}}{\text{C}}} - \underset{\text{F}}{\overset{\text{F}}{\text{C}}} \right]_n$
Natural rubber	<i>cis</i> -(1,4-Polyisoprene)	$\left[\text{CH}_2 - \underset{\text{CH}_3}{\text{C}} = \text{CH} - \text{CH}_2 \right]_n$
Polyisoprene rubber	<i>cis</i> -(1,4-Polyisoprene)	$\left[\text{CH}_2 - \underset{\text{CH}_3}{\text{C}} = \text{CH} - \text{CH}_2 \right]_n$
Neoprene rubber	Polychloroprene	$\left[\text{CH}_2 - \underset{\text{Cl}}{\text{C}} = \text{CH} - \text{CH}_2 \right]_n$
Styrene-butadiene rubber	Poly(butadiene-styrene)	$\left[\text{CH}_2 - \text{CH} = \text{CH} - \text{CH}_2 \right]_4 \left[\text{CH}_2 - \underset{\text{C}_6\text{H}_5}{\text{CH}} \right]_n$
Nitrile rubber	Poly(butadiene-acrylonitrile)	$\left[\text{CH}_2 - \text{CH} = \text{CH} - \text{CH}_2 \right]_5 \left[\text{CH}_2 - \underset{\text{CN}}{\text{CH}} \right]_2$
Polybutadiene	Polybutadiene	$\left[\text{CH}_2 - \text{CH} = \text{CH} - \text{CH}_2 \right]_n$

(From Ref. 22.)

^aX = Cl or Br.

Table 4 Saturated and unsaturated elastomers

Saturated	ASTM abbreviation	Unsaturated	ASTM abbreviation
Butyl	IIR	Natural	NR
Halobutyls (chlorobutyl and bromobutyl)	CIIR, BIIR	Isoprene	IR
Ethylene-propylene-diene monomer rubber	EPDM	Styrene butadiene	SBR
Silicone	Q	Nitrile	NBR
Urethane	U	Neoprene	CR
Fluoroelastomers	FKM	Polybutadiene	BR

Plasticizers are used in rubber compounds to assist in the mixing or molding of the rubber, to soften the final vulcanized rubber, or to add surface lubricity to the surface of the rubber component. Examples are paraffinic wax, silicone oil, paraffinic and naphthenic oils, phthalates, and organic phosphates. Silicone oil is commonly used in syringe pistons that must slide freely within a glass or plastic barrel; it also reduces the coring or fragmentation tendency of vial stoppers.

Fillers are materials that modify rubber characteristics (e.g., hardness) and improve its physical characteristics (e.g., tensile strength), in addition to reducing costs. Rubber is sometimes compounded without the use of fillers; the resultant product is called “gum rubber.” Typical fillers are calcined and hydrated clays, magnesium silicate (talc), magnesium oxide, and silicas. Carbon black, a common filler used to increase the heat resistance in industrial components such as tires, is not used as a filler in pharmaceutical components but it is used in smaller amounts as a black pigment. Polynuclear aromatic (PNA) hydrocarbons are a concern with carbon blacks but the grades used by manufacturers of pharmaceutical components contain very low concentrations.

The pigments used are inorganic salts and oxides, carbon black, or organic dyes that are used for aesthetic or functional purposes (e.g., identification, designating a dosage, etc.). Typical pigments are carbon black, titanium

dioxide, and iron oxide. With these three pigments white, black, red, and many shades of gray and pink can be produced. These pigments are chemically pure and stable, nontoxic, and relatively inexpensive. Other pigments such as phthalocyanines and ultramarine blue can be used for blues and greens, but their color fastness is not as good as the aforementioned pigments.

A typical thermoset rubber compound is shown in Table 5. In terms of percentage by weight, the elastomer and filler are the chief materials used, accounting typically for over 90% of a compound. However, the other “minor” materials are quite necessary in order for the compound to have the necessary chemical, physical, and toxicological properties required for a functional packaging component. For example, without the curing agent the compound would remain a physical mixture of materials that would have the consistency of chewing gum. On the other hand, materials such as pigments may be omitted from the compound with only minor consequences — i.e., the loss of the desired color.

SELECTION OF COMPOUND MATERIALS

Many materials may be used in a rubber compound; however, only a fraction of materials are acceptable in components used for the drug industry. A source of

Table 5 Typical thermoset rubber compound

Material	Percent by weight
Chlorobutyl rubber (elastomer)	52.7
Calcined clay (filler)	39.4
Paraffinic oil (plasticizer)	4.4
Titanium dioxide (pigment)	1.1
Carbon black (pigment)	0.13
Thiuram (curing agent/accelerator)	0.14
Zinc oxide (activator)	1.6
2,6-Di-tert-butyl-4-sec-butyl phenol (antioxidant)	0.53

acceptable materials is the U.S. Code of Federal Regulations (CFR). Since the CFR has no list for drug contact, the drug industry uses the CFR list designated for foods. Applicable sections of 21 CFR are as follows:

Section 175—Indirect Food Additives

Sections 177 and 178—Indirect Food Additives Polymers

Sections 182, 184, 185—Generally Regarded as Safe (GRAS) Lists

Section 177.2600—Rubber Articles Intended for Repeated Use (The primary section containing a list of materials used in rubber formulations for pharmaceutical items.)

There are some cautions with the CFR lists. First, manufacturers may not always submit materials to the Food and Drug Administration (FDA) for listing in the CFR. They may not want to take the time or incur the costs if they see only a limited market for their material in the pharmaceutical or food market. Second, some materials, such as 2-MCBT, which is not permitted by the FDA, are listed but are strongly discouraged. Finally, some materials listed in the CFR, such as food, drug, and cosmetic dyes, are really not applicable for rubber compounds used for pharmaceuticals. Many of these dyes are water-soluble and are not applicable for rubber formulations that come in contact with aqueous solutions, since the dyes could be extracted from the rubber and discolor the drug.

Component manufacturers may use materials not listed in the CFR provided that acceptable toxicity data is available to the reviewing health authority.

TYPES OF RUBBER AND THE MANUFACTURING PROCESS

Elastomeric closures for parenteral products are made from two types of elastomers or rubbers. Thermoset rubber, the most common, undergoes a chemical reaction during the molding or component-forming processing. In this chemical reaction cross-links, or bonds are inserted between the long polymer chains to form a resilient 3D network. Without these cross-links elastomeric closures would have properties resembling those of chewing gum, which is an uncross-linked rubber blended with sugar, flavors, and food coloring. The cross-linking process is not reversible. Once a closure is molded it cannot be remolded into another shape or size. Addition of heat only causes degradation or reversion of the rubber.

Another type of rubber that is used frequently is thermoplastic rubber (1, 2). Components are fabricated in a process that is similar to that used for common hard

plastics, such as polyethylene or polystyrene, but the final product is an elastic material with properties otherwise equivalent to those of thermoset rubbers. No chemical reactions are involved in the processing of a thermoplastic rubber. The fabrication process consists of heating the rubber compound until it liquefies, injecting the liquid into a mold, cooling the mold, and finally removing the closure from the mold. The process is reversible. Closures can be remelted and remolded into different shapes or sizes as desired. Cross-linking in this case is not a chemical process but a physical intertwining of polymeric chains. The resulting intertwined 3D network gives thermoplastic elastomers their elasticity and resiliency.

Currently, thermoplastics account for less than 5% of the elastomeric closures for parenterals. Their limited resistance to heat deformation under stress during autoclave sterilization is the main reason for this limited use. However, thermoplastics have two advantages over thermosets. First, they are chemically less complex and therefore less prone to interact with parenteral medications, and second, they may be manufactured by a simpler and more automated process. Thermoplastic elastomers have found use in baby bottle nipples and dropper bulbs that are not typically heat sterilized under compression.

The manufacturing of rubber components for pharmaceutical applications is a multistep process with controls on each step. Fully validated processes are now common. The process is outlined in Table 6.

There are generally three molding processes used for manufacturing pharmaceutical closures: compression molding, injection molding, and transfer molding. The choice of molding method usually depends on the necessary final dimensional tolerances of the item being molded. Injection molding gives the best dimensional tolerances; however, it is usually the most expensive technique, especially as compared to compression molding. Thermoset rubbers are commonly compression or transfer molded, while thermoplastic rubbers are typically injection molded. The compression molding cycle is illustrated in Fig. 3. Sheets of molded components vary in shape from round to rectangular, and in size from 12 in. in diameter to 36 × 36 inches square, and may contain 50–10,000 components.

A recent trend in rubber component manufacturing is the production of “preprocessed” components. Components are prepared for shipment in either one of two states:

1. *Ready to Sterilize (RtS)*: Typically components are washed, then rinsed with Water For Injection (WFI) to reduce bioburden and endotoxin levels, lubricated with

Table 6 Manufacturing process for rubber components for pharmaceutical use

Process step	Comment
1. Raw material specifications and testing	Tests for identity and purity
2. Weighing of batch ingredients according to DMF-filed “recipe”	Weighted to ±0.2% of nominal batch weight
3. Mixing to get homogeneous batch of material	Mixed either in an internal mixer or on an open mill
4. Testing representative samples of the mixed batch of rubber	Blend of ingredients is tested for cure characteristics; molded test piece is tested for common attributes such as durometer hardness, specific gravity, percentage ash, IR spectrum of pyrolyzate, and UV spectrum of an aqueous extract.
5. Preform material by extrusion, calendering or pelletizing	Material is formed into blocks, trips, sheets or pellets for ease of handling
6. Preformed material is placed on mold or into mold feeder mechanism	
7. Molding by compression, transfer, or injection techniques	Unvulcanized mixtures of materials are shaped into the desired component by heat and pressure; chemical reaction causes cross-links to form between polymer chains
8. Trimming	Convert molded sheets of connected components into individual components
9. Washing	Remove particulate matter, and mold and trim lubricants
10. Rinsing	Reduce surface endotoxins and bioburden when WFI water is used
11. Sterilization	Only done when presterilized components are produced
12. Final QC testing	Tests for identity, dimensional measurements, and physical, chemical and biological tests
13. Packing	A counted number of components are placed in containers, usually bags made of polyethylene or Tyvek®, then sealed in cardboard cartons for shipping
14. Shipping	Cartons are grouped by lot number, palletized and shipped by truck, ship or air

silicone oil, and finally packaged in a classified area (Classes 100–10,000) in Tyvek® bags that can be steam sterilized by the drug or device manufacturer

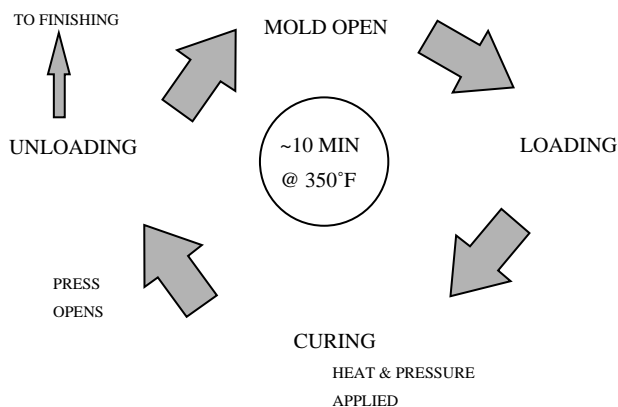


Fig. 3 Compression molding cycle.

before use. Alternately polyethylene bags may be used if the end user is utilizing gamma radiation for sterilization.

2. *Ready to Use (RtU)*: The process is identical to that used for RtS components except that the component manufacturer takes responsibility for component sterilization and the sterilized components are received by the drug or device manufacturer. No additional component processing is required before use.

Both processes typically require extensive validation by the component manufacturer, a description of the process in a Drug Master File (DMF), quality audits by customers, and perhaps FDA approval before RtS or RtU components are utilized. Nevertheless, RtS and RtU are trends that are moving rapidly; most major rubber component manufacturers now market RtS components (3). A list of pharmaceutical rubber manufacturers is shown in Table 7.

STERILIZATION OF RUBBER COMPONENTS

Heat, radiation, and sterilizing gases may be used to sterilize rubber components. However, components, especially vial stoppers, are most frequently sterilized by pressurized steam (autoclaving), a highly effective method and probably the most reliable of available methods when an F_0 value of at least 8 min is reached (4). However, the poor thermal conductivity of rubber and the relatively large mass of wet stoppers placed in a stainless steel container for sterilization require a careful validation of the process. Assurance must be obtained that the moist heat penetrates throughout the mass of stoppers so that every stopper receives a dose equivalent to a minimal F_0 of 8 min. However, rubber closures subjected to elevated temperatures or extended heating even at moderate temperatures degrade or revert, which is most frequently evidenced by stickiness. Therefore, the application of moist heat must be controlled because this degradation is most likely to be greatest at the outer layers of the mass of stoppers. This undesirable effect can be markedly reduced by distributing the stoppers in a shallow tray or otherwise reducing the mass, thus making it possible to reduce the thermal cycle time and the applied heat. Cycle times of 121°C for 30–60 min are usually well tolerated by rubber components made from butyl or halobutyl rubbers. Natural and isoprene rubbers are less tolerant. Increasing the sterilization temperature above 121°C or the time above 60 min is not recommended unless components have been tested to assure that no significant degradation will take place. Dry heat sterilization or drying of components above 105°C is also not recommended without prior testing. Testing may include inspection of the surface for cracks or tackiness, swelling studies to determine state of cure, functional tests such as coring and reseal, and chemical tests such as the United States Pharmacopoeia (USP) protocol or extraction studies performed before and after the sterilization cycle.

Under the usual autoclaving conditions, moist heat does not destroy pyrogens. For closures to be pyrogen-free, a final rinse with WFI before sterilization is necessary.

Ethylene oxide, a sterilizing gas, may be used to sterilize rubber components when they are part of a medical device; however, the gas is readily absorbed by the rubber and sufficient time must be allowed after sterilization for the concentration of residual ethylene oxide to dissipate to acceptable levels (5, 6).

Radiation sterilization, by either gamma or e-beam, also may be utilized to sterilize rubber components. However, some elastomers, such as butyl, chlorobutyl, and bromobutyl, do not tolerate high doses of radiation, while others,

such as natural, isoprene, neoprene, and nitrile rubbers are readily radiation-sterilized without degradation (7, 8). The challenge with radiation sterilization of rubber components in bulk (i.e., cartons or bags) is to obtain a uniform dose of radiation throughout the entire package—enough of a dose to assure the desired sterility assurance level (SAL), usually 10^{-6} , without degrading the components, especially those nearest the radiation source. There are generally two approaches that can be followed. One is to target a uniform dose of 25 kGy, which is generally accepted to assure sterility, while the other is to follow the ANSI/AAMI/ISO 11137:1994 standard and target a dose of radiation that is dependent on the bioburden of the components (9). The first approach is usually overkill but may be acceptable for radiation-resistant rubbers. The second approach usually results in lower doses applied, which may be critical with butyl and other rubbers that are not so radiation resistant. There are several necessary steps if the ANSI/AAMI/ISO standard is used. These are:

- Selection of the SAL, usually 10^{-6}
- Determination of the average bioburden from samples taken from three consecutive production lots
- Establishment of the Verification Dose. This is the dose in kGy found in Table B.1 of the standard that will give a SAL of 10^{-2}
- Confirmation of the Verification Dose by irradiating 100 components and testing each for sterility. No more than two positives are permitted
- Establishment of the Sterilizing Dose from Table B.1 of the standard
- Dose mapping, using dosimeters, on a shipping container filled with components. Irradiate and determine the maximum and minimum doses (D_{\max} and D_{\min}) received in the container
- Calculation of the Target Dose by multiplying the Sterilizing Dose by the ratio, D_{\max}/D_{\min} . The Target Dose is typically increased 6% or more to account for dosimeter error and to add additional sterility assurance.

TESTS AND STANDARDS

In-Process Tests

Several tests may be performed on unvulcanized rubber or on standard-shaped test specimens to measure the properties of a rubber compound (10). These include the following:

1. *Rheometer measurements* measure cure and cure rate characteristics of the rubber. The component

Table 7 Pharmaceutical rubber packaging component manufacturers

Name	Symbol ^a	Address	Phone	Website or e-mail address
Abbott Laboratories	AB	268 E. Fourth Street, Ashland, OH 44805, USA	+419-282-5378	www.abbott.com
AL group Wheaton	AL	618 Beam St., Salisbury, MD 21801, USA	+410-546-6441	www.alcanpackaging.com
Bryant Rubber Corp.	BR	1112 Lomita Blvd., Harbor City, CA 90710, USA	+310-530-2530	www.bryantrubber.com
Daikyo Seiko, Ltd.	DS	38-2, Sumida 3-Chome, Sumida-Ku, Tokyo 131, Japan	+81-3-3614-5461	
Helvoet Pharma	HP	Industrieterrein Kolmen 1519, B-3570 Alken, Belgium	+32-11-59-08-00	www.helvoetpharma.com
Itran-Tompkins Rubber Corp.	IT	375 Metuchen Rd., South Plainfield, NJ 07080, USA	+908-754-8100	www.iranrubber.com
Kokoku Rubber Inc.	KO	1450 E. American Lane, Zurich Towers, # 1545, Schaumburg, IL 60173, USA	+847-517-6770 Ext 12	www.kokokurubber.com
Lexington Medical	LR	663 Bryant Blvd. Rock Hill, SC 29732, USA	+803-366-7036	
The Plasticoid Company	PL	249 High Street, Elkton, MD 21921, USA	+410-398-2800	www.plasticoid.com
Samsung Medical Rubber Co., Ltd.	SA	474-4, Mokne-Dong, Ansan-City, Kyunggi Province, Korea	+82-345-491-8071	www.smrco.co.kr
Seal Line S.p.A.	SL	Via Bernarde, 7, 36040 Montegaldella (VI), Italy	+39-0444/737221	Sealline@keycomm.it
Stelmi Trading International	ST	121, avenue Jean Mermoz, BP-93, F-93127 La Courneuve Cedex, France	+(33)-1-49-92-64-00	www.stelmi.com
West Pharmaceutical Services	WP	101 Gordon Drive, Lionville, PA 19341, USA	1-800-231-3000	www.westpharma.com

^aAbbreviation for reference use in this chapter only.

manufacturer performs this test on unvulcanized rubber. The rheometer measures the viscosity of the rubber as a function of time at a constant temperature. As time increases, the degree of cure or cross-linking increases and thus the viscosity increases.

2. *Durometer hardness* is measured on tests specimens that meet specific standards for shape and thickness (11). Durometer hardness is usually measured using the Shore A scale, which measures relative hardness on a scale of 0 to 100 units. Most rubber components for medical use are found in the 35–60 range with 40–50 typical for rubber vial stoppers. Durometer Hardness may be measured on some actual components if they have a sufficiently large flat surface and thickness, i.e., 28–32 mm IV stoppers.
3. *Compression set* is commonly used as a measure of the dimensional recovery of a rubber compound after compression at a defined level, usually 25%, at a specified time and temperature, usually 24 h at 70°C (12). High compression set values are associated with rubber that “takes a set” or loses its ability to spring back after compression. Low compression set is important for rubber closures and syringe plungers that are heat sterilized while under compression and remain under compression for long periods of time before use but must remain elastic and resilient to maintain seal integrity. Compression Set is measured on dimensionally defined test specimens.
4. *Tensile, modulus, and elongation* are measures of the strength of a rubber compound (13). They are measured on bow tie-shaped specimens that are clamped in a tensile measuring apparatus and stretched.
5. *Water vapor and oxygen transmission (WVT and O₂T)* are commonly measured on thin-film specimens. Butyl and halobutyl compounds have very low WVT and O₂T rates, while rates for natural and isoprene rubbers are higher and for silicone even higher.

Finished Component Tests

Finished component tests may be divided into three categories: those used as routine identity and/or quality control tests, those tests recommended or mandated by government, standards, and compendial groups, and those test that are part of the larger rubber component acceptance and drug/device approval process. In many instances, a test may fall into more than one category.

Identity and quality control tests (14, 15)

Percentage ash is a measure of the nonvolatile materials in a rubber formulation, such as clays and other fillers.

Specific gravity is a measure of the type and quantity of fillers in a formulation.

An *infrared spectrum* of a compound pyrolysate identifies the elastomer qualitatively (natural, butyl, etc.).

An *ultraviolet (UV) spectrum* of an aqueous rubber extract identifies and quantifies the antioxidants, curing agents, accelerators, and other UV-absorbing species extracted from the compound.

Percentage swelling in an organic solvent is a measure of the degree and consistency of cure or cross-linking that determines the physical and functional properties of a component.

These five tests may be used to characterize a rubber formulation or serve, either individually or in combinations, as quality control tests.

Compendial, standard, and government tests

The most influential of these test protocols are the USP (16), the *European Pharmacopoeia* (EP) (17), the *Pharmacopoeia of Japan* (JP) (18), the Organization for International Standardization (ISO) (19), and the Parenteral Drug Association (PDA) (15, 20). USP<381>, *Elastomeric Closures for Injections*, contains five chemical tests and two biological tests. Closures must meet the biological requirements but there are no current specifications for the chemical tests. All USP chemical tests are commonly performed on aqueous extracts but isopropyl alcohol and the drug product vehicle are also permitted. A brief description of the USP<381> tests follows (21).

Turbidity. The clarity of the closure extract is measured with a nephelometer with appropriate standards. Turbidity is a measure of the insoluble extractables from a closure and is affected by the type and amounts of ingredients in a formulation, pretreatments such as washing, extractions, sterilizations, and degree of cure. *Reducing agents.* Organic extractables oxidizable by iodine are determined in this procedure. It is affected by the same variables as turbidity.

Heavy metals. Extractable lead as well as other metals, such as zinc and cadmium, are determined colorimetrically or by atomic absorption.

pH change. The pH of the extract is a measure of the acidic and alkaline water-soluble extractables from a compound.

Total extractables. The sum of the inorganic, organic (nonvolatile), soluble, and insoluble extractables is measured in this test. The weight of total extractables is an indication of the “cleanliness” of a formulation.

Biological tests. USP 24 lists two levels of biological tests—the in vitro tissue culture test and the in vivo systemic injection and intracutaneous tests. A parenteral

closure must pass either type of test to meet USP requirements.

The application of numerical specifications to the USP <381> chemical tests is under discussion and probably will become effective in USP 24 via a supplement. The JP, EP, and ISO have specifications for rubber closures and these test protocols are compared with the USP in Table 8. Not all pharmacopeias and standards groups designate the same tests; therefore, it is important to consult a wide spectrum of references to design a test protocol for specific applications and geographical submissions. Of the four protocols compared, the JP is the most stringent in terms of limits for extractables. The USP, EP, and ISO test protocols are based on a specific closure “area” per volume of water in the extraction step. The JP test protocols, on the other hand, are based on a specific “weight” of closures per volume of water. This difference makes it more difficult for smaller closures than for larger closures in the same rubber compound to meet JP specifications; i.e., 13-mm closures will be less likely to meet JP specifications than 20-mm closures, and 20-mm closures are less likely than 28-mm closures to meet JP specifications.

Although packaging components, such as vial closures, syringe plungers, and needle shields, may meet all compendial and accepted standards, this does not mean that they will be acceptable for use with any specific drug or device. Specific evaluation tests must be performed for that purpose. Compendial and standard tests should be regarded as the first necessary but not sufficient hurdle in the race to gain approval of a component for use with a drug or device.

Specific tests for component evaluation and regulatory approval

There are three requirements for a rubber component:

1. Compatible with the drug
2. Meets functional requirements
3. Provides closure-container seal integrity

Many factors influence the choice of a rubber compound for a particular drug, the most important of which is the solvent vehicle. If the solvent vehicle is an aqueous material, then a butyl, natural, or EPDM may be used. If the solvent vehicle is an oil, then a neoprene or nitrile is utilized.

Configuration is also important and is determined by the required function. A lyophilization stopper, designed to keep out moisture, almost certainly requires a butyl formulation, while a vial stopper for aqueous-based solutions could be formulated from isoprene rubber.

Some preservatives are especially reactive with rubber. Bromobutyl rubbers, but not chlorobutyls, are recommended for drug formulations that contain chlorobutanol. A pH that is either very low or very high affects rubber formulations more than a pH in the range of 5–8. Buffer systems may also affect the choice of rubber formulations. Materials such as phosphates not only attack the rubber at high pHs but also may attack glass as well.

Metallic sensitivities affect compatibility, as drug formulations that are sensitive to divalent cations, such as calcium, zinc, or iron, may not be able to use certain rubber compounds that are cured or pigmented with these materials.

An obvious factor is the need for oxygen or moisture vapor protection. A butyl-based compound is normally chosen whenever protection from materials transmitted either into or from a drug product is required.

When it comes to color preference, most rubber manufacturers of closures used for drugs prefer to use three pigments: iron oxide to produce reds, carbon black for blacks, and titanium dioxide for whites; combinations of these are used for pinks and grays. Organic materials used as pigments are not generally acceptable since they are not very heat stable and are generally more toxic than the three pigments mentioned.

Choice of sterilization method is extremely important in choosing a rubber formulation. Many heat-sensitive drugs, such as proteins, are packaged aseptically; that is, the rubber closure, the vial, and drug are sterilized separately, and then all three items are brought together in a sterile environment to form the final package. The FDA, however, encourages terminal sterilization. In this method, the three materials are brought together and then the entire package, the drug in contact with the vial and closure, is sterilized by heat. This method is much more demanding on the closure than aseptic processing.

Radiation (Co-60) is used to sterilize many rubber items, especially those used in devices. The effect of gamma radiation on rubber closures is a function of the elastomer, dose, and postirradiation time. NR and isoprene are much more resistant to irradiation effects than butyl.

To give the drug manufacturer a high degree of assurance that the closures being investigated for a possible package will be acceptable, a prescreening procedure is commonly used by component suppliers. In this procedure, information and a sample of the drug are obtained from the drug manufacturer. Then three to five possible closures, along with an inert control stopper (a Teflon[®] plug or coated closure), are used to package vials of the drug. The closures are put onto the drug-containing vials using an exaggerated closure area-to-volume ratio (2× to 3× the normal ratio). Vials are stored at higher and/or lower temperatures than

Table 8 Comparison of USP, EP, JP, and ISO test protocols for rubber closures

	USP ^a	EP ^b	JP ^c	ISO ^d
Test types and limits				
Chemical tests	Y	Y	Y	Y
Biological tests	Y	N	Y	N
Functional tests	N	N	N	Y
Test limits	Y, only on Biological Tests	Y	Y	Y
Specific tests				
Alkalinity/pH	Y	Y	Y	Y
Ammonia	N	Y	N	Y
Appearance/turbidity	Y	Y	Y	Y
Ash, total	N	Y	N	Y
Conductivity	N	N	N	Y
Container-closure integrity	N	N	N	Y
Design/dimensional specifications	N	N	N	Y
Elasticity	N	Y	N	Y
Extractable zinc	N	Y	Y	Y
Foam	N	N	Y	N
Fragmentation/coring	N	N	N	Y
Halides	N	N	N	Y
Hardness	N	N	N	Y
Heavy metals	Y	Y	N	Y
Hemolysis	N	N	Y	N
Intracutaneous injection	Y	N	N	N
IR of pyrolysate	N	Y	N	Y
Particulate matter	N	N	N	Y
Penetrability	N	N	N	Y
Pyrogens	N	N	Y	N
Reducing substances	Y	Y	Y	Y
Residue on evaporation	Y	Y	Y	Y
Resistance to steam	N	N	N	Y
Self sealing	N	N	N	Y
Storage	N	N	N	Y
Systemic injection	Y	N	Y	N
Tissue culture	Y	N	N	N
Total Cd and Pb	N	N	Y	N
UV absorbance	N	Y	Y	Y
Volatile sulfides	N	Y	N	Y

^aFrom Ref. 16, Section <381>.^bFrom Ref. 17, Section 3.1.12.^cFrom Ref. 18, Section 49.^dFrom Ref. 19, ISO 8871, ISO 8362-2 & 8362-5, ISO 8536-2 & 8536-6.

the drug package will normally experience in the upright, inverted and on-side positions.

An inert control stopper should be used since a drug may be stable against glass but not against uncoated rubber. Only when an inert control is used will it be possible to determine whether the rubber and/or glass vial is the cause of the drug instability. Ampoules do not make good controls for the prescreening of drugs in vials since the glass for vials may be different from that used for ampoules.

Table 9 summarizes the important drug compatibility factors that must be considered when choosing a rubber component for a specific drug/device application. These factors along with the functional and seal integrity requirements can be used to write specifications for components. The packaging engineer, as with the rubber compounder, faces the challenge of choosing a compound that best meets the overall requirements but, only at best, provides a balance of all the actual chemical and functional needs. For example, an engineer may choose

an isoprene vial closure because multiple punctures with a large cannula are required. Isoprene will provide the low coring and excellent reseal required. But the engineer may give up some shelf life since isoprene is a poor oxygen barrier. Choosing a butyl closure would give additional shelf life due to its better barrier properties, but coring and reseal would not be acceptable. Closure compound choice is always a give-and-take proposition where drug/device requirements must be prioritized before the choice of a rubber compound can be made. Table 10 lists important physical and chemical properties of elastomers (22).

There are four general types of closure–drug interactions:

1. Adsorption occurs when a drug is concentrated at the surface of a closure or vial.
2. Absorption occurs when a drug material is dispersed in the closure matrix.
3. Permeation is the transmission of a drug ingredient through a closure into the atmosphere or transmission of an outside material into the container.
4. Leaching is the process by which closure ingredients are extracted into the drug product.

All four of these interactions commonly occur. No rubber compound is absolutely inert to a drug. In many cases, the extent of the particular interaction is extremely small and may not be measurable, but generally all four are occurring, albeit at a low rate. With proteins, adsorption can be a problem. Many proteinaceous materials made by the biotechnology industry are highly adsorbent onto rubber surfaces and their potency may be readily lost. Other drugs products, especially ones that are very acidic or basic, may attack the stopper and cause the extraction of rubber compound ingredients. Common leachables from rubber closures include low molecular weight elastomer fragments, metal ions, antioxidants, plasticizers, lubricants, curing agents, and accelerators. The rate and relative importance of these four interactions determines the degree of compatibility or incompatibility of a stopper with a drug product.

In May 1999, the FDA issued an updated guidance entitled “Container Closure Systems for Packaging Human Drugs and Biologics—Chemistry, Manufacturing and Controls Documentation,” that listed in tables the packaging information that should be submitted in an application (23). The guidance divides the information into four sections:

1. *Description*—Overall general description of the container-closure system plus specific information on suppliers, materials of construction, and postmanufacturing treatments.

Table 9 Drug compatibility factors affecting the choice of rubber component

Liquid or solid?
Liquid
1. Aqueous—pH, preservative, buffer cosolvent?
2. Oil—Vegetable or Mineral?
Solid
1. Lyophilized or powder fill?
Configuration?
Need for O ₂ , H ₂ O, CO ₂ protection?
Drug type?
Metallic sensitivities?
Method of closure and package sterilization?
Color preference?

2. *Suitability*—This section prescribes that tests must be done to assure protection of the drug product, safety of the packaging component material, compatibility of the component with the drug product, and performance of the component.
3. *Quality control*—The rubber component manufacturer’s release criteria and drug packager’s acceptance are found in this section. Also recommended is a method to monitor the consistency in composition of elastomeric components such as periodic extraction profiles.
4. *Stability*—Testing of the drug product using the packaging component is required. Unlike compendial tests that utilize water for extraction studies of the packaging component, the complete drug product is utilized in these stability studies.

This guideline advocates more information on rubber extractables and the proper use of DMFs.

Knowledge of extraction data from elastomeric components refers not only to the broad (and usually nonspecific) type of extractable data generated in USP<381> testing, but also to the identification and quantification, where necessary, of specific extractable species. Example of nonspecific extractables include turbidity, reducing agents, heavy metals, pH change, and total extractables; They all measure broad types of extractables. Specific extractables include inorganics, such as zinc or lead and organics, such as PNA, stearic acid, nitrosamines, tetramethylthiuram disulfide, or 2,6-di-*tert*-butyl-4-*sec*-butyl phenol. Both liquid (HPLC) and gas chromatography (GC) are well suited for this purpose (24). Drug/device producers who utilize elastomeric packaging components require information from their suppliers on extractables (extractable profiles), including test methods, that are generated in water at various pH values (i.e., 3, 7, and 10) and in organic solvents such as isopropanol.

Table 10 Physical and chemical properties of elastomers^{a,b}

Property	Common name of elastomer (chemical name)							
	Butyl/halobutyl isoprene copolymer	Natural/Isoprene (cis-1,4-polyisoprene)	Neoprene (polychloroprene)	Nitrile (butadiene-acrylonitrile copolymer)	Silicone (polydimethylsiloxane)	Fluoro-elastomers (fluororubber)	Urethane (polyester isocyanate)	EPDM (ethylene propylene diene monomer)
Abrasion resistance	Fair	Good	Fair	Good	Fair	Good	Excellent	Good
Compression set	Poor	Excellent	Good	Good	Poor	Good	Excellent	Good
Coring	Fair	Excellent	Good	Fair	Poor	N.D. ^c	Excellent	Fair
Gas transmission resistance	Excellent	Good	Fair	Good	Poor	Good	Poor	Fair
Heat resistance	Excellent	Good	Good	Good	Excellent	Excellent	Poor	Very Good
Machine-ability	Poor	Good	Fair	Good	Fair	N.D.	Fair	Fair
Moisture vapor resistance	Excellent	Good	Fair	Fair	Poor	Good	Poor	Fair
Ozone resistance	Excellent	Poor	Good	Fair	Excellent	Excellent	Good	Fair
Radiation resistance	Fair to Poor	Good	Good	Good	Fair to Good	Fair to Good	Fair	Poor
Resilience	Poor	Excellent	Good	Good	Good	Fair	Good	Good
Shelf-life	Good	Fair	Good	Fair	Excellent	Excellent	Excellent	Fair
Solvent resistance								
Acid, dilute	Good	Good	Good	Good	Fair	Fair	Poor	Fair
Aliphatic solvents	Poor	Poor	Good	Poor	Poor	Excellent	Excellent	Poor
Alkali, dilute	Good	Good	Good	Good	Good	Good	Poor	Fair
Animal oil	Excellent	Poor	Good	Excellent	Good	Excellent	Excellent	Fair to poor
Aromatic solvents	Good	Good	Poor	Good	Poor	Excellent	Poor	Poor
Chlorinated solvents	Poor	Poor	Poor	Poor	Poor	Excellent	Good	Poor
Mineral oil	Poor	Poor	Good	Excellent	Fair	Excellent	Excellent	Poor
Vegetable oil	Excellent	Poor	Good	Excellent	Excellent	Excellent	Excellent	Fair to poor
Water	Excellent	Good	Fair	Good	Excellent	Good	Poor	Good

^aRatings adapted from Ref. 22.^bRatings expressed are typical for rubber compounds made from the elastomers; they can vary significantly from compound to compound.^cND, not determined.

Armed with this information, they can look for extractables in their drug products.

Confidential packaging component information, such as the compound recipe, may be placed in a Type III DMF so that the FDA can review the information when it reviews the drug application (IND, NDA, ANDA, or BLA) (25). Most elastomeric compounds are filed at the request of a pharmaceutical manufacturer who has chosen to use the rubber closure in one or more drug packages or devices applications. The name of a rubber compound is associated with a precise recipe that designates specific ingredients and quantities. Once a rubber compound is filed in a DMF, no changes can be made in that compound without changing the DMF and notifying all customers on whose behalf the DMF was accessed and reviewed by the FDA. Since changes in a supplier's DMF may require additional stability studies by the drug manufacturer, changes are infrequent.

RECENT ISSUES AND DEVELOPMENTS

Latex Sensitivity

There are medical and regulatory issues surrounding the use of "latex rubber" due to allergic reactions that have resulted in medical emergencies. Even deaths have been noted (26, 27). There are two broad types of rubber—natural and synthetic. Several synthetic rubbers or elastomers are used for pharmaceutical components. These are listed in Table 10. However, there is only one type of commercial NR, which is derived from the rubber tree *Hevea Brasiliensis*. "Latex sensitivity" is associated only with NR and not with the synthetics, although other types of allergic reactions can result from contact with synthetic rubbers. NR is processed and used in two forms—liquid or latex rubber and solid or dry rubber, often referred to as crepe, SMR, and SIR. Specific proteins that are contained in both the latex and dry types cause the sensitivity to NR. Thus, the medical community and regulatory authorities have used the term "latex" for both forms of NR when "latex sensitivity" is discussed. Latex reactions reported in the literature have thus far been attributed to contact with components made from liquid rubber but not from dry rubber. Components made from liquid rubber are made by a dipping process that is best suited for thin-walled items such as gloves, condoms, and catheters. None of the typical rubber packaging components shown in Figs. 1 and 2 is made from latex rubber;—they are all made from dry rubber via a molding process. Many studies have been published regarding the allergic

properties of latex and dry rubber (28, 29) and further studies are in progress to determine if exposure to these dry rubber components can cause allergic latex reactions (30).

There is a great deal of regulatory activity in an attempt to protect the public from unexpected latex reactions. In 1996, the USP proposed a change in section <381> that would prohibit the use of NR in elastomeric closures (31); however, it rescinded this change in April 1997, stating that closure manufacturers should instead devise latex protein limits. A test for the water-soluble protein content of elastomers, based on an ASTM test (2), was published as USP <836> but no specifications or limits were proposed (33). At the same time, the FDA published a final rule that made it mandatory to provide labeling statements on medical devices and packaging components that contain NR (34). This rule was later amended to exclude combination drug/device and biologic/device products such as rubber stoppers and plungers for prefilled syringes (35). The uncertainty about the allergic risk from dry NR components and pending regulations have significantly reduced the use of NR in packaging for new drugs and in devices. Substitutes such as isoprene and SBR rubbers are finding increased use.

Preprocessed Components

Preprocessed closures, commonly referred to as RtS or Ready for Sterilization and RtU or Ready for Use, are an unstoppable trend in pharmaceutical packaging. Information on the manufacturing of these products was described previously in this review. The purpose of preprocessed closures is to reduce total processing costs and improve closure characteristics. Typical RtS closure characteristics are as follows:

Endotoxin Level: < 1 EU/closure

Bioburden Level: < 2 cfu/closure

Silicone Level: 10–40 µg/closure

Visible Particulate Matter: < 20 particles in 25 to 50-µ range; < 2 particles in 50 to 100-µ range; < 1 particle over 100 µ.

Coating and Surface Treatments

Although the material science of rubber compounds has greatly improved, drug-closure-interactions and surface lubricity are problematic for packaging engineers. Surface modifications of closures are frequently necessary to meet acceptance standards set by the USP, EP, JP, and ISO as well as the expectations of regulatory authorities. In practice, both liquids and solids are applied to closure surfaces to minimize interactions and improve lubricity.

Table 11 Commercially available coated and surface treated components

Product name	Material	Primary use	% Surface coated	Supplier ^a
Abboclad	Fluorinated polymer solid film	Compatibility	Plug surface	AB
B2	Silicone, polymerized liquid	Lubricity	Top, plug or both surfaces	WP
Coated stopper	Parylene deposited solid film	Compatibility	Total	IT
Deposition coated stopper	Vapor deposition coated solid film	Compatibility	Total	AB
Flurotec	ETFE solid film	Compatibility	Top, plug or both surfaces	WP & DS
Omniflex plus	Solid fluoropolymer coating	Compatibility	Total	HP
R2	Polymer, nonsilicone solid	Lubricity	Total	ST
SAF	Silicone, high viscosity liquid	Lubricity	Total	HP
Slipcoat	Plasma deposited polymer solid	Lubricity	Total	AL

^aSee Table 7 for supplier information.

A compilation of commercially available coated and surface treated components is shown in Table 11 (36). Although these coating and surface treatments may add significantly to the purchase price of closures they often reduce the total drug product cost by providing the following benefits:

- Increased lubricity, which allows faster processing speeds
- Decreased drug–closure interactions, which permits the marketing of some products not compatible with uncoated rubber and better quality and longer shelf life for others
- Decreased particulate matter, which reduces the number of units rejected for visible particulate matter

Container-Closure Seal Integrity

The FDA Guidance on Container Closure Systems (Ref. May 1999 Guidance) lists sterility or container integrity as an important parameter to be considered in the section on Protection. Seal integrity tests can be done by both physical and microbial methods, but historically, sterility testing alone has been used. A 1998 FDA draft guidance discusses the replacement of the sterility test with an appropriate container-closure integrity test in the stability protocol, permitting an alternative to sterility testing for proving the continued capability of containers to maintain sterility (37). Kirsch et al. (38–41) published a series of four papers that studied mass spectrometry-based helium leak detection, microbial ingress, and vacuum decay and the correlation between these methods. The PDA also published an updated technical report that provides guidance for evaluating pharmaceutical package integrity (42), and Guazzo has published an excellent review article that outlines the advantages and limitations of current methods (43).

SUMMARY

Rubber packaging and device components are an important part of the overall medical delivery system. Without innovative packaging systems, modern drugs would not be available today. Advances in drug development have initiated research in new packaging and delivery systems while the availability of innovative packaging has led to the introduction of new drug therapies. Innovation, while containing costs and conforming to regulations, is the challenge for the 21st century.

REFERENCES

1. Williams, J.L. Medical Applications for Thermoplastic Elastomers Proceedings of the 1st International Conference on Thermoplastic Elastomer Markets and Technology, Schotland Business Research, Inc.: Princeton, NJ, 1988.
2. Naugle, D.G., Djiauw, L.K., *Elastomerics* **1989**, 121, 23–32.
3. Manufacturers HP, ST, and WP, see Table 7 for contact information.
4. Avis, K.E., Smith, E.J., *Elastomeric Parenteral Closures. Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc. New York, 1992.
5. Anderson, S.R., Ethylene Oxide Residues in Medical Materials. *Bull. Parent. Drug Assoc.* **1973**, 27 (2), 49–58.
6. Baan, E. Ethylene Oxide Adsorption and Desorption of Elastomers and Plastics, *Bull. Parent. Drug Assoc.* **1976**, 30 (6), 299–305.
7. Effects of Gamma Irradiation on Elastomeric Closures, Technical Report No. 16. PDA. *J. Parent. Sci. Technol.* **1992**, 52, S1–S13.
8. Langlade, V., Le Gall, P.A., *Complying with the EC Note for Guidance for Gamma Irradiation: Validation of Rubber Closure Radiosterilization*; PDA International Congress: Basel, Switzerland, 1994; 261–273.

9. *Sterilization of Health Care Products—Requirements for Validation and Routine Control: Radiation Sterilization*; ANSI/AAMI/ISO 1137:1994, AAMI: Arlington, VA, 1994.
10. Morton, M. Ed. *Rubber Technology*, 2nd Ed.; Van Nostrand Reinhold: New York, 1973.
11. *Standard Test Method for Rubber Properties: Durometer Hardness*, ASTM D2240-97e1, ASTM: West Conshohocken PA, 1999.
12. *Standard Test Method for Rubber Properties in Compression*, ASTM D 575-91(1996), ASTM: West Conshohocken, PA, 1999.
13. *Standard Test Method for Vulcanized Rubber and Thermoplastic Rubbers and Thermoplastic Elastomers—Tension*, ASTM D 412-98a, ASTM: West Conshohocken, PA, 1999.
14. *Standard Test Methods for Rubber Products—Chemical Analysis*, ASTM D 297-93(1998), ASTM: West Conshohocken, PA, 1999.
15. *Elastomeric Closures: Evaluation of Significant Performance and Identity Characteristics*, Technical Methods Bulletin No. 2, PDA, Bethesda, MD, 1981.
16. *The United States Pharmacopeia 24/ National Formulary 19*, The United States Pharmacopeial Convention, Inc.: Rockville, MD, 2000.
17. Rubber for Closures for Containers for Aqueous Parenteral Preparations and for Powders for Freeze-Dried Products. *European Pharmacopoeia*, Section 3.1.12 in 1999 Addendum European Pharmacopoeia Commission, Strasbourg, France, 1996.
18. Test for Rubber Closures for Aqueous Infusions, *The Pharmacopoeia of Japan, Yakuji Nippo, Ltd.* Section 49, Society of Japanese Pharmacopoeia, Tokyo, Japan, 1996.
19. International Organization for Standardization (ISO), Geneva, Switzerland. Some Important ISO Standards are 8362, Parts 2 & 5, Closures for Injection Vials; 8536, Parts 2 & 6, Closures for Infusion Bottles; and 8871, Part 1, Extractables from Elastomeric Parts for Parenterals. (ISO Standards are in Continual Revision; Contact ISO for Latest Revision.).
20. Extractables from Elastomeric Closures, Analytical Procedures for Functional Group Characterization/Identification. *Technical Methods Bulletin No. 1*, PDA, Bethesda, MD, 1980.
21. *Elastomeric Closures for Injections*; USP<381>, USP 24 The United States Pharmacopeia Convention, Inc., Rockville, MD, 2000.
22. Smith, E.J., Nash, R.J. Elastomeric Closures for Parenterals. *Pharmaceutical Dosage Forms: Parenteral Medications*; Avis, K.E., Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1992; 1.
23. FDA Guidance for Industry, *Container Closure Systems for Packaging Human Drugs and Biologics—Chemistry, Manufacturing, and Controls Documentation*, Food and Drug Administration, Rockville, MD, May 1999.
24. Milano, C.J., Bailey, L.C., Evaluation of Current Compendial Physicochemical Test Procedures for Pharmaceutical Elastomeric Closures and Development of an Improved HPLC Procedure, PDA. J. Pharm. Sci. Tech. **1999**, 53, 202–210.
25. INDIA, Investigational New Drug Application; NDA, New Drug Application; ANDA, Abbreviated New Drug Application; BLA, Biologic License Application.
26. Allergic Reaction to Latex Containing Medical Devices, FDA Medical Alert, MDA 91-1 **1992**, 1–2.
27. Smith, E.J., *Natural Rubber/Latex Closures and Devices: The Issues*; PDA Spring Meeting, San Diego, CA, 1997.
28. Yip, E., Turjanmaa, K., Makinen-Kiljunen, S., The “Non-allergenicity” of NR Dry Rubber Products, with Reference to Type 1 Protein Allergy, *Rubber Developments* **1995**, 48, 48–52.
29. Slater, J.E., Latex Allergy, *J. Allergy Clin. Immunol.* **1994**, 94, 139–149.
30. Hamilton, R.G., Allergenic Content of Elastomeric Closures, Private Communication. The Johns Hopkins University School of Medicine, Baltimore, MD, 1999.
31. Elastomeric Closures for Injections, *Pharmaceutical Forum* **1996**, 22 (5), 2886–2887, USP<381>.
32. *Standard Test Method for Analysis of Protein in Natural Rubber and Its Products*, ASTM D 5712-95, ASTM, West Conshohocken, PA, 1999.
33. Natural Rubber—Water-Soluble Protein Content, *Pharmacopoeial Forum* **1997**, 23 (5), 4681–4683, USP<836>.
34. 21 CFR Part 801 Natural Rubber-Containing Medical Devices, User Labeling, Fed. Reg. **1997**, 62 (189), 51021–51030.
35. 21 CFR Part 8021, Amended Economic Analysis of Final Rule Requiring Use of Labeling on Natural Rubber Containing Devices, Fed. Reg. **1998**, 63 (104), 29552–29590.
36. Smith, E.J. Coatings and Surface Treatments for Rubber Closures for Parenterals, PDA Annual Meeting: Washington, DC, 1999.
37. FDA Draft Guidance for Industry, *Container and Closure Integrity Testing in Lieu of Sterility Testing as a Component of the Stability Protocol for Sterile Products*, Food and Drug Administration: Rockville, MD, 1998.
38. Kirsch, L.E., Nguyen, L., Moeckly, C.S., Pharmaceutical Container/Closure Integrity I: Mass Spectrometry-Based Helium Leak Rate Detection for Rubber-Stoppered Glass Vials, *PDA J. Pharm. Sci. Technol.* **1997**, 51, 187–195.
39. Kirsch, L.E., Nguyen, L., Moeckly, C.S., Gerth, R. Pharmaceutical Container/Closure Integrity II: The Relationship Between Microbial Ingress and Helium Leak Rates in Rubber-Stoppered Glass Vials. *PDA J. Pharm. Sci. Technol.* **1997**, 51, 195–202.
40. Kirsch, L.E., Nguyen, L., Gerth, R. Pharmaceutical Container/Closure Integrity III: Validation of the Helium Leak Rate Method for Rigid Pharmaceutical Containers. *PDA J. Pharm. Sci. Technol.* **1997**, 51, 203–207.
41. Nguyen, L.T., Muangsiri, W., Schiere, R., Guazzo, D.M., Kirsch, L.E., Pharmaceutical Container/Closure Integrity IV: Development of an Indirect Correlation Between Vacuum Decay Leak Measurement and Microbial Ingress, *PDA J. Pharm. Sci. Technol.* **1999**, 53, 211–218.
42. Pharmaceutical Package Integrity, Technical Report No. 27. *PDA J. Parent. Sci. Technol.* **1998**, 52 (S2), 1–48.
43. Guazzo, D.M., Current Approaches in Leak Testing Pharmaceutical Packages. *PDA J. Pharm. Sci. Technol.* **1996**, 50, 378–385.

EFFERVESCENT PHARMACEUTICALS

Nils-Olof Lindberg

Pharmacia AB, Helsingborg, Sweden

Henri Hansson

Galenica AB, Meleon, Malmö, Sweden

INTRODUCTION

Effervescent tablets are uncoated tablets that generally contain acid substances and carbonates or bicarbonates, and that react rapidly in the presence of water by releasing carbon dioxide. They are usually dissolved or dispersed in water before administration (1).

Effervescent mixtures have been known for over 250 years. During the 1930s, the success of Alka Seltzer created a vogue for effervescent products, including tablets (2). Effervescent tablets have been reviewed (3–5).

Effervescent reactions have also been employed in other dosage forms, such as suppositories (laxative effect), vaginal suppositories (mainly contraceptive effect), and drug delivery systems (e.g., floating systems and tablets rapidly dissolving in the saliva).

Effervescent products should be stored in tightly closed containers. Desiccants are usually added to the containers.

PHARMACOPEIAL MONOGRAPHS

Soluble, effervescent tablets are prepared by compression. In addition to active ingredients, they contain mixtures of acids (citric acid, tartaric acid) and sodium bicarbonate (NaHCO_3) that release carbon dioxide when dissolved in water (6). The *United States Pharmacopeia* (USP) 24 includes the following seven monographs: Acetaminophen for Effervescent Oral Solution; Aspirin Effervescent Tablets for Oral Solution; Potassium Bicarbonate Effervescent Tablets for Oral Solution; Potassium Bicarbonate and Potassium Chloride for Effervescent Oral Solution; Potassium Bicarbonate and Potassium Chloride Effervescent Tablets for Oral Solution; Potassium and Sodium Bicarbonates and Citric Acid for Oral Solution; and Potassium Chloride, Potassium Bicarbonate, and Potassium Citrate Effervescent Tablets for Oral Solution (7).

Effervescent tablets as well as effervescent granules and powders are mentioned in the *European Pharmacopoeia*

(Ph. Eur.), although it does not contain any monographs regarding specific drugs (1, 8).

THE EFFERVESCENT REACTION

Acid–base reactions between alkali metal bicarbonate and citric or tartaric acid have been used for many years to produce pharmaceutical preparations that effervesce as soon as water is added. In such systems, it is practically impossible to achieve much more than an atmospheric saturation of the solution with respect to the released carbon dioxide. If the acid dissolves first, then the bulk of the reaction takes place in the saturated solution in close proximity to the undissolved bicarbonate particles. If the bicarbonate dissolves faster, the reaction essentially takes place near the surface of the undissolved acid. Such suspension systems do not favor supersaturation with respect to carbon dioxide because the particulate solids act as nuclei for bubble formations (9).

RAW MATERIALS

General Characteristics

With regard to compressibility and compactibility, the considerations pertaining to raw materials in effervescent products are similar to the ones that prevail in evaluating raw materials intended for conventional tablets. However, poor compactibility cannot usually be compensated for by the use of binders, as this will prevent a rapid dissolution of the effervescent tablet. Addition of a binder is generally not as critical for the dissolution of effervescent granules or powders.

The general tablet compaction process normally is described by a number of sequential phases: rearrangement, deformation (elastic, plastic) of initial particles, fragmentation, and deformation of fragments. Particle surfaces are

brought into close proximity and interparticulate attraction or bonds will be formed (10). Similar conditions will prevail with the effervescent tablets.

A very important property for effervescent products is the adsorption/desorption isotherm of the raw material and, consequently, its moisture content. To avoid a premature effervescent reaction in the tablets, substances with low moisture contents will have to be used. The aqueous solubility is another important property of the substances used in effervescent products. It is also important to use raw materials that are easily wetted. Of course, the taste of the employed substances is important.

Acid Materials

The acidity for the effervescent reaction can be obtained from three main sources: acids, acid anhydrides, and acid salts. Traditional sources of acid materials are the organic acids, citric and tartaric acid; however, some acid salts also are used.

Acids

Citric acid: Citric acid is obtained as a monohydrate or an anhydrate. A variety of particle-size grades are available—colorless, translucent crystals, or white, granular-to-crystalline powder. Citric acid is odorless and has a strong acidic taste. It is soluble in less than 1 part of water and 1 in 1.5 parts of ethanol (11).

Citric acid monohydrate melts at 100°C. It loses water at 75°C, becomes anhydrous at 135°C, and fuses at 153°C. At relative humidities (RH) lower than approximately 65%, it effloresces at 25°C; the anhydrous acid is formed at humidities below approximately 40%. At RH between approximately 65 and 75%, it sorbs insignificant amounts of moisture, but above this, substantial amounts are absorbed (Fig. 1) (11).

Figure 1 also includes the sorption curve of the anhydrate. The anhydrous form melts at 135°C during decomposition (12). At RH approaching 75%, the monohydrate is formed (11).

Information from Heckel plots indicates that anhydrous citric acid is predominantly fragmented during compression (13). The elastic deformation and consequently the elastic recovery during decompression are low (14).

Tartaric acid: Tartaric acid is soluble 1 in 0.75 parts of water, and 1 in 2.5 parts of alcohol (15). It sorbs insignificant amounts of moisture at RH up to approximately 65%, but at RH above approximately 75%, substantial amounts are absorbed (Fig. 1).

Studies indicate that tartaric acid behaves in a manner similar to that of anhydrous citric acid. During compression, the acid fragments predominantly, and the

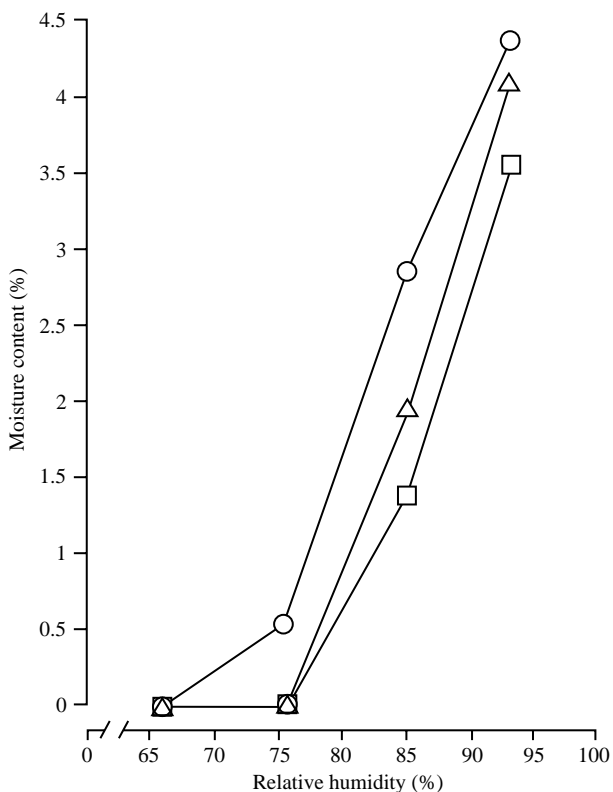


Fig. 1 Sorption isotherms of some hygroscopic acids. Key: x axis = relative humidity, %; y axis = moisture content, %; ○ = citric acid monohydrate; Δ = anhydrous citric acid; □ = tartaric acid. (Adapted from Ref. 16.)

elastic deformation and consequently the elastic recovery were low (14).

A comparison of the formation of carbon dioxide from effervescent tablets based on anhydrous citric acid, ascorbic acid or tartaric acid, and NaHCO_3 in stoichiometric proportions indicated that ascorbic acid and anhydrous citric acid behaved similarly. However, tartaric acid formed the most carbon dioxide, but the disintegration time was longer (16).

Ascorbic acid: Ascorbic acid occurs as white to light yellow crystalline powder or colorless crystals with a sharp, acidic taste and no odor. It is not hygroscopic. Upon exposure to light, it gradually darkens. Ascorbic acid is soluble 1 in 3.5 parts of water and 1 in 50 parts of ethanol (17).

Ascorbic acid particles show an intermediate fragmentation during compaction. The relatively low tablet strength indicates that the attraction forces are relatively weak and not very resistant to stress relaxation and elastic recovery (18).

Ascorbic acid can be used as the acid source. The speed of release of carbon dioxide from a mixture of ascorbic acid and

NaHCO_3 is comparable with that produced by citric or tartaric acid– NaHCO_3 combinations. Since ascorbic acid is less hygroscopic than citric and tartaric acid, using ascorbic acid as the only acid source makes it possible to produce effervescent tablets in a nonairconditional area (19).

Fumaric acid: Fumaric acid is a white, odorless or nearly odorless crystalline powder. It is soluble 1 in 222 parts of water and 1 in 28 parts of ethanol (20). The sorption isotherm indicates that fumaric acid is not a hygroscopic substance (16).

Acetylsalicylic acid (aspirin): Although acetylsalicylic acid is a drug frequently used in effervescent form, it cannot be used as the acid source because of its low water solubility. Additional acid is necessary to decrease the reaction time.

Other acids: Malic acid is hygroscopic and readily soluble in water. It has been suggested for effervescent products (3).

Other acids have been mentioned in connection with effervescent products (3, 5).

Acid anhydrides: The use of acid anhydrides as the acid precursor has been investigated. However, their use in commercial products is limited.

Acid salts: Amino acid hydrochlorides readily release acid when in solution. However, these materials have the disadvantage of being expensive and rather hygroscopic (4). Other suggested acid sources include: sodium dihydrogen citrate (21), a nonhygroscopic substance (16); disodium hydrogen citrate, which is nonhygroscopic below approximately 93% RH/20°C (16); and sodium acid phosphate, which is very soluble in water.

Sources of Carbon Dioxide

Both carbonates and bicarbonates are used as carbonate sources, but the latter is most often used.

Sodium bicarbonate (NaHCO_3)

NaHCO_3 is an odorless, white crystalline powder with a saline, slightly alkaline taste. A variety of particle-size grades of powders and granules are available. The carbon dioxide yield is approximately 52% by weight. At RH below approximately 80% (at room temperature), the moisture content is less than 1%. Above 85% RH, it rapidly absorbs an excessive amount of water and may start to decompose. Its solubility in water is 1 part in 11 parts at 20°C, and it is practically insoluble in 95% ethanol at 20°C. When heated to 250–300°C, NaHCO_3 decomposes and is converted into anhydrous sodium carbonate. However, this process is both time- and temperature-dependent, commencing at about 50°C. The reaction proceeds via surface-controlled kinetics, and when NaHCO_3 crystals are heated for a short period of

time, very fine needle-shaped crystals of anhydrous sodium carbonate appear on the surface (22).

In humid air, there is a slow decarboxylation of NaHCO_3 , where as sodium sesquicarbonate $\text{Na}_2\text{CO}_3 \cdot \text{NaHCO}_3 \cdot 2\text{H}_2\text{O}$ is formed (23).

NaHCO_3 mainly consolidates by plastic deformation and not by fragmentation (18). It is a nonelastic substance (13).

In order to overcome the poor flowability and low compressibility of NaHCO_3 , a spray-drying technique was used. Additives such as polyvinylpyrrolidone and silicon oil were found to be essential to obtain direct compressible spray-dried NaHCO_3 . The product showed good compression characteristics without being transformed into sodium carbonate (24).

Sodium carbonate: Sodium carbonate is commercially available as an anhydrous form and as a monohydrate or a decahydrate. All forms are very soluble in water. The anhydrate is hygroscopic (25).

Potassium bicarbonate: Potassium bicarbonate (KHCO_3) is very soluble in water. When heated to approximately 200°C, it is decomposed, and potassium carbonate, water, and carbon dioxide are formed (26). Consequently, KHCO_3 is less sensitive to heat in connection with drying than is NaHCO_3 . Above approximately 80% RH at 20°C, substantial amounts of water are adsorbed by KHCO_3 .

Potassium carbonate: The moisture scavenging effect of potassium carbonate in effervescent tablets has been investigated (27).

Calcium carbonate: Precipitated calcium carbonate occurs as fine, white, odorless, and tasteless powder or crystals. It is practically insoluble in water and ethanol (95%). Precipitated calcium carbonate is nonhygroscopic (28). Calcium carbonate is a high-density, not very compressible material (29). It is known to consolidate by fragmentation (30).

Other sources: Amino acid–alkali metal carbonate derivatives, such as sodium glycine carbonate, have been suggested as sources of carbon dioxide.

Sodium glycine carbonate is a nonhygroscopic, heat-resistant, stable substance (31). However, the carbon dioxide yield—approximately 18% by weight—is only about one-third of NaHCO_3 .

PRODUCTS

Dosage Forms

Effervescent tablets (1, 6), granules, and powders (8) are mentioned in the pharmacopoeias and exist as products

on the market. The effervescent tablet provides several advantages over conventional oral solid dosage forms. It is administered as a reasonably palatable, sparkling solution. Consequently, it can be given to patients who have difficulties swallowing capsules or tablets. Since the drug is administered as a solution, problems associated with dissolution, that is, absorption rate and extent of bioavailability, are avoided. Drugs that are unstable when stored in aqueous solutions are more often stable in the effervescent tablet.

Effervescent dosage forms have several drawbacks when compared with aqueous solutions and plain tablets. For example, they are relatively expensive to produce due to the use of large amounts of more or less expensive excipients and the necessary special production facilities, as well as high Na^+ and/or K^+ concentrations. In addition, when compared with plain tablets, effervescent tablets are bulky, even though small packages that are easy to carry in a pocket or handbag are available. Finally, it is sometimes difficult to make unpleasant tasting drugs sufficiently palatable in an effervescent form.

When an effervescent product is dropped into a glass of water, the reaction between the acid and the NaHCO_3 is quite rapid, usually completed within 1 minute or less (32). The effervescent reaction is also used in other pharmaceutical dosage forms than the traditional effervescent products. Effervescent laxative suppositories that release carbon dioxide have been thoroughly studied (33). One product has been on the Swedish market for many years. Effervescent vaginal suppositories are described (34). Pulsatile and gastric floating drug delivery systems for oral administration based on a reservoir system consisting of a drug-containing effervescent core and a polymeric coating also have been investigated (35).

Drugs (Product Categories)

Many drugs and drug compositions have been used for effervescent products. Some of these are listed below.

Acetylsalicylic acid (aspirin) is a common drug in many different effervescent products (36, 37).

Paracetamol (acetaminophen) is another analgesic used in effervescent preparations (38).

Effervescent compositions of ibuprofen, another analgesic, are marketed.

Among effervescent antacid preparations, Alka-Seltzer, an effervescent antacid analgesic product, has been available since the 1930s. Pure effervescent antacid products are marketed in many countries.

Effervescent tablets of ascorbic acid, 0.5–1 g, are well known. Other vitamins as well as calcium and some minerals have also been included.

Acetylcysteine, a mucolytic agent that also is used as an antidote for paracetamol overdose, is available as an effervescent tablet.

Effervescent products of water-insoluble drugs have been manufactured. A successful example is the effervescent activated charcoal preparation suggested in the management of theophylline poisoning (39).

Electrolyte Balance Considerations

Effervescent tablets normally have a high sodium content. In most of the effervescent analgesic products in Sweden, the sodium content is approximately 15 mmol. This sodium content may be contraindicated in some patients (e.g., in patients with active sodium-retaining status such as congestive heart failure or renal insufficiency). Otherwise, there are no restrictions concerning the sodium content of effervescent tablets.

Biopharmaceutical Aspects

Drugs are most rapidly absorbed from the gastrointestinal (GI) tract when administered as aqueous solutions. Although dilution of the drug solution in the gastric fluids sometimes results in precipitation, the extremely fine nature of the precipitate permits rapid redissolution (40). The rapid absorption of the aqueous solution is the idea behind effervescent analgesic products, for example. Furthermore, consistent absorption is expected with the solution, as disintegration and dissolution in the GI tract are bypassed.

Effervescence may produce physiological changes within the body. Carbon dioxide bubbling directly onto the intestinal epithelium induced enhanced drug permeability due to an alteration of the paracellular pathway. This, in addition to fluid flow and membrane hydrophobicity concepts, may account for observed increases in drug flux (41).

Buffered effervescent aspirin tablets are generally believed to have a less irritant effect on the gastric mucosa and cause less GI blood loss than conventional tablets. This view has been questioned.

The bioavailability of acetylsalicylic acid from three different dosage forms—two types of effervescent tablets with different buffering properties and tablets of a conventional type—was studied in healthy volunteers. Complete absorption was found for all the preparations

studied. Both effervescent tablets were rapidly absorbed. The buffering properties did not influence the rate of absorption (36).

Effervescent aspirin, soluble aspirin, and soluble aspirin to which sufficient NaHCO_3 had been added to give it the same buffering capacity as the effervescent preparation, were compared in healthy volunteers. There were no significant differences in plasma salicylate levels at any time after taking these preparations (37).

The absorption of the effervescent formulation of paracetamol was compared with that of a plain tablet in normal volunteers. As to the rate of absorption, this was more rapid and consistent from the effervescent preparation than from the plain tablet. This may have important therapeutic implications where a rapid and predictable analgesic effect may be desired (38).

The bioavailability of an effervescent ibuprofen tablet was compared to a sugar-coated tablet. Ibuprofen was absorbed more rapidly from the effervescent tablet but both formulations were bioequivalent in respect to peak plasma concentrations and area under the plasma concentration curves (42).

PROCESSING

Environment

The manufacturing of effervescent tablets requires careful control of environmental factors. As early as the 1930s, it was clear that it was essential to maintain RH throughout the plant of no more than 20%. In addition, a uniform temperature of 21°C also was desirable (2).

A maximum of 25% RH at a controlled room temperature of 25°C or less is usually sufficient to avoid problems caused by atmospheric moisture (3).

Equipment

Conventional processing equipment (mixers, granulators, roller compactors, drying equipment, and mills) can be used to produce effervescent preparations if the influence of atmospheric moisture is considered. As a rule, tablet presses have to be adapted to handle effervescent products, except for tablets with a sufficient proportion of a self-lubricating substance, such as acetylsalicylic acid.

Wet Granulation Methods

The acid and carbonate parts of the effervescent formulation can be granulated either separately or as a

mixture with water (crystal water of citric acid, liquid water, or water vapor), ethanol (possibly diluted with water), isopropanol, or other solvents.

When granulating with solvents without any moisture, no effervescent reaction will occur provided the raw materials are dry and the process is performed in a low humidity atmosphere. However, citric acid will partly dissolve in ethanol or isopropanol, and function as a binder when the solvent is evaporated.

When granulating either with solvents containing water or pure water, the effervescent reaction will start. Care must be taken to maintain adequate control of the process. Vacuum processing is often beneficial due to the ability to control the effervescent reaction and the drying process.

In the fusion method of granulation, the effervescent mixture is heated to approximately 100°C (the melting point of the monohydrate) so that the water of crystallization from hydrous citric acid is released. This process is sporadic and difficult to control, especially in a static bed (3).

By means of high-shear mixers and the heat generated during mixing, it was possible to prepare granular effervescent products in batch sizes of 60–300 kg using the fusion method (43).

Citric acid is moistened and added to the NaHCO_3 . Partial wet fusion occurs, and granules are formed by kneading in a suitable mixer. The granules are tableted while still damp, with the moist citric acid acting as a lubricant. The compressed tablets are transferred immediately and continuously to ovens where they are dried at 70–75°C. Drying also hardens them. As soon as they leave the dryer, the tablets are packed in aluminum foil lined with polyethylene (44).

X-ray diffractometry and infrared (IR) spectrophotometry were used to study the reaction between citric acid and NaHCO_3 when granulating the mixture with water in a high-shear mixer and vacuum drying the wet mass. The contact time before drying varied as did the water content. At low water levels, varying the contact times did not change the citric acid. However, with higher levels of water content, the presence of monocitrates, dicitrates, and tricitrates was verified. The loss of carbon dioxide during granulation occurred in the presence of, especially, dicitrates and tricitrates (45).

Effervescent granules were prepared in a fluid bed granulator/dryer (46).

The drug can be mixed with the effervescent granulate and other excipients or be a part of the granulation. When mixing low proportions of drug with granulate, the risk of segregation must be taken into account.

Dry Granulation

Granulation by slugging (slugs or large tablets that are compressed using heavy-duty tableting equipment) or roller compaction is suitable for materials that cannot be wet granulated. The slugs and the material from the roller compactor are reduced to the proper size. Lubrication is often necessary during slugging but not always with roller compaction. The acidic and basic components may be dry granulated separately or together.

Direct Compression

Some effervescent tablet products are successfully produced by direct compression (e.g., acetylsalicylic acid products). Direct compression normally requires careful selection of raw materials to achieve a free-flowing, nonsegregating, compressible mixture. Effervescent products present the same problems as conventional products in direct compression.

Tableting

The adaptation of a single-punch tablet press for compressing effervescent tablets via external lubrication has been described (5, 47). Only rotary presses are normally used in connection with the commercial production of effervescent tablets. Tablet machine manufacturers have applied various adaptations to their existing equipment to avoid problems due to internal lubrication and punch adhesion. Consequently, many effervescent tablets are produced on rotary presses with external lubrication. Liquid or solid lubricants can be used.

FORMULATION

Excipients (Including Sweeteners and Flavors)

Lubricants

A perfect lubricant (or auxiliary agent, in general) for effervescent products must be nontoxic, tasteless, and water-soluble. Very few traditional lubricants fulfill these requirements.

Intrinsic lubricants are added to the powder mixture and consequently included in the formulation. When added in solid form, the lubricant will have to be finely divided.

Metal stearates, such as magnesium or calcium stearate that serve as lubricants in conventional tablets, are seldom used as intrinsic lubricants in connection with

effervescent tablets due to their insolubility in water. Use of stearates results in an undissolved, foamy, soapy-tasting layer on the surface of the cloudy solution. In addition, normal lubricant concentrations of metal stearates make the tablets hydrophobic, which entails a slow dissolution of the effervescent tablet in the water. However, very low concentrations of metal stearates can be used to improve the rate of solution of effervescent tablets as the tablet will remain immersed in the water during dissolution and not float to the surface the way a tablet without metal stearate would. A floating tablet presents a smaller surface area to the water than a tablet immersed in the liquid.

Sodium stearate and sodium oleate are water-soluble in low concentrations. They have the characteristic soapy taste, which virtually precludes their use in effervescent products.

A combination of 4% polyethylene glycol (PEG) 6000 and 0.1% sodium stearyl fumarate proved to be a good lubricant for ascorbic acid tablets made by direct compression on a small scale (48). Sodium chloride, sodium acetate, and D,L-leucine (water-soluble lubricants) also have been suggested for effervescent tablets (44).

Twenty lubricants for effervescent tablets were tested for lubrication efficiency in direct compression of a standard effervescent formulation. The lubricant concentration was high as compared to traditional tablet lubricants. By increasing the lubricant concentration and the compression force, most lubricants became more effective. The lubricant used in effervescent formulations should combine hydrophobic and hydrophilic properties in order to achieve both good lubrication and a short disintegration time. A medium polar lubricant was the best compromise. Fumaric acid was chosen and its concentration optimized (49). Other research that studied the lubrication of effervescent products indicated optimal concentrations of spray-dried L-leucine and PEG 6000 at levels of 2 and 3%, respectively (50).

Surfactants such as sodium lauryl sulfate and magnesium lauryl sulfate also act as lubricants.

Extrinsic lubrication is provided via mechanisms that apply a lubricating substance, normally paraffin oil, to the tableting tool surface during processing. One method makes use of an oiled felt washer attached to the lower punch below the tip. This washer wipes the die cavity with each tablet ejection. To avoid having tablets stick to the punch faces, materials such as polytetrafluorethylene or polyurethane have been applied to the faces. Another lubrication method sprays a thin layer of lubricant (either liquid or solid lubricant) onto the tool surfaces after one tablet is ejected and before the granulate of the next tablet enters the die cavity.

Products containing acetylsalicylic acid do not usually require additional lubrication.

Glidants

Glidants are usually not necessary. Free-flowing granules, ingredients of appropriate physical form for direct compression, and the large tablet diameters make it possible to exclude the use of glidants.

Antiadherents

The adherence of the granulate or powder mixture to the punch surfaces, so-called picking, can be eliminated by using discs, such as polytetrafluorethylene or polyurethane, cemented to the punch surfaces.

Binders

Binders are commonly used when making conventional tablets. The binders are either added in dry form or dissolved in a suitable solvent and then added in connection with a wet-granulation process. Most binders are polymers and increase the plastic deformation of the formulation.

The use of binders will normally prevent a rapid dissolution of the effervescent tablet. Therefore, many effervescent tablets are formulated without any binder. However, effervescent granules may be formulated with binders since their large surface area, when compared with that of the conventional or the effervescent tablet, will result in rapid dissolution. An effervescent granulation composed of anhydrous citric acid and NaHCO_3 was made with dehydrated alcohol as the granulating liquid. A portion of the citric acid dissolved during the massing and functioned as a binder (51).

In order to compress ascorbic acid from a combination with NaHCO_3 , granulation was required. Common water-soluble binders, such as polyvinylpyrrolidone (polyvidone) or polyvinylpyrrolidone–poly(vinyl acetate)-copolymer, led to a change of color on the part of the ascorbic acid granules. Hydrogenated maltodextrins containing high amounts of maltitol were chosen from a wide range of dextrins and maltodextrins as possible binders. Maltitol was a suitable binder for ascorbic acid effervescent tablets. Formation of crystal bridges of maltitol was the assumed binding mechanism (19). PEG 6000 functions both as a binder and as a lubricant.

Disintegrants or dissolution aids

Disintegrants, which are used in conventional tablets, are not normally used in effervescent tablets because one of the marketing demands is that a clear solution should be obtained within a few minutes after adding the tablet to a glass of cold water.

Diluents

Effervescent products generally do not require diluents. The effervescent materials themselves will have to be added in large quantities.

Sweeteners

Sucrose and other natural sweeteners, such as sorbitol, can be used in effervescent products, although artificial sweetening agents are customary. However, the application of artificial sweeteners is restricted by health regulations. Therefore, the use of such sweeteners will vary from one country to the next based on national standards.

Saccharin or its sodium and calcium salts are used as sweeteners. Aspartame is also employed as a sweetener in effervescent tablets. Earlier, cyclamates and cyclamic acid were the artificial sweeteners of choice, but their use has now been restricted.

Flavors

The simple use of sweetening agents may not be sufficient to render palatable a product containing a drug with an unpleasant taste. Therefore, a flavoring agent can be included. Various dry flavors are available from suppliers. The flavors used must be water-soluble or water-dispersible.

Colors

Water-soluble colors may be added; however, some dyes change color according to pH variations, a consideration that must be noted before a dye is selected.

Surfactants

This type of excipient is sometimes used to increase the wetting and dissolution rate of drugs. Attention must be paid to the formation of foam.

Antifoaming agents

To reduce the formation of foam, and consequently the tendency of drugs to stick to the wall of the glass above the water level, an antifoaming agent, such as polydimethylsiloxane, can be used. However, antifoaming agents do not normally form constituents of effervescent products.

Formulations (Including Optimization)

Literature on formulations of effervescent products is relatively sparse. Table 1 presents some examples of effervescent products on the Nordic market.

A fractional factorial design was employed in the preparation of effervescent aspirin tablets. The optimum conditions for preparing the tablets were determined following the path of steepest ascent (53).

Table 1 Some compositions of effervescent tablets on the Nordic market: Components and weight per tablet

	Product A		Product B		Product C	
	Component	mg	Component	mg	Component	mg
Drugs	Ascorbic acid	1000	Acetylsalicylic acid	500	Paracetamol	500
Excipients			Caffeine	50		
	Citric acid, anhydrous	700	Citric acid, anhydrous	500	Citric acid, anhydrous	1200
	Sodium bicarbonate	490	Sodium bicarbonate	1250	Sodium bicarbonate	1550
	Polyethylene glycol 6000	45	Docusate sodium	0.85	Polyvidone	25
			Sodium benzoate	0.15	Sodium cyclamate	45
	Sorbitol	25			Saccharin sodium	5
	Saccharin sodium	12			Lemon flavor	25
	Riboflavin sodium phosphate (for color)	1			Magnesium stearate	1.4
	Orange flavor	2				

(Adapted from Ref. 52.)

An experiment investigating the effects of tablet manufacturing conditions, tablet formulations, tablet compression pressures, storage conditions, and storage times was performed on five different formulations (54). The effects of two formulation factors (the ratio of citric acid/ NaHCO_3 and the polyvidone content) and two process factors (the temperature and the velocity of the fluidizing air) on granule size, powder content, and dissolution rate of the tablets were studied using factorial design. In addition, the levels of the significant factors were optimized with the path of steepest ascent (46).

Solid dispersions of poorly water-soluble drugs were made by the fusion method. Citric acid was employed in various ratios with NaHCO_3 as the carrier for these drugs (55).

Stability

The greatest problem with effervescent products is the loss of reactivity with time if exposed prematurely to moisture (i.e., the stability of the effervescent system). In addition, the stability of the drug and some excipients, such as flavors, also must be considered.

Effervescent products are not stable in the presence of moisture. Most effervescent products are hygroscopic and can therefore adsorb enough moisture to initiate degradation if they are not suitably packaged.

Tablets made with equivalent amounts of NaHCO_3 and tartaric acid were stored at 70°C . In a closed system, a reaction between the NaHCO_3 and the tartaric acid occurred. When the tablets were stored as an open system,

the weight loss was concluded to be a decarboxylation of the NaHCO_3 (56).

Effervescent compositions may be markedly stabilized if the NaHCO_3 is partly converted to the corresponding carbonate. Usually, the desired degree of stability is attained if approximately 2–10% of the weight of the bicarbonate is converted to the carbonate (57). The addition of sodium carbonate did not by itself improve stability. One explanation for the stabilizing effect caused by heating of the bicarbonate could be that heating causes a uniform distribution of the carbonate on the surface of the bicarbonate so that the water-scavenging efficiency is greater. Another explanation is that the carbonate formed by the rupture of the bicarbonate crystals would be much finer than added crystalline sodium carbonate, however finely ground. A third explanation is the possibility that double salts might be present and that they could be better scavengers than the carbonate itself (56). The moisture scavenging effect of potassium carbonate was determined and the concentration optimized for a specific formulation (27).

The stability of three commercial effervescent and one dispersible aspirin tablet were evaluated by factorially designed experiments. Temperature affected the hydrolysis of all tablets, whereas humidity influenced one product in a plastic tube and one in an aluminum tube (58).

Mercury-intrusion porosimetry and a cantilever beam-proximity transducer balance were used to monitor the stability of selected effervescent tablet systems. An index of reactivity was obtained from the balance measurements. The porosity measurements proved to be useful in elucidating tablet-pore structure changes over time.

Compression pressure and manufacturing conditions were not significant factors in the stability of an effervescent system when nonhygroscopic materials were used (54).

Codeine phosphate in a paracetamol-codeine effervescent tablet was found to react at room temperature with the citric acid constituents to form citrate esters of codeine. The esterification was confirmed in a solid-state reaction at an elevated temperature. Tartaric acid also yielded an ester with codeine phosphate in a similar nonsolvolytic reaction (59).

PRODUCTION

Granulation

At the Pharmacia plant in Helsingborg, Sweden, approximately 1200 kg of effervescent granulate is produced daily. Anhydrous citric acid and NaHCO_3 are massed with ethanol in a planetary mixer and the wet mass is dried on trays. Additional effervescent granulates are produced with vacuum equipment (Topo granulator) where water is the main component of the granulation liquid. The Topo granulator, developed for preparation of granules and coated particles in a vacuum, handles the mixing, granulation, drying, and milling/sieving as a closed system.

The fusion method, which employs heat to liberate water of crystallization from hydrous citric acid in order to effect moistening, was applied by using a high-shear mixer to generate heat (43). Batch sizes of 60 and 300 kg were granulated.

Anhydrous citric acid and NaHCO_3 were granulated with ethanol in a twin-screw extruder at powder flow rates of 60–90 kg/h in a continuous process (51).

The air suspension coating–reacting technique also is used in the production of effervescent granulates.

Tableting

Effervescent tablets are normally produced by machines with external lubrication systems. Most tablet machine manufacturers can add this type of equipment to their rotary machines. Products with a high proportion of acetylsalicylic acid can be manufactured without any traditional lubricants. Consequently, conventional rotary tablet presses can be used. Effervescent acetylsalicylic acid tablets are produced on ordinary high-speed rotary presses at the Pharmacia plant in Helsingborg, Sweden.

Effervescent granules can be tableted while still damp since moist citric acid acts as a lubricant. The compressed

tablets are transferred immediately and continuously to ovens where they are dried. Drying also hardens them (44).

Several types of steel are normally used in the manufacture of compression tooling. Material rich in nickel was found to have the best resistance to rusting induced by a hydrochloride salt, although other factors, such as humidity, temperature, and contact time, also were responsible for the rusting of tooling material (60). This information may be useful when ordering and managing tooling materials for effervescent tablets.

The compression of effervescent mixtures usually results in severe picking and sticking. By means of flat-faced punches with discs of polytetrafluorethylene, the sticking to tablet-punch surfaces is overcome (61). Other nonadherent materials, such as Vulkollan® (a polyethane), Hostalit® (polyvinyl chloride), and Resopal® (a melamine), have been used (62). The disc of the plastic material is attached to the recess of the punch surface by glue or adhesive tape. It should be noted that fragments of the polymer can rub off during compression.

Effervescent tablets were produced using four different formulations that contained citric and/or tartaric acid and NaHCO_3 with polyvidone and PEG 6000. The adhesion of each formulation to the metal faces of the punch tips was determined by means of electron microscopy, surface-roughness measurements, and quantification of punch-weight variations during tablet production. The basic formulations were inherently adhesive and produced tablets with a weak, porous structure; the tablets were rougher than conventional, noneffervescent compressed tablets. Both formulations that contained tartaric acid produced tablets with a lower surface roughness and had less of a tendency to stick to tablet-punch faces than the two formulations that contained citric acid alone. The addition of a water-soluble sucrose ester had a beneficial effect, especially on formulations with inherently high adhesive tendencies (63).

In-Process Quality Control

For a rapid determination of loss on drying, an IR drying balance may be used. In the matter of size distribution, effervescent granulations are controlled by sieve analysis.

During the compression of effervescent tablets, in-process tests are routinely run to monitor the process. These tests include controls of tablet weight, weight variation, thickness, crushing strength, disintegration, and appearance of the tablet. Friability and pH of the solution may be additionally tested. Electronic devices that monitor tablet weight are normally used.

Inspection of the punches is carried out during the manufacturing of the tablets when plastic insertions are used. Inspections ensure that the plastic insertions are intact, i.e., that no loss or damage to the discs has occurred.

Product Evaluations

Both chemical and physical properties have to be considered when evaluating effervescent products. In this review, only the physical properties will be discussed, except where the chemical characteristics are especially influenced by the effervescent base. For more detail, Ph. Eur. includes a special disintegration test for effervescent tablets (1) and granules (8).

Many tests (e.g., titrimetric, gravimetric, colorimetric, and volumetric tests as well as loss-of-weight measurements and pressure measurements) have been proposed in order to determine carbon dioxide content (16, 48, 64). Methods based on monitoring carbon dioxide pressure generation and weight loss have been applied (16, 65).

Results from weight-loss measurements were modeled (65). Research indicates that the determination of water content by Karl Fischer analysis in effervescent tablets was possible after extraction with dioxane (66). NaHCO_3 , which reacts with the Karl Fischer reagent, is insoluble in dioxane and does not interfere during the determination.

Near IR (NIR) is a quick and nondestructive method for the determination of water in effervescent products. In addition, it is suitable for in-process quality control. Measurement of pH of the solution is often performed. The conditions are important for congruent results.

Tablets

The disintegration and dissolution times are very important characteristics of effervescent products. A well-formulated effervescent tablet will disintegrate and dissolve within 1–2 min to form a clear solution. Consequently, the residue of undissolved drug must be minimal. The temperature of the water influences the dissolution time. It is, therefore, important to choose a water temperature that is actually used by consumers (e.g., cold tap water). Ph. Eur. includes a general requirement on disintegration time of 5 min in water 15–25°C (1).

Factors such as crushing strength and friability will influence the possibility of packaging the tablets on packaging lines, as effervescent tablets chip easily at

the edges during handling. When the tablets are filled in tubes, the tablet height is of the utmost importance since the looseness or tightness of the packaging depends on the tablet height. When small or fairly small amounts of drug form part of the formulation, it is essential that content uniformity be carefully supervised.

Powders and Granules

Disintegration and dissolution time is an important characteristic, as is powder weight variation. The Ph. Eur. requirement time for disintegration of granules is 5 min (8).

Production Area

As the mass of an effervescent tablet is, as a rule, many times larger than that of a conventional tablet, larger amounts of raw material will have to be handled when packaging the same number of tablets. Therefore, the production area will be larger, too, unless a compact continuous line has been constructed.

At the Pharmacia plant in Helsingborg, Sweden, all steps during the production of effervescent tablets (i.e., mixing, granulating, drying, milling, final mixing of granulate and other constituents, tableting, and packaging) are performed in dehumidified areas of <25% RH and <25°C. Other companies perform mixing, granulating, drying, and milling at normal humidities but store the final mixture in dehumidified areas while slowly bubbling dehumidified air through the mixture. The mixture is then tableted and packaged in a small, dehumidified area around the tablet and packaging machine.

In direct compression, the mixing can be performed at normal humidities; however, in that case, the mixture is dried (to prevent a premature effervescent reaction) by means of causing dehumidified air to flow through the bed in a suitable container. Tableting and packaging are also performed in the dehumidified area. Thus, the number of manufacturing stages in the low humidity zone is reduced.

PACKAGING

Effervescent tablets should be stored in tightly closed containers or moisture-proof packs (6).

Even the moisture in the air may be enough to initiate the effervescent reaction of an effervescent product if it is not properly protected. When the consumer opens the

container, the effervescent product will again be exposed to the moisture in the air. Consequently, the packaging of all effervescent products is very important. The time between tablet production and start of packaging operation should be kept as short as possible.

Ph. Eur. recommends that effervescent granules and powders be stored in airtight containers (8). In the past, acidic and alkaline components were wrapped separately to prevent effervescent reactions during the storage of powders and granules.

Materials

Effervescent products are usually packed in individual aluminum foil pouches and effervescent tablets are often packed in metal tubes. To avoid excessive laminate stress, the dimensions of the sachets should be adapted to the dimensions of the tablet or the amount of granulate. These pouches are arranged in conveniently sized strips and stacked in a paperboard box.

The metal tube is a multiple-use container sealed with a moisture-proof closure. The tablets are stacked on top of one another. Consequently, a minimum of air surrounds them. The tubes are seamless, extruded aluminum packages. They are closed by tightly fitting plastic snap caps that contain a desiccant chamber. Tubes of plastic materials, such as polyvinyl chloride or polypropylene, have been tested with effervescent tablets. Acceptable stability was obtained with some of these products. Plastic tubes are used more often due to their lower cost and lower noise level during the packaging operation.

Aluminum-foil blisters can provide hermetic packs. Similar protection can be achieved by using a foil-bearing laminate or a strip pack. A special strip pack for effervescent tablets, where each tablet is connected to a desiccant via a channel, has been suggested (67).

The effect of environmental moisture on the physical stability of effervescent tablets in foil-laminate packages containing microscopic imperfections was examined. Physical stability, after storing at different RH and temperature conditions, was assessed by noting whether the tablet components reacted prematurely. A penetrating dye-solution test was used in order to determine whether the foil packages permitted any transmission of moisture. High humidity accelerated the physical deterioration of effervescent tablets when stored in packets of poor integrity (68).

Filling

Packaging operations must be conducted in a low humidity environment if the long-term stability of the product is to

be maintained. The tablets must be hard enough so as not to break during packaging.

Quality Control

Individual foil packets are tested for proper sealing. Several methods of rapid seal integrity testing have been devised, such as the vacuum underwater method, detection of tracer material sealed within the pouch, purging with detectable gas, IR seal inspection, and electronic air-tightness testers (3).

CONCLUSION

The traditional effervescent product is dissolved prior to oral intake. This requires the drug to have an acceptable taste. Since the drug is given as a solution, the absorption is normally rapid and the bioavailability is usually good.

The commercial manufacturing of effervescent products involves controlling air humidity in the production area. Special tablet machines are generally required, and the package is a very important part of the effervescent product. Over-the-counter analgesics have been very successful as effervescent tablets on certain markets.

ACKNOWLEDGMENT

We are indebted to Ms. Margareta Duberg, Diabact AB, Uppsala for her valuable contribution.

REFERENCES

1. *European Pharmacopoeia*, 3rd Ed.; Council of Europe: Strasbourg, France, 1999; 1347, Suppl. 2000.
2. Chilson, F. Effervescent Tablets. *Drug Cosm. Ind.* **1936**, 39, 738–740.
3. Mohrle, R. Effervescent Tablets. *Pharmaceutical Dosage Forms*, 2nd Ed.; Lieberman, H.A., Lachman, L., Schwartz, J.B., Eds.; Marcel Dekker, Inc.: New York, 1989; 1, 285–328.
4. Sendall, F.E.J.; Staniforth, J.N.; Rees, J.E.; Leatham, M.J. Effervescent Tablets. *Pharm. J.* **1983**, 230, 289–294.
5. Schmidt, P.C.; Christin, I. Brausetabletten—Eine Fast Vergessene Arzneiform. *Pharmazie* **1990**, 45, 89–101.
6. *The United States Pharmacopeia USP 24*, United States Pharmacopeial Convention: Rockville, MD, 1999, 2117.
7. *The United States Pharmacopeia USP 24*; United States Pharmacopeial Convention: Rockville, MD, 1999; 18, 167, 1354, 1355, 1356, 1360.

8. *European Pharmacopoeia*, 3rd Ed.; Council of Europe: Strasbourg, France, 1996; 1749, 1752.
9. Higuchi, T.; Perrin, J.; Robinson, J.; Repta, A.J. Carbonation of Aqueous Solutions with Acid Anhydrides. *J. Pharm. Sci.* **1965**, *54*, 1273–1276.
10. Nyström, C.; Karehill, P.-G. The Importance of Inter-molecular Bonding Forces and the Concept of Bonding Surface Area. *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nyström, C., Eds.; Marcel Dekker, Inc.: New York, 1996; 17–53.
11. Wade, A., Weller, P.J., Eds. *Handbook of Pharmaceutical Excipients*, 2nd Ed.; The American Pharmaceutical Association and the Pharmaceutical Society of Great Britain: Washington and London, 1994; 123–125.
12. Kommentar, Böhme, H., Hartke, K., Eds.; *Acidum citricum. Europäisches Arzneibuch*; Wissenschaftliche Verlagsgesellschaft mbH: Stuttgart, and Govi-Verlag GmbH: Frankfurt, 1979; III, 215–219.
13. Duberg, M.; Nyström, C. Studies on Direct Compression of Tablets. XVII. Porosity—Pressure Curves for the Characterization of Volume Reduction Mechanisms in Powder Pocompression. *Powder Technol.* **1986**, *46*, 67–75.
14. Duberg, M.; Diabact, A.B. Personal Communications. Box 303, S-751 05, Uppsala, Sweden.
15. Wade, A., Weller, P.J., Eds.; *Handbook of Pharmaceutical Excipients*, 2nd Ed.; The American Pharmaceutical Association and the Pharmaceutical Society of Great Britain: Washington and London, 1994; 522–523.
16. Schmidt, P.C.; Brögmann, B. Brausetabletten. Welche Säurekomponenten Sind Die Geeigneten. *Dtsch. Apoth.-Ztg.* **1987**, *127*, 991–997.
17. Wade, A., Weller, P.J., Eds.; *Handbook of Pharmaceutical Excipients*, 2nd Ed.; The American Pharmaceutical Association and the Pharmaceutical Society of Great Britain: Washington and London, 1994; 15–18.
18. Nyström, C.; Glazer, M. Studies on Direct Compression of Tablets. XIII. The Effect of Some Dry Binders on the Tablet Strength of Compounds with Different Fragmentation Propensity. *Int. J. Pharm.* **1985**, *23*, 255–263.
19. Schmidt, P.C.; Brögmann, B. Effervescent Tablets: Choice of a New Binder for Ascorbic Acid. *Acta Pharm. Technol.* **1988**, *34*, 22–26.
20. Wade, A., Weller, P.J., Eds.; *Handbook of Pharmaceutical Excipients*, 2nd Ed.; The American Pharmaceutical Association and the Pharmaceutical Society of Great Britain: Washington and London, 1994; 197–198.
21. Rothe W.; Groppenbacher G.; Heinemann H. Brausetabletten, Brausepulver und Verfahren zur Herstellung Derselben. German Patent Application No. 1,962,791.
22. Wade, A., Weller, P.J., Eds. *Handbook of Pharmaceutical Excipients*, 2nd Ed.; The American Pharmaceutical Association and the Pharmaceutical Society of Great Britain: Washington and London, 1994; 436–438.
23. Neumüller, O.A. *Römpps Chemie-Lexikon*, 7th Ed.; Franckh'sche Verlagshandlung: Stuttgart, 1974; 2277.
24. Saleh, S.I.; Boymond, C.; Stamm, A. Preparation of Direct Compressible Effervescent Components: Spray-Dried Sodium Bicarbonate. *Int. J. Pharm.* **1988**, *45*, 19–26.
25. Parfitt, K., Ed. *Martindale: The Complete Drug Reference*, 32nd Ed.; The Pharmaceutical Press: London, 1999; 1630.
26. Neumüller, O.A. *Römpps Chemie-Lexikon*, 7th Ed.; Franckh'sche Verlagshandlung: Stuttgart, 1974; 1697.
27. Wells, M.L.; Wood, D.L.; Sanftleben, R.; Shaw, K.; Hottovy, J.; Weber, T.; Geoffroy, J.-M.; Alkire, T.G.; Emptage, M.R.; Sarabia, R. Potassium Carbonate as a Desiccant in Effervescent Tablets. *Int. J. Pharm.* **1997**, *152*, 227–235.
28. Wade, A., Weller, P.J., Eds.; *Handbook of Pharmaceutical Excipients*, 2nd Ed.; The American Pharmaceutical Association and the Pharmaceutical Society of Great Britain: Washington and London, 1994; 52–55.
29. Ejifor, O.; Esezobo, S.; Pilpel, N. The Plasto-Elasticity and Compressibility of Coated Powders and the Tensile Strengths of their Tablets. *J. Pharm. Pharmacol.* **1986**, *38*, 1–7.
30. Roberts, R.J.; Rowe, R.C. The Effect of Punch Velocity on the Compaction of a Variety of Materials. *J. Pharm. Pharmacol.* **1985**, *37*, 377–384.
31. Fiedler, H.P. *Lexikon der Hilfsstoffe für Pharmazie, Kosmetik und Angrenzende Gebiete*, 2nd Ed.; Editio Cantor: Aulendorf, Germany, 1981; 647.
32. Banker, G.S.; Anderson, N.R. *Tablets: The Theory and Practice of Industrial Pharmacy*, 3rd Ed.; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea & Febiger: Philadelphia, 1986; 293–345.
33. Hakata, T.; Ijima, M.; Kimura, S.; Sato, H.; Watanabe, Y.; Matsumoto, M. Effects of Bases and Additives on Release of Carbon Dioxide from Effervescent Suppositories. *Chem. Pharm. Bull.* **1993**, *41*, 351–356.
34. Kurobe, T.; Kasai, M.; Kayano M. Stable Effervescent Vaginal Suppositories, US Patent 4,853,211.
35. Krögel, I.; Bodmeier, R. Floating or Pulsatile Drug Delivery Systems Based on Coated Effervescent Cores. *Int. J. Pharm.* **1999**, *187*, 175–184.
36. Ekenved, G.; Elofsson, R.; Sölvell, L. Bioavailability Studies on a Buffered Acetylsalicylic Acid Preparation. *Acta Pharm. Suec.* **1975**, *12*, 323–332.
37. Orton, K.; Treharne; Jones, R.; Kaspi, T.; Richardson, R. Plasma Salicylate Levels after Soluble and Effervescent Aspirin. *Br. J. Clin. Pharmacol.* **1979**, *7*, 410–412.
38. Hedges, A.; Clive, M.K.; Maclay, W.P.; Turner, P. A Comparison of the Absorption of Effervescent Preparations of Paracetamol and Penicillin V (Phenoxymethylpenicillin) with Solid Dosage Forms of these Drugs. *J. Clin. Pharmacol.* **1974**, *14*, 363–368.
39. Helliwell, M.; Berry, D. Theophylline Absorption by Effervescent Activated Charcoal (Medicoal®). *J. Int. Med. Res.* **1981**, *9*, 222–225.
40. Gibaldi, M. *Biopharmaceutics. The Theory and Practice of Industrial Pharmacy*, 2nd Ed.; Lachman, L., Lieberman, H.A., Kanig, J.L. Eds.; Lea & Febiger: Philadelphia, 1976; 78–140.
41. Eichman, J.D.; Robinson, J.R. Mechanistic Studies on Effervescent-Induced Permeability Enhancement. *Pharm. Res.* **1998**, *15*, 925–930.
42. Altomare, E.; Wendemiale, G.; Benvenuti, C.; Andreatta, P. Bioavailability of a New Effervescent Tablet of Ibuprofen in Healthy Volunteers. *Eur. J. Clin. Pharmacol.* **1997**, *52*, 505–506.
43. Murray, R.B. New Approach to the Fusion Method for Preparing Granular Effervescent Products. *J. Pharm. Sci.* **1968**, *57*, 1776–1779.
44. Günsel, W.C.; Kanig, J.L. *Tablets. The Theory and Practice of Industrial Pharmacy*, 2nd Ed.; Lachman, L.,

- Lieberman, H.A., Kanig, J.L. Eds.; Lea & Febiger: Philadelphia, 1976; 321–358.
45. Joachim, J.; Kalantzis, G.; Jacob, M.; Mention, J.; Maury, L.; Rambaud, J. Technologie d'un Granule Effervescent. *Etudes Physico-Chimiques. J. Pharm. Belg.* **1986**, *41*, 197–208.
 46. Devay, A.; Uderszky, J.; Racz, I. Optimization of Operational Parameters in Fluidized Bed Granulation of Effervescent Pharmaceutical Preparations. *Acta Pharm. Technol.* **1984**, *30*, 239–242.
 47. Lindberg, N.-O. Preparation of Effervescent Tablets Containing Nicotinic Acid and Sodium Bicarbonate. *Acta Pharm. Suec.* **1970**, *7*, 23–28.
 48. Saleh, S.I.; Aboutaleb, A.; Kassem, A.A.; Stamm, A. A Contribution to the Formulation of Effervescent Tablets by Direct Compression. *Labo-Pharma Probl. Tech.* **1984**, *32*, 763–766.
 49. Röscheisen, G.; Schmidt, P.C. The Combination of Factorial Design and Simplex Method in the Optimization of Lubricants for Effervescent Tablets. *Eur. J. Pharm. Sci.* **1995**, *41*, 302–308.
 50. Rothhäuser, B.; Kraus, G.; Schmidt, P.C. Optimization of an Effervescent Tablet Formulation Containing Spray Dried L-leucine and Polyethylene Glycol 6000 as Lubricants Using a Central Composite Design. *Sur. J. Pharm. Sci.* **1998**, *46*, 85–94.
 51. Lindberg, N.-O.; Tufvesson, C.; Olbjer, L. Extrusion of an Effervescent Granulation with a Twin Screw Xtruder, Baker Perkins MPF 50 D. *Drug Dev. Ind. Pharm.* **1987**, *13*, 1891–1913.
 52. SWEDIS, Medical Products Agency, Box 26, S-751 03. Uppsala, Sweden.
 53. El-Banna, H.M.; Minina, S.A. The Construction and Uses of Factorial Designs in the Preparation of Solid Dosage Forms. Part I: Effervescent Acetylosalicylic Acid Tablets. *Pharmazie* **1981**, *36*, 417–420.
 54. Anderson, N.R.; Banker, G.S.; Peck, G.E. Quantitative Evaluation of Pharmaceutical Effervescent Systems II: Stability Monitoring by Reactivity and Porosity Measurements. *J. Pharm. Sci.* **1982**, *71*, 7–13.
 55. Desai, S.R.; Allen, L.V.; Greenwood, R.B.; Stiles, M.L.; Parker, D. Effervescent Solid Dispersions of Prednisone, Griseofulvin and Primidone. *Drug Dev. Ind. Pharm.* **1989**, *15*, 671–689.
 56. Usui, F.; Carstensen, J.T. Interactions in the Solid State I: Interactions of Sodium Bicarbonate and Tartaric Acid under Compressed Conditions. *J. Pharm. Sci.* **1985**, *74*, 1293–1297.
 57. White, B. Stable Effervescent Compositions and Method of Preparing Same. US Patent 3,105,792.
 58. Bulut, P.; Özer, A.Y.; Sumnu, M.; Hincal, A.A. Evaluation of the Stability of Commercial Effervescent and Dispersible Aspirin Tablets by Factorial Analysis. *S.T.P. Pharm. Sci.* **1991**, *1*, 357–361.
 59. Silver, B.; Sundholm, G. Solid-State Esterification of Codeine Phosphate by the Acid Constituent of Effervescent Tablets. *J. Pharm. Sci.* **1987**, *76*, 53–55.
 60. Narurkar, A.N.; Purkaystha, A.R.; Sheen, P.-C. Effect of Various Factors on the Corrosion and Rusting of Tooling Material Used for Tablet Manufacturing. *Drug Dev. Ind. Pharm.* **1985**, *11*, 1487–1495.
 61. Siegel, S.; Hanus, E.J.; Carr, J.W. Polytetrafluorethylene Tipped Tablet Punches. *J. Pharm. Sci.* **1963**, *52*, 604–605.
 62. Stammberger, W. Entwicklung von Brausezubereitungen. *Acta Pharm. Technol.* **1975**, *21*, 177–184.
 63. Sendall, F.E.J.; Staniforth, J.N. A Study of Powder Adhesion to Metal Surfaces During Compression of Effervescent Pharmaceutical Tablets. *J. Pharm. Pharmacol.* **1986**, *38*, 489–493.
 64. Amela, J.; Salazar, R.; Cemeli, J. Methods for the Determination of the Carbon Dioxide Evolved from Effervescent Systems. *Drug Dev. Ind. Pharm.* **1993**, *19*, 1019–1036.
 65. Anderson, N.R.; Banker, G.S.; Peck, G.E. Quantitative Evaluation of Pharmaceutical Effervescent Systems I: Design of Testing Apparatus. *J. Pharm. Sci.* **1982**, *71*, 3–6.
 66. Spitz, H.D. Determination of Water in Aluminum Chlorohydrate and Effervescent Tablets by Karl Fischer analysis. *J. Pharm. Sci.* **1979**, *68*, 122–123.
 67. Ritschel, W.A. Die Tablette. *Grundlagen und Praxis Des Tablettierens, Granulierens und Dragierens*; Editio Cantor KG: Aulendorf Germany, 1966; 116.
 68. David, S.T.; Gallian, C.E. The Effect of Environmental Moisture and Temperature on the Physical Stability of Effervescent Tablets in Foil Laminate Package Containing Minute Imperfections. *Drug Dev. Ind. Pharm.* **1986**, *12*, 2541–2550.

FURTHER READING

- Mohrle, R. Effervescent Tablets. *Pharmaceutical Dosage Forms*, 2nd Ed.; Lieberman, H.-A., Lachman, L., Schwartz, J.B. Eds.; Marcel Dekker, Inc.: New York, 1989; 1.
- Schmidt, P.C.; Christin, I. Brausetabletten —eine fast vergessene Arzneiform. *Pharmazie* **1990**, *45*, 89–101.
- Sendall, F.E.-J.; Staniforth, J.-N.; Rees, J.-E.; Leatham, M.J. Effervescent Tablets. *Pharm. J.* **1983**, *230*, 289–294.

EXCIPIENTS FOR PHARMACEUTICAL DOSAGE FORMS

Patrick J. Crowley

Luigi G. Martini

GlaxoSmithKline, Harlow, United Kingdom

INTRODUCTION

Medicinal dosage forms, regardless of composition or mode of use, must meet the following requirements that underpin efficacy, safety, and quality:

1. Contain an accurate dose
2. Be convenient to take or administer
3. Provide the drug in a form for absorption or other delivery to the target
4. Retain quality throughout the shelf life and usage period
5. Be manufactured by a process that does not compromise performance and that is reproducible and economical.

Few if any active pharmaceutical ingredients have properties that allow incorporation in units that meet all these criteria. Therefore, it is necessary to add other materials to make good any shortfalls. Consequently, virtually every medicinal product is a combination of the drug substance and excipients. These are indispensable components of medicinal products and, in most cases comprise the greatest proportion of the dosage unit. It goes without saying that knowledge of the composition, function, and behavior of excipients is a prerequisite to the successful design, development and manufacture of pharmaceutical dosage forms.

The requirements listed above can be considered the prime reasons for including excipients in dosage forms since they relate directly to product performance. Issues such as regulatory acceptability, environmental effects and impact on cost of the product are also important selection criteria.

A single chapter cannot do justice to the richness and complexity of the possibilities and constraints associated with using excipients to transform a drug to a dosage form. Each topic merits a chapter, possibly a complete volume in its own right. This chapter provides a general overview of the issues involved in selecting and evaluating excipients. More detailed accounts of individual applications, performance, and associated issues may be found in the references.

ACCURACY OF DOSE

Where the active ingredient is very potent (i.e., dose is low), it may be necessary to disperse the drug in a “diluent” or bulking agent. Otherwise, quantities being filled into capsules or dies for tableting may be so low that normal filling and other process variations translate to excessive variation in unit drug content. Likewise, low-dose medications for inhalation as dry powders may have the drug dispersed in or otherwise associated with an inert “carrier” or flow aid. For a diluent to function in this way it must form a homogenous blend with the drug. Otherwise accuracy of dose cannot be guaranteed (1).

Water may be considered a “diluent” in liquid presentations as it provides the required dose in a volume that can be accurately dispensed or administered. It is also invariably present in medications for topical or transdermal application. Water can be one of the most problematic companion materials in a dosage form because of its capability to promote hydrolysis, act as a vehicle for other molecular interactions, or simply be a medium for microbial growth. Such properties illustrate how a material that resolves one problem may pose others that in turn require the presence of additional excipients.

Liquid or semisolid preparations may require the presence of ancillary excipients to effect solvation or dispersion of the active ingredient. In particular, formulations containing drugs in the suspended state may require viscosity-enhancing agents or other additives to ensure that the drug remains homogeneously dispersed. Otherwise, the accuracy of the dose may be compromised.

USER OR PATIENT CONVENIENCE

Drugs that are bitter or otherwise unpalatable, and administered as oral liquids may be unacceptable, particularly to younger patients. Compliance and therefore efficacy may be compromised unless the product can be made more palatable. Thus, sweeteners, flavors, or

taste-masking agents may be present in liquid oral products, in chewable dosage forms, and in effervescent or dispersible tablets that are constituted as liquids prior to use (2).

Some drugs given by injection cause local pain due to high volume, tonicity, pH, etc. An additive that evinces a local anesthetic effect may relieve such discomfort. Benzyl alcohol is employed for this purpose.

RELEASE OF DRUG FROM THE DOSAGE FORM

Once a medication is ingested, applied to a target area, or otherwise administered, the drug must leave the dosage form for absorption or other delivery to the target. This may involve the following:

1. Dissolution in the gastrointestinal (GI) tract following oral dosage;
2. Partitioning to the skin in the case of topical or transdermal preparations;
3. Passage to pulmonary or nasal cavities (inhalation products).

Excipients can ensure that such delivery is expeditious and consistent. Their presence may be even more crucial with more esoteric forms that must be delivered to a tissue, organ, or even specific cells. Researchers are developing excipients that act as “homing devices” to guide drugs to designated targets. Such approaches will be discussed later in this chapter.

In its simplest form, designing “release” into a dosage form involves adding a disintegrant to the tablet or capsule formulation so that on ingestion the compact breaks up and drug is released for dissolution and absorption. In the case of hydrophobic drugs, dissolution may be aided by wetting agents. More complex release patterns involve using excipients to modify release from the dosage form to delay onset of action or otherwise modify the pharmacokinetics of the drug, thereby maximizing efficacy or minimizing side effects.

Excipients can influence delivery from topical and transdermal medications. The propensity of the drug to migrate from the formulation to the application surface is affected by factors such as lipophilicity of the vehicle, drug solubility in the formulation, and effects of additives on the barrier properties of the skin or mucosal surface.

ORAL ABSORPTION ENHANCEMENT

Oral absorption is indirectly aided by excipients that promote release of drug from the dosage form, or help

dispersion and dissolution prior to passage to the systemic circulation. Excipients that promote absorption per se are less widely used. However, lipids have been used to enhance absorption of hydrophobic active ingredients. Dissolution or dispersion of drug in such materials provides a substrate for lipolysis, resulting in an emulsion of drug and lipid that provides enhanced surface area for dissolution and absorption (3).

Lipids such as oleic acid or its salts are reported to slow gastric emptying and also act as an “ileal brake.” This allows longer time for dissolution and absorption in the small intestine (4, 5). Citric acid and other organic acids also have been shown to slow gastric emptying (6). However, the levels required for such effects may be impractical for most dosage forms.

The small intestine is drained by the hepatic portal vein, making the liver the first “port of call” for orally absorbed drugs. Therefore, high hepatic metabolism will compromise systemic availability. Formulation to enhance lymphatic absorption offers the potential for avoiding such first-pass metabolism. It could also target anticancer agents to lymphatic carcinomas (7). Table 1 lists various materials and associated therapeutic agents that have been formulated for lymphatic delivery.

Oleic acid has been used in a novel approach to boost the bioavailability of propranolol (Fig. 1). The effect was ascribed to preferential uptake by the lymphatic system and avoidance of the extensive first-pass metabolism that would follow passage through the hepatic portal system (8). Formulation with triglycerides also enhanced lymphatic absorption of the antimalarial drug halofantrine (9). However, the low lymph/blood flow ratio limits lymphatic absorption to drugs that are highly lipophilic ($\log P > 5$) and that have significant solubility in long-chain triglycerides (>50 mg/mL).

Strategies to breach physical and enzymatic barriers in the intestinal epithelium that hinder passage to the systemic circulation have included enhancing paracellular flux by disrupting “tight junctions” (10, 11). Inhibition of the P-glycoprotein (PGP) that ejects unrecognized or unwanted materials also has been studied (12). Certain lipids are reported to be PGP inhibitors but there are no reports of successful application to commercial products or use in clinical trials.

Yet another approach to intestinal absorption enhancement concerns the inhibition of intestinal Cytochrome P450 3A4, an enzyme responsible for the prehepatic metabolism of many drugs. Grapefruit juice is reported to be a powerful inhibitor of this enzyme and is known to enhance the bioavailability of cyclosporin, triazolam, nifedipine, and other drugs. Studies have been carried out to identify the components in grapefruit juice that evince

Table 1 Drug carriers for lymphatic targeting

Lymphotropic carrier	Drug	Type of interaction
Dextran sulphate	Bleomycin	Ion-pair
Dextran	Mitomycin C	Covalent binding
β -Cyclodextrin oligomer	1-Hexylcarbamoyl-5-fluorouracil	Hydrophobic inclusion
L-Lactic acid oligomer microsphere	Aclarubicin, cisplatin	Incorporation
Gelatin microsphere	Mitomycin C	Incorporation
Intrinsic protein complex	Vitamin B ₁₂	Complex
Styrene-maleic acid anhydride-co-polymer	Neocartinstation	Covalent binding
Liposome	Ara-C	Encapsulation
S/O emulsion	5-Fluorouracil, Bleomycin	Encapsulation
Lipid mixed micelle	Interferon, TNF	Hydrophobic binding
Chylomicron, LDL	Cyclosporine, vitamin A, coenzyme Q, DDT	Incorporation
Carbon colloid	Ethinylesteradiol 3-cyclopententyl ether Mitomycin C, Aclarubicin	Hydrophobic adsorption

(From Ref. 7.)

this effect (13, 14). It could be argued that inclusion of such materials (thought to be flavinoids), as “excipients” in the dosage form would lead to not only more complete but also more consistent systemic levels by counteracting inconsistencies brought about by enzyme inhibitors in food and drink (such as grapefruit juice). Time will tell whether the ongoing interest in this area will lead to new “excipients” that modulate absorption in this way.

Excipients that are bioadhesive or that swell on hydration can promote absorption by increased contact with epithelial surfaces, by prolonging residence time in the stomach, or by delaying intestinal transit. Cellulose ethers, gums of natural origin, and synthetic acrylic acid polymers have been evaluated for such purposes. The range of materials available and their differing viscoelastic and rheological behaviors mean that it is possible, by

judicious admixture, to develop delivery units with balanced properties so that adhesion, density, hydration, drug release rate, etc. can be tailored to the drug in question and the physiological characteristics of the target delivery site (15).

Enhancers for Other Modes of Absorption

Many physical and enzymatic barriers can prevent successful delivery of active pharmaceutical ingredients by noninvasive, nonoral routes. It is not surprising, therefore, that there is great interest in excipients that can overcome such obstacles.

Transdermal delivery is a case in point. The skin, particularly the stratum corneum presents a formidable barrier to diffusion. Materials used to enhance its permeability have ranged from simple solvents such as ethanol or propylene glycol to aromatic chemicals such as terpenoids. Such penetration enhancers appear to work by disrupting the lipid domains in the stratum corneum that reduce permeability (16). A bespoke penetration enhancer, laurocapram (Azone), was developed in the late 1970s for use in transdermal delivery but its use in commercial products appears to be limited (17).

Entry via nasal or buccal mucosa allows the delivery of peptides or other labile drugs that are highly potent (low-dose drugs) and that do not have steep dose-response relationships. Absorption enhancement requires increased contact time and reduced clearance rate (in the case of nasal delivery), thereby optimizing conditions for mucosal diffusion. Excipients that enhance nasal absorption include

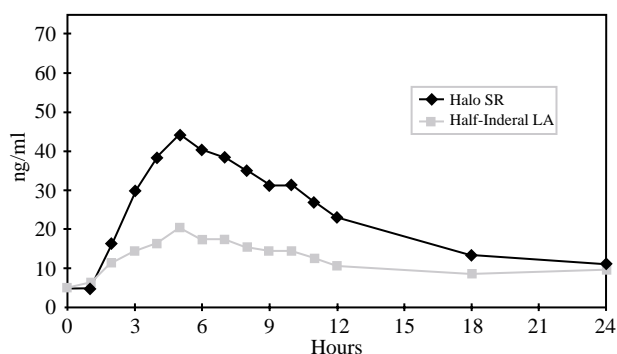


Fig. 1 Effect of oleic acid on propranolol bioavailability. (From Ref. 8.)

phospholipids to enhance mucosal permeability and agents that imbibe water and become mucoadhesive (e.g., glyceryl mono oleate). In addition, the gelling agents hydroxypropyl cellulose and polyacrylic acid promote absorption of insulin in dogs (18). These findings, however, have not been used to find a commercially viable product for intranasal delivery of insulin, presumably due to insulin's narrow therapeutic index. However, intranasal delivery systems for calcitonin, a macromolecule with a much safer dose-response relationship, have been commercialized.

EXCIPIENTS FOR DRUG TARGETING

The 1990s saw an explosion in knowledge and understanding of the roles that natural mediators play in physiological and pathological processes. At the same time, biotechnology has made it feasible to manufacture such mediators relatively cheaply and in large quantities, thereby affording possibilities for use as therapeutic agents. However, effective delivery remains a formidable challenge from the efficacy, safety, and patient-convenience perspective. Most natural mediators are highly potent, extremely labile, and may need to be delivered to a specific organ or cellular target. Conventional oral dosage is not usually feasible due to the hostile environment and enzymatic barriers along the GI tract. Parenteral administration is hardly desirable for chronic therapy. Therefore, many biotechnology products need to be combined with materials that afford protection against destruction, reduce elimination rate, or target a specific site so that activity is enhanced and toxicity minimized. The level of interest and activity in this area supports the view that more effective delivery systems are required if the promise of biotechnology is to be realized. Hence, it is likely that the search for absorption-enhancing excipients for such materials will continue unabated.

Carriers for biopharmaceutical therapeutic agents range from well-established excipients of natural origin to custom-made synthetic materials with putatively enhanced protective or targeting features. Natural or semisynthetic materials predominate however. Sources as diverse as primitive marine plants (chitosans and alginates), plant or animal phospholipids (egg and seed lecithin), and mammalian collagens (gelatin) are being mined for useful delivery or targeting aids, as well as for components of complex formulations such as microemulsions or liposomes. The wide use of biological materials may mean that Mother Nature produces more suitable biopolymers than the synthetic chemist. More plausibly, it may reflect

the need for long and expensive safety evaluation of novel synthetic materials prior to use in man. This hinders timely evaluation other than in vitro or animal models.

More esoteric materials that confer target specificity include glycoproteins, recombinant proteins, or monoclonal antibodies (19). To date, clinical performance of such carrier systems has been disappointing. Further refinement of concepts and materials may be necessary before the performance matches the promise.

Attenuated adenoviruses have been used as vectors where delivery to cell nuclei is required (e.g., in gene medicine). It is a moot point whether these or other targeting or carrier materials are "excipients" part of a prodrug or something in-between. The boundaries between "active" and "inactive" materials are much less clear in such cases. The traditional approach of evaluating a novel entity in its own right in animal safety programs and then formulating with "inert" materials is inappropriate with sophisticated delivery systems because of the important effect of the adjuvant on disposition and kinetics of the active ingredient.

EXCIPIENTS AS STABILIZERS

Product quality can be compromised during manufacture, transport, storage or use. The causes of deterioration can be manifold and product-specific. They include microbial spoilage or chemical transformation of the active or physical changes that alter performance in vivo. Deterioration can compromise safety or make the medication less attractive, which means it may not be used. Excipients can contribute to or cause such changes unless carefully screened for possible interactions in preformulation studies.

Cases where excipients have the opposite effect and stabilize labile drugs are less common. Nevertheless, they have been shown to reduce degradation rates of drugs that are photolabile, oxidizable, or degradable consequent to inter- or intramolecular reactions (20). Stabilization strategies include the following:

1. Formulation with an excipient whose light absorption spectrum overlaps that of the photolabile drug. This is the so-called spectral overlay approach;
2. Using an antioxidant in formulations that are susceptible to degradation by oxidation. This approach has been particularly successful in vitamin-containing products
3. Using an excipient that "hinders" association of groups in the same molecule, in adjacent molecules, or in the vehicle that can interact and cause degradation. There

are several reports of cyclodextrins effecting such “steric stabilizations.” Polyethylene glycol also has been shown to stabilize an ointment formulation by preventing formation of inactive rearrangement products.

Less esoteric but equally important stabilizers include preservatives in liquid products to prevent microbial growth and buffers to provide an environment conducive to good stability where degradation is pH-related. Chelating agents also are used as stabilizers to prevent heavy metals from catalyzing degradation.

EXCIPIENTS AS PROCESS AIDS

The vast majority of medicinal products are manufactured by high-speed, largely automated processes for reasons that are related as much to safety and quality as to cost of goods. Excipients that aid in processing include the following:

1. The almost universal use of lubricants such as stearates in tablets and capsules to reduce friction between moving parts during compression or compaction;
2. Excipients that aid powder flow in tablet or capsule manufacture. Materials such as colloidal silica improve flow from hopper to die and aid packdown in the die or capsule shell. Accuracy and consistency of fill and associated dose is thereby improved
3. Compression aids to help form a good compact, whether on dry granulation (slugging) prior to tableting or on tablet compression. Most are derived from plant, animal, or mineral origin (microcrystalline cellulose, lactose, or magnesium carbonate)
4. Agents such as human or bovine serum albumin that are used in the manufacture of biotechnology-based products. These avoid adsorption of the protein to flexible tubing, filters, and other process equipment
5. Stabilizers to protect the drug from processing conditions that might otherwise be deleterious. It is common to use “cryoprotectants” such as sugars, polyhydric alcohols or dextrans in lyophilized parenteral biotechnology products to prevent inactivation during freezing. A similar approach has also been used to prevent liposomal aggregation and leakage (21).

“Flow aids” also can help performance in cases where the delivery device is an integral part of the medication. Products for pulmonary delivery are often formulated as dry powders that are inhaled via the oral cavity. The fine-particle nature of the medicinal agent, which may be vital for efficient delivery to the

bronchial target area, militates against good flow. Materials such as lactose or mannitol (of appropriate particle size) can enhance flow or act as a “carrier” from the dose unit (usually a capsule) through the inhalation delivery device to the oral cavity on inspiration. They are widely used for these purposes in inhalation formulations of anti-asthmatic agents such as salbutamol and budesonide (22, 23). The recombinant therapeutic proteins human deoxyribonuclease (used to treat cystic fibrosis) and human granulocyte colony stimulating factor (g-CSF) are also formulated with “carriers” to aid pulmonary delivery (24).

The adherence of very fine particles to larger ones can solve segregation problems when mixing powders containing particles of differing size or shape. If the fine particles can associate with their larger companions due to some surface effect, “ordered mixing” ensues and homogeneity is assured during subsequent processing (25).

Process aids do not usually contribute to the performance of the dosage form in terms of quality or in vivo performance. Indeed, lubricants, because of their hydrophobic nature, can hinder disintegration and dissolution of solid dosage forms unless the level and mode of incorporation is carefully characterized and controlled. Thus, in addition to drug-excipient interactions, the potential for interexcipient competition and incompatibility must be considered and studied.

DRUG–EXCIPIENT INTERACTIONS

Despite the earlier account of excipients acting as stabilizers, it is fair to state that there are far more cases on record of excipients adversely affecting quality. Degradation may be caused by interaction between functional groups in the excipient and those associated with the drug. Many small-molecule drugs contain primary, secondary, or tertiary amino groups and these have the propensity to interact with aldehydic groups in sugars or volatile aldehydes present as residues. Chemical interaction can result in degradation of the drug substance to inactive moieties with loss of efficacy where degradation is excessive. Even when degradation is modest, it is possible that the formed degradation products may compromise safety.

Physical interactions between drug and excipient also can compromise quality. Adsorption of drug by microcrystalline cellulose resulted in drug dissolution being less than complete (26). Interaction between chloramphenicol stearate and colloidal silica during grinding led to polymorphic transformation (27).

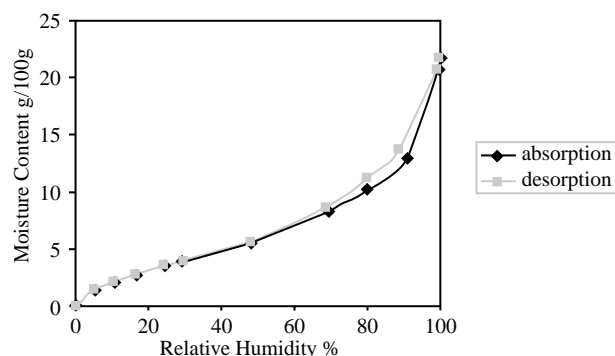


Fig. 2 Water vapor sorption isotherm for microcrystalline cellulose at 25°C. (From Ref. 30.)

Excipients may contribute to degradation even when not directly interacting with active moieties. Soluble materials may alter pH or ionic strength, thereby accelerating hydrolytic reactions in liquid presentations. Such effects may be accentuated during processing. For instance, sterilization by autoclaving, while of short duration, may cause significant degradation product formation because of the high temperature involved. Dextrose is widely used in parenteral nutrition solutions or as a tonicity modifier in other parenterals. Sterilization by autoclaving can cause isomerization to fructose and formation of 5-Hydroxymethyl furfuraldehyde in electrolyte-containing solutions (28). At the other extremes of processing, succinate buffer was shown to crystallize during the freezing stage of lyophilization, with associated reduction of pH and unfolding of gamma interferon (29). It is important to identify and characterize such “process stresses” during dosage-form development and tailor processing conditions accordingly.

Microcrystalline cellulose is a partially depolymerized cellulose that is part-crystalline/part noncrystalline and hygroscopic. Adsorbed water is not held in any “bound” state but will rapidly equilibrate with the environment during processing or storage (30) (Fig. 2). Thus, it is possible that in a dosage form, water can be sequestered by a more hygroscopic active ingredient. If the drug is moisture sensitive, degradation may follow. Stabilization may be possible by drying prior to use, but loss of water may make it a less effective compression aid (31).

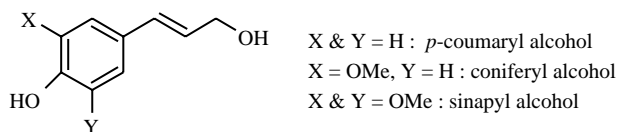


Fig. 3 Potential residues in microcrystalline cellulose.

Microcrystalline cellulose may also contain low levels of nonsaccharide organic residues. These emanate from lignin, a cross-linked biopolymer made up primarily of the three allylic alcohols/phenols in the wood chip starting material (Fig. 3) (32).

It is also possible that degradation products of these phenols, or free radical combinations, may be present, with potential for chemical interaction with the drug.

This focus on residues associated with microcrystalline cellulose is not to denigrate it as an excipient. It is and will remain a most valuable formulation aid that can help compression, disintegration, and flow, as well as acting as a general diluent in solid-dose formulation (33). It can also be a useful additive in liquid products. Indeed, the knowledge available on microcrystalline cellulose interactions probably reflects the level of interest in such a useful material. Rather, the intent is to illustrate how excipients, or residues contained in them, can interact with active ingredients in a number of ways. The first commandment for the formulator is arguably to “know your drug,” but it is also important to be aware of the composition, residues, and other behaviors of excipients.

STABILITY OF EXCIPIENTS

Excipients can lose quality over time. Oils, paraffins, and flavors oxidize; cellulose gums may lose viscosity. Polymeric materials used in film coating or to modify release from the dosage form can age due to changes in glass transition temperature. This can lead to changes in elasticity, permeability, and hydration rate and associated changes in release properties or appearance (34).

Preservatives such as benzoic acid or the para hydroxybenzoates are volatile and can be lost during product manufacture if the process involves heating. Loss during product storage is also feasible if containers are permeable to passage of organic vapors. Acetate buffer is volatile at low pH and can be lost during the drying stages of lyophilization. Such behaviors reinforce the need to know the behaviors of excipients as well as of the active ingredient so that appropriate processing, storage conditions, and “use by” periods are stipulated where necessary.

IMPURITIES IN EXCIPIENTS

Excipients, like drug substances contain process residues, degradation products or other structural deviants formed during manufacture. Historically, it was not unusual for

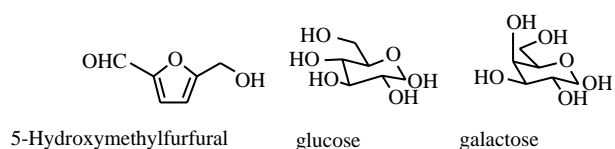


Fig. 4 Potential residues in lactose.

adulterants to be added to “bulk up” the commodity. Thankfully, a combination of better analytical techniques, vendor certification programs, and quality audit systems should mean that adulteration is largely a thing of the past. However, constant vigilance is necessary. As recently as 1996, renal failure in children in Haiti was ascribed to use of glycerol contaminated with diethylene glycol in a liquid paracetamol product (35).

Residues in excipients can affect quality and performance by interacting with the drug or other key components. Reducing sugar impurities in mannitol were responsible for the oxidative degradation of a cyclic heptapeptide (36).

Lactose is one of the most widely used excipients in tablet manufacture. It is available in a number of different forms, differing in hydration and crystal states. Isolation and purification may involve treatment with sulphur dioxide (37). However, there are no reports of complications from residues of this powerful oxidizing agent.

Lactose is a disaccharide comprised of glucose and galactose units. These reducing sugars are reported to be present in spray dried lactose (38), as is the hexose degradant 5-hydroxymethyl furfural (39). This aldehyde has the potential for additional reactions with primary amino groups, Schiff Base formation and color development (Fig. 4) (40).

Drugs containing secondary amine groups also can be degraded. Maillard reaction products have been reported in capsules containing lactose and the antidepressant fluoxetine (41). This reaction is also reported extensively in publications concerned with the food industry. High temperatures and low moisture contents associated with food processing induces caramelization of sugars and oxidation of fatty acids to aldehydes, lactones, ketones, alcohols, and esters (42, 43). It would not be surprising if such degradation products were generated in the same materials used in pharmaceutical dosage forms. Unfortunately, most pharmacopoeial monographs do not list such organic contaminants. Also, some excipient vendors are reluctant to share information on residues and contaminants with customers. The pharmaceutical industry generally represents only a small proportion of business for such commodity providers. Hence, it is difficult to be persuasive on the need for individualized standards and controls. Therefore, the following list (Table 2) cannot be

Table 2 Potential residues in common excipients

Excipient	Residue
PVP, Polysorbates, benzyl alcohol	Peroxides
Magnesium stearate, fixed oils, paraffins	Antioxidants
Lactose	Aldehydes, reducing sugars
Benzyl alcohol	Benzaldehyde
Polyethylene glycol	Aldehydes, peroxides, organic acids
Microcrystalline cellulose	Lignin, hemicelluloses

considered as comprehensive because of this unsatisfactory state of affairs.

Excipient residues may also compromise safety or tolerance. Wool fat or lanolin derived from sheep wool may contain low levels of insecticides from sheep treated for parasites. These insecticide levels are probably too low to cause direct toxicity, but may cause allergic reactions when lanolin in cosmetics or topical medicaments is applied to the skin.

Paradoxically, excipient residues such as antioxidants may inadvertently act as stabilizers of the drug substance. Unheralded removal by the vendor or replacement by a different type of stabilizer could precipitate an unheralded product stability crisis leading to recall from the market. Such a possibility highlights the need for agreed change control systems between the pharmaceutical manufacturer and the excipient vendor.

SAFETY OF EXCIPIENTS

Although excipients have traditionally been considered “inert,” it is now well accepted that some carry the potential for untoward effects. These can range from the inconvenient to the serious, be general or patient-specific, and may or may not be dose-related. The effect may be ascribable to the excipient itself or to a residue from the starting material or the process of manufacture.

Lactose is one of the most widely used tablet excipients. However, 5–10% of the population of the United Kingdom suffers from lactose malabsorption (44), nor is there reason to suppose that this percentage is lower in other countries. Lack of the lactase enzyme leads to fermentation by colonic bacteria, with formation of lactic acid and carbon dioxide causing stomach cramps, diarrhea, and vomiting (45). Whether such clinical symptoms are manifest following ingestion of the levels

normally present in dosage forms is not known, but such phenomena may sometimes explain minor side effects regularly reported during the monitoring that accompanies volunteer Phase I studies.

Malabsorption of the cereal protein gluten is another potential source of untoward effects from excipients. This condition demands a gluten-free diet. Most starches utilized in medicinal products are now gluten-free but gluten-containing materials have been used as film formers in microencapsulated products (46). There is also a possibility that gluten could be present in excipients that utilize cereal derivatives as starting materials or bases (e.g., dry powder flavors that sometimes use maltodextrin bases).

Sucrose is a very effective sweetener, particularly for liquids dosed to children. Its propensity to cause dental caries and the complications it poses in the management of diabetes may have contributed to its progressive removal from medicinal products despite its continuing widespread use in foods and confectionery. Sorbitol is another excellent sweetening agent and has been used as a replacement for sucrose in oral liquid products. It has the propensity to cause diarrhea and flatulence, although the effect may only be manifest at high doses. However, there may be additive effects (e.g., if it is formulated with active ingredients that are also associated with GI intolerance, such as antibiotics).

Synthetic sweeteners have had checkered careers as excipients (47). Cyclamate was banned in the United States following reports of carcinogenicity and withdrawal of generally regarded as safe (GRAS) status in 1969. It remains banned despite additional studies to clarify safety and attempts at reinstatement. It remains acceptable in Europe.

Saccharin is equally controversial. It also is suspected as being a carcinogen due to cyclohexylamine formation, possibly by gut flora, on ingestion (48). It was banned as a food additive by the Food and Drug Administration (FDA) in 1977, but has remained available consequent to regular congressional moratoria on the proposed ban. It is not permitted in Canada except for diabetic beverages and foods.

The flavoring agent sodium glutamate is sometimes used to flavor protein supplements or liver extracts. Flushing, headache, and chest pain have been ascribed to its presence, albeit after food intake rather than medication. This is the background to the so-called Chinese restaurant syndrome (49).

Aspartame is a newer sweetener/flavor enhancer but it too may cause angiodema and urticaria. It is contraindicated in patients suffering from phenylketone urea, as hydrolysis can lead to formation of phenylalanine.

Aspartame also can hydrolyze in solution to form a diketopiperazine derivative and can participate in Michael-type addition reactions with olefines susceptible to nucleophilic attack. The products of such interactions, if they occur, will be drug and formulation-specific, and it is likely that their safety characteristics will be unknown (45).

The use of benzyl alcohol as a local anesthetic was previously discussed. It is also used as a preservative in parenteral dosage forms. However, there is some evidence that benzyl alcohol is neurotoxic and its use is contraindicated in the United Kingdom in children under 3 years of age (50).

The literature is replete with reports of various allergic-type reactions to preservatives (parabens, chlorocresol), antioxidants (propyl gallate, metabisulphite), surfactants and solvents. The list is too long to be discussed in this article but Ref. 48 contains a very useful compilation and discussion of immunotoxic events seen with various dosage form additives.

Many of these studies involved application of copious amounts to animal skin or to human volunteers in Phase I studies. Others concerned reports of reactions in people suffering from pre-existing allergic conditions. Reports of side effects must, therefore, be viewed from such perspectives and the possibilities for side effects weighed against the widespread use of the same materials in food, confectionery, cosmetic, and household products as well as in medicines. This is not to belittle the hypersensitivity and other reported reactions but unless these are put into context, there may be further constraints on excipient usage and unrealistic demands for "totally inert" formulation adjuvants.

Adverse reactions may be caused by the excipient *per se* but by a residue. HIV infection in humans and spongiform encephalopathies in cattle have raised the specter of viral transmission by materials as diverse as human and bovine serum albumin, lactose, gelatin, fatty acids, or their salts, as well as polyols such as glycerol. It is generally accepted that screening procedures for blood donors and the heat treatment usually employed for sterilizing human serum albumin (minimum 10 h at 60°C), originally introduced to guard against hepatitis infection, provides good lethality against the HIV virus. However, the prions associated with spongiform encephalopathies are so resistant to heat and other forms of sterilization that removal is more problematic. Consequently, the use of vegetable sources for fatty acid and organic alcohol-type excipients is becoming more common. Whether gelatin capsules will be replaced by starch or cellulose gum-based alternatives (for the same reason) remains to be seen.

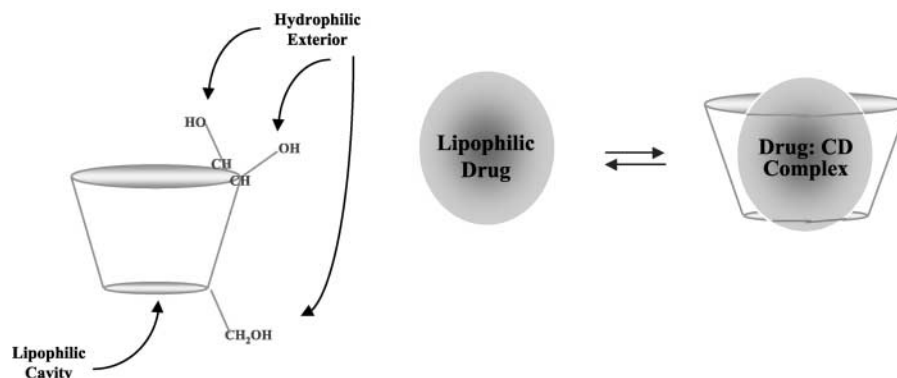


Fig. 5 Mechanism of solubilization by cyclodextrins. (Reproduced from CyDex Inc.)

NOVEL EXCIPIENTS

It might be expected that the increased knowledge of pathological processes and drug-receptor dynamics, along with the relentless pressure for manufacturing efficiencies and economies of scale that have been the hallmark of the 1990s, would also demand and generate new and better formulation aids. This has not happened. Indeed, some have implied that excipients available in 2000 A.D. are not very different from those that were available in 2000 B.C. (51). While clearly calculated to amuse, the assertion contains a grain of truth. Only a handful of novel excipients have emerged over the past 20 years.

The reasons for this are not difficult to understand. Like novel pharmacological agents, a novel excipient must go through numerous safety and metabolic evaluation processes before it can be used in humans. In essence, it would be necessary to apply for a Type 4 Drug Master File in the United States, or a Certificate of Suitability in European Union (EU) countries (50). Such safety and filing programs are expensive and time-consuming. Furthermore, it is difficult to prove “lack of activity” in any material. Excipients are not subject to prescription or pharmacovigilance monitoring, therefore, they need to be “squeaky clean” before “blanket approval” is forthcoming. While a novel excipient can be evaluated at the same time as a novel drug, few organizations wish to put their investment in a novel drug at risk by partnering it with an unproven excipient. Therefore, novel excipients are likely to remain scarce commodities. However, a number of materials considered as “novel” are evincing interest as formulation aids.

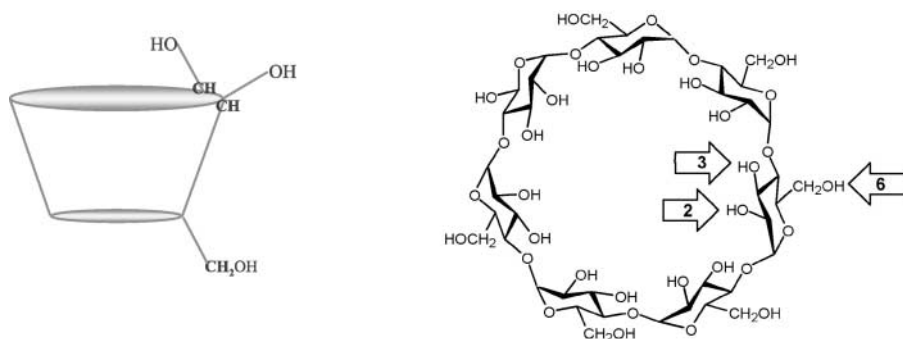
Cyclic Glucose Polymers

Cyclodextrins are not new molecular entities. They were first reported a century ago. However, it is only relatively

recently that their potential as formulation aids has been recognized. Their capability to stabilize labile drugs has already been mentioned. They can also be used to solubilize highly insoluble molecules as, with the insertion of the drug in the annulus, the complex largely acquires the solubility characteristics of the cyclodextrin (52) (Fig. 5). Inclusion complexes have also been used to successfully mask taste or odor, reduce sublimation of drugs with high volatility (53), and enhance thermal stability (54).

The so-called parent cyclodextrins viz the alpha, beta, and other forms (Fig. 6) have properties that may have prevented widespread use as formulation adjuvants. The moderate solubility and the perceived need to form molar complexes meant that their use would be limited to low-dose, highly potent compounds. Furthermore, β -Cyclodextrin in particular could not be used parenterally because of renal nephrotoxicity. This was ascribed to its low solubility possibly associated with the propensity to form a molecular complex with cholesterol *in vivo* and precipitate in the proximal renal tubule. Thus, the potentially most useful application viz dissolution of poorly soluble compounds for injection could not be countenanced. However, β -Cyclodextrin is currently a well-established excipient in oral dosage forms and has recently been allocated monographs (as Betadex) in the European Pharmacopoeia (EP) 2000 and in the U.S. National Formulary, NF 19.

It is also encouraging that derivatized cyclodextrins with greater solubility are now available. The hydroxypropyl and sulphobutyl ether derivatives of β -Cyclodextrin (Fig. 7) have much greater solubilities than the parent material. Indeed, sulphobutyl ether was deliberately developed for use with parenterals in the knowledge that many novel drug substances are poorly soluble. Both these forms have been subjected to comprehensive safety evaluation programs (parenteral in the case of the sulphobutyl ether), and Drug Master Files have been lodged with the FDA. Such



Parent Cyclodextrin	a	b	g
Glucose Units	6	7	8
Molecular Weight	973	1135	1297
Water solubility (g/100 mL)	14.5	1.85	23.2
Cavity Diameter (Å)	4.7-5.3	6.0-6.5	7.5-8.3
Cavity Volume (Å) ³	~174	~262	~472

Fig. 6 Properties and functional groups of some cyclodextrins. (Reproduced from CyDex Inc.)

initiative and commitment on the part of the manufacturers of these newer agents is particularly praiseworthy in light of the costly safety evaluation programs. It may well be that with the availability of such “more suitable” cyclodextrins, they will find a valuable niche in the armamentarium of the formulation scientist. The references cited below comprise two excellent reviews of the promise, properties, and limitations of cyclodextrins (55, 56).

Thus, cyclodextrins are a family of excipients, each with somewhat different properties that allow the

possibility of matching individual cyclodextrins to specific drugs to compensate for a deficiency or to aid performance in some way.

Fluorocarbons

The replacement of chlorofluorocarbon (CFC) propellants with the nonozone-depleting hydrofluorocarbons (HFCs) merit mention for two reasons. First, it illustrates how environmental impact can be an important selection

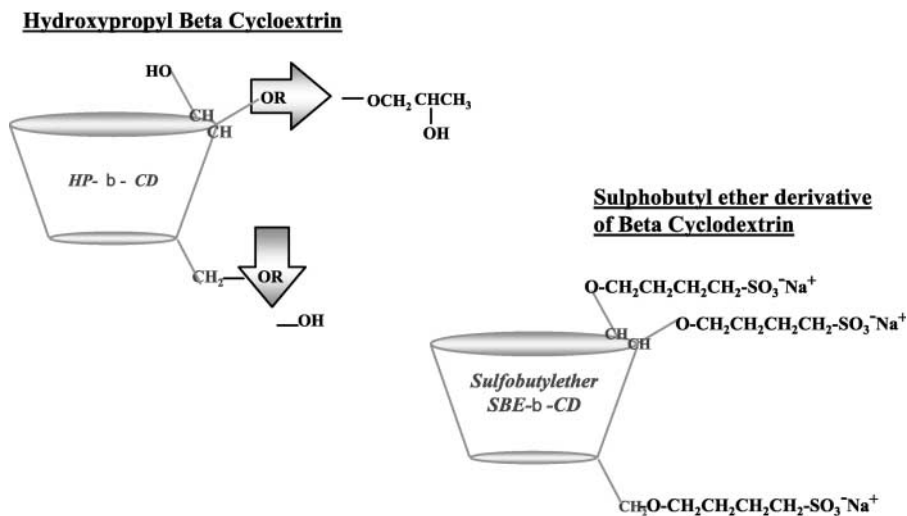


Fig. 7 Functional groups of novel cyclodextrins. (Reproduced from CyDex Inc.)

criterion at a time when “green” issues are high profile. Second, HFCs were developed and evaluated for safety and delivery capability by a consortium of pharmaceutical companies, with costs shared and evaluation programs defined by prior agreement between end-users and propellant manufacturers. Such collaboration could be employed usefully in the future to develop novel excipients for delivery or targeting. The benefits would undoubtedly accrue to all.

Polymeric Targeting or Delivery Aids

Many publications, particularly from academic institutes, contain information concerning synthetic or semisynthetic polymers, which are designed to enhance targeting or delivery properties. However, evaluation of the effect of such material on performance has been invariably confined to in vitro work or perhaps studies in rodents. Clearance for use in humans has not been obtained. Therefore, these substances cannot be considered as excipients that are readily available to the formulation technologist.

If few novel delivery materials become available in the future, the formulation scientist may have to rely on using mixtures of established excipients that, in combination, have properties that are “greater than the sum of the parts” in terms of viscoelasticity, diffusivity, tissue/organ specificity or other desirable targeting or delivery features. Such approaches seem likely to provide considerable scope for creative approaches, and for the formulation technologist, it should be an exciting and fulfilling road to travel.

EXCIPIENT SELECTION

The nature and properties of the active ingredient dictate the choice of an excipient, the dosage form to be elaborated, and the process by which it is manufactured. It is also important to know the patient group and clinical condition. The mode of use of the medication and the envisaged dose must also be considered. Candidate excipients should then be evaluated to demonstrate that they function in the manner intended (do what they are meant to do) and do not adversely interact with the drug, or with other excipients. Obviously, they should not have any pharmacological effect and should not otherwise compromise safety or tolerance.

It is also necessary to consider the regulatory status of excipients and any country-specific requirements or constraints. The U.S. and Japanese regulatory agencies

publish lists of excipients used in medicinal products (57, 58). The materials listed in these compendia can generally be considered suitable for administration by the route for which they are already being used. For materials with no history of previous use, evidence must be provided that they do not compromise patient safety nor induce any other undesirable effects.

SOURCING EXCIPIENTS

Excipients can be crucial determinants of product performance and quality. Thus, they should be sourced directly from a reputable vendor who has quality systems in place to ensure consistent manufacture and control. Procurement from brokers is to be discouraged. Auditing such providers for the presence of quality systems and controls should be the norm, particularly if they are new suppliers to the pharmaceutical industry. A validation program should be put in place to establish reliability of the supply source (59). This program should take the following into account:

1. The nature of the excipient and medicinal product in which it will be used
2. The conditions under which the materials are manufactured and controlled
3. The nature and status of the supplier, and his understanding of the Good Manufacturing Practice (GMP) requirements of the pharmaceutical industry
4. The Quality Assurance system of the manufacturer. Excipients, unlike active ingredients, are not currently subject to regulatory control in terms of GMP unless they are novel materials (in which case preapproval inspection for GMP compliance is necessary). However, the Guide to Good Manufacturing Practice for Bulk Pharmaceutical Excipients, elaborated by the International Pharmaceutical Excipients Councils (IPEC) of Europe and the U.S. while not having any regulatory status, provides much useful information on quality systems and is a good reference for performing audits of excipient facilities (60, 61).

A particular drug or dosage form may have features that rely on the presence of excipients for stabilization, delivery, or other performance parameters. Alternatively, the excipient may need to have additional features to render it suitable for the product in question (e.g., density, absence of a particular residue, etc.). In such cases it may be necessary to agree to extra quality tests and limits over and above those demanded by pharmacopoeias or applied by the vendor.

It is also prudent to be aware of the materials, reagents, and solvent used in the manufacture of the excipient and consider potential interactions between such residues and the active ingredient. It may also be advisable to agree to a Change Control notification procedure with the vendor, to avoid the introduction of new materials in the manufacture of the excipient without prior consideration of the possible impact on the medicinal product.

CONCLUSION

Traditional attitudes that viewed excipients as "inert" materials are long outmoded. It is now well accepted that they are not merely place fillers but can be true "partners" of the active ingredient in many medicinal products and have the potential to enhance or possibly adversely affect performance. As such, their choice, quality control, mode of inclusion, stability, and performance characteristics merit the same attention as the active ingredient. Thus, knowledge of excipients, their foibles, and requirements for handling, processing, and storage are powerful assets in the armamentarium of the pharmaceutical technologist.

REFERENCES

- Orr, N. Quality Control and Pharmaceutics of Uniformity of Content of Medicines Containing Potent Drugs with Special Reference to Tablets. *Progress in the Quality Control of Medicines*; Deasy, P.B., Timoney, R.F., Eds.; Elsevier Press: Amsterdam, 1981; 193–256.
- Roy, G.M. Taste Masking in Oral Pharmaceuticals. *Pharm. Technol. (USA)* **1994**, *18*, 84–90.
- Hutchinson, K. Formulation of Softgels for Improved Oral Delivery of Hydrophobic Drugs. *Excipients and Delivery Systems for Pharmaceutical Formulations*; Royal Society of Chemistry: Cambridge, 1995; 133–147.
- Bates, T.R.; Sequira, J.A. Bioavailability of Micronized Griseofulvin from Corn Oil: In Water Emulsion, Aqueous Suspension and Commercial Tablet Dosage Forms. *J. Pharm. Sci.* **1975**, *64*, 793–797.
- Brown, N.J.; Read, N.W.; Richardson, R.D.; Bogentoft, C. Characteristics of Lipid Substances Activating the Ileal Brake in the Rat. *Gut* **1990**, *31*, 1126–1129.
- Hunt, J.N.; Knox, M.T. The Slowing of Gastric Emptying by Four Strong Acids and Three Weak Acids. *J. Physiol.* **1972**, *222*, 187–208.
- Muranishi, S. Drug Targeting towards Lymphatics. *Adv. Drug Res.* **1991**, *21*, 1–37.
- Barnwell, S.G.; Gauci, L.; Harris, R.; Attwood, D.; Littlewood, G.; Guard, P.; Pickup, M.E.; Barrington, P. J. Controlled Release **1994**, *28* (1–3), 306–309.
- Porter, C.H.; Charman, S.A.; Charman, W.N. Lymphatic Transport of Halofantrine in the Triple-Cannulated Anaesthetized Rat Model: Effect of Lipid Vehicle Dispersion. *J. Pharm. Sci.* **1996**, *85* (4), 351–361.
- Muranishi, S. Absorption Enhancers. *Crit. Rev. Drug Carrier. Syst.* **1990**, *7* (1), 1–33.
- Swenson, E.S.; Curatola, W.J. Means to Enhance Penetration 2. Intestinal Permeability Enhancement for Protein Peptides and Other Polar Drugs: Mechanisms and Potential Toxicity. *Adv. Drug. Del. Rev.* **1992**, *8* (1), 39–92.
- Hidalgo, I.J. Cultured Epithelial Cell Models. *Models for Assessing Drug Absorption and Metabolism*; Plenum Press: New York and London, 1996; 35–48.
- Hukkinen, S.K.; Varhe, A.; Oikkola, K.T.; Neuvonen, P.J. Plasma Concentrations of Triaxolam are Increased by Concomitant Ingestion of Grapefruit Juice. *Clin. Pharmacol. Ther.* **1995**, *58* (2), 127–131.
- Edwards, D.J.; Bellevue, F.H.; Woster, P.M. Identification of 6' 7' Dihydroxybergamotin, a Cytochrome P450 Inhibitor in Grapefruit Juice. *Drug Metab. Dispos.* **1996**, *24* (12), 1287–1290.
- Zaman, M.; McAllister, M.; Martini, L.G.; Lawrence, M.J. The Physicochemical and Biological Factors Influencing Bioadhesion. *Pharm. Tech. Eur.* **1999**, *11* (8-Biopharm. Suppl.), 52–60.
- Hadgraft, J.; Walters, K.A. Skin Penetration Enhancement. *J. Dermatol. Treat.* **1994**, *5* (1), 43–47.
- Wiechers, J. Excipients in Topical Drug Formulations. *Man. Chem.* **1998**, 17–21.
- Rogerson, A.; Parr, G.D. Nasal Drug Delivery. *Routes of Drug Administration*; Florence, A.T., Salole, E.G., Eds.; Wright: London, 1990; 1–29.
- Hnatyszyn, H.J.; Kossovsky, N.; Gelman, A.; Sponsler, E. Drug Delivery Systems for the Future. *PDA J. Pharm. Sci. Technol.* **1994**, *48* (5), 247–254.
- Crowley, P.J. Excipients as Stabilizers. *Pharm. Sci. Tech. Today* **1999**, *2* (6), 237–243.
- Carpenter, J.F.; Crowe, J.H. Modes of Stabilization of a Protein by Organic Solutes during Dessiccation. *Cryobiol.* **1988**, *25*, 459–470.
- Ganderton, D. The Generation of Respirable Clouds from Coarse Powder Aggregates. *J. Biopharm. Sci.* **1992**, (3), 102–105.
- Steckel, H.; Muller, B.W. In Vitro Evaluation of Carrier Particle Size and Concentration on In Vitro Deposition. *Int. J. Pharm.* **1997**, *154*, 31–37.
- Chan, H.K.; Clark, A.R.; Gonda, I.; Mumenthaler, M.; Hsu, C. Spray Dried Powders and Powder Blends of Recombinant Human Deoxyribonuclease (rhDNase) for Aerosol Delivery. *Pharm. Res.* **1997**, *14* (4), 431–437.
- Cook, P.; Hersey, J.A. Powder Mixing in the Tableting of Fenfluramine Hydrochloride: Evaluation of a Mixer. *J. Pharm. Pharmacol.* **1974**, *26*, 298–303.
- Senderoff, R.I.; Mahjour, M.; Radebaugh, G.W. Characterization of Adsorption Behavior by Solid Dosage Form Excipients in Formulation Development. *Int. J. Pharm.* **1992**, *83*, 65–72.
- Forni, F.; Coppi, G.; Iannuccelli, V.; Vandelli, M.A.; Camerani, R. The Grinding of the Polymorphic Forms of Chloramphenicol Stearic Ester in the Presence of

- Colloidal Silica. Acta. Pharmaceutica. Suecica. **1988**, 25 (3), 173–180.
28. Buxton, P.C.; Jahnke, R.; Keady, S. Degradation of Glucose in the Presence of Electrolytes during Heat Sterilization. Eur. J. Pharm. Biopharm. **1994**, 40 (3), 172–175.
29. Lam, X.M.; Constantino, H.R.; Overcashier, D.E.; Nguyen, T.H.; Hsu, C.C. Replacing Succinate with Glycollate Buffer Improves the Stability of Lyophilised Gamma Interferon. Int. J. Pharm. **1996**, 142, 85–95.
30. Hollenbeck, G.R.; Peck, G.E.; Kildsig, D.O. Application of Immersional Calorimetry to Investigation of Solid–Liquid Interaction: Microcrystalline Cellulose–Water System. J. Pharm. Sci. **1978**, 67 (11), 1599–1606.
31. Khan, K.A.; Musikabhumma, P.; Warr, J.P. The Effect of Moisture Content of Microcrystalline Cellulose on the Compression Properties of Some Formulations. Drug Dev. Ind. Pharm. **1981**, 7, 525–538.
32. Lewis, N.G.; Davin, L.B.; Sarkanen, S. *The Nature and Function of Lignins: Comprehensive Natural Products Chemistry*; Pinto, B.M., Ed.; Pergamon Press: Oxford, 1999; 3, Ch. 18.
33. Jivraj, M.; Martini, L.G.; Thomson, C.M. An Overview of the Different Excipients Useful for the Direct Compression of Tablets. Pharm. Sci. Technol. Today **2000**, 3 (2), 58–62.
34. Guo, J.H. Aging Processes in Pharmaceutical Polymers. Pharm. Sci. Tech. Today **1999**, 2 (12), 478–483.
35. Anonymous. Morbidity and Mortality Weekly Reports **1996**, 45 (30).
36. Dubost, D.C.; Kaufman, J.; Zimmerman, J.A.; Bogusky, M.J.; Coddington, A.B.; Pitzenberger, S.M. Characterization of a Solid State Reaction Product from a Lyophilized Formulation of a Cyclic Heptapeptide: A Novel Example of an Excipient-Induced Oxidation. Pharm. Res. **1996**, 12, 1811–1814.
37. Stabilization of Crystalline Lactose, Netherlands Patent Application: NL 8301220 A 841101(DMV-Campina B.V., Netherlands). Patent Application, NL 83-1220 830407.
38. Brownley, C.A., Jr.; Lachman, L. Preliminary Report on the Comparative Stability of Certified Colorants with Lactose in Aqueous Solution. J. Pharm. Sci. **Jan. 1963**, 52, 86–93.
39. Brownley, C.A., Jr.; Lachman, L. Browning of Spray-Processed Lactose. J. Pharm. Sci. **April, 1964**, 53, 452–454.
40. Janicki, C.A.; Almond, H.R., Jr. Reaction of Haloperidol with 5-(Hydroxymethyl)-2-Furfuraldehyde: An Impurity in Anhydrous Lactose. J. Pharm. Sci. **1974**, 63, 41–43.
41. Wirth, D.D.; Baertschi, S.W.; Johnson, R.A.; Maple, S.R.; Miller, M.S.; Hallenbeck, D.K.; Gregg, S.M. Maillard Reaction of Lactose and Fluoxetine Hydrochloride, A Secondary Amine. J. Pharm. Sci. **1998**, 87, 31–39.
42. Aidrian, J. The Maillard Reaction. *Handbook of the Nutritive Value of Processed Food*, Rechcigl, M., Ed.; CRC Press: Boca Raton, FL, 1982; 1, 529–608.
43. Danehy, J.P. Maillard Reactions: Non-Enzymic Browning in Food Systems with Specific Reference to the Development of Flavor. Adv. Food Res. **1986**, 30, 77–138.
44. Neale, G. The Diagnosis, Incidence and Significance of Disaccharidase Deficiency in Adults. Proc. Royal Soc. Med. **1968**, 61, 1099–1102.
45. Weiner, M., Bernstein, I.L., Eds. Adverse Effects: Bulk Materials. *Adverse Reactions to Drug Formulation Agents*; Marcel Dekker, Inc.: New York, 1989; 89–113, Ch. 6.
46. Weiner, M., Bernstein, I.L., Eds. Adverse Effects: Bulk Materials. *Adverse Reactions to Drug Formulation Agents*; Marcel Dekker, Inc.: New York, 1989; 121–128, Ch. 7.
47. Pinco, R.G. Hurdling International Barriers to Existing and New Excipients. World Pharm. Stand. Rev. **1991**, 2, 14–19.
48. Smith, J.M.; Dodd, T.R.P. Adverse Reactions to Pharmaceutical Excipients. Adv. Drug React. Pois. Rev. **1982**, 1, 93–142.
49. Taylor, S.L.; Dormedy, E.S. The Role of Flavoring Substances in Food Allergy and Intolerance. Adv. Food Nutrition Res. **1998**, 42, 1–44.
50. Saiki, J.H.; Thompson, S.; Smith, F.; Atkinson, R. Paraplegia Following Intrathecal Chemotherapy. Cancer **1972**, 29 (2), 370–374.
51. Robertson, M.I. Regulatory Issues with Excipients. Int. J. Pharm. **1999**, 187, 273–276.
52. Staniforth, J.N. Design and Use of Tableting Excipients. Drug Dev. Ind. Pharm. **1993**, 19 (17/18), 2273–2308.
53. Loftsson, T.; Brewster, M. Pharmaceutical Applications of Cyclodextrins. 1. Solubilisation and Stabilization. J. Pharm. Sci. **1996**, 85, 1017–1025.
54. Duchene, D.; Vauton, C.; Glomot, F. Cyclodextrins: Their Value in Pharmaceutical Technology. Drug Dev. Ind. Pharm. **1986**, 12 (11/13), 2193–2215.
55. Cwiernia, B.; Hladon, T.; Stobieki, M. Stability of Diclofenac Sodium in the Inclusion Complex with β -Cyclodextrin in the Solid State. J. Pharm. Pharmacol. **1999**, 51 (11), 1213–1218.
56. Stella, V.J.; Thompson, D.O. Cyclodextrins-Enabling Excipients: Their Present and Future Use in Pharmaceuticals. Crit. Rev. Ther. Drug Carrier Syst. **1997**, 14 (1), 1–104.
57. Stella, V.J.; Rajewski, R.A. Cyclodextrins: Their Future in Drug Formulation and Delivery. Pharm. Res. **1997**, 14 (5), 556–567.
58. *Inactive Ingredients Guide* Division of Drug Information Resources; United States Food & Drug Administration Center for Drug Evaluation Research (CDER): Rockville, MD, Jan 1996.
59. *Japanese Pharmaceutical Excipients Directory 1996* Japanese Pharmaceutical Excipients Council, Eds.; Yakuji Nipo, Ltd.: Tokyo, MHW, Japan, 1996.
60. *Rules and Guidance for Pharmaceutical Manufacturers and Distributors*; Medicines Control Agency (UK), HMSO: London, 1997, Section 5, Annex 8.
61. Mervicill, A. A Good Manufacturing Practices Guide for Bulk Pharmaceutical Excipients. Pharm. Technol. (USA) **1995**, 19, 34–40.
62. Weiner, M., Bernstein, I.L., Eds. *Adverse Reactions to Drug Formulation Agents (A Handbook of Excipients)*; Marcel Dekker, Inc.: New York, 1989.
63. Kibbe, A.J., Ed. *Handbook of Pharmaceutical Excipients*, 3rd Ed.; Pharmaceutical Press: London, 2000.
64. *Inactive Ingredients Guide* Division of Drug Information Resources; United States Food & Drug Administration, Center for Drug Evaluation Research (CDER), Rockville, MD, Jan 1996.
65. *Japanese Pharmaceutical Excipients Directory 1996*, Japanese Pharmaceutical Excipients Council, Eds.; Yakuji Nipo Ltd.: Tokyo, 1996.

EVAPORATION AND EVAPORATORS

David P. Kessler

Purdue University, West Lafayette, Indiana, U.S.A.

INTRODUCTION

In the general sense, *evaporation* refers to any change in phase of a component from liquid to gas. *Vaporization*, sometimes used interchangeably with *evaporation*, is at times specifically used to designate the total change of a liquid phase to gas (vapor).

In this article only the term *evaporation* will be used. Evaporation will be defined as processes carried out in process equipment conventionally classified as *evaporators*. This, in turn, implies that nonequipment-contained classes of evaporation, such as solar ponds and oil tanker spills, will be ignored.

Evaporators are used to increase the concentration of relatively nonvolatile dissolved or suspended components in a solution or slurry (the *liquor*) by evaporating portions of the liquid phase using energy supplied by a *medium*, often steam. The dissolved or suspended components do not appear in the vapor phase to a substantial extent. (If they do, the process is referred to as *distillation*.)

Other methods that will not be discussed here, but also can be used to increase concentration (some with and some without concomitant evaporation) are reverse osmosis, ion exchange, dialysis, electrodialysis, osmotic distillation, and applications that involve fluidized beds, cooling towers, or evaporation of aerosols.

In most evaporators, the solvent or suspending phase is primarily or totally one constituent, most frequently water. The important product in evaporation can be either the more concentrated mixture left behind or the overhead vapor (which is often, but not necessarily, subsequently condensed).

The overhead solvent vapor in solvent recovery processes or boiler water vapor in power plant applications typifies vapor products. *Blowdown* refers to the periodic or continuous purging of the bottoms used to control buildup of undesirable material in the liquid phase when producing a vapor product (1).

Some processes of evaporation can be accompanied by crystallization, as the residual liquor grows more and more concentrated. Carried yet further, evaporation evolves to *drying* (or dehydration, if the constituent removed is

water), as the bottoms product obtained becomes primarily solid rather than liquid.

TYPICAL APPLICATIONS FOR EVAPORATORS

Historically, a classic example of an evaporation process is the production of table salt. Maple syrup has traditionally been produced by evaporation of sap. Concentration of black liquor from pulp and paper processing constitutes a large-volume present application. Evaporators are also employed in such disparate uses as: desalination of seawater, nuclear fuel reprocessing (1), radioactive waste treatment (2, 3), preparation of boiler feed waters, and production of sodium hydroxide (2). They are used to concentrate stillage waste in fermentation processes, waste brines, inorganic salts in fertilizer production, and rinse liquids used in metal finishing, as well as in the production of sugar, vitamin C, caustic soda, dyes, and juice concentrates, and for solvent recovery in pharmaceutical processes.

TYPES OF EVAPORATORS

Extended discussion of types (including photographs and schematic diagrams), design, and operation of evaporators can be found in the literature (4–6).

Because evaporation of a liquid phase usually requires addition of large amounts of thermal energy, the method of transferring this heat to the liquor tends to dominate evaporator capital cost. The source of heat for evaporators is usually a medium such as hot combustion gases or a condensing vapor, typically steam. Molten salts and electrical resistance heaters are less commonly used sources of thermal energy.

Flash evaporators operate by an adiabatic decrease of the pressure on a liquid that has been previously heated. These were first used for production of potable water on ships; now they are used for more general brackish waters and seawater as well as for processed liquids (7).

Disk or *cascade* evaporators use the partial immersion of either disks mounted perpendicular to, or bars mounted parallel to a rotating shaft to carry films of liquid into a hot gas stream (8).

The most efficient method of transferring the energy of a heating medium to the liquor is direct injection of the heating medium. Because of the consequent contamination of the liquor with the heating medium, this method of heat transfer is of relatively less importance in the pharmaceutical industry and will not be discussed here.

The more useful methods for pharmaceutical products maintain purity at the expense of additional resistance to heat transfer by interposing a solid wall of some thermally conductive material between the heating medium and the liquor. The solid wall is usually metallic, but can be coated with materials such as glass, porcelain enamel, or polymers. Glass or ceramic themselves can be used for walls.

The solid wall can be the wall of the evaporator itself, as in *jacketed* evaporators. The area available for heat transfer in jacketed vessels, however, is quite limited. Jacketed vessels frequently incorporate some sort of internal agitator.

Heat transfer can be supplied from within a vessel by a heating coil, but again, the available heat transfer area is not large; however, such coils can be designed in ways that

make their removal for cleaning relatively easy. The alternative is to have the heat exchange external to the main chamber of the evaporator.

Some applications use plate-type exchangers. In plate exchangers, the bounding surface may be in the plate-and-frame form (parallel plates with the heating medium and the liquor flowing in alternate interstitial spaces), or in a spiral-plate configuration that contains a concentric pair of spiral passages (7). Such exchangers can be cleaned easily. They do, however, require a large gasketed area. Fig. 1 shows a typical plate-type evaporator.

The most common geometry for the separating surface between the heating medium and the liquor is probably that of tube bundles, which can be oriented either horizontally or vertically, with the liquor flowing on either the outside or the inside of the tubes. Depending on the application, the tube bundle can be either inside or outside the vessel in which the evaporation takes place.

The heating element of an evaporator is sometimes referred to as a *calandria*. Usually this term is applied to a heating system in which the liquor rises through a vertical tube bundle surrounded by the heating medium and then descends through a central well.

The *short-tube vertical* evaporator is an early type that still sees considerable industrial use. The heating element,

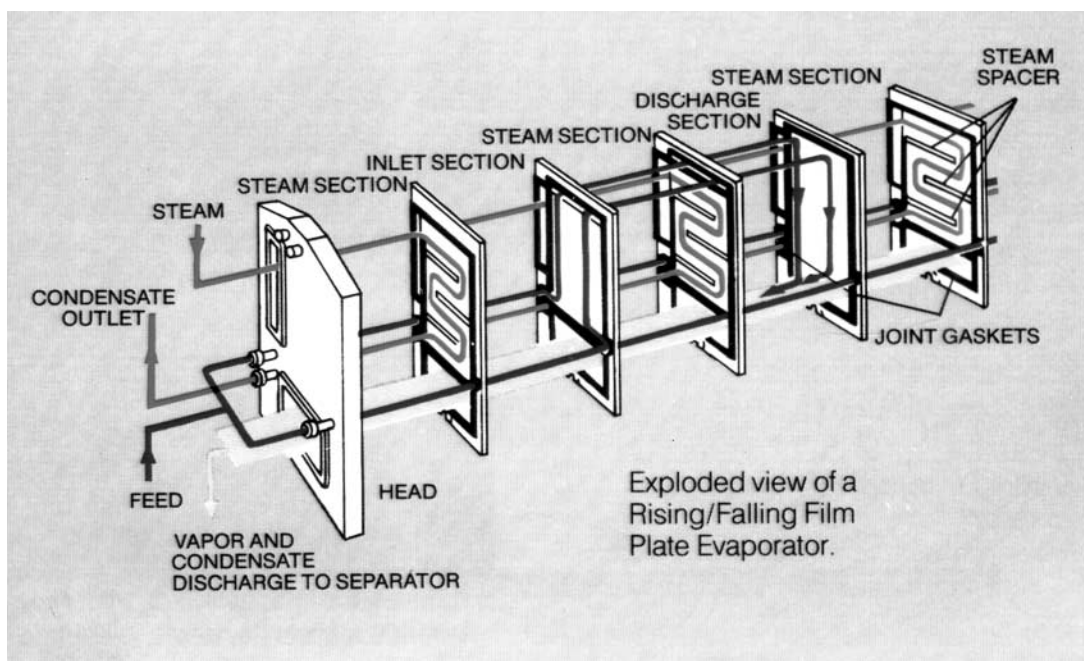


Fig. 1 Exploded view of rising/falling film plate evaporator. (Courtesy of APV Crepaco, Inc., APV Americas, Rosemont, IL—An Invensys Company.)

a vertical bundle of tubes around a center well, is sometimes colloquially referred to as the basket. Circulation is upward through the tubes, the *rising film* mode, and then downward through the central well or *downtake*. Liquid boils in the tubes, which decreases the overall density therein and thus creates the driving force for circulation, since the density of the (nonboiling) fluid in the downtake is greater than that in the tubes.

Mechanical cleaning is fairly easy with such units, and the capital investment is relatively low. Circulation stops, however, if the heat input is interrupted, creating the danger of the settling of any solids suspended in the liquor. This type of unit is not well suited to viscous liquids because of the low heat transfer coefficients associated with the low velocities of natural convection (8). Short-tube vertical evaporators have largely been surpassed by other types, particularly for applications involving liquors that foam, deposit excessive scale, are excessively viscous, or are heat sensitive.

Long-tube vertical evaporators are normally the cheapest per unit of capacity (8). When operated in the rising film mode, temperature variation along the inside of the tubes is both substantial and difficult to predict. The variation in pressure from high at the bottom to low at the top normally means that the liquid enters the bottom of the tubes below its boiling temperature. The liquid is subsequently heated to boiling as it rises, and the boiling temperature simultaneously decreases as the pressure decreases toward the top of the tube (assuming any boiling point rise from increasing concentration is overshadowed by the effect of the reduced pressure on the boiling point).

By operating a long-tube evaporator in the *falling film* mode, the problem of temperature variation induced by pressure differences is mitigated. Here, a film of liquid surrounding a gas core flows down the walls of the tube, so pressure drop is very much less than in the rising film mode. The low residence time of the falling film units makes them useful for heat-sensitive materials, but the necessity of maintaining a film on the walls of the tubes makes feed distribution a problem. They are readily adapted to sanitary processing. Evaporators that combine rising film sections and falling film sections in the same unit are also available.

Forced circulation evaporators have relatively higher heat transfer coefficients, and are somewhat less subject to fouling, salting, and scaling. This advantage is offset by both the cost of external power required for the circulating machinery and a relatively high holdup (8). At times they more frequently experience plugging from deposits detached from the walls of the unit by the force of the circulating fluid. The introduction of a pump may lead to mechanical problems, particularly with liquors that are slurries.

Liquor velocities required to prevent surface deposits are often greater than can be obtained with natural circulation at reasonably low temperature differences (9). In addition to mitigating scale formation, forced circulation also improves the heat transfer coefficient.

For viscous liquids, one way to increase the heat transferred is to improve the heat transfer coefficient by scraping or stirring the fluid adjacent to the wall, as in *agitated film* or *wiped film* evaporators. Accommodation of the mechanical devices used to mix the fluid close to the wall requires a fairly large diameter tube, so these devices tend to consist of only a single tube; thus, heat transfer area is relatively small. The introduction of moving mechanical parts may lead to maintenance problems.

In *horizontal tube* evaporators, the liquor is usually on the outside of the tubes and the heating medium on the inside. Rather than submerging the tubes, the boiling liquid is sometimes sprayed on the outside of the tubes. This gives a performance approaching that of falling film evaporators (8).

Evaporators can be operated at a variety of pressures (9). Reduced pressure, with its concomitant reduction in boiling temperature, offers advantages for heat-sensitive materials and materials that are sensitive to exposure to air.

Evaporation operations are often staged in *multiple effect* systems (1) to achieve better efficiency. Such systems can have a variety of relative directions for flow of liquor and vapor. A typical example of such staging is illustrated in Figs. 2 and 3.

Detailed discussion of the advantages and disadvantages of various types of evaporators is available (8). A table summarizing the advantages and disadvantages of common types of evaporators also is available (10).

Since most evaporators are purchased from outside suppliers either prefabricated or on-site-fabricated, such suppliers can be an excellent source of information on selection of evaporator type. Suppliers and addresses can be found in the literature (11, 12).

DESIGN OF EVAPORATORS

A number of publications have addressed the design of evaporators at both the elementary and complex levels (1, 2, 7–9, 13–18). The reader is referred to these and to their bibliographies for details.

Efficiency of steam-heated evaporators is commonly described in terms of steam economy, defined as the pounds of solvent evaporated per pound of steam consumed (7).



Fig. 2 Two-effect rising/falling film plate evaporator. (Courtesy of APV Crepaco, Inc., APV Americas, Rosemont, IL—An Invensys Company.)

Important to the design of evaporators is the concept of boiling point rise. Boiling point rise, which normally accompanies increasing concentration in a liquor, is defined as the difference between the boiling temperature of the liquor and that of pure water at the same pressure.

A plot of the boiling point of pure water on the abscissa versus the boiling point of the liquor on the ordinate, with lines of constant concentration plotted as a parameter, is known as a Dühring plot. A Dühring plot is unique to the particular chemical species. A monograph summarizing the boiling point rise for a number of inorganic salts, is available (19), as well as information on general equations to predict boiling point rise either empirically or from other thermodynamic properties (9).

During batch operation, the composition of the liquor in the evaporator is continuously changing. For batch

operation in the absence of significant potential and kinetic energy changes, an energy balance on the system constituted by the liquor in the evaporator shows that the sum of the rate of enthalpy removed in the overhead stream and the change in internal energy of the liquor in the evaporator must equal the rate of energy addition as heat:

$$\hat{H}_{ov}\dot{W} + \frac{d(\hat{U}M)}{dt} = \dot{Q} \quad (1)$$

where

\hat{H}_{ov}	enthalpy of overhead vapor (energy/mass)
\dot{W}	mass flow rate of overhead vapor (mass/time)
\hat{U}	internal energy of liquor (energy/mass)
M	mass of liquor (mass)
\dot{Q}	rate of heat input (energy/time)
t	time (time)
()	units

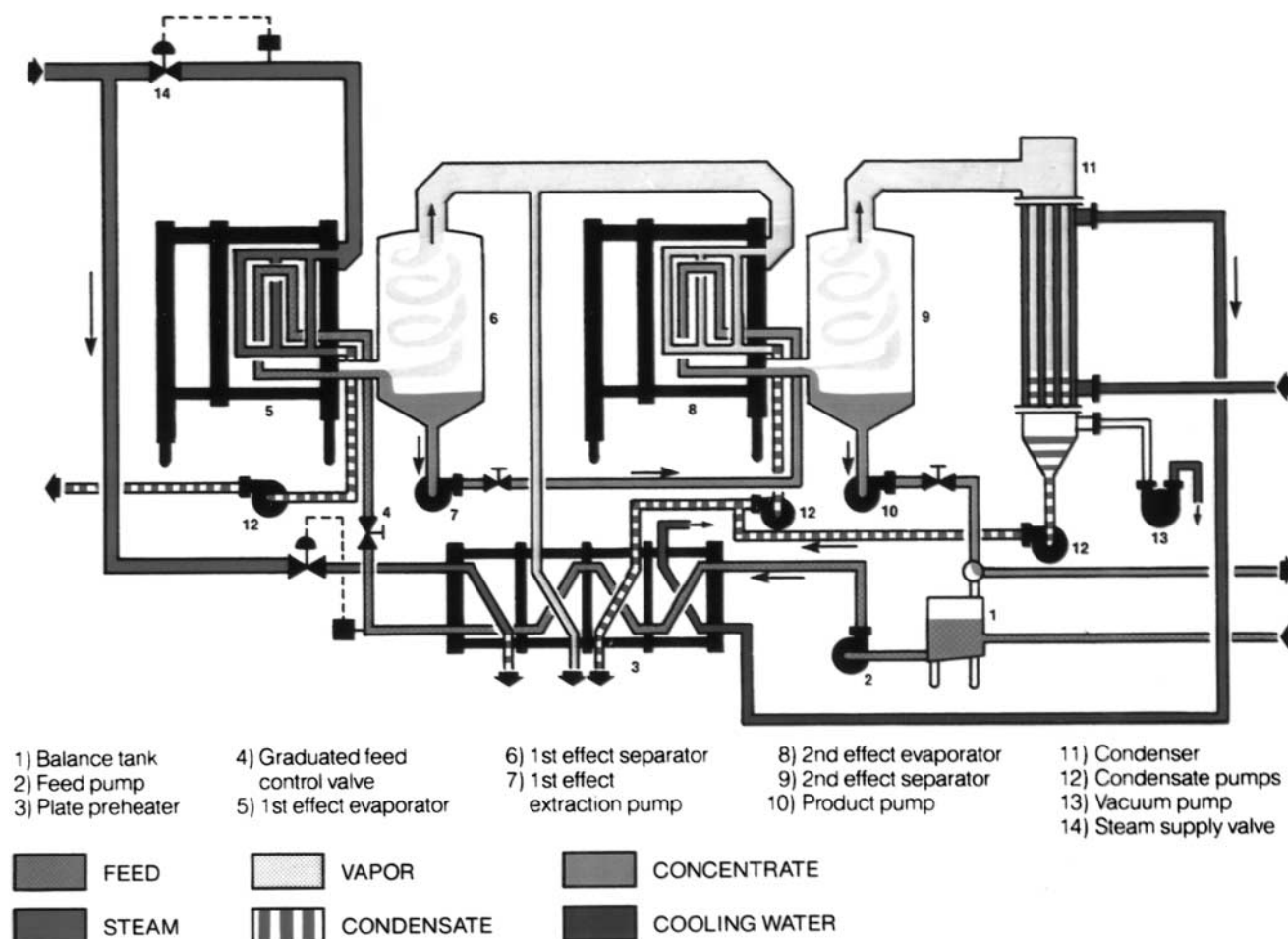


Fig. 3 Two-effect rising/falling film plate evaporator flowsheet. (Courtesy of APV Crepaco, Inc., APV Americas, Rosemont, IL—An Invensys Company.)

A way to estimate vapor enthalpy and data on the internal energy of the liquor as a function of concentration is required.

Even though both the enthalpy and rate of the overhead vapor are often quite constant in time (except for the small sensible thermal energy contribution from the boiling point rise), the rate of heat addition may vary with time because of changes in the overall heat transfer coefficient from surface changes (see below) and because of significant changes in the temperature difference between the liquor and heating medium induced by the boiling point rise. A simple example calculation of batch evaporation can be found in reference (14). The equations involved are similar to those that govern batch distillation.

Many evaporators are operated at substantially steady-state conditions (although heat transfer coefficients will vary slowly with time because of surface changes from scaling, deposition, etc., as noted above). The design of an

evaporator for steady-state operation typically is initiated from the specification of the mass flow rate of material to be processed and its required change in concentration. This information is used in mass balances, which proceed via an energy balance to yield the required rate of heat transfer.

For an evaporator system which has a liquid feed, overhead vapor, and bottoms liquid product, and is used to concentrate species A:

Total mass balance

$$\dot{W}^{(\text{feed})} = \dot{W}^{(\text{overhead vapor})} + \dot{W}^{(\text{bottoms liquid})} \quad (2)$$

Species mass balance

$$\begin{aligned} \dot{W}^{(\text{feed})} x_A^{(\text{feed})} &= \dot{W}^{(\text{overhead vapor})} x_A^{(\text{overhead vapor})} \\ &+ \dot{W}^{(\text{bottoms liquid})} x_A^{(\text{bottoms liquid})} \end{aligned} \quad (3)$$

where

- \dot{W} mass flow rate (total mass/time)
 x mass concentration of species A (mass A/total-mass)

Normally, the feed flow rate and concentration are specified, as are the overhead and bottoms concentrations (the overhead concentration is most often zero). This set of two equations in 2 unknowns can then be solved for the unknown overhead and bottoms mass flow rates.

The required rate of heat transfer can then be calculated from the steady-state energy balance, assuming negligible changes in potential and kinetic energy and no work input (the latter not valid, of course, for vapor recompression evaporators).

$$\dot{Q} = \dot{W}^{(\text{overhead vapor})} \hat{H}^{(\text{overhead vapor})} + \dot{W}^{(\text{bottoms liquid})} \hat{H}^{(\text{bottoms liquid})} - \dot{W}^{(\text{feed})} \hat{H}^{(\text{feed})} \quad (4)$$

where

- \dot{Q} heat transfer rate into system (energy/time)
 \dot{W} flow rate (mass/time)
 \hat{H} enthalpy (energy/mass)

In the above equation, the mass flow rates are known from the mass balance calculation. The enthalpies must be determined from thermodynamic data, which will require specification of the temperature and/or pressure of the stream in combination with the concentrations that were originally specified.

Calculation of the rate of heat transfer required does not of itself determine the heat transfer area required or the configuration thereof. The configuration (tubes, plates, etc.) is typically chosen first, by rules of thumb and experience, depending on the liquor to be processed.

The area is usually calculated, once the configuration is chosen, by using an overall heat transfer coefficient that lumps together all forms of heat transfer in terms of an overall coefficient paired with a characteristic area:

$$\dot{Q} = UAT \quad (5)$$

where

- U overall heat transfer coefficient, valid only when paired with a specified area (energy/time \times length² \times temperature)
 A a specified area normal to heat flow; (e.g., the outside area of the tubes or the inside area of the tubes (length²),
 ΔT Temperature difference between temperature of heating medium and temperature of the liquor

(usually either freestream or bulk temperatures) (temperature).

The overall heat transfer coefficient depends on the properties and flow pattern of the heating medium, the properties and flow pattern of the liquor, the properties of the solid surface that separates the medium and the liquor, and the properties of any deposits at the interface on either side of the separating surface.

In general, heat can be transferred by the three mechanisms of conduction, convection, and radiation (20). In evaporators, the mechanisms of importance are usually convection in the liquor and medium, and conduction through the solid separating them.

In order to calculate the overall heat transfer coefficient in the equation above, the heat transfer properties of the heating medium and the liquor are described in terms of individual heat transfer coefficients, and the heat transfer properties of the separating solid in terms of its thermal conductivity. Deposits at the interface, which one might expect to be described by a thermal conductivity, are usually described instead in terms of either a fouling (heat transfer) coefficient or a "fouling factor" (which is usually defined to have a value of 1000/[fouling coefficient]).

Individual heat transfer coefficients for convection are defined by the equation:

$$\dot{Q} = hAT \quad (6)$$

where

- h individual heat transfer coefficient (energy/time \times length² \times temperature),
 A a characteristic area normal to the flow of heat: typically, the area of the unfouled surface initially in contact with the fluid. In cases where area varies along the heat transfer path, (e.g., tubes), it is necessary to specify on which area the coefficient is based, such as, inside, outside (length²),
 ΔT a specified temperature difference, usually the difference between either the freestream or the bulk temperature and the temperature at the surface.

The rate of heat transfer by conduction can be written in terms of the thermal conductivity:

$$\dot{Q} = kA_{\text{avg}}Tx \quad (7)$$

where

- k thermal conductivity of solid (energy/time \times length \times temperature),
 A_{avg} average area through which heat transfer occurs, e.g., logarithmic mean area for concentric cylinders (length²),

- ΔT temperature drop from one outside surface of the solid to the other outside surface (temperature)
 Δx thickness of solid normal to heat flow (length)

One can relate the overall heat transfer coefficient to the individual steps:

$$\frac{1}{U_{ov}A_{ov}} = \frac{1}{h_{medium}A_{medium}} + \frac{1}{h_{fouling,medium}A_{fouling,medium}} + \frac{x}{kA_{avg,solid}} + \frac{1}{h_{fouling,liquor}A_{fouling,liquor}} + \frac{1}{h_{liquor}A_{liquor}} \quad (8)$$

The overall heat transfer coefficient can be calculated by evaluation of the individual terms in the above equation. Knowledge of the overall temperature drop then permits calculation of the required area. A detailed development of the above is available (5).

The individual heat transfer coefficients for the medium and the liquor can be calculated from general correlations using the properties of the fluid and the velocity fields in the system (20–25). Additional information can be found in the continuing series of the American Society of Mechanical Engineers (26). Detailed studies and bibliographies on prediction of individual heat transfer coefficients for evaporating films on/in horizontal/vertical tubes, can be found in other publications (21, 22, 24–27).

The limiting step in heat transfer is usually the thermal resistance from the liquor itself and/or deposit formation on the liquor side (the last two terms in the previous equation.) On the heating medium side, the heat transfer coefficients are usually for condensation and are therefore substantially larger than those on the liquor side, which are for boiling. The conduction resistance of the tube wall is usually small (but this is not necessarily true of the resistance of deposits that accumulate on the tube wall).

Deposits typically accumulate more on the boiling side than the condensing side. To determine coefficients for systems with substantial surface deposition, experimental data on that particular system is usually necessary.

Heat-sensitive products can break down to insoluble forms that deposit on the surface and inhibit heat transfer. For example, some soluble proteins, when heated, convert to an insoluble form (18). Change in the heat transfer characteristics of the tube surface can also be induced by corrosion.

The liquor can be on either the inside or the outside of the tubes, depending on the physical and chemical characteristics of the materials involved. Considerations include convenience in cleaning deposits from the heat transfer surface, the heat transfer characteristics of the fluids involved, and fluid velocities required for efficient heat transfer or prevention of scale deposition.

A useful table of criteria and data for the rapid design and selection of evaporators is available (10).

Natural circulation evaporators have overall coefficients of the order of $1.1\text{--}3.4 \text{ kW}/(\text{m}^2\text{K}) = 200\text{--}600 \text{ Btu}/(\text{h ft}^2\text{F})$. Adding forced circulation may raise this to the order of $11 \text{ kW}/(\text{m}^2\text{K}) = 2000 \text{ Btu}/(\text{h ft}^2\text{F})$. In agitated-film units, for Newtonian liquids with viscosity of the order of water, coefficients of the order of $2.3 \text{ kW}/(\text{m}^2\text{K}) = 400 \text{ Btu}/(\text{h ft}^2\text{F})$ may be obtained. As the viscosity increases to $10 \text{ Newton sec}/\text{m}^2 = 10,000 \text{ centiPoise}$, the coefficient will drop to the order of $0.7 \text{ kW}/(\text{m}^2\text{K}) = 120 \text{ Btu}/(\text{h ft}^2\text{F})$. More extensive listings of overall coefficients for evaporators may be found in the literature (1, 8, 10, 28).

The heating medium must be at a higher temperature than the liquor, and the resultant temperature driving force must be sufficiently large that excessive area for transfer is not required. One means of increasing the temperature driving force for a heating medium at a given temperature is to reduce the pressure of the liquor side to decrease the boiling temperature. This is a common device in evaporation, and is often accomplished with a steam-jet ejector. Design (and cost) of such ejectors is detailed in the literature (29).

If the overhead vapor is water and the heating medium is steam, the vapor generated is usually comparable in amount and quality to the heating steam used, but is at a lower pressure because the heating steam must be at a higher temperature (therefore, pressure) than the liquor in order to furnish an adequate temperature driving force for heat transfer. A logical question is whether or not to attempt to recover the latent heat in the vapor generated.

The condensing temperature of the vapor does not usually furnish an appropriate temperature difference to permit transferring heat to the liquor (except, perhaps, a small part to sensible heat of the incoming liquor). However, compressing the vapor will raise both its temperature and its condensing temperature, permitting the vapor to be used to transfer its heat of vaporization/condensation to the liquor with consequent recovery of thermal energy.

Unfortunately, the work required for such compression is high-quality (and therefore, expensive) mechanical energy that ends up as low-quality thermal energy. Either thermocompression or mechanical compression may supply the required mechanical energy. Thermocompres-

sion is accomplished by means of high-pressure steam in an ejector system [for design and cost of ejectors see (29)], while mechanical compression is usually done by a centrifugal compressor (1).

An alternative scheme is to split the evaporation process into stages (in different vessels), commonly referred to as *effects*, where the flow of liquor may be in the same direction, the backward direction, or normal to the flow of vapor (1, 9). In such a scheme, the vapor generated in a given effect is used to boil the liquid in a different effect where a proper temperature driving force exists. In multiple-effect evaporation, one exchanges savings in steam costs for increased capital investment in equipment.

Pennink (30) notes that the optimum is usually 3 or 4 effects in a 50,000 lb/h system and 7 to 8 effects in very large systems. Ramakrishna (31) details a shortcut method to estimate the optimum number of effects for multi-effect evaporation, and Ulrich (10) gives a shortcut algorithm for the design of multiple-effect evaporators. A sample calculation can be found in Schilt (15). Cole (18) gives specific steam consumption and minimum evaporation rates for single and multiple-effect evaporators with and without steam-jet thermocompression, as well as data on mechanical vapor recompression cost per ton of water evaporated.

For those who wish to pursue more detailed design, a good starting point for mechanical design of heat exchange equipment is Azbel (9), who also discusses alternative materials of construction. For general discussion of evaporation as well as insight into safety/environmental considerations, see a report from the U.S. Environmental Protection Agency (32). This document discusses failure analysis as applied to evaporators for treatment of effluent from the metal finishing industry. Other useful references, which have extensive bibliographies, are Rubin et al. (33), Knudsen et al. (34), Chisholm (35), and Taborek et al. (36). The continuing series from the American Society of Mechanical Engineers, as exemplified by Shah (37), contains additional information.

COSTS ASSOCIATED WITH EVAPORATORS

The two major costs associated with evaporators, as with any process equipment, are capital investment and operating costs. The best estimate of the installed cost of evaporation systems is, of course, a firm bid from a vendor. The installed cost, however, can be estimated based on the heat transfer surface area, as in Peters and Timmerhaus (38). Costs taken from published references

must be adjusted for changes subsequent to the time of publication. To do this, one may use an index such as the Marshall and Swift all-industry index. The value of this index is published each month in *Chemical Engineering*, a McGraw-Hill publication. Further information on the use of this and other cost indices as well as their histories are available, for example, in Peters and Timmerhaus (38) and Ulrich (10). Variation of purchased evaporator costs with material of construction and pressure can also be found in Ulrich (10).

A chart to estimate the costs of ejectors for thermocompression has been developed (29). Approximate costs of compressors for mechanical vapor recompression also can be found (38–40). In addition, these publications contain details of cost estimation for heat exchangers. For more detailed estimation of cost of shell-and-tube exchangers, see Purohit (41–43). Plate-and-frame and spiral-plate heat exchanger costs can be estimated using Kumana (44).

Operating expenses can be approximated using published techniques (10, 38). These are mainly useful for preliminary cost estimates only. Such items as raw materials, operating labor, utilities, supervisory expenses, maintenance and repairs, etc., may vary greatly depending on the specific process in question.

OPERATION AND CONTROL OF EVAPORATORS

Many of the problems in operation and control of a given evaporator system will be specific to the application. However, all systems need to answer such questions as how to evaluate performance, how best to schedule periodic *boil outs* (cleaning), how to measure and control variables typified by temperature, pressure, fluid level, fluid flow rate, composition, etc., and how to detect faults in evaporator operation quickly and efficiently (45, 46).

At one time, the American Institute of Chemical Engineers issued a guide to performance testing of evaporators (47). The Instrument Society of America issues a book-length guide for the operation and control of evaporators (2), which covers, among other topics, such operational considerations as the optimum use of boil outs. Instrumentation is covered in detail, including variables, the choice of measurement and measurement method, and both design and implementation of the control system including digital techniques. This work contains both chapter bibliographies and a general bibliography.

Operation of an evaporator for minimum cost of production involves scheduling cleaning cycles to remove scale or deposits. Minimizing costs requires the balancing of increased evaporation rate against out-of-service time. Shutdown involves costs associated not only with cleaning, but emptying and refilling as well. The crux of such optimization is the function used to describe the evolution of scale formation with time.

REFERENCES

- Gupta, J.P. Evaporators and Vaporizers. *Fundamentals of Heat Exchanger and Pressure Vessel Technology*; Hemisphere: Washington, 1986; 187–211.
- Nisenfeld, A.E. *Industrial Evaporators: Principles of Operation and Control*; Instrument Society of America: Research Triangle Park, NC, 1985.
- Anthony, D. Improved Evaporators for Radioactive Waste, Chem. Eng. Progress **1983**, 58–63.
- Billet, R. VCH Verlagsgesellschaft MbH. *Evaporation Technology—Principles, Applications, Economics*; Weinheim: Germany, 1989.
- Kakac, S., Ed. *Boilers, Evaporators, and Condensers*; Wiley Interscience: New York, 1991.
- Cheremisinoff, N.P.; Cheremisinoff, P.N. *Heat Transfer Equipment*; Prentice Hall: Englewood Cliffs, NJ, 1993.
- Freese, H.L. *Evaporation, Fermentation and Biochemical Engineering Handbook*; Vogel, H.C., Ed.; Noyes Publications: Park Ridge, NJ, 1983; 227–276.
- Standiford, F.C. Thermal Design of Evaporators. *Perry's Chemical Engineers' Handbook*; Perry, R.H., Green, D.W., Maloney, J.O., Eds.; McGraw-Hill Book Company: New York, 1984; 10–34–10–38.
- Azbel, D. *Heat Transfer Applications in Process Engineering*; Noyes Publications: Park Ridge, NJ, 1984.
- Ulrich, G.D. *A Guide to Chemical Engineering Process Design and Economics*; John Wiley Sons: New York, 1984.
- Evaporators. *Thomas Register of American Manufacturers*; Thomas Publishing Company: New York, 1990; 5, 8584–8595.
- Chemical Processing Guide and Directory. Chem. Process. **1991**, 53 (15), 118–122.
- Myers, A.L.; Seider, W.D. *Introduction to Chemical Engineering and Computer Calculations*; Prentice Hall: Englewood Cliffs, NJ, 1976.
- Blackadder, D.A.; Nedderman, R.M. *A Handbook of Unit Operations*; Academic Press: London, 1971.
- Schilt, E.A. *Evaporation Transfer Operations in Process Industries*; Bhatia, M.V., Ed.; Technomic Publishing Co., Inc.: Lancaster, PA, 1983; 5, 267–299.
- Esplugas, S.; Mata, J. Calculator Design of Multistage Evaporators. *The Chemical Engineering Guide to Heat Transfer*; McNaughton, K.J., Ed.; Hemisphere: Washington, 1986; II, 261–265.
- Krishna, R. Book Review of: Evaporation Technology—Principles, Applications, Economics, by R. Billet. The Chemical Engineer **1989**, (462), 53.
- Cole, J. A Guide to the Selection of Evaporation Plant. The Chemical Engineer **1984**, (404), 20–23.
- Luyben, W.L.; Wenzel, L.A. *Chemical Process Analysis: Mass and Energy Balances*; Prentice Hall: Englewood Cliffs, NJ, 1988.
- Kessler, D.P.; Greenkorn, R.A. *Momentum, Heat, and Mass Transfer Fundamentals*; Marcel Dekker, Inc.: New York, 1999.
- Sandall, O.C.; Hanna, O.T.; et al. Heating and Evaporation of Turbulent Falling Liquid Films. AIChE J. **1988**, 34 (3), 502–505.
- Kocamustafaogullari, G.; Chen, I.Y. Falling Film Heat Transfer Analysis on a Bank of Horizontal Tube Evaporator. AIChE J. **1988**, 34 (9), 1539–1549.
- Han, J.; Fletcher, L.S. Falling Film Evaporation and Boiling in Circumferential and Axial Grooves on Horizontal Tubes. Ind. Eng. Chem. Process Design Dev. **1985**, 24, 570–575.
- Palen, J.W. Falling Film Evaporation in Vertical Tubes. *Heat Exchanger Technology*; Chisholm, D., Ed.; Elsevier Applied Science: London, 1988; 189–218.
- Steiner, D.; Ozawa, M. Flow Boiling Heat Transfer in Horizontal and Vertical Tubes. *Heat Exchangers: Theory and Practice*; Taborek, J., Hewitt, G.F., Afgan, N., Eds.; Hemisphere: Washington, 1983; 19–34.
- Chen, I.Y.; Kocamustafaogullari, G. An Experimental Study and Practical Correlations for Overall Heat Transfer Performance of Horizontal Tube Evaporator Design. *Heat Transfer Equipment Fundamentals, Design, Applications, and Operating Problems*; Shah, R.K., Ed.; The American Society of Mechanical Engineers: New York, 1989; 108, 23–32.
- Han, J.; Fletcher, L.S. Falling Film Evaporation and Boiling in Circumferential and Axial Grooves on Horizontal Tubes. Ind. & Eng. Chem. Process Design Dev. **1985**, 24, 570–575.
- Evans, F.L., Jr. *Ejectors, Equipment Design Handbook for Refineries and Chemical Plants*; Gulf Publishing Company: Houston, 1979; 1, 105–117.
- Evans, F.L., Jr. *Ejectors, Equipment Design Handbook for Refineries and Chemical Plants*; Gulf Publishing Company: Houston, 1979; 1, 105–117.
- Pennink, H. Cost Savings from Evaporator Vapor Compression. Chem. Eng. **1986**, 93, 79–81.
- Ramakrishna, P. Estimate Optimum Number of Effects for Multi-Effect Evaporation. Chem. Eng. **1987**, 94, 82.
- Evaporation Process, Center for Environmental Research Information, National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Capsule Report EPA/625/R-96/008, Cincinnati, OH, 1996.
- Rubin, F.L.; Moak, H.A. Heat-Transfer Equipment. *Perry's Chemical Engineers' Handbook*; Perry, R.H., Green, D.W., Maloney, J.O., Eds.; McGraw-Hill Book Company: New York, 1984; 11–3–11–59.

34. Knudsen, J.G.; Bell, K.J. Heat Transmission. *Perry's Chemical Engineers' Handbook*; Perry, R.H., Green, D.W., Maloney, J.O., Eds.; McGraw-Hill Book Company: New York, 1984; 11-3-11-59.
35. Chisholm, D. *Heat Exchanger Technology*; Elsevier Applied Science: London, 1988.
36. Taborek, J.; Hewitt, G.F.; *Heat Exchangers: Theory and Practice*; Hemisphere Publishing Corporation: Washington, 1983.
37. Shah, R.K. *Heat Transfer Equipment Fundamentals, Design, Applications, and Operating Problems*; The American Society of Mechanical Engineers: New York, 1989.
38. Peters, M.S.; Timmerhaus, K.D. *Plant Design and Economics for Chemical Engineers*; McGraw-Hill, Inc.: New York, 1991.
39. Hall, R.S.; Matley, J. Current Costs of Process Equipment. *The Chemical Engineering Guide to Heat Transfer*; McNaughton, K.J., Ed.; Hemisphere: Washington, 1986; II, 337-357.
40. Hall, R.S.; Matley, J. Current Costs of Process Equipment. *Chem. Eng.* **1982**.
41. Purohit, G.P. Estimating Costs of Shell-And-Tube Heat Exchangers. *The Chemical Engineering Guide to Heat Transfer*; McNaughton, K.J., Ed.; Hemisphere: Washington, 1986; II, 319-330.
42. Purohit, G.P. Costs of Double-Pipe and Multitube Heat Exchangers-Part 1. *The Chemical Engineering Guide to Heat Transfer*; McNaughton, K.J., Ed.; Hemisphere: Washington, 1986; II, 331-334.
43. Purohit, G.P. Costs of Double-Pipe and Multitube Heat Exchangers-Part 2. *The Chemical Engineering Guide to Heat Transfer*; McNaughton, K.J., Ed.; Hemisphere: Washington, 1986; II, 335-336.
44. Kumana, J.D. Cost Update on Specialty Heat Exchangers. *The Chemical Engineering Guide to Heat Transfer*; McNaughton, K.J., Ed.; Hemisphere: Washington, 1986; II, 358-360.
45. To, L.C. Robust Nonlinear Control of Industrial Evaporation Systems. World Scientific: River Edge, NJ, 1999.
46. Dukelow, S.G. *The Control of Boilers*; The Instrument Society of America: Research Triangle Park, NC, 1991.
47. Equipment Testing Procedures Committee. *Evaporators: A Guide to Performance Evaluation*; AIChE Equipment Testing Procedure American Institute of Chemical Engineers: New York, 1978.

EUROPEAN AGENCY FOR THE EVALUATION OF MEDICINAL PRODUCTS (EMA)

David M. Jacobs

David M. Jacobs Consulting, Basel, Switzerland

INTRODUCTION

History of the European Union

After the Second World War, the idea of a “United States of Europe” was promulgated and in 1957, the Treaty of Rome was signed instituting the European Economic Community (EEC) between six countries (Germany, Belgium, France, Italy, Luxembourg, and the Netherlands). In 1973, Ireland, the United Kingdom, (UK) and Denmark joined the EEC. In 1974, heads of state and government decided that a European Parliament should be elected by direct universal suffrage and that it would meet regularly as the European Council (EC) to deal with community affairs and political cooperation. Greece acceded to the EC in 1981, and Spain and Portugal joined in 1986. In 1992, the 12 Foreign Affairs Ministers signed the Maastricht Treaty instituting the European Union (EU), which also included the four freedoms of labor, capital, goods, and services. Finally, Austria, Finland, and Sweden joined the EU in 1995. The Treaty of Amsterdam, which entered into force on May 1, 1999, made further institutional changes such that no draft text can become law without the formal agreement of both the European Parliament and the Council. Thus, the EU now consists of 15 member states, each of which has its own national government and legislative bodies.

The EEC has three important powers:

1. It adopts “European laws” that apply in the 15 countries (“directives” and “regulations”).
2. It disposes of a budget to finance certain programs carried out in its member states.
3. It signs international agreements on cooperation or trade.

All these decisions are taken by common institutions sitting in Brussels, Strasbourg, and Luxembourg. Since 1974, the EC has brought together heads of state and government of the 15 member states, as well as the president of the Commission, to set key guidelines and political goals and to arbitrate on questions for which agreement has not been found within the EU Council of

Ministers. Each member country presides over the Council for a six-month period.

EUROPEAN UNION INSTITUTIONS AND LEGISLATIVE INSTRUMENTS

The European Commission, representing the Community’s interests, draws up common projects and, after a decision has been taken by the EU Council of Ministers, sees that they are properly implemented. It is directed by 20 commissioners and is assisted in its work by a permanent staff of 17,000, most of who are based in Brussels. It is independent of the governments but is subject to control of the European Parliament. It implements common policies and negotiates international agreements. It may bring an action before the European Court of Justice should Community laws not be respected by the member states. It is here, within the Pharmaceuticals and Cosmetics Unit of the Directorate-General for Enterprise (formerly the Directorate-General for Industry, DG III), that European legislation on medicines is drawn up and implemented.

The European Parliament is made up of 626 deputies who are elected by direct universal suffrage every 5 years. It examines all proposals for European directives and regulations, which it may accept, modify, or refuse. It supervises the work of the European Commission, which it can dismiss with a motion of censure, and it votes the annual Community budget.

The EU Council of Ministers meets in order to adopt proposed European directives and regulations in light of the advice given by the European Parliament. The ministers convene depending on the subject that the Council is dealing with and according to their areas of competency (i.e., the ministers of health of the 15 member states are present for a Council dealing with questions of drug regulation or health.) The country presiding over the EC holds the presidency of the EU Council of Ministers. Due to the principle of subsidiarity, Community legislation is only introduced on points of common interest and in order to further the aim of a balanced and dynamic Europe.

The Council of Ministers can adopt several types of legislation, which are more or less restrictive:

- “Regulations” are binding and directly applicable to all citizens.
- “Directives” are binding on all citizens but indirectly (i.e., after they have been “transposed” into the laws of each country).
- “Decisions” are binding and directly applicable but only to the institutions, bodies, businesses, or citizens specifically named.
- “Recommendations,” “advisory opinions,” and “resolutions” are consultative or guidance texts addressed to the states.

EU LEGISLATION FOR PHARMACEUTICAL AND VETERINARY PRODUCTS

The foundation of European pharmaceutical legislation is Directive 65/65/EEC (1), which when promulgated in 1965, applied only to the initial six member states. In this directive, the definition of a medicinal product is given and the data required to obtain approval is described. This original directive is continually updated, amended, and supplemented with subsequent legislation, but remains the basis of pharmaceutical legislation.

Ten years following the first direction, three new directives sought to further promote public health and the free movement of medicinal products within the community. Directive 75/318/EEC (2) set analytical, pharmacotoxicological, and clinical standards for testing proprietary medicinal products. Directive 75/319/EEC (3) established the Committee for Proprietary Medicinal Products (CPMP) and its partial mutual recognition procedure, while Directive 75/320/EEC (4) set up a Pharmaceutical Committee to examine problems in implementing the pharmaceutical directives.

In the years that followed, cooperation between national health authorities at EU level was further encouraged. Two Directives, 83/570/EEC (5) and 87/22/EEC (6), set up the Multistate procedure and the Concertation procedure. These procedures provided a mechanism for exchange of information on all aspects of product licensing between member states and made it easier for national licensing authorities to recognize each other's decisions. In the Concertation procedure, the CPMP was charged with forming an opinion on the feasibility of an application, which, however, was not binding on the member states' national authorities. The Multistate procedure was based on the principle of

recognition of an approval in one member state by the national health authorities in other member states.

The European Agency for the Evaluation of Medicinal Products (EMA) was established by Council Regulation (EEC) No 2309/93 (7) of July 22, 1993, with London chosen as its seat by decision of the Council on October 29, 1993. It began operation on February 1, 1995. Regulation 2309/93 also established the legal basis for a single community-wide centralized procedure for the approval of medicinal and veterinary products.

Simultaneously, Directive 93/39/EEC (8) amended Directives 65/65/EEC (1), 75/318/EEC (2), and 75/319/EEC (3) and established the Decentralized Procedure (commonly known as the Mutual Recognition Procedure).

THE EMA

Mission

The Mission of the EMA is to contribute to the protection and promotion of public and animal health by:

- Mobilizing scientific resources from throughout the EU to provide high quality evaluation of medicinal products, to advise on research and development programs, and to provide useful and clear information to users and health professionals

- Developing efficient and transparent procedures to allow timely access by users to innovative medicines through a single European marketing authorization

- Controlling the safety of medicines for humans and animals, in particular through a pharmacovigilance network and the establishment of safe limits for residues in food-producing animals

Structure

The European system is based on cooperation between the national health authorities of the member states and the EMA. The EMA acts as a focal point of a network that coordinates the scientific resources made available by the member states. This partnership between the EMA, national health authorities, and the EU institutions is crucial to the functioning of the European authorization procedures.

A Management Board supervises the EMA, while its scientific activities are largely carried out through its two scientific committees and their working parties. The Board, scientific committees, and their working parties are supported by the EMA secretariat, headed by an Executive Director.

The Management Board is made up of two representatives from each member state, from the European Parliament and from the European Commission. Representatives of Iceland and Norway, who are members of the European Economic Area (EEA) but not of the EU, also attend meetings of the Board. As of January 1, 2000, these countries formally joined the EMA. The Management Board appoints the Executive Director, and approves the budget and work program each year. On the recommendation of the European Parliament, it gives discharge to the Executive Director for the implementation of the budget.

The principal scientific bodies of the EMA are the CPMP and the Committee for Veterinary Medicinal Products (CVMP). They are made up of two members from each member state as well as from Norway and Iceland, and are appointed to give independent scientific advice to the EMA. The EMA Secretariat comprises four units: administration, evaluation of medicines for human use, technical coordination, and evaluation of medicines for veterinary use.

The Administration Unit is responsible for carrying out administrative and financial functions to ensure that the Secretariat and staff are able to perform their statutory tasks under satisfactory conditions and thus, has two subsections for personnel, budget and facilities, and for accounting.

The Unit for the Evaluation of Medicines for Human Use is responsible for the following:

- management and follow-up of marketing authorization applications under the centralized procedure;
- postmarketing maintenance of authorized medicinal products;
- management of community referrals and arbitrations arising from the mutual recognition procedure; and
- provision of support to European and international harmonization activities of the CPMP and its working parties.

This unit consists of three subdivisions or sectors: for regulatory affairs and pharmacovigilance, for biotechnology and biologicals, and for new chemical substances.

The Unit for the Evaluation of Medicines for Veterinary Use is responsible for the following:

- management and follow-up of marketing authorization applications under the centralized procedure;
- management of applications for the establishment of maximum limits for residues of veterinary medicinal products that may be permitted in foodstuffs of animal origin;
- postmarketing maintenance of authorized medicinal products;

- management of community referrals and arbitrations arising from the mutual recognition procedure; and
- provision of support to European and international harmonization activities of the CVMP and its working parties.

It has two sectors: for CVMP and veterinary procedures, and for safety of veterinary medicines.

The Technical Coordination Unit is responsible for providing logistical support to both human and veterinary medicine evaluation activities as well as a number of general services to the EMA, including document management, conference services, and information technology support. It has four sectors: for inspections, for document management and publishing, for conference services, and for information technology. The sector for inspections coordinates the work of inspectors, the implementation of mutual recognition agreements, and the monitoring of medicines authorized in the community. It provides the secretariat of the Quality Working Party and coordinates the Agency's quality management program.

EMA Scientific Committees

The CPMP and the CVMP are the scientific committees set up to facilitate the adoption of scientific decisions between member states on the authorization of medicinal products on the scientific criteria of quality, safety, and efficacy.

When working for the EMA, members of the CPMP and CVMP act independently of their nominating member state. The scientific committees are aided by a network of approximately 2300 European experts, nominated by the national competent authorities of the member states on the basis of proven experience in the assessment of medicinal products. Experts may serve on working parties or expert groups of the CPMP or CVMP.

The scientific committee decides the appointment of rapporteurs and corapporteurs, i.e., those members of the CPMP or CVMP who take the lead in reviewing a dossier. The committees are required to ensure that all members undertake the role of rapporteur or corapporteur. Compensation is provided to national competent authorities for the services provided by committee members or European experts at the specific request of the agency.

THE CPMP AND THE EVALUATION OF MEDICINAL PRODUCTS FOR HUMAN USE

A CPMP member acts as rapporteur or corapporteur for centralized procedures and the CPMP gives an official

opinion on whether an application for marketing is approvable or not. The EMA is intimately involved in the management of this procedure up to the issue of the marketing authorization. The EMA's involvement also includes preparation of the CPMP opinion in all 11 official EU languages. Quality management standards have been implemented for the preparation of scientific advice and opinions, and a tracking system throughout the life cycle of centrally authorized products has been developed. Postauthorization, variations, and extensions to the license may be submitted and rapporteurs play a major role with these maintenance activities. There is also ongoing activity with regard to adverse drug reaction (ADR) reporting, periodic safety update reports (PSURs), and other follow-up measures. Rapporteurs and corapporteurs are particularly involved in urgent safety restriction procedures.

Pharmaceutical sponsors may seek advice on their development programs from the CPMP. The CPMP has set up a scientific advice review group to strengthen and widen CPMP input and to guarantee the availability of proper expertise. A standard operating procedure for the giving of scientific advice by the CPMP for innovator medicinal products has been adopted.

Working Parties

The CPMP and CVMP each have four working parties, as well as a joint CPMP/CVMP Quality Working Party. There is also an EMA working party on Herbal Medicinal Products.

The CPMP working parties are concerned with biotechnology, efficacy, safety, and pharmacovigilance. The CVMP working parties are concerned with safety of residues, immunological veterinary products, veterinary pharmacovigilance, and efficacy.

These working parties produce position papers, points to consider, notes for guidance, and joint CPMP/CVMP/International Conference on Harmonization (ICH) Guidelines that provide up-to-date scientific opinions on matters of current interest to all member states and pharmaceutical and veterinary manufacturers.

Biotechnology working party

This working party considers aspects of the manufacture and control of biotechnological and biological medicinal products and is also involved in the provision of scientific advice. For example, workshops were held recently on the application of assays for markers of transmissible spongiform encephalopathies (TSE) and on the potential risk of transmitting new variant Creutzfeldt–Jakob Disease (nv-CJD) through plasma-derived medicinal products.

Efficacy working party

Clinical trial methodology and guidelines for special disease-related therapeutic fields are discussed in this party. In cooperation with other working parties, guidance on modified release oral and transdermal dosage forms, on pharmacokinetics, and on clinical investigation of new vaccines, gene therapy, and cell-cultured influenza vaccines has been given.

Pharmacovigilance working party

This working party considers safety-related issues at the request of both the CPMP and national authorities, resulting in the harmonization of the summary of product characteristics and package leaflets of marketed products. Regular video conferences are held with the U.S. Food and Drug Administration (U.S. FDA) to discuss issues of mutual interest. A pilot project was started for the electronic transmission of individual case safety reports with a restricted number of participants from national authorities and marketing authorization holders.

Safety working party

Preclinical and safety issues are discussed, and in cooperation with the Biotechnology Working Party, a note for guidance on the quality, preclinical, and clinical aspects of gene transfer products was produced recently.

Ad hoc and other groups

Ad hoc groups on excipients, Lipodystrophy, and antiretroviral medicinal products have been formed. A multidisciplinary group has been set up to evaluate medicinal products containing thiomersal with a view to limiting exposure to mercury and organomercurial compounds.

Cooperation with Competent Authorities

European Monitoring Centre for Drugs and Drug Addiction (EMCDDA)

The EMA has supported the development of guidelines on risk assessment of new synthetic drugs.

International Conference on Harmonization (ICH and VICH)

The EMA, as one of the six partners in the ICH process, is intimately involved in production and update of ICH guidelines. The Unit for the Evaluation of Medicinal Products for Human Use supports the Steering Committee, the EU topic leaders, the CPMP, and the various working parties in the preparation, review, and administration of ICH guidelines. Similarly, since establishment of the VICH

in 1996, the Unit for Evaluation of Veterinary Medicinal Products supports the Steering Committee and the CVMP as well as the various working parties in this initiative.

Central and Eastern Europe

Many central and eastern European countries (CEEC) are candidates for accession to the EU. The candidates are Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia, and Cyprus. In order to help pharmaceutical registration authorities in these countries prepare for EU membership, a Collaborative Agreement of Drug Regulatory Authorities of European Union Associated Countries (CADREAC) was formed. In addition to this agreement, a simplified procedure for the recognition of centrally authorized medicinal products by the national authorities of CEEC was established. The procedure is optional and is initiated at the request of the marketing authorization holder in the EU.

In addition, under the auspices of a pharmaceutical Pan-European Regulatory Forum (PERF) set up by the European Commission, the EMA administers and provides executive assistance to CEEC and EU regulators in the conduct of working groups and training sessions in order to facilitate the adoption of common technical requirements. Topics include the implementation of Community legislation, pharmacovigilance, and the assessment of dossiers for marketing authorization for quality safety and efficacy.

Meanwhile, to help eliminate technical barriers to pharmaceutical trade with the CEECs, protocols to permit mutual recognition of good manufacturing practice compliance for medicinal products are being negotiated.

THE CVMP AND THE EVALUATION OF MEDICINAL PRODUCTS FOR VETERINARY USE

The CVMP operates in a similar fashion to the CPMP and is heavily involved in the review of centralized procedures for veterinary products. The CVMP has developed a broad range of new guidelines to assist applicants on topics related to research and development and for which no guidance existed previously.

The Unit for the Evaluation of Medicinal Products for Veterinary Use has also been involved in the PERF initiative as well as other activities related to implementation of Community legislation and the quality of medicinal products.

When the EMA opened in January 1995, more than 600 "old" substances remained for which maximum residue limits (MRLs) had to be established. The

assessment of these products was completed before the January 2000 deadline.

Veterinary Working Parties

Similarly to the CPMP Working Parties, the Efficacy, Safety of Residues, Immunologicals, and Pharmacovigilance Working Parties develop guidelines for the testing and reporting requirements of studies for products for veterinary use.

INSTITUTIONAL PARTNERS

The major contact within the services of the European Commission is the Pharmaceuticals and Cosmetics Unit of the Directorate-General for Enterprise; however, there is also continued exchange of information with the Directorate-General for Health and Consumer Protection. Other contacts include the Directorate-General for Research and the Joint Research Center.

European Technical Office for Medicinal Products (ETOMP)

The European Commission Joint Research Center has established a technical office at the EMA responsible for the management of a telecommunications network and other computer technologies to facilitate the dissemination of information on medicinal products. It also manages the EMA Internet website. A new mechanism for the secure exchange of documents through the Internet has been put in place to facilitate, among other things, the transmission of individual case safety reports within the pilot project on pharmacovigilance between EMA, national authorities, and the pharmaceutical industry.

The European Union drug regulatory authorities' network (EudraNet) is an internetworking service provided to EU medicinal regulatory authorities in collaboration with the European Commission Directorate-General for Industry. Part of the EudraNet is accessible to industry and the general public.

Joint Interpreting and Conference Service (JICS)

The JICS of the European Commission serves the institutions of the EU, as well as the decentralized agencies and bodies located in EU member states. A representative of the JICS is based at the EMA to coordinate translation and conference needs. A glossary of specialized and technical EMA terms to assist interpreters at EMA meetings is being developed.

The European Department for the Quality of Medicines (EDQM)

European Pharmacopoeia (EP)

The EP was founded by Belgium, France, Germany, Italy, Luxembourg, Netherlands, Switzerland, and the United Kingdom in 1964, under a Council of Europe Convention, to help standardize their national pharmacopoeias. The EP now has 26 signatories (15 member states, the European community, and 10 other European countries). Its monographs have force of law, replacing the old national pharmacopoeias. Directive 75/318/EEC requires EU pharmaceutical manufacturers to use these monographs when compiling marketing authorization applications. The EMA participates in the work of the EP Commission as part of the EU delegation.

European Network of Official Medicines Control Laboratories (OMCL)

This is a joint project between the EU and the Council of Europe to allow the coordination of laboratory controls between the EU and EFTA members. In 1999, a contract was signed between the EMA and the EDQM to organize sampling and testing of centrally authorized medicinal products by the OMCL network.

EUROPEAN APPROVAL PROCEDURES

There are two European procedures for obtaining a marketing authorization in more than one country belonging to the EU. These are the Centralized Procedure and the Decentralized or Mutual Recognition Procedure.

Centralized Procedure

The Centralized Procedure must be used for biotechnology products and can be used for so-called high technology products as well as for new active pharmaceutical ingredients (i.e., products that have never before been approved for marketing). The Centralized Procedure is laid down in Council Regulation (EEC) N° 2309/93 (7) and Directive 93/41/EEC (9).

In the Centralized Procedure, one license to market the drug in the entire EU is issued and in principle there is only one evaluation of the dossier. In fact, both a rapporteur and corapporteur are appointed, and each assesses the dossier with its own team. The rapporteur and corapporteur are members of the CPMP who are assigned to a particular

dossier by the CPMP. Each member is obliged to act as rapporteur or corapporteur.

Before submission of the dossier, the Sponsor Company contacts the CPMP or CVMP to announce its intention to make a registration submission and to request appointment of a rapporteur. If, as is usually the case, the Sponsor has had contact with national health authorities, it may request that a particular CPMP or CVMP member be appointed as rapporteur. The CPMP/CVMP is not obliged to follow this request, but in many cases either the rapporteur or the corapporteur is the CPMP/CVMP member requested.

After submission, the rapporteur and corapporteur have 120 days to perform their review and to write a draft assessment report. The two assessments are then discussed by the parties and a list of outstanding issues is sent to the sponsor, at which point the clock is stopped. When the answers have been received, the rapporteur has another 30 days to finalize the assessment report, which is sent to the CPMP or CVMP. CPMP/CVMP members also receive a copy of Part I of the dossier and may request the full dossier. After a total of 210 days, the CPMP or CVMP delivers an opinion: favorable or unfavorable.

If the opinion is favorable, the second stage of the procedure, the decision-making process, begins. During the decision-making process, the Commission Services check that the marketing authorization complies with community law and turn the agency opinion into a binding decision for all the member states. Should the CPMP decision be unfavorable, the sponsor may appeal and a second CPMP opinion must be prepared within 60 days.

The agency sends the Pharmaceutical Unit of the Commission its opinion in all 11 community languages together with the Summary of Product Characteristics (SPC), the particulars of the manufacturing authorization holder responsible for batch release and of the manufacturer of the active substance, as well as the labeling and package leaflet. The commission has 30 days to prepare a draft decision. During this period, various commission directorates-general are consulted and are able to give their opinions.

The draft decision is then sent to the Standing Committee on Medicinal Products or the Standing Committee on Veterinary Products for their opinions. Should there be detailed opposition from a member state to the draft commission decision, the standing committee can refer it back to the CPMP if the opposition is scientific. If the matter is nonscientific, a vote is taken and the council decision is made on the basis of a qualified majority. Each member state has a different number of votes depending on size and importance, and the majority of votes must be in favor. If there is no opposition within 30 days, the draft decision is forwarded to the Commission

Secretariat-General for adoption, enabling the Commissioner for Enterprise and the Information Society to issue the final decision. The final decision is published in the Official Journal of the European Communities.

Decentralized Procedure

The Decentralized Procedure is made on the basis of mutual recognition. Council Directive 93/39/EEC (8) has been implemented in all member states in accordance with Directives 65/65/EEC (1) and 75/319/EEC (3). The sponsor makes a submission to the national health authority of one member state, with a request to assess the dossier for mutual recognition. Within 210 days, the so-called Reference Member State (RMS) must approve the application, prepare an assessment report, and agree on an SPC. The clock may be stopped to obtain further information during this time.

The mutual recognition submission can then be made to any number of the other member states, and the RMS sends a copy of the assessment report to the concerned member states (CMS). Within 90 days, member states must raise serious objections and if there are none, each CMS issues a national marketing authorization with an identical SPC.

To facilitate the mutual recognition procedure, a Mutual Recognition Facilitation Group (MRFG) and a Veterinary Mutual Recognition Facilitation Group have been set up, although this was not foreseen in the original directive. These groups meet one day before each CPMP/CVMP meeting. The objections raised are discussed within the group and the RMS tries to reach agreement on the approval possibilities of the dossier and the most appropriate labeling. If necessary, breakout sessions with the sponsor can be held to finalize labeling details.

Should no agreement be reached within the MRFG/VMRFG, the matter is sent to the CPMP/CVMP for an opinion. Thereafter, the procedure is similar to that during the centralized procedure—the end result being a commission decision after which national licenses must be issued within 30 days.

Referrals and Arbitration

A sponsor company or a national authority may make referrals to the EMA under Article 10 of Directive 75/319/EEC, in order to harmonize the summary of product characteristics in all member states for products previously approved under national legislation.

Similarly, where there are public health concerns as a result of pharmacovigilance data, nationally authorized products or products authorized by the mutual recognition procedure may be referred under Articles 12 or 15 of Directive 75/319/EEC. The CPMP/CVMP gives an opinion on variation, suspension, or withdrawal of the marketing authorization in such cases.

REFERENCES

1. Official Journal of the European Communities N° 022, **Feb 9, 1965**, 0369–0373.
2. Official Journal of the European Communities N° L 147, **June 9, 1975**, 0001–0002.
3. Official Journal of the European Communities N° L 147, **June 9, 1975**, 0013–0022.
4. Official Journal of the European Communities N° L 147, **June 9, 1975**, 0023–0024.
5. Official Journal of the European Communities N° L 332, **Nov 28, 1983**, 0001–0002.
6. Official Journal of the European Communities N° L 15, **Jan 17, 1997**, 0038–0041.
7. Official Journal of the European Communities N° L 214, **Aug 24, 1993**, 0001–0002.
8. Official Journal of the European Communities N° N L 214, **Aug 24, 1993**, 0022–0030.
9. Official Journal of the European Communities N° L 214, **Aug 24, 1993**, 0040.
10. EU <http://europa.eu.int>, <http://europa.eu.int/eur-lex>.
11. Pharmaceuticals in the EU. <http://pharmacos.eudra.org>, <http://www.eudraportal.eudra.org>, <http://emea.eu.int>, <http://perf.eudra.org>.
12. National Agencies. <http://heads.medagencies.org>, <http://hevra.org>.
13. ICH. <http://www.ifpma.org/ich1.html>, <http://vich.eudra.org>.
14. European Pharmacopoeia. <http://www.pheur.org>.

EQUIPMENT CLEANING

Destin A. LeBlanc

Cleaning Validation Technologies, San Antonio, Texas, U.S.A.

INTRODUCTION

REGULATORY BACKGROUND

GMP Issues

Cleaning of process equipment has been part of the good manufacturing practices (GMPs) for pharmaceutical manufacturing for many years (1, 2). This has included recommendations for written procedures, cleaning logs, and appropriate design of equipment to facilitate cleaning. Good cleaning practices are necessary to preserve the safety and efficacy of the manufactured drugs and drug products. Possible consequences of inadequate cleaning include cross-contamination (the presence of one drug active in another drug product at an unacceptable level), the presence of foreign material (e.g., a cleaning agent, solvent, or excipient from another drug product), the presence of microbial contamination (numbers and/or species of microbes), or the presence of endotoxins (particularly in parenteral or ophthalmic products). The presence of such contaminants in a drug product may pose safety problems depending on the level of the contaminant. Such contaminants may also affect the efficacy of a drug product; effects could include modifying the bioavailability of the active, the dissolution time of tablets, or the stability of the finished drug. Needless to say, failing to follow GMPs relating to cleaning processes also renders the product “adulterated” and subject to regulatory action.

Expectation of Validation

What is new since about 1990 is the regulatory expectation that certain cleaning processes in pharmaceutical manufacturing be validated. Validation of cleaning processes had been discussed in numerous articles prior to that time (3, 4). However, issues with drug product contamination due to poorly controlled cleaning processes (5), culminating in the Barr Laboratories decision (6), brought this issue to the forefront and clearly established the FDA’s authority to require the validation of cleaning processes. As cleaning is a process, the principles of process

validation apply to the cleaning process. The Barr decision was followed soon by FDA cleaning validation guidance documents in 1992 (7) and 1993 (8). In 1996, the FDA proposed amendments to the GMPs which clearly defined (if approved) validation of cleaning processes as a GMP requirement (9). The U.S. FDA took the lead in requiring validation of cleaning processes, and other agencies also issued similar requirements. This includes the Pharmaceutical Inspection Cooperation Scheme (PIC/S) document PR-1-99 (10), the Draft Annex 15 to the EU GMPs (11), and the Canadian Therapeutic Products Programme “Cleaning Validation Guidelines” (12). Although the initial emphasis was on cleaning validation related to finished drug products, additional guideline documents clarified that cleaning validation should be considered for active pharmaceutical ingredients (13, 14) and for pharmaceutical excipients (15).

Applicability of Cleaning Validation

It should be noted that these cleaning validation requirements apply only to *critical* cleaning processes. Although GMPs require the cleaning of (and cleaning SOPs for) floors, walls, and the outside of process vessels, such processes are not considered critical cleaning processes. The processes that are critical generally include processes for cleaning product-contact surfaces of equipment or utensils. It is these product-contact surfaces that have the possibility of *directly* contaminating the next product made in the same equipment. In addition, the cleaning of nonproduct-contact surfaces that could reasonably *indirectly* contaminate subsequently manufactured products should also be considered for cleaning validation. For example, some companies have, either on their own or because of regulatory requirements, validated the cleaning of internal surfaces of lyophilizers used for production. On the other hand, validation of cleaning between lots of the same product is not necessarily a requirement (16). This is based on the fact that cross-contamination of the active is not an issue. However, other concerns such as contamination with degradation products, with cleaning agent residues, or with microorganisms may suggest that such cleaning is critical, and therefore should be validated.

CLEANING PROCESSES

The overall cleaning process comprises the soiled equipment, a cleaning method with the associated cleaning equipment, a cleaning agent(s), and process parameters (time, temperature, etc.). These factors should all be captured in a cleaning standard operating procedure (SOP).

Equipment Design

As regards the equipment to be cleaned, this depends on both the equipment itself and the residues to be removed. Ideally the equipment to be cleaned has been designed with cleaning in mind. Design characteristics may be different for manual versus automated cleaning. The equipment design may also affect the extent of disassembly of the equipment as part of the cleaning process. Design characteristics that help maximize cleaning include minimizing deadlegs, minimizing cracks and crevices where soils can be trapped, improving accessibility of the cleaning solution to difficult-to-clean portions of the equipment, and providing adequate drainage. Materials of construction should be selected for the equipment based on the expected cleaning process; such materials may affect both chemical (such as acids, alkalis, and solvents) and physical (such as temperature) compatibility. Although it is difficult to cover all the possibilities because of the varieties of equipment to be cleaned, the principle is the same—the selection of the cleaning process will be limited by the original equipment design (except to the extent that such designs can be modified).

Cleaning Methods

Although cleaning methods are sometimes divided into clean-in-place (CIP) and clean-out-of-place (COP) applications, it may be more useful to consider two significant features of cleaning methods to provide broad categorizations of cleaning processes.

Extent of automation

One factor involves the extent of automation. At one extreme of the “automation” continuum is the fully automatic process—no operator intervention is required for preparation of the cleaning solution, for the cleaning cycle, or for any disassembly or reassembly. The only operator requirement might be pushing a button at the beginning of a cycle, recording the cleaning process in the cleaning log book, and perhaps a visual examination as

part of the monitoring procedure at the end of cleaning (and/or before the manufacture of the next product). At the other extreme is the fully manual process. The operator is required for preparation of the cleaning solution, for isolating the system to be cleaned, for applying the cleaning solution, perhaps for applying mechanical action through brushes or wipers, for rinsing the system, and for monitoring process parameters (including the timing of all events). It should be clear that in between these two extremes are various semiautomated processes, which could cover a broad continuum.

Extent of disassembly

A second continuum for cleaning processes involves the degree of disassembly (and consequent reassembly). At one extreme is equipment that requires no disassembly at all (true “clean-in-place”). At the other extreme is equipment that requires disassembly of each component part for cleaning. Disassembly (and reassembly) is preferably avoided for several reasons, including the time it adds to the overall cleaning process (equipment downtime), the concern over damage to the equipment because of stresses during the disassembly/reassembly process, and the concern over incorrect reassembly. However, it should be recognized that there are situations in which partial or complete disassembly of equipment might be required. This includes the removal of filters prior to cleaning, or the opening of a process vessel for placement of a spray device of a portable CIP system into the vessel.

Simplification of cleaning processes

Design of a cleaning process must be taken into consideration not only the nature of the process itself but also the engineering design of the equipment to be cleaned, the various products manufactured in the equipment (such products become “soils” to be cleaned at the end of manufacturing), and the cleaning process parameters (discussed in more detail later). In many pharmaceutical facilities, the objective is to make the process as simple and universal as possible so that one cleaning SOP can be used either for all manufactured products made in the same equipment or for all equipment cleaned in the same process. This simplifies documentation and training and may (because of grouping or bracketing strategies) simplify validation.

Cleaning process steps

The general steps or stages of most cleaning processes involve the following:

Disassembly and isolation: This involves preparation of the equipment for application of the cleaning

solution(s). Disassembly may involve complete disassembly for washing individual parts elsewhere, or may involve partial disassembly, such as removal of filters for separate cleaning elsewhere. The preferred technique for cleaning is to isolate the equipment (or parts thereof) and then clean the entire isolated portion. In a validated process, it is difficult to clean only one portion of a piece of equipment without isolating it (for example, trying to clean a storage vessel only to the level of product in the vessel; the entire vessel, including the vessel dome, should be cleaned).

Prewashing (or prerinsing): In aqueous cleaning, this involves flushing all parts of the system with water (usually at ambient temperature) to physically remove soils that can be readily removed by a flowing water stream. The purpose of the prewash is to minimize soils on the surface for the cleaning step. In this manner, the action of those cleaning agents in the cleaning step are focused on residues that are more tenaciously bound to the surface. In biotechnology manufacture or any manufacture that involves proteinaceous deposits, a second objective of this prewash is to prevent “setting” of those proteinaceous deposits when they are immediately cleaned with a hot water solution. If the cleaning process uses a CIP system, the prewash step is usually a “once through to drain” rather than a recirculating process. The objective is to immediately remove loosely bound soils and discharge them from the equipment rather than to spread them evenly over all equipment surfaces (which would occur to a certain extent in a recirculating system).

Washing: This involves application of the cleaning solution (which may be plain water, but which usually involves some cleaning agent) to all equipment surfaces to effectively remove those soils not removed by the prewash. The washing step may involve continuous application of fresh cleaning solution (such as in a non-recirculating CIP system or in a manual application using a high pressure spray hose), a recirculating application of the cleaning agent in which partially “depleted” cleaning solution is reapplied to surfaces (as in a recirculating CIP system or an automatic machine parts washer), or a static soak of equipment or utensils. The purpose of the washing step is to either dissolve, solubilize, emulsify, suspend, or chemically affect the soils on the surface so they can be readily removed from the equipment either in the washing step (in a non-recirculating process) and/or the rinse step.

Rinsing: The rinsing step is designed to remove both washing solution and associated soluble, solubilized, emulsified, or suspended soils from the equipment. For solvent cleaning, the rinsing solution is usually a fresh application of the same solvent used for the washing step. In aqueous processing, the rinsing solution is usually water. The rinsing step should usually be a non-recirculating

application of the rinsing solution. A general rule of thumb followed for finished product manufacture involving aqueous-based drug products is that the quality of water used in the final rinse should be at least as good as the quality of water used in the manufacture of the next product. The rationale behind this is that any water contaminants in the final rinse left behind on equipment surfaces by the final rinse are identical in quality to water used in manufacturing of the next product. If the drug does not contain water, such as in the manufacture of a synthetic organic active substance, there may be other considerations for the selection of the quality of the final rinse water. A common practice in bulk pharmaceutical manufacture, suitable for most applications where aqueous cleaning is performed, is to use deionized water as a final rinse.

Drying: Drying is an optional step. One factor in whether drying should be done is the time period before the next use of the equipment. Equipment that is to be used immediately (within a few hours) may not have to be dried, particularly if the equipment is effectively drained to minimize any dilution effect of residual rinse water or solvent. However, the effect of residual water on microbial proliferation during extended storage is a significant issue. Options for drying include heated (and optionally filtered) air and the use of a final alcohol/water rinse. The final alcohol/water rinse may also further reduce the bioburden due to the antimicrobial action of the alcohol. This use has to take into consideration the flammability of such a mixture.

Reassembly and storage: These should be part of the cleaning SOP. “Reassembly” may involve removal of temporarily installed cleaning equipment (e.g., the spray device of a portable CIP unit) or reassembly of equipment parts themselves. If the equipment is to be stored for a significant time before reuse, critical elements for storage include whether the equipment is dry, physical protection of equipment from recontamination by use of items such as plastic wrapping, and the room conditions (air quality, temperature, and humidity) where the equipment is stored. Typically it is expected that stored equipment will be tagged as cleaned with an expiration (or “use by”) date. Expiration dates for stored equipment are established based on the possible routes and extent of recontamination during storage. For storage, the focus of regulatory agencies is microbial contamination; however, other types of contamination should also be evaluated.

Automated CIP systems

The discussion of CIP processes deserves special comment because of industry trends to use CIP systems. As used in a broad sense, CIP refers to any system in which the equipment is cleaned with no or minimal disassembly. In a

more narrow sense (and in this sense it is more commonly used now), CIP is used to refer to systems in which one or more spray devices is placed in the equipment to be cleaned. A control unit, comprising a pump, associated valves, and a PLC (programmable logic controller), pumps a cleaning solution from a storage tank through the spray device(s). The spray device(s) is engineered and placed so that solution is either directly sprayed or else sprayed so that the solution cascades down the equipment sidewalls to cover all surfaces of the equipment for effective cleaning. In a non-recirculating CIP system, the cleaning solution passes once through the process vessel and associated piping, and then goes to drain. In a recirculating system, the cleaning solution passes through the process vessel and associated piping and then back to the cleaning solution storage tank. It is then pumped through the spray device again, for multiple passes.

The spray device may be either permanently mounted in the process vessel, or installed for cleaning and then removed for product manufacture. Spray devices may be stationary. Stationary spherical devices, the most common type, are called “spray balls.” Spray balls are usually stainless steel hollow spheres in which holes are drilled. The placement of the holes is designed to provide adequate coverage for the vessel to be cleaned. Stationary spray devices are usually considered “sanitary” because they are self-draining. The other type of spray device is a dynamic (or rotating) spray device. These are similar in principle to a stationary spray device (they are designed to distribute cleaning solution over all surfaces of the process vessel) except that dynamic spray devices will rotate in one or more planes to provide more even distribution of the cleaning solution. Dynamic spray devices also typically operate at high spray pressures, so that the impingement of the cleaning solution on the vessel surfaces provides more mechanical energy to help dislodge residues. Dynamic spray devices are typically not mounted permanently because they are not self-draining (and thus sanitary); however, some newer dynamic devices are claimed to be sanitary.

A key to operation and validation of a spray device is to perform a “coverage” test, such as a “riboflavin test.” Riboflavin is readily water soluble, and also fluoresces under an ultraviolet light. Such a test involves spraying the interior surfaces of the equipment with a dilute solution of riboflavin. A short CIP rinse cycle is then performed using just water in a non-recirculating mode. Following this, the interior surfaces are examined using an ultraviolet light source. If any surfaces fluoresce green, it is an indication that solution coverage in those areas may be inadequate. Poor coverage should require a redesign of the spray device system, either by adding additional spray devices,

using a different spray device, or by drilling additional holes in a stationary spray ball. Such a modified spray system should be retested for adequate coverage. Such riboflavin testing is usually part of the operational qualification (OQ) of the equipment.

Cleaning Agents

Aqueous vs. nonaqueous

In addition to the cleaning method used, the cleaning agents used in the washing step are critical. It should be appreciated that selection of the cleaning method and cleaning agent(s) are somewhat interdependent. Selection of a cleaning method may limit the available cleaning agents that can effectively be used in that process. For example, a CIP process requires a low foaming aqueous cleaning agent, while extent of foam may not be critical for manual cleaning. Cleaning agents may be divided into aqueous and nonaqueous cleaning products. Nonaqueous products are typically solvents, and are more common in cleaning in the bulk manufacture of an active pharmaceutical ingredient (API). Typically, the solvent used for cleaning is the same as that used for manufacture. The cleaning effectiveness depends on the solubility of the residue(s) in the solvent at the temperature of cleaning. Particularly for cleaning of distillation columns, refluxing with a volatile solvent is a common practice for effective cleaning. The trend in the manufacture of APIs is to move away from solvent cleaning to aqueous cleaning. However, it should be recognized that in many cases this is not practical, and even if it is, the aqueous cleaning may be followed by one or more solvent flushes to remove the water from the process vessels.

Types of aqueous cleaning agents

Aqueous processes involve cleaning with water and, optionally, other ingredients to assist in the cleaning process. If aqueous cleaning can be suitably performed, it is preferred over solvent cleaning because of cost issues (including the cost of the solvent as well as the costs of disposal or reclamation of the solvent) and because of environmental issues relating to the use or emissions of solvents. In aqueous processes, the use of water alone should be considered because it eliminates the concerns over having to consider potential contaminants from the cleaning agent during cleaning validation. However, in most cases, the performance characteristics of various aqueous cleaning agents more than overcome the concerns about cleaning agent residues (particularly if the cleaning agents selected are free-rinsing). The successful use of water alone for the washing step depends solely on the

solubility of the residues in water at the temperature of cleaning, and may not typically provide other cleaning mechanisms such as emulsification and dispersion. Therefore, use of water alone may not meet other cleaning objectives such as short processing times.

Another option for aqueous cleaning involves the use of commodity chemicals, including alkalis such as sodium or potassium hydroxide, acids such as phosphoric or citric acid, or sodium hypochlorite solution. These are typically diluted in water at levels of 0.05–1% (w/w), and the resultant solution is typically used at elevated temperatures (45–80°C). Commodity chemicals may provide better cleaning than water alone, and they do so at a relatively inexpensive cost. Residue detection of cleaning agents during validation is relatively straightforward because there is usually only one chemical species to detect from the cleaning agent itself.

A third option for aqueous cleaning is to use a formulated cleaning agent. These formulated products usually contain several functional agents including a surfactant(s), an alkalinity or acidity source, water miscible solvents such as glycol ethers, dispersants such as low-molecular-weight polymers, and various builders such as chelants. The main advantage of such formulated cleaning products is that they are multifunctional because of the variety of components; each component broadens the performance in terms of being applicable on a wider variety of soil types. Well-formulated products thus enable a pharmaceutical manufacturer to use one cleaning agent in one cleaning SOP to effectively clean not only the variety of components in a finished drug product, but also a broader range of finished drugs themselves. It should be noted in the former case that for many (if not most) finished drugs, it is the excipients in the finished drug that are more difficult to clean (as compared to the cleaning of the active ingredient). However, the selection of a formulated cleaning product necessitates that the pharmaceutical manufacturer knows the ingredients in the product, both as a check on the consistency of the formulation over time and to effectively establish residue limits for the cleaning agent.

Basis of selection of cleaning agent

The selection of an aqueous cleaning system is simplified if only water alone, or water and a commodity chemical alone, are used. The cleaning performance can be somewhat predicted based on solubility characteristics (at the appropriate pH) or by consideration of the peptizing performance of alkalinity on protein or the oxidizing action of sodium hypochlorite on denatured protein. In the case of formulated multifunctional cleaning agents, the performance is more difficult to predict based on

chemistry alone, and an acceptable cleaning agent is preferably selected based on experience or on laboratory studies. The selection of cleaning agents is also complicated by the fact that sometimes proper cleaning necessitates the use of two cleaning agents at the same time (a primary cleaning agent and a functional additive of some sort), or by the use of two cleaning agents in succession (for example, the use of an alkaline cleaning product followed by an acidic cleaning product).

Cleaning Parameters

While selection of the cleaning method and cleaning agent(s) is important, equally important are the various parameters to consider in the overall cleaning system. These include cleaning process parameters as well as parameters related to the system actually cleaned. Probably the most important cleaning process parameters are the time of cleaning, the temperature of cleaning, the concentration of the cleaning agent, the water quality, the impingement action of the cleaning solution, and any mixing in the cleaning solution.

Time

Three aspects of time are important to the cleaning process. The first is the time from the end of product manufacture to the beginning of the cleaning process. This is important in validated cleaning because the nature of the soil to be cleaned may change over time. Changes may include the drying of the soil residue (thus possibly making it more difficult to clean) or microbial proliferation (thus increasing the bioburden to be cleaned during the cleaning process). A maximum time between the end of manufacture and the beginning of the cleaning process must be specified, and this maximum time must be considered in the selection of worst case conditions for the validation of cleaning processes. A second aspect of time is the times of the cleaning process steps, as well as the time between steps. This includes specifying the time of the prewash, of the washing step, and of rinsing. These are usually established based on laboratory and scale-up trials in the development of a cleaning SOP. The times between these three steps may be critical; if so a maximum time interval should be specified. For example, in the manual cleaning of larger equipment, the time interval between the washing step and the rinse could be significant if the cleaning agent on the washed part is allowed to dry before the rinsing step starts. Drying after the washing step may redeposit soils and prevent effective rinsing. The expectation for validated cleaning is that the times for the various phases are specified. It is generally unacceptable to specify an open-end time frame such as

“test until clean” (that is, continue repeating the cleaning process until tests indicate the equipment is clean) in a validated process. Such performance is indicative of an uncontrolled cleaning process. The third aspect of time is the time of storage of cleaned equipment. Although recontamination of equipment is generally known to be event related rather than time related, time is known to affect microbial proliferation. For this reason it is expected for validated cleaning processes that an expiration date (or “use by” date) for cleaned equipment be established.

Temperature

A second important process parameter is the temperature, not only of the cleaning solution, but also of the prewash and rinse solution. The solution temperature can significantly affect cleaning performance, including the rate of solubility and the extent of hydrolysis. Control of temperature during cleaning is preferable. However, it should be recognized that consistency is more important than just constancy of temperature. A consistent decrease in temperature (due to the lack of a heat exchanger in a cleaning circuit) may be acceptable for validated cleaning, provided that the decrease in temperature is consistent from one cleaning event to the next. Temperature of the prewash is generally ambient to prevent setting of certain residues at higher temperatures. The temperature of the rinse is probably least critical. However, it should be recognized that the higher temperature of a rinse might facilitate faster rinsing. In addition, if the temperature of a first rinse is significantly lower than the temperature of the cleaning solution, the temperature “shock” may cause a cleaning solution containing emulsified soils to “break,” thus redepositing soils on the equipment surfaces. Temperature should be controlled within reasonable limits, for example, within 5°C of the control point.

Cleaning agent concentration

Cleaning agent concentration should be specified and controlled. Cleaning agent concentration can usually be controlled by diluting based on weight or volume, or by diluting to a known control point, such as to a known conductivity. Within reasonable limits for aqueous cleaning, higher cleaning agent concentrations result in more effective cleaning. Concerns with higher cleaning agent concentration include deleterious effects on equipment and safety issues in manual cleaning.

Water quality

In certain circumstances, water quality can be critical for cleaning performing. If the washing step involves the use of surfactants for cleaning, the presence of hard-water ions (calcium and magnesium) is well known to interfere with

effective detergency. Additionally, in the presence of alkalinity sources (which raise the pH), calcium ions will precipitate as calcium carbonate. Such deposits, if not removed from the equipment surfaces, can contribute to the equipment being judged visually dirty. Some formulated cleaners will contain chelants (such as salts of ethylenediamine tetra-acetic acid) to minimize such possibilities. For most cleaning applications, pharmaceutical manufacturers will also use the same quality of water (Purified Water or Water for Injection) that is used for manufacture of the drug product. Lesser quality water, such as tap water, can be used provided the water quality (both chemical and microbiological) is carefully monitored. In addition, if the tap water quality may vary (due to seasonality or source, for example), the worst-case water conditions must be considered for validation purposes.

Impingement

The impingement of the cleaning solution refers to the physical action of a cleaning solution as it hits the surface from a spray application. Such a spray application may include that from a spray device in a CIP system, or may be from a high-pressure hose spray application. Impingement provides mechanical action to help dislodge residues from surfaces. Such impingement can be beneficial if the dislodged residues can then be suspended, emulsified, or otherwise carried away from the equipment surfaces and removed from the cleaning system. Dislodging residues and just displacing them to another location on equipment surfaces may not prove beneficial. In some circumstances, impingement with a solution containing added cleaning agents may be preferred.

Mixing

Mixing refers to the movement within the cleaning solution itself. With a static application of a cleaning solution, as the soils on the surfaces dissolve, emulsify, or otherwise migrate into the cleaning solution, a concentration gradient of saturated or partially saturated cleaning solution is established near the equipment surfaces. This concentration gradient minimizes the chemical cleaning action. Mixing eliminates this concentration gradient and places fresh cleaning solution (or at least a less saturated cleaning solution) in contact with soils on surfaces. This optimizes the cleaning process. It is desirable that mixing be such that the cleaning solution experiences turbulent flow.

The six parameters discussed above are parameters that usually can be controlled by proper design of the cleaning process. Other parameters that are important for cleaning are things that are controlled more by equipment design or by manufacturing process design. Those characteristics

include the nature of the equipment surfaces, the physical nature of the soil, and the amount of soil.

Nature of the surface

In removing manufactured product soils from equipment, the nature of the surface may also affect the cleaning process. This includes any special factors in the adhesion of the soil to different surfaces. Different surfaces include differences in type, such as stainless steel, glass, and various plastics. Effective removal of soils from all representative surface types is usually considered in a sampling plan for cleaning validation. Different surfaces also include the roughness or smoothness of the surface itself. Although there is controversy on this, as a general rule for most surfaces involved in pharmaceutical manufacturing, smoother surfaces are more easily cleaned. This may be related to the fact that rougher surfaces have cracks or crevices where soils can more easily “hide.” For example, etched glass surfaces are generally more difficult to clean as compared with highly polished glass surfaces. A third factor in considering the nature of the surface is the chemical or physical compatibility of the cleaning solution with the surface itself. The objective in cleaning is to remove the soils and restore the surface to its original condition (or as close to that condition as practical). Two examples of substrate compatibility issues are the repeated use of high levels of hypochlorite on stainless steel (leading to rouge formation) and the use of high levels of aqueous alkalis (sodium or potassium hydroxide) at high temperatures for prolonged periods on glass-lined vessels (leading to etching of the glass surfaces). Other issues might be temperature compatibility of plastics or of gasket materials. Although some deleterious effects may be expected in any cleaning process, the process should be designed to clean effectively and yet keep substrate compatibility issues to a minimum.

Condition of soil

A second factor to consider is the soil condition itself. Three “states” of the soil may be considered—freshly deposited soil, dried soil, and baked-on soils. The difference between the last two is that drying just involves the removal of water without any chemical changes in the soil. Baking usually involves not only the removal of water but also a significant chemical change in the soil. Such chemical changes usually result in the soil being more difficult to remove. For example, drying sugar on a surface may render it slightly more difficult to remove; however, baking it at elevated temperatures will caramelize the sugar and render it extremely difficult to remove. The condition of the soil may change because of manufacturing process conditions, such as product splashing onto a vessel

dome that is steam jacketed and baking onto the surfaces. It also may change because of a time delay after manufacture and before cleaning, allowing the product to dry out. It should be noted that merely drying of certain polymers on surfaces might render them extremely difficult to remove. For example, dried solutions of carboxymethylcellulose (CMC) can be extremely difficult to remove from surfaces.

Although it may be difficult to control the extent of drying or baking in certain processes, these phenomena should be evaluated, and if they do occur, the cleaning process should be designed to remove those soils in the more difficult dried or baked conditions.

Amount of soil

A third factor to consider is the amount of soil on the surface. As a general rule, the greater the amount of soil on the surface, the more difficult the cleaning. For freshly deposited soils, this may not be a serious issue if the bulk of the soil can be readily removed in the prewash. On the other hand, with dried or baked-on soils, the prewash may have little benefit in reducing the amount of soil on the surface. Unfortunately, surfaces that are most likely to have larger amounts of soils (dead legs, cracks, crevices, low flow areas) are also those that are more difficult to clean because of accessibility of the cleaning solution to the surfaces. For cleaning process design purposes, worst cases in soil amounts should be considered.

CLEANING STRATEGIES IN LIGHT OF VALIDATION

Although cleaning processes should be primarily based on what is necessary for good cleaning, they may be modified somewhat based on the regulatory needs for validation. As most pharmaceutical companies will want to validate a cleaning process and not have to do additional significant revalidation work in the near future, this may limit the selection of cleaning agents. As a key part of any validated process is consistency and control, cleaning SOPs for validated processes will also generally have more detail and specificity.

Cleaning for Multiproduct Equipment

Several strategies are possible for cleaning of equipment used to make two or more different products. One option is to optimize a cleaning process for each product made on the equipment. This may mean different cleaning agents for cleaning after each manufactured product, although usually what it means is that the same cleaning agent is

used under different process conditions (such as time and/or cleaning agent concentration). Each manufactured product will have its specific cleaning SOP. Another option is to use only one cleaning process for all products manufactured on that individual piece of equipment. One cleaning SOP (with all process conditions the same) is used for all manufactured products. Such a strategy allows for the possibility of “grouping” or “bracketing” for validation protocol purposes. However, it should be recognized that the decision to use one cleaning SOP for all manufactured products has implications for both cleaning and for validation purposes. A strategy of “one SOP for everything” has advantages for cleaning in terms of simplifying documentation and simplifying training. However, such a strategy can be pursued regardless of whether one adopts grouping strategies for validation or not. Clearly, one can also adopt a hybrid strategy, in which several manufactured products are cleaned with one SOP and another group of products (manufactured on the same equipment) is cleaned with a different SOP.

Cleaning in Campaigns

Cleaning between lots of the same product made successively on the same equipment in a campaign may allow for less aggressive cleaning procedures. The reason is that in such cleaning there is no concern about cross-contamination with an active from a different product. However, there are concerns about cleaning. First is the issue of lot integrity—how much of the active or product from one lot can come along with a different lot and be considered different lots for such purposes as recalls? In addition, while cross-contamination is not an issue, other issues such as contamination from residues of cleaning agents and microbial contamination should also be considered. Another issue in campaigns in which cleaning is minimal is the possibility of degradation product accumulating on the equipment.

VALIDATION ISSUES

IQ/OQ/PQ

Cleaning validation is a type of process validation, and the principles of process validation (17) apply equally to a cleaning process. This includes installation qualification (IQ), operational qualification (OQ), and process or performance qualification (PQ). IQ and OQ should focus on the equipment used for the cleaning process, such as a CIP skid, a spray device, or the monitoring equipment (such as a conductivity probe).

PQ involves performance of the cleaning procedure three consecutive times and evaluating the success of the cleaning procedure, usually by measuring the amount or degree of potential contaminants on the cleaned equipment surfaces. The cleaning SOP should be challenged during the three PQ runs, using (as much as possible) process conditions within the normal ranges that are more likely to induce failure. For example, if the time from the end of manufacture until the beginning of cleaning is specified as a maximum of 12 h, then at least one of the PQ runs should be performed at that maximum time to demonstrate adequate performance.

Cleaning validation is different from other types of process validation in that with cleaning validation both the product cleaned as well as the next product manufactured must be considered. The cleaning SOP is primarily based on what is required to remove the manufactured product. However, the types and acceptable levels of residues following cleaning are also determined by the nature of the next product manufactured in the cleaned equipment. For this reason, cleaning validation is more dependent on what other products are made on the same equipment. Furthermore, the addition of a new product to equipment previously validated for cleaning with multiple manufactured products requires a reevaluation of that previous validation work to determine whether or not the previously validated residue acceptance limits are still applicable in light of a new “next product.”

Residue Limits

Validating a cleaning process includes selecting target residues and setting limits for those residues following the cleaning process. Target residues are selected based on possible residues that can be left after the cleaning process. This requires an understanding of the cleaning process, and may require an investigation into possible degradation products that may occur during the cleaning process. Acceptable levels of those specific residues are based on what could occur should those residues contaminate the subsequently manufactured product (18, 19). Analytical determinations of residues are usually required. In addition to those measurements, it is expected that the equipment will be visually clean. Examination of equipment for visual cleanliness requires training of the observers and may require auxiliary lighting. A visual examination may be supplemented by use of a video camera for recording purposes or by use of a borescope for pipes. In some cases, equipment may be disassembled for visual examination (and optionally for analytical sampling) to determine cleanliness.

Limit in next product

It is important in any discussion of “residue limits” to understand that limits for a cleaning process may be expressed in different ways. This includes the limit of the residue in the subsequently manufactured product, the limit of the residue on the cleaned equipment surfaces, and the limit of the residue in the analyzed sample. These are all related, but they are usually different numbers. For an active ingredient in the cleaning of a finished drug product, the limit in the next product is usually calculated based on application of a safety factor (usually 0.001 or lower) to the minimum daily dose of that active in the maximum daily dose of the subsequently manufactured product. The active or level of active in the subsequently manufactured product is irrelevant unless there is information about unusual deleterious interactions. This calculation is also independent of manufacturing issues such as batch size and equipment surfaces areas, and can be calculated solely on information about the dosing of the two products as follows:

$$L_1 = \frac{\text{MinDA} \times \text{SF}}{\text{MaxDSP}} \quad (1)$$

where L_1 is the limit of the active in the next product, MinDA is the minimum (daily) dose of the active (the target residue), MaxDSP is the maximum (daily) dose of the subsequently manufactured drug product, and SF represents an appropriate safety factor. Care needs to be paid to selection of units; the L_1 limit is usually expressed in $\mu\text{g/g}$ (or ppm).

Limit per surface area

The next limit calculated is usually the limit per equipment surface area. This is calculated based on the limit in the next product, the batch size of the subsequently manufactured product, and the equipment shared surface area. This is expressed as:

$$L_2 = \frac{L_1 \times \text{BS}}{\text{SSA}} \quad (2)$$

where L_2 is the limit per surface area, BS is the batch size, and SSA is the shared surface area. Units should be consistent, and the L_2 limit is usually expressed in units of $\mu\text{g}/\text{cm}^2$.

Limit in analytical sample

The next limit is the limit in the analytical sample. If the sampling method involves swabbing, the surface area swabbed and the amount of diluent used for desorbing the swab must be considered. The limit per swab sample is then calculated as:

$$L_3 = \frac{L_2 \times \text{SA}}{\text{AD}} \quad (3)$$

where L_3 is the limit per analytical sample, SA is the swabbed area, and AD is the amount of diluent for swab elution. Here again units need to be consistent, and the L_3 limit is usually expressed as $\mu\text{g/g}$ or $\mu\text{g/mL}$. It should be clear that the limit in the analytical sample can be manipulated by changing the area sampled (higher areas result in larger limits per analytical sample) or the amount of diluent used (lower amounts result in larger analytical sample limits). If a sampling rinse is used (in place of swabbing), SA effectively becomes the total surface area of the equipment, and AD becomes the volume of solution used for the sampling rinse.

Nondose limits

For residues (such as cleaning agents) that do not have a defined dose, some measure of toxicity, such as an acceptable daily intake (ADI), is used for residue limit purposes. If the subsequently manufactured product is an in vitro diagnostic (IVD), and has no defined dose, then some evaluation of the effects of target residues on the performance or stability of the IVD product should be performed. These nondose factors are used only for the L_1 limit; there are no changes for calculation of L_2 and L_3 limits.

Limits for multiple subsequent products

When a residue limit is to be calculated for a product where there may be more than one subsequently manufactured product, calculations should be made to compare the surface area residue limits (L_2 limits) by using each subsequent product. If the manufacturing order is not to be restricted, the cleaning validation of the first product should be established using the lowest surface area limit.

Sampling Procedures

Sampling procedures for cleaned surfaces can be divided into four types. Direct surface sampling involves a fiberoptic probe (such as a near infrared probe) that is placed directly on the surface. An output is provided as to the type of residue and the level. Such systems are currently in development (20), but are not commercially practical. Swab sampling involves wiping a fixed area of the surface with a premoistened swab. The swabbing procedure is designed to remove any residues from the surface, and the swab is then placed in diluent to desorb the residue from the swab to the diluent. The residue is then measured in the diluent by a suitable analytical technique.

Such swabbing is commonly called “direct surface sampling,” although it clearly is an indirect measure. Rinse sampling involves flushing the equipment surface with a fixed amount of rinse solution (aqueous or solvent), capturing the rinse solution, and then measuring the target residue in the rinse solution. A true sampling rinse is distinct from the final process rinse, and may involve a solution different from that for the process rinse. A fourth sampling procedure is placebo sampling. This involves making a placebo of the subsequently manufactured product in the cleaned equipment. Following manufacture of the placebo, the placebo is sampled and analyzed for the target residue. Any target residue in the placebo would come from the cleaned equipment, and it could be expected that the level present in the placebo would be the level present in any such subsequently manufactured product. Placebo sampling is not widely used because of regulatory concerns related to uniformity of contamination of the placebo from the equipment surfaces and the analytical challenge of finding low levels of residues in placebos.

Analytical Methods

Relationship to target residue

The analytical method selected to measure the target residue must provide a direct measurement of that target residue. When regulatory authorities first began requesting that cleaning be validated, some companies merely tested the rinse water by USP Purified Water specifications to determine if the equipment was clean. The rationale was that the effluent met the same standard as the incoming water. Regulatory authorities (quite rightly) rejected such arguments (because of the possibility of unacceptable levels of potent drugs being present, and because of the possibility that the target residue not being removed in the rinsing procedure), and requested that analytical techniques target the specific residues of concern. However the requirements for analytical methods for residue determination are slightly different from methods for actives level determination in finished product in one important way. For finished product actives determination, a method is required to unequivocally measure the active in the presence of known potential interferences and provide an exact level of the active present. For cleaning validation residue analysis, it is not so as important to know exactly how much residue is present as to know that the amount present is below the acceptance criteria in the validation protocol. For this reason both specific and nonspecific analytical methods can be used for residue detection purposes.

Specificity of methods

Specific methods are preferred because they can more accurately provide information for evaluating potential problems. Because they are designed to eliminate the effects of potential interferences, they can more reliably meet the acceptance criteria. The most common method for residue determination for cleaning validation purposes is the high performance liquid chromatography (HPLC) procedure. In contrast, nonspecific methods such as total organic carbon (TOC) can only provide an upper limit value of the target residue, provided there are no negative interferences (that is, all interferences contribute positively to the analytical response). For TOC, this is usually the case. If the analytical response is treated as if the response comes only from the target residue, then an upper limit calculation of the target residue can be obtained. If such upper limit calculation is below the acceptance criterion, then it is safe to claim that the residue is within acceptance limits. Nonspecific methods such as TOC are more commonly used in biotech manufacture, where proteinaceous actives are readily degraded by the cleaning procedure. In such cases, the TOC values are treated as if the carbon were due solely to the protein active. Actually, some of the carbon may be due to the cleaning agent, and some may be due to the excipients or processing aids.

Validation of methods

It is expected that any analytical method chosen be validated, including an evaluation of specificity, sensitivity (limit of detection and limit of quantitation), accuracy, precision, range, and linearity (21). The range validated is preferably a range around the expected value in the analytical sample. However, it is wise to also include values up to the acceptance limit in the analytical sample.

Sampling/Analytical Method Recovery

The sampling method chosen must be challenged in combination with the analytical procedure to determine the recovery of the sampling method. This is typically a laboratory study involving spiking a model surface with the target residue and performing the sampling procedure on the surface and measuring the residue with the analytical procedure (22). The amount of residue measured is compared to the amount spiked to give a percent recovery. Recoveries of greater than 80% are considered good, but recoveries of greater than 50% are acceptable. As the analytical values have to be transformed by the recovery values, it is desirable to obtain as high a recovery as consistently possible.

MICROBIAL CONTROL ISSUES

Issues in Cleaning

GMPs require that procedures be in place to limit objectionable microorganisms in both nonsterile and sterile drug products. This should be interpreted to include both the number of organisms as well as the type (species) of organism. Protection of subsequently manufactured product from microbial contamination can be accomplished in part by effective cleaning, by a separate sanitizing step, and/or by storage procedures. In many cases effective microbial control is achieved by a good aqueous cleaning process. The conditions of cleaning can either physically remove microbes, or these conditions (hot alkaline or acidic aqueous conditions) can be conducive to the destruction of microbes. The use of hypochlorite for removal of denatured proteins also serves as an effective oxidizing biocide. If cleaning alone does not achieve adequate microbial reduction, the use of either a chemical sanitizer or elevated temperature (steam or hot air) can be considered. Chemical sanitizers include hydrogen peroxide, peracetic acid, quaternary ammonium chlorides, and alcohols; as a general rule phenolic sanitizers are not used for process equipment because of the difficulty of rinsing from equipment surfaces. If the sanitizer leaves a residue, then a final rinse should be considered to reduce that residue to an acceptable level. Acceptable levels of microbes in the cleaned equipment can be established depending on the nature of the subsequently manufactured product. Limits for nonsterile products can be established based on accepted levels of microorganisms in nonsterile products. Such calculations usually result in acceptance levels that are considerably above what can be routinely achieved with good cleaning procedures. Limits for equipment surfaces used for manufacture of terminally sterilized products are usually related to assumptions of the maximum bioburden for product sterilization purposes. Limits for equipment surfaces used for manufacture of aseptically produced products are usually related to assumptions of the maximum bioburden for equipment sterilization purposes.

Issues in Storage

One major regulatory issue in the cleaning of equipment is the possible microbial proliferation due to improper storage, such as in a wet condition or with pools of water. The preferred method for dealing with such concerns is to effectively dry the equipment before prolonged storage. An alternative (but less desirable option) is to include an additional cleaning and/or sanitizing step after storage and

before the next use of the equipment. If this alternative is used, the measurement of both chemical and microbial residues should be performed at the end of this cleaning/sanitizing step.

VALIDATION MAINTENANCE

Once a cleaning process has been appropriately validated, steps should be taken to help insure that the cleaning process remains consistent and in control. Steps that are taken to help assure this include regular monitoring, a change control system, training, and revalidation.

Monitoring

The testing that is done for routine monitoring of the cleaning process should be distinguished from the testing that is part of PQ process qualification. Monitoring tests are usually done on each individual cleaning run. Tests are selected which could be indicative of a cleaning system that is either out of control, or could be trending out of control. Examples of process parameters that could be monitored include the concentration of the diluted cleaning agent, temperature of the cleaning solution, times of various process steps, pressure at a spray device, flow rates, volumes of solutions used, and conductivity of the final rinse water. As these monitoring steps should be part of the cleaning SOP, they should also be performed during the three PQ runs. Some monitoring can give pass/fail information, which clearly indicates the cleaning process is out of control, requiring an investigation and correction of the problem. For example, a higher than specified pressure at a spray device may suggest that spray nozzles were blocked, and that perhaps cleaning coverage was inadequate. This would require an immediate investigation of whether the equipment was adequately cleaned. Such equipment should not be used until a confirmation of adequate cleaning is performed, and the cause of the high pressure corrected. On the other, hand monitoring of the final rinse water conductivity or TOC may show results which do not necessarily suggest that cleaning was inadequate, but rather may display a trend which suggests that cleaning may become inadequate if such a trend continues. This is the value of action and alert limits for monitoring, and of control charts which show trends. It is possible that in certain situations the full range of testing done in the PQ runs would be repeated. However, the value of validation is that consistency has been demonstrated, and the emphasis in monitoring should be tests that might indicate a process change.

Change Control

Validated cleaning processes should be subject to change control. Changes include unplanned and planned changes. Examples of unplanned changes include the failure of a process pump, the clogging of a spray device, and the discontinuance of a cleaning agent by a supplier. The keys to change control are to evaluate the effects of any change, correct the changed item (if possible), implement any increased monitoring as needed, and document the procedure. For example, failure of a pump in a CIP skid may just require a switch of like for like, IQ on the new pump, optionally OQ on the new pump, and proper documentation. Clogging of a spray device may require cleaning of the device and an investigation of the cause of clogging. This should be followed by preventive measures, such as installation of a filter screen to remove material with the potential to clog and some kind of preventive maintenance to regularly clean the screen and inspect the spray device. In both these cases, it may not be necessary to repeat a PQ run. On the other hand, a slight compositional change of the cleaning agent may require laboratory studies to suggest equivalence, followed by one PQ run to confirm equivalence. A change in the manufacturing process itself, and the possible effect of such a change on the cleanability of the equipment, should not be neglected when considering change control for a cleaning process. A change such as increased processing temperature may modify the condition of the soil and, therefore, make it more difficult to clean. In all cases documentation of changes according to a change control SOP is mandatory.

Training

Training operators in the cleaning SOP is an important part of validation maintenance, particularly for manual cleaning methods. Training in a manual method should include a classroom discussion of the method, observation of the SOP being performed by a trained operator, and then demonstration of proficiency by performance of the SOP by the trainee. Training should always follow revision of the SOP, and retraining should follow any deviations that were attributable to operator error.

Revalidation

There are two aspects to revalidation of a previously validated cleaning process. First is revalidation upon any significant change. What is "significant" is a matter of professional judgment. However, a change in cleaning method, such as from manual cleaning to automated

cleaning will generally require revalidation even though the cleaning agent and process parameters are the same. In essence this is not really revalidation but rather validation of a new cleaning process. The other aspect of revalidation is based on the evaluation of the consistency and control of a cleaning process on a regular basis to confirm that the process is still under control. The time of this periodic revalidation should be specified in a cleaning validation policy (such as in the cleaning validation master plan), and typically is every one or two years. Such a periodic revalidation involves an evaluation of the monitoring data, change control, cleaning process deviations, and quality records of products manufactured after the cleaning process. If the monitoring data is adequate, the change control is minor, any deviations have had attributable to corrected causes, and there have been no product quality problems possibly related to cleaning, then all this information is suggestive of a cleaning process that is still under control. In such a case it is possible to document such findings in a revalidation report with a conclusion that the cleaning process is still under control and that the original validation work is still applicable. On the other hand, if the monitoring data show continual trends which require corrections, if numerous individual changes have been made (each of which was acceptable) but the overall cleaning process is now seen as significantly different, if deviations in the cleaning process have either not been attributable to a cause or the cause has not been corrected, and/or if there have been product quality issues related to the cleaning process, then such an investigation may result in a repeat of one or more PQ runs. In such a case, usually there will be some laboratory or pilot scale evaluations before PQ runs are performed.

REFERENCES

1. Current Good Manufacturing Practice in Manufacturing, Processing, Packing, or Holding of Drugs: General, 21 CFR210, 1 April 1997 (revised).
2. Current Good Manufacturing Practice for Finished Pharmaceuticals, 21 CFR 211, 1 April, 1997 (revised).
3. Harder, S.W. The Validation of Cleaning Procedures. *Pharm. Technol.* **1984**, 8 (5), 29–34.
4. Mendenhall, D.W. Cleaning Validation. *Drug Dev. Ind. Pharm.* **1989**, 15 (13), 2105–2114.
5. FDA. *Guide to Inspections of Validation of Cleaning Processes*; FDA Office of Regulatory Affairs: Rockville MD, 1993.
6. United States vs. Barr Laboratories, 812F, Supp. 458 (DNJ 1993).
7. FDA. *Mid-Atlantic Region Inspection Guide: Cleaning Validation*, July 28; 1992.

8. FDA. *Guide to Inspections of Validation of Cleaning Processes*; FDA Office of Regulatory Affairs: Rockville, MD, 1993.
9. FDA. *Current Good Manufacturing Practice: Proposed Amendment of Certain Requirements for Finished Pharmaceuticals*; Federal Register, 1996; 61, 20103.
10. Recommendations on Cleaning Validation. *Document PR 1/99-2. Pharmaceutical Inspection Cooperation Scheme*; Geneva, Switzerland, 1 April, 2000.
11. Draft 4 of Annex 15 to 1997 EU Guide to Good Manufacturing Practice—Eudralex Volume 4, *Validation Master Plan/Design Qualification/Installation and Operational Qualification/Non-Sterile Process Validation/Cleaning Validation*, European Commission, Working Party on Control of Medicines and Inspections, 17 September 1999. <http://dg3.eudra.org/pharmacos/docs/doc99/GMPAnn15.pdf> (accessed September 2000).
12. Cleaning Validation Guidelines, Health Canada Therapeutic Products Programme, 1 May, 2000. http://www.hc-sc.gc.ca/hpb-gps/therapeut/zfiles/english/guides/validate/validation_guide_e.pdf (accessed September 2000).
13. ICH Steering Committee, Draft Consensus Guideline: Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients, 19 July, 2000. www.ifpma.org/pdfifpma/Q7Astep2.pdf (accessed September 2000).
14. PhRMA Quality Committee, Bulk Pharmaceuticals Working Group. PhRMA Guideline for the Validation of Cleaning Procedures for Bulk Pharmaceutical Chemicals. *Pharm. Technol.* **1997**, 21 (9), 56–73.
15. Good Manufacturing Practice Guide for Bulk Pharmaceutical Excipients. *International Pharmaceutical Excipient Council*; Arlington VA, 1995.
16. FDA. *Guide to Inspections of Validation of Cleaning Processes*; FDA Office of Regulatory Affairs: Rockville, MD, 1993.
17. FDA, Guideline on General Principles of Process Validation, May 1987. <http://www.fda.gov/cder/guidance/pv.htm> (accessed September 2000).
18. Fourman, G.L.; Mullen, M.V. Determining Cleaning Validation Acceptance Limits for Pharmaceutical Manufacturing Operations. *Pharm. Technol.* **1993**, 17 (4), 54–60.
19. LeBlanc, D.A. Establishing Scientifically Justified Acceptance Criteria for Cleaning Validation of Finished Drug Products. *Pharm. Technol.* **1999**, 23 (10), 136–148.
20. FTIR.com, Cleaning Validation and Coat Weight Analysis., <http://www.ftir.com/remspec/grazing05html.htm> (accessed September 2000).
21. Kirsch, R.B. Validation of Analytical Methods Used in Pharmaceutical Cleaning Assessment and Validation. *Analytical Validation in the Pharmaceutical Industry* (Suppl. to *Pharm. Technol.*) **1998**, 22 (10), 40–46.
22. Kirsch, R.B. Validation of Analytical Methods Used in Pharmaceutical Cleaning Assessment and Validation. *Analytical Validation in the Pharmaceutical Industry* (suppl. to *Pharm. Technol.*) **1998**, 22 (10), 40–46.

FURTHER READING

- Bismuth, G.; Neumann, S. *Cleaning Validation: A Practical Approach*; Interpharm Press: Denver, CO, 1999.
- Brunkow, R.; Delucia, D.; Haft, S.; Hyde, J.; Lindsay, J.; McEntire, J.; Murphy, R.; Myers, J.; Nichols, K.; Terranova, B.; Voss, J.; White, E. *Cleaning and Cleaning Validation: A Biotechnology Perspective*; PDA: Bethesda, MD, 1996.
- LeBlanc, D.A. *Validated Cleaning Technologies for Pharmaceutical Manufacturing*; Intepharm Press: Denver CO, 2000.
- Points to Consider for Cleaning Validation*; PDA Technical Report No. 29; Bethesda, MD, August 1998.
- Verghese, G. Developing a Validatable Cleaning Process. *Proceedings of the 1999 Interphex Conference*; New York, April 20–22, 1999; Reed Exhibition Companies: Norwalk CT, 1999; 461–469.

ENZYME IMMUNOASSAY AND RELATED BIOANALYTICAL METHODS

John W.A. Findlay

Pharmacia, Skokie, Illinois, U.S.A.

Ira Das

St. Louis, Missouri, U.S.A.

INTRODUCTION

Enzyme immunoassay, a bioanalytical method incorporating an antigen–antibody reaction to capture the analyte of interest and an enzyme reporter system to detect the captured analyte, is one of the most widely used immunoassay formats. The method is sometimes applied only qualitatively to indicate the presence of an antigen in a matrix. However, in the more common quantitative implementation, a calibration (standard) curve is incorporated, from which the concentration of the analyte in unknown samples is interpolated. In the decades since the development of a radioimmunoassay for insulin by Yalow and Berson (1), immunoassays have been widely applied in support of medical practice and drug development. However, in recent years, there has been a decline in the application of immunoassays to the quantitation of low-molecular-weight xenobiotics, primarily due to the advent of liquid chromatography–mass spectrometry (LC–MS) methods, which have high sensitivity and specificity. This is particularly so for support of early drug discovery, where assay development times of as little as a day and analytical run times of only a few minutes per sample make LC–MS ideally suited to fast delivery of results to discovery scientists. Nonetheless, the remarkable specificity of antibodies allows their application in well-characterized immunoassays to the support of Phase III and Phase IV clinical trials as a cost-effective alternative to LC–MS methods. In addition, these methods are still widely used for therapeutic drug monitoring and analysis of low-molecular-weight hormones, such as steroids, in support of medical diagnostics. Immunoassays remain the method of choice for the quantitation of protein macromolecules and antibodies in complex matrices. Another major application of immunoassays is in the detection and quantitation of biomarkers, which are evolving to be of pivotal importance in the evaluation of pharmacological, toxicological, and clinical activities of candidate drugs (2).

Immunoassays generally vary in the type of critical antibody binding reagent or the detection and reporter systems used to monitor the end-point of the binding reaction. These different types of immunoassays have many characteristics in common; therefore, this chapter will include discussions of both enzyme immunoassays and other closely related methods. The enzyme immunoassay technique has been the subject of several textbooks, monographs, and review articles, including an excellent, comprehensive discussion in an earlier edition of this series (3). Thus, this chapter does not provide an in-depth review of the mechanistic details for producing and processing antibodies as reagents or on assay conditions for enzyme immunoassay. Rather, the intent is to present this technique in the context of several primary topics, namely the range of bioanalytical applications, the different, and sometimes additional, validation considerations imposed upon an enzyme immunoassay and its fraternal immunoassay methods, and some newer techniques that are complementary to enzyme immunoassay and offer potential performance enhancements. The chapter is written from the perspective of bioanalysis in biological fluids and does not address in any detail other applications of enzyme immunoassay, such as support of process control or product release, although such topics have been addressed elsewhere (4).

FORMAT OF ENZYME IMMUNOASSAYS

The format of an enzyme immunoassay refers to the configuration in which the components of the assay are assembled for routine application. Once this format has been established and assay conditions defined during assay development, they must remain unchanged through validation and subsequent application to sample analysis. Enzyme immunoassay formats fall broadly into two categories, namely heterogeneous and homogeneous. In a heterogeneous assay, at least one key reagent is

immobilized on a solid surface and there is at least one “washing” step before the final detection step. In contrast, in a homogeneous assay, all reagents are in solution together and there is no “washing” step prior to signal generation and detection. Both categories of assay include formats described as competitive and noncompetitive. In a competitive assay, there is direct competition between the labeled and the unlabeled antigen (analyte or ligand) in solution or, in some cases, between immobilized and soluble antigen for a limited number of antibody binding sites. In noncompetitive assays, antibody binding sites to capture and detect the antigen are not limiting because the antigen is incubated with excess capture antibody and enzyme-labeled detection antibody. An example of a competitive homogeneous assay format is the enzyme-multiplied immunoassay (EMIT) system (5), in which enzyme-labeled antigen competes directly in solution with unlabeled antigen in the biological sample (or calibration standard and quality control samples) for a limited number of antibody binding sites. The reaction endpoint is detected and quantitated spectrophotometrically without any intervening wash steps. This assay configuration is shown in Fig. 1. Enzyme-linked immunosorbent assay (ELISA) is an example of a heterogenous noncompetitive immunoassay. In this format, the primary antibody against the analyte of interest is immobilized on a solid plastic surface, usually a multiwell (or microtiter) plate. The biological sample is dispensed into the multiwell plate and incubated. The immobilized antibody then captures the analyte of interest, and the excess analyte is removed by washing. The antigen–antibody complex is then detected by two-step incubation with conjugated antibody and its substrate. First, an enzyme-labeled antibody, directed against the captured analyte, sandwiches the immobilized antibody–antigen complex. In a second incubation with an appropriate enzyme-specific substrate solution, a colored (or fluorescent or chemiluminescent) product is generated and quantitated spectrophotometrically. This assay format is depicted in Fig. 2. The ELISA can also be established in a competitive heterogenous format in which the antigen is immobilized on a multiwell plate and competition is established between the immobilized antigen and the antigen in solution for a limited number of binding sites on the primary antibody, also in solution. Following a fixed incubation period, the plates are washed and incubated with excess enzyme-labeled secondary antibody (directed against immunoglobulin from the species in which the primary, anti-antigen antibody was generated). The endpoint of this competition is then detected, following appropriate washing steps, by incubations with a signal-generating substrate. This assay format is depicted in Fig. 3.

It should be noted that the relationship between the final signal output and concentration of the analyte (dose–response) may be one of direct or inverse proportionality, and is dependent on the specific assay format. In addition, a number of different reporter enzymes may be used (e.g., horseradish peroxidase, alkaline phosphatase, β -galactosidase), along with a number of different signaling systems (e.g., substrates that yield chromogenic or fluorescent or chemiluminescent products, activation of signaling enzymes, amplification by biotin–avidin system or polymerase chain reaction).

ENZYME IMMUNOASSAYS FOR LOW-MOLECULAR-WEIGHT ANALYTES

Although the advent of sensitive LC–MS assays with short development times has reduced the need for immunoassays for low-molecular-weight compounds, the sensitivity, high-throughput and relatively low-cost characteristics of these assays still allow them to play an important role in some cases. Immunoassay support at the drug discovery stage may still be viable in such areas as the evaluation of biomarkers or determination of peptides, in which the elimination of sample cleanup prior to assay may constitute a valuable advantage of immunoassay over LC–MS. In addition, for low-molecular-weight therapeutic candidates, immunoassays can be used to support late-stage (Phases III and IV) clinical trials, when the metabolic pathways and the use of concomitant medications have been clearly defined. Critical to the application of immunoassay for analysis of low-molecular-weight compounds is the development of an antibody with clearly defined specificity for the analyte of interest. Low-molecular-weight compounds (less than 1000–2000 Da) are generally nonimmunogenic, or only weakly immunogenic, when administered directly to animals. To elicit an immune response producing antibodies suitable for use as reagents in immunoassays, these compounds must be conjugated to a carrier protein, e.g., bovine serum albumin (BSA), prior to immunization. This subject has been reviewed in depth previously (3). The immune system of the immunized animal responds to the carrier, and then secondarily to the hapten (the analog of the analyte bound to the carrier protein) attached to it. Although antibodies to the carrier protein are often produced in large amounts, they are easily removed by such techniques as affinity chromatography (6). Conversely, the desired antibodies to the analyte may be removed by affinity chromatography, which may also improve their specificity by eliminating or reducing

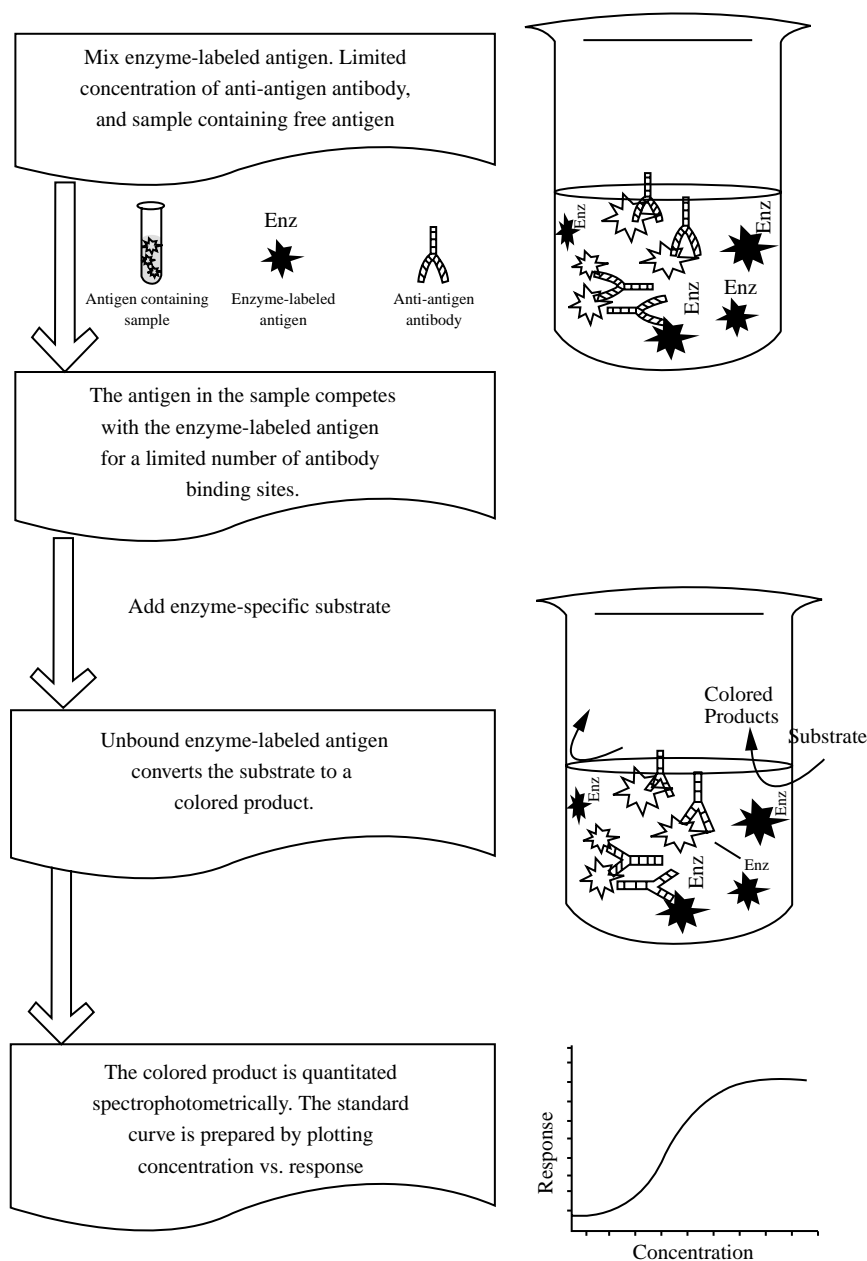


Fig. 1 An illustration of enzyme-multiplied immunoassay.

antibody populations that may cross-react with closely related chemicals, such as metabolites or degradation products. However, in many cases, carrier antibodies do not interfere in the assay to quantitate the analyte of interest and, therefore, anti-antigen antibodies in crude antiserum may be used in immunoassays without further purification. The site of attachment of the hapten to the carrier protein (either directly with the molecule via functional groups that are suitable for chemical coupling, or via a synthetic analog prepared to incorporate chemical

coupling functionality) will determine the specificity profile of the resulting antiserum. As a result, the site of attachment of hapten to carrier protein must be selected judiciously, considering all available knowledge of the metabolism of the compound in the animal system under investigation. Metabolic changes closest to the site of attachment will be poorly discriminated from the parent molecule, whereas metabolic changes distant from the conjugation site will be distinguished most clearly (i.e., will have lowest cross-reactivity). The ultimate

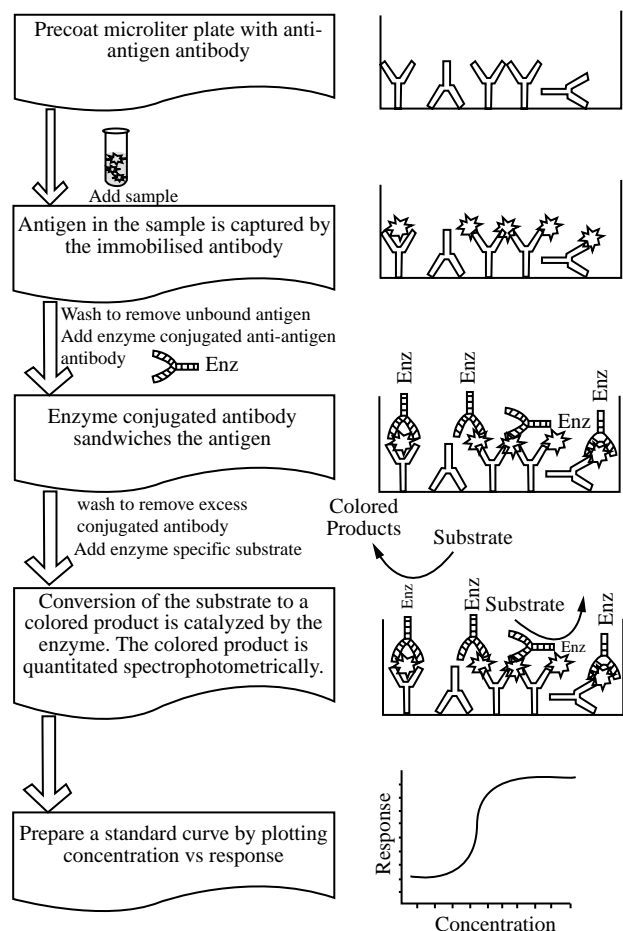


Fig. 2 An illustration of enzyme-linked immunosorbent assay.

result of this strategy is the ability of antibodies, and the resulting assays, to distinguish between enantiomers of chiral drugs (7). Although coupling methods generally have involved condensation reactions of hapten carboxylic or amino groups (or activated derivatives thereof) with amino or carboxylic groups on proteins, newer conjugation methods continue to be developed. These include the use of two-level, heterobifunctional agents such as *N*-(*m*-aminobenzoyloxy) succinimide, first to form a peptide bond to carrier protein and then, following diazotization of the aromatic amino group, to couple to a suitable hapten containing an imidazole, phenol, or indole residue. An example is the coupling of thyroid stimulating hormone to BSA (8). Another example involves interesting steroid chemistry in the preparation of an 11-Alpha-(3-Sulfanylpropyl)oxy hapten analog of the 3-Sulfamate ester of estradiol, and its subsequent coupling to bovine gamma globulin via a heterobifunctional crosslinker (9).

IMMUNOASSAYS FOR MACROMOLECULES

Immunoassays for macromolecules generally fall into one of two categories, namely those for endogenous proteins applied in support of clinical medicine (for example, assays for gonadotrophins or insulin) or those for new, genetically engineered proteins. The advent of genomics, proteomics, and recombinant technology has greatly advanced our understanding of the potential roles of regulatory proteins in the pathogenesis and modulation of diseases, and led to a sharp increase in the number of biological therapeutics under development. Several regulatory proteins are being developed as therapeutic agents and some have reached the market. Fig. 4 indicates the wide distribution of these agents across many different therapeutic categories, with a concentration in the field of cancer treatment. These products are often recombinant analogs of endogenous proteins, with the resulting challenge of developing enzyme immunoassays that are specific enough to distinguish between native and recombinant molecules.

Unlike low-molecular-weight xenobiotics, macromolecules are often immunogenic. An immune response is often elicited against several sites on the molecules called epitopes or antigenic determinants. The number of epitopes per antigen is determined by the size and complexity of the molecule. These epitopes can be linear, consisting of as few as four amino acids in a sequence, or conformational, involving different regions of the molecule in a three-dimensional configuration. Use of antibodies against conformational epitopes can be problematic if the antigen of interest loses its three-dimensional structure in vivo, or during sample analysis. In general, assay formats for macromolecules use at least two antibodies that react with two different regions of the molecule. Use of two antibodies against different regions of the molecule can confer additional specificity and may distinguish the analyte of interest from its metabolites or isoforms.

IMMUNOASSAYS FOR ANTIBODIES

Many of the new biological agents under development are recombinant versions of naturally occurring human proteins, or analogs of human proteins containing minor changes in their primary sequence, or differences in the extent of post-translation modifications. Marketed recombinant proteins include hormones (insulin, erythropoietin and growth hormone), enzymes (DNAase, asparaginase), cytokines (interleukins 1, 2, 11, interferon), growth factors

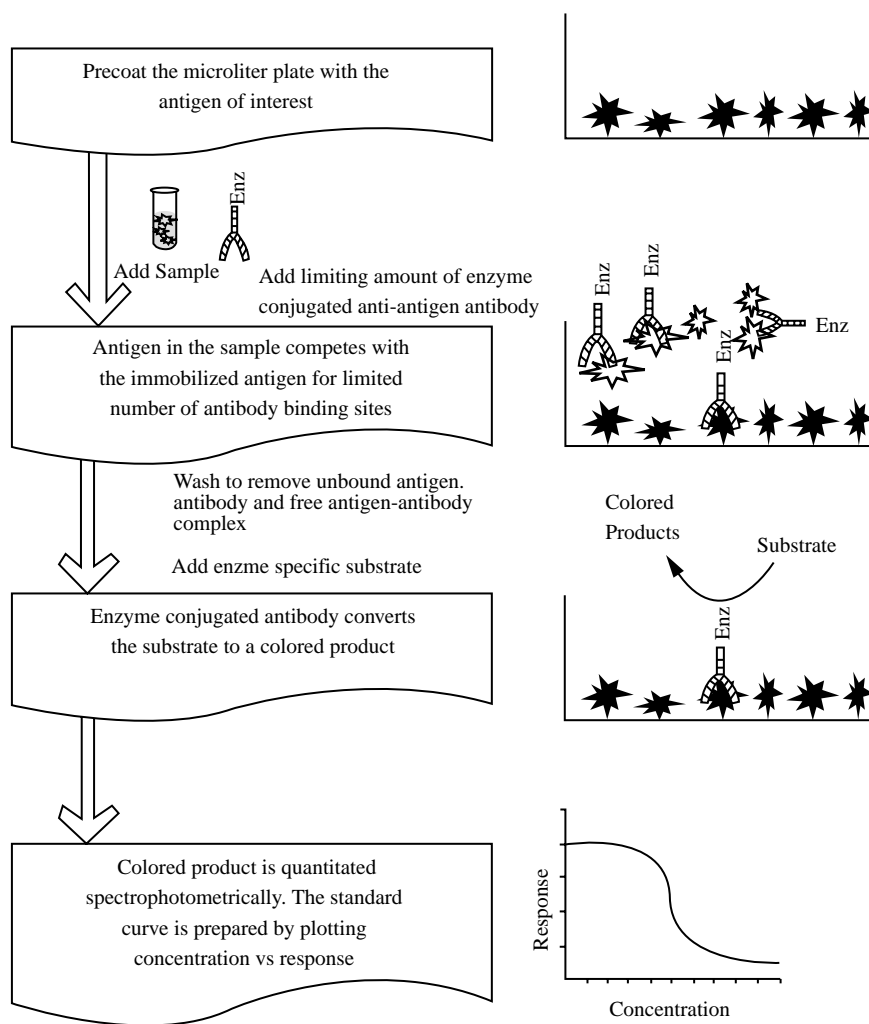


Fig. 3 An illustration of competitive enzyme-linked immunosorbent assay.

(G-CSF, GM-CSF), clotting factors (Factor VIII) and vaccines (hepatitis B). Some other biotechnology products are novel fusion proteins, such as etanercept (Enbrel[®]), whereas yet others are therapeutic antibodies, for example muromonab (Orthoclone OKT3), abciximab (ReoPro[®]), and trastuzumab (Herceptin[®]). Administration of these recombinant proteins to animals and humans may result in their recognition by the host's immune system as "non-self," resulting in an antibody response. Several factors contribute to the potential immunogenicity of these molecules, including the structure of the protein (including post-translation modifications), the presence of protein fragments or protein aggregates in the administered formulation, and the cell substrate or media components that may co-purify with the therapeutic agent. Clinical factors, such as genetic background, disease state or immune status, may also influence the immunogenicity of

a biological product, as well as the route and frequency of dose administration.

In both preclinical and clinical studies, evaluation of immune response to the administered product is necessary to evaluate accurately the safety, pharmacokinetic and pharmacodynamic response as anti-drug antibodies can bind the drug and neutralize the therapeutic effect, or eliminate it by Fc receptor-mediated uptake and destruction in the reticuloendothelial system. Conversely, the pharmacodynamic response could be enhanced if the distribution or clearance of the drug-antibody complex is altered (10). Antibodies to therapeutic agents may also react with the endogenous analog protein, abrogate its activity, and precipitate a severe adverse event. The presence of such antibodies may also interfere with the immunoassay for the quantitation of the therapeutic agent in biological matrices. In addition, the presence of

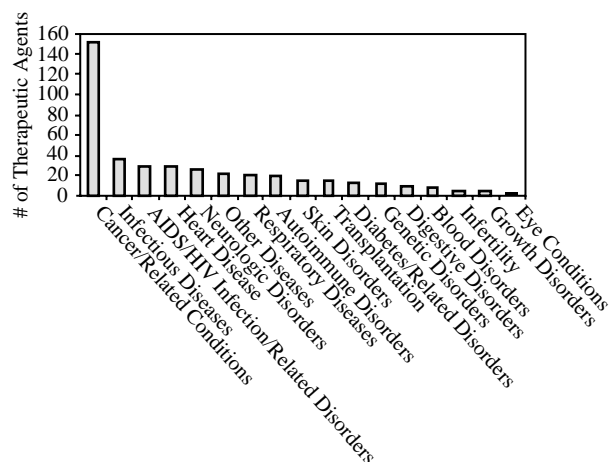


Fig. 4 Biotechnology products under development. (From New Medicines in Development, Biotechnology, PhRMA, April 1998.)

pre-existing antibodies (autoantibodies) to endogenous proteins can further complicate the quantitation of the molecule in biological fluids and its safety assessment in preclinical and clinical studies. Anti-drug antibodies may also interfere in imaging and diagnostic procedures utilizing antibodies; for example, human anti-mouse antibodies (HAMA) in the serum of patients treated with a murine antibody-based therapeutic may interfere in diagnostic assays using murine monoclonal antibodies.

Although several immunoassay formats [precipitin reactions, agglutination, radioimmunoassay (RIA), immunoradiometric assay (IRMA), western blot] have been used to detect and quantitate antibodies, enzyme immunoassay is the most commonly applied method. Several factors should be considered when developing an assay to detect the antibody response to a therapeutic agent. It is essential to understand the purpose for the antibody detection assay as it will influence the selection of the assay format. If the aim of the work is to detect only high-affinity antibodies with concomitant high specificity, a competitive assay format would be most appropriate. However, if all antibodies to the administered molecules (regardless of their affinity) are to be detected, best results will be obtained with a noncompetitive ELISA. In addition, random orientation of the antigen to expose all potential epitopes should be confirmed during assay development. These formats are illustrated in Figs. 5, 6. Product characteristics, such as the impurity profile, protein structure, and the presence of fragments and aggregates of the administered protein, need to be understood so that the most appropriate antigen is used in the assay. The potential loss of epitopes when the

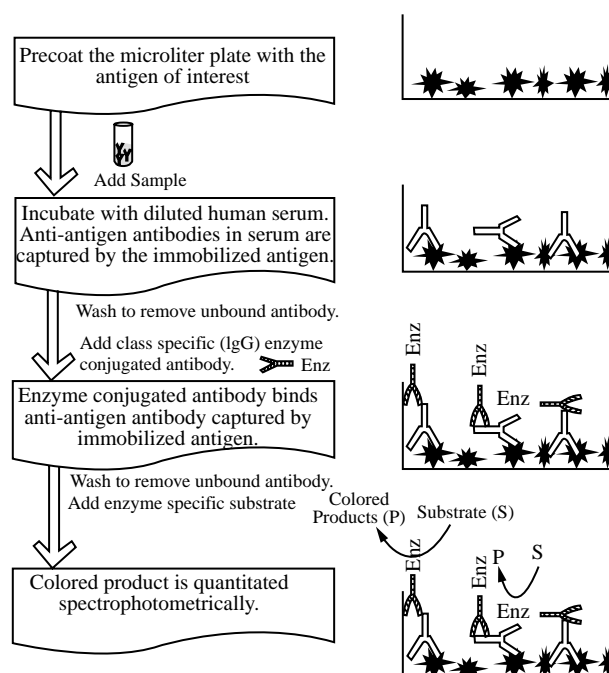


Fig. 5 An illustration of antigen capture antibody detection assay.

protein is directly adsorbed on the surface of the plate, the possibility of circulating antigen-antibody complexes, circulating antigen aggregates, and the source species of antibodies for capture and detection in the assay should also be evaluated. In some instances, cross-reactivity of the detected antibody with subclasses of the macromolecule (e.g., interferon) may need to be evaluated because the binding affinities of each subtype may vary greatly and may influence the interpretation of the results. Pre-existing antibodies can also limit the application of another antibody raised in the same species as that used in the assay configuration (e.g., as the detection antibody). Concomitant medications or high levels of the circulating macromolecular therapeutic agent may also be potential interferents in the assay.

Whether these assays should be established as truly quantitative is controversial. Although a quantitative response seems, from a bioanalytical perspective, desirable, the difficulties in so establishing the assay are significant and the added value is debatable. The primary challenge in developing a quantitative antibody assay is the lack of well-characterized, species-specific, polyclonal anti-drug antibody reference materials to be used as calibration standards (11). Heterologous polyclonal antibody could be used as reference standard (e.g., monkey antibody for human studies), or even nonspecific human IgG. However, although a true quantitative titer

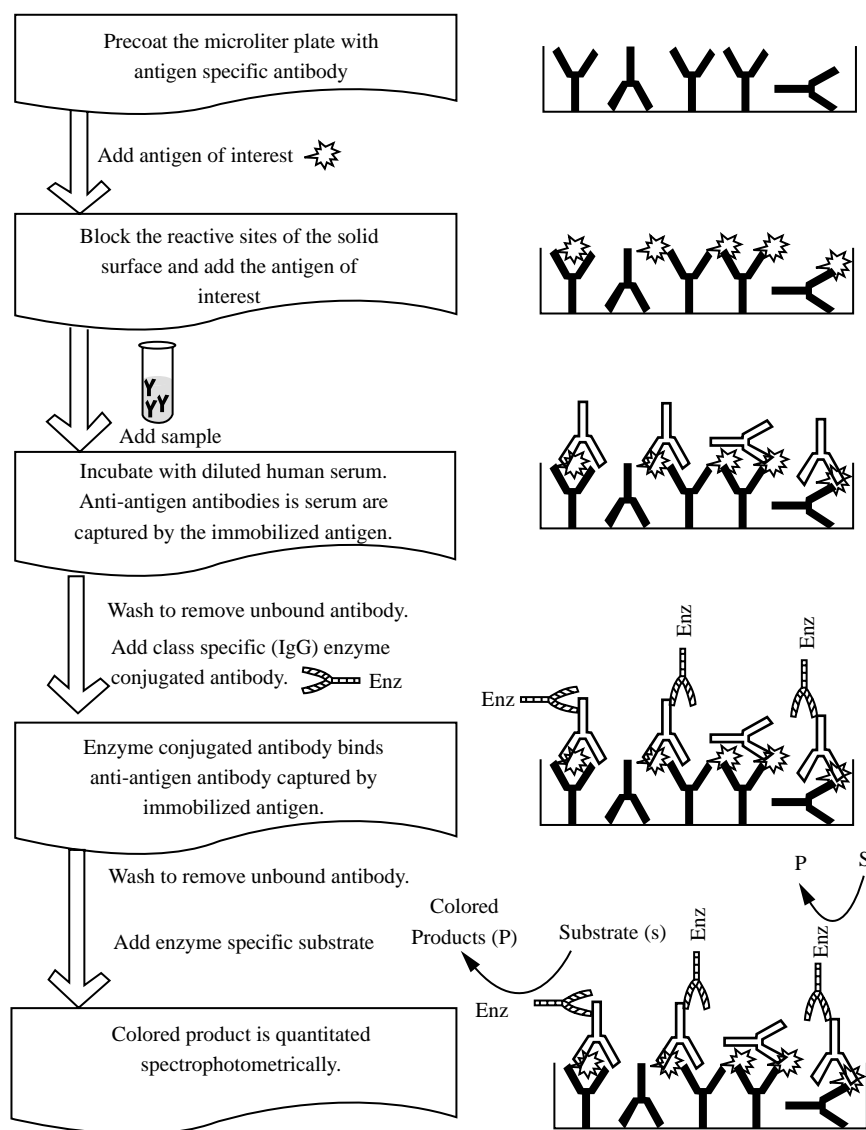


Fig. 6 An illustration of antibody detection method using an antigen–antibody capture configuration.

value may be calculated with reference to a standard of one of these types, the greater value of ELISA assays for antibodies to macromolecules lies in comparative titers determined over time following administration of the candidate therapeutic protein, in conjunction with the assessment of any clinical sequelae.

IMMUNOASSAYS FOR BIOMARKERS

A biomarker is defined as an *in vivo* biological response to a disease or a toxicological event. Pharmacological markers are a subset of biomarkers that respond to drug intervention,

whereas surrogate markers are markers that predict clinical endpoints. In a clinical setting, biomarkers may indicate the response of a disease to therapeutic intervention. In the pharmaceutical industry, correctly chosen biomarkers may help with compound selection, identify mechanism of action, predict dose, demonstrate efficacy and, in cases where sufficiently sensitive assays are unavailable, substitute for direct pharmacokinetic evaluation. Identification of relevant biomarkers at the discovery stage will optimize target selection and validation, facilitate pre-clinical toxicological evaluation, and markedly shorten the time needed to demonstrate proof of concept in early clinical trials. Rapid and effective implementation of appropriate biomarker assays will result in major resource

savings in drug development. A biomarker may be a physiological measurement, such as blood pressure or heart rate, an enzyme activity, or a quantifiable discrete molecule. Biomarkers can be either low-molecular-weight analytes or large macromolecules. Many of these molecules (especially macromolecules) are measured by immunoassay, particularly by enzyme immunoassay.

Many biomarkers are found in blood and urine, where the basal levels may be low or high, depending on the nature of the marker. Thus, one of the additional challenges in establishing enzyme immunoassays for these analytes is obtaining analyte-free matrix for use in preparation of calibration standard and quality control samples. Appropriate analyte-free matrix may sometimes be prepared by extraction of the analyte from matrix by such techniques as affinity chromatography (6) or charcoal adsorption, but care must be taken to ensure that the processed matrix is still representative of the matrix to be analyzed. Alternative solutions to this problem include use of the same matrix type from a different species in which the endogenous biomarker does not occur, pooled baseline samples of low biomarker concentration, or use of a protein-containing buffer. In each case, dilutional linearity of study samples, containing high concentration of the analyte, with the assay matrix should be demonstrated. However, although calibration standards may be prepared in an alternative matrix, whenever possible the quality control samples should be prepared in the matrix to be analyzed so that they reflect the assay performance for the study samples. The concentration range for the calibration standards should attempt to bracket the anticipated concentrations of the biomarkers in both physiological (including such factors as circadian rhythm and intra-individual variability) and pathological states. The concentration levels in the quality controls should also reflect concentration levels in physiological and pathological disease states. Physiological levels may be affected by gender, race, intra-individual variation, circadian rhythm, and seasonal variations, whereas the levels in pathological samples may be affected by stage of the disease, intercurrent disease, current therapies, and overall patient status.

Another challenge in developing assays for biomarkers is the inconsistent availability of well-characterized reference standards. For many macromolecular biomarkers, purified species-specific reference materials are not available. If available, proteins are often not characterized by a standardized method. Different sources of antibodies may also give different results for the same lot of reference standard. It is suggested that crossover studies using standards, quality control pools, and some study samples should be conducted when the source or lot number of reference standard or antibody is changed.

The data from crossover studies will help to normalize study results obtained with different lots of reagents. In addition, biomarkers may exist in multiple forms, such as isoforms, or have homology with other biomarkers in the same class of molecules. Thus, an immunoassay may be developed to measure all forms nonspecifically or target a specific isoform. Cross-reactivity of the antibody selected for use in a biomarker assay should be rigorously tested for potential cross-reactivity with other isoforms and homologous molecules.

VALIDATION OF ENZYME IMMUNOASSAYS

Validation of bioanalytical methods has been a subject of increasing attention over the last 10–15 years. A conference addressing this issue, cosponsored by the US Food and Drug Administration (FDA), the Health Protection Branch (HPB) Canada, the American Association of Pharmaceutical Scientists (AAPS), the European Federation of Pharmaceutical Societies and the United States Pharmacopoeia (USP), was convened in Arlington, Virginia, in December 1990. The proceedings of this meeting, which came to be known as the Crystal City meeting, were published (12) and have subsequently been used as an informal guideline for bioanalytical method validation.

Since 1990, bioanalytical method validation has been a topic of discussion at the International Conference on Harmonization, and was also the topic of a draft guidance document from the FDA (13). Although, many of the validation considerations for chromatographic assays also apply to immunoassay validation, some of the unique considerations for immunoassays were not addressed in the draft guidance. The present authors, along with a number of pharmaceutical industry colleagues, have recently discussed validation issues specific to immunoassays (14). Topics covered in that review included the proper use of quality control (QC) samples for acceptance of assay runs, and statistical aspects of assay validation. Specifically addressed was the issue of how differences in bioanalytical techniques should be considered when developing validation acceptance criteria. In recent years, several publications have reviewed issues related to validation of immunoassays both broadly (15, 16), and specifically for assays for macromolecules (17). In 2000, two conferences were held to “revisit” the issue of bioanalytical methods validation in general (18), as well as specifically for assays for macromolecules, primarily immunoassays and cell-based assays (19).

As for any bioanalytical method, the extent of validation for an immunoassay should be related to the

intended application of the assay. Thus, if an immunoassay is intended to support rapid screening in discovery R&D, the characterization of specificity and the accuracy and precision specifications may be less stringent than if the assay is used to support preclinical and clinical development studies. Indeed, an assay for discovery support may be designed to detect active metabolites as well as parent molecule, so that an estimate of total, circulating, pharmacologically active agents may be made. However, at the development stage, such an assay may be applicable only with clear definition of the cross-reactivities of both the parent and the active metabolite in the assay.

DIFFERENTIATING CHARACTERISTICS OF IMMUNOASSAYS AND THEIR IMPACT ON VALIDATION

The key difference between chromatographic assays and immunoassays is the biological nature of the critical binding reagent in an immunoassay, namely the antibody. As antibodies are produced in biological systems, lot-to-lot variability may occur, and this is greatest for polyclonal antibodies because they are produced in whole animals. Monoclonal antibodies, produced from a single cell line in an *in vitro* biological system, tend to have much lower variability between different production batches. Immunoassays have at least one timed incubation period, which means that variations in binding affinity and avidity between different lots of antibody reagents can result in differences in rates at which equilibrium is reached. Another important difference between immunoassay and chromatographic assays is the use of analyte–protein and antibody–protein conjugates in immunoassays. During validation, the stability of these critical reagents should be demonstrated to ensure that their degradation does not adversely affect assay performance. The availability of reference standards may also differentiate chromatographic and immunoassay procedures. Thus, although well-characterized reference standards are readily available for low-molecular-weight xenobiotics, it is sometimes much more difficult to obtain similarly well characterized reference materials for macromolecule immunoassays. Macromolecular products are not always available in a highly purified state, and are often characterized in terms of biological activity rather than percentage purity. In some cases, more widely studied proteins are available as reference standards from independent agencies such as the World Health Organization.

Another important difference between immunoassays and most chromatographic assays, with significant implications for validation, is the nonlinear nature of the relationship between concentration and response for immunoassays. For optimum calibration curve fit, it is sometimes appropriate to include calibration points above and below the defined limits of quantitation. The calibration curve fit algorithm should evaluate the overall fit of the experimental data with and without the use of these additional standards in the asymptotic regions of the calibration curve. In general, the authors have observed that the use of additional standards in the asymptotic regions improves the accuracy and precision of the assay at the limits of quantitation. The option to include additional standards is not addressed in the draft FDA guideline for bioanalytical method validation (13), which states that the lowest and highest concentrations on the calibration curve should serve as the limits of quantitation.

Precision, accuracy and specificity also raise some interesting and different considerations for immunoassay methods. The biological nature of the reagents and the antibody–antigen reaction can potentially confer higher imprecision on immunoassays, and a larger number of validation runs may be needed to determine the true precision of the method. This higher imprecision also increases the likelihood that a higher number of assay runs will not meet the so called 4-6-20 rule relating to quality control acceptance criteria for method implementation (20, 21), leading to a recommendation (14, 19) that the 20% limit for accuracy be relaxed to 25% for immunoassays. However, as many immunoassays can also have precision comparable to that of chromatographic methods, the acceptance criteria should be determined based on the demonstrated capability of the method during validation and not set arbitrarily to 4-6-25 for all immunoassays.

Although antibodies can be exquisitely specific, the biological nature of these reagents also poses some new specificity issues for immunoassays when compared to chromatographic assays. In immunoassays, the analyte of interest is usually detected and quantitated directly (i.e., without prior extraction) in complex biological matrices such as serum, plasma, or urine. Furthermore, the specificity of the assay can potentially be compromised if cross-reacting metabolites of the analyte of interest are present in the study sample. Nonspecificity in immunoassays can arise from a variety of sources, but may be broadly classified as “specific” or “nonspecific” nonspecificity. Specific nonspecificity can arise from interferents that share similar physicochemical characteristics with the analyte of interest, and include metabolites, degraded forms of the analyte, impurities, or concomitant medications. For macromolecules, post-translation modified proteins,

protein aggregates or host anti-idiotypic antibodies may affect the specificity of the assay in a “specific” manner. Nonspecific nonspecificity arises from a variety of factors unrelated to the analyte, and is often referred to as “matrix effects.” Matrix effects may be due to hemolysis, lipemia, ionic strength differences, pH, serum proteins such as complement or rheumatoid factor, anticoagulants, binding proteins, autoantibodies or heterophilic anti-IgG antibodies. Nonspecific interferences may arise from the matrix chosen for preparation of calibration standards and quality control samples, or the study sample itself. Evaluation of the matrix should include comparison of the concentration–response relationship in spiked (and unspiked) matrix to that in a buffer matrix. Dilution with buffer may adequately decrease the intensity of matrix effects; however, if this approach fails, sample cleanup or full analyte extraction from the matrix may be needed. In these cases, it is important to treat all samples, including calibration standards, quality control or validation samples and study samples, identically, to obviate the need for recovery corrections in the assay.

VALIDATION CONSIDERATIONS FOR IMMUNOASSAYS FOR LOW-MOLECULAR-WEIGHT XENOBIOTICS AND MACROMOLECULES

An essential prerequisite component of small-molecule immunoassay validation is the demonstration of specificity for the analyte of interest. Assay interference (cross-reactivity) due to known metabolites, concomitant medications and, in some cases, endogenous molecules, should be evaluated. These experiments should assess the cross-reactivity of the potential interferents individually and in combination with each other and the analyte of interest to simulate the most likely biological milieu for the analyte. For small molecule immunoassays, the specificity of the assay should be established, whenever possible, by conducting a comparator study over a specified number of analytical runs, using a different, validated and specific bioanalytical method, such as LC–MS. Samples for these comparative analyses should be selected from an earlier study (incurred samples), and should have been collected at two or more time points following drug administration (e.g., approximate time of maximum plasma concentration and a succeeding time corresponding to several elimination half-lives) to allow performance of the immunoassay to be evaluated in an environment of increasing metabolite(s) concentration. If the immunoassay meets the predefined criteria for accuracy and precision

relative to the reference method, the immunoassay may be considered equivalent to the reference method.

The metabolism/catabolism of protein drugs and protein drug candidates is generally much less clearly elucidated than that of conventional small-molecule therapeutics. The dearth of information regarding the metabolism of protein-based drug candidates can hamper efforts to develop a highly specific and accurate enzyme immunoassay. In addition, there are few, sufficiently sensitive, comparator methods available to perform comparative analysis. Approaches to define specificity and, thus, predict reliability of a given assay, include epitope mapping experiments or the use of chromatographic separation prior to immunoassay, although the latter approach can be cumbersome and may result in markedly reduced assay sensitivity. Consequently, complete investigation of assay specificity for a therapeutic protein is more difficult than for a low-molecular-weight xenobiotic. The immunoreactivity of the analyte may be decreased with relatively minor changes in the region of antigenic determinants, such as a change in the amino acid sequence or by oxidation/deamination of an amino acid. On the other hand, proteolytic formation of major fragments of the parent protein may result in retention of the antigenic determinants and preserve the immunoreactivity of the protein metabolite. Such structural changes may or may not result in changes in biological activity, so that immunoreactivity may or may not correlate with biological activity in the study samples collected after administration of the protein. If assay sensitivity permits, methods such as chromatography or electrophoresis, coupled with immunoassay, may shed some light on the structural nature of the compounds in the study samples giving a positive response in the immunoassay. For many recombinant proteins, antigenic determinants of the therapeutic agent are often indistinguishable from those of the endogenous equivalent protein and, therefore, assays for such recombinant proteins are prone to interference from endogenous analog proteins. Such an occurrence poses challenges for the validation analyst in the selection of the preferred homologous matrix for the preparation of calibration standards, validation pools and quality control samples. The problem is accentuated when the administered doses of the exogenous product are so low that circulating concentrations are not increased markedly over the background endogenous concentrations. Clearly, in cases where in vivo concentrations of the exogenous product are very high after dosing, the contribution of low basal levels of the endogenous analyte to the total measured concentration may be small enough to be ignored. An alternative approach to this problem is to remove the endogenous analyte from the matrix by one of a number of methods, as discussed previously (under the heading

“Immunoassays for Biomarkers”) or to use a heterologous biological matrix devoid of the specific interfering endogenous substance for the preparation of calibration standards and control samples. In all of these approaches, it is important that the final prepared calibration curve reflects negligible bias due to the presence of endogenous analyte.

VALIDATION CHALLENGES FOR ENDOGENOUS ANALYTE AND BIOMARKER IMMUNOASSAYS

Immunoassays are often developed for the quantitation of endogenous equivalents of therapeutic molecules and biomarkers. Challenges in developing immunoassays for these compounds are similar to those experienced with macromolecules. As the biomarker is normally always present in the matrix of interest, it is difficult to obtain analyte-free matrices. Standard curves and lower limit of quantitation (LLOQ) validation pools may be prepared by choosing and pooling matrix from individuals with low baseline concentrations, diluting baseline samples with a protein-based buffer, or using an alternative species matrix with negligible concentrations of the analyte. The upper limit of quantitation (ULOQ) can be established by fortifying the baseline sample with the analyte of interest. Whenever possible, the appropriate biological matrix should be used for QC sample preparation. This may be a systemic matrix such as whole blood or plasma, or target-specific matrix such as sputum, cerebrospinal fluid, aqueous humor, platelets, T-cells or tissues. Alternatively, if no matrix effects can be demonstrated, quality control pools may be prepared in an “analyte-free” protein-based buffer.

As with macromolecules, obtaining well-characterized reference material can be difficult. Whenever possible, reference materials should also be species-specific. In situations where well-characterized standards are not available, crossover studies should be conducted to permit normalization of data obtained using reference standards from different vendors or different lots.

VALIDATION CONSIDERATIONS FOR IMMUNOASSAYS FOR ANTIBODIES

Anti-drug antibodies are polyclonal in nature and rarely does one have access to species-specific (especially human) anti-drug antibodies to prepare a calibration curve. Despite the problems associated with developing a quantitative anti-drug antibody assay, most validation parameters for immunoassays still apply. Although true accuracy cannot be determined, relative recovery using

quality control samples can be monitored through the life of the assays. Assay specificity evaluation should include assessment of any nonspecific binding of the antibody to the microtiter plate, potential interference of the administered protein, endogenous protein analogs, concomitantly administered drugs, and antigen–antibody complex or cross-reacting antibodies that may be present in the sample under evaluation.

A major issue with anti-drug antibody assays is definition and interpretation of a positive antibody response. During validation, a negative cut-off value, to distinguish antibody negative and positive results, should be determined by evaluating the analytical noise (imprecision) of the assay and the background absorbance readings from individual baseline samples from healthy volunteers and patients from the appropriate disease population. The recommended number of baseline samples is at least 25 (preferably 100) from each of the volunteer and patient groups. If the background responses tend to cluster closely together and the assay is precise, a negative cut-off can be defined as the mean absorbance ± 3 SD at a given dilution factor(s). If, however, there is greater variability in the background absorbance, developing a negative cut-off is more difficult. The cut-off should be based on an acceptable level of false negative and false positive results. The assay validation scheme should include a process to distinguish true responses from false positives. This is particularly important as auto-antibodies against various proteins may be present in otherwise healthy individuals. Several approaches can be taken to elucidate whether an apparent antibody response is truly positive, such as an alternative method for detecting the antibody (e.g., western blotting). In some cases, when antibody response is evaluated against several different antigens, true positives may be distinguished from false positives by cross-reactivity patterns. Finally, examination of the response prior to drug administration, and the change in this response over time, with continued exposure to the agent, will normally distinguish between true and false positives. In addition, results of *in vitro* neutralizing activity assays and clinical effects (*in vivo* neutralization) may provide further support for the presence of clinically significant, drug-specific antibodies.

MODERN TECHNOLOGICAL ADVANCES RELATED TO ENZYME IMMUNOASSAY

Immunoassays are inherently sensitive and specific. However, with continued need to develop increasingly sensitive assays to support preclinical and clinical

studies, there have been ongoing efforts to enhance the capabilities of these techniques. Advances in critical binding reagents, detection systems, new assay formats and automation have resulted in improved immunoassay technology.

Critical Binding Reagents

Although most immunoassays have used polyclonal antibodies as the critical binding reagents, development of monoclonal antibodies by Kohler and Milstein in 1975 (22), has resulted in their widespread use, particularly in assays for macromolecules. Their unique epitope specificity conveys advantages in double antibody immunoassays for proteins, where one monoclonal antibody may be used to capture the protein by a specific subunit or epitope, and another, directed against a different region or subunit of the protein, may be used to detect it. Use of antibodies against specific regions of the molecule can enhance the specificity of the assay such that one can distinguish the parent molecule from its catabolic products, one isoform from another, or an individual protein from its family members. Although monoclonal antibodies are highly specific, their affinity for the antigen is generally lower than that observed with polyclonal antibodies. Consequently, competitive immunoassays established with monoclonal antibodies as the critical binding reagent are generally less sensitive than those using polyclonal antibodies. For this reason, the application of monoclonal antibodies to competitive immunoassays for low-molecular-weight analytes has been limited.

A nonantibody binding reagent that has received increasing attention recently is the aptamer. Aptamers are oligonucleotide sequences that bind ligands or antigens in a way that is similar in many respects to antibody–ligand interactions (23). Thus, aptamers have been shown to bind with high affinity and selectivity to molecules as diverse as proteins and low-molecular-weight ligands. Aptamer libraries have also been generated, from which members with the desired binding properties may be identified and their concentrations enriched for use in binding assays. The use of these molecules as complements to, or substitutes for, antibodies has been reviewed elsewhere (24). These reagents offer the promise of similar sensitivity and specificity achievable with antibodies, without the need for time-consuming *in vivo* work to generate them.

Phage display technology provides a source of recombinant antibodies with defined affinity and specificity for use in immunoassay, without the need for extended immunization of animals. This approach

involves genetic manipulation of the coat proteins of the filamentous phage, a bacteriophage that lives on *Escherichia coli*. In one approach, the coding sequences of the antibody variable regions (Fv) are first isolated from spleen cells of immunized mice. The coding sequence for a single chain Fv fragment is then fused to the phage coat protein. With the assistance of a series of molecular biology steps, a library containing millions of single-chain antibodies can be displayed on the surface of the phage particles and released into the medium. Through a series of binding and elution steps (“panning”), the mixture of antibodies with the desired affinity and selectivity may be sequentially enriched. To obtain single-chain antibodies of suitable affinity and selectivity against new drug entities, one can repeatedly and rapidly screen the recombinant antibody phage library and avoid the traditional, time-consuming process of antibody production. A detailed discussion of antibody phage display technology is provided in reviews by Hoogenboom (25) or Peterson (26).

Molecularly imprinted polymers (MIPs) also offer some potential as synthetic alternative binding reagents (27) for assay of small molecules. These binding reagents are synthesized by polymerization of functional monomers (e.g., methylacrylic acid, 4-vinyl pyridines) in the presence of the ligand (antigen), which acts as a template. Depending on experimental conditions during polymerization, the template ligand may interact with the monomers by either noncovalent interactions, reversible covalent interactions, or metal-ion-mediated interactions, with the noncovalent approach being most commonly used. Upon completion of the polymerization reaction, the ligand may be washed out to leave its imprint in the polymer. The binding properties of these molecular imprints are characterized by remarkable specificity for the ligand originally imprinted. These MIPs have been applied in place of biologically derived antibodies for the binding assay of a range of low-molecular-weight analytes following extraction of the analyte from the biological matrix. The assays were initially conducted in an organic solvent. Under these conditions, a good correlation was found between the MIP binding assay and an established radioimmunoassay for theophylline (28). Subsequent developments of aqueous assay conditions led to a MIP-based assay for propranolol directly in plasma (29). These high-affinity binding reagents have high chemical and thermal stabilities, resulting in long shelf lives at ambient temperature, an advantage over antibodies. Additionally, although most MIP-based assays have employed radiolabeled analyte as the detection system, some recent studies have successfully used a fluorescence

detection system. However, there are currently some limitations with this technology. The MIPs often have lower binding affinities than do antibodies, resulting in lower assay sensitivities than immunoassays have. The use of MIPs is also limited to analytes stable in organic matrices, and further research is necessary to establish them fully in aqueous media, so that they are competitive with conventional immunoassays for the direct analysis of biological fluids.

Detection Systems

The need for greater sensitivity to monitor extremely low concentrations of either highly potent therapeutic agents, endogenous biomarker molecules, or environmental toxicants has been the primary driver in the development of newer detection systems as more sensitive alternatives to the chromogenic substrates normally used in enzyme immunoassay.

There are numerous examples in the literature of fluorescence being used in place of ultraviolet light absorption as the end-point detection system for an immunoassay. In particular, methods using time-resolved fluorescence detection (30) offer high sensitivity while largely avoiding the problem of background fluorescence in complex matrices by allowing this short-lived fluorescence to decay, before fluorescence of the labeled antibody complex is measured. These procedures frequently employ lanthanide chelates as the long-lived fluorophores. An interesting recent example of this method is the time-resolved fluoroimmunoassay for plasma enterolactone, a lignan produced from fiber-rich foods by intestinal bacteria, and thus claimed to be a biomarker of a healthy diet (31). This assay used a derivative of enterolactone coupled with a europium chelate as the fluorophore, and achieved sensitivity of 1.5 nmol/L. Examples from the pharmaceutical field include the assays for enalaprilat in human serum (32) and sampatrilat in human plasma (33), which are characterized by lower limits of quantitation of 200–500 pg/mL. In the technique of fluorescence polarization, detection is based on the change in polarization of light emitted by a fluorophore molecule when bound to an antibody. This change in polarization is correlated with the concentration of unlabeled antigen, and a standard curve is developed to interpolate the analyte concentration in an unknown sample. This method, which has the advantage of being homogeneous and easily automatable, has been widely available commercially for some time for such applications as therapeutic drug monitoring, but more recently has seen new application to high throughput screening in drug discovery (34).

Chemiluminescence (35) offers yet another sensitive detection system, which is easily implemented with simple instrumentation, but suffers to some extent from background interference in complex matrices. A recent example of an enzyme immunoassay with chemiluminescence as the detection system is the assay for 8-oxoguanine in DNA (36), which uses a secondary antibody conjugated with peroxidase–anti-peroxidase complex and a substrate solution containing hydrogen peroxide, luminol and *p*-Iodophenol.

The utility of chemiluminescence as a detection system has been extended greatly with the development of electrochemiluminescence. Although electrochemiluminescence has been studied since the 1960s, only relatively recently has the system been commercialized by the IGEN Corporation (Rockville, MD). In their system (37), a precursor molecule, tripropylamine (TPA) diffuses to an electrode surface to be activated, resulting in the excitation of a reporter molecule, ruthenium *tris*-bipyridyl. When the reporter molecule returns to ground state from the excited state, it emits a photon of light at a specific wavelength, which is detected by a sensitive photomultiplier tube. The system includes an electrochemical flow cell and magnetic bead technology to trap the ruthenium-tagged molecules on the electrode and thus allow the electrochemical cycle to proceed. Thus, in a typical immunoassay for a macromolecule, two different anti-analyte antibodies (recognizing different epitopes) may be used. One antibody may be labeled with biotin, which complexes with streptavidin-complexed magnetic beads, whereas the second antibody is labeled with the ruthenium complex. During incubation, the analyte is sandwiched between the two antibodies, after which the mixture is drawn into the flow cell, and the antigen–antibody complex is trapped on the electrode surface by magnetic forces. After washing, the amount of ruthenium complex in the trapped antigen–antibody complex is measured by activating the electrode and quantitating the emitted light. This method has been applied to bioanalysis of many analytes; an example is the biotin–avidin coupled assay for interferon alfa-2b in human serum, with a sensitivity of 4 IU/mL (38).

Detection systems have extended into the area of biosensors and immunosensors, with the application of surface plasmon resonance (SPR) in immunochemistry. SPR is one of a number of optical immunosensor techniques (39) in which a change in the resonance angle of incident light occurs when antigen–antibody binding takes place. In an instrument such as the Biacore, a typical experimental design might involve adsorption of an antibody to the gold or silver surface of a microcell, which is backed by a prism or diffraction grating. When a solution containing the antigen of interest flows through

the cell, the formation of the antigen–antibody complex results in a change in the angle of the reflected light (resonance angle) at the metal surface. The shift in the resonance angle has been reported to have a linear relationship to the concentration of antigen added to the system. As no labeled reagents are needed for this method, SPR can be quite simple; however, the technique cannot distinguish between antigen recognition and nonspecific binding, and poor sensitivity can also be a limiting factor. However, recent advances include the incorporation of liposomes linked to a sandwich immunoassay format, resulting in picomolar sensitivity in an assay for interferon (40). In another application for a low-molecular-weight xenobiotic, sulfamethazine, in milk, the analyte was covalently coupled to the gold surface of the sensor chip (41). The final response was the result of competition between covalently bound sulfamethazine and free antigen in calibration solutions and study samples for binding sites on polyclonal antibodies, also in solution. The assay sensitivity was in the range of 1.7–8.0 $\mu\text{g/kg}$.

Assay Configuration and Automation

When immunoassay specificity has not been inherent in the antibody employed in the assay, separation steps such as high performance liquid chromatography have been applied prior to immunoassay. One of the more promising of such coupled methods is capillary electrophoretic immunoassay (CEIA) (42). This method offers a number of potential advantages, including a smaller sample size and lower reagent consumption, simple and readily automatable process, potential for simultaneous determination of multiple analytes, and a broad range of detection techniques. When coupled with laser-induced fluorescence as the detection method and enzyme amplification, CEIA appears to be competitive with standard immunoassay techniques, with assay sensitivity in the 10-pM range. The resolving capabilities of CEIA can separate antibody-bound from free antigen, followed by application of the detection method. Thus, the technique is configured for on-line application. Although clearly having the potential for high throughput, CEIA has been applied only in a serial mode to date, analyzing one sample at a time. CEIA has been applied widely to the characterization of antibodies, as well as to immunoassay of a number of low-molecular-weight analytes, including digoxin, morphine and cortisol.

Immunoassay has achieved considerable success in the medical diagnostic arena largely due to its facile adaptation to automation, high sample throughput and relatively low per sample cost. Issues and challenges involved in the automation of immunoassays have been

addressed recently by Bock (43). An interesting recent application of enzyme immunoassay in an automated mode has been in the support of high throughput screening in drug discovery (44). In this case, the assay reagents are incorporated into a gel matrix rather than in a multiwell format, which permits 1,000–10,000 assays to be run per day, with the assistance of automation, by a single technician. The development of multianalyte immunoassays in miniaturized, microarray formats has also been reported (45).

The move toward automation has also led to development of online and flow injection immunoassays. These methods involve sequential injection of assay reagents and antigen–antibody reactions in flowing systems. One such system with some promise is flow injection renewable surface immunoassay (FIRSI), with fluorescence detection (46). In this flowing system, antibody-coated beads are retained on a flat surface adjacent to the detector; labeled and unlabeled analyte are then injected and flow over the beads, while reaction occurs. The beads are then washed, the final antibody-bound reading occurs at the detector and the flow is reversed to remove the beads in preparation for the next injection. Similar systems, sometimes using magnetic particle-coupled immunoglobulin to facilitate the separation of antibody-bound and free ligand (47), have been used in conjunction with electrochemiluminescence (48) or laser-induced fluorescence (49) as the detection methods.

CONCLUSIONS

Enzyme immunoassay is widely used, both in competitive and noncompetitive formats, for the bioanalysis of a broad range of low-molecular-weight compounds and macromolecules. Through the use of fluorogenic substrates and amplification systems such as avidin–biotin, the sensitivity of enzyme immunoassay has been developed to equal or exceed that of radioimmunoassay (50). The technique has found particularly wide applicability in the determination of new recombinant proteins, in demonstrating antibody responses to macromolecules, and in the measurement of biomarkers of disease, as well as in diagnostic medicine.

As for all bioanalytical methods applied to support of drug development, validation of immunoassays is important. However, several validation issues need special attention for immunoassays. These include stability of the critical reagents, the curvilinear nature of the calibration curve, the greater variability of immunoassays, and, particularly important, the specificity of the assay.

Finally, a number of newer binding reagents, detection methods, assay configurations, and automation applications are being investigated to develop further the potential of immunoassays. These include binding reagents requiring little or no animal immunization for their production, such as phage display antibody libraries, aptamer libraries, and synthetic molecular imprints. Detection methods such as electrochemiluminescence and surface plasmon resonance hold promise of sensitivity equal to, or better than, that of radioimmunoassay without the limitations of radioactivity. Progress has also been made in automation and miniaturization of immunoassays, as well as online techniques, such as CEIA and flow injection methods. These efforts will continue to ensure that improvements in sensitivity and specificity of immunoassays will be widely available through commercialization.

REFERENCES

1. Yalow, R.S.; Berson, S.A. Immunoassay of Endogenous Plasma Insulin in Man. *J Clin. Invest.* **1960**, *39*, 1157–1175.
2. Colburn, W.A. Selecting and Validating Biologic Markers for Drug Development. *J. Clin. Pharmacol.* **1997**, *37* (5), 355–362.
3. Tai, H.H. Enzyme Immunoassay, 1st Ed.; *Encyclopedia of Pharmaceutical Technology*, Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1992; 5, 201–233.
4. Mattiasson, B.; Hakanson, H. Immunochemically Based Assays for Process Control. *Adv. Biochem. Eng. Biotechnol.* **1992**, *46*, 81–102.
5. Rubenstein, K.E.; Schneider, R.S.; Ullman, E.F. "Homogeneous" Enzyme Immunoassay: New Immunochemical Technique. *Biochem. Biophys. Res. Commun.* **1972**, *47* (4), 846–851.
6. Hage, D.S. Affinity Chromatography. *Clin. Chem.* **1999**, *45* (5), 593–615.
7. Got, P.A.; Scherrmann, J.M. Stereoselectivity of Antibodies for the Bioanalysis of Chiral Drugs. *Pharm. Res.* **1997**, *14* (11), 1516–1523.
8. Fujiwara, K.; Matsumoto, N.; Masuyama, Y.; Kitagawa, T.; Inoue, Y.; Inouye, K.; Hougaard, D.M. New Hapten-Protein Conjugation Method Using *N*-(*M*-aminobenzoyloxy) Succinimide as a Two-Level Heterobifunctional Agent: Thyrotropin-Releasing Hormone as a Model Peptide without Free Amino or Carboxyl Groups. *J. Immunol. Methods* **1994**, *175* (1), 123–129.
9. Schwarz, S.; Schumacher, M.; Ring, S.; Nanninga, A.; Weber, G.; Thieme, I.; Undeutsch, B.; Elger, W. 17 β -Hydroxy-11 α -(3'-Sulfanyloxypropyl)oxy-estra-1,3,5(10)-trien-3-yl sulfamate—A Novel Hapten Structure: Toward the Development of a Specific Enzyme Immunoassay (EIA) for Estra-1,3,5(10)-trien-3-yl sulfamates. *Steroids* **1999**, *64* (7), 460–471.
10. Working, P.K.; Cossum, W.A. Clinical and Preclinical Studies with Recombinant Human Proteins: Effect of Antibody Production. *Pharmacokinetics and Pharmacodynamics: Peptides, Peptoids and Proteins*; Garzone, P.D., Colburn, W.A., Mokotoff, M., Eds.; Ch. 13; Harvey Whitney Books: New York, 1991; 201–233.
11. Food and Drug Administration, Center for Biologics Evaluation and Review. Meeting of the Biological Response Modifiers Advisory Committee, Bethesda, Maryland, July, 15, 1999.
12. Shah, V.P.; Midha, K.K.; Dighe, S.; McGilveray, I.J.; Skelly, J.P.; Yacobi, A.; Layloff, T.; Viswanathan, C.T.; Cook, C.E.; McDowell, R.D.; Pittman, K.A.; Spector, S. Analytical Methods Validation: Bioavailability, Bioequivalence and Pharmacokinetic Studies. *Pharm. Res.* **1992**, *9*, 588–592.
13. Food and Drug Administration, Draft Guidance for Industry on Bioanalytical Methods Validation for Human Studies, Docket No 98D-1168, Federal Register **1999**; 517.
14. Findlay, J.W.A.; Smith, W.C.; Lee, J.W.; Nordblom, G.D.; Das, I.; DaSilva, B.S.; Khan, M.N.; Bowsher, R.R. Validation of Immunoassays for Bioanalysis: A Pharmaceutical Industry Perspective. *J. Pharm. Biomed. Anal.* **2000**, *21*, 1249–1273.
15. Findlay, J.W.A.; Das, I. Some Validation Considerations for Immunoassay. *J. Clin. Ligand Assay* **1997**, *20* (1), 49–55.
16. Findlay, J.W.A. Validation in Practice—Experience with Immunoassay. *Bio-International 2. Bioavailability, Bioequivalence and Pharmacokinetic Studies*; Blume, H.H., Midha, K.K., Eds.; Medpharm Scientific Publishers: Stuttgart, 1995; 361–370.
17. Findlay, J.W.A.; Das, I. Validation of Immunoassays for Macromolecules from Biotechnology. *J. Clin. Ligand Assay* **1998**, *21* (2), 249–253.
18. Shah, V.P.; Midha, K.K.; Findlay, J.W.A.; Hill, H.M.; Hulse, J.D.; McGilveray, I.J.; McKay, G.; Miller, K.J.; Patnaik, R.N.; Powell, M.L.; Tonelli, A.; Viswanathan, C.T.; Yacobi, A. Bioanalytical Method Validation: A Revisit with a Decade of Progress. *Pharm. Res.* **2000**.
19. AAPS Workshop. Bioanalytical Methods Validation for Macromolecules, Arlington, VA, 1–3 March 2000, American Association of Pharmaceutical Scientists: Alexandria, VA.
20. Smith, W.C.; Sittampalam, G.S. Conceptual and Statistical Issues in the Validation of Analytic Dilution Assays for Pharmaceutical Application. *J. Biopharm. Stat.* **1998**, *8* (4), 509–532.
21. Kringler, R.O. An Assessment of the 4-6-20 Rule for Acceptance of Analytical Runs in Bioavailability, Bioequivalence, and Pharmacokinetic Studies. *Pharm. Res.* **1994**, *11* (4), 556–560.
22. Kohler, G.; Milstein, C. Continuous Culture of Fused Cells Secreting Specific Antibody of Predefined Specificity. *Nature* **1975**, *256* (5517), 495–497.
23. Ellington, A.D.; Szostak, J.W. In Vitro Selection of RNA Molecules that Bind Specific Ligands. *Nature* **1990**, *346* (6287), 818–822.
24. Jaysena, S.D. Aptamers: An Emerging Class of Molecules That Rival Antibodies in Diagnostics. *Clin. Chem.* **1999**, *45* (9), 1628–1650.
25. Hoogenboom, H.R.; de-Bruine, A.P.; Hufton, S.E.; Hoet, R.M.; Arends, J.W.; Roovers, R.C. Antibody Phage Display

- Technology and its Applications. *Immunotechnology* **1998**, *4* (1), 1–20.
26. Peterson, N.C. Recombinant Antibodies: Alternative Strategies for Developing and Manipulating Murine-Derived Monoclonal Antibodies. *Lab. Anim. Sci.* **1996**, *46* (1), 8–14.
27. Andersson, L.I. Molecular Imprinting for Drug Bioanalysis: A Review on the Application of Imprinted Polymers to Solid-Phase Extraction and Binding Assay. *J. Chromatogr. B* **2000**, *739* (1), 163–173.
28. Vlatakis, G.; Andersson, L.I.; Muller, R.; Mosbach, K. Drug Assay Using Antibody Mimics Made by Molecular Imprinting. *Nature* **1993**, *361* (6413), 645–647.
29. Bengtsson, H.; Roos, U.; Andersson, L.I. Molecular Imprint-Based Radioassay for Direct Determination of S-Propriolol in Human Plasma. *Anal. Commun.* **1997**, *34* (9), 233–235.
30. Dickson, E.F.; Pollack, A.; Diamandis, E.P. Ultrasensitive Bioanalytical Assays Using Time-Resolved Fluorescence Detection. *Pharma. Ther.* **1995**, *66* (2), 207–235.
31. Adlercreutz, H.; Wang, G.J.; Lapcik, O.; Hampl, R.; Wahala, K.; Makela, T.; Lusa, K.; Talme, M.; Mikola, H. Time-Resolved Fluoroimmunoassay for Plasma Enterolactone. *Anal. Biochem.* **1998**, *265* (2), 208–215.
32. Yuan, A.S.; Gilbert, J.D. Time-Resolved Fluoroimmunoassay for the Determination of Lisinopril and Enalaprilat in Human Serum. *J. Pharm. Biomed. Anal.* **1996**, *14* (7), 773–781.
33. Venn, R.F.; Bernard, G.; Kaye, B.; Macrae, P.V.; Saunders, K.C. Clinical Analysis of Sampatrilat, a Combined Renal Endopeptidase and Angiotensin-Converting Enzyme Inhibitor II: Assay in Plasma and Urine of Human Volunteers by Dissociation Enhanced Lanthanide Fluorescence Immunoassay (DELFI). *J. Pharm. Biomed. Anal.* **1998**, *16* (5), 883–792.
34. Nasir, M.S.; Jolley, M.E. Fluorescence Polarization: An Analytical Tool for Immunoassay and Drug Discovery. *Comb. Chem. High Throughput Screen* **1999**, *2* (4), 177–190.
35. Baeynes, W.R.; Schulman, S.G.; Calokerinos, A.C.; Zhao, Y.; Garcia-Campana, A.M.; Nakashima, K.; De Keukeleire, D. Chemiluminescence-Based Detection: Principles and Analytical Applications in Flowing Streams and in Immunoassays. *J. Pharm. Biomed. Anal.* **1998**, *17* (6/7), 941–953.
36. Bruskov, V.I.; Masalimov, Z.K.; Usacheva, A.M. Chemiluminescence Enzyme Immunoassay of 8-Oxoguanine in DNA. *Biochemistry (Moscow)* **1999**, *64* (7), 803–808.
37. Yang, H.; Leland, J.K.; Yost, D.; Massey, R.J. Electrochemiluminescence: A New Diagnostic and Research Tool. *Biotechnology* **1994**, *12* (2), 193–194.
38. Obenauer-Kutner, L.J.; Jacobs, S.J.; Kolz, K.; Tobias, L.M.; Borden, R.W.A. A Highly Sensitive Electrochemiluminescence Immunoassay for Interferon Alfa-2b in Human Serum. *J. Immunol. Meth.* **1997**, *206* (1-2), 25–33.
39. Rabbany, S.Y.; Donner, B.L.; Ligler, F.S. Optical Immunosensors. *Crit. Rev. Biomed. Eng.* **1994**, *22* (5-6), 307–346.
40. Wink, T.; van Zuilen, S.J.; Bult, A.; van Bennekom, W.P. Liposome-mediated Enhancement of the Sensitivity in Immunoassays of Proteins and Peptides in Surface Plasmon Resonance Spectrometry. *Anal. Chem.* **1998**, *70* (5), 827–832.
41. Gaudin, V.; Pavy, M-L. Determination of Sulfamethazine in Milk by Biosensor Immunoassay. *J. AOAC Int.* **1999**, *82* (6), 1316–1320.
42. Bao, J.J. Capillary Electrophoretic Immunoassays. *J. Chromatogr. B* **1997**, *699* (1/2), 463–480.
43. Bock, J.L. The Era of Automated Immunoassay. *Am. J. Clin. Pathol.* **2000**, *113* (5), 628–646.
44. Karet, G. More Compound Tested with Less Automation. *Drug Dis. Dev.* **June-July 2000**, 48–51.
45. Silzel, J.W.; Cercek, B.; Dodson, C.; Tsay, T.; Obrenski, R.J. Mass-Sensing, Multianalyte Microarray Immunoassay with Imaging Detection. *Clin. Chem.* **1998**, *44* (9), 2036–2043.
46. Pollema, C.H.; Ruzicka, J. Flow Injection Renewable Surface Immunoassay: A New Approach to Immunoanalysis with Fluorescence Detection. *Anal. Chem.* **1994**, *66* (11), 1825–1831.
47. Sole, S.; Alegret, S.; Cespedes, F.; Fabregas, E.; Diez-Caballero, T. Flow Injection Immunoanalysis Based on a Magnetoimmunosensor System. *Anal. Chem.* **1998**, *70* (8), 1462–1467.
48. Arai, K.; Takahashi, K.; Kusu, F. An Electrochemiluminescence Flow-Through Cell and its Applications to Sensitive Immunoassays Using *N*-(aminobutyl)-*N*-ethylsoluminol. *Anal. Chem.* **1999**, *71* (11), 2237–2240.
49. Tao, L.; Kennedy, R.T. On-line Competitive Immunoassay for Insulin Based on Capillary Electrophoresis with Laser-Induced Fluorescence Detection. *Anal. Chem.* **1996**, *68* (22), 3899–3906.
50. Hashida, S.; Hashinaka, K.; Ishikawa, E. Ultrasensitive Enzyme Immunoassay. *Biotechnol. Ann. Rev.* **1995**, *1*, 403–451.

EXTRUSION AND EXTRUDERS

J.M. Newton

University of London, London, United Kingdom

INTRODUCTION

Extrusion is the process of forming a raw material into a product of uniform shape and density by forcing it through an orifice or die under controlled conditions. An extruder consists of two distinct parts: a delivery system which transports the material and sometimes imparts a degree of distributive mixing, and a die system which forms the material into the required shape. Extrusion may be broadly classified into molten systems under temperature control or semisolid viscous systems. In molten extrusion, heat is applied to the material in order to control its viscosity to enable it to flow through the die. Semisolid systems are multiphase concentrated dispersions containing a high proportion of solids mixed with a liquid phase. Extrusion is achieved by formulation to control the viscosity of the semisolid mass.

Extrusion is a continuous process that affords a consistent product at high throughput rates. The process has diverse applications in a range of industries utilizing extrusion equipment specifically designed or adapted to form a particular product. A description of the different types of extruders is given here, along with details that illustrate the versatility of extrusion processing.

THEORY AND CHARACTERIZATION OF EXTRUSION

The various types of extruders have the common feature of forcing the extrudate from a wide cross section through the restriction of the die. The force required and the characteristics of the extrudate produced are dependent on the rheological properties of the extrudate, the design of the die, and the rate at which the material is forced through the die. The theoretical approach to understanding the systems, therefore, is generally associated with dividing the process of flow into three sections: 1) entry into the die, 2) flow through the die, and 3) exit from the die.

Extrusion is dependent on the material, and the technique varies with the material studied. With regard to pharmaceuticals, most systems consist of particles dispersed in a fluid and, although consideration will be

given to plastics, the main emphasis is on paste extrusions. These differ in the fact that a fluid is present between solid particles. The relative position of solid and liquid can change during the various stages of the extrusion process, and hence produce effects different from those associated with single-phase systems.

If the die is considered as a simple capillary flow, the relationship between the rate of shear (γ) and die wall shear stress (τ_w) can be described by Eq. 1:

$$\tau_w = \gamma P \cdot R/2L \quad (1)$$

where P is the pressure drop across the length of capillary L and radius R . Corrections for entrance effects modify this equation by considering an increase in the length of the capillary to give Eq. 2:

$$\tau_w = \gamma P/2(L/R + n_b) \quad (2)$$

where n_b is the Bagley entrance correction (1). Determination of n_b can be made by measuring the pressure necessary to force extrudate through dies of different length-to-radius (L/R) ratio. Extrapolation of the graphs to zero pressure values gives the value of n_b as the intercept on the L/R axis (1).

Han and Charles (2) found experimentally that the exit pressure is actually above atmospheric pressure and proposed modification of Eq. 2 to Eq. 3 corrected for exit pressure losses:

$$\tau_w = (\gamma P - P_e)/2(L/R + n_b) \quad (3)$$

where P_e is the exit pressure. This value is difficult to determine and, because it is considerably lower than the pressure loss upstream and through the die, it is usually neglected. The upstream pressure loss can be considerable and can be determined as the intercept on the pressure axis at zero L/R ratio (the Bagley equation), giving Eqs. 4 and 5:

$$\tau_w = (P_T - P_0)R/2L \quad (4)$$

$$P_T = P_0 + 2\tau_w(L/R + n_b) \quad (5)$$

The upstream pressure loss includes pressure losses due to kinetic energy, head effects, elastic losses, and turbulence. Harrison (3) found for a series of pharmaceutical systems that the value of P_0 increases with increasing rate of passage through the die.

Rheological Curves

In addition to determination of the upstream pressure loss P_0 and the end correction n_b , Eq. 5 can be seen to provide a value for the shear stress τ_w in such systems. The slope of the graph P_T versus L/R has a gradient of $2\tau_w$; that is, the die-wall shear stress. The rate of shear at the die wall— $(dv/dr)_w$ —can be derived from the Hagen–Poiseuille's law, as in Eq. 6:

$$(dv/dr)_w = 4Q/R^3 \quad (6)$$

where Q is the volumetric flow rate and R the radius of the die. This assumes that the flow is Newtonian. If this is not the case, Jastrzebski (4) suggested that a correction should be made for the rate of shear, as in Eq. 7:

$$-(dv/dr)_w = \left(\frac{3n' + 1}{n'} \right) \frac{Q}{\pi R^3} \quad (7)$$

where n' is the degree of non-Newtonian flow; it is determined from the gradient of the graph of log-shear stress as a function of the log apparent shear rate. Wilkinson (5) has also indicated that these equations assume that: 1) the flow is laminar, 2) there is no slip at the die wall, and 3) the rate of shear depends only on the shear stress at the point of measurement and is independent of time.

Determination of shear rate versus shear stress curves by application of the ram extruder allow characterization of the rheological properties of the extruded material according to the basic type of curve, as expressed by Eqs. 8–11.

Newtonian:

$$\sigma_w = \gamma' \eta \quad (8)$$

where η is the apparent viscosity and γ' is the rate of shear.

Bingham body:

$$\tau_w = \sigma_y + \gamma' U \quad (9)$$

where σ_y is the stress necessary to be exceeded before Newtonian flow commences, yield value, and U is the plastic viscosity.

Power-law model:

$$\tau_w = K \gamma'^{n'} \quad (10)$$

where K is the power-law viscosity constant and n' is the degree of non-Newtonian flow. For values of n' less than 1, the material becomes less viscous with increasing shear rate (shear thinning), and for values of n' greater than 1, the viscosity increases with increasing shear rate (shear thickening).

Herschel–Buckley model:

$$\tau_w = \tau_y + K \gamma'^n \quad (11)$$

which allows for a system that has a yield value and a shear rate dependent on viscosity.

The application of these types of flow curves requires homogeneous materials that do not change in consistency with extrusion. Harrison et al. (6) found this not to be the case, and suggested that this was due to the presence of plug flow within the extrudate bulk and slip flow at the die wall.

The inability of the standard rheological models to quantitatively describe the process of flow into, through, and out of the die requires an alternative treatment. From a study of ceramic catalyst pastes, Benbow and Ovensten (7) and Benbow (8) assumed that there was broad plug flow at the center of the extrudate, with shearing occurring within a thin liquid layer at the die wall. Assuming that this layer behaves as a Newtonian liquid of thickness x and viscosity η , and that the initial shear stress to induce flow is τ_0 , the total die-wall shear stress τ_w at a given extrudate velocity V is given by Eq. 12:

$$\tau_w = \tau_0 + (\eta/x)V \quad (12)$$

The values of η and x cannot be determined directly and, therefore, Benbow et al. (9) introduced the term β , the die land viscosity factor, to replace η/x , as in Eq. 13:

$$\tau_w = \tau_0 = \beta V \quad (13)$$

Incorporation of this expression into Eq. 4 yields:

$$P_\tau = P_0 + 2(L/R)(\tau_0 + \beta V) \quad (14)$$

The value of τ_0 can be determined by plotting the extrusion pressure against the extrudate velocity V for extrusion through dies of constant value of L . The extrapolated value of extrusion pressure at $V = 0$ gives the value P_{0v0} at zero velocity, as shown by Eq. 15:

$$P_{\tau v0} = P_{0v0} + 2(L/R)\tau_0 \quad (15)$$

Thus, a graph of $P_{\tau v0}$ versus L/R provides the value of τ_0 as equal to half the slope. The value of β can be calculated by rearranging Eq. 14 to Eq. 16:

$$\beta = (P_{\tau_w} - P_{0w}) - (P_{\tau v0} - P_{0v0})2(L/R)V \quad (16)$$

where P_{τ_w} is the total extrusion pressure at extrudate velocity V , and P_{0w} is the upstream pressure loss at extrudate velocity V .

Further characterization of the system was suggested by Benbow (8) and Benbow and Bridgwater (10) in terms of a yield value σ_y associated with the convergence of flow

from the wide cross section of the feed to the narrow cross section of the die. This takes the form of Eq. 17:

$$P_0 = \sigma_y \ln(A_0/A) \quad (17)$$

where A_0 is the initial cross-sectional area and A that of the die. If the original and final cross sections are circular, Eq. 18 holds:

$$P_0 = 2\sigma_y \ln(D_0/D) \quad (18)$$

where D_0 and D are the barrel and die diameters, respectively. For materials that deform plastically and are time independent, the value of σ_y can be calculated from the intercept of the pressure axis divided by twice the natural log reduction ratio (D_0/D) for plots of P against L/R .

By combining this concept with those expressed above, Benbow et al. (9) and Benbow and Bridgwater (10) further modified the Bagley equation to Eq. 19:

$$P_T = 2(\sigma_{y0} + \alpha V) \ln(D_0/D) + 2(L/R)(\tau_0 + \beta V) \quad (19)$$

If the die land velocity factor β varies with the extrusion rate or the liquid layer at the die wall is non-Newtonian, Eq. 19 must be further modified to Eq. 20:

$$P_T = 2\sigma_y \ln(D_0/D) + 2(L/R)(\tau_0 + \beta^* V^{1-n}) \quad (20)$$

where β^* is a modified power-law constant and n the degree of non-Newtonian flow. If the flow velocity into the die is also dependent on the velocity of flow, Benbow et al. (9) and Benbow and Bridgwater (10) propose replacement of the yield value δ_y , by two empirical parameters, the initial die entry yield stress σ_{y0} and the die-entry yield-stress velocity factor α . Substituting in Eq. 18 gives Eq. 21:

$$P_0 = 2(\sigma_{y0} + \alpha V) \ln(D_0/D) \quad (21)$$

The fully corrected Bagley equation now becomes Eq. 22:

$$P_T = 2(\sigma_{y0} + \alpha V) \ln(D_0/D) + 2(L/R)(\tau_0 + \beta V) \quad (22)$$

The value of σ_{y0} can be obtained as the intercept from the derived zero-velocity graph of P_{0v0} as a function of L/R . The value of α , for a given system, is obtained from Eq. 23:

$$\alpha = (P_{0w} - P_{v0}) / (2 \ln(D_0/D)V) \quad (23)$$

where P_{0w} and P_{0v0} are obtained as described previously.

If the systems are treated as polymer melts instead of as paste (i.e., homogenous systems with no fluid migration during extrusion), further characterization of the wet masses can be achieved (11). The flow of melts through a capillary rheometer can be considered to show flow

streamlines converging and then accelerating, which according to Cogswell (12), is extensional flow. He separated the flow field into shear and tensile deformation and then described their calculation from the following equations:

$$TS = \frac{3}{8}(n+1)P_0 \quad (24)$$

where TS is the tensile stress (i.e., stretching), n is the power law index, and P_0 is the die-entrance press drop,

$$ESR = \frac{4\pi\gamma}{3(n+1)} = \frac{\gamma}{2} \tan \theta \quad (25)$$

where ESR is the tensile stretch rate, τ is the shear stress at the die wall, γ is the shear strain rate, and θ is the half angle of natural convergence.

$$EV \frac{TS}{ESR} \quad (26)$$

where EV is the apparent extensional (elongational) viscosity.

Such an approach depends on the flow fitting the power law model (i.e., Eq. 10), and that flow is not dominated by wall slip.

In addition to elongation flow, material can also exhibit elastic behavior. Two parameters that have been proposed (13, 14) to quantify this property are (1) recoverable shear RS and (2) compliance C . These can be derived from:

$$RS = \frac{P}{4\pi} \quad (27)$$

and

$$C = \frac{P}{4\tau^2} \quad (28)$$

Chohan (14) has used these to study the flow of branched polyethylene melt, and while what is exactly implied by these terms at high stretch rates is not clear, they are undoubtedly related to the elastic behavior of the material. The higher the values of each, the greater will be the elastic nature of the material.

MEASUREMENT OF RHEOLOGICAL PROPERTIES

The application of the theoretical treatment depends on the ability to measure the extrusion force and rate. Most commercial extruders do not allow for these types of measurement. Normal rheological equipment, such as cup-and-bob or cone-and-plate, do not have a suitable

geometry or instrumentation to handle materials of the consistency normally used. A ram extruder is a suitable experimental design.

The ram extruder, designed by Benbow and Ovenston (7), operates on a prefilled system and is used for experimental and small-scale extrusion (Fig. 1). It consists of a stainless steel barrel (2.54 cm internal diameter, approximately 20 cm in length), which acts as the material reservoir. The base is constructed to enable interchangeable dies, with central capillaries of varying dimensions to be bolted on. A rubber ring is inserted between the barrel and die to ensure a watertight connection. The piston, or ram, is a stainless steel rod that fits loosely into the barrel. A fluon ring positioned at its lower end provides a low-friction seal to prevent material escaping above the point where the piston moves down the barrel. The extrusion is a noncontinuous operation; first the material (50–100 g) is packed into the barrel and partially consolidated to a plug by inserting the piston. It is possible to add temperature control to the barrel to extrude materials that are thermosensitive. The barrel-and-die assembly is mounted on a rigid metal C-piece,

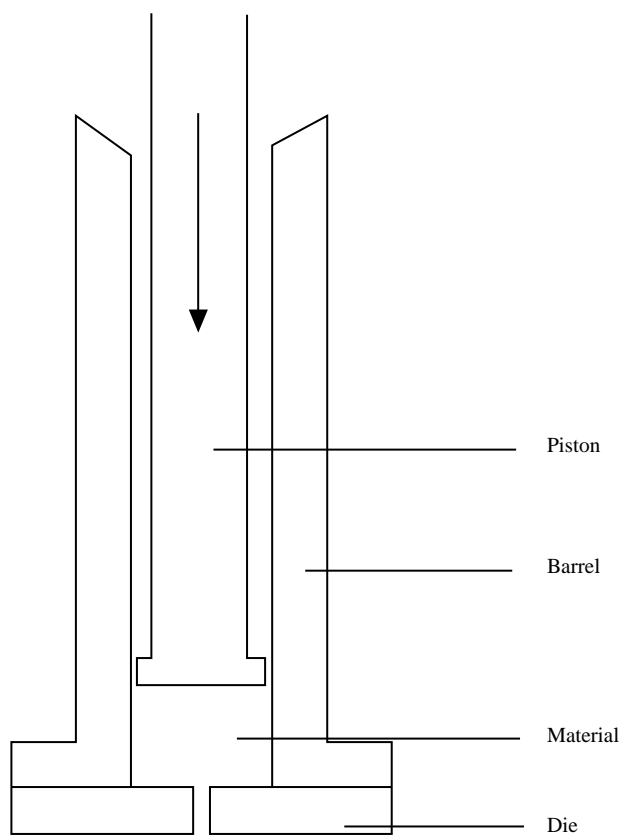


Fig. 1 Diagram of a ram extruder.

and a load is applied to the piston sufficient to extrude the material through the die. The ram extruder can be used in conjunction with an instrumented press. The piston is attached to the cross-head that may be driven down at various constant rates, and its displacement monitored by an attached displacement transducer. Output from this and the load cell is fed into an x - y chart recorder or computer. This arrangement enables the force acting on the material during extrusion to be recorded as a function of the displacement of the piston, and a force–displacement profile is produced.

A typical extrusion mixture produces three distinct regions, as shown in Fig. 2. In the compression stage, the piston descends into the barrel and consolidates the material into a plug prior to flow. This results in a large change in displacement accompanied by a small change in load. Eventually, the material is compressed to its minimum volume and maximum density. At this point, the pressure builds up while the material density is maintained. This is shown in the profile by a large increase in load accompanied by a minimal change in displacement. At the end of the compression stage, the pressure applied to the mass increases until it is high enough for the material to yield and commence flow. This is followed by a period of steady-state flow in which the force required to maintain the extrusion remains constant as the displacement increases. Forced flow occurs when steady-state flow can no longer be maintained. It leads to a gradual rise in extrusion force with displacement. This occurs often toward the end of the extrusion and is caused by the close proximity of the ram tip to the die face.

The force–displacement profile is altered by varying one of the extrusion parameters, such as the die diameter, L/R , or extrusion rate. For a given mixture, the relationship between the steady-state extrusion force, the die L/R at constant die diameter, and the extrusion rate can be represented graphically, as shown in Fig. 3. This is known as the Bagley plot (1). Used in this way, the extruder operates on a principle similar to that of a capillary rheometer, and expressions derived from capillary rheometry (described previously) may be used to characterize the properties of the wet powder mass. After conversion of the steady-state force values to pressure values, the slope of the relationship between the pressure and L/R is numerically equivalent to twice the value of the mean die-wall shear stress (according to Eq. 4). Plotting these values against the corresponding apparent die-wall shear rates (derived from Eq. 6) results in a flow curve that is unique for a particular wet-mass formulation (Fig. 4). The materials exhibit non-Newtonian flow and shear-thinning properties.

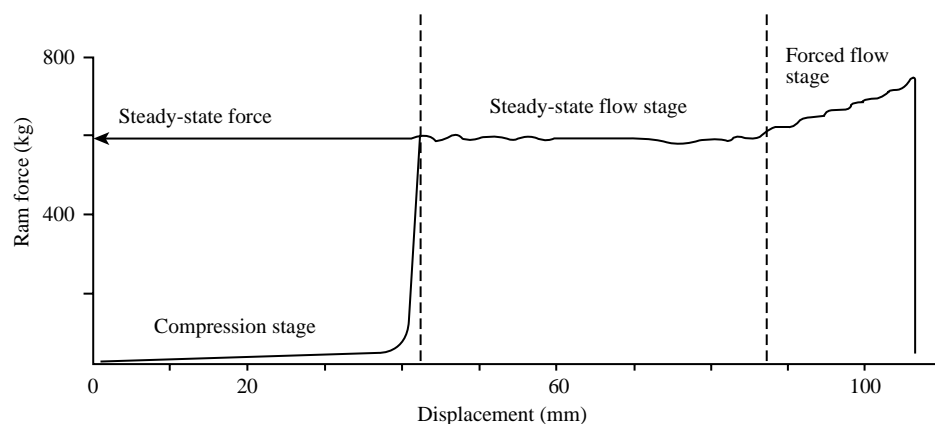


Fig. 2 Force-displacement profile for microcrystalline cellulose-lactose-water mixture.

PRACTICAL CHARACTERIZATION OF EXTRUSION SYSTEMS

The expression of the extrusion properties of pharmaceutical systems by numerical values could aid formulation. To be able to apply the theoretical approaches described previously, it is important to ensure that the restrictions of the systems are considered. One major problem with paste systems is that when subjected to pressure, there is phase separation resulting

in variations in the composition of the mass as it is being extruded. This can be detected by collecting the extrudate and measuring its water content (15, 16). Alternatively, magnetic resonance imaging has been used to quantify the water distribution within the barrel (17) and within the extrudate (18).

The extent to which die-wall slip is involved can be assessed by using dies of different lengths and diameters. An important characteristic that can be observed in the extrudate is its quality in terms of surface structure. Harrison et al. (19) have shown how this can vary from a smooth, regular surface via a rough, "shark-skinned" extrudate. There is obvious need to prevent this

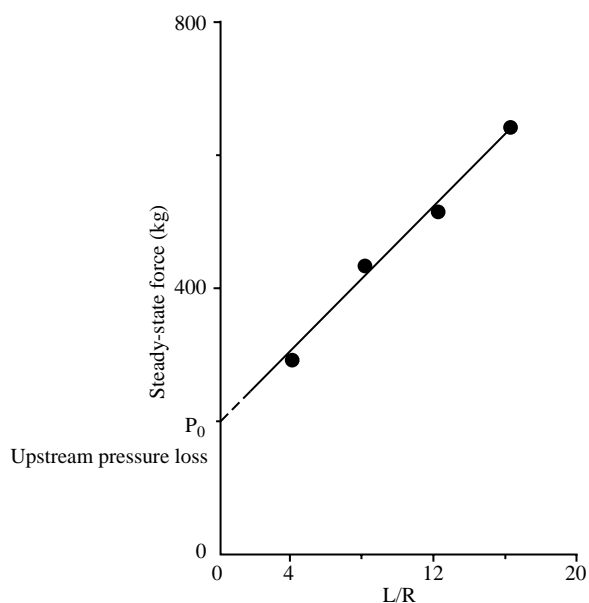


Fig. 3 Steady-state extrusion force as a function of the length-to-radius ratio of the die for microcrystalline cellulose-lactose-water (5:5:6) at constant die diameter (1.5 mm) and extrusion rate (20 cm/min).

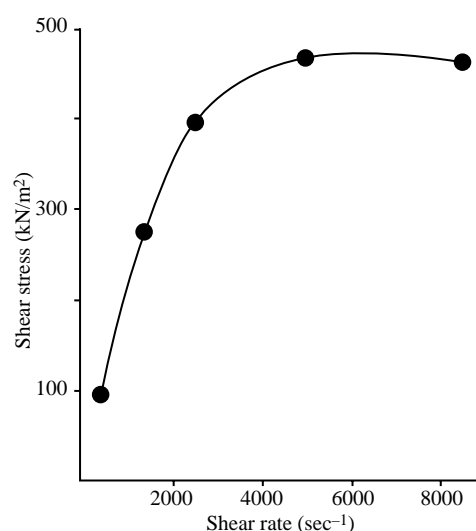


Fig. 4 Typical shear stress-shear rate flow curve for an extrusion mixture containing 50% microcrystalline cellulose extruded through a 1.5-mm diameter die.

phenomenon if extrudate of the correct quality is to be produced. The occurrence of the surface defects is associated with both the composition of the material and the operating conditions (e.g., die length and diameter and the rate of extrusion) (Fig. 5). Raines et al. (20) were able to relate the quality of the surface of the extrudate to the value of the yield stress at zero velocity δ_{y0} , in that those systems with a high value were smooth and regular while those with low values were shark-skinned.

Of the systems studied, most pastes show non-Newtonian behavior. This has important consequences for extruder design and operating conditions as material that are shear-rate dependent require careful handling. To date, most reported rheological investigations indicate that paste systems are shear thinning (i.e., their viscosity decreases with an increase in shear rate) (Fig. 4). Their extent of property can be quantified by obtaining the value of n' , the slope of the log shear-rate/log shear-stress graph. There is also some evidence that paste systems show plug flow (i.e., the central core of the extrudate moves at a constant velocity), while there is a thin layer of moisture at the die wall where shear takes place (21).

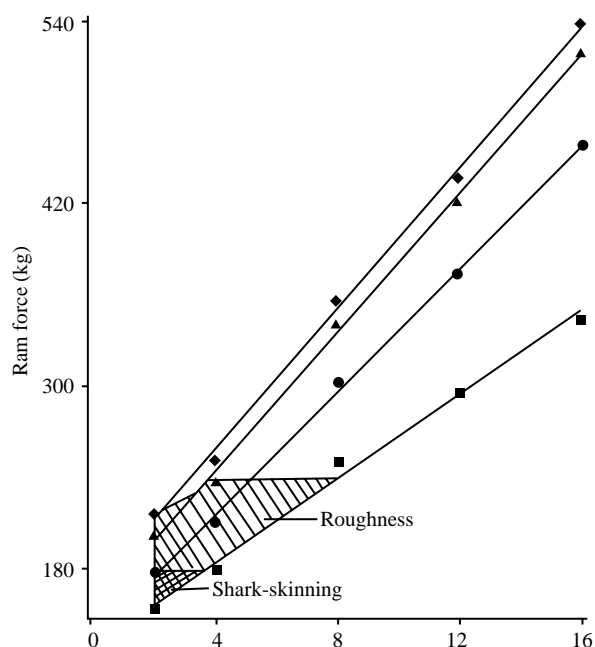


Fig. 5 A graph of ram force as a function of length-to-radius ratio, depicting conditions under which surface defects occur when extruding microcrystalline cellulose–lactose–water (5:5:6). Die diameter = 1.5 mm; ram speed cm/min: ■, 5; ●, 10; ▲, 20; ◆, 40.

FORMULATION

Extrusion mixtures are formulated to produce a cohesive plastic mass that remains homogeneous during extrusion. The mass must possess inherent fluidity, permitting flow during the process and self-lubricating properties as it passes through the die. The resultant extrudate must remain nonadhesive to itself and retain a degree of rigidity so that the shape imposed by the die is retained. Precise formulation requirements depend upon subsequent processing. Extrudate that is simply to be cut to short lengths to form cylindrical granules that are dried in a fluid-bed drier can be less rigid than extrudate intended for complex processing such as spheronization, where the extrudate undergoes a series of subtle shape changes.

The requirements for spheronization of the cylindrical extrudate are as follows:

1. The extrudate must possess sufficient mechanical strength when wet, yet it must be brittle enough to be broken down to short lengths in the spheronizer, but not to be so friable that it disintegrates completely. To achieve a narrow size distribution of spheres, the extrudate is ideally reduced to cylindrical rods of uniform length equal to approximately one and a half times their diameter (22).
2. The extrudate must be sufficiently plastic to enable the cylindrical rods to be rolled into spheres by the action of the friction plate in the spheronizer.
3. The extrudate must be nonadhesive to itself in order that each spherical granule remains discrete throughout the process.

A typical extrusion mixture might contain the following ingredients:

Drug	50–90%
Extrusion aid	
Microcrystalline cellulose, bentonite	5–90%
Binder	
Polyvinylpyrrolidone (PVP)	
Sodium carboxymethylcellulose (SCMC)	
Hydroxypropyl methylcellulose (HPMC)	
Fluid	
Water or solvent	

Extrusion offers the advantage of incorporating a relatively high proportion of active ingredient, up to 90%, in the final product. However, the physicochemical properties of the drug determine to a large extent the maximum quantity that can be included in a particular formulation. An extrusion aid is essential; microcrystalline cellulose is commonly used (23). The function of

microcrystalline cellulose is two-fold: it controls the movement of water through the wet powder mass during extrusion, and modifies the rheological properties of the other ingredients in the mixture, conferring a degree of plasticity which allows it to be readily extruded. This interaction with the liquid phase is both a physical and chemical phenomenon. The microscopic structure of microcrystalline cellulose is a random aggregation of filamentous microcrystals that create a high internal porosity and a large surface area, approximately 130–270 m²/g (24). This provides highly absorbent and moisture-retaining characteristics that are often unaffected by the extrusion process. This could be the essential quality that makes microcrystalline cellulose a unique material for extrusion. Bentonite and kaolin also have been used. Inclusion of 5–10% can significantly improve the extrusion properties of mixtures containing high proportions of drug. Recent work (25) has shown that it is possible to reduce the quantity of microcrystalline cellulose by adding glyceryl monostearate.

Additional ingredients may or may not be necessary. A binder increases plasticity and reduces extrudate friability, particularly when the content of microcrystalline cellulose is low. Natural or synthetic polymers, such as gelatin, PVP, or SCMC, may be incorporated into the mixture as a solid during dry mixing or in solution in the liquid phase. Commercial preparations of microcrystalline cellulose that are already combined with polymers are available. Examples include Avicel RC and Avicel CL grades of microcrystalline cellulose combined with SCMC (FMC Corporation). Variations in the type of microcrystalline cellulose significantly change the rheological properties of the mixture, and therefore, the extrusion characteristics (20). The differences between shear stress–shear rate flow curves of mixtures of microcrystalline cellulose–lactose–water (5:5:6) containing different particle sizes of microcrystalline cellulose and different quantities of SCMC are distinct but different. Inclusion of a polymer in the wet mass produces marked rheological differences. This has implications in the choice of formulations, since the extrudates formed from these various microcrystalline cellulose mixtures behave differently during subsequent processing, such as cutting, spheronization, and drying.

The mixture of dry ingredients is blended with water or a solvent such as ethanol (26) to form a dense cohesive mass suitable for extrusion. The liquid content of the wet powder mass and its distribution are highly critical and should be controlled so that they produce an extrudate that possesses the ideal characteristics. In general, these wet mixes have a much higher moisture content, typically 20–30 wt%, than is required for conventional (tablet)

granulations, the aim being to produce as dense a material as possible for passing through the extruder. Fluffy and incompletely wetted masses feed poorly and cause problems by creating excessive pressure and friction within the equipment. On spheronizing, they tend to produce large quantities of fines, and the “dry” extrudate is insufficiently plastic, forming dumbbell-shaped or ovoid pellets which never round off into spheres. On the other hand, if the mixture is too wet, it produces an extrudate that adheres to the spheronizer plate and to itself. This product tends to aggregate uncontrollably or at best produce spheres of wide-size distribution as the material is transferred from pellet to pellet via the plate motion.

The possible processability of different drugs by this approach has not yet been fully established. It is not possible to relate the pK_a , and freezing point depression, or to relate the ability to produce uniform pellets from a spheronization grade of microcrystalline cellulose (27). However, a relationship between the water solubility and the water level required by a formulation for equal parts of a series of model drugs and microcrystalline cellulose has been established (28).

INDUSTRIAL APPLICATIONS

Plastics

Extrusion technology is extensively applied in the plastics and rubber industries where it is one of the most important fabrication processes. Examples of products made from extruded polymers include pipes, hoses, insulated wires and cables, plastic and rubber sheeting, and polystyrene tiles. The most common extruder employed is the single-screw type (Fig. 6) with either cold or hot feed, which requires the polymer to be heated prior to processing. The extruder consists of a rotating screw inside a stationary cylindrical barrel. The barrel is often manufactured in sections that are bolted or clamped together. Usually, the inner surface of the barrel is grooved to reduce slippage and increase pumping capability. An end-plate die, connected to the end of the barrel, determines the configuration of the extruded product.

The extruder is conventionally divided into three sections: feed zone, transition zone, and metering zone. Resin granules are fed from a hopper directly into the feed section, which has deeper flights or flights of greater pitch. This geometry enables the feed material to fall easily into the screw for conveying along the barrel. The pellets are transported as a solid plug to the transition zone where they are mixed, compressed, melted, and

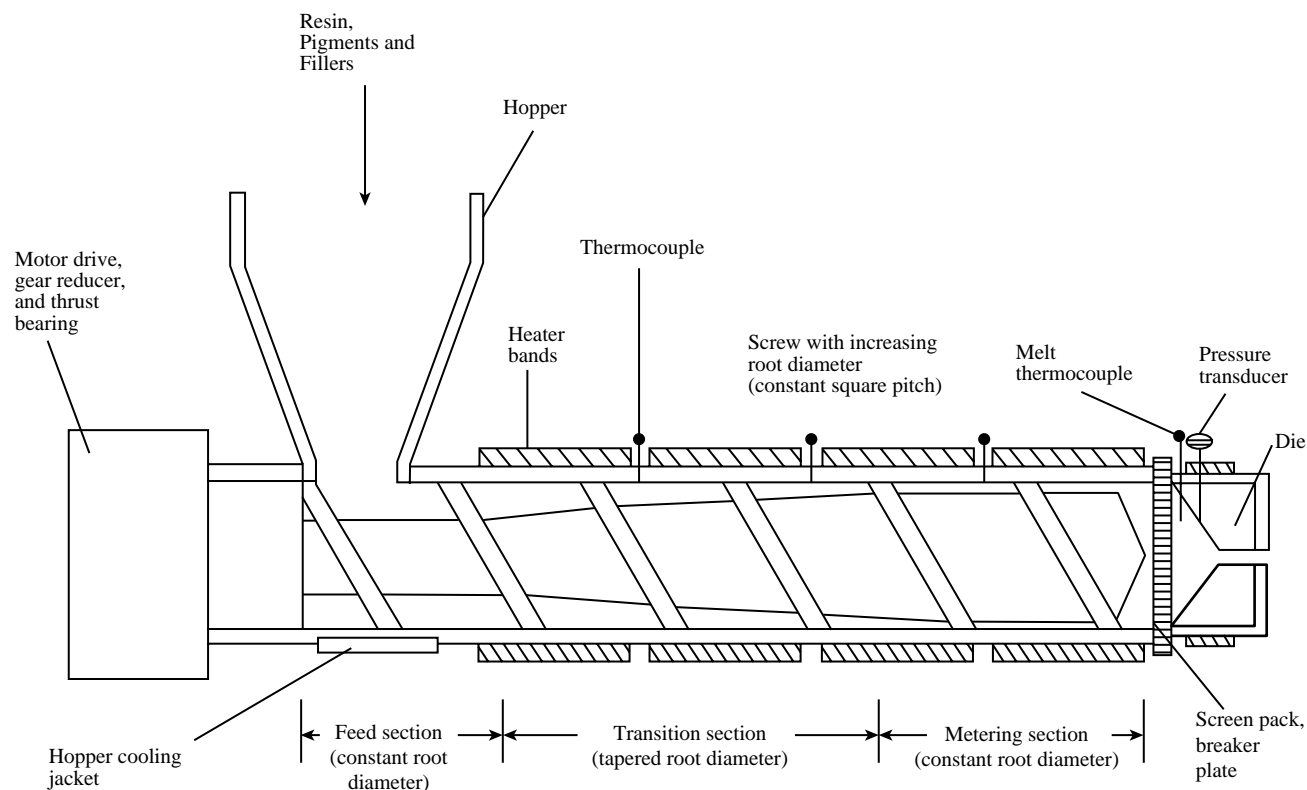


Fig. 6 Component parts of a single-screw extruder.

plasticized. Compression is developed by decreasing the thread pitch but maintaining a constant flight depth or by decreasing flight depth while maintaining a constant thread pitch (29). Both methods result in increased pressure as the material moves along the barrel. Most of the heat required to melt the material is supplied by the heat generated by friction as the resin granules are sheared between the rotating screw and the wall of the barrel. Additional heat may be supplied by electric heaters mounted on the barrel. The melt moves by circulation in a helical path by means of transverse flow, drag flow, pressure flow, and leakage: the latter two mechanisms reverse the flow of material along the barrel. The material reaches the metering zone in the form of a homogeneous plastic melt suitable for extrusion. For an extrudate of uniform thickness, flow must be consistent and without stagnant zones right up to the die entrance. The function of the metering zone is to reduce pulsating flow and ensure a uniform delivery rate through the die cavity. Some applications require a strainer plate fitted between the extruder and die plate to remove solid impurities or lumps of incompletely melted resin.

Polymers with a wide range of viscoelastic and melt viscosities cannot be processed with a single screw. Most

commercial extruders are, therefore, modular in design, providing a choice of screws or interchangeable sections that alter the configuration of the feed, transition, and metering zones. This makes it possible to modify the process to meet particular requirements, for example, from a standard to a high shear or high output extrusion. Modified screw designs allow the extruder to perform a mixing role in addition to extrusion, so that the material can be colored and blended. The various screw and die designs available and practical considerations of thermoplastic extrusion are reviewed by Whelan and Dunning (30). Extrusion processing requires close monitoring of the various parameters that affect polymer extrusion: viscosity, variation of viscosity with shear rate and temperature, elasticity, extensional flow, and slippage of the material over hot metal surfaces. Equations used to describe flow are included in the section on the "Theory and Characterization of Extrusion" presented earlier. Recent advances in the design and operation of extruders allow in-process monitoring and control of parameters, such as the temperature in the extruder, head, and die; pressures in extruder and die; wall thickness and other dimensions; "haul-off" speed and extrusion speed; and power consumption.

The process described above is known as profile or line extrusion in which the shape of the extrudate is determined by the die. The extruded profile proceeds horizontally to the cutoff equipment, which controls its length. It is then cooled to a solid state, usually by spraying with or immersion in water, and passed through a haul-off unit. Finally, it is cut to the required length or coiled. The downstream auxiliaries (e.g., such as haul-off equipment for handling the extrudate stream, collection machinery for winding or coiling continuous lengths of tubing or profiles, cropping and cooling equipment, and systems for monitoring the diameter and wall thicknesses of pipes on-line) are as important as the extruder itself (30, 31). Tubes and pipes and other solid cross-sections are mainly produced by profile extrusion. Profiles may be further processed, for example, as in film extrusion, blow molding, or injection molding (32).

Film extrusion

The polymer melt is extruded through a long slit die onto highly polished cooled rolls that form and wind the finished sheet. This is known as cast film. Plastic packaging film is also formed by blow extrusion, where tubular film is produced by extruding the melt, usually vertically, through an annular-shaped slit die. The extruded tube is inflated by air to form a large cylinder. The bubble is cooled externally by an airstream directed onto its surface and is collapsed on passing between a pair of rollers before being wound up. Film made by the casting process generally has better optical properties than blown film, but is less strong mechanically. Cast films usually require edge trimming at additional cost.

Blow molding

The plastic is heated to a melted or viscous state and a section of molten polymer tubing (parison) is extruded usually downward from the die head into an open mold. The mold is closed around the parison, sealing it at one end. Compressed air is blown into the open end of the tube, expanding the viscous plastic to the walls of the cavity, thus forming the desired shape of the container. The material cools in the cavity and solidifies. The mold is opened and the molding is removed. This technique is used for the manufacture of bottles, toys, and large containers.

Injection molding

The molten plastic is extruded into a cavity mold at high pressure. The material cools in the cavity and solidifies. The mold is then opened and the article is removed. Very intricate configurations can be obtained by this technique

(e.g., to provide intricate and strong components for the electronic, telecommunications, and clock-making industries).

Plastics that are commonly processed by extrusion include acrylics (polymethacrylates, polyacrylates) and copolymers of acrylonitrile; cellulose (cellulose acetate, propionate, and acetate butyrate); polyethylene (low and high density); polypropylene; polystyrene; vinyl plastics; polycarbonates; and nylons. The material properties and extrusion properties have been reviewed by Whelan and Dunning (30). Additives that may be included to modify or enhance properties (33) include lubricants and antislip agents to assist processing during extrusion; plasticizers to achieve softness and flexibility; stabilizers and antioxidants to retard or prevent degradation; and dyes and pigments.

Food

In principle, any food that can be formed into a paste can be processed by an extruder. Food extrusion has been utilized since the 1930s for pasta production. Modern equipment and processing techniques allow the manufacture of complex products in a variety of shapes and sizes. Raw materials such as cereals, oil seed, and protein, along with carbohydrates and water mixtures, can be converted into products such as meat substitutes, pet foods, and snack meals. A widely used and versatile technique combines cooking and extrusion in a so-called extrusion cooker (34). It has the potential to manufacture a range of novelty or specialty products, such as breakfast cereals (expanded and shaped cereals), shaped and filled snacks, protein-fortified and precooked pasta products, and precooked meat pieces for convenience foods. The process is highly economical, and provides mixing, high temperature–short duration cooking, texturizing, and shaping of the food in one step. The equipment closely resembles the screw extruders used in the processing of thermoplastics. The screw is designed to create varying zones along the barrel, allowing the food substance to be processed in stages. The solid and liquid starting materials are fed from a hopper to the feed zone of the extruder and conveyed to the transition zone. Here the materials may be compressed, mixed, sheared, and heated to form a viscous plastic dough. In the metering zone, the plastic mass is subjected to further heating and shearing before being pumped into the die to form the shaped product. The pressure drop on leaving the die causes superheated water to flash off the molten material. If the dough contains starch, gelatinization will result in an expanded porous product with a crunchy texture (35). Finally, the product may be cut, shaped by passing through rollers, dried, and packed.

The viscosity of the dough may vary more than an order of magnitude during the extrusion cooking process as a result of changes in shear rate, temperature, moisture content, and induced physicochemical changes such as protein denaturation, polysaccharide gel formation, and reorientation of molecules (36). For this reason, success in food extrusion requires accurate monitoring and control of feed rate, screw speed, temperature, and moisture to produce and control desired product characteristics. Knowledge of the viscous rheology of the food mixture in the metering section immediately prior to extrusion is of particular importance. However, this is not easily predicted since, unlike the case of homogeneous or simple mixtures of polymers where the major change is melting, food doughs are of such complexity that the exact chemical composition and structure cannot readily be determined. Efforts have been made to develop semiempirical models derived from plastics extrusion to describe the apparent viscosity of cooking doughs, which may be useful in evaluating food formulations (37, 38). Remsen and Clarke (36) used an Instron capillary viscometer and amylograph to describe the relationship between the viscosity of a typical soy flour dough and the applied shear rate, temperature, and time-temperature history. Fletcher et al. (39) investigated the viscous dough rheology of maize mixtures as a function of the extrusion variables (pressure, shear, and temperature). They used an instrumented single-screw extruder fitted with slit dies, and related the results to the product properties. The advantage with this method is that the food material receives a deformation history corresponding to the extrusion cooking process, which is otherwise difficult to replicate in a laboratory rheometer.

Animal Feed Production

In the animal feed industry, extrusion is applied as a means of producing pelletized feeds, commonly in the form of short cylindrical rods of 4–8 mm in diameter. Pellets are a convenient means of precisely controlling the animal's diet. They offer several advantages (40). The quantity of feed the animal receives is better controlled by pellets than a loose-mix feed, and a complex diet of controlled composition is easily produced. The pellet feed can contain as many as 30 single ingredients mixed in the correct proportions. The animal is obliged to chew pellet feed with improved palatability and therefore, digestion. During extrusion, the feed mixture is compressed, resulting in a densified product that requires less storage space.

Pellets are prepared from a mixture of raw materials of varying chemical composition (starch, oil, fiber, and

moisture) and physical characteristics (particle size, bulk density, and moisture-retention properties). The composition of a typical poultry feed is often complex (41). The raw material properties determine the quality of pellet formation. Equipment performance and pellet quality (friability, size uniformity) can be improved by a small amount of extrusion or pelleting aid (binders, lubricants) (42). Additives commonly used in the feed industry include molasses with binding properties when activated by steam; fatty acid lubricants to reduce product-metal friction when extruding or pressing through long dies; lignosulfates (organic materials derived from lignin in trees that improve pellet quality and throughput rates); and mineral binders, such as ball clay and bentonite, or cellulose binders, such as sodium carboxymethylcellulose. It should be noted that in small quantities cellulose binders can improve the pelleting process and reduce pellet friability.

The pelleting process (40) consists of blending and conditioning the feed mixture immediately prior to pressing, pressing itself, cutting of the pellets, and cooling. The complexity of the feed mixture, composed of a number of ingredients of different particle sizes and densities in varying proportions, requires thorough blending to ensure homogeneity. The product is conditioned by adding moisture, typically up to 15%, and heating to a controlled temperature in order to gelatinize the starch or convert it to simple sugars. This reaction causes the starch to act as a binder and converts the meal into a physical state suitable for pressing. The most efficient means of conditioning and heating is by steam. Optimal conditioning parameters, moisture content of the material, temperature, and duration of heating depend on the composition of the mixture (42). For example, high starch-low fiber meals require temperatures of 80–85°C, whereas feeds that contain heat-sensitive ingredients, such as milk and sugar, have a temperature limit of 55°C (39).

Extrusion Pressing

According to Sebestyen (40), pellet mills may be classified into disk-die presses or ring-die pellet mills. In the former, the die consists of a circular plate resting in the horizontal plane into which holes are drilled in a regular pattern. A set of rollers move around the upper surface of the disk, sweeping the meal in their path through the holes and compressing it to form pellets or cubes. Rotating adjustable knives located beneath the disk cut the extrudate to an appropriate length. In another design, the plate revolves while the rollers and knives remain fixed.

Ring-type pellet mills have a radially arranged die resting in the horizontal plane with rollers rotating and revolving along the inner surface. The rolls are offset from the die face, leaving a slight clearance that allows buildup of a thin product layer, optimizing throughput efficiency. The peripheral velocity of the rollers depends upon the die diameter; that is, higher speeds are required for smaller-diameter holes and lower speeds for larger-diameter holes.

Cooling

On leaving the extruder, the warm pellets are pliable and prone to abrasion and deformation. Therefore, a final processing stage is required to harden the pellets (40). Cooling equipment placed directly beneath the mill employs ambient or chilled air to reduce the temperature and remove excess moisture from the final product.

PHARMACEUTICAL INDUSTRY

Extrusion processes are applied within the pharmaceutical industry to produce a variety of dosage forms such as suppositories, implants, and granulations.

The large-scale manufacture of suppositories and pessaries uses either the fusion method where the drug is dispersed in a molten base and the mixture poured into molds to solidify, or the cold compression method (43, 44). In the latter process, the medicament and cold-grated base, usually theobroma oil or witepsol base, are intimately mixed and placed in a cylinder. The mass is extruded by means of a piston through small holes that connect with the mold. The cavities are filled by pressure with the mass which is prevented from escaping by movable end plates. The plates are removed and the suppositories ejected by further extrusion. The extrusion equipment is chilled to prevent melting of the

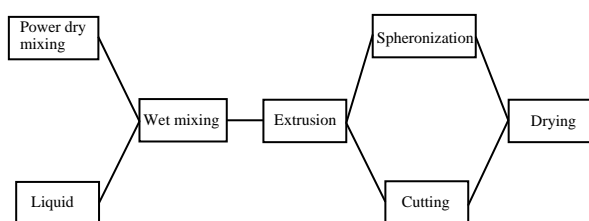


Fig. 7 Schematic of extrusion processing in the pharmaceutical industry.

components due to the heat generated by the friction of compression.

The most important application of extrusion in the pharmaceutical industry is in the preparation of granules or pellets of uniform size, shape, and density that contain one or more drugs. The process involves a preliminary stage in which dry powders, drug, and excipients are mixed by conventional blenders, followed by addition of a liquid phase and further mixing to ensure homogeneous distribution (Fig. 7). The wet powder mass is extruded through cylindrical dies or perforated screens with circular holes, typically 0.5–2.0 mm in diameter, to form cylindrical extrudates. These may be further processed, for example, by cutting and drying to yield granules, or by spheronization (24) to yield spherical granules followed by drying. The spheroids are usually coated with a polymer to control the rate of drug release and filled into hard gelatin capsules to yield a multiple-unit dosage form.

Extruders Used for Pharmaceuticals

Commercial extruders may be classified according to the die design and the feed mechanism that transports the material to the die region.

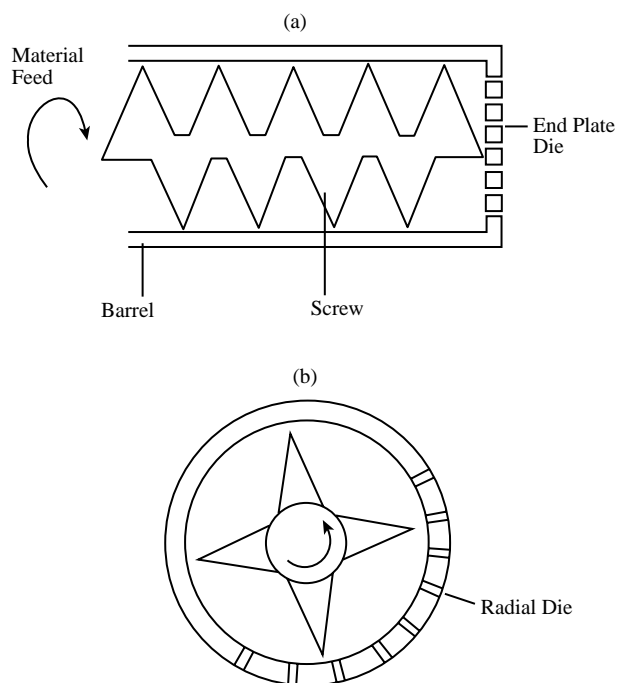


Fig. 8 Screw extruder with (a) end-plate die and (b) radial-screen die.

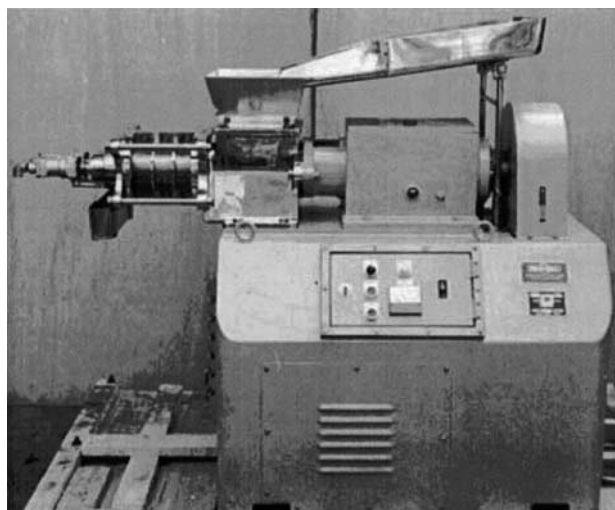
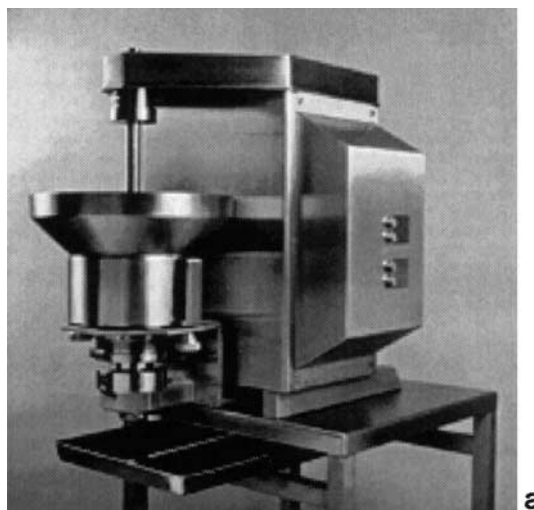


Fig. 9 Twin-screw extruder with radial-screen die. (Manufactured by Fuji Paudal Company, Japan.)

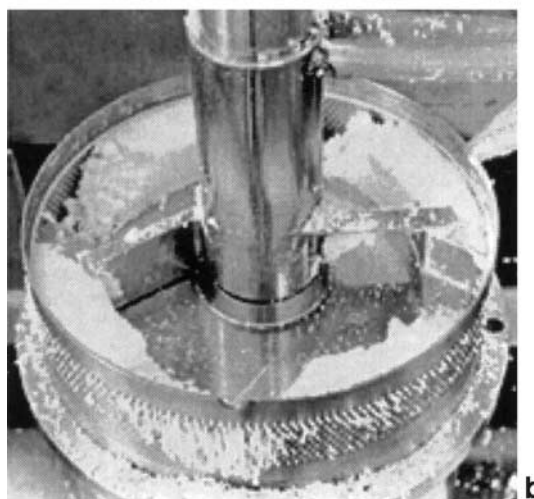
Screen extruders

Screen extruders utilize a screw-feed mechanism consisting of single or twin helical screws rotating in a barrel to convey the damp mass from a feed hopper to the die zone. The die consists of a thin steel plate perforated with numerous holes, which is positioned radially or axially to the screw feed (Fig. 8). The advantages of this arrangement are high continuous throughput rates, from 5 kg/h of wet mass for a laboratory-scale single-screw extruder, up to 800 kg/h for a larger twin-screw design. The screens are easily cleaned and interchanged; they have holes of varying diameter beginning at 0.5 mm and are available commercially. The disadvantage of this type of equipment, however, is that the screw mechanism can exert a high pressure on the material, generating excessive friction and heat as the wet mass passes between the screw and barrel. This is particularly the case with axially orientated dies. These extruders tend to have a high dead volume that contains stagnant material between the feed screws and the screen. Consideration should be given to this if the wet powder mass contains ingredients that are unstable when wetted with water. The low L/R of the die holes can also result in low compaction in the extrudate and distortion of the surface finish, known as shark-skinning. This problem can sometimes be overcome by varying the throughput rate, which will be discussed later.

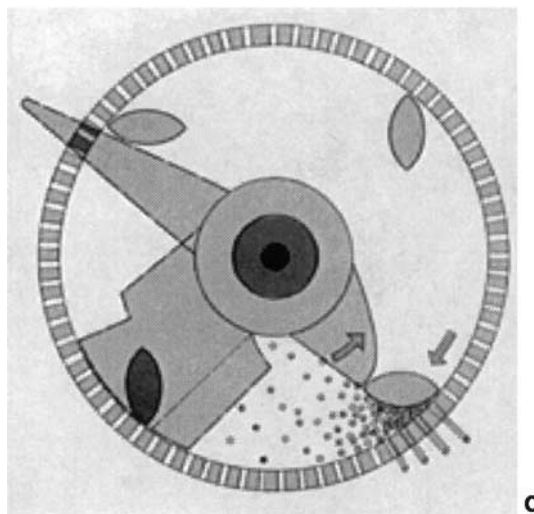
The twin-screw design and radial-die screen assembly of an extruder manufactured by the Fuji Paudal Company of Japan is shown in Fig. 9. Water can be circulated



a



b



c

Fig. 10 The NICA system radial-screen extruder. (a) Assembled unit. (b) Dismantled to show extrusion mechanism. (c) Cross section indicating working principle.

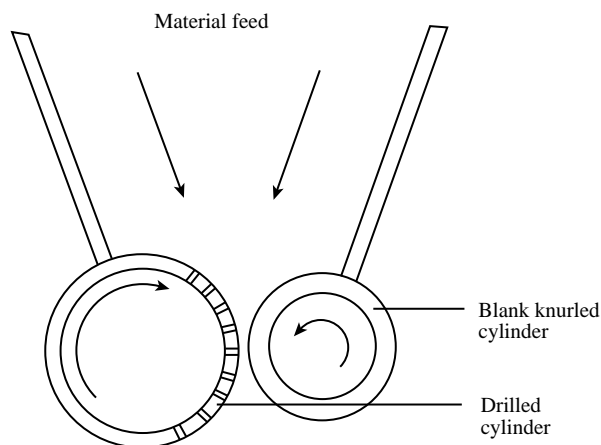


Fig. 11 The rotary-cylinder-type extruder.

through the hollow extrusion rotors to maintain a constant temperature in the extrusion zone. This is a useful facility when processing heat-sensitive materials and for controlling temperature, extrudate moisture levels, and viscosity. Interchangeable screens are available with die holes ranging from 0.5 to 1.5 mm in diameter, allowing the production of a wide range of extrudates. Explosion-proof motors, with fixed or variable speed drive, are fitted for safe processing of wet masses granulated with inflammable solvents. All components in contact with process materials are constructed of high-grade stainless steel. An additional feature is the option of fitting an axial die plate in cases where a denser extrudate is required.

A screen extruder that operates with a novel mechanism is the Nica System Extruder (Fig. 10), manufactured in Sweden. It consists of a radial screen encircling an extrusion rotor and a rotating disk fitted with angled impeller blades or baffles. Above this is a counterrotating central feed blade. Speeds of both the extrusion rotor and the feed blade are variable. Material, such as gravity fed from a hopper, is swept into the blades and pressed through the holes in the screen. According to the manufacturer, there are several advantages to this equipment. First, pressure is exerted on only a small quantity of mass and only at the point of extrusion; that is, just between the baffles and the screen. Second, temperature increase is minimal, and a moisture gradient between the wet mass and extrudate is avoided. Because of this, cooling facilities are not necessary. The dead volume, located in front of each baffle, is limited and may be as low as 15 g per baffle. A small extruder with output up to 4 kg/min is available for development and small-scale production.

For larger production, an extruder with output of up to 12 kg/min is available.

Rotary-cylinder extruder

The working principle of this machine is based on two counterrotating cylinders (Fig. 11). The granulating cylinder is perforated and acts as the die. The diameter and the L/R of the holes can be varied. The holes are spaced further apart and are drilled rather than of punched sheet construction as in the screen-type extruders. The other cylinder is solid and acts as a pressure cylinder. Material is gravity fed from a hopper to the die region between the cylinders and adheres to the knurled surface of the solid cylinder, building up a thin layer that is pressed through the die cylinder. Although the extrusion is a continuous process, actual material flow through each hole is intermittent due to the rotation of the die. Pressure is built up in the perforations, which compacts the wet mass and forces the extrudate to the interior of the cylinder. This pressure is dependent upon the diameter and the length of the perforations. Hence, with die holes of high L/R , this system can achieve good densification of the wet powder mass. This is important in giving the most granules mechanical strength and stability for further processing. Another advantage is the lack of a dead zone, which is limited to the thin layer of material adhering to the pressure cylinder. Since the cylinders only apply pressure to a small quantity of material in the feed zone, there is little tendency for creating moisture gradients. However, cleaning of the granulating cylinder can be troublesome. The material remaining in the die holes can be difficult to dislodge, particularly when the L/R is high. Furthermore, the granulating cylinders are expensive because of the high costs of drilling stainless steel.

A cylinder extruder manufactured by Alexanderwerk (Germany) is shown in Fig. 12. With this equipment, the feed stock material can be metered to the working area. On the smallest machines this is accomplished by a rotary-table feed hopper. On larger machines, the feed rate is controlled by screw feeders sited through the feed hopper. The throughput rate depends on the diameter and L/R of the die holes, as well as on the feed rate. Laboratory-scale extruders with a throughput range of 30 to 50 kg/h use granulation cylinders 70 mm in diameter. Production-scale equipment with a larger granulating cylinder (186 mm diameter) can achieve an output of 100–105 kg/h. Interchangeable cylinders with die holes of 1.0–5.0 mm are available. A multiple-unit assembly consisting of three parallel

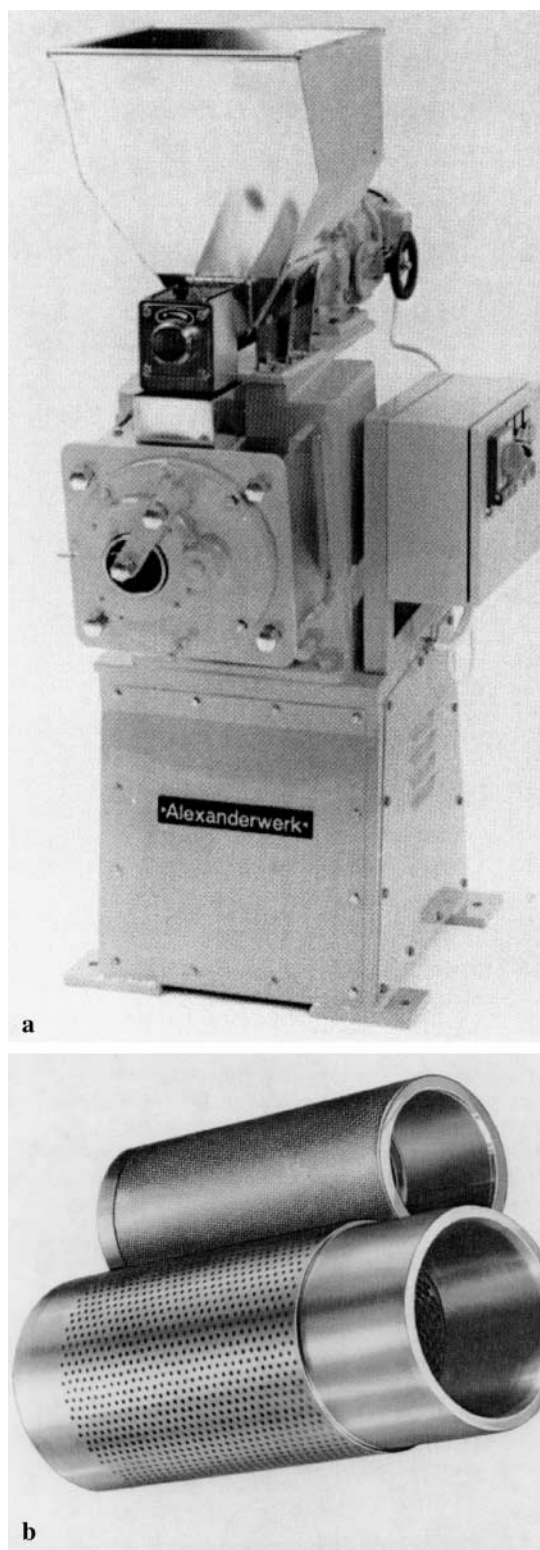


Fig. 12 The cylinder extruder, manufactured by Alexanderwerk, Germany. (a) Laboratory-scale extruder. (b) Die cylinder and pressure cylinder.

extrusion heads can achieve even higher throughput rates of up to 3000 kg/h.

When scaling up production, the die cylinder dimensions cannot be increased proportionately, since the wall thickness and therefore, die L/R become too high. This results in excessive extrusion forces imparted on the material. This is overcome by using special counter-bored die cylinders with reduced depth perforations to provide optimal extrusion conditions at scale-up. The temperature increase in the extrudate is minimized by circulating cool water through the pressure cylinders. A scraper blade attachment inside the perforated cylinder cuts off the extrudate to shorter lengths, making it more manageable.

Rotary-gear extruder

The rotary-gear extruder operates on a similar concept to that of the cylinder extruder. It consists of two hollow counterrotating gear cylinders with counterbored dies, described by the manufacturer (Bepex, Berwind Corporation) as nozzles, that are drilled into the cylinders between the teeth (Fig. 13). The material, gravity fed from a hopper, is drawn in by the toothed cylinders and pushed through the nozzles into the center of the cylinders, where scrapers cut off the extrudate. The product is compacted as it passes through the nozzles, and thereby forms a dense extrudate. The density depends on the nozzle L/D (the ratio of nozzle length to nozzle diameter). Higher throughput rates can be attained with this type of extruder, since output is achieved through both rotating-gear wheels. The equipment and the gear-toothed cylinders are shown in Fig. 13b.

Interchangeable gear cylinders are available with variable nozzle L/D ratios by counterboring from the inside of the rollers to reduce the die length or by using replaceable nozzle inserts to increase the die length. The diameter of the holes can be varied from 1.0 to 10.0 mm to produce a range of pellet sizes. The throughput rate can be controlled by varying the cylinders' rotation speed and the corresponding feed rate. Throughput capacity ranges from 20 kg/h for the small-scale laboratory extruders to approximately 1000 kg/h for production equipment. For large equipment or materials with poor flow characteristics, agitators can be installed in the hopper to prevent bridging; special hoppers with conical feed screws fitted with additional wipers are used for highly viscous products. The machine can be furnished with cooling equipment, circulating water through the compaction gears for processing materials that need temperature control. An alternative pharmaceutical gear extruder, similar in design to the above, has recently been marketed by G.B. Caleva Ltd., United Kingdom.

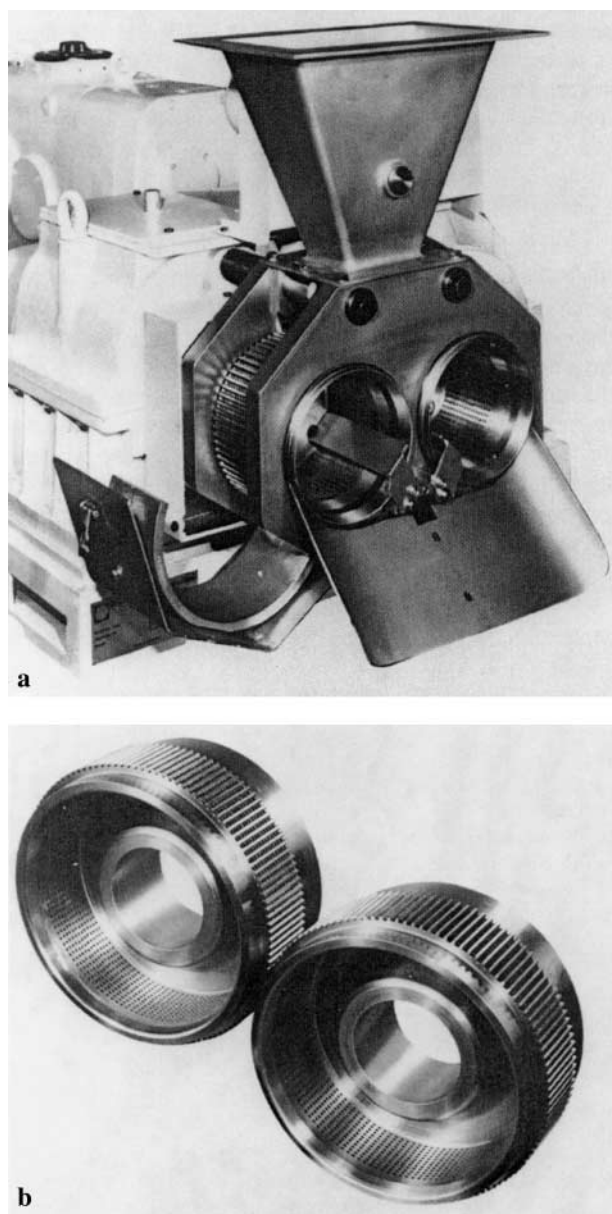


Fig. 13 (a) Rotary-gear extruder. (b) Gear-toothed cylinders.

Ram extruder

Industrial ram extruders are commonly used in the plastics and rubber industries for the preparation of warm strip feed for large cold-fed screw extruders and for forming strips or slugs for feeding injection molding and compression molding machines. They are used in the extrusion of specialized substances that require critical in-process control or that are not readily amenable to processing by screw extruders. Examples include the extrusion of waxlike substances such as coloring crayons,

dental waxes, and rocket propellant, and in the extrusion of moist powders and claylike materials, such as blackboard chalks. Ram extrusion allows control of parameters, such as temperature, size, and weight of extrudate. An example of a high performance ram extruder, as manufactured by Borwell International Ltd., United Kingdom, is shown in Fig. 14. It consists of a chrome-plated barrel positioned in a thermostatically controlled storage tunnel. A range of dies can be fitted to the extruder head. Material is loaded into the barrel by manual or mechanical means and vacuum is applied to eliminate air from the system. The material is extruded by means of an hydraulically powered ram, with the hydraulic (oil) fluid being passed through a special valve system to sense changes in the plasticity of the material and compensate ram pressure to achieve an even extrusion through the hole. A multispeed cutter mounted on a fly wheel severs the extrudate at the die face. The volume of the extrudate is a function of the cut speed and the set ram speed, and can be controlled to a high degree of accuracy within $\pm 1\%$. A continuous operation is possible with the help of a twin barrel and ram arrangement in which material is fed to each barrel in turn by a screw system. Various sizes of extruder are available, from 4.5-L barrel capacity up to 80 L, offering production rates up to a maximum of 800 kg/h.

Choice of extruder

The selection of the extruder design is based on the principal requirements of the extrudate and the nature of further processing. For the production of uniform granules to be dried in a fluid-bed drier, a low-compaction system, such as that provided by the various types of screen extruders may be suitable. Cylinder or gear-type extruders may be more appropriate when aiming for a densified extrudate, such as that required for spheronization. Ram-extrusion systems, which allow precision control of extrudate density, size, and shape, are ideal for the extrusion and forming of pharmaceutical polymers of the type used for subdermal implants.

Consideration should be given to the availability of small-scale equipment, which is vital for development work prior to scale-up on pilot- or production-scale machines. Equipment choice is not necessarily based on maximum throughput rate, since the subsequent processing stages (e.g., cutting, spheronization, and drying) are batch processes and are therefore, a rate-limiting factor in production. Since extrusion is a continuous process, it allows adequate production rates for most purposes with any of the above mentioned extruder types.

The equipment must comply with the code of Good Manufacturing Practice (GMP) standards within the

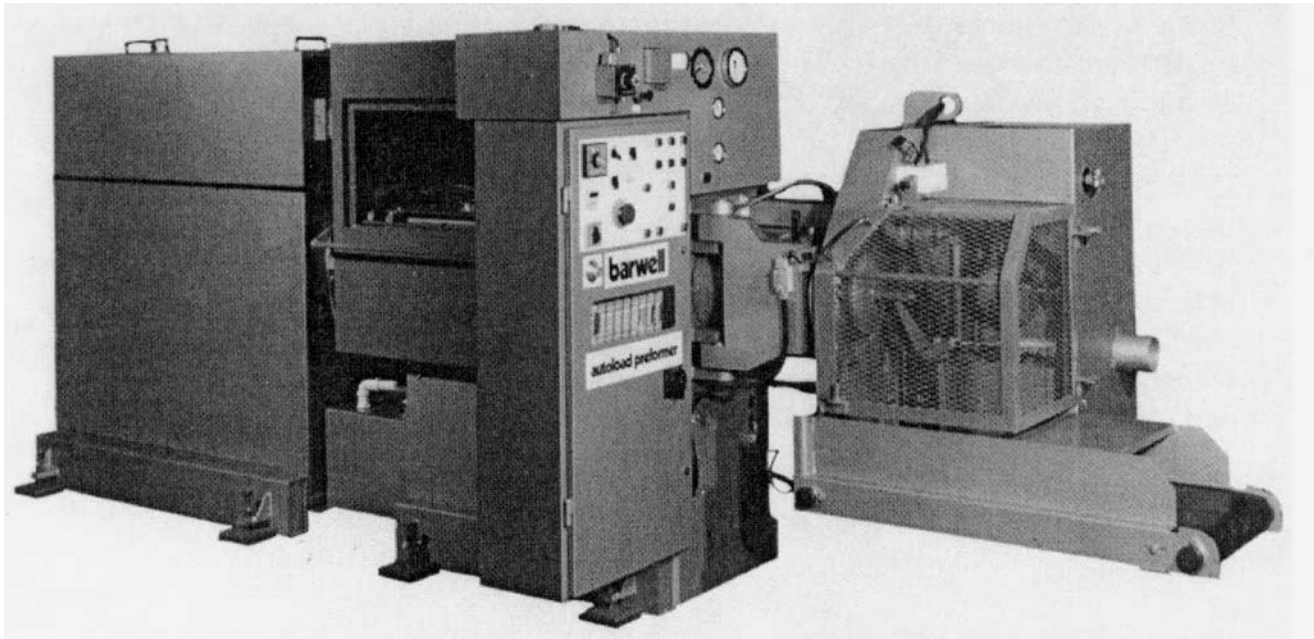


Fig. 14 The Barwell ram extruder.

pharmaceutical industry. Machines should be constructed of durable material with smooth surfaces to discourage adhesion of extraneous material and facilitate cleaning. Materials used for equipment construction must not affect the product. They should be corrosion resistant and able to withstand cleaning disinfectants. All surfaces and parts in contact with the product must therefore be of high-grade stainless steel and designed to exclude contamination of process material or product by lubricant during manufacture. Controls should be accessible by means of recessed contact buttons.

NOMENCLATURE

A_0	initial cross-sectional area	n_b	Bagely entrance correction
A	die cross-sectional area	P	pressure drop along die
C	compliance	P_e	die exit pressure
D	die diameter	P_0	upstream pressure loss
D_0	barrel diameter	P_{0w}	pressure drop at zero velocity
$(dv/dr)_w$	rate of shear at the die wall	P_T	total pressure drop
ESR	tensile stretch rate	P_{t_v}	pressure drop at velocity v
EV	apparent elongational viscosity	Q	volumetric flow rate of extrudate
K	power law viscosity constant	R	radius of capillary (die)
L	length of capillary (die)	RS	recoverable shear
n	degree of non-Newtonian flow (power law index)	T	tensile stress
		U	plastic viscosity
		V	extrudate velocity
		x	thickness of Newton liquid layer
		α	die entry yield stress velocity factor
		β	die and velocity factor
		γ	rate of shear
		η	apparent viscosity
		σ_y	yield value
		σ_{y0}	yield value at zero velocity
		θ	half angle of convergence
		τ	shear stress
		τ_0	die wall shear stress at zero velocity
		τ_w	die wall shear stress
		τ_y	shear stress to be exceeded before flow commences

REFERENCES

1. Bagley, E.B. *J. Appl. Phys.* **1957**, 28, 624–627.
2. Han, C.D.; Charles, M. *Trans. Soc. Rheol.* **1971**, 15, 371–384.
3. Harrison, P.J. *Extrusion of Wet Powder Masses*; University of London: 1982.
4. Jastrzebski, Z.D. I & E C Fundamentals, **1967**, 6(3), 445–454.
5. Wilkinson, W. *Non-Newtonian Fluids*; Pergamon Press: London, 1960.
6. Harrison, P.J.; Newton, J.M.; Rowe, R.C. *J. Pharm. Pharmacol.* **1984**, 36, 796–798.
7. Benbow, J.J.; Ovenston, A. *Trans. Br. Cer. Soc.* **1968**, 67, 543–567.
8. Benbow, J.J. *Chem. Eng. Sci.* **1971**, 26, 1467–1473.
9. Benbow, J.J.; Oxley, E.W.; Bridgwater, J. *Chem. Eng. Sci.* **1987**, 42, 2151–2162.
10. Benbow, J.J.; Bridgwater, J. *Chem. Eng. Sci.* **1987**, 42, 915–919.
11. Cohan, R.K.; Newton, J.M. *Int. J. Pharm.* **1996**, 131, 201–207.
12. Cogswell, F. *Polym. Eng. Sci.* **1972**, 12, 64–72.
13. Bealy, J.M.; Wisbren, K. *Rheology of Molten Plastics Processing*; Van Norstrand: New York, 1990.
14. Chohan, R.K. *J. Appl. Polym. Sci.* **1994**, 54, 487–494.
15. Beart, L.; Remon, J.P.; Knight, P.; Newton, J.M. *Int. J. Pharm.* **1992**, 86, 187–192.
16. Tomer, G.; Newton, J.M. *Int. J. Pharm.* **1999**, 182, 71–77.
17. Tomer, G.; Newton, J.M.; Kinechesh, P. *Pharm. Res.* **1999**, 16, 666–671.
18. Tomer, G.; Mantle, M.D.; Gledde, L.F.; Newton, J.M. *Int. J. Pharm.* **1999**, 189, 19–28.
19. Harrison, P.J.; Newton, J.M.; Rowe, R.C. *J. Pharm. Pharmacol.* **1985**, 37, 81–83.
20. Raines, C.L.; Newton, J.M.; Rowe, R.C. *Extrusion of Microcrystalline Cellulose. Rheology of Food, Pharmaceutical and Biological Materials with General Rheology*; Carter, R.E., Ed.; Elsevier Applied Science: London, 1990; 268–287.
21. Harrison, P.J.; Newton, J.M.; Rowe, R.C. *J. Pharm. Pharmacol.* **1984**, 36, 796–798.
22. Fielden, K.E. *Extrusion and Spheronization of Microcrystalline Cellulose/Lactose Mixtures*; University of London: 1987.
23. Newton, J.M. *Extrusion/Spheronization. Powder Technology and Pharmaceutical Processing*; Ghulia, D., Delevil, M., Paercelt, Y., Eds.; Elsevier Biomedical: Amsterdam, 1994; 391–401.
24. Battiste, O.A. *Microcrystalline Polymer Science*; McGraw-Hill: London, 1975.
25. Basit, A.; Newton, J.M.; Lacey, L.F. *Pharm. Dev. Tech.* **1999**, 4, 499–505.
26. Tufvesson, C.; Lindberg, N.O.; Olbjer, L. *Drug Develop. Ind. Pharm.* **1987**, 13 (9–11), 1891–1913.
27. Jover, I.; Podczek, F.; Newton, J.M. *J. Pharm. Sci.* **1990**, 85, 700–705.
28. Lustig-Gustafsson, C.; Johal, H.K.; Podczek, F.; Newton, J.M. *Eur. J. Pharm.* **1999**, 8, 147–152.
29. Johnson, P.S. *Developments in Extrusion Science and Technology*; Polysay Ltd.: South Samia, Ontario, Canada, 1982.
30. *The Dynisco Extrusion Processors Handbook*; Whelan, T., Dunning, D., Eds.; London School of Polymer Technology, Polytechnic of North London: London, 1988.
31. Smoluk, U.R. *Modern Plastics Mt.* **1988**, 18 (12), 49–56.
32. Schott, H. *Polymer Science. Physical Pharmacy—Physical Chemical Principles in the Pharmaceutical Sciences*; Martin, A., Swarbrick, J., Cammarata, A., Eds.; Ch. 22; Lea & Febiger: Philadelphia, 1983.
33. Giles, R.L.; Pecina, R.W. *Plastic Packaging Materials. Remington's Pharmaceutical Sciences*; Gennaro, A.R., Ed.; Ch. 81; Mack Publishing Co.: Easton, PA, 1985.
34. Senouci, A.; Smith, A.; Richmond, P. *Chem. Eng.* **1985**, 417, 30–33.
35. Owusu-Ansah, J.; Van de Voort, F.R.; Stanley, D.W. *Can. Inst. Food Sci. Technol. I.* **1984**, 17 (2), 65–70.
36. Remsen, C.H.; Clark, J.P. *J. Food Process Eng.* **1978**, 2, 39–64.
37. Hermann, D.V.; Harper, J.M. *Food Sci.* **1974**, 39, 1009–1104.
38. Harper, J.M. *Crit. Rev. Food Sci. Nutr.* **1979**, 11 (2), 155–215.
39. Fletcher, S.I.; McMaster, T.J.; Richmond, P.; Smith, A.C. *Chem. Eng. Commun.* **1985**, 32, 239–262.
40. Sebestyen, E. *Flour and Animal Feed Milling* **1974**, 156 Oct, 24–25, Nov. 1974, 22–25; Dec. 1974, 16–18.
41. Young, L.R.; Pfost, H.B.; Feyerherm, A.M. *Trans. Am. Soc. Agr. Eng.* **1963**, 50, 144–151.
42. Payne, J.D. *Milling Feed Fert.* **1978**, 161, 34–41.
43. Anshel, J.; Lieberman, H.A. *Suppositories. The Theory and Practice of Industrial Pharmacy*; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Ch. 8; Lea & Febiger: Philadelphia, 1976.
44. Hadgraft, J.W. *Rectal Administration. In Bentley's Textbook of Pharmaceutics*; Rawlins, E.A., Ed.; Ch. 25; Bailliere Tindall: London, 1977.

EXPIRATION DATING

Leslie C. Hawley

Mark D. VanArendonk

Pharmacia Corporation, Kalamazoo, Michigan, U.S.A.

INTRODUCTION

Expiration dating of pharmaceuticals corresponds to the determination of a retest period for drug substances and an expiration dating period or shelf-life for drug products. The shelf-life, or expiration dating period, of a drug product is defined as the time interval that a drug product is expected to remain within an approved shelf-life specification, provided that it is stored according to label storage conditions and that it is in the original container closure system. The Expiry/Expiration Date is the actual date placed on the container/labels of a drug product designating the time during which a batch of the drug product is expected to remain within the approved shelf-life specification if stored under defined conditions and after which it must not be used. To arrive at an expiration date, it must be determined first for how long and under what conditions a pharmaceutical formulation can meet all of its quality specifications. In general, this issue is answered through stability testing that monitors chemical and physical product attributes as a function of time, temperature, and other environmental factors. It is not the intent of this article to discuss the mechanics of starting a testing program or to provide an exhaustive discussion of all the testing requirements necessary to establish an expiration dating period. Rather, this discussion focuses on changes in worldwide regulatory requirements pertaining to expiration dating since the last edition of this publication (1), the statistical treatment of stability data to determine a shelf-life for a given drug product, the impact of postapproval changes on the expiration date of a drug product, and finally, a discussion of unresolved issues in expiration dating.

REGULATORY CONSIDERATIONS

Before 1990 there was no official harmonization of stability testing requirements for countries throughout the world. In the United States, the major reference documents outlining regulations or guidelines pertaining

to expiration dating at that time were the Code of Federal Regulations (CFR) (2), the *United States Pharmacopeia* (USP) (3), and the 1987 FDA Guideline for Submitting Documentation for the Stability of Human Drug and Biologics (4). In Europe there was the European Community Guide to Good Manufacturing Practices for Medicinal Products (5), which contained minimal information pertaining to expiration dating. More extensive European guidelines at that time were found in the Note for Guidance, Stability Tests on Active Substances and Finished Products (6). In Japan, the primary reference for the determination of expiration intervals for the approval of a drug product was given in a guideline entitled, at that time, Draft Policy to Handle Stability Data Required in Applying for Approval to Manufacture (Import) Drugs (7) and Draft Guidelines on Methods to Perform Stability Test (8). Very little information pertaining to expiration dating was available for Zone III/IV countries. There were many similarities among the requirements for each of the regions of the world, but there were also many differences. Because the stability testing requirements for many countries and also the regulations pertaining to expiration dating periods were different, a systematic approach to establishing stability protocols adequate for worldwide registration and harmonization of expiration dates in the different regions of the world was virtually impossible.

Since 1990, significant gains in the harmonization of technical requirements for registration of pharmaceuticals for human use were achieved with the inception of the International Conference on Harmonization (ICH). The ICH convened for the first time in 1990, and the first draft of the ICH Tripartite Guideline on Stability Testing of New Drug Substances and Products was issued on March 23, 1992. In that Guideline, the stability testing requirements for the United States, European Union, and Japan are defined. Approved ICH guidances include (9):

- Q1A Stability Testing of New Drug Substances and Products
- Q1B Photostability Testing of New Drug Substances and Products

Q1C Stability Testing for New Dosage Forms

Q5C Stability Testing of Biotechnological Biological Products

Additional DRAFT ICH guidances currently in the early review stages are:

Q1D Reduced Designs for Stability Testing of Medicinal Products

Q1E Statistical Analysis and Data Evaluation

In addition, both European and U.S. regulators have published regional guidances that complement or supplement the ICH guidances regarding the stability requirements for establishing the expiration dating of pharmaceutical products. Harmonization of stability testing requirements for Climate Zone III/IV countries has also begun (10). Below, the stability testing requirements and regulations pertaining to expiration dating are broken down by major geographical regions.

United States

The current U.S. requirements pertaining to expiration dating and the stability testing necessary to establish expiration dates of pharmaceutical products are covered in the USP (11), The Code of Federal Regulations, the ICH guidances mentioned previously, and the Draft FDA Guidance for the Industry (12). The three relevant sections of the CFR cover: 1) where expiration dates must appear on a product (§201.17); 2) what products are required or exempted from having an expiration date (§211.137); and 3) requirements for stability tests and the testing program (§211.166). Because the pertinent sections of the CFR for human pharmaceuticals are brief, they are presented later in this article. The USP covers the same general information as the CFR. The FDA Draft Guidance for the Industry titled, Stability Testing of Drug Substances and Drug Products (12) has replaced the original guidance entitled, Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics, published in February 1987. The guidance is intended to be a comprehensive document that provides information on all aspects of stability data generation and use. It references and incorporates substantial text from ICH Q1A, Q1C, Q1B, and Q5C. The guidance provides recommendations regarding the design, conduct, and use of stability studies necessary to establish expiration dates and to support various regulatory applications.

§201.17 DRUGS: LOCATION OF EXPIRATION DATE

When an expiration date of a drug is required, e.g., expiration dating of drug products required by §211.137 (see the next section) of this article, it shall appear on the immediate container and also the outer package, if any, unless it is easily legible through such outer package. However, when single-dose containers are packed in individual cartons, the expiration date may properly appear on the individual carton instead of on the immediate product container (43 FR 45076, Sept. 29, 1978).

§211.137 EXPIRATION DATING

1. To assure that a drug product meets applicable standards of identity, strength, quality, and purity at the time of use, it shall bear an expiration date determined by appropriate stability testing described in §211.166.
2. Expiration dates shall be related to any storage conditions stated on the labeling, as determined by stability studies described in §211.166.
3. If the drug product is to be reconstituted at the time of dispensing, its labeling shall bear expiration information for both the reconstituted and unreconstituted drug products.
4. Expiration dates shall appear on labeling in accordance with the requirements of §201.17 of this article.
5. Homeopathic drug products shall be exempt from the requirements of this section.
6. Allergenic extracts that are labeled "No U.S. Standard of Potency" are exempt from the requirements of this section.
7. New drug products for investigational use are exempt from the requirements of this section, provided they meet appropriate standards or specifications as demonstrated by stability studies during their use in clinical investigations. When new drug products for investigational use are to be reconstituted at the time of dispensing, their labeling shall bear expiration information for the reconstituted drug product.
8. Pending consideration of a proposed exemption, published in the Federal Register, September 29, 1978, the requirements in this section shall not be enforced for human over the counter (OTC) drug products if their labeling does not bear dosage limitations and if they are stable for at least 3 years as supported by appropriate stability data (43 FR 45077,

Sept. 29, 1978, as amended at 46 FR 56412, Nov. 17, 1981; 60 FR 4091, Jan. 20, 1995).

§211.166 STABILITY TESTING

1. There shall be a written testing program designed to assess the stability characteristics of drug products. The results of such stability testing shall be used in determining appropriate storage conditions and expiration dates. The written program shall be followed and shall include:
 - Sample size and test intervals based on statistical criteria for each attribute examined to assure valid estimates of stability
 - Storage conditions for samples retained for testing
 - Reliable, meaningful, and specific test methods
 - Testing of the drug product in the same container-closure system as that in which the drug product is marketed
 - Testing of drug products for reconstitution at the time of dispensing (as directed in the labeling), as well as after they are reconstituted.
2. An adequate number of batches of each drug product shall be tested to determine an appropriate expiration date and a record of such data shall be maintained. Accelerated studies, combined with basic stability information on the components, drug products, and container-closure system, may be used to support tentative expiration dates, provided full shelf-life studies are not available and are being conducted. Where data from accelerated studies are used to project a tentative expiration date that is beyond a date supported by actual shelf-life studies, there must be stability studies conducted, including drug product testing at appropriate intervals, until the tentative expiration date is verified or the appropriate expiration date determined.
3. For homeopathic drug-products, the requirements of this section are as follows:
 - There shall be a written assessment of stability, based at least on testing or examination of the drug product for compatibility of the ingredients and based on marketing experience with the drug product to indicate that there is no degradation of the product for the normal or expected period of use.
 - Evaluation of stability shall be based on the same container-closure system in which the drug-product is being marketed.

4. Allergenic extracts that are labeled “No U.S. Standard of Potency” are exempt from the requirements of this section (43 FR 45077, Sept. 29, 1978, as amended at 46 FR 56412, Nov. 17, 1981).

EUROPEAN UNION

In addition to the ICH, the Committee for Proprietary Medicinal Products, under the European Agency for the Evaluation of Medicinal Products (EMA) (13), issued, in 1997 and 1998, a number of stability-related guidances to be used for establishing stability testing protocols for drug products to be filed in European countries. The documents complement and supplement the ICH guidelines and are listed below:

1. Reduced Stability Testing Plan-Bracketing and Matrixing: Annex to Note for Guidance on Stability Testing of New Drug Substances and Products (14)—CPMP/QWP/157/96
2. Note for Guidance on Maximum Shelf-Life for Sterile Products for Human Use After First Opening or Following Reconstitution (16)—CPMP/QWP/159/96
3. Note for Guidance on Declaration of Storage Conditions for Medicinal Products in the Products Particulars (17)—CPMP/QWP/609/96
4. Note for Guidance on Stability Testing of Existing Active Substances and Related Finished Products (18)—CPMP/QWP/556/96
5. Note for Guidance on Stability Testing for Type II Variation to a Marketing Authorization (19)—CPMP/QWP/576/96
6. Note for Guidance for In-Use Stability Testing of Human Medicinal Products—Annex to Note for Guidance on Stability Testing of Existing Active Substances and Related Finished Products and Note for Guidance on Stability Testing of Existing Active Substances and Related Finished Products—CPMP/QWP/2934/99

Japan

In addition to ICH guidelines, the specific Japanese stability testing requirements for establishing expiration dating are described in the regional guidelines. In particular, the Japanese Ministry of Health and Welfare (MHW) has issued a guidance entitled, *Handling of Data on Stability Testing Attached to Applications for Approval to Manufacture or Import Drugs* (22). For the

most part, the MHW guidance complements the ICH guidelines. The MHW guidance provides more detailed requirements for “stress” testing than that found in the ICH and the guidances for stability testing used in United States or European Union (see above). In fact, specific stress testing conditions and testing duration and frequency are provided. In contrast, European Union and the U.S. regional guidances do not recommend specific testing conditions and schedules and specify that stress testing is usually accomplished during routine formulation and analytical methods development. This discrepancy between the regional requirements in Japan versus those in European Union and the United States is a notable opportunity for harmonization.

Climate Zone III/IV Countries

Currently countries in climate Zones III and IV are not member countries of the ICH, although it has been recommended by Grimm (14) that the ICH Tripartite Guideline be extended to include Zone III/IV countries. Needless to say, great strides in the harmonization of stability testing requirements and expiration dating for climate Zone III and IV countries has begun. The definitive source on stability testing requirements can be found in a set of guidelines published by the World Health Organization, entitled, *Guidelines for Stability Testing of Pharmaceutical Products Containing Well-Established Drug Substances in Conventional Dosage Forms* (15). This document outlines the conditions and testing required for Zone III and IV countries. Also contained in that document are a list of recommended storage conditions to be prominently indicated on the label of pharmaceutical products distributed in climate Zone III and IV countries. Another set of guidelines relevant to stability testing in climate Zone III and IV is a document issued by MERCOSUL (“Mercado Comunal del Sur” or “Common Market of the South”), entitled *Pharmaceutical Products Stability* (10). The MERCOSUL document was specifically intended for use as a guidance for designing stability testing programs by countries that are members of MERCOSUL, Argentina, Paraguay, Uruguay, and Brazil. The stability testing requirements described in the MERCOSUL guidelines essentially match the conditions and testing requirements in the WHO document. The stability testing requirements for Zone III and IV, for which the WHO guidelines are considered the definitive requirements, are just now being implemented throughout the industry to support marketing of pharmaceutical products in climate Zone III and IV countries.

STATISTICAL DETERMINATION OF SHELF-LIFE FROM LONG-TERM STORAGE DATA

The current approach to analyzing stability data to predict shelf-life involves statistical analysis of long-term storage data. The accelerated data are generally used as supportive data. In particular, it is recommended in ICH Q1A, CPMP QWP/556/96, and the FDA Draft Guidance that a shelf-life of a drug product be determined by the point at which the 95% one-sided confidence limit for the mean degradation curve intersects the acceptable lower specification limit. Typically, statistical analysis to determine shelf-life is applied to at least 12 months of long-term storage data on three production scale batches of drug product. Data from individual batches are first treated by least-squares fitting. The true line representing degradation behavior is not known a priori, but it is estimated by least-squares analysis. The confidence interval about the least-squares line from fitting the stability data is also obtained. As noted above, it is the intercept between the line representing the confidence limit and the upper or lower registration limit that yields the expiration period. If the application of appropriate statistical tests to the slopes and zero time intercepts of the regression lines for the individual batches indicate that the batch-to-batch variability is small (e.g., p values for level of significance of rejection of more than 0.25), and therefore, the data pool, it is advantageous to then combine all the data into one overall estimate for shelf-life. If it is inappropriate to combine data from several batches, the overall expiration dating period is determined by the minimum time a batch may be expected to remain within acceptable and justified limits (9, 12). Detailed information about the statistical evaluation of long-term stability data using regression analysis, tests for batch similarity, and data pooling is described in Carstensen (16), Wessels (17), Lin (18), Bancroft (19), and the FDA guidance (12).

Software packages that handle the statistical evaluation of stability data per the ICH or FDA guidelines are sold commercially. Also, the FDA has developed and makes available its own drug formulation stability test program, STAB, which can be downloaded from the FDA website (20). This statistical approach to calculating the expiry dating period from long-term data applies to new drug substances and their products (9, 12) as well as products containing established unstable substances (15).

Often the expiration dating period determined by the statistical analysis described above results in extrapolation

of the regression lines to times for which there are no stability data available. For extrapolation beyond the observed range to be valid, the assumed degradation relationship must continue to apply through the estimated expiration dating period. As stated in the FDA *Draft Guidance on Stability Testing* (13), an expiration dating period assigned in this manner should always be justified and supported by accelerated test data and is considered tentative until confirmed through full long-term stability data from at least three production batches reported through annual reports. For known stable products, a more flexible approach is taken. According to the WHO guideline (15), one may directly assign a 24-month shelf-life to a product provided that the active ingredient is stable i.e., stability studies have been performed per tabulated accelerated test conditions for Zones II and IV with no significant changes; supporting data indicate that similar formulations have been assigned a shelf-life of 24 months or more; and the manufacturer continues to perform real-time studies until the proposed shelf-life has elapsed. For products (simple dosage forms) covered under ANDAs, the FDA also permits a direct assignment of a shelf-life as long as 24 months at labeled storage conditions if the accelerated stability data at 0, 1, 2, and 3 months and available long-term stability data are satisfactory (12). CPMP guideline QWP/556/96, however, does not follow this approach, and suggests a statistical analysis of real-time data for the determination of shelf-life, even for existing drug substances and products (13).

How the actual expiration date is calculated from the expiration dating period determined from the statistical analysis described above is also of importance. The FDA draft guidance and CPMP/QWP/486/95 both state that the computation of expiration period of the drug product should begin no later than the time of release of the batch (12, 13). The date of release generally should not exceed 30 days from the production date regardless of packaging date (12). If the release date of the lot fails the 30-day test, a different standard applies in fixing the expiration period. In such cases, the expiration date is calculated from within 30 days of the manufacture of the lot, rather than from the release date (12, 13, 21). If the expiration date includes only a month and year, the product should meet specifications through the last-day-of-that-month guidance (21). The data generated in support of the assigned expiration dating period should be from long-term studies under the storage conditions recommended in the labeling (12).

Arrhenius analysis, described below, of multitemperature data cannot be used alone to set a definitive shelf-life

at a different temperature than that for which long-term data are available. This is primarily because discrepancies in degradation mechanisms at different temperatures makes the Arrhenius law invalid for prediction of a definitive shelf-life from accelerated data. Although Arrhenius analysis is not used as a primary method for proposing a tentative expiration dating period based on accelerated data, application of this analysis to accelerated data can be used as supportive data in conjunction with extrapolation of long-term data. As described below, Arrhenius analysis can be used to anticipate the impact of storage condition or climate zone on the expiration dating interval.

IMPACT OF STORAGE CONDITION OR CLIMATIC ZONE ON EXPIRATION DATING INTERVAL

Arrhenius analysis of a given degradation process yields a relationship between temperature and rate constant. The results of such an analysis can be useful in predicting the impact of changes in storage conditions or climatic zone on expiration dating interval of a given drug-product. In particular, Arrhenius analysis can be used to predict the relationship between the expiration dating period and the label storage conditions. In practice a given formulation and container/closure system may have different expiration periods assigned to them, depending on where, i.e., in which "climate zone," they are being marketed. The Arrhenius relationship between temperature and rate constant is derived below.

For zero-order processes, by which many chemical reactions can be modeled, the rate of disappearance of the reactant A is constant and independent of its concentration, as shown in Eq. 1:

$$-dA/dt = k \quad (1)$$

Solving Eq. 1 yields Eq 2:

$$A = A_0 - kt \quad (2)$$

where,

A = the amount of A remaining at time t ,
 A_0 = the initial amount of A, and
 k = rate constant.

For reactions following zero-order kinetics, a plot of A versus t yields a straight line whose slope is equal to $-k$.

For first-order processes, the rate of disappearance of the reactant A is proportional to the concentration of A at any time t , as shown Eq. 3.

$$-dA/dt = kA \quad (3)$$

The solution of the Eq. 3 yields Eq. 4:

$$\ln A/A_0 = -kt \quad (4)$$

or

$$\ln A = -kt + \ln A_0$$

For stability data that follows first-order kinetics, a plot of $\ln A$ versus t will yield a straight line whose slope is k . The rate constant k , in almost all cases, is a function of the temperature T . For most pharmaceutical products, as T is increased, the rate constant and, therefore, the rate of degradation increases. This is the basis for the well known Arrhenius relationship that states that for a given chemical reaction, the empirical relationship between k and T may be described as in Eq. 5:

$$k = b_0 e^{-E/RT} \quad (5)$$

or

$$\ln k = \ln b_0 - E/RT$$

where,

T = Kelvin temperature,

E = activation energy,

R = ideal gas constant, and

b_0 = constant depending on the molecule of interest.

If a particular kinetic process follows Arrhenius' Law, then a plot of $\ln k$ versus $1/T$ will yield a straight line with a slope of E/R .

A good example of the application of Arrhenius analysis is a pharmaceutical product whose shelf-life is limited by degradation of the active ingredient A . Using stability-indicating assays, the loss of A with time at 25°C has been shown to follow zero-order kinetics between 100 and 90% of the labeled amount. In addition, data from accelerated studies at 40, 50, and 60°C show the same zero-order behavior. If the rate constants from the 25, 40, 50, and 60°C experiments follow Arrhenius' Eq. 5, then from the straight line plot of $\ln k$ vs. $1/T$ for component A in this matrix, the rate constant at any temperature in the data range may be obtained by interpolation. Thus, knowing k for any temperature and the time to be spent at that temperature, the stability performance may be calculated by Eq. 2. In other words, Arrhenius analysis can be used to predict

an expiration date at any temperature owing to storage or climatic zone under which the Arrhenius behavior is valid. Although useful for obtaining an understanding of the impact of temperature changes on expiration dating, Arrhenius analysis breaks down when the degradation mechanisms occurring at one temperature are different from the degradation mechanisms occurring at other temperatures.

Grimm has used the Arrhenius relationship to determine "predictive factors" that can be used to predict the shelf-life at long-term storage in various climate zones from accelerated data (14). For example, the degradation rate at 40°C can be determined by multiplying the time it takes the drug to degrade by 5% at 25 or 30°C by "predictive factors" for those temperatures. Using an activation energy of 83 kJmole⁻¹ Grimm derives predictive factors of 5 and 3.3 for 25 and 30°C, respectively. Obviously, the use of "predictive factors" is only an approximation and breaks down when Arrhenius behavior at different temperatures is not followed.

Impact of Post-Approval Changes on Expiration Dating Period

Often, after a pharmaceutical drug product and its associated dating period and storage condition have been approved, there is a desire to change some aspect of the product (e.g., production process, formulation, packaging, etc.). When a postapproval change occurs, the expiration dating period approved for the original product must be justified for the product after the postapproval change. To justify an expiration date after a postapproval change has been made, various amounts of new or different stability testing are required. Since the last edition of this article, there has been significant activity in the development of new regulations and refinement of existing regulations corresponding to stability testing required for postapproval changes.

The *ICH Topic Q1C; Stability Testing: Requirements for New Dosage Forms* is an annex to the parent ICH guidance (issued 5/1997) Q1A and addresses the recommendations for the data that should be submitted regarding stability of new dosage forms by the owner of the original application after the original submission for new active substances and medicinal products. In the United States, the FDA has issued *Scale-Up and Post-Approval Changes* (SUPAC) guidelines that dictate the stability testing requirements for various postapproval changes for U.S. filings. The available SUPAC guidances are listed below (23):

1. SUPAC-IR: Immediate-Release Solid Oral Dosage Forms: Scale-Up and Post-Approval Changes: Chemistry, Manufacturing and Controls, In Vitro Dissolution Testing, and In Vivo Bioequivalence Documentation (Manufacturing Equipment Addendum Issued 1/1999).
2. SUPAC-MR: Modified Release Solid Oral Dosage Forms Scale-Up and Postapproval Changes: Chemistry, Manufacturing and Controls: In Vitro Dissolution Testing and In Vivo Bioequivalence Documentation (Issued 10/6/97, Manufacturing Equipment Addendum Issued 1/1999).
3. SUPAC-SS: Nonsterile Semisolid Dosage Forms; Scale-Up and Post-Approval Changes: Chemistry, Manufacturing and Controls; In Vitro Release Testing and In Vivo Bioequivalence Documentation (Issued 5/1997; Posted 6/16/1997, Addendum Issued 12/1998).

In Europe, two regulations have been introduced to address “variations” to medicinal products, Regulation (EEC) No. 541/95 and 542/95. For the implementation of the procedures set out by these regulations, a number of guidances have been prepared and an application form for a variation to a marketing authorization has been issued.

In the United States, the approach for determination of stability testing required for a given change follows a tiered approach, with more additional stability data obviously being required for more significant postapproval changes (SUPAC reference). Table 1, reproduced from the

FDA Guidance on Stability Testing, describes the five stability data package types required to support a postapproval change.

For each type of postapproval change, the above table is used to help determine what type of Stability Data Package should be filed. A discussion of the type of Stability Data Package necessary for different changes is found in the FDA Draft Guidance for Stability Testing (12) and dosage form-specific information is available in the SUPAC guidances listed above. For certain changes no prior approval is needed, and a Changes-Being-Effectuated supplement can be filed. This is not the case in Europe, where all postapproval changes require an amendment to be approved before marketing medicinal product manufactured with post-approval change.

EXTENSIONS OR REDUCTIONS IN EXPIRATION DATING

Often after extensive stability data have been collected, there is a desire to change the expiration dating of a product. In particular, the available stability data support either an extension or a reduction in the expiration date approved in the original filing. Changes in expiration dating is covered in the FDA guidance.

For example, the methods for obtaining an extension is outlined in the FDA guidance (12):

Table 1 Stability data package types to support a postapproval change

Stability data package	Stability data at time of submission	Stability commitment
Type 0	None	None beyond the regular annual batches
Type 1	None	First (1) production batch and annual batches thereafter on long-term stability studies
Type 2	Three months of comparative accelerated data and available long-term data on 1 batch ^a of drug product with the proposed change	First (1) production batch ^b and annual batches thereafter on long-term stability studies ^c
Type 3	Three months of comparative accelerated data and available long-term data on 1 batch ^a of drug product with the proposed change	First (3) production batch ^b and annual batches thereafter on long-term stability studies ^c
Type 4	Three months of comparative accelerated data and available long-term data on 3 batches ^a of drug product with the proposed change	First 3 production batch ^b and annual batches thereafter on long-term stability studies ^c

^aPilot scale batches acceptable.

^bIf not submitted in the supplement.

^cUsing the approved stability protocol and reporting data in annual reports.

1. Can be described in an annual report if criteria set forth in the approved protocol are met. In this case, the extension of the expiration dating period must be based on full long-term stability data obtained from at least three production batches (21 CFR 314.70).
2. A prior approval supplement is necessary to extend a tentative expiration date based on three pilot-scale batches. The expiration dating remains tentative until confirmed with full long-term data from at least three production batches.

The methods for shortening an expiration date is through a Changes-Being Effected supplement (21 CFR 314.70 or 21 CFR 601.12).

UNRESOLVED ISSUES FOR THE NEW MILLENNIUM

Two major issues still require attention: first, complete harmonization of stability testing requirements throughout the world remains to be achieved. In particular, benefits could be obtained by harmonizing the "stress" testing requirements for different ICH member countries, complete harmonization of stability testing requirements for all Zone III/IV countries, and more extensive harmonization of stability testing conditions between Zone III/IV countries, where 30C/70%RH is the long-term storage condition for stability studies, and ICH member countries, where 30C/60%RH is the intermediate storage condition. The second unresolved issue is worldwide harmonization of label storage statements. Because the ultimate objective of a stability study is to predict a product's shelf-life at a particular label storage condition, it makes sense that label storage statements be harmonized much the way stability testing requirements have been harmonized. Unfortunately, this is not the case. Q1A suggests that labeling be based on each country's requirements. These requirements are different for most countries. For example, label statements in the United States satisfy the USP, whereas in Europe, no labeling is required. In Japan, labeling is necessary only if the product's stability is less than 3 years. In the WHO guidance, labeling is also defined by the requirements of the country. As can be seen, harmonization of label storage statements still needs to be accomplished. More thorough harmonization of stability testing requirements throughout the world and worldwide harmonization of label storage statements is highly desirable to streamline stability testing programs designed to support a global marketplace.

CONCLUSIONS

Worldwide harmonization of stability testing requirements has made significant progress, but it is still not complete (e.g., labeling requirements are country-specific). In the last decade, the ICH guidances and complementary regional guidelines have significantly enhanced the level of harmonization of stability testing requirements for Europe, Japan, and the United States. The WHO guideline for stability testing of pharmaceutical products containing well established drug substances in conventional dosage forms and the MERCOSUR guidance have begun to accomplish harmonization of stability testing requirements in Zone III/IV. There is still a considerable amount of harmonization to be achieved in Zone III/IV countries. As Zone III/IV countries continue to harmonize, it will be interesting to see whether a new intermediate condition is adopted by the ICH that replaces 30C/60%RH, which would allow worldwide harmonization of testing requirements to support determination of expiration dating periods for each of the four climate zones of the world.

REFERENCES

1. VanArendonk, M.; Dukes, G.R. *Expiration Dating. Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1995; 5, 379–394.
2. Federal Register 601. January 15, 1971.
3. U.S. Pharmacopeia XXII. *National Formulary XVII*; U.S. Pharm. Convention, Inc.: Rockville, MD, 1990; 10.
4. Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drugs and Biologics, Washington, DC, 1987.
5. Guide to Good Manufacturing Practices for Medicinal Products. The Rules Governing Medicinal Products in the European Community, Vol. IV. Luxembourg: Office for Official Publications of the European Communities: Luxembourg, 1989.
6. Stability Tests on Active Substances and Finished Products, Notes for Guidance Concerning the Application of Section F of Part I of the Annex to Directive 75/318/EEC. III/66/87-EN, Rev. 4, Brussels, 1988, 7.
7. Notification No. 698, Pharmaceutical Affairs Bureau, Japanese Ministry of Health and Welfare, Tokyo, 1980.
8. Draft Policy to Handle Stability Data Required in Applying for Approval to Manufacture (Import) Drugs and Draft Guidelines on Methods to Perform Stability Test, Ministry of Health and Welfare: The First Evaluation and

- Registration Division, Pharmaceutical Affairs Bureau, 1990.
9. <http://www.ifpma.org/ich5q.html#stability> (accessed April 2000).
 10. MERCSUL/GMC/P.RES No. 53/96.
 11. U.S. Pharmacopeia XXIV. *National Formulary XIX*; U.S. Pharm. Convention, Inc.: Rockville, MD, 2000.
 12. <http://www.fda.gov/cder/guidance/1707dft.pdf> (accessed April 2000).
 13. <http://www.eudra.org/emea.html> (accessed April 2000).
 14. Grimm, W. Extensions of the International Conference on Harmonization Tripartite Guideline for Stability Testing of New Drug Substances and Products to Countries of Climatic Zones III and IV. *Drug Dev. Ind. Pharm.* **1998**, *24*, 313–325.
 15. WHO Technical Report Series 863, World Health Organization Geneva, Switzerland, 1996
 16. Carstensen, J.T.; Nelson, E.W. *J. Pharm. Sci.* **1976**, *65*, 311.
 17. Wessels, P., et al. Statistical Evaluation of Stability Data for Pharmaceutical Products for Specification Setting. *Drug Dev. Ind. Pharm.* **1997**, *23*, 427–439.
 18. Lin, K.K.; Lin, T-Y.D.; Kelley, R.E. Stability of Drugs: Room Temperature Tests. *Statistics in the Pharmaceutical Industry*; Buncher, C.R., Tsay, J.-Y., Eds.; Marcel Dekker, Inc.: New York, 1994; 419–444.
 19. Bancroft, T.A. Analysis and Inference for Incompletely Specified Models Involving the Use of Preliminary Test(s) of Significance. *Biometrics* **1964**, *20* (3), 427–442.
 20. <http://www.fda.gov> (accessed April 2000).
 21. Washington Drug Letter, No. 36. 1999, *31*.
 22. Handling of Data on Stability Testing Attached to Applications for Approval to Manufacture of Import Drugs, Notification No. 413; Pharmaceutical Affairs Bureau, Japanese Ministry of Health and Welfare, Tokyo, 1992.
 23. <http://www.fda.gov/guidance/1721fnl.pdf> (accessed April 2000).

EXPERT SYSTEMS IN PHARMACEUTICAL PRODUCT DEVELOPMENT

Raymond C. Rowe

Ronald J. Roberts

AstraZeneca, Macclesfield, Cheshire, United Kingdom

INTRODUCTION

The process of formulation, whether it be for oral products (e.g., tablets and capsules), parenterals [e.g., intravenous (iv) injections], or any one of the myriad of pharmaceutical products, is generically the same. The process begins with some form of product specification and ends with the generation of one or more formulations that meet the requirements. Although the formulation consists of a list of ingredients and their proportions together with some processing variables where appropriate, the specification can vary considerably from one application to another. In some cases it may be very specific, expressed in terms of a performance level when subjected to a specific test, or quite general. It may also contain potentially conflicting performance criteria that the formulator may need to redefine in the light of experience. Figure 1 shows a typical formulation process broken down into its constituent tasks and subtasks (1).

In designing a formulation, the formulator must take into account the properties of the active ingredient as well as possible chemical interactions between it and the other ingredients added to improve processability and product properties since these may result in chemical instability. There may even be interactions between added ingredients, leading to physical instability. Commercial factors as well as the policy of the industry toward ingredient usage are important influences, as are production factors in the intended markets. The formulator may also routinely access databases on previous formulations as well as make use of mathematical models. During the formulation process, specific tests may need to be run to evaluate the properties of the proposed formulation and an analysis of unexpected results may lead to an adjustment of the ingredients and/or their levels.

TECHNOLOGY

There is a wide divergence of views as to what defines an expert system. Examples include the following:

1. "An expert system is a knowledge-based system that emulates expert thought to solve significant problems in a particular domain of expertise" (2).
2. "An expert system is a computer program that draws on the knowledge of human experts captured in a knowledge base to solve problems that normally require human expertise" (3).

In its simplest form, an expert system has three major components: 1) an interface, a monitor, and keyboard that allow two-way communication between the user and the system; 2) a knowledge base where all the knowledge pertaining to the domain is stored; and 3) an inference engine where the knowledge is extracted and manipulated to solve the problem at hand. Inferencing strategies may be either forward chaining, which involves the system reasoning from data and information gained by consultation from the user to form a hypothesis, or backward chaining, which involves the system starting with a hypothesis and then attempting to find data and information to prove or disprove the hypothesis. Both strategies are included in most expert systems.

Knowledge in any domain takes the form of facts and heuristics; the former being valid, true, and justifiable by rigorous argument, the latter (often referred to as rules of thumb) being the expert's best judgment in any particular circumstance and hence justifiable only by example. Associated with these are the terms data and information, the former referring to facts and figures, the latter being data transferred by processing such that the data are meaningful to the person receiving the information. Knowledge can, therefore, be regarded as information combined with heuristics and rules. It is the objective of the knowledge engineer to acquire or elicit this knowledge and structure it in a computer-readable format.

Knowledge acquisition is probably one of the most difficult stages in the development of an expert system. It is both time-consuming and tedious as well as being expensive and often difficult to manage. However, it is a necessary element in the building of an expert system and, if done well, will undoubtedly lead to potentially useful systems. The basic model of knowledge acquisition is one of a team process whereby the knowledge engineer

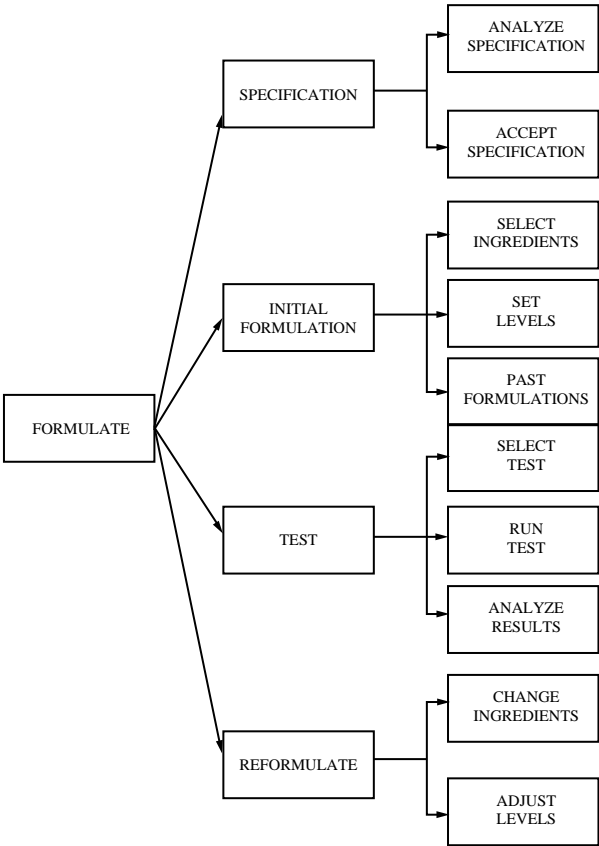


Fig. 1 Tasks and subtasks in the formulation process. (From Ref. 1.)

mediates between the expert(s), the users, and the knowledge bases. The knowledge engineer must acquire or elicit knowledge from not only the expert(s) but also from all the other potential sources, including written documents (research reports, reference manuals, and operating procedures policy statements) as well as consultants, users, and managers. In the case of experts, knowledge is usually acquired through face-to-face interviews. While this process is tedious and can place great demands on both the expert and knowledge engineer, it requires little equipment (e.g., tape recorder or notebook), is highly flexible, and often yields a considerable amount of useful information. At all times, there must be empathy between the participants and in many cases it is helpful to have two knowledgeable engineers present at the interview.

A technique that is often used in the acquisition process is the rapid prototyping approach. In this approach, the knowledge engineer builds a small prototype system as early as possible. This is then shown to both the expert and user, who can suggest modifications and additions. Here the system grows incrementally as more information and

knowledge are gained. This methodology has been used successfully in the development of systems for the formulation of pharmaceuticals.

Once acquired, there are many ways of representing the knowledge in the knowledge base, including production rules, frames, semantic networks, decision tables, and trees and objects (2). Probably the most common methodology is the production rule, which expresses the relationship between several pieces of information by way of conditional statements that specify sections under certain sets of conditions, for example:

IF	(condition 1)
AND	(condition 2)
OR	(condition 3)
THEN	(action)
UNLESS	(exception)
BECAUSE	(reason)

Each rule implements an autonomous piece of knowledge and is easy to understand. Unfortunately, complex knowledge can require large numbers of rules, causing the system to become difficult to manage. The decision as to which method of knowledge representation should be adopted is dependent primarily on the complexity of the domain.

Expert systems can be developed using either conventional computer languages, special purpose languages, or with the assistance of development shells or tool-kits. Conventional languages such as PASCAL and C have the advantages of wide applicability and full flexibility to create the control and inferencing strategies required. They also are well supported and easy to customize. However, considerable amounts of time and effort are needed to create the basic facilities.

Specialized languages, such as LISP (a recursive language and the primary one for artificial intelligence research), PROLOG (a language based on first-order predicate logic), and SMALLTALK (an object-orientated language), have been used extensively in the development of expert systems. They have the advantages of applicability and flexibility of the conventional languages but are faster to implement.

Expert system shells and tool kits are sets of computer programs written in either conventional or specialized languages that can form an expert system when loaded with the relevant knowledge. They compromise on applicability and flexibility but allow more rapid development. Many offer basic facilities, including the means to prepare and store knowledge as a set of rules and to make deductions by chaining the rules together in an inferential process.

Shells differ in their secondary characteristics, such as user interfaces, operating speeds, the method of knowledge representation, and the associated algorithmic and arithmetic computational facilities. It is not surprising, therefore, that formulation is a highly specialized task that requires specific knowledge and often years of experience. This kind of expertise is not easily documented and hence, senior formulators often spend considerable amounts of time training new personnel. In addition, retirement or personnel moves can lead to a loss of irreplaceable commercial knowledge. Computer technology in the form of expert systems provides an affordable means of capturing this knowledge and expertise in a documented form available to all.

One such shell of specific importance in product formulation is Product Formulation Expert System (PFES), now termed Formulogic, developed by Logica UK Ltd. Formulogic is a reusable software kernel and associated methodology to support the generic formulation process. It arose from research work during the mid-1980s and is now used for a variety of formulation support tools across a range of industry sectors, most notably pharmaceuticals (4, 5).

Formulogic is designed specifically for building formulation systems. Its formulation capability is generic, i.e., it is not specific to any particular domain. Individual formulation applications are developed using the shell by defining characteristics of the domain and the corresponding approach to formulation. The end result is a decision support tool for formulators that provides assistance in all aspects of the formulation development process. Formulogic provides the expert system developer with the knowledge representation structures that are common to most product formulation tasks so that a new application can be developed rapidly and efficiently. The architecture of Formulogic comprises three levels—the Physical Level, the Task Level, and the Control Level.

The Physical Level contains all the “nuts and bolts” of the formulation domain in a number of information sources, including a database. The Physical Level is accessed from the Task Level via a query interface. The physical net contains the domain knowledge in a number of objects. An object consists of a set of attributes, each of which may have zero or more values. The objects are arranged in a classification hierarchy. Subobjects, which descend from another object, inherit their attributes and values.

The Task Level is where the formulation problem-solving activity takes place. The formulation process is driven via the generation of a hierarchy of tasks. A task represents some well-defined activity. The hierarchy has an indefinite number of task levels. Domain knowledge about the formulation application is distributed throughout

the hierarchy, with more abstract knowledge represented towards the top of the hierarchy and more specific knowledge toward the bottom.

The task decomposition allows the problem-solving process to be largely decoupled between tasks and also facilitates reasoning about subtasks. An important principle is that tasks plan about and directly manipulate only their immediate subtasks. Recursive application of this principle is the key to integrated behavior of a formulation system as a whole. The task tree is built on dynamically, depending upon the specification at hand as the problem-solving process proceeds. This is different to the object hierarchy where the structure is fixed for a particular domain.

Agendas, a particularly important class of objects used for communication between tasks, also are part of the Task Level. Agendas are important because the user exercises control over the formulation process principally through their manipulation. For parallel reasoning and backtracking by the user it is necessary to maintain more than one world. A world contains a formulation object and specification object together with agendas (in other words, a complete description of the state of the formulation process at any one time). Tasks run on a world or set of worlds; it is meaningless for tasks to run without reference to worlds.

Tasks perform their function by the execution of processes. Each task contains several types of processes. The precondition process assesses whether or not the task can play a sensible role in the current context. The action process performs the primary work of the task, which can include scheduling subtasks to be run next. The monitor process executes between each of the subtasks, typically to assess their result. Finally, the postcondition process assesses the success of the task as a whole immediately prior to completion.

Each process consists of a set of production rules; a rule is said to fire if its condition is true and its exception is false. When a rule fires, its action is executed. The reason is for information only and is not interpreted by Formulogic.

Rules are assigned a priority that reflects the order in which they should be considered. When executing a rule set, rules are tried in order until no more rules can fire. To find a suitable rule, Formulogic orders the rules first on priority and then on specificity. The first rule having a true condition and false exception is fired, and its action executed. The process of finding the next rule that can fire then starts all over again. To avoid looping, rules are only allowed to fire once with the same set of variable bindings.

The Control Level is concerned with the mechanics of running the Task Level. It contains no domain knowledge

and requires no design amendment when a new formulation system is built. Although the Task Level decides which tasks need running and the order in which they should run, the Control Level deals with the mechanics of actually running them and of passing control to them.

The typical functionality of a completed application is as follows:

1. The user enters the product specification, which forms the starting point of the formulation. In a tablet formulation, for example, this would consist of the drug details (unless already known to the system) and the dose.
2. Formulogic steps through a series of tasks to select ingredients and determine their quantities based on the product specification. It achieves this by following the rules and other knowledge that have been built into the system during development. An initial formulation is produced.
3. If the system performs reformulation in addition to producing initial formulations, then the user can enter the test results that have been performed on the product. Formulogic will then determine what kinds of problems with the formulation are indicated by the test results. The user can agree with the system's analysis of the problems or modify them as he or she sees fit. It uses the problem summary to make recommendations about what ingredients need to be added or what quantities need to be altered, and again the user can override the recommendations if he or she wishes. Once the user is happy with the recommendations, Formulogic will produce a modified formulation that meets the new requirements.

At any point during a session, the user can ask for an explanation of the results, browse the system's knowledge, or revert to an earlier stage of the process to modify the specification and obtain another formulation. The user can also save formulations (along with their associated product specifications) and generate printed reports.

APPLICATIONS

The first recorded reference to the use of expert systems in pharmaceutical product formulation was by Bradshaw on the April 27, 1989, in the London Financial Times (6), closely followed by an article in the autumn of the same year by Walko (7). Both refer to the work then being undertaken by personnel at ICI/Zeneca Pharmaceuticals (now AstraZeneca), United Kingdom (UK) and Logica UK Ltd. to develop expert systems for formulating pharmaceuticals using Formulogic. Since that time,

several companies and academic institutions have reported on their experiences in this area (Table 1). This article will review these applications.

The Boots System

Although not strictly developed for pharmaceutical formulation, this system has been included since it is the only one known for formulating topicals. It was developed to assist formulators in the formulation of sun care products—a highly skilled occupation requiring 3–4 years of experience to attain a reasonable level of experience.

Implemented with the use of Formulogic, the system, called SOLTAN, uses knowledge captured by interviewing senior formulators. Ingredients, processes, and relationships of the formulation are represented in a way that reflects their groupings and associations in the real world. In addition, existing information sources, such as databases, are presented in a frame-based semantic network that can be manipulated by the problem-solving knowledge of the domain. Tasks are structured in a hierarchy that is built up dynamically depending on the specification at hand as the problem-solving process proceeds. Knowledge about the formulation is distributed throughout the task hierarchy, with strategic knowledge represented toward the top of the hierarchy and tactical knowledge towards the bottom.

The system was originally developed to formulate sun oils (solutions of ultraviolet absorbers in emollients) but has been rapidly extended, with the incorporation of basic emulsion technology, to cover oil-in-water lotions. Subsequently, the system has been further expanded to incorporate water-in-oil, oil-in-water, and water-in-silicone

Table 1 Published applications of product formulation expert systems in pharmaceuticals

Company/institution	Domain	Development tool
Boots Company	Topicals	Formulogic
Cadila Laboratories (India)	Tablets	PROLOG
University of Heidelberg	Aerosols	C/SMALLTALK
	Tablets	
	Capsules	
	Injections	
University of London/ Capsugel	Capsules	C
Sanofi Research	Capsules	Formulogic
Zeneca Pharmaceuticals	Tablets	Formulogic
	Parenterals	
	Film coatings	

creams and lotions. It can now be used to formulate all types of skin care products, not just sun care products (8). The system won second prize in the UK DTI Manufacturing Intelligence Awards in 1991. It is the only system for which the developers have given details of costings and quantitative benefits.

The Cadila System

Cadila Laboratories Ltd. of Ahmedabad, India have developed an expert system for the formulation of tablets for active ingredients based on their physical, chemical, and biological properties (9). The system first identifies the desirable properties in the excipients for optimum compatibility with the active ingredient and then selects those that have the required properties based on the assumption that all tablet formulations comprise at least one binder, one disintegrant, and one lubricant. Other excipients such as diluents (fillers) or glidants are then added as required.

Knowledge is acquired through "active collaboration" with domain experts over a period of 6–7 months. It is structured in two knowledge bases in a spreadsheet format. In the knowledge base concerning the interactions between active ingredients and excipients, the columns represent the properties of the excipients with descriptors of "strong," "moderate," and "weak." The rows represent the properties of the active ingredients, e.g., functional groups (primary amines, secondary amines, highly acidic etc.), solubility (very soluble, freely soluble, soluble, sparingly soluble, slightly soluble, very slightly soluble, insoluble), density (low, moderate, high), etc. Each cell in the spreadsheet then represents the knowledge of the interaction between the various properties. Production rules derived from this knowledge are in the following forms:

IF	(functional group of active ingredient is "primary/secondary amine")
THEN	(add "strong" binder)
AND	(add "strong" disintegrant)
AND	(avoid lactose)

or

IF	(functional group of active ingredient is "highly acidic")
THEN	(add "moderate" binder)
AND	(add "moderate" disintegrant)
AND	(avoid starch)

or

IF	(active ingredient is soluble)
THEN	(add "weak" binder)
AND	(add "weak" disintegrant)

A similar approach is used for the knowledge base concerning the excipients, where the columns now represent details (e.g., name, minimum, maximum, and normal concentrations) of the excipients and the rows their properties (e.g., type and the descriptors—strong, moderate, and weak). Each cell in the spreadsheet then represents the name and the amount to be added to the formulation.

The system, written in PROLOG, is menu-driven and interactive with the user. The user is first prompted to input all the known properties of the new active ingredient. If the properties have descriptors, the user can select the appropriate ones. All information can be edited to correct errors. The expert system then consults the knowledge bases, suggesting compatible excipients and a formulation. If the latter is unacceptable, the system provides alternative formulations with explanations. All formulations can be stored along with explanations, if necessary. The user is able to update the knowledge base via an interface with a spreadsheet. An example of a formulation generated for the analgesic drug paracetamol, or acetaminophen, (dose 500 mg) is shown in Table 2. It is interesting to note that the diluent/filler is unnamed; it can be assumed that it will not be lactose since the relevant production rule indicates that there would be an interaction with the secondary amine in paracetamol. Furthermore, an examination of formulations on the market indicates that

Table 2 An example of a tablet formulation for the analgesic drug paracetamol as generated by the Cadila system

<i>Input</i>		
Dose	500 mg	
Functional group	Secondary amines	
Solubility	Sparingly soluble	
Density	Moderate	
Hygroscopicity	Moderate	
Dissolution	Slow	
Desired tablet weight	570 mg	
<i>Output</i>		
Active agent	Paracetamol	500.0 mg
Binder	Pregelatinized starch	43.7 mg
Disintegrant	Sodium starch glycolate	5.0 mg
Lubricant	Stearic acid	2.5 mg
Diluent/filler	Unnamed	20.0 mg
	Tablet weight	571.2 mg

(From Ref. 9.)

none contain lactose and that some contain mixtures of maize starch, sodium starch glycolate, stearic acid, magnesium stearate, and microcrystalline cellulose, adding further credibility to the Cadila system.

When first implemented, the prototype system had 150 rules, but this has expanded rapidly to approximately 300 rules in order to increase reliability. This is expected to increase further over time. The system is regarded as being highly successful, providing competitive advantage to the company (9).

The Galenical Development System

The Galenical Developmental System (GSH) was developed by personnel in the Department of Pharmaceutical and Biopharmaceutics and the Department of Medical Informatics at the University of Heidelberg, Germany. It is designed to provide assistance in the development of a range of formulations, starting from the chemical and physical properties of an active ingredient. The project was initiated in 1990 under the direction of Stricker (10), and in the interim has been extensively revised and enhanced (11, 12). Originally implemented using object-oriented C on a workstation, the system was recently upgraded using SMALLTALK V running under the Windows operating system on a personal computer.

Various forms of knowledge representation are used depending on the type of knowledge. Knowledge about objects (e.g., functional groups in the active ingredient, excipients, processes, etc.), their properties, and relationships are represented in frames using an object-oriented approach. Causal relationships are represented as rules, functional connections as formulas, and procedural knowledge as algorithms. The system currently has knowledge bases for aerosols, IV injection solutions, capsules (hard-shell powder), and tablets (direct compression). Each knowledge base incorporates information on all aspects of that dosage formulation (e.g., properties of the excipients to be added, compatibility, processing operations, packaging, and containers and storage conditions), with each aspect given a reliability factor (Sicherheitsfaktor) to indicate its accuracy/reliability. In the original version of the system values for each factor varied between 0 and 9 (10); however, values between 0 and 1 are used currently. The values are propagated using the arithmetic minimum rule and are not used for any decisions. They only serve as indicators of the accuracy/reliability of the knowledge.

The approach used in the system is the decomposition of the overall process into individual distinct development steps, with each step focusing on one problem associated

with a subset of its specifications or constraints for the formulation. A problem is considered solved if its predicted outcome satisfies its associated constraints. The problems are worked through in succession, with care being taken that any solution should not violate any constraints from previous steps. For simplicity, the developers imposed a predefined ordering onto the development steps, providing a backtracking mechanism to go back to a previous step or abort. This ordering minimizes dependency between development steps, which might result in an action causing a constraint previously satisfied to be violated. It also reduces the complexity of the problem to be solved.

The procedural model for one development step (e.g., for the choice of an excipient) is shown in Fig. 2. In any development step the first decision is whether or not to proceed with any action since the problem may have already been solved in previous steps. This is done by comparing the predicted or relevant properties of the current formulation with the initial specification. If the answer is negative, then further action is required; if positive, the problem has been solved. Once this has been decided, actions need to be defined and ranked. Knowledge for this is by means of hierarchically structural rule sets to form a decision tree where each leaf node consists of a subgroup of actions and each branch a rule. The rules in a rule set are ordered as the simplest and most straightforward way of handling conflict. Ranking numbers are used as the basis for the selection strategy. The concept is to search for the best alternative in terms of the highest score, these being the sum of the values of the constants to be resolved within the development step (e.g., solubility, compressibility, etc.) and their weights indicating their respective importance (11, 12). It should be noted that this method of ranking is different from that used originally by Stricker et al. (10), where the lowest score was regarded as the best alternative.

Once the action is selected, the decision is checked in terms of whether or not the measure has adverse effects on the active ingredient in terms of physical or chemical incompatibility. This does not necessarily mean a rejection of the action since knowledge of compatibility is generally of a qualitative nature with little quantification to denote severity. Hence, the overall decision is left to the user.

The amounts of excipients to be used are calculated by formulae with rule-based mechanisms for selecting the appropriate formula. A rule-based mechanism is also used to determine the appropriate function for predicting the property of the intermediate formulation. This is necessary for checking whether or not the original specifications have been satisfied and the action is successful. If

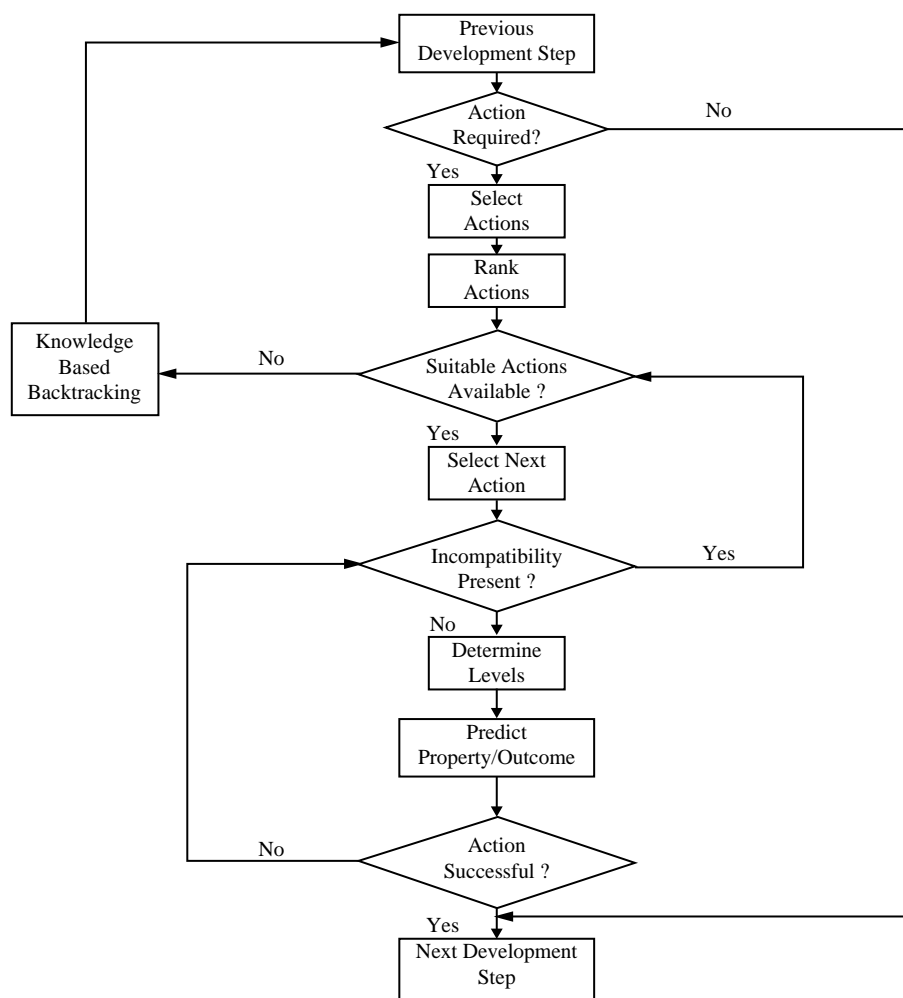


Fig. 2 Procedural model for a development step as used in GSH. (Modified after Ref. 12.)

negative, the chosen action is rejected and the next action in the ranking is tried. It is possible that none of the ranked actions is successful. If this is the case, then knowledge-based backtracking is used to determine which of the previous development steps to return to. Usually, background pharmaceutical knowledge is applied to determine why the current development step was unsuccessful, and a new development step that can solve the root cause of the problem is chosen.

In any expert system, explanations of the decisions made are important, both for instruction of the user and for maintenance of the system. Explanations in GSH take several forms. There are explanations for the development steps and their ordering provided by the designer of the knowledge base. Detailed explanations of the rules activated, formulae used, or individual scores of actions can be generated if required, and canned text and literature references are provided for general knowledge.

A simplified task structure for generating an iv injection solution is shown in Fig. 3. The input to the system includes all the known properties of the active ingredient to be formulated (e.g., solubility, stability, impurities, pK_a , presence/absence of functional groups, etc.) with user-defined labels that relate the specific drug property to the required product property. Use of the system results in the production of four packages—the product formulation, the production method, the recommended packaging and storage conditions, and predicted product properties. All the outputs are provided with reliability factors. An example for an IV injection solution of the cardiac drug digoxin is shown in Table 3, and an example for a hard gelatin capsule of the antifungal drug griseofulvin is shown in Table 4. Comparison of a 0.1 mg. commercial formulation of digoxin with that shown in Table 3 indicates that the same cosolvent is used (1,2-propandiol, presumably to enhance solubility) and ethanol. However,

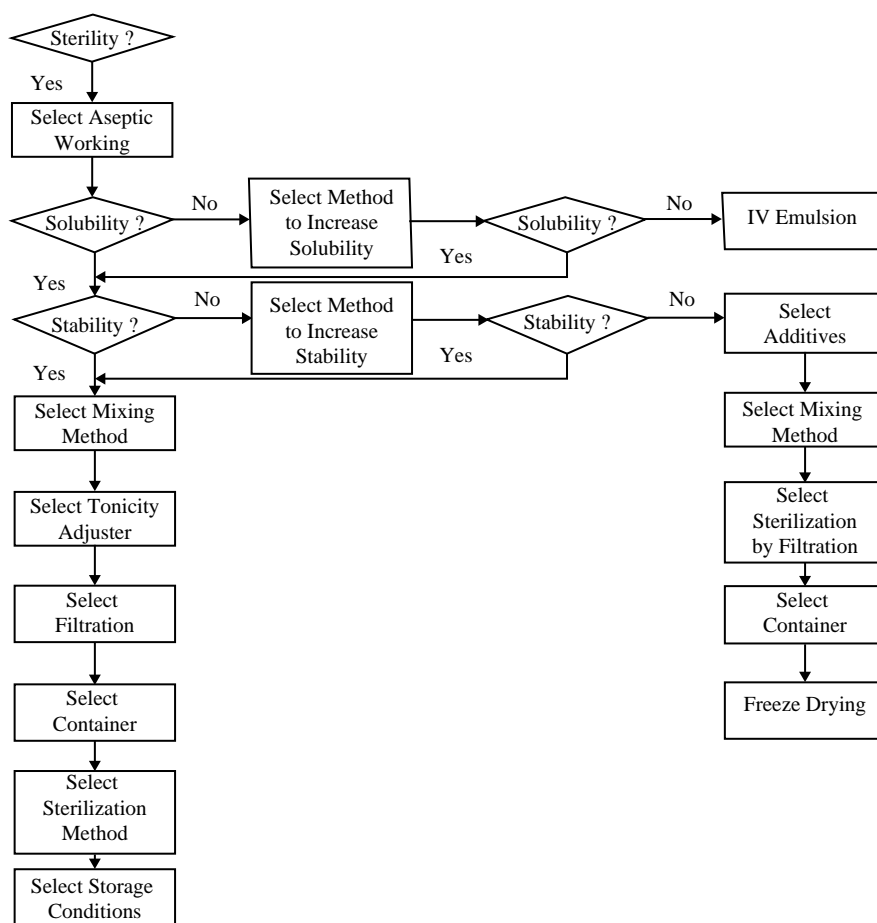


Fig. 3 Structure of the expert system described by Stricker et al. (10) for the formulation of intravenous injection solutions.

the commercial formulation is more sophisticated since it also contains a buffer (disodium hydrogen phosphate/citric acid).

At present, knowledge bases for aerosols, IV injection solutions, hard gelatin capsules, and direct compression tablets have been completed. Other knowledge bases for coated forms, granules, freeze-dried formulations, and pellets are in different stages of development. Trials have demonstrated that the system proposes formulations that are acceptable to formulators, and in December 1996, the system was first introduced for field trials in a pharmaceutical company.

The University of London/ Capsugel System

This system is designed to aid the formulation of hard gelatin capsules (13–15). It was developed as part of a Ph.D. program by Lai, Podczek, and Newton at the School of Pharmacy, University of London, and was

supported by Daumesnil of Capsugel, Switzerland, together with personnel from the University of Kyoto, Japan and the University of Maryland, United States. The system (Fig. 4) is unique in that its knowledge base is broad and contains the following:

1. A database of literature references associated with the formulation of hard gelatin capsules, which is permanently updated through monitoring of current literature.
2. Information on excipients used and their properties. This database currently contains information on 72 excipients and is frequently updated. Data can be retrieved via a menu.
3. An analysis of marketed formulations from Germany, Italy, Belgium, France, and the United States. This is used to identify trends in formulation and identify guidelines on the use of excipients. Currently, the database contains information on 750 formulations of 250 active ingredients. It is frequently updated and data can be retrieved via a menu.

Table 3 An example of an intravenous injection solution formulation for the cardiac drug digoxin as generated by GSH

<i>Formulation</i>			
Active	Digoxin	0.1 mg	
Solvent 1	1,2-Propandiol	0.5 mL	
	Water for injection to	1.0 mL	
<i>Packaging</i>			
Brown glass ampules			
<i>Product properties</i>			
Properties	Specification	Actual	R.F. ^a
Active (mg)	0.095	0.098	1.0
Volume (mL)	1.0	1.0	1.0
pH	3–9	7.0	1.0
Freezing point depression (°C)	0.5–20	13.2	0.8
Shelf life at 25°C (years)		5.0	1.0
Decomposition at 25°C (mol)		1.8	0.7

^aR.F., reliability factor.
(From Ref. 10.)

- Experience and nonproprietary knowledge obtained over a period of 18 months from a group of industrial experts from Europe, the United States, and Japan.
- Results from statistically designed experiments that identify factors that influence the filling and in vitro release performance of model active ingredients.

The system uses production rules with a decision tree implemented in C, coupled with a user interface through which the user can access both the databases and develop new formulations. To assist in collecting all necessary input data, a questionnaire has been designed. Called the expert system input package, it requires information on the physical properties of the active ingredient (e.g., dose, particle shape,

particle size, solubility, wettability, adhesion to metal surfaces, melting point, and bulk density), compatibility of the active ingredient with excipients (e.g., fillers/diluents, disintegrants, lubricants, glidants, and surfactants), and properties of excipients used by the user and manufacturing conditions (e.g., capsule sizes, fill weights, densification techniques, granulation techniques) used by the user.

From this data the system uses a variety of methods to evaluate and predict properties of mixtures of the active ingredient and the excipients. For instance, it uses Carr's compressibility index (16) to predict the flow properties that are used to give an indication of the ability to produce a uniform blend, and the Kawakita equation (17) to predict a maximum in the volume reduction of the powder

Table 4 An example of a hard gelatin capsule formulation for the antifungal drug griseofulvin as generated by GSH

<i>Formulation</i>		
Active	Griseofulvin	150.0 mg
Diluent	Microcrystalline cellulose (PH102)	199.2 mg
Lubricant	Magnesium stearate	3.5 mg
<i>Production process</i>		
High shear mixer for deagglomeration, premix, and main mix.		
Add lubricant, planetary mixer at 12 rpm for 3 min.		
Capsule-filling machine type 1.		
<i>Packaging</i>		
Foil blisters (PVC and aluminum foil)		

(From Ref. 12.)

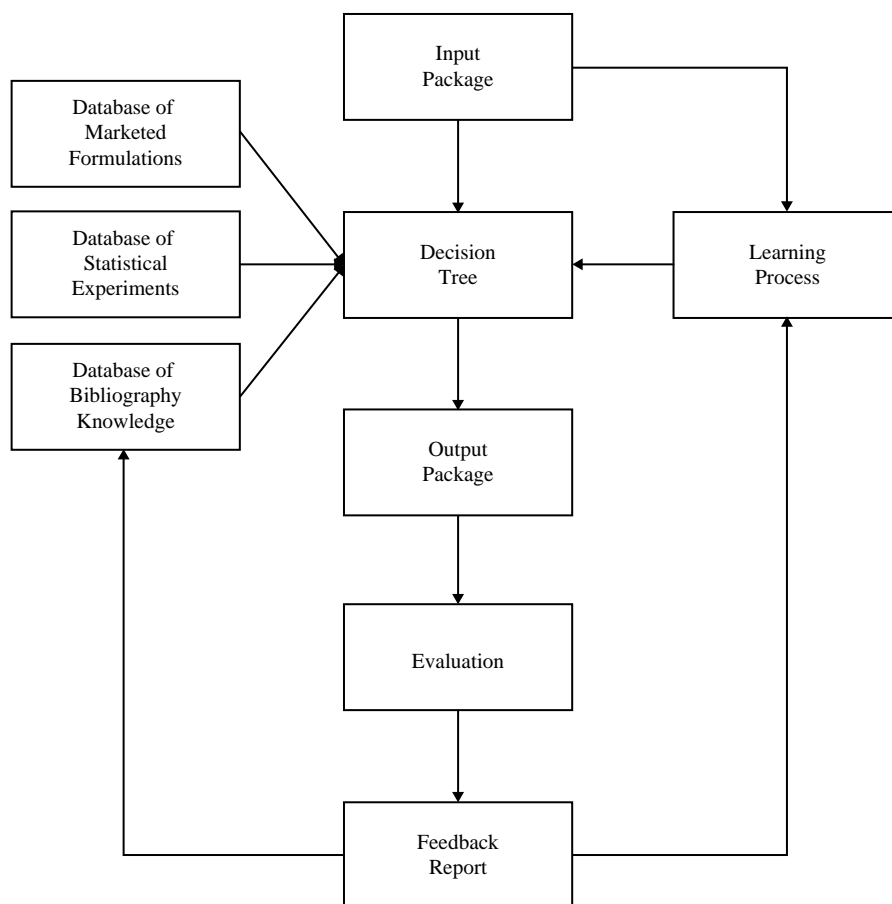


Fig. 4 Structure of the University of London/Capsugel system as described by Lai et al.

achievable by the application of low pressure. The packing properties are obviously important to give the volume that a given weight of powder occupies in order to indicate the size of capsule shell that can be used. When wet granulation is offered as the preferred method of densification, the system only offers advice on the choice of a granulating liquid and binder; no choice on the granulation procedure is offered.

The system provides an output package that includes a formulation (Table 5) with any necessary powder processing and filling conditions, the required capsule size, a statistical design to quantitatively optimize the formulation, specification of excipients used, recommended tests to validate the formulation, and a complete documentation of the decision process.

An interesting addition to the system is a semiautomatic learning tool. This monitors user habits and collects data about the use of excipients. Statistical analysis is performed on these data, allowing agreed alterations to be made to the database. The user is also asked to reply to a

questionnaire regarding the recommended formulation and its performance. The data are analyzed by the expert system founder group, and provide the background for further alterations and developments.

Field trials have proved that the system does provide reasonable formulations (18).

The Sanofi System

Personnel at the Sanofi Research Division of Philadelphia recently developed an expert system for the formulation of hard gelatin capsules based on specific preformulation data of the active ingredient (19). Using Formulogic, the system generates one first-pass clinical capsule formulation with as many subsequent formulations as desired to accommodate an experimental design. The latter are produced as a result of the user overruling decisions made by the system.

Knowledge acquisition is obtained by meetings between formulators, with a knowledge engineer present as a

Table 5 An example of a hard gelatin capsule formulation for a model drug as generated by the university of londoncapsugel system

<i>Tablet properties (inputs)</i>			
Dose (mg)	50.0		
Solubility	Insoluble		
Particle size (μm)	5.0		
Minimum bulk density (g ml^{-1})	0.4		
Tapped bulk density (g ml^{-1})	0.7		
Carr's compressibility (%)	42.857		
<i>Formulation</i>			
		wt%	mg/capsule
Active	Drug	39.7	50.0
Filler	Starch:lactose (1:2)	56.8	71.6
Disintegrant	Croscarmellose sodium	2.0	2.5
Lubricant	Magnesium stearate	1.0	1.3
Surfactant	Sodium lauryl sulfate	0.5	0.6
Capsule weight			126.0
Capsule size			No. 4

(From Refs. 13 and 14.)

consultant. Meetings are limited to 2 h, with minutes being taken and reviewed by all attendees. Meetings are specific to one topic defined in advance. A rapid prototyping approach is used to generate the expert system.

Knowledge in the system is structured using the strategies implemented in Formulogic, i.e., objects and production rules. The latter are as follows:

Tasks are scheduled dynamically. An outline of the task structure used is given in Fig. 5.

IF	(electrostatic properties of a drug are problematic)
THEN	(add glidant)
UNLESS	(glidant has already been added)

The user is first prompted to enter specified preformulation data on the active ingredient (e.g., acid stability, molecular weight, wettability, density, particle size, hygroscopicity, melting point, solubility, etc.) and known excipient incompatibilities together with the required dose. At the initial formulation task, the capsule size is selected together with the process and milling requirements. The excipient classes are selected, with some excipients being excluded, others prioritized, and their amounts determined. At the display reports task, three reports are provided, one providing the preformulation data as given, the second giving the recommended formulation, including the amounts of the excipients and processing/milling requirements, and the third providing the explanation of the decisions and reasoning used by the system. On the

first-pass through the system, the selection of the possible processing, milling, and excipient options are automatic. On subsequent passes, the selections are optional, allowing the user to generate a number of formulations.

An example of a formulation generated by this system, for the nonsteroidal anti-inflammatory drug naproxen, is given in Table 6. This example, as well as others, was considered acceptable by experienced formulators for manufacture and initial stability evaluation.

Unfortunately, the authors (19) do not provide any further details on the state of the system except to imply that formulation evaluation and preformulation tasks could be accommodated with the possible development for other formulation types such as tablets, liquids, and creams.

The Zeneca Systems

Work on expert systems within ICI/Zeneca Pharmaceuticals (now AstraZeneca) began in April 1988, with the initiation of a joint project between the Pharmaceutical and Corporate Management Services departments to investigate the use of knowledge-based techniques for the formulation of tablets. Since that time, work has proceeded with the successful development of expert systems for formulating tablets, parenterals, and tablet film coatings. All have been implemented using Formulogic from Logica UK Ltd., although elements of the system developed for

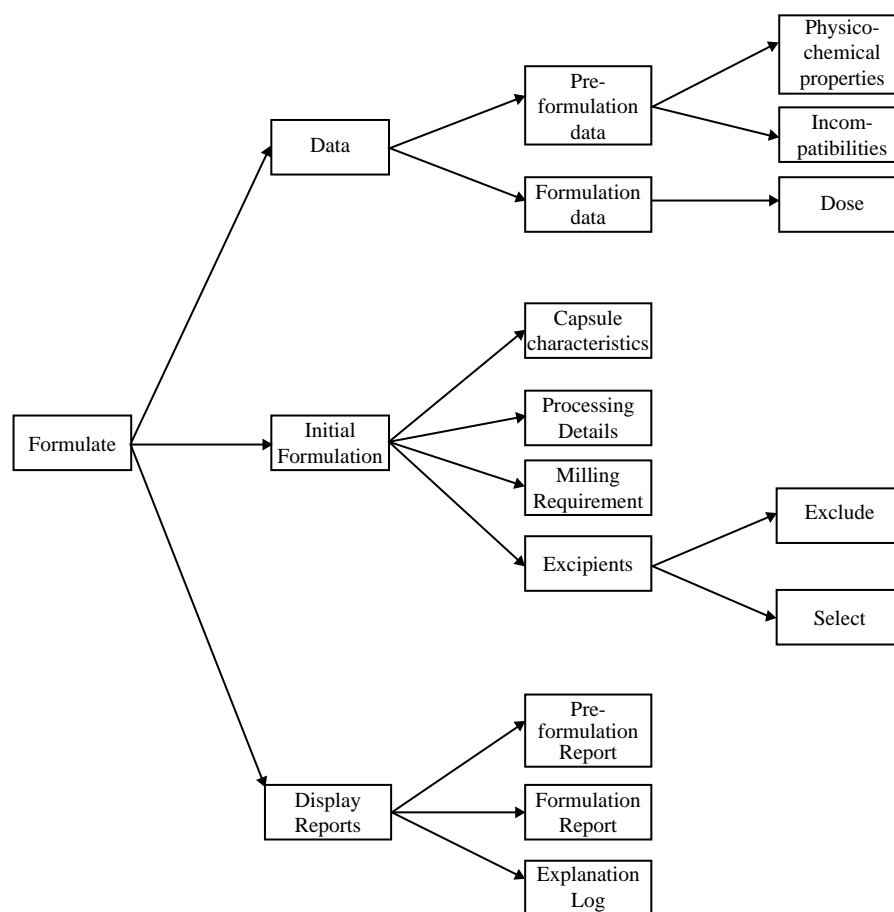


Fig. 5 Task structure for the formulation of hard gelatin capsules as used by Bateman et al. (19).

tablet film coatings were originally prototyped using a rule induction tool to produce decision trees.

Delivery of the first usable system for tablet formulation was in January 1989 (20, 21). All the knowledge was acquired from two primary experts in the field of tableting—one with extensive heuristic knowledge and the other with extensive research knowledge—and structured into Formulogic using specialist consultancy support. Consultancy time for the initial system was in the order of 30 man days, 20% of which was involved in three 2-day visits to the laboratories, incorporating three 90-min interviews with the experienced formulator plus members of the research group, the demonstration of prototype systems, and the validation of the previously acquired knowledge with the expert and other members of the department. Sixty percent of the time was involved in system development and 20% in writing the final report.

After commissioning and extensive demonstration to management and formulators throughout the company during 1989, the system was enhanced by the addition of a

link to a database in January 1990, and the installation of a formulation optimization routine in September 1990. A major revision was initiated in February 1991, following a significant change in formulation practice. Total consultancy time for these enhancements was of the order of 30 man-days. In August 1991, the system was completed and handed over to the formulators both in the United Kingdom and the United States.

The completed system is shown in Fig. 6 (20). It is divided into three stages: 1) the entry of the data, product specification, and strategy; 2) the identification of the initial formulation; and 3) the formulation optimization as a result of testing the initial formulation. The sequence is as follows:

1. The user enters all the relevant physical, chemical, and mechanical properties (e.g., solubility, wettability, compatibility with excipients, and deformation behavior) of the new active ingredient to be formulated into the database. The data may be

Table 6 An example of a hard gelatin capsule formulation for the antiinflammatory drug naproxen as generated by the system described by Bateman et al.

<i>Selected drug properties (inputs)</i>		
Molecular weight		230.26
Melting point (°C)		155
Solubility in water (mg ml ⁻¹)		0.01
Wettability		Poor
Water stability		Poor
Photostability		Poor
Susceptible to hydrolysis		No
Primary/secondary amines		No
Hygroscopicity		Class 1
Poured density (g cm ⁻³)		0.366
Electrostatic problems		No
Unmilled particle size (μm)		36
<i>Formulation</i>		
Active	Naproxen	100 mg
Diluent	Lactose (hydrous)	224 mg
Disintegrant	Microcrystalline cellulose (PH105)	60 mg
Surfactant	Sodium lauryl sulfate	4 mg
Lubricant	Talc	12 mg
<i>Production information</i>		
Milling	Jet milling of drug	
Capsule	Size 0 colored opaque	
Process	Direct blend	
<i>Explanation log</i>		
A colored opaque capsule used because drug is unstable to light.		
Drug requires milling as it has a medium particle size and is insoluble.		
A surfactant is required because drug has poor wettability.		

(From Ref. 19.)

numerical or symbolic (e.g., for solubility in water the data can be entered as mg ml⁻¹ or as the descriptors “soluble,” “partially soluble,” “insoluble,” etc. The data are obtained from a series of proprietary preformulation tests carried out on the active ingredient as received (i.e., 5 g of the drug milled to a specified particle size). These tests include excipient compatibility studies whereby the drug is mixed with the excipient and stored under specified conditions of temperature and humidity for one week, the proportion of drug remaining being analyzed by HPLC and expressed as a percentage. The deformation properties essential for the evaluation of compactibility are assessed using yield pressure and strain rate sensitivity measured via a compression simulator (22).

- The user enters the proposed dose of the active ingredient and a target tablet weight is calculated using both a formula determined from an extensive study of

previously successful formulations and certain rules governing minimum weights for ease of handling and maximum weights for ease of swallowing.

- The user selects a strategy dependent on the number of fillers (one or two).
- The system selects the filler(s), disintegrant, binder, surfactant, glidant, and lubricant, and their recommended concentrations based on a combination of algorithms, formulae, and mixture rules governing their compatibility and functionality. Tasks in this process are dynamically scheduled depending on the problem to be solved. If the system is unable to decide between two excipients, both of which satisfy all the embedded rules, then the user is asked to select a preference.
- The recommended initial formulation is displayed, including final tablet weight, recommended tablet diameter calculated compression properties, and all relevant data (Table 7). This is normally printed for inclusion in a laboratory notebook, file, etc. If required,

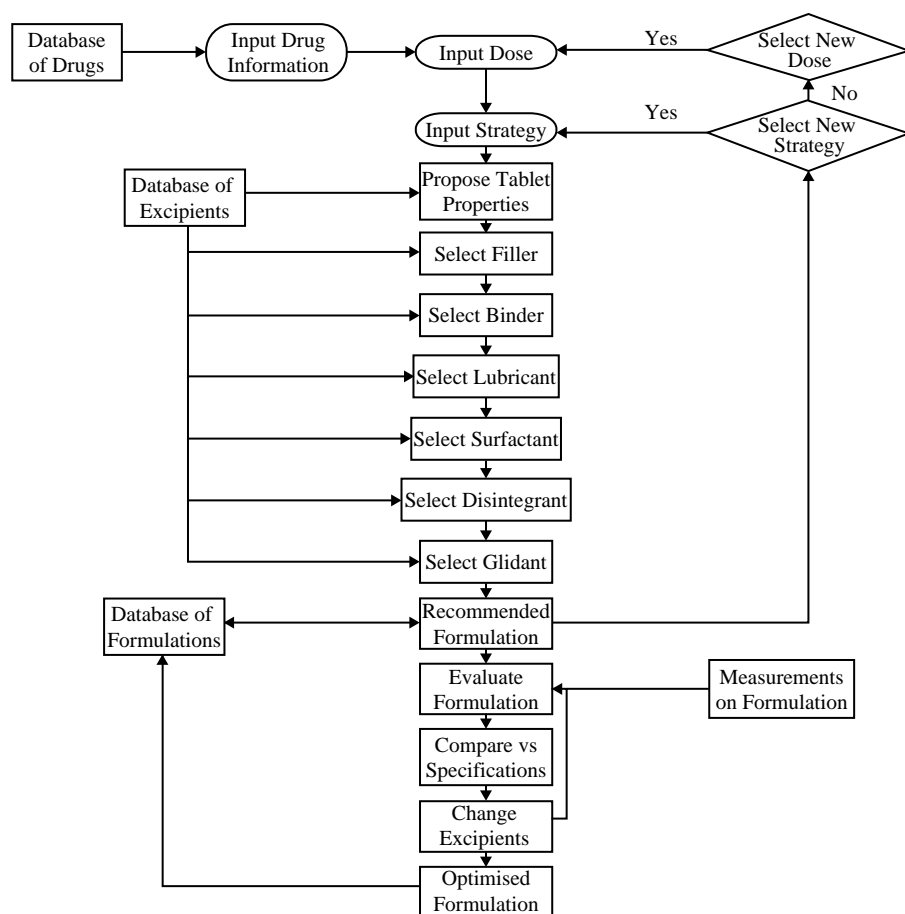


Fig. 6 Structure of the expert system developed by Rowe (20) for the formulation of tablets.

- the data may be stored in a database for future reference, necessary if the formulation optimization route is used.
6. The user enters results from testing tablets prepared using the initial formulation. These may include disintegration time, tablet strength, tablet weight variation, and presence of any defects (e.g., capping, lamination, etc.). The results are compared with specifications, and any problems identified are confirmed with the user. Recommendations for modifications to the formulation are then listed. This routine is fully interactive with the user, who is asked to confirm or contradict/change the advice given.
 7. After agreement is reached, the system modifies the formulation accordingly and displays it as described earlier. This routine may be used as many times as required; each time, the system iterates on the previously modified formulation.

Two “help” routines are embedded in the system, one to provide on-line help in the use of the system, the other to

provide an insight into the rationale behind adoption of the specific rules/formulae/algorithms used. The user is able to browse the knowledge base at will but is not able to edit it without privileged access. Explanations for any recommendations made by the system can be easily accessed, if required. Hypertext links are used throughout. Two screen images from the system are shown in Fig. 7 to illustrate the operation of the system.

The system is well used and is now an integral part of the development strategy for tablet formulation. To date, it has successfully generated formulations for more than 40 active drugs. In many cases, the initial formulations have been acknowledged as being on a par with those developed by expert formulators. Consequently, the formulation optimization routine is now considered redundant and is used very rarely.

Following the successful implementation of the tablet formulation expert system, a request was made for the development of a similar system for parenteral formulations. This project was initiated in April 1992, and

Table 7 Examples of tablet formulations for a model drug as generated by the system described by Rowe

<i>Drug properties (inputs)</i>			
Solubility (mg ml ⁻¹)		0.1	
Contact angle		82°	
Yield pressure (MPa)		50	
Strain rate sensitivity (%)		50	
+ Excipient compatibilities			
<i>Formulation</i>			
		Quantity (mg)	Quantity (mg)
Active	Drug A	50.0	150.0
Filler	Lactose monohydrate	166.9	—
Filler	Dicalcium phosphate dihydrate	—	165.7
Disintegrant	Croscarmellose sodium	4.8	7.0
Binder	Polyvinylpyrrolidone	4.8	—
Binder	Hydroxypropylmethyl cellulose	—	7.0
Surfactant	Sodium lauryl sulfate	0.7	1.1
Lubricant	Magnesium stearate	2.4	3.5
Tablet weight		230.0	335.0
		Predicted properties	Formulation
		50 mg	150 mg
Tablet diameter (mm)		8.0	10.0
Yield pressure (MPa)		139	238
Strain rate sensitivity (%)		20.8	5.1

(From Ref. 20.)

completed in August 1992 (23). The structure of the system is shown in Fig. 8. It is designed for formulating a parenteral for either clinical or toxicological studies in a variety of species (dog, man, mouse, primate, rabbit, or rat) by a variety of routes of administration (iv, intramuscular, subcutaneous, interperitoneal), supplied in either a single or multidose container. Knowledge was acquired from two domain experts using a series of interviews. The sequence is as follows:

1. The user enters all known data on the solubility (aqueous and nonaqueous), stability in specified solutions, compatibility, pK_a , and molecular properties of the active ingredient (molecular weight, log P , etc.). As with the system for tablet formulation, the data may be numerical or symbolic. All relevant properties of additives used in parenteral formulation (e.g., buffers, antioxidants, chelating agents, antimicrobials, and tonicity adjusters) are present in the knowledge base.
2. The selection first attempts to optimize the solubility/stability of the active drug at a range of pH using a variety of formulae and algorithms together with specific rules before selecting a buffer to achieve that

pH. If problems still exist with solubility and stability, then formulation variants (e.g., oil-based or emulsion formulations—not implemented in the present system) are recommended.

3. The system then selects additives, depending on the specification required (e.g., an antimicrobial will only be added if a multidose container is specified or a tonicity adjuster will only be added if the solution is hypotonic). The selection strategy is generally on the basis of ranking with some specific rules.
4. The recommended formulation is displayed with all concentrations of the chosen ingredients expressed as percentage weight by volume (w/v) together with the calculated tonicity, proposed storage conditions, and predicted shelf life (Table 8). Specific observations on the sensitivity of the formulation to metals, hydrolysis, light, and oxygen also are included. This is normally printed for inclusion in a laboratory notebook, file, etc. If required, the formulation may be stored in a database for future reference.

As with the system developed for tablet formulations, this system contains extensive “Help” routines. No

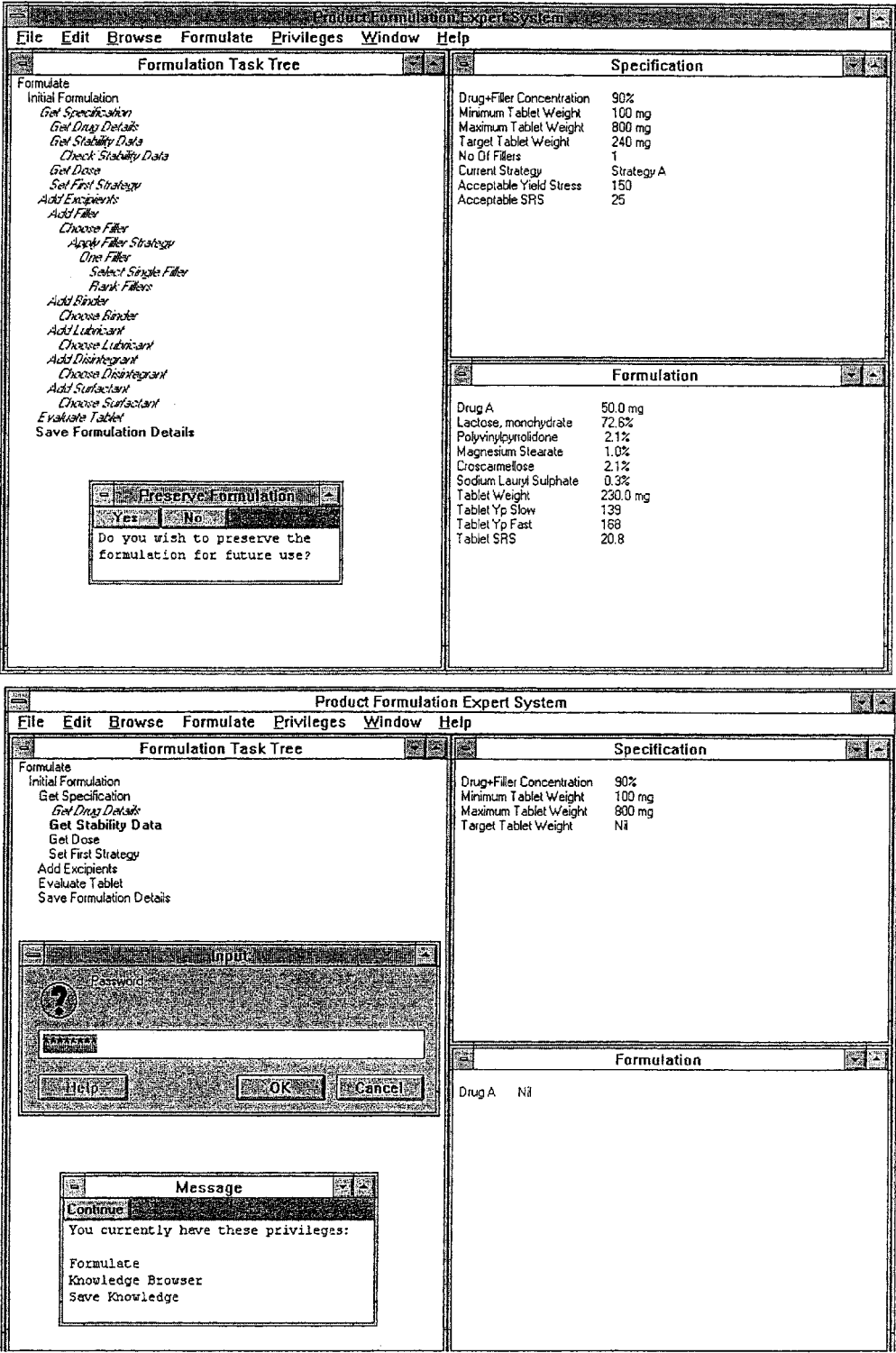


Fig. 7 Screen images for the tablet formulation expert system as described by Rowe (20). (a) Shows user interface with windows for the formulation task tree, specification, formulation, and current task. Tasks are displayed in various formats—current task in highlighted or bold text, completed in italics, and future in standard text. (b) Shows the security of the system. The input of a password displays the user privileges.

Table 8 An example of a formulation of an intravenous injection for clinical trials in man as generated by the system described by Rowe et al.

<i>Drug properties (inputs)</i>		
Drug type		Acid
pK _a		4.5, 3.5
Molecular weight		458.5
Solubility (mg ml ⁻¹)	pH 3.0	0.5
	pH 4.0	1.5
	pH 5.0	7.0
	pH 7.0	40.0
Sensitivity	Light, metal, oxygen	
<i>Formulation</i>		Quantity (% w/v)
Active	Drug (10 mg/ml)	1.00
Buffer	Disodium hydrogen phosphate anhydrous	0.87
Buffer	Hydrochloric acid	q.s.
Chelating agent	Disodium edetate	0.02
	Water for injection to	100.00
<i>Predicted solution properties</i>		
pH	7.4	
Tonicity	Hypertonic (1.6)	
Storage temperature (°C)	25	
Atmosphere for filling	Nitrogen	
Shelf life (years)	>5	

(From Ref. 23.)

formulation optimization routines are included, although a routine to develop a placebo formulation to match the active formulation is included. The system is used to recommend parenteral formulations for a wide range of investigational drugs.

Work on expert systems in the specific domain of tablet film coating was initiated in April 1990, using a rule induction tool in order to develop a system for the identification and solution of defects in film-coated tablets. Although not strictly a formulation expert system, the developed system did contain knowledge whereby a given formulation known to cause defects could be modified to provide a solution. The completed system described by Rowe and Upjohn (24, 25) is a perfect illustration of fault diagnosis with a rule-based decision tree including both forward and backward chaining. Total development time was approximately 1 man-month using both documented knowledge (26) and expert assistance.

The system (Fig. 9) is divided into three parts: identification, solution, and information/references. In the first part, a question and answer routine is used to ascertain the correct identification of the defect. The decision tree used for this process is shown in Fig. 10.

At this point, the user is asked to confirm the decision by comparing the defect with a picture or photographs stored in the database. In the second part, the user is asked to enter all relevant process conditions and formulation details regarding the best way of solving the defect. This may be a change in the process conditions, as in the case of defects occurring with an already registered formulation, or a change in the formulation, as in the case of defects occurring at the formulation development stage. In the third part, the user is able to access data and knowledge regarding each defect. These are in the form of notes, photographs, and literature references connected by hypertext links.

In 1994, due to the successful implementation of both this system and that used to formulate tablets, it was decided to initiate work on a new system that would recommend film-coating formulations for the generated tablet formulations, combined with a reformulation routine based on the film defect diagnosis system (27). The structure of the new system is shown in Fig. 11. The knowledge for the system was acquired by interviewing two domain experts, one with extensive heuristic knowledge and the other with extensive research knowledge. The sequence is as follows: unital)

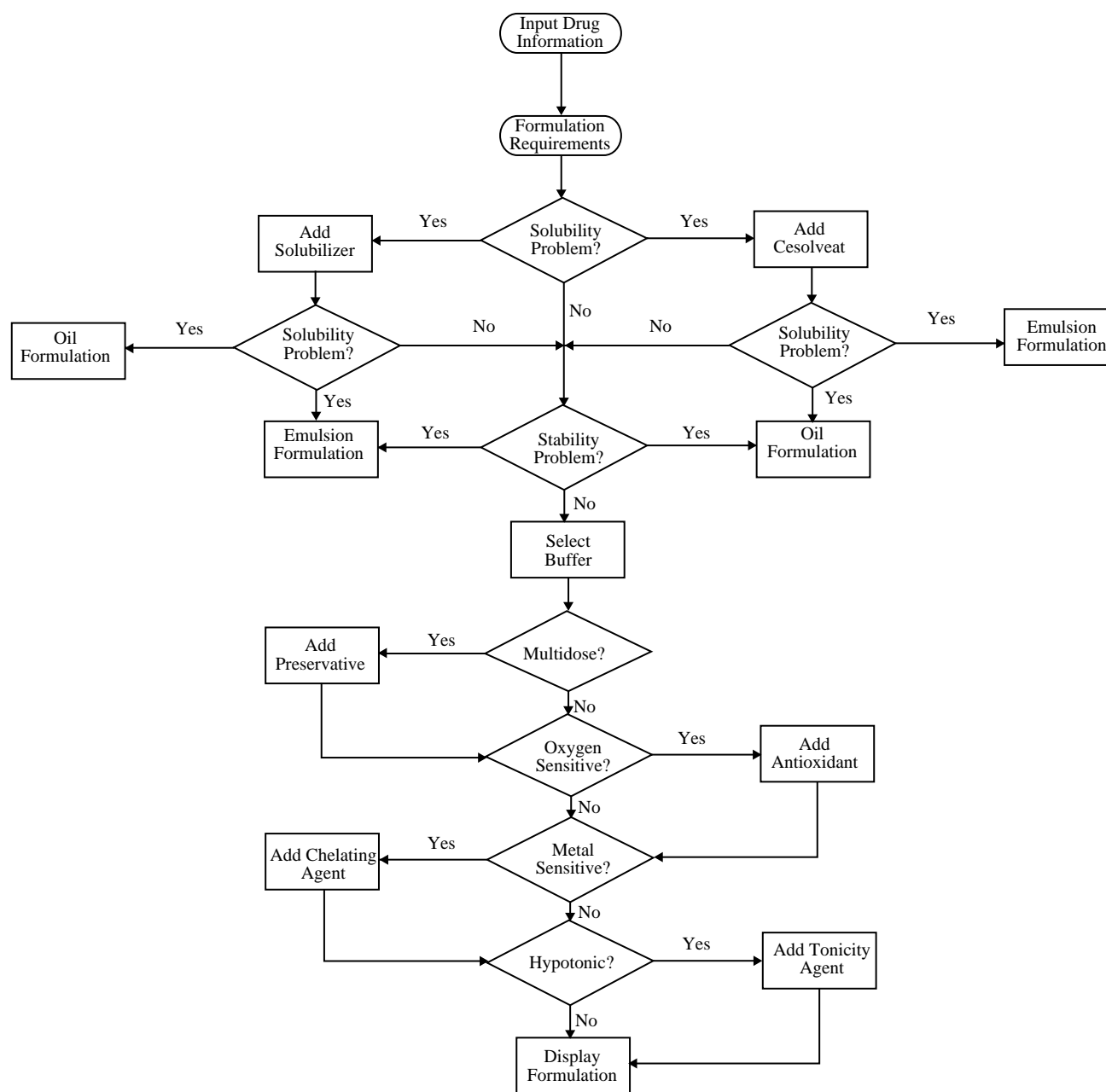


Fig. 8 Structure of the expert system developed by Rowe et al. (23) for the formulation of parenterals.

1. The user enters details of the tablet formulation (e.g., dose of active ingredient and all excipients) together with all tablet properties (e.g., diameter, thickness, strength, friability, color, shape, and the presence/ absence of intagliations).
2. The user enters specifications for the film coating formulation (i.e., immediate release/controlled release, enterosoluble, white or colored).
3. The system first checks that there are no inconsistencies between the input details and the required specification (e.g., tablets with high friabilities are extremely difficult to film coat). If positive, a warning is displayed.
4. The system calculates the surface area of the tablet and selects the required polymer at the recommended level to form a film of reasonable thickness.

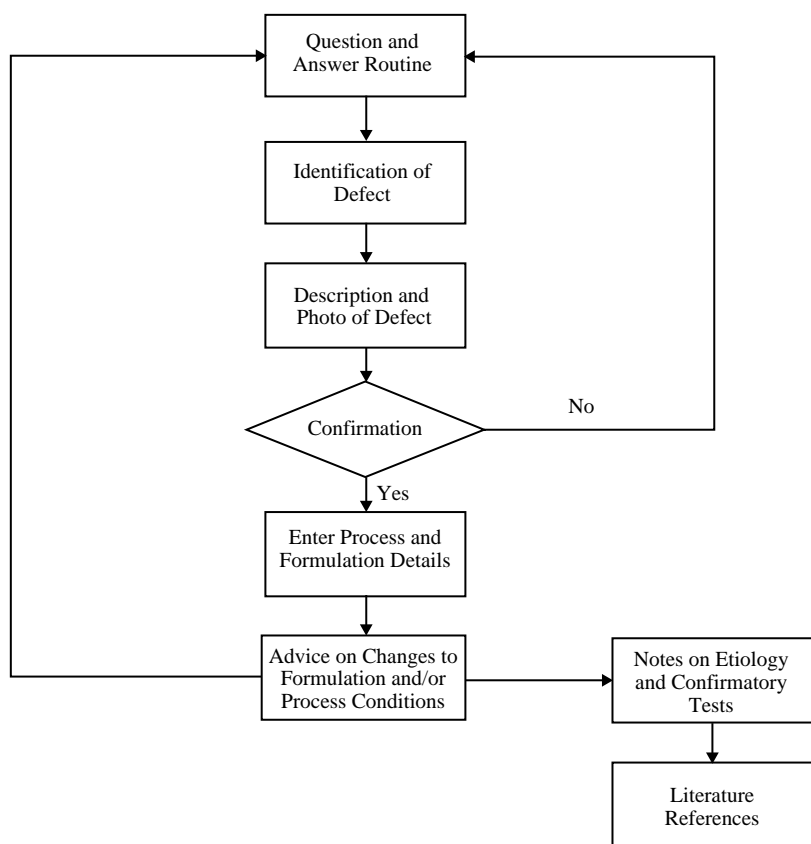


Fig. 9 Structure of the expert system for the identification and solution of film coating defects as described by Rowe and Upjohn (24, 25).

Table 9 An example of a formulation of a white film coating for a tablet of a model drug as generated by the system described by Rowe et al.

<i>Tablet properties (inputs)</i>		
Tablet core formulation	Drug A 50 mg	
Punch shape	Normal concave	
Weight (mg)	230	
Diameter (mm)	8.0	
Thickness (mm)	3.5	
Surface area (cm ²)	1.49	
<i>Formulation</i>	mg/tab	mg/cm ²
Polymer Hydroxypropyl methylcellulose (6 cps)	6.14	4.12
Plasticizer Polyethylene glycol (PEG 400)	1.23	0.82
Pigment Titanium dioxide (Anatase)	5.63	3.78
<i>Predicted film properties</i>		
Thickness (μm)	45	
Opacity (%)	94.9	
Crack velocity (ms ⁻¹)	5.71	

(From Ref. 27.)

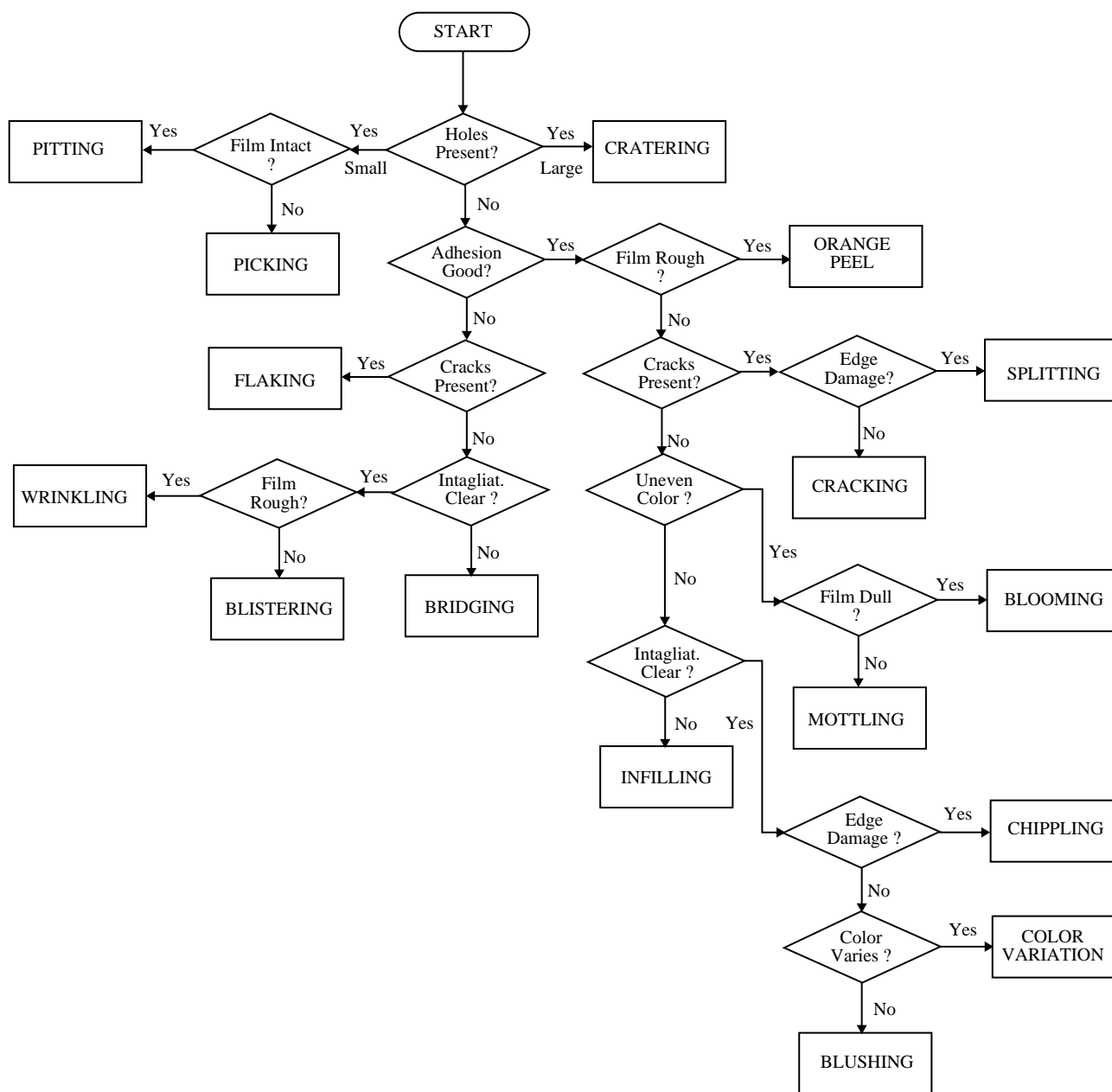


Fig. 10 A decision tree for the identification of film defects on film-coated tablets. (From Ref. 25).

5. The system selects a plasticizer and checks that there are no stability/compatibility problems. If positive, the user is asked to select an alternative plasticizer.
6. The system defines the target opacity of the film coating and decides if an opaque coating is required. The opacity is assessed by means of the contrast ratio defined as the ratio of the reflectance of the film when viewed with a black background to that viewed with a white background (the higher the value the more opaque the film) (28, 29). If positive and the

specification has been set as white, the system uses specifically developed algorithms (30, 31) to calculate if the target specification can be achieved within certain predefined formulation limits of the volume concentration of titanium dioxide and film thickness. If negative, the user is provided with a series of options to continue with the predefined limits, change the limits, or select a colored film coating.

7. The system selects the pigments to achieve the target specification and determines the amount of water

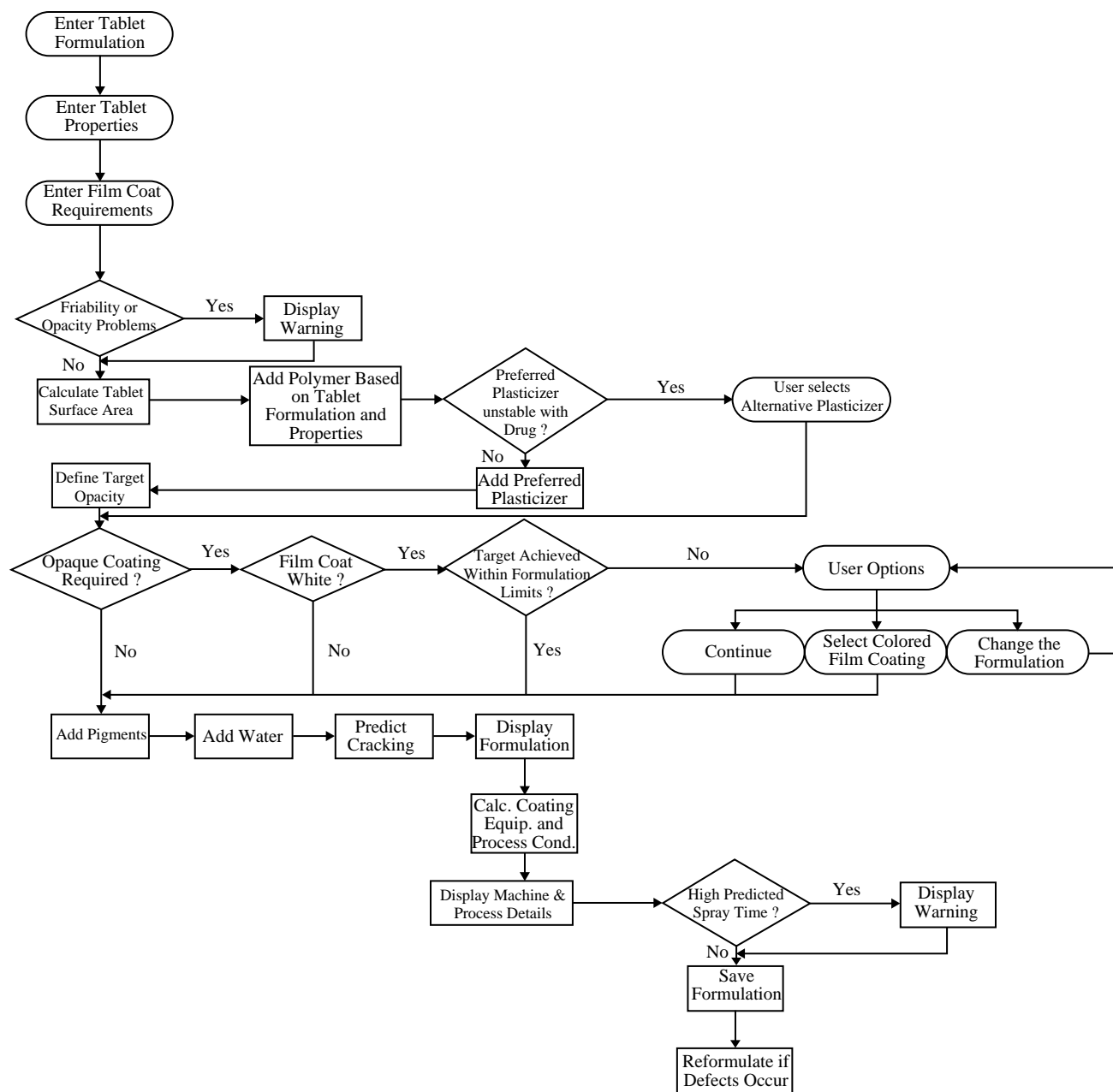


Fig. 11 Structure of the expert system developed by Rowe et al. (27) for the formulation of tablet film coatings.

(the system has been developed for aqueous film coating only).

8. The system accesses a simulation program written in C to predict the cracking propensity of the film formulation (32–34).
9. The recommended formulation is displayed (Table 9) and includes predicted film thickness, opacity, and cracking propensity. Standard machine settings and process details also are displayed with warnings

if the total spray time is judged to be excessive. This is normally printed for inclusion in a laboratory notebook or file. If required and in particular if reformulation is likely to be necessary, the data may be stored in a database for future reference.

10. If a reformulation is necessary due to the presence of defects after coating, then the system uses a modified form of the defect diagnosis system to recommend changes in the formulation and/or process conditions.

This system has proved successful in initial trials, especially in the formulation of opaque films for drugs that are either unstable to light or colored, producing mottled tablets. The calculations concerning the achievement of the target opacity within predefined limits have enabled formulators to make informed decisions regarding the use of white or colored film coatings. The system is now an integral part of the development strategy for film-coated tablets and now has a common database with the tablet formulation system.

BENEFITS

Expert systems have many benefits. These include the following:

1. *Knowledge protection and availability.* The existence of a coherent, durable knowledge base not affected by staff turnover (20). The developers of the University of London/Capsugel system have reported the benefit of being able to use knowledge from experts from many industrial companies in Europe, the United States, and Japan (14, 15). The developers of both the Cadila and the Sanofi systems have reported the benefit of the prompt availability of information and the rapid access to physical chemical data of both drugs and excipients, reducing the time spent searching the literature (9, 19).
2. *Consistency.* All systems generate robust formulations with increased certainty and consistency. This is seen as a distinct benefit where regulatory issues are important.
3. *Training aid.* All systems have been used to provide training for both novice and experienced formulators. The developers of the SOLTAN system have stated that experienced formulators use their expert systems to expose themselves to new raw material combinations with which they are not familiar. Bateman et al. (19) suggested the documentation used in the development of the Sanofi system be adapted to train novice formulators.
4. *Speed of development.* Reduction in the duration of the formulation process has been reported by many (8, 9, 20). Wood (8) reported that formulators who use SOLTAN can produce a formulation in 20 min that might otherwise have taken 2 days. Ramani et al. (9) reported a 35% reduction in the total time needed to develop a new tablet formulation.
5. *Cost savings.* Cost savings can be achieved not only by reducing the development time but also by the more effective use of materials, especially if material cost and controls are included in the system. Ramani et al. (9) reported that use of their system has been a benefit in planning the purchase and stocking of excipients.

The developers of SOLTAN reported that formulations generated by their system are cost effective not only for savings in raw material costs but also because fewer numbers of ingredients are used as compared to traditional formulations. Several users have also reported a decrease in the size of raw material inventories since their expert systems only use those materials specified in the database.

6. *Freeing experts.* The implementation of expert systems in product formulation has inevitably allowed expert formulators to devote more time to innovation (8, 20). The developers of the SOLTAN system reported that the time saved using their expert system typically releases about 30 days of formulating time per year per formulator. Of course, experienced formulators originally involved in training also will have more time to devote to innovation.
7. *Improved communication.* Rowe (20) reported that expert systems in his company have provided a common platform from which to discuss and manage changes in working practice and to identify those critical areas requiring research and/or rationalization. The developers of SOLTAN reported that use of their system has made them scrutinize the way in which they formulated products, highlighting shortfalls from the ideal. They also report that they have discovered previously unknown relationships between ingredients and properties in their products. The benefit of an expert system in promoting discussion also was reported by Bateman et al. (19).

Of all the systems in product formulation, only one has provided costings and undertaken a cost benefit analysis. The developers of SOLTAN estimated the overall cost of their system to be £10,400 for hardware and software, £6000 for consultancy, and £9000 for expert's time, making a total of £25,400. Annual cost savings in the region of £200,000 were reported, delivering a payback of approximately 3 months.

It is interesting to note that where expert systems have been implemented in product formulation, early skepticism among potential users has generally changed to a mood of enthusiastic participation. It is unlikely that expert systems will ever replace expert formulators, but as a decision support tool they are invaluable, delivering many measurable and intangible benefits.

CONCLUSION

Expert systems have been developed by a number of pharmaceutical companies and academic institutes in

order to cover the most common formulation types. Only those that have been mentioned in the open literature have been discussed, although it is generally known that SmithKline Beecham, Glaxo Wellcome, Eli Lilly, and Pfizer also have developed systems. It is possible that many more systems exist, but reticence with regard to publication abounds, and it is difficult to estimate exactly the number developed.

ACKNOWLEDGMENTS

This article has been abstracted from *Intelligent Software for Product Formulation* by R.C. Rowe and R.J. Roberts (1998), with permission from Taylor and Francis Ltd., London.

REFERENCES

1. Bold, K. Expertensysteme Unterstützen Bei der Produktformulierung. *Chem. Ztg.* **1989**, *113*, 343–346.
2. Sell, P.S. *Expert Systems—A Practical Introduction*; Camelot Press: Basingstoke, 1985.
3. Partridge, D.; Hussein, Ch.M. *Knowledge-Based Information Systems*; McGraw Hill: New York, 1994.
4. Turner, J. Manufacturing Intelligence. **1991**, *8*, 12–14.
5. Bentley, P. Product Formulation Expert System. *Intelligent Software for Product Formulation*; Rowe, R.C., Roberts, R.J., Eds.; Taylor and Francis: London, 1998; 27–41.
6. Bradshaw, D. The Computer Learns from the Experts. *Financial Times London* **April 27th 1989**, 26.
7. Walko, J.Z. Turning Dalton's Theory into Practice. *Innovation* **1989**, *18*, 24.
8. Wood, M. Expert Systems Save Formulation Time. *Lab-Equipment Digest* **Dec. 1991**, 17–19.
9. Ramani, K.V.; Patel, M.R.; Patel, S.K. An Expert System for Drug Preformulation in a Pharmaceutical Company. *Interfaces* **1992**, *22* (2), 101–108.
10. Stricker, H.; Haux, R.; Wetter, T.; Mann, G.; Oberhammer, L.; Flister, J.; Fuchs, S.; Schmelter, V. Das Galenische Entwicklungs-System Heidelberg. *Pharm. Ind.* **1991**, *53*, 571–578.
11. Stricker, H.; Fuchs, S.; Haux, R.; Rossler, R.; Rupprecht, B.; Schmelter, V.; Wiegel, S. Das Galenische Entwicklungs-System Heidelberg-Systematische Rezepturentwicklung. *Pharm. Ind.* **1994**, *56*, 641–647.
12. Frank, J.; Rupprecht, B.; Schmelter, V. Knowledge-based Assistance for the Development of Drugs. *IEEE Expert* **1997**, *12* (1), 40–48.
13. Lai, S.; Podczek, F.; Newton, J.M.; Daumesnil, R. An Expert System for the Development of Powder Filled Hard Gelatin Capsule Formulations. *Pharm. Res.* **1995**, *12* (9), S150.
14. Lai, S.; Podczek, F.; Newton, J.M.; Daumesnil, R. An Expert System to Aid the Development of Capsule Formulations. *Tech. Eur.* **1996**, *8* (9), 60–68.
15. Newton, J.M.; Podczek, F.; Lai, S.; Daumesnil, R. The Design of an Expert System to Aid the Development of Capsule Formulations. *Formulation Design of Oral Dosage Forms*; Hashida, M. Ed.; Yakugyo-Jiho: 1998, 236–244.
16. Carr, R. Evaluating Flow Properties of Solids. *Chem. Engineer* **1965**, *18*, 163–168.
17. Ludde, K.H.; Kawakita, D. Die Pulverkompression. *Pharmazie* **1996**, *21*, 393–403.
18. Kashiara, T.; Yoshioka, M. Assessment in Japanese Focus Group. *Formulation Design of Oral Dosage Forms*; Hashida, M., Ed.; Yakugyo-Jiho: 1998; 244–253.
19. Bateman, S.D.; Verlin, J.; Russo, M.; Guillot, M.; Laughlin, S.M. The Development and Validation of a Capsule Formulation Knowledge-Based System. *Pharm. Technol.* **1996**, *20* (3), 174–184.
20. Rowe, R.C. An Expert System for the Formulation of Pharmaceutical Tablets. *Manufacturing Intelligence* **1993**, *14*, 13–15.
21. Rowe, R.C. Expert Systems in Solid Dosage Development. *Pharm. Ind.* **1993**, *55*, 1040–1045.
22. Rowe, R.C.; Roberts, R.J. The Mechanical Properties of Powders. *Advances in Pharmaceutical Sciences*; Ganderton, D., Jones, T., McGinity, J., Eds.; Academic Press: New York, 1995; 7, 1–62.
23. Rowe, R.C.; Wakerly, M.G.; Roberts, R.J.; Grundy, R.U.; Upjohn, N.G. Expert System for Parenteral Development. *J. Pharm. Sci. Technol.* **1995**, *49*, 257–261.
24. Rowe, R.C.; Upjohn, N.G. An Expert System for Identifying and Solving Defects on Film-Coated Tablets. *Manufacturing Intelligence* **1992**, *12*, 12–13.
25. Rowe, R.C.; Upjohn, N.G. An Expert System for The Identification and Solution of Film Coating Defects. *Pharm. Tech. Int.* **1993**, *5* (3), 34–38.
26. Rowe, R.C. Defects in Film Coated Tablets—Aetiology and Solution. *Advances in Pharmaceutical Sciences*; Ganderton, D., Jones, T., Eds.; Academic Press: New York, 1992; 6, 65–100.
27. Rowe, R.C.; Hall, J.; Roberts, R.J. Film Coating Formulation Using an Expert System. *Pharm. Tech. Eur.* **1998**, *10* (10), 72–82.
28. Rowe, R.C. The Opacity of Tablet Film Coatings. *J. Pharm. Pharmac.* **1984**, *36*, 569–572.
29. Rowe, R.C. Quantitative Opacity Measurements on Tablet Film Coatings Containing Titanium Dioxide. *Int. J. Pharm.* **1984**, *22*, 17–23.
30. Rowe, R.C. Knowledge Representation in the Prediction of the Opacity of Tablet Film Coatings Containing Titanium Dioxide. *Eur. J. Pharm. Biopharm.* **1995**, *41*, 215–218.
31. Rowe, R.C. Predicting Film Thickness on Film Coated Tablets. *Int. J. Pharm.* **1996**, *133*, 253–256.
32. Rowe, R.C.; Roberts, R.J. Simulation of Crack Propagation in Tablet Film Coatings Containing Pigments. *Int. J. Pharm.* **1992**, *78*, 49–57.
33. Rowe, R.C.; Roberts, R.J. The Effect of Some Formulation Variables on Crack Propagation in Pigmented Tablet Film Coatings Using Computer Simulation. *Int. J. Pharm.* **1992**, *86*, 49–58.
34. Rowe, R.C.; Rowe, M.D.; Roberts, R.J. Formulating Film Coatings with the Aid of Computer Simulations. *Pharm. Technol.* **1994**, *18* (10), 132–139.

EXCIPIENTS—THEIR ROLE IN PARENTERAL DOSAGE FORMS

Sandeep Nema

Pharmacia Corporation, Skokie, Illinois, U.S.A.

Ron J. Brendel

Mallinckrodt, Inc., St. Louis, Missouri, U.S.A.

Richard J. Washkuhn

Lexington, Kentucky, U.S.A.

INTRODUCTION

The term pharmaceutical excipient or additive denotes compounds that are added to the finished drug product for a variety of reasons. Most often excipients are major components of the drug product, with the active drug molecule present in a small percentage. Excipients also have been referred to as inactive or inert ingredients to distinguish them from the active pharmaceutical ingredients. However, in many instances excipients may not be as inert as some scientists believe. Several countries have restrictions on the type or the amount of excipient that can be included in the formulation of parenteral drug products due to safety issues. For example, in Japan, amino mercuric chloride, or thimerosal use is prohibited, even though these excipients are present in several products in the United States.

As defined in the *European Pharmacopoeia* (EP) 1997 and the *British Pharmacopoeia* (BP) 1999, "Parenteral preparations are sterile preparations intended for administration by injection, infusion, or implantation into the human or animal body" (1, 2). However, for the purposes of this article, only sterile preparations for administration by injection or infusion into the human body will be surveyed. Injectable products require a unique formulation strategy. The formulated product has to be sterile, pyrogen free, and in the case of solutions, free of particulate matter. No coloring agent may be added solely for the purpose of coloring the parenteral preparation. Preferably, the formulation should be isotonic, and depending on the route of administration, certain excipients may not be allowed. For a given drug, the risk of an adverse event may be higher or the effects may be difficult to reverse if it is administered as an injection versus a nonparenteral route, since the injected drug bypasses natural defense barriers. The requirement for sterility demands that the excipient be able to withstand terminal sterilization or aseptic processing. These factors limit the choice of excipients available to the formulator.

Generally, a knowledge of which excipients have been deemed safe by the Food and Drug Administration (FDA) or are already present in a marketed product provides increased assurance to the formulator that these excipients will probably be safe for their new drug product. However, there is no guarantee that the new drug product will be safe as excipients are combined with other additives and/or with a new drug molecule, creating unforeseen potentiation or synergistic toxic effects. Regulatory bodies may view favorably an excipient previously approved in an injectable dosage form and will frequently require less safety data. A new additive in a formulated product will always require additional studies adding to the cost and timeline of product development.

In Japan, if the drug product contains an excipient with no precedence of use in that country, then the quality and safety attributes of the excipient must be evaluated by the Subcommittee on Pharmaceutical Excipients of the Central Pharmaceutical Affairs Council concurrently with the evaluation of the drug product application (3). Precedence of use means that the excipient has been used in a drug product in Japan, and will be administered via the same route and in a dose level equal to or greater than the excipient in question in the new application.

This chapter is a comprehensive review of the excipients included in the injectable products marketed in the United States, Europe, and Japan. A review of the literature indicates that only a few articles that specifically deal with the selection of parenteral excipients have been published (4–9). However, excipients included in other sterile dosage forms not administered parenterally, such as solutions for irrigation, ophthalmic or otic drops, and ointments, will not be covered.

Several sources of information were used to summarize the information compiled in this chapter (4–7, 10–14). Formulation information on the commercially available injectable products was entered in a worksheet. Tables presented in this chapter are condensed from this

worksheet. Each table is categorized based on the primary function of the excipient in the formulation. For example, citrates are classified as buffers and not as chelating agents, and ascorbates are categorized as antioxidants, although they can serve as buffers. This classification system minimizes redundancy and provides a reader-friendly format. The concentration of excipients is listed as percent weight by volume (w/v) or volume by volume (vol%). If the product was listed as lyophilized or powder, the percentages were derived based on the reconstitution volume commonly used. The tables list the range of concentration and examples of products containing the excipient, especially those that use an extremely low or high concentration.

TYPES OF EXCIPIENTS

Solvents and Cosolvents

Table 1 lists solvents and cosolvents used in parenteral products. Water for injection is the most common solvent but may be combined or substituted with a cosolvent to improve the solubility or stability of drugs (15, 16). The dielectric constant and solubility parameters are among the most common polarity indices used for solvent blending (17, 18). Ethanol and propylene glycol are used either

alone or in combination with other solvents in more than 50% of parenteral cosolvent systems. Surprisingly, propylene glycol is used more often than polyethylene glycols (PEGs) in spite of its higher myotoxicity and hemolyzing effects (19–22). The hemolytic potential of cosolvents is as follows (19):

Dimethyl acetamide < PGE400 < Ethanol
< Propylene glycol
< Dimethyloxide

It is possible that the presence of residual peroxide from the bleaching of PEG or the generation of peroxides in PEG may result in the degradation of the drug in the cosolvent system. It is important to use unbleached and/or peroxide-free PEGs in the formulation.

Oils such as safflower and soybean are used in total parenteral nutrition products, where they serve as a fat source and as carriers for fat-soluble vitamins. The *U.S. Pharmacopeia* (USP) requirement for injectable oils is as follows:

- A. Fixed oils (of vegetable origin)
- Saponification value (185–200)
 - Iodine number (79–128). (*The Japanese Pharmacopoeia* (JP) recommends value between 79–137.)

Table 1 Solvents and cosolvents

Excipient	Frequency	Range	Example
Almond oil	1	ND	Poison Ivy Extract (Parke Davis)
Benzyl benzoate	3	20–44.7% w/v	Delestrogen [®] 40 mg/ml (Bristol Myers) 44.7% w/v
Castor oil	1	ND	Delestrogen [®] 20 mg/ml (Bristol Myers) 44.7% w/v
Cottonseed oil	2	73.6–87.4% w/v	Depo-Testadiol [®] (Upjohn) 87.4% w/v
<i>N,N</i> -Dimethylacetamide	2	6–33% w/v	Busulfex [®] (Orphan Medical) 33%
Ethanol	26	0.6–100%	Prograf (Fujisawa) 80 vol%, Alprostadil (Bedford Lab) 100%
Glycerin (glycerol)	12	1.6–70% w/v	Multitest CMI [®] (Pasteur Merieux) 70% w/v
Peanut oil	1	ND	Bal in Oil [®] (Becton Dickinson)
Polyethylene glycol			
PEG	5	0.15–50%	Secobarbital sodium (Wyeth-Ayerst) 50%
PEG 300	3	50–65%	VePesid [®] (Bristol Myers) 65% w/v
PEG 400	3	18–67 vol%	Busulfex [®] (Orphan Medical) 67%
PEG 600	1	5% w/v	Persantine [®] (Dupont-Merck)
PEG 3350	5	0.3–3%	Depo-Medrol [®] (Upjohn) 2.95% w/v
Poppyseed oil	1	ND	Ethiodol [®] (Savage)
Propylene glycol	29	0.2–80%	Ativan [®] (Wyeth-Ayerst) 80%
Safflower oil	2	5–10%	Liposyn II [®] (Abbott) 10%
Sesame oil	6	100%	Solganal Inj. [®] (Schering)
Soybean oil	4	5–20% w/v	Intralipid [®] (Clintec) 20%
Vegetable oil	2	ND	Virilon IM Inj. [®] (Star Pharmaceuticals)

ND, No data available.

- Test for unsaponifiable matter
 - Test for free fatty acid
 - Solid paraffin test at 10°C
 - Acid value NMT 0.56 (*JP* only)
- B. Synthetic mono- and diglycerides of fatty acids (which are liquid and remain so when cooled to 10°C)
- Iodine number (<140)
 - Solid paraffin test at 10°C

The oils also are used to dissolve drugs with low aqueous solubility and provide a mechanism to slowly release drug over a long period of time. Deterioration of fixed oils, which leads to rancidity and production of free fatty acids, must be avoided in injectable products. Also the fixed oils or fatty acid esters must not contain mineral oil or paraffin which the body cannot metabolize.

Polymeric and Surface Active Compounds

Table 2 includes a broad category of excipients whose function in formulation could be as follows:

1. To impart viscosity or act as suspending agents such as carboxy methyl cellulose, sodium carboxy methyl

cellulose, acacia, Povidone, hydrolyzed gelatin, and sorbitol.

2. To act as solubilizing, wetting, or emulsifying agents such as Cremophor EL, sodium desoxycholate, Polysorbate 20 or 80, PEG 40 castor oil, PEG 60 castor oil, sodium dodecyl sulfate, lecithin, or egg yolk phospholipid.
3. To form gels such as when aluminum monostearate is added to fixed oil to form a viscous or gel-like suspension medium.

Polysorbate 80 is the most common and versatile solubilizing, wetting and emulsifying agent. Again, one must be concerned about the level of residual peroxides present in polysorbates and protecting them from air to prevent further oxidation (23). Polysorbate 80 is polyoxyethylene sorbitan ester of oleic acid (unsaturated fatty acid) while polyoxyethylene Polysorbate 20 is sorbitan ester of lauric acid (saturated fatty acid). Thus, stability differences could occur in the drug product formulated with Polysorbate 80 versus Polysorbate 20. One example is Neupogen[®] which when exposed to a high concentration of Polysorbate 20 exhibited substantially less oxidation than when exposed to a similar concentration of Polysorbate 80 (24).

Table 2 Solubilizing, wetting, suspending, emulsifying, or thickening agents

Excipient	Frequency	Range	Example
Acacia	2	7%	Tuberculin Old Test [®] (Lederle) 7%
Aluminum monostearate	1	2%	Solganal Inj. [®] (Schering) 2%
Carboxy methyl cellulose	4	0.50–0.55%	Bicillin [®] (Wyeth-Ayerst) 0.55%
Carboxy methyl cellulose, sodium	19	0.15–3.0%	Nutropin Depot [®] (Genentech) 3%
Cremophor EL ^a	3	50–65% w/v	Sandimmune [®] (Sandoz) 65% w/v
Desoxycholate sodium	1	0.4% w/v	Fungizone [®] (Bristol Myers) 0.41% w/v
Egg yolk phospholipid	3	1.2%	Intralipid [®] (Clintec) 1.2%
Gelatin, Hydrolyzed	1	16% w/v	Cortone [®] (Merck) 16% w/v
Lecithin	8	0.4–1.2% w/v	Diprivan [®] (Zeneca) 1.2% w/v
Pluronic F-68	1	—	Fluosol [®] (Alpha Therapeutics)
Polyoxyethylated fatty acid	2	7–12% w/v	AquaMephyton [®] (Merck) 7% w/v Aquasol A parenteral [®] (Astra) 12%
Polysorbate 80 (Tween 80)	48	0.004–100%	Taxoterer [®] (Aventis) 100%
Polysorbate 20 (Tween 20)	9	0.01–0.4%	Calcijex [®] (Abbott) 0.4% w/v
PEG 40 castor oil ^b	1	11.5 vol%	Monistat [®] (Janssen) 11.5 vol%
PEG 60 castor oil ^c	1	20% w/v	Prograf [®] (Fujisawa) 20% w/v
Povidone (Polyvinyl pyrrolidone)	7	0.5–0.6% w/v	Bicillin [®] (Wyeth-Ayerst) 0.6% w/v
Sodium dodecyl sulfate (Na lauryl sulfate)	1	0.018% w/v	Proleukin [®] (Cetus) 0.018% w/v
Sorbitol	3	25–50%	Aristrospan [®] (Fujisawa) 50 vol%

^aCremophor EL, Etocas 35, polyethoxylated castor oil, polyoxyethylene 35 castor oil.

^bPEG 40 castor oil, polyoxyl 40 castor oil, castor oil POE-40, Croduret 40, polyoxyethylene 40 castor oil, Protachem CA-40.

^cPEG 60 hydrogenated castor oil, Cremophor RH 60, hydrogenated castor oil POE-60, Protachem CAH-60.

Table 3 Chelating agents

Excipient	Frequency	Range	Example
Calcium disodium EDTA ^a	9	0.01–0.1%	Wydase [®] (Wyeth-Ayerst) 0.1% w/v
Disodium EDTA	38	0.01–0.11%	Calcijex [®] (Abbott) 0.11% w/v
Sodium EDTA	1	0.20%	Folvite [®] (Lederle) 0.20%
DTPA ^b	1	0.04%	Magnevist [®] (Berlex) 0.04%

^aEDTA = Ethylenediaminetetraacetic acid.^bDTPA = Diethylenetriaminepentaacetic acid; pentetic acid.

Chelating Agents

Only a limited number of chelating agents are used in parenteral products (Table 3). They serve to complex heavy metals and therefore can improve the efficacy of antioxidants or preservatives. Citric acid, tartaric acid and some amino acids also can act as chelating agents. There have been some misunderstandings concerning the use of EDTA (as calcium salt) as an approved injectable product in Japan. Currently in Japan, some drug products that contain calcium disodium EDTA are on the market and this excipient is also listed as an official excipient (see Table 11). An advantage of calcium EDTA over tetrasodium salt is that calcium EDTA does not contribute sodium and does not chelate as much calcium from the blood.

A complexing agent should not be used in metallo-protein formulations, where the protein subunits are held

by the metal (25). The EDTA, in rare instances, can increase the oxidation rate due to binding of the EDTA–metal complex to protein, resulting in site-specific generation of radicals (26).

Antioxidants

Antioxidants are used to prevent the oxidation of active substances and excipients in the finished product. There are three main types of antioxidants:

1. *True antioxidants*: They act by a chain-termination mechanism by reacting with free radicals, e.g., butylated hydroxytoluene.
2. *Reducing agents*: They have a lower redox potential than the drug and get preferentially oxidized, e.g.,

Table 4 Antioxidants and reducing agents

Excipient	Frequency	Range	Example
Acetone sodium bisulfite	4	0.2–0.4% w/v	Novocaine [®] (Sanofi-Winthrop) 0.4% w/v
Ascorbate (sodium/acid)	8	0.1–4.8% w/v	Vibramycin [®] (Pfizer) 4.8% w/v
Bisulfite sodium	31	0.02–0.66% w/v	Amikin [®] (Bristol Myers) 0.66% w/v
Butylated hydroxy anisole (BHA)	3	0.00028–0.03% w/v	Aquasol A [®] (Astra) 0.03% w/v
Butylated hydroxy toluene (BHT)	3	0.00116–0.03% w/v	Aquasol A [®] (Astra) 0.03% w/v
Cystein/Cysteinate HCl	3	0.07–1.3% w/v	Acthrel [®] (Ferring) 1.3% w/v
Dithionite sodium (Na hydrosulfite, Na sulfoxylate)	1	0.10%	Numorphan [®] (Endo Lab) 0.10%
Gentisic acid	1	0.02% w/v	OctreoScan [®] (Mallinckrodt) 0.02% w/v
Gentisic acid ethanolamine	1	2%	M.V.I. 12 [®] (Astra) 2%
Glutamate monosodium	2	0.1% w/v	Varivax [®] (Merck) 0.1% w/v
Formaldehyde sulfoxylate sodium	9	0.02–0.5% w/v	Terramycin solution (Pfizer) 0.5% w/v
Metabisulfite potassium	1	0.10%	Vasoxyl [®] (Glaxo-Wellcome) 0.10%
Metabisulfite sodium	32	0.02–1% w/v	Intropin [®] (DuPont) 1% w/v
Monothioglycerol (Thioglycerol)	6	0.1–1%	Terramycin solution (Pfizer) 1%
Propyl gallate	2	0.02%	Navane [®] (Pfizer) 0.02%
Sulfite sodium	7	0.05–0.2% w/v	Enlon [®] (Ohmeda) 0.2% w/v
Tocopherol alpha	1	0.005% w/v	AmBisome [®] (Fujisawa) 0.005%
Thioglycolate sodium	1	0.66% w/v	Sus-Phrine [®] (Forest) 0.66% w/v

ascorbic acid. Thus, they can be consumed during the shelf-life of the product.

3. *Antioxidant synergists*: These enhance the effect of antioxidants, e.g., EDTA.

Table 4 summarizes the antioxidants, their frequency of use, concentration range, and examples of products containing them. Sulfite, bisulfite, and metabisulfite constitute the majority of antioxidants used in parenteral products despite several reports of incompatibility and toxicity (27, 28). Butylated hydroxy anisole, butylated hydroxy toluene, alpha tocopherol, and propyl gallate are primarily used in semi/nonaqueous vehicles because of their low aqueous solubility (29). Ascorbic acid/sodium ascorbate may serve as an antioxidant, buffer and chelating agent in the same formulation. Some amino acids such as cysteine also function as effective antioxidants.

The Committee for Proprietary Medicinal Products (CPMP) guideline calls for a full explanation and justification for including antioxidants in the formulation (30). It further states that antioxidants should only be included in a formulation if it has been proven that their use cannot be avoided. Thus, it is imperative to first try inert gas (nitrogen or argon) in the headspace to prevent oxidation. If the antioxidant has to be included, its concentration must be justified in terms of efficacy and safety. Antioxidants such as sulfites and metabisulfites are especially undesirable.

Some antioxidants possess antimicrobial properties, such as propyl gallate and butylated hydroxy anisole, which are somewhat effective against bacteria. Butylated hydroxy toluene has demonstrated some antiviral activity. Compatibility of antioxidants with the drug, packaging system and the body should be studied carefully. For example, tocopherols may be absorbed onto plastics;

ascorbic acid is incompatible with alkalis, heavy metals, and oxidizing materials such as phenylephrine, and sodium nitrite; and propyl gallate forms complexes with metal ions such as sodium, potassium and iron.

Preservatives

Benzyl alcohol is the most common antimicrobial preservative present in parenteral formulations (Table 5). This observation is consistent with other surveys (6, 31). Parabens are the second most common preservatives. Surprisingly, thimerosal is also common, especially in vaccines, even though some individuals are sensitive to mercurics. Several preservatives can volatilize easily (such as benzyl alcohol, and phenol) and, therefore, should not be used for lyophilized dosage form. Chlorocresol is purported to be a good preservative for parenterals, but our survey did not find any examples of commercial products containing chlorocresol. The British Pharmaceutical Codex and Martindale list chlorocresol as a preservative to be used in multidose aqueous injections at concentrations of 0.1% but no examples of injectable products have been provided (32, 33).

Antimicrobial preservatives are allowed in multidose injections to prevent growth of microorganisms that may accidentally enter the container during withdrawal of the dose. However, they are discouraged from being used in single-dose injections in the United States while the EP and BP allow aqueous preparations, that are manufactured using aseptic techniques, to contain suitable preservatives. It should be emphasized that preservatives should never be used as a substitute for inadequate good manufacturing practices (GMP). BP and EP prohibit antimicrobials from single-dose

Table 5 Antimicrobial preservatives

Excipient	Frequency	Range	Example
Benzalkonium chloride	1	0.02% w/v	Celestone Soluspan® (Schering) 0.02% w/v
Benzethonium chloride	4	0.01%	Benadryl® (Parke-Davis) 0.01% w/v
Benzyl alcohol	85	0.75–10%	Progesterone (United Res) 10%
Chlorbutanol	18	0.25–0.5%	Codine phosphate (Wyeth-Ayerst) 0.5%
<i>m</i> -Cresol	5	0.1–0.35%	Humalog® (Lilly) 0.35%
Myristyl gamma-picolinium chloride	2	0.0195–0.169% w/v	Depo-Provera® (Pharmacia-Upjohn) 0.169% w/v
Paraben methyl	52	0.05–0.18%	Inapsine® (Janssen) 0.18% w/v
Paraben propyl	44	0.01–0.1%	Xylocaine w/ Epinephrine (Astra) 0.1% w/v
Phenol	50	0.2–0.5%	Calcimar® (Rhone-Poulanc) 0.5% w/v
2-Phenoxyethanol	4	0.50%	Havrix® (SmithKline Beecham) 0.50% w/v
Phenyl mercuric nitrate	3	0.001%	Antivenin® (Wyeth-Ayerst) 0.001%
Thimerosal	48	0.003–0.012%	Atgam® (Pharmacia-Upjohn) 0.01%

Table 6 Maximum permissible amount of preservatives and antioxidants

Excipient	Maximum limit in USP
Mercurial compounds	0.01%
Cationic surfactants	0.01%
Chlorobutanol	0.50%
Cresol	0.50%
Phenol	0.50%
Sulfur dioxide or an equivalent amount of the sulfite, bisulfite, or metabisulfite of potassium or sodium	0.20%

injections where the dose volume is greater than 15 mL or if the drug product is to be injected via intracisternal, or any route which gives access to the cerebrospinal fluid (CSF). Toxicity is the primary reason for minimizing the use of antimicrobial preservatives. For example, many individuals are allergic to mercury preservatives while benzyl alcohol is contraindicated in children under the age of 2. USP has also placed some restrictions on the maximum concentration of preservatives allowed in the formulation to address toxicity and allergic reactions (Table 6). The World Health Organization (WHO) has set an estimated total acceptable daily intake for sorbate (as acid, calcium, potassium and sodium salts) as not more than 25 mg/kg body weight. The efficacy of the preservative should be assessed during product development using Antimicrobial Preservative Effectiveness Testing (PET) (34–36). Thus, an aqueous-preserved parenteral product can be used up to a maximum of 28 days after the container has been opened (37). Obviously, 28 days has to be justified by performing PET on the finished product in the final package. On the other hand, unpreserved products preferably should be used immediately following opening, reconstitution, or dilution.

Buffers

Buffers are added to a formulation to adjust the pH in order to optimize solubility and stability. For parenteral preparations, the pH of the product should be close to physiologic pH. The selection of buffer concentration (ionic strength) and buffer species is important. For example, 5–15 mM of citrate buffers are used in the formulation but increasing buffer concentration to >50 mM will result in excessive pain on subcutaneous injection and toxic effects due to the chelation of calcium in the blood.

Buffers have maximum buffer capacities near their pK_a . For products that may be subjected to excessive temperature fluctuations during processing (such as sterilization or lyophilization), it is important to select buffers with a small $\Delta pK_a/^\circ\text{C}$. Thus, Tris, whose $\Delta pK_a/^\circ\text{C}$ is large ($-0.028/^\circ\text{C}$), the pH of buffer made at 25°C will change from 7.1 to 5.0 at 100°C . This may dramatically alter the stability or solubility of the drug. Similarly, the best buffers for a lyophilized product may be those that show the least pH change upon cooling, that do not crystallize out, and that can remain in the amorphous state protecting the drug. For example, replacing succinate with glycolate buffer improves the stability of lyophilized interferon- γ (38). During the lyophilization of mannitol that contains succinate buffer at pH 5, monosodium succinate crystallizes, reducing the pH and resulting in the unfolding of interferon- γ . This pH shift is not seen with glycolate buffer.

Table 7 lists buffers and chemicals used to adjust the pH of formulations and the product pH range. Phosphate, citrate, and acetate are the most common buffers used in parenteral products. Mono- and diethanolamines are added to adjust pH and form corresponding salts. Hydrogen bromide, sulfuric acid, benzene sulfonic acid, and methane sulfonic acids are added to drugs which are salts of bromide (Scopolamine HBr, Hyoscine HBr), sulfate (Nebcin, Tobramycin sulfate), besylate (Tracrium Injection, Atracurium besylate) or mesylate (DHE 45 Injection, Dihydroergotamine mesylate). Glucono delta lactone is used to adjust the pH of Quinidine gluconate. Benzoate buffer, at a concentration of 5%, is used in Valium Injection. Citrates are a common buffer that can have a dual role as chelating agents. The amino acids lysine and glycine, function as buffers and stabilize proteins and peptide formulations. These amino acids are also used as lyo-additives and may prevent cold denaturation. Lactate and tartrate are occasionally used as buffer systems. Acetates are good buffers at low pH, but they are not generally used for lyophilization because of potential sublimation of acetates.

Bulking Agents, Protectants, and Tonicity Adjusters

Table 8 lists additives that are used to modify osmolality, and as bulking or lyo/cryoprotective agents. Dextrose and sodium chloride are used to adjust tonicity in the majority of formulations. Some amino acids such as glycine, alanine, histidine, imidazole, arginine, asparagine, and aspartic acid are used as bulking agents for lyophilization and also can serve as stabilizers, and/or as buffers.

Table 7 Buffers and pH-adjusting agents

Excipient	pH Range	Example
Acetate		
Sodium	3.7–4.3	Syntocinon [®] (Novartis)
Acetic acid	3.7–4.3	Syntocinon [®] (Novartis)
Glacial acetic acid	3.5–5.5	Brevibloc [®] (Ohmeda)
Ammonium	6.8–7.8	Bumex Injection [®] (Roche)
Ammonium sulfate	—	Innovar [®] (Astra)
Ammonium hydroxide	—	Triostat [®] (Jones Medical)
Arginine	7.0–7.4	Retavase [®] (Boehringer)
Aspartic acid	5.7–6.4	Pepcid [®] (Merck)
Benzene sulfonic acid	3.25–3.65	Nimbex [®] (Glaxo Wellcome)
Benzoate Sodium/acid	6.2–6.9	Valium [®] (Roche)
Bicarbonate	5.5–11.0	Cenolate [®] (Abbott)
Boric acid/sodium		Comvax [®] (Merck)
Carbonate, sodium	5.0–11.0	Hyperab [®] (Bayer)
Citrate		
Acid	3.0–5.5	DTIC-Dome [®] (Bayer)
Sodium	3.5–6.5	Amikin [®] (Bristol Myers)
Disodium	—	Cerezyme [®] (Genzyme)
Trisodium	—	Cerezyme [®] (Genzyme)
Diethanolamine	9.5–10.5	Bactim IV [®] (Roche)
Glucono delta lactone	5.5–7.0	Quinidine Gluconate (Lilly)
Glycine/glycine HCl	6.4–7.2	Hep-B Gammagee [®] (Merck)
Histidine/histidine HCl	6.5	Doxil [®] (Sequus)
Hydrochloric acid	6.0–7.6	Amicar [®] (Immunex)
Hydrobromic acid	3.5–6.5	Scopolamine (UDL)
Lactate sodium/Acid	2.7–5.7	Innovar [®] (Janssen)
Lysine (L)	—	Eminase [®] (Roberts)
Maleic acid	3.0–5.0	Librium [®] (Roche)
Meglumine	6.5–8.0	Magnevist [®] (Berlex)
Methanesulfonic acid	3.2–4.0	DHE-45 [®] (Novartis)
Monoethanolamine	8.0–9.0	Terramycin (Pfizer)
Phosphate		
Acid	6.5–8.5	Saizen [®] (Serono Labs)
Monobasic potassium	6.7–7.3	Zantac [®] (Glaxo-Wellcome)
Monobasic sodium ^a	6.0–8.0	Pregnyl [®] (Organon)
Dibasic sodium ^b	6.7–7.8	Zantac [®] (Glaxo-Wellcome)
Tribasic sodium	—	Synthroid [®] (Knoll)
Sodium hydroxide	Broad range	Optiray [®] (Mallinckrodt)
Succinate sodium/Disodium	5.0–6.0	AmBisome [®] (Fujisawa)
Sulfuric acid	3.0–6.5	Nebcin [®] (Lilly)
Tartrate sodium/acid	2.7–6.2	Methergine [®] (Novartis)
Tromethamine	6.0–7.5	Optiray [®] (Mallinckrodt)

^aSodium biphosphate, sodium dihydrogen phosphate, or Na dihydrogen orthophosphate.^bSodium phosphate, disodium hydrogen phosphate.

Monosaccharides (dextrose, glucose, maltose, lactose), disaccharides (sucrose, trehalose), polyhydric alcohols (inositol, mannitol, sorbitol), glycols (PEG 3350), Povidone (polyvinylpyrrolidone, PVP) and proteins (albumin, gelatin) are commonly used as lyo-additives.

Hydroxyethyl starch (hetastarch) and pentastarch, which are currently used as plasma expanders in commercial injectable products such as Hespan and Pentaspan, also are being evaluated as protectants during freeze-drying of proteins.

Table 8 Bulking agents, protectants, and tonicity adjusters

Excipient	Example
Alanine	Thrombate III [®] (Bayer)
Albumin	Bioclate [®] (Arco)
Albumin (human)	Botox [®] (Allergan)
Amino acids	Havrix [®] (Smith Kline Beecham)
Arginine (L)	Activase [®] (Genentech)
Asparagine	Tice BCG [®] (Organon)
Aspartic acid (L)	Pepcid [®] (Merck)
Calcium chloride	Phenergan [®] (Wyeth-Ayerst)
Cyclodextrin-alpha	Edex [®] (Schwartz)
Cyclodextrin-gamma	Cardiotec [®] (Squibb)
Dextran 40	Etopophos [®] (Bristol Myers)
Dextrose	Betaseron [®] (Berlex)
Gelatin (cross-linked)	Kabikinase [®] (Pharmacia-Upjohn)
Gelatin (hydrolyzed)	Acthar [®] (Rhone-Poulanc Rorer)
Lactic & glycolic acid copolymers	Lupron Depot [®] (TAP)
Glucose	Iveegam [®] (Immuno-US)
Glycerine	Tice BCG [®] (Organon)
Glycine	Atgam [®] (Pharmacia-Upjohn)
Histidine	Antihemophilic Factor, human (Am. Red Cross)
Imidazole	Helixate [®] (Armour)
Inositol	OctreoScan [®] (Mallinckrodt)
Lactose	Caverject [®] (Pharmacia-Upjohn)
Magnesium chloride	Terramycin Solution (Pfizer)
Magnesium sulfate	Tice BCG [®] (Organon)
Maltose	Gamimune N [®] (Bayer)
Mannitol	Elspar [®] (Merck)
Polyethylene glycol 3350	Bioclate [®] (Arco)
Polylactic acid	Lupron Depot [®] (TAP)
Polysorbate 80	Helixate [®] (Armour)
Potassium chloride	Varivax [®] (Merck)
Povidone	Alkeran [®] (Glaxo-Wellcome)
Sodium chloride	WinRho SD [®] (Univax)
Sodium cholesteryl sulfate	Amphotec [®] (Sequus)
Sodium succinate	Actimmune [®] (Genentech)
Sodium sulfate	Depo-Provera [®] (Pharmacia-Upjohn)
Sorbitol	Panhematin [®] (Abbott)
Sucrose	Prolastin [®] (Bayer)
Trehalose (alpha, alpha)	Herceptin [®] (Genentech)

PVP has been used in injectable products as a solubilizing agent, a protectant and as a bulking agent. Only pyrogen-free grade, with low molecular weight (*K* value less than 18) should be used in parenteral products to allow for rapid renal elimination. PVP not only solubilizes drugs such as rifampicin, but it also can reduce the local toxicity as seen in oxytetracycline injection.

Many proteins can be stabilized in the lyophilized state if the stabilizer and protein do not phase separate during freezing or the stabilizer does not crystallize out.

In the case of Neupogen[®] (GCSF), the original formulation was modified by replacing mannitol with sorbitol to prevent the loss of activity of liquid formulation in case of accidental freezing (24). Mannitol crystallizes if the solution freezes while sorbitol remains in an amorphous state protecting GCSF. Similarly, it is useful that the drug remains dispersed in the stabilizer upon freezing of the solution. Thus, Cefoxitin, a cephalosporin, is more stable when freeze-dried with sucrose than with trehalose, although the glass transition temperature and structural relaxation time is much

greater for trehalose than sucrose (39). FTIR data indicated that the trehalose–cefoxitin system phase separated into two nearly pure components, resulting in no protection (stability). Similarly, dextran was not found to be as useful a cryoprotectant for protein as sucrose because dextran and protein underwent phase segregation as the solution started to freeze. The mechanism of cryoprotection in the solution has been explained by the preferential exclusion hypothesis (40).

Trehalose is a nonreducing disaccharide composed of two D-glucose monomers. It is found in several animals that can withstand dehydration and therefore was suggested as a stabilizer of drugs that undergo denaturation during spray or freeze-drying (41). Herceptin® (Trastuzumab) is a recombinant DNA-derived monoclonal antibody (MAb) that is used for treating metastatic breast cancer. The MAb has been stabilized in the lyophilized formulation using α,α -trehalose dihydrate. Trehalose has also been used as a cryoprotectant to prevent liposomal aggregation and leakage. In the dried state, carbohydrates such as trehalose, and inositol, exert their protective effect by acting as a water substitute (42).

Additives may have to be included in the formulation to adjust the specific gravity. This is important for drugs that upon administration may come in contact with CSF. CSF has a specific gravity of 1.0059 at 37°C. Solutions that have the same specific gravity as that of CSF are termed isobaric, while those solutions that have specific gravity greater than that of CSF are called hyperbaric. Upon administration of a hyperbaric solution in the spinal cord, the injected solution will settle and will affect spinal nerves at the end of the spinal cord. For example, Dibucaine hydrochloride solution (Nupercaine® 1:200) is isobaric, while Nupercaine 1:500 is hypobaric (specific gravity of 1.0036 at 37°C). Nupercaine heavy solution is made hyperbaric by addition of 5% dextrose solution, and this solution will block (anesthetize) the lower spinal nerves as it settles in the spinal cord.

Special Additives

Special additives serve special functions in pharmaceutical formulations (Table 9). The following is a summary of special additives along with their intended use:

1. Calcium gluconate injection (American Regent) is a saturated solution of 10% w/v. Calcium D-saccharate tetrahydrate 0.46% w/v is added to prevent crystallization during temperature fluctuations.
2. Cipro IV® (Ciprofloxacin, Bayer) contains lactic acid as a solubilizing agent for the antibiotic.

3. Premarin Injection® (Conjugated Estrogens, Wyeth-Ayerst Labs) is a lyophilized product that contains simethicone to prevent the formation of foam during reconstitution.
4. Dexamethasone acetate (Dalalone DP, Forest, Decadron-LA, Merck) and Dexamethasone sodium phosphate (Merck) are available as a suspension or a solution. These dexamethasone formulations contain creatine or creatinine as additives.
5. Adriamycin RDF® (Doxorubicin hydrochloride, Pharmacia-Upjohn) contains methyl paraben, 0.2 mg/mL to increase dissolution (43).
6. Ergotrate maleate (Ergonovine maleate, Lilly) contains 0.1% ethyl lactate as a solubilizing agent.
7. Estradurin Injection® (Polyestradiol phosphate, Wyeth-Ayerst Labs) uses Niacinamide (12.5 mg/ml) as a solubilizing agent. Hydeltasol® also contains niacinamide. The concept of hydrotropic agents to increase water solubility has been tried on several compounds, including proteins (44, 45).
8. Aluminum, in the form of aluminum hydroxide, aluminum phosphate or aluminum potassium sulfate, is used as adjuvant in various vaccine formulations to elicit an increased immunogenic response.
9. Lupron Depot Injection® is lyophilized microspheres of gelatin and glycolic–lactic acid for intramuscular (IM) injection. Nutropin Depot consists of polylactate–glycolate microspheres.
10. Gamma cyclodextrin is used as a stabilizer in Cardiotec® at a concentration of 50 mg/mL.
11. Alprostadil (Edex®, Schwartz) is a lyophilized product of Prostaglandin E₁ in α -cyclodextrin inclusion complex. The complex has better stability and aqueous solubility than the drug itself.
12. Itraconazole (Sporanox®, Janssen) is solubilized as a molecular inclusion complex using hydroxypropyl- β -cyclodextrin.
13. Sodium caprylate (sodium octoate) has antifungal properties, but it is also used to improve the stability of albumin solution against the effects of heat. Albumin solution can be pasteurized by heating at 60°C for 10 h in the presence of sodium caprylate. Acetyl tryptophanate sodium is also added to albumin formulations.
14. Meglumine (*N*-methylglucamine) is used to form in situ salt. For example, diatrizoic acid, an X-ray contrast agent, is more stable when autoclaved as meglumine salt than as sodium salt (46). Meglumine is also added to Magnevist®, a magnetic resonance contrast agent.
15. Tri-*n*-butyl phosphate is present as an excipient in human immune globulin solution (Venoglobulin®). Its exact function in the formulation is not known, but it may serve as a scavenging agent.

Table 9 Special additives

Excipient	Example
Acetyl tryptophanate	Human Albumin (American Red Cross)
Aluminum hydroxide	Recombivax HB [®] (Merck)
Aluminum phosphate	Tetanus Toxoid Adsorbed (Wyeth-Ayerst)
Aluminum potassium sulfate	TD Adsorbed Adult (Pasteur Merieux)
ϵ -Aminocaproic acid	Eminase [®] (Roberts)
Calcium D-saccharate	Calcium Gluconate (American Regent)
Caprylate sodium	Human Albumin (American Red Cross)
8-Chlorotheophylline	Dimenhydrinate [®] (Steris)
Creatine	Dalalone DP [®] (Forest)
Creatinine	Decadron [®] (Merck)
Cholesterol	Doxil [®] (Sequus)
Cholesteryl sulfate sodium	Amphotec [®] (Sequus)
Alpha-cyclodextrin	Edex [®] (Schwartz)
Gamma-cyclodextrin	Cardiotec [®] (Squibb)
Hydroxypropyl beta cyclodextrin	Sporanox [®] (Janssen)
Distearyl Phosphatidylcholine	DaunoXome [®] (Nexstar)
Distearyl Phosphatidylglycerol	MiKasome [®] (NeXstar)
L-Alpha-Dimyristoylphosphatidylcholine	Abelcet [®] (The Liposome Co.)
L-Alpha-Dimyristoylphosphatidylglycerol	Abelcet [®] (The Liposome Co.)
Dioleoylphosphatidylcholine (DOPC)	DepoCyt [®] (Chiron)
Dipalmitoylphosphatidylglycerol (DPPG)	DepoCyt [®] (Chiron)
MPEG-distearoyl phosphoethanolamine	Doxil [®] (Sequus)
Diatrizoic acid	Conray [®] (Mallinckrodt)
Ethyl lactate	Ergotrate maleate (Lilly)
Ethylenediamine	Aminophylline (Abbott)
L-Glutamate sodium	Kabikinate [®] (Pharmacia-Upjohn)
Hydrogenated soy phosphatidylcholine	Doxil [®] (Sequus)
Iron ammonium citrate	Tice BCG [®] (Organon)
Lactic acid	Cipro IV [®] (Bayer)
D,L-Lactic and glycolic acid copolymer	Zoladex [®] (Zeneca)
Meglumine	Magnevist [®] (Berlex)
Niacinamide	Estradurin [®] (Wyeth-Ayerst)
Paraben methyl	Adriamycin RDF [®] (Pharmacia-Upjohn)
Protamine	Insulatard NPH [®] (Novo Nordisk)
Simethicone	Premarin Injection [®] (Wyeth-Ayerst)
Saccharin sodium	Compazine Injection [®] (Smith Kline Beecham)
Tri- <i>n</i> -butyl phosphate	Venoglobulin [®] (Alpha Therapeutic)
Triolein	DepoCyt [®] (Chiron)
von Willebrand factor	Bioclate [®] (Arco)
Zinc	Lente Insulin [®] (Novo Nordisk)
Zinc acetate	Nutropin Depot [®] (Genentech)
Zinc carbonate	Nutropin Depot [®] (Genentech)
Zinc oxide	Humalog [®] (Lilly)

16. von Willebrand factor is used to stabilize recombinant antihemophilic factor (Bioclate[®]).
17. Maltose serves as a tonicity adjuster and stabilizer in immune globulin formulation (Gamimune N[®]).
18. Epsilon amino caproic acid (6-amino hexanoic acid) is used as a stabilizer in anistreplase (Eminase Injection[®]).

19. Zinc and protamine have been added to insulin to form complexes and control the duration of action.

The FDA has published the “Inactive Ingredient Guide” which lists all excipients in alphabetical order (14). Each ingredient is followed by the route of administration, and

Table 10 List of excipients from the 1996 FDA Inactive Ingredient Guide

Benzyl chloride	Poloxamer 165
Butyl paraben	PEG 4000
Caldiamide sodium	Polyoxyethylene fatty acid esters
Calteridol calcium	Polyoxyethylene sorbitan monostearate
Cellulose (microcrystalline)	Polyoxyl 35 castor oil
Deoxycholic acid	Polysorbate 40
Dicyclohexyl carbodiimide	Polysorbate 85
Diethyl amine	Potassium hydroxide
Disofenin	Potassium phosphate, dibasic
Docusate sodium	Sodium bisulfate
Edamine	Sodium chlorate
Exametazime	Sodium hypochloride
Glucaptate sodium	Sodium iodide
Glucaptate calcium	Sodium pyrophosphate
Glucuronic acid	Sodium thiosulfate, anhydrous
Guanidine HCl	Sodium trimetaphosphate
Iofetamine HCl	Sorbitan monopalmitate
Lactobionic acid	Stannous chloride
Lidofenin	Stannous fluoride
Medrofenin	Stannous tartrate
Medronate disodium	Starch
Medronic acid	Succimer
Methyl boronic acid	Succinic acid
Methyl cellulose	Sulfurous acid
Methylene blue	Tetrakis (1-isocyano-2-methoxy-2-methyl-propionate) copper (I) Tc
<i>N</i> -(Carbamoyl-methoxy polyethylene glycol 2000)-1,2- distearoyl	Thiazoximic acid
<i>N</i> -2-Hydroxyethyl piperazine <i>N'</i> -2'-ethane sulonic acid	Urea
Nioxime	Zic acetate
Nitric acid	Zinc chloride
Oxyquinoline	2-ethyl hexanoic acid
Pentate (DTPA) calcium trisodium	PEG vegetable oil

in some cases, the range of concentration used in the approved drug product. However, this list does not provide the name of commercial product(s) corresponding to each excipient. Table 10 summarizes all the excipients included in the “Inactive Ingredient Guide” that do not appear in the Physician’s Desk Reference (PDR), GenRx, or Handbook of Injectable Drugs.

Similarly, in Japan the “Japanese Pharmaceutical Excipients Directory” is published by the Japanese Pharmaceutical Excipients Council, with the cooperation and guidance of the Ministry of Health and Welfare (47). This directory divides the excipients into:

1. *Official*. Those 590 excipients that have been recognized in the JP, Japanese Pharmaceutical Codex, and Japanese Pharmaceutical Excipients, and

for which testing methods and standards have been determined. Table 11 summarizes official excipients used in parenteral products.

2. *Nonofficial Excipients*. These 522 excipients are used in pharmaceutical products sold in Japan and will be included in the official book or in supplemental editions. The nonofficial excipients, used in parenteral products, are listed in Table 12.

REGULATORY PERSPECTIVE

The International Pharmaceutical Excipients Council (IPEC) has classified excipients into the following four classes, based on available safety testing information (48):

Table 11 Official Japanese pharmaceutical excipients

Name	Uses	Administration route
Acacia	Diluent, dispersing agent	im
Acetic acid	Buffer agent, solvent, stabilizer	iv, im, sc
L-Alanine	Stabilizer	iv, im,
Aluminum monostearate	Dispersing agent, stabilizer, vehicle	other inj.
Aluminum potassium sulfate	pH adjustment, stabilizer	im, sc
Aminoacetic acid	Buffering agent, solubilizer, stabilizer, suspending agent, vehicle	iv, im, sc, ic
Anhydrous citric acid	Buffer agent, pH adjustment, solubilizing agent, stabilizer	iv, im, other inj.
Anhydrous disodium hydrogen phosphate	Buffering agent, pH adjustment, solubilizer, stabilizer, suspending agent	iv, im, sc, other inj.
Anhydrous sodium dihydrogen phosphate	Buffering agent, pH adjustment, stabilizer	iv, im, other inj.
Arginine hydrochloride	Buffering agent, solubilizing agent, stabilizer	iv, im, sc
Ascorbic acid	Antioxidant, buffering agent, Stabilizer	iv, im, sc, ia
L-Aspartic acid	Solubilizer, stabilizer, vehicle	iv, im
Benzylkonium chloride	Buffering agent, preservative, stabilizer	
Benzethonium chloride	Dispersing agent, preservative, stabilizer	iv, im, other inj.
Benzoic acid	Buffering agent, preservative, stabilizer	iv, im
Benzyl alcohol	Preservative, solubilizer, solvent, stabilizer	iv, im, sc, ic, other inj.
Benzyl benzoate	Antiseptic, solubilizer, solvent	im
Calcium bromide	Isotonicity, stabilizer	iv
Calcium chloride	Isotonicity, suspending agent	iv
Calcium disodium edetate	Stabilizer	iv, ic, ia, is, other inj.
Calcium gluconate	Buffering agent, stabilizer	iv, im, sc
Calcium oxide	Solubilizing agent	iv
Calcium D-saccharate	Stabilizer	iv
Camellia oil	Solvent	im, sc
Carmellose sodium	Emulsifying agent, solubilizing agent, stabilizer, suspending agent	im, ic, sc, other inj.
Castor oil	Solubilizer, solvent	im
Chlorobutanol	Buffering agent, preservative	iv, im, sc
Citric acid	Antioxidant, buffering agent, pH adjustment, preservative, solubilizing agent, stabilizer	iv, im, sc, ia,
Concentrated glycerin	Isotonicity, solubilizer, stabilizer	iv, im, sc
Creatinine	Buffering agent, stabilizer	iv, im, ic, sc, other inj.
Cresol	Preservative	iv, im, sc
L-Cystine	Stabilizer	iv
Dehydrated ethanol	Solubilizer, solubilizing agent, solvent	iv, im, sc
Dextran 40	Stabilizer, vehicle	iv, im
Dextran 70	Stabilizer	sc
Dibasic potassium sulfate	Buffering agent, pH adjustment	iv, im, sc
Dibasic sodium citrate	Buffering agent, vehicle	iv
Dibasic sodium phosphate	Buffering agent, pH adjustment, solubilizing agent, stabilizer, vehicle	iv, im, sc, ia, is, ic, other inj.
Dilute hydrochloric acid	Buffering agent, pH adjustment, solubilizer, stabilizer	iv, im, sc
N, N-Dimethylacetamide	Solvent	iv

(Continued)

Table 11 Official Japanese pharmaceutical excipients (*Continued*)

Name	Uses	Administration route
Glucose	Buffering agent, isotonicity, solubilizer, stabilizer, vehicle	iv, im, sc, ic
Glycerin	Dispersing agent, isotonicity, preservative, solubilizing agent, solvent, stabilizer, suspending agent, vehicle	iv, im, sc, other inj.
Heparin sodium	Stabilizer	iv
L-Histidine	Stabilizer	iv
Hydrochloric acid	pH adjustment, solubilizing agent, stabilizer	iv, im, sc, ia, is, ic, other inj.
N-Hydroxyethyl lactamide solution	Solubilizing agent	iv
Hydroxypropylcellulose	Emulsifying agent, solubilizer, stabilizer, suspending agent, vehicle	im
Isotonic sodium chloride solution	Isotonicity, solvent	iv, im, sc, ia, ic, other inj.
Lactic acid	Buffering agent, pH adjustment, solubilizer, stabilizer,	iv, im, sc
Lactose	Dispersing agent, suspending agent, vehicle	iv, im, sc, ia, ic, other inj.
Lidocaine	Solubilizing agent, solvent	iv, im
L-Lysine-L-Glutamate	Solubilizing agent, stabilizer	iv
Lysine hydrochloride	Stabilizer	iv
Macrogol 400 (PEG 400)	Solubilizing agent	iv
Macrogol 4000 (PEG 4000)	Isotonicity, solubilizer, solvent, stabilizer, suspending agent, vehicle, wetting agent	iv, im, sc
Magnesium chloride	Isotonicity, solubilizing agent, stabilizer	iv
Magnesium gluconate	Stabilizer	iv
Magnesium sulfate	Stabilizer	iv, im, sc
Maleic acid	Buffering agent, pH adjustment, stabilizer	im
Maleic anhydride	Solubilizer, stabilizer	iv
Maltose	Stabilizer	iv, im, sc, ic, other inj.
D-Mannitol	Isotonicity, solubilizing agent, stabilizer	iv, im, sc, ic, other inj.
Meglumine	pH adjustment, solubilizing agent	iv
Mepylcaine hydrochloride	Soothing agent	iv, im, sc
Methanesulfonic acid	pH adjustment	im, sc
L-Methionine	Stabilizer	dental inj.
Methyl parahydroxybenzoate	Preservative, stabilizer	iv, im, sc, ic, other inj.
Monobasic potassium phosphate	Buffer agent, isotonicity, pH adjustment, solubilizing agent, stabilizer	iv, im, sc, ic
Monoethanolamine	Buffer agent, pH adjustment, solubilizer, stabilizer	iv
Monopotassium L-Glutamate monohydrate	Preservative, stabilizer	sc
Monosodium L-Glutamate monohydrate	Buffer agent	iv, im, sc
Nicotinamide	Isotonicity, solubilizing agent, stabilizer	iv, im, sc, other inj.
Peanut oil	Solubilizer, solvent, suspending agent, vehicle	im
Peptone, caesin	Stabilizer	sc
Phenol	Antiseptic, preservative	ic, im, sc, ic, other inj.

(Continued)

Table 11 Official Japanese pharmaceutical excipients (*Continued*)

Name	Uses	Administration route
Disodium edetate	Antioxidant, antiseptic, preservative, stabilizer	iv, ia, other inj.
Phosphoric acid	Buffer agent, isotonicity, pH adjustment, solubilizing agent, stabilizer	iv, im, sc
Polyoxyethylene hydrogenated castor oil 60	Dispersing agent, emulsifying agent, solubilizing agent, stabilizer, surfactant, suspending agent, vehicle	iv, im, sc
Polyoxyethylene hydrogenated castor oil 51	Dispersing agent, emulsifying agent, solubilizing agent	iv, im, sc, is
Polyoxyethylene [160] Polyoxypropylene [30] glycol	Dispersing agent, emulsifying agent, solubilizer, stabilizer, suspending agent, vehicle, surfactant, wetting agent	iv
Polysorbate 80	Dispersing agent, emulsifying agent, solubilizer, surfactant, stabilizer, suspending agent, vehicle, wetting agent	iv, im, sc, ic, other inj.
Potassium sulfate	Stabilizer	local anesthetic inj.
Powdered acacia	Dispersing agent, suspending agent	im, sc
Propylene glycol	Dispersing agent, isotonicity, preservative, solubilizer, solvent, stabilizer, suspending agent, vehicle, wetting agent	iv, im, sc
Propyl parahydroxybenzoate	Antiseptic, preservative, stabilizer	iv, im, sc, ic, other inj.
Protamine sulfate	Prolongating agent	sc
Purified gelatin	Base, stabilizer, suspending agent, vehicle	iv, im, sc
Purified soybean lecithin	Dispersing agent, emulsifying agent, solubilizer, stabilizer	iv
Purified yolk lecithin	Emulsifying agent	iv
Sesame oil	Base, solubilizing agent, solvent, stabilizer, vehicle	iv, im, sc, other inj.
Sodium acetate	Buffer agent, pH adjustment, solubilizing agent, stabilizer	iv, im, sc, other inj.
Sodium acetyl tryptophan	Stabilizer	iv, sc
Sodium benzoate	Antiseptic, buffer agent, preservative, solubilizer, stabilizer	im
Sodium bicarbonate	Buffer agent, isotonicity, pH adjustment, solubilizer, stabilizer	iv, im, sc, is, ic, other inj.
Sodium bisulfite	Antioxidant, isotonicity, stabilizer	iv, im, sc, other inj.
Sodium bromide	Isotonicity	iv, im, sc
Sodium caprylate	Stabilizer	iv, sc
Sodium carbonate	Buffer agent, pH adjustment, solubilizing agent, stabilizer	iv, im, sc
Sodium chloride	Base, buffering agent, isotonicity, solubilizer, stabilizer, suspending agent, vehicle	iv, im, sc, ia, is
Sodium chondroitin sulfate	Stabilizer	iv
Sodium citrate	Antiseptic, buffer agent, isotonicity, pH adjustment, solubilizer, stabilizer	iv, im, sc, ic, ia, is, other inj.
Sodium desoxycholate	Solubilizing agent	ic, iv, is
Sodium dihydrogen phosphate dihydrate	Buffer agent, isotonicity, pH adjustment, stabilizer	iv, im, sc, ia, is, other inj.
Sodium formaldehydesulfoxylate	Stabilizer, isotonicity, pH adjustment	iv, im
Sodium hydroxide	Solubilizer	iv, im, sc, ia, ic, other inj.

(Continued)

Table 11 Official Japanese pharmaceutical excipients (*Continued*)

Name	Uses	Administration route
Sodium salicylate	Antiseptic, preservative, solubilizing agent, stabilizer	iv, im, sc
Sodium thiomalate	Antioxidant, stabilizer	im
Sodium thiosulfate	Solubilizer, stabilizer	iv, im, sc
Sorbitan sesquioleate	Base, emulsifying agent, solubilizing agent, stabilizer, surfactant, vehicle	iv, im
D-Sorbitol	Dispersing agent, isotonicity, plasticizer, preservative, solubilizing agent, stabilizer	iv, im, sc, other inj.
D-Sorbitol solution	Base, isotonicity, solubilizing agent, stabilizer, vehicle	im, sc, other inj
Soybean oil	Base, solubilizer, solvent, vehicle	iv
Stannous chloride	Reducing agent	iv
Sucrose	Base, stabilizer, vehicle	iv, sc
Tartaric acid	Buffering agent, pH adjustment, solubilizing agent, stabilizer, vehicle	iv, im
Thimerosal	Preservative	iv, im, sc
Thioglycolic acid	Solubilizing agent, stabilizer	iv, im, sc
Tribasic sodium phosphate	Buffering agent, pH adjustment	iv, im, sc
Trometamol (Tromethamine)	Buffering agent, solubilizing agent, stabilizer	iv, im, sc, ia, is, ic
Urea	Solubilizing agent, stabilizer, wetting agent	iv, im, sc
Water for injection	Solubilizer, solvent	iv, im, sc, ia, is, ic, other inj.
Xylitol	Isotonicity, stabilizer, vehicle	iv, im, other inj.
Zinc acetate	Stabilizer	sc
Zinc chloride	Stabilizer	im, sc
Zinc oxide	Dispersing agent, stabilizer, vehicle	sc

1. *New chemical excipients*: Require a full safety evaluation program. The estimated cost of safety studies for a new chemical excipient is approximately \$35 million over 4–5 years. European Union (EU) directive 75/318/EEC states that new chemical excipients will be treated in the same way as new actives. In the United States a new excipient requires a Drug Master File (DMF) to be filed with the FDA. Similarly, in Europe a dossier needs to be established. Both the DMF and dossier contain relevant safety information. The IPEC Europe has issued a draft guideline (Compilation of Excipient Masterfiles Guidelines) which provides guidance to excipient producers on how to construct a dossier that will support a Marketing Authorization Application (MAA) while maintaining the confidentiality of the data.
2. *Existing chemical excipient—first use in man*: Implies that animal safety data exist since data may have been used in some other application. Additional safety information may have to be gathered to justify its use in humans.

3. *Existing chemical excipient*: Indicates that it has been used in humans but change in route of administration (say from oral to parenteral), new dosage form, higher dose, etc. may require additional safety information.
4. *New modifications or combinations of existing excipients*: A physical interaction NOT a chemical reaction. No safety evaluation is necessary in this case.

Simply because an excipient is listed as Generally Recognized As Safe (GRAS) does not mean that it can be used in parenteral dosage form. The GRAS list may include materials that have been proven safe for food (oral administration) but have not been deemed safe for use in an injectable product. This makes it difficult for the formulation development scientist to choose additives during the dosage form development.

Many pharmacopeial monographs for identical excipients differ considerably with regards to specifications, test criteria, and analytical methods. Thus, if a pharmaceutical manufacturer is going to supply a product

Table 12 Non-official Japanese pharmaceutical excipients

Excipients	Uses	Administration
Aluminum chloride	Potentiating agent	im, sc
Aluminum hydroxide	Adsorbent	sc, im
Aminoethyl sulfonic acid	Buffer, isotonicity, stabilizer, vehicle	iv, im
Ammonium acetate	pH adjusting agent	im
Anhydrous stannous chloride	Reducing agent	iv
L-Arginine	Buffer, stabilizer, solubilizer	iv, im, sc
Asepsis sodium bicarbonate	Stabilizer	iv
Butylhydroxyanisol	Antioxidant, stabilizer	iv
<i>m</i> -Cresol	Preservative	iv, im, sc, ic
L-Cysteine	Stabilizer	iv
Cysteine hydrochloride	Antioxidant, stabilizer	iv, im
Dichlorodifluoromethane	Propellant	iv
Diethanolamine	Buffer, solubilizer, stabilizer	iv
Diethylenetriaminepentaacetic acid	Stabilizer	iv
Ferric chloride	Stabilizer	iv
Highly purified yolk lecithin	Emulsifier	iv
Human serum albumin	Preservative, stabilizer	iv, im, sc
Hydrolyzed gelatin	Stabilizer	sc
Inositol	Stabilizer, vehicle	iv, im
Lidocaine hydrochloride	Soothing agent	im
D,L-Methionine	Stabilizer	im, sc
Monobasic sodium phosphate	Buffer, isotonicity, adjust pH	iv, im, sc
Oleic acid	Dispersing agent, solvent	iv
Phenol red	Coloring agent	sc
Polyoxyethylene castor oil	Base, emulsifying agent, solubilizing agent, stabilizer	iv
Polyoxyethylene hydrogenated castor oil	Base, emulsifying agent, solubilizer, stabilizer, suspending agent, vehicle	iv
Polyoxyethylene sorbitan monolaurate	Emulsifying agent, solubilizing agent, surfactant	iv, im, sc
Potassium pyrosulfite	Stabilizer	iv, sc, im
Potassium thiocyanate	Stabilizer	iv
Purified soybean oil	Solubilizer	iv
Sodium acetate, anhydrous	Buffer, pH adjuster, solubilizing agent, stabilizer	im
Sodium carbonate, anhydrous	Buffer, solubilizing agent	iv, im, ic
Sodium dihydrogen phosphate monohydrate	Buffering agent	ic
Sodium gluconate	Stabilizer, vehicle	iv, im
Sodium pyrophosphate, anhydrous	Dispersing agent, isotonicity, stabilizer	iv, im, is
Sodium sulfite	Antioxidant, stabilizer	iv
Sodium thioglycolate	Antioxidant, stabilizer	iv, im, sc
Sorbitan esters of fatty acids	Emulsifying agent, solubilizing agent, surfactant, stabilizer, suspending agent, vehicle	iv
Succinic acid	pH adjusting agent	iv
α -Thioglycerol	Antioxidant	iv, im, sc
Triethanolamine	Buffer, pH adjuster, solubilizing agent, stabilizer	iv
Zinc chloride solution	Stabilizer	sc

throughout the world, the manufacturer will have to repeat testing on the same excipient several times in order to meet USP, JP, EP, BP, and other pharmacopoeias. EP, JP and

USP are the main driving bodies within the International Conference on Harmonization (ICH) that are working on several of the commonly used excipients in order to

achieve a single monograph for each excipient. For example, benzyl alcohol undergoes degradation by a free radical mechanism to form benzaldehyde and hydrogen peroxide. The degradation products are much more toxic than the parent molecule. The *USP*, *JP*, and *EP* require three different chromatographic systems to test for organic impurity (mainly benzaldehyde). The harmonized monograph of benzyl alcohol will eliminate unnecessary repetition, which does not contribute to the overall quality of the product (49). The following 11 pharmacopoeial tests can be substituted by a single gas chromatography (GC) method:

EP:

- Benzaldehyde, related substance (GC)
- Halogenated compounds and halides (colorimetric test)
- Assay (hydroxyl value)

JP:

- Limit test for benzaldehyde
- Limit test for chlorinated compounds
- Distillation range Assay (hydroxyl value)

NF/USP:

- Benzaldehyde (HPLC)
- Halogenated compounds and halides (colorimetric test)
- Organic volatile impurities (GC)
- Assay (hydroxyl value)

The harmonization process is just beginning and is a major step in the right direction.

Another area where regulatory bodies are focusing their attention is the manufacturing process used to produce excipients. The IPEC has undertaken major initiatives to improve the quality of additives and has published "Good Manufacturing Practices Guide for Bulk Pharmaceutical Excipients" (50). The excipients may be manufactured for the food, cosmetic, chemical, agriculture, or pharmaceutical industries, and the requirements for each area are different. The purpose of this guide is twofold: 1) to develop a quality system framework that can be used for suppliers of excipients and which will be acceptable to the pharmaceutical industry, and 2) to harmonize the requirements in the United States, Europe, and Japan.

The United States, Europe, and Japan require that all excipients be declared on the label if the product is an injectable preparation. The European guide for the label and package leaflet also lists excipients, that have special issues. These are addressed in an Annex (51). Table 13 contains a summary of some of these ingredients, which are commonly used as parenteral excipients and the corresponding safety information that

should be included in the leaflet. Similarly, 21 CFR 201.22 requires prescription drugs containing sulfites to be labeled with a warning statement about possible hypersensitivity. An informational chapter in USP <1091> entitled "Labeling of Inactive Ingredients" provides guidelines for labeling of inactive ingredients present in dosage forms.

CRITERIA FOR THE SELECTION OF EXCIPIENT AND SUPPLIER

During the development of parenteral dosage forms, the formulator selects excipients that will provide a stable, efficacious, and functional product. The choice, and the characteristics of excipients should be appropriate for the intended purpose.

An explanation should be provided with regard to the function of all constituents in the formulation, with justification for their inclusion. In some cases, experimental data may be necessary to justify such inclusion, e.g., preservatives. The choice of the quality of the excipient should be guided by its role in the formulation and by the proposed manufacturing process. In some cases, it may be necessary to address and justify the quality of certain excipients in the formulation (52).

Normally, a pharmaceutical development report is written in the United States, which should be available at the time of Pre-Approval Inspection (PAI). The development report contains the choice of excipients, their purpose and levels in the drug product, compatibility with other excipients, drug or package system, and how they may influence the stability and efficacy of the finished product.

The following key points should be considered in selecting an excipient and its supplier for parenteral products:

1. Influence of excipient on the overall quality, stability, and effectiveness of drug product.
2. Compatibility of excipient with drug and the packaging system.
3. Compatibility of excipient with the manufacturing process. For example, preservatives may be adsorbed by rubber tubes or filters, acetate buffers will be lost during lyophilization process, etc.
4. The amount or percentage of excipients that can be added to the drug product. Table 6 summarizes the maximum amount of preservatives and antioxidants allowed by various pharmacopoeias.

Table 13 Excipients for label and corresponding information for leaflet

Name	Information on leaflet
Arachis oil	Whenever arachis oil appears, peanut oil should appear beside it (because some individuals are sensitive to peanuts)
Benzoic acid and benzoates	It may increase the risk of jaundice in newborn babies
Benzyl alcohol	Contraindicated in infants or young children; up to 3 years old
Boric acid its salts and esters	Contraindicated in infants or young children; up to 3 years old
Dimethyl sulfoxide	Can cause stomach upset, diarrhea, drowsiness, and headache
Lactose	Unsuitable for people with lactose insufficiency, galactosemia, or glucose/galactose malabsorption syndrome
Organic mercury compounds	Can cause kidney damage
Parahydroxybenzoate and their esters	Known to cause urticaria. Generally delayed type reactions, such as contact dermatitis
Phenylalanine	Harmful for people with phenylketonuria
Polyethoxylated castor oils	Warning for parenterals only—hypersensitivity, drop in blood pressure, inadequate circulation, dyspnea, hot flushes
Potassium	For products administered iv—can cause pain at the site of injection or phlebitis
Sodium	May be harmful to people on low sodium diet
Sorbitol	Unsuitable in hereditary fructose intolerance
Sucrose (saccharose)	Unsuitable in hereditary fructose intolerance, glucose/galactose malabsorption syndrome, or sucrase-isomaltase deficiency
Sulphites (metabisulphites)	Can cause allergic-type reactions including anaphylactic symptoms and bronchospasm in susceptible people, especially those with a history of asthma or allergy
Urea	For products given iv—may cause venous thrombosis or phlebitis

- Route of administration. The *USP*, *EP*, and *BP* do not allow preservatives to be present in injections intended to come in contact with brain tissues or CSF. Thus intracisternal, epidural, and intradural injections should be preservative free. Also, it is preferred that a drug product to be administered via intravenous (iv) route be free of particulate matter. However, if the size of the particle is well controlled, like in fat emulsion or colloidal albumin or amphotericin B dispersion, it can be administered by iv infusion.
- Dose volume. All LVPs and those SVPs where the single dose injection volume can be greater than 15 ml are required by the *EP/BP* to be preservative free (unless justified). The *USP* recommends that special care be observed in the choice and the use of added substances in preparations for injections that are administered in volumes exceeding 5 ml.
- Whether the product is intended for single or multiple dose use. According to *USP*, single dose injections should be free of preservative. The *FDA* takes the position that even though a single dose injection may have to be aseptically processed, the manufacturer should not use a preservative to prevent microbial growth. European agencies have taken a more lenient attitude on this subject.
- The length or duration of time that the drug product will be used once the multidose injection is opened.
- How safe is the excipient?
- Does the parenteral excipient contain very low levels of lead, aluminum, or other heavy metals?
- Does a dossier or DMF exist for the excipient?
- Has the excipient been used in humans? Has it been used via a parenteral route and in the amount and concentration that is being planned?
- Has the drug product that contains this excipient been approved throughout the world?
- What is the cost of the excipient and is it readily available?
- Is the excipient vendor following the *IPEC GMP* guide? Is the vendor *ISO 9000* certified?
- Will the excipient supplier certify the material to meet *USP*, *BP*, *EP*, *JP*, and other pharmacopoeias?
- Has the supplier been audited by the *FDA* or the company's audit group? How did it fare?

Presence of impurities in excipients can have a dramatic influence on the safety, efficacy or stability of the drug product. Monomers or metal catalysts used during a polymerization process are toxic and can also destabilize the drug product if present in trace amounts. Due to safety

concerns, the limit of vinyl chloride (monomer) in polyvinyl pyrrolidone is nmt 10 ppm, and for hydrazine (a side product of polymerization reaction) nmt 1 ppm. Monomeric ethylene oxide is highly toxic and can be present in ethoxylated excipients such as PEGs, ethoxylated fatty acids, etc.

The FDA has issued a guidance suggesting that animal-derived materials such as egg yolk lecithin, and egg phospholipid) used in drug products, originating from Belgium, France, and the Netherlands between January and June 1999 should be investigated for the presence of dioxin and polychlorinated biphenyls. The contamination in the animal-derived product was probably due to contaminated animal feed.

Excipients manufactured by fermentation processes, such as dextrose, citric acid, mannitol, and trehalose, should be specially controlled for endotoxin levels. Mycotoxin (highly toxic metabolic products of certain fungi species) contamination of an excipient derived from natural material has not been specifically addressed by regulatory authorities. The German health authority issued a draft guideline in 1997 where a limit was specified for Aflatoxins M₁, B₁, and the sum of B₁, B₂, G₁, and G₂ in the starting material for pharmaceutical products.

Heavy metal contamination of excipients is a concern, especially for sugars, phosphate, and citrate. Several rules have been proposed or established. For example, the EP sets a limit of nmt 1 ppm of nickel in polyols. California Proposition 65 specifies a limit of nmt 0.5 µg of lead per day per product (53). Similarly, the FDA has proposed a guideline that would limit the aluminum content for all LVPs used in TPN therapy to 25 µg/L (54). Furthermore, it requires that the maximum level of aluminum in SVPs intended to be added to LVPs and pharmacy bulk packages, at expiration date, be stated on the immediate container label.

Physical and chemical stability of the excipient should be considered in assigning a reevaluation date. Since many drug products have a small amount of active and a comparatively high amount of excipients, degradation of even a small percentage of excipient can lead to levels of impurities sufficient to react or degrade a large percentage of active material. For example, benzyl alcohol decomposes via free radical mechanism in the presence of light and oxygen, to form benzaldehyde (x% of benzaldehyde is approximately equivalent to 1/3 x% of hydrogen peroxide). Hydrogen peroxide can rapidly oxidize sulfhydryl groups of amino acids such as cysteine present in peptides or proteins.

It is essential that adequate research and thought be given in the selection of a pharmaceutical excipient

supplier. It is not uncommon for the supplier to change its manufacturing process to make products more efficiently (i.e., less costly). Normally, excipients are low-value, high-volume products that are used by several industries. The pharmaceutical industry, in general, is not the major customer of excipients (in terms of volume of material purchased). For example, the pharmaceutical industry uses approximately 20% of gelatin produced. Of this 20%, most is for production of oral dosage forms. The parenteral portion is approximately 5% of this 20%. Therefore, it is extremely important that the drug manufacturer has a contract with the excipient supplier, that prohibits the supplier from making any change in the process/quality of the material without informing their customers well in advance. Also, the pharmaceutical manufacturer should investigate all the alternate sources that could be used in case of an emergency. A change in the supplier should not be made without consulting the pertinent regulatory bodies, since such an event may require prior regulatory approval.

The pharmaceutical manufacturer should have an active Vendor Certification Program. The manufacturer also should assure that the vendor is ISO 9000 certified. An audit of the excipient manufacturer is essential, since the pharmaceutical industry is ultimately responsible for the quality of the drug product that includes the excipient(s) as one of the components. The IPEC GMP guide may be used as an audit tool, since it is written in the format of ISO 9000 using identical nomenclature and paragraph numbering. The audit may ensure that the quality is being built into the excipient that may be difficult to measure later by quality control on receipt of the material. This is especially true for parenteral excipients where not only chemical, but also microbiological attributes are critical. Bioburden and endotoxin limits may be needed for each of the excipients and several guidelines are available to establish the specifications (55, 56).

Recent events in Haiti highlight the importance of assuring the quality of excipients to the same degree that one normally does for active ingredients. From November 1995 through June 1996, acute anuric renal failure was diagnosed in 86 children. This was associated with the use of diethylene glycol-contaminated glycerin used to manufacture acetaminophen syrup (57).

SAFETY ISSUES

Reference 58 is an excellent resource on the safety and adverse reaction to several excipients. Sensitization reactions have been reported for the parabens, thimerosal,

and propyl gallate. Sorbitol is metabolized to fructose and can be dangerous when administered to fructose-intolerant patients. Table 13 also lists safety concerns.

Progress in drug delivery systems and new protein-s/peptides being developed for parenteral administration has created a need to expand the list of excipients that can be safely used. An informational chapter included in the USP 24, presents a scientifically based approach for safety assessment of new pharmaceutical excipients (59). This chapter is based on the excipient safety evaluation guidelines prepared by The Safety Committee of the International Pharmaceutical Excipient Council, with appropriate redaction. Table 14 summarizes the approach in developing a new excipient.

Currently, there are some concerns regarding Transmissible Spongiform Encephalopathies (TSE) via animal-derived excipients such as gelatin. TSEs are caused by prions that are extremely resistant to heat and normal sterilization processes. TSEs have a very long incubation time with no cure and include diseases such as the following:

- Scrapies in sheep and goats
- Bovine spongiform encephalopathy (BSE), otherwise known as Mad Cow Disease, in cattle
- Kuru disease in humans
- Creutzfeld-Jacob disease (CJD) in humans, which has been attributed to repeated parenteral administration of growth hormone and gonadotropin derived from human pituitary glands.

Several guidelines have been issued that address the issue of animal-derived excipients and scientific principles to minimize the possible transmission of TSEs via medicinal products (60, 61). The current situation indicates that there are negligible concerns for lactose, glycerol, fatty acids, and their esters, but the situation is less clear for gelatin. In this scenario, if one has a choice, then it may be beneficial to select nonanimal-derived excipients. The use of bovine serum albumin (BSA) or human serum albumin (HSA) is of concern because they can be derived from virus-contaminated blood. The risk of TSEs from excipients can be greatly reduced by controlling the following parameters:

1. Source of animal should be from countries where BSE has not been reported.
2. Animals used should be young.
3. Category III or IV animal tissue should be used in manufacture (60).
4. A production process that is likely to destroy TSE agents should be utilized.

Table 14 Summary of safety evaluation of excipient

Tests	Injectable route ^a
Baseline toxicity data	
Acute oral toxicity	Required
Acute dermal toxicity	Required
Acute inhalation toxicity	Conditional
Eye irritation	Required
Skin irritation	Required
Skin sensitization	Required
Acute injectable toxicity	Required
Application site evaluation	Required
Phototoxicity/photoallergy	Required
Genotoxicity assays	Required
ADME/PK-intended route	Required
28-day toxicity (2 species) intended route	Required
Additional data: Short or intermediate term repeated use	
90-day toxicity (most appropriate species)	Required
Embryo-fetal toxicol	Required
Additional assays	Conditional
Genotoxicity assays	Required
Immunosuppression assays	Required
Additional data: Intermittent long-term or chronic use	
Chronic toxicity (rodent, nonrodent)	Conditional
Reproductive toxicity	Required
Photocarcinogenicity	Conditional
Carcinogenicity	Conditional

^aTerm injectable includes routes such as iv, sc, intrathecal, etc.

Amendment to the European Commission directive 75/318/EEC would require manufacturers to provide a “Certificate of Suitability” or the underlying “scientific information” in the form of a marketing variation to attest that their pharmaceuticals are free of TSEs.

FUTURE DIRECTION

Several new excipients are being evaluated in order to increase the solubility or improve the stability of parenteral drugs. Cyclodextrins have been tried for the above reasons. Currently, there are two FDA approved parenteral products that have utilized α and γ -cyclodextrins. β -cyclodextrin is unsuitable for parenteral administration because it causes necrosis of the proximal kidney tubules upon IV and subcutaneous administration (62). Hydroxypropyl β -cyclodextrin (HP β CD) and sulfobutylether β -cyclodextrin

(SBE-7- β -CD) have shown the most promise. CaptisolTM is the trade name of SBE-7- β -CD and is anionic. Currently, two CaptisolTM based small molecule IV and IM drug formulations are in Phase III clinical trials in the United States. One parenteral formulation that utilizes HP β CD (Cavitron[®]) is in Phase II/III clinical trials, and another (Sporanox) has been approved by the FDA. Manufacturers of HP β CD and SBE-7- β -CD have established a DMF with the FDA. A detailed review of cyclodextrins was recently published (63, 64). It should be noted, however, that cyclodextrin also can accelerate the degradation of drug product (65) and can sequester preservatives, rendering them ineffective (66).

Chitosan, β -1,4-linked glucosamine, is a naturally occurring, biodegradable, nontoxic polycationic biopolymer. It is being investigated for its potential as a cross-linked matrix of microspheres to deliver antineoplastic drugs. Because of its charge, it can trap several drugs and can bind strongly with cancer cells, thereby minimizing drug toxicity and enhancing therapeutic efficacy (67). Chitosan also has been shown to stabilize liposomes.

Biodegradable polymeric materials such as polylactic acid, polyglycolic acid, and other poly-alpha-hydroxy acids have been used as medical devices and as biodegradable sutures since the 1960s (68). Currently, the FDA has approved for marketing, only devices made from homopolymers or copolymers of glycolide, lactide, caprolactone, *p*-dioxanone, and trimethylene carbonate (69). Such biopolymers are finding increased application as a matrix to deliver parenteral drugs for prolonged delivery (70). At least four drug products—Lupron Depot[®], Decapeptyl[®], Nutropin Depot[®], and Zoladex[®]—have been approved. These four drug products are microspheres in PLG, polylactic acid (PLA), or the PLGA matrix. Polyglycolic acid (PGA) is highly crystalline (approximately 50%) with a high melting point (220–225°C). PLA can be produced by the polymerization of L-lactic acid (LPLA), D-lactic acid (DPLA), or a blend of D- and L- lactic acid (DLPLA). LPLA is 37% crystalline while DLPLA, is amorphous. The degradation time of LPLA is much slower than that of DPLA requiring more than 2 years. By copolymerizing lactic and glycolic acid, polymeric matrices with a wide range of properties (tensile strength, crystallinity, and degradation rate) can be obtained. Decapeptyl[®] is approved in France and is a microsphere for IM administration. It contains drug in a matrix of PLGA and Carboxymethyl cellulose with mannitol and polysorbate 80.

Polyanhydrides degrade primarily by surface erosion and possess excellent in vivo compatibility. In 1996 the

FDA approved a polyanhydride-based drug delivery system to the brain of chemotherapeutic agent BCNU, which is currently being manufactured by Guilford Pharmaceutical, Inc.

Several phospholipid-based excipients are finding increased application as solubilizing agents, emulsifying agents, or as components of liposomal formulation. The phospholipids occur naturally and are biocompatible and biodegradable. Examples include egg phosphatidylcholine, soybean phosphatidylcholine, hydrogenated soybean phosphatidylcholine (HSPC), DMPC, DSPC, DOPC, DSPE, DMPG, DPPG, and DSPG. SpartajectTM technology uses a mixture of phospholipids, to encapsulate poorly water-soluble drugs, to form micro-suspensions that can be injected intravenously. Busulfan[®] drug product uses this technology and is currently undergoing Phase I clinical trials. Many liposomal and liposomal-like formulations (DepoFoam[®]) are either approved (Depo-Cyt[®]) or are undergoing clinical trials to reduce drug toxicity, improve drug stability, prolong the duration of action, or to deliver drug to the central nervous system (71). Two amphotericin formulations have been approved in the United States. They are liposomal, or a lipid complex between the antifungal drug and positively charged lipid. Amphotec[®] is a 1:1 molar ratio complex of amphotericin B and cholesteryl sulfate while Abelcet[®] is a 1:1 molar complex of amphotericin B with phospholipids (seven parts of L- α -dimyristoylphosphatidylcholine and L- α -dimyristoylphosphatidyl glycerol).

Poloxamers or pluronics are block copolymers comprised of polyoxyethylene and polyoxypropylene segments. They exhibit reverse thermal gelation and are being tried as solubilizing, emulsifying, and stabilizing agents. Thus, a depot drug delivery system can be created using pluronics whereby the product is a viscous injection that gels upon IM injection (72). Pluronic can prevent protein aggregation or adsorption/absorption and can help in the reconstitution of lyophilized products. Pluronic F68 (Polaxamer-188), F38 (Poloxamer-108), and F127 (Poloxamer-407) are the most commonly used pluronics. For example, liquid formulation of human growth hormone and Factor VIII can be stabilized using pluronics. Fluosol[®] is a complex mixture of perfluorocarbons, with a high oxygen-carrying capacity emulsified with Pluronic F-68, and various lipids. It was recently approved by the FDA for adjuvant therapy to reduce myocardial ischemia during coronary angioplasty. A highly purified form of Poloxamer 188 (FlocorTM), intended for IV administration, is undergoing Phase III clinical trials for various cardiovascular diseases. Purification of Poloxamer 188 has been shown to reduce nephrotoxicity.

Poloxamers and other polymeric materials such as albumin may coat the micro- or nano particle, alter their surface characteristics and reduce their phagocytosis and opsonization by the reticuloendothelial system following IV injection. Such surface modifications often result in prolongation in the circulation time of intravenously injected colloidal dispersions (73). Poloxamers also have been used to stabilize suspension such as NanoCrystal™ (74).

The first successful development of an injectable perfluorocarbon-based commercial product was achieved by the Green Cross Corporation in Japan, when it made Fluosol-DA®, a dilute (20% w/v) emulsion based on perfluorodecalin and perflurotripropylamine emulsified with potassium oleate, Pluronic F-68, and egg yolk lecithin. These perfluorocarbons are inert and also can be used to formulate nonaqueous preparations of insoluble proteins and small molecules (75). Perfluorocarbons also have been approved by the FDA for use in one ultrasound contrast agent, Optison®, which is administered via the IV route. Optison® is a suspension of microspheres of HSA with octafluoropropane. Heat treatment and sonication of appropriately diluted human albumin, in the presence of octafluoropropane gas, is used to manufacture microspheres in the Optison® injection. The protein in the microsphere shell makes up approximately 5–7 (wt%) of the total protein in the liquid. The microspheres have a mean diameter range of 2.0–4.5 µm with 93% of the microsphere being less than 10 µm.

Sucrose acetate isobutyrate (SAIB) is a high viscosity liquid system that converts into free-flowing liquid when mixed with 10–15% ethanol (76). On subcutaneous or IM injection, the matrix rapidly converts to a water-insoluble semi-solid, that is capable of delivering proteins and small molecules for a prolonged period. SAIB is biocompatible, and biodegrades to natural metabolites. This is a fairly new matrix and three INDs have been filed for veterinary applications. It has not been used in humans.

Several other biodegradable, biocompatible, injectable polymers are being investigated for drug delivery systems. They include polyvinyl alcohol, block copolymer of PLA–PEG, polycyanoacrylate, polyanhydrides, cellulose, alginate, collagen, gelatin, albumin, starches, dextrans, hyaluronic acid and its derivatives, and hydroxyapatite (77).

REFERENCES

1. Parenteral Preparations. *European Pharmacopoeia*, 3rd Ed.; Council of Europe: Strasbourg, 1997; 1765.
2. Parenteral Preparations. *British Pharmacopoeia*; Stationary Office: London, 1999; II, 1575.
3. Uchiyama, M. Regulatory Status of Excipients in Japan. *Drug Inf. J.* **1999**, *33*, 27–32.
4. Nema, S.; Washkuhn, R.J.; Brendel, R.J. Excipients and their Use in Injectable Products. *PDA J. Pharm. Sci. Technol.* **1997**, *51* (4), 166–71.
5. Wang, Y.J.; Kowal, R.R. Review of Excipients and pHs for Parenteral Products Used in United States. *J. Parenter. Sci. Technol.* **1980**, *34* (6), 452.
6. Powell, M.F.; Nguyen, T.; Baloian, L. Compendium of Excipients for Parenteral Formulations. *PDA J. Pharm. Sci. Technol.* **1998**, *52* (5), 236–311.
7. Wang, Y.J.; Hanson, M.A. Parenteral Formulations of Proteins and Peptides: Stability and Stabilizers. *J. Parenter. Sci. Technol.* **1988**, *42* (supplement), S4–S26.
8. Boylan, J.C.; DeLuca, P.P. Formulation of Small Volume Parenterals. *Pharmaceutical Dosage Forms: Parenteral Medications*, 2nd Ed.; Avis, K.E., Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.; New York, 1992; 1, 173–48.
9. Strickley, R.G. Parenteral Formulations of Small Molecules Therapeutics Marketed in the United States (1999)—Part 1. *PDA J. Pharm. Sci. Technol.* **1999**, *53* (6), 324–349.
10. *Physician's Desk Reference*, Medical Economics Co.: Montvale, 1994, 1996, 1998 & 1999.
11. Trissel, L.A. *Handbook on Injectable Drugs*, 10th Ed.; American Society of Health-System Pharmacists, Inc.: Bethesda, 1998.
12. Kibbe, A.H. *Handbook of Pharmaceutical Excipients*, 3rd Ed.; The Pharmaceutical Press: London, 2000.
13. Mosby's, GenRx (Ed.) 8th Ed. Mosby-Year Book, Inc.: St. Louis MO, 1998.
14. Inactive Ingredient Guide, Division of Drug Information Resources, Food & Drug Administration, CDER, January 1996.
15. Sweetana, S.; Akers, M.J. Solubility Principles and Practices for Parenteral Drug Dosage Form Development. *PDA J. Parenter. Sci. Technol.* **1996**, *50* (5), 330–342.
16. Yalkowsky, S.H.; Roseman, T.J. Solubilization of Drugs by Cosolvents. *Techniques of Solubilization of Drugs*; Marcel Dekker, Inc.: New York, 1981; 91–134.
17. Rubino, J.T.; Yalkowsky, S.H. Cosolvency and Cosolvent Polarity. *Pharm. Res.* **1987**, *4* (3), 220–230.
18. Hancock, B.C.; York, P.; Rowe, R.C. The Use of Solubility Parameters in Pharmaceutical Dosage Form Design. *Int. J. Pharm.* **1997**, *148*, 1–21.
19. Reed, K.W.; Yalkowsky, S. Lysis of Human Red Blood Cells in the Presence of Various Cosolvents. *J. Parenter. Sci. Technol.* **1985**, *39* (2), 64.
20. Brazeau, G.A.; Fung, H. Use of an In-vivo Model for the Assessment of Muscle Damage from Intramuscular Injections: In-vitro–In-vivo Correlation and Predictability With Mixed Solvent Systems. *Pharm. Res.* **1989**, *6* (9), 766.
21. Brazeau, G.A.; Cooper, B.; Svetic, K.A.; Smith, C.L.; Gupta, P. Current Perspectives on Pain Upon Injection of Drugs. *J. Pharm. Sci.* **1998**, *87* (6), 667.
22. Yalkowsky, S.H.; Krzyzaniak, J.F.; Ward, G.H. Formulation-Related Problems Associated with Intravenous Drug Delivery. *J. Pharm. Sci.* **1998**, *87* (7), 787.
23. Johnson, D.M.; Gu, L.C. Autoxidation and Antioxidants. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: NY, 1988; 1, 415–449.

24. Herman, A.C.; Boone, T.C.; Lu, H.S. Characterization, Formulation, and Stability of Neupogen® (Filgrastim), a Recombinant Human Granulocyte-Colony Stimulating Factor. *Formulation, Characterization, and Stability of Protein Drugs: Case Histories*; Pearlman, R., Wang, Y.J., Eds.; Plenum Press: NY, 1996; 9, 325.
25. Fatouros, A.; Osterberg, T.; Mikaelsson, M. Recombinant Factor VIII SQ: Influence of Oxygen, Metal Ions, pH and Ionic Strength on its Stability in Aqueous Solution. *Int. J. Pharm.* **1997**, *155*, 121–131.
26. Stadtman, E.R. Metal Ion Catalyzed Oxidation of Proteins: Biochemical Mechanism and Biological Consequences. *Free Radical Biol. Med.* **1990**, *9*, 315.
27. Munson, J.W.; Hussain, A.; Bilous, R. Precautionary Note for Use of Bisulfite in Pharmaceutical Formulations. *J. Pharm. Sci.* **1977**, *66* (12), 1775–1776.
28. Enever, R.P.; Po, A.L.W.; Shotton, E. Factors Influencing Decomposition Rate of Amitriptyline Hydrochloride in Aqueous Solution. *J. Pharm. Sci.* **1977**, *66* (8), 1087–1089.
29. Akers, M.J. Antioxidants in Pharmaceutical Products. *J. Parenter. Sci. Technol.* **1982**, *36* (5), 222–228.
30. *Note for Guidance on Inclusion of Antioxidants and Antimicrobial Preservatives in Medicinal Products* CPMP: January 1998.
31. Dabbah, R. The Use of Preservatives in Compendial Articles. *Pharmacopeial Forum* **1996**, *22* (4), 2696.
32. *Martindale: The Extra Pharmacopoeia*, 31st Ed.; Royal Pharmaceutical Society: London, 1996; 1128.
33. *British Pharmaceutical Codex*; Royal Pharmaceutical Society: London, 1973; 100.
34. USP (51) Antimicrobial Effectiveness Testing. *United States Pharmacopeia*; 24; US Pharmacopeial Convention, Inc.: Rockville, 2000, 1809.
35. Efficacy of Antimicrobial Preservation. *European Pharmacopoeia*, 3rd Ed.; Council of Europe: Strasbourg, 1997; 286.
36. Dabbah, R. Harmonization of Microbiological Methods—A Status Report. *Pharmacopeial Forum* **1997**, *23* (6), 5334–5344.
37. *Note for Guidance on Maximum Shelf-life for Sterile Products for Human Use After First Opening or Following Reconstitution*; CPMP: July 1998.
38. Lam, X.M.; Costantino, H.R.; Overcashier, D.E.; Nguyen, T.H.; Hsu, C.C. Replacing Succinate with Glycolate Buffer Improves the Stability of Lyophilized Interferon- γ . *Int. J. Pharm.* **1996**, *142*, 85–95.
39. Pikal M. The Correlation of Structural Relaxation Time With Pharmaceutical Stability Freeze-Drying of Pharmaceuticals and Biologicals Conference Brownsville VT Sept. 23–26 1998.
40. Arakawa, T.; Kita, Y.; Carpenter, J.F. Protein–Solvent Interactions in Pharmaceutical Formulations. *Pharm. Res.* **1991**, *8* (3), 285–291.
41. Miller, D.P.; Anderson, R.E.; de Pablo, J.J. Stabilization of Lactate Dehydrogenase Following Freeze-Thawing and Vacuum-Drying in the Presence of Trehalose and Borate. *Pharm. Res.* **1998**, *15* (8), 1215–1221.
42. Carpenter, J.F.; Crowe, J.H. Modes of Stabilization of a Protein by Organic Solutes During Desiccation. *Cryobiology* **1998**, *25*, 459–470.
43. Baumann, T.J.; Smythe, M.A.; Kaufmann, K.; Miloboszewski, Z.; O'Malley, J.; Fudge, R.P. Dissolution Times of Adriamycin and Adriamycin RDF. *Am. J. Hosp. Pharm.* **1988**, *45*, 1667.
44. Jain, N.K.; Jain, S.; Singhai, A.K. Enhanced Solubilization and Formulation of an Aqueous Injection of Piroxicam. *Pharmazie* **1997**, *52* (12), 942–946.
45. Meyer, J.D.; Manning, M.C. Hydrophobic Ion Pairing: Altering the Solubility Properties of Biomolecules. *Pharm. Res.* **1998**, *15* (2), 188–192.
46. Wang, Y.J.; Dahl, T.C.; Leesman, G.D.; Monkhouse, D.C. Optimization of Autoclave Cycles and Selection of Formulation for Parenteral Product, Part II: Effect of Counter-Ion on pH and Stability of Diatrizoic Acid At Autoclave Temperatures. *J. Parenter. Sci. Technol.* **1984**, *38* (2), 72.
47. Japan Pharmaceutical Excipient Council, Eds. *Japanese Pharmaceutical Excipients Directory 1996*; Yakuji Nippo, Ltd.: Tokyo, 1996.
48. Excipients in Pharmaceutical Dosage Forms: The Challenge of the 21st Century Conference Proceedings Nice France May 14–15 1998.
49. Benzyl Alcohol. *Pharmacopeial Forum* **Sept.–Oct. 1995**, *21* (5), 1240.
50. USP (1078) Good Manufacturing Practices for Bulk Pharmaceutical Excipients. *United States Pharmacopeia*, 24 Ed.; US Pharmacopeial Convention, Inc.: Rockville, 2000; 2040.
51. Excipients in the Label and Package Leaflet of Medicinal Products for Human Use. *The Rules Governing Medicinal Products in the European Union*; Guidelines: Medicinal Products for Human Use, European Commission: September 1997; 3B.
52. *Note for Guidance on Development Pharmaceuticals Committee for Proprietary Medicinal Products (CPMP)*: July 1998.
53. Paul, W.L. Excipient Intake and Heavy Metals Limits. *Pharmacopeial Forum* **1995**, *21* (6), 1629.
54. Aluminum in Large and Small Volume Parenterals Used in Total Parenteral Nutrition. *Federal Register* January 5 **1998**, *63* (2), 176–185.
55. *Guideline on Validation of the Limulus Amebocyte Lysate Test as an End-Product Test for Human and Animal Parenteral Drugs, Biological Products, and Medical Devices*; Food & Drug Administration: December 1987.
56. Opalchenova, G.A. Comparison of the Microbial Limit Tests in the British, European, and United States Pharmacopeias and Recommendation for Harmonization. *Pharmacopeial Forum* **1994**, *20* (4), 7872–7877.
57. US Department of Health and Human Services. Morbidity and Mortality Weekly Report. August 2 **1996**, *45* (30), 649–650.
58. Weiner, M.; Bernstein, I.L. *Adverse Reactions to Drug Formulation Agents: A Handbook of Excipients*; Marcel Dekker, Inc.: New York NY, 1989.
59. USP (1074) Excipient Biological Safety Evaluation Guidelines. *United States Pharmacopeia*, 24 Ed.; US Pharmacopeial Convention Inc.: Rockville, 2000; 2037.
60. *Note for Guidance on Minimizing the Risk of Transmitting Animal Spongiform Encephalopathy Agents Via Medicinal Products*. CPMP; April 21, 1999.
61. The Sourcing and Processing of Gelatin to Reduce the Potential Risk Posed by Bovine Spongiform Encephalo-

- pathy (BSE) in FDA-Regulated Products for Human Use, Guidance for Industry, US Dept. of Health and Human Services, FDA, Sept. 1997.
62. Frank, D.W.; Gray, J.E.; Weaver, R.N. Cyclodextrin Nephrosis in the Rats. *Am. J. Pathol.* **1976**, *83*, 367.
 63. Stella, V.J.; Rajewski, R.A. Cyclodextrins: Their Future in Drug Formulation and Delivery. *Pharm. Res.* **1997**, *14* (5), 556–567.
 64. Thompson, D.O. Cyclodextrins-Enabling Excipients: Their Present and Future Use in Pharmaceuticals: Critical Reviews in Therapeutic Drug Carrier Systems. **1997**, *14* (1), 1–104.
 65. Loftsson, T.; Johannesson, H.R., *Die Pharmazie* **1994**, *49*, 292–293.
 66. Lehner, S.J.; Muller, B.W.; Seydel, J.K. Effect of Hydroxylpropyl-Beta-Cyclodextrin on the Antimicrobial Action of Preservatives. *J. Pharm. Pharmacol.* **1994**, *46*, 186–191.
 67. Felt, O.; Buri, P.; Gurny, R. Chitosan: A Unique Polysaccharide for Drug Delivery. *Drug Develop. Ind. Pharm.* **1998**, *24* (11), 979–993.
 68. Jain, R.; Shah, N.H.; Malick, A.W.; Rhodes, C.T. Controlled Drug Delivery by Biodegradable Poly(Ester) Devices: Different Preparative Approaches. *Drug Develop. Ind. Pharm.* **1998**, *24* (8), 703–727.
 69. Middleton, J.C.; Tipton, A.J. Synthetic Biodegradable Polymers as Medical Devices. *Med. Plastics Biomaterial* **1998**, *5* (2).
 70. Pettit, D.K.; Lawter, J.R.; Huang, W.J.; Pankey, S.C.; Nightlinger, N.S.; Lynch, D.H.; Schuh, J.A.C.L.; Morrissey, P.J.; Gombotz, W.R. Characterization of Poly(glycolide-co-D,L-lactide)/Poly(D,L-lactide) Microspheres for Controlled Release of GM-CSF. *Pharm. Res.* **1997**, *14* (10), 1422–1430.
 71. Katre, N.V.; Asherman, J.; Schaefer, H. Multivesicular Liposome (DepoFoam™) Technology for the Sustained Delivery of Insulin-Like Growth Factor-I. *J. Pharm. Sci.* **1998**, *87* (11), 1341–1346.
 72. Wang, P.; Johnston, T.P. Sustained-Release Interleukin-2 Following Intramuscular Injection in Rats. *Int. J. Pharm.* **1995**, *113* (1), 73–81.
 73. Moghimi, S.M. Mechanisms Regulating Body Distribution of Nanospheres Conditioned with Pluronic and Tetronic Block Co-Polymers. *Adv. Drug Deliv. Rev.* **1995**, *16*, 183–193.
 74. Zheng, J.Y.; Bosch, H.W. Sterile Filtration of NanoCrystal™ Drug Formulations. *Drug Develop. Ind. Pharm.* **1997**, *23* (11), 1087–1093.
 75. Knepp, V.M.; Muchnik, A.; Oldmark, S.; Kalashnikova, L. Stability of Nonaqueous Suspension Formulations of Plasma Derived Factor IX and Recombinant Human Alpha Interferon at Elevated Temperatures. *Pharm. Res.* **1998**, *15* (7), 1090–1095.
 76. Sullivan, S.A.; Gilley, R.M.; Gibson, J.W.; Tipton, A.J. Delivery of Taxol and Other Antineoplastic Agents from a Novel System Based on Sucrose Acetate Isobutyrate. *Pharm. Res.* **1997**, *12* (11), 291.
 77. Gombotz, W.R.; Pettit, D.K. Biodegradable Polymers for Proteins and Peptide Drug Delivery. *Bioconjugate Chem.* **1995**, *6*, 332–351.

EXCIPIENTS—SAFETY TESTING

Marshall Steinberg

International Pharmaceutical Excipients Council-Americas, Arlington, Virginia, U.S.A.

Florence K. Kinoshita

Hercules Incorporated, Arlington, Virginia, U.S.A.

INTRODUCTION

The safety issues concerning pharmaceutical excipients can be classified into three categories: quality, toxicology, and improper use (1). Various regulatory directives address the quality category. In addition, the International Pharmaceutical Excipient Council's (IPEC) guideline publications address this issue by following the Organization of International Standardization (ISO) 9000 structure. IPEC is an industry association with worldwide membership that includes over 200 pharmaceutical, chemical, and food processing firms that develop, manufacture, sell, and use pharmaceutical excipients. IPEC comprises three regional organizations located in the United States, Europe, and Japan each with the same objectives.

Quality

The aforementioned IPEC guidelines that address quality include:

- Good Manufacturing Guide for Bulk Pharmaceutical Excipients
- Good Manufacturing Practices (GMP) Audit Guideline for Distributors of Bulk Pharmaceutical Excipients
- IPEC-Americas Significant Change Guide for Bulk Pharmaceutical Excipients
- GMP Audit Guideline for Suppliers of Bulk Pharmaceutical Excipients
- New Excipient Safety Evaluation Guidance
- IPEC-Americas Guide for the Development of an Impurity Profile
- Format and Required Content of Certificates of Analyses

The IPEC-Americas Safety Guidelines (modified) are presented as an information chapter in United States Pharmacopoeia (USP) 24/NF 19 (2). The Good Manufacturing Guide or Practices for Bulk Pharmaceutical

Excipients also has been published as an information chapter in USP 24/NF 19. The guideline also serves as a basis for the World Health Organization (WHO) guidance to its national members (3). The intention of IPEC is to ensure that these guidelines reflect the concerns and intentions of responsible parties in the United States, the European Union (EU), and Japan. In other words, the guidelines are harmonized so that excipients that meet the requirements of a harmonized monograph can be sold and used in these three areas of the world. The use of these and other national guidelines ensure the quality of excipients.

These guidelines, using an ISO 9000 format, not only provide a way to assess whether systems are in place, but provide a means for evaluating the effectiveness of the systems. They also provide guidance on how to conduct an audit of a manufacturing operation that produces excipients (4) and in turn, give guidelines on auditing their distribution and repackaging (5).

International Conference on Harmonization Residual Solvent Guidance

Excipient impurity profiles and how to evaluate this important aspect of excipient manufacture, particularly in light of the International Conference on Harmonization (ICH) guidance published in 1999 (6), also are addressed. Care must also be taken that residual solvent levels do not exceed those proscribed in the ICH Guidance for Residual Solvents published in 1999. Solvents are divided into three classes:

1. *Class 1 solvents: Solvents to be avoided.* These include known human carcinogens, strongly suspected human carcinogens, and environmental hazards.
2. *Class 2 solvents: Solvents to be limited.* These include nongenotoxic animal carcinogens or possible causative agents of other irreversible toxicity, such as neurotoxicity or teratogenicity and solvents suspected of other significant but reversible toxicities.

3. *Class 3 solvents: Solvents with low toxic potential.* These include solvents with low toxic potential to man; no health-based exposure limit is needed. Class 3 solvents have permitted daily exposures of 50 mg or more per day.

IPEC Significant Change Guidance

Two areas of concern to excipient makers and users have been those of significant change and certificates of analyses. Any change by the manufacturer of an excipient that alters an excipient's physical or chemical property from the norm or that is likely to alter the excipient's performance in the dosage form is considered significant (7). Regardless of whether there is a regulatory requirement to notify the local regulatory authority, the manufacturer has an obligation to notify its customers of significant change so that the customer can evaluate the change on the customer's products. The Significant Change Guidance establishes uniform considerations for evaluating significant changes involving the manufacture of bulk excipients. The types of changes that might be considered include:

- Site
- Scale
- Equipment
- Process
- Packaging
- Specifications

The requirement for evaluating the impact of change on the excipient begins at the processing step from which GMP compliance begins, as noted in the IPEC Good Manufacturing Guide or Practices Guide for Bulk Pharmaceutical Excipients, or later in the process.

The evaluation criteria in the guideline include:

1. Changes in the chemical properties of the excipient owing to the change
2. Changes in the physical properties of the excipient owing to the change
3. Changes in the impurity profile of the excipient owing to the change
4. Changes in the functionality of the excipient owing to the change
5. Changes in the moisture level of the excipient owing to the change
6. Changes in the bioburden of the excipient owing to the change

The guideline also provides for consideration of objective criteria when considering changes to the

impurity profile of an excipient as a result of any change. IPEC-Americas has developed a guide for the preparation of an impurity profile for excipients. The profile addresses the following:

1. All specified organic impurities
2. Unidentified organic impurities at or above 0.1% whether specified or not, unless the impurity has an established pharmacological effect or is known to be unsafe at a lower level
3. Residual solvents
4. Inorganic impurities
5. Toxic impurities

The content of the impurity profile varies with the nature of the excipient, the raw materials used in its manufacture, and its chemical composition. Changes are considered significant whenever a new impurity is introduced at or above the 0.1% concentration or when an impurity previously present at or above 0.1% disappears.

IPEC Certificate of Analysis Guidance

The second issue involves the certificate of analysis that the manufacturer must provide to the formulator when shipping the excipient. Most often, a certificate of analysis does not contain information developed as a result of analysis of the specific batch of material being delivered. The analysis may have either been conducted on previous individual batches or on a mixture of aliquots of previous batches. No guidelines regarding exactly what should be found in the certificate and how it should be presented have been established. This is addressed in the guideline (8). At the time of this writing, the frequency of sampling has not been resolved with the U.S. Food and Drug Administration (FDA). Some believe that in the face of no significant changes it should not be necessary to sample each manufactured batch, but that there is a need only for sufficient sampling to ensure that statistical significance of sampling results can be met. What to do if the manufacturing process is continuous rather than a batch process would fall under the same criteria except that the sampling frequency would probably be based on time/volume rather than batches.

EXCIPIENT USAGE

There are roughly 8000 "nonactive" ingredients being used in food, cosmetics, and pharmaceuticals worldwide (1). In 1996, approximately 800 excipients were used in

marketed pharmaceutical products in the United States (1). Although the FDA maintains a “list” of inactive ingredients, the EU and other European countries do not have official published lists, although steps are being taken to rectify this situation.

Few excipients are manufactured specifically for pharmaceutical use. Many are manufactured for other purposes (e.g., food, cosmetics, paint thickeners, construction, etc.). For their use in pharmaceuticals, additional quality, functionality, and safety requirements must be met.

Improper Use of Excipients

The improper use of excipients is addressed, to a certain extent, by the package inserts found in the formulated products. The challenge is to educate consumers and health providers to read and comply with the information contained in these inserts.

DEFINITION OF AN EXCIPIENT

For toxicological purposes, it may be inappropriate to define excipients as inert ingredients. It may be more appropriate to define an excipient (9) as “Any substance other than the active drug or pro-drug which has been appropriately evaluated for safety and is included in a drug delivery system to either:

1. Aid processing of the system during manufacture
2. Protect, support, or enhance stability, bioavailability
3. Assist in product identification
4. Enhance any other attribute of the overall safety and effectiveness of the drug product during storage and use.”

As the fourth definition indicates, excipients include a multiplicity of activities from mold releasers to absorption enhancers, and more recently include substances that permit large molecule (e.g., proteins) to be absorbed from the gastrointestinal tract without degradation. Most actions by an excipient are mechanical rather than pharmacological.

APPROVAL MECHANISMS FOR EXCIPIENTS

Currently, regulatory agencies have not established safety-testing guidelines specifically for excipients (10–13). Under U.S. law, a new pharmaceutical excipient, unlike an

active drug, has no regulatory status unless it can be qualified through one or more of the approval mechanisms available for components used in finished drug dosage forms. These approval mechanics include:

- Generally Recognized As Safe (GRAS) determination pursuant to 21 Code of Federal Regulations (CFR) 182, 184, and 186
- Approval of a food additive petition under 21 CFR 171
- As contained in a New Drug Application (NDA) approval for a specific drug product and for a particular function or use in that dosage form

Within the EU there is a directive that makes it clear that new chemical excipients will be treated in the same way as new actives (14).

TOXICITY TESTING

The very nature of excipients, for the most part, represents unique problems in testing for toxicity. The actions sought for many excipients are mechanical rather than physiological. Exceptions to this are flavors (12). A most desirable description of some excipients would include being pharmacologically inert and mechanically functional. An alternative would be one where the toxic dose was so high as to be meaningless while still retaining functionality requirements through a range of high and low doses. The acceptable risk for a traditional excipient, when compared to an active principle in a formulation, is generally several orders of magnitude different. Unless an excipient has some very unique properties, it is unlikely that a new excipient would be developed that did not have a large safety factor for toxicity and side effects under conditions of use.

As excipients become more complex and are required to perform functions not required in the past, it is conceivable that a distinction will have to be made between excipients and what might be termed “co-drugs.” The use of monoclonal antibodies to deliver an active principle to a specific tissue site might be considered an example of this diversity.

In 1994, as part of the IPEC-Americas program to obtain stand-alone status for excipients, a safety committee was formed. The committee was composed of men and women from a variety of medical and chemistry disciplines who were directed to develop safety-testing guidelines for new excipients. These guidelines were published in 1996 (12). At that time, regulations in most developed countries did not address registration of an excipient as a separate entity. For example, the drug

master file for an excipient in the United States is reviewed only as part of the NDA process. Inherent to the current process is the assumption that the use of an excipient in an approved drug dosage form ensures its acceptance in other dosage forms and its ultimate inclusion in the National Formulary (NF). The NF monographs provide standards/specifications for identity, purity, and analysis. Priority for inclusion is given to formulations with approved NDAs and those approved for use in foods. The FDA favors the use of commercially established excipients, such as food additives and substances that have been designated GRAS.

The guidelines developed by IPEC-Americas (12) provide for a tier approach to required testing. The tests to be conducted are based upon the route of application of the

formulated drug and the duration of use. A base set of data is required for all candidate excipients. The guidelines require a review of the chemical and physical properties of the excipient and a review of the scientific literature, exposure conditions (including dose, duration, frequency, route, and user population), and absence or presence of pharmacological activity.

Alternatives to the use of living animals are encouraged wherever these procedures have been validated. The information will provide sufficient information upon which to base a safety judgment and the data will be acceptable to a regulatory agency. The studies should also follow the appropriate legal and professional codes (15) in the conduct of all tests and should meet the Good

Table 1 Summary IPEC-America safety testing guidelines

Routes of exposure for humans:						
Tests	Oral	Mucosal transdermal/ injectable topical			Inhalation/ intranasal	Ocular
<i>Baseline toxicity data</i>						
Acute oral toxicity	R	R	R	R	R	R
Acute dermal tox.	R	R	R	R	R	R
Acute inhalation tox.	C	C	C	C	R	C
Eye irritation	R	R	R	R	R	R
Skin irritation	R	R	R	R	R	R
Skin sensitization	R	R	R	R	R	R
Acute parenteral tox.	—	—	—	R	—	—
Application site eval.	—	R	R	R	R	—
Pulmonary sensitization	—	—	—	—	R	—
Phototoxicity/allergy	—	—	R	—	—	—
Bacterial gene mutation	R	R	R	R	R	R
Chromosomal damage	R	R	R	R	R	R
ADME—intended route	R	R	R	R	R	R
28-day toxicity (2 species) intended route	R	R	R	R	R	R
<i>Additional data: Short- or intermediate-term repeated use</i>						
90-day toxicity (most appropriate species)	R	R	R	R	R	R
Embryo-fetal toxicity	R	R	R	R	R	R
Additional assays ^a	C	C	C	C	C	C
Genotoxicity	R	R	R	R	R	R
Immunosuppression (3)	R	C	C	R	R	R
<i>Additional data: Intermittent long-term or chronic use</i>						
Chronic toxicity (rodent Nonrodent)	C	C	C	C	C	C
1-generation reproduction	R	R	R	R	R	R
Photocarcinogenicity	—	—	C	—	—	—
Carcinogenicity	C	C	C	C	C	C

Note: R, required; C, conditional

^aAdditional assays are dependent on the judgment of the data evaluator. They may include, but are not limited to screening for endocrine modulators or tests to determine if findings in animals are relevant to humans.

(From Ref. 12, p. 53.)

Laboratory Practices of the agency-(ies) to which the data will be submitted.

The base set of data is designed to provide fundamental information regarding acute toxicity by the oral route and/or intended dose route (Table 1). Skin and eye irritation testing should be conducted irrespective of the route of use of the candidate excipient. These data are intended to protect researchers during the research and production life of the material.

Absorption/distribution/metabolism/excretion/pharmacodynamics are considered fundamental data, as are mutagenicity tests (e.g., Ames test, *in vivo* chromosome aberration test, and mouse micronucleus test). Twenty-eight day repeated dosing studies in two species by appropriate route(s) also should be performed in a rodent and a nonrodent species, respectively.

One of the unique aspects of the IPEC approach is that not all tests are required. Some of the tests are conditional upon findings in other test procedures. Specific attention is paid to the route of exposure as well as to tests that might be required as potential exposure duration is increased. Emphasis is placed on the fact that the route of exposure for the test animals should be the same as the route of exposure anticipated in humans. Strict attention is paid to the type of exposure. For example, a protocol for study of a product intended for inhalation therapy that results in prolonged exposure of up to several hours per day will differ from that used to evaluate a material that would be used in a product resulting in exposure to several metered doses each day. Some tests may have to be conducted using a route of exposure different from the intended use route. This may be due to the nature of the test animal (e.g., reproductive tests in rabbits may require that the dosage route be other than inhalation, if inhalation is to be the route of use of the formulation containing the excipient).

The IPEC-Americas publication emphasizes that untrained people should not use its guidelines. In addition, the guidelines are not to be used as a checklist. They are to be used by professionals qualified to make the necessary judgments concerning what is referred to as “Conditional” tests. The conduct of these conditional tests is dependent on the results obtained from other required tests. It was considered that given the specificity of some of the cellular and subcellular techniques available and the variety of test animals being developed, that the traditional long-term imprecise test procedures may produce irrelevant information compared to that available from other test procedures. Also, some chemical families produce false positive or questionable results in certain species and the development of these types of data only serve to confound and require additional testing to clarify the questionable results.

It is conceivable that some excipients may not require the standard 2-year, two rodent species carcinogenicity studies. Such excipients include those that are not absorbed (or are rapidly metabolized and/or rapidly excreted), that do not exhibit toxicity in 90-day studies, and those that are negative for genotoxicity. This is the approach taken by the IPEC-Americas Safety Committee and one of the reasons that the 1996 peer-reviewed journal publication (12) indicates that the conduct of rodent carcinogenicity studies is conditional. The carcinogenicity studies that are conditional are the traditional 50 animals/sex/group rodent studies conducted for 18 or 24 months or variations thereof. The decision to make these tests conditional was also predicated on the fact that other models, that provided adequate information upon which to base a safety judgment regarding carcinogenic potential, were available.

GENETICALLY ENGINEERED ANIMAL MODELS

The use of genetically engineered animals has the potential to supplant some of the traditional long-term (2-year) rodent studies. Mouse models have been developed for use as mechanistic models in cancer research. Potential alternatives to the 2-year rodent oncogenicity bioassay include the p53 knockout mouse and the Tg.AC mouse (16).

The use of these mouse models is based on the observation that human neoplasms commonly demonstrate molecular alterations in tumor suppressor genes and/or oncogenes. In normal tissues, tumor suppressor genes (such as p53 and Rb) serve as negative regulators of cell proliferation. Inactivation or loss of tumor suppressor activity through gene mutation or deletion results in loss of this critical regulatory function and may lead to uncontrolled cell proliferation.

Loss of tumor suppressor gene is the most common genetic alteration found in human cancers. Deletion of one or both alleles of p53 (p53 knockout mice) increases the incidence of neoplasia and decreases latency of tumor development. When p53 knockout mice are exposed to genotoxic agents, they rapidly develop neoplasms in a range of tissues. Sensitive targets in p53 mice are often comparable to those in “normal” mice and hence, their utility as a model.

The Tg.AC mouse is used as a skin tumorigenesis model, and when exposed to phorbol ester tumor promoters and other nongenotoxic agents, is rapidly induced. When fully validated, a test battery, including the

Table 2 Summary of IPEC-Europe excipient testing guidelines

Tests	Routes of exposure for humans:					
	Oral	Mucosal	Transdermal	Dermal/topical	Parenteral	Inhalation/intranasal
<i>Step 0</i>						
ADME	R	R	R	R	R	R
<i>Step 1</i> (Basic set)						
Acute oral toxicity (intended route)	R	R	R	R	R	R
Eye irritation	—	R	R	R	R	R
Skin irritation	—	R	R	R	R	R
Skin sensitization	R	R	R	R	R	R
Acute parenteral toxicity	—	—	—	—	R	—
Application site evaluation	—	R	R	R	R	—
Pulmonary sensitization	—	—	—	—	—	—
Phototoxicity/photo- allergy	—	—	—	—	—	—
Ames test	R	R	R	R	R	R
Chromosome damage	R	R	R	R	R	R
Micronucleus	R	R	R	R	R	R
4 weeks toxicity	R	R	R	R	R	R
2(species)—intended route						
<i>Step 2</i>						
90-day toxicity (most appropriate species)	R	R	R	R	R	R
Teratology (rat and rabbit)	R	R	R	R	R	R
Genotoxicity assays	R	R	R	R	R	R
<i>Step 3</i>						
6—9 months chronic toxicity (Rodent, nonrodent)	C	C	C	C	C	C
Segment I	R	R	R	R	R	R
Segment III	C	C	C	C	C	C
Photocarcinogenicity	—	—	C	C	—	—
Carcinogenicity	C	C	C	C	C	C

Note: R, required C, conditional.
(From the IPEC Europe Safety Committee. The Proposed Guidelines for the Safety Evaluation of New Excipients. European Pharmaceutical Review, Nov 1997.)

heterozygous p53 knockout mouse and the Tg.AC mouse, may provide a model which will identify both genotoxic and nongenotoxic carcinogens and reduce the in-life time to conduct studies for carcinogenicity to as little as 6 months.

SUMMARY

The tests suggested by IPEC-Americas are summarized in Table 1 (12). The “R” represents required tests and the “C” represents tests that are conditional based on intended use and the results of previous tests. The tests suggested by IPEC-Europe (18) are found in Table 2 and differ slightly from Table 1. The decision whether or not to perform “C” labeled tests requires the judgment of a trained professional. Both IPEC test models are also predicated on obtaining chemical, pharmacological, and physical data from other investigators involved in the development of candidate excipients. Information developed by chemists, pharmacologists, and other disciplines is invaluable in estimating the hazards associated with a new compound.

Testing in humans, using the IPEC-Americas model, either as part of a clinical trial or as a stand-alone procedure, should be conducted as soon as warranted by the animal data. Critical evaluation of the base- set data may support the use of a candidate excipient intended for use once or twice in a lifetime. If one conducts the studies listed in Table 1, section 2, critical evaluation of the data may support the use of the new excipient in a variety of products intended for limited repeated intake, for example an antibiotic. If the Absorption, Distribution, Metabolism, Excretion/Pharmacokinetics (ADME/PK) studies show that the excipient is not absorbed, review of the other data may permit inclusion in a product intended to be used for 30–90 consecutive days. For longer-term usage, the tests listed in Table 1, section 3 must be considered. One-generation reproduction studies must be conducted to assess any excipient-induced effects/disturbances in mating behavior, development/maturation of gametes, fertility, and preimplantation/implantation loss of embryos. Should the data continue to support some concern for either reproductive or developmental toxicity, a segment III study might be appropriate (3, 19, 20).

Specific details regarding test methodology are not provided in the guideline. Test procedures generally recognized by experts and the regulatory agencies should be used. Each test should be designed to address a specific issue and the data should be evaluated accordingly. Care

should be taken when evaluating animal data to ensure that toxicological findings are not unique to the particular test species and therefore not relevant to the human experience.

Finally, it is important that a material being evaluated for safety is the same as that which will be used in pharmaceutical preparations. The manufacturer of the excipient must follow GMPs. A complete audit trail must be available from the time of manufacture until the product is made available to the consumer. IPEC-Americas has developed a third-party audit program that follows the guidelines enumerated above. The program is conducted by the International Pharmaceutical Excipients Audits, Inc. (IPEA) and is the only program of its type that focuses only on the quality of pharmaceutical excipients. The program is designed to prevent problems with excipients, such as the one that occurred in Haiti in 1996, where 80 children died because the glycerol in their cough medication was mostly glycol.

Ultimately, the safety of an excipient in a formulation requires the following:

1. That the excipient being used is the same excipient that was tested for safety.
2. The test procedures were adequate to evaluate safety and are acceptable to relevant regulatory authorities.
3. The excipient is as specified.
4. The formulated pharmaceutical is used as specified.
5. The concentration of the excipient in the formulation takes into account appropriate test data.

REFERENCES

1. De Jong, H.J. The Safety of Pharmaceutical Excipients. *Therapie* **1999**, 54 (11).
2. The United States *Pharmacopoeia* 24/*National Formulary* 19; 2000.
3. Principles for the Safety Assessment of Food Additives and Contaminants in Food. World Health Organization. *Environ. Health Criter.* **1987**, 70 .
4. *Good Manufacturing Practices Guide for Bulk Pharmaceutical Excipients*; The International Pharmaceutical Excipients Council: Arlington, VA, 2000.
5. *The International Pharmaceutical Excipients Council Guideline for Distribution of Bulk Pharmaceutical Excipients*; 2000.
6. Guideline for Residual Solvents, *International Conference on Harmonization*; 1999.
7. IPEC-Americas Significant Change Guide for Bulk Pharmaceutical Excipients, The International Pharmaceutical Excipients Council of the Americas, 2000.
8. Guidance on the Required Content and Format of Certificates of Analysis, The International Pharmaceutical Excipients Council of the Americas, 2000.

9. Blecher, L. Excipients: The Important Components. *Pharm. Process* **1993**.
10. Cooper, J. Inert Components of Pharmaceutical Preparations. *Drug Dev. Ind. Pharm.* **1979**, 5, 293.
11. Smith, J.M.; Dodd, T.R.P. Adverse Drug Reactions. *Acute Poisoning Review* **1982**, 1, 93.
12. Steinberg, M.; Borzelleca, J.F.; Enters, E.K.; Kinoshita, F.K.; Loper, A.; Mitchell, D.B.; Tamulinas, C.B.; Weiner, M.L. A New Approach to the Safety Assessment of Pharmaceutical Excipients. *Regul. Toxicol. Pharmacol.* **1996**, 24 (2), 149.
13. Steinberg, M. ICH Guidelines and Safety: An Update. *Pharm. Tech.* **1999**.
14. E.U. Directive 75/318 European Economic Community, 1995.
15. Guiding Principles in the Use of Animals in Toxicology Society of Toxicology, 1996.
16. *Life Sciences News Letter*; ITT Research Institute: Chicago IL, 2000.
17. Moreton, C.R. New Excipients: From Idea to Market. *Eur. Pharma. Rev.* **1997**, 2 (3).
18. The IPEC Europe Safety Committee. The Proposed Guidelines for the Safety of New Excipients. *Eur. Pharm. Rev.* **1997**, November.
19. Bass, R.; Ulbrich, B.; Hildebrandt, A.G.; Weissinger, J.; Doi, O.; Baeder, C.; Fumero, S.; Harada, Y.; Lehmann, H.; Manson, J.; Neubert, D.; Omori, Y.; Palmer, A.; Sullivan, F.; Takayama, S.; Tanimura, T. Draft Guideline on Detection of Toxicity of Reproduction for Medicinal Products. *Adv. Drug React. Toxicol. Rev.* **1991**, 9 (J), 127.
20. Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Foods, Redbook II: Draft, *U.S. Food and Drug Administration (USFDA), Bureau of Foods*; 1993.
21. Weiner, M.L.; Kotloskie, L.A. *Excipient Toxicity and Safety*; Marcel Dekker, Inc.: New York, 1999.

EXCIPIENTS—POWDERS AND SOLID DOSAGE FORMS

Hak-Kim Chan
Nora Y.K. Chew

University of Sydney, Sydney, New South Wales, Australia

INTRODUCTION

Excipients are the additives used to convert pharmacologically active compounds into pharmaceutical dosage forms suitable for administration to patients (1). Although excipients are the nonactive ingredients, they are essential in the successful production of acceptable solid dosage forms such as tablets and powders. For example, the lack of filling materials would make it exceedingly challenging, if not impossible, to produce a 1 mg dose tablet of a potent drug.

The following general criteria are essential for excipients (2): physiological inertness; physical and chemical stability; conformance to regulatory agency requirements; no interference with drug bioavailability; absence of pathogenic microbial organisms; and commercially available at low cost.

In reality, no single excipient would satisfy all the criteria; therefore, a compromise of the different requirements has to be made. For example, although widely used in pharmaceutical tablet and capsule formulations as a diluent, lactose may not be suitable for patients who lack the intestinal enzyme lactase to break down the sugar, thus leading to the gastrointestinal tract symptoms such as cramps and diarrhea. The role of excipients varies substantially depending on the individual dosage form.

EXCIPIENTS IN TABLETS AND CAPSULES

For tablets and capsules, excipients are needed both for the facilitation of the tableting and capsule-filling process (e.g., glidants) and for the formulation (e.g., disintegrants). Except for diluents, which may be present in large quantity, the level of excipient use is usually limited to only a few percent and some lubricants will be required at <1%. Details of the types, uses, and mechanisms of action of various excipients for tablet and capsule production have been discussed at length in other articles

in this encyclopedia.^a The types and functions of excipients for tablet production are summarized in Table 1. Although binders, lubricants, and antiadherents are specific for making tablets, other excipients in Table 1 are also used in capsule production for reasons similar to those for tablets.

It is worth noting that some of these tableting excipients may exert effects in opposition to each other. For example, binders and lubricants, because of their respective bonding and waterproofing properties, may hinder the disintegration action of the disintegrants. In addition, some of these tableting excipients may possess >1 function that may be similar (e.g., talc as lubricant and glidant) or opposite (e.g., starch as binder and disintegrant) to each other. Furthermore, the sequence of adding the excipients during tablet production depends on the function of the excipient. Whereas the diluents and the binders are to be mixed with the active ingredient early on for making granules, disintegrants may be added before granulation (i.e., inside the granules), and/or during the lubrication step (i.e., outside the granules) before tablet compression.

EXCIPIENTS IN FREEZE-DRIED (LYOPHILIZED) POWDERS

Freeze-dried (lyophilized) powders are obtained by the process of freeze-drying (lyophilization), which involves freezing of an aqueous-based drug solution in a glass vial followed by sublimation of the ice in a vacuum (3). Because the process is carried out at low temperatures, it is most suitable for heat-sensitive compounds. Antibiotics, such as cephalosporins, are among the preparations commonly prepared by freeze-drying (4). An interesting finding of excipients in freeze-drying is related to breakage of the glass vial (5). This was observed in excipients, such as mannitol, which would undergo mechanical expansion during warming after fast freezing.

Excipients are used in freeze-drying for various purposes. They act as bulking agents to give a pleasing

^aSee *Tablet Compression*, page 2669; *Tablet Formulation*, page 2701.

Table 1 Summary of types and functions of tableting excipients

Excipient	Functions	Examples
Diluent	To act as a bulking agent or filling material	Sugars, lactose, mannitol, sorbitol, sucrose Inorganic salts, primarily calcium salts Polysaccharides, primarily microcrystalline celluloses
Binders and adhesives	To hold powders together to form granules for tableting	Sugars, glucose, syrup Polymers, natural gums, starch, gelatin or synthetic celluloses, polyvinylpyrrol-pyrrolidone (PVP), poly-methacrylate (Eudragit TM)
Glidants	To improve the flow of granules from the hopper to the die cavity to ensure uniform fill for each tablet	Fine silica, magnesium stearate, purified talc
Disintegrants	To facilitate the breakup of a tablet in the gastrointestinal tract	Starch and derivatives (polyplasdone XL) Microcrystalline cellulose
Lubricants	To reduce the friction between the granules and the die wall during compression and ejection of the tableting process	Clays, algin, gums, surfactants Water-insoluble: metal stearates, stearic acid, talc Water-soluble: boric acid, sodium chloride, benzoate and acetate, sodium or magnesium lauryl sulfate
Antiadherents	To minimize the problem of picking, i.e., portion of the tablet face picked out and adhered to the punch face during tableting	Carbowax 4000 or 6000 Talc, cornstarch, metal stearates, sodium lauryl sulfate
Colorants	For identification purposes and visual marketing values	Natural pigments Synthetic dyes
Flavors and sweeteners	To improve the taste of chewable tablets	Natural, e.g., mannitol Artificial, e.g., aspartame

appearance to the freeze-dried products. Buffers are present to control the pH of the products that are stable only within a narrow pH range in solution, both during freezing and the subsequent reconstitution. However, it is important to realize that certain buffers, such as the phosphate buffer, in which $\text{Na}_2\text{HPO}_4 \cdot \text{H}_2\text{O}$ crystallizes during freezing, cause a pronounced drop in pH (6). This can lead to deleterious effects on the active ingredients, according to the pH dependence of the product stability. Other excipients that may be present in freeze-dried powders include: solubility enhancers (e.g., surfactants or cosolvents), osmotic agents (e.g., saline and sugars), antioxidants (e.g., ascorbic acid), and preservatives for multiple-injection containers (e.g., benzyl alcohol and chlorobutanol). In addition, freeze-dried biological powders may also contain excipients that function to reduce protein adsorption onto the container surface (e.g., surfactants and albumins) (7). A particularly important

use of excipients for therapeutic protein formulations is the stabilization of the protein molecules in the dry state, as discussed later.

Therapeutic Protein Formulations

Therapeutic proteins are usually prepared in liquid formulations or as freeze-dried powders that are to be reconstituted immediately before use. A number of the proteins have been found to be unstable when dried alone, with aggregation being a major problem. It has been found that stability can be greatly improved if the proteins are dried in the presence of certain excipients (8, 9). However, not all excipients that can stabilize protein against aggregation are suitable. Other considerations required of the excipients for use in protein pharmaceuticals include:

Redox reaction potential: reducing sugars such as lactose and sucrose may not be suitable if they react with the protein (e.g., via the lysine residue) resulting in protein glycosylation [e.g., lactosylation of recombinant deoxyribonuclease I by lactose (10)] and other reaction products. However, glycosylation alone may not necessarily be a problem if the glycosylated proteins do not cause toxicity and immunogenicity while maintaining the therapeutic efficacy.

Parenteral use suitability: excipients such as trehalose, which has not been used in any products acceptable by regulatory authorities, may create concerns over toxicity.

Nonparenteral use feasibility: for example, inhalation drug delivery; lactose has been used for marketed aerosol products and hence may be more suitable for inhalation protein formulations.

Table 2 gives some examples of excipients used as stabilizers for proteins in freeze-dried formulations. Among others, saccharides are the most widely used excipients for stabilizing freeze-dried therapeutic proteins. There are exceptions to the need for stabilizing excipients, e.g., recombinant (α -Antitrypsin was stable when freeze-dried alone or with lactose, sucrose, and polyvinylpyrrolidone (11).

The mechanism of the protective effects imparted by the excipients has not been fully elucidated. Empirical observations have pointed to the following contributing factors: formation of a glassy state of the protein–excipient system; crystallinity of the excipients; hydrogen bonding between the excipient and protein molecules; and residual water content.

Glass is an amorphous or noncrystalline solid. It is characterized by the glass transition temperature above which the glassy state softens to the rubbery state. Protein stabilization imparted by excipients can be achieved when the freeze-dried powders are held below the glass transition temperature (T_g) of the protein–excipient systems (i.e., in the glassy state). Of particular relevance to the protein stability is that in the glassy state, the diffusion rate and mobility of the molecules are much less than those in the rubbery state. Thus, any physicochemical reactions leading to protein degradation will be diminished as the protein molecules are “frozen” in the glass formed by the excipients (12).

In contrast to the amorphous excipients, crystalline excipients, such as mannitol, were reported to reduce the stability of proteins (13). Mannitol can be used if the powder is rendered amorphous by the presence of other excipients such as glycine (14). Evidence for protein stabilization by hydrogen bonding has mainly come from

the Fourier transformed infrared (FTIR) spectroscopy (15), which provides information on the protein secondary structures. The amide I absorption band (approximately $1600\text{--}1700\text{ cm}^{-1}$) of freeze-dried proteins with excipients was found to bear more similarities than the freeze-dried proteins alone to the native proteins in the aqueous environment. This has been explained by sustenance of the native protein structures by protein–excipient hydrogen bonding in the dry powders. However, FTIR measurements were mostly carried out in compressed potassium bromide disks containing the protein. The integrity of the compressed proteins has been largely overlooked (16).

Water affects the stability of proteins by enhancing the mobility of the protein molecules (17). It has been established that an optimal level of water is required to maintain stability of proteins during storage (18). Moisture was known to increase the mobility of the surface groups of protein as measured by solid-state nuclear magnetic resonance spectroscopy (19, 20). The distribution of water between the protein and the excipients in a freeze-dried powder depends on the crystalline or amorphous nature of the excipients (21). For example, if a protein is formulated with an amorphous excipient and stored in a sealed container, water would distribute according to the water affinity of the protein and excipients (21). When the amorphous excipient crystallizes (e.g., because of elevated temperatures), it will expel its sorbed water, which may cause stability problems in the protein (8).

EXCIPIENTS IN POWDER AEROSOL FORMULATIONS

Pharmaceutical inhalation aerosols are widely used for treatment of diseases such as asthma and chronic bronchitis. There are three basic types of aerosol products: the propellant-driven metered-dose inhalers, the dry powder inhalers, and the nebulizers (33). Because of the ozone-depleting and greenhouse effects of the chlorofluorocarbon (CFC) propellants, interest in the dry powder aerosols has risen in recent years.

The main use of excipients in the dry powder inhaler formulations has been to act as carriers for the active ingredients (Table 3). The performance of a dry powder system depends on both the aerosol device and the powder formulation. To generate respirable aerosols, powder formulations must meet two opposing criteria: the particles have to be sufficiently fine (e.g., $<7\text{ }\mu\text{m}$) for lung deposition, and yet coarse enough for optimal flow in device (and capsule)-filling and emptying. To achieve this, the drug is blended with coarse inert excipient carriers (34).

Table 2 Some examples of excipients used as protectants for freeze-dried protein and peptide formulations

Protein	Excipients and uses	Reference
Recombinant human growth hormone (rhGH)	Mannitol and glycine as amorphous excipients to prevent human growth hormone (hGH) aggregation.	14
Bovine and human insulins	Trehalose as a lyoprotectant, reserves the secondary structure of rhGH.	22
	Dextrin, Emdex™ (spray-dried dextrose) and hydroxypropyl β-cyclodextrin minimized insulin aggregation	23
Recombinant factor IX	Polysorbate 80 as protectant for freezing; sucrose as protectant for drying; histidine as pH buffer; glycine for cake appearance	24
Recombinant human interleukin-6	Aggregation prevented by amorphous trehalose, sucrose or a combination of sucrose, and glycine or mannitol	25
Recombinant human interleukin-1 receptor antagonist	Sucrose, sorbitol, trehalose and alanine as protectants against aggregation and deamidation; mannitol and glycine as bulkingagent; sodium citrate as buffer	26
FK906 tripeptide	Sugars (sucrose, lactose, trehalose, maltose), polymer (dextran) and salts (NaCl, KCl) to modify the glass transition temperatures of the freeze-dried powders	27
Recombinant human albumin	Organic acid excipient molecules with either a carboxyl group or an amino group present at C-1 position completely stabilized rHA against aggregation	28
Lactate dehydrogenase	Polyethylene glycol as protectant for freezing; sugars (mannitol, lactose, trehalose) as lyoprotectants against loss of bioactivity	29
Alkaline phosphatase	Lactose and trehalose maintain activity longer at elevated temperatures than mannitol	30
Recombinant bovine somatotropin, lysozyme	Both the excipient type (sucrose, sorbitol, glycerol) and moisture content affected protein degradation	31
Hemoglobin	Mannitol protected protein from phase separation induced damage during freeze drying	32
Recombinant human factor XIII	Trehalose and sucrose preserved the native dimeric structure of the protein and prevented aggregates formation	

Table 3 Some examples of excipients used for dry powder aerosols

Active ingredient	Excipient carrier	Reference
Salbutamol sulfate	Lactose (63–90 μm): regular, spray-dried, and recrystallized	34
Budesonide	Lactose (α-monohydrate (<32 μm, 63–90 μm, 125–180 μm)	39
rhDNase	Lactose (50 wt% < 42 and 115 μm)	38
	Mannitol (50 wt% < 43 μm)	
	Sodium chloride (50 wt% < 87 μm)	
Bovine serum albumin–maltodextrin (50–50)	Lactose (α-monohydrate (63–90 μm)	40
	Fine particle lactose (76 wt% < 10 μm)	
	Micronized polyethylene glycol 6000 (97.5 wt% < 10 μm)	
Recombinant human granulocyte-colony stimulating factor-mannitol	Polyethylene glycol 8000 (38–75 μm, 90–125 μm)	41

Thus, the primary reason of using excipient carrier is to enhance flowability of the drug powder. The excipient carriers are large particles which, because of their sizes ($>50\text{ }\mu\text{m}$), would not be inhaled into the lung. They provide surfaces for the fine drug particles to adhere (Fig. 1), forming an interactive powder mix that would have an improved flowability than the drug alone for handling. On dispersion of the powder by air flow, the fine drug particles are detached from the carriers for inhalation. In an ideal drug-carrier system, the adhesion of the drug to the carriers is strong enough to prevent demixing during filling, handling, and storage, but not so strong as to prevent the generation of fine drug particles by detachment from the carrier during inhalation.

Another reason for using excipient carrier is to improve the availability of fine drug particles in the aerosol cloud. Surface texture of excipients appears to play a prominent role. The fine particle fraction of the antiasthmatic drug salbutamol sulfate was significantly higher with the recrystalline lactose as carrier than with regular or spray-dried lactose. The difference was attributed to the lower surface rugosity (roughness) of the recrystalline lactose (35). With another antiasthmatic compound, salmeterol xinofoate, a formulation using lactose carrier produced a higher fine particle fraction than formulations containing sucrose or spray-dried sorbitol (36). The implication is that using a suitable excipient as carrier it is possible to generate the desirable amount of fine drug particles in an aerosol with a minimal inspiratory effort. Recombinant human deoxyribonuclease I (rhDNase), the first therapeutic protein approved by the Food and Drug Administration (FDA) in the United States for inhalation use in the treatment of cystic fibrosis (37), generated a twofold increase in the fine particle fraction in the aerosol when blended with excipients lactose, mannitol, or sodium chloride (38). In this case, the increase was independent of both the type and relative amount of the excipient used.

Besides surface texture, excipient particle size also plays an important role in the fine particle generation as shown by budesonide, where the highest fine particle fraction was obtained with small-sized ($<32\text{ }\mu\text{m}$) lactose as the carrier (39). Additionally, fine particle excipients such as fine lactose or polyethylene glycol were reported to improve the performance of carrier-based protein dry powder aerosols (40). However, there are some cases where carriers improved total powder emission but reduced the percent of active powders in the aerosol (41). To be useful carriers, the excipients must be physically stable. The important physicochemical characteristics for drug carrier selection are discussed in Ref. 42.

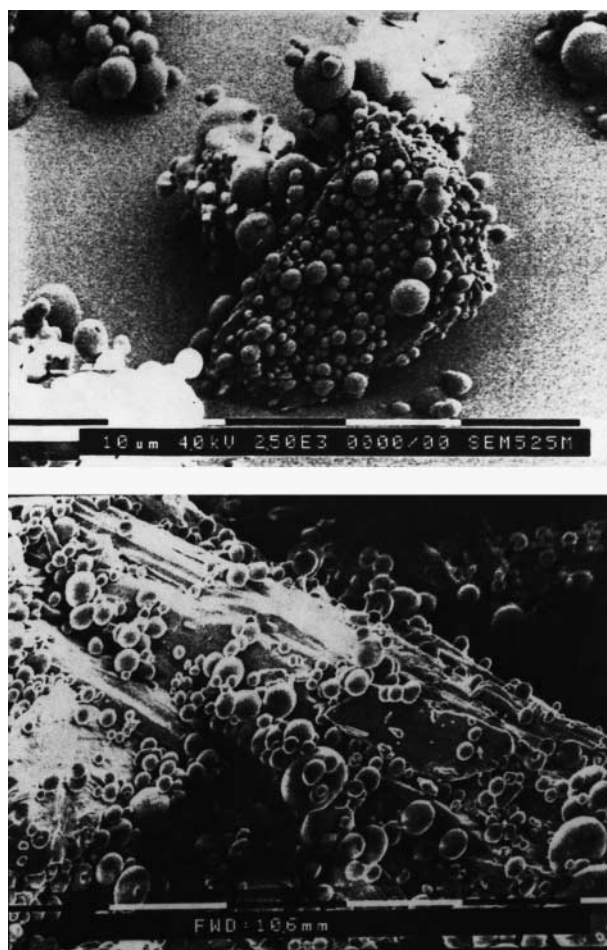


Fig. 1 An example of excipient as carrier for drug particles. Scanning electron micrographs showing adhesion of recombinant human deoxyribonuclease I (rhDNase) particles to lactose (*top*) and mannitol (*bottom*).

In addition to being used as carrier, excipients can enhance the aerosol performance by cospray-drying with the active ingredient. In this case, instead of being external to the drug particles, the excipient exists with the active ingredient in the same particle. For example, using sodium chloride as a crystalline excipient, the fine particle fraction of rhDNase in the aerosol was increased linearly with the amount of excipient present (38). The enhancement was correlated with the degree of crystallinity of the powder in Fig. 2.

As pointed out at the beginning, excipients are not the active ingredients and should be physiologically inert. However, a special use of excipients in dry powder aerosols has been for bronchial provocation testing in asthmatics (43, 44) and for the enhancement of mucociliary clearance in both normal and asthmatic subjects (45, 46). In both cases they acted as the active

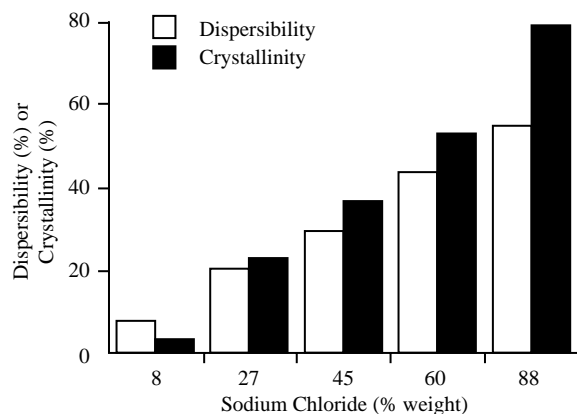


Fig. 2 Relationship between dispersibility (expressed as percent weight of particles less than 7 μm in the aerosol) and crystallinity (by X-ray powder diffraction) of rhDNase powders with different sodium chloride contents. (Adapted from Ref. 38.)

ingredients. The excipients are osmotic agents such as sodium chloride and mannitol. They change the osmolarity of the airway fluid, leading to the physiologic effects of enhanced clearance in the lungs or bronchoconstriction in hyperresponsive subjects.

EXCIPIENTS IN SPRAY-DRIED POWDERS

Spray-drying is a process where a drug solution is atomized to fine droplets followed by evaporation in a stream of warm air to form dry particles (47). The properties of the spray-dried products are controlled by both the process and formulation parameters (48). During the process, the active ingredients are subjected to mechanical shears from atomization and heat stress from the drying air at elevated temperature. Because of the tremendous surface area exposure of the atomized droplets, the drug will also be subjected to degradations such as oxidation and surface denaturation. Excipients can be used as stabilizers or protectants against degradation of the active ingredients. Autoxidation of the analgesic and anti-inflammatory agent aminopyrine was eliminated by excipients such as antioxidants, chelating agents, and clay (49, 50). Denaturation of the model protein (β -Galactosidase) was prevented in the presence of trehalose as an excipient (51). Sucrose was found to minimize the degradation product methemoglobin when oxyhemoglobin was spray-dried (52). Recombinant human growth hormone (rhGH), degraded by aggregation resulting from surface denaturation during spray-drying, was successfully stabilized by the

surfactant polysorbate 20 (53). Lactose has been found to protect spray-dried rhDNase against aggregation during storage (54, 55). As a cospray-dried excipient, sodium chloride was reported to increase the dispersibility of the spray-dried rhDNase powder to form aerosols for inhalation (38).

Excipients such as colloidal silica have been reported to increase the flow of spray-dried aminopyrine-barbital powders (49). In contrast, formulations of spray-dried salicylic-acid-containing gelatin and polyvinyl alcohol as excipients were less free-flowing (56). Gum arabic and polyvinylpyrrolidone prevented the sublimation of salicylic acid during spray-drying. For vitamin E acetate, the cospray-dried excipients affected both the powder flowability and drug release properties. Hydroxypropyl cellulose improved the drug release properties; AerosilTM (colloidal silica) enhanced the powder flow. A balance between these two physical parameters was achieved with approximately 6:1 weight ratio of cellulose to Aerosil (57).

Although it does not effectively protect ascorbic acid against oxidative degradation, colloidal silicon dioxide was found to increase the yield of spray-dried powder (58).

For polymorphic compounds, such as sulfa drugs, talc excipients induced polymorphic transformation of sulfamethoxazole during the process of microencapsulation by spray-drying (59).

Particle size and true density of spray-dried sodium salicylate were affected by binder excipients (56). Drug distribution in spray-dried tolbutamide particles was dependent on the disintegrant excipients used. The drug distributed throughout the particles with low-substituted hydroxypropyl cellulose as excipient but only deposited on the surface with pregelatinized corn starch (60).

Excipients like dibutyl phthalate were used as plasticizers for controlled-release microspheres of theophylline and sulfamethazine prepared by spray-drying (61). Likewise, citric acid was used as plasticizer for spray-dried sodium carboxymethyl cellulose and hydroxypropylmethyl cellulose microspheres containing theophylline (62). Excipients were found to affect the release rate of theophylline with citric acid and triethylene citrate giving the slowest and fastest rate, respectively, as compared with polyethylene glycol and glycerin excipients.

EXCIPIENTS IN CONTROLLED RELEASE SOLID DOSAGE FORMS

Polymeric excipients are commonly used for controlled-release formulations either as a coating around a drug core by microencapsulation or as a matrix in which the drug is

embedded. Depending on the release profile requirement, polymeric excipients are traditionally classified as hydrophilic or hydrophobic. Some representative coating materials include water-soluble resins (e.g., gelatin, starch, polyvinylpyrrolidone, water-soluble celluloses), water-insoluble resins (e.g., polymethacrylate, silicones, water-insoluble celluloses), waxes and lipids (e.g., paraffin, beeswax, stearic acid), enteric resins (e.g., shellac, cellulose acetate phthalate) (63). (Further details on polymers for controlled release systems can be found under “Biopolymers for Controlled Drug Delivery” in the first edition of this encyclopedia series.) Here the focus is on some recent applications of excipients in biologicals.

Live rotavirus vaccine was developed for oral delivery to prevent infections by the virus in young children (64). However, incorporation of live rotavirus into poly (DL-lactide-co-glycolide) microspheres or alginate microcapsules was reported to result in a significant loss of rotavirus infectivity. The loss was reduced by stabilization of the rotavirus vaccine with an excipient blend of cellulose, starch, sucrose, and gelatin at a mass ratio of 30:30:30:10 in granules or tablets (64).

Transforming growth factor (TGF)- β 1, a cytoprotectant against the toxicity caused by cell cycle-specific drugs, was encapsulated in alginate beads as a potential oral delivery system to release TGF- β 1 in the gastrointestinal tract. However, the TGF- β 1 was interacting with alginate, which prevented the release of the protein. Polyacrylic acid, as a polyanion excipient, was used to shield the TGF- β 1 from interacting with the alginate (65).

Glucose at concentrations >10% was used to achieve adequate reconstitution of freeze-dried biodegradable poly-DL-lactide nanoparticles with conservation of the encapsulated cyclosporin A (66). Glucose and trehalose were also found to be the most efficient cryoprotectors for the lyophilization process, whereas trehalose was used for spray-drying, in the production of solid lipid nanoparticles (67).

Tetanus toxoid (the vaccine for tetanus) encapsulated in polyester microspheres was produced for single-injection immunization (68, 69). The entrapment efficiency of the protein vaccine was significantly improved by coencapsulation with excipients such as trehalose and (γ -Hydroxypropyl) cyclodextrin. However, these excipients did not impart stabilizing effect on tetanus toxoid. In contrast, bovine serum albumin was found to be the most prominent stabilizer for protein in the body after administration by injection.

It is important to point out that the stabilizing effects of excipients were sometimes reported for the formulations in vitro rather than in the in vivo conditions. However, the

degree of retention of the native protein structure in the dry state may not be a general indication of stability for the ‘wetted’ solid within polymer controlled-release devices in the body. In the case of tetanus toxoid, it was shown that the extent of structural alternations in the presence of 1:5 (gram excipient:gram protein) sodium chloride, sorbitol, or polyethylene glycol did not correlate with stability conferred toward moisture-induced aggregation (70).

Surfactant and polyethylene glycols (PEG) excipients have been used in microencapsulation of macromolecules for various effects. For example, Tween 20, at the critical micelle concentration and at a molar concentration of protein:surfactant of 1:0.018 or larger, was found to increase the encapsulation efficiency of β -Lactoglobulin in poly (DL-Lactide-co-glycolide) microspheres (71). The initial burst release was reduced with increasing Tween 20 concentration, and the effect was attributed to reduction of the number of pores and channels inside the microspheres. For gene therapy, the release of biologicals encapsulated in microspheres can be significantly improved by adding surfactant during microencapsulation, as recently exemplified by the enhancing effect of polyvinyl alcohol on the release of adenovirus from PLGA microspheres (72). PEG 400 has been used to improve the stability of the protein, nerve growth factor (NGF) during the microencapsulation by a double emulsion method. It stabilized the protein by reducing the contact with the organic solvent in the process. Furthermore, the presence of NaCl in the microencapsulation process has been shown to modify the microsphere structures, leading to a reduction of the initial release rate of NGF (73).

EXCIPIENTS AND FORMULATION INCOMPATIBILITY

During formulation design some excipients may be incompatible with the active ingredient or with other excipients. Excipient incompatibility problems are, in fact, widely published and date back to the mid-1950s. For example, as a tableting excipient, lactose could react via its aldehyde group with both primary (1) and secondary (74) amines by the Maillard-type condensation reaction. Sorbitol, another excipient sugar, is hygroscopic at relative humidity >65%, which should thus be avoided during manufacturing. Calcium salts are other widely used tableting excipients. However, calcium carbonate is incompatible with acids or acidic drugs because of the acid–base chemical reaction. Calcium salts are also incompatible with tetracyclines because of the formation

of calcium–tetracycline complexes. Details of reactivities and incompatibilities of individual excipients are given in Ref. 1. Incompatibility attributable to excipients is commonly studied under accelerated testing conditions or using thermal analyses such as differential scanning calorimetry. However, the results of this rapid testing could be misleading and thus of very limited value (75).

Besides direct excipient–drug interactions, excipients can lead to instability of the active ingredient by an indirect role through moisture distribution. Residual water content is known to affect the stability of solid dosage forms and powders (76). Decomposition of cephalothin sodium and benzylpenicillin potassium decomposition in freeze-dried preparation was believed to be partly attributed to the effect of water binding to excipients (4). The degradation rate of cephalothin sodium increased with the water content of excipients corn starch and celluloses (77). The results were correlated with the water mobility in the presence of the excipients (4, 77). A study of the effect of various excipients on the solid-state crystal transformation of the antimalarial compound mefloquine hydrochloride revealed that microcrystalline cellulose promoted the transformation from form E into form D (78). However, methylcellulose, hydroxyethylcellulose, β -Cyclodextrin, cospovidone, and hydrous lactose had no effect. The effect was again explained by the difference in the water uptake behavior by the excipients. Aspirin was formulated with a sugar diluent containing approximately 8% moisture, which did not cause instability problems (79). This was ascribed to the moisture present in the formulation being unavailable to react with the aspirin. The availability of moisture associated with excipients in a formulation can thus be manipulated to control the hydration rate of the active ingredient as in the case of nitrofurantoin, with crystalline lactose giving the fastest and microcrystalline cellulose giving the slowest rate (80). The rate of hydrolysis of methylprednisolone sodium succinate was higher when cofreeze-dried with mannitol than with lactose (81). This correlated with the rate of crystallization of mannitol in the formulation and its subsequent effect on the water distribution in the solid. The stabilizing potency of excipients on recombinant human albumin against aggregation also correlated with the water-sorbing capacity of the excipients (27).

Instability attributable to excipient-mediated water distribution in solids and powders has been explained by excipient physical properties (21, 82–84). Crystalline materials will not uptake moisture until the deliquescent point is reached. In contrast, amorphous excipients will absorb water until their glass transition temperatures fall below the ambient temperature when the mobility of the molecules has increased so much that excipient

crystallization will occur to expel the absorbed water from the crystal lattice. Before crystallization, these excipient materials will act as buffers or sorbents to hold the excess moisture which, depending on the water activity, may not be accessible to the active ingredient that is thus be protected from moisture-mediated decomposition. However, when excipient crystallization occurs, the expelled water will become available to react, leading to instability of the drug.

CONCLUSION

Although excipients are the nonactive ingredients, they are indispensable for the successful production of acceptable solid dosage forms. The important roles played by excipients in tablets and capsules, freeze-dried, and spray-dried powders, as well as powder aerosol formulations, were discussed. Some recent applications of excipients in controlled, release formulations for biologicals were also highlighted. Finally, incompatibility problems attributable to excipients were considered with an emphasis on the indirect role of excipients through moisture distribution.

REFERENCES

1. Wade, A., Weller, P.J., Eds.; *Handbook of Pharmaceutical Excipients*, 2nd Ed.; American Pharmaceutical Association: Washington DC, 1994.
2. Bandelin, F.J. Compressed Tablets by Wet Granulation. *Pharmaceutical Dosage Forms: Tablets*, 2nd Ed.; Lieberman, H.A., Lachman, L., Schwartz, J.B., Eds.; Marcel Dekker, Inc.: New York, 1989; 1.
3. Pikal, M.J. Freeze-Drying of Proteins. Part I: Process Design. *BioPharmacology* **1990**, 18–27.
4. Oguchi, T.; Yamashita, J.; Yonemochi, E.; Yamamoto, K.; Nakai, Y. Effects of Saccharides on the Decomposition of Cephalothin Sodium and Benzylpenicillin Potassium in Freeze-Dried Preparations. *Chem. Pharm. Bull.* **1992**, 40, 1061–1063.
5. Williams, N.A.; Guglielmo, J. Thermal Mechanical Analysis of Frozen Solutions of Mannitol and Some Related Stereoisomers: Evidence of Expansion During Warming and Correlation With Vial Breakage During Lyophilization. *J. Parenteral Sci. Technol.* **1993**, 47, 119–123.
6. Pikal, M.J. Freeze –Drying of Proteins. Part II: Formulation Selection. *BioPharmacology* **1990** October, 26–30.
7. Crommelin, D.J.A., Sindelar, R.D., Eds.; *Pharmaceutical Biotechnology* Harwood Academic Publishers: The Netherlands, 1997; 72–74.

8. Carpenter, J.F.; Pikal, M.J.; Chang, B.S.; Randolph, T.W. Rational Design of Stable Lyophilized Protein Formulations: Some Practical Advice. *Pharm. Res.* **1997**, *14*, 969–975.
9. Arakawa, T.; Prestrelski, S.J.; Kenney, W.C.; Carpenter, J.F. Factors Affecting Short-Term and Long-Term Stabilities of Proteins. *Adv. Drug Delivery Rev.* **1993**, *10*, 1–28.
10. Quan, C.; Wu, S.; Hsu, C.; Canova-Davis, E. *Protein Sci.* **1995**, *4* (suppl), 490T.
11. Vemuri, S.; Yu, C.-D.; Roosdorp, N. Effect of Cryoprotectants on Freezing, Lyophilization and Storage of Lyophilized Recombinant Alpha-Antitrypsin Formulations. *PDA J. Pharm. Sci. Technol.* **1994**, *48*, 241–246.
12. Frank, F. Long-Term Stabilization of Biologicals. *Biotechnology* **1994**, *12*, 253–256.
13. Izutsu, K.; Yoshioka, S.; Terao, T. Decreased Protein-Stabilizing Effects of Cryoprotectants Due to Crystallization. *Pharm. Res.* **1993**, *10*, 1232–1237.
14. Pikal, M.J.; Dellerman, K.M.; Roy, M.L.; Riggins, R.M. The Effects of Formulation Variables on the Stability of Freeze-Dried Human Growth Hormone. *Pharm. Res.* **1993**, *8*, 427–436.
15. Carpenter, J.F.; Prestrelski, S.J.; Dong, A. Application of Infrared Spectroscopy to Development of Stable Lyophilized Protein Formulations. *Eur. J. Pharmacol. Biopharmacol.* **1998**, *45*, 231–238.
16. Chan, H.-K.; Ongpipattanakul, B.; Au-Yeung, J. Aggregation of Rh DNase Occurred During the Compression of KBr Pellets Used for FTIR Spectroscopy. *Pharm. Res.* **1996**, *13*, 238–241.
17. Hageman, M.J. Water Sorption and Solid-State Stability of Proteins. *Stability of Protein Pharmaceuticals, Part A*; Ahern, T.J., Manning, M.C.; Eds.; Plenum Press: New York, 1992; 273–309.
18. Hsu, C.C.; Ward, C.A.; Pearlman, R.; Nguyen, H.M.; Yeung, D.A.; Curley, J.G. Determining the Optimum Residual Moisture in Lyophilized Protein Pharmaceuticals. *Dev. Biol. Standard* **1992**, *74*, 255–271.
19. Yoshioka, S.; Aso, Y.; Kojima, S. Determination of Molecular Mobility of Lyophilized Bovine Serum Albumin and (γ -Globulin by Solid-State ^1H NMR And Relation to Aggregation-Susceptibility). *Pharm. Res.* **1996**, *13*, 926–930.
20. Separovic, F.; Lam, Y.H.; Ke, X.; Chan, H.-K. A Solid-State NMR Study of Protein Hydration and Stability. *Pharm. Res.* **1998**, *15*, 1816–1821.
21. Chan, H.-K.; Au-Yeung, J.K.-L.; Gonda, I. Water Distribution in Freeze-Dried Solids Containing Multiple Components. *Pharm. Res.* **1996**, *13* (suppl), S-216.
22. Katakam, M.; Banga, A.K. Aggregation of Insulin and Its Prevention by Carbohydrate Excipients. *PDA J. Pharm. Sci. Technol.* **1995**, *49*, 160–165.
23. Bush, L.; Webbs, C.; Bartlett, L.; Burnette, B. The Formulation of Recombinant Factor IX: Stability, Robustness, and Convenience. *Sem. in Hematol.* **1998**, *35* (suppl 2), 18–21.
24. Luckel, B.; Bodmer, D.; Helk, B. A Strategy for Optimizing the Lyophilization of Biotechnology Products. *Pharm. Sci.* **1997**, *3*, 3–8.
25. Chang, S.B.; Beauvais, R.M.; Dong, A.; Carpenter, J.F. Physical Factors Affecting the Storage Stability of Freeze-dried Interleukin-1 Receptor Antagonist: Glass Transition and Protein Conformation. *Arch. Biochem. Biophys.* **1996**, *331*, 249–258.
26. Jang, J.W.; Kitamura, S.; Guillory, J.K. The Effect of Excipients on Glass Transition Temperatures for FK 906 in the Frozen and Lyophilized States. *PDA J. Pharm. Sci. Tech.* **1995**, *49*, 166–174.
27. Costantino, H.R.; Langer, R.; Klivanov, A.M. Aggregation of a Lyophilized Pharmaceutical Protein, Recombinant Human Albumin: Effect of Moisture and Stabilization by Excipients. *Biotechnology* **1995**, *13*, 493–496.
28. Prestrelski, S.J.; Arakawa, T.; Carpenter, J.F. Separation of Freezing-and Drying-Induced Denaturation of Lyophilized Proteins Using Stress-Specific Stabilization. II. Structural Studies Using Infrared Spectroscopy. *Arch. Biochem. Biophys.* **1993**, *303*, 465–473.
29. Ford, A.W.; Dawson, P.J. The Effect of Carbohydrate Additives in the Freeze-Drying of Alkaline Phosphatase. *J. Pharm. Pharmacol.* **1993**, *45*, 86–93.
30. Bell, L.N.; Hageman, M.J.; Muraoka, L.M. Thermally Induced Denaturation of Lyophilized Bovine Somatotropin and Lysozyme As Impacted Moisture and Excipients. *J. Pharm. Sci.* **1995**, *84*, 707–712.
31. Heller, M.C.; Carpenter, J.F.; Randolph, T.W. Protein Formulation and Lyophilization Cycle Design: Prevention of Damage Due to Freeze-Concentration Induced Phase Separation. *Biotechnol. Bioeng.* **1999**, *63*, 166–174.
32. Kreilgaard, L.; Frokjaer, S.; Flink, J.M.; Randolph, T.W.; Carpenter, J.F. Effects of Additives on the Stability of Recombinant Human Factor XIII During Freeze-Drying and Storage in the Dried Solid. *Arch. Biochem. Biophys.* **1998**, *360*, 121–134.
33. Clark, A.R. Medical Aerosol Inhalers: Past, Present, and Future. *Aerosol Sci. Technol.* **1995**, *22*, 374–391.
34. Ganderton, D. The Generation of Respirable Clouds from Coarse Powder Aggregates. *J. Biopharm. Sci.* **1992**, *3*, 101–105.
35. Kassem, N.M.; Ganderton, D. The Influence of Carrier Surface on the Characteristics of Inspirable Powder Aerosols. *J. Pharm. Pharmacol.* **1990**, *42*, 11.
36. Mackin, L.A.; Rowley, G.; Fletcher, E.J. An Investigation of Carrier Particle Type, Electrostatic Charge and Relative Humidity on In-Vitro Drug Deposition from Dry Powder Inhaler Formulations. *Pharm. Sci.* **1997**, *3*, 583–586.
37. Cipolla, D.C.; Clark, A.R.; Chan, H.-K.; Gonda, I.; Shire, S.J. Assessment of Aerosol Delivery Systems for Recombinant Human Deoxyribonuclease. *STP Pharma. Sci.* **1994**, *4*, 50–62.
38. Chan, H.-K.; Clark, A.R.; Gonda, I.; Mumenthaler, M.; Hsu, C. Spray Dried Powders and Powder Blends of Recombinant Human Deoxyribonuclease (rhDNase) for Aerosol Delivery. *Pharm. Res.* **1997**, *14*, 431–437.
39. Steckel, H.; Muller, B.W. In Vitro Evaluation of Dry Powder Inhalers II: Influence of Carrier Particle Size and Concentration on in Vitro Deposition. *Int. J. Pharm.* **1997**, *154*, 31–37.
40. Lucas, P.; Anderson, K.; Staniforth, J.N. Protein Deposition from Dry Powder Inhalers: Fine Particle Multiplets As Performance Modifiers. *Pharm. Res.* **1998**, *15*, 562–569.
41. French, D.L.; Edwards, D.A.; Niven, R.W. The Influence of Formulation on Emission, Deaggregation and Deposition of

- Dry Powders for Inhalation. *J. Aerosol Sci.* **1996**, *27*, 769–783.
42. Byron, P.R.; Naini, V.; Phillips, E.M. Drug Carrier Selection Important Physicochemical Characteristics, *Respiratory Drug Delivery V*; Dalby, R.N., Byron, P.R., Farr, S.J., Eds.; Interpharm Press: Illinois, 1996; 103–113.
43. Anderson, S.D.; Brannan, J.; Spring, J.; Spalding, N.; Rodwell, L.; Chan, H.-K.; Gonda, I.; Walsh, A.; Clark, A.R. A New Method for Bronchial Provocation Testing in Asthmatic Subjects Using a Dry Powder of Mannitol. *Am. J. Crit. Care Med.* **1997**, *156*, 758–765.
44. Chew, N.Y.K.; Chan, H.-K. Dispersion of Mannitol Powders As Aerosols: Influence of Particle Size, Air Flow and Inhaler Device. *Pharm. Res.* **1999**, *16*, 1098–1103.
45. Daviskas, E.; Anderson, S.D.; Brannan, J.D.; Chan, H.-K.; Eberl, S.; Bautovich, G. Inhalation of Dry Powder Mannitol Increases Mucociliary Clearance. *Eur. Respir. J.* **1997**, *10*, 2449–2454.
46. Brannan, J.D.; Anderson, S.D.; Koskela, H.; Chew, N. Responsiveness to Mannitol in Asthmatic Subjects With Exercise—And Hyperventilation-Induced Asthma. *Am. J. Respir. Crit. Care Med.* **1998**, *in press*, 158.
47. Masters, K. *Spray Drying Handbook*, 4th Ed.; Wiley & Sons: New York, 1985.
48. Wendel, S.; Celik, M. An Overview of Spray-Drying Applications. *Pharm. Technol.* **1997**, (Oct), 124–156.
49. Kawashima, Y.; Lin, S.Y.; Ueda, M.; Takenaka, H.; Ando, Y. Direct Preparation of Solid Particulates of A minopyrin–Barbital Complex (Pyribital) from Droplets by a Spray-Drying Technique. *J. Pharm. Sci.* **1983**, *72*, 514–519.
50. Kawashima, Y.; Lin, S.Y.; Ueda, M.; Takenaka, H. Preparation of Solid Particulates of Amino–Pyrene-Barbital Complexes (Pyribital) Without Autooxidation by a Spray Drying Technique. *Drug Dev. Ind. Pharm.* **1983**, *9*, 285–302.
51. Broadhead, J.; Rouan, Edmond S.K.; Hau, I.; Rhodes, C.T. The Effect of Process and Formulation Variables on the Properties of Spray Dried Galactosidase. *J. Pharm. Pharmacol.* **1994**, *46*, 458–467.
52. Labrude, P.; Rasolomanana, M.; Vigneron, C.; Thirion, C.; Chaillot, B. Protective Effect of Sucrose on Spray Drying of Oxyhemoglobin. *J. Pharm. Sci.* **1989**, *78*, 223–229.
53. Mumenthaler, M.; Hsu, C.C.; Pearlman, R. Feasibility Study on Spray-Drying Protein Pharmaceuticals: Recombinant Human Growth Hormone and Tissue-Type Plasminogen Activator. *Pharm. Res.* **1994**, *11*, 12–20.
54. Clark, A.R.; Dasovich, N.; Gonda, I.; Chan, H.-K. The Balance Between Biochemical and Physical Stability for Inhalation Protein Powders: RhDNase As An Example. *Respiratory Drug Delivery V*; Dalby, R.N., Byron, P.R., Farr, S.J., Eds.; Interpharm Press: Illinois, 1996; 167–174.
55. Chan, H.-K.; Gonda, I. Solid State Characterization of Spray-Dried Powders of Recombinant Human Deoxyribonuclease (rhDNase). *J. Pharm. Sci.* **1998**, *87*, 647–654.
56. Kawashima, Y.; Matsuda, K.; Takenaka, H. Physicochemical Properties of Spray-Dried Agglomerated Particles of Salicylic Acid and Sodium Salicylate. *J. Pharm. Pharmacol.* **1972**, *24*, 505–512.
57. Takeuchi, H.; Hsasaki, T.; Niwa, T.; Kawashima, Y.; Uesugi, K.; Kayano, M.; Miyake, Y. Preparation of Powdered Redispersible Vitamin E Acetate Emulsion by Spray-Drying Technique. *Chem. Pharm. Bull.* **1991**, *39*, 1528–1531.
58. Moura, T.F.; Gaudy, D.; Jacob, M.; Terol, A.; Pauvert, B.; Chauvet, A. Vitamin C Spray Drying: Study of the Thermal Constraints. *Drug Dev. Ind. Pharm.* **1996**, *22*, 393–400.
59. Takenaka, H.; Kawashima, Y.; Lin, S.Y. Polymorphism of Spray-Dried Microencapsulated Sulfamethoxazole With Cellulose Acetate Phthalate and Colloidal Silica, Montmorillonite, or Talc. *J. Pharm. Sci.* **1981**, *70*, 1256–1260.
60. Takeuchi, H.; Handa, T.; Kawashima, Y. Enhancement of the Dissolution Rate of a Poorly Water-Soluble Drug (tolbutamide) by a Spray-Drying Solvent Deposition Method and Disintegrants. *J. Pharm. Pharmacol.* **1987**, *39*, 769–773.
61. Palmieri, G.F.; Wehrle, P.; Stamm, A. Evaluation of Spray-Drying As a Method to Prepare Microparticles for Controlled Drug Release. *Drug. Dev. Ind. Pharm.* **1994**, *20*, 2859–2879.
62. Wan, L.S.; Heng, P.W.; Chia, C.G. Citric Acid As a Plasticizer for Spray-Dried Microcapsules. *J. Microencapsulation* **1993**, *10*, 11–23.
63. Baker, J.A. Microencapsulation. *The Theory and Practice of Industrial Pharmacy*, 3rd Ed.; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea & Febiger: PA, 1986; 415–416.
64. Duncan, J.D.; Wang, P.X.; Harrington, C.M.; Schafer, D.P.; Matsuoka, Y.; Mestecky, J.F.; Compans, R.W.; Novak, M.J. Comparative Analysis of Oral Delivery Systems for Live Rotavirus Vaccines. *J. Controlled Release* **1996**, *41*, 237–247.
65. Mumper, R.J.; Hoffman, A.S.; Puolakkainen, P.A.; Bouchard, L.S.; Gombotz, W.R. Calcium-alginate Beads for the Oral Delivery of Transforming Growth Factor- β (TGF- β): Stabilization of TGF- β by the Addition of Polyacrylic Acid Within Acid-Treated Beads. *J. Controlled Release* **1994**, *30*, 241–251.
66. Molpeceres, J.; Aberturas, M.R.; Chacon, M.; Berges, L.; Guzman, M. Stability of Cyclosporin-Loaded Poly-Sigma-Caprolactone Nanoparticles. *J. Microencapsulation* **1997**, *14*, 777–787.
67. Muller, R.H.; Dingler, A.; Weyhers, H.; Muhlen, A.; Zur, W.; Mehnert, W. Solid Lipid Nanoparticles—A Novel Carrier Systems for Cosmetics and Pharmaceuticals. *Pharmazeutische Industrie* **1997**, *59*, 614–619.
68. Johansen, P.; Men, Y.; Audran, R.; Corradin, G.; Merkle, H.P.; Gander, B. Improved Stability and Release Kinetics of Microencapsulated Tetanus Toxoid by Coencapsulation of Additives. *Pharm. Res.* **1998**, *15*, 1103–1110.
69. Audran, R.; Men, Y.; Johansen, P.; Gander, B.; Corradin, G. Enhanced Immunogenicity of Microencapsulated Tetanus Toxoid With Stabilizing Agents. *Pharm. Res.* **1998**, *15*, 1111–1116.
70. Constantino, H.R.; Schwendeman, S.P.; Griebenow, K.; Klibanov, A.M.; Langer, R. The Secondary Structure and Aggregation of Lyophilized Tetanus Toxoid. *J. Pharm. Sci.* **1996**, *85*, 1290–1293.
71. Rojas, J.; Pinto-Alphandary, H.; Leo, E.; Pecquet, S.; Couvreur, P.; Fattal, E. Optimization of the Encapsulation

- and Release of Beta-Lactoglobulin Entrapped Poly (DL-lactide-co-glycolide) Microspheres. *Int. J. Pharm.* **1999**, *183*, 67–71.
72. Matthews, C.B.; Jenkins, G.; Hilfinger, J.M.; Davidson, B.L. Poly- L-Lysine Improves Gene Transfer With Adenovirus Formulated in PLGA Microspheres. *Gene Ther.* **1999**, *6*, 1558–1564.
 73. Pean, J.-M.; Boury, F.; Venier-Julienne, M.-C.; Menei, P.; Proust, J.-E.; Benoit, J.-P. Why Does PEG 400 Coencapsulation Improve NGF Stability and Release from PLGA Biodegradable Microspheres. *Pharm. Res.* **1999**, *16*, 1294–1299.
 74. Wirth, D.D.; Baertshi, S.W.; Johnson, R.A.; Maple, S.R.; Miller, M.S.; Hallenbeck, D.K.; Gregg, S.M. Maillard Reaction of Lactose and Fluoxetine Hydrochloride a Secondary Amine. *J. Pharm. Sci.* **1998**, *87*, 31–39.
 75. Monkhouse, D.C.; Maderich, A. Whither Compatibility Testing. *Drug Dev. Ind. Pharm.* **1989**, *15*, 2115–2130.
 76. Zografi, G. Status of Water Associated With Solids. *Drug Dev. Ind. Pharm.* **1988**, *14*, 1905–1926.
 77. Aso, Y.; Yoshioka, S.; Terao, T. Effect of the Binding of Water to Excipients As Measured by ^2H -NMR Relaxation Time on Cephalothin Decomposition Rate. *Chem. Pharm. Bull.* **1994**, *42*, 398–401.
 78. Kitamura, S.; Chang, L.-C.; Guillory, J.K. Polymorphism of Mefloquine Hydrochloride. *Int. J. Pharm.* **1994**, *101*, 127–144.
 79. Snavelly, M.J.; Price, J.C.; Jun, H.W. The Stability of Aspirin in a Moisture Containing Direct Compression Tablet Formulation. *Drug Dev. Ind. Pharm.* **1993**, *19*, 729–738.
 80. Otsuka, M.; Matsuda, Y. The Effect of Humidity of Hydration Kinetics of Mixtures of Nitrofurantoin Anhydride and Diluents. *J. Pharm. Bull.* **1994**, *42*, 156–159.
 81. Herman, B.D.; Sinclair, B.D.; Milton, N.; Nail, S.L. The Effect of Bulking Agent on the Solid-State Stability of Freeze-Dried Methylprednisolone Sodium Succinate. *Pharm. Res.* **1994**, *11*, 1467–1473.
 82. Zografi, G.; Grandolfi, G.P.; Kontny, M.J.; Mendenhall, D.W. Prediction of Moisture Transfer in Mixtures of Solids: Transfer Via the Vapor Phase. *Int. J. Pharm.* **1988**, *42*, 77–88.
 83. Saleki-Gerhardt, A.; Stowell, J.G.; Byrn, S.R.; Zografi, G. Hydration and Dehydration of Crystalline and Amorphous Forms of Raffinose. *J. Pharm. Sci.* **1995**, *84*, 318–323.
 84. Chan, H.-K.; Au-Yeung, K.-L.; Gonda, I. Development of a Mathematical Model for the Water Distribution in Freeze-Dried Solids. *Pharm. Res.* **1999**, *16*, 660–655.
 85. Wade, A., Weller, P.J.; Eds.; *Handbook of Pharmaceutical Excipients*, 2nd Ed.; American Pharmaceutical Association: Washington, DC, 1994.
 86. Lieberman, H.A., Lachman, L., Schwartz, J.B., Eds.; *Pharmaceutical Dosage Forms: Tablets*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1989.

Fluid Bed Processes for Forming Functional Particles

Yoshinobu Fukumori

Hideki Ichikawa

Kobe Gakuin University, Kobe, Hyogo, Japan

INTRODUCTION

Fluid bed processors have been used for drying, agglomeration, and coating. In addition to a simple fluid bed, the tumbling, agitating, centrifugal, and spiral flow fluid bed and the spouted bed with or without the draft tube have been developed for improving the process performance. Agglomerates produced from the simple fluid bed process are usually characterized with their soft, porous properties. This is advantageous in their application to tableting because the easy deformation of the agglomerates leads to the hardening of the tablets. The fluid bed processes are also favorably applied to the coating of the pharmaceutical particles including the agglomerates and crystals. Among the many types of surface modification processes, the fluid bed processes are characterized with their easy, simple mechanical formation of multilayers on the particles. This has been producing many types of functional pharmaceutical particulate system for the purpose of efficient drug delivery, including the enteric-coated particles and the sustained, prolonged, and delayed release systems.

These agglomeration and coating technologies, which are established and well experienced in the pharmaceutical area, have been extended to other areas such as food and agricultural industries. However, the current fluidization technology has a limit in the size of the particles that can be processed.^[1] The extension of the size limit to a smaller range will lead to the wider applications. From practical aspects, the fluidized bed processes will be described below.

FLUID BED PROCESSORS

Typical fluid bed processors are illustrated in Fig. 1. Fig. 1A is a simple fluid bed, usually with a conical shape at the bottom of the chamber, leading to inducing spouted particle flow more or less depending on the angle of the conical chamber. The spray is supplied from the top toward the fluidized bed in agglomeration and into the bed in a tangential direction in coating. Elutriated particles from the fluid bed are trapped mostly by the bag filter and are returned to the bed by periodical shaking of the bag filter with or without a pause of fluidization air input and

spray in the agglomeration, but always without a pause in the coating. Fig. 1B is an example of the tumbling fluid bed processors. A typical construction of the tumbling fluid bed processor automatically operated is shown in Fig. 2. The fluidization air is supplied through the slit between the turntable and the chamber. The tumbling forces the particles to be centrifuged toward the chamber wall; then, the particles are blown up by the slit air. These make a circulating fluidized bed, in which the spray is supplied by the top or the tangential mode. The rolling of the particles on the turntable makes the wet particles roundish and compact. Fig. 1C is a typical assembly of the Wurster process, a kind of a spouted bed process assisted with a draft tube. The particles fluidized in the annular part between the draft tube and the chamber are introduced into the draft tube because of the accelerated airflow from the bottom and are then spouted from the draft tube. The particle velocity is reduced in the upper expansion chamber, leading to the return of the particles to the fluid bed in the annular part. During this circulating particle flow, the particles are sprayed in the draft tube; at the same time, they have a chance to be exposed to the spray air jet, which can exert a strong separation force to the particles. Usually, the particles are easily agglomerated or coagulated during spraying, but the above strong air jet can make the particles disintegrated. These characteristics of the Wurster process make it possible for the coating of the finer particles^[2] or for the production of the finer agglomerates for the purpose of agglomeration^[3] compared with the other type of the fluid bed processes.

The types of fluidization, as proposed by Geldart,^[4] depend on the particle size and density. Because the particle density of the pharmaceutical powders is mostly around 1.5 g/cm³, the particles are categorized into larger than 900 μm (D type), 900–150 μm (B type), 150–20 μm (A type), and smaller than 20 μm (C type) particles in the Geldart's fluidization map. The A and C types are cohesive and adhesive; therefore they are agglomerated into the D and B types to achieve free-flowing properties. This is the main purpose of agglomeration. The D type cannot be fluidized because of too large particle size; therefore the coating is simply processed using a rotating pan. The B type always exhibits bubbling fluidization which induces inhomogeneous particle flow; this is practically



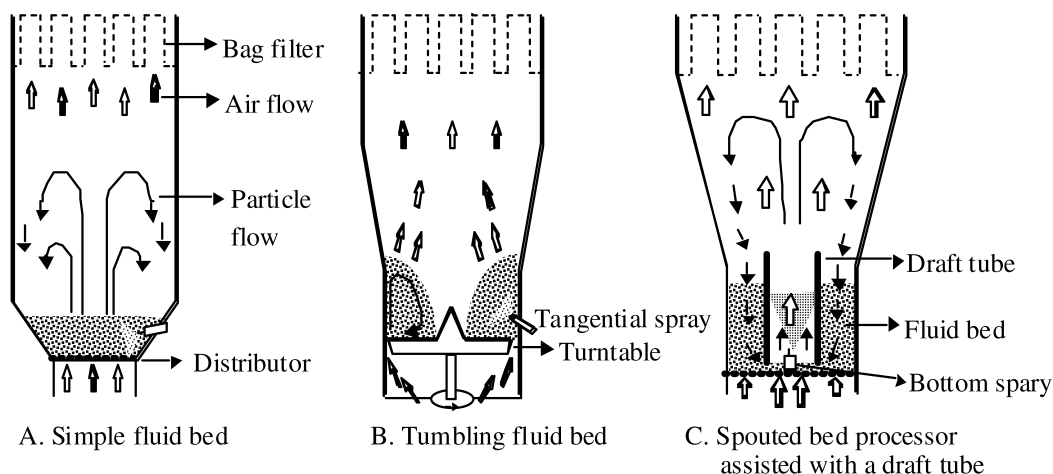


Fig. 1 Schematic drawing of the typical fluid bed processors.

disadvantageous because such a flow especially in the spray zone leads to poor coating performance in the yield and homogeneity of the product. In order to achieve a homogeneous particle flow of the B-typed particles, particle flow patterns, different from the simple fluidization such as those in the tumbling, centrifugal, and spiral fluid

bed and the spouted bed, have been required. This is the reason why many different types of the fluid bed processor have been developed so far. The A-typed particles can be fluidized homogeneously without air bubbles, but the separation force from fluidization is not sufficient to avoid agglomeration. This prevents the fluid bed coating

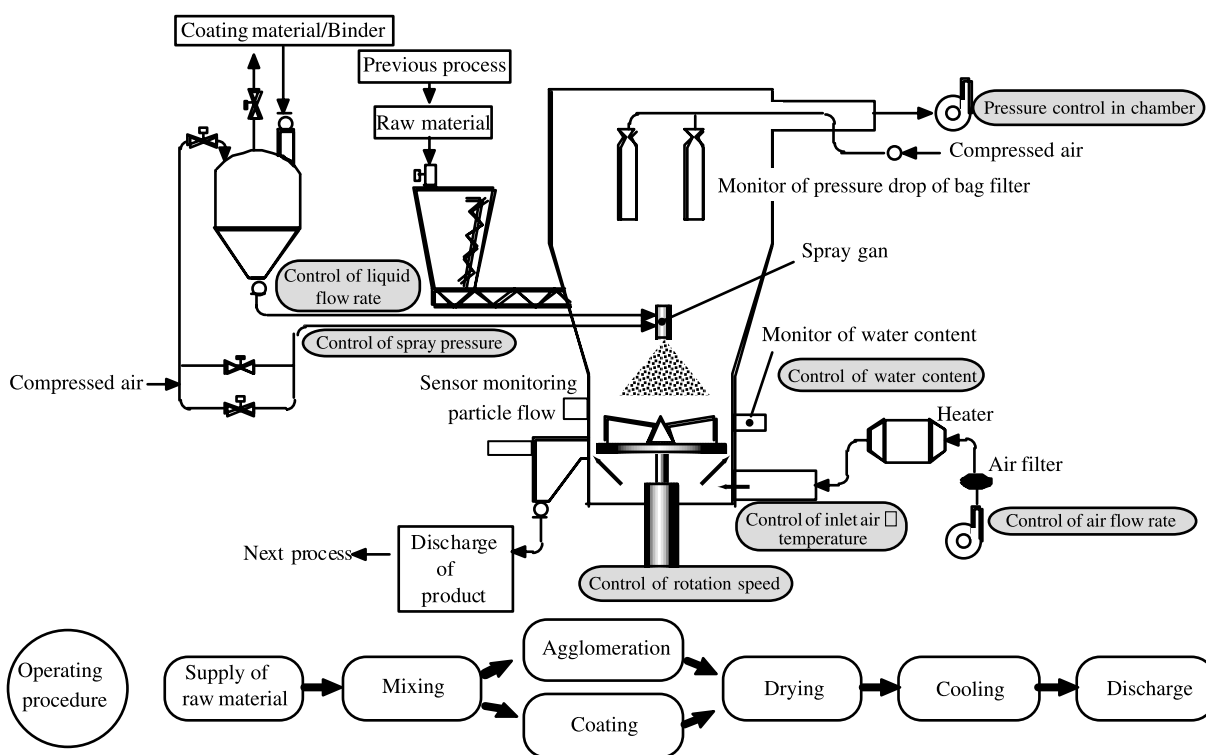


Fig. 2 Tumbling fluid bed processor (Agglomaster, Hosokawa Micron Corp.) operated automatically.

process from being industrially applied to these small particles despite many researches attempting the fine particle-coating technologies. The type C particles cannot be fluidized because of their small size.

When the particles are fluidized or spouted under spraying, separation force is exerted to the particles more or less agglomerated by the spray solution. Balance of the

separation force and the binding strength of the binder or coating material determines the degree of agglomeration, i.e., agglomerate size.^[5,6] Each fluid bed processor can generate its inherent separation force, surely depending on the operating conditions such as the airflow pattern and the inlet airflow rate. The excessive separation force can even disintegrate the core particles. Thus, depending on

Table 1 Binders, coating materials, and cores

Type	Brand	Supplier	Solvent	Soluble in:	Main component
Spray solution	Kolidon VA64	BASF	Water	Water	6:4 poly(VP/VA)
	Kolidon 20, 30	BASF	Water	Water	Polyvinylpyrrolidone
	Opadry	Colorcon	Water	Water	Mixture of water-soluble polymer, plasticizer, and pigment
	TC-5	Shin-Etsu	Water	Water	Hydroxypropylmethylcellulose
	HPC-H, M, L, SL, SSL	Nippon Soda	Water	Water	Hydroxypropylcellulose
	CMEC	Freund	Water-ethanol	Intestine	Carboxymethylethylcellulose
	HPMCP	Shin-Etsu	Water-ethanol	Intestine	Hydroxypropylmethylcellulose phthalate
	Eudragit E100	Röhm	Organic solvent	Stomach	Poly(BMA/MMA/DAEMA)
	Eudragit L/S	Röhm	Organic solvent	Intestine	Poly(MMA/MAA)
	Eudragit RS100/RL100	Röhm	Insoluble	(Insoluble)	1:2:0.1/1:2:0.2 Poly(EA/MMA/TAMCl)
Spray dispersion	EC N-10F	Shin-Etsu	Water	Insoluble	Ethylcellulose ground powder
	Aquacoat	FMC	Water	Insoluble	Ethylcellulose pseudolatex
	Surelease	Colorcon	Water	Insoluble	Ethylcellulose pseudolatex
	Eudragit RS30D	Röhm	Water	Insoluble	1:2:0.1 Poly(EA/MMA/TAMCl) pseudolatex
	Eudragit RL30D	Röhm	Water	Insoluble	1:2:0.2 Poly(EA/MMA/TAMCl) pseudolatex
	Eudragit NE30D	Röhm	Water	Insoluble	2:1 Poly(EA/MMA) latex
	Acoat	Shin-Etsu	Water	Intestine	Hydroxypropylmethylcellulose acetate succinate
	Kollicoat MAE30D/DP	BASF	Water	Intestine	1:1 Poly(EA/MAA)
	Eudragit L30D	Röhm	Water	Intestine	1:1 Poly(EA/MAA)
	Lubri wax 101/103	Freund		Insoluble	Hydrogenated castor/rape oil
Dry powder	Polishing wax PEP 101	Freund		Insoluble	Carnauba wax
		Freund		Insoluble	Poly(ethyleneoxide/propylene-oxide)
Core	Nonpareil 101/103/105	Freund			Granules of sucrose-starch/sucrose/lactose-microcrystalline cellulose
	Celphere 102/203/305/507	Asahi Kasei			Granules of microcrystalline cellulose

BMA: butyl methacrylate, EA: ethyl acrylate, DAEMA: dimethylaminoethyl methacrylate, MAA: methacrylic acid, MMA: methyl methacrylate, TAMCl: trimethylammonioethyl methacrylate chloride, VA: vinyl acetate, VP: vinylpyrrolidone.



the requisite of the final agglomerate size or the core particle size to be discretely coated, the optimal processor should be selected in practice.

MATERIALS FOR AGGLOMERATION AND COATING

Table 1 lists the typical binders and the coating materials for agglomeration and coating. The water-soluble polymers are used as a binder in agglomeration, and some of them are also used as a coating material. As each type of the water-soluble polymers becomes higher in molecular weight, it contributes more to the increase in the viscosity of its solution and the interparticle binding strength, leading to more enhanced particle growth in agglomeration and coating.^[5,6]

The commercially available polymeric dispersions, which have been most widely used as a coating material, are classified into three types based on the preparation methods:^[7] 1) latexes synthesized by emulsion polymerization; 2) pseudolatexes prepared by emulsion processes such as emulsion solvent evaporation, phase inversion, and solvent change; and 3) dispersions of micronized polymeric powders.

Eudragit L30D and NE30D are acrylic copolymer latexes synthesized by emulsion polymerization.^[8] The

particle sizes of these latexes are in submicron order. L30D is a copolymer of ethyl acrylate (EA) as an ester component with methacrylic acid (MA) (MA/EA 1:1). It is used for enteric coating because of the presence of carboxyl groups in the copolymer. NE30D is a copolymer of ester components only, EA and MMA (2:1). The films formed from NE30D have a very low softening temperature and hence are flexible and expandable even under indoor conditions.

Cellulose derivatives cannot be synthesized directly in the latexes; therefore they are prepared as pseudolatexes (Aquacoat, Aquateric,^[9] and Surelease^[10]) or micronized powders [Acoat hydroxypropylmethylcellulose acetate succinate (HPMCAS)^[11] and EC N-10F]. While the pseudolatexes can be prepared as submicron particles, the micronized powders have mean particle sizes of a few micrometers. Poly vinyl acetate phthalate (VAP) is also supplied as a micronized powder (Coateric).^[10] Eudragit RS and RL are terpolymers of EA and MMA as ester components with trimethylammonioethyl methacrylate chloride (TAMCl) as hydrophilic quaternary ammonium groups; RS and RL are 1:2:0.1 and 1:2:0.2 terpoly(EA/MMA/TAMCl), respectively. Because Eudragit RS and RL contain MMA-rich ester components (EA/MMA 1:2), they have softening temperatures higher than those of NE30D (EA/MMA 2:1) and form hard films under indoor conditions. Eudragit RS and RL powders are easily trans-

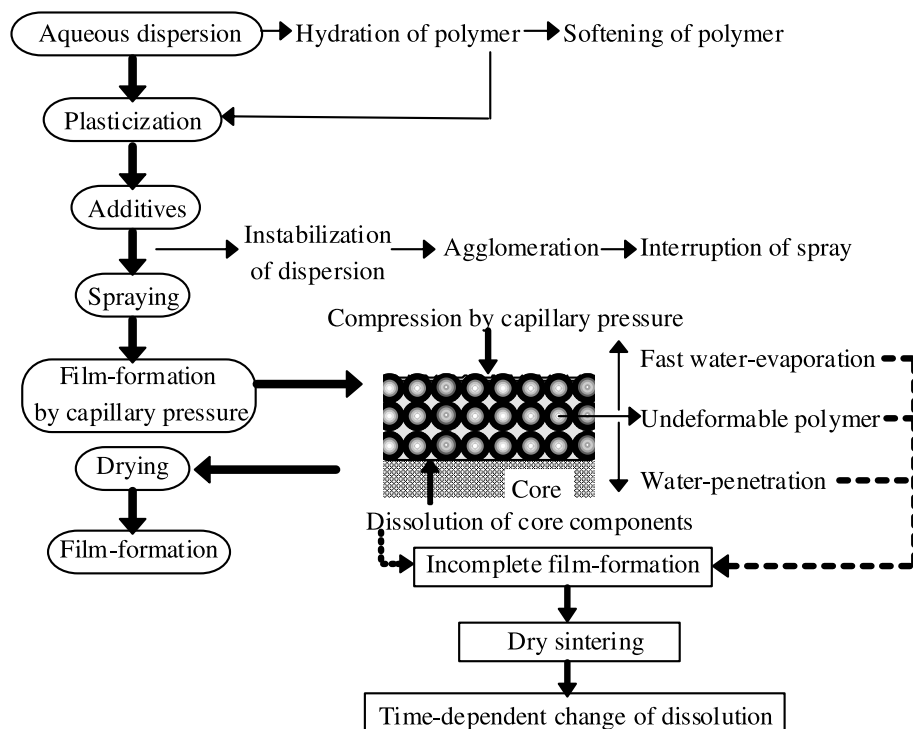


Fig. 3 Film formation from the aqueous polymeric dispersions.

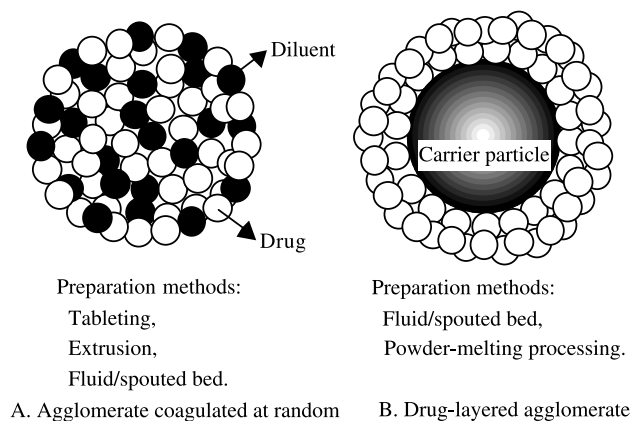


Fig. 4 Structure of agglomerates.

formed into pseudolatexes by emulsifying their powders in hot water without additives.^[8] It is costly to ship the aqueous dispersions around the world; therefore the Aquateric pseudolatex is supplied as a spray-dried powder. It is redispersed just before use.^[9]

A variety of additives are incorporated into the dispersions as surfactants (Tween 80, sodium lauryl sulfate, polyoxyethylene nonyl phenyl ether, cetyl alcohol, and Pluronic F-68), plasticizers (dibutyl sebacate, oleic acid, and Myvacet 9-40), pigments, anti-adherents (fumed silica), anticoagulant (Myvacet 9-40), preservatives (sorbic acid), and stabilizers (ammonia).^[7]

The film formation process from the aqueous dispersion is shown schematically in Fig. 3. The mechanisms of film formation from the aqueous polymeric dispersions have been discussed for a long time, and many theories have been proposed. The detail was reviewed by Muroi^[12] from a basic point of view. Film forma-

tion in pharmaceutical applications was discussed by Lehmann,^[8] Steuernagel,^[9] and Fukumori.^[7] Fusion and film formation of the polymeric particles during the coating process can be explained by the wet sintering theory for particles suspended in water, the capillary pressure theory for particle layers containing water in various degrees of saturation, and the dry sintering theory for dry particle layers.

PARTICULATE DESIGN AND PREPARATION

The typical particle structures produced from the agglomeration process are shown in Fig. 4. Simple agglomerates (Fig. 4A) are prepared by fine drug powder(s) being mixed with fine powders of additives in the fluid bed, and then binder solution being sprayed from the top in the simple fluid bed (Fig. 1A) followed by drying. In this case, because the particles are agglomerated in the fluidization airflow, they are usually very porous. The tumbling fluid bed or the spouted bed process assisted with draft tube (Fig. 1B and C) can be applied if more compact agglomerates are required. When fine drug powder is agglomerated with the coarse carrier particles (cores in Table 1), drug-layered agglomerates shown in Fig. 4B can be prepared.^[13] In the latter process, the tumbling fluid bed is often used because tumbling on the turntable contributes to the efficient layering of the fine powder. The drug-layered agglomerates are often used as core particles in the coating process.

Among the many kinds of the surface modification process, the fluid bed processes can most easily produce multilayered particle structure with each layer being monolithic, random multiphase structure, ordered multiphase structure, and so on (Fig. 5). The combination of

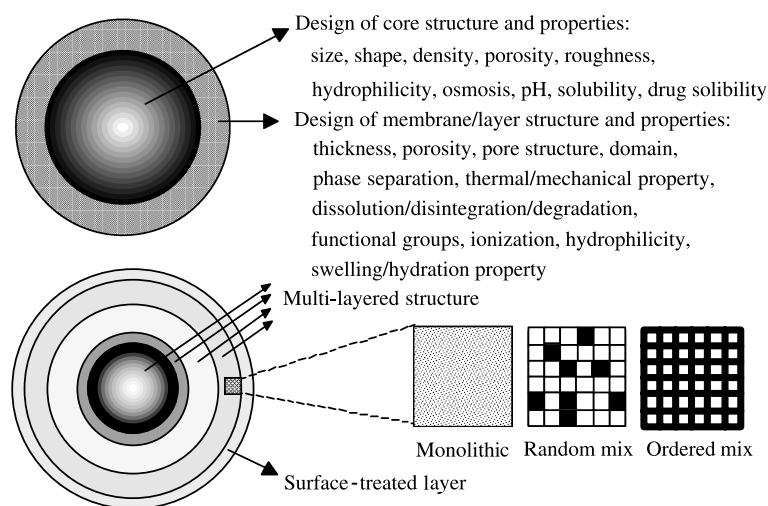


Fig. 5 Particulate structures produced by the fluid bed process.



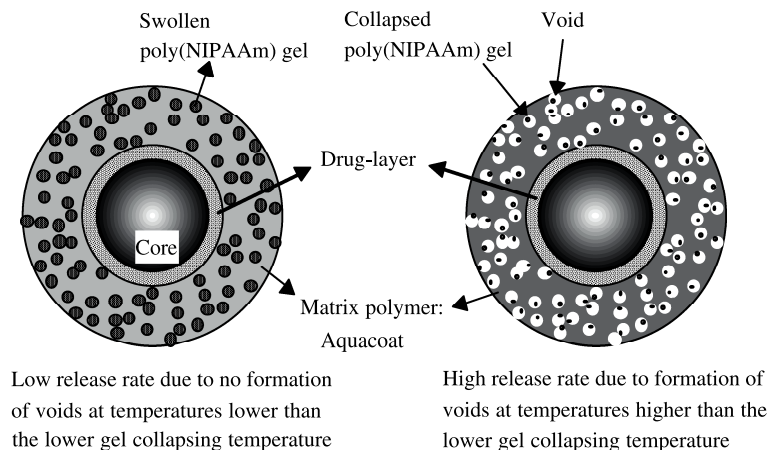


Fig. 6 Schematic diagram of the structure of the particle for positively thermosensitive release of drug utilizing gel-collapsing behavior of poly(*N*-isopropylacrylamide).

the different components and layers can produce almost infinite types of functional particles. As an example, designs and preparations of several thermosensitive controlled-release particles will be described below.

Controlled-release technology based on the external temperature-activated release can find application in diverse industrial fields.^[1,14–16] In the pharmaceutical area, for example, the deviation of the body temperature from the normal temperature (37°C) in the physiological presence of the pathogens or pyrogens can be utilized as a useful stimulus that induces the release of the therapeutic agents from a thermosensitive controlled-release system. Physically controlled temperature using a heat source such as the microwaves from outside the body can also be used for temperature-activated antitumor drug release combined with the local hyperthermic treatment of cancer.

The membranes of the thermosensitive controlled-release microcapsules were constructed by a random mixing Aquacoat (Table 1) with the latex particles having poly(EA/MMA/2-hydroxyethyl methacrylate) core and poly(*N*-isopropylacrylamide (NIPAAm)) shell. This is an example where the membrane has the random two-phase structure as shown in Fig. 5. The microcapsules exhibited a thermosensitive release of water-soluble drug.^[17] The mechanism is explained in Fig. 6. When the temperature was changed in a stepwise manner between 30 and 50°C, the microcapsules showed an “on–off” pulsatile release. This “on–off” response was reversible.

Alternatively, this type of thermosensitive microcapsule can be prepared even with already established pharmaceutical ingredients. As is well known, hydroxypropyl cellulose (HPC), the commonly used binder and coating substrate (Table 1), has a lower critical solution temperature (LCST) around 44°C, and its water solubility dras-

tically changes across the LCST. Negatively thermosensitive drug release microcapsules with an HPC layer were thus designed by utilizing the thermally reversible dissolution/precipitation process resulting from the LCST phenomena^[18] (Fig. 7). This is the case that the particles are constructed with multilayers of monolithic structure (Fig. 5). The release rate from the microcapsules with sandwiched HPC layer was found to be drastically decreased when the temperature came near the LCST: the release rate at 50°C was approximately 10 times lower than that at 30°C. In contrast, no negatively thermosensitive release was obtained in the microcapsules without the outer coat of ethyl cellulose: the drug release rate was monotonously increased as the temperature increased.

Unlike many types of the millisized devices that have been developed so far, the microparticles mentioned here were around 100 µm in diameter, and the estimated membrane thickness was only several tens of microns. Such small dimensions of the thermosensitive materials incorporated into the microcapsule membranes provided a

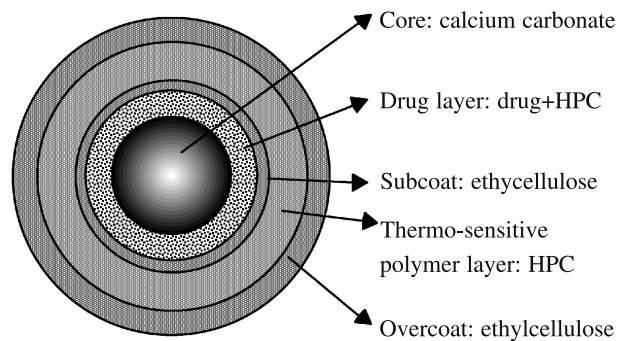


Fig. 7 Particle exhibiting negatively thermosensitive drug release.

sharp release rate change in response to the temperature change. This may be one of the advantages of the fine particle-coating technology. The above-demonstrated flexibility in designing the membrane structures also offers considerable advantages over conventional microencapsulation methods.

CONCLUSION

Using the fluid bed processes and processors along with the appropriate materials and their well-designed formulation and particulate structure make it possible to prepare highly functional particles as demonstrated here. However, this method has an unavoidable limit in the size of particles that can be efficiently processed because a steady circulation or fluidization of the particles smaller than 20 μm is still difficult. In order to expand the applications of this process, some new or improved fluidization technologies will be required.

ACKNOWLEDGMENT

The work described in this paper is financially supported in part by a grant-in-aid for Scientific Research (C) (12672098) from Japan Society for the Promotion of Science.

REFERENCES

1. Jono, K.; Ichikawa, H.; Miyamoto, M.; Fukumori, Y. A review of particulate design for pharmaceutical powders and their production by spouted bed coating. *Powder Technol.* **2000**, *113*, 269–277.
2. Ichikawa, H.; Tokumitsu, H.; Jono, K.; Osako, Y.; Fukumori, Y. Coating of pharmaceutical powders by fluidized bed process. VI. Microencapsulation using blend and composite latices of copoly(ethyl acrylate-methyl methacrylate-2-hydroxyethyl methacrylate). *Chem. Pharm. Bull.* **1994**, *42* (6), 1308–1314.
3. Ichikawa, H.; Fukumori, Y. Microagglomeration of pulverized pharmaceutical powders using the Wurster process. I. Preparation of highly drug-incorporated, subsieve-sized core particles for subsequent microencapsulation by film-coating. *Int. J. Pharm.* **1999**, *180*, 195–210.
4. Geldart, D. Types of gas fluidization. *Powder Technol.* **1973**, *7*, 285–292.
5. Fukumori, Y.; Ichikawa, H.; Jono, K.; Takeuchi, Y.; Fukuda, T. Computer simulation of agglomeration in the Wurster process. *Chem. Pharm. Bull.* **1992**, *40* (8), 2159–2163.
6. Fukumori, Y.; Ichikawa, H.; Jono, K.; Fukuda, T.; Osako, Y. Effect of additives on agglomeration in aqueous coating with hydroxypropyl cellulose. *Chem. Pharm. Bull.* **1993**, *41* (4), 725–730.
7. Fukumori, Y. Coating of Multiparticulates Using Polymeric Dispersions. Formulation and Process Consideration. In *Multiparticulate Oral Drug Delivery*; Ghebre-Sellassie, I., Ed.; Marcel Dekker: New York, 1994; 79–111.
8. Lehmann, K.O.R. Chemistry and Application Properties of Polymethacrylate Coating. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*; McGenity, J.W., Ed.; Marcel Dekker: New York, 1989; 153–246.
9. Steuernagel, C.R. Latex Emulsions for Controlled Drug Delivery. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*; McGenity, J.W., Ed.; Marcel Dekker: New York, 1989; 1–62.
10. Moore, K.L. Physical Properties of Opadry, Coateric, and Surelease. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*; McGenity, J.W., Ed.; Marcel Dekker: New York, 1989; 303–316.
11. Nagai, T.; Sekikawa, F.; Hoshi, N. Applications of HPMC and HPMCAS Aqueous Film Coating of Pharmaceutical Dosage Forms. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*; McGenity, J.W., Ed.; Marcel Dekker: New York, 1989; 81–152.
12. Muroi, S. *High Polymer Latex Adhesives*; Kobunshikankokai: Kyoto, Japan, 1984.
13. Fukumori, Y.; Yamaoka, Y.; Ichikawa, H.; Fukuda, T.; Takeuchi, Y.; Osako, Y. Coating of pharmaceutical powders by fluidized bed process. II. Microcapsules produced by layering of fine particles and subsequent aqueous enteric coating. *Chem. Pharm. Bull.* **1988**, *36* (4), 1491–1501.
14. Kaneko, Y.; Sakai, K.; Okano, T. Temperature-Responsive Hydrogels as Intelligent Materials. In *Biorelated Polymers and Gels*; Okano, T., Ed.; Academic Press: Boston, 1998; 29–70.
15. Ichikawa, H.; Fukumori, Y. New applications of acrylic polymers for thermosensitive drug release. *STP Pharma* **1997**, *7* (6), 529–545.
16. Peppas, N.A.; Bures, P.; Leobandung, W.; Ichikawa, H. Hydrogels in pharmaceutical formulations. *Eur. J. Pharm. Biopharm.* **2000**, *50* (1), 27–46.
17. Ichikawa, H.; Fukumori, Y. A novel positively thermosensitive controlled-release microcapsules with membrane of nano-sized poly(*N*-isopropylacrylamide) gel dispersed in ethylcellulose matrix. *J. Controlled Release* **2000**, *63*, 107–119.
18. Ichikawa, H.; Fukumori, Y. Negatively thermosensitive release of drug from microcapsules with hydroxypropyl cellulose membranes prepared by the Wurster process. *Chem. Pharm. Bull.* **1999**, *47* (8), 1102–1107.

FLOW PROPERTIES OF SOLIDS

Stephen A. Howard

Purdue Pharma L.P., Ardsley, New York

INTRODUCTION

The preparation of essentially all dosage forms involves the handling of solid materials. Among all finished products, solid dosage forms are the most predominant in terms of volume and value. The importance of solid-handling properties, especially flow properties, cannot not be overemphasized. The flow properties of solids have great impact on the tableting and encapsulation processes since these dosage form manufacturing processes require the flow of powder materials from a storage container to filling stations, such as tablet dies or capsule fillers. Weight uniformity of course is dependent on the uniform and rapid flow of powders. The flow properties of solids also have great influence on the mixing and demixing of powders that take place before tableting or encapsulation. As is pointed out by Von Behren (1) in process steps affected by the flowability of formulations include mix uniformity and flow and of course the weight and pressure settings used. It should also be pointed out that the speed of production is also greatly affected by the formulation's flow characteristics. For the final product, Von Behren (1) lists, weight, content uniformity, hardness and disintegration/dissolution as being affected by formulation flow.

Different flow properties are required at different stages of processing and should be carefully taken into consideration during formulation and process validation. Kinetics of mixing is influenced by the physical state of the active constituent (2). Particle size of excipients has a significant effect on the content uniformity of ethinyl estradiol powder mixes and tablets. As the particle sizes of excipients increase, the degree of mixing decreases. The more free-flowing the excipient, the more abrasive the particle, facilitating the breakdown of drug agglomerates (3). On the other hand, segregation studies carried out in a two-dimensional segregation cell suggest that materials with good flow show demixing tendencies, whereas powder blends with poor flow are less apt to separate (4). Fast flow is not always helpful in weight uniformity of the finished product. Hauer et al. (5) showed that none cohesive free flowing powders gave poorer capsule fill uniformity since they were less cohesive and flowed out of

the dosators. Knowing your process and determining the characteristics needed to optimize that process is critical.

Characterization of powders is essential to quality control of raw materials, active or excipient, in order to maintain product uniformity. Flow-property studies of powder materials facilitate the scientific design of formulations and processing equipment, such as the design of mass-flow hoppers. This article reviews the flow properties of powder materials. The factors affecting the flow properties of solids are briefly discussed first, followed by measurement of flow properties.

FACTORS INFLUENCING THE FLOW OF SOLIDS

Nature of Powders and Granulations

Powders are generally considered as two-phase assemblies of discrete particles with interactions between gas and solid internal surfaces. A material is classified as a powder if it is composed of dry, discrete particles with a maximum dimension of less than 1000 μm according to British Standard 2955 (6). Powders differ from other physical states of matter since they are nonhomogenous in nature but consist of discrete solid particles of different sizes and shapes interdispersed with a gaseous phase. Powders are similar to solids in that they can exhibit both elastic recovery and brittle fracture. Unlike solids, however, powders can expand or contract when stressed. Preconditioning can change the nature of the material. The condition and duration preconditioning forces are applied to a powder can determine if and how a powder will flow. Under stress, powders can flow like a liquid. However, unlike liquids, they do not flow if the stresses are too small. When powder materials do flow, the stresses are not dependent on the rate of flow as they are with a liquid. In order to characterize the properties of an assembly of particles or a bulk mass, the collective properties of constituent particles within their gaseous environment must be determined. The solid-handling properties of a bulk mass are influenced by any factor that can have an effect on the particle-particle interactions of constituent particles. Factors associated with the nature of the particles

and their surfaces such as size, shape, surface morphology, packing conditions, and interparticle forces must therefore be considered. To make the situation more complex, the interparticle forces can be of a number of types: mechanical forces, surface tension, electrostatic forces, van der Waals forces, solid-bridge forces, or plastic welding forces; none of these can be readily quantified. The properties and phenomena associated with an assembly of particles are (7):

- Particle size distribution and specific surface area;
- Particle shape distribution;
- Cohesion, strength, and adhesion;
- Packing properties (bulk density, porosity);
- Rate and compressibility of packing;
- Flowability and failure properties;
- Segregation; and
- Angle of internal friction. A combination of these properties determines the behavior of bulk material.

Particle Size and Size Distribution

All matter interacts. As the dimensions of particles increase and the particles change in nature, the forces acting on them change. Fine powder particles less than 100 μm in diameter are acted upon primarily by surface forces. Particles above 1000 μm in diameter are governed by gravitational forces. Therefore, the balance of interaction forces determines powder behavior. With relatively small particles, the flow through an orifice may be restricted because the cohesive forces between the particles are of the same magnitude as gravitational forces. Since the latter forces are a function of the diameter raised to the third power, they become more significant as the particle size increases and flow is facilitated. Too large a particle however with respect to the orifice through which it has to flow can cause arching that can block flow from hopper or into a die or capsule dosator. The properties of solids that determine the magnitude of particle–particle interactions have been reviewed (8, 9).

In general larger particle flow faster than smaller particles. In the TSI Aero-Flow brochure (10) two sieved fractions of lactose, one larger than 38 μm and one less than 38 μm , were compared to the unfractionated lactose. As expected the fine fraction showed the poorest flow while the larger fraction showed the best flow characteristics in the Aero-Flow rotating drum tester while the unfractionated sample was between the two extremes. With wider particle distributions and gravimetric test methods such as flow from a hopper, density differences can be a determining factor until the cohesive nature of the fine particles decrease flow. Particle size and size

distribution has been systematically investigated for their effects on flow using a flat-bottom flowmeter (11). Increase in the content of fines increases the flow rate to a maximum value, followed by a decrease in flow rate if the content of fines is further increased. For a given concentration of fines, the flow rate increases to a maximum as the diameter of the fines decreases to approximately 90 μm ; further reduction in fine diameter reduces the flow rates. Equations of the Brown–Richards type relationship are adequate for modeling the static flow in such systems. Equation (1) relates the flow rate Q to powder density ρt and circular aperture diameter D_A :

$$D_A K [4Q/60\pi \cdot \rho t \cdot \sqrt{g}]^{1/n} \quad (1)$$

where K and n are material-dependent constants and g is the gravitational force.

In a similar study of the factors influencing the flow of lactose granules, a strong negative correlation was found between the flow rate of granules and proportion of particles less than 150 μm (100 mesh) in size (12).

A decrease in particular size of salicylic acid resulted in a decrease in the angle of internal flow Ψ , which was derived empirically as a measure of cohesiveness (13). The value of Ψ was found to be dependent on the relative proportion and particle size of salicylic acid and lactose when blends of these powders were studied. Geoffroy and Carstensen (14) have used shear cell measurements of sodium chloride, granular dicalcium phosphate (DiTab) and hydroxyapatite (TriTab) to examine the affect particle size has on the constants in the shear cell. They modified the Warren–Springs equation (Eq. 5 found in the Shear Cell section of this article) which relates cohesion and tensile strength of cohesive powders. They showed that cohesion stress (C) is related to tangential force (T) by the equation $C = \alpha T^\beta$ where α and β are constants dependent on the materials used. They also demonstrated relationship of α to particle diameter. They incorporated this into the Warren–Springs equation so that particle size could be better accounted for in analyzing shear cell data.

Shape Factors and Surface Morphology

Particle sizes combined with shape factors have been the subject of many of the recent studies regarding flow of solids. Sphericity, circularity, surface-shape coefficient, volume-shape coefficient, and surface-volume-shape coefficient are some of the most commonly used shape factors. It is generally accepted that the flowability of powders decreases as the shapes of particles become more irregular. Efforts to relate various shape factors to powder

bulk behavior have become more successful recently, primarily because of the fact that shape characterization techniques and methods for physically sorting particles of different shapes are improving. This is primarily thanks to the use of fractal geometry.

Fractal geometry was developed by B.B. Mandelbrot and was most fully described in a 1983 text (15). Fractal geometry essentially describes the space filling ability of a rugged line or surface by adding a fractional number to the to the topological dimension of a system as illustrated in Fig. 1. Brian Kaye and co-workers have extended Mandelbrot's concepts to describe the structure of rugged shaped powder grains and later to the distribution function of a powder (16–19). The use of fractal geometry and chaos theory has been used to develop a test method, using a rotating drum (10, 18, 19) and this test method will be described later in the article.

Hickey and Concessio (20) studied a series of powders of pharmaceutical interest using a vibrating spatula. In all

of the powders studied, mass flow rates increased as the particle size increased. Sodium chloride, which did not exhibit fractal behavior (very low fractal number), showed no irregular flow patterns. Lactose, cromolyn sodium and charcoal that did exhibit fractal behavior at low stride lengths, demonstrated irregular flow patterns. Kaye in unpublished data described the avalanching behavior of five lactose powders. Though similar in particle size, the various lactose samples showed different morphology and different avalanche behavior. Brittain (21) tested anhydrous and fast flow lactose, using Carr index and showed that the Fast-Flo lactose yielded higher flow capacity. He attributed this to the morphology of the two materials and the rounded edges in the Fast-Flo material. Cartillier and Tawashi (22) also related the flow properties and packing characteristics of seven different lactose powders as measured by the angle of repose and flow through an orifice to particle morphology, especially micro morphology as determined through a scanning electron

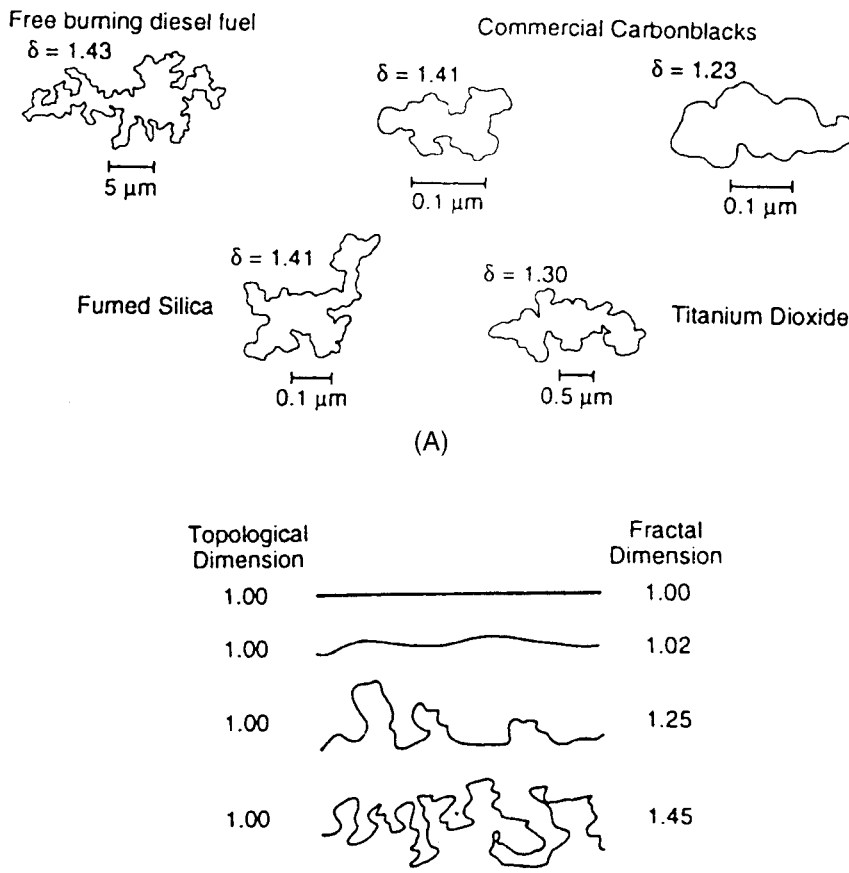


Fig. 1 (A) Profiles of fumed pigment fine particles where the fractal dimension, δ , describes the ruggedness of the profile. (B) Four lines with identical topological dimensions with varying degrees of ruggedness as seen from their corresponding fractal dimensions. (From Ref. 18.)

microscope, and expressed in various parameters including fractal dimensions.

The angle of repose of granules prepared by five different methods was found to be primarily a function of surface roughness. The Hausner ratio (discussed later) has been related to the morphological properties of sands (23).

The shapes of components being mixed have a great impact on the mixing rate and the physical stability of the resultant mixture. Lactose and calcium carbonate of different particle shapes were mixed in a Y-cone mixer. The time required to achieve an acceptable standard deviation of mixing σ_A increased with the irregularity of the particles of both components, and the mixtures containing irregularly shaped particles segregated less on subsequent vibration (24).

Some work has shown a direct correlation between shape factor and the flow properties of powders. The flowability of fine powders, as measured by a shear-cell as well as by Carr's method, was found to increase with increasing sphericity, where the sphericity is indicated by a shape index Ψ approaching one, as measured by an image analyzer (25). Huber and co-workers (26) derived an equation in which flow rate was correlated to the volume specific surface as measured by laser diffractometry. Reasonable predictions were made for individual powders as well as binary and ternary mixtures.

Moisture and Static Charge

The affect of humidity can vary drastically from powder to powder. Absorbed moisture in solids can exist either in the unbound state or as part of crystal structure. It exerts its effect directly by changing the surface properties of the particle. It can also affect flow properties indirectly and permanently through the formation of granules, which are held, together by solid bridges generated by hydration and dehydration of a binder. The process of wet granulation can be viewed as an intentional use of moisture (or other organic liquid) in controlling powder flow properties. Moisture significantly influences powder flowability as measured by the tensile strength of powders by forming liquid bridges. The increase in tensile strength has been translated into increase in torque or power consumption in a mixer and has been utilized for the monitoring of wet-granulation processes. At higher moisture content and higher packing densities, liquid bridges may progress from pendular to funicular bonds. The effect of moisture varies, depending on the degree of packing or the porosity of the powder bed. For a porous and cohesive material, flowability is not affected by moisture since the moisture can penetrate to the inside of particles without the formation of liquid bridges. A single lot of

microcrystalline cellulose NF was placed in various humidity conditions and the moisture content determined through loss of drying (27). Flow parameters were determined through the compressibility index determined by tap density and also through the use of a shear cell. It was shown that increased moisture decreased flow and that once the moisture content of the microcrystalline cellulose exceeded 5% its flow was predicted to be poor. Differences in flow rates of microcrystalline cellulose through an orifice were attributed partially to differences in moisture contents that affected cohesiveness (28). Cohesiveness of two grades of microcrystalline cellulose (Avicel PH101 and Emcocel) measured with a sandwich rheometer, peaks at approximately 20–25 wt% moisture content. Avicel was found to be more cohesive than Emcocel at a moisture content less and 30 wt%, whereas at higher moisture content the cohesive behavior was comparable (29).

Very low moisture can hinder flow since you are more likely to develop electrostatic charging. Particles acquire static charge most commonly through grinding, attrition, and collision, a phenomenon generally known as triboelectrification. Surface charge can also be generated by the sudden separation of their closely contacted, dissimilar surfaces, as in the case of sieving, mixing, or the movement of dry particles through a hopper or over a belt. Excipient powders are generally charged negatively in contact with metal or glass surfaces, whereas many are charged positively in contact with plastic surfaces (30). Most pharmaceutical excipients have low resistivity and therefore lose electrostatic charge through earth leakage relatively quickly. Electrostatic charge interactions can be controlled with beneficial results in terms of improved physical stability of normally unstable, segregating systems. Physical stability can be improved with or without permanent electrification (31).

Powder Cohesion and Storage Compaction

The storage condition a powder is placed under will have a great affect on the flow characteristics it exhibits. As stated in recent article by Marinelli (32) "as a solid remains at rest in a bin or hopper, it can become more cohesive and difficult flowing." Hopper and bin load levels, vibratory forces, time in storage, temperature of storage as well as the intrinsic cohesiveness of the material will alter its flow characteristics. Shear cells have been designed to measure the cohesion of materials under load and after preconditioning. The nature of these tests and how they work is described in more detail later on in this chapter.

Effects of Temperature

The cohesion of powder as measured in a Jenike or annular shear cell decreases as the temperature is decreased (33). This is attributed to the reduction in plasticity and in the inability of asperities on the surface of neighboring particles to form welded bonds as the temperature of the sample is lowered (34). Similarly, flow rates of powdered sugars and fatty acids through a circular orifice decrease with increasing temperature (35). An increase in tensile strength of powders is also observed for lactose and griseofulvin with increasing temperature and explained on the same basis (36). Pilpel (37) noted that an important property of a powder determining whether or not it was sensitive to temperature changes when studying flow behavior was its homologous temperature. The homologous temperature for any given material is the investigational temperature as a fraction of the melting point of the material in absolute degrees (Kelvin). Stearic acid that has a melting point of 343 K when tested at 25°C (298 K) would have a homologous temperature of 298/343 K equal to 0.87. In his investigations Pilpel found that powders started to have less flowability at homologous temperatures above 0.9. As you would suspect materials with lower melting points will have more flow problems than would higher melting point solids. Pilpel also noted that when powders that would not flow out of a funnel at room temperature were cooled to -25°C they flowed easily.

MODELS OF HOPPER FLOW

The design of hoppers and bins has been studied extensively in the last 35 years especially in the mining, food and chemical industry. Due to the large volumes of powders handled in these industries the design of storage and handling materials are a major investment. During the late 1950s and early 1960s Jenike employed a soil mechanics approach to powder handling, developing a logical, theoretical basis for bulk solids flow (38). In general, two flow patterns can be used to describe material flow through a hopper, although systems exist in between. When materials flow by gravity, the entire content of the hopper moves together and the material discharges on a first-in first-out basis. This is the case of mass flow. When the movement is restricted to a central region, leaving a relatively stationary zone in the periphery, the situation is known as funnel flow or core flow. Alternatively, flow blockage occurs by doming or bridging where a stable arch is formed across the outlet of a hopper (Fig. 2a). With highly cohesive powders or poorly designed hoppers when the gate is opened only the material above the opening

flows out and then no further flow occurs. This is an extreme funnel flow condition known as rat holing or piping (Fig. 2b). Mechanical and structural arches are stable obstructions formed when particles come together to form a weight bearing arch, much in the same way a stone bridge is formed. This type of arch can be avoided by making sure the opening is of sufficient width to allow the particles to pass through. For cones the rule of thumb is that the opening should be six to eight times the largest particle size (39), 10 times has been mentioned by Jenike (38). The more troublesome arch is a cohesive arch. This type of arch is formed through compaction and cohesion of the material in the hopper. Bates (40) defines cohesive arching as "the formation of a stable flow obstruction over an outlet or within a flow channel due to the bulk strength of the material exceeding the unconfined failure strength at which the span of the arch would collapse because of the stresses acting on the mass." The design of a mass-flow hopper is based on the postulate that gravity flow of a solid in a channel takes place, provided that the yield strength that the solid develops as a result of consolidation is insufficient to support an obstruction to flow. The calculations for hopper dimension are based on an assumption of the stress pattern that may exist in the hopper in the static condition that is expressed as a flow factor (ff), which is the ratio of the maximum principal stress at two difference points, as shown by

$$ff = \frac{\text{Maximum principal stress encountered within the hopper}}{\text{Maximum principal stress at free surface}} \quad (2)$$

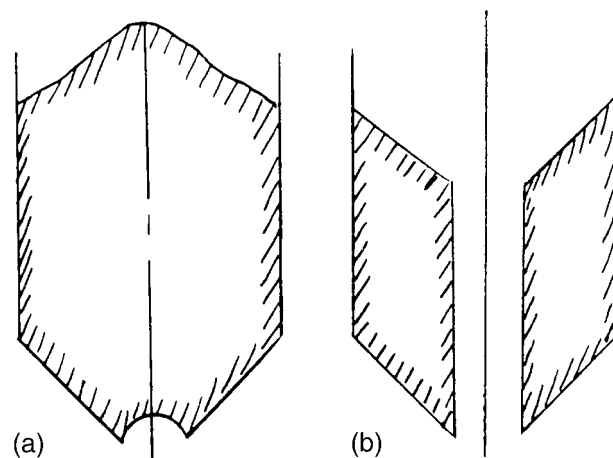


Fig. 2 Flow patterns: (a) doming; (b) piping or rat holing in obstructed flow.

The hopper design, therefore, requires the maximum principal stress at the outlet to be greater than the unconfined yield stress of the powder material. The two stresses may be compared graphically by plotting the flow function of the material (FF) and the flow factor of the hopper (ff) on the same fc vs. σ_m plot. The point at which the hopper stress and the material strength are equal is known as the “critical unconfined yield stress;” it must be determined when calculating the size of the hopper outlet. The hopper flow factor is obtained from figures described by Jenike (38) and come from shear cell measurements described later on in this chapter.

If the hopper-flow factor lies above the powder-flow function, cohesive arching is not possible with the material in that hopper. If the hopper factor lies below the powder-flow function, arching of material in the hopper is possible even for large-outlet sizes.

Various shear cells have been developed to aid in obtaining the needed information to determine the functions necessary to design mass flow hoppers, bins and silos. In addition to powder shear, wall friction testing (41, 42) and compressibility (43) as well as other testing all enter into hopper design. This chapter can only briefly touch on the theories and test methods that go into hopper design. Various recent texts (44–46) give a much more in depth treatment of the subject. Various web sites such as www.powderandbulk.com have been the source of updates in this area.

Prescott and Hossfeld (47) have written an article relating hopper design directly to pharmaceuticals, most specifically to the tabletting process. Mechanisms of segregation in the hopper are discussed and case histories of how hopper designs and inserts, which aid flow, can help solve these problems. Since the hoppers for tablet presses as well as capsule filling equipment are specifically designed for this equipment and since in the pharmaceutical industry a variety of materials must be handled by the same equipment, the use of inserts might be a relatively inexpensive fix for flow problems. Bates (48) discusses the use of inserts in hoppers and Troxel (49) also discusses the use of mechanical flow aids such as vibrators, air cannons as well as chemical flow aid, which are discussed in the next section.

FLOW ENHANCERS (GLIDANTS)

To improve the flowability of powders and granulations, a small amount of a flow agent, or glidant, is often added, usually in powder form. Commonly used glidants are colloidal silicone dioxide, talc, and starch. Several

postulates have been proposed for the mechanism of glidant action (50–52).

- Dispersion of static charge from the surface of host particles;
- Distribution of glidant in the host particles;
- Preferential adsorption of gases and vapors otherwise adsorbed onto the host particle;
- Physical separation of particles and subsequent reduction in van der Waals interaction; and
- Adsorption of glidant particles to granulation surfaces in such a way that friction between particles and surface rugosity are minimized.

The effects of a glidant on the flowability of a powder depend on many factors, such as physical and chemical affinity for the powder, the average particle size and shape in relation to those of the powder, concentration of the glidant and degree of mixing, as well as the moisture content. To be effective, in general, the glidant particles should be very much smaller than those of the host powder in order to coat them completely, smoothing out irregularities in their shape, and reducing the frictional and adhesive forces that operate between them. In almost all systems, there is an optimum concentration above which the glidant ceases to be effective. If too much is added, powder flowability may decrease, and it is therefore necessary to control the addition carefully for the best results. Glidants probably have mainly a mechanical action. They adhere to the surfaces of host powders, smoothing out irregularities and reducing their tendency to interlock mechanically during movement and flow. If the particles of the glidants are assumed to be spherical in shape with a radius r and are uniformly closely packed to those of the host, which are also assumed to be spherical with the radius R , optimum flow should occur when the mixture contains approximately $(2\pi r(R+r)^2/100/R^3)\sqrt{3}\%w/w$ of a glidant (53). A 100- μm powder requires about 3% of a 1- μm glidant, which is in reasonable agreement with experimental observations. The type of glidant, the mixing factors, and the nature of the formulation all will have affects on the reliability of this equation.

Colloidal Silcion dioxide is an agent that has been commonly used as a glidant. Flow rate, density, as well as other related measurements, were conducted on four different excipients. The addition of 1% colloidal silicon dioxide increased the flow of the poorer flowing excipients but had lesser affects on the exipients that were of larger size and were better flowing (54). Increasing amounts of chlorpromazine hydrochloride markedly decreased the flow characteristics of all the various blends. The particle size of the excipient blend seemed to grow larger with the addition of the chlorpromazine hydrochloride since it is

believed that the drug stuck to the surface of the excipients. The addition of 1% colloidal silicon dioxide negated the negative affects of the drug. The addition of the colloidal silicon dioxide decreased the particle size increases caused by the addition of drug. It is believed therefore that a possible mechanism of the silicon dioxide was to prevent the drug from coating the excipient particles and reducing flow. Scanning electron microscopy corroborated this conclusion. In another work (55) the amount of silicon dioxide was varied from 0–2%. It was shown that 0.5% addition of the glidant caused a marked reduction to the weight deviation of tablets produced on a single punch tablet press. The standard deviation of tablet weight did not further reduce when greater than 0.5% of colloidal silicon dioxide was added to the formulation showing a limit to its effectiveness. In a recent article (56), 2% colloidal silicon dioxide was shown to be affective in improving the flow of micronized ibuprofen powder in powder layering.

Magnesium stearate, which is better know as a lubricant, has also been shown to have glidant properties. Shear cell studies with eight different materials with varying size and shape showed that the optimum amount of magnesium stearate that affected flow was related directly to particle size (57). In another study mixing time and type of mixer had a marked affect on the flow of lactose containing 1% magnesium stearate (58). These experiments demonstrate that magnesium stearate becomes increasingly affective as a glidant as it coats the material. The surface area needed to be covered and the shear and duration of the mixing are primary variables in determining the optimal glidant concentration of magnesium stearate.

MEASUREMENTS OF FLOW PROPERTIES

Flow Through an Orifice

During the manufacture of tablets, granulations must first flow through a stationary orifice, the outlet of a hopper, followed by flowing into moving orifices, that is, tablet dies. The first type of flow, static flow, has been investigated extensively. Equations developed for modeling static flow, W_s , through an orifice of diameter P , are often based on the Brown–Richards equation and are written as

$$W_s = f(d)P^{b(d)} \quad (3)$$

Generally f and n denote “function of” and $n(d)$ is generally 2.5. Moreover n has also been shown to be a function of the particle diameter d . Conditions for static

flow where the Brown–Richards equation is applicable are the following (59):

- The flow rate decreases as the particle size is increased;
- The orifice size is at least six times greater than the particle size;
- The granulation height is at least two times greater than the orifice diameter; and
- The ratio of the orifice diameter to granulation hopper diameter is less than 0.5.

The dimensions of a flowmeter should therefore be carefully considered prior to its construction. For example, a flat-bottom flowmeter with a diameter of 6.0 cm, and circular orifices with diameters of 1.428, 0.925, and 0.635 cm and a column height exceeding 18 cm were employed as reported by Danish and Parrott (11). For a flowmeter consisting of a hopper with interchangeable attachments, a top-loading balance with output capabilities, and a digital analog converter, a strip-chart recorder was employed in the evaluation of formulations (60). Criteria for acceptance of prototype formulations were developed, including uniform flow of >200 g/min through a 30° cone, and low through a 12.7-mm straight-bore attachment of $7 > 2000$ g/min.

Some formulations can flow too well and upper limits are required based on the capacity of the tablet press under consideration.

A flowmeter is generally capable of measuring the flow rate of powders that flow well. For cohesive materials the minimum orifice diameter necessary to induce free flow was reported to be a better index of flowability (61).

Several studies were conducted to investigate the relationship between the orifice flow rate and the failure properties of powder materials. It was indicated that a simple linear relationship between flow rate and tensile strength would exist only when the powder system is sufficiently cohesive to produce measurable tensile strength of a consolidated powder but sufficiently free-flowing to produce a gravity flow under an unknown and changing consolidation state (62). For less easily flowing powders the hopper may be vibrated to facilitate flow in order to measure flow rates.

A flow device was developed with movable orifices that attempted, through the geometry of the system, to eliminate wall friction as a factor (63). Most consistent results were obtained when a stainless steel cylinder was used as the main hopper. The orifice was a movable end piece with different size holes. In order to select the best excipients for the direct compression of microtablets a funnel flow device, using interchangeable orifices was used. The lengths of the orifices were also varied to keep the angle of inclination constant (64). It was shown that

flow rates could be estimated even for narrow orifices. The Carr index and Hausner ratios, which will be discussed in the next section, of the excipients was also determined and used in the excipient evaluation. In a more recent paper (65) a funnel method was used to evaluate a number of sustained release theophylline microtablet formulations. Colloidal silicon dioxide was added to the nonflowing granulations and then measurements were taken. The angle of repose was also measured for all of these formulations. A number of papers in the last 10 years, in addition to the ones already discussed, use flow rate measurements to evaluate pharmaceutical formulations and excipients (66–73) and almost all combine the hopper flow measurements with other measurements, such as angle or repose and Carr Index. In summary, flowmeters are a practical way of screening powders with good flow and distinguishing good flowing materials from poorer flowing materials. Since all materials don't flow well enough for evaluation, flowmeter measurements are often used in combination with other measurements for formulation flow evaluation.

Empirical Measurements of Powder or Granular Properties

A number of tests such as contact angle, bulk density, and tap density are relatively simple tests that, though not direct measurements of flow, have been found often to be predictive of the flow characteristics of materials. This is because these measurements are highly dependent on particle size and shape and the cohesive nature of the material, all of which are chief variables in the flowability of solids. These tests are commonly used in combination and are still routinely being conducted and related to flow in every-day pharmaceutical development.

Angle of Repose and Other Handling Angles

If a powder is allowed to flow onto a flat surface, a pile or heap of powder is formed. A material that is not cohesive and flows well, spreads out, forming a low heap. More cohesive materials form higher heaps, which are less spread out. The angle of repose is defined as the angle of the free surface of a pile of powder to the horizontal plane. Pharmaceutical powders give two main types of angle of repose: the poured angle, which is the angle measured on a pile poured freely onto a flat surface; and the drained angle, which is the angle measured on the conical surface of powder in a flat-bottomed container, if the powder is discharged through an orifice in the base. The poured angle of repose can be determined by any

device that allows a standard method of pouring the powder onto the flat surface and measuring the angle of the resulting heap. A protractor is commonly used or the base of the heap is controlled and the height of the heap is measured (Fig. 3) (74).

The drained repose angle is obtained with flow devices as described previously. The measurement is affected by the degree of consolidation of the material in the hopper. The drained angle is usually larger than the poured angle for the same powder. Just as the flow rate is affected by the size of the orifice and the particle size of the powder, so is the drained angle. Pilpel (75), recognizing the effect wall friction could have on the drained angle, reported on a device in which the powder is initially placed in a large container with a built-in platform. The powder is allowed to flow out at the bottom, leaving an undisturbed conical heap on the platform, thus eliminating wall friction from the measurement. Carstensen et al. (76) demonstrated that the frictional force between the support base and the granulation plays a significant part in the repose angle of a heap of granules on a plane surface. This means that even the Pilpel device as well as poured-angle measurements would be affected by frictional factors. Jones and Pilpel (77) tested a number of devices and showed that the measurements obtained were highly dependent on the device. In a review article (78), Pilpel states that since the angular properties of powders are so dependent on the details of measurements, they are useful only in a qualitative manner. For this reason, the shear cell has replaced the repose angle measurement as a quantitative method to detect the cohesive characteristics of powder.

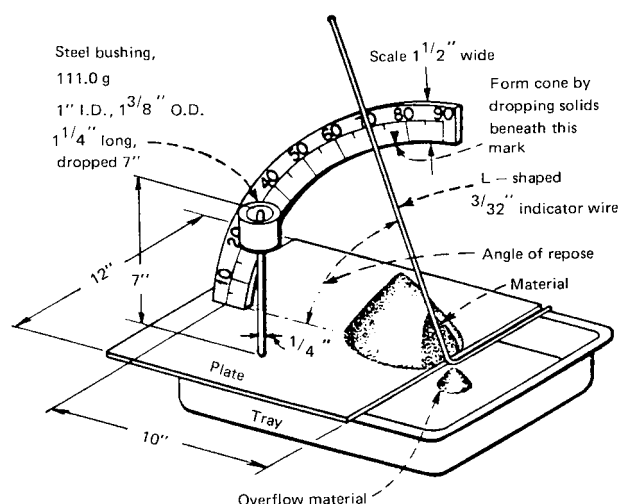


Fig. 3 The angle-of-repose plate incorporates a protractor, an indicator wire, and a jarring device for angle-of-fall tests. (From Chem. Engi. 1965, 18, 164.)

In addition to measurements of poured and drained repose angles, several other handling angles have been utilized. By placing the powder on a smooth surface (slide) and finding the minimum slope that causes the powder to slide, the angle of the slide is obtained. It is highly dependent on the material over which it slides (79), and can be useful in designing hoppers or conveyers. Another angle that has been utilized is the angle of spatula, which is measured by lifting powder with a flat spatula and measuring the angle the heap of powder forms with the spatula surface. Carr (74) standardized this method and utilized it with several other parameters to characterize powders. Characterization numbers are discussed later. The angle of spatula serves as a very simple but rough method of characterizing a powder with respect to flow. A vibrating spatula has also been used in recent years and this will also be discussed in a later section (20, 80).

Dahlinder et al. (61) developed a new device to measure the drained angle of repose. It consists of a split cylinder, which contains a flat circular platform. To measure the angle of repose, the cylinder is filled with powder. The cylinder is split, allowing the powder to fall into a heap on the platform. This device is similar to Pilpel's (75) but eliminates wall friction during the flow by utilizing the split cylinder to initiate flow. Dahlinder tested sodium chloride, microcrystalline cellulose, and aspirin powder as well as lactose, lactose-cornstarch, and aluminum hydroxide-magnesium carbonate granules. The resulting angle-of-repose measurements were reproducible (relative standard deviation about 2%). The authors concluded that this method could be used to characterize even fairly cohesive materials. Nyqvist and Nicklasson (81) utilizing Dahlinder's repose-angle device tested direct-compression lactose containing small concentrations of various actives and found a linear relationship of the angle of repose to the coefficient of tablet-weight variation. Nyqvist (82) found that when testing penicillin granulations the Dahlinder repose-angle results did not give any specific ranking order for the pure drug substance or the granulation. In contrast, annular shear cell data were able to predict the flow behavior for tablet production. Most recently Parrott (83) utilized a poured angle measurement to compare Soludex, a corn-based maltodextrin, to nine commercial excipients. He also used flow through a fixed orifice, and found that the flowmeter represented the flow of the materials better than the repose angle. However, several ingredients did not flow at all through the orifice, and the angle of repose represented the only comparative value for the 10 excipients investigated.

Heistand (8) and Carstensen (84) related the repose angle to the intrinsic cohesion and frictional coefficient of

powders. As mentioned previously, the dependence on the technique used makes this method much less reliable for this purpose than a shear cell. In addition, the repose angle does not quantitatively treat the compressive forces of the flow process in the same way as the shear cell. As Pilpel mentioned (79), angular tests are applicable to relatively free-flowing powders containing particles larger than 100 μm . Such powders cannot be investigated in the scientifically more satisfactory shear cell and tensile strength test apparatus because of their low cohesion and tensile strength. Furthermore, the particles become crushed on consolidation. Today angular measurements continue to be used with mixed results, as shown by the previous examples. The test, usually in combination with other tests, provides a simple and sometimes useful method for monitoring the characteristics of powders and granules for quality control purposes.

Packing Properties and Bulk Densities

The bulk density of a powder is obtained by dividing its mass by the bulk volume it occupies. The volume includes the spaces between particles as well as the envelope volumes of the particles themselves. The true density of a material (i.e., the density of the actual solid material) can be obtained with a gas pycnometer. The bulk density of a powder is not a definite number like true density or specific gravity but an indirect measurement of a number of factors, including particle size and size distribution, particle shape, true density, and especially the method of measurement. Although there is no direct linear relationship between the flowability of a powder and its bulk density, the latter is extremely important in determining the capacity of mixers and hoppers and providing an easily obtained valuable characterization of powders.

Aerated bulk density

This is the bulk density of a powder after it has been allowed to aerate, that is, in most cases flow. A number of commercial devices are used to obtain this value. These devices allow the powder to flow from a fixed height, usually through a set of screens, into a container of fixed volume. An excess of powder is used and cleared from the top of the container, the tared container is then weighed. The bulk density is the weight of the powder divided by the volume of the container. The device can be as complex as a series of vibrating screens or as simple as pouring powder through a screen into a tared graduated cylinder. The cylinder is placed on a tapping device to measure the tap or, as Carr (74) defines it, the packed bulk density. Some tap the cylinder three times in order to level the

powder and get a more uniform result. The leveling of the powder can also be critical in other devices. The author utilized a device known as a Scott volumeter in which powder is passed through a series of glass baffles before falling into a tared cubical container. The excess powder is removed, so that the powder is flush with the container top. Considerable operator variability between tests has been noted, and it was found that the pressure in removing the excess powder varied between operators; this was the chief source of variability. A standard wiper method was developed and solved the problem.

Tapped or compressed bulk density

Tap density is the bulk density of a powder which has been compacted by tapping or vibration following a specified procedure. A large number of machines are available to measure tap density; some use a fixed volume of powder and some a fixed weight of powder. The sample is dropped (tapped) a set distance at a set frequency for a fixed number of times. Vibrators of known frequency have also been used for compaction. A simple device, developed by Neumann in 1953 (85) is shown in Fig. 4. It uses a cut cam, which allows the powder to drop exactly 1 cm each revolution. In tap devices the machine can be set to stop after a certain number of taps and the volume is measured. The sample is usually tapped or vibrated until an equilibrium volume is obtained and at that point the final tap density is determined. The rate of tap density can also be measured by recording the density after a given number of taps or vibrating for a given amount of time, and repeating the process until no change in volume occurs.

Density and flow

Carr (74) reported that the more a material is compacted in a compaction or tap bulk density test, the poorer its flow properties. He defines compressibility C by

$$\%C = 100(P - A)/P \quad (4)$$

where P is the final tap or compacted bulk density and A the aerated bulk density. A more commonly used term is the Hausner ratio, which is simply P/A , the final tap density divided by the aerated bulk density. This ratio was introduced by Hausner in 1967 (86) to characterize metal powders, but is commonly used today for pharmaceutical powders. The higher the Hausner ratio, the poorer the flow. Neumann (52) reported that the higher the rate of packing (fewer taps to reach equilibrium), the better the flowability. He also experimented with weights to compact the powder. He plotted the log of the relative volume obtained with a given weight vs. the log of the weight used; linear plots are produced. Chowhan and Chow (87)

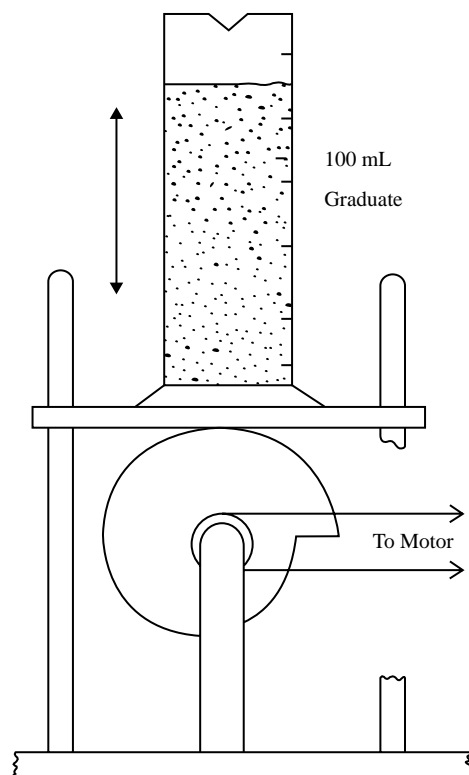


Fig. 4 Tapping device. (From Ref. 85.)

used a modified version of Neumann's device in which they placed loosely packed powders and powder mixtures in cylindrical containers and applied a series of loads to the surface of the powder beds. They tested a drug with various excipients and plotted the log of the relative volume vs. the log of the relative weight utilized. They also obtained linear plots and called the intercept on the relative volume axis the powder-consolidating ratio. This value yielded a linear correlation with weight variation for capsules filled on a Zanasi capsule-filling machine.

Varthalis and Pilpel (88) derived a relationship between tensile strength and packing rate with lactose and paracetamol and, alternatively, lactose and oxytetracycline. From this relationship they developed a measure of powder flow and packing in terms of the "angle of internal flow," an empirically derived parameter, for the rate of change of bulk density with tamping. Newton and Bader (13) utilized this parameter and related it to the fill weight of capsules containing aspirin and lactose. Attempts to relate it to in vitro release of the aspirin were not successful. In the first of a series of articles regarding lactose coated with nonionic surfactant, Sakr and Pilpel (89) utilized tap density to explain the surface effects of the coating. Yamashiro et al. (90) used very short

tap intervals (stopped and measured after one or two taps) to test an equation stating that the number of taps divided by the relative volume of the powder after these taps is linearly related to the number of taps. When this relationship holds, the flow rate of a material can be related to the proportionality constants.

In conclusion, tap-density methods are a quick and inexpensive way to characterize materials. A strictly empirical relationship exists between the degree of compaction and powder flow. If viewed in this light, tap density can be a useful measurement technique.

CLASSIFICATION SYSTEMS

In 1965 Carr (74) proposed a characterization system that has been used extensively to classify pharmaceutical powders and granulations. The method consists of a point system that weighs four factors equally, giving each a maximum of 25 points with a maximum of 100 for a perfectly flowing material. The first measurement is the angle of repose. A diagram of the apparatus used by Carr and a discussion of angle of repose appear earlier (Fig. 3). According to Carr, the contact angle is a simple and easy method of indirectly measuring the following properties affecting flow: shape, size, porosity, cohesion, fluidity, surface area, and bulk density. The second factor, also discussed previously, is compressibility as obtained from tap density. Carr defined compressibility by Eq. 4 and claimed that the percent compressibility indirectly provides an excellent representation of uniformity in size and shape, deformability, surface area, cohesion, and moisture content.

The third factor in Carr's characterization scheme is the angle of spatula. As mentioned previously, Carr designed a special method for measuring this angle. According to him, the angle of spatula is an indirect measure of cohesion, surface area, size, shape, uniformity, fluidity, porosity, and deformability.

The last factor is either the cohesion or the coefficient of uniformity. Cohesion is used with powders (very fine particles) or with materials on which an effective cohesion force can be measured. The uniformity coefficient is used with granular and powdered granular materials on which an effective surface cohesion cannot be measured.

The procedure for finding the apparent surface cohesion involves determining the retention of material on a nest of 250-, 150-, and 74- μm (60-, 100-, and 200-mesh) screens over a bottom pan. A weight of 2 g of test powder is recommended to be placed on the 250- μm (60-mesh) screen, followed by vibrating the nest of screens for

20–120 s. The amount of material left on each screen is weighed and rated in points or percentage. Accordingly: each 0.1 g on the 250- μm (60-mesh) screen corresponds to five points or 5%; each 0.1 g on the 150- μm (100-mesh) screen corresponds to three points; and each 0.1 g on the 74- μm (200-mesh) screen corresponds to one point. If the entire amount of material passes through the 74- μm (200-mesh) screen, the cohesion is zero. Electrostatic attraction of some particles and the tendency of smaller particles to adhere to larger particles can skew the result. The particle must be 74- μm (200-mesh) material in order for the test to be attempted.

The uniformity coefficient is arrived at by dividing the width of the sieve opening that passes 10% of the sample. It is determined from a screen analysis of the material. The more uniform a mass of particles is in both size and shape, the more flowable it is likely to be.

The point score for the evaluation of flowability of dry solids is given in Table 1. In order to measure the flowability of a powder utilizing this method, each test is made and the points can be found in the table. The total number of points would place the sample into one of the seven categories in the left column of the table. Because of its simplicity and the fact that commercial devices are available to conduct these tests (this test is ASTM test designation D 6393-99), this system is widely used in the pharmaceutical industry both in preformulations, formulation research, and in quality control.

Other classification systems are used less frequently. Carr (74) also devised a system to classify materials as to their floodability. He defines the floodability of a material as its tendency to flow like a liquid because of the natural fluidization of a mass of particles by air. In order to so classify a material, the flowability is determined utilizing the method just described. This value is equivalent to a measurement Carr calls the angle of fall, angle of difference, and dispersibility. Though referred to in any of the papers mentioned here, this system is much less utilized than the flowability measurements. Geldart (91) reported on a characterization system of powders according to their ability to aerate and later Molerus (92) modified this system. In a more recent symposium this method of powder classification was examined (93–95).

As mentioned in previous sections of this article, the Carr index as well as other empirical tests such as the Hausner ratio continue to be used, often in combination with other tests to characterize and predict the flow of pharmaceutical excipients and formulations. In addition to papers already discussed in this article a number of papers have appeared in the more recent pharmaceutical literature in which these methods have been used to classify the flow characteristics of pharmaceutical

Table 1 Point scores for evaluation of flowability of dry solids

Flowability and performance	Angle of repose		Compressibility		Angle of spatula		Uniformity coef. ^a		Cohesion ^b	
	Deg.	Points	%	Points	Deg.	Points	Units	Points	%	Points
Excellent, 90–100 pts.	25	25	5	25	25	25	1	25		
Aid not needed	26–29	24	6–9	23	26–30	24	2–4	23		
Does not arch	30	22.5	10	22.5	31	22.5	5	22.5		
Good, 80–89 pts.	31	22	11	22	32	22	6	22		
Aid not needed	32–34	21	12–14	21	33–37	21	7	21		
Does not arch	35	20	15	20	38	20	8	20		
Fair, 70–79 pts.	36	19.5	16	19.5	39	19.5	9	19		
Aid not needed, but vibration may be necessary	37–39	18	17–19	18	40–44	18	10–11	18		
	40	17.5	20	17.5	45	17.5	12	17.5		
Passable, 60–69 pts.	41	17	21	17	46	17	13	17		
Borderline, material may hang up	42–44	16	22–24	16	47–59	16	15–16	16		
	45	15	25	15	60	15	17	15	<6	15
Poor, 40–59 pts.	46	14.5	26	14.5	61	14.5	18	14.5	6–9	14.5
Agitation and vibration required	47–54	12	27–30	12	62–74	12	19–21	12	10–29	12
	55	10	31	10	75	10	22	10	30	10
Very poor, 20–39 pts.	56	9.5	32	9.5	76	9.5	23	9.5	31	9.5
More positive agitation needed	57–64	7	33–36	7	77–89	7	24–26	7	32–54	7
	65	5	37	5	90	5	27	5	55	5
Very, very poor, up to 19 pts.	66	4.5	38	4.5	91	4.5	28	4.5	56	4.5
Special agitation and hopper design required	67–89	2	39–45	2	92–99	2	29–35	2	51–79	2
	90	0	>45	0	>99	0	>36	0	>79	0

^a Used with granular and powdered granular materials.^b Used with powders or where an effective cohesion can be measured.(From Chem. Eng. **1965**, 18.)

powders (93–101). Sugar based excipients (96), granulated lactitol (101), microcrystalline cellulose codried with β -cyclodextrins (100) and various phyllosilicate powders such as bentonites (97) were characterized through Carr index measurements. With the phyllosilicate powders the authors noted that hopper flow values were of little use due to very high standard deviations and the angle of repose measurements and Carr index and Hausner Ratio were considered more relevant to their characterization. In the microwave drying of high shear granulations, the Carr index decreased with increasing microwave power due possibly to an increase in dust formation (98). Hausner ratio and slowing of the rate of taping was used to show that moisture increased the interparticular friction of Avicel PH 302 (99). Angle of repose along with flow through multi size orifices were used to characterize the flow of several dextrins (93). Another example of the use of Carr's indexing method is given in a paper by Vennat

et al. (102) where the flowability and floodability index of a group of direct-compression excipients and procyanidins are reported. The material was tested in the commercial Hosokawa powder characteristic tester. Only angle of repose and taped and bulk densities were used to classify the rheology of Avicel PH-200 and Cellactose (94, 95).

In summary, Carr's flow characterization method is frequently used for characterizing powders. No one method alone gives a total picture of flow and hence a combination of methods is often used. If known forces are being utilized during the flow process, a shear cell would be the method of choice when a scientifically based characterization is needed. If granulations are being tested with good flow characteristics, a flowmeter might be the method of choice. A combination of empirical tests, for example, Carr's flowability test, would be very useful as a simple screening device or as a quality control test for multiple lots. The selection of the method should be based

on the amount of information needed and the material to be tested, and may be more importantly, the equipment available.

SHEAR CELLS

The flow through an orifice is an excellent method to compare materials with good flow. Pharmacy, however, most often deals with cohesive powders that do not readily flow. The flow of these materials is increased by employing force feeders, resulting in a dynamic process. Angle-of-repose measurements that do not utilize force and that are often a measurement of stagnant powder have therefore limited application. Shear testers, that measure the frictional characteristics of a powder bed under load, yield valuable information with regard to powder flow in high-speed tablet- and capsule-filling equipment.

Data Treatment

A number of different types of shear cells are discussed in this section. Each of these cells is designed to condition the powder under a known force, and measure the force needed to shear the powder bed under a load force equal to or less than the original conditioning force. For each conditioning force, the force needed to shear the bed for a series of loads is plotted against the load force in a similar manner in which a frictional coefficient is determined for a nonpowdered material. Such a plot is known as a yield locus (Fig. 5), and it is often obtained for a number of different conditioning forces. Thus, if the forces acting on a material in a given process can be approximated, intrinsic information regarding the frictional and cohesive nature of the material can be obtained that should have great relevance to real processes.

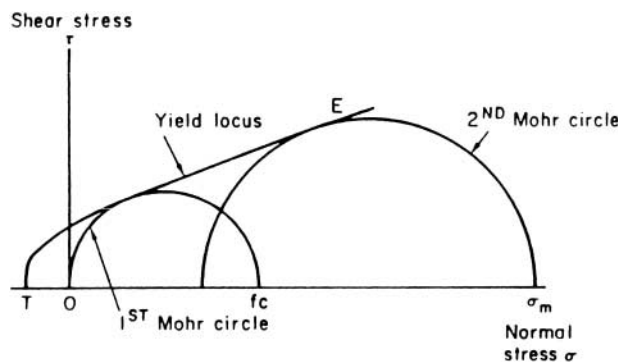


Fig. 5 Mohr circle construction to obtain the major normal stress, σ_m , and the unconfined field stress, f_c . (From Ref. 78.)

The most common method of treating shear data is the Mohr circle analysis, also shown in Fig. 5. The x -axis represents the stress normal to the shearing power (load stress), and the y -axis represents the shear stress. The curve represents the yield locus for a cohesive solid. The intersections of the Mohr circles are drawn in such a way that they are tangent to the yield locus. These circles represent the total forces on the powder bed at the point of shear for any direction. The point at which the Mohr-circle-drawn tangent to the upper end of the yield locus (2nd Mohr circle) intersects the x -axis at the highest point is known as the principal normal stress σ_m . This value represents the maximum normal stress under which the powder was consolidated before it yielded and changed volume. Drawing a Mohr circle that is tangent to the yield locus and passes through the origin gives a value known as the unconfined yield locus f_c (the point where the Mohr circle intersects with x -axis). This value represents the maximum principal stress acting on a free surface necessary to cause failure. Jenike (38) related this value to the strength of an arch in a hopper. For a cohesive solid, the intersection of the yield locus with the y -axis (shear stress) is considered the cohesion of the powder for that yield locus. The point at which the yield locus intersects the negative x -axis is considered the tensile strength of the material compressed under the normal load used to generate the locus. Most often a series of conditioning forces is used, giving a series or "family" of yield loci (Fig. 6a). A plot of the unconfined yield locus f_c vs. the principal consolidating force for a series of yield loci has been used by Jenike (38) as a flow (or failure) function. Each of these values has been used to classify material as to their flowability.

Eq. 5 describes the yield locus. As early as 1965, Eq. 5 was established empirically from a study of the shapes of yield loci for more than 30 powders (103):

$$\ln[\tau/C] = (1/\eta) \ln[(\sigma + T)/T] \quad (5)$$

where τ is the shear stress, σ the normal stress, T the tensile strength, C the cohesion, and η the shear index. It is often referred to as the Warren–Springs equation and has been confirmed by numerous authors (104–106). Earlier in this article it was mentioned that Geoffroy and Carstensen (14) modified this equation taking particle size of the material into affect. Hiestand (106) states that the Warren–Springs equation is suitable for describing the yield locus for failure in shear. However, contrary to common usage, the term T is not the tensile strength but the internal cohesion, the magnitude of which may be much less than the tensile strength. As stated by Hiestand, this is contrary to the common usage of the equation. He based his conclusion primarily on experimental data with sitosterols. In contrast,

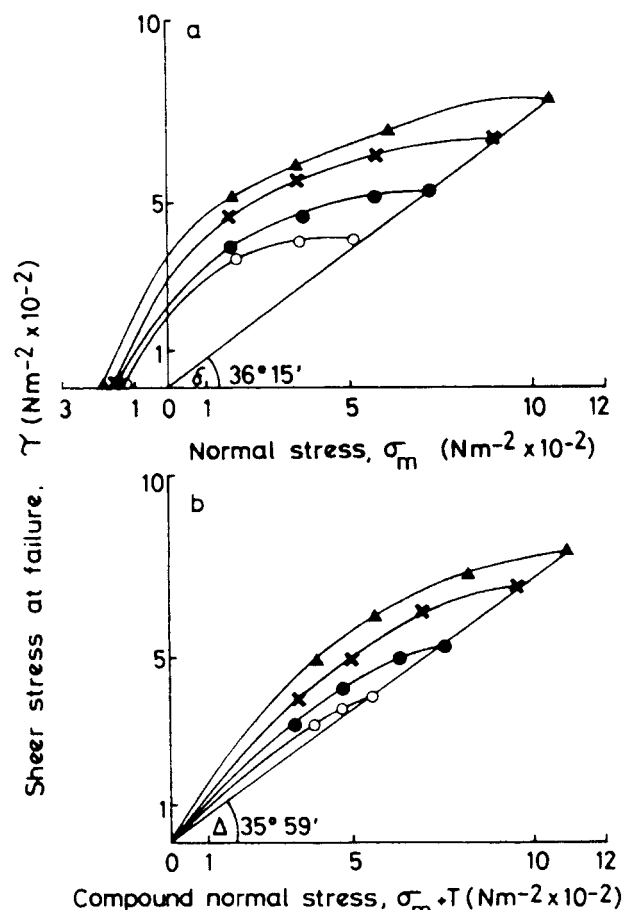


Fig. 6 Families of yield loci for lactose powder coated with $0.5 \times 10^{-5} \text{ mol g}^{-1}$ of light paraffin at 20°C . $P_f \blacktriangle$, 0.346; \times , 0.282; \bullet , 0.279; \circ , 0.261; (a) normal stress; (b) compound stress. (From Ref. 107.)

Irono and Pilpel (107) modified the treatment of the shear data (Fig. 6b). In this figure they added tensile strength values obtained independently in the tensile-strength measuring to the normal load in the shear cell loci plots.

The actual method used to treat the data varies with respect to the nature of the shear cell utilized. Direct measurements of the various functions such as tensile strength and unconfined yield strength have been taken. The following section describes the methods being utilized and their advantages and disadvantages and how they have been applied to the flow of pharmaceutical powders. A detailed description of data evaluation is not included. Several excellent reviews of shear and tensile measurement are available and can be found in the references. Johanson, who teamed with Jenike in conducting much important work in the field, has published an interesting historical perspective of the field (108).

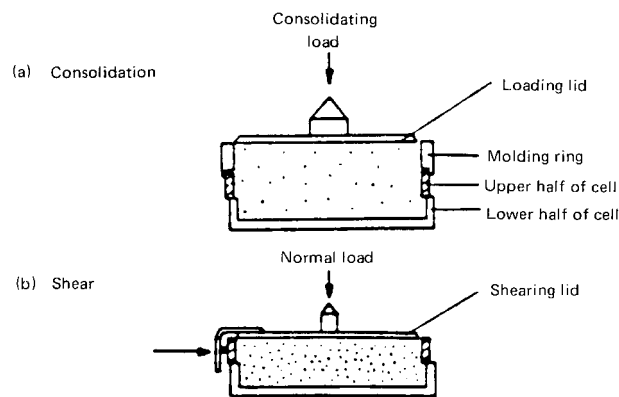


Fig. 7 The Jenike shear cell. (From York, P. *Int. J. Pharm.* 1980, 6, 101.)

Jenike Shear Cell

This instrument was developed first for soil testing and adapted by Jenike to measure cohesive powders and relate the data to hopper flow and hence hopper design. This is essentially the standard cell (Fig. 7) and the basis of much of the data solids-handling work. It is split horizontally; the lower half is fixed and the upper half is moveable at a constant low rate. The cells are first filled with sample and then consolidated by placing a load on the loading lid (Fig. 7a) and rotating the lid backward and forward through a slight angle. The loading lid and molding ring are removed and the powder scrapped level with the top of the upper half of the cell. The shearing lid is placed in position and the cell is sheared by moving the upper half over the lower fixed portion. The load is equal to or lower than the consolidated load. The process is repeated with the same consolidating load but with lower shearing loads until the yield locus is obtained. If a family series of yield loci are to be developed, a series of consolidating forces are used. A flow-factor plot should be based on at least three consolidation pressures. Needless to say, this process is extremely time-consuming and tedious. Another disadvantage is that correct consolidation of the sample is difficult to achieve. For this reason the results may be very operator-dependent. Correct consolidation is generally established by preliminary experiments, which examine the shape of the stress-strain curve obtained during shearing. Despite these difficulties, the Jenike shear cell has been used extensively and is commercially available. A number of papers have appeared in the pharmaceutical literature describing utilization of this device (103–116). Doelker (109) in comparing various microcrystalline cellulose types states that techniques such as the Hausner ratio and Carr index are poorly reproducible and can be compared only if determined by

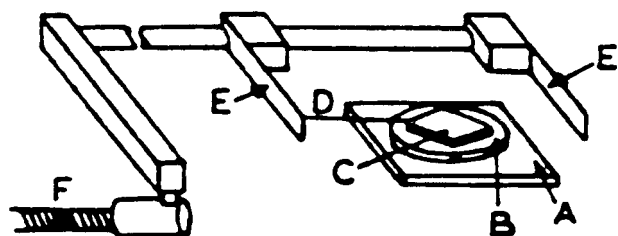


Fig. 8 The plate-type shear cell. Key: (A) lower plate; (B) template; (C) upper plate; (D) tow line; (E) cantilever strain gauge; (F) screw jack. (From York, P. *Int. J. Pharm.* **1980**, *6*, 100.)

the same authors. He states "A more trustworthy method for evaluating the flow characteristics of cohesive powders is to use shear cells." He goes on to review various flow measurements of microcrystalline cellulose as well as other commonly used direct compression excipients.

Plate-Type Shear Cell

Nash et al. (117) and later Hiestand et al. (118,119) developed a very simple test device that can be easily constructed (Fig. 8) and was used to evaluate pharmaceutical powders. The powder is placed between two plates, usually with the surface in contact with the powder consisting of sandpaper or other rough surface to ensure that the powder shears in the bed rather than at the surface of the plate. A consolidating weight is placed on the upper plate. Final consolidation is achieved by shearing, using very short movements in such a way that a reading from the strain gauge is obtained. This is repeated until the shear force, read on a chart recorder, reaches a plateau value. A lower weight is then placed on the upper plate and a lower point in the yield locus is obtained through a single shear movement. The entire process must be repeated for each point in the yield locus.

As with the Jenike cell, this process is time-consuming. Amidon and Houghton (111), however, used a single yield locus with this cell for comparative purposes. Hiestand et al. (120), in comparing this apparatus to the Jenike cell, claimed that this simple shear cell can be used to provide characterization of the unconfined yield strengths of powders. The results from the two devices are not identical. However, as much as the Hiestand device requires less powder and the consolidation step is more automated and consistent, it provides an inexpensive alternative to the Jenike-type cell to characterize pharmaceutical powders. Amidon and Houghton used the cell to examine the effect of moisture

on the powder flow properties of microcrystalline cellulose (27).

Ring or Annular Shear Cell

The ring or annular shear cell, was developed by Carr and Walker (121) as early as 1968. In recent years this tester has undergone a number of modifications. Peschl (122, 123) has developed an annular shear cell in which the sample and shear cell consists of a full circle. This contrasts to the earlier cells that have a band of sample on the outer portion of the circle. This was done to eliminate wall friction. It is also rotated very slowly, since at low speed, velocity variability becomes more negligible in the shear measurement. In this way a full ring can be utilized and speed differences in the outside and inside of the ring become negligible. Schulze (124–126) made the latest modifications to the ring cell tester (Schulze tester RST – 01.01). In the Schulze ring shear tester the sample is placed in an outer circular channel. An angular lid, which is attached to a crossbeam, lies on top of the sample. Small bars are attached to the bottom side of the lid and the bottom of the cell to prevent the powder from sliding against the lid or the bottom. The shear cell is rotated while the lid is prevented from rotating by cross beams attached to fixed beam. The movement of the cell with respect to the fixed lid causes the powder bed to shear and load cells attached to tie rods measure the force needed to initiate this. To exert weight on the sample, weights are hung from a crossbeam. This can be done during the shearing and conditioning of the powder sample. The cell can also be removed and time consolidation can be conducted by placing weights on the sample out side of the test device, in a similar mannner to that with the Jenike cell. An automatic version of this tester (126) has been developed in which the computer can automatically add loading to the sample and condition it. The instrument can be operated in a manual, semiautomatic or totally computer run mode. Both the earlier annular shear cells, the Peschl or Schulze's RST –01.01 cell have several important advantages over the Jenike-type cell. They offer a constant area of shear. Handling is easier and consolidation and shear are quicker, since after the bed is consolidated, a full locus can be generated without reconsolidating after each load. The consolidation process becomes more automated and uniform, eliminating much of the operator variability in the measurement process. Peschl reported a relationship between tablet weight variation and internal friction as measured by his device (122, 123). The Peschl shear tester has been utilized for shear testing and quality control of pharmaceuticals by at least one U.S. pharmaceutical company as well as in Europe. Nyqvist (82, 127) and

Nyqvist and Nicklasson (81) utilized an annual shear cell with direct-compression lactose and various actives as well as high-dose penicillin products and found that the device was an excellent predictor of flow on a tablet press as related to tablet weight variation. From the shear-cell data it was actually possible to predict the frequency of tablet machine adjustment on a rotary tablet press in the production of penicillin products (82). In 1985 Baichwal and Augsburg (128) employed an annual shear cell to quantify the amount of friction between pharmaceutical powders with various lubricants and a smooth metal surface. More recently Podczek and Miah (57) used a commercial version of a Carr annular shear cell to measure the flow factors for unlubricated and lubricated powders. Schulze (125, 126) demonstrated the ability to differentiate the flow function of lactose with and without the presence of 1% active drug. This same laboratory has also produced a ring shear tester to measure wall friction of bulk solids (129). Schulze (130) provides an excellent review of all the previously mentioned shear devices as well as contact angle, and paddle type testers, to be discussed later in this chapter. He also lists several devices not included in this discussion.

Bi- and Triaxial Shear Cells and Other Related Tests

There are several shear tests as well other related tests, which have been used in bulk solids handling and in the design of hoppers and silos but have not been used extensively to study pharmaceuticals. The monoaxial shear tester, Johanson Hang-up Indicizer and compressibility test are examples of test devices that have been used in the bulk handling industries and are described in Schulze's chapter (130). Enstad and Feise (131) further discuss the uniaxial shear tester which tests powder compacted in one direction. The major discussion in their chapter (131) is a discussion of biaxial shear testers. Though not mentioned in the pharmaceutical literature, this test has received a good deal of attention in the recent powder technology literature with the development of a number of new test devices in the 1990s. There are essentially three types of biaxial testers, those being rigid boundary, flexible boundary, and mixed boundary testers. The rigid boundary tester is the one used most for the measurement of mechanical behavior of particulate solids. The sample is brick shaped and the set-up insures that the sample will always retain a rectangular cross-section. In the flexible tester the walls are flexible and made of rubber. Pressure can be exerted via air pressure by inflating chambers within the flexible walls. The mixed boundary

system has only been used at very high pressure with primarily sand. Shearing at constant volume, testing time of loading and varying the direction of loading, and measuring the affects on the stress and strain relationships in the bed are the chief function of this test method. Most of the work with this tester has been with limestone and sand and the results have been used in the design of hoppers. The triaxial shear tester, which was introduced into soil mechanics in the 1920s, has not been used to characterize pharmaceutical powders to any great extent but has been discussed in the powder-handling literature. Luong (132) provides a review of this test method. Kolymbas and Wu (133) gives an extensive discussion of the experimental techniques and the potential errors occurring with the device and how to adjust for them. In a triaxial tester the sample is subjected to pressure in three directions, a pressure chamber usually keeps two of which constant. The cells discussed previously were biaxial with pressure by the normal load and in the direction of shear. In the triaxial shear cell, the sample is placed in a cylindrical rubber membrane and enclosed by rigid end cups. The sample is consolidated by maintaining the same pressure in all three directions, which does not induce shear but volumetric strain. Pressure is usually exerted at the end cups with some sort of a piston that moves in one direction while the pressure chamber maintains the pressure in the other directions. The pressure changes and the volume changes are monitored in all directions. A Mohr-circle treatment of the data is usually made. Certain factors, which are used to evaluate materials, are more directly obtained with this device than with biaxial test devices. The triaxial shear device is still primarily a research tool, and many variations have been constructed. It is less appropriate for flow measurement of pharmaceuticals since it is designed for relatively high pressures.

Johanson (134) developed a very simple device they called the Johanson Indicizer system, or the Johanson Hang-up Indicizer, which they claim can predict material bridging in a hopper. This device was reviewed by Bell et al. (135). In private conversations Johanson stated that they have employed this device with a number of pharmaceuticals. It simply compacts very small amounts of powder into a potential bridge and measures the force to break the bridge and allow the powder to flow. It is not designed to give any detailed information on the powder but only provides a test for powder bridging. This device is commercially available.

Tensile strength can be obtained from shear measurements via the Warren-Springs equation or through extension of the yield locus. The direct measurement of the tensile strength of powders with the help of commercial split-cell devices is common. The powder

is consolidated to a known density, usually through weights. One half of the cell is stationary and the other half is movable on a low-friction surface. The force needed to separate the powder bed at a given density is recorded. In general, a straight-line log-log relationship exists between the tensile strength and the packing fraction (density of sample divided by the true density). Hiestand et al. (106) stated that no method, using loosely packed powder beds, as just described, is highly successful. For this reason, he measured the tensile strength of compacts. Despite his opinion, the measurement of the tensile strength of powder beds continues to be utilized by pharmaceutical scientists. Chowhan and Yang (136) found a linear relationship between tensile strength and the coefficient of variation of the fill weight of capsules. In a separate paper (62) they reported that the flow rate of powder mixtures containing simple glidants, such as corn starch and microcrystalline cellulose, at different concentrations is linearly related to tensile strength. Ho et al. (110) on the other hand, found no relationship between tensile strength and tablet weight variation for a number of direct-compression excipients.

In general tensile strength is not used alone to predict flow. Danjo et al. (25) utilized tensile strength measurement, shear cell measurement, and Carr's flow factor to investigate the effect of particle shape on powder flow.

In summary, various shear and tensile strength measuring devices have been developed, and many are commercially available. Though they are more complex and often more difficult to use than the other methods, they yield data on a more scientific basis, which allows a mechanistic approach to flow problems.

AVALANCHE BEHAVIOR AND POWDER RHEOMETERS

Earlier in this article, in the section on the Affect of Particle Shape, the principle of fractal geometry was introduced. As discussed previously Hickey and Concessio (20) utilized fractal analysis to characterize flow of pharmaceutical powders, using a vibratory spatula. Crowder and Hickey developed a vibratory spatula device that was able to obtain flow data much more rapidly (137). Using this device and fractal geometry, they were able to quantify the differences in the flow properties of sprayed and nonspray dried lactose (80). Kaye et al. also used the vibratory flow of solids to measure avalanches down a shoot (18, 19). Eventually Kaye et al. would use a rotating drum method (138) to study solid flow through avalanching behavior. This device is commercially

available (10) as the Aero-Flow™ automated powder flowability analyzer and is being used in the pharmaceutical industry today.

The theoretical basis for this device can be found in fractal dimensions and deterministic chaos theory. Kaye (139) uses a term known as "fractal dimensions in space" in which he treats fractal data not as a physical structure but as a pattern of events that can be used to describe chaotic systems. He describes deterministic chaos, in the same text, as an interaction of events, which are described by deterministic physics. The interaction of the events become so complex that the system may as well be chaotic. For this reason he has combined fractal analysis with chaos theory to develop an understanding of dynamic particle flow. Hickey and Concessio (140) used this same

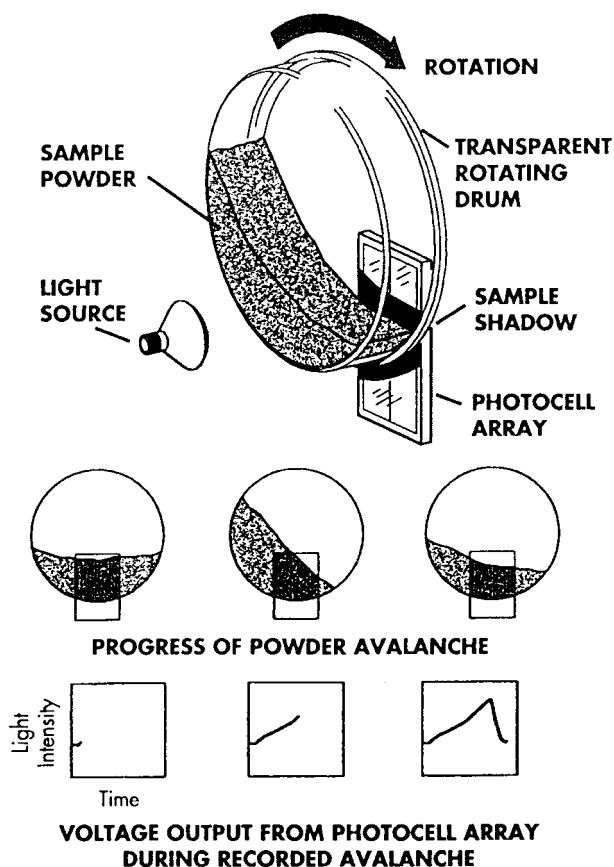


Fig. 9 Descriptive drawing of a rotating drum avalanche test device. This device is commercially available as the TSI Aero-Flow automated powder flowability analyzer. The top pictures show the device. The progress of powder avalanche shows the position of the powder in the drum as the drum is rotated and the lower voltage output shows the output of the device during each drum rotation. (From Ref. 10.)

approach in characterizing lactose, using a rotating drum device. In their device they utilized a video camera to measure the dynamic angle of repose with time. The commercial device developed by Kaye works, using light transmission and a series of photocells to measure the light transmission. The device can be seen in Fig. 9.

As can be seen in Fig. 9 the powder is placed in the drum and the drum is allowed to rotate. The powder lifts up and when the dynamic angle of repose is exceeded, the powder sample will avalanche downward. The movement of the powder blocks the photocells from obtaining light and a voltage output, which can be related to the avalanching of the powder is recorded (Fig. 9). A set of data for lactose can be seen in Fig. 10(141).

The flowability of materials according to the literature is related to the mean time for avalanching to occur. The shorter the time the more free flowing. The scatter is related to cohesiveness (10) or an irregularity factor (143). Lower scatter values would indicate a less cohesive material and would predict more regular flow while greater scatter would be indicative of a more cohesive material and the increased likelihood of irregular flow patterns. The example in Fig. 10 shows a lactose sample that was sieved into two fractions: one above 38 μm and one below 38 μm . The middle data is from unfractionated lactose. As can be seen from Fig. 10 that as expected the finer material

had a larger mean and larger scatter showing it to be a poorer flowing material. The larger fraction showed the shortest mean time to avalanche and the smallest scatter showing that it would be the better flowing material. As expected the unfractionated lactose yielded data in between the two fractions. The flow properties of various lots of lactose monohydrate and paracetamol were studied using the Aero-Flo device and lot to lot variations were noted (142). Trobridge et al. (143) tested five grades of lactose with a variable commercial automated drum tester and found that the mean time to avalanche correlated well with flow performance on a tablet press when the drum speed was 180 s/rotation (spr). When the speed of rotation was set at 200 spr slippage was observed with one of the lactose grades and a different rank order was obtained. It was concluded that the fixed speed unit data identified the two optimal formulations but did not distinguish between marginal flow and no flow, but the variable speed unit did. Crowder and Hickey (144) reviewed the physics of powder flow as related to pharmaceutical solids. Included in the article is a section, which explains many of the terminology of complex systems. They include a discussion of avalanching measurement and the use of chaos theory in measuring flow including rotating drum devices. Included in the discussion is powder and granular mixing as well as milling, neither of which is covered in

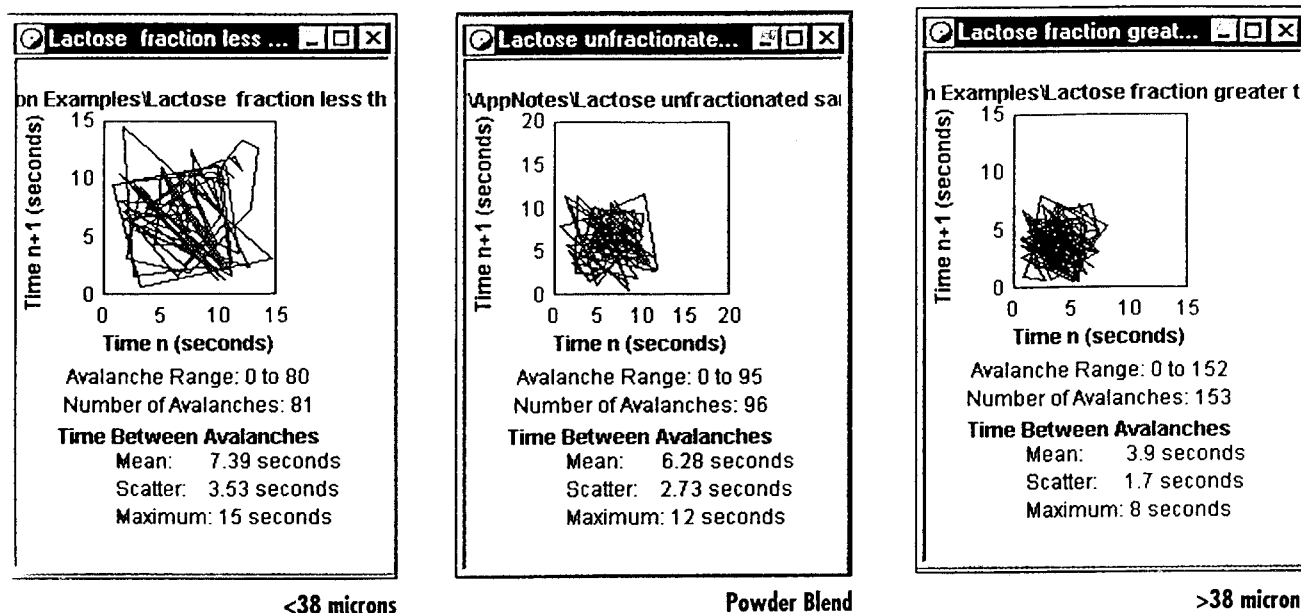


Fig. 10 Rotating drum avalanche data from a TSI Aero-Flow automated powder flowability analyzer. The first frame shows the data output for lactose, which passed through a 38- μm screen (less than 38- μm). The output data for the original lactose sample (not screened) is shown in the middle frame. The last frame shows the flow data for the lactose sample remaining on the 38- μm screen (greater than 38- μm). (From Ref. 10.)

this article, but is also effected by or affect the flow characteristics of solids.

Paddle or blade rheometers are commonly used in to measure the shear of liquids. Brabender (145) developed such a system for solids in which a plate compacts the solid and the torque to initiate turning is measured. Podczek (146, 147) has tested seven different suppliers' microcrystalline cellulose with a new commercial blade type rheometer. In their experiments they were able to measure very small differences between these lots, which they related to actual capsule filling and tablet experiences. The measuring device is commercially available and is now marketed as the FT3 Powder Rheometer (148). (Podczek articles list the instrument as Wet and Dry Powder Rheometer Fingerprint, ManUmit Products Ltd., Bourne End, U.K.) In this instrument the powder is placed in a circular vessel with a closed bottom. The blade enters the powder and moves downward or upward in a helical motion while the force on the blade shaft is recorded. The helical path along which the blade moves is dependent on the axial and rotational speeds and the direction the blade rotates. The angle of approach the blade makes with the powder can be varied, and the direction and angle of the measurement will allow the measurement of compaction, shear and slicing of the powder bed (terms and conditions defined in Podczek's papers). The torque data is treated so as to give theoretically the largest torque exerted by the whole powder column on the rotor blade for a defined test condition. In this way powder shear can be measured under a variety of shear conditions, including various downward compaction modes and upward expansion modes. Podczek (146, 147) found that a downward compaction mode was most sensitive to small differences in the behavior of the various microcrystalline cellulose samples. These differences were not apparent using the Carr compressibility index. Other conditions such as slicing the bed was much less discriminatory. Freeman (149) reported on testing this device with talcum, zinc oxide, and an un-named pharmaceutical. By testing a material and finding the proper test conditions, this device could provide a simple and highly sensitive control test since it is relatively operator independent and yields quantitative values.

CONCLUSION

The final question that should be answered regarding these test methods is which one or ones should you use. Velasco et al. (150) compared data from static angle of repose, dynamic angle of repose, Carr compressibility

index, flow meter, and ring shear test for Ludipress® and Maltrin® M 150 and found a good correlation between all the methods. Tan and Newton (151) compared Carr compressibility, Hausner ratio, angle of repose, and the Jenike flow factor and demonstrated good correlation between all the test methods for five pharmaceutical excipients. They found no correlation between the angle of internal flow and the angle of effective friction. As mentioned previously, differences were seen from slow avalanche results, and faster avalanche results, and other physical measurements (143). As discussed above, rheometer measurements at different conditions (146, 147) and physical measurements for the same powders didn't always correlate.

Two very recent papers compare a number of flow measurement techniques and use statistics for comparison and to devise a new flow index. Lee et. al. (152) compared results from avalanche testing, Carr's compressibility index and critical orifice diameter for six pharmaceutical excipients. Statistical analysis established that there are relationships and similarities between the ranking of powder flow properties between these three methods. They also used a dual approach, which combines visual observation of the type of motion of the powder bed in rotation with the numerical descriptors such as mean time to avalanche and scatter. They found this dual approach was found to be more accurate in the assessment of powder flow than using the numerical descriptors alone. Taylor et al. (153) tested 41 pharmaceutical blends including pure excipients and active blends, using a vibrating spatula, avalanching, critical orifice, angle of repose, and compressibility index. An empirical composite index was established and powder flow was ranked in accordance with formulator experience. Principal components analyses of the angle of repose, percent compressibility, and critical orifice of the powders were also performed. Using principal components analysis, the results of these three tests were statistically weighted to provide a weighted composite index that the authors showed to be the best predictor of flow for these pharmaceuticals. The vibrating spatula and avalanche data were not consistent with formulator experience and cited vendor references for flow. The authors stated that the results of these two test methods might have more relevance if they had used further data treatment such as fractal analysis.

The process of solid flow is extremely complex and dependent on many material and process variables. The answer to the question of which is the best method to use is therefore dependent on your process and what you are trying to predict or measure and select the method or combination of methods that best meets your needs.

REFERENCES

1. Von Behren, Dale, A. Physical Characterization of Excipients in Practice. *Pharm. Technol.* **1996**; 87–89, June.
2. Cartilier, L.H.; Moës, A. Effect of Flowing Adjuvants on the Homogeneity and the Kinetics of Mixing of Low Dosage Cohesive Powder Mixtures. *Drug Dev. Ind. Pharm.* **1986**, 12 (8, 9), 1203–1218.
3. Sallam, E.A.; Orr, N.A. Studies Relating to the Content Uniformity of Ethinylestradiol Tablets 10 µg: Effect of Particle Size of Excipients. Third Expo./Congr. Int. Technol. Pharm. **1983**, 2, 28–37.
4. Samyn, J.C.; Murphy, K.S. Experiments in Powder Blending and Unbending. *J. Pharm. Sci.* **1974**, 63, 370–375.
5. Hauer, V.B.; Remmele, T.; Züger und Sucker, H. Gezieltes Entwickeln und Optimieren von Kapselformulierungen mit Einer Instrumentierten Dosierröhrchen-Kapselabfüllmaschine. *Pharm. Inc.* 55 Nr. 5 **1993**, .
6. British Standard 2955, British Standards Institute, London, 1958.
7. Standly-Wood, N. WS-16, Powder and Bulk Solids Conference and Exhibitions, Chicago, 1989.
8. Hiestand, E.N. Powders: Particle–Particle Interactions. *J. Pharm. Sci.* **1966**, 35, 1325–1344.
9. Carstensen, J.T.; Ertell, C.; Geoffrey, J.-M. Physico-Chemical Properties of Particulate Matter. *Drug Dev. Ind. Pharm.* **1993**, 19 (1, 2), 195–219.
10. TSI Aero-Flow™ Automated Powder Flowability Analyzer, TSI Inc.
11. Danish, F.Q.; Parrott, E.L. Flow Rates of Solid Particulate Pharmaceutical. *J. Pharm. Sci.* **1971**, 60, 548–554.
12. Gold, G.; Duvall, R.N.; Palermo, B.T.; Slater, J.G. Powder Flow Studies III. *J. Pharm. Sci.* **1968**, 57, 667–671.
13. Newton, J.M.; Bader, F. The Angle of Internal Flow as Indicator of Filling and Drug Release Properties of Capsule Formulations, *J. Pharm. Pharmacol.* **1987**, 39, 164–168.
14. Geoffroy, J.-M.; Carstensen, J.T. Modified Warren-Springs Equation, *Powder Technol.* **1993**, 76, 135–140.
15. Mandelbrot, B.B. *The Fractal Geometry of Nature*; Freeman, W. Ed.; San Francisco, 1983.
16. Kaye, B.H. *Characterization of Powders and Aerosols*; Wiley-VCH: Weinheim, 1999.
17. Kaye, B.H. *Chaos and Complexity: Discovering the Surprising Patterns of Science and Technology*; VCH: Weinheim, 1993.
18. Kaye, B.H. Characterizing the Flowability of Powder Using the Concepts of Fractal Geometry and Chaos Theory. Part. Part. Syst. Charact. **1997**, 14, 53–66.
19. Kaye, B.H. A New Approach to Powder Rheology. *Pharm. Technol.* **March**, 1994; 116–126, .
20. Hickey, A.J.; Concessio, N.M. Flow Properties of Selected Pharmaceutical Powders from a Vibrating Spatula. Part. Part. Syst. Charact. **1994**, 11, 457–462.
21. Brittain, H.G. Raw Materials. *Drug Dev. and Ind. Pharm.* **1989**, 15 (13), 2083–2103.
22. Cartilier, L.H.; Tawashi, R. Effect of Particle Morphology on the Flow and Packing Properties of Lactose. *S.T.P. Pharma Sci.* **1993**, 3 (3), 213–220.
23. Pitkin, C.; Carstensen, J.T. Effect of Particle Shape on Some Bulk Solid's Properties. *Drug Dev. Ind. Pharm.* **1990**, 16 (1), 1–12.
24. Wong, L.W.; Pilpel, N.J. Effect of Particle Shape on the Mixing of Powders. *J. Pharm. Pharmacol.* **1990**, 42, 1–6.
25. Danjo, K.; Kinoshita, K.; Kitagawa, K.; Lid, I.K.; Sunada, H.; Otsuka, A. *Chem. Pharm. Bull.* **1989**, 37 (11), 3070–3073.
26. Huber, G.M.W.; Becker, R.; Müller, R.H. Zusammenhang zwischen Fließeigenschaften und Oberfläche Pulverförmiger Rezepturen. *Pharm. Ind.* **1994**, 56 (4), 389–392.
27. Amidon, G.E.; Houghton, M.E. The Effect of Moisture on the Mechanical and Powder Flow Properties of Microcrystalline Cellulose. *Pharm. Res.* **1995**, 12 (6), 923–929.
28. Doelker, E.; Mordier, D.; Iten, H.; Humbert-Droz, P. Comparative Tableting Properties of Sixteen Microcrystalline Celluloses. *Drug Dev. Ind. Pharm.* **1987**, 13 (9–11), 1847–1875.
29. Heng, P.W.S.; Staniforth, J.N. The Effect of Moisture on the Cohesive Properties of Microcrystalline Celluloses. *J. Pharm. Pharmacol.* **1987**, 40, 360–362.
30. Staniforth, J.N.; Rees, J.E. Electrostatic Charge Interactions in Ordered Powder Mixes. *J. Pharm. Pharmacol.* **1982**, 34, 69–76.
31. Staniforth, J.N. British Pharmaceutical Conference Science Award Lecture 1986 Order out of Chaos. *J. Pharm. Pharmacol.* **1987**, 39, 329–334.
32. Marinelli, J. Factors that Impact a Bulk Solid's Flowability, (Internet). 2000; 1–2.
33. York, P.; Pilpel, N. Effect of Temperature on the Frictional, Cohesive, and Electrical Conducting Properties of Powders. *Mater. Sci. Eng.* **1972**, 9 (5), 281–291.
34. Pilpel, N.; Britten, J.R. Effects of Temperature on the Flow and Tensile Strengths of Powders. *Powder Technol.* **1979**, 22 (1), 33–44.
35. Onyekweli, A.O.; Pilpel, N. Cooling as a Possible Method for Increasing the Flowability of Certain Pharmaceutical and other Powders. *J. Pharm. Pharmacol.* **1980**, 32, 120–125.
36. Jayasinge, S.S.; Pilpel, N.; Harwood, C.F. Effect of Temperature and Compression on the Cohesive Properties of Particulate Solids. *Mater. Sci. Eng.* **1970**, 5 (5), 287–294.
37. Pilpel, N. Cool Powders Run Fast. *New Scientist* **1981**; 313–315.
38. Jenike, A.W. Storage and Flow of Solids. *Bull.* 123 **1964**, Utah Engineering Experimental Station, University of Utah, Salt Lake City.
39. Marinelli, J. Mass Flow Design Considerations? (Archiving—Part 1), (Internet). 2000; 1–2.
40. Bates, L. Flow in Bulk Storage Bins, (Internet). 2000; 1–4.
41. Marinelli, J. Wall Friction Effects, (Internet). 2000; 1–3.
42. Marinelli, J. How Wall Friction Effects Hopper Angles, (Internet). 2000; 1–3.
43. Marinelli, J. Compressibility —What is it? (Internet). 2000; 1–3.
44. Brown, C.J., Nielson, J., Eds.; *Silos Fundamentals of Theory, Behaviour and Design*, E & FN Spon: London and New York, 1998.
45. Marinelli, J.; Carson, J.W. Bulk Powders. *Encyclopedia of Chemical Technology*, 4th Ed.; 1996; 19, 1114–1141.

46. Mulcahy, D.E., Ed. *Materials Handling Handbook*, 1998.
47. Prescott, J.K.; Hossfeld, R.J. Maintaining Product Uniformity and Uninterrupted Flow to Direct-Compression Tableting Presses. *Pharm. Technol.* **1994**, *18* (6), 98–114.
48. Bates, L. The Use of Inserts in Hoppers, (Internet). Powder and Bulk Dot Com2000, 1–3.
49. Troxel, T.G. Flow Aids—What to Use and when to Use them. Powder/Bulk Solids Adv. Dry Process2000, 12–16.
50. Jones, T.M. Symposium on Powders, Society of Cosmetic Chemists of Great Britain, Dublin, Ireland, 1969
51. Peleg, M.; Mannheim, C.H. Effect of Conditioners on the Flow Properties of Powdered Sucrose. *Powder Technol.* **1972**, *7*, 45–50.
52. Neumann, B.S. *Advances in Pharmaceutical Sciences*. Bean, H.S., Beckett, A.H., Charles, J.E., Eds.; Academic Press: London, 1965, *2*, 181–207.
53. Jones, T.M.; Pilpel, N. Some Physical Properties of Lactose and Magnesia. *J. Pharm. Pharmacol.* **1965**, *17*, 440.
54. Borerro, J.M.; Muñoz Ruiz, A.J.; Jiménez-Castellanos, M.R. Relationship between the Flow Characteristics and the Intrinsic Factors of Chlorpromazine Hydrochloride and its Mixtures with Diluents as Direct Compression Formulations. *Boll. Chim. Farmaceutico. Anno.* **1994**, *133* (5), 294–300.
55. Chang, R.-K.; Leonzio, M.; Hussain, M.A. Effect of Colloidal Silicone Dioxide on Flowing and Tableting Properties of an Experimental, Crosslinked Polyalkylammonium Polymer. *Pharm. Dev. Technol.* **1999**, *4* (2), 285–289.
56. Nastruzzi, C.; Cortesi, R.; Esposito, E.; Genovesi, A.; Spadoni, A.; Vecchio, C.; Menegatti, E. Influence of Formulation and Process Parameters on Pellet Production by Powder Layering Technique. *AAPS Pharm. Sci. Tech.* **2000**, *1* (2), 1–22.
57. Podczek, F.; Miah, Y. The Influence of Particle Size and Shape on the Angle of Internal Friction and the Flow Factor of Unlubricated and Lubricated Powders. *Elsevier Int. J. Pharm.* **1996**, *144*, 187–194.
58. Van Ooteghem, M.; De Winter, B.; Ludwig, A. Influence of the Mixing Conditions on the Flow Properties of Powders to be Filled into Hard Gelatin Capsules. *Acta Pharm. Jugosl.* **1998**, *38*, 287–295.
59. Laughlin, S.; Carstensen, J.T. Relationship between Flow Rates of Granular Powders through Stationary and Moving Orifices. *J. Pharm. Sci.* **1981**, *70*, 711–713.
60. Joyce, M.A.; Liebert, R.T.; Carter, R.W. Eastern Regional Meeting, American Association of Pharmaceutical Scientists, Atlantic City, NJ, 1988.
61. Dahlinger, L.E.; Johansson, M.; Sjogren, J. Comparison of Methods for Evaluation of Flow Properties of Powders and Granulates. *Drug Dev. Ind. Pharm.* **1982**, *8* (3), 455–461.
62. Chowhan, Z.T.; Yang, I.C. Powder Flow Studies IV. Tensile Strength and Orifice Flow Rate Relationships of Binary Mixtures. *Int. J. Pharm.* **1983**, *14*, 231–242.
63. Muñoz-Ruiz, A.J.; Jiménez-Castellanos, M.R. Integrated System of Data Acquisition for the Measurement of Flow Characteristics. *Pharm. Technol. Int.* 1993; 21–26.
64. Flemming, J.; Mielck, J.B. Requirements for the Production of Microtablets: Suitability of Direct-Compression Excipients Estimated from Powder Characteristics and Flow Rates. *Drug Dev. Ind. Pharm.* **1995**, *21* (19), 2239–2251.
65. Rey, H.; Wagner, K.G.; Wehrle; Schmidt, P.C. Development of Matrix-Based Theophylline Sustained-Release Microtablets. *Drug Dev. Ind. Pharm.* **2000**, *26* (1), 21–26.
66. Moneghini, M.; Carcano, A.; Perissutti, B.; Rubessa, F. Formulation Design Studies of Atenolol Tablets. *Pharm. Dev. Technol.* **2000**, *5* (2), 297–301.
67. Faham, A.; Prinderre, P.; Farah, N.; Eichler, K.D.; Kalantzis, G.; Joachim, J. Hot-Melt Coating Technology. I. Influence of Compritol 888 Ato and Granule Size on Theophylline Release. **2000**, *26* (2), 167–176.
68. Fernández-Arévalo, M.; Vela, M.T.; Rabasco, A.M. Rheological Study of Lactose Coated and Acrylic Resins. *Drug Dev. Ind. Pharm.* **1990**, *16* (2), 295–313.
69. Timmins, P.; Delargy, A.M.; Minchom, C.M.; Howard, J.R. Influence of Some Process Variables on Product Properties for a Hydrophilic Matrix Controlled Release Tablet. *Eur. J. Pharm. Biopharm.* **1992**, *38* (3), 113–118.
70. Mollan, M.J., Jr.; Çelik, Characterization of Directly Compressible Maltodextrins Manufactured by Three Different Processes. *Drug Dev. Ind. Pharm.* **1993**, *19* (17, 18), 2335–2358.
71. Onunkwo, G.C.; Udeala, O.K. Studies of *Rauwolfia vomitoria* Root III. Flow Properties of R. vomitoria Granulations. *S.T.P. Pharma Sci.* **1995**, *5* (4), 296–301.
72. Goracinova, K.; Klisarova, L.; Simov, A.; Fredro-Kumbaradzi, E.; Petruševska-Tozi, L. Characterization of Fluid Bed Prepared Granulates with Verapamil Hydrochloride as Active Substance. *Acta. Pharm.* **1996**, *46*, 147–153.
73. Gohel, M.C.; Patel, L.D.; Modi, C.J.; Jogani, P.D. Functionality Testing of a Coprocessed Diluent Containing Lactose and Microcrystalline Cellulose. *Pharm. Technol.* 1999; 44–46, Yearbook.
74. Carr, R.L., Jr. Classifying Flow Properties of Solids. *Chem. Eng.* **1965**, *7*, 163–167.
75. Pilpel, N. Flow Properties of Non-Cohesive Powders. *Chem. Process Eng.* **1965**, *46*, 167.
76. Carstensen, J.T.; Lai, T.Y.F.; Toure, P.; Sheridan, J. Repose Angles as a Function of the Supporting Surface. *Int. J. Pharm.* **1980**, *5*, 157–160.
77. Jones, T.M.; Pilpel, N. Some Angular Properties of Magnesia and their Relevance to Material Handling. *J. Pharm. Pharmacol.* **1966**, *18* (Suppl.), 182–189.
78. Pilpel, N. Cohesive Pharmaceutical Powders. *Advances in Pharmaceutical Sciences*; Bean, H.S., Beckett, A.H., Charles, J.E. Eds.; Academic Press: London, 1971, *3*, 179–219.
79. Augenstein, D.A.; Hogg, R. An Experimental Study of the Flow of Dry Powders over Inclined Surfaces. *Powder Technol.* **1978**, *19*, 205–215.
80. Crowder, T.M.; Hickey, A.J. An Instrument for Rapid Powder Flow Measurement and Temporal Fractal Analysis. Part. Part. Syst. Charact. **1999**, *16*, 32–34.
81. Nqvist, H.; Nicklasson, M. Flow Properties of Compressible Lactose Containing Small Quantities of Drug Substances. *Drug Dev. Ind. Pharm.* **1985**, *11* (4), 745–759.
82. Nqvist, H. Measurement of Flow Properties in Large Scale Tablet Production. *Int. J. Pharm. Tech. Prod. Mfr.* **1984**, *5* (3), 21–24.

83. Parrott, E.L. Comparative Evaluation of a New Direct Compression Excipient, Soludex 15. *Drug Dev. Ind. Pharm.* **1989**, *15* (4), 561–583.
84. Carstensen, J.T. *Solid Pharmaceuticals: Mechanical Properties and Rate Phenomena*; Academic Press: New York, 1980.
85. Neuman, B.S. Ch. 10. *Flow Properties in Disperse Systems*; Herman, J.J., Ed.; North-Holland Publishing Co. Amsterdam, 1953.
86. Hausner, H.H. Friction Conditions in a Mass of Metal Powder. *Int. J. Powder Metall.* **1967**, *3* (4), 7–13.
87. Chowhan, Z.T.; Chow, Y.P. Evaluation of Pharmaceutical Availability from the Calculation of Drug Levels and Release Profiles. *Int. J. Pharm.* **1980**, *4*, 317–326.
88. Varthalis, S.; Pilpel, N. Anomalies in Some Properties of Powder Mixtures. *J. Pharm. Pharmacol.* **1976**, *28*, 415–419.
89. Sakr, F.M.; Pilpel, N. The Tensile Strength and Consolidation of Lactose Coated with Non-Ionic Surface-tants II-Tablets. *Int. J. Pharm.* **1982**, *10*, 43–56.
90. Yamashiro, M.; Yuasa, Y.; Kawakita, K. An Experimental Study on the Relationship between Compressibility, Fluidity and Cohesion of Powder Solids at Small Tapping Numbers. *Powder Technol.* **1983**, *34* (2), 225–231.
91. Geldart, D. Types of Gas Fluidization. *Powder Technol.* **1973**, *7* (5), 285.
92. Molerus, O. Interpretation of Geldart's Type A, B, C and D Powders by Taking into Account Interparticle Cohesion Forces. *Powder Technol.* **1982**, *33* (1), 81–87.
93. Munoz, R.; Velasco, A.; Antequera, M.V.; Monedero Perales, M.; del Carmen, J.-C.; Ballesteros, M.R. Powder Flow and Compression Characteristics of Dextrins for Direct Compression. *Congr. Int. Technol. Pharm.* **1992**, *4*, 131–139.
94. Flores, L.E.; Arellano, R.L.; Esquivel, J.J.D. Lubricant Susceptibility of Cellactose and Avicel PH-200: A Quantitative Relationship. *Drug Dev. Ind. Pharm.* **2000**, *26* (3), 297–305.
95. Flores, L.E.; Arellano, R.L.; Esquivel, J.J.D. Study of Load Capacity of Avicel PH-200 and Cellactose, Two Direct Compression Excipients, Using Experimental Design. *Drug Dev. Ind. Pharm.* **2000**, *26* (4), 465–469.
96. Mulderrig, K.B. Placebo Evaluation of Selected Sugar-Based Excipients. *Pharm. Technol.* **2000**, *24* (5), 34–42.
97. Viseras, C.; López-Galindo, A. Characteristics of Pharmaceutical Grade Phyllosilicate Powders. *Pharm. Dev. Technol.* **2000**, *5* (1), 47–52.
98. Kiekens, F.; Córdoba-Díaz; Remon, J.P. Influence of Chopper and Mixer Speeds and Microwave Power Level during the High-Shear Granulation Process on the Final Granule Characteristics. *Drug Dev. Ind. Pharm.* **1999**, *25* (12), 1289–1293.
99. Nicholas, V.; Chamblin, O.; Andrès; Rochat-Gonthier, M.-H.; Pourcelot, Y. Preformulation: Effect of Moisture Content on Microcrystalline Cellulose (Avicel PH-302) and its Consequences on Packaging Performances. *Drug Dev. Ind. Pharm.* **1999**, *25* (10), 1137–1142.
100. Tsai, T.; Wu, J.-S.; Ho, H.-O.; Sheu, M.-T. Modification of Physical Characteristics of Microcrystalline Cellulose by Codrying with β -Cyclodextrins. *J. Pharm. Sci.* **1998**, *87* (1), 117–122.
101. Armstrong, N.A. Direct-Compression Characteristics of Granulated Lactitol. *Pharm. Technol.* 1998; 84–92.
102. Vennat, B.; Gross, D.; Pourrat, A.; Pourrat, H. A Dosage Form for Procyanidins High-dose Tablets by Direct Compression. *S.T.P. Pharma.* **1988**, *4* (5), 378–383.
103. Ashton, M.D.; Cheng, D.C.H.; Farley, R.; Valentin, F.H.H. Some Investigations into the Strength and Flow Properties of Powders. *Rheol. Acta* **1965**, *4* (3), 206–217.
104. Eaves, T.; Jones, T.M. *Rheol. Acta*, **1971**, *10*, 127–134.
105. Kocova, S.; Pilpel, N.A. Failure Properties of Lactose and Calcium Carbonate Powders. *Powder Technol.* **1972**, *5* (6), 329–343.
106. Heistand, E.N.; Poet, C.B. Tensile Strength and Compressed Powders and an Example of Incompatibility as End-point and Shear Yield Locus. *J. Pharm. Sci.* **1974**, *63* (4), 605–612.
107. Irono, C.I.; Pilpel, N. Effects of Paraffin Coatings on the Shearing Properties of Lactose. *J. Pharm. Pharmacol.* **1982**, *34*, 480–485.
108. Johanson, J.R. Theory of Bulk Solids Flow a Historical Perspective. *Int. J. Bulk Solids Storage in Silos* **1987**, *3* (1), 1–15.
109. Doelker, E. Comparative Compaction Properties of Various Microcrystalline Cellulose Types and Generic Products. *Drug Dev. Ind. Pharm.* **1993**, *19* (17, 18), 2399–2471.
110. Ho, R.; Bagster, D.F.; Crooks, M.J. Flow Studies on Directly Compressible Tablet Vehicles. *Drug Dev. Ind. Pharm.* **1977**, *3* (5), 475–487.
111. Amidon, G.E.; Houghton, M.E. Powder Flow Testing in Preformulation and Formulation Development. *Pharm. Manu.* July 1985; 21–31.
112. Kata, M. Shear Cell Investigation of Powder Mixtures Containing Micronized Drugs. *Acta Pharm. Technol.* **1979**, *25* (3), 203–216.
113. Laloge, M.; Chulia, D.; Guillemoteau, J.Y.; Verain, A. *S.T.P. Pharma.* **1988**, *4*, 319–324.
114. Cohard, C.; Gonthier, Y.; Chulia, D.; Verain, A. Comparative Rheological Study of Different Materials with the Jenike Shear Cell. *J. Pharm. Belg.* **1984**, *39* (4), 209–216.
115. Chulia, D.; Verian, A. Une Nouvelle Cellule a Cisaillement Adaptee Aux Problemes Pharmaceutiques, (A New Shear-cell Adapted to Pharmaceutical Problems). *Ann. Pharm. Franc.* **1983**, *41*, 15–24.
116. Sanz Urgoti, E.; Vasquez Lopez, F. Reologia De Materiales Pulverulentos – I. Aproximacion a su Estudio. *Cienc. Ind. Farm.* **1986**, *5*, 350–353.
117. Nash, J.J.; Leiter, G.G.; Johnson, A.P. *Ind. Eng. Chem. Prod. Res. Dev.* **1965**, *4*, 140.
118. Hiestand, E.N.; Wilcox, C.J. Some Measurements of Friction in Simple Powder Beds. *Pharm. Sci.* **1968**, *57*, 1427.
119. Hiestand, E.N.; Wilcox, C.J. Shear Cell Measurements of Powders: Proposed Procedures for Elucidating the Mechanistic Behavior of Powder Beds in Shear. *J. Pharm. Sci.* **1969**, *58*, 1403.
120. Hiestand, E.N.; Valvani, S.C.; Peot, C.B.; Strzelinski, E.P.; Glasscock, J.F. Shear Cell Measurements of Powders: Determination of Yield Loci. *J. Pharm. Sci.* **1973**, *62* (9), 1513–1517.

121. Carr, J.F.; Walker, D.M. Annular Shear Cell for Granular Materials. *Powder Technol.* **1968**, *1*, 369–373.
122. Peschl, I.A.S.Z. Quality Control of Powders for Industrial Application. *Powder Handling Process* **1989**, *1* (4), 357–363.
123. Peschl, I.A.S.Z. Quality Control of Powders for Industrial Application, Proceeding of the 14th Annual Meeting Powder and Bulk Solids Rosemont, IL: May 1989, 517–536
124. Schulze, D. Development - Application of a Novel Ring. *Aufbereitungs-Technik* **2000**, *35* (10), 524–535.
125. Schulze, D. Flowability and Time Consolidation Measurements Using a Ring Shear Tester. *Powder Handling Process* **1996**, *8* (3), 221–226.
126. Schulze, D. *First European Symposium Process Technology in Pharmaceutical and Nutritional Sciences*; 1998, 276–285.
127. Nqvist, H. *Int. J. Pharm. Prod. Mfr.* **1984**, *5* (1), 13–17.
128. Baichwal, A.R.; Augsburger, L.L. Development and Validation of a Modified Annular Shear Cell to Study Frictional Properties of Lubricants. *Int. J. Pharm.* **1985**, *26*, 191–196.
129. Behres, M.; Klasen, C.-J.; Schulze, D. Development of a Shear Cell for Measuring the Wall Friction of Bulk Solids with a Ring Shear Tester. *Powder Handling Process* **1998**, *10* (4), 405–409.
130. Schulze, D. Ch. 2.2. *Measurements of the Flowability of Bulk Solids, Silos, Fundamentals of Theory, Behaviour and Design*; Brown, C.J., Nielson, J., Eds.; E. & FN Spon: London and New York, 1998; 18–52.
131. Enstad, G.G.; Feise, H. Ch. 2.3. *Flow Property Testing of Particulate Solids by Uniaxial and Biaxial Testers*; Brown, C.J., Nielson, J., Eds.; E. & FN Spon: London and New York, 1998; 53–64.
132. Luong, M.P. Ch. 2.4. *Reflections on Triaxial Testing, Rheology and Flowability, Silos, Fundamentals of Theory, Behaviour and Design*; Brown, C.J., Nielson, J., Eds.; E. & FN Spon: London and New York, 1998; 65–75.
133. Kolymbas, D.; Wu, W. Recent Results of Triaxial Tests with Granular Materials. *Powder Technol.* **1990**, *60* (2), 99–119.
134. Johanson, J.R. The Johanson Indicizer™ System vs. the Jenike Shear Tester. *Bulk Solids Handling* **1992**, *2*, 237–240.
135. Bell, T.A.; Ennis, B.J.; Grygo, R.J.; Scholten, W.J.F.; Schenkel, M.M. Practical Evaluation of the Johanson Hang-up Indicizer. *Bulk Solids Handling* **1994**, *14* (1), 117–125.
136. Chowhan, Z.T.; Yang, I.C. Powder Flow Studies III: Tensile Strength, Consolidation Ratio, Flow Rate and Capsule Filling–Weight Variation Relationships. *J. Pharm. Sci.* **1981**, *70* (8), 927–930.
137. Crowder, T.M.; Hickey, A.J. A Semiconductor Strain Gauge Based Instrument for Rapid Powder Flow Rate Measurement. *Respiratory Drug Delivery* **1998**, *6* 297–298.
138. Kaye, B.H.; Gratton-Liimatainen, J.; Faddis, N. Studying the Avalanching Behaviour of a Powder in a Rotating Disc. *Part. Part. Syst. Charact.* **1995**, *12*, 232–236.
139. Kaye, B. Moving and Storing Powders. Book in Preparation.
140. Hickey, A.J.; Concession, N.M. Chaos in Rotating Lactose Powder Beds. *Part. Sci. Tech.* **1996**, *14*, 15–25.
141. Kaye, B.H. Sampling and Characterisation Research: Developing Two Tools for Powder Testing. *Powder and Bulk Eng.* **1996**.
142. Trowbridge, L.; Grimsey, I.; York, P. Assessing Powder Flow from Avalanching Behaviour. *Pharmaceutical Technology*; School of Pharmacy, University of Bradford: BRADFORD BD7 1DP, API Ref. Lib. #80.
143. Trowbridge, L.; Williams, A.C.; York, P.; Worthington, V.L.; Dennis, A.B. A Comparison of Methods for Determining Powder Flow: Correlation with Tableting Performance. *Pharm. Res.* **1997**, *14* (11 (supplement)), 415.
144. Crowder, T.M.; Hickey, A.J. The Physics of Powder Flow Applied to Pharmaceutical Solids. *Pharm. Technol.* **2000**, *24* (2), 50–58.
145. Brabender, O.H.G. Duisburg: Flowability Test, Specification Sheet No. 2124, 1982.
146. Podczek, F. Rheological Studies of the Physical Properties of Powders Used in Capsule Filling—Part I. *Pharm. Technol. Europe* **1999**, *11* (9), 16–24.
147. Podczek, F. Rheological Studies of the Physical Properties of Powders Used in Capsule Filling—Part II. *Pharm. Technol. Europe* **1999**, *11* (10), 34–42.
148. TheNew FT3 Powder Rheometer—Freeman Technology.
149. Freeman, R. The Flowability of Powders, International Conference on Powder and Bulk Solids Handling, 2000.
150. Velasco, A.; Muñoz-Ruiz, A.J.; Perales, M.C.; Muñoz, N.; Jiménez-Castellanos, M.R. Evaluation of an Adequate Method of Estimating Flowability According to Powder Characteristics. *Int. J. Pharm.* **1994**, *103*, 155–161.
151. Tan, S.B.; Newton, J.M. Powder Flowability as an Indication of Capsule Filling Performance. *Int. J. of Pharm.* **1990**, *61*, 145–155.
152. Lee, Y.S.L.; Poynter, R.; Podczek, F.; Newton, J.M. Development of a Dual Approach to Assess Powder Flow from Avalanching Behavior. *AAPS Pharm. Sci. Tech.* **2000**, *1* (3), 1–14, Article 21 (<http://www.pharmscitech.com>).
153. Taylor, M.K.; Ginsburg, J.; Hickey, A.J.; Gheyas, F. Composite Method to Quantify Powder Flow as a Screening Method in Early Tablet or Capsule Formulation Development. *AAPS Pharm. Sci. Tech.* **2000**, *1* (3), 1–21, Article 18 (<http://www.pharmscitech.com>).

FLAVORS AND FLAVOR MODIFIERS

John M. Lipari

Thomas L. Reiland

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

INTRODUCTION

The use of flavors and flavor modifiers to improve the taste and aroma of foods and pharmaceuticals is an art that dates back several centuries. In large measure, the practice is still the same today and, except for the advent of new semisynthetic flavoring agents with improved stability, the field has remained relatively unchanged. In the analytical arena, the story is different. Sophisticated instrumentation methods have been developed to characterize, purify, and manufacture flavoring agents that are similar, in many respects, to those occurring in nature. The technology continues to evolve at an accelerated pace, resulting in several stable, potent, and unique flavors, which are now available to target both foods and pharmaceuticals. This article discusses flavors and flavoring agents typically used in industry and highlights formulation variables that could affect the performance of flavors in finished products. Where necessary, pertinent literature for further reading is cited.

DEFINITION OF FLAVOR

Flavor is the complex effect of three components: taste, odor, and feeling factors. It is usually associated with the pleasure of savoring food or beverages and has, subsequently, suffered from considerable imprecision in definition. Flavor is a sensation with multidimensional components involving subjective and objective perceptions. The sensory perceptions are both qualitative as well as quantitative and, therefore, can be measured. *Webster's New Collegiate Dictionary* defines flavor as the "... quality of something that affects the sense of taste, ... the blend of taste and smell sensations evoked by a substance in the mouth." This definition is correct, but incomplete, and should be redefined to include feeling factors.

Taste

Taste consists of four primary sensations: sweet, sour, bitter, and salty. Correspondingly, there are four different

kinds of taste buds. These sensations are elicited by the tongue and interpreted by the brain. Certain areas of the tongue respond more readily to specific tastes (1) than others. Sweet sensations are most easily detected at the tip of the tongue, whereas bitter ones are most readily detected at the back of the tongue. Sour sensations occur at the sides of the tongue, but salty sensations are usually detected at both the tip and at the sides of the tongue. During ingestion, taste buds react to soluble substances. The resulting sensations are transmitted to the brain by the ninth cranial (glossopharyngeal) nerve. The tenth and twelfth cranial nerves participate in this sensory reaction, but their role is limited (2).

Odor

The odor component of flavor is due to conscious or subconscious reactions to volatile substances, without which most foods would be lacking in taste appeal. By closing the nostrils while eating a mouthful of some flavored substance and immediately following this with another mouthful with the nostrils open, it may be shown that food could be rendered tasteless, as is often experienced by people suffering from the effects of a head cold.

There are many varieties of odorants, but a universally accepted structure-activity relationship of these has not been established. Yet, there is evidence that odor may involve specific receptor interactions (3), suggesting that structural properties of odorants may be important in eliciting specific odor sensations.

Feeling Factors

"Mouth feel" factors are critical in flavor perception. Examples include astringency, pepper bite, menthol cooling, and texture (e.g., softness or hardness as in candy). Sensations, such as crunch after biting into a crisp stick of celery or an apple, contribute to the overall flavor of foods. These mouth feel factors are also important in improving the organoleptic qualities of pharmaceuticals.

FLAVORING AGENTS

Flavoring agents may be classified as natural, artificial, or natural and artificial (N&A) by combining the all-natural and synthetic flavors (4). Pharmaceutical flavors are available as liquids (e.g., essential oils, fluid extracts, tinctures, and distillates), solids (e.g., spray-dried, crystalline vanillin, freeze-dried cinnamon powders, and dried lemon fluid extract), and pastes (e.g., soft extracts, resins, and so-called concretes, which are brittle on the outside and soft on the inside). Liquid flavors are by far the most widely used because they diffuse readily into the substrate. They are available both as oily (e.g., essential oils) or nonoily liquids. Their texture is generally dependent on the solvent within which they are prepared. Fluid extracts may contain a single ingredient or a variety of compounded ingredients. Tinctures are obtained by maceration or percolation of specific herbs and spices in alcohol.

Essential oils boil at elevated temperatures, but many cannot be directly distilled without decomposition. Vacuum, steam, and fractional or molecular distillation are often used for their manufacture. Fractional distillation removes traces of water, resinous materials, colors, terpenes, and sesquiterpenes from the distillate. This process improves solubility and enhances flavor intensity. Sesquiterpeneless oils are more soluble than terpeneless oils because of the removal of head and tail fractions (e.g., waxy residues). Most common sesquiterpeneless oils used in the pharmaceutical industry include oil of orange and oil of lemon.

Oils and juices are obtained from plant sources by expression. Citrus essential oils are almost exclusively obtained by this method. Thoroughly washed unripe citrus fruits are cold pressed manually, or mechanically, to rupture oil cells in the rind. The oil is collected by draining and centrifuging. Manual operation is labor intensive and has been replaced by machines.

Natural Flavoring Agents

Natural ingredients have been used since antiquity to flavor foods and to make early “medicines” palatable. Honey was and remains a sweetener and flavoring agent. Wine was used as a crude infusorial in medicinal herbs. Modern use of natural flavors in pharmaceuticals is limited, because they are often unstable and their quality is unpredictable from season to season. The most commonly used natural flavors are terpeneless citrus oils, which are stable if well protected from light and air. A variety of other natural flavors are used in the food and

pharmaceutical industries; some of the more common flavors are described below.

Anise (*Pimpinella anisum*, Umbelliferae)

Anise is a herbaceous annual cultivated extensively in Europe. The essential oil is obtained by steam distillation of dried fruits (seeds). The distillate is a clear-to-pale yellowish oil. It solidifies at low temperatures and has a characteristic sweet licorice-like odor and flavor. Its main constituents include anethol (approximately 90%), methylchavicol, *p*-methoxyphenylacetone, and acetic aldehyde. Anise oil is used frequently at concentrations of up to approximately 3000 ppm in liquid preparations.

Cardamon (*Elettaria cardamomum*, Zingiberaceae)

Cardamon is cultivated in India and Sri Lanka. The essential oil is obtained by steam distillation of comminuted seeds to yield a greenish-yellow liquid with a warm, spicy, aromatic odor and flavor. The main constituents of the oil are limonene, cineol, D- α -terpineol, and terpinyl acetate. Cardamon is generally used at concentrations of approximately 5–50 ppm.

Wild Cherry (*Prunus serotina*, Rosaceae)

Wild cherry is a large tree, native to southern Canada. It is widespread in the United States and Europe. The bark, small branches, and twigs are used to prepare the fluid extract and tincture. The main constituent of wild cherry extract is the glucoside prunasin, which on enzymatic hydrolysis yields prussic acid, glucose, and benzaldehyde. Also present are coumarin, phytosterols, benzoic acid, and fatty acids (e.g., oleic, linoleic, and palmitic acids). It has a characteristic sweet, tart, cherry-like flavor. Wild cherry bark extract is commonly used at concentrations of approximately 50–800 ppm in foods and pharmaceuticals.

Lemon (*Citrus limonum*, Rutaceae)

Lemon is an evergreen shrub or tree native to the Far East; it was introduced to the Mediterranean regions at the time of the Crusades. The leaves, fruits, and rind are used either whole or pressed in foods and in liquid or solid pharmaceutical products. The essential oil of lemon is obtained by cold expression. Approximately 40 constituents have been identified, with 90% being limonene. Fluid extracts and tinctures are obtained from the dried peel.

Lemon petitgrain is obtained by steam distillation of the leaves. For flavoring, it must be terpeneless. The main constituents are D- α -pinene, camphene, D-limonene, dipentene, L-linalol, nerol, and citral. Lemon petitgrain oil is used in a wide variety of applications. Typical concentrations range from 1 to 35 ppm. The essential oil and extract of lemon are generally used at higher concentrations that may range up to 1000 and 10,000 ppm, respectively. All lemon oil derivatives have the characteristic lemon odor and a slightly bitter flavor.

Orange, Bitter (*Citrus aurantium*, Rutaceae)

Bitter orange is a tall tropical tree that can grow up to approximately 10 m (33 ft.) high. The tree is native to the Far East and is cultivated extensively throughout the Mediterranean, Guinea, the West Indies, West Africa, and Brazil. The leaves and twigs produce essential petitgrain oil following steam distillation. *Neroli bigarade* essential oil is produced from the blossoms by steam distillation. The peel is expressed and steam distilled to produce essential oil of orange. The main constituent of orange oil is D-limonene, with various acids, aldehydes, and diesters. Essential oil of orange is widely used in foods and pharmaceuticals at concentrations of up to 500 ppm.

Orange, Sweet (*Citrus sinensis*, var. *aurantium dulcis*, Rutaceae)

Sweet orange is an evergreen tree that grows to approximately 6 m (20 ft.) high. It is generally of oriental origin and is cultivated extensively in the Mediterranean, Florida, and California. A petitgrain oil is obtained from the leaves and twigs, but its production is low because of limited use, primarily in the perfumery industry.

Essential oil of sweet orange is obtained by expression. Its physical–chemical properties (e.g., specific gravity, optical rotation, and refractive index) vary according to origin. The oil contains more than 90% limonene, in addition to relatively high quantities of decylic, octylic, nonylic, and dodecylic aldehydes, and citral esters. It has a characteristic odor and a mildly bitter, astringent flavor; it is generally used at concentrations of up to 500 ppm.

Peppermint (*Mentha piperita*, Labiatae)

Peppermint is a herbaceous plant that grows to approximately 81 cm (32 in.) high. The essential oil is obtained by steam distillation of the flowering plant tops. It is cultivated in Europe, North and South America, and

Japan. The main constituents of the essential oil are α - and β -pinene, limonene, cineol, ethyl amylcarbinol, menthone, menthol, isomenthol, menthyl acetate, and piperitone. It has a strong mint odor with a sweet balsam taste masked by a strong cooling effect. It is widely used in foods, as well as in liquid pharmaceuticals, to 8000 ppm.

Artificial Flavoring Agents

Unlike natural flavoring agents, synthetic flavors are usually stable. The development of synthetic flavors paralleled the development of instrumental analysis, in which active ingredients in natural flavors are identified and reconstructed synthetically with reasonable accuracy. Exact duplication of a natural flavor is, however, difficult because often minor components are the most important contributors to the overall flavor profile. These minor components are not easily identified. For example, the major components of vanilla are vanillin and ethyl vanillin. However, the flavor nuances of the vanilla bean have never been successfully matched in artificial (synthetic) vanilla.

Natural and Artificial (N&A) Agents

In N&A flavor systems, natural flavors are combined with synthetic ingredients to enhance flavor balance and fullness. These flavors are generally classified according to type and taste sensation. Table 1 contains a list of N&A flavor components that elicit various sensory properties, all of which are commonly used in food and drug compounding. It is not an exhaustive list because manufacturers regard their flavor formularies as proprietary. Many N&A flavors may be chemically and structurally similar, but vary significantly in taste and aroma. Similar flavors from various vendors might vary significantly in composition. Of interest is the fact that a relatively small change in chain length can have a profound impact on flavor type. Minor changes, such as the conversion of allyl benzoate to cyclohexyl esters (e.g., a caproate or valerate), transform a basic cherry flavor to peach or pineapple.

In situ conversion of essential N&A flavor components from one molecular form to another, as a result of ion pairing, is common in food and drug products. Therefore, the inadvertent conversion of flavors between types during drug formulation studies (e.g., effect of pH, salts, and temperature) can present a serious challenge in flavor-quality assessment. The fact that one and the same N&A flavor component can deliver several flavor and odor impressions implies that a blend of several flavor compounds would be preferable. Such blends show improved stability. In addition, flavor impressions from

Table 1 Primary taste and flavor characteristics of typical flavor ingredients

Ingredient	Primary taste			Flavor characteristic
	Sweet	Bittersweet	Bitter	
Allyl benzoate		X		Cherry
Allyl caproate		X		Pineapple
Allyl cyclohexylbutyrate		X		Pineapple
Allyl cyclohexylcaproate		X		Peach/apricot
Allyl cyclohexylvalerate		X		Apple
Allyl phenoxyacetate		X		Honey/pineapple
Anethol	X			Anise
Anisyl alcohol	X			Peach
Anisyl formate	X			Strawberry
Benzyl isobutyrate	X			Strawberry
Benzyl salicylate	X			Raspberry
Cinnamaldehyde		X		Cinnamon/melon
Cinnamyl anthranilate		X		Grape
Citral		X		Lemon
Citronellyl formate		X		Plum
Cyclohexyl caproate		X		Peach/cognac
Decyl formate	X			Grape
Diacetyl	X			Butter
Diphenyl ether	X			Black currant
Ethyl valerate		X		Banana/apple
Eugenol			X	Clove buds
Geraneol			X	Rose-like
α -Ionone	X			Raspberry
Isoamyl salicylate		X		Strawberry
Isobutyl anthranilate		X		Grape/strawberry
Isopropyl valerate		X		Apple
Linalyl anthranilate	X			Orange
Methyl ionone	X			Raspberry/currant
Methyl propionate	X			Black currant
Methyl undecyl ketone		X		Coconut
Musk ambrette	X			Peach
Nerol			X	Rose-like
Neryl acetate	X			Raspberry
Neryl butyrate	X			Cocoa
Propenyl guaethol	X			Vanilla
Propyl isobutyrate	X			Pineapple
Rhodinol			X	Rose
Santaly acetate		X		Apricot
Terpenyl butyrate		X		Plum
Tetrahydrofurfuryl propionate		X		Chocolate/apricot
Vanillylidene acetone		X		Vanilla
Yara yara	X			Strawberry

N&A flavor blends are usually not dominated by a single component. For these reasons, single natural and artificial flavor ingredients are seldom, if ever, used alone in a finished product.

Tables 2 and 3 list the components thus far qualitatively identified (4) in two common fruits, the raspberry and the

strawberry. These tables illustrate the complexity in compounding synthetic systems to mimic natural types. For this reason, there has been a steady rise in the use of N&A flavors, in addition to their superior performance, when compared to natural flavors. Also, the quality and uniformity of the N&A flavor is greater than that of natural

Table 2 Natural components of raspberry aroma

Acids	Carbonyls	Esters	Alcohols
Acetic	Acetaldehyde	Butyl acetate	1-Butanol
Butyric	Acetone	Ethyl acetate	<i>trans</i> -2-Buten-1-ol
Caproic	Acetyl methyl carbinol	Ethyl butyrate	Ethanol
Caprylic	Acrolein	Ethyl crotonate	Geraniol
Formic	Diacetal	Ethyl propionate	1-Hexanol
2-Hexenoic	β , β -Dimethylacrolein	2-Hexenyl acetate	<i>cis</i> -3-Henen-1-ol
3-Hexenoic	Hexanal	2-Hexenyl butyrate	Methanol
Isobutyric	2-Hexenal	Hexyl acetate	3-Methyl-3-buten-1-ol
Isovaleric	<i>cis</i> -3-Hexenal	Hexyl butyrate	1-Pentanolol
Propionic	4-(<i>p</i> -Hydroxyphenyl)-2-butanone	Isoamyl acetate	1-Penten-3-ol
Valeric	α -Ionone	Isopropyl butyrate	
	β -Ionone	Methyl butyrate	
	2-Pentanone	Methyl caproate	
	2-Pentanal	Methyl caprylate	
	Propanal	Propyl acetate	

(From Ref. 4.)

flavors, and lower concentrations of N&A flavors are often used to achieve the same effect as obtained with natural flavors. Table 4 shows a typical formulation of a commercial N&A strawberry flavor. It contains a small proportion of natural flavors; the bulk of the ingredients are synthetic.

Another advantage of N&A flavors is the broad spectrum of flavoring agents from which the formulator can develop an entirely new flavor system that is unique, not available naturally. A flavor extensively used in foods and pharmaceutical products is tutti-frutti—bubble gum (Table 5).

In summary, there are a variety of flavor types: natural, synthetic, and semisynthetic. Appropriate use concentrations depend on many factors, including product characteristics, such as composition, physical state, shelf life, pH, processing temperature, storage conditions, and reactivity of components. Flavor concentrations also depend on the market sector for which the product is targeted. The age of the user and the mode of use are two examples of user-dependent variables that have significant bearing on the type, concentration, and nature of the flavor selected for product development.

FLAVOR SELECTION IN PHARMACEUTICAL PREPARATIONS

A number of criteria are used to select flavors during formulation. Different flavor concentrations produce

highly subjective sensations. Specific requirements for balance and fullness are dependent, in part, on the drug substance and the physical form of the product. For this reason, when selecting a flavor system, the compounding pharmacist must take into account several variables upon which a desired response would depend. Some of these are product texture (e.g., viscosity of formulation, solid or liquid), water content, base vehicle or substrate, and taste of the subject drug. Notable specific examples to consider are:

- Immediate flavor identity from the formulation as it is ingested;
- Compatible mouth feel factors and rapid development of a fully blended flavor in the mouth during ingestion of the product;
- Absence of “off” notes in the mouth and a mild transient aftertaste during ingestion of the product.

The selection of a flavor system, thus, requires an extensive evaluation of a number of organoleptic qualities. Vehicle components within which the drug is presented have a significant bearing on the performance of the flavor system. Of these, the sweetener is perhaps the most relevant.

Sweeteners

The most commonly used sweeteners are sucrose, glucose, fructose, sorbitol, and glycerin. Using sucrose (sugar) as a standard, with 100 units of sweetness, Table 6 lists the

Table 3 Natural components of strawberry aroma

Acids	Esters	Alcohols
Acetic	Butyl acetate	Benzyl alcohol
Benzoic	Ethyl acetate	1-Borneol
Butyric	Ethyl acetoacetate	Butanol
Caproic	Ethyl benzoate	2-Butanol
Cinnamic	Ethyl butyrate	Ethanol
Formic	Ethyl capronate	2-Heptanol
Isobutyric	Ethyl cinnamate	Hexanol
Isovaleric	Ethyl crotonate	<i>trans</i> -2-Hexanol
α -Methylbutyric	Ethyl formate	<i>p</i> -Hydroxyphenyl-2-ethanol
Propionic	Ethyl isobutyrate	Isoamyl alcohol
Salicylic	Ethyl isovalerate	Isobutanol
Succinic	Ethyl α -methylbutyrate	Isofenchyl alcohol
<i>n</i> -Valeric	Ethyl propionate	Methanol
Isovaleric	Ethyl salicylate	1-Pentanol
	Ethyl valerate	Penten-1,3-ol
	<i>trans</i> -2-Hexenyl	Phenyl-2-ethanol
	<i>trans</i> -2-Hexenyl acetate	<i>n</i> -Propanol
	Hexyl acetate	DL- α -Terpineol
	Hexyl butyrate	<i>cis</i> -Terpineol hydrate
	Isoamyl acetate	
	Isopropyl butyrate	
	Methyl acetate	
	Methyl butyrate	
	Methyl capronate	
	Methyl isobutyrate	
	Methyl- α -methylbutyrate	
	Propyl acetate	
Carbonyl compounds		Others
Acetaldehyde		Acetals
Acetophenone		Acetoin
Acetone		γ -Decalactone
Acrolein		1,1-Diethoxyethane
<i>n</i> -Butanal		1,1-Dimethoxyethane
Crotonal		Dimethoxymethane
Diacetyl		Dimethyl sulfide
2-Heptanone		Hydrogen sulfide
<i>cis</i> -3-Hexal		1-Ethoxy-1-propoxyethane
Methyl-3-butanone		Maltal
2-Pentanone		1-Methoxy-1-ethoxyethane
2-Pentanal		Methyl sulfide

(From Ref. 4.)

relative intensities of other sweeteners. Sweetness intensity changes with concentration. It has been estimated (5) that the sweetness of glucose relative to cane sugar is 53 at a concentration of 8% but increases to 88 at a concentration of 35%. Sweetener intensity increases with concentration but reaches a maximum where feeling factors become more prominent. Sugar (sucrose), at a concentration of 30%, is intensely sweet. Yet, its sweet

intensity at concentrations 50% or higher is not perceptibly different, although distinct mouthfeel characteristics (e.g., syrupy, salivation) become pronounced. This is due to osmotic effects on mucous membranes within the oral cavity.

Glycerin, glucose, sorbitol, and sucrose have limited use in solid dosage forms (e.g., tablets) because the materials are hygroscopic. Mannitol is used more often in

Table 4 A typical formula composition of natural and artificial strawberry flavor

Ingredient	Parts by weight
Amyl acetate	34.0
Amyl butyrate	15.0
Anethole	1.5
Butyric acid	15.0
Cinnamyl valerate	9.5
Diacetyl	10.0
Ethyl amyl ketone	15.0
Ethyl cinnamate	52.0
Ethyl methylphenylglycidate	260.0
Ethyl valerate	60.0
Lemon essential oil	1.0
Maltol	70.0
Methyl heptene carbonate	0.5
Neroli essential oil	0.5
γ -Undecalactone	58.5
α -Ionone	6.5
Amyl valerate	15.0
Jasmine absolute	85.0
Cinnamyl isobutyrate	7.0
Cognac essential oil	1.5
Ethyl acetate	50.0
Ethyl butyrate	30.0
Ethyl heptylate	2.5
Ethyl propionate	15.0
Hydroxyphenyl-2-butanone	0.5
Methyl anthranilate	6.5
Methyl cinnamate	35.5
Methyl salicylate	6.5
Orris resinoid	1.5
Vanillin	70.0
Solvent {ethylene glycol ethyl ester}	1060.0
Total	2000.0

(From Ref. 4.)

tablet manufacture. Besides being less hygroscopic, it has a negative heat of solution. For this reason, chewable tablets containing mannitol have a pleasant cooling sweet taste, which complements flavor quality. The artificial sweetener saccharin is widely used in foods and pharmaceuticals. It is approximately 350 \times as sweet as sugar. It is sweet at very low concentrations (equivalent to about 5–10% sugar) but bitter at higher concentrations. Approximately 20% of the population are “saccharin sensitive;” that is, they perceive saccharin to be bitter even at low concentrations. Upon repeated tasting, saccharin becomes less sweet and increasingly bitter. By the third or fourth tasting, solutions of relatively low concentrations are often no longer sweet to the saccharin-sensitive person.

Table 5 Formula and composition of natural and artificial tutti-frutti flavor

Ingredient	Parts by weight
Amyl acetate	300.0
Amyl butyrate	48.0
Ethyl butyrate	36.0
α -Ionone	120.0
Jasmine absolute (10 % in alcohol)	0.1
Lemon essential oil	1.0
Orris resinoid	80.0
Imitation rose (10 % in alcohol)	28.0
Rum ether	100.0
γ -Undecalactone	18.0
Vanillin	11.0
Alcohol (solvent)	257.0
Total	1000.0

(From Ref. 4.)

The artificial sweeteners, cyclamate and aspartame, are about 30 \times as sweet as sugar, but like saccharin, their sweet-bitter profiles are concentration dependent. Aspartame does not have a significant bitter aftertaste when compared to saccharin and has gained in popularity. Cyclamates were banned in the 1970s because of carcinogenic concerns, which have, subsequently, been shown to be overstated.

Monoammonium glycyrrhizinate has a lingering sweet aftertaste, which can be exploited for taste-masking products with a mildly bitter aftertaste. It is also effective in enhancing chocolate flavor. Glycerin is commonly used for its solvent effect on many compounds, as well as its humectant effect. Sugar syrups promote significant “cap-locking”—the crystallization of the sugar on the cap and bottle thread, but the addition of glycerin (10–20%) minimizes this effect. Glycerin is seldom used as a single sweetener in pharmaceuticals because it has a characteristic mouth-warming and burning effect.

Flavor Enhancers and Potentiators

Flavor enhancers are used universally in the food and pharmaceutical industries. Sugar, carboxylic acids (e.g., citric, malic, and tartaric), common salt (NaCl), amino acids, some amino acid derivatives (e.g., monosodium glutamate—MSG), and spices (e.g., peppers) are most often employed. Although extremely effective with proteins and vegetables, MSG has limited use in pharmaceuticals because it is not a sweetener. Citric acid

Table 6 Intensity values for frequently used sweeteners

Sweetener	Intensity
Sorbitol	60
Mannitol	50
Hydrogenated starch hydrolysate	30–40
Maltitol solution	70–80
Maltitol	90
Xylitol	100
Erythritol	60–70
Glycerin	55–75
Sucrose	100
Fructose	117
Maltose	30

(From SPI Polyols, Inc., Polyols Comparison Chart, Revised 6/99.)

is most frequently used to enhance taste performance of both liquid and solid pharmaceutical products, as well as a variety of foods. Other acidic agents, such as malic and tartaric acids, are also used for flavor enhancement. In oral liquids, these acids contribute unique and complex organoleptic effects, increasing overall flavor quality. Common salt provides similar effects at its taste threshold level in liquid pharmaceuticals. Vanilla, for example, has a delicate bland flavor, which is effectively enhanced by salt.

Taste-Masking Agents

The flavoring industry has many proprietary products purported to have excellent taste-masking properties (4), which have been used with some success. Yet, there are a number of natural and artificial flavors that can be generally described to possess similar taste-masking effects (Table 7).

Table 7 Agents for masking and complementing the basic tastes

Basic taste	Masking agent
Sweet	Vanilla, bubble gum, grape, other fruits
Acid	Lemon, lime, orange, cherry, grapefruit
Metallic	Berries, mints, grape, marshmallow, gurana
Bitter	Licorice, coffee, chocolate, mint, grapefruit, cherry, peach, raspberry, orange, lemon, lime

Of the many tastes that must be masked in pharmaceuticals, bitterness is most often encountered; to mask it completely is difficult. A tropical fruit has been used for centuries in central Africa to mask the bitter taste of native beers. This so-called “miracle berry” contains a glycoprotein that transiently and selectively binds to bitter taste buds. Due to stability challenges, attempts to isolate the compound for commercial exploitation have been unsuccessful. Yet, many fruit syrups are relatively stable in pharmaceuticals if formulated with antimicrobial preservative agents. Syrups of cinnamon, orange, citric acid, cherry, cocoa, wild cherry, raspberry, or glycyrrhiza elixir can be used to effectively mask salty and bitter tastes in a number of drug products (6). The extent to which taste-masking may be achieved is not usually predictable due to complex interactions of other flavor elements in these products. The degree to which bitterness may be masked by these agents ranks in a descending order: cocoa syrup is most effective, followed by raspberry syrup, cherry, cinnamon, compound sarsaparilla, citric acid, licorice, aromatic elixir, orange, and wild cherry.

Sour and metallic tastes in pharmaceuticals also can be reasonably masked. Sour substances containing hydrochloric acid are most effectively neutralized with raspberry and other fruit syrups. Metallic tastes in oral liquid products (e.g., iron) are usually masked by extracts of gurana, a tropical fruit. Gurana flavor is used at concentrations ranging from 0.001 to about 0.5% and may be useful in solid products as well (e.g., chewable tablets and granules).

FLAVOR MODIFICATION TECHNOLOGIES

Solubility-Limiting Methods

Many drugs are reasonably soluble in water and ionize extensively at physiologic pH. Drugs with an offensive taste usually demonstrate negative organoleptic properties after dissolving in saliva during ingestion. Chemical modification, such as derivatization or lipophilic counterion selection, where possible, may be an effective method for reducing aqueous solubility and taste. This is exemplified by erythromycin, a partially soluble, bitter-tasting macrolide anti-infective. The solubility of erythromycin monohydrate is approximately 2 mg/mL in water at a pH of approximately 7. When converted to erythromycin ethylsuccinate, the aqueous solubility of the drug is less than 50 µg/mL. This form is practically tasteless as a ready-made liquid or a chewable tablet. The rate of dissolution in body fluids (e.g., saliva) may be further reduced by controlling formulation pH, solids content,

and temperature. This technique can be applied to a number of drugs whose taste profiles are dependent on aqueous solubility.

Vesicles and Liposomes

Host–guest systems (7) (e.g., phospholipids and certain surfactants) form spherical or ellipsoidal, closed, bilayer structures called vesicles. These structures often comprise several compartments within which a drug could be trapped, either as a solution or a dispersion. Under various conditions, these vesicles form closed systems, which are ideal vehicles for taste masking or for modulated release of drug in vivo. It is a challenge to formulate drugs with these flavor-masking methods without altering the regulatory status of the product (e.g., chemical designation of the active substance, in vitro dissolution kinetics, physical or chemical stability, and bioavailability). Various manufacturers (e.g., American Lecithin Company, Oxford, Connecticut) offer a complete line of phospholipids (purified and solubilized in various carrier systems) for use in the food and pharmaceutical industries.

Microencapsulation and Coated Systems

Recently, a great deal of attention has been focused on the usefulness of coated fine particles in achieving pharmaceutical objectives. By coating drug particles with an appropriate polymer system, desirable properties can be imparted to the dosage form, eliminating undesirable properties, such as taste. Coating drug particles significantly modulates drug release while improving taste, stability, and other handling characteristics (e.g., flow and compression). Commercial particle coaters make it possible to coat fine drug particles, achieving slow release and taste masking in oral formulations. Examples for which particle coating has been used to introduce unique line extensions to the marketplace include Theo-Dur[®] and Depakote Sprinkle[®]. In the case of Theo-Dur[®], theophylline is sprayed onto sugar beads followed by a polymer to control drug release. By encapsulating a drug substance, this process prevents interaction of the drug with taste receptors, thus eliminating bitterness. Other frequently used microencapsulation methods include spray drying, spray congealing, coacervation and phase separation, interfacial polymerization, and extrusion.

Complexation and Chemical Modification

The use of ion exchange resins to form drug adsorbates for sustained release (8, 9) was closely associated with

Strassenburgh Laboratories, an affiliate of Pennwalt Corporation, which was granted several patents in this area (10, 11). Their first significant application involved amphetamine adsorbed onto a sulfonic acid cation exchange resin (Biphedamine[®]) for use in appetite suppression (12). Over the years, several products have been introduced commercially since the initial work with amphetamine (13–16); examples include Ionamin[®] (phentermine: Medeva Pharmaceuticals, Inc.), Tussionex[®] (hydrocodone polistirex and chlorpheniramine polistirex: Medeva Pharmaceuticals, Inc.), and a variety of cough-cold products, including phenylpropanolamine, chlorpheniramine, and dextromethorphan (17). This technology is applicable to taste masking as well.

The mechanism of drug release from the sustained–release complex (e.g., an ion exchange resinate consisting of a drug with a bitter taste) is ideal for liquids, when formulated either as granules for reconstitution or ready-made suspensions. By retaining a low counter-ion concentration in the product, almost all of the drug may be retained in the matrix, so that upon ingestion, ions of the body trigger the release mechanism through a dynamic equilibration process. Slow equilibrating complexes that provide low diffusivity of drug to the taste buds (e.g., low aqueous solubility of the drug) can eliminate bitterness and other offensive organoleptic drug properties. Other complexation phenomena employed in formulation work for flavor enhancement are: inclusion complexes (e.g., cyclodextrins and their derivatives), matrices, and physical complexes with waxy substances (e.g., polyethylene glycols).

More recently, pharmaceutical manufacturers have introduced various technologies for coating drug particles with semipermeable polymeric membranes designed to provide controlled release in vivo. Coatings of neat drugs and their adsorbates (18–21) combined controlled–release characteristics with the benefit of taste masking, caused by the effective reduction of dissolved drug concentrations in the mouth. Taste evaluation of a variety of these preparations showed a significant reduction in bitterness (19–21), suggesting that coatings and adsorbates have potential in the flavor enhancement of drugs with offensive tastes.

CONCLUSIONS

The use of flavors and flavor modifiers in pharmaceutical formulations is of considerable importance in promoting drug products. Flavors are also key factors in promoting patient compliance, because products with offensive taste

are likely to be objectionable. Taste sensations are, however, wholly subjective, and of the many objective methods thus far used, none can adequately and completely characterize taste and aroma sensations without some bias.

It is also certain that a fair proportion of the population is indifferent to taste and lacks the acuity necessary for distinguishing small differences in taste and aroma between samples (22, 23). Furthermore, clear flavor performance differences can be obtained during pharmaceutical product development, by techniques designed to promote taste acceptance. The use of flavors, flavor modifiers, and other methods for flavor enhancement, such as physical and chemical manipulations of drugs, may be potential methods for the development of products with superior market preference characteristics.

ACKNOWLEDGMENTS

The authors are grateful to Akwete L. Adjei and Richard Doyle for their contributions to the 1st Edition of this article and to Tracy B. Lynch for her assistance in the final preparation of this manuscript.

REFERENCES

1. Beidler, L.M. *Olfaction and Taste*; Zotterman, Y., Ed.; Pergamon Press, Inc.: New York, 1963.
2. Pfaffmen, C.J.J. *Neurophysiol.* **1955**, *18*, 429–444.
3. Amerine, M.A.; Pangborn, R.M.; Roessler, E.B. *Principles of Sensory Analysis of Food*; Academic Press: New York, 1965; 193–206.
4. Furia, E., Bellanca, N., Eds. *Fenaroli's Handbook of Flavor Ingredients*; CRC Press: Cleveland OH, 1971.
5. Renner, H.D. *Confect. Prod.* **1939**, *5*, 255–256.
6. In *Remington's Pharmaceutical Sciences*; Gennaro, A.R., Ed.; Ch. 69 Mack: Easton, PA, 1980.
7. Fendler, J.H. Membrane Mimetic Chemistry. *Chem. Eng. News* **1984 Jan 2**, 25–38.
8. Martin, G.J. *Ion Exchange and Adsorption Agents in Medicine*; Little, Brown, & Co.: Boston, MA, 1955.
9. Calmon, C.; Kressman, T.R.E. *Ion Exchangers in Organic and Biochemistry*; Interscience: New York, 1957.
10. Keating, J.W. US Patent 2,990,332, 1961; US Patent 3,143,465, 1964.
11. Hays, E.E. US Patent, 3,035,979, 1962.
12. Deeb, G.; Becker, B. Absorption of Sustained-Release Oral Amphetamine Preparations in the Rat. *Toxicol. Appl. Pharmacol.* **1960**, *2*, 410.
13. Swift, J.G. A Study of Sustained Ionic Release Anti-histamines. *Arch. Int. Pharmacodyn.* **1960**, *124*, 341.
14. Wulff, O.J. Prolonged Antitussive Action of a Resin-Bound Noscapine Preparation. *Pharm. Sci.* **1965**, *54*, 1058.
15. Brudney, N. Ion-exchange Resin Complexes in Oral Therapy. *Can. Pharm. J.* **1959**, *59*, 245.
16. Schlichting, D.A. Ion Exchange Resin Salts for Oral Therapy I. *J. Pharm. Sci.* **1962**, *51*, 134.
17. Raghunathan, Y.; Amsel, L.; Hinsvark, O.; Bryant, W. Sustained-release Drug Delivery System I: Coated Ion-Exchange Resin System for Phenylpropanolamine and Other Drugs. *J. Pharm. Sci.* **1981**, *70*, 379.
18. Borodkin, S. US Patent 3,947,572, 1976.
19. Borodkin, S.; Sundberg, D. US Patent 3,594,470, 1971.
20. Borodkin, S.; Sundberg, D.P. *J. Pharm. Sci.* **1971**, *60*, 1523.
21. Borodkin, S.; Yunker, M.H. Interaction of Amine Drugs with a Polycarboxylic Acid Ion-Exchange Resin. *J. Pharm. Sci.* **1970**, *59*, 481.
22. Cameron, C.W. *The Taste Sense and the Relative Sweetness of Sugars and Other Sweet Substances*, Scientific Report Series No. 9, Sugar Research Foundation, Inc.: New York, 1947.
23. *Flavor Quality: Objective Measurement*; Symposium, Division of Agricultural and Food Chemistry; American Chemical Society: Washington, D.C., 1976.

FURTHER READING

- Flavor: Its Chemical, Behavioral, and Commercial Aspects. In *Proceedings of the Arthur D. Little, Inc., Flavor Symposium, 1977*; Apt, C.M., Ed.; Westview Special Studies in Science and Technology: Boston, MA, 1977.
- Basic Principles of Sensory Evaluation; ASTM Special Technical Publication, No. 433; American Society for Testing and Materials: Philadelphia, PA, 1973.
- Correlation of Subjective-Objective Methods in the Study of Odors and Taste; ASTM Special Technical Publication, No. 440; American Society for Testing and Materials, Philadelphia, PA, 1968.
- Food Acceptance Testing Methodology; Advisory Board on Quartermaster Research and Development; National Academy of Sciences-National Research Council: Chicago, IL, 1954.
- Fenaroli's Handbook of Flavor Ingredients*; Furia, E., Bellanca, N., Eds.; CRC Press: Cleveland, OH, 1971.
- Gorman, W. *Flavor, Taste and the Psychology of Smell*; Charles C. Thomas: Springfield, IL, 1964.
- Moncrieff, R.W. *The Chemical Senses*; Leonard Hill: London, 1967.
- Modifying Bitterness*; Roy, Glenn, Ed.; Technomic Publishing Company, Inc.: Lancaster, PA, 1997.
- Flavor Quality: Objective Measurement. Scanlan, ACS Symposium Series, Scanlan, R.A., Ed.; American Chemical Society: Washington, D.C., 1977; 51.

FLAME PHOTOMETRY

Thomas M. Nowak

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

INTRODUCTION

The use of flame photometry as a quantitative tool can be traced to work by Kirchhoff and Bunsen in the early 1860s (1). Its modern history begins, however, in the 1940s, when instruments became available that successfully addressed the problems of reproducible sample introduction and detection. Flame photometry soon developed into a reliable analytical technique for the determination of several cations of pharmaceutical interest, notably sodium, potassium, and lithium. The technique is useful in the analysis of bulk drugs, dosage forms, and clinical samples such as blood and urine.

This article focuses primarily on “traditional” low-temperature flame photometry. High-temperature flame photometry has evolved into separate techniques, typically identified by their temperature sources (e.g., inductively coupled plasma-atomic emission spectrometry, ICP-AES 2). Some references to other related analytical tools, including high-temperature flame photometry, are made here to establish perspective.

PRINCIPLE OF OPERATION

Flame photometry, as with other spectrophotometric techniques, takes advantage of the unique spectral properties of each element when it is energized about its ground state to an excited state. When the energized electrons return to their ground state, they emit light at discrete wavelengths. The emitted light is optically filtered and photometrically detected. Quantification is based on a calibration curve of emission of intensity versus analyte concentration.

The analyte is introduced as a homogeneous solution into the flame as an aerosol. The flame provides sufficient energy to yield free gaseous atoms in the ground state. The amount of energy provided by the flame additionally allows a small fraction of available atoms to be energized above the ground state. Typically, the flame is produced using a mixture of air and propane (or air and butane), which provides a temperature of approximately 1900°C

(3). In high-temperature flame photometry, the percentage of atoms is increased. In both high- and low-temperature flame photometry, the atoms of interest are energized as a result of the temperature of the flame, furnace, or plasma. This is in contrast to atomic absorption spectrophotometry, which uses light of a discrete wavelength to energize the analyte atoms. The temperature of the flame in atomic absorption is primarily used to yield a sufficient number of free atoms in their ground state.

The general viability of low-temperature flame photometry depends on two factors. First, the alkali and alkaline earth metals of analytical interest (sodium, potassium, lithium, cesium, rubidium, magnesium, calcium, strontium, and barium) reach their excited states at relatively lower temperatures than do most other elements. Second, the emission wavelengths offer enough resolution such that optical filtering can be accomplished at a relatively low cost.

INSTRUMENTATION AVAILABLE

The first commercially available flame photometer was introduced in the 1940s by the Perkin–Elmer Corporation. In 1948, Beckmann Instruments, Inc., introduced a flame attachment that could be used with their popular model D. U. spectrophotometer (1). By the late 1950s, instruments had been developed that used lithium as an internal standard to maximize precision. Autodilution features and microprocessor-controlled operations became widely used options in the 1970s. The most recent significant development was the introduction of cesium as the internal standard, by Instrumentation Laboratory, Inc. (Figs. 1–3). This development makes concurrent lithium determinations more practical.

With the use of fuels that produced hotter flames, earlier flame photometers became useful for analyzing elements beyond the alkali and alkaline earth metals. The development of atomic absorption spectrophotometers in the late 1960s provided the analytical chemist with a better tool for many of these applications. Later developments in high-temperature flame photometry narrowed the analytical applications



Fig. 1 Flame photometer using cesium as the internal standard. (Courtesy of Instrumentation Laboratory, Inc.)

of low-temperature flame photometry even further. The utility of the flame photometer to the clinical chemist, however, was not diminished until the development of ion-selective electrode (ISE) analyzers, which began in the mid-1970s. Although the niche for traditional flame photometers has narrowed, owing to the emergence of other technologies, flame photometry is the method of choice in a variety of applications.

APPLICATIONS

The analysis of clinical samples represents a typical application of flame photometry. Concentrations of sodium, potassium, and lithium in blood and urine are well within instrument working ranges. The specificity of the technique is a distinct advantage. Automated models of flame photometers, available during the past 25 years, are typically designed to serve the needs of the clinical chemist. Instrument calibration protocols are built into instruments to facilitate the timely analysis of sodium, potassium, and lithium in clinical samples.

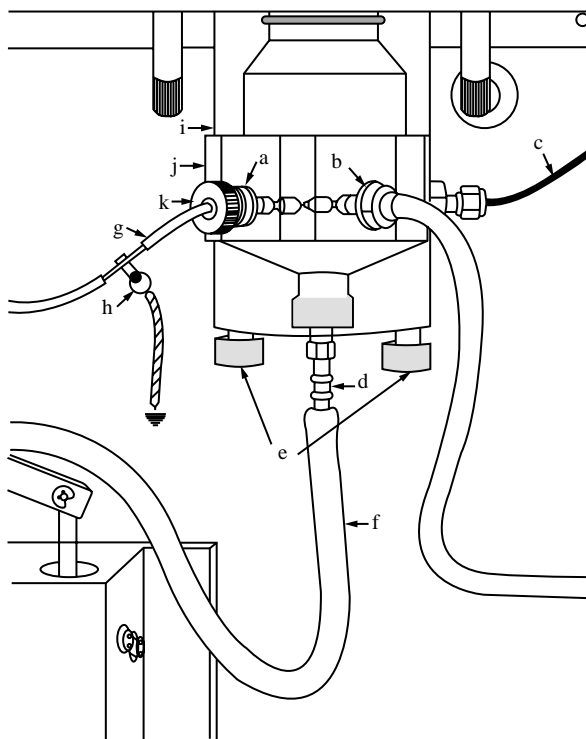


Fig. 2 Atomizer of IL 943 Flame Photometer; a) sample orifice assembly; b) air orifice; c) gas tube assembly; d) atomizer bowl drain; e) atomizer thumb screws; f) U-tube; g) sample injection nozzle tubing; h) ground fitting; i) top atomizer assembly; j) bottom atomizer assembly; k) adjustment for aspiration rate setting. (Courtesy of Instrumentation Laboratory, Inc.)

Other applications in the scope of pharmaceutical analysis include the analysis of sodium and potassium in injectable formulations and of trace amounts in bulk drugs, in dissolution experiments, and in content uniformity testing (4).

There are at least 14 USP, NF, or BP bulk drug monographs that use flame photometry either to control sodium potassium as an impurity or to assay for the primary ion (Table 1) (5, 6). An external standard method procedure is referenced in both the USP and the EP. The USP chapter, "Flame Photometry for Reagents," first appeared in USP XVII (1965) (7).

There are at least 25 USP or BP formulation monographs that use flame photometry to assay ions of interest (Table 2) (8). This technique is applicable to a variety of situations because of the relatively low cost per sample (in analyst time, instrument capital expense, and testing supplies); reasonable precision (typical relative standard deviation values are 0.6% for sodium, 1% for potassium, and 2% for lithium); low sample volume

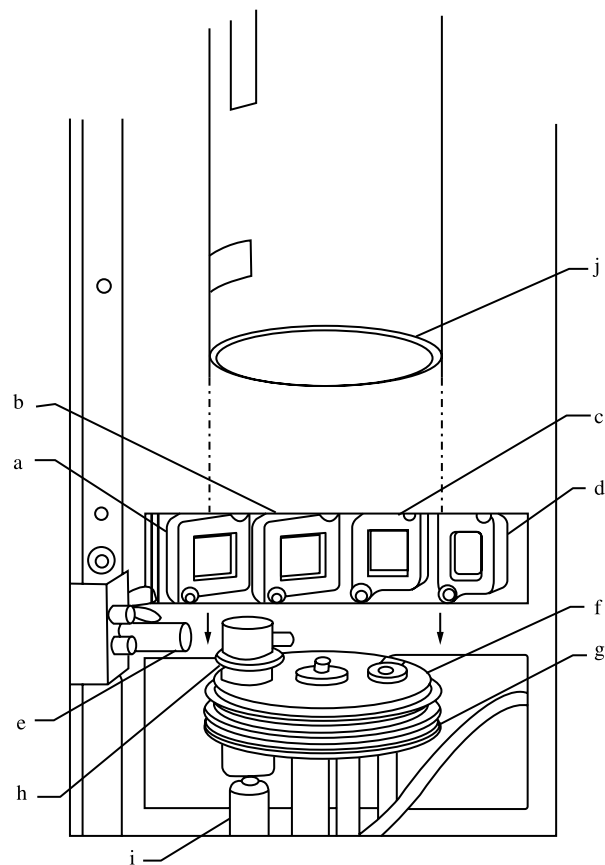


Fig. 3 Flame housing of IL 943 flame photometer; a) sodium filter, 589 nm; b) potassium filter, 776 nm; c) lithium filter, 670 nm; d) cesium filter, 852 nm; e) ignition detector; f) burner assembly; g) rubber gasket; h) spark electrode; i) ignition coil wire; j) chimney. (Courtesy of Instrumentation Laboratory, Inc.)

requirements (as low as 10 µl in some cases); and ease of operation.

COMMON SOURCES OF ANALYTICAL ERROR

In most applications, interferences are rare. Each new matrix requires some investigation, however, because problems can occur, especially when one element is to be determined in the presence of a large excess of another element. In a typical analysis, a liquid sample is diluted in a 1:100 or 1:200 ratio with diluent containing a lithium or cesium salt. Adequate dilution of a sample represents one of the most effective means of overcoming interference problems. Generally, samples are diluted to contain less than 10% by weight of total solids (not including the dilution with internal

Table 1 USP, NF, or EP bulk drug monographs

Monograph	Assay
Cellulose sodium phosphate, USP	11% sodium
Chlorophyllin copper complex sodium, USP	6% sodium
Lithium carbonate, EP	0.03% sodium
Lithium carbonate, USP	0.1% sodium
Lithium citrate, USP	Lithium
Lithium hydroxide, USP	Lithium
Magaldrate, USP	0.11% sodium
Polacrillin potassium, NF	Potassium
Potassium acetate, EP	0.5% sodium
Potassium acetate, USP	0.03% sodium
Potassium chloride, EP	0.1% sodium
Potassium citrate, EP	0.3% sodium
Potassium nitrate, EP	0.1% sodium
Sodium chloride, EP	0.05% potassium

standard). If dilution does not eliminate interference, atomic absorption should be used. It offers advantages for the determination of magnesium, calcium, and zinc in many matrices (9, 10). In some instances, however, sample pretreatment can make it possible to obtain good results with flame photometry (11).

With the development of ion-selection electrode technology (ISE), a means became available to directly measure (no dilution) sodium and potassium in the presence of clinical samples containing a significant amount of protein or lipids. Because of nonaqueous components in the sample matrix, the volume occupied by sodium and potassium ions is less than the total volume of the sample. When using a technique that requires dilution (flame photometry) or utilizes dilution (indirect-ISE), a lower concentration is observed than that obtained with direct-ISE. In as much as the bias can be clinically significant (up to 7% in some instances) it is important that the method used be taken into account (12, 13).

THE FUTURE

Low-temperature flame photometry is a mature technology and not likely to see many significant new applications. Advances in high-temperature flame photometry and atomic absorption techniques appear, well-suited to most of the new challenges in elemental analysis. Indeed, most of the pioneer commercial suppliers of flame photometry instrumentation have abandoned the market. Clinical laboratories are using ion-selective electrode analyzers more and more for

Table 2 USP, NF, or EP bulk drug monographs

Monograph	Use
Anticoagulant citrate phosphate dextrose adenine solution, USP	Sodium
Citric acid, magnesium oxide, and sodium carbonate irrigation, USP	Sodium assay
Half-strength lactated Ringer's and dextrose injection, USP	Sodium and potassium assay
Lactated Ringer's injection, USP	Lithium assay
Lithium carbonate capsules, USP	Lithium assay
Lithium carbonate tablets, USP	Lithium assay
Lithium carbonate extended-release tablets, USP	Lithium assay
Lithium citrate syrup, USP	Lithium assay
Modified lactated Ringer's and dextrose injection, USP	Sodium and potassium assay
Multiple electrolytes and dextrose injection type 4, USP	Sodium assay
Oral rehydration salts, BP	Sodium and potassium assay
Potassium and sodium bicarbonates and citric acid effervescent tablets, USP	Sodium and potassium assay
Potassium chloride and glucose intravenous infusion, BP	Potassium assay
Potassium chloride and sodium chloride intravenous infusion, BP	Sodium and potassium assay
Potassium chloride in sodium chloride injection, USP	Sodium and potassium assay
Potassium chloride in dextrose and sodium chloride injection, USP	Sodium and potassium assay
Potassium chloride, sodium chloride, and glucose intravenous infusion, BP	Sodium and potassium assay
Potassium citrate and citric acid oral solution, USP	Potassium assay
Protein hydrolysate injection, USP	Sodium and potassium assay
Ringer's injection, USP	Sodium and potassium assay
Ringer's irrigation, USP	Sodium and potassium assay
Sodium acetate injection, USP	Sodium assay
Sodium citrate and citric acid oral solution, USP	Sodium assay
Tricitrates oral solution, USP	Sodium and potassium assay
Tromethamine for injection, USP	Sodium and potassium assay

sodium, potassium, and lithium determinations. A core of existing applications, however, will support the viability of low-temperature flame photometry into the foreseeable future.

REFERENCES

- Gardiner, K.W. Flame Photometry. *Physical Methods in Chemical Analysis*; Berl, W.G., Ed.; Academic Press: New York, 1972; III.
- DiPietro, E.S.; Bashor, M.M.; Stroud, P.E.; Smarr, B.J.; Burgess, B.J.; Turner, W.E.; Neese, J.W. *Sci. Total Environ.* **1988**, *74*, 249–262.
- Dean, J.A. *Flame Photometry*; McGraw-Hill: New York, 1960.
- Soltero, R.A. *Clin. Chem.* **1985**, *31* (6), 1094.
- United States Pharmacopeia, The National Formulary*, 24th Rev.; 19th Ed., National Publishing: Philadelphia, 1999.
- European Pharmacopoeia 2000*; 3rd Ed., European Department for the Quality of Medicines: Strasbourg France, 1999.
- The United States Pharmacopeia XVII*; Mack Printing Co: Easton, PA, 1965.
- British Pharmacopoeia, 1998*; Her Majesty's Stationery Office: London, 1999; Version 2.1.
- Williams, W.D. Flame Photometry and Atomic Absorption Spectrophotometry. *Practical Pharmaceutical Chemistry*, 3rd Ed.; Beckett, A.H., Stenlake, J.B., Hayworth, A., Eds.; Athlone Press: University of London, 1976; 2, 297–310.
- Oberdier, J.O. Atomic Absorption Spectrophotometry. *Encyclopedia of Pharmaceutical Technology*; 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 2000; 1.
- Nielsen, I. *J. Clin. Chem. Clin. Biochem.* **1986**, *24* (5), 353–354.
- Burnett, D.; Ayers, G.J.; Rumjen, S.C.; Woods, T.F. *Ann. Clin. Biochem.* **1988**, *25* (1), 102–109.
- Worth, H.G.J. *Ann. Clin. Biochem.* **1985**, *22* (4), 343–350.

FILTERS AND FILTRATION

Maik W. Jornitz

Sartorius Group, Germany

INTRODUCTION

The separative process of filtration is widely used within the biopharmaceutical industry to remove contaminants from liquids, air, and gases, such as particulate matter but especially microorganisms. Microorganism removal is either required to achieve a sterile filtrate or, if the pharmaceutical product is thermally sterilized, to reduce the bioburden and, therefore, avoid elevated levels of endotoxins—the debris of gram-negative organisms (1).

There are many filter configurations within the industry, such as sheet or modular depth filter types for prefiltration purposes, flat filter membranes mainly for microbial detection and specifications, and, most commonly, filter cartridges containing either depth filter fleeces or membrane filters. Such membrane filters are available in a large variety of membrane polymers for different applications. These materials are discussed later in the chapter.

Sterilizing grade membrane filters are defined by the FDA Guideline on Sterile Drug Products Produced by Aseptic Processing (2) by being able to retain 10^7 *Brevundimonas diminuta* (formerly *Pseudomonas diminuta*) organisms per square centimeter of filtration area at a differential pressure of 2 bar (3). Such retention efficiency has to be validated, using the actual drug product and the process parameters, due to the possibility of an effect to the filters compatibility and stability and/or the microorganisms size and survival rate (4, 5). Performing these so-called product bacteria challenge tests became a regulatory demand (6) and, therefore, belong to a standard filter validation. Before these challenge tests can be performed, several parameters, for example, bactericidal effects of the product, have to be evaluated. The recently published PDA Technical Report 26 (7) describes the individual parameters—the possible effects and mechanisms to be used to perform challenge tests. Additionally, the report discusses filtration modes, sterilization, and integrity testing.

FILTER TYPES

One can differentiate filters in different distinctive types, commonly in membrane and depth filters. Depth filters

retain contaminants within the depth of the filter matrix. Contaminants have to move through the tortuous path of the fiber matrix and eventually will collide with a fiber and separate from the medium. Due to the depth retention, such filters have a very high dirt-load capacity and are able to separate a high load of contaminants of different sizes. Depth filters are utilized for coarse particle removal, polishing filtration, and, especially, to protect final membrane filters reverse osmosis or deionizing units. Depth filters can greatly enhance the membrane filter's total throughput capability. Therefore, before utilizing a filtration process in a new application, filterability trials are commonly performed to evaluate the optimal prefilter–final filter combination to achieve the lowest cost per liter ratio and highest yield. Initial filterability tests are done with 47-mm composites (flat filter discs of the filter cartridge device to be used). Having found the optimal combination of prefilter retention to final filters pore size, pleated small-scale devices are used to achieve appropriate filter sizing parameters. Filterability trials are performed with automatic test rigs, which utilize a balance as a load cell to measure the filtrate collected. Commonly, the balance is connected to a computer system, which uses a specific software showing the flow rate, total throughput, and differential pressure graphs and offer a report including such.

As depth filters retain contaminants within the fiber matrix, membrane filters are surface retentive filters and, therefore, have the distinct disadvantage to clog faster. The filter industry, therefore, pleats such membranes to install a higher effective filtration area into a filter device. Still, such effort has its limitation, due to a maximum allowable pleat density. Having reached the limit, the only option is the use of prefilters or membranes of different pore sizes to gain a fractionate retention and, therefore, a prolonged lifetime of the filter. Some membrane filter configurations have such membrane or depth filter prefilters built into the filter cartridge. This is convenient for the filter user in respect to lowered hardware costs. Additional prefilter housings are not necessary. In comparison to depth filters, membrane filters have a narrow pore size distribution, which results in a by far sharper retention rate. Pore size ratings are facilitated to differentiate membrane filters

and the performance of such. Commonly, a sterilizing grade filter is labeled 0.2 μm when it retains 10^7 B.diminuta/cm². Another advantage and necessity of membrane filters is the fact that these are integrity testable. Therefore, flaws or defects can be detected, which is critical, due to the function of membrane filters, mainly in separating microorganisms from pharmaceutical solutions.

Membrane filters are made in a wide variety of pore sizes (Fig. 1). The effective pore size for membranes vary, and membranes can be used in reverse osmosis (RO), nanofiltration (NF), ultrafiltration (UF), and microfiltration (MF). RO membranes are widely used in water treatment to remove ionic contaminations from the water. These membranes have an extreme small pore size and, therefore, require excellent pretreatment steps to reduce any fouling or scaling of the membrane, which would reduce the service lifetime. RO membranes are used by extensive pressures on the upstream side of the filter membrane to force the liquids through the pores.

The retention ratings of UF filters are also not measured in pore size but rather in MWCO (molecular weight cut-off), i.e., the molecular weight of the substance to be retained. UF filter systems are most often used in cross-flow (tangential flow) systems. The feed stream is directed over the actual membrane to diminish blockage of the membrane. Depending on the pressure conditions, the fluid (permeate) penetrates through the membrane, whereby the remaining fluid is recirculated (retentate). UF filter systems find applications in concentration, diafiltration, and removal steps within pharmaceutical downstream processing. MF can be used as dead-end filtration (the feed is directed to the membrane, resulting into a filtrate, separated from the contaminant) or tangential flow mode. The tangential flow characteristic for MF is commonly used for cell or cell debris removal in downstream processing. MF membranes typically differ from UF membranes in

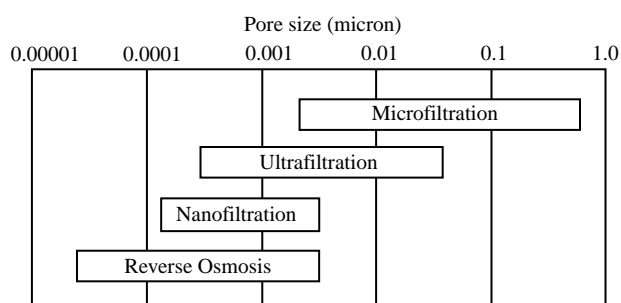


Fig. 1 Typical pore sizes for membranes used in reverse osmosis, nanofiltration, ultrafiltration, and microfiltration.

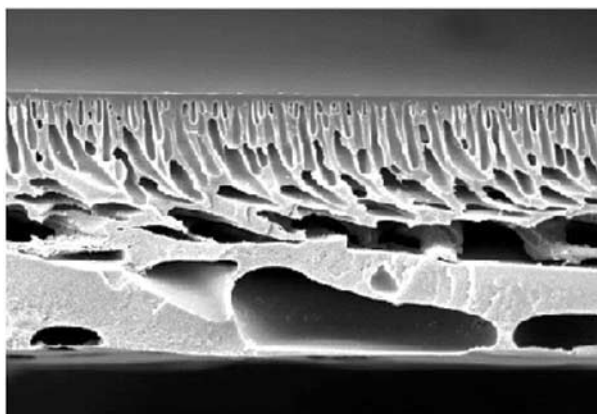


Fig. 2 Skin layer structure of a UF membrane. (Courtesy of Sartorius AG.)

the morphology of the membrane's cross-sectional cut. The symmetry of sterilizing-grade microfilters usually ranges from being uniform to being slightly asymmetric. Ultrafilters, on the other hand, are highly asymmetric, with the rejecting layer consisting of a tight skin (0.5–10 to μm thick) supported by a thick spongy structure of a much larger pore size (Fig. 2).

MF is used in a large variety of filtration applications, from fine cut prefiltration to sterilizing grade filtration in aseptic processing. Often sterilizing grade filters are the terminal step before filling or final processing of the drug product. MF is available for air and gas applications and liquid clarification or sterilization. For the different applications, specific membrane configurations and materials have been developed.

FILTER MATERIALS

There are a variety of different depth filter and membrane filter materials used within the pharmaceutical processes. Depth filter are fibrous materials: for example, Polypropylene, Borosilicate, or Glassfibre fleeces (Fig. 3). Borosilicate and Glassfibre materials are highly adsorptive and commonly used to remove colloidal substances, like iron oxide from water or colloidal haze from sugar solutions.

Prefilters can also contain membranes instead of fibrous depth filter material. Such membrane material are commonly mixesters of cellulose or pure cellulose acetate. The cellulose mixester filter material contains a high degree of cellulose nitrate, which again is highly adsorptive. Such prefilters have a very sharp retention rating and, therefore, are used in applications in which the



Fig. 3 SEM of the random fiber matrix of a depth filter. (Courtesy of Sartorius AG.)

contaminant has a narrow size distribution and/or a sterilizing grade filter has to be protected, due to the fact that such a filter cannot be changed during the filtration process. Membrane prefilters, when blocked, can be

exchanged during the filtration process, due to the final filter downstream.

Most commercial UF and NF membranes and many MF membranes are made by the phase-inversion process, where a polymer is dissolved in an appropriate solvent along with appropriate pore-forming chemical agents. The polymer solution is cast into a film, either on a backing or freestanding, and then the film is immersed in a nonsolvent solution that causes precipitation of the polymer. Such membranes are Polyamides, such as Nylon, Polyethersulfon (PESU), or Polyvinylidene fluoride (PVDF). Cellulosic membranes, such as cellulose nitrate, acetate, or regenerated cellulose, are casted as a cellulose–water–solvent mixture onto a belt and transported through heated tunnels. The resulting evaporation process produces the porous structure of the membrane seen in Fig. 4.

Other techniques for membrane formation include stretching the polymeric film, commonly Polytetrafluoroethylene (PTFE), while it is still in a flexible state and then annealing the membrane to “lock in” and strengthen the

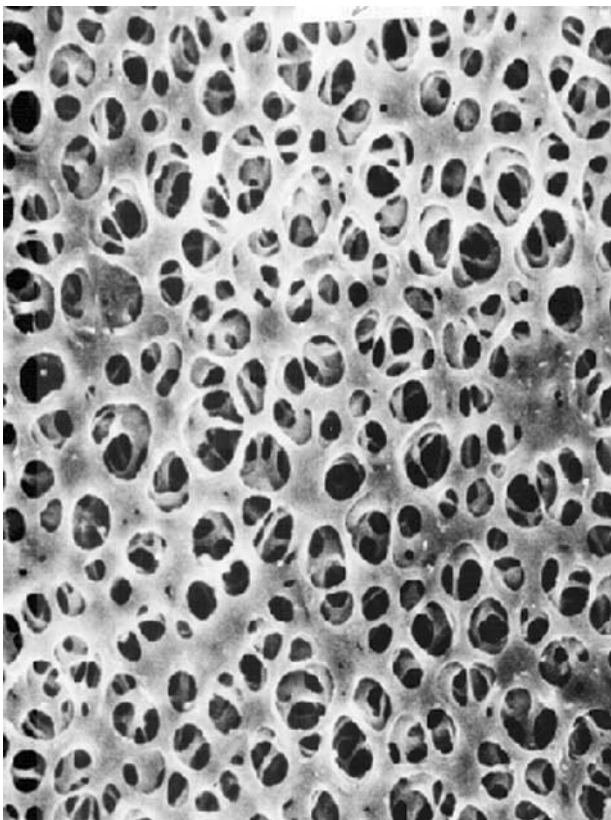


Fig. 4 Porous structure of Celluloseacetate. (Courtesy of Sartorius AG.)

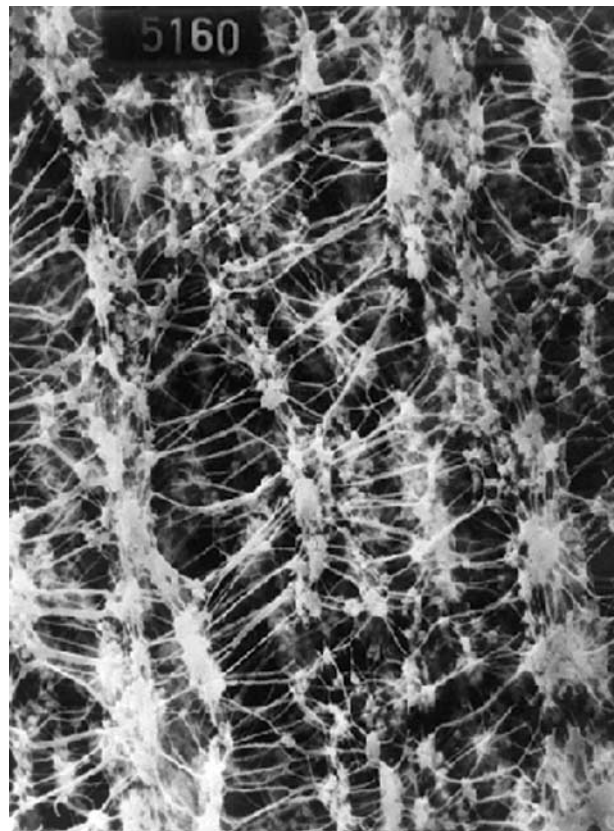


Fig. 5 PTFE membrane structure. (Courtesy of Sartorius AG.)

pores in the stretched membrane. The stretching process results into a distinctive membrane structure of PTFE nodes, which are interconnected by fibrils, (Fig. 5).

PTFE membranes are highly hydrophobic and, therefore, are used as air filters. Air filters have to be highly hydrophobic to avoid water blockage due to moisture or condensate, especially after steam sterilization of these filters. Water blockage could be detrimental, if the filter is, for example, used in a tank venting application to overcome condensation vacuum of a nonvacuum resistant tank. If the filter would not allow a free flow of air into the tank, it may implode. Therefore, vent filters for this application have to be chosen and sized with care. PTFE membranes are also highly mechanical and thermal resistant, which is required, because such filters are used over several months, withstanding multiple steam-sterilization cycles. Especially in large-scale fermentation, these filters are used over several months, avoiding unwanted infections of the fermenter's or bioreactor's cell line.

Finally, track-etched MF membranes are made from polymers, such as polycarbonate and polyester, wherein electrons are bombarded onto the polymeric surface. This bombardment results in "sensitized tracks," where chemical bonds in the polymeric backbone are broken. Subsequently, the irradiated film is placed in an etching bath (such as a basic solution), in which the damaged polymer in the tracks is preferentially etched from the film, thereby forming cylindrical pores. The residence time in the irradiator determines pore density, and residence time in the etching bath determines pore size. Membranes made by this process generally have cylindrical pores with very narrow pore-size distribution, albeit with low overall porosity. Furthermore, there always is the risk of a double hit, i.e., the etched pore becomes wider and could result in particulate penetration. Such filter membranes are often used in the electronic industry to filter high-purity water.

Table 1 lists the different membrane polymers available and the advantages and disadvantages, which depend on the properties of the polymer. The table shows that there is no such thing as a membrane polymer for every application. Therefore, filter membranes and the filter performance have to be tested before choosing the appropriate filter element.

FILTER CONSTRUCTION

Filters are available in several constructions, effective filtration areas, and configurations. Depending on the

individual process, the filter construction and setup will be chosen to fit its purpose best. Most commonly used for RO filters are tubular devices, so-called spiral wound modules due to the spiral configuration of the membrane within the support construction of such device. UF systems can be found as a spiral wound module, a hollow fiber, or a cassette device. The choice of the individual construction depends on the requirements and purposes towards the UF device. Similar to the different membrane materials, UF device construction has to be evaluated in the specific applications to reach an optimal functioning of the unit. Microfilters and depth filters can be lenticular modules or sheets but are mainly cylindrical filter elements of various sizes and filtration areas, from very small scale of 300 cm² to large scale devices of 36 m². A 10-inch high cylindrical filter element can be seen in Fig. 6.

These filter elements are installed into stainless steel filter housings by pushing the double O-ring cartridge adapter into the housing base plate recess. The filter housing is then assembled and connected, and the filter is flushed with water and steam-sterilized, either by in-line steaming or autoclaved. If filter housings are not available or not preferred, disposable filters can be used. The filter element is welded into a plastic housing, usually Polypropylene, and after every filtration process, discarded. The advantage of such a disposable filter device is the reduced cleaning validation effort, and the user does not come in contact with the filtered product. Such disposable filters can be autoclaved, but not in-line steam-sterilized, due to the pressure-temperature ratio of the housing polymer. Most often, such disposable filters are used for scale-up filtration tests, due to the ease of use and the availability of a band of effective filtration areas.

FILTER VALIDATION

Pharmaceutical processes are validated processes to assure a reproducible product within set specifications. Equally important is the validation of the filters used within the process, especially the sterilizing grade filters, which, often enough, are used before filling or the final processing of the drug product. In its Guideline on General Principles of Process Validation, 1985 (8), and Guideline on Sterile Drug Products Produced by Aseptic Processing, 1987 (2), the FDA makes plain that the validation of sterile processes is required by the manufacturers of sterile products.

Sterilizing grade filters are determined by the bacteria challenge test. This test is performed under strict

Table 1 Advantages and disadvantages of various membrane polymers

Membrane material	Advantages	Disadvantages
Cellulose acetate	Very low nonspecific adsorption (nonfouling) High flow rates and total throughputs	Limited pH compatibility Not dry autoclavable
Cellulose nitrate (nitrocellulose)	Good flow rate and total throughputs	High nonspecific adsorption Limited pH compatibility Not dry autoclavable
Regenerated cellulose	Very low nonspecific adsorption (nonfouling) Very high flow rates and total throughputs	Limited pH compatibility Not dry autoclavable
Modified regenerated cellulose	Very low nonspecific adsorption (nonfouling) Moderate flow rates and total throughputs, especially with difficult to filter solutions Broad pH compatibility	Ultrafilters not dry autoclavable
Nylon 66	Good solvent compatibility Good mechanical strength Broad pH compatibility Dry autoclavable	High nonspecific protein adsorption Low hot-water resistance Moderate flow rate and total throughput
Polycarbonate	Good chemical compatibility	Moderate flow rates Low total throughputs Difficult to produce
Polyethersulfone	High flow rates and total throughputs Broad pH compatibility	Moderate-to-low nonspecific adsorption, depending on surface modifications. Limited solvent compatibility

(Continued)

Table 1 Advantages and disadvantages of various membrane polymers (*Continued*)

Membrane material	Advantages	Disadvantages
Polypropylene	Excellent chemical resistance High mechanical resistance	Hydrophobic material High nonspecific adsorption due to hydrophobic interactions
Polysulfone	High flow rates and total throughputs Broad pH compatibility	Moderate-to-high nonspecific adsorption Limited solvent compatibility
Polytetrafluoro-ethylene	Excellent chemical resistance High mechanical resistance	Hydrophobic material High nonspecific adsorption due to hydrophobic interactions High-cost filter material
Polyvinylidene-difluoride	Low nonspecific adsorption Dry autoclavable Good solvent compatibility	Moderate flow rate and total throughput Hydrophobic base, made hydrophilic by chemical surface treatment; may lose hydrophilic modification due to chemical attack High-cost filter material

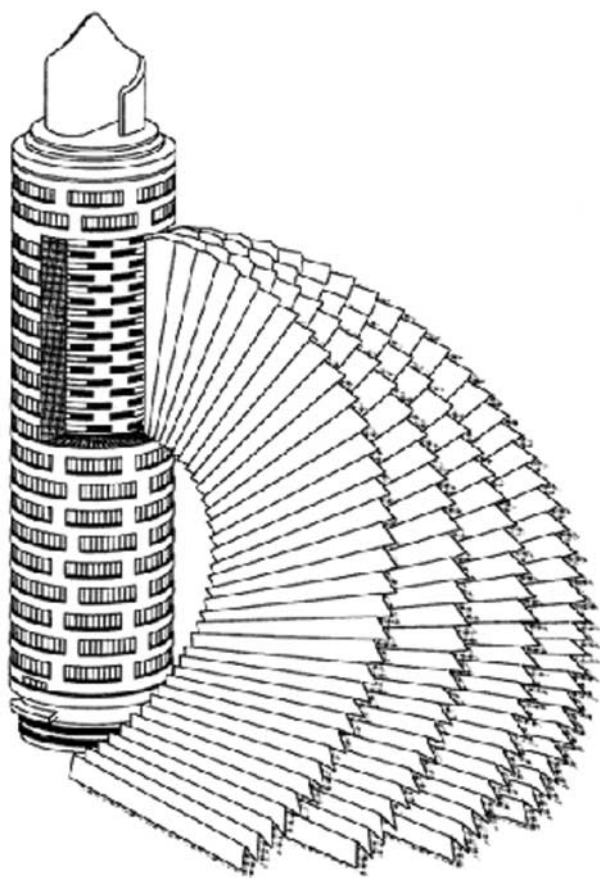


Fig. 6 10-inch standard filter element with pleated membrane and protection fleeces. (Courtesy of Sartorius AG.)

parameters and a defined solution (ASTM F 838-83) (3). In any case, the FDA nowadays also requires evidence that the sterilizing grade filter will create a sterile filtration, no matter the process, fluid or bioburden, found (6, 9). This means that bacteria challenge tests have to be performed with the actual drug product, bioburden, if different or known to be smaller than *B. diminuta* and the process parameters. The reason for the requirement of a product bacteria challenge test is threefold. First, the influence of the product and process parameters to the microorganism has to be tested. There may be cases of either shrinkage of organisms due to a higher osmolarity of the product or prolonged processing times. Second, the filter's compatibility with the product and the parameters has to be tested. The filter should not show any sign of degradation due to the product filtered. Additionally, rest assurance is required that the filter used will withstand the process parameters; e.g., pressure pulses, if happening, should not influence the filter's performance. Third, there are two

separation mechanisms involved in liquid filtration: sieve retention and retention by adsorptive sequestration (1, 8, 10–12). In sieve retention, the smallest particle or organism size is retained by the biggest pore within the membrane structure. The contaminant will be retained, no matter the process parameters. This is the ideal. Retention by adsorptive sequestration depends on the filtration conditions. Contaminants smaller than the actual pore size penetrate such and may be captured by adsorptive attachment to the pore wall. This effect is enhanced using highly adsorptive filter materials, for example, Glassfibre as a prefilter or Polyamide as a membrane. Nevertheless, certain liquid properties can minimize the adsorptive effect, which could mean penetration of organisms. Whether the fluid has such properties and will lower the effect of adsorptive sequestration and may eventually cause penetration has to be evaluated in specific product bacteria challenge tests. Table 2 shows the advantages and disadvantage of both separation mechanisms.

Before performing a product bacteria challenge test, it has to be assured that the liquid product does not have any detrimental, bactericidal or bacteriostatic, effects on the challenge organisms. This is done utilizing viability tests. The organism is inoculated into the product to be filtered at a certain bioburden level. At specified times, the log value of this bioburden is tested. If the bioburden is reduced due to the fluid properties, a different bacteria challenge test mode becomes applicable (7). If the reduction is a slow process, the challenge test will be performed with a higher bioburden, bearing in mind that the challenge level has to reach 10^7 per square centimeter at the end of the processing time. If the mortality rate is too high, the toxic substance is either removed or product properties are changed. This challenge fluid is called a placebo. Another methodology would circulate the fluid product through the filter at the specific process parameters as long as the actual processing time would be. Afterwards, the filter is flushed extensively with water and the challenge test, as described in ASTM F838-38, performed. Nevertheless, such a challenge test procedure would be more or less a filter compatibility test.

Besides the product bacteria challenge test, tests of extractable substances or particulate releases have to be performed (7, 8, 13). Extractable measurements and the resulting data are available from filter manufacturers for the individual filters. Nevertheless, depending on the process conditions and the solvents used, explicit extractable tests have to be performed. These tests are commonly done only with the solvent used with the drug product but not with the drug ingredients themselves, because the drug product usually covers any extractables during measurement. Such

Table 2 Advantages and disadvantages of separation mechanisms

Retention mechanism	Advantages	Disadvantages
Sieve retention	Reliable at worst case product properties Reliable separation even at high flows and pressure conditions Blockage, i.e., exhaustion, can be determined No unspecific adsorption, minimal loss of desired product, and little adsorptive fouling	Retentive only at the specific pore size rating
Adsorptive sequestration	It is possible to retain particles smaller than the filter's indicated pore size Separation of colloidal substances is possible In some case, pyrogens can be removed	Highly influenced by product specific properties Separated particles can be shed by varying process conditions Saturation of the active sites cannot be determined, no warning Unspecific adsorption will result in product losses and fouling Lower reliability in terms of absolute separation

tests are conducted by the validation services of the filter manufacturers using sophisticated separation and detection methodologies, as GC-MS, FTIR, and RP-HPLC. These methodologies are required, due to the fact that the individual components possibly released from the filter have to be identified and quantified. Elaborate studies, performed by filter manufacturers, showed that there is neither a release of high quantities of extractables (the range is ppb to max ppm per 10-inch element) nor have toxic substances been found (13).

Particulates are critical in sterile filtration, specifically of injectables. The USP 24 (*United States Pharmacopoeia*) and BP (*British Pharmacopoeia*) quote specific limits of particulate level contaminations for defined particle sizes. These limits have to be kept and, therefore, the particulate release of sterilizing grade filters has to meet these requirements. Filters are routinely tested by evaluating the filtrate with laser particle counters. Such tests are also performed with the actual product under process conditions to prove that the product, but especially process conditions, do not result in an increased level of particulates within the filtrate.

Additionally, with certain products, loss of yield or product ingredients due to adsorption shall be determined (14, 15). For example, preservatives, like benzalkoniumchloride or chlorhexadine, can be adsorbed by specific filter membranes. Such membranes need to be saturated by the preservative to avoid preservative

loss within the actual product. This preservative loss, e.g., in contact lens solutions, can be detrimental, due to long-term use of such solutions. Similarly, problematic would be the adsorption of required proteins within a biological solution. To optimize the yield of such proteins within an application, adsorption trials have to be performed to find the optimal membrane material and filter construction.

Cases that use the actual product as a wetting agent to perform integrity tests require the evaluation of product integrity test limits (7, 17). The product can have an influence on the measured integrity test values due to surface tension, or solubility. A lower surface tension, for example, would shift the bubble point value to a lower pressure and could result in a false negative test. The solubility of gas into the product could be reduced, which could result in false positive diffusive flow tests. Therefore, a correlation of the product as a wetting agent to the, water wet values has to be done, according to standards set by the manufacturer of the filter. This correlation is carried out by using a minimum of three filters of three filter lots. Depending on the product and its variability, one or three product lots are used to perform the correlation. The accuracy of such a correlation is enhanced by automatic integrity test machines. These test machines measure with highest accuracy and sensitivity and do not rely on human judgement, as with a manual test (7). Multipoint

diffusion testing offers the ability to test the filter's performance and, especially, to plot the entire diffusive flow graph through the bubble point. The individual graphs for a water-wet integrity test can now be compared to the product wet test and a possible shift evaluated. Furthermore, the multipoint diffusion test enables the establishment of an improved statistical base to determine the product wet versus water-wet limits (16, 17).

FILTER INTEGRITY TESTING

Sterilizing grade filters require testing to assure the filters are integral and fulfill their purpose. Such filter tests are called integrity tests and are performed before and after the filtration process. Sterilizing grade filtration would not be admitted to a process if the filter would not be integrity tested in the course of the process. This fact is also established in several guidelines, recommending the use of integrity testing, pre- and post-filtration. This is not only valid for liquid but also for air filters.

Examples of such guidelines are:

1. *FDA Guideline on Sterile Drug Products Produced by Aseptic Processing* (1987): Normally, integrity testing of the filter is performed after the filter unit is assembled and prior to use. More importantly however, such testing should be conducted after the filter is used in order to detect any filter leaks or perforations that may have occurred during filtration.
2. *The Guide to Inspections of High Purity Water Systems, Guide to Inspections of Lyophilization of Parenterals*, and also the CGMP document 212.721 Filters state the following:
 - a. The integrity of all air filters shall be verified upon installation and maintained throughout use. A written testing program adequate to monitor integrity of filters shall be established and followed. Results shall be recorded and maintained as specified in 212.83.
 - b. Solution filters shall be sterilized and installed aseptically. The integrity of solution filters shall be verified by an appropriate test, both prior to any large-volume parenteral solution filtering operation and at the conclusion of such operation before the filters are discarded. If the filter assembly fails the test at the conclusion of the filtering operation, all materials filtered through it during that filtering operation should be rejected. Rejected materials may be refiltered using filters

whose integrity has been verified provided that the additional time required for refiltration does not result in a total process time that exceeds the limitations specified in 212.111. Results of each test shall be recorded and maintained as required in 212.188(a)

3. ISO 13408-1 *First Edition, 1998-08-1, Aseptic Processing of Health Care Products*, Part 1: General requirements: Section 17.11.1 Investigation, m. pre- and post-filter integrity test data, and/or filter housing assembly:

20.3.1. A validated physical integrity test of a process filter shall be conducted after use without disturbing the filter housing assembly. Filter manufacturer's testing instructions or recommendations may be used as a basis for a validated method. Physical integrity testing of a process filter should be conducted before use where process conditions permit. "Diffusive Flow," "Pressure Hold," and "Bubble Point" are acceptable physical integrity tests.

20.3.2. The ability of the filter or housing to maintain integrity in response to sterilization and gas or liquid flow (including pressure surges and flow variations) shall be determined.

4. USP 23, 1995, P. 1979. *Guide to Good Pharmaceutical Manufacturing Practice* (Orange Guide, U.K., 1983):

PDA (Parenteral Drug Association), Technical Report No. 26, *Sterilizing Filtration of Liquids* (March 1998):

Integrity tests, such as the diffusive flow, pressure hold, bubble point, or water intrusion tests, are nondestructive tests, which are correlated to the destructive bacteria challenge test with $10^7/\text{cm}^2$ B. diminuta (1, 8). Derived from these challenge tests, specific integrity test limits are established, which are described and documented within the filter manufacturers' literature. The limits are water-based; i.e., the integrity test correlations are performed using water as a wetting medium. If a different wetting fluid, such as a filter or membrane configuration, is used, the integrity test limits may vary. Integrity test measurements depend on the surface area of the filter, the polymer of the membrane, the wetting fluid, the pore size of the membrane, and the gas used to perform the test. Wetting fluids may have different surface tensions, which can depress or elevate the bubble point pressure. The use of different test gases may elevate the diffusive gas flow. Therefore, appropriate filter validation has to be

established to determine the appropriate integrity test limits for the individual process.

Bubble Point Test

Microporous membranes will fill their pores with wetting fluids by imbibing that fluid in accordance with the laws of capillary rise. The retained fluid can be forced from the filter pores by air pressure applied from the upstream side. The pressure is increased gradually in increments. At a certain pressure level, liquid will be forced first from the set of largest pores, in keeping with the inverse relationship of the applied air pressure P and the diameter of the pore, d , described in the bubble point equation:

$$P = \frac{4\gamma \cos \theta}{d} \quad (1)$$

where γ is the surface tension of the fluid, θ is the wetting angle, P is the upstream pressure at which the largest pore will be freed of liquid, and d is the diameter of the largest pore.

When the wetting fluid is expelled from the largest pore, a bulk gas flow will be detected on the downstream side of the filter system (Fig. 7). The bubble point measurement determines the pore size of the filter membrane, i.e., the larger the pore the lower the bubble point pressure. Therefore, filter manufacturers specify the

bubble point limits as the minimum allowable bubble point. During an integrity test, the bubble point test has to exceed the set minimum bubble point.

Diffusion Test

A completely wetted filter membrane provides a liquid layer across which, when a differential pressure is applied, the diffusive airflow occurs in accordance with Fick's law of diffusion (Fig. 8). This pressure is called test pressure and commonly specified at 80% of the bubble point pressure. In an experimental elucidation of the factors involved in the process, Reti (18) simplified the integrated form of Fick's law to read as follows:

$$N = \frac{DH(p_1 - p_2)}{L} \cdot \rho$$

where N is the permeation rate (moles of gas per unit time), D is the diffusivity of the gas in the liquid, H is the solubility coefficient of the gas, L is the thickness of liquid in the membrane (equal to the membrane thickness if the membrane pores are completely filled with liquid), P ($p_1 - p_2$) is the differential pressure, and ρ is the void volume of the membrane, its membrane porosity, commonly around 80%.

The size of pores only enter indirectly into the equation; in their combination, they comprise L , the thickness of the

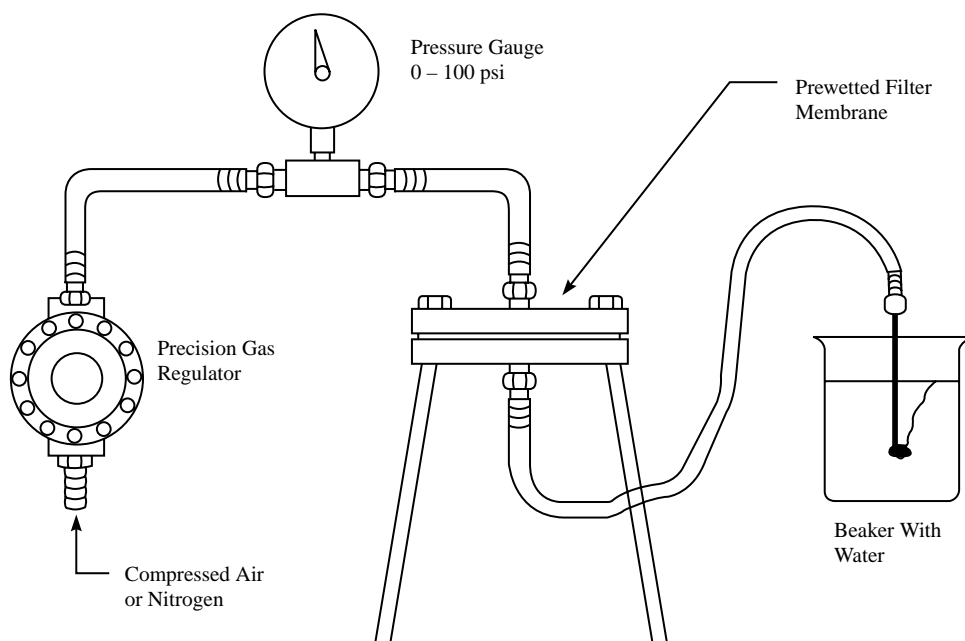


Fig. 7 Manual bubble point test set-up. (Reprinted from Technical Report No. 26, Sterilizing Filtration of Liquids © 1998 by PDA.)

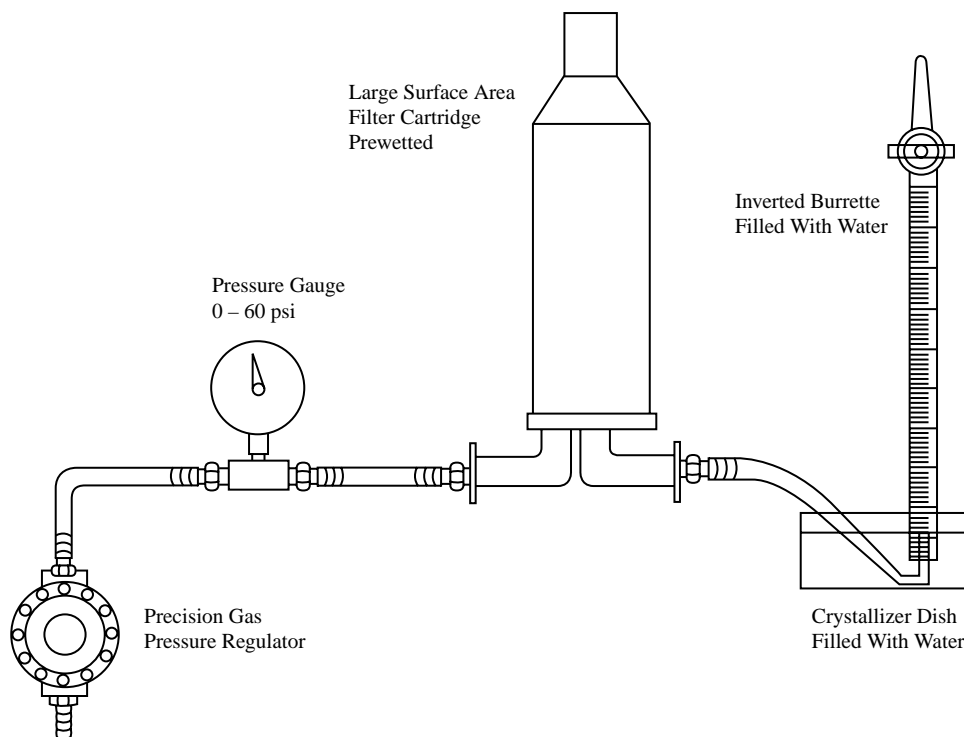


Fig. 8 Manual diffusive flow test set-up. (Reprinted from Technical Report No. 26, Sterilizing Filtration of Liquids © 1998 by PDA.)

liquid layer, the membrane being some 80% porous. The critical measurement of a flaw is the thickness of the liquid layer. Therefore, a flaw or an oversized pore would be measured by the thinning of the liquid layer due to the elevated test pressure on the upstream side. The pore or defect may not be large enough that the bubble point comes into effect, but the test pressure thins the liquid layer enough to result into an elevated gas flow. Therefore, filter manufacturers specify the diffusive flow integrity test limits as maximum allowable diffusion value. The larger the flaw or a combination of flaw, the higher the diffusive flow.

Pressure Hold Test

The pressure hold test is a variant of the diffusive airflow test (19). The test set-up is arranged as in the diffusion test except that when the stipulated applied pressure is reached, the pressure source is valved off (Fig. 9). The decay of pressure within the holder is then observed as a function of time, using a precision pressure gauge or pressure transducer.

The decrease in pressure can come from two sources: 1) the diffusive loss across the wetted filter. Because the upstream side pressure in the holder is constant, it decreases progressively as all the while diffusion takes

place through the wetted membrane; and 2) the source of pressure decay could be a leak of the filter system set-up.

An important influence on the measurement of the pressure hold test is the upstream air volume within the filter system. This volume has to be determined first to specify the maximum allowable pressure drop value. The larger the upstream volume, the lower will the pressure drop be. The smaller the upstream volume, the larger the pressure drop. This also means an increase in the sensitivity of the test, and also an increase of temperature influences, if changes occur. Filter manufacturers specify maximum allowable pressure drop values.

Water Intrusion Test

The water intrusion test is used for hydrophobic vent and air membrane filters only (20–23). The upstream side of the hydrophobic filter cartridge housing is flooded with water. The water will not flow through the hydrophobic membrane. Air or nitrogen gas pressure is then applied to the upstream side of the filter housing above the water level to a defined test pressure. This is done by way of an automatic integrity tester. A period of pressure stabilization takes place over time frame, by the filter manufacturer's recommendation, during which the

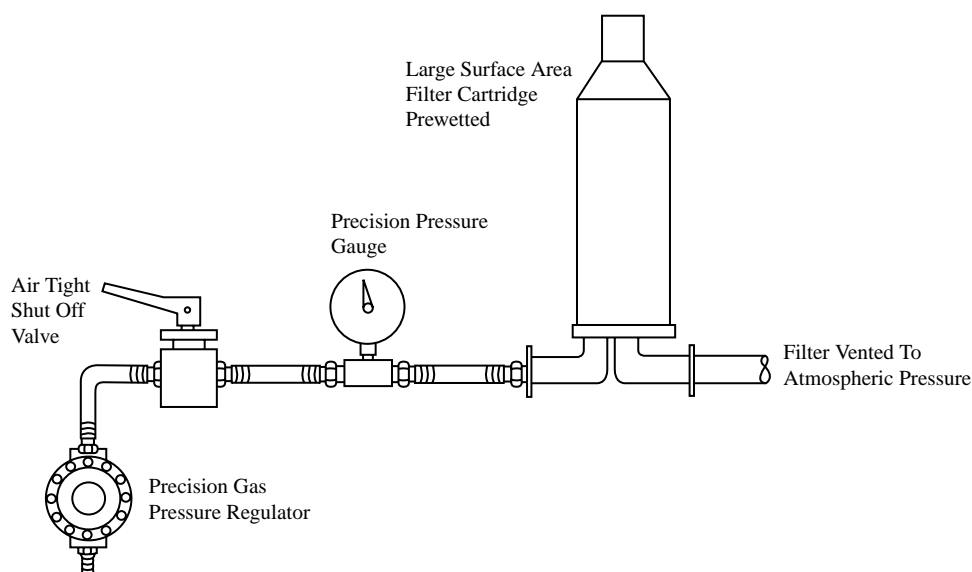


Fig. 9 Manual pressure-hold test set-up. (Reprinted from Technical Report No. 26, Sterilizing Filtration of Liquids © 1998 by PDA.)

cartridge pleats adjust their positions under imposed pressures. After the pressure drop thus occasioned stabilizes, the test time starts, and any further pressure drop in the upstream pressurized gas volume, as measured by the automatic tester, signifies a beginning of water intrusion into the largest (hydrophobic) pores, water being incompressible. The automated integrity tester is sensitive enough to detect the pressure drop. This measured pressure drop is converted into a measured intrusion value, which is compared to a set intrusion limit, which has been correlated to the bacteria challenge test. As with the diffusive flow test, filter manufacturers specify a maximum allowable water intrusion value. Above this value, a hydrophobic membrane filter is classified as nonintegral.

REFERENCES

1. *Filtration in the Biopharmaceutical Industry*; Meltzer, T.H., Jornitz, M.W., Eds.; Marcel Dekker, Inc.: New York, 1998.
2. FDA, Center for Drugs and Biologics and Office of Regulatory Affairs. *Guideline on Sterile Drug Products Produced by Aseptic Processing*; 1987.
3. *Standard Test Method for Determining Bacterial Retention of Membrane Filters Utilized for Liquid Filtration, Standard F838-83*; Revised 1988 American Society for Testing and Materials (ASTM), 1983.
4. Meltzer, T.H.; Jornitz, M.W.; Mittelman, M.W. Surrogate Solution Attributes and Use Conditions: Effects on Bacterial Cell Size and Surface Charges Relevant to Filter Validation Studies. *PDA J. Sci. Technol.* **Jan/Feb 1998**, 52 (1).
5. Levy, R.V. The Effect of PH, Viscosity, and Additives on the Bacterial Retention of Membrane Filters Challenged With *Pseudomonas Diminuta*. In *Fluid Filtration: Liquid*; American Society for Testing and Materials (ASTM): Washington DC, 1987; II.
6. *Validation of Microbial Retention of Sterilizing Filters*; PDA Special Scientific Forum: Bethesda, MD, July 1995; 12–13.
7. Technical Report No. 26, Sterilizing Filtration of Liquids. *PDA J. Pharm. Sci. Technol.* **1998**, 52 (S1).
8. FDA, Center for Drugs and Biologics and Office of Regulatory Affairs. *Guideline on General Principles of Process Validation*; Washington, DC, September 1985.
9. Jornitz, M.W.; Meltzer, T.H. *Sterile Filtration—A Practical Approach*; Marcel Dekker, Inc.: New York, 2001.
10. Tanny, G.B.; Strong, D.K.; Presswood, W.G.; Meltzer, T.H. Adsorptive Retention of *Pseudomonas Diminuta* by Membrane Filters. *J. Parent. Drug Assoc.* **1979**, 33, 40–51.
11. Emory, S.F.; Koga, Y.; Azuma, N.; Matsumoto, K. The Effects of Surfactant Types and Latex-Particle Feed Concentration on Membrane Retention. *Ultrapure Water* **1993**, 10 (2), 41–44.
12. Osumi, M.; Yamada, N.; Toya, M. Bacterial Retention Mechanisms of Membrane Filters. *PDA J. Pharm. Sci. Technol.* **1996**, 50 (1), 30–34.
13. Reif, O.W.; Sölkner, P.; Rupp, J. Analysis and Evaluation of Filter Cartridge Extractables for Validation in Pharmaceutical Downstream Processing. *PDA J. Pharm. Sci. Technol.* **1996**, 50, 399–410.
14. Hawker, J.; Hawker, L.M. Protein Losses During Sterilization by Filtration. *Lab. Practises* **1975**, 24, 805–814.

15. Brose, D.J.; Henricksen, G. A Quantitative Analysis of Preservative Adsorption on Microfiltration Membranes. *Pharm. Tech Europe* **1994**, 42–49.
16. Waibel, P.J.; Jornitz, M.W.; Meltzer, T.H. Diffusive Airflow Integrity Testing. *PDA J. Pharm. Sci. Tech.* **1996**, 50 (5), 311–316.
17. Jornitz, M.W.; Brose, D.J.; Meltzer, T.H. Experimental Evaluations of Diffusive Airflow Integrity Testing. *PDA J. Parenter. Sci. Technol.* **1998**, .
18. Reti, A.R. An Assessment of Test Criteria in Evaluating the Performance and Integrity of Sterilizing Filters. *Bull. J. Parenter. Drug Assoc.* **1977**, 31 (4), 187–194.
19. Madsen, R.E., Jr.; Meltzer, T.H. An Interpretation of the Pharmaceutical Industry Survey of Current Sterile Filtration Practices. *PDA J. Pharm. Sci. Technol.* **1998**, 52 (6), 337–339.
20. Jornitz, M.W.; Waibel, P.J.; Meltzer, T.H. The Filter Integrity Correlations. *Ultrapure Water* **1994 Oct.**, 59–63.
21. Meltzer, T.H.; Jornitz, M.W.; Waibel, P.J. The Hydrophobic Air Filter and the Water Intrusion Test. *Pharm. Tech.* **1994**, 18 (9), 76–87.
22. Tarry, S.W.; Henricksen, G.; Prashad, M.; Troeger, H. Integrity Testing of EPTFE Membrane Filter Vents. *Ultrapure Water* **1993**, 10 (8), 23–30.
23. Tingley, S.; Emory, S.; Walker, S.; Yamada, S. Water-Flow Integrity Testing: A Viable and Validatable Alternative to Alcohol Testing. *Pharm. Tech.* **1995**, 19 (10), 138–146.

Film Coating of Oral Solid Dosage Forms

Linda A. Felton

*University of New Mexico College of Pharmacy, Albuquerque,
New Mexico, U.S.A.*

INTRODUCTION

Polymeric materials have been used to coat pharmaceutical solid dosage forms for decorative, protective, and functional purposes. These thin polymeric film coatings may improve the esthetic appearance and provide easy identification of drug products. Polymer coatings may be used to enhance the chemical stability of the drug in a dosage form by providing a physical barrier to environmental storage conditions.^[1,2] Application of a polymeric film may also improve the physical stability of solid dosage forms by increasing the strength and fracture resistance of the solid.^[3] Reactive components within a single dosage form may be separated using polymeric film coatings. Taste and odor masking of drugs may also be accomplished with the use of polymeric film coatings.^[4] One of the most common reasons for the application of polymeric coatings is to alter the release characteristics of drugs.^[5–7] This article discusses formulation development, processing, and testing of polymeric films and film-coated products.

FUNDAMENTALS OF FILM FORMATION

In the pharmaceutical industry, polymeric films are generally applied to solid dosage forms using a spray-atomization technique. The polymer is dissolved or dispersed in aqueous or organic solvents prior to spraying. The solid cores are often preheated in the coating equipment prior to initiation of the coating process. This prewarming stage is especially important in the coating of soft gelatin capsules.^[8,9] The coating solution or dispersion is atomized with air into small droplets, which are then delivered to the surface of the substrate. Upon contact, the atomized droplets spread across the substrate surface. The solvent may penetrate into the core, causing surface dissolution and physical mixing at the film–tablet interface.^[10] As the solvent begins to evaporate, the polymer particles densely pack on the surface of the solid. A schematic of the film formation process for an aqueous polymeric dispersion is shown in Fig. 1.

Upon further solvent evaporation, the particles flow together due to the cohesive forces between the polymer spheres, a process known as coalescence. Heat is generally added to the coating equipment to facilitate solvent evaporation and film formation.^[11]

Immediately following the completion of the coating process, coated solids are generally stored at temperatures above the glass transition temperature of the polymer to further promote coalescence of the film^[6] and ensure a homogeneous distribution of the plasticizer.^[12] During this postcoating storage or curing stage, the microstructure of the polymer is altered,^[13] and the mechanical, adhesive,^[14] and drug release properties^[15,16] of the film are correspondingly affected. The time required to obtain a stable film without further aging effects is dependent on a number of factors, including the type and concentration of plasticizer,^[17,18] the bed temperature used during coating,^[16] and the storage conditions.^[11] Fig. 2 shows the influence of curing on theophylline release from pellets coated with an acrylic polymer.^[15] Drug release rates slowed during storage at elevated temperatures and higher concentrations of the plasticizer in the coating reduced the time necessary for complete film coalescence.

While the majority of published research has shown that curing slows drug release, Bodmeier and Paeratakul^[13] found that curing could enhance or retard drug release, depending on curing conditions, plasticizer concentration, and the physicochemical properties of the drug. In that study, chlorpheniramine maleate was found to have a low affinity for the ethyl cellulose polymer and curing slowed drug release. In contrast, ibuprofen was found to have a high solubility in ethyl cellulose and longer curing times resulted in faster drug release. Drug crystals were found on the surface of the coated pellets following curing and the researchers attributed these findings to diffusion of the drug into the film. A subcoat or intermediate seal coat was suggested to minimize interaction between the drug and the cellulosic polymer.

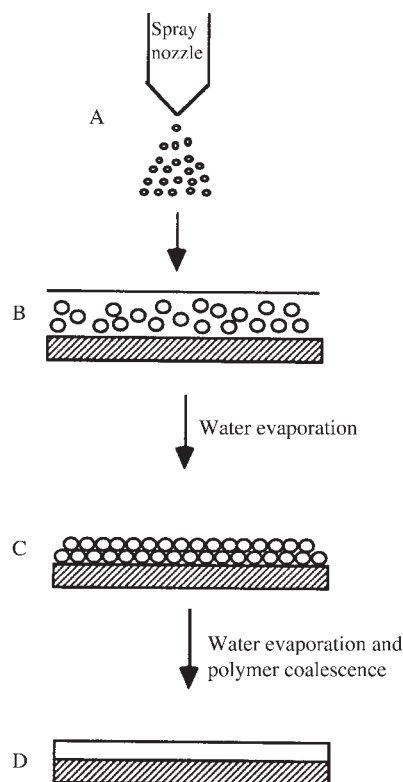


Fig. 1 Schematic representation of the film formation process for an aqueous polymeric dispersion: (A) atomization of the polymeric dispersion; (B) deposition of the polymeric dispersion on the substrate surface; (C) packing of the polymer spheres with water filling the void spaces; (D) formation of continuous polymeric film.

COATING EQUIPMENT

There are three types of coating equipment used to apply polymeric materials: conventional coating pans, perforated coating pans, and fluidized beds. More detailed discussions of the various coating equipments may be found elsewhere.^[19–21] The conventional coating pan system consists of a round coating pan that rotates on an inclined axis. Tablets in the coating pan tumble due to pan rotation. Heat is blown across the surface of the tumbling tablets and exhaust air is withdrawn. The conventional coating pan was originally used in the sugar coating process, where syrups were ladled onto the substrates. Aqueous polymeric film coating, however, requires more rapid solvent evaporation and the drying efficiency of conventional coating pans was improved by adding perforations. These perforations allow air to be forced through the tablet bed. A schematic of a perforated coating pan is shown in Fig. 3. As with conventional coating systems, pan rotation causes the solid cores to continually

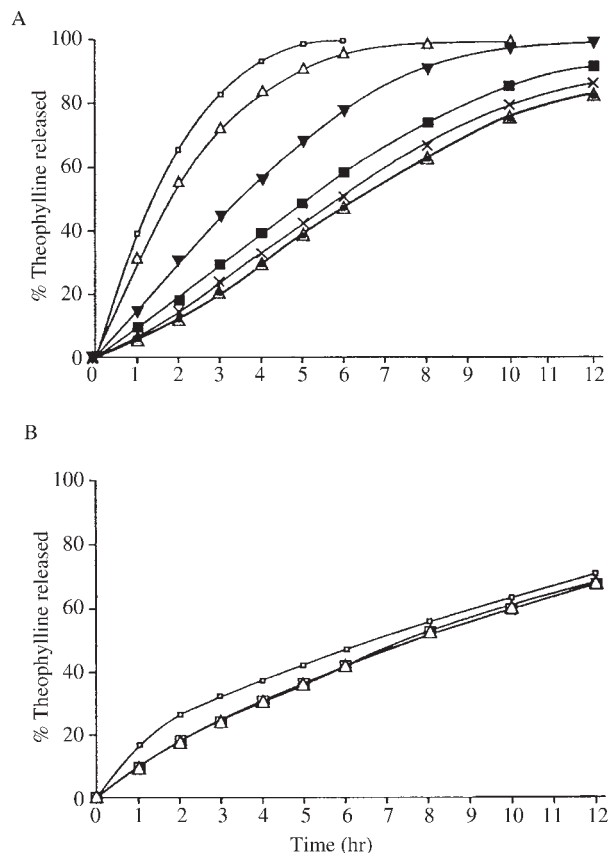


Fig. 2 Change in drug release from theophylline pellets coated with Eudragit® RS 30 D containing 5% Pharmacoat® 606 and either (A) 20% triethyl citrate or (B) 30% triethyl citrate after storage at 40 °C/50% relative humidity. For (A): (□) initial; (Δ) 2 hr; (▼) 6 hr; (■) 24 hr; (×) 3 days; (◼) 7 days; (▲) 10 days. For (B): (□) initial; (Δ) 3 hr; (■) 6 hr; (◆) 12 hr. (From Ref. 15.)

move during the coating process. The application zone is the area in the pan where the atomized polymer is sprayed onto the substrate surface. Tablets must make multiple passes through the application zone to form the film. Baffles in the perforated coating pans contribute to the tumbling action of the tablets and facilitate uniform film coverage.

In contrast to the coating pans, small particles such as beads, pellets, granules, and powders are generally coated using the fluidized bed or air suspension method, which utilizes a carrier gas to keep the cores in motion. The high air current makes this technique more efficient at water removal.^[22] The bottom spray technique is one of the most common fluidized bed application methods and a schematic of this process is shown in Fig. 4A. A perforated distribution base plate allows sufficient air into the product container to force the particles up into a cylindrical

Film Coating of Oral Solid Dosage Forms

3

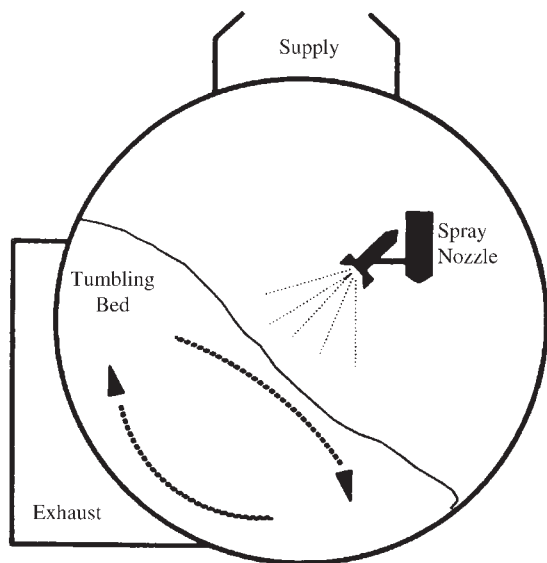


Fig. 3 Schematic of a perforated coating pan apparatus. (From Ref. 19.)

or slightly conical coating chamber, known as a Wurster insert. These inserts have been recently improved to keep the particles away from the spray nozzle until the spray pattern is fully developed.^[23] The bottom spray method provides ideal conditions for complete film coalescence.^[24] The particle size of the pellets, however, has been shown to influence fluidization patterns, which may ultimately affect the thickness of the film coating.^[25]

Two additional spray techniques have been used in fluidized bed coating. As shown in Fig. 4B, the top spray method, also known as the granulator mode, sprays the polymeric material countercurrently into the fluidizing particles. Because the coatings produced are not uniform in thickness, the top spray technique is suitable for taste

and odor masking, where drug release is not dependent on film thickness.^[19] The rotary or tangential spray technique, shown in Fig. 4C, uses a rotating disk to add a centrifugal force to fluidization and gravity. While films produced using this method are similar to those created with the Wurster system,^[23] the drying efficiency is improved and higher spray rates may be employed. Using the tangential spray system, film thickness has been shown to be independent of the particle size of the cores.^[25] One of the drawbacks of this application technique, however, is a greater potential for adhesion of particles to the upper walls of the product chamber.

Irrespective of the type of coating equipment used, polymeric solutions and dispersions are generally applied using a spray-atomization technique and two types of spray nozzles are employed. With pneumatic nozzles, high-pressure air is passed across the fluid stream as it exits the nozzle opening. In contrast, hydraulic nozzles rely on the fluid being pumped at relatively high pressures through a small opening. One of the advantages of pneumatic nozzles is that the atomized droplet size can be controlled independently of the polymer flow rate, whereas changing the spray rate of a hydraulic nozzle without adjusting the nozzle will result in changes in the atomization spray pattern.^[19]

A variety of pumps may be used to deliver the coating material to the spray nozzle. The peristaltic pump is ideal for delivering latex and pseudolatex polymeric dispersions that may coagulate at high pressure. This pump is the most commonly used and is also the easiest to clean. To control the delivery of the liquid polymeric material more precisely, a gear pump may be employed. Problems with undissolved solids in the coating formulation, however, may arise due to the tight tolerances between the two gears. The gear pump is also more difficult to clean

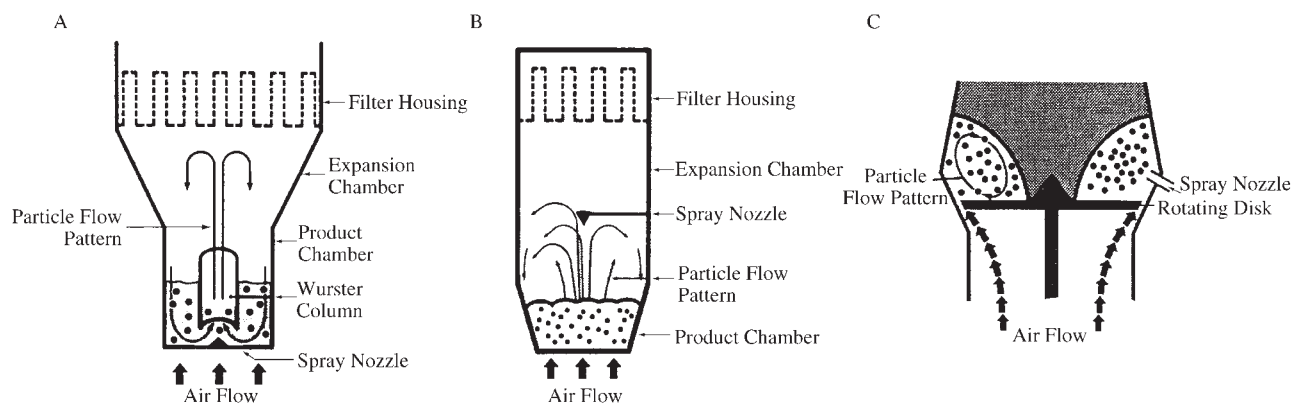


Fig. 4 Schematic of a fluidized bed coating apparatus: (A) bottom spray with Wurster column insert; (B) top spray technique; (C) tangential spray technique. (From Ref. 25.)

compared to the peristaltic system. A piston pump utilizes both air and hydraulic systems. One of the advantages of the piston pump is that minor clogs in the nozzle may be easily cleared due to the pressure reserve. Polymeric materials, however, may coagulate due to the high pressures used and the piston system is quite difficult to clean.

To overcome problems associated with traditional coating processes, several novel coating methods have been proposed. Obara et al.^[26] demonstrated the “dry coating” technique, where the polymer powder is fed directly to the tablet or pellet bed with simultaneous spraying of a plasticizing agent. Although modifications may be necessary, this technique can be used in fluidized bed and perforated coating pan methods, as shown in Fig. 5A and B, respectively. This method uses no organic solvents or water. The researchers reported that processing times were dramatically reduced, although higher amounts of an enteric polymer were required for gastric resistance. An electrostatic coating process has also been described.^[27]

FILM COATING MATERIALS

Film Formers

Materials used to coat pharmaceutical products are primarily based on acrylic and cellulosic polymers and

the aqueous solubility characteristics of these compounds generally dictate their uses. Sustained release coatings are water-insoluble or swellable films through which the medicament slowly diffuses. Drug products coated with these polymers are administered less frequently and the high peak plasma concentrations associated with side effects may be reduced or eliminated. Common sustained release polymers commercially available include ethyl cellulose and water-insoluble polymethacrylates. These polymeric materials may also be used as binders in sustained release matrix tablets and in other novel drug delivery systems. In contrast, water-soluble polymers, including hydroxypropyl cellulose, hydroxypropyl methylcellulose, sodium carboxymethylcellulose, and polyvinyl pyrrolidone, are often used for rapidly disintegrating film-coated tablets. These materials have also been added to the water-insoluble polymers to accelerate drug release from sustained release films.^[17,28]

Enteric film coatings exhibit pH-dependent solubility and have been used to protect drugs from degradation in the stomach.^[29] Enteric coatings have also been employed to decrease the incidence of gastric irritation from drugs^[30] or target drug delivery to the small intestines. In the low pH of the stomach, mixed acid and acid ester functional groups on the enteric polymers are unionized, and therefore, insoluble. As the pH increases in the intestinal tract, these functional groups ionize and the polymer becomes soluble. Thus, an enteric polymeric film allows the coated solid to pass through the stomach intact and release the medication in the small intestines. Common enteric polymers commercially available include cellulose acetate phthalate, hydroxypropyl methylcellulose phthalate, hydroxypropyl methylcellulose acetate succinate, polyvinyl acetate phthalate, and several methacrylic acid copolymers.

Solvents

Many solvent systems have been used to disperse or dissolve polymers for pharmaceutical film coating purposes. The primary criteria for the selection of a solvent for a particular polymer system include solvency, volatility, toxicity, and pollution control. The most superior films, showing the greatest combined strength of cohesiveness, have been reported when the coating solution solvation and polymer chain extension are at a maximum.^[31] Film coating technology, however, has shifted toward aqueous-based systems for environmental and economic reasons and the majority of polymeric materials used today are applied as aqueous-based solutions and dispersions. With aqueous-based systems,

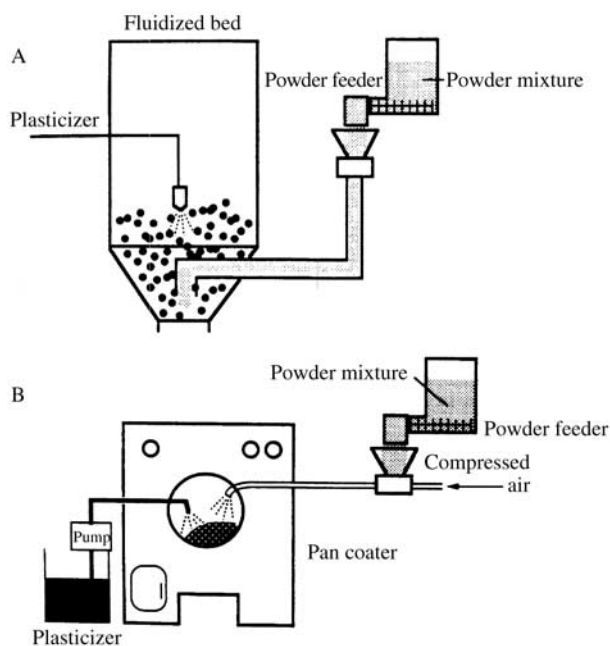


Fig. 5 Schematic of the “dry coating” method: (A) fluidized bed apparatus; (B) pan coating apparatus. (From Ref. 26.)



the risk of explosion is diminished, costs of disposing of the solvents are reduced, and concerns of potential toxicities due to residual solvents within the film are eliminated. Many commercial polymeric systems are available as aqueous latex and pseudolatex dispersions, where colloidal polymer particles are suspended in water. Latexes are obtained by emulsion polymerization, whereas emulsification of polymeric solutions is used to produce pseudolatex dispersions. Pseudolatexes may also be prepared from suspensions of spray-dried or mechanically milled solid polymer particles. The viscosity of these latex and pseudolatex dispersions is independent of the molecular weight of the polymer and only slight increases in viscosity occur with increased polymer concentration.^[32]

Plasticizers

Many pharmaceutical polymers exhibit brittle properties and require the addition of a plasticizing agent to obtain an effective coating, free of cracks, edging, or splitting. Plasticizers function by weakening the intermolecular attractions between the polymer chains and generally cause a decrease in the tensile strength and the glass transition temperature and an increase in the flexibility of the films.^[33] Plasticizers are necessary components to reduce brittleness, improve flow, impart flexibility, and to increase toughness, strength, and tear resistance of the film.^[34] These plasticizers play a critical role in the performance of polymeric film coatings.^[15,35]

Plasticizers are generally nonvolatile, high boiling, nonseparating substances that, when added to polymers, change certain physical and mechanical properties of that material. Plasticizers used in a polymeric system should be miscible with the polymer and exhibit little tendency for migration, exudation, evaporation, or volatilization. Many compounds can be used to plasticize polymers, including water. Phthalate esters such as diethyl phthalate, sebacate esters such as dibutyl sebacate, and citrate esters such as triethyl citrate and tributyl citrate are commonly used as plasticizing agents. Various glycol derivatives including propylene glycol and polyethylene glycol have also been used to plasticize polymeric films. In addition, surfactants, preservatives, and other compounds have been shown to function as plasticizing agents in cellulosic and acrylic polymers.^[36,37]

To be effective, a plasticizer must partition from the solvent phase into the polymer phase and subsequently diffuse throughout the polymer to disrupt the intermolecular interactions.^[38] The rate and extent of this partitioning for an aqueous dispersion have been found to be dependent on the solubility of the plasticizer in water

and its affinity for the polymer phase. The partitioning of water-soluble plasticizers in an aqueous dispersion occurs rapidly, whereas significantly longer equilibration times are required for water-insoluble plasticizing agents.^[8,39] For aqueous-based dispersed systems, water-insoluble plasticizers should be emulsified first and then added to the polymer.^[40] Sufficient time must be allowed for plasticizer uptake into the polymer phase prior to the initiation of coating. If insufficient time for plasticizer partitioning is given, the unincorporated plasticizer droplets, as well as the plasticized polymer particles, will be sprayed onto the substrates during the coating process, resulting in uneven plasticizer distribution within the film, which could potentially cause changes in the polymer properties of the film over time. Siepmann et al.^[41] developed a plasticizer uptake model to predict the minimum stirring time necessary for complete partitioning of plasticizers into aqueous polymeric dispersions. A postcoating thermal treatment may help reduce or eliminate the plasticization time effects.^[18]

The effectiveness of a plasticizing agent is dependent, to a large extent, on the amount of plasticizer added to the film coating formulation and the extent of polymer–plasticizer interaction. Forces involved in polymer–plasticizer mixtures include hydrogen bonding, dipole–dipole, and dipole-induced dipole interactions, as well as dispersions forces. Many experimental methods to determine the extent of polymer–plasticizer interactions have been reported in the pharmaceutical literature, including torsional braid pendulum,^[42] vapor pressure depression,^[43] osmotic pressure,^[44] swelling tests,^[45] gas–liquid chromatography,^[46] viscometry,^[47] melting point depression,^[48] nuclear magnetic resonance,^[49] and Fourier transform infrared.^[50] The degree of plasticizer–polymer interactions has been extensively characterized using differential scanning calorimetry and the decrease in the glass transition temperature of the polymer with the addition of a plasticizing agent is a common measure of plasticizer effectiveness.^[33,51,52]

Other Additives

In addition to the polymer, plasticizer, and solvent, other water-soluble and water-insoluble compounds such as pigments, antiadherents, surfactants, and antifoaming agents may be added to coating formulations to improve the appearance of the final dosage form, to facilitate processing, and to reduce the tackiness of the films. Drugs have also been incorporated directly into film coating formulations.^[32,53] The volume concentration, size and size distribution, chemical properties and surface charge of the additives, and the extent of polymer–particle interaction

may significantly affect the mechanical, adhesive, and drug-release properties of the coatings.^[54–56] Several classes of commonly used film coating additives are discussed below.

Antiadherents

During coating, curing, and storage, many pharmaceutical film coatings become sticky and the coated particles may agglomerate. The degree of tackiness correlates with the minimum film forming temperature of the polymer and has been shown to increase with increasing plasticizer concentration.^[57] To reduce the stickiness of the film and minimize agglomeration of the coated substrates, antiadherent compounds are generally included in coating formulations and talc is one of the most common antiadherents used. However, the high levels of talc (up to 100% w/w, based on dry polymer weight) required to reduce the tackiness of the coating may result in clogging of the spray nozzle and particle sedimentation. Furthermore, talc is hydrophobic and the addition of talc to polymeric films has been shown to decrease water vapor permeability^[2] and the dissolution rate of drugs.^[58] Other studies, however, have shown that talc may increase dissolution, presumably by forming cracks in the coating.^[59] Talc has also been found to affect the mechanical, thermal, and adhesive properties of polymeric films.^[60,61] Glyceryl monostearate (GMS) has been used as an alternative to talc.^[62] Wesseling et al.^[57] showed that 5% GMS was as effective as 50% talc in reducing the tackiness of several acrylic polymeric films.

Colorants

Pigments used in pharmaceutical systems include aluminum lakes of water-soluble dyes, opacifiers such as titanium dioxide, and various inorganic materials including the iron oxides. Reports of color migration and stability issues have curtailed the use of water-soluble dyes in pharmaceutical products.^[63] Pigments have been employed to provide easy product identification and present a more pharmaceutically elegant dosage form. In addition, titanium dioxide and other opacifying agents have been incorporated into film coating formulations to improve the stability of light-sensitive drugs.^[1,64]

Incorporation of pigments into polymeric dispersions may result in incompatibilities, such as coagulation of the polymer or flocculation of the pigments.^[65] These instabilities are related to the size and surface charge of the components in the colloidal polymer formulation as well as the pH of the medium.^[66] Other excipients may be

added to the coating formulation to stabilize the polymeric dispersions.^[40] Sodium carboxymethylcellulose, for example, has been used at a concentration of approximately 6% (w/w, based on dry polymer weight) to stabilize pigmented acrylic dispersions.^[67] The polymer chains of this anionic cellulose ether lead to sterical stabilization of the pigment and prevent flocculation by increasing the viscosity of the mixture. Other approaches to stabilize aqueous polymeric dispersions have also been suggested.^[68]

Pigments differ significantly in their physical properties, including density, particle shape, particle size, and morphology, and these differences contribute to the complex relationship with aqueous film coatings. The size, shape, surface chemistry, and concentration of the pigments have been shown to affect polymer properties. Felton and McGinity,^[69] for example, found an inverse relationship between the particle size of the pigment and film–tablet adhesion. These researchers theorized that larger particles disrupt the interfacial bonding between the polymer and the surface of the tablet to a greater extent than the smaller particles. In another study, Rowe^[70] attributed the increased modulus of elasticity of polymeric films containing colorants to the shape of the pigment particle and to the extent of particle–polymer interaction. Maul and Schmidt^[71] investigated the effects of several pigments of similar size but differing in surface polarity on drug release. Fig. 6 shows that pigments with polar surfaces (such as titanium dioxide, iron oxide, and mica) produced films that were less permeable than when the hydrophobic talc was incorporated into the coating. Other studies have shown that the addition of titanium dioxide to acrylic and cellulosic films increases water vapor permeability^[2,72] and enhances polymer adhesion.^[69,73]

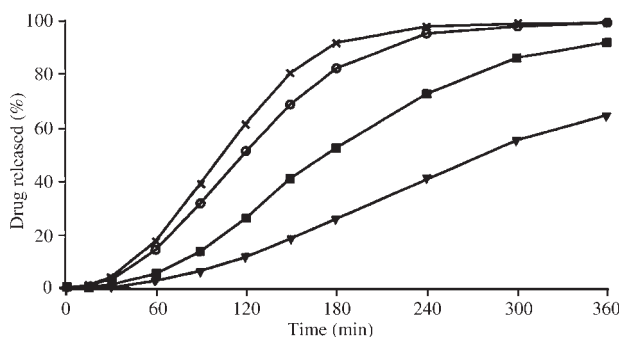


Fig. 6 Influence of surface chemistry of pigments on theophylline release from pellets coated with Eudragit® RS 30 D containing dibutyl phthalate as a plasticizer. (▼) Titanium dioxide as fine platelets; (■) mica; (×) talc; (○) red iron oxide fine platelets. (From Ref. 71.)

Film Coating of Oral Solid Dosage Forms

Surfactants

Surfactants have been added to coating formulations to improve substrate wettability,^[74] facilitate spreading of the polymeric material on the surface of the substrate,^[34] and homogenize the coating mixtures.^[75] Various polymer properties may be affected by the addition of surfactants to polymeric film coatings. Lindholm et al.^[76] for example, showed faster drug release rates occurred as the concentration of polysorbate 20, a hydrophilic nonionic surfactant, was increased in ethyl cellulose films. These researchers suggested that the surfactant leached from the coating during dissolution, creating pores in the film through which the drug diffused.

EVALUATION OF POLYMERIC FILMS

Determination of the drug release or dissolution properties of solid dosage forms is probably the most common test performed on pharmaceutical systems and methods for quantifying drug release are discussed elsewhere.^[77] For film-coated drug products, the polymer, film thickness, and plasticizer type and concentration have been shown to influence dissolution.^[15,78,79] The addition of soluble and insoluble excipients in the film coating formulation,^[80,81] processing parameters used during coating,^[16,82] post-coating drying,^[13,18] and storage conditions^[83] have also been shown to affect drug release.

The drug release characteristics and long-term stability of coated products may be predicted by evaluating isolated free films for permeability and mechanical strength. Methods commonly used to prepare isolated films include casting and spraying techniques. With the casting method, the polymeric solution or dispersion is cast onto a smooth surface of Teflon[®] or aluminum and the solvent is slowly evaporated. Use of insoluble excipients in the film coating formulation, however, may result in two different film surfaces being formed as the solvent is evaporated. Free films may also be prepared using an atomization spray process and a schematic of the apparatus is shown in Fig. 7. This spray-atomization technique is similar to processes used to coat pharmaceutical solids. The spray box consists of a rotating drum covered with a nonstick material that is encased in a hot box and the polymeric material is sprayed onto the rotating drum.^[84] Films prepared from aqueous polymeric dispersions using the spray method are generally more uniform and exhibit more consistent and reproducible mechanical properties than films prepared by the casting technique.^[85]

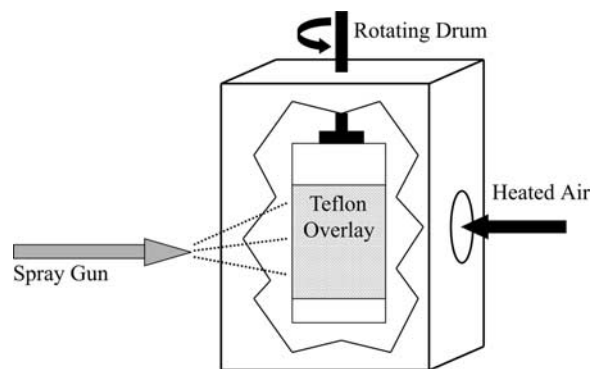


Fig. 7 Schematic of a spray box apparatus.

Water Vapor Permeability

Polymeric films have been used as barriers to protect moisture-sensitive drugs from atmospheric water and improve the stability of drugs that degrade by hydrolytic mechanisms.^[86] The water vapor permeability coefficient is used to evaluate the effectiveness of a particular film coating as a barrier to water. Water vapor permeability is commonly evaluated using isolated free films to construct water vapor transmission cells.^[53,87] A schematic of the apparatus is shown in Fig. 8.^[88] A water vapor transmission cell consists of a glass cylinder containing a saturated salt solution to produce a specific internal vapor pressure. The cylinder is then sealed with the polymeric film. Rubber rings and aluminum caps are often used to prevent water vapor egress through any areas other than the film. The transmission cells are placed in a humidity-controlled chamber and weighed periodically over a specified time period. These water vapor transmission cells rely on a vapor pressure gradient to achieve a linear weight gain. From the daily weight gains, the water vapor permeability constant can be calculated using Eq. 1:

$$P_{\text{erm}} = \frac{WL}{A \Delta P} \quad (1)$$

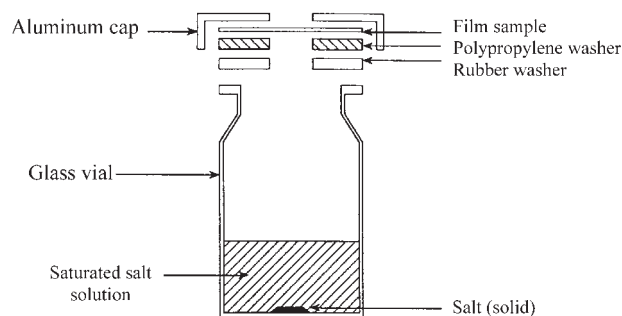


Fig. 8 Schematic of a water vapor transmission cell. (From Ref. 88.)

where P_{erm} is the moisture permeability constant, W is the amount of moisture transmitted per unit time, L is the thickness of the film, A is the area of the film exposed, and ΔP is the vapor pressure gradient. A method to determine water vapor permeation of applied films has also been reported.^[31]

Water vapor permeation has been shown to be dependent on the film composition,^[31,89] additives in the film,^[72] and the solvent system in which the polymer is dissolved or dispersed.^[2] A number of researchers have investigated the effects of plasticizers in coating formulations on the permeability of polymeric films.^[51,84,90] Water vapors have been shown to permeate more rapidly through films containing hydrophilic plasticizers,^[2] whereas the inclusion of a hydrophobic plasticizer in the coating has been found to exert no significant effects on water vapor permeability.^[91]

Thermal Analysis

Thermal analysis is a term used to describe a number of analytical experimental techniques that investigate polymer properties as a function of temperature. Some of the properties that can be determined include enthalpy, mass, melting temperature, heat of fusion, and the glass transition temperature. Hatakeyama and Quinn^[92] provide an excellent description of thermoanalytical techniques.

The glass transition temperature (T_g) is probably the most common physical property determined for amorphous polymeric materials. At temperatures below the T_g of the polymer, the material generally behaves as a hard and brittle glass, whereas the behavior of the polymer changes to soft and elastic at temperatures above the T_g . The degree of excipient–polymer interactions has been extensively characterized using differential scanning calorimetry. The effectiveness of a plasticizer in a film coating formulation has been related to its ability to decrease the T_g of the polymer.^[15,51,52] Table 1 shows the T_g of an acrylic polymer plasticized with several citrate esters.^[93] The addition of 10% plasticizer in the film coating produced a significant decrease in the T_g of the polymer for all compounds studied and higher concentrations of the plasticizer further lowered the T_g of the film. These data demonstrate that the citrates effectively plasticized the acrylic polymer.^[33] In contrast to plasticizing agents, other additives in film coating formulations, such as talc, titanium dioxide, and lactose, may increase the T_g of the film^[69,94] due to a restriction in the mobility of the polymer chains or a rise in the crystallinity of the polymer.^[95]

Table 1 Glass transition temperature (T_g) of Eudragit® RS 30 D polymeric films as a function of plasticizer type and concentration

Plasticizer	T_g of Eudragit® RS 30 D ^a	
	10% Plasticizer	20% Plasticizer
Triethyl citrate	34.3	12.8
Acetyl triethyl citrate	37.0	17.5
Tributyl citrate	38.2	20.5
Acetyl tributyl citrate	38.2	22.2
Triacetin	42.2	27.4

^aUnplasticized film is 55.0 °C.

(From Ref. 93.)

Mechanical Properties

Basic research with polymeric solutions and dispersions has focused on the mechanical properties, including tensile strength, Young's modulus, and elongation. Mechanical testing has been used to evaluate the effectiveness of plasticizers,^[33,90] as well as the permeability of film coatings^[96] and the dissolution characteristics of film-coated solids.^[97,98] Mechanical data obtained from tensile testing of free films have been used to predict the incidence of film defects in coated tablets.^[99] These mechanical data have also been used to make predictions regarding the long-term storage stability of film-coated solid dosage forms.^[100]

Common methods used to evaluate mechanical properties of polymeric films include microindenter probe analysis,^[101] puncture and shear tests,^[35,102] and stress relaxation.^[103] The stress–strain test is probably one of the most popular and widely used techniques to determine the mechanical properties of polymeric materials of pharmaceutical interest.^[104,105] Stress–strain testing generally consists of applying an axial load to an isolated free film and measuring the load and deformation simultaneously. The stress–strain test will provide a generalized curve from which several useful properties can be determined. A typical stress–strain curve obtained from tensile testing is shown in Fig. 9 and illustrates the terminology used. During the tensile test, initially there is a linear portion of the curve where the elongation of the polymeric film is directly proportional to the applied stress, and the slope of this linear region is known as Young's modulus. Young's modulus is a measure of the stiffness of the film or the ability of the film to withstand high stress while undergoing little elastic deformation. The greater the slope of the line, the higher the modulus and the greater the stiffness of the polymeric material. The plateau area where

Film Coating of Oral Solid Dosage Forms

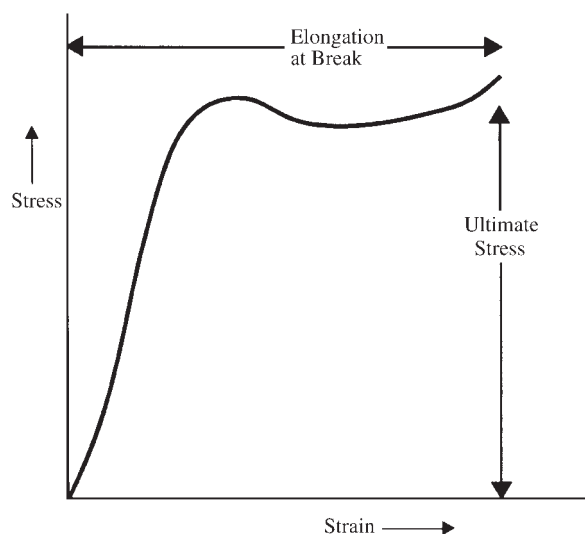


Fig. 9 Example of a stress–strain curve obtained in the tensile testing of isolated free films.

the film first undergoes a marked increase in strain without a corresponding increase in stress is known as plastic deformation. During plastic deformation, the structure of the polymeric film changes as the polymer chains orient themselves parallel to the direction of flow. As the film continues to be elongated, a point is reached at which the film breaks, known as the elongation at break. The stress value at this point is referred to as the ultimate tensile strength, which is a measure of the ability of the solid to withstand fracture. The area under the curve represents the work done and is referred to as the toughness of the polymer.

Using tensile testing of free films, researchers have shown that plasticizer type and concentration influence the mechanical properties of both acrylic and cellulosic polymers.^[33,106] Fig. 10 shows the influence of dibutyl

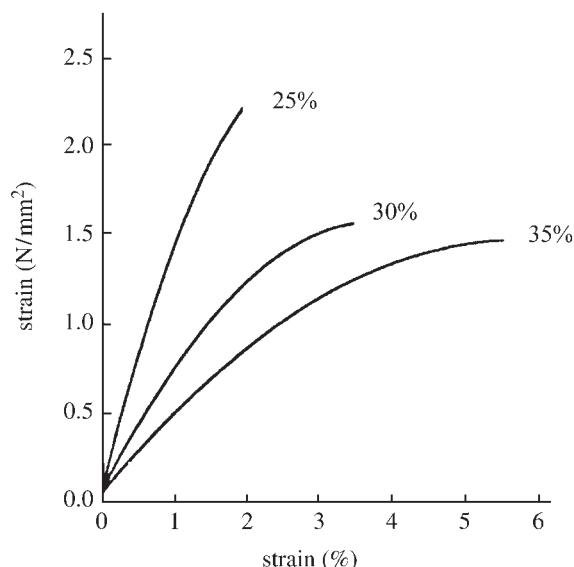


Fig. 10 Influence of dibutyl sebacate concentration on the stress–strain curve of isolated free films prepared from an ethyl cellulose aqueous dispersion (Aquacoat®). (From Ref. 107.)

sebacate concentration on the tensile properties of free ethyl cellulose films.^[107] As the concentration of the plasticizer was increased from 25% to 35%, the elongation at break increased and the slope of the curve decreased, indicating the film became more flexible. Storage conditions and aging have also been found to affect the mechanical properties of polymeric films.^[108,109]

The mechanical properties determined by tensile testing of isolated free films in the dry state may not predict the behavior of the coated dosage form when in contact with biological fluids.^[102] During the dissolution process, aqueous fluids will diffuse through the film coating into the solid core and the polymeric film will be in a hydrated or wet state. Table 2 shows the mechanical

Table 2 Mechanical properties of dry and hydrated films prepared from different polymeric dispersions plasticized with 20% (w/w) triethyl citrate (standard deviation in parentheses)

Polymer dispersion (film thickness, μm)	Puncture strength (MPa)		Elongation (%)	
	Dry	Wet	Dry	Wet
Aquacoat (309)	0.34 (0.11)	0.10 (0.02)	1.34 (0.18)	0.13 (0.02)
Surelease (394)	0.23 (0.44)	0.74 (0.10)	0.62 (0.12)	4.89 (0.90)
Eudragit NE 30 D (314)	2.16 (0.19) ^a	1.58 (0.10) ^a	> 365.00	> 365.00
Eudragit RS 30 D (309)	1.99 (0.23)	0.93 (0.04)	142.83 (4.32)	38.41 (4.65)
Eudragit RL 30 D (316)	1.81 (0.11)	1.60 (0.14)	126.31 (8.04)	13.02 (2.45)
Eudragit L 30 D (264)	0.83 (0.05)	1.78 (0.09) ^a	0.46 (0.25)	> 365.00

^aFilms did not rupture.

(From Ref. 35.)

properties of dry and wet acrylic and cellulosic films determined using a puncture test technique and these data clearly demonstrate that the strength and flexibility of hydrated films may be quite different from films in the dry state. For example, while brittle when dry, Eudragit L 30 D was significantly more flexible when hydrated, presumably due to the plasticizing effect of water.^[35] Bodmeier and Paeratakul^[102] also found that the solubility of the plasticizer incorporated in the film coating formulation influenced the mechanical properties of hydrated films. Triethyl citrate, a water-soluble plasticizer, leached from an acrylic film upon exposure to water, thus causing significant differences in the mechanical strength of dry and hydrated films. In contrast, no significant differences in the mechanical properties of wet and dry films containing the water-insoluble and nonleaching plasticizer acetyl tributyl citrate were noted.

While tensile testing of free films has been the traditional method used for studying the mechanical properties of polymers, compression tests have been to a lesser extent.^[110] Many similarities exist between tension and compression tests, including the manner of conducting the test and the analysis and interpretation of the results. Uniform displacement rates are applied in a manner similar to a tensile test, except for the direction of loading. One of the advantages of compression testing is that the effects of substrate materials, storage conditions, and physical aging on the mechanical properties of applied polymeric films may be investigated. Compression testing of coated solids has also been used to determine qualitatively the adhesive properties of polymers, with single peak failure indicative of high substrate/coating adhesion.^[3] Fig. 11 shows the force–deflection profiles obtained from compression testing of enteric-coated soft gelatin capsules. In most cases, the capsule shell and the polymeric film fractured simultaneously, as shown in Fig. 11A, indicating good adhesion. Fig. 11B shows that some of the coated soft gelatin capsules exhibited a break in the film followed by rupture of the gelatin shell, which the authors attributed to poor film–capsule adhesion.^[9]

Compression testing of coated pellets has recently received attention in the pharmaceutical literature, as researchers compress these pellets into the more tamper-resistant tablet dosage form.^[111,112] Methods to quantify the compression properties of coated pellets have been discussed elsewhere.^[113,114] The compressional forces used during tableting are critical and, if these forces exceed the strength of the coating, the film will fracture and faster dissolution rates will result.^[115] Due to the intensive contact during compression, slower drug release may also occur as the films from the coated pellets fuse together to form a matrix tablet.^[40] Readily compressible

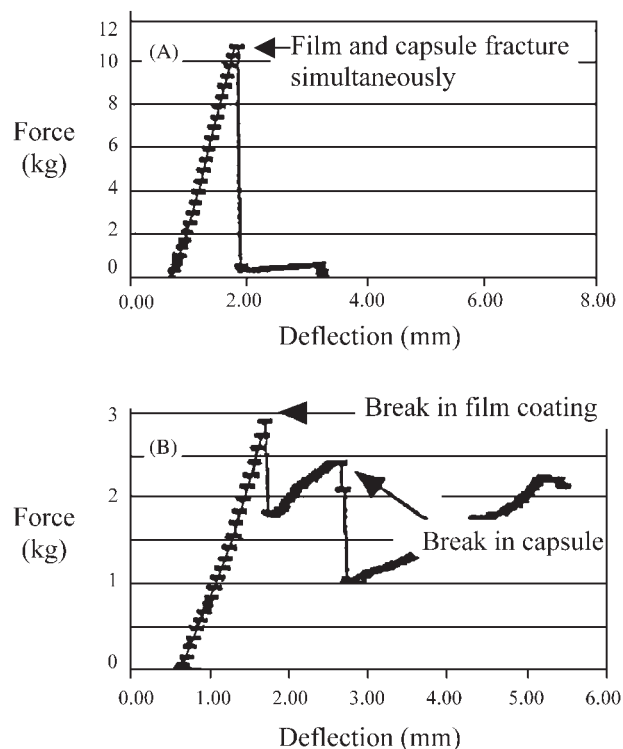


Fig. 11 Force–deflection profiles of soft gelatin capsules containing polyethylene glycol as a fill liquid and coated with Eudragit® L 30 D-55 obtained from compression testing: (A) polymer plasticized with 20% triethyl citrate; (B) polymer plasticized with 20% tributyl citrate. (From Ref. 9.)

excipients, such as starch and microcrystalline cellulose, have been dry blended with coated pellets prior to tableting to eliminate the possible fracture or fusing of the polymeric films. These excipients prevent direct contact of the coating and reduce friction during compression. The addition of at least 15% (w/w) microcrystalline cellulose to coated acetyl salicylic acid pellets was shown to prevent the formation of matrix tablets.^[116]

Polymer Adhesion

Good adhesion between a polymer and the surface of a solid is a major prerequisite for the film coating of pharmaceutical dosage forms.^[117] Poor adhesion may result in flaking or peeling of the coating from the solid substrate during storage, which could significantly jeopardize film functionality. Loss of adhesion may lead to an accumulation of moisture at the film–tablet interface, affecting the stability of drugs susceptible to hydrolytic degradation.^[118] Loss of adhesion may also compromise the mechanical protection the coating

provides to the solid substrate.^[3] In addition, experiments on adhesion may be useful to the pharmaceutical scientist during preformulation studies to investigate the relationship between tablet excipients and polymeric coating formulations.^[119]

Force involved in adhesion

The two major forces that have been found to affect polymer–tablet adhesion include the strength of the interfacial bond and the internal stresses within the coating. For pharmaceutical products, hydrogen bond formation is the primary type of interfacial contact between the tablet surface and the polymer. Dipole–dipole and dipole-induced dipole interactions also occur to a lesser extent. Factors that affect the type or the number of bonds formed between the polymer and the solid surface will influence film adhesion.^[117]

The second major factor influencing polymer adhesion is the internal stresses within a film. When a polymeric solution or dispersion is applied to a substrate, internal stresses inevitably develop within the coating.^[120] These stresses include stress due to shrinkage of the film upon solvent evaporation, thermal stress due to the difference in thermal expansion of the film and the substrate, and volumetric stress due to the volume change when a substrate swells upon storage. Several researchers have developed equations to estimate the total stress within a film.^[120,121] The total internal stress, P , may be calculated using Eq. 2:

$$P = \frac{E}{3(1-\nu)} \left[\frac{\phi_s - \phi_r}{1 - \phi_r} + \Delta\alpha_{\text{cubic}} \Delta T + \frac{\Delta V}{V} \right] \quad (2)$$

where E is the elastic modulus of the film, ν is the polymer's Poisson's ratio, ϕ_s is the volume fraction of the solvent at the solidification point of the film, ϕ_r is the volume fraction of the solvent remaining in the dry film at ambient conditions, $\Delta\alpha_{\text{cubic}}$ is the difference between the cubical coefficient of thermal expansion of the film coating and the substrate, ΔT is the difference between the glass transition temperature of the polymer and the temperature of the film during manufacturing and storage, ΔV is the volumetric change of the tablet core, and V is the original volume of the tablet core. While this equation has been derived for polymeric solutions, the theory is applicable to polymeric dispersions as well. From Eq. 2, the total stress within a film is directly proportional to the elasticity of the polymer. Therefore, factors that influence the elastic modulus of the polymer may affect film–tablet adhesion.

Methods to assess adhesion

The small size of the tablet and the nonuniform surface roughness of the substrate have presented significant challenges to the pharmaceutical scientist in assessing the adhesive properties of polymers. In the 1970s, film–tablet adhesion was assessed using a peel test, where a modified tensile tester peeled the film from the surface of the tablet at a 90° angle.^[122] The peel angle, however, is dependent on both the elasticity of the film and the uniformity of adhesion, which may produce significant deviations in the data.^[123] More recently, several variations of the butt adhesion technique have been reported in the pharmaceutical literature.^[122,127,128] Although similar to the peel test, the entire film is removed normal to the surface of the tablet, rather than sections of the film being peeled. The butt adhesion technique eliminates variations due to the elasticity of the film and is less influenced by the uniformity of adhesion. Using the butt adhesion test, force–deflection profiles can be generated, an example of this is shown in Fig. 12. This graph, similar to stress–strain curves commonly generated in the tensile testing of free films, permits the visualization of the development of the force within the sample during the adhesion test. In addition to the force of adhesion, the elongation at adhesive failure, the modulus of adhesion, and the adhesive toughness of the film can be determined. The elongation at adhesive failure is the distance the upper plate traveled at film separation. Analogous to the elongation at break obtained from tensile testing of free films, elongation at adhesive failure reflects the ductility of the polymeric film.

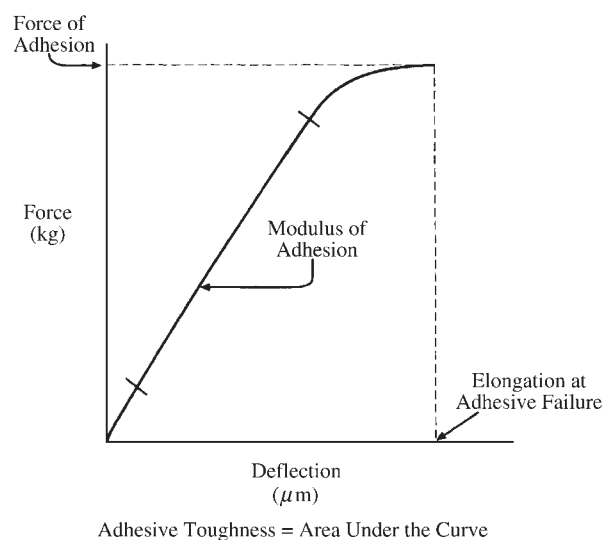


Fig. 12 Example of a force–deflection profile obtained from a butt adhesion experiment.

The modulus of adhesion is the slope calculated from the linear region of the force–deflection diagram and may be compared to Young's modulus obtained from tensile testing of free films. The adhesive toughness is the work required to remove the film from the tablet surface and may be calculated from the area under the force–deflection profile.

Substrate variables affecting adhesion

The surface roughness of the tablet compact and the force of compression used during tableting will affect polymer adhesion, by altering the effective area of contact between the film coating and the surface of the solid. The compressional force used during tableting has been found to significantly influence adhesion of organic-based cellulosic films, as shown in Fig. 13.^[123] Above a critical force, increased compression during the tableting process resulted in decreased adhesion. Below the critical compression pressure, the tablet laminated, rather than the film being separated from the tablet surface, and cohesive failure occurred. In the same study, a relationship between tablet porosity and polymer adhesion was found, as shown in Fig. 13, which the authors attributed to the extent of solvent penetration into the substrate cores and the physical mixing at the film–tablet interface. Using aqueous-based acrylic dispersions, Felton and McGinity^[119] also reported stronger adhesion when the polymer was applied to rougher tablet surfaces.

Because adhesion is primarily due to hydrogen bond formation, hydrophobic excipients in the tablet may decrease adhesion by presenting a surface consisting of mainly nonpolar hydrocarbon groups. As shown in Fig. 14, increased concentrations of stearic acid, a commonly used

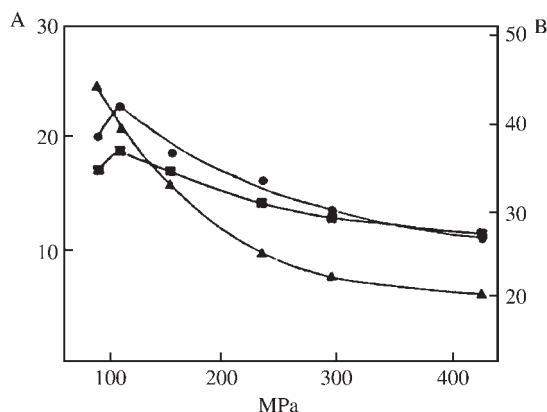


Fig. 13 Influence of compression pressure on the (▲) porosity and adhesion of film formulations containing (●) low viscosity and (■) high viscosity hydroxypropyl methylcellulose: (A) porosity (%); (B) adhesion (kPa). (From Ref. 123.)

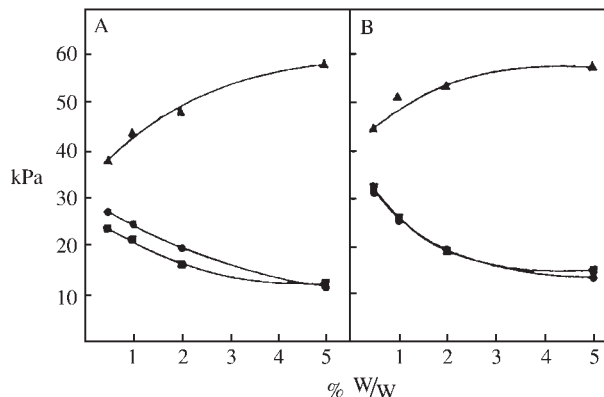


Fig. 14 The effect of lubricant concentration (% w/w) on the measured adhesion (kPa) of hydroxypropyl methylcellulose films: (A) Pharmacoat® 606; (B) Methocel® 60HG viscosity 50; (▲) stearic acid; (●) magnesium stearate; (■) calcium stearate. (From Ref. 124.)

lubricating agent that has a free polar carboxyl group, improved adhesion of an organic-based cellulosic polymeric film.^[124] When more hydrophobic lubricating agents were added to the tablet compacts, polymer adhesion decreased. Increased concentrations of magnesium stearate in tablets have also been shown to significantly lower adhesion of aqueous-based acrylic and cellulosic films.^[119,125] In contrast, the addition of microcrystalline cellulose to tablets improved film–tablet adhesion, presumably due to the saturation of the tablet surface by hydroxyl groups of the excipient.^[124,125]

Film coating variables affecting adhesion

Several studies have shown that the solvent system used in the coating formulation affects polymer adhesion.^[126] The solvent will interact with the polymer and will affect the random coil structure of the polymer chains. It is generally accepted that the greater the polymer–solvent interaction, the greater the end-to-end distance, thus exposing more of the polymer that is capable of interacting with and binding to the surface of the solid. Nadkarni et al.^[127] suggested that the solubility parameter of the solvent be used as a qualitative measure to the extent of polymer solvation, with greater polymer solvation resulting in greater film–tablet adhesion. In 1988, Rowe^[128] developed equations using the solubility parameters of the tablet excipients and the polymer to predict trends in film–tablet adhesion. Wood and Harder^[122] used contact angle measurements, as an indication of surface wettability, to predict polymer adhesion. In contrast, Felton et al.^[74] added surfactants to acrylic dispersions to alter substrate wettability. These researchers suggested that tablet wettability is not a valid

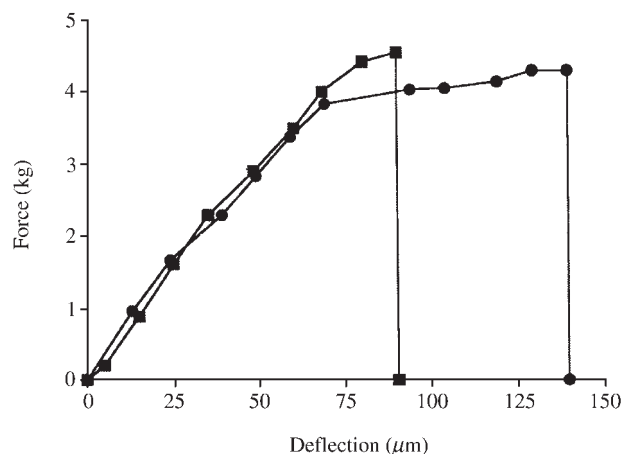
Table 3 Influence of the plasticizer in the coating formulation on the force of adhesion and the glass transition temperature (T_g) of an acrylic resin copolymer (standard deviation in parentheses)

Plasticizer	Force of adhesion (kg)	T_g (°C)
Triethyl citrate	4.85 (0.27)	36.5 (1.1)
Polyethylene glycol 6000	4.32 (0.25)	38.6 (2.5)
Tributyl citrate	3.81 (0.30)	51.2 (2.2)
Dibutyl sebacate	3.48 (0.33)	62.0 (3.6)

(From Ref. 129.)

predictor of adhesion of polymeric dispersions, because it is the polymer that binds to the substrate, not the solvent.

One of the two major forces influencing polymer adhesion is the internal stress within the polymeric film.^[120] The addition of plasticizing agents to coating formulations generally decreases the internal stress in the film by decreasing both the elastic modulus (E) and the glass transition temperature (T_g) of the film coating (Eq. 2). Felton and McGinity^[129] demonstrated a relationship between adhesion and the T_g of the polymer. As shown in Table 3, film–tablet adhesion was strongest in the more plasticized films. The researchers attributed these findings to the extent of the polymer–plasticizer interactions and the effectiveness of the plasticizing agent in lowering the internal stresses within the film coating. Higher concentrations of plasticizers within polymeric films caused a slight decrease in the force of adhesion and a significant increase in the elongation at adhesive failure and adhesive toughness, as shown in Fig. 15.^[129] These findings were

**Fig. 15** Influence of triethyl citrate plasticizer concentration on the force–deflection profile of Eudragit® L 30 D-55 applied to tablet compacts: (■) 20% plasticizer; (●) 30% plasticizer. (From Ref. 129.)

again attributed to a lowering of the internal stresses within the film and an increase in the elasticity of the polymer. Interestingly, adhesion was also found to be dependent on the physicochemical properties of the plasticizer in conjunction with the surface hydrophobicity of the tablet compacts, with films containing hydrophobic plasticizers being less sensitive to the hydrophobicity of the substrate.^[129]

Insoluble additives in coating formulations, including talc, titanium dioxide, and other pigments have been found to influence polymer adhesion.^[118] These insoluble excipients generally cause an increase in the internal stresses within the film and are thought to embed themselves at the film–tablet interface.^[70,117] The extent to which excipients incorporated into film coating formulations may affect polymer adhesion is dependent on the particle size, concentration, and morphology of the insoluble agents in the coating.^[56,69,118,125]

Processing parameters affecting adhesion

Dimensional changes in the tablet core that occur during the coating process will influence the internal stresses within the films of the final coated products and may ultimately affect polymer adhesion.^[130] The extent of substrate swelling is dependent on the temperature used during processing. Rowe^[131] suggested that the excipients in the tablet core and the polymeric coating materials should have similar coefficients of thermal expansion to minimize the development of internal stresses in the film during processing.

The postcoating drying or curing stage of processing may affect polymer adhesion. As the solvent evaporates and the polymer droplets coalesce, the number of potential polymer–substrate binding sites increases, as shown in Fig. 16. Felton and Baca^[14] showed that film–tablet adhesion increased during curing, and the force of adhesion was dependent on the hydrophobicity of both the substrate and the plasticizing agent in the film formulation. Environmental storage conditions have also been shown to influence adhesion.^[9,73,118,126,129]

PROBLEMS ENCOUNTERED DURING FILM COATING

Defects in the Film

Coated solids are generally evaluated visually to detect defects in the film^[132] and a list of common defects is presented in Table 4.^[133] Some defects may be esthetic in nature while others are more serious and may compromise the functionality of the film coating. Manipulating

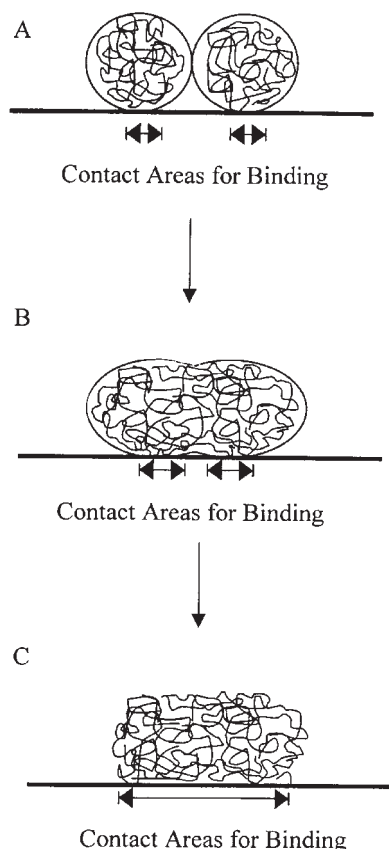


Fig. 16 Schematic of the increase in potential polymer-substrate interactions as film formation proceeds: (A) closely packed polymer spheres; (B) initiation of coalescence and polymer chain interdiffusion; (C) complete film formation. (From Ref. 14.)

processing conditions, adjusting the coating formulation, or reformulating the substrate may resolve these problems. Picking, for example, occurs when tablets briefly stick together during the coating process. This defect generally occurs in an over-wet tablet bed and may be eliminated by reducing the spray rate or increasing the bed temperature.^[63] Picking may also be reduced by adding antiadherents to the coating formulation. Mottling is a blotched or spotted film coating and is associated with inadequate pigment dispersion or color migration. If the colorant is uniformly dispersed in the coating formulation, color variations within the film may be curtailed by increasing the rotational pan speed, using baffles in the coating pan, or increasing the number of spray guns used.^[134,135] A rough surface, similar to the surface of an orange, is known as the orange peel effect and occurs when the spray droplets become too viscous to spread across the substrate. Perhaps, more serious and likely to compromise film functionality, cracking and peeling of the coating occur when the internal stresses exceed the tensile strength of the film and reformulation of the coating is often required.^[132]

Characteristics of the Substrate

The physical and chemical characteristics of the substrate are important considerations in film coating. Substrates, for example, must be sufficiently robust to physically withstand tumbling in the coating apparatus. Weak tablets may chip or fracture during coating and slower rotational

Table 4 List of common defects that may occur during film coating

Defect	Description
Blistering	Film becomes locally detached, forming a blister
Blooming	Dulling of coating
Blushing	White specks or haziness in film
Bridging (of the intagliation)	Film pulls out of intagliation or monogram forming a bridge across the mark
Chipping	Film becomes chipped
Color variation	Intertablet variation in color
Cracking	Film cracks across the crown of the tablet
Cratering	Volcanic-like craters in film
Flaking	Film flakes off exposing tablet surface
Infilling	Intagliation filled by solidified foam
Mottling	Uneven distribution of color (intratablet)
Orange Peel	Film rough and nonglossy surface like skin of orange
Peeling	Film peels back from edge exposing tablet surface
Picking	Isolated areas of film pulled away from surface
Pitting	Pits in surface of tablet core without disruption of film
Roughness	Film rough and nonglossy
Wrinkling	Film with wrinkled appearance

(From Ref. 133.)

pan speeds may be required. Interactions between the substrate core and the film coating may occur as the polymeric solution or dispersion physically mixes at the interface.^[95,98,136–138] Pitting has been reported to occur when coating tablets containing low melting excipients, such as stearic acid.^[139] Drugs may also interact with excipients in the film coating formulation.^[140] A seal coat or subcoat may be necessary to separate the reactive components and eliminate these types of interactions.

Sustained release wax matrix tablets are generally considered difficult to coat with aqueous polymeric dispersions due to the poor wettability of the hydrophobic tablet surface.^[141] The physicochemical characteristics of the substrate have also been shown to affect the mechanical and adhesive properties of the polymer.^[69,119] Adhesion problems have been reported when coating both hard and soft gelatin capsules.^[8,9,142] The capsule shell is relatively smooth and generally provides less surface area for interfacial contact compared to tablet compacts.^[143,144] The hydrophobicity of the fill liquid of soft gelatin capsules in conjunction with the plasticizer used in the coating formulation has been shown to influence polymer adhesion.^[9] Adhesion may be improved by adding polyethylene glycol 400 or 6000 to the coating formulation or using subcoats of hydroxypropyl cellulose or hydroxypropyl methylcellulose.^[117]

In addition to adhesion problems, other difficulties attributed to the physical properties of gelatin have been reported in coating capsules with aqueous polymeric formulations. The capsule halves may separate during the coating process due to the tumbling movements of the coating pan. During coating, the capsule shell may become either sticky due to solubilization of the gelatin or brittle due to water evaporation. Thus, the bed temperature and the polymer spray rate become even more critical parameters in capsule coating processes. In addition, a prewarming stage is required when coating soft gelatin capsules.^[8,9] The temperature of the fill liquid must be raised to that of the bed temperature to allow for uniform drying of the film.

Physical Aging

Another critical problem that may occur with film-coated products is a change in drug release rates over time and is generally associated with aging. The majority of polymers used in pharmaceutical products are amorphous and are not at thermodynamic equilibrium at temperatures below their glass transition temperatures. During time, amorphous polymers undergo a slow transformation toward a thermodynamic equilibrium. A schematic representation of this process is shown in Fig. 17. As temperatures are

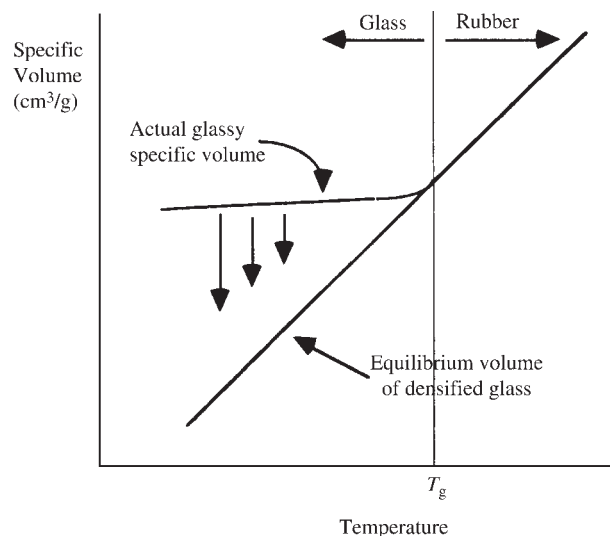


Fig. 17 Schematic of the origin of physical aging in amorphous polymers.

cooled to below the glass transition temperature of the polymer, the free volume of the polymer will slowly relax toward a lower free energy state over time. This process is referred to as physical aging. Ali and Sheldon^[145] demonstrated that a structural reorganization of the polymer chains occurs during aging. Researchers have shown that aging significantly impacts the long-term stability of film-coated products, affecting the permeability, mechanical, adhesive, and drug release properties.^[100,108,146,147] Complete coalescence of the film coating is critical to minimize physical aging effects.^[148] Finally, it should be noted that polymers used in the coating of pharmaceutical dosage forms are sensitive to environmental storage conditions, and temperature and humidity can significantly alter polymer properties.^[15,18,83,100,108,109]

CONCLUSIONS

Cellulosic and acrylic polymeric films have been used to coat pharmaceutical drug products for decorative, protective, and functional purposes. These materials are generally applied using a spray-atomization technique. Additional excipients, including plasticizers, pigments, antiadherents, and surfactants, may be incorporated into the coating formulation to aid in processing or to improve the esthetic appearance of the coated solid. The addition of these excipients, however, may alter polymer properties and affect drug release rates. In the development of a film coating formulation, various polymer properties are commonly evaluated and these data are used to make predictions regarding the dissolution characteristics and



long-term stability of the final product. Adhesion of the polymer is critical to the performance of the film and may be affected by the physicochemical properties of the substrate, additives in the coating formulation, and processing conditions. Subcoats have been used to improve polymer adhesion as well as to separate reactive components in the film and substrate. Defects in the polymeric film may be eliminated by adjusting processing parameters or by reformulating the coating and/or the substrate. Aging of coated solids may cause changes in the drug release characteristics and polymeric films should be completely coalesced to minimize aging problems.

REFERENCES

1. Bechard, S.R.; Quraishi, O.; Kwong, E. Film Coating: Effect of Titanium Dioxide Concentration and Film Thickness on the Photostability of Nifedipine. *Int. J. Pharm.* **1992**, *87*, 133–139.
2. List, P.H.; Kassis, G. On the Permeability of Various Tablet Coatings to Water Vapour and Oxygen. *Acta Pharm. Technol.* **1982**, *28*, 21–33.
3. Stanley, P.; Rowe, R.C.; Newton, J.M. Theoretical Considerations of the Influence of Polymer Film Coatings on the Mechanical Strength of Tablets. *J. Pharm. Pharmacol.* **1981**, *33*, 557–560.
4. Roy, G.M. Taste Masking in Oral Pharmaceuticals. *Pharm. Tech.* **1994**, *18* (Apr), 84–97.
5. Tang, L.; Schwartz, J.B.; Porter, S.C.; Schnaare, R.L.; Wigent, R.J. Drug Release from Film-Coated Chlorpheniramine Maleate Nonpareil Beads: Effect of Water-Soluble Polymer, Coating Level, and Soluble Core Material. *Pharm. Dev. Technol.* **2000**, *5* (3), 383–390.
6. Ghebre-Sellassie, I.; Iyer, U.; Kubert, D.; Fawzi, M.B. Characterization of a New Water-Based Coating for Modified-Release Preparations. *Pharm. Tech.* **1988**, *12*(Sep), 96–106.
7. Gazzaniga, A.; Iamartino, P.; Maffione, G.; Sangalli, M.E. Oral Delayed-Released System for Colonic Specific Delivery. *Int. J. Pharm.* **1994**, *108*, 77–83.
8. Felton, L.A.; Haase, M.M.; Shah, N.H.; Zhang, G.; Infeld, M.H.; Malick, A.W.; McGinity, J.W. Physical and Enteric Properties of Soft Gelatin Capsules Coated with Eudragit L 30 D-55. *Int. J. Pharm.* **1995**, *113*, 17–24.
9. Felton, L.A.; Shah, N.H.; Zhang, G.; Infeld, M.H.; Malick, A.W.; McGinity, J.W. Physical-Mechanical Properties of Film-Coated Soft Gelatin Capsules. *Int. J. Pharm.* **1996**, *127*, 203–211.
10. Felton, L.A.; Perry, W.L. A Novel Technique to Quantify Film-Tablet Interfacial Thickness. *Pharm. Dev. Technol.* **2002**, *7* (1), 1–5.
11. Hutchings, D.; Kuzmak, B.; Sakr, A. Processing Considerations for an EC Latex Coating System: Influence of Curing Time and Temperature. *Pharm. Res.* **1994**, *11* (10), 1474–1478.
12. Lippold, B.H.; Sutter, B.K.; Lippold, B.C. Parameters Controlling Drug Release from Pellets Coated with Aqueous Ethyl Cellulose Dispersions. *Int. J. Pharm.* **1989**, *54*, 15–25.
13. Bodmeier, R.; Paeratakul, O. The Effect of Curing on Drug Release and Morphological Properties of Ethylcellulose Pseudolatex-Coated Beads. *Drug Dev. Ind. Pharm.* **1994**, *20* (9), 1517–1533.
14. Felton, L.A.; Baca, M.L. Influence of Curing on the Adhesive and Mechanical Properties of an Applied Acrylic Polymer. *Pharm. Dev. Technol.* **2001**, *6* (1), 1–9.
15. Amighi, K.; Moes, A. Influence of Plasticizer Concentration and Storage Conditions on the Drug Release Rate from Eudragit RD30D Film-Coated Sustained-Release Theophylline Pellets. *Eur. J. Pharm. Biopharm.* **1996**, *42* (1), 29–35.
16. Williams, R.O., III; Liu, J. Influence of Processing and Curing Conditions on Beads Coated with an Aqueous Dispersion of Cellulose Acetate Phthalate. *Eur. J. Pharm. Biopharm.* **2000**, *49*, 243–252.
17. Frohoff-Hulsmann, M.A.; Schmitz, A.; Lippold, B.C. Aqueous Ethyl Cellulose Dispersions Containing Plasticizers of Different Water Solubility and Hydroxypropyl Methylcellulose as Coating Material for Diffusion Pellets. I. Drug Release Rates from Coated Pellets. *Int. J. Pharm.* **1999**, *177*, 69–82.
18. Wessling, M.; Bodmeier, R. Influence of Plasticization Time, Curing Conditions, Storage Time, and Core Properties on the Drug Release from Aquacoat-Coated Pellets. *Pharm. Dev. Technol.* **2001**, *6* (3), 325–331.
19. Mehta, A. Processing and Equipment Considerations for Aqueous Coatings. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*, 2nd Ed.; McGinity, J.W., Ed.; Marcel Dekker, Inc.: New York, 1997; 287–326.
20. Bauer, K.H.; Lehmann, K.; Osterwald, H.P.; Rothgang, G. Equipment for Sugar Coating and Film Coating Processes. In *Coated Pharmaceutical Dosage Forms*; CRC Press: Boca Raton, FL, 1998; 153–184.
21. Christensen, F.N.; Bertelsen, P. Qualitative Description of the Wurster-Based Fluid-Bed Coating Process. *Drug Dev. Ind. Pharm.* **1997**, *23* (5), 451–463.
22. Harris, M.R.; Ghebre-Sellassie, I.; Nesbitt, R.U. A Water-Based Coating Process for Sustained Release. *Pharm. Tech.* **1986**, *10*(Sep), 102–107.
23. Jones, D. Air Suspension Coating for Multiparticulates. *Drug Dev. Ind. Pharm.* **1994**, *20* (20), 3175–3206.
24. Hossain, M.; Ayres, J.W. Variables That Influence Coat Integrity in a Laboratory Spray Coater. *Pharm. Tech.* **1990**, *14*(Oct), 72–82.
25. Wesdyk, R.; Joshi, Y.M.; Jain, N.B.; Morris, K.; Newman, A. The Effect of Size and Mass on the Film Thickness of Beads Coated in Fluidized Bed Equipment. *Int. J. Pharm.* **1990**, *65*, 69–76.

26. Obara, S.; Maruyama, N.; Nishiyama, Y.; Kokubo, H. Dry Coating: An Innovative Enteric Coating Method Using a Cellulose Derivative. *Eur. J. Pharm. Biopharm.* **1999**, *47*, 51–59.
27. Houlton, S. Electrostatics in Continuous Tablet Coating. *Manuf. Chem.* **1998**, *69*, 13–16.
28. Takka, S.; Rajbhandari, S.; Sakr, A. Effect of Anionic Polymers on the Release of Propranolol Hydrochloride from Matrix Tablets. *Eur. J. Pharm. Biopharm.* **2001**, *52*, 75–82.
29. Levine, M.M.; Ferreccio, C.; Cryz, S.; Ortiz, E. Comparison of Enteric-Coated Capsules and Liquid Formulation of Ty21a Typhoid Vaccine in Randomized Controlled Field Trial. *Lancet* **1990**, *336*, 891–894.
30. Petroski, D. Endoscopic Comparison of Various Aspirin Preparations. Gastric Mucosal Adaptability to Aspirin Restudied. *Curr. Ther. Res. Clin. Exp.* **1989**, *45*, 945–954.
31. Banker, G.S.; Gore, A.Y.; Swarbrick, J. Water Vapour Transmission Properties of Applied Polymer Films. *J. Pharm. Pharmacol.* **1966**, *18* (Suppl.), 205S–211S.
32. Bodmeier, R.; Paeratakul, O. Propranolol HCL Release from Acrylic Films Prepared from Aqueous Latexes. *Int. J. Pharm.* **1990**, *59*, 197–204.
33. Gutierrez-Rocca, J.C.; McGinity, J.W. Influence of Water Soluble and Insoluble Plasticizers on the Physical and Mechanical Properties of Acrylic Resin Copolymers. *Int. J. Pharm.* **1994**, *103*, 293–301.
34. Banker, G.S. Film Coating Theory and Practice. *J. Pharm. Sci.* **1966**, *55* (1), 81–89.
35. Bodmeier, R.; Paeratakul, O. Mechanical Properties of Dry and Wet Cellulosic and Acrylic Films Prepared from Aqueous Colloidal Polymer Dispersions Used in the Coating of Solid Dosage Forms. *Pharm. Res.* **1994**, *11* (6), 882–888.
36. Lindholm, T.; Juslin, M.; Lindholm, B.-A.; Poikala, M.; Tiilikainen, S.; Varis, H. Properties of Free Ethyl Cellulose Films Containing Surfactant and Particulate Matter. *Pharm. Ind.* **1987**, *49* (7), 740–746.
37. Wu, C.; McGinity, J.W. Non-traditional Plasticization of Polymeric Films. *Int. J. Pharm.* **1999**, *177*, 15–27.
38. Iyer, U.; Hong, W.-H.; Das, N.; Ghebre-Sellassie, I. Comparative Evaluation of Three Organic Solvent and Dispersion-Based Ethylcellulose Coating Formulations. *Pharm. Technol.* **1990**, *14*, 68–86.
39. Bodmeier, R.; Paeratakul, O. The Distribution of Plasticizers Between Aqueous and Polymer Phases in Aqueous Colloidal Polymer Dispersions. *Int. J. Pharm.* **1994**, *103*, 47–54.
40. Lehmann, K.O.R. Chemistry and Application Properties of Polymethacrylate Coating Systems. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*, 2nd Ed.; McGinity, J.W., Ed.; Marcel Dekker, Inc.: New York, 1997; 101–176.
41. Siepmann, J.; Paeratakul, O.; Bodmeier, R. Modeling Plasticizer Uptake in Aqueous Polymer Dispersions. *Int. J. Pharm.* **1998**, *165*, 191–200.
42. Sakellariou, P.; Hassan, H.; Rowe, R.C. Plasticization and Interactions of Polyethylene Glycol 6000 with Hydroxypropyl Methylcellulose/Polyvinyl Alcohol Blends. *Int. J. Pharm.* **1994**, *102*, 207–211.
43. Tait, P.J.T.; Abushihada, A.M. Comparative Studies on the Use of Gas Chromatographic and Vapour Pressure Techniques for the Determination of the Interaction Energy Parameter. *Polymer* **1977**, *18*, 810–816.
44. Spurlin, H.M.; Martin, A.F.; Tennent, H.G. Characterization of Cellulose Derivatives by Solution Properties: Plasticizers as Solvents. *J. Polym. Sci.* **1946**, *1* (2), 63–74.
45. Chi, M.S. Compatibility of Cross-Linked Polymers with Plasticizers by Glass Transition Temperature Measurement and Swelling Tests. *J. Polym. Sci. Pol. Chem. Ed.* **1981**, *19*, 1767–1779.
46. Edelman, A.; Fradet, A. Inverse Gas Chromatography Study of Some Triacetin–Polymer Systems. *Polymer* **1989**, *30*, 317–323.
47. Entwistle, C.A.; Rowe, R.C. Plasticization of Cellulose Ethers Used in the Film Coating of Tablets. *J. Pharm. Pharmacol.* **1978**, *31*, 269–272.
48. Anagnostopoulos, C.E.; Coran, A.Y. Polymer–Diluent Interactions. II. Polyvinyl Chloride–Diluent Interactions. *J. Polym. Sci.* **1962**, *57*, 1–11.
49. Belfiore, L.A.; Cooper, S.L. Thermodynamic Miscibility in Polymer–Polymer and Polymer–Diluent Blends. *J. Polym. Sci. Polym. Phys. Ed.* **1983**, *21*, 114–115.
50. Theodorou, M.; Jasse, B. Fourier-Transform Infrared Study of Conformational Changes in Plasticized Poly-(Vinyl Chloride). *J. Polym. Sci. Polym. Phys. Ed.* **1983**, *21*, 2263–2274.
51. Johnson, K.; Hathaway, R.; Leung, P.; Franz, R. Effect of Triacetin and Polyethylene Glycol 400 on Some Physical Properties of Hydroxypropyl Methylcellulose Free Films. *Int. J. Pharm.* **1991**, *73*, 197–208.
52. Verhoeven, J.; Schaeffer, R.; Bouwstra, J.A.; Junginger, H.E. The Physico-chemical Characterization of Poly(2-Hydroxyethyl Methacrylate-co-Methacrylic Acid): 2. Effect of Water, PEG 400 and PEG 6000 on the Glass Transition Temperature. *Polymer* **1989**, *30*, 1946–1950.
53. Jenquin, M.R.; Sarabia, R.E.; Liebowitz, S.M.; McGinity, J.W. Relationship of Film Properties to Drug Release from Monolithic Films Containing Adjuvants. *J. Pharm. Sci.* **1992**, *81* (10), 983–989.
54. Maejima, T.; McGinity, J.W. Influence of Film Additives on Stabilizing Drug Release Rates from Pellets Coated with Acrylic Polymers. *Pharm. Dev. Technol.* **2001**, *6* (2), 211–221.
55. Bianchini, R.; Resciniti, M.; Vecchio, C. Technological Evaluation of Aqueous Enteric Coating Systems with and without Insoluble Additives. *Drug Dev. Ind. Pharm.* **1991**, *17* (13), 1779–1794.
56. Felton, L.A.; McGinity, J.W. Influence of Insoluble Excipients on Film Coating Systems. *Drug Dev. Ind. Pharm.* **2002**, *28*, 225–243.



57. Wesseling, M.; Kuppler, F.; Bodmeier, R. Tackiness of Acrylic and Cellulosic Polymer Films Used in the Coating of Solid Dosage Forms. *Eur. J. Pharm. Biopharm.* **1999**, *47*, 73–78.
58. Fassihi, R.A.; McPhillips, A.M.; Uraizee, S.A.; Sakr, A.M. Potential Use of Magnesium Stearate and Talc as Dissolution Retardants in the Development of Controlled Drug Delivery Systems. *Pharm. Ind.* **1994**, *56* (6), 579–583.
59. Chang, R.K.; Hsiao, C. Eudragit RL and RS Pseudolat- tices: Properties and Performance in Pharmaceutical Coating as a Controlled Release Membrane for Theophyl- line Pellets. *Drug Dev. Ind. Pharm.* **1989**, *15* (2), 187–196.
60. Okhamafe, A.O.; York, P. Effect of Solids–Polymer Interactions on the Properties of Some Aqueous-Based Tablet Film Coating Formulations. II. Mechanical Characteristics. *Int. J. Pharm.* **1984**, *22*, 273–281.
61. Okhamafe, A.O.; York, P. Relationship Between Stress, Interaction and the Mechanical Properties of Some Pigmented Tablet Coating Films. *Drug Dev. Ind. Pharm.* **1985**, *11* (1), 131–146.
62. Petereit, H.-U.; Assmus, M.; Lehmann, K. Glyceryl Monostearate as a Glidant in Aqueous Film-Coating Formulations. *Eur. J. Pharm. Biopharm.* **1995**, *41* (4), 219–228.
63. Porter, S.C. Tablet Coating. *Drug Cosmet. Ind.* **1981**, *128* (5), 46–93.
64. Matsuda, Y.; Itooka, T.; Mitsuhashi, Y. Photostability of Indomethacin in Model Gelatin Capsules: Effects of Film Thickness and Concentration of Titanium Dioxide on the Coloration and Photolytic Degradation. *Chem. Pharm. Bull.* **1980**, *28* (9), 2665–2671.
65. Rowe, R.C. The Prediction of Compatibility/Incompat- ibility in Blends of Ethyl Cellulose with Hydroxypropyl Methylcellulose or Hydroxypropyl Cellulose using 2- Dimensional Solubility Parameter Maps. *J. Pharm. Pharmacol.* **1986**, *38*, 214–215.
66. Nyamweya, N.; Mehta, K.A.; Hoag, S.W. Characteriz- ation of the Interactions Between Polymethacrylate-Based Aqueous Polymeric Dispersions and Aluminum Lakes. *J. Pharm. Sci.* **2001**, *90*, 1–11.
67. Lehmann, K.; Rothgang, G.; Bossler, H.; Dreher, D.; Petereit, H.U.; Liddiard, C.; Weisbrod, W. *Practical Course in Lacquer Coating*; Rohm Pharma, GmbH: Weiterstadt, 1989.
68. Wong, D. Practical Considerations for Stabilization of Aqueous Colloidal Polymer Dispersions. *Pharm. Tech.* **1997**, *21* (May), 38–50.
69. Felton, L.A.; McGinity, J.W. Influence of Pigment Concentration and Particle Size on Adhesion of an Acrylic Resin Copolymer to Tablet Compacts. *Drug Dev. Ind. Pharm.* **1999**, *25* (5), 599–606.
70. Rowe, R.C. Modulus Enhancement in Pigmented Tablet Film Coating Formulations. *Int. J. Pharm.* **1983**, *14*, 355–359.
71. Maul, K.A.; Schmidt, P.C. Influence of Different-Shaped Pigments and Plasticizers on Theophylline Release from Eudragit RS30D and Aquacoat ECD30 Coated Pellets. *STP Pharma Sci.* **1997**, *7* (6), 498–506.
72. Parker, J.W.; Peck, C.E.; Banker, G.S. Effects of Solids- Loading on Moisture Permeability Coefficients of Free Films. *J. Pharm. Sci.* **1974**, *63* (1), 119–125.
73. Lehtola, V.-M.; Heinamaki, J.T.; Nikupaavo, P.; Yliruusi, J.K. The Mechanical and Adhesion Properties of Aqueous-Based Hydroxypropyl Methylcellulose Coating Systems Containing Polydextrose and Titanium Dioxide. *Drug Dev. Ind. Pharm.* **1995**, *21* (6), 675–685.
74. Felton, L.A.; Austin-Forbes, T.; Moore, T.A. Influence of Surfactants in Aqueous-Based Polymeric Dispersions on the Thermo-Mechanical and Adhesive Properties of Acrylic Films. *Drug Dev. Ind. Pharm.* **2000**, *26* (2), 205–210.
75. Lehmann, K. The Application and Processing of Acrylic Coatings in Form of Aqueous Dispersions Compared with Organic Solutions. *Acta Pharm. Fenn.* **1982**, *91* (4), 225–238.
76. Lindholm, T.; Lindholm, B.-A.; Niskanen, M.; Koski- niemi, J. Polysorbate 20 as a Drug Release Regulator in Ethyl Cellulose Film Coatings. *J. Pharm. Pharmacol.* **1986**, *38*, 686–688.
77. Hanson, W.A. *Handbook of Dissolution Testing*; Aster Publishing Corporation: Eugene, 1991.
78. Li, S.P.; Feld, K.M.; Kowarski, C.R. Preparation and Evaluation of Eudragit Acrylic Resin for Controlled Drug Release of Pseudoephedrine Hydrochloride. *Drug Dev. Ind. Pharm.* **1991**, *17* (12), 1655–1683.
79. Goodhart, F.W.; Harris, M.R.; Murthy, K.S.; Nesbitt, R.U. An Evaluation of Aqueous Film-Forming Dispersions for Controlled Release. *Pharm. Tech.* **1984**, *8* (Apr), 64–71.
80. Tan, Y.T.F.; Heng, P.W.S.; Wan, L.S.C. Mixed Polymer Coating for Modified Release from Coated Spheroids. *Pharm. Dev. Technol.* **1999**, *4* (4), 561–570.
81. Li, S.P.; Mehta, G.N.; Buehler, J.D.; Grim, W.M.; Harwood, R.J. The Effect of Film-Coating Additives on the In Vitro Dissolution Release Rate of Ethyl Cellulose- Coated Theophylline Granules. *Pharm. Tech.* **1990**, *14* (Mar), 20–24.
82. Lorck, C.A.; Grunenberg, P.C.; Junger, H.; Laicher, A. Influence of Process Parameters on Sustained-Release Theophylline Pellets Coated with Aqueous Polymer Dispersions and Organic Solvent-Based Polymer Sol- utions. *Eur. J. Pharm. Biopharm.* **1997**, *43*, 149–157.
83. Munday, D.L.; Fassihi, A.R. Changes in Drug Release Rate: Effect of Stress Storage Conditions on Film Coated Mini-Tablets. *Drug Dev. Ind. Pharm.* **1991**, *17* (15), 2135–2143.
84. Heinamaki, J.T.; Lehtola, V.-M.; Nikupaavo, P.; Yliruusi, J.K. The Mechanical and Moisture Permeability Prop- erties of Aqueous-Based Hydroxypropyl Methylcellulose Coating Systems Plasticized with Polyethylene Glycol. *Int. J. Pharm.* **1994**, *112*, 191–196.

85. Obara, S.; McGinity, J.W. Properties of Free Films Prepared from Aqueous Polymers by a Spraying Technique. *Pharm. Res.* **1994**, *11* (11), 1562–1567.
86. Udeala, O.K.; Aly, S.A.S. Degradation Kinetics of Thiamine Hydrochloride in Directly Compressed Tablets. III. Water Vapour Transmission Through Free and Applied Eudragit Films. *Drug Dev. Ind. Pharm.* **1989**, *15* (11), 1797–1825.
87. Banker, G.S.; Gore, A.Y.; Swarbrick, J. Water Vapour Transmission Properties of Free Polymer Films. *J. Pharm. Pharmacol.* **1965**, *18*, 457–466.
88. Nagai, T.; Obara, S.; Kokubo, H.; Hoshi, N. Application of HPMC and HPMCAS to Aqueous Film Coating of Pharmaceutical Dosage Forms. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*, 2nd Ed.; McGinity, J.W., Ed.; Marcel Dekker, Inc.: New York, 1997; 177–225.
89. Baert, L.; Remon, J.P. Water Vapour Permeation of Aqueous Based Ethylacrylate Methylmethacrylate Copolymer Films. *Int. J. Pharm.* **1993**, *99*, 181–187.
90. Guo, J.-H. Effects of Plasticizers on Water Permeation and Mechanical Properties of Cellulose Acetate: Antiplasticization in Slightly Plasticized Polymer Film. *Drug Dev. Ind. Pharm.* **1993**, *19* (13), 1541–1555.
91. Benita, S.; Dor, P.; Aronhime, M.; Marom, G. Permeability and Mechanical Properties of a New Polymer: Cellulose Hydrogen Phthalate. *Int. J. Pharm.* **1986**, *33*, 71–80.
92. Hatakeyama, T.; Quinn, F.X. *Thermal Analysis. Fundamentals and Applications to Polymer Science*; John Wiley & Sons, Ltd.: Chichester, 1994.
93. O'Donnell, P.B.; McGinity, J.W. Mechanical Properties of Polymeric Films Prepared From Aqueous Polymeric Dispersions. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*, 2nd Ed.; McGinity, J.W., Ed.; Marcel Dekker, Inc.: New York, 1997; 517–548.
94. Okhamafe, A.O.; York, P. The Glass Transition in Some Pigmented Polymer Systems Used for Tablet Coating. *J. Macromol. Sci. Phys.* **1984–1985**, *B23* (4–6), 373–382.
95. Okhamafe, A.O.; York, P. Thermal Characterization of Drug/Polymer and Excipient/Polymer Interactions in Some Film Coating Formulations. *J. Pharm. Pharmacol.* **1989**, *41*, 1–6.
96. Hjartstam, J.; Borg, K.; Lindstedt, B. The Effect of Tensile Stress on Permeability of Free Films of Ethyl Cellulose Containing Hydroxypropyl Methylcellulose. *Int. J. Pharm.* **1990**, *61*, 101–107.
97. Dittgen, M. Relationship Between Film Properties and Drug Release from Acrylic Films. *Drug Dev. Ind. Pharm.* **1985**, *11* (Feb–Mar), 269–279.
98. Lin, S.-Y.; Lee, C.-L.; Lin, Y.-Y. Drug–Polymer Interaction Affecting the Mechanical Properties, Adhesion Strength and Release Kinetics of Piroxicam-Loaded Eudragit E Films Plasticized with Different Plasticizers. *J. Control. Release* **1995**, *33*, 375–381.
99. Rowe, R.C. The Cracking of Film Coatings on Film-Coated Tablets—A Theoretical Approach with Practical Implications. *J. Pharm. Pharmacol.* **1981**, *33*, 423–426.
100. Guo, J.-H.; Robertson, R.E.; Amidon, G.L. Influence of Physical Aging on Mechanical Properties of Polymer Free Films: The Prediction of Long-Term Aging Effects on the Water Permeability and Dissolution Rate of Polymer Film-Coated Tablets. *Pharm. Res.* **1991**, *8* (12), 1500–1504.
101. Rowe, R.C. Microindentation—A Method for Measuring the Elastic Properties and Hardness of Films on Conventionally Coated Tablets. *J. Pharm. Pharmacol.* **1976**, *28*, 310–311.
102. Bodmeier, R.; Paeratakul, O. Dry and Wet Strengths of Polymeric Films Prepared from an Aqueous Colloidal Polymer Dispersion, Eudragit RS30D. *Int. J. Pharm.* **1993**, *96*, 129–138.
103. Gibson, S.H.M.; Rowe, R.C.; White, E.F.T. The Mechanical Properties of Pigmented Tablet Coating Formulations and Their Resistance to Cracking. II. Dynamic Mechanical Measurement. *Int. J. Pharm.* **1989**, *50*, 163–173.
104. Aulton, M.E.; Abdul-Razzak, M.H. The Mechanical Properties of Hydroxypropylmethylcellulose Films Derived from Aqueous Systems. Part 2: The Influence of Solid Inclusions. *Drug Dev. Ind. Pharm.* **1984**, *10* (4), 541–561.
105. Obara, S.; McGinity, J.W. Influence of Processing Variables on the Properties of Free Films Prepared from Aqueous Polymeric Dispersions by a Spray Technique. *Int. J. Pharm.* **1995**, *126*, 1–10.
106. Lin, S.Y.; Lee, C.J.; Lin, Y.Y. The Effect of Plasticizers on Compatibility, Mechanical Properties, and Adhesion Strength of Drug-Free Eudragit E Films. *Pharm. Res.* **1991**, *8*, 1137–1143.
107. Hutchings, D.; Clarson, S.; Sakr, A. Studies of the Mechanical Properties of Free Films Prepared Using an Ethylcellulose Pseudolatex Coating System. *Int. J. Pharm.* **1994**, *104*, 203–213.
108. Gutierrez-Rocca, J.C.; McGinity, J.W. Influence of Aging on the Physical–Mechanical Properties of Acrylic Resin Films Cast from Aqueous Dispersions and Organic Solutions. *Drug Dev. Ind. Pharm.* **1993**, *19* (3), 315–332.
109. Wu, C.; McGinity, J.W. Influence of Relative Humidity on the Mechanical and Drug Release Properties of Theophylline Pellets Coated with an Acrylic Polymer Containing Methylparaben as a Non-traditional Plasticizer. *Eur. J. Pharm. Biopharm.* **2000**, *50*, 277–284.
110. Porter, S.C. The Effect of Additives on the Properties of an Aqueous Film Coating. *Pharm. Tech.* **1980**, *4*(Mar), 67–75.
111. Maganti, L.; Celik, M. Compaction Studies on Pellets: II. Coated Pellets. *Int. J. Pharm.* **1994**, *103*, 55–67.
112. Bodmeier, R. Tableting of Coated Pellets. *Eur. J. Pharm. Biopharm.* **1997**, *43*, 1–8.



113. Aulton, M.E.; Dyer, A.M.; Khan, K.A. The Design of a Controlled-Release Drug Delivery System for Ibuprofen in the Form of a Tablet Comprising Compacted Polymer-Coated Millispheres. *Drug Dev. Ind. Pharm.* **1994**, *20* (20), 3069–3104.
114. Felton, L.A.; Shah, N.H.; Zhang, G.; Infeld, M.H.; Malick, A.W.; McGinity, J.W. Compaction Properties of Individual Non-pareil Beads Coated with an Acrylic Resin Copolymer. *STP Pharma. Sci.* **1997**, *7* (6), 457–462.
115. Bechard, S.R.; Leroux, J.C. Coated Pelletized Dosage Form: Effect of Compaction on Drug Release. *Drug Dev. Ind. Pharm.* **1992**, *18* (18), 1927–1944.
116. Lopez-Rodriguez, F.J.; Torrado, J.J.; Torrado, S.; Escamilla, C.; Cadorniga, R.; Augsburger, L.L. Compression Behavior of Acetylsalicylic Acid Pellets. *Drug Dev. Ind. Pharm.* **1993**, *19* (12), 1369–1377.
117. Felton, L.A.; McGinity, J.W. Adhesion of Polymeric Films to Pharmaceutical Solids. *Eur. J. Pharm. Biopharm.* **1999**, *47* (1), 1–14.
118. Okhamafe, A.O.; York, P. The Adhesion Characteristics of Some Pigmented and Unpigmented Aqueous-Based Film Coatings Applied to Aspirin Tablets. *J. Pharm. Pharmacol.* **1985**, *37*, 849–853.
119. Felton, L.A.; McGinity, J.W. Influence of Tablet Hardness and Hydrophobicity on the Adhesive Properties of an Acrylic Resin Copolymer. *Pharm. Dev. Technol.* **1996**, *1* (4), 381–389.
120. Rowe, R.C. A Reappraisal of the Equations Used to Predict the Internal Stresses in Film Coatings Applied to Tablet Substrates. *J. Pharm. Pharmacol.* **1983**, *35*, 112–113.
121. Okutgen, E.; Hogan, J.E.; Aulton, M.E. Quantitative Estimation of Internal Stress Development in Aqueous HPMC Tablet Film Coats. *Int. J. Pharm.* **1995**, *119*, 193–202.
122. Wood, J.A.; Harder, S.W. The Adhesion of Film Coatings to the Surfaces of Compressed Tablet. *Can. J. Pharm. Sci.* **1970**, *5* (1), 18–23.
123. Fisher, D.G.; Rowe, R.C. The Adhesion of Film Coatings to Tablet Surfaces—Instrumentation and Preliminary Evaluation. *J. Pharm. Pharmacol.* **1976**, *28*, 886–889.
124. Rowe, R.C. The Adhesion of Film Coatings to Tablet Surfaces—The Effect of Some Direct Compression Excipients and Lubricants. *J. Pharm. Pharmacol.* **1977**, *29*, 723–726.
125. Lehtola, V.-M.; Heinamaki, J.T.; Nikupaavo, P.; Yliruusi, J.K. Effect of Some Excipients and Compression Pressure on the Adhesion of Aqueous-Based Hydroxypropyl Methylcellulose Film Coatings to Tablet Surface. *Drug Dev. Ind. Pharm.* **1995**, *21* (12), 1365–1375.
126. Fung, R.M.; Parrott, E.L. Measurement of Film-Coating Adhesiveness. *J. Pharm. Sci.* **1980**, *69* (4), 439–441.
127. Nadkarni, P.D.; Kildsig, D.O.; Kramer, P.A.; Banker, G.S. Effects of Surface Roughness and Coating Solvent on Film Adhesion to Tablets. *J. Pharm. Sci.* **1975**, *64*, 1554–1557.
128. Rowe, R.C. Adhesion of Film Coatings to Tablet Surfaces—A Theoretical Approach Based on Solubility Parameters. *Int. J. Pharm.* **1988**, *41*, 219–222.
129. Felton, L.A.; McGinity, J.W. Influence of Plasticizers on the Adhesive Properties of an Acrylic Resin Copolymer to Hydrophilic and Hydrophobic Tablet Compacts. *Int. J. Pharm.* **1997**, *154*, 167–178.
130. Okutgen, E.; Hogan, J.E.; Aulton, M.E. Effects of Tablet Core Dimensional Instability on the Generation of Internal Stresses Within Film Coats. Part 1. Influence of Temperature Changes During the Film Coating Process. *Drug Dev. Ind. Pharm.* **1991**, *17* (9), 1177–1189.
131. Rowe, R.C. The Adhesion of Film Coatings to Tablet Surfaces—A Problem of Stress Distribution. *J. Pharm. Pharmacol.* **1981**, *33*, 610–612.
132. Rowe, R.C.; Forse, S.F. The Effect of Polymer Molecular Weight on the Incidence of Film Cracking and Splitting on Film Coated Tablets. *J. Pharm. Pharmacol.* **1980**, *32*, 583.
133. Rowe, R.C.; Upjohn, N.G. An Expert System for the Identification and Solution of Film-Coating Defects. *Pharm. Tech.* **1993**, *5* (Sep), 130–138.
134. Signorino, C.A.; Forcellini, L.J. Evaluating the Uniformity of Aqueous Film Coating. *Pharm. Tech.* **1996**, *20*, (Suppl.) 48–53.
135. Rowe, R.C. Defects in Film-Coated Tablets: Aetiology and Solutions. In *Advances in Pharmaceutical Sciences*; Ganderton, D., Jones, T. Eds.; Academic Press: London, 1992; Vol. 6, 65–99.
136. Lin, S.Y.; Perng, R.I. Solid-State Interaction Studies of Drugs/Polymers. I. Indomethacin/Eudragit E, RL or S Resins. *STP Pharma Sci.* **1993**, *3* (6), 465–471.
137. Okhamafe, A.O.; York, P. Studies on Interaction Phenomena in Aqueous-Based Film Coatings Containing Soluble Additives Using Thermal Analysis Techniques. *J. Pharm. Sci.* **1988**, *77* (5), 438–443.
138. Sarisuta, N.; Kumpugdee, M.; Muller, B.W.; Puttipatkhachorn, S. Physico-chemical Characterization of Interactions Between Erythromycin and Various Film Polymers. *Int. J. Pharm.* **1999**, *186*, 109–118.
139. Rowe, R.C.; Forse, S.F. Pitting—A Defect on Film-Coated Tablets. *Int. J. Pharm.* **1983**, *17*, 347–349.
140. Schmid, S.; Muller-Goymann, C.C.; Schmidt, P.C. Interactions During Aqueous Film Coating of Ibuprofen with Aquacoat ECD. *Int. J. Pharm.* **2000**, *197*, 35–39.
141. Porter, S.C. Use of Opadry, Sureteric, and Surelease for the Aqueous Film Coating of Pharmaceutical Oral Dosage Forms. In *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*, 2nd Ed.; McGinity, J.W., Ed.; Marcel Dekker, Inc.: New York, 1997; 327–372.
142. Murthy, K.S.; Enders, N.A.; Mahjour, M.; Fawzi, M.B. A Comparative Evaluation of Aqueous Enteric Polymers in Capsule Coatings. *Pharm. Tech.* **1986**, *10* (Oct), 36–46.
143. Thoma, K.; Oschmann, R. Investigations of the Permeability of Enteric Coatings. Part 5: Pharmaceutical—Technological and Analytical Studies of Enteric-Coated Preparations. *Pharmazie* **1991**, *46*, 278–282.



144. Osterwald, H.P. Experience with Coating of Gelatin Capsules with Driacoater and WSG Apparatus, Especially Rotor WSG. *Acta Pharm. Technol.* **1982**, 28, 329–337.
145. Ali, M.S.; Sheldon, R.P. Structural Changes in Glassy Polymers. *J. Appl. Polym. Sci.* **1970**, 14, 2619–2628.
146. Delporte, J.P. Effects of Aging on Physico-chemical Properties of Free Cellulose Acetate Phthalate Films. *Pharm. Ind.* **1979**, 41 (10), 984–990.
147. Sinko, C.M.; Yee, A.F.; Amidon, G.L. The Effect of Physical Aging on the Dissolution Rate of Anionic Polyelectrolytes. *Pharm. Res.* **1990**, 7 (6), 648–653.
148. Guo, J.-H.; Robertson, R.E.; Amidon, G.L. An Investigation into the Mechanical and Transport Properties of Aqueous Latex Films: A New Hypothesis for Film-Forming Mechanism of Aqueous Dispersion System. *Pharm. Res.* **1993**, 10 (3), 405–410.

Freeze Drying, Scale-Up Considerations

Edward H. Trappier

Lyophilization Technology, Inc., Ivyland, Pennsylvania, U.S.A.

INTRODUCTION

Product presentation, formulation, and processing are integral in their influences on the outcome of manufacturing a lyophilized product. Activities in development involve knowledge of the clinical and stability requirements, formulation and product design, and investigation into suitable processing conditions. These activities incorporate the awareness that the product is expected to be integrated into large-scale routine manufacturing of commercial product. Multiple steps and increasing batch sizes are often necessary as the product progresses through clinical studies and market introductory batches. Aspects of equipment, processing capabilities, processing influences, and finished product attributes all need to be considered in the scale-up to increasingly larger batch sizes on the pathway of taking a product to market.

DEVELOPMENT AND SCALE-UP PATHWAY

In the progression of bringing a new product to the market, scale-up from the initial development batches to commercial manufacturing is an important activity. This progression includes transferring the technology for producing a product after sufficient initial development, eventually integrating the procedures into a commercial manufacturing environment. As part of later development activities, materials for stability and early clinical studies are commonly prepared on a small scale. This comprises the first transfer of the processing methods into a manufacturing environment. Following successful clinical results, the product presentation and any refinements to the manufacturing procedures are implemented, and the process is duplicated on a larger scale in a manufacturing environment for full commercial production.

With an approved investigational new drug (IND), a new drug entity is developed into a new product with administration to human patients for establishing safety, efficacy, and dosing regime in phase I through phase III clinical studies. Initial development activities focus on preparing materials for conducting these studies. Meeting the needs of supplying the clinical material often results in initial product presentations that may be different than the final product that is introduced to the market. As clinical

trials progress, active pharmaceutical ingredient (API) manufacturing, analytical methods, product design, and manufacturing procedures are developed and refined. With the success of early clinical trials and entering into phase III clinical studies, product preparation and specifications are finalized. The API is well characterized, analytical methods are validated, product presentation and specifications finalized, as well as finished product processing procedures established.

Initial development activities are focused on preparing to supply early clinical requirements and product needs in designing the product formulation and presentation. The product presentation is based upon optimal administration to the patient in a clinical setting and stability requirements for the active ingredient. Lyophilization is used as a method of preservation for products that are sensitive to the presence of water and have limited stability as a liquid ready-to-use preparation. This low temperature vacuum drying process removes the water, leaving a dried solid, resulting in sufficient stability to allow long-term storage. At the time of use, the product is reconstituted with a diluent that returns all the attributes of a liquid ready-to-use product, yielding the preparation suitable for administration. One advantage to a lyophilized preparation is the different concentrations of API that may be obtained by varying the diluent volume for reconstitution, along with the volume used for administration, to allow for different dosages from a single preparation. Once the initial product presentation suitable for the early clinical studies is identified, processing procedures can then be developed. The focus for the initial lyophilization process is robustness and ease of scale-up rather than optimal parameters for routine manufacturing. In practice, the presentation initially developed may not necessarily be the final product marketed; the product presentation may change as a result of clinical studies. Continued development activities often then encompass a series of refinements to the product and procedures. In the pathway of bringing a new product to the market, scale-up and technology transfer may therefore occur multiple times.

As the product progresses through stages of development and clinical studies, product and process design are refined. Processing experience is gained on increasingly larger scale batches as the demands for clinical material increases. Finally, phase III clinical studies require

product quantities that may approach batch sizes closer to those that are typical of full-scale commercial manufacturing. Fig. 1 describes objectives and environment in the progression of steps in taking a product from initial development to manufacturing-scale operations. Initial success along with expanded indications throughout the life cycle of a product may continue to require larger batch sizes to meet increasing market demand. With the evolution of the product from development to clinical studies, followed by initial introduction and potential growth with market expansion, scale-up begins with smaller and progresses to increasingly larger batch sizes.

Design of a preparation and development of a suitable lyophilization process is first conducted in the laboratory. Product design focuses on the formulation and presentation. Stabilizing the API, imparting necessary features for patient administration, and constructing desirable dried product attributes are the objective in developing an appropriate formulation. Coupled to those objectives are the concentration, dictating product presentation parameters of fill volume, as well as selection of the container and closure. The formulation and presentation both influence the conditions required for the lyophilization process.

In development, variations in critical process parameters are explored. Preparation for integrating a new product and the associated new procedures into manufacturing warrants a thorough understanding of the product and methods necessary to produce material. These manufacturing methods need to be effective and adequately controllable for both a small scale for early clinical studies, later to be refined for batches to be processed in routine manufacturing of commercial product. Manufacturing capabilities may be an existing in-house operation or at another facility providing contracted manufac-

turing services. Where the finished product is produced often influences the accessibility of information in learning the capabilities of manufacturing and the ease of technology transfer during the scale-up.

The focus in initial development of a potential new product is designing a product and suitable processing methods for the early clinical studies. With clinical success yet established, lyophilization process development is focused on short-term objectives and the efforts limited: time is of the essence. Results of early clinical studies direct refining the product design and processing procedures as clinical trials progress. Priority is given to designing a product presentation that is suitable for the intended route of administration, convenient to use with sufficient flexibility in dose administration in the clinic, and provides sufficient stability for the duration of the clinical studies. Often with limited knowledge, experience, and amounts of API at the time of early clinical studies, there is little opportunity for extensive product design and process development. The objective is to focus on an adequate product design for completing the studies and sufficiently conservative process for ease of technology transfer and scale-up.

At these early development stages, an “optimized” process is one that is safe, robust, efficient, and effective with minimal risk of difficulties in producing clinical materials, and not necessarily what would be considered the shortest. Challenges of transferring the product from early development to producing early clinical supplies involve availability of raw material, appropriate product design, and manufacturing capability. Often at the time clinical studies are beginning, processing methods for raw material are still evolving, the API material is not yet well characterized, and is in limited supply. With pressures to

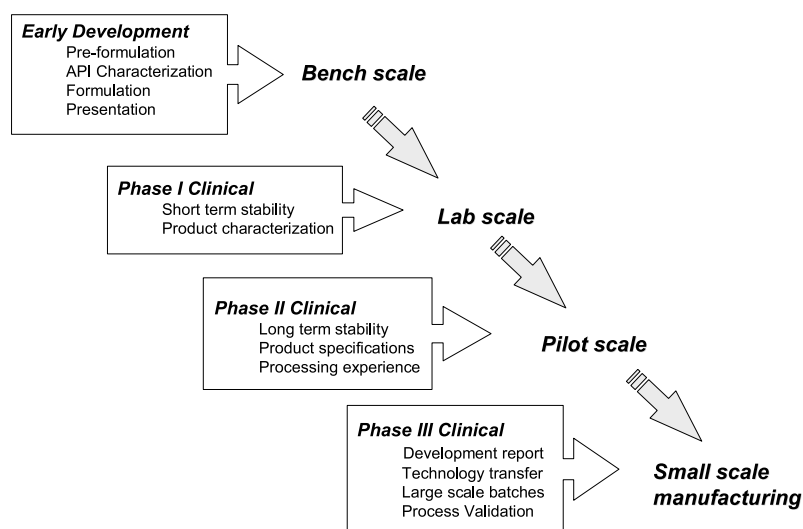


Fig. 1 Objectives and operations in progression to manufacturing scale. (View this art in color at www.dekker.com.)



begin clinical studies, the timeline of a project is often accelerated and the transfer to a manufacturing facility is considered the last link in a series of efforts. With decisions made on the basis of only a few development studies and limited data, the objective of the scale-up of early clinical manufacturing is to manage the level of risk with a process that is suitable but not necessarily conducive for large-scale commercial manufacturing.

As clinical studies progress, development efforts can begin to focus on the final product design and process parameters established through expanded studies in the laboratory. Objectives for these studies become improvements in the procedures that are well suited for routine processing in a manufacturing environment. These activities would also include process validation studies. The scope and extent of the validation studies expand as batch sizes increase with the progression from development to commercial manufacturing. Fig. 2 provides a graphic representation of the development efforts and activities along with the extent of validation appropriate as API production, finished product manufacturing, and analytical methods are refined in preparation for increasing batch sizes progressing toward large-scale commercial manufacturing.

The success of development efforts is measured by the ease of technology transfer and scale-up to large-scale manufacturing. Small-scale studies in the laboratory need to incorporate influencing factors of large-scale manufacturing. The intent is to develop a process that is safe,

effective, efficient, and sufficiently robust where the manufacturing technology can be transferred to routine large-scale commercial production. Product-design objects are efficacy, stability, safety, and ease of use. Process design entails establishing desirable finished product attributes and ease of reproducible manufacturing that consistently results in high quality and yield, at a low unit cost. The formulation, product design, and process all have an influence on each other, with consideration of the impact of one on another being of significant importance during development.

PRODUCT DESIGN

Characteristics of the product and functions in early clinical studies are the driving force for initial product development. Often, there is a nominal knowledge of the API characteristics at such an early juncture. Preformulation studies provide attributes such as solubility, effects of pH, and predict sensitivities to environmental conditions. Solubility, along with pharmacologic characteristics, dictates the volume in the product container. Sensitivities to environmental conditions include temperature, light, oxygen, and contact surfaces. Specific to lyophilized preparations, there is also the potential for degradation via hydrolysis reactions when in the presence of water. Each of these aspects, in particular the sensitivities to environmental

Progression of product to market introduction; development efforts and extent of validation with relative batch sizes.

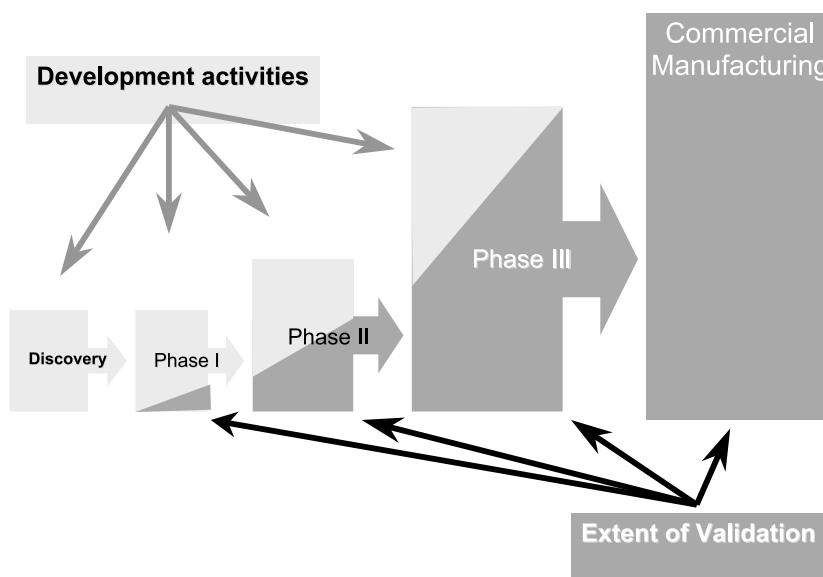


Fig. 2 Batch size, development efforts, and extent of validation. (View this art in color at www.dekker.com.)



conditions, needs to be considered during preformulation and development studies in preparation for scale-up.

Aspects of selecting a product container include both initial fill volume for dispensing in manufacturing and volume used for reconstitution, along with ease of use in a clinical setting. Later considerations include batch size capacity in a lyophilizer during commercial manufacturing. Smaller-diameter containers yield increased number of units per shelf surface area in a lyophilizer, and therefore larger batch size. This has an impact on manufacturing unit costs as well as the ability to later meet growing market demands. There may be occasion that the container size changes as the product moves from early clinical studies to product approval for commercial manufacturing. Such changes should be expected and are often unavoidable. Source of the container may also change as the product is scaled up. It is therefore prudent to evaluate different sources of containers during development, qualifying at least a primary and a secondary supplier in preparation for routine manufacturing. This would also apply to elastomeric closures used to complete the primary barrier of the container-closure system for the finished product. Product compatibility, manufacturability, and container-closure integrity all need to be evaluated for both primary and alternate components, avoiding single-source materials.

FORMULATION DESIGN AND PRODUCT CHARACTERIZATION

The nature of the product, requirements for stabilization, route of administration, and processing are all factors that influence the design of the product formulation. Specific details of these aspects are beyond the scope of this presentation and have been addressed by Pikal elsewhere in this encyclopedia. As product development and scale-up occur through the progression of preparing material for clinical studies and subsequent commercial manufacturing, it is expected that a growing body of data is generated. Such data are important in characterizing the product and understanding the critical needs, sensitivities, and behavior of the product. Analysis of a lyophilized product provides insights that are helpful in scaling up to larger batch sizes with the potential of being processed in varying environments. Along with evaluating attributes of a lyophilized preparation, finished product specifications are developed.

As upstream manufacturing progresses from small- to larger-scale processing, quality attributes of the API material may change. These changes may influence the behavior in the finished product. Analysis of the API for monitoring any changes encompasses potency and purity.

Low-temperature thermal analysis is also warranted for assessing any impact of changes in the API. Changes in quality and purity can have an influence on the phase transition of the solution, influencing the behavior during freezing and the threshold temperature during drying. Quality and purity attributes of the raw material need to be evaluated for the impact on downstream processing and finished product attributes as the upstream processing for the API is scaled-up or significant changes are made.

The formulation and product presentation may require adjustments as the product progresses through clinical studies. Concentration of the active and dose content may be adjusted for materials in later clinical studies. If the changes are nominal, then there may be little need for adjustment to the process. Significant changes warrant further development as the lyophilization processing conditions require refinement and may require subsequent scale-up studies.

PRODUCT PREPARATION

Conditions having no apparent impact during development activities can become more significant during larger-scale manufacturing. These include length of time the product exists as an aqueous solution and exposure to widely different environmental conditions. As batch sizes increase through clinical materials preparation to large-scale manufacturing, the time interval from when the API is first compounded to form a solution to when the product is frozen and lyophilized becomes extended. A small batch may be formulated on the bench requiring only a few hours to compound, filter, fill, and begin lyophilization. In routine manufacturing, the bulk solution may be formulated and held for hours prior to sterile filtration and filling. The difference in time required for preparation of a batch may warrant reducing the temperature of the solution to minimize degradation reactions. If Arrhenius behavior applies for temperature dependence of rate constants for reactions, degradation would be significantly reduced by maintaining the bulk solution at 5°C during storage rather than 25°C. Such practices are appropriate when the stability of the product as a bulk solution is limited.

The vessel used for the compounding may be different during development simply because of batch size. Compatibility with these different materials is important to consider in preparation for scale-up. Bench-scale studies and small clinical batches may be compounded in glass Erlenmeyer flasks or glass carboys. Commercial-scale manufacturing would typically use larger stainless steel tanks. Compatibility with the different materials that the product comes into contact with is important to evaluate.



Volume-to-surface-area ratios may change based on the configuration of the bulk solution container. This affects exposure to the air and absorption of oxygen and carbon dioxide. To limit the effect of such sensitivities as with oxygen, an antioxidant may be included in the formulation. Extent of exposure to light may also be different when preparing larger-quantity bulk solutions. Such sensitivities are important to explore in preformulation and early development studies and considered during scale-up.

LYOPHILIZATION PROCESS DEVELOPMENT

Part of the development studies focused on defining the critical process parameters to be controlled and that achieve reproducibility of the process and consistency of the finished product. An integral part of the development activities for a lyophilized product is establishing appropriate process parameters that encompass loading the liquid-filled containers through stoppering and unloading dried product from the lyophilizer. The lyophilization process encompasses loading, freezing, and any type of thermal treatment, primary and secondary drying, stop-

Table 1 Lyophilization process description: model presentation

Step	Processing parameters
Loading	Soak at 5°C ($\pm 5^\circ$) and 1 atm for 2 hr
Freezing	Ramp shelf to -40°C at an average controlled rate of 30°C/hr Control shelf at target set point of -40°C ($\pm 5^\circ$) for 3 hr
Primary drying	Evacuate chamber, control at a target set point of 80 μHg (± 20 μHg). Ramp shelf to -10°C at an average controlled rate of 30°C/hr Control shelf at target set point of -10°C ($\pm 5^\circ$) for 15 hr
Secondary drying	Control chamber pressure at a target set point of 80 μHg (± 20 μHg). Ramp shelf to 30°C at an average controlled rate of 30°C/hr Control shelf at target set point of 30°C ($\pm 5^\circ$) for 5 hr
Stoppering	Control chamber pressure at a target set point of 1 atm Control shelf at target set point of 30°C ($\pm 5^\circ\text{C}$).

Table 2 Lyophilization process control recipe: lysozyme model presentation

	Process conditions		Segment
	Shelf temperature	Pressure	Time
Loading	Soak: 5°C	atm	2 hr
Freezing	Ramp: 30°C/hr	atm	1.5 hr
	Soak: -40°C	atm	3 hr
Primary drying	Ramp: 30°C/hr	80 μHg	1 hr
	Soak: -10°C	80 μHg	15 hr
Secondary drying	Ramp: 30°C/hr	80 μHg	0.6 hr
	Soak: 30°C	80 μHg	5 hr
Stoppering	Soak: 30°C	atm	Total: 27.2 hr

pering, and unloading. Critical processing parameters for these steps are shelf temperature, chamber pressure, and time. These parameters are considered to be independent as they are variables under direct control and affected only by the control implemented during the process. Table 1 outlines a process description that details the independent parameters used for processing a model presentation consisting of a lysozyme formulation representing a protein preparation. A corresponding recipe used in programming an automated control system is illustrated in Table 2. Fig. 3 represents a graphical recording from lyophilizing a batch when executing processing parameters outlined in Table 2. The graph illustrates the implementation of the independent parameters of shelf temperature, chamber pressure, and time, along with the resultant product temperature profile. Note that the product is maintained below a threshold temperature to assure drying with retention is accomplished. The sudden rise in product temperature, termed a "break," reflects the sublimation front passing the temperature sensor. If the sensor is located in the bottom of the container, then this break indicates that all of the ice in the container has sublimed. The product temperature then increases, approaches the shelf temperature, at which time it is appropriate to implement the parameters suitable for desorption in secondary drying. Understanding the impact of the critical process parameters on product behavior is important in developing a sufficiently robust process and controlling the variables during scale-up.

Numerous studies have demonstrated the impact of shelf temperature on product temperature and rates of sublimation. As DeLuca demonstrated in his studies presented in 1984, the effect is principally on the rate of sublimation, with some proportionate contribution to an increase in product temperature.^[1] Summarized in Table 3, the data demonstrate the effect of increasing the shelf

Processing parameters and resulting product temperature profile for a Lysozyme preparation during loading, freezing, primary and secondary drying. Note achieving a product temperature below a threshold in freezing and due to sublimation of ice in primary drying. A sudden increase or "break" in temperature during primary drying indicates the completion of sublimation.

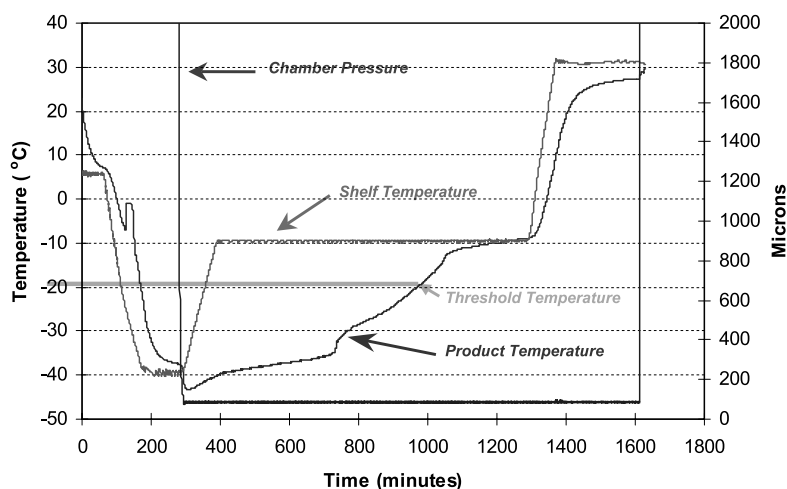


Fig. 3 Lyophilization process: lysozyme model preparation. (View this art in color at www.dekker.com.)

temperature by 10°C, yielding a substantial increase in sublimation, sometimes nearly doubling the rate. An increase in product temperature also resulted, although the product increased no more than 2 to 3°C. As sublimation consumes heat, the heat energy removed is replaced with that provided by the shelf. The higher the shelf temperature, the greater the amount of heat provided to the product, and therefore a greater potential rate of sublimation can be achieved. With heat from the shelf driving the rate of sublimation, the evolution of water vapor from the sublimation front and the product temperature resulting from attempting to reach steady-state conditions can sometimes be a fragile balance. A limit to the amount of

heat supplied for promoting rates of sublimation is imparted by the dried layer above the sublimation front. Pikal et al. studied this relationship, modeling the effect of solutes in a formulation on the mass transfer of water vapor through the dried layer above the sublimation front.^[2] In their studies Pikal et al. characterized models of crystalline and amorphous solutes predicting resistance to mass transfer of water vapor. They also noted influences of concentration to mass transfer. There has long been and continues to be interest in modeling the effects of these independent parameters of shelf temperature, chamber pressure, and time during loading and freezing through primary and secondary drying.^[3-7] This interest is in both processing rates and impact on finished product attributes such as residual moisture.

Nail investigated the effect of chamber pressure on heat transfer and the relative rate of sublimation.^[8] Chamber pressure has a prominent effect on product temperature. Shelf temperature is the principal influence on sublimation rate; chamber pressure has a strong influence on heat transfer from the shelf to the vial and the relative difference between the vapor pressure of ice and the environment. The combined effects yield a direct impact on product temperature. The driving force of sublimation is the difference between the vapor pressure of ice in the product and the pressure of the environment. An increase in the differential increases the propensity of ice to sublime. For example, if the product temperature is

Table 3 Predicting drying rate and product temperature at various shelf temperatures

Shelf temperature setpoint	Average product temperature	Average sublimation rate
50°C	-10°C	2.43 L/hr
40°C	-12°C	2.15 L/hr
30°C	-14°C	1.51 L/hr
20°C	-17°C	1.15 L/hr
10°C	-20°C	0.68 L/hr

Source: From Ref. [1].



–32°C with a corresponding vapor pressure of ice being 321 μHg when the pressure in the product chamber is 300 μHg , the small pressure differential presents a weak driving force for the ice to vaporize. The increase in heat transfer along with the nominal rate of sublimation results in increased product temperature. With a chamber pressure of 50 μHg , a greater pressure differential would result in a stronger propensity for ice to vaporize. Greater sublimation coupled with reduced heat transfer due to lower pressure would therefore reduce the product temperature. Pikal et al. correlated the rate of sublimation to chamber pressure at a constant shelf temperature, as summarized in Table 4.^[9] The magnitude of the difference between the chamber pressure and the vapor pressure of ice at the sublimation front and the pressure differential through the layer of dried solutes above the sublimation front have a significant impact on the mass flow of water vapor evolved from sublimation at the ice–vapor interface as it travels through the dried layer. As the ice–vapor interface forming the sublimation front progresses through the initial volume dispensed into the product container, the distance the water vapor travels increases. An increased pressure differential across this distance created with a lower chamber pressure improves the flow of water vapor through the dried layer. This, in turn, decreases the water vapor concentration at the ice–vapor interface, therefore increasing the propensity of ice to sublime.

All the factors that influence desorption to achieve sufficiently low residual moisture for complete drying are not yet well understood and are an area of increasing interest to investigators. Historically, conventional wisdom purports that a lower pressure readily removes any water bound to the product after the ice has sublimed. This water may be associated with solutes where the water may be part of the amorphous composition or simply adsorbed onto the surface of the solutes. Conditions of shelf temperature and chamber pressure and their effect on achieving low residual moisture are not yet well understood and of increasing interest. Studies by Pikal et al. suggest that

reduced pressure in secondary drying does not have a significant effect on rates of desorption, at least for the model preparations and processing conditions studied.^[10]

The time associated with the different conditions is also a critical independent processing parameter. Achieving complete solidification is assured with a sufficiently low temperature for a minimum time. Factors that influence sublimation of ice during primary drying and achieving appropriate residual moisture content in secondary drying for adequate stability include sufficient time to complete the respective process step. The simplicity of time as a controlled variable is not to minimize the importance of the process parameter. Each of these three parameters, defined during development, is indeed critical and needs to be sufficient for processing, irrespective of what equipment and scale the product is processed. This includes processing at a smaller scale in early clinical materials processing or at a larger scale in routine manufacturing for marketed product.

Dependent process variables are those that result from implementing the various independent parameters under direct control during processing. As independent parameters are altered there will be some effect on the process conditions as reflected by the dependent variables. Proper control of independent variables will result in achieving the acceptable dependent variables, assuring a high level of confidence that the finished product exhibits the expected attributes. Adequately defined and executed process parameters would be expected to yield a finished product of predictable quality, purity, efficacy, and stability.

Product temperature is a principal dependent variable that reflects the progression throughout the lyophilization process. During freezing, the product is chilled to a final temperature for sufficient time to achieve the necessary solidification. The appropriate temperature required is determined by the character of the formulation and measured during thermal analysis studies completed as part of development. It is also prudent to verify the physicochemical properties if the product formulation is altered when the product progresses through clinical trials.

Even subtle differences in transferring and scaling up the process can have substantial effects in achieving this critical balance. Success during scale-up results from sufficient robustness instilled in the process during development. Initial process development activities for early clinical studies are indeed minimal. As clinical studies progress, further development work is appropriate where a process more suitable for routine manufacturing may be defined. This includes process robustness studies to establish viable ranges for critical parameters. Upon identifying target processing conditions for the independent

Table 4 Predicting drying rate and product temperature at various chamber pressures

Chamber pressure (μHg)	Average sublimation rate
400	0.55–0.61 g/hr
200	0.53–0.55 g/hr
100	0.42–0.45 g/hr
68	0.37–0.39 g/hr

Ranges for 20-mL tubing vials from different manufacturers.
Source: From Ref. [9].

Table 5 Critical process parameter boundary conditions for a proven acceptable range

Process condition	Product loading	Cooling rate	Freezing	Primary ramp	Primary drying	Secondary ramp	Secondary drying
<i>Shelf temperature</i>							
High	10°C	0.5°C/hr	− 35°C	0.5°C/hr	− 5°C	0.5°C/hr	35°C
Target	5°C	0.5°C/hr	− 40°C	0.5°C/hr	− 10°C	0.5°C/hr	30°C
Low	0°C	0.5°C/hr	− 45°C	0.5°C/hr	− 15°C	0.5°C/hr	25°C
<i>Chamber pressure</i>							
High				100 μHg	100 μHg	100 μHg	100 μHg
Target	atm	atm	atm	80 μHg	80 μHg	80 μHg	80 μHg
Low				60 μHg	60 μHg	60 μHg	60 μHg
<i>Time</i>							
	2 hr		3 hr		15 hr		5, 7, and 9 hr

variables of shelf temperature, chamber pressure, and time, further studies may focus on identifying suitable boundary conditions for establishing a proven acceptable range.^[11] Variable processing conditions for such studies outlined in Table 5 and graphically represented in Fig. 4 reflect desirable ranges that yield a sufficiently robust process more easily integrated into a commercial manufacturing setting. These studies may be conducted as part of the development studies or treated separately as process validation activities. Considering these studies as either part of development or validation is not as important as the need to have completed such studies prior to scale-up to manufacturing. The greatest benefit is in

having a process that is sufficiently robust where slight variations and influences due to scale-up and transfer become inconsequential.

FINISHED PRODUCT ATTRIBUTES

The attributes of the dried product are dictated by the nature of the formulation and conditions used during processing. It is therefore desirable, if not critical, that identical processing conditions be executed on a large scale such that the quality attributes achieved during development are replicated. Any differences in the

Boundary conditions surrounding target parameters establishing a proven acceptable range listed in Table 5 for processing the model Lysozyme model preparation illustrated in Figure 3.

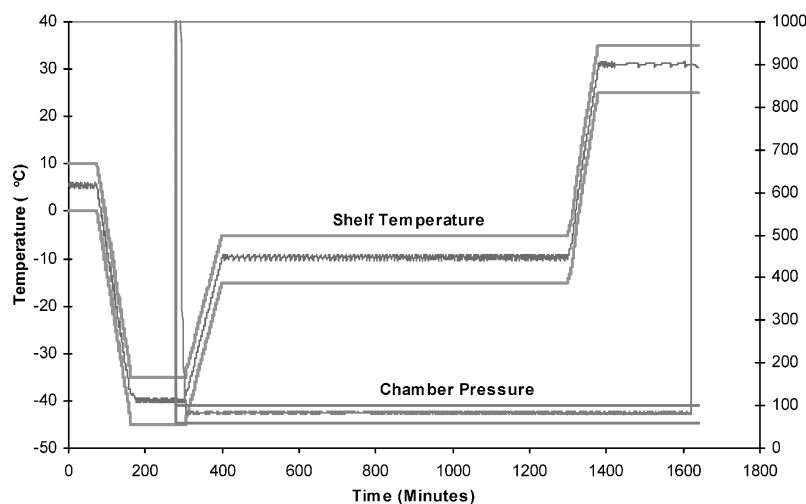


Fig. 4 Lyophilization proven acceptable range: lysozyme model preparation. (View this art in color at www.dekker.com.)



processing conditions can result in differences in subsequent processing steps as well as finished product characteristics. A comprehensive product evaluation scheme would include analytical assessment of the physico-chemical product aspects along with those unique to lyophilized product. These include physical appearance, reconstitution time, and quality of the constituted solution, along with residual moisture.

Lyophilized preparations have a unique set of dried finished product attributes in addition to those of liquid, ready-to-use products.^[12] Physical appearance of the dried cake, residual moisture, reconstitution time, and clarity of the constituted solution are product attributes important to quantify. The end result of lyophilization is to preserve attributes the product exhibited when first prepared after reconstitution of the dried product. Validation studies demonstrate and document the capability when starting materials are of a known acceptable quality and an appropriate and well-controlled process is reproduced, the finished product will be of predictable quality, purity, efficacy, and stability. As a minimum, chemical analysis may be used to show that the desired product attributes were preserved during processing and therefore the pro-

cessing conditions were adequate. Measuring the potency of the product is important in verifying that both the filling and freeze drying operations were appropriate and controlled, producing a product with the required efficacy. It is also important to show that any degradation during processing has been avoided. Specific analytical techniques are used to assure this objective has been achieved and the product meets the predefined quality attributes.

The appearance of the dried product, in itself, is not considered a critical product attribute. It does, however, reflect success in processing and finished quality attributes. An objective and a measure of success of appropriate processing conditions is establishing and retaining the quality attributes imparted during freezing. Freeze drying with retention and avoidance of collapse through primary and secondary drying, as described by MacKenzie, reflects the level of success in proper processing.^[13] If appropriate processing conditions are implemented that assure adequate solidification during freezing and the product is maintained below a threshold temperature during primary and secondary drying, the structure established during freezing will be preserved. Product warming above a threshold temperature where the

Dried cake appearance reflects the size and structure of the ice formed during initial freezing. Variation in dried cake physical appearance for a polysaccharide model preparation reflecting differences in ice crystal size. Note slight shrinkage of the dried cake from the original liquid fill volume characteristic of the model composition.

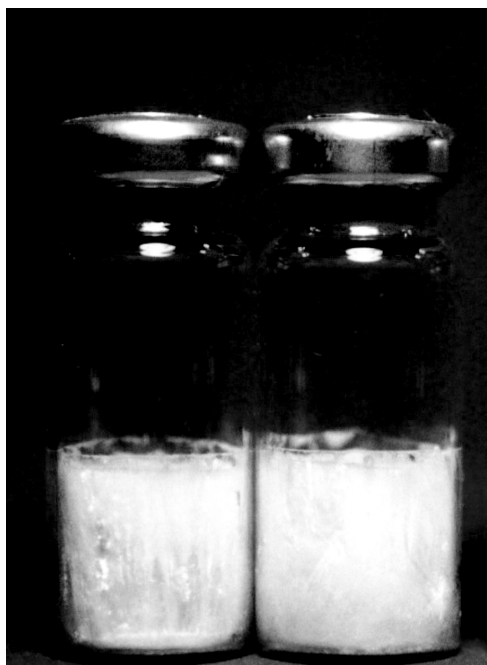


Fig. 5 Dried cake appearance. (View this art in color at www.dekker.com.)

material reverts to a liquid by either exceeding a glass transition causes a collapse of the structure or a eutectic temperature resulting in a melt. Either event alters the structure of the material and affects product behavior during further processing and finished product attributes. A change in the structure may impede sublimation and desorption, resulting in higher residual moisture, as well as influence reconstitution. Any change in the structure of the dried product is detected with physical inspection, assessing the dried cake appearance.

Physical appearance of the dried material can sometimes vary substantially. This is in part dependent upon the nature of the formulation along with techniques used in processing. The appearance of the dried cake is a mirror image of the ice crystal formed during the initial freezing. Fig. 5 illustrates the variation in dried product appearance due to differences in structure resulting from varying ice crystal size. If the ice crystals are large the cake may have a more coarse structure. Conversely, if the ice crystals are small, then the cake structure may be very fine. The structure of the ice is influenced by the shelf cooling rate during freezing. This, in turn, is dependent on the temperature at which the nucleation of ice occurs and on the

rate of ice crystal growth. Searles et al. investigated such influences of cooling rate and effect of thermal treatment on expected differences in ice crystal size.^[14] If the cooling rate is different for the product processed in the manufacturing unit, then the physical appearance may be different.

The physical appearance is not considered a critical product attribute in itself. Rather, the physical appearance, reflecting attributes such as either a coarse or fine structure of the dried cake, may also be associated with other attributes such as reconstitution time. For example, if large ice crystals form during the solidification as the product is chilled during freezing, large voids will remain in the dried product when the ice crystals have sublimed. These large voids then provide an avenue through which the diluent permeates the dried cake upon reconstitution. Changes in cooling rates with a potential effect on ice crystal size and the resultant change in dried product structure may therefore impart differences in ease of reconstitution. A time specification for the dried product to completely go into solution needs to be identified during development.

Ease of reconstitution is a desirable product attribute, although it is secondary to the quality, purity, efficacy,

Visible insoluble material of the constituted solution implicating a physico-chemical change during processing or upon reconstitution.

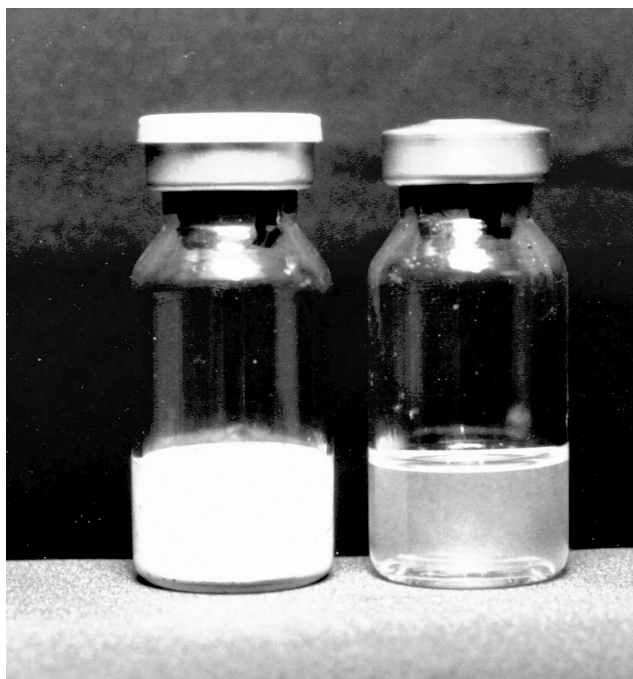


Fig. 6 Reconstituted solution. (View this art in color at www.dekker.com.)



and stability. Reconstitution will sometimes indicate the presence of material that has collapsed and failed to dry with retention of the structure established during freezing. Collapse and melt back is associated with causing portions of the cake to dissolve more slowly and extend the time required to go into solution completely. The material polymerizing, peptides or proteins forming aggregates, or liposomes coupled or clustered may also be causes of a haze, cloudy solution or insoluble materials during reconstitution. Poorly soluble material or incomplete dissolution similar to that observed in Fig. 6 often indicates that there has been a significant physicochemical change in the product. The end point of the reconstitution is the formation of a clear solution, no less clear than the diluent used for reconstitution.

A target range of allowable residual moisture needs to be correlated to long-term storage stability and is necessary to define during development. Prevention of degradation by hydrolysis is circumvented by the removal of any appreciable water that may become involved in chemical reactions during storage. Stability in the dried state is assessed during development. The residual moisture of the dried product is dictated by the extent of desorption during secondary drying. The rate of desorption is strongly influenced by the product temperature and time in secondary drying. Achieving a residual moisture within the range established in development and correlated to processing conditions verifies that the time necessary for sufficient desorption was adequate when processing the larger batch size.

PROCESS DEVELOPMENT CONSIDERATIONS

There may often be many unknowns about suitable manufacturing operation while in the midst of product and process development activities. For preparation of clinical materials, the product may be produced in-house or at a contract facility. There are multiple possibilities for manufacturing product on a commercial scale: The product may be integrated into an existing in-house operation, manufactured at a yet unidentified contract manufacturing site, or the new product may warrant expansion of an existing operation or construction of a new facility. A direction may well be yet unknown at the time product and process development is ongoing. Such a circumstance requires some assumptions and foresight in selecting parameters when designing a process that would be best suited for routine large-scale manufacturing. Whether a manufacturing site is known or has yet to be identified, and unless there are extraordinary needs dictated by the

product, designing a robust process increases the potential that technology transfer will be successful.

Intending that the process will be suitable for routine manufacturing, an understanding of an existing manufacturing operations capability provides valuable insight, direction, and focus to the development studies. With such understanding, the boundaries and parameters that envelope processing procedures suitable for the current manufacturing operation can be studied in development. If not existing, then an approximation of “typical” capabilities for a manufacturing operation is helpful. With planned expansion of operations required for commercial manufacturing, questions of “What do you need?” and “What can you provide?” for processing capabilities echo between development, engineering, and operations staff. To be complete and comprehensive, the perspective of this presentation on the topic will be development and scale-up into an existing operation. This will at least allow a comprehensive treatment for subjects of interest in scale-up. It is also important to realize that current operations should not necessarily be limited to critical product or processing needs. An example is with the use of organic solvents in a formulation to improve product structure or processing rates.^[15,16] In such circumstances the need of stabilizing the product and impact on the process needs to be weighed in deciding on the direction during development. It is important to also appreciate that the current capabilities of any operation should not stifle creative and beneficial advancements in manufacturing technology.

For the first round of development it is prudent to attempt to fit a product and process within an existing operational capability. The first source of information to seek is the operational qualification (OQ) data quantifying the functional capabilities and performance of the lyophilizer. Table 6 outlines a list of performance capabilities of

Table 6 Performance capabilities from operational qualification studies

<i>Shelf temperature:</i>				
Range	Low		High	
Rates	Cooling		Heating	
Control	Low	Intermediate		High
Uniformity	Low	Intermediate		High
<i>Chamber pressure:</i>				
Range	Low		High	
Control	Low	Intermediate		High
<i>Sublimation/condensation rates:</i>				
	Shelf temperature setpoint			
	Chamber pressure setpoint			
	Rate achieved			
Total ice sublimed/condensed				

Copyright © Marcel Dekker, Inc. All rights reserved.

interest that allows a sufficient assessment of equipment functions and performance.

Shelf temperature has the greatest influence on the processing rates and ultimate product temperatures at the completion of each process step. The achievable shelf temperature capability ranges from the coldest temperature for freezing to the warmest temperature for primary and secondary drying. High rates of cooling and warming reflect the most aggressive changes in shelf temperature throughout the process. Control capabilities include the variation in controlling at a specific temperature and any relative deviation from the desired set point. This capability can also be influenced by the type and location of the temperature sensor. Shelf-temperature variation influences the batch uniformity and is important at the low temperatures used for freezing, intermediate temperatures often used for primary drying and relatively high temperatures for secondary drying.

Left uncontrolled, chamber pressure reflects the water vapor in transit from the product from which ice is subliming to the condenser where the water vapor is converted back to ice. Chamber pressure has the greatest influence on product temperature. A pressure suitable for a formulation containing sucrose and having a phase transition or collapse temperature of -32°C may be as low as 60 μHg . Preparations with high phase transition temperatures such as those that contain mannitol that tends to crystallize and having a eutectic melt as warm as -2.6°C may be processed using chamber pressures such as 400 μHg . Controlling the chamber pressure during primary drying has been widely used throughout the industry for more than 20 years, as noted previously. The effect and benefit of chamber pressure control during secondary drying is still a controversy.

There are three general methods used to control the chamber pressure. They include injecting nitrogen into the product chamber, injecting air or nitrogen into the inlet of the vacuum pump, and closing the valve between the chamber and condenser or condenser and vacuum pump to decrease the flow of water vapor. It is a common and preferred practice to control the chamber pressure by introducing filtered nitrogen into the chamber using a proportional valve. The behavior of the nitrogen when injected into the chamber, relative effect on rates of sublimation during primary drying as compared to the other methods, influence on desorption during secondary drying, and effect on conversion of water vapor to ice at the condenser surface all need to be considered when utilizing pressure control by the various methods.

Sublimation and condensation rate studies are part of a comprehensive OQ and reflect the capability of the equipment to complete the primary drying. The amount of

ice sublimed from the shelf and collected on the condenser with a constant shelf temperature and chamber pressure indicates the level of performance that can be achieved under load conditions. The rate reflects the capability of the shelves along with the heat transfer system to supply sufficient energy for sublimation. Evolution of corresponding quantities of water vapor challenges the effectiveness and capacity of the condensing system. The quantity of ice sublimed and condensed reflects both the obtainable rates and capacities. Sublimation of ice from the shelves reflects the capability to supply heat to the product. The amount of ice sublimed can also indicate the uniformity when rates from each shelf are compared. For an external condenser the pressure differential between the chamber and condenser reflects any inhibition to water vapor flux because of the equipment configuration when the vapor is in transit from the product to the condenser. The quantity of ice collected on the condenser is a measure of the effectiveness in trapping and removing water vapor and the ability to hold the expected quantity of ice. The sublimation and condensation rate studies are therefore useful in evaluating the equipment capabilities and gaining confidence that the lyophilizer can achieve the required process parameters.

Capabilities of the lyophilizer to be used for larger-scale manufacturing become the extent of processing parameters and therefore provide guidelines for process development studies. For transparent scale-up the parameters investigated during laboratory studies need to be within the known operating capabilities of the manufacturing equipment. Controlling the parameters investigated during process development to within the capabilities of the larger production unit eliminates the need for adjustments to the process when scaling up to larger batches in a manufacturing environment.

Validation data for large-scale processing in manufacturing is also a wealth of information on the processing capability and performance. Historical manufacturing experience with other products provides an indication of the extent of batch-to-batch variation during routine manufacturing. This would include an assessment of both processing capabilities as well as finished product characteristics. The level of control and reproducibility, along with consistency of finished product, can be evaluated to assess capabilities and predict results during scale-up. Batch uniformity with respect to variation that may be attributed to different positions in the lyophilizer included in validation studies can also be useful. Environmental influences and their effect on rates of sublimation relative to position on the shelf have been illustrated by Greiff.^[17] In order to encompass the variation, a sufficiently large batch is necessary to evaluate the impact. As API is often



in limited supply for development studies, a sufficiently large batch may consist of vials filled with only the excipients or use of a surrogate formulation. Realizing that the number of vials in the development study may be small, evaluating the uniformity in behavior during processing and finished product attributes provides insight on the expected variation throughout a larger-scale batch.

Uniformity in the product behavior during processing and consistency in finished product is strongly influenced by nucleation and ice crystal growth during freezing. The randomness of nucleation, an event based on probability, and subsequent ice crystal growth can be a significant influence when solidification is complete during freezing and product temperature during drying. Fig. 7 reflects the distribution of times and temperatures at which nucleation occurs for a limited number of vials within a small-scale batch being monitored using fine-gauge thermocouples carefully positioned in the center at the bottom. Note also the increase in the temperature of the monitored vials reflecting in a second and later increase in temperature due to nucleation and ice crystal growth in neighboring

vials. This delay in the solidification of water in vials without a thermocouple suggests a greater extent in supercooling, imparting the associated attributes expected with smaller ice crystal size. This event occurs with a variation within any group of vials within any single batch. Such variation in nucleation within a group of containers is evidenced with a single batch and therefore within any group of batches. The range is apparent when comparing such events that occur among multiple batches using the same processing parameters. This variation is illustrated in Fig. 8 for four batches of a model formulation chilled using the same parameters.

This variation in product temperatures also occurs during primary drying as drying progresses and reflects what may be considered typical vial-to-vial variation during lyophilization. As ice crystal size dictates the surface area of the remaining solutes there may also be an effect on ease of removing residual moisture in secondary drying. For the same batch illustrated in Fig. 7 above and exhibiting a variation in nucleation, product temperatures will also vary during drying, as apparent in the temperature

Temperature profile during freezing of multiple vials within a batch. Variation in time and range in temperature for nucleation of ice and ice crystal growth for vials monitored using a thermocouple. Increases in temperature following ice crystal growth reflects nucleation and ice crystal growth in vials that do not contain a thermocouple surrounding those being monitored.

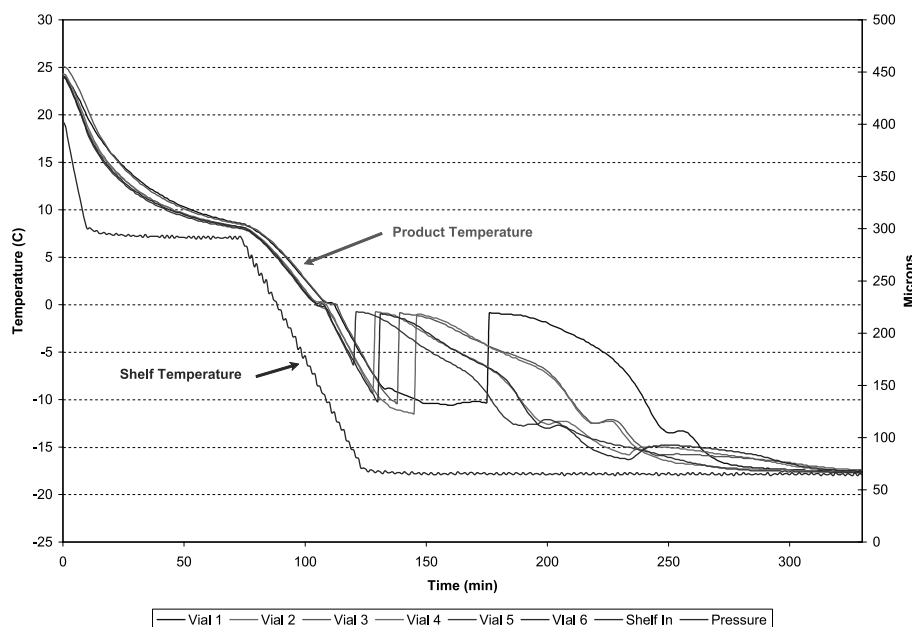


Fig. 7 Variation in time and temperature for nucleation and ice crystal growth within multiple vials of a single batch. (View this art in color at www.dekker.com.)

Temperature profile during freezing of multiple vials within multiple batches. Variation in time and range in temperature for nucleation of ice and ice crystal growth for vials containing a thermocouple and those not containing a thermocouple surrounding those monitored.

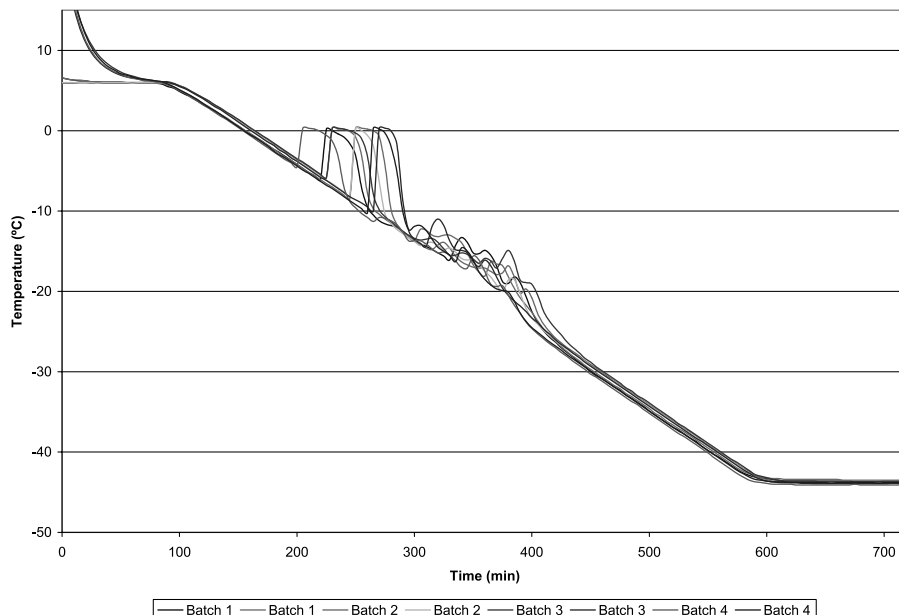


Fig. 8 Variation in time and temperature for nucleation and ice crystal growth within multiple vials of multiple batches. (*View this art in color at www.dekker.com.*)

profile in Fig. 9. This vial-to-vial variation occurs, even in the most carefully controlled laboratory conditions. Such behavior will also occur in routine manufacturing and is important to consider while establishing parameters when developing a process. It is therefore imperative that the process be sufficiently robust to accommodate the inherent variation within a single batch and that occurs from batch to batch.

Compiling the data on equipment capabilities and experience of validation along with routine manufacturing provides a framework in which to select parameters for study during process development. Parameters controlled to within the operational capabilities of the manufacturing equipment minimized any adjustments to the process and continued development efforts during scale-up. Suitable parameters to accommodate variation that may occur within a larger lyophilizer and in routine manufacturing can be considered in the process design during development. Realizing that the product and process will exist in a manufacturing environment for the majority of the product life, the objective of the development studies is to establish a process that will easily integrate into a manufacturing operation during scale-up.

EQUIPMENT INFLUENCES

In circumstances where a new product is being integrated into existing manufacturing operations, the parameters of the process need to be within the manufacturing equipment performance capabilities. Therefore the capability of the equipment used during development and that to be used in manufacturing should be evaluated. This includes equipment design and configuration, control of critical processing parameters, and operating capacities. Understanding how the equipment for development and manufacturing functions and capabilities in implementing the required processing conditions is the first step in transferring a process from one operation to another.

It is not unusual that process development is complete before a manufacturing site is identified. Therefore the guidance for identifying appropriate process parameters for use in manufacturing may not be available. From a manufacturing perspective it is prudent to understand the processing capabilities for the research lyophilizer along with the actual parameters used, as well as any difference on the processing conditions the product experiences. For example, if left uncontrolled during development studies,



Temperature profile and behavior during freeze drying of multiple vials within a batch corresponding to the variation in nucleation of ice and ice crystal growth shown in Figure 7.

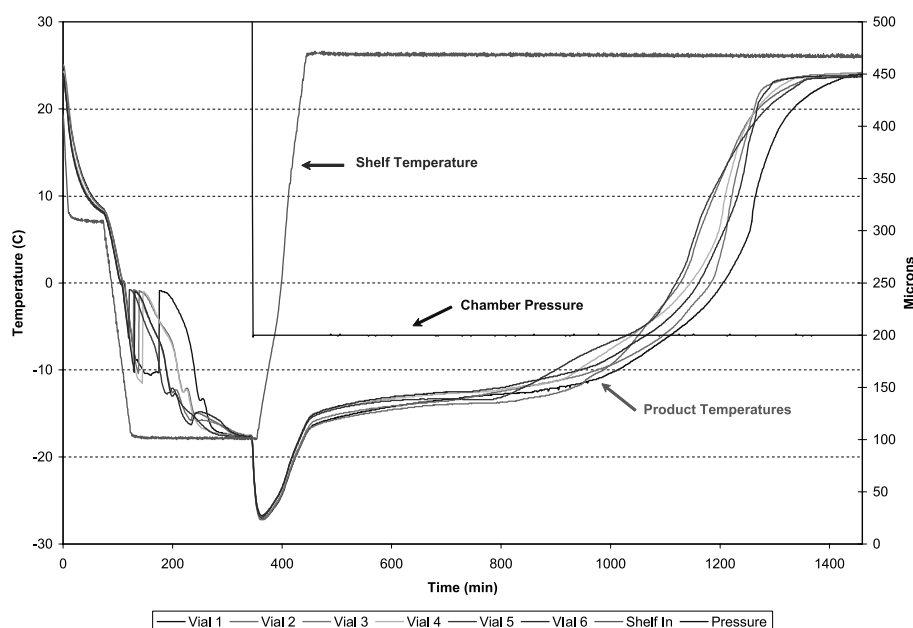


Fig. 9 Variation in behavior during processing multiple vials of a single batch. (*View this art in color at www.dekker.com.*)

cooling and heating rates can vary, dependent upon unit manufacturer, equipment component configuration, and lyophilizer capacity. When transferring a process, the processing conditions specified during development need to be replicated in routine manufacturing. Typical capabilities of the development unit that are important to assess are the same as those when evaluating large-scale manufacturing equipment, as presented in Table 5. Other considerations are instruments and methods used for process control and monitoring. A suitable production lyophilizer utilizes a resistance temperature device (RTD) located in a thermowell positioned in the heat transfer fluid just prior to where the fluid enters into the chamber and is the control point for the shelf temperature. This location is the most effective to use for control, as it provides the most consistent measurement, accommodating dynamics of the equipment, variations in processing, and differences in thermal load. Research units may have the sensor in a different location or a different method for monitoring and controlling the shelf and implementing temperature control. The impact of such a difference is the level of control as well as an offset to the actual shelf temperature. For example, an RTD located in a heat transfer fluid storage tank that is located upstream of the refrigeration heat exchanger and heater

would reflect different dynamics and therefore a different “shelf” temperature.

A second major variable is method of pressure measurement. Research units often use thermoconductivity type pressure instruments to monitor the chamber pressure. These instruments were designed for measuring processes conducted at reduced pressures, such as those in the electronics industry. These instruments measure the differences in thermoconductivity of an atmosphere within the lyophilizer. Given the ideal gas law where $PV = nRT$, the pressure decreases with a decrease in n , the number of molecules, where the molecular density, the number of molecules per unit volume is reduced, with all other variables constant. The thermoconductivity of an atmosphere decreases with a decrease in the molecular density and can be correlated to pressure. As these instruments are calibrated against nitrogen, the main constituent of a normal atmosphere, when sensing the pressure of a different atmosphere, such as that composed of water vapor, these types of instruments will indicate a different pressure. These instruments are prone to errors when the composition of the atmosphere varies. Sublimation during primary drying yields a composition of the atmosphere in the chamber high in water vapor, resulting in errors in measure pressure as high as 60%.

Position of the sensor can also have an influence on the indicated pressure. Monitoring and control of the pressure that comprises the environment to which the product is exposed is necessary. Therefore the sensor needs to be positioned on the product chamber. If the sensor is positioned on the condenser vessel or in the piping going to the vacuum pump, then the measured pressure will not reflect the amount contributed by the water vapor present during the sublimation of ice. Research lyophilizers sometimes have the pressure sensors located on the vacuum line and actually read the pressure of the pump rather than the chamber.

EXECUTING SCALE-UP ACTIVITIES

Knowledge of manufacturing capacities and capabilities, appropriate and adequate processing conditions for a robust process, coupled with careful control of starting material and processing conditions are necessary for successful technology transfer in achieving the predefined processing conditions and predicted finished product qualities. Sufficient knowledge of the product and process gained in development prevents scale-up from becoming a period of discovery. Processing at larger batch sizes provides an opportunity to verify the suitability of the processing parameters identified during development and focus on assessing the distribution of behavior and uniformity of finished product qualities at a larger scale.

A complete and comprehensive development report is an invaluable resource for technology transfer. This document preserves the body of knowledge and experience gained during product design and process development activities. Information including characteristics of the API, formulation, and finished product outlined in the preceding sections are part of a comprehensive development report. As well, the knowledge and experience in initial scale-up for processing clinical material is also valuable. The expected behavior, finished product attributes, and any difficulties that arise in scaling-up provide useful insight as larger batch sizes are integrated into routine manufacturing.

Realistically, subtle differences between conditions and outcomes during development and those in large-scale manufacturing will exist. Influences of preparing a product and processing in a manufacturing environment, in conjunction with the expanded quantity of finished product, combine to become significant factors. Batch sizes and quantities of finished units are often nominal as compared to that of large-scale manufacturing. Development studies and early clinical manufacturing requirements may only require relatively small batch sizes. As such, if the incidence of an event or result is one in a

thousand and the batch size is a few hundred vials, it may only be an occasional observation or result. If the frequency is one in a few thousand, it may not become evident until large-scale batches are prepared in manufacturing. With manufacturing batch sizes approaching as large as tens of thousands of units being common, a statistically small incidence becomes a meaningful quantity of finished product.

The statistical distribution of variables simply due to larger quantities of finished product units will all be factors. These factors can be compounded such that any one alone has little or no impact. In combination and with potential of synergistic effects, the outcome of scale-up may yield results that are beyond any previous experiences. The level of success, even with factors or influences that may not be foreseen, is strongly dependent on the extent of knowledge and preparation during development as well as refinements implemented during the steps in preparation of clinical materials, leading up to integrating a new product into routine manufacturing for supplying the marketplace.

For scale-up of the lyophilization process, critical process parameters of shelf temperature, chamber pressure, and time are compared to those established during development and intended to be implemented using a larger lyophilizer. Some reasonable variation due to the dynamics of larger size and capacity equipment would be expected. For example, there may be an initial overshoot for a short duration when changing processing parameters. This is typical of proportional type control, where the amplitude and frequency of oscillations around the set point decreases until constant control at the set point is achieved. Proportional control is used for both shelf temperature and chamber pressure. These oscillations may be observed at the completion of a ramp leading to a controlled shelf temperature at a specific set point or when changing chamber pressure control, as in the transition from primary to secondary drying. These oscillations, if they are of nominal extent and duration, have little, if any, effect on product temperature during the process. Product during freezing and drying may also exhibit different behavior. Placement of product temperature sensors is a major influence. Precise placement is relatively easily accomplished in a development laboratory setting. Using good aseptic technique in processing sterile product limits the manipulation necessary for specific placement of a sensor in a vial. Even with such difference in placement, accurate temperature values are assured at the most critical times in the process; the end of loading, freezing, primary, and secondary drying. It is important to verify the product has reached critical temperatures at the end of each of these steps to assure the conditions are adequate before progressing to the next step. Verifying that the



product is chilled below the phase transition temperature and maintained below the threshold temperature identified during development are important assessments in evaluating the product temperature profile during scale-up batches.

Product behavior for a sterile product in manufacturing is also influenced by the conditions imparted due to requirements for aseptic processing. Materials prepared under laboratory conditions in a development environment are typically processed differently than in a manufacturing environment. For example, the level of cleanliness and bioburden are different. Product attributes, such as the nucleation and crystal growth of ice, would occur at different temperatures and rates. This has an influence on product behavior during processing and finished product attributes, as discussed earlier.

Verifying processing conditions are suitable and within reasonable range, finished product attributes are assessed and compared with expected results. Attributes associated with lyophilized preparations such as physical appearance, reconstitution, and residual moisture are important to evaluate and compare differences between results from development studies and those in a manufacturing environment. In addition, potency and purity, along with those attributes specific to a lyophilized dosage form, are compared to the specifications identified during development and refined as clinical studies progress.

Processing a larger number of units as batch sizes increase provides an opportunity to evaluate process capability. Beyond evaluating processing conditions and finished product attributes processed in a manufacturing environment, larger-scale batches provide an opportunity to evaluate the process dynamics and resulting product on a larger scale. It is prudent to evaluate a larger number of product samples during the initial large-scale manufacturing as compared to the relative limited number samples during routine manufacturing. With a larger number of samples, statistical methods provide a greater capability of quantifying the variation in product characteristics and quality attributes. Kieffer and Torbeck present methods and illustrations of establishing sample size and statistical approaches in conducting process capability studies using a step-by-step procedure useful for statistical evaluation when scaling up to larger batch sizes.^[18] When processing larger batch sizes, reproducibility of processing conditions can be coupled with evaluating batch uniformity. This may include randomized sampling of a larger number of finished product units and using statistical methods to compare results within some desired distribution. Use of statistical methods provide a more quantitative measure at a larger scale for each scale-up batch and evaluate the results achieved between batches. This leads to accomplishing a higher degree of assurance as the product moves

toward integrating the production into commercial manufacturing.

CONCLUSION

Ease and success of scale-up depends on the knowledge and experience gained beginning with development. Scaling up as the product progresses from development, through clinical studies, ultimately to large-scale commercial manufacturing occurs in multiple steps. During this progression there are often refinements to the product and process along the way. The starting materials should be well characterized. This includes the quality and purity of the API, adjustments in API, and formulation excipient concentrations, as well as the container-closure components. Requirements for preparing a batch and accommodating any product sensitivity need to be well defined for assuring the quality of the starting solution. A sufficiently robust process, with target parameters identifying "ideal" conditions, along with definition of a proven acceptable range within boundary conditions needs to be readily executed within a manufacturing environment.

The statistical distribution of variables due to larger quantities of finished product units can become a significant influence factor. This factor may be compounded such that any one alone has little or no impact. In combination and with potential of synergistic effects, the outcome of scale-up may yield results that are beyond any previous experiences. The level of success, even with factors or influences that may not be foreseen when preparing for scale-up in manufacturing as batch sizes increase, is strongly dependent on the extent of knowledge and preparation during development. This is also influenced by the extent of refinements implemented during the steps in preparation of clinical materials, leading up to integrating a new product into routine large-scale manufacturing.

REFERENCES

1. DeLuca, P.P. Freeze-Drying Pharmaceuticals and Biologicals. In *Pharm Tech Conference Proceedings*; Advanstar Communications: Cleveland, 1984; 184–194.
2. Pikal, M.J.; Shah, S.; Senior, D.; Lang, J.E. Physical chemistry of freeze-drying: Measurement of sublimation rates for frozen aqueous solutions by a microbalance technique. *J. Pharm. Sci.* **1983**, 72 (6), 635–650.
3. Rowe, T.W. Freeze Drying of Biological Materials: Some Physical and Engineering Aspects. In *Current Trends in Cryobiology*; Smith, A.U., Ed.; Plenum Press: New York, 1970; 60–138.



4. MacKenzie, A.P. The physico-chemical basis of the freeze drying process. *Dev. Biol.* **1976**, *36*, 51–67.
5. Pikal, M.J.; Roy, M.L.; Shah, S. Heat and mass transfer in vial freeze drying of pharmaceuticals: Role of the vial. *J. Pharm. Sci.* **1984**, *73*, 1224–1237.
6. Nail, S.L.; Gatlin, L.A. Freeze Drying: Principles and Practice. In *Pharmaceutical Dosage Forms: Parenteral Medications*, 2nd Ed.; Avis, K.E., Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1993; Vol. 2, 163–234.
7. Pikal, M.J. Freeze drying of proteins: Part I. Process design. *Biopharm* **1990**, *3* (9), 18–27.
8. Nail, S.L. The effect of chamber pressure on heat transfer in the freeze drying of parenteral solutions. *J. Parenter. Drug Assoc.* **1980**, *34* (5), 358–368.
9. Pikal, M.J.; Roy, M.L.; Shah, S. Mass and heat transfer in vial freeze-drying of pharmaceuticals: Role of the vial. *J. Pharm. Sci.* **1984**, *73* (9), 1224–1237.
10. Pikal, M.J.; Shah, S.; Roy, M.L.; Putman, R. The secondary drying stage of freeze drying: Drying kinetics as a function of temperature and chamber pressure. *Int. J. Pharm.* **1990**, *60*, 203–217.
11. Trappler, E.H. Validation of Lyophilization. In *Validation of Pharmaceutical Processes: Sterile Products*, 2nd Ed.; Carleton, F.J., Agalloco, J.P., Eds.; Marcel Dekker, Inc.: New York, 1999; 721–750.
12. Daukas, L.A.; Trappler, E.H. Assessing the quality of lyophilized parenterals. *Pharm. Cosmet. Qual.* **1998**, *2* (5), 21–25.
13. MacKenzie, A.P. Collapse During Freeze Drying-Qualitative and Quantitative Aspects. In *Freeze Drying and Advanced Food Technology*; Goldblith, W.A., Rey, L., Rothmayer, W.W., Eds.; Academic Press: New York, 1975; 278–307.
14. Searles, J.A.; Carpenter, J.F.; Randolph, T.W. Annealing to optimize the primary drying rate, reduce freezing-induced drying rate heterogeneity, and determine T_g' in pharmaceutical lyophilization. *J. Pharm. Sci.* **2001**, *90* (7), 872–887.
15. Seager, H.; Taskis, C.B.; Syrop, M.; Lee, T.J. Structure of products prepared by freeze-drying solutions containing organic solvents. *J. Parenter. Sci. Technol.* **1985**, *39* (4), 161–179.
16. Kasraian, K.; DeLuca, P. Effect of *tert*-butyl alcohol on the resistance of the dry product layer during primary drying. *Pharm. Res.* **1995**, *12* (4), 491–495.
17. Greiff, D. Factors affecting the statistical parameters and patterns of distribution of residual moistures in arrays of samples following lyophilization. *J. Parenter. Sci. Technol.* **1990**, *44* (3), 118–129.
18. Kieffer, R.; Torbeck, L. Validation and process capability. *Pharm. Technol.* **1996**, *22* (6), 66–76.



Request Permission or Order Reprints Instantly!

Interested in copying and sharing this article? In most cases, U.S. Copyright Law requires that you get permission from the article's rightsholder before using copyrighted content.

All information and materials found in this article, including but not limited to text, trademarks, patents, logos, graphics and images (the "Materials"), are the copyrighted works and other forms of intellectual property of Marcel Dekker, Inc., or its licensors. All rights not expressly granted are reserved.

Get permission to lawfully reproduce and distribute the Materials or order reprints quickly and painlessly. Simply click on the "Request Permission/Order Reprints" link below and follow the instructions. Visit the [U.S. Copyright Office](#) for information on Fair Use limitations of U.S. copyright law. Please refer to The Association of American Publishers' (AAP) website for guidelines on [Fair Use in the Classroom](#).

The Materials are for your personal use only and cannot be reformatted, reposted, resold or distributed by electronic means or otherwise without permission from Marcel Dekker, Inc. Marcel Dekker, Inc. grants you the limited right to display the Materials only on your personal computer or personal wireless device, and to copy and download single copies of such Materials provided that any copyright, trademark or other notice appearing on such Materials is also retained by, displayed, copied or downloaded as part of the Materials and is not removed or obscured, and provided you do not edit, modify, alter or enhance the Materials. Please refer to our [Website User Agreement](#) for more details.

Request Permission/Order Reprints

Reprints of this article can also be ordered at

<http://www.dekker.com/servlet/product/DOI/101081EEPT120014502>



Gelatin-Containing Formulations: Changes in Dissolution Characteristics

Saranjit Singh
Sariputta P. Pakhale

National Institute of Pharmaceutical Education and Research, Nagar, India

INTRODUCTION

Dissolution parameter or the dissolution profile is one of the important specifications for oral solid dosage forms. Even during aging, dissolution characteristics are required to remain either unchanged or within specifications laid down with reference to the results observed during bioequivalence, comparative bioavailability, or clinical studies. The product can be rejected and recalled if dissolution instability is shown during storage under conditions defined on the label. In that respect, the manufacturer of a pharmaceutical product has a responsibility not only from ethical, moral, and legal standpoints, but also for economical reasons to ensure that the drug is released reproducibly from the dosage form during the shelf life.

The problem of alteration in dissolution characteristics on aging is typical to gelatin-based dosage forms. It is routinely faced by the manufacturers and, therefore, is a matter of concern. Fortunately, a lot of understanding has been developed over the period, since the time the problem was first recognized. The purpose of this article is to put the problem in the right perspective based on recent developments and the review of the reports in literature.

THE PROBLEM OF PELLICULIZATION

Gelatin is hydrolyzed by most of the proteolytic systems to yield amino components. Further, it reacts with acids and bases, aldehydes and aldehydic sugars, anionic and cationic polymers, electrolytes, metal ions, plasticizers, preservatives, and surfactants. Even, exposure to stress conditions of humidity, temperature, and/or light leads to perceptible changes.

The exposure of gelatin formulations to environmental factors and/or chemical catalysts results in formation of a swollen, very thin, tough, rubbery, water-insoluble membrane, also known as "pellicle." This membrane acts as a barrier and restricts the release of the drug. It is

not disrupted easily by gentle agitation and as a result dissolution characteristics of formulations containing gelatin in the outer layer change and Q -values often drop to the point of rejection.^[1,2] An example is shown in Fig. 1.^[3] The plot depicts a comparison of dissolution behavior of fresh gelatin capsules against 1-yr old capsules and tablets devoid of gelatin. Evidently, 1-yr old capsules show reduced dissolution rate, in particular.

Fortunately, several studies have shown that the problem of fall in dissolution rate of gelatin formulations has a little consequence on in vivo drug bioavailability.^[4–6] Fig. 2 depicts both dissolution and bioavailability profiles for fresh drug capsules and those stored at room temperature for 11 mo.^[7] It shows a drop in in vitro dissolution by ~50% (Fig. 2a), but no significant change in C_{\max} or t_{\max} values (Fig. 2b).

The absence of in vivo effect is attributed to digestion of the denatured gelatin by enzymes present in GIT. This is the reason that the problem was not even exposed before 1960s when simulated gastric and intestinal fluids were used as dissolution media. There are large numbers of studies in literature, which have shown that adverse effects on dissolution are virtually eliminated when the products are tested in the dissolution media containing enzymes.^[8–12] An example highlighting the corrective influence of enzymes on reduced dissolution of stressed ibuprofen hard gelatin capsules is shown in Fig. 3.

The studies of the type shown in the figure indicated that if enzymes were able to alleviate the impeding barrier exerted upon the drug molecule by highly cross-linked gelatin capsule wall, so a dissolution test in the presence of enzyme could avoid the time and cost of bioequivalence studies. Accordingly, attention was called on having a two-tier dissolution test for specific evaluation of gelatin products.^[12] The test was eventually developed by a FDA's Industry Gelatin Capsule Working Group in which USP was also a participant. It was included in the 25th edition of USP.^[13] The test encompasses initial dissolution study in the plain medium as specified in the individual monograph, followed by a second dissolution in the medium containing enzymes. Two types of enzymes,

2

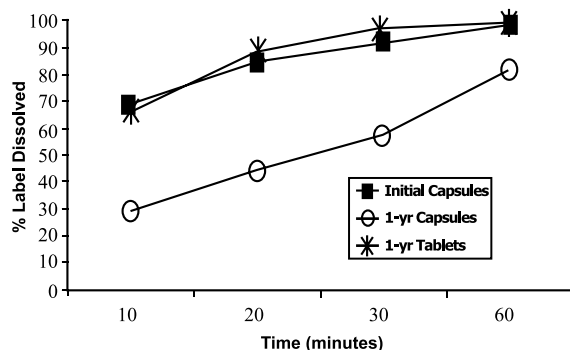


Fig. 1 Dissolution profiles for fresh and 1-yr old capsules and tablets. The 1-yr old capsules are evidently associated with lower dissolution rate. (From Ref. [3], by courtesy of Russell Publishing.)

Gelatin-Containing Formulations: Changes in Dissolution Characteristics

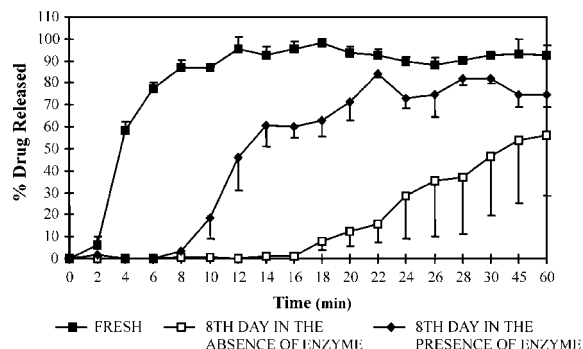


Fig. 3 Release profiles of fresh hard gelatin capsules (■) and those stored for 8 day at accelerated conditions of 40°C/75% RH in the presence of light (□, ◆). It is evident that dissolution rate improves in the presence of enzyme.

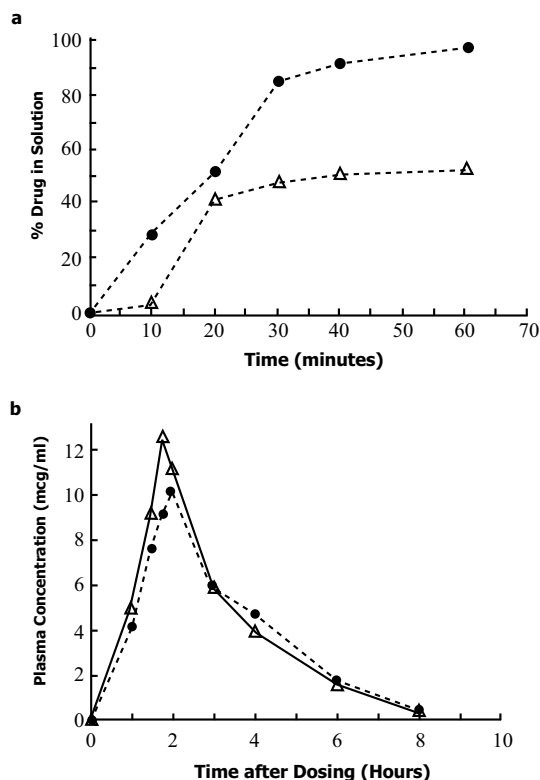


Fig. 2 Comparison of dissolution (a) and bioavailability (b) of capsules containing a high dose hydrophobic drug. Key: ●, fresh capsule; △, capsules stored for 11 mo at room temperature. The plot shows that while dissolution is affected on prolonged storage, there is no corresponding change in bioavailability profiles. (From Ref. [7], p. 16, by courtesy of Wiley-Liss Inc., a subsidiary of John Wiley & Sons Inc.)

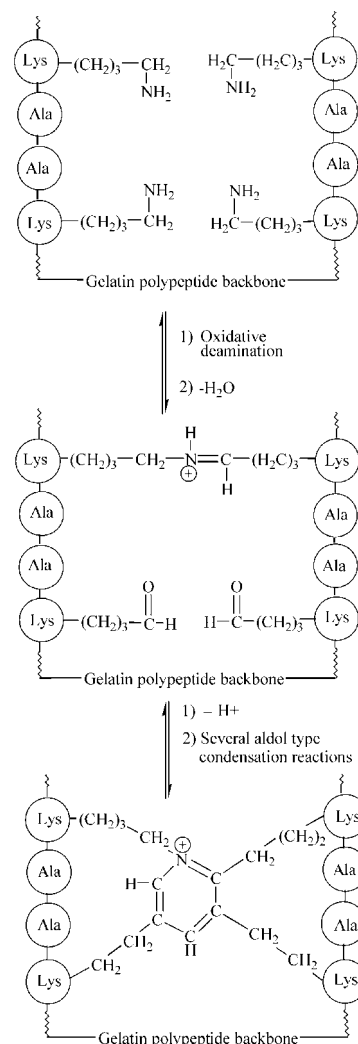


Fig. 4 Cross-linking reaction involving free amino groups of lysine. (From Ref. [1], p. 916. Reprinted by courtesy of Wiley-Liss Inc., a subsidiary of John Wiley & Sons, Inc.)

pepsin and pancreatin, are recommended, depending upon the pH of the dissolution medium. Purified pepsin resulting in an activity of 750,000 units or less per 1000 mL is suggested for the conditions where a monograph recommends water or a medium with a pH less than 6.8. For the medium with pH of 6.8 or greater, pancreatin is added at not more than 0.05 g per 1000 mL. Efforts currently are directed towards extension of the USP two-tier test to formulations containing insoluble drugs for which the use of nonionic surfactants combined with pepsin has been explored.^[8]

Nevertheless, the problem of pellicle formation and eventual fall in dissolution rate is still of concern, as the drug bioavailability may be influenced if there is a severe challenge. There is a report where exposure of phenytoin capsules to high humidities resulted in poor dissolution as well as destruction of clinical efficacy.^[14] Moreover, the enzyme test is not official in pharmacopoeias other than USP and the products stand a chance of being recalled, if the normal pharmacopoeial dissolution limits are not met.

THE CHEMISTRY OF PELLICLE FORMATION

Gelatin is a mixture of water-soluble proteins derived from collagen. It is a linear polymer with molecular weight ranging between 15,000 and 250,000. The protein fractions consist of variety of amino acids joined together by an amide linkage. The amino acids include glycine 25.5%; proline 18.0%; hydroxyproline 14.1%; glutamic acid 11.4%; alanine 8.5%; arginine 8.5%; aspartic acid 6.6%; lysine 4.1%; leucine 3.2%; valine 2.5%; phenylalanine 2.2%; threonine 1.9%; isoleucine 1.4%; methionine 1.0%; histidine 0.8%; tyrosine 0.5%; serine 0.4%; cystine 0.1%; and cysteine 0.1%.^[1] The pellicle formation is attributed mainly to trifunctional amino acids, especially lysine. Some reports have also suggested involvement of histidine and arginine.

The changes take place through three main mechanisms.^[1] First is the oxidative deamination of lysine residues, which are proximal to each other, resulting in terminal aldehyde groups. The aldehyde groups combine with free ϵ -amino group of a neighboring lysine to yield

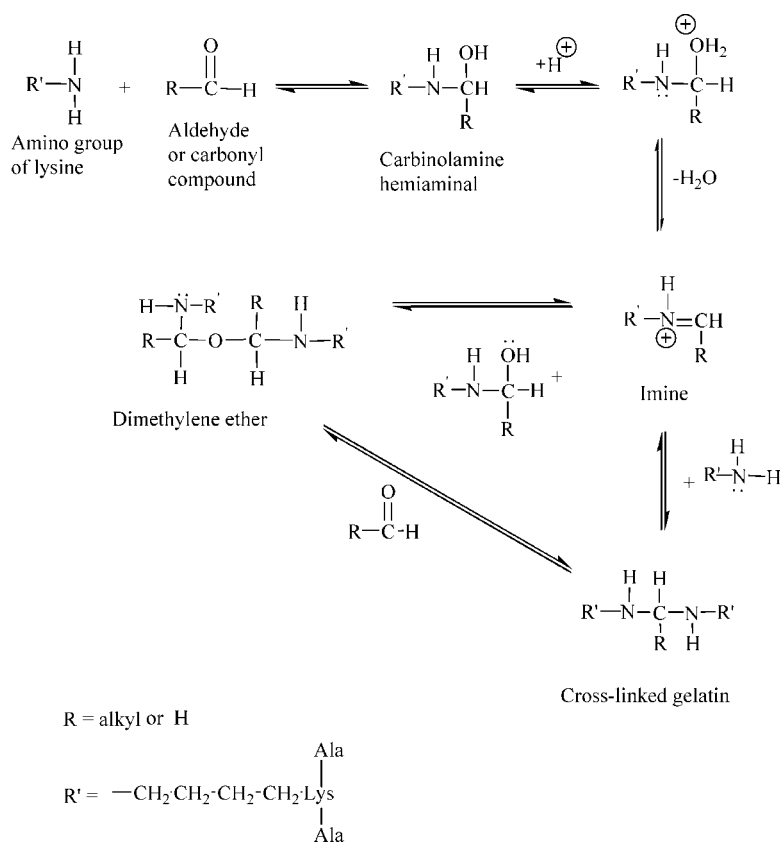


Fig. 5 Cross-linking reactions of gelatin mediated through cationic imine formation. (From Ref. [1], p. 917. Reprinted by courtesy of Wiley-Liss Inc., a subsidiary of John Wiley & Sons, Inc.)

an imine, which subsequently undergoes a series of aldol type condensation reactions to produce a cross-linked product containing pyridinium ring(s). This mechanism is described in Fig. 4. The second mechanism involves reaction of the lysyl ϵ -amino groups with an external aldehyde, present in the formulation as an impurity or generated in situ on exposure of formulations to adverse environmental conditions. This reaction yields a hydroxymethylamino derivative, which loses water to form a cationic imine. The latter reacts with another

hydroxymethylamino lysine residue to form dimethylene ether, which eventually rearranges to form a methylene link between two lysyl ϵ -amino groups, resulting in development of a cross-link. The third mechanism is the formation of an aminal, the amine form of an acetal, which is generated on reaction of cationic imine intermediate (see upper point) with a free amino group. The pH of surroundings plays an important role in this reaction. The last two mechanisms are depicted in Fig. 5.

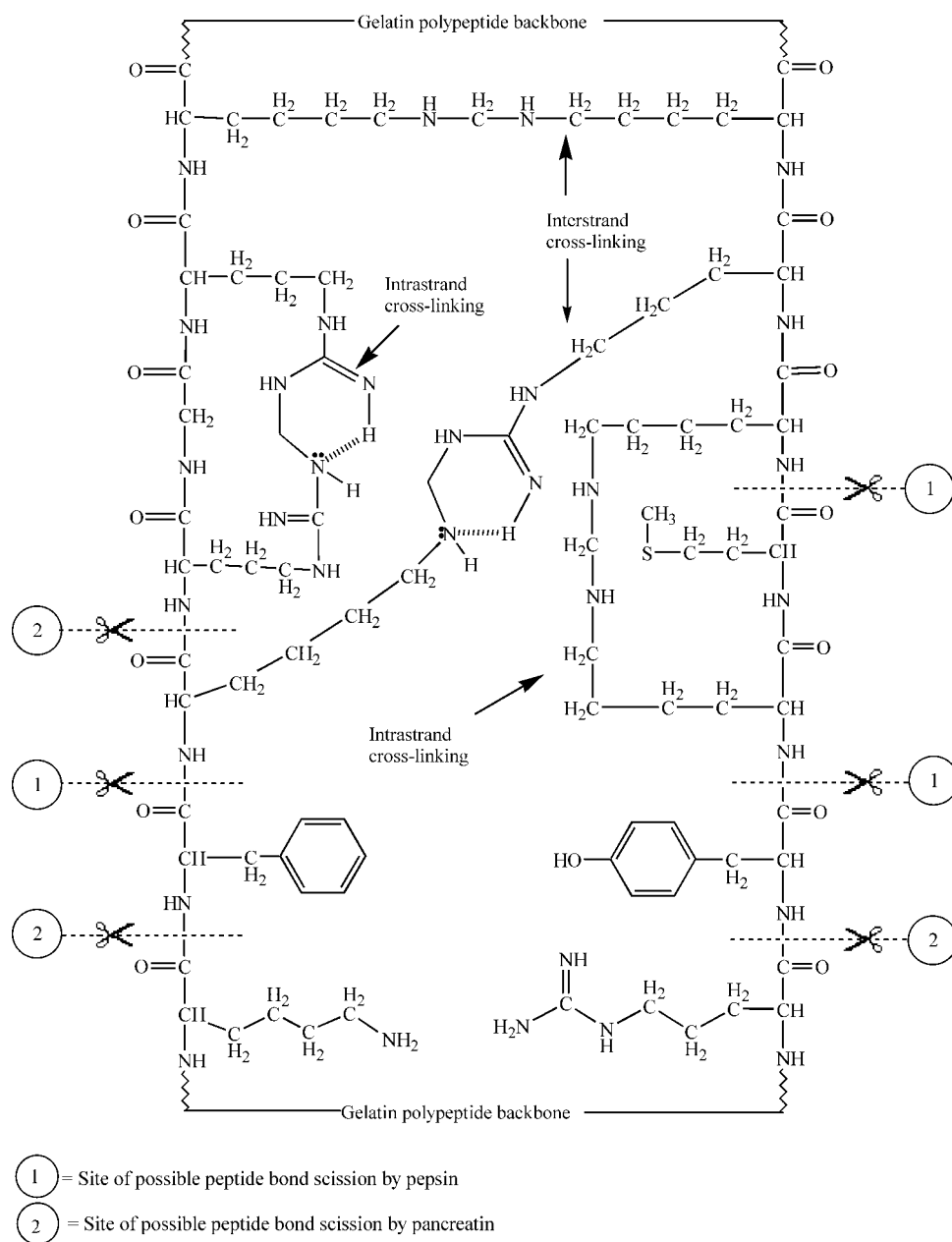


Fig. 6 Cross-linking through formation of interstands and intrastands bridges across polypeptide backbone. (From Ref. [1], p. 918. Reprinted by courtesy of Wiley-Liss Inc., a subsidiary of John Wiley & Sons, Inc.)



The cross-linking of the gelatin polypeptides may occur either through bridging within the same polypeptide strand (intrastrand, intramolecular cross-linking) or between amino acid residues from two neighboring peptide strands (interstrand, intermolecular cross-linking).^[1] The formation of strands is shown in Fig. 6. As a result, the interparticulate bonds formed in the original compact are removed and replaced by new bonds resulting in the dosage form that has different porosity and pore structure. Hence different in vitro release pattern is obtained as compared to the original.^[17]

TYPE OF GELATIN PRODUCTS THAT SHOW CHANGES IN DISSOLUTION BEHAVIOR

The altered dissolution behavior due to pelliculization occurs mainly in those products that contain gelatin in the outer layer. Typical dosage forms are hard and soft gelatin capsules and sugar-coated tablets.

Table 1 lists some of the examples where dissolution problems have been reported with hard gelatin capsules. Fig. 7 depicts the behavior in one such case. Hard capsules normally contain ~13%–16% water, which acts as a plasticizer and imparts flexibility. The moisture variations in the range of 12%–18% do not seriously impair the shell structure, however, below 12%, the shells become brittle and get easily ruptured. The capsules become moist, soft, and distorted when the moisture rises above 18%. Even otherwise, moisture can be transferred from shells to deliquescent and hygroscopic contents or the reverse flow of moisture can result from contents to shell in case of efflorescent ingredients, causing potential softening and sticking of the capsules. The transfer of moisture between capsule contents and the shell is proposed to be one of the reasons for changed properties and eventual retardation of drug release.^[22] For example, Geogarakis, Hatzipantou, and Kountourelis^[23] observed a significant retardation in dissolution rate of ampicillin trihydrate hard gel capsules on storage under varying humidities (50%–90% RH).

Table 1 Literature reports on changes in dissolution characteristics of hard gelatin capsules

Drug	Storage conditions				Effect on Dissolution	Reference
	Temperature (°C)	% RH	Illumination	Time period		
Gemfibrozil	37	—	—	1 mo, 2 mo, and 3 mo	—	5
Poorly water soluble drug	37	80	—	On-going stability study	Significant ↓ at 1 mo	8
	45	—	—		—	
Etodolac	40	75	—	8–20 weeks	Significant ↓ in all	15
	40	75	—		Significant ↓	
Chloramphenicol	25	49	—	32 weeks	No change	16
		66			No change	
		80			No release up to 1 hr	
Nitrofurantoin	40	79	—	2 and 10 weeks	Significant ↓ in 10 week samples	17
Hydrophobic drug in various colored capsules	—	80	Ambient light	2 weeks	Significant ↓	18
		80	Fluorescent	2 weeks	Significant ↓	
Hydrophobic drug in clear capsules	—	80	UV	2 day	Significant ↓	
		80	Ambient light	4 weeks	No change	
Triamterene/hydrochloro-thiazide	40	85	Fluorescent	4 weeks	No change	19
		85	—	4 weeks	Significant ↓ for both drugs	
Acetaminophen	40	75	—	55 day	Significant ↓	20
	25	60	—	52 weeks	Significant ↓	
Hard gelatin capsules shells	81	37	—	12–14 weeks and 21 weeks	Significant ↓ in all	21

↓ = Decrease in dissolution.

RT = Room temperature.

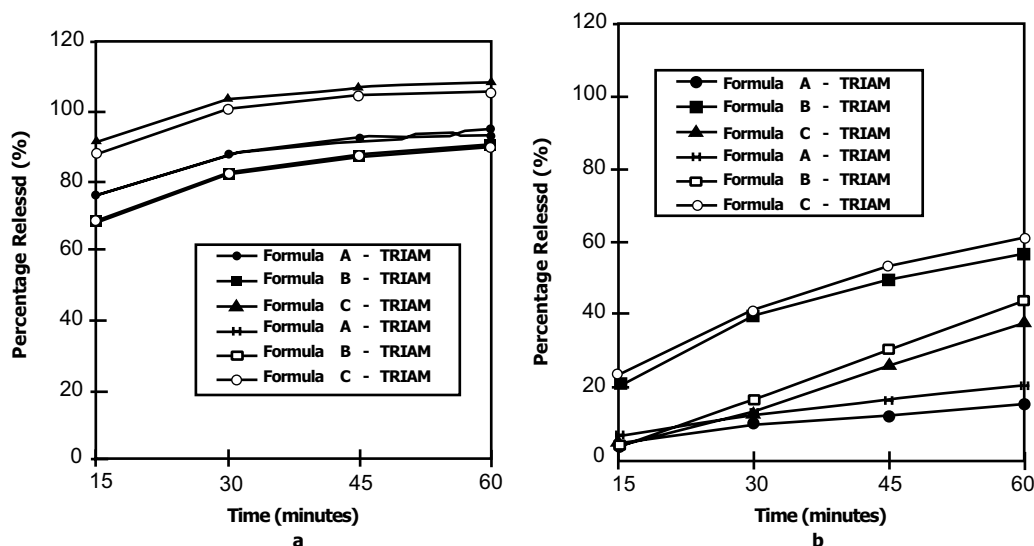


Fig. 7 Dissolution profiles of hard gelatin capsule formulations, three each containing triamterene (TRIAM) and hydrochlorothiazide (HCTZ). Key: Fresh capsules (a); and capsules stored at 40°C/85% RH for 4 weeks (b). Formula A is control, Formula B contains glycine, and Formula C contains citric acid. Comparison of the two figures clearly shows fall in dissolution on exposure of all formulations to high temperature and humidity conditions. (From Ref. [19], p. 497 by courtesy of Marcel Dekker, Inc.)

The observed behavior was attributed to the agglomeration and subsequent caking of the capsule contents due to moisture transfer from the shell.

The reports from literature on dissolution problems with soft gel capsules are summarized in Table 2. Soft-shell capsules, like hard capsules, are also influenced by environmental and chemical factors, but the extent of pelliculization is higher due to a larger mass of gelatin in soft capsules.^[30] This is well projected in Fig. 8. The plot gives a comparison of release profiles of amoxicillin

trihydrate from hard and soft gelatin capsules exposed to similar type of storage conditions.^[29] Evidently, soft gel capsules show more severe change. Under high humidity, soft capsules become soft, tacky, and bloated, offering a possibility not only of migration of moisture from shell to capsule contents, but also of reverse flow of chemical catalysts, when present in solvents used in preparation of drug solutions or dispersions. The shells of soft capsules also contain plasticizers, typically glycerine or sorbitol, PEG, and ethers of polyethylenated glycosides, along with

Table 2 Literature reports on changes in dissolution characteristics of soft gelatin capsules

Drug	Storage conditions			Time period	Effect on dissolution	Reference
	Temperature (°C)	% RH	Illumination			
Acetaminophen	40	75		55 day	Significant ↓	20
	25	60		52 weeks	Significant ↓	
Digoxin	5, 25, and 37	—	—	1, 3, 6, and 10 weeks	Significant ↓ in 10 weeks sample	24
Medium chain triglycerides	40 or more	—	—	6 mo	Significant ↓	25, 26
Acetaminophen and nifedipine	25	60	—	2–26 weeks	Significant ↓	27
	40	75			Significant ↓	
Vitamins	40	75	—	6 mo and 24 mo	Significant ↓ in all	28
Nimesulide	40	75	Fluorescent and UV	8 day	Significant ↓	29

↓ = Decrease in dissolution.

RT = Room temperature.

Gelatin-Containing Formulations: Changes in Dissolution Characteristics

7

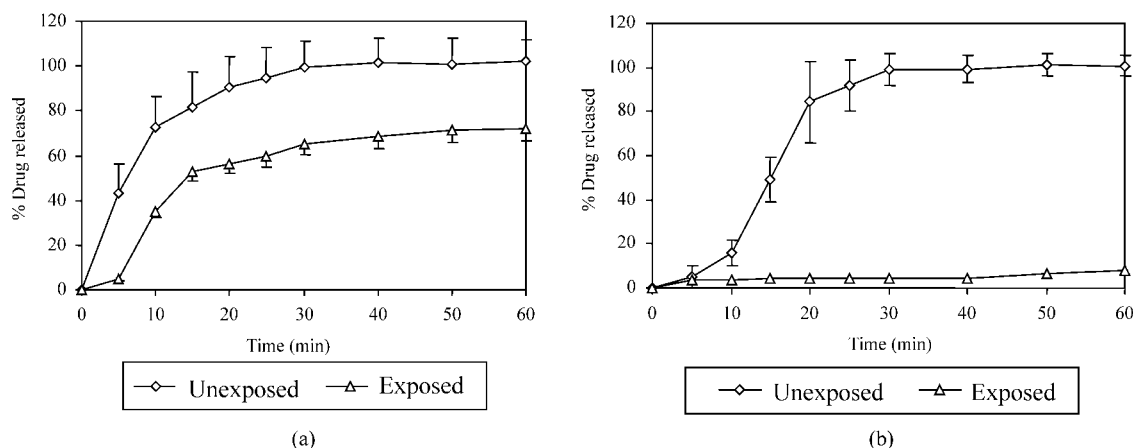


Fig. 8 Release profiles of hard (a) and soft gelatin (b) capsules of amoxycillin trihydrate in unexposed condition and after exposure to 40°C/75% RH/light for 8 day. The soft gelatin capsules show more drastic fall in dissolution upon exposure than the hard gels.

gelatin and water and any catalytic impurities present in these materials can also contribute to chemical and dissolution instability.^[31,32]

Table 3 gives a summary of the known dissolution problems with sugar-coated tablets. The progressive decrease in disintegration and dissolution has been ascribed to adherence of gelatin subcoat to the tablet core.^[33] This was also a conclusion made in a study where most of the core tablets were found to be dry even at the end of the dissolution test.^[34] Fig. 9 shows an example of sugar-coated chloroquine phosphate tablets exposed to 40°C and 75% RH. Clearly, the dissolution rate decreases with time.^[12]

There also exists a report in literature on increase in both disintegration and dissolution time even in tablets containing gelatin as a binder. This happened when tablets were stored at 50°C/83% RH and 70°C/96% RH for 7 weeks.^[37] However, not much decrease in dissolution was observed in another study, when similar dosage form was

stored under milder condition of 40°C/75% RH and light for 3 weeks.^[38] It shows that pellicle formation may occur even in other gelatin-containing dosage forms, when exposed to severe storage conditions. Further, it confirms the contention that the problem of pellicle formation is an easy occurrence in dosage forms containing gelatin in the outer layer.

In one of the studies, sensitive gelatin dosage forms were also evaluated for the extent of cross-linking in the packaged form.^[29] Fig. 10a–c depicts the behavior of dissolution of hard and soft gelatin capsules in blisters, after exposure to accelerated conditions of temperature and humidity in a photostability chamber. The data for fresh and unpacked formulations are also included for comparison. Fig. 10a shows no significant change in the dissolution of packed amoxycillin trihydrate hard gelatin capsules as compared to the unexposed product, while the dissolution evidently is delayed in case of directly exposed capsules. The behavior of two soft gelatin

Table 3 Literature reports on changes in dissolution characteristics of sugar-coated tablets

Drug	Storage conditions				Effect on dissolution	Reference
	Temperature (°C)	% RH	Illumination	Time period		
Acetaminophen	RT	—	—	7 mo	No change	9
	RT	High	—	3 mo, 5 mo, and 7 mo	Significant ↓	
Phenylbutazone	20, 37, 50	—	—	Varying between 2 and 14 weeks	↓ in those stored at 50°C for 14 weeks	33
Valproic acid	45	—	—	1 mo, 2 mo, and 3 mo	Significant ↓ in 2 mo and 3 mo samples	34
Ibuprofen	40	75	—	4 weeks	Significant ↓	35
Riboflavin	37	75	—	—	Significant ↓	36
	45	—	—	—	Significant ↓	

↓ = Decrease in dissolution.

RT = Room temperature.

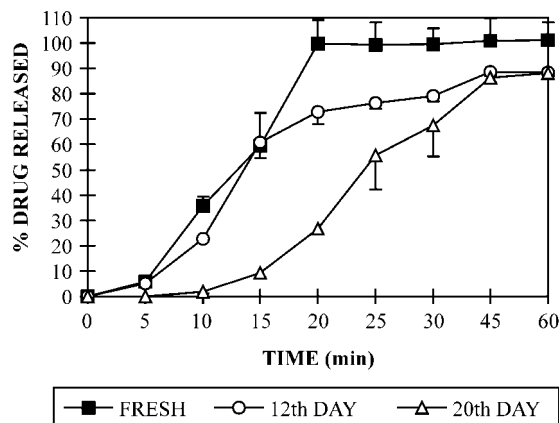


Fig. 9 Release of chloroquine phosphate from marketed sugar-coated tablets on storage at 40°C and 75% RH up to 20 day. Evidently, the dissolution rate falls with increase of the duration of storage under accelerated conditions of temperature and humidity.

capsules in blisters (Fig. 10b and c) shows that changes in dissolution, from minor to complete, occur in packed formulations as compared to unexposed samples. The unpacked capsules of course undergo drastic changes. The comparison of the three figures confirms the previous contention that soft gels are more severely cross-linked than hard gels, and it additionally shows that blister packs may provide complete to nil protection, depending upon the drug and the nature of capsules, whether hard or soft. The release profiles of trifluoperazine sugar-coated tablets packed in strip packaging (Fig. 10d) show that the release of the drug from both the strip-packed and unpacked tablets is delayed by around 20 min, after which the profiles merge with those of the unexposed samples. The lag period perhaps appears out of the influence of temperature, as strips are expected to be impervious to both humidity and light.

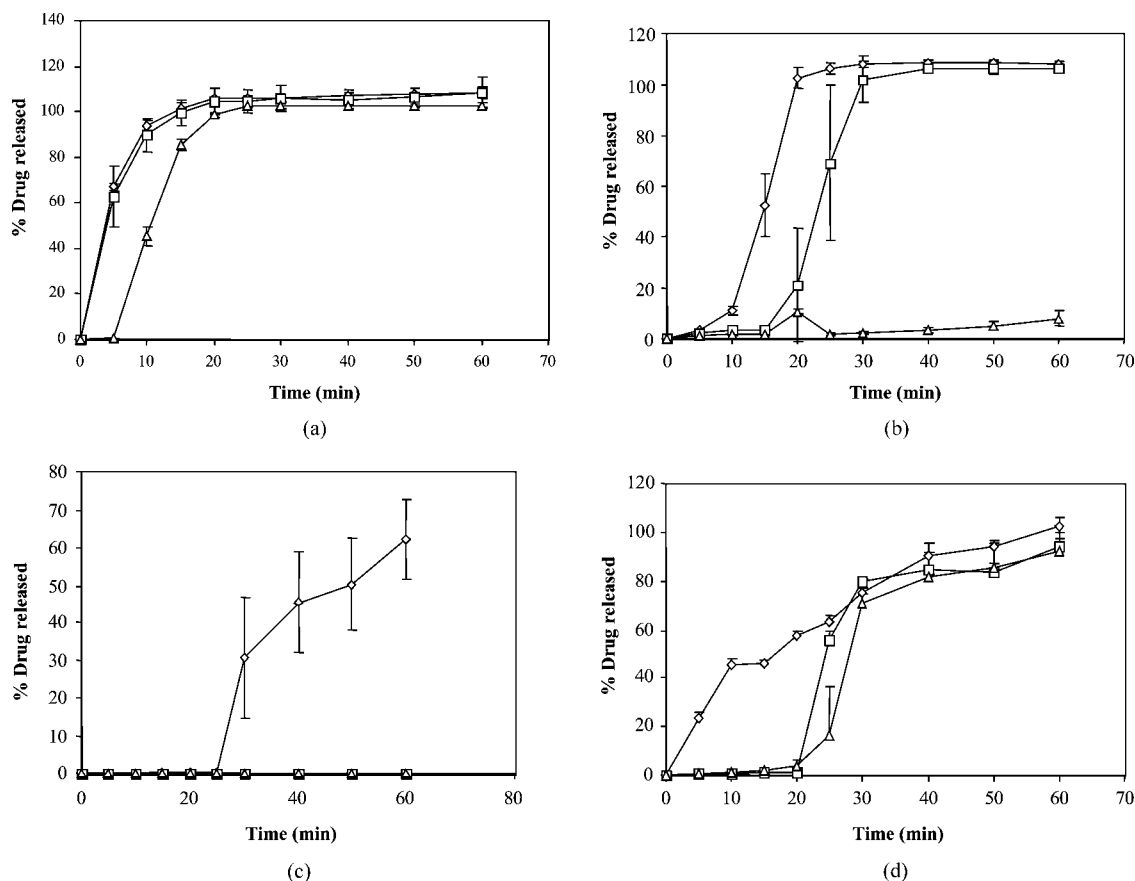


Fig. 10 Release profiles of hard gelatin capsule containing amoxicillin trihydrate (a); soft gelatin capsules containing cephalixin (b); soft gelatin capsules containing nimesulide (c); and sugar-coated tablets containing trifluoperazine (d) upon storage at 40°C/75% RH in the presence of light.

FACTORS INDUCING THE CHANGE

Tables 1–3 and Figs. 1, 2a, 3, 7b, and 8–10 show that a significant fall in dissolution can take place in gelatin formulations, when exposed to normal or accelerated environmental conditions of temperature, humidity, and/or light for varied durations. It clearly means that the pellicle formation is catalyzed by environmental factors, either alone or in combination.

The cross-linking reaction is normally slow at room temperatures, and it may take several weeks or even months before a perceptible change is observed. High temperature accelerates pellicle formation through the process of denaturation by increasing the rate of cross-linking reactions and a typical example is depicted in Fig. 11.^[25] The pellicle formation is also accelerated when high humidity is combined with high temperature. Humidity plays an independent role, influencing pellicle formation in several ways, like indirect catalysis of imine formation, catalysis of excipient decomposition yielding products, which cause cross-linking of gelatin, and as a vehicle for denaturation of gelatin. The rate of cross-linking is further influenced by UV and visible irradiation, when combined with humidity^[18] or humidity and temperature.^[38]

Apart from the environmental factors, chemical compounds also induce cross-linking of gelatin. Supporting this is the mechanism in Fig. 5, which highlights the role of external aldehydes. Among the low molecular weight aldehydes, formaldehyde is most important as it is released in dosage forms from plasticizers and preservatives, fats, and polyethylenated compounds such as PEG, ethers of PEG, and aliphatic alcohols or phenols, polyethylenated glycerides, nonionic surfactants (polysorbates, esters of unsaturated fatty acids), and corn starch.^[4–6,39,40] The corn starch at times contains traces of stabilizer hexamethyl

tetramine, which decomposes under humid conditions to form ammonia and formaldehyde.^[4] Accordingly, a lot of work has been done to establish correlation between the concentration of formaldehyde and the extent of reduction in dissolution of gelatin-containing preparations.^[21,41–44] Other aldehydes reported to influence integrity of gelatin include furfural, acrolein, glutaraldehyde, and glyceryl aldehyde.^[1,5,6,19,27,45,46] An interesting case is the negative effect on the dissolution rate of gelatin capsules in the presence of rayon coiler, which is filled in the headspace of HDPE bottles containing capsules.^[45,47] The rayon produces furfural, which when present in saturated vapor phase, rapidly insolubilizes gelatin capsules. The reported effect is depicted in Fig. 12.

There are several other chemicals, which induce pelliculization of gelatin. These are saccharides, e.g., glucose and aldose sugars^[11]; imines and ketones^[21]; dyes like FD&C red No. 3 and 40^[10,48]; hydrogen peroxide, benzene, sulfonic acids, *p*-toluene sulfonic acid^[19,49]; carbodiimides, e.g., 1-ethylene 3-(3-dimethylamino propyl) carbodiimide hydrochloride, guanidine

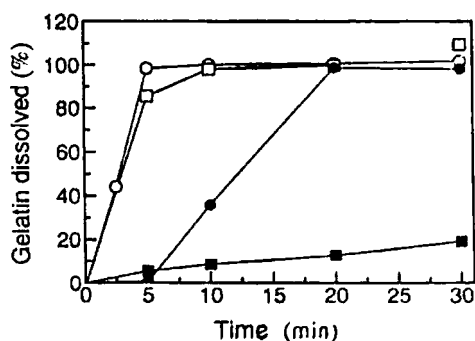


Fig. 11 The behavior of dissolution of gelatin shell with time at different temperatures. Key: ○, initial; □, sample stored for 6 mo at 25°C; ●, sample stored for 6 mo at 40°C; and ■, sample stored for 6 weeks at 60°C. (From Ref. [25], p. 1498. Reprinted by courtesy of the authors and The Pharmaceutical Society of Japan.)

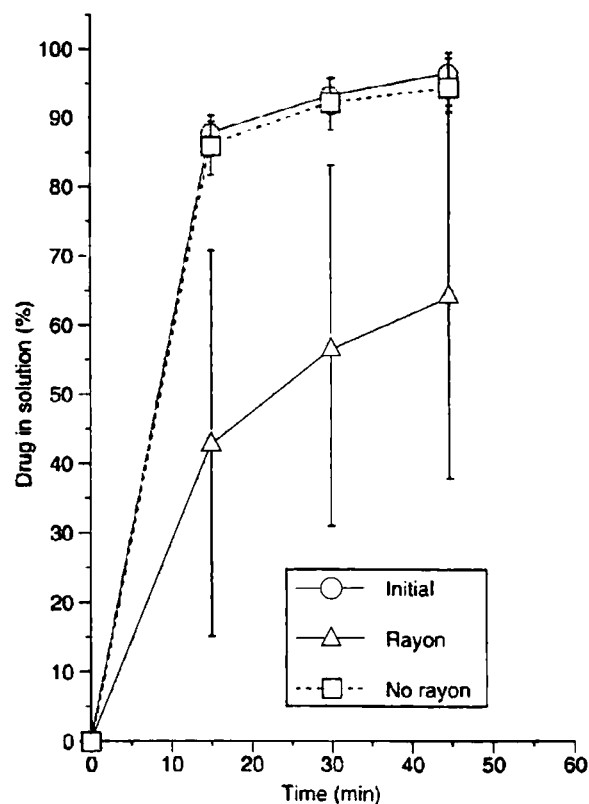


Fig. 12 A typical example of the decrease in the rate of drug dissolution in the presence of rayon after storage of formulation at 40°C/75% RH for 2 mo. (From Ref. [47], p. 82. Reprinted by courtesy of Advanstar Communications.)

hydrochloride^[19,46]; terephthaloyl chloride^[50]; and calcium carbonate.^[51,52]

INDUCTION OF PELLICLE FORMATION

It is always desirable to have some sort of preassessment whether dosage forms containing gelatin would show a decrease in dissolution rate due to pellicle formation with time on storage. Sometimes, the formulations are also intentionally subjected to stress conditions to judge the efficacy of enzymes in the dissolution medium for their capacity to overcome pellicle formation. For this purpose, a few approaches have been used successfully.

One example is the addition of formaldehyde at concentrations of 0 ppm, 20 ppm, 80 ppm (or 120 ppm) in lactose and filling of this material in soft and hard gel capsules.^[20,21,27,53] Formulations, so prepared, are stored at room temperature and/or accelerated conditions of 40°C and 75% RH. The samples are drawn at various times and subjected to dissolution in plain and enzyme containing media. Comparison of results gives an idea on the susceptibility of the formulation towards pelliculization. However, one disadvantage of this method is that at times notable retardation of drug dissolution occurs only after storage of the product at room temperature/accelerated conditions for several days or months.^[20,27,30] The long period of storage prevents carrying out of repeated trials, particularly during formulation and packaging development.

A better, versatile, and rapid method is the preparation of films of raw gelatin material (according to the formula for hard or soft gelatin capsules) and exposing small

pieces to formaldehyde vapors in a closed chamber for 12 hr. The set up and conditions were described in a recent report.^[43] Some workers also use a dipping method to expose gelatin film to aldehyde.^[26,50,54] The major advantage of the direct formaldehyde exposure methods is that it takes shorter time, of only a few hours. The method was used recently for classifying raw gelatins in order of their sensitivity to cross-linking.^[43]

Another method, which also gives similar results, involves simultaneous exposure of gelatin films to all the three environmental factors, viz., temperature, humidity, and light. The films or pieces are exposed for 8 day in a photostability chamber at 40°C/75% RH, with a total illumination of 2 million lux hr visible light and UV light of $>200 \text{ W hr m}^{-2}$.^[43] The samples are withdrawn and subjected to dissolution studies. Although the test takes 8 day, the major advantage of this test method is that it simulates the environmental conditions, to which the product is expected to be exposed during its manufacturing, transportation, distribution, and storage. Another benefit is that formulations can be directly studied in unpacked and packed state.

EVALUATION OF THE EXTENT AND NATURE OF CROSS-LINKING

There are several ways to investigate the extent and nature of cross-linking undergone by gelatin and its products as a consequence of exposure to environmental factors and chemical catalysts. Table 4 provides the list of methods and their applications.

A simple method to determine the extent of gelatin cross-linking is the determination of the solubility of

Table 4 Methods for the determination of the extent and nature of cross-linking in gelatin films and formulations

Technique	Application
Monitoring of solubility and dissolution	Determination of the extent of cross-linking
Gravimetric analysis, a protein assay method (involving color reaction with bicinchoninic acid) and UV absorbance measurements at 214 nm	Determination of the extent of noncross-linked gelatin
Chemical analysis using TNBS reagent	Loss of ϵ -amino groups upon cross-linking
Fluorescence spectrophotometry	Conformational changes of gelatin in gel state
¹³ C-NMR spectroscopy	Determination of the types of cross-links developed during reaction with formaldehyde
FT-Infrared (FT-IR) spectroscopy	Determination of the types of cross-linking
FT-Near Infrared (FT-NIR) spectroscopy	Determination of water uptake by the dosage form, extent of cross-linking and monitoring of migration of formaldehyde from contents to the shell
Magnetic Resonance Imaging (MRI)	Study of diffusion of ions into the gelatin–water matrix

the films or dissolution of formulations.^[43,55] The fall in solubility/dissolution has been shown to be linearly correlated to the extent of cross-linking.^[43] Thus a good idea on the extent of change undergone by films or formulations at any time under the specific challenge can be obtained from these simple tests. There also exist some indirect methods like gravimetric analysis, a protein assay method (involving color reaction with bicinchoninic acid), and UV absorbance measurements at 214 nm to determine gelatin in the dissolved or an undissolved state.^[21,56] A furthermore in-depth study can be done by the use of a chemical assay employing 2,4,6-trinitrobenzenesulfonic acid (TNBS).^[21,46,54,56] The reagent reacts with primary amino groups of gelatin and helps to know the loss of ϵ -amino groups, which participate in the cross-linking process.

Determination of intrinsic fluorescence can be used to investigate conformational changes undergone by gelatin gels.^[57] On the other hand, the mechanism and site of development of cross-links can be studied by the use of ¹³C-NMR spectroscopy. This technique has been successfully used for the determination of the involvement of amino groups in lysine-lysine, lysine-arginine, and arginine-arginine cross-links, subsequent to reaction with formaldehyde.^[58-61] Using the same method, Gold, Smith, and Digenis^[59] established that pancreatin, a proteolytic enzyme present in the gastrointestinal tract, depolarizes cross-linked gelatin. Some workers have also explored the use of FT-IR and FT-NIR spectroscopy. Salsa, Pina, and Teixeira-Dias^[62] employed FT-IR for the determination of the types of cross-links developed during reaction with formaldehyde. The FT-NIR spectroscopy was used for the determination of water uptake,^[63] extent of cross-linking, and for monitoring migration of formaldehyde from capsules contents into the gelatin shell.^[41] The combination of NIR spectroscopy with principal component analysis, a multivariate analysis, was found to give good predictable results.^[41,63] Recently, Magnetic Resonance Imaging (MRI) has been found suitable for the study of diffusion of ions into the gelatin-water matrix and the setting of gelatin on cross-linking.^[64]

APPROACHES FOR THE DEVELOPMENT OF STABLE GELATIN FORMULATIONS

Looking into the wide-spread use of gelatin-containing dosage forms and the commonality with which the problem of cross-linking or pellicle formation is observed, there exists a good deal of interest in developing stable formulations that would not show problem of reduced dissolution with time. The solutions, of course, lie in

countering the influence of environmental and/or chemical catalytic factors, discussed above.

The simplest approach is the use in formulations of gelatin grades that are resistant to the influence of environmental factors as well as chemical catalysts. Such gelatins are sold commercially. The enquiry about the resistance of the gelatin raw material to cross-linking should be made from the manufacturer/supplier before purchase of the material. Generally, gelatins with bloom strength lower than 250 are less prone to pellicle formation and there exists no difference with respect to cross-linking among the types A and B of gelatins.^[65]

Otherwise, the appearance of problem due to exposure to accelerated and stress environmental conditions (Tables 1-3) can largely be overcome by storing dosage forms containing gelatin in the outer layer under cool, dry, and dark conditions. Instructions in this regard can be added on the labels of the gelatin-based products. The manufacturers at their level can help control the problem by paying particular attention to moisture, as it induces pellicle formation in a variety of ways. The humidity is required to be controlled even during the manufacture of gelatin-based formulations. The products should also be packed in moisture and light resistant packaging. For example, the stability of products packed in blisters can be improved by the use of water-impermeable films and laminates, like polyvinyl chloride (PVC)-polyvinylidichloride (PVDC), PVC-PVDC-PE, Aclar, etc. instead of the permeable PVC films used conventionally. To prevent from light, colored blister films can be used and additionally, individual blister packs should be supplied in duplex cartons. Unfortunately, this is not the practice in many parts of the world. Other better alternate is the use of strip or Alu/Alu packing.

The other solution is the careful selection of ingredients and excipients that are devoid of catalytic impurities or which do not decompose to catalytic products. One simple example here is that of polyethylene glycols, which are routinely used as solvents for drugs in soft gels. The glycols normally contains small amounts of aldehydic impurities and the best way to prevent their interaction with gelatin shell is the use of aldehyde free grade of glycol or heating of commercial grade solvent to expel volatile aldehydic impurities. Similarly, the use of excipients like corn starch can be avoided, till absolutely necessary.

Alternately, specific stabilizers can also be added into the formulations. The protection against aldehydes, whether present initially in the formulation or released in situ, can be provided by the addition of aldehyde scavengers like, lysine, phenylamine, glutamine, hydroxylamine hydrochloride, *p*-amino benzoic acid, glycine, etc.^[49] The in situ release of aldehydes as a consequence to degradation of the contents can be controlled by

the manipulation of pH using buffering agents. Carboxylic acids, such as benzoic acid, fumaric acid, maleic acid, citric acid, etc. are helpful in this direction. The synergistic combination of an amino acid and buffers in preventing pellicle formation has also been explored.^[19] Fig. 13 shows that addition of glycine and citric acid together in the formulation fill proves more effective in preventing dissolution instability than when glycine or citric acid are added alone (Fig. 7).

Some studies have also reported successful prevention of cross-linking using direct inhibitors like semicarbazide hydrochloride, hydroxylamine hydrochloride, piperazine hydrate, pyridine, piperidine, glycerine, *p*-aminobenzoic acid, etc.^[19,49] Another approach, specific to filled hard gelatin capsules, is the addition of disintegrants to the fill powder blend. Capsule formulations containing 10 or more percent of disintegrant is reported to withstand the stress of high humidity storage conditions, presumably due to the porous nature of the capsule fill.^[66] The use of titanium oxide, iron oxide, and color pigments in the capsular dosage forms containing gelatin offers a good protection against cross-linking introduced by light. A good example is that of curcumin, which at a content of 0.4% in the capsule shell, was able to result in a threefold or higher increase in the half-life of the test compounds.^[67] Dyes, like FD&C Yellow No. 5 and Blue No. 1, have also been shown to protect the dosage forms from light.^[18] The synthetic iron oxides, being potent absorbers of wavelengths below 400 nm, can be employed successfully,

except that they have an attached caution that excess iron content (> 15 ppm) can result in discoloration of soft gelatin capsules.^[32]

PATENT STATUS OF STABILIZATION APPROACHES

The use of glycine–citric acid synergistic combination in fills of gelatin capsular products for the purpose of protection against cross-linking is covered by US patent No. 5,674,106^[68] and World patent No. 9733568.^[69] There also exist a few patents where stabilization has been attempted by addition of the stabilizers in the films or gelatin raw material. The incorporation of glutamic acid, tryptophan, or nitrilotrismethylene phosphoric acid or a mixture thereof into gelatin before forming the final product has been claimed to impart improved stability against storage under hot and humid conditions and/or aldehydes.^[70] Addition of a polypeptide content from 15%–70% based on total weight of peptide and gelatin was found to prevent gelatin from becoming insoluble with time without deteriorating the shape retentivity of capsules.^[71]

FUTURE CONSIDERATIONS

Till date, the problem of pelliculization of gelatin products has found major solutions in: 1) understanding of

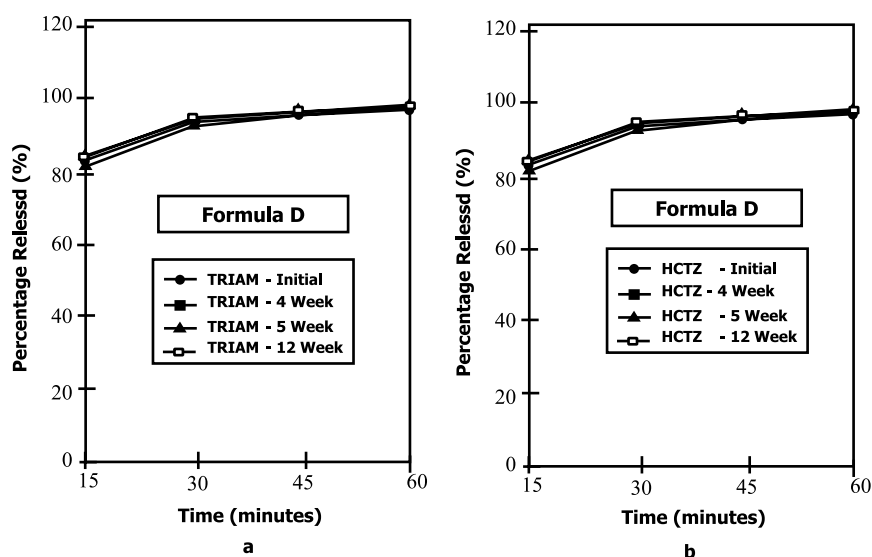


Fig. 13 The dissolution behavior of triamterene (TRIAM) (a) and hydrochlorothiazide (HCTZ) (b) formulations containing glycine and citric acid (Formula D) on storage at 40°C/85% RH up to 12 weeks. This figure is correlated to Fig. 7 and is from the same reference. The overlapping curves in both cases show that the amino acid/buffer combination is able to neutralize the adverse effect of storage under high humidity and temperature. The figure further depicts that neutralization remains effective even on storing the capsules as long as 12 weeks. (From Ref. [19], p. 498 by courtesy of Marcel Dekker, Inc.)

the chemistry; 2) identification of stabilizers that can be added in formulation fills or films; 3) development of rapid test methods for evaluation of possibility of pellicle formation; and 4) introduction of two-tier dissolution test in USP. Still, a lot has to be done, of which foremost is the adoption of two-tier test by International pharmacopoeias, other than USP. More rigorous efforts need to be done towards stabilization approaches targeted to raw gelatin so that all manufacturers of gelatins around the world are able to offer stable raw material for pharmaceutical use. Concurrently, there is a need for identification of newer types of efficient stabilizers that can be added to the gelatin film during formation of hard and soft gel capsules or in gelatin solutions meant for application as a subcoat in sugar-coated tablets. The hard gel capsule manufacturers produce billions of capsules every year and if solution is offered to them in way of stable gelatin raw material and stabilizers that can be added in the fill/film formula, this can have a very wide impact. The same concept can also be applied at the time of manufacture of soft gels. In case, successes are achieved on this front, the problem is likely to be eliminated to a large extent.

REFERENCES

1. Digenis, G.A.; Gold, T.B.; Shah, V.P. Cross-Linking of Gelatin Capsules and Its Relevance to Their In Vitro–In Vivo Performance. *J. Pharm. Sci.* **1994**, *83* (7), 915–921.
2. Carstensen, J.T.; Rhodes, C.T. Pellicule Formation in Gelatin Capsules. *Drug. Dev. Ind. Pharm.* **1993**, *19* (20), 2709–2712.
3. Lozano, R.; Agorrody, M.; Beraud, B.; Dauphin, J.-F. Gelatin Cross-Linking in Capsules and Its Effects on Dissolution. *Am. Pharm. Rev.* **2001**, *4* (4), 8–16.
4. Mohamad, H.; Renoux, R.; Aiache, S.; Aiache, J.M.; Kantelip, J.P. Investigation of Drug Biopharmaceutical Stability: Application to Capsules of Tetracycline Hydrochloride. Part 2. In Vivo Study. *S.T.P. Pharm. Pratiq.* **1986**, *2*, 630–635.
5. Chafetz, L.; Hong, W.; Tsilifonis, D.C.; Taylor, A.K.; Philip, J. Decrease in the Rate of Capsule Dissolution Due to Formaldehyde from Polysorbate 80 Autoxidation. *J. Pharm. Sci.* **1984**, *73* (8), 1186–1187.
6. Mohamad, H.; Renoux, R.; Aiache, S.; Aiache, J.M.; Kantelip, J.P. Investigation of Drug Biopharmaceutical Stability. Part 3. Study of Ampicillin Trihydrate Capsules (In Vitro/In Vivo Investigation). *S.T.P. Pharm. Pratiq.* **1986**, *2*, 912–917.
7. Murthy, K.S.; Ghebre-Sellassie, I. Current Perspectives on the Dissolution Stability of Solid Oral Dosage Forms. *J. Pharm. Sci.* **1993**, *82* (2), 113–126.
8. Shaw, J.M.; Sullivan, B.M.; Bowen, W.E.; Martin, G.P.; Reed, R.A. Studies Directed Towards the Development of Enzymatic Dissolution Test Methods for Cross-Linked Hard Gelatin Capsule Formulations Containing Poorly Water-Soluble Drug Substances. Abstracts of the AAPS Annual Meeting, New Orleans, LA, Nov 14–18, 1999; American Association of Pharmaceutical Scientists: Arlington, VA, 1999; Abstract No. 3358.
9. Dahl, T.C.; Sue, I.L.T.; Yum, A. The Effect of Pancreatin on the Dissolution Performance of Gelatin-Coated Tablets Exposed to High-Humidity Conditions. *Pharm. Res.* **1991**, *8* (3), 412–414.
10. Murthy, K.S.; Reisch, R.G.J.; Fawzi, M.B. Dissolution Stability of Hard-Shell Capsule Products, Part II: The Effect of Dissolution Test Conditions on In Vitro Drug Release. *Pharm. Technol.* **1989**, *13* (6), 53–58.
11. Hom, F.S.; Veresh, S.A.; Miskel, J.J. Soft Gelatin Capsules I: Factors Affecting Capsule Shell-Dissolution Rate. *J. Pharm. Sci.* **1973**, *62*, 1001–1006.
12. Manikandan, R. Studies on the Retardation of Dissolution of Gelatin Containing Formulations on Aging. M. Pharm. (Formulation) Thesis, NIPER, S.A.S. Nagar, India, 1999.
13. United States Pharmacopoeia. 25th Edition, Physical Test and Determinations. 1999, p. 7516.
14. Martin, E.D.; Fraser, R.J.L.; Camens, I. Storage of Phenytoin Capsules. *Med. J. Aust.* **1985**, *143*, 634–635.
15. Dey, M.; Enever, R.; Kraml, M.; Prue, D.G.; Smith, D.; Weierstall, R. The Dissolution and Bioavailability of Etodolac from Capsules Exposed to Conditions of High Relative Humidity and Temperatures. *Pharm. Res.* **1993**, *10* (9), 1295–1300.
16. Khalil, S.A.H.; Ali, L.M.M.; Khalek, A.M.M.A. Effects of Aging and Relative Humidity on Drug Release. Part 1: Chloramphenicol Capsules. *Pharmazie* **1974**, *29* (1), 36–37.
17. Gouda, H.W.; Moustafa, M.A.; Al-Shora, H.I. Effect of Storage on Nitrofurantoin Solid Dosage Forms. *Int. J. Pharm.* **1984**, *18*, 213–215.
18. Murthy, K.S.; Enders, N.A.; Fawzi, M.B. Dissolution Stability of Hard-Shell Capsule Products, Part I: The Effect of Exaggerated Storage Conditions. *Pharm. Technol.* **1989**, *13* (3), 72–84.
19. Adesunloye, T.A.; Stach, P.E. Effect of Glycine/Citric Acid on the Dissolution Stability of Hard Gelatin Capsules. *Drug Dev. Ind. Pharm.* **1998**, *24* (6), 493–500.
20. Meyer, M.C.; Straughn, A.B.; Mhatre, R.M.; Hussain, A.; Shah, V.P.; Bottom, C.B.; Cole, E.T.; Lesko, L.L.; Mallinowski, H.; Williams, R.L. The Effect of Gelatin Cross-Linking on the Bioequivalence of Hard and Soft Gelatin Acetaminophen Capsules. *Pharm. Res.* **2000**, *17* (8), 962–966.
21. Ofner, C.M., III.; Zhang, Y.E.; Jobeck, V.C.; Bowman, B.J. Cross-Linking Studies in Gelatin Capsules Treated with Formaldehyde and in Capsules Exposed to Elevated Temperature and Humidity. *J. Pharm. Sci.* **2001**, *90* (1), 79–88.
22. Chang, R.K.; Raghavan, K.S.; Hussain, M.A. A Study on Gelatin Capsule Brittleness: Moisture Transfer Between the Capsule Shell and Its Content. *J. Pharm. Sci.* **1998**, *87* (5), 556–558.
23. Georgarakis, M.; Hatzipantou, P.; Kountourelis, J.E. Effect of Particle Size, Content in Lubricant, Mixing Time and Storage Relative Humidity on Drug Release from Hard

- Gelatin Ampicillin Capsules. *Drug Dev. Ind. Pharm.* **1988**, *14* (7), 915–923.
24. Johnson, B.F.; Mcauley, P.V.; Smith, P.M.; French, J.A.G. The Effects of Storage upon *In Vitro* and *In Vivo* Characteristics of Soft Gelatin Capsules Containing Digoxin. *J. Pharm. Pharmacol.* **1977**, *29*, 576–578.
25. Hakata, T.; Sato, H.; Watanabe, Y.; Matsumoto, M. Effect of Storage Temperature on the Physicochemical Properties of Soft Gelatin Capsule Shells. *Chem. Pharm. Bull.* **1994**, *42* (7), 1496–1500.
26. Hakata, T.; Sato, H.; Watanabe, Y.; Matsumoto, M. Effect of Formaldehyde on the Physicochemical Properties of Soft Gelatin Capsule Shells. *Chem. Pharm. Bull.* **1994**, *42* (5), 1138–1142.
27. Bottom, C.B.; Clark, M.; Carstensen, J.T. Dissolution Testing of Soft Shell Capsules—Acetaminophen and Nifedipine. *J. Pharm. Sci.* **1997**, *86* (9), 1057–1061.
28. Sukur, K. Disintegration Test Media for Soft Gelatin Capsules—Need for Global Harmonization. Abstracts of the 1999 AAPS Annual Meeting, New Orleans, LA, Nov. 14–18, 1999; American Association of Pharmaceutical Scientists: Arlington, VA, 1999; Abstract No. 3350.
29. Venugopal, K. Studies on Stability of Gelatin Based Products. M. Pharm. (Formulation) Thesis, 2000, NIPER, S.A.S. Nagar, India.
30. Brown, J.; Madit, N.; Cole, E.T.; Wilding, I.R.; Cade, D. The Effects of Cross-Linking on the *In Vivo* Disintegration of Hard Gelatin Capsules. *Pharm. Res.* **1998**, *15* (7), 1026–1030.
31. Seager, H. Soft Gelatin Capsules: Solution to Many Tableting Problems. *Pharm. Technol.* **1985**, *9*, 84–104.
32. Jimerson, R.F. Soft Gelatin Capsule Update. *Drug Dev. Ind. Pharm.* **1986**, *12* (8&9), 1133–1144.
33. Barrett, D.; Fell, J.T. Effect of Aging on Physical Properties of Phenylbutazone Tablets. *J. Pharm. Sci.* **1975**, *64* (2), 335–337.
34. Shah, I.G.; Parsons, D.L. Preparation and In-Vitro Evaluation of Powder Solution Tablets of Valproic Acid. *Drug Dev. Ind. Pharm.* **1993**, *19* (13), 1587–1602.
35. Romero, A.J.; Grady, L.T.; Rhodes, C.T. Dissolution Testing of Ibuprofen Tablets. *Drug Dev. Ind. Pharm.* **1988**, *14* (11), 1549–1586.
36. Khalil, S.A.H.; Barakat, N.S.; Boraie, N.A. Effect of Aging on Dissolution Rates and Bioavailability of Riboflavin Sugar-Coated Tablets. *S.T.P. Pharm. Sci.* **1991**, *1*, 189–194.
37. Asker, A.F.; Abdel-Khalek, M.M.; Machloof, I. Effect of Scaling Up and Formulation Factors on the Qualities of Prednisone Tablets. *Drug Dev. Ind. Pharm.* **1981**, *7* (1), 79–111.
38. Singh, S.; Manikandan, R.; Singh, S. Stability Testing for Gelatin-Based Formulations: Rapidly Evaluating the Possibility of a Reduction in Dissolution Rates. *Pharm. Technol.* **2000**, *24* (5), 58–72.
39. Jones, B.E. *Hard Capsules—Development and Technology*; Ridgway, K., Ed.; The Pharmaceutical Press: London, U.K., 1987; 39–48.
40. Doelker, E.; Vial-Bernasconi, A.C. Shell-Content Interactions in Gelatin Capsules and a Critical Evaluation of Their Effects on Drug Availability. *S.T.P. Pharm.* **1988**, *4*, 298–306.
41. Gold, T.B.; Buice, R.G.; Lodder, J.R.A.; Digenis, G.A. Determination of Extent of Formaldehyde-Induced Crosslinking in Hard Gelatin Capsules by Near-Infrared Spectrophotometry. *Pharm. Res.* **1997**, *14* (8), 1046–1050.
42. Gold, T.B.; Buice, R.G.; Lodder, J.R.A.; Digenis, G.A. Detection of Formaldehyde-Induced Crosslinking in Soft Elastic Gelatin Capsules Using Near-Infrared Spectrophotometry. *Pharm. Dev. Technol.* **1998**, *3* (2), 209–214.
43. Venugopal, K.; Singh, S. Evaluation of Gelatins for Crosslinking Potential. *Pharm. Technol. Drug Delivery* **2001**, Supplement, 32–37.
44. Cade, D.; Madit, N.; Cole, E. Development of a Test Procedure to Consistently Cross-Link Hard Gelatin Capsules with Formaldehyde. *Pharm. Res.* **1994**, *11*, S-147.
45. Schwier, J.R.; Cooke, G.G.; Hartauer, K.J.; Yu, L. Rayon: Source of Furfural—A Reactive Aldehyde Capable of Insolubilizing Gelatin Capsules. *Pharm. Technol.* **1993**, *17*, 78–80.
46. Ofner, C.M., III.; Bubnis, W.A. Chemical and Swelling Evaluations of Amino Group Crosslinking in Gelatin and Modified Gelatin Matrices. *Pharm. Res.* **1996**, *13* (12), 1821–1827.
47. Hartauer, K.J.; Bucko, J.H.; Cooke, G.G.; Mayer, R.F.; Schwier, J.R.; Sullivan, G.R. Effect of Rayon Coiler on the Dissolution Stability of Hard-Shell Gelatin Capsules. *Pharm. Technol.* **1993**, *17*, 76–83.
48. Cooper, J.J.W.; Ansel, H.C.; Cadwallader, D.E. Liquid and Solid Interactions of Primary Certified Colorants with Pharmaceutical Gelatins. *J. Pharm. Sci.* **1973**, *62* (7), 1156–1164.
49. Marks, C.A.; Tourtellotte, D.; Andux, A. A Phenomenon of Gelatin Insolubility. *Food Technol.* **1968**, *22*, 1433–1436.
50. Guyot, M.; Fawaz, F.; Maury, M. In Vitro Release of Theophyllin from Cross-Linked Gelatin Capsules. *Int. J. Pharm.* **1996**, *144*, 209–216.
51. El-Fattah, S.A.; Khalil, S.A.H. Variations in Dissolution Rates of Sugar-Coated Chlorpromazine Tablets. *Int. J. Pharm.* **1984**, *18*, 225–234.
52. Ray-Johnson, M.L.; Jackson, I.M. Temperature-Related Incompatibility Between Gelatin and Calcium Carbonate in Sugar-Coated Tablets. *J. Pharm. Pharmacol.* **1976**, *28*, 309–310.
53. Digenis, G.A.; Sandefer, E.P.; Page, R.C.; Doll, W.J.; Gold, T.B.; Darwazeh, N.B. Bioequivalence Study of Stressed and Unstressed Hard Gelatin Capsules using Amoxycillin as a Drug Marker and Gamma Scintigraphy to Confirm Time and GI Location of *In Vivo* Capsules Rupture. *Pharm. Res.* **2000**, *17* (5), 572–582.
54. Vandelli, M.A.; Rivasi, F.; Guerra, P.; Forni, F.; Arletti, R. Gelatin Microspheres Crosslinked with D, L-Glyceraldehyde as a Potential Drug Delivery System: Preparation, Characterisation, In Vitro and In Vivo Studies. *Int. J. Pharm.* **2001**, *215*, 175–184.



55. Welz, M.M.; Ofner, C.M., III. Examination of Self-crosslinked Gelatin as a Hydrogel for Controlled Release. *J. Pharm. Sci.* **1992**, *81* (1), 85–90.
56. Bubnis, W.A.; Ofner, C.M., III. The Determination of ϵ -Amino Group in Soluble and Poorly Soluble Proteinaceous Materials by a Spectrophotometric Method Using Trinitrobenzenesulfonic Acid. *Anal. Biochem.* **1992**, *207*, 129–133.
57. Liu, W.G.; Yao, K.D.; Wang, G.C.; Li, H.X. Intrinsic Fluorescence Investigation on the Change in Conformation of Cross-Linked Gelatin Gel During Volume Phase Transition. *Polymer* **2000**, *41*, 7589–7592.
58. Taylor, S.K.; Davidson, F.; Ovenall, D.W. Carbon 13 -Nuclear Magnetic Resonance Studies on Gelatin Crosslinking by Formaldehyde. *Photogr. Sci. Eng.* **1978**, *22*, 134–138.
59. Gold, T.B.; Smith, S.L.; Digenis, G.A. Studies on the Influence of pH and Pancreatin on 13 C-Formaldehyde-Induced Gelatin Crosslinks Using Nuclear Magnetic Resonance. *Pharm. Dev. Technol.* **1996**, *1* (1), 21–26.
60. Albert, K.; Peters, B.; Bayer, E.; Treiber, U.; Zwilling, M. Crosslinking of Gelatin with Formaldehyde; A 13 C NMR Study. *Z. Naturforsch.* **1986**, *41b*, 351–358.
61. Albert, K.; Bayer, E.; Worsching, A.; Voge, H. Investigation of the Hardening Reaction of Gelatin with 13 C Labeled Formaldehyde by Solution and Solid State 13 C NMR Spectroscopy. *Z. Naturforsch.* **1991**, *46b*, 385–389.
62. Salsa, T.; Pina, M.E.; Teixeira-Dias, J.J.C. Crosslinking of Gelatin in the Reaction with Formaldehyde: An FTIR Spectroscopic Study. *Appl. Spectrosc.* **1996**, *50* (10), 1314–1318.
63. Buice, R.G.; Gold, T.B.; Lodder, R.A.; Digenis, G.A. Determination of Moisture in Intact Gelatin Capsules by Near-Infrared Spectrophotometry. *Pharm. Res.* **1995**, *12* (1), 161–163.
64. www.cis.rit.edu/people/faculty/hornak/jph-part-2.htm. (accessed 2001).
65. Ramarao, K.V.; Singh, S. Sensitivity of Gelatin Raw Materials to Cross-Linking: Influence of the Type of Gelatins and Bloom Strength. *Pharm. Technol.* **2002**, *26* (12), 42, 44, 46.
66. Dahl, T.C.; Sue, I.T.; Yum, A. The Influence of Disintegrant Level and Capsule Size on Dissolution of Hard Gelatin Capsules Stored in High Humidity Conditions. *Drug Dev. Ind. Pharm.* **1991**, *17* (7), 1001–1016.
67. Thoma, K. Photodecomposition and Stabilization of Compounds in Dosage Forms. *The Photostability of Drugs and Drug Formulations*; Taylor & Francis: London, 1996; 111–140.
68. Adesunloye, A.T.; Stach, P.E. Filled Gelatin Capsules. US 5,874,106, February 23, 1999.
69. Adesunloye, A.T.; Stach, P.E. Filled Gelatin Capsules Having a Reduced Degree of Cross-Linking. WO 9,733,568, September 18, 1997.
70. Cade, D.; Madit, N. Process for Stabilizing Gelatin Products, US 5, 620, 764, April 15, 1997.
71. Tatematsu, S.; Omata, K.; Hashimoto, T.; Takahashi, M.; Ito, N. Gelatin Composition, EP 0,35,982, October 11, 1989.

GELS AND JELLIES

Clyde M. Ofner III

University of the Sciences in Philadelphia, Philadelphia, Pennsylvania, U.S.A.

Cathy M. Klech-Gelotte

McNeil Consumer Healthcare, Fort Washington, Pennsylvania, U.S.A.

INTRODUCTION

The word “gel” is derived from “gelatin,” and both “gel” and “jelly” can be traced back to the Latin *gelu* for “frost” and *gelare*, meaning “freeze” or “congeal” (1). This origin indicates the essential idea of a liquid setting to a solid-like material that does not flow, but is elastic and retains some liquid characteristics. The distinction between gel and jelly remains somewhat arbitrary, with some differences based on the field of application. The food industry uses the term “gelatin jelly” whereas the pharmaceutical industry uses the term “gelatin gel.”

Use of the term “gel” as a classification originated during the late 1800s as chemists attempted to classify semisolid substances according to their phenomenological characteristics rather than their molecular compositions (2). At that time, analytical methods needed to determine chemical structures were lacking. The USP (3) defines gels (sometimes called jellies) as semisolid systems consisting of either suspensions made up of small inorganic particles, or large organic molecules interpenetrated by a liquid. Where the gel mass consists of a network of small discrete particles, the gel is classified as a two-phase system. Single-phase gels consist of organic macromolecules uniformly distributed throughout a liquid in such a manner that no apparent boundaries exist between the dispersed macromolecules and the liquid.

Single-phase gels and jellies can be described as three-dimensional networks formed by adding macromolecules such as proteins, polysaccharides, and synthetic macromolecules to appropriate liquids. In pharmaceutical applications, water and hydroalcoholic solutions are most common. Many polymer gels exhibit reversibility between the gel state and sol, which is the fluid phase containing the dispersed or dissolved macromolecule. However, formation of some polymer gels is irreversible because their chains are covalently bonded. The three-dimensional networks formed in two-phase gels and jellies are formed by several inorganic colloidal clays. Formation of these inorganic gels is reversible.

Gels are generally considered to be more rigid than jellies because gels contain more covalent crosslinks, a

higher density of physical bonds, or simply less liquid. Gel-forming polymers produce materials that span a range of rigidities, beginning with a sol and increasing in rigidity to a mucilage, jelly, gel, and hydrogel. Table 1 lists monographs for the 23 gel drug products, three jelly drug products, and three nondrug gels listed in the USP.

This review focuses mainly on water-based gels and jellies. Gel structure, the basis for understanding the physical properties associated with gels, is examined first, followed by the rheology of gels. Specific natural, semisynthetic, and synthetic gel-forming polymers and inorganic clays are then discussed along with their pharmaceutical applications.

GEL MICROSTRUCTURE

Substances that form aqueous gels are usually hydrophilic polymers capable of extensive solvation. At certain temperatures and polymer concentrations, and, in some cases, with the addition of ions, a three-dimensional network is formed. Although polymer gels vary considerably in chemical structure, they all behave as elastic solids at low applied stresses, even though they primarily consist of liquid. The differences in chemical composition, however, result in several types of gel microstructure. Pharmaceutical gels may be loosely categorized on the basis of their network microstructure according to the following scheme suggested by Flory (2):

1. Covalently bonded polymer networks with completely disordered structures
2. Physically bonded polymer networks, predominantly disordered but containing ordered loci
3. Well-ordered lamellar structures, including gel mesophases formed by inorganic clays

Covalently Bonded Structures

Covalently crosslinked gel networks are irreversible systems. They are typically prepared from synthetic

Table 1 Gel and jelly monographs in the USP 24

Monograph title	Drug product or nondrug product
Aluminum Hydroxide Gel	Drug product
Aluminum Phosphate Gel	Drug product
Aminobenzoic Acid Gel	Drug product
Benzocaine, Butamben and Tetracaine Hydrochloride Gel	Drug product
Benzocaine Gel	Drug product
Benzoyl Peroxide Gel	Drug product
Betamethasone Benzoate Gel	Drug product
Silica Gel	Nondrug product
Clindamycin Phosphate Gel	Drug product
Desoximetasone Gel	Drug product
Desamethasone Gel	Drug product
Dimethyl Sulfoxide Gel	Nondrug product
Diclonine Hydrochloride Gel	Drug product
Erythromycin and Benzoyl Peroxide Topical Gel	Drug product
Erythromycin Topical Gel	Drug product
Flucinonide Gel	Drug product
Hydrocortisone Gel	Drug product
Hydroxypropyl Cellulose Ocular System	Drug product
Lidocaine Hydrochloride Jelly	Drug product
Metronidazole Gel	Drug product
Naftifine Hydrochloride Gel	Drug product
Phenylephrine Hydrochloride Nasal Jelly	Drug product
Porous Silica Gel	Nondrug product
Pramoxine Hydrochloride Jelly	Drug product
Salicylic Acid Gel	Drug product
Fluoride and Phosphoric Acid Gel	Drug product
Stannous Fluoride Gel	Drug product
Tolnaftate Gel	Drug product
Tretinoin Gel	Drug product

hydrophilic polymers in one of two ways, details of which may be found in a comprehensive treatise (4–6). Because the resulting gel matrices are often highly rigid, these gels have been classified as hydrogels.

In the first method of preparation, infinite gel networks arise from the nonlinear copolymerization of two or more monomer species, with one being at least trifunctional. Both the direction and position by which each polymer chain grows during the reaction are random, resulting in the final microstructure of these gels being completely disordered. The gel point for copolymerization between equimolar concentrations of two monomer species can be predicted using the modified Carothers equation (7):

$$\bar{X}_n = \frac{2}{(2 - \rho f_{av})}$$

where \bar{X}_n is the number-average degree of polymerization, ρ is the fractional conversion, and f_{av} is the average functionality of the monomers involved. The gel point is

reached when $X_n \rightarrow \infty$, indicating that the critical conversion for gelation (ρ_c) is equal to $2/f_{av}$ (7). In practice, however, gel points tend to be overestimated by the equation.

The other method for preparing chemically crosslinked gel structures involves covalent crosslinking of individual linear or branched polymer chains, using a low concentration of crosslinking agent (5). The crosslinking ratio X , which is simply the mole ratio of crosslinking agent to polymer repeat units, can be used to characterize the resulting three-dimensional network.

Other parameters to characterize crosslinking are the equilibrium swelling ratio, the molecular weight between crosslinks, the so-called mesh size between crosslinks, and the crosslinking density (8). The equilibrium swelling ratio Q_m is the ratio of the volume of the swollen polymer to the volume of the dry polymer. It decreases as the extent of crosslinking increases. The molecular weight between crosslinks M_c , developed from the Flory equations for

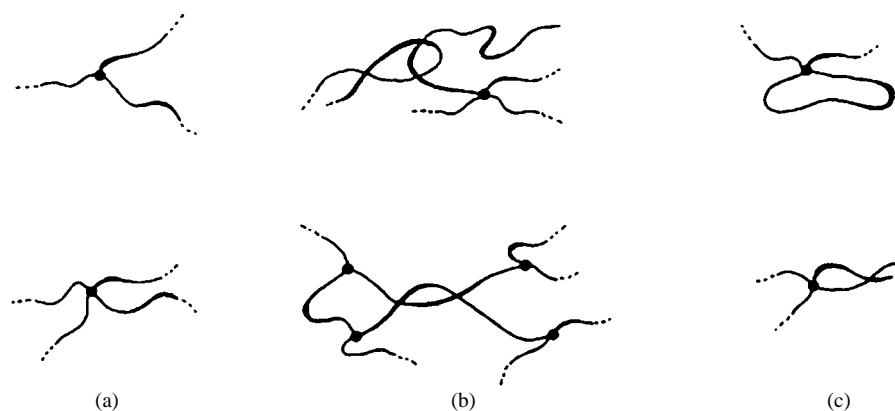


Fig. 1 Chemical and physical crosslinks associated with covalently bonded polymer gels. (a) Bi- and trifunctional chemical crosslinks; (b) simple and trapped physical entanglements; (c) ineffective chemically bonded loop and dangling ends.

equilibrium swelling (9), is also indirectly proportional to crosslinking extent. The mesh size ζ is an estimate of the distance between crosslinks based on the root-mean-square end-to-end distance of the random coil polymer between crosslinks. Estimates of the crosslinking density usually involve more rigorous treatments that require theoretical expressions (6). These models recognize the discrepancy between the amount of crosslinking agent added to the reaction and the number of effective chemical crosslinks actually produced, in addition to the presence of physical crosslinks formed by chain entanglements. If the crosslinking site is known, the covalent crosslinking density may be determined without theoretical models by chemical analysis of these sites (10). Examples of chemical and physical crosslinks, which may exist in a gel network of covalently linked linear chains, are illustrated in Fig. 1.

Physically Bonded Structures

Physically bonded gel networks are reversible systems; factors such as temperature and ion additives can induce a transition between the sol and gel phases. These gels are formed primarily by natural organic polymers (proteins and polysaccharides) and semisynthetic cellulose derivatives. Gels of some synthetic, hydrophilic polymers are also included in this class.

Polymer chains exist most often in the sol as random coils, which undergo conformational transitions to yield a gel (11). Such transitions may involve large ordered sections of one or more chains, which fold into a single, double, or triple helix. The three-dimensional network is then formed by cooperative association of several sections into higher ordered regions called junction zones (12). Many junction zones are dispersed throughout the

amorphous domains of the network, thus giving mechanical strength to the gel.

The microstructures of physically bonded gels are much more complex than those of the disordered, chemically crosslinked gels. The spatial arrangements assumed by polymer chains in forming junction zones may differ, as well as the secondary intermolecular forces that hold these zones together. The physical properties of gels, including rigidity, melting temperature, and yield point, are related to the type of junction zone formed. Several types have been identified or hypothesized; they are briefly reviewed here.

The particular organization of polymer chains in a junction zone depends on the chemical structure of the repeating unit. For example, sulfated polysaccharides (e.g., agar and K-carrageenan) that contain an assortment of sulfated galactose residues form double helices, two or more of which aggregate into multihelical junction zones (12). However, the presence of a few contaminant residue units greatly reduces gelling ability. The residues produce kinks that effectively block helix formation in large sections of chains, indicating that steric fit is critical to gel formation (12). The microstructure of these gels is schematically represented in Fig. 2.

Gels composed of semisynthetic cellulose derivatives, including sodium carboxymethylcellulose, contain microcrystalline junction zones. Residual crystallinity in the form of chain bundles can survive the derivatization processing of cellulose (13). The bundles are connected through common chains to yield a gel, and the ultimate gel strength depends on the efficiency by which the bundles were previously dispersed throughout the sol (13). Furthermore, microcrystalline domains, forming between segments of stereoregular synthetic polymers such as polyvinyl alcohol (14), enhance the gel's rigidity.

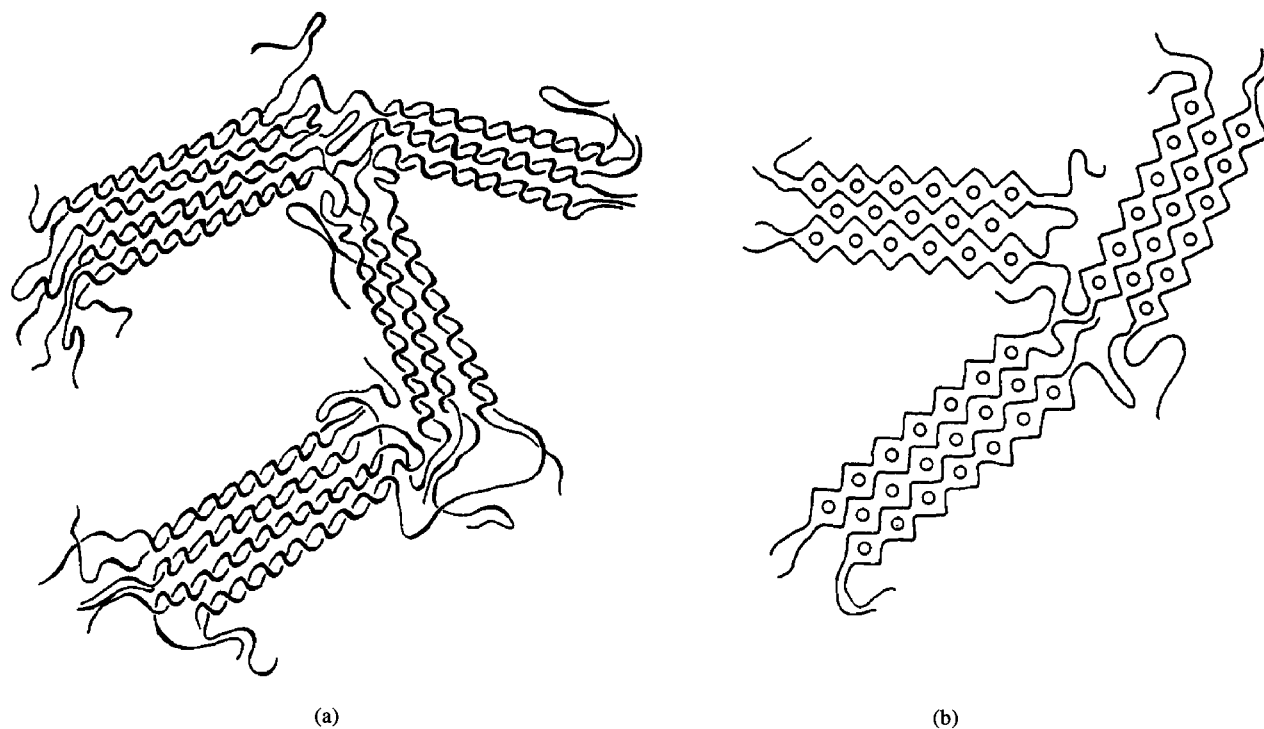


Fig. 2 Microstructures associated with physically bonded polymer gels. (a) Multihelical junction zones of agar gels; (b) egg-box model junction zones of calcium alginate gels.

Micelle-like junction zones are formed by methylcellulose and polyethyloxyene polypropyloxyene block copolymers (poloxamers). Although the polymers differ in chemical structure, both have hydrophobic regions in their chains: the di- and trimethyl-D-glucose residues of methylcellulose and the polypropyloxyene block of poloxamer. Another feature common to the two polymers is that their gels exhibit inverted thermal reversibility, that is, they gel with heating and melt with cooling. Both the inverted temperature behavior and the presence of hydrophobic regions in the polymers provide evidence for the formation of micelle-like junction zones (12, 15). Water molecules structured around the hydrophobic regions of polymer chains in a sol become disordered with increases in temperature. Newly exposed hydrophobic regions attract one another to form bonds, whereas hydrophilic areas rearrange to maximize their contact with the aqueous medium. The resulting micelle-like structures continue to grow in size and number at higher temperatures, leading eventually to gel formation.

Other junction zones require the presence of multivalent ions to form ion bridges between polymer chains. An egg box model was proposed by Grant et al. (16) for the formation of calcium alginate gels, in which calcium cations are cooperatively bound between ionized carboxyl

groups located on the polyglucuronate sequences of alginic acid. The cations are coordinated in the interstices of ordered segments of the polysaccharide chains (Fig. 2b). Other mechanisms for the involvement of ions in gelation may be similar but less precisely understood. For instance, guar gum forms cohesive gels only in the presence of borate ions (17).

Finally, some physically bonded gel networks are held together by simple entanglement couplings between individual polymer chains rather than by large aggregate junction zones. Above a critical concentration, long chains of hydrophilic polymers, such as hyaluronic acid and carbomer, are apparently forced through the domains of other chains because of their large molecular volumes (18). An increasingly intertwined network occurs with higher polymer concentrations to produce a gel. Although these gels are highly disordered compared with other physically bonded gels that have ordered junction zones, they are considered here because they form reversible gel microstructures. Yet, some researchers believe (12) that ordered regions must be present to impart rigidity in entangled gels, but such regions may be relatively dispersed and unstable, making them difficult to identify. Others propose (19) that higher order exists from interchain bridging by complexation with solvent molecules.

Molecular associations between polymer segments occur through the cooperation of several intermolecular forces such as hydrogen bonding, van der Waals forces, and electrostatic attractive and repulsive forces. The disruption of junction zones is associated with a high activation energy, further indicating that many intermolecular forces cooperate to retain the structure of each junction (12).

Well-Ordered Gel Structures

Under suitable conditions, certain silica, alumina, and clay aqueous dispersions form rigid gels or lyogels. When clays belonging to the smectite class, such as bentonite, aluminum magnesium silicate, hectorite, and laponite, come into contact with water, they undergo interlayer swelling spontaneously, followed by osmotic swelling to produce a gel (20). The plate-like clay particles associate into an ordered, extended network for which two models have been described (21). The “house of cards” model is based on attraction between weak positively charged particle edges and negatively charged particle faces; edge-to-edge associations of particles into flat ribbons have also been proposed.

Highly ordered lamellar gel microstructures are formed by certain surfactants and mixtures of a surfactant and long-chain fatty alcohols in water. Using small angle X-ray scattering (SAXS), an ordered lamellar stack lattice model was proposed for the gel formed by 10% w/w cetostearyl alcohol containing 0.5% cetrimide surfactant (22). In contrast, the microstructure of a Brij 96 gel depends on the surfactants concentration. A hexagonal liquid-crystalline gel structure was detected by SAXS at concentrations of 40 to 60 wt% in water, whereas an extended lamellar structure was detected at higher concentrations (70 to 85 wt%) (23).

PHYSICAL PROPERTIES OF GELS AND JELLIES

The physical properties of gels and jellies can be classified into two groups: transitional properties (including gel point, retrogradation, and syneresis) and rheological properties (including rigidity, yield point, and rupture strength). The experimental techniques used to characterize these physical properties can be similarly classified (24). Spectrophotometric and thermal techniques are used to identify gel microstructures (physical junction zones) and their related transitional properties. For example, nuclear magnetic resonance (NMR) spectroscopy

measures the structural and dynamic characteristics of the polymer just prior to aggregation and gel formation, and circular dichroism (CD) spectroscopy measures the conformational changes of the polymer during network formation (e.g., helix–coil transitions). Mechanical techniques are used to determine rheological properties of gels. These techniques employ either small-deformation measurements that yield viscoelastic parameters or large-deformation measurements that generate complete stress-strain profiles, which include failure parameters.

Transitional Properties

Sol–gel transition (gel point)

Sol–gel transition may be dependent on polymer concentration or temperature. Spectrophotometric methods, as mentioned previously, are used to probe sol–gel transitions that depend on the critical gelling concentration. Thermal methods, including differential scanning calorimetry, are used to measure sol–gel transitions, or melting temperatures, of thermoreversible gels. A relationship for estimating the heat of gelation was derived by Eldrige and Ferry (25) in which the dependence of the sol–gel transition temperature on polymer concentration was considered.

The critical gelling concentration is the concentration below which no macroscopic gel is formed under the prevailing experimental conditions (24); rather, a sol is formed by the polymer and solvent. This concentration depends on polymer–polymer and polymer–solvent interactions, the hydrophilic–lipophilic character of the polymer, and the molecular weight and flexibility of the chain (26). Furthermore, polymers that require ions to form gels have critical gelling concentrations that depend on the concentration of these additives and many other variables. Table 2 lists ranges of minimal concentrations for substances to gel in water.

There are two thermal gel points associated with thermoreversible gels. Shifts in temperature may cause gel formation at the setting point or gel liquification at the melting point. In addition, temperature hysteresis may occur in some gels in which the gel-setting point is lower than the melting point. The hysteresis behavior indicates that junction zones constitute a family of associations rather than a set of identical crosslinkages (12). Agarose gels show temperature hysteresis; the gel sets at about 40°C and melts at about 90°C.

Physical aging

Most gels have structures that have not attained equilibrium; different preparative methods and conditions

Table 2 Gelling concentrations for substances used in pharmaceutical products

Substance	Gel-forming concentrations (wt %)	Required additives
Proteins		
Collagen	0.2–0.4 ^a	
Gelatin	2–15	
Polysaccharides		
Agar	0.1–1	
Alginates	0.5–1	Ca ²⁺
	5–10	Na ⁺
K-Carrageenan	1–2	K ⁺
Gellum gum (low acetyl)	0.5–1	Ca ²⁺
Glycyrrhiza	2	
Guar gum	2.5–10	
	0.25	Borate ion
Hyaluronic acid	2	
Pectins (low methoxy)	0.8–2	Ca ²⁺
Starch	6	
Tragacanth gum	2–5	
Semisynthetic polymers (cellulose derivatives)		
Carboxymethylcellulose	4–6	Na ⁺
	10–25	Na ⁺
Hydroxypropylcellulose	8–10	
Hydroxypropylmethylcellulose	2–10	
Methylcellulose	2–4	
Synthetic polymers		
Carbomer	0.5–2	
Poloxamer	15–50	
Poyacrylamide	4	
Polyvinyl alcohol	10–20	
Inorganic substances		
Aluminum hydroxide	3–5	
Bentonite	5	
Laponite	2	Electrolytes
Surfactants		
Cetostearyl alcohol	10	Cetrimide
Brij 96 ^b	40–60	

^a Adjusted to pH > 4 and warmed to 37°C.^b Brij 30–99 surfactants are polyoxyethylene-alkyl ethers.

influence the gel state. The gel physically ages as it moves toward equilibrium, making the history of a gel sample an important consideration when measuring physical properties. Aging reflects changes in gel microstructure, where noncovalent crosslinks are breaking and reforming. Furthermore, instabilities caused by the nonequilibrium state arise in some polymer gels; two examples are retrogradation and syneresis.

Retrogradation is the spontaneous reversion of a polymer solution to a gel on standing. Polyvinyl alcohol dissolved in water undergoes retrogradation, whereby the stereoregular chains form microcrystalline aggregates as

the solution ages (14). Polyvinyl alcohol gels also retrograde, forming crystalline domains. Amylase, which is the linear polysaccharide fraction of starch, undergoes retrogradation, reducing the physical stability of starch solutions and gels over time.

Syneresis is the process whereby liquid is liberated spontaneously from the gel matrix. This instability arises from the nonequilibrium state of the gel established as it sets or because of a change in external conditions. At equilibrium, elastic contraction forces of polymer chains are usually balanced by solvent swelling forces, resulting from an osmotic pressure differential (27). With changes in

temperature, for example, the osmotic pressure shifts, causing an elastic contraction of polymer chains. The contractive response squeezes excess liquid out of the matrix. Agar and carrageenan are examples of gels that exhibit syneresis.

Rheological Properties

Like the transitional physical properties, the rheological properties of gels are not easily characterized because they depend strongly on the attributes of the polymer, history of the gel sample, and experimental conditions. Most often, the apparent viscosity or gel strength increases with an increase in the effective crosslink density of the gel or in the concentration and average molecular weight of the polymer. However, a rise in temperature may increase or decrease the apparent viscosity, depending on the molecular interactions between the polymer and solvent. In addition, the direction of change in apparent viscosity may not be readily predictable when additives such as ions, nonelectrolytes, solvents or nonsolvents, and other compatible polymers are mixed with a gel.

Viscoelasticity

Under an applied stress, which is the force per unit area, ideal liquids flow and perfectly elastic solids deform. The rate of shear, $d\gamma/dt$, is the measure of liquid deformation (flow): viscosity η is a liquid's resistance to flow, strain γ is the measure of solid deformation, and the shear modulus G is the resistance to strain for stresses applied tangentially to the solid's surface. Clearly, gels are semisolids that have both solid and liquid character under stress—they are viscoelastic substances.

Depending on their rheological properties, physically bonded gels can be divided into three groups: entanglement networks, strong gels, and weak gels (24). Entanglement networks behave as dilute solutions when diluted below their critical gelling concentrations, whereas strong gels have stress-strain profiles that include rupture points. Examples of polymers that form entanglement networks are guar gum and hyaluronic acid; strong gels are formed by agar, calcium alginate, gelatin, and pectin.

Weak gels are also entanglement networks, but they undergo specific molecular interactions that increase their strength (24). Therefore, the rheological properties of weak gels are intermediate between those of entanglement networks and strong gels. Xanthan gum and carbomer are examples of polymers that form weak gels. Hyaluronic acid generally forms an entanglement network, but under specific conditions (pH 2.5 and 0.15 *M* salt), it forms a weak gel (28).

Types of flow

The flow associated with entanglement networks and weak gels can be readily measured with continuous-shear instruments, such as the Ferranti-Shirley cone and-plate viscometer. These instruments characterize the gel behavior over a range of shear rates; a complete rheogram is usually generated for a particular gel in order to identify its flow behavior. If the shear stress versus shear rate curve is concave toward the shear-rate axis, the gel exhibits pseudoplastic flow. The internal gel microstructure breaks down with increases in shear rate, which lowers the apparent viscosity (it "shear-thins" the gel). However, if a gel does not flow at low stresses, but only above a finite stress, it exhibits plastic flow. The yield point, which is a measure of gel strength, is the stress at which flow begins. In general, entanglement networks and weak gels exhibit pseudoplastic and plastic flow, respectively.

Entanglement networks and weak gels may exhibit thixotropy, or time-dependent flow, in addition to exhibiting either pseudoplastic or plastic flow. Thixotropy, which is noted in a rheogram as a hysteresis loop, occurs because the gel requires a finite time to rebuild its original structure that breaks down during continuous shear measurements. The degree of thixotropy depends on gel type, sample history, and experimental conditions.

Dynamic oscillatory techniques use small-deformation measurements to determine the viscoelastic properties of gels within the linear region. Unlike continuous shear instruments, oscillatory instruments (e.g., the Weissenberg rheogoniometer) have the advantage of not altering the gel microstructure because small deformations are made with a sinusoidally oscillating stress or strain. For an ideal elastic solid, the stress and strain are in phase, whereas for an ideal liquid, the stress and strain have a phase difference of 90°. Viscoelastic materials, including gels, have phase angles between 0 and 90°. The parameters obtained through oscillatory testing are the storage modulus G' which reflects solid-like properties, and the loss modulus G'' , which reflects liquid-like properties. These moduli depend on the frequency of oscillation.

Details of both continuous shear and dynamic oscillatory testing of pharmaceutical semisolids, including gels, have been addressed in reviews (29, 30), where the fundamental viscoelastic principles are further developed along with a discussion of data analyses and interpretations.

Rigidity

The modulus of rigidity, or shear modulus (G), is defined as the ratio of shear stress to strain. It is a measure of a gel's ability to resist deformation. The minimum rigidity for a strong gel to resist deformation under its own weight is

equal to about $g\rho l$ which is the product of the acceleration due to gravity (g), density (ρ), and a linear dimension (l) of the sample. Therefore, the minimum rigidity is about 100 Pa (10^3 dyn/cm) for a gel sample 1 cm long.

An empirical measure of gel rigidity is the measurement of gelatin gel strength, also known as Bloom strength. The procedure is described in the USP (31). It is the determination in a Bloom gelometer of the gram weight needed to depress a standardized plunger 4 mm into the surface of a 6.67% (w/w) gel after maturing the gel at 10°C for 17 h. The Bloom gelometer, however, has largely been replaced by the Stevens LFRA/Voland Texture Analyzer and the TA.XT2 Texture Analyzer (Micro Systems).

Rupture strength

Rupture strength is equal to the stress at which a strong gel ruptures or fails rather than undergoing further strain. The rupture strength is determined by large-deformation measurements on instruments such as the Instron tester, where a tensile stress is applied to the sample. However, strong gel samples can only be tested in tension if they can support their own weight (24), and most physically crosslinked gels are relatively weak, making this type of test difficult.

GEL-FORMING SUBSTANCES AND THEIR PHARMACEUTICAL USES

Gel-forming hydrophilic polymers are typically used to prepare lipid-free semisolid dosage forms, including dental, dermatological, nasal, ophthalmic, rectal, and vaginal gels and jellies. Gel vehicles containing therapeutic agents are especially useful for application to mucous membranes and ulcerated or burned tissues because their high water content reduces irritancy. Furthermore, these hydrophilic gels are easily removed by gentle rinsing or natural flushing with body fluids, reducing the propensity for mechanical abrasion. The superior optical clarity of synthetic polymer gels, such as those composed of poloxamer and carbomer, has led to the current interest in developing therapeutic ophthalmic gels.

Unconventional routes of drug administration by using gels and jellies are also being explored. Thus, two nasal jellies were developed and marketed. The intranasal vitamin B-12 gel, Nascobal (Schwarz Pharma), is used as a dietary supplement. The gel base is composed of a hydrophilic cellulose derivative, the exact nature of which is not disclosed. However, the gel is apparently odorless and nonirritating, and adheres well to the mucous

membrane. Neo-Synephrine Viscous (Sanofi Winthrop) is a water-soluble nasal jelly formulated with methylcellulose; it contains the decongestant phenylephrine hydrochloride.

In addition to serving as drug-containing vehicles, some gels have other important functions. For example, a soft, flexible gel applied to burned skin can prevent excessive water loss by forming a physical barrier. Ocular gel inserts are designed to lubricate the eye continuously and promote healing. Still other gels are intended for lubricating surgical and medical instruments in order to minimize local irritation.

Many gel-forming substances are available for preparing pharmaceutical gels and jellies. Although these substances share some common physical characteristics, the intended use may require gelling attributes of a certain substance or blend of substances. Table 3 lists the favorable properties of pharmaceutical gels for particular applications.

Proteins

Collagen

Collagen is the major connective tissue protein in animals. The collagen molecule is considered to be a block copolymer, formed with blocks of glycine (33%) and proline and hydroxyproline (23%), between blocks of the remaining amino acids (44%). Tropocollagen is the asymmetrical subunit of collagen, which is made up of three peptide chains wound in a triple helix. The nonhelical portions of collagen, so-called telopeptides, are susceptible to enzyme digestion. Native collagen is insoluble in water, but 2% to 3% may be solubilized in dilute acidic solutions. Collagen may also be solubilized by treatment with proteolytic enzymes, which digest the telopeptide portions (e.g., Vitrogen, Collagen Corp.).

The asymmetrical molecular structure of collagen permits the formation of rigid gels; highly rigid structures are obtained at 0.1% concentration by UV irradiation or chemical crosslinking with aldehydes (32). These gels are used as biomaterials for vitreous replacements. Pepsin-solubilized collagen forms clear, rigid gels; they are prepared with a 0.2% to 0.4% collagen solution, which consists of unassociated molecules and small fibrils when cooled to $\sim 5^\circ\text{C}$. As the temperature is raised to 37°C , the collagen molecules aggregate, forming subfibrils that further aggregate to form the gel network (33). The rigidity is then increased by lightly crosslinking with aldehydes.

Optically clear collagen gels have been considered for use in ophthalmic drug delivery. Ocular inserts of succinylated or methylated collagen are patented as soluble

Table 3 Pharmaceutical gels

Applications	Favorable properties
Dental	Highly thixotropic Optimal viscosity for filling fissures Adherent to enamel surfaces Optically clear Water soluble Orally digestible
Dermatological	Thixotropic Good spreadability Greaseless (especially for acne preparations) Easily removed Emollient Demulcent (especially for abraded tissue) Nonstaining Compatible with a number of excipients Water soluble <i>or</i> miscible
Nasal	Adherent Odorless Nonirritating Water soluble
Ophthalmic	Optically clear Sterile Mucomimetic Lubricating Demulcent Nonirritating or nonsensitizing Water soluble or miscible
Surgical and medical procedures	Lubricating Adherent to instrument surfaces Maximal contact with mucous membranes Nonirritating
Vaginal	Acid stable Adherent Does not liquify at body temperature Slow dissolving Lubricating Greaseless and nontacky Nonirritating

devices for drug delivery (34). Constant, controlled release of pilocarpine hydrochloride was maintained for 5 to 15 days with crosslinked collagen and other collagen-derivative inserts (35). A corneal shield made of non-crosslinked collagen is marketed as Bio-Cor by Bausch & Lomb. This shield is placed over an eye injury and, as collagen slowly dissolves, the cornea is continuously lubricated to promote healing. Collagen implants (Lacrimedics) are also used for relief of dry eye by partially blocking tear removing canals; the implants dissolve within 7 to 10 days. Collagen gels are also used in hemostasis. A lightly crosslinked sterile pad, Instat hemostat (Johnson &

Johnson), assists in forming blood clots to arrest bleeding during surgery.

Gelatin

Gelatin is denatured collagen, which is hydrolytically degraded under acid or alkaline conditions to produce Type A or B gelatins, respectively. The amino acid content of acid-processed gelatin is virtually identical to that of collagen, yielding an isoelectric point, *pI*, between 8 and 9. In contrast, alkaline processing reduces the ratio of amide groups to carboxyl groups, thereby shifting the *pI* to about 4 or 5.

Gelatin forms elastic gels reversibly by cooling solutions that contain a sufficient concentration. The gel microstructure consists of a three-dimensional network held together by junction zones in which gelatin chains have partly refolded into the triple helix of the parent collagen molecule (36). The physical properties associated with gelatin gels depend on protein concentration, average molecular weight, temperature, pH, and additives (37).

The swelling kinetics of uncrosslinked type B gelatin films in the absence (38) and presence of additives (39) was studied. The water-swelling kinetics of chemically crosslinked matrices (40) were considered for controlling the release of therapeutic agents. Heat-crosslinked gelatin matrices were also examined for drug release (41). Both, the molecular weight between crosslinks, M_c , and the mesh size between crosslinks, ζ , were used to characterize gelatin gel matrices (42) during release of the macromolecular solute dextran. The crosslinking extent and density in gelatin matrices was directly determined by chemical analysis of the uncrosslinked primary amino groups and compared with swelling parameters of crosslinking (10).

Although gelatin has been used by the pharmaceutical industry for mainly producing soft and hard gelatin capsules, some commercial products use a hydrated gel form. The sterile product, H.P. Acthar Gel (Rorer), contains 16% gelatin for sustaining the release of adrenocorticotrophic hormone from an intramuscular or subcutaneous injection. A sterile, absorbable gelatin sponge, Gelfoam (UpJohn), which has a lightly crosslinked matrix, is used during surgery to absorb blood and promote clotting. An absorbable gelatin film, Gelfilm Ophthalmic (Upjohn), is available for use in ocular surgery. The ocular implant requires 2 to 5 months for complete absorption.

Polysaccharides

Alginates

Alginic acid is processed from brown seaweed, using a dilute acid, followed by alkalization with soda ash to yield the water-soluble salt form, sodium alginate. Alginic acid is a linear glycuronoalcan composed of polymannuronic acid blocks (M), polygluconic acid blocks (G), and mixed blocks of these two uronic acids (MG) (43). The gelling properties and gel microstructure of the various salt forms of alginic acid depend on the M, G, and MG block content. High-M alginates form turbid, weak gels, and high-G alginates form transparent, brittle gels. Gelation depends on the cation type; sodium alginate gels are water-soluble, whereas calcium alginate gels are insoluble, yet swell in water. Many "egg box" junction zones are formed

by calcium cations and G blocks to create such rigid gel networks.

Sodium and calcium alginate are used in commercial pharmaceutical formulations. Sodium alginate gels have superior spreading and lubricating properties; they are also nontacky and tasteless, and have emollient qualities. Moreover, sodium alginate is compatible with many compounds such as starch, sodium carboxymethylcellulose, pectin, carrageenan, 25% ethanol, 4% sodium chloride, and most alkali salts (44). Taking advantage of this compatibility and the favorable gel properties, sodium alginate is formulated with sodium carboxymethyl cellulose in the nonirritating, water-soluble lubricating jelly, Ortho Personal Lubricant (Ortho). Calcium alginate is used as a wound dressing for varicose and decubitus ulcers (45), where it forms a hydrophilic gel over the wound by absorbing exudate, which promotes healing. The high rigidities of calcium alginate gels make them suitable for preparing dental impressions and as matrix barriers for controlling drug delivery (46). Theophylline release was studied from granules coated with sodium alginate and calcium lactate through insoluble gel formation (47).

Carrageenan

Carrageenan, a sulfated polysaccharide extracted from red seaweed, may be separated into a number of fractions depending on the species, season, and environment. The three major fractions, lambda-, iota-, and kappa (κ)-carrageenan, contain an alternating sequence of (1 \rightarrow 4)-linked (β -D-galactopyranosyl and (1 \rightarrow 3)-linked α -D-galactopyranosyl residues, but they differ in the degree and sites of sulfation (48). Of these three fractions, only lambda-carrageenan cannot form gels.

Under optimal ionic strength and polymer concentration, κ -carrageenan forms rigid, thermoreversible gels in the presence of cations. The elastic moduli of these gels depend on the type of cation and follow the Hofmeister series: $\text{Cs}^+ > \text{Rb}^+ > \text{K}^+ \gg \text{Na}^+ > \text{Li}^+$. The κ -carrageenan gel also undergoes syneresis, the spontaneous liberation of liquid from the matrix. Early models for gelation envisioned multihelical junction zones composed of polysaccharide chains folded into double helices. This model was extended to accommodate cations in that the formation of the helical domains would require mediation by the cations themselves (49). However, other researchers propose (50) an indirect role for cations in which they alter solvent structure around the κ -carrageenan helices, thereby inducing gel formation. Differential scanning calorimetry and isothermal titration calorimetry were used to study mono- and divalent cation binding and induced conformation changes in κ -carrageenan (51).

Gels made with κ -carrageenan and potassium ions have excellent lubricity and emollient properties. Because of such favorable gelling characteristics, these gels are used as vehicles for the topical administration of drugs and as gelling agents in other pharmaceutical preparations. They can also be used in combination with sodium carboxymethylcellulose, which interferes with gel formation resulting in a variety of consistencies and textures. Theophylline and diclofenac sodium release were studied from tablet matrices containing κ -carrageenan and iota-carrageenan with KCl or CaCl_2 (52).

Hyaluronic acid

Hyaluronic acid is a linear glycosaminoglycan (an aminopolysaccharide), which is an important component of synovial fluid and the cellular matrix of connective tissue. It comprises a biological gel that supports cells, maintains tissue hydration, and functions as a lubricant and shock absorbent in joints (53). Hyaluronic acid forms transparent, rigid gels at 2% concentration, although they are not as rigid as agar and carrageenan gels. The gel microstructure is constructed by entanglement couplings among the long polysaccharide chains, which have molecular weights ranging from about 1×10^6 to 4×10^6 Da, depending on the source of animal tissue.

A 1% viscoelastic jelly of the sodium salt is commercially available as Healon (Kabi Pharmacia), which is widely employed in ophthalmic surgery. Sodium hyaluronate jelly provides a nonirritating viscoelastic medium for separating tissues during surgery and prevents postoperative adhesion formation (53). Investigators in a clinical trial of intraarticular injections of 500–730 kDa sodium hyaluronate (Hyalgan, Sanofi) for osteoarthritis concluded that this treatment might delay the structural progression of the disease (54). When hyaluronic acid was tested as an ophthalmic gel (55), its mucoadhesive properties improved the ocular bioavailability of tropicamide.

Pectins

Pectins are a complex group of polysaccharides, the structures of which vary with plant source. They are heteropolysaccharides that are extracted from apple pomace and citrus fruit rinds and consist mainly of polygalacturonic acids. The degree of esterification determines the gelation process and, hence, the eventual commercial use. High-methoxy (HM) pectin gels in the presence of high concentrations of sucrose at acidic pH. The sucrose dehydrates pectin to a gel. High-methoxy pectins are used in sweet food products such as fruit jams and preserves.

In contrast, low-methoxy (LM) pectins gel in the presence of divalent cations, especially calcium, by the “egg box” mechanism proposed for alginates. Moreover, calcium pectinate gels prepared at neutral pH are heat stable, whereas acidic pH gels are thermoreversible (56). Gel strength depends on the extent of esterification (levels from 30% to 50% are optimal), the distribution of ester groups on the chain, and the average molecular weight. LM pectins have been used traditionally in antidiarrheal formulations with kaolin. HM pectins were evaluated in controlled release matrix formulations (57). Pectin microspheres were reported to improve ophthalmic bioavailability of piroxicam in rabbits compared with commercial piroxicam eye drops (58).

Starch

Starch is the principal polysaccharide of higher plants and can be extracted from a number of sources, including corn, wheat, and potato. Starch granules can be fractionated into two structurally distinct polysaccharides: 30% amylose (linear polymer) and 80% amylopectin (branched polymer). Starch is insoluble in cold water because of extensive hydrogen bonding between polysaccharide chains. However, the hydrogen bonds break at temperatures above the gelatinization range of 60°C to 70°C, allowing the granules to swell. The starch sol can then form a gel on cooling. The type of gel depends on the starch species; corn starch forms opaque, rigid gels, whereas potato starch forms clear, nonrigid gels (59).

The microstructure of starch gels is an example of a matrix strengthened by filler. Heating causes the amylopectin granules to become swollen and porous, whereas the linear amylose chains dissolve. As the solution cools, amylose sets into a gel matrix that threads through the porous amylopectin granules, thereby producing a reinforced gel (60).

Aqueous starch solutions exhibit retrogradation in which crystalline aggregates form spontaneously over time with Brownian movement. As the solution ages, it becomes opalescent, followed by the gradual precipitation of starch. Starch gels are also unstable because of the formation of crystalline aggregates; however, they undergo syneresis. The amylose fraction is responsible for these instabilities; amylose chains are linear, enabling them to become parallel in order to form aggregates. Starch has been used extensively as a pharmaceutical excipient in tablets, serving as a filler, binder, and disintegrant. Starch gels have found limited use as skin emollients. Starch has been examined as a mucoadhesive in microparticles to improve protein uptake across the nasal mucosa (61).

Table 4 Miscellaneous polysaccharide gelling agents and their uses

Agent	Notes	Product or experimental use	Ref.
Agar	Soluble at $> 90^{\circ}\text{C}$	Molecular structure	(62)
	Rigid gel from $\sim 0.1\%$	Bulk laxative	(63)
	Multihelical junction zones	Rectal administration of insulin	(64)
		Suppository base with NaCMC	(65)
Gellan gum	Good optical clarity	Timoptic-XE (Gelrite; Merck)	
	Forms gels with cations	In situ gelling polymer for ophthalmic delivery	(66)
Glycyrrhiza	Gels are pseudoplastic and thixotropic	Enhanced colon calcitonin absorption	(67)
	Efficient bitter taste masker	Vehicle for antiviral idoxuridine	(68)
Guar gum	Liquifaction at $< \text{pH } 7$		(69)
	Nonionic	Sustained release of quinidine sulfate	(70)
	Binds free water to eliminate syneresis	Colon specific drug carrier	(71)
Tragacanth gum	Acid stable ($\text{pH} \sim 2$)		(72)
	Restore vaginal acidity	Aci-Jel (Ortho)	
	Gel vehicle	Ephedrine sulfate jelly, USP	

Miscellaneous carbohydrates

Table 4 lists less frequently used agents and a few notes about their properties. Also listed are products or experimental uses and references.

Semisynthetic Polymers (Cellulose Derivatives)

Carboxymethylcellulose, sodium

Sodium carboxymethylcellulose (NaCMC) is a carboxymethyl ether of cellulose, the ubiquitous polysaccharide composing the fibrous tissue of plants. The hydroxyl groups on the 2-glucopyranose residues of cellulose are replaced by carboxymethyl groups; the number of replacements is known as the degree of substitution DS. Both the DS and polymer chain length determine the solubility, viscosity, and gel strength of NaCMC. It dissolves in water and mixtures of water with the lower alcohols and glycerin. Aqueous gels are stable from pH 2 to 10 but susceptible to microbial growth.

The microstructure of NaCMC gels was examined on freeze-fractured samples with a transmission electron microscope (73). A fine, quasi-crystalline structure consisting of filaments with thicknesses of 2 to 3 nm was identified. These filaments are microaggregates of individual polymer chains. The rigidity of NaCMC gels can be increased by adding multivalent cations such as Al^{3+} or Fe^{2+} , resulting in ionic bridges between the cations and ionized carboxyl groups (74). These gels are stable indefinitely.

Because of its acid stability, NaCMC is suitable for therapeutic vaginal gels. The base of Ortho-Gynol contraceptive jelly (Ortho) is formulated with NaCMC, water, and propylene glycol and adjusted to pH 4.5 with acetic acid. NaCMC also composes the gel base of Glucose (Paddock), which contains 40% dextrose and is used to treat hypoglycemia. In a study of NaCMC gels as the vehicle for a recombinant human platelet-derived growth factor for treatment of non healing diabetic ulcers, the gel vehicle alone was reported to have a higher healing rate compared to good wound care alone (75).

Hydroxypropyl cellulose

Hydroxypropyl cellulose is produced by substituting propylene oxide for hydroxyl groups on alkalinized cellulose. This cellulosic derivative is soluble in cold water and many polar organic solvents such as methanol, ethanol, isopropanol, and propylene glycol (76). The alcohol solubility increases with increasing degrees of hydroxypropyl substitution. Aqueous preparations of hydroxypropyl cellulose are susceptible to hydrolysis (acid unstable); therefore, a pH of 6 to 8 provides optimum stability. Photodegradation and limited biological degradation may also occur.

Hydroxypropyl cellulose gels on heating and forms thermoreversible gels. The pharmaceutical gel products containing hydroxypropyl cellulose take advantage of its compatibility with alcohol. The microviscosity of hydroxypropyl cellulose gels was evaluated for prediction of drug diffusion rates (77). Examples of commercial products with alcoholic and hydroalcoholic gel bases are given in Table 5.

Table 5 Pharmaceutical gel products containing hydroxypropyl cellulose or hydroxypropyl methylcellulose

Brand name	Manufacturer	Therapeutic agent; activity	Alcohol content (%)
Hydroxypropyl cellulose			
Compound W Gel	Whitehall	Salicylic acid; keratolytic	67.5
DuoPlant Gel	Schering-Plough	Salicylic acid; keratolytic	57.6
Erygel	Herbert	Erythromycin; antibiotic for acne	92
Hydrisalic Gel	Pedinol	Salicylic acid; keratolytic	100
Keralyt Gel	Summers	Salicylic acid; keratolytic	19.4
Lacrisert Insert	Merck & Co.	Hydroxypropyl cellulose; lubricant	0
Retin-A Gel	Ortho	Retinoic acid; antiacne	90
Hydroxypropyl methylcellulose			
ArthriCare Triple-Medicated Gel	Commerce	Methyl salicylate; menthol, methyl nicotinate; liniment	
Persa-Gel W	Ortho Derm	Benzoyl peroxide; antibacterial	
Xylocaine Jelly	Astra	Lidocaine HCl; anesthetic	

Hydroxypropyl methylcellulose

Hydroxypropyl methylcellulose (HPMC) is produced similarly to hydroxypropyl cellulose except that methyl chloride is included in the reaction. The composition of the substituted hydroxyl groups ranges from 3% to 12% hydroxypropyl and from 19% to 30% methyl. HPMC is soluble in cold water and the polyethylene glycols up to 600 Da, but, in contrast to hydroxypropyl cellulose, HPMC is not soluble in alcohol. It is also less susceptible to hydrolysis, and is stable from pH 3 to 11.

Aqueous gels are formed on heating; the gel point ranges from 50°C to 90°C, depending on the grade (78). Addition of small amounts of water-miscible solvent, such as ethanol and the glycols, raises the gel point. Examples of HPMC-based pharmaceutical gels that contain only water are included in Table 3. The good lubricating and adherent qualities of HPMC gels are exploited in Xylocaine 2% Jelly (Astra), which is used to minimize discomfort associated with medical procedures. A 2% solution (80 kDa) is commercially available as an ophthalmic surgical aid (OcuCoat; Storz). The influence of indomethacin, propranolol HCl, and tetracycline HCl was studied on the properties and swelling characteristics of matrix gels containing hydroxypropyl methylcellulose and methylcellulose (79).

Methylcellulose

Methylcellulose is a cellulose ether in which methyl groups have been substituted for hydroxyl groups on the 2-glucopyranose residues. It is soluble in cold water at low methoxy contents; increased substitution increases the solubility in hydroalcoholic and alcoholic solutions (80).

Aqueous solutions of methylcellulose gel on heating, whereby micelle-like junction zones form throughout the network. Gel strength and gelling temperature depend on the concentration, degrees of substitution, and average molecular weight. The gelling temperature can be lowered by adding sugar or most electrolytes, which reduce polymer hydration.

High viscosity grades of methylcellulose are used in pharmaceutical gels. The gels are demulcent and have good surfactant properties, which permit easy spreading on body tissues. Therefore, methylcellulose gels are used as dressings for burned tissue because they minimize water loss and are easily removed. The high viscosity grades are used in ophthalmic preparations such as Murocel artificial tears (Bausch & Lomb) and Neo-Synephrine viscous solution (Sanofi Winthrop). Other pharmaceutical applications include the bulk-forming laxative, Citrucel (SK-Beecham), and lubricating jellies for surgical and medical procedures. A methylcellulose gel was investigated for topical administration of tetracycline HCl (81). The bioadhesive properties of 3% methylcellulose gels on slowing mucociliary clearance were examined using a rate model (82).

Synthetic Polymers

Carbomer

Carbomer is a synthetic polyacrylic acid resin, which is copolymerized with about 0.75 to 2% polyalkylsucrose (83). This is the reason why aqueous dispersions of carbomer must be protected against microbial growth. Carbomer is a high molecular weight polymer that contains carboxylic acid groups on about two thirds of its

repeat units. Gels are formed on neutralization between pH 5 and 10 with metal hydroxides or amines such as diisopropanolamine and triethanolamine. Neutralization expands the long chains of carbomer by charge repulsion to produce an entangled gel network. Because electrostatic repulsion plays a critical role in forming a gel, viscosity and gel strength depend on both pH and salt content.

The molecular size of carbomer is important in determining gel properties and applications. Carbomer 934 and 940, with average molecular weights of 3×10^6 and 4×10^6 Da, respectively, are most commonly used by the pharmaceutical industry. Both grades have favorable rheological properties for topical applications; the gels undergo plastic flow (84) and have temperature-stable viscosities (85). The gels can be formulated with large quantities of alcohol, but the alcohol dehydrates the polymer network and lowers the viscosity. Ostrenga et al. conducted a systematic investigation of propylene glycol/water vehicles gelled with 1% carbomer to optimize percutaneous absorption of fluocinolone acetonide and fluocinonide (86). These investigators showed under specified conditions that an increasing partition coefficient of the steroid between the vehicle and isopropyl myristate increased the uptake of drug into the skin. These studies laid the ground for commercial use of such gels for a range of steroids.

Carbomer is also compatible with dimethyl sulfoxide. The veterinary product, Domoso (Syntex Veterinary Labs), is a topical gel consisting of carbomer 934 and 90% dimethyl sulfoxide. Carbomer 940 gels exhibit superior optical clarity compared with 934 gels (87), and can be used in ophthalmic preparations. However, only carbomer 934 is approved for internal use. The mucoadhesive properties of carbomer were studied with triamcinolone acetonide (88), with acyclovir (89, 90) and to improve protein uptake across the nasal mucosa (61). Pharmaceutical gel products containing carbomer in their formulations are listed in Table 6. All are dermatological gels except for Anbesol Gel (Whitehall) and Pilopine HS (Alcon).

Poloxamer

Poloxamer is the generic name for a series of block copolymers that are composed of one polypropylene oxide block sandwiched between polyethylene oxide blocks. For example, poloxamer 188 can be written as (PEO)₇₅-(PPO)₃₀-(PEO)₇₅. The poloxamers serve as high molecular weight surfactants because the PEO blocks are hydrophilic, whereas the PPO blocks are hydrophobic.

At relatively high concentrations (>20%), poloxamers form thermoreversible gels; however, they gel on heating rather than cooling. The amphiphilic nature supports the

Table 6 Pharmaceutical gel products containing carbomer

Brand name	Manufacturer	Therapeutic agent; activity
Carbomer 934		
Anbesol Gel	Whitehall	Benzocaine, anesthetic; antiseptic
Cleocin T Gel	Upjohn	Clindamycin; antibiotic for acne
Persa-Gel W	Ortho	Benzoyl peroxide; antibacterial
Therapeutic Mineral Ice Gel	Bristol-Meyers	Menthol; liniment
Ben-Gay Vanishing Scent Gel	Pfizer	Menthol, camphor; liniment
Sportsman Ice Gel	Thompson	Menthol; liniment
Carbomer 940		
Benzac W	Galderma	Benzoyl peroxide; antibacterial
Benzagel	Dermik	Benzoyl peroxide; antibacterial
Benzamycin	Dermik	Erythromycin, benzoyl peroxide; antibiotic, antibacterial
Desquam-X 5	Westwood Squibb	Benzoyl peroxide; antibacterial
Double Ice ArthriCare Gel	Commerce	Menthol, camphor, liniment
Estar Gel	Westwood Squibb	Coal tar; antipsoriasis
Flex-all 454 Gel	Chattam	Menthol, methyl salicylate; liniment
Lidex Gel	Syntex	Fluocinonide; anti-inflammatory
Ordor Free ArthriCare Rub	Commerce	Menthol, methyl nicotinate, capsaicin; liniment
Persa-Gel	Ortho Derm	Benzoyl peroxide; antibacterial
Pilopine HS	Alcon	Pilocarpine HCl; miotic
PrameGel	GenDerm	Pramoxine HCl; antipruritic
Topicort Gel	Hoecht Marion Roussel	Desoximetasone; anti-inflammatory

gelling mechanism of poloxamers, where micelle-like junction zones form at or above room temperature. The junction zones consist of large populations of micelle-like structures, which apparently form a viscous, liquid crystalline phase (91, 92). Poloxamers can also form gels in dilute hydroalcoholic solutions.

Poloxamer gels have many characteristics favorable for use as artificial skin, which is helpful in treating third-degree burns. The gels are nontoxic, enhance healing by controlling water, heat, and electrolyte loss, and provide detergent activity on wound debris (93). Because of poloxamer's inverted thermoreversibility, cool solutions can be poured onto damaged tissue, forming gels when warmed to body temperature. The gels are easily removed by rinsing with cool water.

Poloxamer gels mimic mucus and are optically clear, which makes them suitable for ophthalmic drug delivery. A poloxamer gel formulation containing pilocarpine showed improved bioavailability over an aqueous solution of the drug (94). Release kinetics of lidocaine (91), diclofenac (95), and hydrocortisone (95) from topical poloxamer gels were also assessed. Subcutaneous injection of insulin loaded poloxamer gels prolonged the hypoglycemic effect of insulin in rats (95). Commercial topical gels include AquaTar (Allergan Herbert), a poloxamer-407 base that contains coal tar, and Benzac W (Galderma), a blend of poloxamer-182 and carbomer-940 gels, that contains benzoyl peroxide.

Polyacrylamide

Polyacrylamide (PAAm) is a hydrophilic polymer that absorbs and retains large volumes of water. However, aqueous solutions of PAAm, especially the high molecular weight species, undergo physical aging, which results in a decrease in viscosity. PAAm is soluble in hydrophilic nonaqueous liquids such as glycerol, but insoluble in methanol and ethanol.

Polyacrylamide forms water-based gels at concentrations around 4% w/v, which exhibit pseudoplastic behavior. A PAAm ophthalmic gel containing pilocarpine was compared with other gel vehicles; the ocular bioavailability for the PAAm gel was three times greater than that of the aqueous control solution (97). The kinetics of ibuprofen release for crosslinked PAAm gels was studied (98). A kinetic model was proposed for swelling induced loading of insulin into crosslinked PAAm gels (99).

Polyvinyl alcohol

Polyvinyl alcohol (PVA) is a hydrophilic, synthetic polymer that is prepared indirectly by hydrolyzing polyvinyl acetate. It cannot be produced by direct polymerization of vinyl alcohol because the monomer is

unstable (100). The chemical structure of PVA is simple, consisting of a carbon-chain backbone with alternating hydroxyl groups; it favors the formation of crystalline aggregates in solutions, gels, and solids. PVA is soluble in water, glycerin, and mixtures of water with the lower alcohols; however, it can be precipitated from aqueous solutions with sulfates and phosphates. The gelation of PVA occurs in more concentrated solutions; the gel point of 10% PVA in water is around 14°C (101). In general, gel strength depends on the degree of crystallinity of a particular sample.

Polyvinyl alcohol is used in ophthalmic preparations, serving as a mucus mimicking (mucomimetic) agent in artificial-tear formulations such as Liquifilm Tears and Liquifilm Forte (Allergan). Physically crosslinked PVA hydrogels, prepared by low-temperature crystallization, were tested as vehicles for the rectal administration of indomethacin (102). Chemically crosslinked PVA hydrogels have been considered for soft contact lenses. Both, the molecular weight between crosslinks, M_c , and the mesh size between crosslinks, ζ , were used to characterize a covalently crosslinked polyvinyl alcohol gel (8) during evaluation of bovine serum albumin (BSA) release. PVA hydrogel nanoparticles loaded with albumin into polylactic/glycolic acid microspheres released albumin up to two months (103). The release of water-soluble pseudoephedrine was studied from compressed swellable-soluble PVA matrices (104).

Inorganic Substances

Aluminum hydroxide

Aluminum hydroxide forms a two-phase gel consisting of a network of discrete solid particles in water. Aluminum hydroxide gels are soluble in acidic and extremely basic media, and are compatible with many additives, including glycerin, saccharin, and some preservatives (105).

The aluminum hydroxide gel exhibits thixotropic behavior. In addition, gel stability is enhanced by polyols such as mannitol and sorbitol (106). However, unlike other gels, aluminum hydroxide gels do not have demulcent properties; they are used mainly as an oral antacid preparations.

Smectite clays

Bentonite and hectorite clays consist primarily of hydrated aluminum and magnesium silicates, respectively. Bentonite is recognized for its swelling capacity; one gram can absorb up to 11 ml of water. A commonly used smectite clay is aluminum magnesium silicate (Veegum, R.T. Vanderbilt Co.). These clays have

plate-like particles, 1 to 2 μm in size, that form well-ordered gel mesophases spontaneously on contact with water. The gels are formed at about 5% concentration, and are presumably stabilized by opposite electrostatic charges between particle edges and faces (21). Therefore, these smectite gels are incompatible with strong electrolytes, and in particular with di- and trivalent cations. The gels also exhibit plastic flow; they have static yield values and are thixotropic.

Laponite clay also belongs to the smectite class, but is a synthetic gelling agent. Like bentonite, laponite swells considerably in water—but only a 2% concentration is needed to form a gel. Laponite does not contain impurities, which is an advantage over the natural clays; however, some electrolyte must be included in the water to support the gel microstructure (107).

Smectite clays are used by the pharmaceutical industry as mainly suspending and thickening agents. Because of the net negative charge of these clays, cationic drugs can be adsorbed by an ion-exchange mechanism. Release of metronidazole from bentonite complexes was inhibited at acidic pH, but increased significantly at pH 7 (108). Simple gels containing hectorite and gelatin were investigated for rectally administering insulin to rabbits (109). Finally, the iontophoretic transdermal transport of calcium from a paste of calcium-enriched bentonite was studied using excised pig skin (110).

REFERENCES

1. Webster's Seventh New Collegiate Dictionary; Merriam: Springfield, MA, 1971.
2. Flory, P.J. Discuss. Faraday Soc. **1974**, 57, 7.
3. The United States Pharmacopeia 24/The National Formulary 19; Pharmacopeial Convention: Rockville, MD, 2000; 2111–2112.
4. Peppas, N.A. *Hydrogels in Medicine and Pharmacy*; CRC Press: Boca Raton, FL, 1987; I.
5. Peppas, N.A. *Hydrogels in Medicine and Pharmacy*; CRC Press: Boca Raton, FL, 1987; II.
6. Peppas, N.A. *Hydrogels in Medicine and Pharmacy*; Peppas, N.A., Ed.; CRC Press: Boca Raton, FL, 1987; III.
7. Alger, M.S.M. *Polymer Science Dictionary*; Elsevier Applied Science: London, 1989; 69.
8. Gander, B.; Gurny, R.; Doelker, E.; Peppas, N.A. Pharm. Res **1989**, 6, 578–584.
9. Bray, J.C.; Merrill, E.W. J. Appl. Polym. Sci. **1973**, 17, 3779–3794.
10. Ofner, C.M., III; Bubnis, B.A. Pharm. Res. **1992**, 13, 1821–1827.
11. Bryce, T.A.; McKinnon, A.A.; Morris, E.R.; Rees, D.A.; Thom, D. Discuss. Faraday Soc. **1974**, 67, 331.
12. Rees, D.A. Adv. Carbohydr. Chem. Biochem. **1969**, 24, 267.
13. deButts, E.H.; Hudy, J.A.; Elliot, J.H. Ind. Eng. Chem. **1957**, 49, 94.
14. Nagy, M. Colloid Polym. Sci. **1985**, 263, 245.
15. Vadnere, M.; Amidon, G.; Lindenbaum, S.; Haslam, J.L. Int. J. Pharm. **1984**, 32, 207.
16. Grant, G.T.; Morris, E.R.; Rees, D.A.; Smith, P.J.C.; Thom, D. FE.B.S. Lett. **1973**, 32, 1.
17. Noble, O.; Turquois, T.; Tavel, F.R. Carbohydr. Polym. **1990**, 12, 203.
18. Laurent, T.C. *Chemistry and Molecular Biology of the Intercellular Matrix*; Balazs, E.A., Ed.; Academic Press: New York, 1970; 2, 703.
19. Tanaka, F. Macromolecules **1990**, 23, 3790.
20. van Olphen, H. *An Introduction to Clay Colloid Chemistry*; Interface Publishers: New York, 1963; 67–68.
21. Schott, H. J. Colloid Interface Sci. **1968**, 26, 133–139.
22. Barry, M.D.; Rowe, R.C. Int. J. Pharm **1989**, 53, 139.
23. Tiemessen, H.L.G.M.; Bodde', H.E.; van Mourik, C.; Junginger, H.E. Progr. Colloid Polym. Sci. **1988**, 77, 131.
24. Clark, A.H.; Ross-Murphy, S.B. Adv. Polym. Sci. **1987**, 83, 57.
25. Eldridge, J.E.; Ferry, J.D. J. Phys. Chem. **1954**, 58, 975.
26. Florence, A.T.; Atwood, D. *Physicochemical Principles of Pharmacy*; Chapman and Hall: New York, 282.
27. Florence, A.T.; Atwood, D. *Physicochemical Principles of Pharmacy*; Chapman and Hall: New York, 285.
28. Gibbs, D.A.; Merrill, E.W.; Smith, K.A.; Balazs, E.A. Biopolymers. **1968**, 6, 777.
29. Barry, B.W. *Dermatological Formulations: Percutaneous Absorption*; Marcel Dekker, Inc.: New York, 1983.
30. Schott, H. Rheology. *Remington: The Science and Practice of Pharmacy*; Gennaro, A.R., Ed.; Mack: Easton, PA, 1995.
31. The United States Pharmacopeia 24/The National Formulary 19; Pharmacopeial Convention: Rockville, MD, 2000; 2049.
32. Stenzel, K.H.; Miyata, T.; Rubin, A.L. Annu. Rev. Biophys. Bioeng. **1974**, 3, 255.
33. Wallace, D.G.; Condell, R.A.; Donovan, J.W.; Paivinen, A.; Rhee, W.M.; Wade, S.B. Biopolymers. **1986**, 25, 1875.
34. Miyata, T.; Rubin, A.L.; Stenzel, K.H.; Dunn, M.W. U.S. Patent. 4,164,559, 1979.
35. Vasantha, R.; Sehgal, P.K.; Rao, K.P. Int. J. Pharm. **1988**, 47, 95.
36. Veis, A. *Macromolecular Chemistry of Gelatin*; Academic Press: New York, 1964.
37. Ward, A.G.; Courts, A. *The Science and Technology of Gelatin*; Academic Press: New York, 1977.
38. Ofner, C.M., III; Schott, H. J. Pharm. Sci. **1986**, 75, 790–796.
39. Ofner, C.M., III; Schott, H. J. Pharm. Sci. **1987**, 76, 715–723.
40. Klech, C.M.; Li, X. J. Pharm. Sci. **1990**, 79, 999.
41. Welz, M.M.; Ofner, C.M., III. J. Pharm. Sci. **1992**, 81, 85–90.
42. Mwangi, J.W. *Ph.D. Dissertation*; University of the Sciences in Philadelphia: 2001.
43. Morris, V.J. Gelation of Polysaccharides. *Functional Properties of Food Macromolecules*; Mitchell, J.R.,

- Ledward, D.A., Eds.; Elsevier Applied Science: London, 1986, 123.
44. Cottrell, I.W.; Kovacs, P. Algin. *Food Colloids*; Graham, H.D., Ed.; AVI: Westport, CT, 1977; 450.
 45. Thomas, S. Pharm. J. **1985**, 235, 188.
 46. Julian, T.N.; Radehaugh, G.W.; Wisniewski, S.J. J. Contr. Release **1988**, 7, 165.
 47. Kaneko, K.; Kanada, K.; Oouchi, K.; Saito, N.; Ozeiki, T. Arch. Pract. Pharm. **1999**, 59, 8–16.
 48. Kennedy, J.F.; White, C.A. *Bioactive Carbohydrates in Chemistry, Biochemistry, and Biology*; Ellis Horwood: Chichester, UK, 1983; 163.
 49. Morris, E.R.; Rees, D.A.; Robinson, G.R. J. Mol. Biol. **1980**, 138, 349.
 50. Morris, V.J.; Chilvers, G.R. Carbohydr. Polym. **1983**, 3, 129.
 51. Winstead, D.A. *Ph.D. Dissertation*; Philadelphia College of Pharmacy and Science: 1997.
 52. Picker, K.M. Drug Dev. Ind. Pharm. **1999**, 25, 339–346.
 53. Balazs, E.A.; Band, P. Cosmet. Toilet. **1984**, 99, 65.
 54. Listrat, V.; Ayrat, X.; Patarnello, F.; Bonvarlet, J.P.; Simonnet, J.; Amor, B.; Dougados, M. Osteoarthritis Cartilage **1997**, 5, 153–60.
 55. Saettone, M.F.; Chetoni, P.; Torracca, M.T.; Burgalassi, S.; Giannaccini, B. Int. J. Pharm. **1989**, 51, 203.
 56. Gidley, M.J.; Morris, E.R.; Murray, E.J.; Powell, D.A.; Rees, D.A. IBJM **1980**, 2, 332.
 57. Sungthongjeen, S.; Pitaksuteepong, T.; Somsiri, A.; Sriamornsak, P. Drug Dev. Ind. Pharm. **1999**, 25, 1271–1276.
 58. Giunchedi, P.; Conte, U.; Chetoni, P.; Saettone, M.F. Eur. J. Pharm. Sci. **1999**, 9, 1–7.
 59. Heckman, E. Starch and its Modifications for the Food Industry. *Food Colloids*; Graham, H.D., Ed.; AVI: Westport, CT, 1977; 478.
 60. Rink, S.G.; Stainsby, G. Prog. Food Nutr. Sci. **1982**, 6, 323.
 61. Witschi, C.; Mrsny, R. J., Pharm. Res. **1999**, 16, 382–390.
 62. Foord, S.A. *Ph.D. Dissertation*; University of Bristol; 1980.
 63. Franz, G. Adv. Polym. Sci. **1986**, 76, 3.
 64. Losse, G.; Muller, F.; Fischer, S.; Wunderlich, G.; Hacker, E. Pharmazie **1989**, 44, 331.
 65. Singh, K.K.; Deshpande, S.G.; Baichwal, M.R. Indian Drugs **1994**, 31, 149–154.
 66. Rozier, A.; Mazuel, C.; Grove, J.; Plazonnet, B. Int. J. Pharm. **1989**, 57, 163.
 67. Imai, T.; Sakai, M.; Ohtake, H.; Azuma, H.; Otagiri, M. Pharm. Res. **1999**, 16, 80–89.
 68. Segal, R.; Pisanty, S. J. Clin. Pharm. Therm. **1987**, 12, 165.
 69. Goldstein, A.M.; Alter, E.N. Guar Gum. In *Industrial Gums*; Whistler, R.L., Ed.; Academic Press: New York, 1959, 338.
 70. Bamba, M.; Puisieux, F.; Marty, J.P.; Carstensen, J.T. Int. J. Pharm. **1979**, 2, 307.
 71. Gliko-Kabir, I.; Yagen, B.; Penhasi, A.; Rubinstein, A. Pharm. Res. **1998**, 15, 1019–1025.
 72. Meer, W.A. Plant Hydrocolloids. *Food Colloids*; Graham, H.D., Ed.; AVI: Westport, CT, 1977; 527.
 73. Mullet, T.; Hakert, H.; Eckert, T. Colloid Polym. Sci. **1989**, 267, 230.
 74. Greminger, G.K.; Savage, A.B. Methylcellulose and its Derivative. *Industrial Gums*; Whistler, R.L., Ed.; Academic Press: New York, 1959; 665.
 75. The Effect of Sodium Carboxymethylcellulose Gel in Patients with Nonhealing Lower Extremity Diabetic Ulcers, American College of Surgeons, Chicago, IL October, 13, 1997.
 76. Kibbe, A.H. *Handbook of Pharmaceutical Excipients*; American Pharmaceutical Association and the Pharmaceutical Press (London): Washington, DC, 2000; 244–248.
 77. Alvarez-Lorenzo, C.; Gomez-Amoza, J.L.; Martinez-Pacheco, R.; Souto, C.; Concheiro, A. Int. J. Pharm. **1999**, 180, 91–103.
 78. Kibbe, A.H. *Handbook of Pharmaceutical Excipients*; American Pharmaceutical Association and the Pharmaceutical Press (London): Washington, DC, 2000; 252–255.
 79. Mitchell, K.; Ford, J.L.; Armstrong, D.J.; Elliot, P.N.; Rostron, C. Int. J. Pharm. **1993**, 100, 165–173.
 80. Greminger, G.K.; Savage, A.B. Methylcellulose and its Derivative. *Industrial Gums*; Whistler, A.L., Ed.; Academic Press: New York, 1959, 579.
 81. Kubis, A.; Dybek, K.; Krutul, H. Pharmazie **1987**, 42, 519.
 82. Zhou, M.; Donovan, M.D. Int. J. Pharm. **1996**, 135, 115–125.
 83. Kibbe, A.H. *Handbook of Pharmaceutical Excipients*; American Pharmaceutical Association and the Pharmaceutical Press (London): Washington, DC, 2000; 79–82.
 84. Barry, B.W.; Meyer, M.C. Int. J. Pharm. **1979**, 2, 1.
 85. Penn, L.E. Gel Dosage Forms: Theory, Formulation, and Processing. *Topical Drug Delivery Formulations*; Osborne, D.W., Amann, A.H., Eds.; Marcel Dekker, Inc.: New York, 1990; 381–338.
 86. Ostrenga, J.; Steinmetz, C.; Poulsen, B. J. Pharm. Sci. **1975**, 60, 1175–1179.
 87. Lochhead, R.Y.; Hemker, W.J.; Castarieda, J.Y. Cosmet Toilet. **1987**, 120, 89.
 88. Shin, S.C.; Kim, J.Y.; Oh, I.J. Drug Dev. Ind. Pharm. **2000**, 26, 307–312.
 89. Rossi, S.; Bonferoni, M.C.; Ferrari, F.; Caramella, C. Pharm. Dev. Tech. **1999**, 4, 55–63.
 90. Bonferoni, M.C.; Rossi, S.; Ferrari, F.; Caramella, C. Pharm., Dev. Tech. **1999**, 4, 45–53.
 91. Chen-Chow, P.C.; Frank, S.G. Int. J. Pharm. **1981**, 8, 89.
 92. Attwood, D.; Collett, J.H.; Tait, C.J. Int. J. Pharm. **1985**, 26, 25.
 93. Nalbandian, R.M.; Henry, R.L.; Balko, K.W.; Adams, D.V.; Neuman, N.R. J. Biomed. Mater. Res. **1987**, 6, 1135.
 94. Miller, S.C.; Donovan, M.D. Int. J. Pharm. **1982**, 12, 147.
 95. Tomida, H.; Shinohara, M.; Kuwada, N.; Kiryu, S. Acta Pharm. Suec. **1987**, 24, 263.
 96. Barichello, J.M.; Morishita, M.; Takamyam, K.; Nagai, T. Int. J. Pharm. **1999**, 184, 189–198.
 97. Saettone, M.F.; Giannaccini, B.; Guiducci, A.; Savigni, P. Int. J. Pharm. **1986**, 31, 261.
 98. Hussain, M.D.; Rogers, J.A.; Mehvar, R.; Vucathala, G.K. Drug Dev. Ind. Pharm. **1999**, 25, 265–271.
 99. Chen, L.L. Pharm. Dev. Tech. **1998**, 3, 241–249.
 100. Molyneus, P. *Water-Soluble Synthetic Polymers and Behavior*; CRC Press: Boca Raton, FL, 1983; I, 122.
 101. Pritchard, J.G. *Polyvinyl Alcohol: Basic Properties and Uses*; MacDonald Technical and Scientific: London, 1970.

102. Morimoto, K.; Nagaysau, A.; Fukanoki, S.; Morisaka, K.; Hyon, S.; Ikada, Y. *Pharm. Res.* **1989**, *6*, 338.
103. Wang, N.; Wu, X.S.; Li, J.L. *Pharm. Res.* **1999**, *16*, 1430–1435.
104. Quintanar-Guerrero, D.; Ganem-Quintanatar, D.; Raygoza-Trejo, D.; Doelker, E. *Drug Dev. Ind. Pharm.* **1999**, *25*, 169–175.
105. Nairn, J.G. Solutions, Emulsions, Suspensions, and Extracts. *Remington: The Science and Practice of Pharmacy*; Gennaro, A.R., Ed.; Mack: Easton PA, 2000, 1515–1518.
106. Shah, D.N.; White, J.L.; Hem, S.L. *J. Pharm. Sci.* **1981**, *70*, 1101.
107. Neumann, B.S.; Sansom, K.G. *J. Soc. Cosm. Chem.* **1970**, *21*, 237.
108. Shrivastava, R.; Jain, S.R.; Frank, S.G. *J. Pharm. Sci.* **1985**, *74*, 214.
109. Ritschel, W.A.; Ritschel, G.B.; Sathyan, G. *Res. Commun. Pathol. Pharmacol.* **1988**, *62*, 103.
110. Szan, A.; Papp, L. *J. Control. Rel.* **1998**, *56*, 239–247.

BIBLIOGRAPHY

- Drug Facts and Comparisons. In *Facts and Comparisons*; St. Louis MO, 2000.
- Graham, H.D. *Food Colloids*; AVI: Westport, CT, 1977.
- Kennedy, I.F.; White, C.A. *Bioactive Carbohydrates in Chemistry, Biochemistry, and Biology*; Ellis Horwood Limited: Chichester UK, 1983.
- Kibbe, A.H. *Handbook of Pharmaceutical Excipients*; American Pharmaceutical Association and The Pharmaceutical Press (London): Washington DC, 2000.
- Mitchell, J.R.; Ledward, D.A. *Functional Properties of Food Macromolecules*; Elsevier Applied Science: London, 1986.
- Physicians' Desk Reference for Nonprescription Drugs*; Medical Economics Company Inc. Oradell NJ, 1998.
- Physicians' Desk Reference*; Medical Economics Company: Oradell NJ, 2001.

GENERIC DRUGS AND GENERIC EQUIVALENCY

Arthur H. Kibbe

Wilkes University, Wilkes-Barre, Pennsylvania, U.S.A.

INTRODUCTION

All drugs that are approved for sale generally carry at least two names. The drugs are given a proprietary or trade name given by the company that first develops them. These companies often are referred to as the innovator company. The drug is assigned a nonproprietary or generic name, which is agreed to by the WHO International Nonproprietary Nomenclature (INN) Committee and the U.S. Adopted Names Council (USAN). A new drug is usually first marketed with some patent protection and at a price that, at a minimum, recoups the cost of development over the remaining life of the patent or other exclusivity arrangement. Eventually, protection from competition is lost to other pharmaceutical companies, often companies or divisions of companies that specialize in marketing off-patent drugs. These companies or divisions are called generic companies. They can apply to the appropriate regulatory body such as the Food and Drug Administration (FDA) for permission to market the same active ingredient under its nonproprietary or generic name. The generic manufacturer is not required to do a complete clinical trial to prove effectiveness and safety because that has already been well established for the drug. However, it is required to show that the new drug product is equivalent to the original drug product. For the purposes of this article, we define the drug as the chemical that has the pharmacological effect and the drug product as a dosage form that contains the drug and other ingredients or excipients that allow the formulation of the dosage form. There is a large economic incentive for the development of generic drug products, especially for highly successful drug products. The pharmaceutical company that first brought the product to market maintains the price at the original level or higher to continue the cash flow into the company. This allows the other companies to develop a formulation of the drug and to win approval to market with the knowledge that, even at a fraction of the selling price of the innovator's product, the company can make a good profit. Some innovators defend their market share by arguing quality and reliability. The FDA must act as an impartial arbitrator of this debate. The debate is clearly about money, but is argued in a scientific forum. The key question is, "Are we

sure that the two products, if used in the same way in the same patient, will yield the same result." If a drug product is subject to this debate, the innovator always says "no" and the second and subsequent manufacturers always say "yes." In the United States, the FDA sets the standards against which the question is resolved, and scientists take sides usually on the issue of "are the current FDA standards good enough." If the FDA gives an "A" rating to a drug product, it is in effect telling the prescriber that the drug product will yield the same therapeutic and side-effects profile as the innovator drug product. The Orange Book specifies the equivalence rating from the FDA. Almost all generic drug products currently marketed are rated A; the FDA has not approved a generic without an A rating in decades. Finally, the consumer pays the price, either in the unnecessarily high cost of drugs if unnecessary studies are performed and generic competition delayed or in risky drug substitution if the FDA is too relaxed in its standards. The tests required by the FDA have changed over the years. They have become more proscriptive and are based on sound statistical grounds. The FDA has also increased the level of oversight of the pharmaceutical companies that manufacture generic equivalents of innovator products. Thus, the regulatory process has become more stringent, and the level of assurance that the public has that a generic product is both safe and effective has gone up. The FDA has often stated that there are no known therapeutic failures from switching among products that have been ruled as equivalent by the FDA.

LEGISLATIVE AND REGULATORY HISTORY

In the early 1970s, most states had ant substitution laws that required the dispensing of the innovator product when the prescriber wrote for a drug by trade name. Most physicians had learned only the trade name of the drug product, and these laws ensured that generic substitution would be at a minimum (1). The American Pharmaceutical Association (APhA) along with other groups pushed for the repeal of these laws and opened the way for the growth

of the generic industry. The lack of bioequivalence data available at that time led to the formation of the Generic Drug Bureau within the Food and Drug Administration. As a result of the efforts of that group, the FDA produced a book, *Approved Drug Products With Therapeutic Equivalence Evaluations*, in the late 1960s. This became known as the Orange Book because of the cover color. The book has been published annually with monthly updates. The contents are now available on the FDA Website (2).

In 1984, the Drug Price Competition and Patent Term Restoration Act was passed. This act, also known as the Waxman–Hatch Bill of 1984, encouraged the development of new innovative drugs by established procedures, extended patent rights, and facilitated the FDA approval process for generic drugs (3). To address the first goal, the law created a mechanism to extend the period of patent protection for manufacturers of innovative new drugs generally ensuring at least 5 years of market exclusivity after approval. To address the second goal, the law established an Abbreviated New Drug Application (ANDA) for applications after 1962. Drugs chemically equivalent to those previously approved by a full application process need only be proven bioequivalent, not clinically equivalent. Depending on the drug, proof of bioequivalence can involve in vitro dissolution studies, in vivo single-dose bioavailability studies, in vivo multidose bioavailability studies, or a combination of these. However, in vitro dissolution studies alone are not adequate proof of bioequivalence for purposes of an ANDA.

SCIENTIFIC BASIS FOR GENERIC DRUG PRODUCT EQUIVALENCY: BIOAVAILABILITY–BIOEQUIVALENCY

The goal of the testing of generic products is not to establish the clinical usefulness of the drug but only to ensure that the generic product or new formulation has the same relative bioavailability as or is bioequivalent to the innovator product.

Bioavailability has been defined as a measure of the rate and extent of absorption of a drug into the systemic circulation after administration of a dosage form. An intravenous i.v. dose is considered by definition to be 100% bioavailable. All other routes of administration will produce a total bioavailability less than or equal to that of the i.v. dose. Thus, only a drug that is completely absorbed into the systemic circulation can have the extent of bioavailability equal to the dose stated on the label. In addition to the extent of absorption, the rate of absorption

plays a key role when evaluating the potential therapeutic impact of a particular dosage form. Knowledge of the time to onset of drug action, which is directly related to rate of absorption, is a significant concern, especially in acute clinical situations such as asthma attack, hyperglycemic shock, and pain.

The bioavailability of drugs from specific dosage forms is affected by the nature of the inactive ingredients or pharmaceutical excipients and the process used in its formulation. (For additional information, see Bioavailability of Drugs and Bioequivalency in this Encyclopedia.) When comparing similar dosage forms from different manufacturers or different lots from the same manufacturer, it is most useful to determine the relative bioavailability of the two products or lots. Some scientists have attempted to establish an in vitro test that could successfully predict in vivo bioavailability. However, to date, none has been developed.

Pharmacokinetics means the application of kinetics to drugs. It can be defined as the study of the time course and fate of drugs in the body. Teorell is often given credit for the origin of pharmacokinetics with his publications, *Kinetics of Distribution of Substances Administered to the Body* (4, 5). This science is the theoretical support for the use of bioequivalency testing to establish therapeutic equivalence among dosage forms of the same drug. The first approach to a pharmacokinetic understanding of drugs in the body, called compartment analysis, considered the body as a group of compartments through which the drug must pass. The compartment itself does not exist but represents the average of many processes that give rise to the observed phenomenon. The size of the imaginary compartment can be calculated and is useful in understanding the process of absorption, distribution, and elimination or metabolism of the drug. Regardless of the model used, a plot of the plasma concentration of the drug versus time yields a curve that can be described by a polyexponential equation. The area under that concentration–time curve (AUC) is directly related to the amount of drug absorbed. The time to reach peak concentration and the peak concentration itself are related to both the dose and the rate of absorption.

An important limitation of compartment analysis is that it cannot be applied universally to any drug. A simpler approach that is useful in the case of bioequivalency testing is the model independent method. It is based on statistical-moment theory. This approach uses the mean residence time (MRT) as a measure of a statistical half-life of the drug in the body. The MRT can be calculated by dividing the area under the first-moment curve (AUMC) by the area under the plasma curve (AUC) (6).

(See other articles in this Encyclopedia for more detailed discussion of these subjects.)

MEASUREMENT OF RELATIVE BIOAVAILABILITY OR BIOEQUIVALENCY

Drug products often undergo bioavailability testing in the early stages of development. Changes in formulation necessitated by results of clinical trials or stability testing or changes in the availability of excipients or changes in suppliers of excipients often require that the manufacturer perform a relative bioavailability or bioequivalency test to ensure that subsequent lots of a product will yield the same amount of active ingredient at the same rate as was possible in earlier formulations.

Bioequivalency studies are usually performed on young, healthy, male adult volunteers under controlled dietary conditions and fixed activity levels. This is because the goal of the study is not to establish the clinical usefulness of the drug but only to ensure that the two formulations have the same relative bioavailability or are bioequivalent.

Key Parameters

When assessing bioequivalence, the following three parameters that characterize the plasma or blood concentration–time profile of the administered drug are usually measured:

1. Peak height, C_{\max} , represents the highest concentration of the drug in the systemic circulation;
2. Time to peak, t_{\max} , represents the time for peak height to occur after the drug was administered;
3. Area under the curve, AUC, represents the total integrated area under the concentration–time curve.

The first two parameters are indicators of absorption rate, whereas the third is directly proportional to the extent of drug absorbed into the systemic circulation from the dosage form. Figure 1 is an example of a concentration–time curve for a single dose of drug to a subject.

Although it is theoretically possible to determine the rate and extent of absorption of a drug by measurement of the rate and extent of the appearance of the drug in the urine, this is not considered as reliable a method for evaluation of a drug product's bioequivalency as are blood level data. Thus, the studies commonly performed to demonstrate bioequivalence fall into two categories: single-dose and multidose or steady-state studies. There are advantages and disadvantages to each. Single-dose studies are less expensive and expose healthy volunteers

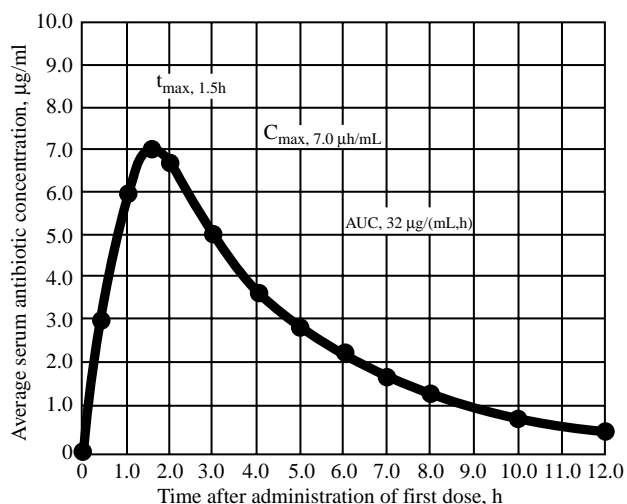


Fig. 1 Blood concentration curve.

to less drug during the course of the study. However, these studies require more sensitive analytical methods and have higher subject-to-subject variability. In both cases, a cross-over study design is used to control for sequence effects. The study is designed to control for or take into account as many variables as possible. The subjects are randomly assigned to groups. Blood samples are obtained from each subject before dosing and at fixed time intervals after dosing. Currently, the data are then analyzed using appropriate statistical ANOVA. The results must meet FDA guidelines for mean and 90% confidence interval for each of the three key parameters. For oral solid dosage forms, the FDA requires that for a product to be considered bioequivalent, the ratio of the parameter for the two products, together with their 90% confidence interval, must fall between 0.8 and 1.25, using log-transformed data. This, in effect, means that drug products that differ by more than 10% in their rate and extent of absorption will not be approved as generic equivalents.

CURRENT SCIENTIFIC ISSUES

Two issues have been raised recently with regard to the approval of generic drugs. The first has to do with the issue of "Narrow Therapeutic Index Drugs," and the second has to do with the use of individual bioequivalence in place of average bioequivalence. The former concern has been addressed in detail by Drs. Benet and Goyan (7). They concluded that narrow-therapeutic-range drugs were the least likely to have therapeutic failures among generic

drugs, with proof of bioequivalency. The use of average bioequivalence data is under attack. This is because of the concern that there might be a significant subject-by-product interaction. Regulatory agencies now assume that this is not the case (8). The advantage of using individual bioequivalence studies is the reassurance that if subject/product interactions do occur, the study design would control for them, and a more statistically valid measure of the rate and extent of absorption of the drug from the two product would be determined. Some of the disadvantages associated with shifting from average to individual bioequivalence testing are cost, numbers of subjects needed, and diversity of the study population required. [See other articles that address the impact of the new metrics on the reliability and cost of the performance of bioequivalence testing (9–12).]

THE CHANGING POLITICAL ECONOMY OF GENERIC DRUGS

The modern generic drug industry in the United States really only dates from the passage of the Waxman–Hatch Act in 1984. Within 5 years of passage, generic drugs captured 40% of the market for prescriptions written in the United States. Since that time, the generic drug market share has stabilized between 40 and 50% of the prescriptions written. However, the dollars paid for generic drugs are only 10% of the total sales of drugs in the United States. That statistic alone tells us that the consumer receives enormous benefit from the substitution of therapeutically equivalent generic drugs when available.

A horrendous scandal hit the industry in the late 1980s, wherein firms representing 75% of the production of the generic industry pled guilty to one or more criminal charges involving filing false applications with the FDA, paying illegal gratuities to FDA personnel, and/or related crimes to gain an unfair competitive advantage in the emerging marketplace. Surprisingly, this scandal produced only a small delay in the market share march of generic drugs and only a temporary loss of consumer confidence in generic products.

The scandal was tied to a phenomenon that still dominates the business strategies of generic drug firms to this day: the need to obtain approval to manufacture and distribute before other firms enter the market. Because of the “commodity” nature of the business and the relative ease of entry into the industry, firms devote most of their resources and managerial talent to obtaining first or second approvals from the FDA for their products. Once a generic

drug has four or more competitors, it is no longer profitable for additional generic companies to enter the market.

Generic drug manufacturers typically will continue to manufacture drugs that produce little or no profit because large purchasers that are their prime customers (chain drugs stores, buying groups for smaller community pharmacies, etc.) prefer to buy from companies that can supply most of the common generic drugs. For example, if a generic drug firm no longer produces amoxicillin because it can make more money by shifting its antibiotic production facilities to, for example, a cephalosporin drug for which it has less competition, a large chain may choose to buy its entire generic antibiotic line from another company that supplies both.

The profitable generic drug companies are profitable because they have found a strategy to maintain some control over the price of their products. In the early years (1984–1988), the best way to get “first approval” from the FDA apparently was to be first to file, to get assays or bioequivalence studies done on difficult to duplicate drugs, or to find some way to get an expedited approval from inside the agency. Unfortunately, this sometime involved payoffs to FDA review chemists (those FDA experts assigned the task of evaluating biostudy results, the crucial piece of a generic drug application, remained remarkably free of the scandal). More often, it involved submitting false information to the FDA (including, in a few cases, false biostudies). Many generic drug firms did not survive the scandal, and others survived only after the previous management and ownership were purged from the firms.

For a short period, it was believed that the profitable segment of the business involved not production but distribution. After all, if commodity prices approach marginal cost and the marginal cost of manufacturing drugs is minimal, but the price to the consumer remains significantly more than marginal, there must be middlemen somewhere making the money. Clearly, those middlemen were not in the retail pharmacy where profits continued to be squeezed. Distributors were thought to be the new profit centers. But a funny thing happened on the way to that particular bank...

Consumers became outraged at the rapid increase in the price of pharmaceuticals as the innovator companies (and some generic firms) rushed to raise prices and as generic drug company after generic drug company was pushed out of the industry in the wake of investigations by a Congressional committee and a federal grand jury. Second, the Administration, in response to public concern about the cost of pharmaceuticals, pressured the pharmaceutical industry and forced lower prices and significant rebates to the federal and state government

programs that paid for drugs. Wholesale distributors of all drugs subject to the federal rebates suffered.

Finally, the firms that thought they could profit most from the scandal entered the market. These were innovator firms, many of which had already played a significant role in the distribution of generics. Ultimately, the profit margins from generic drug sales were not sufficient to carry the overhead of the branded companies, and most left the market or returned to their distributor role. Even in the case of the firms manufacturing and marketing generic versions of their own branded products, giving them significant advantage over the remaining pure generic firms in developing and filing of the ANDAs with the FDA and the added advantage of relatively less scrutiny from the scandal-rocked agency, most had exited the marketed by the end of the decade.

Some innovator firms entered the generic drug market so that they could have a product line consistent with their new business strategy: disease state management. This strategy, a function of the rise of HMOs and the return of the concept of scarcity to prescription drug dispensing, was intended to involve the development of a continuum of drug therapies for the treatment of a specific illness (diabetes, depression, etc.), wherein the patient would be tried on the older, less-expensive drug first and, if it did not work, the next most cost-effective drug would be administered and so on until the least cost-effective drug would be the treatment of last resort. Unfortunately, the branded companies that selected this strategy found themselves competing with doctors, hospitals, and insurance companies for control of the treatment regime of individual patients, a losing proposition for the entity with the least amount of information about and access to the individual patient.

Another factor in reducing prices of all drugs that had some form of competition was the rise of the HMO and its pharmaceutical watchdog, the pharmacy benefit manager (PBM). These PBMs create a formulary of approved drugs (drugs for which they would reimburse partially or fully) based on bids from competing companies.

Much of the public's confusion regarding generic drugs arose from a practice of the PBMs to pressure doctors to substitute different chemical entities in the same therapeutic class for the prescribed medicine. Such a switch is called a therapeutic substitution as opposed to the switching among manufacturers of therapeutically equivalent drugs (generics and the innovator drug or other FDA "AB"-rated substitutes). Therapeutic substitution involves a switch to a different drug, whereas generic substitution involves a switch to the same drug from a different manufacturer. If a patient is switched

between FDA "AB"-rated drugs, the FDA offers the assurance that they can expect the same therapeutic and side-effect profile as the brand drug or another "AB"-rated generic drug. The FDA offers no such assurance if the switch occurs among different drugs, even if they are in the same therapeutic class. For example, aspirin and Tylenol may be equally effective in the treatment of headache, but the FDA makes no such certification, whereas it makes exactly that certification for Bayer aspirin and Safeway aspirin.

The dominance of the HMO (and related organizations) and their PBMs (and related organization types) served to accelerate the substitution of generic drugs at the turn of the 21st century. However, even that pressure could not slow the re-emergence of a high rate of price increase, greater than consumer or comparable wholesale prices as a whole, in prescription drugs. Innovator companies learned that establishing very high prices for "breakthrough" drugs could more than compensate for the loss of patent protection on a highly profitable drug.

Furthermore, the United States is the only developed country in the world that has chosen not to explicitly control the price of any drug product and has used its market power as a huge buyer relatively sparingly. Consequently, U.S. prices for drug products still under patent are usually substantially above those charged anywhere else in the world. Generic prices approach cost except for those few generics that have managed to eliminate or limit for a specific period competition from other generics.

Those generic drug firms that prospered in this restrictive price environment all had one or more niche drugs that were immune from corrosive price competition. Some companies mastered a manufacturing process that produced bioequivalent medicine that the innovator itself found difficult to master lot to lot. Others took advantage of certain exclusivity provisions in the law for those that challenged a product patent in court, ostensibly to cover the cost of litigation. In other cases, the settlement of those cases provided some form of licensing or distribution rights that permitted the sale of a generic product while the patent was still valid. Finally, a fortunate firm might find itself in possession of the exclusive right to purchase the raw material from the only source available to generic drug manufacturers.

All generic drug firms capable of generating the necessary cash to develop and market new drugs are moving toward that lucrative market. For the time being, the United States has chosen to use the market mechanism as its only important control on drug prices. Generics are the competition, and competition is our only real form of price control.

According to the Congressional Budget Office (CBO), consumers saved \$8–10 billion in 1994 because of the use of generic drugs. In that same 1998 report, CBO cited the Waxman–Hatch Act, generic substitution laws passed by the states, and government health programs as seminal events leading to the acceptance of generic drugs and the resulting savings.

REFERENCES

1. Knoben, J.E.; Scott, G.R.; Tonelli, R.J. Overview of the FDA Publication Approved Drug Products with Therapeutic Equivalence Evaluations. *Am. J. Hosp. Pharm.* **1990**, *47* (12), 2696–2700.
2. Food and Drug Administration. *Approved Drug Products with Therapeutic Equivalence Evaluations*; <http://www.fda.gov/cder/ob/> FDA: Rockville; Available from the Government Printing Office: Washington, DC, 2000.
3. Weaver, L.C. Drug Cost Containment—The Case for Generics: Situation in the U.S.A. *J. Soc. Admi. Pharm.* **1989**, *6* (1), 9–13.
4. Teorell, T. Kinetics of Distribution of Substances Administered to the Body. I. The Extravascular Modes of Administration. *Archives Internationales de Pharmacodynamie et de Therapie.* **1937**, *57*, 205–225.
5. Teorell, T. Kinetics of Distribution of Substances Administered to the Body. II. The Intravascular Mode of Administration. *Archives Internationales de Pharmacodynamie et de Therapie.* **1937**, *57*, 226–240.
6. Yamaoka, K.; Nakagawa, T.; Uno, T. Statistical Moments in Pharmacokinetics. *J. Pharmacokinet. Biopharm.* **1978**, *6* (6), 547–558.
7. Benet, L.Z.; Goyan, J.E. Bioequivalence and Narrow Therapeutic Index Drugs. *Pharmacotherapy* **1995**, *15* (4), 433–440.
8. Patnaik, R.N.; Lesko, L.J.; Chen, M.L. Individual Bioequivalence. New Concepts in the Statistical Assessment of Bioequivalence Metrics. FDA Individual Bioequivalence Working Group. *Clin. Pharmacokine.* **1997**, *33* (1), 1–6.
9. Midha, K.K.; Rawson, M.J.; Hubbard, J.W. Individual and Average Bioequivalence of Highly Variable Drugs and Drug Products. *J. Pharm. Sci.* **1997**, *86* (11), 1193–1197.
10. Snikeris, F.; Tingey, H.B. A Two-Step Method for Assessing Bioequivalence. *Drug Inf. J.* **1994**, *28* (3), 709–722.
11. Holder, D.J.; Hsuan, F. A Moment-Based Criterion for Determining Individual Bioequivalence. *Drug Inf. J.* **1995**, *29* (3), 965–979.
12. Mohandoss, E.; Chow, S.C.; Ki, F.Y. Application of Williams' Design for Bioequivalence Trials. *Drug Inf. J.* **1995**, *29* (3), 1029–1038.

FREEZE DRYING

Michael J. Pikal

University of Connecticut, Storrs, Connecticut, U.S.A.

INTRODUCTION

Freeze drying, also termed “lyophilization,” is a drying process employed to convert solutions of labile materials into solids of sufficient stability for distribution and storage. A typical production scale freeze dryer consists of a drying “chamber” containing temperature-controlled shelves, which is connected to a “condenser” chamber via a large valve. The condenser chamber houses a series of plates or coils capable of being maintained at very low temperature (i.e., less than -50°C). One or more vacuum pumps in series are connected to the condenser chamber to achieve pressures in the range of 0.03–0.3 Torr in the entire system during operation. A commercial freeze dryer may have 10–20 shelves with a total load of the order of 50,000 vials. The objective in a freeze-drying process is to convert most of the water into ice in the *freezing stage*, remove the ice by direct sublimation in the *primary drying stage*, and finally remove most of the unfrozen water in the *secondary drying stage* by desorption. The water removed from the product is reconverted into ice by the condenser.

In a typical freeze-drying process, an aqueous solution containing the drug and various formulation aids, or “excipients,” is filled into glass vials, and the vials are loaded onto the temperature-controlled shelves. The shelf temperature is reduced, typically in several stages, to a temperature in the vicinity of -40°C , thereby converting nearly all the water into ice. Some excipients, such as buffer salts and mannitol, may partially crystallize during freezing, but most “drugs,” particularly proteins, remain amorphous. The drug and excipients are typically converted into an amorphous glass also containing large amounts of unfrozen water (15–30%) dissolved in the solid (i.e., glassy) amorphous phase. Thus, most of the desiccation actually occurs during the freezing stage of the freeze-drying process. After all water and solutes have been converted into solids, the entire system is evacuated by the vacuum pumps to the desired control pressure, the shelf temperature is increased to supply energy for sublimation, and primary drying begins. Due to the large heat flow required during primary drying, the product temperature runs much lower than the shelf temperature.

The removal of ice crystals by sublimation creates an open network of “pores,” which allows pathways for escape of water vapor from the product. The ice–vapor boundary (i.e., the boundary between frozen and “dried” regions) generally moves from the top of the product toward the bottom of the vial as primary drying proceeds. Primary drying is normally the longest part of the freeze-drying process. Primary drying times of the order of days are not uncommon, and in rare cases, weeks may be required for a combination of poor formulation and suboptimal process design. Although some secondary drying does occur during primary drying (i.e., desorption of water from the amorphous phase occurs to a limited extent once the ice is removed from that region), the start of secondary drying is normally defined, in an operational sense, as the end of primary drying (i.e., when all ice is removed). Of course, because not all vials behave identically, some vials enter secondary drying whereas other vials are in the last stages of primary drying. When the judgment is made that all vials are devoid of ice, the shelf temperature is typically increased to provide the higher product temperature required for efficient removal of the unfrozen water. The final stages of secondary drying are normally carried out at shelf temperatures in the range of 25 – 50°C for several hours. Here, because the demand for heat is low, the shelf temperature and the product temperature are nearly identical.

As freeze-drying plants are very expensive and process times are often long, a freeze-dried dosage form is relatively expensive to produce. Indeed, because of both cost and ease of use, a ready-to-use solution is the preferred option for a parenteral dosage form, particularly if the solution can withstand terminal heat sterilization. When an aqueous solution does not have sufficient stability, the product must be produced in solid form. At least for small molecules, stability normally increases in the following order: solution < glassy solid < crystalline solid (1–3); this is likely a result of restricted motion in solids with the high degree of order in the crystalline solid limiting reactivity even further. As many pharmaceuticals cannot be produced on a commercial scale by crystallization, a glassy solid may be the only solid-state option.

Freeze drying (4–7) and spray drying (8–10) are drying methodologies in common use in the pharmaceutical industry, and are suitable for the production of glassy solids. Freeze drying is a low-temperature process. In general, a formulation can be dried to 1% water or less, without any of the product exceeding 30°C. Thus, conventional wisdom states that freeze-drying is less likely to cause thermal degradation than a high-temperature process such as spray drying. Historically, freeze-drying is the method of choice for products intended for parenteral administration. Sterility and relative freedom from particulates are critical quality attributes for parenterals. Largely because the solution is sterile filtered immediately before filling into the final container, and further processing is relatively free of exposure to humans, a freeze-drying process maintains sterility and particle-free characteristics of the product much more easily than do processes that must deal with dry powder handling issues, such as dry powder filling of a spray-dried or bulk-crystallized powder. Indeed, with modern robotics automatic loading systems (11), humans can be removed from the sterile processing area entirely, at least in principle. Furthermore, as the vials are sealed in the freeze dryer, moisture control and control of headspace gas can easily be controlled, an important advantage for products whose storage stability is adversely affected by residual moisture and/or oxygen. As the critical heat and mass transfer characteristics for freeze-drying are nearly the same at the laboratory scale as in full production, resolution of scale-up problems tends to be easier for a freeze-drying process than for spray drying, at least in our experience. Also, development of a freeze-dried product requires less material for formulation and process development, a particularly important factor early in a project.

While freeze-drying has a long history in the pharmaceutical industry as a technique for stabilization of labile drugs, including proteins, many proteins suffer irreversible change, or degradation, during the freeze-drying process (12–16). Even when the labile drug survives the freeze-drying process without degradation, the resulting product is rarely found perfectly stable during long-term storage, particularly when analytical techniques with a sensitivity to detect low levels of degradation ($\approx 0.1\%$) are employed. Both small molecules (1–3, 17) and proteins (18–21) show degradation during storage of the freeze-dried glass. In some cases, instability is serious enough to require refrigerated storage (18, 19, 22).

Stability problems are most often addressed by a combination of formulation optimization and attention to process control. Lyoprotectants are added for stability during the freeze-drying process as well as to provide

storage stability, and the level and type of buffer is optimized. Optimization of the freezing process may be critical; control of product temperature during drying is critical for products that tend to suffer cake collapse during primary drying, and control of residual moisture is nearly always critical for storage stability. Formulation and process are interrelated: A bad formulation can be nearly impossible to freeze dry, and even with a well designed formulation, a poorly designed process may require more than a week to produce material of suboptimal quality. Although blind empiricism may, in time, yield an acceptable formulation and process, an appreciation for the materials science of amorphous systems and some understanding of heat and mass transfer relevant to freeze-drying are needed for efficient development of freeze-dried products. Obviously, one also requires at least a phenomenological understanding of the major degradation pathways specific to the drug under consideration.

The objective of this article is to present the scientific and engineering fundamentals most useful in the development of formulations and processes for the manufacture of freeze-dried pharmaceuticals. Generalizations are illustrated with specific examples from the literature, but no attempt is made to survey all published works. Most of the section on the freeze-drying process applies equally well to small molecules and proteins, whereas most of the section on formulation and stability is specific to proteins.

THE FREEZE-DRYING PROCESS

Freezing

Freeze concentration

The objective of freezing is to remove most of the water from the system by formation of ice and to convert all solutes into solids, either crystalline solids or a glass. Once the sample is a solid, primary drying may begin. As a sample is cooled, the system remains liquid well below the equilibrium freezing point, but with sufficient supercooling, nucleation of ice proceeds rapidly and growth of ice crystals begins. As liquid water converts to ice, all solutes are concentrated in the regions between ice crystals, ultimately concentrating until they crystallize or until the system increases in viscosity sufficiently to transform into a solid amorphous system, or glass. The example shown in Fig. 1 is representative of a solute that does not crystallize during freeze-drying. The system supercools about 10°C to a temperature of -15°C before ice nucleation becomes rapid enough to generate crystals

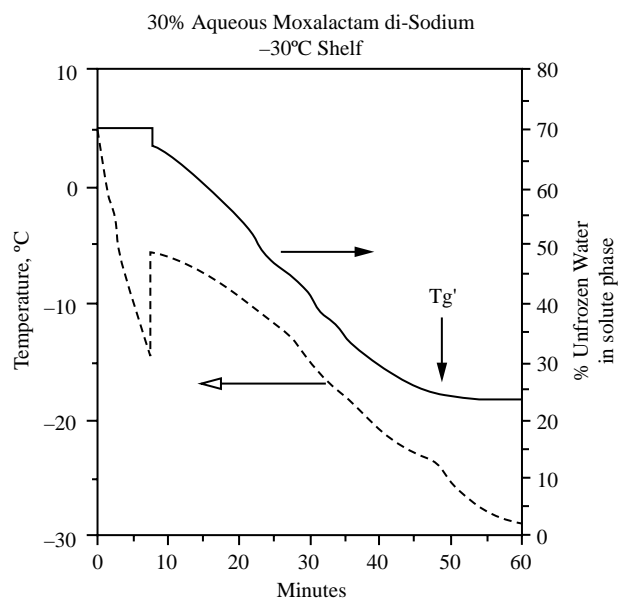


Fig. 1 Freeze concentration in an amorphous system. The product temperature is shown as a broken line whereas the percentage of unfrozen water is shown as a solid line. (From Ref. 6.)

of ice. The evolution of heat caused by the sudden crystallization of ice increases the sample temperature to roughly the equilibrium freezing point (-5°C). The sudden ice crystallization also results in a small but sharp decrease in the percentage of water in the solute phase (given on the right vertical axis). After the initial ice nucleation and crystallization at 10 min, the product cools with continuous conversion of water to ice, thereby decreasing the amount of water in the solute phase and increasing the concentration of solute in the remaining solution. When the product temperature reaches about -24°C , denoted as T_g' , the concentration of water in the solute phase has been reduced to about 24%, and further reduction in water concentration is curtailed on the time scale of freezing, even though the sample temperature decreases further to about -30°C at the end of the freezing process at 60 min. Here, the solute phase has been concentrated from about 30% solute to about 76% solute, and most of the water has been separated from the solute phase as ice. Had the initial concentration been much lower, say 1%, the final concentration would still be 76%, although the freezing profile would differ quantitatively from that shown in Fig. 1. Clearly, most of the desiccation in a freeze-drying process occurs during the freezing process. While desiccation is indeed the objective, it must be recognized that concentration of all solutes during freezing will dramatically increase the probability of

bimolecular collisions, which could produce unexpected instability in the partly frozen system. If a system, initially at 1% solute, is susceptible to a second-order degradation reaction with an activation energy of 25 kcal/mol, a reduction in temperature from 5°C to -24°C would reduce the rate constant by a factor of 200, but an increase in concentration from 1% to 76% would increase the concentration factor in the rate equation by a factor of 5800, yielding a factor of 29 net increase in reaction rate. Although this example is an oversimplification, it is significant to note that increases in second-order degradation rates have been observed in model systems during freeze concentration (23).

T_g' is the temperature at which a sharp increase in baseline occurs during a differential scanning calorimeter (DSC) scan of a frozen solution, suggesting a sharp increase in heat capacity, and has been referred to as the glass transition temperature of the maximally concentrated freeze concentrate (24). A sharp decrease in electrical resistance of the frozen system is noted at the same temperature (25). Structural collapse of a cake structure in the dried region, indicating viscous flow, occurs during primary drying at the collapse temperature, denoted by T_c , which is only slightly higher than T_g' (25). It is clear that T_g' corresponds to the temperature at which mobility in the system is manifested on an experimental timescale. Effectively, the system behaves as a solid below T_g' .

Various electrolytes, such as buffer salts and NaCl, if present in the formulation, will also concentrate during the freezing process. Such exposure to high concentrations of electrolytes (i.e., about 6 molal NaCl during freezing a 0.15 M NaCl solution) might contribute to the destabilization of the native conformation, thereby leading to degradation during freezing.

Crystallization of excipients and consequences

In the case of protein drugs, the drug does not crystallize, but other solute components may (or may not) crystallize, depending on their nature and concentration, other formulation components, and the details of the freezing process. High initial concentrations and slow freezing tend to favor crystallization. Depending on the intended role of the excipient, crystallization may be desirable or undesirable, and it may be important just when in the process crystallization does occur. For example, if mannitol is intended only as a bulking agent, and is not expected to play a role in stabilization of the drug, crystallization is desirable as crystalline mannitol can easily be freeze dried at high temperatures to give an elegant product, meaning a short freeze-drying cycle and freedom from product defects such as collapse. Here, the drug form is that of an amorphous coating on the

crystalline mannitol, with stability properties normally close to that of a system freeze dried without the mannitol. If mannitol remains amorphous during freezing and crystallizes as the product temperature is increased in early primary drying, extensive vial breakage can occur (26). Conversely, if mannitol is intended to serve as a stabilizer during drying and/or storage, crystallization is undesirable. Complete crystallization yields a system equivalent to a physical mixture of glassy drug particles and crystalline mannitol, with no opportunity for a stabilizing interaction except at the phase boundary between the particles, no dilution of drug molecules, and no separation of "reactants" in the protein phase from the protein molecules. In short, any plausible stabilization mechanism requires that at least some of the mannitol remain molecularly dispersed in the amorphous drug phase. Of course, just as crystalline mannitol has some solubility in any liquid, mannitol will have a nonzero thermodynamic solubility in the drug phase. Thus, even if mannitol crystallization proceeds to thermodynamic equilibrium, a small amount of mannitol will remain in the drug phase. However, most of the stabilization potential of the mannitol will be lost upon crystallization. As mannitol does tend to crystallize nearly completely when present as the major formulation component, mannitol is generally a poor choice as a stabilizer for freeze drying.

Although buffers are included in a formulation to maintain constant pH, selective crystallization of the less soluble buffer component during freezing can result in massive pH shifts in the freeze concentrate. Figure 2 shows the pH changes observed during equilibrium freezing of several pure buffer systems (27). Freezing

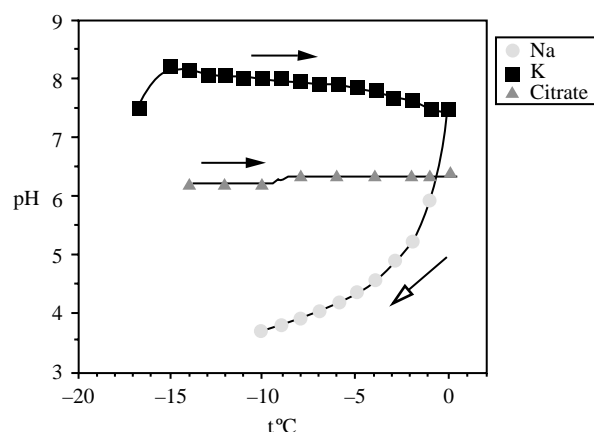


Fig. 2 Shifts in pH during freezing arising from buffer salt crystallization. Near-equilibrium conditions were achieved by seeding with ice and salt. circles: sodium phosphate; squares: potassium phosphate; triangles: citrate. (Data from Ref. 27.)

was carried out while seeding with ice and crystalline buffer salts, so the data represent near-equilibrium behavior. Over the temperature range studied, the citrate buffer system showed no significant pH shift. The sodium phosphate buffer system shows a dramatic decrease in pH of about 4 pH units due to crystallization of the basic buffer component, $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$. Conversely, the potassium phosphate system shows only a modest increase in pH of about 0.8 pH unit. Under nonequilibrium conditions (i.e., no seeding) and with lower buffer concentrations, the degree of crystallization is less, and the resulting pH shifts are moderated (28). Table 1 shows data accumulated (29, 30) during freezing of phosphate buffer solutions in large volumes at cooling rates intended to mimic freezing in vials. For the concentrated buffer solutions (100 mM), the frozen pH values are close to the equilibrium values given in Fig. 2. However, lowering the buffer concentration by an order of magnitude considerably reduces the pH shift observed during freezing. It should also be noted that the small pH shift for potassium phosphate buffer noted in Fig. 2 is a result of the starting pH being 7.5. As shown in Table 1, if the initial pH is 5.5, the 100 mM potassium phosphate buffer increases by 3.1 pH units during freezing. In short, the frozen system pH is near 8.7, regardless of whether the initial pH is 7.0 or 5.5. Clearly, if drug stability is sensitive to pH shifts, buffer crystallization must be avoided. In our experience, the best solution is to formulate such that the weight ratio of buffer to other solutes is very low (18, 22). However, the precise meaning of "very low" varies with the nature and amount of the other solutes as well as with the nature of the buffer.

Instability during freezing and T_g'

Most small molecules are stable during the freezing process, even though storage stability may be marginal. Proteins show a wide range of formulation-sensitive instability. Some proteins survive freezing with little or no

Table 1 Shifts in pH during nonequilibrium freezing with phosphate buffer systems

Concentration (mM)	Initial pH	Frozen pH	Δ pH
Sodium phosphate buffer			
100	7.5	4.1	-3.4
8	7.5	5.1	-2.4
Potassium phosphate buffer			
100	7.0	8.7	+1.7
100	5.5	8.6	+3.1
10	5.5	6.6	+1.1

(From Refs. 29 and 30.)

measurable loss in activity, whereas others are irreversibly deactivated by the freezing process (12–21, 31–33). An environment quite different from a dilute aqueous solution is created during freezing. All solute species are dramatically concentrated, ionic strength increases, the pH may shift, and above all, hydrophobic interactions that stabilize the native conformation in water are reduced or eliminated as bulk water is removed from the protein phase. In addition, just as proteins undergo thermal denaturation at elevated temperature, proteins also undergo spontaneous unfolding at very low temperature, called “cold denaturation” (34, 35). Estimated cold denaturation temperatures are often below the T_g' of a protein formulation and therefore of questionable relevance to freeze-drying. However, these estimates are based upon thermodynamic parameters measured in dilute aqueous solutions. The impact of perturbations caused by freeze concentration is largely unknown.

Because of these many freezing stresses that may decrease the free energy of denaturation, thermodynamic destabilization of the native conformation during freezing is not surprising. Indeed, it appears surprising that all proteins do not spontaneously unfold and degrade during freezing. However, for a destabilized protein to unfold and engage in subsequent irreversible reactions, the rate of unfolding must be fast relative to the timescale of freezing. Even in dilute aqueous solution at room temperature, protein unfolding may involve time constants of the order of hours (36, 37). In the freeze concentrate approaching T_g' , at temperatures $\approx 30^\circ\text{C}$ lower than room temperature and viscosities about 10 orders of magnitude greater than in a dilute aqueous solution, one would expect greatly reduced unfolding rates, perhaps sufficiently reduced to prevent degradation during freezing. Kinetic studies of unfolding near T_g' would resolve this question. Unfortunately, such data are not available. However, it has been argued on theoretical grounds that rate processes, in general, should be slowed greatly as the system temperature approaches a glass transition temperature (33). Experimental studies of three different reactions in frozen maltodextrin systems lend support to this view, although not all reactions show the same strong temperature dependence near a glass transition (38). As a general rule, it would seem prudent to minimize the time a protein spends in a freeze concentrate above T_g' (i.e., minimize freezing time), and primary drying should be carried out near or below T_g' . However, these precautions do not always guarantee stability.

Optimum freezing rate

Frequently, the freezing process is characterized by specification of freezing rate, where in most cases it

is really the temperature change of the heat sink that is specified, or at best, the cooling rate of the solution is given. However, there are at least two freezing parameters that are needed to define the freezing process—the degree of supercooling and the rate of ice crystallization. The degree of supercooling is the difference between the equilibrium freezing point and the temperature at which ice crystals first form in the sample. The degree of supercooling determines the number of nuclei and, therefore, determines the number of ice crystals formed in the sample. A high degree of supercooling produces a large number of ice crystals, and as the total amount of water that freezes is fixed, the ice crystals produced after completion of freezing are small in size. Size of ice crystals impacts on process design and may also affect product quality, as will be discussed later. The rate of ice crystal growth determines the residence time of the product in a freeze-concentrated fluid state. A rapid growth rate minimizes the residence time, normally allowing less degradation during the freezing process. In practical freeze-drying applications, heat transfer limits the rate of ice growth. Therefore, in vials and pans where heat removal is through the container bottom, rapid ice growth is facilitated by a small fill volume to container area ratio (i.e., small fill depth) and good contact between the container bottom and the freeze dryer shelf. In general, rapid ice growth is also promoted by a low shelf temperature. However, if a warm vial with a large fill depth is placed directly on a very cold shelf, a very high degree of supercooling may be produced near the vial bottom before convective mixing cools the upper portion of the solution. Ice then forms near the vial bottom whereas the remainder of the solution remains completely liquid. Freezing then continues slowly from bottom to top as a freezing front. Such a freezing pattern normally produces a dried cake with two distinct structures: a region of very fine pores near the bottom, with most of the cake having very large pores due to the very large ice crystals that formed during advance of the freezing front. The product may be perceived as “lacking in elegance.” At the least, a process designed to produce a fast freeze gives instead a slow freeze.

A high degree of supercooling produces small ice crystals and small pore size in the dried layer, which results in a high resistance to water vapor transport during primary drying (39). Consequently, long primary drying times result. Small ice crystals also mean a high specific surface area in the dried product, which decreases secondary drying time (40). These trends suggest that a moderate degree of supercooling is optimal. While the size of ice crystals would not normally be expected to impact on product quality, it is clear that if a protein were to

denature at the aqueous ice interface, small ice crystals would mean a large interfacial area and more denaturation, perhaps leading to increased aggregation during freezing. There is now clear evidence that, at least in some cases, proteins may indeed denature at the aqueous-ice interface (41, 42).

In general, the optimum freezing process produces moderate and uniform supercooling and fast ice growth. Uniformity within a given vial avoids heterogeneous cake structure, and uniformity between vials minimizes variation in drying rates and sample temperatures. The combination of moderate supercooling and fast ice growth is difficult to achieve as, in usual practice, the cooling rate of the heat sink (i.e., the shelf) is the only controllable factor in freezing. A general procedure that normally gives satisfactory results may be summarized as follows. First, the product load is cooled to a temperature below the equilibrium freezing point but above the temperature where experience has demonstrated that ice nucleation does not occur (i.e., typically $\approx -5^\circ\text{C}$), and the load is equilibrated at this temperature for a short time (i.e., 15–30 min). Next, the shelf temperature is decreased quickly toward the final temperature (i.e., to -40°C at $\approx 1^\circ\text{C}/\text{min}$). At least with a small fill depth (≈ 1 cm), ice nucleation usually occurs uniformly with moderate supercooling, yet ice growth proceeds relatively rapidly. When product temperature is monitored with temperature sensors in selected vials, the time at which all measured product temperatures are below a given temperature (i.e., time when the product is several degrees above the final shelf temperature) is recorded. Since vials containing temperature sensors often freeze sooner than the batch as a whole (43), the product load is allowed to soak for a fixed time (about an hour) to allow all vials to “catch up” and reach the desired temperature at which all product is in the solid state.

Primary Drying

Mass transfer resistance, product temperature, and drying rate

The effect of product temperature and mass transfer resistance on sublimation rate may be mathematically illustrated (39, 44) by expressing the sublimation rate per vial, dm/dt , in terms of the driving force for transport of water vapor from the ice-vapor interface to the chamber, $P_0 - P_c$,

$$\frac{dm}{dt} = \frac{P_0 - P_c}{R_p + R_s} \quad (1)$$

where P_0 is the equilibrium vapor pressure of ice at the temperature of the frozen product, P_c is the pressure in

the drying chamber, R_p is the resistance of the dried product layer, and R_s is the stopper resistance. As P_0 increases exponentially with temperature, it is obvious that the driving force for sublimation, and, therefore, also the sublimation rate, increases dramatically as the product temperature increases (about a factor of two per 5°C increase in temperature). The decrease in primary drying time with increasing product temperature is illustrated by Fig. 3. These data represent calculations based upon integration of Eq. 1 for a hypothetical (but typical) product at both a small (0.5 cm) and large (2.0 cm) fill depths. The value of chamber pressure used varies due to an attempt to maintain the condition $P_0 \gg P_c$, but yet employ chamber pressures near the optimum range for uniform heat transfer (≈ 0.1 – 0.2 Torr) when possible (44). At the lowest product temperatures, the chamber pressures selected are well below the optimum for uniform heat transfer, but much higher pressures would give very long drying times. At each temperature, drying time is roughly proportional to the square of the fill depth, and target temperatures below -40°C become impractical for 2-cm fill depths.

It should be noted that Eq. 1 assumes that the gas in the vial is essentially 100% water vapor. Both theoretical and experimental evidence indicate that, during most primary drying conditions of practical interest, the gas in the vial and the gas in the drying chamber is nearly all water vapor, even when the chamber pressure is being controlled by an inert gas leak (44). The molar flow rate of inert gas (i.e., N_2) leaked into the drying chamber is generally much smaller than the molar flow rate of water

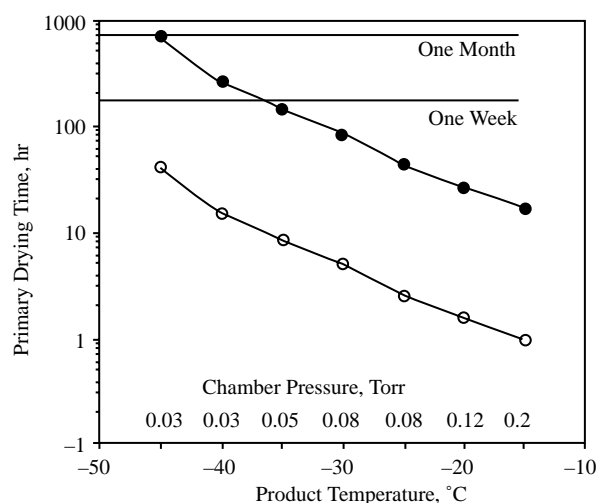


Fig. 3 Calculated primary drying times for a typical product as a function of product temperature. Open circles = 0.5-cm fill depth; filled circles = 2.0-cm fill depth.

vapor during primary drying, so the composition of gas in the drying chamber remains mostly water vapor. The inert gas “overwhelms” the vacuum pumps and, therefore, increases the total pressure in pumping system, which then causes a “back-up” of water vapor into the drying chamber.

Clearly, fast freeze-drying demands both high target product temperature and a small fill depth, which unfortunately is not always possible. To maintain product elegance, and in some cases to minimize degradation, primary drying must be carried out at product temperatures below the collapse temperature. As noted earlier, the collapse temperature and T_g' are closely related, with the collapse temperature normally being several degrees higher than T_g' . In practice, recognizing that not all vials freeze dry at exactly the same temperature, the target product temperature is chosen several degrees below the collapse temperature to provide a safety margin. A collapse temperature is a property of all components in the amorphous phase, and consequently the collapse temperature is highly formulation dependent (45, 46).

Although the product temperature is generally the most important factor in determining the rate of primary drying, product resistance is also an important parameter. The dried product resistance, R_p depends on the cross-sectional area of the product, A_p , by the relation $R_p = \hat{R}_p/A_p$, where \hat{R}_p is the area normalized product resistance, which is independent of the sample area but depends on both the nature of the product and the thickness of the dried product. Thus, R_p depends on the container used in that A_p is fixed by the internal diameter of the vial. The units used for \hat{R}_p (39, 44) are $\text{cm}^2 \text{ Torr hr g}^{-1}$. With this choice of units, the numerical value of \hat{R}_p represents the approximate time (in hours) to freeze dry a 1 cm thick sample at a temperature of -20°C , if the resistance would remain constant over the duration of primary drying. As one might expect, the resistance increases as the dry-layer increases in thickness, and therefore, resistance increases as primary drying proceeds. However, the relationship between resistance and thickness is not usually a direct proportion, although the resistance is often roughly linear in dry-layer thickness (39, 44).

Variation of product resistance with concentration is illustrated by the data in Fig. 4, where values of the mean product resistance, $\langle \hat{R}_p \rangle$, are given for various product types at different total solute concentrations. The mean product resistance is the mean resistance over a dry layer thickness interval of 0–1 cm. The scatter reflects formulation-specific effects whereas the general trend with increasing concentration illustrates the tendency of higher concentrations to produce higher resistance and therefore, longer drying times. A typical value for the

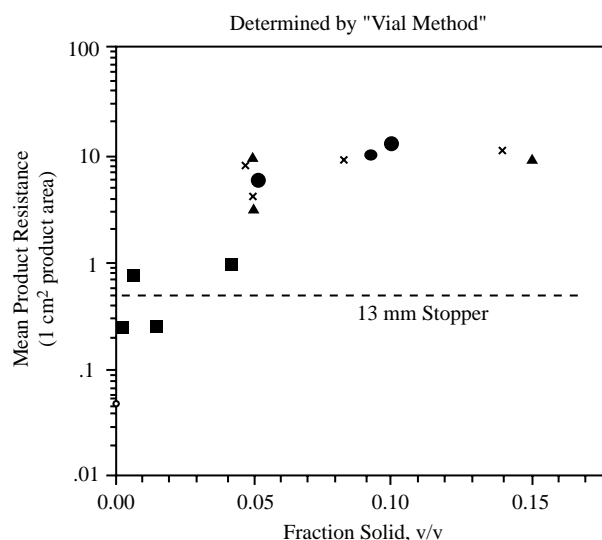


Fig. 4 Mean dry product resistance values for various products. Squares = proteins; filled circles = peptides; triangles = carbohydrates; open circle = pure ice; crosses (x) = miscellaneous types. (M.J. Pikal, Eli Lilly & Co., unpublished results.)

resistance of a small 13-mm finish stopper is shown as a straight line. Note that except for very dilute solutions, $R_p \gg R_s$.

Process control: Product temperature measurement

In most commercial freeze-drying processes, chamber pressure, shelf temperature, and time are the only controllable process parameters. Product temperature is not directly controlled. It is the balance between heat and mass transfer that determines the product temperature (47). Obviously, shelf temperature is important in determining the heat transfer and product temperature. However, because much of the heat is transferred through the gas phase (i.e., collisions of gas molecules with the hot shelf surface and the cold vial bottom), heat transfer as well as mass transfer (Eq. 1) is determined, in part, by the chamber pressure. Therefore, product temperature is determined by shelf temperature, chamber pressure, the heat-transfer characteristics of the vials, and the mass-transfer characteristics of the product and semistoppered vials.

Conventionally, product temperature is monitored in a small number of vials by placing temperature sensors at the bottom center of the vials. It is assumed that at least the average of the measured temperatures is representative of the rest of the vials. The trend in product temperature with time is often used to determine the end of primary drying. That is, when the measured product temperature shows a sharp increase near the anticipated end of primary drying,

and begins to approach the shelf temperature, it is assumed that all ice in that vial has been removed. Generally, for that vial, this assumption is correct. When all vials containing temperature sensors are judged dry by this criterion, one might be tempted to assume that all vials in the batch are dry. This assumption is not generally correct. Vials containing temperature sensors are not representative of the batch as a whole. There exists a temperature and drying rate bias between the monitored vials and the rest of the batch. Monitored vials usually freeze sooner with less supercooling than does the batch as a whole (43). This effect is particularly significant in production where due to the particle-free environment, the introduction of a temperature sensor introduces a significantly higher level of heterogeneous nucleation sites, thereby causing nucleation of ice at a lower temperature than in vials without temperature sensors. Observations made during a production run on moxalactam are illustrated in Fig. 5 (43). Monitored vials supercool to around -15°C and then begin freezing. The vials not monitored do not begin freezing until much later, at which time the product temperature must be close to the shelf temperature of -25°C . Thus, nonmonitored vials undergo a much greater degree of supercooling, have smaller ice crystals, smaller pores, more resistance to mass transfer, freeze dry at higher temperature, and require more time to dry than the rest of the vials. The temperature bias during primary drying is usually small, $\approx 1^{\circ}\text{C}$, but the drying time bias is 10% of the primary drying time, which can be quite significant for long primary drying times (43). To compensate for this bias in drying time, a soak period of 10–15% of the primary drying time is imposed following the time when all monitored vials are judged dry. Only after this soak period is the shelf temperature increased for

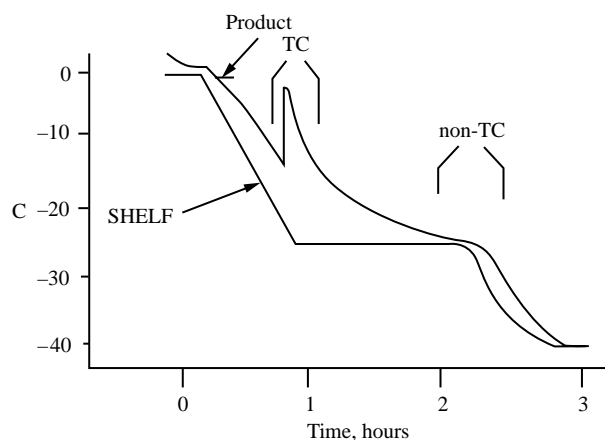


Fig. 5 Experimental observation of freezing bias: A production run of moxalactam di-sodium. (From Ref. 42.)

secondary drying. Premature increase of the shelf temperature carries a high risk of collapse. Thus, although monitoring the temperature of the product in vials does provide some information on the progress of primary drying, the information provided is far from perfect!

Location of the monitored vials in the vial array on the shelf is also an important factor in recording representative data. Vials on the edge of an array (i.e., facing the dryer door or walls) normally are not representative of the vials in the interior of an array, which are surrounded by other vials (44). Due to differences in radiative heat transfer, such vials normally dry faster at a higher temperature. With a shelf temperature around 5°C , temperature bias is about 1°C , with drying time bias about 10% (44), and the effect increases in magnitude as the difference between shelf temperature and ambient temperature increases and chamber pressure decreases. However, to minimize the risk of loss of sterility in surrounding vials, the temperature sensors are frequently placed in vials on the edge of the vial array facing the door. Again, the monitored vials are not representative of the batch as a whole. Clearly, minimizing the risk of product contamination with microorganisms is extremely important. From a sterility assurance viewpoint, the ultimate low-risk process is achieved with fully automatic loading systems, which, in principle, can eliminate the need for human presence in the sterile block except in emergency situations. However, it is difficult—perhaps impossible—to place temperature sensors in product vials when using an automatic loading system. Given the problems noted above with the routine use of product temperature sensors, it would seem reasonable to employ product temperature sensors only in development and validation, where vials near the center of an array could be monitored. With a robust process and suitable validation data, it may be assumed that production batches would run with the same product temperature history as the development and validation batches. Thus, product temperature would not be monitored during routine manufacturing. The simplest process design involves using a fixed shelf temperature:chamber pressure:time program.

Designing an *efficient* and *robust* process based upon a fixed shelf temperature:chamber pressure:time program is not a difficult assignment when the primary drying time is short. For example, if the primary drying time of an average batch is 6 h, designing for worst case of perhaps twice the primary drying time of an average batch would extend the average process time by only 6 h; however, if the primary drying time of an average batch is 4 days, designing for worst case will result in a significant reduction of production capacity. By using methodology

for remote sensing of the end point of primary drying (43), which will be discussed later, this process inefficiency can be virtually eliminated. Moreover, it is both possible and practical to monitor product temperature without placing temperature sensors in the product vials (48). This methodology, called manometric temperature measurement, is based on an analysis of the rate of pressure increase in the chamber when the valve separating the drying chamber from the condenser chamber is periodically quickly closed (and the pressure control system is deactivated) for a brief period of time (≈ 15 s). When the drying chamber is thus isolated from the condenser chamber, the chamber pressure increases due to four effects (Fig. 6): 1) sublimation at the ice–vapor interface at constant product temperature and transfer of water vapor through the dried cake; 2) dissipation of the temperature gradient in the frozen layer, thereby increasing the temperature at the ice–vapor interface; 3) heat flow from the shelf to the product, thereby increasing the temperature at the ice–vapor interface; and 4) natural air leaks from the surroundings into the chamber. The pressure rise data is fitted to a theoretical function to obtain the average product temperature at the ice–vapor interface. Effects 3 and 4 above both produce a linear increase in pressure with time whereas the other two effects are nonlinear. Effect 1 is dominant (Fig. 6), but a consideration of the other effects is necessary to obtain accurate product temperatures (48). Manometric temperature measurement gives

the temperature at the ice–vapor interface, which is the temperature relevant to collapse, and most important, gives a representative temperature of the product vials without risk of sterility compromise. The freeze dryer, however, must be computer-controlled so that product temperature data can be obtained in real time via mathematical analysis of the pressure rise data. Manometric temperature measurement can also be used to sense the endpoint of primary drying as well as providing accurate dried layer resistance data (48).

Process control and chamber pressure

Pressure control may be accomplished by one of three methods: controlled nitrogen leak, conductance control, or condenser temperature control. Pressure control by “controlled nitrogen leak” is accomplished by opening/closing a needle valve connected to a nitrogen source at atmospheric pressure in response to the deviation of the measured chamber pressure from the set point. Very fine pressure control, within ± 5 mTorr, can easily be achieved. Typically, sterile nitrogen is leaked into the drying chamber, but equally fine pressure control is achieved by leaking nitrogen into the vacuum line near the vacuum pumps. Although, to my knowledge, no relevant data exists, current dogma suggests that leaking into the drying chamber is preferred because the risk of sterility compromise is less. Conductance control is based upon modulating the resistance of the chamber to condenser pathway by partially closing/opening the valve separating the drying chamber from the condenser chamber (49). One disadvantage of this technique is the inability to control pressure during secondary drying. Once evolution of water vapor slows to very low rates (during secondary drying), the chamber pressure will reduce to whatever ultimate vacuum the system will produce that day. Although control of chamber pressure to regulate heat input is only necessary in primary drying, one could argue that a process that does not control pressure throughout the process is not fully reproducible. In general, this argument is without scientific foundation, but there are special circumstances where product contamination by adsorption of volatile stopper impurities (or other foreign vapors in the freeze dryer) may occur, and such contamination is more serious at very low chamber pressures (50). Here, one would want to control chamber pressure in secondary drying at a somewhat elevated level (≈ 200 mTorr). Chamber pressure can also be controlled by control of condenser temperature, provided the freeze dryer design provides for fine control of condenser temperature (49). Control of condenser temperature controls the vapor pressure of ice on the condenser, thereby controlling the partial pressure of water in the drying chamber. Due to

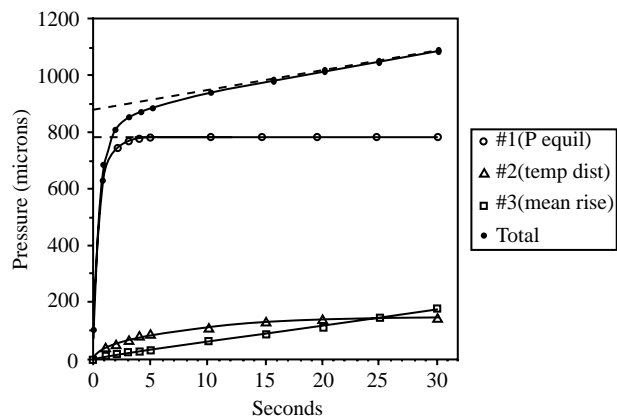


Fig. 6 Calculated contributions to the pressure rise in a manometric temperature measurement experiment. Calculations were made for a typical product with an initial ice temperature of -20°C , corresponding to an initial vapor pressure of 775 mTorr. Open circles = effect 1, sublimation; open triangles = effect 2, dissipation of temperature gradient; open squares = effect 3, heat flow from shelf to product; filled circles = sum of all effects. (Adapted from Ref. 48.)

natural air leaks and the finite pressure difference between condenser and chamber required for mass transfer, the controlled chamber pressure will be somewhat higher than the vapor pressure of ice on the condenser, perhaps 20–50 mTorr higher. Thus, to control chamber pressure at 200 mTorr, the condenser temperature would be controlled at a temperature around -35°C . With this mode of pressure control, the vapor in the drying chamber remains essentially 100% water vapor throughout the process, including secondary drying. Conversely, with pressure control by a nitrogen leak, the vapor changes from essentially 100% water vapor in primary drying to mostly nitrogen during secondary drying (Fig. 7). Thus, a process run with a nitrogen leak to provide pressure control is not necessarily the same as the corresponding process run with pressure control via control of condenser temperature. In general, the difference is not expected to be of practical significance. With a final product temperature of 25°C or greater in secondary drying, a chamber pressure of 200 mTorr pure water vapor produces a relative humidity of less than 0.8%, which is not high enough to impede secondary drying of most materials. However, chamber pressure control via condenser temperature control will not produce the change in gas composition noted in Fig. 7,

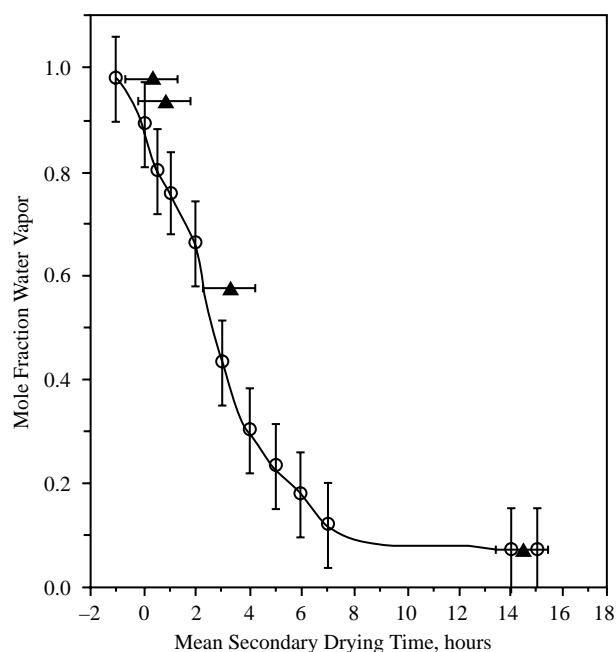


Fig. 7 Typical variation in gas composition in the drying chamber during secondary drying. Open circles: experimental values determined in a laboratory freeze dryer; filled triangles: theoretical values calculated from mass transfer theory and freeze dryer characteristics. (M.J. Pikal, Eli Lilly & Co., unpublished results.)

as the system passes from primary drying to secondary drying, which does limit the process control options for determining the end point of primary drying.

Process control: Effect of condenser performance

Depending upon dryer design and operating conditions, the condenser temperature may vary over a wide range. As long as the condenser temperature remains sufficiently low to allow control of the chamber pressure at the desired set point, the temperature of the condenser has no impact on the freeze-drying process (i.e., no change in any of the variables in Eq. 1). For example, if the chamber pressure is being controlled (via a nitrogen leak) at the target pressure of 0.10 Torr with a condenser temperature of -50°C , reduction of the condenser temperature to -70°C will have no effect on the process. Here, the reduced partial pressure of water at the condenser will be compensated by an increased partial pressure of nitrogen arising from an (automatic) increase in nitrogen leak rate. Although a detailed analysis of condenser performance is complex (49) and beyond the scope of this chapter, it should be noted that under conditions of very high sublimation rate, the condenser system may be overloaded, resulting in loss of pressure control (i.e., the chamber pressure increases beyond control). Unless the shelf temperature is sharply decreased, a “run-away” condition may develop with loss of product temperature control and ultimately, loss of the batch due to product collapse or ice melt. An overloaded condenser may reflect uneven ice build-up at the condenser plates caused by suboptimal design or operation, but may also arise from operation beyond the design capability of the refrigeration system (49). That is, the refrigeration system may not be able to remove heat from rapidly condensing water vapor (i.e., from very high sublimation rate) and yet maintain the condenser plate temperature low.

Process control: Determination of the end point of primary drying

Typically the shelf temperature setting used in primary drying is much lower than the shelf temperature required for efficient removal of residual water during secondary drying. Because an increase in shelf temperature before all vials have completed primary drying carries a high risk of product collapse, some indicator of the end of primary drying is needed for optimum process control. Of course, one can increase the temperature for secondary drying at a fixed time, but as discussed earlier, such a process is not an optimum one (i.e., the process is normally longer than necessary). Product temperature response is the most common indicator of the end of primary drying. Here, the time is noted when all product temperature sensors

approach the shelf temperature being used in primary drying. Next, a delay time or soak period is introduced to compensate for the fact that the vials containing temperature sensors are not typical of the batch as a whole. After this delay time, the shelf temperature is increased from the primary drying setting to the setting used for secondary drying. As determination of the appropriate delay time is difficult, the delay times used are often quite arbitrary. The principle problem is that freezing bias normally differs considerably between development and manufacturing. Thus, the optimum delay time for manufacturing cannot be determined in most development laboratories. It is obvious that use of product temperature sensors to determine the end point of primary drying is not entirely satisfactory. As suggested earlier, manometric temperature measurement may be used to determine the end point of primary drying. A far simpler method is to base the determination of the end point upon a real time measurement of the vapor composition in the freeze dryer chamber. As primary drying ends, and the process passes into secondary drying, the composition of the vapor in the drying chamber shifts from nearly pure water vapor to nearly pure nitrogen (Fig. 7), assuming chamber pressure is being controlled via a nitrogen gas leak. A sensitive and inexpensive measurement of vapor composition is provided by an electronic moisture sensor with output in dew point or partial pressure of water (43, 51). An electronic moisture sensor has the sensitivity to determine presence of residual ice in less than 1% of the vials (43), and under some conditions can also be employed to determine the end point of secondary drying—i.e., when the residual moisture has been reduced to below the target levels (51).

Secondary Drying: Desorption of Water from the Freeze Concentrate

Product temperature control and glass transitions

The concept that one must maintain the product temperature below the collapse temperature during primary drying is now well accepted. Exceeding the collapse temperature will cause loss of product elegance, which will vary from moderate deformation and shrinkage to a complete melt and deposition along the walls and bottom of the vial. Structural collapse also may occur during secondary drying or during storage of the dried product, particularly if the dried product contains high levels of residual moisture, and the storage temperature is high. If the product temperature exceeds its glass transition temperature, cake shrinkage and deformation will occur. Such an event is a less common problem during secondary

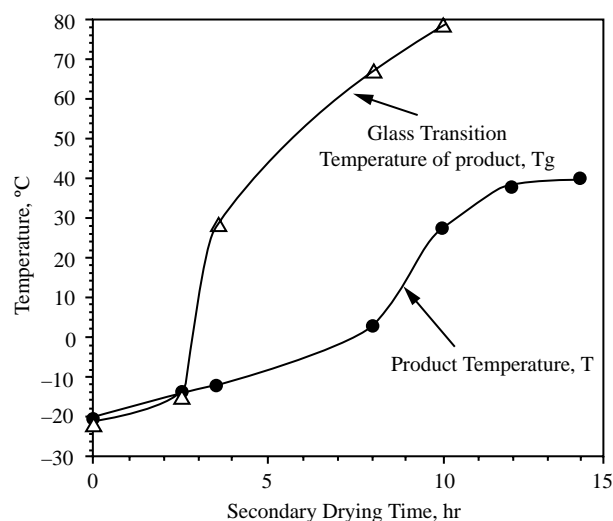


Fig. 8 Variation of glass transition temperature and product temperature for moxalactam di-sodium during secondary drying. (Calculated from data in Refs. 25 and 40.)

drying than is collapse during primary drying, largely because most secondary drying processes are extremely conservative during the early stage where a glass transition is most likely. As the water content of the amorphous phase decreases during drying, the glass transition temperature increases very sharply. Some representative data are shown in Fig. 8. If the shelf temperature remains at the setting used in primary drying for the first few hours of secondary drying, as is common practice, the glass transition temperature will nearly always rise much faster than will the product temperature. However, if one were to optimize secondary drying, and therefore, eliminate most of the dead time in early secondary drying, the potential for a glass transition would increase dramatically. Here, a knowledge of the glass transition temperature as a function of water content would be required for process optimization.

Optimum residual moisture

The optimum residual moisture for a given product must be established by empirical studies. Certainly, to eliminate the possibility of a structural collapse during storage, the water content needs to be low enough so that the glass-transition temperature is well above the highest temperature relevant to product distribution and storage. Moreover, in our experience with both small molecules and proteins, in-process degradation is generally relatively insensitive to the final moisture content, and storage stability normally improves as residual moisture decreases, although the relationship between stability and residual water is not

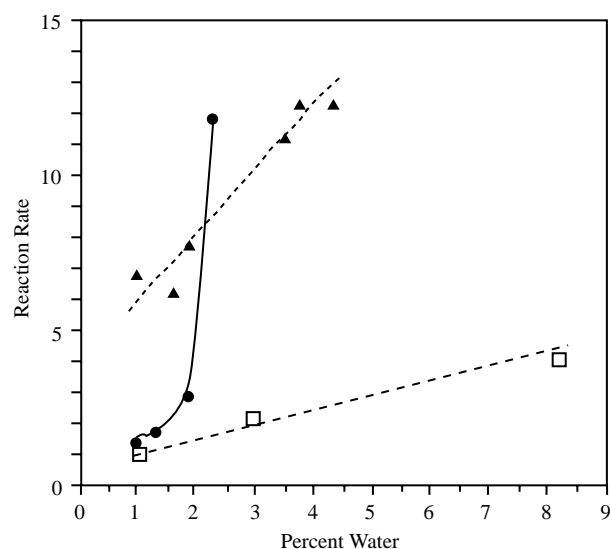


Fig. 9 The effect of residual water on the stability of freeze-dried proteins. Circles = human growth hormone deamidation and oxidation at 25°C, %/month (From Ref. 19); squares = aggregation of human serum albumin at 50°C, %/day (Ref. 52); triangles = hemoglobin oxidation at 23°C in a sucrose formulation, %/year (Ref. 53).

necessarily linear. Data for human growth hormone (hGH) formulated with glycine and mannitol (19), human serum albumin (52), and hemoglobin formulated with sucrose (53) illustrate the range of behavior normally encountered (Fig. 9). The rate of chemical degradation of hGH at 25°C (methionine oxidation and asparagine deamidation) increases in highly nonlinear fashion by nearly an order of magnitude as the moisture content varies from 1% to 2.5% (19); the rate of aggregation of human serum albumin increases linearly with increasing moisture content (52); and the rate of hemoglobin oxidation at room temperature doubles as the residual water content increases from 1% to 4% (53). Although the optimum moisture content for storage stability may be zero, the gain in stability between about 1% water and zero water content is often not sufficient to justify the additional processing complications in achieving and maintaining extremely low water contents. An exception is storage stability of a monoclonal antibody–vinca alkaloid conjugate formulation where stability is sensitive to very low levels of residual moisture (22).

The decrease in storage stability accompanied by an increase in residual moisture is often interpreted in terms of molecular mobility in the solid. That is, higher water content facilitates the mobility needed to support reactivity of the protein. One interpretation states that above monolayer levels of water, the protein has increased

conformational flexibility, and the additional water has the ability to mobilize water and other potential reactants in the amorphous protein phase (54). Both effects are expected to increase the rate of protein degradation. Alternatively, water plasticizes the amorphous phase, thereby lowering the glass transition temperature, T_g . Indeed, if sufficient water is present to lower T_g below the storage temperature, the amorphous phase would be in a fluid state, with greater mobility and greater reactivity than when stored below T_g in a glassy solid state. A system stored above its T_g would also suffer cake collapse and loss of pharmaceutical elegance. Although the monolayer and the glass transition interpretations are similar in that they both attribute the increased reactivity to water induced mobility in the amorphous phase, they differ in that a system above the monolayer level of water is not necessarily above the glass-transition temperature. Furthermore, as the level of water equivalent to monolayer coverage is essentially independent of temperature, the monolayer concept predicts that the trend in stability with water content is independent of temperature. For example, if monolayer coverage is equivalent to 10% water, protein reactivity would increase sharply above 10% water at all temperatures. Conversely, the glass transition interpretation states that the key stability variable is the difference between the glass transition temperature and the storage temperature, $T - T_g$. Here, reactivity increases sharply when $T > T_g$, so the sudden onset of reactivity depends on both T and the water content through the effect of water content on T_g .

Although most empirical and theoretical evidence suggest that storage stability improves monotonically as the water content decreases, we must also acknowledge the common assumption that there is a critical nonzero level of water that a protein requires for conformational stability in the freeze-dried solid, and therefore, some intermediate level of residual water would be optimum for stability. Although direct experimental evidence for this assumption is meager, some systems do have inferior stability when highly desiccated. Influenza virus is clearly less stable when the residual moisture deviates from the optimum 1.7% (55), but the direct relevance of this observation to protein formulations is uncertain. Aggregation of excipient-free tissue plasminogen activator at high temperature is faster at low water content (56), and aggregation of excipient-free hGH during storage at 25°C is faster in samples that, at some time in their history, have been highly desiccated (19).

Factors impacting drying rate

Water removal during secondary drying involves diffusion of water in the glassy solid, evaporation at the solid–vapor

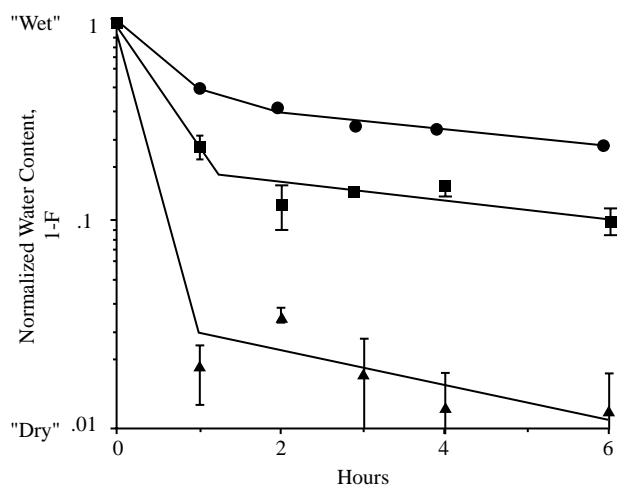


Fig. 10 Examples of the kinetics of secondary drying. Triangles = mannitol (crystalline); squares = poly (vinylpyrrolidone); circles = moxalactam di-sodium (amorphous). All solids were prepared by freeze-drying a 5% aqueous solution from a 1-cm fill depth, followed by hydration to a uniform moisture level of $\approx 7\%$. The quantity, F , is the fractional attainment of equilibrium, which corresponds to near zero water content. The secondary drying conditions were: product temperature = 18°C ; chamber pressure = 200 mTorr. (From Ref. 40.)

boundary, and flow through the pore structure of the dried cake. During the early portion of secondary drying, water content is high, and the drying rate is high in spite of the relatively low temperature of the solid (40). An illustration of secondary drying kinetics under isothermal conditions is given in Fig. 10 (40) for samples initially about 7% water. Mannitol is crystalline whereas povidone and moxalactam are 100% amorphous. Even at the start of the experiment, the glass transition temperatures of povidone and moxalactam are well above the sample temperature of 18°C . Thus, water is being removed from either an essentially crystalline solid (mannitol) or glassy amorphous solids (povidone and moxalactam). The symbol F represents the fractional attainment of equilibrium, which in these experiments is near zero water content. Thus, $1-F$ represents a normalized water content with a value of unity representing no progress in drying. Although the quantitative aspects of drying behavior are obviously specific to the product, in each case, the water content is reduced sharply during the first 1–2 h, followed by a period of much lower drying rate. If the drying rate were directly proportional to the residual water content, drying kinetics would be first order, and the curves in Fig. 10 would be straight lines. A simple diffusion model based upon constant diffusion constant and constant area

for diffusion would also predict linear semilog drying curves (40). Obviously, the curves are not straight lines. The residual water content appears to approach a plateau level, which is specific to the product. For the mannitol data, the plateau level of residual water is very low (≈ 0.1 – 0.2%) and probably represents occluded water in the crystalline mannitol. For the amorphous samples, povidone and moxalactam, the plateau levels are quite high, $\approx 1\%$ for povidone and $\approx 3\%$ for moxalactam (i.e., a significant fraction of the water appears to be bound). However, calculations based upon equilibrium water desorption isotherms and the measured partial pressures of water in the drying chamber demonstrate that the residual water present at the end of 6 h is far above the equilibrium water content (40). Thus, the residual water is *not* bound in a thermodynamic sense, but rather is kinetically trapped, a result, at least in part, of decreasing effective surface area for desorption of water as the smaller more rapidly drying solid particles are dried (40). For a given product, the plateau level of water decreases as the drying temperature is increased, decreases as the specific surface area of the solid increases, but is relatively insensitive to the thickness of the dried cake (40).

Traditional freeze-drying practice has often used very low chamber pressures during secondary drying, presumably in the belief that the rate of secondary drying would be accelerated by the use of the low pressures. However, it is found (40) that the rate of secondary drying is insensitive to chamber pressure, at least with pressures in the range of 0–0.2 Torr (40). This empirical observation is consistent with the conclusion that the rate-limiting mass transfer process for drying an amorphous solid is either diffusion in the solid or evaporation at the solid–vapor boundary, probably the latter (40). As the use of very low pressures in secondary drying may facilitate contamination of the product by adsorption of impurity gases from the stopper or other sources (50), it is clear that relatively high chamber pressures (≈ 0.1 – 0.2 Torr) should generally be used for secondary drying. Of course, extremely high chamber pressures (≈ 1 Torr) should perhaps be avoided. With very high chamber pressures, flow of water vapor through the pore structure would become rate limiting, and the very high pressure would then decrease the rate of drying.

Changes in moisture content during storage

Often one employs a heroic secondary drying process to reduce the residual water content to very low levels only to find that the moisture content increases during storage. Although moisture transfer from ambient through the stopper is possible, the increase in moisture content is most

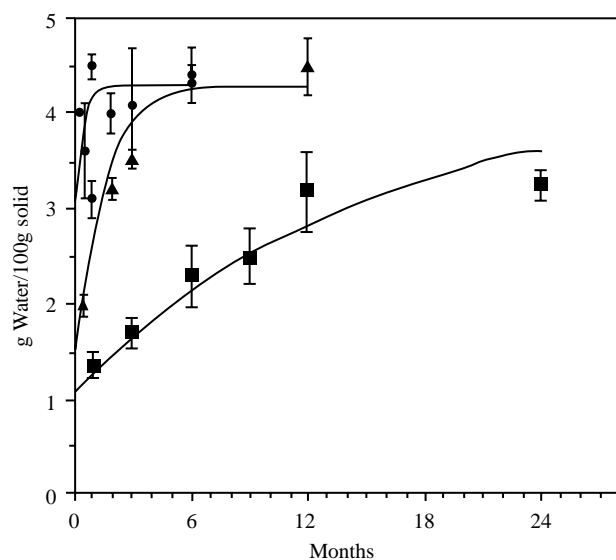


Fig. 11 Kinetics of water transfer from stoppers to 25 mg freeze dried lactose. The stoppers were 13 mm finish West 1816 gray butyl stoppers that were steam sterilized and vacuum dried for 1 hour. Circles = 40°C; triangles = 25°C; squares = 5°C. (From Ref. 57.)

often related to moisture release by the stopper (57). This phenomenon is illustrated in Fig. 11, where the time dependence of moisture content in 25 mg of freeze-dried lactose is given. Here, the stoppers were gray butyl stoppers that had been previously steam sterilized, followed by 1 h of vacuum drying to remove excess water. The symbols represent experimental data for 40°C, 25°C, and 5°C storage, and the smooth curves represent the best fit to a theoretical model (57). It should be noted that at least at 40°C and 25°C storage, the moisture content increases sharply from 1%, but then reaches a plateau value of about 4.3%, suggesting that equilibrium has been reached. That is, at the plateau value of 4.3% water, the activity of the water in the stopper is equal to the activity of water in the lactose and mass transfer ceases. In other studies, we have noted samples deliberately prepared to be initially of high water content have lost water during storage. In short, there is moisture equilibration between moisture in the stopper and moisture in the product that may either increase the moisture content of a dry product or decrease the moisture content of a wet product. The plateau or equilibrium level of water appears to be independent of temperature. The equilibrium values for 40°C storage and 25°C storage are identical, and the moisture content values for 5°C storage appear to be approaching the same equilibrium value of 4.3%. The rate of approach to equilibrium, however, is strongly temperature dependent. The equilibrium level of water is

strongly dependent upon the stopper treatment history. Stoppers vacuum dried for 8 h after steam sterilization allowed only a small increase in moisture content during storage. The equilibrium water content is lower for a higher mass of solid, and is slightly higher for a more hygroscopic product (57).

The moisture exchange between stopper and product may have important stability consequences. Absorption of moisture from the stopper may result in a product which has, after time, a moisture content high enough so that the glass transition temperature is below the storage temperature. Structural collapse will result, and stability of the product may well be compromised. Moreover, because the glass transition temperature often changes dramatically in a narrow range of water content, the onset of instability could be quite sudden. Moisture exchange may be moderated by extensive high temperature vacuum drying of the stoppers after steam sterilization. A better solution is to employ alternate rubber formulations that are less prone to release water to the product. Such stopper formulations are now available from several vendors.

FORMULATION AND STABILITY

Function of Excipients

Most freeze-dried products contain several components in addition to the drug or active component. These additional components, called excipients, are intended to serve a specific function, normally related to stability or process, and may constitute the major fraction of the freeze-dried solid.

With some products, the quantity of drug per vial is extremely small. Here, bulking agents such as mannitol or glycine are used to provide product elegance (i.e., satisfactory appearance) as well as to provide sufficient cake mechanical strength to avoid product blow-out. When a very dilute solution is freeze dried, the flow of water vapor during primary drying may generate sufficient force on the fragile cake to break the cake structure and carry some of the product out of the vial with the water vapor stream. Product blow-out normally occurs only when the solution to be freeze dried is very low in total solids (1% or less). Thus, with low dose drugs, a bulking agent may be a critical formulation component.

Collapse temperature modifiers are excipients that will increase the collapse temperature to allow more efficient freeze-drying without collapse. Such materials may also function as bulking agents and/or stabilizers. Dextran ($T_c = -10^\circ\text{C}$), hydroxyethyl starch ($T_c = -10^\circ\text{C}$), ficoll ($T_c = -20^\circ\text{C}$), gelatin ($T_c = -8^\circ\text{C}$) are examples.

Although none of these materials are commonly used in parenteral formulations, dextran and hydroxyethyl starch are used in large quantities as IV therapeutic agents, and therefore, would presumably be acceptable as excipients. Usually, the collapse temperature of a mixture is intermediate between the collapse temperatures of the individual components (58), but collapse temperatures of candidate formulations cannot be predicted with high accuracy.

Macroscopic collapse may often be avoided by use of a crystalline matrix component. Here a readily crystallizable component is added at a relatively high level (i.e., more than 50% of total solids and ideally much more) such that a crystalline matrix is formed. Freeze drying such a formulation amounts to freeze-drying with microscopic or partial collapse (i.e., complete collapse of the amorphous phase), but cake structure is maintained by the crystalline component. Thus, both elegance and good reconstitution time will be maintained, and the samples will normally dry to low residual water with ease. However, if collapse of the amorphous phase containing the protein leads to degradation, this method for circumventing collapse is not viable. Glycine and mannitol are commonly used as crystalline matrix components. This excipient function is, in reality, a special case of the use of a bulking agent.

Particularly with proteins, excipients are often added to prevent degradation, as for example, lyoprotectants, antioxidants, nonionic surfactants, metal ion chelators, and other proteins such as BSA in diagnostic products. In many cases, the stabilizer may also serve as a bulking agent. Glycine, mannitol, sucrose, and lactose are perhaps the most commonly used stabilizers. However, as lactose is a reducing sugar and commonly reacts with proteins, its use must be questioned. Both mannitol and glycine tend to crystallize, and are, therefore, generally poor choices as stabilizers when used alone. Sugars, in particular sucrose, are often effective lyoprotectants, as well as enhancers of storage stability.

Buffers are often added to control pH, but caution must be exercised when buffers are used in a formulation to be freeze dried. As discussed earlier, crystallization of either buffer component (acid or base) during freezing may cause a significant pH shift during freezing, thereby causing greater pH variation than would have been obtained in an unbuffered system.

Products intended for human use are occasionally formulated with NaCl or glycerol to make the reconstituted product isotonic. This practice is normally not a requirement for IV drugs, but can be quite important in minimizing pain on injection of subcutaneous and IM doses. In general, and in particular with proteins, isotonic

adjustment is best accomplished by including the tonicity modifier in the diluent rather than in the freeze-dried product. Sodium chloride and glycerol can lower the collapse temperature significantly, and sodium chloride may cause aggregation of the protein during freeze drying.

With multidose products, there is a need for use of a preservative to prevent microbial growth during the period of product use. Mixtures of ethyl- and methyl-parabens are a common choice as are phenol and *m*-Cresol. As preservatives are used at extremely low levels (i.e., $\leq 0.1\%$ w/w in solution), they normally would not alter the collapse characteristics of the formulation. However, since the preservative is not needed during the freeze-drying process, to keep the formulation for freeze-drying as simple as possible, preservatives are best introduced via the diluent intended for reconstitution.

Surfactants may be added at low levels (i.e., $\approx 0.05\%$ w/w in solution) for several purposes. Surfactants may aid reconstitution if the drug does wet well, and surfactants are often added to low dose products to minimize losses due to surface adsorption. Surfactants may also be effective as stabilizers in low dose protein systems.

Relationships Between Formulation and Process

Particularly for freeze-dried products, formulation and process are interrelated. Properties of the formulation, in particular the collapse temperature, will have a significant impact on the ease of processing. An efficient process is one that runs at a high product temperature. However, the temperature cannot be too high or product quality will be compromised. As the glass transition temperature depends on chemical composition of the amorphous phase, T_g' and collapse temperature are strongly formulation dependent. Collapse temperatures for common excipient systems vary from less than -50°C to around -10°C (Table 2).

The collapse temperature depends upon the composition of the amorphous phase, and crystallization of one or more components may significantly alter the collapse temperature. In this way, process variations that induce crystallization may alter the physical state of the formulation. For example, human growth hormone (hGH) formulated with glycine and mannitol in a hGH:glycine:mannitol weight ratio of 1:1:5 may form a completely amorphous system if frozen very quickly. Here, the collapse temperature is -24°C . However, a slower freeze allows crystallization of most of the mannitol, resulting in a collapse temperature greater than -5°C (18). Another example is provided by the glycine:sucrose system (64–68) where glycine is present in excess of the sucrose. If frozen quickly, glycine remains

Table 2 Collapse temperature [T_c (°C)] and Glass Transition [T_g' (°C)] data for selected excipients^a

Material	T_g' (°C)	Reference	T_c (°C)	Reference
Polymers				
BSA	-11	46		
Dextran	-10	46, 59	-10	60, 61
Ficoll	-19	59	-20	45
Gelatin	-9	59	-8	45
PVP (40k)	-20	59	-23	45
Saccharides and polyols				
Dextrose	-44	59		
Hydroxypropyl β -cyclodextrin			-18	60
Lactose	-28	46, 59	-31	60, 61
Mannitol	-35, -28	46, 59		
Raffinose	-27	62	-26	45
Sorbitol	-46	59	-45	45
Sucrose	-32, -35	46, 63	-34, -32	60, 61
Trehalose	-27, -29	46, 59	-34	60
Amino acids				
β -Alanine	-65	46		
Glycine	(-62)	64		
Histidine	-33	46		
Salts and buffer components				
Acetate, potassium	-76	46		
Acetate, sodium	-64	46		
CaCl ₂	-95	46		
Citric acid	-54	46		
Citrate, potassium	-62	46		
Citrate, sodium	-41	46		
HEPES	-63	46		
NaHCO ₃	-52	46		
Phosphate, KH ₂ PO ₄	-55	46		
Phosphate, K ₂ HPO ₄	-65	46		
Phosphate, NaH ₂ PO ₄	-45	46		
Tris base	-51	46		
Tris HCl	-65	46		
Tris acetate	-54	46		
ZnCl ₂	-88	46		

^aCollapse temperature data were obtained with freeze-drying microscopy and T_g' data were obtained by DSC at roughly 10°C/min heating rates and represent mid-points of the glass transition region. Molecular weight is given for polymers when the data are sensitive to molecular weight. When significant differences exist between laboratories, both values are given. Values in parenthesis were estimated by extrapolation from noncrystallizing mixtures to the pure compound.

amorphous, and the glycine:sucrose freeze concentrate has a very low collapse temperature (i.e., roughly -45°C, depending upon the exact composition). However, if glycine is allowed to crystallize, the glass-transition temperature is essentially that of a sucrose freeze concentrate (i.e., about -34°C). Further, as the structure is maintained by the crystalline glycine, exceeding T_g' does not result in macroscopic collapse, and the system may be freeze dried without apparent collapse even at temperatures exceeding -15°C. Thus, crystallization can transform a formulation from one that is nearly

impossible to freeze dry in a commercial operation to one where freeze-drying is relatively easy. Clearly, if both glycine and sucrose are included in the formulation, the level of glycine must be either very low or very high relative to sucrose. If the level of glycine in the sucrose phase is very low, glycine will not crystallize but the impact on T_g' will be minimal. If the level of glycine is high, glycine can be induced to crystallize nearly completely (66, 68), thereby minimizing the level of glycine in the amorphous phase and maximizing the T_g' of the formulation.

The glass-transition temperature of a multicomponent amorphous system may be estimated from glass transition temperatures of the individual amorphous components (69). The simplest expression, commonly referred to as the Fox equation, is

$$\frac{1}{T_g} = \frac{w_1}{T_{g1}} + \frac{w_2}{T_{g2}} \quad (2)$$

where w_i is the weight fraction of component i and T_{gi} is the glass transition temperature of pure component i . Generalization of Eq. 2 to systems of more than two components is obvious. The effect of a second solute component on a formulation may be roughly calculated from Eq. 1 if T_{gi} in Eq. 2 is identified with T_g' of aqueous component i , and w_i are weight fractions of solute relative to the total mass of solute. Although Eq. 2 does not strictly apply to a freeze concentrate containing two (or more) solute components and water, such calculations from T_g' data are sufficiently accurate to be of practical use. For example, the effect of glycine on T_g' of aqueous sucrose systems (65) is fully consistent with Eq. 2.

Bulking Agents

General considerations

Bulking agents are intended to be inert and simply function as fillers to increase the density of the product cake. Product elegance is improved and product blow-out is prevented, but the bulking agent is not intended to provide enhanced chemical or physical stability of the drug substance. Amorphous excipients can function as bulking agents, but most amorphous excipients have relatively low collapse temperatures and therefore require low drying temperatures and long processing times. For example, although lactose has been used as a bulking agent, the relatively low collapse temperature (-31°C) requires long processing times. In addition, lactose is a reducing sugar and will chemically react with amine functionality, and therefore, cannot be used with many drugs. Although not in common use as a freeze-drying excipient, hydroxyethylstarch is an example of an inert amorphous additive with a high collapse temperature and therefore, could function as a bulking agent without requiring long processing times. However, hydroxyethylstarch tends to undergo some cake shrinkage and cracking during drying and therefore, provides a less elegant product than do the common crystalline bulking agents, glycine and mannitol. Since sorbitol is simply an isomer of mannitol, one might expect sorbitol could serve as a bulking agent. However, sorbitol does not easily crystallize during freeze-drying, and amorphous sorbitol has a

collapse temperature around -45°C (Table 2), thereby preventing its use at high levels in formulations for freeze-drying. Human serum albumin (HSA) or in diagnostic products, bovine serum albumin (BSA), have seen use as bulking agents and/or stabilizers. However, given current concerns regarding excipients isolated from human sources, use of HSA in new pharmaceutical products should be questioned. Mannitol is by far the most commonly used bulking agent. A mannitol-based formulation is elegant, reconstitutes quickly, and except for the potential of vial breakage (26, 70, 71), is generally easy to freeze dry without risk of product defects. Vial breakage is minimized by small fill depths, lower mannitol concentration, slow freezing, and avoiding freezing temperatures less than about -25°C until crystallization is complete. Glycine functions well as a bulking agent. Glycine crystallizes easily to form an elegant product that reconstitutes quickly and does not induce vial breakage. However, a glycine cake is more fragile than a mannitol cake, and is generally perceived as being somewhat less elegant than a mannitol cake.

Polymorphism in crystalline bulking agents

In general, the polymorphism issue of greatest significance in freeze-drying is whether or not a given component is crystalline or amorphous. Which crystalline polymorph forms is usually of secondary importance. However, it should be noted that both mannitol and glycine do exhibit crystalline polymorphism in freeze-dried systems (72–75). The only crystal polymorphism issue of known significance is the formation of a hydrate of mannitol (73, 75). Under some conditions, not fully understood, mannitol forms a hydrate that does not easily desolvate to an anhydrate during secondary drying. Secondary drying at temperatures in excess of 50°C appear to be necessary to desolvate the hydrate. Thus, samples of freeze-dried mannitol may have high residual water content if secondary drying temperatures were moderate. If this residual water were to remain bound in the hydrate crystal lattice, the high residual water content would simply be a curiosity. However, during storage, particularly at elevated temperature, the hydrate slowly desolvates, thereby releasing water to the amorphous drug phase. Such a scenario may compromise storage stability, particularly during accelerated stability tests.

Conditions for crystallization of bulking agents

A number of factors are critical for crystallization of the bulking agent. First, the nature of the bulking agent is of obvious importance. Unless the bulking agent will readily crystallize from an aqueous system at ambient temperatures, crystallization during freezing is unlikely. Although

both mannitol and glycine do readily crystallize during freezing, there are many other solutes that readily crystallize and are at least potentially acceptable as excipients for parenteral drugs. Mannitol and glycine are simply the most commonly used bulking agents. Crystallization is favored by higher concentrations of crystallizable solute, and perhaps most important, the bulking agent must be the major solute component to obtain reliable crystallization. The presence of high levels of other solutes, particularly those that remain amorphous, will generally impede or prevent crystallization of the bulking agent. As a general formulation rule, one should employ the crystallizable bulking agent at a concentration (weight percent) that is at least a factor of three greater than the sum of the concentrations of all other solute components. In addition, because resistance of the dry layer to flux of water vapor increases as the concentration of total solids increases, thereby prolonging primary drying, one should avoid total solids concentrations much above 100 mg/ml when possible. A rigid adherence to both of these rules would suggest that one would not employ a bulking agent if the drug concentration were much in excess of 25 mg/ml. Indeed, because a bulking agent is normally not needed to avoid product blow-out at drug concentrations above 25 mg/ml, one should question the use of bulking agents in such applications.

Surfactants

Generally, surfactants are employed to reduce surface adsorption losses of low-dose drugs, to improve wetting and reconstitution behavior, and to stabilize proteins during freezing. In nearly all applications, it is a nonionic surfactant that is used, and by far the most commonly used surfactant is polysorbate 80. However, it should be noted that the popularity of polysorbate 80 is due more to its extensive history of use and presumed greater acceptance by regulatory agencies than to its demonstrated superior performance in a given application. In practice, the surfactant level is very low, typically 0.01–0.1% w/v.

Proteins and surface active drugs often adsorb on filters, solution-processing equipment, and container surfaces, thereby producing loss of drug. Such losses are normally only of practical significance if the drug concentration is very low (i.e., when the total amount of drug loss is a significant percentage of the drug in solution). Addition of a low concentration of surfactant will frequently reduce the level of drug adsorption simply because the surfactant is preferentially adsorbed at the surfaces. With proteins, surface-adsorption effects may be considerably more

complex and damaging. Many proteins adsorb at interfaces, particularly the air–water interface, suffer a conformational change, desorb from the interface, and either refold or combine with other conformationally altered molecules, thereby leading to irreversible aggregation.

Human growth hormone aggregates extensively during shaking an aqueous solution, but this aggregation can be nearly eliminated by adding Polysorbate 20 to the formulation (76). It should be noted that although other surfactants (including polysorbate 80) were studied, Polysorbate 20 was the most effective stabilizer. In this study, all surfactant levels studied were above the surfactant critical micelle concentration (CMC), but the rate of aggregation decreases linearly as a function of increasing polysorbate 20:hGH mole ratio until it exceeds 2.0. This linearity provides a clue to the stabilization mechanism. As surfactant adsorption to the air–water interface would be expected to depend on the surfactant monomer concentration, and the monomer concentration is constant above the CMC, a stabilization mechanism involving adsorption of surfactant to the air–water interface would be expected to produce a roughly constant stabilization effect at all surfactant concentrations above the CMC. The data are not consistent with this prediction. However, it is known that polysorbate 20 binds to hGH with a stoichiometry of about 2.5 to 4 (77). Thus, it seems plausible that stabilization during shaking involves binding of surfactant to the surface of the protein (77).

Aggregation of hemoglobin during freeze–thaw is essentially eliminated by addition of polysorbate 80 at concentrations from 0.0125% to 0.1% (78). Without polysorbate 80, particle formation after five freeze–thaw cycles increased by more than an order of magnitude when freezing to -20°C and by about a factor of five when freezing to -80°C . However, use of polysorbate 80 reduced particulate generation at both freezing temperatures to a level where particulate levels after freeze–thaw were not significantly different than before freezing. Attempts to demonstrate binding of the surfactant to hemoglobin were unsuccessful, suggesting that the stabilization mechanism does not involve binding to the protein. Rather, it was proposed that stabilization involves surfactant adsorption at the interface, thereby preventing the protein from adsorbing at interfaces.

Although one cannot always expect protein aggregation to be eliminated as effectively by addition of surfactant as demonstrated in the above examples, it is certainly prudent to test the effect of surfactants in situations where protein aggregation is a problem during freeze–thaw or during solution handling.

Stabilizers

Types of stabilizers

A stabilizer is simply a formulation component without physiological effect that is added to the formulation to maintain the physical or chemical stability of the drug substance. As discussed in the previous section, the function of a surfactant may be to stabilize, and as control of pH is often critical to stability, one might consider buffers to be stabilizers. However, in this section we will address stabilizers other than surfactants or buffers. Further, as most of the stabilization literature deals with proteins, our discussion will focus on proteins, although in principle many stabilization principles apply to small molecules as well.

Minimizing oxidation

Although oxidation is a very common degradation pathway in pharmaceutical systems, one might expect that oxidation problems in freeze-dried products would be easily solved by elimination of oxygen from the system during processing; that is, because the product is dried in a vacuum environment and the vials either sealed in vacuum or in an atmosphere of an inert gas (i.e., nitrogen), no oxygen will be present to support an oxidation reaction. Indeed, storage stability of human growth hormone is greatly improved in vials back-filled with nitrogen compared to vials back-filled with oxygen (19). However, significant oxidation was also observed in samples back-filled with nitrogen. In practice, complete elimination of oxygen from the solution being processed is unlikely, and some oxygen may well be trapped in the amorphous solute phase after freezing and not be completely removed during drying. Further, due to a variety of causes including diffusion from the atmosphere over time, oxygen content in the vial headspace may easily reach levels of about 1%. Even 1% oxygen in the vial headspace is sufficient, from a stoichiometric viewpoint, to produce substantial decomposition, particularly with high-molecular-weight drugs. For example, with a 5-mg dose of human growth hormone ($M = 22.5$ kD) in a 5-cc vial, 1% oxygen in the headspace represents a factor of ten excess of the oxygen required for complete oxidation of the protein.

As ground-state molecular oxygen is not particularly reactive, presence of oxygen is not the only requirement for a rapid oxidation reaction. Activation of molecular oxygen to more reactive species requires light (i.e., photoactivation to singlet oxygen, $^1\text{O}_2$), or presence of a reducing agent and trace levels of transition metal ions (i.e., iron and/or copper), which can then convert molecular oxygen into more reactive oxidizing agents

such as superoxide radical ($\text{O}_2^{\cdot-}$), hydroxyl radical ($\cdot\text{OH}$), or hydrogen peroxide (H_2O_2) (79). Transition metal ions are often present in excipients, and processing in stainless steel equipment can lead to significant iron contamination. As the role of the transition metal ion is catalytic, only trace levels are required. The reducing agent is consumed in the conversion of molecular oxygen into reactive oxygen species, so higher levels of reducing agent are required for significant degradation of a drug. However, higher levels could still represent contamination from impurities in the drug and/or excipients. In the example of 5 mg hGH, a low-molecular-weight reducing agent ($M \approx 100$) present at a level of 0.1% w/w of the amount of hGH could lead to oxidation of more than 10% of the protein. The reducing agent could also originate from a misguided attempt to suppress oxidation by addition of an antioxidant such as ascorbate. Ascorbate may function as an antioxidant in some circumstances but will also function as an effective prooxidant and reduce molecular oxygen to reactive oxygen species in the presence of transition metal ions such as iron or copper (79, 80). Finally, peroxides are common contaminants in polyethylene glycols and nonionic surfactants and can serve as the oxidant. Thus, an attempt to stabilize protein conformation by addition of these materials may well chemically destabilize if oxidation is a possible degradation pathway.

A number of amino acid residues are subject to oxidation. Metal-catalyzed oxidation of methionine to methionine sulfoxide and other products is perhaps the most common pathway. Oxidation of cysteine to form either nonnative intra- or intermolecular disulfide bonds is also common, and histidine residues are also easily oxidized in metal-catalyzed pathways, with tryptophan and tyrosine being degraded by light-catalyzed oxidations (79). Mechanisms are complex, and even with a given reaction such as methionine oxidation, the active oxygen species and reaction product(s) varies with the experimental conditions (79, 80); thus, stabilization is difficult. Certainly, oxygen content in the solution being processed should be minimized, and the product should be sealed in the freeze dryer with either vacuum or nitrogen in the headspace. Due to the negative impact of even small amounts of transition metal ions, prooxidants, and peroxides, the bulk drug substance and excipients need to have rigid specifications. Formulation pH and the type of buffer salt may also be important (80). The optimum conditions vary with the application, and some empirical screening experiments are necessary to optimize a given formulation.

Specific chemical components may also be added in an attempt to retard oxidation. The classical solution, addition

of an antioxidant, must be approached with caution as the antioxidant may function as a prooxidant in a metal-catalyzed conversion of molecular oxygen into reactive oxygen species. Likewise, addition of a metal-complexing agent such as EDTA does not always retard oxidation. In fact, oxidation may be accelerated by complexing the transition metal (80). With most metal-catalyzed oxidations, the oxidation mechanism is a site-specific one, where the reduced form of the metal binds to a residue such as histidine on the protein and converts oxygen to reactive oxygen species at that site, which in turn oxidizes another residue nearby (81). In cases where complexation with EDTA facilitates oxidation, it is assumed that the EDTA-metal complex binds more strongly to the protein and is therefore more effective in generation of reactive oxygen species near a reactive residue (81). However, addition of EDTA does retard oxidation in some cases (82), and it would seem prudent to screen several complexing agents for impact on oxidation. Scavengers for reactive oxygen species, hydroxyl radical, $\cdot\text{OH}$, and singlet oxygen, $^1\text{O}_2$, may also retard oxidation (79, 83), but the results are highly specific to the system. For example, for iron-catalyzed oxidation of a methionine-containing peptide, thiourea was found effective for both ascorbate and dithiothreitol prooxidant systems but mannitol was effective only for the dithiothreitol prooxidant system (83).

Stabilization during freezing

In the context of proteins, stability has two distinct meanings. The term pharmaceutical stability refers to the ability of a protein to be processed, distributed, and used without irreversible change in primary structure, conformation, or state of aggregation. We refer to pharmaceutical instability as degradation. However, the phrase protein stability is also commonly used to describe the position of the equilibrium between native and unfolded conformations. If a protein formulation requires a high level of chemical denaturant, or a high temperature, to shift the equilibrium between native and unfolded in favor of the unfolded state, the protein is said to be stable. This meaning of stability I call "thermodynamic stability." Thermodynamic instability involves physical changes, somewhat analogous to a thermodynamic change of state. Pharmaceutical instability may be purely a result of a physical change (i.e., noncovalent aggregation), but may also involve changes in primary structure or, in other words, chemical degradation. Certainly, chemical transformations such as oxidation and deamidation are degradations as is aggregation to form insoluble precipitates.

Pharmaceutical stability and thermodynamic stability are not necessarily directly related. For example, a protein

may exhibit thermodynamic instability during freeze-drying and unfold, but if no irreversible reactions occur during storage or during reconstitution, the reconstituted protein may refold completely within seconds and, therefore, display perfect pharmaceutical stability. A protein that remains in the native state and is thermodynamically stable may still degrade via chemical reactions such as deamidation and oxidation over storage times of years, particularly if the reactive moiety is located on the protein surface. Conversely, thermodynamic instability may well be a prelude to degradation. Certainly, an unfolded protein could expose normally buried and protected methionine and asparagine residues to the solution environment and render these residues more reactive. Also, degradation via irreversible aggregation is believed to often proceed through unfolded or partially unfolded conformations as intermediates (84).

Many proteins survive the freeze-drying process with little or no degradation, whereas other proteins exhibit significant degradation and loss of activity during processing. Degradation during the freeze-drying process may arise during freezing and/or during drying, and it is useful to establish when in the process degradation occurs. As a measure of the degradation during freezing, freeze-thaw stability studies are carried out, and to estimate (roughly) the degradation during drying, stability during freeze-drying is compared to stability during freeze-thaw. The basic assumption is that degradation during thawing is comparable to degradation during reconstitution; therefore, the difference in activity between a freeze-dried-reconstituted sample and a freeze-thawed sample is a measure of the loss in activity during drying. This assumption is likely a reasonable approximation for a fast thawing process. One observation of particular significance is that some excipients stabilize during both freezing and drying (called lyoprotectants), whereas others stabilize only during freezing (called cryoprotectants) (41, 85, 86).

If degradation occurs only during freezing and practical variations in the freezing process do not eliminate the problem, it is obvious that a cryoprotectant system is required. However, as a number of distinct stresses may develop during freezing, the proper choice of cryoprotectant is not always obvious. Provided some information regarding the degradation pathway is available, the stabilization strategy could be quite specific. For example, if oxidation is suspected to be a major pathway, the recommendations regarding minimization of oxidation should be followed. If aggregation is a major problem, the use of nonionic surfactants should be considered, or if it is known that a change in pH caused by buffer crystallization is a problem, steps to minimize pH change during freezing

need to be taken. Use of an alternative buffer system or simply use of very low levels of the buffer might well eliminate the pH change and solve the stability problem. Alternatively, addition of an excipient (i.e., sucrose or trehalose) that will interfere with buffer crystallization will also prevent excessive pH shift during freezing (41, 85, 86).

In addition to the specific stabilization strategies noted, one may also employ a general stabilization strategy that is based upon addition of components that normally increase the thermodynamic stability of the protein. That is, addition of a component that will increase the free energy of denaturation will moderate the effect of various freezing stresses that cause a decrease in the free energy of denaturation, with the net result that even during freezing, the native conformation will remain the dominant conformation and degradation will be reduced. The assumption is that increasing thermodynamic stability will also increase pharmaceutical stability. A number of solute types (i.e., amino acids, saccharides, polyols, polyethylene glycols, and some other classes) have been found effective both in increasing thermodynamic stability and in increasing pharmaceutical stability during freezing (85, 86). These solutes are referred to as excluded solutes because such solutes are present in lower concentration near the surface of a protein. The thermodynamic consequence of such exclusion is that the protein is preferentially hydrated, and the chemical potential of the protein is increased. It may be argued that the increase in the chemical potential will be much greater for the unfolded state than for the native state, thereby increasing the free energy of denaturation and stabilizing the native conformation (85, 86). Although the increase in protein chemical potential due to solute exclusion does not always correlate quantitatively with the effectiveness of that solute as a cryoprotectant, it would be prudent to test several pharmaceutically acceptable excluded solutes (i.e., glycine, mannitol, disaccharides, polyethylene glycols) as potential cryoprotectants in cases where freezing instability might involve conformational destabilization. Obviously, polyethylene glycols would not be a good choice if oxidation were an issue. For cryoprotection, it is the concentration of the cryoprotectant in solution that is the key concentration variable, regardless of the concentration of protein in the formulation. With nonpolymer excluded solutes, relatively high concentrations (i.e., 0.2–0.5 *M*) are normally required for effective cryoprotection (85).

Stabilization during drying

In addition to the stresses that develop during freezing, drying imposes an additional stress associated with

essentially complete removal of water. Indeed, the water substitute hypothesis (41, 85–87) is based upon the proposition that a significant thermodynamic destabilization occurs when the hydrogen bonding between protein and water is lost during the last stages of drying. The use of a water substitute as a lyoprotectant allows a hydrogen-bonding interaction between protein and the water substitute, which thermodynamically stabilizes the native conformation and preserves activity. A water substitute is a moiety that is capable of hydrogen bonding to the protein surface much as water and stabilizes via a thermodynamic mechanism; that is, stabilization is achieved by maintaining the free energy of unfolding very high such that essentially all of the protein is maintained in the native conformation. However, it must be recognized that most observations can also be rationalized in terms of a purely kinetic stabilization mechanism. In general, drying is conducted at temperatures sufficiently low that the protein exists in a glassy solid state where molecular mobility is greatly restricted. It may be argued (41) that a good stabilizer is a component that effectively couples protein dynamics to the dynamics of the glass such that even if the protein is destabilized thermodynamically, motion in the protein is too slow to allow significant unfolding during the drying operation. Thus, the protein conformation is stabilized regardless of the free energy of unfolding.

Nonreducing di- and tri-saccharides such as sucrose, trehalose, or raffinose are normally good drying stabilizers (41, 85, 86). They qualify as good water substitutes and also form glasses which, via hydrogen bonding, can couple protein dynamics to matrix dynamics. Under conditions of acid pH, particularly below pH 5, trehalose is a much more effective stabilizer than are sucrose or raffinose. Sucrose and raffinose hydrolyze to their reducing sugar components much more rapidly in the solid state than does trehalose, thereby leading to degradations initiated by reducing sugars (88). The key concentration variable in drying stabilization is the weight ratio (or mole ratio) of saccharide to protein, with the weight ratio of stabilizer to protein normally being between 1:1 and 10:1. The principle here is to insure that the protein is in a matrix of the stabilizer. Thus, with dilute protein systems, good stabilization can often be obtained at rather low concentrations of stabilizer. Here, it may be desirable to employ both a bulking agent such as glycine or mannitol, as well as a stabilizer. This strategy can yield both an elegant cake structure as well as optimal stability without the high solute concentrations than can slow primary drying.

Stabilization during storage

Although the destabilization stresses during storage remain much the same as the drying stresses, the time

scale of storage is obviously much longer than the timescale of freeze-drying. In addition, the freeze-dried product is normally being stored well below its glass transition temperature, meaning molecular mobility is greatly restricted (41, 85, 86). Because of these differences, the dominant degradation pathway during storage may differ from the degradation pathway during freeze-drying. For example, deamidation of asparagine residues is a very common degradation pathway during storage but rarely is a problem during processing. This observation does not imply that deamidation is faster in the dry glassy solid than in solution, but rather is simply the result of the much longer time scale of storage. Indeed, because of the need to insure stability over the shelf life of about 2 years, storage stability normally presents a more serious problem than does stability during processing. Degradation during storage may arise from chemical degradation, such as oxidation and asparagine deamidation, and/or aggregation to noncovalent or covalent aggregates (41, 85, 86). Aggregation often means formation of a population of very large insoluble aggregates that precipitate from solution. It must be admitted, however, that the observation that the level of aggregation increases during storage does not necessarily mean that the bimolecular process of aggregation occurs in the solid state. It is also possible that only conformational changes occur during storage. Aggregation of a fraction of the conformationally modified protein could then occur instantaneously during reconstitution. The remaining fraction would then refold to the native conformation. As aggregation is normally measured in solution, it is not possible to determine just when the aggregation step actually does occur. When the amount of aggregation measured depends on the reconstitution procedure and/or the composition of the diluent (i.e., presence of surfactant) used to reconstitute (32), one would conclude that, at least in part, aggregation of (partially) unfolded protein occurs during reconstitution.

Stabilization strategies focus on three general factors: (1) specific chemical requirements, (2) physical state of the solid, and (3) structure of the protein. Depending on the degradation pathway(s) and mechanisms, any one or all of these factors could be critical. Certainly, if oxidation is a major degradation pathway, stabilization requires attention to the chemical composition of the system, including both the effects of impurities and addition of stabilizers as discussed earlier. Although pH has no real meaning in the solid state, the degree of ionization in the solid does reflect the pH of the solution from which the solid was prepared, and pH of the solution can be a major variable. Because differences in polarity of the environment between a dry solid and an aqueous solution may mean that the optimum

pH for solid stability is somewhat different than the optimal pH for a solution, some solid-state screening experiments are necessary to define the precise optimum pH for stability of the freeze-dried solid.

The physical state of the system is of obvious importance. We freeze dry to increase storage stability relative to the solution state, assuming that the restricted molecular mobility of the solid state will lead to stabilization. In general, an amorphous solid (i.e., glass) is several orders of magnitude more stable than the corresponding aqueous solution, and at least with small molecules, crystalline drugs are much more stable than glasses (89). However, proteins cannot be crystallized during freeze-drying, so the only solid-state option is a glass. Forming a glass does not insure adequate stability. Pure dried proteins are glasses and often do not have sufficient stability for long-term storage at the temperatures of interest, and excipients are added to the formulation to stabilize. Because a physical mixture of protein and excipient will not provide additional stabilization, it is necessary that the proposed stabilizer system remain at least partially amorphous and in the same phase as the protein (41, 85, 86). Further, to avoid chemical reactions between the excipient and protein, the excipient system must be inert. The most common violation of inertness is the use of reducing sugars such as lactose in an attempt to stabilize the protein. For example, while lactose will stabilize hGH against aggregation, an adduct of lactose and hGH quickly forms during storage of the freeze-dried solid (18).

Drying also removes water, a reactant in a hydrolysis reaction, and stability is normally best at very low water content (41, 85, 86). However, the effect of residual

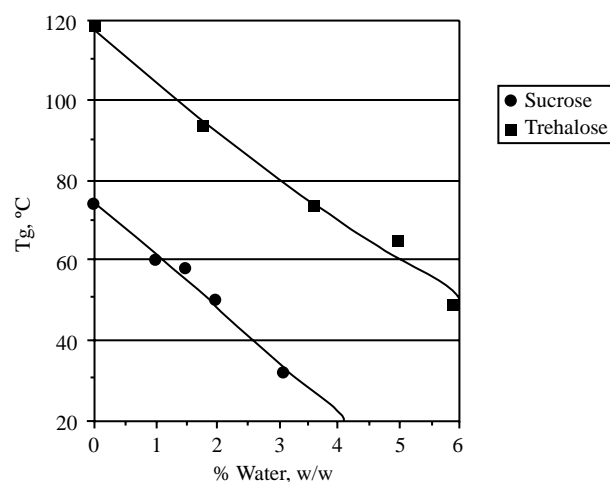


Fig. 12 The effect of residual water on the glass transition temperatures of sucrose and trehalose. (From Ref. 90.)

moisture on stability is far more complex than its possible role as a reactant. Residual water has a major effect on the physical state of the system in that increasing levels of residual moisture decreases the glass transition temperature of the amorphous system, thereby making the system less solid-like at a given temperature, and eventually causing the glass to transform into a fluid at the storage temperature. Figure 12 illustrates the effect of residual moisture on the glass transition temperature of two common stabilizers, sucrose and trehalose (90). For both examples, small amounts of residual water greatly depress the glass transition temperature, but since dry sucrose has a much lower glass transition temperature than does dry trehalose, the water content that gives a glass transition temperature at room temperature, W_g (25 °C), is much lower for sucrose. Thus, in applications where maintaining very low water

content is difficult, trehalose is a better choice for a stabilizer than is sucrose.

Generally, chemical and physical stability decrease sharply as the system enters the fluid state (41, 86, 91, 92). Chemical degradation of hGH in a trehalose formulation illustrates this behavior (92). The trehalose formulation shows a well-defined glass transition temperature with DSC, with the expected decrease in T_g as the water content is increased. The rate constant for chemical degradation is essentially independent of water content while the system remains glassy, but at least for the 50°C data, degradation increases sharply as increasing water content depresses the system glass transition temperature significantly below the storage temperature (Fig. 13). Rates of aggregation show essentially the same behavior (92). However, as stability is not sensitive to water content in the glassy state, stability is not correlated with $T - T_g$ below the glass transition temperature. The study of

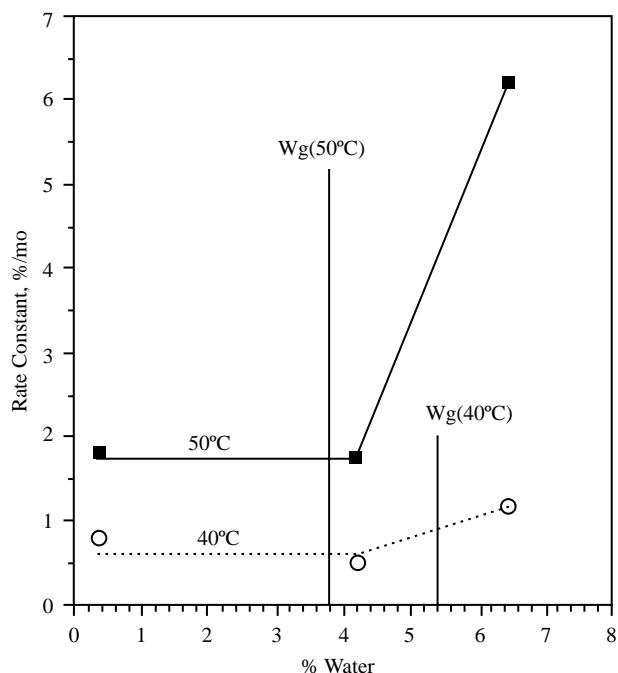


Fig. 13 Chemical degradation in freeze dried hGH formulated with trehalose as a function of water content at 40°C and 50°C. The pseudo first-order rate constant for degradation (%/month) is given for the combination of asparagine deamidation and methionine oxidation. The formulation is hGH:trehalose in a 1:6 weight ratio with sodium phosphate buffer (pH 7.4) at 15% of the hGH content. The highest moisture content samples were collapsed after storage at both 40°C (moderate collapse) and 50°C (severe collapse). The water content that reduces the glass transition temperature of the formulation to the storage temperature is denoted “Wg.” Open circles = 40°C storage; filled squares = 50°C storage. (From Ref. 86.)

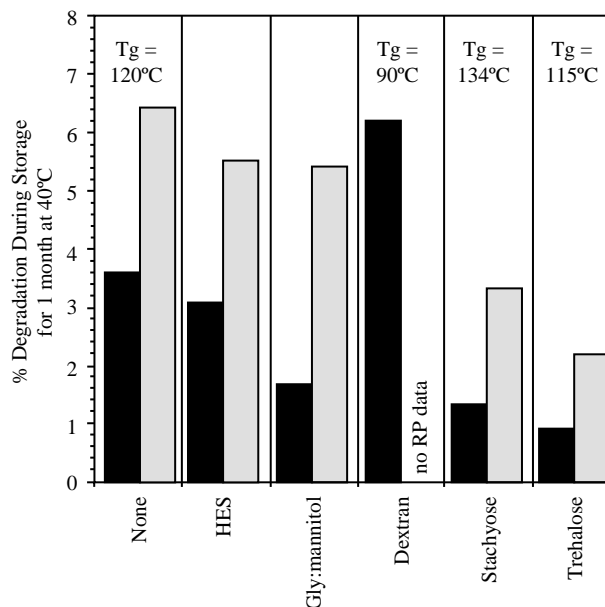


Fig. 14 The effect of excipients on the storage stability of freeze-dried human growth hormone (hGH). Samples were stored for 1 month at 40°C. Solid bars: aggregation (primarily dimer), shaded bar = chemical degradation via methionine oxidation and asparagine deamidation. The glass transition temperatures of the initial freeze-dried formulations are given above the bars when a glass transition temperature could be measured by DSC. The glycine:mannitol formulation is a weight ratio of hGH:glycine:mannitol of 1:1:5, the dextran formulation is 1:6 hGH:dextran 40, none means no stabilizer, and the others are 1:1 hGH:stabilizer. All formulations contain sodium phosphate buffer (pH 7.4) at 15% of the hGH content. Initial moisture contents are all $\approx 1\%$. (From Ref. 86.)

formulation effects on hGH stability suggests the same conclusion. Chemical and aggregation stability of hGH in several formulations is compared in Fig. 14 (92). The glycine:mannitol formulation is a 1:1:5 weight ratio of hGH:glycine:mannitol, where only the mannitol is crystalline. The other formulations are 100% amorphous. Hydroxyethyl starch (HES), stachyose, and trehalose are formulated in a 1:1 weight ratio of excipient:hGH, whereas the dextran formulation is 6:1 dextran:hGH. While the concept that an excipient system must remain at least partially amorphous to improve protein stability is not in question, it is clear that remaining amorphous is not a sufficient condition for stability. Apparent aggregation in the dextran formulation is greater than in the pure protein. Hydroxyethyl starch shows a slight improvement in stability over the pure protein, but is not nearly as effective as the glycine:mannitol formulation, and increasing the level of HES to 3:1 HES:hGH does not improve stability (92). Conversely, both stachyose and trehalose provide better stability than the glycine:mannitol system does, with trehalose superior to stachyose. All systems are glasses at the storage temperature of 40°C, and for those formulations where glass transition temperatures are available, it is clear that storage is well below the T_g , and there is no simple relationship between T_g and stability. Although one might speculate that a glass is more solid, and therefore more stable, the larger the difference between T_g and the storage temperature, the data are not consistent with this speculation. Comparing the stachyose and trehalose formulations, which are both 1:1 formulations with hGH, the T_g of the stachyose formulation is nearly 20°C higher than that of the trehalose formulation, but trehalose offers slightly better stability than does stachyose. These observations and other similar results (92) suggest that while it is necessary for the formulation to have a T_g well above the highest anticipated storage temperature for both elegance and stability reasons, stability well below T_g is not directly related to the precise difference between storage temperature and T_g .

The lack of a clear correlation between stability and T_g for systems well below T_g is not entirely unexpected. Molecular mobility slows greatly below T_g , but does not approach zero until some temperature much lower than T_g (41, 86). Thus, dynamics in the glass depend not only on the value of T_g but are also extremely sensitive to other characteristics of the glass as well as the thermal history of the glass (41, 93). Even if the major component of the glassy matrix (i.e., the sucrose) were to have essentially zero mobility (i.e., negligible translational and rotational motion), mobility of small molecules (i.e., diffusion of water) could still be significant, and mobility of potentially reactive groups on the protein (i.e., asparagine side chain)

could be sufficient to reorient into the transition state and react (41). Thus, it is not only necessary that the matrix itself be solid-like, but effective coupling between matrix mobility and the mobility critical for degradation is needed for solid-like stability behavior (41). Although the molecular characteristics required for effective coupling are not fully understood, it does seem that hydrogen bonding between protein and stabilizer provides one coupling mechanism (41, 86). Thus, disaccharides and trisaccharides are effective stabilizers.

With proteins, structure is also a critical stability variable. It is now common knowledge that proteins often suffer significant conformational changes on freeze-drying that may be moderated by addition of stabilizers to the formulation (41, 85, 86). In principle, different conformations may have different stability characteristics with the native conformation normally believed to represent the most stable conformation (41, 85, 86). Indeed data for rIL-2 show a strong correlation between storage stability and structure as measured by Fourier transform infrared spectroscopy (FTIR) (21), with a more native conformation associated with greater storage stability. Likewise, formulations of freeze-dried hGH having more native-like conformations are more stable to both aggregation and chemical degradation during storage (92). Thus, degradation of a protein in any given formulation is a function of the distribution and reactivities of the protein substates created during the freeze-drying process, with greater stability being associated with the more native-like substates. From this viewpoint, a good stabilizer is one that maximizes the population of native-like substates. Empirical evidence suggests that disaccharides perform this function quite well, regardless of what mechanism might be postulated to explain the observation (41, 85, 86).

In summary, guidelines for stabilization during storage involve the following principles: 1) optimize pH and address specific chemical effects such as oxidation; 2) disperse the protein in an inert glassy matrix with strong coupling between protein dynamics and matrix dynamics to form an amorphous phase such that at all residual water contents and storage temperatures of interest the protein phase will be well below its glass-transition temperature (note that this requires both protein and stabilizer exist in the same glassy phase); 3) employ a formulation (and process) such that the dried protein will retain the native conformation.

Retention of native conformation, coupling of protein mobility with matrix mobility, and formation of a single phase with the stabilizer all require an excipient that interacts, probably via hydrogen bonding, with the surface of the protein. Little or no interaction would likely lead to

Table 3 Glass transition temperatures, (T_g), of selected excipients measured by differential scanning calorimetry^a

Compound	T_g (°C)	Reference
Citric acid	11	94
Glycine	(≈ 30) ^b	66
Lactose	114	95
Maltose	100	95
Mannitol	13	74, 96
Raffinose	114	95
Sorbitol	-1.6	96
Sucrose	75	95
Trehalose	118	95
Maltodextrin 860	169	95
PVP K90	176	95

^aConsult the references for details of the techniques.

^bValue in parentheses is extrapolated from mixtures using the Fox equation and is highly approximate.

phase separation, poor coupling of protein mobility with matrix mobility, and no opportunity for the excipient to stabilize the native conformation by either water substitution or immobilization of the protein. Disaccharides (i.e., sucrose and trehalose) and trisaccharides (i.e., raffinose) seem to satisfy these criteria and are generally good stabilizers. Polymers are generally much less effective. A product glass-transition temperature well above all anticipated storage conditions is an important product quality attribute. Glass transition temperatures of selected amorphous excipients are given in Table 3. Both trehalose and raffinose have much higher glass transition temperatures than does sucrose, and in this sense, would be better choices for stabilizers. However, as long as the residual water content is maintained low and the product is not plasticized by other low-molecular-weight formulation components, a sucrose formulation will be well below its glass-transition temperature at all practical storage temperatures. Due to a very low T_g , sorbitol should not be used as a major formulation component. Incomplete crystallization of glycine or mannitol may also lead to low product glass-transition temperatures. Lactose and maltose have high glass transition temperatures but are reducing sugars and, therefore, are poor stabilizer candidates. As noted earlier, applications at low pH pose problems for both sucrose and raffinose due to rapid hydrolysis to the reducing sugar components. Trehalose is a better choice for low pH applications.

As with drying stabilization, stabilization of a protein for storage with saccharides generally requires a weight ratio of stabilizer to protein between 1:1 and 10:1, with better stabilization at the higher excipient levels. At least for hGH, the rate constants for degradation (chemical

degradation and aggregation) decrease linearly on a plot of the logarithm of rate constant as a function of stabilizer:hGH weight ratio (92). With high dose products where the protein concentration in the fill solution is high, use of high ratios of stabilizer to protein may not be feasible due to the extremely high concentration of stabilizer that would be required.

REFERENCES

1. Pikal, M.J.; Lukes, A.L.; Lang, J.E. Thermal Decomposition of Amorphous Beta-Lactam Antibacterials. *J. Pharm. Sci.* **1977**, *66*, 1312–1316.
2. Pikal, M.J.; Lukes, A.L.; Lang, J.E.; Gaines, K. Quantitative Crystallinity Determinations of Beta-Lactam Antibiotics by Solution Calorimetry: Correlations With Stability. *J. Pharm. Sci.* **1978**, *67*, 767–773.
3. Pikal, M.J.; Dellerman, K.M. Stability Testing of Pharmaceuticals by High-Sensitivity Isothermal Calorimetry at 25°C: Cephalosporins in the Solid and Aqueous Solution States. *Int. J. Pharm.* **1989**, *50*, 233–252.
4. Pikal, M.J. Freeze Drying of Proteins, Part I: Process Design. *Biopharm.* **1990**, *3* (8), 18–27.
5. Pikal, M.J. Freeze Drying of Proteins: Part II. Formulation Selection. *Biopharm.* **1990**, *3* (9), 26–30.
6. Pikal, M.J. Freeze Drying of Proteins: Process, Formulation, and Stability. In *Formulation and Delivery of Proteins and Peptides*; Cleland, J.L., Langer, R., Eds.; ACS: Washington DC, 1994; 120–133.
7. MacKenzie, A.P. Factors Affecting the Mechanism of Transformation of Ice into Water Vapor in the Freeze Drying Process. *Ann. N. Y. Acad. Sci.* **1965**, *125*, 522–547.
8. Broadhead, J.; Rouan, S.K.E.; Rhodes, C.T. The Spray Drying of Pharmaceuticals. *Drug Dev. Ind. Pharm.* **1992**, *18*, 1169–1206.
9. Mumenthaler, M.; Hsu, C.C.; Pearlman, R. Feasibility Study on Spray-Drying Protein Pharmaceuticals: Recombinant Human Growth Hormone and Tissue-Type Plasminogen Activator. *Pharm. Res.* **1994**, *11*, 12–20.
10. Masters, K. Applications of Spray-Drying in the Food Industry, in the Pharmaceutical-Biochemical Industry. In *Spray-Drying Handbook*; Longman Scientific and Technical: Essex UK, 1991; 491–676.
11. Parenteral Lyophilization Facilities: An Innovative Approach to Loading and Unloading Operations Proceedings of the International Congress, on Advanced Technologies for Manufacturing of Aseptic and Terminally Sterilized Pharmaceuticals and Biopharmaceuticals Basel Switzerland Feb. 17–19, 1992; 4–30.
12. Carpenter, J.; Crowe, J.; Arakawa, T. Comparison of Solute-Induced Protein Stabilization in Aqueous Solution and in the Frozen and Dried States. *J. Dairy Sci.* **1990**, *73*, 3627–3636.
13. Tanaka, R.; Takeda, T.; Miyajima, K. Cryoprotective Effect of Saccharides on Denaturation of Catalase by Freeze Drying. *Chem. Pharm. Bull.* **1991**, 1091–1094.
14. Hellman, K.; Miller, D.; Cammack, K. The Effect of Freeze Drying on the Quaternary Structure of L-Asparaginase

- from Erwinia Carotovora. *Biochim. Biophys. Acta.* **1983**, 749, 133–142.
15. Ressing, M.E.; Jiskoot, W.; Talsma, H.; van Ingen, C.W.; Beuvery, E.C.; Crommelin, D.J.A. The Influence of Sucrose, Dextran, and Hydroxypropyl- β -Cyclodextrin as Lyoprotectants for a Freeze Dried Mouse IgG_{2a} Monoclonal Antibody (MN12). *Pharm. Res.* **1992**, 9, 266–270.
 16. Izutsu, K.; Yoshioka, S. Stabilization of Protein Pharmaceuticals in Freeze Dried Formulations. *Drug Stabil.* **1995**, 1, 11–21.
 17. Bell, L.N.; Hageman, M.J. Differentiating Between the Effects of Water Activity and Glass Transition Dependent Mobility on a Solid State Chemical Reaction: Aspartame Degradation. *J. Agric. Food Chem.* **1994**, 42, 2398–2401.
 18. Pikal, M.J.; Dellerman, K.M.; Roy, M.L.; Riggan, R.M. The Effects of Formulation Variables on the Stability of Freeze Dried Human Growth Hormone. *Pharm. Res.* **1991**, 8, 427–436.
 19. Pikal, M.J.; Dellerman, K.M.; Roy, M.L. Formulation and Stability of Freeze Dried Proteins: Effects of Moisture and Oxygen on the Stability of Freeze Dried Formulations of Human Growth Hormone. *Develop. Biol. Standard* **1991**, 74, 323–340.
 20. Townsend, M.W.; DeLuca, P.P. Use of Lyoprotectants in the Freeze Drying of a Model Protein. Ribonuclease A. *J. Parenter. Sci., Technol.* **1988**, 42, 190–199.
 21. Prestrelski, S.J.; Pikal, K.A.; Arakawa, T. Optimization of Lyophilization Conditions for Recombinant Interleukin-2 by Dried-State Conformational Analysis Using Fourier-Transform Infrared Spectroscopy. *Pharm. Res.* **1995**, 12, 1250–1259.
 22. Roy, M.L.; Pikal, M.J.; Rickard, E.C.; Maloney, A.M. The Effects of Formulation and Moisture on the Stability of a Freeze Dried Monoclonal Antibody-Vinca Conjugate: A Test of the WLF Glass Transition Theory. *Develop. Biol. Standard* **1991**, 74, 323–340.
 23. Kiovsky, T.E.; Pincok, R.E. Kinetics of Reactions in Frozen Solutions. *J. Am. Chem. Soc.* **1966**, 88, 7704–7710.
 24. Franks, F. Freeze Drying: From Empiricism to Predictability. *Cryo-Letters* **1990**, 11, 93–110.
 25. Pikal, M.J.; Shah, S. The Collapse Temperature in Freeze Drying: Dependence on Measurement Methodology and Rate of Water Removal from the Glassy Phase. *Int. J. Pharm.* **1990**, 62, 165–186.
 26. Williams, N.A.; Dean, T. Vial Breakage by Frozen Mannitol Solutions: Correlation With Thermal Characteristics and Effect of Stereoisomerism, Additives, and Vial Configuration. *J. Parenter. Sci. Technol.* **1991**, 45, 94–100.
 27. Larsen, S.S. Studies on Stability of Drugs in Frozen Systems. *Arch. Pharm. Chem. Sci. Ed.* **1973**, 1, 41–53.
 28. Murase, N.; Franks, F. Salt Precipitation During the Freeze-Concentration of Phosphate Buffer Solutions. *Biophys. Chem.* **1989**, 34, 293–300.
 29. Gomez, G.; Rodriguez-Hornedo, N.; Pikal, M.J. Effect of Freezing on the pH Of Sodium Phosphate Buffer Solutions. *Pharm. Res.* **1994**, 11, S-265, PPD 7364.
 30. Szkudlarek, B.A.; Rodriguez-Hornedo, N.; Pikal, M.J. Analysis of pH Changes of Potassium Phosphate Buffer Salt Solutions During Freezing. *Pharm. Res.* **1994**, 11, S-228, PPD 7215.
 31. Arakawa, T.; Prestrelski, S.; Kinney, W.; Carpenter, J.F. Factors Affecting Short-Term and Long-Term Stabilities of Proteins. *Adv. Drug Delivery Rev.* **1993**, 10, 1–28.
 32. Carpenter, J.F.; Prestrelski, S.; Arakawa, T. Separation of Freezing- and Drying-Induced Denaturation of Lyophilized Proteins by Stress-Specific Stabilization: I. Enzyme Activity and Calorimetric Studies. *Arch. Biochem. Biophys.* **1993**, 303, 456–464.
 33. Franks, F.F. Product Stability During Drying. *BioPharm Conference, Proceedings 1993*; Advanstar Communications: Eugene Oregon, 1993; 78–87.
 34. Franks, F. *Biophysics and Biochemistry at Low Temperature*; Cambridge University Press: London, 1985.
 35. Privalov, P. Cold Denaturation of Proteins. *Biochem. and Mol. Biol.* **1990**, 25, 281–305.
 36. Tsong, T.Y.; Baldwin, R.L. Effects of Solvent Viscosity and Different Guanidine Salts on the Kinetics of Ribonuclease a Chain Folding. *Biopolymers* **1978**, 17, 1669–1678.
 37. Kiefhaber, T.; Quaas, R.; Hahn, U.; Schmid, F.X. Folding of Ribonuclease T₁. 1. Existence of Multiple Unfolded States Created by Proline Isomerization. *Biochemistry* **1990**, 29, 3053–3061.
 38. Lim, M.; Reid, D. Studies of Reaction Kinetics in Relation to The T_g' Of Polymers in Frozen Model Systems. *Water Relationships in Food*; Levine, H., Slade, L., Eds.; Plenum Press: New York, 1991; 103–122.
 39. Pikal, M.J.; Shah, S.; Senior, D.; Lang, J.E. Physical Chemistry of Freeze Drying: Measurement of Sublimation Rates for Frozen Aqueous Solutions by a Micro Balance Technique. *J. Pharm. Sci.* **1983**, 72, 635–650.
 40. Pikal, M.J.; Shah, S.; Roy, M.L.; Putman, R. The Secondary Drying Stage of Freeze Drying: Drying Kinetics as a Function of Temperature and Chamber Pressure. *Int. J. Pharm.* **1990**, 60, 203–217.
 41. Pikal, M.J. Mechanisms of Protein Stabilization During Freeze Drying and Storage: The Relative Importance of Thermodynamic Stabilization and Glassy State Relaxation Dynamics. In *Freeze Drying/Lyophilization of Pharmaceutical and Biological Products*; Rey, L., May, J., Eds.; Marcel Dekker, Inc.: 1999.
 42. Chang, B.S.; Kendrick, B.S.; Carpenter, J.F. Surface-Induced Denaturation of Proteins During Freezing and its Inhibition by Surfactants. *J. Pharm. Sci.* **1996**, 85, 1325–1330.
 43. Roy, M.L.; Pikal, M.J. Process Control in Freeze Drying: Determination of the End Point of Sublimation Drying by an Electronic Moisture Sensor. *J. Parenter. Sci. Technol.* **1989**, 60–66.
 44. Pikal, M.J.; Roy, M.L.; Shah, S. Mass and Heat Transfer in Vial Freeze Drying of Pharmaceuticals: Role of the Vial. *J. Pharm. Sci.* **1984**, 73, 1224–1237.
 45. MacKenzie, A.P. Basic Principles of Freeze Drying for Pharmaceuticals. *Bull. Par. Drug Assoc.* **1966**, 20, 101–129.
 46. Chang, B.S.; Randall, C.S. Use of Subambient Thermal Analysis to Optimize Protein Lyophilization. *Cryobiology* **1992**, 29, 632–656.
 47. Pikal, M.J. Use of Laboratory Data in Freeze Drying Process Design: Heat and Mass Transfer Coefficients and the Computer Simulation of Freeze Drying. *J. Parenteral Sci., Technol.* **1985**, 39, 115–138.

48. Milton, N.; Nail, S.L.; Roy, M.L.; Pikal, M.J. Evaluation of Manometric Temperature Measurement as a Method of Monitoring Product Temperature During Lyophilization. *PDA J. Pharm. Sci. Technol.* **1997**, *51*, 7–16.
49. Kobayashi, M. Development of New Refrigeration System and Optimum Geometry of the Vapor Condenser for Pharmaceutical Freeze Dryers Proceedings of the 4th International Drying Symposium, Kyoto, Japan, July 9–12, 1984; Toei, R.R., Mujumdar, A., Eds.; 2, 464–471.
50. Pikal, M.J.; Lang, J.E. Rubber Closures as a Source of Haze in Freeze Dried Parenterals: Test Methodology for Closure Evaluation. *J. Parenter Drug Assoc.* **1978**, *32*, 162.
51. Bardat, A.; Biguet, J.; Chatenet, J.; Courteille, F. Moisture Measurement: A New Method for Monitoring Freeze Drying Cycles. *J. Parenter Sci. Technol.* **1993**, *47*, 293–299.
52. Moreira, T.; Cabrera, L.; Gutierrez, A.; Cadiz, A.; Castellano, M. Role of Temperature and Moisture on Monomer Content of Freeze Dried Human Albumin. *Acta Pharm. Nord.* **1992**, *4*, 59–60.
53. Pristoupil, T.; Kramlova, M.; Fortova, H.; Ulrych, S. Haemoglobin Lyophilized With Sucrose: The Effect of Residual Moisture on Storage. *Haematologia* **1985**, *18*, 45–52.
54. Hageman, M.J. The Role of Moisture in Protein Stability. *Drug Dev. Ind. Pharm.* **1988**, *14*, 2047–2070.
55. Greiff, D. Protein Structure and Freeze Drying: The Effects of Residual Moisture and Gases. *Cryobiology* **1971**, *8*, 145–152.
56. Hsu, C.; Ward, C.; Pearlman, R.; Nguyen, H.; Yeung, D.; Curley, J. Determining the Optimum Residual Moisture in Lyophilized Protein Pharmaceuticals. *Develop. Biol. Standard* **1991**, *74*, 255–271.
57. Pikal, M.J.; Shah, S. Moisture Transfer from Stopper to Product and Resulting Stability Implications. *Develop. Biol. Standard* **1991**, *74*, 165–179.
58. MacKenzie, A.P. Collapse During Freeze Drying Qualitative and Quantitative Aspects. *Freeze Drying and Advanced Food Technology*; Goldblith, S.A., Rey, L., Rothmayr, W.W., Eds.; Academic Press: London, 1975; 277–307.
59. Her, L.M.; Nail, S.L. Measurement of Glass Transition Temperatures of Freeze-Concentrated Solutes by Differential Scanning Calorimetry. *Pharm. Res.* **1994**, *11*, 54–59.
60. Pikal, M.J.; Shah, S.; Unpublished Data Eli Lilly & Co.
61. MacKenzie, A.P.; as Cited in Ref. 4.
62. Slade, L.; Levine, H. Beyond Water Activity: Recent Advances Based on an Alternative Approach to the Assessment of Food Quality and Safety. *Crit. Rev. Food Sci. Nutr.* **1991**, *30*, 115–360.
63. Pikal, M.J.; Chang, L.Q.; Unpublished Data University of Connecticut.
64. Shalaev, E.Y.; Kanev, A.N. Study of the Solid-Liquid State Diagram of the Water–Glycine–Sucrose System. *Cryobiology* **1994**, *31*, 374–382.
65. Shalaev, E.; Malakhov, D.V.; Kanev, A.N.; Tuzikov, F.V.; Varaskin, N.D.; Vavilin, V.I. Study of the Phase Diagram Water Fraction of the System Water–Glycine–Sucrose by DTA and X-ray Diffraction Methods. *Thermochim. Acta* **1992**, *196*, 213–220.
66. Lueckel, B.; Bodmer, D.; Helk, B.; Leuenberger, H. Formulations of Sugars with Amino Acids or Mannitol-Influence of Concentration Ratio on the Properties of the Freeze-Concentrate and the Lyophilizate. *Pharm. Develop. and Technol.* **1998**, *3*, 325–336.
67. Suzuki, T.; Franks, F. Solid–Liquid Phase Transitions and Amorphous States in Ternary Sucrose–Glycine–Water Systems. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 3283–3288.
68. Kassraian, K.; Spitznagel, T.; Juneau, J.; Yim, K. Characterization of the Sucrose/Glycine/Water System by Differential Scanning Calorimetry and Freeze Drying Microscopy. *Pharm. Develop. and Technol.* **1998**, *3*, 233–239.
69. Hancock, B.C.; Zografi, G. The Relationship Between the Glass Transition Temperature and the Water Content of Amorphous Pharmaceutical Solids. *Pharm. Res.* **1994**, *11*, 471–477.
70. Williams, N.A.; Lee, Y.; Polli, G.P.; Jennings, T.A. The Effects of Cooling Rate on Solid Phase Transitions and Associated Vial Breakage Occurring in Frozen Mannitol Solutions. *J. Parentes. Sci. and Tech.* **1986**, *40*, 135.
71. Williams, N.A.; Guglielmo, J. Thermal Mechanical Analysis of Frozen Solutions of Mannitol and Some Related Stereoisomers: Evidence of Expansion During Warming and Correlation with Vial Breakage During Lyophilization. *J. Parenter Sci., and Technol.* **1993**, *47*, 119–123.
72. Akers, M.J.; Milton, N.; Byrn, S.R.; Nail, S.L. Glycine Crystallization During Freezing: The Effects of Salt Form, pH, and Ionic Strength. *Pharm. Res.* **1995**, *12*, 1457–1461.
73. Yu, L.; Milton, N.; Groleau, E.; Mishra, D.; Vansickle, R. Existence of a Mannitol Hydrate During Freeze Drying and Practical Implications. *J. Pharm. Sci.* **1999**, *88*, 196–199.
74. Kim, A.I.; Akers, M.J.; Nail, S.L. The Physical State of Mannitol After Freeze Drying: Effects of Mannitol Concentration, Freezing Rate, and a Noncrystallizing Cosolute. *J. Pharm. Sci.* **1998**, *87*, 931–935.
75. Cavatur, R.K.; Suryanarayanan, R. Characterization of Phase Transitions During Freeze Drying by In Situ X-ray Powder Diffractometry. *Pharm. Develop. and Technol.* **1998**, *3*, 579–586.
76. Bam, N.B.; Cleland, J.L.; Yang, J.; Manning, M.C.; Carpenter, J.F.; Kelley, R.F.; Randolph, T.W. Tween Protects Recombinant Human Growth Hormone Against Agitation-Induced Damage Via Hydrophobic Interactions. *J. Pharm. Sci.* **1998**, *87*, 1554–1559.
77. Bam, N.B.; Randolph, T.W.; Cleland, J.L. Stability of Protein Formulations: Investigation of Surfactant Effects by a Novel EPR Spectroscopic Technique. *Pharm. Res.* **1995**, *12*, 2–11.
78. Kerwin, B.A.; Heller, M.C.; Levin, S.H.; Randolph, T.W. Effects of Tween 80 and Sucrose on Acute Short-Term Stability and Long-Term Storage at -20°C Of a Recombinant Hemoglobin. *J. Pharm. Sci.* **1998**, *87*, 1062–1068.
79. Li, S.; Schoneich, C.; Borchard, R. Chemical Instability of Protein Pharmaceuticals: Mechanisms of Oxidation and Strategies for Stabilization. *Biotech. Bioeng.* **1995**, *48*, 490–500.
80. Li, S.; Schoneich, C.; Wilson, G.; Borchardt, R. Chemical Pathways of Peptide Degradation. V. Ascorbic Acid

- Promotes Rather than Inhibits the Oxidation of Methionine to Methionine Sulfoxide in Small Model Peptides. *Pharm. Res.* **1993**, *10*, 1572–1579.
81. Stadman, E.R. Metal Ion-Catalyzed Oxidation of Proteins: Biochemical Mechanism and Biological Consequences. *Free Radical Biol. Med.* **1990**, *9*, 315–325.
 82. Hall, P.; Roberts, R. Methionine Oxidation and Inactivation of α_1 -Proteinase Inhibitor by Cu^{2+} And Glucose. *Biochim. Biophys. Acta* **1992**, *1121*, 325–330.
 83. Li, S.; Schoneich, C.; Borchardt, R. Chemical Pathways of Peptide Degradation. VIII. Oxidation of Methionine in Small Model Peptides by Prooxidant/Transition Metal Ion Systems: Influence of Selective Scavengers for Reactive Oxygen Intermediates. *Pharm. Res.* **1995**, *12*, 348–355.
 84. Brems, D.N. Solubility of Different Folding Conformers of Bovine Growth Hormone. *Biochemistry* **1988**, *27*, 4541–4546.
 85. Carpenter, J.R.; Pikal, M.J.; Chang, B.S.; Randolph, T.W. Rational Design of Stable Lyophilized Protein Formulations: Some Practical Advice. *Pharm. Res.* **1997**, *14*, 969–975.
 86. Pikal, M.J. Freeze Drying of Proteins. In *Peptide and Protein Delivery*; Lee, V.H.L., Ed.; *in press* Marcel Dekker, Inc.: New York.
 87. Arakawa, T.; Kita, Y.; Carpenter, J.F. Protein –Solvent Interactions in Pharmaceutical Formulations. *Pharm. Res.* **1991**, *8*, 285–291.
 88. Pikal, M.; Busse, J.; Kovach, P.; Unpublished Results Eli Lilly & Co.
 89. Pikal, M.J. Impact of Polymorphism on the Quality of Lyophilized Products. In *Polymorphism in Pharmaceutical Solids*; Brittain, H.G., Ed.; Marcel Dekker, Inc.: New York, 1999.
 90. Saleki-Gerhardt, A.; University of Wisconsin-Madison: 1993.
 91. Duddu, S.; Zhang, G.; Dal Monte, P. The Relationship Between Protein Aggregation and Molecular Mobility Below the Glass Transition Temperature of Lyophilized Formulations Containing a Monoclonal Antibody. *Pharm. Res.* **1997**, *14*, 596–600.
 92. Pikal, M.J.; Roy, M.L.; Rigsbee, D.R.; Unpublished Data Eli Lilly & Co.
 93. Shamblin, S.L.; Tang, X.; Chang, L.; Hancock, B.C.; Pikal, M.J. Characterization of the Time Scales of Molecular Motion in Pharmaceutically Important Glasses. *J. Phys. Chem. B.* **1999**, *103*, 4113–4121.
 94. Lu, Q.; Zografi, G. Properties of Citric Acid At the Glass Transition. *J. Pharm. Sci.* **1997**, *86*, 1374–1378.
 95. Taylor, L.; Zografi, G. Sugar –Polymer Hydrogen Bond Interactions in Lyophilized Amorphous Mixtures. *J. Pharm. Sci.* **1998**, *87*, 1615–1620.
 96. Yu, L.; Mishra, D.; Rigsbee, D. Determination of the Glass Properties of D-Mannitol Using Sorbitol as an Impurity. *J. Pharm. Sci.* **1998**, *87*, 774–777.

FOOD AND DRUG ADMINISTRATION: ROLE IN DRUG REGULATION

Roger L. Williams

U.S. Pharmacopeia, Rockville, Maryland, U.S.A.

OVERVIEW

With a FY2000 budget of \$395 billion and employees numbering over 61,000, the Department of Health and Human Services (DHHS) is one of the largest departments in the administrative branch of the United States Government. It is responsible for the activities of 11 operating divisions,^a including the Food and Drug Administration (FDA) (1). Although FDA's budget and staff are comparatively small (FY2000 \$1.1 billion and 9000 employees), FDA has broad authority and impact. FDA regulates approximately \$1 trillion worth of products that are sold in the United States, accounting for about 25% of products purchased by consumers annually and affecting some 95,000 businesses whose manufactured goods fall under FDA regulation. FDA activities work to assure that: 1) foods are safe and wholesome; human and veterinary drugs, human biological products, and medical devices are safe and effective; cosmetics are safe; and radiation-emitting consumer products are safe; 2) regulated products are honestly, accurately, and informatively represented; and 3) regulated products are in compliance with FDA regulations and guidelines, noncompliance is identified and corrected, and any unsafe or unlawful products are removed from the marketplace.

FDA is directed by the Commissioner of Food and Drugs whose appointment by the DHHS Secretary is subject to confirmation by the United States Senate. The Office of the Commissioner includes eight subsidiary offices. These are Chief Counsel; Equal Opportunity; Administrative Law Judge; Senior Associate Commissioner; International and Constituent Relations; Policy, Planning, and Legislation; Management and Systems; and Science Coordination and Communication. In addition, the Office of the Commissioner directs the activities of five

Centers that are responsible for many of the primary regulatory activities of the Agency. These are the Center for Biologics Evaluation and Research (CBER); the Center for Devices and Radiologic Health; the Center for Drug Evaluation and Research (CDER); the Center for Food Safety and Applied Nutrition; and the National Center for Toxicological Research. The first three of these Centers direct activities that result in the availability of many therapeutic products—drugs, biologics, and devices—to treat and prevent human disease. The Commissioner of Food and Drugs also directs the activities of the Office of Regulatory Affairs, which is responsible for FDA's enforcement activities and oversees the activities of more than 1100 field inspectors.

FDA activities are complex, challenging, science-based, and continually changing to meet societal needs and expectations. This article provides a brief overview of these activities, with the understanding that careful study and analysis may be needed to fully understand and adhere to the science, technical and legal conditions that underlie FDA's actions. Many articles and books have been published delineating FDA's statutory and regulatory mandates, to which the reader is referred (2a, 2b, 2c). FDA's web page at <http://www.fda.gov> also provides additional useful information.

LEGISLATION

The Food, Drug, and Cosmetic Act

FDA operates in accordance with provisions of the Federal Food, Drug, & Cosmetic Act (FFDCA) (3) the Public Health Service Act (4), and other laws (5). Although the principle regulatory agency for food and drugs, FDA works cooperatively with other federal agencies, such as the Environmental Protection Agency and the Department of Agriculture, and with state governments. Regulation of food and drugs by FDA dates back almost 100 years to passage of the Federal Food and Drugs Act in 1906. This law was enacted because of widespread concern about patent medicines and food quality in the US. It created

^aThe remaining are the National Institutes of Health, the Centers for Disease Control and Prevention, the Agency for Toxic Substances and Disease Registry, the Indian Health Service, the Health Resources and Services Administration, the Substance Abuse and Mental Health Services Administration, the Agency for Healthcare Research and Quality, the Health Care Financing Administration, the Administration for Children and Families, and the Administration on Aging.

important concepts that continue to the present, namely, that foods and drugs marketed in the United States should not be misbranded, i.e., should not make unsubstantiated claims or otherwise present misleading information, and should not be adulterated. The next major drug regulatory legislation after 1906 was the FFDCA itself, which was enacted in 1938 in response to the Elixir of Sulfanilamide tragedy. Using the excipient diethylene glycol to solubilize the sulfa drug, sulfanilamide, the Elixir was a potent nephrotoxin that killed over 100 children and adults. It was marketed without provision of any kind of information to FDA. In response, the 1938 law created a safety standard that requires performance of adequate tests using reasonable methods to demonstrate that a product is safe. An important feature of the 1938 legislation was and remains a requirement for premarket provision of safety information to FDA, with authority given to the Agency to confirm that the information did in fact support a determination of a safety. The 1938 approach was based on notification instead of premarket approval. If an applicant did not hear from the FDA within 60 days of filing an application it could market the product. The 1938 legislation also established the general approach, still current, that FDA for the most part responds to data developed by sponsors and applicants. FDA usually does not by itself develop data for a regulatory submission, although it frequently engages in sampling and testing of unapproved and approved products. The 1938 legislation introduced other approaches that continue to this date. For example, the concept that certain therapeutic products should bear adequate directions for use and that, in certain instances, some products may be administered only by prescription (Rx Legend) was introduced in the 1938 legislation. This distinction between Rx and over-the-counter (OTC) drugs was further elaborated in the 1951 Durham–Humphrey amendments to the FFDCA.

Many major amendments have been made to the FFDCA since 1938. The most important of these were the 1962 Harrison–Kefauver amendments that arose as a result of the thalidomide tragedy. These amendments created many provisions that form the basis for modern drug regulation by the Agency. The legislation established a premarket process that allows the FDA to judge the safety and efficacy of drugs before they can be legally marketed. It also created a requirement for submission of an Investigational New Drug (IND) application to allow distribution and study of an unapproved new drug. To document efficacy of a new drug, the legislation created an ‘efficacy standard’ requiring substantial evidence of effectiveness based on adequate and well-controlled efficacy studies submitted by an applicant. The 1962 legislation also stated that drugs be produced in

accordance with current Good Manufacturing Practices and expanded substantially the information required in product labeling.

Beyond the 1938 establishing legislation and the 1962 amendments, many other significant legislative changes to the FFDCA have occurred. These include the 1972 Drug Listing Act (established a notification process for commercially marketed products); the 1976 Medical Device Act (created Class I, II, and III types of devices based on risk, with premarket clearance required for Class III); the Drug Price Competition and Patent Term Restoration Act of 1984 (finalized approaches that allow marketing of therapeutically equivalent generic drugs coupled with patent term extension and exclusivity provisions to reward innovation); the Orphan Drug Act of 1983 (creates incentives to develop drugs for rare diseases); the Drug Export Act of 1996 (allows export of unapproved products with certain stipulations); the Prescription Drug Marketing Act of 1987 (protects against diversion of prescription drug products from well-controlled distribution channels); the Generic Drug Enforcement Act of 1993 (debars individuals convicted for illegal activities related to the approval of Abbreviated New Drug Applications); the Prescription Drug User Fee Act (requires payment for review of new drug and analogous applications and certain supplements, plus annual establishment and product fees); and the Dietary Supplement and Health Education Act of 1994 (creates a food category for dietary supplements and establishes a premarket notification process for dietary supplements entering the market after 1994). The most recent legislative amendments to the FFDCA, which also affects the Public Health Service Act, is the 1997 Food and Drug Administration Modernization Act (FDAMA) (6, 7). Many of the elements of drug regulation, as exemplified in the provisions of the FFDCA, have arisen as a result of a finding of or concern for societal risk. In contrast, FDAMA was designed to address a somewhat different perception of risk, namely, that FDA reform was needed to accelerate the availability of new medicines. This perception arose from the belief that an excessively restrictive Agency could create risk by reducing the availability of new therapeutic products. FDAMA focused on improving all aspects of FDA’s regulatory activities, including drugs (Title I), devices (Title II), foods (Title III), coupled with more general changes and requirements (Title IV). FDAMA codified many FDA initiatives previously expressed in regulations or guidance, including: 1) harmonization of measures to regulate the manufacture of drugs and biologics; 2) elimination of the need for insulin batch certification; 3) withdrawal of the distinction between antibiotics and drugs; 4) strategies to streamline approval of drug and antibiotic

manufacturing changes; 5) reduction in the need for environmental assessments; and 6) FDA's rules on accelerated approval of specified investigational drugs. FDAMA also codified FDA's practice of allowing one clinical investigation as the basis for product approval in certain circumstances, while generally preserving the 1962 standard for more than one adequate and well-controlled studies to prove efficacy. FDAMA also changed FDA's policies in many important areas, including allowance of a firm to disseminate peer-reviewed journal articles about an off-label indication of its product with certain stipulations, and allowance of drug companies to provide economic information about their products to formulary committees, managed care organizations, and similar large-scale buyers of health-care products. FDAMA created a special exemption for pharmacy compounding with certain stipulations to prevent pharmaceutical manufacturing under the guise of compounding.

The Public Health Service Act

An additional important set of Federal laws, even older than the FFDCA, relates to the regulation of biologic products. Following the deaths of 12 children from poor quality diphtheria antitoxin, Federal laws were created in 1902 to require the licensing of biologic products. FDA now regulates these under the provisions of the Public Health Service Act, which defines a biologic product as "a virus, therapeutic serum, toxin, antitoxin, vaccine, blood, blood component or derivative, allergenic product, or analogous product, or arsphenamine or derivative of arsphenamine, applicable to the prevention, treatment or cure of a disease or condition of human beings." Biologic products are generally derived from living organisms. Because biologic products are also defined as "drugs" and/or "devices," they are subject to the adulteration, misbranding, and registration provisions of the FFDCA. The importance of the 1902 legislation expanded with availability of therapeutic products produced through recombinant biotechnology approaches. These products fall under the Public Health Service Act as "analogous products" and may be subject to the jurisdiction of CBER. Inter-center agreements at FDA allow review and approval of drugs and biologics produced through recombinant technology in other centers as well.

REGULATIONS

FDA implements the statutory provisions of the FFDCA and PHS Act and associated laws through regulations. Regulations are rules that generally have the force of law.

They provide more explicit information about how a business or manufacturer should conduct their operations and submit information to FDA to be in compliance with the law. For example, the stipulation for adequate and well-controlled investigations stated in the 1962 amendments to the FFDCA was elaborated in a regulation that provides more explicit statements on what constitutes an adequate and well-controlled study. Other important regulations issued by FDA over the last several decades include the 1981 regulation Protection of Human Subjects; Informed Consent; Standards for Institutional Review Boards (clarifies or creates requirements for informed consent and institutional review boards to protect human subjects participating in FDA regulated research) and several regulations designed to accelerate the availability of investigational and approved drugs to treat life-threatening illnesses such as HIV and cancer (1987 Treatment Use of Investigational New Drugs, 1988 Procedures for Subpart E Drugs, 1992 Accelerated Approval, 1992 Parallel-Track Mechanism). A public process exists that allows interested parties to view and comment on preliminary (Announced Notice of Proposed Rule-Making) and draft (Proposed Rule-Making) FDA regulations before they are finalized (Final Rule). Provisions of the Federal Administrative Procedures Act govern the overall process, including the Agency's response to public comments. FDA regulations are published and updated annually in the Code of Federal Regulations (CFR).

GUIDANCES FOR INDUSTRY

FDA uses guidance documents as a means of communicating to regulated industry and the public at large about ways to meet its governing laws and implementing regulations. While guidance documents have been used over the years under many names, FDA began producing them regularly in the mid-1980s as one of three phases of a comprehensive effort to improve the IND/NDA process. Guidances assist sponsors and applicants in understanding how applications should be formatted and what they should contain, how regulated industry should comply with regulatory directives, how inspections should be conducted, and how to comply with many other regulatory activities. The use of guidances has increased as a result of international harmonizing activities, such as the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), the Veterinary International Conference on Harmonization for veterinary products, and the Global Harmonization Task Force for devices. To facilitate

development, issuance, and use of guidance documents, the Agency published a 1997 Good Guidance Practices document (62 FR 8967) that articulated the purpose, definition, legal effects, procedures for development, standard elements, implementation, dissemination, and appeals for Agency guidances. As part of this effort, the FDA committed to publish semiannually possible guidance topics or documents for development or revision during the next year, and to seek public comment on additional ideas for new or revisions of existing guidance documents (63 FR 59317). Many guidances have been produced by FDA over the last several years to assist sponsors and applicants. Although FDA guidances do not have the force of law, they indicate the Agency's best judgment about the amount and type of information needed to satisfy the Agency's legal and regulatory requirements. Regulated industries and businesses do not need to follow guidance recommendations if they wish to employ alternative methods that are acceptable to FDA. In accordance with a requirement of the 1997 FDAMA legislation, FDA is converting its Good Guidance Practices document into a regulation (8).

GUIDANCES FOR REVIEWERS

While most FDA guidances are directed to regulated industries, some are directed to Agency review staff in the form of Good Review Practices documents. These guidances instruct FDA review staff on how to conduct a high quality, consistent, timely review of an application or supplement. Good Review Practices documents may be viewed as one element in a series of quality control elements for regulatory review processes that also include secondary and tertiary supervisory oversight of a primary review, review templates, training, internal standard operating procedures, and many other initiatives as well. Good Review Practices documents complement Agency guidances and are also developed in accordance with the Agency's Good Guidance Practices document. Although Good Review Practices documents and internal standard operating procedures are directed at agency staff, they are of substantial interest to regulated industry and serve as a means of promoting transparency about Agency functions to pharmaceutical sponsors and the public at large.

DRUG REGULATION AT FDA

The FFDCA defines drugs as "articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or other animals" and "articles (other than

food) intended to affect the structure or any function of the body of humans or other animals." The therapeutic use defines whether an article is a drug. Thus, FDA may consider a food, cosmetic, and dietary supplement to be a drug if they are associated with a therapeutic claim. The FFDCA also defines a drug as an article recognized in the official United States Pharmacopeia, official Homeopathic Pharmacopeia of the United States or official National Formulary, or any supplement to these documents. The concept of new drugs arose with the 1938 legislation. FDA allows "grandfathered" drugs marketed in the United States prior to 1938 to remain in the market providing they are generally recognized as safe (GRAS) and effective (GRAE). After the 1938 and 1962 legislation, all new drugs require Agency review and approval before they can be marketed. Recognizing that "new drugs" may be considered by several FDA Centers, information about most new drugs, including both prescription pioneer, generic equivalents, and OTC drugs, is submitted in new drug applications (NDAs) or abbreviated new drug applications (ANDAs), and supplements and annual reports for these applications to CDER. Most of the following discussion thus relates to the regulation of new drugs as executed by this Center. Many elements considered in this discussion are also applicable to biologic products submitted to CBER. Note: Two licenses are necessary to manufacture and distribute some biological products—one for the product (Product Licensing Application/PLA) itself and one for the establishment where it is manufactured (Establishment Licensing Application/ELA). Others only require one application, termed a Biologics Licensing Application, which covers both the product and its manufacture.

DRUG DEVELOPMENT AND APPROVAL

Discovery, development, regulatory assessment with possible approval, and post-marketing manufacture, distribution, and marketing of a new drug is a complex series of activities that is highly resource intensive. Modern drug discovery and development occurs primarily in laboratories of the pharmaceutical industry and in academic and government research centers as well. The loss rate is large, with thousands of drugs screened in early laboratory and animal studies before a few are considered suitable for studies in humans. Information about early discovery and nonclinical animal studies is submitted to the Agency in an Investigational New Drug (IND) application. This part of the regulatory process relies on notification, so that a sponsor may proceed with clinical studies in humans if the Agency does not respond with

30 days of submission. Regulations giving provisions of the IND process are provided at 21 CFR 312 and in further agency guidances (e.g., the FDA guidance *Content and Format of INDs for Phase 1 Studies of Drugs, Including Well-Characterized Therapeutic Biotechnology-Derived Products* (9)). CDER receives approximately 1500 INDs each year, most of which represent individual investigator INDs. Approximately 400 of these are commercial ones submitted by pioneer manufacturers. A key provision of the IND process—and for many clinical studies of approved drugs as well—is the protection of human research subjects, which should occur according to the stipulations of the 1981 informed consent and institutional review board regulations.

After filing an IND, a sponsor conducts nonclinical and clinical studies to assess the safety, efficacy, and quality of an investigational new drug. Information from these studies is collected into an NDA and submitted to FDA for review. These studies involve characterization of the new drug for its important safety, efficacy, and quality attributes. The IND process may take many years and include scores of nonclinical and clinical studies. The studies move through discrete phases. The first phase (Phase 1) group of studies focus on safety coupled with pharmacokinetic and pharmacodynamic studies in small numbers of usually healthy subjects. The studies continue with patient studies to explore efficacy (Phase 2, sometimes termed proof of concept studies) and conclude with additional studies in larger numbers of patients to assess safety further and confirm efficacy (Phase 3). Many additional studies may be performed in association with these primary studies to assess the influence of concomitant medications (drug–drug interaction studies), bioavailability and bioequivalence, subpopulation effects, and other findings that may be useful to practitioners, patients, and consumers in understanding how to use the new drug optimally. While many drugs follow the sequences defined in Phases 1–3, this is not always the case. Depending on an investigational drug's intended indications, its safety profile, its therapeutic need and many other factors, many modifications to the general sequence may occur.

Most NDAs and ANDAs are approved at FDA by CDER. In a single year, the Center approves approximately 100 NDAs, with approximately 30% of these representing previously unapproved new molecular entities (NMEs). The remainder represent line extensions, e.g., new dosage forms, new routes of administration. In addition, CDER approves many thousands of supplements to NDAs each year, most of which represent manufacturing changes but many of which represent supplements providing information about new uses (efficacy supplements) or additional safety information. CDER also has

responsibility for considering annual reports that are required for each approved NDA and ANDA. With a review staff of approximately 1700, CDER is organized into several main units. One of the most important of these is the Office of Review Management that oversees the activities of five Offices of Drug Evaluation (I–V). These Offices in turn are responsible for the activities of 15 review divisions that are organized primarily according to therapeutic groups and drug classes. For example, the Division of Cardio-Renal Drug Products reviews applications that are for new drugs used to treat cardiovascular and renal diseases. The Office of Review Management also comprises an Office of Biostatistics and an Office of Post-Marketing Drug Risk Assessment. The former office reviews statistical analyses in applications and supplements, while the latter office focuses on post-marketing adverse drug reports. A further primary unit in CDER is the Office of Pharmaceutical Science. This Office has responsibility for the Offices of New Drug Chemistry, Clinical Pharmacology and Biopharmaceutics, Office of Testing and Research, and Office of Generic Drugs. Additional administrative and regulatory activities in CDER, including those involving compliance, are handled at the level of the Center. CDER works closely with many Centers at FDA, including CBER and CDRH, and perhaps most notably with the Office of Regulatory Affairs, which is responsible for assuring that firms manufacture products in accordance with current Good Manufacturing Practices (cGMPs). Additional inspections to assure compliance with Good Laboratory Practices (GLPs) and Good Clinical Practices (GCPs) are conducted by CDER.

INDs and NDAs

From a science and technical standpoint, a good conceptual understanding has emerged in the last several decades about the IND processes that result, finally, in a safe, effective, and good quality new drug product. According to this understanding, a drug substance (active ingredient/active moiety and nonactive components, e.g., impurities, residual solvents) is combined with excipients to create a pharmaceutical product with defined identity, strength, quality, purity, and potency. An associated aspect of quality relates to product performance. Performance is assessed by product quality bioavailability and relative bioavailability (bioequivalence) studies. These studies measure the rate and extent of release of the active ingredient from a drug product and its subsequent availability to one or more sites of action. At these sites of action, the active ingredient and/or its metabolites produce the safety and efficacy outcomes reflected in product labeling. Safety, efficacy, and quality topics in

drug development and regulation can be expressed via a set of questions (primary question, test to address the question, and confidence needed in analysis of test outcome) that allow a basis for mutual understanding. From a regulatory perspective, the first set of questions can usually be stated simply (does the new drug have good quality? is it safe? is it effective?). These are the questions that drive nonclinical and clinical characterization studies conducted during the IND period and that lead to a set of data that is submitted in an NDA. Following regulatory assessment and approval if indicated, these characterization studies result in product labeling that provide instructions for use by the practitioner for prescription drug products and by the patient and/or consumer for OTC products. With Agency approval, an applicant may manufacture, distribute, advertise, and sell the approved new drug product throughout the United States market.

An NDA submission includes the following six technical sections: 1) Clinical; 2) Human pharmacokinetics and bioavailability; 3) Chemistry, manufacturing and controls; 4) Microbiology; 5) Nonclinical pharmacology and toxicology; and 6) Statistics. Further information about some of these sections is considered briefly in the following sections. More detailed information is provided in FDA regulatory documents and in internet and print publications. Examples include *the CDER Handbook* (10) and a recently updated document entitled *From Test Tube to Patient: Improving Health through Human Drugs* (11).

Clinical Information

FDA's focus on efficacy began in 1962, was further elaborated in regulations, was modified by FDAMA, and has been considered in detail in an Agency guidance entitled *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* (12). Data from clinical studies containing efficacy data are a primary component of any NDA. These data are judged by FDA review staff to establish effectiveness of a new drug for one or more indications. A core set of guidances, many developed in ICH, also provide basic recommendations on the conduct of clinical studies needed to establish efficacy. In addition, FDA has published many guidances that provide recommendations on clinical trial design approaches, therapeutic endpoints, and other factors to consider in planning clinical safety and efficacy studies for specific disease categories or drug classes. Depending on the drug class, therapeutic indication, and therapeutic need, FDA has substantial latitude in defining the types and amount of information needed to establish safety and efficacy. While full documentation of clinical benefit information, such as reduction in morbidity and mortality,

may be needed to support approval of many new drugs, FDA may rely on lesser information for investigational drugs where a critical therapeutic need exists. For example, FDA may rely on surrogate markets instead of clinical benefit markets to allow approval, with the understanding that additional clinical benefit information may be requested after marketing.

Unlike many other regulatory agencies, FDA encourages frequent meetings with sponsors during the IND period so that good communication occurs about the information that will be needed in an NDA. In the final analysis, this information must convince Agency review staff that adequate and well-controlled studies provide substantial evidence of effectiveness and that study results also show that a product is safe under the conditions of use in the proposed labeling. Overall, the benefits arising from use of the new drug product must outweigh its risks. Risk/benefit judgments are frequently challenging for the applicant, the FDA, and the public at large. For drugs used to treat benign, self-limited conditions, little risk may be tolerated. Larger degrees of risk may be acceptable for drugs used to treat serious and life-threatening illnesses. Both during and after approval, information about safety of a new drug is required through many Agency laws and regulations. During the IND period, many of these requirements are designed not only to provide information about adverse drug reactions but also to protect human subjects participating in clinical trials. After approval, the agency requires product manufacturers and certain health care facilities to report adverse drug events. FDA developed a MedWatch program with a common reporting form and contact points to facilitate reporting of serious adverse events by health care professionals and consumers. MedWatch covers not only new drug products regulated by CDER but also biologics and medical devices. Information from safety reporting may enter product labeling and otherwise impact on the availability of a new drug product to practitioners and consumers. Although rare, withdrawal of an approved new drug product from the market may occur if a pharmaceutical manufacturer and the FDA agree that its benefits no longer outweigh its risks.

Pharmacokinetic and Bioavailability Studies

Pharmacokinetic and bioavailability studies and other clinical pharmacology studies have taken on increasing importance in the set of nonclinical and clinical studies that are performed during the IND period. This has occurred in part as a result of increasing capability to measure an active ingredient/moiety and its metabolites in accessible biologic fluids over time (pharmacokinetic studies). It has also occurred with increasing capability to

study the time course of drug effects (pharmacodynamics) relative to exposure, which can be expressed in terms of either dose or systemic concentration. In addition, most investigational new drugs enter the clinic with a better mechanistic understanding of how positive and negative effects occur in relation to the pathophysiology of disease. A focus of these studies relates to an understanding of an optimal dosage regimen in the population and in an individual. Clinical pharmacology information supports adjustment in dosage regimens based on intrinsic (e.g., genetic polymorphisms, age, gender, height, body mass and composition, and organ dysfunction) and extrinsic factors (e.g., diet, smoking, alcohol intake, concomitant medications). Adjustments in a dosage regimen according to these factors have become of increasing importance. For example, several drugs have been withdrawn from the market in recent years because of dangerous adverse drug reactions (mibefradil, terfenadine, hismanal, cisapride). Dose optimization pharmacokinetic and pharmacodynamic studies, including subpopulation and drug–drug interaction studies, are now performed frequently, in addition to the more routine absorption, distribution, metabolism, excretion (ADME) pharmacokinetic studies that have been performed for many years. A good understanding of exposure–response relationships may also support risk/benefit judgments of safety/efficacy data.

Bioavailability and Bioequivalence

Product quality bioavailability and bioequivalence studies are also an important part of the information needed to support an FDA approval (13). For most orally administered drugs, BA and BE measures are frequently expressed in terms of systemic exposure measures such as area under the plasma concentration–time curve (AUC) and maximum concentration (C_{\max}). These measures of systemic exposure link with safety and efficacy outcomes that may be expressed in terms of biomarkers, surrogate endpoints, or clinical benefit endpoints. Studies that can meet the intent of these regulations for orally administered and certain other drug products have been elaborated in greater detail in an FDA guidance entitled *Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations* (14). Bioavailability and bioequivalence studies document that the performance of a drug product is reliable and consistent. This is important not only as part of the new drug approval process but also in the presence of postapproval changes in the components and composition of an approved drug product and/or its method of manufacture. Depending on the magnitude of these changes, redocumentation of bioequivalence may be needed after approval. Relative BA

studies are useful in comparing the systemic exposure profiles of different dosage forms. In this context, BA information, sometimes together with pharmacokinetic and pharmacodynamic and other data, can be used to link the performance of two different dosage forms and assure comparable clinical outcomes.

Chemistry, Manufacturing and Controls, and Microbiology

During the IND period, sponsors characterize the drug substance and drug product sufficiently so that important quality attributes are established and controlled. This effort focuses on: 1) the drug substance, to assure identity and strength of the active ingredient(s) and to control impurities arising from production and/or degradation; 2) the drug product, to assure the identity and strength of the active ingredient(s) and to monitor degradants that may arise during manufacture and storage; 3) the container–closure system, to protect the drug product during storage; 4) stability testing to assure maintenance of quality attributes during shelf-life; and 5) container labeling. For sterile pharmaceutical products, special approaches are needed. Full understanding of the manufacturing processes for a finished drug product also requires an understanding and application of in-process controls and of the quality of manufacturing materials even when they are not present in the final drug product. Using characterization data, a sponsor develops a set of specifications to assure the identity, strength, quality, purity, and potency of the product and to allow batch release into the marketplace. A specification is defined as a list of tests, references to analytical procedures to evaluate those tests, and the appropriate acceptance criteria. Specifications allow a determination that a particular drug substance or drug product can be considered acceptable for its intended use. Specifications are needed for the drug substance, the drug product, and the container and closure. They may also be needed for intermediates, raw materials, reagents, and other components, including container and closure systems and in-process materials. Specifications are one part of a total control strategy for the drug substance and drug product designed to ensure product quality and consistency. Adherence to current Good Manufacturing Practices is an important part of this overall strategy. Based on characterization and specification setting processes, pioneer manufacturers compile information about the quality of starting materials and their manufacture into a finished dosage form. FDA chemists review this information to assure that critical quality attributes are controlled. In addition, the chemistry review also focuses on in-process controls and validation of

analytical procedures and, for sterile drug products, process validation to assure sterility. Compendial drug substance and excipient monographs as well as general tests and procedures in the *United States Pharmacopeia and National Formulary (USP-NF)* are frequently cited in an NDA and considered during the chemistry review.

Nonclinical Pharmacology and Toxicology Studies

The purpose of nonclinical animal safety studies is to support estimation of initial starting doses in humans and to characterize toxic effects with regard to target organs, exposure, dose dependence, and reversability. This information is provided in a series of studies that focus on single and repeated dose toxicity, reproduction toxicity, genotoxicity studies, local tolerance studies and, for drugs with especial concern and/or that are intended for long-term use, on carcinogenicity. Other nonclinical studies in animals may focus on safety effects on vital organ systems and pharmacokinetic (ADME) studies. The timing of nonclinical studies is important to assure optimal and safe performance of clinical studies. A series of guidelines have become available through ICH that provide guidance on the types of nonclinical studies and their timing needed to support clinical studies of an investigational agent and product labeling for an approved new drug. These cover studies on the following general topics: carcinogenicity, genotoxicity, pharmacokinetics, toxicity, reproductive toxicity, study of biotechnology products, safety pharmacology, and timing of nonclinical and clinical studies. Because this information is required to allow advance of investigational studies in the clinic, it will be submitted in reports and updates during the IND process. It may be summarized and analyzed as well in the NDA.

Biostatistics

Nonclinical and clinical development of an investigational agent requires a series of exploratory and confirmatory studies that rely on adequate statistical analyses. Careful attention is needed with regard to the many aspects of trial design to assure unbiased, robust conclusions regarding safety and efficacy. Important elements of a statistical analysis are discussed in an FDA guidance entitled *Guidance on Statistical Principles for Clinical Trials* (15).

NDA Review

Review of the information in an NDA is conducted according to time frames that are stipulated by law. These

time-frames vary depending on whether the drug merits a priority or standard review. Priority drugs represent a substantial advance over available therapy. FDA commits to reviewing these types of drugs within six months. Standard drugs are defined as having therapeutic qualities similar to an already marketed drug. FDA commits to reviewing these in 10–12 months. The outcome of the FDA's deliberations are expressed in action letters. If submitted information establishes the effectiveness and safety of the new drug, the FDA will issue an approval action letter that permits the applicant to market the approved new drug in the United States. If the information does not establish effectiveness and safety, FDA may issue a nonapprovable action letter with a request for further information. If satisfactory, this information may subsequently support FDA's issuance of an approval letter. Over the years, approximately 60–80% of NDAs are approved, with the number rising in recent years presumably as applicants develop a better understanding of the information needed in an application to establish safety, efficacy, and quality. As part of an approval, FDA may request that an applicant provide additional information as part of an approval commitment (Phase 4 studies).

To assist the Agency in its deliberations, FDA established a system of advisory committees in 1964. These committees review data and provide recommendations to the Agency about whether or not an approval should proceed, or whether some other Agency action is or is not appropriate. The Agency's advisory committees, which function generally under the 1972 Federal Advisory Committee Act, have been modified over the years. Some of these modifications occurred as result of a 1992 Agency requested review by the Institute of Medicine, which coincided with an internal Agency evaluation, and some were put in place in the 1997 FDAMA legislation. The Agency's current advisory committee system involves many committees that meet to discuss a new drug's safety and efficacy, a new indication for an already approved drug, or a special science topic or adverse event profile. FDA's advisory committees provide only recommendations, which the Agency usually follows but at times may not. Membership in an advisory committee is chosen to reflect a needed constituency for a given topic, and includes both consumer and industry representatives. Care is taken to avoid conflicts of interest, with exclusion of a member if needed or, if an individual's view is important, with waiver to allow participation according to specified criteria. Advisory committee charters must be renewed biennially by DHHS and the General Services Administration.

After an NDA is approved, a pharmaceutical manufacturer may promote and advertise the approved new drug in accordance with provisions of the FDCA and its

implementing regulations. An approved application may be supplemented with new safety, efficacy, or manufacturing information.

ABBREVIATED NEW DRUG APPLICATIONS

The 1984 Drug Price Competition and Patent Term Restoration amendments to the FDCA created an abbreviated mechanism for the approval of generic copies of drug products approved for safety and efficacy via the NDA process. Provisions of the amendments and its implementing regulations require that a generic applicant demonstrate that its product is the same as that of the corresponding innovator drug (the reference listed drug) in terms of active ingredient(s), strength, dosage form, and route of administration. These stipulations are termed pharmaceutical equivalence. In addition, the applicant must demonstrate that the labeling of its proposed generic version is comparable to that of the innovator product and that the generic product is bioequivalent to the reference listed drug. With this approach, the requirement for extensive nonclinical and clinical testing for a generic product is frequently obviated. Information developed by a generic applicant is submitted for Agency review in an ANDA if acceptable, the application is approved and the generic copy is deemed interchangeable with the corresponding reference product under specified conditions of use. As part of the 1984 legislation, an extensive series of requirements regarding patent certification and exclusivity was developed. These approaches were part of a general intent of the 1984 legislation to balance incentives for innovation with the societal need for low-cost duplicates of pioneer products.

OVER-THE-COUNTER DRUG PRODUCTS

Based on distinctions created in the 1938 FDCA and the 1951 Durham–Humphrey Act, new drugs are categorized as either prescription and nonprescription or OTC. OTC drug products are deemed sufficiently safe for self-use. While FDA applies the same standards of safety and efficacy to prescription and nonprescription new drugs, it regulates them in two ways. One is by a monograph system and the other is through switching a prescription drug approved under an NDA to nonprescription status (Rx to OTC Switch). The OTC monograph system arose out of a need to document efficacy for the many thousands of new drugs that had been approved for safety only between 1938 and 1962. This was accomplished through the Drug

Efficacy Safety Implementation (DESI) program. As part of the effort, FDA initiated an OTC Drug Review in 1972 that resulted in the development of over 100 monograph categories (e.g., antacids, laxatives) for over 500 active ingredients that were marketed between 1938 and 1962 in approximately 700,000 different dosage forms. A manufacturer may market an OTC product in one of these categories without submitting information to FDA providing the manufacture and marketing of the product conform to stipulations in the monograph and to *USP-NF* substance and product monographs if available. In addition, experience with a prescription drug after its approval may result in an understanding that its can be used safely and effectively without professional supervision. FDA has allowed this type of Rx to OTC switch, either for approved prescription new drugs or for new indications, on 66 occasions in the last five years.

THE UNITED STATES PHARMACOPEIA

Practitioners established the *United States Pharmacopeia* in 1820 to promote the availability of unadulterated and appropriately named and prepared therapeutic products. With the addition of an information component beginning in 1980, USP's role expanded from establishing standards for healthcare articles to providing useful information to assist practitioners, consumers, and patients in optimal use of therapeutic products. Section 201 (j) of the Federal Food, Drug and Cosmetic Act defines the *United States Pharmacopeia (USP)* and the *National Formulary (NF)* as official compendia. These texts provide quality standards for therapeutic products and excipients approved under the provisions of the Food, Drug & Cosmetic Act, and other therapeutic products as well. The availability of a *USP-NF* monograph requires that any drug marketed under the monograph name must comply with the specifications, irrespective of whether the article bears the *USP-NF* designation. Through the adulteration and misbranding provisions of the Food, Drug & Cosmetic Act, FDA can take enforcement action against firms whose drug products do not comply with a *USP* or *NF* standard. The letters “USP” or “NF” are not trademarked and can be utilized by companies for non-drug products if they wish as a representation of the quality of their products subject only, for the most part, to regulatory constraints. Manufacturers of drug are not required to comply with USP standards but if they choose not to do so they are required to label their product as “not USP” and indicate how their product differs on the container label. For the most part, manufacturers choose to adhere to USP standards rather than conform to this requirement.

ADDITIONAL INFORMATION

In regulating drugs, FDA is continually changing and growing as a result of societal needs, rapidly changing science and technology, and health care delivery challenges. It also must consider international activities and harmonization and carve-outs where regulatory focus is diminished or clarified. Key to all of FDA's activities is the availability of resources to allow performance of its statutory functions. Some of these issues are discussed briefly in the following paragraphs.

Resources

Until 1992, resources for FDA came from appropriated tax dollars allocated at the beginning of each year through budget processes of the Administration and Congress. In 1992, Congress enacted the Prescription Drug User Fee act (PDUFA) that authorized FDA to collect fees from the prescription drug industry to augment FDA's appropriated resources. The additional resources were to be used to expedite the review of NDAs and Biologics Licensing Applications (BLAs) so that prescription drug products could reach the marketplace more quickly. The first PDUFA program, termed PDUFA I, was enacted for a period of five years, ending in 1997. In that year, as part of FDAMA (Title 1/Subtitle A—Fees Relating to Drugs), PDUFA was reauthorized, as PDUFA II, for five more years. PDUFA I included performance goals for FDA that were coupled with FDA managerial reforms. Via fees charged for application review, more than 900 employees have been added to the Agency's new drug and biologics review programs. In 1999, approximately \$125 million in user fees were collected from the pharmaceutical industry. With these resources, CDER and CBER have generally met or exceeded the performance objectives of PDUFA I and II. PDUFA II has increased both the resources available to the FDA as well the performance objectives, e.g., by 2002, when PDUFA II ends, 90 percent of standard original NDAs and BLAs filed during FY 2002 will be reviewed and acted on within 10 months of receipt.

PDUFA provides resources only for the review of NDAs, leaving many of areas and processes of the Center still supported by appropriated dollars. Failure to increase Agency appropriations in the last several years have left many non-PDUFA components at FDA with substantially lower funding at the end of the 1990s than at the beginning. While PDUFA resources are useful, consumer groups and other have expressed concern over both undue emphasis on review timeliness and industry influence, given that a substantial fraction of the budget for new drug

reviews is now provided by industry. Recent market withdrawals have heightened this concern (16, 17), although FDA has provided data to indicate that market withdrawals diminished in frequency since PDUFA was introduced (18).

International Harmonization

FDA engages in many international activities that result in information exchange, working closely with the World Health Organization and in different types of bilateral, trilateral, and multilateral arrangements. One of the most substantial examples of this effort is ICH (19). A primary objective of ICH is to avoid duplicative animal and human testing and to come to a common understanding of technical requirements to support the registration processes in the three ICH nations/regions, which include the European Union, Japan, and the United States. ICH began in 1989 as a collaborative effort between representatives from the Food and Drug Administration working with representatives from the European Commission (at the time the Commission of the European Communities) and the Japanese Ministry of Health and Welfare, and also with representatives from the three corresponding pharmaceutical manufacturers associations, the Japanese Pharmaceutical Manufacturers Association, the US Pharmaceutical Research and Manufacturers of America, and the European Federation of Pharmaceutical Industries Associations. Observers to ICH include representatives from the World Health Organization (WHO), the European Free Trade Association, and Canada's Therapeutic Products Program. ICH has established a Steering Committee, which meets twice yearly, moving sequentially through each ICH area, to provide oversight to the ICH Expert Working Groups that focus on specific topics for harmonization. Secretariat support to ICH is provided by the International Federation of Pharmaceutical Manufacturers Association provides two representatives to the Steering Committee. The ICH Expert Working Groups focus on Efficacy (clinical safety and efficacy topics), Safety (nonclinical safety), Quality, and Regulatory Communications. ICH has established a process for guidance development that begins with identification of a topic area for harmonization, for which a concept paper is developed, collection of background information, and formation of an Expert Working Group to draft the guidance (Step 1). This initial process yields a draft guidance that moves through a multistep process that, if successful, yields a final guidance that becomes part of the regulatory machinery in the European Union, the United States, and Japan (Steps 2–5).

With this approach, ICH differs from guidances developed by the World Health Organization, which do not necessarily become binding on a regulatory agency. In the 10 years since its inception, ICH has resulted in the preparation of over 40 finalized guidelines and position papers that are designed to guide drug development and registration activities in the European Union, Japan, and the United States. A more recent focus of ICH has been the development of a Common Technical Document that will provide a core set of information to support an NDA or BLA in the US, and the corresponding application documents in Europe and Japan. ICH has confronted several issues and challenges during the initial 10 years of its existence. For example, ICH is highly resource intensive and has at times excluded interested parties and stakeholders. The issue of exclusion has been dealt with at least in part via the efforts of WHO, which has disseminated ICH draft and finalized guidelines to WHO Member States. In addition, at biennial meetings of WHO's International Conference of Drug Regulatory Authorities, presentations on the status of ICH topics has become an increasingly important part of the program.

Regulatory Control

A key question for any society is the scope of responsibility that it gives to its regulatory agencies. This scope is defined through legislative and administrative actions, through judicial decisions, and through regulatory action or inaction. In the last decade, legislative decisions have limited and/or clarified the scope of FDA's responsibilities in two important areas. Examples include the regulation of dietary supplements and pharmacy compounding. Before 1994, FDA regulated dietary supplements according to the provisions of the 1958 Food Additive Amendments to the FFDCA, which required pre-market safety review for all new ingredients, including dietary supplements. In the 1994 Dietary Supplement Health and Education Act (DSHEA), Congress added provisions to the FFDCA that eliminated requirements for pre-market documentation of safety for dietary supplements and dietary ingredients of dietary supplements (20). In accordance with the 1958 legislation, these requirements continue to apply to other new food ingredients or for new uses of old food ingredients. DSHEA and the 1990 Nutritional Labeling and Education Act (NLEA) also extended the definition of a dietary supplement from vitamins, minerals and, proteins to herbs or similar nutritional substances and substances such as ginseng, garlic, fish oils, psyllium, enzymes, glandulars, and mixtures of these. Many provisions of DSHEA are now being implemented via FDA regulations.

A particularly challenging one relates to regulations that distinguish between dietary supplement claims and drug claims. The FFDAC defines a drug as an article intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or animals. A product may be subject to regulation as a drug if it makes a claim that is other than food, and it is intended to affect the structure or any function of the body of man or other animals. Manufacturers of a dietary supplement may claim an affect on the structure and function of the body as long as they do not make drug claims. FDA has been challenged to provide a clear distinction between drug and dietary supplement claims in product labeling. This challenge reflects the boundary between food and drug claims and relates to the limitations and/or clarification in the scope of FDA's responsibilities expressed in NLEA and DSHEA. This distinction is currently provided in an FDA regulation entitled Regulations on Statements Made for Dietary Supplements Concerning the Effect of the Product on the Structure or Function of the Body (21).

Section 127 of FDAMA added section 503A of the FFDCA to clarify the status of pharmacy compounding of a drug by a pharmacist or physician on a customized basis for an individual patient. Section 503A defines pharmacy compounding to allow exemptions from the Good Manufacturing Practices, full disclosure requirements, and new drug provisions of the FFDAC. To qualify for these exemptions, a compounded drug product must satisfy several requirements delineated in Section 127. The general objective of section 127 is to allow a pharmacist and or physician to engage in the legitimate practice of compounding and to clarify the distinction between compounding and pharmaceutical manufacturing that comes under the oversight of FDA.

SUMMARY

Drugs to prevent and/or treat disease are a critical component of the U.S. health care system. The U.S. total health care bill is well over \$1 trillion, with medical products accounting for approximately 8% of this total expenditure. Both the total health care cost and the fraction of the cost represented by medical products is expected to rise substantially in the coming years. Availability of most medical products is regulated closely by FDA and is one activity in a complex set of activities by which a new drug reaches a patient. Generally, these activities may be viewed in terms of discrete yet overlapping efforts that include drug discovery, nonclinical and clinical development programs, regulatory assessment, and utilization.

Revolutionary changes are occurring in each of these four areas. Drug discovery is driven by an increasingly detailed understanding of human physiology and pathophysiology arising from the molecular biology revolution and many associated diagnostic and therapeutic advances. Molecular modeling and combinatorial chemistry combined with high through-put screens for positive and negative drug effects and biopharmaceutical properties have expanded the number of lead candidates for nonclinical and clinical development. An improved understanding of the mechanistic basis for drug absorption and disposition and drug action provides better information about drug safety and efficacy. The regulatory review process at FDA has become more timely, in part as a result of availability of resources from users' fees charged to pharmaceutical sponsors submitting NDAs for prescription drugs and certain supplements to these applications.

The utilization of drugs to maintain health and treat disease has undergone profound change in the medical community as a result of many factors, including the need to contain costs yet assure that patients have access to the latest medical treatments. Overarching these changes in discovery, development, assessment, and utilization have been revolutions in information technology, materials science, management, and many other areas. All activities associated with getting a medical product to a patient are further affected, sometimes profoundly, by national and international societal and political factors. The intensity with which FDA functions relates to countervailing forces that work to promote the availability of the latest medical products versus forces that work to focus on the safety of these products. Globalization of the pharmaceutical industry has resulted in a need to harmonize on regulatory requirements and recommendations for the development of a new drug product. From a public policy perspective, FDAMA represents societal encouragement, expressed through Congressional legislation, for FDA to act more rapidly in making regulatory decisions. Availability of users' fees has provided FDA with resources to achieve this objective. Changes arising from FDAMA and users' fees appear to represent a major transformation in societal thinking about drug regulation, which in the past has focused on an expectation the FDA should function as a gatekeeper to keep unsafe medical products and foods from the market. FDAMA and users' fees work to shift that focus to facilitation of the availability of medical and other products regulated by FDA. These changes arose in the decade of the 1990's as a result of many factors, including a long history of charges that FDA created a drug lag, availability of an increasing number of important new medical products, a rise in activism from patients with serious and life-threatening diseases such as AIDS and

cancer, and many other factors as well. As with all societal directives, opinions and forces may cause a reversal in Agency approaches. Concerns are raised now about drug safety and certain market withdrawals of unsafe drugs and drug combinations. The charge, generally refuted by FDA (18), is that a more rapid regulatory process leads to the availability of unsafe medical products.

REFERENCES

1. FDA Mission Statement, <http://www/fda.gov>.
- 2a. Cooper, R.M., Ed.; *Food and Drug Law*, 5th Ed. Food and Drug Law Institute: Washington, DC, 1991.
- 2b. Beers, D.O. *Generic and Innovator Drugs—a Guide to FDA Approval Requirements*; Aspen Law Business: New York, 1998.
- 2c. Temple, R. Development of Drug Law, Regulations, and Guidance in the United States. *Principles of Pharmacology, Basic Concepts and Clinical Applications*; Munson, P.L., Ed.; Ch. 113 Chapman & Hall: New York, 1996, Revised Reprint.
3. 21 U.S.C. Sections, 321–397.
4. 42 U.S.C. Sections 241 (Research and Investigations), 242 (a) (Controlled Substances), 242l (International Cooperation), 262–263 (Biological Products), 264 (Interstate and Foreign Infectious Disease Control Functions that Relate to the Law Enforcement Functions of FDA) (1994), See also 21 C.F.R. 5.10 (a), 1998.
5. Horton, L. Mutual Recognition Agreements and Harmonization. *Seton Hall Law Rev.* **1998**, 29, 692–735; Orphan Drug Act of 1983, Pub. L. No. 97-414, 96 Stat. 2049, 1983; Drug Price Competition and Patent Term Restoration Act of 1994, Pub. L. No. 98-417, 98 Stat. 1585, 1984; Food Drug Modernization Act of 1997, Pub. L. No. 105-115, 111 Stat. 2296, 1997.
6. Pub. L. No. 105-115, 111 Stat., 2296–1997.
7. FDAMA and Related Document, <http://www.fda.gov/opacom/7modact.html>.
8. Federal Register **2000**, 65 (30).
9. <http://www.fda.gov/cder/guidance>.
10. <http://www.fda.gov/cder/handbook/index.htm>.
11. <http://www.fda.gov/cder/about/whatwedo/testtube.pdf>.
12. <http://www.fda.gov/cder/guidance-clinical/medical>.
13. 21 Code of Federal Regulations 320.
14. <http://www.fda.gov/cder/guidance/biopharmaceutics>.
15. <http://www.fda.gov/cder/guidance/91698>.
16. Withdrawals of FDA-Approved Drugs Raise Questions. *Mayo Clin. Health Letter* **1998**, 16 (4).
17. Kleinke, J.D.; Gottlieb, S. Is FDA Approving Drugs Too Fast. *BMJ* **1998**, 70, 405–406.
18. Friedman, M.A.; Woodcock, J.; Lumpkin, M.M.; Shuren, J.E.; Hass, A.E.; Thompson, L.J. The Safety of Newly Approved Medicines—Do Recent Market Removals Mean There is a Problem? *JAMA* **1999**, 281, 1728–1734.
19. ICH information, <http://www.ifpma.org/ich1.html>.
20. A Summary of The Dietary Health Supplement and Education Act is Available. <http://vm.cfsan.fda.gov/dms/dietsupp.html>.
21. 21 CFR 101, Federal Register, **2000**, 65 (4), 1000–1050.

GOOD MANUFACTURING PRACTICES (GMP)—AN OVERVIEW^a

Ira R. Berry

Duramed Pharmaceuticals, Inc., Somerset, New Jersey, U.S.A.

INTRODUCTION

The Current Good Manufacturing Practice (cGMP) regulations for finished pharmaceuticals that have been promulgated by the U.S. Food and Drug Administration (FDA) have been a subject of active discussion since they were first published with the passage of the Kefauver–Harris Drug Amendments in 1962 (1, 2). GMPs were intended to establish minimum manufacturing and control practices for the pharmaceutical industry and focus on what needed to be done rather than how it should be done. Failure to comply with the current Good Manufacturing Practice regulations as set forth in the “Code of Federal Regulations,” 21 CFR Parts 210 and 211 (3), constitutes adulteration of a drug that is entered into interstate commerce and is therefore subject to regulatory action. These requirements apply to human and animal drugs. The regulations in Part 210 are introductory in nature; Part 211 contains the more detailed and descriptive regulations and which will be discussed in this chapter.

In the late 1970s, the FDA organized a task force to study the GMPs. Revised GMPs were published in September 1978, and became official in March 1979. At that time, the FDA also considered establishing more specific GMP regulations for products such as small-volume parenterals, medicinal gases and drug substances, to supplement the existing umbrella regulations.

Today, separate GMPs are in effect for biologics and foods but have not yet been promulgated for small-volume parenterals, medicinal gases or drug substances. In attempting to create regulations for specific products, the FDA concluded that it would be better to first issue guidances and guidelines rather than to revise regulations. Thus, what is put forth in 21 CFR Part 211 is supplemented with a number of guidances, guidelines and Compliance Policy Guides (4–11). There remain some differences, however, between guidances and guidelines from the Center for Biologics Evaluation and Research (CBER), the Center for Drug Evaluation and Research (CDER), and the Compliance Policy Guides, sometimes leaving a firm’s

cGMP status subject to the interpretation of a field investigator.

Based on the amount of time needed to promulgate a revision of the regulations, it is understandable that it is preferable to work with guidances, guidelines, and compliance policy guides. Current GMPs are supposed to be, as their title indicates, a description of the current manufacturing and control practices that are acceptable for a pharmaceutical company selling products in the United States. Although these cGMPs are not enforced in some foreign countries, an FDA inspection in a foreign country, based on current GMPs, can be the key to importing and marketing a product in the United States.

The FDA is required to inspect a firm every 2 years for compliance to cGMPs. With the advent of programs such as the new drug preapproval inspection program implemented in 1990, inspections may be more frequent and have expanded into areas not previously investigated regularly by the FDA, such as clinical manufacturing. An unsatisfactory inspection can delay approval of new products and lead to further regulatory action by the FDA, such as seizure and injunction, for existing products. The penalties can apply to the individual or both the firm and individuals.

The GMPs as set forth in 21 CFR Part 211 also have been applied to drug substances and clinical products. Guidelines and guidances have been issued to describe the FDA interpretation of 21 CFR Part 211 pertaining to drug substances and the production of investigational drugs and reinforce the agency’s understanding that cGMPs are applicable (12, 13). The FDA has reinforced the connection between registration of drugs and the manufacture of active pharmaceutical ingredients by the issuance of guides to industry (14–16). Recently, the FDA has issued for comment a draft guidance for “Good Manufacturing Practice for the Manufacturing, Processing, and Holding of an Active Pharmaceutical Ingredient” (17). Finalization of these draft cGMP principles is being written into a guideline that is being coordinated through the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH) toward publication of a guidance for active pharmaceutical ingredients that will be standardized

^aRevised from Ref. 1.

and followed by manufacturers in the United States, Europe, and Japan (18). As a sidenote, active pharmaceutical ingredients have also been called drug substances and bulk pharmaceutical chemicals.

The “Status of Current Good Manufacturing Practice Regulations for Finished Pharmaceuticals” is as follows:

Section 210.1(a) The regulations set forth in this part and in parts 211 through 226 of this chapter contain the minimum current good manufacturing practice for methods to be used in, and the facilities or controls to be used for, the manufacture, processing, packing, or holding of a drug to assure that such drug meets the requirements of the act as to safety, and has the identity and strength and meets the quality and purity characteristics that it purports or is represented to possess.

(b) The failure to comply with any regulation set forth in this part and in parts 211 through 226 of this chapter in the manufacture, processing, packing, or holding of a drug shall render such drug to be adulterated under section 501(a)(2)(B) of the act and such drug, as well as the person who is responsible for the failure to comply, shall be subject to regulatory action.

The “Applicability of Current Good Manufacturing Practice Regulations for Finished Pharmaceuticals” is as follows:

Section 210.2(a) The regulations in this part and in parts 211 through 226 of this chapter as they may pertain to a drug and in parts 600 through 680 of this chapter as they may pertain to a biological product for human use, shall be considered to supplement, not supersede, each other, unless the regulations explicitly provide otherwise. In the event that it is impossible to comply with all applicable regulations in these parts, the regulations specifically applicable to the drug in question shall supersede the more general.

(b) If a person engages in only some operations subject to the regulations in this part and in parts 211 through 226 and parts 600 through 680 of this chapter, and not in others, that person need only comply with those regulations applicable to the operations in which he or she is engaged.

This article reviews Part 211, Current Good Manufacturing Practice for Finished Pharmaceuticals. Title 21, Parts 600 through 680 for biological products, supplement but do not supersede the regulations in this part, unless the regulations explicitly provide otherwise. The

focus of Good Manufacturing Practice for all products is on a quality control unit that has the responsibility and authority to approve or reject all components, drug product containers, closures, in-process materials, finished product, and production and control documentation. “Quality Control Unit” refers to any person or organizational element designated by the firm to be responsible for the duties relating to quality control. Specific subparts of Part 211 are summarized and described later.

This article is not intended to reproduce the complete GMPs, but certain parts are excerpted for emphasis.

SUBPART A: GENERAL PROVISIONS

This subpart reinforces Part 210, which is an introduction to the more detailed and specific practices described in Part 211. It establishes the scope of GMP and carries forth the definition of terms in Part 210.

SUBPART B: ORGANIZATION AND PERSONNEL

Section 211.22: Responsibilities of Quality Control Unit

(a) There shall be a quality control unit that shall have the responsibility and authority to approve or reject all components, drug product containers, closures, in-process materials, packaging material, labeling, and drug products, and the authority to review production records to assure that no errors have occurred or, if errors have occurred, that they have been fully investigated. The quality control unit shall be responsible for approving or rejecting drug products manufactured, processed, packed, or held under contract by another company.

(b) Adequate laboratory facilities for the testing and approval (or rejection) of components, drug product containers, closures, packaging materials, in-process materials, and drug products shall be available to the quality control unit.

(c) The quality control unit shall have the responsibility for approving or rejecting all procedures or specifications impacting on the identity, strength, quality, and purity of the drug product.

(d) The responsibilities and procedures applicable to the quality control unit shall be in writing; such written procedures shall be followed.

The intent of this subpart is to ensure that there is a group within the organization that can review and judge the acceptability of procedures used to produce pharmaceutical products on an independent basis, as well as judging the products themselves, before they are entered into interstate commerce. The FDA has emphasized separation of the quality control unit from production (organizationally). In addition, the FDA considers the organizational level to which the quality control unit reports very important.

From a legal perspective, the chief executive officer (CEO) or president of a firm is considered the most responsible official and thereby becomes the most liable. Therefore, he is subject to criminal prosecution should the organization be found to violate the Food, Drug and Cosmetic (FDC) Act. One of the most serious infractions is fraud, that is, the intent to mislead the FDA. Hence, it is incumbent on the CEO to have well-qualified personnel in the organization and an organizational structure that reinforces quality.

Section 211.25: Personnel Qualifications

This section emphasizes the training of personnel both in cGMP and in their specific responsibilities with regard to manufacturing, processing, packing or holding of a drug product and functions to provide assurance that the drug product has the safety, identity, strength, quality, and purity that it purports or is represented to possess. This section also requires that there be a sufficient number of qualified personnel.

(a) Each person engaged in the manufacture, processing, packing, or holding of a drug product shall have education, training, and experience, or any combination thereof, to enable that person to perform the assigned functions. Training shall be in the particular operations that the employee performs and in current good manufacturing practice (including the current good manufacturing practice regulations in this chapter and written procedures required by these regulations) as they relate to the employee's functions. Training in current good manufacturing practice shall be conducted by qualified individuals on a continuing basis and with sufficient frequency to assure that employees remain familiar with cGMP requirements applicable to them.

(b) Each person responsible for supervising the manufacture, processing, packing, or holding of a drug product shall have the education, training, and experience, or any combination thereof, to perform assigned functions in such a manner as to provide assurance that the drug product has the safety, identity, strength, quality, and purity that it purports or is represented to possess.

(c) There shall be an adequate number of qualified personnel to perform and supervise the manufacture, processing, packing, or holding of each drug product.

This section makes it clear that the quality control unit is not the only group responsible for the quality of products and conformance with GMP. Because the quality of a product must be "built in," control (at the manufacturing level) of raw materials and process control are important.

SUBPART C: BUILDINGS AND FACILITIES

This section requires that the buildings and facilities are adequate, provide specifically defined areas for certain operations and are designed to prevent mix-ups. Included are design and construction features; lighting; ventilation, air filtration, air heating and cooling; plumbing; sewage and refuse disposal; washing and toilet facilities; sanitation; and maintenance. Lighting, ventilation, air filtration, and air heating and cooling must be adequate. Again, the word adequate is used frequently. This is where an individual investigator's and firm's interpretations can differ.

This section also requires written procedures associated with sanitation and that the facilities should be maintained in a good state of repair. Although it may seem obvious that maintenance should be performed regularly, it can happen that preventative maintenance programs compete with production requirements for attention; however, an in-depth preventative maintenance program should be in place.

SUBPART D: EQUIPMENT

This section addresses equipment design, size, and location, as well as construction, cleaning and maintenance. Similar to the requirements for buildings and facilities, it is necessary to provide appropriate equipment for the manufacture of a product and ensure that the equipment

material of construction is not reactive, additive, or absorptive. In the 1978 version of the GMPs, requirements for equipment cleaning and use logs, as well as written procedures for equipment cleaning and maintenance were added. These requirements aid in the investigation and solution of problems by identifying batches that may also be implicated in a particular problem.

This section also covers the use of automatic, mechanical, or electronic equipment. Additional information relating to compliance with this section can be found in Compliance Policy Guides 7132a.07 (5), 7132a.08 (6), 7132a.11 (8), 7132a.12 (9), and 7132a.15 (10). Compliance Policy Guide 7132a.15 interprets Section 211.68 (b) as requiring maintenance of the program as part of the validation package for computer systems. Reliance on validated computer systems in place of several manual checks has become widely accepted.

SUBPART E: CONTROL OF COMPONENTS AND DRUG PRODUCT CONTAINERS AND CLOSURES

This section relates to the receipt, identification, storage, handling, sampling, testing, and approval or rejection of components and drug product containers and closures, and the requirements for written procedures for each. It also covers the use of approved materials, retesting of approved material, and prevention of use of rejected materials. Although the requirements of this section indicate that each lot be appropriately identified as to its status and that materials in different statuses be stored separately, the implementation of computerized warehouses has made it possible to eliminate status labels and physical separation of quarantined and approved materials. Rejected materials are usually handled separately. These practices are not to imply that a computerized system can be used without appropriate assurance of controls.

The current GMP requirement that materials must be tested or examined for all specifications and released prior to use is in conflict with the philosophy of vendor certification, which is based on a consistent, reliable record of good quality. Only vendors with well-controlled processes and a good record of acceptable batches qualify for such a program. Thus, a material could be put into use based on the quality record of the supplier (vendor), even if testing is only for identification [211.84(d) (2)].

This section also requires the use of oldest approved stock first, retesting of approved stock “as appropriate,” and controls for drug product containers and closures. It

prohibits use of rejected components and drug product containers and closures.

SUBPART F: PRODUCTION AND PROCESS CONTROLS

This section focuses again on the need for written procedures and formal authorization by the quality control unit for any deviation from written procedures. Areas covered are addition of components; calculation of yield; equipment identification; sampling and testing of in-process materials and drug products; time limitations on production; control of microbiological contamination; and reprocessing.

Many drug companies are using electronic means of verifying component names or item codes, receiving and control numbers, weights, or measures, and even the verification of component addition to a batch. There is a range of acceptability on the part of the FDA of electronic means of verification and batch documentation; however, the validation of such systems must be performed to accept electronic means of identification and verification.

The process controls required in this section should be based on process capabilities rather than conforming with a checklist based on the regulations. This would mean that tests not typically used for a particular dosage form may be appropriate, whereas other more commonly used tests may be without any value. This not only depends on the validation of the process but also equipment and process qualification. In addition, the need for microbiological controls can be greatly reduced by knowing whether a product supports microbial growth and whether the environment in the production area is maintained at a sufficiently low bioburden. Clearly, certain products require close attention to the production environment because of the ingredients and the end use.

Many firms use the so-called clean-zone concept, in which the restrictions on personnel entering a production area and the required protective clothing are based on the nature of a product—whether the product is prone to the growth of microbes or whether it is required to be sterile. Even for products not required to be sterile or that are not supportive of microbial growth, this concept controls the production environment through reduction of bioburden.

Reprocessing frequently receives considerable attention from the FDA. Over the years, reprocessing appears to have decreased, not only because of FDA pressures, but also because more products and processes are being validated and better controls are being exercised during production. At times, however, there is the need to reprocess, but it requires authorization of the quality control unit.

For a product covered by a New Drug Application (NDA) or Abbreviated New Drug Application (ANDA), provision for reprocessing must be included in the approved registration document. Although it is not always possible in the filing of an NDA or ANDA to foresee all reasons why a product may need to be reprocessed, a procedure for reprocessing can be evaluated and included in the registration document. If not included in the approved registration document, the regulations require submission of a supplemental application and prior approval in order to market a reprocessed batch.

To some people, batch or lot yield may seem to be more of a business concern rather than a regulatory or technical matter. However, GMPs require that yield tolerances be established and that yields outside of the tolerances be investigated. The need for an investigation is to determine that yields outside of normal limits can be an indication of problems during production that would not be evident with routine testing. A minor deviation may be relatively insignificant and could simply mean that the yield tolerances need to be reevaluated, a procedure that should be followed periodically.

It may seem that the identification of equipment in the processing record is also a superfluous burden. If several pieces of equipment have been shown to be used interchangeably, one might question the reason for this additional documentation; however, when a problem arises, it is necessary to know exactly which equipment was used. It may be possible to trace this back by reviewing equipment cleaning and use logs, but the investigation is simplified by having this information in the batch record. Recording variable batch information concerning the equipment, such as tablet compressing speeds, is also necessary.

SUBPART G: PACKAGING AND LABELING CONTROLS

This subpart covers one of the aspects of pharmaceutical production that has received much attention because of recalls, including an increase in recalls related to labeling errors or product mix-ups associated with the packaging and labeling operation. Specific requirements identified in this section recently include the following.

Section 211.122: Materials Examination and Usage Criteria

- (f) Use of gang-printed labeling for different drug products, or different strengths or net contents of the

same drug product, is prohibited unless the labeling from gang-printed sheets is adequately differentiated by size, shape, or color.

(g) If cut labeling is used, packaging, and labeling operations shall include one of the following special control procedures:

1. Dedication of labeling and packaging lines to each different strength of each different drug product;
2. Use of appropriate electronic or electromechanical equipment to conduct a 100% examination for correct labeling during or after completion of finishing operations; or
3. Use of visual inspection to conduct a 100% examination for correct labeling during or after completion of finishing operations for hand-applied labeling. Such examination shall be performed by one person and independently verified by a second person.

(h) Printing devices on, or associated with, manufacturing lines used to imprint labeling upon the drug product unit label or case shall be monitored to assure that all imprinting conforms to the print specified in the batch production record.

Section 211.125: Labeling Issuance

Section 211.125 has been amended by revising paragraph (c) to read as follows:

(c) Procedures shall be utilized to reconcile the quantities of labeling issued, used and returned, and shall require evaluation of discrepancies found between the quantity of drug product finished and the quantity of labeling issued when such discrepancies are outside narrow preset limits based on historical operating data. Such discrepancies shall be investigated in accordance with Section 211.192. Label reconciliation is waived for cut or roll labeling if a 100% examination for correct labeling is performed in accordance with Section 211.122(g)(2).

Section 211.130: Packaging and Labeling Operations

Section 211.130 has been amended by redesignating paragraphs (b), (c), and (d) as paragraphs (c), (d), and (e), respectively, and by adding a new paragraph (b) to read as follows:

(b) Identification and handling of filled drug product containers that are set aside and held in unlabeled condition for future labeling operations to preclude mislabeling of individual containers, lots, or portions of lots. Identification need not be applied to each individual container but shall be sufficient to determine name, strength, quantity of contents, and lot or control number of each container.

These requirements reflect an increased use of electronic means to ensure correct labeling and tight controls on the practice of filling containers that will be labeled at a later date. A time-consuming operation required in the current GMPs is associated with the reconciliation of labels. The recalls and associated investigations demonstrate that unless 100% accountability can be achieved in the reconciliation process, there will not be an effective means of ensuring correct labeling.

Section 211.132 was revised on February 2, 1989, to describe tamper-resistant packaging and labeling requirements for over-the-counter (OTC) human drug products. Compliance Policy Guide 7132a.17 (11) was issued in 1992 to describe the standardized tamper-resistant packaging requirements. This section also covers information concerning requests for packaging and labeling exemptions. It allows changes in packaging and labeling to comply with the requirements for OTC products subject to approved NDAs to be implemented prior to FDA approval as provided for in Section 314.70(c). Manufacturing changes to provide for sealed capsules require prior FDA approval under Section 314.70(b).

Section 211.132 states that none of the requirements for “special packaging” (child-resistant packaging), as defined in Section 310.3 (1) and required under the Poison Prevention Packaging Act of 1970, are affected.

Subpart G also covers drug product inspection and expiration dating. The expiration date that is required in Section 211.137 relates to stability studies performed on the drug product described in 21 CFR 211.166. It requires that expiration dates be related to storage conditions stated on the product labeling.

Furthermore, the programs established are to use stability-indicating methods, under controlled conditions, in the marketed container–closure system and on an adequate number of batches to determine the appropriate expiration date. The FDA has issued guidelines on stability testing which outline in more detail the requirement to establish a stability program to determine and support the expiration date of a product (4a, 4b, 7). A new draft guidance for stability testing was published by the FDA in 1998 (19), and discussions with comments to finalize this guidance are still continuing.

SUBPART H: HOLDING AND DISTRIBUTION

This section covers warehousing and distribution and the procedures required for the quarantine of drug products before release by the quality control unit, storage of drug products under appropriate conditions, procedures to ensure use of the oldest approved stock first, and a system for documenting the distribution of each lot of drug product. This is another area where computerized systems are being used extensively. During inspections, the FDA review includes evaluation of the validation of any computerized systems and controls.

SUBPART I: LABORATORY CONTROLS

This entire section refers to the requirements covering the testing of drug products and their components prior to release for distribution. It also covers stability testing and special testing, including testing for penicillin, if a reasonable possibility exists that a nonpenicillin drug product has been exposed to cross-contamination with penicillin and laboratory animals. Additional information can be found in the Good Laboratory Practices, 21 CFR 58.

Reserve samples are required to be maintained for active ingredients and drug products. These specific requirements are elucidated in 211.170. The section on reserve samples also requires that a visual inspection of reserve samples of drug products be conducted at least once a year for evidence of deterioration.

Fundamental to the testing requirements is the need for validated methods with established and documented accuracy, sensitivity, specificity, and reproducibility. It is also necessary to have meaningful sampling and testing plans that meet statistical quality control criteria. Judgments made with regard to sampling procedures should be based on the quality of the process control or the reliability of the vendor who supplies a raw material, drug substance, or packaging component.

SUBPART J: RECORDS AND REPORTS

This section details the records and reports required to be maintained for pharmaceutical drug products, their components, and the equipment used in the processing of a drug product. Through these records, the entire history

of a batch can be traced. The records cover equipment cleaning and use logs; component, container, closure, and labeling records; master production and control records and production record review; laboratory records; distribution records; and complaint files. Because this amount of recordkeeping can be voluminous, Section 211.180(d) allows for microfilm, microfiche, or other accurate reproductions of the original records for storage. Many firms are using the electronic generation of batch and analytical records. It is important to be able to retrieve all of the above records easily during an FDA inspection. Electronic methods must be supplemented with proper procedures to ensure that the records do not deteriorate over a period of time and can be retrieved when the computer systems used to generate the records have been revised or replaced.

This section also requires a master production and control record for each product, from which the batch production and control records are generated. These records must include complete instructions concerning the manufacture of a batch and precautions to be followed. Prior to the commercial distribution of a drug product into interstate commerce, all executed production and control records must be reviewed. If there is a discrepancy or a failure of any batch or any of its components to meet specifications, there must be an investigation and a written report of the findings. The investigation are to extend to other batches of the same or other drug products that may have been associated with the out of specification batch or discrepancy.

Another part of this section covers complaint files, which are reviewed regularly during FDA inspections. In fact, a complaint file review may be the sole reason for an inspection if the FDA receives a complaint directly from a pharmacist, which may be a cause for concern. Sometimes the FDA will visit a firm to follow up on a complaint, even though the firm may not have been informed by the complainant. In the event that a complaint is received by a firm, it should be evaluated and a response sent to the complainant. It may be necessary also to conduct an investigation and prompt further action regarding the product or batch in the marketplace.

SUBPART K: RETURNED AND SALVAGED DRUG PRODUCTS

This section requires that extensive records be maintained on returned drug products including ultimate disposition.

Again, if the reason that a drug product is returned implicates other batches, an investigation is to be conducted in accordance with 211.192. Drug product salvaging is not allowed for drug products that have been subjected to improper storage conditions. If there is a question as to whether drug products have been subjected to such conditions, they may be salvaged only if there is evidence from laboratory tests that all applicable standards of identity, strength, quality, and purity have been met. In addition, evidence is required from the inspection of the premises that the drug products and associated packaging were not subjected to improper storage conditions as a result of a disaster or accident. Understandably, the value of the material to be salvaged is taken into consideration when such rigorous requirements exist for salvaging.

CONCLUSION

Compliance with cGMP requires that responsible employees in a firm be knowledgeable about the practices that other firms follow in order to comply. FDA investigators visit many firms and find a broad picture of current manufacturing and control practices. Thus, Current Good Manufacturing Practices are “state of the art,” constantly changing. To be in regulatory compliance, a firm must review their procedures and systems regularly and revise them as necessary.

REFERENCES

1. Vickory, H.; Nally, L. Good Manufacturing Practices: An Overview. *Encyclopedia of Pharmaceutical Technology*; Marcel Dekker, Inc.: New York, 1993; 7, 109–120.
2. Berry, I.R., Nash, R.A., Eds. *Pharmaceutical Process Validation*; Marcel Dekker, Inc.: New York, 1993.
3. *Code of Federal Regulations*; Title 21, Parts 210 and 211, U.S. Government Printing Office: Washington, DC, 1999.
4. *Compliance Policy Guides*, Requirements for Expiration Dating and Stability Testing (1995). 7132a.04.
5. *Compliance Policy Guides*; Computerized Drug Processing; Input/Output Checking, 1987; 7132a.07.
6. *Compliance Policy Guides*; Computerized Drug Processing; Identification of “Persons” on Batch Production and Control Records, 1987; 7132a.08.
7. *Compliance Policy Guides*; Lack of Expiration Date or Stability Data, 1995; 7132a.10.
8. *Compliance Policy Guides*; Computerized Drug Processing; CGMP Applicability to Hardware and Software, 1987; 7132a.11.

9. *Compliance Policy Guides*; Computerized Drug Processing; Vendor Responsibility, 1987; 7132a.12.
10. *Compliance Policy Guides*; Computerized Drug Processing; Source Code for Process Control Application Programs, 1987; 7132a.15.
11. Tamper-Resistant Packaging Requirements for Certain Over-the-Counter (OTC) Human Drug Products. 1992, 7132a.17.
12. Berry, I.R.; Harpaz, D., Eds.; *Validation of Bulk Pharmaceutical Chemicals*, 1st Ed.; Interpharm Press: Buffalo Grove, IL, 1997.
13. Berry, I.R.; Harpaz, D. Eds.; *Validation of Active Pharmaceutical Ingredients*, 2nd Ed.; Interpharm Press: Englewood, CO, 2000.
14. *Guide to Inspections of Bulk Pharmaceutical Chemicals*; U.S. Government Printing Office: Washington, DC, 1991.
15. *Guide to Inspections of Sterile Drug Substance Manufacturers*; U.S. Government Printing Office: Washington, DC, 1994.
16. *Guide to Inspections of Foreign Pharmaceutical Manufacturers*; U.S. Government Printing Office: Washington, DC, 1996.
17. *Guidance for Industry: Manufacturing, Processing, or Holding Active Pharmaceutical Ingredients*; U.S. Department of Health and Human Services: Rockville, MD, 1998.
18. ICH Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients Draft, International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use, London, UK, 2000.
19. Stability Testing of Drug Substances and Drug Products. *Compliance Policy Guides*; Draft Guidance U.S. Department of Health and Human Services: Rockville, MD, 1998.

GOOD LABORATORY PRACTICE (GLP)—AN OVERVIEW

Nigel J. Dent

Country Consultancy Ltd., Northamptonshire, United Kingdom

OVERVIEW

The Good Laboratory Practice Guidelines (GLP) have been in existence for nonclinical safety studies since 1976 (1–8). They have progressed through various transitional phases to become guidelines in some countries and regulatory/statutory instruments in others.

The current document is the Organisation for Economic Co-operation and Development (OECD) Principles of GLP (9) and is currently accepted as the industry standard. This was reviewed and published in January 1997 (10, 11) but must be used in conjunction with the appropriate Scientific Guidelines for the scientific side of the study, i.e., the OECD Toxicology Guidelines etc. (12). This sets out to cover all nonclinical safety studies and gives guidance as to how these studies should be conducted in conjunction with the appropriate regulatory toxicology guidelines and, on that basis, when encompassed in the various Directives of the European Union (EU) (13) or in other Memorandum of Understanding, allow data generated under this program to be mutually accepted by other OECD countries. One must not forget, however, the other equally important Guidelines and Regulations of other countries, such as those of the U.S. Food and Drug Administration (FDA) (14) and the U.S. Environmental Protection Agency (EPA) (15–20), and similar organizations in Japan (21). All have basically similar rules and, being members of the OECD, data generated to the OECD principles will generally be accepted in the United States and Japan. The EPA regulations used to be quite different and were applied to agrochemical and pesticide products. However, having been revised recently, they have been brought in line with the documents of other agencies.

The guidelines themselves, with the exception of those in countries where they are featured as regulations, are, as stated, guidelines to the conduct of the study and aim to cover compliance with the GLP principles but in no way do they dictate how the science will be performed.

It must be remembered that compliance is monitored by adherence to GLP, whereas the regulatory authority and the receiving authority of the dossier when submitted for the application of a marketing permit or similar document review the science.

OBJECTIVE OF THE GUIDELINES

The general objective of the guidelines originates in the very early 1970s, when one pharmaceutical company in particular and a contract research organization (CRO) generated data that, when submitted to the FDA, gave them cause for concern in the accuracy of the data presented and, in certain instances, the honesty of the submission.

At that time, a full review of companies and institutions conducting nonclinical safety studies (toxicology) was undertaken by the FDA (22) and although, in general, the industry was found to be credible, one company (Industrial Biotest) was found to be generating extremely poor-quality data, in many instances, in a fraudulent manner. This, therefore, caused the FDA to put together and implement the GLPs (23, 24).

Over the next 10 years, many countries introduced similar good-practices guidelines. The EU (25) in general produced its guidelines and eventually, despite the fact that the world was operating according to similar principles, a standard document was produced by the OECD in the early 1980s and became the industry standard. The reason this was of benefit to the whole industry was because this now precluded the fact that every submitting company would have to be inspected by each relevant monitoring authority and, when implemented into several directives and legal statutes (26, 27) within the OECD (particularly in the United Kingdom) (28), this allowed data generated by one company to be accepted by several receiving authorities without further inspection.

The objective of the GLPs is to ensure that a standard approach is undertaken covering traceability and accountability and, while still allowing freedom for the scientists, to impose certain restrictions on the generation of data and the experimental work.

It must be remembered that GLP is merely common sense in a formal environment. The key phrases that are currently seen in a GLP environment include good documentation, good training, maintenance and calibration of all equipment, the archiving and storing of data in a formal and retrievable manner, and the use of high-quality, validated equipment and accredited test systems (animals).

This in general is merely good science, and the GLPs have further enhanced this by the addition of an independent Quality Assurance Unit (QAU) and a study director/principal investigator who jointly controls and oversees the project and involves the management in putting together adequate resources and assuming overall responsibility for the study.

This can be seen as good science, with several slight enhancements. The details of these individual subjects are addressed later in this article.

WHO DOES IT AFFECT?

Any company or institution performing nonclinical safety studies for the submission of data for a new chemical entity; a new biological, immunological, pesticide, veterinary or agrochemical product; or, for that matter, a similar product that will eventually appear in the marketplace and be consumed by the general public must adhere to GLP in the conduct of their nonclinical safety study experimentation.

Within a company, every person from senior management to the junior technician is bound by these GLPs and must exhibit clear understanding and training in these practices.

To ensure that the practices are followed, a regulatory inspection takes place on a 2-year basis in most countries, and the objective of this is to review, as an independent group, how these good practices are being followed. Certification or a guarantee that the company is operating according to these standards is the benchmark standard. This is also addressed later in this article.

As we move into the twenty-first century, it is quite apparent that the industry will shrink as mergers and acquisitions (29–33) take place, and, with this, the emergence of the now familiar CRO will become ever more popular in the conduct of nonclinical safety studies. It is, therefore, very important that, in this area, the sponsor has the assurance that these facilities are operating not only to the highest standard of science but also in compliance with GLP and that, as a subcontractor, the data they generate will be equally accepted as if the data were generated by the company itself.

WHY HAVE IT?

In general terms, for companies conducting nonclinical safety studies (34), it is, a regulatory requirement, and without this certificate or certification of compliance, data

will generally not be accepted by the receiving/regulatory authorities.

However, one should not embark on the process of obtaining or working to GLP with this sole aim in mind. It should be used as an ongoing improving and quality standard for the laboratory.

In fact, it is the author's experience over the past 5 years, that many companies have gone far beyond the requirements of GLP compliance (35) and that the overall concept of good scientific design and good science has been superseded by the desire merely to obtain compliance. It is quite often seen that an extremely poor quality scientific study has been conducted in complete compliance with GLP. It has been seen on several occasions in a laboratory, for example, where the refrigerator has been located far from its permitted limits; where the temperature has been diligently recorded, signed, and dated as required by GLP, but where no attempt has been made to either document the excursions outside the accepted range or to rectify the problem. The operative was merely under the impression that as long as temperature is recorded, this is GLP despite the damage that excursions outside the temperature range may have caused to any investigational product stored in the refrigerator.

Over the years, those scientists who have worked according to the principles of GLP now readily admit without any prompting that they are unsure how they conducted scientific studies before the advent of these good practices. The ability to reconstruct studies, to work to a standard format across several differing laboratories or countries, and to be able to prove beyond reasonable doubt that these were the values obtained and the results submitted.

Certainly, data with a GLP compliance statement are being accepted more readily by the receiving authorities, which has led to fewer repeated studies. This, in turn, is helping to achieve the aim of all scientists in reducing the use of animals.

From a company's point of view, working according to the principles of GLP shows that it has an attitude that is both ethical and moral to the production of scientific data with products that will eventually enter the human food chain or be of benefit to mankind.

HOW IS IT ENFORCED?

In the OECD countries, for at least 14 years, an Inspectorate has been set up, varying in inspector numbers from several hundred in the United States to one or two in

countries not conducting a great deal of scientific nonclinical research. All countries, however, have a regulatory group that, in some instances, also acts as the receiving authority for the review of data, and reports to the GLP Monitoring Authority. This regulatory group visits on a 2-year basis or, in Germany, a 4-year basis, those companies that have claimed compliance and will then be on a rolling program of review (36).

Unlike its role in many areas of regulatory compliance, it is still the responsibility of the sponsoring company to claim compliance from the Monitoring Authority. This claim is made for a particular company, laboratory, and/or series of tests. From the date of compliance when a letter is written to the Monitoring Authority, data generated from then are assumed by that company to be in compliance with GLP. This claim is then verified in a visit from the regulatory inspector. The inspection may be performed by one or two persons for 1 to 5 days. At the end of the inspection, an exit meeting is held, and the company is usually given an indication of its performance. Noncompliance points are noted in writing and discussed, and a report is then prepared. In view of the findings, three levels of compliance can be obtained:

1. Sufficient deviations have been seen to question the integrity of the data and, therefore, a complete rejection of the claim of compliance is made, with a revisit necessary.
2. Minor points of compliance have been seen that can be handled in a specified period in which case, the laboratory is placed under the category, of pending compliance.
3. Very minor points of compliance have been seen, which, when addressed in writing by the management in a 1-month period with supportive paperwork, etc., lead to the company being given a Statement of Compliance, a Certificate of Compliance, or an indication that the laboratory is in compliance. It depends on the specific country whether a Certificate of Compliance is given. If a certificate is given, it generally states that on the particular day that the inspection took place, the laboratory was found to be in compliance with the OECD Principles of GLP. Also, the address of the facility is given as well as a listing of areas in which compliance has been confirmed. This could be stated as *analytical support facilities, acute toxicology, mutagenicity*, or similar designations.

Naturally, the benchmark standard is either the OECD Guidelines or similar standards in Japan or the United States. The Inspectorate carries out inspections against these documents. It could be said that often it is merely a

review of the procedure and an opinion of compliance given by the inspector versus the interpretation of the individual conducting the experimental work. To try to overcome this criticism and to ensure that all inspectors work according to a standard format, over the past 4 years, the OECD has instituted a series of mutual joint visits (MJVs) (37).

The process of an MJV is that a company is inspected by its local inspector and that the inspector is accompanied by inspectors from two other countries as observers. At the conclusion of the inspection, the company is given their findings by its local inspector and, outside that meeting, a review of the performance of the inspector with positive and negative points is given by the two observing inspectors. To ensure continuity, one of these three inspectors would then be on the next MJV.

In the past, Memorandums of Understanding (MoUs) (38) have been instituted between certain major countries, such as Japan and the UK, the United States and Japan or Canada, etc. However, these have generally fallen into nonuse for a variety of reasons, especially in Europe, where it is now, or has been for some time, not possible for a country to negotiate directly with another country. Brussels, however, being the center of the European Community, has to carry out that discussion with a proposed partner in another country. As such, at the time of producing this overview, very few MoUs are currently in force.

WHAT IS GLP?

As noted in the Overview, GLP is a series of guidelines that cover the conduct and data production for nonclinical safety studies.

The OECD covers a series of activities and personnel. Responsibilities, training, quality assurance (QA), standard operating procedures (SOPs), study plans and study reports, data production and recording, equipment maintenance and calibration, computers and validation, test systems and test substances, and archiving are the primary areas covered by the GLPs.

A very brief overview of each of these areas is given hereafter.

Responsibilities

The prime players in a GLP scenario would be the management, the sponsor, the study director, the principal investigator, and the QA.

In a hierarchical structure, management would be totally responsible for the conduct of the work and for the

assurance that resources have been made available and that an active role is played by these people in overseeing the conduct of scientific research.

The sponsor is the company that places a contract with a CRO or requests from within a company that work in another department be undertaken. The sponsor is the person who is supplying the money and the request for the work.

The study director is the prime player and is ultimately responsible for the production of the study plan, the conduct of the study, and the overseeing or production of the final report. Naturally, a large amount of delegation may take place; however, this must always be in writing, and the overall responsibility for the conduct of the study; the daily contact with the study staff; the prevention of recording of problems and the assurance that the study has been conducted in line with the study plan, the GLP, and the scientific guidelines solely belongs to this individual.

The Principal Investigator is the next in line of responsibility after the Study Director in a multisite study. For example, they could be the person seen in a field study situation where the crop-spraying, for example, may be undertaken at a place remote from the GLP designated site where the study director works. The principal investigator is therefore the person responsible initially for that portion of the work, although under the direct control of the study director. It may also be that, within a company, work is subcontracted to the Analytical Department, for example, for the analysis of formulated material. The person responsible for this particular aspect of the scientific work is the principal investigator, who is involved in the study plan and responsible to the study director. Another typical scenario is work conducted in a CRO under the control of the study director, where samples of plasma are taken for toxicokinetics, for example, and these samples analyzed by the sponsor. The sponsor's analyst, therefore, may well be designated the principal investigator.

Quality assurance

This is an independent group that does not become involved in the conduct of the study but merely reviews the data, experimental work, and documents produced to ensure compliance with the SOPs, the study plans, and GLP. Other activities such as training and assistance in interpreting GLPs, etc., may be the responsibility of the QAU. (For more information, see the following section in this article or Section II, 1.4 pp. 17–19 of the OECD GLPs.)

Training and recording

It is the responsibility of the management to ensure that training takes place and the responsibility of the Study

Director to assure that the individuals conducting the work are adequately trained and have adequate records. At a minimum, there must be a CV, a training record, and a very clear job description. Specifically, with regard to a study director, there must be explicit details of how the study director position can be met and the responsibilities of that individual in carrying out the relevant duties.

There should be procedures detailing how the training will take place; recording of the training must be made on a regular basis, the records must be stored in archives and regularly updated, and a complete and historical review of the trainee's activities, previous training, and ability to conduct the work according to GLP must be documented. (See the OECD Principles, Section I, 2a–d.)

Quality Assurance

This function, as has already been stated, is an independent review. The responsibilities here start with a review of the study plan and continue through the review of the study in the in-life phase, data audits, and the final study report audit.

In addition to these, systems audits and process audits can be undertaken.

The aim of QA is to assure the management that compliance with GLP is maintained throughout the entire study, that the data integrity is maintained, and that compliance with the SOPs and the study plan is adhered to by all experimental study staff. The study audit is a specific audit of the study in direct relation to the study plan. A systems audit, however, rather than proceeding in a vertical line, takes a horizontal line across all studies and would include such tasks as archiving, training, SOPs, general computer validation, animal house operation, and management activities. These are but a few areas that would constitute a systems audit but, hopefully, gives an idea of the type of activities across studies that would be audited.

Process audits, on the other hand, have specifically been addressed in the revised 1997 GLPs, and these are basically aimed at auditing short-term studies of a repetitive nature, generally undertaken by similar teams of people. Here, that the system is working and that parts of the process are reviewed over a quoted period in the QA SOP are assured. The aim is to ensure that all critical aspects of this process are reviewed through different studies over a period of time. This, then, does not necessitate QA review of all short-term studies on every occasion, nor does it require the review of such areas as analytical analysis on a batch-by-batch basis or the analysis of hematology or biochemistry samples each time these come up for analysis.

QA itself is required to produce SOPs that clearly detail operation, method of selection of critical phases, and studies and to report its results to the management.

After every audit, a report is produced that is then discussed with the study director and circulated to the management with the overall agreement from the study director relating to the audit findings and their explanation of the resolution. (For an in-depth review of QA, see Principles of GLP, Section II, 2.1–2.2, p. 20.)

Standard Operating Procedures (SOPs)

These generally have been likened to a complete documented history of the entire aspect of conducting nonsafety studies. Any activity needs to be described in one of these documents. There may be a compilation of activities, or they may address single items such as the calibration and use of an electronic balance. They must be produced by the individual most familiar with the task, agreed on by the management, and countersigned by a person senior to the author.

Once produced, SOPs must be reviewed on a regular basis, (approximately every 2 years) and any changes to these procedures must be made in writing, with the agreement of all parties and circulated to each owner or user of the SOP. The SOP itself must be filed in the archive and additional copies produced. An SOP management system must be set up, whereby a responsible person knows the whereabouts of all SOPs and can retrieve and replace them with amended or superseded revisions and can make sure that they are reviewed regularly and disposed of when no longer required.

The SOP must appear immediately in the area adjacent to the workplace to be readily available to all persons. Frequently, SOPs are the basis of training, and most companies now have SOP-based training schemes. The content and receipt of the SOP should be acknowledged immediately on receipt and a training program set up whereby confirmation of the understanding and the ability to perform the duties stated in the SOP is documented in the appropriate training record.

SOPs should be adequately controlled to prevent unauthorized photocopying, which may lead to the possibilities of a superseded copy not being administered to the known recipients. Someone making an illegal photocopy would not be on the distribution list and, therefore, would not always receive updated versions, with the possibility that an outdated method could be used.

The requirement for archiving historical copies is one of the key attributes of GLP in that traceability can be seen as originating at the archive. The dates and historical

record of the SOPs can prove irrevocably that a particular action was the method in use at the time.

SOPs can be paper-based or electronic. The trend toward electronic record-keeping is becoming more common in laboratories. The only requirement made by the inspectorate is that accurate, controlled copies are available on the electronic media and that prevention of copying or unauthorized changing are built into the SOP system. (See GLP Principles Section II, 7, p. 24.)

Study Plans and Reports

Before any study can be undertaken satisfactorily, a study plan must be produced. The study plan is merely an indication of all of the activities that will take place, resources required, time frames, and objectives. The study plan can be likened to a road map that, when given to all the participants, will allow them to start at the beginning and to proceed through the various mazes to the final completion point indicated by the study report. The one golden rule in GLP is one study plan, one study director, and one report.

The study report itself is a mirror image of all the headings in the study plan and serves to confirm that the objectives of the study have been met and that the results and discussions of the data presented give an indication of the outcome of the particular experimental work.

Both the study plan and the report are audited by QA, and each study plan and report are generally determined by the company's format.

Typical headings must be given in both documents. (These can be found in Section II, 8.1–8.3, pp. 25–27 and 9.1–9.2, p. 28, of the OECD Principles of GLP.)

Data

Raw data, or source data, are generally considered the first records made, either electronically in computer-readable form, or records created the first time that the “pen hits the paper.”

These should be original signed and dated recordings that may be on any type of media. Cases in which media such as heat-sensitive paper contain the result, should be photocopied in the event of deterioration over a time.

Electronic data can be regarded as the disk, CD-ROM, or similar media provided that this material, when reintroduced to the computer and the software, can generate the images stored on the disk or electronic media in a 100% readable form.

There are many types of electronic media, machines, and source data or raw data within the toxicological environment. However, ironically, the most common storage media and the most common raw data are paper.

Paper and its storage partner, microfilm, have been around for many years, and their stability and reproducibility are well known. Other electronic media, however, do not have the same capability of reproduction known over a long period, and, thus, most industries and companies prefer paper.

In regard to the data they should be recorded promptly, legibly, and signed and dated, and any corrections should be made in a format to allow the original record to be seen, the change described and justified where applicable, and the change signed and dated by the individual making the revision. This procedure, whether on paper or via computer, should have the same standards. With use of the computer, an audit trail is necessary to identify the change and the person making it, along with the reason.

Equipment

As can be imagined, equipment in a toxicological study may be varied, simple, or complex. As such, it is difficult to describe each individual type of equipment in this limited space.

GLP requires that equipment be maintained, calibrated, and generally demonstrated as fit for use.

Equipment such as high-pressure liquid chromatography (HPLC) should have system-suitability checks, installation qualifications, and operational qualifications performed at a minimum.

Other equipment such as centrifuges and balances should be maintained and calibrated and, with regard to the latter, regular checks should be made with known, standardized, regularly calibrated weights. These should be placed on the balance with a frequency to guarantee that data from the machine are accurate. Even if the balance is an electronic calibrating balance, regular manual check weights should be applied.

Each piece of equipment should have a log book that gives a historical record of its use, breakdown, repair, and service. Generally, it is acceptable that these log books be placed by the equipment generating critical data to be presented in the final report.

All equipment should be clearly identified as to the time that it started producing raw data for experimental use and, when no longer required, the equipment should be removed from the laboratory or suitably labeled “not for GLP use.”

The calibration and validation of equipment have been addressed extensively but, equipment that can be shown as “fit for use,” within the GLP environment is generally acceptable to most regulatory inspectors. (For additional information on the GLP requirements for equipment, see Principles, Section II, 4, p. 22.)

Computers

Over the past few years, computers have played a very important role in many aspects of toxicology. The general trend in the industry and particularly from the Inspectorate is to ensure that they are fully validated.

Validation, however, means different things to different people. Some companies and their Information Technology Group (ITG) will dismantle the computer and its software components, reconfigure them, test them, and then reinstall them. Others will take a more realistic approach and work on the basis that the computer was brought in for a specific task and, is considered validated provided that task is completed with the aid of the computer in a reproducible and acceptable manner.

However, in the most simplistic form, validation could be covered by “evidence that the computer will perform the task for which it was purchased and, more importantly, continue to perform that task for the foreseeable future.” In other words, as with other equipment, is the computer fit for purpose?

Several documents have been written from a regulatory standpoint, the most useful being Monograph 10 of the OECD Principles, *Application of GLP to Computer Systems* (39). Many books are available and vary in detail and content to cover everything that one would wish to know about computers, but were afraid to ask, down to the simple documentation giving the essentials for validation and providing a disk with the SOPs to comply with GLP! Several examples are listed in the References (40–42).

The prime concern of the Inspectorate is that the user responsible for performing the validation and producing the report is in control of the equipment and can ensure and prove the integrity of the data when entered into the computer and regenerated in some other form.

It is generally accepted in the industry that acceptance testing is perfectly satisfactory for most computers and assures that the computer, when installed on company premises, will perform the function for which it was purchased. However, each computer must be viewed in the role it will play in the company and suitable testing must be conducted to ensure that the data and integrity are of the highest quality and that total control over output is maintained.

As with all equipment, computer maintenance and calibration records are of paramount importance. If in-house software programs are produced, they are tested and validated, and the source code is made available. One of the key elements required in the computer record-keeping is that of change control and password protection and training. (See GLP Principles Section II, 7(b), p. 25.)

One very important rule is that *the electronic signatures rule* (43), and it must be observed when data are signed off electronically. This FDA requirement became effective August 20th, 1997, and covers all data for which signatures are made electronically and requires that the FDA is officially notified.

Test Systems

This really is a slightly complex name for what is generally considered the animal subject. Test system, however, has been utilized because in many instances in toxicology, GLP now applies to such subjects as ground water, soil, insects such as earthworms and honey bees, and microorganisms such as daphnia and, therefore, the use of the word *animal* is not always applicable.

The main criteria are that the origin of the test system is known with its breeding history, where applicable, that these are purchased from well-known and, if possible, accredited suppliers, and that the quarantine period is observed to ensure that test systems are of high quality and fit for use.

Care, husbandry, intermediate sacrifice if the test system is found to be “in extremis,” and humane sacrifice before necropsy are essentials for the test system. Separate housing among species and experimentation is critical, and all aspects of manipulation of the animal from clinical observations, dosing, and special tests such as electrocardiogram (ECG) need to be well documented and outlined in SOPs.

Animal husbandry itself, the animal room, and the animal room diary giving an indication of exactly what occurred in the room and to the animal are essential items of documentation. Unique identification is also of paramount importance with the animals, cages, and the location of the cages.

Full and documented history of heating and ventilation are required and, in barrier-maintained rooms, signed and dated records of positive to negative pressures are to be kept. These records should also reference any malfunction and its rectification. Furthermore, *excursions outside the permitted range* must be documented, and the effect on the study and data integrity must be identified and addressed by the study director in the final report. (Additional information is available in the GLP Principles, Section, II, 3.2, p. 21.)

Test Substance

In most instances, the test substance can be the new chemical entity (NCE) or an existing product; a comparator, pharmaceutical, veterinary, or agrochemical product; or even a device. Knowledge of the composition,

characterization, stability, and other physiochemical properties is essential. Stability, however, may be determined as the short-term studies progress, with the proviso that the overall stability is known, along with full characterization, by the time long-term toxicity studies are carried out. Stability testing may well be carried out in parallel as long as the stability of the active ingredient and formulated product is known sufficiently to allow for control of the dosing to be done within the period of stability known at that time.

One of the key elements of test substance control is accountability. A record of the amount received for toxicity testing should be accurately recorded, and 100% accountability of that product throughout the life of the testing is an essential element of GLP.

Again, formulation of the product is required to be covered in detail, and, in many companies, the elements of GLP are the benchmark standards when dealing with test substance. Use of the test substance in the animal facility, the maintenance of homogeneity of suspensions, the mixing of the product in feed, and the testing of the product are all essential. This is one particular area in which within the toxicology testing area, support functions such as analytical studies then come under GLP. These functions will be required to test formulations and feedstuffs, etc. to ensure that the correct amount of the active ingredient is present as determined by the study plan for the various dosing groups. This requires that a validated method be available before any work is carried out, with the ability to analyze samples of the formulated product or plasma samples for toxicokinetics as the study progresses.

It is required that a reserve or retention sample of the active ingredient is retained. This should be retained for as long as it affords reasonable testing and within the expiry period determined by the analytical facility. It is also required that a retention sample be retained for studies that are not considered to be short term. This is one particular area in which revision of the GLPs is sometime not well-thought through. Originally, it had been stated that reserve samples should be taken for studies exceeding 4 weeks. This was subsequently revised to specify “studies that are not considered to be short-term.” The glossary in the GLPs defines a short-term study as “a study of short duration with repetitive processes.” This, one must admit, does not give a lot of guidance!

(See additional information on systems in GLP Section II, 3.3, p. 21 and 6.1–6.2, p. 23.)

Archives

Having addressed all the various aspects of the study, one can see that much documentation, tissues, slides, and wax

blocks could well be accumulating. The requirement is to store this material for “a period of time.” Again, very little guidance is given in the GLP, and one is referred to the national guidelines for the storage of data. However, it is of great importance that this material is maintained in good condition in a retrievable format for at least 15 years for or 2 years past the availability of “the product,” whichever is longer.

Generally, companies themselves are maintaining that material for far longer, or for 2 years past the availability of the product.

All the material must be retained in a secure location for easy access, under the responsibility of a management-designated archivist and deputy. The security aspect of the archive should preclude damage from outside sources, fire, water, rodents, etc.

The entire aspect of archiving is basically one of common sense, and guidance on how to archive these materials can be obtained from government agencies that store personnel records or from libraries. [Additional information on archiving is available in GLPs under Section II, 3.4, p. 22 and 10, p. 29 and in the British Standard for Archiving (44).]

HOW CAN COMPLIANCE BE MAINTAINED WITHIN A FACILITY?

Compliance, having been granted after an inspection, should be monitored on a daily basis. However, it is the author's opinion that many companies standards of compliance relax after the initial certificate has been granted only to find that, 2 years later, for example, an enormous rush 1 month before an announced inspection is required to generate the appropriate documentation and to update the system. It is suggested that QC reviews be carried out on a regular basis to ensure that points likely to detract from the overall compliance are reviewed regularly and that project meetings be held where QA is invited to give a précis of the regular points seen during audits so that these can be addressed and rationalized.

Training and retraining, along with an awareness of the requirement to comply with GLP, are of immense importance. New equipment, major SOP revisions, and transfer of technicians or scientists among departments are always good signs that additional training be carried out. It must be remembered that GLP is team work. It is no good considering that there are the scientists and technicians on the one hand and, QA on the other. There is also no point in considering that whatever

happens and however little QC is carried out, QA will discover all the mistakes in the final report and review. Remember, it is not QA's problem; that department's role is to ensure that compliance has been maintained; QC and data-checking are the responsibilities of every member of the staff team.

Improvement targets should be set in line with quality-control manuals used in other accreditation systems. It is always a good point to review internally and on a regular basis: 1) problems that have been encountered in experimentation; 2) audit findings; 3) ways to improve work by looking at new systems and reviewing SOPs to ensure that these are current; and 4) and areas where improvements can be made.

An example can be taken from the accreditation systems (45), in which, in addition to QA audits, departments become involved in self-inspection. Each department can identify a QA representative whose daily responsibility is to review compliance issues, to look at the overall quality policy of the company, and to ensure that between QA audits, self-inspection is performed and that a departmental review is made of these findings with action points and a time plan identified.

The primary impetus for the maintenance of compliance, however, is the regular external inspection by the Inspectorate. In addition, it is now becoming frequent for independent consultants to be brought in to do preregulatory inspections. Whichever way one views the system, whether through consultation or by assigning a department to perform inspections in one area and to conduct audits in another, the regular review of compliance should be maintained.

When using CROs, the whole aspect of auditing takes on a different light. Here, subcontracting is usually performed because of internal pressures, shortage of space, or lack of in-house expertise. Dealing with CROs is no different than setting up an in-house GLP system. The CRO should be regarded as an extension of the facility in which the sponsor is conducting its own research.

PITFALLS AND BENEFITS

In conclusion, it is worthwhile to address the pitfalls and benefits of operating according to the principles of GLP.

A pitfall could be seen as a restriction on the scientist against performing free research. It could also be seen as an intrusion by an independent body looking at why problems occur and at the sorts of problems that occur and carrying out regular reviews with senior management about these problems. Costs will increase because of time pressures

and the necessity of involving third-party reviews. The recording of data will now be subject to more QC, more required approvals, extra costs, and, generally, more data presented. Time must be taken to write and review SOPs. This in itself can be a very costly exercise; the author knows of one company that, having spent more than 6 months writing its SOPs, classed them as capital pieces of equipment and put a value of \$5500 on that volume.

Other companies and personnel may encounter similar pitfalls. The list is not intended to be exhaustive but merely to indicate areas in which additional time, money, and resources will be allocated. However, on the positive side, benefits can be seen immediately.

In talking to many people who have operated under the GLP system for the past 20 years, it is generally heard that the system allows for a better standard of research, less repeated work, the ability to have full accountability and traceability of everything within the experimental phase, and the knowledge that all documentation produced at the end of the study is now safe and secure in the archive and can be readily accessed for regulatory review or inspection.

Fewer studies are being repeated, and, therefore, the immediate benefit is the lowering of subject usage.

The fact that data, when generated with a certificate or a compliance statement, will now be accepted by all OECD member countries means that once the study is completed and the regulatory submission made, the time for acceptance several countries (if submissions are made in a multistate procedure) will be reduced dramatically.

Finally, it is considered that the initial bureaucratic straitjacket of GLP when thrust on the international research community in 1976 has rapidly turned full circle and now is seen as the quality standard to which all companies in all countries want to aspire.

From that point of view, all nonclinical safety studies, when conducted according to the principles of GLP and adequately addressing science as well as compliance, can achieve a very high success rate both in the outcome of the science and in the acceptance of data for a regulatory submission.

REFERENCES

1. U.S. Food and Drug Administration (FDA), Nonclinical Laboratory Studies. Proposed Regulations for Good Laboratory Practice Regulations. Proposed Rule, Federal Register, 1976; 41, 51206–51228.
2. U.S. Food and Drug Administration (FDA), Nonclinical Laboratory Studies. Good Laboratory Practice Regulations. Final Rule, Federal Register, 1978; 43, 59986–60020.
3. U.S. Food and Drug Administration (FDA), Good Laboratory Practice for Nonclinical Laboratory Studies: Amendment of Good Laboratory Practice Regulations. Amendment, Federal Register, 1980; 45, 24865.
4. U.S. Food and Drug Administration (FDA), Good Laboratory Practice Regulations. Proposed Rule, Federal Register, 1984; 49, 43530–43537.
5. U.S. Food and Drug Administration (FDA), Good Laboratory Practice Regulations. Final Rule, Federal Register, 1987; 52, 33768–33782.
6. U.S. Food and Drug Administration (FDA), Good Laboratory Practice Regulations. Minor Amendments. Final Rule, Federal Register, 1989; 54, 15923–15924.
7. U.S. Food and Drug Administration (FDA), Good Laboratory Practice Regulations. Removal of Examples of Methods of Animal Identification. Final Rule, Federal Register, 1991; 56, 32088.
8. U.S. Food and Drug Administration (FDA), Good Laboratory Practice Regulations. Technical Amendment, Federal Register, 1994; 59, 13200.
9. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, ENV/MC/CHEM (98)17, As revised in 1997. <http://www.oecd.org.ghs/ehsmono>.
10. Dent, N.J. Comparison Between the Principles of GLP 1992 as Defined in Monograph 45 and the Draft Revision of the Principles of GLP As Published by OECD on 6th January 1997. *Qual. Assurance J.* **1997**, 2, 7–12.
11. Dent, N.J. European Regulatory Compliance Issues: Good Research Practices. *J. Am. Coll. Toxicol.* **1994**, 13 (1), 79–85.
12. *OECD Guidelines for the Testing of Chemicals*; OECD: Paris, 1993.
13. Directives EEC/87/18, EEC/9, EEC/88/320. Official J. European Commun. **1992**.
14. U.S. Food and Drug Administration, Good Laboratory Practice for Nonclinical Laboratory Studies: Title 21, Part 58, Code of Federal Regulations, FDA, 1993.
15. U.S. Environmental Protection Agency (EPA), Pesticide Programs: Good Laboratory Practice Standards. Final Rule, Federal Register, 1993; 48, 53946–53969.
16. U.S. Environmental Protection Agency (EPA), Toxic Substance Control: Good Laboratory Practice Standards. Final Rule, Federal Register, 1993; 48, 53922–53944.
17. U.S. Environmental Protection Agency (EPA), Toxic Substance Control Act (TSCA): Good Laboratory Practice Standards. Final Rule, Federal Register, 1989; 54, 34034–34050.
18. U.S. Environmental Protection Agency (EPA), Federal Insecticide, Fungicide and Rodenticide Act (FIFRA): Good Laboratory Practice Standards. Final Rule, Federal Register, 1989; 54, 34052–34074.
19. Good Laboratory Practice Standards (FIFRA), EPA. Federal Register 1993; Title 40, Part 160, Code of Federal Regulations, Final Rule revised as of July 1.
20. Good Laboratory Practice Standards (TSCA), EPA. Federal Register 1993, Title 40, Part 792, Code of Federal Regulations, Final Rule revised as of July 1.
21. Shillam, K.W. GLP Legislation 16-29. *Good Laboratory and Clinical Practices: Techniques for the QA Professional*; Carson, P.A., Dent, N.J., Eds.; GLP Regulations Management Briefings—Post Conference Report. U.S. Food and Drug Administration: Rockville, MA, 1979.

22. Lepore, P.D. Overall Responsibilities for GLP 29-39. *Good Laboratory and Clinical Practices: Techniques for the QA Professional*; Carson, P.A., Dent, N.J., Eds.
23. Lepore, P.D. *Good Laboratory Practice Regulations: Questions and Answers*; Available through FDA FOI, FDA, (HFI-35); Rockville, MA, 1979.
24. Lepore, P.D. *Good Laboratory Practice Regulations: Questions and Answers*; Available through FDA FOI, FDA, (HFI-35); Rockville, MA, 1981.
25. Broad, R.D.; Dent, N.J. An Introduction to Good Laboratory Practice 3–16. *Good Laboratory and Clinical Practices: Techniques for the QA Professional*; Carson, P.A., Dent, N.J., Eds.
26. Commission Directive 1999/11/EC of 8th March 1999. Official J. European Commun. **1999**, L77/8 23.03.99.
27. Commission Directive 1999/12/EC of 8th March 1999. Official J. European Commun. **1999**, L77/8 23.03.99.
28. Good Laboratory Practice Regulations 1999. Statutory Instrument 1999 No. 3106. <http://www.hmso.go.uk/si/si1999/19993106.htm>.
29. Dominated by the Urge to Merge. *Scrip Magazine* **January 1999**, 46.
30. A Strategic Approach to R&D Outsourcing. *Scrip Magazine* **July/August 1999**, 29.
31. R&D Trends Gives Clues to Future. *Scrip Magazine* **July/August 1999**, 17.
32. How Much of a Threat is Virtual Pharma? *Scrip Magazine* **November 1999**, 41.
33. The Right Treatment for the Right Patient. *Scrip Magazine* **January 2000**, 11.
34. OECD Glossary. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. ENV/MC/CHEM(98)17. As revised in 1997. <http://www.oecd.org.ehs/ehsmono>.
35. Dent, N.J. Forget Compliance and Concentrate on Science. *Qual. Assurance J.* **1998**, 3, 103–108.
36. Personal Communication from OECD.
37. Personal Communication from OECD.
38. Personal Communication from OECD.
39. OECD GLP Consensus Document on the Application of the Principles of GLP to Computerised Systems. Environment Monograph (10), 116, OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. ENV/MC/CHEM(98)17. As revised in 1997. <http://www.oecd.org.ehs/ehsmono>.
40. *Computerized Data Systems for Non-Clinical Safety Assessment-Current Concepts and QA*; The Drug Information Association: Maple Glen, PA.
41. Double, M.E.; McKendry, M. *Computer Validation Compliance—A QA Perspective*; Interpharm Press: Buffalo Grove, IL.
42. Chamberlain, R. *Computer Systems Validation for Pharmaceutical and Medical Device Industries*; Interpharm Press: Buffalo Grove, IL.
43. U.S. Food and Drug Administration, Electronic Records; Electronic Signatures; Final Rule, Federal Register, 1997; 62, 13429–13466.
44. Lawrence, R.G. Archives 155-171. *Good Laboratory and Clinical Practices: Techniques for the QA Professional*; Carson, P.A., Dent, N.J., Eds.; 155–171.
45. Piton, A. Quality Assurance: The Present and Future Role of International Standardization. *Implementing International Good Practices—GAPs, GCPs, GLPs, GMPs*; Dent, N.J., Ed.; 251–265.
46. Kracht, W.R. Implementing the ISO 9000 Quality Standards in a GMP or GLP Environment. *Implementing International Good Practices—GAPs, GCPs, GLPs, GMPs*; Dent, N.J., Ed.; 267–277.

BIBLIOGRAPHY

- Carson, P.A., Dent, N.J., Eds.; *Good Laboratory and Clinical Practices—Techniques for the Quality Assurance Professional*; Butterworth.
- Compliance Programme Manual. <http://www.fda.gov.ora.cpgm.default.htm>.
- Dent, N.J. *Implementing International Good Practices—GAPs, GCPs, GLPs, GMPs*; Interpharm Press Inc.: Buffalo Grove, IL USA, 1993.
- EPA Homepage. <http://www.epa.gov>.
- EuropeanQuality. <http://www.euroqual.org>.
- GALPs. http://www.epa.gov/doc/irm_galp.
- GLP Regulations Questions and Answers. <http://www.fda.gov/cder/guidance/index.htm>.
- Good Research Practices—A Practical Guide to the Implementation of the GXP*s.
- Good Laboratory Practice on Line. <http://www.glp guru.com>.
- International Organization for Standardization. <http://www.iso.ch>.
- OECD Homepage. <http://www.oecd.org>.
- Title 21 Code of Federal Regulations Part 58. <http://www.access.gpo.gov/nara/cfr/index.html>.
- U.S. Federal Register—Environmental Issues. <http://www.epa.gov.fedrgstr>.

GOOD CLINICAL PRACTICES (GCP)—AN OVERVIEW

Richard A. Guarino

Oxford Pharmaceutical Resources, Inc., Totowa, New Jersey, U.S.A.

BACKGROUND

The Good Clinical Practice (GCP) regulations section in the Code of Federal Regulations (CFR) (21 CFR 312) (1) outlines the respective responsibilities of the clinical investigator, the drug sponsor, and the clinical study monitor involved in investigational new product development. These obligations, along with each participant's moral and ethical responsibilities for the safety of subjects who participate in clinical studies, comprise the essence of GCPs. GCPs have long been the norm for the investigator, as written in the 1572 Form (2); however, the first proposed regulations pertaining to investigator, sponsor, and monitor were first circulated in 1977 and 1978. In 1987, 10 years later, GCPs were published as final regulations in the CFR. Today, investigators, sponsors, and monitors are obligated by law to follow these GCPs.

To conduct clinical research that meets the requirements of the FDA for new product approval, it is essential to understand GCP regulations and their subsequent impact on the clinical development process of drugs, devices, and biologics.

The 1987, Investigational New Drug (IND), regulations specified within the current CFR identify (more clearly than in previously proposed GCP guidelines) the delegation of responsibilities in the conduct of clinical trials. Not only do investigators have a key responsibility in assessing patients' efficacy and safety response to new drugs, devices, or biologics, but the sponsor and monitor also have equal responsibility for the patients' safety and welfare. The key players who are obligated under GCP regulations are described below and will be referenced throughout this article:

Investigator. An investigator is the individual who conducts a clinical investigation (i.e., under whose immediate direction the drug is administered or dispensed to the subject). If an investigation involves many physicians at a particular institution, one physician is designated as the Principal Investigator (PI) and they are the responsible leader of the team of investigators. The subinvestigator is

any other individual member of that team as identified by the PI. These individuals are usually licensed physicians or individuals working under a licensed physician (3).

Sponsor/Investigator. A sponsor/investigator is an individual who both initiates and conducts an investigation (i.e., under whose immediate direction the investigational drug is administered or dispensed). This category refers mostly to physician investigators who are conducting clinical research under an investigator IND (4).

Sponsor. A sponsor is an individual or organization that takes responsibility for and initiates a clinical investigation. This may be an individual, pharmaceutical company, governmental agency, academic institution, or a private or other organization (5).

Monitor. A monitor is the person selected by the sponsor who is qualified by training experience to facilitate and oversee the progress of the investigation (6).

INVESTIGATOR OBLIGATIONS

In 21 CFR 312.53 (7), the regulations deal with the descriptive information provided on form FDA 1572, the Statement of Investigator form. Also included in 21 CFR 312.53 are the selection requirements for clinical investigators. Previously, to conduct studies designated as Phases 1 and 2, investigators were required to complete a Statement of Investigator form FDA 1572; investigators conducting studies designated as phase 3 or phase 4 completed a different Statement of Investigator form, which was known as form FDA 1573. As a result of the IND rewrite regulations, form FDA 1573 is no longer used for any clinical studies. At present, for Phases 1–4, only the Statement of Investigator form FDA 1572 is required. This document states the obligations of investigators conducting clinical research. In addition to the general

information on the 1572, new information includes the following: the name and address of any clinical laboratory facility, the address of the Institutional Review Board (IRB) (8) responsible for the review and approval of the protocol, the patient consent form, and the individual investigators participating in the study. This document also states that the sponsor is charged with the responsibility of selecting qualified investigators, who are defined as those who are capable of conducting the study by virtue of their training and experience. By using the phrase “training and experience,” the FDA means that clinical investigators conducting a study of a particular disease should have enough experience in that clinical specialty to observe correctly the signs, symptoms, and progress of the disease being treated with a new investigational drug. For example, if a new drug is designed for an Obstetrics/Gynecology practice, a pediatrician would not be expected to have the expertise to assess this drug, nor would a cardiologist have expertise in evaluating a gastrointestinal drug.

Investigators are defined as those who have signed and completed form FDA 1572 or sub- or coinvestigators listed on that form, who are considered to have the academic and experiential qualifications for participating in the clinical program.

The “fine print” on the reverse side of form FDA 1572 is a written agreement whereby the investigators assure the sponsor that they will conduct the study in accordance with the appropriate study plan (i.e., the protocol) (9) and will observe the GCP tenets. Implicit in this agreement is the fact that the Investigator will have obtained signed Informed Consent (IC) (10) forms from patients or subjects participating in the clinical research under their jurisdiction. Form 1572 also charges the investigator with the reporting of adverse experiences that occur during the investigation and provides assurance that the investigator has read and understood the investigator’s brochure. In addition, he or she assures that all individuals participating in the supervision of any clinical study, under the direction of the investigator, are aware of their responsibilities. Once form 1572 has been signed by the investigator, he or she further assures compliance with the requirements of providing study materials, protocols, and other pertinent information to an authorized IRB for review. This information, along with a curriculum vitae, should be provided along with the assurance that the investigational plan set forth in the study protocol will be complied with.

To summarize, the primary responsibilities of investigators in clinical trials are the ethical and moral obligations to all the participating patients and subjects in the study. Investigators must provide a measure of safety for each participant in the study so that the patient is

protected ethically and morally from any endangerment that might occur during a trial using an investigational drug. After the investigator’s responsibilities are outlined and he or she has signed form FDA 1572, any additional information from the sponsor that might be necessary should be requested and any concerns regarding procedures should be raised. An investigator is responsible for: 1) ensuring that an investigation is conducted according to the signed investigator statement, the investigational plan, and applicable regulations; and 2) for protecting the rights, safety, and welfare of subjects participating in a clinical investigation on any unapproved product. Also, the investigators must maintain complete control and accountability of the experimental products under investigation. An investigator shall obtain the informed consent of each human subject to whom the drug is administered and shall administer the drug only to subjects under the investigator’s supervision or under the supervision of a subinvestigator responsible to the investigator. The investigator shall not supply the investigational drug to any person not participating in the clinical program.

The investigators are required to maintain adequate records of the disposition of the experimental medications, including dates, quantity, and use by subjects (11). If the investigation is terminated, suspended, discontinued, or completed, the investigator shall account for and return the unused supplies to the sponsor, or otherwise provide written documentation for disposition of the unused supplies of the drug. An investigator is required to maintain accurate case histories designed to record all observations and other pertinent data on each individual treated with the investigational drug. (Usually, this is accomplished by completing case report forms and maintaining medical records).

All investigators shall retain records of all subjects enlisted in investigational trials for 2 years after a new drug application (NDA) is approved for the indication for which the drug is being investigated. If no application is to be filed or if the application is not approved for such indication, records must be maintained 2 years after the investigation is discontinued or the IND is closed and the FDA has notified the sponsor of the status of the application (12).

The investigator shall furnish all reports to the sponsor of the drug. The sponsor is responsible for collecting and evaluating the results obtained. The sponsor also is required to submit annual reports to the FDA on the progress of the clinical investigations.

Investigators shall promptly report to the sponsor any adverse effect that may reasonably be regarded as caused by, or probably caused by, the investigational drug. If the

adverse effect is serious (13) the investigator shall report the adverse effect immediately. (See chapter on ADR reporting.)

An investigator shall provide the sponsor with an adequate report shortly after completion of the investigator's participation in the study.

OTHER INVESTIGATOR RESPONSIBILITIES

The investigator must assure that an IRB complies with the regulations established in the CFR and that the IRB is responsible for the initial and continuing review and approval of the proposed clinical study. The investigator must also assure that he or she will promptly report all changes in the research activity and all unanticipated problems involving risk to human subjects Adverse Reactions (ADRs) or others to the IRB. In addition, the investigator will not make any changes in the research protocol without IRB approval, except where necessary to eliminate apparent immediate hazards to human subjects.

An investigator will on request from any properly authorized officer or employee of the FDA, at reasonable times, permit such officer or employee to have access to, copy, and verify any records or reports made by the investigator. The investigator is not required to divulge subject names, unless the records of particular individuals require a more detailed study of the cases or unless there is reason to believe that the records do not represent actual case studies or do not represent actual results obtained.

SPONSOR OBLIGATIONS

The sponsor's primary responsibility is clearly delineated in 21 CFR 312.50 (14) and ensures that clinical studies are conducted in compliance with FDA regulations. The sponsor is responsible for selecting qualified investigators and for providing them with the information they need to conduct an investigation in accordance with the published regulations. Usually, the sponsor accomplishes this task by supplying the potential investigator with an investigator's brochure and a protocol of the clinical investigation on the agent to be investigated. An investigator's brochure (15) contains all information from nonclinical studies and reports and any previous human efficacy and safety study reports that reflect previous experiences of patients of the investigational agent.

Of primary interest in the obligations is the option of a sponsor to transfer total or partial responsibility for the

conduct of a clinical study to a Contract Research Organization (CRO) (16). During the last decade, CROs have played a significant role in new drug development. However, CROs who contract with sponsor companies are obligated under the same GCP regulations as defined in this chapter. A CRO may be the sponsor or the monitor with equal obligations as defined in 21 CFR 312. The current regulations noted in 21 CFR 312.52 (17) are specific and require that any transfer, whether in total or in part, be described in writing and agreed to by both parties. The FDA states that any obligations not specifically described by the sponsor in the written transfer of responsibilities will be considered as not transferred to the CRO; the liability for these undefined responsibilities, therefore, remains with the sponsor. The FDA further requires the CRO (once any transfer of responsibilities has been made by the sponsor) to comply with all applicable regulations and notes that the CRO is subject to the same regulatory actions as a sponsor if a CRO does not satisfy FDA regulations in the fulfillment of its contracted duties. As a result of these regulations, it is possible for a CRO to act on behalf of a sponsor once this legal transfer of obligations has been completed. Although the CRO must assure complete compliance with the responsibilities assigned, it remains the sponsor's responsibility to ensure the quality and integrity of data generated under the supervision of a CRO. In this situation, the sponsor would be expected to act as a quality assurance (18) auditor of the data, even though assignment for the conduct of a study has been delegated to the CRO. It is important to note the following: that any such transfer shall be described in writing; if not all obligations are transferred, the description of the specific obligations being assumed by the CRO must be clearly stated. Any obligation not covered by the written description shall be deemed not to have been transferred. The regulations also charge the sponsor with responsibility for the inventory and control of the drug. Only investigators participating in a clinical trial may receive and have access to investigational drug and materials.

SPONSOR AND MONITOR OBLIGATIONS

One of the most important responsibilities of the sponsor is to monitor the progress of every clinical investigation conducted under its direction (21 CFR 312.56) (19). A monitor's obligations, under the auspices of the sponsor, are to ensure that the deficiencies created during the conduct of clinical investigations are corrected or justified by the investigator and that the investigator adheres to the

investigational plan. The appointed monitors for any clinical investigation conducted under a sponsor's IND have an obligation to assure that an investigator is complying with the signed Form FDA 1572 and the general investigational plan and that the clinical protocol is being followed. If an investigator does not correct his or her errors and mistakes and no improvement is noted in the progress of the study, the monitor shall promptly secure compliance or discontinue shipment of the investigational new drug to the investigator and end the investigator's participation in the clinical program. In addition the monitors, while monitoring the progress of a clinical investigation, must evaluate the evidence relating to safety and effectiveness. At the same time, sponsors shall make such reports to the FDA regarding information relevant to the safety of the drug, as they are required to do under section 312.32 (20) of the FDA regulations.

When a monitor reports an adverse effect to a sponsor during an investigational study, it is the sponsor's obligation to determine whether there is an unreasonable and significant risk to the subject or patient (21). At that time, the sponsor must determine if the investigational study is to be discontinued. Important among the procedures of reporting adverse effects is the sponsor's obligation to the FDA, the IRB, and to all investigators who, at any time, participate in clinical studies and who are prescribing the experimental drug. Subsequent to this, the sponsor should furnish the FDA with a full report of the sponsor's actions and shall determine whether or not to discontinue the investigation. If the decision to discontinue is made, based on the seriousness of the ADRs reported, the studies should be terminated as soon as possible and no later than 7 days after making the decision.

It is important to understand that the obligations of monitors include the responsibility for assuring that all records and data recorded on case report forms reflect valid data gathered by the investigator and that they coincide with corresponding medical and hospital records of the candidate participating in the investigational study. Detailed auditing and documentation assure the sponsor that the monitor is overseeing the clinical data collected by the investigator and that GCPs are being followed (22). One misconception of many monitors who audit clinical investigations is that their only task is to assure correct entry of data. In fact, it is of extreme importance among the monitor's obligations to note any adverse effects or any deviations in laboratory values that could signify a safety problem to investigational study subjects. This is especially true in large multiclinic studies, in which many centers are conducting investigational studies following the same protocol and many monitors are auditing data. If any abnormal reactions or laboratory deviations are noted

from center to center, the monitors should compare observations and assess an accumulative percentage of occurrence of these deviations. At times, a sporadic, apparently minor deviation can turn out to be a significant deviation when calculated across all centers. If monitors are astute, they can often prevent recurrence of adverse events that might jeopardize the safety of the subjects participating in investigational drug studies.

Another responsibility of the monitor is to assure maintenance of accurate records showing the receipt, shipment, or other disposition of the investigational drug (23). These records are required to include, as appropriate, the name of the investigator to whom the drug is shipped, the date, the quantity, and the batch number of each shipment. The monitor/sponsor shall also assure the return of all unused supplies of the investigational drug from each investigator whose participation in the investigation is discontinued or terminated. The sponsor may authorize an alternative disposition of unused supplies of the investigational drug, provided this alternative disposition does not expose humans to risks. Although the overall responsibilities are assigned to sponsors, it is the monitors' underlying responsibility for drug accountability.

In turn, the investigators, during experimental research, are also responsible for record retention similar to that of the sponsor. They are required to maintain adequate records of the disposition of the drug, including dates, quantity, and use by the subjects or patients. The investigator is also obligated, if he or she is terminated, suspended, or discontinued or if he or she has completed a study, to return all unused supplies of the drug to the sponsor or otherwise provide documentation of how the unused supplies of the drug were disposed. (It is recommended always to return the unused study medication to the sponsor). An investigator is required to prepare and maintain adequate and accurate case histories (designed to record all observations and other data pertinent to the investigation) on each individual treated with the investigational drug. The monitor should assure that all the previous procedures are adhered to and reported in a timely fashion.

An often-neglected investigator responsibility is the requirement to submit periodic reports to the sponsor. An investigator should be prepared to provide the sponsor with progress reports. These should include an update of the ongoing investigational trial. Annual reports to the FDA on the progress of the clinical investigations are required to be submitted by the sponsor. These reports contain information based on the investigators' progress reports (24). Safety reports are another issue. An investigator should promptly report to the sponsor any adverse events that may reasonably be regarded as caused by or likely caused by the investigational drug. Alarming adverse events (i.e., severe

adverse reactions that jeopardize a patient's safety in any way) must be reported immediately by the investigator to the sponsor (25). Lastly, when an investigator has completed or terminated an investigational study, a final report shall be provided to the sponsor. This comprehensive report should be submitted to the sponsor shortly after completion of an investigator's participation in the investigation. The report summarizes the final observations of the study and any adverse events that occurred during the course of the clinical investigation. Monitors should also be responsible for encouraging investigators to complete and submit all the reports listed above. Constant follow-up may be necessary by the monitor if these investigator responsibilities are to be fulfilled. In most cases, the clinical monitor usually will provide the investigator with these reports.

Legal repercussions can occur from any neglect of the obligations by investigators, sponsors, or monitors. The CFR stipulates in 21 CFR 312.58 (26) that the FDA can inspect the sponsor's records or reports on request from any properly authorized officer or employee of the FDA. These inspections normally occur at reasonable times and permit the FDA to have access to copy and verify any records and reports relating to a clinical investigation conducted under an IND. On written request by the FDA, the sponsor may be asked to submit the records, reports, or copies of them to the FDA. Under these regulations, the sponsor is also obligated to discontinue shipments of the drug to any investigator who has failed to maintain or make available records or reports of the investigation. Subsequently, an investigator may, on request from any properly authorized officer or employee of the FDA, at reasonable times, permit such an officer or employee to have access to or copy and verify any records or reports made by the investigator. The investigator is not required to divulge subject or patient names unless the records of particular individuals require a more detailed study of the cases.

GCP NONCOMPLIANCE

What are the consequences if an investigator has repeatedly or deliberately either failed to comply with these GCP requirements or has submitted false information in any report to the sponsor (27)? Initially, the Center for Drug Evaluation and Research (CDER) or the Center for Biologics Evaluation and Research (CBER) will furnish the investigator with written notice of the matter complained of and offer the investigator an opportunity to explain the matter in writing or at the option of the investigator, grant an informal conference. If the

explanation offered by the investigator is not accepted by the CDER or the CBER, the investigator will then be given an opportunity for a regulatory hearing. At this hearing, the issue of whether the investigator is entitled to receive investigational drugs will be addressed. After evaluating all available information, including any explanation presented by the investigator, the FDA commissioner determines whether the investigator has repeatedly or deliberately failed to comply with the GCP requirements or has deliberately or repeatedly submitted false information to the sponsor in any required report. The commissioner will then notify the investigator and the sponsor of any investigation in which the investigator has been named as a participant that the investigator is not entitled to receive investigational drugs. The investigation can not be terminated without reasonable cause as set forth by the commissioner and committee. Sponsor can also suspend shipment of drugs to the investigator for noncompliance to the protocol. If there is reasonable cause for this action, the investigator becomes subject to further investigation for each IND and each approved application submitted to the FDA containing data reported by this investigator. Therefore, every investigational study conducted by this investigator will be examined to determine whether the investigator has submitted unreliable data. Other investigations that are conducted under the same protocol will be temporarily put on hold. Conversely, the commissioner may determine, after eliminating the unreliable data by the investigator, that the remaining data justify continuing other of the same investigations at other sites. However, if a danger to the public health exists, the commissioner will terminate the IND immediately; the sponsor will be notified and will have an opportunity for a regulatory hearing before the FDA on the question of whether the IND should be reinstated. If the commissioner determines that the data submitted are unreliable and that the data submitted by the investigator cannot be justified, the commissioner will proceed to withdraw approval of the drug product in accordance with the provisions of the Food and Drug Cosmetic Act (FD&C). As a result, an investigator who has been deemed to be ineligible to receive investigational drugs will be blacklisted and unable to participate in any experimental studies. The investigator may be reinstated when the commissioner determines that the investigator has presented adequate assurances that the investigator will use investigational drugs in compliance with FDA regulations.

In conclusion, before an investigator accepts the responsibilities to conduct a clinical investigation with an IND drug, he or she must be aware of the legal obligations he or she has agreed to when form FDA 1572 is signed. Investigators must comply with the protocol and the rules, regulations, and guidelines of GCPs.

Investigators must realize that they are subject to a federal offense and can jeopardize their reputation and, ultimately, their ability to conduct further clinical research. Investigators must know that the precise collection of data is mandatory in the conduct of clinical research. Research must be designed to assess the efficacy of the product and, above all, to assure that the safety of the patient remains the primary concern.

Sponsors' and monitors' responsibilities in complying with GCPs are also subject to serious repercussions under 21CFR 312.58. FDA inspectors are allowed to examine sponsors' files and the interventions of monitors' site visits to assure that GCP compliance was executed. Case report forms and clinical results are subjected to the same scrutiny that are applied to the investigators' responsibilities. If during an FDA inspection discrepancies are found in any form among the investigator, sponsor, and, when appropriate, the CRO (i.e., its documents), all three parties will be held responsible, and the IND will be placed on hold until the findings are resolved.

Investigators', sponsors', and monitors' obligations must be fulfilled by complying with GCP rules and regulations. Sponsors' and monitors' consistent and persistent managing roles are vital in assuring that each person involved in conducting clinical studies meet his or her legal obligations. The success of any clinical program will depend on the cooperation, understanding, and compliance of this triad working together. With this agreement of responsibilities and a well-organized clinical plan, the results can only conclude valid data in support of a new drug application.

REFERENCES

1. *Code of Federal Regulations*, Title 21, Part 312, (annually).
2. Guarino, R.A. The Investigational New Drug Application and the Investigator's Brochure. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 74–75.
3. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 326.
4. Guarino, R.A. The Investigational New Drug Application and the Investigator's Brochure. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 56.
5. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 329.
6. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 330.
7. *Code of Federal Regulations*, Title 21, Part 312.53, (annually).
8. Guarino, R.A. Institutional Review Board/Independent Ethics Committee and Informed Consent: Protecting Subjects Throughout the Clinical Research Process. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 271.
9. Guarino, R.A. Clinical Research Protocols. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 219.
10. Guarino, R.A. Institutional Review Board/Independent Ethics Committee and Informed Consent: Protecting Subjects Throughout the Clinical Research Process. *New Drug Approval Process*, 3rd Ed. Marcel Dekker, Inc.: New York, 2000; 100, 278.
11. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 328.
12. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 328.
13. Guarino, R.A. Adverse Reactions and Interactions of Drugs. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 247.
14. *Code of Federal Regulations*, Title 21, Part 312.50, (annually).
15. Guarino, R.A. The Investigational New Drug Application and the Investigator's Brochure. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 86.
16. Guarino, R.A. Working with a CRO. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 439.
17. *Code of Federal Regulations* Title 21, Part 312.52, (annually).
18. Guarino, R.A. Quality Assurance. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 349.
19. *Code of Federal Regulations*, Title 21, Part 312.56, (annually).
20. *Code of Federal Regulations*, Title 21, Part 312.32, (annually).
21. Guarino, R.A. Adverse Reactions and Interactions of Drugs. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 262–263.
22. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 330–334.
23. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 331.
24. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 332.
25. Guarino, R.A. Adverse Reactions and Interactions of Drugs. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 260.

26. *Code of Federal Regulations*, Title 21, Part 312.58, (annually).
27. Guarino, R.A. Obligations of the Investigator, Sponsor, and Monitor. *New Drug Approval Process*, 3rd Ed.; Marcel Dekker, Inc.: New York, 2000; 100, 332–333.
28. Guarino, R.A. Clinical Research: The Obligations of the Investigator, Sponsor, and Monitor. *Drug Info. J.* **1978**, *12* (3), 157.
29. Guarino, R.A.; Cox, B.S. *Practice and Application of Good Clinical Practices*; Cooperation with St. John's University: New York, Oct 1982; The American Federation for Clinical Research Eastern Section.
30. Guarino, R.A. *Practice and Application of Good Clinical Practices. Concepts and Strategies in New Drug Development*; Clinical Pharmacology and Therapeutic Series Praeger Publishers: New York, 1983; 4.
31. Guarino, R.A. *Role of the MD in Protocol Design*; Applied Clinical Trials Conference and Exhibition: Reston, VA, 1993.
32. Guarino, R.A. *Clinical Study Conduct/Procedures*; Drug Information Association Meeting: Chicago, IL, July 1993.
33. Stuart, J.P. *Clinical Trials. A Practical Approach*; John Wiley & Sons: New York, 1983.
34. MacKintosh, D.R.; Zepp, V.J. GCP Responsibilities of Principal Investigators: Beyond the 1572. *Appl. Clin. Trials* **Nov. 1996**, 32–40.
35. MacKintosh, D.R.; Zepp, V.J. Source Documentation: A Key to GCP Compliance in Clinical Trials. *Appl. Clin. Trials* **Mar. 1996**, 42–46.
36. Souza, K.; Small, R.D. The use of Total Quality Management Principles in Placing and Monitoring Clinical Trials. *Drug Info. J.* **1995**, *29*, 695–703.

GLASS AS A PACKAGING MATERIAL FOR PHARMACEUTICALS

Claudia C. Okeke

United States Pharmacopeia, Rockville, Maryland, U.S.A.

INTRODUCTION

Glass has been used for over 6000 years, dating to ancient times. Through the years and with more knowledge of its technology, glass has become the most widely used drug packaging material. The origin of the first synthetic glass is unknown; however, Egyptians were known to mold figurines from sand and silicon dioxide (1). The American Heritage dictionary defines glass as a large class of materials with highly variable mechanical and optical properties that solidify from the molten state without crystallization, that are typically based on silicon dioxide, boric oxide, aluminum oxide, or phosphorous pentoxide, that are generally transparent or translucent, and that are regarded physically as supercooled liquids rather than fine liquids (2).

ASTM defines glass as an inorganic product of fusion that has cooled to a rigid condition without crystallizing (3). ASTM further states that glass is typically hard and brittle and has a conchoidal fracture. Glass may be colorless or colored. It is transparent but may be made opaque or translucent (3). Glass is noncrystalline and is amorphous in structure and may be formed from both organic and inorganic materials.

Glass is manufactured at very high temperatures, and at very hot temperatures, it has the properties of a viscous liquid. With the properties of a viscous liquid, hot glass can be formed into many commonly used forms with precision and accuracy. As a result of its noncrystalline nature, it affords a unique transparent property that is maintained because it does not crystallize upon cooling.

There are a variety of uses of glass, one use being a pharmaceutical packaging material. Glass is favored over other types of packaging material because its transparent property enables it to provide good visualization of contained material. Another good quality of glass is its excellent resistance to attack by most liquids, and, therefore, it resists interaction with contained products. It is also totally impermeable to gases, and it can be sterilized with any appropriate process. Glass, when properly colored, also provides protection of a product from light.

GLASS COMPOSITION

There are two glass compositions generally used—soda-lime and borosilicate. The soda-lime compositions are used for ordinary tableware, food, and beverage products, and window glass, among many others. Borosilicate glass is not widely used, but it is more durable and heat-resistant than soda-lime glass. This glass is used for laboratory and scientific glassware, and heat-resistant cookware, among many others. Borosilicate glass affords properties that make it a preferred composition for certain pharmaceutical containers. Glass composition will be discussed in detail below (4).

GLASS MANUFACTURE AND COMPOSITION

Raw Materials

Silicon dioxide (SiO_2) also known as silica, is the principal constituent of glass. Glass is made by melting its ingredients at very high temperatures, such as 1550°C . Because the melting point of silica alone is so high, and, therefore, commercially difficult to melt and form into containers, other oxides are added to silica to lower the melting point and make it easier to fabricate and use commercially (5). These oxides include sodium oxide (Na_2O); aluminum oxide (Al_2O_3); potassium oxide (K_2O); boron oxide (B_2O_3); and calcium oxide (CaO). The addition of one or more of these oxides reduces the melt viscosity for fabrication purposes. Other materials may be added to change a color, facilitate melting, or for other reasons. Sodium and potassium oxides are obtained as the product of chemical processing of naturally occurring materials, and an abundant source is the brines of Searles Lake in California. Sodium is primarily added to silica to lower the melting temperature and aid in the removal of gas bubbles in a reasonable time. Potassium oxide is used in smaller amounts. All of these additives change the properties of silica in such a way that the resulting glass is less resistant to attack by aqueous solutions. Other additives such as calcium oxide, magnesium oxide, and aluminum oxide are also used in various amounts to affect

the properties of the glass. Silica has a very low coefficient of thermal expansion, and the addition of an alkali such as sodium oxide increases the thermal expansion.

When fluxing action is needed during melting and glass with high thermal expansion is undesirable, boric oxide is added forming borosilicate glasses, which have lower thermal expansion and improved resistance to attack by aqueous solutions because of lower sodium oxide contents.

Conditions necessary for glass formation may be deduced from either geometric or bond strength considerations. Silica sand deposit is available in all parts of the world. Silica sand is mined by hydraulic dredging, and any impurities are eliminated through further processing. One impurity, iron oxide, may be present, and, if present, may affect the final color of the glass (4). There are other oxides that are potential glass formers and may be used in glass formation: GeO_2 , P_2O_5 , As_2O_5 , P_2O_3 , As_2O_3 , Sb_2O_3 , V_2O_5 , Sb_2O_5 , Nb_2O_5 , and Ta_2O_5 (6). Pharmaceutical containers may require amber glass to provide protection for light-sensitive pharmaceutical products. Amber glass is available in both soda-lime and borosilicate. The amber glasses have negligible transmission in ultraviolet (UV) and near UV regions and, hence, provide the required protection. In these cases, glass is interacted with iron oxide to provide amber color.

For pharmaceutical products that undergo sterilization by means of ionizing radiation, cerium oxide of about 1% or less is added to the glass formulation. This is because ionizing radiation type sterilization [dosages as high as 20–30 kGy (2 to 3 Mrad)] affects ordinary glass by darkening the glass and, thus, inhibits the final inspection of the glass products.

The mechanism by which ionization radiation acts on glass is by dislodging the electrons in the glass structure (usually intermediate density and low cost glasses) to form color centers and creating changes to the multivalent ions in such a way that they absorb visible. The result is a formation of neutral atoms such as sodium that can agglomerate into colloid-like configurations. Cerium oxide, which is multivalent when added to the formulation, counteracts the darkening effect of radiation by capturing these free electrons and minimizing the deleterious effects without adding color to the glass (4).

All raw materials used to formulate a glass composition should be characterized using certain specifications of particle size, distribution, overall purity, and specific impurity content. It is necessary to match the particle size of the raw materials as much as possible to allow trouble-free melting.

Also classified as a raw material is cullet. Cullet is a scrap glass of desired composition, which results from scrap generated during forming operations and kept

strictly segregated by composition. Cullet is then recycled through the melting process. The use of cullet conserves raw materials and aids in the melting process. A typical diagram illustrating the process in glass manufacturing is shown in Fig. 1 (1).

FORMULATION—BATCH MIXING, CHARGING, AND MELTING

In a typical factory, a batch formulation is developed that provides the desired glass composition when melted. The raw materials are stored in large capacity silos. Depending on the type of plant facility, the raw material may be proportioned and mixed by a computer programmed with the batch formulation. The computer functions to control the automatic material transport system, weighing equipment, withdrawing the correct amount the raw materials from the perspective silos, and discharging them into a large mixing vessel. After mixing for the desired time, the batch is discharged into cans of mixed batch. These cans are transported to the charging end of the melting furnace. Complete records are maintained for the weighing, mixing, and transportation of batch materials to the furnace.

Usually, when the intimately mixed batch is charged into the hot furnace, a series of melting, dissolution, volatilization, and redox reactions occur between the materials in a certain order and at the appropriate temperature (7). At the furnace, the batch contents are discharged through a screw feeder into the furnace, with temperatures capable of exceeding 1700°C. The batch floats on the surface of the glass already in the furnace and gradually melts into the desired glass composition. The temperature of the furnace is necessary to facilitate melting and producing glass of the desired homogeneity in a commercially acceptable time. The glass is allowed to stay in the furnace until a satisfactory homogeneous product is formed, after which the temperature is lowered at the discharge end to achieve a glass viscosity that allows the desired forming operations to take place.

In most cases, the glass made cannot be characterized by a batch or lot designation but only by the date and time of withdrawal from the furnace. This is because the glass-melting process is a continuous process, even though the glass materials are charged in discrete amounts. The residence time in the furnace permits extensive mixing to occur and, therefore, erases all identity of the batch from individual cans (4).

Devitrification is the uncontrolled formation of crystals in glass during melting, forming, or secondary processing. The optical properties, mechanical strength, and some-

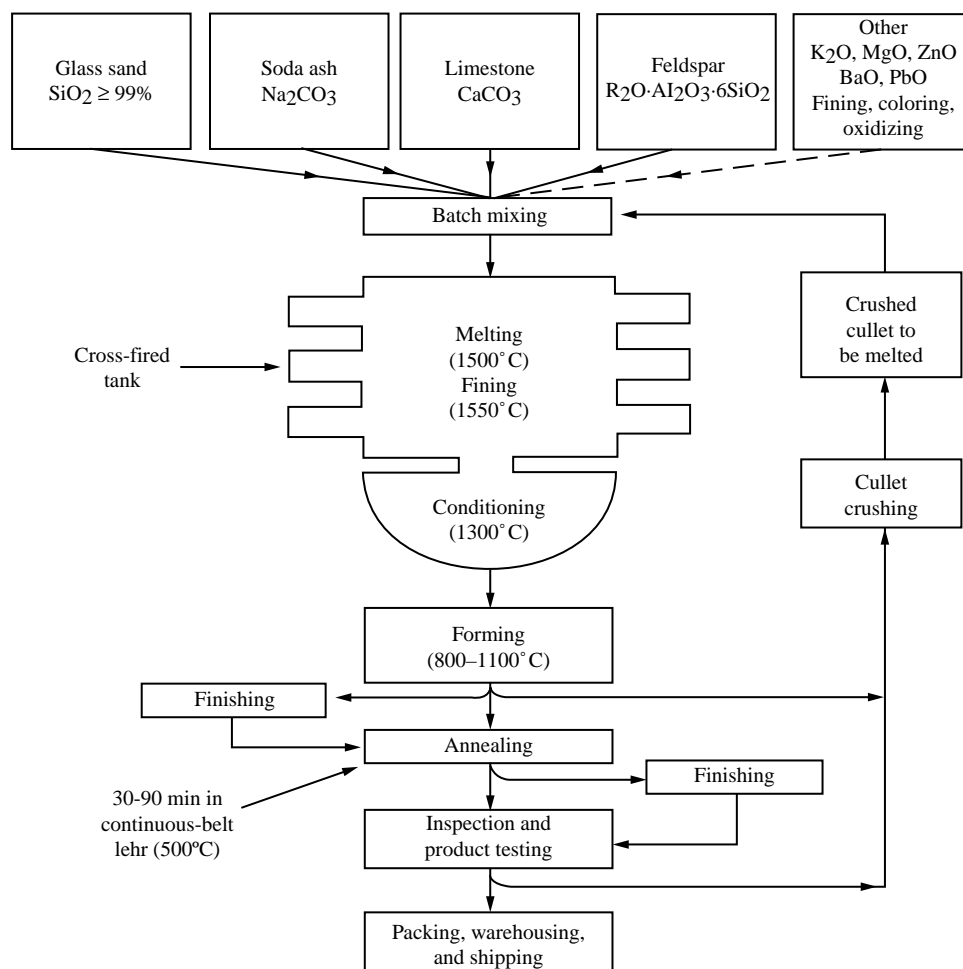


Fig. 1 Glass manufacture. Temperatures used may vary depending on the type of glass. (From Ref. 1).

times the chemical durability of the glass can be adversely affected by devitrification. These unwanted crystals grow homogeneously within the glass or heterogeneously at the air-glass or refracting-glass interface. Devitrification occurs mainly in glasses where the optimum temperatures for maximum nucleation rate and for maximum growth rate nearly coincide. If these glasses are held too long in this critical temperature range or are cooled too slowly through it, the glass starts to crystallize. For soda-lime glasses, the crystal phase, devitrite, forms between 850–900°C. However, since the glass is already quite viscous, the critical temperature range is short, and devitrification is not much of a problem (1–3).

QUALITY OF A GLASS PRODUCT

Following a successful product fabrication, there are some qualities to observe, one of which is total homogeneity of

product. Most glass manufacturers try to ensure that composition of the glass is homogeneous throughout. However, gas bubbles that did not escape the body of glass while in the furnace may cause a nonhomogeneous composition to exist.

Glass bubbles in manufacturing result from chemical reactions that take place during melting. Glass bubbles evolve due to the following: 1) because of gas formation from decomposition of the carbonates or sulfates or both; 2) from air trapped between the grains of the fine-grained batch materials; 3) from water evolved from the hydrated batch materials; and 4) from the change in oxidation state of some of the batch materials, such as red lead. However, with enough time in the furnace, the air bubbles escape the melt by rising slowly to the top. In some cases, the gas bubbles may be withdrawn with glass from the furnace. These bubbles are called seeds or blisters. They are not desirable and may be a cause for rejection of the final glass product, depending on the size and extent. When the

surface of a glass final product is broken by the bubble, it is considered a functional defect. However, if the bubble is surrounded by glass, it is considered a cosmetic defect, and while generally it may not be hazardous, it may still be rejected.

Cord is another attribute in glass formation that is not desirable and should be prevented. Cords are formed when glass is not stirred properly at appropriate places in the furnace. It is referred to as the act of improper homogenization. Typically, in this case, glass melting occurs at a very high temperature, and some constituents vaporize from the surface glass and form regions of viscous glass. When small amounts of this glass mix into the body of the melt, they appear as very narrow and long inhomogeneous regions, referred to as cord (4).

Batch segregation, melt segregation, volatilization, and temperature fluctuations, as well as refractory corrosion from tank-lining material, cause stria or cord formation as well. These can also be prevented by melt homogenization and vigorous fining action. The homogeneity leading to cord formation can be removed by diffusion and flow and by vigorous fining, along with convection current mixing process before the glass is cooled. Homogeneity can also be improved by mechanical and static mixers that continuously shear the glass.

The bubble and cord defects that occur during the melting process are the two most commonly encountered. Melting defects should be held to a minimum as much as possible.

Fining is the physical and chemical process of removing glass bubbles (seeds and blisters) from the molten glass melt. Typically, fining agents that react at higher temperatures than are needed for melting are used. Some examples of fining agents are sulfates and sodium-potassium nitrates in combination with arsenic or antimony trioxides. Arsenic trioxide is used for melting glasses at higher temperatures, e.g., 1450–1500°C, whereas antimony is used for lower melting glasses at 1300–1400°C. As glasses cool, oxygen bubbles are removed by the reaction with arsenic or antimony trioxide to form pentoxide (1).

Gas may be evolved from a typical amber soda-lime glass batch, and fining agents are usually employed to resolve the problem (8). A typical batch mixture for amber soda-lime container glass is shown in Table 1.

QUALITY CONTROL IN BATCHING AND MELTING

Producing glass containers for parenteral products requires strict control of all process aspects, starting with raw

materials and ending with the testing of glass as a material and as containers or parts of medical devices. The objectives of a quality-control program should be to maintain glass produced within physical and chemical specifications and to prevent off-specification material from reaching the pharmaceutical manufacturer.

Normal practice is to make sure that the raw material meets desired specifications and maintains its integrity during loading and shipping. There are physical as well as chemical requirements associated with each raw material, and these requirements are met by controlling for desired behavior during batching and melting and by the desired properties of the final glass product. The chemical impurity content is usually dictated by the desired attributes of the glass container. For example, raw materials of low iron content are used in order to provide a virtually colorless final product. Raw materials of low sulfur content (in the case of borosilicates) are used to minimize gas bubble formation during melting. Substances that would be harmful if extracted from the glass during terminal sterilization and subsequent storage (e.g., lead or arsenic) should be absent (4).

Most raw material suppliers provide certificates of analysis for the material they provide, and the information is checked by further testing on receipt. Records of performance of raw materials are maintained. The vendor's production facilities are usually inspected to ensure that mixing of raw materials of different grades cannot occur during loading and shipping. Ideally, this is a desirable situation and should be the goal of quality-oriented raw material suppliers and the glass manufacturers.

Scrap glass, or cullet, is also considered a raw material. Proper handling of cullet to prevent mixing of compositions and contamination by foreign materials requires constant attention and is just as important as preventing contamination of a raw material received from outside the plant.

Special considerations should be applied to ensure that all glass batch formulations are prepared by qualified technologists. The formulation for each composition is given in a batch sheet, which specifies the glass type, melting tank, date issued, amounts of raw materials, total batch weight, and weight of glass expected from the batch. In some companies, batch sheets are under the strict control of the glass-technology department, and each sheet denoting a change requires approval by several levels of department management. Changes are made primarily to manage the cullet supply, which can vary. Batch sheets usually provide a continuous time record of a given composition melted in a given furnace, and the batch sheets are supplied to the department responsible for actual

Table 1 Typical batch mixture for amber soda-lime container glass

Batch materials	Oxides supplied (kg)						LOI ^a (kg)
	Weight (kg)	SiO ₂	Al ₂ O ₃	CaO	Na ₂ O	FeO	
Sand, SiO ₂	300	299.3	0.2	—	—	0.03	0.5
Soda ash, Na ₂ CO ₃	100	—	—	—	58.3	—	41.7
Aragonite, CaCO ₃ ^b	90	—	—	49.0	—	0.02	40.7
Feldspar, SiO ₂ ·Al ₂ O ₃ mineral ^c	40	26.4	7.6	0.4	1.3	0.03	0.1
Salt cake, NaCl	4	—	—	—	2.1	—	1.9
Powdered coal ^d	9	—	—	—	—	—	9
Iron pyrites, FeS ₂	1.4	—	—	—	—	0.84	0.6
Cullet	460	333.7	9.2	48.8	67.2	1.03	0.1
Total	1004.4	659.4	17.0	98.2	128.9	1.95	94.6
Yield of glass, (kg), and wt% oxides present	909.8	72.48	1.87	10.79	14.17	0.21	

^aLoss on ignition. Generally, the oxides of carbon and sulfur (plus some chlorine, depending on the fining agent) volatilize during melting.

^bAlso 0.2 kg MgO or 0.02 wt%.

^cAlso 4.1kg K₂O or 0.45 wt%.

^dUsed primarily to reduce the Fe₂O₃ to FeO to give the characteristic amber color, although the redox state of the glass melt also influences the fining reactions.

(From Ref. 8.)

weighing and mixing of the raw materials. In most companies, batching and mixing operations are usually under computer control, involving automated weighing, mixing, and conveying operations. Great care is to be exercised in entering the batch formulations into the computer, with continual checking for accuracy.

The next step is the testing of the glass produced. The testing program may be based on the following: 1) glasses are melted in large furnaces in a continuous operation, and the chemical and physical properties of any glass are dependent on one another; 2) continuous melting in large furnaces does not create abrupt compositional changes as might be experienced with discrete batch melting; 3) changes take place slowly in the furnace because of the size and volume of glass contained; therefore, periodic sampling and testing is a necessary production control. Measurement of selected physical properties serves to determine the values of all other relevant physical properties and also serves to define the chemical composition (4).

Samples of glass can be obtained daily from every producing position and subjected to a variety of tests. Each test is performed sufficiently often to ensure product quality. The following properties are determined:

- Seal stress
- Softening (viscosity) point
- Annealing and strain points
- Density
- Light transmission

- Chemical durability
- Chemical analysis

For most manufacturers, seal stress is determined on a daily basis and is a primary indicator of any shift in glass composition. Typically, a sample of daily production is flame-sealed or fused to a reference glass of the same composition having the desired target properties. Any deviation in composition is reflected in the generation of stress in the seal area. The physical property inferred from this test is thermal expansion, which is highly sensitive to changes in alkali content that, in turn, influence the chemical durability. If the test results show sufficient deviation from internal specifications, corrective action is initiated by issuing a slightly modified batch sheet. The effects of the corrective action should be observed in the next day's production.

The other tests mentioned may be performed in time intervals ranging from weekly to biweekly and monthly. The results of all these tests provide a framework describing glass quality on an ongoing basis. Thus, if the daily seal stress and the weekly viscosity points and biweekly chemical durability tests are satisfactory, the monthly chemical analysis will be right on target. It also follows that glass produced at any time, although not specifically tested for durability, will, in fact, be satisfactory if the other tests are satisfactory. In addition, an interlocking database can be used to certify to the pharmaceutical manufacturer that glass produced at any time meets specified requirements.

Process of Forming Pharmaceutical Containers

Pharmaceutical containers can be made by a blow-molding process and by a glass rod or tube-shaping process. Commonly, the large-volume parenterals of 100 ml or more are made by a blow-molding process, whereas small-volume parenterals of 100 ml or less are made by the tube-shaping process.

Although for the purpose of this section, discussion will be limited to pharmaceutical containers, molten glass can be formed in all kinds of shapes and sizes, such as bottles, jars, etc. Molten glass is either molded, drawn, rolled, or quenched, depending on the desired shape and use.

Blow-Molding Process

The blow-molding process uses hot glass as it leaves the furnace. The glass must be at the correct temperature and viscosity for forming to be successful. This process describes a typical blow-molding process. At the exit end of the furnace, the glass flows through an orifice in the bottom of the furnace section called the feeder. A pair of cooled blades cuts the glass stream into discrete chunks called gobs. The gobs contain the correct volume of glass to make one container. The gob is delivered into blank molds where it is settled with compressed air and, ultimately, forced to conform to the interior shape of a cast-iron mold that represents the bottle. This preformed (cast-iron mold shape) stage is obtained with a counter-blow of compressed air. The formed bottle is inverted and transferred into the blow mold where it is further finished by blowing. The mold is hinged so that it opens and allows the removal of the bottle. The bottle is sent through a controlled cooling process, known as annealing, which allows the bottle to cool down to room temperature in a stress-free condition. A press-and-blow machine, such as the Hartford-Empire machine, is an example of a blow machine used for making articles such as drinking glasses. A typical process of blow-and-blow machine is shown in Fig. 2.

Blow-molded containers can be made in clear glass or glass that has been colored for protection of the product from light. In comparison to the tubing process which will be discussed later, the walls of blown containers are usually thicker than those of the tubing process, and, hence, blow-process containers have greater impact strength. However, containers made from tubing process have more uniform distribution of glass in the walls. This may be critical, depending on its intended use, and may impact on optical-inspection equipment. Blow process containers also have thicker bottoms, and this could affect heat transfer during lyophilization.

Tubing Process

The Danner process is a common method used for glass tubing, which is a mechanical process. There is also the hand drawn process described below. Glass tubing can be made with a precisely controlled outside diameter (OD) and wall thickness. There is an upper size limit of about 40-mm OD, which is not a limitation on the tubing-forming process, but on the subsequent forming operations used to make containers. In the hand drawn process, about 15 m of 140-mm OD, 2-mm wall thickness tubing can be made on one draw by a gaffer and two helpers. Gathers of glass up to 30–35 kg can be made with bubble in their center at the end of the blow pipe. This is rotated in place by the gaffer, and rotated and stretched out by one of the helpers walking away from it. The size and the diameter of the tubing is a function of the helper's walking speed and is further controlled by a second helper that cools the tubing by fanning at specified times. Up to 150 m of the much smaller diameter thermometer tubing can be made in this manner from one gather (1, 4, 9).

In the Danner process, a continuous stream of glass, of the proper temperature, flows slowly and controllably onto a rotating mandrel, which is an inclined cylinder made of refractory ceramic materials or of platinum alloy. Glass tubing is drawn off the end of the mandrel, which is tapered, while air blowing through the mandrel helps to maintain tubing dimensions. This air is pressurized to keep the tubing from collapsing toward the end. The glass temperature controls the diameter and the wall thickness of the tubing, the inflating air pressure, and the rate of withdrawal of glass from the cylinder. A schematic representation of the Danner process is shown in Fig. 3.

When tubing has cooled sufficiently, it is cut into lengths convenient to handle and is used as feed stock for the container-forming process. Hairpin cords of inhomogeneity and volatilization from the surface are capable of occurring during the tubing process. The tubing process can be used to make ampuls, vials, syringe barrels, cartridges, and a variety of containers used in medication delivery (4).

Another type of tubing process is the Vello process. In the Vello process, the molten glass passes down through an annulus between a horizontal ring and a vertical bell. The moving stream connects to a nearly horizontal runway, and as in the Danner process, it is drawn off at a suitable rate for dimensional specification and stability. The Vello process forms tubing faster because the glass is cooled more quickly to the appropriate forming viscosity in a stationary forehearth of the desired dimension. However, the Vello

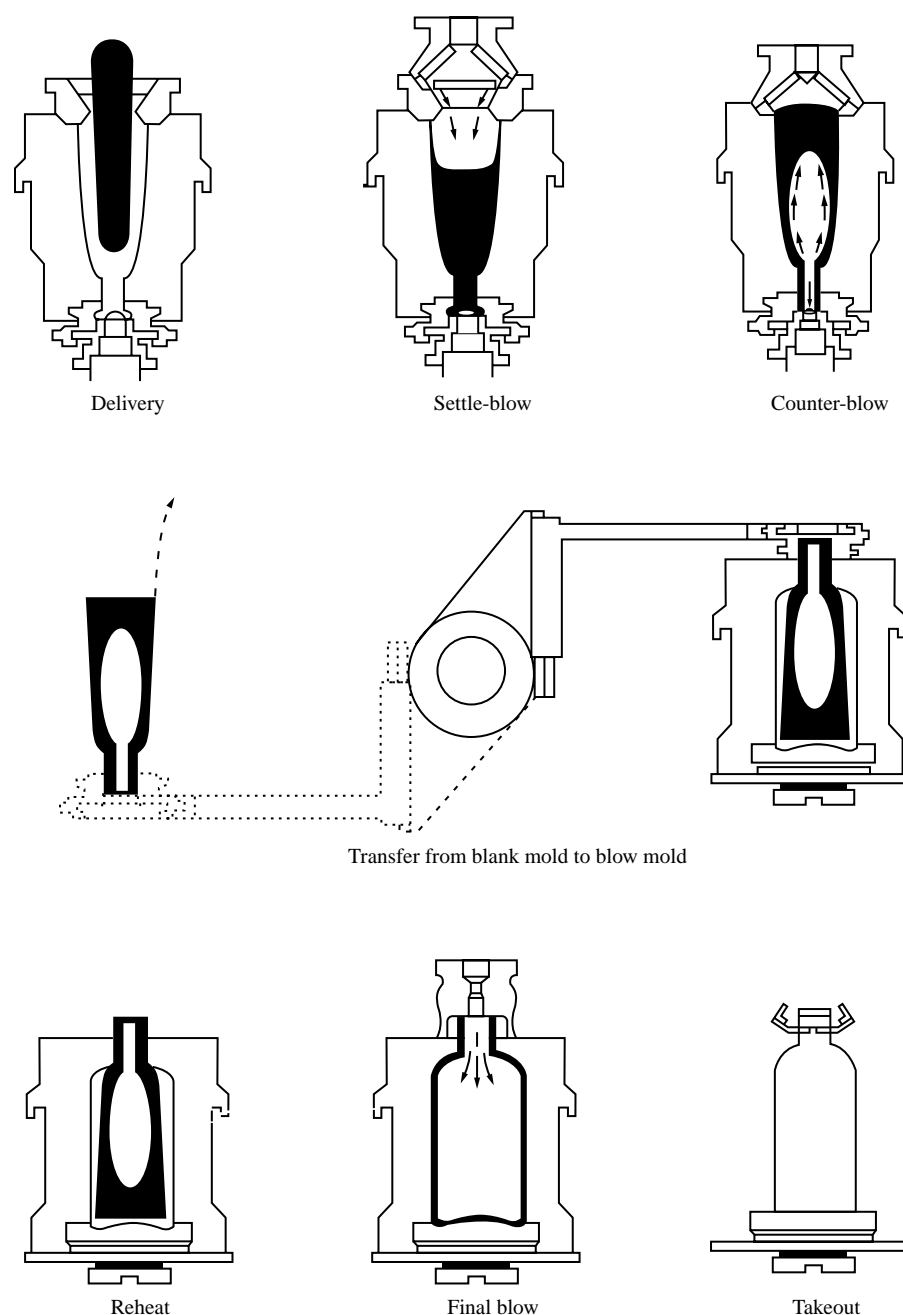


Fig. 2 The H.E. IS blow-and-blow machine (9). The gob is delivered into a blank mold, settled with compressed air, and then preformed with a counter-blow. The parison or preform is then inverted and transferred into the blow mold where it is finished by blowing. (From Ref. 9.)

process is more difficult to operate and is most suitable for longer production runs having few size changes (1).

In comparison to sized molded containers from the blow-molded process, tubing process containers are made lighter in weight and provide more precisely controlled

dimensions than blow-molded containers. The walls and bottom thickness are more uniform, and, hence, are more suited for use in automatic inspection Systems. Clear and light resistant glass compositions can be used in the tubing process.

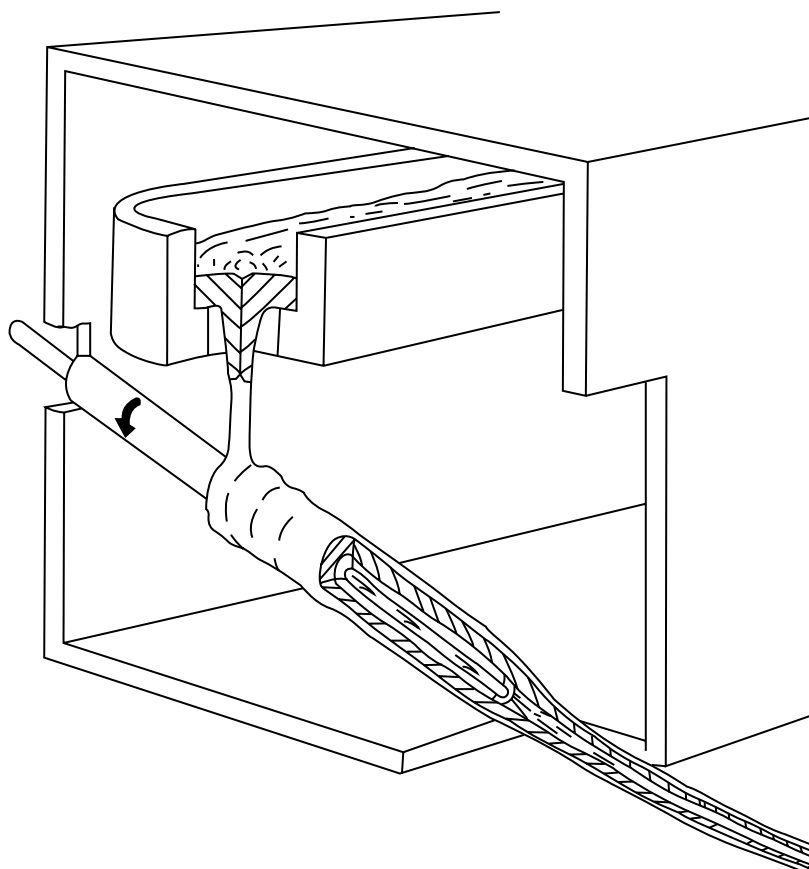


Fig. 3 The Danner process for manufacture of tubing.

GLASS COMPOSITION, PROPERTIES, AND CLASSIFICATION

As mentioned earlier, the focus in this article is on soda-lime-silica glasses and borosilicate glasses.

Soda-Lime-Silica Glasses

These compositions are for glass for every day use. The batch materials for soda-lime glasses are easily available at reasonably moderate cost. These materials readily meet at moderate temperatures. These glasses have a relatively high thermal expansion coefficient and only moderate resistance to attack by a contained product.

The actual compositions of soda-lime are usually more complex than the term soda-lime suggests. As mentioned earlier, these glasses may contain MgO , Al_2O_3 , BaO , or K_2O and various colorants, in addition to Na_2O , CaO , and SiO_2 . Alumina increases the durability of soda-lime glasses, whereas MgO prevents devitrification. Soda-lime glasses account for 90% of all glass produced and are used

for containers, flat glass, pressed and blown ware, and some types of lighting products (1).

In soda-lime glasses, the amber color is developed by the interaction of iron oxide and a small amount of sulfur (a few tenths of a percent), in the presence of a reducing agent during melting. As such, soda-lime amber glasses are called reduced ambers. Typical compositions of soda-lime and borosilicate glasses are presented in Table 2 (4).

Borosilicate Glasses

Borosilicate glasses are developed to meet more stringent requirements than soda-lime glasses. Because of their composition, these glasses require higher melting temperatures and more costly batch materials. As a result, these glasses are more expensive than soda-lime glasses.

Borosilicate glasses are also known as heat-resistant glasses because they have a lower coefficient of thermal expansion than soda-lime glasses and, hence, are more resistant to fracture from rapid temperature changes

(thermal-shock resistance). Borosilicate glasses are used widely as laboratory apparatus, chemical process piping and drain lines, and for baking and cooking dishes in the home. Borosilicates have excellent resistance to chemical attack, and this augments their use as laboratory apparatus and in packaging of pharmaceuticals as containers. The fluxing action of the boron facilitates melting by weakening the network. This is due to the presence of planar three coordinate boron, which weakens the silicate network at high temperature. Borosilicates are divided into two groups based on the coefficient of thermal expansion.

In borosilicate glasses, the amber color is developed through interaction between iron oxide and titanium oxide or iron oxide and manganese oxide. Reducing agents are not used in borosilicate glasses for color development. Borosilicate ambers are called oxidized ambers.

PROPERTIES OF GLASS

Rheology, the melting, forming, annealing procedures, and limitations of use at high temperature, is determined by the viscosity of the glass. Viscosity is measured between 10(13) and 10 Pa.5 [10(14) and 100 P]. Also, viscosity of glasses is compared qualitatively. The addition of modifiers to a glass can alter the viscosity at certain temperatures. At low temperature, the effects of a modifier on viscosity are controlled by its coordination number. Modifiers with high coordination numbers tend to

increase low temperature viscosity due to packing restraints (1).

ASTM provides definition for the procedures and selected reference points discussed in this chapter. Some frequently used reference points are annealing points and strain points. Below are presented some of the ASTM (3):

- *Annealing*: a controlled cooling process for glass designed to reduce residual stress to a commercially acceptable level and to modify structure.
- *Annealing Point (A.P.)*: that temperature corresponding either to a specific rate of elongation of a glass fiber when measured by Test Method C336 or a specific rate of midpoint deflection of a glass beam when measured by Test Method C598. At the annealing point of glass, internal stresses are substantially relieved in a matter of minutes.
- *Annealing Range*: the range of glass temperature in which stress in glass can be relieved at a commercially practical rate. For purposes of comparing glasses, the annealing range is assumed to correspond with the temperature between the A.P. and the strain point (St.P.).
- *Strain Point (St.P.)*: that temperature corresponding to a specific rate of elongation of a glass fiber when measured by Test Method C336 or a specific rate of midpoint deflection of a glass beam when measured by Test Method C598.
- *Softening Point (S.P.)*: that temperature at which a glass fiber of uniform diameter elongates at a specific rate under its own weight when measured by Test Method C338. The viscosity at the softening point depends on

Table 2 Typical compositions and thermal expansions of soda-lime and borosilicate pharmaceutical glasses

Weight %	Soda-Lime			Borosilicate				
	Blow molded		Tubing	Blow molded		Tubing		
	Clear	Amber	Clear	Clear	Amber	Clear	Clear	Amber
SiO ₂	73.0	71.9	67.7	67.8	66.7	80.4	72.0	69.2
B ₂ O ₃	—	—	1.5	13.6	9.5	12.9	11.4	10.4
Al ₂ O ₃	0.8	1.9	2.8	5.8	5.3	2.6	6.8	5.4
CaO	10.7	10.6	5.7	1.1	1.8	<0.05	0.5	0.4
MgO	0.4	0.8	3.9	<0.1	<0.1	<0.05	0.2	0.3
BaO	—	—	2.0	2.4	1.2	—	—	2.1
Na ₂ O	13.5	14.1	15.6	8.3	7.4	4.0	6.4	6.0
K ₂ O	0.3	0.3	0.6	0.8	1.0	—	2.4	2.3
TiO ₂	—	—	—	—	—	—	—	2.8
Fe ₂ O ₃	—	0.3	—	—	1.3	—	—	1.0
MnO	—	—	—	—	6.0	—	—	—
Thermal expansion, cm/cm/°C, 0–300°C	88	90	93	56	59	33	55	54

(From Ref. 4.)

the density and surface tension. For example, for a glass of density 2.5 g/cm^3 and surface tension 300 dynes/cm , the softening point temperature corresponds to a viscosity of $10^{6.6} \text{ Pa}\cdot\text{s}$.

Discussion

At the strain point, internal stresses are substantially relieved in a matter of hours (3).

The Margules Viscometer, a calibrated instrument that measures the force exerted by molten glass on a rotating spindle can be used to measure viscosity. Glass is usually melted and fined at viscosities between 5 and 50 Pa·s (50–500 P). However, the forming and final viscosity requirements may differ. Hard glasses usually have a high softening point whereas soft glasses have a lower softening point. The length of a glass (i.e., long or short) can be used to explain the differences between the strain point and the softening point of glass. A long glass usually has a large difference between the strain point and the softening point, meaning that it solidifies slower than a short glass as the temperature decreases (1).

The temperature required for glass to form into useable shapes is usually above 1000°C for borosilicates and less than 1000°C for soda-lime. The annealing process is very effective in relieving the residual stresses contained in formed glass (10). The temperature range for proper stress-relief annealing of borosilicate is $600\text{--}650^\circ\text{C}$ and less for soda-lime. These forming stresses are relieved within minutes at these temperatures, and a subsequent slow cooling to room temperature retains the stress-free state.

Thermal Expansion

The value of thermal expansion is important in determining how a glass container survives sudden changes in temperature, that is, thermal shock resistance. The thermal expansion of a glass determines the range of materials to which it can safely be sealed. The thermal expansion characteristic indicates how much change in dimension (e.g., length) the article undergoes upon a change in temperature. The change in dimension in turn determines the stress generated in the glass. The ability of a glass as a heat exchanger thermal barrier and its ease of melting and forming depend on its heat-transfer properties and emissivity. In most cases, glass expands when heated and contracts when cooled. If the thermal cycle is slow enough, there is no hysteresis effect. The slope of linear expansion vs. temperature is known as the thermal expansion coefficient, α , which is virtually constant between 0 and 300°C for most glasses. However, as the temperature of the glass rises to near the set point (strain

point $+5^\circ\text{C}$), the thermal expansion increases more rapidly. Glasses used to the extreme limits are vulnerable to thermal shock, and tests should be made before adapting the final design for any use.

For soda-lime glasses, when expansion is high, the stress is high, and the probability of fracture is greatly increased. Borosilicate glasses have lower coefficients of expansion than soda-lime compositions and, hence, can withstand larger temperature changes without fracture. This is important in processes where relatively rapid temperature changes occurs such as in dry-heat sterilization and lyophilization, among other processes.

Stresses caused by steady-state thermal gradients may or may not cause failure, depending on the degree of constraint imposed by some parts of the item upon others or by external mounting. Thermal stress resistance (face-to-face temperature differentials) that causes tensile stress is observed in some types of glasses. When glass is suddenly cooled, as by the removal from a hot oven, tensile stresses are introduced in the surfaces and compensating compressional stresses in the interior. On the other hand, sudden heating leads to surface compression and internal tension. In both cases, stresses are temporary and disappear once temperature uniformity is reached. Also, because glass fractures only in tension, usually at the surface, the temporary stresses from sudden cooling are much more damaging than those from sudden heating, assuming all surfaces are heated and cooled at the same time. Thermal shock endurance is generally determined by empirical testing because the strength of glass is greater under momentary stress than under prolonged load. Resistance to breakage can be determined by heating the ware to some appropriate temperature then plunging it into cold water. For example, a resistance of 150°C means that no breakage occurred on heating to 150°C and plunging the glass into water at 15°C . A much higher value of thermal shock can be recorded when other cooling media such as air or oil are used (1).

The thermal shock that a container receives as its temperature is lowered depends on the temperature differential and the time required to reach the lowest temperature. Glass as a material can withstand very low temperatures, but sufficient time must be allowed to reach the low temperature in order to avoid breakage. Although it might seem that thick glass walls withstand thermal shock better because of higher strength, the fact is that thin walls resist thermal shock better because temperature change is accommodated more rapidly, lessening the stress created by the temperature differential (4).

The thermal expansion coefficient depends on the glass composition. It is usually assumed that the lower the expansion the better, but this is not necessarily true. It is

true, if soda-lime glass with an expansion of about 90 is compared to borosilicates with expansion coefficients of 33–55. However, there is little to choose from between borosilicates in this range of expansion values. Glasses with expansion coefficients in this range perform satisfactorily under most circumstances. If the application requires cooling rates higher than normally experienced, the lowest expansion glass is of course the choice.

GLASS CONTAINER CHARACTERISTICS: COMPARISON OF BLOW-MOLDING AND TUBING PROCESS

As described earlier, some containers are made from tubing and some by the blow-molding process. Containers made from tubing have a maximum capacity of about 50 ml and a 40-mm OD. Blow molding is more suitable for larger sizes and is used for containers greater than 100 ml in volume. This flexibility in size causes great variation in wall and bottom thickness and weight (4).

Tubing used for containers is produced with closely held dimensions, including wall thickness, wall uniformity, cross section, and straightness. Because these dimensional attributes are strictly controlled, smaller volume containers can be made to much better dimensional precision and accuracy. At the same time, containers made from tubing are likely to exhibit different glass defects than blow-molded containers, mainly because of the kind of defects associated with the manufacture of tubing. Gas bubble inclusions, for example in blow-molded bottles, have a spherical or ovoidal shape, whereas those made from tubing have “air line” gas bubbles that have been stretched out during the forming of the tubing.

Another distinction between the two types of containers is the interior surface composition. For glass to be readily formed, it must be heated to a sufficiently high temperature. Because of this, the more volatile glass constituents tend to escape the surface and condense in cooler regions. This effect is minimal in blow molding, because every part of the forming process leads to lower temperatures, reducing the tendency to vaporize. Containers from tubing, on the other hand, are formed by reheating tubing in specific regions of a tubing length. The result is that some parts of a vial, for example, experience very high temperatures, whereas other parts are barely above room temperature. A typical example of this is the forming of the vial bottom. The glass temperatures required cause vaporization of sodium and boron oxides, together with some chlorides; these compounds condense on the interior vial sidewall, just above the bottom. These so-called

forming deposits or “bloom” can be removed by reevaporation during the annealing step, or if this process is incomplete, by washing later.

A special discussion of tubing containers for lyophilization is warranted because a proper design embodies a combination of container and product characteristics that should be taken into account. Blow-molded bottles for lyophilization, although subject to some manufacturing control, cannot be made as uniquely as tubing containers. The same design considerations apply but are under much less control.

The following considerations have been found to be important in the design and use of containers for lyophilization:

- Product characteristics
- Amount of fill
- Thermal shock
- Glass composition
- Container wall thickness
- Container contours
- Container surface damage

The great variability in products to be lyophilized implies that there will be behavior variations in these products as they freeze. Any glass container considered for use should be evaluated with the actual product, rather than trying to simulate product behavior with test solutions.

The result of product freezing and thawing is internal pressure generation. If the container is overfilled, excessive pressure can result with subsequent container breakage. Fills of less than 50% of the container volume are recommended, with an optimum of about 35%.

The contour of the container is important in resisting the forces encountered during product freezing and in ensuring adequate heat transfer through the bottom. There should be as gradual a transition as practicable between the sidewall and the bottom, to reduce the stress concentration effect caused by this angle. A sharp, re-entrant angle between the sidewall and bottom is the worst condition. In addition, a flat bottom with a little “push up” in the center is best for heat transfer. This condition is in concert with a gradual transition in the heel region; both conditions improve performance.

Everything mentioned above as factors for satisfactory performance can be negatively affected by damage to the surface of the container. The container must be as free as possible of scratches, scuffs, impact damage, etc., if the design criteria are to be effective. Maintaining a damage-free glass surface is also a requirement of the filling line, where contact with sharp metal objects should be minimized. The factors discussed above demonstrate that other factors are of importance beside low glass thermal

expansion, usually the first and sometimes the only criterion considered (4).

Chemical Property—Durability

ASTM defines chemical durability as:

the lasting quality (both physical and chemical) of a glass surface. It is frequently evaluated, after prolonged weathering or storing, in terms of chemical and physical changes in the glass surface, or in terms of changes in the contents of a vessel (3).

One of the main reasons for using glass compared to other containers or packaging material is due to its resistance to chemical corrosion. The chemical durability of a glass varies from highly soluble to highly durable, depending on its composition and the solvent used. Glass used in packaging parenteral products and in direct contact with parenteral liquids or solids must have good chemical durability. Analysis are usually based on measurements of weight loss, changes in surface quality of the glass or finished container, or analysis of solutions that were in contact with a glass. A method of determining the durability of a glass is by subjecting grains of uniform size distribution to accelerated attack by high purity water at a temperature characteristic of terminal steam sterilization. In addition, glass compositions can be directly compared with respect to their resistance to attack by aqueous solutions because glass grain are prepared and sized uniformly. These processes will establish the intrinsic durability of the glass. However, the presence of other factors, such as glass constituents that may condense on the surface during high temperature forming process or other volatile deposits, can reduce or improve the durability.

When glass is attacked by water under accelerated test conditions, the pH of the solution increases as a result of an ion exchange process between the alkali (primarily sodium) content of the glass and the hydrogen ions in solution. The higher the pH, the less durable the glass. At large increases in pH, there are high effects on the contained parenteral product and there is an accelerated attack on glass. The extent of pH change can then be determined by titration of attacking solution with diluted acid. Glasses are rated based on the volume of diluted acid required to neutralize the extracted alkali. Additionally, the extent of attack can be determined by the analysis of the solution for specific constituents, including sodium, boron, aluminum, calcium, and silicon. The amount of acid required to neutralize the extract solution will correlate with the amount of specific glass constituents found in the solution by direct analysis. The most durable glass will require smaller volumes of acid and lower concentration of the constituents in solution.

These observations apply generally to both glass grain tests and container tests, except that the acid volumes and constituent concentration are much lower when testing containers because of a much lower glass surface-to-solution volume ratio.

The reaction of acids with glass may be either a leaching process or a complete dissolution process. Acids such as hydrofluoric acid attack silica glasses by dissolving the silica network. Other acids such as hydrochloric acid or nitric acid may react by dissolving certain glasses. However, the reaction mechanism is by selective extraction of alkali and the substitution of protons in a diffusion controlled process.

The reaction of bases with most silicate glasses produces dissolution rates when tested in 5% NaOH solution at 95°C. The mechanism also involves a complete dissolution process as that described for acid. Weaker alkaline solutions may both leach and dissolve and sometimes show greater dependence on glass composition. And, in the case of strong alkali solutions, the rate of attack doubles for each 10 K increase in temperature or each increase in pH unit. Usually higher alkali durability glasses are used for laboratory wares.

Also, as stated earlier, the attack of water, is related to the leaching mechanism described for acid. Low alkali, high alumina, or borosilicate glasses generally have high water durability. Weathering of glass is the result of the action of water, carbon dioxide, and other constituents. Water initially is absorbed by the glass and then exchanged for alkalies that form alkali salt solutions and, if left in contact with glass, may cause additional damage. As a result, weathering resistance may not correlate with acid durability. Test methods to accelerate the weathering process are designed in chambers at 50°C and 98% rh.

A comparison of the chemical durability of soda-lime glasses and borosilicate glasses show that borosilicates are far more durable than soda-limes, requiring from 10 to 20X less acid to neutralize solutions in glass-grain tests. This is due to the significantly lower alkali content of the borosilicates. The same result applies when comparing containers made from borosilicates and soda-lime. Thus, when product-container interactions must be limited, as in the case of parenterals, the compositions of choice are the borosilicates.

SPECIAL TREATMENTS

Special treatments include treatments to the container after it has been formed into its final shape. These processes can be performed before or after the annealing process. The

processes discussed here are concerned with printing and “sulfur” treatment to improve the chemical durability of the interior container surface.

Printing Treatment (4)

The printing of pharmaceutical containers may involve either applying the printed information as a step in the manufacture of the container or applying of labels as the filled container is processed by the pharmaceutical manufacturer. The printing of information on the container while still in the hands of the glass manufacturer is significant. A number of issues are connected with this, one of which is the strict accountability by the glass producer for containers printed with information on drug type, lot number, etc. Stringent safeguards must be employed during printing to ensure the integrity of the container lots.

Material for printing on glass can include ceramic glazes, epoxy formulations cured by heat, and ultraviolet curing formulations. The latter two are applied after annealing, as the temperatures applied would destroy these materials. Ceramic glazes are applied before the annealing step and have properties such that the annealing temperatures are sufficient to fire on the glaze.

Ceramic glazes provide the most durable type of printing. They are vitreous or glass-like in nature, and form a strong bond with the glass substrate. Ceramic glazes are chemically the most durable. They provide the greatest resistance to abrasion and the highest hardness. However, the glaze and the glass have to be carefully matched. A great disparity in the thermal expansions of the two materials causes problems. The glaze should have a greater expansion than the underlying glass so that as the container cools from the firing process, the glass container develops compressive stress at the glaze–glass interface, rather than tensile stress. Glass is much stronger in compression than in tension. The glaze develops fine cracks or a crazed appearance during cooling, but the underlying glass will not be significantly damaged.

Ceramic glazes are used in ampuls to introduce a controlled-break site. A band of paint is applied at the constriction of the ampule where controlled breakage is desired, so the contents can be withdrawn. The band is used to make Colorbreak ampuls, and its function is to act as a stress concentrator when bending stress is applied to the ampul to break off the stem for product withdrawal. Very consistent ampul break forces are achieved by the use of ceramic bands.

Interior Surface Treatments (4)

Sometimes forming deposits become fused onto the glass surface and are difficult to remove, in effect compromising the durability of the interior surface. These effects are overcome by chemical neutralization of these deposits through surface treatment processes. At the same time, the intrinsic ability of a glass surface to withstand attack by aggressive products is also improved. Thus, there is a two-fold benefit in treatment processes for containers made from tubing: to remove the residual forming deposits and to improve the basic durability of the glass surface. These benefits are derived concurrently during treatment. Blow-molded containers, although not subject to the forming deposit problem, can also benefit from surface treatment.

If an aggressive product of pH 8 and above is to be packaged in a borosilicate container, surface treatment is required. Otherwise, the glass is attacked by the product, resulting in contamination by both soluble and insoluble reaction products. The latter are manifested by entities variously called flakes, shimmers, etc. These flakes are essentially silica that has been stripped off the surface by the attack on the glass. Their presence is direct evidence of excessive attack. Proper surface treatment enhances resistance to attack and results in a glass container that can contain products of pH 8 and higher without being damaged.

The basis of all treatment processes is the removal of alkali from the glass surface, resulting in improved resistance to attack by aqueous solutions. Alkali or sodium removal accomplishes this by greatly hindering the ion exchange process responsible for glass attack. Exchange of sodium ions in the glass for hydrogen ions in solution results in a pH rise in the solution; this can accelerate attack on the basic glass structure. If hydrogen ions cannot readily leave the solution because there are very few labile sodium ions with which to exchange, there will be little pH rise and, hence, little attack.

A common means of removing surface sodium is by reaction with sulfur dioxide in the presence of oxygen or by reaction with sulfur trioxide. The reaction product is sodium sulfate.

Surface treatment of blow-molded containers consists of several stages. As the freshly formed bottle leaves the mold and before it enters the annealing lehr or oven, it is still high in temperature. The bottle is filled with one or several of the following gaseous mixtures: sulfur trioxide, sulfur dioxide and oxygen, or 1,1-difluoroethane. If only moderate treatment is desired, one step suffices. The reaction of these gases with surface sodium starts immediately because of the glass temperature and

continues as the bottle enters the annealing process at its elevated temperatures. The reaction product of sodium sulfate is clearly seen in the cool bottle as a whitish haze on the inside surface just prior to packing into cartons. It is easily removed by subsequent washing processes. Blow-molded bottles treated by both of these steps can achieve high durability. These two methods, however, are restricted to large containers with relatively wide openings. Smaller containers made from tubing with much more restricted entry must be handled in a different way.

The effective treatment of containers made from tubing is a much more crucial issue because of the adverse effects of residual forming deposits and because the high pH products are usually packaged in small volume ampuls and vials made from tubing. The same basic treatment scheme is used, that of reacting the surface sodium with a sulfur compound at elevated temperatures. A 3% ammonium sulfate solution is injected into the container just prior to entering the annealing lehr; the volume injected is several tenths of a milliliter. The ammonium sulfate decomposes during the annealing process, with temperatures of 600–650°C being typical. The solution is thought to decompose, resulting in the formation of sulfuric acid vapors or possibly sulfur trioxide, which react with the surface sodium and any residual forming deposits, giving the characteristic sodium sulfate haze.

Containers given an effective surface treatment show a pronounced improvement in chemical inertness. Products packaged in such containers remain unaffected for long periods of time. The improvement of the inside surface, resulting from the surface treatment, is permanent and is not destroyed even by repeated autoclaving.

An additional reason for sulfur treating the interior surface of bottles is to improve the durability of soda-lime glass bottles. Normally, the surface inertness of soda-lime glass does not approach that of borosilicate. Sulfur treatment of a soda-lime surface can improve this situation to the point where products not normally considered for soda-lime containers can be packaged in them. The reason for choosing a treated soda-lime bottle for a mild product rather than a borosilicate bottle is based on economics, as soda-lime bottles are less expensive than borosilicate. It should be noted that this approach is used primarily for blow-molded bottles. There is compendial recognition and control of this, using treated soda-lime bottles in this way. Suffice it to say for the present that these bottles are known as Type-II bottles, according to the *United States Pharmacopeia* (USP). The USP and other methods of test and classification of glass for pharmaceutical products are discussed later.

Putting aside the question of testing and classifying Type-II bottles for the moment, it is of interest to consider test methods that determine the effectiveness of interior surface treatment of borosilicate containers made from tubing. It was stated earlier that a measure of durability, or inertness, is the pH rise of a contained solution. There are several commonly used ways of determining this. These methods are based on the pH rise of a water solution in the container after it has been autoclaved at 121°C for 60 min. The pH rise can be evaluated by titrating the solution with dilute acid in the presence of an indicator to a neutral end point. Another way is to include an indicator in the original water fill, autoclave, and observe the color change indicative of pH change. An example of this method uses bromothymol blue indicator adjusted to an initial pH of 5.8–6. The color of the indicator after autoclaving is compared to a series of standards having a range of pH up to at least 7.5 and the pH rise estimated. The autoclave cycle time and temperature used for these tests is that of the *Water Attack Test* as set forth in the USP described later in this chapter.

A solution autoclaved in a well-treated container consumes very small amounts of dilute acid upon titration or shows very small pH rise by color change of the indicator. Actual values of acid required, or pH rise, depend on various factors, including container size, as it controls the interior surface area-to-solution volume ratio (4).

Classification of Glass and Glass Containers—Compendial Perspective

The chemical specifications of glassware was first contained in USP XII in which specifications of glassware as containers for injections were provided. In subsequent revisions changes appeared and definitions of four types of glassware are described in USP XIX (11, 12).

Glass containers for pharmaceutical use are glass articles in direct contact with pharmaceutical preparation. In addition to the USP, the *European Pharmacopoeia* (EP) and other pharmacopeias have grouped glass containers suitable for packaging pharmacopeial preparations into the four different classifications specified in USP XIX up to USP 24 (13). The classes are based on the degree of chemical or hydrolytic resistance of these glasses to water attack. The degree of attack is determined by the amount of alkali released from the glass under the influence of the attacking medium under conditions specified. The quantity of alkali used is extremely small in some cases. These tests designs and glass classification are described from USP 24–NF 19

under section <661> *Containers* (13). These tests are designed to be conducted in areas relatively free from fumes and excessive dust.

Glass Types

The *USP* and *EP* have provided similar classifications that are summarized below.

Type I glass containers

Type I glass containers are comprised of a borosilicate glass with about 80% SiO_2 and 10% B_2O_3 and smaller amounts of Al_2O_3 and Na_2O . It is inert and has the lowest coefficient of thermal expansion. It is least likely to crack when a sudden temperature differential occurs. It is commonly used to make ampuls and vials for parenteral use. It is used for solutions that can dissolve basic oxides to cause an increase in pH that could alter the efficacy or potency of the drug (5).

USP describes Type I glass as:

Highly resistant borosilicate glass, and usually used for packaging acidic and neutral parenteral preparations. Also, where stability data demonstrates their suitability, Type I are used for alkaline parenteral preparations (13).

EP describes Type I glass as:

Neutral glass with high hydrolytic resistance due to the chemical composition of the glass itself. Type I are suitable for all preparations whether or not for parenteral use and for human blood and blood components (14).

Type II glass

A dealkalized form of soda-lime glass with higher levels of Na_2O and CaO . It is less resistant to leaching than Type I but more than Type III. However, to make Type II and other types more resistant to leaching, the surface can be treated with SO_2 to convert surface oxides present to soluble salts that are then washed off. This surface treatment is effective for containers used once and those repeatedly exposed to heat. Type II has a lower melting point than Type I and, therefore, is easier to fabricate. It has a higher coefficient of thermal expansion, and is used in solutions that can be buffered to maintain a pH below 7 (5).

[*USP*] Soda-lime glass that is suitably dealkalized and is used for packaging acidic and neutral parenteral preparations, and also where stability data demonstrates their suitability, is used for alkaline parenteral preparations (13).

[*EP*] Soda-lime silica glass with high hydrolytic resistance resulting from suitable treatment of the surface. These containers are suitable for acidic and neutral, aqueous preparations for parenteral use (14).

Type III glass

A soda-lime glass containing same amount of sodium and oxide levels as in Type II but contains more leachable oxides of other elements. And because of its high reactivity, it is used to package anhydrous liquids and other dry products (5).

[*USP*] These are soda-lime glass containers that are usually not used for parenteral preparations, except where suitable sensitivity test data indicates that Type III is satisfactory for the parenteral preparations that are packaged therein (13).

[*EP*] These are soda-lime glasses with only moderate hydrolytic resistance. They are suitable for nonaqueous preparations for parenteral use, for powders for parenteral use, and for preparations not for parenteral use (14).

Type IV or NP glass

[*USP*] These are general purpose soda-lime glass. They are intended for packaging nonparenteral articles, i.e., those intended for oral or topical use (13).

[*EP*] These are soda-lime silica glass with low hydrolytic resistance. These are suitable for solid preparations that are not for parenteral use and for some liquid or semi-solid preparations that are not for parenteral use (14).

Tests

These glass containers for pharmaceutical use have to comply with relevant tests such as tests for hydrolytic resistance for *EP* (14), and tests chemical resistance for *USP* (13). The test procedure and methods are slightly different for each of the pharmacopeias. However, this article will emphasize the test procedure provided in the *USP 24-NF 19*.

For the four types of glasses, there are designated relevant test types and expected limits. These are provided in Table 3.

USP has provided procedure and test requirements for three types of tests. These are the *Powdered Glass Test*, the *Water Attack* test, and the *Arsenic* test. These tests,

Table 3 USP glass types, test type, and limits

Glass type	Test type	Limits	
		Size ^a (ml)	ml of 0.020 N Acid
I	Powdered glass	All	1.0
II	Water attack	100 or less	0.7
		Over 100	
III	Powdered glass	All	8.5
NP	Powdered glass	All	15.0

^aSize indicates the overflow capacity of the container.
(Courtesy of USP 24–NF19 © USP.)

Apparatus, and Reagents for the tests are described below (13).

Apparatus used for these tests^a

Autoclave: An autoclave capable of maintaining a temperature of 121 ± 2.0°C, equipped with a thermometer, a pressure gauge, a vent cock, and a rack adequate to accommodate at least 12 test containers above the water level is used.

Mortar and pestle: A hardened-steel mortar and pestle, made according to the specifications in the accompanying illustration.

Other equipment: Sieves, about 20.3-cm (8-inch), made of stainless steel including the Nos. 20, 40, and 50 sieves, along with the pan and cover (see *Openings of Standard Sieves* (811)), 250-ml conical flasks made of resistant glass aged as specified, a 900-g (2-lb) hammer, a permanent magnet, a desiccator, and an adequate volumetric apparatus are used.

Reagents used for these tests^b

High-purity water: The water used in these tests has a conductivity at 25°C, as measured in an in-line cell just prior to dispensing, of not greater than 0.15 μs per cm (6.67 Megohm-cm). There must also be an assurance that this water is not contaminated by copper or its products (e.g., copper pipes, stills, or receivers). The water may be prepared by passing distilled water through a deionizer cartridge packed with a mixed bed of nuclear-grade resin, then through a cellulose ester membrane having openings not exceeding 0.45 μm. Do not use copper tubing. Flush the discharge lines before water is dispensed into test vessels. When the low conductivity specification can no longer be met, replace the deionizer cartridge.

Methyl red solution: Dissolve 24 mg of methyl red sodium in purified water to make 100 ml. If necessary,

neutralize the solution with 0.02 N sodium hydroxide or acidify it with 0.02 N sulfuric acid so that the titration of 100 ml of High-purity water, containing 5 drops of indicator, does not require more than 0.020 ml of 0.020 N sodium hydroxide to effect the color change of the indicator, which should occur at a pH of 5.6 (13).

Powdered glass test for Types I, II, and NP glasses^c

Rinse thoroughly with purified water six or more containers selected at random, and dry them with a current of clean, dry air. Crush the containers into fragments about 25 mm in size, divide about 100 g of the coarsely crushed glass into three approximately equal portions, and place one of the portions in the special mortar. With the pestle in place, crush the glass further by striking three or four blows with the hammer. Nest the sieves, and empty the mortar into the No. 20 sieve. Repeat the operation on each of the two remaining portions of glass, emptying the mortar each time into the No. 20 sieve. Shake the sieves for a short time, then remove the glass from the Nos. 20 and 40 sieves, and again crush and sieve as before. Repeat again this crushing and sieving operation. Empty the receiving pan, reassemble the nest of sieves, and shake by mechanical means for 5 min or by hand for an equivalent length of time. Transfer the portion retained on the No. 50 sieve, which should weigh in excess of 10 g, to a closed container, and store in a desiccator until used for the test.

Spread the specimen on a piece of glazed paper, and pass a magnet through it to remove particles of iron that may be introduced during the crushing. Transfer the specimen to a 250-ml conical flask of resistant glass, and wash it with six 30-ml portions of acetone, swirling each time for about 30 s and carefully decanting the acetone. After washing, the specimen should be free from agglomerations of glass powder, and the surface of the grains should be practically free from adhering fine particles. Dry the flask and contents

^aFrom USP 24–NF 19, © USP.
^bFrom USP 24–NF 19, © USP.

^cFrom USP 24–NF 19, © USP.

for 20 min at 140°C, transfer the grains to a weighing bottle, and cool in a desiccator. Use the test specimen within 48 h after drying.

Procedure

Transfer 10.0 g of the prepared specimen, accurately weighed, to a 250-ml conical flask that has been digested (aged) previously with High-purity water in a bath at 90°C for at least 24 h or at 121°C for 1 h. Add 50.0 ml of High-purity water to this flask and to one similarly prepared to provide a blank. Cap all flasks with borosilicate glass beakers that previously have been treated as described for the flasks and that are of such size that the bottoms of the beakers fit snugly down on the top rims of the containers. Place the containers in the autoclave, and close it securely, leaving the vent cock open. Heat until steam issues vigorously from the vent cock, and continue heating for 10 min. Close the vent cock, and adjust the temperature to 121°C, taking 19–23 min to reach the desired temperature. Hold the temperature at $121 \pm 2.0^\circ\text{C}$ for 30 min, counting from the time this temperature is reached. Reduce the heat so that the autoclave cools and comes to atmospheric pressure in 38–46 min, being vented as necessary to prevent the formation of a vacuum. Cool the flask at once in running water, decant the water from the flask into a suitably cleansed vessel, and wash the residual powdered glass with four 15-ml portions of High-purity water, adding the decanted washings to the main portion. Add 5 drops of Methyl Red Solution, and titrate immediately with 0.020 *N* sulfuric acid. If the volume of titrating solution is expected to be less than 10 ml, use a microburet. Record the volume of 0.020 *N* sulfuric acid used to neutralize the extract from 10 g of the prepared specimen of glass, corrected for a blank. The volume does not exceed that indicated in Table 1 for the type of glass concerned (13).

Water attack at 121°C for Type II glasses^d

Rinse thoroughly twice 3 or more containers, selected at random, with high-purity water.

Procedure: Fill each container to 90% of its overflow capacity with High-purity water, and proceed as directed for *Procedure* under *Powdered Glass Test*, beginning with “Cap all flasks,” except that the time of autoclaving shall be 60 min instead of 30 min, and ending with “to prevent the formation of a vacuum.” Empty the contents from 1 or more containers into a 100-ml graduated cylinder, combining, in the case of smaller containers, the contents of several containers to obtain a volume of 100 ml. Place the pooled specimen in a 250-ml conical flask of resistant

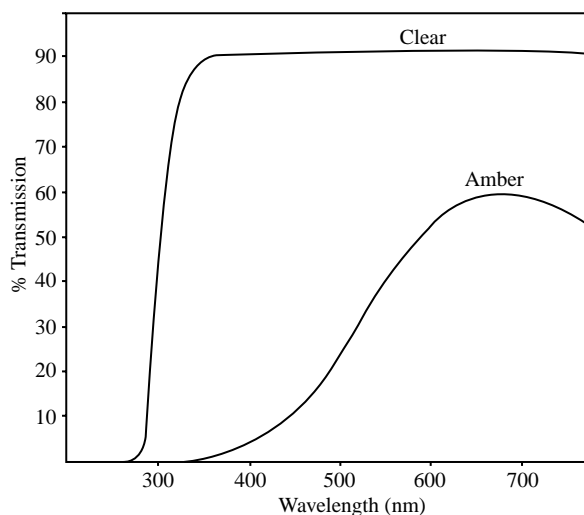


Fig. 4 Typical light transmission of clear and amber borosilicates.

glass, add 5 drops of Methyl Red Solution, and titrate, while warm, with 0.020 *N* sulfuric acid. Complete the titration within 60 min after opening the autoclave. Record the volume of 0.020 *N* sulfuric acid used, corrected for a blank obtained by titrating 100 ml of High-purity water at the same temperature and with the same amount of indicator. The volume does not exceed that indicated in Table 1 for the type of glass concerned (13).

Arsenic test^d

For the *Test Preparation*, 35 ml of the water from one Type I glass container or, in the case of smaller containers, 35 ml of the combined contents of several Type I glass containers, are used and prepared as directed for *Procedure* under *Water Attack at 121°C* and the procedure described for *Arsenic* test in *USP 24–NF 19* general chapter <211> *Arsenic* is then followed. The limit provided for this test is 0.1 ppm.

Light transmission test^d

In addition to the above-mentioned tests, compendial limits are provided for light transmission for colored light protecting glass containers. These containers intended to provide protection from light or supplied as “light Resistant” are expected to meet the requirements for *Light transmission* in this section. Light Transmission characteristics of typical clear and amber borosilicates are shown in Fig. 4.

In this test, a spectrophotometer of suitable sensitivity is used to a cut section of the glass container. The transmittance of the section is measured, and the observed light transmission is not expected to exceed the limits provided in Table 4.

^dFrom *USP 24–NF 19*, © USP.

Table 4 Limits for glass types I, II, and III

Normal size (in ml)	Maximum percentage of light transmission at any wavelength between 2990 and 450 nm	
	Flame-sealed containers	Closure-sealed containers
1	50	25
2	45	20
5	40	15
10	35	13
20	30	12
50	15	10

(From *USP 24–NF 19*, © USP.)

These methods of testing glass are basically similar to other compendial limits and other standards on glass such as those in International Organization for Standardization (ISO), German DIN, and the ASTM. The expected limit and test result expectations and procedures may vary for each standard of compendial method.

There are differences in testing glass as a container compared to glass as a material (glass-grain tests). The glass surface area-to-solution volume ratio is higher in grain tests, resulting in higher concentrations of glass constituents in solution after autoclaving. This facilitates solution analysis and differentiation between glasses. The other major difference is the presence of forming deposits, in the case of containers made from tubing, which influences test results. Blow-molded container test results are not affected by deposits.

SUMMARY

Glass as a packaging material has many advantages. Glass provides:

- Total impermeability to gaseous environmental contaminants
- Total impermeability to loss of essential volatile ingredients by diffusion through the container walls
- Excellent clarity and attractive sparkle
- Ease of cleaning and sterilizing with heat
- Resistance to attack by all liquids except HF and other caustic products
- A variety of shapes and it can accept a wide variety of closure types
- Ease of filling, closing, unscrambling, labeling, and cartoning

- Good compressional strength to allow efficient storage, especially allowing cartons of glass to be stacked high in warehouse
- Ease of hot-filling

Glass has disadvantages that include:

- Breakage—when it breaks, it shatters into numerous sharp fragments
- Weight—with a density of 2–2.5 g/ml, along with brittle character thick container walls, they become quite heavy
- Cost—in some cases, they are very expensive because of their weight and the type of fabrication involved (5)

REFERENCES

1. Boyd, D.C.; Danielson, P.S.; Thompson, D.A. *Glass. Encyclopedia of Chemical Technology*; Kroschwitz, J.I., Howe-Grant, M., Eds.; John Wiley, Inc.: New York, 1994; 12, 555–623.
2. *Glass The American Heritage Dictionary*, 2nd College Ed.; Houghton Mifflin Company: Boston, MA, 1991; 561.
3. ASTM Standard C162-88, Standard Definitions of Terms Relating to Glass and Glass Product. *Annual Book of ASTM Standards*; 1999; 15.02, 1–14.
4. Abendroth, R.P. Glass as a Packaging Material for Pharmaceuticals. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1993; 7, 79–100.
5. Jenkins, W.A.; Osburn, K.R. Packaging Drugs and Pharmaceuticals, Drug Packaging Materials. *Cooperation with Institute of Packaging Professionals*; Technomic Publishing Co.: Lancaster, PA, 1993; 99–102.
6. Goldschmidt, V.M. *J. Soc. Glass Technol.* **1927**, *11*, 337.
7. Woolley, F.E. *Engine Red Materials Handbook*; ASM International: Bilthoven, The Netherlands, 1991; 4, 386–393.
8. Spinosa, W.C.; Stephen, P.M.; Schorr, J.R. *Review of Literature on Control Technology Which Abates Air Pollution and Conserves Energy in Glass Melting Furnaces*; Nov 11, EPA-600/2-77-005, 2-76-269, 2-76-032b Corning, Inc.: Battelle, Columbus, OH, 1977.
9. Giegerich, W.; Trier, W. *Glass Machine Construction and Operation of Machines for the Forming of Hot Glass*; Kreidl, N.J., Ed.; Springer-Verlag: Berlin, 1969.
10. Pincus, A.G.; Holmes, T.R. *Annealing and Strengthening of Glass Industry*; Magazines for Industry, Inc.: New York, 1977.
11. *Containers for Injection, USP XII*; United States Pharmacopeia: Rockville, MD, 1942; 567–576.
12. *Containers, USP XIX*; United States Pharmacopeia: Rockville, MD, 1975; 643–647.
13. *Containers <661>, USP 24–NF 19*; United States Pharmacopeia: Rockville, MD, 2000; 1930–1935.
14. *Containers*; European Pharmacopoeia, Council of Europe: 67075 Strasbourg Cedex France, 1997; 172–185.

GERIATRIC DOSING AND DOSAGE FORMS

Cheryl A. Wiens

University of Alberta, Edmonton, Alberta, Canada

Carol A. Borynec

University Hospital, Edmonton, Alberta, Canada

DEMOGRAPHICS

The word *geriatric* refers to individuals who are over age 65 years. This has been described as the most heterogeneous population because it constitutes fit, physiologically healthy patients to extremely frail, debilitated patients in long-term care facilities (LTCF).

Currently, greater than 12% of the population in the United States (1) is aged 65 years and over. This population is prescribed approximately one-third of all prescription medications (2). On average, geriatric patients use four to five medications, and greater than 50% use over-the-counter (OTC) medications. The use of medications is even higher in the LTCF population.

The elderly are also more likely to suffer from chronic disease and multiple disease states and therefore receive long-term treatment with medications. Considering the high medication utilization of geriatric patients, it is pertinent to focus on geriatric dosing and dosage forms.

PHYSIOLOGIC CHANGES WITH AGING

Although development of disease is not considered normal with aging, there are a number of physiological changes which can predispose the elderly to developing illness (Table 1). Many of the age-related physiological changes have been reviewed previously (3, 4). Because of the physiological changes that occur with the aging process, there are changes in the pharmacokinetics and pharmacodynamics of medications.

PHARMACOKINETICS

Pharmacokinetics refers to how the body handles medications. As the body changes with age, its ability to absorb, distribute, and break down drugs can change significantly. Unfortunately, in some cases, it is difficult to

predict how significantly aging will affect some people, thus, dosing changes alone cannot be based solely on age. Pharmacokinetics can be broken down into four processes: 1) absorption (of the medication through the gastrointestinal tract, a mucous membrane, or the skin); 2) distribution (of the medication throughout the body); 3) metabolism (most often through the liver); and 4) elimination (most often through the kidneys).

Absorption

There appears to be little change in absorption of medications in the elderly (3, 5, 6). Dry skin may decrease absorption of topically applied medications. A decrease in stomach acid secretion may affect absorption of certain medications when given orally; however, this is not consistent for all geriatric patients. Other changes that may contribute to altered absorption may include dietary changes, a decrease in salivary secretion, or use of dentures (3). Absorption is generally dependent on the concomitant disease states of the patient (3). For example, if a patient had congestive heart failure, with significant congestion of the gastrointestinal (GI) tract, then less medication and nutrients would be absorbed through the GI tract. The decreased absorption is not related to age but results from concomitant systemic or GI diseases.

Distribution

Physiological changes owing to the aging process will result in an increase in lipid mass and a decrease in lean body mass (3, 5, 6). Fat-soluble medications may distribute more widely and remain in the body longer in an elderly patient, whereas water-soluble medications have a smaller volume of distribution because of decreased lean mass.

Distribution of medications is also determined by protein binding in the serum (6, 7). The most common protein is albumin, which binds predominantly acidic drugs such as phenytoin. Another common protein is α -1 acid glycoprotein, an acute-phase reactant, which has a higher affinity for basic drugs such as lidocaine. The most

Table 1 Age-related physiological changes in the elderly

Organ system	Change
Body composition	↓ lean body mass ↑ body fat ↓ body water
Cardiovascular	↓ cardiac output ↓ stress response (β -response blunted)
Central nervous system	↓ peripheral conduction velocity ↓ weight, volume of brain
Endocrine	↓ hormonal secretions menopausal changes ↑ incidence diabetes, thyroid atrophy
Gastrointestinal	↓ secretions ↓ rate of stomach emptying ↓ rate of intestinal transit time ↓ liver volume and blood flow
Genitourinary	Atrophy vagina, prostate
Immune	↓ cell-mediated immunity
Pulmonary	↓ elasticity, chest wall compliance ↓ alveolar surface
Renal	↓ nephrons ↓ creatinine clearance ↓ renal blood flow ↓ glomerular filtration rate (GFR)
Sensory changes	↓ accommodation of lens of eye
Skeletal	↓ skeletal bone mass
Skin and hair	↓ hydration of skin ↓ in dermal thickness

↓ Indicates decrease; ↑ indicates increase.

important consideration for protein levels when considering drug distribution is to take into account the health status of the individual. An elderly patient may experience only a small decline in protein levels owing simply to age. However, chronic inflammatory disease states may significantly increase the levels of α -1 acid glycoprotein, thereby decreasing the free level of medication in the blood stream. When considering albumin, it is important to take into account chronic disease states and nutritional status. Poor nutritional intake will result in a lower level of albumin and, therefore, a higher free level of drug. In most cases a higher free level of drug does not result in a clinically significant response (because the liver most often compensates for the increased drug levels and simply metabolizes the excess levels with hepatic reserve); however, this may be quite important for medications with a narrow therapeutic index. Medications such as warfarin require very careful monitoring to maintain the International Normalized Ratio (INR) within a very narrow

therapeutic range. In this case, if warfarin increases even slightly, the response could be far greater than expected.

Second, tissue protein binding of medication can affect drug disposition in the elderly patient. For example, amiodarone binds strongly to tissue proteins, which greatly increases its volume of distribution. In the elderly or chronically ill patient, tissue proteins may change or decline, which can also affect the serum levels of medications.

Finally, free drug levels may be affected by interaction with other medications, which may cause displacement of a drug off the protein. Considering that the elderly are taking more medications, they are more likely to experience protein-displacement drug interactions. However, protein-displacement interactions are not considered as clinically significant as metabolism interactions.

Metabolism

Liver blood flow decreases with age, but overall it does not appear to affect the metabolizing capacity of the hepatocytes. Medications that are dependent on liver blood flow for break down may have a somewhat delayed metabolism (3, 5, 6). There is some decline in phase I metabolism but not necessarily in phase II metabolism (5). Phase I metabolism involves the oxidation/reduction reactions commonly needed to metabolize medications. Phase II reactions, including conjugation reactions, appear to remain grossly intact in the elderly. This means that medications that are metabolized by phase II metabolism are preferred over agents that require phase I metabolism. In elderly patients who appear frail or suffer from chronic disease, both phase I and phase II metabolisms may show a decline. However, age alone does not appear to cause a decline in phase II metabolism. It is important to consider the chronic disease states or lifestyle factors, such as diet, smoking, and ethanol intake, in predicting the metabolizing capability of a patient (3, 5, 6). Patients who have poor diet and negative lifestyle factors such as cigarette smoking and high consumption of alcohol are less likely to have hepatic metabolizing reserve and are more likely to suffer adverse reactions from medication because of relative overdosing (or incapacity to metabolize a standard dose of medication). Although the elderly may have little change in their metabolizing capacity owing to age only (6) applying the above principles to dosing is recommended.

Elimination

Renal function generally decreases with age (3, 6). It is not possible to predict which patients will not experience a decline in renal function, even though it has been estimated that up to 30% of the elderly have no change in their renal

function. The most commonly used nomogram to predict renal function is the Cockcroft–Gault equation (8), which factors in age, causing a decline in renal function. Unfortunately, this equation was not originally validated in an elderly, frail, or female population. Furthermore, predicting renal function is most difficult in frail patients. Careful monitoring for response to the medication and side effects is necessary to avoid an excessive dosage. It is also difficult to predict renal function in these clinical situations: malnourished patients who have low muscle mass and, therefore, low serum creatinine; patients who are missing a limb; and individuals who do not have stable renal function (creatinine not having reached steady state).

Medications that are eliminated renally should be given in a reduced dosage or given less frequently than in younger patients. However, the exact dosing cannot be predicted accurately in many subpopulations of the elderly including the very old (>85 years) nor in the frail patients. In these populations, it is appropriate to monitor drug response based on clinical improvement or signs and symptoms of toxicity (3, 5, 6). Monitoring drug levels at steady state may also be necessary.

PHARMACODYNAMICS

Pharmacodynamic changes in the elderly are also documented (5, 9); however, they are not as well-researched or defined as are pharmacokinetic changes. Pharmacodynamics refers to the change in response to the medication. Possible mechanisms include (9):

- Change in receptor density
- Change in receptor affinity and receptor characteristics
- Postreceptor alterations (change in transduction signal coupling or amplification)
- Desensitization of receptors
- Altered negative feedback responses
- Target tissues exhibit intrinsic change

These alterations in homeostatic function can result in significant impairment or adverse consequences if these changes are not accounted for while prescribing certain medications (Table 2).

Changes in Sensitivity

The elderly are more likely to be sensitive to a number of different classes of medication because of changes in pharmacodynamics (5, 9–11). It is very important to monitor and adjust the dosage of certain classes of medications in the elderly owing to their increased or decreased sensitivity (Table 3).

ADVERSE DRUG REACTIONS (ADR) / ADVERSE DRUG EVENT (ADE)

As Lamy (12) has pointed out, there is much debate regarding the definition of an adverse drug reaction (ADR) or an adverse drug event (ADE). An ADR has been defined by the World Health Organization (WHO) as a noxious, unintended effect of a drug that occurs in doses normally used in humans for the diagnosis, prophylaxis, or treatment of disease. An ADR also can be defined as an undesirable effect of drug therapy beyond its anticipated therapeutic effects occurring during clinical use. An ADE is defined as a noxious and unintended patient event caused by a drug (e.g., laboratory abnormalities, symptoms, signs, etc.). In the outpatient geriatric population, approximately 30% will experience an ADR and are responsible for approximately 10–20% of hospital admissions (12–14). In hospitalized patients, the incidence of an ADR or ADE is estimated to be 10–25% (15). Studies have found that the majority of events are avoidable (12), and unfortunately, many patients (estimated to be one-third) do not fully recover from the ADR.

ADRs are also a concern because of the associated costs, primarily related to hospitalization. ADRs increase length of stay and risk of mortality, which overall increase costs of the ADR treatment. The elderly with ADRs also visit their physician more frequently than elderly patients who are not experiencing ADRs (16). The elderly patient may not bring up concerns with healthcare professionals (12) and may go on to accept the ADR as a part of normal treatment. The incidence of ADR therefore may be significantly underreported in the elderly population.

Assessing the risk of ADRs in the elderly can be difficult because most clinical trials exclude the elderly because of comorbidities or age. There is also a bias in prescribing, thus, elderly patients at risk of experiencing adverse effects from a drug may not even be exposed to the medication. This provides an underestimate of the true incidence of ADRs in an entire population who may take the medication (5, 12, 17).

The risk of experiencing an ADR is related to the diagnosis, types of medications prescribed, and the number of drugs used (5, 12–14, 17). The most significant factor is multiple drug use and is related to number of coexisting diseases. When the number of medications is decreased, it has been shown that ADRs also decrease (12). It appears that the number of medications increases the risk of ADR exponentially (14).

It is debatable if age alone is an independent risk factor for ADR; however, the elderly are more susceptible to ADRs because of physiological changes resulting in

Table 2 Altered homeostatic mechanisms in the elderly

Altered function	Mechanism	Examples of drugs
Anticoagulation	Poor hepatic production of coagulation factors	Anticoagulants
Arrhythmias	Poor dietary intake	Thrombolytics
Higher cognitive function	Cardiac hypersensitivity	Antiarrhythmic medication
	Central cholinergic transmission	Central anticholinergics
	Neuronal loss	Stimulants β -agonists
	Receptor downregulation	
Orthostasis	Blunting of β response (no tachycardia)	Blood pressure medications
	Changes in vascular tree	Tricyclic antidepressants
	Changes in autonomic nervous system	Antipsychotics
	Receptor down regulation	Diuretics
Postural control	\downarrow D2 receptors in the striatum	Sedative hypnotics
	\uparrow Risk of sway	
Tardive dyskinesia	Impaired dopamine-synthesizing neurons	Traditional antipsychotics
Thermoregulation	Poor temperature-regulating mechanisms: \downarrow shivering \downarrow metabolic rate \downarrow vasoconstriction \downarrow thirst response \downarrow subjective awareness of temperature	Long-term antipsychotic therapy Medications affecting awareness, mobility, muscular activity, vasoconstrictor mechanisms CNS medications Phenothiazines Barbiturates Benzodiazepines Tricyclic antidepressants Narcotics Alcohol
Visceral muscle function	Visual disturbances (pupillary autonomic responses) Bladder instability (bladder capacity, detrusor contractions) Intestinal motility decreased	Anticholinergic medications

\downarrow Indicates decrease; \uparrow indicates increase.

Table 3 Medications affected by changes in sensitivity

Medication (class)	Altered effect	Recommendation
ACE-I	Greater sensitivity to BP reduction Increased risk hyperkalemia	Lower initial dose
Antiarrhythmics (class I)	Longer elimination $t_{1/2}$	Lower initial dose
Antiarrhythmics (class III)	Longer elimination $t_{1/2}$	Lower initial dose
Anticholinergics	Increased sensitivity to anticholinergic effects, increased risk confusion	Avoid if possible
Antipsychotics	Greater risk orthostasis and movement disorders	Use lowest initial dose when necessary
		Lower initial dose
β -blockers	Low potency agents—more likely to cause confusion	
Benzodiazepines	Greater sensitivity to BP reduction Longer elimination $t_{1/2}$ Greater sensitivity to sedating effects	Lower initial dose Use shorter-acting hydrophilic agents such as lorazepam, oxazepam
Corticosteroids	Increased sensitivity to GI complications	Lower initial dose
Digoxin	Increased risk osteoporosis Decreased volume of distribution	Minimize duration of therapy if possible
Diuretics	↓ responsiveness owing to decline in renal function	Lower loading dose
NSAIDs	Greater risk of GI complications	Lower maintenance dose Enhanced monitoring of renal function, electrolytes
		Used only after acetaminophen for OA
Opioids	Increased sensitivity to effect Longer elimination $t_{1/2}$	Recommended with GI-protective agent or use a COX-II inhibitor
SSRIs	Longer elimination $t_{1/2}$	Lower initial dose
Tricyclic antidepressants	Greater sensitivity to anticholinergic effects, arrhythmic effects	Avoid using fluoxetine
Warfarin	Increased sensitivity to anticoagulant effect	Preference for desipramine, nortriptyline (better tolerated secondary amines)

ACE-I, indicates angiotensin-converting enzyme inhibitor; SSRI, selective serotonin receptor inhibitor; NSAID, nonsteroidal anti-inflammatory drug; OA, osteoarthritis.

altered pharmacokinetics or pharmacodynamics. Some studies have shown that when controlling for confounders such as clinical status of the patient, number of medications, and length of hospital stay, age alone does not predict ADRs (12, 17).

It is important to note that withdrawal of medications may also lead to ADR (18). Certain settings (e.g., LTCF) encourage prescribers to remove medications, some of which may have been used by the patient for many years. If the withdrawal of therapy is not appropriate or is done too quickly, this may cause ADR for the patient (18, 19).

The most common drugs involved in ADRs are cardiovascular medications, aspirin, NSAIDs, and psychotropic agents. However, it should be noted that these are also the most commonly used medications in the elderly (14).

The prevention of ADRs in the elderly has been inadequately studied. Considering that the majority of ADRs are type A (dose-related) it would appear that most ADRs in the elderly could be prevented. Additional study on this subject is necessary to determine the most effective interventions in decreasing the incidence of ADRs in the elderly. A review by Atkin (14) has addressed studies that suggest improvements such as better history-taking and record-keeping by physicians to minimize the number of medications used by patients. Other suggestions include requesting that patients always bring their medications to medical appointments and using medications that would be considered lower risk in this population.

There are a number of different ways to assess and identify an ADE. Listed below are a number of tables that will assist in interpreting signs or symptoms that may indeed be an adverse drug event. Table 4 identifies disease states that may contribute to similar symptoms that may be misdiagnosed as an ADE. Other syndromes that are most commonly associated with toxicity or side effects from specific medications are shown in Table 5. Finally, medication classes that most commonly causes ADEs in the elderly are presented in Table 6.

Falls

Falls are common in the elderly and are often multifactorial (20–22). There is a concern about patients who fall because the complications are significant. Complications include fractures, soft tissue injuries, immobilization and hospital-acquired illness, institutionalization, and even death from additional complications such as pneumonia (20). It is difficult to determine the percentage of patients who fall because many falls do not result in injury and are never reported.

The cost involving treatment of falls is staggering (21, 22), including costs for hospitalization and acute

care, rehabilitation, and institutionalization, if necessary. It has been estimated that falls cost \$12.6 billion in lifetime expenses for persons older than 65 years of age. There is also a tremendous emotional impact on the patients, causing many patients to lose their sense of security and to remain housebound.

Although certain illnesses contribute to falls, medications have been shown to cause falls independent of other factors. The most commonly offending drugs are benzodiazepines because they have been shown to increase falls and hip fractures (23). An association between the dose (the higher the dose, the more likely the fall), duration of use, and type of benzodiazepine (e.g., long-acting medications) has also been reported. Other classes of medications that increase the risk of falls include tricyclic antidepressants, SSRIs, and opioid analgesics (23, 24).

It may be entirely appropriate to continue using these medications if they are prescribed. To ensure safe drug use, patients should be cautioned about the risk of falling and encouraged to use extra precautions. High-risk medication should be administered only when needed, rather than on a scheduled basis, because this will hopefully decrease the overall dose. In the elderly, a short-acting phase II metabolized benzodiazepine such as oxazepam or lorazepam would be preferred if a benzodiazepine is necessary.

Delirium

Drug-induced changes in cognition are common and disturbing ADRs for both the patient and the patient's family or caregiver. Delirium is commonly reported on hospital admission (approximately 10–15% of elderly patients meet criteria for diagnosis of delirium) (25, 26). There is also a significant occurrence of delirium in elderly patients undergoing various types of surgery, including general surgery open heart surgery and hip fracture repair (may be as high as 50%) (25, 26). The outcome after diagnosis of delirium is poor. At 1 month, the mortality has been shown to be as high as 14%, and at 6 months mortality is 22%. If the patient survives, change in cognitive symptoms may last for 3 months postdischarge. Furthermore, these patients are at an increased risk for other complications, including prolonged hospital stay, functional decline, and institutionalization.

Medications may cause, or at least exacerbate, delirium in many patients. Polypharmacy has been shown to be a risk factor for delirium. There are a number of different types of medications that may cause delirium. Many have anticholinergic properties, whereas others have a yet-unknown mechanism for causing delirium (Table 4). Patients who present with acute changes in mental status

Table 4 Presenting symptoms and associated adverse drug events (ADE)

Symptom	Underlying condition	Drug cause of ADE
Disturbed mental status/delirium	Hypoglycemia Hypothyroidism B12 deficiency Uremia	Anticholinergic medications and/or properties Antipsychotics Antihistamines Antiparkinsonian agents Antispasmodics Ophthalmic preparations OTC sleep/allergy medications Tricyclic antidepressants Other medications Analgesics/NSAIDs Anticonvulsants Corticosteroids (high dose) Digoxin H2 blockers Insulin Muscle relaxants Narcotics Psychotropics (anxiolytic, antidepressant, antipsychotic) Sedative/hypnotic Sulfonylurea
Depression	Hypercalcemia Hypo-hyperthyroidism	Hypnotics Amiodarone Lipid-soluble β -blockers
Fatigue	CHFGI bleed Anemia Hypothyroidism	Diuretics β -blockers Hypnotics Muscle relaxants
Falls/syncope	Structural CNS lesion Dehydration Hypokalemia Arrhythmias	Antiarrhythmics Levodopa Diuretics Antihypertensives TCA Sedatives Antipsychotics Hypoglycemics Alcohol

or delirium must be evaluated medically and should include a review of their medications. Typically, all medications should be discontinued and restarted only if considered life-saving at the time. Long-term chronic medications should be held because it is important to resolve delirium and attempt to identify which medications may be the offending agent (25–27).

Drug–Drug Interactions

Drug interactions are very common in the elderly, attributable in part to the high number of medications

prescribed in this population. Drug interactions are considered one type of adverse drug event and have been found to increase with the number of medications prescribed (28). Drug interactions are also a common cause of hospital admission (29).

The elderly are particularly susceptible to interactions because of the changes in physiology, multiple physician prescribing, increased use of OTC medications, nonadherence to complex regimens, and multiple disease states that complicate the handling of medications.

Medication interactions have been found to be poorly monitored. Medications are often necessary, and

Table 5 Common adverse effects

Adverse effect	Possible medications
Anticholinergic effects	Antihistamines
Delirium	Tricyclic antidepressants
Dry mouth	Medications for urge incontinence
Constipation	
Urinary retention	
Tachycardia	
Disturbed vision	
Blurring	See anticholinergic effects
Dry eyes	Glaucoma medications Ocular lubricants
Extrapyramidal effects	Antipsychotics
Tremor	Antiemetic (metoclopramide)
Pseudoparkinsonism	
Akathisia	
Sedation	Antipsychotics Sedative/hypnotics Tricyclic antidepressants Muscle relaxants
Orthostatic hypotension	Antihypertensives Diuretics Levodopa Tricyclic antidepressants

Table 6 Medication classes commonly affecting mobility in the elderly

Medication class	Mobility ADE
Antihypertensives	Postural hypotension
Antipsychotics	Postural hypotension Sedation Extrapyramidal effects Falls
Narcotics	Sedation Confusion ↓ coordination (falls)
Sedative/hypnotics	Sedation Weakness Confusion ↓ coordination (falls)
Tricyclic antidepressants	Postural hypotension Sedation Arrhythmias Falls

↓ Indicates decrease.

interactions can be handled in a manner in which the patient does not experience adverse effects and yet receives the benefit of appropriate medications.

There are various types of drug interactions. One classification is pharmacokinetic or pharmacodynamic interactions. Pharmacokinetic interactions describe an interaction with absorption, metabolism, or elimination. Medications can also displace each other through competitive protein binding. Many pharmacokinetic interactions result through competition or interaction at the cytochrome enzyme system called the CYP450 system (28). Numerous isozymes, or subtypes, have been discovered. Each isozyme is responsible for metabolizing specific medications. These isozymes may be inhibited or induced by certain disease states or by other medications. Pharmacodynamic interactions refer to additive or antagonistic activity or action of one medication on another. The study of pharmacokinetic interactions is often done in younger, healthy patients and may not reflect what commonly occurs in an elderly patient. It may also be difficult to predict the extent of the interaction; however, some studies have resulted in specific recommendations on how to adjust the dose of interacting medications.

Table 7 contains a list of drug classes and clinically significant drug interactions. This is not a comprehensive list, and the reader is referred to the appropriate chapter dealing with pharmacokinetic drug interactions.

Drug–Food Interactions

Table 8 lists commonly used drugs and potential food interactions. This is not a comprehensive list, but rather highlights many of the clinically significant food–drug interactions (30).

Recent concerns with food products include grapefruit juice, which has been shown interact with many different medications through the CYP450 3A4 isoenzyme. This may increase the levels of medication such as cyclosporine, dihydropyridine calcium channel blockers, midazolam, triazolam, and astemizole (31).

DRUG MANAGEMENT PROBLEMS IN THE ELDERLY

Adherence

Traditionally called compliance, the concept of not taking medication properly is now called nonadherence. However, these words are often used interchangeably. The documented rates range from 25 to 50% in the elderly (32). Intentional nondadherence, or consciously making a decision to not take the medications as prescribed, involves not taking the medication at all or using only

Table 7 Drug–drug interactions of commonly used medications in the elderly

Drug	Interacting drug	Mechanism/effect	Management
Analgesics NSAIDs	Warfarin	Increased risk of bleeding	Avoid use of NSAIDs while using warfarin
	Corticosteroids	Increased fluid retention; increased risk of GI bleeding	Avoid use of NSAIDs while using corticosteroids; use selective COX-II inhibitors if necessary
Opioids	Alcohol	Increased risk of CNS depression; respiratory depression	Avoid use of multiple CNS depressants; decrease doses of CNS depressants when used concomitantly
	Antipsychotics		
	TCA		
Acetylcholinesterase inhibitors	Sedative/hypnotics		
	Phenothiazine antipsychotics; tricyclic antidepressants; anticholinergic medications		
Acetylcholinesterase inhibitors (A.Ch.E.-I) (donepezil, tacrine, rivastigmine)	Bethanechol	Antagonize acetylcholine increasing effect of A.Ch.E.-I	Avoid use of anticholinergics when using A.Ch.E.-I
Donepezil	Succinylcholine	Additive cholinergic side effects (e.g., nausea, abdominal cramping)	Avoid use of bethanechol with A.Ch.E.-I
	Ketoconazole	Exaggerated muscle relaxation	Discontinue A.Ch.E.-I before surgery
Tacrine	Quinidine	Inhibited metabolism of donepezil	Monitor for donepezil toxicity; dose reduction not empirically necessary
	Fluvoxamine	Decreased metabolism of tacrine leading to increased levels of tacrine	Avoid use of interacting medications; monitor for signs of toxicity of tacrine
	Cimetidine		
	Quinolones		
Antibiotics Aminoglycosides	Theophylline	Inhibition of theophylline metabolism	Decrease dose of theophylline while using tacrine
	Diuretics	Increased risk of dehydration, renal impairment	Monitor renal function, hydration closely
	Cisplatin	Increased risk of nephrotoxicity	Monitor renal function closely
Quinolones	Vancomycin		
	Iron supplements antacids (aluminum-calcium-or magnesium-containing)	Impair absorption of quinolone	Avoid taking quinolone within 2 h of these medications
Ciprofloxacin	Theophylline	Reduced clearance of theophylline and caffeine; increased risk of CNS toxicity	Decrease dose of theophylline; avoid caffeine intake
	Caffeine		
Sulfonamides	Warfarin	Displacement of medications from protein binding sites; transient increase in drug levels and toxicity	Monitor for toxicity of displaced medications
	Sulfonyleurea		
	Phenytoin		
	Methotrexate		

(Continued)

Table 7 Drug–drug interactions of commonly used medications in the elderly (*Continued*)

Drug	Interacting drug	Mechanism/effect	Management
Isoniazid (INH)	Phenytoin	Inhibition of metabolism—increased toxicity of anticonvulsants	Monitor levels of anticonvulsants during INH therapy
	Carbamazepine	Increased risk of hepatotoxicity	Avoid use of ethanol during INH treatment
Anticonvulsants Phenytoin	Ethanol		Monitor for phenytoin toxicity; reduce dose as needed
	Amiodarone	Increased level of phenytoin	
	Cimetidine		
	Fluoxetine		
	Fluconazole		
	Ketoconazole		
	Carbamazepine	Decreased level of phenytoin	Increase dose of phenytoin to reach therapeutic blood level
	Rifampin		
	Theophylline		
	Isoniazid	Increased risk of hepatotoxicity	Monitor hepatic function closely while using this combination
Carbamazepine	Erythromycin	Increased levels, toxicity of carbamazepine	Monitor toxicity of carbamazepine; dosage may require reduction
	Cimetidine		
	Diltiazem		
Gabapentin	Phenytoin	Decreased levels of carbamazepine	Increase dose of carbamazepine as necessary
	Phenobarbitone		
	Theophylline		
Valproic acid	Antacids (aluminum- or magnesium-based)		
	Carbamazepine	Decreased levels of gabapentin	Do not administer antacids within 2 h of gabapentin
	Aspirin	Decreased level of valproic acid; enzyme induction	Increase dose of valproic acid as needed
Antidepressants SSRI (citalopram, fluvoxamine, fluoxetine, sertraline, paroxetine)	Rifampin	Increased free level of valproic acid—risk of valproic acid toxicity	Monitor valproic acid toxicity over time—if toxicity continues, consider decreasing dose
		Decreased clearance of valproic acid	Decrease dose of valproic acid while taking rifampin
	Monoamine oxidase inhibitors (MAO-I)	Increased toxicity of MAO-I (e.g., seizures, confusion, hyperpyrexia)	Avoid combination of SSRI and MAO-I
	Tricyclic antidepressant (TCA)		
	Diazepam	Increased toxicity of TCA	Avoid combination of SSRI and TCA
		Increased levels of diazepam	Avoid diazepam use if possible in the elderly; decrease dose if used with fluoxetine
		Increased effect of diazepam	Start with smallest dose of trazodone; decrease dose of trazodone if necessary
	Trazodone	Increased levels of trazodone	

(Continued)

Table 7 Drug–drug interactions of commonly used medications in the elderly (*Continued*)

Drug	Interacting drug	Mechanism/effect	Management
Sertraline Fluvoxamine	No clinically significant interactions		
	Astemizole	Increased levels of astemizole; increased risk of cardiac arrhythmia	Avoid use of astemizole with fluvoxamine; use alternate antihistamine
Paroxetine Citalopram	Benzodiazepines (triazolam, alprazolam)	Increased levels of benzodiazepine	Decrease dose of benzodiazepines
	Theophylline, caffeine	Increased levels of theophylline or caffeine	Decrease dose of theophylline by approximately 1/3; avoid caffeine intake
	Warfarin	Increased effect of warfarin; increased risk of bleeding	Monitor INR; decrease dose of warfarin if necessary
	Phenytoin Cimetidine	Decreased effect of phenytoin Increased levels of citalopram	Monitor phenytoin; increased dose may be necessary Monitor for side effects of citalo pram; reduce dose if necessary
Nefazodone	Azole (e.g., fluconazole, ketoconazole) and erythromycin	Increased levels of citalopram	Monitor for side effects of citalo pram during course of antibiotics
	Benzodiazepines (alprazolam, triazolam, midazolam)	Increased levels of benzodiazepines	Decrease dose (50–75%) of benzodiazepines
Bupropion	Levodopa	Increased effect from levodopa	Monitor for Levodopa toxicity; dose of Levodopa should be reduced if necessary
	Anticonvulsants	If anticonvulsant is used to control seizure disorder, bupropion may increase risk of seizures	Avoid use of bupropion in patients with history of seizure disorder
Venlafaxine Mirtazapine	MAO-I	Increased risk of toxicity	Avoid use of MAO-I while taking venlafaxine
	No clinically significant interactions		
TCA	Anticholinergic agents (e.g., antihistamines, low-potency neuroleptics)	Enhanced anticholinergic response	Avoid anticholinergic TCAs (use desipramine, nortriptyline when TCA is necessary); avoid combination with other anticholinergic agents
Cardiovascular			
Antihypertensives	NSAIDs	Decreased renal function, fluid retention; increased BP	Avoid use of NSAIDs if possible; use minimal dose when necessary
	Vasodilators (e.g., nitroglycerin)	Additive drop in BP; risk of orthostasis and falls	Lower BP to goal only; use nonpharm techniques to prevent falls
ACE-I	K ⁺ sparing diuretics; K ⁺ supplements	Increased K ⁺ retention; risk of cardiac arrhythmias	Monitor K ⁺ ; avoid K ⁺ -sparing diuretics while on ACE-I
Digoxin	Amiodarone	Increased level of digoxin	Lower dose of digoxin; smaller dose of amiodarone
	Diuretics	Hypokalemia, digoxin toxicity	Monitor K ⁺ ; supplement when necessary
Amiodarone	Warfarin	Increased levels of warfarin, potentiation of anticoagulant	Monitor INR; decrease warfarin dose
	Digoxin	Digoxin levels increased; increased risk of digoxin toxicity	Monitor digoxin levels, signs/symptoms of toxicity; decrease digoxin dosing

(Continued)

Table 7 Drug-drug interactions of commonly used medications in the elderly (*Continued*)

Drug	Interacting drug	Mechanism/effect	Management
Diuretics	Quinidine	Increased (free) levels of quinidine	Decrease dose of quinidine by 30–50%
	Procainamide	Increased levels of procainamide	Decrease dose of procainamide by 30% when starting amiodarone
	Phenytoin	Increased levels of phenytoin	Monitor phenytoin levels and s/s of toxicity; decrease phenytoin dosing if necessary
	NSAID/aspirin	Decreased fluid removal, minimized response of BP lowering effect	Avoid NSAID use if possible in the elderly
Warfarin	Lithium	Competition for lithium elimination; increased lithium levels	Avoid use of diuretics while using lithium; lower dose of lithium if necessary
	Amiodarone	Increased anticoagulant effect	Lower dose of warfarin (titrate to goal INR)
	NSAIDs/aspirin	Increased risk of bleeding	Minimize NSAID use while on warfarin
	Estrogen supplements	Pharmacodynamic interaction; procoagulant effect of estrogen	Increased dose of warfarin may be needed; estrogens contraindicated in patient with coagulation disorder
Endocrinology	Antibiotics (e.g., cotrimoxazole, ciprofloxacin, fluconazole, ketoconazole)	Increased levels, effect of warfarin	Monitor INR carefully during course of antibiotics; reduce dose of warfarin if necessary
	Calcium supplements	Decreased bioavailability of L-thyroxine	Take thyroid supplement on an empty stomach
	Sucralfate		
	Cholestyramine		
Sulfonylurea (e.g., glyburide, glipazide)	Colestipol	Alteration in L-thyroxine requirements (may ↑ or ↓)	Monitor thyroid function and adjust L-thyroxine as necessary
	Iron supplements	May block hypoglycemic responses	Use cardioselective β-blockers (e.g., metoprolol, atenolol); patient education to monitor for sweating
	Aluminum hydroxide	Enhanced hypoglycemic response	Monitor for signs/symptoms of hypoglycemia; manipulate diet or reduce dose of sulfonylurea
	Amiodarone		
Metformin	β-blockers		
	Fluconazole		
	High-dose salicylates		
	Sulfonamides		
Gastrointestinal	Contrast dye	Lactic acidosis and organ failure	Avoid use of contrast dye while on metformin; avoid renally toxic medications while on metformin
	Cimetidine	Decreased elimination, increased levels of metformin	Avoid use of cimetidine while taking metformin; consider using ranitidine or famotidine
	Carbamazepine		
	Diazepam	Inhibition of metabolism	Avoid use of cimetidine if possible; use ranitidine or famotidine
Cimetidine	Glipizide		
	Phenytoin		
	Theophylline		
	TCA		
Warfarin	Warfarin		

(Continued)

Table 7 Drug–drug interactions of commonly used medications in the elderly (*Continued*)

Drug	Interacting drug	Mechanism/effect	Management
Proton pump inhibitors	DigoxinIntraconazole	Decreased absorption of medications that are pH-dependent	Monitor response to medication; increase in dose may be necessary
	Iron salts		
Metoclopramide	Ketoconazole	Antagonism of dopamine by metoclopramide	Avoid use of metoclopramide in patients with Parkinson’s disease
	Dopamine replacement		
	Parkinson’s disease medications	Medications that slow GI motility antagonize the effect of metoclopramide	Avoid use of antagonistic medications
	Anticholinergics		
Narcotics			
Neuroleptics			
Neuroleptics	Antiparkinson’s medications	Neuroleptics may exacerbate or cause Parkinsonian movements	Avoid use of neuroleptics (except clozapine) in Parkinson’s patients
	Antihypertensives	Increased risk of orthostasis, hypotension	Monitor BP in patients; reduce dosage of medications as necessary to prevent falls, dizziness
Haloperidol			
	Carbamazepine	Increase metabolism, decreased effectiveness of haloperidol	Increase haloperidol dose as necessary
Respiratory			
Theophylline	Phenytoin	Decrease serum levels of theophylline	Increase theophylline dosing as needed
	Rifampin		
	Carbamazepine		
	Allopurinol	Increase serum levels of theophylline	Decrease theophylline dosing to prevent toxicity
	Cimetidine		
	Ciprofloxacin		
Corticosteroids	Erythromycin	Increased hyperglycemia from steroids; loss of control of blood glucose levels	Monitor serum glucose closely; increase hypoglycemic medications or add insulin as necessary
	Antidiabetic medications	Increased GI intolerance; risk of GI bleeding	Avoid use of concomitant therapy if possible
	NSAIDs	Increased risk of hypokalemia	Monitor K ⁺ and replace as necessary
	Amphotericin B loop diuretics		

ACE-I, indicates angiotensin converting-enzyme inhibitor; BP, blood pressure; CNS, central nervous system; GI, gastrointestinal; INR, international normalization ratio; K⁺, potassium; NSAID, nonsteroidal anti-inflammatory drug; SSRI, selective serotonin receptor inhibitor; TCA, tricyclic antidepressant.

Table 8 Clinically significant drug and food interactions in the elderly

Class of drug	Interaction	Clinical significance	Recommendation
Anti-infectives			
Antibacterials			
Azithromycin	Food decreases absorption	May lower concentration of antibiotic by >50%	Space medication and food at least 2 h apart
Erythromycin	Food decreases absorption; food may decrease some GI upset	Most patients have GI complaints and may prefer to take erythromycin with food; however, this may lead to decreased absorption	Space medication and food 2 h apart if patient can tolerate; if GI complaints, take erythromycin with small snack
Fluoroquinolones (e.g., ciprofloxacin)	Dairy products, iron, multivalent cations decrease absorption of fluoroquinolones	Quinolones ineffective because of chelation	Administer medication at least 2 h apart from any foods
Metronidazole	Alcohol-containing products may cause a disulfiram-like reaction	Nausea, vomiting, vasodilation may be experienced by patients	Avoid all alcohol-containing products while taking metronidazole
Penicillins (oral)	Food decreases absorption	May lower concentration of antibiotic	Space medication and food at least 2 h apart
Tetracyclines	Dairy products, iron, multivalent cations decrease absorption of tetracycline	Tetracycline ineffective because of chelation	Administer medication at least 2 h apart from any foods containing cations
Antifungals			
Griseofulvin	High-fat meal increases absorption	If medication given without a fatty meal, griseofulvin ineffective	Administer with fatty foods
Ketoconazole	Possible insulin-sparing effect	Patient may experience lower blood glucose concentrations on the same insulin dose	Monitor for decreased insulin requirements
	Interference with vitamin D and steroid metabolism	Osteomalacia may develop if long-term administration of ketoconazole is used	Ensure adequate vitamin D and calcium intake Monitor calcium and phosphorus levels if ketoconazole is used long-term
Antituberculars			
Isoniazid	May decrease pyridoxine (vitamin B6); food may decrease absorption of isoniazid	Peripheral neuropathy may develop; decreased absorption of isoniazid may lead to therapeutic failure	Supplement individuals with vitamin B6 50 mg daily; administer isoniazid at least 2 h apart from any foods
Rifampin	May decrease effect of vitamin D	Individuals may be predisposed to osteomalacia	Ensure adequate vitamin D and calcium intake during long-term therapy

(Continued)

Table 8 Clinically significant drug and food interactions in the elderly (*Continued*)

Class of drug	Interaction	Clinical significance	Recommendation
Anticoagulants Warfarin	Vitamin K—green leafy vegetables and green tea antagonize warfarin	Large amounts vitamin K can decrease INR leading to therapeutic failure of warfarin	Avoid excessive consumption of green leafy vegetables; do not alter diet once anticoagulation with warfarin is stabilized
Anticonvulsants Phenytoin	Reduced levels of vitamin D due to increased metabolism of active vitamin D metabolites Administering medication with enteral feeds binds phenytoin Decrease in folate concentrations	Osteomalacia and rickets have been reported with long-term anticonvulsant therapy; additional risk factors for these conditions (e.g., poor sunlight exposure) increases risk Enteral feeds can significantly lower absorption of phenytoin Long-term therapy may put patients at risk of folate deficiency	Ensure adequate supplementation of vitamin D, 400 IU daily Hold enteral feeds for minimum of 1 h while administering phenytoin Close monitoring of gingival hyperplasia and megaloblastic anemia if patient is using long-term phenytoin therapy; consider supplementing folate starting at 1 mg daily for prevention
Antidepressants MAO-I (e.g., isocarboxazid, phenelzine, tranylcypromine) Tricyclic antidepressants (e.g., amitriptyline, desipramine, nortriptyline, etc.)	Tyramine-containing foods lead to excessive sympathetic stimulation Medications inhibit breakdown of food because of decreased saliva production; food in general may cause xerostomia	Hypertensive crisis can develop May cause difficulty in swallowing food	Avoid aged, fermented, pickled, or smoked foods Suck on sugarless hard candy, ice chips, or chew sugarless gum
Antihistamines Chlorpheniramine, dimenhydrinate, diphenhydramine, etc.	Medications inhibit breakdown of food because of decreased saliva production; food in general may cause	May cause difficulty in swallowing food	Suck on sugarless hard candy, ice chips, or chew sugarless gum
Autonomic medications Ephedrine, pseudoephedrine, amphetamines	Caffeine may cause increased nervousness and insomnia	Most significant if taken before bedtime	Minimize daily caffeine intake while using stimulating medications

(Continued)

Table 8 Clinically significant drug and food interactions in the elderly (Continued)

Class of drug	Interaction	Clinical significance	Recommendation
Cardiovascular medications ACE-I (e.g., lisinopril, enalapril, ramipril, etc.) Captopril	Salt substitutes containing potassium	Potassium-containing substitutes and an ACE-I can result in hyperkalemia	Avoid potassium-containing salt substitutes
	Food decreases absorption of captopril	Taking with food may minimize blood pressure-lowering effects	Space medication and food 2 h apart or take at the same time every day
	Zinc deficiency with high-dose, long-term use	Most pronounced in individuals receiving at least 150 mg daily; taste impairment a marker of zinc deficiency	Monitor for taste impairment; zinc supplementation has not been definitively shown to be of benefit
	High-fiber, high-pectin foods delay and decrease absorption	May minimize the benefit of digoxin	Avoid administering digoxin with high-fiber foods; administer at the same time every day
Potassium-sparing diuretics (e.g., triamterene)	Potassium-containing salt substitutes may lead to additive effect of increased potassium	Excessive amounts of potassium may lead to cardiac arrhythmias	Avoid potassium-containing salt substitutes
Corticosteroids Hydrocortisone, prednisone, methylprednisolone	Decreased glucose tolerance with carbohydrates	Rapid, small elevation in blood glucose can be seen. Glucose levels do not normalize until steroid therapy is discontinued	Monitor blood glucose levels in persons with diabetes
	Triglycerides and cholesterol may increase	Long-term use may accelerate atherosclerotic processes	Monitor patients at risk of heart disease or elevated cholesterol/triglyceride levels on long-term steroid therapy
	Possible insulin sparing effect with ketoconazole	Patient may experience lower blood glucose concentrations on the same insulin dose	Monitor for decreased insulin requirements
Diabetes medications Insulin	Interaction with the absorption of vitamin B12	Patient may experience vitamin B12 deficiency	Monitor levels closely in patients predisposed to vitamin B12 deficiency
Metformin			
Gastrointestinal medications Misoprostol	Increases GI motility, nausea, distress if taken on empty stomach	Taken without food can dramatically increase GI side effects	Take with food

(Continued)

Table 8 Clinically significant drug and food interactions in the elderly (*Continued*)

Class of drug	Interaction	Clinical significance	Recommendation
Sucralfate	Food (protein) decreases binding of sucralfate to gastric mucosa	Ineffective binding leads to lack of protection for gastric ulcers	Give sucralfate at least 1 h before meals to allow adequate binding to the gastric mucosa
Lipid lowering medications			
Cholestyramine	Binds iron, folic acid, essential fatty acids, vitamin A	Cholestyramine can significantly lower absorption of nutrients	Consider vitamin supplementation and monitor iron status for long-term use of cholestyramine
	Calcium absorption may be impaired	Long-term use of cholestyramine may increase risk of osteoporosis	Ensure adequate calcium, vitamin D intake by spacing supplements at least 2 h from administration of cholestyramine
Lovastatin	Improved absorption if given with food	If taken on an empty stomach lipid-lowering effect may be minimized	Take with evening meal or snack to maximize absorption of lovastatin
Pulmonary medications			
Theophylline time-release	High-fat meals—rate of absorption can be affected	Elevated theophylline levels may result, possibly causing tachycardia, palpitations, irritability, and tremor	Avoid administering medication with high-fat foods, altering diet while on the medication, or take 1 h before eating
Pain medications			
NSAID (e.g., ibuprofen, naproxen, piroxicam)	Gastric irritation may occur from direct GI contact or through systemic mechanisms (i.e., irritation may occur with suppositories or enteric coated products)	Severe gastric irritation may occur if taken on an empty stomach	Take with food to minimize gastric irritation
Parkinson's medications			
Anticholinergic medications (e.g., benztropine, procyclidine)	Medications inhibit breakdown of food because of decreased saliva production; food in general may cause xerostomia	May cause difficulty in swallowing food	Suck on sugarless hard candy, ice chips, or chew sugarless gum
Levodopa	Competition for absorption with protein	Minimize or nullify benefit of levodopa	Space medication and food 2 h apart or administer with nonprotein meal
Psychiatric medications			
Lithium	Sodium and lithium compete for reabsorption/elimination	Altering sodium intake may alter lithium level	Do not change diet once lithium is stabilized

the amount believed necessary. Causes of intentional or unintentional nonadherence include medication regimens that are difficult to fit into the patient's lifestyle, unwanted effects, expensive medications, or patient's inability to self-administer the product (33, 34). Most important, if patients do not believe they need the medication, they will be less likely to take the prescription.

Risks for nonadherence have been studied, and some of the findings include the complexity of dosing schedule (e.g., four times daily versus once daily dosing); frequent changes in medication; substitution of medication (therapeutic or generic); multiple medications; unpleasant side effects; difficulty in opening containers; cost; difficult routes of administration; inadequate patient education/understanding; and cognitive, visual, or physical function impairment (31, 33, 34). Patients who already have vision or mobility problems may become noncompliant when some of the factors noted above are added to their existing challenges with medication management.

Factors that have been linked to improving adherence include a belief that the medication is important, a belief that it is effective, and regularly attending the same medical clinic and pharmacy (34).

A number of studies have also tried to determine indicators of poor self-medication management. Some of these indicators are cognitive impairment (MMSE <24), physical dependency (Katz \geq 1), and poor self-reported medication management (35).

When prescribing medications for the elderly, it is important to consider physical disability, visual impairment, the shape or color of medications (ability to differentiate), and the belief system of the patient (35, 36). It would be beneficial to have healthcare professionals review the technique of using certain devices and to discuss the use of compliance aids.

Various compliance aids have been studied in an attempt to increase adherence to prescribed regimens (36–38). These aids include alarm clock devices, blister packing, calendar packaging, and more. Some providers feel they greatly assist patients who otherwise would become confused or could not manage multiple medications on their own, which would increase their risk of an adverse event or nonadherence. It has also been argued that these aids can be difficult to use, and the patients who benefit most are often those who would likely manage well without any aids. There are studies providing support for both views, and the subject requires additional study.

It is important to consider that even with compliance aids, the patients must receive appropriate counseling and instruction on how to manage the devices properly (37, 38). The aids must also be recommended and agreed to by the patient for them to be of use. For example, if a patient feels

they have lost some of their independence by having to use blister packaging, they are still not likely to become compliant. However, if the patient has agreed that an aid would be helpful and save them time, they are more likely to appreciate the additional assistance and will use the product to increase adherence. Patient preference, agreement, and education are still the most important factors in improving adherence. Additional factors that must be considered are the medication effects. If a patient discontinues a prescription because of side effects, packaging the medications into a compliance aid will not improve adherence. Before a compliance aid is considered, the reason for nonadherence must be carefully reviewed (39) because an aid may not be necessary at all and may offend patients if they think the provider feels they are becoming cognitively impaired.

Polypharmacy

When translated, polypharmacy simply means multiple medications. However, the clinical interpretation of this term usually refers to the use of multiple medications in an inappropriate, illogical, or harmful manner to the patient.

It is expected that the elderly have a higher incidence of chronic disease and will be prescribed more medications to manage their diseases. It is very easy to quickly reach a state of polypharmacy. Risk factors include multiple doctors, diseases, and pharmacies (40). Patients often present with a variety of complaints but may lack a diagnosis, triggering the prescription of medications that are used to manage the complaints that do not resolve the illness. Patients also use numerous nonprescription medications, which may go unmonitored and yet have the potential to interact with prescribed medicines. Patients seeing multiple doctors is often a concern because specialists may not take into consideration other prescribed medications before ordering new drugs for the patient. Similarly, some physicians may not be aware that their patients are seeing other doctors, causing difficulty in monitoring prescriptions in these of patients.

Although it is somewhat controversial, certain personality traits in elderly patients may predispose them to using more medications (36, 40). Elderly patients and their physicians may also be reluctant to discontinue medications. Often, patients with these views feel it is necessary to use medications and to obtain more prescriptions each time they visit a physician. This may also tie in with the attitudes of healthcare providers, who may be pressured to prescribe more medications for the patient. Finally, polypharmacy may be common in the elderly because they often borrow medication from family or friends, sometimes because of lack of education about the risks of inappropriate medication use.

Tablet Splitting

Tablet splitting is common in the elderly population (41) often because of cost-savings and the production of larger dosage strengths to meet the needs of younger patients. The prescriber may feel that the patient requires much less medication than a typical dose and consequently prescribes a dose that the patient must split. This is done frequently with medications that are adjusted throughout the course of treatment, such as warfarin for anti-coagulation. Tablet splitting is also common when titrating medications that cannot be started in full doses because of intolerance. A common example of this is the use of β -blockers post-MI when only small doses are tolerated at the onset, but when the goal is to reach a much higher dose.

This topic has been studied only in a small number of studies and often focuses on younger subjects (42). The process of tablet splitting is also not generalizable to the elderly population because the muscle strength, fatigue, and duration of time to split are often measured in young patients doing repetitive tablet-splitting tasks, which does not reflect the typical scenario in an elderly patient's home.

Underutilization/Undertreatment

Both undertreatment and underutilization of medications are common problems in the elderly. It is often a result of stereotyping by healthcare professionals, lack of education of managing medications in the elderly, and lack of research (4). Examples include treatment of hypertension, prescribing hormone therapy, osteoporosis, treating MI, lipid management, pain management, and depression.

DETERMINING MEDICATION APPROPRIATENESS IN THE ELDERLY

There have been many attempts to characterize and quantify the "appropriateness" of medications in the elderly. The first such scale was developed by Hanlon and is called the Medication Appropriateness Index (43). It has been used most frequently in the Veterans Affairs population in the United States.

Beers et al. (44) developed a list of medications that were considered inappropriate in a nursing home population. Their most recent list further describes appropriate doses of medications that should be used in the elderly (44). For example, digoxin is suggested to be given at 0.125 mg daily because most elderly patients either

become toxic or do not require higher doses for efficacy. Other lists have been developed (11) and updated (10). These lists are generally considered guidelines and should be used with professional judgment when prescribing and managing dosing in the elderly. Other tools have also been studied and are being developed. A review of different instruments has been written by Shelton et al. (45).

DOSAGE FORMS

Geriatric patients are able to use any type of dosage form available; none are contraindicated based on age alone. The most frequently used dosage form in the elderly is the oral route; however, other dosage forms may be used based on disease state. For example, chronic obstructive pulmonary disease is common in the elderly, and the route of choice for administration of medications is by inhalation. Therefore, this population frequently uses inhalers. Another common dosage form used in geriatric patients is eyedrops. Dosage forms requiring hand strength or coordination may be difficult to administer in some geriatric patients with certain conditions such as arthritis, weakness after a stroke, dementia, or other impairments. The challenges with each of these dosage forms are presented later.

Inhalers

Lung diseases such as COPD and asthma are common in the elderly population. The cornerstone of treatment of these conditions is to use medications administered by inhaler, using either a turbuhaler, nebulizer, metered-dose inhaler (MDI), diskhaler, or other breath-activated device. These devices often require good eyesight, hand-eye coordination, hand strength to depress the MDI, and intact cognition to manage the device. Age itself has been found to be a predictor of poor technique; however, other studies (46) found that other disease states that increase with age are more significant predictors of success in using MDIs. A patient with poor hand strength, such as a patient with rheumatoid or osteoarthritis, may have difficulty managing the inhalers. Patients who are blind or have decreased vision may also have difficulty organizing their inhalers in the proper order or manipulating any of the devices. Cognitive decline is also more common in elderly patients and tends to increase with age. This has been found to be a determinant of poor technique when using inhalers (46). It is important for healthcare professionals to assess the ability of patients to use inhalation devices before they are prescribed. Repeated counseling and education are

necessary to maintain good technique and appropriate use of these devices in the elderly population (38, 46, 47).

Eye Drops

Disease states frequently requiring medication such as eye drops are common in the elderly. Many elderly patients also use nonprescription natural tear replacement products for lubrication. It is therefore common to have elderly patients using eye drops, even though there are many challenges and difficulties in administering medication by this route. The majority of patients, when asked, admit to having difficulty administering eye drops (48). There is difficulty in raising arms above one's head, which can result from limited range of motion secondary to arthritis or deconditioning. It is also necessary to be in the supine position or to tilt one's head far back to keep the drops in the eye. This is often difficult for patients because of neck osteoarthritis or stiffness. It is also very challenging to aim the drops accurately. A number of devices have been studied, but none make compression of the drops easier (48). It is often difficult for patients to compress the bottle to release the drops, and some patients do not use correct technique and expel more than the required amount of drops. Some of the reasons for difficulty in administering drops are similar to those listed above for difficulty with actuating MDIs. Most of the devices appear to help improve aim only, which is just one part of the problem for elderly patients. Difficulty with eye drops commonly results in treatment failure as patients become nonadherent or administer drops incorrectly and receive less benefit than could be expected. It is important to review technique and to prescribe appropriate compliance aids to make administration easier for patients because often, there are no other options but to use eye drops for ophthalmic conditions.

CONCLUSION

Considering the diversity and challenges in dealing with the geriatric population, it is important to be aware of the physiological changes that occur with aging and the complications that can arise when these changes are not considered when dosing medications. It is also important to be aware of the potential problems that can arise in treating geriatric patients with multiple disease states because they are more likely to experience adverse events and to have more drug-related problems such as nonadherence. Although each patient should be managed individually, there are important principles that should be considered when working with geriatric patients to optimize the dosing and dosage forms used in this population.

REFERENCES

1. U.S. Census Bureau, U.S.A. Statistics in Brief—Population and Vital Statistics. <http://www.census.gov/statab/www/part1.html> (accessed July 26, 2000).
2. Ostrom, J.F.; Hammarlund, E.R.; Christensen, D.B. Medication Usage in an Elderly Population, *Med. Care*. 1985; 23, 157–164.
3. Lamy, P.P. Comparative Pharmacokinetic Changes and Drug Therapy in an Older Population, *J. Am. Geriatr. Soc.* 1982, 30 (11 Suppl.), S11–S19.
4. Hanlon, J.T.; Ruby, C.M.; Shelton, P.S.; Pulliam, C.C. Geriatrics. *Pharmacotherapy and Pathophysiologic Approach*, 4th Ed.; DiPiro, J.T., Talbert, R.L., Yee, G.C., Matzke, G.R., Wells, B.G., Posey, L.M., Eds.; Appleton & Lange: Stamford, CT, 1999; 52–61.
5. Montamat, S.C.; Cusack, B.J.; Vestal, R.E. Management of Drug Therapy in the Elderly. *New Engl. J. Med.* 1989, 321 (5), 303–309.
6. Parker, B.M.; Cusack, B.J.; Vestal, R.E. Pharmacokinetic Optimisation of Drug Therapy in Elderly Patients. *Drugs Aging* 1995, 7 (1), 10–18.
7. Grandison, M.K.; Boudinot, F.D. Age-Related Changes in Protein Binding of Drugs. *Clin. Pharmacokinet.* 2000, 38 (3), 271–290.
8. Cockcroft, D.W.; Gault, M.H. Prediction of Creatinine Clearance from Serum Creatinine. *Nephron* 1976, 16, 31–41.
9. Swift, C.G. Pharmacodynamics: Changes in Homeostatic Mechanisms, Receptor and Target Organ Sensitivity in the Elderly. *Br. Med. Bull.* 1990, 46 (1), 36–52.
10. Beers, M.H. Explicit Criteria for Determining Potentially Inappropriate Medication Use by the Elderly. *Arch. Intern. Med.* 1998, 157, 1531–1536.
11. McLeod, P.J.; Huang, A.R.; Tamblyn, R.M.; Gayton, D.C. Defining Inappropriate Practices in Prescribing for Elderly People: A National Consensus Panel. *Can. Med. Assoc. J.* 1997, 156 (3), 385–391.
12. Lamy, Peter P. Adverse Drug Effects. *Clin. Geriatr. Med.* 1990, 6 (2), 293–307.
13. Hanlon, J.T.; Schmader, K.E.; Koronkowski, M.J.; Weinberger, M.; Landsman, P.B.; Samsa, G.P.; Lewis, I.K. Adverse Drug Events in High Risk Older Outpatients. *J. Am. Geriatr. Soc.* 1997, 45, 945–948.
14. Atkin, P.A.; Veitch, P.C.; Veitch, E.M.; Ogle, S.J. The Epidemiology of Serious Adverse Drug Reactions Among the Elderly. *Drugs Aging* 1999, 14 (2), 141–152.
15. Lazarou, J.; Pomeranz, B.H.; Corey, P.N. Incidence of Adverse Drug Reactions in Hospitalized Patients, a Meta-Analysis of Prospective Studies. *J. Am. Med. Assoc.* 1998, 279, 1200–1205.
16. Veehof, L.J.G.; Stewart, R.E.; Meyboom-de Jong, B.; Haaijer-Ruskamp, F.M. Adverse Drug Reactions and Polypharmacy in the Elderly in General Practice. *Eur. J. Clin. Pharmacol.* 1999, 55, 533–536.
17. Beyth, R.J.; Shorr, R.I. Epidemiology of Adverse Drug Reactions in the Elderly by Drug Class. *Drugs Aging* 1999, 14 (3), 231–239.
18. Gerety, M.B.; Cornell, J.E.; Plichta, D.T.; Eimer, M. Adverse Events Related to Drugs and Drug Withdrawal in Nursing Home Residents. *J. Am. Geriatr. Soc.* 1993, 41, 1326–1332.

19. Graves, T.; Hanlon, J.T.; Schmader, K.E.; Landsman, P.B.; Samsa, G.P.; Pieper, C.F.; Weinberger, M. Adverse Events after Discontinuing Medications in Elderly Outpatients. *Arch. Intern. Med.* **1997**, *157*, 2205–2210.
20. Cumming, R.G. Epidemiology of Medication-Related Falls and Fractures in the Elderly. *Drugs Aging* **1998**, *12* (1), 43–53.
21. Monane, M.; Avorn, J. Medications and Falls. *Clin. Geriatr. Med.* **1996**, *12* (4), 847–858.
22. King, M.B.; Tinetti, M.E. Falls in Community-Dwelling Older Persons. *J. Am. Geriatr. Soc.* **1995**, *43*, 1146–1154.
23. Leipzig, R.M.; Cumming, R.G.; Tinetti, M.E. Drugs and Falls in Older People—A Systematic Review and Meta-Analysis, I. Psychotropic Drugs. *J. Am. Geriatr. Soc.* **1999**, *47*, 30–39.
24. Leipzig, R.M.; Cumming, R.G.; Tinetti, M.E. Drugs and Falls in Older People—A Systematic Review and Meta-Analysis, II Cardiac and Analgesic Drugs. *J. Am. Geriatr. Soc.* **1999**, *47*, 40–59.
25. Inouye, S.K. Delirium in Hospitalized Older Patients. *Clin. Geriatr. Med.* **1998**, *14* (4), 745–764.
26. Flacker, J.M.; Marcantonio, E.R. Delirium in the Elderly. *Drugs Aging* **1998**, *13* (2), 119–130.
27. Moore, A.R.; O-Keeffe, S.T. Drug-Induced Cognitive Impairment in the Elderly. *Drugs Aging* **1999**, *15* (1), 15–28.
28. Seymour, R.M.; Routledge, P.A. Important Drug–Drug Interactions in the Elderly. *Drugs Aging* **1998**, *12* (6), 485–494.
29. Doucet, J.; Chassagne, P.; Trivalle, C.; Landrin, I.; Pauty, M.D.; Kadri, N.; Menard, J.F.; Bercoff, E. Drug–Drug Interactions Related to Hospital Admissions in Older Adults—A Prospective Study of 1000 Patients. *J. Am. Geriatr. Soc.* **1996**, *44*, 944–948.
30. Thomas, J.A.; Burns, R.A. Important Drug–Nutrient Interactions in the Elderly. *Drugs Aging* **1998**, *13* (3), 199–209.
31. Fuhr, U. Drug Interactions with Grapefruit Juice Extent, Probable Mechanism and Clinical Relevance. *Drug Safety* **1998**, *18* (4), 251–272.
32. Morrow, D.; Leirer, V.; Sheikh, J. Adherence and Medication Instructions Review and Recommendations. *J. Am. Geriatr. Soc.* **1988**, *36*, 1147–1160.
33. Conn, V.; Taylor, S.G.; Stineman, A. Medication Management by Recently Hospitalized Older Adults. *J. Community Health Nurs.* **1992**, *9* (1), 1–11.
34. Stewart, S.; Pearson, S. Uncovering a Multitude of Sins: Medication Management in the Home Post Acute Hospitalisation Among the Chronically Ill. *Aust. NZ J. Med.* **1999**, *29* (2), 220–227.
35. Ruscin, J.M.; Semla, T.P. Assessment of Medication Management Skills in Older Outpatients. *Ann. Pharmacother.* **1996**, *30*, 1083–1088.
36. Thwaites, J.H. Practical Aspects of Drug Treatment in Elderly Patients with Mobility Problems. *Drugs Aging* **1999**, *14* (2), 105–114.
37. Cramer, J.A. Optimizing Long-Term Patient Compliance. *Neurology* **1995**, *45* (Suppl 1), S25–S28.
38. Sexton, J.; Gokani, R. Pharmaceutical Packaging and the Elderly. *Pharm. J.* **1997**, *259*, 697–700.
39. Weintraub, M. Compliance in the Elderly. *Clin. Geriatr. Med.* **1990**, *6* (2), 445–452.
40. Stewart, R.B.; Cooper, J.W. Polypharmacy in the Aged Practical Solutions. *Drugs Aging* **1994**, *4* (6), 449–461.
41. Rochon, P.A.; Clark, J.P.; Gurwitz, J.H. Challenges of Prescribing Low-Dose Drug Therapy for Older People. *Can. Med. Assoc. J.* **1999**, *160* (7), 1029–1031.
42. McDevitt, J.T.; Gurst, A.H.; Chen, Y. Accuracy of Tablet Splitting. *Pharmacotherapy* **1998**, *18* (1), 193–197.
43. Hanlon, J.T.; Schmader, K.E.; Samsa, G.P.; Weinberger, M.; Uttech, K.M.; Lewis, I.K.; Cohen, H.J.; Feussner, J.R. A Method for Assessing Drug Therapy Appropriateness. *J. Clin. Epidemiol.* **1992**, *45* (10), 1045–1051.
44. Beers, M.H.; Ouslander, J.G.; Rollinger, I.; Reuben, D.B.; Brooks, J.; Beck, J.C. Explicit Criteria for Determining Inappropriate Medication Use in Nursing Home Residents. *Arch. Intern. Med.* **1991**, *151*, 1825–1832.
45. Shelton, P.S.; Fritsch, M.; Scott, M.A. Assessing Medication Appropriateness in the Elderly: A Review of Available Measures. *Drugs Aging* **2000**, *16* (6), 437–450.
46. Gray, S.L.; Williams, D.M.; Pulliam, C.C.; Sirgo, M.A.; Bishop, A.L.; Donohue, J.F. Characteristics Predicting Incorrect Metered-Dose Inhaler Technique in Older Subjects. *Arch. Intern. Med.* **1996**, *156*, 984–988.
47. Daniels, S.; Meuleman, J. Importance of Assessment of Metered-Dose Inhaler Technique in the Elderly. *J. Am. Geriatr. Soc.* **1994**, *42*, 82–84.
48. Winfield, A.J.; Jessiman, D.; Williams, A.; Esakowitz, L. A Study of the Causes of Non-Compliance by Patients Prescribed Eyedrops. *Br. J. Ophthalmol.* **1990**, *74* (8), 477–480.
49. Kane, R.L.; Ouslander, J.G.; Abrass, I.B. *Essentials of Clinical Geriatrics*, 4th Ed.; McGraw-Hill: New York, 1999.
50. Swonger, A.K.; Burbank, P.M. *Drug Therapy and the Elderly*; Jones and Bartlett Publishers: Boston, 1995.
51. Luisi, A.F.; Owens, N.J.; Hume, A.L. Drugs and the Elderly. *Reichel's Care of the Elderly Clinical Aspects of Aging*, 5th Ed.; Gallo, J.J., Busby-Whitehead, J., Rabins, P.V., Silliman, R.A., Murphy, J.B., Reichel, W., Eds.; Lippincott Williams & Wilkins: Baltimore, 1999; 59–87, Chap. 5.
52. Beyth, R.J.; Shorr, R.I. Medication Use. *Practice of Geriatrics*, 3rd Ed.; Duthie, E.H., Jr., Katz, P.R., Eds.; W.B. Saunders Company: Philadelphia, 1998; 38–47, Chap. 5.
53. Avorn, J.; Gurwitz, J.H. Principles of Pharmacology. *Geriatric Medicine*, 3rd Ed.; Cassel, C.K., Cohen, H.J., Larson, E.B., Meier, D.E., Resnick, N.M., Rubenstein, L.Z., Sorenson, L.B., Eds.; Springer: New York, 1997, 755–770, Chap. 5.
54. Schwartz, J.B. Clinical Pharmacology. *Principles of Geriatric Medicine and Gerontology*, 4th Ed.; Hazzard, W.R., Blass, J.P., Ettinger, W.H., Jr., Halter, J.B., Ouslander, J.G., Eds.; McGraw-Hill: New York, 1999; 303–331, Chap. 5.
55. Abrams, W.B.; Beers, M.H.; Berkow, R. *The Merck Manual of Geriatrics*, 3rd Ed.; Merck & Co. Inc.: Whitehouse Station: New Jersey, 2000.
56. Hansten, P.D.; Horn, J.R.; Koda-Kimble, M.A.; Young, L.Y. *Hansten & Horn's Managing Clinically Important Drug Interactions*; Applied Therapeutics Inc.: Vancouver, Washington, 1998.
57. Delafuente, J.C.; Stewart, R. *Therapeutics in the Elderly*, 3rd Ed.; Williams & Wilkins: Baltimore, 2000.
58. Kinirons, M.T.; Crome, P. Clinical Pharmacokinetic Considerations in the Elderly—An Update. *Clin. Pharmacokinet.* **1997**, *33* (4), 302–312.



Genetic Aspects of Drug Development

Werner Kalow

University of Toronto, Toronto, Ontario, Canada

INTRODUCTION

The science of genetics is currently changing in a way that we cannot neglect when looking at drug development: Traditional genetics was solely concerned with structural variation of genes. New techniques led to the recognition that the expression of genes can vary.

Traditional medicine was not much concerned with genetics, except for genetic diseases. However, genetic diseases are always rare because disease-causing gene alterations render them evolutionary disadvantageous; hence such alterations tend to be eliminated by Darwinian selection. However, it is now realized that most or all common diseases are contributed to by genetic factors. It is thus predictable that genetic considerations will become increasingly important in medicine.

Most drugs on the market nowadays were developed without much concern for genetics, or perhaps with restricted concerns, although genetic factors are well known to be able to alter an individual's response to a drug; a person may show a lack of response or an over-response that in some cases has been fatal. Such gross differences are hoped to be eliminated in the future or at least curtailed because the new genomic techniques may increase the knowledge of genes that can affect both drug safety and efficacy. The present choice of a drug for treating any disorder is not based on causes but on diagnostics and statistics. This is hoped to be changed to become "personalized medicine," which means choosing the right drug for a given patient on the basis of that patient's genetic makeup. While many genes can contribute to all common diseases, but only one or a few are likely to be the main contributors to the disease in that patient. In other words, the hope is to choose a drug for treatment that affects the impact of the particular gene, which causes or enhances the disease in that person. The common diseases contributed to by multiple genes include cardiovascular, asthmatic, and schizophrenic disorders, among many others. The rise of genomic techniques will convert single-gene pharmacogenetics into multiple-gene pharmacogenomics.

In order to realize the hope, a number of tasks will have to be performed. A primary task will be the identi-

cation of the genes whose malfunction contributes to, or can cause, the disease to be treated. Once the gene of interest is identified by genomic methods, it will be necessary to define its function; this will usually require knowledge of the protein produced by the gene via translation. By establishing the normal gene function, recognition of the malfunction is facilitated and its contribution to the disease may become clear. It then becomes possible to look for chemicals that can attach to the gene or to the protein, thereby eliminating or relieving the faulty gene function. Finally, the chemical may be converted to a drug.

All these processes require scientific views. The purpose of this article is to provide an overview at the scientific backgrounds for this study and for the required procedures.

BACKGROUND DATA

Variability of Gene Function

Genes determine the appearance, actions, and much of the fate of any individual, be it a bacterium, a plant, or a mammal. Our genes make any human different from anyone else. This means that the genes cannot be similar in different individuals. In fact, most subject-to-subject differences depend, and evolution depended, on variability of genes, that is, on gene mutations.

Most mutations represent the replacement of one base by another; a minority is caused by deletions of some base pairs, by frameshift, by polypeptide chain elongation due to mutation of a terminator codon, or by recombinational events with mutation-like effects. The replacement of one base for another is currently indicated by the term "single-nucleotide polymorphism," abbreviated as "SNP."^[1,2] It occurs at a frequency of roughly one per every 500–1000 base pairs of genomic DNA, which means that most genes carry one or two SNPs, although many are without any functional significance.^[3]

Techniques developed in recent years have opened a totally new aspect of genetics, that is, the study of gene expression.^[4–6] That is, the DNA of a given gene may lead to the production of varying numbers of its derivative RNA, the protein-forming nucleic acid. As a consequence,



the function of a gene may vary, independent of gene structure.

Gene expression is controlled by transcription.^[7] Transcription is the formation of one-stranded RNA from DNA, the gene consisting of a two-stranded nucleotide coil. The process involves unwinding of DNA. It is catalyzed by RNA polymerases, enzymes controlled by complicated factors.^[8] Posttranscriptional processing of the messenger RNA (mRNA) involves the removal of introns by splicing. The formation of protein from RNA by translation is again a complex process.^[9] Translation can be inhibited by antisense oligonucleotides.^[10,11] Thus, gene expression involves complex processes. It can be affected by hormones, diseases, and (importantly) by drugs.

Furthermore, gene expression can be controlled by epigenetics, a potentially reversible change in DNA methylation.^[12,13] Its normal functions are regulation of gene activity during aging, tissue-specific DNA modification, and X-chromosome inactivation. Epigenetics may contribute to cancer control.^[14] In short, a look at gene structure and mutations may no longer be good enough to alone serve the hunter for personalized medicine.

Interactions of Genes and Drugs

Pharmacogenetics started with the discovery of variable responses of different people to a given drug, as summarized in papers by Motulsky^[15] and Vogel,^[16] and in the book by Kalow.^[17] Before genes could be tested, allelic variations were phenotypically indicated as a cause by the occurrence of abnormal drug responses within families or groups of twins. Also, response abnormalities in different human populations were known early, at first only assuming a genetic cause of such population differences.

All these early reports of inborn differences dealt with distinct abnormalities in drug response that were monogenic (Mendelian) in character. It is a different matter when we observe that only a few people respond exactly alike to a given drug. This observation was mathematically formalized by the introduction of the concept of "ED50," meaning the dose of a drug that had the desired effect in 50% of population.^[18] The utility of the ED50 indicates that an unequal response of individual animals or humans to any drug is a rule and is often very large. This variability can be graphically represented by a normal (Gaussian) distribution curve with ED50 as the mean. The first evidence that genetic factors can also be involved came from systematic studies of twins, which indicated genetic control of various drug metabolizing capacities.^[19]

Thus, the genetics of a drug response can be of two kinds, which deserve a clear distinction: First, the particular genetic variant of a particular gene may strongly affect the

response to one drug, or to a series of drugs. Second, there can be multifactorial variation of a drug response caused by effects or influences of many genes, usually more or less modified by environmental factors. The examples of these two kinds of variation in order to clarify their impact are given below.

The first case is well illustrated by the studies of CYP2D6, a drug-metabolizing P450 cytochrome that is prominent in human liver. This protein contributes to the metabolism of about 60 different drugs.^[20] So far, 74 alleles have been found in that gene, most often several mutations (up to 8) per allele. The difference in the consequences of CYP2D6 mutations deserves attention: A few of the mutations prohibit the formation of CYP2D6, others totally eliminate its biological activity, some of them merely decrease its activity,^[21] others change its activity toward selected drug substrates,^[22,23] and gene duplication may cause greatly increased activity in some subjects.^[24]

The absence of CYP2D6 activity occurs in about 7% of Caucasians, in less than 1% of Asians and most Africans,^[25] although it is absent in about 20% of San Bushmen.^[26] A mutation identified in about 50% of Chinese subjects causes a substantially decreased average CYP2D6 activity in China compared to Europe.^[27] Of seven precisely investigated CYP2D6 mutations, a few were found in Europe, others in China and Africa, but only one (G4268C) was found everywhere (in Europe, China, Japan, and Africa), suggesting that this is an old mutation that occurred prior to the migration of *Homo sapiens* from Africa.^[25] Thus, differences in the timing of mutational events may account for most interethnic differences of drug-metabolizing enzymes because such mutations tend to be biologically irrelevant in the absence of drugs.

These data show that there can be great complexity even if we deal with a single variable gene as in traditional pharmacogenetics. It is a relatively simple process to count the absence of enzyme activity ("the poor metabolizer phenotype") in one population, but data thus obtained, for instance, in Europe, may be useless everywhere else.

Multifactorial control of drug responses cannot be described in similarly clear terms, but some examples can be given. For instance, one may look at the two groups of homozygous subjects with respect to debrisoquine's metabolism, i.e., there are those homozygous for the wild-type CYP2D6 and those without the enzyme. In each group, there is a large interindividual variability of debrisoquine's persistence. This can be accounted for only by the presence and variability of numerous factors other than the CYP2D6 genotypes contributing to the drug's elimination.^[25] Multifactorial control is present whenever response curves show a Gaussian distribution. Vesell^[28]



cited examples to show that the metabolism of various drugs was mainly under genetic control, even when the metabolism was not due to an identifiable single gene. Table 1 indicates the potential complexity of drug metabolizing control even if there is only one identified catalyst.

Because of old and available methodologies, drug metabolism and other forms of drug elimination are particularly well-investigated fields. However, the variability of drug effects can also depend very much on genetic variation of drug targets, such as drug receptors.^[29] Kim and Wilkinson^[30] described the variation of drug transporters and the effects.

Drugs can also influence gene expression. A good recent example is the determination of changes produced by rifampin. Rae et al.^[31] used microassays to measure the effects of that drug on the overall pattern of mRNA expression of drug-metabolizing enzymes in human hepatocytes. There was no effect on CYP2E1 or CYP2D6. The drug caused a 6.5-fold increase in the expression of CYP2C8 and a 3.7-fold increase of CYP2C9, but had no effect on CYP2C18 mRNA. CYP3A4 showed the highest level of 55.1-fold induction.

It is now recognized that drugs can affect the activity of many genes by enhancing or inhibiting their expression. In fact, we often might be unsure whether a drug effect represents an interaction with a protein or with its gene. For instance, it seems that the addictive effect of nicotine or other drugs results from altered gene expression produced by these drugs.^[32]

It has to be mentioned that nongenetic or environmental influences may also have great impact. Starvation, malnutrition, and protein deficiency can all cause a decreased rate of drug disposition,^[33] but so can even

relatively minor food deficiencies.^[34] Some foods or drugs cause enzyme induction^[35] and others cause enzyme inhibition.^[36] Infections, other diseases, and many factors that can affect hormonal functions may alter a drug response. It is well established that cigarette smoking causes induction of some drug-metabolizing enzymes. However, the extent of smoking varies between individuals and between populations, in part because of genetic variability of the nicotine-metabolizing cytochrome CYP2A6;^[37] this is an example showing that genetic predisposition can affect the magnitude of an environmental impact.^[38] Many drug responses are influenced by gender and age; both factors are controlled by genes but also by time.

Regardless of all these complexities, the fact that drug action always depends more or less on genes is important, although the control may be exercised by one or by a multiplicity of genes.

Genes and Diseases

The term "genetic diseases" refers to diseases that are essentially caused by faulty function of a single gene. Although they are Mendelian disorders, their phenotypic expression can vary from case to case.^[39,40] Examples are, for instance, cystic fibrosis, neurofibromatosis, Marfan syndrome, and Huntington disease. All these diseases are rare because the producing genetic variants are disadvantageous from an evolutionary point of view. Because of the rarity, there are few broad-based efforts to find drugs for their treatment. These limitations also express societal and economic concerns.

All common diseases are caused, or contributed to, by many genes, often interacting with environmental factors.^[41] Examples are, for instance, arthritis, atherosclerosis, asthma, schizophrenia, and Alzheimer's disease. Many cases are cited in the book *The Genetic Basis of Common Diseases*.^[42] Loring et al.^[43] found 118 differently expressed genes in Alzheimer brains. Gharavi and Lifton^[44] describe the hereditary basis of blood pressure variation and hypertension: Identified to date are mutations in seven genes that raise blood pressure and in nine genes that lower blood pressure in humans. Many of these genes control renal salt handling. Some cases suggest the presence of angiotensinogen variation, or changes in beta-adrenergic or dopamine receptor system. There are also Mendelian (monogenically controlled) cases of human hypertension, indicating that the disease is not necessarily controlled by many genes. As indicated in the four-volume book by Scriver et al.,^[45] many diseases show similar complications.

It is a principal difficulty that medical diagnosis is currently almost always dictated by the appearance of any

Table 1 Factors that may cause reduced metabolism of a drug

1. Alterations of the enzyme's gene
 - Structural variation of the gene through mutation
 - Reduced expression of the gene (caused by gene–gene interaction, effects on the gene by drugs, hormones, or foods), by epigenetics, or by unstable RNA
 - Reduced function: failing transcription or translation (protein formation)
2. Alterations of protein concentration or function
 - Enzyme destruction or instability reduces its concentration
 - Enzyme inhibition by drugs, hormones, foods, and infections
 - Mistaken protein folding
3. Reduced access of drug to enzyme
 - Excessive drug binding (e.g., in tissues or to albumin)
 - Genetic variation of a transporter protein

A Question: Which of these factors cause the typical age-related reduction of enzyme activity?



disease, perhaps by a single prominent symptom, but usually does not reveal the cause of the disease. This is a shortcoming that is hoped to be overcome by defining the genes which are altered in a case of common disease. However, knowing the crucial genes will allow the investigation of their function, and knowing the function may reveal the contribution of the gene's failure or its action to the disease. The final question is which gene contributes to the disease of which patient.

Disease Genetics Determines Drug Action

Multiple treatments are available for most of the common diseases, but they often fail, and a given patient may benefit from only one or the other of the recommended drugs. In some cases, genetic factors affecting the disease are known to determine whether a drug is beneficial or useless for that patient. Let us consider a few examples given below.

Data on asthma are a good example of the complexities of interacting genes, diseases, and drugs. First, there is no universally accepted definition of asthma. Twin studies suggest that asthma is to 0.72%–0.8% a genetic disease,^[46] but inheritance does not follow a clear Mendelian pattern, and environmental factors modulate the clinical expression.^[47] Thus, asthma is multifactorial. Disease expression is influenced by interactions between multiple major and minor genes, and modulated by environmental factors.^[48]

Drazen et al.^[49] emphasize the heterogeneity of therapeutic responses in asthma. Current asthma therapy employs several different drugs aimed at different targets. Inhaled glucocorticoids like beclomethazone are used as anti-inflammatory drugs, acting on glucocorticoid receptor alpha.^[50] β_2 -Adrenoceptor agonists like salbutamol protect against histamine-induced bronchoconstriction.^[51] Cysteinyl-leukotriene inhibitors, i.e., ACE-inhibitors like fosinopril, counteract slow-reacting substances of anaphylaxis.^[52] Among additional factors, driving and maintaining the asthmatic inflammatory process appear to be an abnormal or inadequately regulated CD4+ T-cell immune response.^[53] The variable platelet-activating factor (PAF) acetylhydrolase and oxidized derivatives of phosphatidylcholines play important roles as causative factors in asthma and other diseases.^[54] The gene NOS1 regulating neuronal nitric oxide is associated with allergic asthma, as shown in studies of mice and humans.^[55]

Other broad-based examples besides that of asthma could surely be found. However, there are other established though more restricted examples. Well known is the failure of treatment of Alzheimer's disease by tacrin if the patient has a variant of ApoE4, an apolipoprotein. Tacrin

is a cholinesterase inhibitor, one of the first anti-Alzheimer drugs,^[56] chosen to elevate the brain levels of the transmitter acetylcholine, a transmitter that was low because of deficient formation by the enzyme choline acetyltransferase.^[57] The contribution of a genetic variant of ApoE4 to Alzheimer's disease was an independent later observation.^[58] Then, the discoverer of this contribution noticed that patients with the ApoE4 variant tended to respond poorly to tacrin.^[59]

Another example is the genetic variability of the effects of neuroleptic drugs in schizophrenia. Cichon et al.^[60] have shown that this variable responsiveness can be caused by genetic variation of several of the P450 cytochromes, plus variation of dopamine and serotonin receptors. Because variation of these receptors seems to contribute to the schizophrenic disease, this is another example of interaction of genetic effects, diseases, and drug response.^[61–63]

All these observations support confidence in research aimed at the development of personalized medicine, that is, at the development of treatments that affect the function of genes which contribute most to a patient's disease.

THE TASKS

Searching for Genes of Interest

Foremost will be the search for mutations that are responsible for, or substantially contribute to, the occurrence of the common disease of interest. Association studies^[64] will provide the link between the disease and a genetic variant. Of the millions of DNA sequence variants in the human genome, the majority is single-nucleotide polymorphism (SNP). Determination of SNP-based genetic profiles may be viewed as individual fingerprints, indicating relative genetic contribution of the risk for various illnesses. This is the driving force behind intense efforts to establish the technology for large-scale analysis of SNPs.

Many technological descriptions of gene searching were published recently. For instance, Kwok^[65] summarized a number of genotyping methods currently in use. All effective designs require genotyping of a large number of individuals with a large number of markers. Therefore, the whole genome must be searched for the chance to find responsible genes. There is multiplex genotyping^[66] and minisequencing.^[67] Medlin^[68] used "inkjet technology" to propel DNA nucleotides onto slides. Qi et al.^[69] described a flexible, non-gel-based detection method, using thermostable ligation for allele ligation and rolling circle amplification (RCA) for signal enhancement. Tsongalis et al.^[70] described a system utilizing DNA polymerase-

mediated pyrophosphorolysis and a luciferase detection reaction. Patil et al.^[71] set out to identify all SNPs in human chromosome 21 and to group them into blocks of haplotypes. They established human–rodent hybrid cell lines, each containing one copy of human chromosome 21 from a different individual. Using high-density oligonucleotide arrays, they obtained complete DNA sequences in that human chromosome. Further methods will be developed and described. Nevertheless, the search for genes underlying complex traits has been difficult and often disappointing.^[72] The main reason for these difficulties is that several genes, each with rather small effect, might be interacting to produce the trait.

Probably just as important as the search for structural variants of genes will be the exploration of gene expression data. Gene expression can be compared between cells derived from particular tissues, or obtained from the same tissue at different times. If the tissue was disease free on first but not on second sampling, and if the expression of some genes were found to differ between the samples, one would know which genes are altered by the disease, or which may be its cause. By the same method, one could learn which genes are affected by a drug or toxin, or which change with age, or which are expressed in one but not in another tissue. One would know the gene with which to deal, and one would start to understand the biological role of a gene. It could mean identification and validation of new molecular targets for drug development, and prediction of potential side effects during preclinical development and toxicology studies.^[73,74]

Gene expression studies require microarray technology. Recently, high throughput technologies for biochemical analysis of gene expression have come into use.^[75–78] Serial analysis of gene expression (SAGE) is a widely used method.^[6] Burgess^[79] spotted different probes on a glass microscope slide and labeled the RNA isolated from the cells with two different fluorochromes before being hybridized to the microarray. Gene expression as whole profiles can replace the conventional focus on an individual gene.^[80]

Conceptually different approaches to the development of this technology resulted in the generation of two different array formats: oligonucleotide arrays and cDNA arrays.^[81] Oligonucleotide probes represent mispaired bases in DNA segments. Lee et al.^[82] stressed the importance of studying replications to insure data reliability, while Herwig et al.^[83] presented a statistical evaluation of expression data. Brazma et al.^[84] proposed a standard for presenting and exchanging such data in form of “minimum information about a microarray experiment” (MIAME). Haverty et al.^[85] established the “human gene expression index” (www.HugeIndex.org) to serve as a public

repository for gene expression data on normal human tissues. In short, investigations of gene expression represent a major new scientific interest.

Gene Function

In biological perspective, identification of an important gene becomes truly significant only after its function is recognized. This is well acknowledged; in answer to the entry “gene function,” PubMed cited over 60,000 entries on the Internet.

Because the human and mouse genomes are 80% identical, while many single genes are completely identical, studies of mice are widely used.^[43] The use of mice has also the advantage that the function of a gene can be demonstrated by in vitro or in vivo knockout, i.e., experimental gene elimination. Yoshihara et al.^[86] found a significant number of fly homologs of human neurological disease loci; they suggested *Drosophila* as an important disease model for human neuropathology.

There is much debate over the utility of multiple locus association analyses in the identification of genomic regions harboring sequence variants that influence complex traits.^[87] There are new methods of testing the statistical association between haplotypes and a wide variety of traits.^[88] The identification of thousands of genes generated by the various genome projects is a challenge.^[89]

A crucial aspect of studying gene function is identification of the protein formed by the gene in question; determination of protein function is a traditional aspect of biology. The science of protein identification and their systematic, large-scale analysis is “proteomics.” Protein profiles can be correlated with specific effects of disease and with alterations of function.^[90–92] The importance of the subject is well illustrated by the recent creation of a journal entitled “Proteomics.”

The biological utilization of the many types and amounts of information gathered by different methods is aided by the complex science of “bioinformatics,” i.e., the use of computers for the storage or analysis of biological information. Also there is a journal by that name. There is a recent catalog of the computer-based resources that provide access to many of the collections of information on bioinformatics,^[93] and lists of key words and related items.^[94]

Drug Design

Once the functional genomic studies have revealed all pertinent information, i.e., the gene of interest, its biological role, and the dependent protein, one can start to think of the chemical structures that a drug must have to

affect gene function.^[26] There are different methods of virtual screening as discovery tools for lead structure.^[95]

A good example is the development of novel inhibitors of acetylcholinesterase, using a structure-based design approach.^[96] Starting with chimeras of tacrine, the authors conducted a molecular modeling study using the x-ray structure of the apoenzyme; they determined several favored attachment positions and other characteristics. They synthesized several matching compounds and determined their inhibitory potencies. They were found to be in the low nanomolar range.

If the three-dimensional structure of a given protein is known, this information can be directly exploited for the retrieval and design of new ligands. Nuclear magnetic resonance spectroscopy and x-ray crystallography are useful for molecular modeling.^[97] The knowledge of the crystal structure of the target protein or of a model derived from it is very helpful. This information can be used to stimulate thinking of improving an existing ligand, or of developing new alternative bonding skeletons. The features of the protein binding pocket can be translated into queries used for virtual computer screening of large compound libraries or to design novel ligands de novo.^[98] The need for structural information about drug-protein complexes is fundamental for drug discovery.^[98] A leads optimization procedure is described by Williams et al.^[99] The current state of design methodology for inhibitors specifically targeting protein surfaces and computational methods has been reviewed by Zeng.^[100]

Perspectives of the requirements for the development of leads and drugs have been summarized in an overview paper.^[101] The chemical features should be amenable to optimization, the patient situation should be favorable, absorption and other distribution properties should be good. A set of measurable items can help to appraise such determinants.

CONCLUSIONS

There are large-scale efforts in all scientifically advanced countries to promote personalized medicine. This new medicine would leave behind the traditional ways of choosing drugs for treatment, ways that were based on population averages and visual disease diagnostics. Genetic diagnostics would become the fulcrum for clinical decisions. The tasks needed to achieve these changes are enormous, and substantial time for achievement will be required; yet, current scientific knowledge tells us that the goal makes sense, and that major searching efforts are worthwhile. Medical and public perceptions of the hoped-

for changes are problems that need to be addressed in an open discussion.

REFERENCES

1. Shi, M.M. Enabling Large-Scale Pharmacogenetic Studies by High-Throughput Mutation Detection and Genotyping Technologies. *Clin. Chem.* **2001**, *47*, 164–172.
2. Grant, D.M.; Phillips, M.S. Technologies for the Analysis of Single-Nucleotide Polymorphisms: An Overview. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 183–190.
3. Chasman, D.; Adams, R.M. Predicting the Functional Consequences of Non-synonymous Single Nucleotide Polymorphisms: Structure-Based Assessment of Amino Acid Variation. *J. Mol. Biol.* **2001**, *23*, 683–706.
4. Kaufman, R.J. Overview of Vector Design for Mammalian Gene Expression. *Mol. Biotechnol.* **2000**, *16*, 151–160.
5. King, H.C.; Sinha, A.A. Gene Expression Profile Analysis by DNA Microarrays: Promise and Pitfalls. *J. Am. Med. Assoc.* **2001**, *14*, 2280–2288.
6. Madden, S.L.; Wang, C.; Landes, G. Serial Analysis of Gene Expression: Transcriptional Insights into Functional Biology. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 223–252.
7. Rudner, D.Z.; Losick, R. Morphological Coupling in Development: Lessons from Prokaryotes. *Dev. Cell* **2001**, *1*, 733–742.
8. Oelgeschlager, T. Regulation of RNA Polymerase II Activity by CTD Phosphorylation and Cell Cycle Control. *J. Cell Physiol.* **2002**, *190*, 160–169.
9. Moore, M.J.; Rosbash, M. Cell Biology. Tapping into mRNA Export. *Science* **2001**, *30*, 1841–1842.
10. Myers, K.J.; Dean, N.M. Sensible Use of Antisense: How to Use Oligonucleotides as Research Tools. *Trends Pharmacol. Sci.* **2000**, *21*, 19–23.
11. Cooper, S.R.; Taylor, J.K.; Miraglia, L.J.; Dean, N.M. Pharmacology of Antisense Oligonucleotide Inhibitors of Protein Expression. *Pharmacol. Ther.* **1999**, *82*, 427–435.
12. Wolffe, A.P.; Matzke, M.A. Epigenetics: Regulation Through Repression. *Science* **1999**, *15*, 481–486.
13. Nakao, M. Epigenetics: Interaction of DNA Methylation and Chromatin. *Gene* **2001**, *31*, 25–31.
14. Muylers-Chen, I.; Paro, R. Epigenetics: Unforeseen Regulators in Cancer. *Biochim. Biophys. Acta* **2001**, *30*, 15–26.
15. Motulsky, A.G. Pharmacogenetics and Ecogenetics in 1991. *Pharmacogenetics* **1991**, *1*, 2–3.
16. Vogel, F. Moderne Probleme der Humangenetik. *Ergebnisse der Inneren Medizin und Kinderheilkunde* **1959**, *12*, 65–126.



17. Kalow, W. *Pharmacogenetics. Heredity and the Response to Drugs*, W.B. Saunders Company: Philadelphia, 1962.
18. Trevan, J.W. The Error of Determination of Toxicity. *R. Soc. Lond., Proc. 1 Ser. B* **1927**, *101*, 483–514.
19. Vesell, E.S.; Page, J.G. Genetic Control of Drug Levels in Man: Antipyrine. *Science* **1968**, *161*, 72–73.
20. Kalow, W.; Grant, D.M. Pharmacogenetics. In *The Metabolic & Molecular Bases of Inherited Disease*, Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D., Childs, B., Kinzler, K.W., Vogelstein, B., Eds.; McGraw-Hill: New York, 2001; 225–255.
21. Fukuda, T.; Nishida, Y.; Imaoka, S.; Hiroi, T.; Naohara, M.; Funae, Y.; Azuma, J. The Decreased In Vivo Clearance of CYP2D6 Substrates by CYP2D6*10 Might Be Caused Not Only by the Low-Expression but Also by Low Affinity of CYP2D6. *Arch. Biochem. Biophys.* **2000**, *15*, 303–308.
22. Woolhouse, N.M.; Eichelbaum, M.; Oates, N.S.; Idle, J.R.; Smith, R.L. Dissociation of Co-regulatory Control of Debrisoquine/Phenformin and Sparteine Oxidation in Ghanaians. *Clin. Pharmacol. Ther.* **1985**, *37*, 512–521.
23. Masimirembwa, C.; Hasler, J.; Bertilssons, L.; Johansson, I.; Ekberg, O.; Ingelman-Sundberg, M. Phenotype and Genotype Analysis of Debrisoquine Hydroxylase (CYP2D6) in a Black Zimbabwean Population. Reduced Enzyme Activity and Evaluation of Metabolic Correlation of CYP2D6 Probe Drugs. *Eur. J. Clin. Pharmacol.* **1996**, *51*, 117–122.
24. Johansson, I.; Lundqvist, E.; Bertilsson, L.; Dahl, M.L.; Sjoqvist, F.; Ingelman-Sundberg, M. Inherited Amplification of an Active Gene in the Cytochrome P450 CYP2D Locus as a Cause of Ultrarapid Metabolism of Debrisoquine. *Proc. Natl Acad. Sci.* **1993**, *90*, 11825–11829.
25. Kalow, W. Interethnic Differences in Drug Response. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 109–134.
26. Winkler, F.K.; Banner, D.W.; Bohm, H.J. Structure-Based Approaches in Modern Drug Discovery Research. *Ernst Schering Res. Found. Workshop* **2001**, *34*, 123–142.
27. Johansson, I.; Oscarsson, M.; Yue, Q.-Y.; Bertilsson, L.; Sjoqvist, F.; Ingelman-Sundberg, M. Genetic Analysis of the Chinese CYP2D Locus. Characterization of Variant CYP2D6 Genes Present in Subjects with Diminished Capacity for Debrisoquine Hydroxylation. *Mol. Pharmacol.* **1994**, *46*, 452–459.
28. Vesell, E.S. Polygenic Factors Controlling Drug Response. *Med. Clin. North Am.* **1974**, *58*, 951–963.
29. Weber, W.W. Pharmacogenetics—Receptors. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 51–80.
30. Kim, R.B.; Wilkinson, G.R. Pharmacogenetics of Drug Transporters. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 81–108.
31. Rae, J.M.; Johnson, M.D.; Lippman, M.E.; Flockhart, D.A. Rifampin is a Selective, Pleiotropic Inducer of Drug Metabolism Genes in Human Hepatocytes: Studies with cDNA and Oligonucleotide Expression Arrays. *J. Pharmacol. Exp. Ther.* **2001**, *299*, 849–857.
32. Noble, E.P. Addiction and Its Reward Process Through Polymorphisms of the D2 Dopamine Receptor Gene: A Review. *Eur. Psychiatry* **2000**, *15*, 79–89.
33. Anderson, K.E. Influences of Diet and Nutrition on Clinical Pharmacokinetics. *Clin. Pharmacokinet.* **1988**, *14*, 325–346.
34. Herken, H. Untersuchungen an Serumproteinen bei Eiweißmangelernährung. *Ärzt. Wochenschr.* **1949**, *4*, 297–302.
35. Okey, A.B. Enzyme Induction in the Cytochrome P-450 System. In *Pharmacogenetics of Drug Metabolism. Int. Encycl. Pharmacol. Therap. Sect. 137*, Kalow, W., Ed.; Pergamon Press: New York, 1992; 549–608.
36. Fuhr, U.; Klittich, K.; Staib, A.H. Inhibitory Effect of Grapefruit Juice and Its Bitter Principal, Naringenin, on CYP1A2 Dependent Metabolism of Caffeine in Man. *Br. J. Clin. Pharmacol.* **1993**, *35*, 431–436.
37. Messina, E.S.; Tyndale, R.F.; Sellers, E.M. A Major Role for CYP2A6 in Nicotine C-Oxidation by Human Liver Microsomes. *J. Pharmacol. Exp. Ther.* **1997**, *282*, 1608–1614.
38. Sellers, E.M. Pharmacogenetics and Ethnoracial Differences in Smoking. *J. Am. Med. Assoc.* **1998**, *8*, 179–180.
39. Dipple, K.M.; McCabe, E.R. Phenotypes of Patients with “Simple” Mendelian Disorders Are Complex Traits: Thresholds, Modifiers, and Systems Dynamics. *Am. J. Hum. Genet.* **2000**, *66*, 1729–1735.
40. Beutler, E. Discrepancies Between Genotype and Phenotype in Hematology: An Important Frontier. *Blood* **2001**, *1*, 2597–2602.
41. Thomson, G. Mapping of Disease Loci. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 337–362.
42. King, R.A.; Rotter, J.I.; Motulsky, A.G. (Eds.). *The Genetic Basis of Common Diseases*, 2nd edition. Oxford University Press, New York, 2002.
43. Loring, J.F.; Wen, X.; Lee, J.M.; Seihamer, J.; Somogyi, R. A Gene Expression Profile of Alzheimer’s Disease. *DNA Cell Biol.* **2002**, *20*, 683–695.
44. Gharavi, A.; Lifton, R. The Inherited Basis of Blood Pressure Variation and Hypertension. In *The Metabolic & Molecular Bases of Inherited Disease*, Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D., Eds.; McGraw-Hill: New York, 2001; 5399–5417.
45. Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D., Eds.; *The Metabolic & Molecular Bases of Inherited Disease*; McGraw-Hill: New York, 2001.
46. Lindpaintner, K. Genetics in Drug Discovery and Development: Challenge and Promise of Individualizing Treatment in Common Complex Diseases. *Br. Med. Bull.* **1999**, *55*, 471–491.



47. Rojas-Martinez, A.; Santillan, A.A.; Delgado-Enciso, I.; Barrera-Saldana, H.A. Genetic Aspects of Asthma. *Rev. Invest. Clin.* **2000**, *52*, 441–450.
48. Barnes, K.C. Gene–Environment and Gene–Gene Interaction Studies in the Molecular Genetic Analysis of Asthma and Atopy. *Clin. Exp. Allergy* **1999**, *4*, 47–51.
49. Drazen, J.M.; Silverman, E.K.; Lee, T.H. Heterogeneity of Therapeutic Responses in Asthma. *Br. Med. Bull.* **2000**, *56*, 1054–1070.
50. Adcock, I.M. Glucocorticoid-Regulated Transcription Factors. *Pulm. Pharmacol. Ther.* **2001**, *14*, 211–219.
51. Bijl-Hofland, I.D.; Cloosterman, S.G.; Folgering, H.T.; van Den Elshout, F.J.; van Weel, C.; van Schayck, C.P. Inhaled Corticosteroids, Combined with Long-Acting beta(2)-Agonists, Improve the Perception of Bronchoconstriction in Asthma. *Am. J. Respir. Crit. Care Med.* **2001**, *1*, 764–769.
52. Leff, A.R. Regulation of Leukotrienes in the Management of Asthma: Biology and Clinical Therapy. *Annu. Rev. Med.* **2001**, *52*, 1–14.
53. Miller, A.L. The Etiologies, Pathophysiology, and Alternative/Complementary Treatment of Asthma. *Altern. Med. Rev.* **2001**, *6*, 20–47.
54. Stafforini, D.M. PAF Acetylhydrolase Gene Polymorphisms and Asthma Severity. *Pharmacogenetics* **2001**, *2*, 163–175.
55. Grasemann, H. Genetics of the Neuronal NO Synthase (NOS1) in the Etiology of Bronchial Asthma. *Pneumologie* **2001**, *55*, 390–395.
56. Summers, W.K.; Majovski, L.V.; Marsh, G.M.; Tachiki, K.; Kling, A. Oral Tetrahydroaminoacridine in Long-Term Treatment of Senile Dementia, Alzheimer Type. *N. Engl. J. Med.* **1986**, *13*, 1241–1245.
57. White, P.; Hiley, C.R.; Goodhardt, M.J.; Carrasco, L.H.; Keet, J.P.; Williams, I.E.; Bowen, D.M. Neocortical Cholinergic Neurons in Elderly People. *Lancet* **1977**, *26*, 668–671.
58. Poirier, J.; Davignon, J.; Bouthillier, D.; Kogan, S.; Bertrand, P.; Gauthier, S. Apolipoprotein E Polymorphism and Alzheimer's Disease. *Lancet* **1993**, *18*, 697–699.
59. Poirier, J.; Delisle, M.C.; Quirion, R.; Aubert, I.; Farlow, M.; Lahiri, D.; Hui, S.; Bertrand, P.; Nalbantoglu, J.; Gilfix, B.M. Apolipoprotein E4 Allele as a Predictor of Cholinergic Deficits and Treatment Outcome in Alzheimer Disease. *Proc. Natl Acad. Sci. U.S.A.* **1995**, *19*, 12260–12264.
60. Cichon, S.; Nothen, M.M.; Rietschel, M.; Propping, P. Pharmacogenetics of Schizophrenia. *Am. J. Med. Genet.* **2000**, *97*, 98–106.
61. Scharfetter, J. Dopamine Receptor Polymorphisms and Drug Response in Schizophrenia. *Pharmacogenomics* **2001**, *2*, 251–261.
62. Bantick, R.A.; Deakin, J.F.; Grasby, P.M. The 5-HT1A Receptor in Schizophrenia: A Promising Target for Novel Atypical Neuroleptics? *J. Psychopharmacol.* **2001**, *15*, 37–46.
63. Schafer, M.; Rujescu, D.; Giegling, I.; Guntermann, A.; Erfurth, A.; Bondy, B.; Moller, H.J. Association of Short-Term Response to Haloperidol Treatment with a Polymorphism in the Dopamine D(2) Receptor Gene. *Am. J. Psychiatry* **2001**, *158*, 802–804.
64. Devlin, B.; Roeder, K.; Bacanu, S.A. Unbiased Methods for Population-Based Association Studies. *Genet. Epidemiol.* **2001**, *21*, 273–284.
65. Kwok, P.Y. Methods for Genotyping Single Nucleotide Polymorphisms. *Annu. Rev. Genomics Hum. Genet.* **2001**, *2*, 235–258.
66. Ross, P.L.; Hall, L.; Haff, L.; Garvin, A. Multiplex Genotyping by Specialized Mass Spectrometry. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 201–222.
67. Sitbon, G.; Syvanen, A.-C. Multiplex Fluorescent Minisequencing Applied to the Typing of Genes Encoding Drug-Metabolizing Enzymes. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 191–200.
68. Medlin, J. Array of Hope for Gene Technology. *Environ. Health Perspect.* **2001**, *109*, A34–A37.
69. Qi, X.; Bakht, S.; Devos, K.M.; Gale, M.D.; Osbourn, A. L-RCA (Ligation-Rolling Circle Amplification): A General Method for Genotyping of Single Nucleotide Polymorphisms (SNPs). *Nucleic Acids Res.* **2001**, *15*, 29.
70. Tsongalis, G.J.; Rainey, B.J.; Hodges, K.A. READIT: A Novel Technology Used in the Interrogation of Nucleic Acid Sequences for Single-Nucleotide Polymorphisms. *Exp. Mol. Pathol.* **2001**, *71*, 222–225.
71. Patil, N.; Berno, A.J.; Hinds, D.A.; Barrett, W.A.; Doshi, J.M.; Hacker, C.R.; Kautzer, C.R.; Lee, D.H.; Marjoribanks, C.; McDonough, D.P.; Nguyen, B.T.; Norris, M.C.; Sheehan, J.B.; Shen, N.; Stern, D.; Stokowski, R.P.; Thomas, D.J.; Trulson, M.O.; Vyas, K.R.; Frazer, K.A.; Fodor, S.P.; Cox, D.R. Blocks of Limited Haplotype Diversity Revealed by High-Resolution Scanning of Human Chromosome 21. *Science* **2001**, *23*, 1719–1723.
72. Hoh, J.; Wille, A.; Ott, J. Trimming, Weighting, and Grouping SNPs in Human Case–Control Association Studies. *Genome Res.* **2001**, *11*, 2115–2119.
73. Clarke, P.A.; te Poele, R.; Wooster, R.; Workman, P. Gene Expression Microarray Analysis in Cancer Biology, Pharmacology, and Drug Development: Progress and Potential (2). *Biochem. Pharmacol.* **2001**, *15*, 1311–1336.
74. Herzig, U.; Cadenas, C.; Sieckmann, F.; Sierralta, W.; Thaller, C.; Visel, A.; Eichele, G. Development of High-Throughput Tools to Unravel the Complexity of Gene Expression Patterns in the Mammalian Brain. *Novartis Found. Symp.* **2001**, *239*, 129–146.
75. Heck, D.E.; Roy, A.; Laskin, J.D. Nucleic Acid Microarray Technology for Toxicology: Promise and Practicalities. *Adv. Exp. Med. Biol.* **2001**, *500*, 709–714.
76. Raetz, E.A.; Moos, P.J.; Szabo, A.; Carroll, W.L. Gene Expression Profiling. *Methods and Clinical Applications*



- in Oncology. *Hematol. Oncol. Clin. North Am.* **2001**, *15*, 911–930.
77. Bubendorf, L. High-Throughput Microarray Technologies: From Genomics to Clinics. *Eur. Urol.* **2001**, *40*, 231–238.
78. Nakanishi, T.; Oka, T.; Akagi, T. Recent Advances in DNA Microarrays. *Acta Med. Okayama* **2001**, *55*, 319–328.
79. Burgess, J.K. Gene Expression Studies Using Microarrays. *Clin. Exp. Pharmacol. Physiol.* **2001**, *28*, 321–328.
80. Jousen, A.M.; Huang, S. Possibilities of Broad Spectrum Analysis of Gene Expression Patterns with cDNA Arrays. *Ophthalmologie* **2001**, *98*, 568–573.
81. Harkin, D.P. Uncovering Functionally Relevant Signaling Pathways Using Microarray-Based Expression Profiling. *Oncologist* **2000**, *5*, 501–507.
82. Lee, M.L.; Kuo, F.C.; Whitmore, G.A.; Sklar, J. Importance of Replication in Microarray Gene Expression Studies: Statistical Methods and Evidence from Repetitive cDNA Hybridizations. *Proc. Natl Acad. Sci.* **2000**, *29*, 9834–9839.
83. Herwig, R.; Aanstad, P.; Clark, M.; Lehrach, H. Statistical Evaluation of Differential Expression on cDNA Nylon Arrays with Replicated Experiments. *Nucleic Acids Res.* **2001**, *1*, 29.
84. Brazma, A.; Hingamp, P.; Quackenbush, J.; Sherlock, G.; Spellman, P.; Stoeckert, C.; Aach, J.; Ansorge, W.; Ball, C.A.; Causton, H.C.; Gaasterland, T.; Glenisson, P.; Holstege, F.C.; Kim, I.F.; Markowitz, V.; Matese, J.C.; Parkinson, H.; Robinson, A.; Sarkans, U.; Schulze-Kremer, S.; Stewart, J.; Taylor, R.; Vilo, J.; Vingron, M. Minimum Information About a Microarray Experiment (MIAME)—Toward Standards for Microarray Data. *Nat. Genet.* **2001**, *29*, 365–371.
85. Haverty, P.M.; Weng, Z.; Best, N.L.; Auerbach, K.R.; Hsiao, L.L.; Jensen, R.V.; Gullans, S.R. HUGEINDEX: A Database with Visualization Tools for High-Density Oligonucleotide Array Data from Normal Human Tissues. *Nucleic Acids Res.* **2002**, *1*, 214–217.
86. Anonymous. Mouse Models of Human Diseases: What Can We Learn? *Trends Genet.* **2001**, *17*, S1.
87. Yoshihara, M.; Ensminger, A.W.; Littleton, J.T. Neurobiology and the Drosophila Genome. *Funct. Integr. Genomics* **2001**, *1*, 235–240.
88. Schaid, D.J.; Rowland, C.M.; Tines, D.E.; Jacobson, R.M.; Poland, G.A. Score Tests for Association Between Traits and Haplotypes When Linkage Phase Is Ambiguous. *Am. J. Hum. Genet.* **2002**, *70*, 425–434.
89. Borrebaeck, C.A. Tapping the Potential of Molecular Libraries in Functional Genomics. *Immunol. Today* **1998**, *19*, 524–527.
90. Witzmann, F.A. Proteomics. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 253.
91. Bals, R.; Jany, B. Identification of Disease Genes by Expression Profiling. *Eur. Respir. J.* **2001**, *18*, 882–889.
92. Husi, H.; Grant, S.G. Proteomics of the Nervous System. *Trends Neurosci.* **2001**, *24*, 259–266.
93. Tang, S.; Helme, D. Bioinformatics WWW Resources. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 291–310.
94. Hyndman, D.L.; Mitsuhashi, M. Applied Bioinformatics. In *Pharmacogenomics*, Kalow, W., Meyer, U.A., Tyndale, R.F., Eds.; Marcel Dekker, Inc.: New York, 2001; 311–336.
95. Langer, T.; Hoffmann, R.D. Virtual Screening: An Effective Tool for Lead Structure Discovery? *Curr. Pharm. Des.* **2001**, *7*, 509–527.
96. Doucet-Personeni, C.; Bentley, P.D.; Fletcher, R.J.; Kinkaid, A.; Kryger, G.; Pirard, B.; Taylor, A.; Taylor, J.; Viner, R.; Silman, I.; Sussman, J.L.; Greenblatt, H.M.; Lewis, T. A Structure-Based Design Approach to the Development of Novel, Reversible AChE Inhibitors. *J. Med. Chem.* **2001**, *27*, 3203–3215.
97. Forster, M.J. Molecular Modelling in Structural Biology. *Micron* **2002**, *33*, 365–384.
98. Gane, P.J.; Dean, P.M. Recent Advances in Structure-Based Rational Drug Design. *Curr. Opin. Struct. Biol.* **2000**, *10*, 401–404.
99. Williams, G.A.; Dugan, J.M.; Altman, R.B. Constrained Global Optimization for Estimating Molecular Structure from Atomic Distances. *J. Comput. Biol.* **2001**, *8*, 523–547.
100. Zeng, J. Mini-Review: Computational Structure-Based Design of Inhibitors That Target Protein Surfaces. *Comb. Chem. High Throughput Screen.* **2000**, *3*, 355–362.
101. Oprea, T.I.; Davis, A.M.; Teague, S.J.; Leeson, P.D. Is There a Difference Between Leads and Drugs? A Historical Perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308–1315.

HISTORY OF DOSAGE FORMS AND BASIC PREPARATIONS

Robert A. Buerki

The Ohio State University College of Pharmacy, Columbus, Ohio, U.S.A.

Gregory J. Higby

American Institute of the History of Pharmacy, Madison, Wisconsin, U.S.A.

INTRODUCTION

The creation and manufacture of dosage forms has been at the center of pharmacy practice for the past thousand years. For American pharmacists of the nineteenth century, *secundem artem*, or the acronym "S.A." in physicians' prescriptions, instructed them to use their special skills "according to the art" of their profession to compound a medicine; it was out of this art, rather than science, that almost all of today's major dosage forms arose. Tablets, capsules, injectables, and oral solutions were all known to pharmacists and physicians a century ago. In addition, there were scores of specialized dosage forms that attempted to meet the medical needs of patients, even if the drugs administered in these doses were ineffective or designed to treat symptoms rather than the underlying disease. The origins of most of these dosage forms are lost in history. For this reason, the authors have elected to forego a contrived narrative tying together the few facts at hand with an equally large amount of speculation about the history of dosage forms. Rather, we have assembled a glossary of terms used in orthodox Western medicine to describe both common and unusual modes of drug administration.

For most of its history, the field of pharmacy was much more concerned with drug preparations than with the resulting dosage forms. Up to the sixteenth century, almost all drugs were derived from plants and were made into preparations that served as the ingredients for medicines; these preparations were called *galenicals*, after the great central figure of Western therapeutics, Claudius Galen of Pergamon (131–201). Strictly speaking, galenicals are pharmaceutical preparations obtained by macerating or percolating crude drugs with alcohol or some other menstruum to remove only the desired principles and leave the inert constituents undissolved. Examples of galenicals include decoctions, extracts, fluidextracts, fluidglycerates, infused oils, infusions, oleoresins, resins, tinctures, and vinegars. The term is used very loosely today to designate any type of simple pharmaceutical preparation, whether it

is an extract of a crude drug or a solution of chemicals. Because galenicals were often administered without alteration, the pursuit of new extraction and other preparative techniques sometimes led to developments in dosage forms as well. The goal of medicine preparers since Galen's time has been to create dosage forms that are stable, free from inert material, therapeutically efficacious, and concentrated to facilitate handling and administration.

Tables 1 and 2 provide a comprehensive classification of ancient and modern pharmaceutical preparations recognized in pharmacopeias and other official and unofficial compendia, but do not make the distinction between dosage forms prepared by maceration or percolation (galenicals) and other chemically or physically similar dosage forms. Preparations are divided into liquids and solids. The liquids are further subdivided as: 1) general aqueous solutions and preparations, 2) sweet or viscid aqueous solutions and preparations, 3) general nonaqueous solutions and preparations, 4) alcoholic or ethereal solutions and preparations, 5) oleaginous solutions and preparations, and 6) parenteral solutions and preparations. The solids are subdivided as: 1) medicated solids, 2) medicated particulate solids, 3) medicated solid applications, 4) oral individual dosage forms, and 5) nonoral individual dosage forms. The dosage forms are arranged alphabetically and include Latin titles and synonyms. The historical development of individual dosage forms is traced to about 1950; it was about then that modern pharmaceutical science was applied in depth to the problems of dosage forms, and the term itself gained general currency in the literature of pharmacy.

DOSAGE FORMS AND BASIC PREPARATIONS

Abstracts: Abstracts are powdered extracts of crude drugs prepared by percolating the drug with an appropriate menstruum, reserving a certain portion, evaporating the weak percolate to a thin extract, blending this extract with the reserve portion and lactose, and evaporating the

Table 1 Liquid pharmaceutical preparations

<i>General aqueous solutions and preparations</i>			
Acids, diluted	Enemas	Milks	Vapors
Aerosols	Fomentations	Solutions	Vinegars
Baths	Gargles	Solutions, irrigating	Washes
Decoctions	Infusions	Solutions, nasal	Washes, mouth
Douches	Inhalations	Solutions, ophthalmic	Washes, aromatic
Draughts	Insessia	Sprays	
Drops	Juices	Suspensions	
<i>Sweet or viscid aqueous solutions and preparations</i>			
Condita	Linctus	Mixtures	Shampoos
Confections	Lohochs	Mucilages	Syrups
Conserves	Lotions	Oxymels	
Honeys	Magmas	Quiddonies	
<i>General nonaqueous solutions and preparations</i>			
Fluidglycerates	Glycerites	Paints, medicinal	
<i>Alcoholic or ethereal solutions and preparations</i>			
Collodions	Elixirs	Juleps	Tinctures
Cordials	Essences	Solutio	Wines
Drops, toothache	Fluidextracts	Spirits	
<i>Oleaginous solutions and preparations</i>			
Balsams	Liniments, dental	Oleates	Petroxolins
Emulsions	Oils	Oleoresins	
Liniments	Oils, infused	Oleovitamins	
<i>Parenteral solutions and preparations</i>			
Ampuls	Injections	Serums	

mixture until dry. The mass is weighed and enough lactose added to make the finished product exactly half the weight of the drug from which it is derived. The resulting product represent twice the strength of the original drug. Although they were convenient for compounding prescriptions, abstracts gained little popularity. Made official in the *United States Pharmacopoeia VI* published in 1880, abstracts were dropped in the next revision (1).

Acetum: See *Vinegars*.

Acids, diluted: Diluted acids are acid preparations, usually 5–10% strength, used for both internal and external medicines. Acids became official with the first *United States Pharmacopoeia* (1820). The term comes from the Latin *acidus* meaning “sharp” or “sour” (2, 3).

Acidum dilutum: See *Acids, diluted*.

Aerosols: Aerosols are a system of finely divided liquid or solid particles dispersed in and surrounded by a gas. The roots of the modern aerosol go back to 1862 and J.D. Lynde, who received a patent for a valve, complete with dip tube, designed to dispense an aerated liquid from a bottle.

In 1899, Helbing and Pertsch added liquefied gases. Patents in the early twentieth century usually were related to spraying perfumes. Aerosol fire extinguishers came into use in the 1930s, and insecticide sprays (bug bombs) appeared during World War II. Introduced in 1947, medical aerosol use increased greatly during the 1950s. In 1952, fewer than half a million such aerosols were produced; by 1963, this figure had risen to almost 40 million (4).

Ampulla: See *Ampuls*.

Ampuls: Ampuls are small, flask-shaped, hermetically sealed glass containers containing a sterile medicinal liquid intended for hypodermic injection, either subcutaneously, intramuscularly, or intravenously. Also, ampul is the class name adopted by the *National Formulary V (N.F.)* (1926) for the solutions in these containers. The ampul was invented in 1886 by the French pharmacist Stanislas Limousin (1831–1887) in response to a need by physicians to conserve their stocks of injectable solutions, which were difficult to transport and deteriorated rapidly due to the development of mold. In his classical essay, “Ampoules hypodermiques; nouveau mode de préparation des solutions hyperdermiques,” published in *Archives of*

Table 2 Solid pharmaceutical preparations

<i>Medicated solids</i>			
Bandages	Electuaries	Lamels	Resins
Cones, medicated	Extracts	Moxa	Silk, oiled
Dressings, medicated	Gums	Papers, medicated	Soaps
Dressings, protective	Inhalants	Papers, waxed	
<i>Medicated particulate solids</i>			
Abstracts	Oil sugars	Salts	Teas
Cucufa	Powders	Salts, artificial	Triturations
Insufflations	Precipitates	Salts, effervescent	
Magisteries	Saccharures	Salts, smelling	
<i>Medicated solid applications</i>			
Applications	Epithema	Ointments, ophthalmic	Rubifacients
Auristilla	Frontalia	Pastes	Sacculi
Cataplasms	Gelatins	Pencils, medicated	Scutum
Caustics	Gels	Plasma	Spasmadraps
Cements	Glycerogelatins	Plasters	Stypes
Cerates	Inunctions	Plasters, adhesive	Swabs
Cerecloths	Litus	Plasters, blister	Vesicatories
Creams	Mulls	Plasters, porous	
Dentifrices	Ointments	Pomatum	
<i>Oral individual dosage forms</i>			
Bacillules	Granules	Pills	Triturates, tablet
Cachets	Masses	Powders, divided	Troches
Capsules	Parvules	Tablets	Wafer envelopes
Capsules, soft	Pastilles	Tablets, hypodermic	Wafers
Dragées	Pearls	Tablets, poison	
Globules	Pellets	Tablets, solution	
<i>Nonoral individual dosage forms</i>			
Bougies	Pessaries	Politzer plugs	Suppositories

Pharmacy (1886), Limousin outlined the essential directions for their manufacture:

These ampoules have the form of a small ovoid balloon. They are terminated by a tapered glass tube, and their capacity is a little greater than one cc.

I sterilize the inside of these small containers, using the method of M. Pasteur, by submitting them in an oven to a temperature of about 200 degrees [Celsius]. I then fill them with the medicated solution, be it by introducing the point of the hot ampoule into the cold liquid, or be it by injecting the hot liquid by means of a small injector at the highest point of the ampoule.

The ampoule being filled, I close it over the oxidation flame by holding the open end of the tube into the uppermost part of the flame.

Although great advances have been made in the techniques and mechanics of ampul production,

Limousin's simple rules capture the basic underlying principles. In the United States, ampuls, or "hermetically sealed containers which are filled with a medicinal liquid in a sterile condition, intended for parenteral use," became official in the *National Formulary V* (1926); that same year, the *United States Pharmacopeia X* included a chapter on sterilization but no monograph for individual ampuls. Iodine Ampuls, N.F., containing Iodine Tincture, *United States Pharmacopeia*, in sealed containers, intended to be broken and the liquid applied topically for the emergency disinfection of cuts or wounds, remained official through the *National Formulary XIII* (1970). The French term *ampoule* came from the Latin *ampulla*, which originally designated an earthen jar container for perfume (2, 4–6).

Antiseptic cottons: See *Dressings, medicated*.

Antiseptic gauzes: See *Dressings, medicated*.

Antiseptic pencils: See *Pencils, medicated*.

Apozemes: See *Decoctions*.

Applicatio: See *Applications*.

Applications: Any preparation for external use (5).

Aqua: See *Waters*.

Artificial salts: See *Salts, artificial*.

Astringent pencils: See *Pencils, medicated*.

Auristilla: Auristilla is a preparation used for medication of the ear canal. The *National Formulary V* (1926) Compound Oil of Hyoscyamus is an illustration of the type, closely resembling the Baumé Tranquille of the French *Codex* (1908) (5).

Bacillula: See *Bacillules*.

Bacillules: Bacillules are rod-shaped lozenges, prepared by massing the lozenge material, rolling it into cylinders or pill pipes, and cutting the cylinders into sections approximately twice the length of the diameter. Licorice lozenges are frequently prepared in this form, a popular example of which was formerly known as Wister's lozenge (see *Troches*) (5).

Bacillum: See *Bougies*.

Balneum: Balneum is a bath for general application (see *Baths*) (5).

Balsams: Balsams are natural solutions of resin in an essential oil, which may also contain benzoic or cinnamic acids. The term has ancient roots in words referring both to spices and to embalming. In premodern times, balsams (*baumé* in French) were any resinous vegetable juices or gums, acquiring their modern meaning in the 1800s. Several official preparations, such as Balsam of Copaiba, do not meet this definition but still carry the name because of their outward similarities with true balsams (1, 7).

Balsamum: See *Balsams*.

Bandages: Bandages are strips or ribbons of muslin gauze or other material employed in surgery for the retention of dressings and for the compression, protection, or support of diseased or injured parts. Bandages may be classified as inelastic, semielastic, elastic, or splint bandages. Inelastic bandages include muslin ribbon or roller bandages, gauze bandages, and

water dressing bandages; semielastic bandages include flannel bandages; elastic bandages include rubber, rubberized, or crepe bandages; splint bandages include plaster-of-Paris bandages and crinoline bandages (gauze stiffened with dextrin) or starch bandages (gauze stiffened with starch paste), used to form the base upon which plaster-of-Paris is applied (5).

Baths: The word baths refers to the external application of water to the body, one of the oldest therapeutic techniques. Although some drugs were added, the simple water bath, or balneum, was most common. In the nineteenth century, when the therapeutic application of water was at its peak, baths were divided into hot or vapor (above 36.1°C), warm (29.4–36.1°C), tepid (18.3–29.4°C), and cold (0–18.3°C). The hot bath was a stimulant; the warm bath was soothing; the tepid bath was for treating skin problems; and the cold bath could be used as a stimulant, tonic, or sedative, depending on the administration technique (8).

Blister plasters: See *Plasters, blister*.

Boluses: Boluses are large pills, over 325 mg (5 grains) in weight. The term comes from the Greek *bolos* meaning “lump” (see *Pills*) (2, 5).

Bougia: See *Bougies*.

Bougies: Bougies are instruments or shaped, solid medications for insertion into the urethra or other body cavities. The term comes from the French *bougie*, signifying a thin wax candle named for the Algerian city, Bougie (2, 5).

Buginarium: A nasal bougie (see *Bougies*) (5).

Cachets: Cachets are lenticular or spoon-shaped rimmed disks pressed from rice-flour wafer sheets, used to administer bitter or nauseating drug powders. The powder is deposited in dry concave cachets, and the rims are moistened with water; empty convex cachets are placed on top, sealing the cachets and enveloping the powder. The term “cachet” is from the French *cache*, “to hide.” The *cachet de pain* was invented by the French pharmacist Stanislas Limousin (1831–1887) in 1873 as an improvement over wafers and wafer envelopes. Limousin also developed a perforated board, accommodating three sizes of cachets, a powder measurer, powder funnels, and wooden “wetter and pressers” to speed extemporaneous cachet production by the pharmacist. A popular brand of cachets and cachet apparatus was manufactured around

1885 under the trade name Konseals by the J.M. Grosvenor Company of Boston around 1885. The Konseal apparatus consists of three perforated, nickel-plated metal plates, hinged together to form a cover plate, a base plate, and a shield plate. Saucer-shaped rice-flour Konseals are pressed into the perforations on the cover and base plates, while the shield plate is folded back over to protect the sealing edges of the Konseals in the base plate. The Konseals in the base plate are filled with powdered drug with the help of special funnels and are tamped down with thimble compressors. The shield plate is lifted, and a moistened roller is passed over the edges of the empty Konseals in the cover plate, which is closed over the base plate, sealing the Konseals. The Konseals are made of thinner material than ordinary cachets, and the finished product is less bulky and neater in appearance. Johann Schmidt later introduced rimless Dry Seal Cachets resembling flattened capsules, which were sealed by pressure without moistening (see *Wafers*) (2, 4, 5).

Cachets, dry seal: See *Cachets*.

Caementum: See *Cements*.

Cambric, oiled: See *Silk, oiled*.

Capsula: See *Capsules*.

Capsula amylacea: Starch capsule (see *Cachets*).

Capsula dura: Hard capsule (see *Capsules*).

Capsula gelatina: Gelatin capsule (see *Capsules*).

Capsula mollis: Soft capsule (see *Capsules, soft*).

Capsules: Capsules are telescoping, interconnecting shells of hard gelatin and sugar used for the administration of solids, masses, and liquids. The term comes from the Latin *capsula*, which is the diminutive of *capsa*, meaning “box.” Successor to the soft capsules invented by the French pharmacist François Mothes in 1833, the telescoping hard gelatin capsule was invented and patented in 1847 by James Murdoch of London. In the United States, the New York firm of H. Planten manufactured two-part capsules sometime after 1836, but their usefulness was impaired by their poor fit. In 1863, the firm developed jujube paste capsules intended for dispensing powders alone, offering them to the trade before 1870. Another manufacturer, Dundas Dick, experimented in the same direction and secured a patent on cone-shaped capsules in 1865. Twelve years later, inspired by reports from Italy,

the Detroit pharmacist F.A. Hubel made molds of iron wire mounted in blocks of wood, which could be dipped in a gelatin solution. Using pins of different diameters for the body and cap allowed Hubel to produce capsule sections which would telescope into each other, producing a small cylinder closed at both ends—the prototype of the modern hard gelatin capsule. Hubel sold his entire output to Parke, Davis & Co. in 1875, securing the first in a series of patents in 1877. Despite this protection, several small competing firms soon emerged that made capsules for other companies to sell or for their own sales organizations, most of which were consolidated by James Wilkie into the U.S. Capsule Company. Around 1901, this firm and its subsidiary, the M.L. Capsule Company, were purchased by Parke, Davis & Co., which expanded and improved its manufacturing processes under Wilkie’s supervision. Wilkie is also credited with introducing phosphorbronze wire for capsule molds, a material superior to iron and one which remained in use until replaced by stainless steel in the 1930s. Capsule manufacturing received its greatest impetus in the 1920s when automatic filling devices were perfected. Pharmacists extemporaneously fill capsules at the prescription counter by “punching” capsule bodies into a smooth layer of medicated powder and filler at a uniform level of compression, replacing the caps, and checking the weight of the filled capsule (see *Capsules, soft*) (2, 9–11).

Capsules, elastic: See *Capsules, soft*.

Capsules, enteric-coated: See *Enteric-coated doses*.

Capsules, gelatin: See *Capsules*.

Capsules, glutoid: See *Enteric-coated doses*.

Capsules, hard: See *Capsules*.

Capsules, soft: Soft capsules are elastic globular shells of gelatin, containing sufficient glycerin to retain permanent flexibility, and intended primarily for the administration of irritating or nauseating oily liquids. They were invented in 1833 by French pharmacist François Mothes; an improved capsule was patented by Mothes and Joseph Dublanc the following year and perfected in 1840 by the court apothecary Adolph Steeger of Bucharest. Soft capsules are easier to swallow than hard capsules and are desirable for administering large volumes of liquids. The original Mothes capsules were hollow globes of soft gelatin with an opening at the top for filling. Once filled, the capsules are sealed with a drop of hot gelatin. In 1846, Giraud introduced olive-shaped elastic capsules with an elongated neck, made by dipping molds into a warm

glycerogelatin solution. After drying, the molds were stripped and the capsule necks sealed for future use. Pharmacists cut off the necks of the capsules as needed, filled them with a dropper or syringe, and sealed them with a drop of warm glycerogelatin or the blade of a hot spatula. The first official formula for the manufacture of soft gelatin capsules appeared in the French *Codex Medicamentarius* of 1866 (see *Capsules*) (4, 5, 10, 5, 12).

Capsules, starch: See *Cachets*.

Carbasus: Gauze see *Dressings, medicated*.

Cataplasma: See *Cataplasms*.

Cataplasms: Cataplasms are moist substances intended for external application and are of a consistency as to accommodate themselves accurately to the surface to which they are applied, without being so liquid as to spread over the neighboring parts, or so tenacious as to adhere firmly to the skin. Cornelius Celsus (50–25 B.C.) described early Roman softening ointments or malagma (from the Latin *malasso*, meaning “soften”) made by boiling flour in water to make a stiff paste that was admixed with melted gums or wax. The word “cataplasm” derives from two Greek words, *kata*, meaning “down,” and *plasso*, meaning “to mold.” Modern cataplasms are made by rubbing together glycerin and a dry powder, such as kaolin, to make a very firm mixture. Since cataplasms (or poultices) were commonly used in the United States in domestic practice and made in the home, they were rarely prepared by pharmacists. A kaolin cataplasm became official in the *United States Pharmacopeia VIII* (1905) but was transferred to the *National Formulary IV* in 1916 (2, 4, 5, 13).

Caustic pencils See *Pencils, medicated*.

Caustics: Caustics are local remedies that destroy life on the part of the body to which they are applied. The term comes from the Greek *kaustikos*, meaning “to burn.” The strongest common caustic (*causticum commune acerrimum*) was potassium hydroxide, or caustic potash, used to form issues on the body or to open abscesses. The *United States Dispensatory* (1836) advised that the most convenient mode of employing a caustic to form an issue was:

To apply to the skin a piece of linen spread with adhesive plaster, having a circular opening in its centre corresponding to the intended size of the issue, and then to rub upon the skin within the opening a piece of the caustic previously moistened

at one end. The application is to be continued until the life of the part is destroyed, when the caustic should be carefully washed off by a wet sponge or wet tow, or neutralized by vinegar.

Caustic potash was sometimes used to remove strictures in the urethra or, in solution, as an application to the spine in treating tetanus. A contemporary caustic, silver nitrate, was described in the seventh century A.D. by Geber; it remains official in the *United States Pharmacopeia* as Fused Silver Nitrate. Christopher Glaser, apothecary to Louis XIV, first prepared it in sticks for use as a caustic (2, 5, 13).

Cements: The 7th edition of the *National Formulary* (1942) listed a formula for a Cement of Zinc Compounds and Eugenol, which was widely used by dentists as a temporary filling. The cement was supposed to exert a sterilizing effect and protect the dentine from further destruction. The formula passed out of the *National Formulary* with its 9th edition (1950) (14).

Cerates: Cerates are unctuous substances for external application as dressings for inflamed surfaces. Derived from the ancient Greek *keroma* (from *keros*, “wax”), cerates are generally made with oil, lard, or petrolatum as a basis, with sufficient wax, paraffin, spermaceti, or resin added to raise the melting point of the oils and fats employed. Cerates should be of such consistency that they may be easily spread at ordinary temperatures upon muslin or a similar material with a spatula and, yet, not so soft as to liquefy and run when applied to the skin. The most widely used cerate was Ceratum Cantharidis, a blistering plaster made of Spanish flies that was official in the *United States Pharmacopeia* and later in the *National Formulary* until 1950 (see *Plasters, blister*) (5).

Ceratum: (See *Cerates*).

Cerecloths: Cerecloths are an early form of dipped plasters, described by William Salmon in his *Pharmacopoeia Londonensis* (1691), used chiefly to lay upon issues—small ulcers produced by caustics or cutting, the discharge from which was encouraged to fulfill certain therapeutic indications. Also known as spasmodraps or sparadraps (see *Plasters*) (15).

Cereolus: A cereolus is an urethral bougie (see *Bougies*) (5).

Charta: The term comes from the Greek *chartes* or papyrus (see *Papers, powder*) (2).

Charta amylacea: Starch papers (see *Cachets*).

Charta cerata: (see *Papers, waxed*).

Chrisma: A salve (see *Ointments*) (5).

Clysters: See *Enemas*.

Coatings, pill: Rhazes (850–923) used a mucilage of psyllium seed to coat offending pills; a century later, Avicenna (980–1037) introduced silver and gold coatings for pills not merely to mask bad taste but to enhance the supposed medicinal effect. Later, the influential seventeenth-century Parisian physician and pharmacist Jean de Renou recommended that pills with a bitter taste should be gilded and mixed among some powdered spices. Coating pills with gold and silver leaf was commonly practiced in France, other parts of Europe, and the United States until well into the nineteenth century but then fell into disuse. Mohr, Redwood, and Procter's classic *Practical Pharmacy* (1849) reported that foil-coated pills were "still occasionally administered, but much less frequently than formerly." Sugar-coated and gelatin-coated pills had their first acceptance as an indirect consequence of the invention of the gelatin capsule. In 1838, M. Garot, a Parisian pharmacist, coated offending pills with gelatin. A year earlier, the French pharmacist Labelonie had recommended that pills of cubeb and copaiba be covered with sugar, a process that was patented by Adolphe Fortin of Paris in 1837. Over the next several years, other French pharmacists (Deschamps, Bousquet, Mayer, and Roman) secured patents for coating pills with various combinations of sugar, honey, and acacia. By 1862, England's Bernard Proctor could list 45 distinct processes for coating pills with a variety of substances. In the United States, the first manufacture of sugar-coated pills was associated with the patent-medicine industry, probably as early as 1845. New Yorkers Cornelius V. Clickener and Zadoc Porter and a Philadelphia physician named Swayne all claimed to be the "inventor" of the sugar-coated pill but doubtlessly adapted French technology. By 1857, five different sugar-coated patent-medicine pills were available to American pharmacists, but no process was available for pharmacists who wished to sugar-coat pills extemporaneously. About that same time, Henry A. Tilden and William R. Warner independently developed processes for manufacturing sugar-coated pills and began selling them in bulk; in 1866, Warner began manufacture under his own firm name, William R. Warner & Company. That same year, Henry Wathew of Philadelphia developed and patented a prototype of the angled coating pans still employed today.

By the mid-1870s, a variety of small mechanical coating pans were available to American pharmacists, although the quality of the extemporaneous coatings that could be achieved was generally inferior to that of the manufactured variety, which caused pharmacists to relinquish the practice (16–18).

Coatings, tablet: After the introduction and acceptance of the compressed tablet as a dosage form superior to pills, the same techniques that had been employed for sugar-coating pills were adapted to tablets (see *Coatings, pill*). Moreover, compression coating techniques were also introduced by Charles Carter of Philadelphia as early as 1858. In 1896, P.J. Noyes invented an apparatus incorporating a movable die cavity into which the coating material was fed; the powder was compressed to form a coating around the pill or tablet by striking the upper punch with an automatic hammer. The following year, Noyes patented an improved machine that not only applied the coating but also performed the preliminary compression of the tablet. F.J. Stokes received a patent for a similar machine in 1917, although the compression coating of tablets did not become widely adopted until the 1950s (18).

Collodions: Collodions are liquid, external preparations with a base of pyroxylin dissolved in a mixture of alcohol and ethyl oxide or similar solvent. They were used medicinally soon after the discovery of gun cotton by Schoebein in 1846, first as a surgical dressing by Maynard in 1847 (19).

Collodium: See *Collodions*.

Collunarium: See *Solutions, nasal*.

Collutorium: See *Washes, mouth*.

Collyrium: See *Solutions, ophthalmic*.

Compressed pills: See *Tablets*.

Compressed tablets: See *Tablets*.

Condita: Condita are comprised of candied or preserved roots or fruits. They are made by boiling the plant parts until tender, soaking them in hot syrup, and then pouring off the syrup; the procedure was repeated several times, reusing the syrup after boiling it down to a thick consistency. Condita can also be a compound of wine, honey, and spices (often pepper). Condita are also known as confitures or sweetmeats (18, 20).

Cones, medicated: Cones are light, porous hemispherical masses of sucrose and egg albumin, used as a vehicle for homeopathic medications. The cones, also called disks, are designated (in millimeters) according to size by the diameter of the base. The common size (No. 6) should absorb about 2 drops of dispensing alcohol. Cones are medicated by adding a sufficient quantity of the dilution to saturate them and pouring off the excess liquid (5).

Confectio: See *Confections*.

Confections: Confections are saccharine, soft solids, in which one or more medicinal substances are incorporated to provide an agreeable form of administration and a convenient method for preservation. In the thirteenth century, some apothecaries were called *confectionarii* from *confectio* meaning “a composition.” Confections are made by adding medicinal ingredients in either the form of a smooth paste, a fine powder, or a liquid to a basis of finely powdered sugar. Confection of Rose and Confection of Senna were official in the *National Formulary* through the 5th edition (1926) (2, 5).

Confitures: See *Condita*.

Conserva: See *Conserves*.

Conserves: Conserves are confections prepared from fresh medicinal agents and refined sugar, beaten into a uniform mass. The *United States Dispensatory* (1836) noted that “as active substances even thus treated undergo some change, and those which lose their virtues by desiccation cannot be long preserved, the few preserves now retained are intended rather as convenient vehicles of other substances, than for separate exhibition.” Conserves are also known as preserves (also see *Confections*) (5, 13, 20).

Cordiale: See *Cordials*.

Cordials: Cordials are sweetened alcoholic preparations of high alcoholic content. They are also a tonic medicine formulated to stimulate the heart (5, 20).

Cottons, antiseptic or medicated: See *Dressings, Medicated*.

Creams: Creams are semisolid emulsions, usually medicated, intended for external application. The term “cream” has been used to refer to a wide variety of opaque, soft semisolids or thick liquids intended for external use. The *British Pharmacopoeia* classifies creams as medicated

liquid emulsions consisting of a mixture of anhydrous lanolin, olive oil (or other fixed oil), and lime water; milk of magnesia is often referred to as a cream, and many cosmetic preparations are called creams. Claudius Galen (131–201) prepared the first *unguentum refrigerans* or “cold cream,” an ointment containing olive oil, rose oil, white wax, and a small quantity of water. This was a prototype for other cosmetic ointments introduced by Johann Mesue, Jr., in the thirteenth century. A modern version, consisting of almond oil, spermaceti, white wax, and rose water, passed into the *United States Pharmacopoeia* as “Rose Water Ointment” or “Galen’s Cerate.” Pharmaceutical creams became official with the introduction of the formula for a “Sun Cream” in the *National Formulary VIII* (1946), a product designed to prevent sunburn but permit tanning. Cosmetic creams, which are usually not medicated, include preparations classified as all-purpose creams, baby creams, barrier creams, bleaching creams, cleansing (or rolling) creams, cold (or fatty) creams, foundation (or vanishing) creams, hair creams, and hand creams (4, 5, 21).

Cremor: See *Creams*.

Cucufa: Cucufa is a cap, dusted on the inside with medicinal powder, which is applied to the head to strengthen the brain (22).

Curatio: See *Dressings, medicated*.

Decoctions: Decoctions are a solution of vegetable drugs, obtained by boiling the substances in water; they are also known as apozemes (8).

Decoctum: See *Decoctions*.

Dental liniments: See *Liniments, dental*.

Dental wax: See *Drops, toothache*.

Dentifrices: Dentifrices are powders, pastes, washes, or medicated soaps used for cleaning the teeth. Dentifrices are usually flavored with aromatic oils, frequently contain soap, almost always some form of chalk, and are applied with a toothbrush. The *National Formulary* carried official formulas for a liquid dentifrice through its 5th edition (1926) and a powder dentifrice through its 11th edition (1960) (5).

Dentifricium: See *Dentifrices*.

Dentilimentum: See *Liniments, dental*.

Dermatologic pastes: See *Pastes*.

Diluted acids: See *Acids, diluted*.

Disks: See *Cones, medicated*.

Dispensing tablets: See *Tablets, solution*.

Douches: Douches are a column of fluid, of a certain nature and temperature, allowed to fall on a part of the body. Air can also be used as a douche (20).

Dragées: Dragées are candied or preserved roots and fruits described by the Arab Najm ad-dyn Mahmoud in the eighth century and reintroduced in the eighteenth century by the famous French pharmacist Moyse Charas (see *Confitures*). By the middle of the nineteenth century, the term was extended to include a type of sugar-coated pill formed by repeatedly shaking slightly moistened, tiny 6-mg (1/10 grain) sugar granules (or nonpareils) in a basin of finely powdered drug mixed with sugar. After a sufficient number of layers had been built up, the dragées would receive a final coating of sugar or copal and tolu balsam, a painstaking process described by Ernest Agnew in the *American Journal of Pharmacy* (1870) (see *granules*) (18, 23).

Draughts: Draughts are liquid medicines usually prepared to be taken in a single dose or "draught." Draughts are also known as potions (5).

Dressings, medicated: Medicated dressings are external applications resembling ointments in consistency but remaining semisolid at body temperature. Paraffin Dressing, formerly official in the *National Formulary VI* (1936), was employed as an air-excluding, soft, pliable, analgesic, splint-like covering for surfaces denuded by burns. A wide range of materials have been used as coverings or protectives to apply heat or medicaments to a diseased or injured part, to prevent wound infection, and to absorb and prevent decomposition of wound discharges. Antiseptic or Medicated Cottons include borated cotton, iodoform cotton, iodized cotton, and styptic cotton; Antiseptic or Medicated Gauzes include borated gauze, corrosive sublimate gauze, carbolated gauze, iodoform gauze, and picric acid gauze; Antiseptic or Medicated Lints include borated lint; Cellulose Waddings or wood wools include moss, peat, sawdust, jute, and oakum or marine lint. See the *British Pharmaceutical Codex* (1924) for representative formulas (also see *Bandages* and *Dressings, protective*) (5, 24).

Dressings, protective: Protective dressings are comprised of a wide range of materials employed as wound coverings, either to shield the parts from external infection or to prevent the escape of fluids contained in the dressing. Protective dressings are also used as a covering for poultices and for the retention of heat. Common protective dressings include oiled silk, oiled muslin or cambric, waxed paper, gutta percha tissue, rubber dams, mackintosh or jaconet, and rubber sheeting (5).

Drops: Drops are pharmaceutical mixtures meant to be given in small amounts. Before the twentieth century, the term applied to solutions used in small quantities expressed in "drops." These were commonly strong medicines "dropped" into water, such as Vinegar of Opium or "black drop." In modern pharmacy, the term became more associated with the need to get a medicine into an appropriately small amount of vehicle for application to the eye (ophthalmic), ear (otic), or passages of the nose (nasal). As a dosage unit, the drop is troublesome because it can vary greatly in size, depending on the size of the dropper orifice and the surface tension of the liquid. The *United States Pharmacopeia IX* (1910) set the official dropper at 20 drops per gram of water at 15°C \pm 10% (3, 20).

Drops, ear: See *Drops*.

Drops, eye: See *Drops*.

Drops, nose: See *Drops*.

Drops, toothache: Drops for toothache are comprised of a solution of phenol in oil of cinnamon and methyl salicylate. The solution entered the *National Formulary V* (1926). A mixture of phenol, creosote, or volatile oils dissolved in paraffin, with a few filaments of cotton added, and molded into sticks constitutes Dental Wax (5).

Ductum: See *Douches*.

Dry seal cachets: See *Cachets*.

Ear drops: See *Drops*.

Eclectus: See *Linctus*.

Eclegma: See *Lohoch*.

Effervescent salts: See *Salts, effervescent*.

Elaeosacchara: See *Oil sugars*.

Elastic capsules: See *Capsules, soft*.

Electuaries: Electuaries are confections prepared from dried medicinal agents, especially powders, combined with syrup or honey in order to render them pleasant to the taste and convenient for internal use. The *United States Dispensatory* (1836) noted that electuaries “should not be so soft ... as to allow the ingredients to separate, nor so firm ... as to prevent them from being swallowed without mastication.” French writers recommend using brown sugar syrup to prepare electuaries, because it is less apt to crystallize than that made from refined sugar. The term comes from the Greek words *ek*, meaning “out,” and *leichein*, “to lick” (2, 13) (see *Confections*).

Electuarium: See *Confections*.

Elixirs: Sweetened, hydroalcoholic, flavored liquid medicines, which became popular in mid-nineteenth-century America. The word is derived apparently from the Arabic *al-iksir*, which is an Arabic form of the Greek, *xirion*. Originally the term meant “dry powder.” Elixirs came into medicine through their connection with alchemy. Elixir Rubrum, one of the most renowned alchemical compounds, could supposedly turn mercury to gold or prolong life. The term was picked up by followers of Paracelsus (1493–1541) and became applied to liquid preparations. European elixirs were generally bitter. One of the first American elixirs was Cordial Elixir of Quinine (ca. 1838), made by John T. Heinitch of Lancaster, Pennsylvania. After the Civil War an “elixir craze” began, which led to scores of companies competing for business. As much as any other development, the “craze” led to the publication of the first *National Formulary* in 1888 (25, 26).

Embrocatio: See *Liniments*.

Emplastrum: See *Plasters*.

Emulsions: Emulsions are a preparation consisting of two immiscible liquids, usually water and oil, one of which is dispersed as small globules in the other. Before the late seventeenth century, the term only applied to natural emulsions, such as ground almonds and water, which resembled milk. In 1674, a physician named Grew reported the preparation of oils and egg yolk to the Royal Society of Great Britain. In the 1700s, other emulsions were made with acacia, honey, tragacanth, and other natural emulsifying agents. In the 1800s, the wet-gum (ca. 1850) and dry-gum (ca. 1870) methods were established as standard preparation techniques. Interest in medicinal

emulsions peaked in the early to mid-twentieth century with the development of several new emulsifying agents. Originally listed under “Mixtures” in the *United States Pharmacopeia*, they are a separate entry in the 7th revision (1890) (4).

Emulsum: See *Emulsions*.

Enemas: Enemas are injections of liquid, either plain or containing drugs, into the rectum and colon to empty the lower intestine or to introduce food or medicine for therapeutic purposes. Enemas are one of the most ancient and widely used methods of introducing therapeutic substances into the body. The origins of use are lost in prehistory, but written records in Egypt before 1000 B.C. describe enemas being used to both cleanse the bowel and administer medicines. These early enemas consisted of three parts: a vehicle (usually water, beer, or milk), an emollient (usually oil or honey), and a medicinal substance. The Greek historian Herodotus (fifth century B.C.) attributed the general good health of the Egyptians to their use of enemas and claimed they had achieved their expertise through their experiences with injecting embalming fluids via the anus. Pliny (first century A.D.), however, argued that the Egyptians had learned to administer enemas by watching the ibis use its curved bill to inject itself with Nile River water as a purge. Hippocrates recommended enemas to treat fevers and constipation. Other authors of antiquity, such as Galen and Oribasius, wrote at length about what substances could be introduced via enemas. Enema apparatus was first described in detail by Arabian physicians of the eleventh and twelfth centuries. Albulcasis described a device made of an anal tube or funnel attached to a bag made from an animal bladder or sheep skin. The metal piston syringe came into use in the 1400s and supplanted these clumsy bags. The enema syringe soon became the object of much medical ingenuity, particularly in France. In 1480, Louis XI suffered a severe stroke and recovered, giving credit to the enemas he had received, beginning a 400-year period of fascination with the enema among the French, who refined the syringe apparatus. Molière referred to enemas repeatedly in his works, and Ambrose Paré devised a syringe instrument about 1580 that allowed enema self-administration. During the seventeenth century, other special apparatus were designed to allow administrations of tobacco smoke via enemas. The Dutch anatomist Regner de Graaf completed the first book-length study of enemas in 1668. In the mid-eighteenth century, gum rubber began to be used in enema apparatus, replacing skin bags. In 1820, John Read developed a two-way syringe with ball valves (the first modern stomach

pump), which was also used for enemas. Spring-loaded syringes and other advances followed, but by the late nineteenth century the dangers of improper enema use became apparent, especially the consequences of high-pressure administration. By the mid-twentieth century, enemas were again administered with simple funnels and gravity, a reflection of their prehistoric origin. Enemas are also known as clysters or glysters (27, 28).

Enteric-coated capsules: See *Enteric-coated doses*.

Enteric-coated doses: Enteric-coated doses are dosage forms that have been coated or chemically treated to prevent disintegration in the stomach but which disintegrate in the intestinal tract. Enteric coating is employed when the medicinal substances would be decomposed or rendered inactive by gastric enzymes or when they would irritate the gastric mucosa. The *Pacific Medical and Surgical Journal* (1867) noted that collodion protects pills from dissolving in the stomach, but gastric insolubility as a basis for medication is generally credited to the German dermatologist Paul Unna, who introduced keratin-coated pills in 1884; Ceppi introduced salol as an enteric coating in 1891. Glutoid capsules are a special form of enteric coating prepared by subjecting soft or hard gelatin capsules to the action of formaldehyde until they become insoluble in the stomach but not in the intestine. The hardening process was developed in Switzerland by Weyland in 1895 and patented by Hausmann in Germany that same year. Pharmacists have extemporaneously coated capsules, pills, and tablets with salol, keratin, casein, and shellac, or with a mixture of *n*-butyl stearate, carnauba wax, and stearic acid to create enteric coatings. Early patented enteric coatings included keratin (Pohl, Germany, 1885), fat-covered capsules enclosed in membranous sacs (Webb and Webb, Great Britain, 1906), and benzoin-coated capsules (Horigan, 1928). Later patented coatings utilized cellulose nitrate and acetate (Volwiler, 1928), cellulose esters and ethers with saponifiable organic compounds (Glaessner, Austria, 1931), ammoniacal bleached shellac (Wruble, 1933), stearic acid, carnauba wax, petrolatum, elm bark, and agar (Miller, 1935, and Worton, 1938), and abietic, oleic, and benzoic acids with methyl abietate (Eldred, 1937). Since 1940, research on enteric coatings has focused on the synthesis of resinous polymers, which are insoluble in acids, such as cellulose acetate phthalate (Hiatt, 1940) and a glycerol-stearic acid-phthalic anhydride ester (Volweiler and Moore, 1940) (4, 5, 12, 17).

Enteric-coated pills: See *Enteric-coated doses*.

Enteric-coated tablets: See *Enteric-coated doses*.

Epispastics: Epispastics are local remedies, the application of which produce a serous discharge beneath the cuticle, forming a blister. Epispastics are also known as vesicatories (13).

Epithema: Epithema consists of topical applications other than ointments or plasters. Liquid epithema include fomentations, soft epithema include cataplasms, and dry epithema include bags filled with dried drugs (20).

Escharotica: See *Caustics*.

Escharotics: Caustic applications (see *Caustics*) (5).

Essences: The term essences sometimes refers to a volatile oil or to a simple tincture; most often, the term is used interchangeably with spirits (20).

Essencia: See *Essences*.

Extracta: See *Extracts*.

Extracts: Extracts are either pasty or semisolid masses or dry, solid, or powdered products prepared by exhausting drugs with appropriate solvents, carefully evaporating the products to fixed standards. An extract is intended to preserve the useful constituents of a drug in a concentrated, relatively uniform, permanent condition, and in a form suitable for medication. The *Edinburgh Pharmacopoeia* (1817) and *Dublin Pharmacopoeia* (1826) distinguish between extracts prepared from infusions, decoctions, or tinctures, and those prepared from the expressed juices of plants, calling the latter *succi spissati* or inspissated juices. Three forms of extracts are recognized: semiliquid or those of syrupy consistency; plastic masses known as pilular or solid extracts; and dry powders known as powdered extracts (4, 5, 13).

Extractum: See *Extracts*.

Eye Drops: See *Drops*.

Eye Washes: See *Solutions, ophthalmic*.

Fluidextracts: Fluidextracts are concentrated liquid preparations representing the therapeutically active principles of vegetable drugs. They are formulated in such a way that the activity of one gram of the drug is contained in one milliliter of the fluidextract. They are generally prepared by some form of percolation, using

alcohol in the menstruum. Fluidextracts first became official in 1850 when five were entered in the *United States Pharmacopoeia* by their heyday in the late nineteenth century, almost 100 were official. Joseph Remington called fluidextracts “American preparations,” because they were developed by such native pharmaceutical scientists as William Procter, Jr., and Edward Squibb. Fluidextracts were perhaps the ultimate galenical class because of their permanence, concentrated form, and uniform relationship between the strength of the extract and the drug it contained. With the dominance of synthetic chemical drugs in the twentieth century, fluidextracts all but disappeared, except for those used for flavoring purposes (1, 19, 29).

Fluidextractum: See *Fluidextracts*.

Fluidglycerates: Fluidglycerates are a class of fluidextracts in which a mixture of glycerin and water is used as the primary menstruum during percolation instead of alcohol and water. The preparation of these extracts was suggested by Beringer in 1908. They were briefly official from the 5th–7th editions of the *National Formulary* (1926–1942) (3).

Fluidglyceratum: See *Fluidglycerates*.

Fomentations: Fomentations consists of an external application of cloths dampened with hot water or a medicinal decoction. Narcotic drugs were sometimes used. Dry fomentations were heated bricks wrapped in cloth and applied externally (5, 20).

Fomentum: See *Fomentations*.

Fotus: See *Fomentations*.

Frontalia: Medicines applied to the forehead (20).

Gargarisma: See *Gargles*.

Gargles: A gargle is a liquid medicine intended to be retained in the mouth and placed in contact with the back of the throat by throwing back the head and agitated by air released from the larynx (20).

Gauzes, antiseptic or medicated: See *Dressings, medicated*.

Gelatin capsules: See *Capsules*.

Gelatina: See *Gelatins*.

Gelatins: Gelatins are semisolid gelatinous preparations for internal or external use (5).

Gelatum: See *Gels*.

Gels: Gels are semisolid organic or inorganic colloids rich in liquid, consisting of hydrated threads or granules of the dispersed phase intimately associated with the dispersion medium. Although Francesco Selmi studied inorganic colloids in the 1840s, modern colloid science began in 1861 with the work of Thomas Graham, who investigated diffusion and dialysis and introduced such terms as colloid, glue, sol, gel, peptization, and syneresis. In the early 1900s, Freundlich introduced the terms *lyophilic* and *lyophobic* to describe colloids in which the dispersed phase has a high or a low affinity, respectively, for the dispersion medium. In 1950, Weiser divided gels into inorganic gels, which include gelatinous precipitates (such as Milk of Magnesia) and inorganic jellies (such as Bentonite Magma), organic gels or jellies (such as Pectin Paste), and crystalline or amorphous jellylike networks in which both solid and liquid phases are continuous (see *Magma*) (4).

Globules: Globules are spheres of sugar (see *Pills*). They can also be round or oval glycerogelatin capsules (see *Pearls*) (5).

Glutoid capsules: See *Capsules, glutoid*.

Glycerinum: See *Glycerites*.

Glycerites: Glycerites are solutions of medicinal substances in glycerin introduced in the 5th revision of the *United States Pharmacopoeia* (1873). Glycerites afford a rapid and simple method of making aqueous solutions of phenol, tannic acid, tar, and other substances that are not otherwise easily soluble. This class of preparations is called glycerins in Great Britain (4, 5).

Glyceritum: See *Glycerites*.

Glycerogelatins: Glycerogelatins are soft, medicated masses, usually molded into the form of blocks, which melt at body temperature and have as a base a mixture of gelatin, glycerin, and water; Glycerated Gelatin United States Pharmacopoeia is generally used as a base. At the time of application, the blocks are melted, and the liquid applied to the skin with a soft brush. Four glycerogelatins were introduced into the 3rd edition of the *National Formulary* (1906), and a general formula remained official

through the 8th edition (1946). Under the title *Gelatinum*, the *British Pharmaceutical Codex* (1922) included preparations that were either similar to glycerogelatin, that more closely resembled dermatologic pastes, or that were intended for internal administration and called jellies (4, 5).

Glycerogelatinum: See *Glycerogelatin*.

Glysters: See *Enemas*.

Gossypium: Cotton (see *Dressings, medicated*).

Granula: Granular effervescent salts (see *Salts, effervescent*).

Granules: Granules are small spheres of sugar pellets that are saturated with liquid medication before being swallowed. They can also be very small pills of 0.06 g or less, sometimes called parvules, made by slightly moistening sugar granules (or nonpareils) with a syrup in which an active ingredient has been dissolved. Granules containing 1 mg of powerful medications, such as arsenious acid or aconite, found favor among physicians in Europe, particularly in Italy, because they provided powerful drugs in small, precise doses (see *Dragées*; also see *Pills*) (4, 23).

Gums: Gums are mucilaginous, amorphous, transparent, or translucent glucosidal principles of plants used internally (as demulcents or expectorants), externally (as emollients or protectives), or for their emulsifying action. Three distinctive types of gums are recognized: arabin, which is completely soluble in water; bassorin and cerasin, which are partially soluble or swell in contact with water; and mucilages and pectins, which swell to form jellies. Many gums are complex mixtures of several of these types (5).

Guttae: See *Drops*.

Hard capsules: See *Capsules*.

Haustus: See *Draughts*.

Homeopathic tinctures: See *Tinctures*.

Homeopathic triturations: See *Triturations*.

Honey: Preparations with honey as the vehicle (8).

Infused oils: See *Oils, infused*.

Infusions: Infusions are aqueous solutions obtained by soaking vegetable drugs in cold or hot (not boiling) water (8).

Infusum: See *Infusions*.

Inhalants: Inhalants are products consisting of finely powdered or liquid drugs that are carried into the respiratory passages by the use of powder blowers or low-pressure aerosol containers holding a suspension of the drug in a liquefied propellant. A dry inhalation is a product consisting of finely powdered drugs that are carried into the respiratory passages with the help of special devices. Inhalants can also refer to drugs or a combination of drugs which, by virtue of their high vapor pressure, can be carried by an air current into the nasal passages where they exert their effect. In the latter form of inhalant, the drug is absorbed on fibrous material and enclosed in an inhaler—a plastic or metal tube fitted with a cap to prevent loss of medicament when not in use. The patient removes the cap, inserts the nasal tip into a nostril, and breathes the air drawn through the inhaler to obtain the drug (4).

Inhalatio: See *Inhalations*.

Inhalations: Inhalations are medicinal agents administered by breathing in gases and vapors. A wide variety of techniques and purposes fall under this category. The isolation of pure gases by Priestley, Scheele, and others in the late eighteenth century motivated Thomas Beddoes to found the Pneumatic Institute (1798) in England to treat lung diseases. The gases used for inhalation therapy included oxygen, nitrous oxide, and ether. Gases were applied as surgical anesthesia in the 1840s largely through the efforts of the Americans Crawford Long, Horace Wells, and W.T.G. Morton; their innovations were the first great contributions to medical science by citizens of the United States. The therapeutic use of gases continued throughout the nineteenth century, especially for the treatment of tuberculosis. During this period, devices were designed to vaporize liquid drugs with steam for the treatment of lung disorders. The first apparatus developed to atomize medicinals, made by Berson in 1860, was a combination inhaler and atomizer operated by steam that was generated by heating water in a closed vessel. These devices were quite popular for home use, and their modern counterparts remain a part of home health care. The term *inhalation* has also been applied to preparations more properly called inhalants, such as Compound Eucalyptus Inhalation, or sprays, such as

Epinephrine Inhalation (see *Inhalants and sprays*) (24, 30, 31).

Inhalations, dry: See *Inhalants*.

Injectio: See *Injections*.

Injections: Injections are sterile solutions or suspensions used for administering pharmaceutical preparations by intravenous, subcutaneous, intramuscular, and intraspinal injection. The term *injections* is also the class name adopted by the 5th edition of the *National Formulary* (1926) for solutions in ampuls intended for hypodermic injection. Although ancient humans may have invented the concept of introducing drugs through punctures in the skin by attempting to recreate the effects of venomous snake and insect bites through the use of poisoned arrows, perhaps the first introduction of medication through the skin for medicinal purposes was inoculation for smallpox. Human inoculation with the smallpox virus by pricking the body with needles dipped in pus from an active case of the disease was practiced for centuries among people of the Orient but was only introduced into Western medicine about 1717. In 1796, Edward Jenner (1749–1823) performed his first vaccination with material from a cowpox sore. In 1657, Sir Christopher Wren was the first to inject a drug intravenously, a practice successfully adopted by the English practitioner Johan D. Major in 1662 under the title “*chirurgica infusoria*.” Physicians experimented with the injection of water, opium, arsenic, cinnamon, oil of sulfur, and other substances with limited and often fatal results. Injections of purging medicines, such as jalap resins, were particularly popular in the treatment of syphilis. Nevertheless, the successful utilization of intravenous injection awaited the proof of the germ theory of disease and the discovery of sterile methods by Pasteur, Koch, Lister, and others in the latter half of the nineteenth century; the introduction of the hypodermic syringe, which was suggested by Charles G. Pravez of Lyons in 1853, popularized by Alexander Wood of Edinburgh and Charles Hunter of London in 1855–1858, and improved by Luer in 1894; and the invention of the ampul by Stanislas Limousin in 1886. A committee of the Royal Medical and Chirurgical Society of London gave approval to hypodermic injections in 1867, the same year the *British Pharmacopoeia* published a monograph for the first official injection, *Injectio Morphinae Hypodermica*. No attempt was made to sterilize these solutions, but E.R. Squibb (1873) and others recommended that parenteral solutions could be preserved by the addition

of small amounts of carbolic acid, salicylic acid, chloroform, or camphor. At about the same time (1875), John Tindall developed the process of sterilization by discontinuous heating, which bears his name. Nevertheless, hypodermic routes of administration were slow to gain widespread recognition, for physicians became increasingly aware of fevers and other toxic symptoms following the use of crudely prepared injections. Ehrlich’s introduction of hypodermic injections of salvarsan for syphilis (1910) provided the strongest impetus for the development of parenteral administration, stimulating a series of rapid advancements in technique. In 1911, Martindale and Wynn discussed the pharmaceutical manipulation of salvarsan, emphasizing sterilization and aseptic techniques. That same year, Hort and Penfold applied the term *pyrogens* to describe substances that cause a febrile reaction upon injection. They found that distilled water sealed in sterile containers gave rise to toxic symptoms, whereas freshly distilled water caused no reaction; later that same year, Wechsellmann showed that febrile reactions could be eliminated if solutions were made from sterile distilled water. Seibert confirmed these findings in 1923, noting that poorly constructed stills could produce pyrogenic distilled water. In 1930, Rademaker formulated a rigid set of aseptic precautions and rules, governing the preparation of parenteral fluids, which are still valid today, and setting the stage for modern intravenous medicines. Injections are also known as parenterals (4, 5, 24, 32–33).

Insessia: Insessia refers to a vapor bath, usually administered by having the patient sit on a perforated chair, beneath which is placed a large container filled with hot water or a hot decoction of a plant drug (20).

Insufflatio: See *Insufflations*.

Insufflations: Insufflations are powders used for blowing into the nose, preferably by means of one of the various kinds of powder blowers or insufflators made for the purpose. They may also be applied directly in the way in which snuff is usually taken. Also, a snuff. The term comes from the Latin *insufflare* meaning “to breath into” or “to blow into” (2, 5).

Inunctions: Inunctions are ointments applied with friction, intended for local application and quick absorption. The term was formerly applied to preparations consisting of wool fat in which mercury or other medicinal agents were incorporated. A Compound Menthol Inunction (retitled Compound Menthol Ointment in 1936)

remained official in the *National Formulary* until 1960 (5, 24).

Inunctum: See *Inunctions*.

Irrigating solutions: See *Solutions, irrigating*.

Irrigatio: See *Solutions, irrigating*.

Jellies: See *Gels*.

Juices: Juices are liquids obtained by expression from the fresh parts of plants. The *British Pharmaceutical Codex* (1949) contained monographs for the juices of garlic, lemon, and taraxacum. In the United States, Cherry Juice remained official through the 16th edition of the *National Formulary* (1985) as a flavoring agent (4).

Juices, inspissated: Extracts prepared from the expressed juices of plants (see *Extracts*).

Juleps: A julep is a sweet drink, usually a demulcent, acidulous, or mucilaginous mixture. Much more popular in Europe than in the United States, a typical julep of the late nineteenth century (*mistura gummosa*) contained 10 parts acacia triturated with 30 parts syrup of acacia, 100 parts water, and 10 parts orange flower water (3, 20).

Konseals: Konseals is a trade name for the brand of cachets and cachet apparatus manufactured by the J.M. Grosvenor Company of Boston about 1885, originally introduced as "Morstadt's cachets" by Karl Morstadt of Prague (see *Cachets*) (5).

Lamella: See *Lamels*.

Lamels: Lamels are small disks, about 3 mm in diameter, cut or stamped from thin films of glycerinated gelatin, containing definite quantities of various medicaments used in ophthalmology. Lamels, or eye disks, are applied with a camel's-hair brush to the inner surface of the lower eyelid, where they are immediately dissolved in the lachrymal fluid. Four lamels were official in the *British Pharmacopoeia* until 1953 (24).

Lavatio: See *Washes, mouth*.

Ligamentum: See *Bandages*.

Linctus: A linctus is a thick viscid liquid that must be licked from a spoon or licorice stick, from the Latin *lingere*, "to lick" (see *Lohochs*) (5).

Liniments: Liniments are external preparations of a consistency thicker than water, but thinner than ointments, usually applied to the skin with a gentle rubbing of the hands. Liniments are among the oldest of dosage forms, along with related forms, such as plasters and ointments. The term came to its present use about 1600. Drying liniments are preparations which dry when smeared on the skin, forming a medicated film removable by water (5, 8, 19).

Liniments, dental: Dental liniments were introduced in the *National Formulary V* (1926) through cooperation between dental and pharmaceutical authorities; these concentrated, often poisonous liniments, were designed to be rubbed into the gums (5, 19).

Liniments, drying: See *Liniments*.

Linimentum: See *Liniments*.

Linimentum: exsiccantum: Drying liniment (see *Liniments*).

Lints, antiseptic or medicated: See *Dressings, medicated*.

Lintum: Lint (See *Dressings, medicated*).

Liquor: See *Solutions*.

Litus: A litus is fluid preparation applied with a brush (see *Paints, medicinal*) (5).

Lohochs: Lohochs are thick syrupy medicines, usually used to fight a cough, to be sipped slowly or licked, sometimes called a looch. If in the latter dosage form, it is usually called a linctus (5).

Looch: See *Lohochs*.

Lotio: See *Lotions*.

Lotions: Lotions are fluid preparations, usually containing suspended insoluble material and applied externally. They are different from liniments by being aqueous, rather than oleaginous or alcoholic in nature. Lotions were official in the first *National Formulary* (1888). In the mid-nineteenth century, lotions were often applied by wetting linen and placing on the affected area (1, 20).

Lozenges: See *Troches*.

Magisteries: Certain precipitates from saline solutions bore this title, but it was usually applied to secret remedies (20).

*Magma*s: In modern pharmacy, the term magma means an aqueous preparation containing precipitated inorganic material in a fine state of subdivision. *Magma Bismuthi* and *Magma Magnesia*e (milk of magnesia) were the first official magmas in the *United States Pharmacopeia IX* (1916). In the nineteenth century and previous eras, this term referred to the residue obtained after expressing organic substances to extract their fluid parts, usually referred to as a marc (19, 20).

Massa: See *Masses*.

Masses: Masses are plastic, semisolid pharmaceutical preparations composed of active medicinal substances combined with a diluent or filler and an excipient, capable of being shaped into pills with little or no further treatment. The three essential requirements of a pill mass are: adhesiveness (the mass must be sufficiently adhesive to retain its shape and yet be soft enough to be worked by the fingers or suitable apparatus into the desired form); firmness (the mass must possess sufficient firmness to permit the pills to retain their shape); and plasticity (a natural result of the proper degree of adhesiveness and firmness). Two masses appear to have survived the changes in modern medicine: Mass of Mercury (or Blue Mass), a cholagogic preparation last official in the *National Formulary IX* (1950), and Ferrous Carbonate Mass (or Vallet's Mass), a hematinic last official in the *National Formulary X* (1955) (4, 5, 24).

Medicated cottons: See *Dressings, medicated*.

Medicated dressings: See *Dressings, medicated*.

Medicated gauzes: See *Dressings, medicated*.

Medicated oils: See *Oils*.

Medicated paints: See *Paints, medicated*.

Medicated papers: See *Papers, medicated*.

Medicated pencils: See *Pencils, medicated*.

Mel: See *Honeys*.

Mellita: See *Honeys*.

Milks: Historically, the term *milk* has been applied generally to any liquid that possesses the outward appearance of milk, such as milk of magnesia. Legend holds that the class of beauty preparations called toilet milks may have arisen from Cleopatra and her milk baths. The modern milks are oil-in-water emulsions, named for their appearance and use as additives to baths. Moreover, actual milk was used and modified, especially with the addition of malt, as a medicinal beverage. Fermented Milk, or kumyss, is fresh cow's milk, to which sugar and yeast are added for fermentation. Fermented milk was official in the 3rd through the 5th edition of the *National Formulary* (1906–1926). Humanized Milk was a combination of cow's milk and fresh cream, plus Humanizing Milk Powder, which contained a small amount of Compound Pancreatic Powder and lactose, made official in *National Formulary III* (1906) (7, 34).

Mistura: See *Mixtures*.

Mixtures: Mixtures are aqueous preparations for internal use containing insoluble, nonfatty substances. They differ from emulsions in containing no fat and from liniments in being used internally. Both mixtures and emulsions were originally grouped under *Mistura* in the *United States Pharmacopoeia*; in the 7th revision (1890), emulsions were given separate recognition. Examples include Compound Mixture of Opium and Glycyrrhiza (Brown Mixture), Carminative Mixture (Dalby's Carminative), Mixture of Copaiba (Lafayette Mixture), Mixture of Copaiba and Opium (Chapman's Mixture), Mixture of Magnesia, Asafetida, and Opium (Dewees' Carminative), Oleobalsamic Mixture (Hoffman's Balsam), Compound Mixture of Opium and Chloroform (Squibb's Diarrhoea Mixture), and Expectorant Mixture (Stokes' Expectorant) (3, 5).

Mouth washes: See *Washes, mouth*.

Moxa: Moxa are cones of combustible matter used for cauterization by burning (see *Cones*). Moxibustion, the burning of moxa, was an ancient method of counter-irritation or cautery arising out of China. Small cones of combustible organic material (originally *Artimesia moxa* or common mugwort) were placed on certain areas of the skin, ignited, and allowed to burn down, leaving a blister. Moxa entered Western medicine in the seventeenth century as a treatment for gout but fell into disuse a century later along with other forms of cautery (5, 28).

Mucilages: Mucilages are viscid preparations made by dissolving or suspending gummy substances in water.

The term comes from the Latin *mucus*. The gummy substances, if natural, are carbohydrates obtained from the exudates of trees or shrubs. Gums have been used since ancient Egypt. Hippocratic works (ca. 400 B.C.) mention acacia. Mucilage of Acacia was official in the first *United States Pharmacopoeia* (1820), with Mucilage of Tragacanth becoming official in the 1st revision (1830). Both have been used as thickening agents or to prevent the creaming of emulsions (2, 19).

Mucilago: See *Mucilages*.

Mulla: See *Mulls*.

Mulls: Mulls are ointments of high fusion points, containing the desired medicinal agent, and spread on soft muslin or mull in a manner similar to that of ordinary spread plasters. The most suitable base for preparing mulls is a mixture of suet and lard, with the occasional addition of wax or lead oleate plaster. Mulls are prepared extemporaneously by tacking unsized mull over a sheet of moistened parchment paper and spreading a melted, partially cooled ointment over the mull with a broad, flat bristle brush and smoothing the surface with two warmed elastic spatulas. When cooled, the mull is covered with waxed paper and rolled into a cylinder for dispensing. Salicylic Acid Mull, Salicylated Creosote Mull, Corrosive Mercuric Chloride Mull, and Zinc Mull were last official in the *National Formulary V* (1926) (5).

Muslin, oiled: See *Silk, oiled*.

Nasal solutions: See *Solutions, nasal*.

Nebula: See *Sprays*.

Nodulus: Nodulus is an abbreviation for *nodulus uterinus*, a form of uterine bougie.

Nose drops: See *Drops*.

Oculentum: See *Ointments, ophthalmic*.

Odontalgicum: See *Toothache drops*.

Oil sugars: Mixtures of sugar with fixed and volatile oils, rendering the oils miscible to water to an certain extent, offered as a convenient mode of administering medicines to children. The *National Formulary* offered a general formula for 2% oil sugars through its 7th edition (1946) (5, 13).

Oiled silk: See *Silk, oiled*.

Oils: Any liquid that greases; i.e., leaves, when dropped on a cloth, a stain which water does not wash out; this stain makes paper translucent. If a solid substance exhibits similar properties, it is called a fat. Oils are called volatile or fixed, according to whether this stain disappears on warming or is permanent because of the nonvolatility of the oil. A few medicated oils, such as Phenolated Oil N.F. and Phosphorated Oil N.F., were briefly official (3, 31).

Oils, infused: Oleaginous preparations for external use made by macerating a drug with alcohol and ammonia water, and digesting the mixture with sesame oil at 60–70°C until the alcohol and ammonia water have evaporated. The most common, Infused Oil of Hyoscyamus, was used to make Compound Oil of Hyoscyamus. This preparation was popular in France for the treatment of earache under the name *baumé tranquille* (3).

Ointments: Semisolid preparations intended for application to the skin with or without inunction. In addition to serving as vehicles for the topical application of medicinal substances, ointments may also serve as emollients for the skin and as protectives to prevent contact of the skin surface with aqueous solutions and skin irritants. Ointments are made by fusion, incorporation, or chemical reaction. Ancient humans attributed special powers to the fats of animals and humans, and their mixtures with resins, waxes, powdered herbs, and minerals represent one of the earliest dosage forms employed. A greaseless ointment, consisting of hartshorn beaten up with incense and flour and mixed with sweet ale appears in the Ebers papyrus (1500 B.C.). The Greeks did not distinguish between liquid or semisolid preparations used for an ointment, but rather according to use or ingredients. For example, *malagma* were softening ointments, whereas *keroma* were wax ointments, the predecessor of the later cerates. Plant mucilages, balsams, and oils mixed with wax were also classified as ointments. Galen's rose water ointment (or cold cream) was an early departure (second century A.D.) from the entirely fatty type of preparation (see *Creams*). By the thirteenth century, pharmacists distinguished between *olea* or oils (liquid oily ointments), *emplastra* or plasters (masses sticking firmly to the skin), and *unguenta* or ointments (semisolid smears), a concept that remained unchanged for nearly 500 years. The first *United States Pharmacopoeia* (1820) recognized lard as the chief ingredient of the first official ointments, which were rendered to the consistency of butter by the addition of suet, wax, or spermaceti. By the middle of the nineteenth century, however, natural ointment bases began to be replaced by artificial bases, introduced with special regard

to the purposes they were to serve. Schacht introduced Glycerite of Starch in 1858, a translucent jelly prepared by heating glycerin and starch in certain proportions to a certain temperature, and W.A. Miller introduced Petrolatum in 1873 as "Cosmolin and Paraffin Ointment," both of which preparations were adopted by the *United States Pharmacopeia* in 1880. In 1885, the pharmacologist Oscar Liebreich rediscovered the therapeutic value of wool fat, the *oesypus* of the ancient Greeks, which he called lanolin; it was recognized by the *Pharmacopeia* in 1893. Later, the Russian chemist Lifschuetz discovered that the emulsification power of lanolin depended upon the free alcohols he had isolated as a group (1895–1898). In 1907, the dermatologist Paul G. Unna introduced eucerin, a new ointment base consisting of 1 part of Lifschuetz's alcohols, 20 parts of paraffin ointment, and 20 parts of water, the forerunner of the American "Aquaphor." Between 1920 and 1944, hydrogenated oils, sulfated and sulfonated hydrogenated oils, as well as stearic acid, sodium stearate, self-emulsifying glyceryl stearate mixtures, polymers of glycols (such as polyethylene glycol 4000), and esters of these glycols (such as polyethylene glycol monostearate) became important, followed by such modern bases as Plastibase, attapulgate, Veegum, guar gum, Carbopol, and the silicones, which were introduced between 1945 and 1959 (4, 5, 24).

Ointments, ophthalmic: Sterile ointments designed for application to the eyelids. Petrolatum, petrolatum-mineral oil, and petrolatum-anhydrous lanolin bases are often used in ophthalmic ointments because of their low irritating potential. Finely powdered, sterile active ingredients are aseptically incorporated into a sterile base, using sterile utensils, and dispensed in sterile ophthalmic-tipped tubes to reduce the possibility of contamination (4).

Ointments, softening: See *Cataplasms*.

Oleates: Usually liquid preparations made by dissolving alkaloids in oleic acid. Oleate of Mercury, however, is an ointment-like product of mercuric oxide in oleic acid. Oleic acid was named by the pharmacist Chevreul after olives (2, 5).

Oleinata: See *Oleates*.

Oleatum: See *Oleates*.

Oleoresina: See *Oleoresins*.

Oleoresins: Extracts of plant drugs prepared by percolation using a selective solvent (usually ether or acetone),

followed by complete removal of the solvent by evaporation. The term first arose in the 1820s (Buchner and Peschier), and the oleoresins first appeared in the *United States Pharmacopeia* of 1860 through the efforts of William Procter, Jr. (19).

Oleosacchara: See *Oil sugars*.

Oleovitamins: Preparations using fish liver oil, fish liver oil diluted with an edible vegetable oil, or a solution of vitamin concentrate in fish liver oil or in an edible vegetable oil. Oleovitamins were created during World War II to fill a therapeutic gap created by the interruption in cod liver oil supplies. The class became official in the second supplement to the *United States Pharmacopeia XI* (1942) as a source of vitamins A and D (31).

Oleum: See *Oils*.

Oleum infusum: See *Oils, infused*.

Ophthalmic ointments: See *Ointments, ophthalmic*.

Ophthalmic solutions: See *Solutions, ophthalmic*.

Orbicules: Spherical globules of sugar (see *Pills*) (5).

Oxymels: Acid-honey preparations containing acetic acid or vinegar (see *Honeyes*) (4).

Paints, medicinal: Liquid medicinal preparations possessing antiseptic, caustic, soothing, or stimulating properties, usually applied by means of a brush. Paints intended to remain in contact with a specified surface are usually prepared with collodion, glycerin, glycerin and water, egg albumin in alcohol, or gutta percha. Paints intended to be absorbed are prepared with oleic acid or fatty oils. Caustic substances are usually applied dissolved in distilled water, alcohol, or ethereal vehicles, whereas resinous substances, such as benzoin, storax, tolu balsam, or sandarac dissolved in ether, are employed as bases for medicated varnishes, and used for application to the skin and raw mucous surfaces (5).

Papers, medicated: Preparations intended primarily for external application, either by saturating paper with medicinal substances or by applying the latter to the surface of the paper by the addition of some adhesive liquid. Potassium Nitrate Paper remained official in the *National Formulary* through its 5th edition (1926), but the most widely used paper is Mustard Paper, commonly

called mustard plaster, a mixture of powdered black mustard and a solution of rubber spread on paper, cotton cloth, or other fabric, which remained official in the *National Formulary* through its eleventh edition (1960). Mustard poultices or cataplasms were formerly called sinapisms, after the botanical name for black mustard, *Sinapis nigra*, and were described in the *United States Dispensatory* (1836) as

Powerfully rubefacient... usually becoming insupportably painful in less than an hour... . As a general rule, the poultice should be removed when the patient complains much of the pain; and in cases of insensibility should not, unless greatly diluted, be allowed to remain longer than one, or at most two hours, as violent inflammation, followed by obstinate ulceration, is apt to occur.

Home-made mustard plasters (equal parts of mustard and flour, moistened with tepid water to form a paste and applied to the skin in a muslin bag) still play a role in folk medicine (5, 13, 24).

Papers, powder: See *Powders, divided*.

Papers, waxed: Parchment-like paper treated with melted wax or paraffin used largely as an economical substitute for more expensive protective dressings (see *Dressings, protective*) (5).

Parenterals: See *Injections*.

Parogenum: See *Petroxolins*.

Parvula: See *Parvules*.

Parvules: Small sugar-coated pills, of 0.06 g or less, sometimes incorrectly called granules. (also see *Granules*) (4, 5).

Pasta: See *Pastes*.

Pasta dermatologica: Dermatologic pastes (see *Pastes*).

Pastes: Ointment-like mixtures of starch, dextrin, zinc oxide, sulfur, calcium carbonate, or other medicinal substances made into a smooth paste with glycerin, soft soap, petrolatum, lard, or other fats, and medicated with antiseptic or astringent agents, designed for external use. Early pastes, such as *Pasta Glycyrrhizae* and *Pasta Althaeae*, were internal preparations, most of which were of gum-like consistency. The modern pastes were introduced by the noted dermatologists Paul G. Unna

and Oskar Lassar around 1900. Dermatologic Pastes normally contain a higher proportion of powdered material than that included in ointments and are less greasy but more absorptive than other preparations for external application. The *British Pharmaceutical Codex* groups Witch Hazel Cream, Vanishing Cream, Tannic Acid Jelly, Catheter Lubricant, and a wide assortment of “medicated preparations for external application, employed principally as antiseptic, caustic, cooling, protective or soothing dressings” under the title “Pastes.” Pastes entered the *National Formulary III* in 1906 (4, 5, 24).

Pastes, dermatologic: See *Pastes*.

Pastilles: A form of lozenge, particularly those which are chocolate flavored (see *Lozenges*); also, combustible cones of aromatic drugs used for fumigation (see *Cones*). The term came into English usage around 1650 from the French *pastille*, which was derived from the Latin *pastillus*, meaning “little loaf” (2, 5).

Pastillus: See *Pastilles* and *Cones*.

Pearls: Round or oval capsules made by enclosing liquids, solids, or tablets in a shell of glycerogelatin material. Pearls are less elastic than soft capsules and contain no air space, the glycerogelatin shell being completely filled with the medicinal substance. Pharmacists made pearls extemporaneously by laying a softened sheet of glycerogelatin over a warmed molding plate containing a specified number of semicircular (or other shaped) depressions. The sheet was covered with a measured quantity of medicinal liquid, and the liquid with a second sheet of glycerogelatin to exclude air. The whole assembly was covered with a matching molding plate, and compressed with a mechanical press to form the pearls, an exacting and time-consuming process requiring special apparatus. This process has been superseded in industry by a continuous automated process in which a liquid is injected between two ribbons of gelatin while passing between revolving dies (also see *Capsules, soft*) (4, 5).

Pelleta: See *Pellets*.

Pellets: Small spheres of sucrose saturated with an alcoholic tincture, primarily used in homeopathic medicine. Pellets are made in different sizes, designated according to the diameter of ten pellets measured in millimeters. *Remington's Practice of Pharmacy* (1926) states that pellets “should be made of the purest materials,

should be perfectly white and odorless and able to withstand all the tests prescribed for sucrose or cane sugar” (also see *Globules*) (5, 24).

Pencils, antiseptic: See *Pencils, medicated*.

Pencils, astringent: (See *Pencils, medicated*).

Pencils, caustic: (See *Pencils, medicated*).

Pencils, medicated: Cylinders used in dermatologic practice to apply medicinal agents directly to the skin. The medicinal agent is incorporated into a paste consisting of starch, dextrin, tragacanth, and sucrose with sufficient water to form a plastic mass, which is rolled into cylinders about 5 mm in diameter, cut into sections about 5 cm long, dried on parchment paper at room temperature, and wrapped in tinfoil. Medicated pencils intended as a caustic application, such as sticks of silver nitrate, are sometimes referred to as escharotica. Salicylic Acid Pencils, the last official medicated pencils, appeared in the *National Formulary V* (1926); also known as Antiseptic, Astringent, Caustic, Salve, or Styptic Pencils (5).

Pencils, salve: See *Pencils, medicated*.

Pencils, styptic: See *Pencils, medicated*.

Perles: See *Pearls*.

Pessaries: Medicated vaginal suppositories, globular or oviform in shape, weighing between 4 g (if made from oil of theobroma or cocoa butter) and 10 g (if made from glycerated gelatin). The term derives from a Greek word describing the small stones used for playing the game of draughts (5, 35).

Pessarium: See *Pessaries*.

Pessum: See *Pessaries*.

Petroxolins: Fluid preparations for external use with a base of liquid petrolatum and ammonium oleate. The preparations became official in the 6th edition of the *National Formulary* (1936) (3).

Pigmentum: See *Paints, medicinal*.

Pill coatings: See *Coatings, pill*.

Pills: Small, solid masses of a globular, ovoid, or lenticular shape intended for oral administration. Pills are

prepared by incorporating medicinal agents with other materials to form a cohesive, plastic mass, which is divided into the requisite number of portions, each of which is formed into the desired shape. Pills usually range in weight from 0.10 to 0.30 g. Exceptionally large pills of 0.60 g or more are referred to as boluses; very small sugar-coated pills of 0.06 g or less are known as parvules or granules; pellets, globules, or orbicules are small spheres of sugar saturated with an alcoholic tincture, largely used in homeopathic medicine. When the pill came into use in ancient Mesopotamia and Egypt, it offered for the first time a definite dose corresponding with a desired therapeutic action. The Greeks named the little balls of medicine *katapotia* (“something to be swallowed”), later Latinized to *catapotium*; by the first century A.D., the term *pilula* came into use in Rome. Pills persisted as a major dosage form for a remarkably long period of time. For example, over half of the prescriptions dispensed at a Charlestown, Massachusetts, pharmacy during the years 1872–1875 were for pills. Moreover, certain combinations of drugs persisted for thousands of years in pill form. The Pills of Rufus, for example, were originally a *hiera* (bitter powder) made into pill form by the Arabs and popularized by Avicenna (980–1033); a modern version, Pills of Aloe and Mastic, were last official in the 8th revision of the *United States Pharmacopeia* (1905) (4, 5, 36–37).

Pills, compressed: See *Tablets*.

Pills, enteric-coated: See *Enteric-coated doses*.

Pilula: See *Pills*.

Pilula comprimata: Compressed pills (see *Pills*).

Pilula enterica: Enteric-coated pills (see *Pills*).

Pilules: See *Pellets*.

Plasma: Nonfatty unctuous preparations (5).

Plasters: Substances intended for external application, made of such materials and of such consistency as to adhere to the skin. Adhesive plasters afford protection and mechanical support, whereas medicated plasters furnish an occlusive and macerating action, bringing the medication into close contact with the skin; when spread on perforated cloth, the product is called a porous plaster. Plasters are among the most ancient of all pharmaceuticals. Primitive humans may have used plasters of mud and leaves to help heal wounds or relieve pain. The Ebers papyrus (1500 B.C.)

describes several plasters and poultices for treating burns, and the Greeks assigned a special place within their temples where plasters were prepared. Indeed, the word plaster is derived from the Greek *emplastron* meaning “to smear on” or “to mold on.” The famous diachylon plaster, made from oil, litharge, and certain plant juices was compiled by Menecrates, physician of the Emperor Tiberius about 39 A.D. and passed on in verse form to Claudius Galen (131–201), who developed a number of practical formulas for plasters and procedures for their preparation which endured for centuries. A modern version of diachylon plaster survived as Lead Oleate Plaster (or Lead Plaster) in the *United States Pharmacopeia* and later in the *National Formulary* through its 8th edition (1946). Originally, plasters were spread directly on the affected part of the body; by the sixteenth century, plaster material was being spread on linen or leather or dipped. In 1514, Giovanni da Vigo popularized spread plasters called sparadraps or spasadraps; in 1691, William Salmon described dipped cerecloths in his *Pharmacopoeia Londonensis*. Plasters were spread with the aid of an offset spatula or plaster iron heated by means of a spirit lamp, an arduous and time-consuming process which exhausted the patience of pharmacists who struggled trying to keep the refractory plaster masses warm and malleable. Although Elisha Perkins of Baltimore is credited with obtaining the first American patent for a manufactured plaster (1830), John C. De La Cour of Camden, New Jersey, is generally credited as among the earliest producers of machine-spread adhesive plasters (1836), one of the first commercially prepared dosage forms manufactured in the United States. In 1852, an English apothecary named Mather patented a method of spreading plasters on leather by the use of heated rollers, producing a thinner and more uniform product, leading Edward Parrish to observe in his classic text, *American Pharmacy* (1856), that “the spreading of plasters which was formerly an important part of the business of the apothecary has now ... been monopolized by manufacturers who bring machinery to their aid.” The only plaster which still finds significant use today is Salicylic Acid Plaster, *United States Pharmacopeia*—the common corn plaster (also see *Plasters, adhesive*, and *Plasters, porous*) (4, 5, 15, 24, 38).

Plasters, adhesive: A mixture of rubber, resins, and waxes, with a filler of absorbent powder, such as zinc oxide, orris root, or starch, mechanically mixed and spread on cotton cloth. Early adhesive plasters were composed of resin and litharge or diachylon spread on calico, muslin, or linen. In 1843, B.C. Rowland of Liverpool, England, reported on an “India Rubber Court Plaster” made by

dissolving India rubber in naphtha or turpentine and spreading the liquefied rubber on silk or satin. Two years later, Horace Day and William Shecut of New York City patented a combination of India rubber and gums as a plaster mass. In 1852, Benjamin Nickels of Surrey, England, patented an “elastic plaster” combining adhesive material on an elastic fiber. In 1863, Joshua Melvin of Lowell, Massachusetts, patented the manufacture of adhesive plaster in roll or cylindrical form. By the early twentieth century, two types of rubber-based adhesive plasters emerged: surgeons’ adhesive plaster, a plain yellow-colored mass, and zinc oxide adhesive plaster, a white mass containing zinc oxide. Modern adhesive plasters, consisting of vinyl resin, plasticizers, and other chemical additives have an excellent ability to remain adhered under severe conditions of moisture and heat, and rarely cause skin irritation. Adhesive plaster remained official through the 19th revision of the *United States Pharmacopeia* (1975) (also see *Plasters* and *Plasters, porous*) (5, 15, 24).

Plasters, Blister: Plasters designed to produce inflammation, blisters, or issues (running sores). Many blistering agents were used to prepare blister plasters, but the potent and powerful cantharides (or Spanish flies), was the most widely applied. It was used internally by the ancients, who were well aware that it produced hematuria even when applied externally. Aretaeus introduced an external preparation of cantharides as a blistering agent in the second century to the shaved head to relieve headache. This treatment was also used for epilepsy and vertigo, and persisted through the centuries, but only after the patient had been given milk to drink for 3 days to protect the bladder from injury. “In many constitutions the strangury will ensue, especially where the discharge of serous juice is too great,” an English textbook on materia medica advised in 1730. “But, however it be, such applications are necessary when a patient proves delirious, as frequently happens in high fevers.” A century later, the *United States Dispensatory* (1836), in discussing the use of Cerate of Cantharides, “the common blistering plaster of the shops,” remarked that

When the full operation of the flies is desirable, and the object is to produce a permanent effect, the application should be continued for twelve hours... . It should then be removed, and followed by a bread and milk poultice, or some other emollient dressing, under which the cuticle rises, and a full blister is usually produced. By this management the patient will escape strangury, and the blister will very quickly heal after the discharge of the serum.

Elisha Perkins of Baltimore obtained a patent in 1830 for an apparatus to prepare blister plasters. George W. Carpenter of Philadelphia sold and recommended Perkin's blister cloth in 1831 as "a very convenient article for the country physician, being ready spread for immediate use." *Ceratum Cantharides* was official in the *United States Pharmacopeia* and, later, in the *National Formulary* until 1950 (13, 15, 28).

Plasters, porous: Adhesive plasters spread on perforated cloth. Sir William Butts, royal physician to Henry VIII, devised a "spasmodrap or dypped plaster" which was to be poked "full of smalle hoolys." In 1845, Horace Day and William Shecut of New York City patented a rubber-based "porous plaster" rendered full of minute holes to "allow the free escape of the perspiration." In 1854, Somerville Scott Alison of London patented a porous or "perforated Lambskin... prepared according to the process called chamois curing." The porosity is considered a mechanical advantage in that it prevents the plaster from slipping from the point of application, each opening serving as a stop. Porous plasters are also far more comfortable than the nonporous variety, which they have superseded (also see *Plasters* and *Plasters, adhesive*) (5, 15).

Politzer plugs: Greased pellets of cotton about the size of a coriander seed with a thread attached, for insertion into the ear as a protective. The pellets were named after Adam Politzer (1835-1920), an Austrian otologist (1, 2).

Pomatum: Fats saturated with odorous principles (5).

Potio: See *Draughts*.

Potions: See *Draughts*.

Potus: See *Draughts*.

Poultices: Originally spelled "pultes" in sixteenth-century England, the term came from the Latin *puls* or *pultes* meaning "a pottage of meal" (also see *Cataplasms*) (2).

Powder papers: See *Papers, divided*.

Powders: Intimate mixtures of dry, powdered medicinal substances reduced to a fine powder by the processes of comminution and trituration. Bulk powders are divided into two categories: simple, consisting of one substance, and compound, consisting of two or more powders mixed together. One of the most ancient compound powders was *hiera picra* ("sacred bitters"), a mixture of aloe and canella introduced about 500 B.C. as a laxative, the prototype of a

large number of bitter powders containing aloe as the principal ingredient and bearing the general title *Hiera*. This powder was listed in various pharmacopoeias and was last recognized in the 4th edition of the *National Formulary* (official until 1926). Powders were designed originally as a convenient mode of administering hard vegetable drugs such as roots, barks, and woods; powders were also found to be convenient for the dispensing of insoluble chemical compounds such as calomel, bismuth salts, mercury, and chalk. Famous compound powders of the past include: Compound Powder of Glycyrrhiza, a variant of Compound Senna Powder recognized by the first *London Pharmacopoeia* (1618); Powder of Ipecac and Opium, or Dover's Powders, introduced by the English physician Thomas Dover as *Pulvis Diaphoreticus* in the early eighteenth century; Aromatic Powder of Chalk, a simplified version of a complex confection devised by Sir Walter Raleigh during his imprisonment and introduced into the *London Pharmacopoeia* of 1721 as *Confectio Raleighana*; and Antimonial Powder, patented in 1747 as Dr. James's Fever Powder. The *United States Dispensatory* (1836) noted that "the form of powder is convenient for the exhibition of substances which are not given in very large doses, are not very disagreeable to the taste, have no corrosive property, and do not deliquesce rapidly on exposure." Today, drugs not available in capsule or tablet form can still be conveniently administered in powdered form by placing them on the back of the tongue and swallowing them with water. Flavored powders are particularly useful for children who might have difficulty swallowing a tablet or capsule (also see *Powders, divided*) (4, 5, 13, 24).

Powders, divided: Intimate mixtures of dry, powdered medicinal substances intended for oral administration, divided into single doses, each of which is folded into a small sheet of glassine paper. One of the oldest of dosage forms, divided powders have largely been replaced by capsules or tablets. Nevertheless, the preparation of powders permits the drugs to be reduced to a very fine state of subdivision, a physical condition which frequently intensifies their therapeutic activity, a factor in increasing the efficacy of homeopathic triturations, calomel and sodium bicarbonate mixtures, and Dover's Powder (Powder of Ipecac and Opium). Divided powders also furnish a convenient means for administering drugs that are not excessively bitter, nauseous, or otherwise offensive to the taste. One of the most durable divided powders is Seidlitz Powders, a saline cathartic originated and patented by Thomas Savory in 1815. Savory claimed that the powders owed their value to the mineral properties of the Seidlitz

spring in Germany, which contains magnesium sulfate. The powders consist of sodium bicarbonate (wrapped in blue paper) and tartaric acid and potassium and sodium tartrate (wrapped in white paper), each of which are dissolved separately in water and then mixed. The formula was exposed in a book of recipes for patent medicines published by the Philadelphia College of Pharmacy in 1824, and remained official as Compound Effervescent Powders in the *United States Pharmacopeia*, and later in the *National Formulary* through its 12th edition (1965); also known as *Powder papers* (5).

Precipitates: Drugs prepared by separating particles from a previously clear liquid by physical or chemical means. Precipitation usually occurs when a hot saturated solution of an amorphous substance is allowed to cool or when a liquid in which the dissolved substance is insoluble is added to its solution. Pharmacists formerly used the process of precipitation as a convenient method of obtaining solid substances in fine particles (precipitated calcium carbonate), to purify solids (precipitated calcium phosphate), or to prepare mercury salts. White precipitate (ammoniated mercury) was first described by Beguin in 1632, a soluble double chloride of mercury and ammonium known to the alchemists as *sal alembroth* and *sal sapientiae*, respectively. Red precipitate (red mercuric oxide) was known to alchemists as *hydragyrum precipitatum per se* or “precipitate per se”; yellow precipitate is a synonym for yellow mercuric oxide (5, 24).

Preserves: See *Conserves*.

Protective dressings: See *Dressings, protective*.

Pulvis: See *Powders*.

Quiddonies: Preparations with a vehicle of thick, quince-flavored syrup (also see *Syrups*) (39).

Resina: See *Resins*.

Resins: Solid preparations consisting chiefly of the resinous principles from vegetable bodies. The officially prepared resins differ from alcoholic extracts in that the latter contain all of the alcohol-soluble principles in the drugs, whereas the resins contain only the alcohol-soluble principles that are insoluble in water. The term probably arose from the Greek *rheos*, “to flow,” referring perhaps to the flow of pine resin commonly observed (2, 5).

Rotula: Globules or orbicules (see *Pills*) (5).

Rubificients: Local remedies which produce redness and inflammation of the skin. The word comes from two Latin words: *ruber*, meaning “red,” and *facio*, meaning “to make” (2, 13).

Saccharures: Preparations made by saturating sucrose with a tincture, drying it, and grinding the mixture to a powder (5).

Saccelli amylacea: See *Wafer envelopes*.

Sacculi: Abbreviation for *sacculi medicinales*, or “bags of drugs” (20).

Sal: See *Salts*.

Sal effervescens: See *Salts, effervescent*.

Sal factitium: See *Salts, artificial*.

Salts: Compounds formed by the union of acids and bases, by the action of alkalies upon metals, or by the direct union of elements. The term is often incorporated in the common name of salts used as pharmaceuticals: bitter salts, epsom salt, or Seidlitz salt (magnesium sulfate), preparing salt (sodium stannate), Preston’s salts (ammonium chloride), Rochelle salt or Seignette’s salt (potassium and ammonium tartrate), salt of Mars (ferrous sulfate), salt of Saturn (lead acetate), salt of tartar (potassium carbonate), salt of tin (stannous chloride), salt of wisdom (mercury bichloride and ammonium chloride), sore-throat salt (fused potassium nitrate), vinegar salts (calcium acetate), and vomiting salt (zinc sulfate). The term is also applied to some acids, such as salt of lemon or sour salt (citric acid), salt of sorrel (oxalic acid), and spirit of salt (muriatic acid) (5, 13).

Salts, artificial: A mixture of the more important chemical salts naturally present in several of the well-known mineral springs of Europe. “These are properly labeled artificial,” notes *Remington’s Practice of Pharmacy* (1926), “and if used to prepare effervescent salts or mineral waters they should be sold only as an imitation of the genuine” (5).

Salts, effervescent: Granular effervescent salts were formerly made by mixing dry powders with dry tartaric acid and sodium bicarbonate and moistening the mixture with strong alcohol. The pasty mass was passed through a sieve, and the granules dried quickly in a hot room, sifted, and filled into bottles, which were hermetically sealed to prevent the access of moisture. This method was greatly

improved upon by mixing the powders in a flat enameled dish and heating in an oven to about 100°C or by heating the mixture in a deep jacketed kettle or in a pill-coating pan, heated, as it revolves, by a gas flame. When the mixture becomes moist, it is manipulated with a wooden spatula to make it uniform in consistency, and rubbed through a coarse tinned iron sieve; the granules obtained are dried slowly at a low heat in an oven (5).

Salts, smelling: Ammonia-based preparations used as a restorative in “hysterical syncope” (fainting). Dry smelling salts (or vinaigrettes) are composed of ammonium chloride and potassium carbonate, perfumed with lavender; liquid smelling salts are composed of ammonium carbonate dissolved in stronger ammonia water and alcohol, and perfumed with oils; solidified smelling salts are similar preparations solidified with stearic acid (5).

Salve pencils: See *Pencils, medicated*.

Salves: This term probably arose from the Anglo-Saxon *sealf* or the German *salbe*, both meaning “ointment” (see *Ointments*) (2).

Scutum: Abbreviation for *scutum stomachicum*, a large plaster, applied to the breast or stomach (20).

Semicupia: A half-bath, used interchangeably with *Insessia* (20).

Sericum oleatum: See *Silk, oiled*.

Serums: Serum therapy came into prominence in the 1890s when Emil von Behring (1854–1917) extended Pasteur’s theory of attenuated viruses. Behring demonstrated that the serum of animals immunized against attenuated diphtheria toxins can be used as a preventive or therapeutic inoculation against diphtheria in other animals through a specific neutralization of the toxin of the disease. In 1894, Behring began to produce his new antitoxic serum on a grand scale; it soon became recognized as the specific treatment for diphtheria. Antisera act by combining with the toxin in the blood of the patient, rendering it inert. Scarlet fever, tetanus, erysipelas, botulism, and gas gangrene have been successfully treated by antitoxic serums prepared in this manner. Antibacterial serum is produced by injection of an animal with successive doses of bacteria. The immune substances thus formed act by enhancing phagocytosis, destroying the bacteria. Pneumonia, streptococcic infection, and spinal meningitis have been aided by the use of this type of serum. Mixed serum

contains both antitoxic and antibacterial immune bodies; the serum used to treat scarlet fever is of this type (3, 40).

Shampoos: A wash for the hair or soap for hair washing. The term comes from the Hindustani word *tshampa*, meaning “to squeeze” or “to press,” probably associated with hot oriental baths (2).

Silk, oiled: A thin, very soft, and pliable protective dressing made of fine silk, coated with a flexible linseed-oil varnish. Oiled Silk is available in semitransparent or opaque form, made by the addition of talc or starch in the final coating. Oiled Muslin or Oiled Cambric is similar to oiled silk, except that the basic fabric is of glazed cotton; it is thicker and heavier and consequently less pliable than oiled silk (see *Dressings, protective*) (5).

Sinapismus: See *Plasters, blister*.

Sindon oleata: Oiled muslin or Oiled cambric (see *Silk, oiled*).

Snuffs: See *Insufflations*.

Soaps: A class of chemical substances which are metal salts of fatty acids. Pliny (first century A.D) records that the ancient Romans learned the preparation of soap from Nordic tribes, who used a pomade prepared from goat fat and the calcined ashes of beechwood. *Sapo*, the Latin word for soap, is derived from the Nordic *sepe*. The chemistry of soaps was elucidated in the early nineteenth century by French chemist M.E. Chevreul (1786–1889). Soaps may be divided into two classes: soluble soaps (or detergent or cleansing soaps), which are compounds of fatty acids with alkali metals, particularly sodium and potassium; and insoluble soaps, which are compounds of fatty acids and metals of any other group, such as Lead Oleate Plaster or Lime Liniment, a calcium soap of linseed oil. Soluble soaps include Hard Soap (or Castile Soap), prepared from olive oil and sodium hydroxide, official through the 11th edition of the *National Formulary* (1960), and Soft Soap (or Green Soap), prepared from linseed oil, glycerin, and dekanormal solutions of sodium and potassium hydroxide; the latter continues to hold official status (4, 5).

Soft capsules: See *Capsules, soft*.

Solutio: Dental preparations official only in the 5th edition of the *National Formulary* (1926), consisting of a solution of a resinous material dissolved in chloroform (3).

Solution tablets: See *Tablets, solution*.

Solutions: Liquid preparations that contain one or more substances dissolved in a solvent and, by reason of their ingredients or method of preparation, do not fall into some other category of preparation. From the seventeenth century on, the term *liquor* denoted liquid solutions. By the early twentieth century, liquor usually referred to aqueous solutions of nonvolatile substances. For example, Solution of Magnesium Citrate, *United States Pharmacopeia*, had the Latin title *Liquor Magnesii Citratis*. Beginning with the *United States Pharmacopeia XVI* and the *National Formulary XI* (1960), the term liquor was dropped and solutions were listed by their active ingredient or ingredients (19).

Solutions, irrigating: A preparation to be applied continuously by means of a special device for the purpose, as in the treatment of wounds with Dakin's solution (5).

Solutions, nasal: Solutions of drugs for instilling in the nose rather than spraying are generally a modern development. The first nasal solutions were formulated with menthol and thymol dissolved in light mineral oil. Later isotonic aqueous solutions were designed as drops (3).

Solutions, ophthalmic: Originally called collyria, ophthalmic solutions arose from the eye washes of the ancient world. Early preparations were not the sterile, buffered solutions of today. The term collyrium comes from the Greek *kollurion*, which Hippocrates used to designate fatty suppositories for gynecological purposes. They were formed into sticks, from which a paste was made with a liquid. Eventually, more liquid was added and the paste thinned to a lotion. This lotion was used as an eye wash, and term came to be used exclusively for this type of preparation. Remington's formula for a collyrium of 260 mg (4 grains) of sodium borate in 30 mL (1 oz) of camphor water was a standard from 1886 into the mid-twentieth century (2, 41).

Sovella: See *Tablets, solution*.

Spasmadraps: Pieces of linen or other cloth dipped in or spread with a medicinal plaster, popularized in 1514 by Giovanni da Vigo (1460–1525), physician to Pope Julius II. From the Latin *spasma* meaning “healing powder” and the French *drap* meaning “cloth.” Also known as sparadraps or cerecloths (see *Plasters*) (15).

Species: See *Teas*.

Spirits: Solutions of volatile substances (usually volatile oils) in alcohol. Some editions of the *National Formulary*

stated that “spirits of volatile oils” contained 6.5% of the volatile oil, but that figure was later rejected as too low. Although used internally, several spirits were used medicinally by inhalation or as flavorings (4, 31).

Spiritus: See *Spirits*.

Sprays: Medicated liquids prepared for dispersal by atomizers or nebulizers, usually on external surface or mucous membranes of the respiratory tract. In the *United States Pharmacopoeia*, sprays were called *inhalatio*; the *National Formulary* referred to them as *nebulae*. Sprays of the early twentieth century were formulated with aromatics dissolved in light mineral oil. As injuries from inhaled oils became apparent, especially among children, these sprays were displaced by buffered aqueous solutions (1, 31).

Starch capsules: See *Cachets*.

Steam: See *Vapors*.

Steatina: A salve mull (see *Mulls*) (5).

Stilus dissolubilis: Dissolving pencils (see *Pencils, medicated*). Pencils containing a caustic or an astringent (5).

Stilus medicatus: See *Pencils, medicated*.

Stilus unguentis: Salve pencils, cooling, antiseptic, or astringent (see *Pencils, medicated*) (5).

Stupa: See *Stypes*.

Stypes: Cloths wrung out of hot water and sprinkled with a counterirritant (2).

Styptic pencils: See *Pencils, medicated*.

Succus: See *Juices*.

Succus spissatus: Condensed juice (see *Extracts*).

Sugar plums: See *Condita*.

Suppositories: Conical or ovoid medicated solids intended for insertion into one of the several orifices of the body, excluding the mouth. Suppository use has been known as early as 2600 B.C., and was recommended in the works of Hippocrates (ca. 400 B.C.). The term derives from the Latin *suppositus*, meaning “to place under.” Premodern suppositories were made by hand using soap

or other semisolid fatty substances as the main vehicles. They were not commonly used until the seventeenth and eighteenth centuries and did not become popular until the mid-nineteenth century and the advent of cocoa butter as vehicle. In 1766, Antoine Baumé described a suppository mold which used liquefied cocoa butter. This technique was popularized in America by Alfred B. Taylor about 1852 using paper cones as molds. Metal molds were introduced about 1860, although many pharmacists continued to form suppositories by hand without heat. Cold compression of cocoa butter was made possible through the introduction of metal suppository presses about 1868, although the first popular compression mold was not introduced until 1879. After 1870, mixtures of glycerin and gelatin came to be used as vehicles for suppositories, beginning a quest for the perfect vehicle that continues to the present (2, 19, 42).

Suppositorium: See *Suppositories*.

Suspensio: See *Suspensions*.

Suspensions: Heterogeneous systems containing coarsely dispersed material that settles. A wide variety of pharmaceutical preparations have been used as suspensions, for example, White Lotion, Magma of Bismuth, and Compound Mixture of Opium and Glycyrrhiza (Brown Mixture). In addition, several official ointments are suspensions of solids in a semisolid base. A large number of suspensions are categorized as mixtures in the *United States Pharmacopeia* and the *National Formulary* (19).

Swabs: Ampuls containing Iodine Tincture, *United States Pharmacopeia*, covered with gauze or other absorbent material, and used for first-aid treatment. Iodine Swabs, later called Iodine Ampuls, were official in the *National Formulary* through its 13th edition (1970). When iodine tincture is required for first aid, the tip of the ampul is broken and the gauze absorbs the iodine and provides a means of applying it directly to the wound. Some Iodine Ampuls are in the form of fine capillary tubes, which are broken when needed and the tincture applied directly (24).

Sweetmeats: See *Condita*.

Syrups: A nearly saturated aqueous solution of sugar (usually sucrose) with or without medicinal or flavoring ingredients. Syrups are usually divided into flavored, containing a fruit or aromatic substance for a pleasant taste, and medicated, containing a drug. Simple Syrup,

United States Pharmacopeia was commonly used in the preparation of pill masses and other mixtures (3, 8, 19).

Syrupus: See *Syrups*.

Tabella: See *Tablets*.

Tablet Triturates: See *Triturates, tablet*.

Tablets: Dosage forms prepared by molding or compressing medicinal substances in dies. Tablets vary widely in shape, the most common form being discoid, and range from 0.06 to 0.60 g in weight. Jean de Renou applied the Latin word *tabella* to a special type of troche in 1608; Burroughs Wellcome & Company coined the term “tablet” in 1878 to refer to its brand of compressed pills; the term is derived from the French *tablette*, meaning “shelf” and the Latin *tabula*, meaning “board.” In 1843, the English apothecary William Brockedon patented a device for compressing medicinal agents commonly employed in pills and lozenges without the use of liquid adhesive agents; the resulting product was known as compressed pills. The Philadelphia druggist Jacob Dunton invented a similar device in 1864, marketing his own compressed pills in 1869; Joseph Remington devised a similar machine in 1875 to allow the retail druggist to “manufacture his own medication called for on prescription.” Each of these devices consisted of a compression cylinder and lower die (to hold the medicinal substance) as well as an upper die which was struck with a mallet to compress the material. More reliable compression was achieved by using the screw devices invented by Germany’s Professor Rosenthal (1874) and perfected by Austria’s Carl Engler (1907). Another advancement was the lever device introduced by Philadelphia’s Bennett L. Smedley (1879). The first rotary tablet machine was developed in 1872 by Henry Bower, an employee of the Philadelphia drug manufacturer John Wyeth; two years later, Joseph A. McFerran received a patent for the first fully automatic tablet machine (2, 4, 43).

Tabletta: See *Tablets*.

Tablets, compressed: See *Tablets*.

Tablets, dispensing: See *Tablets, solution*.

Tablets, enteric-coated: See *Enteric-coated doses*.

Tablets, hypodermic: Molded tablet triturates intended to be dissolved in water to make a solution to be injected parenterally. The usual weight of hypodermic tablets is about 0.03 g, which distinguishes them from

ordinary tablet triturates that weigh about 0.06 g. Formerly prepared extemporaneously by pharmacists, modern compressed hypodermic tablets are not intended to be sterile, although they are manufactured under strict conditions as a precaution against contamination (see *Triturates, tablet*) (4).

Tablets, molded: See *Triturates, tablet*.

Tablets, poison: Tablets of mercury bichloride in an angular, not discoid shape, blue in color, each having the word "POISON" and the skull-and-crossbones design distinctly stamped upon it. A unique one-product classification, poison tablets (or *Toxitebellae*) first became official in the *United States Pharmacopoeia IX* (1916), and two strengths remained official in the *National Formulary* through its 10th edition (1950); the larger tablets remained official through the 12th edition (1965). Diluted in a solution of 1:1000 concentration, mercury bichloride is an antiseptic used chiefly for the disinfection of inanimate objects and the unabraded skin (4, 5, 24).

Tablets, solution: Molded or compressed tablets containing large amounts of potent substances not intended for administration, but rather as a convenience in dispensing; also known as Dispensing Tablets. To lessen the risk of their being dispensed by mistake for other tablets, dispensing tablets are always of angular rather than discoid shape. They are usually scored to facilitate division into more or less accurate fractions. Also tablets to be dissolved in water for external use (4).

Teas: Coarsely powdered mixtures of dried herbs intended for medicinal teas or poultices; also known as Species. The *National Formulary* recognized an Emollient Species, used as a cataplasm; a Laxative Species (St. Germain Tea); and a Pectoral Species (Breast Tea) for a "catarrhal condition of the respiratory tract" through its 5th edition (1926). Similar teas from home-grown herbs persist as common household remedies. True tea from China was introduced to England by Christopher Borough in 1379. The English word probably comes from the Dutch *thee* (2, 5, 24).

Tinctura: See *Tinctures*.

Tinctures: Alcoholic or hydroalcoholic solutions of drugs, usually of plant origins. The term comes from the Latin *tingere*, "to dye or soak in color." Tinctures, as alcoholic solutions, entered medical practice in the thirteenth century through the efforts of Raymond Lull and Arnald of Villanova. Paracelsus (1493–1541) was a

strong advocate for tinctures; inasmuch as he was controversial, his advocacy probably discouraged their incorporation into compendia until the 1700s. In the 1800s, wine remained the prime hydroalcoholic vehicle in the United States. After the turn of the century, however, tinctures (and elixirs) displaced those wines because of their wide variation in strength. Moreover, prohibition impeded their widespread use. As galenicals declined throughout the twentieth century, tinctures lingered on as an official class, mainly as flavorings, for example, Tincture of Orange Peel. Homeopathic Tinctures are generally prepared by long maceration of freshly dried succulent plants or their parts in alcohol, the completed tincture being made to represent one part of the dry crude material in each ten parts of the completed preparation (2, 5, 19).

Tinctures, homeopathic: See *Tinctures*.

Tincture triturations, homeopathic: See *Triturations*.

Toothache drops: See *Drops, toothache*.

Toxitebella: See *Tablets, poison*.

Triturates, tablet: Small, disk-like masses of medicinal powders prepared by forcing a moistened tablet mass into a die by manual pressure and allowing them to dry and harden. The basis of tablet triturates is usually a mixture of lactose and sucrose in a 5:1 ratio, moistened with a volatile liquid such as alcohol. Tablet triturates were introduced in New York in 1878 by Dr. Robert W. Fuller as a palatable and convenient means of administering potent drugs by mouth. Fuller's original triturates consisted of triturations of metallic, mineral, and vegetable matter, mixed into a paste with alcohol or water, and molded into the desired shape. In 1882, Fuller described a perforated, hard rubber tablet triturate mold with a corresponding pegged plate, which was practically identical to those available today. Tablet triturates served the purposes of homeopathic physicians well and undoubtedly helped to further the use of homeopathic doses. Twentieth-century pharmacists also prepared hypodermic tablets (used for preparing hypodermic injections) utilizing the technique. Today, tablet triturates and hypodermic tablets are formed by compression and are termed molded tablets (4, 5, 43).

Trituratio: See *Triturations*.

Triturations: Dilutions of potent powdered drugs prepared by intimately mixing them with a suitable diluent,

usually lactose, in a definite proportion by weight, usually 10%, used as a dispensing aid. Such poisonous substances as strychnine sulfate, arsenic, mercury bichloride, and atropine are much more accurately dispensed using this technique, the pharmacist weighing a multiple of the prescribed drug in a triturated form. Homeopathic triturations were formerly prepared by triturating one part of a drug into 99 parts of lactose over a period of at least one hour; homeopathic tincture triturations were prepared by mixing 10 mL of a homeopathic "strong tincture" with 10 g of lactose and triturating the mixture gently until dry. Although such powdered triturations were being replaced by commercially prepared tablet triturates by the mid-1920s, a general formula for triturations, specifying geometric dilution, remained official in the *United States Pharmacopeia* through its 14th revision (1950) (5).

Triturations, homeopathic: See *Triturations*.

Trituraations, tincture: See *Triturations*.

Troches: Solid dosage forms in the form of small disks, cylinders, or tablets, intended to be placed in the mouth and allowed to dissolve or disintegrate slowly. The term is derived from the Greek *trochos*, meaning "round" or "circular." They were subsequently called *pastils* in French and lozenges in English. One of the earliest troches (500 B.C.) was *terra sigillata*, or sealed earth, a product composed of clay from the island of Lemnos and goat's blood, rolled and cut into disks and impressed with a seal; by the Middle Ages, a variety of troche presses were employed. In 1856, Edward Parrish described an apparatus for rolling and cutting troches consisting of a rolling-board, wooden roller, and cutting punch. During the next two decades, F.L. Slocum (1879), F.E. Harrison (1880), Wallace Procter (1894), and nearly a dozen others patented similar machines. The 4th edition of the *National Formulary* (1916) featured nine formulas for troches; by the mid-1930s, troches were being replaced by tablets: the 6th edition (1936) featured only Troches of Elm. Modern troches consist of powdered drugs bound with sugar and tragacanth or incorporated into a hard candy or glycerogelatin base (2–4).

Trochiscus: See *Troches*.

Unguenta extensa: See *Mulls*.

Unguentum: See *Ointments*.

Vapor siccus: See *Inhalations, dry*.

Vapors: Steam, plain or medicated, generated by the use of steam or boiling water or by the use of a specially constructed apparatus (5).

Vesicatories: Local remedies, the application of which produces a serous discharge beneath the skin, forming a blister. Also known as epispastics (13).

Vials: See *Ampuls*.

Vinaigrettes: See *Salts, smelling*.

Vinegars: Infusions or solutions of drugs in vinegar or acetic acid; one of the oldest methods of drug preparation. Ancients knew that vinegar was often a better solvent than water and a preservative as well. The first preparation listed in the first *United States Pharmacopeia* (1820) was *Acetum Opii*, or Vinegar of Opium. The English word derives from the French *vin*, "wine," and *aigre*, "sour" (2, 8, 19).

Vinum: See *Wines*.

Wafer capsules: See *Cachets*.

Wafer envelopes: Preformed envelopes of rice flour used to administer bitter or nauseating drugs. Developed by Johann Schmidt as *sacculi amylacei*, wafer envelopes marked an improvement of convenience over wafer sheets. Pharmacists often furnished empty wafer envelopes to their patients, who transferred doses to them from prepared powder papers (see *Wafers*) (5).

Wafers: Flat sheets of rice flour used to administer nauseating drugs. When dry, wafer sheets are nonadhesive, stiff, somewhat brittle, and slightly thicker than ordinary cardboard. Powders are administered by floating a piece of wafer sheet upon water until it becomes thoroughly softened, passing a tablespoon underneath and lifting it out, and depositing the powder in the center and folding over the corners to thoroughly enclose the powder. If water is poured into the spoon, the concealed powder can be swallowed without any disagreeable taste being perceived. Wafer sheets are made by pouring a mixture of rice flour and water upon hot greased plates or rolling it between two hot, polished, revolving cylinders (5).

Washes: Aqueous preparations designed to cleanse specific parts of the body. Examples include enemas, eye drops, mouth washes, and nasal washes. General washes made official include Alkaline Aromatic Solution N.F. and Antiseptic Solution N.F. (31).

Washes, eye: See *Solutions, ophthalmic*.

Washes, mouth: Hydroalcoholic solutions of soap flavored with essential oils for cleansing the oral cavity; they became first official in the *National Formulary V* (1926) (31).

Waters, aromatic: Saturated solutions usually of volatile oils or similar substances in distilled water. Aromatic waters such as rose water were used in antiquity. Distilled waters containing volatile oils reached their therapeutic peak in the early sixteenth century. Although their therapeutic use declined in modern times, they continued to be used as flavorings. Hamamelis water (witch hazel) has lingered on as an aftershave and astringent (19).

Wax, dental See *Drops, toothache*.

Waxed papers: See *Papers, waxed*.

Wines: Alcoholic liquids prepared from drugs by the process of solution, maceration, or percolation, differing from tinctures only in that wine is used as a solvent or menstruum instead of various strengths of alcohol; one of the oldest liquid preparations, since the alcoholic content of the wine improved its solvent characteristics in many cases. The *National Formulary IV* (1916) recognized 15 wines, but the passage of Prohibition convinced the revisors to drop red and white wine, as well as all medicated wines, from that compendium (3).

REFERENCES

1. Arny, H.V. *Principles of Pharmacy*; W. B. Saunders Company: Philadelphia, 1918.
2. Wain, H. *The Story Behind the Word*; Charles C. Thomas: Springfield IL, 1958.
3. Arny, H.V. *Principles of Pharmacy*; With the Collaboration of Robert P. Fischelis W.B. Saunders Company: Philadelphia, 1937.
4. *American Pharmacy: An Introduction to Pharmaceutical Technics and Dosage Forms*; Sprowls, J.B., Jr., Lyman, R.A., Eds.; J. B. Lippincott Company: Philadelphia, 1966.
5. Cook, E.F.; LaWall, C.H. *Remington's Practice of Pharmacy*; J. B. Lippincott Company: Philadelphia, 1926.
6. Bender, G.A. *Great Moments in Pharmacy: The Stories and Paintings in the Series A History of Pharmacy in Pictures by Parke, Davis & Company*; Northwood Institute Press: Detroit, 1966.
7. *Kremers Files for the History of Pharmacy [A Topical Reference Collection]*; School of Pharmacy, University of Wisconsin-Madison.
8. Wood, H.C.; Remington, J.; Sadtler, S. *The Dispensatory of the United States of America*; J. B. Lippincott Company: Philadelphia, 1886.
9. Urdang, G. The Invention of Gelatin Capsules. *Pharm. Arch.* **1943**, *14* (4), 58–59.
10. Alpers, W.C. Gelatine Capsules: 1. History of the Capsule; 2. Filling the Capsule with Powders or Pill Mass. *Proc. Am. Pharm. Assoc.* **1896**, *44*, 175–185.
11. Stadler, L.B. Innovations in Dosage Forms: The Gelatin Capsule. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1959**, *20* (12), 723–724.
12. DuMez, A.G. A Contribution to the History of the Development of the Enteric Capsule. *J. Am. Pharm. Assoc.* **1921**, *10* (5), 372–376.
13. Wood, G.B.; Bache, F. *The Dispensatory of the United States of America*; Grigg & Elliott: Philadelphia, 1836.
14. Wood, H.C.; Osol, A. *The Dispensatory of the United States of America*; J. B. Lippincott Company: Philadelphia, 1943.
15. Griffenhagen, G. The Lost Art of Plaster Spreading. *Am. Prof. Pharmacist* **1957**, *23* (2), 139–143.
16. Kremers, E.; Urdang, G. *History of Pharmacy: A Guide and a Survey*; J. B. Lippincott: Philadelphia, 1940.
17. Thompson, H.O.; Lee, C.O. History, Literature, and Theory of Enteric Coatings. *J. Am. Pharm. Assoc. (Sci. Ed.)* **1945**, *34* (5), 135–138.
18. Sonnedecker, G.; Griffenhagen, G. A History of Sugar Coated Pills and Tablets. I. Invention and Acceptance. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1957**, *18* (8), 486–488, II. Methods and Equipment. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)*, **1957**, *18* (9), (1957) 553–555.
19. *American Pharmacy: Fundamental Principles and Practices*; Lyman, R.A., Ed.; J. B. Lippincott Company: Philadelphia, 1945.
20. *Medical Lexicon: A Dictionary of Medical Science*; Dunglison, R., Ed.; Enlarged and Thoroughly Revised by Robley Dunglison Henry C. Lea: Philadelphia, 1874.
21. Harry, R.G. *Modern Cosmeticology: The Principles and Products of Modern Cosmetics*; Revised by J. B. Wilkinson in Cooperation with R. Clark, E. Green, and T. P. McLaughlin Chemical Publishing Company, Inc.: New York, 1962; I.
22. Lémery, N. *Pharmacopée Universelle*; P. Gossey: Paris, 1729.
23. Agnew, E. Medical Dragées and Granules. *Am. J. Pharm.* **1870**, *42* (3), 270–272.
24. *Remington's Practice of Pharmacy*; Martin, E.W., Cook, E.F., Leuallen, E.E., Osol, A., Tice, L.F., Van Meter, C.T., Eds.; Mack Publishing Company: Easton, PA, 1961.
25. Lloyd, J.U. *Elixirs and Flavoring Extracts: Their History, Formulae, and Methods of Preparation*; William Wood & Company: New York, 1892.
26. Higby, G.J. *One Hundred Years of the National Formulary*; American Institute of the History of Pharmacy: Madison, WI, 1989.
27. Taber, C.W. *Taber's Cyclopedic Medical Dictionary*; F. A. Davis Company: Philadelphia, 1961.
28. Brockbank, W. *Ancient Therapeutic Arts: The Fitzpatrick Lectures Delivered in 1950 and 1951 at the Royal College of Physicians*; William Heinemann Medical Books, Ltd. London, 1954.

29. Remington, J.P. *The Practice of Pharmacy*; J. B. Lippincott Company: Philadelphia, 1905.
30. Davis, A.B. The Development of Anesthesia. *Am. Sci.* **1982**, 70 (5), 522–528.
31. Cook, E.F.; Martin, E.W. *Remington's Practice of Pharmacy*; Mack Publishing Company: Easton, PA, 1951.
32. Griffenhagen, G.B. The History of Parenteral Medication. *Bull. Parenter. Drug Assoc.* **1962**, 16 (2), 12–19.
33. Groves, M.J. *Parenteral Products: The Preparation and Quality Control of Products for Injection*; William Heinemann Medical Books, Ltd. London, 1962.
34. Toilet Milks. *Chemist Druggist* **1937**, 127 (2998), 89.
35. Griffenhagen, G. Tools of the Apothecary: 4. Suppository Molds. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1956**, 17 (6), 402–403.
36. Griffenhagen, G. Tools of the Apothecary: 5. Pill Tiles and Spatulas. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1956**, 17 (7), 464–465.

FURTHER READING

- Adrian, L.A. *Étude historique sur les Extraits Pharmaceutiques cont. la Description des div. Procédés et Appareils...*; Paris, 1889.
- An Account of the Rise and Attempts of a Way to Convey Liquors Immediately into the Mass of Blood. *Philos. Transac.* **1665**, 1 (December 4).
- Ansel, H.C. *Introduction to Pharmaceutical Dosage Forms*; Lea & Febiger: Philadelphia, 1981.
- Barach, A.L. Symposium —Inhalation Therapy: Historical Background. *Anesthesiology* **1962**, 23 (4).
- Bartholow, R. *Manual of Hypodermic Medication*; Lippincott: Philadelphia, 1869.
- Boberg, E.J. *Gelatin Capsules*; University of Wisconsin, 1912.
- Boston Society for Medical Improvement. *Report of a Committee... on the Alleged Dangers Which Accompany the Inhalation of the Vapor of Sulphuric Ether*; David Clapp: Boston, 1861.
- Bringinghurst, F. On Pill Machines. *Proc. Am. Pharm. Assoc.* **1867**, 15.
- Brockbank, W. The Ancient Art of Enema Administration. *Ancient Therapeutic Arts: The Fitzpatrick Lectures Delivered in 1950 and 1951 at the Royal College of Physicians*; William Heinemann Medical Books, Ltd. London, 1954.
- Brockbank, W.; Corbett, O.R. De Graaf's 'Tractatus De Clysteribus.' *J. Hist. Med.* **1954**, 9 (2).
- Brockbank, W. The Less Ancient Art of Intravenous Injection of Drugs. In *Ancient Therapeutic Arts: The Fitzpatrick Lectures Delivered in 1950 and 1951 at the Royal College of Physicians*; William Heinemann Medical Books, Ltd.: London, 1954.
- Brockbank, W. *Ancient Therapeutic Arts: The Fitzpatrick Lectures Delivered in 1950 and 1951 at the Royal College of Physicians*; William Heinemann Medical Books, Ltd.: London, 1954.
- Brown, H.M. The Beginnings of Intravenous Medication. *Ann. Med. Hist.* **1917**, 1 (2).
- Buess, H. *Die Historischen Grundlagen der Intravenösen Injektion. Ein Beitrag zur Medizingeschichte der 17. Jahrhunderte*; H. R. Sauerländer & Co. Aarau Switzerland, 1946.

- Burrin, P.L. Innovations in Dosage Forms: Some Unique Dosage Forms. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1959**, 20 (12).
- Caujolle, F. Vegetable Essences in French Pharmacopoeia. *Indian Pharm.* **1954**, 10 (3).
- Collard, E. Il Y a Pommade Et Pommade. *Rev. d'Hist. Pharm.* **1962**, 50 (172).
- Crellin, J.K.; Scott, J.R. Pharmaceutical History and Its Sources in the Wellcome Collections: III. Fluid Medicines, Prescription Reform and Posology 1700–1900. *Med. Hist.* **1970**, 14 (2).
- Cyr, G.N. Innovations in Dosage Forms: Innovations in Percolation. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1959**, 20 (12).
- Diepgen, P. *Das Analzäpfchen in der Geschichte der Therapie*; Stuttgart, 1953.
- Dobler, F. Die Tinctura Bei Theophrastus Paracelsus. Experimentelle Überprüfung Seiner Haupttinkturen. Band 13. *Die Vorträge der Hauptversammlung der Internationalen Gesellschaft für Geschichte der Pharmazie*, 1958.
- Dragstedt, C.A. Intravenous Injections. *Am. Prof. Pharm.* **1948**, 14 (2).
- Dutton, W.F. *Intravenous Therapy*; F. A. Davis Co.: Philadelphia, 1924.
- Eisenberg, L. History of Inhalation Therapy Equipment. *Int. Anesthes. Clin.* **1966**, 4 (3).
- Ellis, E.T. A New Method of Making Suppositories. *Am. J. Pharm.* **1879**, 51 (4).
- England, J.W. An Historical Note on the Official Rosin Cerates. *J. Am. Pharm. Assoc.* **1933**, 22 (11).
- Evans, A.J.; Train, D.A. *A Bibliography of the Tableting of Medicinal Substances*; The Pharmaceutical Press: London, 1963.
- Evolution of the Hypodermic Syringe and Other Instruments for Parenteral Therapy. *Chem. Drug* **1953**, 159 (3824).
- Fairthorne, R.F. On the Preparation of Suppositories. *Am. J. Pharm.* **1871**, 43 (10).
- Feinstein, K. *Thoretische und praktische Untersuchungen ueber das Perkulationsverfahren nebst einem Ueberblick ueber dessen Entwicklung*; Zurich, 1936.
- First Rotary Tablet Press Traced to Wyeth, 1872. *Pulse of Pharmacy [Wyeth]* **1959**, 13 (1).
- Foote, P. *Tablets*; Bulletin of the University of Wisconsin, Serial No. 1566, Madison, 1928.
- Friedenwald, J.; Morrison, S. The History of the Enema With Some Notes on Related Procedures: Part I. *Bull. Hist. Med.* **1940**, 8 (1), 68–114, Part II, *ibid.*, 8 (2), **1940**.
- Furst, P.T.; Coe, M.D. Ritual Enemas. *Nat. Hist.* **1977**, 86 (3).
- Garrison, F.H. *An Introduction to the History of Medicine*; W. B. Saunders Company: Philadelphia, 1929.
- Goodman, H. The Hypodermic Syringe: The Early History With Notes on the Forgotten Anatomical Instrument. *Med. Times (N.Y.)* **1956**, 84 (6).
- Griffenhagen, G. Tools of the Apothecary: 6. Pill Machines. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1956**, 17 (8).
- Griffenhagen, G. A History and Evolution of the Suppository Mold. *Am. J. Pharm.* **1953**, 125 (3).
- Griffenhagen, G. Tools of the Apothecary: 10. Lozenges, Tablets, and Capsules. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1956**, 17 (12).
- Guitard, E.-H. L. 'emplâtre. *Rev. d'Hist. Pharm.* **1960**, 48 (167).

- Halsband, R. New Light on Lady Mary Wortley Montagu's Contribution to Inoculation. *J. Hist. Med.* **1953**, 8 (4).
- Hamarnah, S. At the Smithsonian—Exhibits on Pharmaceutical Dosage Forms. *J. Am. Pharm. Assoc.* **1962**, NS2 (8).
- Hamarnah, S. Early Arabic Pharmaceutical Instruments. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1960**, 21 (2).
- Harrop, J. Monograph on Fluid Extracts. *Solid Extracts and Oleoresins*; Columbus OH, 1895.
- Hehre, F.W. Early Events in the Development of Inhalation Therapy. *Conn. Med. J.* **1958**, 22 (11).
- Heizer, R.F. The Use of the Enema Among the Aboriginal American Indians. *Ciba Symp* **1944**, 5 (11).
- Helfand, W.H.; Cowen, D.L. Evolution of Pharmaceutical Oral Dosage Forms. *Pharm. Hist.* **1983**, 25 (1).
- Hoff, E.C.; Hoff, P.M. The Life and Times of Richard Lower, Physiologist and Physician, 1631–1691. *Bull. Inst. Hist. Med.* **1936**, 4 (7).
- Holland, M.O. Collyria. *Am. Prof. Pharmacist* **1938**, 4 (6).
- Howard, Jones N. A Critical Study of the Origins and Early Development of Hypodermic Medication. *J. Hist. Med.* **1947**, 2 (2).
- Jonas, A.S. The History of Plaster of Paris Bandages. *Surg. Gynec. Obstet.* **1956**, 102 (2).
- Kane, H.H. *The Hypodermic Injection of Morphia, Its History, Advantages and Dangers*; New York, 1880.
- Kebler, L.F. The Tablet Industry—Its Evolution and Present Status; the Composition of Tablets and Methods of Analysis. *J. Am. Pharm. Assoc.* **1914**, 3 (6).
- Kebler, L.F.; *ibid.* **1914**, 3 (7).
- Kebler, L.F.; *ibid.* **1914**, 3 (8).
- Kilmer, F.B. Manufacture of Medicinal Plasters. *Am. J. Pharm.* **1910**, 82 (9).
- Kirkby, W. The Story of the Pill from the Days of the Ebers Papyrus to Recent Times. *Chem. Drug.* **1939**, 130 (Special Issue).
- Kratzenstein, C.G. *Abhandlung von dem Nutzen der Electricität in der Arznuwissenschaft*; (facsimile, c. 1978); Halle, 1744–1745.
- Krüger, M. *Zur Geschichte der Elixiere, Essenzen und Tinkturen*; Bd. 10, *Veröffentlichungen aus dem Pharmaziegeschichtlichen Seminar der Technischen Hochschule*; Braunschweig, 1968.
- Krüger, M. *Zur Geschichte der Elixiere, Essenzen, und Tinkturen*; Bd. 10, *Veröffentlichungen aus dem Pharmaziegeschichtlichen Seminar...*; Braunschweig, 1968.
- Lee, C.O. Emulsions. University of Wisconsin: 1930.
- Leigh, J.M. The Evolution of Oxygen Therapy Apparatus. *Anaesthesia* **1974**, 29 (4).
- Lémery, N. *Pharmacopée Universelle*; P. Gossey: Paris, 1729.
- Lieberman, W. The History of the Enema. *Ciba Symp* **1944**, 5 (11).
- Lieberman, W. Notes and Literary References to the Clyster. *ibid.* **1944**, 5 (11).
- Lieberman, W. A Chronology of the Enema. *ibid.* **1944**, 5 (11), 1712.
- Lillico, J. The Use of Enemata by Primitive Peoples. *Ann. Med. Hist. (Third Series)* **1941**, 3 (1).
- Lloyd, J.U. *Elixirs, Their History, Formulae and Methods of Preparations, Including Practical Processes for Making the Popular Elixirs of the Present Day and Those Which Have Been Official in the Old Pharmacopoeias, with a Resume of Unofficial Elixirs from the Days of Paracelsus*; Cincinnati, 1883.
- Lloyd, J.U. The Chapman Suppository Mold. *Proc. Am. Pharm. Assoc.* **1902**, 50.
- Mattison, R.V. On Suppository Molds. *Proc. Am. Pharm. Assoc.* **1875**, 23.
- Manufacturing Industries of the Drug Trade: Plantin's American Medical Capsulery and Laboratory, 1836–1896. *Pharm. Era* **1896**, 16 (27).
- Mendelsohn, S. A Short History of Elixirs. *Manuf. Chem.* **1952**, 23 (12).
- Meredith, D.T. Innovations in Dosage Forms: Friable Pills and Escaps. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1959**, 20 (12).
- Moeller, H.W. Innovations in Dosage Forms: The Story of Lozenges. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1959**, 20 (12).
- Mohr, F.; Redwood, T.; Procter, W., Jr. *Practical Pharmacy: The Arrangements, Apparatus and Manipulations of the Pharmaceutical Shop and Laboratory*; Lea and Blanchard: Philadelphia, 1849.
- Morton, W.J. *The Invention of Anaesthetic Inhalation*; Appleton: New York, 1880.
- Nielsen, H. *Ancient Ophthalmological Agents*. A Pharmacohistorical Study of the Collyria and Seals for Collyria Used During Roman Antiquity, as well as of the Most Frequent Components of the Collyria. In *Acta Historica Scientiarum Naturalium et Medicinalium*; Odense University Press: Odense Denmark, 1974; 31.
- On a New Suppository Mold. *Am. J. Pharm.* **1867**, 49 (2).
- Paris, J.A. *Pharmacologia; or the History of Medicinal Substances*; F. & R. Lockwood: New York, 1822.
- Parrish, E. *Introduction to Practical Pharmacy*; Blanchard and Lea: Philadelphia, 1856.
- Parrish, E.; Bakes, W.C. Fancy and Fashion in Pharmacy: Suppositories. *Am. J. Pharm.* **1861**, 28 (1).
- Penn, H.P. The Geoffrey Kaye Museum Collection of Portable Ether Inhalers. *Anaesth. Intensive Care* **1975**, 3 (4).
- Petit, P. Vers Un Renouveau Des Préparations Magistrales Grâce à La Formation Continue. *Le Pharmacien De France* No. 12 (June, 1974), P. 569, Through. *Rev. d'Hist. Pharm.* **1975**, 63 (227).
- Randone, M.; Ponte, P.L. Un Metodo Di Cura Del '700: Le Pillole Di Mercurio Di Agostino Belloste. *Minerva Med* **1967**, 58 (77, Suppl.).
- Remington, J.P. On An Improved Pill Press. *Am. J. Pharm.* **1876**, 48 (3).
- Remington, J.P. *The Practice of Pharmacy: A Treatise on the Modes of Making and Dispensing Official, Unofficial, and Extemporaneous Preparations, with Descriptions of Their Properties, Uses, and Doses. Intended as a Hand-book for Pharmacists and Physicians and a Text-book for Students*; et seq J. B. Lippincott Company: Philadelphia, 1887.
- Richtmann, W.O. *Bibliography of Aromatic Waters*; Chronologic, 1809 to 1900: Milwaukee, 1902.
- Rodman, R.W. The Gilded Pill. *J. Am. Pharm. Assoc. (Pract. Pharm. Ed.)* **1943**, 4 (7).
- Runne, H. Die Entwicklung Der Praktischen Pharmazie in Den Letzten Vierhundert Jahren. *Arch. Pharm.* **1931**, 269.
- Rusek, V. Vyvoj Injekcni Formy Leciv. I. Ceskoslovenská Farmacie [Prague] **1960**, 9 (10), II. [on technology] **12** (1), (1963). (In Czech, with English and German abstracts.)
- Scheel, P.; Dieffenbach, J.F. *Die Transfusion des Blutes und die Einspritzung der Arzneien in die Adern. Historisch und in*

- Rucksicht auf die practische Heilkunde bearbeitet*; (Reprinted 1972) F. Brummer: Copenhagen, Berlin, 1802/1803–1828.
- Seel, E. Ueber Arzneitabletten. Pharm. Zenthalle **1906**, 47 (43).
- Siedler, P. sterile Ampuls. *Die chemischen Arzneimittel der letzten 113 Jahre...*; Bornträger: Berlin, 1914.
- Siedler, P. Gelatine Capsules. *Die chemischen Arzneimittel der letzten 113 Jahre...*; Berlin, 1914.
- Simpson, J.A.; Weiner, E.S.C. *The Oxford English Dictionary*; Clarendon Press: Oxford; Oxford University Press: New York, 1989.
- Sullivan, W.N. The Coupling of Science and Technology in the Early Development of the World War II Aerosol Bomb. *Mil. Med.* **1971**, 136 (2).
- Taylor, A.B. Suppositories. *Am. J. Pharm.* **1852**, 24 (3).
- The Means of Fumigating Infected Chambers, Etc. *The Family and Ship Medicine Chest Companion: Being a Compendium of Domestic Medicine ...*; Philadelphia, 1851.
- Thomas, K.B. The Early Use of Chloroform; with Some Notes on Certain Apparatus Designed for its Delivery. *Anaesthesia* **1971**, 26 (3).
- Thomas, K.B. An Ether Vaporiser of 1847. *Anaesthesia* **1968**, 23 (4).
- Thompson, C.J.S. The Pomander, a Link in the History of Preventive Medicine. *d'Hist. Med.* 2e Congres. International (also appeared as a separate, Evreux, 1922).
- Tui, C. Practical Aspects of Pyrogen Problems. *J. Am. Pharm. Assoc. (Pract. Pharm., Ed.)* **1944**, 5 (2).
- Urdang, G. Compressed Tablets [Historically Considered]. *What's New [Abbott]* **1943**, (76).
- Van Itallie, P.H. Innovations in Dosage Forms: Pioneers of Tablet-Making. *J. Am. Pharm. Assoc. (Pract. Pharm., Ed.)* **1959**, 20 (12).
- Van, I.; tallie, H. History of Syringes and Injection Therapy. *S. D. J. Med. Pharm.* **1965**, 18 (1).
- Weiser, H.B. Gels I: Gelatinous Precipitates and Jellies of Inorganic Substances and Gels II: Organic Jellies. In *A Textbook of Colloidal Chemistry*; John Wiley & Sons, Inc.: New York, 1949.
- Wood, J.R. *Tablet Manufacture, Its History, Pharmacy and Practice*; Lippincott: Philadelphia, 1906.
- Wright, S.P. On Suppositories. *Am. J. Pharm.* **1870**, 42 (3).
- Zekert, O.F. Zur Geschichte Der Extrakte. *Pharm. Monatsh.* **1923**, 4 (3).

HEALTH CARE SYSTEMS: WITHIN THE UNITED STATES

Henri R. Manasse, Jr.

American Society of Health-System Pharmacists, Bethesda, Maryland, U.S.A.

INTRODUCTION

A national health care system reflects the social, political, economic, and cultural character of a nation. A nation's historical roots and dominant values shape policies and directions for the organization, quality, financing, and access to health care services. These factors determine who gets what kind of care—at which locations, for what price, and paid by whom.

The distinctive historical antecedents of American cultural and social development have shaped the present health care system. These contexts have led to a health care system that is uniquely American in character and composition. Although the issues currently facing the American health care system bear some similarity to those in other developed, industrialized nations, many of the factors are unique to the United States.

Social values, that is, the collective societal beliefs about the nature of the human being and the structure of a society, play a strong role in the development of national policies. Political and economic decisions rest in large measure on the prevailing values held in a society. Hence, if a predominant social value rests on the notion that all societal members have a right to health care, political and economic policy developments will follow suit. One way of examining these contexts is to look at a spectrum of social values.

Donabedian (1) has proposed that such a spectrum might be considered from two polar positions: libertarianism versus egalitarianism. Dougherty (2) adds the dimensions of utilitarianism and contractarianism. The essence of these taxonomies of social values is that they characterize specific sets of beliefs and values held by a wide array of individuals.

Libertarian philosophical thought places major emphasis on personal achievement and freedom from political intervention. It holds that individuals should be free to exert their rational capacity to evaluate and determine what is good for them. They can then further act on these determinations for themselves from their own personal, fiscal, physical, and human resources. To this view, Dougherty (2) adds:

Because they can think, persons can understand their circumstances and the alternatives available to them. Because they can choose, persons can act to affirm or change their circumstances. Because they can think and choose, persons are free to create their own life plans and the values of which they are made.

It follows then, that predominant libertarian values are deeply entrenched in the notion of the self-made person and that social rewards should only accrue if they are deserved and earned. The role of government is, therefore, limited to those functions that absolutely do not abridge the rights of the individual to exert his or her own will for what he or she believes to be best. Moreover, government's role would be limited to those functions and needs for which individuals could not provide (national defense, negotiation of treaties, etc.).

Egalitarian principles focus on the equal moral standing of all individuals regardless of achievement or station in life. This philosophy also centers on the right to equal opportunity and to the extent possible, to be free from need and want. Thus, egalitarianism (2) can be viewed as follows:

Practically, this means an equal right to a reasonable share of those basic goods and services known to be necessary for a decent human life, including a right to a job, minimum income support, or provision in kind of the goods necessary for life, as well as a right to a range of social and health care services designed to prevent and minimize psychological and physical suffering, disabilities, and premature death.

Egalitarian values place specific demands on government and political policy to construct broad services and support systems so that all members of society are provided with equal opportunity designed to prevent and minimize psychological and physical suffering and disabilities, and to achieve one's life's aims. In this fashion, government would act on the entitlements due to all members of society. Such entitlements might be derived from legal or other forms of social consensus.

This range of social values from libertarianism to egalitarianism holds differing beliefs about equality,

justice, opportunity, rights, and the functional responsibilities of government. When this spectrum of social values is held over health, health care, and the administration and financing of health care services, it is not surprising that a vastly different array of designs emerge. Because health and health care are so tightly wound into personal, cultural, and social beliefs, it is not surprising that such a vast array of health care systems and notions about health have emerged across the world.

America's historical foundations have leaned strongly to the libertarian philosophical viewpoint (3). The influence of the "Protestant ethic" from Europe, coupled with the opportunities that a fresh land provided to "become one's own person," are strongly borne out in American society. An unbridled, free-market economy and freedom from governmental intervention in the daily lives of the citizenry are strong values that have been integrated into the American lifestyle and American political economic thought. The notion of "pulling yourself up by your bootstraps" succinctly reflects these dominant social, political, and economic values. Such antecedents are reliable markers for characterizing America's health care system. Consequently, it is a fascinating mosaic of pluralistic approaches. It is a strongly market-driven, industrialized system, which, at the same time, may be described as one of the world's best and one of the world's most troubled systems.

The United States does not have a universal health insurance program characteristic of many developed nations. Nor does it have national health care services like that of the United Kingdom and other nations. Except for those persons in the United States who possess special legal entitlements, the American health care system is largely a private enterprise; in other words, an industrialized system. The providers, payers, and institutions of care represent a rich mixture of private agents, corporations, insurance systems, and governmental agencies. There is not a singular rationalizing source for setting broad-based national policy and direction for the health care system as a whole. Rather, the vast market place of ideas has a variety of options in order to implement any proposal for which someone will pay. Relman has termed this approach "the industrialization of health care" (4).

The role of the national and state governments in the health care system is limited to those entitlement programs that have been legislated into federal or state law or where there is a federal and state partnership. Federal involvement in the provision of health care services began with the U.S. Public Health Service (PHS), an agency of the U.S. government. The PHS was established in 1798 to provide essential health care services to merchant marine

personnel and members of the U.S. armed forces. Subsequent federal involvement in the provision of and the payment for health care has incrementally increased to include care for individuals with special entitlements. The latter include veterans of the armed forces, the elderly, indigent people, Native Americans, persons with HIV/AIDS, certain disabled individuals, and qualifying persons with end-stage renal disease. For example, qualified veterans of the armed forces have access to a federal system of hospitals, clinics, and long-term care facilities under the Department of Veterans Affairs (a cabinet-level agency of the executive branch of the federal government). Since 1965, the federal government sponsors Medicare, a health insurance program for the elderly (65 years of age and over and later the disabled). The federal government also cost-shares with participating state governments to provide the Medicaid program (also enacted in 1965). The latter is an insurance program for health services directed toward qualifying indigent people. In Medicare and Medicaid, institutional and individual providers participate as contractors under a set of specific conditions for participation.

State and local (city and county) governments have limited roles in the provision of health care services. State, county, and city health departments are as differently organized and functioning, as there are states, counties, and cities in the United States. These agencies reflect and represent the special needs of the geographic areas and demographic compositions of their respective domains. Hence, the functioning and expanse of services offered by the New York City Department of Health is vastly different from a similar agency in rural Montana.

This unique approach to the application of a health care system must also be examined in light of the diversity of the demography and geography of the United States. Approximately 273 million people inhabit the United States across a geographic expanse of 3.5 million square miles of land. Ranging from the deserts of Nevada to the Rocky Mountains of Colorado and Wyoming to the tropics of Florida and the oceanic seaboards of the east, west, and southern coasts, American geography and topography is expansive (5). Hence, a substantial challenge to the delivery of health care services exists in this array of geographical areas.

The American population is equally diverse and expansive. There are almost 35 million people who are age 65 or older. African Americans constitute 12.8% of the population, Asian and Pacific Islanders 4%, American Indians 0.9%, and Caucasians make up 82% of the population (5). Because the United States is largely a nation of immigrants, there are literally hundreds of additional ethnic groups that are part of the American

social fabric. As of March 1997, 25.8 million individuals in the United States were foreign-born, which represents a 30% increase from 1990, when there were 19.8 million foreign-born individuals in the United States. Mexico was the place of origin for 7 million or 28% of the total foreign born population in 1997 (5). During 1996 and 1997, 1.3 million people moved to the United States from abroad, and 92% of those individuals moved to metropolitan areas. Additionally, during this time period, 3 million people left the central cities and 2.8 million moved to the suburbs (6). The health care system of the United States should then be viewed in the following context:

- A diverse spectrum of social values, which have historically pointed more toward libertarianism than egalitarianism
- Limited roles of the federal, state, and local governments in the provision of, and payment for, health care services
- A pluralistic, free-market approach to the provision of health care services
- A geographically diverse and substantive land mass
- A culturally diverse and numerically large population whose characteristics are changing toward more elderly and racial and ethnic heterogeneity

It is critical that the reader be sensitive to these contextual variables to understand the American health care system and how health care policy is shaped and implemented in the United States.

THE ORGANIZATION OF U.S. HEALTH CARE SERVICES

Health care services in the United States are provided by a broad array of facilities, which are financed from a variety of payment sources. As of 1998, there were 6021 hospitals (7), 1,012,582 hospital beds, 33,765,940 admissions, and 241,574,380 inpatient days. In 1998, the average length of stay in community hospitals was 6 days, whereas it was 7.7 days in 1975 (7).

It is also notable that the numbers of hospitals in urban and rural settings are shrinking. In 1993, there were 3012 urban hospitals and 2249 rural, whereas in 1998, there were 2816 urban and 2199 rural hospitals (13). The numbers of public acute care hospitals decreased from 1390 in 1993 to 1260 in 1997 (8). Closure of hospitals and simultaneous reductions in hospital beds has occurred in inner city areas where care is provided for large numbers of indigent patients. Such closures are related to the high costs of care, which are not concomitantly reimbursed by state and federal sources

either because the individuals are not eligible or because payment rates do not cover the costs incurred. Small, isolated rural hospitals are facing similar economic and, hence, survival difficulties. The plight of rural hospitals is of special significance because their survival is often linked to the economic and social survival of a rural community.

While the world's population grows at an annual rate of 1.7%, the population over 65 increases by 2.5% per year. There are just fewer than 600 million people over the age of 60 in the world. Approximately 360 million of the world's over 60 population lives in the developing world, in which 7.5% of the population is elderly. In contrast, 18.3% of the population is elderly in the developed world. The most rapid changes are occurring in some developing countries where an increase of 200–400% in the elderly population is predicted over the next 30 years (9). Because of the growth of the elderly population, there has been an increase in the demand for geriatric and long-term care facilities. Over the next several decades, the elderly's health care consumption in the United States will be approximately \$25,000 per person (in 1995 dollars) compared to \$9200 in 1995 (10). In this respect, the United States is following the trends exhibited in most developed industrialized countries.

The increased utilization of health care services by the elderly is expected to put additional strains on an already besieged health care system. Increasing the life span, either through preventive measures or through other acts of distributive justice, solves some problems while creating others. This astounding paradox will assuredly complicate the political and social processes of decision making. Equally likely will be the burdens these phenomena add to an already overburdened national economy.

In the last several years, it is the substitutability that has been emphasized, as more and more procedures are performed in outpatient settings. Many services previously performed in the hospital now take place in physician offices. In 1996, there were 734,493,000 visits to the physician, with an average of 3.4 per person (11), and the most frequent principal reason for a visit was a general medical examination, with a total of 54.7 million in 1996. Also in 1996, there were 67.2 million visits to outpatient departments, and 40.3 million inpatient surgery procedures were performed (11). This analysis points to the increasing importance of the ambulatory care setting as a place for rendering care. The relevance of outpatient care will continue to grow as more medical procedures are performed outside hospitals and greater emphasis is placed on preventive care. Outpatient visits in community hospitals alone have advanced from 263,631,000 in 1986 to 301,329,000 in 1990 to 474,193,000 in 1998 (7).

The National Association of Home Care estimates that more than 20,000 providers deliver home care services to approximately 8 million individuals each year (12). According to the Health Care Financing Administration (HCFA), the average number of home health visits a year per Medicaid beneficiary was 80, compared to 27 visits in 1989. Additionally, the number of home health agencies participating in Medicare has increased from almost 5000 in 1988 to over 10,000 in 1997 (13). Care of patients in home settings is likely to expand as data further suggest reduced cost for such care without compromising quality. Technological and scientific developments related to providing sophisticated treatments in the home will also stimulate growth in this sector of health services.

TRENDS IN HEALTH INSURANCE COVERAGE

According to the President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry, there are five trends that summarize the characteristics of health insurance plans of the late 1990s:

- Increased complexity and concentration of health plans
- Increased diversity of health insurance products
- Increased focus on network-based delivery
- Shifting financial structures and incentives between purchasers, health plans, and providers
- The development of clinical infrastructure for utilization management and quality improvement (14)

In response to rapidly increasing health care costs, private insurance companies and employers (who pay the premiums in whole or in part for their employees) have increased their part in implementing cost-containment strategies. A dramatic effort has been the application of business principles to purchasing and vendor selection and payment for and selection of health care providers and institutions of care.

Private employers, the federal government, and state and local governments invest significant financial resources in health care purchasing expenditures. In 1995, private employers contributed \$183.8 billion to private health insurance premiums, whereas the federal government spent \$11.3 billion on private health insurance premiums, and state and local government spent \$47.1 billion (14). In 1995, more than 83% of the insured population was covered by private insurance, whereas about 31% was enrolled in a public program, such as Medicare or Medicaid.

Probably the most significant change in the American health care system in recent years is the

development of managed care. In managed care settings, the covering company is responsible for providing services, whereas, at the same time, it is exposed to the financial risks of unanticipated services. Health Maintenance Organizations (HMOs) contract with hospitals and certain physician providers for services within a negotiated schedule of fees. HMOs and other such managed care organizations specify where and by whom care is to be given. The latter is a radical departure from the historically preeminent "freedom of choice" that patients and care providers enjoyed under the traditional indemnity and fee-for-service reimbursement programs. The traditional method of paying for medical services is fee-for-service when the provider charges a fee for each service provided, and the insurer pays all or part of that fee.

Managed care is an umbrella term for HMOs and all health plans that provide health care in return for preset monthly payments and coordinate care in a defined network of primary care physicians and hospitals. A network includes physicians, clinics, health centers, medical group practices, hospitals, and other providers that a health plan selects and contracts with to care for its members. An HMO is an organization that provides health care in return for preset monthly payments. Most HMOs provide care through a network of physicians, hospitals, and other medical professionals that their members must use in order to be covered for that care.

There are a number of different types of HMOs. A staff model HMO is an HMO in which the physicians and other medical professionals are salaried employees, and the clinics or health centers in which they practice are owned by the HMO. A group model HMO is made up of one or more physician group practices that are not owned by the HMO but operate as independent partnerships or professional corporations. The HMO pays the groups at a negotiated rate, and each group is responsible for paying its doctors and other staff as well as covering the cost of hospital care or care from outside specialists. An Independent Practice Association (IPA) generally includes large numbers of individual private practice physicians who are paid either a fee or a fixed amount per patient to take care of the IPA's members. A Preferred Provider Organization is a network of doctors and hospitals that provides care at a lower cost than through traditional insurance. PPO members have more health coverage when they use the PPO's network and pay higher out-of-pocket costs when they receive care outside the PPO network (15).

An integrated health system is a network that provides a coordinated continuum of services and is clinically and fiscally accountable for outcomes. There was a significant

growth of integrated health systems during the late 1990s. In 1997, there were 228 integrated systems and, in 1998, there were 266, representing an increase of almost 17% (16). Simultaneously, there has been a disintegration of systems when mergers fail and disassemble. Iglehart comments on how managed care has changed the face of health care:

Before the emergence of managed care, it was largely physicians, acting individually on behalf of their patients, who decided how most health care dollars were spent. They billed for their services, and third-party insurers usually reimbursed them without asking any questions, because the ultimate payers—employers—demanded no greater accounting. Now, many employers have changed from passive payers to aggressive purchasers and are exerting more influence on payment rates, on where patients are cared for, and on the content of care. Through selective contracting with physicians, stringent review of the use of services, practice protocols, and payment on a fixed, per capita basis, managed-care plans have pressured doctors to furnish fewer services and to improve the coordination and management of care, thereby altering the way in which many physicians treat patients. In striving to balance the conflicts that arise in caring for patients within these constraints, physicians have become “double agents.” The ideological tie that long linked many physicians and private executives—a belief in capitalism and free enterprise—has been weakened by the aggressive intervention of business into the practice of medicine through managed care (17).

There has been a recent challenge to the core tenet of managed care that centralized decision making could deliver improved care at a reduced cost. In November 1999, a large health care company decided to allow physicians to choose what care patients need without the insurance company’s intervention or approval. This action opens the door to further discussions about how managed care principles are utilized. Regardless of managed care’s future course, cost containment measures will be necessary to prevent an explosion of health care costs. The demand for cost containment will need to be weighed against the imperative to insure that patients have access to care. Paul Ellwood, often referred to as the “father of the HMO,” believes that there will be a new era in which patients, not employers and government purchasers, will have power (18). Regardless, the weight of political and consumer pressures, along with experience and economic efficiency, will determine the future of managed care.

HEALTH CARE FINANCING

The expenditures for health care in the United States have grown from \$51 billion in 1967 (6.3% of GNP) to over \$1 trillion in 1997 (14% of GDP).^a In 1997, on a per capita basis, \$4090 was spent on health care (19) and 0.64 per day/capita was spent on prescription drugs (20). This is substantially higher than that of other industrialized nations. When comparing health expenditures in the major industrialized countries comprising the Organization for Economic Cooperation and Development (OECD), for example, dramatic differences in per capita expenditures are noted (21). Such differences also exist in the percentage share of GDP spent on health care (21), and relative growth in health care expenditures over time varies greatly among these countries (21, 22).

The Health Care Finance Administration asserts that national health expenditures are projected to total \$2.2 trillion and reach 16.2% of the GDP by 2008. The growth in health spending is projected to average 1.8 percentage points above the growth rate of the GDP for 1998–2008. This differential is higher than recent experience but remains below the historical average for 1960–1997, where growth in health spending exceeded growth in GDP by close to three percentage points. There are a number of factors that contribute to the projected acceleration, including:

- An increase in private health insurance underwriting cycle
- A slower growth in managed care enrollment
- A movement towards less restrictive forms of managed care
- A continued trend toward increased state and federal regulation of health plans

The growth of health care expenditures without a concomitant gain in health status of the population is receiving more and more attention on the governmental and corporate agenda. On the governmental level, an increasing proportion of federal and state budgets is being allocated to health care. In the private sector, corporations and individuals are bearing larger proportions of health care costs. Although no particular percentage of GDP has been determined to be an acceptable or unacceptable expenditure for health care services, the fact is that costs are increasing and the health care sector is gaining an increasing share of the economy. This follows several other interesting trends. During the period of 1961 to 1997,

^aThe GNP is the total annual flow of goods and services in a nation’s economy. Most industrial countries now use GDP, which measures the value of all goods and services produced within a nation, regardless of the nationality of the procedure.

national health expenditures as a percentage of GNP rose from 5.4% to over 14%. In the same period, dramatic differences occurred in the source of revenues for health care expenditures. The pattern of spending these resources also changed significantly (13).

In 1960, 49% of health care revenues came from out-of-pocket payments from individuals. Out-of-pocket spending is defined as expenditures for coinsurance and deductibles required by insurers, as well as direct payments for services, which are covered by a third party. In 1990, individual consumers spent \$144.4 billion directly for out-of-pocket payments for personal health services (23). This accounted for 38% of all personal health spending. In 1998, consumers spent \$183.7 billion in out-of-pocket payments, which accounts for 33% of the \$558.7 billion in personal health spending (23).

Consumers have spent and continue to spend less of their own personal money for health care services. This decrease in personal spending has been shifted largely to third parties, such as private health insurance, government programs, philanthropic organizations, and other sources. It is evident that the shift away from personal, out-of-pocket health spending has resulted in greater consumption of health care services. This transition reflects the general maxim in health care economics that the consumption of health care services is probably insatiable (24). Moreover, unlike other sectors of the economy and the laws of economics they obey, prices for health care services do not fall with increased consumption or purchasing.

According to Iglehart, the decline in personal spending is "attributed in large part to the growth in health maintenance organizations (HMOs), which traditionally offer broad benefits with only modest out-of-pocket payments. In the past few years, however, most HMO enrollees have had increased cost-sharing requirements, as employers and health plan managers have sought to constrain spending even further. Out-of-pocket payments are still considerably less in an HMO than with indemnity insurance (17)." However, "The overall declines in per capita out-of-pocket spending mask the financial difficulties of many poor people and families. A recent study estimated that Medicare beneficiaries over 65 years of age with incomes below the federal poverty level (in 1997 the level was \$7755 for individuals and \$9780 for couples) who were also eligible for Medicaid assistance still spent 35% of their incomes on out-of-pocket health care costs. Medicare beneficiaries with incomes below the federal poverty level who did not receive Medicaid assistance spent, on average, half their incomes on out-of-pocket health care costs (17)."

Historically, a lack of public insurance programs created obstacles to health care services. For those who could not

afford to pay for private insurance, the costs associated with health care were larger than most could afford. After lengthy debate, the U.S. Congress passed legislation in 1965 that established Medicare and Medicaid. Medicare covers over 95% of the elderly in the United States as well as many individuals who are disabled. Coverage for the disabled began in 1973 and is divided in two parts: 1) hospital insurance and 2) supplementary medical insurance. The total disbursement for Medicare in 1997 was \$213.575 billion, and there were 36,460,143 enrollees, of which 32,164,416 were elderly.

The total expenditure for the Medicaid program was \$160 billion in 1996. Of the total amount spent in 1996, Medicaid payments for nursing facilities and home health care totaled \$40.5 billion for more than 3.6 million recipients. The average cost per recipient in 1996 was \$12,300, and almost 45% of the total cost of care for individuals using nursing homes and Medicaid was paid for home health care (13).

Since the enactment of Medicare and Medicaid, there have been various legislative and administrative changes. The Balanced Budget Act of 1997 enacted the most significant changes to Medicare and Medicaid since its inception, including a capped allocation of monetary resources to states and the addition of the Children's Health Insurance Program. The Children's Health Insurance Program set aside \$24 billion over 5 years for states to provide health care to over 10 million children who are not eligible for Medicaid.

In 1960, public programs paid for one quarter (24.5%) of all health care spending; by 1988, this share had increased to 42.1%. Together Medicare and Medicaid financed \$351 billion in health care services in 1996, which is more than one-third of the nation's total health care bill. Additionally, it represents three-quarters of all public spending on health care. There has been a significant increase in Medicare managed care enrollment—from 3.1 million at the end of 1995 to 6.3 million in 1999, leaving approximately 33 million beneficiaries in a traditional fee-for-service Medicare program.

An area of controversy is the limitation on coverage for prescription drugs. Spending on prescription drugs is the fastest-growing piece of personal health expenditures, amounting to \$78.9 billion in 1997. Additionally, spending for prescription drugs has increased at double-digit rates: 10.6% in 1995, 13.2% in 1996, and 14.1% in 1997 (17). The reason for this rapid growth, according to Iglehart, includes: "Broader insurance coverage of prescription drugs, growth in the number of drugs dispensed, more approvals of expensive new drugs by the Food and Drug Administration, and direct advertising of pharmaceutical products to consumers. The use of some new drugs reduces

hospital costs, but not enough to offset the increase in expenditures for drugs (17).” In the year 2000, 86% of health care plans will have an annual limit on brand and generic drugs, and there will be increased use of copayments for prescription drugs (25).

The budget cuts imposed by Congress in 1997 to help balance the budget have restricted the fees that caregivers receive for the elderly and disabled. When federal health programs cut funding significantly, as occurred in the Balanced Budget Act of 1997, the resulting cutbacks at the institutional and health-system level trickled down to providers’ abilities to provide an acceptable level of service designed to protect patient safety and foster appropriate medication use. Partial restoration of the Balanced Budget Act in 1999 addressed the transition to an outpatient prospective payment system for hospitals, payments to skilled nursing facilities and home health agencies, payments for indirect medical education, and a number of rural health care provisions.

The dramatic shift of third parties (government, private health insurance) toward paying for a greater and greater proportion of personal health care services has led to a paradigm shift in attitudes and actions toward health care financing and cost control. Several approaches have been adopted in the governmental sector to slow the increases in costs and expenditures. The most dramatic of these has been the introduction in 1983 of the prospective payment system (PPS) to curb the growth in hospital costs and expenditures. By imposing prospective limits on Medicare payments to hospitals through a system of reimbursing average costs of specific diagnoses, hospital utilization has decreased dramatically. The average length of stay and admission rates in community hospitals of elderly patients (those covered by Medicare) dropped sharply after the introduction of PPS (13).

Because of cost-containment strategies of both the private and governmental sectors, hospital utilization has declined. This has resulted in a decline in the number of patient beds, the average length of stay, and patient bed census (7). The present predominant view is that hospitalization of any patient, regardless of revenue source, is to be avoided wherever possible. Only those patients for whom hospitalization can be fully justified are admitted.

As much as the financing of America’s health care system is a major issue on the policy agenda of the nation, so too is the continuous question about the relationship between the costs and the outcomes of care. As costs increase, the numbers of policy analysts, organizations, and governmental agencies calling for a better definition of the cost-outcome relationship has sharply risen.

Cost-effectiveness and cost-benefit analyses are frequently mentioned in academic and policy-analysis

circles. These notions center on careful examination of the costs and their corresponding outputs. Eisenberg (26) defines cost-effectiveness analysis as the measure of the net cost of providing service (expenditures minus savings) as well as the results obtained (e.g., clinical results measured singly or a series of results measured on some scale). Cost–benefit analysis determines whether the cost is worth the benefits by measuring both in the same units (26). Such analyses will be critical, as future policy decisions are made with regard to the collection, allocation, and utilization of finite resources in the health care system for the enhancement of health status of the American people.

Private-sector strategies and governmental plans to curb health care costs have not escaped criticism. Ginsberg, for example, argues that the notion of “for profit” hospital chains has severe limitations with respect to garnering large proportions of market share and, consequently, greater profits (27). He bases this view on the limited amount of private funding available for hospital care. On the other hand, he sees this sector as being able to grow in the area of nursing homes and other businesses related to the care of the elderly.

ACCESS TO HEALTH CARE SERVICES IN THE UNITED STATES

There are three classes of individuals who have open access to and can derive some form of services from America’s health care system:

- Those who receive support from governmental sources because of specific entitlements (indigents, elderly, and veterans)
- Those who are provided with basic health insurance coverage from their employers
- Those who choose to cover their expenses from out-of-pocket payments

There are, however, those who have no specific financial support or capacity to pay for health care services and who are not eligible for any type of entitlements. These individuals must rely on some form of charity care or services. In addition, there are those who, for reasons of geographic remoteness or total inability to gain access, have no access to health care services. This group represents a complex, resource-based demand model, which also has an equally complex pattern of health care system and services-utilization requirements.

With increasing health care costs and consequent increases in insurance premium costs, gaining access to

health care services without incurring personal costs has become more difficult. Not all services are covered for individuals in the federal Medicare and Medicaid programs. Moreover, there are strict limitations on the extent of services offered in these programs. A similar set of restrictions may be found in private-sector health care coverage strategies. Because few insurance programs and none of the federal programs provide coverage for unlimited long-term care, all but the very rich are at risk of financial ruin.

The health care lexicon includes two new terms to reflect these problems: underinsured and uninsured. The underinsured may include the “working poor,” those individuals who have jobs and may be covered by a very limited, if any, health insurance program by their employers. They are likely low wage earners and those receiving incomes at, or slightly above, the poverty level. Typically, they do not qualify for Medicaid entitlements, do not have employer-paid health insurance benefits, and cannot afford (or choose not to purchase) third-party coverage for payment of health care services.

There are no specific policy plans available to finance uninsured and underinsured care. Whether planned as charity care or unplanned as financial loss, the “price tag” for uncompensated care in the United States was \$18.5 billion in 1997, which is 6% of the total of hospital expenses (28). This percentage has remained constant since 1984, when the percentage of total expenses for uncompensated care was also 6% (8).

Reduced payments and high levels of uncompensated care have led to the closing of hospital facilities in both urban and rural blighted areas, making access to care even more difficult for some. Whiteis and Salmon (29) refer to this phenomenon as “disinvestment in the public goods.” Because privately owned and not-for-profit hospitals and private clinics, pharmacies, and physician’s offices must rely on their own financial soundness, any threat to that foundation may lead to closure.

The amount of uncompensated care is magnified in areas where serious social problems exist because health status is directly related to social status. Health status should be examined in broad terms by reviewing morbidity and mortality data available for the whole population. The life expectancy of people who live in the United States has grown by almost 10 years, from 68.5 years in 1936 to 76.1 years in 1996. Women were expected to live to 79.1 years in 1996, whereas the average for men was 73.1 years (11). The leading causes of death in 1996 among people living in the United States were (11):

1. Heart disease (733,361 deaths)
2. Cancer (539,533 deaths)

3. Stroke (169,942 deaths)
4. Pulmonary diseases (108,027 deaths)
5. Accidents (94,948 deaths)
6. Pneumonia and flu (63,727 deaths)
7. Diabetes (61,787 deaths)
8. AIDS (31,130 deaths)
9. Suicide (30,903 deaths)
10. Liver disease (25,047 deaths)

Infant mortality, another measure of the health status of a nation, stated as the number of deaths per live births, was 7.2 per 1000 live births in 1997 compared to 9.9 per 1000 live births for 1988. Overall, these figures are comparable to those of the major, industrialized nations of the world.

Major morbidity in the United States is currently centered on diseases of life style. These morbidities contrast sharply with disease patterns prevalent during the early part of the 20th century. Outside of AIDS and other sexually transmitted diseases, infectious diseases represent a small proportion of prevalent morbidity. Rather, life-style diseases, associated with smoking, poor nutrition, a sedentary life style, alcohol and other chemical consumption, homicides, suicides, and accidents, represent the majority of morbidity in the United States. Significant preventive strategies can markedly reduce the incidence, prevalence, and mortality associated with these health care problems.

Not surprising, in areas with high concentrations of indigent people, there are similarly high concentrations of uninsured individuals requiring intense health care services. These areas exist in both rural and urban settings. Emergency rooms have become a major resource for primary health care services in areas where physician office services or other service providers (clinics) are not available because of location, cost, or quality. Emergency rooms have also become providers of high-intensity care for victims of gun shot wounds, drug overdoses, communicable diseases, and other trauma associated with poor social conditions. Much of the care in emergency rooms is uncompensated because the quality and amount exceed the allowable reimbursement. Some trauma centers in economically blighted areas have been closed (30).

Hospitals in inner cities and blighted rural areas also care for a higher proportion of “at-risk” patients than hospitals in the for-profit sector generally located in more affluent areas (29). In fact, affluent hospitals sometimes “dump” their uncovered patients on charity care and other public hospitals in order to reduce their financial risks. This, however, increases the financial risks of public or charity hospitals. Again, the reimbursement levels under present schemes for large numbers of “at-risk” patients simply do not cover costs; thus, the United States has

witnessed hospital closings, particularly in those areas where such loss is most noticeable (30).

American health policy continues to grapple with these issues related to the underinsured and the uninsured (31). A multiple-tiered health care system based on social class and ability to pay is unacceptable in a nation that boasts incomparable riches and political agendas of democracy and rights. Ginsberg (27) notes:

Despite all our efforts of recent years, then, health care costs continue to increase.... There is undoubtedly waste in the health care system, but no solid proposals have been advanced to recapture the \$100 billion, plus or minus, that some believe can be saved. I believe that we will not reshape our national health policy agenda unless and until we achieve a broad consensus on the key issues. Do the American people, for example, desire to ensure access to health care for the entire population? In that case they must agree to pick up a sizable additional tab, which they have thus far avoided.

The issue of quality health care has become an increasing issue of concern in the face of cost constraints and limited access to health care. The *President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry* (32) states that "the purpose of the health care system must be to continuously reduce the impact and burden of illness, injury and disability and to improve the health and functioning of the people of the U.S." According to the Commission, there are basic characteristics of health care that, as a nation, we should strive to achieve. The Commission has created "Guiding Principles for the Consumer Bill of Rights and Responsibilities" for the health care of people in the United States. These include the following:

- All consumers are created equal.
- Quality comes first.
- Preserve what works.
- Costs matter.

THE FUTURE OF HEALTH CARE

Suggestions for broad reform, which address the financial, access, and quality of care issues for America's health care system, have emerged during the past decade. Iglehart emphasizes the irony of the American health care system. He writes (17):

By many technical standards, U.S. medical care is the best in the world, but leaders in the field declared

recently at a national round table that there is an "urgent need to improve health care quality." The stringency of managed care and a low inflation rate have slowed the growth of medical spending appreciably, but a new government study projects that health care expenditures will soon begin escalating again and will double over the next decade. In short, the American system is a work in progress, driven by a disparate array of interests with two goals that are often in conflict: providing health care to the sick, and generating income for the persons and organizations that assume the financial risk.

The President's Commission (32) outlines areas in which the American health care system could be improved in light of the reality that many individuals receive substandard care and 44.3 million individuals are without health insurance coverage. This commission outlines several types of quality problems including avoidable errors, underutilization of services, overuse of services, and variation in services. Based on the reality of these quality problems, the Commission recommended that the initial set of national aims should include (32):

- Reducing the underlying causes of illness, injury and disability
- Expanding research on new treatments and evidence on effectiveness
- Ensuring the appropriate use of health care services
- Reducing health care errors
- Addressing oversupply and undersupply of health care resources
- Increasing a patient's participation in his or her care

The President's Commission engages a broad consumer advocacy movement in public and private sectors calling for a major reform of the U.S. health care system to improve access to care for more individuals living in America. Consistent with previous patterns, however, these calls have only led to incremental adjustments in policy and slight quality changes in direction. The major problems, for the most part, remain unaffected. Although broad based health care reform efforts have been unsuccessful, market forces and more targeted legislation and regulatory efforts have changed the face of health in the 1990s.

The 1993-94 Clinton health care reform plan, in its ideology, provided an ambitious plan to eliminate the enormous problem of lack of access to health care. It proposed to guarantee comprehensive health benefits for all American citizens and legal residents, regardless of health or employment status. The proposal was unsuccessful due to a number of factors, including its vast scope, the complicated nature of the plan, and an underestimation of

the politics involved with radically reforming health care. The failure of the Clinton administration health care reform agenda and the subsequent events to revise the American health care system are important lessons of health-care-system related politics.

Unfortunately, since the failure of the Clinton Administration plan in 1994, the number of uninsured individuals in America has grown. According to the Census Bureau, 44.3 million people are uninsured, comprising about 16.3% of the population. Of those uninsured, 15.4% are under 18 years of age, and the largest percentage is among individuals between 18 and 24 years of age. People of Hispanic origin make up 35.3% of those uninsured and 43% of the total uninsured population are not citizens of the United States (6).

The number of uninsured persons is expected to continue to grow. Proposals for health care reform to combat this problem include President Clinton's proposal for Medicare buy-in proposals for "middle aged" adults and House Majority Leader Dick Armey's (R-TX) proposal for a refundable tax credit to pay for insurance for the uninsured. The 2000 presidential campaign opened the debate for legislation that will improve health care coverage for the uninsured. This public debate on how to enhance access to care will stimulate creative ways to improve the U.S. health care system. However, rhetoric is not enough; it needs to be translated into programs that attack the problem.

The essence of the health care financing dilemma is related to how much a nation wishes to spend, on whom these funds are to be expended, and by what methods a relationship among cost, quality, and outcomes might be determined. In a time when advancing science and technology is flourishing in the health care field, "high tech" medicine will continue to evolve with an ever-increasing price tag. Furthermore, the costs of unanticipated and complex disease problems (e.g., HIV/AIDS) add to the unpredictability of health care system costs. This is all to say that most policy makers understand what needs to be done. They are in a quandary, however, in finding the appropriate and acceptable solution. Hence, it is likely that costs and expenditures will continue to rise (and, thereby, increase the percentage of GNP that will be spent for health care) and that solutions may become even more elusive.

Although some might argue that the available resources for expenditures on health care are ultimately limited, few are able to say exactly where that limit is or should be. In the United States, there has been an expansion of technologies and procedures based on scientific advancements without a concomitant development of a moral and ethical policy for determining who might be best served by such advancements. Rationing of health care services or

otherwise limiting access to high cost services, for example, has resulted from political policy rather than from deliberated public policy and rational decision making. This is most notably evidenced in the Medicaid component of the U.S. health care system.

As cost pressures continue to mount, there will likely be a return to having patients pay more of the health care expenditure dollar from their own resources. This will take the form of higher deductibles and co-insurance payments. Perhaps returning the burden of health care financing to the individual will raise the collective consciousness of American society that "there is no such thing as a free lunch" insofar as using and paying for health care services is concerned. Certainly, this phenomenon has occurred in social welfare "reform" in which the programs that have had mixed success have been restructured to "roll" participants off of welfare to work.

On the other hand, there are perhaps no solutions forthcoming on some of the problems represented in the arena of health care financing. As Hardin suggests, there is indeed a class of human problems that have no technical solution (33). In using Hardin's analogies, Hiatt (34) suggests that "nobody would quarrel with the proposition that there is a limit to the resources any society can devote to medical care, and few would question the suggestion that we are approaching such a limit. The dilemma confronting us is how we can place additional stress on the medical commons without bringing ourselves closer to ruin."

CONCLUSION

These are the principal contemporary features of the U.S. health care system. A massive societal structure is at once saviour, behemoth, juggernaut, and question mark. It certainly will be in a constant state of flux and gradual change. It therefore bears constant vigilance and careful guidance by those who derive their livelihoods from it and those who are the beneficiaries of its caring. Most importantly, it will require significant pressure from those who are disenfranchised from it.

REFERENCES

1. Donabedian, A. *Aspects of Medical Care Administration: Specifying Requirements for Health Care*; Harvard University Press: Cambridge, 1973.
2. Dougherty, C.J. *American Health Care: Realities, Rights, and Reforms*; Oxford University Press: New York, 1998.

3. Bella, R.N.; Madsen, R.; Sullivan, W.M.; Swidler, A.; Tipton, S.M. *Habits of the Heart*; Harper & Row: New York, 1985.
4. Relman, A. Reforming the Health Care System. *New Engl. J. Med.* **1990**, *323* (14), 991–992.
5. www.census.gov (accessed Oct 1999).
6. www.doc.gov (accessed Oct 1999).
7. *American Hospital Association, Hospital Statistics*, Health Forum: Chicago, IL, 2000.
8. *Modern Healthcare*, By the Numbers 1999 Edition (Supplement to Modern Healthcare), July 19, 1999.
9. www.who.gov (accessed Nov 1999).
10. Fuchs, V.R. Health Care for the Elderly: How Much? Who Will Pay for it? *Health Aff.* **1999**, *18* (1), 11–21.
11. www.cdc.gov/nchs (accessed Oct 1999).
12. www.nahc.org (accessed Dec 1999).
13. www.hcfa.gov (accessed Dec 1999).
14. Quality First: Better Health Care for all Americans. *The President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry*, 45.
15. www.NCQA.org (accessed Oct 1999).
16. Landis, N. *Am. J. Health Syst. Pharm.* **1999**, *56*, 1392.
17. Iglehart, J. The American Health Care System: Expenditures. *New Engl. J. Med.* **1999**, *340* (1).
18. New Health Care Model Balances Cost, Quality, Shifts Power to Patients. PR news wire. Nov. 1999.
19. www.hcfa.gov (accessed Oct 1999).
20. www.bea.doc.gov (accessed Oct 1999).
21. Schieber, G.J. Health Expenditures in Major Industrialized Countries, 1960–87. *Health Center Financing Review* **1990**, *11* (4), 159–167.
22. www.phrma.org (accessed Oct. 1999).
23. Smith, S.; Freeland, M.; Heffler, S.; McKusick, D. Health Tracking: Trends. *Health Aff.* **1998**, *17* (5), 128–140.
24. Klarman, H.E. *The Economics of Health*; Columbia University Press: New York, 1965.
25. www.hcfa.org (accessed Oct. 1999). Medicare + Choice: Changes for the Year 2000 Executive Summary.
26. Eisenberg, J.M. Clinical Economics: A Guide to the Economic Analysis of Clinical Practice. *JAMA* **1989**, *262*, 2879–2886.
27. Ginsberg, E.N. For Profit Medicine: A Reassessment. *New Engl. J. Medicine* **1988**, *319* (12), 757–761.
28. www.aha.org (accessed Nov 1999).
29. Whiteis, D.G.; Salmon, J.W. *The Corporate Transfer of Health Care*; Baywood Publishing Company: Amityville NY, 1990.
30. Christianson, J.B. Institutional Alternatives to the Rural Hospital. *Health Care Financing Review* **1990**, *11* (3), 87–97.
31. McLennan, K. Care and Cost. *Current Issues in Health Policy*; Westview Press: Boulder CO, 1989.
32. Quality First: Better Health Care for all Americans, The President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry.
33. Hardin, G. The Tragedy of the Commons. *Science* **1968**, *162*, 1243–1248.
34. Hiatt, H.H. Protecting the Medical Commons: Who is Responsible. *NEJM* **1975**, *293*, 235–241.
35. Quality First: Better Health Care for All Americans, Final Report to the President of the United States. The President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry.
36. To Err is Human: Building a Safer Health System, Institute of Medicine. National Academy of Sciences, 1999.
37. Kronick, R.; Todd, G. Explaining the Decline in Health Insurance Coverage, 1979–1995. *Health Aff. Mar/April*, **1999**, 30–47.
38. Merrill, R.A. Modernizing the FDA. *An Incremental Revolution. Health Affairs*, Mar/April, 1999; 96–111.
39. Robinson, J.C. *The Future of Managed Care Organization. Health Affairs*, Mar/April, 1999; 7–24.
40. Grunback, K. Primary Care in the United States—The Best of Times, The Worst of Times. *New Engl. J. Med.* **1999**, *341* (26).
41. Smith, B.M. Trends in Health Care Coverage and Financing and their Implications for Policy. *New Engl. J. Med.* **1997**, *337* (14).
42. Vladeck, B.C. *The Political Economy of Medicare. Health Affairs*, Jan/Feb, 1999; 22–36.

HEALTH CARE SYSTEMS: OUTSIDE THE UNITED STATES

Albert I. Wertheimer

Temple University, Philadelphia, Pennsylvania, U.S.A.

Sheldon X. Kong

Merck & Co. Inc., Whitehouse Station, New Jersey, U.S.A.

INTRODUCTION

It is quite fascinating how the organization, structure, and financing of health care services can be so very diverse in different countries around the world. One might think that leaders and policymakers would be aware of each other's national health systems and, by emulating the best features, that they would tend to move toward harmonization and greater similarity.

Actually, this assumption is false. National health care systems vary widely and are more related to variables in each country (1). In fact, the health system in a given country is a mirror of how that society functions at large. Health care delivery systems must be compatible with the: 1) *economic system*: socialist, capitalist, or mixed; 2) *political system*: major or minor role of degree of government centralization; 3) *wealth of the country*: use of primary care facilities, access to specialists and tertiary care facilities; 4) *traditions and conventions as seen in their history*—fundamental, visible things are difficult to change; 5) *geography*: whether the majority of the population is located in a few metropolitan areas, with the remainder scattered in rural areas, or whether the population is spread over hundreds of islands; 6) *infrastructure*: roads, communication systems, and air service; and 7) *extent of and belief in high technology* (2).

There are other factors as well: the system from a previous colonial power, extent of literacy and education, and relationships with outside countries, to name a few.

BACKGROUND

The remainder of this article examines the health care delivery systems in six very different countries. Even though Canada and the United States are similar countries with a shared border and language and with open

communication, their health care delivery systems could not be any more different. Each side of the border is aware of what happens on the other side, however, a series of complex and powerful forces keep them moving in their own directions.

We look at six countries very briefly in this article to highlight the incredibly diverse approaches to health service organization and financing. In essence, most health systems fit into one of the following models:

1. State ownership and control—The best examples are the British National Health Service and the Swedish system in which clinics, hospitals, and most service providers are owned and operated by the government (3).
2. State health insurance program—Here, the government is the sole or major payer. However, some of the facilities and resources are in nongovernment hands. This is the case in much of Europe (4).
3. Mixed systems—This is seen in much of Asia and Central America and usually where there is a small wealthy class and a massive lower class. The lower class receives care from public facilities, and the small upper class uses private-sector, fee-for-service, and self-paid care.

Other scenarios fit into this category as well. The United States has several independent health care systems including the military, veterans, Medicaid (a federal program for the medically indigent), Medicare (a federal insurance program for those 65 years of age and older), private-sector for-profit, and not-for-profit clinics, hospital chains, managed-care organizations, religious, prison health, and university teaching facilities (5).

4. Exclusively private sector—This category is shrinking as nations realize that health maintenance and disease prevention/wellness are important to their national goals of strength and productivity. Switzerland would still fit into this category, where most health care resources are in private hands (6).

SPECIMEN NATIONAL SYSTEMS

Canada

Organization

Canada uses a national health service, which provides medical services and hospital care to its entire population. The individual provincial governments operate health plans that conform to national legislation but can differ in various aspects. This “Medicare” program guarantees comprehensiveness, universal access, portability, and public administration (7).

Health Canada is the national, federal health agency; however, the operation of health service provision is delegated to the provincial governments, which control virtually 100% of Canada’s hospitals. There is a gatekeeper primary health care system, with GPs (general practitioners) or primary care family doctors serving as the entry point. Access to specialists, diagnostic testing, hospitals, and others is through the GP. Individual citizens have the freedom to choose their own doctors, 95% of whom are self-employed in private practice. The provincial government pays these doctors on a fee-for-service basis.

The individual provincial governments offer different supplemental benefits not covered by the national Medicare program, such as drugs, dental care, and vision care to the poor, elderly, and other specific groups. Supplemental benefits for the typical, employed, and nonelderly person come from the purchase of supplemental health insurance from private sources (8).

Pharmaceuticals

Canada created the Patented Medicine Prices Review Board (PMPRB) in 1987 to guarantee that pharmaceutical products would not have excessive prices in Canada. The board reviews prescribed and over-the-counter (OTC) prices and publishes annual guidelines for manufacturers. Compliance with PMPRB guidelines is voluntary; however, since 1993, the board has the authority to reduce excessive prices and return the excess amount to the government, and to punish the manufacturer.

The PMPRB compares prices in Canada with those in seven industrialized nations (France, Germany, Italy, Sweden, Switzerland, the United Kingdom, and the United States) to ensure that Canadian prices are in line with those of comparable countries. There is some controversy that existing drug products are well-controlled regarding prices, but that such is not the case with newly introduced pharmaceuticals.

Further controls exist at the provincial level at which each province maintains a published formulary of drugs that are reimbursable along with the reimbursement level. Quebec, observers perceive, lists nearly all new drug products, whereas Ontario appears to be slow to list newly approved products. Each province has additional control mechanisms. Ontario requires the first generic drug to be at least 40% less costly than the branded originator product. Some components of the reference price system are seen in British Columbia and Newfoundland.

There is growing harmonization among the provinces; however, there is still no national, standardized, and interchangeable list of drugs for ambulatory care use. In hospitals, drugs that are administered are paid for by Medicare. Each province has interesting and different features in its drug benefit plan.

The Prince Edward Island plan pays for seniors; welfare recipients; nursing home patients; and those with rheumatic fever, diabetes, tuberculosis, multiple sclerosis, AIDS, and several other conditions. New Brunswick has an annual copayment cap for seniors and for organ transplant recipients and for selected other patient categories. A copayment is set at approximately \$9 (Canadian) but is waived for some groups in Quebec, along with an annual copay ceiling of \$750.

Other interesting features of the Canadian system include its 1998 mutual recognition agreement with the EU, prohibition of prescription drug advertising to consumers, a 20-year patent exclusivity period, and the establishment of the PMPRB to ensure fair pricing of medications (9, 10).

Republic of South Africa

Organization

The Republic of South Africa (RSA) has a most diverse health care environment, with world-class practice and facilities in wealthy urban areas and some of the most primitive care in poor remote villages, with a vast array between these extremes. Primary care is now the focus of the ANC government in an effort to correct years of neglect and undemocratic practices under the earlier apartheid-oriented regimes. Public health services are being brought to the Black townships as rapidly as resources permit (11).

However, there are virtually no funds for new drugs against HIV infection in patients, a problem most prevalent in the RSA. To maximize the value of its drugs budget, the RSA has enacted legislation to create an Essential Drugs List for the public sector, along with generic substitution authority, the removal of some pharmacists’ unique

professional privileges, and legislation permitting the parallel importation of pharmaceutical products already registered in the RSA. Obviously, this conserves resources, stretching them for more patients, but this angers the RSA and multinational pharma firms.

South Africa is still the wealthiest country in Africa, with a (1997) GDP at approximately \$130 billion. It must be noted, though, that aggregate numbers hide massive racial differences. It is improving, but the standard of living for Blacks is yet only slightly better than it is in neighboring countries, whereas whites enjoy a standard of living similar to that found in North America or Western Europe. An unemployment rate of over 30% (mostly among Blacks) exacerbates the fiscal situation (12).

Routine immunizations for children, conforming to the World Health Organization (WHO) recommended schedule is the governmental policy, but it is not yet accomplished in all regions. Infectious diseases including HIV remain a serious challenge. Planning and budgeting for resource allocation are difficult because accurate census figures do not exist. Total health expenditures appear to be in the area of \$300 per person per year, and it is estimated that the private sector accounts for greater than 50% of total expenditures.

Public-sector expenditures emphasize primary care, lately, at the expense of tertiary care facilities. Private-sector spending is primarily through private "medical schemes." These are nonprofit organizations supported by employer associations and employees. There are slightly fewer than 200 of these schemes, providing insurance and care payment for nearly 3 million workers and their 5 million dependents (of a total estimated RSA population of 40 million). The largest area of medical scheme expenditure is for medicines, which causes the pressures on pharmaceutical pricing addressed below. After drugs, the next largest expenditures are for private hospitals, medical specialists, general practitioners, and dentists (13).

The RSA Department of Health (DOH) has totally restructured the previous apartheid system of racial and provincial health systems into a coordinated national health program operated through health regions and local health districts. Still, there are major differences in knowledge, education, expectations, and wealth within different subpopulations (14, 15).

Pharmaceuticals

Until recently, manufacturers were free to establish their desired price for a drug. Wholesalers and retailers added what they chose to reach the retail selling price for medications. In 1997, a proposed scheme of prices extending to the retailer was agreed on, but resistance was met from the Pharmaceutical Manufacturers Association(PMA). In the

legislation, a pricing board composed of members selected by the Minister of Health would establish prices for each product and a maximum selling price. Public-sector primary care drugs are reimbursed 100% by the government. Hospital care outpatient drugs can have copayments. The Essential Drugs List would be the core of what is to be available at public facilities, but there appears to be a long way to go before most of these agents will be regularly available on a consistent basis at primary care centers or at public hospitals (13).

The parallel importation of RSA-registered drugs available at lower prices abroad is the basis for PMA litigation against the Drug Legislation of 1997. In addition to the price-setting committee, DOH efforts to encourage the use of generic drugs has proven to be a source of conflict. Other features of the new legislation bar dispensing samples or making bonus payments to dispensers of medicines; the creation of a Code of Ethics for pharmaceutical marketing; and a series of safety regulations, dealing primarily with limiting practice to fully qualified and licensed professionals.

There is a fast lane for new drug approvals if the product is already in at least one of the following jurisdictions: the United Kingdom, Canada, United States, Sweden, or Australia. Approximately 85% (by value) of pharmaceuticals go through the nearly 3,000 community pharmacies. Yet, approximately 80% of the population rely on the public sector for drugs, received through clinics, hospitals, primary care posts, or military facilities. Although there is a 20-year patent period of exclusivity/protection, the parallel imports option effectively defeats this protection.

It will be interesting to see how the access to drugs, price controls, and quality improvement forces will interact and what the actual situation will be in South Africa in the coming years, especially as the country complies with intellectual property and World Trade Organization policies and rules (16).

Japan

Organization

After North America and before Western Europe, Japan is the second largest pharmaceutical market in the world. Its population of 126 million spends \$70 billion on pharmaceuticals each year. On average, each Japanese resident spends \$2000 each year on health care with \$550 of that on pharmaceuticals. Perhaps the primary single features of the Japanese market are the above-average proportion of elderly in the population and the higher than usual consumption of drugs. It has been estimated that by the year 2050, nearly 30% of the population will be older

than 65 years of age. The high consumption rate is attributed to drugs being injected and/or sold by the physician, a practice used, in part, to increase the total price of an office visit (17).

The primary funding source for health services in Japan is the Social Insurance System (SIS), made up of employee programs that pay for nearly 55% of care. The Medical Service for the Aged program covers another 35% of care. Private expenditures and a very small portion for public health promotion and disease prevention make up the difference. The Ministry of Health and Welfare (MHW) maintains overall responsibility for health care services and functions via a number of bureaus. Numerous sources comment that regulations are difficult to understand and interpret, often overlapping, and that this serves as a barrier to foreign firms desiring to enter a market. Physicians, for example, are authorized to own and operate hospitals, effectively excluding corporate owners or physicians not licensed in Japan (18).

Universal health insurance was established in 1961. Nearly the entire population is covered through the employer plans or through programs for the unemployed, retired, or self-employed. Employees pay 10% of the cost of treatments, up to an annual ceiling, and also pay a portion of their premiums, with their employers.

Pharmaceuticals

The MHW sets prices for reimbursable drugs (those approved for the Social Insurance System). Physicians, clinics, and private hospitals are reimbursed at a price slightly higher than their actual acquisition cost. The government has scheduled annual reductions in the reimbursement prices to reduce this source of additional income to physicians. Patients make copayments of 20%, although for children and low-income elderly the copayment is waived, and recently a plan to eliminate copayments for persons 70 years of age and older was introduced.

The MHW reductions of 5–10% of the prices of existing drug products appear to have had the opposite of the intended impact. Doctors are prescribing more of the newest, high-priced pharmaceuticals that have not had their margins reduced yet, thereby earning a bigger amount from the wider difference between their actual cost and the listed reimbursement amount.

With regard to generic drugs, astute observers believe that the Japanese government wants its R&D-intensive firms to be successful. A regulation requires generics to be priced at not less than 40% of the innovator brand price. It is reasonable to assume that the margins (Yakkasa) for physicians are lower with generic drugs, and that these margins will continue into the future, as will the reference price scheme (19).

There is a Japanese pharmacopeia that sets official standards and diverse government agencies that perform tasks undertaken by an FDA. It is rumored that the Japanese will establish a Western-style FDA in the near future.

One of the most disliked regulations in the view of foreign and multinational pharmaceutical companies is the requirement for duplicative clinical trials with humans in Japan, because those carried out elsewhere are not recognized. Also of interest is the fact that Japan, like Korea and Taiwan, has no separation between prescriber and dispenser of drugs. Called “Bungyo,” it is a major source of revenue for doctors and clinics. Fewer than 20% of prescriptions ever reach a pharmacy for dispensing (19).

Good post-marketing surveillance practices (GPMSP) rules have been in place since 1993. Postmarketing experience reports are to be sent to a government agency. Both GPMSP and periodic safety reporting requirements are in place that require a review of the product each year while it is in its re-examination period, immediately after marketing approval. Unlike in the United States, where a new drug application is approved for an indefinite period, in Japan, there is a periodic full reassessment. Such re-evaluations are conducted every 5 years once the initial re-examination period for a drug product has ended.

Drug products are distributed primarily via the 2000 wholesalers, and in addition, there exists a small second channel with drugs going directly to hospitals, GPs, and pharmacies. There are approximately 66,000 pharmacies, most of which are family-owned independents. There are chains as well. However, a growing market for OTCs is found in convenience stores.

Physicians administer and sell drugs to patients as a highly profitable sideline. The incentive is for the physician to use as much of the most costly drug products as possible. There is only a small OTC market, because physicians try to prescribe and dispense as much as is possible. Other than some concern about a drug lag, the pharmaceutical environment in Japan is robust. Periodically, there are calls to separate prescribing and dispensing; however, this is not likely in the near future given the powerful forces backing the status quo (20).

United Kingdom

Organization

With a population of more than 60 million and GDP per capita of more than US \$22,000, the United Kingdom is one of the richest nations in the world. It is one of the G7 countries, a member of the European Union, and a member of the Organization for Economic Co-operation and Development (OECD).

In 1996, total health care expenditure in the United Kingdom was approximately 7.0% of the GDP. Public expenditure by the National Health Service (NHS) accounts for most of the health care costs. The NHS was set up after World War II, with the aim of unifying health care services by voluntary and local hospitals. The NHS offers free health services to all U.K. residents, funded through general taxation.

Two of the major characteristics of the U.K. health care system include health authorities responsible for hospital services and GP fundholders responsible for primary care. In 1996, 100 health authorities became operational in England, responsible for the provision of NHS hospital and community health services covering geographic boundaries with populations ranging from 125 thousand to over 1 million. There are four levels of hospital services. At the community level, community hospitals offer basic medical care for the treatment of acute cases and patients requiring convalescent and long-term/terminal care. General practitioners are the key staff here. At the district level, district general hospitals operate the key acute units, serving an average population of a quarter-million. At the regional level, major specialty services such as neurosurgery, open-heart surgery, and radiotherapy are provided. At the national level, highly specialized hospitals provide complex services for parts or for the entire country (21).

GPs are the gatekeepers and fundholders of the health care system. The principle of fundholding is that GPs manage their own budgets. They can obtain a defined range of services from hospitals and manage patients at the GP level whenever possible to reduce costs. In the late 1990s, GPs fundholders were organized into Primary Care Groups (PCGs). These networks of GPs cover wide geographic areas with an average population of 100,000. In 1999, there were 481 PCGs in England and Wales, and all have unified budgets (e.g., drugs, hospital care services). With a population of a small to medium-sized HMO in the United States, these PCGs have a very broad influence on patient health care and the selection of drugs through formularies.

Pharmaceuticals

The regulatory authority in the United Kingdom is the Medicines Control Agency (MCA) under the Department of Health. The agency's responsibilities include drug licensing, clinical trials licensing, pharmacovigilance and drug safety, communication and provision of information on medicines, inspection of facilities and enforcement of regulations, and the *British Pharmacopoeia*. The United Kingdom is a reference member state for the European Union mutual recognition procedure. The European Union's pharmaceutical registration system came into

effect for all member countries in 1995. The aim of the EU system is to harmonize pharmaceutical regulations throughout the EU. The centralized registration procedure is handled by the European Medicines Evaluation Agency (EMA). Authorization through the central registration procedure is immediately valid in all EU member countries. The decentralized procedure relies on the principle of mutual recognition. After registration has been obtained in a member country under the centralized procedure, application may be made for registration in one or more other member countries via the decentralized procedure (21).

The majority of pharmaceuticals are distributed through wholesalers to retail pharmacies, with large pharmacy chains now dominating the market. There are approximately 11,000 community pharmacies in the United Kingdom (21). In recent years, pharmacy services are increasingly available in supermarkets at the expense of local independent pharmacies.

Total expenditure on pharmaceuticals in the United Kingdom amounted to approximately 8650 million pounds in 1999, accounting for approximately 17% of the total health expenditure (21). The NHS covers prescription drugs. However, the government does not reimburse for over-the-counter (OTC) products. The Department of Health indirectly controls pharmaceutical prices. Because the price control scheme is related to profit control, rather than to the prices of individual products, pharmaceuticals are relatively free-priced in the United Kingdom. The government operates a negative list for products that are not reimbursable. The cost of most licensed prescription products is fully reimbursed. However, cost constraints and prescribing budgets mean that GPs will often prescribe a generic when one is available. As a result, new prescription drugs usually have a slower penetration rate in the United Kingdom than in the United States. The recently introduced National Institute for Clinical Excellence (NICE) will add more barriers to the introduction of new pharmaceutical products in the United Kingdom.

National Institute for Clinical Excellence

Funded by the government, the National Institute for Clinical Excellence (NICE) was set up as a Special Health Authority in the United Kingdom in 1999 and, as such, it is a part of the National Health Service (NHS). It was set up to "provide the NHS [patients, health professionals, and the public] with authoritative, robust and reliable guidance on current best practice." Its key functions are "to appraise the clinical benefits and the costs of those [health care] interventions and to make recommendations." Guidance is issued from each appraisal based on the clinical benefits, cost-effectiveness, and total economic impact on the

National Health Service. The government does not have to adhere to the recommendations by the NICE in its guidance and financial payment to health care providers. However, many believe that a negative recommendation from the NICE will have a detrimental impact on the pricing, reimbursement, and sales of the appraised product not only in the United Kingdom but also throughout Europe, Australia, and Canada.

The guidance covers both individual health technologies (including medicines, medical devices, diagnostic techniques, procedures, and health promotion) and the clinical management of specific conditions. The Institute may recommend a technology for general use, for specific indications, or for defined subgroups of patients. Based on the appraisal, a therapeutic intervention (e.g., drug) will be classified into one of three categories: category A, routine use in the NHS; category B, further trials needed; and category C, not recommended for routine use in the NHS.

The NICE has a board reflecting a range of expertise including the clinical professions, patients and user groups, NHS managers, and research bodies. The Board ensures that the NICE conducts its business on behalf of the NHS in the most effective manner. Details of the appraisal process and membership of the Appraisals Committee are available on the NICE Web site (www.nice.org.uk). Because the NICE was new at the time of completion of this article, its impact on the pharmaceutical industry is still not clear.

Germany

Organization

With a population of approximately 82 million in 1998 and a GDP per capita of more than \$26,000, Germany is one of the world's largest economies and health care markets. The population enjoys a generally good standard of health with a high degree of public awareness about health-related issues. Life expectancy in Germany is among the highest in the world. In 1997, the life expectancy for males was 74 years and for females 80. Approximately 15.8% of the population were over 65 years in 1997, and it has been projected that by 2020, the number of German inhabitants aged over 60 years will be 28.2% (22).

In 1997, health expenditures in Germany totaled \$298 billion, equal to 14.2% of the GDP. The health care system in Germany is decentralized, and health care expenditures are covered by a variety of sources/payers. The statutory insurance system (GKV) represents the biggest proportion of the total care coverage (for almost 50%). Employers, government budget, private households, private insurance, retirement insurance, and accident insurance cover the

remaining 50% of the health care expenditures. The largest spending sector is hospital expenditure, representing 34.3% of the total GKV health care expenditures (22).

The federal government has little executive responsibility for the provision of health care in Germany. Its primary responsibility is to provide a regulatory framework within which the individual Länder have to operate. The health ministries of the individual Länder are responsible for implementing the federal legislation, enacting their own legislation, supervising subordinate authorities and the medical profession, hospital planning, and regional administration.

Hospitals in Germany can be classified into three major categories based on ownership: public, nonprofit, and private. In 1997, the public sector operated approximately 40% of general hospitals, and nonprofit organizations operated another 40%. However, the number of privately owned facilities has been increasing steadily over the past decade.

The number of practicing doctors has risen steadily for the past 10 years. More than 70% of the practicing doctors are specialists, with general medicine as the largest specialty. Fewer than 30% of doctors practice without any specialty.

Pharmaceuticals

Germany is a reference member of the EU pharmaceutical registration system. The European Medicines Evaluation Agency (EMA) handles the centralized registration and the decentralized registration procedures in individual countries. After marketing authorization of a product with a new active substance has been granted in one country, the mutual recognition procedure is compulsory in other member countries. The mutual recognition procedure is also compulsory for line extensions and generic products. Marketing authorization approvals in Germany are valid for 5 years and renewable thereafter in 5 year periods.

Germany is the home of some major multinational pharmaceutical companies such as Aventis, BASF, Bayer, Boehringer Ingelheim, Merck KGaA, and Schering AG. VFA is the research-based manufacturers' association, whereas the Bundesverband der Pharmazeutischen Industrie (BPI) represents small and medium-sized companies. Because North America is the largest pharmaceutical market in the world, many of the VFA pharmaceutical companies locate their key operations in the United States. Exports to Western European countries represent a major source of income for many of the German pharmaceutical companies.

The pharmaceutical market in Germany is one of the largest in the world. Based on drug use per capita, Germany is second only to Japan in the consumption of pharmaceuticals. The principal distribution channels for pharmaceuticals in Germany are public retail pharmacies and hospital

pharmacies. In 1998, there were 47,322 pharmacists in Germany, equal to 0.6 pharmacists per thousand population (22). Public (retail) pharmacies employed 96% of all pharmacists in 1998 and they obtained their supplies primarily from wholesalers. Prescribed drugs, including both branded and generic products, can only be dispensed in a pharmacy with a doctor's prescription. The generics market in Germany is one of the largest and fastest-growing in Western Europe, representing approximately one-third of the European generics markets. OTC products can be divided into three overlapping categories: prescription OTC medicines, nonprescription OTC medicines, and freely available OTC products that can be sold freely through all retail outlets such as health food stores, supermarkets, and other retail outlets.

Mexico

Organization

Mexico is a federal republic of 31 states and a federal district. The population was officially estimated to be 97.7 million in 1997. GDP per capita was estimated at approximately US \$4400 in 1998. As a developing nation, communicable diseases are still one of the major causes of mortality, although chronic and degenerative diseases have become the leading cause of death during the past decade.

One of the major challenges for the government is to address the inadequacies of the Mexican health care system. Approximately 10 million people have virtually no access to regular basic health care services, and another 20 million people have less than adequate access. In 1996, the total health care expenditure in Mexico was equivalent to approximately 4.6% of GDP. Spending by the public sector accounted for approximately 60% in 1996 (23).

There are three sectors in the Mexican health care system: public, social security, and private. The public sector is primarily directed and operated by the Secretariat of Health. The public sector of health services is under the Secretariat of Health and is coordinated by over 200 health districts. The Federal District Department provides health care services to some 3.2 million people in Mexico City. The Mexican Social Security Institute (IMSS) Solidarity program covers another 10 million people in rural areas.

The social security system covers health services for government employees, managed by the Social Insurance Institute of State Employees (ISSSTE), and for private-sector workers, managed by the Mexican Social Security Institute (IMSS). The two agencies operate their own networks of hospitals and clinics and provide similar benefits. Some other smaller social security agencies exist, providing medical services for special groups such as the army, navy, and state oil company personnel.

The private (commercial) sector includes private hospitals, doctor's offices, and practitioners of traditional medicine. Charity organizations such as the Red Cross also play a role in the Mexican health care system.

Pharmaceuticals

The regulatory authority in Mexico is the Dirección General de Control de Insumos para la Salud (DIGECIS). The Health Secretariat issues pharmaceutical registration. Safety and efficacy must be proven by phase III clinical trials in Mexico to register drugs that are new to the Mexican market. All major pharmacopoeia (*International Pharmacopoeia*, *US Pharmacopoeia*, *British Pharmacopoeia*, *French Pharmacopoeia*, *Swiss Pharmacopoeia*, *European Pharmacopoeia*, and *Japanese Pharmacopoeia*) are acceptable in Mexico.

Most domestic producers in Mexico are wholly owned or licensed subsidiaries of multinational pharmaceutical firms. Exports have been growing fast, with other Latin American countries as the major destination markets. However, the United States is the major supplier of pharmaceutical imports in Mexico.

Pharmaceuticals in Mexico are subject to government price control. The private sector accounts for approximately 85% of the pharmaceutical market. Prescription drugs account for the majority of the pharmaceutical market, with antibiotics as one of the largest classes. Because the use of generics is still a relatively new phenomenon, most of the prescribed pharmaceuticals are branded products. OTC products represent approximately one-fifth of the total pharmaceuticals market.

SUMMARY

As presented, these six representative countries use vastly different organizations, financing mechanisms, goals, and provision structures. In fact, few systems around the world are identical because the systems represent the values and priorities and political as well as economic leanings and traditions of that country. If there were one perfect system, we would be seeing migration toward that model. However, because this is not the case, it is reasonable to assume that most of the various systems encountered around the world are at least satisfactory in their foundations and macrolevel characteristics, even if some of the operating details are not always popular (24).

The world is full of interesting additional approaches that a serious student of this subject might wish to explore further. Some of these include the "need clause" used in Norway, where, for example, their FDA had the authority

to refuse to accept and review a new drug because Norway already had six benzodiazepines on the market. The FDA deemed that sufficient unless the sponsoring company knew of a new indication or other therapeutic breakthrough from its use. The Swedes bought all of the then-existing community pharmacies in the country in 1970 to rationalize distribution, and service level and to create a monopsonistic body for negotiating with manufacturers in price-setting. The French and others place new drugs into one of several reimbursement categories. Clearly, life-saving drugs are put in the 100% reimbursement (to the patient) category. Most others strive for the 70% reimbursement category; however, if the manufacturer cannot agree on a price satisfactory to the Social Security agency, the product will be placed in a lower reimbursement category, effectively hampering its market success. This is a powerful bargaining chip for the government to contain drug prices.

It will be interesting to watch the future in this area to see how medications previously requiring a doctor's prescription that move to OTC status are handled, and how nutraceuticals, herbals, homeopathic, and naturopathic drugs, without the benefit of rigorous, randomized clinical trial or outcome data are handled as well. Similarly, we can be certain that there will be excitement galore when the nations in Central America and the Middle East decide to control pharmaceuticals and to end the practice of lay-person purchases of virtually any product without the benefit of a physician's order. Separation of pharmacy and physician functions will occur in the Far East in the not too distant future, causing even more excitement or grief.

If logic dictates, we should expect to see in the future a trend to offer incentives for prescribers who use the most cost-beneficial products (bonuses) and disincentives for patients (reimbursement level co-payment differences) and physicians when less than optimal choices are made. Irrespective of whatever does actually occur, it will be most interesting to observe.

REFERENCES

1. Roemer, M.I. *National Health Systems of the World*; Oxford University Press: Oxford, UK, 1993; 2, 61.
2. Fry, J.; Farndale, W.A.J. *International Medical Care*; Washington Square East: Wallingford, PA, 1972; 367.
3. In *OECD Health Systems: Facts and Trends*; OECD: Paris, 1993; 100–161.
4. Hewitt, M. *International Health Statistics*; Office of Technology Assessment, U.S. Congress: Washington, D.C., 1993; 76–86.
5. Elling, R.H. *International Health Perspectives*; Springer: New York, 1977; 4, 17–25.

6. Joseph, S.C.; Koch-Weser, D.; Wallace, N. *Worldwide Overview of Health and Disease*; Springer: New York, 1977; 7–43.
7. Korman, R.A. *Academic Reference Manual: Canadian Health Care Information*; IMS Health: Mississauga, Ontario, 1999; 80–162.
8. *World Pharmaceutical Markets: Canada*; Espicom Business Intelligence: Chichester, UK, May, 1999; 9–48.
9. Alleyne, G.A.O. *Health Statistics from the Americas*; Pan American Health Organization: Washington, DC, 1998; 17–37.
10. Alleyne, G.A.O. *Health in the Americas*; Pan American Health Organization: Washington, DC, 1998; 1, 325–337.
11. Monekosso, G.L. *Eighth Report on the World Health Situation; African Region*; WHO: Brazzaville, 1994; 2, 2–37.
12. Nokagima, H. *Eighth Report on the World Health Situation*; 147–165 Global View: Geneva, 1993; 1, 37–59.
13. *World Pharmaceutical Markets: South Africa*; Espicom Business Intelligence: Chichester, UK, Feb 1999; 3–13.
14. *World Development Report 1993*, World Bank; Jamison, D.T., Ed.; 108–170 Oxford University Press: Oxford, 1993; 3–65.
15. *World Tables 1995*; World Bank: Washington, DC, 1995; 22–66.
16. Basch, P.F. *Textbook of International Health*; Oxford: New York, 1990; 144–326.
17. *Bartholomew's Mini World Factfile*; Harper Collins: London, 1995; 44, 102, 175.
18. *Eighth Report on the World Health Situation: Western Pacific Region*; World Health Organization: Manila, 1993; 7, 79–83.
19. *World Pharmaceutical Markets: Japan*; Espicom Business Intelligence: Chichester, UK, Jan 1999; 10–23.
20. SCRIP: London UK, 1998, 1999, and 2000.
21. *World Pharmaceutical Markets: United Kingdom*; Espicom Business Intelligence: Chichester, UK, Feb 1999; 17–70.
22. *World Pharmaceutical Markets: Germany*; Espicom Business Intelligence: Chichester, UK, Feb 1999.
23. *World Pharmaceutical Markets: Mexico*; Espicom Business Intelligence: Chichester, UK, Feb 1999.
24. Brudon, P. *The World Drug Situation*; WHO: Geneva, 1988; 7–108.

FURTHER READING

- Alexander, T.J. *Internal Markets in the Making: Health Systems in Canada, Iceland, and the UK*; OECD: Paris, 1995.
- Saltman, R.B.; Figueras, J.; Sakellarides, C. *Critical Challenges for Health Care Reform in Europe*; Open University Press: Buckingham, UK, 1999.
- Schneider, M.; Dennerien, R.K.; Kose, A.; Scholtes, L. *Health Care in the EC Member States*; Elsevier: Amsterdam, 1992.
- Spivey, R.N.; Wertheimer, A.I.; Rucker, T.D. *International Pharmaceutical Services*; Haworth: Binghamton, NY, 1992.
- The Use of Essential Drugs: 6th Report of the WHO Expert Committee*; WHO: Geneva, 1995.
- Van de Water, H.; Van Herten, L.M. *Health Policies on Target*; THO Prevention and Health: Leiden, 1998.



Headspace Oxygen Analysis in Pharmaceutical Products

Allen C. Templeton

Robert A. Reed

Merck & Co., Inc., West Point, Pennsylvania, U.S.A.

INTRODUCTION

Oxidative chemical degradation of drug substances in pharmaceutical formulations is well documented.^[1] Although exact mechanistic details on what governs the promotion of reactions between drug substances (RH) and molecular oxygen in pharmaceutical formulations is not fully understood, it is generally thought that such reactions fall under the category of autoxidation processes^[1,2]:

Initiation $\text{RH} \rightarrow \text{R}\cdot$

Propagation $\text{R}\cdot + \text{O}_2 \rightarrow \text{ROO}\cdot$

$\text{ROO}\cdot + \text{RH} \rightarrow \text{ROOH} + \text{R}\cdot$

Termination $\text{ROO}\cdot \rightarrow \text{Molecular products}$

Note in the above reaction scheme that molecular oxygen is involved in the propagation step of the reaction and is integral to the catalytic cycle responsible for the generation of drug substance oxidative degradation in pharmaceutical formulations.

Oxidative degradation of the active drug substance in a pharmaceutical formulation leads to a lowering of drug potency and potentially reduced product shelf life.^[2] The rate of oxidative degradation is proportional both to the structural susceptibility of the specific drug substance to auto-oxidize and storage (reaction) conditions such as temperature, humidity, oxygen concentration, and time. Other deleterious effects of oxidative processes that have been noted include product discoloration, changes in dissolution rate/profile, precipitation, and the generation of foul odors and flavors.^[1] Most importantly, oxidative degradation products generated in the final pharmaceutical product upon storage may also have adverse pharmacological properties, including those related to toxicity or adverse side effects.

In addition to oxidative degradation of the finished product upon shelf storage, drug oxidation during the manufacturing process can also pose significant challenges for readily oxidized drug substances. The effects are especially pronounced for formulations where

manufacturing unit operations are performed in solutions and/or suspensions. In these cases, care may be required to minimize or eliminate exposure to oxygen during key processing steps. Inert gas purges to lower solution oxygen levels and inert blanketing of process vessels have long been employed in parenteral manufacturing processes. Besides purging at key manufacturing steps, e.g., during drug compounding and holding in solution prior to lyophilization, packaging with an inert blanket over the final product has also been employed for parenteral formulations susceptible to oxidation. By removing from the package one of the two key reactants in the autoxidation cycle, important gains can be made in product shelf life and the quality of the product reaching the consumer. Packaging oxygen-sensitive products under an inert atmosphere, often termed modified atmosphere packaging (MAP), is a process widely adopted to increase the shelf life of other oxygen-sensitive commercial products, including foodstuffs and medical devices. While MAP has taken hold in the parenteral arm of the pharmaceutical industry, there are relatively few examples of solid dosage forms that are packaged under reduced oxygen levels.

“Smart” packaging approaches like MAP may allow for the commercial development of oral formulations with drug compounds particularly prone to oxidation and thus allow for the availability of critical therapeutic agents in a high quality, stable, and convenient dosage form. The development cycle time of such compounds could also improve, in that, efforts to stabilize the compound by both formulation and packaging can be successfully integrated. Thus, MAP implementation for solid dosage forms is clearly on the horizon. Keys to successfully deploy MAP for any oxygen-sensitive product are: 1) utilization of packaging that presents a high barrier to oxygen permeation; 2) incorporation of an inert gas flushing step compatible with the speed of current product packaging lines; 3) selection of an appropriate inert gas (e.g., argon, nitrogen, carbon dioxide, etc.); and 4) rapid means to measure residual headspace oxygen content in a variety of pharmaceutical packaging configurations—ranging from very low (tens–hundreds of μL) to relatively high volume (mL – L) packages.

ANALYTICAL APPROACHES FOR HEADSPACE MEASUREMENTS

Fast, sensitive methods for monitoring headspace oxygen levels would critically benefit MAP development and deployment in all three areas listed above, aiding in each of the phases involved with packaging R&D, manufacturing process development, and final product quality assurance. In terms of packaging R&D, robust oxygen measurement methods would allow for a real-time screening of the suitability of packaging systems (all components of a container and closure) to serve MAP packaging applications. The most obvious example of this would be the acquisition of oxygen ingress kinetics for the packaging system under both realistic and accelerated storage conditions. Experimental data could be used to develop predictive models around each critical packaging system component to support future MAP development programs. For manufacturing process development, scaling and validating a manufacturing process that allows maximum package throughput per unit time while achieving and maintaining critical threshold levels for oxygen would benefit from rapid and accurate oxygen determination in the development feedback loop. Finally, tools to provide quality assurance/control information (at-, on-, or off-line) would ensure that the oxygen content of the inert atmosphere in the packaging container headspace is below the predetermined specification for the product. An assessment of container closure integrity could also be determined using such testing, augmenting or replacing

current microbial/dye ingress measurement approaches based on sterility assessment.

The rapid and accurate determination of headspace oxygen levels in the various packaging configurations used in the pharmaceutical industry poses a number of unique analytical challenges that are beyond the capabilities of a single measurement technique. The purpose of this article is to provide a background on various measurement options available to the packaging R&D technologist or laboratory analyst in the context of pharmaceutical packaging applications. The article is organized by technique with a brief introduction to measurement principles, followed by a critical discussion of pharmaceutical packaging applications.

A REVIEW OF HEADSPACE OXYGEN ANALYSIS TECHNIQUES

Headspace Gas Chromatography (GC)

Principles of operation

When a problem arises in a pharmaceutical research laboratory that requires analysis of a headspace sample, the first tool that probably comes to mind is the tried-and-true approach of headspace GC. This is for good reason, as the literature points to a vast body of knowledge, stretching back 40 yr, on the theory and application of this methodology.^[3] While not taking away the importance

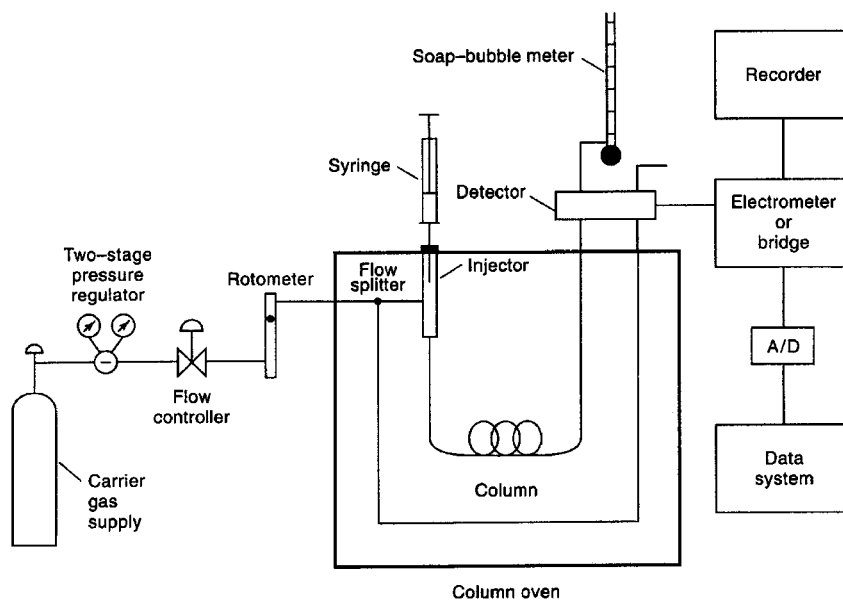


Fig. 1 A block schematic of the components that comprise a typical GC instrument. (From Ref. [12], p. 41. Copyright ©2002 Advanstar Communications Inc. Advanstar Communications Inc. retains all rights to this material.)

Headspace Oxygen Analysis in Pharmaceutical Products

3

of headspace GC as an analytical problem-solving tool, which stretches far beyond the bounds of oxygen analysis, a number of problems exist where other techniques can provide similar or better results for specific headspace analytes, often in a more timely fashion. The following description of headspace GC is only intended as a cursory overview; refer the substantive literature of the field for greater detail on the subject.^[3]

GC is inherently a technique for the study of volatile compounds. Fig. 1 shows a block schematic of the components that comprise a typical GC instrument equipped with split injection port and capillary column. For headspace sampling, a gaseous sample is introduced by injection into an inert moving gas stream called the mobile phase or carrier gas.

The mobile phase transports the sample onto a stationary phase where the components of the sample are separated by means of selective partitioning between stationary and mobile phases. The time referred to as retention time (t_r), which elapses between sample introduction and when peaks appear at the detector, is characteristic of the properties of the analyte molecules. The area of the respective peaks as they appear in the resultant chromatogram is proportional to the analyte concentration in the sample, as related by a measurable response factor. For the analysis of headspace oxygen in pharmaceutical applications, a sample of headspace (tens of μL) is removed from a package either by withdrawing a portion of sample using a gas-tight syringe or by using

the GC autosampler to pressurize the container with mobile phase to displace an aliquot of sample. Performing the experiment in either manner would mean operation in what is referred to as the static mode of headspace sampling, wherein the aim is to sample the equilibrium contents of a sample headspace at a given instance.

Fig. 2 shows an example chromatogram of lab air illustrating the determination of headspace oxygen using a megabore capillary column with thermal conductivity detection. As in any type of chromatographic analysis, selection of the appropriate column to achieve separation of the sample gas mixture into its various components is pivotal. In this case, a commercially available thin-film porous layer open tubular (PLOT) capillary column consisting of a fused silica capillary tube coated with a homogeneous 50- μm thick zeolite molecular sieve (5/pore size) film was used to separate oxygen from argon and nitrogen in an air sample during a 10-min analysis. The mode of separation is based on the size of the analyte molecule or atom, and retention time is as follows: Nitrogen > Oxygen > Argon. The net area and percent area of oxygen are determined from the following equations:

$$\text{Area}_{\text{Net},x} = \frac{(\lambda_{\text{He}} - \lambda_x)}{\lambda_{\text{He}}} \times \text{Area}_{\text{Raw},x}$$

$$\text{Area}_{\text{Net},x} \% = \frac{\text{Area}_{\text{Net},x}}{(\text{Area}_{\text{Net},\text{Argon}} + \text{Area}_{\text{Net},\text{Nitrogen}} + \text{Area}_{\text{Net},\text{Oxygen}})} \times 100\%$$

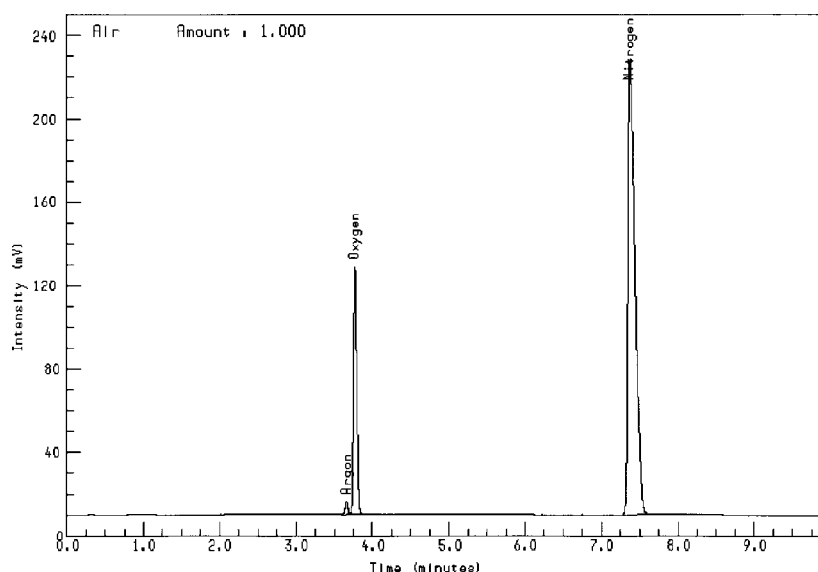


Fig. 2 An example chromatogram illustrating the determination of headspace oxygen by GC using a PLOT molecular sieve column with thermal conductivity detection. Chromatographic conditions were carrier gas: helium (2 mL min^{-1}); oven temperature: 26°C ; inlet: 160°C , split mode, 10:1 split ratio, split flow of 20 mL min^{-1} ; injector: 160°C ; run time: 10 min; TCD detector: 160°C . (From Ref. [12], p. 41. Copyright ©2002 Advanstar Communications Inc. Advanstar Communications Inc. retains all rights to this material.)

where x is argon, nitrogen, or oxygen, $\text{Area}_{\text{Raw},x}$ the raw peak area of x ($\mu\text{V sec}$), $\text{Area}_{\text{Net},x}$ the net peak area of x after conversion, λ_{He} the conductivity of helium ($154.6 \text{ W m}^{-1} \text{ K}^{-1}$), and λ_x the conductivity of argon ($17.8 \text{ W m}^{-1} \text{ K}^{-1}$), nitrogen ($25.9 \text{ W m}^{-1} \text{ K}^{-1}$)*, or oxygen ($26.2 \text{ W m}^{-1} \text{ K}^{-1}$)*.

A potential drawback of GC or any technique requiring use of a syringe for sample injection is the potential for contamination. Contamination can result from: 1) leakage of the package during sample removal; 2) syringe leakage during sample transport from package to instrument; 3) septum port leakage during sample introduction; and/or 4) contamination from residual air remaining in the syringe barrel. Residual air contamination from syringe barrel leads to a direct correlation between contamination and syringe barrel volume.

Applications

The utilization of headspace GC to analyze oxygen in pharmaceutical packages and samples offers several advantages, such as the widespread availability of instruments and user familiarity. For sample containers of appropriate dimensions, the autosampling capability of commercial headspace GCs makes the instrument particularly appealing for analyzing samples en masse. GC autosampler vial dimensions are not typically consistent with common pharmaceutical package sizes. Thus, the utility of the autosampler is typically limited to either a few samples of appropriate dimension or samples that can be transferred to headspace autosampling vials; the latter would defeat the purposes of many investigations. Withdrawing and injecting samples manually using a gas-tight syringe can pose various problems in addition to those noted above; for example, when working with samples that are under vacuum, pressure equalization with ambient air can lead to sample dilution and erroneously large values for oxygen. Thus, care must be employed to use a gas-tight locking syringe and institute appropriate controls, as described in the previous discussion. Moreover, care must also be taken when removing sample from packages without introducing ambient atmospheric leakage during the sampling process; application of a self-adhesive rubber septum prior to puncturing the sample is usually sufficient as a preventative measure when puncturing through aluminum foil (e.g., foil induction sealed HDPE bottles). Samples sealed with rubber stoppers (e.g., lvo vials) usually provide an adequate barrier to leakage during sampling. Sampling is a problem common to all destructive headspace oxygen analysis methods and points out the difficulties in trying to quantitatively measure small quantities of a gaseous analyte without contamination with air. Despite problems with sample removal, the low absolute sample volume

requirements (tens of μLs) make static headspace GC attractive and potentially useful for pharmaceutical packages and samples of all types/sizes. Following sample introduction, the major limitation of the technique for analysis of oxygen is the requirement of a relatively lengthy separation step (few minutes at best) to be oxygen selective, coupled with needed assessment of appropriate standards and demonstration of system suitability throughout analysis of a sample set. Lastly, although traditionally a GC instrument has a large footprint that would make at-line work unfeasible, recent commercial development of portable GC units (sometimes termed μGCs) is beginning to make at-line sampling more feasible.

μGC systems operate under the same principles as do conventional headspace GC analyzers. The key difference is scale, and scale translates into sample volume requirements. The instrument typically is comprised of miniaturized injector, column, and detector all mounted in a single module or channel. The sample is introduced through the inlet via an internal diaphragm pump connected to the end of the module. Due to the extremely low volume of the sample tubing and microscale nature of the instrumentation components, analysis volumes can be of the order of tens of microliters and analysis times less than 60 sec. Thus, the analysis time disadvantages of conventional headspace GCs are overcome with a μGC . An additional significant advantage with μGCs is that typical systems are two channel and allow for the simultaneous determination of two headspace analytes (e.g., oxygen and moisture).

Electrochemical Methods

Principles of operation

The use of electrochemical methods for oxygen analysis also has long historical precedence both from a scientific and commercial perspective. The basis of operation of most commercial electrochemical headspace oxygen analyzers is a solid-state zirconium oxide ion-selective electrode.^[4] In this sort of sensor, oxygen ions are able to migrate into the defect sites in the ceramic lattice structure of zirconia at elevated temperatures ($>400^\circ\text{C}$). As has been known since the time of Nernst (1899), when gases of differing concentrations reside on opposite sides of an electrode membrane, a measurable potential (E') difference is generated. If one side of the membrane is exposed to gas of reference oxygen concentration $[\text{O}_2]_{\text{REF}}$, then the potential generated in the presence of a gas of unknown oxygen concentration $[\text{O}_2]_{\text{UNK}}$ is proportional to the difference in oxygen concentration across the membrane, as given by the Nernst equation^[4]:

$$E = \frac{RT}{nF} \ln \frac{[\text{O}_2]_{\text{REF}}}{[\text{O}_2]_{\text{UNK}}}$$

where E is the potential (V), R the gas constant (J K^{-1}), T the temperature (K), n the number of electrons transferred,^[4] and F the Faraday's constant (J mV^{-1}).

In practice, the measurement procedure involves puncturing the package and removing (usually with a small diaphragm pump) 2 mL–4 mL of headspace for purging and subsequent analysis in a small cell containing the electrode. The analysis is rapid, typically of the order of 5 sec. Importantly, the measurement is not adversely affected by humidity, but the sensor can be damaged if inadvertently exposed to liquid samples. Commercial instruments, which allow the simultaneous measurement of oxygen and carbon dioxide levels in headspace samples, are available with sampling accessories for a variety of package containers and sample types (e.g., aluminum cans, beer bottles). There also exist commercial electrochemical oxygen analysis techniques that are based on the direct reduction of oxygen at a planar gold electrode (Clark techniques or modifications thereof). The latter may prove useful for specific headspace oxygen analysis problems, such as monitoring oxygen during a processing step.

Applications

The sample volume requirements of electrochemical methods limit the use of the analysis approach to pharmaceutical packages and sample configurations of larger dimensions containing ≥ 2 mL headspace volume, such as HDPE bottles, glass vials, and blister packages (individual blisters) of sufficient volume. The rugged character of such instruments and their ease of use make the technique particularly suited for use in a manufacturing setting.

Frequency Modulation Spectroscopy (FMS)

Principle of operation

Recently commercialized for package inspection applications, FMS is a high-sensitivity laser absorption technique that is useful for nondestructive monitoring of gas concentrations in optically transparent containers.^[5–8] Oxygen absorbs near infrared (NIR) light in a band of rotational transitions centered at 762 nm. Oxygen absorbance measurements, at this wavelength on typical NIR spectrometers that use incoherent white light sources, lack the sensitivity to provide a useful measure of headspace oxygen levels due to low frequency lamp intensity fluctuations. However, the measurement signal-to-noise (S/N) can be vastly improved (100–10,000X) by utilizing laser light sources and frequency modulation detection techniques. When a tunable diode laser is modulated using a radio frequency oscillator, the detection bandwidth can be shifted to high frequency where intensity

fluctuations, inherent to low frequency measurements, are minimized.

The measurement is conducted by frequency modulating a diode laser by superimposing a radio frequency oscillation, Ω , onto the diode injection current. The spectral output of a frequency modulated diode laser, shown at the top of Fig. 3, consists of a carrier frequency, ω_c and side band frequencies, $\omega_c \pm \Omega$. When the laser is scanned through the wavelength of oxygen absorbance, the amount of light absorption, which is proportional to the gas concentration, is “written” into the side band frequencies by recording the difference in absorption between the two sidebands. The differential absorption information is recovered using phase sensitive detection techniques. The demodulated absorption line shape is shown in Fig. 4. The amount of light absorbed is directly proportional to oxygen concentration. The gas density, n , is related to the peak-to-peak signal amplitude, ΔI , by Beer's law, which for a weakly absorbing molecule, is given by:

$$n = (\Delta I \pi / I_0 S x) \text{cm}^{-3}$$

where ΔI is the change in intensity of light after passing through the container (W cm^{-2}), Γ the full width at half maximum of the absorption signal (cm^{-1}), π a constant, I_0 the incident laser intensity (W cm^{-2}), S the integrated absorption cross section ($\text{cm}^2 \text{cm}^{-1}$), and x the container diameter (cm). The measured density in a sample vial is referenced to a standard and displayed as a concentration (%).

Applications

Application of the technique is limited to headspace monitoring in packages, which are optically transparent (glass vials, ampoules, bottles—including colored glassware such as amber) or translucent (LDPE, but not HDPE packages). Other containers with various optical densities and corresponding transmittance properties could also be potentially sampled, but would need to be evaluated on a case-by-case basis. Packages are inserted into a sample holder that can be customized for cell volumes of ~ 1 mL–100 mL, which typically present cell pathlengths of 1 cm–3 cm.

The principal advantage of the method is that it can nondestructively analyze oxygen contents within sealed containers without change to the sample. Thus, a single package can be used to study oxygen concentration profiles at periodic intervals and under various storage conditions without the need to destructively insert a probe or remove the headspace contents for analysis. The value of this cannot be overstated in terms of absolute sample requirements, sample preparation time, and

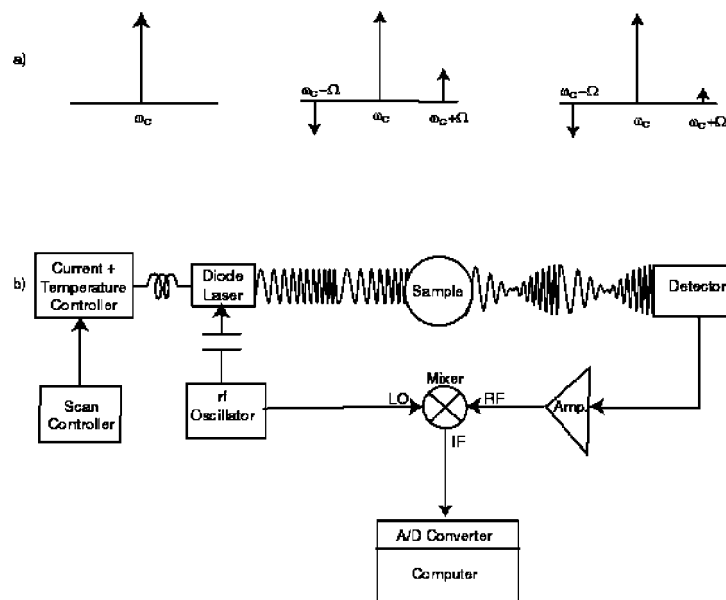


Fig. 3 a) Frequency and intensity profile of a diode laser beam, from left to right: unmodulated, modulated with no absorption, and modulated with absorption by the upper sideband. b) Schematic diagram of an instrument for FMS. The frequency modulated diode laser output is converted to an amplitude modulation after passing through a gas sample, which absorbs at a particular wavelength. The amplitude modulation is proportional to gas concentration and can be phase sensitively detected and related to oxygen concentration. (From Ref. [12], p. 41. Copyright ©2002 Advanstar Communications Inc. Advanstar Communications Inc. retains all rights to this material.)

sample-to-sample variability. From a quality assurance perspective, substantial cost-savings could also be realized from a nondestructive analysis approach. Development of on-line instrumentation would also be of significant value

for performing 100% inspection of the headspace oxygen contents of some parenteral packages.

Fluorescence Quenching Methods

Principles of operation

Oxygen sensing technologies based on luminescence quenching are also a relatively new commercial development and the scientific underpinnings for their operation is described extensively in a recent review article.^[9] The purpose of this discussion will be to briefly introduce this oxygen measurement technology; the reader is referred to more specialized literature for more details. In one commercial configuration (schematic shown in Fig. 5), fluorescence is used to measure the partial pressure of oxygen by using a bifurcated optical fiber to transmit an excitation source from a blue/green LED ($\lambda_{\max} = 470$ nm) to a thin-film coating applied to the tip of the fiber. A ruthenium complex dispersed in the thin film is excited and produces a fluorescence emission maximum at 600 nm that is collected at the tip and carried back to a CCD detector. The intensity of fluorescence observed is inversely proportional to the amount of oxygen in a liquid or gaseous sample.

Luminescence quenching sensors can be explained by the following processes, where M represents

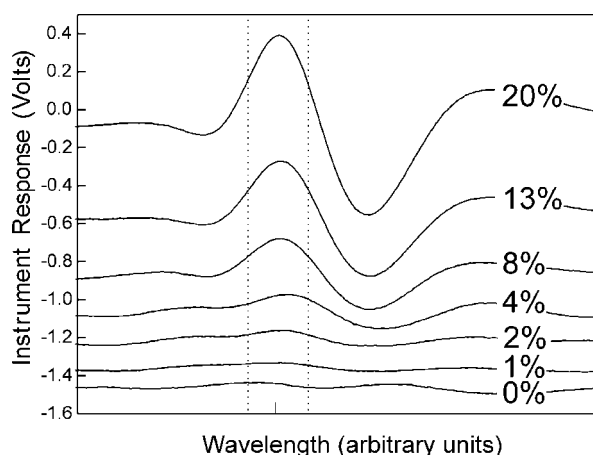


Fig. 4 Frequency modulation signals from oxygen absorption. The peak-to-peak amplitude of each spectra is proportional to oxygen concentration (noted to the right of each scan). (From Ref. [12], p. 41. Copyright ©2002 Advanstar Communications Inc. Advanstar Communications Inc. retains all rights to this material.)

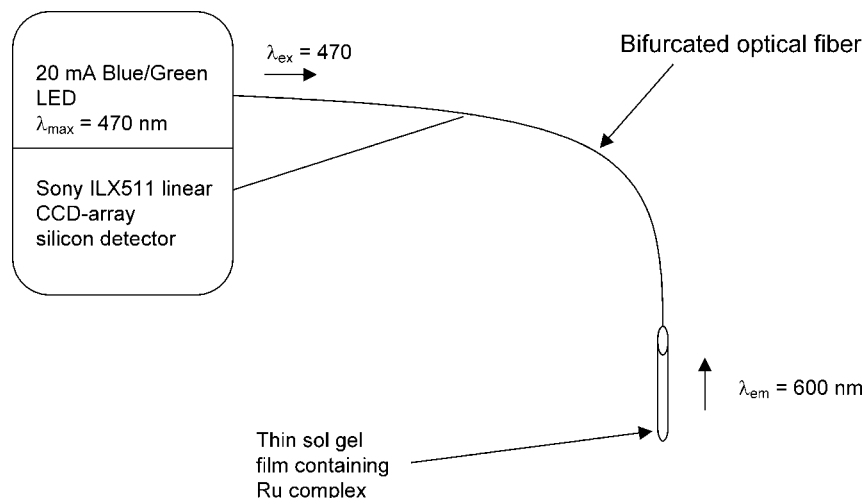


Fig. 5 Schematic of headspace oxygen analyzer based on fluorescence quenching. (From Ref. [12], p. 41. Copyright ©2002 Advanstar Communications Inc. Advanstar Communications Inc. retains all rights to this material.)

the luminescing molecule and Q is a quencher:

Photon Absorption (I_a) $M + h\nu \rightarrow M^*$

Luminescence (k_r) $M^* \rightarrow M + h\nu$

Nonradiative decay (k_{nr}) $M^* \rightarrow M + \Delta$

Dynamic Quenching (k_q) $M^* + Q \rightarrow M + Q^*$

Quenchers act by depleting the excited-state population (M^*), thus lowering the intensity of luminescent emission or shortening decay time. In the case of commercial oxygen sensors, M is a fluorescent ruthenium dye whose fluorescence intensity is quenched by oxygen in a manner commonly referred to as “dynamic fluorescence quenching.” In terms of the process described above, a collision of an oxygen molecule (Q) with the fluorophore in its excited state (M^*) leads to a nonradiative energy transfer and a loss of fluorescence intensity.

The amount of fluorescence intensity quenching can be quantitatively related to the partial pressure of oxygen in a sample from the simplified Stern-Volmer equation^[9]:

$$I_0/I = 1 + Kp_{O_2}$$

where I is the quenched fluorescence signal intensity, I_0 the unquenched fluorescence signal intensity, K the quenching constant that is related to particular fluorophore employed as M, and p_{O_2} the oxygen partial pressure.

An expanded version of the Stern-Volmer equation containing a factorial expansion of these same terms can also be employed to relate I_0 to oxygen levels in a more exact manner. Standard operation includes, at minimum, a daily calibration at a number of oxygen concentrations (at a fixed pressure and temperature) and use of the measured

calibration coefficients to determine the concentration of oxygen in an unknown sample.

Frequent recalibration is necessary since the sensors tend to degrade under normal operating conditions. This is particularly true of new sensors, which may require “burn in” over a few hours operation before stable operation is achieved. Analyses are accomplished by inserting a needle probe into a sample vial and waiting a fixed time for equilibration of the sensor to the headspace environment. The equilibration time is a function of the chemical composition of the protective overcoat applied to the sensor tip to exclude ambient light (a Teflon[®] AF coating was used in the present study). It is well known^[9] that the accuracy of fluorescence quenching is limited by resolution (random noise), deviations from the Stern-Volmer relationship, and calibration error. All these factors can promote measurement inaccuracy.

Applications

Fluorescence quenching sensors do not actually consume headspace oxygen and thus can be miniaturized for use on very small volume samples. However, the technique is still destructive as the tip of a 300 μm probe must still be used to puncture the package and come in contact with the relevant headspace, allowing an effective smallest volume that can be probed of about 100 μL . Some attractions of the technique include: 1) rapid response times (5 sec–2 min) that are limited to the time required for equilibrium to be reached between oxygen in the sample headspace and the sensing film by way of diffusion through the protective overcoat; 2) very low sample volume requirements ($\sim 100 \mu\text{L}$) that are limited only by

the contact dimensions of the probe; and 3) low cost and small instrument footprint. The low sample volume requirement makes it one of the two instruments capable (the other is headspace GC) of rapidly measuring oxygen levels in individual blister cavities of typical dimensions employed in pharmaceutical packaging. The small size and easy setup also make this instrument very attractive from a portability perspective for use in potential at-line packaging measurements. Fluorescence quenching instruments are also available in multichannel configurations that allow for the simultaneous measurement of several samples and the data logging component of the software allows for programmed measurements at fixed time intervals.

The various limitations of the method should also be noted. Care must be taken to calibrate across the intended measurement temperature range of samples, as fluorescence quenching levels change as a function of temperature. Temperature effects the measurement by impacting fluorescence decay time and the collisional frequency of oxygen with the fluorophore. Requirements to maintain the sample at $\pm 1^\circ\text{C}$ or to concurrently measure temperature while conducting the measurement curtail the practical utility of the approach. Experience in our own lab has also pointed out a need to consistently use longer signal integration times over the life of any given probe to achieve sufficient signal intensities over background noise. The decreased response factor observed could be due to changes in the ruthenium dye (e.g., degradation) over time. Higher signal intensities and/or

lower noise levels are desirable because the S/N ratio is proportional to the square root of integration time. Moreover, as can be seen from the Stern-Volmer equation, the rate of change of signal intensity with quenching will be largest at lower oxygen levels. Further development of the instrumentation for oxygen-specific measurements would greatly benefit the overall utility of the technique.

Quadrupole Mass Spectrometric Analyzers

Principles of operation

Quadrupole mass analyzers are by far the most common type of mass spectrometer in use today and the literature on these type of analyzers is extensive.^[10,11] Quadrupole mass analyzers are often thought of as mass filters because they can be tuned to transmit ions of a narrow range of mass/charge (m/z) ratios. Fig. 6 shows a generalized block schematic of a quadrupole mass spectrometer. A typical quadrupole instrument separates ions with different masses by application of a combination of static and radio frequency electric fields to four cylindrical rods. A headspace gas sample is introduced at an inlet and fed into an ion source where electrons are emitted from a filament and ionize the sample gas. The sample ions are then accelerated in an electrical field and are injected into the opening at the center of the rods. In the simplest systems, one pair of rods is connected and attached to the positive terminal of a d.c. power source, the other connected pair is

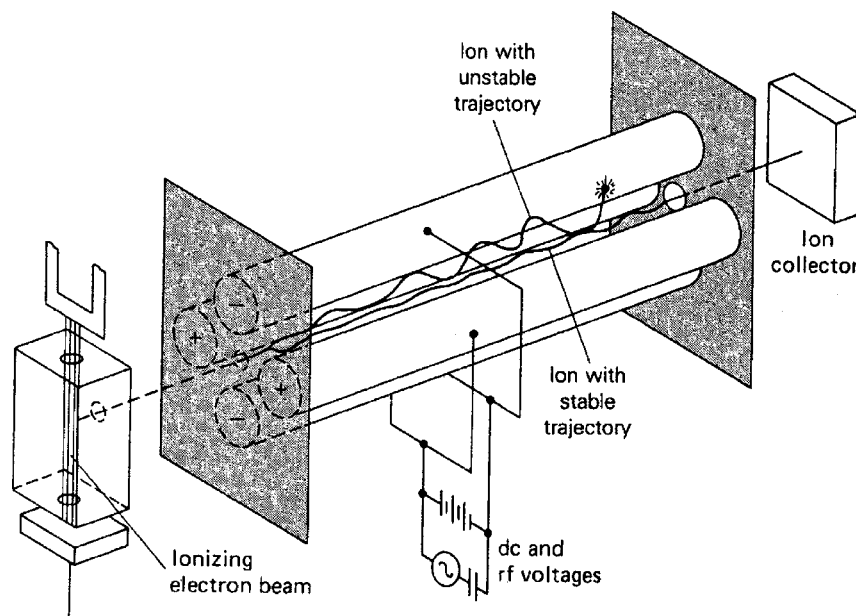


Fig. 6 Block schematic of a quadrupole mass spectrometer. (From Ref. [12], p. 41. Copyright ©2002 Advanstar Communications Inc. Advanstar Communications Inc. retains all rights to this material.)

attached to the negative end of the same source. A variable rf field with phase differences of 180° is also applied to each connected pair of rods with ac power supply. In order for an ion to travel through the quadrupole to the detector, the ion must have a stable trajectory in the presence of these applied voltages and electric field, as described by the so-called Mathieu equations of motion for ions in an electric field. For the purpose of this brief discussion, the application of a particular set of voltages to the rods allows mass filtering and the transmission of a band of ions having a limited range of m/z values. This is usually accomplished by scanning the rf voltage and d.c. voltages from zero to a predetermined maximum value while maintaining a fixed ratio. In this manner, a mass spectrum of analyte ions reaching the detector can be recorded at a Faraday cup detector.

Applications

The application of quadrupole mass analyzers in a residual gas analysis (RGA) mode has long precedence in the semiconductor industry, where residual contaminant gas levels in various ultrahigh vacuum chambers must be monitored and controlled. The principal advantage of using quadrupole mass spectrometry for pharmaceutical applications lies in the fact that many analytes can be simultaneously monitored and quantitated (2–300 m/z) using a single injection of headspace sample. Like other destructive techniques, the largest drawback for the instrumentation is sample introduction. A number of sample introduction strategies are currently under exploration to address this issue and improve the utility of the approach. The technique would likely be useful in process monitoring applications where the detailed contents of an inert gas blanket beyond that of simple oxygen levels are required.

A COMPARISON OF TECHNIQUES

The key to matching up a technique with a specific application is to evaluate the type of information desired, instrumentation availability, and the type of packaging configuration or sample under exploration. In MAP development, the studies may range from research into the oxygen permeability properties of various packaging systems, packaging system selection, process development and implementation to quality assurance. In the evaluation of various packaging systems, the questions may be more directed toward understanding the rates of oxygen ingress (or consumption) going from very low levels ($<1\%$ oxygen) to ambient oxygen concentrations. The types of tools used in the research laboratory setting should be re-evaluated when placed into a manufacturing setting for

process development and implementation. Lastly, in a QA/QC setting, other considerations such as long-term method robustness, stability, ease of calibration and use, oxygen levels to be routinely analyzed, and other practical considerations will help guide tool selection. The examination of tool selection criteria in the forthcoming discussion will be guided by: 1) performance considerations; 2) other relevant technique attributes; and 3) the type of package under consideration.

A recent research article provided a comparison of the techniques.^[12] Key findings from this study are summarized below:

(a) *Linearity*. All the techniques reviewed provide measurements with good linearity across the measurement range 0.00%–20.0%, all displaying $R^2 \geq 0.99$.

(b) *Accuracy*. Generally speaking, for most MAP programs the aim will be to reduce oxygen values from ambient to a predetermined specified level. Often, the measurement technique may not be required to give the absolute oxygen level in the container, but rather to demonstrate sufficient accuracy as to indicate that the oxygen content of the package is below the target threshold, thus serving as a limit test. Thus, the importance of the measurement technique to measure at lower levels (e.g., $<5\%$) is more important than capabilities at higher levels. In the study noted, the performance of the techniques was examined by comparing the absolute differences observed between theory and experiment for a low level oxygen standard (0.00%). The degree of deviation (from least to greatest) observed between predicted and observation for low oxygen level standards followed the trend:

Quad MS < Electrochemical < FMS < μ GC < GC

< Fluorescence Quenching.

(c) *Limit of Quantitation*. The lower limits of quantitation can be either estimated or obtained from the vendor. For critical absolute oxygen concentration measurements at very low levels, the lower limits of quantitation should be established on the specific system under study. Sampling procedure plays a critical role in the quality of the data for oxygen levels below 2% and limits the accuracy of the data.

Knowing the performance limitations of a particular instrument is key for interpreting the results obtained from a series of measurements. As discussed next, other considerations beyond performance also have bearing on technique selection. Table 1 attempts to summarize these additional points to consider when matching the analytical requirement to a headspace oxygen analysis tool. Considerations described here include analysis time, ease of calibration/use, commercial availability, expense,

Table 1 Summary of some of the attributes of various headspace oxygen sampling tools

Instrument	Analysis time	Ease of calibration/use ^a	Instrument availability	Expense ^b (K)	Required sample volume	Destructive
GC	Minutes	3	Commercial	40 +	10 μ L–30 μ L	Yes
μ GC	Seconds	2	Commercial	15–20	10 μ L–30 μ L	Yes
Electrochemical	Seconds	1	Commercial	4–10	1 mL–3 mL	Yes
Fluorescence quenching	Minutes	3	Commercial	2–5	None ^c	Yes
FMS	Seconds	2	Semicommercial	35–50	None ^d	No
Quad MS	Minutes	3	Semicommercial	25–50	10 μ L–30 μ L	Yes

^a 1 = easiest to calibrate and use, 2 = more difficult to calibrate and use, and 3 = most difficult to calibrate and use.^b Prices estimated are US Dollars.^c A sufficient sample volume, estimated at $\sim 100 \mu$ L, must be available so that a probe can be inserted, which makes contact with the sample.^d Instrument is nondestructive and can be used on vials in which a laser can be transmitted.

sample volume requirements, and sample introduction. As Table 1 shows, analysis times for the six techniques range from minutes (GC, fluorescence quenching, quadrupole MS) to seconds (μ GC, electrochemical methods, FMS). All the instruments are available commercially, although at various stages of development, and vary in expense from 2 to 50K (US Dollars). FMS and Quad MS are the least commercially developed of the techniques at this time and more work is needed before routine use will be possible. One highly subjective term, ease of calibration/use, is included in Table 1 and reflects the authors' judgments on the relative ease of training others to operate the equipment. Other issues such as versatility, portability, and suitability for use in a manufacturing setting are also important considerations when making a decision about which instrument to use. For example, headspace GC, μ GC, and quadrupole MS have a great number of uses beyond oxygen monitoring and there may be a desire to acquire other information simultaneously (moisture, carbon dioxide, other volatiles). The only nondestructive technique of which the authors are currently aware is FMS, which can only be used with package types that allow adequate transmittance of the instrument probe beam.

Lastly, another way to approach the decision about which tool to choose can be addressed from the perspective of which package type from which oxygen levels are desired. For the sake of brevity, only the three most common pharmaceutical packaging (plastic bottles, glass vials/ampoules, blister packages) configurations are discussed.

(a) Plastic bottles with foil induction inner seal: Plastic (e.g., HDPE) bottles with foil induction inner seals (most common are 30 mL, 75 mL, 120 mL volumes) are commonly used to package pharmaceutical products. For the rapid and accurate determination of headspace oxygen analysis in a package of this configuration, any of the six

techniques discussed in this report are applicable, with the exception of FMS. Electrochemical methods are the best suited for large volume (>2 mL) packages and display an unparalleled ease of use. These instruments also tend to be amongst the most rugged of all the instrumentation described herein and seem to be the most well-suited for use by operators with a limited amount of technical training.

(b) Glass vials and ampoules: Glass vials and ampoules are another commonly used pharmaceutical packaging configuration (especially for parenteral formulations). Any of the six techniques discussed in this report are applicable for analysis of headspace oxygen levels as long as sufficient volume is available. FMS offers considerable advantages for this package type since the approach is nondestructive. Moreover, FMS can be used in vials filled with a variety of media, including sample types such as powder that may coat the walls of the container. As long as a portion of the analysis beam is transmitted, the differential absorptive signal processing inherent to the technique can be used to yield oxygen concentration values. The potential for nondestructive, on-line 100% inspection of packaging container headspace is an intriguing possibility for the approach.

(c) Blister packages: The small volume of individual blister cavities make this the most challenging of the various pharmaceutical packages to analyze for headspace oxygen levels. The challenges are twofold: the extremely small volume available for sampling and sample removal/introduction without contamination. Amongst the instruments that are applicable for small volume measurements (headspace GC, fluorescence quenching, quadrupole MS), all suffer drawbacks as measurement tools for this purpose from a sampling perspective. While future development in this area is needed, headspace GC and μ GC appear to be most suitable at this time in that self-adhesive septa can be placed on the package backing



and a small portion of sample manually withdrawn for injection. Other viable approaches, including nondestructive techniques, are currently under exploration.

CONCLUSION

This entry has provided an introduction to the principles of six measurement approaches for analyzing the headspace oxygen levels in various pharmaceutical packages. Various performance criteria have been applied to examine the fit of each particular tool to typical pharmaceutical package types. Future research on the topic would help address better the robustness of sample introduction techniques for small volume samples and perhaps the development of sensor technologies that would allow oxygen levels to be remotely interrogated and profiled over time from a small device integrated or inserted into the package.

ACKNOWLEDGMENTS

The authors wish to thank Yieng-Hau R. Han, Rajiv Mahajan, and Rey T. Chern of Merck Research Laboratories for insightful discussion. In addition, a special thanks goes to Jim Veale and Bobby Anderson of Lighthouse Instruments, Inc. for reviewing the section on FMS.

REFERENCES

1. Johnson, D.M.; Gu, L.C. Autoxidation and Antioxidants. In *Encyclopedia of Pharmaceutical Technology*, 1st Ed.;

Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1997; Vol. 1, 415–450.

2. Connors, K.A.; Amidon, G.L.; Kennon, L. *Chemical Stability of Pharmaceuticals*; Wiley: New York, 1979; 80–98.
3. Kolb, B.; Ettre, L.S. *Static Headspace Gas Chromatography: Theory and Practice*; Wiley: New York, 1997; 1–55.
4. Wang, J. *Analytical Electrochemistry*; Wiley: New York, 2000; 350–379.
5. Bjorklund, G.C. Frequency Modulation Spectroscopy: A New Method for Measuring Weak Absorptions and Dispersions. *Opt. Lett.* **1980**, 5 (1), 15–17.
6. Janik, G.R.; Carlisle, C.B.; Gallagher, T.F. Two-Tone Frequency-Modulation Spectroscopy. *J. Opt. Soc. Am. B* **1986**, 3 (8), 1070–1074.
7. Silver, J.A. Frequency-Modulation Spectroscopy for Trace Species Detection: Theory and Comparison Among Experimental Methods. *Appl. Opt.* **1992**, 31 (6), 707–717.
8. Cooper, D.E.; Martinelli, R.U. Near-Infrared Diode Lasers Monitor Molecular Species. *Laser Focus World* **1992**, 28 (11), 133–146.
9. Demas, J.N.; DeGraff, B.A.; Coleman, P.A. Oxygen Sensors Based on Luminescence Quenching. *Anal. Chem.* **1999**, 71 (23), 793–801.
10. Drinkwine, M.J.; Lichtman, D. *Partial Pressure Analyzers and Analysis*; Whetten, N.R., Long, R., Jr., Eds.; American Vacuum Society: New York, 1978; 456–479.
11. Ferran, R.J.; Boumsellek, S. High-Pressure Effects in Miniature Arrays of Quadrupole Analyzers for Residual Gas Analysis from 10⁻⁹ to 10⁻²Torr. *J. Vac. Sci. Technol. A* **1996**, 14, 1258–1265.
12. Templeton, A.C.; Han, Y.-H.R.; Mahajan, R.; Chern, R.T.; Reed, R.A. Rapid Headspace Oxygen Analysis for Pharmaceutical Packaging Applications. *Pharm. Technol.* **2002**, 26 (7), 44–61.

HARMONIZATION OF PHARMACOPEIAL STANDARDS

Lee T. Grady
Jerome A. Halperin

United States Pharmacopeia, Rockville, Maryland, U.S.A.

INTRODUCTION

The USP and NF^a standards of strength, quality, purity, and packaging and labeling are recognized by the U.S. Federal Food, Drug, and Cosmetic Act and Amendments since 1906. These requirements are enforced by the Food and Drug Administration (FDA), a party in the harmonization of requirements for drugs.

Although originally founded as an organization to standardize medicines in the United States, the USP and its products and services are now known and utilized throughout the world. In today's transitional and global economy for pharmaceuticals, the USP has a strong international presence and influence. Economic forces are driving major trading parties to affiliate to reduce trade barriers. Integral to this process is harmonization of requirements, regulations, and standards governing the approval and marketing of drugs, devices, etc., by governments. The mission of the USP is to promote the public health through establishing and disseminating legally recognized standards of quality and information for the use of medicines and related articles by health care professionals, patients, and consumers.

This mission is not limited to the United States. Almost from its beginning in 1820, the USP has been aware of and part of international initiatives affecting pharmacopeial standards and their use by governments and professional organizations to control drug quality. That early commitment to internationalism has now grown to formal, on-going projects of harmonization with the pharmacopeias of Europe and Japan and agreements with pharmacopeias of Argentina, Brazil, and Mexico. International programs in drug information exist with a number of multinational organizations, foreign governments, and professional groups.

^aThroughout this article, the abbreviation USP, when used alone, signifies the United States Pharmacopeial Convention, Inc. The abbreviation *USP* in italics and followed by Roman numerals signifies a particular revision of the *United States Pharmacopeia*. The abbreviation USP–NF signifies the *United States Pharmacopeia–National Formulary*, two books in a single binding.

Harmonization is wanted strongly for the role of pharmacopeias in product registration exercises. That is, product development can proceed as is, without the later repeating of studies or testing to support registration in other than the original region. The primary beneficiaries, thus, are international companies. But harmonization has an independent value in facilitating international commerce in existing products, especially excipients.

Globalization

Twenty years ago, the vast majority of all drug substances and finished dosage forms used in the United States were prepared in the United States. For various reasons before globalization, the United States lost synthetic chemical operations due to stringent environmental requirements. Also, some manufacturing was lost because of inducements of low taxes by other governments to pharmaceutical manufacturers to relocate to their countries. Neither of these factors is the same as globalization. Globalization may be the single most important, historical *worldwide* trend at this time. The entire structure of international commerce and the allocation of capital and expertise are features of globalization. In this regard, the pharmaceutical industry is not particularly different from any other industry. Today's development of drugs or biological products may occur in the United States, Europe, or a combination of nations, and no one nation can be pointed to as to the innovator of a particular drug. Also, international companies would prefer to market the minimum number of formulations worldwide. But product registration of formulations is a very complex area, and the pharmacopeias are one evidence of the fact that different formulations may require different test methods. The end point of harmonization must be monographs that are acceptable to the registration authorities in different regions. Differences in pharmacopeial standards could be seized upon to create technical barriers to trade.

Table 1 List of Interpharmacopeial Open Conferences

Conference	Location	Date
Joint Pharmacopeial Open Conference on International Harmonization of Excipient Standards	Orlando, Florida, U.S.A.	January to February 1991
Interpharmacopeial Open Conference on Harmonization of Biotechnology-derived Products Standards	Verona, Italy	April 1993
Second Joint Pharmacopeial Open Conference on International Harmonization of Excipient Standards	St. Petersburg, Florida, U.S.A.	January to February 1994
Joint Pharmacopeial Open Conference on Sterility/Preservatives	Barcelona, Spain	February 1996
International Harmonization—General Monographs on Dosage Forms and Pharmacotechnological Test Methods	Seville, Spain	October 1998

Alternative Methods

The three regional pharmacopeias—the *USP*, the *European Pharmacopoeia (EP)*, and the *Japanese Pharmacopoeia (J)* allow an individual laboratory, able to do the official method, to validate an alternative method of analysis. The latter is chosen usually for speed, convenience, or economy but also to incorporate an existing database when a new or revised pharmacopeial method is adopted. Under those provisions, a laboratory can validate a method from another pharmacopeia, thereby avoiding duplication of routine work. In all three cases, only the official method could be used in compliance or contest. One point of harmonization is to avoid even the more remote instances of duplicative testing, in addition to international product registration.

There should be support at both the national and international community levels for pharmacopeial harmonization. There should be support both for harmonization of excipients and for harmonization of common general tests and assays. In doing so, one should prefer meaningful standards, not the lowest common denominator or the most stringent. Complicating the International Conference on Harmonization 1 was the attempt to be entirely consistent with the concept of forward harmonization.

Major support for pharmacopeial harmonization would come from increased cooperation and contribution to pharmacopeias on all the nonharmonization work. Harmonization takes away scarce resources from pharmacopeias, and there are other constituencies of the pharmacopeias to be served. This is an obvious consequence of the fact that pharmacopeial standards

apply to products already in the marketplace, both brand name and generic.

Harmonization has three essential values. The first is the facilitation of international commerce. The second is the facilitation of product registration processes in multiple nations. The third is to reduce duplicative testing costs. The facilitation of registration is for any new molecular entity—a onetime event for each country; whereas for the facilitation of international commerce and reduction, the duplicative testing remains throughout the lifetime of the product.

Pharmacopeial harmonization is challenging. The fact is that differences exist because of the different histories of the pharmacopeias. There are many factors. The most obvious are: content, language, legalities, speed, and the audiences for the standards. The *USP* applies also to the practice of pharmacy, both in a community pharmacy and hospital, and, thus, the standards set are appropriate to those environments.

There is an obverse to harmony and that is disharmony. An example of disharmony is the need to repeat tests using rabbits for pyrogen where testing for bacterial endotoxin is otherwise prescribed. This represents the most extreme disharmony of methods. But there is a greater disharmony; i.e., reaching different conclusions whether to pass/fail the specimen. In this case, the quality control professionals must make a judgment whether or not this material can be sold in one or more regions. Functionally equivalent to harmonization is the absence of disharmony. Because of a difference in policy, pharmacopeias may differ on adoption of a test. If certain tests are considered necessary by one pharmacopeia in order to protect the consumer, then it is appropriate for

that pharmacopeia to adopt the test without reference to any other region.

Reference Standards

Most discussions of harmonization revolve around excipients or general tests and assays. But the performance of even a harmonized method using different reference standards is not an optimal situation. In fact, harmonization of reference standards preceded many of the harmonization efforts of the last 10 years. Pharmacopeias and the World Health Organization (WHO) have, in the past, shared bulk materials to create their individual reference standards. Where a drug exists as a highly purified crystal, then the difference in pharmacopeial reference standards is administrative and legal in that no difference in results in laboratories is to be seen. This is not the case with mixtures, such as an antibiotic reference standard, which may be established based on different microbiological assays. Hormone records were part of the very earliest reference standard programs.

Biotechnology-derived products have led to renewed interest in establishing reference standards based on the same bulk of material. Thus, a single formulation, assay, and reference standard may be the fact worldwide. This situation can become complex, such as with insulin, where both biotechnology-derived insulin and animal-source insulin are in the marketplace at the same time.

HISTORY OF HARMONIZATION

One of the earliest references to USP's commitment to international harmonization may be found in the historical introduction to the 3rd revision (1) of the *USP* (1851): "The new Dublin and London Pharmacopeias were compared with our own, with a view of introducing uniformity wherever more important considerations did not seem to forbid the requisite modifications." Note that uniformity for its own sake was not the *sine qua non*.

Awareness of the Committee of Revision in 1851 of the importance of keeping the *Pharmacopeia* up-to-date may have been enhanced because of the enactment of the Drug Import Act in 1848, which mandated that drugs imported into the United States had to meet the standards of the country of origin and had to comply with the standards of the *USP* or one of the major European pharmacopeias. The U.S. Customs Service

established laboratories at major port cities and analyzed the imported drugs according to the declared standard. Ties to European medicine remained strong due to Americans traveling to foreign countries for study.

Harmonization of pharmacopeial standards as a practical matter began at the International Congresses of Pharmacy between 1865 and 1910 (2), but the first formal attempt can be traced to 1902. USP President Horatio C. Wood, M.D. and Frederick M. Power, Ph.D., an American chemist of the Wellcome Chemical Research Laboratories of London, were appointed by the U.S. Secretary of State as delegates to represent the U.S. government at the *International Conference for the Unification of the Formulae for Heroic Medicines*, a conference of 19 countries from Europe and North America (3). The second conference occurred in 1918; the third, in 1925, was attended by 31 countries from all continents except Asia and Australia, and was drafted a new "International Convention," which came in force in 1929. It revised the 1902 agreements on 77 "heroic" medicines and introduced the concept of maximum dose. It also requested that the League of Nations create a permanent secretariat of pharmacopeias (4). Andrew G. DuMez, Ph.D., represented the USP, and was officially appointed by the U.S. Public Health Service to represent the United States at this conference (4, 5). An expert committee of the League of Nations planned a third conference for 1938, but it was never convened because of World War II (2).

Other attempts to exchange information among committees for revision of pharmacopeias were attempted through the International Congresses of Pharmacy. Joseph B. Remington, Ph.D., Chairman of the USP Committee of Revision, attending the 1913 conference in The Hague as a delegate of the American Pharmaceutical Association, introduced a resolution to establish an International Bureau of Information to provide information to pharmacopeial revision committees in every country and to operate a testing laboratory. Remington was named to a committee to implement the plan as put forth by Prof. Alexander T. Schirch of Berne, Switzerland (6).

Latin America

Seeking to establish dialogues with Central and South America, the USP, in 1905, responded to a request for a Spanish edition of the *Pharmacopeia* by contracting with Dr. Jose Guillermo Diaz, Dean and Professor of the College of Pharmacy of the University of Havana, Cuba. Support for this project may have come from a

resolution adopted by the Second International Sanitary Convention of the American Republics in 1905, which read in part (3): “Resolved, that a translation of this *United States Pharmacopeia* into the Spanish language would prove of great benefit to the medical profession and pharmacists in each of the republics represented in this Convention.”

The Spanish edition of *USP VIII* was published in 1908. It was adopted by Cuba, Puerto Rico, and the Philippines. Addressing the Convention of 1910, Joaquin Bernardo Calvo, representing Costa Rica, stated (3): “offering to our physicians and pharmacists who do not speak English the *Pharmacopeia* of the United States translated into Spanish; it is one of the most useful works of its kind, if not the most useful, among those published up to the present date.” Spanish editions continued to be published through *USP XV* in 1955. The Spanish edition of *USP XI* was adopted as the official pharmacopeia first by Costa Rica.

In his report to the Convention of 1960, however, Secretary Adley B. Nichols stated (7): “The distribution picture of the *USP* in Spanish has not been satisfactory for some time, and this is especially the case with *USP XV*. In no country is there a marked demand for the translation. Apparently the English language is sufficiently widely known today to permit the use of the readily available English edition.” Perhaps the best summary of why the *USP* produced a Spanish edition of the pharmacopeia can be found in the words of Dr. Charles H. LaWall, Chairman of the Committee of Revision: “The publication of the Spanish edition can never be considered financially advantageous to the Convention, but it should be continued as a patriotic duty and in recognition of the in-use of the book in the Spanish-speaking American countries.”

The *USP* again published in Spanish in 1995, with semiannual supplements since then. This was now possible through “machine translation.” The situation today is mixed, but the English version maintains its importance in Latin America.

MODERN FORUMS FOR HARMONIZATION OF DRUG QUALITY STANDARDS

Forums for harmonization emerged immediately after a *USP* Open Conference in May of 1989 in Williamsburg, Virginia (8). It was concluded there that a thrust should be made toward harmonizing excipients and, possibly, test methods. This position was laid before two international meetings in 1989. The first in Strasbourg,

France, in June of that year, celebrated the 25th anniversary of the *EP* (9). The second, in September, in Tokyo, Japan, sponsored by the Pharmaceutical Manufacturers Associations of Tokyo and Osaka, focused on a theme of drug quality and the role of the pharmacopeias in the year 2000 (10). At both of these well-attended conferences, the representatives of the industry spoke of the need for harmonization of standards among the pharmacopeias representing the major drug discovery and drug manufacturing areas of the world—the United States, Europe, and Japan—to facilitate international commerce in pharmaceuticals.

The areas of pharmacopeial standards most frequently cited at those meetings as in need of harmonization were pharmaceutical excipients and analytical tests and assays. Excipients posed the greatest barrier to commerce as a result of a patchwork of standards in the *USP*, *EP*, and *JP* for a small universe of substances and many natural products of animal, mineral, and vegetable origin that are shared throughout the world. Standards for these common substances reflect cultural, scientific, and temporal differences in how and when these standards were established and last revised. Similarly, differences in tests and assays among the three compendia frequently resulted in situations where meeting the standards of a pharmacopeia in one sector would not predict meeting the standards for that same substance in another sector. That is, testing for the same parameter by another method, often resulting in extra expenditures for capital equipment for laboratories, as well as extra time and resources for conducting the tests and in maintaining trained analysts for different procedures.

PHARMACOPEIAL DISCUSSION GROUP

Founded in Tokyo in September 1989 as the “Quadripartite Group,” the Pharmacopeial Discussion Group (PDG) was originally composed of representatives of the BP, EP, JP, and USP. The current group includes members from the EP, JP, and USP. At its first meeting at the USP headquarters in Rockville, MD, in March 1990, important agreements were reached:

- To meet twice yearly in a small group consisting of the senior staff executive and scientific officers of each pharmacopeia.
- To implement the concept of forward harmonization, which was agreed in Tokyo (10). *Forward harmonization* has three characteristics: a preference for the selection of methods that would be acceptable well into

the future; retaining of any standard meaningful to an individual pharmacopeia; and unilateral progress not inhibited by trying to have every new advance occur simultaneously in every pharmacopeia.

- To include two additional concepts of harmonization: *prospective*—to avoid conflicts among pharmacopeial standards before they occur, and *retrospective*—to resolve conflicts among existing pharmacopeial standards.
- To solicit advice from the pharmaceutical industry and government regulatory agencies on candidate articles for harmonization and their relative priorities.
- To convene open international pharmacopeial conferences.

Articulating the three concepts for harmonization was particularly important. Prospective and retrospective concepts clarify the distinction between work required to avoid conflict when establishing standards for pharmacopeial articles (for which standards do not exist or where few standards exist among the pharmacopeias), from work required to reconcile differences among well-established standards for articles that may have been in the pharmacopeias for considerable time. Prospective harmonization was inaugurated for biotechnology-derived products. Retrospective harmonization focused on pharmaceutical excipients and analytical tests and methods. Forward harmonization expresses a philosophy and environment for harmonization consistent with advances in pharmaceutical analysis.

Establishing a process for harmonization requires recognizing that each pharmacopeia is a sovereign entity, and has certain authorities and obligations derived from the legislation or treaty that created it that harmonization processes must recognize. Complicating the process was the realization that, different from anything that had been attempted before, a forward-moving process was being devised involving the revision systems of three pharmacopeias, each having evolved in different cultures and histories over periods ranging now from 30–180 years, resulting in profound differences in pharmacopeial policies.

Importantly, this was a voluntary effort; there was no legislative or treaty mandate to harmonize. In fact, no organization can compel harmonization. Complicating the process are the differences worldwide trend among the times for appearance of a first monograph, revision publication schedules, public notice and comment opportunities, and updating provisions. The ideal system would allow closely concurrent, if not simultaneous, actions by each pharmacopeia. The realities of level of funding, publication, and acceptance procedures, however, fall short of ideal. As a matter of fact, harmonization takes

resources away from all other pharmacopeial programs because no specified support is received.

The first attempt at soliciting advice for pharmacopeial priorities was the joint issuance of a letter in May 1990 by the USP in English, by the EP in English and French, and by the JP in Japanese (11). It went to the regulatory agencies and pharmaceutical industry associations in the countries and regions served by each of the pharmacopeias. Reflecting the sentiments of the speakers at the 1989 conferences, the letter was devoted to pharmaceutical excipients and asked only three questions: 1) Which excipients have been a source of problem or delay? 2) Has it been necessary to repeat stability or bioavailability studies because of differences in standards for excipients? 3) What are candidates for the top 10 excipients for harmonization? It also asked respondents to identify those specifications, tests, and assays in the monographs for these excipients that are most important to be harmonized.

Responses were returned from individual pharmaceutical companies through their industry associations and the regulatory agencies in the respective sectors. The complete response was compiled by the USP, and a list of the top-10 excipients was developed as the focal point for initial harmonization efforts, after review by each pharmacopeia. The idea of a “*lead*,” subsequently called the *coordinating pharmacopeia*, was adopted, which would take leadership for the revision of the monographs for the excipients for which it had volunteered. The initial list of assignments included: magnesium stearate, microcrystalline cellulose, lactose, starch, cellulose derivatives, sucrose, povidone, stearic acid, calcium phosphate, and polyethylene glycol (12).

The three pharmacopeias have periodicals in which the respective publics are informed of any proposed changes or additional standards (13). Standards do not get out of synchronization through ignorance, because the other two pharmacopeias are familiar with upcoming harmonization-related text in each other's periodicals. And the industry by and large subscribes to all three periodicals and, therefore, should be kept abreast of developments in harmonization. The disconnect arises out of the working speeds and legal procedures of each.

Prospective harmonization is particularly successful when dealing with biotechnology-derived products (14, 15). That is because there are only a couple of manufacturers involved. There should be no reason for the pharmacopeias to arrive at different standards proposed for any particular medicine. This is in stark contrast to the situation where there are multiple manufacturers of drugs no longer covered by patent protection. There is no possibility of harmonization of the some 4000 monographs for individual substances and preparations.

The USP must pay strict attention to the legal situation in the United States. Here, one cannot write "lock-out specifications," thereby keeping somebody out of the business of pharmaceuticals. The other two pharmacopeias, confronted with the same situation, came to the same place in due course, but not at the speed that the USP demonstrates in avoiding any possibility of lock-out specifications. Where there is a valid medical or pharmaceutical reason for specifications that lock out competitors, then the pharmacopeia will set such standards.

One expression frequently heard is to "essentially harmonize." To essentially harmonize is the practical limit of what is possible. It is necessary to rate harmonization on a scale from 0 to 100% harmonized. But in passing judgment, an old expression pertains: "Where you sit is where you stand." The USP's scale for harmonization takes the point of view of a laboratory supervisor who schedules work, training, and capital goods purchases. This would seem to be the point of view that is of most practical value. To assign a quantitative characterization, one must make judgments as to the significance of differences. A completely harmonized requirement would use the same method and establish the same limits. Another, slightly less harmonized requirement would use the same method, but two pharmacopeias would have different limits where one set of limits is nested within the other. If those limits are not nested, then there is a degree of disharmony. Summaries of the overall state of excipient harmony, both from the PDG list and of 200 excipients that appear in *NF* and elsewhere, have appeared in *Pharmacopeial Forum* (16).

The reason a scale for harmonization is necessary is the fact that each pharmacopeial monograph may contain some 10 requirements, and individual monographs will be harmonized on perhaps 7, 8, or even all 10 of those requirements. So where 8 requirements out of 10 are harmonized, it is reported as an 80% harmonized monograph (16).

The goal of harmonization is to bring the policies, standards, monograph specifications, analytical methods, and acceptance criteria of pharmacopeias into agreement. We recognize such unity may not always be achievable. Where unity cannot be achieved, harmonization means agreement based upon objective comparability and a clear statement of any differences. The goal, therefore, is harmony, not unison.

INTERPHARMACOPEIAL OPEN CONFERENCES

Copresented by the BP, EP, JP, and USP, the first Interpharmacopeial Open Conference on Standards for

Excipients was convened in Orlando, Florida, from January 30 to February 1, 1991. Attended by 165 participants, representation included 11 countries, 59 pharmaceutical or excipient manufacturers or suppliers, 3 regulatory agencies (FDA, EEC, and HPB), and 7 pharmacopeias (the copresenters and the French, Italian, and Spanish Pharmacopeias) (17). In preparation for this conference, the USP convened open meetings on magnesium stearate and lactose, attended by almost every major manufacturer from Europe and the USP.

The conference endorsed the goals of the pharmacopeias to improve and harmonize standards for existing excipients, to focus on testing methods and address specifications after test methods had been agreed upon, and to develop functionality tests, including particle size, surface area, and density.

Implementation of the open conference recommendations led to a change in configuration of the membership in the pharmacopeial harmonization process. Whereas the BP had been an independent member of the Quadripartite Group from its inception, implementation of harmonization of standards and tests for excipients was recognized as a regional matter under the aegis of the EP, and the BP's independent membership in the process ended. The resulting group of the USP, JP, and EP became known as the PDG and has continued its efforts in tripartite configuration.

A second Joint Pharmacopeial Open Conference on International Harmonization of Excipients was held in St. Petersburg, Florida, in January, 1994 (18). Major progress was achieved and established the principle that conferences were the key component of harmonization. Progress on pharmacopeial harmonization was reported at the International Conference on Harmonization in Brussels in November 1991, and activity of the PDG to fulfill its goals is high. Several tangible and important milestones have been reached. First and most important is the recognition by the JP of its need to develop a vehicle for public notice and comment for pharmacopeial revision, and the essentiality of such a vehicle to the communication process. Recognizing that need, the JP reached a decision to publish the *Japanese Pharmacopoeial Forum* (JPF) on a quarterly basis. The first edition in January 1992 (11) included proposals for the revision of magnesium stearate and lactose monographs. Notable also is the fact that matters in JPF relating to international harmonization are printed in English, whereas domestic revision matters appear in Japanese.

A second milestone was a letter by the PDG in May 1992 soliciting further candidates, beyond excipients. Responses to that inquiry focused primarily on tests and assays. Replies were ranked by order of priority. The priority of

excipients was expanded to the top 25, based on further analysis of responses. The lists of combined assignments and priorities for pharmacopeial harmonization appeared in the forum publications of the pharmacopeias (12, 13).

JP took another step toward harmonization by announcing that it would implement an annual supplement program to update the *JP* between editions, beginning with an October 1993 supplement to *JP XII* (11).

Refinement of the process of pharmacopeial harmonization occurs ongoing (see Fig. 1 and Appendix which details the process as of December 1999). Accommodating the revision processes, time requirements, and publication schedules of three revision systems and nine publications proved infeasible as initially envisioned. It required continual adjustment as issues reach stages in revision that had not been foreseen or did not fit within existing systems. Also complicating progress was that, although the USP in 1980 established the first expert group that focused on excipients (Subcommittee on Pharmaceutical Ingredients), the other pharmacopeias did not have task groups readily at hand devoted specifically to excipients.

Experience gained by harmonizing the first of the excipients (lactose and magnesium stearate) showed that, because so many parties are affected and multiple expert groups must be convened, forward, retrospective harmonization is intrinsically a multiyear process.

TESTS AND STANDARDS

Impurities in Excipients

Limit tests have a long standing in pharmacopeias. For some (heavy metals is an example), the sensitivity of the method was the basis for the standard. Modern limits in the *USP–NF* are toxicity based. There is divergence in harmonization because of toxicity-based rather than method-based standards. The modern basis avoids the exclusion of safe products from the marketplace, whereas the older approach could lead to lock-out specifications considered technical barriers to trade.

Biotechnology-Related Standards

Implementation of prospective harmonization began formally with a second conference on standards for biotechnology-derived products. The conference was attended by about 150 scientists and regulators from 20 countries. With participation by experts on pharmacopeial revision bodies from each of the three pharmacopeias, the conference produced a series of recommendations relating to informational chapters, general chapters on tests and

assays, and group and individual monographs for selected biotechnology-derived human drugs and biologics (14).

The introduction of biotechnology-derived products presented a decisive moment (15): an opportunity to avoid conflicting standards through the commitment of the pharmacopeias to the process of harmonization for pharmacopeial standards for an emerging technology. Included in this opportunity are the practical values obtained from common reference standard materials.

The complexity of the technology, in concert with medical conditions, and clinical environments, and the desire for instant globalization of return on investment call for facilitation and avoidance of ambiguities by uniform standards worldwide.

Biotechnology-derived drugs are of mutual interest to the USP and the International Conference on Harmonization on Drug Quality (ICH-Q6). The relationship is uncertain between compendial standards and the development and approval of biotechnology-derived drugs by regulatory agencies. The *USP* contains many general chapters (i.e., “horizontal standards” such as stability, injections, and bacterial endotoxins), which are cited extensively by the biotechnology industry, and an informational chapter on biotechnology-derived products that explains terminology and facilitates communication.

Status of Interpharmacopeial Harmonization

Supplements to *USP24–NF19* contain updates of an informational chapter by this name. It lists all of the projects undertaken by the PDG. Appendixes list the projects as of January 2000, and identify the coordinating pharmacopeia. However, these do not report the official status of each project because this changes with each supplement.

Because revision programs work on different schedules, one should compare only the current texts of two or more pharmacopeias whenever divergence of mandatory requirements is an issue. The official pharmacopeial texts should be compared.

Harmonization proposals do not have official status. The work of the PDG is finished at Stage 5B. The progress of the harmonization projects can and should be verified in the most recent number of the periodicals: *Pharmacopeial Forum*, *Japanese Pharmacopoeial Forum*, and *Pharmeuropa*. For standards in force in *USP24–NF19*, see also the latest supplement or Interim Revision Announcement.

Equivalence

Tests and assays of the EP and the JP that have been elaborated through the PDG Procedure are considered as

equivalent to *USP–NF*, except as noted in Chapter <1196> entitled *Status of Interpharmacopeial Harmonization* found in supplements to *USP24–NF19*. Because the legal status of each may not be at the same stage, a precautionary check should be made to support any plans or actions.

Equivalence is attributed to those monographs, tests, or assays that have arrived at Stage 5B. The nature and reason for divergences are expected to be described in the three pharmacopeial periodicals. Only those stated exceptions are considered nonequivalent.

INTERNATIONAL CONFERENCE ON HARMONIZATION

Founded in 1990, the International Conference on Harmonization (ICH) is comprised of the pharmacopeial manufacturers associations in Europe (EFPIA), Japan (JPMA), the United States (PMA), and the drug regulatory agencies in Europe (EEC), Japan (MHW), and the United States (FDA), with the International Federation of Pharmaceutical Manufacturers Association (IFPMA) serving as secretariat. Pharmacopeias are not members of the ICH, where membership is reserved for three PMAs and three regulatory agencies. Invited observers include Canada, WHO, and the European Free Trade Association (EFTA).

With expert working groups in the areas of drug efficacy, drug safety, and drug quality, the ICH is the foremost opportunity for harmonization among the leading drug regulatory and manufacturing groups in the world. The ICH Expert Working Group on Drug Quality (EWG-Q) includes the topic Q4, “Pharmacopeias.”

The pharmacopeias have worked with the ICH process to facilitate the international environment of pharmaceutical research and product registration. On the other hand, the additional situation for compendia is that the standards that they have published now apply to all of the products already marketed. In that case, a company has testing history in their quality control (QC) departments. The QC departments are the most conservative elements within the pharmaceutical industry—an attribute necessary to their task. QC departments are reluctant to change methods when they feel that their products are properly represented by their current suite of tests. Thus, there is resultant tension between trying to develop harmonized standards that facilitates one area of activity in the world of pharmaceuticals and not disturbing a satisfactory marketplace. A vast amount of progress has been made in harmonization of pharmacopeial methods.

At the first biennial meeting of the ICH in Brussels in November 1991, the United States, Japanese, and

European Pharmacopeias presented papers relating to progress being made in the harmonization of pharmacopeial standards for excipients (19). Other topics, such as stability, validation, impurities, and biotechnology, were established. The pharmacopeias are involved in all these issues. In fact, *USP* general chapters served as background for harmonization for some of these topics.

The ICH Steering Committee responded favorably to a request by *USP* for *observer status*, recognizing it as a nongovernment, nonindustry body with official status under U.S. statutes. Now, each of the three pharmacopeias can participate in EWG-Q activities and in ICH biennial meetings as independent bodies.

Stability is a key quality concern that is addressed in various ways by pharmacopeial standards. It was also the first subject for which the EWG-Q developed a guideline. The *USP* has redefined the concept of *Controlled Room Temperature* in terms of a Mean Kinetic Temperature of 25°C, which is identical with the long-term storage temperature promoted by ICH. Furthermore, ICH has advised against including recommended storage label statements that could conflict with United States, or other regions’ practices. Thus, the *USP* standards and the ICH guideline agree on this overarching concern for the stability of pharmaceuticals. The *USP* actively participated in achieving this desirable outcome.

Validation of analytical procedures is intrinsic to both new drug approval and compendial revision. The *USP* had already established an informational chapter, <1225> entitled *Validation of Compendial Methods* before the international harmonization effort began—a joint effort among the Pharmaceutical Research Manufacturers Association, the FDA, and the *USP*. It was useful in the work of the EWG-Q to prepare a document on validation of analytical procedures that concentrated on the submission of a new drug to the reviewing authorities. Later, in response to demands of users, the ICH document was expanded to more readily meet the scope of *USP* <1225>. Differences that arose were resolved, thereby securing the harmonized situation. The emergent ICH document and the *USP* chapter are harmonized in breadth and detail in such a way that the vocabularies of validation and the underlying analytical strategies are in concert.

A recommendation by the PDG to establish a harmonized procedure for stability to light was taken up by the EWG with Japanese participants responsible for a first draft of a guideline. A harmonized procedure emerged, again for the purpose of new drug approval.

Impurities are of many kinds, and, therefore, the issues for harmonization are numerous. Everyone concerned agrees that toxic impurities must be controlled at very low

levels, and the analytical difficulties in measuring low levels limit choice. It follows that harmonization of methods is straightforward, once the objectives of analysis are laid out. Both, the USP and ICH limit measurement to impurities at or above 0.1%, which is of practical significance in everyday commerce in bulk pharmaceutical chemicals. In addition, the ICH guideline refers to the pharmacopeial limits on toxic impurities, such as heavy metals. Thus, on all critical issues, there is no conflict between the ICH guideline for new chemical entities and established USP impurity policies, which apply to hundreds of drugs already on the market. The USP further identifies as signal impurities those that are distinctly informative of the purification or decomposition of a drug substance. The remaining impurities are considered in nonspecific categories of ordinary impurities and a labeling requirement for other impurities and are limited to 2% total. ICH sets no such ceiling limit. These policies do not require disclosure of proprietary synthesis or purification details, yet they accomplish the necessary task of limiting bias, thereby assuring meaningfulness of tests and assays. But here ICH guidelines demand identification of each impurity and of individual limits, requiring detailed lot-to-lot bookkeeping on all impurities. Impurities in excess of 0.1% are to be “qualified” (i.e., toxicity considered); USP policy requires notification to USP of any known toxic impurities.

For some years, the USP had a requirement on organic volatile impurities. When the EP was adopting a similar requirement, it was clear that toxicologists on different sides of the ocean would come to different conclusions as to appropriate limits on solvents. At the request of the pharmacopeias, this topic was taken up by (EWG-Q). The effort was successful, and a guideline on residual solvents emerged. Relative to the existing USP limits, it was necessary to revise some upward and others downward, but none to any great extent. This new effort was 15 years subsequent to the initial exercise. However, there remains an unresolved solvent and that is benzene.

Specifications Documents

The ICH–EWG–Q produced two guidelines, Q6A and Q6B, dealing with specifications to support a new drug registration. Biotechnology-related specifications are treated by Q6B, and all others that were previously the subject of EWG-Q guidelines are treated by Q6A. Pharmacopeial methods are intrinsic to these guidelines.

ICH–EWG–Q developed a list of 12 general chapters (tests and assays) that were deemed critical to new product registration and urged the PDG to concentrate on prompt

harmonization (see Appendix III). One of these, *Antimicrobial Preservatives Effectiveness*, was dropped in 1999 when nonharmonizability was clear because of differences in the essence of utilization of the same microbiological procedure. Of the 11 remaining chapters, all but two (*Microbial Limits of Non-sterile Articles* and *Dose Uniformity*) were harmonized with regard to scientific content by the end of 1999; only the necessary publication sequences were unfinished.

The Q6A and Q6B documents were preceded by the North American Conference on Specifications. In all cases, no method selection, scope of application, or overall policy (dissolution, impurities, particulate matter, etc.) is at odds with USP—a remarkable situation in view of the breadth of the topics covered. The main difference was and is the proportion of active ingredient in a formulation that triggers choice of determination of content or of weight to establish uniformity of dosage units.

Concordance

In the future, harmonization could be accelerated by reference to laboratory data, rather than by trying to achieve harmonized texts, tests, or assays that establish the same attribute of an article (i.e., water by titration vs. loss on drying). Inherent in this allowance is the assumption that concordant methods can be shown to yield comparable outcomes with regard to acceptable identification, strength, quality, purity, bioavailability, or labeling in the context of the monograph of a recognized article. In the event of a dispute, however, only the result obtained by the procedure given in the appropriate pharmacopeia is conclusive.

This concordance rests on the probable presumption of good manufacturing practices in production and control, now a reasonable presumption in today’s international environment.

The use of concordant methods does not necessarily require identical reagents, procedures, or measurements. Official procedures of pharmacopeias, per se, require no validation, but validation of the applicability of official procedures to each preparation (formulation) is to be presumed (e.g., the presence of interfering ingredients).

Concordance would facilitate international commerce in official articles by allowing the reduction of duplicative testing or delay in national product registration or approval proceedings. No provision of the USP General Notices is abrogated, and the allowance, therein, for alternative methods is not foreclosed by a monograph citation to the concordance.

As stated in the General Notices, an article is recognized in *USP–NF* when a monograph for the article

is published in it. Each monograph contains standards that define an acceptable article, and gives tests and assays and other specifications designed to demonstrate that the article is acceptable. Monographs and their interpretation are subject to the provisions of the General Notices and general chapters. The tests and assays required in a single monograph are intertwined, therefore, and may further state any necessary variation from the designated general chapters or the General Notices. That is, pharmacopeial monographs define interdependent attributes, each drawn in light of other monograph requirements and the General Notices and must be viewed in total. Consequently, allowances for concordance would only be utilized when supported by an initial verification through duplicate testing of the same specimens and evidence that both methods were validated.

APPENDIX I: WORKING PROCEDURES OF THE PDG

Stage 1: Identification

PDG identifies subjects to be harmonized and nominates a coordinating pharmacopeia for each subject.

Stage 2: Investigation

The coordinating pharmacopeia for a subject to be harmonized collects the information on the existing specifications in the three pharmacopeias, on the grades of products marketed, and on the potential analytical methods. For new products or new methodologies, existing information in the scientific literature or from manufacturers is collected and analyzed. The coordinating pharmacopeia prepares a draft monograph or chapter, accompanied by a report giving the rationale for the proposal, with validation data where appropriate and available. Stage 2 ends with the proposal draft, which is mentioned in this procedure as the Stage 3 draft. The Stage 3 draft, accompanied by supporting comments or data that explain the reasons for each test method or limit proposed, is sent by the coordinating pharmacopeia to the secretariats of the other two pharmacopeias.

Stage 3: Proposal

The three pharmacopeias publish the Stage 3 draft in the next available issue of their forums, in the style provided by the coordinating pharmacopeia. If necessary, questions are addressed to the readers of the forums when specific

issues require their advice, information, or data. In *Pharmeuropa* and the *Japanese Pharmacopoeial Forum*, the Stage 3 draft is published in a specific section entitled *International Harmonization*. In the *Pharmacopeial Forum (USP)*, the Stage 3 draft is published in the *Pharmacopeial Previews* section. The draft is published in its entirety. The corresponding secretariats may have to add information needed for the understanding of implementation of the texts, e.g., the addition of the description of an analytical method or of reagents that did not exist in the pharmacopoeia. Comments by readers of these forum resulting from this preliminary survey are to be sent to their respective pharmacopeial secretariat, preferably within 4 months of publication. The period for public review and comment should not exceed 6 months, however. Each pharmacopeia analyzes the comments received and submits its consolidated comments to the coordinating pharmacopeia within 2 months of the end of the public review/comment period. The coordinating pharmacopeia reviews the comments received and prepares a harmonized document (Stage 4 draft) accompanied by a commentary discussing comments received regarding the previous text and providing reasons for action taken in response to those comments. The Stage 4 draft together with the commentary is sent to the secretariats of the other pharmacopeias (end of Stage 3).

Stage 4: Official Inquiry

The Stage 4 draft is published in the *forum* of each pharmacopeia, with the style adapted to that of the pharmacopeia concerned. In *Pharmeuropa* and the *Japanese Pharmacopoeial Forum*, the Stage 4 draft together with the commentary is published in a specific section entitled *International Harmonization*. In *Pharmacopeial Forum*, the Stage 4 draft is published in the *In-Process Revision* section. Comments regarding this draft are to be sent by readers of the forum to their respective pharmacopeial secretariat, preferably within 4 months and at most within 6 months of publication in the forum. Each pharmacopeia analyzes the information received and submits its consolidated comments to the coordinating pharmacopeia within 2 months of the end of the review/comment period. The coordinating pharmacopeia reviews the comments received and prepares a draft harmonized document (Stage 5A draft), accompanied by a commentary discussing comments received regarding the previous text and providing reasons for action taken in response to those comments. The Stage 5A draft, together with the commentary, is sent to the secretariats of the other two PDG members (end of Stage 4).

Stage 5: Consensus

- a. *Provisional.* The stage 5A draft is reviewed and commented on by the other two pharmacopeias within 4 months of receipt. The three pharmacopeias shall do their utmost to reach full agreement at this stage, with a view to reaching a final consensus. If the consensus is reached, a Stage 5B draft is developed. In those rare instances where consensus is not reached or where novel, unanticipated serious issues are identified by any of the parties and need to be considered, depending on the complexity and gravity of the issues, more than one mechanism for resolution may be adopted. This includes the coordinating pharmacopeia call for a meeting of experts from the three pharmacopoeias to search for a consensus. This group prepares a modified 5A document (5A-2) that will be published in the three forums for public comments. These are then reviewed by the group of experts that finalized the consensus document (5B).
- b. *Final.* When consensus has been reached, the Stage 5B draft (consensus document) is sent by the coordinating pharmacopeia to the other pharmacopeias for final sign-off.

Note: The last two stages of the implementation of the “harmonized” chapters and monographs take place individually, according to the procedures established by each pharmacopeial organization.

Stage 6: Adoption

The document is submitted for adoption to the organization responsible for each pharmacopeia. Each pharmacopeia incorporates the harmonized draft according to its procedure. If necessary, the Stage 5B draft can be adopted, with specific amendments identified as such, corresponding to a general policy in the territory of the pharmacopeia in question. The monographs may, therefore, be harmonized, without being identical in every respect. Adopted texts are published by the pharmacopeias in the Supplements or, where applicable, in a new edition/revision. If a consensus has not been reached at Stage 5A, the pharmacopeias prepare together an article on divergences to be published by all three pharmacopeias in the respective forums.

Stage 7: Date of Implementation

The pharmacopeias will inform each other of the date of implementation in the particular region.

APPENDIX II: EXCIPIENT HARMONIZATION

Excipient	Coordinating pharmacopeia
Alcohol	EP
Benzyl alcohols	EP
Dehydrated alcohol	EP
Calcium disodium edetate	JP
Calcium phosphate, dibasic	JP
Calcium phosphate, dibasic (anhydrous)	JP
Carboxymethylcellulose, calcium	USP
Carboxymethylcellulose, sodium	USP
Carboxymethylcellulose, sodium (cross-linked)	USP
Cellulose (microcrystalline)	USP
Cellulose (powdered)	USP
Cellulose acetate	USP
Cellulose acetate phthalate	USP
Citric acid (anhydrous)	EP
Citric acid (monohydrate)	EP
Crospovidone	EP
Ethylcellulose	EP
Hydroxyethylcellulose	EP
Hydroxypropylcellulose	USP
Hydroxypropylcellulose (low-substituted)	USP
Hydroxypropylmethylcellulose	JP
Hydroxypropylmethylcellulose phthalate	USP
Lactose (anhydrous)	USP
Lactose (monohydrate)	USP
Magnesium stearate	USP
Methylcellulose	JP
Methyl parahydroxybenzoate	EP
Petrolatum	USP
White Petrolatum	USP
Polyethylene glycol	USP
Polysorbate 80	EP
Povidone	JP
Saccharin, calcium ^a	USP
Saccharin (free)	USP
Saccharin, sodium	USP
Silicon dioxide	JP
Silicon dioxide (colloidal)	JP
Sodium chloride	EP
Sodium starch glycolate	USP
Starch, corn (maize)	USP
Starch, potato	EP
Starch, rice	EP
Starch, wheat	EP
Stearic acid	EP
Sucrose	EP
Talc	EP
Titanium dioxide	JP
Ethyl parahydroxybenzoate	EP
Propyl parahydroxybenzoate	EP
Butyl parahydroxybenzoate	EP
Glycerol	USP

^a The JP declines to participate in harmonization, so the USP and EP will harmonize bilaterally.

Note: PDG affirmed that harmonization was not to be undertaken in view of the different drinking waters standards elaborated by various governments. See the monograph for Purified Water.

APPENDIX III: STATUS OF GENERAL TESTS AND ASSAYS

	Coordinating pharmacopeia	ICH-Q6 lists
Dissolution ^a	EP/USP	*
Disintegration	EP/USP	*
Dose uniformity ^b	JP/USP	*
Color and clarity	EP	*
Extractable volume	EP	*
Heavy metals	USP	*
Particulate matter	EP	*
Residue on ignition—sulfated ash	JP	*
Sterility	EP	*
Bacterial endotoxin	JP	*
Microbial contamination	EP	*
Preservative effectiveness ^c	EP	*
Particle size distribution estimation by analytical sieving	USP	*
Inhalations	EP	
Bulk density and tapped density	EP	
Optical microscopy—powder fineness	USP	
Powder flowability	USP	
Specific surface area	EP	
Tablet friability	USP	

^aApparatuses 1, 2, and 4 are harmonized. Not all apparatuses appear in other pharmacopeias, and decision rules are harmonized. Approach to selection of media is not harmonized and may not be a valid subject in view of population and formulation differences.

^bIncludes content uniformity and weight variation.

^cIn view of two-test nature of an otherwise highly similar procedure used in different modes in US and Europe, this is nonharmonizable on test times and decision values.

REFERENCES

1. The United States Pharmacopeial Convention, Inc. *The United States Pharmacopeia*; 3rd rev., Lippincott, Grambo and Co.: Philadelphia, 1851.
2. Proceedings of the First International Conference on Harmonization. D'Arcy, P.F., Harron, D.W.G., Eds.; Queen's University of Belfast: 1992; 135–182.
3. The United States Pharmacopeial Convention, Inc. *Abstract of the Proceedings of the National Convention of 1910 for Revising the Pharmacopeia*; Washington, D.C., 1910.
4. The U.S. Pharmacopeial Convention, Inc. *Abstract of the Proceedings of the United States Pharmacopeial Convention of 1930*; Washington, D.C., 1931.
5. Anderson, L. *Unpublished manuscript. USP history*; 1906–1929; 55–56.
6. The United States Pharmacopeial Convention, Inc. *Abstract of the Proceedings of the Decennial Meeting, The United States Pharmacopeial Convention 1940*; Washington, D.C., 1941.
7. The United States Pharmacopeial Convention, Inc. *Abstract of the Proceedings of the Decennial Meeting, The United States Pharmacopeial Convention 1960*; Washington, D.C., 1961.
8. The United States Pharmacopeial Convention, Inc., Proceedings of the USP Open Conference of Revision Issues. Rockville, MD, 1989.
9. Proceedings of the 20th Anniversary of the European Pharmacopoeia. *Pharmeuropa* **1990**, 2 (1), 266–268.
10. PMA of Tokyo and Osaka, Conference on International Harmonization of Pharmaceutical Quality-Vision of Pharmacopoeia in the 21st Century. PMA: Tokyo, 1990.
11. Society for the Japanese Pharmacopoeia. *Jap. Pharmacop. Forum* **1992**, 1 (1), 20–29.
12. *Pharmacop. Forum* **1993**, 19 (4), 5849.
13. *Pharmeuropa* **1993**, 5 (2), 65.
14. Proceedings of Interpharmacopeial Open Conference on Standards for Biotechnology-Derived Products, *Pharmeuropa*, 1993 Special Edition
15. Dabbah, R.; Grady, L.T. *Curr. Opin. Biotechnol.* **1998**, 9, 307–311.

16. Cecil, T.L.; Paul, W.L.; Grady, L.T. Update of the Degree of Harmonization Survey of Excipient Monographs. *Pharmacop. Forum* **1997**, *23* (2), 3895–3902.
17. The United States Pharmacopeial Convention, Inc., Proceedings of the Joint Pharmacopeial Open Conference on International Harmonization of Excipient Standards. First Conference: Rockville, MD, 1991.
18. *Ibid* **1994**, Second Conference.
19. Proceedings of The First International Conference on Harmonization, Brussels, 1991.
20. Proceedings of The Second International Conference on Harmonization, Orlando, FL, 1993.
21. Proceedings of The Third International Conference on Harmonization, Yokohama, 1995.
22. Proceedings of The Fourth International Conference on Harmonization, Brussels, 1997.

Handling Hazardous Chemicals and Pharmaceuticals

Antonio Conto

ChemSafe Sas, Colletterto Giacosa (TO), Italy

INTRODUCTION

Over the last few years, legislation on the safety of chemical substances for humans and for the environment has been completed. Chemicals manufactured, imported and, in any way, handled by pharmaceutical and chemical companies are covered by a number of laws, which have been reinforced at all national levels.^[1,2] The United States and European requirements differ in some ways as to the application of safety parameters, but a good reciprocal acceptance on safety data exists. The safety of chemicals is evaluated through a number of physico-chemical, toxicological, and ecotoxicological studies standardized by the international evaluation agencies. Risk assessment (RA) of these chemicals is performed based on such experimental data, the use of the chemical, and possible exposure to humans and the environment.^[3] In Europe, there is a long-standing harmonized system (since 1981) that evaluates the final classification and labeling of a chemical and the so-called new or existing chemicals. In the United States, chemical safety is regulated by the Toxic Substance Control Act (TSCA) and the Code of Federal Regulation (CFR),^[4] and is managed by the Environmental Protection Agency (EPA). The Food and Drugs Administration (FDA) focuses on the registration of novel drugs in the American market. One basic step of the approach for handling a hazardous chemical is its classification and subsequent labeling. The classification is reported in the Safety Data Sheet (SDS), which is a tool that informs workers and operators about the possible hazards and risks related to the substance. The handling of hazardous chemicals must also observe the occupational exposure limits as well as medical advice rules, which provide first aid in case of accidental spillage or exposure to the chemical. To handle a hazardous chemical correctly, we need to know a lot about it [e.g., legislation that covers its handling, methods for testing the hazards and consequent risks for humans, possible efficacy (in case of active principle ingredient, or API), epidemiological data, and classification and labeling]. Pharmaceuticals are substances intended for human therapy, but because they are internally handled during the manufacturing phases (production, formulation, packaging, etc.), concerns for the safety of workers and the environment arise.

In this article, we will go through the safety concepts that can be applied to a chemical substance to give the readers a general overview of the matter. The safety of hazardous pharmaceuticals is a field that can be included in this approach, and many suggestions or proposals regarding pure chemicals will be discussed. The term “chemicals” is used interchangeably with pharmaceuticals because all pharmaceuticals are, in essence, chemicals.

DEFINITIONS

Table 1 shows some definitions that can help the readers.

LEGISLATION AND REQUIREMENTS

We will refer to two main approaches to the safety of new chemicals: that applied in the United States and that applied in Europe.

United States

The so-called premanufacturing notification (PMN) is regulated by the EPA through the TSCA, which has been in force since January 1, 1977.^[1] The notification process and the assessment system for new chemicals became operative in July 1979. All substances imported into the United States (including pharma intermediates, which are considered chemicals from a regulatory point of view) have to be listed in the TSCA Inventory of Chemical Substances, unless they are imported in low quantities for research and development (R&D) purposes. Requests for limited notifications may be accepted for the following types of products: new chemicals imported or produced in quantities of 1000 kg/year or less, new chemicals used for the production or processing of instant photographic or peel-apart film articles, and some polymers.

The PMN approach also applies for known substances destined for a significant new use (Significant New Use Rule). If the substance is on the list and it is still destined for the use shown (no SNUR), the production/importation can take place. If the substance is not listed, the PMN



Table 1

Chemical substances	Chemical elements and their compounds in the natural state, or obtained by any production process, including any additive necessary to preserve the stability of the products and any impurity deriving from the process used
Hazardous pharmaceuticals	Substances that can be considered of concern based on intrinsic toxic characteristic, or in relation to their exposure to humans and the environment
Preparations	Mixtures or solutions composed of two or more substances
Placing on the market	The making available of the substance to third parties
NONS	Notification of new substances
EINECS	European Inventory of Existing Chemical Substances
ELINCS	European List of Notified Chemical Substances
PMN	Premanufacturing notification (USA)
USA OSHA	Occupational Safety and Health Agency
USA NIOSH	National Institute for Occupational Safety and Health
ICCA	International Council of Chemical Association
HPV	High production volume
QSAR	Quantitative structure–activity relationship
LD ₅₀	Lethal dose that kills 50% of the experimental animals
NOAEL	No Observed Adverse Effect Level
NOEL	No Observed Effect Level
OEL	Occupational Exposure Level
SDS	Safety Data Sheet
PPE	Personal protection equipment
API	Active principle ingredient
RA	Risk assessment

process is followed. The PMN must be presented at least 90 days before the production of the new substance. New substances presented with a PNM are added to the list only after a 90-day review period and after the producer has submitted the so-called Notice of Commencement (NOC) within 30 days of the start of production. The notifier is asked for the following information: Chemical Abstracts Service (CAS) number, chemical name, structural and molecular formulas, trade name, spectral analysis, type and percentage of impurities, and use. Initially, no experimental data are requested; but if valid data are available, they must be included. Existing chemicals (those already listed) can be classified and consequently handled based on criteria given by the Occupational Safety and Health Agency (OSHA) and the National Institute for Occupational Safety and Health (NIOSH).

Europe

The European procedure involves the conduction of a testing program (toxicology, ecotoxicology and physico-chemical properties) based on the tonnage of the substance (annual or total production) to be notified in the European Union (EU) market. A final dossier containing all the data on safety, production, storage, handling, and disposal must be prepared.^[2]

The dossier must also include a proposal for the classification and labeling of the test substance, along with an RA for humans and the environment. The levels of notification requested in Annexes VIIC, VIIB, and VIIA of EEC Directive 92/32 are to be considered standard programs (minor changes may be introduced depending on the physicochemical features of the substance); for subsequent levels (Annex VIII), all the European Authorities must discuss and agree on the testing program to be carried out. The ecotoxicological profile of the substance from the previous level is obviously taken into account.

The EU Directive 67/548 and its latest amendment (EEC 92/32) consider “new” those chemicals not listed in European Inventory of Existing Chemical Substances (EINECS) and not belonging to the following categories for which specific community laws are in place: pharmaceuticals such as API, pesticides, cosmetics, biocides, and food additives. Exemptions may be obtained for products under R&D or for process-oriented R&D (PORD) products.

Particular attention is paid to the use of animals in testing; a product already notified cannot undergo a new testing program, but the two applicants are obliged to share toxicology data.

Existing chemicals are covered by specific regulations that set up a priority list system.^[5] Preparations are

classified and labeled by adopting a theoretical approach based on the hazard of each single component as detailed in the European Directive 45/1999.^[6]

TESTING ON CHEMICALS

The actual evaluation of the possible hazards of chemicals and the risk to humans handling such chemicals is based on data obtained from animal studies.^[7,8] This approach is constantly under discussion in terms of the ethical use of animals and some difficulties in adapting animal data to humans.^[9] Thanks to years of research, a huge amount of data on chemicals already exists, and the availability of data banks means that it is easy to access. Nevertheless, many chemicals are still unclassified for safety, and much research still needs to be done. Over the last 3 or 4 years, some industry associations have launched programs focused on testing chemicals to cover the lack of safety information, namely ICCA and HPV initiatives.^[10] Furthermore, some theoretical new tools such as the family approach and the quantitative structure–activity relationship (QSAR) are now available. These approaches are now under validation processes, which hopefully will lead to their use for regulatory purposes.

The current testing package, which checks the hazards of a chemical, consists of the following studies:

- Oral, dermal, and inhalation acute toxicity testing.
- Dermal and ocular irritation studies.
- Skin sensitization studies.
- Mutagenicity studies (basically Ames and chromosomal aberration).
- Repeated dose toxicity (subacute and chronic) studies.
- Reproductive and carcinogenicity studies.
- Ecotoxicology studies for the environmental impact.
- Studies to identify endocrine-disrupting chemicals.
- Physicochemical profile.

The research program has a tiered approach starting from the acute profile up to the most important studies such as chronic toxicology, carcinogenicity, and reproductive studies.

At the end of the process, we will be able to classify the chemical substance and to understand which handling safety procedures are the most suitable.

The aim of the experimental studies is to determine a number of toxicological parameters that will serve as a basis in classifying a substance.^[7] The most important are:

- LD₅₀, the lethal dose that kills 50% of the treated animals; this is determined by acute studies.
- The potential to be irritating to skin and eyes, and the sensitization effects.

- The No Observed Adverse Effect Level (NOAEL) and the No Observed Effect Level (NOEL) determined by repeated dose studies.
- Some intrinsic physicochemical properties that can lead to risks for humans or the environment.

Normally, those chemicals classified as highly toxic, highly sensitizing, carcinogenic, or affecting the reproductive/fertility field are considered very dangerous and very risky for humans, and special containment and safety procedures are required for them. In some cases, they are banned from the market or from use in chemical industries.

CLASSIFICATION AND LABELING

The classification of a chemical substance is focused on three main areas: physicochemical properties, toxicology, and ecotoxicology.

Based on current regulations, a substance may be classified as detailed in Table 2, which reports the possible classification class and the related letter for the label on the packaging.^[11]

The next step is to label the substance based on the classification obtained. The label must include all necessary information to inform the users about the hazards of the substance, as indicated in Table 3.

SAFETY DATA SHEET

The SDS is a document for users, which contains all the information on the substance. In particular, it provides information that is useful for the handling of chemicals.^[12] It is normally composed of a number of paragraphs focused on specific information.

Table 4 shows the structure of a typical SDS.

Table 2

Possible classification	Letter
Explosive	E
Oxidizing	O
Highly flammable	F
Extremely flammable	F ⁺
Toxic	T
Very toxic	T ⁺
Corrosive	C
Harmful	Xn
Irritant	Xi
Dangerous to the environment	N



Table 3

Information to be reported	Explanation
Name of the chemical	Chemical name of the substance
Name of the manufacturer	Full address including emergency numbers
Symbol	A picture requested by laws clearly showing the type of danger (a flame, a skull, etc.)
Letter	The abovementioned letter of danger identification (E, O, F, T, etc.)
R phrases	Risk phrases describing the danger (e.g., R25, toxic if swallowed)
S phrases	Safety phrases in case of accidental exposure (e.g., S24, avoid contact with skin)
Other indications	If needed

HANDLING

The handling of hazardous chemicals or pharmaceuticals is a consequence of all the abovementioned activities. It is obviously affected by different parameters that enter into the final evaluation. The first is the intrinsic hazard of the chemical. This is evaluated, as discussed, through a number of appropriate tests. However, the knowledge of such elements is not sufficient to calculate the risk of the chemical for humans. A simple rule says that the final risk is related to the exposure of the chemical to humans. This leads us to believe that handling is something that can considerably decrease the rate of exposure and, consequently, the risk. Low toxic substances with high ex-

posure can be more dangerous than highly toxic substances with a low (or no) exposure.^[7] This means that handling/manipulation procedures are the key factor to reducing possible effects on workers or, if not applied correctly, to amplifying adverse effects on humans that can lead to very severe health problems, especially after long exposures.

When approaching the problem, the first parameter that needs to be carefully evaluated is the possible route of exposure.^[13] Inhalation route related to the presence of thin powders is the most dangerous way of affecting the human body. This can cause respiratory problems in the lungs, or sensitizing problems, or, in worst cases, cancer of the lungs. The second route of exposure is, without

Table 4

Paragraph number	Title	Information included
1	Identification of the substance and company	Chemical name, company name, reference person, and emergency numbers
2	Composition/information on ingredients	Chemical names
3	Hazard identification	Description of possible hazards based on testing
4	First aid measures (FAMs)	Detailed explanation of FAM related to different exposure routes
5	Fire-fighting measures	Based on the recorded physicochemical properties
6	Accidental release measures	Including personal and environment precautions; cleaning procedures
7	Handling and storage	How to handle and store with PPE description
8	Exposure control/personal protection	Including, if available, TWA, TLV, OEL limits
9	Physical and chemical properties	Based on testing
10	Stability and reactivity	If available
11	Toxicological information	Based on testing
12	Ecological information	Based on testing
13	Disposal consideration	Possibility of recycling/neutralizing waste residuals, etc.
14	Transport information	Classify for transport (UN numbers, etc.)
15	Regulatory information	Classification and labeling are reported here
16	Other information	If necessary
	Disclaimer	If deemed necessary by the company

Table 5

Exposure route	PPE	First aid measures
Inhalation	Dust mask, self-contained breathing apparatus in the worst cases	Remove the persons from the exposed area to fresh air immediately
Dermal	Protective clothing, gloves	Remove contaminated clothes and shoes immediately; wash the area with soap or mild detergent and rinse thoroughly until no evidence of substance remains (15–20 min)
Eyes	Protective glasses	Wash immediately with plenty of water or normal saline; keep eyelid open
Ingestion	Face mask	Wash mouth with water if the person is conscious; treat symptomatically and supportively

doubt, the dermal route, through which local effects (irritation, dermatitis, and redness) or systemic effects (toxicity or sensitization) can be determined. The third route is the oral route, which is less common but causes high toxicity in case of accidental ingestion of the chemical. This route is used less among skilled company workers but maybe used in the public at large due to accidental reasons.

Compounds classified as harmful or toxic in these routes of exposure must be handled with specific procedures and personal protection equipment (PPE) in almost all phases of their manipulation during chemical processes. Compounds classified as highly toxic, highly sensitizing, carcinogenic, or affecting the reproductive system should be used in sealed areas or closed systems with no possible exposure to humans by any route.

Table 5 shows the containment/handling and first aid procedures in relation with the different possible routes of exposure of humans (assuming the substance shows adverse effects).

Medical advice from the company medical doctor is required if any symptom appears within a few minutes after exposure, or in cases of considerable acute exposure.

The medical doctor is responsible for adopting all the suitable therapies based on the symptoms observed and in relation to the information written on the label of the chemical.

Based on physicochemical testing, a different approach to the manipulation of the chemical is needed; this is more related to storage conditions than to the handling itself. Particular care must be taken with explosives, flammables, and oxidizing substances in terms of labeling, packaging, storing, and handling. Table 6 shows the containment/handling procedures based on the intrinsic characteristics of the substance.

Special handling provisions are foreseen for those chemicals classified as dangerous to the environment. Do not allow the substance to enter the drainage system, surface water, ground water, and soil. Therefore, storage must be severely controlled. The release of the substance in the environment is severely banned and reference to local/national rules/laws must be followed.

All residuals from empty containers and cleaning of reactors must be stored, well labeled, and disposed of by specialized waste disposal companies. Incineration is normally suggested as the best method of disposal. The amount of waste must be controlled and kept to a minimum.

All substances classified as highly toxic, highly sensitizing, carcinogenic, or affecting the reproductive system must be used in a sealed area or in a closed system in the chemical plant to avoid any possible exposure to humans. If the chemical plant has some weak points where a possible leakage or exposure could occur, all the PPE must be available.

Table 6

Characteristics	PPE	First aid measures
Explosive	Protective clothing	Avoid shock, friction, and heating; store the sample clearly labeled in a well-ventilated room
Flammable	Protective clothing	Avoid any contact with flame sources; store the sample tightly closed and clearly labeled in a controlled area; avoid smoking
Oxidizing	Protective clothing	Avoid contact with flammable materials; store the sample tightly closed and clearly labeled in a controlled area; avoid smoking

$$\frac{\text{NOEL or NOAEL} \times 70 \text{ (50) kg (average worker weight)}}{\text{serum half-life} \times \% \text{ absorption} \times \text{respiration over an 8-hr workday} \times \text{safety factor}}$$

OCCUPATIONAL EXPOSURE LEVELS

The so-called Occupational Exposure Level (OEL) is a key parameter for people responsible for the safety issues in pharmaceutical and chemical industries.^[13,14] The assessment and the application of the OEL are particularly important also for hazardous pharmaceuticals that are routinely handled within the pharmaceutical companies. There are different approaches to determine the OEL of a certain chemical and its relation to the handling/use of the substance. We are oriented to applying a tiered approach, which assesses all possible parameters that compose the hazard evaluation of a chemical substance.

Such a tiered approach is carried out in different phases:

Phase I: Identify NOEL–NOAEL in animal studies and also the possible effects in healthy human beings; kinetics/absorption data are also important.

Phase II: Find out all occupational toxicology data from available sources, treating the route of exposure as highly significant.

Phase III: Identify available epidemiological data (post-marketing surveillance and industrial hygiene).

Phase IV: Apply suitable and good safety factors.

Phase V: Calculate OEL using the following formula: (see equation above).

The final calculated value of the OEL will be considered during all procedures involving the chemical, in particular the possible human exposure consequent to its handling and use.

CONCLUSION

As we realize from the information provided in this article, the handling of hazardous chemicals and pharmaceuticals is a process involving many different areas of expertise. A company that intends to adopt a serious policy concerning the safety of their compounds (raw materials, intermediates, or bulks/API) needs to set up a multidisciplinary team with experts in pharmacology (for efficacy and epidemiology data), industrial toxicology (for classification), medicine (in relation to health), and management (in relation to legal responsibility in case of accidents). The key area is clearly industrial toxicology because of its capacity to interpret toxicological data de-

rived from various experimental activities and its use for safety. The presence of an industrial toxicologist is common in large organizations, but infrequent in small/medium companies where safety issues will become a problem in the near future.

The last step that is becoming more and more requested by competent authorities in terms of chemical safety is the preparation of a final RA on the basis of the intrinsic hazard features of the studied chemical, its use, and its possible exposure to humans. The final RA evaluation will give us the index of possible risk for humans and the containment measures to be adopted when the risk is too high.

The handling of substances classified as highly toxic, highly sensitizing, carcinogenic, or affecting the reproductive system should be kept to a minimum and, in some cases, banned if replacement is possible.

REFERENCES

1. Toxic Substance Control Act, **January 1, 1977**.
2. EEC Directive 32/92 (Seventh Amendment of EEC Directive 67/548), L154, **June 5, 1992**.
3. ECC Directive 67/93, L227, **September 8, 1993**.
4. Code of Federal Regulation, paragraph 21, USA.
5. EEC Directive 01/90, L146C, **June 15, 1990**, and EEC Directive 793/93, L84, **April 5, 1993**.
6. EEC Directive 45/1999, L200, **May 31, 1999**.
7. Klaasen, C.; Amdur, M.O.; Doull, J. *Casarett and Doull's Toxicology The Basic Science of Poisons*, 3rd Ed.; Macmillan Publishing Company: 1990; 902–916.
8. Tuffery, A.A. *Laboratory Animals, An Introduction for New Experimenters*; Department of Biological Sciences, NE Surrey, College of Technology: Ewell, Surrey, UK, 1987.
9. Willard, R.; Chappel, W.R.; Mordenti, J. Extrapolation of toxicological and pharmacological data from animals to humans. *Adv. Drug Res.* **1991**, 20, 1–116.
10. Conto, A.; Gnemi, P. New initiatives for chemical safety. *Chem. Mark. Rep.* **February 14, 2000**, 257 (7).
11. EU Directive 59/2001, L225/1, **August 6, 2001**.
12. ECC Directive 91/155, L76, **March 3, 1991**.
13. Conto, A. In *Classification of Drugs and Active Principles*, Workshop on "Cross Contamination" organized by the Italian Pharmaceutical Association (AFI), Pomezia, Rome, February 8, 2002.
14. Libermann, D.; Lockwood, C.; McConeel-Meachen, M., et al. Barrier isolation technology, a safe and effective solution for providing pharmaceutical; development facilities. *Pharm. Eng.* **July/August 2001**, 21–30.

INHALATION, DRY POWDER

Lynn Van Campen and
Geraldine Venthoye

Inhale Therapeutic Systems, Inc., San Carlos, California, U.S.A.

INTRODUCTION

Drug delivery to the lung has been historically aimed at controlling local respiratory disease where the central airways may be as suitable a target for drug deposition as the deeper lung. With little exception, currently marketed inhalation products provide therapy for asthma, chronic obstructive pulmonary disease (COPD), and bronchitis. More recently, however, there has been significant interest in using the lung, and in particular the deep lung alveolar surface, as a portal to the systemic circulation for drugs not readily administered orally. Therapeutic targets have therefore broadened substantially because pulmonary delivery is recognized for its potential to provide a noninvasive alternative to injection. New delivery systems capable of delivering drug in particles or droplets small enough to reach the peripheral or deep lung are under intense development to meet these new therapeutic targets.

Delivery of drugs to the lung depends on administration by any one of three methods: nebulizer, metered-dose inhaler (MDI), or dry powder inhaler (DPI). The nature of the drug substance and its therapeutic target may dictate which lung-delivery dosage form is more appropriate for the drug. Nebulization, for example, requires that the drug dissolve well in an aqueous medium at a concentration suitable for convenient dosing. Drugs developed as MDIs must dissolve or suspend well in a nonaqueous propellant medium at a concentration appropriate for doses metered in volumes generally less than 100 μl . For DPIs, the physical properties of the drug substance determine the ease with which processing will yield a stable powder that can be effectively aerosolized in milligram quantities by the inhaler device to deliver the proper drug dosage.

Nebulizers

Nebulizers have a long history in pulmonary delivery. Although generally effective, traditional nebulizer systems require lengthy (10–20 min) administration periods during which drug solution is delivered with relative inefficiency using an external power source. More convenient hand-held systems are currently in

development and offer the convenience of portability and metered-dose administration (1). Depending on drug solubility and dose, these systems may require multiple actuations to deliver an effective dose.

MDIs

Since the 1970s, MDIs have dominated inhalation delivery, especially in the United States. MDIs are more convenient to use than nebulizers, generally offering 100–300 metered doses per pocket-sized canister. Limitations to the reliability of their therapeutic effectiveness, however, typically arise from the need for the patient to coordinate MDI actuation with breath inhalation (2) and from the deposition of a sometimes significant amount of drug, driven by the propellant blast, to the back of the throat instead of the lung. Then, in the late 1980s, the chlorofluorocarbon (CFC) propellants used in MDIs were identified as agents contributing to the depletion of the ozone layer. This led to industry-wide reformulation efforts still underway to replace CFCs with environmentally more friendly propellants.

DPIs

The technical challenges of MDI reformulation have contributed to a growing interest in the potential of DPI technology for the development of new products that satisfy similar therapeutic, market, and environmental needs.

The DPI device presents medication to the patient as a dry powder in a form that can be inhaled orally for delivery to the target lung tissues. The delivery system should assist in the generation of very fine particulates of medication in a way that enables them to avoid the impaction barriers that normally operate in the lung to prevent the ingress of potentially harmful particles. These barriers include the oropharynx and, for deep lung delivery, the air-conducting bronchi and bronchioles.

Studies have shown that to clear the oropharyngeal impaction barrier (comprising the mouth, throat, and pharynx), particles with aerodynamic diameters smaller

than 5 μm are required (3, 4). Only particles with aerodynamic diameters less than 3 μm will reach the terminal bronchi and the alveoli in significant numbers (5). Therefore, the particle diameter required to be produced by the delivery system will depend to a great extent on the intended target lung tissue. Lung deposition is also affected substantially by the specific inhalation dynamics of the patient, which, in turn, are influenced by the delivery device. This article addresses various attributes of the dry powder inhalation product, from intrinsic material properties to final product performance.

More simple in concept than implementation, DPI technology is rapidly expanding to address a broadening therapeutic need as well as market opportunity. Characteristics of the ideal DPI system will include most or all of the following attributes:

- Simple and comfortable to use;
- Compact and economical to produce;
- Highly reproducible fine-particle dosing;
- Reproducible emitted dose;
- Physically and chemically stable powder;
- Minimal extrapulmonary loss of drug, with low oropharyngeal deposition, low device retention, and low exhaled loss;
- Multidose system;
- Powder protected from external environment and can be used in all climates and protected from moist exhaled air;
- Overdose protection; and
- Indicate number of doses delivered and/or remaining.

Outline

Fine particle powders can be produced by various methods such as micronization or spray drying. The physicochemical nature of these fine particles generally defines the stability of the bulk powder, which in turn is critical to the long-term effective performance of the dry powder product. The section *Fine Particles and the Solid State* is an introduction to better understanding the fundamental properties that underlie the behavior of bulk powders. Commentary on the various means of producing fine powders follows in the section *Powder Production: Formulation and Processing*.

Drug containment in DPIs falls into two categories: unit dose, in which the dose is premeasured during manufacture, and reservoir, in which the drug dose is metered during dose administration. Some devices store multiple unit doses for convenience. These are addressed in the section *Filling and Packaging*. The next section, *Devices: Forming the Dry Powder Aerosol*, considers the various

means of aerosolizing powder in the context of device design history and functionality. Advantages and disadvantages of different design types are considered.

Performance and Regulatory Requirements describes various ways of characterizing the dry powder aerosols and provides information on product quality performance requirements and how these attributes must be reflected in registration applications.

The Role of DPIs in Therapy briefly addresses factors affecting the therapeutic profiles of drugs delivered by DPI profiles. The article concludes with *DPIs: A Burgeoning Industry*, which surveys the DPI products on the market in 1999 and a selection of those known to be under development. Thoughts on the potential of the DPI dosage form in future therapeutic applications conclude the article.

FINE PARTICLES AND THE SOLID STATE

Crystalline and Amorphous (Glassy) States

Pharmaceutical solids can generally be described as either crystalline or amorphous (or glassy). In fact, the actual solid phase composition of a pharmaceutical formulation is usually characterized by an intermediate composition composed of both crystalline and amorphous character. In a multicomponent system such as a solid formulation comprising drug and excipient(s), certain components or even a single component may be amorphous. Because the amorphous form of a material is always a less stable, higher-energy form than its crystalline counterpart, the distinction between these forms relates to thermodynamic stability of the solid.

Crystalline materials are characterized by a three-dimensional, long-range order that translates into a distinct and unique molecular pattern that can be characterized by X-ray diffraction (XRD) (6). The molecular arrangement of the glassy state resembles that of the liquid state and lacks three-dimensional order. Thus, the classically glassy state has been designated as amorphous, that is, without structure. Pharmaceutical operations commonly used in the manufacture of DPI formulations, such as milling, spray drying, and lyophilization (freeze drying), produce materials possessing amorphous character (7).

Milling of crystalline materials introduces or increases amorphous character as the result of the significant mechanical activation that takes place during the process, including friction, deformation, attrition, and agglomeration (8). The extent of disorder, or amorphous character, introduced by the milling process depends on the behavior

of the material and its inherent resistance to the milling-imposed stresses, the amount of energy imposed by the process, and the time scale of energy release. Solid particles formed from the liquid phase, as in spray drying or freeze drying product from solution, are predominantly amorphous materials.

Crystalline materials exhibit a characteristic melting point at which they convert into a liquid form, whereas amorphous materials show no such defined transition. Rather, amorphous materials change on heating from a brittle glassy state to a rubbery state over a narrow temperature range known as the glass transition temperature, or T_g . Orders of magnitude of change in properties such as viscosity and molecular mobility take place near the T_g for amorphous materials; the temperature dependence of these properties near the T_g is typically non-Arrhenius (9). Glassy to rubbery transition is also associated with a stepwise change in the heat capacity. Thus, the T_g of an amorphous material may be determined using a differential scanning calorimeter where a stepwise change in heat flow (corresponding to a change in heat capacity) is observed during sample heating (10). For many hydrophilic drugs and excipients, water acts as a plasticizer, increases molecular mobility, and reduces the glass transition temperature. It is not uncommon for as little as 2–3 wt% water to depress the T_g by 30–40°C. The physical and chemical stability of a glassy material decreases as it approaches the T_g because of that mobility. Thus, the presence of water promotes instability. The rate at which a compound crystallizes from an amorphous state will likewise increase as the temperature of storage approaches T_g and as moisture content is increased (11, 12).

Effect of Physical State on Stability of Dry Powder Formulations

The impact of even subtle changes in physical properties of a DPI formulation can lead to substantial changes in aerosol behavior. Moisture uptake by the hydrophilic components of the formulation can lead to surface dissolution and liquid bridging between particles. This in turn leads to crystal growth, particle fusion, and an increase in particle size, which can result in strongly diminished aerosol performance (13). Powder densification under vibration during unit dose or reservoir filling, as well as product shipping, can also affect the observed aerosol behavior of dry powders.

Because of their greater molecular mobility in the solid state, amorphous systems generally exhibit greater physical and chemical instability at any given temperature

compared with their crystalline counterparts. Thus, DPI formulations are desirably prepared in a crystalline state. The low molecular weight of drugs in DPI products currently marketed supports their crystalline nature. Since the late 1980s, there has been increased interest in delivering drugs of biological origin, such as proteins for systemic uptake. These molecules typically do not crystallize and tend to remain amorphous. In freeze-dried and spray-dried biologicals for pulmonary delivery, excipients that act as protectants such as sugars must also remain amorphous to interact with the protein and/or provide a rigid matrix around the protein molecules to restrict and stabilize their motion. As with any amorphous product, physical change can be minimized by storage at temperatures well below the T_g and protection from moisture during handling and storage.

The chemical stability of an amorphous formulation also is usually a function of its storage temperature relative to T_g . The enhanced molecular mobility achieved near the glass transition translates into an increase in translational diffusion-dependent degradation pathways such as aggregation in proteins. It should be recognized that the reaction kinetics near the T_g do not obey Arrhenius kinetics and that extrapolation of the accelerated stability data generated near the T_g to stability at the storage temperature should be viewed with extreme caution. Amorphous materials must be stored well below the glass transition (at least 10°C and typically 40–50°C below T_g) to maintain their physical and chemical stability.

When dealing with partially crystalline materials such as those produced by milling, the impact of water uptake is exaggerated. The amorphous component likely absorbs greater quantities of water than its crystalline counterpart, leading to reduced T_g , increased molecular mobility and both physical and chemical instability.

Bulk Powder Properties

The respirable powders of a DPI cannot be characterized adequately by single particle studies alone; bulk properties must also be assessed because they contribute to ease of manufacture and affect maximal system performance. Primary bulk properties include particle size, particle size distribution, bulk density, and surface area. These properties, along with particle electrostatics, shape, surface morphology, etc., affect secondary bulk powder characteristics such as powder flow, handling, consolidation, and dispersibility.

The characterization and control of primary particle size and the particle size distribution of drug-containing particles are perhaps the most important factors in the

design and manufacture of dry powders for inhalation. The size, density, and shape of a particle determine its aerodynamic behavior and therefore its likelihood of depositing in the desired region of the lung. The mean (average) or median (50th percentile) particle size may be based on the number of particles, the mass (or volume) of particles, or even the surface area of particles. The particle size distribution describes the range or frequency of particle sizes occurring in a sample or the width of the particle size distribution in a sample. The variation in these size parameters owing to sampling errors must be considered in characterizing blended powders, powders stored for prolonged periods, or powders that have been mechanically agitated.

Mass median diameter (MMD) is the most common descriptor of primary particle size and may be determined by sieving or centrifugal sedimentation. Volume median diameter, as determined by laser diffraction, may be used as an approximation of MMD provided the particle density is known and does not vary with size and the particle shape is near spherical. The MMD of a powder can be used as a predictor of aerodynamic diameter by:

$$\text{MMAD} = \text{MMD} \cdot \rho_{\text{true}}^{1/2}$$

where MMAD is the mass median aerodynamic diameter, and ρ_{true} is the true density of the particle, usually determined by helium pycnometry. Cohesion/adhesion between particles normally results in the MMAD being larger than predicted. Values of MMAD less than 5 μm are considered necessary to facilitate airborne particle transit past the larynx and deposition within the lung. Powders intended for delivery to the deep lung, such as treatments for asthma or for systemic delivery, require aerodynamic behavior reflected by MMAD values between 1 and 3 μm (14). Particles of MMAD less than approximately 0.5 μm are likely to be exhaled.

Surface area is a bulk powder characteristic directly dependent on particle size distribution, porosity, and morphology. It is commonly determined by nitrogen adsorption, whereby the adsorption isotherm data are fit to a suitable mathematical model from which surface area is derived. If the particle size distribution is sufficiently narrow and the particles are not hollow or porous, the surface area can be used as a measure of change in average particle rugosity or shape. Surface area may be a more sensitive means of monitoring process control during fine particle manufacturing (jet milling or spray drying) than are particle sizing techniques.

Bulk powder density, porosity, and consolidation rate are used as characteristics of powder structure and ease of flow. These properties are typically more difficult to

determine for fine respirable powders than for coarse particles owing to the formation of bridging structures caused by high interparticulate interaction. These forces must be overcome by introducing energy, such as ultrasonic vibration or mechanical agitation, to fluidize micron-sized powders in a controllable manner. Carrier-based powder formulations are designed in part to overcome the inherent cohesion of micron-sized particles. In these formulations, the microfine drug adheres to larger carrier particles, improving powder flow and metering capability. For effective delivery, the drug particles must, of course, separate from their carrier on aerosolization and/or inhalation.

Pelletization is often used to improve the flow properties of micron-sized powders. Pelletization converts an ensemble of single particles into larger agglomerates through the formation of weak solid bridges between particles. This process also results in increased bulk powder density (ρ_{bulk}). The solid bridges formed during pelletization may aid powder flow and metering but must be overcome during aerosolization (15).

Bulk powder density must be distinguished clearly from the true density of particles. Bulk powder density is simply the mass of a powder bed divided by its volume. The volume of the powder bed includes the spaces between agglomerates, between primary particles, and the volume of micropores within the particles. These voids within the powder bed volume are collectively the powder porosity. Powder porosity (F) is calculated as:

$$F = (1 - \rho_{\text{true}}/\rho_{\text{bulk}})$$

The average number of contact points between particles increases as bulk density increases, and the interparticulate forces at these contact points must be overcome to produce a dispersed aerosol cloud. Therefore, a powder of low bulk density may be more easily dispersed as an aerosol than an otherwise identical powder of high bulk density.

However, a powder with low bulk density may be more prone to consolidation than a powder with high bulk density. Powder consolidation can be envisioned as a process of densification or packing of the particles. Consolidation occurs most rapidly during the powder agitation that accompanies powder filling or product shipping; even the imperceptible vibrations a powder experiences during seemingly static storage cause consolidation with time. The rate and extent of consolidation are dependent on particle size distribution and particle shape and can be used to describe the dynamic behavior of powder structures. The extent of powder consolidation, or compressibility, can be

evaluated by performing tap density measurements, and using the following relationship:

$$100 \times (\rho_{\text{tap}} - \rho_{\text{aerated}}) / \rho_{\text{tap}} = \% \text{ compressibility}$$

where ρ_{tap} is the tapped bulk density, and ρ_{aerated} is the aerated bulk density.

An understanding of the behavior of a powder during the manufacturing process, e.g., during blending, may aid in the identification of an optimal filling and packaging process. Blended powders may also undergo segregation of active particles from carrier particles concurrently with consolidation. Although segregation can lead to poor drug content uniformity, it is in fact desirable in the case of aerosolizing a powder composed of drug blended with a larger particle size carrier.

POWDER PRODUCTION: FORMULATION AND PROCESSING

The primary factor influencing the manufacture of DPI powders is the need to produce material that can penetrate into the lung. The manufacturing of fine particles is challenging, especially with regard to reproducibility. This challenge has resulted in the development of various approaches to the controlled production of fine particles, primarily depending on the nature of the drug. Of the processes described later, micronization and blending and, more recently, spray drying are used most often.

Once manufactured, small particles present another challenge. At small particle diameters, gravity ceases to be the major force exerted on the particles, and instead, interparticle forces become more prominent. The resultant increase in the cohesive and adhesive nature of the particles produces problems such as poor flowability, fillability, and dispersibility. These problems are typically minimized by blending with larger, less cohesive excipient particles such as lactose or pelletization of the individual drug particles. The cohesive nature of particles can be reduced further by modifying the particle surface, the goal of several emerging technologies.

Secondary processing techniques are often employed in powder production to ensure that the stability of the manufactured drug product is ensured. These major technologies are addressed in more detail later.

Formulation

Formulation of dry powders for inhalation must rely on a very short list of excipients to fulfill the customary roles of diluent, stabilizer, solubilizer, processing aid, and property

modifier (e.g., flow sustain release agent). Only a few materials are approved in the United States for use in inhalation products, and of those (e.g., propellants, surfactants), many are of little help in dry powder formulation.

Where dose requirements and drug properties allow, drug may be processed in the absence of any excipient, e.g., Astra's Pulmicort Turbuhaler. Most DPIs marketed or under development, however, rely on the addition of lactose as filler and flow enhancer (see *Blending*). Given the proprietary nature of product development, it is not known what additional excipient materials will immerge in the future as safe for inhalation. It is likely, however, that the expansion of inhalation technology to systemic delivery will call for the addition of sugars, buffer salts, and other excipients common to parenteral dosage forms to the list of acceptable inhalation excipients.

Controlled Crystallization or Precipitation

Crystallization, or precipitation, is the process by which particles are produced from solution of the material in a suitable solvent. The level of control over this process determines the physical nature and size of the finished particles. Most pharmaceutical bulk material is produced through crystallization as the final stage of the manufacturing process. The formation of a stable, crystalline material is normally the target of this final step.

In the production of materials for use in DPI products, however, the particle size of the crystallized product is normally too large. Subsequent reduction in particle size is then necessary and can significantly alter the physical nature of the material (16).

Micronization

Micronization is a high-energy particle-size reduction technique that can convert coarse-diameter particles into particles of less than 5 μm in diameter. Different types of equipment can micronize particles, for example, jet or fluid energy mills and ball mills. Although the different equipment have different operating parameters, the fundamental method of reducing the particle size is the same. All techniques involve applying a force on the particle, typically in the form of a collision, either particle–particle or particle–equipment. The force acts at imperfections in the crystal surface, initiating crack propagation through the particle. As the size of the particle decreases, the number of imperfections decreases, thereby making the task of reducing particle size more difficult.

Micronization has been used for the past 50 years to produce small particles for inhalation therapy. However, only in recent years have batch-to-batch reproducibility and stability problems been associated with the technique. Batch-to-batch variations can be caused by morphological differences in starting material; thus, it is critical that a reproducible raw material supply be available. Stability issues typically derive from changes to the varying quantities of amorphous material that are produced by the micronizing process on the surface of the resulting particles (17). This can be minimized through careful control of the micronization process, including processing conditions, batch size, and feed rate, or by the addition of a secondary processing procedure. In addition, micronization can cause decomposition of some materials (18). The issues associated with micronization are forcing many companies to investigate alternative methods of producing small particles.

Blending

The most commonly used method for improving the flowability, fillability, and dispersibility of small cohesive particles is blending the drug with excipient particles, most commonly lactose, of considerably larger particle size. Typically, these large excipient particles are greater than 60 μm , and the small drug particles are less than 5 μm . The objective of the mixing process is to produce an ordered powder in which the small particles attach themselves to the surface of larger “carrier” particles. The challenge is to ensure that the force of adhesion between the drug and carrier is strong enough to withstand segregation during blending and product storage and weak enough to allow separation of the drug particles from the carrier surface on aerosolization (19, 20). During formulation feasibility, the blends are made by mortar and pestle and/or geometric mixing in a tumbling blender. For high-volume production, the process generally involves a high-shear mixer.

The final product performance of a powder blend in a DPI is ultimately dependent on the individual drug and carrier properties as well on the process by which they are blended (21, 22). Small changes in carrier morphology can result in significant variations in the dose received by a patient (16). Again, control of the raw material supply is critical to successful product development. Moreover, secondary processing may be required to ensure that carrier particles behave consistently from batch to batch. Steps that involve transport or storage of the finished blend should be monitored closely to avoid segregation, which occurs when the drug separates from the carrier or when

carriers of different sizes separate. Segregation can be minimized by the careful selection of formulation and process equipment. For example, hopper design can play a significant role in minimizing segregation.

Pelletization

Pelletization, which often does not require the use of excipients, may offer an alternative to blending for high-dose therapeutics. The process involves deliberate agglomeration of the fine drug material into less cohesive, larger units (23). Pelletization is usually achieved by vibratory sieving or any process that tumbles powder. All processes require particular attention to time and energy parameters to ensure a consistent product. The resultant pellets must be used in a system capable of deaggregating to an appropriate particle size for aerosol drug delivery (15).

Secondary Processing

As discussed, materials used in dry powder inhalation are predominantly crystalline in nature, with varying degrees of amorphicity. This typically results from the high-energy milling process, which introduces regions of amorphous material within a crystalline material. Occasionally, however, the converse is true. Minimizing any change over time and ensuring that the material is physically stable before final packaging are major formulation challenges. These stability issues tend to be physical in nature, but occasionally chemical changes such as impurity formation also occur.

The technique generally used to minimize the degree of change in crystallinity of the milled product is to eliminate the water or other solvents from the product, usually by packaging the material within a suitable barrier (for example, aluminum foil laminate). Other techniques include the production of a 100% crystalline material, which may eliminate the effects of moisture. This technique, however, may require a secondary production stage of annealing or a quarantine period to allow the product to equilibrate under controlled storage conditions.

The final measure of crystallization effects is assessed by appropriate rigorous stability data [for example, 6 months accelerated stability at 40°C/75% relative humidity (RH)]. This may seem excessive, but measuring the degree of crystallinity is inherently difficult because of low analytical sensitivity (for example, amorphous components below 5 wt%), and pure single-phase standards are difficult to prepare and, subsequently, to measure.

Spray Drying

Spray drying, a process typically used in the production of coarser (up to 500 μm) food, pharmaceutical, and industrial powders, can also be used to prepare microparticulate powders for DPIs (13, 24–26). A typical first step involves creating a solution of the excipients and drug. Dissolving the excipients and drug ensures a uniform distribution of all the excipients and the active drug in the finished powder in contrast to the heterogeneous nature of blended powders. The solution is then atomized and mixed with a drying medium, usually air, or an inert gas if the feed consists of an organic solvent. The solvent is evaporated and removed from the drug solids.

Each spray-dried droplet forms a single particle whose size is determined by the droplet size, the dissolved solids of the feed solution, and the density of the resulting solid particle. For a given formulation and process, both the solid content and density of the powder remain constant within a batch and from batch to batch. Therefore, the distribution of the primary particle size is determined by the droplet size distribution. A narrowly distributed particle size can be achieved with a well-designed atomizer and well-controlled atomizer process parameters.

The droplet has a relatively short residence time (on the order of seconds) in the spray dryer, which minimizes the degradation of heat-sensitive components. In addition, the drug is exposed to a temperature much lower than that at the drying inlet owing to the cooling effect of the solvent evaporation. Control of droplet residence time and the lower temperature defines the amorphous versus crystalline nature of the material.

A spray dryer consists of a feed tank, a rotary or nozzle atomizer, an air heater, a drying chamber, and a cyclone to separate the powder from the air. A rotary atomizer uses centrifugal energy to form the droplet. Pressure nozzle atomizers feed solution to a nozzle under pressure, which forms the droplet. Two fluid nozzles feed solution separately into a nozzle head, which produces high-speed atomizing air that breaks the solution into tiny droplets. Both the feed solution and the drying air are fed into the drying chamber in a standard cocurrent flow (27).

Lyophilization

Lyophilization, although a relatively expensive process, can be a good process for relatively unstable compounds. In lyophilization, the solvent (usually water) is frozen and then removed by sublimation in a vacuum environment. The low temperature maintained

during the entire process minimizes thermal degradation of the drug compound.

Typically, the drying process can be divided into primary and secondary phases. During the primary phase, the drug solution is filled into vials and then placed within a temperature-controlled drying chamber. There, the solution is frozen according to physiochemical principles as the shelf temperature is lowered to below freezing. The shelf temperature is subsequently increased but maintained below the freezing point. A vacuum is applied to the chamber to sublimate the solvent. This phase of the drying process extracts the majority of the solvent (50–80%).

During the secondary drying phase, the remainder of the solvent is removed at an elevated but still subfreezing temperature. During freezing, supercooling is necessary to encourage crystallization of the drug compound. The extent to which the compound is supercooled depends on the nature of the compound, the temperature program of the shelf, the heat transfer properties of the container, and the presence of particulates in the solution. The degree of supercooling determines the size of the solvent crystal and, subsequently, the size of the channel formed during primary drying. Consequently, the degree of supercooling affects the rate of sublimation, the rate of secondary drying, and, eventually, the surface area of the finished powder. A goal of secondary drying is to minimize product moisture content.

Therefore, it is most important to select a cooling temperature profile to achieve the desired objective(s). The objective could be simply to achieve a uniform degree of supercooling and freezing or to add an annealing process to allow the solute to crystallize or the ice crystal to grow. The possibility of allowing a long annealing process gives great flexibility to achieve the desired solid-state property for the powder.

The lyophilized cake must then be milled. Compared with particles generated by spray drying, the particle size of milled lyophilized powders generally has a broader distribution than does spray-dried powder, which is formed one particle at a time in a continuous process. Despite the longer processing time necessary to create a dry powder through lyophilization (and the consequent economic implications), this process can provide the formulator with better control of the powder in its solid state.

Supercritical Fluid Technology

Extraction by supercritical fluids, carbon dioxide and propane in particular, is currently being investigated as a means of controlling the size and shape of particles

for inhalation. Supercritical fluids are liquids above their critical pressure and temperature (28). Under these conditions, the molecules exhibit the flow, polarity, and solvency properties common of liquids but have the diffusivities and reactivities characteristic of gases.

Precipitation of the particles occurs by two methods involving atomization of a feed: 1) rapid expansion of supercritical solutions containing dissolved drug, and 2) gas antisolvent recrystallization, the supercritical fluid acting as an antisolvent for dissolved drug contained in droplets of another miscible or partially miscible liquid (for example, ethanol, methanol, and acetone).

The second technique, sometimes described as SEDS (solution-enhanced dispersion by supercritical fluids), has been scaled up successfully for an inhalation application to pilot plant manufacture. As with spray drying, this technique is a single-step process. The drug material must show solvency in the cosolvent but complete insolubility in the supercritical carbon dioxide. The resultant solvent-free particles are less cohesive than micronized material as high crystallinity is achieved, leading to decreased charging effects (29). Particle size distributions for these powders are reported to be narrow, with small median aerodynamic diameters ($<2.5\ \mu\text{m}$). In addition, regular particle morphologies are obtained for these thermodynamically stable powders, making them amenable to further processing steps and handling (30).

FILLING AND PACKAGING

The primary consideration when developing systems for packaging dry powders for dose delivery is the goal of delivering the correct drug dosage to the patient. When dealing with drug powders intended for use in DPIs, some basic issues must be considered—specifically, that the drug usually must be delivered in a small volume, which is often difficult to handle owing to small particle size.

The greatest challenge faced in developing packaging systems for dry powders relate to maintaining dispersibility in packaging, which can be affected by compression and electrical charge. Compression of the drug powder, which can be a consequence of excessive handling, can result in an unintended increase in drug concentration. The small drug particles are also vulnerable to alteration in electrical charge, which can result from the motion of particles, against both themselves and the packaging equipment, and from the unintended absorption of water by the drug powder.

Package dose metering can be accomplished by weight or by volume. Dry powders developed for DPIs are formulated to deliver a specific dose of drug per a given unit of drug powder. Drug powders can be packaged either in unit dose or in reservoir systems, each of which has certain advantages (Table 1).

Unit-Dose Systems

Unit-dose systems package drug powders into individual-use packages that contain a known quantity of drug. Patients may use single or multiple units of the drug to obtain a given dose. The greatest advantage of unit-dose systems is that a greater degree of control at the manufacturing level can be maintained. Individual drug doses can be metered by weight or volume. Although metering by weight results in a high degree of accuracy, it is a slow process. Thus, more commonly, unit-dose packaging is metered by volume.

Metering by volume, although offering a reliable means for high-volume production, has disadvantages related to the dispersibility of the drug powder, which in turn can affect the accuracy and precision of the drug dose. Dispersibility can be managed by devising filling processes that optimize powder flow, which will vary by drug compound; by minimizing handling and thereby compression of the drug powder; and by minimizing the relative motion of drug particles against other drug particles and against the filling equipment. Minimizing drug powder motion can reduce electrostatic charging of particles and consequent equipment malfunctions and problems with packaging, which include dose-extraction difficulties and particle dispersion at the time of drug administration.

Unit-dose systems typically rely on blister packaging or capsules to contain the drug until it is dispersed by the delivery device. Blister packages have several advantages over capsules. Those constructed of aluminum are generally impervious to moisture. Inner linings of either polyvinyl chloride or polypropylene create means for sealing the package. Gelatin capsules generally contain approximately 12% water under ambient conditions and thus are a potential source of moisture to powder not equilibrated to ambient relative humidities.

Reservoir Systems

Reservoir systems offer the advantage of variable dosing, generate less waste, are less expensive to manufacture, and are simpler to use than unit-dose systems. Relying on a metering system contained within the delivery device, they

Table 1 Primary packaging for DPI drug formulation

Dosing system	Advantages	Disadvantages
Unit-dose	Simpler, cheaper device, less prone to malfunction Protects powder up to the time when it will be delivered to the patient as an aerosol	Patient must handle and load individual unit-dose packages into the device before dosing Dose titration is limited to dose-quantity available from drug supplier (similar to pills)
Multidose	More convenient for the patient	The device becomes more complex because means to load multiple doses are required Also, means for displaying number of doses left are required Device may be more prone to malfunction owing to jamming or improper indexing
Reservoir	Multidose and dose titration easy to implement Convenient—no separate unit-dose blisters to worry about	Powder not generally well protected after reservoir is opened; physical and/or chemical characteristics may deteriorate with time Biological contamination may be an issue Metering of the dose is carried out by the device, which increases the device complexity; metering often is not adequately controlled because the physical characteristics of the powder are often unknown at the time of dosing

may be less precise in their drug delivery. Because the drug reservoir must be accessed repeatedly, these systems encounter an increased difficulty in maintaining moisture level. Maintaining a highly flowable drug powder in this system may also lead to greater drug formulation challenges.

DEVICES: FORMING THE DRY POWDER AEROSOL

Design Objectives and Constraints

The ultimate goal of all pulmonary delivery devices is to reproducibly deposit the required quantity of drug in the target lung tissues. Many factors influence the selection of a particular DPI design, including the characteristics of the drug to be delivered, its powder formulation, and its associated therapeutic regime. Other factors that must be considered include drug cost, desired dose, market factors, and expected degree of patient compliance.

Drug cost plays an important role in determining the economic feasibility of a device by determining how much of a drug may be lost in administration and routine device use. Drug dose and side effects may determine the reproducibility bounds of particle size and mass of drug delivered that are necessary for effective

therapy. The drug target tissue may determine the importance of achieving a very small particle size; local delivery of drugs to the upper airway may allow for a larger particle size than delivery of particles to the deep lung for systemic absorption. The degree of cohesive-ness of the powder particles will determine how much energy the device must transfer to produce a particle of a given size. The therapeutic regime and a range of market factors, such as degree of convenience and cost of other available drugs and/or therapies, may determine how portable or inexpensive the device must be. The anticipated degree of patient compliance may favor some technological solutions over others.

DPIs cannot be considered devices alone but must be considered as components of a larger delivery system, which also includes the formulation of the drug powder, its manufacturing processes, and packaging.

Functional Description

Several DPI designs have been proposed, developed, and successfully marketed in the past 3 decades. Although these devices vary widely in characteristics and operation, they all perform certain basic functions. A DPI must extract the dose from the bulk powder drug package, generate a fine cloud of drug particles by deagglomerating the powder and diluting it with air, and deliver the drug cloud to the patients' airways.

Bulk Powder Drug Package

The first task for a device in delivering a dose is extracting the dose from the drug package. As addressed earlier, two primary alternatives are available: 1) a number of doses may be stored in a powder reservoir, and 2) each dose may be individually packaged as a unit dose. Reservoir systems are inherently multidose. With unit-dose systems, the device may require individual loading of a unit-dose package before inhalation or loading of several unit-dose packages into the device for multidosing.

Energy Sources

Energy input is required to extract the powder from its packaging, generate the fine particle cloud, and dilute it with air. Historically, this energy comes from the patient’s inhalation effort (Table 2). In some cases, the energy for extracting/metering the powder comes from the mechanical manipulation of the device by the patient (31, 32). More recent designs use concepts borrowed from the MDI industry or novel approaches involving other technologies.

Uncontrolled vs. on-demand aerosol generation

DPI devices that rely on patient inhalation are inherently on-demand, that is, only when the patient inhales is the aerosol delivered. In contrast, MDIs are inherently uncontrolled, and the patient’s breathing maneuver has to be carefully synchronized to the aerosol generation event for effective dosing (Table 3).

DPI devices that rely on sources of energy other than patient inhalation effort may face similar problems as MDIs. Devices may “trigger” the aerosol generation event, which then follows uncontrollably. Alternatively, the

delivery of energy to the powder may be modulated by the device and controlled based on monitoring of the patient inhalation maneuver. Table 3 gives some advantages and disadvantages to both approaches.

Homogeneous Powders and Blends

Dry powders must be able to flow readily to leave the capsule or powder reservoir but also must generate a fine aerosol so that the patient can inhale a proper dose. These two requirements are often difficult to achieve simultaneously. Fine powders tend to be cohesive and have poor flow properties. Blending with a carrier phase, pelletization, and other approaches have been used to overcome these limitations. The use of blends and homogeneous powders is compared in Table 4 from a DPI device perspective.

PERFORMANCE AND REGULATORY REQUIREMENTS FOR DPIS

Performance Characterization

Two critical attributes characterize the performance of DPIS: the uniformity of the delivered dose and the aerodynamic assessment of particle size distribution. To determine delivered-dose uniformity, an apparatus capable of quantitatively retaining the delivered dose leaving the device is used. For aerodynamic particle size assessment, a multistage liquid impinger or cascade impactor is used. All aerosol performance testing must be conducted under defined temperature and humidity conditions.

Table 2 Energy sources for delivery

Patient inhalation		Other (air pump, metered propellant, electrical, other)	
Advantages	Disadvantages	Advantages	Disadvantages
No need to coordinate aerosol generation with patient inhalation	Delivery, dispersion performance and thus dose is affected by the patient’s ability to inhale at a suitable high flow rate	Decouples aerosol generation from patient ability to perform a correct inhalation maneuver	Adds complexity and cost to the device by increasing the number of subsystems in it
Device is generally very simple in many cases, no moving parts are involved in powder deagglomeration		Allows the extraction and deagglomeration of more cohesive powders because additional energy can be applied in the process	

Performance Specifications

Currently, the United States, European, and British pharmacopoeias specify different requirements for delivered-dose uniformity. Table 5 details these requirements as well as proposed U.S. Food and Drug Administration (FDA) expectations (33). The *Japanese Pharmacopoeia* does not specify a delivered-dose uniformity requirement. Current compendia should be consulted as references.

Of the four pharmacopoeias, the *U.S. Pharmacopeia* (USP) has the strictest requirements for delivered-dose uniformity. Although the *British Pharmacopoeia* (BP) allows the same performance range, the USP defines the range around the label claim, and the BP defines the range around the average value. FDA expectations for delivered-dose uniformity is currently tighter than those specified in all the pharmacopoeias.

The various pharmacopoeias outline appropriate methods for aerodynamic assessment of particle size distribution. The USP defines the size distribution through mass median aerodynamic diameter (MMAD) and geometric standard deviation (GSD). None of the pharmacopoeias specify a requirement for particle size. However, the particle size specifications that are set should be appropriate for the intended use of the product. For example, if the particles are intended to reach the deep lung, the MMAD of particles exiting the device should be less than 5 μm . In general, the smaller the aerosol MMAD, the greater the deposition in the lung.

Two recent trends originating with the FDA may influence the assessment and reporting of aerodynamic

particle size distribution. The first is determining the particle size distribution from a single or unit-emitted dose. This may pose an analytical challenge in some cases because the amount of active ingredient in each stage may be present only in trace amounts. The second is setting a drug quantity specification for each stage of the impinger or cascade impactor.

RELEASE AND STABILITY TESTING PARAMETERS

Various dry powder attributes are assessed at release and on stability. These include physical characteristics such as powder appearance, content uniformity, delivered dose uniformity, and particle size distribution. Chemical attributes that may be assessed include drug content, purity, and identity as well as the water content of a powder. Dry powders may also undergo microscopic evaluation for foreign particulate matter, unusual agglomeration, and particle size. Microbial limits also should be examined, including the total aerobic, yeast, and mold counts. The presence of specific pathogens should be ruled out. The dry powders also may be dissolved to test for pH level. In addition, certain compendial requirements for content and delivered-dose uniformity should also be measured.

The USP and *European Pharmacopoeia* (EP) propose that the total aerobic count not exceed 100 CFU/g, that the total yeast count and mold count not exceed 10 CFU/g, and that no specific pathogens be detectable. Specifications for the other attributes should be based on the intended use

Table 3 Control of aerosol generation

Uncontrolled		On-demand	
Advantages	Disadvantages	Advantages	Disadvantages
Device is simpler; no feed-back systems are required to monitor the patient.	More prone to patient misuse	Aerosol is delivered when patient can inhale it most effectively	Device is more complex because feed-back systems (mechanical or electronic) are required
“Violent” aerosol generation processes are allowable, permitting the delivery of large amounts of energy to the powder in a very short period of time	Typically requires the use of a holding chamber to store the aerosol in between generation and patient inhalation, resulting in a larger-size device	Better dose control	Energy delivery to the powder has to be very well-controlled
			Device may be more prone to failure

Table 4 Influence of powder behavior on device design

Blends (lactose carrier)		Homogeneous powders	
Advantages	Disadvantages	Advantages	Disadvantages
Powder can be easily extracted from its packaging	On delivery, coughing and other unpleasant sensations may be induced because the carrier particles deposit in the mouth and throat	Little mouth and throat deposition; patient does not “feel” he/she is inhaling an aerosol	Formulation process becomes a key factor in the development of the product; the properties of the compound to be delivered dominate the performance of the resulting powder
Inclusion of the carrier phase usually facilitates dispersion	A larger amount of powder needs to be moved and dispersed; in terms of energy requirements, unit dispersion may be offset by the increase in payload		

and the historical performance of the product. As with other dosage forms, specifications must be met throughout the intended shelf life of the product.

The International Conference on Harmonization (ICH) has identified stability requirements for room temperature storage and testing intervals. It recommends that dry powders be stored at 25°C and 60% RH for real-time conditions; at 40°C and 75% RH for accelerated conditions; and 30°C and 60% RH if significant change is observed at accelerated conditions. The ICH recommends testing samples every 3 months for the first year, every 6 months for the second year, and yearly thereafter. In addition to these requirements, the FDA suggests a storage condition at 25°C and 75% RH if significant change is observed at the accelerated condition. Six-month data would be required at the time of the New Drug Application (NDA) submission, and the study must cover 1 year.

When an NDA is submitted, the FDA requires that 12 months of data be collected at real-time conditions and 6 months of data be collected under accelerated conditions. If significant change is observed at 6 months for the accelerated condition, 6-month data at the 30°C and 60% RH condition must be submitted, and the study must cover 1 year.

ROLE OF DPIS IN THERAPY

Some direct comparisons of DPI and MDI for the same drug have been made in the interest of developing alternative but comparable products for patients. The therapeutic performance of inhalation delivery systems is

as dependent on the patient as it is on the product itself. Therefore, some demonstration of clinical comparability is generally required to support product substitution.

The deposition pattern of the inhaled dry powder aerosol can be strongly influenced by the patient's inhalation dynamics and lung anatomy. At high inhalation flow rates, a given particle will have a greater tendency to impact the back of the throat or to deposit in the upper airways. For those delivery systems requiring high flow to deaggregate the powder particles, deep lung deposition is less accessible. The proliferation of device designs has been in part the result of attempts to minimize dosing variability regardless of source (34).

The target for lung deposition varies depending on the therapy under consideration. In the treatment of asthma by β -adrenergic agonists, the Central airways are generally targeted. On the other hand, therapies intended to treat alveolar disease, chronic obstructive pulmonary disease, or systemic conditions must reach the peripheral regions of the deep lung.

Among the pharmacokinetic advantages offered by delivery to the lung are fast onset of action and lack of first-pass effect. Doses to the lung can prove 10–20 times more effective than oral dosing and for local therapy can result in substantially reduced side effects.

In a study, three different fluticasone propionate products—an MDI and two DPI products, Diskhaler and Diskus—were directly compared in healthy volunteers and patients (35). The systemic drug bioavailability was highest for the MDI, whereas the bioavailability was similar for the two DPIs. The pharmacokinetic results are consistent with the in vitro evaluation in which the MDI gave the highest fine particle dose (FPD), whereas the two

Table 5 Product quality requirements

Pharmacopoeia	First-stage testing (N = 10)	Second-stage testing (additional 20)
United States	NMT (Not more than) 1 of 10 outside the range of 75.0–125.0% of label claim None outside the range of 65.0–135.0% of label claim; if 2–3 are outside of 75.0–125.0% and none are outside of 65.0–135.0% proceed to second stage	NMT 3 of 30 outside the range of 75.0–125.0% of label claim; none outside the range of 65.0–135.0% of label claim
British	NMT 1 of 10 outside the range of 75–125% of average value None outside the range of 65–135% of average value; if 2–3 are outside of 75–125% and none are outside of 65–135%, proceed to second stage	NMT 3 of 30 outside the range of 75–125% of average value; none outside the range of 65–135% of average value
European	NMT 1 of 10 outside the range of 65–135% of average value; none outside the range of 50–150% of average value If 2–3 are outside of 65–135% and none are outside of 50–150%, proceed to second stage	NMT 3 of 30 outside the range of 65–135% of average value; none outside the range of 50–150% of average value
FDA, proposed ^a	NMT 1 of 10 outside the range of 80–120% of label claim None outside the range of 75–125% of label claim; if 2–3 are outside of 80–120% and none are outside of 75–125%, proceed to second stage	NMT 3 of 30 outside the range of 80–120% of label claim; none outside the range of 75–125% of label claim
Japan	No delivered-dose uniformity specification	N/A

^a(From Ref. 30.)

Table 6 Earliest dry powder inhalation systems

Year introduced	Name	Manufacturer	Indication	Packaging/ metering	Energy source(s)	Blend
1949	Aerohaler	Abbott	Asthma, COPD	Unit dose; “sifter” cartridge	Mechanical, patient inspiration	No
1971	Spinhaler	Fisons (now Aventis)	Asthma, COPD	Unit dose; hard gelatin capsule	Mechanical, patient inspiration	Yes
1977	Rotahaler	Allen and Hanburys (now Glaxo)	Asthma, COPD	Unit dose; hard gelatin capsule	Mechanical, patient inspiration	Yes
1988	Turbuhaler	Astra	Asthma, COPD	Reservoir	Mechanical, patient inspiration	No
	Diskhaler	Allen and Hanburys (now Glaxo)	Asthma, COPD	Multidose blister	Mechanical, patient inspiration	Yes
	Inhalator	Boehringer–Ingelheim	Asthma, COPD	Unit dose; hard gelatin capsule	Mechanical, patient inspiration	Yes

Table 7 More recent dry powder inhalation systems^a

Name	Manufacturer	Packaging/metering	Energy source(s)
Pulvinal	Chiesi	Reservoir	Mechanical, patient inspiration
Easyhaler	Orion	Reservoir	Mechanical, patient inspiration
Clickhaler	ML Labs	Reservoir	Mechanical, patient inspiration
Discus	Glaxo	Multidose blister	Mechanical, patient inspiration
Monohaler	Astra	Unit dose	Mechanical, patient inspiration
AIR ^b	Alkermes	Unit dose	Mechanical, patient inspiration
Spiros ^b	Dura	Multidose blisters	Mechanical, not driven by patient inspiration
Inhance TM Pulmonary Delivery System	Inhale	Unit dose	Mechanical, not driven by patient inspiration

^aThis list is not exhaustive. Many other manufacturers, in both the United States and Europe, are also developing dry powder inhalation drug-delivery systems.

^bIn clinical trials; not yet on the market.

DPIs had similar FPD values. It was also reported that in a separate study using healthy volunteers, the pharmacokinetics of a nonchlorofluorocarbon formulation have been shown to be similar to that for the original chlorofluorocarbon formulation for a fluticasone propionate MDI product. Consistent with their similar pharmacokinetic results, the clinical performance of Diskhaler and Diskus is similar in both children and adults (36, 37).

DPIs: A BURGEONING INDUSTRY

The first commercially available DPI system appeared on the market in 1949, developed and marketed by Abbott

under the name Aerohaler. Like all pulmonary drug-delivery methods that existed before now, it delivered small molecule compounds (bronchodilators or inhaled corticosteroids) to the airway (not necessarily to the deep lung) for the treatment of asthma or chronic obstructive pulmonary disease. Table 6 lists some of the early DPI systems and their basic characteristics.

Table 7 presents some of the newer entrants in the DPI field, most still focusing on local delivery of small molecule drugs to the airway for asthma or COPD but some in clinical trials for systemic delivery of macromolecules such as insulin via the deep lung.

New DPI technologies in development by Inhale Therapeutic Systems and Alkermes are enabling the delivery of macromolecules to the deep lung. Leading this

field is Inhale's insulin product, currently in the phase 3 trials with Pfizer. In the next years, it is expected that dry powder inhalation will become a broadly accepted and effective means of delivering a wide variety of therapeutics—antibiotics, analgesics, antibodies, hormones, proteins, and perhaps gene therapeutics. The potential of this technology continues to be explored.

ACKNOWLEDGMENT

We thank D.B. Bennett, S. Duddu, S. Fong, S. Wong, J. Lord, J. Parks, C. Schuler, and S. White for their contributions to this article.

REFERENCES

1. Steed, K.P.; Freund, B.; Zierenberg, B.; Newman, S.P. Proceedings of Drug Delivery to the Lungs VI London UK, 1995; 14–15.
2. Crompton, G.K. The Adult Patients' Difficulties with Inhalers. *Lung* **1990**, 659–662.
3. Stahlhofen, W.; Gebhart, J.; Heyder, J. Experimental Determination of the Regional Deposition of Aerosol Particles in the Human Respiratory Tract. *J. Am. Int. Hyg. Assoc. J.* **1980**, 41, 385–398.
4. Byron, P.R. Pulmonary Targeting with Aerosols. *Pharm. Tech.* **1987**, 11, 42–56.
5. Wong, D.Y.T.; Wright, P.; Aulton, M.E. Influence of Drug Particle Size on the Performance of Dry Powder Inhalations of Nedocromil Sodium. Proceedings of the 14th Pharmaceutical Technology Conference, Barcelona, Spain, 1995; 3, 86–108.
6. Ford, J.L.; Timmins, P. *Pharmaceutical Thermal Analysis, Techniques And Applications*; Ellis Horwood: Chichester, UK, 1989.
7. Elamin, A. *Effect of Milling and Spray Drying on Water Interactions and Physicochemical Properties of Pharmaceutical Materials*; Uppsala University: Sweden, 1994.
8. Sebhathu, T.; Angberg, M.; Ahlneck, C. Assessment of the Degree of Disorder in Crystalline Solids by Isothermal Microcalorimetry. *Int. J. Pharm.* **1994**, 104, 135–144.
9. Oksanen, C.A.; Zografi, G. The Relationship Between the Glass Transition Temperature and Water Vapour Absorption by Polyvinyl Pyrrolidone. *Pharm. Res.* **1990**, 7, 654–657.
10. Saleki-Gerhardt, A.; Ahlneck, C.; Zografi, G. Assessment of Disorder in Crystalline Solids. *Int. J. Pharm.* **1994**, 101, 237–247.
11. Hancock, B.C.; Zografi, G. The Relationship Between the Glass Transition Temperature and the Water Content of Amorphous Pharmaceutical Solids. *Pharm. Res.* **1994**, 11, 471–477.
12. Hancock, B.C.; Shamblin, S.L.; Zografi, G. Molecular Mobility of Amorphous Pharmaceutical Solids below their Glass Transition Temperatures. *Pharm. Res.* **1995**, 12, 799–806.
13. Venthoye, M.G. *Characterization of an Amorphous Dry Powder Aerosol System*; Ph.D. Thesis, London University, 1997.
14. Curry, S.H.; Taylor, A.J.; Evans, S. Br. *J. Clin. Pharmacol.* **1995**, 2, 267–270.
15. Wetterlin, K.I. Design and Function of the Turbohaler®. A *New Concept in Inhalation Therapy*; Newman, S.P., Moren, F., Crompton, G.K., Eds.; Medicom: Amsterdam, 1987, 85–89.
16. Kassem, N.M. *Generation of Deeply Inspirable Clouds From Dry Powder Mixtures*; Ph.D. Thesis, University of London, 1990.
17. Ward, G.H.; Schultz, R.K. Process-Induced Crystallinity Changes in Albuterol Sulphate and its Effect on Powder Physical Stability. *Pharm. Res.* **1995**, 12, 773–779.
18. Rogerson, C. The Design and Production of Microparticles for Inhalation. Proceedings of Recent Advances in Dry Powder Inhalers. London, 1996.
19. Moren, F. In Vitro and In Vivo Performance of Powder Inhalers, Proceedings of Respiratory Drug Delivery II Interpharm Press: Buffalo Grove, IL, 1990.
20. Zanen, P.; Van Spiegel, P.I.; Van der Kolk, H.; Tushnizen, E.; Enthoven, R. The Effect of Inhalation Flow on the Performance of a Dry Powder Inhalation System. *Int. J. Pharm.* **1992**, 81, 199–203.
21. Kassem, N.M.; Ho, K.K.L.; Ganderton, D. The Effect of Air Flow and Carrier Size on the Characteristics of an Inspirable Cloud. *J. Pharm. Pharmacol.* **1989**, 41 (Suppl), 14p.
22. Ganderton, D.; Kassem, N. Dry Powder Inhalers. *Advan. Pharm. Sci.* **1992**, 6, 165–191.
23. Bell, J.H. Pelletised Medicament Formulations, Fisons Patent, Application No. 152047, 1975.
24. Vidgren, M.T.; Vidgren, P.A.; Paronen, T.P. Comparison of Physical and Inhalation Properties of Spraydried and Mechanically Micronised Disodium Cromoglycate. *Int. J. Pharm.* **1987**, 35, 139–144.
25. Chawla, A. *Spray-Dried Powders For Use In Dry Powder Aerosol Formulation*; Ph.D. Thesis, University of London, 1993.
26. Chawla, A.; Taylor, K.M.G.; Newton, J.M.; Johnson, M.C.R. Production of Spraydried Salbutamol Sulphate for Use in Dry Powder Aerosol Formulation. *Int. J. Pharm.* **1994**, 108, 233–240.
27. Masters, K. *Longman Scientific and Technical. Spray Drying Handbook*, 5th Ed.; John Wiley & Sons: New York, 1991.
28. Phillips, E.M.; Stella, V.J. Rapid Expansion from Supercritical Solutions: Application to Pharmaceutical Processes. *Int. J. Pharm.* **1993**, 94, 1–10.
29. York, P. Solid-State Properties of Powders in the Formulation and Processing of Solid Dosage Forms. *Int. J. Pharm.* **1983**, 14, 1–28.
30. York, P. New Approaches to the Preparation of Particles for Inhalation. Proceedings of Recent Advances in Dry Powder Inhalers. London, 1996.
31. Schultz, R.K.; Miller, N.C.; Smith, D.K.; Ross, D.L. Powder Aerosols with Auxiliary Means of Dispersion. *J. Biopharm. Sci.* **1992**, 3, 115–122.

32. Hill, M. Characteristics of An Active Multiple Dose Dry Powder Inhaler, Proceedings of Respiratory Drug Delivery. IV Interpharm Press: Buffalo Grove, IL, 1994; 109–116.
33. Guidance for Industry. *Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Drug Products, Chemistry, Manufacturing, and Controls Documentation Draft*; U.S. Food and Drug Administration: Department of Health and Human Services: Rockville, MD, October 1998.
34. Timsina, M.P.; Martin, G.P.; Marriott, C.; Ganderton, D.; Yianneskis, M. Drug Delivery to the Respiratory Tract Using Dry Powder Inhalers. *Int. J. Pharm.* **1994**, *101*, 1–13.
35. Johnson, M. Fluticasone Propionate: Pharmacokinetic and Pharmacodynamic Implications of Different Aerosol Delivery Systems, Respiratory Drug Delivery VI, **May 3–7, 1998**, 61–70
36. Cater, J.I.; Vare, M.; Peters, W.J.; Olsson, B.; Gomez, E. Comparison of the Efficacy of Fluticasone Propionate Given Twice Daily via the DiskusTM/AccuhalerTM and the DiskhalerTM in Patients with Asthma. *Eur. Resp. J.* **1995**, *8*, 427S.
37. Bousquet, J.; Tosserard, B.; Medley, H.V. Double-Blind Parallel Group Study to Compare the Long Term Clinical Efficacy and Safety of Two Different Methods of Administering Inhaled Fluticasone Propionate in Chronic Severe Asthmatic Patients. *Eur. Resp. J.* **1995**, *8*, 427S.

IMMUNOASSAY

Stephen G. Schulman

G. J. P. J. Beernink

University of Florida, Gainesville, Florida, U.S.A.

INTRODUCTION

Immunoanalytical methods that are based on the selective, reversible binding of small molecules (drugs) or macromolecules by biologically derived antibodies have revolutionized the field of biomedical analysis. Immunoassays have allowed the determination of very small amounts of analytes that were previously unassayable in biological matrices by other techniques. Since the original work on the analysis of insulin by Berson and Yalow (1), immunoassay methods have been developed for the determination of a wide variety of drugs, pesticides, hormones, and biological proteins. Immunoassays are relatively simple procedurally. This has led to the development of many "kit-type" immunoassay systems that are used routinely for home diagnostics.

Antibodies that are used as reagents in immunoassays are, in general, molecules of the immunoglobulin G (IgG) type. They are produced by white blood cells in response to foreign substances introduced in mammalian species. Millions of years of vertebrate evolution have developed immunoglobulins into exquisitely discriminating devices capable of recognizing subtle differences between molecules; for example, a mouse can generate millions of different immunoglobulin specificities. These immunoglobulins combine specifically with the substances (antigens) that elicited their formation. This then triggers processes by which the foreign antigens are cleared from the organism, which is the ultimate goal of the immune process (2).

These molecules are heterogeneous, bifunctional glycoproteins in which the variable amino acid sequence in the polypeptide component provides its biologic activity. This polypeptide component is made up of two heavy or H chains (50,000 Da) and two light or L chains (20,000 Da), held together by disulfide bonds. The two binding sites of the antibody molecule appear to reside on the NH₂ terminal ends of the polypeptide chains (3). Antibodies produced by injection of foreign antigens into a host animal are structurally heterogeneous and respond to different aspects of the same antigen with different binding strengths and specificities. If the antibody-

producing blood cell can be isolated in a pure cell culture, however, only one particular antibody structure will be produced (4). Antibodies harvested from pure cell cultures are referred to as monoclonal antibodies, are homogeneous and react with only one or a few closely related antigens. The antibody producing cells which can be grown in culture are produced by cell fusion techniques and provide superior antibodies for use in immunoassays because of their identical specificities and binding strengths. Antigens must have molecular weights in excess of 10⁴ Da in order to evoke antibody production. However, small molecules, such as drugs, can be bound to macromolecular carriers and some of the antibodies produced from these will respond to the drug (hapten).

All immunoassay procedures take advantage of the specific reactions between antibodies and antigens. They involve measurement, directly or indirectly, of the extent of binding between antibodies (reagents) and antigens (analytes). Labels are used in conjunction with the antigens or antibodies so that the concentrations of molecular species can be measured instrumentally. Labels are entities that impart some measurable signal, such as radioactivity, fluorescence, chemiluminescence, or electrochemical or enzyme activity to the antibody or antigen to which it is attached. The determination of the extent of antibody binding requires measurements of the amounts of labeled antigen or antibody in the complexed (bound) and in the free forms. This is generally expressed as the bound/free (b/f) concentration ratio, which is related to the concentration of analyte (5). The measured signal can be either directly or inversely proportional to the b/f ratio depending on the chemistry of the system used. There are two ways to determine the amount of antigen present: one using a limited amount of reagent (competitive assays), the other using an excess amount of reagent (noncompetitive assays).

The competitive assay uses a limited amount of antibody, which is insufficient to bind all of the antigen. The antigen competes with a fixed amount of labeled antigen for the limited number of antibody binding sites. From the proportion of bound (or free) labeled antigen, the concentration of unlabeled antigen can be determined.

The noncompetitive assay uses an excess of antibody. Different approaches to detect the bound antigen have been developed, the most common use an antibody, in excess, coupled to a solid phase. The bound antigen is then detected with a second antibody labeled in a way that aids detection (e.g., radioactive, fluorophore, etc.). The amount of antigen in the sample is then directly proportional to the amount of labeled antibody captured on the solid phase.

Both assays require differentiation of bound label from free label. This can be achieved by two methods: heterogeneous assay or homogeneous assay. Heterogeneous assay separates bound label from free label using a means of removing the antibody. Homogeneous assay compares the signal of the label when antigen is bound to antibody compared to when the antigen is free.

RADIOLABELED IMMUNOASSAYS

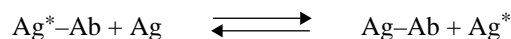
The earliest (1) immunoassays made use of radioactively labeled antigens or antibodies. These analytes are referred to as radiolabeled immunoassays. Antibody binding sites are extremely specific and to retain this specificity the best option would be to replace a nonradioactive isotope in the tracer molecule by its radioisotope (e.g., replace hydrogen by ^3H). However when the substitution is made in a part of the molecule away from the antibody binding site, the choice of radioisotope can be governed by other considerations, such as half-life, availability, high activity, and radiochemical purity.

The most common radioactive label is ^{125}I , which is bound to antibodies and antigens by a variety of techniques including chloramine-T, iodogen, or lactoperoxidase iodination. Iodination with ^{125}I is the preferred radiolabeling technique because the isotope is a γ -emitter, inexpensive, can be easily detected, has an appropriate half-life for most analytical purposes, and can be obtained in preparations with high specific activity. Problems occasionally associated with radioiodination include loss of immunoreactivity due to the size of the label or chemical alteration of reagents due to the high energy of decay (6). Use of beta-emitting isotopes (^{14}C and ^3H) may resolve these problems, but require more complicated scintillation counting for label measurement.

Labeled Antigen Radioimmunoassays

Labeled antigen radioimmunoassays involve competition between a labeled antigen and an unlabeled antigen

(analyte) for a limited number of antibody-combining sites as shown in the following reactions:



where Ag^* is labeled reagent antigen, Ag is analyte antigen, and Ab is antibody against Ag .

The Ag^* and Ab are analytical reagents with concentrations fixed so that when a sample containing Ag is added, competition between Ag and Ag^* for Ab binding sites occurs. Increasing concentrations of Ag result in a lesser degree of binding of Ag^* and the measurement of radioactivity of the binding or free fraction can be used to determine the amount of Ag present in a sample.

Determination of the distribution of Ag^* between the antibody-bound and free fractions, in radioimmunoassay procedures, requires a separation step which isolates Ag^* from Ag^*-Ab . Once separated, measurement of the radioactivity of the label in one or both the fractions provides a signal that is approximately exponentially related to the concentration of unlabeled antigen. The graphical relationship can have either a positive or a negative slope depending on whether the antibody-bound or free fraction is measured. The exact shape of the calibration curve however, is dependent on the antibody-binding equilibrium constant (7).

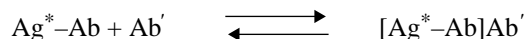
A variety of methods has been applied to the separation of bound and free Ag^* . These include precipitation, solid phase attachment, capillary electrophoresis, chromatography, and microfiltration. Originally, precipitation and solid-phase extraction were the most common types of separations techniques. However the ease of automation of capillary electrophoresis and flow-injection analysis (chromatography) makes these two techniques very interesting.

Precipitation

Precipitation techniques can be categorized into two general classes, nonspecific and specific. The nonspecific separations involve the addition of a salt or solvent that decreases the solubility of the antigen-antibody complex under conditions that do not affect the free-labeled antigen. After addition, the immune complexes can be precipitated by centrifugation and the radioactivity in either the supernatant solution or the precipitate can be measured. Examples of precipitation reagents used in immunoassays include alcohol, ammonium sulfate, polyethylene, and dioxane. Care must be taken to avoid coprecipitation of the unbound label.

Specific precipitation is a technique that has been devised to overcome some of the problems associated with nonspecific techniques. In this approach, a second

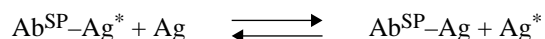
antibody (Ab'), specific to the analytical antibody is added in excess to cause precipitation of the primary antigen–antibody complex as shown below:



The method has been termed the double-antibody technique and can be used for a wide variety of analytes (8). The double antibody techniques generally require more time because the second antibody reaction can require days to reach equilibrium. The speed of precipitation clearly depends on the concentration of the second antibody. Polyethylene glycol has been used as a cosolvent to increase the precipitation rate.

Solid phase techniques

Many early immunoassays used solid phases to separate the labeled antigen from the complex by differential adsorption of the former. Examples of the solid phases used are dextran-coated charcoal, ion-exchange resins, and cellulose powder. Once the labeled antigen has been adsorbed, the bound, labeled antigen could be decanted to allow the measurement of the radioactivity of the bound and/or free-labeled materials separately. Because of problems with nonspecific adsorption and errors due to stripping of the labeled antigen from the solid, these types of separations have been largely replaced by those employing solid phases in which the reagent is covalently bound to the solid, as for example, in the solid phase attachment of a specific antibody (Ab^{SP}), which binds labeled and unlabeled antigen competitively as shown below:



The solid phase to which the antibody is bound can be either suspended in solution, on particles such as cellulose, agarose, or dextran beads or can be attached to the surface of a test tube or a microtiter plate. The stationary solid phases eliminate the centrifugation step that is necessary with the suspended beads although plastic solid phases have a limited capacity of binding proteins.

Capillary electrophoresis

This technique has proven to be a powerful separation technique for the separation of macromolecules, such as antibodies, for two reasons: the near flat plug flow profile and the small diffusion constant of the antibodies. These characteristics eliminate band broadening. With both superior separation power and high detection sensitivity, capillary electrophoresis (CE) can separate free Ab and Ag

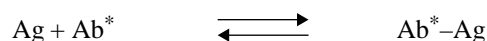
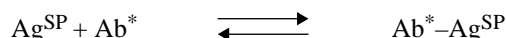
from bound Ab and Ag rapidly and is suitable for immunoassays. CE can combine immunologic recognition with on-line quantitation, microscale analysis, and automatic instrumentation to offer unique advantages for immunoassays. In immunoassays CE can be coupled to all of the existing CE detection techniques from UV and laser induced fluorescence (LIF) to mass spectrometry (MS). CE coupled to LIF has been applied to the determination of a variety of compounds including therapeutic drugs, peptides, and antibodies (9).

Flow-injection analysis

This technique introduces a sample into a flow of reagents. The reaction proceeds during transport in a reactor, after which the product is measured downstream in a detector. By combining immunoassays with flow-injection analysis (FIA), the long incubation times usually associated with heterogeneous assays become irrelevant, since FIA exploits nonequilibrium conditions. The washing steps are performed automatically by the continuous flow of buffer. Besides the irrelevance of incubation times, another advantage over normal immunoassays is that the reactor can be used several times because regeneration by using a suitable agent (e.g., glycine HCl) is possible. Sequential injections can be used to deliver reagents, substrate (if necessary), and regeneration agents to the immunoreactor. Many different assays using electrochemical, fluorescent and chemiluminescent detection have been developed.

Labeled Antibody Radioimmunoassays (Immunoradiometric Assays)

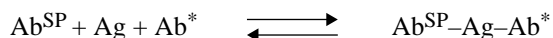
The problem of binding proteins to solid phases can be avoided by employing a solid phase antigen (Ag^{SP}) and a labeled antibody (Ab^*) as shown in the following reactions:



Where the unlabeled antigen (the analyte) competes for antibody binding sites with the solid phase antigen. This approach may have some difficulties caused by the requirement for purification of the immunospecific antibody. Because monoclonal antibodies are produced as essentially immunospecifically pure populations, they are ideal for labeled antibody techniques.

Another approach that utilizes labeled antibodies is the so-called “sandwich technique” (10). In this method, the

analyte antigen is bound both by an antibody that is attached to a solid phase (Ab^{SP}) and a labeled antibody as follows:



The reaction is noncompetitive and thus, the concentration of unlabeled antigen has a direct relationship with the amount of label bound in the complex. The sandwich radioimmunoassays are both sensitive and convenient but are limited to analytes which are, at least, bivalent; i.e., antigens that can provide at least two sites for antibody attachment. Also, highly purified antibodies are required for the sandwich technique.

ENZYME IMMUNOASSAYS

Early work demonstrated the use of enzymes coupled to antibodies or antigens as reagents in immunoassay. Enzyme activity can be measured in a variety of ways, each with certain advantages, which makes a variety of enzymes good labeling substances. Most assay methods are based on spectroscopic properties derived from an enzymatically transformed substrate. These methods are colorimetry, fluorometry, luminometry, and electrometry.

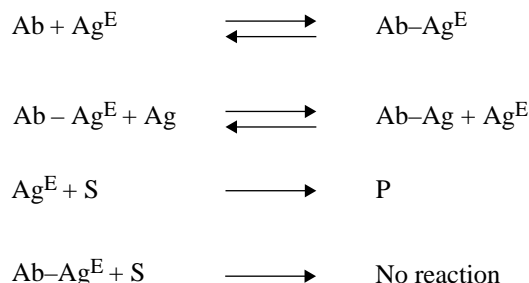
The development of enzyme immunoassays was pioneered in 1972 by Engvall and Perlmann (11) and by Van Weemen and Schuurs (12) and is translated into a wide variety of enzyme-based systems used both in research and in routine analysis with sensitivities approaching those of radioimmunoassays. Enzyme immunoassays (EIAs) can be divided into two major classes, homogeneous and heterogeneous.

Homogeneous Enzyme Immunoassays

Homogeneous immunoassay (HOIA) does not require physical separation of the free and antibody-bound antigen because the measured physical signal derived from the antibody-bound, labeled material may be significantly different from that of the unbound entity. There may be an enhancement or an inhibition of enzyme activity upon binding of the antibody to the antigen. HOIA are simple to perform and automation can be carried out easily (13). Elimination of the separation step avoids a major source of imprecision in the assay. However selectivity may be compromised since interfering substances are not eliminated in the separation step.

HOIA using an enzyme labeled antigen

Rubenstein et al. (14) described a HOIA method for morphine using lysozyme as the enzyme label. The covalent enzyme labeled antigen (Ag^E) competes with sample antigen (Ag) for a limited concentration of antibody (Ab) to form a complex. The resultant complex exhibits very little enzyme activity because of either steric hindrance (14) or allosteric inhibition (15) caused by the bound antibody. In the presence of Ag there is competition for the Ab leaving more Ag^E uncomplexed and free to catalyze the conversion of substrate to product. Thus, the enzyme activity, which can be measured by either the appearance of product (P) or disappearance of substrate (S), is directly proportional to the amount of free antigen in the sample (16).

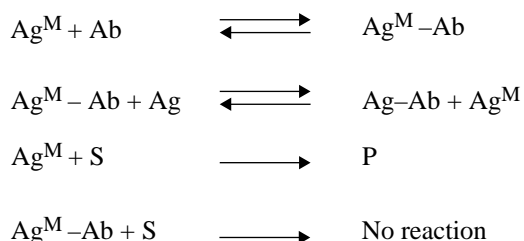


HOIAs based on this principle have been developed for a number of therapeutic agents and endogenous compounds using glucose-6-phosphate dehydrogenase (G-6PDH) under the name EMIT^R (Enzyme Mediated Immunoassay Technique). Enzyme activity is conveniently measured by absorption spectroscopy following the production of NADH from NAD which absorbs light strongly at 340 nm. The assay is rapid and sensitive to picomole levels (17).

HOIA using an antigen labeled enzyme modulator

This method is based on stability of an antigen labeled with an enzyme modulator (Ag^M) to modulate the activity of an indicator enzyme. The Ag^M competes with free antigen (Ag) for a limited amount of antibody (Ab). The Ag^M , on binding with Ab , is unable to modulate the activity of the indicator enzyme. As the concentration of analyte increases it competes successfully for binding sites on the antibody, leaving more Ag^M free to complex with indicator enzyme, thereby, modulating its activity (18, 19). A positive modulator will increase while a negative modulator will decrease enzyme activity. Thus, in the case of a positive modulator the enzyme activity will be directly proportional to the concentration of the analyte and

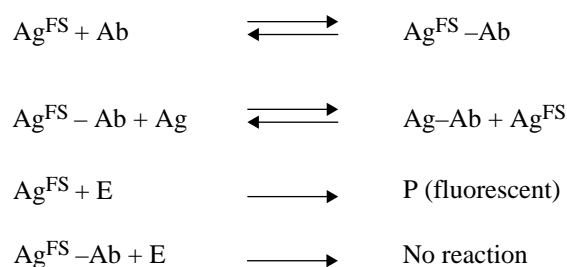
for a negative modulator the activity will be inversely proportional to the concentration of the analyte.



Based on this principle, practical assays for human serum thyroxine (20) and theophylline (21) have been developed and different distinct classes of modulators have been investigated.

HOIA using an antigen labeled with a fluorogenic enzyme substrate

This method was originally developed by Burd and Wong (22). The antigen is linked to a fluorogenic enzyme substrate (Ag^{FS}) to form a stable conjugate, which competes with the sample antigen (Ag) for a limited concentration of antibody (Ab). The antigen conjugated substrate is a fluorogenic substrate for the enzyme that reacts only when it is not bound to the Ab. Thus, at high concentrations of sample antigen, more of the Ag^{FS} would remain free to act as the substrate for the enzyme and more products would be formed. Thus, fluorescence intensity increases with increasing concentration of the sample antigen.



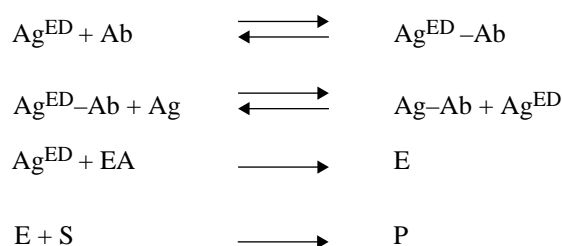
A derivative of umbelliferyl- β -galactoside serves as a fluorogenic substrate for *E. coli* β -galactosidase in this system (23) and solid-phase reagent strips based on this method have been developed (24).

HOIA using cloned enzyme donor immunoassay (CEDIA^{R})

This technique uses two different inactive enzyme fragments, a large fragment called enzyme acceptor (EA) and a very small fragment called enzyme donor (ED)

(25). Those fragments can associate to give an active enzyme.

In this technique, the antigen is labeled with enzyme donor (Ag^{ED}) capable of contributing to enzyme activation with enzyme acceptor (AD). When the labeled antigen binds to an antibody (Ab), the enzymatic activity is lost. The Ag^{ED} conjugate competes with sample antigen (Ag) for antibody complex formation. An increase in sample antigen concentration, therefore, leads to increased displacement of Ag^{ED} , leaving more free Ag^{ED} to associate with EA, resulting in more enzyme activation to form product (P). The enzyme activity is directly proportional to the amount of free sample antigen.



CEDIA s are available for the important relevant therapeutic drugs and most have picomolar detection limits.

Heterogeneous Enzyme Immunoassays

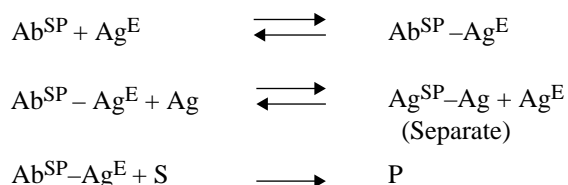
Heterogeneous immunoassays (HEIA) have at least one separation step which allows the differentiation of bound from free material. The enzyme-linked immunosorbent assay (ELISA), is a heterogeneous immunoassay which has been widely used. In this method, either antigen or antibody is immobilized on a solid phase. An essential difference from HOIA is that in HEIA the enzyme label is designed to retain its activity even after its reaction with the antibody. In comparison with RIA, non isotopic HEIA such as enzyme labeled assay has several important advantages, including better reagent shelf life, fewer health hazards and simpler equipment required. HEIA can be divided into the classes of competitive and noncompetitive assays.

Competitive assays

Enzyme-labeled antigen conjugate: In this assay, the enzyme labeled antigen (Ag^{E}) competes with sample antigen for a limited amount of antibody which has been immobilized on a solid-phase, for example, polystyrene (Ab^{SP}). After incubation, the unbound Ag^{E} is separated by washing with a detergent solution. The solid phase Ab^{SP} containing bound labeled and unlabeled antigen is incubated with a substrate (S) and the product

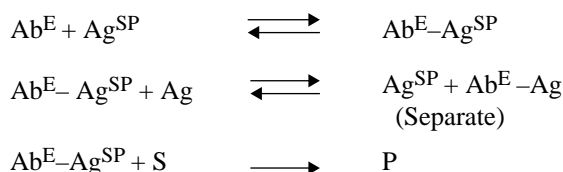
concentration is determined using a colorimeter or fluorimeter.

The enzyme activity or product concentration is inversely proportional to the concentration of sample antigen.



This method is very important for the measurement of low-molecular mass analytes such as steroids or melatonin or drugs such as cocaine (26). The most sensitive of these assays can detect <1 fmol analyte.

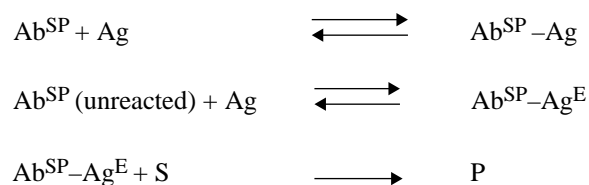
Enzyme-labeled antibody: This assay employs labeled antibody (Ab^{E}) and the antigen is attached to the solid phase (Ag^{SP}). The binding of Ag^{SP} to Ab^{E} is decreased by the addition of sample Ag.



The enzyme activity is inversely proportional to the concentration of sample Ag. Human IgG at the picomole level has been quantified in less than 1.5 h with this method (27). Estrone-3-glucuronide at the femtomole level has been quantified with this method.

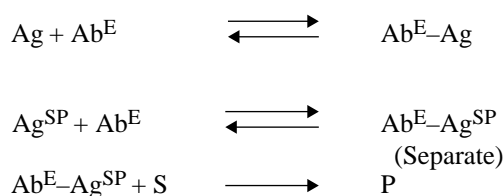
Noncompetitive assays

Enzyme-labeled antigen: In this method, the sample is first incubated with a moderate excess of solid-phase immobilized antibody (Ab^{SP}). After washing, excess enzyme-labeled antigen (Ag^{E}) is allowed to bind to unreacted Ab^{SP} .



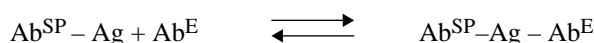
The concentration of P is inversely proportional to the concentration of standard or test antigen.

Enzyme-labeled antibody: The sample antigen (Ag) is incubated with a moderate excess of enzyme-labeled antibody (Ab^{E}). The mixture is then added to an excess of immobilized antigen (Ag^{SP}) to remove unreacted Ab^{E} .



The enzyme activity is inversely proportional to the concentration of sample and the procedure has been used to measure α -fetoprotein (28).

Sandwich or double antibody: This method is used with antigens having multiple antibody binding sites (epitopes). Immobilized unlabeled antibody (Ab^{SP}), in excess, is incubated with sample or antigen. After washing, the antibody-antigen complex is then incubated with an excess of enzyme-labeled antibody (Ab^{E}) which binds to one or more antigenic sites to form a sandwich type complex.



In this case, the concentration of enzyme product is directly proportional to the concentration of sample antigen.

Sandwich type assay is well suited for quantifying antigens with multiple antigenic determinants, such as antibodies, rheumatoid factors, polypeptide hormones, proteins, and hepatitis B surface antigens. The results obtained with sandwich type immunoassays are comparable to those obtained with those using radiolabels in term of precision, convenience and sensitivity. Highly sensitive thyroid-stimulating hormones assays are developed, as well as assays to measure estrogen and estradiol at attomole levels (29).

Avidin-Biotin Systems

An integral necessity of every enzyme immunoassay is the availability of enzyme-labeled antibody or antigen conjugates. The assays discussed in the first part of this article used chemical methods for forming these conjugates. An alternative approach has been widely employed that is based on the avidin-biotin reaction. Avidin is a protein of molecular weight about 60,000 that is found in egg white. It consists of 4 identical subunits. Each subunit binds biotin with an extremely high affinity. A dissociation constant of 10^{-15} corresponding to a dissociation half-life of about 160 days was reported. Hence, from a practical point of view, this binding can be regarded as almost irreversible in nature.

The rationale behind the use of the biotin-avidin system in immunochemistry is that avidin binds biotin

substituted structures. Peptides and proteins can be easily biotinylated via their amino groups by using activated biotin-*N*-hydroxysuccinimide.

Reagents for biotinylation of proteins and peptides are commercially available and allow the fast and efficient derivatization of antibodies and enzymes without the loss of enzyme or antibody activity. Several reagents are also available for the biotinylation of sulfhydryl groups aldehydes, nucleic acids and carbohydrates. Hence, in addition to proteins and peptides, a variety of antigens can be labeled with biotin.

Avidin can be coupled to enzymes by covalent binding techniques or by employing the so-called avidin–biotin complex (ABC). In the latter case, a complex between avidin and polybiotinylated enzyme is formed. The ratios of avidin to polybiotinylated enzymes are chosen in such a way that the resulting complex incorporates a number of enzyme molecules, but retains free biotin binding sites for the interaction with biotinylated antibody or antigen structures. A great selection of ABC complexes for a great variety of different enzymes is commercially available. The possibility of easily making biotinylated antibody or antigen structures and the availability of a variety of enzyme-labeled avidin derivatives or complexes facilitates the flexible design of enzyme immunoassays.

The most frequently used assay version uses a primary antibody which is immobilized to a solid support (e.g., the surface of a microtiter plate). In the first step, the antigen (analyte) is added to the wells. After the antigen is bound to the immobilized antibody, the wells are washed and a biotinylated second antibody is added to the tubes. The resulting sandwich between immobilized antibody; antigen and biotinylated antibody is incubated with the ABC complex. After washing, the antibody–antigen–biotinylated antibody–ABC complex is detected by addition of the enzyme substrate. The non competitive nature of this assay in conjunction with the extremely high amplification of the signal by the ABC complex results in an assay (in the femtomole range), that often exceeds the sensitivity of radioimmunoassays.

Competitive immunoassays based on the avidin–biotin approach have also been described. In this case, the antibody is immobilized on microtiter plates. Wells are incubated with a constant concentration of biotinylated antigen in the presence of different concentrations of standards or the sample. After washing, the ABC complex is added. The formed antibody–biotinylated antigen–ABC sandwich is then detected by the addition of enzyme substrate. The immobilized enzyme–activity is inversely proportional to the concentration of analyte in the sample, resulting in a typical sigmoidal calibration curve.

FLUORESCENCE IMMUNOASSAYS

Fluorescence immunoassay (FIA) entails the measurement of the photoexcited fluorescence from an antibody or antigen which participates in an immunochemical binding system, and whose spectral properties vary with the concentration of the analyte. Because few naturally occurring antigens and antibodies have the requisite fluorescent properties, they are usually substituted with fluorescent labels. Fluorescent labels are used in homogeneous and heterogeneous immunoassay systems and may be bound to competing (labeled and unlabeled) antigens; antibodies or solid phases or they may exist in solution as parts of enzyme substrates.

The fluorogenic molecules used to covalently label the antigens or the antibodies to be used in a fluorescence immunoassay are called fluorescent probes or labels. Fluorescent probes are small molecules whose fluorescent properties are altered subsequent to interactions with proteins or other macromolecules.

In quantitative immunoassay using fluorescent labels, there are several chemical and spectroscopic properties that the fluorochrome should possess. The labeled ligand should have a relatively high water solubility because immunoassays are usually carried out in the aqueous environment in which the antibodies are stable. The presence of certain functional groups on the fluorochrome will facilitate its conjugation to the ligand. Derivatives of the fluorochromes containing reactive groups such as acid chlorides, isothiocyanates and diazonium salts can be used for the conjugation of the label to the ligand. A major concern in labeling a ligand with a fluorescent probe is the possibility of altering the specificity of the ligand for its antibody. Consequently the site of the conjugation chosen should provide maximum exposure of functional groups necessary for antibody recognition. In addition, the stability and shelf life are also important considerations.

The spectral characteristics of greatest importance in the selection of a fluorochrome are the molar absorptivity at the selected wavelength of excitation, the quantum yield of fluorescence of the labeled species, the spectral regions of absorption and emission of radiation, and the Stokes' shift. Ideally, the fluorescent label should have an absorption spectrum with a high molar absorptivity in the visible region of the electromagnetic spectrum, well removed from the excitation spectra of proteins and other endogenous interferences normally present in biological fluids. The emission wavelength should also lie well into the visible region with Stokes' shift (displacement of the fluorescence maximum from the longest wavelength absorption maximum) of at least 50 nm. A high quantum

yield of fluorescence in the antibody-bound or free labeled ligand is also desirable.

The labels most commonly used are fluorescein isothiocyanate (FITC) and a number of reactive rhodamine dyes. Fluorescein isothiocyanate has a relatively high quantum yield and can be conjugated to drugs and other ligands under fairly mild conditions (30, 31).

Ligand labeling with fluorescent metal chelates has created a versatile class of fluorescent probes. The chelates of rare earth metals have unique emission characteristics in that, upon excitation of aromatic portions of the ligands of the lanthanide complex, the energy of excitation is efficiently transferred to the lanthanide ion. This causes $f-f$ transitions that produce very narrow almost line-like emission bands that permit all of the emitted light to be collected by the detector with narrow emission slits. In addition, the rare earth chelates possess large quantum yields in combination with very large Stokes' shifts. The excitation region of these chelates is fairly broad and these aspects of the lanthanide luminescence permit excellent sensitivity and selectivity by enabling the use of fairly wide bandwidths for excitation and narrow bandwidths for emission. With the proper combination of rare earth chelate and fluorimetric technique, very sensitive immunoassays can be developed that avoid a large number of interferences commonly encountered in the immunoassay process.

Homogeneous Fluorescence Immunoassay

Because of the extreme sensitivity of the quantum yield and spectral position of fluorescence to the microenvironment of the fluorophore, fluorescence immunoassay often lends itself well to homogeneous techniques. The elimination of the necessity of a step for the separation of bound and free ligand represents one of the major advantages of fluorescence immunoassay and provides the opportunity for simple, fast and reliable quantitation. The free and antibody-bound ligands reside in different microenvironments. In the aqueous environment, the free labeled antigen will experience strong polarizing forces as a result of interactions with water molecules. These forces will be exerted to different degrees in the ground and excited states of the labeled antigen because these electronic states have different dipole moments. Also, certain functional groups on the fluorophore may be free to rotate prior to and subsequent to the fluorescent transition in water. In the hydrophobic environment of an antibody binding site, the dielectric strength is low and rotation of functional groups on the fluorophore is severely restricted. The solvation and restricted rotational freedom of the antibody-bound labeled antigen usually cause this species

to fluoresce at shorter wavelengths than the free labeled antigen because the relative stabilization of the excited state of the latter by strong electrostatic and electromeric interactions is much greater than in the bound labeled antigen. Moreover, the weak solvation and restricted rotational freedom of the antibody-bound probe also cause the bound probe to fluoresce more intensely than the free probe because the bound probe is somewhat shielded from internal conversion which competes with fluorescence for deactivation of the excited state.

If the fluorescent emission spectrum of the bound labeled ligand is sufficiently displaced, enhanced or decreased in intensity (quenched) relative to that of the free labeled ligand, the resulting spectroscopic measurements can be used for quantitation without a separation step. Additionally, the techniques previously described in enzyme immunoassays, such as reactant-labeled immunoassay, can form the bases of fluorescent immunoassay. For example the fluorophore, from whose optical properties quantitation is derived, can be generated or consumed in an enzymatic reaction.

Two rather interesting variations on fluorescence immunoassay that do not require the fluorescence spectra of free and bound labeled materials to have different spectral positions have fairly recently become rather popular in clinical analysis. These are fluorescence polarization immunoassay and time-resolved fluorescence immunoassay.

Fluorescence polarization immunoassay (FPIA)

The physical principle underlying fluorescence polarization immunoassay involves the selective elimination of light waves whose electric vectors do not all lie in a single plane. This is accomplished by passing the exciting light through a polarizing filter. The resulting polarized radiation will selectively excite (photoselect) those molecules whose absorption transition moments have a significant component in the plane of the electrical vector of the exciting beam (32). As a result, molecules excited with polarized light will emit radiation that is polarized in the same direction as the exciting light, to a degree inversely related to the amount of Brownian rotation occurring during the interval between absorption and emission of light (33). This means that the photoselected molecules originally excited by polarized light and having fairly small volumes (i.e., free labeled antigen) will have random orientations with respect to the plane of polarization of the exciting light because they will rotate faster than they fluoresce. They therefore, will display very little polarized fluorescence. However, photoselected molecules having very large volumes such as the antibody

proteins and their complexes, will rotate at a rate comparable to or slower than the rate at which they fluoresce. Consequently, randomization of fluorescent transitions moments will not occur in these large molecules and substantial fluorescence polarization will be observed.

On binding of an antigen to an antibody there will be a reduction or a restriction in the rotational Brownian motion of the fluorescent label. This will cause considerable polarization of the fluorescence along or perpendicular to the optical axis of the excitation polarizer, depending upon whether the fluorescence transition moment of the molecule is oriented closer to 0 or 90° to the transition moment associated with the absorption band excited. Let us first consider the case where the transition moments for excitation and fluorescence are parallel (or nearly so).

If a second polarizing film (emission polarizer) is placed between the fluorescing sample and the photodetector of the fluorimeter, with its optical axis perpendicular to that of the polarizing film between the lamp and the sample, a much greater fraction of the highly polarized fluorescence from the antibody-labeled ligand will be filtered than would be of the unpolarized fluorescence from the same concentration of free labeled ligand excited under the same conditions. If the optical axes of both polarizers are parallel and the excitation and emission moments of the fluorophore are parallel, or nearly so, the emission polarizer will pass relatively more radiation from the bound labeled ligand than from the free labeled ligand, to the detector, because the unpolarized emission will be dispersed over all angles to the optical axis of the emission polarizer and some will therefore be filtered. Regardless of the orientation of the optical axis of the second polarizer with respect to the first, the fluorescence intensity registered by the detector should, ideally, be the same for unpolarized fluorescence (i.e., that of the free labeled ligand) while, in the case of parallel absorption and fluorescence transition moments, the intensity of the polarized fluorescence, from the bound labeled ligand, measured when the optical axes of the polarizers are parallel (F_{\parallel}) should be greater than when the optical axes of the polarizers are perpendicular (F_{\perp}). In the case of perpendicular absorption and fluorescence transition moments, for the free labeled ligand $F_{\parallel} = F_{\perp}$ (unpolarized fluorescence) and for the bound labeled ligand $F_{\perp} > F_{\parallel}$ (polarized fluorescence). We now define the degree of polarization as:

$$P = (F_{\parallel} - F_{\perp}) / (F_{\parallel} + F_{\perp}) \quad (1)$$

For a free labeled ligand $F_{\parallel} = F_{\perp}$ so that $P = 0$ at all excitation wavelengths. For a bound labeled ligand it is possible to have $+1/2 > P > -1/3$, depending upon the

wavelength of excitation. However, if a system is contrived which originally contains all antibody-bound, labeled ligand, upon addition of the unlabeled ligand the labeled drug will be displaced. The relative increase in unpolarized fluorescence and decrease in polarized fluorescence from the solution will cause a net decrease in P as calculated from its operational definition in Equation (1). If all the labeled ligand were ultimately displaced from the antibody complex, P would fall to zero. Depending on the extent of binding of the labeled ligand, the degree of polarization varies between some non zero value and zero. This permits the construction of a calibration curve of degree of fluorescence polarization versus concentration of unlabeled ligand and permits the execution of a homogeneous immunoassay.

The use of polarized fluorescence for the quantitation of several antigen-antibody reactions is widely used. Two assays of particular interest are those applied to the measurement of serum levels of gentamicin and phenytoin (30, 34, 35). The use of fluorescence polarization as a method for routine drug level determinations is limited by the light energy losses in the polarizing films and the background interferences that result in reduction of sensitivity.

Time-resolved fluoroimmunoassay

All of the fluorimetric techniques, so far considered, have been based on the measurement of the intensity of fluorescence produced under "steady state" conditions. "Steady-state" fluorimetry is derived from the excitation of the sample with a continuous temporal output of exciting radiation. The lamps and their power supplies used in conventional fluorimeters are sources of continuous radiation. After a short period of initial excitation of the sample, a steady state is established in which the rate of excitation of the analyte is equal to the sum of the rates of all processes deactivating the lowest excited singlet state (fluorescence, internal conversion, and intersystem crossing). When the steady state is established, the observed fluorescence intensity becomes time invariant and produces the temporally constant signal which is measured by the photodetector. With the development of modern electro-optics however, it has become possible to excite a potentially fluorescent sample with a thyratron pulsed flash lamp which emits its radiation in bursts of 2–10 ns duration with about 0.2 ms between pulses or with a pulsed laser whose pulses occur with durations upward of a few picoseconds. A fluorescent sample excited with such a pulsed source will not fluoresce continuously. Rather, its fluorescence intensity, excited by a single pulse will decay exponentially until the next pulse again excites the sample. The pulsed source then acts very

much as does a mechanical chopper in phosphorimetry. The fluorescence from the sample excited by the pulsed source can be represented, after detection, as a function of time on a fast sampling oscilloscope or on an x - y plotter used in conjunction with a multichannel pulse analyzer. The former approach is called pulsed-source fluorimetry and the latter, time-correlated single photon counting. In either case, fluorescence with decay times much longer than the lamp pulse characteristics can be treated in the same way that radioactive decay curves are analyzed. A semilogarithmic plot of fluorescence intensity against time will yield a straight line (or a series of intersecting lines if several fluorophores have comparable, but not identical decay times) whose slope is proportional to the decay time and whose vertical axis intercept can be compared with that of a standard solution of the fluorophore for quantitative analysis. If, however, the lamp pulse time and the decay-time of the fluorophore are comparable, the lamp characteristics must be subtracted from the observed signal to obtain the fluorophore's decay characteristics. This is usually accomplished by using a computer to solve a deconvolution integral representing the composite temporal characteristics of the lamp and the fluorophore output.

The pulsed-source (time-resolved) method, then effects spectroscopic separation of the emission of several fluorescing species by taking advantage of differences in their decay times rather than their fluorescence intensities. This means that several overlapping fluorescences, such as those of free and antibody-bound ligand can be quantified simultaneously. Lanthanide chelate-labeled antibodies, which have long lifetimes (in the $1 = \text{ms}$ to $1 = \mu\text{s}$ range) form the basis of a time-resolved FIA, for which instrumentation is commercially available.

Heterogeneous Assays

Most research and development of fluorescence immunoassays, to date, have been concentrated in the area of homogenous assay. Unfortunately, homogeneous assays may not always provide optimum analytical sensitivity. This may be due to the endogenous background fluorescence of proteinaceous materials, Rayleigh and Raman scatter caused by proteins or to the lack of environmental sensitivity of the emission from the fluorescent label. Notwithstanding the speed and simplicity of the homogenous assay procedures, it is occasionally desirable to circumvent these problems by physically separating the antibody-ligand complex from other species in the sample before fluorimetric quantitation.

The separation of the antibody-bound ligands from free ligands and other fluorescing species present in solution can be accomplished by various methods based on the

chemical, physical or immunological differences between the free ligand and the antibody-ligand complex and include gel permeation chromatography, chemical precipitation with inorganic salts or organic solvents and double antibody methods as well as the use of a solid phase support to which the antibody is either adsorbed or bonded covalently. In the latter method, the solid material may be paper discs, the walls of test tubes, glass or plastic beads, cross-linked dextrans or agaroses. Once the antibodies are immobilized on the solid support, the labeled and unlabeled ligands are introduced and allowed to compete for available binding sites on the antibody. The bound, labeled fraction is then separated from the free labeled ligand by washing. The labeled ligand complexed to the antibody can then be measured directly, without removal from the solid phase, by a fluorimeter with a front-surface fluorescence attachment (36).

CHEMILUMINESCENCE IMMUNOASSAYS

In all exergonic reactions energy is released. This energy is generally emitted as heat. In some reactions, however, this energy is released as light, a phenomenon known as chemiluminescence. During the course of chemiluminescence reactions, one or more of the resulting products is formed in an electronically excited state. Light is then emitted from the excited molecules by the process of fluorescence. Chemiluminescent reactions are almost invariably oxidation-reduction reactions. A number of compounds that show chemiluminescence of an intensity that is suitable for analytical detection systems or that are catalysts or reagents in chemiluminescent reactions have been investigated to be used in chemiluminescence immunoassays. Among these are synthetic organic compounds (e.g., phthalazinediones, acridinium esters), cofactors in bioluminescent reactions (NAD and ATP) and enzymes (peroxidase, oxidases, kinases, luciferases).

Covalent linking to either the antigen or antibody is carried out by chemical modification of the label (e.g., diazotisation or reaction with isothiocyanate, *N*-hydroxysuccinimide, hemisuccinate, imidoesters), by chemical modification of antigen or antibody (e.g., by reaction with hemisuccinate, glutaraldehyde), or by conjugation using bifunctional reagents (e.g., mixed anhydride carbodiimide, bis (N-hydroxy succinimides) or azidosuccinimides).

One of the most intensely chemiluminescent compounds is luminol (5-amino-2,3-dihydrophthalazine-1, 4-dione). If the amino group is linked to position 6 of the aromatic ring, the molecule is called isoluminol.

Isoluminol can be easily attached to antibodies or antigens via alkyl spacer groups. Aminobutylethylisoluminol (ABEI) has been shown to be very effective as a chemiluminescence labeling group, since the activity is generally not changed by coupling to low- or high-molecular-weight compounds. It can be easily coupled to activated carboxylic groups of immunologically relevant compounds using the mixed anhydride reaction (37). Proteins such as antibodies can be labeled sufficiently with ABEI through reaction of free amino groups of the proteins with an ABEI-isothiocyanate derivative (38).

The oxidative reactions of luminol and isoluminol derivatives at high pH result in the formation of 3-aminophthalate or 4-aminophthalate and nitrogen via an electronically excited state. The transition from the excited to the ground state induces the emission of light having a wavelength maximum of 425 nm. Quantitation is possible at picomolar or even attomolar levels of the aminophthalhydrazide.

In aprotic solvents, only oxygen and a base are necessary for the oxidation. In protic solvents a catalyst has to be present, in addition. As catalysts, enzymes such as horseradish peroxidase are frequently employed.

Acridinium esters are used also in immunological procedures. They have the advantage that no catalyst is necessary for the luminescence reaction. The solution has to be adjusted to a pH of 6–7 before the oxidant peroxide is added in order to ensure the highest activity.

Direct Chemiluminescence Immunoassays

The above reactions can be used in an immunological assay by coupling the chemiluminescent compound to an antigen or antibody; under the assumption that the immunological and chemiluminescent properties of the derived coupling product are not substantially changed. The approach to coupling chemiluminescent molecules to the required ligand obviously depends on the nature of the two species. When both species are small molecules, conventional synthetic organic chemistry can be used. A problem arises however when one or both molecules are more complex (e.g., proteins). Since the functionality of these molecules is determined by the chemical and physical environment, the range of chemistries that can be used is limited. As an example of direct chemiluminescence immunoassay, antibody might be immobilized onto a solid support and antigen allowed to bind to this antibody. After subsequent washing, a second chemiluminescent-labeled antibody that recognizes a different epitope of the antigen is added. After further washing, the chemiluminescent label is activated and the emitted light is quantified.

Since this assay is based on the sandwich principle, increasing amounts of analyte will produce increased light emission.

A chemiluminescent assay using competitive immunological techniques can be easily designed when chemiluminescent antigens are employed. In this case chemiluminescent antigen and sample compete for immobilized antibodies. Depending on the nature of the reaction antigen, bound and unbound label can also be separated by adsorption onto charcoal.

Indirect (Enzyme-Mediated) Chemiluminescence Immunoassays

An indirect chemiluminescence immunoassay is an assay, with another component than the primary chemiluminescent emitter coupled to the antigen or antibody. This can be a cofactor or a catalyst or even a molecule capable of converting a nonchemiluminescent precursor to a chemiluminescent or potentially chemiluminescent species. Most indirect assays are enzyme mediated.

One widely used enzyme mediated chemiluminescence immunoassay (39) uses the firefly enzyme luciferase that catalyzes the oxidation of D-Luciferin in the presence of ATP. D-luciferin, but not luciferin esters such as phosphates, is oxidized in the presence of the enzyme. An antigen or antibody-enzyme conjugate (e.g., alkaline phosphatase conjugate) is bound to a solid support by an antigen-antibody reaction. The immobilized enzyme enzymatically releases D-luciferin, which subsequently is quantified in a luciferase based lumino-metric assay, here described for the determination of alkaline phosphatase.

The chemiluminescent signals of luminol derivatives can be enhanced by the addition of firefly luciferin to the reaction mixture using hydrogen peroxide as oxidant and horseradish peroxidase as catalyst (40). This causes prolongation of the light production. Benzothiazoles, such as dehydroluciferin and 6-hydroxybenzothiazole derivatives also enhance the light emission by a factor of about 500 to 1000-fold and show a more constant light emission over time than luminol. A relatively constant emission of light over a period of 15 min has been observed for these derivatives (41). This constant and prolonged light emission pattern simplifies the analytical procedures as multiple reactions can be initiated outside of the luminometer without the necessity of initiating the chemiluminescence reaction in front of the photomultiplier.

Another chemiluminescent enzyme system is based on the use of stabilized dioxetane substrates. Dioxetanes are intermediates in many chemiluminescent reactions. It's possible to synthesize stabilized dioxetanes (phosphatase

and β -galactose moieties) that do not spontaneously react. When exposed to the right enzyme (alkaline phosphatase and β -galactosidase, respectively) the dioxetane will be destabilized and spontaneously undergo a chemiluminescence reaction (42).

ELECTROCHEMICAL IMMUNOASSAYS

Electrochemical immunoassays include a wide variety of devices based on the coupling of immunological reactions with electrochemical transduction. All of them involve the immobilization of an immunoreagent component on the surface of the electrode transducer. Electrochemical detection is based on the direct intrinsic redox behavior either of an analyte species or of some reporter molecule. For the detection no expensive equipment is needed, with the measurement of either a simple current or a voltage charge. Different electrochemical detection strategies are used, but amperometric detection is most widely used. Potentiometric and conductometric detection are applied in different assays as well.

For the detection different electrode supports and a great variety of immobilization procedures have been used. Gold electrodes, screen-printed electrodes and carbon materials (such as graphite and glassy carbon) have been frequently used. The sensing phase is constructed onto the surface of the electrode by means of covalent linkage, physical adsorption or membrane entrapping of the specific immunoreagent.

The major advantage of electrochemical immunoassays is the fact that it's based on the use of a nonoptical detection system, which makes it possible to detect signals in the presence of whole blood samples. There are three different ways in which electrochemistry is used in a detection system.

Direct monitoring of the antigen–antibody reaction

This is by far the simplest approach to electrochemical immunoassay, because there is no label needed. There are several methods demonstrated by analyzing simple solutions, however, considerable work will be necessary for detection when sensors are exposed to biological sample matrices.

Using electroactive compounds as labels

In order to be used as an immunoassay label, an electrochemically active compound has to possess suitable electrochemical properties. It has to be soluble in aqueous media and should be stable in solution over a wide pH range. To be detectable, it must allow highly selective

electrochemical detection or possess chemical properties to allow selective membranes to be used in the measurement electrode.

One of the advantages of an intrinsically electroactive label is that there is no need for substrate, cofactor or the special incubation steps for an enzyme label. However a consequence is the loss of enzyme signal amplification which increases the demands on detection sensitivity. Human serum albumin has been detected in a competitive heterogeneous assay using indium ions (43).

Using enzymes as amplification labels with the enzyme monitored by measurement of an electro-active product or substrate

The direct detection of electrochemical labels entails problems with sensitivity. For this reason the majority of electrochemical immunoassay development has focused on the measurement of enzyme labels by detection of electroactive products arising from enzyme catalyzed reactions. A wide variety of enzyme labels have been used for electrochemical immunoassays. These include glucose oxidase, glucose-6-phosphate dehydrogenase and alkaline phosphatase (44).

Glucose oxidase catalyses the conversion of glucose to gluconic acid with reduction of oxygen to hydrogen peroxide. In principle, either the consumption of oxygen or the liberation of hydrogen peroxide can be monitored to detect the activity of glucose oxidase. However due to the high background level of oxygen in biological fluids, hydrogen peroxide monitoring is more often used. Other glucose oxidase assays have been used (e.g., ferrocene and 1,4-benzoquinone), where oxygen is replaced by an alternative electron acceptor because of faster reaction rates, lower oxidation potential and the detection system is less sensitive to the levels of oxygen in the sample. A homogenous assay based on ferrocene was developed for digoxin (45).

Glucose-6-phosphate dehydrogenase catalyses the conversion of glucose-6-phosphate to 6-phosphogluconate while reducing NAD^+ to NADH. However the direct electrochemical measurement of NAD(P)H is rather difficult, owing to by a high overvoltage that is required for the electrode reaction and by electrode fouling. By using redox mediators (e.g. quinones, ferrocenes, and phenoxazine) these problems can be overcome. Theophylline, and phenytoin have been detected using glucose-6-hydrogenase.

Alkaline phosphatase is one of the most suitable enzymes for electrochemical immunoassays owing to its high turnover number and broad substrate specificity. Different substrates have been used, but 4-aminophenyl phosphate is most suitable, since the reaction product,

4-aminophenol is easily oxidized without fouling of the electrode surface. Thyroxine-binding globulin, cortisol, and prostatic acid phosphatase have been detected by using alkaline phosphatase.

LIPOSOME IMMUNOASSAYS

Liposomes are formed when phospholipid molecules spontaneously self-assemble in aqueous solution to produce spheres in which an aqueous cavity is enclosed by one or more phospholipid bilayer membranes. This cavity can be used to entrap a variety of materials. For immunoassays, a variety of detectable labels, called markers, have been encapsulated. The markers used range from inorganic salts [e.g., KCl, $K_4Fe(CN)_6$] for electrochemical detection, through widely used fluorescent markers (e.g. calcein, carboxyfluorescein, sulforhodamine B and chelates) to enzymes (e.g., alkaline phosphatase, horseradish peroxidase, and glucose-6-phosphate dehydrogenase). A wide range of detection methods and substrates is available for these enzymes. The amount of the encapsulated molecule is directly related to the amount of antigen. To achieve the lowest limit of detection the amount of encapsulated molecule (marker) should be as high as possible. Fluorescent markers are most commonly used because these markers can be measured easily and very sensitively. Besides these mostly small fluorescent molecules larger molecules, as enzymes and enzyme cofactors are also entrapped. The advantage of using larger molecules is that they tend to leak less through the liposome membrane. Most markers are water soluble, but lipophilic markers (e.g., perylene derivatives) can also be incorporated in liposomes (46). These markers are mainly incorporated in the membrane which eliminates leakage.

The amount of liposome-encapsulated marker has to be related to the amount of antigen, therefore the liposome has to be conjugated to the antigen. This can be done either directly or indirectly. With the direct approach the liposome is bound to the antigen. The indirect approach either uses an antibody or a secondary molecule bound to the liposome.

Homogeneous Liposome Immunoassays

Because of the special structural characteristics of liposomes it's possible to develop adapted schemes exploiting these characteristics. It's essential that the liposomes retain their structural integrity in the presence of lytic agents, which might be present in biological fluids.

However most liposome immunoassays use a natural lytic agent, the two most frequently used are complement (47) and mellitin (48).

When foreign cells enter the human body they are captured by antibodies, which attach themselves to the cell membrane surface. Complement binds to these antibodies in a specific order, after which the target cells are lysed. Because the liposome bilayer is structurally similar to the cell wall, complement can be used to completely lyse antibody-bound liposomes. Attention has to be paid to the fact that although complement is specific, some liposomes are susceptible to complement lysis, without being bound to an antibody. Most liposome immunoassays are homogeneous complement-based assays.

Mellitin is also able to lyse liposomes completely, although rather slowly, it is found in bee venom and its biological task is cell destruction. Cytolysin mediated assays use a conjugate of mellitin with the antigen. The conjugate is free and able to lyse the liposomes or bound to an antibody and unable to lyse the liposomes. The activity is thus reversed, proportional to the amount of antibody present in the sample. The disadvantage of mellitin-based assays is that the range of antigens is limited in size, because the presence of a large antigen will inhibit the activity of mellitin. Biotin and digoxin have been analyzed by a homogeneous liposome immunoassay using mellitin as lytic agent.

Homogeneous liposome immunoassays using lytic agents other than complement and mellitin have also been used. Phospholipase C catalyzes the dephosphorylation of phospholipids, which in turn destabilizes the liposome. The assay is based on the inhibition of the lytic activity by an antibody binding to an antigen conjugated to phospholipase C (49). Gentamicin is analyzed by this method.

Assays that use liposomes coated with murine monoclonal antibodies and magnetic particles coated with antimurine monoclonal antibodies, have been developed. Incubation of these two components results in destabilization of the liposome by binding to the magnetic particles. As a result the entrapped marker will be released (50).

Heterogeneous Liposome Immunoassays

Compared to homogeneous assays only a small number of heterogeneous assays have been developed. As all heterogeneous assays, a heterogeneous liposome immunoassay or a liposome immunosorbent assay always use one or more steps to separate the specifically bound liposomes from the free liposomes by washing. After washing lysis of the liposomes can be performed with a detergent which effectively and quickly lyse all liposomes,

such as Triton X-100. These assays can be regarded as modifications of enzyme-linked immunoassays.

We have presented in this article a brief overview of existing immunological assays. New drug developments will, in the future, require more versatile and sensitive assays. It can be said with certainty that immunological assays will retain their importance for achieving these goals.

REFERENCES

1. Berson, S.A.; Yalow, R.S. Quantitative Aspects of the Reaction Between Insulin and Insulin-Binding Antibody. *J. Clin. Invest.* **1959**, *38*, 1996–2016.
2. Male, D.; Champion, B.; Cooke, A. Maturation of the Immune Response. *Advances Immunology*; J. B. Lippincott Co.: Philadelphia, 1987; 10.1.
3. Edelman, G.M.; Gall, W.E. Antibody Problems. *Ann. Rev. Biochem.* **1969**, *38*, 415–466.
4. Kohler, G.; Milstein, C. Continuous Cultures of Fused Cells Secreting Antibody of Predefined Specificity. *Nature* **1975**, *256*, 495–497.
5. Campfield, L.A. Mathematical Analysis of Competitive Protein Binding Assays. *Principles of Competitive Protein Binding Assays*; Odell, W., Franchimont, P., Eds.; John Wiley & Sons: New York, 1983; 125–148.
6. Izzo, J.L.; Roncone, A.; Izzo, M.J.; Bale, W.F. Relationship between Degree of Iodination of Insulin and its Biological Electrophoretic and Immunochemical Properties. *J. Biol. Chem.* **1964**, *239*, 3749–3754.
7. Potts, J.T., Jr; Sherwood, L.M.; O'Riordan, L.L.H.; Aurbach, G.D. Radioimmunoassay of Polypeptide Hormones. *Advan. Intern. Med.* **1967**, *13*, 183–240.
8. Midgley, A.; Hepburn, M.R. Use of Double Antibody Method to Separate Antibody Bound From Free Ligand in Radioimmunoassay. *Methods in Enzymology: Immunochemical Techniques*; Langone, J.J., Ed.; Academic Press: New York, 1980; 74, 266–273.
9. Bao, J.J. Capillary Electrophoretic Immunoassays. *J. Chromatogr. B.* **1997**, *699*, 463–480.
10. Salmon, S.E.; Smith, B.A. Sandwich Solid Phase Radioimmunoassays for Characterization of Human Immunoglobulins Synthesized In-Vitro. *J. Immunol.* **1970**, *104*, 665–672.
11. Engrall, E.; Perlman, P. Enzyme – Linkend Immunosorbent Assay (ELISA): Quantitative Assay of Immunoglobulin G. *Immunochemistry* **1971**, *8*, 871–874.
12. Van Weeman, B.K.; Schuurs, A.H.W.M. Immunoassay Using Antigen-Enzyme Conjugates. *FEBS Letter* **1971**, *15*, 232–236.
13. Brunk, S.D.; Hadjiioannou, T.P.; Hadjiannou, S.I.; Malmstadt, H.V. Adaptation of Emit Technique for Serum Phenobarbital and Diphenylhydantoin Assays to Miniature Centrifugal Analyzer. *Clin. Chem.* **1976**, *22*, 905–907.
14. Rubenstein, K.E.; Schneider, R.S.; Ullman, E.F. Homogeneous Enzyme Immunoassay – New Immunochemical Technique. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 846–851.
15. Rowley, G.L.; Rubenstein, K.E.; Huisjen, J.; Ullman, E.F. Mechanism by Which Antibodies Inhibit Hapten-Malate Dehydrogenase Conjugates. *J. Biol. Chem.* **1975**, *250*, 3759–3766.
16. Ngo, T.T. *Enzyme-mediated Immunoassay*; Plenum: New York, 1985.
17. Chang, J.; Gotcher, S.; Gunshaw, J.B. Homogeneous Enzyme-Immunoassay for Theophylline in Serum and Plasma. *Clin. Chem* **1982**, *28*, 361–367.
18. Ngo, T.T.; Lenhoff, H.M. Enzyme Modulators as Tools for the Development of Homogeneous Enzyme Immunoassays. *FEBS Letter* **1980**, *116*, 285–288.
19. Ngo, T.T.; Lenhoff, H.M. Antibody Induced Conformational Restriction as Basis for New Separation-Free Enzyme-Immunoassay. *Biochem. Biophys. Res.* **1983**, *116*, 1097–1103.
20. Finley, P.R.; Williams, R.J.; Lichti, D.A. Evaluation of a New Homogeneous Enzyme-Inhibitor Immunoassay of Serum Thyroxine with Use of a Bichromatic Analyzer. *Clin. Chem.* **1980**, *26*, 1723–1726.
21. Blecka, L.J.; Shaffar, M.; Dworschack, R. Inhibitor Enzyme Immunoassays for Quantitation of Various Haptens: A Review. *Immunoenzymatic Techniques*; Avrameas, S., Dmet, P., Mosseyeff, R., Feldman, G., Eds.; Elsevier: Amsterdam, 1983; 207.
22. Burd, J.F.; Wong, R.C.; Feeney, J.C.; Carrico, R.C.; Boguslaski, R.C. Homogeneous Reactant-Labeled Fluorescent Immunoassay for Therapeutic Drugs Exemplified by Gentamicin Determination in Human-Serum. *Clin. Chem.* **1977**, *23*, 1402–1408.
23. Bogulaski, R.C.; Li, T.M.; Benoric, J.L.; Ngo, T.T.; Burd, J.F.; Carrico, R.C. Substrate Labeled Homogeneous Fluorescent Immunoassays for Haptens and Proteins. *Immunoassays: Clinical Laboratory Technique for the 1980's*; Nakamura, R.M., Dito, W.R., Tucker, E.S., III Eds.; Alan Liss: New York, 1980; 45–64.
24. Walter, B.; Greenquist, A.C.; Howard, W.E., III. Solid-phase Reagent Strips for Detection of Therapeutic Drugs in Serum by Substrate-Labelled Fluorescent Immunoassay. *Anal. Chem.* **1983**, *55*, 873–878.
25. Engel, W.D.; Khanna, P.K. Cedia In Vitro Diagnostics with a Novel Homogeneous Immunoassay Technology – Current Status and Future Prospects. *J. Immunol. Meth.* **1992**, *150*, 99–102.
26. Spiehler, V.; Fay, J.; Rogerson, R.; Schorenforfer, D.; Nied Bala, R.S. Enzyme Immunoassay Validation for Qualitative Detection of Cocaine in Sweat. *Clin. Chem.* **1996**, *42*, 34–38.
27. Halliday, M.I.; Wisdom, G.B. Competitive Enzyme-Immunoassay Using Labelled Antibody. *FEBS Letter* **1978**, *96*, 298–300.
28. Masseyeff, R. Assay of Tumour-Associates Antigens. *Scand. J. Immunol.* **1978**, *8* (Suppl. 7), 83–90.
29. Klein, K.O.; Baron, J.; Colli, M.J.; McDonnell, D.P.; Cutter, G.B. Estrogen-Levels in Childhood Determined by Ultrasensitive Recombinant Cell Bioassay. *Clin. Invest.* **1994**, *94*, 2475–2480.
30. Watson, R.A.A.; Landon, J.; Shaw, E.J.; Smith, D.S. Polarization Fluoro-Immunoassay of Gentamicin. *Clin. Chim. Acta.* **1976**, *73*, 51–55.

31. Shaw, E.J.; Watson, R.A.A.; Landon, J.; Smith, D.S. Estimation of Serum Gentamicin by Quenching Fluoro-Immunoassay. *J. Clin. Path.* **1977**, *30*, 526–531.
32. Weber, G. Polarization of the Fluorescence of Macromolecules. *Biochem. J.* **1952**, *51*, 145–155.
33. Chen, R.F. Extrinsic and Intrinsic Fluorescence of Proteins. *Practical Fluorescence: Theory, Methods and Techniques*; Guilbault, G.G., Ed.; Ch. 12 Marcel Dekker, Inc.: New York, 1973.
34. McGregor, A.R.; Crockall-Grenning, J.O.; Landon, J.; Smith, D.S. Polarization Fluoro-Immunoassay of Phenytol. *Clin. Chim. Acta.* **1978**, *83*, 161–166.
35. O'Neal, J.S.; Schulman, S.G. Fluorescence Polarization Immunoassay of Phenytoin Employing a Sulfonamido Derivative of 2-Naphthol-8-Sulfonic Acid as a Label. *Anal. Chem.* **1984**, *56*, 2888–2891.
36. Blanchard, G.C.; Gardner, R. Two Immunofluorescent Methods Compared With a Radial Immunodiffusion Method for Measurement of Serum Immunoglobulins. *Clin. Chem.* **1978**, *24*, 808–814.
37. Kohen, F.; De Boever, J.; Kim, J.B. Recent Advances in Chemiluminescence Based Immunoassays for Steroid-Hormones. *J. Steroid Biochem.* **1987**, *27*, 71–79.
38. Patel, A.; Campbell, A.K. Homogeneous Immunoassay Based on Chemiluminescence Energy-Transfer. *Clin. Chem.* **1983**, *29*, 1604–1608.
39. Geiger, R.; Miska, W. Bioluminescence Enhanced Enzyme-Immunoassay: New Ultrasensitive Detection Systems for Enzyme Immunoassays. *J. Clin. Chem. Biochem.* **1987**, *25*, 31–38.
40. Whitehead, T.P.; Thorpe, G.H.G.; Carter, T.J.N.; Groucutt, C.J.; Kricka, L.J. Enhanced Luminescence Procedure for Sensitive Determination of Peroxidase – Labeled Conjugate in Immunoassay. *Nature (London)* **1983**, *305*, 158–159.
41. Thorpe, G.H.G.; Moseley, S.B.; Kricka, L.J.; Stott, R.A.; Whitehead, T.P. Enhanced Luminescence Determination of Horseradish Peroxidase Conjugates: Application of Benzothiazole Derivatives as Enhancers in Luminescence Assays on Microtitre Plates. *Anal. Chim. Acta.* **1985**, *170*, 101.
42. Bronstein, I.; Voyta, J.C.; Thorpe, G.H.G.; Kricka, L.J.; Edwards, B. Chemiluminescent Enzyme-Immunoassay of Alpha-Fetoprotein Based on an Adamantyl Dioxetane Phenyl Phosphate Substrate. *Clin. Chem.* **1989**, *35*, 1441–1446.
43. Doyle, M.J.; Halsall, H.B.; Heineman, W.R. Enzyme-Linked Immunosorbent Assay with Electrochemical Detection for Alpha-1-Acid Glycoprotein. *Anal. Chem.* **1982**, *56*, 2318–2322.
44. Heineman, W.R.; Halsall, H.B. Strategies for Electrochemical Immunoassay. *Anal. Chem.* **1985**, *12*, 1321–1331.
45. Suzawa, T.; Ikariyama, Y.; Aizawa, M. Multilabeling of Ferrocenes of a Glucose-Oxidase Digoxin Conjugate for the Development of a Homogeneous. *Anal. Chem.* **1994**, *66*, 3889–3894.
46. Schott, H.; Von Cunow, D.; Langhals, H. Labeling of Liposomes with Intercalating Perylene Fluorescent Dyes. *Biochim. Biophys. Acta.* **1992**, *1110*, 151–157.
47. Kinsky, S.C. Antibody-Complement Interaction with Lipid Model Membranes. *Biochim. Biophys. Acta.* **1972**, *265*, 1–23.
48. Litchfield, W.J.; Freytag, J.W.; Adamich, M. Highly Sensitive Immunoassays Based on Use of Liposomes Without Complement. *Clin. Chem.* **1984**, *30*, 1441–1445.
49. Kim, C.K.; Park, K.M. Liposome Immunoassay (LIA) for Gentamicin Using Phospholipase-C. *Immunol. Methods* **1994**, *170*, 225–231.
50. Wright, S.E.; Huang, L.; J. Liposome Res. **1992**, *2*, 257–273.

HYDROLYSIS OF DRUGS

Kenneth A. Connors

University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.

Jason M. LePree

Boehringer Ingelheim Pharmaceuticals, Ridgefield, Connecticut, U.S.A.

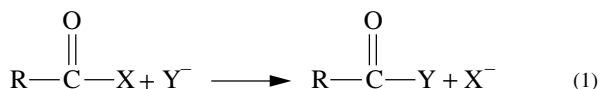
INTRODUCTION

The term “hydrolysis” describes a chemical reaction in which a chemical bond is split (lysis) via the addition of water. This reaction is one of the most important routes of drug decomposition, as it occurs frequently in active ingredients and excipients of pharmaceutical dosage forms. A search of the scientific literature, performed with a popular electronic database and using the combination of keywords “hydrolysis” and “drugs,” yielded over 3300 articles from January 1989 to September 1999. A more selective analysis of a collection of stability data on 91 drugs (1) shows that 70 of these undergo hydrolytic degradation reactions. Of these, 61% can be classified as reactions of carboxylic acid derivatives, 4.3% of phosphoric acid derivatives, 20% of carbonyl derivatives, and 14.3% are nucleophilic displacements, often on the aliphatic carbon. These classes are discussed in the subsequent sections of this chapter.

HYDROLYTIC REACTIONS AND MECHANISMS

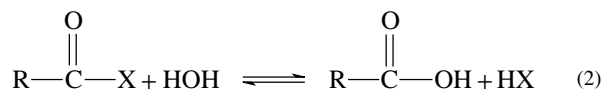
Carboxylic Acid Derivatives

Some drugs contain an acyl group, RCO^- in the compound RCOX , where X is called the leaving group. The most important types of acyl or carboxylic compounds, also called carboxylic acid derivatives are given in Table 1. The characteristic reaction of acyl compounds involves cleavage of the $\text{C}-\text{X}$ bond. The net result of the attack of a nucleophile Y^- on an acyl compound is cleavage of the $\text{C}-\text{X}$ bond and formation of the $\text{C}-\text{Y}$ bond; these reactions, shown in Eq. 1, are, therefore, called acyl transfers.

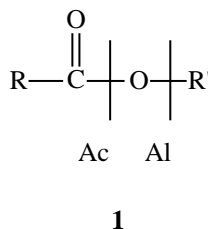


In a hydrolysis, H_2O (or OH^-) is the attacking nucleophile, and hydrolysis of a carboxylic acid derivative

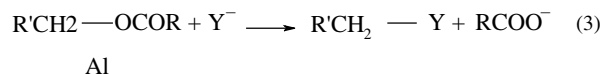
is an acyl transfer to water, as in Eq. 2.



An important example of acyl transfer reactions is ester hydrolysis, where two sites of bond cleavage, both leading to the same product, are possible. The possibilities, shown in structure 1, are denoted acyl-oxygen fission (Ac) and alkyl-oxygen fission (Al).



Alkyl-oxygen fission constitutes a nucleophilic substitution on carbon, shown in Eq. 3.



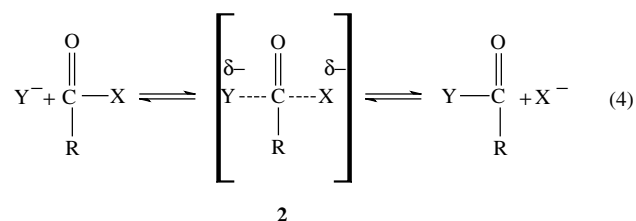
Several experimental methods have been devised to establish the type of cleavage in ester hydrolysis; in nearly all cases, it is found to be acyl-oxygen cleavage via nucleophilic attack on the carboxyl group. Ester reaction mechanisms are often discussed in terms introduced by Ingold (2), who classified the type of bond fission (Ac or Al), the reaction molecularity (1 or 2), and the ionic form of the substrate (A for the conjugate acid $\text{RC}(\text{OH})\text{OR}^+$ and B for the conjugate base RCOOR). Ingold's classification is shown in Table 2.

For the bimolecular reaction with Ac cleavage, two reasonable mechanisms have been suggested. The first is a direct displacement analogous to the $\text{S}_{\text{N}}2$ mechanism of aliphatic nucleophilic substitution. This route is shown in Eq. 4; structure 2 is the transition state (\ddagger), although it is

Table 1 Carboxylic acid derivatives

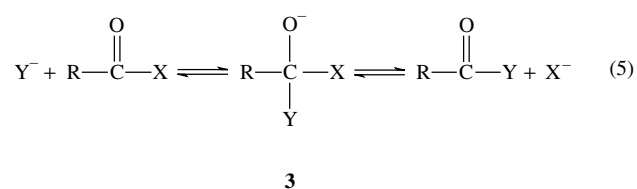
Structure	Chemical class	Comments
$\begin{array}{c} \text{O} \\ \parallel \\ \text{RC}-\text{Cl} \end{array}$	Acid chloride	
$\begin{array}{c} \text{O} \quad \text{O} \\ \parallel \quad \parallel \\ \text{RC}-\text{O}-\text{CR} \end{array}$	Acid anhydride	
$\begin{array}{c} \text{O} \quad \text{O} \\ \parallel \quad \parallel \\ \text{RC}-\text{N}-\text{CR} \end{array}$	Imide	May also be cyclic and may be substituted
$\begin{array}{c} \text{RCH}-\text{CO} \\ \diagdown \quad \diagup \\ (\text{CH}_2)_n-\text{NH} \end{array}$	Lactam	May be substituted
$\begin{array}{c} \text{RCH}-\text{CO} \\ \diagdown \quad \diagup \\ (\text{CH}_2)_n-\text{O} \end{array}$	Lactone	
$\begin{array}{c} \text{O} \\ \parallel \\ \text{RC}-\text{OH} \end{array}$	Carboxylic acid	
$\begin{array}{c} \text{O} \\ \parallel \\ \text{RC}-\text{OR}' \end{array}$	Ester	Aliphatic or aromatic, depending on R'
$\begin{array}{c} \text{O} \\ \parallel \\ \text{RC}-\text{NH}_2 \end{array}$	Amide	Unsubstituted as shown or substituted by replacement of H
$\begin{array}{c} \text{O} \\ \parallel \\ \text{RC}-\text{SR}' \end{array}$	Thiol ester	

oversimplified because it does not incorporate electronic redistribution of the carbon–oxygen π double bond.



The second mechanism is the two-step addition–elimination pathway shown in Eq. 5; it invokes the

formation, by nucleophilic addition to the carboxyl group, of tetrahedral intermediate, **3**.

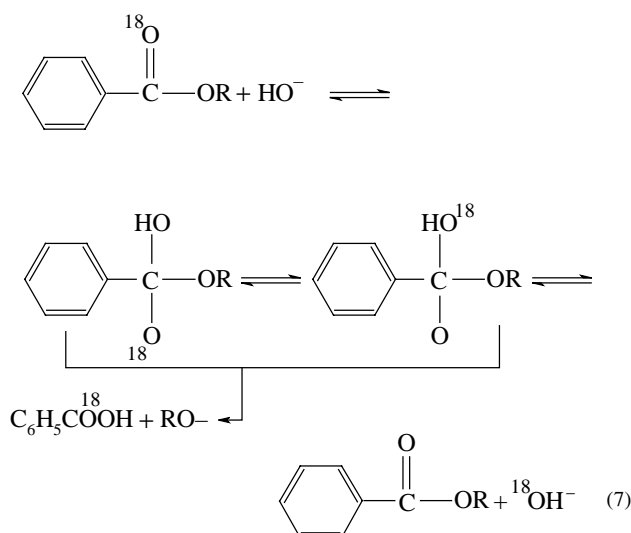
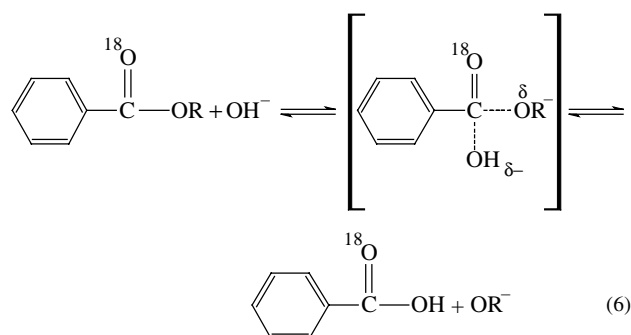


In 1951, Bender provided the first diagnostic test of mechanism for these reactions (3). The alkaline hydrolysis of several benzoate esters, a $\text{B}_{\text{Ac}}2$ reaction, was studied with

Table 2 Classification of ester hydrolysis reactions

Conjugate form	Type of fission	
	Acyl	Alkyl
Acid	A _{Ac} 1	A _{Al} 1 (S _N 1)
	A _{Ac} 2	A _{Al} 2 (unknown)
Base	B _{Ac} 1 (unknown)	B _{Al} 1 (S _N 1)
	B _{Ac} 2	B _{Al} 2 (S _N 2)

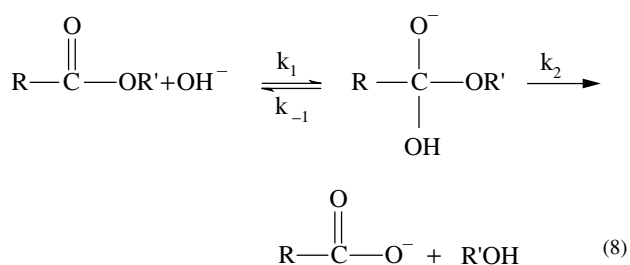
esters enriched in ^{18}O at the carbonyl oxygen. If the S_N2 mechanism in Eq. 6 is operative, no ^{18}O exchange with the solvent during hydrolysis would be expected, because the carbonyl oxygen does not engage in any reversible step. The tetrahedral intermediate mechanism, however, provides a route for ^{18}O exchange, concurrent with hydrolysis, by the symmetrical partitioning shown in Eq. 7.



When ethyl benzoate carbonyl- ^{18}O was subjected to alkaline hydrolysis, samples of ester isolated during the

progress of the reaction were found to contain less ^{18}O than the initial ester. This demonstration of concurrent oxygen exchange and hydrolysis is strong (though not definitive) evidence for the tetrahedral-intermediate mechanism. In other reactions, evidence of different kinds supports the tetrahedral-intermediate mechanism; for example, the intermediate has been detected spectroscopically in certain instances. In some reactions, it has been possible to identify a change in the rate-determining step as the pH is changed; this observation can be explained only by invoking a two-step reaction mechanism.

The two-step addition-elimination is taken here as a mechanistic description of the hydrolysis reaction. Alkaline ester hydrolysis is the example given in Eq. 8.



The differential rate equation for the intermediate (I) is given by Eq. 9.

$$\frac{d[\text{I}]}{dt} = k_1 [\text{OH}^-] [\text{RCOOR}] - (k_{-1} + k_2) [\text{I}] \quad (9)$$

Because the intermediate is consumed as rapidly as it is produced, its concentration is invariant over the time course of the reaction, and the steady state approximation can be applied to give Eq. 10.

$$[\text{I}] = \frac{k_1 [\text{RCOOR}] [\text{OH}^-]}{k_{-1} + k_2} \quad (10)$$

The rate of ester hydrolysis is $v = k_2 [\text{I}]$ or as in Eq. 11.

$$v = \frac{k_1 [\text{RCOOR}] [\text{OH}^-]}{k_{-1}/k_2 + 1} \quad (11)$$

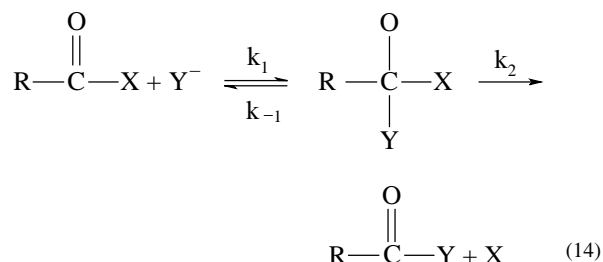
The ratio k_{-1}/k_2 describes the "partitioning" of the tetrahedral intermediate between the reactant and the product states. The experimental rate equation is shown in Eq. 12.

$$v = k [\text{RCOOR}] [\text{OH}^-], \quad (12)$$

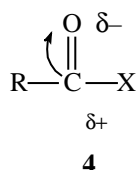
where k is the observed second-order rate constant. It follows that Eq. 13 holds that

$$k = \frac{k_1}{(k_{-1}/k_2) + 1} \quad (13)$$

If $k_2 \gg k_{-1}$, then $k \approx k_1$ and the first step is the rate-determining step (rds) of the reaction; if $k_2 \ll k_{-1}$, then $k \approx k_1 k_2 / k_{-1}$, and the second step (decomposition of the tetrahedral intermediate) is the rds. Eq. 13 provides a basis for the interpretation of structural effects on reactivity. Eq. 8 can be generalized to give Eq. 14

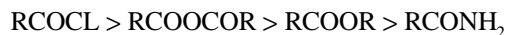


Eq. 13 applies to this kinetic scheme. Changes in the structures of R, X, and Y^- affect k through their separate effects on k_1 and k_{-1}/k_2 (4). The carboxyl group is polarized, as shown in structure 4.

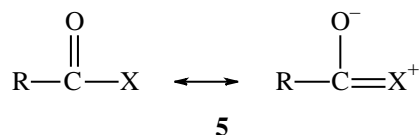


Attack by the nucleophile Y^- , which may be OH^- or H_2O in hydrolyses, therefore, takes place at the electron-deficient carboxyl carbon and is aided by electron withdrawal at R and X. Such effects increase k_1 , and, hence, k . At the same time, these structural effects influence k_{-1}/k_2 . Increased electron withdrawal by R should make it more difficult for both X^- and Y^- to leave the tetrahedral intermediate, resulting in decreases in both k_{-1} and k_2 , with little change in the ratio k_{-1}/k_2 . Structural changes that increase the electron-withdrawing ability of X, however, tend to decrease k_{-1} and increase k_2 , thus decreasing k_{-1}/k_2 . The resultant effect on k depends upon the magnitude of k_{-1}/k_2 relative to unity, as in Eq. 13; in most cases, it appears k is dominated by k_1 .

The previous discussion can be used to interpret the hydrolytic behavior of a series of compounds that possess the same acyl group but varied leaving groups. For example, the order of hydrolytic reactivity for an amide, an ester, an anhydride, and an acid chloride is:

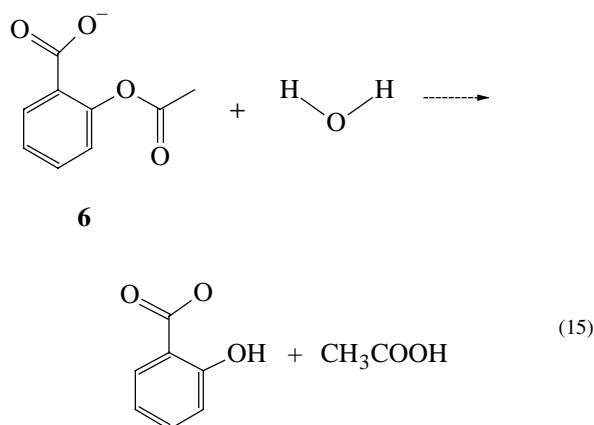


Such a series exhibits great variations in k_1 and k_{-1}/k_2 . The effect on k_1 in this series is a combination of inductive and resonance effects; the latter releases electrons from the leaving group to the acyl group, as indicated in structure 5.



These effects make k_1 smaller for amides than for esters, for example. At the same time k_{-1}/k_2 is affected by the ability of the leaving group (its "nucleofugality"). In the reverse of the k_2 step, that is, an attack by the nucleophile X^- , the more basic X^- , the more effective it is as a nucleophile. Hence, the more basic the leaving group X^- , the less effective it is as a leaving group, and the smaller is k_2 . The order of increasing basicity of the leaving groups in the above series is $\text{Cl}^- < \text{RCO}_2^- < \text{RO}^- < \text{NH}_2^-$, and, therefore, k_2 is smaller for the amide than for the ester and so on. The ratio k_{-1}/k_2 is highest for the amide, and k , the hydrolytic rate constant, is the smallest. Because the basicity of the leaving group can be measured by the pK_a of its conjugate acid, correlations of reactivity (expressed as $\log k$) with the pK_a of the leaving group (for a constant acyl group) are often successful. Thus, aromatic esters (leaving group a phenoxide ion) are more reactive than are aliphatic ester (leaving group an alkoxide ion.)

A very important example of ester hydrolysis of a pharmaceutical is that of the anion of aspirin, 6, given in Eq. 15



The marked instability of aspirin is due to two structural features: one, that it is an aromatic ester and, for reasons

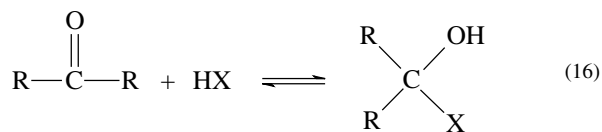
previously discussed, it is, therefore, more labile than an aliphatic ester and two, the ortho relationship of the acetoxy group to the carboxylate. Owing to this proximity, aspirin is subject to intramolecular catalysis of the ester hydrolysis. The pK_a of aspirin is 3.6, and, therefore, it exists predominately in the anionic form above pH 5. This intramolecularly catalyzed reaction accounts for the hydrolytic instability of aspirin in neutral solutions. Additional pharmaceutical examples of hydrolysis of a lactone (a cyclic ester), an amide, and a lactam (a cyclic amide) are warfarin, acetaminophen, and the penicillins, respectively. Equations for these reactions are given in the first edition of this encyclopedia (5).

Phosphoric acid derivatives can be conveniently discussed with the carboxylic acid derivatives, owing to their similar reaction mechanisms. The structural class is extremely important biologically (6), but few pharmaceutical examples within it exist. For discussions of hydrolytic reactions of the phosphoric acid derivatives, the reader is referred to the first edition of this encyclopedia (5).

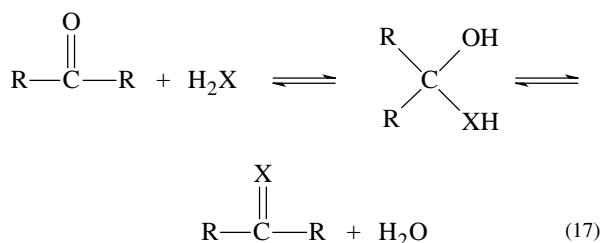
Derivatives of Carbonyl Compounds

The characteristic reaction of an acyl (carboxyl) compound $RCOX$ is cleavage of the $C-X$ bond. The reactions of carbonyl compounds, such as aldehydes $RCHO$ and ketones R_2CO , do not involve cleavage of the $C-R$ or $C-H$ bonds. Many drugs are derivatives of carbonyl compounds, though their parentage may not be obvious.

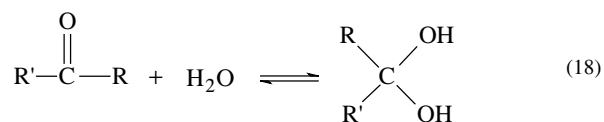
Carbonyl addition reactions take place with bond formation between the carbonyl carbon (which is electrophilic) and the nucleophilic portion of an adding species. The reaction of a reagent with a single dissociable proton can be described by Eq. 16.



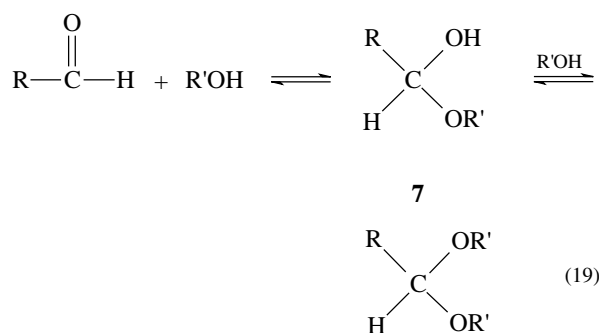
This is a simple addition. If the reagent has two protons, a two-step addition-elimination occurs, as in Eq. 17.



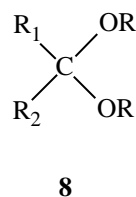
Reaction (17) is a condensation reaction. The reverse of a condensation reaction is a hydrolysis of the $C-X$ bond, the process of interest here. This is shown in Eq. 18, where the product of the reaction is a hydrate.



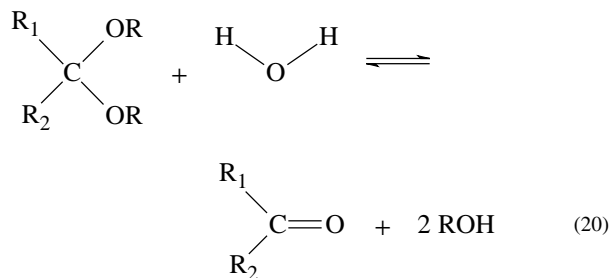
Carbonyl compounds can form hemiacetals, **7**, which in the presence of acids can form acetals, as shown in Eq. 19, and they are subject to hydrolysis.



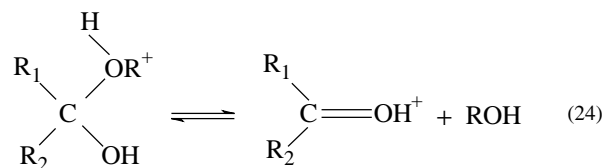
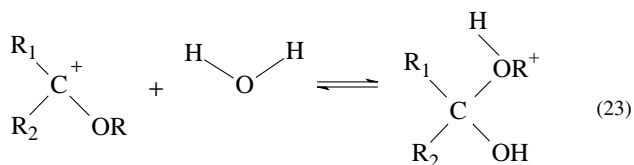
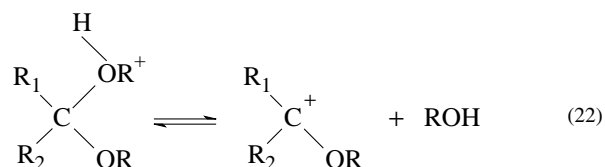
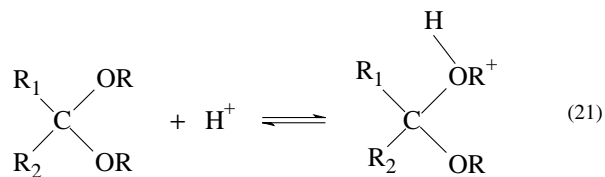
The hydrolysis of acetals (**8**, $R_2 = H$), ketals, and ortho esters (**8**, $R_2 = OR$) has been thoroughly investigated (7).



The overall reaction is given in Eq. 20:

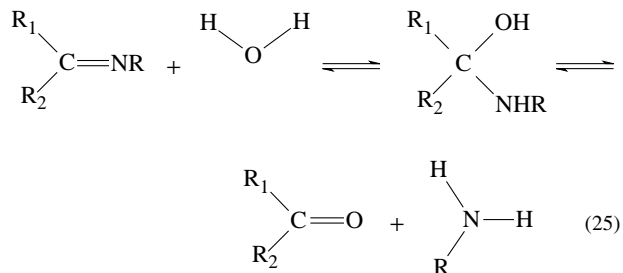


The pathway, a unimolecular reaction of the conjugate acid (or A1 mechanism), is shown in Eqs. 21–24.



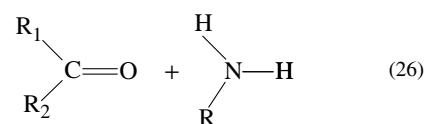
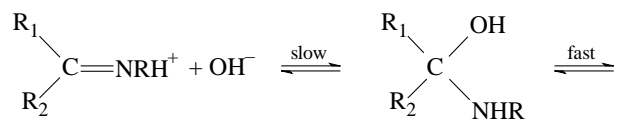
Much of the interest in acetal chemistry is related to carbohydrates, because polyhydroxyaldehydes exist predominantly as cyclic hemiacetals (8). Digoxin, for example, possesses acetal functions subject to hydrolysis.

The hydrolysis of imines is the reverse of the addition of amines to carbonyls, and the two-step scheme of Eq. 25 applies.

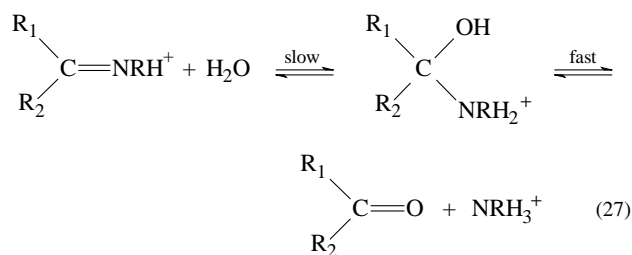


The kinetic behavior of this process depends upon the pH of the medium. In the alkaline region, the addition of

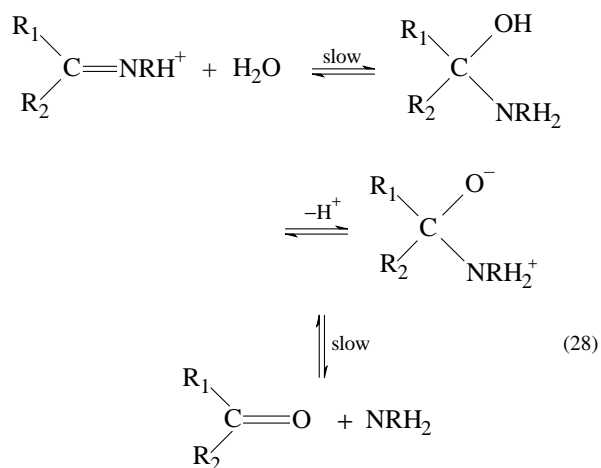
water is rate-determining, and the reaction probably involves attack of the hydroxide ion on the protonated amine (which is kinetically equivalent to attack by water on the neutral imine), as shown in Eq. 26.



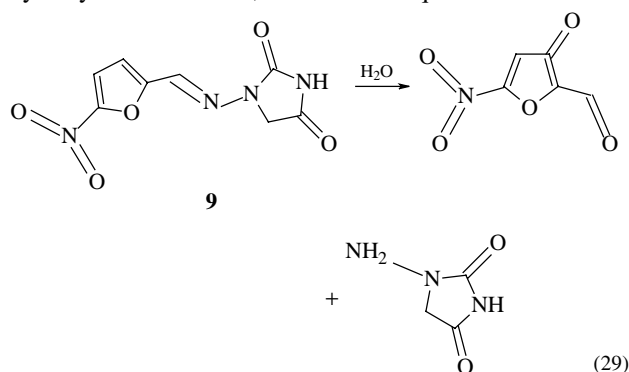
At a somewhat lower pH, as the fraction of protonated imine increases (typical $\text{p}K_a$ values are 6–7) and the concentration of hydroxide decreases, the attack of water on the protonated imine becomes important, as shown in Eq. 27.



The rate becomes briefly pH independent when all of the imine is protonated, but with further decrease in pH, the rate-determining step shifts to loss of amine from the carbinolamine intermediate, as in Eq. 28.



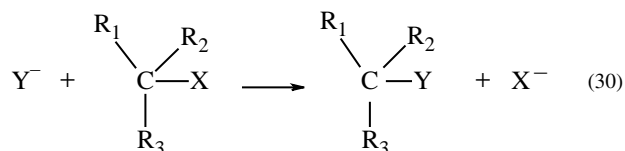
The hydrolysis of nitrofurantoin, **9**, an example of the hydrolysis of an imine, is shown in Eq. 29.



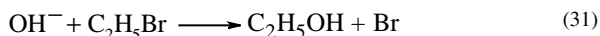
Nucleophilic Substitution

Several classes of nucleophilic substitution reactions can be distinguished (in addition to similar processes, such as acyl transfer to a nucleophile): substitution on aliphatic carbons, on aromatic carbons, and on elements other than carbon, such as metals in metal-ligand coordination complexes. Of these, substitution on aliphatic carbon is most frequently encountered.

The reaction is of the type shown in Eq. 30



where the nucleophile Y^- may be anionic or neutral. In hydrolysis reactions, the nucleophile is H_2O or OH^- ; the hydrolysis of an alkyl halide, given in Eq. 31 is typical:



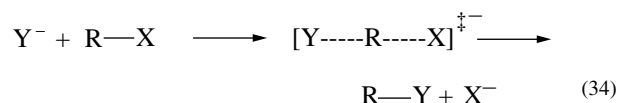
In aliphatic nucleophilic substitutions two mechanistic routes have been clearly identified; one is the two-step process shown in Eqs. 32, 33.



The first step, which is rate-determining, is an ionization to a carbocation intermediate that reacts with the nucleophile in the second step. Because the transition state for the rds includes $R-X$ but not Y^- , the reaction is unimolecular and is labeled S_N1 (substitution nucleophilic

unimolecular). First-order kinetics are observed, with the rate being independent of the nucleophilic identity and concentration.

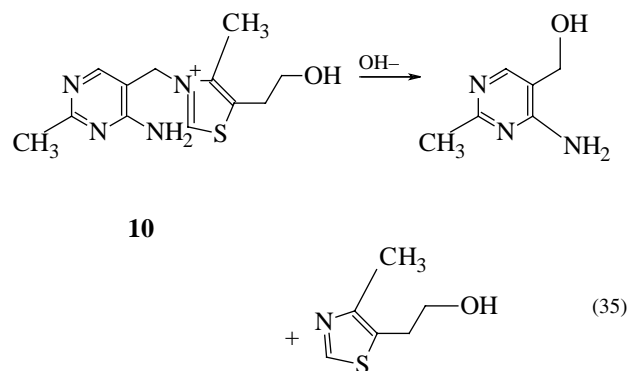
The second mechanism is a one-step direct displacement reaction, shown in Eq. 34.



This bimolecular process is called the S_N2 mechanism. It yields second-order kinetics, unless the nucleophile is the solvent, in which case apparent first-order kinetics are seen. The species in brackets is the transition state.

The S_N reactions have been very carefully studied (9, 10). The S_N1 mechanism is favored by structural features that lead to stabilization of the carbocation, and, therefore, tertiary aliphatic substrates tend to react via the S_N1 route. Primary substrates react by the S_N2 mechanism. Secondary substrates may react by the S_N1 or S_N2 mechanism, depending the details of the reaction; alternatively, secondary substrates may proceed by a "borderline" mechanism having some features of both S_N1 and S_N2 . In solvolyses, the distinguishing characteristic of the S_N2 route is the covalent participation in the rds. The extent of such participation, which is absent in a pure S_N1 reaction, can be estimated by studying solvent effects (11) or the pressure effects (12) on reaction rates.

A pharmaceutical example of a nucleophilic substitution reaction is thiamine hydrochloride, **10**, which is cleaved, as shown in Eq. 35. The cationic moiety can support the electron transfer and is a good leaving group.



HYDROLYTIC KINETIC PHENOMENA

In the experimental study and interpretation of hydrolytic reaction rates, many kinetic effects provide opportunities

for probing the detailed nature of the reaction or have significant practical consequences on drug stability. The topics of temperature dependence, catalysis and solvent effects, were discussed in the first edition of this encyclopedia (5).

REFERENCES

1. Connors, K.A.; Amidon, G.L.; Stella, V.J. *Chemical Stability of Pharmaceuticals*, 2nd Ed.; Wiley-Interscience: New York, 1986; 1–864.
2. Ingold, C.K. *Structure and Mechanism in Organic Chemistry*; Ch. XIV; Cornell University Press: Ithaca New York, 1953; 751–791.
3. Bender, M.L. Oxygen Exchange as Evidence for the Existence of an Intermediate in Ester Hydrolysis. *J. Am. Chem. Soc.* **1951**, *73*, 1626–1629.
4. Bender, M.L. Mechanisms of Catalysis of Nucleophilic Reactions of Carboxylic Acid Derivatives. *Chem. Revs.* **1960**, *60*, 53–113.
5. Connors, K.A. Hydrolysis of Drugs. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Marcel Dekker, Inc.: New York, 1993; 8, 1–29.
6. Bruice, T.C.; Benkovic, S. *Bioorganic Chemistry*; W.A. Benjamin: New York, 1966; 2, 1–786.

7. Cordes, E.H. Mechanism in Catalysis for the Hydrolysis of Acetols, Ketols and Ortho Esters. *Prog. Phys. Org. Chem.* **1967**, *4*, 1–44.
8. Capon, B. Mechanism in Carbohydrate Chemistry. *Chem. Revs.* **1969**, *69*, 407–498.
9. Ingold, C.K. *Structure and Mechanism in Organic Chemistry*; Ch. VII; Cornell University Press: Ithaca, New York, 1953; 306–418.
10. Bunton, C.A. *Nucleophilic Substitution at a Saturated Carbon Atom*; Elsevier: Amsterdam, 1963; 1–172.
11. Bentley, T.W.; Schleyer, P.V.R. Medium Effects on the Rates and Mechanisms of Solvolytic Reactions. *Adv. Phys. Org. Chem.* **1977**, *14*, 1–67.
12. Asano, T.; LeNoble, W.J. Activation and Reaction Volumes in Solution. *Chem Revs.* **1978**, *78*, 407–489.

BIBLIOGRAPHY

- Connors, K.A. *Chemical Kinetics: The Study of Reaction Rates in Solution*; VCH: New York, 1990; 1–480.
- Jencks, W.P. *Catalysis in Chemistry and Enzymology*; McGraw-Hill: New York, 1969; 1–836.
- Maskill, H. *The Physical Basis of Organic Chemistry*; Oxford University Press: London, 1986; 1–490.
- Stewart, R. *The Proton: Applications to Organic Chemistry*; Academic Press: Orlando, FL, 1985; 1–234.

Hydrogels

María Dolores Blanco

Rosa Olmo

José María Teijón

Universidad Complutense de Madrid, Madrid, Spain

INTRODUCTION

Hydrogels are three-dimensional and hydrophilic polymer networks capable of swelling in water or biological fluids and retaining a large amount of fluids in the swollen state.^[1] The water content in the equilibrium of swelling affects different properties of the hydrogels: permeability, mechanical properties, surface properties, and biocompatibility. The utility of hydrogels as biomaterials lies in the similarity of their physical properties with those of living tissues. This resemblance is based on their water content, soft and rubbery consistency, and low interfacial tension with water or biological fluids.^[2]

Hydrogels can be prepared from a wide variety of materials, of natural origin, obtained from plants and animals, as well as from materials prepared by the modification of the aforementioned natural structures and from synthetic polymeric materials. Among the natural polymers, proteins such as collagen,^[3] and polysaccharides such as chitosan^[4] or hyaluronic acid^[5] are used. Regarding synthetic polymers, due to the singular properties that characterize them, as a consequence of their vast structural versatility, they are the materials that have experienced a greater growth and development in terms of their practical applications.^[6]

Hydrogels, which in a dehydrated state have a crystalline aspect and are called xerogels, apart from their immediate definition of being gels containing water, are polymers or copolymers with particular characteristics: They are hydrophilic and insoluble in water, and they swell in its presence (or in the presence of any aqueous fluid), increasing in volume and becoming soft and elastic, but keeping their shape, until reaching a physical–chemical balance. These characteristics are a consequence of diverse factors:

- Their hydrophilic nature is due to the presence of polar groups, such as $-\text{OH}$, $-\text{COOH}$, $-\text{CONH}_2$, $-\text{HSO}_3$ along the polymer chains.
- Their insolubility in water is due to the existence of a three-dimensional cross-linked network in their

structure. This cross-linking can be due to weak cohesive forces (forces such as van der Waals and hydrogen bridges) and due to covalent or ionic bonds.

- Their elastic consistency is determined by the hydrophilic monomer used and the low density of the cross-linking of the polymer.
- The conservation of the shape is a result of the balance between the osmotic forces originated by the water upon entering the polymer and the cohesive forces exerted by the polymeric chains that resist that expansion.^[7]

SYNTHESIS

Hydrogels, based on synthetic monomers, are synthesized through condensation polymerization or through free-radical polymerization. In terms of condensation polymerization,^[8] it is characteristic of compounds that contain reactive functional groups and it is produced between pairs of these functional groups, with the consequent release of small molecules, such as H_2O , NH_3 , or HCl . A clear example is the synthesis of poly(ethylene glycol) (PEG), a polymer soluble in water with numerous biomedical applications^[9] such as wound dressings and drug-delivery systems. Its molecular weight is below 20,000; polymers with higher molecular weights are called poly(ethylene oxide) (PEO). Poly(ethylene glycol) hydrogels are often prepared by radiation cross-linking of high molecular weight PEO^[10] or by chemical cross-linking by reaction of the hydroxyl groups on the ends of PEG. Poly(ethylene glycol) star polymer hydrogels have also been prepared by using γ -irradiation.^[11,12]

Another form of polymeric synthesis is by free-radical polymerization.^[13] These types of polymers are macromolecules formed by the incorporation of unsaturated monomer molecules, in other words, with a double bond, to the active center of a chain in growth. In most cases, these macromolecules are vinyl polymers. In Fig. 1, the chemical structure of some monomers used in the synthesis of hydrogels is shown.

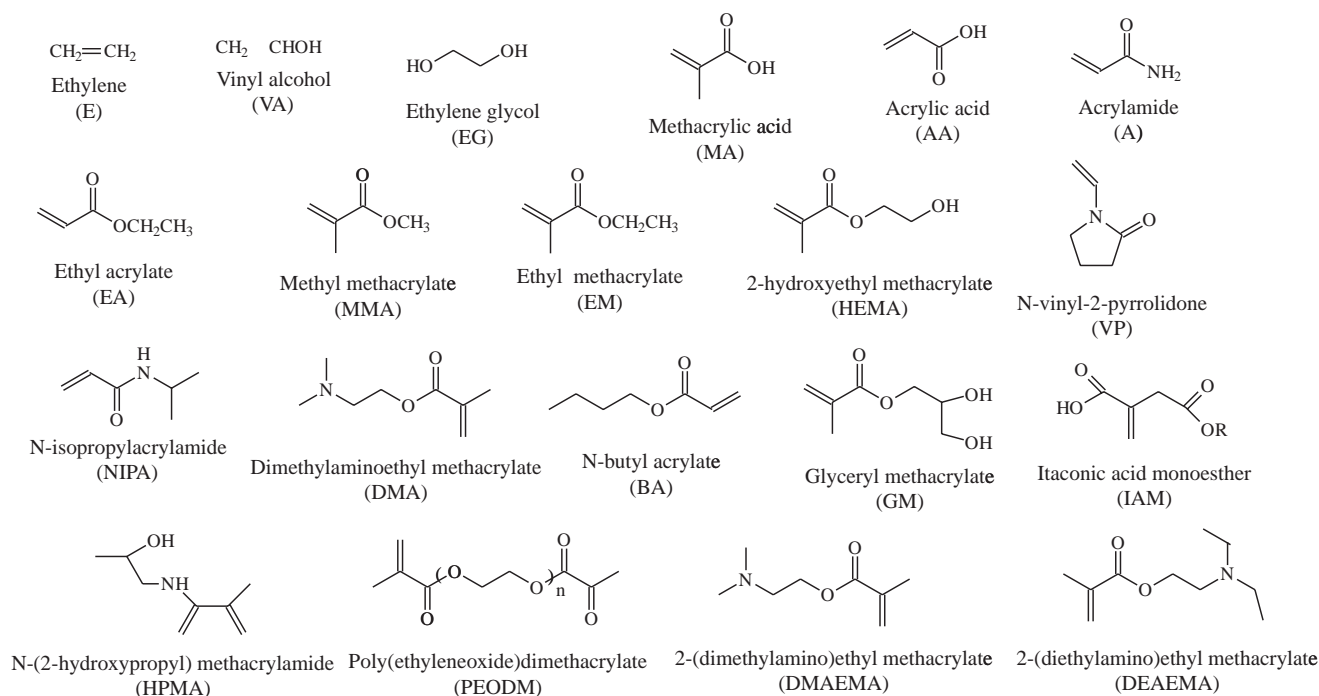


Fig. 1 Chemical structures of some monomers used in the synthesis of hydrogels.

In the formation of addition polymers, a chain reaction is produced with three well-differentiated stages: (a) initiation, (b) propagation, and (c) termination. The beginning of the polymerization requires an activation of the monomers, whether it is thermal, photochemical, by free radicals, or ionic, which allows the binding of the activated monomer in the propagation stage; this way, lineal or branched polymeric chains are produced, depending on the chemical characteristics of the monomer used.

Regarding the initiation process of polymerization, it can be started by γ -radiation. It is a method that has been used for the synthesis of hydrogels of PEO^[12] as well as hydrogels based on vinyl monomers;^[14] in this latter case, azo-compounds such as 2,2-azo-isobutyronitrile (AIBN)^[15] or 2,2'-azobis(2-amidine-propane) dihydrochloride or V-50,^[16] and aqueous salt solutions such as aqueous ammonium peroxodisulfate^[17] are also used. Among the monomers most used in the preparation of hydrogels through free-radical polymerization are 2-hydroxyethyl methacrylate (HEMA)^[18] and *N*-vinyl-2-pyrrolidone (VP).^[19]

On many occasions, the hydrogels are made up of more than one monomer, in most cases by two types of different monomers, in this case being copolymers.^[8] The chemical structure of the copolymers depends on the concentration of the different monomers in the feed mixture and on the reactivity of the functional groups

that participate in the polymerization. In general, the copolymerization allows the mechanical properties and the swelling of the hydrogels to be modified. So, the copolymerization of HEMA with VP generates hydrogels with a higher water content in the swelling equilibrium than those of poly(2-hydroxyethyl methacrylate) PHEMA and whose mechanical properties worsen when the percentage of VP in the copolymer is increased.^[20] However, the copolymers of HEMA and methyl methacrylate (MMA), a monomer less hydrophilic than VP, are characterized by water content in the swelling equilibrium less than PHEMA hydrogels.^[21]

In the synthesis of a large number of hydrogels, cross-linking agents are used, which interconnect the lineal polymeric chains establishing a three-dimensional network of strong chemical bonds among them.^[18,19] It is necessary that the polymer has certain groups in its structure that can be used as anchor points in order to form the network. The cross-linking agent, whose choice depends on the selected monomers, must have at least two reactive groups in its structure, in order to be able to cross-link different polymeric chains, normally tetrafunctional and hexafunctional compounds, such as ethylene glycol dimethacrylate (EGDMA)^[22] and 1,1,1-trimethylolpropane trimethacrylate (TPT),^[23] although other cross-linking agents have also been used such as ethylenediaminetetraacetic dianhydride (EDTAD)^[24] and pentaerythritol triacrylate

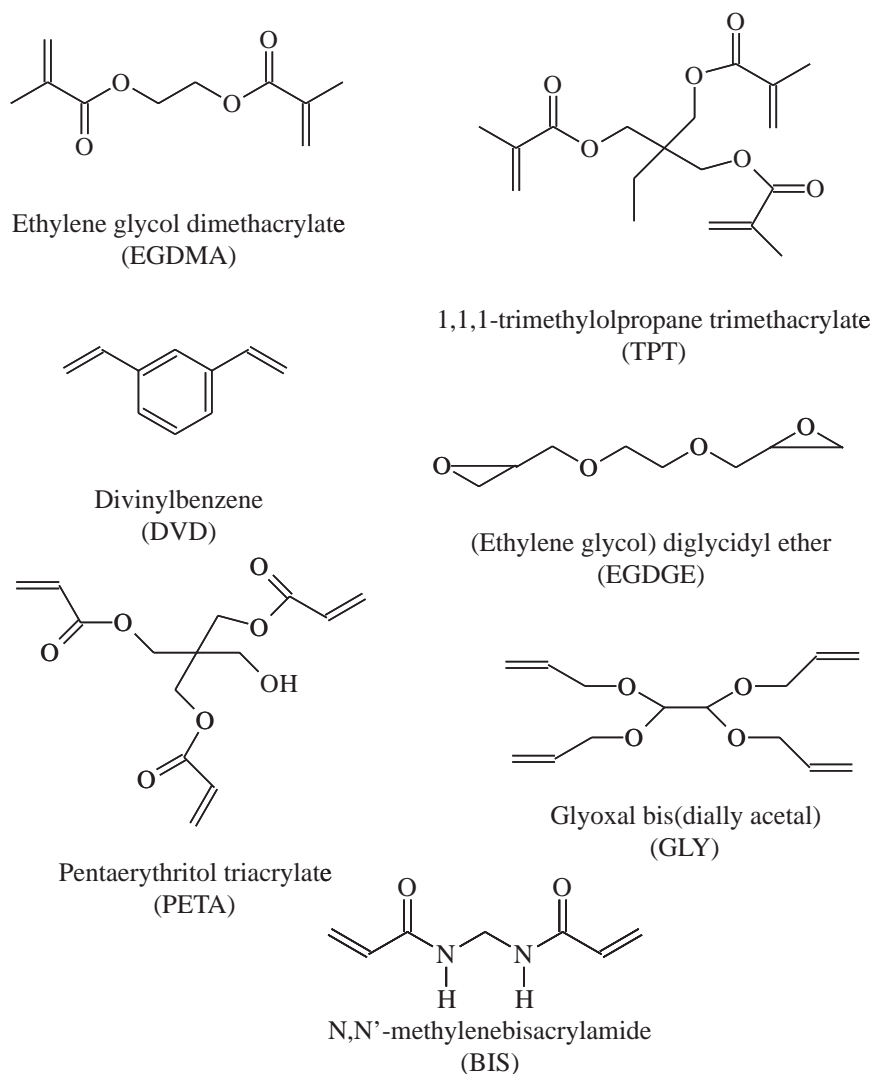


Fig. 2 Chemical structures of some cross-linking agents used in the synthesis of hydrogels.

(PETA)^[25] (Fig. 2). The cross-linked polymers exhibit considerable differences in their properties depending on the degree of cross-linking and the preparation method. In general, the degree of cross-linking determines the degree of swelling, the pore size, the total surface area, and the mechanical resistance of the polymer.

The large deformation capacity of the elastomeric materials is due to the flexible nature of the polymeric chains that form them. This molecular flexibility (microscopic) allows its large macroscopic deformation capacity. If the chains that make up the elastomeric material are connected among themselves through covalent links, which connect them to each other, then the solid becomes a three-dimensional molecular network, the deformation of the polymeric matrix is only temporary

(while the application of deforming force lasts), and the recovery is reversible. These bonds or cross-links give continuance to the material, preventing the sliding of some chains with respect to others. The elastic properties of the material depend on the proportion of cross-links that are introduced in order to form the three-dimensional network (Fig. 3).

The *cross-linking density* is the number of points of cross-links present per unit volume of material. If the cross-linking density is low, the network will be very open and therefore, the chain sections that join two consecutive cross-links are long and their flexibility is hardly limited, causing the material to exhibit elastic behavior. If the cross-linking density is high, the network will be closely woven. In this case, the chain sections between cross-links

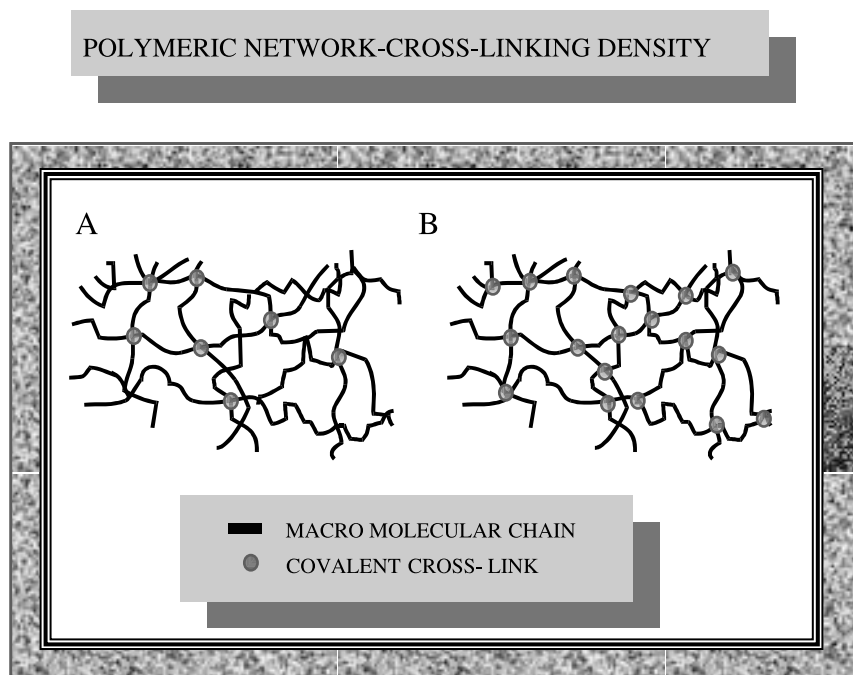


Fig. 3 Cross-linking density of a polymeric network: (A) low cross-linking density; (B) high cross-linking density.

are short and the chains are anchored by many points, causing the material to lose flexibility, and making it more rigid or stiffer.

However, the preparation of nonbond cross-linked polymer hydrogels of VP by copolymerization with the cross-linking agent cyclic octaethylene glycol fumarate, a 29-membered ring, having a polymerizable double bond has also been described. Treading of polymer chains into the macrocyclic rings is the cause for the cross-linking.^[26]

Regardless of the polymerization reaction that leads to the synthesis of a hydrogel, and that is basically conditioned by the type of monomer or monomers that make it up, there are several types of polymerization processes for the preparation of hydrogels: bulk polymerization, suspension polymerization, and emulsion polymerization. In the bulk polymerization, the system is made up of the monomers, the cross-linking agent, if a cross-linked gel is desired, and the initiator of the reaction, which are situated in a recipient, where the polymerization reaction is unleashed; this technique allows one to obtain hydrogels with a wide variety of morphologies in terms of the mold used for the synthesis, the most common being the preparation of cylinders^[15,16] and films.^[27] Some variants of this type of polymerization processes allow one to obtain macroporous

hydrogels.^[28] Macroporosity can be introduced into hydrophilic copolymer membranes by means of porosity technology; with this approach, monomers are copolymerized around a crystalline matrix, which is either dispersed or leached from the membrane after polymerization. The preparation of macroporous membranes by the freeze–thaw method requires the presence of a chemically inert material, which can form a crystalline structure around which the monomers can be polymerized; in the case of hydrophilic monomers, aqueous systems can be used, where the ice crystals constitute the crystalline matrices. Applying these technologies, Professor Tighe's group has prepared macroporous membranes of poly(HEMA) or poly(acrylic acid).^[28] Other researchers have prepared porous networks of poly(HEMA) using a thermally initiated free-radical solution polymerization.^[29]

In suspension polymerization, the drops of monomer, containing the initiator, are dispersed in a concentrated aqueous saline solution, which produces hydrogel beads. Applying this technique, Professor Ping I. Lee^[30] has synthesized PHEMA hydrogels. Hydrogels can also be prepared through emulsion polymerization,^[1,8] this way, emulsification methods can be used to obtain spherically shaped hydrogels from cross-linked polymers or copolymers.^[31]

PROPERTIES

There is a direct relationship between the properties of a hydrogel and its structure, in such a way that both characteristics cannot be considered in an isolated way because its polymeric composition and the synthesis method decisively influence the structure of the polymeric matrix as well as the final properties that the gel will have (Fig. 4). Therefore, when the properties of the hydrogels are presented, reference must be made to the structural parameters that condition them.

Swelling

Hydrogels are characterized, first by their capacity to absorb water or aqueous solutions. The water content in the swelling equilibrium of a hydrogel is affected, fundamentally, by the nature of the monomer or monomers that make it up, by the type and density of the cross-link, and by other factors such as temperature, ionic strength, and pH of the hydration medium.

The amount of the aqueous medium incorporated in a hydrogel is determined, generally, gravimetrically and can be expressed in several ways. The fractional hydration

$(W)^{[14]}$ can be obtained as follows:

$$W = (w_1 - w_0)/w_1$$

where w_1 and w_0 are the weights of swollen and dry gels, respectively.

The swelling ratio $(r)^{[16]}$ is expressed as:

$$r = (w_1 - w_0)/w_0$$

At a particular temperature, the volume fraction of polymer (ϕ_2) within a hydrogel at swelling equilibrium is given by:

$$\phi_2 = (D_0/D)^3$$

in which D_0 and D are the diameters of dry and swollen equilibrium discs, respectively, which can be determined using a micrometer or photographically.^[32]

Hydrogels formed from hydrophilic monomers contain larger percentages of water. So, for example, the content of water in the equilibrium of HEMA and VP copolymers increases when the content of VP in the gel is increased, obtaining percentages of fractional hydration of 38.3 wt% for hydrogels with a 25 wt% of VP in the feed mixture of polymerization and of 59.3 wt% for

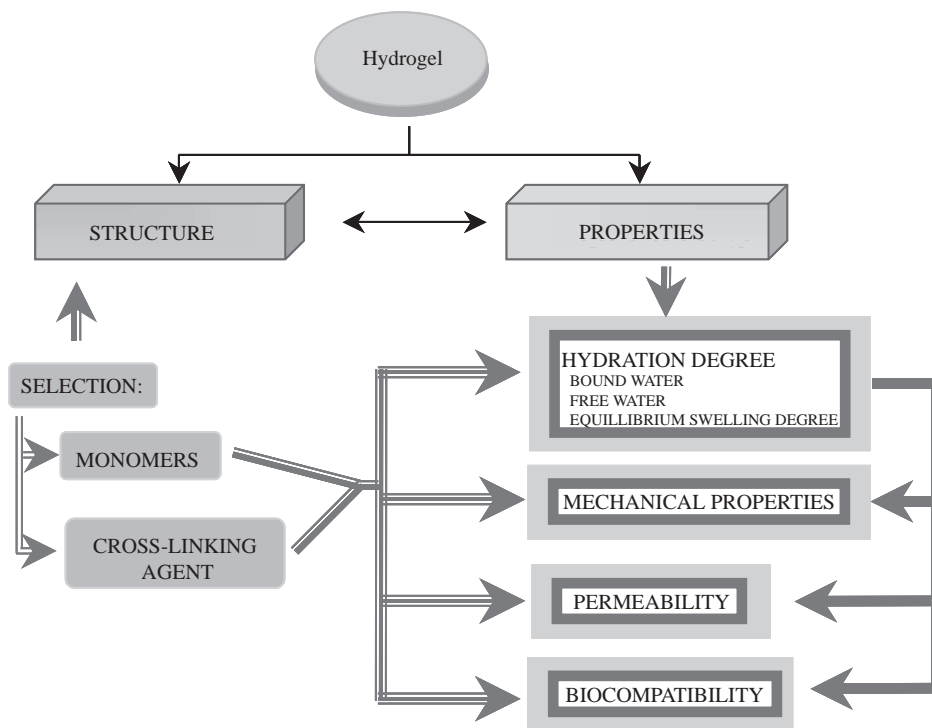


Fig. 4 Relationship between structure and properties of hydrogels. The selection of monomers and cross-linking agent determines hydrogel properties. On the other hand, a hydrogel with specific properties must be synthesized using specific monomers and cross-linking agent.

those that contain 50 wt% of VP, both cross-linked with 5 wt% of EGDMA.^[33]

Another variable that directly influences the capacity of the water absorption of hydrogels is the degree of their cross-linking. The increase in the content of cross-linking agent decreases the percentage of fractional swelling. So, for example, for PHEMA hydrogels, these values vary from 36 wt% to 29 wt% when the amount of the cross-linking agent (EGDMA) is increased from 0.5 wt% to 5 wt%.^[22] This same effect has been described for numerous hydrogels, for example, copolymers of HEMA and acrylamide (A) cross-linked with EGDMA,^[34] and copolymers of *N*-isopropylacrylamide (NIPA) and acrylic acid (AA) cross-linked with *N,N'*-methylene bisacrylamide (BIS) or with glyoxal bis(diallyl acetal) (GLY).^[35]

Likewise, the type of cross-linking agent also affects the water content in swelling equilibrium, so, for example, PHEMA hydrogels cross-linked with TPT (1 wt% of TPT) reach percentages of fractional swelling which are lower ($W = 27\text{--}23$ wt%) than those cross-linked with EGDMA, in the same conditions of synthesis.^[23] Also, copolymers of NIPA and AA experience greater swelling when they are cross-linked with GLY than with BIS.^[35]

There are hydrogels whose swelling is conditioned by the temperature and by the pH of the medium. Poly(*N*-isopropylacrylamide) (PNIPA) and its copolymers have been extensively studied in the development of thermoresponsive hydrogels. Cross-linked PNIPA exhibits a volume phase transition temperature (VPTT), at around 32°C in aqueous media. At temperatures below its VPTT, the cross-linked PNIPA matrix is superabsorbent and swells to form a hydrogel. On increasing the temperature above the VPTT, the hydrogel undergoes a rapid shrinkage. The thermosensitivity of the hydrogel is due to the hydrophilic–hydrophobic balance of its constituent polymer chains and is directly related to the lower critical solution temperature (LCST) phenomenon that is exhibited by linear PNIPA in aqueous solution.^[36] Furthermore, diverse copolymeric hydrogels of PNIPA have been synthesized, which are temperature sensitive.^[37,38] In addition, temperature-sensitive hydrophilic gel microcapsules of poly(*L*-lysineisopropylamide-terephthalic acid), containing water, have been obtained.^[39] The microcapsule changes its size between 33 and 35°C. Below 33°C the microcapsules are fully spherical while they are aggregated above 35°C; this thermosensitive morphological change is reversible. Another type of temperature-sensitive hydrogels has been prepared by the group of Yoshida and Carenza, which has synthesized poly(acryloyl-*L*-proline methyl ester)-based hydrogels cross-linked with TPT,^[40,41] which swell between 74% and 63%

(swollen hydrogel volume/dry hydrogel volume, %) at 4°C, and they shrink at 37°C.

The pH sensitivity of the hydrogels is due to the presence of weakly acidic and/or basic functional groups on the backbone. Their water uptake properties are provided by the ionization of the functional groups, which depends on the pH and on the ionic strength of the external medium. The Falamarzian group has synthesized diverse hydrophobic/polyelectrolyte copolymer hydrogels. They have reported that poly(methyl methacrylate/dimethylaminoethyl methacrylate) [MMA/DMA] cross-linked with divinylbenzene (DVB) hydrogels are polycationic pH-sensitive hydrogels, which are quite hydrated in acidic pHs while glassy in alkaline and neutral pHs.^[42,43] Also, chemically derived cross-linked polyacrylamide-grafted guar gum microgels possessing weakly anionic groups have been developed;^[31] due to the presence of ionizable carboxylic functional groups, these microgels are sensitive to pH and ionic strength of the external media. The rate of solvent diffusion increases considerably at higher pH, i.e., above the pK_a ($pK_a = 4.6$) of the anionic microgels. At higher pH, above the pK_a , the $-\text{COOH}$ groups may dissociate, thereby increasing the osmotic pressure inside the microgels resulting in a higher swelling. The electrostatic repulsion might cause the cross-linked network to swell, thereby allowing water molecules to enter into the matrix. Another type of pH-sensitive hydrogel is made up of copolymers of *N*-acryloyl-*N'*-methyl piperazine and MMA cross-linked with EGDMA.^[44] The high swelling observed at low pH ($\text{pH} = 2.6$) is due to the protonation of the tertiary amine group of the piperazine unit, which causes electrostatic repulsion thus leading to increased swelling. The swelling of the gel goes down in a solution of pH 7. Likewise, the hydrogels synthesized by cross-linking chitosan and polyvinyl pyrrolidone (PVP) blends with glutaraldehyde to form a semi-interpenetrating polymer network are pH sensitive;^[45] an acidic environment ($\text{pH} 1$ and 2) has a pronounced effect on the swelling profiles of hydrogels compared with neutral or alkaline media. Increased swelling of hydrogels, under acidic conditions, is due to the protonation of a primary amino group on chitosan.

Mechanical Properties

The mechanical behavior of the hydrogels can be described by the theories of rubber elasticity and viscoelasticity,^[46] which are based on time-independent and time-dependent recovery of the chain orientation and structure, respectively. Mechanical properties due to rubber elastic behavior of hydrogels can be determined

by tensile measurements, while the viscoelastic behavior can be determined through dynamic mechanical analysis.

The Huglin's group has determined the mechanical properties of diverse hydrogels through compression–strain measurements.^[32,35,47,48] Young's moduli of elasticity (E) can be obtained as the slopes in plots of stress (τ) vs. strain ($\lambda - 1$), where τ is the applied force per unit area of hydrogel and λ is the ratio of deformed length to undeformed length of hydrogel. These plots are linear over the range of strain covered.

The compression moduli (G) are obtained as the slope of τ vs. $(\lambda - \lambda^{-2})$ plot, at low strains: $\tau = G(\lambda - \lambda^{-2})$. Once the compression modulus is known, the effective cross-linked density (ν_e) can be determined from the following equation:

$$G = RT \nu_e \phi_2^{-1/3}$$

The effective cross-linked density is an experimental quantity, unlike the theoretical cross-linking density (ν_l), which is given by

$$\nu_l = Cf/2$$

where C (mol dm^{-3}) is the concentration of added cross-linking agent and f is its functionality.

Starting from the values of ν_e , the molar mass between cross-links (M_c) can be calculated:

$$M_c = \rho / \nu_e$$

where ρ is the density of the hydrogel, which can be obtained from the mass and the volume. The molar mass between cross-links can also be determined through the Peppas–Merrill relationship.^[49]

Likewise, starting from the values of ν_e , the polymer–water interaction parameter (χ) can be calculated using the following expression:

$$\ln(1 - \phi_2) + \phi_2 + \chi \phi_2^2 + \nu_e V_1 (\phi_2^{1/3} - 2\phi_2 f^{-1}) = 0$$

where V_1 is the molar volume of water ($\text{dm}^3 \text{mol}^{-1}$) at a specific temperature.

The mechanical properties of the hydrogels depend, fundamentally, on their composition and structure, and on many occasions they are weak, for which many of these materials cannot be used in load-bearing applications such as the replacement of damaged tissues. Although the quantity of water absorbed in the polymeric matrix of a hydrogel plays a very important role in its diffusion and permeability properties as well as its biocompatibility, it can negatively affect its mechanical properties in many cases. Furthermore, the state of the water in the hydrogel also affects its mechanical properties, because the increase

in the proportion of freezing water in the gel causes a worsening in its mechanical resistance, since that water is going to act as a plasticizer, decreasing the glass transition temperature of the system.^[21] The strategies to improve the mechanical properties of the hydrogels are, mainly, the cross-linking and the copolymerization with hydrophobic monomers.

The highly hydrophilic hydrogels, such as those consisting of PVP cross-linked with EGDMA (0.5 wt%–5 wt%),^[47] characterized by small values of the interaction parameter χ ($\chi = 0.49$ – 0.57), present small Young's modulus (0.019 MPa–0.504 MPa) that improve when the percentage of the cross-linking agent is increased. The copolymerization of VP with a hydrophobic monomer such as *n*-butyl acrylate (BA)^[48] notably improves the mechanical properties of hydrogels, whose compression modulus (G) increases when the amount of BA in the hydrogel is increased, from 0.046 MPa for hydrogels with 2 wt% of BA to 0.256 MPa for those with 60 wt% of BA, all of them cross-linked with 1 wt% of TPT, which corresponds to a decrease in their elasticity ($E = 0.148$ MPa– 0.0829 MPa). Also in these hydrogels, the increase in the cross-linking agent decreases elasticity, increasing the compression modulus ($G = 0.439$ MPa for hydrogels with 60 wt% BA and 2.5 wt% TPT), which is correlated with a more effective cross-linking density ($\nu_e = 0.115 \text{ mol dm}^{-3}$ – $0.195 \text{ mol dm}^{-3}$ for hydrogels with 60 wt% BA and 1 wt%–2.5 wt% TPT, respectively) and therefore will lower the values of molar mass between cross-links ($M_c = 9700 \text{ g mol}^{-1}$ – 5700 g mol^{-1} for hydrogels with 60 wt% BA and 1 wt%–2.5 wt% TPT, respectively).

The type of cross-linking agent also has an important effect on the mechanical properties of the hydrogels. Thus, copolymeric hydrogels of NIPA and AA^[35] have better mechanical properties when they are cross-linked with BIS, than with glyoxal bis(diallyl acetal) (GLY). Therefore, for hydrogels with 10 mol% of AA, the compression modulus is 6.68 kPa when they are cross-linked with 2.5 mol% of BIS and 1.44 kPa when it is done with the same percentage of GLY. The compensation is that hydrogels cross-linked with GLY exhibit much higher swelling ratios.

The mechanical resistance of the hydrogels is, therefore, very variable, existing mechanically weak hydrogels such as alginate hydrogels ionically cross-linked with CaCO_3 in the presence of D-glucono- δ -lactone, whose compression modulus reaches a value of 175 kPa.^[50] However, others are significantly mechanically resistant, such as those formed of PVP / β -chitosan in the presence of glutaraldehyde^[51] or the semi-interpenetrating hydrogels based on poly(3-hydroxybutyrate) and net-PEG,^[52] whose

tensile strength is 43.13 MPa and 350.6 MPa, in the dry state, respectively. Hydrogels of poly(vinyl alcohol) (PVA) called SalubriaTM^[53] have compressive modulus between 1 MPa and 18 MPa, and shear tangent modulus between 0.1 MPa and 0.4 MPa, which is within the range of the modulus of articular cartilage. Values of network parameters of different hydrogels are shown in Table 1.

BIOMEDICAL APPLICATIONS OF HYDROGELS

Hydrogels have numerous applications in the field of biomedicine because, as it has been previously indicated, their characteristics can be modulated in terms of their composition and preparation method. This way, matrices with elevated swelling levels can be obtained. In addition, they present mechanical properties which are suitable for a specific application, which along with their characteristics of being made up of biostable or biodegradable devices, it makes these elements very versatile and with very diverse applications, some of which are shown in the following paragraphs.

Contact Lenses and Ocular Implants

Soft contact lenses are one of the most widely used applications of hydrogels. One of the main characteristics of these lenses is their comfort because they perfectly adapt to the global ocular curvature, and furthermore they allow atmospheric oxygen to reach the cornea by dissolving in the water of the lens and then transported by diffusion, until reaching the permeability limits close to those of a hypothetical lens of distilled water. Studies carried out by the Tighe group^[21] show that there is a correlation between the equilibrium water content of a hydrogel and its permeability to oxygen; thus the copolymeric hydrogels of HEMA and VP with equilibrium water content of 80% have a high permeability to oxygen, somewhat higher than copolymeric hydrogels of HEMA–VP–styrene, whose equilibrium water content is around 70%.

The first contact lenses were developed at the beginning of the 1970s by Wichterle as a consequence of his work with hydrogels based on glycolmethacrylates.^[54] The first soft contact lenses marketed with real success were those made up of PHEMA, whose water content was 38%. Several companies subsequently developed a range of

Table 1 Network parameters of some hydrogels. ϕ_2 : volume fraction of polymer; ν_t : theoretical cross-linking density; ν_e : effective cross-linking density; E : Young's modulus; G : compression modulus; M_c : molar mass between cross-links; χ : polymer–water interaction parameter

Hydrogel	ϕ_2	ν_t (mol dm ⁻³)	ν_e (mol dm ⁻³)	E (kPa)	G (kPa)	M_c (g mol ⁻¹)	χ	Ref.
PVP–EGDMA 0.5%	0.039	62×10^{-3}	5.7×10^{-3}	19		216,000	0.491	[47]
PVP–EGDMA 5%	0.227	616×10^{-3}	114×10^{-3}	504		10,700	0.572	
VP/BA (98/2)–TPT 1%	0.100	94.3×10^{-3}	40.4×10^{-3}	148	46	30,200	0.507	[48]
VP/BA (40/60)–TPT 1%	0.758	86.2×10^{-3}	114.8×10^{-3}	82.9	256	9,700	1.148	
VP/BA (40/60)–TPT 2.5%	0.782	215.6×10^{-3}	194.8×10^{-3}	1399	439	5,700	1.208	
NIPA/AA (90/10)–BIS 2.5%	0.15	0.508	18.4×10^{-3}	22	6.68	60.4	0.57	[35]
NIPA/AA (90/10)–GLY 2.5%	0.18	1.05	5.71×10^{-3}	4.54	1.44	204	0.57	
LH alginate(4.5 w/v)– 1.5 × CaCO ₃					175			[50]
PVA (75% water)					1,000–18,000			[53]
PEO (1.5 g dL ⁻¹)–0.47 kGy hr ⁻¹						12,000		[12]
PEO (1.5 g dL ⁻¹)–6.1 kGy hr ⁻¹						10,200		
PEO (2 g dL ⁻¹)–0.47 kGy hr ⁻¹						9,800		
PEO (2 g dL ⁻¹)–6.1 kGy hr ⁻¹						6,100		
PVA–EDTAD 5%			5.5×10^{-3}	3				[27]
PVA–EDTAD 25%			202×10^{-3}	199				
PEGDA:PETA (10:0)						62.4		[25]
PEGDA:PETA (10:2)						39.8		

VP: *N*-vinyl-2-pyrrolidone; PVP: polyVP; EGDMA: ethylene glycol dimethacrylate; BA: butyl acrylate; TPT: 1,1,1-trimethylolpropane trimethacrylate; NIPA: *N*-isopropyl acrylamide; AA: acrylic acid; BIS: *N,N'*-methylene bisacrylamide; PVA: polyvinyl alcohol; EDTAD: ethylenediaminetetraacetic dianhydride; PEGDA: polyethyleneglycol diacrylate; PETA: pentaerythritol triacrylate.

hydrogel lens materials containing various monomers such as VP, methacrylic acid (MAA), MMA, and glyceryl methacrylate (GM), among others,^[55,56] in the attempt of obtaining materials with an elevated water content and suitable mechanical properties, which allow them to resist the force of the eye lid, along with an elevated permeability to oxygen (Table 2). The permeability of oxygen is a very important factor when considering that hydrogel lenses can induce limbal hyperemia, which is a consequence of a localized hypoxia.^[57] It is considered that the oxygen transmissibility should be greater than $125 \times 10^{-9} \text{ Dk L}^{-1}$ in order to prevent stromal anoxia in the closed eye.^[58] This is achieved with hydrogel materials of high oxygen permeability such as silicone hydrogel lenses with values of Dk of 140 (Lotrafilon A—CIBA Vision).^[57]

The substitution of the crystalline lens for a transparent intraocular lens (IOL) has become an integral part of almost all cataract operations. Because the implantation of the IOL is done through a small incision, foldable lenses have been developed, some of which are PHEMA hydrogels (refractive index, 1.44; water content, 38%) or HEMA copolymers with MMA (refractive index, 1.47; water content, 20%) or with 6-hydroxyhexyl methacrylate (refractive index, 1.47; water content, 18%),^[59] because there is evidence that hydrophilic materials are less damaging to the corneal endothelium and produce lower levels of inflammatory response.

The vitreous humor of the eye is a hydrophilic gel with water content above 99%, formed by a network of collagen and hyaluronic acid, which in some pathological situations

becomes dysfunctional. For this reason, some hydrogels have been tested as substitutes for the vitreous humor of the eye, among them hydrogels of PVP, polyacrylamide (PA) and PVA, although no truly satisfactory result has been achieved.^[60]

Tissue Prosthesis and Tissue Regeneration

Among the applications of hydrogels as materials for tissue replacement, its possible use for replacing articular cartilage stands out. The importance of this use is obvious if we consider the millions of people who suffer from degenerative joint diseases, which make it necessary to have effective treatments available to repair cartilage damaged by pathological processes or from injuries. Among the possible treatments would be the replacement of the cartilage by arthroscopy, using a material that is similar to natural cartilage. This natural cartilage consists of a matrix of proteoglycans, able to retain water molecules due to the negative charges of the glycosaminoglycans present in their structure. These water molecules are eliminated from the matrix when it is exposed to pressure, this way cushioning the force exerted on it. For this reason, it can be considered that, from a practical point of view, the cartilage functions biomechanically as a specialized, physicochemically active biological hydrogel,^[61] and hydrogels can be used as model systems for quantitatively studying the influence of the intrinsic swelling parameter on the consolidation behavior, and thus obtain information on the fundamental parameters that control the biomechanical properties of the articular cartilage.

Table 2 Selected soft contact lens materials

Trade Name	Principal components	Water content (%)	Oxygen permeability (Dk)	Refractive index	Manufacturer
Optima FW	Polymacon (HEMA)	38.6	8.4×10^{-11}	1.43	Bausch & Lomb
Soflens66	Alphafilcon A (HEMA/VP/TBCM)	66	32×10^{-11}	1.39	Bausch & Lomb
Soflens	Hilafilcon A (HEMA/VP)	70	33×10^{-11}	1.38	Bausch & Lomb
Purevision	Balafilcon A (SVC/VP/VA)	36	99×10^{-11}	1.426	Bausch & Lomb
Focus	Nelfilcon A (PVA/NFMA)	69	26×10^{-11}	1.38	Ciba Vision
Cibasoft	Tefilcon (HEMA)	37.5	8.9×10^{-11}	1.43	Ciba Vision
Focus Monthly Visitint	Vifilcon A (HEMA/EDMA/MAA)	55	16.0×10^{-11}	1.415	Ciba Vision
Focus Night and Day	Lotrafilcon A [fluoro-silicone hydrogel]	24	140×10^{-11}	—	Ciba Vision
Acuvue	Etafilcon A (HEMA/MAA)	58	28×10^{-11}	1.40	Vistakon
Horizon 55EW	Methafilcon A (HEMA)	55	18.8×10^{-11}	1.41	Westcon Contact Lens
Proclear	Omafilcon (HEMA/MCOE-PC)	59	—	—	Biocompatibles International

HEMA: 2-hydroxyethyl methacrylate; VP: *N*-vinyl-2-pyrrolidone; TBCM: 4-*t*-butyl-2-hydroxycyclohexyl methacrylate; SVC: silicone vinyl carbamate; VA: vinyl alanine; NFMA: *N*-formylmethyl acrylamide; EDMA: ethylene dimethacrylate; MAA: methacrylic acid; MCOE-PC: 2-methacryloyloxyethylphosphorylcholine.

Regarding what was previously presented, it is understood that there is a clear interest in finding a material, with hydrogel characteristics, that can substitute polyethylene (PE), which is used in partial or total replacement of articular cartilage. For this reason, researchers at the University of Kyoto have studied hydrogels of PVA with mechanical characteristics and suitable lubrication because these hydrogels have a thicker fluid film under pressure than PE and also their wear factor is approximately five times higher than that of PE. In the *in vivo* studies, after maintaining the implants up to 52 weeks, no tissue degeneration was seen in the surrounding tissue^[62] and the biocompatibility with respect to the implants of ultrahigh molecular weight polyethylene (UHMWPE) seems to improve.^[63,64] Other researchers have come close to preparing a material that substitutes this tissue through the preparation of semi-interpenetrating hydrogels, with mechanical properties that are similar to those of cartilage.^[65] The mechanical characteristics of friction and wear of hydrogels of PHEMA have also been studied as possible substitutes of articular cartilage.^[66] However, the most promising hydrogels in this application seem to be those of PVA, prepared by freeze–thaw techniques,^[67] one of which is called Salubria[™]^[53] and its water content and mechanical characteristics are notably similar to those of articular cartilage.

Other studies are centered on the development of hydrogels that include helicoidal fibers of poly(ethylene terephthalate), in an attempt to simulate natural collagen fibers, in order to generate a material that has mechanical characteristics close to those of the ligaments and can be used as connective replacement tissue.^[68]

Hydrogels are also used as breast prostheses, due to their swelling characteristics, which make them similar to live tissue. Medical follow-ups of implanted women indicate that these implants remain soft after 18–48 months.^[69] However, all the results are not equally concordant because breast prostheses with textured surfaces and filled with PVP hydrogel undergo a significant increase in volume when implanted in patients as a consequence of osmosis.^[70]

There is high interest in the use of hydrogels in the regeneration of the central nervous system (CNS). Therefore, chemically cross-linked PHEMA tubes have been synthesized by applying centrifugal forces, whose outer diameter is 2.4 mm and wall thickness is 40 μm –400 μm , which could be used for guided regeneration in the nervous system.^[71] On the other hand, the Woerly group has studied hydrogels as macromolecular networks that may act as extracellular matrices to induce tissue growth and organization during the healing process, thus promoting tissue restructuring and axonal regeneration in

the damaged CNS. They have studied the possibility of tissue replacement in the CNS by means of polymer–cell hybrid constructs in which dissociated cells are included within a stable three-dimensional polymer matrix with sufficient void volume for cell biomass accumulation and a sufficient surface area for cell growth.^[72] For this, they have used hydrogels of *N*-(2-hydroxypropyl) methacrylamide (HPMA) cross-linked with BIS, which contain collagen, in which embryonic neurons, neonatal astrocytes, and Schwann cells have been immobilized, which in most cases were viable and were differentiated within the polymeric matrices. Initial studies of implantation of these hydrogels in rat neocortex showed that the polymers did indeed interact with the host as evidenced by the migration of host nonneuronal cells into the gels.^[73] The inclusion of amino sugar or peptide sequences in this type of hydrogels increases their adhesion properties with host neural tissue, apart from experiencing vascularization and being infiltrated by host nonneuronal cells, such as astrocytes and macrophages; the presence of a high number of axons has also been detected.^[74] This type of hydrogels containing nervous tissue cells can be used for repairing and regenerating the CNS.

Tissue Engineering

The objective of tissue engineering is to develop parts of the body as alternatives for tissue and organ transplants. It is essential to have a scaffolding material that guides cellular adhesion, the growth and the formation of new tissues in this development. Recently, hydrogels have been explored as scaffolding materials for applications in tissue engineering.

Among the possible materials that can be used for the preparation of hydrogels used in tissue engineering are the polysaccharides. The formation of hydrogels based on polysaccharides can be achieved, in general, through hydrogen links or by ionic interactions. Hydrogels of agarose and chitosan are among the first, and hydrogels of alginates and carrageenans are among the second.^[75] Chitosan can be processed through freezing and lyophilizing chitosan–acetic acid solutions in suitable molds, in order to form porous structures that can be used in cell transplantation and tissue regeneration,^[76] because the size and the orientation of the pores can be controlled, and as a result the mechanical properties of chitosan scaffolds. One of the applications described for chitosan-based hydrogels is in cartilage tissue engineering.^[75] Alginate hydrogels ionically cross-linked with calcium carbonate, calcium sulfate, and D-glucono- δ -lactone have also been developed, which generate homogenous matrices, with

appropriate mechanical properties that allow the incorporation of cells, specifically osteoblasts, and their culture in three dimensions.^[50]

Another type of hydrogel is that of poly(ethylene oxide) dimethacrylate (PEODM), obtained by photopolymerization in the presence of 2-hydroxy-1-[4-(hydroxyethoxy)phenyl]-2-methyl-1-propanone as UV photoinitiator, which encapsulates chondrocytes and can be used as scaffolds for cartilage growth.^[77]

With the aim of increasing cellular adhesion or helping a biospecific cellular adhesion, synthetic biomaterials modified with peptide sequences that constitute ligands of cellular adhesion have been prepared.^[78] Following this strategy, hydrogels of polyethylene glycol diacrylate with acryloyl-polyethylene glycol-RGDS have been synthesized, where RGDS is a peptide, which include TGF- β 1 (transforming growth factor- β 1) conjugated with PEG, because TGF- β increases the production of extracellular matrix proteins by vascular smooth muscle cells. The smooth muscle cells from the thoracic aorta of rats included in the PEG-based hydrogel grow better in this type of hydrogels, compared to those that lack TGF- β . Furthermore, this compound significantly improves the mechanical properties of the hydrogel.^[79]

The use of hydrogels in the preparation of scaffolds for other applications such as dental pulp tissue replacement^[80] or mandible condylar reconstruction has also been tackled.^[81]

Biosensors

Hydrogels are used in the preparation of biosensors, acting as supports for the immobilization of enzymes. The Veronese group has prepared diverse biosensors for enzyme immobilization, one of them being an amperometric sensor constructed by using PEG-modified glucose oxidase immobilized in a PVA cryogel membrane, obtained by a freezing–thawing cyclic process. This sensor allows for the determination of glucose electrochemically by measuring the hydrogen peroxide production as a result of the enzymatic reaction, which can be used in the determination of serum glucose.^[82] Likewise, considering the clinical interest in determining acetylcholine, they have developed a specific biosensor through the immobilization of acetylcholinesterase and choline oxidase (as PEG-modified choline oxidase) in a PVA cryogel membrane. This biosensor allows amperometrical determination of the hydrogen peroxide formed after the hydrolysis of acetylcholine by the acetylcholinesterase, which produces choline, in which the choline oxidase acts on generating hydrogen peroxide.^[83] The PEGylation step

is the critical step to increase the enzyme size needed to improve retention inside the gel and to convey at the same time the solubility of the enzyme inside the matrix. This fact is also confirmed in the immobilization of lipase, previously conjugated with PEG, in PVA hydrogels, designed for biocatalytic applications.^[84]

Because the determination of hydrogen peroxide is in itself of interest, hydrogel-modified graphite electrodes have been prepared through the inclusion of diverse redox enzymes in hydrogels of poly(vinylimidazole) complexed with Os(4,4'-dimethylbipyridine)₂Cl cross-linked with poly(ethylene glycol) diglycidyl ether (PEGDGE), as biosensors for the detection of hydrogen peroxide. Of all the biosensors prepared, the most sensitive H₂O₂ biosensor was that contained the newly purified sweet potato peroxidase.^[85]

Another compound whose measurement is of interest in clinical and forensic laboratories, as well as in food and beverage industries, is ethanol. Hydrogels of poly(carbamoyl)sulfonate (PCS) have been used for the immobilization of alcohol oxidase in order to make a biosensor that acts as an amperometric transducer for the detection of alcohols.^[86]

Wound Dressings

Hydrogels are used to aid in the regeneration of ulcerous and burn wounds because they create a permanent moist medium in the wound, therefore, stimulating the cellular activity in all the stages of the cicatrization process. Furthermore, they absorb exudation and secretions. The moist atmosphere created by the hydrogel softens and detaches the necrotic tissue, helping tissue regeneration in the granulation stage, absorbing the secretions, and creating a favorable atmosphere for the division and cellular activity of the epithelial tissue. At the same time, this type of dressing does not adhere to the wound, which is an advantage for the patient. Likewise, they prevent bacterial contamination of the wounds, being able to even incorporate some type of antibiotic, and furthermore they are permeable to oxygen. To attain satisfactory conditions at the wound site, the dressing should possess a suitable water vapor transmission rate.^[87] Studies carried out with experimental animals confirm that wound dressing based on hydrogels induced a significantly enhanced healing rate in the impaired wound.^[88]

Several laboratories have marketed hydrogel-type dressings (Table 3), all of them characterized by their high water content, greater than 95% in weight in some cases. In terms of their composition, there is a wide variety, they consist of: polyurethane, polyacrylate, polyacrylamide and agar, cross-linked polymers derived from acry-

lates substituted with groups of amine and nitrile, chitosan and PVP, as well as with derivatives of hyaluronic acid with or without alginates.

Dressings prepared from gelatine sponges that incorporate epidermic growth factor (EGF) are being studied, to help cellular proliferation and the epithelial regeneration process, and that are situated between elastomeric polyurethane membranes. The *in vivo* studies in rabbits indicate that these dressings provide a higher degree of reduction in the wound areas.^[89] Wound dressings with antimicrobial compounds have also been prepared; this wound dressing has the capacity to absorb the wound exudates thanks to its layer of carboxymethyl-chitin hydrogel, while its antimicrobial capacity resides in the chlorhexidine gluconate impregnated in a layer of chitosan acetate foam.^[90] Other wound dressings are made up of cross-linked hydrogels formed from blends of PVP with PEG and agar.^[91] Polyvinyl alcohol and PVP blended hydrogels,^[92] as well as PEO and PEO/PVA blend hydrogels^[93] for wound dressings have also been prepared. Poly(2-hydroxyethyl methacrylate)-based artificial skin for advanced wound dressing usage through the inclusion of weaved and knitted fabrics and fibers in soft PHEMA matrix has also been prepared.^[94]

Drug-Delivery Systems

The efficacy of the drug of a pharmaceutical product depends, to a large extent, on its dosage form, in other words, on its bioavailability. Currently, it is generally accepted that the pharmacological activity of a compound is not, in itself, enough to ensure a good therapy, but an

optimal therapeutic effect depends on appropriate dosage systems, on the type of formulation, and on the selection of a specific release system for the drug.

The efficacy of a drug in a specific application requires the maintenance of some appropriate concentration levels (generally plasmatic) of this compound during a prolonged period of time. However, the conventional administration of drugs gives, in most cases, a very poor control of the concentration of these substances in plasma, because they give rise to variations in the concentration of the bioactive product once a specific dose has been applied. The conventional dosage systems can give rise to alternative periods of inefficacy or toxicity. These difficulties have called for the development of new administration techniques for bioactive compounds, directed towards attaining that with a single dose the concentration of the drug in the organism (or even better in their target) is maintained out of the extreme limits of inefficacy and toxicity, during the appropriate period of time.^[95]

Drug-delivery systems allow the release of a drug during a dilated period of time or at a specific moment in the treatment. Furthermore, through these systems, the drug can be released in its target in the organism, being able to decrease its side effects and also prevent biological barriers that stop it from acceding the target.^[96] Among the possible drug-delivery systems, hydrogels have been extensively studied as drug-delivery devices^[96,97] due to their excellent capacity of absorbing water and their swelling and permeability characteristics that enable them to undergo structural changes in response to different physical, chemical, or biological stimuli,^[97,98] for which they have been named “intelligent” or stimuli-responsive drug-delivery systems.

Table 3 Selected wound dressing materials

Trade name	Principal components	Manufacturer
Tenderwet	Polyacrylate	Hartmann
Hidrosorb	Polyurethane	Hartmann
Geliprem	Polyacrylamide/agar	Inibsa
Hyflex	Hypan [®] (acrylate derivatives composed of several sequences of units with pedant hydrophilic groups and several sequences of units with pedant nitrile groups)	Hymedix
Aquatix II	Chitosan derivatives /PVP, polyethylenimine/PVP	Hydromer
Clear Hydrogel Films	Polysaccharide-based hydrogels	Veraco International
Hyalofill	Hyaff (hyaluronic acid-based hydrogels)	Convatec, S.A.
Jaloskin	Hyaff	Convatec, S.A.
Hyalogran	Hyaff	Convatec, S.A.
Hyalgin	Hyaff	Convatec, S.A.

General characteristics

Hydrogels loaded with an active substance are especially useful in the administration of this substance in accessible regions of the organism, such as the buccal, nasal, and vaginal region, as well as for oral or transdermal administration. They can also be placed subcutaneously or in the environs of the target of the bioactive compound, through a surgical incision.^[97]

The inclusion of a drug or active substance in a hydrogel can be carried out in two ways. One of them is through physical methods, introducing the dry gel (xerogel) in a concentrated solution of the drug, followed by the dehydration of the hydrogel once it has reached the swelling equilibrium,^[14,15] in other words, by equilibrium partitioning; or, if the drug is stable enough in the required synthesis condition, including it in the initial polymerization feed mixture, in order to, this way, directly obtain the gel loaded with the drug.^[17] The second form of inclusion of a substance in a hydrogel uses chemical mechanisms, consisting of the immobilization of the substance in the hydrogel matrix through the formation of ester, anhydride or amide type bonds that are later hydrolyzed in aqueous medium, this way getting the release of the drug.^[99]

The release of a drug dispersed in a hydrogel, which is initially in a dehydrated state (xerogel), is produced through the diffusion of the drug from and through the polymeric matrix, initially in a glassy state, by diffusion to the crosscurrent of water or biological fluids.^[100] When the xerogel loaded with the drug is put in contact with the aqueous medium, the water penetrates the matrix and the gel begins to swell, decreasing the glass transition temperature of the polymer and reaching a rubbery state, which increases the permeability of the polymer to the drug and allows it to spread to the exterior.^[101]

In order to interpret the kinetics of swelling of a hydrogel, as well as the release of a drug from a hydrogel, the concept of that sorption of water into glassy hydrogels generally exhibits anomalous behavior ranging from Fickian to Case II diffusion has been used. Fickian and non-Fickian behavior, for swelling as well as for drug release, depends on the rate of polymer relaxation at the glassy–rubbery transition at the swelling interface. To characterize the various types of solute diffusion in polymers, the following equation can be used, for at least the short-time period:^[101]

$$M_t/M_\infty = kt^n$$

where M_t and M_∞ are the mass of water taken up at time t and infinite time in sorption experiments, respectively.

In drug release experiments, M_t and M_∞ are the amounts of drug released at time t and the maximum amount of drug released, respectively, k is a constant which incorporates the structural and geometrical characteristics of the device, and n is the diffusional exponent that determines the release mechanism. For devices with slab geometry, $n = 0.5$ represents a Fickian diffusion mechanism, in this case the drug release is controlled by diffusion; while that when $n = 1$ the drug release is independent of time, and corresponds to a zero-order kinetics, and is called Case II transport. In this case the drug release is controlled by swelling. Values of n between 0.5 and 1 correspond to an anomalous transport, which indicates that the drug release is produced by mechanisms controlled by diffusion as well as by swelling. In the case of polymeric devices with geometries in the shape of a sphere or cylinder, $n = 0.43$ (sphere) or 0.45 (cylinder) indicates Fickian diffusion, while $n = 0.85$ (sphere) or 0.89 (cylinder) corresponds to drug release mechanism of Case II transport; the anomalous transport would be indicated by intermediate values of n .^[102]

Another criterion for predicting if the transport in polymeric gels is controlled by diffusion (Fickian) or by relaxation, is to determine the diffusional Deborah number (De), which is a ratio between the characteristic polymer relaxation time of the polymer (λ) when it is subject to a swelling stress and a characteristic diffusion time (θ), defined as the coefficient between the square of the sample thickness (h) and the coefficient of water diffusion in the polymeric gel (D):^[103]

$$De = \lambda/\theta = \lambda/(h^2/D)$$

If the swelling process is controlled by the relaxation time ($De \gg 1$) or by water diffusion ($De \ll 1$), the diffusion is Fickian. However, if $De \approx 1$, both processes will be produced simultaneously and it will be an anomalous transport.

Another parameter that indicates the characteristics of diffusion in polymers is the swelling interface number (Sw), a dimensionless number that describes the balance between solvent penetration and drug release:^[103]

$$Sw = v\delta/D$$

where v is the velocity of the moving glassy/rubbery front, δ is the thickness of the swollen gel layer, and D is the diffusion coefficient of drug in the polymer. When $Sw \gg 1$ the swelling front advances more quickly than the drug release, for which the release is Fickian. On the contrary, values of $Sw \ll 1$ indicate a zero-order release, since the transport speed of the drug through the solvated region

is quicker than the speed of the glassy–rubbery front. When $Sw \approx 1$ anomalous behavior prevails.

When the systems have a Fickian diffusion, the solution of the differential form of Fick's Law, considering one-dimensional diffusion from a very thin gel phase with slab geometry and a constant diffusion coefficient D , is:^[104]

$$\frac{M_t}{M_\infty} = 1 - \sum_{n=0}^{\infty} \frac{8}{(2n+1)^2 \pi^2} \exp\left(\frac{-D(2n+1)^2 \pi^2 t}{h^2}\right)$$

where h is the thickness of the dry hydrogel and t is the time.

This equation can be approached for early and moderate times of diffusion through the reduced expression:

$$M_t/M_\infty \approx 4(Dt/\pi h^2)^{1/2}, \quad M_t/M_\infty \leq 0.6$$

One criterion of Fickian behavior holds that M_t/M_∞ vs. $t_{1/2}$ plot should be linear up to 60% reduced sorption.

Diffusion coefficients can also be calculated from the late-time approximation:

$$M_t/M_\infty = 1 - [(8/\pi^2) \exp(-\pi^2 Dt/h^2)],$$

$$M_t/M_\infty \geq 0.4$$

This equation is more appropriate for determining drug diffusion coefficients of high-molecular-weight drugs.^[105]

Values of the diffusional exponent n close to 0.50 have been determined for the swelling of hydrogels of poly(2-hydroxyethyl methacrylate-*co*-acrylamide) [p(HEMA-*co*-A)],^[34] as well as for poly(2-hydroxyethyl methacrylate-*co*-methyl methacrylate) [p(HEMA-*co*-MMA)],^[105] among others. In general, the increase of the hydrophilic monomer in a copolymeric hydrogel gives rise to higher diffusion coefficients, which indicate a faster swelling as, for example, hydrogels with 80 wt% of HEMA and 20 wt% of A, cross-linked with 5 wt% of EGDMA, have diffusion coefficients for saline solution (NaCl, 0.9 wt%) uptake of $9.27 \times 10^{-11} \text{ m}^2 \text{ sec}^{-1}$ ($9.27 \times 10^{-7} \text{ cm}^2 \text{ sec}^{-1}$), while that for hydrogels of 60 wt% HEMA and 40 wt% A, with the same percentage of cross-linking agent, the diffusion coefficient is of $15.0 \times 10^{-11} \text{ m}^2 \text{ sec}^{-1}$ ($15.0 \times 10^{-7} \text{ cm}^2 \text{ sec}^{-1}$).^[34] The increase of the hydrophobic monomer (MMA) in copolymeric gels of p(HEMA-*co*-MMA) decreases the value of the diffusion coefficient for water uptake, whose values are $3.03 \times 10^{-7} \text{ cm}^2 \text{ sec}^{-1}$ for the samples with 75 mol% of HEMA, and $1.78 \times 10^{-7} \text{ m}^2 \text{ sec}^{-1}$ for those with 60 mol% of HEMA.^[105]

Drug release from a hydrogel is determined by the physical and chemical characteristics of the polymeric

matrix and by the nature of the drug included. In general, the increase of the hydrophilic monomer in a copolymer helps the drug release,^[34] and the decrease of the pore size of the matrix, produced by an increase of the cross-linking, slows the release, which results in smaller values of the diffusion coefficient of the drug.^[33] The chemical nature of the drug also conditions those diffusion coefficient values. Thus, for example, the release of cytarabine from hydrogels of poly(acrylamide-*co*-monomethyl itaconate) [p(A-*co*-MMI)], specifically from the 75A/25MMI composition, has diffusion coefficient values of $8.25 \times 10^{-11} \text{ m}^2 \text{ sec}^{-1}$,^[106] while the diffusion coefficient for the release of 5-fluorouracil, from the same hydrogels is $13.3 \times 10^{-11} \text{ m}^2 \text{ sec}^{-1}$.^[107] A complete study of the effect of the size of different drugs in the value of their diffusion coefficient from different hydrogels has been carried out by the Peppas' group.^[105] Some drugs released from different hydrogels are shown in Table 4.

Stimuli-responsive devices

Because hydrogels can be synthesized from monomers that give them sensitivity to different environmental stimuli, this characteristic can be used to release drugs or bioactive compounds included in hydrogels in response to those stimuli.

One of the stimuli that a hydrogel can be sensitive to is temperature. Within this group, we find examples such as thermosensitive hydrogels of poly(acryloyl-L-proline-methyl ester) [p(A-ProOMe)] which have been used for releasing gentamicin, used in the treatment of infectious diseases, isoniazid, for the treatment of tuberculosis, and insulin, for the treatment of diabetes.^[41] Hydrogel shrinking or swelling was found to play a relevant role in high-molecular-weight drug release. Thus, in the case of insulin, the release is affected by the temperature-sensitive properties of the hydrogel. However, these properties do not permit the control of the release of compounds of low molecular weight, such as gentamicin or isoniazid from these hydrogels. There are other compounds that have also been released from thermosensitive hydrogels. Such is the case of streptokinase, an antithrombotic agent, from block copolymers of poly(*N*-isopropyl acrylamide) and poly-(methacrylic acid) (PMAA).^[108] Likewise, the hormone calcitonin has been included in poly(*N*-isopropyl acrylamide-*co*-butylmethacrylate-*co*-acrylic acid) hydrogel, which responds to changes of pH and temperature.^[109]

Some of the temperature-sensitive hydrogels are also pH sensitive.^[108,109] The pH sensitivity of the hydrogels is due to the presence of weakly acidic and/or basic functional groups on the backbone, and in many cases,

Table 4 Drugs released from different hydrogels

Hydrogel	Drug	Reference
Poly(2-hydroxyethyl methacrylate)	Thiamine HCl	[30]
	L-Ascorbic acid	[15]
	Cytarabine	[17,18]
	5-Fluorouracil	[22,23]
Poly(2-hydroxyethyl methacrylate- <i>co</i> -acrylamide)	5-Fluorouracil	[126]
	Cytarabine	[34]
Poly(2-hydroxyethyl methacrylate- <i>co</i> -methyl methacrylate)	Theophylline, triamterene, oxprenolol, buflomedil, vitamin B ₁₂ , inulin, myoglobin	[105]
Poly(2-hydroxyethyl methacrylate- <i>co</i> - <i>N</i> -vinyl-2-pyrrolidone)	Cytarabine	[33,127]
Poly[2-hydroxyethyl methacrylate- <i>co</i> - <i>N,N'</i> -dimethyl- <i>N</i> -methacryloyloxyethyl- <i>N</i> -(3-sulfopropyl) ammonium]	Sodium salicylate	[14]
Poly(methyl methacrylate- <i>co</i> -dimethylaminoethyl methacrylate)	Aminopyrine, caffeine, theobromine	[110]
Poly(ethylacrylate methyl methacrylate) (Eudragit)	Metoprolol tartrate	[117]
Poly(acrylic acid- <i>co</i> -acrylamide)	Methotrexate	[128]
Poly(acrylamide- <i>co</i> -monomethyl itaconate)	5-Fluorouracil	[107]
	Cytarabine	[106]
Poly(acrylic acid- <i>co</i> - <i>n</i> -alkyl methacrylate)	Theophylline, aminophylline	[16]
Modified polyacrylamide- <i>g</i> -guar gum	Diltiazem, nifedipine	[31]
Poly(acryloyl-L-proline-methyl ester)	Gentamicin, isoniazid, insulin	[41]
Poly(methacrylated inulin- <i>co</i> -bis(methacryloylamino) azobenzene- <i>co</i> -2-hydroxyethyl methacrylate)	Prednisolone	[124]
Poly(ethylene oxide)	Salicylic acid	[12]
Poly(ethylene glycol) diacrylate	Bovine serum albumin (BSA)	[25]
Chitosan	Capsaicin, nomivamide	[129]
Chitosan/poly(<i>N</i> -vinyl-2-pyrrolidone)	Amoxicillin	[45]

they have been designed for the release of substances in areas of the organism with characteristic pH values, such as the stomach, where the pH is very acidic, or the intestine whose pH is neutral or basic. Thus, diltiazem hydrochloride and nifedipine, two antihypertensive drugs, have been released from cross-linked hydrogels of polyacrylamide-grafted guar gum (pAAm-*g*-GG) in simulated gastric and intestinal pH conditions,^[31] observing that the release was faster in pH 7.4 than in an acidic medium. The diffusion characteristics of aminopyrine, caffeine and theobromide, all of them weak bases whose maximum solubility reaches pH 1.2, and that have different solubility in water, from p(MMA/DMA) hydrogels cross-linked with DVB,^[110] a polycationic hydrogel that is in a hydrated state at acid pH and is dehydrated and stiff at neutral pH, have also been studied. A pH-sensitive hydrogel whose maximum swelling produces low acidic conditions is the semi-interpenetrant gel based on chitosan/PVP,^[45] which has been studied as a device for releasing the antibiotic amoxicillin in the acidic conditions of the stomach.

Other hydrogels have been designed incorporating different compounds, generally enzymes, which allow

them to respond to specific molecules. Thus, hydrogels of poly(diethylaminoethyl-*g*-ethylene glycol) [p(DEAEM-*g*-EG)] that contain immobilized glucose oxidase, which gives them glucose sensitivity, have been synthesized. Catalase has also been immobilized in these gels, with the intention that it decreases the amount of hydrogen peroxide formed in the reaction of the glucose oxidase and the amount of available oxygen is increased for the formation of acid. The enzymes are active in the gels and produce an acidic environment in the hydrogel, which causes the swelling of the polymeric matrix.^[111] The aim of these systems is to achieve a feedback-controlled release of insulin, in such a way that it can control the release of insulin absorbed in the gels through changes in their swelling modulated by glucose.

Other systems have been designed that respond, for example, to calcium, such as in the case of a starch-cellulose matrix that contains α -amylase, and whose degradation is produced in the presence of calcium, which allows the release of the drugs incorporated in the matrix.^[112] Hydrogels based on PVA that release antibiotics only in the presence of infection, due to the fact that they are thrombin sensitive,^[113] have also been

synthesized. Another system for the release of drugs or bioactive substances consists of an antigen-responsive hydrogel formed by an antibody-grafted linear polyacrylamide and the cross-linked polyacrylamide grafted with the corresponding antigen.^[114]

Routes of drug delivery

The aim of many of the hydrogels that have been designed for controlled drug release is to have their administration through different routes, for example, oral, nasal, vaginal, subcutaneous, or transdermal.^[97]

The oral administration of drugs through hydrogels is one of the routes that has aroused the most interest among researchers, who have tackled this form of administration, mainly through two strategies.

Mucoadhesive hydrogels. One of these strategies is the development of mucoadhesive hydrogels^[115] that interact with the mucus as a result of physical entanglement and secondary bonding, mainly through hydrogen bonding and van der Waals forces, due to the presence of hydroxyl, carboxyl, amine, and amide groups in the surface of the polymeric matrix,^[116] this way prolonging the residence time of the dosage form on the site of absorption.

The use of the buccal cavity for placing devices of controlled drug release allows it to avoid the first-pass metabolism and prevents the degradation of the drug in the gastrointestinal tract. With this aim, Eudragit buccal patches, a neutral copolymer of poly(ethylacrylate methylmethacrylate), with hydrophilic polymers that confer it bioadhesiveness, for the controlled release of metoprolol, a drug for the treatment of hypertension, have been prepared. The release of metoprolol is modified by the addition of hydrophilic polymers, due to their capacity of absorbing water.^[117]

Diverse materials that stimulate the bioadhesion with mucus have been studied. The bioadhesiveness of chitosan microspheres and starch microspheres has been studied through nasal administration in human volunteers, with the aim of characterizing them as bioadhesive nasal delivery systems. The studies indicate that chitosan delivery systems can reduce the rate of clearance from the nasal cavity, notably increasing the contact time between the release system and the nasal cavity.^[118] The adhesion characteristics of polyampholytic hydrogels formed by blends of deacetylated and *O*-carboxymethylated chitins have also been studied, observing that they are governed by the aqueous content of these hydrogels, those with a lower water content presenting greater adhesion strength.^[119]

The bioadhesive properties of hydrogels of poly(*N,N*-dimethylaminoethyl methacrylate-*co*-methyl methacrylate) [p(DMA/MMA)] cross-linked with DVB have been studied in vitro using pig gastric, sublingual, vaginal, and intestinal mucus. These hydrophobic polybasic gels have good adhesive properties in all of the mucus studied, which makes them candidates for the controlled release of drugs in these places.^[120]

Due to the turnover of the mucus gel layer of the mucus surface, which in some tissues such as the intestine is very fast, the adhesion of mucoadhesive devices of drug release is more difficult. This fact, along with the nonspecificity of adhesion of these devices has caused this to be researched in the development of release systems that include lectins and lectin-like molecules, which can directly interact with the cellular membrane in such a way that a cytoadhesion independent of mucus turnover is produced.^[121]

Sensitivity to pH changes. The second strategy developed for the oral administration of drugs through hydrogels is based on their sensitivity to changes in pH in different areas of the digestive tract,^[31,45,110,122,123] which allows them to swell to acid or neutral pH and to release the bioactive compound in areas such as the stomach or intestine.

Another approximation to make the oral administration of drugs through hydrogels specific has consisted of the design of drug release systems in the colon, taking advantage of the presence of specific bacteria, which form part of the bacteria flora, in this region. These bacteria secrete enzymes capable of hydrolyzing some polysaccharides, such as dextrano or inulin, and they are also able to reduce azo-compounds. These azo-inulin and azo-dextran gels can be hydrolyzed specifically by inulinase and dextranase, respectively, in the colon.^[124,125]

Another route of drug administration using hydrogels is done through subcutaneous implantation. In this case, the amount of water of the hydrogel is an important factor when considering its biocompatibility, because the greater the amount is, the more similar the hydrogel is to the individual's tissues. The presence or not of a thin fibrous capsule around the hydrogel discs implanted subcutaneously largely depends on their water content in the swelling equilibrium, which also affects the period of time necessary to its formation.^[126] When the drugs are administered through hydrogels subcutaneously, their plasmatic half-life is notably increased. Thus, for example, the administration of 5-fluorouracil through hydrogels of p(HEMA-*co*-A), cross-linked with EGDMA, increases the half-life of the drug 98 times ($t_{1/2} = 25.11$ hr), compared to its administration by intraperitoneal injection ($t_{1/2} = 0.257$ hr).^[126] An equivalent behavior is noticeable

when the antineoplastic cytarabine is administered subcutaneously through its inclusion in hydrogels of PHEMA and poly(2-hydroxyethyl methacrylate-*co*-*N*-vinyl-2-pyrrolidone) [p(HEMA-*co*-VP)] cross-linked with EGDMA, because its presence is detected in plasma between 4 and 16 days, despite the fact that the plasmatic half-life of it is 2.6 hr when it is administered by injection.^[127]

The administration of drugs through the skin through transdermal release from hydrogels means an advantage in the treatment of some illnesses. Thus, the iontophoretic delivery of methotrexate from hydrogels of AA and its copolymers with acrylamide has been studied, because this antineoplastic is used in the treatment of psoriasis. These hydrogels seem to be suitable for the topical administration of methotrexate, which would prevent secondary effects, fundamentally hepatotoxic, which are caused when it is administered orally.^[128] Other drugs included in hydrogels, whose transdermal administration has been studied, are capsaicin and nonivamide, which are used topically in the treatment of rheumatoid arthritis.^[129] Likewise, chitosan-based transdermal drug-delivery systems have been designed for releasing propranolol hydrochloride,^[130] a drug that is used in the treatment of tachycardia, trembling, and cardiac arrhythmias.

Hydrogels can be synthesized with a wide variety of characteristics, obtaining biostable as well as biodegradable polymeric matrices, whose mechanical properties and degree of swelling fit with a specific application. This makes them a group of materials with a wide and diverse spectrum of uses, which gives them a promising future in the field of biomedicine.

ACKNOWLEDGMENTS

This work was supported by grant number MAT2000-0768-CO2-01 from Ministerio de Ciencia y Tecnología (Spain).

REFERENCES

1. Bouwstra, J.A.; Junginger, H.E. Hydrogels. In *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C. Eds; Marcel Dekker, Inc.: New York, 1993; Vol. 7, 441–465.
2. Blanco, M.D.; García, O.; Trigo, R.M.; Teijón, J.M.; Katime, I. 5-Fluorouracil release from copolymeric hydrogels of itaconic acid monoester. I. Acrylamide-*co*-monomethyl itaconate. *Biomaterials* **1996**, *17* (11), 1061–1067.
3. Rao, J.K.; Ramesh, D.V.; Rao, K.P. Implantable controlled delivery systems for proteins based on collagen-pHEMA hydrogels. *Biomaterials* **1994**, *15* (5), 383–389.
4. Roberts, G.A.F. *Chitin Chemistry*; The Macmillan Press Ltd.: London, 1992.
5. Luo, Y.; Kirker, K.R.; Prestwich, G.D. Cross-linked hyaluronic acid hydrogel films: New biomaterials for drug delivery. *J. Control. Release* **2000**, *69*, 169–184.
6. Lowman, A.M.; Peppas, N.A. Hydrogels. In *Encyclopedia of Controlled Drug Delivery*; Mathowitz, E. Ed.; John Wiley & Sons: New York, 1999; Vol. 2, 397–418.
7. Peppas, N.A.; Mikos, A.G. Preparation Methods and Structure of Hydrogels. In *Hydrogels in Medicine and Pharmacy*; Peppas, N.A. Ed.; CRC Press: Boca Raton, FL, 1986; Vol. 1, 1–27.
8. Katime, I. *Química Física Macromolecular*; Servicio Editorial de la Universidad del País Vasco: Bilbao, Spain, 1994.
9. Harris, J.M. Introduction to Biotechnical and Biomedical Applications of Poly(ethylene glycol). In *Poly(ethylene glycol) Chemistry: Biotechnical and Biomedical Applications*; Harris, J.M. Ed.; Plenum Press: New York, 1992; 1–14.
10. Stringer, J.L.; Peppas, N.A. Diffusion of small molecular weight drugs in radiation-crosslinked poly(ethylene oxide) hydrogels. *J. Control. Release* **1996**, *42*, 195–202.
11. Peppas, N.A.; Keys, K.B.; Torres-Lugo, M.; Lowman, A.M. Poly(ethylene glycol)-containing hydrogels in drug delivery. *J. Control. Release* **1999**, *62*, 81–87.
12. Savas, H.; Güven, O. Investigation of active substance release from poly(ethylene oxide) hydrogels. *Int. J. Pharm.* **2001**, *224*, 151–158.
13. Ebdon, J.R., Ed.; *New Methods of Polymer Synthesis*; Blackie & Sons Ltd.: Glasgow, 1991.
14. Blanco, M.D.; Rego, J.M.; Huglin, M.B. Drug release with simultaneous dimensional changes from a new copolymeric hydrogel. *Polymer* **1994**, *35*, 3487–3491.
15. Trigo, R.M.; Blanco, M.D.; Huerta, P.; Olmo, R.; Teijón, J.M. L-Ascorbic acid release from PHEMA hydrogels. *Polym. Bull* **1993**, *31*, 577–584.
16. Katime, I.; Novoa, R.; Zuluaga, F. Swelling kinetics and release studies of theophylline and aminophylline from acrylic acid/*n*-alkyl methacrylate hydrogels. *Eur. Polym. J.* **2001**, *37*, 1465–1471.
17. Teijón, J.M.; Trigo, R.M.; García, O.; Blanco, M.D. Cytarabine trapping in poly(2-hydroxyethyl methacrylate) hydrogels: Drug delivery studies. *Biomaterials* **1997**, *18*, 383–388.
18. Trigo, R.M.; Blanco, M.D.; Teijón, J.M.; Sastre, R. Anticancer drug, ara-C, release from PHEMA hydrogels. *Biomaterials* **1994**, *15*, 1181–1186.
19. Davis, T.P.; Huglin, M.B.; Yip, D.C.F. Properties of poly(*N*-vinyl-2-pyrrolidone) hydrogels crosslinked with ethyleneglycol dimethacrylate. *Polymer* **1988**, *29*, 701–706.

20. Davis, T.P.; Huglin, M.B. Studies on copolymeric hydrogels of *N*-vinyl-2-pyrrolidone with 2-hydroxyethyl methacrylate. *Macromolecules* **1989**, *22*, 2824–2829.
21. Baker, D.A.; Corkhill, P.H.; Ng, C.O.; Skelly, P.J.; Tighe, B.J. Synthetic hydrogels: 2. Copolymers of carboxyl-, lactam- and amide-containing monomers—Structure/property relationships. *Polymer* **1988**, *29*, 691–700.
22. García, O.; Trigo, R.M.; Blanco, M.D.; Teijón, J.M. Influence of degree of crosslinking on 5-fluorouracil release from poly(2-hydroxyethyl methacrylate) hydrogels. *Biomaterials* **1994**, *15*, 689–694.
23. García, O.; Blanco, M.D.; Gómez, C.; Teijón, J.M. Effect of the crosslinking with 1,1,1-trimethylolpropane trimethacrylate on 5-fluorouracil release from poly(2-hydroxyethyl methacrylate) hydrogels. *Polym. Bull.* **1997**, *38*, 55–62.
24. Ruiz, J.; Mantecón, A.; Cádiz, V. Synthesis and properties of hydrogels from poly(vinyl alcohol) and ethylenediaminetetraacetic dianhydride. *Polymer* **2001**, *42*, 6347–6354.
25. Mellott, M.B.; Searcy, K.; Pishko, V. Release of protein from highly cross-linked hydrogels of poly(ethylene glycol) diacrylate fabricated by UV polymerization. *Biomaterials* **2001**, *22*, 929–941.
26. Zilkha, A. Non-bond crosslinked polymer hydrogels. *Eur. Polym. J.* **2001**, *37*, 2145–2146.
27. Ruiz, J.; Mantecón, A.; Cádiz, V. Synthesis and properties of hydrogels from poly(vinyl alcohol) and ethylenediaminetetraacetic dianhydride. *Polymer* **2001**, *42*, 6347–6354.
28. Oxley, H.R.; Corkhill, P.H.; Fitton, J.H.; Tighe, B.J. Macroporous hydrogels for biomedical applications: Methodology and morphology. *Biomaterials* **1993**, *14*, 1064–1072.
29. Dziubla, T.D.; Torjman, M.C.; Joseph, J.I.; Murphy-Tatum, M.; Lowman, A.M. Evaluation of porous networks of poly(2-hydroxyethyl methacrylate) as interfacial drug delivery devices. *Biomaterials* **2001**, *22*, 2893–2899.
30. Lee, P.I. Dimensional changes during drug release from a glassy hydrogel matrix. *Polym. Commun.* **1983**, *24*, 45–47.
31. Soppimath, K.S.; Kulkarni, A.R.; Aminabhavi, T.M. Chemically modified polyacrylamide-*g*-guar gum-based crosslinked anionic microgels as pH-sensitive drug delivery systems: Preparation and characterization. *J. Control. Release* **2001**, *75*, 331–345.
32. Huglin, M.B.; Rehad, M.M.A.M.; Zakaria, M.B. Thermodynamic interactions in copolymeric hydrogels. *Macromolecules* **1986**, *19*, 2986–2991.
33. Blanco, M.D.; Trigo, R.M.; García, O.; Teijón, J.M. Controlled release of cytarabine from poly(2-hydroxyethyl methacrylate-*co*-*N*-vinyl-2-pyrrolidone) hydrogels. *J. Biomater. Sci. Polym. Ed.* **1997**, *8*, 709–719.
34. Sastre, R.L.; Blanco, M.D.; Gómez, C.; Socorro, J.M.; Teijón, J.M. Cytarabine trapping in poly(2-hydroxyethyl methacrylate-*co*-acrylamide) hydrogels: Drug delivery studies. *Polym. Int.* **1999**, *48*, 843–850.
35. Xue, W.; Champ, S.; Huglin, M.B. Network and swelling parameters of chemically crosslinked thermoreversible hydrogels. *Polymer* **2001**, *42*, 3665–3669.
36. Li, S.K.; D'Emanuele, A. On-off transport through a thermoresponsive hydrogel composite membrane. *J. Control. Release* **2001**, *75*, 55–67.
37. Xue, W.; Champ, S.; Huglin, M.B. Thermoreversible swelling behaviour of hydrogels based on *N*-isopropylacrylamide with a zwitterionic comonomer. *Eur. Polym. J.* **2001**, *37*, 869–875.
38. Ju, H.K.; Kim, S.Y.; Lee, Y.M. pH/temperature-responsive behaviour of semi-IPN and comb-type graft hydrogels composed of alginate and poly(*N*-isopropylacrylamide). *Polymer* **2001**, *42*, 6851–6857.
39. Makino, K.; Fujita, Y.; Takao, K.; Kobayashi, S.; Ohshima, H. Preparation and properties of thermosensitive hydrogel microcapsules. *Colloids Surf. B: Biointerf.* **2001**, *21*, 259–263.
40. Martellini, F.; Higa, O.Z.; Takacs, E.; Safran, A.; Yoshida, M.; Kataskai, R.; Carenza, M. Thermally reversible gels based on acryloyl-L-proline methyl ester. *Radiat. Phys. Chem.* **1999**, *55*, 185–192.
41. Caliceti, P.; Salmaso, S.; Lante, A.; Yoshida, M.; Katakai, R.; Martellini, F.; Mei, L.H.I.; Carenza, M. Controlled release of biomolecules from temperature-sensitive hydrogels prepared by radiation polymerization. *J. Control. Release* **2001**, *75*, 173–181.
42. Siegel, R.A.; Falamarzian, M.; Firestone, B.A.; Moxley, B.C. pH-controlled release hydrophobic/polyelectrolyte copolymer hydrogels. *J. Control. Release* **1988**, *8*, 179–182.
43. Falamarzian, M.; Varshosaz, J. The effect of structural changes on swelling kinetics of polybasic/hydrophobic pH-sensitive hydrogels. *Drug. Dev. Ind. Pharm.* **1998**, *24*, 667–669.
44. Gan, L.H.; Deen, G.R.; Gan, Y.Y.; Tam, K.C. Water sorption studies of new pH-responsive *N*-acryloyl-*N*-methyl piperazine and methyl methacrylate hydrogels. *Eur. Polym. J.* **2001**, *37*, 1473–1478.
45. Risbud, M.V.; Hardikar, A.A.; Bhat, S.V.; Bhonde, R.R. pH-sensitive freeze-dried chitosan-polyvinyl pyrrolidone hydrogels as controlled release system for antibiotic delivery. *J. Control. Release* **2000**, *68*, 23–30.
46. Anseth, K.S.; Bowman, C.N.; Brannon-Peppas, L. Mechanical properties of hydrogels and their experimental determination. *Biomaterials* **1996**, *17*, 1647–1657.
47. Davis, T.D.; Huglin, M.B.; Yip, D.C.F. Properties of poly(*N*-vinyl-2-pyrrolidone) hydrogels crosslinked with ethyleneglycol dimethacrylate. *Polymer* **1988**, *29*, 701–706.
48. Huglin, M.B.; Rehab, M.M.A.M. Mechanical and thermodynamic properties of butyl acrylate-*N*-vinylpyrrolidone hydrogels. *Polymer* **1987**, *28*, 2200–2206.

49. Peppas, N.A.; Merrill, E.W. Poly(vinyl alcohol) hydrogels: Reinforcement of radiation-crosslinked networks by crystallization. *J. Polym. Sci. Polym. Chem.* **1976**, *14*, 441–445.
50. Kuo, C.K.; Ma, P.X. Ionically crosslinked alginate hydrogels as scaffolds for tissue engineering: Part 1. Structure, gelation rate and mechanical properties. *Biomaterials* **2001**, *22*, 511–521.
51. Risbud, M.V.; Bhat, S.V. Properties of polyvinyl pyrrolidone/ β -chitosan hydrogel membranes and their biocompatibility evaluation by haemorheological method. *J. Mater. Sci. Mater. Med.* **2001**, *12*, 75–79.
52. Hao, J.; Deng, X. Semi-interpenetrating networks of bacterial poly(3-hydroxybutyrate) with net-poly(ethylene glycol). *Polymer* **2001**, *42*, 4091–4097.
53. Stammen, J.A.; Williams, S.; Ku, D.N.; Guldberg, R.E. Mechanical properties of a novel PVA hydrogel in shear and unconfined compression. *Biomaterials* **2001**, *22*, 799–806.
54. Wichterle, O.; Lim, D. Hydrophilic gels for biological use. *Nature* **1960**, *185*, 117–118.
55. Pedley, D.G.; Skelly, P.J.; Tighe, B.J. Hydrogels in biomedical applications. *Br. Polym. J.* **1980**, *12*, 99–110.
56. Lloyd, A.W.; Faragher, R.G.A.; Denyer, S.P. Ocular biomaterials and implants. *Biomaterials* **2001**, *22*, 769–785.
57. Toit, R.; Simpson, T.L.; Fonn, D.; Chalmers, R.L. Recovery from hyperemia after overnight wear of low and high transmissibility hydrogel lenses. *Curr. Eye Res.* **2001**, *22*, 68–73.
58. Harvitt, D.; Bonanno, J. Re-evaluation of the oxygen diffusion model for predicting minimum contact lens Dk/t values needed to avoid corneal anoxia. *Optom. Vis. Sci.* **1999**, *76*, 712–719.
59. Kohnen, T. The variety of foldable intraocular lens materials. *J. Cataract Refract. Surg.* **1996**, *22*, 1255–1257.
60. Chirila, T.V.; Hong, Y.; Dalton, P.D.; Constable, I.J.; Refojo, M.F. The use of hydrophilic polymers as artificial vitreous. *Prog. Polym. Sci.* **1998**, *23*, 475–508.
61. Broom, N.D.; Oloyede, A. The importance of physico-chemical swelling in cartilage illustrated with a model hydrogel system. *Biomaterials* **1998**, *19*, 1179–1188.
62. Oka, M.; Noguchi, T.; Kumar, P.; Ikeuchi, K.; Yamamuro, T.; Hyon, S.H.; Ikada, Y. Development of an artificial articular cartilage. *Clin. Mater.* **1990**, *6*, 361–381.
63. Noguchi, T.; Yamamuro, T.; Oka, M.; Kumar, P.; Kotoura, Y.; Hyon, S.; Ikada, Y. Poly(vinyl alcohol) hydrogel as an artificial articular cartilage: Evaluation of biocompatibility. *J. Appl. Biomater.* **1991**, *2*, 101–107.
64. Oka, M.; Ushio, K.; Kumar, P.; Ikeuchi, K.; Hyon, S.H.; Nakamura, T.; Fujita, H. Development of artificial articular cartilage. *Proceedings of the Institution of Mechanical Engineers. Part H. J. Eng. Med.* **2000**, *214*, 59–68.
65. Corkhill, P.H.; Fitton, J.H.; Tighe, B.J. Towards a synthetic articular cartilage. *J. Biomater. Sci. Polym. Ed.* **1993**, *4*, 615–630.
66. Freeman, M.E.; Furey, M.J.; Love, B.J.; Hampton, J.M. Friction, wear, and lubrication of hydrogels as synthetic articular cartilage. *Wear* **2000**, *241*, 129–135.
67. Gu, Z.Q.; Xiao, J.M.; Zhang, X.H. The development of artificial articular cartilage—PVA—hydrogel. *Biomed. Mat. Eng.* **1998**, *8*, 75–81.
68. Ambrosio, L.; De Santis, R.; Iannace, S.; Netti, P.A.; Nicolais, L. Viscoelastic behaviour of composite ligament prostheses. *J. Biomed. Mater. Res.* **1998**, *42*, 6–12.
69. Ersek, R.A.; Salisbury, A.V. Textured surface, nonsilicone gel breast implants: four years' clinical outcome. *Plast. Reconstr. Surg.* **1997**, *100*, 1729–1739.
70. Benediktsson, K.; Perbeck, L.G. Fluid retention in Bioplasty Misti Gold II breast prostheses with development of capsular contracture. *Scand. J. Plast. Reconstr. Surg. Hand Surg.* **2000**, *34*, 65–70.
71. Dalton, P.D.; Shoichet, M.S. Creating porous tubes by centrifugal forces for soft tissue application. *Biomaterials* **2001**, *22*, 2661–2669.
72. Woerly, S.; Plant, G.W.; Harvey, A.R. Neural tissue engineering: From polymer to biohybrid organs. *Biomaterials* **1996**, *17*, 301–310.
73. Woerly, S.; Plant, G.W.; Harvey, A.R. Cultured rat neuronal and glial cells entrapped within hydrogel polymer matrices: A potential tool for neural tissue replacement. *Neurosci. Lett.* **1996**, *205*, 197–201.
74. Plant, G.W.; Woerly, S.; Harvey, A.R. Hydrogels containing peptide or aminosugar sequences implanted into the rat brain: Influence on cellular migration and axonal growth. *Exp. Neurol.* **1997**, *143*, 287–299.
75. Suh, J.K.F.; Matthew, H.W.T. Application of chitosan-based polysaccharide biomaterials in cartilage tissue engineering: A review. *Biomaterials* **2000**, *21*, 2589–2598.
76. Madihally, S.V.; Matthew, H.W. Porous chitosan scaffolds for tissue engineering. *Biomaterials* **1999**, *20*, 1133–1142.
77. Bryant, S.J.; Anseth, K.S. The effects of scaffold thickness on tissue engineered cartilage in photocrosslinked poly(ethylene oxide) hydrogels. *Biomaterials* **2000**, *22*, 619–626.
78. Cook, A.D.; Hrkach, J.S.; Gao, N.N.; Johnson, I.M.; Pajvani, U.B.; Cannizzaro, S.M.; Langer, R. Characterization and development of RGD-peptide-modified poly(lactic acid-co-lysine) as an interactive resorbable biomaterial. *J. Biomed. Mater. Res.* **1997**, *35*, 513–523.
79. Mann, B.K.; Schmedlen, R.H.; West, J.L. Tethered-TGF- β increases extracellular matrix production of vascular smooth muscle cells. *Biomaterials* **2001**, *22*, 439–444.
80. Bohl, K.S.; Shon, J.; Rutherford, B.; Mooney, D.J. Role of synthetic extracellular matrix in development of engineered dental pulp. *J. Biomater. Sci. Polym. Ed.* **1998**, *9*, 749–764.
81. Weng, Y.; Cao, Y.; Silva, C.A.; Vacanti, M.P.; Vacanti, C.A. Tissue-engineered composites of bone and cartilage for mandible condylar reconstruction. *J. Oral Maxillofac. Surg.* **2001**, *59*, 185–190.

82. Doretto, L.; Ferrara, D.; Gattolin, P.; Lora, S.; Schiavon, F.; Veronese, F.M. PEG-modified glucose oxidase immobilized on a PVA cryogel membrane for amperometric biosensor applications. *Talanta* **1998**, *45*, 891–898.
83. Doretto, L.; Ferrara, D.; Lora, S.; Schiavon, F.; Veronese, F.M. Acetylcholine biosensor involving entrapment of acetylcholinesterase and poly(ethylene glycol)-modified choline oxidase in a poly(vinyl alcohol) cryogel membrane. *Enzyme Microb. Technol.* **2000**, *27*, 279–285.
84. Veronese, F.M.; Mammucari, C.; Schiavon, F.; Schiavon, O.; Lora, S.; Secundo, F.; Chilin, A.; Guiotto, A. Pegylated enzyme entrapped in poly(vinyl alcohol) hydrogel for biocatalytic application. *Il Farmaco* **2001**, *56*, 541–547.
85. Gaspar, S.; Popescu, I.C.; Gazaryan, I.G.; Bautista, A.G.; Sakharov, I.Y.; Mattiasson, B.; Csöregi, E. Biosensors based on novel plant peroxidases: A comparative study. *Electrochim. Acta* **2000**, *46*, 255–264.
86. Patel, N.G.; Meier, S.; Cammann, K.; Chemnitz, G.C. Screen-printed biosensors using different alcohol oxidases. *Sensors Actuators B* **2001**, *75*, 101–110.
87. Wu, P.; Fisher, A.C.; Foo, P.P.; Queen, D.; Gaylor, J.D.S. In vitro assessment of water vapour transmission of synthetic wound dressings. *Biomaterials* **1995**, *16*, 171–175.
88. Kietzmann, M. Improvement and retardation of wound healing: Effects of pharmacological agents in laboratory animal studies. *Vet. Dermatol.* **1999**, *10*, 83–88.
89. Ulubayram, K.; Cakar, A.N.; Korkusuz, P.; Ertan, C.; Hasirci, N. EGF containing gelatin-based wound dressings. *Biomaterials* **2001**, *22*, 1345–1356.
90. Loke, W.K.; Lau, S.K.; Yong, L.L.; Khor, E.; Sum, C.K. Wound dressing with sustained anti-microbial capability. *J. Biomed. Mater. Res.* **2000**, *53*, 8–17.
91. Higa, O.Z.; Rogero, S.O.; Machado, L.D.B.; Mathor, M.B.; Lugão, A.B. Biocompatibility study for PVP wound dressing obtained in different conditions. *Radiat Phys. Chem.* **1999**, *55*, 705–707.
92. Razzak, M.T.; Darwis, D.; Sukirno, Z. Irradiation of polyvinyl alcohol and polyvinyl pyrrolidone blended hydrogel for wound dressing. *Radiat. Phys. Chem.* **2001**, *62*, 107–113.
93. Yoshii, F.; Zhanshan, Y.; Isobe, K.; Shinozaki, K.; Makuuchi, K. Electron beam crosslinked PEO and PEO/PVA hydrogels for wound dressing. *Radiat. Phys. Chem.* **1999**, *55*, 133–138.
94. Young, C.D.; Wu, J.R.; Tsou, T.L. High-strength, ultra-thin and fiber-reinforced pHEMA artificial skin. *Biomaterials* **1998**, *19*, 1745–1752.
95. Vert, M. Polyvalent polymeric drug carriers. *Crit. Rev. Ther. Drug Carrier Syst.* **1986**, *2*, 291–327.
96. Kumar, M.N.V.R.; Kumar, N. Polymeric controlled drug-delivery systems: Perspective issues and opportunities. *Drug Dev. Ind. Pharm.* **2001**, *27*, 1–30.
97. Peppas, N.A.; Bures, P.; Leobandung, W.; Ichikawa, H. Hydrogels in pharmaceutical formulations. *Eur. J. Pharm. Biopharm.* **2000**, *50*, 27–46.
98. Pillai, O.; Panchagnula, R. Polymers in drug delivery. *Curr. Opin. Chem. Biol.* **2001**, *5*, 447–451.
99. Levenfeld, B.; San Román, J.; Bunel, C.; Vairon, J.P. Polymers with pharmacological activity, 5. Hydrolytical behaviour of polymethacrylic hydrogels bearing “Paracetamol” side groups. *Makromol. Chem.* **1991**, *192*, 793–803.
100. Korsmeyer, R.W.; Peppas, N.A. Solute and penetrant diffusion in swellable polymers. III. Drug release from glassy poly(HEMA-*co*-NVP) copolymers. *J. Control. Rel.* **1984**, *1*, 89–98.
101. Lee, P.I. Kinetics of drug release from hydrogel matrices. *J. Control. Rel.* **1985**, *2*, 277–288.
102. Ritger, P.L.; Peppas, N.A. A simple equation for description of solute release. II. Fickian and anomalous release from swellable devices. *J. Control. Rel.* **1987**, *5*, 37–42.
103. Korsmeyer, R.W.; Peppas, N.A. Macromolecular and Modeling Aspects of Swelling-Controlled Systems. In *Controlled Release Delivery Systems*; Roseman, T.J., Mansdorf, S.Z. Eds; Marcel Dekker, Inc.: New York, 1983; 77–90.
104. Crank, J. *The Mathematics of Diffusion*; Chapter 4, 2nd Ed.; Clarendon Press: Oxford, 1975; 4.
105. Brazel, C.S.; Peppas, N.A. Mechanisms of solute and drug transport in relaxing, swellable, hydrophilic glassy polymers. *Polymer* **1999**, *40*, 3383–3398.
106. Blanco, M.D.; Gómez, C.; García, O.; Teijón, J.M. Ara-C release from poly(acrylamide-*co*-monomethyl itaconate) hydrogels: In vitro and in vivo studies. *Polym. Gels Networks* **1998**, *6*, 57–69.
107. Blanco, M.D.; García, O.; Trigo, R.M.; Teijón, J.M.; Katime, I. 5-Fluorouracil release from copolymeric hydrogels of itaconic acid monoester. I. Acrylamide-*co*-monomethyl itaconate. *Biomaterials* **1996**, *17*, 1061–1067.
108. Vakkalanka, S.K.; Brazel, C.S.; Peppas, N.A. Temperature and pH-sensitive terpolymers for modulated delivery of streptokinase. *J. Biomater. Sci. Polym. Ed.* **1996**, *8*, 119–129.
109. Serres, A.; Baudys, M.; Wankim, S. Temperature and pH-sensitive polymers for human calcitonin delivery. *Pharm. Res.* **1996**, *13*, 196–201.
110. Varshosaz, J.; Falamarzian, M. Drug diffusion mechanisms through pH-sensitive hydrophobic/polyelectrolyte hydrogel membranes. *Eur. J. Pharm. Biopharm.* **2001**, *51*, 235–240.
111. Podual, K.; Doyle, F.J., III; Peppas, N.A. Glucose-sensitivity of glucose oxidase-containing cationic copolymer hydrogels having poly(ethylene glycol) grafts. *J. Control. Release* **2000**, *67*, 9–17.
112. Goldbart, R.; Kost, J. Calcium responsive bioerodible drug delivery system. *Pharm. Res.* **1999**, *16*, 1483–1486.
113. Suzuki, Y.; Tanihara, M.; Nishimura, Y.; Suzuki, K.; Kakimaru, Y.; Shimizu, Y. A new drug delivery system with controlled release of antibiotic only in the presence of infection. *J. Biomed. Mater. Res.* **1998**, *42*, 112–116.

114. Miyazaki, T.; Asami, N.; Urgami, T. A reversible antigen-responsive hydrogel. *Nature* **1999**, *399*, 765–769.
115. Nagai, T.; Machida, Y. Buccal delivery systems using hydrogels. *Adv. Drug Deliv. Rev.* **1993**, *11*, 179–191.
116. Peppas, N.A.; Buri, P.A. Surface, interfacial and molecular aspects of polymer bioadhesion on soft tissues. *J. Control. Release* **1985**, *2*, 257–275.
117. Wong, C.F.; Yuen, K.H.; Peh, K.K. Formulation and evaluation of controlled release Eudragit buccal patches. *Int. J. Pharm.* **1999**, *178*, 11–22.
118. Soane, R.J.; Frier, M.; Perkins, A.C.; Jones, N.S.; Davis, S.S.; Illum, L. Evaluation of the clearance characteristics of bioadhesive systems in humans. *Int. J. Pharm.* **1999**, *178*, 55–65.
119. Zhao, X.; Kato, K.; Fukumoto, Y.; Nakamae, K. Synthesis of bioadhesive hydrogels from chitin derivatives. *Int. J. Adhes. Adhes.* **2001**, *21*, 227–232.
120. Quintanar-Guerrero, D.; Villalobos-García, R.; Alvarez-Colín, E.; Cornejo-Bravo, J.M. In vitro evaluation of the bioadhesive properties of hydrophobic polybasic gels containing *N,N*-dimethylaminoethyl methacrylate-*co*-methyl methacrylate. *Biomaterials* **2001**, *22*, 957–961.
121. Lehr, C.M. Lectin-mediated drug delivery: The second generation of bioadhesives. *J. Control. Release* **2000**, *65*, 19–29.
122. Torres-Lugo, M.; Peppas, N.A. Molecular design and in vitro studies of novel pH-sensitive hydrogels for the oral delivery of calcitonin. *Macromolecules* **1999**, *32*, 6646–6651.
123. Shantha, K.L.; Harding, D.R.K. Preparation and in vitro evaluation of poly(*N*-vinyl-2-pyrrolidone–polyethylene glycol diacrylate)–chitosan interpolymeric pH-responsive hydrogels for oral drug delivery. *Int. J. Pharm.* **2000**, *207*, 65–70.
124. Maris, B.; Verheyden, L.; Van Reeth, K.; Samyn, C.; Augustijns, P.; Kinget, R.; Van den Mooter, G. Synthesis and characterisation of inulin-azo hydrogels designed for colon targeting. *Int. J. Pharm.* **2001**, *213*, 143–152.
125. Stubbe, B.; Maris, B.; Van den Mooter, G.; De Smedt, S.C.; Demeester, J. The in vitro evaluation of “azo containing polysaccharide gels” for colon delivery. *J. Control. Release* **2001**, *75*, 103–114.
126. Blanco, M.D.; García, O.; Gómez, C.; Sastre, R.L.; Teijón, J.M. In vivo drug delivery of 5-fluorouracil using poly-(2-hydroxyethyl methacrylate-*co*-acrylamide) hydrogels. *J. Pharm. Pharmacol.* **2000**, *52*, 1319–1325.
127. Blanco, M.D.; Trigo, R.M.; Teijón, C.; Gómez, C.; Teijón, J.M. Slow releasing of ara-C from poly(2-hydroxyethyl methacrylate) and poly(2-hydroxyethyl methacrylate-*co*-*N*-vinyl-2-pyrrolidone) hydrogels implanted subcutaneously in the back of rats. *Biomaterials* **1998**, *19*, 861–869.
128. Alvarez-Figueroa, M.J.; Blamco-Méndez, J. Transdermal delivery of methotrexate: Iontophoretic delivery from hydrogels and passive delivery from microemulsions. *Int. J. Pharm.* **2001**, *215*, 57–65.
129. Wang, Y.Y.; Hong, C.T.; Chiu, W.T.; Fang, J.Y. In vitro and in vivo evaluations of topically applied capsaicin and nonivamide from hydrogels. *Int. J. Pharm.* **2001**, *224*, 89–104.
130. Thacharodi, D.; Rao, K.P. Development and in vitro evaluation of chitosan-based transdermal drug delivery systems for the controlled delivery of propranolol hydrochloride. *Biomaterials* **1995**, *16*, 145–148.

HOT-MELT EXTRUSION TECHNOLOGY

Michael Repka

The University of Mississippi, University, Mississippi, U.S.A.

John J. Koleng, Jr.

Feng Zhang

Jim W. McGinity

The University of Texas at Austin, Austin, Texas, U.S.A.

INTRODUCTION

Hot-melt extrusion is one of the most widely applied processing techniques in the plastics industry. Joseph Brama invented the extrusion process for the manufacturing of lead pipes at the end of the eighteenth century. However, hot-melt extrusion was not applied in the plastics' industry until the mid-nineteenth century, when it was first introduced into a wire insulation polymer coating process. Today, hot-melt extrusion is not only widely applied in the production of polymeric articles but also in polymer production and compounding. Currently, more than half of all plastic products, including plastic bags, sheets, and pipes, are manufactured by this process (1).

For pharmaceutical systems, several research groups have recently demonstrated that the hot-melt extrusion technique is a viable method to prepare granules, sustained release tablets, and transdermal and transmucosal drug delivery systems. Molten thermoplastic polymers during the extrusion process can function as "thermal binders and/or drug release retardants" once they exit the extruder and solidify. For film processing, a polymer can be shaped into a film via the heating process rather than through the traditional solvent-cast technique.

For pharmaceutical applications, hot-melt extrusion offers many advantages over traditional processing techniques. The advantages are as follows: 1) neither solvents nor water are used in this process; 2) fewer processing steps are needed, and, thus, time-consuming drying steps are eliminated; 3) there are no requirements on the compressibility of the active ingredients, and the entire procedure is simple, continuous, and efficient; 4) the intense mixing and agitation during processing cause suspended drug particles to de-aggregate in the molten polymer, resulting in a more uniform dispersion of fine particles; and 5) the bioavailability of the drug substance could be improved when it is dispersed at the molecular level in hot-melt extruded dosage forms. To produce the

pharmaceutical dosage forms via hot-melt extrusion, a pharmaceutical grade polymer must be selected that can be processed at a relatively low temperature due to the thermal sensitivity of most drugs. All components must be thermally stable at the processing temperature during the short duration of the heating process.

PROCESS AND EQUIPMENT

Hot-melt extrusion equipment consists of an extruder, downstream auxiliary equipment, and other monitoring tools used for performance and product quality evaluation (2). The extruder is typically composed of a feeding hopper, barrel, screw, die, screw driving unit, and a heating/cooling device. A diagram of a typical extruder is shown in Fig. 1. Downstream equipment is used to collect the extrudates prior to further processing. Monitoring devices on the equipment include temperature gauges, a screw speed controller, an extrusion torque monitor, and pressure gauges.

During the hot-melt extrusion process, different zones of the barrel are preset to specific temperatures before the extrusion process. A blend of the thermoplastic polymers and other processing aids is then fed into the barrel of the extruder through the hopper. The materials are transferred inside the heated barrel by a rotating screw. Temperatures at different sections of the barrel are normally controlled by electrical heating bands, and the temperature is monitored by thermocouples. The materials inside the barrel are heated mainly by the heat generated due to the shearing effect of the rotating screw and the heat conducted from the heated barrel. The molten mass is eventually pumped through the die, which is attached to the end of the barrel. The extrudates are then subject to further processing by auxiliary downstream devices.

During a continuous extrusion process, the feed stock is required to have good flow properties inside the hopper.

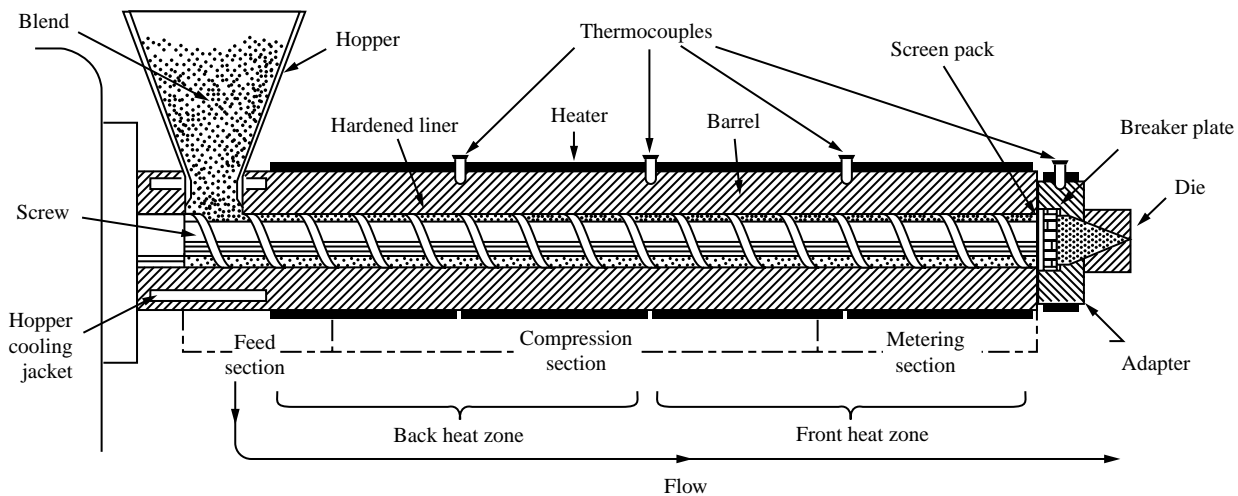


Fig. 1 Schematic diagram of a single-screw extruder. (From Ref. 11.)

For the material to demonstrate good flow, the angle between the side wall of the feeding hopper and a horizontal line needs to be larger than the angle of repose of the feed stock. In the case of cohesive materials or for very fine powders, the feed stock tends to form a solid bridge at the throat of the hopper, resulting in erratic powder flow. For these situations, a force-feeding device is sometime used.

The design of the extrusion screw has a significant influence on the efficiency of the hot-melt extrusion process. The function of the screw is to transfer the material inside the barrel and then to mix, to compress, and to melt the polymeric materials and to pump the molten mass through the die. Several parameters are used to define the geometrical features of the screw (Fig. 2).

Most screws are made from stainless steel that is surface-coated to withstand friction and potential surface erosion and decay that may occur during the extrusion process. Based on the geometrical design and the function of the screw at each section, an extruder is generally divided into three zones: feeding section, melting or compression section, and metering section, as seen in Fig. 1. Only single-screw extruders were used during the early days of this technology. Twin-screw extruders were invented in the late 1930s. The two screws can either rotate in the same direction (co-rotating extruder) or in the opposite direction (counter-rotating screw). Twin-screw extruders possess many advantages when compared to single-screw extruders, such as easier material feed, more intensive mixing, less tendency to overheat the materials, and a shorter residence time.

The purpose of the feeding section is to compact and to transfer the feed stock into the barrel of the machine. The channel depth (Fig. 2) is normally greatest in this section.

The performance of the feeding section depends on the external friction coefficient of the feed stock at the surface of the screw and barrel. The friction at the inner surface of the barrel is the driving force for the material feed, whereas the friction at the surface of the screw restricts the forward motion of the material. A high friction coefficient in the barrel and a low friction coefficient at the screw surface would contribute to a more efficient transfer of the materials in the feed section. Other properties of the feed stock, such as bulk density, particle size, particle shape, and material compactability, can also have an impact on the performance of the feeding section. The transfer of the

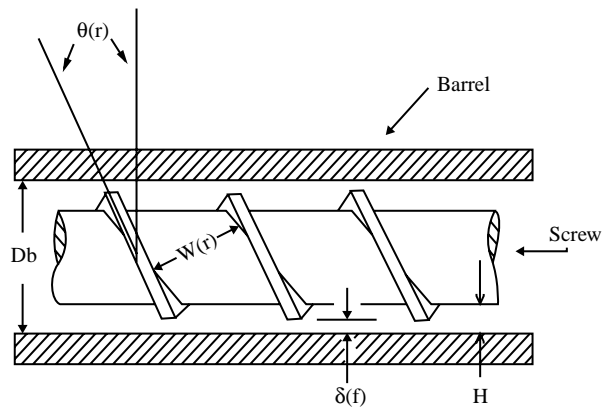


Fig. 2 Geometrical diagram of an extruder screw: 1) diameter of the barrel (D_b): inside diameter of the barrel; 2) channel depth (H): distance from screw roots to barrel inner surface; 3) flight clearance ($\delta(r)$): the distance in between the flight and the barrel inner surface; 4) channel width ($W(r)$): distance in between two neighboring flights; and 5) helix angle ($\theta(r)$): angle formed in between the flight and the direction perpendicular to the screw axis.

material should be efficient in order to maintain an increase in pressure in the compression zone and the metering zone. The pressure rise in these zones should be high enough to provide an efficient output rate of the extrudate. It is also possible to finetune the barrel temperature at the feeding section in order to optimize the friction at the surface of the barrel. Inconsistent material feed may result in a “surge” phenomenon that will cause cyclical variations in the output rate, head pressure, and product quality.

The polymer will begin to melt once the material enters the compression section of the extruder. The temperature of the melting section is normally set at 30–60°C above the glass transition temperature of amorphous polymers or the melting point of a semicrystalline polymer.

Under typical processing conditions, polymer melts behave as pseudoplastic fluids. The viscosity of a pseudoplastic fluid is shear rate dependent and is described by the following power law (Eq. 1):

$$\eta = K \times \dot{\gamma}^{n-1} \quad (1)$$

where η represents the viscosity of the polymer melt; $\dot{\gamma}$ is the shear rate imposed on the polymer; K depends on the properties of the polymer and is an exponential function of the temperature; n is a constant (in the range of 0.25–0.9 for the polymer melt) and depends only on the properties of the polymer.

A minimum temperature of the barrel must be reached in order to extrude the polymer. Otherwise, the shear torque on the screw may overload the capacity of the driving unit due to the high viscosity of the polymer melt. The dependence of polymer melt viscosity on the temperature at a given shear rate follows the Arrhenius equation (Eq. 2):

$$\eta = K' \times \exp(Ea/RT) \quad (2)$$

In Eq. 2, K' is a constant, depending on the structure and the molecular weight of the polymer; Ea is the activation energy of the polymer for the flow process, and it is a constant for the same type of polymer; R is the gas constant; and T is the temperature in degrees Kelvin.

The heat conducted from the barrel contributes to the melting process. However, most of the heat is generated from viscous heat dissipation due to the shearing of the polymer melt. Viscous heat generation is a process of transforming the mechanical energy into the thermal energy of the polymer melt. The rate of heat generation per unit volume due to the viscous heat dissipation is calculated from Eq. 3:

$$\dot{E} = m \times \dot{\gamma}^{n+1} \quad (3)$$

where m is a constant, $\dot{\gamma}$ represents the shear rate, and n is the power law constant (3).

The efficiency of the melting process depends on the properties of the polymer and the geometrical design of the extruder. Generally, the melting process of polymers of low viscosities and high thermal conductivities is a more efficient process. Changes in the screw design are sometimes warranted to improve the melting process. A solidified plastic or polymer component may block the channel if the melting step is insufficient. This could also result in a “surge” of the extrudate.

Thermoplastic polymers primarily exist in a molten state in the metering section. The output rate of the extrudate is highly dependent on the channel depth and the length of the metering section of the screw. The metering section has a shallow channel. Similar to the feeding section, the relative motion between the stationary barrel and the rotating screw metering section results in a velocity vector in the down channel direction, are known as “drag flow.” However, the flow of the polymer melt down in the channel is restricted due to a pressure build-up in the die. This decrease in output rate of the polymer melt due to this pressure is called “pressure flow.” The total flow rate is the difference between the “pressure flow” and the “drag flow” (5).

The die is attached to the end of the barrel. The geometrical design of the die will control the physical shape of the molten extrudate. The cross section of the extrudate increases due to swelling as the molten mass exits the die. Due to its viscoelastic properties, the polymer melt is able to recover some of the deformation imposed by the screw inside the barrel during the extrusion. This is referred to as “die swelling.”

Hot-melt extrusion processing conditions depend on the chemical stability and physical properties of the thermal polymer such as molecular weight, glass transition temperature, and the melting point (in the case of a semicrystalline polymer). During the melt extrusion process, polymers are subjected to a mechanical shear stress imposed by the rotating screw and thermal stress due to the relatively high processing temperature. Under these stresses, polymeric materials may undergo physical chain scission, chemical depolymerization, or thermal degradation. Techniques, such as differential scanning calorimetry and gel permeation chromatography, are widely used to monitor the stability of the polymer under the melt extrusion process. In order to improve the stability of the polymer during the extrusion process, plasticizers, antioxidants, and other additives are often included in the formulation.

Different types of downstream processing equipment are necessary for the hot-melt extrusion process. For extruded film preparations, chill rolls are used to cool

down and control the film temperature before it is taken up by the roller. The thickness of the film can also be controlled by adjusting the rotating speed of the chill rolls. Control of the chill roll temperature also influences the properties of the film. When the extrudate is in a rod shape, it can be chilled through a water bath or a cooling air tunnel and cut into granules by a pelletizer.

MATERIALS USED IN HOT-MELT EXTRUSION

The materials used in the production of hot-melt extruded dosage forms must meet the same levels of purity and safety as those used in traditional dosage forms. Most of the compounds used in the production of hot-melt extruded pharmaceuticals have been used in the production of other solid dosage forms such as tablets, pellets, and transdermals. The materials used in hot-melt extruded products must possess some degree of thermal stability in addition to acceptable physical and chemical stability. The thermal stability of each individual compound and the composite mixture should be sufficient to withstand the production process.

Hot-melt extruded dosage forms are complex mixtures of active medicaments and functional excipients. The functional excipients may be broadly classified as matrix carriers, release modifying agents, bulking agents, and miscellaneous additives. The selection and use of various excipients can impart specific properties to hot-melt extruded pharmaceuticals in a manner similar to those in traditional dosage forms.

The properties of the active drug substance often limit the formulation and preparation options available to the pharmaceutical scientist in the development of an acceptable dosage form. Hot-melt extrusion offers many benefits over traditional processing techniques. This is a relatively new technique to the pharmaceutical industry. The process is anhydrous, thus avoiding any potential drug degradation from hydrolysis following the addition of aqueous or hydroalcoholic granulating media. In addition, poorly compactable materials can be incorporated into tablets produced by cutting an extruded rod, thus eliminating any potential tableting problems seen in traditional compressed dosage forms. As an initial assessment, the thermal, chemical, and physical properties of the drug substance must be characterized. Depending on the unique properties of the drug substance and the other excipients in the formulation, the drug may be present as undissolved particles, a solid solution, or a combination in the final dosage form. The state of the drug in the dosage

form may have a profound impact on the processability and stability of the product.

In addition to thermal degradation, the active compound may interfere with the functionality of the other components in the formulation. Oxprenolol hydrochloride was shown to melt under the hot-melt extrusion processing conditions, thus decreasing the viscosity of the extrudate to yield a material with poor handling properties (6). In similar work preparing dosage forms by injection molding, Cuff and Raouf (7) reported that fenoprofen calcium inhibited the hardening of a PEG–MCC matrix, thus resulting in an unusable product. Lidocaine was shown to effectively lower the T_g of Eudragit® E/HDPE films (8), and hydrocortisone demonstrated a time-dependent lowering of the glass transition temperature of hydroxypropylcellulose (HPC) films (9). These changes may be beneficial, as in the last two examples, or detrimental, as in the first two examples.

As mentioned previously, the drug may be present in one of several forms in the final product. The advantages and disadvantages of each form have been discussed in both injection molding (7) and melt extrusion (10) systems. Solid dispersion systems may be more stable and more easily processed than solid solution systems, but solid solution systems may be produced that are transparent and have increased bioavailability of poorly soluble compounds. Figs. 3 and 4 show the X-ray diffraction patterns for a polymer film containing lidocaine. The absence of the lidocaine peaks in the extruded samples, and the reported decrease in polymer T_g in these systems confirms that the drug exists in a solid solution within the film matrix.

In hot-melt extruded drug delivery systems, the active compound is embedded in a carrier formulation comprised of one or more meltable substances and other functional excipients. The meltable substances may be polymeric materials (6–12) or low melting point waxes (13, 14). The selection of an appropriate carrier compound is important in the formulation and design of a hot-melt extruded dosage form. The properties of the carrier material often dictate the processing conditions necessary for the production of the dosage unit, and the physical and chemical properties of the carrier often modulate the release of the active compound from the final dosage form. Table 1 lists some of the properties of various carrier compounds used in the production of hot-melt extruded dosage forms.

The use of polymeric carriers usually requires the incorporation of a plasticizer into the formulation in order to improve the processing conditions during the manufacturing of the extruded dosage form or to improve the physical and mechanical properties of the final product. The choice of a suitable plasticizer will depend on many

factors, such as plasticizer-polymer compatibility and plasticizer stability. Triacetin (6), citrate esters (8, 9), and low molecular weight polyethylene glycols (6, 9, 12) have been investigated as plasticizers in hot-melt extruded systems. As mentioned previously, certain drug compounds were reported to function as plasticizers in different hot-melt extruded systems (6, 8, 9). The plasticizer functions to reduce the glass transition temperature T_g of the polymer and to reduce the processing temperatures necessary for production. The reduction in polymer T_g is dependent upon the plasticizer type and level. Fig. 5 demonstrates the effectiveness of various plasticizers in lowering the T_g of hydroxypropylcellulose films. The reduction in processing temperatures may improve the stability profile of the active compound (9) and/or the polymeric carrier (6, 12). Plasticizers also reduce the shear forces needed to extrude a polymer, thus improving the processing of certain high molecular weight polymers (9, 12). Table 2 demonstrates the profound influence on the incorporation of PEG 3350 into hot-melt extruded tablets containing PEO (MW 1,000,000) and

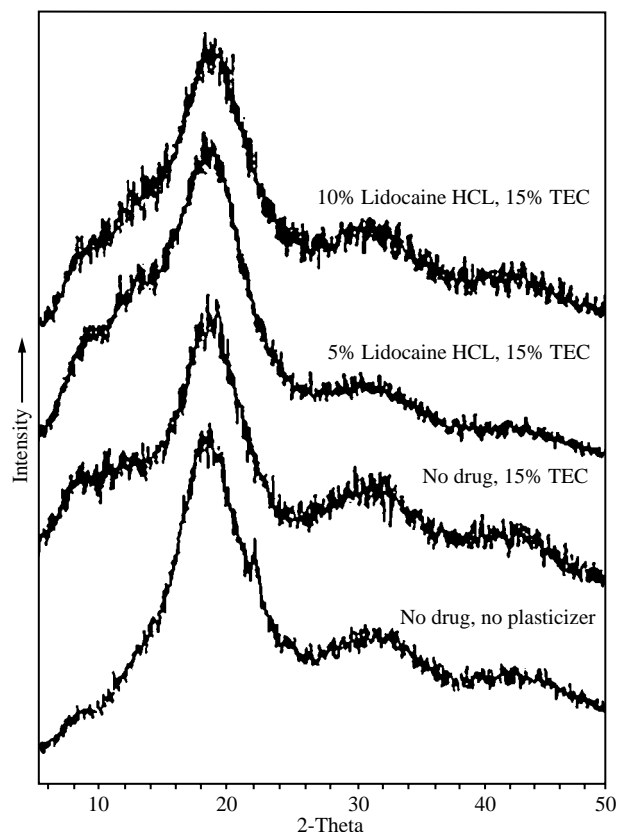


Fig. 3 Wide angle X-ray diffraction scans of extruded Eudragit E films showing the effect of plasticizer and drug on the packing of polymer film. (From Ref. 8.)

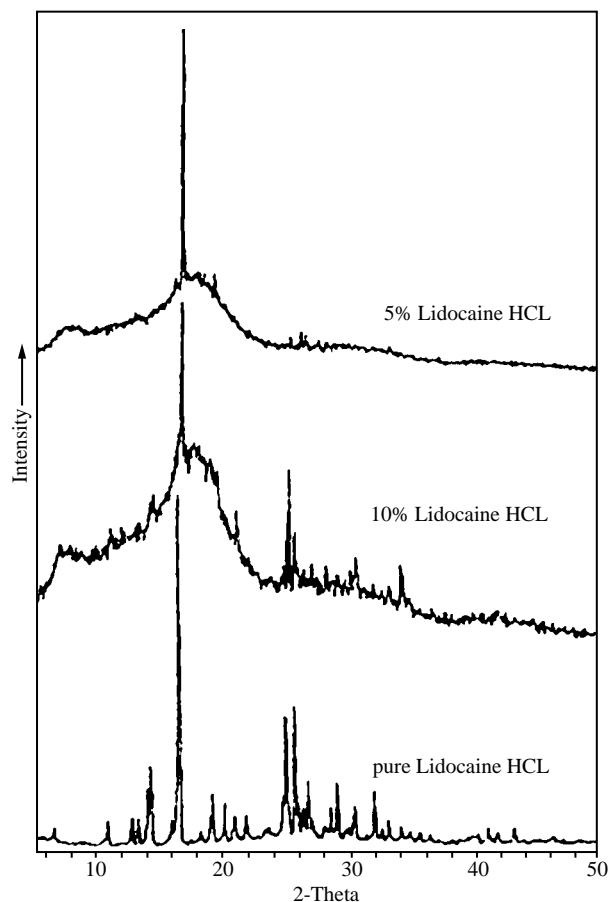


Fig. 4 Wide angle X-ray diffraction scans of a physical blend of Eudragit E100 with 5 and 10% lidocaine HCL compared with lidocaine HCL crystals. (From Ref. 8.)

chlorpheniramine maleate. The incorporation of PEG 3350 into the carrier matrix reduces the temperatures necessary for processing and prevents the degradation of the polymer as noted by an equivalent molecular weight. The stability of a plasticized polymethacrylate system has been reported where the residual monomer content only increased 0.003% following extrusion using triacetin (6). Ethyl cellulose was also reported to be fairly stable following processing (11).

The thermochemical stability and volatility of the plasticizer during processing and storage must also be taken into consideration (9, 15–17). Table 3 illustrates these factors. Repka and McGinity demonstrated that the amount of plasticizer remaining in hot-melt extruded films over time was a function of the plasticizer type. The amount of plasticizer recovered in these films was compared to the initial theoretical amount added in all formulations. HPC films containing polyethylene glycol 400 and 8000 (PEG 400, PEG 8000), triethyl citrate (TEC)

Table 1 Properties of selected carriers used in the design and production of hot-melt extruded dosage forms

Chemical name	Trade name	Molecular weight range (s)	Melting point range (s)	Glass transition temperature	Reference
Polymeric carriers					
Ethylcellulose	Ethocel [®] Aqualon [®] EC	N-7 to N-100 based on viscosity	Decomposes at >190°C	130–133°C	6, 41
Hydroxypropyl cellulose	Klucel [®]	80,000–1,150,000	Chars at 260–275°C	Softens at 130°C	9, 42
Polyethylene glycol	Carbowax [®]	1000–20,000	37–63°C	–17°C for MW 6000	7, 15, 43
Polyethylene oxide	Polyox [®] WSR	100,000–7,000,000	65–80°C	–40 to –60°C	4, 44
Polymethacrylates	Eudragit [®] RSPM Eudragit [®] E	>100,000	—	52°C 40°C	6, 8, 44
Chemical name	Melting point Range	Saponification Value	Reference		
Non-polymeric carriers					
Carnauba Wax	81–86°C	78–95	13, 14, 45		

and acetyltributyl citrate (ATBC) were tested at these two conditions: 25°C/0% RH and 25°C/50% RH for 1-week (initial), 3-month, and 6-month testing intervals. As can be seen from the results, both the citrate esters had a minimal loss at both testing conditions for the duration of the study. The most stable plasticizer for this 6-month study was PEG 8000. However, the lower molecular weight PEG had the greatest loss of any of the other three plasticizers investigated. At the more rigorous testing conditions of 25°C/50% RH, only 63% PEG 400 was detected in the HPC films for the duration of the study. These results indicate the tendency of PEG 400, and potentially other plasticizers, to evaporate or degrade under hot-melt extrusion processing and storage. It is also evident that these findings for PEG 400 explain changes upon mechanical testing of PEG 400-incorporated HPC films (9).

Plasticizers may also be incorporated into hot-melt extruded dosage forms to improve the physical–mechanical properties of the final dosage form. In transdermal films, the addition of a plasticizer to the polymer matrix can improve the film's flexibility (8, 9). Plasticizers may also influence the product's tensile strength and elastic modulus. In addition to the plasticizer's effect on the performance of hot-melt extruded polymer systems, the thermal history of the material may impact the properties of the final dosage form. The properties of high molecular weight polymers often depend upon the processing and storage conditions experienced prior to incorporation into the final dosage forms. Exposure to heat and/or moisture may influence the stability and processing of different batches of polymeric materials. There have been several reports on the influence

of heating and cooling on the physical and chemical properties of PEG-containing systems (16–21).

For systems employing nonpolymeric carrier materials, similar concerns must be addressed. The compatibility between the active substance and carrier should be addressed. The incorporation of a low-melting point compound into a low-melting point wax may result in the formation of a eutectic mixture or a reduction in the melting point of the material that may prevent the formation of a solid dosage form. The production of granules using carnauba wax has been reported (13, 14). The granules contained diclofenac sodium and could be produced at temperatures less than the reported melting point of the wax material. The use of waxes and wax-based materials may be advantageous due to their reported chemical inertness (13).

The drug release rate from hot-melt extruded dosage forms is highly dependent upon the characteristics of the carrier material. Most of the materials reported for use in hot-melt extruded dosage forms are water insoluble (6, 11, 13, 14) or have slow hydrating or gelling rates (9, 12). To improve or to modulate drug release from these systems, functional excipients may be added. Depending upon the physical and chemical properties of these additional excipients, various release profiles may be achieved. Follonier and coworkers (11) investigated a variety of compounds in several polymeric systems using diltiazem hydrochloride as a model compound. The additives were incorporated into the formulation in an effort to increase the drug release rate by increasing the porosity of the pellet during dissolution. Viscosity inducing agents were incorporated

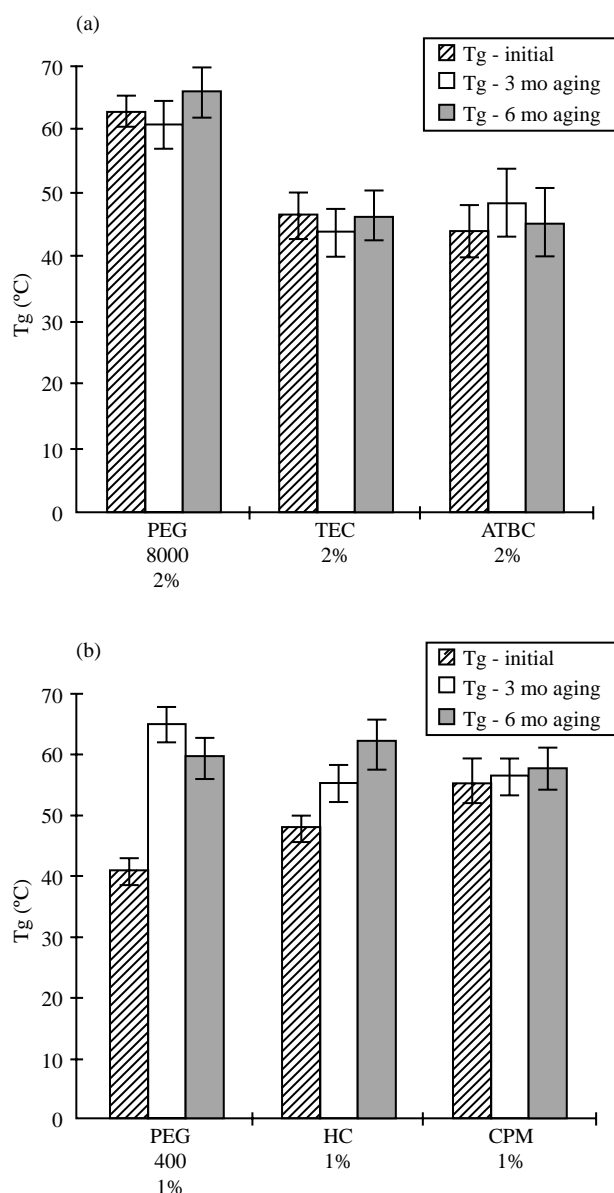


Fig. 5 (a) Glass transition temperatures of HPC films containing PEG 8000 2%, TEC 2%, and ATBC 2% at 1 week, 3 months, and 6 months; (b) Glass transition temperatures of HPC films containing PEG 400 1%, HC 1%, and CPM 1% at 1 week, 3 months, and 6 months. (From Ref. 9.)

in the polymer matrix to limit the burst effect often seen with matrix systems. The use of ionic or pH-dependent polymers in the carrier matrix may allow for zero-order drug release or pH-dependent drug delivery. Swelling agents, such as AcDiSol[®] and Explotab were also investigated as a method to modulate drug release. In contrast, it has been demonstrated that Explotab could be used as a “super-absorbent” in HPC hot-melt

extruded films to facilitate moisture or exudate uptake in wound care applications (22). A similar approach of drug release modification was applied to wax containing systems (13, 14). Hydroxypropylcellulose, Eudragit L, and sodium chloride were incorporated into diclofenac sodium/carnauba wax matrices. Increasing the content of the cellulose derivative or methacrylic acid copolymer resulted in a substantial increase in the release of diclofenac sodium. The release of diclofenac sodium from hydroxypropyl cellulose/wax matrices was less pH dependent than the system containing wax/Eudragit L because the methacrylic acid copolymer is insoluble in water or in solutions with pH < 6. The effect of sodium chloride was less pronounced and was attributed to the negligible swelling effect of this material.

The properties of poly(vinyl acetate) as a carrier for theophylline from matrix dosage forms prepared by hot-melt extrusion has been investigated (23). The influence of granule size and drug loading level on the drug release properties was studied. The thermal stability of poly(vinyl acetate) was also investigated. In this study, Zhang and McGinity ground the rod shaped extrudates and then compressed them into tablets with various combinations of microcrystalline cellulose. The influence of granule size on the release rate of theophylline from the compressed tablets containing the hot-melt extruded granules is seen in Fig. 6. As the size of hot-melt extruded theophylline/PVAc granules was increased, there was a significant decrease in the release rate of the theophylline. Because the drug was released from the matrix by a diffusion mechanism, the decrease in the drug release rate from the tablets containing larger granules was concluded to be a result of a longer diffusion pathway. The influence of theophylline loading on the release properties of these tablets containing extruded granules is shown in Fig. 7. The PVAc was demonstrated to have a high solids carrying capacity when processed by hot-melt extrusion. It is important to note that a powder blend containing 50% theophylline could be readily processed utilizing this hot-melt technique.

These same investigators studied the stability of PVAc to shear stress using a Plasticorder[®] rheometer (23). As shown in Fig. 8, the initial peak indicated the heating process of the added PVAc. The torque gradually decreased as the temperature of the polymer melt in the chamber was increased. No further change in the torque was observed after the initial heating step. This finding confirmed that PVAc was not susceptible to degradation by either thermal or shearing stress under the processing conditions.

Other materials may also be included in the formulation of hot-melt extruded dosage forms. These miscellaneous compounds include embodying agents and antioxidants.

Table 2 Stability of polyethylene oxide (MW 1,000,000) in hot-melt extruded tablet formulations containing 6 wt% chlorpheniramine maleate processed at different temperatures

Formulation	PEO (wt%)	PEG 3350 (wt%)	Zone 1 temperature	Zone 2 temperature	Die temperature	Weight average MW	Standard deviation
1	94	0	100°C	115°C	130°C	710,000	50,000
2	54	40	70°C	75°C	85°C	900,000	100,000
Pure polymer	100	—	—	—	—	920,000	80,000

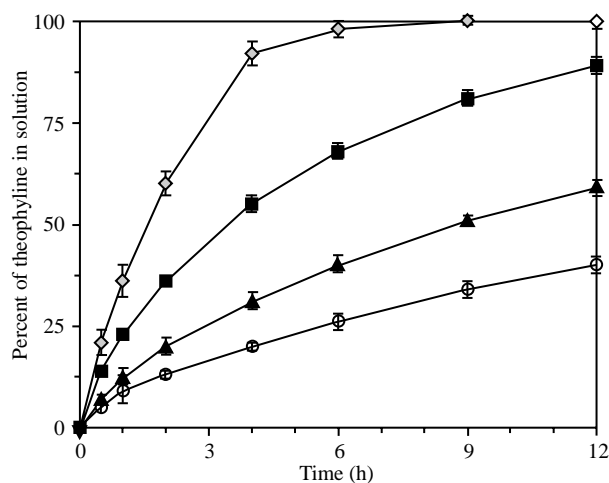
(From Ref. 12.)

Table 3 Stability of plasticizers in hot-melt extruded HPC films processed at 180°C as a function of storage conditions ($n = 6$)

Plasticizer type	% Theoretical (25°C/ 0% RH)			% Theoretical (25°C/ 50% RH)		
	1 wk	3 mon	6 mon	1 wk	3 mon	6 mon
TEC	94.4 (1.83)	93.6 (0.77)	92.3 (1.20)	94.1 (2.21)	93.1 (1.97)	92.8 (2.33)
ATBC	92.8 (2.11)	91.4 (1.65)	90.0 (1.77)	91.5 (1.36)	90.1 (2.87)	88.3 (2.53)
PEG 8000	97.3 (0.38)	97.1 (1.36)	96.6 (0.34)	97.1 (0.56)	96.5 (1.73)	96.6 (1.43)
PEG 400	88.7 (3.15)	79.1 (3.82)	74.3 (2.67)	86.1 (2.18)	66.5 (2.08)	63.0 (1.98)

Standard deviation denoted in parentheses.

(From Ref. 17.)

**Fig. 6** Influence of the granule particle size on the theophylline release properties of the tablets containing hot-melt extruded granules ($n = 3$). Tablets: 20% extruded granules (Formula 3: 25% theophylline, 2% PEG, and 73% PVAc), 79.5% Avicel® PH 200, and 0.5% magnesium stearate (♦) Less than 125 μm ; (■) 180–212 μm ; (▲) 300–425 μm ; (○) 500–600 μm . (From Ref. 23.)

Cuff and Raouf (7) reported the incorporation of microcrystalline cellulose into PEG 8000 matrices in order to improve the formulation viscosity and plasticity of the resulting tablets formed by injection molding. Excessive temperatures needed to process unplasticized or underplasticized cellulose-based polymers (i.e., hydroxypropyl-cellulose or ethyl cellulose) may lead to polymer oxidation. One manufacturer of these materials recommends the incorporation of an antioxidant, such as butylated hydroxytoluene or ascorbic acid, into formulations containing low molecular weight hydroxypropyl-cellulose (24). Similarly, a combination of an antioxidant, light absorber, and acid acceptor is recommended for systems employing ethyl cellulose (25). Poly(ethylene oxide) films have been reported to be protected from free radical and oxidative degradation by the incorporation of an antioxidant (26).

The materials used in the production of hot-melt extruded dosage forms are the same pharmaceutical compounds used in the production of more traditional systems. Thermal stability of the individual compounds is a prerequisite for the process, although the short

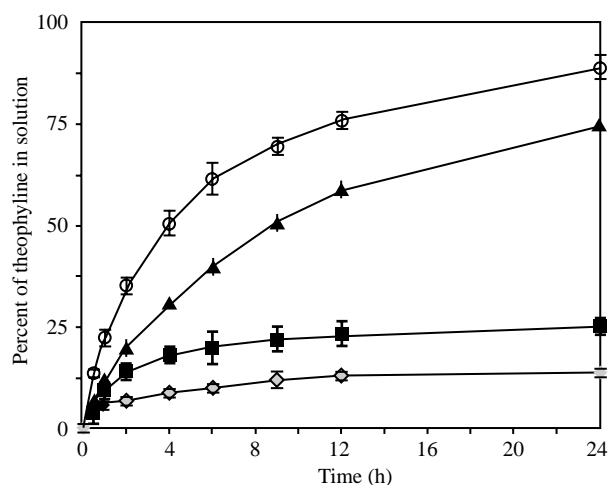


Fig. 7 Influence of drug loading levels on theophylline release properties of the tablets containing hot-melt extruded granules ($n = 3$). Tablets: 20%, 300–425 μm extruded granules (Formule 1–4: theophylline, 2% PEG and PVAc qs to 100%), 79.5% Avicel PH 200, and 0.5% magnesium stearate. (◆) 5% loading; (■) 15% loading; (▲) 25% loading; (○) 50% loading. (From Ref. 23.)

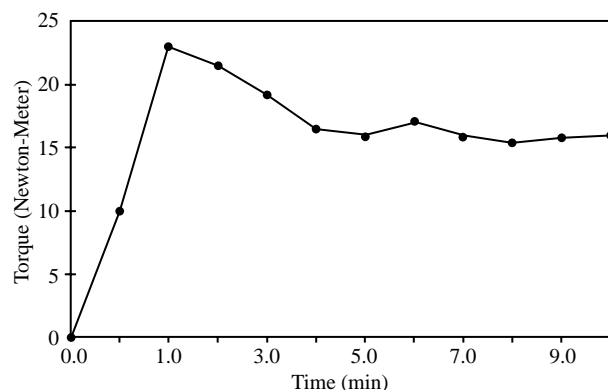


Fig. 8 Heat and shear stability of poly(vinyl acetate) monitored via Plasi-corder rheometer. (From Ref. 23.)

processing times encountered in this process may not limit all thermolabile compounds. The incorporation of plasticizers may lower the processing temperatures encountered in hot-melt extrusion, thus reducing drug and carrier degradation. Drug release from these systems can be modulated by the incorporation of various functional excipients. The dissolution rate of the active compound can be increased or decreased depending on the properties of the rate-modifying agent. For systems that display oxidative or free radical degradation during processing or storage, the addition of antioxidants, acid acceptors, and/or light absorbers may be warranted.

APPLICATIONS

For over two decades, the value of “continuous processing” in the pharmaceutical industry has been recognized. The potential of automation, reduction of capital investment, and the reduction in labor costs has made hot-melt extrusion worthy of consideration (27).

Conventional extrusion/spheronization is a significant process in obtaining controlled release pellets, as are solution/suspension techniques (28, 29). The quality of these pellets or granules is important, as was recognized by Gamlen et al. (10) and Lindberg et al. in the mid-1980s (30, 31). However, control of porosity, content uniformity, consistent pellet size distribution, as well as achieving a true continuous process, were not easily attained. Pellet technology has advanced in the 1990s with the advent of new processing equipment. Such developments have given the pharmaceutical scientist numerous opportunities to apply scientific principles to the design of novel dosage forms.

Until recently, hot-melt extrusion had not received much attention in the pharmaceutical literature. Pellets comprising cellulose acetate phthalate were prepared using a rudimentary ram extruder in 1969 and studied for dissolution rates in relation to pellet geometry (32). More recently, production of matrices based on polyethylene and polycaprolactone were investigated using extruders of laboratory scale (33, 34). Mank et al. reported in 1989 and 1990 on the extrusion of a number of thermoplastic polymers to produce sustained release pellets (35, 36). A melt-extrusion process for manufacturing matrix drug delivery systems was reported by Sprockel and coworkers (37). As one can see, a review of the pharmaceutical scientific literature does not elucidate many applications for hot-melt extrusion in this field.

Follonier and coworkers in 1994 investigated the possibility of using hot-melt extrusion technology to produce sustained-release pellets (6). Fig. 9 exhibits the ram extruder used in their investigations. Again, it was the researchers' goal to provide a product in a simple, single, and continuous manner. Thermal degradation was recognized as a limitation of this hot-melt process. Diltiazem hydrochloride, a relatively stable, freely soluble drug, was incorporated into their polymer-based pellets for sustained-release capsules. Prior to formulation, polymers and plasticizers were selected to optimize the possibility of a successful product. In this report, ethyl cellulose (EC), cellulose acetate butyrate (CAB), poly(ethyl acrylate/methyl-methacrylate/trimethyl ammonio ethyl methacrylate chloride) (Eudragit RSPM), and poly(ethylene-co-vinyl acetate) (EVAC) were the polymers utilized. Plasticizers used included triacetin and

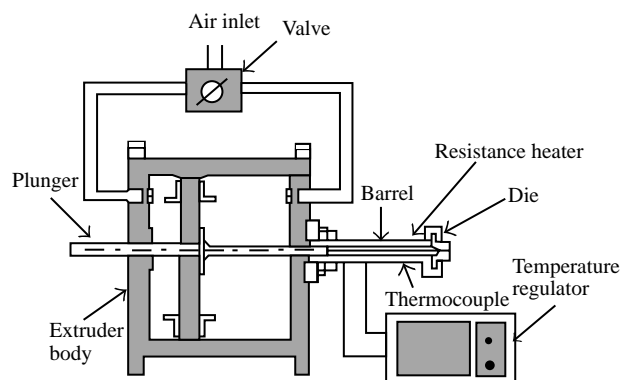


Fig. 9 Schematic diagram of a laboratory ram extruder. (From Ref. 6.)

diethyl phthalate. The porosity of the formulations was assessed by mercury porosimetry. Pellets that were produced exhibited a rather smooth surface and low porosity, as demonstrated in Fig. 10. The *in vitro* release of diltiazem was biphasic, with the CAB and EVAC pellets giving the slowest release rate, as seen in Fig. 11, 12. They also found that the stability of Eudragit RSPM was adequate for extrusion at a temperature of 130°C. Not surprisingly, the type and percent of plasticizer used, drying time of the polymers, extrusion temperatures, and plasticization times varied with each formulation. In a latter study, Follonier et al. examined different parameters influencing the release of diltiazem hydrochloride from hot-melt extruded pellets (11). These parameters included polymer type, drug/polymer ratio, and pellet size. The authors also incorporated various polymer excipients into the pellet formulations to vary the drug release rate, such as croscarmellose sodium (Ac-Di-Sol®) and sodium starch glycolate (Explotab). These pellets could be applicable for incorporation into hard gelatin capsules. With optimization of techniques and formulations, it is apparent that hot-melt extrusion of these and other sustained-release pellets is a viable drug delivery technology.

Currently, the most frequent means of producing thin films for transdermal/transmucosal drug delivery and wound care is via film casting from organic or aqueous solvents (8). However, it is recognized that there are numerous problems with these types of films. For example, Gutierrez-Rocca and McGinity showed that physical aging of both aqueous- and solvent-cast acrylic films resulted in a decrease in elongation or elasticity and an increase in the tensile strength. The changes in the mechanical properties were related to a relaxation of the polymer chains toward a state of equilibrium (16). Also, it has been demonstrated that the type and level of

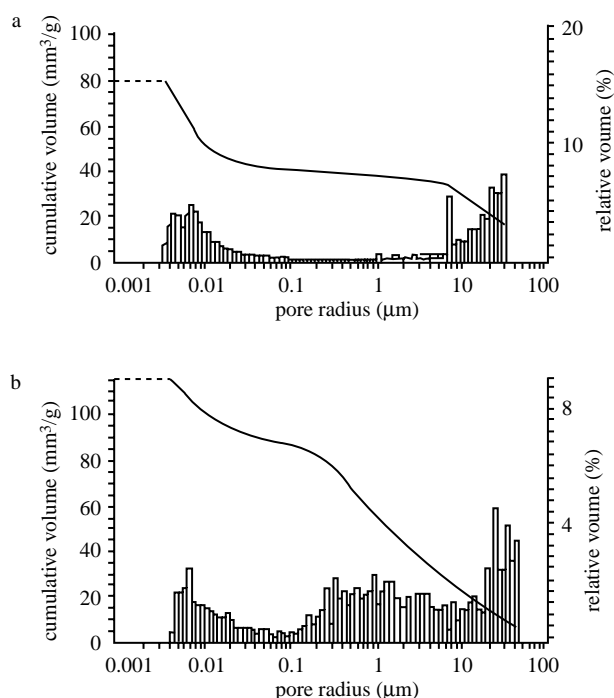


Fig. 10 Mercury infusion curves and pore size distributions of Eudragit RSPM (a) and Elvaxr 40W (b) based pellets containing diltiazem hydrochloride. (From Ref. 6.)

plasticizers, curing time, and temperatures will have a significant effect on the dissolution rate of drugs from films formed from aqueous dispersions (38–40).

Aitken-Nichol and coworkers investigated the viability of hot-melt extrusion technology in 1996 for the production of thin, flexible, acrylic films for topical drug

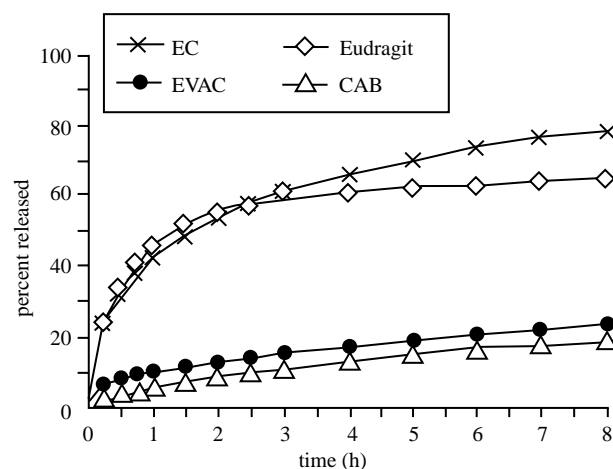


Fig. 11 Release profiles of diltiazem hydrochloride from extruded pellets based on various polymers (polymer/drug ratio: 1:1; size: 2 × 2 mm). (From Ref. 6.)

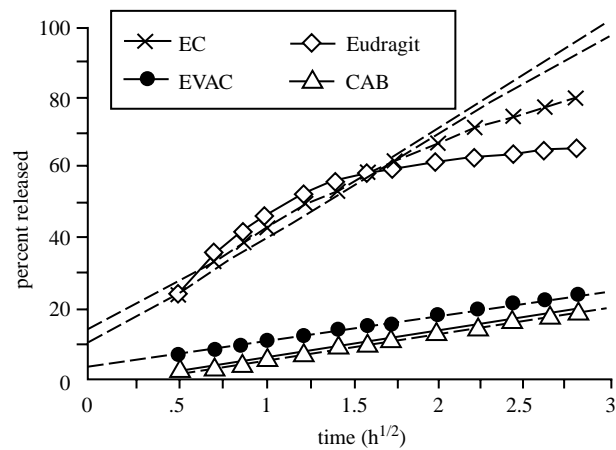


Fig. 12 Diltiazem hydrochloride released (%) as a function of the square root of time and calculated regression lines (—). (From Ref. 6.)

delivery (8). One of the advantages that the researchers pointed out was that the manufacturing process is not restricted by solvent concerns. Although a truly continuous operating mode was not utilized in this investigation, the potential for one definitely exists. The investigators compared a cast film with various extruded films. The data in Table 4 illustrate the effects of types and levels of plasticizers and model drugs used in this study on the glass transition temperature (T_g) and the mechanical properties of high density polyethylene (HDPE) and Eudragit E-100 extruded films. Eudragit E-100 was the primary thermoplastic polymer extruded. The authors found that

hot-melt extrusion was a viable technology for the production of free films of this acrylic resin. Although triethyl citrate was an acceptable plasticizer for this polymer, these researchers showed that lidocaine HCl was also able to plasticize the acrylic films. The authors concluded that the differences in the dissolution and ductile properties between cast films and extruded films were due to the amount of drug dissolved in the polymer. The dissolution rate of lidocaine HCl was affected by the drug loading, in contrast to the solvent-cast films tested.

Transdermal and transmucosal drug delivery systems are frequently produced by films cast from organic or aqueous solvents. Repka and coworkers discussed the numerous disadvantages accompanying these techniques, including long processing times, environmental concerns, and excessive costs (9). Hot-melt extrusion technology was used by these researches to produce HPC films utilizing a Killion extruder. Various plasticizers and two model drugs were incorporated into the HPC films. The influence of these plasticizers and drugs on the physical–mechanical properties of the films was investigated. The authors observed that a pure HPC film could not be produced without the incorporation of a plasticizer due to the high stress exhibited in the extruder. The results in Table 5 illustrate the effects of a number of conventional plasticizers and two model drugs on the three mechanical properties of HPC films. With the exception of PEG 400, all plasticizers investigated demonstrated adequate stability for the duration of the study. PEG 400, although initially exhibiting excellent plasticizer qualities for the HPC films, was found to be unstable in all parameters

Table 4 Mechanical properties of Eudragit E100 films containing diphenhydramine HCL (DPH) and lidocaine HCL (L-HCL)

Polymer	Plasticizer	Drug	T_g (°C)	d-spacing (Å)	Peak stress(c) σ (kg/cm ²)	Elongation at break ϵ (%)
E100	none	none	40	4.76	N/A	N/A
E100-Ex	15% TEC	none	18	4.92	13.4	59.3
E100-Ex	12% triacetin	none	25	4.76	29	47.9
E100-Ex	15% TEC	5% DPH	20	N/A	12.9	53.5
E100-Cast	15% TEC	5% L-HCL	20	5.03	3.65	549
E100-Ex	15% TEC	5% L-HCL	21	4.79	9.88	218
E100-Ex	15% TEC	10% L-HCL	10.5	4.80	2.47	376.8
HDPE-E100	none	5% L-HCL	35 ^a	4.76 ^b	77.7 ^c	110.0 ^c
1:1—Ex					3.0 ^d	3.0 ^d
Ex—Extruded						

^a T_{neck} at 111°C
^bCrystalline peaks at 4.1, 3.60, and 2.49-Å.
^cTested in the direction of orientation.
^dTested perpendicular to orientation.
(From Ref. 8.)

Table 5 Influence of plasticizers and drugs on tensile strength (TS), percent elongation (%E), and Young's modulus (YM) of HPC extruded films stored at 25°C

	TS (initial)	TS (3 mon)	TS (6 mon)
PEG 8000 2%	13.7 (1.1)	13.2 (1.3)	12.2 (0.7)
TEC 2%	17.2 (1.7)	18.9 (1.1)	20.8 (0.7)
ATBC 2%	26.1 (2.6)	19.2 (1.5)	19.6 (1.4)
PEG 400 1%	37.6 (3.5)	29.9 (2.7)	27.9 (2.1)
HC 1%	26.7 (2.7)	33.0 (2.7)	34.1 (3.8)
CPM 1%	32.7 (3.4)	32.9 (3.1)	30.8 (1.8)
	%E (initial)	%E (3 mon)	%E (6 mon)
PEG 8000 2%	5.01 (0.4)	4.45 (0.6)	4.39 (0.7)
TEC 2%	5.29 (0.4)	5.37 (0.4)	5.08 (0.6)
ATBC 2%	6.02 (0.6)	6.13 (0.5)	6.40 (0.7)
PEG 400 1%	6.62 (0.6)	5.25 (0.6)	5.05 (0.5)
HC 1%	5.40 (0.4)	4.82 (0.4)	4.55 (0.4)
CPM 1%	5.26 (0.4)	5.03 (0.7)	4.87 (0.7)
	YM (initial)	YM (3 mon)	YM (6 mon)
PEG 8000 2%	4.25 (0.2)	4.31 (0.6)	4.21 (0.4)
TEC 2%	4.43 (0.3)	4.11 (0.4)	4.09 (0.4)
ATBC 2%	4.75 (0.6)	3.09 (0.5)	3.19 (0.4)
PEG 400 1%	4.05 (0.3)	7.13 (0.3)	6.15 (0.6)
HC 1%	6.57 (0.3)	7.98 (0.3)	8.60 (0.3)
CPM 1%	5.38 (0.3)	4.44 (0.4)	4.64 (0.3)

Standard Deviations denoted in parenthesis; $n = 6$.
(From Ref. 9.)

tested. The influence of processing temperature and storage time on the two model drugs was investigated and is outlined in Table 6. Besides the fact that CPM proved to be an excellent plasticizer for HPC, providing

mechanically stability for the hot-melt extruded film, it also proved to be chemically stable for up to 12 months. In addition, it was demonstrated by differential scanning calorimetry that CPM was fully dissolved in the HPC film

Table 6 Influence of processing temperature and storage at 25°C on percentage of chlorpheniramine maleate (CPM) and hydrocortisone (HC) remaining in HPC extruded films ($n = 6$)

(°C)	CPM (1 wk)	CPM (6 mon)	CPM (12 mon)
170	98.6 (1.6)	98.5 (1.9)	98.4 (1.9)
180	98.5 (1.9)	98.5 (2.1)	98.1 (2.4)
190	98.1 (2.2)	97.6 (2.7)	97.3 (2.5)
200	97.9 (2.3)	97.7 (2.2)	97.6 (2.3)
(°C)	HC (1 wk)	HC (6 mon)	HC (12 mon)
170	93.9 (2.3)	92.4 (3.0)	91.4 (2.4)
180	87.7 (2.2)	83.1 (2.8)	79.9 (2.2)
190	75.9 (3.2)	70.7 (2.7)	71.6 (2.4)
200	70.2 (3.8)	68.8 (3.3)	62.9 (3.1)

Standard deviations denoted in parenthesis.
(From Ref. 9.)

up to the 10% level (Fig. 13) (17). As can be seen, CPM is amorphous within the HPC film thus, a solid solution exists. Hydrocortisone was shown to be a good plasticizer comparable to that of the conventional plasticizers studied however, the chemical stability of HC incorporated into the HPC films was demonstrated to be a function of processing temperature and residence time in the extruder.

Repka et al. (9) also found an inverse relationship during film testing. All extruded films exhibited a marked decrease in tensile strength, in contrast to a large increase in percent elongation, when testing was performed perpendicular to flow vs. in-the-direction of flow, as seen in Fig. 14. This is in contrast to the findings of Aitken-Nichol et al. (8). These researchers reported a 25- to 35-fold increase in both tensile strength and percent elongation in a high density polyethylene/Eudragit E100 (50:50 polymer ratio containing 5% lidocaine HCl) extruded film when tested in the direction of orientation versus perpendicular to orientation (15). The differences between the findings in the two studies, however, may be explained by the poor compatibility of the two polymers. Such mechanical property differences illustrate the importance of "flow orientation" when designing delivery systems or wound care applications utilizing extruded films.

Vitamin E TPGS NF (D- α -tocopheryl polyethylene glycol 1000 succinate) has been utilized for numerous applications in pharmaceutical dosage forms. Its chemical structure contains both a lipophilic and a hydrophilic moiety, making it similar to a conventional surface-active agent. Due to TPGS's unique properties as a solubilizer,

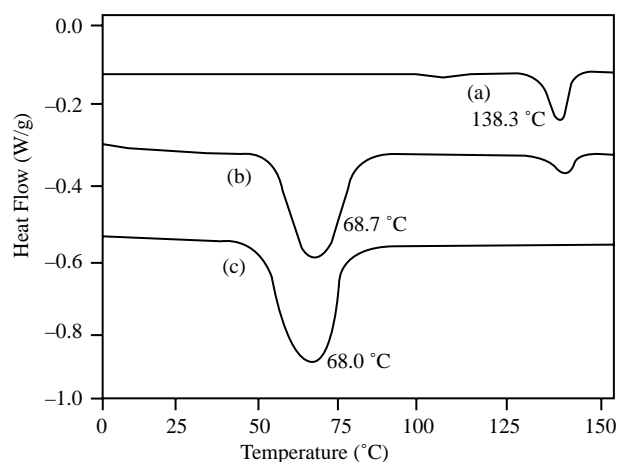


Fig. 13 Differential scanning calorimetry profiles of hot-melt extruded films containing chlorpheniramine maleate and hydroxypropylcellulose: (a) chlorpheniramine maleate (CPM); (b) 10% CPM and 90% HPC (Klucel HF) physical mix; and (c) 10% CPM and 90% HPC (Klucel HF) extruded film. (From Ref. 17.)

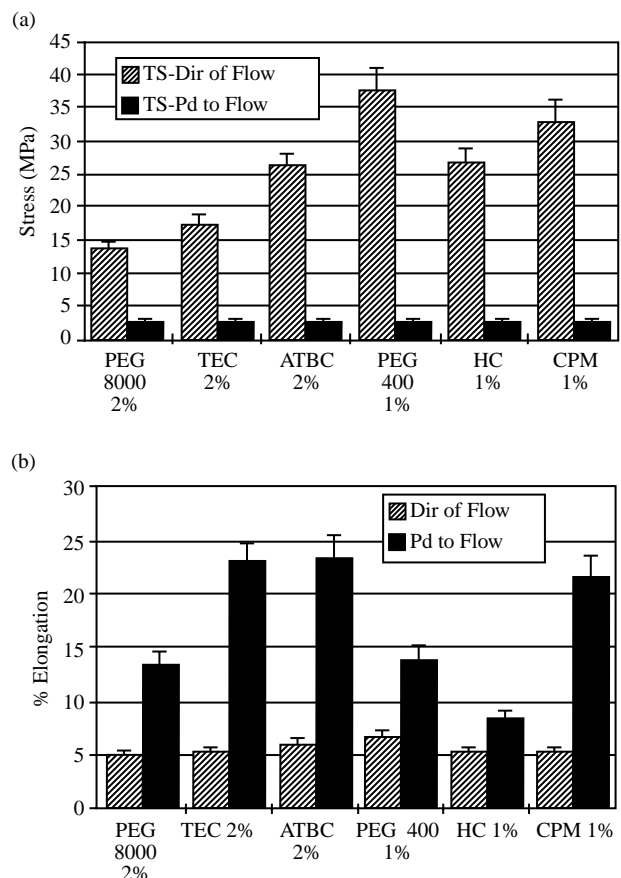


Fig. 14 (a) Tensile strength of HPC films containing various plasticizers and drugs tested in direction of flow and perpendicular to flow; (b) Percent elongation of HPC films containing various plasticizers and drugs tested in direction of flow and perpendicular to flow. (From Ref. 9.)

absorption enhancer, and a potential controlled drug release vehicle, transdermal and transmucosal applications have been shown to be possible via hot-melt extrusion technology (41). Repka and McGinity prepared films containing hydroxypropylcellulose and polyethylene oxide (PEO) using a Randcastle extruder (Model # 750) with and without Vitamin E TPGS as an additive. As can be seen from Fig. 15, the addition of 1, 3, and 5% TPGS, respectively, decreased the glass transition temperature of the extruded films containing either a 50:50 or 80:20 ratio of HPC to PEO in an almost linear fashion. The presence of 3% Vitamin E TPGS lowered the T_g over 11°C when compared to the HPC/PEO 50:50 blend film without TPGS, thus functioning as a plasticizer. The films containing 3% Vitamin E TPGS had similar mechanical properties to that of the films containing 3% PEG 400 and a three-fold increase in percent elongation when compared to the films containing TEC 3% and ATBC 3%. In addition,

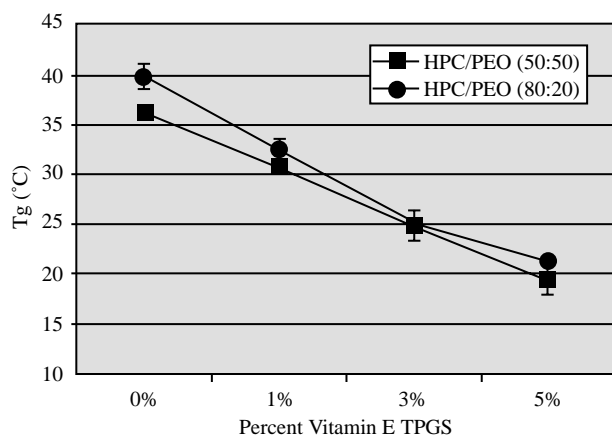


Fig. 15 Glass transition temperatures of films containing different levels of Vitamin E TPGS incorporated into two formulations of HPC/PEO hot-melt extruded films ($n = 4$). (From Ref. 41.)

the Vitamin E TPGS facilitated the processing of the HPC/PEO films by decreasing the barrel pressure, drive amps, and torque of the extruder equipment (Table 7). The unique properties of Vitamin E TPGS can function to promote more applications and opportunities in wound care and in transdermal and transmucosal drug delivery.

Another application of hot-melt extrusion was described by Zhang and McGinity (12). These researchers investigated the properties of polyethylene oxide (PEO) as a drug carrier and studied the release mechanism of chlorpheniramine maleate (CPM) from matrix tablets. In these extruded tablets, PEG 3350 was included as a plasticizer to facilitate processing. In this study, the stability of the primary extruded polymer, PEO, as a function of processing temperature was reported (Table 2). Again, polymer type, temperature, and residence time in the extruder was shown to be of great importance. These authors also reported that the drug, polymer, and other ingredients must be stable at the elevated processing

temperature during the approximately 2 min that the powder blend is processed through the equipment. In addition, the researchers showed that additional mixing of the components occurred in the barrel of the extruder, because the content uniformity of the extruded tablets was within 99.0–101.0% of the theoretical content. The profiles in Fig. 16 illustrate the influence of PEG 3350 on the release of CPM from the extruded matrix tablets. It can be seen that as the percent of PEG 3350 increases, the release rate of CPM increases. Polyethylene glycol is composed of the same structural unit as PEO but has a lower molecular weight than PEO. Thus, PEG 3350 hydrated and dissolved faster than the PEO. The hydration and dissolution rates of the entire matrix system were, thus, accelerated due to the presence of the plasticizer. The influence of drug loading on the release of CPM is shown in Fig. 17. When the drug content was increased from 6 to 12%, no change in the percentage of drug release with respect to time was observed. There was only a slight increase when the drug loading reached 20%. This study conveys the reproducibility of dissolution data for the tablets produced by hot-melt extrusion.

A bioadhesive hot-melt extruded film for intraoral drug delivery and the processing thereof has been patented (42,43). Applications of these films may be utilized in transmucosal drug delivery or even transdermal systems. The films may be produced separately and layered after extrusion, or in some cases, a multilayered system may be extruded in one continuous process. Currently on the market is an extruded film device that is utilized as a denture adhesive. This system includes thermoplastic polymers that have a bioadhesive quality when the film is wetted. Before application and wetting, however, this thin film may be held in one's hand and shaped or cut. This device is again produced by a one-step, continuous process using hot-melt extrusion technology.

A polymeric film that possesses inherent bioadhesive properties has the added benefit of simplifying the dosage

Table 7 Processing conditions for hot-melt extruded films containing a 50:50 ratio of hydroxypropylcellulose to polyethylene oxide with vitamin E TPGS as an additive ($n = 4$)

Vitamin E TPGS (%)	Melt temperature (°C)	Pressure (psi)	Drive amps	Torque (N.m)	Screw speed
0	180	>3000	>4.00	overload	25
0	190	2100	3.84	40	40
1	180	1800	3.06	33	40
3	180	1500	2.61	28	40
5	180	1100	2.07	21	40

(From Ref. 17.)

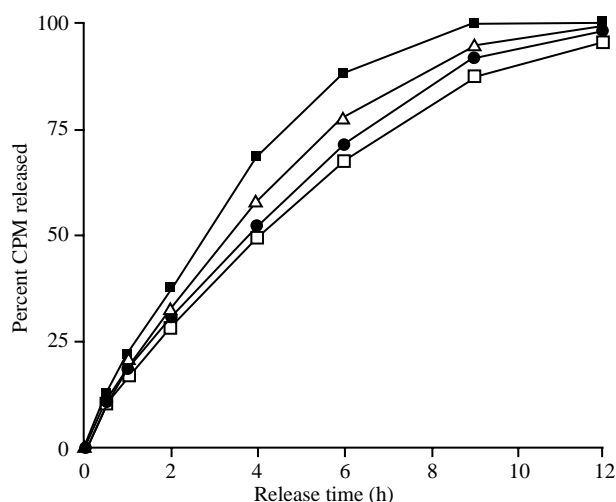


Fig. 16 Influence of polyethylene glycol (3350) on the release of chlorpheniramine maleate from matrix tablets using USP method II at 37°C and 100 rpm in 900 ml purified water: □, 6% CPM, 0% PEG (3,350) and 94% PEO (1.0 m); ●, 6% CPM, 0% PEG (3,350) and 88% PEO (1.0 m); △, 6% CPM, 20% PEG (3,350) and 74% PEO (1.0 m); and ■, 6% CPM, 40% PEG (3,350) and 54% PEO (1.0 m). (From Ref. 12.)

form design and reducing the preparation cost, due to the elimination of the adhesive layer in the system. It is desirable for the film to have adequate adhesion strength so that desirable retention at the application site can be

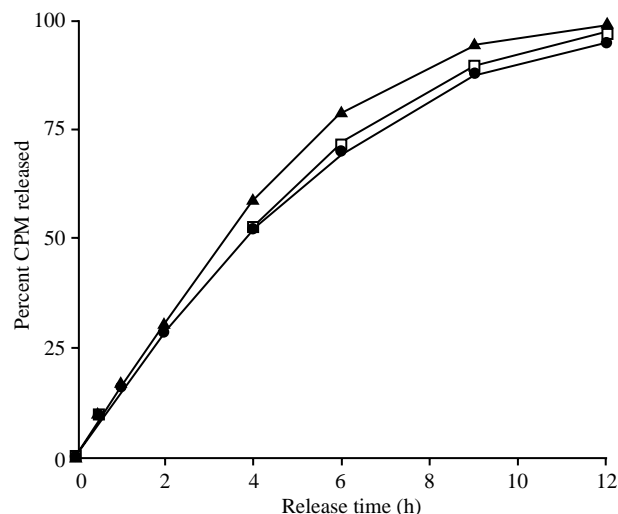


Fig. 17 Influence of drug loading on the release of chlorpheniramine maleate from matrix tablets using USP method II at 37°C and 100 rpm in 900 ml purified water during the first 6 h: ●, 6% CPM, 94% PEO (1.0 m); □, 12% CPM, 88% PEO (1.0 m); and ▲, 20% CPM, 80% PEO (1.0 m). (From Ref. 12.)

achieved. Bioadhesion, however, has been a difficult phenomenon to measure. The test method itself can be problematic as well as the specific property being measured. Physical-mechanical And bioadhesive characteristics are important parameters in the product development of films, for transdermal, transmucosal, and wound care applications. Repka and McGinity conducted bioadhesion testing of hot-melt extruded HPC films, with various additives in human subjects, utilizing a Chatillon testing apparatus (17, 22). Fig. 18 shows a schematic of the equipment used for this testing. Also illustrated is an example of a force deflection profile obtained from bioadhesion experiments in this study (Fig. 19). These researchers found that the force of adhesion, elongation at adhesive failure, and modulus of adhesion are a function of the type of additive in the extruded film. Fig. 20 shows the force of adhesion of the hot-melt extruded films tested. It can be seen that the force of adhesion was highest for the films containing carbomer 971P (Carbopol® 971P) and a polycarbophil (Noveon AA-1®). This study demonstrates that a single layer HPC film may be produced with the bioadhesive incorporated into the matrix, thus eliminating a separate “adhesive layer” and simplifying the process of transdermal and transmucosal delivery systems.

Miyagawa, Sato, and coworkers studied the controlled-release and mechanism of release of diclofenac in studies conducted in 1996 and 1997 (13, 14). These researchers utilized a twin-screw compounding extruder to prepare wax matrix granules composed of carnauba wax, the model drug, and other rate controlling agents. Their first investigation showed that a wax matrix with high mechanical strength could be obtained even at temperatures lower than the melting point of the wax. Dissolution release profiles of diclofenac from the wax matrix granules were strongly influenced by the formulation of the

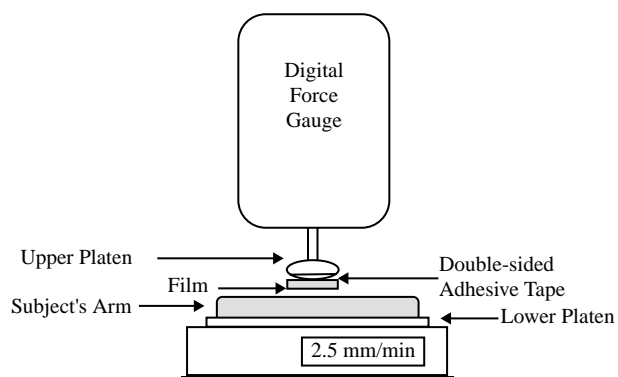


Fig. 18 Schematic of the Chatillon apparatus used to perform butt bioadhesion experiments in vivo in human subjects. (From Ref. 22.)

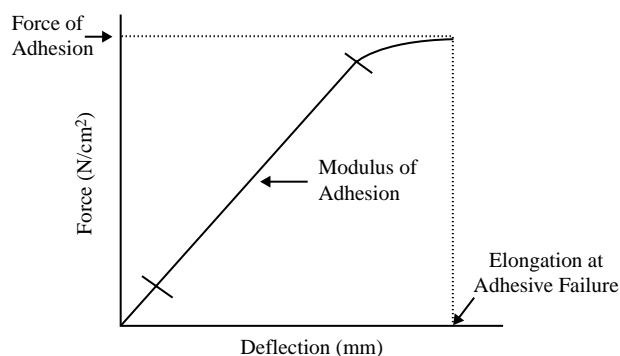


Fig. 19 Example of a force-deflection profile obtained from a butt bioadhesion experiment utilizing hot-melt extruded films in vivo using a Chatillon digital force gauge attached to a motorized test stand. (From Ref. 22.)

granules. The rate-controlling additives that were varied in the formulations included hydroxypropylcellulose, methacrylic acid copolymer (Eudragit L-100), and sodium chloride. The authors emphasized the advantages of using the twin-screw extruder for wax matrix tablets, such as the utilization of low temperatures, high kneading and dispersing ability, and low residence time of materials in the extruder. The investigators concluded in the second study (14) that the selection of rate-controlling agents based on their physicochemical properties, i.e., solubility and swelling characteristics, had a significant impact on the properties of wax matrix granules prepared from this extrusion process.

Koleng and McGinity (44) utilized hot-melt extrusion technology for the preparation of rapid release granules. In this investigation, a hot-melt extrusion process was used to granulate acetaminophen and filler excipients with low molecular weight poly(ethylene glycol)s. The resultant granules were then combined with additional excipients (disintegrant and lubricant) and compressed into tablet

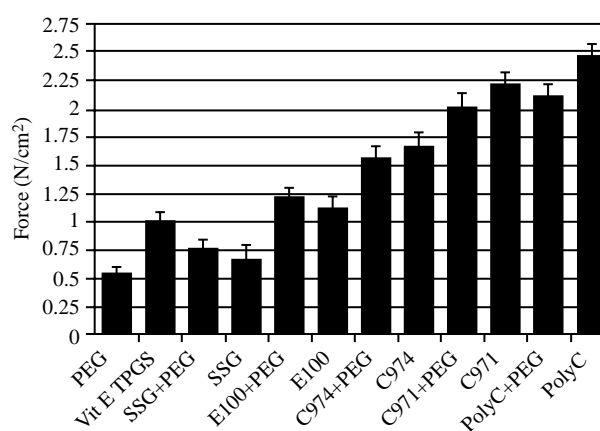


Fig. 20 Force of adhesion of hydroxypropylcellulose hot-melt extruded films containing various polymer additives (12 subjects, $n = 6$). (From Ref. 17.)

compacts. Table 8 lists the release characteristics of the bulk granules and tablets produced from hot-melt extruded granules for formulations containing 15–25% polyethylene glycol 6000. The granules displayed improved drug release compared to the tablets. Tablets containing 15% poly(ethylene glycol) released greater than 80% of the incorporated acetaminophen after 30 min, as required for acetaminophen tablets in the USP 23.

SUMMARY

Although a relatively new technology in the pharmaceutical industry, hot-melt extrusion has been visualized and employed to revolutionize the production of many different dosage forms and systems. It has demonstrated to be applicable to immediate release and sustained release dosage forms including granules, pellets, and tablets. It has

Table 8 Release of acetaminophen from hot-melt extruded granules and tablets prepared from hot-melt extruded granules containing poly(ethylene glycol) 6000 as a thermal binder in 900 ml of 50 mM phosphate buffer (pH 5.8) at 37°C and a paddle speed of 50 rpm

Time (min)	15% PEG 6000		20% PEG 6000		25% PEG 6000	
	Granules (%)	Tablets (%)	Granules (%)	Tablets (%)	Granules (%)	Tablets (%)
5	85.0	18.1	77.1	15.5	79.3	17.4
15	95.5	53.0	99.0	45.3	94.2	35.8
30	98.5	91.4	100.0	77.7	100.0	65.5
45	99.2	95.9	—	95.1	—	86.2
60	100.0	100.0	—	100.0	—	97.8

also been shown to provide numerous advantages in the production of thin films for both drug delivery and wound care applications. New chemical entities that demonstrate a low bioavailability due to solubility issues are prime candidates for hot-melt technologies. These drugs and pharmaceutical devices encompass both prescription products and over-the-counter medications. Hot-melt extrusion technologies may offer numerous advantages over traditional methods. Shorter and more efficient processing times to the final product, environmental advantages due to elimination of solvents in processing (including the possibility of recycling), and increased efficiency of drug delivery to the patient make hot-melt extrusion an exciting challenge for the pharmaceutical scientist.

REFERENCES

- Kaufman, H.S.; Falcetta, J.J. Introduction to Polymer Science and Technology. *An SPE Textbook*; John Wiley & Sons: New York, 1997.
- Kruder, G.A. *Encyclopedia of Chemical Technology*; John Wiley & Sons: New York, 1996.
- Rauwendaal, C. *Polymer Extrusion*; Hanser/Gardner Publications: Cincinnati OH, 1994.
- Nielsen, L.E. *Polymer Rheology*; Marcel Dekker, Inc.: New York, 1977.
- Tadmor, Z.; Klein, I. *Engineering Principles of Plasticating Extrusion*; Van Nostrand Reinhold Company: New York, 1970.
- Follonier, N.; Doelker, E.; Cole, E.T. Evaluation of Hot-Melt Extrusion as a New Technique for the Production of Polymer-Based Pellets for Sustained Release Capsules Containing High Loading of Freely Soluble Drugs. *Drug Dev. Ind. Pharm.* **1994**, *20* (8), 1323–1339.
- Cuff, G.; Raouf, F. A Preliminary Evaluation of Injection Molding as a Technology to Produce Tablets. *Pharm. Tech.* **June**, **1998**, 97–106.
- Aitken-Nichol, C.; Zhang, F.; McGinity, J.W. Hot-Melt Extrusion of Acrylic Films. *Pharm. Res.* **1996**, *13*, 804–808.
- Repka, M.A.; Gerding, T.G.; Repka, S.L.; McGinity, J.W. Influence of Plasticizers and Drugs on the Physical-Mechanical Properties of Hydroxypropylcellulose Films Prepared by Hot-Melt Extrusion. *Drug Dev. Ind. Pharm.* **1999**, *25* (5), 627–635.
- Grünhagen, H.H.; Müller, O. Melt Extrusion Technology. *Pharm. Manu. Int.* **1995**, *1*, 167–170.
- Follonier, N.; Doelker, E.; Cole, E.T. Various Ways of Modulating the Release of Diltiazem Hydrochloride from Hot-Melt Extruded Sustained Release Pellets Prepared Using Polymeric Materials. *J. Controlled Release* **1995**, *36*, 243–250.
- Zhang, F.; McGinity, J.W. Properties of Sustained Release Tablets Prepared by Hot-Melt Extrusion. *Pharm. Dev. Tech.* **1998**, *4* (2), 241–250.
- Miyagawa, Y.; Okabe, T.; Yamaguchi, Y.; Miyajima, M.; Sunada, H. Controlled-Release of Diclofenac Sodium from Wax Matrix Granule. *Int. J. Pharm.* **1996**, *138*, 215–254.
- Sato, H.; Miyagawa, Y.; Okabe, T. Dissolution Mechanism of Diclofenac Sodium from Wax Matrix Granules. *J. Pharm. Sci.* **1997**, *86*, 929–934.
- Price, J.C. Polyethylene Glycol. *Handbook of Pharmaceutical Excipients*; Wade, A., Weller, P.J., Eds.; American Pharmaceutical Association: Washington DC, 1994; 355–361.
- Gutierrez-Rocca, J.C.; McGinity, J.W. Influence of Aging on the Physical–Mechanical Properties of Acrylic Resin Films Cast from Aqueous Dispersions and Organic Solutions. *Drug Dev. Ind. Pharm.* **1993**, *19*, 315–332.
- Repka, M.A. *Physical–Mechanical and Chemical Properties of Topical Films Produced by Hot-Melt Extrusion*; The University of Texas at Austin: Austin Texas, 2000.
- Lin, C.; Cham, T. Compression Behavior and Tensile Strength of Heat-Treated Polyethylene Glycols. *Int. J. Pharm.* **1995**, *118*, 169–179.
- Craig, D.Q.M.; Newton, J.M. Characterisation of Polyethylene Glycols Using Differential Scanning Calorimetry. *Int. J. Pharm.* **1994**, *74*, 33–41.
- McGinity, J.W.; Maincent, P.; Steinfink, H. Crystallinity and Dissolution Rate of Tolbutamide Solid Dispersions Prepared by the Melt Method. *J. Pharm. Sci.* **1984**, *73*, 1441–1444.
- Fassihi, A.R.; Davies, P.J.; Parker, M.S. Effect of Punch Pressure on the Survival of Fungal Spores During the Preparation of Tablets from Contaminated Raw Materials. *Zbl. Pharm.* **1997**, *116*, 1267–1271.
- Repka, M.A.; McGinity, J.W. Physical–Mechanical, Moisture Absorption and Bioadhesive Properties of Hydroxypropylcellulose Hot-Melt Extruded Films. *Biomaterials* **2000**, *21* (14), 1509–1517.
- Zhang, F.; McGinity, J.W. Properties of Hot-Melt Extruded Theophylline Tablets Containing Poly (vinyl Acetate). *Drug Dev. Ind. Pharm.* **2000**, *29* (6), 938–948.
- Klucel® Hydroxypropylcellulose; Physical and Chemical Properties. *Technical Bulletin*; Hercules, Inc.: Wilmington DE, 1997.
- Aqualon® Ethylcellulose (EC); Physical and Chemical Properties. *Technical Bulletin*; Hercules, Inc.: Wilmington DE, 1996.
- Bailey, F.E.; Koleske, J.V. *Poly(ethylene oxide)*; Academic Press: New York, 1976.
- Gamlen, M. Continuous Extrusion Using a Raker Perkins MP50 (Multipurpose) Extruder. *Drug Dev. Ind. Pharm.* **1986**, *12*, 1701–1713.
- Reynolds, A.; Manf. Chem. **1970**, *41*, 40–44.
- Harris, M.R.; Ghebre-Sellassie, I. Aqueous Polymeric Coating for Modified Release Oral Dosage Forms. *Aqueous Polymeric Coatings For Pharmaceutical Dosage Forms*; McGinity, J.W., Ed.; Marcel Dekker, Inc.: New York, 1997; 79.
- Lindberg, N.O.; Tufvesson, C.; Olbjer, L. Extrusion of an Effervescent Granulation with a Twin Screw Extruder, Baker Perkins MPF 50 D. *Drug Dev. Ind. Pharm.* **1987**, *13*, 1891–1913.
- Lindberg, N.O.; Tufvesson, C.; Holm, P.; Olbjer, L. Extrusion of an Effervescent Granulation with a Twin Screw Extruder, Baker Perkins MPF 50 D. Influence on

- Intragranular Porosity and Liquid Saturation. *Drug Dev. Ind. Pharm.* **1998**, *14*, 1791–1798.
32. Rippie, E.G.; Johnson, J.R. Regulation of Dissolution Rate by Pellet Geometry. *J. Pharm. Sci.* **1969**, *58*, 428–431.
33. Shivanand, P.; Hussain, A.S.; Sprockel, D.L. Factors Affecting Release of KCl from Melt Extruded Polyethylene Disks. *Pharm. Res.* **1991**, *8*, S–192.
34. Prapaitrakul, W.; Sprockel, D.L.; Shivanand, P.; Sen, M. Development of a Drug Delivery System Through Melt Extrusion. *Abstracts of the 4th AAPS, Pharm. Res*; Atlanta, GA; **6**, S98.
35. Mank, R.; Kala, H.; Richter, M. Darstellung Wirkstoffhaltiger Extrusionformlinge Auf Der Basis Von Thermoplasten., Teil 1, Untersuchungen Zur Wirkstoffliberation. *Pharmazie* **1989**, *44*, 773–776.
36. Mank, R.; Kala, H.; Richter, M. Darstellung Wirkstoffhaltiger Extrusionformlinge Auf Der Basis Von Thermoplasten., Teil 2, Untersuchungen Zur Optimierung Der Wirkstofffreigabe. *Pharmazie* **1990**, *45*, 592–593.
37. Sprockel, O.; Sen, M.; Shivanand, P.; Prapaitrakul, W. A Melt-Extrusion Process for Manufacturing Matrix Drug Delivery Systems. *Int. J. Pharm.* **1997**, *155*, 191–199.
38. Steuernagel, C.R. Latex Emulsions for Controlled Drug Delivery. *Aqueous Polymeric Coatings For Pharmaceutical Dosage Forms*; McGinity, J.W., Ed.; Marcel Dekker, Inc.: New York, 1989.
39. Factors Affecting Drug Release from Sustained-Release Film-Coated Pellets Using Acrylic Aqueous Dispersions 6th International Conference on Pharmaceutical Technology, Paris, France, June 2–4, 1992.
40. Schmidt, P.C.; Niemann, F. The MiniWiD-Coater. III. Effect of Application Temperature on the Dissolution Profile of Sustained-Release Theophylline Pellets Coated with Eudragit RS 30 D. *Drug Dev. Ind. Pharm.* **1993**, *19*, 1603–1612.
41. Repka, M.A.; McGinity, J.W. Influence of Vitamin E TPGS on the Properties of Hydrophilic Films Produced by Hot-Melt Extrusion. *Int. J. Pharm.* **2000**, *202* (1–2), 63–70.
42. Schiraldi, M.T.; Perl, M.M.; Rubin, H. Bioadhesive Extruded Film for Intra-Oral Drug Delivery and Process. US Patent 4,713,243, December 15, 1987.
43. Schiraldi, M.T.; Perl, M.M.; Rubin, H. Bioadhesive Extruded Film for Intra-oral Drug Delivery and Process. US Patent RE33,093, October 17, 1989.
44. McGinity, J.W.; Koleng, J.J. Preparation and Evaluation of Rapid-Release Granules Using a Novel Hot-Melt Extrusion Technique, Abstracts of the 16th Pharmaceutical Technology Conference, Athens Greece, 1997; **2** 153.
45. Dahl, T.C. Ethylcellulose. *Handbook of Pharmaceutical Excipients*; Wade, A., Weller, P.J., Eds.; American Pharmaceutical Association: Washington DC, 1994; 186–190.
46. Harwood, R.J.; Johnson, J.L. Hydroxypropylcellulose. In *Handbook of Pharmaceutical Excipients*; Wade, A., Weller, P.J. Eds.; American Pharmaceutical Association: Washington DC, 1994; 223–228.
47. Braun, D.B.; DeLong, D.J. Polyethers (Ethylene Oxide Polymers). *Encyclopedia of Chemical Technology*; Kirk-Othmer, Ed.; John Wiley & Sons, Inc.: New York; 1982; 616–632.
48. Shukla, A.J. Polymethacrylates. *Handbook of Pharmaceutical Excipients*; Wade, A., Weller, P.J., Eds.; American Pharmaceutical Association: Washington DC, 1994; 362–366.
49. Kobayashi, N.H. Carnauba Wax. *Handbook of Pharmaceutical Excipients*; Wade, A., Weller, P.J., Eds.; American Pharmaceutical Association: Washington DC, 1994; 552–553.

HOMOGENIZATION AND HOMOGENIZERS

Venkatesh Naini

Barr Laboratories, Inc., Pomona, New York, U.S.A.

Shailesh K. Singh

Wyeth-Ayerst Research, Peral River, New York, U.S.A.

INTRODUCTION

Homogenization encompasses techniques of emulsification of one liquid into another, dispersing solid particles uniformly in a product, and disrupting cell membranes. Traditionally, homogenizers have been used in the pharmaceutical industry for emulsification. However, they are finding increasing applications in the manufacture of liposomes (1), nanosuspensions (2), solid-lipid nanoparticles (3), tablet coating dispersions (4), micro-encapsulation (5), and in cell disruption for harvesting therapeutic proteins in cell cultures (6).

Pharmaceutical emulsions are generally classified as oil-in-water (o/w) or water-in-oil (w/o) systems, where the first component represents the dispersed phase, although more complex systems are feasible. The first step in the process of emulsification involves application of mechanical energy (homogenization) to break up the dispersed phase and form a stable emulsion. Homogenization is also used for particle size reduction in pharmaceutical suspensions. Important factors controlling the formation of pharmaceutical emulsions and dispersions are mechanical and/or formulation related. Mechanical forces during homogenization cause droplet or particle size reduction by shear, turbulence, impact, and cavitation (7). Shear is caused by elongation and subsequent breakup of droplets, due to acceleration of a liquid. Cavitation is caused by an intense pressure drop, leading to formation of vapor bubbles in the liquid, which implode causing shock waves in the fluid. This leads to disruption of droplets, particles, and cell membranes. Homogenizers, available from different manufacturers operate using a combination of these forces (8).

This review will focus on commonly used homogenizers in the pharmaceutical industry viz. high-pressure homogenizer, rotor-stator homogenizer, microfluidizer, and ultrasonic homogenizer.

HIGH-PRESSURE HOMOGENIZATION

Auguste Gaulin introduced the first high-pressure homogenizer in 1900 for homogenizing milk (9). The basic high-pressure homogenizer consists of a positive displacement pump attached to a homogenizing valve assembly (Fig. 1). The pump forces liquid into the valve area at a high pressure. As the product is forced through the adjustable gap (D), its velocity increases tremendously with a corresponding decrease in pressure. The emerging product then impinges on the impact ring (C). This sudden change in energy causes increased turbulence, shear, and/or cavitation, resulting in droplet size reduction and uniform dispersion of particles. High-pressure homogenizers are used in emulsification (10–12), preparation of microparticles and nanodispersions (13–16), liposomes (1, 17, 18), and in cell disruption (6, 19). A laboratory scale model of the high-pressure homogenizer is shown in Fig. 2. For emulsion processing a single-stage or two-stage valve assembly can be used, where 10% of the total pressure is applied at the second stage. Another commonly used approach is the multiple-pass homogenization, if a very narrow particle size distribution is needed. This can be achieved by using a series of homogenizers or processing several discrete passes through the same machine.

Several factors affect the final emulsion formulation obtained using high-pressure homogenization. More importantly the level and type of surfactant (11, 12), level of the oily phase, and homogenization process parameters, such as pressure, number of cycles, or discrete passes through the homogenizer, play a significant role. Pandolfe studied the effect of several of these factors and found that the premix condition (prior to homogenization), emulsifier concentration, and energy input by the homogenizer, significantly affected the quality of the final emulsion (11). The effect of increasing homogenizing pressure on emulsions with

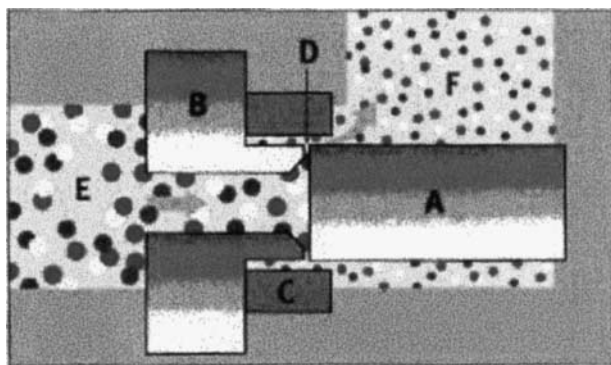


Fig. 1 Homogenizing valve assembly in a high-pressure homogenizer. (With permission: APV Homogenizer Group, Wilmington, MA.)

different levels of oily phase, in poor premix (turbine stirrer at 1000 rpm) and good premix (prehomogenized at 500 psi) samples, is shown in Fig. 3. Although droplet size reduction is seen at all conditions, more effective

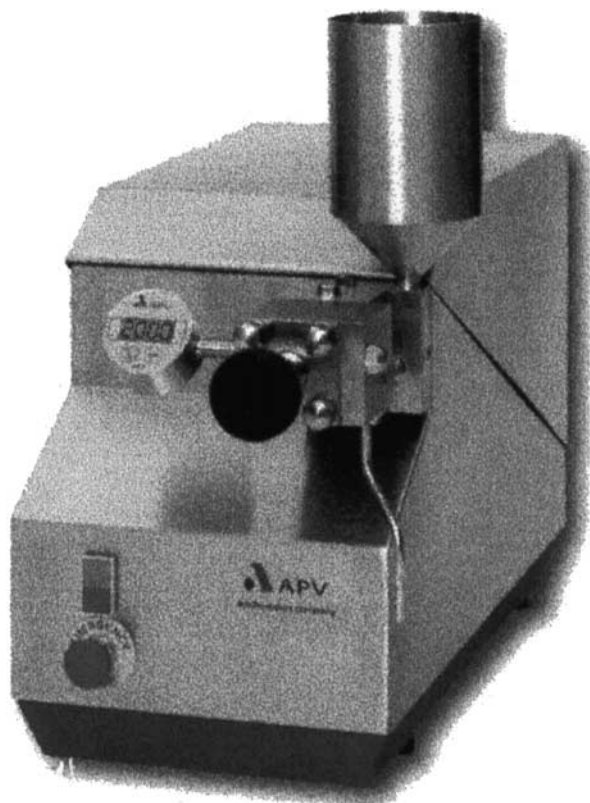


Fig. 2 Laboratory scale model of a Gaulin type high-pressure homogenizer. (With permission: APV Homogenizer Group, Wilmington, MA.)

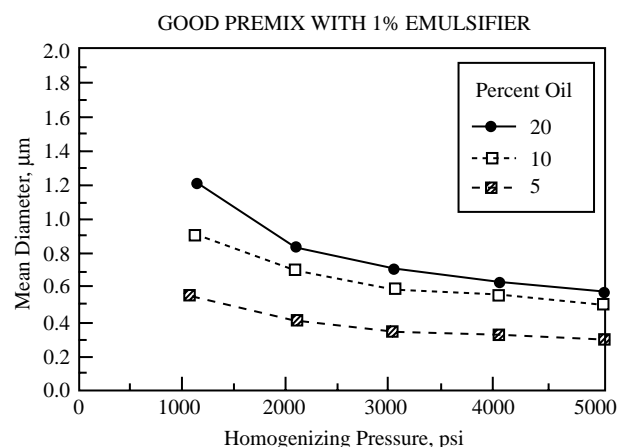
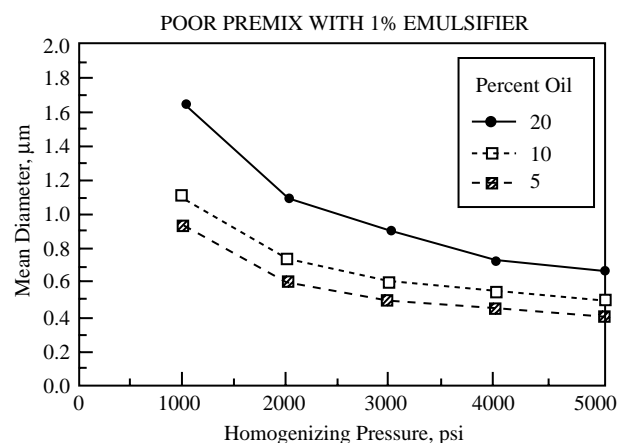


Fig. 3 Mean droplet diameter versus homogenizer pressure for emulsions with 1% emulsifier and 5%, 10%, or 20% oily phase. (Adapted from Ref. 11.)

formulations can be obtained by using the lowest amount of oily phase at the highest homogenization pressure with a properly premixed dispersion. It was also concluded that increasing homogenization pressure could effectively reduce the amount of emulsifier required in a formulation. Because of its efficient droplet size reduction, high-pressure homogenization can be used for preparing parenteral fat emulsions (10). Here the requirement is that number of droplets or particles above 1 μm should be limited and no particle should be larger than 5 μm .

Calvor and Muller (13) used high-pressure homogenization and a novel method to prepare biodegradable microparticles of poly(D,L-lactide) (PLA) and poly(D,L-lactide-co-glycolide) (PLGA). They heated the drug-polymer-containing suspensions above the glass transition temperature (T_g) of the polymer, followed by high-pressure homogenization. Above the T_g of the

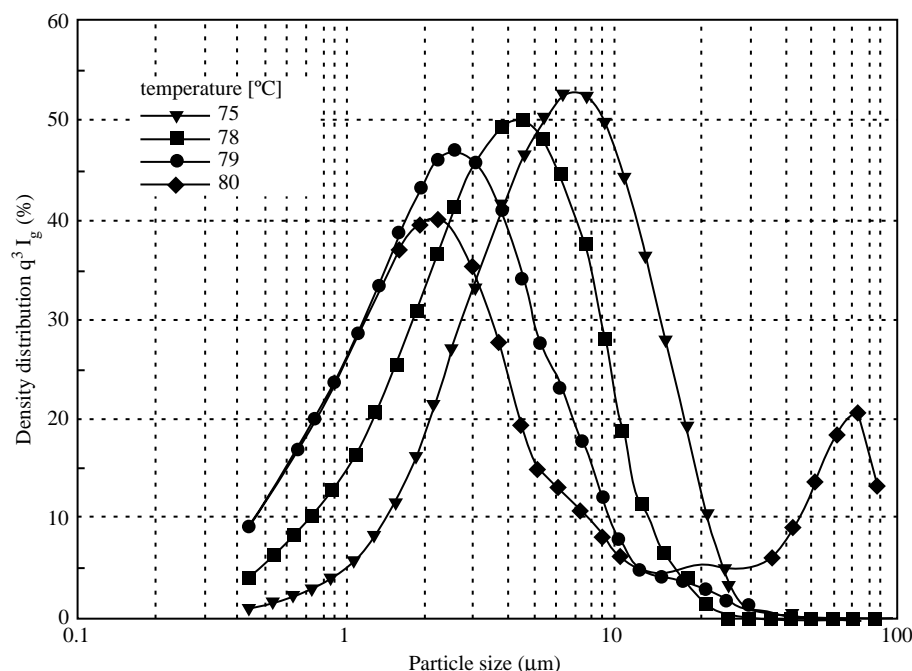


Fig. 4 Volume size distribution of poly(D,L-lactide-co-glycolide) particles at a homogenization temperature of 75°C (▼), 78°C (■), 79°C (◆) and 80°C (●). (Adapted from Ref. 13.)

polymer, the viscosity of the dispersed phase was lowered leading to efficient droplet size reduction by homogenization. Fig. 4 shows the effect of homogenization temperature on particle size reduction in PLGA ($T_g \sim 40^\circ\text{C}$) containing systems. Traditionally, Gaulin-type homogenizers are known to uniformly disperse and breakup particle agglomerates. Actual breakup of primary particles in the suspension was considered highly unlikely. However, more recently, high-pressure homogenization was successfully employed to prepare nanosuspensions of poorly soluble drugs, starting from micron-sized material (14). Some factors controlling size reduction of drug particles during homogenization include, gap width in the homogenizer, particle shape and size in the feed material, and fragility of drug crystals. Nanosuspensions have promising applications in injectable formulations of poorly soluble drugs or reformulating solution parenterals, which contain toxicologically less favorable excipients. They can also be used to enhance saturation solubility and bioadhesive properties of drugs in the GIT, leading to better bioavailability following oral administration (2). Due to the abrasive nature of suspended drug particles, erosion of contact surfaces and heavy-metal contamination is a major concern during high-pressure homogenization. A study by Krause and coworkers

found that heavy-metal contamination was minimal (<1 ppm) in nanosuspensions after being homogenized at 1500 bar for 50 cycles (20).

Liposomes are phospholipid vesicles containing an aqueous compartment surrounded by one or more bilayers (1). They are finding increasing application as carriers for small molecule drugs, controlled release and targeting of protein and peptide therapeutics, and as immunological adjuvants in vaccines (17). Due to the realization that liposomes need to be produced on a large-scale, high-pressure homogenizers are well suited for industrial production under aseptic conditions. The effects of shear and cavitation during homogenization usually form small unilamellar vesicles (SUVs). The number of passages through the homogenizer and pressure used affects the vesicle sizes (17, 18). For a certain combination of lipid and water, increasing homogenization pressure produces smaller and narrower (decreasing polydispersity) vesicles with an optimum diameter. Any further increase in the number of passes results in broader size distributions due to coalescence (17). Bachmann and coworkers, used a continuously operating high-pressure homogenizer to scale-up production of liposomes using the “one-step” method, where SUVs are prepared from powdered lipid and aqueous drug solution (18). Encapsulation efficiency and entrapped aqueous volume of the vesicles are

Table 1 Encapsulation efficiency and entrapped volumes of vesicles prepared using a high-pressure homogenizer

Number of cycles	40 MPa		Homogenizing pressure (70 MPa)	
	Encapsulation efficiency (%)	Aqueous volume entrapped (l/mol)	Encapsulation efficiency (%)	Aqueous volume entrapped (l/mol)
1	11.8	0.91	11.6	0.89
5	10.4	0.80	9.0	0.69
10	8.0	0.62	7.5	0.58
20	7.6	0.58	6.1	0.47

(From Ref. 18.)

summarized in Table 1. Liposomes prepared with phosphatidylcholine fraction of soybean lecithin (SPC) were homogenized at 40 MPa or 70 MPa, with the higher pressure producing much smaller vesicles. However, extensive recirculation decreased these differences after several passes through the homogenizer. In addition, encapsulation efficiencies decreased at higher pressures and repetitive processing (Table 1).

Recent advances in biotechnology have produced several new protein drugs from mammalian and bacterial cell cultures. High-pressure homogenization is widely used to harvest intracellular proteins and enzymes of interest in cell cultures (6, 19). Lander et al. conducted a mechanistic study of cell disruption caused by homogenization (19). Shear and cavitation were found to play an important role in cell membrane disruption and release of intracellular contents. High-pressure homogenizers for cell disruption applications use special valve assemblies for efficient rupture of cell walls (6). Other process parameters to consider during cell disruption are viscosity of the cell suspension, flow rate, and number of passes through the homogenizer.

MICROFLUIDIZATION

The microfluidizer is a high-pressure homogenizer that works, on a different principle. The pre-homogenized liquid is forced through an interaction chamber using a high-pressure pump. The interaction chamber consists of ceramic microchannels, which cause the liquid feed to split into two streams. These streams are then recombined at very high velocities producing forces of shear, impact, and cavitation, which cause droplet or particle-size reduction in emulsions and suspensions. A complete

description of the operation of the microfluidizer is summarized in US Patent 4,533,254 (21). A schematic diagram of the microfluidizer process is shown in Fig. 5. These homogenizers are commercially available from Microfluidics Corporation (Newton, MA). Microfluidizers are capable of handling emulsions (21–23), artificial blood (24, 25), suspensions (26), and liposomes (27–30). A microfluidizer that can operate at process pressures of up to 40,000 psi is shown in Fig. 6.

Because of their efficient droplet size reduction and ease of scale-up, microfluidizers are frequently used to prepare parenteral feeding emulsions (22, 23). Droplet diameters were directly related to the process pressure used, number of passes through the microfluidizer, and concentrations of emulsifier and oily phase in the emulsion (22). Reduction in droplet size of a 10% emulsion as a function of number of passes through the microfluidizer is shown in Fig. 7 (22). When the homogenizer was operated at its maximum operating pressure of 10,000 psi, droplet diameters decreased from 380 nm for a single pass to a plateau of 250 nm after four cycles, with further processing having no significant effect. Lidgate et al. used a microfluidizer to prepare an o/w parenteral emulsion for use as a vaccine adjuvant and compared its stability to emulsions prepared by other methods (23). Stress tests to induce creaming were used to test emulsions produced by various techniques. Microfluidization produced a superior parenteral emulsion compared to a homogenizer mixer. Stability correlated well with increasing number of microfluidizer cycles used to process the emulsion (23).

The microfluidizer has distinct advantages over conventional milling processes in particle size reduction of pharmaceutical suspensions. Absence of heavy-metal contaminants due to surface erosion and easy scale-up to production were observed when using the microfluidizer

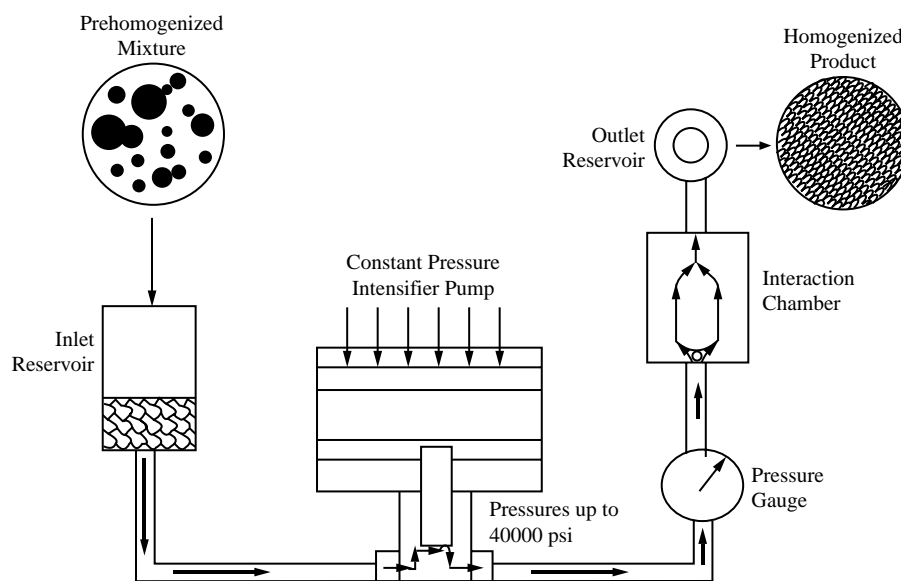


Fig. 5 Microfluidizer processor flow diagram. (Microfluidics Corporation, Newton, MA.)

for preparing radiopaque suspensions (25). Several reports deal with scaled up production of liposomes using the microfluidizer (28–30). In one study, liposome dispersions with relatively high lipid concentrations (400 $\mu\text{mol/ml}$) could be processed to narrow size

distributions using the microfluidizer (26). In addition, a US Patent (4,776,991) describes the large-scale production of liposome encapsulated hemoglobin for use as a blood substitute (30). Liposomes produced with high-pressure homogenizers usually result in small unilamellar vesicles (SUVs). Their main disadvantages are low encapsulation efficiency and tendency to leak their contents more often than multilamellar vesicles (MLVs). Sorgi and Huang used the microfluidizer to prepare cationic liposomes, where the active component is not encapsulated, but forms

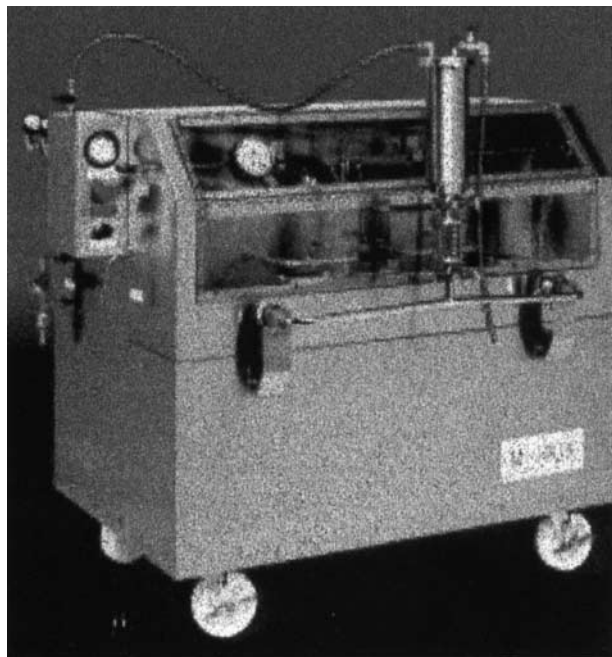


Fig. 6 The M-140K Microfluidizer processor. (Microfluidics Corporation, Newton, MA.)

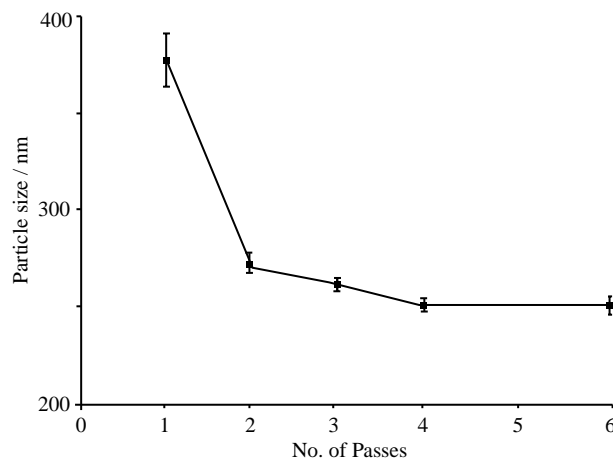


Fig. 7 Emulsion droplet size versus number of passes through the microfluidizer for a 10% oily phase emulsion processed at 10,000 psi. (Adapted from Ref. 22.)

a complex with the liposome using charge interaction (29). The microfluidizer was well suited for preparation and scale-up of cationic liposomes of a plasmid DNA, which was successfully used in gene therapy clinical trials (29).

ROTOR-STATOR HOMOGENIZATION

The rotor-stator homogenizer is one of the most commonly used pieces of equipment in the pharmaceutical industry. Although they have limited capability in achieving very fine droplets or particles, rotor-stator mixers are capable of handling liquids at much higher viscosities, compared to high-pressure homogenizer and the microfluidizer. A rotor-stator homogenizer consists of an impeller in close tolerance to a stationary housing, which restricts the flow of liquid caused by the impeller movement. Shear and impact comminute particles and droplets caught between the rotor and stator (5, 31). The colloid mill is an extreme example of the rotor-stator homogenizer, where the gap between the rotating truncated cone (rotor) and its housing (stator) is adjustable. However, the colloid mill suffers from disadvantages like generation of excessive heat and incorporation of air in the finished product. Various geometries and configurations of the mixing head in the rotor/stator design are available from different manufacturers (8). They can be used in the batch mode and continuous, or "in-line," mode. An "in-line" rotor/stator homogenizer is depicted in Fig. 8. Parameters affecting final product quality in rotor/stator homogenization are homogenization intensity, residence time of product in the shearing field, viscosity of the dispersed and continuous phases, surfactant concentration, rotor/stator design,

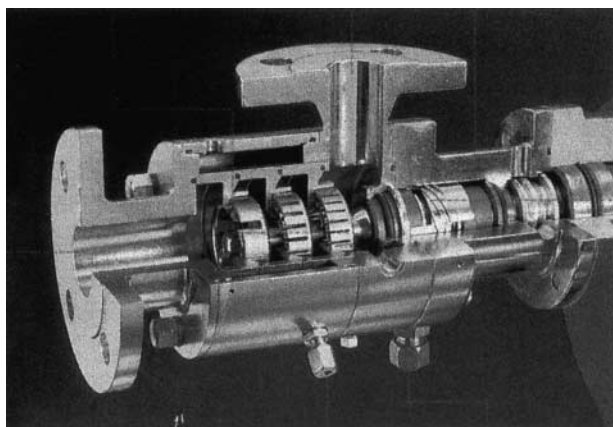


Fig. 8 An in-line rotor/stator homogenizer. (IKA Works, Wilmington, NC.)

volume of the mixer, and volume ratio of the two phases (5, 31).

Djakovic and coworkers, studied several factors such as homogenization time, emulsifier concentration, and homogenization intensity to determine optimal parameters for emulsification, using a rotor-stator homogenizer (32). Mean droplet diameter and polydispersity were used as measures of final product quality. For constant homogenization intensity (rpm) and mixing time, droplet size and polydispersity decreased with increasing emulsifier concentration before reaching an optimum level (32). Maa and Hsu (5) compared the batch mode and a flow-through apparatus, using rotor/stator homogenization for microencapsulation. Emulsion droplets obtained using the flow-through method were consistently higher than the batch mode. Since emulsification is effected by residence time of liquid in the shearing field, the flow-through method induced lower shear compared to the batch mode. However, using effective recirculation in the flow-through mode can overcome this problem (31).

ULTRASONIC HOMOGENIZATION

Sonication emulsifies primarily by cavitation. An ultrasonic homogenizer consists of a generator, converter, and horn tip (8, 31). The converter consists of a piezoelectric quartz crystal, which transforms electrical energy into high intensity vibrations and transmits them to the horn tip immersed in the liquid. Droplet size reduction occurs

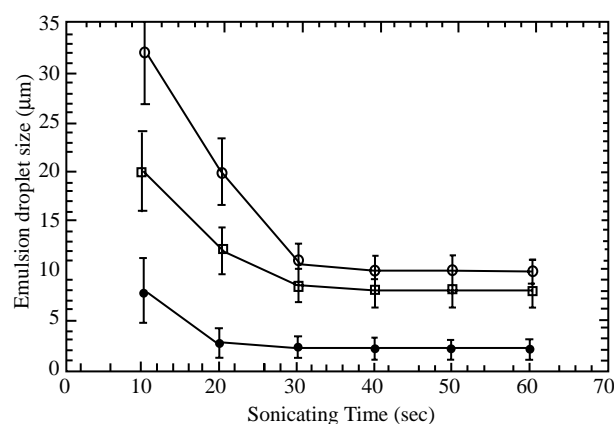


Fig. 9 Effect of sonication time on emulsion droplet size for 0.4 g/ml of poly(methyl methacrylate)/methylene chloride solution in 6% of polyvinyl alcohol (PVA) solution at a volume ratio of 15:2 (ml/ml) sonicated at 20% (○), 50% (□), and 100% (●) of full power. (Adapted from Ref. 31.)

Table 2 List of homogenization equipment supplied by various manufacturers

Type of homogenizer	Model (manufacturer)	Mode of operation (Batch/continuous)	Operating parameters	Capacity	Applications
High-pressure	APV Model 2000 (APV Homogenizers)	Laboratory scale; batch or continuous	Maximum pressure: 30000 psi	Batch (100 ml) Continuous (11 L/h)	Emulsions, nanodispersions, liposomes, ointments, cell disruption, vaccines, parenteral emulsions
	Gaulin and Rannie Models (APV Homogenizers)	Production scale; batch or continuous	Maximum pressure: 21750 psi	Up to 50000 L/h	
	Ariete Model NS8315 (Niro Soavi)	Production scale; batch or continuous	Pressure range: 2000–15000 psi	8000–50000 L/h	
	M-110Y (Microfluidics)	Laboratory scale; batch or continuous	Pressure range: 3000–23000 psi	Batch (>60 ml) Continuous (250–600 ml/min)	Dispersions, emulsions, cell disruption, encapsulation, liposomes, vaccines, parenteral emulsions
Microfluidizer	M-140K (Microfluidics)	Laboratory scale; batch or continuous	Pressure range: 8000–40,000 psi	Batch (1000 ml) Continuous (500 ml/min)	
	M-210EH (Microfluidics)	Production scale; batch or continuous	Pressure range: 2500–30,000 psi	Batch (3.8 L) Continuous (5.7 L/min)	
	Silverson Model GX25	Production scale; batch	Maximum rpm: 3600	2400 L (low viscosity) 400 L (high viscosity)	Emulsions dispersions, pastes, creams, lotions
	Ultra-Turrax UTL (IKA Works, Inc.)	Production scale; continuous	Maximum rpm: 6000	3500 L/h	
Ultrasonic	Microson XL2007 (Misonix, Inc.)	Laboratory scale; batch	Power: 100 W; Frequency: 22.5 kHz	<50 ml	Emulsions, dispersions, cell disruption
	Floccell 800D (Misonix, Inc.)	Production scale; continuous	Power: 475 W; Frequency: 20 kHz	With booster horn 40 L/min	

mainly by intense shock waves generated in close proximity to the tip. For large-scale applications, ultrasonic homogenizers can be used in the continuous mode with a flow-cell (33). In general, droplet size reduction in ultrasonic homogenizers is affected by sonication intensity, viscosity of the mixture, emulsifier concentration, and time of sonication (31, 34). Fig. 9 shows the effect of increasing sonication power on emulsion droplet size in liquid-liquid emulsification (31). Higher sonicating power resulted in smaller emulsion droplets. At all power levels the droplet size reduced dramatically, initially followed by a leveling-off phase. Sonication is comparable to rotor-stator homogenization if sufficient power is used. However, as liquid viscosity increases, rotor-stator homogenization is more efficient due to shear effects (31).

EQUIPMENT CONSIDERATIONS

A variety of homogenizers capable of performing a range of processes are available. Important considerations during formulation development include the feasibility and availability of pilot and production scale equipment, which can reproduce the same results. A sampling of homogenizers available from selected manufacturers is given in Table 2. High-pressure homogenizers and microfluidizers are available in a wide range of capabilities ranging from bench-top models to production equipment capable of handling large amounts of material. These homogenizers are limited in their handling of high viscosity fluids compared to rotor-stator homogenizers, which are designed to handle even thick pastes and creams. However, particle size reduction is more efficiently carried out using high-pressure homogenizers and microfluidizers. Ultrasonic homogenizers are also capable of handling large volumes by using the continuous or flow-through approach. However higher intensities are needed to cause cavitation and droplet reduction in high viscosity fluids. As indicated in Table 2, high-pressure homogenizers and microfluidizers can be easily adapted for batch or continuous processing. Rotor-stator homogenizers are available either as batch processors or in-line dispersers, which can be used continuously or by recirculating the product.

REFERENCES

1. Brandl, M.M.; Bachmann, D.; Drechsler, M. Liposome Preparation Using High-Pressure Homogenizers. *Liposome Technology*, 2nd Ed.; Gregoriadis, G., Ed.; CRC Press: Boca Raton, 1993; 1, 49–65.
2. Muller, R.H.; Peters, K. Nanosuspensions for the Formulation of Poorly Soluble Drugs I. Preparation by a Size Reduction Technique. *Int. J. Pharm.* **1998**, *160*, 229–237.
3. Muller, R.H.; Mehnert, W.; Lucks, J.-S.; Schwarz, C.; Muhlen, A.; Weyhers, H.; Freitas, C.; Ruhl, D. Solid Lipid Nanoparticles (SLN)—An Alternative Colloidal Carrier System for Controlled Drug Delivery. *Eur. J. Pharm. Biopharm.* **1995**, *41* (1), 62–69.
4. Bodmeier, R.; Chen, H. Hydrolysis of Cellulose Acetate Butyrate Pseudolatexes Prepared by a Solvent Evaporation—Microfluidization Method. *Drug Dev. Ind. Pharm.* **1993**, *19* (5), 521–530.
5. Maa, Y.-H.; Hsu, C.C. Liquid-Liquid Emulsification by Rotor/Stator Homogenization. *J. Control. Release* **1996**, *38*, 219–228.
6. Pandolfe, W.D. *Cell Disruption by Homogenization*; APV Homogenizers: Wilmington, MA, 1996; 1–20.
7. Walstra, P. Formation of Emulsions. *Encyclopedia of Emulsion Technology: Basic Theory*; Becher, P., Ed.; Marcel Dekker, Inc.: New York, 1979; 1, 58–128.
8. Scott, R.R. A Practical Guide to Equipment Selection and Operating Techniques. *Pharmaceutical Disperse Systems*; Lieberman, H.A., Rieger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1989; 2, 1–71.
9. Ed. *Processing of Emulsions and Dispersions by Homogenization*; APV Homogenizers: Wilmington, MA, 1996; 1–23.
10. Bock, T.K.; Lucks, J.-S.; Kleinbudde, P.; Muller, R.H.; Muller, B.W. High Pressure Homogenization of Parenteral Fat Emulsions—Influence of Process Parameters on Emulsion Quality. *Eur. J. Pharm. Biopharm.* **1994**, *40* (3), 157–160.
11. Pandolfe, W.D. Effect of Premix Condition, Surfactant Concentration, and Oil Level on the Formation of Oil-in-Water Emulsions by Homogenization. *J. Dispersion Sci. Tech.* **1995**, *16* (7), 633–650.
12. Daniels, R.; Schulz, M.B. Hydroxypropylmethylcellulose (HPMC) As Emulsifier for Submicron Emulsions: Influence of Molecular Weight and Substitution Type on Droplet Size After High-Pressure Homogenization. *Eur. J. Pharm. Biopharm.* **2000**, *49*, 231–236.
13. Calvor, A.; Muller, B.W. Production of Microparticles by High-Pressure Homogenization. *Pharm. Dev. Tech.* **1998**, *3* (3), 297–305.
14. Grau, M.J.; Kayser, O.; Muller, R.H. Nanosuspensions of Poorly Soluble Drugs—Reproducibility of Small Scale Production. *Int. J. Pharm.* **2000**, *196*, 155–157.
15. Liedtke, S.; Wissing, S.; Muller, R.H.; Mader, K. Influence of High Pressure Homogenization Equipment on Nanodispersions Characteristics. *Int. J. Pharm.* **2000**, *196*, 183–185.
16. Lamprecht, A.; Ubrich, N.; Perez, M.H.; Lehr, C.-M.; Hoffman, M.; Maincent, P. Influences of Process Parameters on Nanoparticle Preparation Performed by a Double Emulsion Pressure Homogenization Technique. *Int. J. Pharm.* **2000**, *196*, 177–182.
17. Brandl, M.; Bachmann, D.; Drechler, M.; Bauer, K.H. Preparation by a New High Pressure Homogenizer Gaulin Micron Lab 40. *Drug Dev. Ind. Pharm.* **1990**, *16* (14), 2167–2191.

18. Bachmann, D.; Brandl, M.; Gregoriadis, G. Preparation of Liposomes Using a Mini-Lab 8.30 H High-Pressure Homogenizer. *Int. J. Pharm.* **1993**, *91*, 69–74.
19. Lander, R.; Manger, W.; Scouloudis, M.; Ku, A.; Davis, C.; Lee, A. Gaulin Homogenization: A Mechanistic Study. *Biotechnol. Prog.* **2000**, *16*, 80–85.
20. Krause, K.P.; Kayser, O.; Mader, K.; Gust, R.; Muller, R.H. Heavy Metal Contamination of Nanosuspensions Produced by High-Pressure Homogenization. *Int. J. Pharm.* **2000**, *196*, 169–172.
21. Cook, E.J.; Lagace, A.P. Apparatus for Forming Emulsions. US Patent 4,533,254, Aug 6, 1985.
22. Washington, C.; Davis, S.S. The Production of Parenteral Feeding Emulsions by Microfluidizer. *Int. J. Pharm.* **1988**, *44*, 169–176.
23. Lidgate, D.M.; Fu, R.C.; Fleitman, J.S. Using a Microfluidizer to Manufacture Parenteral Emulsions. *Biopharm.* **1989**, *10*, 28–33.
24. Zheng, S.; Zheng, Y.; Beissinger, R.L.; Wasan, R.T.; McCormick, D.L. Hemoglobin Multiple Emulsion as an Oxygen Delivery System. *Biochim. Biophys. Acta.* **1993**, *1158*, 65–74.
25. Zheng, S.; Beissinger, R.L.; Wasan, D.T.; Sehgal, L.R.; Rosen, A.L. Oxygen Carrying Multiple Emulsions. US Patent 5,438,041, Aug 1 1995.
26. Illig, K.J.; Mueller, R.L.; Ostrander, K.D.; Swanson, J.R. Use of Microfluidizer for Preparation of Pharmaceutical Suspensions. *Pharm. Tech.* **1996**, *10*, 78–88.
27. Talsma, H.; Ozer, A.Y.; van Bloois, L.; Crommelin, D.J.A. The Size Reduction of Liposomes with a High Pressure Homogenizer (Microfluidizer®). Characterization of Prepared Dispersions and Comparison with Conventional Methods. *Drug Dev. Ind. Pharm.* **1989**, *15* (2), 197–202.
28. Vemuri, S.; Yu, C.; Wangsatorntanakun, V.; Roosdorp, N. Large-Scale Production of Liposomes Using a Microfluidizer. *Drug Dev. Ind. Pharm.* **1990**, *16* (15), 2243–2256.
29. Sorgi, F.L.; Huang, L. Large Scale Production of DC-Chol Cationic Liposomes by Microfluidization. *Int. J. Pharm.* **1996**, *144*, 131–139.
30. Farmer, M.C.; Beissinger, R.L. Scaled-up Production of Liposome-Encapsulated Hemoglobin. US Patent 4,776, 991, Oct 11, 1988.
31. Maa, Y.-H.; Hsu, C.C. Performance of Sonication and Microfluidization for Liquid–Liquid Emulsification. *Pharm. Dev. Tech.* **1999**, *4* (2), 233–240.
32. Djakovic, L.M.; Dokic, P.P.; Sefer, I.B. Mathematical and Experimental Essentials of the Emulsification Process: Optimal Parameters Determination. *J. Disp. Sci. Tech.* **1989**, *10* (1), 59–76.
33. *Continuous Ultrasonic Processing Cell*; Misonix Corporation: Farmingdale, NY, 1998.
34. Higgins, D.M.; Skauen, D.M. Influence of Power on Quality of Emulsions Prepared by Ultrasound. *J. Pharm. Sci.* **1972**, *61* (10), 1567–1570.

LENS CARE PRODUCTS

Masood Chowhan

Ralph Stone

Alcon Laboratories, Fort Worth, Texas, U.S.A.

INTRODUCTION

Contact lenses are made of polymeric materials designed and fabricated to correct vision. Because these lenses are removed from the eye after a prescribed wear time, lens care products are required to clean, disinfect and rinse them prior to reinsertion to avoid ocular infections and other complications. Lens care products are also required to enhance the comfort of lens wear.

HISTORICAL OVERVIEW

Lens care products are relatively new compared to many other pharmaceutical products. Leonardo da Vinci was the first to conceive the concept of the contact lens. In 1508 he illustrated the concept of vision involving “upside down” images with a water-filled sphere covering the eye. However, the actual development of contact lenses did not occur until about 100 years ago (1887–1888) when scleral contact lenses were fabricated. The three innovators credited for this are Dr. A.E. Fick, a physician in Zurich, F.A. Mueller, a maker of prosthetic eyes in Germany, and Dr. Eugene Kalt, a French physician. The earlier lenses were made of glass. In the late 1930s (1937–1939), Mullen, Oberg and Gyrrfy are credited with fabricating plastic contact lenses made from methyl methacrylate (PMMA). However, Kevin Tuohy, who filed a patent for contact lens design in 1948, is recognized as the “father” of modern day corneal contact lenses. These early lenses were rigid and uncomfortable with very low oxygen transmission. Advances in rigid lens technology have provided materials capable of oxygen transmission required to maintain corneal health. In the 1960s Otto Victerle developed the hydrophilic soft contact lens from polyhydroxyethyl methacrylate (HEMA). Hydrophilic soft contact lenses are the primary lenses available today. Since then, significant technological advances have been made in contact lens material, designs, and manufacturing processes. In 1990, the estimate of contact lens wearers in the United States alone was around 30 million.

The commercialization of the first pharmaceutical quality lens care products occurred in the 1950s. Harry Hind, a pharmacist and founder of the Barnes-Hind Company, has been credited as one of the first to develop and commercialize a wetting and storage solution for the rigid PMMA plastic lenses. Prior to his efforts, the literature mentions formulation of a saline solution containing sodium bicarbonate to be used with scleral lenses made from glass.

CURRENT CONTACT LENS MATERIALS AND FUTURE DIRECTIONS

Most of the materials currently used in fabricating contact lenses have been available since the mid-1960s with the exception of polymethyl methacrylate. Lens materials can be broadly classified as follows:

Rigid Gas-Permeable Lenses:

- Cellulose acetate butyrate
- Silicone
- Silicone acrylate
- Fluoro silicone acrylate
- *t*-Butylstyrene
- *t*-Butylstyrene-*co*-silicone acrylate

Soft Hydrophilic Lenses:

- Polyhydroxyethyl methacrylate
- Polyhydroxyethyl methacrylate-*co*-methacrylic acid
- Polyglyceryl methacrylate
- Polyhydroxyethyl methacrylate-*co*-polyvinylpyrrolidone
- Polyvinylpyrrolidone-*co*-methyl methacrylate
- Polyhydroxyethyl methacrylate-*co*-silicone (silicone hydrogel)

In addition to these materials, several others, such as polyurethanes, polysulfones, polyvinyl alcohol, and various copolymers, have been tried or are under development. Recently, silicon hydrogel lenses with high oxygen permeability were marketed. The future trends in

material development will continue to include polymers which have a high degree of oxygen permeability, resist accumulation of metabolic products of the cornea, and materials that resist lens deposits and bacterial attachment on lens surfaces to minimize the potential for ocular infection.

Classification of Contact Lenses Currently Marketed

During a period beginning from mid-1970, numerous hydrophilic and rigid gas-permeable lenses were introduced into the market. Although many had similar basic chemical compositions, they contained different additives designed to achieve desirable properties or to avoid infringement of existing patents. Such a proliferation of contact lens materials created confusion regarding Food and Drug Administration (FDA) criteria for approval of contact lenses and their care products. In the mid-1980s, the FDA worked with the contact lens manufacturers and evolved a classification for soft contact lenses based on the ionic or nonionic nature of polymers constituting the lens material and the water content. A classification was also worked out for rigid gas-permeable (RGP) lenses based on the chemical nature of the polymers. It is interesting to note that contact lenses and their care products were originally considered as drugs. However, upon passage of the U.S. Medical Device Act in 1996, contact lenses were reclassified as devices. Contact lenses and their care products were considered Class-III devices, which mandate the filing of a premarket approval application and obtaining FDA approval prior to marketing. Recently these products were reclassified as Class II devices and are currently cleared for marketing under the 510(k) premarket notification section of the regulations.

LENS CARE PRODUCTS BY FUNCTIONAL PURPOSE

Marketed lens care products fall mainly into the following categories: cleaners, disinfectants, lubricants, and multipurpose products. Cleaners are subdivided into daily or weekly cleaners. Disinfectants comprise solutions containing chemical antimicrobial agents, which do not require heating the lenses, and preserved or unpreserved saline solutions, which are used with an electrical thermal device for lens disinfection. These products are also used to rinse contact lenses. Lens lubricants are intended to enhance the comfort of lens wear and are used prior to insertion and during wear. Multipurpose solutions are intended to

accomplish two or more of the functions described earlier (cleaning, rinsing, and disinfection).

Rigid lens care also includes conditioning solutions to make the basic hydrophobic polymers wettable when placed on the eye.

PRODUCTS FOR CLEANING SOFT CONTACT LENSES

Lens Deposits

Composition

Basically there are two types of deposits: those resulting from tear components and those derived from other sources. Tear components especially proteins can accumulate on the lens surface. These proteins can denature or change conformations during absorptions on over tissue. Most deposits, with the exception of those that are tenaciously bound to the lens, can be cleaned easily with a surfactant-type of daily cleaner. Deposits resulting from tear components include proteinaceous deposits such as lysozyme, lactoferrin, albumin, globulins, etc. Proteinaceous deposits are present on all types of lenses. However, the amounts differ, depending on the number of ionic charges on and in the lens, the pore size, and the relative hydrophobicity of the polymers. For example, conventional nonporous and uncharged hard PMMA lenses with a hydrophobic surface attract very little proteinaceous deposits. Among the soft contact lenses with hydrophilic surfaces, the extent of deposits differs among various groups. For example, Group-4 lenses, which exhibit considerable negative charges due to methacrylic acid content, interact readily and heavily with a positively charged protein (lysozyme). Group-2 hydrogel lenses have no ionic charges, but can acquire substantial amounts of protein because of their large pore size. Besides protein deposits, lipid deposits are also found on contact lenses. These deposits are more common with rigid gas-permeable lenses because of their lipophilic nature. Such deposits may include cholesterol esters, wax esters, triglycerides, sterols, fatty acids, etc. Calcium present in tears results in calcium carbonate or phosphate-type of deposits as well as so-called mixed deposits (calcium bonded to organic compounds). Such deposits are common mainly in high water content soft contact lenses and are difficult to remove without damaging the contact lenses. Other deposits result from the patient environment. These include deposits resulting from cosmetics, make-up, and hair-spray, as well as materials from the wearing environment such as pollen dust and debris.

Table 1 Daily cleaners with shearing particles

Trade name	Manufacturer	Type of particle	Used with
OPTI-CLEAN	Alcon	Nylon	Soft, PMMA, RGP
OPTI-CLEAN II	Alcon	Nylon	Soft, PMMA, RGP
OPTI-FREE daily cleaner	Alcon	Nylon	Soft, PMMA, RGP
Boston cleaner	Polymer technology	Silica	PMMA, RGP
Boston advance cleaner	Polymer technology	Silica	PMMA, RGP

Problems associated with lens deposits

Cleaning is one of the most important steps in contact lens care. It helps in the removal of surface debris and contaminating microorganisms, thus facilitating the disinfection process. Improperly cleaned lenses can cause discomfort, red eye, decrease in visual acuity, and giant papillary conjunctivitis (GPC). The last often requires discontinuance of lens wear, at least until the symptoms clear. The change from heat to cold disinfection technology and the introduction of disposable lenses may have reduced the incidence of GPC. However, since these lenses can be worn on an extended basis for up to seven days without cleaning, GPC can still occur, as has been noted in the literature.

Classification of Lens Cleaners

Daily cleaners versus weekly cleaners

Daily cleaners generally contain surfactants and are used every day. They may also contain abrasive (deposit-shearing) particles, which enhance product performance. Commonly used daily cleaners with deposit shearing particles are listed in Table 1.

There are two types of weekly cleaners: those containing enzymes and those containing concentrated surfactants. Products containing enzymes for daily use are usually accepted by the patients and recommended by practitioners. Commonly used enzymatic products are listed in Table 2.

In-the-eye versus out-of-the-eye cleaners

Most of the cleaners marketed are out-of-the-eye cleaners; however, in recent years there has been a trend to try to develop cleaners for use while the lenses are inserted. A specific instance where such products could be beneficial is the case of the extended-wear lenses, which are not removed daily but are worn up to a week at a time. Generally, these cleaners are less effective in removing deposits already formed on the lens surface. However, they may play a role in retarding deposit formation.

Consumer versus professional use cleaners

Cleaners for lens wearers are used either on a daily or weekly basis and are fairly innocuous. Even upon gross misuse, they are not likely to be sight-threatening. Professional cleaners, however, are potent as well as toxic if not used properly. They are also more likely to damage the lens if used too frequently.

Active Components of Lens Cleaners

Surfactants

Surfactants are broadly classified into nonionic, anionic, cationic, and amphoteric types. Nonionic and amphoteric surfactants are most commonly used in contact lens cleaners, because strong anionic surfactants are generally toxic to the cornea. The recommended procedure by the manufacturer removes several types of deposits, except the most tenaciously bound and denatured proteins, lipids, and mucins. Surfactant-type cleaners are also effective in removing greater than 99.9% of microorganisms contaminating the lens. There are several mechanisms for their effectiveness, which include displacement of contaminants from the surface by mechanical force after the surface debris has been loosened as a result of a reduction in the interfacial tension. In addition, surfactants act by emulsification and micellar solubilization. They may also play a role in preventing or retarding deposition of contaminants.

Enzymes

Enzymes are biochemical molecules responsible for catalyzing reactions in which certain chemical bonds are broken. Their mechanism of action in cleaning involves attacking substrate protein, lipid, and mucin deposits, and fragmenting them into smaller molecules which are readily removed by the mechanical action of rubbing with the fingers and rinsing. Marketed products contain different enzymes, such as papain, pancreatin, and subtilisin. Papain and subtilisin are only proteolytic in nature, whereas pancreatin is a broad-spectrum enzyme

Table 2 Enzymatic products for contact lenses

Trade name	Manufacturers	Enzyme	Source	Dosage form
OPTI-FREE enzymatic cleaner	Alcon	Pancreatin	Mammals	Tablet
OPTI-ZYME enzymatic cleaner	Alcon	Pancreatin	Mammals	Tablet
SupraClens daily protein remover	Alcon	Pancreatin	Mammals	Liquid
Allergan enzymatic contact lens cleaner	Allergan	Papain	Plant	Tablet
ProFree/GP weekly enzymatic cleaner	Allergan	Papain	Plant	Tablet
Ultrazyme enzymatic cleaner	Allergan	Subtilisin-A	Microrganisms	Tablet
ReNu effervescent enzymatic cleaner	Bausch & Lomb	Subtilisin	Microrganisms	Tablet
ReNu thermal enzymatic cleaner	Bausch & Lomb	Subtilisin	Microrganisms	Tablet
Sensitive eyes enzymatic cleaner	Bausch & Lomb	Subtilisin	Microrganisms	Tablet

containing protease, lipase, and amylase enzymes that digest proteins, lipids, and mucins. Enzyme cleaners are effective in attacking all lens proteins, including the removal of tenaciously bound and denatured deposits, that cannot be removed by surfactant cleaners. Traditional enzymatic cleaning of contact lenses is recommended usually once a week. The soaking time varies from 15 min to overnight, followed by a disinfection process. However, more recently products have been introduced that can be used simultaneously during disinfection on a daily process. Certain enzyme products are recommended for this single-step cleaning and disinfection, using a heat or chemical regimen that enhances convenience and increases user compliance.

Oxidizing agents

Oxidizing agents such as sodium perborate and sodium percarbonate have also been used in cleaning contact lenses. None of these products are currently marketed for that purpose in the United States. Products marketed earlier were withdrawn because of their deleterious effects on lens polymers.

Deposit-shearing particles

Deposit-shearing particles are incorporated in suspension form in some daily cleaners. These formulations are more effective than daily surfactant cleaners as they are capable of removing tenaciously bound and denatured deposits. Some of the marketed products contain polymeric beads or silica. When used as recommended, these products are very effective and do not scratch the lens surface.

Chelating agents

Chelating agents such as disodium edetate (EDTA) are commonly used in lens care products to enhance the antimicrobial activity of preservatives and remove calcium and magnesium from the lens. EDTA is the most effective

chelating agent known for calcium and magnesium. Other chelating agents have been used such as phosphonates, which are most effective against iron.

Solvents

Solvents such as isopropyl alcohol have been incorporated in daily cleaners to aid in removing lipid type of deposits. Such solvents have been reported to affect certain lens materials, especially silicone acrylate rigid gas-permeable lenses.

Accessory Cleaning Products

Hand soaps

In the daily care of contact lenses, wearers are instructed to clean their hands with soap and to dry hands their hands with lint-free towels prior to handling their lenses. Selection of the specific soap product is important. In addition, to cleaning the hands thoroughly, it should be rinsing and should not cause ocular irritation even if residual amounts are transferred to the lens. With these considerations in mind, some hand soaps have, therefore, been designed specifically for contact lens users.

Cleaning devices

The general method of daily lens cleaning involves rubbing lenses between the index finger and thumb or placing the lens in the palm of the hand and rubbing with the index finger after applying cleaning solution to the surface. However, there are also special devices available in the market for cleaning lenses. They are said to clean lenses more effectively and avoid potential scratching by the fingers. These devices involve mechanical agitation and are manually or electrically operated. Ultrasound devices have been used mainly by lens practitioners in their office. None of those devices per se are effective in removing tenaciously bound, denatured deposits.

PRODUCTS FOR DISINFECTING CONTACT LENSES

Disinfection of contact lenses is an important step in preventing ocular irritation, red eye, and potential loss of eyesight due to corneal ulcers resulting especially from *Pseudomonas aeruginosa* infections. The pore openings of hydrophilic soft contact lenses are estimated to be between 3.0 and 7.0 nm, and are considerably smaller than the average bacterial particle size of 0.2–1.0 μm or fungus particle size of 2–6 μm . Even viruses ranging in particle size from 25 to 200 nm are large in size compared to the pore openings of soft lenses. None of the microorganisms can penetrate an intact lens matrix. However, when lenses are not properly cared for, some fungus growth facilitates the penetration of fungal hyphae into the matrix.

Contact lenses and their cases are frequently contaminated by microbes. Although studies have indicated that as many as 30% of lens cases are contaminated, the incidence of permanent ocular damage due to this is very low. Nevertheless, it is critical to properly instruct lens wearers and emphasize the importance of disinfection in order to avoid the potential risk of ocular infection or damage to the eyesight.

Thermal Versus Chemical Disinfection

Both thermal and chemical methods are commonly used for the disinfection of soft contact lenses. With the former, a case containing the lenses immersed in saline solution is heated by an electrical unit with a predesigned heating cycle. The current FDA requirement for thermal disinfection by saline solution requires a minimum temperature of 80°C for 10 min within the contact lens case. This ensures elimination of vegetative forms of ocular pathogens but not the spores.

The chemical method involves antimicrobial compounds with an adequate antimicrobial spectrum and biocidal action. The FDA guidelines include the methods for determination of efficacy for all disinfecting solutions. The initial testing process is defined as the elimination or reduction of microorganisms achieved over the disinfection period. The FDA guidelines also provide a method for manufacturers to conduct use tests on purposely contaminated contact lenses, following a complete disinfection regimen, which includes cleaning and rinsing. This test is generally known as the FDA regimen test. The FDA guidelines specify the types and levels of organisms as well as the details of the test. Other methods using ultraviolet light, microwave, and ultrasonics, have been

tried for lens disinfection but are not widely applied because of ineffectiveness or deleterious effects on lens materials.

The advantage of the thermal method is that it ensures complete elimination of vegetative forms of microorganisms, whereas chemical disinfectants may encounter some resistant organisms. Although the thermal method is preferable from the microbiological viewpoint, it has several disadvantages: It is a complex method involving the use of electrical devices. The failure of an electrical heating unit to perform properly presents a potential risk of ocular infection. Malfunctioning units and improper use may result in electrical shock to users and fires have been reported. If the saline solution in the lens container evaporates during the heating cycle as a result of carelessness in not properly tightening the lens case cap, the lens might be damaged. Thermal disinfection has been cited in shortening lens life and enhancing the formation of deposits on the surface. This occurs especially if the lens has not been properly cleaned prior to thermal disinfection. The method is not practical for campers who are frequently without an electrical outlet.

Chemical disinfection, on the other hand, is not as effective in killing organisms as thermal disinfection, but has several advantages: It is simple to use, thereby ensuring greater user compliance. Lens life is longer with chemical disinfection as lenses are not subjected to daily heat treatment. The method results in fewer deposit problems as surface debris left on the lens surface due to improper cleaning is not baked by heat.

The choice between thermal and chemical disinfection depends, to a large extent, on the recommendation of the lens practitioners. Factors involved include the wearer's sensitivity to preservatives, needs, personal hygiene habits, and product cost. Today fewer heat disinfection units are available.

Thermal Disinfection

Soft contact lenses were introduced in the United States in 1972. At that time, the thermal disinfection method was the only method available. It uses either preserved or unpreserved saline solution.

Unpreserved versus preserved saline solutions

Prior to the commercial availability of pharmaceutically prepared saline solutions, they were prepared by the lens wearer using salt tablets and distilled water. This method was undesirable and created many problems as the pH and osmolarity of such solutions were not controlled, which often resulted in parameter changes in some soft lenses. The major problem stemmed from noncompliance.

In order to reduce cost, tap water was often used instead of distilled water for preparing saline solutions, which resulted in mineral deposits on the lens surface. Microorganisms proliferated in nonsterile saline solution prepared with distilled water stored for a number of days and instilled in eyes directly to hydrate contact lenses while they were worn. Such gross misuses resulted in ocular infection affecting vision. By the mid-1980s products label and package inserts warned against the use of tap water for contact lens care. In the mid-1970s, pharmaceutically prepared sterile saline solution preserved with thimerosal was introduced to eliminate the disadvantages of home-made saline. However, thimerosal often caused a brownish and grayish lens discoloration and also led to red-eye and sensitization reactions in some patients. Soon after the introduction of pharmaceutically prepared preserved saline solutions, most salt tablets were withdrawn from the market. They were, however, reintroduced because of the red-eye and sensitization problems associated with thimerosal-preserved salines. In 1987, the FDA asked companies marketing salt tables for their voluntary withdrawal because of several incidences of keratitis caused by *Acanthamoeba* species, resulting in significant loss of vision or eyesight of several lens wearers. Most of these cases were associated with the use of salt tablets to prepare home-made saline solution.

In the United States today saline solutions are marketed in preserved and unpreserved forms; both are sterile and pharmaceutically prepared. The unpreserved solutions are available in unit dose or multidose plastic containers and multidose aerosols. In the latter instance, the pressurized container and the construction of the valve prevents inoculation of microorganisms into the container during its use, thus maintaining sterility more effectively in comparison to multidose nonpreserved saline solution in plastic containers. Nonpreserved saline solution in an aerosol container is, therefore, preferred over nonpreserved saline solution because it eliminates the potential of irritation and sensitization reactions caused by thimerosal and sorbic acid preserved salines. Such reactions may result in considerable patient discomfort and require temporary discontinuance of lens wear. However, the newer preservatives Polyquad and Dymed do not cause significant levels of such reactions. The disadvantage of unpreserved saline solution in unit dose is higher cost. There is also a potential risk of ocular infection as it is a common practice among lens wearers to leave their lenses in a lens case containing an unpreserved saline for several weeks, often without appropriate disinfection. Unpreserved saline does not protect against proliferation of microorganisms.

In contrast, preserved saline helps to prevent microbial growth when lenses are not worn and stored in lens cases. They are also less costly in comparison to unit dose or multidose pressurized nonpreserved saline solution containers. As already noted, the principal disadvantage of preserved saline solution is that some of the preservatives, such as thimerosal and sorbic acid, have the potential of causing irritation and sensitization in some patients. However, these reactions are not sight-threatening. In most instances these symptoms clear on discontinuance of the product without requiring any drug therapy. Again, this problem now appears to have been eliminated or greatly minimized with the introduction of newer preservatives such as Polyquad and Dymed.

Both preserved and unpreserved saline solutions are multifunctional solutions. In addition to thermal disinfection, they are also used to dissolve enzyme tablets in cleaning contact lenses, as a rinsing solution following cleaning and chemical disinfection, and as a lens storage solution.

Thermal disinfecting units

The FDA guidelines require that thermal disinfecting units must attain a minimum temperature of 80°C for 10 min in the saline solution, which is placed in the lens cases. There are several units on the market, which meet these requirements; they vary in their time-temperature profiles. Certain lenses such as those belonging to the FDA Group-4 classification (ionic, high-water-content lenses) do not withstand repeated heat treatment and tend to discolor. In general, FDA Group-4 lenses are not heat disinfected. Units on the market today have a thermostat, which cuts off the electrical current when a certain temperature is reached, eliminating the need for patients to switch it off. However, malfunctioning of units occasionally does occur.

Units on the market today are much different from those available earlier. The first unit was analogous to a "baby bottle warmer." Lenses were placed in a lens case, which was placed in a reservoir of water in an electrical heating unit. The temperature in the reservoir reached almost the boiling point of water and the time for disinfection was 30 min. To ensure not only disinfection but also sterilization of lenses, a unit was introduced capable of achieving 120°C for 20 min to ensure complete elimination of vegetative organisms as well as spores. This unit, however, was not successful since the high temperature was detrimental to many lens polymers. The current trend is toward developing a heating unit, which reaches temperatures below 80°C but ensures elimination of vegetative forms of ocular pathogens.

Chemical Disinfecting Agents (Oxidizing and Nonoxidizing)

There are two categories of chemical disinfecting solutions: The first category comprises those containing nonoxidizing chemical antimicrobial agents, which are nontoxic at the concentration level used in the products. The second category constitutes those containing oxidizing agents, which are toxic at the level used for disinfection but are degraded to a nontoxic level during the disinfection process over a course of time or by use of a second step that involves a neutralizing ingredient. Disinfecting solutions containing oxidizing agents (specifically hydrogen peroxide) gained popularity among lens practitioners in the mid- to late-1980s because of the absence of traditional preservatives, which often caused red eye and delayed hypersensitivity. However, the neutralizing solution used with hydrogen peroxide often contains preservatives that can cause ocular reactions. Lens practitioners liked the concept of hydrogen peroxide decomposing to innocuous water and oxygen. Fewer reactions have been observed with such products, but the long term toxic effects on the eye of the free radicals which are generated by low concentrations of undergraded oxidizing agents are not well known. Hydrogen peroxide products can cause severe toxic reactions if the products are not used properly and the patient inserts a lens without neutralizing the hydrogen peroxide.

The ideal chemical disinfecting agent should possess the following properties: It should have excellent wearer's acceptance in terms of being nonirritating, nonsensitizing, and easy to use. It should be relatively nontoxic compared to the earlier preservatives in terms of cytotoxicity, including its effects on epithelial and endothelial cells as well as its ability to maintain mitotic activity of corneal epithelial cells. It must have an adequate antimicrobial spectrum and be able to eliminate ocular pathogens in short lens-soaking regimens. It should not bind or bind minimally to the lens surface. It should be compatible with the lens and not cause discoloration or alter the tint of colored contact lenses.

A hydrogen peroxide disinfection system for soft lenses was tried initially in the early 1970s. The system failed to gain FDA approval because of the potential toxic nature of the chemical, the complexity of several steps, and the cost. The system was approved by the FDA in the early 1980s after having undergone significant refinements compared to the original system.

All products currently marketed in the United States under the oxidizing agent category contain hydrogen peroxide. In the international market, products containing chlorine-releasing agents are also available. These

products are generally indicated for the disinfection of hydrophilic soft contact lenses and are contraindicated for rigid gas-permeable lenses. Most of the hydrogen peroxide is decomposed by catalytic degradation (platinum ring); chemical neutralization using pyruvate, sodium bisulfite, or sodium thiosulfate; dilution and rinsing; and enzymatic neutralization (catalase).

Hydrogen peroxide, on degradation, forms water and oxygen and, hence, it is perceived by practitioners as a superior product. However, most of the neutralizers used in the second step contain preservatives such as thimerosal and sorbic acid, and stabilizers such as stannates or phosphonates. Consequently, they have the associated disadvantages of those ingredients. On the other hand, not having a preservative in a neutralizer makes these products vulnerable to microbial growth on accidental contamination. A single-step product containing hydrogen peroxide is also available on the market, which is convenient, but has the same disadvantage of not having any preservative effect at the end of the disinfection cycle. The disinfecting time recommended by various companies ranges from 10 min to an overnight soak. This category of disinfecting agents has a better antimicrobial spectrum and a faster kill rate. However, their shortcomings included toxicity if the regimen was not followed properly, complexity of use, and zero to minimal protection against microbial recontamination once the disinfection cycle was complete.

The first generation nonoxidizing chemical disinfection solution contained a combination of antimicrobial agents incorporating thimerosal with chlorhexidine or alkyltriethanol ammonium chloride. These solutions are now not commonly used because of the thimerosal problems as discussed previously. The newer antimicrobial agents, Polyquad and Dymed, because of their molecular structure and large molecule size have a better profile. These antimicrobial agents were introduced in the late 1980s and their long-term use has not caused reactions similar to those observed for older antimicrobial agents. Products with these agents currently dominate the lens care market.

PRODUCTS FOR ENHANCING SOFT CONTACT LENS WEAR COMFORT

Factors Contributing to Wear Comfort

It has already been noted that hydrogel contact lenses are inherently more comfortable than rigid (RGP or PMMA) lenses. This is related to the former's superior flexibility and hydrophilic character which permits incorporation of substantial amounts of water (38–74%) into the lens

material. However, after periods of wear time, some lenses may experience changes in hydration, that may be related to deposits, environmental (e.g., temperature and humidity) changes and improper care. In particular, "dry spots" may become evident on the lens with attendant reduction in comfort and visual acuity. When this happens, the wearer may benefit from periodic administration of rewetting (or soothing or comfort) solutions onto the lens while being worn. These solutions are usually low-viscosity aqueous compositions containing polymers or surfactants, which enhance the wettability of the surface, facilitating the spreading tears, and enhancing the stability of the tear film. They may also provide cushioning and lubricating actions, lessening impact, and reducing the frictional forces of the eyelids as they move across the lens on the corneal surface. The frictional forces would be especially important in instances where deposits or debris on the lens are present in sufficient amounts to cause physical irritation to the ocular tissue. In addition to wetting, cushioning, and lubricating, the ability of the solutions to facilitate removal of contaminants and retard further soilage are also desirable attributes. It is also desirable, in terms of convenience to the lens wearer, that the frequency of administration of drops of the above-mentioned types be minimal. Therefore, the use of polymers and surfactants that associate with the (deposited) lens and resist removal by the rinsing action of the tears is called for (i.e., polymers and surfactants with good substantivity).

Although rewetting efficacy is usually the primary requirement for hydrogel lens wearers, all three actions (rewetting, cushioning, and lubrication) may be of considerable significance for RGP or PMMA lens wearers. The practitioner should seek to help the wearer find the product best suited for his or her specific comfort needs. Solutions designed for hydrogel lenses generally have lower viscosities, whereas solutions for rigid hard lenses usually have higher viscosities. Although high viscosities can be of distinct benefit in enhancing cushioning action, a solution that is too viscous can cause blurred vision and hinder normal lid movement to an undesirable extent.

Components of Lens Comfort Solutions

Among the polymers used in lens comfort solutions are polyvinyl alcohol, polyvinylpyrrolidone, dextran, and various cellulose derivatives such as hydroxyethyl cellulose, hydroxypropyl cellulose, and hydroxypropyl methylcellulose. Surfactants include certain poloxamer and poloxamine compounds. Other normal components comprise appropriate preservative(s) as well as buffering and tonicity-adjusting agents.

LENS CARE PRODUCTS FOR RIGID GAS-PERMEABLE LENSES

Several types of RGP lenses are marketed (Table 2) and others are under development. Because of their rigid nature, they have certain characteristics in common with conventional hard PMMA lenses. In many cases, care products available for use with the latter have been found to be also suitable for the care of the RGP lenses. Although they may not have optimal characteristics, they are still preferred by many practitioners in comparison to products that were originally designed for use with soft lenses and were subsequently approved by the FDA also for use with RGP materials.

Most categories of products indicated for the care of soft contact lenses are applicable to rigid gas-permeable lenses. These include daily cleaners, wetting and cushioning drops, and weekly enzymatic cleaners. While disinfecting solutions are necessary for RGP lens care, they are often positioned as conditioning solutions. Conditioning is important in providing a hydrophilic lens surface during wear. Because RGP lenses do not withstand heat, saline solutions (both preserved and unpreserved) are not needed for these lenses, except for dissolving enzymatic cleaners for weekly cleaning or rinsing.

Because RGP lenses are not as porous or water absorbing as soft lenses, wearers do not experience the problems that are specific to soft lenses, resulting from preservatives penetrating and concentrating within the lens polymer matrix, which often cause toxic and hypersensitivity-type reactions. However, certain preservatives bind with RGP surfaces and can create clinical problems. Binding may involve ionic and/or hydrophobic interactions.

In addition to accumulation of proteinaceous deposits, such as those occurring on soft hydrophilic lenses, the molecular make-up of many RGP lenses also tends to attract lipid deposits, such as cholesterol esters, wax esters, triglycerides, etc. This is especially true of the more hydrophobic materials such as silicone acrylates with high Dk (oxygen permeability) values. Accordingly, more recent developments in material science related to contact lenses have resulted in materials such as fluorosilicone acrylates and fluorocarbons with purportedly less propensity for deposits.

Wearing rigid lenses is much less comfortable than wearing soft contact lenses. In fact, this is generally perceived to be the major factor limiting the growth of the RGP lens market segment. Consequently, the need for superior wetting, cushioning, and lubricating products is clearly recognized. Superior combination products are also

required for RGP lenses, which provide convenience and enhance product performance.

MULTIPURPOSE SOLUTIONS

Multipurpose solutions are designed to increase wearer compliance and the convenience of product use. Such solutions are not commonly used for the conventional PMMA hard lenses and rigid gas-permeable lenses. They combine two or more basic functions of lens care, including cleaning, disinfection, soaking, wetting, and lubricating. Combination of these functions in a single product may compromise certain aspects of product efficacy. For instance, a solution designed to clean and disinfect may not clean as well the cleaner would alone. However, a combination cleaning and disinfection solution provides the convenience of a single step and would be particularly useful for wearers whose lenses do not attract deposits as readily because of their tear chemistry.

Multipurpose solutions for soft lenses are primarily limited to cleaning and disinfecting. Wetting and lubricating combinations are not a major need because of the inherent hydrophilic nature of soft lenses that makes them comfortable to wear. The technology advances made in identifying preservatives with broader spectrum and capacity along with the ability of chemicals to clean while disinfecting has further simplified care of lenses. Recently, a new product has been developed in a multipurpose solution format that has allowed removal of the rubbing step.

PACKAGING OF CONTACT LENS PRODUCTS

All lens care products are packaged in plastic containers or pressurized metal containers with the exception of enzymatic or disinfecting tablets. The tablets are generally effervescent and packaged in laminated foil or blister packs to ensure adequate shelf life. The packaging materials normally used by the pharmaceutical industry for effervescent tablets are adequate for this purpose. The plastic containers are usually fabricated from low-density polyethylene, high-density polyethylene, or polypropylene materials. Many of the containers are opaque for protection against light. Several types of colorant mixtures, which usually contain titanium, are used in squeeze bottles, and the composition and thickness of the container wall should be designed to allow easy delivery. Another important

consideration is the orifice at the tip, which allows the desired product to flow. For instance, a disinfecting or saline solution needs to be delivered in a large volume of 3 to 5 ml to fill the lens case. A steady stream is desirable and acceptable here, but not for products like wetting, lubricating, and cushioning drops, which are directly instilled in the eye. For these products, the tip has to be designed to allow drop-by-drop instillation. Formulation characteristics such as viscosity and surface tension determine the tip design. The tip must be smooth and rounded as it can come into contact with the eye. The caps for the bottle are normally constructed of polystyrene or polypropylene material.

Because all lens care products, with the exception of the tablet dosage form, have to be sterile, the containers must be sterilized prior to filling unless the process involves form, fill, and seal technology. Containers are usually sterilized by ethylene oxide or gamma irradiation. The latter method is preferred because of stringent government regulations and requirements regarding ethylene oxide residues and its degradation products. Recent years have also seen strict controls regarding the exposure of workers to ethylene oxide and its by-products. Terminal sterilization of the final product is normally not done for lens care products, with the exception of nonpreserved saline solutions in aerosol containers, which are sterilized by gamma irradiation.

ACCESSORY CONTACT LENS PRODUCTS

Accessory contact lens products include cases and devices for cleaning contact lenses and facilitating insertion.

Lens Cases

Lens cases are utilized for disinfection and storage while the lenses are not being worn. Lens cases may have a single compartment with a barrel shape, generally holding 7 to 10 ml of solution, or they may have two compartments in a flat case design, each compartment holding 3 to 5 ml of solution. Lens cases are also used by lens manufacturers as mailers to ship RGP and PMMA lenses in the dry state. Recently, shipping the lenses in conditioning solution has been approved. On the other hand, hydrophilic soft contact lenses are generally shipped in sealed sterile glass vials containing buffered isotonic saline solution. Plastic lens cases are fabricated from polymeric materials such as polyethylene, polypropylene, polysulfone, or polycarbonate.

Devices Facilitating Cleaning

Cleaning devices are useful for lens wearers who lack the manual dexterity to clean lenses with their fingers. They are also used for cleaning lenses that could be scratched easily by fingertips. The consumer versions provide mechanical action by manual swirling or by agitation with the help of an electrical motor or sonication. The professional versions are usually ultrasonic-type devices. None of these devices are capable of removing tenaciously bound protein. The reservoir of the cleaning device is filled with special cleaning fluid or a saline with a few drops of a surfactant-type daily lens cleaner. The device should allow easy lens placement and retrieval and minimize the potential of damage. Lens baskets that allow for rinsing after the cleaning regimen without additional handling are a desirable feature.

Devices Facilitating Lens Insertion

These devices are generally helpful to elderly patients, especially those wearing aphakic lenses. They consist usually of a rubber bulb with a suction cup. They can cause severe corneal damage if used improperly as well as ocular infection if they are not properly cleaned and disinfected.

Future Directions in Lens Care Products

Of the persons fitted with contact lenses (both soft and hard), 40% discontinue wearing them within the first three years. One of the reasons cited is the time and effort required in taking care of them. Therefore, lens care systems will be developed that are simple and more convenient to use, requiring fewer products, and less time. The new products will minimize patient problems through safer preservatives, which are nontoxic even on misuse and do not produce hypersensitization. More effective cleaners as well as products which minimize protein and microbial attachment to lens surfaces will also be forthcoming as well as cleaners that can be instilled in the eye during lens wear to retard deposit formation on extended-wear lenses or to clean lenses while they are being worn. Other future product types may include special artificial tears and comfort drops for older patients who are prone to dry eye. There is also a need for diagnostic products that can detect potential problems before they are manifested clinically. Universal products that could be used with all contact lenses (i.e., hard PMMA, RGP, and soft lenses) may be desirable; however, it is unlikely that such products will be available as different contact lenses vary significantly in their chemical and surface characteristics. The main

emphasis in the future will be on the development of convenient, easy-to-use products that can increase patient compliance and reduce the dropout rate, while ensuring desirable efficacy without compromising product safety. Products specifically designed for the care of disposable and frequent replacement lenses will also be forthcoming.

COMPONENTS OF LENS CARE PRODUCTS

Active Components

Active components play a primary role in the intended use of a product. The active components in lens care products are usually limited to either one or two chemical entities. However, a product may contain several active components if it is designed for multiple functions or indications.

Chemicals with disinfecting capability

Active chemical entities must have bactericidal and fungistatic properties. However, cidal properties for fungal organisms is preferable. The product performance criteria encompassing the types of organisms to be tested, levels of inoculum, and method of testing are defined by guidelines developed by the FDA. Current chemical systems have disinfection times of 4–6 h. This fits with most wearing schedules, because more than 90% of patients remove their lenses overnight between wear periods. Overnight removal seems to help the eye recover from the stress of lens wear. Recently, some efforts to move to shorter disinfection time are under way to increase convenience. The products will require dramatically enhanced disinfection efficacy than currently available or strict compliance to a regimen to achieve the required level of disinfection.

Surface-active and other agents with cleaning capability

Surfactants of various types have been traditionally used for cleaning conventional PMMA hard lenses. They are considered effective in removing surface deposits on these lenses, such as cosmetics, hair spray, mascara, etc. However, they are not effective in removing tenaciously bound deposits (e.g., proteins, lipids, lipoproteins, mucoproteins) that are commonly encountered with soft hydrophilic contact lenses and rigid gas-permeable lenses. Such deposits are more effectively removed by products containing enzymes, strong oxidizing agents, or suspended abrasive (deposit-shearing) particles. The types of surfactants used and representative products on the market are discussed later under Products for

Cleaning Soft Contact Lenses. A few other agents such as citrate and some phosphates have been used as cleaning agents.

Components with wetting, lubricating, and cushioning capabilities

Polymers and surfactants are the two main classes of compounds used as wetting, lubricating, and cushioning agents. Contingent on the nature of the specified polymers, they are used in various combinations to achieve desired product characteristics.

Both synthetic and natural polymers are commonly used in lens care products. These agents can provide a cushioning effect as a result of increased viscosity. The need for products with such action is greater for the wearers of conventional hard PMMA lenses and rigid gas-permeable lenses, for these are inherently not as comfortable as soft contact lenses because of the lens design and the physiochemical nature of the polymers. Viscosity-building agents provide the necessary initial coatings on the lens surface before it is coated by natural tear components. Besides contributing to viscosity, the ability of polymers to adsorb on lens surfaces and to elicit surface-active (wetting) properties are important considerations.

Surface-active agents are adsorbed on the lens surface and allow ready spreading of tears when the lenses are inserted, thus making them more comfortable to wear. The use of surface-active agents to impart wettability to the lens surface is of lesser value for products used for hydrophilic soft contact lenses, which have built-in wetting characteristics because of their water content. Addition of surfactants may be of value, however, in retarding deposit formation or for cleaning while the lenses are worn. The methodology for measuring both advancing and receding contact angles, as an indication of wetting efficacy, has been standardized for contact lenses. Various polymers have different degrees of wettability as measured by contact angle using a goniometer. However, the *in vitro* contact-angle measurement of contact lenses made from various polymers is of little clinical value. No discernible differences in *in vivo* contact angle can be detected on insertion of a contact lens in the eye following a few blinks that result in coating the lens surface with tear components. Surfactants are usually combined with polymers to impart substantivity and cushioning characteristics. The types of polymers and surfactants used in representative products in the market are discussed later under Products for Enhancing Soft Contact Lens Wear Comfort. Again, the need for surface-active agents which can facilitate wetting of lenses and spreading of tears is greater for

PMMA and rigid gas-permeable lenses because of their hydrophobic surface characteristics.

Ancillary Components

Preservatives

Preservatives are used in almost all multidose contact lens products. Because the potential for misuse of products by lens wearers is significant, preservatives prevent the potential proliferation of microorganisms. A contaminated product could ultimately lead to an ocular infection and possible loss of eyesight. Preservatives are considered active components when incorporated in products for the purpose of disinfecting contact lenses. The preservative must possess the antimicrobial activity described in the FDA guidelines for contact lens care.

The FDA has issued guidelines for preservative efficacy, which include an additional safety factor to compensate for potential misuse of products. These requirements are more stringent than the requirements of the *United States Pharmacopeia* (USP) for preservative efficacy. The FDA requires rechallenging the preserved products on day 14 with a defined inoculum of microorganisms; this is not required by the USP. The preservatives generally used are mainly the same as those used for thermal or chemical disinfection of contact lenses. The concentrations could be different than in the chemical disinfection of contact lenses. The concentration is generally higher in cleaners and comfort-enhancing solutions because of possible binding with polymer and surfactant components, which result in a decrease in preservative efficacy.

Different countries have different standards for preserving contact lens solutions; the requirements of the *British Pharmacopoeia* are the most stringent. Many marketed products do not meet these requirements. The three major preservative efficacy tests, which must be considered in developing products, are delineated in the United States Pharmacopoeia (modified test to meet FDA guidelines), the *British Pharmacopoeia* and the *German Pharmacopoeia* (DAB). In terms of difficulty to comply, the *British Pharmacopoeia* test is the most stringent and the modified USP test the least stringent. The main differences among the test requirements are the exposure times and use of *Escherichia coli* as a challenge organism. The *British Pharmacopoeia* utilizes a 6-h criterion for antimicrobial activity, whereas the USP uses a 14-day criterion. Recently, an International Standard has been approved which is comparable to the US FDA procedure. The types of preservatives used are

described in the tables related to various product categories.

Buffers for adjusting pH

The use of buffers and pH adjustment is an important consideration in lens care products. It is a general practice that all products which are likely to come in direct contact with ocular tissues should be buffered for ocular comfort around physiologic pH and preferably in the range 6–8.0. The most commonly used buffers in contact lens care products are phosphates and borates. Buffers used occasionally are acetate, citrate, and others. Besides buffers, sodium hydroxide and hydrochloric acid are generally used to achieve a desirable pH in the final product. They are also used to adjust the final pH in products, which do not have any buffering system. The selection of an appropriate buffering system should consider the pH necessary for optimal performance of the product, as well as products stability and potential incompatibility with other components of the product.

Although it is desirable to have a product as close to physiologic pH as possible, it is often essential to formulate a product outside the physiologic pH range in order to achieve the desired stability of the product, optimal efficacy, or appropriate solubility of active and ancillary components. Products formulated outside the physiologic pH range should have low buffer capacity to allow quick equilibration to tear pH by the bicarbonate buffer system present in the tears. The maintenance of pH close to physiologic pH is essential for products intended for soaking and disinfection of hydrophilic soft contact lenses in order to maintain the parameters of some lenses, especially those belonging to FDA Group 4. Such changes in parameters can cause not only discomfort but can also produce blurred vision.

Tonicity-adjusting agents

Contact lens products should be formulated as closely as possible to the tonicity of tears. This is important for optimal comfort. Prolonged exposure to hypotonic solutions can induce edema in corneal epithelial cells, which can cause blurred vision and discomfort, whereas prolonged exposure to hypertonic solutions can cause corneal epithelial cells to shrivel and cause discomfort by exposing nerve endings. None of the currently available contact lens solutions have been responsible for such symptoms as most of them are formulated close to isotonicity and many contact lens products have minimal contact with the cornea. Nevertheless, maintaining the tonicity of products close to the isotonic value of 280 (± 50) mOsm/kg is important for optimal comfort as well as for maintaining the integrity of certain hydrophilic soft

contact lenses (especially those belonging to FDA Group 4). The type of compounds commonly used for imparting isotonicity include buffering agents, sodium chloride, potassium chloride, propylene glycol, mannitol, dextrose, etc.

Viscosity-building agents

Viscosity-building agents such as synthetic and natural polymers are used as active ingredients in solutions providing comfort and rewetting products. However, they are also used as ancillary agents in contact lens cleaning products. These agents allow better cleaning of contact lenses by enhancing the contact of the cleansing agent with the soiled lens and by facilitating the process of rubbing the lens between fingers or between the palm of the hand and an index finger. Highly viscous or gel-type cleaners are also available. However, they are not very popular as they are difficult to rinse and may cause ocular irritation. The types of polymers used as ancillary agents are the same as those used for solutions providing comfort or rewetting solutions.

PHARMACEUTICAL TECHNOLOGY CONSIDERATIONS IN PRODUCT DESIGN

With few exceptions, contact lens products are sterile solutions. The sterility requirements are important because of the potential of sight-threatening ocular infections. The sterility test procedures and pass–fail criteria as described in the *United States Pharmacopeia* must be met for FDA approval of contact lens products. The technology practiced in the development of pharmaceutical products, such as injectable and large volume intravenous fluids, is generally acceptable, with somewhat less stringent requirements for contact lens care products because most of these products do not come into direct contact with the eye (i.e., its interior cells and fluids). Products that are packaged in unit dose containers or multidose pressurized aerosol containers do not have to be preserved. However, multidose products other than aerosol containers should be preserved and should pass the FDA preservative efficacy test. Manufacturers intending to develop lens care products should consult with the FDA for the latest guidelines.

Apart from solutions, there are other pharmaceutical dosage forms for contact lens products that are less commonly used. They include gel- and suspension-type cleaners, powders, and tablet dosage forms for enzymatic cleaners or disinfectants. All of these products are formulated as sterile products with the exception of powders and tablets. These, however, must comply with

USP requirements for bioburden with added specifications for absence of colony-forming units of *Staphylococcus aureus* and *Pseudomonas aeruginosa* among other organisms. The choice of components in formulating these products is much more restrictive in comparison to pharmaceutical products intended for systemic or topical use because of the potential of causing irritation to the sensitive ocular tissues.

Chemical components must be nonirritating and compatible with ocular tissue as well as with lenses. The choice of components for tablets, such as binding and lubricating agents, is much more restrictive, as these ingredients must be soluble and form a clear solution when dissolved in saline solutions. Formulation components generally used and acceptable for ophthalmic products may be unacceptable for contact lens products, as many contact lens materials can concentrate components used in contact lens products as much as several hundredfold. On insertion of the lenses, these ingredients are released in toxic levels to the cornea, causing minor to severe ocular reactions. The binding and release of ingredients can be further complicated by the condition of the lens. For instance, thimerosal and chlorhexidine, commonly used ingredients in contact lens products, concentrate differently in new, deposit-free lenses as compared to used lenses with protein deposits on the surfaces. Appropriate studies must be designed to address these issues to minimize the potential of ocular irritation. Certain ingredients such as thimerosal, when used initially are well tolerated. However, on prolonged exposure, some lens wearers develop hypersensitivity, resulting in intolerance of thimerosal products. Such intolerance is difficult to predict and currently no satisfactory method is available to predict delayed hypersensitivity-type reactions, which are modulated immunologically. The commonly used guinea pig maximization test is not predictive of delayed hypersensitivity reactions.

As noted previously, formulating products close to physiological pH is desirable. However, for optimal product performance or meeting regulatory requirements, it is sometimes necessary to formulate products outside the physiologic pH range. Sorbic acid, a commonly used preservative for lens care products, is a marginally effective antimicrobial agent with a pKa value of 4.8. A product fails the FDA preservative efficacy test if it is formulated around a physiologic pH of 7–7.4 at a concentration of 0.1 sorbic acid normally used in marketed products. To maximize its antimicrobial activity without compromising ocular comfort to a significant extent, it is necessary to formulate such products around pH 6.5–6.8. Thus, a consideration of the dissociation constants of preservatives is essential as the antimicrobial activity of

many preservatives depends on the undissociated species, which should be present in an adequate amount at physiological pH without causing ocular irritation.

The product design should also consider the nature of the lens polymer and its surface charges. For example, FDA Group-4 lenses, which carry negative surface charges, can react with positively charged product components, resulting in severe ocular toxic reactions. Such toxic reactions can be prevented by incorporating nonionic surfactants to form micelles with the cationic components, thus minimizing surface interactions and toxicity. If a cationic agent in the product is a primary disinfecting agent, efforts to minimize toxicity or surface interactions often result in reduced antimicrobial performance. It is, therefore, essential to minimize surface interactions and yet have a product, which meets the requirements for disinfection or preservation.

The hydrophilic–lipophilic balance (HLB) and molecular dimensions of a preservative should also be considered in the design of contact lens products. For example, chlorobutanol (a nonionic preservative used in ophthalmic products) penetrates into the matrix of rigid gas-permeable (RGP) lens materials because of the molecule's substantial lipophilic characteristics. The preservative present in the lens matrix may change the parameters or may be gradually released during wear, causing irritation and toxicity to the corneal cells. On the other hand, the molecular dimensions of a preservative are very important in designing a product for hydrophilic, soft contact lenses. Generally, preservatives with a high molecular weight and appropriate molecular configuration are less likely to penetrate the porous matrix of hydrogel lenses. Products containing preservatives like Dymed and Polyquad, with molecular weights above 1000, are currently available.

ACKNOWLEDGMENT

The authors wish to acknowledge the unstinting assistance of Cathy Hughes in the preparation of this article.

BIBLIOGRAPHY

- Bailey, N.J. Contact Lens Spectrum **1987**, 2 (7), 6–31.
- Bennett, E.S.; Grohe, R.M. *Rigid Gas-Permeable Contact Lenses*; Professional Press Books/Fairchild Publications: New York, 1986.
- British Pharmacopoeia 1988*; [Appendix SVIC (Efficacy of Antimicrobial Preservatives in Pharmaceutical Products)] Her Majesty's Stationery Office: London, 1988; 2, A200–A203.

- Chowhan, M.; Bilbault, T.; Quintana, R.P.; Rosenthal, R.A. *Contactologia* **1993**, *15*, 190–195.
- US Patent, 5,370,744, Alcon Laboratories, Inc.
- CLMA Standards Subcommittee, Technical Affairs Committee (Q.A. Cappelli, chair). Industry Accepted Reference Procedure for Determining Wetting Angle **1979**, *4* (4), 35–39, Contact Lens Manufacturers Association (CLMA), Chicago [cf. *Contact Lens Forum*].
- Dabezie, O.H., Jr. Ed. *Contact Lenses: The CLAO Guide to Basic Science and Clinical Practice*; Little, Brown and Company: Boston, 1989; 1 and 2.
- Deutsches Arzneibuch*; [Anhang VIII.N1 (Prüfung auf ausreichende Konservierung)] 9. Ausgabe, Deutscher Apotheker Verlag: Stuttgart, 1986; 369–370.
- Food and Drug Administration, Draft Testing Guidelines for Class III Soft (Hydrophilic) Contact Lens Solutions United States Food and Drug Administration, Silver Spring, MD, 1985.
- Food and Drug Administration, Guidance Document for Class III Contact Lenses United States Food and Drug Administration, Silver Spring, MD, 1989.
- Gossel, T.A.; Wuest, J.R. Contact Lenses and Lens Care Products. In *Handbook of Nonprescription Drugs*; Feldmann, E.G., Ed.; American Pharmaceutical Association: Washington, 1990, 601–631.
- Harris, J.K. Solutions for Cleaning, Disinfection, and Storage. In *Contact Lenses*; Aquavella, J.V., Rao, G.N., Eds.; J.B. Lippincott Company: Philadelphia, 1987; 226–263.
- Hartstein, J., Ed. *Extended Wear Lenses for Aphakia and Myopia*; The C.V. Mosby Company: St. Louis, 1982.
- Hartstein, J.; Swanson, K.V.; Harris, C.R. *Contemporary Contact Lens Practice*; Mosby-Year Book, Inc.: St. Louis, 1991.
- Holly, F.J., Ed. *The Preocular Tear Film in Health, Disease, and Contact Lens Wear*; Dry Eye Institute, Inc.: Lubbock Texas, 1986.
- Houlsby, R.D.; Ghajar, M.; Chavez, G. J. Am. Optom. Assoc. **1988**, *59*, 184–188.
- Kreiner, C.F. *Kontaktlinsen-Chemie*; Median-Verlag: Heidelberg, 1980.
- Lippman, J.I. *CLAO J.* **1990**, *16*, 287–291.
- Lowther, G.E. Preparations Used with Contact Lenses. In *Clinical Ocular Pharmacology*; Bartlett, J.D., Jaanus, S.D., Eds.; Butterworth's: Boston, 1989; 337–353.
- MacKeen, D.L. Am. Pharm. **1986**, *NS26* (10), 27–31.
- Mandell, R. *Contact Lens Practice*; Charles C. Thomas: Springfield, IL, 1981.
- Philips, A.J., Stone, J., Eds. *Contact Lenses: A Textbook for Practitioner and Student*; Butterworth's: Boston, 1989.
- Randeri, K.J.; Quintana, R.P.; Chowhan, M.A. Contact Lenses Cleaning. In *Contact Lenses: The CLAO Guide to Basic Science and Clinical Practice*; Kastl, P.R., Ed.; Kendall/Hunt Publishing Company: 1995; 215–236.
- Ruben, M. *Color Atlas of Contact Lenses & Prosthetics*; The C.V. Mosby Company: St. Louis, 1989.
- Tripathi, R.C.; Tripathi, B.J.; Silverman, R.A.; Rao, G.N. Int. Ophthalmol. Clin. **1991**, *31*, 91–120.
- The United States Pharmacopoeia*, *24*, *NF19*; United States Pharmacopeial Convention, Inc.: Rockville, MD, 2000; 1809–1823.
- White, P.; Scott, C. Contact Lenses & Solutions Summary. In *Supplement to Contact Lens Spectrum*; Viscom Publications, Inc.: Norwalk, CT, Feb. 1991.
- Yeager, M.D.; Benjamin, W.J. Int. Contact Lens Clin. **1987**, *14* (8), 61.

LIPID EXCIPIENTS IN PHARMACEUTICAL DOSAGE FORMS

Alan L. Weiner

Alcon Research, Ltd., Fort Worth, Texas, U.S.A.

THE ROLE OF LIPIDS

Biologically, lipids function as structural elements in plants and animals, transport vehicles, mediators of chemical reactions, energy sources, and as messenger molecules. It is from an understanding of these functions that lipids have been applied to pharmaceutical applications. Lipids can be appropriately manipulated to conform to various physical states such as solutions, suspensions, emulsions, creams, gels, or solids. Because a combined lipid-drug formulation will present a very different conformation and/or molecular size than the drug alone when introduced into a biological environment, changes in both the distribution and the recognition of the drug by the host defense system will occur. These changes may be either intended or undesired.

Overall, the uses of lipids in pharmaceutical dosage forms can be grouped into four categories: 1) improvement in the processing or stability of the formulation in the preferred physical state; 2) enhancement or reduction in cellular or systemic absorption of the drug from the formulation; 3) more effective drug targeting; and 4) sustained or more controlled delivery of the drug. In this article, the problems and opportunities in utilizing lipids for these applications will be explored from various aspects of pharmaceutical formulation, analysis methods, and other commercial issues.

CATEGORIES OF LIPIDS IN PHARMACEUTICAL PREPARATIONS

The principal material categories discussed in this article are illustrated in Fig. 1. Categories shown are: fatty chain acids, salts, alcohols or amines, oils and waxes, phospholipids, glycolipids, neutral lipids, and nonlinear chain compounds, such as sterols.

Fatty Acids and Derivatives

Fatty acids, salts, alcohols, and amines have been utilized in pharmaceutical formulations. Fatty acids are present in cosmetics, ointments, and suppositories and are used in

tablet coating applications and as carriers in inhalant products. Fully saturated fatty acids are solid materials at chain lengths above eight, whereas longer chain polyunsaturated forms may exist as liquids, unless the double bonds are conjugated. Fatty acid salts are used widely in tableting applications and often include magnesium, calcium, and aluminum stearates. The handling properties and physical interaction with other solids make them ideal as conditioning agents to effect even distribution of particles, improve compressibility, and control the eventual release of active agent. The fatty alcohols, including cetyl and palmityl alcohol, are well known and are used extensively in ointments or creams as an emollient or emulsion modifier. Fatty amines are generally utilized as precursor compounds in coupling reactions to produce lipophilic drug derivatives. Fatty amines, like the acids, are solid at higher chain lengths and insoluble in aqueous solution. Finally, aldehydes of fatty chain compounds have application in fragrances and flavorings. The unsaturated liquid forms of fatty aldehydes are most often employed.

Oils, Waxes, and Neutral Lipids

Fatty acids, when esterified to glycerol, form mono-di- and triglycerides. Depending on the number of substituted fatty acids, the carbon chain lengths and the unsaturation level, the resultant product may exist in either liquid state (oil) or solid state (wax). Products in this category may either be naturally derived or synthetically produced and, thus, a wide diversity of products is commercially available. Oils have drug-carrying and solubilization functions in oral, topical, or injectable products. Waxes are presently used in topical and oral preparations to improve desired physical properties and control dissolution of the final product. Common materials in this group that are present in commercialized pharmaceutical preparations include castor oil, hydrogenated vegetable oil wax, paraffin, carnauba wax, white wax, olive oil and olive oil ethyl ester, mineral oil, petrolatum, cetyl ester wax, and beeswax. Neutral lipids, such as steroids, are also of value in pharmaceutical practices. For example, cholesterol and cholesterol esters can be exploited to broaden liposome phase transitions and to allow easier manipulation and or emulsification.

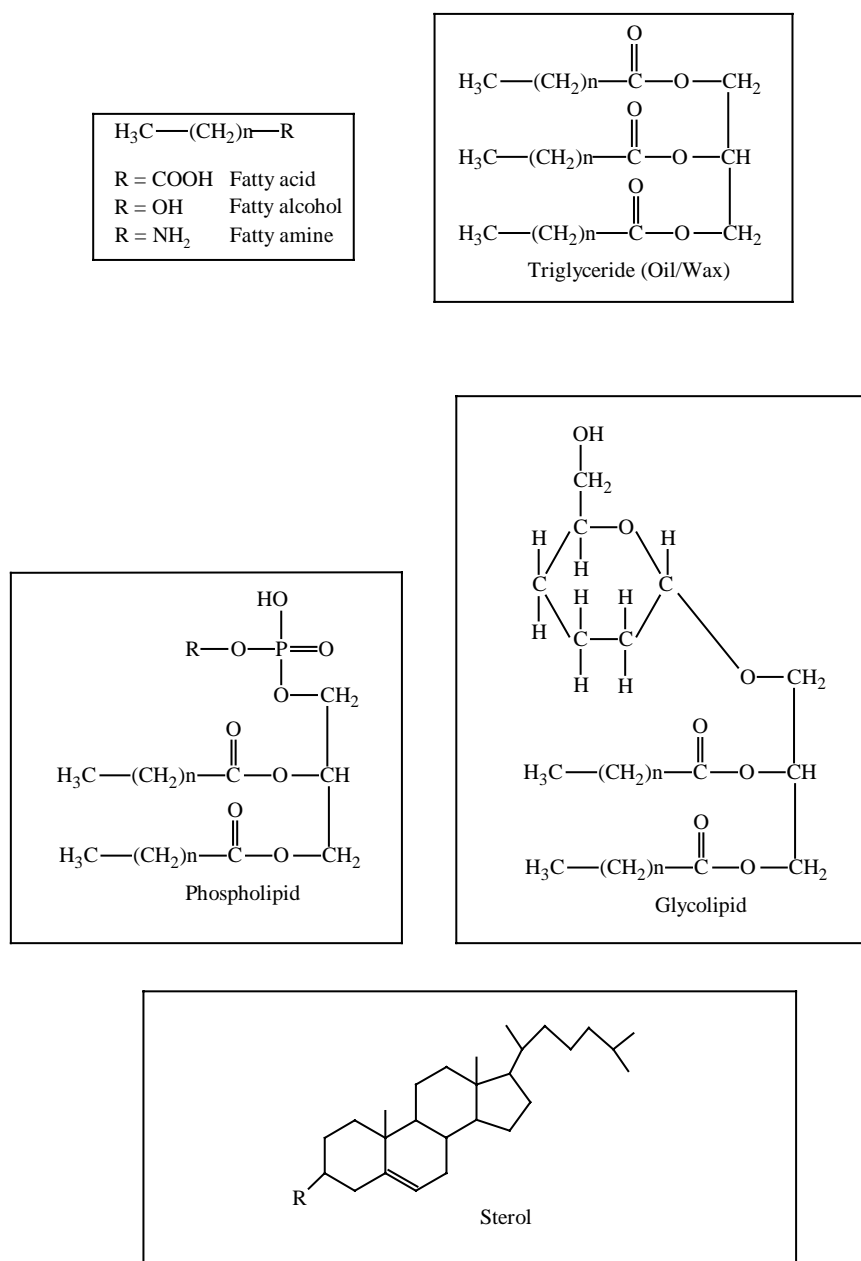


Fig. 1 Lipid structures.

Phospholipid Compounds

The complexity of glycerides advances by modification of the terminal hydroxyl with phosphate linked head groups to form phospholipids. Common phospholipid head groups include choline, ethanolamine, serine, inositol and inositol phosphates, glycerol, and glycerol esters. As with the triglycerides, numerous species are possible by various

combinations of different headgroups and fatty acyl substitution at the 1st and 2nd positions of the glycerol backbone. Fluidity differences are evident as a function of the gel to liquid crystalline transition temperatures (Table 1). Solubility of phospholipids is intimately linked to the conformation of the aggregate material rather than strictly a chemical function of the molecule. Monoacyl phospholipids, which tend to form micelles, are usually more readily

Table 1 Transition temperatures of common phospholipids

Head group	Fatty acid component	Transition temperature (°C)
Hydrogen (phosphatidic acid)	Dimyristoyl	51
	Dipalmitoyl	67
Choline	Dioleoyl	−22
	Natural (mixed unsaturated)	−15
	Dilauryloyl	−1.8
	Dimyristoyl	23
	Dipalmitoyl	41
Ethanolamine	Distearoyl	55
	Dimyristoyl	50
	Dipalmitoyl	66
Glycerol	Dioleoyl	−18
	Dilauryloyl	4
	Dimyristoyl	23
	Dipalmitoyl	41
	Distearoyl	55
Serine	Natural (mixed unsaturated)	7
	Dimyristoyl	38
	Dipalmitoyl	51
Sphingomyelin	Natural	32
	Dipalmitoyl	41
	Distearoyl	57

soluble in aqueous solution. Diacyl phospholipids generally form liquid crystalline suspensions as long as the temperature is held at or above the phase transition. Overall, because phospholipids are amphipathic, they function well as emulsifying or dispersing agents. Although most often found in topical products, phospholipids have been employed in oral capsule formulations as well as in liposomal parenteral formulations.

Glycolipid Compounds

Glycolipids include compounds formed by the linkage of a sugar moiety to a glyceride backbone. Classical glycolipids such as glycosyl glycerides or ceramides are not common in pharmaceutical preparations primarily due to low abundance and high cost; however, glycosyl modifications can be used to impart targeting mechanisms to aggregate lipid formulations.

LIPID CONFORMATION

The colligative and solubilizing properties of lipids have been known since the early part of the 20th century. What had been perceived as clusters of amphipathic

molecules (soaps) in aqueous solution are now better understood as micelles (1–3). Micellar solubilization of pharmaceuticals is now a familiar classical technique, which has been reviewed in detail (4, 5). Additional understanding of drug dissolution and impact of lipid conformations can be gained from studies that utilized macro and microemulsions to solubilize drugs (5–9).

Beyond the concept of micelles and emulsions, the combination of lipid and aqueous components can result in formulations that exhibit a variety of physical states. These result from lipid orientations, which may be either random, in thermotropic crystalline lattices, or in aggregate structures with long-range order (lyotropic liquid crystals) (10). Figure 2 illustrates possible lipid conformations with micellar and emulsion/surface monolayer orientations being the earliest of recognized microstructures. Lipid molecules are also capable of packing tightly together to give more ordered states, such as a bilayer. In fact, depending on the components in the formulation, lipids may contribute to distinct phase diagrams containing up to at least six different regions (11). The “neat” or lamellar (liposomal) phase (12) physically can appear as a suspension to a highly viscous cream (13), depending on the final volume. Lamellar organizations further include liquid crystalline,

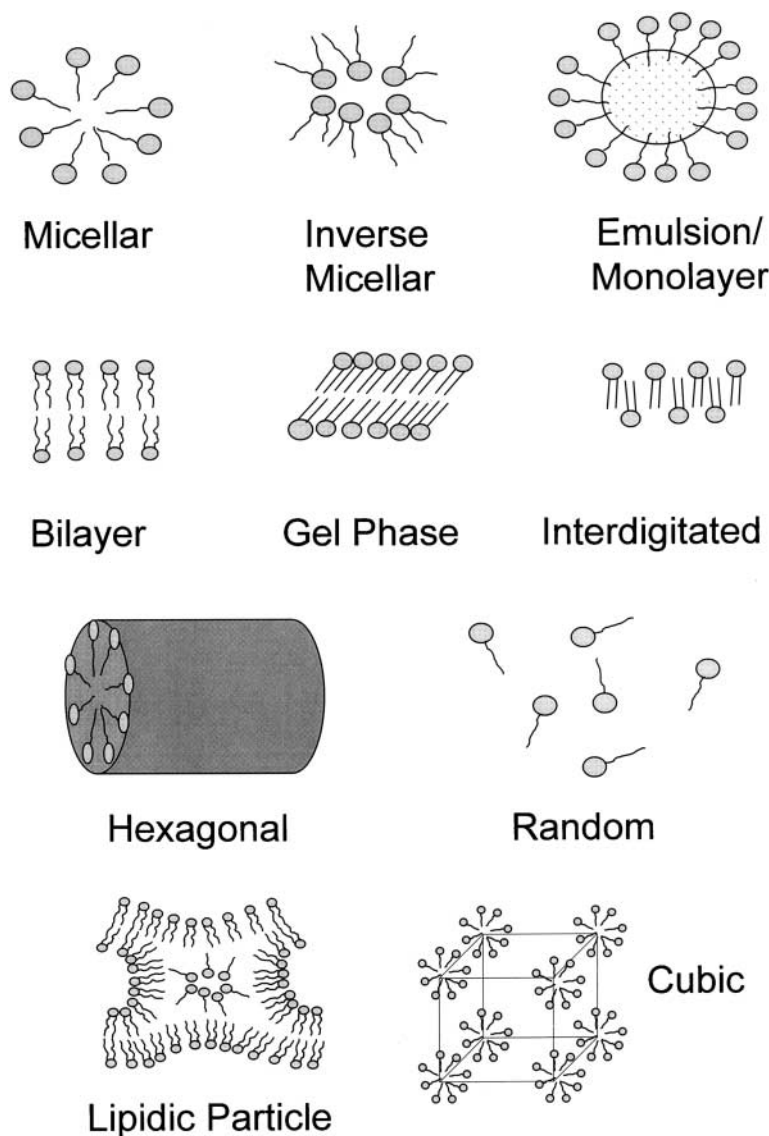


Fig. 2 Conformational states and spatial organizations of lipids. Small circles indicate the polar head group region of the molecules, whereas hydrophobic chains are represented as single lines.

gel state, and interdigitated phases. The middle phase (hexagonal) is a cylindrical arrangement with a hydrocarbon core, surrounded by an interfacial layer of hydrated polar groups. This orientation can be reversed to yield the hexagonal II structure, which appears stiff, and suspends as hard particles in water and will not dilute in additional aqueous solution. Lipids can also organize into cubic structures that appear clear, brittle, and viscous in character (14). Mixed conformations of lipids have also been observed, as is the case for lipidic particles, with inverse micelles distributed between a lamellar matrix.

These different aggregated states are sometimes reached by transitions through other phases simply by changes in temperature or by nature of the external environment (such as solvent concentration). Phase transitions of lipids can be regulated by changes in pH, temperature, cation concentrations, and presence of amphipathic additives. A comprehensive database on lipid phase transitions has been recently compiled by Caffrey (15). Polymorphic transitions are common in the preparation of lipid-containing formulations and can account for the differences in physical handling properties of the material during manufacture in which varying

amounts of solvents, or energy input, are encountered at the different process steps. Geometric preferences of lipids can be better understood by considering the relative volume occupied by each portion of the molecule. Cullis et al. (16) and Gruner et al. (17) have described such molecular shape analysis of lipids in detail (Fig. 3).

The various macromolecular aggregates of lipids also make it possible to associate with drugs through physical entrapment. In these cases, the concentration of lipid associated drug is dictated by the size or conformation of the drug and the spaces created in the lipid aggregate. Entrapment levels of drug can vary from less than 1% to as high as complete encapsulation. Excluding chemical interactions or solubilization phenomena, drug encapsulation efficiency is linked to both the lipid conformation and the manufacturing method, which juxtaposes the two components. This is particularly true of liposome formulations in which factors of temperature, pressure, solute concentration, solvent volumes, or energy application all contribute to the final amount of drug incorporated within the lipid matrix (18).

MANUFACTURING OF LIPID-CONTAINING PRODUCTS

Manufacturing issues should be evaluated as part of overall lipid formulation development. Although lipid addition may be desirable for pharmacological purposes there are often commercial issues which present hurdles to eventual marketing. Among those factors are the availability and cost of raw materials, handling of intermediate mixed-phase systems, particulate control, stability of the lipid ingredients, sterilization, packaging interactions, qualification of noncompedial ingredients, and development of validated methods for analysis of either raw material or product.

Availability and Cost of Lipid Raw Materials

For commercial formulations, raw material sources should be of high purity and quality consistently meeting defined specifications, preferably compendial when available. Fortunately, high-purity lipids have

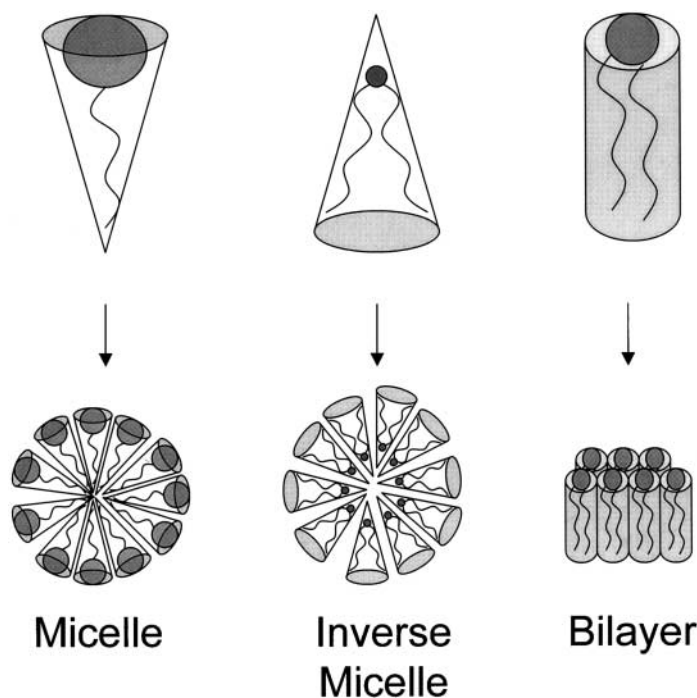


Fig. 3 Geometric models representing packing behavior of micelles, inverse micelles, and lamellar structures. Lipid molecules with domains that occupy a large lateral distance compared to adjacent domains of the molecule will conform to corresponding shapes and packing behavior. Micelles and inverse micelles approximate cone shape, whereas lamellar systems resemble cylinders. (Adapted from Refs. 16 and 17.)

become increasingly available in large supply, and in particular, phospholipid supply has advanced in recent years due to the approval and marketing of parenteral liposome formulations.

Large supply of naturally derived lipids can be obtained from plants in which many oils and fatty acids can be readily extracted and purified. Animal sources (e.g., eggs or milkfats) are used to derive complex lipids such as phospholipids and cholesterol. Yield from natural sources is dependent on the weight-percent composition and the efficiency of the extraction procedure. The constitution of fatty acids in vegetable oils varies widely from different sources (19). For example, oleic acid is present at 64.6% by weight in olive oil but is present at only 0.7% in palm kernel oil. Similarly, castor oil triglyceride is comprised of almost entirely ricinoleic chains. There are numerous raw material suppliers of oils and oil fractions worldwide. As such, the relative cost of bulk purified fractions and their derivatives such as salt forms and alcohols is quite low (\approx \$50–150/kg). Costs for high purity synthetic phospholipids ($>97\%$) continue to drop when compared to prices 8 to 10 years earlier (18, 20, 21). This is primarily due to the advent of marketed liposome products. Phospholipids are available worldwide from companies such as Genzyme (U.S./U.K.), Avanti Polar Lipids (U.S.), Matreya Inc. (U.S.), American Lecithin Co. (U.S.), Lipoid GmbH (Germany)/Vernon Walden (U.S.), Lucas Meyer (Germany), Nichiyu Liposome/Nippon Oil and Fat (Japan) and Northern Lipids (Canada). Also now available in larger supply are PEGylated phospholipids, lysopholipids, and various cationic lipids used in gene therapeutic applications.

Compounding of Lipids into Formulations

Lipids may be supplied either as lyophilized or dried powders, as liquids, or dissolved or suspended in an appropriate solvent. Dried powders are typically waxy or sticky in character, making aliquoting of the product troublesome. Lipid lyophilates can also be hygroscopic, thus presenting some difficulty for fluidized bed operations. Fluid bed applications however, may be successful if the concentration of lipid is held fairly low (22).

Handling of lipid products in a liquid state allows for more accurate compounding. Solubilization vehicles may involve dissolution into either polar or nonpolar solvents or surfactant solutions. Aside from safe handling, a primary concern when introducing lipids

from solvent environments is the compatibility of the solvent with processing equipment, in particular the gaskets and seals found in the pumps, transfer lines, mixers, and extractors. Elevated temperatures, which are used to dissolve or maintain the phase transition of the lipid or to initiate solvent evaporation, can further induce problems for gaskets and seals. In systems with nonvolatile solvents and detergents, removal of the dissolution vehicle can be achieved with flow centrifugal separation, chromatography or diafiltration.

Introducing a lipid-containing solvent to an aqueous solution of drug will require some means for energy input to mix the components, whether by simple rotating blade stirring, more vigorous homogenization, or milling, by sonication and filtration, or by high-pressure extrusion (23). Unless cosolvation is achieved a suspension of nonmiscible aqueous and nonpolar phases will offer more resistance to flow and may present problems for extrusion through filters, which tender a surface that is of opposite character to one of the phases. Binding of active agent, such as proteins, must also be considered (24). The conformational state of the lipid and the concentration also will govern the workability of the formulation. Lamellar orientations will flow fairly readily at concentrations below about 400 mg/ml whereas cubic or hexagonal phases will present difficulty even at very low concentrations.

Particulate and Particle Size Control

For lipids dissolved in solution, particulate control may be necessary, depending on administration route. For parenterals the large- or small-volume parenteral particulate tests of the USP are applicable, as would be the Ph. Eur. clarity test and JP foreign matter insoluble test. For suspended or opaque lipid formulations, compendial particle size tests and specifications for suspensions will apply (i.e., light obscuration or microscopy). Size reduction of lipid formulations can be achieved with a variety of processing equipment including paddle stirrers, homogenization and high shear mixers, Microfluidizer, colloid mills, ultrasonic and piezoelectric emulsifiers, and pressurized filtration or membrane extrusion.

Lipid Stability

A characteristic of lipid products, particularly those with unsaturated lipids is peroxide formation with oxidation (25). Free radicals such as ROO^\bullet , RO^\bullet , and OH^\bullet can

damage the drug and induce toxicity. Lipid peroxides may also form due to autoxidation, which increases with unsaturation level. Hydrolysis of the lipid may be accelerated due to the pH of the solution (26) or from processing energy such as ultrasonic radiation (27). Antioxidants (i.e., α -tocopherol, propyl gallate, ascorbate, or BHT) may be required.

Sterilization Methods

Because of possible oxidation, steam or heat sterilization of lipids may not be an attractive option. Validations at nonstandard autoclave cycles or F_0 values may prove valuable. In some instances, as in ophthalmic ointments made with white petrolatum or oleagenous components, the base can be sterilized by dry heat (160–180°C for 1–3 h) and combined aseptically with the sterile drug and additives (28). Ointments with lipids such as lanolin, petrolatum, and mineral oil have been terminally sterilized by cobalt-60 gamma irradiation with success (28–30). In less viscous solutions, end terminal filtration with aseptic fill is often necessary, requiring justification to regulatory authorities for lower sterility assurance levels. High pressures may be required to force lipid through membranes, thus adequate filter integrity checks are critical. High-pressure extrusion through straight channel membranes (polycarbonates/ceramics) may reduce the pressures and improve flow rates (23).

Packaging and Processing Surface Considerations

Lipids will be attracted to hydrophobic surfaces where losses might be expected. Overall binding will be affected if the molecule also contains ionic regions. Generally, lipids are chemically compatible with most plastics or glass used in packaging. Residual solvents that partition into the lipid phase may interact with plastic resins such as polystyrene or polycarbonate or with rubberized seals. Extractables from plastic packaging may be accelerated by the presence of lipid. Extraction of silicone polymer into lipid phases also occurs from silanized surfaces or coated gaskets. Compendial tests and regulations on extractables and impurities will apply here. Presence of metals coming from process or package surfaces are a potential source of catalysis leading to degradation of lipid components. In addition, reducing oxygen headspace should be considered to minimize potential degradation. Inert gas purging can improve lipid stabilities.

ANALYSIS OF LIPID PHARMACEUTICAL PRODUCTS

Because lipids exhibit primary through quaternary conformations, the analysis of lipid-containing pharmaceutical products must therefore involve both chemical and physical determinations to define the product on the molecular level as well as the aggregate state of the product. Stability-indicating assays can be developed from both approaches.

Chemical Determinations

Lipids can be measured by standard analytical methods such as wet chemistry, HPLC, thin-layer chromatography, and gas chromatography (31). Because lipid formulations may have several components and because the formulation may have an overall aggregate structure, it is usually necessary to develop further strategies for isolation and detection of the lipid ingredients (32). Dissociation of the individual components using temperature, solvents, detergents, reducing or oxidizing agents, or mechanical disruption prior to analysis may be necessary. Further separation or extraction of the components before analysis is required if interference is encountered. With lipid solvent extraction, there is a greater likelihood for losses of material. More often than not, a combination of techniques will be required for validation of the product. There are worldwide compendial procedures for analysis of various lipids by wet chemistry procedures to determine concentration, end group analysis, acid value, hydroxyl value, iodine value, and saponification value. A listing of the USP/NF methods for lipids is shown in Table 2.

HPLC Methods

Analysis of nonpolar lipids by HPLC is best carried out using normal phase columns. However, for mixed phases with polar drug or drug within an aqueous phase some compromise may be necessary. Good separation of polar and neutral lipids with a C8 column and a four-solvent mobile phase has been reported (33). Elution of neutral lipids like triglycerides from C18 columns is slow, however, good resolution can be achieved. Mobile-phase development is usually necessary to effect a high degree of separation and resolution of similarly eluting components (34). Heated columns are beneficial for increasing temperature above the phase-transition temperature of the lipid and thus minimize clogging. Lipids with predominantly saturated chain lengths will

Table 2 USP 24/NF 19 methods for lipids

Lipid category	Method
Steroids and steroid esters	Cholesterol, betamethasone, calcifediol, cholcalciferol, clocortolone pivate, cortisone acetate, dehydrocholic acid, desoximetasone, desoxycortisone (acetate and pivalate), dexamethasone, difluorosone diacetate, dihydrotachysterol, dydrogesterone, ergocalciferol, estradiol, estriol, conjugated estrogens, estrone, estropipate, ethynodiol diacetate, fludrocortisone acetate, flumethasone pivalate, flunisolide, fluocinolone acetate, fluocinonide, fluoromethalone, fluoxymesterone, flurandrenolide, halcinonide, hydrocortisone, levonorgestrel, medroxyprogesterone, meprednisone, mestranol, methylprednisone, methyltestosterone, mibolerone, mometasone furoate, nandrolone (decanoate and phenpropionate), norethynodrel, norgestrel, oxymethalone, prednisone, progesterone, rimexolone, spironolactone, stanozolol, testolactone, testosterone, trenbolone acetate, and triamcinolone
Fatty-acids, salts and esters	Aluminum monostearate, calcium stearate, ethyl oleate, isopropyl myristate, isopropyl palmitate, magnesium stearate, oleic acid, polyoxyl 40 stearate, propionic acid, sodium stearate, stearic acid, purified stearic acid, and zinc stearate
Fatty alcohols	Benzyl alcohol, butyl alcohol, cetostearyl alcohol, cetyl alcohol, cetyl esters wax, lanolin alcohols, octyldodecanol, oleyl alcohol, and stearyl alcohol
Oils and oil esters	Almond oil, castor oil, cod liver oil, corn oil, cottonseed oil, diacetylated monoglycerides, ethiodized oil injection, glyceryl behenate, glyceryl monostearate, hydrogenated castor oil, hydrogenated vegetable oil, light mineral oil, mineral oil, mono- and diglycerides, mono- and diacetylated monoglycerides, oil-soluble vitamins, olive oil, orange flower oil, peanut oil, peppermint oil, perflubron, persic oil, polyoxyl 35 castor oil, polyoxyl 40 hydrogenated castor oil, rose oil, safflower oil, sesame oil, soybean oil, squalane, tocopherols excipient, vitamin E, and vitamin E PEG succinate
Phospholipids	Lecithin
Waxes	Caranuba wax, emulsifying wax, hard fat, hydrophilic ointment, hydrophilic petrolatum, microcrystalline wax, paraffin, petrolatum, rose water ointment, synthetic paraffin, white wax, yellow ointment, and yellow wax

have poor ultraviolet (UV) absorption even at the lower wavelengths (190 nm). Alternative detectors should then be considered.

Thin Layer Chromatography

One- and two-dimensional thin-layer silica gel chromatography remains a cornerstone of lipid analysis (31, 35). Sensitivity by this method is typically as low as 2 µg/spot. Lipid visualization can be achieved by many methods, including iodine vapor or charring of plates following exposure to sulfur-dichromic acid, cupric reagents, Phospray, or α-naphthol. Other useful reagents include stains and dyes such as fluorescamine, rhodamine 6G, bromothymol blue, molybdenum blue, phosphomolybdic acid, and silver nitrate, which may be impregnated into the silica. Validation of lipid

purity and quantitation is performed using gel or plate scanners.

Gas Chromatography (GC)

Capillary and packed (GC) columns are of value in the analysis of complex mixtures of lipids. The best capillary column length will depend on the complexity of the material injected, however, 30-m columns are often employed. In packed columns, many types of stationary phases are available for lipid separation, and these include silicone and alkylated or cyanogenated derivatives, polyesters, polyglycol, and carboranes. It is also common to derivatize the fatty-acid side chains to the corresponding methyl esters by reaction in BF₃/methanol prior to chromatographic analysis to achieve more distinct and uniform separations.

Physical Determinations

The aggregate states of lipids are discernable from measurements provided by a host of analytical devices, which yield information on physical properties. These techniques may also be used to give clues as to the interaction of the lipid carrier and the drug.

Particle sizing

Lipid particulate analysis can be achieved with most commercial laser particle counters. Compendial requirements for suspensions also allow for sizing via electron microscopy determinations, which can provide for qualitative assessments in addition to quantitation. Lipid suspension particles can be detected by negative-stain, freeze-fracture, critical-point drying and scanning techniques.

Nuclear magnetic resonance (NMR)

NMR is a valuable technique in the analysis of lipid phases (36). More specifically, proton, deuterium, carbon-13, fluorine-19, and phosphorus-31 NMR have been utilized for analysis of the dynamic and motional properties of lipids, lipid diffusion, ordering properties, head-group hydration, lipid asymmetry, quantitation of lipid composition, and head-group conformation and dynamics. Cullis et al. (16) and Gruner et al. (17) have shown the importance of P-31 NMR as a tool in the determination of phase properties and lipid asymmetry and the identification of bilayer, hexagonal, and isotropic phases.

Electron spin resonance (ESR)

ESR is used to give information on the local environment of a lipid molecule (37). In normal state these molecules inherently do not exhibit ESR spectrums. However the necessary signal can be generated from reporter labels such as nitroxide or doxyl probes, which can either be linked directly to the lipid molecule or prelinked to a lipid chain and then partitioned into the lipid aggregate formulation. Data from ESR spectrums are valuable for determining molecular properties of the formulation, such as phase transitions and separations, order parameters (anisotropy), polarity, lateral diffusion, segregation and clustering, surface and transbilayer potentials, permeability and internal volumes, surface determinants (antigens), pH gradients, lipid asymmetry, flip flopping, and fusion.

Differential scanning calorimetry

Differential scanning calorimetry is a well known technique in the study of the thermal behavior of lipids

and can be used to assess purity and stability of lipids, perturbation of aggregate structures, phase transition temperatures, lipid mixing behavior, and influence of other molecules and ions on structure (38).

X-ray and neutron diffraction

Diffraction patterns of lipid solutions can yield strong evidence for the presence of specific repeating conformational structures as well as the spacing between lipid molecules in organized films or layers (12, 39–41).

Spectroscopic analysis

Appraisals of the optical properties of lipid solutions and dispersions will provide information on concentrations, aggregation and stability, phase transitions, densities, and repeating structures (42–44). Measurements of refractive index, scattered light intensity (polarized and depolarized), and birefringence are relatively easy laboratory methods on which certain product specifications may be based. Also, fluorescent techniques can readily provide information on lipid movements and transfer of lipid between particles (45, 46).

COMMERCIAL AND EXPERIMENTAL LIPID DOSAGE FORMS

The key purposes of lipid materials in dose forms include: 1) improving the solubility or physical workability of the drug for ease of administration or to enhance stability; 2) augmentation or reduction in absorption of the drug from the formulation; 3) specific drug targeting to maximize response and minimize side effects; and 4) controlled or slow delivery of the drug from the formulation. Currently, the largest use of lipids in pharmaceuticals is in products for oral or topical dosage administration, although there is growing use in parenteral, pulmonary and nasal products.

The mechanics of drug delivery from lipid systems is governed by five structural features: primary structure, that is, chemical or molecular interactions, secondary organization into aggregate structures such as inverse micelles, tertiary organization of the aggregate projected into three dimensions such as three dimensional inverse micelles or hexagonal phase tubules; quaternary associations, agglomeration or interaction of the three dimensional structures such as lateral stacking of hexagonal tubules, and final packing of the molecules in solution (concentration effects) to produce liquids, creams, solids, etc. Mammalian systems have natural degradative pathways for the rapid metabolism of most

lipid raw materials. However, the release of drug to a biological system will be controlled at least as much by these structural features as by simple natural degradative mechanisms.

Lipophilic Derivatives and Prodrugs

Lipophilic derivatives and prodrugs are a prudent tactic to alter the normal interaction of drug compounds with cells and cellular barriers. The type of modification desired may be a permanent one if the drug compound maintains its activity following conjugation, or a reversible type subject to biological cleavage. A variety of strategies for chemical coupling of lipids to drugs can be developed to produce the desired modified product (20). Briefly, nonreversible reaction approaches include conjugation to drugs with amine residues either via glutaraldehyde, lipid anhydrides or halides, or succinimidyl derivatives of the fatty acyl chain via carbodiimide activation. Nonreversible conjugations have also been accomplished via carbodiimide-activated carboxyl moieties of drugs reacted with amine containing lipids. Permanent conjugation to phenolic residues of drugs can be accomplished with diazo derivative of lipids bearing available amines. This approach has been successfully employed to prepare lipophilic derivatives of peptides containing tyrosine residues. Lipid conjugation reactions have also taken advantage of available sulfhydryl groups on drugs, particularly for modification of antibodies. On the other hand, bioreversible conjugates have been made via use of the Schiff base reaction to couple an aldehyde-bearing lipid to amines present on the drug. It is possible to convert the

hydroxyl groups of more abundant glycolipids to the corresponding aldehydes using periodate. An extension beyond the Schiff base reaction is the Mannich base condensation, which introduces a nucleophilic reactant such as an enolate anion or amide to the reaction between the aldehyde and amine. In these reactions the amine component is generally a lipophilic primary or secondary amine. Extensive literature is available on the production and biological activities of lipophilic *N*-Mannich base derivatives and stabilized α -acycloalkyl forms. The effect of lipophilic derivatized drugs given by various administration routes has been discussed in detail in several reviews (47–50).

Permeation Enhancers

There are a number of excellent reviews on the requirements for enhancing drug permeability across lipophilic biological barriers, with particular reference to the importance of the lipophilic properties of formulations (51–55). A listing of some penetration enhancing lipids can be found in Table 3. Administration sites include buccal, oral, nasal, ocular, transdermal, rectal, and pulmonary.

Vehicles for Dispersion

There is fairly extensive use of lipid materials in oral liquid dose forms strictly as vehicles. Primary applications have been as surfactants to promote drug suspension or dissolution (fatty glycols and fatty acids), as flavoring agents (natural or synthetic oils), and as thickening agents (hydrogenated oils). Aspects of

Table 3 Lipid permeation enhancers

Category	Compounds
Ionic lipids	Lauryl sulfate, dodecyl-2-pyrrolidone, <i>N</i> -dodecyl azacycloheptan-2-one (Azone), <i>N</i> -dodecyl- <i>N,N</i> -dimethyl betaine, calcium dodecylbenzene sulfonate, dioctyl sodium sulfosuccinate, dodecyl <i>N,N</i> -dimethylamino (acetate or propionate), and cetyltrimethylammonium bromide
Steroid and steroid esters	Cholate, deoxycholate, taurocholate, glycocholate, taurodeoxycholate, sodium taurodihydrofusidate, and cholesterol esters
Fatty-acids/fatty-acid esters/fatty alcohols	Oleic acid, lauric acid, capric acid, heptanoic acid, stearic acid, palmitoleic acid, palmitelaidic acid, octadecanoic acid, sucrose laurate, and isopropyl myristate
Phospholipids	Phosphatidylcholines, lysophosphatidylcholine, monooleoyl phosphatidylethanolamine
Oils	Monoolein, cocoa butter, cardamom oil, tricaprylin, mineral oil, terpenes, and terpenoids

pharmaceutical oral suspensions have been discussed in greater depth (56). Solution based-lipid vehicles have been applied in softgel applications, the primary category being oils that are compatible with gelatins and that have had applications as both lubricants for processing of the gelatin sheets and as drug vehicles for liquid fill operations into the softgel. High lipid concentrations may allow for higher ethanol content in softgel fills (57).

Lipids may also be used to create solutions for injectable products, particularly intravenous preparations. Fatty acids, fatty glycols, and fatty alcohols may be used to enhance the dissolution of certain insoluble drugs, act as preservatives, or function as active agents as demonstrated by Scleromate injection (Glenwood, Tena-fly, NJ), a mixture of fatty acid salts derived from cod liver oil. As another example, benzyl alcohol preservative has been useful in formulations for water insoluble drugs such as etoposide (VePesid[®], Bristol-Meyers Squibb, Princeton, NJ) and is a primary active agent in Zilactin gel (Zila, Phoenix AZ). Polyoxyethylated fatty-acid derivative has been used in the dissolution of phytonadione, a lipid-soluble vitamin (AquaMEPHYTON, Merck, Sharp & Dohme, Rahway, NJ) for subcutaneous or intramuscular injection. Injectable amphotericin B is solubilized with sodium desoxycholate in Fungizone (Bristol-Meyers-Squibb, Princeton, NJ). Polyoxyethylated or PEGylated castor oils are also used in dissolution of injectable drugs such as cyclosporine (Sandimmune, Novartis/Sandoz, East Hanover, NJ), paclitaxel, and teniposide (Taxol, and Vumon, Bristol-Myers Squibb) and miconazole (Monistat, Janssen Pharmaceutica, Piscataway, NJ). Clear colloidal dispersions are also possible using a wide variety of phospholipids, cholesterol esters, and tocopherol esters (13, 18).

Solid- and Liquid-Crystalline Suspensions, Creams, and Gels

Suspensions and creams are most often developed for aqueous insoluble drugs, as typified by ophthalmic and otic preparations of corticosteroids. In suspensions such as Cortisporin or Pediotic (Glaxo Wellcome, Research Triangle, NC), components such as cetyl alcohol, glyceryl monostearate, mineral oil, and propylene glycol are commonly used to effect a homogeneous suspension of drug particles.

The creation of fine suspensions may also be necessary to administer a highly insoluble product by parenteral injection. In these instances lipids may be used as either wetting agents or suspension vehicles (58). Lecithin (phosphatidylcholine) is a suitable agent for wetting or

suspending of drug particles in either aqueous or nonaqueous solutions. This common formulation additive is used for injectable long-acting IM suspensions of penicillin (e.g., Bicillin L-A[®], Wyeth-Ayerst, Philadelphia, PA) for the bronchodilator inhalation aerosol, Atrovent[®] (Boehringer Ingelheim, Ridgefield, CT), and for the otic suspension Cipro HC (Bayer, Germany). Naturally derived and synthetic lecithins, in mixtures with neutral lipids, also serve as the active ingredient in lung surfactants products such as Exosurf (Glaxo Wellcome), Survanta (Ross, Columbus OH), and Infasurf (Forest Labs, St. Louis, MO).

Lipid materials are used extensively as vehicles in topical creams, ointments, gels, and lotions, and usually serve as the base material for many such preparations. These formulations are principally presented as emulsions and may contain fatty acids, fatty-acid salts, fatty alcohols, petroleum based and natural oils, waxes, fatty glycols, lanolin, and other hydrophobic surfactants. Such emulsions are common in dermal topicals, for which the list of products is too numerous to elaborate. Some of these ingredients also are present as vehicles in suppositories such as semisynthetic glycerides existing in Nembutal (Abbott Laboratories, North Chicago, IL), hydrogenated vegetable oil as found in Dulcolax (Novartis Consumer Health), and glycerides of fatty acids or oils that can be found in vaginal suppositories such as Prostin E2 (Pharmacia and Upjohn, Kalamazoo, MI) and Crinone Gel (Wyeth-Ayerst). Mineral oil is frequently present in transdermal products such as Catapres-TTS (Boehringer Ingelheim) and Estraderm (Novartis Pharmaceuticals Corp). In addition, numerous ocular or topical ointments, such as Tobradex (Alcon Labs, Fort Worth, TX) and Nitro-Bid (Hoechst Marion Roussel, Kansas City, MO), use white petrolatum and mineral oil as a base. Short-chain triglycerides such as triacetin can be found in products such as Prepidil Gel (Pharmacia-Upjohn, Kalamazoo MI), a cervically administered prostaglandin.

Oil-in-water emulsions have also been applied for intravenous use. Commercial parenteral emulsions include Dizac (Schein, Florham Park, NJ) and Diprivan (Zeneca Pharmaceuticals, Wilmington, DE). Many commercial or experimental parenteral products have been based on vegetable oil (most often soybean, safflower, or cottonseed) stabilized with phosphatides and monoglycerides, which nicely match the hydrophile-lipophile balance (HLB) requirements of those oils ($\approx 6-7$). Further prospects for expanded nontoxic parenteral emulsions may come with use of other phosphatide-based surfactants with high HLB values (24, 59, 60). A number of excellent reviews and articles on the application of parenteral

emulsions for drug delivery have been written (61–65). Drug delivery from oil/phosphatide emulsions stems from earlier development and marketing of intravenous nutrient emulsion products such as Intralipid (KabiVitrum, Clayton, NC) and Aminosyn II, (Abbott Laboratories), which are sterile and nontoxic. The development of these emulsions as well as distribution profiles of these and similar emulsion compositions have been studied (63, 64, 66–68). A cursory scan of the literature on parenteral oil emulsion formulations will obtain studies on amphotericin B, prostaglandin E1, halothane, pregnanolone, paclitaxel, perilla ketone, penclomedine, F-octylbromide, flurbiprofen axetil (Lipfen), lasalocid, lignin, podophyllotoxin, tacrolimus, doxorubicin, epirubicin, menatetrenone, chlorpheniramine maleate, naproxin, cyclosporin A, propranolol, testosterone, benzocaine, phenylazoaniline, palmitoylrhizoxin, pilocarpine, diazepam, and various peptide and proteins for vaccine delivery.

Liposomal Dosage Forms

Extensive studies on liposomes date back to the 1960s. Many good comprehensive review references exist on compositions and manufacturing of liposomes (18, 23, 24, 69–73) including the article “Liposomes as Pharmaceutical Dosage Forms” in this volume. The earliest commercial liposomal formulations were developed for veterinary application (Novasome, IGI, Vineland, NJ) or over-the-counter cosmetic creams promoted for improved hydration (L’Oreal, Paris and Dior, Paris). More recently, parenteral liposome formulations of amphotericin B, doxorubicin, and daunorubicin have been approved and marketed (ABELCET, Elan, the Liposome Co., Inc, Princeton, NJ; AmBisome and DaunoXome, Nexstar/Fujisawa, Deerfield Park, IL; Amphotec and Doxil, Sequus/Alza, Menlo Park, CA), with others on the horizon for applications in photodynamic therapy (74). Although the vast majority of liposome preparations are constructed from phospholipids, other nonphospholipid materials can be used either alone or in mixtures to form bilayer arrays. One such example is Amphotec, which utilizes sodium cholesteryl sulfate as the primary lipid. Other liposome forming materials may include but are not limited to fatty-acid compositions, ionized fatty acids, or fatty acyl amino acids, longchain fatty alcohols plus surfactants, ionized lysophospholipids or combinations, nonionic or ionic surfactants and amphiphiles, alkyl maltosides, α -tocopherol esters, cholesterol esters, polyoxyethylene alkyl ethers, sorbitan alkyl esters, and polymerized phospholipid compositions (20).

Low Density Lipoprotein Carriers

Lipoproteins are naturally occurring particulate emulsion carriers for the transport of cholesterol and other lipids such as triglycerides in the blood. Because low-density lipoprotein (LDL) particle clearance is receptor mediated, they have been proposed as drug carriers for targeting applications, specifically for targeting of cytotoxic agents to tumor cells, delivery of antiviral agents to parenchymal liver cells, targeting of immunomodulators, antiviral and antiparasitic drugs to Kupffer and endothelial cells, and as gene vectors (75–78).

Solid Dosage Forms

The main use of lipids in solid form has been for oral tableting applications. Fatty-acid salts such as magnesium and calcium stearates, and various waxes and glycerides are most often used as conditioners and binders during compaction and provide more even, controlled, or slower disintegration of the tablet once administered. Extensive literature is available in which these materials are discussed in context of the preparation of oral tablets or capsules (79–81). The use of lipids for solid-implant formulations has also been investigated. Materials such as cholesterol, and high-melting-point fatty-acids, fatty anhydrides, and glycerides have been utilized in compressed implants to prolong systemic delivery of drugs (82, 83). Lipids are also well suited for suppository and vaginal insert formulations.

Safety of Lipid Products

The safety of a lipid product will in part be a reflection of 1) the purity of the compounds administered; 2) biological toxicity of the basic chemical ingredients; and 3) reactions to structural presentations of the lipid. Noncompendial lipid materials will necessitate significant toxicology testing.

Contaminants and impurities

Sensitive analytical procedures will be required to distinguish contaminants from lipid peaks within the preparation. By-products may be present from the synthetic processes used to produce the lipids or if copurified from the natural source. For example, common impurities in synthetic diacyl chain lipids are the monoacyl forms, which are generally more toxic to biological systems. Endotoxins and pyrogens either may be detectable in a lipid preparation (23, 84) or difficult to detect due to the lipids shielding against the analytical reagents (85).

Immune reactivity

Consideration should be given to immune reactivities when administering lipids to mammalian systems. Oils are well known for their adjuvancy; different oils will produce varying levels of reactivity (86). Biological responses such as leukocyte attraction, encysting of the oil, and edema reactions vary in severity simply as a function of the chemical nature or purity of the oil itself. Importantly, most lipids, including normal endogenous compounds possess antigenic potential to varying degrees (87). These possible reactivities should be monitored as part of the overall clinical design.

Conformational considerations

There may be biological sensitivities to the conformational presentation and sizing of the lipid formulation. For example the ability of antibodies to distinguish between lamellar organizing lipids and hexagonal-phase lipids are known and may form the basis for certain types of autoimmune dysfunction (88). Biological factors such as reticuloendothelial cell recognition of particles provide further impetus for control of the particle size and stability.

REFERENCES

- Harkins, W.D.; Mattoon, R.W.; Corrin, M.L. Structure of Soap Micelles Indicated by X-rays and the Theory of Molecular Orientation. I. Aqueous Solutions. *J. Am. Chem. Soc.* **1946**, 68, 220–228.
- Jones, M.N., Chapman, D., Jones, M., III Eds.; *Micelles, Monolayers and Biomembranes*; John Wiley & Sons: New York, 1994; 264.
- Mittal, K.L., Shah, D.O. Eds., *Surfactants in Solution*; Plenum Publishing Corp.: New York, 1992; 720.
- Florence, A.T. Drug Solubilization in Surfactant Systems. *Techniques of Solubilization of Drugs*; Yalkowsky, S.H., Ed.; Marcel Dekker, Inc.: New York, 1981; 15–89.
- Mittal, K.L. *Micellization, Solubilization and Microemulsions*; Plenum Publishing Corp.: New York, 1977; vol. 1 and 2, 945.
- Gelbart, W.M., Ben-Shaul, A., Roux, D., Eds. *Micelles, Membranes, Microemulsions, and Monolayers (Partially Ordered Systems)*; Springer Verlag: New York, 1994; 608.
- Shah, D.O., Ed. *Micelles, Microemulsions and Monolayers: Science and Technology*; Marcel Dekker, Inc.: New York, 1998; 610.
- Sjoblom, J., Ed. *Emulsions and Emulsion Stability*; Marcel Dekker, Inc.: New York, 1996; 474.
- Shinoda, K.; Friberg, S. *Emulsions and Solubilization*; John Wiley & Sons: New York, 1986; 174.
- Tyle, P. Liquid Crystals and Their Applications in Drug Delivery. In *Controlled Release of Drugs: Polymers and Aggregate Systems*; Rosoff, M., Ed.; VCH Publishers: New York, 1989; 125–162.
- Lo, I.; Florence, A.T.; Treguier, J.P.; Seiller, M.; Puisieux, F. The Influence of Surfactant HLB and the Nature of the Oil Phase on the Phase Diagrams of Nonionic Surfactant-Oil-Water Systems. *J. Colloid Interface Sci.* **1977**, 59, 319–327.
- Gruner, S.M. Material Properties of Liposomal Bilayers. *Liposomes: From Biophysics to Therapeutics*; Ostro, M.J., Ed.; Marcel Dekker, Inc.: New York, 1987; 1–38.
- Weiner, A.L. Lamellar Systems for Drug Solubilization. *Liposomes: From Biophysics to Therapeutics*; Ostro, M.J., Ed.; Marcel Dekker, Inc.: New York, 1987; 339–369.
- Brown, G.H.; Wolken, J.J. *Liquid Crystals and Biological Structures*; Brown, G.H., Wolken, J.J., Eds.; Ch. 2 and 3 Academic Press: New York, 1979.
- Caffrey, M. *Lipidat a Database of Thermodynamic Data and Associated Information on Lipid Mesomorphic and Polymorphic Transitions*; CRC Press Inc.: Boca Raton, FL, 1993; 305.
- Cullis, P.R.; de Kruijff, B.; Hope, M.J.; Verkleij, A.J.; Nayar, R.; Farren, S.B.; Tilcock, C.; Madden, T.D.; Bally, M.B. Structural Properties of Lipids and their Functional Roles in Biological Membranes. *Membrane Fluidity in Biology. Concepts of Membrane Structure*; Academic Press: New York, 1983; 1, 39–81.
- Gruner, S.M.; Cullis, P.R.; Hope, M.J.; Tilcock, C.P.S. Lipid Polymorphism: The Molecular basis of Nonbilayer Phases. *Ann. Rev. Biophys. Biophys. Chem.* **1985**, 14, 211–238.
- Weiner, A.L.; Cannon, J.B.; Tyle, P. Commercial Approaches to the Delivery of Macromolecular Drugs with Liposomes. *Controlled Release of Drugs: Polymers and Aggregate Systems*; Rosoff, M., Ed.; VCH Publishers: New York, 1989; 217–253.
- Bailey, A.E. Composition and Characteristics of the Individual Fats and Oils. *Industrial Oils and Fat Products*; Interscience Publishers, Inc.: New York, 1951; 967.
- Weiner, A.L. Lipids in Pharmaceutical Dosage Forms. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, Boylan, Eds.; Marcel Dekker, Inc.: New York, 1993; Vol. 8, 417–476.
- Weiner, A.L. Developing Lipid-Based Vehicles for Peptide and Protein Drugs. Part I: Selection and Analysis Issues. *Biopharm.* **1990**, 3 (3), 27–32.
- Chen, C-M.; Alli, D. Use of Fluidized Bed in Liposome Manufacturing. *J. Pharm. Sci.* **1987**, 76, 419.
- Martin, F.J. Pharmaceutical Manufacturing of Liposomes. *Specialized Drug Delivery Systems. Manufacturing and Production Technology*; Tyle, P., Ed.; Marcel Dekker, Inc.: New York, 1990; 267–316.
- Weiner, A.L. Liposomes for Protein Delivery. *Immuno-methods* **1994**, 4, 201–209.
- Bressler, R. Fatty-Acid Oxidation. *Lipid Metabolism*; Florkin, M., Stotz, E.H., Eds.; Comprehensive Biochemistry Ch. 8 Elsevier Publishing Co. Amsterdam, 1970; 18.
- Ho, R.J.Y.; Schmetz, M.; Deamer, D.W. Non-enzymatic Hydrolysis of Phosphatidylcholine Prepared as Liposomes and Mixed Micelles. *Lipids* **1987**, 22, 156–158.
- Jana, A.K.; Agarwal, S.; Chatterjee, S.N. Ultrasonic Radiation Induced Lipid Peroxidation in Liposomal Membrane. *Radiat. Environ. Biophys.* **1986**, 25, 309–314.

28. Turco, S.J.; King, R.E. *Sterile Dosage Forms. Their Preparation and Clinical Application*; Lea & Febiger: Philadelphia, 1974; 326–328.
29. Tsuji, K.; Goetz, J.F.; Vanmeter, W. Effect of 60 Co-Irradiation on Penicillin G Procaine in Veterinary Mastitis Products. *J. Pharm. Sci.* **1979**, *68*, 1075–1080.
30. Nash, R.A. Radiosterilized Tetracycline Ophthalmic Ointment. *Bull. Par. Drug Assoc.* **1974**, *28*, 181–187.
31. Kates, M. *Techniques in Lipidology. Isolation, Analysis and Identification of Lipids*, 2nd Revised Ed.; Elsevier: Amsterdam, 1986; 464.
32. Kuksis, A.; Myher, J.J. General Strategies For Practical Chromatographic Analysis of Lipids. *Journal of Chromatography Library Monographs*; Kuksis, A., Ed.; (Chromatography of Lipids in Biomedical Research and Clinical Diagnosis) Elsevier: Amsterdam, 1987; Vol. 37, 1–47.
33. McCracken, M.S.; Holt, N.J. High-Performance Liquid Chromatographic Determination of Lipids in Vesicles. *J. Chromatog.* **1985**, *348*, 221–227.
34. Jungalwala, F.B. High-Performance Liquid Chromatography of Phosphatidylcholine and Sphingomyelin with Detection in the Region of 200 nm. *Biochem. J.* **1976**, *155*, 55–60.
35. Fried, B.; Sherma, J. *Thin Layer Chromatography*; Marcel Dekker, Inc.: New York, 1986; 394.
36. Browning, J.L. NMR Studies of the Structural and Motional Properties of Phospholipids in Membranes. *Liposomes: From Physical Structure to Therapeutic Applications*; Knight, C.G., Ed.; Elsevier/North-Holland Biomedical Press: Amsterdam, 1981; 189–242.
37. Marsh, D.; Watts, A. ESR Spin Label Studies of Liposomes. *Liposomes: From Physical Structure to Therapeutic Applications*; Knight, C.G., Ed.; Elsevier/North-Holland Biomedical Press: Amsterdam, 1981; 139–188.
38. Mabrey-Gaud, S. Differential Scanning Calorimetry of Liposomes. *Liposomes: From Physical Structure to Therapeutic Applications*; Knight, C.G., Ed.; Elsevier/North-Holland Biomedical Press: Amsterdam, 1981; 106–138.
39. Luzzati, V.; Tardieu, A. Lipid Phases-Structure and Structural Transitions. *Ann. Rev. Phys. Chem.* **1974**, *25*, 79–94.
40. Mitsui, T. X-Ray Diffraction Studies of Membranes. *Adv. Biophys.* **1978**, *10*, 97–135.
41. Franks, N.P.; Lieb, W.R. X-ray and Neutron Diffraction Studies of Lipid Bilayers. *Liposomes: From Physical Structure to Therapeutic Applications*; Knight, C.G., Ed.; Elsevier/North-Holland Biomedical Press: Amsterdam, 1981; 243–272.
42. Mishima, K.; Satoh, K.; Ogiwara, T. Optical Birefringence of Phosphatidylcholine Liposomes in Gel Phases. *Biochim. Biophys. Acta* **1977**, *898*, 231–238.
43. Yi, P.N.; MacDonald, R.C. Temperature Dependence of Optical Properties of Aqueous Dispersions of Phosphatidylcholine. *Chem. Phys. Lipids* **1973**, *11*, 114–134.
44. Powers, L.; Pershan, P.S. Monodomain Samples of Dipalmitoyl Phosphatidylcholine with Varying Concentrations of Water and Other Ingredients. *Biophys. J.* **1977**, *20*, 137–152.
45. Tanaka, Y.; Schroit, A.J. Calcium/Phosphate-Induced Immobilization of Fluorescent Phosphatidylserine in Synthetic Bilayer Membranes: Inhibition of Lipid Transfer Between Vesicles. *Biochemistry* **1986**, *25*, 2141–2148.
46. Nichols, J.W.; Pagano, R.E. Resonance Energy Transfer Assay of Protein-Mediated Lipid Transfer Between Vesicles. *J. Biol. Chem.* **1983**, *258*, 5368–5371.
47. Bundgaard, H. Prodrugs as a Means to Improve the Delivery of Peptide Drugs. *Adv. Drug Del. Rev.* **1992**, *8* (1), 1–38.
48. Lee, V.H.L.; Li, V.H.K. Prodrugs for Improved Ocular Drug Delivery. *Adv. Drug Del. Rev.* **1989**, *3* (1), 1–38.
49. Sloan, K.B., Ed. *Prodrugs. Topical and Ocular Drug Delivery*; Marcel Dekker, Inc.: New York, 1992; 313.
50. Roche, E.B. *Bioreversible Carriers in Drug Design: Theory and Application*; Pergamon Press: New York, 1987; 292.
51. Lee, V.H.L.; Yamamoto, A.; Kompella, U.B. Mucosal Penetration Enhancers for Facilitation of Peptide and Protein Drug Absorption. *CRC Critical Reviews in Therapeut. Drug Carriers* **1991**, *8* (2), 91–192.
52. Lee, V.H.L. Enzymatic Barriers to Peptide and Protein Absorption and the Use of Penetration Enhancers to Modify Absorption. *Delivery Systems for Peptide Drugs*; Davis, S.S., Illum, L., Tomlinson, E., Eds.; Plenum: New York, 1986; 87–104.
53. Wearley, L.L. Recent Progress in Protein and Peptide Delivery by Noninvasive Routes. *CRC Critical Reviews in Therapeut. Drug Carriers* **1991**, *8* (4), 331–394.
54. Bodde, H.E.; Verhoeven, J.; van Driel, L.M.J. The Skin Compliance of Transdermal Drug Delivery Systems. *CRC Critical Reviews in Therapeut. Drug Carriers* **1989**, *6* (1), 87–115.
55. Knepp, V.M.; Hadgraft, J.; Guy, R.H. Transdermal Drug Delivery: Problems and Possibilities. *CRC Critical Reviews in Therapeut. Drug Carriers* **1987**, *4*, 13.
56. Carstensen, J.T. *Theory of Pharmaceutical Systems*; (General Principles) and 2 (Heterogeneous Systems) Academic Press: New York, 1972 and 1973; Vol. 1.
57. Weiner, A.L. Integrity Protected Gelatin. US Patent 5,376,381, 1994.
58. Wang, Y.-C.J.; Kowal, R.R. Review of Excipients and pH's for Parenteral Products Used in the United States. *J. Parent. Drug Assoc.* **1980**, *34*, 452–462.
59. Weiner, A.L. Emulsions. US Patent 5,626,873, 1997.
60. Weiner, A.L. Emulsions. US Patent 5,171,737, 1992.
61. Collins-Gold, L.C.; Lyons, R.T.; Bartholow, L.C. Parenteral Emulsions for Drug Delivery. *Adv. Drug Del. Rev.* **1990**, *5* (3), 189–208.
62. Benita, S., Ed. *Submicron Emulsions in Drug Targeting and Delivery (Drug Targeting and Delivery)*; Harwood Academic Publishers/Gordon and Breach Publishing Group: Newark, NJ, 1999; Vol. 9, 352.
63. Davis, S.S.; Washington, P.W.; Illum, L.; Liversidge, G.; Sternson, L.; Kirsh, R. Lipid Emulsions as Drug Delivery Systems. *Ann. N.Y. Acad. Sci.* **1987**, *507*, 75–88.
64. Davis, S.S.; Hadgraft, J.; Palin, K. Medical and Pharmaceutical Applications of Emulsions. In *Encyclopedia of Emulsion Technology*; Beecher, P., Ed.; Marcel Dekker, Inc. New York, 1985; Vol. 2, 159–238.
65. Singh, M.; Ravin, L. Parenteral Emulsions as Drug Carrier Systems. *J. Parenter. Sci. Technol.* **1986**, *40*, 34–41.

66. Hansrani, P.K.; Davis, S.S.; Groves, M.J. The Preparation and Properties of Sterile Intravenous Emulsions. *J. Parent. Sci. and Technol.* **1983**, *37*, 145–150.
67. Davis, S.S.; Hansrani, P.K. The Evaluation of Parenterally Administered Emulsion Formulations. *Radionuclide Imaging in Drug Research*; Wilson, C.G., Hardy, J.G., Frier, M., Davis, S.S., Eds.; Croom Helm, 1982; 217–241.
68. Davis, S.S.; Illum, L. Colloidal Delivery System. *Site Specific Drug Delivery*; Tomlinson, E., Davis, S.S., Eds.; John Wiley & Sons: Chichester, 1986; 93–110.
69. Shek, P.N., Ed. *Liposomes in Biomedical Applications (Drug Targeting and Delivery)*; Harwood Academic Publishers/Gordon and Breach Publishing Group: Newark, NJ, 1995; 6, 304.
70. Ostro, M.J., Ed. *Liposomes: From Biophysics to Therapeutics*; Marcel Dekker, Inc.: New York, 1987; 393.
71. Janoff, A.S., Ed. *Liposomes: Rational Design*; Marcel Dekker, Inc.: New York, 1998; 451.
72. Lasic, D.D.; Papahadjopoulos, D., Eds. *Medical Applications of Liposomes*; Elsevier Science Ltd.: Amsterdam, 1998; 779.
73. Barenholtz, Y.; Crommelin, D.J.A. Liposomes as Pharmaceutical Dosage Forms. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J., Eds.; Marcel Dekker, Inc.: New York, 1993; 9, 1–39.
74. Husain, D.; Miller, J.W.; Michaud, N.; Connolly, E.; Flotte, T.J.; Gragoudas, E.S. Intravenous Infusion of Liposomal BPD for Photodynamic Therapy of Experimental Choroidal Neovascularization. *Arch. Ophthalmol.* **1996**, *114*, 978–985.
75. de Smidt, P.C.; van Berkel, Th.J.C. LDL-mediated Drug Targeting. *CRC Critical Reviews in Therapeut. Drug Carriers* **1990**, *7* (2), 99–120.
76. Bijsterbosch, M.K.; van Berkel, Th.J.C. Native and Modified Lipoproteins as Drug Delivery Systems. *Adv. Drug Del. Res.* **1990**, *5* (3), 231–252.
77. Hara, T.; Tan, Y.; Huang, L. In Vivo Gene Delivery to the Liver Using Reconstituted Chylomicron Remnants as a Novel Nonviral Vector. *Proc. Nat. Acad. Sci.* **1997**, *94* (26), 14547–14552.
78. de Smidt, P.C.; Bijsterbosch, M.K.; van Berkel, Th.J.C. LDL as a Carrier in Site Specific Delivery. *Targeted Therapeutic Systems*; Tyle, P., Ram, B.P., Eds.; Marcel Dekker, Inc.: New York, 1990; 355–383.
79. *Tableting Specification Manual*, 4th Edn.; American Pharmaceutical Association, Tableting Specification Steering Committee: Washington, D.C., 1995; 114.
80. Banker, G.S.; Rhodes, C.T. *Modern Pharmaceutics*, 3rd Ed.; Marcel Dekker, Inc.: New York, 1996; 943.
81. Lieberman, H.A.; Lachman, L.; Schwartz, J.B. *Pharmaceutical Dosage Forms: Tablets*; Marcel Dekker, Inc.: New York, 1990; 1–3.
82. Wang, P.Y. Lipids as Excipient in Sustained Release Insulin Implants. *Int. J. Pharmaceut.* **1989**, *54*, 223–230.
83. Joseph, A.A.; Hill, J.L.; Patel, J.; Patel, S.; Kincel, F.A. Sustained Release Hormonal Preparations XV: Release of Progesterone from Cholesterol Pellets In Vivo. *J. Pharm. Sci.* **1977**, *66*, 490–493.
84. Kuo, H.S.; Muthua, S.C.; Thompson, C.R. Fat Emulsion Pyrogenicity Test. US Patent 4,245,044, 1981.
85. Dijkstra, J.; Mellors, J.W.; Ryan, J.L.; Szoka, F.C. Modulation of the Biological Activity of Bacterial Endotoxin by Incorporation into Liposomes. *J. Immunol.* **1987**, *138*, 2663–2670.
86. Leenaars, M.; Koedam, M.A.; Hendriksen, C.F.; Claassen, E. Immune Responses and Side Effects of Five Different Oil-Based Adjuvants in Mice. *Vet. Immunol. Immunopathol.* **1998**, *61* (2–4), 291–304.
87. Alving, C.R.; Richards, R.L. Immunologic Aspects of Liposomes. *Liposomes*; Ostro, M.J., Ed.; Marcel Dekker, Inc.: New York, 1983; 209–287.
88. Rauch, J.; Janoff, A.S. Antibodies against Phospholipids other than Cardiolipin: Potential Roles for Both Phospholipid and Protein. *Lupus* **1996**, *5* (5), 498–502.

LAMINAR AIRFLOW EQUIPMENT: APPLICATIONS AND OPERATION

Gregory F. Peters

Lab Safety Corporation, Des Plaines, Illinois, U.S.A.

INTRODUCTION

The number and complexity of pharmaceutical manufacturing and compounding processes requiring protection from airborne contaminants has increased substantially in recent years. Because pathogenic viable and nonviable contamination may be readily introduced into a patient along with a therapeutic parenteral drug, the sterility and purity of parenterals must be controlled in the manufacture and assured in the compounding of these products. Laminar airflow (LAF) equipment is widely used as an engineering control in aseptic processing to provide a production environment free of airborne and resulting surface contamination by microorganisms, pyrogenic and drug residues, and other materials that present a risk of intravascular infection, pyrogenic response, or occlusion of the peripheral vasculature. With the growth of the small- and intermediate-size generic drug manufacturing, drug repackaging, and diverse hospital pharmacy and home health care IV admixtures, compounding industries, clean space design and management has become the direct responsibility of an increasing number of middle- and line-management personnel. As such, a working knowledge of LAF theory, aseptic processing, and clean space management is integral to the conceptualization, construction, and operation of a safe and effective clean space, and an important consideration in the selection or retention of the clean space manager and operative personnel.

CONTAMINATION CONTROL

All manufactured products are vulnerable to contamination by a myriad of aerosolized contaminants, including microorganisms, pyrogenic dust, ash, pollen, smoke, hydrocarbons, and other chemicals that are omnipresent in the environment (Fig. 1). Because of the potential dangers to the patient resulting from a parenteral product containing even minute quantities of these contaminants, exceptional measures are required to exclude them from the finished product. Careful planning is essential

in preventing contamination of the environment leading to occupational exposure of personnel to hazardous substances, which are routinely manipulated in pharmacy operations. The primary objective of the aseptic process is control and elimination of viable contaminants. These contaminants are numerous and varied, normally consisting of bacteria, fungi, and viruses. Viruses are usually short-lived upon exposure to air, and require host systems to remain viable and reproduce; they are generally not of direct concern, except in excluding their vector, or transport mechanism. Aseptic processes are designed to exclude bacteria and fungi as well as their spore forms and breakdown products. Many of these contaminants are naturally airborne and occur commonly in the atmosphere. However, the contamination of greatest concern in aseptic processing is the endogenous microbiologic material generated by the operative personnel and others involved in the manipulation of parenteral products. This type of contamination is easily aerosolized and introduced into air currents by normal "shedding" of endogenous microbiota, and by mechanical means as microscopic droplets of sputum, produced by talking, laughing, or sneezing. Endogenous contamination is generated in enormous quantities on the skin^a (1) and may be deposited on the surfaces of containers, equipment, gowns, and materials introduced into the aseptic work field during the course of manipulations.

Waterborne contaminants may also be introduced by contaminated cleaning agents or poor cleaning and sanitizing techniques. All types of gross microbiologic contamination are found on work surfaces, gloves, and compounding materials following contact with contaminated objects. Pyrogenic and nonpyrogenic dust must also be excluded from parenteral products. Even minute quantities of this material may cause acute inflammation or abscess at the injection site, and induce a life-threatening pyrogenic response. In addition, evidence of long term dangers of fiber emboli in producing pulmonary and

^aThe average adult human sheds 25,000–50,000 *Staphylococcus epidermidis* particles per minute; one person in five is a carrier and active producer of *Staphylococcus aureus*, the organism responsible for toxic shock syndrome, a bacteremia fatal in 50–90% of reported cases.

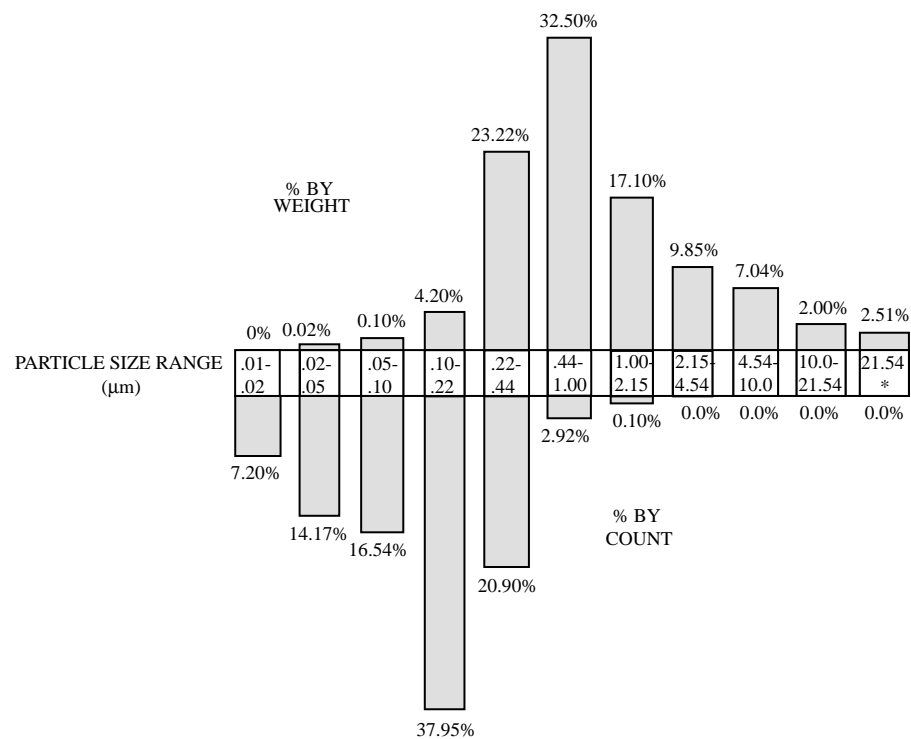


Fig. 1 Distribution by size and weight percent of particulates in normal atmospheric air. (Courtesy of American Air Filter Research, Louisville, Kentucky.)

cerebral granulomas, stenosis, and occlusion of micro-circulation, as well as clouded vision, and neurological sequelae in patients receiving particulate-contaminated intravenous admixtures has emerged (2, 3). Additionally, drug product residues remaining in the aseptic processing field as a result of improper line-clearance are a threat to the patient for whom they are not intended.

History

LAF is an “offshoot technology” stemming from development of the high efficiency particulate air (HEPA) filter, spearheaded by the U.S. Army Chemical Corps, Naval Research Laboratories, and the Atomic Energy Commission in the 1940s and 1950s. Known as an absolute filter, the HEPA filter was further developed by the nuclear industry to provide fail-safe removal of extremely hazardous microscopic airborne particulates^b at nuclear facilities. In the late 1950s, a proliferation of laminar flow clean benches (LFCBs), incorporating a HEPA filter, made it possible for the hospital pharmacy to achieve a small, but exclusive compounding environment of sterile air and sanitized, bacteriostatic worksurfaces in

^b99.97% of particulates smaller and larger than 0.3 microns.

which to prepare small numbers of individual, patient-specific sterile products.

On a broader scale, pharmaceutical manufacturers were beginning to utilize absolute filtration as a primary engineering control in the maintenance of large, carefully controlled clean spaces in the batch production of quality-controlled parenteral products. In this application, LAF was supplied directly to production lines and extended critical worksurfaces within defined, nonturbulent entrance and exit planes as parallel or “columnated” airflow (misnamed “laminar flow”).^c This highly controlled laminar airstream was supplied to the critical worksurface, in addition to conventionally supplied turbulent airflow to the general space, provided through terminal diffusers for filtration of the balance of room air. In this manner, the “stepped” control of all critical, as well as support areas was achieved.

In the mid-1970s, increased control of manufacturing process air quality became possible with the refinement of LAF clean space design, the growing body of

^cLaminar flow is defined as a fluid stream having discernible differentiations (laminations or layers) of velocity, pressure, temperature, or other characteristic, whereas parallel flow is uniform throughout its vertical, horizontal and longitudinal extent.

historical process quality data, the refinement of statistical process controls (SPCs), and the emergence of industry operating standards (4). In 1979, concern for the safety of pharmacy personnel compounding anti-neoplastic drugs in LFCBs was expressed, following studies that clearly identified the potential health risks to these personnel (5). The introduction and widespread use of the Class II laminar flow biological safety cabinet (BSC) by oncology/hematology pharmacy occurred in the late 1970s, and was almost universal by the end of the 1980s. Additional studies conducted in the late 1990s clearly indicated that, in spite of pharmacy-wide implementation of Class II BSC containment technology, the same potential health risks identified in 1979 continued to exist in the workplace (6). The swift, anecdotal response in some segments of the pharmacy industry was to strongly recommend immediate retrofit of the Class II design with the Class III BSC, or Barrier Isolator (sometimes referred to as a “barrier hood”) in the glovebox or half-suit configuration, as an alternative to “open” Class II units used by pharmacy in a manner that had ostensibly failed to protect workers and the environment. This recommendation, however, was not based upon adequate and careful engineering failure analysis (FA) of the Class II design as implemented in pharmacy operations, by which the Practice might rationally evaluate and address the reasons for such failure. Neither, as of this writing, have any systematic, scientifically based feasibility studies, validation and monitoring protocols, or operating specifications leading to proper, corrective retrofit with the Class III or barrier isolator designs been developed as a corollary to such FA. (It should be noted that not all “barrier isolators” or “gloveboxes” are Class III BSCs.) These steps are absolutely essential in preventing a repeat failure of any type of BSC in the engineering control of hazardous substances manipulated by pharmacy personnel, and the resulting characterization of the IV pharmacy as a serious occupational risk environment.

THE HEPA FILTER

The essential element common to all LAF equipment is the HEPA filter.

Construction

The HEPA filter is normally constructed of borosilicate microfibers, formed into a flat sheet by a process similar to papermaking (Fig. 2). This sheet is pleated to increase the

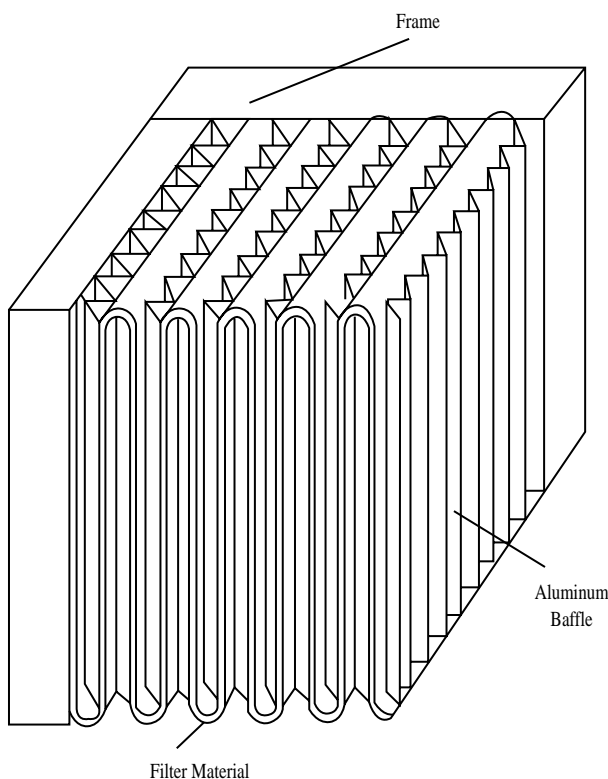


Fig. 2 HEPA filter cross-section showing pleated construction. (Courtesy of the Baker Company, Sanford, Maine.)

overall filtration surface area, in order to minimize the static pressure drop across the filter to a predetermined value^d at a given airflow specification. The pleats are separated by serrated aluminum baffles or stitched fabric ribbons, which direct airflow through the filter. This combination of pleated sheets and baffles facilitates maximum exposure of the upstream filtration area to the airstream and is referred to as the filtration medium. It is constructed in a predetermined size, and installed into an outer frame made of fire-rated particle board, aluminum, or stainless steel. The frame-media junctions are permanently glued or “pot-sealed” to ensure a leakproof bond. The frame is fitted with a continuous, closed-cell neoprene gasket or other suitable occlusive seal to provide a gas-tight installation of the filter into the air handling system.

Filtration Efficiency

Refinement of HEPA filter manufacturing and testing technology and development of the ultra low particulate air (ULPA) filter, have led to an increase of absolute

^dNormally 0.50–1.2 inches. water column (WC) or water gauge (WG).

filtration retention efficiency of greater than two orders of magnitude above 99.97%. HEPA filtration efficiencies range from a minimum of 99.97 to 99.99%, with ULPA efficiencies above 99.9999% for particulates larger and smaller than 0.3 μm in diameter (Fig. 3). (HEPA filters used in pharmaceutical and pharmacy-compounding applications rarely exceed 99.99%.) Expressed another way, the HEPA filter is capable of trapping and retaining 999,700–999,900 of every 1,000,000 particles smaller and larger than 0.3 μm in diameter.

Filtration Mechanisms

The three principal filtration mechanisms (Fig. 4) by which aerosols are collected on the HEPA filtration medium are:

- Inertial impaction, where particle inertia causes it to leave the flow streamlines and impact on the fiber;
- Interception, a screening effect dependent upon particle fiber-size relationships; and
- Diffusion, a Brownian motion diffusion of very small particles due to molecular bombardment (7).

Other filtration mechanisms, such as sedimentation and electrostatic attraction, provide some degree of aerosol collection, however most particulates are removed from the airstream by the above methods.

Shipment, Storage, and Handling

The HEPA filter is extremely fragile and should be shipped, stored, and handled in the same manner as delicate instrumentation. Personnel responsible for receiving and handling HEPA filters should receive training in proper handling technique. All incoming HEPA filters should be visually inspected for apparent damage due to

mishandling, and all damage described in detail on the shipping documentation prior to acceptance.

Filtration Performance Certification

Individual HEPA filter efficiencies are established by an exacting challenge of the filter frame and medium, usually incorporating a “cold-boil,” polydisperse aerosol of dioctylphthalate (DOP) or equivalent (8), which is introduced into the upstream plenum-side of the filter in a manner that ensures even distribution of the test aerosol behind the filter, at its rated airflow.

Following verification of acceptable airflow velocities, expressed by the manufacturer in linear feet per minute (LFPM) or meters per second (m/s), leakage is determined by measuring the penetration of the test aerosol through the filter as a percentage of the upstream plenum aerosol concentration, using an aerosol photometer (4) or optical particle counter (9) to carefully scan the entire filter media face and frame. Penetration of the test aerosol at or above 0.01% of the upstream concentration is considered a leak requiring repair. Leakage should be repaired only with room temperature-vulcanizing (RTV) silicone caulk, which is easily applied, and exhibits long term stability and resistance to deformation. Repairs to the filtration medium should be made in the manner specified in the literature (10). Individual repair “patches” should not exceed 1.5 in. in length or width, nor should the sum total of all repairs exceed 3% of the total area of the filter face. During operation and at all times no object, drug residue, or debris should be permitted to come into contact with the HEPA filter.

LAF AS A BARRIER TECHNIQUE

LAF has long been the primary method of controlling airborne contamination in the aseptic processing of pharmaceutical and pharmacy products. Also known as nonturbulent, or unidirectional airflow, LAF is generally defined as “HEPA-filtered air having parallel streamlines, flowing in a single pass and direction through a clean zone” (4). LAF is technically defined as fluid flow without macroscopic fluctuations, which generally occur when the Reynolds number^e is less than 2000. Industry standards require that 80% or more of the total airflow exhibit this characteristic, in order to meet the

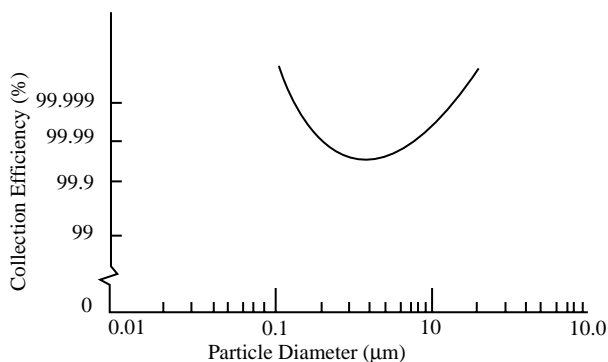


Fig. 3 Theoretical HEPA filter-collection efficiency. (Courtesy of the Baker Company, Sanford, Maine.)

^eThe “Reynolds number” is the ratio of inertial to viscous forces in a pipe or duct.

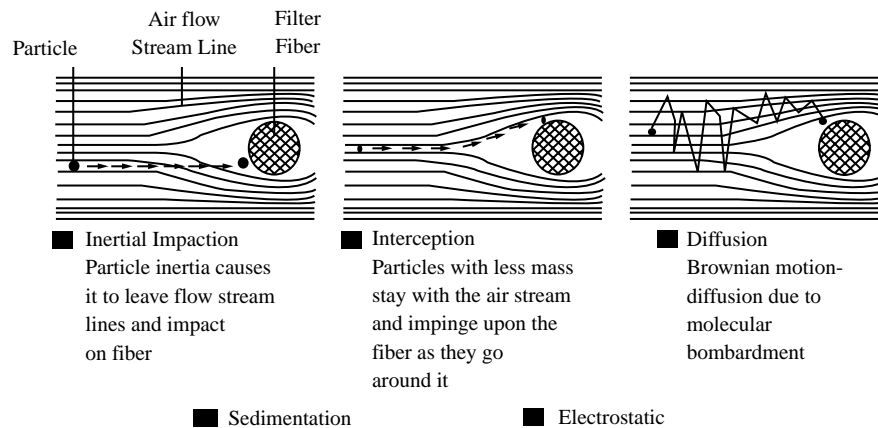


Fig. 4 Air filtration theory particle-collection mechanisms. (Courtesy of the Baker Company, Sanford, Maine.)

definition of LAF (11). LAF is necessary to maintain the most stringent air cleanliness classes^f (4) in the production of “first air”^g at the critical worksurface of the LFCB and BSC, and in operational cleanrooms (4,11). It is important to note that the terms “laminar airflow” and “Class 100” are not interchangeable; LAF first air incorporating properly validated HEPA filtration produces air cleanliness of approximately Class 1, nearly two orders of magnitude cleaner than Class 100. Class 100 nonlaminar airflow does not constitute first air within a critical work zone, and should not be substituted in applications requiring or designating LAF.

Conventional Airflow

Conventional airflow (also known as turbulent, or non-unidirectional airflow) incorporates HEPA filters, located in-duct, or as room terminal filtration modules (TFMs; Fig. 5). Often confused with LAF, conventional airflow does not meet that definition because it allows multiple-pass circulating characteristics or a nonparallel airflow direction, or both. This type of airflow is incapable of producing first air, and is normally used as secondary or “buffer” filtration in treating a processing or compounding space that contains laminar airflow devices (LAFDs) to maintain primary critical work surface conditions, or in treating other processing or

support areas about which a definitive air cleanliness statement must be made. Properly designed, a conventional airflow system is effective in maintaining the less-stringent air cleanliness classes^f in operational cleanrooms (4, 11).

Advantages

When used properly by trained personnel, employing adequate process controls, the LAF environment provides a reliable barrier to measurable airborne viable and nonviable, solid particulate contamination, which may defeat the aseptic process. LFCBs, BSCs, and heating, ventilation, and air conditioning (HVAC) installations are easily validated. These systems normally continue in operation with little or no variation in output quality for long periods of time, and are easily maintained and tested.

Limitations

Any discussion of LAF equipment must include consideration of aseptic technique and the aseptic process as a whole. Often relied upon as an infallible process support, LAF is in fact a fragile, protective envelope of slow-moving, aerosol-free air^h (4) the effect of which is easily disrupted and defeated by improper placement of processing materials, poor manufacturing and personnel practices, inadequate aseptic technique, or failure to maintain the HVAC components. Although the laminar slip stream itself is free of particulates, it does not eliminate particulates or other surface contaminants present in the aseptic field, or contamination introduced into the aseptic field on the surfaces of improperly prepared

^f“... Most stringent air cleanliness classes ...” are defined as class 10–100, (F.S. 209e); Classes C, D, and E (British Standard 5295); Classes 4–5 (EN ISO); Grades A–B (EC). “...less stringent air cleanliness classes...” are defined as Class 1,000–100,000 (F.S. 209e); Classes F through K (B.S. 5295); Classes 6–8 (EN ISO); Grades C–D (EC).

^g“First air” is uninterrupted air issuing directly from a HEPA filter in a laminar-airflow environment.

^hAirflow velocity exiting an unobstructed workstation should be maintained at 90 LFPM average with a uniformity within $\pm 20\%$ across the entire area of the exit.

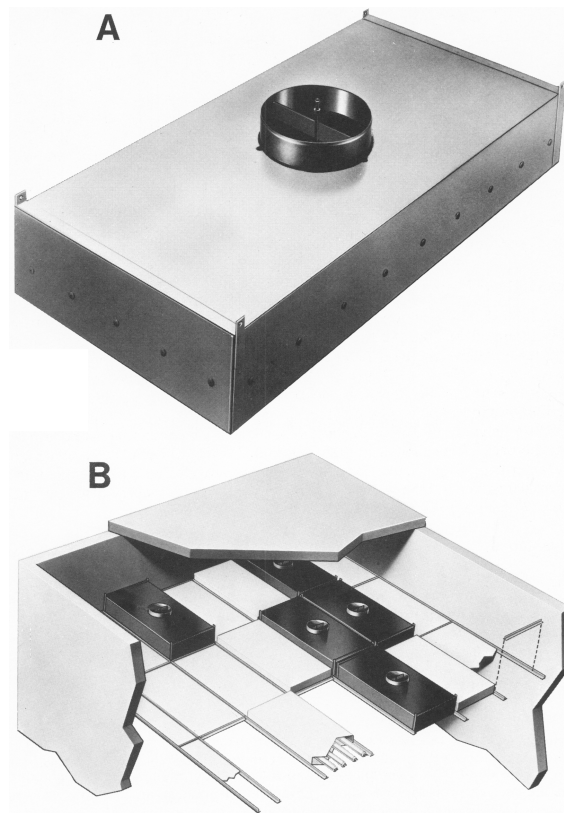


Fig. 5 (A) The self-contained terminal HEPA filtration module (TFM). (B) Arrangement of TFMs in the T-Bar ceiling in a conventional-flow application. (Courtesy of American Air Filter, Inc., Louisville, Kentucky.)

processing materials, and their manipulation and storage within the aseptic environment. Neither is LAF a substitute for proper manipulative and aseptic technique. Excessive confidence in any LAF system should not lead to the neglect of proper precleaning and staging of processing materials, personnel selection, training, and validationⁱ (12); effective routine housekeeping and maintenance procedures (13); and aseptic process monitoring and auditing methods.^j This may result in a breakdown of the process and a compromise of product integrity. In addition,

ⁱ1) Personnel selection and screening criteria; 2) a formalized training program to include a period of supervised manufacturing or clinical experience culminating in a recommendation for validation by the supervisor; 3) a personnel validation method, including a written exam with a required passing grade, and a practical assessment of aseptic technique, utilizing sterile microbiologic growth media in a process simulation incorporating all processing steps encountered by the candidate during the actual processing operation.

^jQuality assurance (QA) testing and recertification of LAF systems and other engineering controls, which support the aseptic process (process auditing), irrespective of quality control, or procedures to demonstrate conformance with product specifications of identity, purity, sterility, and apyrogenicity (product auditing).

the use of improper manipulative technique in BSCs and containment systems may result in a compromise of the waste-stream, resulting in cumulative contamination of operative personnel and the environment.

Operating procedures

Although the design of each type of LAFD dictates certain specific operating procedures, several general principles apply to all LAFDs (9).

Cleaning and preparation: LFCB: The LFCB should be allowed to run for at least 30 min before the commencement of aseptic operations. All work-zone accessible surfaces, with the exception of the filter-protective screen, should be cleaned and sanitized by application and recovery of a low-residuing, water-base disinfectant cleanser (household bleach or other hypochlorite solutions should not be used at any time on stainless steel surfaces), followed by application of 70% ethyl or isopropyl alcohol sprayed evenly length-wise across the work surface from the back of the cabinet to the front, and allowed to dry (14). (The combination of a water-base disinfectant cleanser with alcohol provides

both the broadest antimicrobial action and widest range of surfactant and solubility factors for recovery of surface residues during the cleaning process.)

BSC: The BSC should operate continuously to ensure containment of hazardous substances. All work-zone interior surfaces with the exception of the HEPA filter-protective screen should be cleaned and sanitized in the manner of the LFCB in the proper order to ensure protection of operator's garb from contaminants and cleaning residues during the cleaning process. This will also prevent transfer of drug residues to the general environment (14).

Staging

After drying of the work surface, operations in the LAFD should begin with staging of all working materials for introduction into the aseptic work area (14). This should include:

- assembly of required working materials: drug components, syringes and sterile fluid pathways, diluents, dispensing and venting devices, wipes, final containers, etc.
- preparation, in a Class 100,000 environment (4), of all working materials to go into the aseptic field by sanitizing with 70% ethyl or isopropyl alcohol spray, wiping all containers, careful removal of gross contamination and filth from any container that does not have an outer wrapper, removal of inner containers from their outer wrappers, and placement of all materials on a sanitized, stainless steel surface. This is preferably a cart or tray at a close proximity to the LAFD for direct loading of materials into the aseptic work zone.

Materials such as paper, labels, writing implements, etc., should not be placed into the work zone.

Aseptic manipulation

Prior to introduction of the working materials into the LAFD, the gloved hands of the operator should be thoroughly washed and rinsed to remove dry lubricants, sanitized by spraying with 70% alcohol, and allowed to dry in the laminar airstream. The working materials may then be transferred to the work zone, and aseptic manipulations begun. The operator's hands should be slowly inserted into, and removed from the laminar airstream, in order to minimize backwash and cross-stream contamination of the work zone. Working materials should be arranged in such a way that work progresses laterally, from right to left or left to right, so that non-interrupted first air is continuously supplied to the critical surfaces of all working materials at all

times. If the operator must leave the work zone area, his or her gloved hands should be resanitized with 70% ethyl or isopropyl alcohol prior to reentering the work zone. This practice takes little time and minimizes contamination of the work zone by endogenous and residual environmental flora. An alcohol-spray bottle or other suitable dispenser should be provided close to the work zone entrance for this purpose, with the alcohol filtered and the dispenser sanitized each time the dispenser is refilled. Good aseptic technique is essential to retain the sterility of compounded products, and properly fitting surgeon's gloves, mask, laboratory coat (or arm barrier), and hair cover are recommended for all pharmacy operators working in the LAFD^a (1, 14).

LFCB

The oldest and most basic LAFD is the LFCB, universally referred to as a "hood"—an enclosed work area with its own HEPA-filtered air supply (Fig. 6). It provides only product protection by capturing room air, passing it through a HEPA filter, and directing the filtered air horizontally or vertically uniformly across the work surface toward the operator at a constant speed.

Limitations

The LFCB should not be used in operations requiring manipulation of cytotoxic, radioactive, microbiologic, or other hazardous materials, which may become aerosolized and aspirated by the operator. Reconstitution and manipulation of antineoplastics and vesicants, mass reconstitution of antibiotics, antivirals, vaccine formulation, and similar manipulations should be done in a laminar flow BSC.

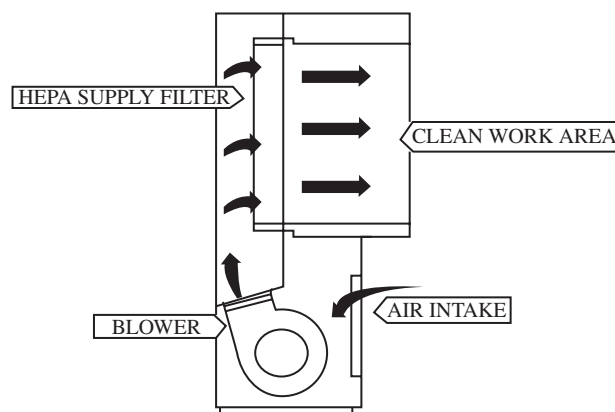


Fig. 6 The LFCB airflow profile. (Courtesy of the Baker Company, Sanford, Maine.)

Placement of materials in the work zone

Working materials placed into the LFCB should be positioned a minimum of three object-diameters in from the open end of the unit when the object is exposed to the laminar airstream on all sides, and a minimum of six diameters in from the open end when working at either end of the unit or when the object is exposed to the laminar airstream on only one side (14). This general method counteracts backwash contamination, which may compromise the aseptic field. The introduction of any large object (automated, high speed compounding devices, water baths, carboys, etc.), obstruction, or complicated process into the laminar airstream may likewise induce backwash contamination, and should be validated to maintain cohesion of the airstream and exclude both backwash and cross-stream contamination during the process. This may be accomplished by the use of visual tracers, such as smoke sticks or other smoke-producing devices, to introduce quantities of smoke both upstream of the critical worksurface and at the open end of the unit during a process validation or qualification run for visual observation of airstream behavior. Work should take place not less than 6 in. in from the open end of the work surface. Although engineering improvements in the laminar airstream recovery have been made, the laminar airstream is slow-moving,^k and care should be exercised in the location, operation, and maintenance of this unit.

Location and performance testing

The LFCB should be located in an area free of ventilation or other air currents, steady or intermittent, which might hamper the laminar airstream by backwash effect.^l Performance testing of the LFCB by a qualified inspector is recommended at least every 6 months, and servicing or replacement of the unit prefilters every 60 days or less.

Applications

The LFCB is used in hospital, clinical, and home health care pharmacies to provide a sterile field in which to conduct aseptic manipulations in the compounding of large-volume parenterals (LVPs) in the form of IV admixtures, hyperalimentation, and small-volume parenterals (SVPs) in the form of piggybacks, syringes, or other parenteral products of less than 250 ml, and for general sterile manipulation of nonhazardous materials.

^kNormally 90 LFPM \pm 20% or as specified by the user.

^l“Backwash contamination” is the general term given to the reflux entry of unfiltered room air into the LAFD work zone.

In industry, the LFCB is used to conduct small batch sterile filling operations, in the general manipulation and isolation of nonhazardous materials, and in quality assurance/quality control (QA/QC) sterility testing.

Laminar Flow BSC

Pharmaceutical and clinical research in the past four decades has led to the development of drug products and other hazardous substances, the manufacture, handling, and compounding of which are considered hazardous to operative personnel in both the short and the long term (15). In addition to protection of the purity and sterility of the product, it is necessary to consider protection of personnel and the environment. The need to protect both the product and personnel has resulted in the adaptation and use of a variety of laminar flow BSCs in the manufacture and compounding of numerous biological, radioactive, cytotoxic, allergenic, and antibiotic drug products. It is necessary that proper containment and barrier techniques be followed in the preparation, operation, and cleanup of any BSC. Although both are considered to be LAFDs, a clear distinction between the LFCB and BSC should be made in the training of operative personnel. There should be no generalized “grouping” of BSCs and LFCBs as “hoods” requiring similar use and maintenance patterns; each type of LAFD has airflow patterns, containment characteristics, and operating requirements unique to its design (Fig. 7). Operative personnel must understand that the containment and barrier techniques used in the operation of the BSC are significant not only as protection against sudden, overt contamination, but more importantly, as barriers to traces of residual contamination to which constant exposure may present long term health risks (15,16). Current pharmacy practices and contamination control manipulative techniques have been shown to be inadequate to contain hazardous substances in all phases of compounding and administration, and a pharmacy-wide study and standardized, remedial training programs must be quickly developed and carried out to protect operative personnel (6, 17).

Classification

BSCs are divided into three classes. Class I and Class II are used for low to moderate risk agents, and Class III for high risk agents. The risk level serves as a guide for pharmaceutical manufacturing and pharmacy compounding operations based on a reasonable extrapolation from the personnel risk levels, containment models, and product/personnel protection factors encountered in

operations using the biological agents for which this type of equipment was originally designed.

The BSC in pharmacy operations

Although the BSC was not specifically designed for pharmaceutical manufacturing or pharmacy compounding operations, reasonable analogies in the management of hazardous aerosols may be made, justifying use of the BSC in such operations. Such was the case in 1979, following a comprehensive study (5) that determined that pharmacy and nursing personnel were experiencing occupational exposures to antineoplastic agents following their compounding in LFCBs and routine administration. As a prudent response to these findings, recommendations based upon such analogies by segments of the pharmacy community facilitated pharmacy-wide use of Class II BSCs. The Class II design was quickly implemented without attendant feasibility studies, validation and monitoring protocols, or standardized operating, cleaning, and line-clearance procedures, resulting in the installation and operation of this equipment with wide variation in its effectiveness from institution to institution.

Recent studies have indicated the unabated occupational exposure of pharmacy personnel to these substances, causing an inference of the inadequacy of the Class II system in controlling environmental contamination (6). These findings have again prompted vendors and segments of the pharmacy community to aggressively promote use of the Class III BSC, or barrier isolator, postulating that a so-called "closed system" would unfailingly prevent the transfer of hazardous agents to the environment, thus preventing personnel exposures. Pharmacy-wide promotion of a containment technology is

thus being repeated without the benefit of comprehensive engineering studies to both identify the reasons for any "failure" of the Class II system, and to provide standardized procedures to properly implement and operate the Class III system. It must be noted that Class II systems are completely effective in the management of particulate and aerosol contaminants in the microbiologic and toxicologic procedures for which they were originally designed. These contaminants include numerous dangerous agents, for which exacting, albeit decades-old manipulative and management techniques remain effective. No applications' failure of the Class II system has yet been reported by these industries, implying that it is improper use of this equipment and/or poor manipulative, product transfer, and cleaning techniques by pharmacy and nursing personnel, rather than any inherent design or applications' flaw in the Class II system, which would account for unabated occupational exposures (17). At this writing, insufficient evidence appears to exist supporting the remedial, pharmacy-wide retrofit of the Class II system with the Class III BSC, or barrier isolator. The pharmacy community should, rather than prematurely dispose of an effective and functional containment system for a more complex and expensive design, seek to identify and correct the cause(s) of such personnel exposures. A controlled study measuring the nature and magnitude of personnel exposures prior to, and following implementation of carefully designed and executed training exercises, containment procedures, and waste-streaming techniques should be carried out prior to any measures to retrofit these engineering controls.

Following any scientific determination of the inadequacy of the Class II technology in providing the necessary

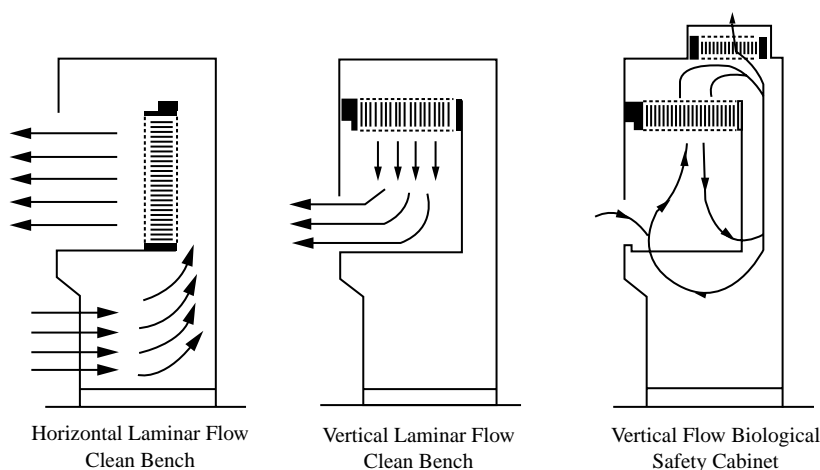


Fig. 7 Airflow patterns in horizontal- and vertical-flow LFCBs and vertical-flow laminar flow BSC. (Courtesy of the Baker Company, Sanford, Maine.)

levels of operator and environmental protection, the Class III and barrier isolator systems should, without presumptions, be carefully evaluated for use in this application. Such evaluation should systematically encompass the full range of performance, operational, and maintenance factors, which bear upon gloveboxes or other closed systems, such as lowered productivity because of the cumbersome nature of gloves, half-suits, and reduction of dexterity; difficulty in effective cleaning, resulting in increased residual and cross-contamination of products; inferior product protection characteristics of turbulent Class 100 supply airflow; lowered throughput; increased compounding time and errors; and the manner in which these factors may inhibit the product protection capabilities of the Class III design.

Suitability

Prior to choosing a BSC for any pharmaceutical manufacturing or pharmacy compounding operation, all risks should be assessed by a qualified process engineer, safety officer, or industrial hygienist, ensuring that the equipment meets occupational safety as well as process requirements.

Class I BSC: The Class I BSC is an open containment unit suitable for work involving agents of low to moderate risk to the user and environment, where there is a need for containment but none for product protection or isolation. The Class I BSC provides protection to personnel using the cabinet by means of constant, controlled airflow into the work area and away from the operator, preventing the escape of aerosols through the front opening. It is of limited use in manufacturing and has no reported use in the current practice of pharmacy.

Class II BSC: The Class II BSC (Fig. 9) provides product, personnel, and environmental protection, and is the most common BSC employed in pharmaceutical manufacturing and pharmacy-compounding operations. The Class II BSC has several subclassifications, based upon cabinet ventilation design (18) (Table 1). The Class II BSC (Fig. 8), the most widely used by hospital and home-care pharmacies, features a front access opening

with carefully maintained inward airflow for replacement of air exhausted from the cabinet, a HEPA-filtered vertical laminar flow airstream within the entire work area, and HEPA-filtered exhaust air. The vertical laminar flow airstream and front access opening are common to all Class II cabinets, although LAF velocities and patterns, HEPA filter sizes and position, ventilation rates, and cabinet exhaust methods vary considerably in different designs (Fig. 7).

Class III BSC: The Class III BSC (Fig. 9) provides the highest level of personnel, product, and environmental protection from high-risk microbiologic and toxicologic agents. It is usually employed in pharmaceutical manufacturing operations involving weighing, diluting, and high volume aerosol generation of high-risk agents, as well as for handling contaminants that are slowly or rapidly vaporized. This type of laminar flow device affords the maximum containment and product protection barrier, and is used only in cases of extreme exposure hazard or product sensitivity.

The Class III BSC is a gas-tight enclosure, utilizing total air displacement ventilation that protects personnel from exposure to the products contained within the enclosure, the product from contaminants found in the ambient environment, and the environment from release of potentially hazardous substances. The Class III BSC is used where absolute containment of hazardous agents is required, and is normally configured with glove ports housing gas-tight, full length latex, neoprene, PVC, urethane, or laminated polymer gloves (19).

Operating procedure

In addition to cleaning all accessible work zone surfaces, the Class II BSC worksurface tray should be lifted up and back, and the area under the tray should be thoroughly cleaned with the same frequency as the other user-accessible worksurfaces; this is of particular importance in preventing the buildup of potentially harmful product residues. All materials used in cleaning and sanitizing should be treated as toxic waste, and disposed of in accordance with state and local

Table 1 BSC cabinet ventilation

(Type) Class II	Cabinet air characteristics	Air recirculated (%)
Type A	30% Vented back into room	70
Type B3	30% Ducted to outdoors	70
Type B1	70% Ducted to outdoors	30
Type B2	100% Ducted to outdoors	0

*

ordinances. Following complete drying of the work surface, operations may begin by staging all working materials for introduction into the aseptic work area. The working materials may then be transferred to the work zone, and aseptic manipulations begun. Work should be performed only on the work surface, taking care not to handle or store materials on or near the ventilation grilles. First air should be maintained at the critical surfaces of all working materials in the aseptic field; interference with the vertical flow of first air by passing anything over critical orifices or septa in the aseptic field must be prevented. Good aseptic technique is essential to retain the sterility of compounded products^m (14) and surgeon's gloves, mask, laboratory coat (or arm barrier), and hair cover are recommended for all pharmacy operators working in the BSC.ⁿ

Location

The location, operation, and maintenance of the BSC should be carefully planned. Similar to LFCB, the BSC should be located in an area free of steady or intermittent air currents, which might defeat the laminar airstream by a backwash effect^l. The BSC should be connected to an adequate power source, having the minimum possible current fluctuation. The effects of voltage variation on cabinet performance can be pronounced (Fig. 10A), causing a variation in intake and supply velocities sufficient to result in an unacceptable unit performance (Fig. 10B), thereby compromising personnel and product protection design features (Table 2). Performance testing of the BSC by a qualified inspector is recommended at least every 6 months and always after relocation of the unit. Appropriate surface decontamination is recommended prior to moving or refiltering the unit.^o

^mThe importance of good aseptic technique is occasionally deemphasized by personnel compounding cytotoxic and antibiotic agents in the BSC, based on the belief that these substances are themselves toxic or germicidal to any microbiologic contaminants. It should be noted that numerous microbiologic organisms remain viable and replicate in cytotoxic compounds, and that no antibiotic has a universal antimicrobial action. Proper aseptic technique is therefore mandated in all operations carried out in the LAF environment.

ⁿThese are basic barriers, proven to aid in the retention of gross endogenous contamination at close working proximity to the critical field, but as importantly, these barriers protect the operator from direct contact with potentially hazardous substances, and should be used in accordance with established guidelines (see "Regulatory Issues").

^oAlthough microbiologic decontamination of the BSC is not required, except as used in the manipulation of microbiologic agents, proper containment technique should be employed when refiltering or servicing these units.

Table 2 Protection provided by various BSC designs

Design	Personnel	Product	Environmental
Clean benches		•	
Class I	•		•
Class II			
Type A	•	•	•
Type B1	•	•	•
Type B2	•	•	•
Type B3	•	•	•
Class III	•	f(design) ^a	•

^aCharacteristic a function of design

Applications

The BSC is used in hospital, clinical, and home health care pharmacies to provide a sterile field in which to conduct aseptic manipulations in the compounding of LVPs, SVPs, antineoplastics, antibiotics, antivirals, and vaccines, the direct exposure to which may be hazardous to the operator and the environment. Most hospital and clinical pharmacy operations are currently carried out in a Class II(B3) BSC, which is vented to the outside by a dedicated, nonrecirculating HVAC exhaust, directly connected to the BSC.^p Although direct connection to an HVAC exhaust is not required for proper Operation of the Class II(a) BSC in the containment of biological aerosols, such direct connection is recommended in all pharmaceutical manufacturing or pharmacy-compounding applications. The Class I BSC is almost never used in pharmacy compounding operations, and the Class III BSC technology is currently being evaluated for use in pharmacy compounding following comprehensive evaluation of any Class II FA, and outcomes of feasibility studies of the Class III system in such operations.

In industry, the BSC is used to conduct small batch sterile-fill operations, manipulation (weighing and

^pCurrent guidelines may recommend use of the Class 11 Type B2 (total exhaust) BSC in certain manufacturing and compounding applications, but proper operation of this unit is difficult to maintain because of non-interlocked operation of separate BSC supply and in-house exhaust air-handling systems. B2s in the configurations currently available frequently develop a supply-exhaust flow imbalance, which may readily compromise product or personnel protection. Until a reliable system for direct interlock of the supply and exhaust air handlers is available that synchronizes the operation of the components, facilitating changes in operation by either air handler being proportionately matched by the other, use of the Class-11 Type B2 BSC is not recommended for pharmaceutical manufacturing and pharmacy compounding applications as a containment LAFD.

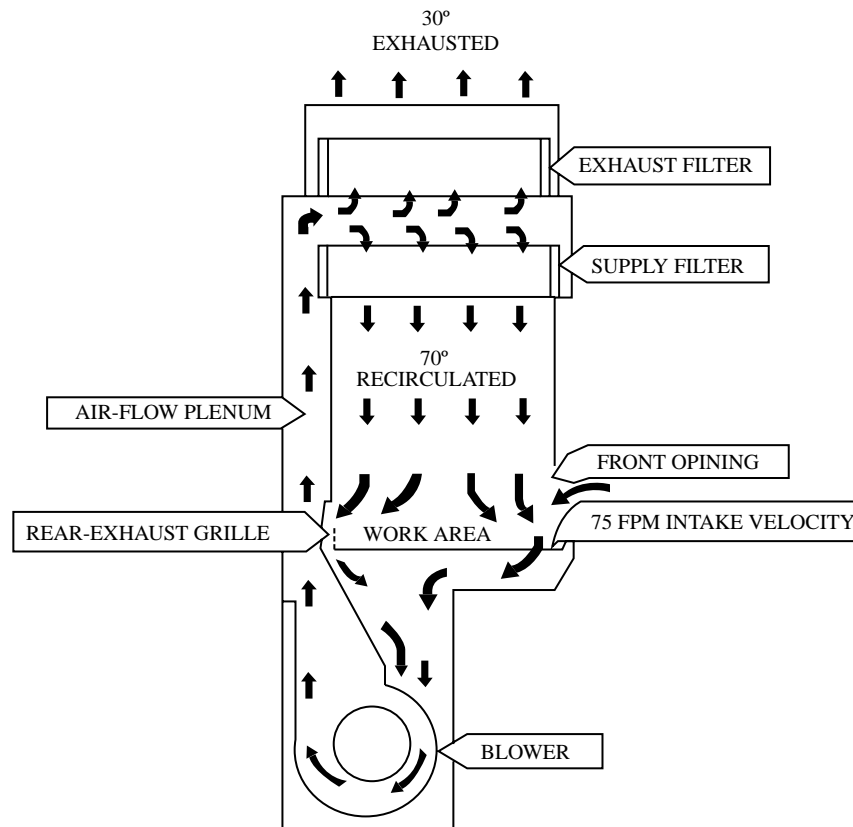


Fig. 8 Airflow patterns of the Class II(a) laminar flow BSC. (Courtesy of the Baker Company, Sanford, Maine.)

pouring), isolation of hazardous materials, and in QA/QC testing applications. All classes of BSC are encountered in pharmaceutical manufacturing operations for a wide variety of processing applications.

Terminal HEPA Filtration Module

The terminal HEPA filtration module (TFM) is a self-contained HEPA filter and plenum unit (Fig. 11), which may be used to provide laminar or conventional airflow⁴ to a clean space, or may be dedicated as a LAF workstation (20). The TFM is available with a 10 in. (optional 12 inch) collar for connection by a circular supply duct to a central air handling system (Fig. 5A) or as a free-standing, fully powered unit containing a motor

⁴Laminar-airflow room is defined as a cleanroom in which filtered air entering the room makes a single pass through the work area in a parallel-flow pattern, with a minimum of turbulent flow areas. Laminar-flow rooms must have HEPA filter coverage of at least 80% of the ceiling (as vertical flow), or one wall (as horizontal flow), producing a uniform and parallel airflow (net filter medium face area versus gross area = 0.80).

and blower. It is normally installed in a "T-bar" grid ceiling system, suspended by seismic restraints from the architectural ceiling or building supports (Fig. 5B), which constitutes the inner clean space ceiling. Permanently installed or modular air handling systems are available that provide treatment (heating, air conditioning, dehumidification, etc.) of recirculated and make-up air to the clean space. Self-powered TFMs do not provide supply air treatment, and care should be taken in installing this type of unit because of high noise level and heat output factors, which may fatigue operative personnel or adversely affect temperature-sensitive stored drug products.

The TFM may also be used to provide a Class 1–100 work surface of almost any size, and may be installed in the ceiling (vertical downflow), or in a wall (horizontal flow). Using this design, shrouded laminar flow Class 1–100 "first air" is provided directly to the critical worksurface in the same quantity and quality as by a LFCB. Outflow from the critical worksurface is used to treat the general space, in combination with additional room "buffer" HEPA filtration, thus

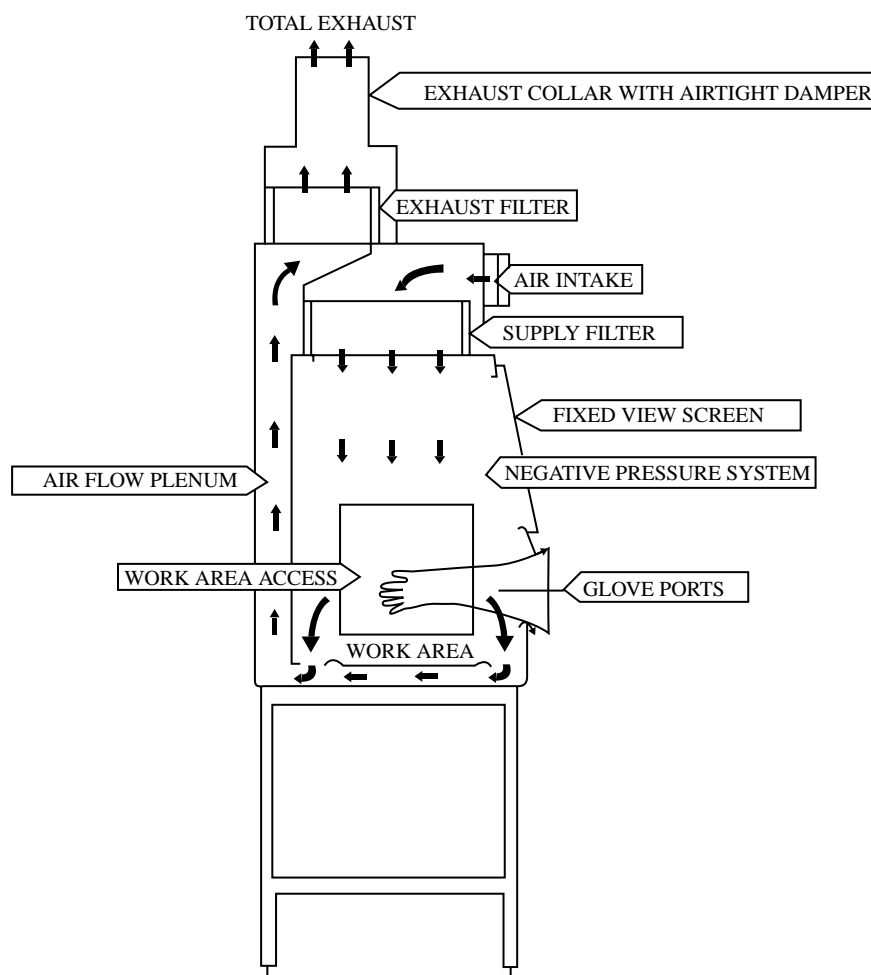


Fig. 9 Airflow patterns of the Class III laminar flow BSC. (Courtesy of the Baker Company, Sanford, Maine.)

providing both the highly-controlled critical work surface, and the less-controlled general room area. Several inexpensive concept designs are available to develop reliable work surfaces and buffer areas, which support both pharmacy and pharmaceutical manufacturing operations (20).

In-duct HEPA filtration

In this type of filtration the HEPA filter is placed within the supply duct system, providing Class 100 air from the point immediately downstream of the filter to the supply diffusers within the general clean space. This arrangement is used where a comparatively small volume of supply air is required for the clean space operation, when filter handling and disposal is critical (requiring remote “bag in/out” containment isolation or other special handling considerations), or where space

limitations prevent the installation of TFM. Induct HEPA filtration provides conventional airflow and is recommended as secondary or “buffer” filtration in clean spaces where critical work zones are treated using an LAFD. In-duct HEPA filtration is not recommended for LAF applications to provide required Class 1–100 conditions at critical work surfaces, where uninterrupted “first air” is necessary to achieve the highest levels of contamination control.

The plenum ceiling

The plenum ceiling is an arrangement of ductless terminal HEPA filter modules which, in a manner similar to TFM, are installed in a “T-bar” grid system. These filter modules are not connected directly to an air handling system by means of ducts. The uniform flow of air through all modules results rather from a

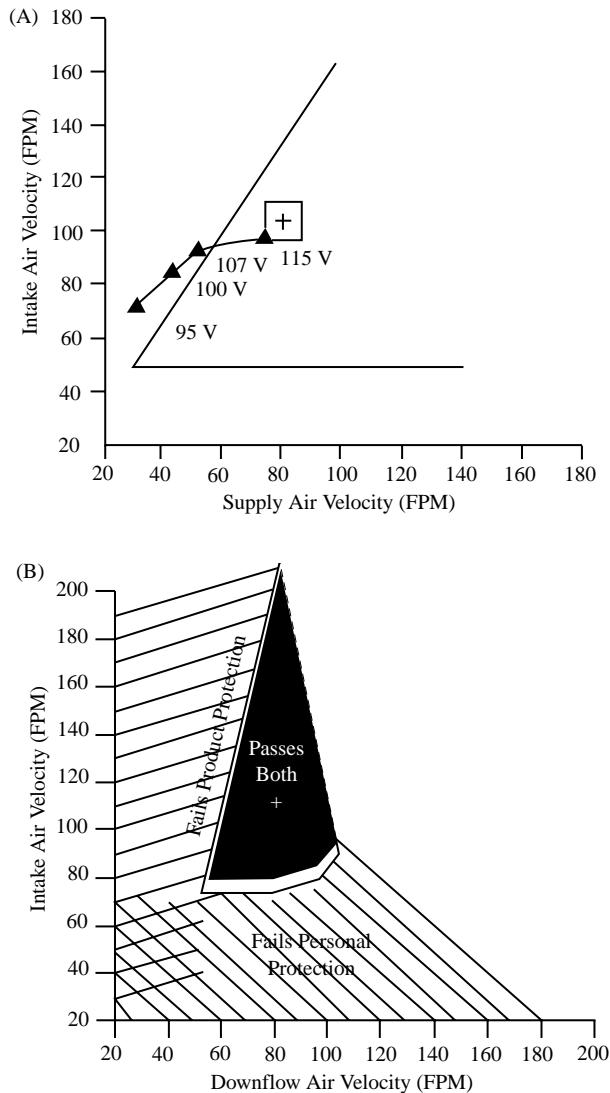


Fig. 10 (A) Effects of voltage variation on cabinet performance; National Sanitary Foundation (NSF) range, +: Normal setpoint; —: Performance envelope; ▲—▲ Airflow balance. (B) Performance envelopes for Class II BSCs determined by conducting a series of microbiological aerosol tests at a variety of airflow settings. (Courtesy of the Baker Company, Sanford, Maine.)

pressurized common space or “plenum” immediately above the entire T-bar ceiling assembly. This plenum is supplied directly by an air handler, eliminating the need for extensive ductwork. The plenum ceiling is normally used in overhead space-limited LAF applications.

The cleanroom

The cleanroom is a dedicated clean space with exacting as-built, at-rest, and operational specifications of

airborne and surface cleanliness, temperature, relative humidity, and lighting and noise levels, in which specific critical operations are carried out (11). The cleanroom may be designed to provide vertical or horizontal LAF throughout the entire room for large-scale operations requiring extended critical work zones (sterile conveyors, sterile fill operations, etc.), or may be designed as a conventionally-supplied, controlled secondary or “buffer” area, housing one or more smaller LAFD for aseptic processing steps requiring comparatively limited critical work zones.

The cleanroom is normally constructed in an existing building or structure as a core unit, surrounded by supporting access and staging anterooms, service chases, and machine rooms necessary to support the aseptic operation, supply working materials, and remove finished products and waste materials without cross-contamination or interference with the critical work stream.

TESTING AND CERTIFICATION OF LAF SYSTEMS

Thorough, periodic testing of LAF equipment is necessary to optimize performance and demonstrate compliance with established operating procedures and industry standards.

Certification

The term “certification” is widely used in connection with this type of testing. Certification may, however, only be provided by a registered testing or metrology laboratory, or other organization that derives its certification authority directly from a legitimate regulatory agency (FDA, EPA, AIHA, etc.) and which is subject to review. For a test procedure to qualify as a performance certification, the procedure must be performed and documented in accordance with the current good laboratory practices (21). All on-site testing of this nature constitutes a laboratory field certification of performance, with each test report accession-numbered in the laboratory test notebook as an actual laboratory test procedure. If the vendor of this service does not have such regulatory oversight, the buyer must qualify the vendors adequacy of experience, equipment, and efficacy of all test procedures to be performed (22), regardless of the vendor’s professional affiliations or memberships.

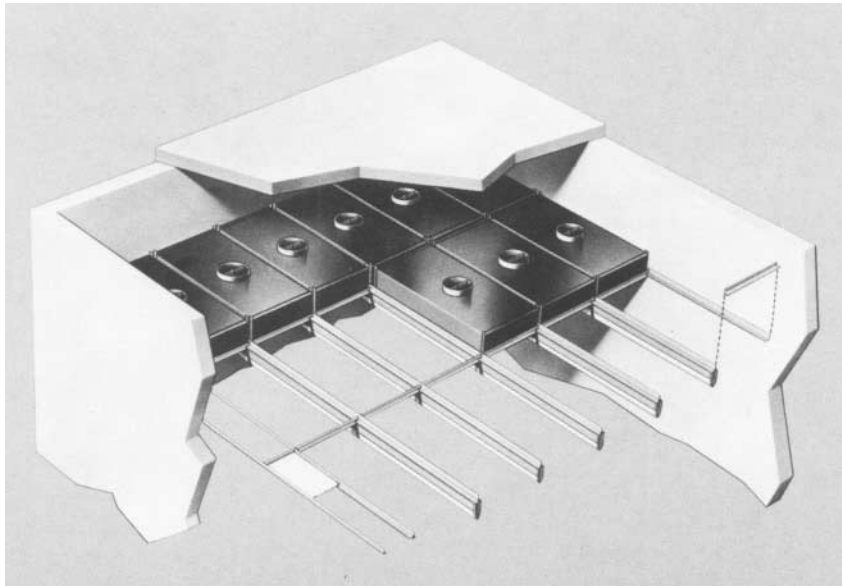


Fig. 11 Self-contained terminal HEPA filtration modules installed in a T-bar suspended plenum ceiling in a laminar-flow application. (Courtesy of American Air Filter, Inc., Louisville, Kentucky.)

Vendor Qualification

Although large manufacturers often have in-house personnel and facilities available to carry out LAF equipment testing, this task is frequently performed by an outside testing contractor, whose procedures and findings are relied upon without question by the buyer. Because the pharmaceutical manufacturer and pharmacy practitioner bear the ultimate responsibility to regulatory groups for ensuring compliance with all testing standards and requirements, it is essential that they select a qualified testing contractor. The buyer should interview the prospective contractor and review all aspects of the tester's training and experience, the test procedures and equipment to be used, equipment calibration and traceability, and the proposed documentation (22). The prospective contractor should be prepared to provide both company and individual resumes of qualifications, including at least three references of clients for whom the contractor has recently provided similar services (20).

Test Equipment Calibration and Traceability

The LAF equipment tester should have the equipment necessary to carry out the required challenge of the LAF system (20). This equipment should be in good working order and calibrated to a National Institute of

Standards and Testing (NIST)-traceable^r standard at least annually, or more often as indicated by the manufacturer's specifications or equipment performance (23). NIST-traceable calibration ensures the uniformity, accuracy, and serviceability of all test equipment as well as proper maintenance and care by the user. The NIST-traceability of equipment calibration is substantiated by a detailed calibration certification letter, issued by the calibrating authority for each piece of equipment.

Certification Standards

The primary standard for testing LAF equipment is a standard operating procedure (SOP), unique to the equipment or system. This SOP is normally established by the using organization to define the calibration and testing requirements of a specific LAF system, and should include the scope, intent, and frequency of testing; the equipment to be used in conducting the test; test equipment calibration; test methods and performance criteria; documentation of results; and corrective actions and acceptance. The SOP should cite the applicable references and industry standards from which it has been derived. In the absence of an organizational SOP, LAF equipment should be tested in accordance with industry standards applicable to the specific LAF

^rFormerly the National Bureau of Standards (NBS), Washington, DC.

application or area of operation (Table 3). In this case, regulatory groups will defer to these standards for proof of proper LAF system function and serviceability.

Test Reports

A comprehensive report should be issued upon completion of the test procedures, which should include all values measured for compliance, a listing of all test equipment, calculations, conversions, and all appropriate statistical justification along with comments pertaining to system function and operation. A dated test-completion or certification sticker should be affixed to the LAF unit referring any examining authority to the completed test report. All reports should include floor plans or maps of the clean space, which identify sample locations, probe heights, unusual performance characteristics, system adjustments, and repairs.

Test Procedures

Air velocity

Measurement of air velocity and laminar profile quality is the first step in LAF system testing. Using the appropriate instrument and measurement technique (4, 11, 18), the velocity of the laminar airstream for all LAFDs, and conventional buffer supply airflow volume for all controlled areas should be measured and adjusted to conform with pertinent SOPs or industry standards, consistent with the system manufacturers specifications and limitations (20).

HEPA filter performance

Following verification of the proper velocities, the HEPA filter should be aerosol-challenged in accordance with industry standards (4, 11, 18). Because this challenge is based upon the effectiveness of the filter in retaining aerosols, the upstream concentration of test aerosol should be verified before commencing this test and should not be assumed to be adequate, regardless of the circumstances.

Particle counting

Only through discrete particle counting can air cleanliness be verified, and the cleanliness class of the sampled environment established (4). Periodic in-process monitoring of workstations, buffer rooms, anterooms, production areas, and any other area about which a definitive air cleanliness statement is made or reasonably assumed, should be carried out in accordance with

SOPs or industry standards (4, 11). A discrete particle counter (DPC) with an adequate sampling rate, calibration features, and dynamic range should be used for sample acquisition, based on the specified air cleanliness level (4, 17).

Noise levels

Noise in excess of recommended levels within the LAF environment or clean space may cause operator discomfort and premature fatigue, and may indicate HEPA filter failure, or a malfunction of the LAFD or clean space air handling unit (20, 24). Cabinet pressure integrity and vibration analysis are recommended for all biological safety cabinets.

Visible and ultraviolet light levels

Visible light levels should be monitored for operator comfort and total visibility of the worksurface (20, 22). Ultraviolet germicidal light should be measured for effectiveness in all LAFDs having this feature (18).

LAMINAR FLOW CLEAN SPACE PROJECT DEVELOPMENT

As an engineering control, the demands normally placed upon an LAF system should not exceed the system's ability to provide a sterile, aerosol-free work area in which to conduct the desired aseptic operations with a high degree of confidence. Neither should a facility be overbuilt; unnecessary complication of clean space operation to achieve control levels that provide no demonstrable improvement in the process or finished product is counterproductive and an expensive waste of resources. To develop a system that is adequate for the aseptic processing task at minimum expense, a definitive, phased approach to design and implementation of the LAF system should be taken, regardless of the size or complexity of the system. The method described below ensures that any project, from a simple hood installation to a complex process cleanroom, will have the most comprehensive development and planning, thus maximizing the probability of a successful outcome.

Conceptualization

As the initial step, a thorough and complete conceptualization of the process, and the steps that will be taken to achieve the desired results should be defined in a protocol.

Table 3 Regulations and guidelines pertaining to LAF systems^{a,b}

	Guidelines										Regulations							
	ASHP (Proposed)	EN ISO 14644-1 TO 6 (Testing HEPA Filters)	IES-RP-GC-001-86 (Testing cleanrooms)	IES-RP-GC-006-84-T (Testing cleanrooms)	IES-RP-GC-018.2 (Cleanroom Housekeeping and monitoring)	NSF 49	USP 1074 (Proposed)	CGMP-CFR	FED. STD. 209c	CGMP-CFR	JCAHO	Regional Pharmacy Practice Acts	Regional Department of Public Health	OSHA 8.1.1	NRC, FEDERAL	NRC, STATE	Municipal Codes	
Pharmacy (IV/TPN)	•						•				•	•	•	•	•	•	•	
Pharmacy (CYTA)	•				•						•	•	•		•	•	•	
Home health care (IV/TPN)		•	•								•	•	•		•	•	•	
Home health care (CYTA)		•	•								•	•	•		•	•	•	
Nursing home (IV/TPN)											•	•	•		•	•	•	
Nuclear pharmacy	•						•				•	•	•		•	•	•	
Pharmaceutical manufacturing		•										•	•		•	•	•	
Nuclear pharmaceutical manufacturing		•	•				•				•	•	•		•	•	•	
Pharmaceuticals repackaging		•	•				•				•	•	•		•	•	•	

^a Courtesy Lab Safety Corp., Chicago.
^b Reactor products on site: nonagreement states.
TPN: Total parenteral nutrition; CYTA: Cytotoxic agent; CGMP: Current Good Manufacturing Practices; CGLP: Current Good Laboratory Practices;
CFR: Code of Federal Regulations; NRC: Nuclear Regulatory Commission.

This document is to be used as the basis of communication with design and mechanical engineers in developing specifications for an effective and workable design, to be constructed, validated, and operated within existing cost constraints.

Protocol and Project Management

The protocol should be developed under the supervision of a project manager, selected on the basis of experience and understanding of the product, existing and intended markets, regulatory issues, the intended aseptic process, the engineering controls normally required for such a process, and the design, construction, validation, operational, and maintenance methods to be employed (25). The manager represents the owners' interests, and acts as liaison in all dealings with outside vendors and contractors. He or she should seek direct input from key personnel and representatives of all groups having direct involvement with the planned clean space (26) in meetings and discussions, as the concept evolves, with appropriate sign-off by the participants on the finished protocol. This method prevents after-the-fact, unforeseen demands upon system operation, or dissatisfaction with system performance by essential personnel. Conceptualization is based upon the following considerations:

- Space constraints vs. process requirements.
- Product or process quality statements.
- Output capacity and growth expectations.
- Industry standards and regulatory issues (Table 3).
- Cleanliness class(es) required. Identification of critical work zones and areas requiring LAF, rather than mere characterization of these areas as Class 100.
- Work streaming. Inclusion of the necessary QA/QC steps in the overall process.
- Necessary process equipment (Table 4).
- Identification of all process steps requiring an LAF environment to facilitate the desired cleanliness class. What facility and equipment performance alert and action limits are necessary, and how are these to be monitored?
- Process equipment portability and flexibility requirements. Will the process(es) be expanded or modified at some time in the future? What equipment must be permanently installed and "hard wired" within the clean space?
- Process equipment service and maintenance requirements.
- Number of personnel required.

- Personnel disciplines. What type and arrangement of anterooms are required for scrubbing, gowning, and storage of barrier materials?
- Process materials staging and waste management.
- Health and safety requirements.
- Facility housekeeping and maintenance procedures.
- Facility validation, testing, and recertification.
- Documentation, i.e., the necessary process records, logs, labels, etc., and how, when, and where these documents are produced and used.

Regulatory Issues

Operation of LAF systems in the course of pharmaceutical manufacturing and pharmacy compounding operations is ultimately scrutinized by several regulatory and quasi-regulatory groups responsible for the control and oversight of these industries. Because these regulatory groups require, almost without exception, operation of manufacturing or compounding systems in accordance with "current industry standards," the distinction between regulations and guidelines has become unclear. Guidelines, as a reflection of the most current application of any technology, often take on the weight of regulations as determinants of industry standards. As such, operation in accordance with specific industry guidelines is often required to demonstrate compliance with nonspecific or generalized regulations—i.e., current good manufacturing practices (cGMPs) or Joint Commission on Accreditation of Health Care Organizations (JCAHO) Standards. Table 3, although not intended to be all inclusive, provides a list of regulations and guidelines that are pertinent to the operation of LAF systems in several different pharmaceutical manufacturing and pharmacy-compounding applications.

Design

Based upon the protocol, the facility and process design is the next step. Whether a simple hood installation to an existing space, an upgrade retrofit of an existing space, or construction of a totally new space, the design responsibility has to be assigned and a designer chosen. The designer's obligations are considerable and should be thoroughly understood by the individual or firm retained to provide the clean space design (20).

Construction

Following completion of the design, the building contractor is chosen. The construction phase is normally

Table 4 Equipment Features and Performance^{a,b,c}

	Cleanliness Class-Fed Std 209°C				Applications										Airflow and Equipment Configuration						
	Class 10	Class 100	Class 1,000	Class 100,000	Pharmacy IV corresponding	TPN (HA) compounding	CYT(A) compounding	Product protection	Personnel protection	Technology, screening	Sterility testing	Work station	Cleanroom	Laminar flow	Conventional flow	First air work station	Terminal filtration	In-duct filtration	Ventable to room	Direct connection	Thimble unit
LFCB	2	1			1	1	3	1		1	1	1	1	1	1	1		1			
LFBS Class I								1	1	3	1	1					1	1	1		
LFBS Class IIA	2	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LFBS Class IIB	1	1 ^d			2	2	1	1	1	1	1	1	1	1	2	2					
LFBS Class 3																					
glovebox	2	1			1	1	1	1	1	1	1	1	2	1	1	1	1	1	1		
In-duct HEPA filter		2	1	4	4	4	4			4	1	1	1	1	1	1	1	1			
Terminal HEPA filter module																					
(TFM)	2	2	1	4	4	4				4	2	1	2	1	2	1	1	1			
Filter plenum ceiling	2	1			2							1	1	1	2	1	2	1			

^a See also Appendix B.
^b 1: Primary engineering control; 2: With optimization or modification;
3: Obsolete application; 4: Secondary or “buffer” control.
^c Courtesy Lab Safety Corp., Chicago.
^d Early designs.

carried out with the input and assistance of the designer in conducting vendor audits for the selection of contractors with sufficient experience and facilities to complete the building tasks efficiently. Materials and workmanship audits should be carried out periodically by the project manager and the designer.

Validation and certification

Validation is described as proof that the system performs as stated. As an engineering control, the LAF system must demonstrably support the intended aseptic or controlled process. Validation of the aseptic manufacturing process and the LAF systems that support terminal sterilization in pharmaceutical manufacturing applications should be carried out in accordance with industry standards (1, 27–31). Such validation should be accomplished in three phases, consisting of installation qualification (IQ), operational qualification (OQ), and process qualification (PQ), with full and detailed documentation of all activities and outcomes (20).

Cleanroom Construction and Validation: A Low-Cost Approach

The following is a consolidation of steps leading to the construction and operation of a low-cost clean space of the type required for small- and intermediate-size drug repackagers, home health care pharmacies, and small pharmaceutical manufacturers.

Cleanroom design

The cleanroom should be conceptualized, designed, constructed, validated, operated, and maintained in a manner that supports the aseptic process. A construction plan, including reasonable time frames for the acquisition of labor and materials should be developed, followed, and updated. All work and materials should comply with local building codes and safety ordinances.

HEPA Filtration

Computation of the air change rate (AC) necessary for the desired cleanliness class is the initial step in determining the ceiling module filter density required for the clean space, and should be calculated accurately, with sufficient redundancy to assure the required airflow and resulting cleanliness levels (20, 32,33).

Air Handling

Air handling requirements should be calculated, and air handler type and capacity should be matched to the AC rate, heating, cooling, and dehumidification requirements of the facility (34), as calculated from the reflected ceiling plan. At almost any intended level of processing, low cost, “turn-key” air handling systems are available with a minimum of lead time. They are easily installed, operated, and maintained. The installation of a modular air handler of this type provides conditioned air, allowing operation of the core and anteroom areas as an independent facility, without dependence upon a central building air handling apparatus for process control. This type of low-cost, modular air handler, in a single-or multiple-unit installation, facilitates complete treatment and conditioning of the supply air to be provided in quantities adequate to maintain the most stringent air cleanliness classes for workstations and general room air, as well as temperature and relative humidity at recommended core and anteroom internal operating pressures. In addition, several manufacturers of this type of unit have experienced, on-staff mechanical engineers qualified to assist the buyer in determining the exact air treatment and handling requirements for a specific application at no additional charge. Qualification of the knowledge and experience of all individuals involved in the conceptualization and design phases is necessary, and references should be obtained and checked by the buyer.

Differential Pressurization

The cleanroom facility should be carefully designed to control the ingress of contaminants, and be positively pressurized to the surrounding area in accordance with industry standards. The core and anterooms are positively pressurized by varying the amount of incoming “make-up” air (20). In the case of “soft wall” clean space facilities, such pressurization (potential outflow) is not possible, and a sufficient amount of constant, active outflow (kinetic outflow) should occur to prevent ingress of contaminants.

Lighting and Electrical

Electrical service should be provided in the normal fashion in accordance with local building and electrical codes. Clean space service requires no special treatment, and should provide a 25–50 A surplus over worst-case processing demands, to allow for additional electrical

equipment and source power fluctuations. Fixture types, sealable outlets, and hard-wiring of processing equipment should be selected and carried out in accordance with industry standards and electrical codes (20).

Walls and Windows

Walls should be typical 5/8-in. drywall over metal studs at 16-in. centers, and may be insulated or not as deemed appropriate. The inner clean space wall should be finished with enameled panels (available at most building supply outlets) that are butted and sealed with RTV silicone, to provide gas-tight seams between the wall panel, ceiling T-bar, and floor base junctions. (Clear RTV silicone is recommended, neatly and sparingly “mopped in,” with all residues cleared away while still wet; this method creates an invisible, water-tight seal facilitating long term serviceability and ease of cleaning.) The outer walls should be painted with cleanable epoxy enamel or suitably covered, and a long, narrow observation window (recommended aspect 1:8–10, without ledge), or a series of double-pane, ledgeless windows should be installed and finished to provide complete observation of the operational process by prospective clients, supervisory personnel, and others.

Floor

The anteroom and core flooring should be an attractive, highly durable, one-piece vinyl or other suitable floor material, providing a minimum of seams. This flooring should ideally radius in a cove at the wall base and continue upward to a height of 8 inches. to 1 ft above the floor. At that point it should be capped with a suitable beading, sealed by RTV silicone, neatly “mopped in,” and allowed to dry. No drain or other floor opening should be permitted in the core; a small, single drain may be permitted in the anteroom in the event carts or other materials are to be cleaned and staged in that area. The floor should be routinely cleaned by a wet-vac recovery of a spread floor cleaning solution (with hose and head located within the clean space, with pass-through connection to a vacuum source located and vented outside the clean space).

Ceiling

The anteroom and core ceilings should be of typical 1.5 in. T-Bar construction, with seismic restraints adequate

to support the HEPA terminals, lighting fixtures, and ceiling panels. The panels may be obtained from clean space suppliers in the form of finished “cleanroom” ceiling panels, or by reduction of additional enameled wall panels to 2 × 4 ft (nominal) panels, affixed by permanent adhesive to standard, plain face (nontextured), fire-rated, 0.5 inches. commercial ceiling panels, which are easily installed and RTV silicone-sealed in place. This allows installation of a permanent, gas-tight ceiling, washable like the walls and floor, and attractive in that it provides visual continuity of the wall material.

Doors and Pass-Throughs

Doors and pass-throughs should facilitate easy, integral entry and exit of personnel and working materials (20).

Construction

Construction of the facility should be based upon the completed design and incorporate normal construction methods, tools, and techniques. Special preparation of components is necessary, including a general cleaning and protection of components and equipment from potential sources of contamination during the construction phase. The working materials should be protected from atmospheric dust, sawdust, grease, aerosols of oil, and other residues, which may be encountered during construction, using “clean construction” methods (20).

Validation

Following completion of all testing and certification of the facility, documentation should be developed and retained, which may be used as proof of performance in accordance with the protocol in the as-built configuration (4, 11). Periodic retesting and monitoring at specified intervals in the at-rest and operational modes (4, 11) should be carried out to maintain operation of the facility in accordance with SOPs or industry standards.

Operation

Operation of the facility in accordance with validation conditions, through the use of appropriate SOPs and industry standard quality management parameters (28) should be commenced following validation, incorporating the necessary alert and action limits, monitoring

procedures and maintenance steps to be carried out and included in the production cycle documentation. All microbiologic monitoring of the clean space ante-rooms, core, worksurfaces and personnel barriers should be carried out in a systematic manner by experienced personnel (35).

Personnel Selection and Validation

Because the effectiveness of operative personnel is potentially the greatest variable in any controlled process, a selection, training, examination, and grading system should assure the suitability of candidates, the adequacy of training, and the ultimate uniformity and consistency of clean space operating procedures (12, 14, 15, 36).

Maintenance

Clean space maintenance SOPs and documentation should be developed and followed in strict accordance with industry standards (13) to ensure the consistency of operation in accordance with validation conditions.

REFERENCES

1. Friebe, W.R. Validation of Aseptic Processing Operations. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1988; 1, 355–359.
2. Kleinberg, M.; Shatsky, F.; Lumkin, B. Particulate Matter in In-Line Burettes. *Particulate and Microb. Control* **1983**, 2 (5), 77–150.
3. Wilson, J. Infection Control in Intravenous Therapy. *Heart Lung* **1976**, 5 (3), 430–436.
4. Commissioner, Federal Supply Service, General Services Admin. *Clean Room and Workstation Requirements, Controlled Environment*; U.S. Government Printing Office: Washington, 1976.
5. Falck, K.; Grohn, P.; Sorsa, M. Mutagenicity in Urine of Nurses Handling Cytostatic Drugs. *Lancet* **1979**, 1, 1250–1251.
6. Connor, T.H.; Anderson, R.W.; Sessink, P.J.; Broadfield, L.; Power, L.A. Surface Contamination with Antineoplastic Agents in Six Cancer Treatment Centers in Canada and the United States. *AJHP* **1999**, 56 (14), 1427–1432.
7. *Laminar Flow Biological Safety Cabinets, A Training Manual for Biomedical Investigators*; National Cancer Institute: Washington, 1972.
8. Hinds, W.C.; Macher, J.M.; First, M.W. Size Distribution of Aerosols Produced by the Laskin Aerosol Generator Using Substitute Materials for DOP. *J. Am. Ind. Hyg. Assoc.* **1983**, 44 (7), 495–500.
9. *Recommended Practice for Laminar Flow Clean Air Devices*; Institute of Environmental Sciences: Mt. Prospect, IL, 1986.
10. *Recommended Practice for HEPA Filters*; Institute of Environmental Sciences: Mt. Prospect IL, 1989.
11. *Recommended Practice for Testing Cleanrooms*; Institute of Environmental Sciences: Mt. Prospect IL, 1984.
12. Dirks, I.; Smith, F.M.; Furtado, D. Methods for Testing Aseptic Technique of Pharmacy Personnel. *Am. J. Hosp. Pharm.* **1982**, 39, 457–459.
13. *Cleanroom Housekeeping-Operating and Monitoring Procedures*; Institute of Environmental Sciences: Mt. Prospect IL, 1989.
14. McKeon, M.R.; Peters, G.F. *VALITEQ Aseptic Technique Validation System Compounding Manual*; Boylan, J.C., Ed.; Lab Safety Corp. Des Plaines IL, 1999; 35–37.
15. OSHA. *Work Practice Guidelines for Personnel Dealing with Cytotoxic [Antineoplastic] Drugs*; U.S. Dept. of Labor: Washington, 1986.
16. American Society of Hospital Pharmacists. ASHP Technical Assistance Bulletin on Handling Cytotoxic and Hazardous Drugs. *Am. J. Hosp. Pharm.* **1990**, 47, 1033–1049.
17. Galatowitsch, S. Technique May Be Culprit Behind Class II BSC Contamination. *Cleanrooms* **1999**, 13(11), 1, 4–45.
18. National Sanitation Foundation. *NSF Standard 49*; Advisory Committee for Biohazard Cabinetry: Ann Arbor, MI, 1995.
19. Farquharson, G.J. Isolators for Pharmaceutical Applications. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1999; 18, 121–136.
20. Peters, G.F. Laminar Airflow Equipment: Engineering Control of Aseptic Processing. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1993; 8, 317–359.
21. Food and Drug Administration. *Current Good Laboratory Practices*; Washington, 1992.
22. Bryan, D.; Marback, R. Laminar-Airflow Equipment Certification: What the Pharmacist Needs to Know. *Am. J. Hosp. Pharm.* **1984**, 41, 1343–1349.
23. *Recommended Practices for Equipment Calibration or Validation Procedures*; Institute of Environmental Sciences: Mt. Prospect IL, 1986.
24. OSHA. *Standard 1910.95*; U.S. Dept. of Labor: Washington, 1971.
25. Facility Design Considering Safety, Monitoring and Detection: A Team Approach (Verbal Communication) Microcontamination Conference, Santa Clara, CA, 1988.
26. Whyte, W. *Cleanroom Design*; Wiley: New York, 1991.
27. Food and Drug Administration. *Current Good Manufacturing Practice for Finished Pharmaceuticals*; Washington, 1992.

28. Kozicki, M.; Hognie, S.; Robinson, P. *Cleanrooms*; Van Nostrand, Reinhold, Eds.; New York, 1991.
29. Parenteral Drug Association. *Validation of Aseptic Filling for Solution Drug Products*; Philadelphia, 1980.
30. American Society of Hospital Pharmacists. Technical Assistance Bulletin on Quality Assurance for Pharmacy-Prepared Sterile Products. *Am. J. Hosp. Pharm.* **1993**, *50*, 2386–2398.
31. *Sterile Drug Products for Home Use*; The United States Pharmacopoeial Convention, Inc.: 1998.
32. Dixon, A. *Cleanroom Management Manual*; Cleanroom Management Association: Tempe, AZ, 1991.
33. *Guideline on Sterile Drug Products Produced by Aseptic Processing*; Food and Drug Administration: Washington, 1987; 20–27.
34. High Performance-Low Energy-Cost Cleanroom: A Case Study Microcontamination Conference, Santa Clara, CA 1990.
35. Peters, G.F.; McKeon, M.R. Microbiologic Monitoring of Aseptic and Controlled Processes. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 2000; 19, 239–278.
36. Dixon, A. Training Cleanroom Personnel. *J. Parent. Sci. Technol.* **1991**, *45* (6), 276–278.

ISOMERISM

Thomas N. Riley

Jack DeRuiter

William R. Ravis

C. Randall Clark

Auburn University, Auburn, Alabama, U.S.A.

BASIC PRINCIPLES

Isomers are defined as molecules of identical atomic compositions (molecular formulas), but with different bonding arrangements of atoms or orientation of their atoms in space. Based on this definition, several types of isomerism are possible including constitutional, configurational, and conformational isomerism. Constitutional isomers (also called structural or positional isomers) are molecules with the same atomic composition but different bonding arrangements between atoms, as illustrated by the examples of catechol, resorcinol, and hydroquinone (Fig. 1). All of these compounds have the same atomic composition ($C_6H_6O_2$), but different bonding arrangements of atoms and are thus distinct chemical entities with different chemical and physical properties.

Configurational isomers are defined as molecules of identical atomic composition and bonding arrangements of atoms, but different orientations of atoms in space, and these different orientations cannot interconvert freely by bond rotation. Since these types of isomers differ only in relative spatial orientations of atoms, they are commonly referred to as stereoisomers. Configurational stereoisomers are subcategorized as optical isomers (enantiomers) or geometric isomers (Fig. 2), depending upon the hybridization state and geometry of the atoms that impart the properties of stereoisomerism and the overall structure of the molecule. Stereoisomers of this type are distinct chemical entities that may have different chemical and physical properties.

Conformational isomers (conformers) are stereoisomeric forms characterized by different relative spatial arrangements of atoms that result from rotation about sigma bonds. Thus, unlike configurational isomers, conformers are interconverting stereochemical forms of a single compound. The nature of conformational and configurational stereoisomerism, as well as the role of stereoisomerism in drug activity is the subject of this article.

Optical Activity and Molecular Structure

Modern stereochemistry originated with the research of Malus in 1808 who discovered that plane-polarized light is generated when a beam of light is passed through calcium carbonate. In 1813, the mineralogist Biot reported that asymmetrically cut quartz crystals rotate the plane of a beam of polarized light. It also was noted that certain organic liquids, as well as solutions of certain organic compounds, can rotate the plane of polarized light. Biot attributed this effect on plane-polarized light to a property of the individual organic molecules through which the light is passed, a property now referred to as optical activity. The concept of optical activity was extended by Herschel in 1812, who observed that hemihedral quartz crystals, having odd faces inclined in one direction, rotated the plane of polarized light in one direction, whereas crystals whose odd faces were inclined in the opposite direction rotated plane-polarized light to the same extent but in the opposite direction.

Pasteur refined the observations of the mineralogists by proposing a link between optical activity and molecular structure. His landmark work of 1847 was based on earlier observations by Biot that chemically identical salts of tartaric acid rotated plane-polarized light differently. Pasteur discovered that two distinct crystalline forms of tartaric acid salt could be obtained from solutions of the optically inactive salt of "paratartaric acid" (also known as racemic acid), and that one crystal form has hemihedral faces that inclined to the right, whereas the other has faces that inclined to the left. He separated the distinct crystalline salts forms and observed that they, unlike paratartaric acid, are optically active; solutions of the left-handed crystals rotate the plane of polarized light to the right, and solutions of the right-handed crystals rotate the light to the same degree, but in the opposite direction. Pasteur further demonstrated that the left- and right-handed crystals were mirror images of each other and concluded that this property must reflect the handedness of the molecules that constitute the crystals.

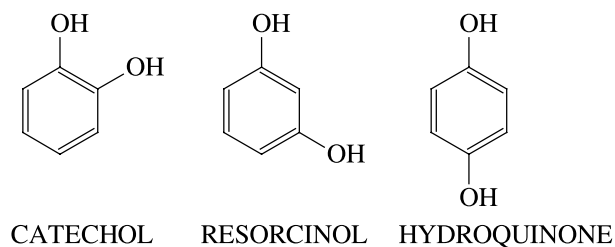


Fig. 1 Constitutional isomers.

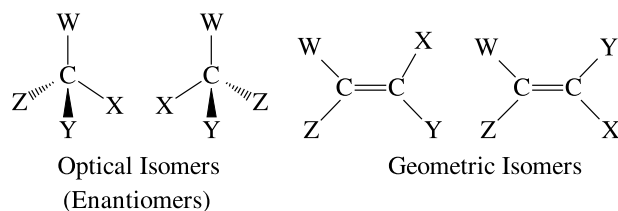


Fig. 2 Stereoisomers.

The molecular basis for the left- and right-handedness of distinct crystals of the same chemical substance and the associated differences in optical rotation was developed from the hypothesis of Paterno (1869) and Kekulé that the geometry about a carbon atom bound to four ligands is tetrahedral. Based on the concept of tetrahedral geometry, Van't Hoff and LeBel concluded that when four different groups or atoms are bound to a carbon atom, two distinct tetrahedral molecular forms are possible, and these bear a nonsuperimposable mirror-image relationship to one another (Fig. 3). This hypothesis provided the link between three-dimensional molecular structure and optical activity, and as such represents the foundation of stereoisomerism and stereochemistry.

Chirality and Optical Isomers (Enantiomers)

The property of nonsuperimposability became known as chirality, and molecules containing asymmetrically substituted carbons are referred to as chiral molecules. The term chiral was derived from the Greek word meaning "hand" and was applied as a description of the left- and right-handedness of crystal structure resulting from molecular asymmetry. The individual mirror image forms

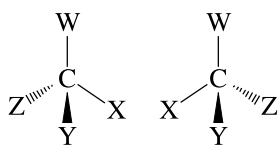


Fig. 3 Tetrahedral geometry and optical isomerism.

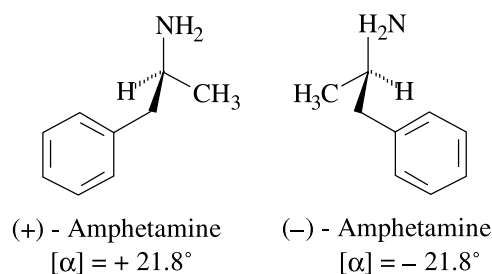


Fig. 4 Amphetamine enantiomers.

of a chiral molecule are called optical isomers because they rotate the plane of polarized light (are optically active) and differ in structure only in the orientation of atoms or groups about the asymmetric carbon (are isomers). Today, optical isomers are more commonly referred to as enantiomers or an enantiomeric pair.

Generally, optical isomers or enantiomers have identical physical and chemical properties, for example, the enantiomeric forms of amphetamine (Fig. 4) have identical melting points, pK_a , solubilities, etc. There are, however, two important differences in properties between the members of an enantiomeric pair. First, each member rotates the plane of polarized light to the same degree, but in opposite directions. The enantiomer rotating the plane to right (clockwise) is designated as the dextrorotatory (d) or (+)-enantiomer. The other enantiomer rotates the plane to the left (counterclockwise) and is designated as the levorotatory (l) or (–)-enantiomer. This is illustrated in Fig. 4 for the enantiomers of amphetamine, where the enantiomer with the specific optical rotation of (+)-21.8° is designated as dextrorotatory, whereas the mirror enantiomer with a specific rotation of (–)-21.8° is called levorotatory. A second difference between enantiomers is their interactions with other chiral substances. For example, enantiomers may have different solubilities in chiral solvents, they may react at different rates in the presence of an optically active reagent or enzyme, and many have different affinities for chiral surfaces and receptors.

Most optically active drugs are chiral as a result of the presence of an asymmetrically substituted tetrahedral carbon atom. However, chirality can result from the

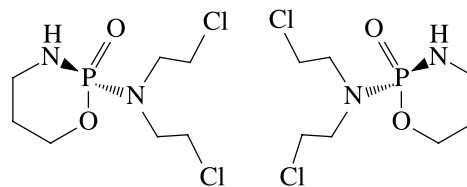


Fig. 5 Optical isomers of cyclophosphamide.

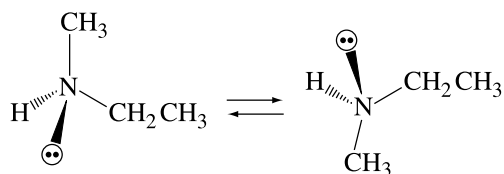


Fig. 6 Pyramidal inversion of chiral nitrogen.

presence of other asymmetrically substituted atoms within molecules as illustrated below including phosphorous (Fig. 5), nitrogen (Fig. 6), and sulfur (Fig. 7).

Definitions and Nomenclature

As discussed in the preceding section, each member of an enantiomeric pair rotates the plane of polarized light to the same degree, but in opposite directions (dextrorotatory and levorotatory). However, the amount of optical rotation is not constant for an individual enantiomer but rather is dependent on the solvent, concentration, temperature, the wavelength of light used, and the path length of the sample cell employed to determine the rotation. Thus, meaningful optical rotation comparisons for chiral compounds are only possible when optical activities are determined under specified conditions. Such conditions are defined as specific rotations $[\alpha]$ and are expressed for solutions and neat liquids in Eq. 1 and 2, respectively.

$$[\alpha]_{\lambda}^t = \frac{100\alpha}{l \cdot c} \quad (1)$$

$$[\alpha]_{\lambda}^t = \frac{\alpha}{l \cdot d} \quad (2)$$

where α is the measured rotation; t , the temperature; λ the wavelength; c , the concentration; d , the density; and l , the length.

Specific rotation data may assist in the identification of a specific enantiomer, or may be used to determine the optical purity (enantiomeric purity) of a mixture of enantiomers. Optical purity is defined as the percent excess of one enantiomer over another in a mixture and is expressed in Eq. 3:

$$\text{Optical purity} = \frac{[d] - [l]}{[d] + [l]} \cdot \frac{\alpha_{\text{obs}}}{\alpha_o} \quad (3)$$

Based on Eq. 3, a mixture consisting of equal amounts of each enantiomer would have no net optical rotation; the optical rotation of one enantiomer is cancelled by the rotation of the other enantiomer. Such a mixture is referred to as a racemic mixture or racemate.

Other terms commonly applied in discussions of optically active compounds include resolution and

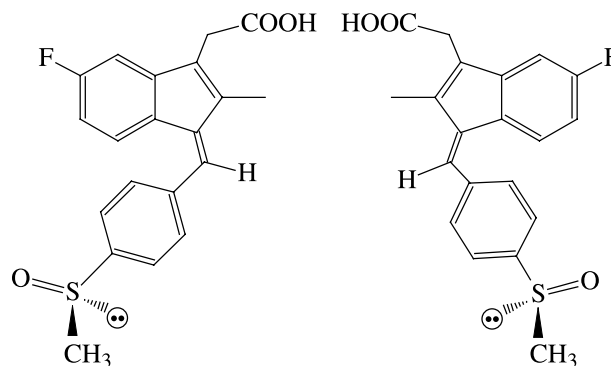


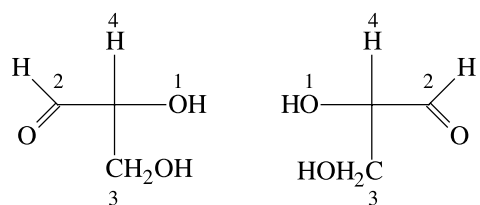
Fig. 7 The chiral sulfoxide sulindac.

racemization. Resolution describes the processes whereby a racemic mixture is separated (resolved) into its component enantiomers. Racemization refers to the conversion of either enantiomer into equal parts (racemic mixture) of both enantiomers.

Over the years, several nomenclature systems have been developed to characterize the relationship between enantiomers. The system based on optical activity and the classification of enantiomers as dextrorotatory [d or $(+)$] or levorotatory [l or $(-)$] already has been described. However, this system of nomenclature is of limited applicability because the sign of rotation, $(+)$ or $(-)$, does not predict the absolute configuration or the relative spatial arrangement of atoms in the enantiomers. In an attempt to designate the precise configurations about carbon centers of asymmetry, the Cahn–Ingold–Prelog R/S system have been developed and adopted as the most commonly used nomenclature system for isomers.

In applying the Cahn, Ingold, and Prelog R/S system the compound is oriented in a Fischer projection and the four groups or atoms bound to an asymmetric carbon are ranked by the following set of sequence rules (Fig. 8):

1. Substituents are ranked (1–4) by the atomic number of the atom directly joined to the chiral carbon.
2. When two or more of the atoms connected to the chiral carbon are the same, the atomic number of the next adjacent atom determines the priority. If two or more atoms connected to the second atom are the same, the third atom determines the priority, etc.
3. All atoms except hydrogen are formally given a valence of 4. When the actual valence is less than 4 (N, O), phantom atoms are assigned an atomic number of zero and therefore rank the lowest.
4. A tritium atom has a higher priority than deuterium, which has a higher priority than hydrogen. Similarly, any higher isotope has a higher priority than any lower one.



S-Glyceraldehyde *R*-Glyceraldehyde

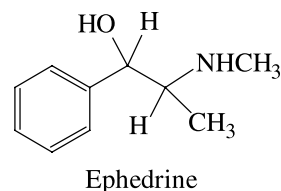
Fig. 8 Cahn–Ingold–Prelog sequence rule.

5. Atoms with double and triple bonds are counted as if they were connected by two or three single bonds. Hence a CC is regarded as a carbon bound to two carbons; and a CO is regarded as a carbon bound to two oxygens.

Once the four groups bound to the chiral carbon are ranked, the compound is oriented in such a way that the lowest priority group (4) is projected away from the observer. Then, if the other groups (1–3) are oriented by priority in a clockwise fashion, the molecule is designated as *R* (rectus), and if counterclockwise, as *S* (sinister). These sequence rules are applied in the assignment of the absolute configurations for the enantiomers of glyceraldehyde in Fig. 8. According to the first rule, the highest priority substituent (1) is the hydroxy group (OH) and the lowest priority group (4) is the hydrogen atom; since the atomic number of carbon is higher than that of hydrogen but lower than that of oxygen, the two carbon substituents (CHO and CH₂OH) are assigned intermediate priorities. To determine the priority relative to the two carbon substituents, both the 2nd and 5th rules must be applied. The aldehyde carbon is part of a carbonyl (C=O) moiety which, by rule 5, is equivalent to a carbon bound to two oxygen atoms. The alcohol carbon (CH₂OH) is bound to one oxygen and two hydrogens. The “two oxygens” of the aldehyde take priority over the single oxygen of the alcohol moiety; the aldehyde is assigned priority 2, and the alcohol priority 3. With all substituents ranked and the enantiomers oriented in such a way that the lowest priority group (4) is projected away from the observer, the configurations can be assigned. The enantiomer in which the substituents are oriented by priority in a clockwise fashion, is designated as *R*, and the enantiomer in which the substituents are oriented by priority in a counterclockwise direction is designated as *S*.

Compounds with Multiple Centers of Asymmetry

Many stereoisomeric drugs contain more than one asymmetrically substituted atom. For example, ephedrine has two chiral centers (Fig. 9). In this case, a greater



Ephedrine

Fig. 9 Ephedrine having multiple chiral centers.

number of configurational isomers is possible; the maximum number possible is 2^n , where n is the number of chiral atoms. Hence, the maximum number of possible configurational isomers for ephedrine is four, which are designated *RR*, *SS*, *RS*, and *SR* (Fig. 10). The *RS* and *SR* isomers are nonsuperimposable mirror images and hence are enantiomers. The same relationship exists for the *RR* and *SS* isomers. The relationship between each member of an enantiomeric pair and each member of the other enantiomeric pair is diastereomeric; they are nonsuperimposable nonmirror images. Thus, the *RS* isomer is a diastereomer of the *RR* and *SS* isomers, and the *SS* isomer is a diastereomer of the *RS* and *SR* isomers. In this case, the *SR* and *RS* enantiomers are referred to as ephedrines, whereas the *RR* and *SS* enantiomers are called pseudo-

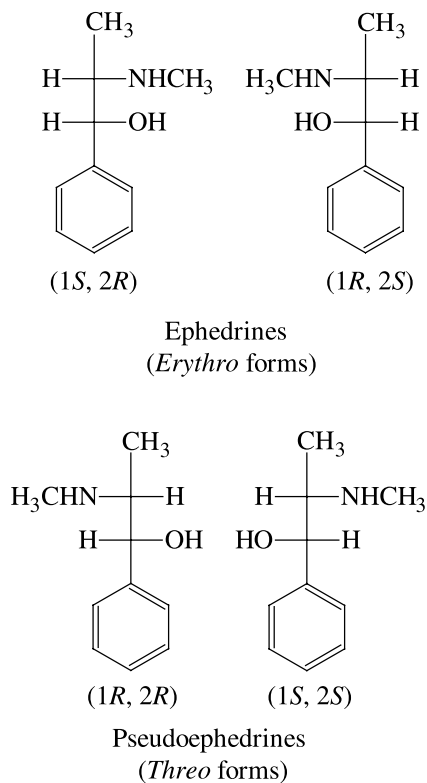


Fig. 10 Ephedrine and pseudoephedrine stereoisomers.

ephedrine. Diastereomers, unlike enantiomers, differ in their physicochemical properties including solubilities, acid-base strengths, melting points, etc.

In addition to the *R* and *S* designations, compounds with two chiral centers may also be identified by stereochemical nomenclature that describes the entire system. For example, the *erythro* and *threo* nomenclature derived from carbohydrate chemistry may be employed to describe the relative positions of similar groups on each chiral carbon. Thus, the ephedrine isomers are designated as *erythro* forms since the similar groups (OH and NHCH₃) are on the same side of the vertical axis of the Fischer projection, and the pseudoephedrine isomers are designated as *threo* forms since like groups are on opposite sites of the vertical axis of the projection (Fig. 10).

It is important to note that the 2ⁿ rule predicts only the maximum number of stereoisomers possible in compounds with more than one center of chirality. For example, some compounds with two asymmetrically substituted carbon atoms may have only three stereoisomeric forms. This

occurs when three of the substituents on one asymmetric carbon are the same as those on the other asymmetric carbon, as shown for the antitubercular ethambutol (Fig. 11). In this case, one stereoisomer has a plane of symmetry even though two asymmetric atoms are present. Such an isomer is referred to as a *meso* compound and is optically inactive. Therefore, when a plane of symmetry is present in a compound with two centers of asymmetry, only three stereoisomeric forms are possible.

Geometric Isomerism

Geometric isomerism was first defined by Wislicenus in 1887 as isomerism occurring in compounds where rotation is restricted by double bonds or ring systems. Geometric isomers do not rotate the plane of polarized light (unless they also contain a chiral center), and hence are not optically active.

Geometric isomerism resulting from restricted rotation about double bonds

The sp² hybridized carbon atoms of alkenes (olefins) and the atoms or groups attached to these carbons all lie in the same plane, and rotation around the double bond is restricted. As a result, stereoisomerism is possible when each carbon atom of the double bond is asymmetrically substituted. Because geometric isomers are nonsuperimposable, nonmirror images they are classified as diastereomers and therefore possess different physical and chemical properties. A number of drugs contain dissymmetrically substituted carbon-carbon double bonds and therefore can exist as two distinct geometric isomers (Fig. 12).

Several systems of nomenclature have been developed to designate the configuration of geometric isomers. Historically, the *cis-trans* system of nomenclature has been applied most frequently. It was developed to assign the configuration of geometric isomers when each isomer contains a like group or atom on each carbon atom of the double bond. However, in more complex structures this

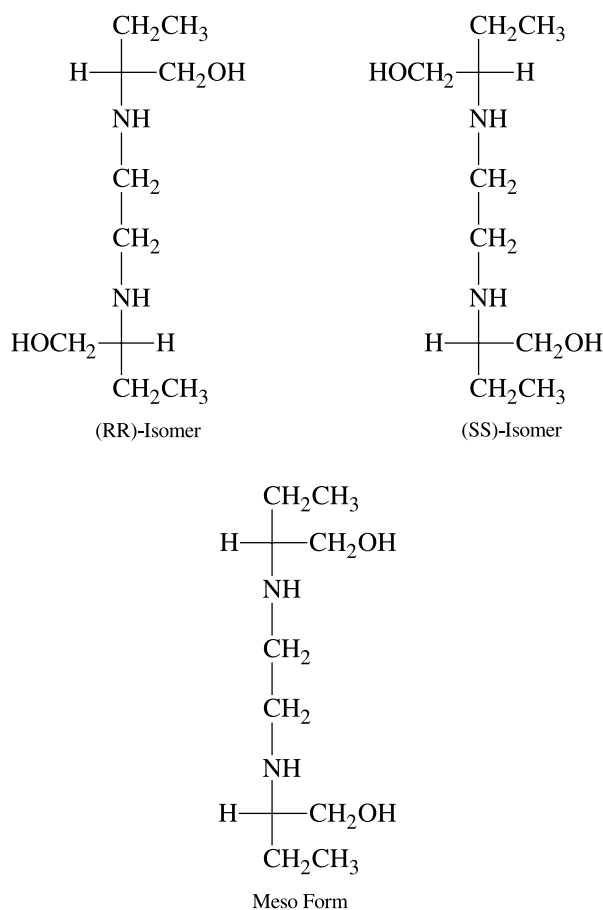


Fig. 11 Ethambutol stereoisomers.

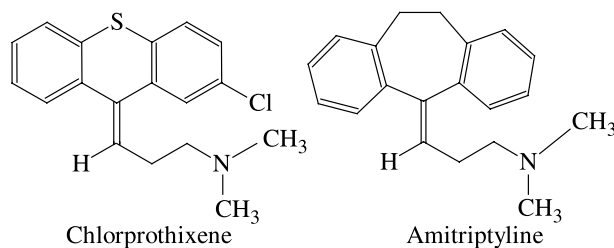


Fig. 12 Pharmacologically active dissymmetrically substituted alkenes.

system cannot be used unambiguously to assign configurations for all geometric isomers. For example, the sp^2 atoms of the antipsychotic agent thiothixene do not contain a like atom or group (Fig. 13). Thus, the *E/Z* notation system was developed to unambiguously assign configurations in all cases of geometric isomerism.

1. Priorities are assigned on the basis of atomic number for the two atoms or groups attached to each carbon of the carbon–carbon double bond. The same priorities apply as in the Cahn–Ingold–Prelog sequence rules.
2. Configuration is assigned based on the relative positions of the highest priority atoms or groups on each carbon of the carbon–carbon double bond. If these groups are on the same side, the *Z* (zusammen) designation is used. If they are on the opposite side, the *E* (entgegen) designation is assigned.

Geometric isomerism is also possible in double-bonded carbon-heteroatom systems such as imines and oximes (CN), and in azo (NN) systems. For example, several cephalosporin derivatives contain an alkoxyimino side chain that may exist in *E* or *Z* isomeric forms; the azo compound prontosil may display similar isomerism (Fig. 14). A *syn-anti* system can also be applied to CN geometric isomers. In these cases, the priority assignments are made as described previously, and the isomer with highest priorities on the same face of the double bond are called *syn*, whereas those with these groups on opposite sides are named *anti*.

In a number of amides, thioamides, and related systems, rotation about the single bond is hindered, and distinct geometric isomers can be observed and even isolated. This type of geometric isomerism is referred to as atropisomerism and results from resonance contributions by the nitrogen atom that imparts significant double bond character to the system, thus slowing rotation. Such is the case for the thioamide aldose reductase inhibitor

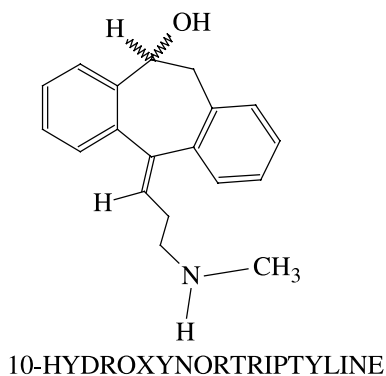


Fig. 13 Geometric isomer nomenclature—thiothixene.

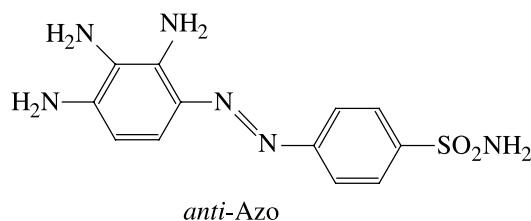
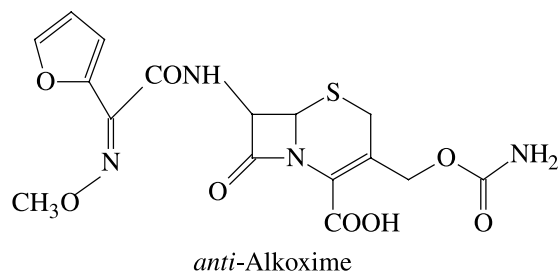
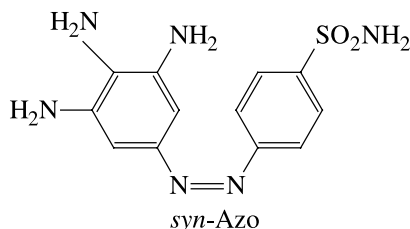
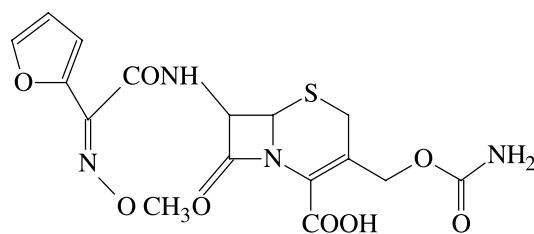


Fig. 14 Geometric isomerism about C-heteroatom double bonds.

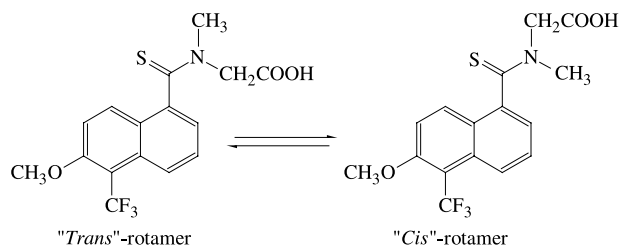


Fig. 15 Atropisomerism, tolrestat.

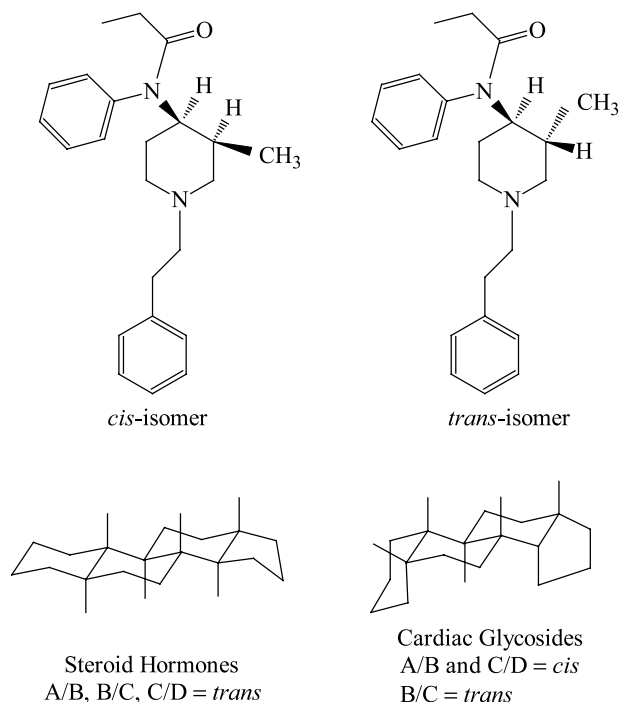


Fig. 16 Geometric isomerism in cyclic drug structures.

tolrestat shown in Fig. 15. Isomers of this type are also called rotamers and are considered to be conformational isomers since they result from rotation about a single bond.

Geometric isomerism in cyclic compounds

The presence of a ring system, either cycloaliphatic or heteroaliphatic, like that of a double bond, prevents rotation, therefore giving rise to geometric isomers when two carbon atoms of the ring are each substituted by different groups. Structures of pharmacologically-important structures that display geometric isomerism are shown in Fig. 16.

When rings are fused through adjacent atoms, the fusion may be *cis* or *trans*, as illustrated by the steroid structure (Fig. 16). In the naturally occurring steroids the ring junctions are all *trans*, except in the case of the cardiac glycosides where both A/B and C/D junctions are *cis*. Rings fused through nonadjacent atoms, or bridged systems, may also display stereoisomerism.

Conformational Isomerism

In the preceding sections the nature and properties of configurational (optical and geometric isomers) were discussed. It is important to remember that configurational isomers are distinct, separable compounds.

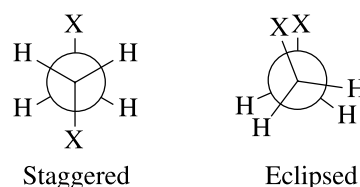


Fig. 17 Newman projection formulas for staggered and eclipsed rotamers.

Conformational isomers are different three-dimensional arrangements in space of the atoms of a single compound or configurational isomer. Such isomers are termed conformers and are interconvertible by free rotation about single bonds.

In alkanes or alkyl systems, an infinite number of conformations is possible as a result of rotation about CC single bonds, and each conformation has a certain potential energy. Two conformational extremes, one of low (minimal steric interaction between bond substituents) and one of high (maximal steric interaction) potential energy, have been described for these systems. These are depicted as Newman projections in Fig. 17. A more recent nomenclature system for conformational isomers arising from energy barriers associated with rotation about a C–C single bond is shown in Fig. 18. In this case, the most stable conformer is the antiperiplanar isomer, where the distance between the two bulky groups (X and Y) is maximized, thus minimizing steric interactions. The least stable conformer is the fully eclipsed or synperiplanar conformer where the two groups are closest together. The other conformers, the anticlinal and gauche or synclinal are of intermediate stability.

Conformational isomerism is believed to be of great significance for drug-receptor and drug-enzyme

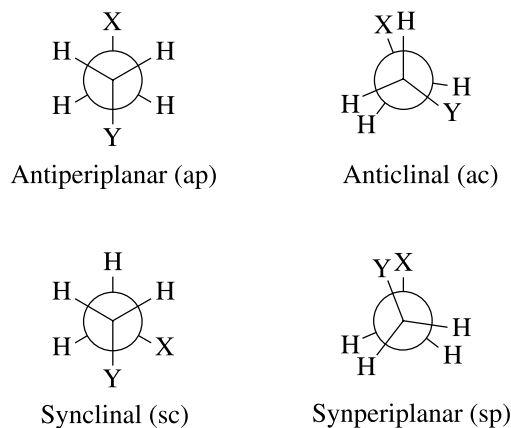


Fig. 18 Conformational isomers.

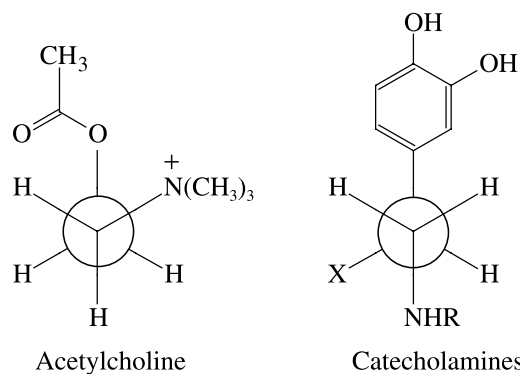


Fig. 19 Pharmacophoric conformations of acetylcholine and norepinephrine.

interactions. For example, experimental data suggest that catecholamines such as norepinephrine and dopamine interact with their receptors in the antiperiplanar conformation. Furthermore, the preferred (pharmacophoric) conformation of acetylcholine at its muscarinic receptor sites is the synclinal conformer (Fig. 19).

Conformational isomerism in cycloalkyl and cycloheteroalkyl structures is characterized by several different conformational extremes. For example, cyclohexane systems can exist in three distinct conformations: boat, twist boat, and chair. Of these, the chair form is the most stable conformation because steric interactions are minimized (Fig. 20). The substituents present on a ring conformer are designated as axial or equatorial, depending on the direction of projection from the average plane of the carbon skeleton. Substituents that project directly up or down from the ring are axial, and those in the plane of the ring are equatorial. Because of the conformational flexibility of cycloalkanes such as cyclohexane, the ring conformation can invert. During inversion, all axial substituents become equatorial and all equatorial substituents become axial.

Ring conformations are believed to be important in drug activity. For example, experimental evidence suggests that the analgesic fentanyl binds to opiate receptors preferentially in the conformation shown in Fig. 21, where the

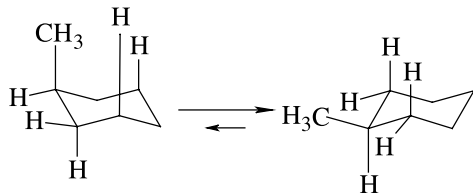


Fig. 20 Conformational isomerism of cycloalkanes.

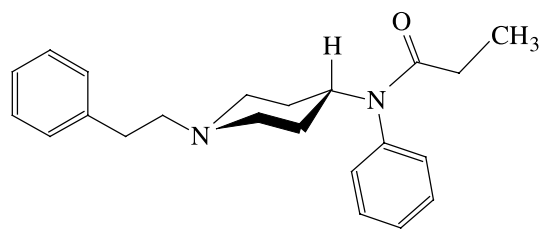


Fig. 21 Conformational isomerism, 3-methylfentanyl.

bulky substituents are positioned equatorially to minimize conformational instability.

PHARMACOLOGICAL ASPECTS OF ISOMERISM

Background and Definitions (1, 2)

Structural and steric complementarities of drug molecules with their target sites of action are essential criteria for the production of a pharmacological effect. The requirement of stereocomplementarity for organic medicinal agents reflects the inherent dissymmetry of the biological system comprised of stereoisomeric proteins, amino acids, lipids, carbohydrates, nucleic acids, etc. Molecular dissymmetry of biological systems was initially demonstrated in the 1880s by Pasteur. Then in 1908 Cushney explained differences in pharmacological actions of epinephrine enantiomers on the basis of dissymmetric drug receptors. Easson and Stedman in 1933 proposed a model for the molecular basis of sympathomimetic amines stereoselectivity at adrenergic receptors based on optimization of the number of interactions for each stereoisomer. Hence, the more efficacious (*R*)-(-)epinephrine isomer can achieve a three-point receptor interaction while the less active (*S*)-(+)-enantiomer and its achiral desoxy analogue, dopamine, can only achieve a 2-point interaction (Fig. 22). While Easson–Stedman's description of stereoisomeric drug action is a rather simplistic approximation of the steric aspects of drug action, it served to stimulate concerted chemical and pharmacological study of the phenomenon of stereoselectivity of drug action. Pfeiffer's rule, an attempt to quantifying drug stereoselectivity, states that, in the case of enantiomeric drugs, the greater the activity of the racemate the higher the ratio of the enantiomer's activity. A quantitative treatment of stereoisomeric drug action is the technique of eudismic analysis, which is most commonly applied to studies of enantiomeric drug action. In an eudismic analysis, the more active enantiomer (e.g., higher receptor

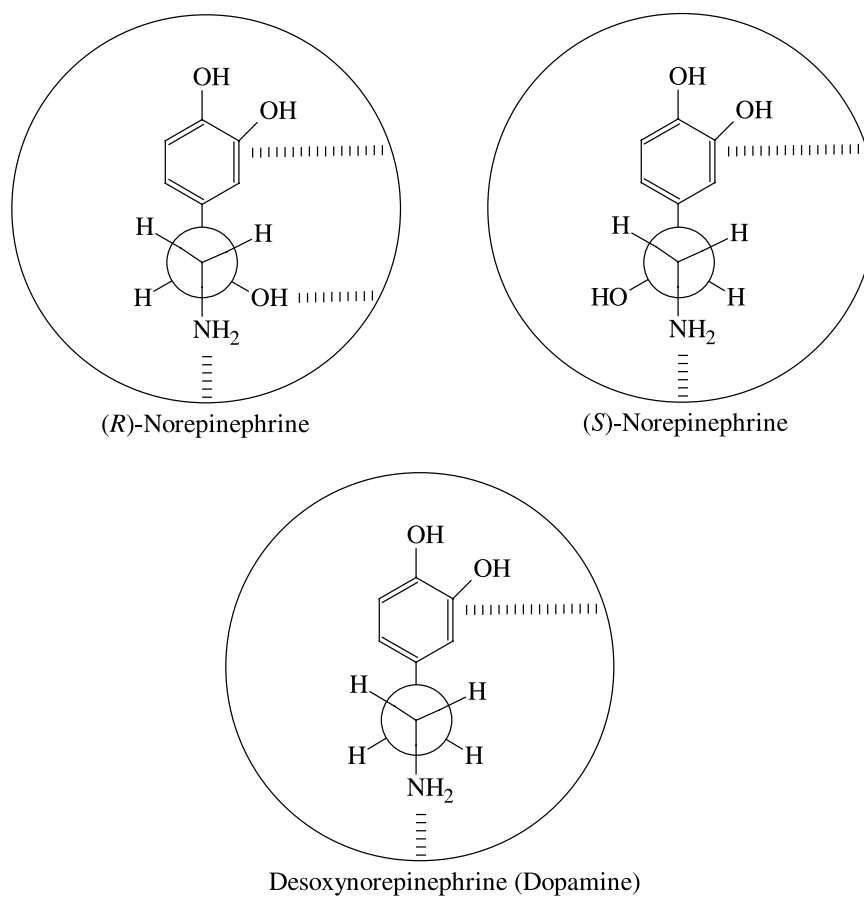


Fig. 22 The Easson-Stedman hypothesis of stereoisomeric drug action.

affinity—aff) is termed the eutomer and its less active mirror image form (e.g., relatively lower receptor affinity) is referred to as the distomer. Stereoselective pharmacological profiles of an enantiomer can be quantitated as a ratio of specific pharmacological actions termed the eudismic ratios (Eq. 4). The logarithm of the eudismic ratio is termed the eudismic index (Eq. 5).

$$\text{Eudismic ratio (ER)} = \text{Aff}_{\text{eu}} + \text{Aff}_{\text{dis}} \quad (4)$$

$$\text{Eudismic index (EI)} = \log \text{Aff}_{\text{eu}} + \text{Aff}_{\text{dis}} \quad (5)$$

Steric Aspects of Drug Action (3–7)

It is estimated that approximately one-half of all drugs worldwide exist as stereoisomers. However, only one-half of stereoisomeric drugs are marketed as the individual stereoisomer and most of the latter are of natural or semi-synthetic origin. There is increasing awareness of the clinical importance of drug stereoselectivity because differences in the behavior of isomers in the chiral living

system can result in significant differences in clinical outcomes. Table 1 presents a number of examples of these difference.

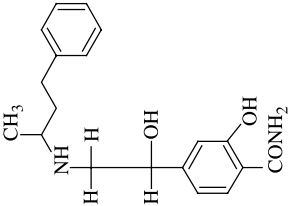
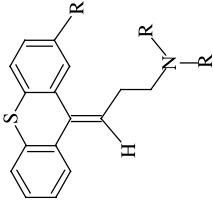
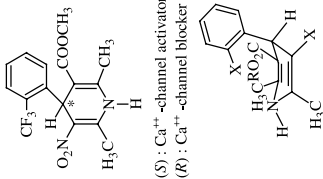
The previous examples have described the varying influences of chirality on pharmacological activity. Table 1 also illustrates the influence of other types of stereoisomerism on the pharmacological profiles of medicinal agents including that of geometric isomerism on the antipsychotic activity of thioxanthene derivatives. The eutomeric relationship of the Z-thioxanthene has been ascribed to its greater conformational complementarity with the pharmacophoric extended conformation of that natural receptor ligand dopamine. Both conformational and configurational isomerism are important in the pharmacological actions of the calcium-channel-blocking 1,4-dihydropyridines. Unsymmetrical ester substitutions at C3 and C5 result in chirality at C4 with the enantiomers of selected derivatives possessing opposite effects on calcium-channel function. Further, the availability of a “boat-like” conformation of the dihydropyridines has been

Table 1 Steric aspects of drug action

Stereoisomeric drug	Structure	Comments
<i>R</i> - and <i>S</i> - α -Methyldopa		Only the <i>S</i> isomer possess pharmacologic activity (stereo specificity of action) because only the <i>S</i> isomer can be stereoselectively bioactivated to (1 <i>R</i> ,2 <i>S</i>)- α -methylnor-epinephrine, a presynaptic α 1-sympatho-mimetic antihypertensive.
<i>R</i> - and <i>S</i> -Flecainide		An example of a lack of stereoselectivity of pharmacological action. The enantiomers of flecainide demonstrates an eudismic ratio of \sim 1.0 on excitable tissue of the canine myocardium.
<i>R</i> - and <i>S</i> -Warfarin		The anticoagulant activity of warfarin is a classic example of stereoselective drug action. <i>S</i> -warfarin in vivo is from two- to five-fold more anticoagulant than its <i>R</i> -enantiomer. This potency difference is coincidentally offset by the two-or five-fold greater plasma clearance of the distomer.
<i>R</i> - and <i>S</i> -Albuterol		The <i>S</i> -isomer of albuterol has been shown to oppose the bronchodilatory effects of the eutomer, <i>R</i> -albuterol and may therefore contribute to paradoxical bronchospasm and the occurrence of severe reagenic-like reactions seen with racemic albuterol.
<i>R</i> - and <i>S</i> -Bupivacaine		Bupivacaine exhibits stereoselectivity of local anesthetic activity, CNS and cardiac toxicity and pharmacokinetics. <i>R</i> -Bupivacaine is more potent, more toxic and more rapidly cleared than its <i>S</i> -isomer.

(Continued)

Table 1 Steric aspects of drug action (Continued)

Stereoisomeric drug	Structure	Comments
Labetalol stereoisomers	 <p>Labetalol stereoisomers</p>	These stereoisomers provide an example of different but complementary pharmacological of stereoisomers contributing to overall therapeutic utility. <i>R,R</i> -labetalol is a selective β -sympatholytic while the <i>S,R</i> -isomer is an α_1 -adrenolytic eutomer.
Antipsychotic thioxanthenes		Z-orientation of the propylamine side chain enhances dopamine receptor site recognition of these agents.
Dihydropyridine derivatives	 <p>(<i>S</i>) : Ca^{++} -channel activator (<i>R</i>) : Ca^{++} -channel blocker</p>	Unsymmetrical ester substitutions at C3 and C5 result in chirality at C4 of the dihydropyridines. Separation of pharmacological actions is seen in appropriately structures optical isomers of chiral dihydropyridines. Further, the availability of a “boat-like” conformation of the dihydropyridines has been proposed to facilitate the calcium channel blocking activity of this class of agents.

proposed to facilitate the calcium channel blocking activity of this class of agents.

Hence it is obvious that stereoisomeric influences as well as structural effects, play a major role in the production of the pharmacological profiles of medicinal agents. Variations noted in vitro that are related to the three-dimensional shape of a drug molecule originate primarily in stereoselective processes involved in the target site. In vivo variations related to stereochemical influences reflect selective interactions of the drug both at its site of action and in the various phases of biodisposition.

STERIC ASPECTS OF PHARMACOKINETICS

The macromolecules of the body can distinguish between drug isomers leading to stereoselective differences in tissue and protein binding, biotransformation, and renal excretion. Slight spatial differences in stereoisomers can have marked effects on the degree of association and interaction with proteins and enzyme systems. Stereoisomeric drugs often display stereoselectivity in pharmacokinetic processes and pathways that require interaction and association with macromolecules and enzymes systems. Furthermore, diastereomers display a higher degree of stereoselective than enantiomers due to their differences in solubility and partitioning. Since enantiomers have identical solubility and partitioning characteristics, their stereo selective pharmacokinetics are expected to reside in more structurally specific processes such as membrane transport, biotransformation, and binding to proteins and tissues.

Drug Absorption

The passive gastrointestinal absorption of enantiomeric drugs would be expected to be similar since the physical properties of partitioning and solubility of enantiomers are the same. Transdermal absorption of drugs such as ketoprofen also appears to be nonstereo specific (8).

Stereoisomers with structural similarities to endogenous entities and nutrients display differences in permeability rates across the gastrointestinal membrane and hence in bioavailability. L-DOPA, which is absorbed by an amino acid transport system, passes the gastrointestinal wall at a rate four to five times that of the D-enantiomer (9). L-methotrexate is absorbed by active processes in the gastrointestinal tract, and the D-isomer is reportedly absorbed by passive absorption. Another example of enantiomeric effects on gastrointestinal absorption is the

reported stereoselective absorption and intestinal biotransformation of cephalixin in rats (10).

In addition to stereospecific membrane permeability, there is a potential for stereoselective drug and dosage form effects. The crystalline forms of racemates may not be the same as the crystal structures of the individual stereoisomers and may be a source of differences in rates of dissolution between racemic and single enantiomer dosage forms. With respect to formulation, stereoselective interactions between solid dosage form excipients that are chiral, such as cellulose derivatives, may have the potential of altering the dissolution rates of stereoisomers. Further investigations of excipients and dosage form influences on the dissolution of chiral drugs are needed.

Drug Distribution

The extent and degree of interactions between chiral macromolecules of the body and stereoisomers is a source of observable differences in isomeric drug distribution. Stereoselectivity in drug distribution may occur when tissue or protein binding or uptake is associated with structurally specific receptor, protein, or enzyme binding. Since only unbound or free drug is susceptible to elimination and distribution to receptors and other tissues and fluids, differences in the protein and tissue binding of stereoisomers are reflected in their overall pharmacokinetic profiles.

Stereoselective interactions with both isolated human albumin and α_1 -acid glycoprotein have been observed with a variety of protein-binding techniques. The interaction of an enantiomer with a plasma protein yields a diastereomeric association. The *S* isomer of warfarin is bound to a greater extent to albumin than the *R* isomer (11). α_1 -Acid glycoprotein binds *S*-propranolol (87.3%) to a slightly higher degree than *R*-propranolol (83.8%), whereas human albumin binds *R*-propranolol more strongly than the *S* form (12).

The extent of binding differences between isomers is most readily noted from ratios of fractions free in the plasma; this ratio can be as high as three (Table 2). Recognizing that plasma protein binding impacts on drug distribution and elimination, differences in plasma protein binding between stereoisomers may lead to misinterpretation of pharmacokinetic comparisons between isomers, unless protein binding of the isomers is considered.

For enantiomeric drugs with low organ clearance, differences in renal or hepatic clearance between stereoisomers may reflect their free fraction in the plasma and not real stereoselectivity of the ability of the organ to remove the free enantiomers (intrinsic clearance) from the plasma. Clearance differences between stereoisomers of

Table 2 Stereoselective biodisposition

Parameter	Isomer	Warfarin	Ibuprofen ^a	Disopyramide ^b	Propranolol ^c	Verapamil ^d
Total clearance	<i>R</i>	3.5 ml/h/kg	68 ml/min	111 ml/min	1210 ml/min	10.24 ml/min/kg
	<i>S</i>	4.9 ml/h/kg	74 ml/min	111 ml/min	1030 ml/min	18.1 ml/min/kg
	<i>R/S</i>	0.71	0.92	1.0	1.17	0.57
Volume of distribution	<i>R</i>	0.154 L/kg	9.9 L	48 L	4.8 L/kg	2.74 L/kg
	<i>S</i>	0.16 L/kg	10.5 L	50 L	4.1 L/kg	6.42 L/kg
	<i>R/S</i>	0.96	0.94	0.96	1.17	0.43
$t_{1/2}$, h	<i>R</i>	35	2	5.2	3.6	4.08
	<i>S</i>	24	1.7	5.5	3.5	4.81
	<i>R/S</i>	1.46	1.18	0.95	1.03	0.85
Unbound, %	<i>R</i>	1.2	—	34	20.3	11.0
	<i>S</i>	0.9	—	22.2	17.6	6.4
	<i>R/S</i>	3.3	—	1.53	1.15	1.72

^a(Ref. 26.)^b(Ref. 14.)^c(Ref. 12.)^d(Ref. 59.)

verapamil and disopyramide may be a function of plasma protein binding differences. In addition, volumes of distribution as well as concentration ratios of stereoisomers in body fluids to total plasma and blood are influenced by plasma protein binding. For example, the larger volume of distribution and greater total body clearance of *R*-disopyramide compared to the *S* isomer may be explained by the lower plasma protein binding of the *R* isomer (13, 14). Also, the higher synovial fluid concentrations of *S*-ibuprofen following administration of the racemate appears to be related to the lower plasma protein binding of this isomer compared to that of the *R* isomer (15).

An unique example of stereoisomer selective tissue uptake of stereoisomers is noted with NSAIDs. The *R* enantiomer of ibuprofen shows preferential uptake into fat following the administration of the racemate and individual isomers (16). However, this apparent difference in fat distribution is probably a consequence of the selective metabolic uptake and formation of the coenzyme A thioester of the *R* isomer, which does not occur with the *S* isomer.

Drug Biotransformation

Numerous metabolic pathways involving mixed-function oxidases, esterases, transferases, and hydroxylases exhibit selectivity toward stereoisomeric substrates. Of all disposition differences that stereoisomers may display, the greatest stereoselectivity is expected in biotransformation, because of the specificity of metabolic enzymes and isoenzymes. The overall differences in hepatic clearance

of stereoisomers reflect not only differences in intrinsic clearance (activity of drug metabolizing enzymes) for the isomers but also the steric effects of plasma protein binding and hepatic blood flow.

When stereoisomers are biotransformed by a variety of pathways, differences in the susceptibility of the separate isomers to these pathways result in stereoselectivity for their metabolite patterns. For example, *S*-warfarin is oxidized to form primarily 7-hydroxy-*S*-warfarin, whereas the *R* enantiomer predominantly undergoes hydroxylation in the 6-position (17, 18). Oxazepam glucuronidation is 3–3.4 times higher for the *S* isomer compared to the *R* isomer in man and dogs (19). Biotransformation may generate an additional chiral center in the drug structure and result in diastereomeric metabolites with markedly different disposition characteristics.

Some of the greatest differences in the pharmacokinetics of stereoisomers can be attributed to stereoselective hepatic biotransformation. The oral clearance of *S*-mephenytoin is 170 times that of the *R* enantiomer in extensive metabolizers of the drug (20). This large difference in clearance is reflected in a 2-h half-life for the *S*-mephenytoin compared to a 76-h half-life for the *R* enantiomer in the same patient group. Interestingly, the half-life of the *S* isomer (63 h) and the *R* isomer (77 h) are similar in poor metabolizers of mephenytoin. Numerous other examples of stereoselectivity in hepatic clearance can be found in the literature (21–23).

Metabolic processes in humans and animals may invert the configuration of several NSAIDs. For many

2-arylpropionic acid NSAIDs in humans, the *S* enantiomer is the pharmacologically active form. Following administration of the racemate or *R* enantiomer, a slow inversion of the *R* to the *S* enantiomer occurs. All the 2-arylpropionic acid NSAIDs are administered as the racemate with the exception of naproxen. Based on investigations of ibuprofen, the *R* enantiomer is inverted to the *S* enantiomer via coenzyme A thioester formation (24, 25). It has been estimated, from pharmacokinetic studies of racemic ibuprofen that approximately 63% of the *R* isomer is inverted to the pharmacologically active *S* form (26). Stereoisomeric inversion is not universal for all 2-arylpropionic acids in humans with reports of inversion for ibuprofen, fenoprofen, and benoxaprofen and little or no evidence of inversion for tiaprofenic acid, ketoprofen, indoprofen, and carprofen (27–33).

Racemic drugs that experience hepatic “first-pass” metabolism of one or both isomers may experience stereoselectivity in biotransformation prior to the drug entering the systemic circulation, resulting in a route of administration effect on plasma concentration ratios and therapeutic response. Following IV doses of propranolol, plasma concentrations of both enantiomers appear to be similar, with only slight differences in total body clearance between the isomers (34). When propranolol is given orally, the higher rate of biotransformation and lower plasma binding of the less active *R* isomer results in lower plasma concentration of this isomer compared to the active *S* isomer. Other drugs administered as racemate such as verapamil, metoprolol, and prilocaine are reported to display stereoselective hepatic first-pass effects on the oral bioavailability of their isomers (35–38).

Renal and Biliary Excretion

The tubular secretory contribution to drug renal clearance has the potential of producing stereoselective renal elimination in the handling of stereoisomers. Enantiomers of terbutaline, disopyramide, and pindolol exhibit stereoselective renal clearance, probably due to differences in tubular secretion of individual stereoisomers (39–41). When examining the renal clearance of unbound disopyramide, the *L*-enantiomer has a renal clearance 29–86% higher than that of the *D*-isomer. The overall clinical significance of renal clearance differences for stereoisomers depends on the importance of renal elimination in the drug's total elimination. Biliary clearance of stereoisomers is difficult to evaluate in light of the techniques required and the many stereoselective factors influencing rate of biliary excretion. For these reasons, uncomplicated examples of stereoselective biliary excretion are not common.

Pharmacodynamics and Stereoselective Pharmacokinetics

The potency ratio (eudismic ratio) for stereoisomeric drugs reflects not only pharmacokinetic but also pharmacodynamic variations for the individual stereoisomers. In many respects, pharmacodynamic differences contribute more to large potency and activity differences of stereoisomers. Stereoselectivity at the level of drug-receptor interactions appears to be more pronounced than that noted for biotransformation or protein-binding interactions. A wide range of drug classes have been identified in which stereoselective pharmacodynamics are seen (4). Evidence for pharmacodynamic differences between stereoisomers is found in studies of isolated tissue and organ preparations and the relationships of drug response to plasma concentrations (42–44).

Analysis of pharmacological effects versus plasma concentrations of stereoisomers have been interpreted in terms of pharmacodynamic stereoselectivity. Pharmacological and toxicologic effects should be related to free plasma concentrations of stereoisomers in order to be able to appropriately compare activity or potency. Comparison of the free warfarin plasma concentration required to produce a 50% inhibition in synthesis of the prothrombin-activity complex reveals that *S*-warfarin has five times the inhibitory activity of the *R* isomer (45). The impact of the higher pharmacodynamic activity of *S*-warfarin on the dose-effect relationships of warfarin stereoisomers is partially offset by the fact that the *S* isomer is metabolically cleared more rapidly.

Pharmacokinetic Implications and Considerations

Recognizing that stereoisomers may vary markedly in their pharmacokinetic and pharmacodynamic profiles raises questions as to the usefulness of drug disposition, bioavailability, and pharmacodynamic modeling studies in which racemic forms are administered and nonstereospecific assays are utilized. Total body clearances, volumes of distribution, and bioavailability parameters become complex when the plasma concentrations of both stereoisomers are added. Clearance and bioavailability values observed after dosing of a racemic agent reflect a value falling between that of the two isomers (46–49). The observed half-life of the racemate is that of the stereoisomer with the longest half-life. The necessity of providing pharmacokinetic and pharmacodynamic data for both stereoisomers as well as their possible interaction could discourage the development of individual stereoisomers of drugs.

Since the administration of a racemate involves two separate drug entities, there can be intersubject variability in pharmacokinetic and pharmacodynamic profiles for both stereoisomers. Intersubject variability in the pharmacokinetics of both isomers leads to a large range of plasma concentration ratios across populations of patients. Factors such as disease, age, genetics, and concurrent drug therapy have been shown to lead to differing effects on stereoisomers of a drug. For example, the ratio of *S*- to *R*-tocainide plasma concentrations may vary over a 1.5- to threefold range (50, 51). *R*-Tocainide is believed to be several times more active than the *S* isomer (52). Factors such as disease, age, genetics, and concurrent drug therapy have been shown to lead to differing effects on stereoisomers of a drug. Liver disease patients have lower *S/R* ratios of plasma concentrations for ibuprofen stereoisomers than healthy subjects and this may be related to decrease conversion of the *S*- to *R*-isomer (53).

A major source of intersubject variability for stereoisomers is polymorphism in drug biotransformation (54). Extensive metabolizers of metoprolol display higher clearances of both stereoisomers than poor metabolizers, but the magnitude of the differences between stereoisomers is not the same within each population (38). Extensive metabolizers of metoprolol require lower total (*R* and *S*) plasma concentration of metoprolol for the same degree of beta blockade (51). In addition, *R*-mephenytoin shows similar oral metabolic clearance in extensive and poor metabolizers, whereas *S*-mephenytoin, which undergoes 4-hydroxylation, has markedly higher clearances in extensive metabolizers (55).

The area of clinical pharmacology that first directed attention to the consequences of stereoisomerism on therapeutic and pharmacokinetics was that of drug interactions, particularly those of the anticoagulant warfarin. Not only may drug interactions be stereoselective, but there is a potential for one stereoisomer to alter the pharmacokinetics and pharmacodynamics of the other. A classical example is the interaction with achiral phenylbutazone, which inhibits the metabolism of active *S*-warfarin but stimulates the metabolism of the less active *R* isomer (56). Other stereoselective drug interactions include the induced elimination of misonidazole by phenytoin (57). Phenytoin enhances the clearance of (4-)-misonidazole by 56%, which is higher than the increase in clearance of (-)-misonidazole.

Identical chemical and physical properties of enantiomers represent a potential source for enantiomer-enantiomer interactions at both pharmacokinetic and pharmacodynamic levels. Whether by competition for plasma- or tissue-binding sites or for drug-metabolizing enzymes, enantiomers may exhibit changes

in pharmacokinetics when administered as a racemate compared to individual stereoisomers. The enantiomers of disopyramide exhibit similar clearance and volumes of distribution when given separately (14). However, when administered as the racemate, the *S* isomer has higher clearance and volume of distribution. Ibuprofen enantiomers have also been shown to undergo enantiomer-enantiomer interactions (58).

In the past, pharmacokinetic and pharmacodynamic investigations of chiral drugs have neglected the influences of stereoisomerism. This is primarily a result of the lack of stereospecific analysis procedures. Nonstereospecific assays give pharmacokinetic and pharmacodynamic information which represents a complex combination of the characteristics of the separate stereoisomers. With the advent of stereospecific analysis procedures a better understanding of drug kinetics and action as possible.

ANALYTICAL METHODOLOGY AND ISOMERISM

A diastereoisomeric interaction is always required for the resolution of enantiomeric substances. This interaction occurs between the enantiomers of interest and a second enantiomeric species often referred to as the chiral selector. The diastereoisomeric interaction between the enantiomers and the chiral selector may involve a covalent bond or other less stable noncovalent associations. In the example below, a mixture of *R* and *S* isomers associates with the *R* isomer of the chiral selector to yield two diastereoisomeric products:



In this process, enantiomers (*R* and *S*) of the heterochiral substance of interest interact with the homochiral selector (*R*) to yield two products having a diastereoisomeric relationship. The formation of diastereoisomeric complexes is the basis for enantiomeric separations. The ability to use smaller differences in the rates of formation, stability, and properties of these complexes has been responsible for the major advances in stereochemical separations.

Chromatographic Techniques

The chromatographic separation of enantiomers, often referred to as enantioseparation, has received a great deal of attention in recent years. Both liquid (LC) and gas (GC) chromatographic procedures are used. The former is

extremely useful for enantioseparations because of the available variations in scale, mechanism, and technique. It has been used in enantioseparations from analytical to preparative in scale, taking advantage of various modes of diastereoisomeric interactions and using elution and displacement techniques. All the chromatographic methods involve diastereoisomeric interactions between the enantiomers of interest and a second chiral substance, the chiral selector. The difference in the interaction of the two enantiomers with the chiral selector is the enantioselectivity, $\alpha = k'_R/k'_S$; it can be thermodynamic or kinetic in nature. The choice of k'_R or k'_S as the numerator of the enantioselectivity expression is made for a ≥ 1.0 . The k' (capacity factor) is defined in the customary manner in Equation (6).

$$k' = (V_R - V_0)/V_0 \quad (6)$$

It represents the number of column volumes required to elute the enantiomer. The volume of mobile phase required to elute the component V_R (the elution volume, retention volume) and the void volume, V_0 , are the chromatographic measurements needed to calculate the capacity factor. The void volume is defined as the volume of mobile phase required to elute an unretained solute.

Chromatographic enantioseparations can be achieved when the diastereoisomeric interaction is established via a precolumn covalent derivatization, or a noncovalent association formed within the chromatographic system through the use of chiral mobile phase additives (primarily an LC technique) or chiral stationary phases.

PRECOLUMN DERIVATIZATION

The reaction of a homochiral derivatizing agent with a heterochiral sample to yield covalently linked diastereoisomeric products is a precolumn technique used for the analysis of the individual enantiomers. The diastereoisomeric products are separable under a variety of chromatographic conditions (GC and LC), including both normal and reversed-phase procedures (59–61). Fig. 23 shows the LC separation of two diastereoisomeric products formed by the reaction of a heterochiral amine and a homochiral derivatization reagent. The derivatization reaction involves the formation of an amide by treating an *N*-substituted *S*-prolyl chloride with racemic amphetamine (Fig. 24).

The configuration of the asymmetric center in the prolyl moiety is homochiral *S* and the racemic amphetamine (*R* and *S*) yields diastereoisomeric amides *S,R* and *S,S*. Thus, these amides have the same configurations (*S*) at the center

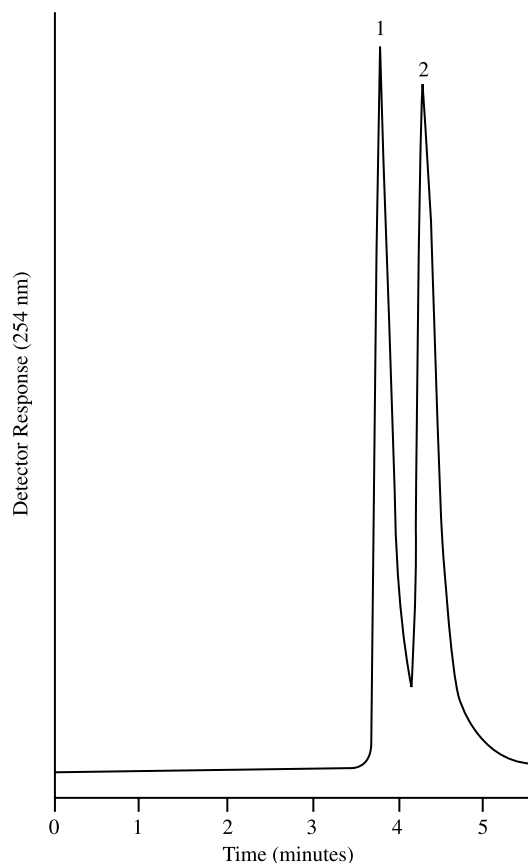


Fig. 23 Normal-phase liquid chromatographic separation of 4-nitrophenylsulfonyl-*S*-prolylam-phenetamine. 1 = *R*-amphetamine, 2 = *S*-amphetamine. (Ref. 60.)

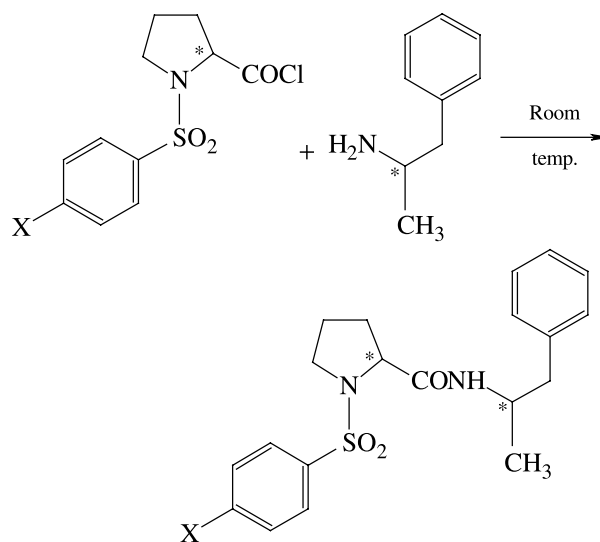


Fig. 24 Precolumn diastereomeric derivatization.

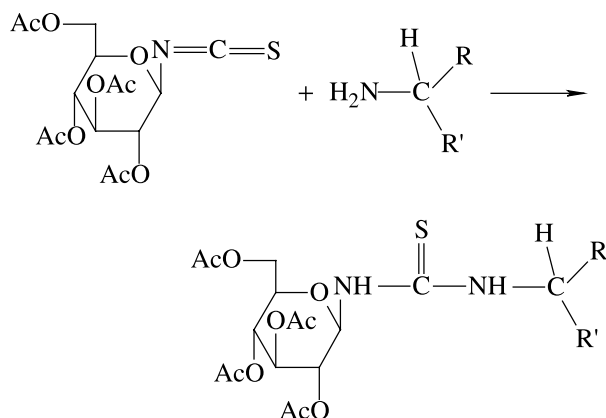


Fig. 25 GITC derivatization of a primary amine.

originating from the prolyl moiety, and differing configuration at the chiral center originating from heterochiral amphetamine. A similar derivatization procedure with *N*-trifluoroacetyl-*S*-prolyl chloride is used in GC to form diastereoisomeric amides. The *N*-TFA moiety imparts volatility and electron capture detectability to the derivatives for GC analysis (62).

The homochiral derivatization reagents are usually modified compounds that are available in high

enantiomeric purity. Cost and availability make naturally occurring substances such as amino acids excellent sources of enantiomeric purity. A reagent widely used in the diastereoisomeric derivatization of amines is 2,3,4,6-tetra-*O*-acetyl-13-D-glucosyl isothiocyanate (GITC) (Fig. 25). The isothiocyanate moiety reacts with primary and secondary amines yielding thiourea products, thus covalently linking the homochiral acetylated glucose and the amine (63). If the amine is racemic, the products are a diastereoisomeric pair. The configurations of all the chiral centers in glucose remain the same and the two configurations in the amine yield diastereoisomeric products (60). Fig. 26 shows the reversed-phase LC separation of diastereoisomeric ureas following derivatization of a racemic amine sample with GITC.

Mobile-Phase Additives

The direct separation of enantiomers by chromatography can be achieved by two fundamentally different processes: The diastereoisomeric interaction can occur between the sample molecules and the stationary phase or between the sample and chiral mobile-phase additives. Consequently, two different results can be obtained: A chiral stationary phase with an achiral mobile phase or an achiral stationary phase with a chiral mobile phase. The use of chiral mobile-phase additives is limited to LC procedures. The most widely applied examples include ligand exchange, ion-pair and inclusion complexation.

The exact sequence of events in the case of chiral mobile-phase additives varies, depending on the affinity of the additive for the achiral stationary phase. The chiral additive may associate with the stationary phase to become a dynamic chiral stationary phase, or the additive and sample molecules may form a diastereoisomeric complex in the mobile phase which associates with the achiral stationary phase.

Cyclodextrins

The cyclodextrins are perhaps the most widely used mobile-phase additives which form diastereoisomeric inclusion complexes. The cyclodextrins (α , β , and γ -cyclodextrins) are made up of α -D-glucose units (6, 7, or 8 α -D-glucose molecules) cyclized to form a truncated cone-like molecule (Fig. 27). The interior cavity of this molecule contains hydrogens and glycoside oxygens and is relatively lipophilic, whereas the exterior of the cyclodextrin molecule is composed primarily of hydroxyl groups and therefore is relatively hydrophilic. The wider end of the cone is composed of secondary hydroxyl groups and the narrow (truncated) end is composed of the primary

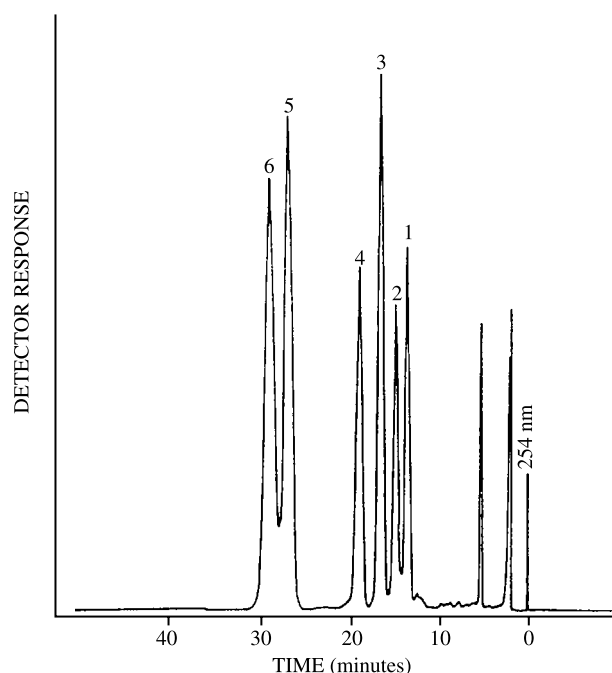


Fig. 26 Reversed-phase liquid chromatographic separation of GITC-derivatized amines. 1 = *R,R*-pseudoephedrine, 2 = *S,S*-pseudoephedrine, 3 = *R,S*-ephedrine, 4 = *S,R*-ephedrine, 5 = *S*-methamphetamine, 6 = *R*-methamphetamine. (From Ref. 61.)

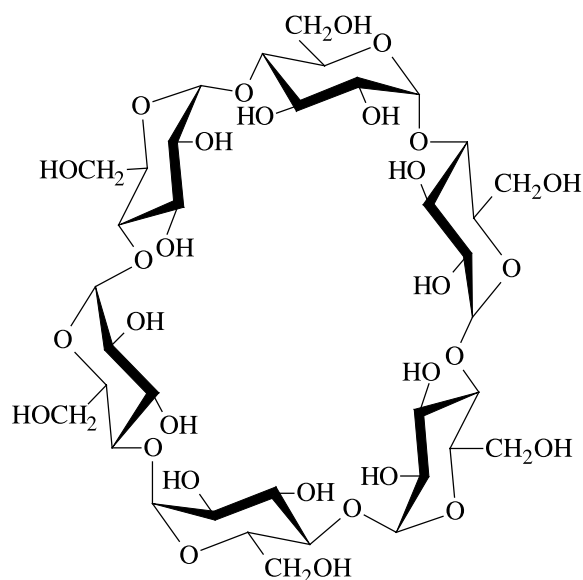


Fig. 27 Structure of α -cyclodextrin.

hydroxyl groups from the individual D-glucose units. Thus, complexation reactions occur only from the secondary OH side. The dimensions of the cavity in the cyclodextrin depend upon the number of glucose units in the compound.

The cyclodextrin-solute inclusion complexation is a reversible reaction with an equilibrium constant similar to that for diffusion-controlled processes. Thus, the effect of underivatized cyclodextrins as mobile-phase additives is often a reduction in column efficiencies. The increase in peak width in cyclodextrin systems compared to that of conventional reversed-phase systems requires a decrease in flow rates for best resolution. Some enantioseparations using this technique report flow rates between 20 and 40 $\mu\text{l}/\text{min}$ (64). Considerable column efficiency changes occur over a flow range of 0.02–2.0 ml/min. In addition to flow rate, the concentration of cyclodextrin in the mobile phase can be adjusted to optimize enantioseparations.

The pH, temperature, and ionic strength of the mobile phase also affect the cyclodextrin-solute complexation and retention properties. Many enantio-separations using cyclodextrin-modified systems involve solutes with an aromatic ring substituent or similar cyclic structure at the chiral center. A variety of chiral barbiturates, hydantoins, and related compounds (65–67) have been resolved by using β -cyclodextrin and alkylated B-cyclodextrin-modified systems.

Ligand Exchange

The ligand-exchange process has been applied as a mobile-phase-additive technique for enantioseparations.

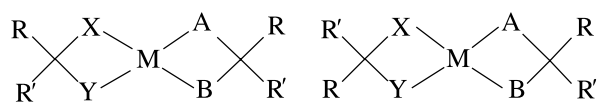
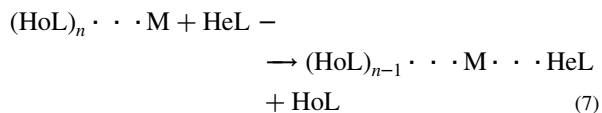


Fig. 28 Diastereoisomeric chelates.

It involves the formation of a dissociable diastereoisomeric complex between a homochiral additive and a heterochiral solute about a central metal ion (Fig. 28). The mobile phase contains both the homochiral ligand and the metal ion as additive components. These species probably exist as the fully complexed species with at least two molecules of the homochiral additive complexed to the metal ion. The ligand exchange with the components of the heterochiral sample enantiomers yields diastereoisomeric products. Obviously the concentration of the homochiral additive ligand is relatively high compared to the heterochiral sample molecule. Thus, the probability of displacing two molecules of homochiral ligands from the metal ion is extremely small, limiting the number of possible products to just two, as seen in Equation (6).



where HoL, is the homochiral ligand; HeL, is the heterochiral ligand, and; M, is the metal.

Ligand exchange has found extensive use in the resolution of amino acids, amino acid derivatives, and other bidentate ligands (68, 69). These separations are done using an aqueous or organic modified aqueous mobile phase in which the metal ion, such as copper(II) from copper sulfate, is soluble. The stationary phase is usually a hydrocarbon such as C_8 . Some of the widely used homochiral ligands such as L-proline and L-hydroxyproline (70) are N-alkylated with large hydrocarbon moieties such as C_8 to enhance the dynamic chiral stationary-phase process. The hydrocarbon moieties of the ligand and the stationary phase associate in the usual solvophobic process, leaving the chelating functional groups exposed to the mobile phase and its components.

The mobile-phase-additive ligand-exchange chromatography has been applied to the resolution of some chiral amino alcohols. The enantiomers of norephedrine, norpseudoephedrine, metaraminol, and phenylephrine were resolved (70) in a reversed-phase procedure using a 0.05 M acetate buffer mobile phase containing N-octylhydroxyproline and 5–8 mM copper(II). All these enantioseparations show base-line resolution

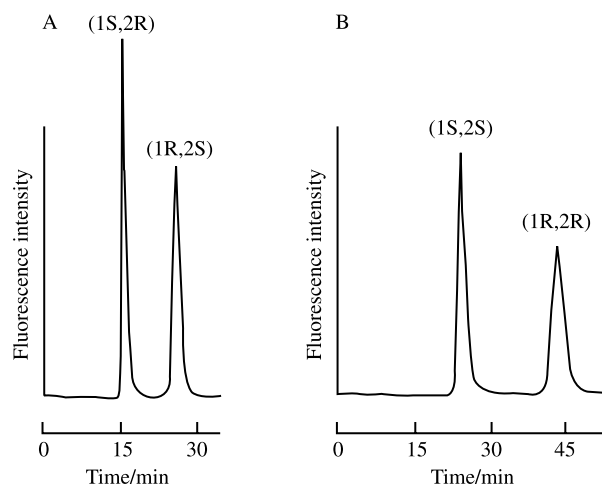


Fig. 29 Ligand-exchange liquid chromatographic separation of norephedrine (A) and norpseudoephedrine (B), using a C_{18} stationary phase coated with *N-n*-dodecyl-L-hydroxyproline and a mobile phase containing 5 mM copper(II). (From Ref. 72.)

with norpseudoephedrine (Fig. 29), yielding an enantioselectivity in the 2.0 range ($\alpha = 2.0$) and the 1S, 2S-isomer eluting before the 1R, 2R-enantiomer.

Diastereoisomeric Ion Pairs

The addition of a homochiral counterion to the mobile phase for the LC resolution of enantiomers can yield enantioseparations with only slight differences in the properties of the resulting diastereoisomeric ion pairs. Derivatives of L-proline and 10-camphorsulfonic acid have been used to resolve enantiomeric amines through the formation of diastereoisomeric ion pairs (71). The concept of reciprocity in these separations (72) has been shown to occur in such a way that if the *R* enantiomer of acid HA resolves base B into its *R*- and *S*-enantiomers, an enantiomer of B (for example, *R*-B) can be used to resolve *R*-HA and *S*-HA.

Chiral Stationary Phases

The use of chiral stationary phases (CSP) in liquid chromatography continues to grow at an impressive rate. These CSPs contain natural materials such as cellulose and starch as well as totally synthetic materials, utilizing enantioselective and retentive mechanisms ranging from inclusion complexation to π -electron interactions. The major structural features found in chiral stationary phases include cellulose, starch, cyclodextrins, synthetic polymers, proteins, crown ethers, metal complexes, and aromatic π -electron systems.

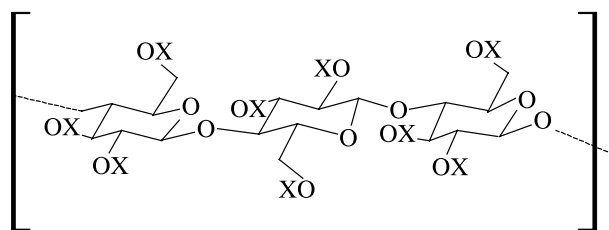


Fig. 30 Cellulose (XH) and derivatives.

Cellulose and Starch

Cellulose and starch are naturally occurring polysaccharides exhibiting enantioselective properties. Cellulose is a linear polymer of D-glucose units with individual fibers arranged in parallel bundles through hydrogen bonding between the fibers (Fig. 30). Starch is a similar material, containing a higher degree of branching with enantioselectivities similar to those of cellulose. The mechanism of retention and enantioselectivity for cellulose involves hydrogen bonding and possibly the formation of inclusion complexes within the cellulose structure. The chromatographic procedures used for these materials involve normal-phase conditions (nonpolar solvents), which allow for strong hydrogen-bond formation between the solute and CSP. A variety of stationary phases of varying enantioselective and retention properties have been prepared by chemical modification of the hydroxyl groups by esters, nitrates, carbamates, and ethers. Because of their excellent chiral recognition properties, these materials have been used to resolve a large number of pharmaceutically significant compounds (73).

Synthetic Polymers

Synthetic polymers can be prepared to contain chirality as is the case for cellulose and other natural polymers. Chirality can be introduced into the monomer before polymerization to yield the chiral polymer. Alternatively polymerization of an achiral monomer in the presence of some chiral catalyst yields the chiral polymer. Poly-methacrylates exhibiting chirality due to single-handed helicity have been prepared via polymerization in the

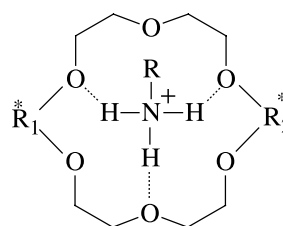


Fig. 31 Crown ether complexes.

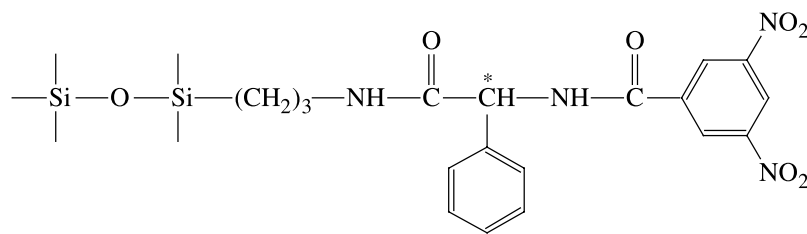


Fig. 32 3,5-Dinitrobenzoylphenylglycine (DNBPE) stationary phase.

presence of a chiral catalyst. These materials are used in liquid chromatography primarily under low-pressure conditions and have shown good resolution for compounds capable of hydrogen-bond formation.

Proteins

Proteins are naturally occurring chiral polymers that interact selectively with chiral solutes. The proteins most widely used as chiral selectors in liquid chromatography are albumin and α_1 -acid glycoprotein. These proteins are covalently linked to silica particles and are used primarily in the reversed-phase mode of separation.

Bovine serum albumin has a molecular weight of about 6.6×10^4 and consists of 581 amino acids. The tertiary structure of the protein is controlled by the 17 disulfide linkages, and at pH 7.0 the net charge on the protein is -18 . The human plasma protein, α_1 -acid glycoprotein, contains five carbohydrate units and has a molecular weight of 4.1×10^4 . The 14 sialic acid residues in the sugar units are involved in binding various cationic species and are believed to play a role in the chiral selectivity of this protein.

The mechanism of solute retention and chiral recognition is similar for both albumin and α_1 -acid glycoprotein and involves charge and dipolar interactions in addition to hydrophobic associations. The mobile-phase variables that can be adjusted to affect retention and chiral recognition include pH, ionic strength, and organic modifier (74). Such changes affect the properties of the solute as well as the stationary-phase protein. Thus, pH changes can alter the retention of a neutral solute by changing the properties of the binding site on the CSP.

Crown Ethers

These macrocyclic ethers assume a crown-like shape in solution with a central cavity capable of containing a small solute. They bind to small cationic species through association with the electron-rich oxygens of the ether

linkage. Chiral crown ethers (Fig. 31) serve as selectors for enantiomeric amines in the protonated state (75). They have been used as mobile-phase additives and been covalently linked to the surface of silica particles. The chromatographic procedure for these materials is primarily the reversed-phase mode, and many separations are done in a purely aqueous mobile phase modified to pH 1–2.0 with perchloric acid.

Chiral recognition in these phases is achieved when a chiral center is introduced into the crown ether to serve as a barrier to one enantiomer of the guest amine. The chiral barrier is usually a large bulky group that selectively affects the association of one enantiomer of the amine.

Aromatic π -electron Systems

These CSPs have in common a strong aromatic π – π interaction, dipolar stacking process as a central component of the retention process. They are often referred to as Pirkle columns based on the pioneering work by Pirkle and coworkers (76, 77). These π -donor and π -acceptor interactions occur though the tendency of some electron-rich aromatic systems to donate π -electrons (π -base) to an electron-deficient acceptor aromatic system (π -acid). These CSPs offer the clear advantage of complete structural characterization of the stationary phase. Since these are totally synthetic phases prepared by linking small chiral substances to silica particles, all the solute-stationary phase interactions can be readily predicted. The solute-CSP interactions include hydrogen bonds, steric, dipole–dipole, and conformational as well as π – π interactions. These phases generally consist of an acylated amino acid covalently linked to aminopropyl silica (Fig. 32). The acyl group is usually an aromatic acid, with the π acceptor 3,5-dinitrobenzoyl group among the most common. Phenylglycine and leucine provide chirality of general utility for the resolution of many chiral substances.

The major advantage of these phases is the ability to manipulate structure and the reciprocity of chiral recognition. The reciprocity approach is not readily available for the naturally occurring polymeric CSPs

such as cellulose, proteins, and others. Those solutes not containing the complementary π -acid or π -base functionality can often be derivatized to incorporate these features into the molecules of interest.

REFERENCES

- Williams, K.M. Molecular Asymmetry and its Pharmacological Consequences. *Adv. Pharmacol.* **1991**, 22, 57–135.
- Lehmann, P.A. Stereoisomerism and Drug Action. *Trends Pharmacol. Sci.* **1986**, 7 (7), 281–285.
- Ariens, E.J. Implications of the Neglect of Stereochemistry in Pharmacokinetics and Clinical Pharmacology. *Drug Intell. Clin. Phar.* **1987**, 21, 827–829.
- Jamali, F.; Mehvar, R.; Pasutto, F.M. Enantioselective Aspects of Drug Action and Disposition: Therapeutic Pitfalls. *J. Pharm. Sci.* **1989**, 78 (9), 695–715.
- Smith, D.F. *Handbook of Stereoisomers: Therapeutic Drugs*; CRC Press, Inc.: Boca Raton, FL, 1989; 405.
- Wainer, I.W.; Drayer, D.E. *Drug Stereochemistry: Analytical Methods and Pharmacology*; Marcel Dekker, Inc.: New York, 1988; 276.
- Hyneck, M.; Dent, J.; Hook, J. Chirality: Pharmacological Action and Drug Development. *Chirality in Drug Design and Synthesis*; Brown, C., Ed.; Academic Press: San Diego, 1990; 1–28.
- Panus, P.C.; Campbell, J.; Kulkarni, S.B.; Herrick, R.T.; Ravis, W.R.; Banga, A.K. Transdermal Iontophoretic Delivery of Ketoprofen Through Human Cadaver Skin and in Humans. *J. Control. Rel.* **1997**, 44 (2), 113–121.
- Wade, D.N.; Mearrick, P.T.; Morris, J.L. Active Transport of L-DOPA in the Intestine. *Nature (London)* **1973**, 242 (5398), 463–465.
- Tamai, I.; Ling, H.Y.; Timbul, S.M.; Nishikido, J.; Tsuji, A. Stereospecific Absorption and Degradation of Cephalexin. *J. Pharm. Pharmacol.* **1988**, 40 (5), 320–324.
- Brown, N.A.; Jahnshen, E.; Muller, W.E.; Wollert, U. Optical Studies on the Mechanism of the Interaction of the Enantiomers of the Anticoagulant Drugs Phenprocoumon and Warfarin with Human Serum Albumin. *Mol. Pharmacol.* **1977**, 13 (1), 70–79.
- Walle, U.K.; Walle, T.; Bai, S.A.; Olanoff, L.S. Stereoselective Binding of Propranolol to Human Plasma, Alpha 1-acid Glycoprotein and Albumin. *Clin. Pharmacol Ther.* **1983**, 34 (6), 718–723.
- Takahashi, H.; Ogata, H.; Shimizu, M.; Hashimoto, K.; Mashuhara, K.; Kashiwada, K.; Someya, K. Comparative Pharmacokinetics of Unbound Disopyramide Enantiomers Following Oral Administration of Racemic Disopyramide in Humans. *J. Pharm. Sci.* **1991**, 80 (7), 709–711.
- Giacomini, K.M.; Nelson, W.L.; Pershe, R.A.; Valdevieso, L.; Turner-Tamayasu, K.; Blaschke, T.F. In Vivo Interaction of the Enantiomers of Disopyramide in Human Subjects. *J. Pharmacokinet. Biopharm.* **1986**, 14 (4), 335–355.
- Day, R.O.; Williams, K.M.; Graham, G.G.; Lee, E.J.; Knihinicki, R.D.; Champion, G.D. Stereoselective Disposition of Ibuprofen Enantiomers in Synovial Fluid. *Clin. Pharmacol. Ther.* **1988**, 43 (5), 480–487.
- Williams, K.; Day, R.; Knihinicki, R.; Duffield, A. The Stereoselective Uptake of Ibuprofen Enantiomers into Adipose Tissue. *Biochem. Pharmacol.* **1986**, 35 (19), 3403–3405.
- Lewis, R.J.; Trager, W.F. Warfarin Metabolism in Man, Identification of Metabolites in Urine. *J. Clin. Invest.* **1970**, 49 (5), 907–913.
- Toon, S.; Low, L.K.; Gibaldi, M.; Trager, W.F.; O'Reilly, R.A.; Motley, C.H.; Goulart, D.A. The Warfarin–Sulfinpyrazone Interaction, Stereochemical Considerations. *Clin. Pharmacol. Ther.* **1986**, 39 (1), 15–24.
- Sisenwine, S.F.; Tio, C.O.; Hadley, F.V.; Liu, A.L.; Kimmel, H.B.; Ruelius, H.W. Species-Related Differences in the Stereoselective Glucuronidation of Oxazepam. *Drug Metab. Disp.* **1982**, 10 (6), 605–608.
- Wedlund, P.J.; Aslanian, W.S.; Jacqz, E.; McAllister, C.B.; Branch, R.A.; Wilkinson, G.R. Phenotypic Differences in Mephenytoin Pharmacokinetics in Normal Subjects. *J. Pharmacol. Exptl. Ther.* **1985**, 234 (3), 662–669.
- Godbillon, J.; Richard, J.; Gerardin, A.; Meinertz, T.; Kasper, W.; Jahnchen, E. Pharmacokinetics of the Enantiomers of Acenocoumarol in Man. *Br. J. Clin. Pharmacol.* **1981**, 12 (5), 621–629.
- Gill, T.S.; Hopkins, K.J.; Bottomley, J.; Gupta, S.K.; Rowland, M. Cimetidine–Nicoumalone Interaction in Man, Stereochemical Considerations. *Br. J. Clin. Pharmacol.* **1989**, 27 (4), 469–474.
- Chandler, M.H.; Scott, S.R.; Blouin, R.A. Age-Associated Stereoselective Alterations in Hexobarbital Metabolism. *Clin. Pharmacol. Ther.* **1988**, 43 (4), 436–441.
- Hutt, A.J.; Caldwell, J. The Metabolic Chiral Inversion of 2-arylpropionic Acids—A Novel Route With Pharmacological Consequences. *J. Pharm. Pharmacol.* **1983**, 35 (11), 693–704.
- Caldwell, J.; Winter, S.M.; Hutt, A.J. The Pharmacological and Toxicological Significance of the Stereochemistry of Drug Disposition. *Xenobiotica* **1988**, 18 (Suppl 1), 59–70.
- Lee, E.J.; Williams, K.; Day, R.; Graham, G.; Champion, D. Stereoselective Disposition of Ibuprofen Enantiomers in Man. *Br. J. Clin. Pharmacol.* **1985**, 19 (5), 669–674.
- Rubin, A.; Knadler, M.P.; Ho, P.P.; Bechtol, L.D.; Wolen, R.L. Stereoselective Inversion of (R)-fenoprofen to (S)-fenoprofen in Humans. *J. Pharm. Sci.* **1985**, 74 (1), 82–84.
- Bopp, R.J.; Nash, J.F.; Ridolfo, A.S.; Shepard, E.R. Stereoselective Inversion of (R)-(–)-benoxaprofen to the (S)-(+)-enantiomer in Humans. *Drug Metab. Disp.* **1979**, 7 (6), 356–359.
- Simmonds, R.G.; Woodage, T.J.; Duff, S.M.; Green, J.N. Stereospecific Inversion of (R)-(–)-benoxaprofen in Rat and Man. *Eur. J. Drug Metab. Pharmacokinetics* **1980**, 5 (3), 169–172.
- Singh, N.N.; Jamali, F.; Pasutto, F.M.; Russell, A.S.; Coutts, R.T.; Drader, K.S. Pharmacokinetics of the Enantiomers of Tiaprofenic Acid in Humans. *J. Pharm. Sci.* **1986**, 75 (5), 439–442.
- Foster, R.T.; Jamali, F.; Russell, A.S.; Alballa, S.R. Pharmacokinetics of Ketoprofen Enantiomers in Young and Elderly Arthritic Patients Following Single and Multiple Doses. *J. Pharm. Sci.* **1988**, 77 (3), 191–195.
- Tamassia, V.; Jannuzzo, M.G.; Moro, E.; Stegnjaich, S.; Groppi, W.; Nicolis, F.B. Pharmacokinetics of the

- Enantiomers of Indoprofen in Man. *Int. J. Clin. Pharmacol. Res.* **1984**, *4* (3), 223–230.
33. Stoltenborg, J.K.; Puglisi, C.V.; Rubio, F.; Vane, F.M. High-Performance Liquid Chromatographic Determination of Stereoselective Disposition of Carprofen in Humans. *J. Pharm. Sci.* **1981**, *70* (11), 1207–1212.
 34. Von Bahr, C.; Hermansson, J.; Tawara, K. Plasma Levels of (+) and (–)-propranolol and 4-hydroxypropranolol after Administration of Racemic (+/–)-propranolol in Man. *Br. J. of Clin. Pharmacol.* **1982**, *14* (1), 79–82.
 35. Tucker, G.T.; Mather, L.E.; Lennard, M.S.; Gregory, A. Plasma Concentrations of the Stereoisomers of Prilocaine after Administration of the Racemate, Implications for Toxicity? *Br. J. Anaesthesia*. **1990**, *65* (3), 333–336.
 36. Echizen, H.; Brecht, T.; Niedergesass, S.; Vogelgesang, B.; Eichelbaum, M. The Effect of Dextro-, Levo-, and Racemic Verapamil on Atrioventricular Conduction in Humans. *Amer. Heart J.* **1985**, *109* (2), 210–217.
 37. Echizen, H.; Vogelgesang, B.; Eichelbaum, M. Effects of D,L-verapamil on Atrioventricular Conduction in Relation to its Stereoselective First-Pass Metabolism. *Clin. Pharmacol. Ther.* **1985**, *38* (1), 71–76.
 38. Lennard, M.S.; Tucker, G.T.; Silas, J.H.; Freestone, S.; Ramsay, L.E.; Woods, H.F. Differential Stereoselective Metabolism of Metoprolol in Extensive and Poor Debrisoquin Metabolizers. *Clin. Pharmacol. Ther.* **1983**, *34* (6), 732–777.
 39. Borgstrom, L.; Nyberg, L.; Jonsson, S.; Lindberg, C.; Paulson, J. Pharmacokinetic Evaluation in Man of Terbutaline given as Separate Enantiomers and as the Racemate. *Br. J. Clin. Pharmacol.* **1989**, *27* (1), 49–56.
 40. Lima, J.J.; Boudoulas, H.; Shields, B.J. Stereoselective Pharmacokinetics of Disopyramide Enantiomers in Man. *Drug Metab. Disp.* **1985**, *13* (5), 572–577.
 41. Hsyu, P.H.; Giacomini, K.M. Stereoselective Renal Clearance of Pindolol in Humans. *J. Clin. Invest.* **1985**, *76* (5), 1720–1726.
 42. Raschack, M. Relationship of Antiarrhythmic to Inotropic Activity and Antiarrhythmic Qualities of the Optical Isomers of Verapamil. *Arch. Pharmacol.* **1976**, *294* (3), 285–291.
 43. Giacomini, K.M.; Cox, B.M.; Blaschke, T.F. Comparative Anticholinergic Potencies of R- and S- Disopyramide in Longitudinal Muscle Strips from Guinea Pig Ileum. *Life Sci.* **1980**, *27* (13), 1191–1197.
 44. Wise, R.; Wills, P.J.; Bedford, K.A. Epimers of Moxalactam, In Vitro Comparison of Activity and Stability. *Antimicrob. Agts. Chemother.* **1981**, *20* (1), 30–32.
 45. Levy, G.; O'Reilly, R.A.; Wingard, L.B., Jr. Comparative Pharmacokinetics of Coumarin Anticoagulants. XXXV. Examination of Possible Pharmacokinetic Interaction between (R)-(+)- and (S)-(–)-Warfarin in Humans. *J. Pharm. Sci.* **1978**, *67* (6), 867–868.
 46. Tucker, G.T.; Lennard, M.S. Enantiomer Specific Pharmacokinetics. *Pharmacol. Ther.* **1990**, *45* (3), 309–329.
 47. Ravis, W.R. Pharmacokinetics and Pharmacodynamics. *Stereoisomerism in Pharmaceuticals*; Riley, T.N., Ed.; Technomic Publishing Company: Lancaster, PA, 1991.
 48. Ravis, W.R.; Owen, J.S. Stereochemical Considerations in Bioavailability Studies. *Generics and Bioequivalence*; Jackson, A.J., Ed.; CRC Press: Ann Arbor, MI, 1994; 113–137.
 49. Mehvar, R.; Jamali, F. Bioequivalence of Chiral Drugs. Stereospecific Versus Non-Stereospecific Methods. *Clin. Pharmacokin.* **1997**, *33* (2), 122–141.
 50. Thomson, A.H.; Murdoch, G.; Pottage, A.; Kelman, A.W.; Whiting, B.; Hillis, W.S. The Pharmacokinetics of R- and S-tocainide in Patients with Acute Ventricular Arrhythmias. *Br. J. Clin. Pharmacol.* **1986**, *21* (2), 149–154.
 51. Sedman, A.J.; Gal, J.; Mastropaolo, W.; Johnson, P.; Maloney, J.D.; Moyer, T.P. Serum Tocainide Enantiomer Concentrations in Human Subjects. *Br. J. Clin. Pharmacol.* **1984**, *17* (1), 113–115.
 52. Block, A.J.; Merrill, D.; Smith, E.R. Stereoselectivity of Tocainide Pharmacodynamics In Vivo and In Vitro. *J. Cardiovasc. Pharmacol.* **1988**, *11* (2), 216–221.
 53. Li, G.; Treiber, G.; Maier, K.; Walker, S.; Klotz, U. Disposition of Ibuprofen in Patients with Liver Cirrhosis. Stereochemical Considerations. *Clin. Pharmacokin.* **1993**, *25* (2), 154–163.
 54. Testa, B.; Mayer, J.M. Stereoselective Drug Metabolism and its Significance in Drug Research. *Prog. Drug Res.* **1988**, *32*, 249–303.
 55. Wedlund, P.J.; Aslanian, W.S.; Jacqz, E.; McAllister, C.B.; Branch, R.A.; Wilkinson, G.R. Phenotypic Differences in Mephenytoin Pharmacokinetics in Normal Subjects. *J. Pharmacol. Exptl. Ther.* **1985**, *234* (3), 662–669.
 56. O'Reilly, R.A.; Trager, W.F.; Motley, C.H.; Howald, W. Stereoselective Interaction of Phenylbutazone with [12c/13c]Warfarin Pseudoracemates in Man. *J. Clin. Invest.* **1980**, *65* (3), 746–753.
 57. Williams, K.M. Kinetics of Misonidazole Enantiomers. *Clin. Pharmacol. Ther.* **1984**, *36* (6), 817–823.
 58. Lee, E.J.; Williams, K.; Day, R.; Graham, G.; Champion, D. Stereoselective Disposition of Ibuprofen Enantiomers in Man. *Br. J. Clin. Pharmacol.* **1985**, *19* (5), 669–674.
 59. Barksdale, J.M.; Clark, C.R. Liquid Chromatographic Determination of the Enantiomeric Composition of Amphetamine and Related Drugs by Diastereomeric Derivatization. *J. Chromatogr. Sci.* **1985**, *23* (4), 176–180.
 60. Noggle, F.T., Jr.; DeRuiter, J.; Clark, C.R. Liquid Chromatographic Determination of the Enantiomeric Composition of Methamphetamine Prepared from Ephedrine and Pseudoephedrine. *Anal. Chem.* **1986**, *58* (8), 1643–1648.
 61. Barksdale, J.M.; Clark, C.R. Synthesis and Liquid Chromatographic Evaluation of Some Chiral Derivatizing Agents for Resolution of Amine Enantiomers. *Anal. Chem.* **1984**, *56* (6), 958–961.
 62. Hengstmann, J.H.; Falkner, F.C.; Watson, J.T.; Oates, J. Quantitative Determination of Guanethidine and Other Guanido-Containing Drugs in Biological Fluids by Gas Chromatography with Flame Ionization Detection and Multiple Ion Detection. *Anal. Chem.* **1974**, *46* (1), 34–39.
 63. Debowski, J.; Sybilska, D.; Jurcyak, J. β -Cyclodextrin as a Chiral Component of the Mobile Phase for Separation of Mandelic Acid Enantiomers in Reversed Phase Systems of High Performance Liquid Chromatography. *J. Chromatogr.* **1982**, *237* (3), 303–306.
 64. Zukowski, J.; Sybilska, D.; Bojarski, J.; Szejtli, J. Resolution of Chiral Barbiturates into Enantiomers by Reversed-Phase High-Performance Liquid Chromatography Using Methylated Beta-Cyclodextrins. *J. Chromatogr.* **1988**, *463* (3), 381–390.

65. Magurie, J.H. Some Structural Requirements of Resolution of Hydantoin Enantiomers with a β -cyclodextrin Liquid Chromatograph Column. *J. Chromatogr.* **1987**, 387 (5), 453–457.
66. Sybilska, D.; Zukowski, J.; Bojarski, J. Resolution of Mephentoin and Some Chiral Barbiturates into Enantiomers by Reversed Phase High Performance Liquid Chromatography via β -cyclodextrin Inclusion Complexes. *J. Liq. Chromatogr.* **1986**, 9 (6), 591–606.
67. Debowski, J.; Sybilska, D.; Jurcyak, J. The Resolution of Some Chiral Compounds in Reversed Phase High Performance Liquid Chromatography by Means of β -cyclodextrin Inclusion Complexes. *Chromatographia* **1982**, 16 (2), 198–200.
68. LePage, J.N.; Lindner, W.; Davies, G.; Seity, D.E.; Karger, B.L. Resolution of the Optical Isomers of Dansyl Amino Acids by Reversed Phase Liquid Chromatography with Optically Active Metal Chelate Additives. *Anal. Chem.* **1979**, 51 (3), 433–435.
69. Horikawa, R.; Sakamoto, H.; Tanimura, T. Separation of α -Hydroxy Acid Enantiomers by High Performance Liquid Chromatography Using Copper (II)-L-Amino Acid Eluent. *J. Liq. Chromatogr.* **1986**, 9 (4), 537–549.
70. Yainazaki, S.; Takeuchi, T.; Tanimura, T. Direct Enantiomeric Separation of Norepinephrine and its Analogues by High Performance Liquid Chromatography. *J. Liq. Chromatogr.* **1989**, 12 (7), 2239–2248.
71. Pettersson, C.; Schill, G. Separation of Enantiomeric Amines by Ion-Pair Chromatography. *J. Chromatogr.* **1981**, 204 (1), 179–183.
72. Pettersson, C.; Schill, G. Chiral Resolution of Amino-alcohols by Ion-Pair Chromatography. *Chromatographia* **1982**, 16 (2), 192–197.
73. Shibata, T.; Mori, K.; Okamoto, Y. Polysaccharide Phases. In *Chiral Separations by HPLC*; Krstulovic, A.M., Ed.; Ellis Horwood, Ltd.: Chichester, U.K., 1989; 336–398.
74. Allenmark, S.; Bomgren, B.; Boren, H. Direct Liquid Chromatographic Separation of Enantiomers on Immobilized Protein Stationary Phases. IV. Molecular Interaction Forces and Retention Behaviour in Chromatography on Bovine Serum Albumin as a Stationary Phase. *J. Chromatogr.* **1984**, 316 (12), 617–624.
75. Shinbo, T.; Yamaguchi, T.; Nishimura, K.; Sugiura, M. Chromatographic Separation of Racemic Amino Acids by Use of Chiral Crown Ether-Coated Reversed-Phase Packings. *J. Chromatogr.* **1987**, 405 (9), 145–153.
76. Pirkle, W.H.; Sikkenga, D.L. Resolution of Optical Isomers by Liquid Chromatography. *J. Chromatogr.* **1976**, 123 (3), 400–406.
77. Pirkle, W.H.; Dazzen, R. Reciprocity in Chiral Recognition: Comparison of Several Chiral Stationary Phases. *J. Chromatogr.* **1987**, 404 (1), 107–111.
78. March, J. *Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*; Wiley-Interscience: New York, 1985; 82–140.
79. Brown, C. *Chirality in Drug Design and Synthesis*; Academic Press: San Diego, 1990; 243.
80. Wainer, I.W.; Drayer, D.E. *Drug Stereochemistry Analytical Methods and Pharmacology*; Marcel Dekker, Inc.: New York, 1988; 276.

ISOLATORS FOR PHARMACEUTICAL APPLICATIONS

Gordon J. Farquharson

Bovis Lend Lease Pharmaceutical, United Kingdom

INTRODUCTION

The use of isolators in research and manufacturing in the health care and life science industries continues to develop rapidly. An overview shows many new applications being conceived, designed, constructed, and set into operation to satisfy many different processing needs. These applications frequently fall outside the accepted concepts and practice found in human scale rooms. It is only in the last 2 or 3 years that the technology has been recognized within cGMP guidelines and regulations. This article attempts to describe the practical experience gained from consultancy and design of isolators from the early 1970s up to and including state of the art applications developed today. For clarity, an isolator can be defined as:

A device creating a small enclosed controlled or clean classified environment in which a process or activity can be placed with a high degree of assurance that effective segregation will be maintained between the enclosed environment, its surroundings, and any personnel involved with the process or manipulation. Isolators can be closed or open designs, and may be maintained at positive or negative pressure to their surroundings.

The following list summarises some essential features that help to expand on the definition as it is applied in the life science and pharmaceutical industry context:

- An isolator is an enclosed controlled environment of minimum volume.
- Isolators segregate people from processes.
- Isolators can contain processes hazardous to the surroundings, processes that are at risk from the surroundings, or in some cases activities where both types of risk coexist.
- Isolator walls or envelopes may be rigid or flexible; typical primary materials are stainless steel and glass, and polyvinyl chloride (PVC) flexible sheet welded to form an enclosure.
- Isolators are internally pressurized with air or inert gas to help achieve the required segregation between inside and outside. Pressurization can be positive or negative.

- Air filtration through High Efficiency Particulate Air (HEPA) or Ultra Low Particulate Air (ULPA) filters (9) is used to control the quality of air entering, leaving, and recirculating. For some applications in which very small inert gas quantities are used, nitrogen, for example, can be delivered through sterilizing-grade membrane filters.
- Isolators are frequently connected intimately to items of process equipment to provide an effective locally controlled environment.

The principal industry drivers generating the interest in isolators in the last few years have been focused on improvements in process integrity. This includes operator protection from potent and hazardous materials and, in the case of sterile products manufacturing, to reduce the potential of contaminated nonsterile units from being produced by a specific process. There are some circumstances in which reductions in occupied space and operational cost savings have been essential objectives.

TECHNICAL GUIDELINES AND STANDARDS

Although guidelines, such as the UK Pharmaceutical Isolator Guideline (1), and standards for microbiological safety cabinets (2, 3) and those for flexible-film isolators (4), satisfy part of the need for effective standards, further support is needed. During 2001, the ISO Technical Committee 209 should publish EN/ISO (DIS) 14644-7 (5). This standard will be entitled "Enhanced Clean Devices." Although it is not pharmaceutical industry specific, it will contain much excellent basic good practice guidance. 2001 should also see publication of PDA's Monograph "Design and Validation of Isolator Systems for the Manufacturing and Testing of Health Care products." Useful sources of reference can also be obtained from the nuclear industry. ISO 10648 "Containment enclosures" contains some valuable sections (6, 7).

For isolators, the barrier between the critical controlled environment and its surroundings is created by a single element, rather than the multilayer protection approach used in cleanroom technology. It is essential, therefore, to realize that the engineering solution, integrity, and

reliability of the final operational isolator will have a direct impact not only on the enclosed process, but also on the quality required of the surrounding environment. This is particularly so when an isolator is used to contain a vulnerable aseptic process. This is usually be the most critical task for isolators, and the one that is most difficult to prove.

The design and construction of isolators should be carried out in an appropriate quality-assured way because the devices are frequently complex and require a high level of documentation to comply with both safety and good quality requirements. ISO 9000 compliant or similar quality assurance systems provide an appropriate management environment in which to design and build systems destined for quality or safety critical applications.

HEALTH CARE INDUSTRY REGULATORY DEMANDS AND EXPECTATIONS

Many operational applications are now in place, particularly in Europe. They can be found in manufacturing, R&D, and QC sectors. Experience continues to develop, and the users are at the forefront of the knowledge. Regulatory guidance can now be found in the European Union Guide to Good Manufacturing Practice (EU GMP) and will shortly appear in "Pharmaceutical Inspection Co-operation Scheme" (PICS) inspection guidelines. It is the responsibility of the isolator user, together with the designer or supplier, to effectively demonstrate an appropriate level of system integrity and performance. Regulators primarily hope to see that the use of isolators has been targeted at improvements in sterility or other attributes of quality assurance, as well as operator health and safety, and environmental protection. Such regulators may be less impressed by issues relating to reduction of unit manufacturing cost or amount of capital employed.

At present, the pharmaceutical industry regulatory requirements refer to isolators specifically in the context of the manufacture of sterile products. There is no reference to their role in broader areas of crosscontamination and operator safety control. Within Europe, the current EU GMP clearly states that isolators might produce improvements in sterility assurance of sterile products, and that aseptic processing manufacturing isolators should be placed in at least a Grade D surrounding environment. The Food and Drug Administration (FDA) requirements are less well defined, but it is likely that in equivalent circumstances, they would like to see an isolator located in a class 100,000 or M6.5 environment "In Operation."

As far as the configuration and performance of isolators is concerned, the pharmaceutical regulators have not made specific demands in their documentation. However, some of the main issues that have been identified by U.S. and European investigators or inspectors include:

- Concern that the use of isolators engenders a false sense of security, and that cGMP standards might be abandoned
- Concern over the effects of vibration caused by the process and the critical environment being physically connected
- Integrity of glove and half-suit systems
- Effective leak testing regimes that should be established
- Effectiveness of the physical cleaning of isolators applied in conjunction with gaseous or aerosol disinfection systems
- Evidence of good ergonomic design
- A well-defined and implemented personnel intervention policy, including intervention recording
- Tailored process simulation programs for aseptic processing
- Inappropriate sterility assurance level claims

APPLICATIONS

Isolator systems can be used for quality-critical, safety-critical, or combined applications. The examples given here are not exhaustive but focus on some of the most important applications, and clearly illustrate the broad range of devices that are created to satisfy particular needs.

Sterility Testing

For many years, isolators have provided a valuable tool for providing very clean conditions in the microbiological laboratory for testing the sterility of the end product. Both flexible-film and rigid-wall devices have been successfully utilized. These isolators are used to carry out manipulations with a growth-promotion medium. Any failures in the security of the isolator are likely to manifest themselves as growth in the culture medium. Such growths would be deemed false positives, and would result in a requirement to investigate the source of contamination. The security and effectiveness of isolators, operating in conjunction with sanitization techniques and transfer systems such as interlocking transfer ports, have been demonstrated to provide a more secure system than the traditional cleanrooms for this type of analytical work.

Sterility testing requires a strict control of microbial contamination challenge from outside the controlled environment, but not of the particulate contamination liberated by the process itself. Hence, a positive-pressure isolator in a controlled environment, using nonunidirectional airflow, is satisfactory. Flexible-film devices using half-suit manipulation techniques are frequently used in these applications. These isolators are usually configured as "closed" isolators because they don't have continuous-process discharges from the contained

volume. Fig. 1 shows an example of an isolator used in sterility testing.

Subdivision and Dispensing of Potent Compounds

As pharmaceutical products contain increasingly potent active constituents and as health, safety, and environmental protection issues increase in importance, isolators have been developed in many shapes and forms to permit

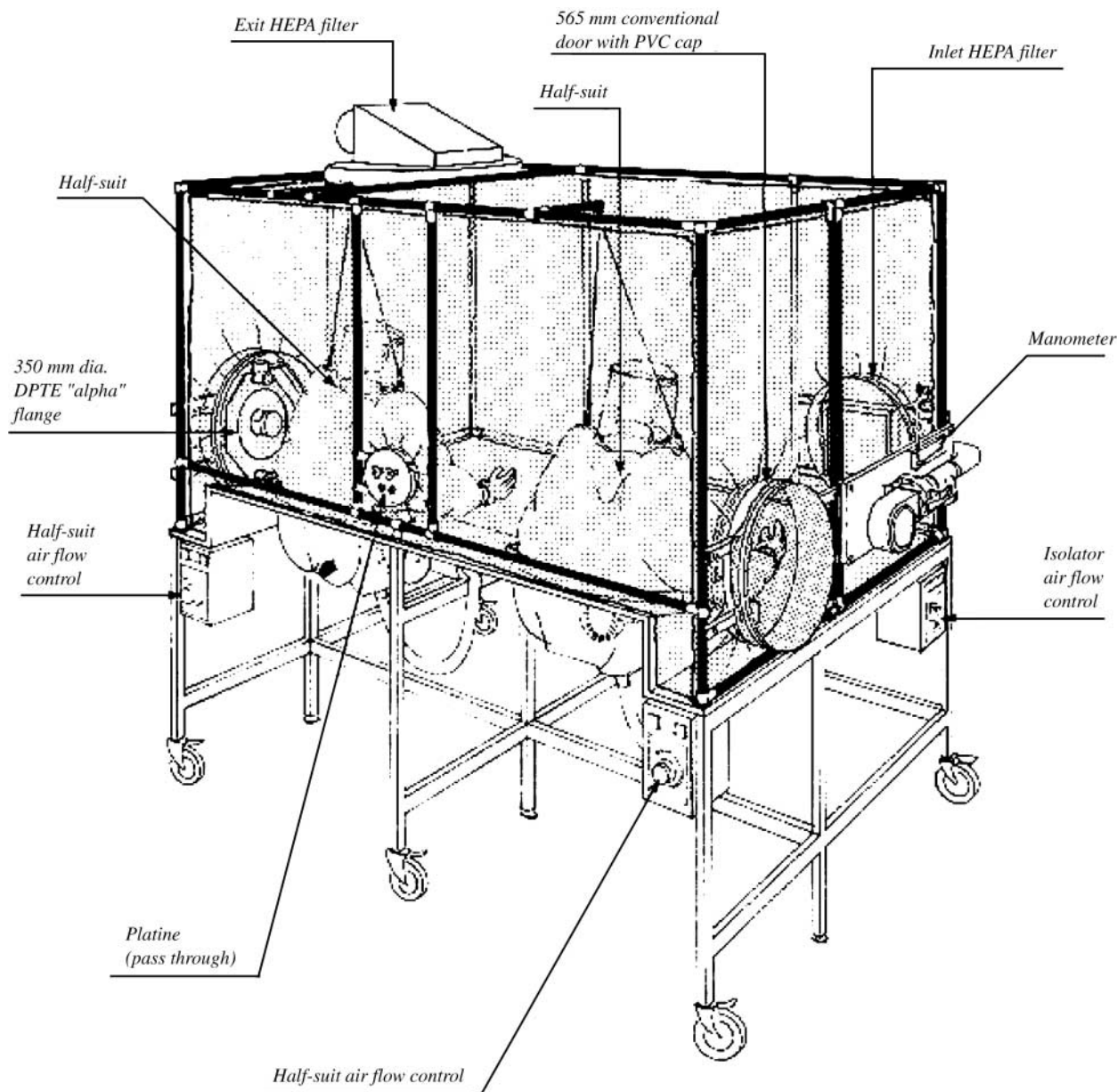


Fig. 1 Flexible-film isolator with half-suit used for sterility testing. (Diagram courtesy of La Calhene.)

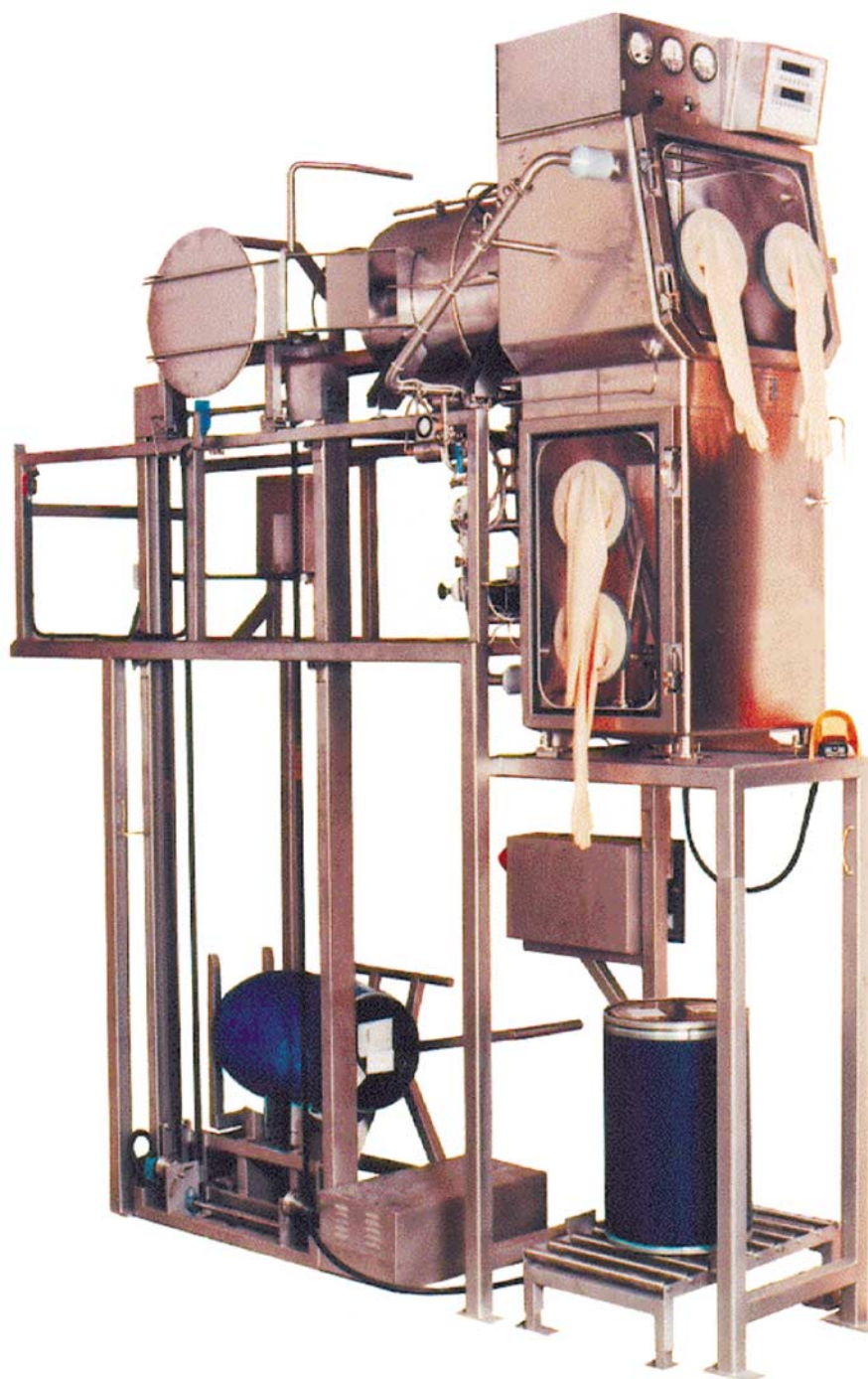


Fig. 2 Potent drug-dispensing isolator. (Photograph courtesy of TPC Microflow Ltd.)

the safe weighing and subdivision of highly active compounds. The most sophisticated applications, such as the subdivision of bulk sterile active compounds, require that the isolator maintain aseptic processing conditions internally at the same time as satisfying the

safety requirements. Fig. 2 shows an isolator device designed to allow a keg of potent raw material to be introduced into an isolator environment, and be securely subdivided into lots suitable for a subsequent formulation batch process. The device includes systems for mechanical



Fig. 3 Specialized isolator for potent drug processing and handling. (Photograph courtesy of TPC Microflow Ltd.)

handling of the kegs, and for washing their exterior to decontaminate them after completion of the manipulation.

Powder Processing Systems

A natural extension from the simple handling of potent compounds is to use barrier technology to provide highly secure mechanisms for processing and more complex manipulation of powders. Isolators of this type and configuration have the objectives of operator and environmental protection as well as the provision of a secure clean and aseptic environment around the process.

The example shown in Fig. 3 is a system used for the drying and handling of bulk powder. It provides a clean classified environment for handling bulk powder, a containment of the potent compound, and a nitrogen environment to allow the safe use of flammable solvents. It is a rigid positive-pressure device located in a EU GMP Grade D cleanroom. The cleanliness of the internal nitrogen environment is maintained using nonunidirectional airflow. Manipulations are achieved using a combination of glove and half-suit systems.

Small-Scale Manipulations

Many isolator applications at the clinical trial scale of manufacturing are based on the same scale of technology used for sterility testing. The aseptic dispensing of pharmaceutical products in hospital pharmacies is also carried out on this scale. Such manufacturing is not carried out on a continuous basis, but in relatively small batches

that can be transferred from the isolator with the help of one of the more secure transfer systems. In this type of application, one isolator is being used to dispense a variety of products or several isolators are used for separate tasks.

In the application shown in Fig. 4, a group of isolators are placed in an EC Grade C working environment (approximately equivalent to ISO 7 or Class 10,000 "At Rest"). The isolators are configured to provide an aseptic environment interfacing with a depyrogenating oven for handling and holding dry-heat-sterilized components. Additional individual isolators are provided in which separate formulation and/or filling operations can be undertaken. Materials are moved from one isolator to another using closed containers that dock with alpha/beta docking ports mounted in the sidewall and floor of the isolators. Internal cleanliness is maintained through positive pressurization and the supply of double HEPA-filtered unidirectional airflow to the critical process zones within the isolators.

LARGE-SCALE ASEPTIC PRODUCTION

Isolators are now used in industrial scale aseptic processing for both formulation and filling. Fig. 5 shows an example of a rigid interconnected network of isolators providing a complete component handling and aseptic filling capability in an integrated line. The internal control of the environment is achieved by a combination of nonunidirectional and unidirectional airflow within positively pressurized cells. The individual cells are

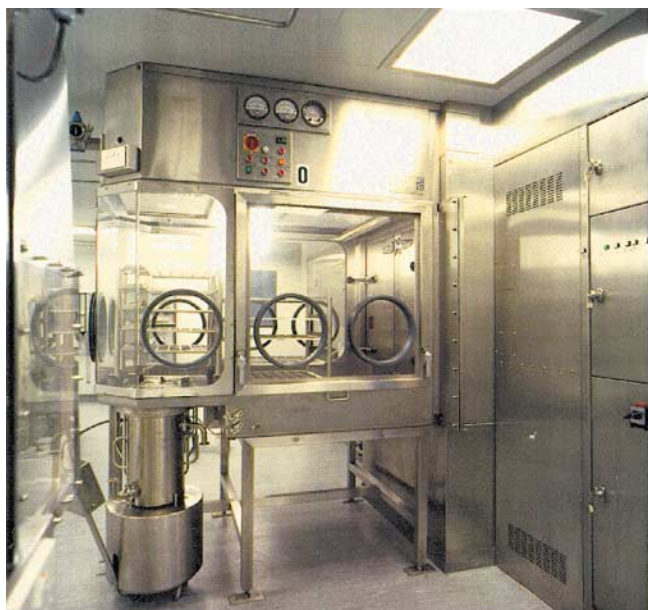


Fig. 4 Isolators used in clinical trials and specials manufacturing (the gloves have been removed for photographic clarity). (Photograph courtesy of Boots Contract Manufacturing and Bovis Tanvec Ltd.)

separated by “airlock” flap valves. The product flow through to aseptic filling operations is largely automatic, with gloves provided for specific manual interventions. The complete line is placed in a room in which the environment is controlled but unclassified in terms of cleanroom standards. Hydrogen peroxide vapor is used for the surface sterilization of the isolator network, and for the continuous surface disinfection of containers of product components that enter the system.

Technical Considerations for the Design, Manufacture, and Testing of Isolators

The susceptibility of the process to contamination is critical in terms of risk assessment. In aseptic processing, for example, open processes are at far greater risk than closed processes. Furthermore, systems requiring complex aseptic assembly prior to use are more difficult to manage in a small isolator environment than if cleaning and sterilization-in-place techniques were employed (referred to as *CIP* and *SIP*). It is essential, therefore, to evaluate all the process steps, including equipment transfers, assembly manipulations, and the processing activity itself. Effective documentation of such analysis is good practice in a validated operational scenario.

Barrier-integrity characteristics

Complex isolators are unlikely to be leak free. It is obvious that the better the barrier integrity is, the less will be the

opportunity for contamination to be transferred across the isolator wall. The integrity is influenced by the basic materials of construction, design effectiveness, damage caused by cleaning, resistance to process chemicals, and robustness. The first fundamental option to consider is a rigid or a soft-wall construction. The former normally utilizes a combination of stainless steel and glass or Perspex windows; the latter uses a welded transparent PVC envelope connected to a stainless steel floor tray or machine bedplate. After choosing either a rigid or soft-wall solution, it is most important to ensure that an effective mechanism is selected to measure the leakage-air tightness of the isolator. Such tests should form part of the construction acceptance testing of a device, and should become part of routine operational leak testing. A pressure-hold test is the simplest method of determining that the complete assembly satisfies a set of acceptance criteria. Provided that the isolator, including its internal and external air systems, is configured in an appropriate way, simple isolating valves can be used to isolate the isolator, thus allowing convenient leakage-rate tests to be carried out. Such tests could ultimately be carried out on a batch basis if required. The most commonly used test methods are:

- Evaluation of pressure decay rate
- Leakage-rate measurement at constant pressure
- Tracer gas leak-rate detection
- Tracer particle transfer measurement (sometimes called a leak-induction test)



Fig. 5 A specialized network of isolators used for large-scale industrial syringe filling; process flow is from left to right. (Photograph courtesy of Medeva Ltd.)

Table 1 illustrates a range of four “Leakage Classes” that will be included in the “Enhanced clean devices” standard, EN ISO (CD) 14644-7. These classes are based on levels of leakage defined in the ISO 10648 (6,7) nuclear containment standards.

Manipulation technique

The manipulation technique influences the overall integrity of an isolator system, as in virtually all cases, the manipulation device presents a potential breach to the barrier of the isolator. Additionally, where the

manipulation involves placing part of the human body within a specialised system component, such as gloves or a half-suit, a greater potential exists for process or product contamination than would occur with tong manipulators, remote manipulators, or robotics devices. Utilizing a simple pressure-hold or leakage test on a glove port and glove before and after a process work session can achieve a high level of glove integrity assurance. Such a technique is virtually impossible with half-suit applications due to their size and construction. Strategic inspections and pressure tests for integrity at

Table 1 Isolator leak rates (tentative levels)

Leakage class reference	Leak rate enclosure air change/hour	Leak rate: percentage enclosure volume/hour %
1	$<5 \times 10^{-4}$ (<0.0005)	<0.05
2	$<2.5 \times 10^{-3}$ (<0.0025)	<0.25
3	$<10^{-2}$ (<0.1)	<10
4	$<10^{-1}$ (<1.0)	<100

less frequent intervals are required for such devices. Alternative manipulation methods using tongs and remote manipulators (mainly for handling radioactive substances) can be of advantage in isolators, but currently are rarely found in pharmaceutical applications. However, it should be noted that tong manipulators, for example, have their own special problems and use a gaiter system to ensure air pressure integrity around the rotation and sliding gimbals. This gaiter is akin to the glove gauntlet, and its integrity and risk of failure are just as important.

Manipulations should always be minimised. An event report should be made, and a preprepared action plan must be implemented should damage occur to a glove. The selection of glove and gauntlet materials is most important. Half-suit and sleeve/glove systems use a technique where the glove can be detached from the sleeve by way of a specialized fitting in the wrist region. Some of these devices have the capability of allowing glove change while the system is in use. Gauntlet or one-piece systems are generally more durable, but do not have the same change in use capability. A balance has to be maintained between durability and "feel." Typical available materials are:

- Latex (natural rubber) allows great dexterity; if thick enough, these can be wear and abrasion resistant. It is chemically resistant to many mild chemicals, detergents, and disinfectants.
- Neoprene (synthetic) has high flexibility and allows good dexterity. It has low tensile strength, and is chemically resistant to materials with the exception of oxidizing agents such as hydrogen peroxide.
- Nitrile (synthetic copolymer) has poor flexibility, but is strong and highly chemical resistant.
- PVC (synthetic polymer) is strong and chemically resistant to materials except ketones and aromatic hydrocarbons; it is inflexible and subject to tearing.
- Urethanes have good abrasion, chemical resistance, and good tensile strength, but are not suitable for high temperatures.
- Laminated polymers exhibit high strength and selective resistance to chemicals; they are subject to failure when used with sharp objects.

A variation of the pressure-hold test can be used to test the integrity of manipulation gloves in situ. This provides a convenient way of determining the integrity of the glove without entering the controlled environment or unnecessarily changing an expensive commodity. An alternative, provided by a particular vendor, involves inert gas purging of a glove volume, followed by measurement of the build-up of oxygen that diffuses through glove pinhole leaks.

Transfer techniques

The transfer of materials into and out of an isolator represents the most likely and most common source of loss of internal environmental integrity. The more secure the transfer system, the less demanding is the surrounding environment. The simplest devices, such as single doors, present very little ability to separate the external from the internal environment. In fact, in these applications, the only facet of the device's performance that provides any protection is outward airflow when the door or cover is open. Security can be improved by a double-door pass-through hatch. The performance and effectiveness of such a device can be improved by introducing mechanical or electromechanical interlocking of the opposing doors. Furthermore, adding positive ventilation of the airlock space to dilute and remove contamination that may enter when the external door is open adds security to this form of transfer. The most secure techniques include a direct process connection to the isolator, interlocked docking port systems (often called alpha/beta systems), and airflow protected tunnels for continuous component discharge. Although no specific tests or standards exist to define the performance of such devices, tests can be adopted from other applications. Containment tests such as those used for open-fronted microbiological safety cabinets can be effectively used to determine a protection factor for airflow-protected product-discharge tunnels. In the case of alpha/beta interlocked docking port systems, certain manufacturers have developed and applied particulate and microbiological challenge tests to determine the effectiveness or a protection factor of these devices. This quantifies the segregation achieved by such a device in operation. Such tests may become the basis of type or performance testing and subsequent selection of specialized transfer devices. A typical alpha/beta port is illustrated in Fig. 6.

Internal pressurization

Internal pressure within an isolator clearly has a major influence on the ability of the isolator to exclude the external environment. The level of the pressurization should also be considered in relation to its ability to withstand the piston effect of rapid glove movement, and whether or not the internal isolator air system should achieve a specific outflow of air in the event of partial or total glove loss. The integrity and performance of pressurisation equally apply to negative-pressure systems. However, negative-pressure systems, used for clean and aseptic processing, are more likely to require a higher class of surrounding environment than that required for equivalent positive pressure systems. The glove piston

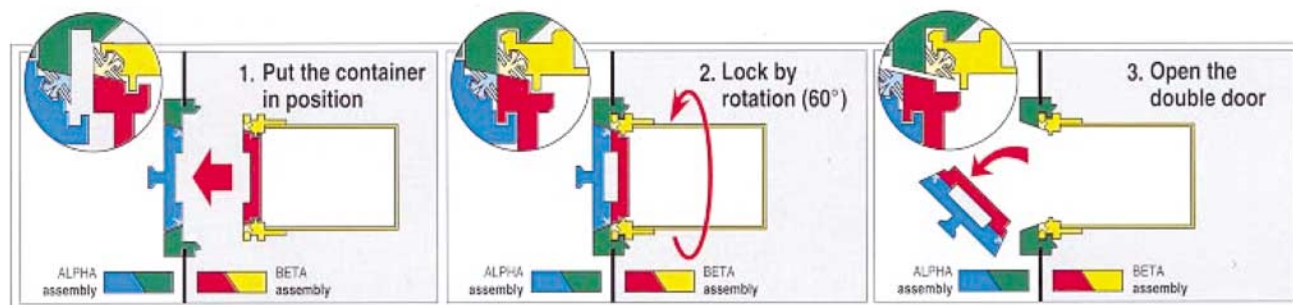
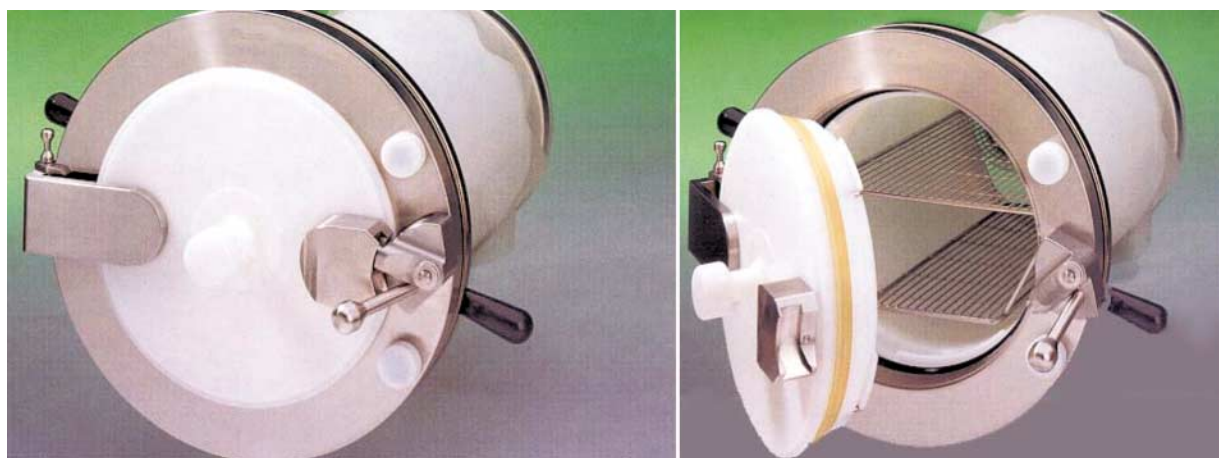


Fig. 6 A typical alpha/beta port. (Illustration of DPTE courtesy of La Calhene.)

effect is relative volume related (i.e., the volume displaced by the glove compared to the volume of the isolator). As a rule of thumb, devices with a pressure of 15–25 Pa compared to the surrounding atmosphere, are likely to be less secure (due to the piston effect) than devices with a pressure of 50–80 Pa compared to the surrounding area. When lost glove airflow protection is required, air velocities of 0.5–0.7 m/s should be considered. If extremely high velocities above these figures occur, reentrainment of external contamination due to high turbulence is a distinct possibility. In testing the effectiveness of the pressurisation, it is necessary to challenge not only the steady state conditions but also the transient conditions. It is therefore expected that a series of tests be carried out to investigate start and stop modes of the isolator control system, the influence of glove or other device manipulation, and of course, the interaction of the process itself. The most demanding processes in aseptic pharmaceutical applications are usually continuous depyrogenation tunnels, where the air leakage into the tunnel due to isolator overpressure, varies with time, and may well vary with different machine settings for various sizes. In this type of application, it is absolutely critical that the qualification of the isolator environment

is carried out in conjunction with all states of operation. Similarly, it is essential that the depyrogenation process in the tunnel is fully qualified with all states of the isolator.

Airflow configuration

As with cleanroom applications, isolators can use unidirectional or non-unidirectional airflow regimes, or in larger systems a combination of both. Isolators are very different from cleanrooms in that generally a physical barrier separates critical from noncritical zones rather than the use of managed airflow. It is possible therefore in isolators to obtain the necessary levels of environmental cleanliness with very low air velocities. Unidirectional airflow systems designed to achieve class 100/M3.5/ISO 5 cleanliness can be achieved with velocities as low as 0.10 m/s. When the velocity is set at such low levels, it is necessary to carefully evaluate the influence of heat sources and process disturbance. Having defined the airflow characteristics required for the isolator, it is necessary to determine both the airflow rate using flow anemometers and the uniformity of the airflow in the case of unidirectional airflow systems. The latter is best achieved using flow visualization smoke-tracing

techniques. This is particularly important if an isolator environment itself contains a mixture of unidirectional and turbulent and/or conventional flow zones.

Filtered-air exchange rate in nonunidirectional flow systems should be based on the rate required to dilute internally generated contamination. The contamination decay rate of an enclosed volume can be used as a measure of effectiveness of the air-movement system in an isolator. This can be helpful for assessing the degassing rate at the conclusion of a gaseous sanitisation or sterilizing procedure. The decay rate can be measured using artificial aerosol generation combined with measuring the decay profile broadly in accordance with the method set out in IES recommended practice 006.2. Alternative gas-decay methods can be adapted from tracer-gas methods used to prove the effectiveness of ventilation systems.

HEPA filtration

The provision and location of air filters and the filtration integrity in isolators are as important as in clean or containment room technology. The filter and its installation must be designed with the utmost care, and special consideration should be given to the effect of vibration transferred from the isolator mechanical systems or the process. Final filters should be placed as close as possible to the critical zone. However, this sometimes makes the task of fitting the filters difficult due to restricted space and access. Maintenance convenience can be improved by locating filters close to, but away from the critical zone, in carefully engineered housings. When this technique is used, the duct between the filter and the isolator should be constructed of non shedding materials. Natural or artificial aerosol challenge tests are the appropriate way to test in situ final HEPA or ULPA filters and existing cleanroom oriented standards, and requirements are directly relevant to isolator applications.

Airborne cleanliness classification

Classification of the working environment by measuring the particulate concentration is normally carried out in accordance with the requirements stated within the accepted airborne-particle classification standards. If the isolator working volume is extremely small, it may be necessary to increase the number of test locations from the single point determined by using the formula within the standards. Two, three, or more locations focused on the critical points of process or product exposure. When continuous or cyclic automatic particle monitoring is being considered, care should be taken to ensure that the volume of air taken as a sample does not adversely influence the isolator pressurization.

Open aperture integrity

Open aperture protection is an important feature of some isolators. In addition to designing and testing for isolator pressurisation, some isolators need to be engineered to maintain segregation in case of partial or total glove loss. This is particularly the case when biological or chemical hazards are present internally, and need to be contained for safety reasons. Furthermore, in many production-scale isolator networks, particularly those filling parenteral containers, it is an advantage to pass the filled closed containers out of the isolator continuously via an airflow protected tunnel. In such cases, there is a protecting air velocity, entering or leaving the isolator, to contain a hazard or minimize the opportunity of internal contamination, respectively. The effectiveness of this inrush or outrush of air (protection factor) can be quantified by a biological or aerosol challenge test derived from adaptation of the method defined in British Standard 5726 (2) or the U.S. Standard NSF 49 (3). Both these standards relate to microbiological safety cabinets where the test is specified for the purposes of quantifying a containment factor. The containment test challenges the aperture with a test aerosol, and measures the quantity that escapes through the opening. The ratio of escape to generated quantity is used to calculate a containment factor. This test method can also be effectively deployed for testing and demonstrating the effectiveness of any other aperture inward or outward and is used to protect and segregate the internal from the external environment. In the classic case, there is a continuous discharge of filled containers across a dead plate, leaving an aseptic processing environment. Here it is important to determine that there is no turbulence causing induction of external contamination into the critical aseptic zone.

CLEANING AND SANITIZATION OF ISOLATORS

Effective cleaning and biodecontamination of the internal surfaces of isolators and their intimate product contact parts is of greatest importance, particularly for aseptic processing. The effectiveness and repeatability of the sanitisation method also has an impact on the quality of the surrounding environment required. This is particularly the case if it is anticipated that batch and machine format changes are carried out with the isolator open. If this procedure is adopted, it is necessary first to minimize the introduction of room contamination into the open isolator, and then to apply an effective cleaning procedure. If some contamination has entered the device, it is essential to use a repeatable and effective sanitization or surface sterilization

method. The requirement to clean and disinfect would be lessened if changeover were achieved with a closed isolator. The least effective processes, such as surface swabbing and aerosol spraying with disinfectant, are unlikely to satisfy the requirements of a repeatable process of high efficacy. However, the use of highly controlled gaseous-phase processes, using materials such as formaldehyde, hydrogen peroxide, peracetic acid, and chlorine dioxide, can be very effective. These gaseous-phase processes are generally intolerant of soiling deposits on the surface. Therefore, the methods must be deployed in conjunction with effective physical cleaning.

DESIGN CONSIDERATIONS FOR THE SURROUNDING ENVIRONMENT

Finally, having taken all the above issues into account, and determined the type and nature of the isolator to be used and the quality of the surrounding environment, it is important to consider some of the broader issues relating to the design of the surrounding environment. It is essential to thoroughly consider all the attributes of the facility in which the isolator is placed, as this can have a major influence on the product. The facility in which the isolator is placed must provide a clear departmental separation from less critical adjacent activities. By layout and configuration, it must achieve the following:

- Provide appropriate physical security for the operation
- Control and manage access of personnel and materials
- Ensure the required background environment is maintained
- Provide the utilities required by the isolator

REFERENCES

1. Isolators for Pharmaceutical Applications. In *The UK Isolator Group*; ISBN 0 11 701829 5 Her Majesty's Stationery Office: London, 1994.
2. Microbiological Safety Cabinets. BS5926: 1992 & BS EN 12469: 2000.
3. NSF 49 Microbiological Safety Cabinets.
4. ACDP UK, Guidance on the Use, Testing, and Maintenance of Laboratory and Animal Flexible Film Isolators, Dec. 1995.
5. Clean Rooms and Associated Controlled Environments. A family of new standards produced by ISO and CEN. EN/ISO, 14644-1–14644-8, (Progressive Publication from May 1999).
6. Refer to ISO 10648 *Containment Enclosures—Part 1: Design Principles*; 1997.
7. Refer to ISO 10648 *Containment Enclosures—Part 2: Classification According to Leak Tightness and Associated Checking Methods*; 1994.
8. U.S. Fed. Standard 209E, Airborne Particulate Cleanliness Classes in Clean Rooms and Clean Zones.
9. Refer to *Air Filtration*, EN 1822 parts 1–5.
10. Air Cleanliness in Clean Rooms. JIS 9920 **1989**.

IONTOPHORESIS

J. Bradley Phipps
Erik R. Scott
J. Richard Gyory
Rama V. Padmanabhan

ALZA Corporation, Mountain View, California, U.S.A.

INTRODUCTION

Iontophoresis is a method of transferring substances to and from the body for therapeutic or diagnostic purposes by applying an electric potential to enhance their movement across biological membranes. The most common applications of iontophoresis involve the delivery of therapeutic substances across the skin, though there are numerous examples of the use of iontophoresis to treat conditions of the eye, ear, nose, and mouth. Iontophoresis can also be used to remove substances (e.g., glucose) from the body. A technique known as microiontophoresis employs a small capillary probe to study cellular function by releasing precise quantities of active substances.

Banga (1) has stated that Veratti made the earliest references to the use of an electric potential to enhance the penetration of charged substances into tissues in the year 1747. Stephane Luduc (2) is generally recognized as the most important early researcher in the field because of his comprehensive studies of iontophoresis described in the 1907 publication "Les Ions et les Medications Ioniques."

Today, iontophoresis of drugs across skin or mucosal membranes is a noninvasive (needleless) method where the rate of delivery is primarily determined by the magnitude of the applied current, making patterned and on-demand delivery possible. Commercially available devices are typically bench-top systems with discrete patches connected to a power supply by electrical cables. However, due to recent innovations in electronic circuitry and battery technology, iontophoretic treatments can be administered with small, integrated patch-like systems.

The most common therapeutic applications of iontophoresis are topical administration of lidocaine as a local anesthetic and dexamethasone for treatment of local inflammation (Iomed, Inc., Salt Lake City, Utah and Empi Corp., Minneapolis, Minnesota). In addition to these therapeutic uses, iontophoretic systems are commercially available for the diagnosis of cystic fibrosis. For example, the CF Indicator[®] (Scandipharm, Birmingham, Alabama)

and the Webster Sweat Inducer (Wescor, Inc., Logan, Utah) deliver pilocarpine to cause local sweating: sweat is collected and analyzed for high levels of chloride. More recently, iontophoresis has been used to extract glucose from the skin as a means of detecting hypo- and hyperglycemia (Cygnus Corp., Redwood City, California).

Although these examples demonstrate the successful commercial use of iontophoretic technology for topical delivery of compounds, transdermal systems for systemic administration of medicinal agents have not been widely employed. There is, however, heightened interest in this field because of potential medical and economical benefits offered by iontophoretic technology, especially for meeting the delivery challenges posed by new biotechnology compounds.

A schematic diagram of a transdermal iontophoretic system on skin is shown in Fig. 1. A source of electrical energy, such as a battery, supplies electric current to the body through two electrodes. The first electrode, called the donor electrode, delivers the therapeutic agent into the body. The second electrode, called the counter or receptor electrode, closes the electrical circuit. Each electrode contacts an ionically conductive reservoir, normally present as a liquid or hydrogel. The reservoirs are placed on the patient's skin and contain either the drug (for the donor electrode assembly) or a biocompatible electrolyte (for the counter electrode assembly).

TRANSPORT MECHANISM

The term iontophoresis encompasses several processes for moving molecules across the skin: electromigration, electroosmosis, and electroporation. Electromigration is the movement of charged ionic species in response to an applied electric field. This process is usually of primary importance for delivering charged drug species. The movement of charged species within a solvent can induce solvent flow by a process known as electroosmosis. This process is useful for delivering both

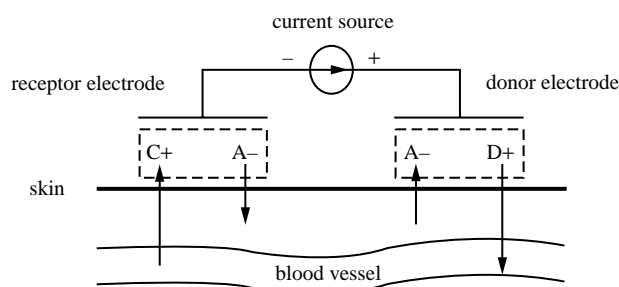


Fig. 1 Schematic diagram of an iontophoretic device on skin indicating the flow of ions in response to an applied voltage.

neutral and charged drug species. Electroporation is the temporary creation of aqueous pores through lipid bilayers by applying high-voltage pulses, typically 50–1000 V, across the bilayers of a biological membrane. This process is useful for delivering large hydrophilic drug species. For any given iontophoretic treatment, one or more of these processes may occur simultaneously and to a varying extent, depending on the magnitude and duration of the applied electric field, the composition of the donor reservoir, and the type of the tissue being treated.

The movement of charges due to electromigration is illustrated in Fig. 1. In this diagram, positively charged drug (D^+) and its counter ion (A^-) are formulated for delivery from the anodic donor reservoir. The cathodic counter reservoir contains biologically acceptable cations (C^+) and anions (A^-). When an electric field is applied, drug ions migrate into the skin and endogenous anions, mostly chloride, migrate from the body into the donor reservoir. Simultaneously, at the cathodic counter electrode, anions migrate from the counter reservoir into the skin, whereas endogenous cations, mostly sodium and potassium, migrate from the body into the counter reservoir. The movement of ions maintains local electroneutrality throughout this process.

For iontophoretic treatments involving placement of the system on the skin, the therapeutic agent in the donor reservoir must cross the outermost layer of the skin, known as the stratum corneum, which is the primary barrier to permeation of substances both into and out of the body. The stratum corneum's excellent barrier properties result from its unique structure: approximately 10–20 layers of flattened, keratin-rich cells cemented together by lipid bilayers composed primarily of ceramides. In general, lipophilic species are capable of traversing the stratum corneum because of their ability to partition into the intercellular lipid region. In contrast, most ionic and polar substances are largely excluded from this region.

At the typical voltages used in iontophoresis (e.g., 2–80 V), the nature and composition of the pathways in the skin remain matters of some debate. However, there is a growing body of evidence that the preferred path for ionic species across the stratum corneum is not spatially homogenous, but rather consists of a distribution of localized regions. These regions include endogenous shunt-like structures across the stratum corneum such as sweat ducts and hair follicles, but may also include pathways not associated with natural shunts. Direct physical measurements of the transport of model permeants through the skin of hairless and nude mice have shown that between 60 and 90% of the overall flux can be explained by such regions (3).

The rate of transport of a charged drug species across a biological membrane is generally modeled by the Nernst–Planck equation. This equation contains terms for diffusion, electromigration, and bulk convection. However, under optimized iontophoretic conditions the electromigration contribution is often much greater than that of the other two, therefore the expression for delivery of an ionic drug species is frequently simplified to include only the electromigration term. The molecular flow of drug is thus related to the electric current, according to Faraday's principle:

$$N = (t_d IM) / (z_d F) \quad (1)$$

where N = total rate of delivery, t_d = transport number (the fraction of charge carried by the drug ion), I = current applied across the skin, M = drug molecular weight, z_d = charge of the drug molecule, and F = Faraday's constant (96,485 coul/Eq).

Many researchers have used in vitro and in vivo studies to demonstrate that the rate of drug delivery is linearly proportional to the applied current over a wide range. Because studies are typically performed with a fixed skin contact area, this proportionality indicates that the transport number is a constant and not dependent on the current density. However, the transport number is unique for each drug, and is a function of the drug's mobility, charge, and concentration, as compared with those of other migrating species. These dependencies are summarized in the following expression:

$$t_d = \frac{\mu_d |z_d| C_d}{\sum_i \mu_i |z_i| C_i} \quad (2)$$

where μ_d , z_d , and C_d are the mobility, charge and molar concentration, respectively, of the drug species, and μ_i , z_i , and C_i are the mobility, charge, and concentration for each mobile ion that competes with the drug for transport across the skin barrier. Competing ions are those in the

formulation that have the same sign of charge as the drug (i.e., competing coions), as well as those ions in the body that have the opposite sign as the drug (competing counterions).

The transport number determines delivery efficiency; that is, the amount of drug delivered per unit charge passed across the skin. Because it is desirable to minimize current (for optimal biocompatibility and battery longevity), it is advantageous to develop a formulation that maximizes t_d . This can be accomplished by maximizing the mobility and concentration of drug species while minimizing, to the greatest extent possible, the mobility and concentration of the competing species. Simple measures to achieve this include incorporating the highest practical concentration of the drug in the formulation, as well as avoiding excipients or impurities that produce mobile coions in the formulation. Even if both of these measures are taken, a t_d value of unity (i.e., all current is carried by the drug ion) is still unattainable in practice, because of endogenous counterions in tissue (e.g., Na^+ and Cl^-). However, the efficiency-lowering effect of competing counterions can be reduced by exploiting the inherent permselectivity of skin. Permselectivity based on charge arises from the principle of Donnan exclusion. The skin has an isoelectric point of about pH 4. Therefore, for a formulation with a pH value below four the skin will have a net positive charge and favor transport of anions; and at a formulation pH > 4 , skin will have a net negative charge and favor transport of cations. Thus, at the physiological pH 7.4, skin should be negatively charged and therefore be *cation selective*.

Efficient iontophoretic delivery depends on the materials and excipients used in the system. Selection of appropriate electrodes and formulations are critical to maximizing the efficiency of iontophoretic treatments. The following sections will summarize key criteria for selection of electrodes and formulations for transdermal applications, however, many of the same principles are applicable to iontophoretic systems used on mucosa or implanted within the body.

ELECTRODES

Electrodes apply the driving force for ion migration and are therefore critical components of the system. They serve as the bridge between the electric circuit and the two reservoirs, and perform both electrical and chemical functions. During iontophoretic therapy, electrodes undergo sustained electrochemical reactions and thus

the migration of reactants is a critical functional consideration. In this aspect, electrodes used in iontophoretic devices are different from those of most other medical electrodes. For example, medical potentiometric electrodes (such as those used for electrocardiograms or electroencephalograms) undergo no net reaction because little or no electrode current is required by the measurement equipment. In other applications such as cardiac pacing or transcutaneous electrical nerve stimulation (TENS), the applied voltage pulses are extremely brief (milliseconds or less) or of alternating polarity so the net quantity of reaction products is not great. For these applications the management of electrochemical reactions is achieved merely by constructing the electrodes from inert materials, for example, gold, platinum, or stainless steel. In contrast, iontophoretic electrodes are usually inherently reactive materials and are chosen for their preferred electrochemical attributes, as is described below. Because of their reactivity, and also because electrodes contact the drug formulation directly and therefore the patient's body indirectly, it can be technically challenging to choose an electrode system that possesses adequate performance while avoiding adverse material and biological interactions during storage and use.

As indicated in Fig. 1, a transdermal iontophoretic system requires that two electrode assemblies contact the patient's skin. The donor electrode (also known as the delivery or active electrode) contacts the drug reservoir. The counter electrode (also known as the return or receptor electrode) contacts the counter reservoir and completes the electrical circuit by providing a path for the current. The two reservoirs are separated from each other and contact skin over a fixed area. The electrodes apply an electric field across the skin by converting electric current supplied by the battery into ionic current moving in the skin and body. In doing so, a Faradaic reaction takes place at the electrode/electrolyte interface. As described previously in this chapter, there is generally a linear dependence of the rate of drug delivery on this current.

The polarity of the donor and counter electrodes depends on the sign of the charge on the species to be delivered. To cause migration of positively charged species from the donor reservoir into the skin, the donor electrode must have a positive polarity (i.e., anode), and the counter electrode must have a negative polarity (i.e., cathode). For negatively charged species, the polarity is reversed, so that the donor electrode is the cathode, and the counter electrode is the anode.

A practical electrode system must meet a variety of performance, compatibility, and physical requirements. When possible, the electrode system should provide for

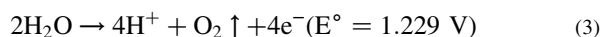
maximum drug delivery efficiency (i.e., the rate of drug delivered per unit current), operate at low voltage (i.e., <1 V), have adequate longevity (e.g., 24 h), and distribute the current evenly over the entire area of skin surface (e.g., 10 cm^2). The electrode will provide maximum drug delivery efficiency if it does not contain or produce any competing ions (i.e., mobile ions of the same charge as the drug ion). For optimum compatibility, the electrodes should be made of materials that are nontoxic and compatible with other formulation components such as the drug, excipients, and matrix material. In addition, they must not generate reaction products that are toxic or that adversely affect drug and excipient stability.

Two classes of electrode are the nonconsumable or inert type and the consumable or sacrificial type. Nonconsumable electrodes are made from nonreactive materials, whereas consumable electrodes are electrochemically active and are structurally altered by the passage of current during treatment.

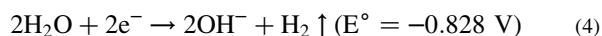
Nonconsumable Electrodes

Early iontophoretic drug delivery systems (IDDS) used materials that were not consumed during use. Commonly used materials included metals such as stainless steel or platinum. Although these nominally inert materials may have long use and storage lives, they also have significant shortcomings.

In accordance with Faraday's Law, the operation of an IDDS requires redox reactions at the electrodes in proportion to the amount of charge passed. For nonconsumable electrodes, contacting an essentially aqueous electrolyte solution, electrolysis of water is the likely redox reaction. Therefore, the reaction at the anode is:



and at the cathode, the most prevalent steady-state reaction is:



Both of these reactions have a number of undesired consequences. The generation of H^+ and OH^- can shift the formulation pH, affecting both delivery efficiency (due to a shift in the skin permselectivity, formation of competitive ions, or alteration of the charge state of the drug) and skin tolerability. The gases that are produced can accumulate on the electrode or skin surface, interfering with the uniformity of current distribution. Furthermore, because these reactions take place at relatively high voltage, there is high power consumption and a risk of electrolytic decomposition of the drug or other excipients.

Some electrodes made from nonnoble metals, such as stainless steel, can release metal ions through direct oxidation at the anode, or indirectly by the creation of a caustic environment at the cathode. For example, nickel and chromium ions were released from a medical-grade steel anode following a few minutes of iontophoresis at $400\text{ }\mu\text{A}/\text{cm}^2$. These ions can be toxic to the body (4).

Chemical methods that mitigate the deleterious effects of unwanted reaction products can be divided into three categories: blocking their migration, neutralizing (e.g., buffering) them, and preventing their formation by addition of sacrificial redox species to the reservoir electrolyte. Migration blocking is achieved by isolating the electrolyte adjacent to the electrode from the drug formulation or from the body by using ion-exchange or size-selective membranes or coatings. For example, an anion-selective coating at a nonconsumable electrode composed of methacrylamidopropyltrimethyl ammonium chloride copolymerized with methyl methacrylate was found to prevent degradation of the drug oxymorphone at electrode potentials up to 800 mV by blocking migration of the drug to the anode surface (5).

Buffering agents can partially compensate for the generation of acid and base, but their duration of efficacy (buffering capacity) is limited by the quantity of buffer species present, and the addition of excess buffer salts can result in ionic competition. Ion-exchange polymers can effectively scavenge, neutralize, and immobilize generated reaction products (6); the advantage over simple buffer salts is that the polymer chains are immobile and therefore noncompeting.

Consumable Electrodes

As described above, approaches have been devised for resolving the disadvantages of nonconsumable electrodes. Although many of these schemes are simple in principle, their practical implementation can be difficult because of various physical, chemical, and biological requirements. A simpler alternative is to use consumable electrodes. Also known as sacrificial electrodes, they are altered by electrochemical reaction during operation. Appropriate consumable electrodes will have redox reactions that take place at low potentials, thus avoiding parasitic reactions (e.g., electrolysis of water, drug, or excipients) at the electrode surface. Also, the reactants and products of the redox reaction must meet formulation and biological compatibility requirements.

A sacrificial electrode has a finite operational lifetime, or capacity, defined as the amount of charge that can be passed before the reactants are effectively depleted. The

capacity of an electrode can be empirically determined, or can be calculated from the following equation:

$$Q = \frac{umnF}{3.6M} \quad (5)$$

where Q = capacity (mAh), u = utilization (fraction of reacting species available for reaction), m = mass of reactant (g), n = number of equivalents of charge per mole of reactant (eq/mol), F = Faraday's constant (96,485 C/eq), M = molecular weight of the reactant (g/mol), and the constant 3.6 is a conversion factor (C/mAh).

If depletion results in an open circuit, the electrode will cease to function. Otherwise if an electrically conductive pathway remains, the electrode will continue to function under suboptimal conditions (i.e., at a higher voltage where other redox reactions, such as water electrolysis, take place).

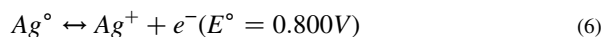
Use of materials and structures that maximize utilization of reactants are preferred, as this allows the electrodes to be thin while making the most economical use of the consumable electrode. Factors that limit utilization include a tendency for passivation of the active component (by formation of a uniform, insoluble, nonconductive product over the active surface), and formation of "islands" (electrical isolation of one or more active portions by nonuniform current distribution).

No single consumable electrode is ideal for all iontophoretic applications. Different materials meet different capacity needs, and because consumable electrodes consist of chemically reactive species, certain materials may be compatible with certain drugs or excipients but not all of them. The most popular electrodes are based on the silver/silver chloride redox couple. Silver and silver chloride have several advantageous characteristics: They are biocompatible, perform well, and have an established history of use in medical applications including sensing electrodes.

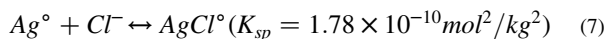
Silver as a Consumable Anode

The use of a silver anode in the presence of chloride or another halide ion in the electrolyte solution is the most commonly used consumable anode for delivery of positively charged drugs.

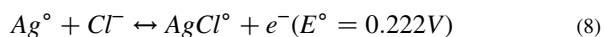
Metallic silver oxidizes according to the following reaction:



However, when chloride is present, it will immediately react with the silver ion, that is:



The complete reaction is therefore:



The final product, silver chloride forms on the surface of the silver anode and is electrically neutral and practically insoluble. Therefore, this reaction couple does not generate species that compete with cationic drugs for delivery. Because the equilibrium potential is low and the reaction is kinetically fast, the silver anode operates at low voltage, avoiding undesirable side reactions such as water splitting or electrochemical degradation of the drug or excipients. The rapid reaction kinetics and immobility of silver chloride make the system highly reversible. A reversible electrode system is especially attractive in an alternating polarity IDDS (i.e., one in which the drug is present in both electrode reservoirs and is delivered alternately from each as the current is reversed). The electrode is effectively discharged on one phase of the cycle and recharged on the other phase, leading to extended capacity.

There is one significant limitation to the use of silver as an anode to form silver chloride. The silver surface of the electrode gradually passivates as the silver chloride layer builds up, causing an increase in discharge voltage with use. This passivation process can limit both the maximum allowable current density and the utilization of silver.

Chloride Ion Management with Silver Anodes

For optimal silver anode function, there must be sufficient chloride ion to react with the silver ion. When the anode is the donor electrode (i.e., delivers positively charged drugs), the addition of chloride salts can lead to ion competition and reduced delivery efficiency. To minimize cation competition the preferred method of adding chloride is by using the drug chloride salt (7). Not all drug substances are readily available as chloride salts, but many can be converted to the chloride form by ion exchange or to the hydrochloride salt form by protonation of amine groups through addition of HCl.

Because chloride ion is consumed as it reacts with silver ions to form insoluble AgCl, its concentration decreases during treatment. Chloride must be present in the electrolyte in sufficient quantity to ensure proper operation of the electrode throughout the therapy. If the concentration of chloride drops to the point where it can no longer scavenge free silver ions, there is the potential for silver to be delivered to the skin. The amount of bulk chloride in the electrolyte required to avoid silver migration depends on many factors, including the volume of the electrolyte, current density,

duration of treatment, the rate at which chloride is replenished from the body (i.e., the chloride transport number), and the configuration of the electrode and reservoir. Because interaction among these factors is often complex, it is difficult to construct a generalized model purely from first principles. Rather, for a given electrode/reservoir combination, it is often simpler to experimentally determine the threshold chloride concentration at which silver migration begins to occur, and then use a mass balance calculation to compute the required starting composition. A helpful indicator of silver migration is the potential of the polarized silver electrode. As chloride becomes less abundant, the silver anode voltage increases in a Nernstian fashion. Empirically, it has been found that operation of the anode above approximately 400 mV (vs. Ag/AgCl standard reference electrode) can lead to free migration of silver ions (5).

If it is impossible to use the chloride salt of a drug or impractical to include it in the amount required to prevent silver migration, two alternative methods can prevent the onset of silver migration without introducing mobile cations into the delivery reservoir. One is to immobilize the silver by using anion-selective membranes or coatings, or by using chelating agents. Another is to precipitate the silver using chloride ion-containing resins (e.g., those with quaternary ammonium chloride functionality) in the donor reservoir (8).

Silver Chloride as a Consumable Cathode

The silver chloride cathode reaction is given by Eq. 8 in reverse; that is, silver chloride is reduced to form metallic silver and chloride ion. The silver chloride cathode shares many of the qualities of the silver anode, with some additional desirable traits: No electrolyte is depleted by its reaction; it is hydrophilic and therefore wetted by the reservoir electrolyte; and the insoluble reaction product, metallic silver, is electrically conductive, eliminating problems of polarization or isolation of the redox species. Because of this combination of properties, the operating voltage of silver chloride decreases with use, and the utilization of a silver chloride cathode is nearly 100%.

Operation of the silver chloride cathode can lead to an accumulation of chloride ions in the electrolyte. When the electrode is a counter electrode, chloride buildup is not a concern. However, when used as a donor electrode for delivery of anionic drugs, the accumulation of chloride ion can lead to decreased drug flux because of ionic competition. For short duration applications, this

effect may be negligible. However, if the molar concentration of chloride approaches some appreciable fraction of that of the drug, substantial competition will occur. A simple, yet not always practical, way to remain below this threshold fraction is to increase drug content. The amount of additional drug required can be computed from a mass balance calculation. The chloride accumulation rate is determined by the current, according to Faraday's law, and this can be compared with the drug content, which decreases over time as drug is delivered into the body.

FORMULATION COMPOSITION

The formulations of an IDDS are the ingredients in the drug and counter reservoirs, which typically consist of a solvent, a drug salt or a biocompatible salt, and a matrix-forming material. A formulation may also include additives such as buffers, antimicrobial agents, antioxidants and additional electrolyte salts or permeation enhancers. All of these can interact in a complex fashion to affect rate of delivery, biocompatibility, and product shelf life.

Solvent

Drug solubility and stability, in addition to solvent biocompatibility, are obvious considerations when selecting a solvent for a pharmaceutical formulation. For an IDDS, the effect of a solvent on the drug charge state is also an important consideration. Although a neutral drug molecule can be transported into the skin by electroosmosis (9), maximal drug-delivery efficiency is usually achieved if the drug has a net electric charge. For this reason, polar solvents with large dielectric constants are preferred. For example, the dielectric constants of water, glycerol, and ethanol are 80, 42, and 24, respectively. Use of solvents with large dielectric constants results in greater dissociation of the drug salt (i.e., less ion-pairing), enhancing drug mobility during application of an electric field.

Because of its large dielectric constant and inherent biocompatibility, water is the most commonly used solvent. Other cosolvents such as ethanol, glycerol, polyethylene glycol, or polypropylene glycol may be added to enhance drug solubility and drug stability, or to reduce the rate of water evaporation. Sanderson and colleagues (10) used a 40:60 mixture of water and ethanol to enhance the solubility of dobutamine hydrochloride and demonstrated a twofold enhancement in dobutamine flux. However, addition of a cosolvent to enhance drug

solubility may, in some cases, decrease the rate of delivery by electromigration if excessive ion-pairing results.

Jadoul and coworkers (11) studied the effect of adding ethanol and propylene glycol (PG) to aqueous solutions of fentanyl and metoprolol. They reported that drug flux was diminished by up to 80% for solutions containing 60 vol% ethanol or PG. A fourfold drop in formulation conductance was also measured, indicating that more ion association was occurring in the cosolvents. In addition, the solvent may have a direct effect on the skin, thus altering its permeability to drug ions (12). In summary, the effect of a solvent or cosolvent on the drug solubility, ion interactions, and skin permeability are important considerations when choosing the formulations for IDDS.

Drug Salt

In addition to the usual solubility, stability, and biocompatibility considerations, several unique aspects should be considered when selecting the drug salt for an IDDS formulation. First, the counterion must be compatible with the electrochemical reactions occurring at the electrode. As noted previously, halide drug salts are preferred when using a silver anode. For example, fentanyl citrate is used in intravenous formulations, but citrate does not form an insoluble salt with electrochemically generated silver cation. For this reason, a formulation containing fentanyl hydrochloride was specifically developed for use in a patient-activated IDDS for treatment of pain (13). Clinical results using this formulation strategy are summarized latter in this chapter.

As mentioned previously, the extent of drug salt dissociation is an important consideration when selecting a solvent. Therefore, for a particular solvent (e.g., water), selection of a drug salt that more fully dissociates will likely result in more efficient drug delivery. Using aqueous solutions of the acetate, sulfate, and hydrochloride salts of morphine, a correlation between drug salt dissociation and transdermal delivery has been observed (14). From conductance measurements, it was determined that morphine hydrochloride was more fully dissociated in water than were the sulfate and acetate salts. The rate of morphine delivery, at currents ranging from 0.1 to 1 mA, was about 60% greater for the hydrochloride salt than for the sulfate and acetate salts.

Ion mobility (the velocity achieved by an ion per unit electric field) is largely determined by its ionic charge and by the extent of its physical interaction with the formulation or skin. The mobilities of a drug ion and its counterion in a formulation are likely to be different than their mobilities in the skin. In addition, the mobilities of ions that are endogenous to the skin

(e.g., Na^+ , K^+ , Cl^- , HCO_3^-) are likely to be different in the two environments. Therefore, during electromigration the ionic composition of the formulation in the vicinity of the skin can be substantially different than the bulk composition. As a result, it has been suggested that the drug counterion can alter the pH of the interface between the formulation and the skin, and thus alter transport efficiency (10). A twofold enhancement in transport efficiency for the succinate salts of verapamil, gallopamil, and nalbuphine relative to the hydrochloride salts was reported. This result was attributed to the ability of the weakly acidic succinate anion to buffer the boundary layer near the skin surface at about pH 4.8, thus avoiding significant hydronium ion competition.

Matrix

Use of drug dissolved in a liquid solvent is generally adequate for in vitro experimentation, but is not optimal for use in a commercial product. Not only must the formulation be biocompatible, but it must also be readily incorporated into the IDDS during commercial-scale manufacturing, be easily applied by the user, and leave little or no residue on the skin. To achieve these goals, two fundamentally different matrix-based formulation strategies have been adopted for use in IDDS.

In one approach, the drug solution is placed on an absorbent porous material. Such materials include hydrophilic fabrics composed of polyester or nylon, and hydrophilic porous films composed of polyurethane, polyvinyl alcohol (PVOH), or cellulose. To improve hydration kinetics and solvent retention, hydrophilic polymers and/or surfactants have been incorporated into the fabric or foam matrices (15). Examples of hydrophilic polymers are polyethylene oxide, PVOH, poly-*N*-vinyl pyrrolidone, polyacrylamide, polyhydroxyethyl methacrylate, and polysaccharides such as hydroxyethyl cellulose, modified starches, or natural gums. Nonionic surfactants such as Tween 20[®], Neodol 91-6[®], or Tergitol 15-S-7[®] can also be added to enhance the rate of hydration. The water-retentive properties of the polymers, combined with the structural integrity of a fabric or porous film, provide a composite matrix material that will readily absorb the drug solution during the manufacturing process or just prior to use by the patient. The addition of solvent just prior to use can enhance drug stability, particularly for polypeptides and proteins.

The second strategy utilizes drug-containing hydrogels and provides an alternative to the absorption of drug solution by porous composite matrices. With the hydrogel approach, drug salt is mixed with a solvent and a

network-forming polymer to create a viscous solution. The solution is then dispensed into a cavity containing an electrode of the appropriate polarity, and the polymer is crosslinked. To minimize degradation of the drug during the crosslinking process, physical crosslinking is preferred over chemical or radiation-induced crosslinking reactions. For example, polyvinyl alcohol can be dissolved in water, mixed with drug salt, and then frozen at about -20°C . When thawed, a soft, cohesive, water-rich hydrogel results (16). Other hydrophilic polymers such as polyvinyl pyrrolidone or polysaccharides (e.g., hydroxypropylmethyl cellulose) can be added to modify the rheological, adhesive, or water-retentive properties of polyvinyl alcohol hydrogels (17).

In general, polar nonionic polymers have been used as the matrix material in formulations for IDDS. Nonionic polymers are preferred because they typically do not have mobile ionic species and do not interact strongly with drug ions. However, results from studies of drug delivery from matrices composed of ionic polymers have been reported. For example, Gupta and coworkers reported a substantial reduction in cromolyn flux when a hydrogel composed of polyglyceryl methacrylate and water was employed, suggesting a strong interaction between the cromolyn anion and polymer (12).

In contrast, the transdermal iontophoretic delivery of the drug cation, hydromorphone, was enhanced by using a hydrogel formulation composed of water and polyacrylamido-methylpropane-sulfonate (poly-AMPS) (8). A hydrogel composed of water and the acid form of poly-AMPS was imbibed with a stoichiometric amount of hydromorphone base to form the hydromorphone salt of poly-AMPS. Hydromorphone hydrochloride was also added, and the hydrogel was placed in contact with a silver anode. Hydromorphone was delivered at a current density of 0.05 mA cm^{-2} through dermatomed pig skin into a 0.1 M sodium chloride solution. The flux of hydromorphone from the poly-AMPS hydrogel was found to be about twice that of a nonionic PVOH hydrogel. This result suggests that migration of mobile chloride ions from the skin into the hydrogel was hindered by the presence of immobilized sulfonate anions (i.e., ionic repulsion or Donnan exclusion).

Excipients

Excipients such as buffers, antimicrobials, antioxidants and chelating agents may be required for optimal drug stability in IDDS formulations. Several unique criteria when selecting excipients must be considered.

As described earlier, excipients can contact the electrodes of the system. Therefore, excipients must be

screened for their compatibility with the electrodes. Sacrificial electrodes (e.g., Ag and AgCl) are particularly reactive. If inherently nonreactive electrodes are used (e.g., platinum or carbon), then the excipient can be exposed to a relatively large electric potential at the electrode/reservoir interface during system use. In such cases, excipients that are inherently stable should be selected. Excipients are typically evaluated for their electrochemical stability using standard potentiometric techniques (e.g., cyclic voltammetry) before being selected for use in an IDDS formulation.

The effect of an excipient on drug transport must also be considered, especially if it is ionic. If the excipient has the same charge as the drug ion, then it will be delivered into the skin with the drug. In addition to the direct competitive effect on drug transport, the excipient may alter the permselectivity of the skin, causing a change in the drug transport efficiency. If an excipient of opposite charge to the drug ion is chosen, then the effect of the excipient counterion on drug transport must be determined.

Because IDDS formulations usually contain water, the use of lipophilic excipients may not be possible. Instead, salt forms of excipients are often employed. However, excipient salts often contain inorganic cations that are usually much more mobile than most drug cations. The detrimental effects of inorganic cations on the flux of drug cations have been well documented (18). In particular, since inorganic cations are depleted from the formulation more rapidly than the drug cations, the flux of the drug will not be constant but rather will increase with time during system use. The competitive effect of buffer anions on transport of anionic drugs has also been reported (19).

Standard phosphate and citrate buffers have been successfully used in formulations for transdermal iontophoresis of drug ions. However, because small inorganic and organic ions frequently have a negative effect on drug flux due to competition, selecting a buffer can be challenging. Several unique buffering strategies have been developed specifically for use in IDDS. In one strategy, zwitterionic buffering agents are used at their isoelectric pH (20). The net zero charge of the zwitterion largely avoids the ion competition effect. Two preferred zwitterionic buffers are *N*-2-hydroxyethylpiperazine-*N*-2-ethane sulfonic acid (HEPES) and 2-*N*-morpholino-propane sulfonic acid (MES).

Alternatively, the delivery of the buffer ion can be largely eliminated by using cationic buffers in the cathode reservoir and anionic buffers in the anode reservoir, so that the buffer ion moves away from the skin when an electric field is applied. In addition to common weak acids such as citric and phosphoric, the use of amino acids in the anode

formulation has been suggested. Amino acids such as cysteine and histidine would be incorporated at neutral or basic pH, where they are predominately anionic (20).

In another buffering strategy, polymeric materials with pendant acid or base groups (e.g., carboxylic, phosphoric, amines) are dispersed in the formulation (10, 20). Examples of such polymeric buffers are polyacrylic acid and methacrylate/divinyl benzene copolymers (e.g., Amberlite IRP-64[®]). The exceedingly high molecular weight of these polymers renders them essentially immobile in an electrical field. The counterion to the ionic resin must still be considered; preferably it should have the opposite charge of the drug to avoid ionic competition. Alternatively, the counterion to the drug ion can be specifically selected for its inherent buffering capability (10).

In all of these strategies, the goal is to minimize the mobility of the buffering agent, or its counterion, relative to that of the drug ion. This is accomplished by choosing buffers with no net ionic charge, choosing buffers whose mobile species have a charge opposite that of the drug ion, or by increasing the molecular weight of the buffering agent.

Excipients may also be included in the formulation to enhance drug delivery efficiency. For example, Sanderson and colleagues (10) reported a threefold enhancement in delivery of dobutamine after the skin site was pretreated with an anionic surfactant, sodium lauryl sulfate (SLS). They attributed the enhancement in flux to an increase in the negative charge on the skin because of neutralization of fixed positive charges in the skin and to hydrophobic binding of the surfactant to the skin. More specifically, Sanderson and coworkers proposed that an increase in negative charge within the transport pathway enhanced the migration of drug cations by hindering the migration of chloride ions from the body. They noted that while SLS is not biocompatible, the charge-alteration strategy may be useful if other more biocompatible surfactants were identified.

Alternatively, Huntington and Cormier (21) used nonionic surfactants, including dodecanol and 1,2-dodecanediol, to enhance the delivery of the anionic drug, ketoprofen. Because nonionic surfactants are not directly affected by the applied electric field, their use may be preferred over ionic surfactants.

Some drugs and excipients may cause excessive skin irritation (22). Researchers have reduced skin irritation in humans by including an antiinflammatory agent in the formulation (23). Using the moderately irritating antiemetic drug, metoclopramide, they demonstrated improved biocompatibility by adding hydrocortisone to the formulation. As little as 0.05% hydrocortisone in the

formulation significantly reduced erythema at the skin site following treatment. They also reported that hydrocortisone had no effect on the transport of the metoclopramide cation.

Formulation pH

Many drugs have a broad pH range in which drug solubility and stability are adequate for transdermal delivery by iontophoresis. However, optimal drug delivery and biocompatibility are usually restricted to a more narrow range of formulation pH. As discussed previously, skin is a permselective membrane with an isoelectric point of about pH 4. For this reason, formulation pH can affect the selectivity of the skin to cations and anions. As formulation pH increases, skin becomes more negatively charged, thus favoring cation transport. Therefore, to maximize the transdermal flux of a cationic drug, the formulation pH should be as basic as is practical, limited by drug solubility, charge state, stability, and biocompatibility. By analogy, for anionic drugs, acidic formulations are generally preferred.

The effect of formulation pH on skin permselectivity has been clearly demonstrated (24): As the pH of a solution containing salicylate anion was increased from pH 4 to 6, and then to 8, the transdermal flux of salicylate anion at 200 $\mu\text{A}/\text{cm}^2$ decreased from 480 to 192, and then to 174 $\mu\text{g h}^{-1} \text{cm}^{-2}$, respectively. In contrast, with an identical increase in pH, the flux of triethylamine cation increased from 117 to 170, and then to 303 $\mu\text{g h}^{-1} \text{cm}^{-2}$, respectively.

A formulation must provide adequate drug transport, while ensuring good biocompatibility. Several investigators have found that formulation pH can have a substantial effect on skin irritation (10, 20, 25). By measuring the redness at treated sites on hairless guinea pigs, Cormier and Johnson (23) found that skin irritation was reduced by choosing different pH ranges for the anode and cathode formulations. For anode formulations, pH values between 4 and 10 produced the lowest skin responses. For cathode formulations, pH values between 2 and 4 were least irritating.

Electrolytes for Counter Reservoir

An appropriate electrolyte for the formulation in contact with the counter electrode must provide sufficient conductivity to minimize the voltage required during system use. The minimum conductivity, σ , needed to limit the voltage drop, ΔV , across a hydrogel formulation with a thickness of L and a cross-sectional area of A at a current of I is given by the expression:

$$\sigma = (IL)/(\delta VA) \quad (9)$$

For example, a voltage drop of less than 1 V will occur across a hydrogel formulation with a thickness of 0.5 cm and a cross-sectional area of 5 cm² at a current of 0.5 mA, if the formulation has a conductivity of at least 50 μ S/cm.

Because formulation conductivity is sensitive to solubility and ion-pairing effects, it can be used to characterize alternative formulations during formulation development. Gangarosa and colleagues (26) and Yoshida and Roberts (19) provide examples of the effect of drug and electrolyte conductivity on formulation performance in IDDS.

In addition to rendering the formulation sufficiently conductive, the ion delivered from the nondrug formulation must be biocompatible. Irritation resulting from use of four inorganic electrolytes has been reported by Anigbogu and coworkers (27). They reported that use of 0.9% NaCl or KCl in the anode reservoir at current densities of 0.5 mA/cm² and 1 mA/cm² for 1 h elicited no skin irritation in rabbits. In contrast, the use of 0.9% CaCl₂ or MgCl₂ caused moderate erythema.

Using weak acids and bases as electrolytes for the counter reservoir formulation at the proper pH (i.e., pH < 4 for cathode formulations and pH > 4 for anode formulations) provides adequate biocompatibility and low skin resistance (20). Low skin resistance is advantageous since less voltage output is required from the control circuit, and therefore a smaller battery may be required, potentially reducing the size and cost of the IDDS.

For current densities at or above 0.2 mA/cm², the sensation associated with transdermal iontophoresis is determined by the type of ion being delivered into the skin. When human subjects compared the sensation experienced during iontophoresis of different salt solutions applied to the right and left forearms, delivery of calcium caused less sensation than delivery of phosphate, magnesium, and zinc, which caused less sensation than delivery of chloride, acetate, citrate, and sulfate, which in turn caused less sensation than delivery of lithium, potassium, and sodium. In general, multivalent ions were found to cause less sensation than monovalent ions (28).

ELECTRONIC COMPONENTS

The unique electrical nature of an IDDS provides an enhanced level of control over drug delivery. In addition, the ability to provide information feedback is not often available in other modes of drug delivery.

There are two general electrical approaches used in iontophoretic treatments: One is to control the applied voltage; the other is to control the applied current. A simple example of a controlled-voltage circuit is a battery that applies a constant voltage to the electrodes. However, as will be pointed out later, the skin resistance is neither constant between individuals nor is it static from the start to finish of a single application. Thus, a constant voltage approach will result in variable current. Since the rate of drug delivered is directly proportional to the current passed, a controlled-current approach is generally preferred.

The electrical control components of an IDDS may be divided into two elements: the power source and the control circuitry.

Power Source

The electrical power source for portable IDDS is a battery. Commercially available primary batteries have cell voltages that range from about 1.2 to 3.0 V. To overcome the resistance of the skin and achieve the desired current, it may be necessary to use higher voltages. Voltage can be increased by combining cells in series, or by using circuitry to "step-up" the voltage at the cost of drawing a larger current from the battery. This additional current drain adds to the total battery capacity needed to provide the required therapy. Thus, the total battery capacity required is the sum of that needed to support the therapy plus any "overhead" associated with the electrical circuitry.

The first objective of the circuitry is to control the amount of drug delivered. For a zero-order (constant-delivery) system, for example, a specified current should be maintained regardless of the resistance of the application site. Computing the power requirement in this case is difficult because the resistive load of the skin

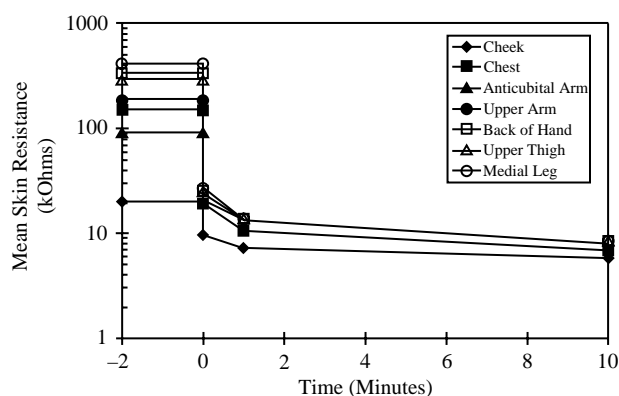


Fig. 2 Skin resistance of various body sites as a function of time of current application. (From Ref. 29.)

depends on both the application site and the time elapsed since current has been initiated (29). Fig. 2 shows the skin resistance at different application sites in response to the application of 1 mA to a 4 cm² area. Two minutes before applying 1 mA, a small current of 10 μ A was applied to determine the skin resistance, which was highly variable and had a mean value of about 200 k Ω . Just 1 s after applying a current of 1 mA, the resistance decreased substantially to 16 ± 4 k Ω . As time progressed, the resistance continued to fall and all sites appeared to approach a quasi-steady-state value of 7.4 ± 1.3 k Ω . The small relative standard deviation (RSD) of the quasi-steady-state resistance of 18% compared with an initial RSD of 92% indicates that all body sites tested tend toward a common resistance, despite the large differences that are present prior to application of current. This fact has been attributed to an increase in ionic conductivity of the skin due to an increase in ion concentration within the stratum corneum and to “activation” of shunt pathways (3).

For a typical initial skin resistance of 200 k Ω , 200 V would be required to achieve a current of 1 mA. Therefore, a typical system having a modest output voltage of 10–20 V will operate below the desired current (i.e., will be noncompliant) during the initial moments of current application. However, because the resistance drops quickly, a voltage capability of about 10–20 V is generally sufficient to achieve compliance within an acceptably short time of about 1 min or less.

Control Circuitry

A controlled current circuit does not necessarily require a high level of electronic sophistication. A simple field-effect transistor with a feedback resistor can maintain the current delivered from a battery at a constant value over a wide range of skin resistances. Other types of delivery requirements such as pulsed current, patient-controlled on-demand dosing, dose titration, ramp-up or ramp-down dosing, and other special waveforms can be addressed with control circuitry. In these cases integrated circuits are usually employed to minimize system size.

Visual and audio feedback information is provided to the user by light-emitting diodes (LED), liquid crystal displays (LCD), and piezoelectric transducers. This information can indicate that the system is working, show errors or warnings, signal system maintenance needs (such as battery replacement), or display the amount of drug that has been delivered.

The ability of an IDDS to electronically control drug delivery gives rise to another potential advantage; the opportunity to create sensor-based or “closed-loop” therapy. If an appropriate sensor is available to measure

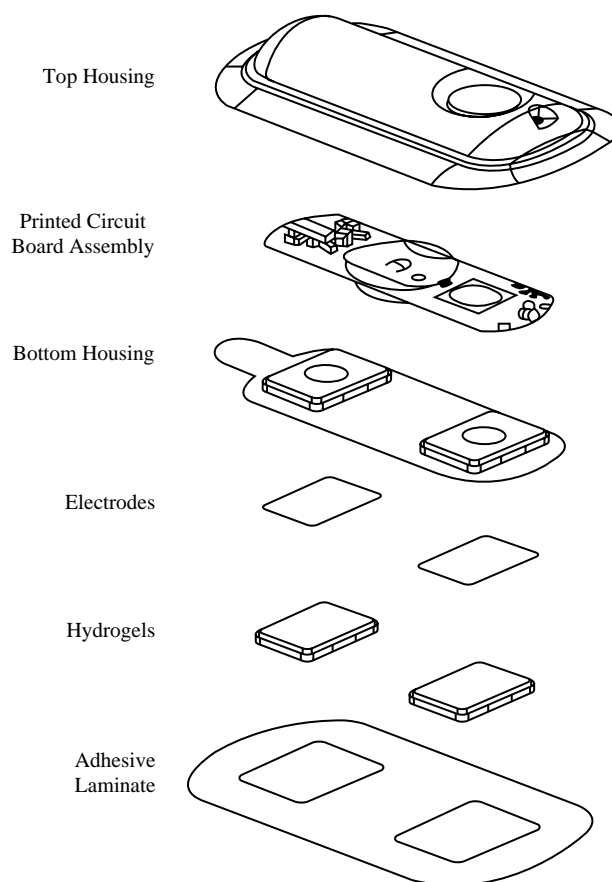


Fig. 3 Exploded view of an iontophoretic drug delivery system showing major components: top housing, printed circuit board assembly, a bottom housing containing reservoirs for placement of electrodes and hydrogels, and an adhesive laminate. (Diagram from International Patent Publication Number WO 96/39222.)

a biological response (e.g., to measure blood glucose levels in response to the delivery of insulin), then circuitry can be used to modulate the output current, and therefore drug delivery. In so doing, an automated, closed-loop system is created that ensures the proper amount of drug is delivered and avoids both over- and under-dosing the patient. As shown in Fig. 3, the electronic components are typically mounted on a printed circuit board and mounted within a protective housing containing the electrodes and formulation hydrogels.

CLINICAL ASSESSMENT OF SYSTEMIC DELIVERY

In the following clinical case studies, a number of delivery attributes discussed in earlier sections are exemplified. In

addition, *in vitro* methodologies are briefly discussed to highlight the good correlation possible with observed *in vivo* results.

Fentanyl Clinical Study

Fentanyl has a molecular weight of 336 Da in the base form, an aqueous solubility of about 24 mg/ml as the HCl salt, and a charge of +1 over a wide pH range. Fentanyl is a synthetic opioid widely used in anesthesia and analgesia. A passive transdermal system is available for the treatment of chronic pain (Duragesic®, Janssen Pharmaceutica). This patch delivers fentanyl through the skin continuously over 72 h. For the management of acute pain, such as post-operative pain, the slow onset of action of fentanyl obtained with passive transdermal delivery is not appropriate. However, fentanyl delivery from an IDDS to quickly attain therapeutic blood levels and a quick onset of analgesia has been demonstrated clinically (30, 31). In addition, the electrical nature of iontophoresis allows the patient to initiate a dose of fentanyl when in pain, which is not possible with passive delivery (32).

The IDDS used in the clinical study consisted of a reusable, battery-operated controller capable of delivering user-actuated doses when connected to a disposable patch-like drug unit. The drug unit had a silver anode and silver chloride cathode in intimate contact with hydrogel reservoirs within a flexible foam housing. The drug unit contained 5 mg (base equivalent) of the hydrochloride salt of fentanyl and the donor hydrogel reservoir had a skin

contact area of 2 cm². The disposable drug units were applied to the upper outer arms of healthy volunteers. The opioid effects of fentanyl were blocked in the volunteers by oral administration of naltrexone every 12 h.

The study was a four-treatment crossover (three with IDDS plus an IV treatment), conducted in 12 subjects. All treatments delivered fentanyl for the first 20 min of every hour for 24 h. The three IDDS treatments employed currents of 150, 200, and 250 μ A, while the IV treatment infused 50 μ g of fentanyl. Blood samples were collected, and serum fentanyl concentrations were measured using a specific radioimmunoassay. Plasma data collected during and immediately following the 1st, 13th, and 25th treatments are shown in Fig. 4. The gradual upward shifts in concentration over time indicate a baseline increase due to incomplete fentanyl clearance between the hourly doses. For all treatment types, serum fentanyl concentrations rapidly increased within a few minutes after the start of each dose.

The average drug input fluxes during the 20-min dosing periods between hours 24 and 25 were calculated to be 81, 108, and 138 μ g/h/cm² for currents of 150, 200, and 250 μ A, respectively. These values, the mean maximum plasma concentration values, and the total AUC values (over the same time period) for the three IDDS treatments all increased proportionally with current. These results agree with theoretical expectations expressed by Equation (1). In addition, the variabilities in the fentanyl pharmacokinetic parameters were similar for the IDDS and IV treatments, indicating that the IDDS doses were delivered with an accuracy similar to the IV infusions.

The *in vivo* flux values for fentanyl agree closely with *in vitro* flux values obtained with human cadaver skin. *In vitro* transdermal experiments with fentanyl were conducted using two-compartment permeation cells.

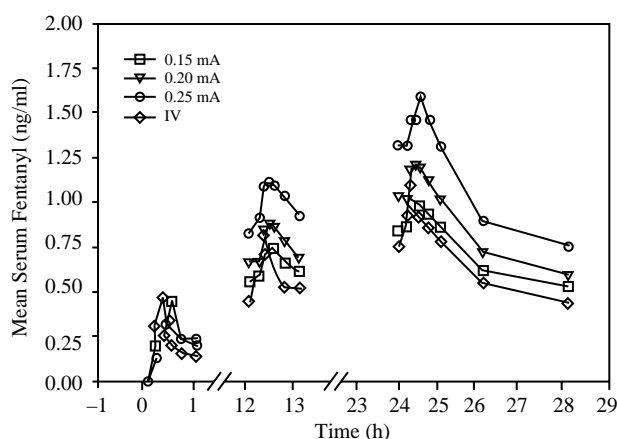


Fig. 4 Mean serum fentanyl concentrations for 12 healthy volunteers receiving fentanyl intermittently (20 min each hour over a 24-h administration period) from an iontophoretic system at three currents, and from IV fentanyl infusion (50 μ g for 20 min hourly). Serum fentanyl concentrations were measured during and immediately after the 1st, 13th, and 25th doses. (From Ref. 31.)

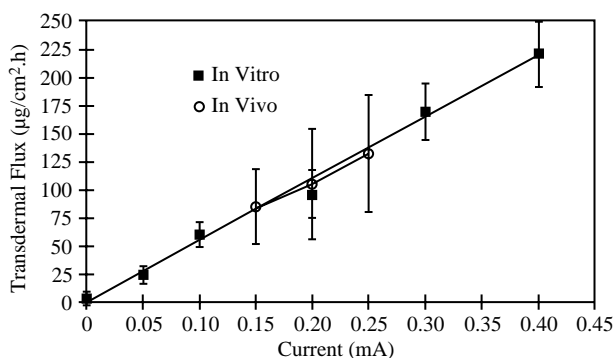


Fig. 5 Comparison of *in vitro* ($n = 5$) and *in vivo* ($n = 13$) fentanyl flux as a function of applied current. (From Ref. 13.)

Heat-stripped human epidermis was oriented so that the stratum corneum surface contacted the fentanyl hydrochloride solution in the donor compartment. The donor solution pH and fentanyl concentration were very similar to the pH and concentration of the donor formulation used in the clinical study.

To establish a good in vitro/in vivo correlation, the composition and ionic strength of the receptor solution must be adjusted (18). For fentanyl, the best correlation was obtained with a receptor solution comprised of a 10-fold dilution of modified Dulbecco' phosphate-buffered saline (without calcium or magnesium) at pH 7.4. The currents tested in vitro ranged from 0 to 400 μ A and were applied to a silver anode and a silver chloride cathode using a constant current power supply. The in vitro data presented in Fig. 5 clearly demonstrate the expected linear dependence of fentanyl delivery on current and the excellent correlation to the in vivo data.

Luteinizing Hormone-Releasing Hormone Clinical Study

Systemic delivery of polypeptides, proteins and oligonucleotides pose many significant delivery challenges to pharmaceutical scientists. Because of their extensive metabolism in the gastrointestinal tract, they are generally not good candidates for oral administration. These compounds typically have large molecular weights, are hydrophilic, and have a pH-dependent electrical charge. These qualities also make them poor candidates for passive transdermal delivery but excellent candidates for iontophoretic delivery (33).

Luteinizing hormone-releasing hormone (LHRH) is a native reproductive hormone containing 10 amino acids with a molecular weight of approximately 1200 Da and a charge of +1 near neutral pH. Pulsatile delivery of LHRH by the hypothalamus stimulates the production of gonadotropins, such as luteinizing hormone (LH), by the pituitary gland for the maintenance of normal female reproductive function. In contrast, continuous secretion of LHRH shuts down the reproductive axis.

It is possible to achieve continuous or pulsed LHRH delivery profiles using transdermal iontophoresis. Heit and colleagues have demonstrated that continuous transdermal iontophoresis of LHRH is possible (34). Continuous transdermal iontophoresis of leuprolide, a nine-amino acid analog of LHRH, has also proved clinically successful (35).

The iontophoretic delivery of discrete pulses of LHRH using an IDDS was investigated in a clinical study with healthy male volunteers (36). The clinical system comprised an 8 cm² hydrogel formulation containing 15 mM LHRH as the hydrochloride salt in contact with a

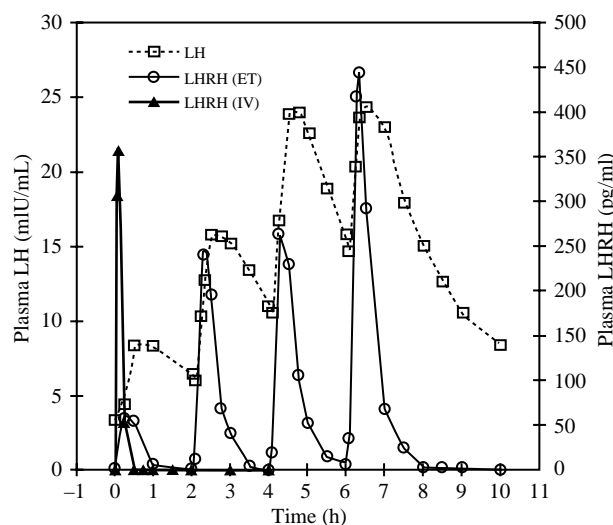


Fig. 6 Mean plasma LHRH concentrations of eight healthy male volunteers receiving four 15-min doses of LHRH at 0.1 mA/cm² every two h, and an IV LHRH bolus dose of 5 μ g. Mean plasma LH levels resulting from the iontophoretic LHRH doses are also plotted. (From Ref. 36.)

silver anode. Plasma LHRH and LH concentrations obtained in eight subjects receiving one 15-min pulse at a current of 0.8 mA (0.1 mA/cm² current density) every 2 h for 8 h are shown in Fig. 6. The plasma LHRH concentration obtained following the administration of a single 5 μ g IV bolus is also shown as a reference. These results demonstrate that pulsatile delivery of LHRH is possible using reasonable iontophoretic conditions. Significantly, the plasma LHRH profiles obtained with an IDDS are comparable with those obtained following IV bolus administration. The four consecutive 15-min doses resulted in plasma LHRH profiles with sharp peaks (C_{\max} up to 450 pg/ml). Plasma LHRH levels rapidly declined to negligible baseline values between doses, consistent with IV administration. The plasma LHRH levels rose and fell much more rapidly than profiles obtained with subcutaneous injections (37).

Following each dose of LHRH, the pharmacodynamic plasma LH response also followed a pulsatile pattern with a mean maximum concentration of about 26 mIU/ml. As was discussed earlier for fentanyl, the variability in the iontophoretic LHRH plasma data was comparable to that obtained with IV administration.

While iontophoresis has not been widely employed, the field is rapidly evolving. The development of small patch-like devices for glucose extraction and for systemic delivery of therapeutic agents is occurring at several companies, including Cygnus, Becton Dickinson, and ALZA. If commercially successful, these efforts would

greatly expand the applications of iontophoresis for systemic therapy.

REFERENCES

- Banga, A.K. *Electrically Assisted Transdermal and Topical Drug Delivery*; Taylor and Francis: Bristol, PA, 1998.
- Leduc, S. *Electric Ions and Their Use in Medicine*. Rebman Ltd.: Liverpool, UK, 1908.
- Scott, E.R.; Phipps, J.B.; White, H.S. Direct Imaging of Molecular Transport Through Skin. *J. Invest. Dermatol.* **1995**, *104*, 142–145.
- Lindblad, L.E.; Ekenvall, L. Electrode Material in Iontophoresis. *Pharm. Res.* **1987**, *4*, 438.
- Phipps, J.B. Electrode and Reservoir Design for Optimal Transdermal Delivery by Iontophoresis. *Transdermal Administration, A Case Study, Iontophoresis*; APGI/CRS European Symposium, Couvreur, P., Duchene, D., Green, P., Junginger, H.E., Eds.; 3–4 March 1997, Paris, France Editions de Sante: Paris, 1997; 30–39.
- Sanderson, J.E.; Deriel, S.R. Method and Apparatus for Iontophoretic Drug Delivery. US Patent 4,722,726, 1988.
- Philips, J.B.; Untereker, D.F. Iontophoresis Apparatus and Methods of Producing Same. US Patent 4,744,787, 1988.
- Philips, J.B.; Howland, W.W.; Jevne, A.H.; Holmblad, C. Device and Method for Iontophoretic Drug Delivery. US Patent 5,558,633, 1996.
- Pikal, M.J. The Role of Electroosmotic Flow in Transdermal Iontophoresis. *Adv. Drug Del. Rev.* **1992**, *9*, 201–237.
- Sanderson, J.E.; Reil, S.D.; Dixon, R. Iontophoretic Delivery of Nonpeptide Drugs: Formulation Optimization from Maximum Skin Permeability. *J. Pharm. Sci.* **1989**, *78*, 361–364.
- Jadoul, A.; Regnier, V.; Preat, V. Influence of Ethanol and Propylene Glycol Addition on the Transdermal Delivery by Iontophoresis and Electroporation. *Pharm. Res.* **1997**, *14*, S308–S309.
- Gupta, S.K.; Kumar, S.; Bolton, S.; Behl, C.R.; Malick, A.W. Effect of Chemical Enhancers and Conducting Gels on Iontophoretic Transdermal Delivery of Cromolyn Sodium. *J. Controlled Release* **1994**, *35*, 229–236.
- Lattin, G.L.; Phipps, J.B.; Southam, M.A.; Klausner, M. Evaluation of Fentanyl Delivery in Humans Using E-TRANS Technology. *Transdermal Administration, A Case Study, Iontophoresis*; APGI/CRS European Symposium, Couvreur, P., Duchene, D., Green, P., Junginger, H.E., Eds.; 3–4 March 1997, Paris, France Editions de Sante: Paris, 1997; 365–368.
- Corish, J.; Corrigan, O.I.; Foley, D. The Iontophoretic Transdermal Delivery of Morphine Hydrochloride and other Salts Across Excised Human Stratum Corneum, Proceedings of the Conference on Prediction of Percutaneous Penetration: Methods, Measurements, Modeling, April 1989 International Conference; Scott, R.C., Guy, R.H., Hagraft, J., Eds.; IBC Technical Services Ltd.: London, 1990; 302–307.
- Beck, J.E.; Lloyd, L.B.; Petelenz, T.J. Iontophoretic Delivery Device with Integral Hydrating Means. US Patent 5,730,716, 1998.
- Hyon, S.H.; Yoshito, I. Transdermal Therapeutic Composition. US Patent 4,781,926, 1988.
- Jevne, A.H.; Vegoe, B.R.; Holmblad, C.M. Hydrophilic Pressure Sensitive Biomedical Adhesive Composition. US Patent 4,593,053, 1986.
- Phipps, J.B.; Gyory, J.R. Transdermal Ion Migration. *Adv. Drug. Del. Rev.* **1992**, *9*, 137–176.
- Yoshida, N.H.; Roberts, M.S. Prediction of Cathodal Iontophoretic Transport of Anions Across Excised Skin from Different Vehicles Using Conductivity Measurements. *J. Pharm. Pharmacol.* **1995**, *47*, 883–890.
- Cormier, M.J.; Ledger, P.W.; Johnson, J.; Philips, J.B.; Chao, S. Reduction of Skin Irritation and Resistance During Electrotansport. US Patent 5,624,415, 1997.
- Huntington, J.A.; Cormier, M. Composition, Device and Method for Electrotansport Agent Delivery. US Patent 5,811,465, 1998.
- Ledger, P.W. Skin Biological Issues in Electrically Enhanced Transdermal Delivery. *Adv. Drug Del. Rev.* **1992**, *9*, 289–307.
- Cormier, M.; Johnson, B. Skin Reactions Associated with Electrotansport. *Transdermal Administration, A Case Study, Iontophoresis*; APGI/CRS European Symposium, Couvreur, P., Duchene, D., Green, P., Junginger, H.E., Eds.; 3–4 March 1997; Paris, France Editions de Sante: Paris, 1997; 50–57.
- Nightingale, J.; Sclafani, J.; Kurihara-Bergstrom, T. Effect of pH on the Iontophoretic Delivery of Ionic Compounds, Proceedings of the 17th International Symposium on Controlled Release of Bioactive Materials, Reno, Nevada, July 22–25, 1990; Lee, H.E., Ed.; Controlled Release Society: Lincolnshire, IL, 1990; 17, 431–432.
- Phipps, J.B.; Cormier, M.; Padmanabhan, R. *In Vivo* Ion Efflux from Humans and the Relationship to Skin Irritation. *Transdermal Administration, A Case Study, Iontophoresis*; APGI/CRS European Symposium, Couvreur, P., Duchene, D., Green, P., Junginger, H.E., Eds.; 3–4 March 1997; Paris, France Editions de Sante: Paris, 1997; 289–291.
- Gangarosa, L.P.; Park, N.H.; Fong, B.C.; Scott, D.F.; Hill, J.M. Conductivity of Drugs Used for Iontophoresis. *J. Pharm. Sci.* **1978**, *67*, 1439–1443.
- Anigbogu, A.; Singh, P.; Liu, P.; Dinh, S.; Maibach, H. Effects of Iontophoresis on Rabbit Skin *In Vivo*. *Pharm. Res.* **1997**, *14*, S-308.
- Philips, J.B. Method for Reducing Sensation in Iontophoretic Drug Delivery. US Patent 5,221,254, 1993.
- Gyory, J.R.; Phipps, J.B. Effect of Current Density and Current Duration on the Membrane Resistance of Skin. *Transdermal Administration, A Case Study, Iontophoresis*; APGI/CRS European Symposium, Couvreur, P., Duchene, D., Green, P., Junginger, H.E., Eds.; 3–4 March 1997; Paris, France Editions de Sante: Paris, 1997; 262–265.
- Gupta, S.K.; Bernstein, K.J.; Noorduyn, H.; Van Peer, A.; Sathyan, G.; Haak, R. Fentanyl Delivery from an Electrotansport System: Delivery is a Function of Total Current, not Duration of Current. *J. Clin. Pharmacol.* **1988**, *38*, 951–958.
- Gupta, S.K.; Southam, M.; Sathyan, G.; Klausner, M. Effect of Current Density on Pharmacokinetics Following Continuous or Intermittent Input from a Fentanyl Electrotansport System. *J. Pharm. Sci.* **1998**, *87*, 976–981.

32. Thysman, S.; Tasset, C.; Preat, V. Transdermal Iontophoresis of Fentanyl: Delivery and Mechanistic Analysis. *Int. J. Pharm.* **1994**, *101*, 105–113.
33. Delgado-Charro, M.B.; Guy, R.H. Iontophoretic Delivery of Nafarelin Across the Skin. *Int. J. Pharmaceut.* **1995**, *117*, 165–172.
34. Heit, M.C.; Williams, P.L.; Jayes, F.L.; Chang, S.K.; Riviere, J.E. Transdermal Iontophoretic Peptide Delivery: *In Vitro* and *In Vivo* Studies with Luteinizing Hormone Releasing Hormone. *J. Pharm. Sci.* **1993**, *82*, 240–243.
35. Meyer, R.B.; Kreis, W.; Eschbach, J.; O'Mara, V.; Rosen, S.; Sibalis, D. Transdermal versus Subcutaneous Leuprolide: A Comparison of Acute Pharmacodynamic Effect. *Clin. Pharmacol. Ther.* **1990**, *48*, 340–345.
36. Scott, E.R.; Phipps, J.B.; Gyory, J.R.; Padmanabhan, R.V. Electrotransport Systems for Transdermal Delivery: A Practical Implementation of Iontophoresis. *Handbook of Pharmaceutical Controlled Release Technology*; Wise, D., Ed.; Marcel Dekker, Inc.: New York, 2000.
37. Handelsman, D.J.; Jansen, R.P.S.; Boylan, L.M.; Spaliviero, J.A.; Turtle, J.R. Pharmacokinetics of Gonadotropin-Releasing Hormone: Comparison of Subcutaneous and Intravenous Routes. *J. Clin. Endocrinol. Metab.* **1984**, *59*, 739–746.
38. Banga, A.K.; Chien, Y.W. Iontophoretic Delivery of Drugs: Fundamentals, Developments and Biomedical Applications. *J. Contr. Rel.* **1998**, *7*, 1–24.
39. Burnette, R.R. Iontophoresis. *Transdermal Drug Delivery: Developmental Issues and Research Initiatives*; Hadgraft, J., Guy, R.H., Eds.; Marcel Dekker, Inc.: New York, 1989; 247–292.
40. Singh, P.; Maibach, H.I. Iontophoresis in Drug Delivery: Basic Principles and Applications. *Crit. Rev. Ther. Drug. Carr. Syst.* **1994**, *11*, 161–213.
41. Tyle, P. Iontophoretic Devices for Drug Delivery. *Pharm. Res.* **1986**, *3*, 318–326.

FURTHER READING

For additional examples of In Vitro and In Vivo Iontophoretic Drug Delivery Studies and Techniques, the Reader is Directed to the Comprehensive Reviews by Banga (1), Banga and Chien (38), Burnette (39), Singh and Maibach (40), and Tyle (41).

INHALATION, LIQUIDS

Michael E. Placke

Jeffrey Ding

William C. Zimlich, Jr.

Battelle Pulmonary Therapeutics, Inc., Columbus, Ohio, U.S.A.

INTRODUCTION

Using the inhalation route to deliver therapeutic aerosols is a common practice in the treatment of patients with various airway diseases. Drug delivery via inhalation offers many advantages in the administration of pharmaceutical compounds, because the drugs are delivered directly to the site of action. Therefore, the required therapeutic dose for each treatment is lower than if the dose is administered via oral or parenteral routes. As a consequence, the inhalation route reduces adverse effects due to systemic absorption and intensifies the amount of drug deposited to the targeted tissue. In the U.S. market, bronchodilator, anticholinergic, anti-inflammatory, and corticosteroid drugs are common inhaled dosage forms used for the treatment of respiratory diseases, such as asthma and COPD. Additional drug classes, such as antibiotics, antifungals, and antiviral compounds, are also used or under development for topical delivery via inhalation. Further, new inhaled applications are being developed for the treatment of systemic diseases. These products include both small and macromolecules. Pulmonary delivery of peptides and proteins is a rapidly growing area of inhaled drug delivery, spearheaded by the development of inhaled insulin for treatment of diabetes.

An aerosol is defined in its simplest form as a collection of solid or liquid particles suspended in a gas such as air. Aerosols are two-phase systems consisting of the particulates and the gas in which they are suspended (1). Aerosol science has been studied for many decades in various disciplines, especially related to environmental sciences, to understand the health effects associated with exposure to air pollutants. These activities have improved our understanding of the dynamic behavior of aerosol particles and their relative effects on human health, but also laid the foundation for the pharmaceutical industry to use inhalation of aerosols for therapeutic purposes. Pharmaceutical aerosols deliver therapeutically active drug compounds to the human respiratory tract for local or systemic actions. This dosage form is intended to be inhaled through either the mouth or the nose. The

increasing use of inhalation therapy has been mainly driven by the confluence of three factors:

1. Advances in aerosol generation technology, which provide more efficient and controlled delivery of aerosolized pharmaceutical compounds into the human respiratory airway;
2. Advances in biotechnology, with the development of new therapeutic agents that are often difficult to deliver by other routes of administration;
3. A better understanding of the pathophysiology of disease, which allows physicians and scientists to envision a broader range of therapeutic options.

Those involved with the pulmonary delivery of therapeutic agents are challenged in all three areas, especially the first two. We can now develop new therapies, but we must be certain that we can reproducibly deliver those therapeutic agents to the desired site, in the dosage and form in which they will be most effective (2).

The benefit of using the respiratory airways as a route for drug delivery has been increasingly explored over the last 10 years (3). The large surface area of the human respiratory system, sometimes compared to the size of a tennis court ($\sim 100 \text{ m}^2$), is an attractive target site for pulmonary delivery of drugs intended for absorption into the blood. Drug administration via inhalation has many advantages for the systemic delivery of pharmaceutical compounds such as proteins and peptides, which are often degraded when delivered through the GI tract and avoid the need for repeated injections (4).

Inhaled aerosols also provide high drug concentrations in the respiratory tract. The treatment of airway diseases by aerosol inhalation is beneficial because of intensified localized drug deposition and fewer side effects. Consequently, smaller doses are needed to achieve the therapeutic level in the lungs via inhalation thereby avoiding high systemic levels of the drug. Therefore, inhaled drug aerosols have become the standard clinical approach for managing asthma, respiratory distress syndrome (RDS), chronic obstructive pulmonary disease (COPD), cystic fibrosis, and other respiratory tract diseases (5). Despite

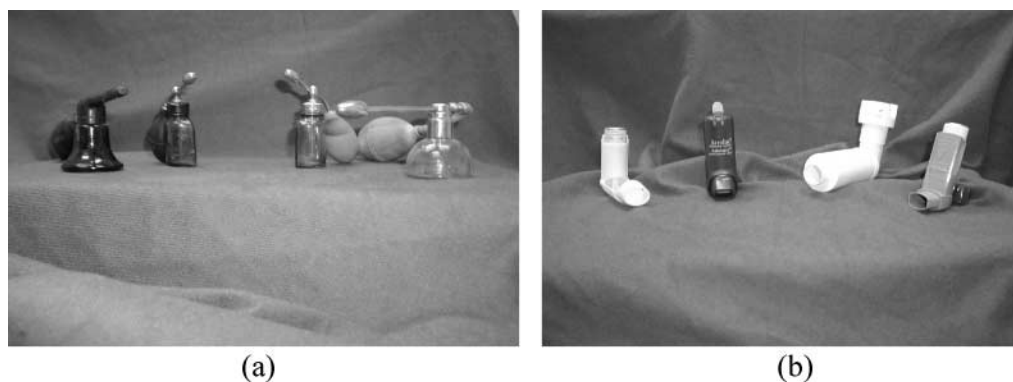


Fig. 1 Drug delivery devices. (a) Squeezing bulb nebulizers (courtesy of Bruce McVeety) and (b) pMDIs.

the spectrum of opportunities that now exists for new inhaled therapies, successful development of pulmonary drug products is a challenging task. The problems and barriers are multidisciplinary in nature, perhaps explaining why this field has not accelerated faster (6).

Using aerosolization to administer drugs to the respiratory airway dates back to the early part of the 20th century, when the pneumatic nebulizer was used as a drug delivery device. Fig. 1a shows some examples of the first types of nebulizers used, the simplest of which were powered by squeezing the bulb attached to the device. Fig. 1b shows more modern units such as pMDIs.

The nebulization principle for the squeeze-bulb nebulizers is the same as many conventional nebulizers. Liquid is forced through a small critical orifice. The particle size of aerosol generated from the early squeeze-bulb nebulizers was generally large (due to the inability to accurately machine small holes) and, therefore, largely unsuitable for human inhalation. However, the concept of delivering drug products via inhalation was an important milestone that established the foundation for developing better pulmonary drug delivery technology.

The modern pressurized metered dose inhaler (pMDI) was developed between 1955 and 1956. However, the pMDI had its roots in research carried out many years before (7). During the 1930s and 1940s, the discovery of liquefied propellants (chlorinated-fluorocarbons or freon propellants) such as CFC12, CFC114, CFC11, and CFC22 was a major step in the realization of a portable inhaler. However, the first pMDI was not invented until 1956, when the metering valve used in the device was developed and patented (7).

The invention of the first pMDI by Riker Laboratories (now 3M Pharmaceuticals) was described in detail recently, and was approved by the FDA in March 1956 (8). The Medihaler-EptTM and Medihaler-IsoTM were launched at the same time. Illustrations from the first brochure of these inhalers are presented in Fig. 2. The Medihaler-Iso was changed to accommodate suspension formulations during 1957. This proved particularly useful for formulating difficult to solubilize drugs, such as many of the steroids later developed for the treatment of asthma. The successful innovation of the pMDI established the foundation for the development of drug delivery via

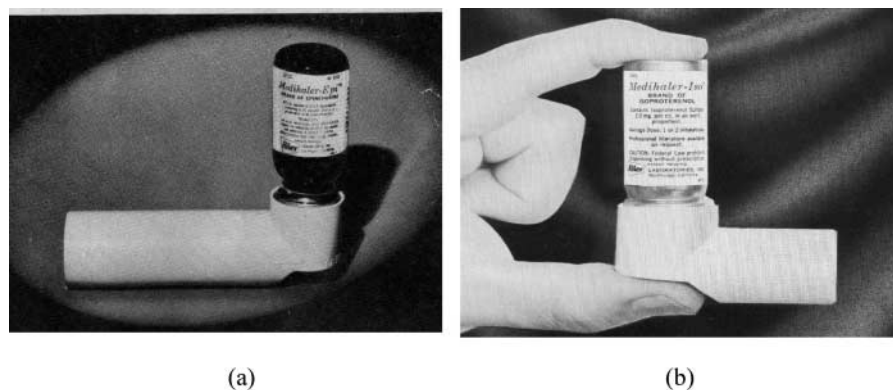


Fig. 2 The metered dose inhaler. (a) Medihaler-Ept and (b) Medihaler-Iso. (Courtesy of Charles Thiel.)

respiratory airways and remained the primary small volume nebulizer for the next 40 years.

Successful administration of any pharmaceutical compound to the human respiratory airway requires the generation and delivery of the drug in the form of an suitable for human inhalation. The human airway is a very complex structure that has evolved to serve two purposes: 1) exchanging gases vital to metabolism (O_2 and CO_2) and 2) defending against foreign particulates from entering the lungs. The generation of drug aerosols and their site of deposition in human lungs are dependent on many factors, such as the physical and chemical properties of aerosol particles, the mechanism of action and quality of performance of the drug delivery device, the patient's breathing pattern, and the morphometry of the human lungs. These factors are significant because:

- Aerosol properties, such as particle size distribution, aerosol velocity, and hygroscopicity, affect aerosol deposition in the human lungs. Aerosol size distribution, including mass median aerodynamic diameter (MMAD) and geometric standard deviation (GSD), is one of the most important variables in governing the site of droplet or particle deposition in the lungs (3).
- Device characteristics play an important role in the delivery of pharmaceutical compounds to human lungs. The mechanism by which aerosols are created affect particle size distribution, kinetics of aerosol delivery, and the reproducibility of drug dosing from the delivery device. These factors are key to controlling the delivery of a constant amount of drug aerosol across a diverse patient population.
- The breathing patterns or maneuvers of a patient can also significantly affect the site of aerosol deposition in the human lungs, even of aerosols with similar characteristics (e.g., MMAD and GSD). Aerosol deposition mechanisms in the airway mainly consist of sedimentation, inertial impaction, and diffusion. Pharmaceutical aerosols deposit mainly by sedimentation and impaction. Therefore, aerosols delivered under high velocity conditions, either due to high ballistic velocity produced by the device or due to high patient inspiratory flow rates, result predominantly in deposition due to inertial impaction, markedly enhancing conducting airway deposition. Conversely, deposition in the peripheral lung can be promoted by decreasing the velocity of an aerosol during inhalation. Further, deposition of these particles that reach the deep lung can be enhanced by a postinspiratory pause (i.e., breath-hold), allowing particles to settle more completely.

- Morphometry of the upper and lower airways will also affect the drug deposition in the lungs. Anatomical conditions such as airway size, disease status, and degree of obstruction are different from patient to patient; thus there will be variability in the site of drug deposition between patients.

The inhalation drug delivery devices currently available commercially consist of metered dose inhalers (MDIs), dry powder inhalers (DPIs), and nebulizers. The advantages and disadvantages of those devices have been reviewed extensively in the literature. This article will focus on the inhalation of aerosols generated from liquid systems using either small volume inhalers, or larger home or clinical nebulizers. Aerosols produced in powder form are covered in another chapter. Various drug delivery devices, including those available commercially and those under development, will be discussed in detail, so that the reader will have a better understanding of the advantages and disadvantages of those devices for pulmonary drug delivery. An emphasis will be given to new pulmonary drug delivery device technology in development, and its potential application to deliver pharmaceutical compounds for both local targeting and systemic absorption of drugs intended to act outside the lung. Over the last 10 years, many innovations have occurred in the area of inhalation drug delivery stimulated by advances in aerosol generation technology and the demand created by new biopharmaceuticals.

COMMERCIALLY AVAILABLE PULMONARY DELIVERY TECHNOLOGY

The inhalation drug delivery devices currently available commercially consist of metered dose inhalers (MDIs), dry powder inhalers (DPIs) and nebulizers. The effective performance (or lack thereof) of the inhalation devices is well recognized as being crucial in how effective a pulmonary delivered drug is in the treatment of diseases, both topical and systemic disorders. In order for a drug to be effective by inhalation, the generated aerosol must be of proper size to be deposited in the targeted location. The patient needs to inhale the drug formulation properly, and the drug must be deposited with sufficient quantity at the action site so that it can be absorbed or distributed effectively. The drug delivery devices available commercially have been used widely for the treatment of airway diseases such as asthma and chronic obstructive pulmonary diseases (COPDs). Such treatment is effective mainly because of the wide therapeutic index of the drug compounds. It is useful to review the advantages and

disadvantages of current device technologies to provide a better understanding of how the devices are used by patients and what kind of improvements are needed to deliver various drug compounds by inhalation. This article is limited to discussion of devices that generate liquid drug aerosols as solutions, suspensions, and emulsions^a

There are several ways to categorize drug delivery devices. The inhalation drug delivery devices discussed in this chapter are categorized according to their operating principles or fundamental mechanism of generating aerosols.

Pressure-Driven Devices

Jet nebulizers

The jet nebulizer has been used for many years to treat airway diseases both in the clinical setting and in home use. The jet nebulizer utilizes compressed gas from an air source either from a hospital air line or from a portable compressor to convert the liquid drug solution into small droplets by forcing the liquid through a small hole known as a critical orifice. Atomization occurs as a result of the disruption of the surface tension holding the liquid together by the action of internal and external forces. In the absence of such disruptive forces, surface tension tends to pull the liquid into the form of a sphere, since this has the minimum amount of surface energy (9). The droplets produced by atomization contain many large drops, which will undergo further disruption to form small particles. The particle size of the droplets will depend not only on the kinetic energy used during the nebulization, but also the pore diameter of the critical orifice and on the physical properties of the liquid formulation.

The jet nebulizer is driven by air pressurized typically at 20–40 psi. The compressed air accelerates through a narrow orifice to break the bulk liquid into sheets, jets, films, or streams. Those ligaments are accelerated to a velocity sufficient to impact on baffles or on the nebulizer wall. The outgoing air becomes saturated with water vapors derived from the liquid retained in the nebulizer, and this has two important consequences:

1. The nebulizer is cooled and reaches an equilibrium temperature approximately 10°C below ambient, so that the patient inhales a relatively cool aerosol cloud.
2. The evaporation of water causes the concentration of solutes to increase with time as the vehicle carrier evaporates (10).

^aDevices producing liquid dry powder aerosols is discussed in *Inhalation, Dry Powder*, page 1529–1544.

Droplet formation: A general theory for the formation of droplets by nebulization has not been fully developed, although the understanding of the nebulization principle has been improved with mathematical modeling. A rigorous mathematical description of droplet formation requires knowledge of distribution of aerodynamic pressure on the droplet surface and the interaction of this pressure with internal forces acting on the droplets. Therefore, before further discussing of the effect of various parameters on the nebulization performance, it is beneficial to consider briefly the various ways in which a single droplet of liquid can break up under the action of aerodynamic forces. Once the droplets are generated using a compressed air source during the primary atomization, unstable droplets experience further disintegration to form smaller particles, which is referred to as “secondary atomization.” The formation of a single droplet under the aerodynamic forces exerted by the external sources, together with the internal forces of the liquid formulation, such as surface tension and viscosity, dictate the formation of the droplet. The relationship between these forces was described by Klusener (11), who offered the following equation:

$$P_1 = P_a + P_\sigma = \text{constant} \quad (1)$$

where P_1 is the internal pressure at any point on the droplet surface, P_a is the external aerodynamic pressure due to the compressed air source, and P_σ is the pressure caused by the surface tension. From the equation, it is observed that a droplet can maintain its stability as long as the change in the air pressure, P_a , is balanced by a corresponding change in the pressure caused by surface tension, so that the internal pressure remain constant. When the aerodynamic pressure is greater than the pressure due to surface tension, the external pressure will deform the droplet in order to reduce the pressure by surface tension to maintain the constant internal pressure. This deformation will continue to break up the droplet into smaller particles (9).

For a formulation with low viscosity, such as aqueous formulation, the viscous force is relatively small compared with the surface tension force. Under such relative conditions, a dimensionless parameter can be obtained, as shown in Eq. 2 (8). For a given liquid with low viscosity, the critical condition for the droplet to break up occurs when the aerodynamic force equals the surface tension, expressed as

$$C_D \frac{\pi d^2}{4} 0.5 \rho V^2 = \pi d \sigma \quad (2)$$

where C_D is the coefficient of drag, d is the diameter of the droplet, ρ and σ are the density and surface tension of

the liquid, respectively. From the above equation one can calculate the maximum stable droplet diameter,

$$d_{\max} = \left(\frac{8\sigma}{C_D \rho V^2} \right) \quad (3)$$

and the critical velocity at which the droplet will disrupt

$$V_{\text{crit}} = \left(\frac{8\sigma}{C_D \rho d} \right)^{0.5} \quad (4)$$

The equation can also be expressed as the dimensionless parameter called the Weber number, as

$$We_{\text{crit}} = \left(\frac{\rho V^2 d}{\sigma} \right)_{\text{crit}} = \frac{8}{C_D} \quad (5)$$

which is the ratio of external aerodynamic force to surface tension force. Using the above equations, the maximum diameter at a particular critical velocity can be estimated. However, it should be noted that the estimated diameter results from only the process of primary atomization. Secondly, atomization will generally decrease and exaggerate the particle size distribution. Additionally, when a droplet impacts a baffle immediately after the primary atomization, the droplet diameter drops sharply. Baffles are used in nebulizer systems to help control the particle size distribution of the droplets so that generated aerosols are within a desired particle size range, usually with mean values that are less than 10 μm and preferably less than 5 μm , for effective deposition in the human lungs.

Effect of physical properties of formulation: The physical properties of the formulation, such as viscosity and surface tension, are important parameters that affect the process of nebulization. Surface tension can be considered as a consolidating influence that attempts to minimize the production of increased surface area, and the liquid viscosity exerts a stabilizing influence by opposing any change in the shape of droplets as they are produced (12). Studies (13) have shown that the droplet size is proportional to the liquid surface tension for low-viscosity liquids, while the density has only a small effect on the droplet size. The effect of viscosity on the droplet size seems to be complex. Research by Searls and Snyder (14) showed that high-viscosity liquids not only have increased nebulization time (decreasing the mass output), but also reduced mean particle size. However, Hinds et al. (15) reported little change in droplet size with viscosity ranged from 17 to over 100 cP using a Laskin aerosol generator operated at 20 psi.

An extensive study was conducted by McCallion et al. (16) investigating the effects of different physicochemical properties on the performance of the jet nebulizer.

Materials with different viscosity and surface tension were nebulized using three different nebulizers, Pari LC (Pari-Werk GmbH, Starnberg, Germany), Sidestream (Medic Aid, Pagham, United Kingdom), and Cirrus (Intersurgical Complete Respiratory Care, Workingham, United Kingdom). However, only data from the Pari LC nebulizer were presented in the paper. The particle size distribution of the aerosols was quantified using laser diffraction, which has been shown by Clack (17) to be a robust and reliable technique having good correlation with in vivo deposition data. Table 1 shows the experimental data for different formulations obtained using the Pari LC jet nebulizer. The mass median diameters were obtained using two nebulization flow rates, 6 and 8 L/min, respectively. The authors concluded that the more viscous fluids tended to produce smaller droplets. Within a specific range of the surface tension, of approximately 70 dyne/cm (70.0–72.9 dyne/cm), the mass median diameter decreased from 3.2 to 2.5 μm as the viscosity increased from 1.0 to about 6.0 centipoise. The same was true for surface tension of approximately 20 dyne/cm: as the viscosity increased, the particle size decreased. The effect of surface tension on particle size distribution was not as clear. However, the authors pointed out that a correlation between surface tension and mass median diameter appeared to exist when the fluid systems were analyzed separately. In glycerol and propylene glycol solutions, surface tension was directly proportional to droplet size, while the converse was true for silicone fluids. A trend was reported for the other two nebulizers tested, but no data were presented (18, 19).

Drug concentration effects: Because of the large volume of air used during nebulization, there is continuous evaporative loss of any solvent, such as water. This phenomenon has been described in several studies (20–23). Theoretically, the increase in drug concentration will continue until the solute concentration in the reservoir reaches the limit of drug solubility, at which point precipitation may occur. In practice, many drugs used in nebulizers (such as bronchodilators) are very hydrophilic, and the nebulizer solutions are formulated at concentrations well below the solubility limit (3), so there is little issue with precipitation. However, this results in changing particle size distribution with variable deposition patterns, and a change in the unit dose given per unit time. Collectively, this produces variable and unpredictable deposited doses for many drug products. The magnitude of the change in solution concentration is difficult to quantitatively predict and depends strongly on the formulation, nebulization condition, and type of nebulizers. Niven (12) has recently derived a mathematical equation based upon the early work by Mercer (21) to

Table 1 Average mass median diameter for fluids nebulized in a Pari LC nebulizer

Fluid	Viscosity (cP)	Surface tension (dyne/cm)	Mass median diameter (μm)	
			6 L/min	8 L/min
Water	1.00	72.8	3.6	3.1
Ethanol	1.19	24.1	3.0	2.5
Glycerol 10%	1.31	72.9	3.1	2.9
Glycerol 25%	2.09	72.2	2.6	2.4
Glycerol 50%	6.03	70.0	2.5	2.0
P. Glycol 10%	1.50	62.0	1.9	1.6
P. Glycol 30%	3.00	52.0	1.6	1.5
P. Glycol 50%	6.50	45.0	1.3	1.2
S.F. 200/0.65 cs	0.49	15.9	3.3	2.9
S.F. 200/1 cs	0.82	7.4	2.4	2.0
S.F. 200/5 cs	4.60	19.7	1.6	1.3
S.F. 200/10 cs	9.40	20.1	1.7	1.7
S.F. 200/20 cs	19.00	20.6	1.7	1.5
S.F. 200/50 cs	48.00	20.8	1.2	1.1
S.F. 200/100 cs	97.00	20.9	1.4	1.4

P. Glycol = Propylene glycol; S.F. = silicon fluid.
(Adapted from Ref. 16.)

describe the relationship of drug concentration in the nebulizer as a function of time. The drug concentration obtained by Mercer (21) was expressed as

$$C(t) = C_0 \left(\frac{V_0}{V_0 - (W + S)Ft} \right)^{(WW+S)}$$

(6)

where *W* and *S* are the solution output and solvent output per liter of air (ml/L air) respectively as shown in Fig. 3. *V*₀ is the initial fill volume into the nebulizer, and *F* is the nebulization flow rate. A more user-friendly version of the equation was derived (12) and expressed in terms of drug mass output that might be of more practical use in predicting drug output from a nebulizer for a range of operation conditions as claimed by the author:

$$M_{out} = M_0 \left[1 - \left(\frac{V}{V_0} \right)^{(WW+S)} \right]$$

(7)

where *M*₀ is the initial amount of drug in the solution, and *V* is the solution volume at any time, *t*, expressed as

$$V = V_0 - [W + S]Ft.$$

(8)

Newman et al. (24) evaluated the concentration change using four different jet nebulizers, with a gentamicin solution, which is used to treat patients for cystic fibrosis. The types of nebulizers used were the Bird micronebulizer, DeVilbiss 646, Bard Inspiron mini-neb, and Medic-Aid Upmist. All were operated at four different air flow

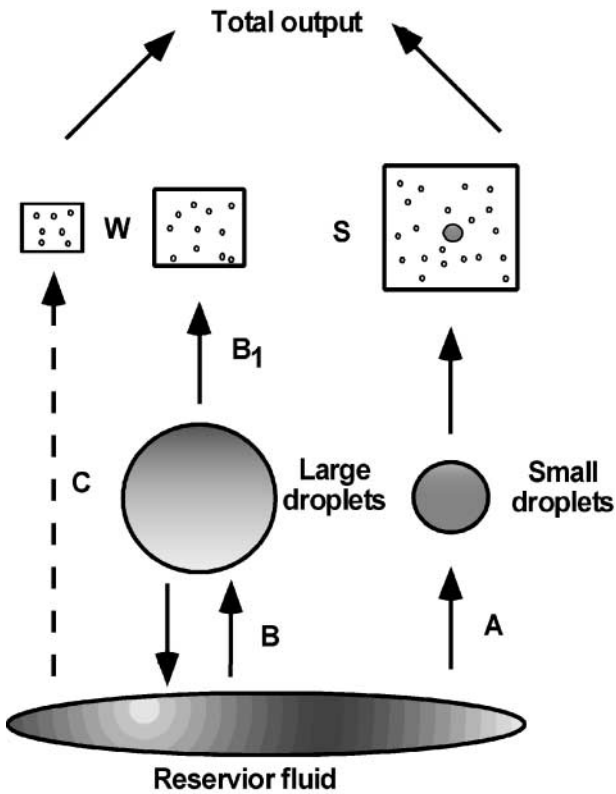


Fig. 3 Schematic diagram of nebulizer output. Droplets small enough will be carried out by air stream (route A), and the large droplets will be either recycled to the nebulizer reservoir (route B) or carried out by the outgoing air stream (route B₁). Some of the solvent will be evaporated (route C). (Adapted from Ref. 12.)

rates: 6, 8, 10, and 12 L/min. Nebulization time was defined as 30 s after the last visible release of aerosol. The concentration of the drug solution left in the nebulizers was then measured using an Osmometer. The initial drug concentration was 40 mg/ml for both fill volumes. Table 2 shows the final concentration of gentamicin solution at different nebulization flow rates.

It was found that the ratio of final to initial drug concentration ranged from 1.35 to 1.70 for the 2 ml fill volume, and from 1.45 to 1.96 for the 4 ml fill volume. The ratio did not vary significantly between the four different nebulization flow rates, but was significantly higher ($p < 0.01$) for the 4 ml fill volume than for the 2 ml fill volume, indicating that the larger volume allowed more time to concentrate the reservoir drug solution. The same phenomenon was observed using a albuterol solution, and the drug concentration increased after the nebulization (25). It was also found that the concentration ratio decreased with the increase in solution fill volume, which is different from the results shown in Table 2. However the discrepancy may result from the different nebulization sources, where dry oxygen versus an electric powered air compressor were used in the experiment (25).

Temperature effects: It is well known that the nebulizer solution cools during nebulization, because a large amount of compressed air is used to generate aerosol particles (20, 21). The temperature of the nebulizer solution decreases several degrees (5–10°C) due to the evaporation of solvent in the solution during the first several minutes of nebulization until a new heat balance is established. Some cooling of the solution in the nebulizer due to adiabatic expansion of the compressed air source may be absorbed by the thermal energy released by break-

up and impact of the liquid (26). The magnitude of temperature change in the nebulization solution depends on many factors, such as heat capacity, properties of nebulizer material, ambient conditions, and the humidity and temperature of the compressed air. Fig. 4 shows the change in solution temperature as a function of nebulization time (20). Four different types of nebulizers were used together with different air compressors to aerosolize a sodium chloride solution. The solutions in the nebulizers were initially at ambient temperatures (approximately of 23°C).

Different nebulizers had different temperature reduction profiles, depending on the compressed air flow rate applied during the nebulization, as pointed out by the authors. The solution temperature falls to a steady value, T_s , which is 5–6°C below the ambient temperature at the nebulization flow rate of 6.3–5.0 L/min, and 11–15°C at 8 L/min air flow rate.

Drug aerosol delivery efficiency and particle size distribution: Delivery efficiency of aerosolized drug product and particle size distribution are the two most important parameters in assessing the performance of a nebulizer for aerosol therapy. The aerosol output from the nebulizer depends mainly on the nebulizer type, drug formulation, solution volume placed in the nebulizer, and the applied air pressure (and hence the airflow through the nebulizer). The simplest method to calculate the output from a nebulizer is to weigh the nebulizer before and after the nebulization, often expressed as the solution volume output per unit time (i.e., mg/min). However, this method of calculation is subject to at least two possible error sources: 1) neglecting the change in solution density when it is used to transfer the mass output per unit time to

Table 2 Final drug concentration (mg/ml) in the nebulizer reservoirs

Nebulizer type	Nebulization flow rate			
	6 L/min	8 L/min	10 L/min	12 L/min
Solution fill volume: 2 ml				
Bird	54.4	62.0	57.2	58.4
DeVilbiss	55.6	60.4	56.0	65.6
Inspiron	58.4	65.2	68.0	68.0
Upmist	64.4	58.0	58.0	58.0
Solution fill volume: 4 ml				
Bird	59.2	65.6	58.0	64.0
DeVilbiss	78.4	77.2	71.6	75.2
Inspiron	71.2	69.2	74.0	74.8
Upmist	76.8	75.6	62.8	62.8

(Adapted from Ref. 24.)

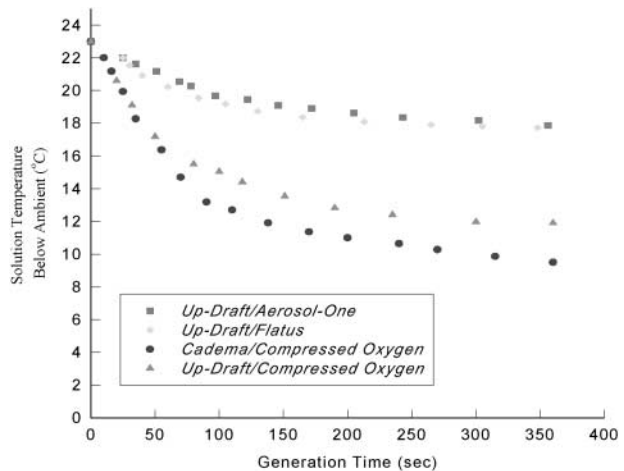


Fig. 4 Nebulizer solution temperature below ambient (23°C) vs. nebulization time. (Adapted from Ref. 20.)

volume output per unit time, and 2) neglecting the change in solution concentration as discussed previously. The delivery efficiency of the nebulizer can be expressed simply as

$$E = \frac{M_p - M_d}{M_p} \quad (9)$$

where E represents the nebulizer delivery efficiency, M_p is the drug amount loaded into the nebulizer before nebulization, and M_d is the drug amount left in the nebulizer after the nebulization. Nebulizers are often operated until there is no visible aerosol or “sputtering,” and a significant reduction in the nebulizer output occurs when the solution volume in the nebulizer reservoir becomes so low that it fails to maintain a continuous liquid supply to the nebulizer orifice. However, these are very crude measures and provide no indication of the amount of aerosolized drug a patient inhales.

The delivery efficiency, or the nebulizer output, can be expressed in many different ways, as widely documented in the literature. Sometimes it is represented as the volume output, or solution mass output. However, it is more practical to use the drug mass emitted from the nebulizer at or near the mouthpiece to estimate the nebulizer output, because the amount of aerosolized drug mass at this point in the system is the best measure of how much drug the patient has available to inhale.

Nebulizer output depends mainly on three variables: the type of the nebulizer, the operating conditions during nebulization, and the solution to be nebulized. Although solution density and viscosity play minor roles in nebulizer output for the majority of aqueous formulation, viscosity becomes an issue when aerosolizing highly viscous solutions. For any given solution and nebulizer, the operating pressure primarily determines the nebulizer output. Increasing the nebulization pressure usually increases nebulizer output. Increasing nebulizer output typically has a clinical benefit: it can reduce the time required to nebulize an effective dose, particularly for young patients who are less tolerant of the treatment procedure. As drug mass delivered per unit time is increased, treatment time is reduced leading to better acceptability and improved clinical outcome (27). Fig. 5 correlates drug mass output as a function of operating pressure (28).

A Pari LC Plus jet nebulizer (Fig. 6) was used in the study. The nebulizer was operated using a specially designed electronic unit to regulate the compressed air supplied to the nebulizer. The nebulizer was operated in a pulse mode, in which the time of nebulization was set at 3 s for each aerosol generation. The drug output was generally

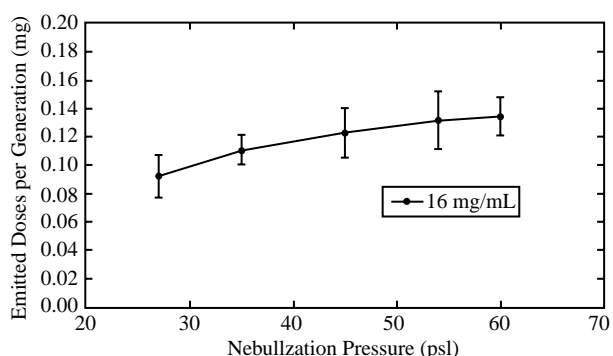


Fig. 5 Emitted dose per generation from the nebulizer vs. the nebulization pressure. (Adapted from Ref. 28.)

greater with each incremental increase in operating pressure. These increases were significant ($p < 0.05$) over a range of nebulization operating pressures of 27–45 psi. Additional increases in operating pressure of 45 to 60 psi did not significantly increase aerosol drug output.

Finlay et al. (29) recently tested 15 different jet nebulizers available commercially to investigate the performance of nebulizers. The drug compound used in the study was a 2.5 ml unit dose of Ventolin® (containing 1 mg/ml of salbutamol sulfate), which corresponds to a 2.5 mg in nominal dose per drug container. Table 3 shows the experimental results for the different nebulizers.

Nebulizers fitted with a T-mouthpiece have an unrestricted flow of ambient air passing through the nebulizer output, supplying inhaled air flow, which effectively increases drug output. In vented nebulizers,



Fig. 6 The diagram of the Pari LC Plus jet nebulizer. (Courtesy of Pari GmbH.)

Table 3 Mass output of different nebulizers

Nebulizer name	Supplier	Flow rate (L/min)	In vitro inhaled dose		Methods ^a
			(mg)	% Nominal dose	
AirLife	Baxter, Valencia, CA	5.5	0.60 ± 0.02	24.1 ± 0.9	A
Disposable Sidestream	Medic-Aid, Pagham, U.K.	6.7	0.35 ± 0.01	13.8 ± 0.8	B
Sidestream	Medic-Aid, Pagham, U.K.	8.0	0.35 ± 0.02	24.1 ± 0.9	B
Ventstream	Medic-Aid, Pagham, U.K.	8.0	0.59 ± 0.01	23.6 ± 0.5	B
LCJet+ (ProNeb)	Pari, Richmond, VA	3.6	0.60 ± 0.02	23.9 ± 0.8	B
LCJet+ (Pulmo-Aid)	Pari, Richmond, VA	4.8	0.65 ± 0.02	26.0 ± 0.6	B
Pulmo-Neb	DeVilbiss, Somerset, PA	6.2	0.60 ± 0.02	24.0 ± 0.8	A
646	DeVilbiss, Somerset, PA	7.8	0.14 ± 0.02	5.5 ± 0.9	B
Raindrop	Puritan Bennett, Lenexa, KS	6.2	0.56 ± 0.02	22.3 ± 0.9	A
Salter	Salter Labs, Arvin, CA	6.1	0.65 ± 0.02	25.9 ± 0.6	B
T-Updraft	Hudson RCI, Temecula, CA	6.1	0.34 ± 0.02	13.8 ± 0.6	A
T-Updraft II	Hudson RCI, Temecula, CA	5.7	0.74 ± 0.02	29.7 ± 0.9	A
Up-Mist	Hospitak, Lindenhurst, NY	7.5	0.48 ± 0.02	19.1 ± 0.9	A
1405	Avion, Burlington, Ont.	6.2	0.47 ± 0.02	18.5 ± 0.9	A
1807	Avion, Burlington, Ont.	6.5	0.50 ± 0.02	19.8 ± 0.8	A

^aA is a conventional T-mouthpiece, and B is a vented nebulizer.
(Adapted from Ref. 29.)

the inhaled air must flow through the droplet production region. Therefore, the breathing pattern of the patient has an effect on the aerosol characteristics produced by such devices. The delivery efficiency of the commercial nebulizers tested ranged from 6 to 30% of the nominal dose loaded into the nebulizers. For conventional nebulizers, the inhaled dose is approximated by the amount of drug leaving the nebulizer divided by 2, since inhalation is assumed to occupy half of the breathing cycle (29). However, this remains a very crude measure of delivered or inhaled dose.

The importance of particle size distribution in determining the site of the deposition of pharmaceutical compounds within the respiratory tract has been discussed extensively in the literature (1–7, 10, 12, 20, 24, 27, 30). For any given inhaled dose of aerosolized drug, the particle size distribution of drug aerosol, aerosol velocity, the breathing pattern (inspiratory flow and volume), and lung morphology, determine the location the aerosol is deposited in the respiratory airways. It is generally well accepted that the particle size of drug aerosols generated for inhalation should be within 1–5 μm in diameter. Because of polydispersity of the aerosol, the mass median aerodynamic diameter (MMAD) and geometrical standard deviation (GSD) are often used to describe the distribution of aerosol particles. The MMAD is defined as the mass median of the distribution of mass with respect to aerodynamic diameter (1). The mass median diameter

(MMD) and GSD of the salbutamol aerosol generated by 15 different nebulizers (29) are shown in Table 4, where $\text{MMD} = \text{MMAD}/(\text{density})^{1/2}$.

Table 4 Particle size distribution of different nebulizers

Nebulizer name	Flow rate (L/min)	Particle size distribution	
		MMD (μm)	GSD
AirLife	5.5	5.8 ± 1.0	1.7
Disposable Sidestream	6.7	4.2 ± 0.2	1.5
Sidestream	8.0	4.3 ± 0.4	1.5
Ventstream	8.0	4.7 ± 0.4	1.5
LCJet + (ProNeb)	3.6	6.2 ± 0.4	1.7
LCJet + (Pulmo-Aid)	4.8	6.2 ± 0.3	1.7
Pulmo-Neb	6.2	5.6 ± 0.2	1.6
646	7.8	6.8 ± 0.6	1.7
Raindrop	6.2	5.0 ± 0.4	1.6
Salter	6.1	6.1 ± 0.9	1.7
T-Updraft	6.1	5.3 ± 0.5	1.7
T-Updraft II	5.7	5.9 ± 0.3	1.7
Up-Mist	7.5	4.5 ± 0.7	1.6
1405	6.2	5.0 ± 0.6	1.6
1807	6.5	6.1 ± 0.5	1.6

(Adapted from Ref. 29.)

The size distribution of aerosols generated using most commercial nebulizers is typically highly variable with somewhat large mean diameters. This not only results in poor transfer efficiency but poor and highly variable respiratory deposition patterns between and within patient populations.

Metered dose inhalers

Pressurized metered dose inhalers (pMDIs) have been widely used in the treatment of asthma and COPD. These devices are more popular than nebulizers and dry powder inhalers. The popularity of the pMDI has steadily increased since the 1950s due mainly to its compact and portable size, self-contained power source, and other rapid therapeutic effects in the use of bronchodilators. pMDIs appear to be simple to use, although this appearance is in fact deceptive. Many patients are not adequately trained to correctly use pMDIs or find it nearly impossible to effectively coordinate their breathing with the high velocity release of the aerosol. High ejection velocity of the aerosol at the mouthpiece (usually near 30 m/s) is the single largest deficit of pMDIs. At these high ballistic velocities, most of the drug aerosol impacts in the mouth or throat and can not be readily entrained within the respiratory flow (even when coordinated with aerosol release). Typically, about 60–90% of the emitted drug will be deposited in the oral-pharyngeal region. This not only reduces the amount of drug that reaches the intended action site, but often contributes to unwanted systemic side effects due to oral absorption. Spacers have been developed that help mitigate this problem. The spacers are placed in front of the pMDIs actuator to reduce the aerosol velocity, and reduce the number of large diameter droplets before entering into the airway. However, this occurs at the expense of delivery efficiency, further reducing the dose of drug inhaled.

A typical pMDI consists of a drug reservoir containing the drug compound suspended or solubilized in liquefied propellant, a metering valve, and an actuator that connects with the metering valve through a stem. The propellants historically have been chlorofluorocarbons (CFCs) which

have high vapor pressures and are generally biologically inert. The pMDI delivers a metered dose of active drug compound using the propellants as a power source. In the rest position, the metering chamber is connected with the formulation reservoir. Depressing the drug canister activates the metering chamber and the formulation reservoir is then closed. Simultaneously, a connection is opened between the valve stem orifice and the metering chamber. The formulation contained in the metering chamber quickly escapes through the stem valve and spray orifice, atomizing the formulation. As the liquid CFCs vaporize, they generate sufficient energy to produce efficient two-phase atomization, without the need for an external power source (7).

The primary factors that will influence the performance of the pMDI are the propellant used, the physical properties of drug formulation, which partially governs droplet formation and drug uniformity, and the device itself.

Propellants: The propellants used in the pMDIs consist of CFCs with different molecular formulas. The propellants commonly used in inhalers to disperse the liquid into aerosols are CFC-11 (CCl_3F), CFC-12 (CCl_2F_2), and CFC-114 ($\text{CClF}_2\text{-CClF}_2$) (31, 32). These three propellants have been used for decades in pMDI products because of their unique properties, which include low toxicity, excellent physical and chemical stability, ideal vapor pressure, and generally good compatibility with drug compounds and surfactants. These propellants possess an appropriate degree of solvency that solubilizes the surfactants typically used in pMDI systems such as oleic acid, sorbitan trioleate, and soya lecithin, yet dissolve low quantities of most drugs. This combination of surfactant solubility and drug insolubility in the pMDI formulation is essential for maintaining uniform suspensions of appropriate particle size. The physical properties of these three propellants are shown in Table 5.

The vapor pressure of the formulation in a pMDI can be adjusted to a targeted value by mixing propellants with different vapor pressures. Propellants offer the additional advantage of having a narrow density range that may be adjusted to balance the density of the drug particle in suspension (33).

Table 5 Physical properties of CFCs propellants commonly used in pMDIs

Propellants	Molecular weight	Density (g/cm ³) at 21°C	Vapor pressure (psig) at 20°C	Boiling point (°C) at 1 atm	Atmospheric life (years)
CFC-11	137.4	1.49	−1.8	23.8	75
CFC-12	120.9	1.33	67.6	−29.8	111
CFC-114	170.9	1.47	11.9	3.6	200–300

Morén (32) investigated the influence of vapor pressure on the characteristics of aerosols generated using pMDIs. In these studies, nine healthy volunteers inhaled eight doses at one-minute intervals coordinated with the dose firing, and the amount of drug deposited in the actuator, extension tube, and mouth was determined spectrophotometrically. It was found that an increase in vapor pressure results in higher initial velocity of the aerosols, smaller initial aerosol particle size, and more rapid propellant evaporation. The amount of drug deposited in the actuator increased at the higher vapor pressure of 502 kPa, which was likely due to the high velocity of the aerosol exiting the spray nozzle and a wide spray angle.

Aerosol formation: Aerosol formation from pMDIs is a complex process influenced not only by the propellant as discussed above, but also by properties of the formulation and device design (e.g., valve volume and orifice size). After aerosol droplets form at the exit of the spray nozzle, an aerosol plume begins to expand, and large particles travel along the axis of the actuator. Fig. 7 shows photographs of aerosol formation using a commercially available pMDI.

Most large droplets fall out from the aerosol cloud, and those droplets are deposited in the mouthpiece or in the oral-pharyngeal region of the respiratory airways when the inhaler is used by a patient. When spacers are used, many large particles are collected within the spacers. The remaining aerosol cloud continues to evaporate, and the

velocity of the aerosol reduces dramatically from about 35 m/s at 5 cm from the spray orifice to about 10 m/s at 20 cm (34). During aerosolization, surfactant molecules, used as dispersing agents and valve lubricants, will influence the evaporation rate of the propellants (33). The dimension and shape of the aerosol cloud significantly influence the transport efficiency of aerosol to the respiratory tract.

Physical stability of drug suspension: The physical stability of the nonpolar drug suspensions formulated in pMDIs again depends on many parameters, such as propellant type, chemical properties of the drug compounds, and the surfactants used in the formulation. A suspension is a liquid system in which insoluble solid particles are dispersed in a liquid medium. Suspensions can be divided into colloidal suspensions, in which the solid particles are less than 1 μm , and coarse suspensions, where particles are larger than 1 μm (35). The physical stability of a suspension can also be classified according to whether the formulation is flocculated or deflocculated. In a deflocculated system, the solid drug particles settle down slowly and form irreversible aggregates. As a result the particles form a distinct cake on the bottom of a container. The deflocculated system is impossible to redisperse by simple shaking, and the suspension is unlikely to be of use after caking. Therefore, it is necessary to develop the flocculated formulation systems for pMDIs. In flocculated systems, the particles form large aggregates, so that the

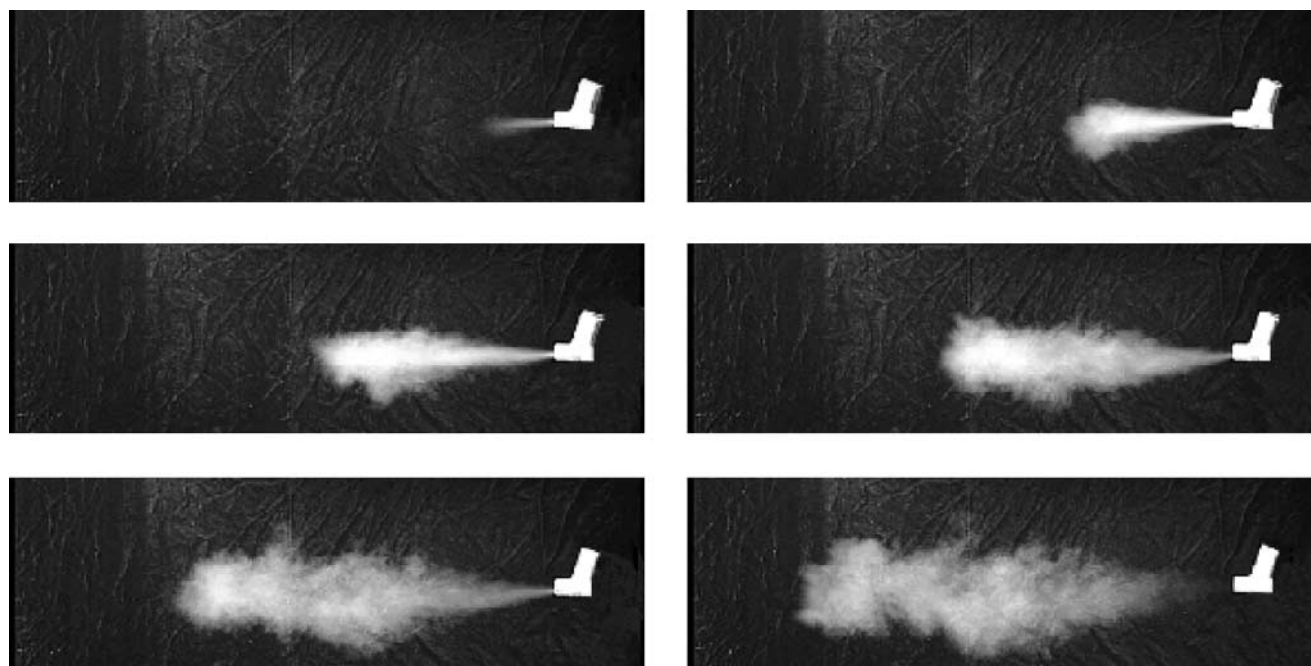


Fig. 7 Formation of the aerosol plume as a function of time. Each frame has a time interval of 30 mins.

solid particles can be resuspended upon shaking. The formulation of a flocculated system requires reducing or neutralizing the zeta potential around the drug compound. This means that repelling forces between the particles are neutralized and the particles can then interact with each other to form large aggregates which rapidly settle but, can be easily redispersed (36). In formulating a suspension for pMDI use, the drug particles are usually less dense compared with the suspension media. Therefore, in an unshaken inhaler the drug gradually separates as a layer, which floats on the propellant and surfactant mixture, and is resuspended immediately after shaking. The solubility of the drug compound in the propellant also plays an important role in the physical stability of the pMDI formulation. An extremely low solubility of drug compound is required in the nonpolar system. Therefore, pMDI formulations usually require a polar salt to be incorporated, because significant drug solubility in the propellant can lead to crystal growth, often associated with the caking of the suspension and a reduced amount of fine particles (37).

Sedimentation ratio is often used to assess suspension stability. Byron (37) reported the sedimentation ratios for a 1% sodium fluoresein suspension formulation with different amounts of surfactant (sorbitan trioleate) after standing for 20 days at room temperature. The suspension formulation with the lowest sedimentation ratio had the best-flocculated system. However, all formulations were easily redispersible: one complete revolution of the container was sufficient to produce a homogeneous dispersion. There was no clear difference in the times taken to reach apparent sedimentation equilibrium. Physical stability of the formulation was determined according to:

- Whether the drug material was adequately deaggregated during manufacture,
- Whether redispersibility was easy and allowed doses to be metered reproducibly,
- Surfactant concentration-drug combination gave the greatest respirable fraction of the aerosols (37).

Determination of dose and particle size distribution:

The amount of drug reproducibly delivered from pMDIs per actuation is one of the most frequently reported performance parameters of these products. Several experimental configurations are used to characterize the pMDI's emitted dose. One of the most commonly used experimental protocols is described in the Official Compendia of Standards published by *U.S. Pharmacopeia* (38). The pMDI device is fitted with a sampling tube at one end, and a vacuum pump is connected with a filter holder at the other end. A continuous airflow is

drawn through the sampling tube to avoid the loss of the drug material into the atmosphere. The active compound delivered from the pMDI is recovered from the filter and the sampling tube to determine the mass output from the device. There are no universal definitions for different terminologies used to express dosages provided by pMDIs, and the labeling requirements for pMDIs differ among nations (37). In the United States, the emitted dose (label claim) is defined as the amount of drug leaving the device, excluding the drug mass left in the actuator and valve stem of the device. However, in other nations, the metered dose is often used, which includes the amount left in the actuator but not that retained in the valve stem (37). Descriptions of the emitted dose often use different terms, which are often used for different purposes (39, 40). Because of the lack of consistency in the terminology used in the literatures and by regulatory agencies, it is often difficult to compare the experimental results between different research groups, devices, and experimental methods and procedures. Therefore, it would be beneficial to harmonize the terminology commonly used to describe drug doses from different parts of the pMDIs, and efforts are currently underway to standardize the way delivered dose and device performance are measured and reported. Table 6 shows the definitions for the various doses frequently used in the literatures.

However, even with a clear understanding of the dose delivered by a pMDI, there is no direct or consistent correlation to the amount of drug delivered to the patient's lungs. A typical drug distribution within patients and from devices is shown in Table 7 (41). The table is based on an experiment conducted on healthy volunteers with an average FEV₁ at 105% of predicted (range 92–119%). A terbutaline formulation was labeled with ^{99m}Tc and administered to patients using a pMDI. The average inspiratory flow rate was 33.8 L/min, (SD = 6.7), and the total inspired volume averaged 2.86 L (SD = 0.87), with an average of (SD = 0.7) 9.1s breath-hold. The inter-subject variability in the patients' breathing patterns was relatively small, contributable to on-line monitoring of breathing patterns using a spirometer.

The amount of drug deposited in the lungs expressed as percentage of emitted dose (including drug left on actuator) was about 16.7% (SD = 9.6), and the percentage of emitted dose deposited in the oral-pharynx was 68.3% (SD = 10.2). The percentage of emitted dose deposited in the actuator was 13.4% (SD = 2.1). The intersubject variability for the percentage of emitted dose found in the actuator and oral-pharynx was 15.3 and 14.9%, and the variability for the total lung deposition was 57.7%. These

Table 6 Definition of various doses

Terminology	Description
Nominal dose (metered dose)	The dose represents the actual amount of the drug loaded into the pMDI device. It can be calculated using drug concentration times the metering valve volume.
Emitted dose (ex-valve dose)	The amount of drug emitted from the device, including the drug deposited on the actuator and inhaled dose.
Inhaled dose (label claim)	The amount of drug emitted from the device, not including the drug left on the actuator. This amount of drug is available for inhalation by patient.
Fine particle dose	The amount of drug with particle diameter less than 5.0 μm , which is a portion of the inhaled dose.
Deposited dose	The amount of drug deposited in the respiratory airways, which can further be divided into the drug amount in the oropharyngeal region and lungs.

large variations in the total lung deposition are typical of clinical results when inhaled drugs are given using pMDIs.

The particle size distribution of drug aerosol generated from the pMDI is another important performance parameter of the drug delivery device, which depends on the combination of the propellants and surfactants, the original size of the drug particle, and the design of the device (metering valve, orifice, actuator). The initial atomization process and subsequent spray evaporation govern the amount of drug aerosol deposited both in the oral-pharyngeal and pulmonary regions. Aerosol particles ejected from the pMDI experience constant changes in size due mainly to the drug formulation. Fig. 8 shows the change in particle size as a function of the distance from the actuator orifice. The experiment was conducted using a pMDI obtained commercially, and a laser diffraction instrument (Malvern, Mastersizer X) was used to determine the size of drug aerosol. The device was

mounted on a platform, and the distances between the actuator and the center of the laser beam were chosen to be about 5, 10, 15, 20, and 25 cm. The pMDI was operated according to the instructions given by the manufacturer, and five measurements were repeated for each chosen distance.

The size of the drug particle was about 8 μm at 5 cm between the actuator orifice, and gradually decreased to about 2.5 μm at 35 centimeters from the actuator orifice due to the evaporational shrinkage. The particle diameter was about 6.5 μm at a distance of 10 centimeters, which was about the distance to the oral-pharyngeal region from the actuator orifice. The amount of drug deposited in the oral-pharyngeal region is mainly due to the inertial impaction; therefore it depends more on particle size than on particle velocity. The inertial impaction is proportional to the square of the particle diameter, and is linearly proportional to the particle velocity. The decreases in particle diameter followed exponentially with the distance

Table 7 Deposition of radioactive drug after inhaling via pMDI

Subject	Actuator	Throat	Total lung	Regional lung			P/C ratio	Exhaled
				Central	Intermediate	Peripheral		
1	15.5	76.3	7.1	1.7	2.4	3.1	1.87	1.1
2	10.4	81.8	6.1	1.2	1.9	3.0	2.54	1.6
3	15.6	66.2	12.5	2.9	4.3	5.4	1.86	5.8
4	12.8	75.6	10.9	3.2	3.9	3.8	1.2	0.8
5	11.7	66.3	21.8	6.5	7.1	8.2	1.27	0.2
6	11.8	57.6	30.0	7.8	9.3	12.9	1.65	0.6
7	13.1	71.7	15.0	4.1	5.5	5.4	1.31	0.3
8	16.0	51.0	30.0	6.3	11.2	12.6	2.00	3.0
Mean	13.4	68.3	16.7	4.2	5.7	6.8	1.71	1.7
SD	2.1	10.2	9.6	2.4	3.3	4.0	0.45	1.9

(Adapted from Ref. 41.)

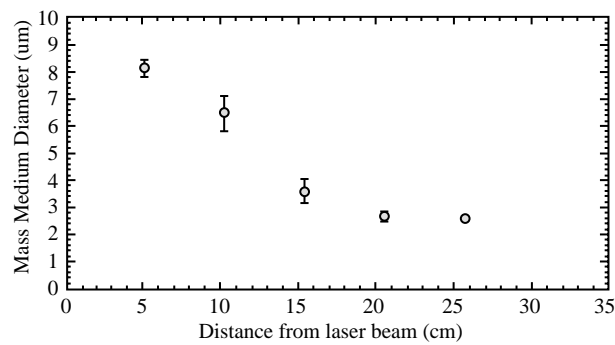


Fig. 8 The change in particle size vs. the distance from the actuator orifice.

from the actuator orifice. The exponential decrease rate of particle diameter differs with different drug formulations (vapor pressure, nonvolatile concentration, etc.).

High Frequency Oscillating Devices

Ultrasonic nebulizer

Ultrasonic nebulizers utilize high frequency sonic energy to convert liquid drug formulations into aerosol clouds. Although these devices have been used in clinical settings since the 1960s, they are not as widely used as the jet nebulizer and pMDIs. Ultrasonic nebulizers typically have higher drug mass output and significantly larger particle size distributions at the point of aerosolization compared with jet nebulizers (42). Advantages and disadvantages of the ultrasonic nebulizer for the delivery of pharmaceutical compounds to the human respiratory airways are listed in Table 8. The output of drug compound leaving the ultrasonic nebulizer is at a relatively slow velocity compared with the jet nebulizer. Therefore, less patient coordination is required during treatment.

The reservoir temperature in the ultrasonic nebulizer increases rapidly during operation. In the study by Phipps et al. (20), the solution in the nebulizer increased from room temperature (~23°C) to greater than 30°C after

about 5 min of operation, and exceeded 40°C after 30 min of nebulization. This significant temperature increase may cause some pharmaceutical compounds to be degraded during aerosolization (43). Changes in drug potency due to delivery by ultrasonic nebulizer have been reported in the literature, with degradation more profound for protein and peptide therapeutics compared to more stable small molecules.

Mechanism of aerosol generation: Ultrasonic nebulizers generate aerosols using a piezoelectric transducer that vibrates when subjected to high-frequency electric energy. When liquid is fed continuously through or over the vibrating surface, a wave pattern is formed as the liquid spreads over the liquid surface. Liquid droplets are formed when the amplitude of the vibrating surface increases to the point at which the capillary wave becomes unstable and collapses (44). An empirical equation to calculate the droplet diameter (*D*) produced from an ultrasonic nebulizer was described by Lang (44). The diameter is approximately proportional to the capillary wave length (λ), density (ρ) and surface tension (σ) of the liquid, and the excitation frequency (*F*).

$$D = 0.34\lambda = 0.996 \frac{\sigma}{\rho F^2}.$$

(10)

As with jet nebulizers and pMDIs, the formulation characteristics include liquid density, surface tension, viscosity, and vapor pressure. Many of these that effect nebulizer performance of formulation properties are described in the above equation. The configuration of the device, such as the location of the baffles, mechanisms in the transport of the liquid formulation to the piezoelectric transducer, and the size of medication chamber, also affect the performance. The units frequency and amplitude of vibration, control the mass output and the particle size distribution permitting some adjustment in aerosol characteristics.

Performance characteristics: Greenspan (42) summarized the performance of commonly used ultrasonic nebulizers. The excitation frequency ranged from 1.3 to about 2.6 MHz, and the MMAD of the produced aerosols

Table 8 Characteristics of ultrasonic nebulizer

Advantages	Disadvantages
Little patient coordination	Large particle size and less efficient
Aerosol exited with low velocity	Microbiological contamination risk
Small dead volume and quite operation	Possibility of degrading drug compound
High dose output and fast drug delivery	Poor portability and is expensive
No propellants requirements	Requirement of electric power supplier

was approximately 1–7 μm . Several ultrasonic nebulizers generated relatively large aerosol particles, which were usually associated with low excitation frequency. Hager et al. (45) characterized the performance of two ultrasonic nebulizers using a pentamidine solution, which is used to treat *Pneumocystis carinii* pneumonia associated with acquired immunodeficiency syndrome (AIDs). The ultrasonic nebulizers used in the study were Fiso Neb (model FZV 40 BAMKI, Fisons) and Porta-Sonic (model 8500GB, Devilbiss). It was found that the mass output for FISO Neb at airflow rate of 6 L/min were 201.4 and 36.7 mg for drug concentrations of 50 and 10 mg/mL, respectively. Corresponding outputs decreased to 85.2 and 23.6 mg for Porta-Sonic nebulizer. However the MMD of aerosol generated from the Fiso Neb was 5.8–6.93 μm , and 1.96–3.04 μm for Porta-Sonic nebulizer. The higher mass output from the Fiso Neb compared to the Porta-Neb was due to the presence of larger aerosol particles produced by the Fiso Neb nebulizer. However, these larger droplets are typically nonrespirable, if they are transported at all to the patient inhalation port. Therefore, the overall dose efficiency of ultrasonic nebulizers tend to be poorer than jet nebulizers.

ADVANCED DRUG DELIVERY TECHNOLOGIES

Using the various drug delivery devices, such as pMDIs, nebulizers, and DPIs, discussed in the first half of this chapter, has had an enormous impact on patient quality of life, as inhaled drugs play an important role in the management of airway diseases. However, past successes in topical delivery of drugs to the lung owe more to the therapeutic properties of the drugs used than to the delivery device itself (30). Currently used delivery devices tend to be inefficient, with small amounts of drug delivered to the site of action, with a large portion of the drug compound lost in the device or deposited orally, often causing unwanted side effects.

The shortcomings described for the current devices limit their ability to meet the demands created by advancements in more potent small molecules and biotechnology therapies. More and more new drugs are potential inhalation candidate therapies to treat various diseases, including nonairway diseases such as diabetes. There is an urgent need for more advanced drug delivery technology, which can deliver the drug more precisely to the targeted area for action, with better efficiency and improved reproducibility. Over the last 10 years, many innovations have occurred in the field of inhalation drug delivery with advances in aerosol technology and electronics, spurred by the demand created by new drug

compounds. Most of the technologies have tried to tackle problems such as reducing the ballistic velocity of aerosol released from pMDIs by breath activated functions and developing new propellants with lower ejection pressures. Focus has shifted to defining and developing a more “ideal” inhaler to administering various pharmaceutical compounds for different diseases. The following describes many of the features these new pulmonary drug delivery devices are designed to include (30).

- Aerosol generation independent of the patient's inhalation.
- Duration of aerosol dose generation should occupy a substantial part of a slow inhalation cycle. A generation time of greater than 1 s will permit the patient to better coordinate aerosol delivery effectively during inhalation.
- The aerosol cloud should consist largely of particles of less than 5 μm in size. Fine aerosols are necessary for delivery to the lung periphery if this is the site of action or if systemic delivery is the objective. The ability to vary and then control the mean particle size distribution of the aerosol would be an added advantage for targeting deposition within different regions of the respiratory tract.
- The velocity of the aerosol should be minimal to reduce oral-pharyngeal deposition and provide maximum delivery to the lungs.
- The inhaler should be simple to use, require little patient coordination and preferably should be breath-actuate.
- The inhaler should have dimensions similar to those of the pMDIs, fitting the pocket or the handbag and permitting discrete use.
- The inhaler should contain a large number of doses (>50 doses).
- A dose counter should be included.
- The device should be easy to manufacture and reasonably priced.
- The device should generate uniform doses throughout its life, be resistant to microbial and other contamination, and have a suitable shelf life.
- Propellants should be avoided, or designed for low environmental impact.

Pressure Driven Devices

New generation pressurized metered dose inhaler

Asthma affects an estimated 100 million people worldwide, including over 15 million in the United States (45), and while the pMDI has many disadvantages, it still is the most prescribed drug product for the treatment of asthma. The long-term dominance of pMDIs in inhalation drug

Table 9 Physical properties of propellants (CFCs and HFAs)

Propellants	Density (g/cm ³) at 21°C	Vapor pressure (psig) at 20°C	Boiling point (°C) at 1 atm
CFC-11	1.49	– 1.8	23.8
CFC-12	1.33	67.6	– 29.8
CFC-114	1.47	11.9	3.6
HFA-134a (replace CFC-12)	1.21	81	– 27
HFA-227 (replace CFC-12, CFC-114)	1.41	43	– 17

delivery is mainly due to its convenience. However, it is extremely difficult to use effectively. Everard (46) pointed out that the pharmaceutical industry elected to perpetuate the failings of the pMDI and develop replacements for CFCs, rather than develop novel devices to meet current demands. Extensive resources have been invested to preserve the pMDI when the industry was challenged with phasing out CFC propellants, and to reduce the oral-pharyngeal deposition by methods such as using attached spacers and breath-actuated devices.

pMDIs with new propellants: Since 1974, the CFCs have been believed to contribute to erosion of the earth's stratospheric ozone. Coyne (47) summarizes the history and progress of the Montreal Protocol negotiations and formation in great detail, and introduces the objectives and history of the Pharmaceutical Aerosol CFC Coalition (PACC) in the United States and the International Pharmaceutical Aerosol Consortium (IPAC) in Europe. The objectives of the organizations are to find replacements for CFC propellants that can be used in pMDIs.

Reformulation of drugs used in pMDIs to avoid CFC propellants presented numerous obstacles, because no approved alternative pharmaceutical propellants were available (48). Replacement of the propellants used in pMDIs requires changes in many fundamental aspects of formulation, device design, device components, and manufacturing. Several classes of propellants have been considered as alternatives for use in pMDIs, but each has its disadvantages, as shown below, in a listing based on Leach (49).

Alternative propellant	Disadvantage(s)
Dimethyl ether	Flammable
Hydrocarbons	Flammable, high volatile organic compound content
Hydrochlorofluorocarbons	Some ozone depleting potential
Compressed gases (e.g., carbon dioxide and nitrogen)	Inconsistent pressure throughout the product life

So far, two alternative propellants, tetrafluoroethane (HFA-134a, CF₃CH₂F) and heptafluoropropane (HFA-227, CF₃CHFCF₃), have been successfully tested and developed as propellants in pMDIs. The physical properties of these two propellants are compared with those of CFCs in Table 9.

There are many similarities between pMDIs formulated with CFCs and HFAs, such as the operating principles and basic components. However, due to differences in the physical properties such as vapor pressure, density, and solvency characteristics, the pMDIs powered with HFA propellants encounter complex formulation challenges and the need for new hardware designs. The surfactants commonly used in the CFC-based pMDIs are oleic acid, sorbitan trioleate, and soya lecithin, which not only help to stabilize the formulation, but also help lubricate the metering valve system. These surfactants exhibit lower solubility when formulated with HFAs compared to CFCs. Therefore, sometimes a co-solvent will be added to the HFA formulation. New surfactants may need to be identified for drug formulations when using HFAs as propellants. However, with each change in product formulation, extensive testing is required before the product is approved for general clinical use.

Therefore, the phasing out of CFCs in pMDIs has provided an opportunity to improve the performance of the pMDIs and address the shortcomings of pMDI technology. Keller (50) compared the performance of three different active compounds, beclomethasone dipropionate (BDP), budesonide, and di-sodium-cromo-glycate (DSCG), formulated with both CFC and HFA propellants. The HFA formulations were compared with three types of commercial devices, Becotide 100®, Pulmicort®, and Intal®. The amount of drug deposited on the mouthpiece and USP throat was measured, and particle size distribution and fine particle dose were also determined accordingly. Table 10 compares the performance of the pMDIs using formulations with CFCs and HFAs propellants.

Table 10 Performance of pMDIs formulation with CFCs and HFAs

Product batch no. Formulation (label claim)	Becotide 100® 10072926 CFCs suspension (100 µg)	BDP 809/L86-04 Non-CFCs solution (100 µg)	Pulmicort® YK702 CFCs suspension (200 µg)
Mouthpiece (µg)	16.6	7.8	21.4
USP-throat (µg)	55.2	22.0	114.0
FPD (µg)	26.9	60.8	43.3
FPF (%)	26.9	60.8	21.7
MMAD (µm)	~3.1	~1.4	~4.5
GSD	1.63	1.87	1.63

Product batch no. Formulation (label claim)	Budesonide 821/372-05 Non-CFCs suspension (200 µg)	Intal® FEF2E CFCs suspension (1000 µg)	DSCG 802/352-01 Non-CFCs suspension (1000 µg)
Mouthpiece (µg)	33.4	419.3	116.8
USP-throat (µg)	46.7	385.7	488.4
FPD (µg)	86.2	147.8	300.0
FPF (%)	43.1	14.8	30.0
MMAD (µm)	~3.1	~4.0	~3.8
GSD	1.55	1.81	1.65

(Adapted from Ref. 51.)

For the BDP formulation, the fine particle dose increased from 27 to about 61 µg and the drug lost in the mouthpiece and USP throat was reduced from 17 µg and 55 µg to about 8 µg and 22 µg respectively, when using the non-CFC formulation. The increase in the fine particle dose resulted from the reduction in the MMAD, which changed from 3.1 to about 1.4 µm. The particle size distribution of a suspension formulation is expected to be controlled mainly by the drug powder particles used in the formulation. For a solution formulation, particle size distribution is dictated by other factors such as formulation and device configurations. The fine particle dose was also increased for the budesonide formulation associated with a decrease in the amount of drug deposited in the USP throat, and the particle size distribution of budesonide delivered from both pMDIs changed slightly since both formulations were suspensions. Similar results were achieved for the DSCG compared with the Intal pMDI, except that the increase in fine particle dose was likely due to the decrease in the drug amount deposited in the mouthpiece.

These performance changes typically reduce the drug deposition in the oral-pharyngeal region, providing an added advantage of producing fewer steroid side effects such as candida infection and hoarseness of the throat (49).

pMDIs with breath activated function: The firing of the pMDI and inhalation by the patient have to be

coordinated or “synchronized,” because the drug aerosol is produced by the pMDI for a very short time only, in the order of milliseconds. This presents a challenge for patients to inhale and capture the aerosol as it is generated. It has been estimated that more than half of the patients (47–89%) receive little benefit from their pMDIs because of failure to coordinate the firing of the pMDIs and inhalation (51,52). A breath-activated pMDI (Autohaler, 3M Riker) has been developed to minimize the need for the actuation-inhalation coordination. This device is intended to reduce the performance variability of the pMDIs caused by differences in use techniques. The Autohaler is triggered by the patient’s inhalation flow rate. To use the Autohaler, the patient turns a level on the top of the device that engages a spring mechanism to load the canister against a vane mechanism that is used between the canister and the actuator. When the patient’s inspiratory flow rate exceeds 30 L/min, the vane moves and allows the canister to be pressed into the actuator to deliver the drug.

Newman (53) conducted a clinical study to compare the performance of conventional pMDIs with the Autohaler using a scintigraphic technique to measure the drug deposition in the different sites of interest. The patients who participated in the study were divided into two groups classified as good coordinators and poor coordinators. Initially, patients used their own inhalers

Table 11 Percentage deposition at different sites

Deposition site	Own MDI	Taught MDI	Autohaler
Good coordinators			
Lungs	18.6 (2.9)	12.8 (1.8)	17.5 (2.8)
Oropharynx	64.4 (3.8)	71.1 (2.5)	61.2 (4.5)
Actuator	16.1 (2.0)	15.8 (1.2)	21.0 (2.8)
Exhaled	0.7 (0.4)	0.3 (0.1)	0.3 (0.1)
Poor coordinators			
Lungs	7.2 (3.4)	22.8 (2.5)	20.8 (1.7)
Oropharynx	67.7 (4.7)	59.3 (2.3)	60.7 (2.2)
Actuator	23.5 (3.8)	17.6 (0.9)	18.2 (1.3)
Exhaled	1.7 (0.9)	0.3 (0.1)	0.2 (0.1)

(Adapted from Ref. 53.)

as baseline deposition values. The depositions were then measured after those patients were instructed about the proper inhalation technique when using their own inhaler and after using the Autohaler. Table 11 shows the percentage of the aerosol dose located at various locations after inhalation by good coordinators ($n = 10$) and by poor coordinators ($n = 8$).

For the good coordinators, the deposition patterns were very similar for the patients using their own pMDIs and Autohaler, in which about 18% of the aerosol dose was deposited in the lungs and about 60% of the dose was deposited in the oral-pharyngeal region. There was a trend toward less deposition in the lungs after the patients had been taught the proper pMDI technique. It was possible that the patients had different abilities to familiarize themselves with the new technique as instructed during the study. There was a clear improvement in the lung deposition for the poor coordinators after they were given instruction or when they used the Autohaler, in which over 20% of the aerosol dose was deposited in the lungs. The increased lung deposition contributed to the improvement in the pulmonary function parameter, in which the FEV₁ increased significantly from the baseline.

New Pressure Driven Devices

Respimat™

The agreement to eliminate the use of CFCs in pMDIs for respiratory medication has stimulated interest in the development of propellant-free inhaled drug delivery technology. The Respimat developed by Boehringer Ingelheim (Ingelheim am Rhein, Germany) is one of the technologies under development (54–58). The device consists of a Uniblock, spring, drug cartridge with a capillary tube, and a nonreturn valve. The Uniblock is the major component of the device, which contains two nozzles positioned at opposing angles. The spring is the power source. Turning the lower part of the device, the rotation movement is transferred to a liner movement, which simultaneously tightens the spring and pushes the capillary tube with the nonreturn valve to the lower position. During this movement, about 13.5 μ l of drug solution is drawn through the capillary to a pump chamber. When the patients press the release button, the mechanical force stored in the spring pushes the capillary tube and nonreturn valve to the upper position, driving the metered volume of drug solution through the two jet nozzles in the Uniblock, which generates two jets of liquid. The two jets are directed toward each other at a carefully controlled angle, and aerosol is generated by the impaction of the two jets.

Newman (55) examined the deposition profiles in patients of several developmental Respimat prototypes using gamma scintigraphic techniques. In these studies, the deposition of the bronchodilator fenoterol and the corticosteroid flunisolide were examined. Fenoterol in the Respimat was formulated in an aqueous medium, and flunisolide was formulated in 96% ethanol. The experimental data from these studies are shown in Table 12.

The performance of the Respimat was clearly improved for the final prototype compared with Prototype III. Drug deposition in the lungs increased from approximately 31 to 39% for the fenoterol, and from about 40 to 45% for the drug compound of flunisolide respectively. At the same time, the oral-pharyngeal deposition decreased from

Table 12 Average deposition data using the Respimat (% of the emitted dose)

Device	Formulation	Lung deposition	Oropharynx deposition	Reference
Prototype III	Fenoterol (100 μ g, aqueous)	31.1	53.7	55
Prototype III	Flunisolide (250 μ g, ethanolic)	39.7	39.9	56
Final Prototype	Fenoterol (100 μ g, aqueous)	39.2	37.1	57
Final prototype	Flunisolide (250 μ g, ethanolic)	44.6	26.2	57

(Adapted from Ref. 55.)



Fig. 9 The schematic diagram of the AeroEclipse nebulizer. (Courtesy of Monaghan/Trudell International.)

approximately 54 to 37% for the fenoterol, and from 40 to about 16% for the flunisolide. Comparing the deposition data for the aqueous and ethanolic formulations, it can be seen that the ethanolic formulation had better performance, which may be due to the different particle size distribution for the formulations. The MMADs of the aerosols used in the Respimat were $2.0 \pm 0.4 \mu\text{m}$ for the aqueous formulation and $1.0 \pm 0.3 \mu\text{m}$ for the ethanolic formulation (53). The overall deposition efficiency of this device was apparently two to three times better than typically seen with pMDIs.

AeroEclipse™

The AeroEclipse (Monaghan/Trudell International) shown in Fig. 9, is a nebulizer powered by compressed air, which can be operated in both continuous nebulization and breath-actuated modes. The nebulizer was characterized,

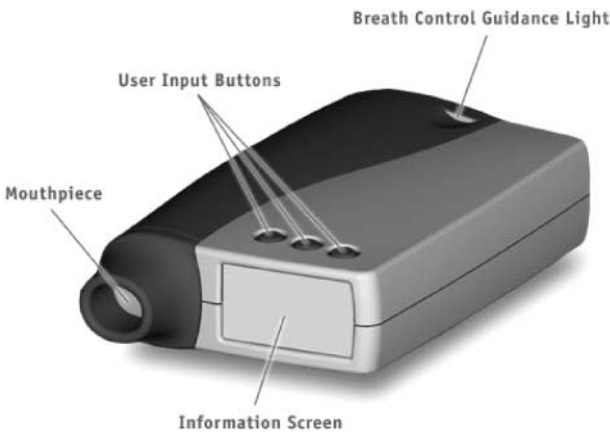


Fig. 10 The metered dose liquid inhaler as shown in Aradigm Corporation’s website (www.aradigm.com).

in vitro, using different breathing patterns, with either continuous nebulization or breath-actuated nebulization (59). A piston pump was used to mimic a range of breathing patterns in pediatric patients.

Table 13 shows the experimental data. The average drug mass for the three breathing patterns were $11.1 \pm 0.74\%$ (P1), $22.9 \pm 2.74\%$ (P2), and $36.3 \pm 1.22\%$ (P3). The average mass median aerodynamic diameter was $3.55 \pm 0.07 \mu\text{m}$ with the geometric standard deviation of 2.55. The test was also conducted to compare the device performance with other small-volume nebulizers (60), and found to be more efficient in the delivery rate (mass out per minute) than that of other SVN’s.

AERx™

The AERx, developed by Aradigm (Hayward, CA), is a metered dose liquid inhaler designed to deliver various pharmaceutical compounds to the peripheral lungs. The system, as shown in Fig. 10, consists of a unit dose disposable container equipped with a nozzle array, a piston assembly, and electronics associated with breath actuation and compliance monitoring functions (61).

The unit dose package contains the unit dose reservoir and an array of laser drilled nozzles, and the reservoir and

Table 13 Breathing patterns

Pattern	Tidal volume (ml)	Breathing frequency (1/min)	Nebulization
P1	50	40	Continuous
P2	200	25	Continuous
P3	440	19	Breath-activated

(Adapted from Ref. 60.)

Table 14 Emitted dose (% of loaded dose) vs. inhalation flow rate ($n = 5$)

Flow rate (L/min)	Emitted dose (% loaded dose)
20	50.6 \pm 2.0
30	68.6 \pm 3.0
45	72.3 \pm 2.5
60	70.9 \pm 1.5
70	68.5 \pm 1.3
85	66.1 \pm 5.1
100	61.5 \pm 1.9

(Adapted from Ref. 61.)

nozzle array are connected with a heat seal that allows the formulation to flow from the reservoir to the nozzle after the seal is ruptured. The piston assembly consists of a motor, a piston, and a cam, which compresses the unit dose packet to extrude the drug under pressure through the nozzle array to produce aerosols suitable for inhalation. The AERx also has internal electronic monitoring, which measures the patient's inspiratory flow rate as a function of time of inspiration and triggers the dispensing of the dose at a predetermined inspiratory flow rate and time for optimal delivery. The dosage administered is also logged, thereby providing a record of treatments and an indication of patient compliance with therapy (62).

The performance of the AERx has been characterized both in vitro and in vivo (63–67). Schuster et al. (61) reported in vitro experimental data of emitted dose as a function of inhalation flow rates from 20 to 100 L/min, and Table 14 shows the emitted dose expressed as the percentage of loaded dose at different inhalation flow rates. The emitted dose was consistent at inhalation flow rates ranging from 30 to 70 L/min, which averaged approximately 60–70% of the loaded dose emitted from the device. The percentage of loaded dose emitted from the device was about 51% at the inhalation flow rate of 20 L/min. When the device was used at this low flow rate, the aerosol was not efficiently entrained into the air, which was confirmed by a large fraction of the drug compound recovered from the air channel opposite the nozzle (61). The emitted dose was also decreased with the increase in the inhalation flow rate, which may have been due to increase in turbulent deposition as explained by the authors.

A study was also conducted both in vitro and in vivo to compare the performance of the AERx with a pMDI device that was operated with a SmartMist™ system for in vivo study using the scintigraphic technique (64). Table 15 shows the in vitro experimental data for both the AERx and the pMDI devices.

Table 15 In vitro performance of solution MDI and AERx system

	Solution MDI ($n = 4$)		AERx system ($n = 4$)	
	Mean	SD	Mean	SD
Emitted dose (%) ^a	63.2	8.2	60.8	7.1
FPF (%) ^b	71.0	5.6	90.6	1.6
MMAD (μ m)	1.2	0.1	2.6	0.1
GSD	1.8	< 0.1	1.5	0.3

^aPercentage of radioactivity contained in the AERx dosage form or ex-valve dose for the MDI.

^bFPF = Amount of aerosolized radioactivity in droplets with less than 5.7 μ m.

(Adapted from Ref. 63.)

The in vitro data showed that the fine particle fraction for the AERx was approximately 91% of emitted dose compared with 71% for the pMDI at the comparable emitted dose. The mass median aerodynamic diameter was 2.6 μ m for the AERx and 1.2 μ m for the pMDI, with a slightly larger GSD. An experiment was conducted to compare the in vivo performance of both devices using scintigraphy (64). The average depositions (expressed as the percentage of loaded dose for the AERx and percentage of ex-valve dose for the pMDI) in the oral-pharynx and stomach were 6.9% with a relative standard deviation (RSD) of 47% for the AERx and 42% with a RSD of 16% for the pMDI respectively. The average depositions in the lungs were 53.3% (RSD = 10.9%) and 21.7 (RSD = 30.9%) for the AERx and the pMDI, respectively. The experimental data also showed that uniform drug distribution in the lungs was achieved while using the AERx ($c/p = 1.15$) as compared with that of the pMDI ($c/p = 1.66$), where the c/p ratio is defined as the ratio of drug deposition in the central region to that of peripheral region.

HaloLite™

HaloLite, shown in Fig. 11, is a hand-held drug delivery system developed by Medic-Aid (Bognor Regis, United Kingdom). The device, which uses compressed air, consists of a medication chamber, a control unit, and an aerosol generation assembly that is operated by a portable, dedicated compressor. The aerosol is generated based upon conventional nebulization principles. The control unit allows the patient to select a budesonide nebulizing suspension, salbutamol, or terbutaline, since the device is calibrated to deliver preset doses of 50 μ g of budesonide, 200 μ g of salbutamol, or 500 μ g of terbutaline from commercially available pre-mixed solution formulations

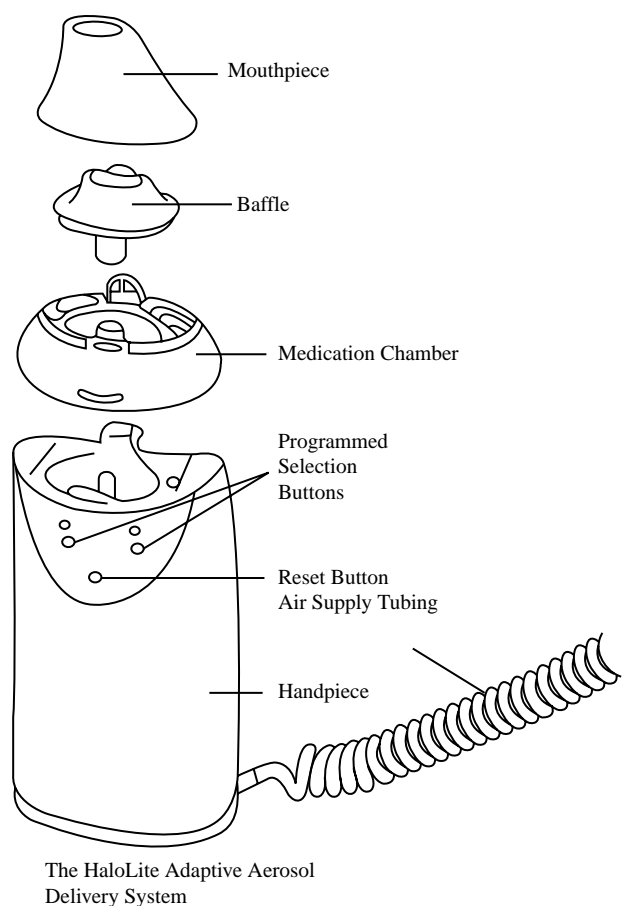


Fig. 11 The schematic diagram of the Halolite. (Courtesy of Medic-Aid.)

of each drug (68). The device has the capability to analyze the patient's breathing pattern (flow rate, frequency, etc.) to determine a) the aerosol pulse time for each inhalation, b) whether the patient is inhaling or exhaling, c) the start of inhalation, and d) trigger the aerosolization process.

In vitro experimental data show that the fine particle fraction ($<5.0\ \mu\text{m}$) ranged from 70 to 80% of the device output, and the coefficient of variation for the dose output was 6% at a preset dose of 63 mg of 1% NaF solution (69). The device was also evaluated in vivo using healthy volunteers and asthmatics, and the coefficient of variations were 11% for the healthy volunteers and 21% for the asthmatics.

New High Frequency Oscillating Devices

AeroDose™

The AeroDose is a battery-operated drug delivery device under development by AeroGen, Inc. (Sunnyvale, CA),

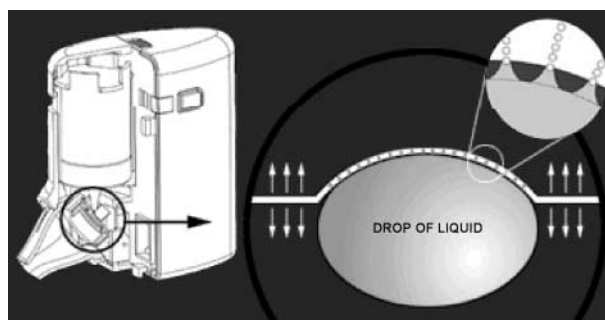


Fig. 12 The schematic diagram of AeroDose™ as shown in AeroGen, Inc.'s website (www.aerogen.com).

which uses vibrating orifice technology to produce drug aerosols suitable for human inhalation. The schematic diagram of the device and its operation principle are illustrated in Fig. 12. The device consists of a drug canister, aerosol generator assembly, and electronic control unit. The aerosol generator contains a dome-shaped aperture plate with 600 tapered holes that vibrate due to the attachment to a piezoelectric crystal. The piezoelectric material will undergo small physical displacement at ultrasonic frequency to generate aerosols with low velocity. During each use, a droplet of formulation ($\sim 15\ \mu\text{L}$) is dispensed from a multidose canister to an aperture plate by a metering pump, as shown in Fig. 12, and an aerosol is then generated by breath activation.

Three AeroDose devices were characterized to determine the in vitro performance using albuterol sulfate with a nominal dose of $120\ \mu\text{g}$ per actuation (70,71). The average liquid volume dispensed for three devices was 15.2 ml with a coefficient of variation (cv) of 2.7%. The emitted dose expressed as percentages of nominal doses were 80% (cv 5.9%), 81% (cv 3.7%), and 79% (cv 4.0%). The MMADs were $2.0\ \mu\text{m}$ (cv 13%), $2.0\ \mu\text{m}$ (cv 5.7%), and $1.9\ \mu\text{m}$ (cv 21%), with GSDs of 2.1 (cv 7.2%), 2.2 (cv 4.5%), and 2.4 (cv 4.7%) for the three devices tested. A lung deposition study using $^{99\text{m}}\text{Tc}$ radiolabeled albuterol sulfate was also conducted to determine the amount of drug deposited in the lungs for both the AeroDose and the conventional pMDI. The average lung deposition was 18% (cv 32%) of the emitted dose for the pMDI and 70% (cv 28%) of the emitted dose or an overall delivery efficiency of approximately 56% for the AeroDose for the six volunteers who participated in the study.

Metered solution inhaler (MSI)

The metered solution inhaler (MSI), under development by Sheffield Pharmaceuticals, uses the same aerosolization principle as the AeroDose. The MSI is a portable, hand-held drug delivery device, as shown in Fig. 13. Using a



Fig. 13 The metered solution inhaler as shown in Sheffield Pharmaceutical's website (www.sheffieldpharm.com).

motorized pump, drug solution is delivered to the surface of a ultrasonic horn powered by a piezoelectric crystal upon actuation, and is then aerosolized in about 1 s (72, 73).

The performance of the MSI was characterized *in vivo* using morphine sulfate for systemic absorption (72). A total of 1.25 mg of morphine sulfate was delivered in five doses of 250 μg per actuation to the healthy volunteers. The average amount of drug deposited was approximately 800 μg in the lungs and less than about 500 μg in the oral-pharynx; therefore the average respiratory deposition efficiency was approximately 64% of the inhaled dose. The average amount of drug deposited in the peripheral respireto region was slightly over 300 μg , a little less than 300 μg was deposited in the intermediate lung, and about 200 μg was deposited in the central region. The ratio of deposition in the peripheral to central regions ranged from 1.1 to 2.2. The morphine sulfate concentration in the plasma reached a peak within 3 min after inhalation, compared with 15 min for the subcutaneous injection. The peak plasma concentration level for the inhalation route was more than double compared to the subcutaneous route.

Vibrating membrane nebulizer

The vibrating membrane nebulizer is currently under development by Pari GmbH, (Germany), which is based

upon the TouchSpray technology developed by The Technology Partnership plc (United Kingdom). The main aerosol generation principle is similar to that of AeroDose, which uses piezoelectric material as a power source to vibrate a surface having on it a droplet of drug solution. The vibrating membrane nebulizer consists of two major components, a membrane and a piezoelectric ring, as shown in Fig. 14. The membrane is a circular, wafer-thin metal plate with small holes. A ring-shaped piezoelectric actuator excites the membrane to vibrate, thus ejecting the fluid through the holes as droplets and creating aerosols. Unlike the AeroDose, which is a single-dose device, the vibrating membrane nebulizer is a continuous generation device similar to the conventional nebulizer, which produces aerosols during both inhalation and exhalation (74, 75).

Experimental data obtained from *in vitro* testing using a functional model of the vibrating membrane nebulizer showed that the average percentage of emitted dose for a continuous aerosolization in approximately 8 min was about 79% of dose loaded in the device, and the output rate of the delivery device was found to be about 238 $\mu\text{g}/\text{min}$. The mass median diameter measured by the Malvern Diffraction Sizer was 4.6 μm . It was also found that the cumulative output from the device was linearly proportional to the time of aerosolization, which implied a constant output rate throughout the entire aerosolization process (75). Experimental data obtained from a prototype device developed based upon the functional model showed a reduction from 79 to about 50% in the emitted dose of the device. The amount of drug left in the device increased



Fig. 14 Hand-held EHD device being developed by BPT.

from 477 to about 1060 mg, which changed from 19 to 43% in terms of percentage to the loaded dose. The increase in the drug amount left in the device reduced the delivery efficiency of the device. The delivery efficiency can be improved by redesigning the flow channel and incorporating a breath-activated function to eliminate the aerosolization process during exhalation, as suggested by the authors (74).

Novel Electrohydrodynamic (EHD) Devices

Electrohydrodynamic (EHD) aerosolization is a process of the disruption of a liquid surface into a spray of droplets when subjected to an electric field. The liquid meniscus at the outlet of a capillary tube takes a conical shape under the action of the electric field. The cone tip breaks up into a spray of fine, charged aerosol (76). The cone-jet mode offers the appealing feature of aerosol monodispersity. It can produce droplets/particles over a wide size range, from submicron to hundreds of microns in size, depending on liquid flow rate, applied voltages, and physical and chemical properties of the liquid. When considering the size range of interest to pulmonary drug delivery (1–5 μm), the production of monodispersed aerosols with relative ease by EHD aerosolization technology is unmatched by any other aerosol generation process. Additionally, the aerosols are generated from capillary nozzles having relatively large diameters (e.g., 100 μm), which are therefore, unlikely to clog at the atomization site or during the metering of each dose. When considering the particle size range of 1–5 μm , the capacity to produce monodispersed aerosols with relative ease by EHD is also unmatched by any other aerosol generation process (77). The particle size distribution of the aerosol can be controlled by adjusting a number of variables, such as physical and chemical properties of drug formulations, operating conditions, and electric field. Another key clinical and technical attribute of the EHD technology is that it delivers a soft (isokinetic) aerosol cloud of nearly monodispersed particles to the patient, because aerosol formation does not require any liquid propellants or other pressurized systems. The EHD process can aerosolize a broad range of drug formulations, requires low precision manufacturing (i.e., low-cost production), and is amenable to miniaturization. These fundamental features underpin the ability of EHD pulmonary drug delivery devices to aerosolize a wide array of drug formulations, and provide accurate, reproducible drug delivery with targeted dosing to the human respiratory tract. Battelle Pulmonary Therapeutics, Inc. (BPT) is currently developing an array of EHD pulmonary devices, ranging from small hand-held

disposable units, home table-top aerosol delivery devices, to large benchtop clinical devices designed to operate continuously with high outputs (Fig. 15).

In vitro performance

Prior to clinical evaluations, comprehensive studies were conducted to characterize prototype EHD pulmonary drug delivery devices for emitted dose uniformity and reproducibility of particle size distribution (78). Table 16 shows the dose uniformity. The nominal dose is defined as the amount of active drug metered as a liquid solution from the device. The emitted dose is the amount of aerosolized drug obtained at the exit of the mouthpiece, and the device delivery efficiency is the ratio of the emitted to nominal dose. The mean delivery efficiency was approximately 94% (SD = 3.2) with a 3.4% coefficient of variation. The delivery efficiency for the two devices tested was 93% (SD = 3.60, $n = 5$) and 95% (SD = 2.8, $n = 5$), respectively. The intradevice variation was 1.7% (coefficient of variation, $n = 2$ prototypes). There were no significant differences ($p > 0.05$) in the drug mass output within the device throughout its life cycle, nor between devices tested in the study.

Fig. 15 shows the drug mass recovered from the various stages of the impactor and device at a nominal drug dose of 75 μg per actuation. The amount of drug deposited on each stage was used to calculate the MMAD and GSD. The calculated MMAD was 2.85 μm with a GSD of 1.6. The fine particle fraction (FPF) of the aerosol was 90% (<5.8 μm) of the emitted dose, and 95% (<5.8 μm) of the dose distal to the USP throat.

In vivo performance

A Phase I clinical study was conducted to evaluate the performance of the prototype EHD pulmonary drug delivery device in healthy volunteers (78). The EHD

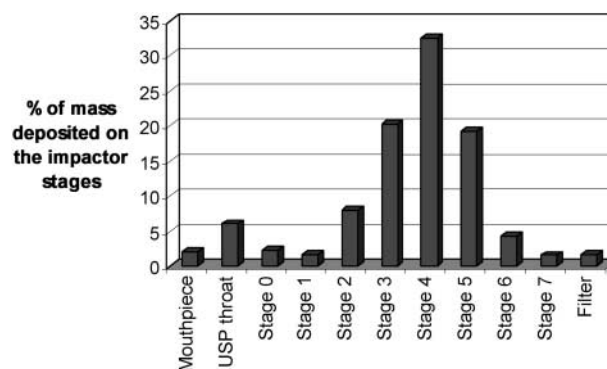


Fig. 15 Particle size distribution of a NCE from EHD device. (Adapted from Ref. 78.)

Table 16 Dose uniformity of a NCE delivered EHD pulmonary drug delivery device

Device ID	Actuation number	Nominal dose (µg)	Emitted dose (µg)	Device delivery efficiency (%)
Device 1	3	75	74.2	98.9
	15	75	68.5	91.3
	30	75	69.1	92.1
	45	75	67.8	90.4
	60	75	67.8	90.4
Device 2	3	75	71.3	95.0
	15	75	67.8	90.4
	30	75	73.4	97.9
	45	75	71.7	95.6
	60	75	71.7	95.6
Mean			70.3	93.7
SD			2.42	3.22
CV(%)			3.44	3.44

(Adapted from Ref. 78.)

device was compared with a commercially available capsule dry powder inhaler (DPI). The NCE was prepared in two formulations. The first was a powder formulation specific for the DPI, and the second was formulated in a liquid solution for the EHD device. Each formulation was radiolabelled with ^{99m}Tc, and the amount of drug deposited in the lungs from each device was quantified using gamma scintigraphy. Each volunteer was administered one dose of drug using the DPI and three different doses using the EHD device. The nominal doses were 1000 µg for the DPI and 150, 250, and 400 µg for the EHD device. Multiple blood samples were collected from each subject up to 8 h after dosing to measure plasma drug

concentration. An interval of 1 week was used between administration of the four doses.

The average whole lung deposition efficiency for the EHD device expressed as the percentage of the emitted dose was approximately 78% (cv 7.3%) for the 400-µg dose.

Fig. 16 shows representative scintigraphic images from both the DPI and the EHD pulmonary drug delivery device. It can be seen that the EHD pulmonary drug delivery device produced a uniform deposition distribution through the lung field with only ~16% deposition in the oral-pharyngeal region, and virtually no drug in the GI tract. However, the amount of drug

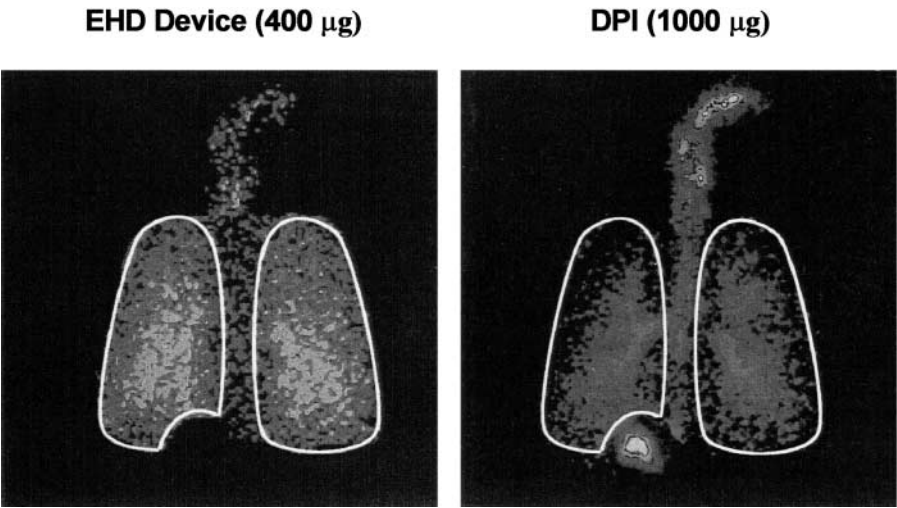


Fig. 16 Scintigraphic images of the EHD device and the DPI. (Adapted from Ref. 78.)

Table 17 Performance comparison of various devices

Delivery system for liquid	System features	Particle size	Device efficiency	In vivo deposition
RespiMat	Mechanical spring	3.5–6.5 μm		30–45% of emitted dose
AeroEclipse™	Compressed air powered	~3.6 μm	11–36% of nominal dose	
AERx	Mechanical, battery powered	2–4 μm	50–72% of nominal dose	~53% of loaded dose
Halolite	Compressed air powered	~3.0 μm		8–10% of nominal dose
AeroDose	Vibrating orifice	~2.0 μm	~80% of nominal dose	~70% of emitted dose
MSI™	Vibrating orifice	FPF >50% of emitted dose		~64% of emitted dose
Membrane nebulizer	Vibrating orifice, continuous mode	NA	50–80% of emitted dose	NA
EHD	Electrospray	2–4 μm , adjustable	~93% of nominal dose	~78% of emitted dose

deposited in the oral-pharyngeal region using the DPI averaged 67% of metered dose, which was nearly 6 times higher compared with that of the high dose delivered using the EHD device. The AUC for plasma drug concentration with the DPI was approximately 20% higher than that of the EHD device for a comparable deposited pulmonary dose. It appeared that the higher systemic exposure resulting from the DPI was likely due to absorption from the throat and GI tract.

Performance Comparisons

Comparison of the performance of various drug delivery devices based upon the data published in the literature is often difficult because of inconsistency in the experimental methods used in the characterization of those devices as well as differences in the devices. However it is also extremely important to understand the different features of the devices to help select the best device for a drug compound intended for delivery through the human respiratory tract. Table 17 summarizes the performance of devices currently under development.

The selection of a drug delivery system suitable for the administration of a particular drug compound to human respiratory airways depends not only on the device performance, but also on many other factors such as drug properties, formulation excipient, intended site of action, marketing preference, and regulatory requirements. It is important to fully understand the drug properties during the formulation development so that a proper excipient system can be identified. The

selection of an excipient for a particular drug also influences the particle size distribution of aerosol generated from a drug delivery device. For example, increasing the volatile components in the solution formulation can reduce the diameter of aerosol particles because of solvent evaporation, which is another variable available to optimize drug formulation and device performance for either local delivery or systemic absorption. The stability of drug product in the long term (usually ~2 years) has also to be considered in the device selection process. Some drug compounds can only be packaged in a unit dose because of instability of the drug product in the multidose reservoir; therefore, a device that can be fitted with a unit dose package is needed for such purpose. The drug delivery devices mentioned above have different features and different requirements for the drug formulation, and there is no single delivery device that can meet all the needs for such requirements. To ensure a successful launch of a new inhalation drug product, it is necessary to conduct a comprehensive evaluation of a potential drug candidate for administering via inhalation.

REFERENCES

1. Hinds, W.C. *Aerosol Technology, Properties, Behavior, and Measurement of Airborne Particles*; 2nd Ed.; John Wiley & Sons, Inc.: New York, 1999.
2. Wood, R.E.; Knowles, M.R. Recent Advances in Aerosol Therapy. *J. Aerosol Med.* **1994**, 7 (1), 1–11.
3. Byron, P.R. Aerosol Formulation, Generation, and Delivery Using Nonmetered System. *Respiratory Drug Delivery*; Byron, P.R. Ed.; CRC Press, Inc.: Boca Raton, 1990, 143–165.

4. Patton, J.S.; Platz, R.M. (A) Routes of Delivery: Case Studies, (B) Pulmonary Delivery of Peptides and Proteins for Systemic Action. *Adv. Drug Delivery Rev.* **1992**, *8*, 179–196.
5. Gupta, P.K.; Adjei, A.L. Therapeutic Inhalation Aerosol. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.K. Eds.; Marcel Dekker, Inc.: New York, 1997, 185–234.
6. Newman, S.P.; Forward. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.K. Eds.; Marcel Dekker, Inc.: New York, 1997, xi–xii.
7. Clark, A.R. Medical Aerosol Inhalers: Past, Present, and Future. *Aerosol Sci. Technol.* **1995**, *22*, 374–391.
8. Thiel, C.G.; From Susie's Question to CFC Free: AN Inventor's Perspective on Forty Years of MDI Development and Regulation. *Respiratory Drug Delivery*; Byron, P.R., Dalby, R.N., Farr, S.J. Eds.; Interpharm Press, 1996, 349–351.
9. Lefebvre, Arthur H. Basic Processes in Atomization. *Atomization and Sprays*; Hemisphere Publishing Co.: Washington, DC, 1989, 27.
10. Newman, S.P.; Pellow, P.G.D. Valuation of Jet Nebulizer. *In Drug Delivery to the Respiratory Tract*; Ganderton, D., Jones, T. Eds.; Ellis Horwood Ltd.: Chichester, England, 1987, 124–132.
11. Klusener, O. The Injection Process in Compressorless Diesel Engines. *VDIZ* **1933**, *77* (7).
12. Niven, Ralph W. Atomization and Nebulizers. *Inhalation Aerosols*; Hickey, A.J. Ed.; Marcel Dekker, Inc.: New York, 1996, 273–312.
13. Rizkalla, A.A.; Lefebvre, A.H.; Influence of Liquid Properties on Air-Blast Atomizer Spray Characteristics. *ASME Gas Turbine Conference* **1974**, 1–5, Paper No. 74-GT-1.
14. Searls, E.M.; Snyder, F.M. Relation of Viscosity to Drop Sizes. *J. Econ. Entomol.* **1936**, *29*, 1167–1170.
15. Hinds, W.C.; Macher, J.M.; First, M.W. Size Distribution of Aerosols Produced by the Laskin Aerosol Generator Using Substitute Materials for DOP. *Am. Ind. Hyg. Ass. J.* **1983**, *44*, 495–500.
16. McCallion, Orla, N.M.; Taylor, Kevin, M.G.; Thomas, Marian; Taylor, Anthony, J. Nebulization of Fluids of Different Physicochemical Properties with Air-Jet and Ultrasonic Nebulizers. *Pharm. Res.* **1995**, *12* (11), 1682–1688.
17. Clark, Andrew, R. The Use of Laser Diffraction for the Evaluation of the Aerosol Clouds Generated by Medical Nebulizers. *Int. J. Pharm.* **1995**, *115*, 69–78.
18. Davis, S.S. Physico-chemical Studies on Aerosol Solutions for Drug Delivery.] I. Water-Propylene Glycol Systems. *Int. J. Pharm.* **1978**, *1*, 71–83.
19. Newman, S.P.; Pellow, P.G.D.; Clarke, S.W. Dropsizes from Medical Atomisers (Nebulizer) for Drug Solution with Different Viscosities and Surface Tensions. *Atomization Spray Technol.* **1987**, *3*, 1–11.
20. Phipps, Paul, R.; Gonda, Igor; Droplets Produced by Medical Nebulizers Some Factors Affecting Their Size and Solute Concentration. *Chest* **1990**, *97* (6), 1327–1332.
21. Mercer, T.T.; Tillery, M.I.; Chow, H.Y. Operating Characteristics of Some Compressed-Air Nebulizers. *Am. Ind. Hyg. Assoc. J.* **1968**, *29*, 66–78.
22. Smye, S.W.; Jollie, M.I.; Cunliffe, H.; Littlewood, J.M. Measurement and Predication of Drug Solvent Losses by Evaporation from a Jet Nebuliser. *Clin. Phys. Physiol. Meas.* **1992**, *13*, 129–134.
23. Ferron, G.A.; Soderholm, S.C. Estimation of the Times for Evaporation of Pure Water Droplets and for Stabilization of Salt Solution Particles. *J. Aerosol Sci.* **1990**, *21*, 415–429.
24. Newman, S.P.; Pellow, P.G.D.; Clay, M.M.; Clarke, S.W. Evaluation of Jet Nebulizers for Use with Gentamicin Solution. *Thorax* **1985**, *40*, 671–676.
25. Wood, J.A.; Wilson, R.S.E.; Bary, C. Changes in Salbutamol Concentration in the Reservoir Solution of a Jet Nebulizer. *Br. J. Dis. Chest* **1986**, *80*, 164–169.
26. May, K.R. The Collision Nebulizer: Description, Performance and Application. *Aerosol Sci.* **1973**, *4*, 235–243.
27. Everard, M.L.; Clark, A.R.; Milner, A.D. Drug Delivery from the Nebulisers. *Arch. Dis. Childhood* **1992**, *67*, 586–591.
28. Ding, J.Y.; Brooker, M.J.; Andre, J.C.; Zimlich, W.C.; Imondi, A.R.; Placke, M.E. Performance Evaluation of the Pari LC Nebulizer Using Different Nebulization Pressures. *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Serentec Press, Inc.: 2000, 349–351.
29. Finlay, W.H.; Stapleton, W.K.; Zuberbuhler, P. Variations in Predicted Regional Lung Deposition of Salbutamol Sulphate Between 19 Nebulizer Types. *J. Aerosol Med.* **1998**, *11* (2), 65–80.
30. Ganderton, D. Targeted Delivery of Inhaled Drugs: Current Challenges and Future Goals. *J. Aerosol Med.* **1999**, *12*, S3–S8.
31. Stapleton, K.W.; Finlay, W.H.; Zuberbuhler, P. An In Vitro Method for Determining Regional Dosages Delivered by Jet Nebulizers. *J. Aerosol Med.* **1994**, *7*, 325–344.
32. Morén, F.; Aerosol Dosage Forms and Formulation. *Aerosols in Medicine, Principle, Diagnosis and Therapy*; Morén, F., Dolovich, M.B., Newhouse, M.T., Newman, S.P. Eds.; Elsevier Science Publisher: Amsterdam, 1993, 321–350.
33. Hickey, A.J.; Evans, R.M. Aerosol Generation from Propellant-Driven Metered Dose Inhalers. *In Inhalation Aerosols*; Hickey, A.J. Ed.; Marcel Dekker, Inc.: New York, 1996, 417–439.
34. Clark, A.R. MDIs: Physics of Aerosol Formulation. *J. Aerosol Med.* **1996**, *9* (S1), s19–s26.
35. Edman, P. Pharmaceutical Formulations—Suspensions and Solutions. *J. Aerosol Med.* **1994**, *7* (S1), s3–s6.
36. Hallworth, G.W. The Formulation and Evaluation of Pressurized Metered-Dose Inhalers. *Drug Delivery to the Respiratory Tract*; Jones, D.G.T. Ed.; Ellis Horwood, Ltd.: Chichester England, 1987, 87–118.
37. Byron, Peter R.; Aerosol Formulation, Generation, and Delivery Using Metered System. *Respiratory Drug Delivery*; Byron, P.R. Ed.; CRC Press, Inc. Boca Raton, 1990, 167–205.
38. Physical Tests and Determinations: <601> Aerosols, The Official Compendia of Standards, USP 24 NF19; U.S. Pharmacopeia and National Formulary, 2000, 1895–1912.
39. Leach, C. Enhanced Drug Delivery through Reformulating MDIs with HFA Propellants—Drug Deposition and Its Effect on Preclinical and Clinical Programs. *In Respiratory*

- Drug Delivery V*; Byron, P.R., Dalby, R.N., Farr, S.J. Eds.; Interpharm Press, 1996, 133–144.
40. Fink, J.B. Metered-dose Inhalers, Dry Powder Inhalers, and Transitions. *Respir. Care* **2000**, *45* (6), 623–635.
 41. Borgström, L.; Newman, S. Total and Regional Lung Deposition of Terbutaline Sulphate Inhaled via a Pressurised MDI or via Turbuhaler. *Int. J. Pharm.* **1993**, *97*, 47–53.
 42. Sterk, P.J.; Plomp, A.; van de Vate, J.F.; Quanjer, P.H. Physical Properties of Aerosols Produced by Several Jet- and Ultrasonic Nebulizers. *Bull. Eur. Physiopathol. Respir.* **1984**, *20*, 65–72.
 43. Greenspan, B.J. Ultrasonic and Electrohydrodynamic Methods for Aerosol Generation. In *Inhalation Aerosols*; Hickey, A.J. Ed.; Marcel Dekker, Inc.: New York, 1996, 313–335.
 44. Lang, R.J. Ultrasonic Atomization of Liquids. *J. Acoust. Soc. Am.* **1962**, *34*, 6–8.
 45. Hager, J.; Gober, K.-H.; Löhr, J.-P.; Dürr, M. Measurement of Particle and Mass Distribution of Pentamidine Aerosol by Ultrasonic and Air Jet Nebulizers. *J. Aerosol Med.* **1992**, *5* (2), 65–79.
 46. Everard, M.L. Aerosol Therapy Past, Present, and Future: A Clinician's Perspective. *Respir. Care* **2000**, *45* (6), 769–776.
 47. Coyne, T.C. Introduction to the CFC Problem. *J. Aerosol Med.* **1991**, *4* (3), 175–180.
 48. Kontny, M.J.; Destefand, G.; Jager, P.D.; McNamara, D.P.; Turi, J.S.; Van Campen, L. Issues Surrounding MDI Formulation Development with Non-CFC Propellants. *J. Aerosol Med.* **1991**, *4* (3), 181–187.
 49. Leach, C.L. Approaches and Challenges to Use Freon Propellant Replacements. *Aerosol Sci. Technol.* **1995**, *22*, 328–334.
 50. Keller, M. Innovations and Perspectives of Metered Dose Inhalers in Pulmonary Drug Delivery. *Int. J. Pharm.* **1999**, *186*, 81–90.
 51. Larsen, J.S.; Hahn, M.; Kochevar, J.W.; Morris, R.J.; Kasier, H.B.; Weisberg, S.C.; Halverson, P.C.; Quessey, S.N. Administration Errors with a Conventional Metered Dose Inhaler Versus a Novel Breath Actuated Device. *Ann. All.* **1993**, *71*, 103–106.
 52. Epstein, S.E.; Manning, C.P.; Ashley, M.J. Corey, P.N.; Survey of the Clinical Use of Pressurized Aerosol Inhaler. *Can. Med. Assoc. J.* **1979**, *120* (7), 813–816.
 53. Newman, S.P.; Weisz, A.W.B.; Talaei, N.; Clarke, S.W. Improvement of Drug Delivery with a Breath Actuated Pressurised Aerosol for Patients with Poor Inhaler Technique. *Thorax* **1991**, *46*, 712–716.
 54. Zierenberg, B. Optimizing the In Vitro Performance of Respimat. *J. Aerosol. Med.* **1999**, *12*, s19–s24.
 55. Newman, S.P. Use of Gamma Scintigraphy to Evaluate the Performance of New Inhalers. *J. Aerosol. Med.* **1999**, *12*, s25–s31.
 56. Steed, K.P.; Towse, L.J.; Freund, B.; Newman, S.P. Lung and Oropharyngeal Depositions of Fenoterol Hydrobromide Delivered from the Prototype III Hand-held Multidose Respimat Nebuliser. *Eur. J. Pharm. Sci.* **1997**, *5*, 55–61.
 57. Newman, S.P.; Steed, K.P.; Reader, S.J.; Hopper, G.; Zierenberg, B. Efficient Delivery to the Lungs of Flunisolide from a New Portable Hand-held Multidose Nebuliser. *J. Pharm. Sci.* **1996**, *85*, 960–964.
 58. Newman, S.P.; Brown, J.; Steed, K.P.; Reader, S.J.; Kladders, H.; Lung Deposition of Fenoterol and Flunisolide Delivered Using a Novel Device for Inhaled Medications. *Chest* **1998**, *113*, 957–963.
 59. Smaldone, G.C. *Enhanced In-Vitro Delivery of Budesonide via Continuous and Breath-Activated Nebulization*; European Respiratory Society, Florence Italy, Aug. 30–Sep. 3, 2000.
 60. Blacker, R.; Morton, R.W.; Mitchell, J.P.; Nagel, M.W.; Hess, D.R. *The Effect of Small Volume Nebulizer (SVN) Design on Fine Particle Mass Delivery of a Bronchodilator* Present at Drug Delivery to the Lungs, London, UK, Dec., 1999.
 61. Schuster, J.A.; Farr, S.J.; Cipolla, D.; Wilbanks, T.; Rosell, J.; Lloyd, P.; Gonda, I. Design and Performance Validation of a Highly Efficient and Reproducible Compact Aerosol Delivery System: AERx™. In *Respiratory Drug Delivery VI*; Dalby, R.N., Byron, P.R., Farr, S.J. Eds.; Interpharm Press, 1998, 83–90.
 62. Dolovich, M. New Propellant-Free Technologies under Investigation. *J. Aerosol Med.* **1999**, s9–s17.
 63. Schuster, J.; Rubsamen, R.M.; Lloyd, P.; Lloyd, J. The AERx™ Aerosol Delivery System. *Pharma. Res.* **1997**, *14*, 354–357.
 64. Farr, S.J.; Warren, S.J.; Lloyd, P.; Okikawa, J.K.; Schuster, J.A.; Rowe, A.M.; Rubsamen, R.M.; Taylor, G. Comparison of In Vitro and In Vivo Efficiencies of a Novel Unit-Dose Liquid Aerosol Generator and a Pressurized Metered Dose Inhaler. *Int. J. Pharm.* **2000**, *198*, 63–70.
 65. Farr, S.J.; Schuster, J.A.; Lloyd, P.; Lloyd, L.J.; Okikawa, J.K.; Rubsamen, R.M. AERx™—Development of a Novel Liquid Aerosol Delivery System: Concept to Clinic. In *Respiratory Drug Delivery V*; Byron, P.R., Dalby, R.N., Farr, S.J. Eds.; Interpharm Press, 1996, 175–185.
 66. Cipolla, D.; Boyd, B.; Evans, R.; Warren, S.; Taylor, G.; Farr, S.J. Bolus Administration of INS365 Studying the Feasibility of Delivering High Dose Drugs Using the AERx™ Pulmonary Delivery System. In *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Interpharm Press, 2000, 231–239.
 67. Chan, H.K.; Daciskas, E.; Eberl, S.; Robinson, M.; Bautovich, G.; Young, I. Deposition of Aqueous Aerosol of Technetium-99m Diethylene Triamine Penta-Acetic Acid Generated and Delivered by a Novel System (AERx™) in Healthy Subjects. *Eur. J. Nuclear Med.* **1999**, *26*, 320–327.
 68. Denyer, J. Adaptive Aerosol Delivery in Practice. *Eur. Respir. Rev.* **1997**, *7*, 388–389.
 69. Denyer, J.; Nikander, K.; Halolite, T.M. A Novel Liquid Drug Aerosol Delivery System. *Respiratory Drug Delivery VI*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Interpharm Press, 1998, 311–314.
 70. De Young, L.R.; Chamber, F.; Narayan, S.; Wu, C. The AeroDose Multidose Inhaler Device Design and Delivery Characteristics. In *Respiratory Drug Delivery VI*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Interpharm Press, 1998, 91–95.
 71. De Young, L.R.; Chamber, F.; Narayan, S.; Wu, C. Albuterol Sulphate Delivery from the AeroDose Liquid Inhaler. In *Respiratory Drug Delivery VI*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Interpharm Press, 1998, 315–318.

72. Hirst, P.H.; Bacon, R.E.; Newman, S.P.; Armer, T.; Mohsen, N.; Pavkov, R.; Byron, D. Deposition, Absorption and Bioavailability of Aerosolized Morphine Sulfate Delivered by a Novel Hand-held Device, The Metered Solution Inhaler (MSI). In *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Interpharm Press, 2000, 467–469.
73. van der Linden, K.; Haack, O.; Ruttel, M. US Patent, 5, 950, 619.
74. Stangl, R.; Luangkhot, N.; Liening-Ewert, R.; Jahn, D. Characterising the First Prototype of a Vibrating Membrane Nebuliser. In *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Interpharm Press, 2000, 455–458.
75. Stangl, R.; Luangkhot, N.; Liening-Ewert, R.; Jahn, D. Characterising the Functional Model of a Vibrating Membrane Nebuliser. *Drug Delivery to the Lungs*; Aerosol Society: London, 1999.
76. Cloupeau, M.; Prunet-Foch, B. Electrostatic Spraying of Liquids in Cone-jet Mode. *J. Electronics* **1989**, 22, 135–159.
77. Gomez, A.; Bingham, L.de Juan; Tang, K. Production of Protein Nanoparticles by Electrospray Drying. *J. Aerosol Sci.* **1998**, 29, 561–574.
78. Zimlich, W.C.; Ding, J.Y.; Busick, D.R.; Moutvic, R.R.; Placke, M.E.; Hirst, P.H.; Pitcairn, G.R.; Malik, S.; Newman, S.P.; Macintyre, F.; Miller, P.R.; Shepherd, M.T.; Lukas, T.M. The Development of a Novel Electrohydrodynamic Pulmonary Drug Delivery Device. In *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J. Eds.; Interpharm Press, 2000, 241–246.

MEDICATION ERRORS

Diane D. Cousins

United States Pharmacopeia, Rockville, Maryland, U.S.A.

INTRODUCTION

A report issued by the Institute of Medicine (1) in November 1999 has drawn unprecedented national attention to the prevalence of medical error, including medication error. The issues and complexities surrounding this problem are compounded by the array of players and solutions that must address it for the public health and safety of the U.S. population. Medical products and their manufacturers are reported not only as part of the problem but also as part of the solution. Ongoing national efforts in medication error reporting and prevention place the pharmaceutical industry far ahead of other medical product manufacturers. These efforts, headed by the *United States Pharmacopeia* (USP), have involved the industry since 1991 by sharing reports received from healthcare practitioners and documenting industry actions to the reported problems. The Institute of Medicine fosters a systems approach to error analysis that focuses on identifying the root cause of error within the system and not on blame of the individual. The report postulates that individuals who commit errors are often well trained, experienced, well intentioned individuals whose misfortune is a result of the unsafe systems in which they operate. With the advent of new technologies and the healthcare-delivery processes surrounding them, new problems are likely to surface, particularly as health systems are redesigned in response to this national call to action. The challenge to the pharmaceutical industry is to learn from its own experiences and the USP national database of errors. The industry will be expected to be knowledgeable of the medication use process for the health setting in which its products are used and to anticipate misuse by designing error out of products.

SCOPE OF THE PROBLEM

Incidence of Medication Errors and Related Morbidity and Mortality

That medication errors occur frequently in U.S. hospitals has been well documented (2–4). In observation studies

carried out between 1962 and 1995 on the rate of administration errors in a variety of inpatient settings, rates ranged from 0 to 59% (5). Estimates that medication errors occur in almost 7% of hospitalized patients have been reported (6). One study found that the frequency of medication errors was 1.4 per admission (4). When approximately 290,000 medication orders were analyzed, Lesar et al. estimated that there were almost two serious errors for every 1000 orders written. Based on a review of death certificates, it was estimated that nearly 8000 people died from medication errors in 1993 compared with almost 3000 people in 1983 (3). Researchers found an error rate at two children's hospitals of 4.7 per 1000 orders (7). Several excellent and comprehensive reviews of the literature on medication errors have recently been published (1, 5).

A variety of error rates for different aspects of the medication use process have been reported. Researchers use different methodologies and definitions of "medication error" and study different aspects of the medication use process (i.e., prescribing, dispensing, and administering). Because there is no national standardization for the denominator used to report medication error rates, the denominator can vary among several: doses dispensed, doses administered, doses ordered, patient days. Therefore, the rates reported in the literature are limited in their use for comparative purposes (5).

Research supports a systems approach to error prevention as well as to investigation of errors (8–11). This means that all aspects of the medication use process, including characteristics of the products themselves, should be explored for ways to improve safety in use.

Cost of Medication Errors

Medication errors are costly to both the patient (direct costs such as additional treatment and increased hospital stay) and to society (indirect costs such as decreased employment, costs of litigation) (1, 5). The cost of medication errors in a 700-bed teaching hospital, based on a study in 11 medical and surgical units in two hospitals over a 6-month period, was estimated at \$2.8 million annually (2). The increased length of stay associated with a medication error was estimated at 4.6 days (2). In a 4-year

study of the costs of adverse drug events (ADEs) in a tertiary care center, 1% of these events were classified as medication errors. The excess hospital costs for ADEs over the study period were almost \$4,500,000, with nearly 4000 days of increased hospital stay (12).

Harm attributable to drugs is a major reason for malpractice claims associated with medical procedures (8). The average compensation for medication errors between 1985 and 1992 was almost \$100,000. Most compensation for medication errors is for larger amounts that are agreed on in out-of-court settlements (5). None of the costs cited above include the cost of patient harm or subsequent hospital admissions (1,3).

USP's EFFORTS TO STANDARDIZE MEDICATION ERRORS

History of USP and Its Involvement in Medication Errors

The USP is a private, not-for-profit organization whose mission is to promote public health through the creation of standards and authoritative information for the use of medicines and related technologies. The USP's authority to set standards is established by the Pure Food and Drug Act and by the Federal Food, Drug and Cosmetic Act. These standards include those for quality, strength, purity, packaging, labeling, and storage of drug products. The USP also creates the official name for drug products and is a member of the United States Adopted Names (USAN) Council that sets the nonproprietary name for drugs in the United States. The USP has been involved in reporting programs for health professionals for nearly 30 years through its USP Practitioners' Reporting Network (USP PRN). These programs support the standards-setting activity by providing practitioner-based experiences about the quality and safe use of medicines in the marketplace.

Nearly a decade ago, the USP agreed to coordinate the medication errors reporting program for the Institute for Safe Medication Practices. The Institute was seeking a home for its grass-roots program and believed the program could have greater impact on the national level. Through the program, the USP hoped to learn of those circumstances in which the product labeling, packaging, or name of product caused or contributed to an error. Then, the USP envisioned setting standards to address the issues and thereby to prevent future errors. In 1994, the USP signed an agreement to purchase the program from the Institute, established the USP Medication Errors Reporting (MER) Program, and began its long-term commitment to

the program as an important part of the USP's standards-setting process. As a condition of the agreement, the ISMP continues to receive copies of reports submitted to this program for its education and advocacy work.

Healthcare professionals report errors in which they are involved as well as errors that they observe or are party to. Reported information forms a database used by the USP to identify problematic situations, to heighten practitioners' awareness of these situations, and to make appropriate interventions regarding issues with drug products.

USP Medication Errors Reporting Program

The prevention of medication errors is the primary objective of the USP MER Program. It collects and analyzes potential and actual medication errors submitted by healthcare practitioners. The program affords healthcare professionals the opportunity to report medication errors and thereby to contribute to improving patient safety by sharing their experiences.

To report an error, practitioners may phone USP toll-free at 1-800-23ERROR. A voice-mail system allows a report to be left 24 hr a day, 7 days a week. Reporters may submit reports anonymously or speak directly to one of USP's health professional staff. Alternatively, a report may be submitted to USP in writing. Report forms (Figs. 1 and 2) may be obtained by calling the USP directly or via an on-demand fax-back system. Practitioners may also access the form online on the USP's website.

Medication error information submitted to the USP is entered into a nationally recognized repository for medication error reporting. This database serves to track, monitor, and analyze medication errors from a systems-based perspective. The USP develops educational resources and materials to disseminate best-practice solutions and error-avoidance strategies to students and practitioners.

The MER Program is presented in cooperation with the Institute for Safe Medication Practices and is a partner in MED WATCH, the FDA's medical products reporting program. Although the FDA does not usually assert jurisdiction over practice issues, which are often involved in medication errors, it is concerned with issues relevant to product quality such as labeling and packaging, and product names, both trade and generic. When medication errors concerning product labeling and packaging are reported through the MER Program, pharmaceutical manufacturers are notified. They respond frequently and voluntarily make changes in labeling and packaging. Depending on the nature of the medication error, the MER Program reports provide material for ongoing discussions between the FDA and manufacturers and, if warranted, for

MEDICATION ERRORS REPORTING PROGRAM

Medication Errors Do Occur

Medication errors can occur anywhere, any time along the drug therapy course, from prescribing through transcribing, dispensing, administering, and monitoring. An error can cause confusion, alarm, and frustration for the health care provider and for the patient. And YES, an error can even cause a death or injury to your patient. The causes of errors are many; for example, lack of product knowledge or training; poor communication; ambiguities in product names, directions for use, medical abbreviations, handwriting, or labeling; job stress; poor procedures or techniques; or patient misuse. Along this continuum, any health care professional may be the cause of or contribute to an actual or potential error.

A Safer Environment for Your Patients

It is important to recognize that health care providers learn from medication errors. By sharing your experience through the nationwide USP Medication Errors Reporting (MER) Program you help your colleagues to gain an understanding of why errors occur and how to prevent them. You can also have a positive impact on the quality of patient care and influence drug standards and information. When others are informed about an error, the chance of recurrence may be lessened. Education regarding medication errors assists health care professionals to avoid errors by recognizing the circumstances and causes of actual and potential errors.

Easy Access

Just call 800-233-7767 to reach a USP health care professional, who will take your report and respond to your concerns. Reports may also be submitted in writing or faxed. All reported information is reviewed by USP for

possible impact on USP standards and information development. Reports are forwarded to the Food and Drug Administration, the ISMP, and when appropriate, the product manufacturer/labeler. If you wish to remain anonymous to any of these sources, the USP will act as your intermediary in all correspondence. While including your identity is optional, it does allow for appropriate follow-up with you to discuss your observations or provide feedback.

USP: A Partner in MEDWATCH

The USP Practitioners' Reporting Network is a partner in MEDWATCH, the FDA's medical products reporting program. As a partner, USP PRN contributes to the FDA's efforts to protect the public health by helping to identify serious adverse events for the agency. This means that your reported information is shared with the FDA on a daily basis, or immediately if necessary.



The USP PRN® is designed to collect experiences and observations from health care providers through three separate reporting programs:

- The USP Drug Product Problem Reporting Program
- The USP Medication Errors Reporting Program
- The USP Veterinary Practitioners' Reporting Program

The Institute for Safe Medication Practices, and the American Veterinary Medical Association cooperate in presenting the USP PRN.

Your Input Could Make the Difference!

USP PRN...CALL US WHEN YOU NEED US.



U.S. Pharmacopeia
12601 Twinbrook Parkway
Rockville, MD 20852-1790



NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST-CLASS MAIL PERMIT NO 39 ROCKVILLE MD

POSTAGE WILL BE PAID BY ADDRESSEE:

DIANE D COUSINS RPh

THE USP PRACTITIONERS' REPORTING NETWORK

12601 TWINBROOK PARKWAY

ROCKVILLE MD 20897-5211



Fig. 1 USP medication errors reporting program form.

	USP MEDICATION ERRORS REPORTING PROGRAM Presented in cooperation with the Institute for Safe Medication Practices The USP Practitioners' Reporting Network SM is an FDA MEDWATCH partner																			
	<div style="display: flex; justify-content: space-between;"> <input type="checkbox"/> ACTUAL ERROR <input type="checkbox"/> POTENTIAL ERROR </div> <p>Please describe the error. Include sequence of events, personnel involved, and work environment (e.g., code situation, change of shift, short staffing, no 24-hr. pharmacy, floor stock). If more space is needed, please attach separate page.</p>																			
<p>Was the medication administered to or used by the patient? <input type="checkbox"/> No <input type="checkbox"/> Yes Date and time of event: _____</p> <p>What type of staff or health care practitioner made the initial error? _____</p> <p>Describe outcome (e.g., death, type of injury, adverse reaction). _____</p> <p>If the medication did not reach the patient, describe the intervention. _____</p> <p>Who discovered the error? _____</p> <p>When and how was error discovered? _____</p> <p>Where did the error occur (e.g., hospital, outpatient or retail pharmacy, nursing home, patient's home)? _____</p> <p>Was another practitioner involved in the error? <input type="checkbox"/> No <input type="checkbox"/> Yes If yes, what type of practitioner? _____</p> <p>Was patient counseling provided? <input type="checkbox"/> No <input type="checkbox"/> Yes If yes, before or after error was discovered? _____</p>																				
<p>If a product was involved, please complete the following:</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 50%; text-align: center;">Product #1</th> <th style="width: 50%; text-align: center;">Product #2</th> </tr> </thead> <tbody> <tr><td>Brand name of product involved</td><td>_____</td></tr> <tr><td>Generic name</td><td>_____</td></tr> <tr><td>Manufacturer</td><td>_____</td></tr> <tr><td>Labeler (if different from mfr.)</td><td>_____</td></tr> <tr><td>Dosage form</td><td>_____</td></tr> <tr><td>Strength/concentration</td><td>_____</td></tr> <tr><td>Type and size of container</td><td>_____</td></tr> <tr><td>NDC number</td><td>_____</td></tr> </tbody> </table>			Product #1	Product #2	Brand name of product involved	_____	Generic name	_____	Manufacturer	_____	Labeler (if different from mfr.)	_____	Dosage form	_____	Strength/concentration	_____	Type and size of container	_____	NDC number	_____
Product #1	Product #2																			
Brand name of product involved	_____																			
Generic name	_____																			
Manufacturer	_____																			
Labeler (if different from mfr.)	_____																			
Dosage form	_____																			
Strength/concentration	_____																			
Type and size of container	_____																			
NDC number	_____																			
<p>If available, please provide relevant patient information (age, gender, diagnosis, etc.). Patient identification not required.</p>																				
<p>Reports are most useful when relevant materials such as product label, copy of prescription/order, etc. can be reviewed.</p> <p>Can these materials be provided? <input type="checkbox"/> No <input type="checkbox"/> Yes If yes, please specify. _____</p>																				
<p>Suggest any recommendations you have to prevent recurrence of this error or describe policies or procedures you have instituted to prevent future similar errors.</p>																				
<p>A copy of this report is routinely sent to the Institute for Safe Medication Practices (ISMP), to the manufacturer/labeler, and to the Food and Drug Administration (FDA). USP may release my identity to: (check boxes that apply)</p> <p><input type="checkbox"/> ISMP <input type="checkbox"/> The manufacturer and/or labeler as listed above <input type="checkbox"/> FDA <input type="checkbox"/> Other persons requesting a copy of this report <input type="checkbox"/> Anonymous to all</p>																				
<p>Your name and title _____</p> <p>Your facility name, address, and ZIP _____</p>		<p>Telephone number (include area code) _____</p>																		
<p>Signature _____</p>		<p>Date _____</p>																		
<p>Return to the attention of: Diane D. Cousins, R.Ph. USP PRN 12601 Twinbrook Parkway Rockville, MD 20852-1790</p>		<p>Call Toll Free: 800-23-ERROR (800-233-7767) or FAX 301-816-8532 USP home page: http://www.usp.org Electronic reporting forms are available. Please call for additional information and/or your free diskette.</p>																		
		<p>Date Received by USP: _____</p> <p>File Access Number: _____</p>																		

 699
 M1306DM

Additional forms can be found in the USP DI Vol. I and Vol. III.

Fig. 2 USP medication errors reporting program form.

regulatory action. Furthermore, reported information identifies broader issues that may become the basis for instituting industry-wide changes. The reported concerns of practitioners have prompted the USP, FDA, and various drug manufacturers to institute numerous changes and improvements to drug products and have contributed to safer medication prescribing and use.

Facility-Based Reporting May Help Define Denominator of Errors

Because of its leadership and experience in the prevention of medication errors, the USP began to receive inquiries from hospitals seeking a nationally standardized database that would help them meet its accreditation requirements and also to compare rates of medication errors among hospitals. Hospitals were willing to share their adverse experiences with other participating hospitals but only if the report could be shared on an anonymous basis. In 1998, the USP developed MedMARxSM, an Internet-accessible database of medication errors for hospitals. Reports submitted to the system are anonymous so that participating hospitals will share information openly. The database is structured to become part of the hospital's internal quality-improvement program and captures not only errors but prevention strategies taken by each hospital in response to errors. This valuable aspect of the national database enables hospitals to practice risk prevention, not just risk management, by learning from the unfortunate experiences of others. It is expected that this database will become a rich repository of information not only for hospitals but for the pharmaceutical industry as well.

USP's First Advisory Panel on Medication Errors

In 1996, the USP created an ad hoc Advisory Panel on Medication Errors. The mission of the Panel was to provide practitioner review of reports received through the USP MER Program and to make recommendations

relative to USP's standards-setting, information, and reporting programs. The Panel chairperson also has a unique opportunity to make broader recommendations through its seat on the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). The chairperson of the Advisory Panel on Medication Errors is an ex-officio nonvoting member of the NCC MERP.

The USP Advisory Panel on Medication Errors is a unique and unprecedented opportunity for healthcare professionals to provide peer review of medication errors occurring nationally and to recommend far-reaching strategies for medication error prevention.

The Panel consists of 12 actively practicing volunteers representing medicine, nursing, and pharmacy. This year a Safe Medication Use Expert Committee will be elected to replace the Panel. For the first time, with the formation of this committee, a formal mechanism will be in place in the standards-development process for the purpose of providing direct practitioner input to standards development for the safer use of pharmaceuticals.

A National Coordinating Council Is Initiated

After a few years operating the MER Program, the USP realized that the solutions addressing the myriad issues identified through the program were beyond the mission of its standards-setting capacity. Indeed, errors proved to be multidisciplinary in origin and multifactorial in cause. These other practice-related and process-related aspects surrounding medication errors needed to be addressed. In 1995, the USP spearheaded the formation of the NCC MERP. The NCC MERP promotes the reporting, understanding, and prevention of medication errors relative to professional practice, healthcare products, procedures, and systems (Table 1). The Council is composed of 20 national organizations and agencies, representative of health professions, licensing boards, healthcare facilities, pharmaceutical manufacturers, regulators, standards-setters,

Table 1 The goals of the NCC MERP are far-reaching and encompass the full spectrum of healthcare goals

Goals of NCC MERP
Examine and evaluate the causes of medication errors
Increase awareness of medication errors and methods of prevention throughout the healthcare system
Recommend strategies relative to system modifications, practice standards and guidelines
Stimulate development and use of medication-error reporting and evaluation systems and stimulate reporting to a national system for review, analysis, and development of recommendations to reduce and prevent medication errors

^a Copyright © 1998–1999 National Coordinating Council for Medication Error Reporting and Prevention. All Rights Reserved.

Table 2 The membership of the NCC MERP is interdisciplinary and represents cross-functional groups in the delivery of healthcare products and services

National Coordinating Council for Medication Error: Reporting and Prevention Organizations Represented

American Association of Retired Persons
American Health Care Association
American Hospital Association
American Medical Association
American Nurses Association
American Pharmaceutical Association
American Society of Consultant Pharmacists
American Society of Health-System Pharmacists
American Society for Healthcare Risk Management
Department of Veterans Affairs
Food and Drug Administration
Generic Pharmaceutical Association
Healthcare Distribution Management Association
Institute for Safe Medication Practices
Joint Commission on Accreditation of Healthcare Organizations
National Association of Boards of Pharmacy
National Council of State Boards of Nursing, Inc.
Pharmaceutical Research and Manufacturers of America
United States Pharmacopeia

and others (Table 2). The USP is a founding member of and Secretariat to the Council.

Since the Council’s formation, it has produced several important work products. Among them are the standardization definition of the “medication error,” the development of a series of recommendations designed to reduce errors in the medication use process, and the adoption of a severity index for categorizing the outcome of medication errors. The “Recommendations to Correct Error-Prone Aspects of Prescription Writing,” the first set of suggestions issued by the Council, included a list of “Dangerous Abbreviations,” abbreviations that are frequently misunderstood or have often been implicated in medication errors and should never be used (Table 3). In addition to being used in prescription writing, these abbreviations can be found in proprietary product names, on manufacturers’ product labels, and in advertising by pharmaceutical manufacturers. The pharmaceutical industry can support this effort by avoiding the use of these abbreviations. The Council also produced an extensive set of recommendations to reduce errors attributable to labeling and packaging. The recommendations are targeted to regulators and standards-setters, healthcare organizations and professionals, and the industry (Tables 4–7). The practical importance of the Council’s

recommendations lies in a joint endorsement by a diverse group of organizations ranging from experts in safety issues to manufacturers of drug products to regulators. The importance of achieving consensus through a collaborative effort by these national leading healthcare and consumer organizations furthers the adoption of nonpunitive, systems-based approaches to reduce medication errors.

ERROR AVOIDANCE STRATEGIES FOR THE INDUSTRY—DESIGNING ERROR OUT OF PRODUCTS

The USP’s medication error-reporting programs have uncovered a number of reported error-prone situations that could help industry consider the problems that should be addressed in advance, starting with the selection of a drug name and including the development of labeling, packaging, and dosing devices. Some of these cases are presented here. In many, the manufacturer corrected design flaws immediately and successfully. These should be considered showcase examples of industry responsiveness. The cases should also serve to teach certain designs in labels or packaging that should be avoided. And finally, the cases demonstrate how products can be misused because of the systems with which they interface.

Characteristics of Product Errors

Keep in mind that the medication use process is a complex continuum that requires the successful interaction of multiple allied health professionals, technology, and the patient. It can be described as a succession of joined, but distinct processes, known as nodes (Table 8). Each node in the medication use process is, in actuality, a discrete system and presents an opportunity for the occurrence and prevention of medication errors.

Medication errors have been defined in many ways depending on research methodologies, incident reporting systems, risk management, or total quality-improvement systems. The USP uses the broad definition of medication error from the NCC MERP.

A medication error is any preventable event that may cause or lead to inappropriate medication use or patient harm, while the medication is in the control of the healthcare professional, patient, or consumer. Such events may be related to professional practice, healthcare products, procedures, and systems including: prescribing; order communication; product labeling, packaging, and nomenclature; compounding; dispensing; distribution; administration; education; monitoring; and use.

Table 3 Dangerous abbreviations to never use owing to misunderstanding or misinterpretation

Abbreviation	Intended meaning	Common error
U	Unit	Mistaken as a zero or a four (4) resulting in overdose. Also mistaken for “cc” (cubic centimeters) when poorly written
µg	Microgram	Mistaken for “mg” (milligrams) resulting in overdose
Q.D.	Latin abbreviation for every day	The period after the “Q” has sometimes been mistaken for an “I”, and the drug has been given “QID” (four times daily) rather than daily
Q.O.D.	Latin abbreviation for every other day	Misinterpreted as “QD” (daily) or “QID” (four times daily). If the “O” is poorly written, it looks like a period or a “I”
SC or SQ	Subcutaneous	Mistaken as “SL” (sublingual) when poorly written
T I W	Three times a week	Misinterpreted as “three times a day” or “twice a week”
D/C	Discharge; also discontinue	Patient’s medications have been prematurely discontinued when D/C (intended to mean “discharge”) was misinterpreted as “discontinue,” because it was followed by a list of drugs
HS	Half-strength	Misinterpreted as the Latin abbreviation “HS” (hour of sleep)
Cc	Cubic centimeters	Mistaken as “U” (unit) when poorly written
AU, AS, AD	Latin abbreviations for both ears; left ear; right ear	Misinterpreted as the Latin abbreviation “OU” (both eyes); “OS” (left eye); “OD” (right eye)

Table 4 Recommendations on labeling and packaging to industry manufacturers of pharmaceuticals and devices (adopted May 12, 1997)

<p>The Council recommends that industry not use any printing on the cap and ferrule of injectables except to convey warnings.</p> <p>The Council encourages industry to employ failure mode and effects analysis in its design of devices, and the packaging and labeling of medications and related devices.</p> <p>The Council encourages industry to employ machine-readable coding (e.g., bar coding) in its labeling of drug products. The Council recognizes the importance of standardization of these codes for this use.</p> <p>The Council encourages printing the drug name (brand and generic) and the strength on both sides of injectables, and IV bags, containers, and overwraps. For large volume parenterals and IV piggybacks (minibags), the name of the drug should be readable in both the upright and inverted positions.</p> <p>The Council encourages industry to support the development of continuing education programs focusing on proper preparation and administration of its products.</p> <p>The Council encourages industry to use innovative labeling to aid practitioners in distinguishing between products with very similar names, for example, the use of tall letters such as VinBLAStm and VinCRIStm.</p> <p>The Council encourages industry to avoid printing company logos and company names that are larger than the type size of the drug name.</p> <p>The Council encourages collaboration among industry, regulators, standards-setters, healthcare professionals, and patients to facilitate design of packaging and labeling to help minimize errors.</p>

Copyright © 1998–1999 National Coordinating Council for Medication Error Reporting and Prevention. All Rights Reserved.

Table 5 Recommendations on labeling and packaging to regulators and standards-setters (adopted May 12, 1997)

<p>The Council recommends that FDA restrict the use of any printing on the cap and ferrule of injectables except to convey warnings.</p> <p>The Council recommends the use of innovative labeling to aid practitioners in distinguishing between products with very similar names, for example, the use of tall letters such as VinBLAStm and VinCRIStm.</p> <p>The Council recommends that FDA discourage industry from printing company logos and company names that are larger than the type size of the drug name.</p> <p>The Council supports the recommendations of the USP-FDA Advisory Panel on Simplification of Injection Labeling. Furthermore, the Council encourages USP/FDA to consider expansion of the concepts of simplification to apply to: package inserts; and labeling of other pharmaceutical dosage forms.</p> <p>The Council encourages further development of FDA’s error prevention analysis efforts to provide consistent regulatory review of product labeling and packaging relative to the error-prone aspects of their design.</p> <p>The Council encourages collaboration among regulators, standards-setters, industry, healthcare professionals, and patients to facilitate design of packaging and labeling to help minimize errors.</p> <p>The Council encourages USP/FDA to examine feasibility and advisability of use of tactile cues in container design and on critical drugs. Such cues may be in the design of the container or embedded in the label.</p> <p>The Council encourages the printing of the drug name (brand and generic) and the strength on both sides of injectables and IV bags, containers, and overwraps. For large volume parenterals and IV piggybacks (minibags), the name of the drug should be readable in both the upright and inverted positions.</p>

Copyright © 1998–1999 National Coordinating Council for Medication Error Reporting and Prevention. All Rights Reserved.

Table 6 Recommendations to health care professionals to reduce errors due to labeling and packaging of drug products and related devices (adopted March 30, 1998)

The Council encourages healthcare professionals to routinely educate patients and caregivers to enhance understanding and proper use of their medications and related devices. Furthermore, the Council encourages healthcare professionals to regularly participate in error prevention training programs and, when medication errors do occur, to actively participate in the investigation.

In addition, the Council makes the following recommendations to healthcare professionals to reduce errors due to labeling and packaging of drug products and related devices:

1. The Council encourages healthcare professionals to use only properly labeled and stored drug products and to read labels carefully (at least three times—before, during, and after use).
2. The Council encourages collaboration among healthcare professionals, healthcare organizations, patients, industry, standard-setters, and regulators to facilitate design of packaging and labeling to help minimize errors.
3. The Council encourages healthcare professionals to take an active role in reviewing and commenting on proposed regulations and standards that relate to labeling and packaging (i.e., Federal Register and Pharmacopeial Forum).
4. The Council encourages healthcare professionals to report actual and potential medication errors to national (e.g., FDA MedWatch Program and/or the USP Practitioners' Reporting Network), internal, and local reporting programs.
5. The Council encourages healthcare professionals to share error-related experiences, case studies, etc., with their colleagues through newsletters, journals, bulletin boards, and the Internet.

Copyright © 1998–1999 National Coordinating Council for Medication Error Reporting and Prevention. All Rights Reserved.

Table 7 Recommendations to healthcare organizations to reduce errors due to labeling and packaging of drug products and related devices (adopted March 30, 1998)

The Council recommends the establishment of a systems approach to reporting, understanding, and prevention of medication errors in health care organizations. The organization's leaders should foster a culture and systems that include the following key elements:

1. an environment that is conducive to medication error reporting through the FDA MedWatch Program and/or the USP Practitioners' Reporting Network;
2. an environment which focuses on improvement of the medication use process;
3. mechanisms for internal reporting of actual and potential errors including strategies that encourage reporting;
4. systematic approaches within the healthcare organization to identify and evaluate actual and potential causes of errors including Failure Mode and Effects Analysis (FMEA) and root cause analysis;
5. processes for taking appropriate action to prevent future errors through improving both systems and individual performance.

In addition, the Council makes the following recommendations to healthcare organizations to reduce errors due to labeling and packaging of drug products and related devices:

1. The Council recommends that healthcare organizations employ machine readable coding (e.g., bar coding) in the management of the medication use process.
2. The Council recommends reevaluation of existing storage systems for pharmaceuticals by healthcare organizations and establishment of mechanisms to insure appropriate storage and location throughout the organization from bulk delivery to point of use. The following issues should be considered when applicable: storage and location that will help distinguish similar products from one another; storage and location of certain drugs, (e.g., concentrates, paralyzing agents) that have a high risk potential; scope, access, and accountability for floor stock medications; safety and accountability of access to pharmaceuticals in the absence of a pharmacist (e.g., floor stock, eliminate access to pharmacy after hours); labeling and packaging of patient-supplied medications.
3. The Council recommends the development of policies and procedures for repackaging of medications that will clarify labeling to help avoid errors.
4. The Council encourages collaboration among healthcare organizations, healthcare professionals, patients, industry, standard-setters, and regulators to facilitate design of packaging and labeling to help minimize errors.
5. The Council recommends that healthcare organizations develop and implement (or provide access to) education and training programs for healthcare professionals, technical support personnel, patients, and caregivers that address methods for reducing and preventing medication errors.

Copyright © 1998–1999 National Coordinating Council for Medication Error Reporting and Prevention. All Rights Reserved.

Table 8 Nodes in the medication-use continuum

Medication use process nodes
Prescribing
Documenting
Dispensing
Administering
Monitoring

Thorough documentation of medication errors provides information about the severity of the error as it relates to the outcome of the patient, the product(s) involved, the level of staff handling the product or processing the order, any contributing factors that may predispose a product to misuse, and the suspected root cause of the error. The USP adds certain codes to MER Program data to characterize the error as it was reported. These codes include the type of error and the possible cause(s) of error. Table 9 lists product characteristics that have been recorded over 9 years to have caused or contributed to a medication error.

The pharmaceutical industry should pay close attention to these items in the earliest stages of product development, including clinical stages. Several years ago the drug zidovudine (an antiviral) was referred to as “AZT” in clinical trials. The abbreviation was brought along as the

product was marketed. However, “AZT” had been a common abbreviation for azathioprine (an immunosuppressant), and several errors were made. The Institute of Medicine report (1) suggests that the FDA develop and enforce standards for the design of drug packaging and labeling that will maximize safe product use. Table 9 should serve as a starting gate of areas to examine.

Case 051133: Poor label design; confusing or incomplete label information; packaging

A pediatric patient was presented to the emergency room (ER) experiencing seizures for which 150 mg of I.V. Cerebyx[®] (fosphenytoin, an anticonvulsant) was ordered. The pharmacy technician took the call for Cerebyx[®] and delivered three 10-ml vials of Cerebyx 50-mg PE (phenytoin sodium equivalents) per milliliter to the ER as a “floor stock” transaction. A nurse then misread the 50 mg PE/ml on the 10-ml container label, making the assumption that the entire vial contained 50 mg PE. The contents of all three vials were prepared for administration. Instead of 150 mg PE, the patient was administered 10 times the intended dose, or 1500 mg of PE. The patient later died. ER staff only discovered the error after the patient’s blood phenytoin levels were returned from the laboratory.

Discussion: Serious medication errors, including some leading to death, have resulted from the interpretation of the Cerebyx[®] product labeling. The terminology

Table 9 Product characteristics from medication errors as reported to USP

Abbreviations—Includes symbols and acronyms used in drug names as well as directions for use.
Dosage form confusion—Confusion due to similarity in color, markings, shape and/or size to another product, or to a different strength of the same product.
Equipment design confusing/inadequate for proper use—Example: administration pump makes it difficult to set precise fluid rate or is confusing to use.
Label (manufacturer’s) design—Physical label design, e.g., contrast of label information and background, letter font, symbol(s), or logo causes information to be overlooked or difficult to read.
Measuring device inaccurate/inappropriate—Scale of graduation markings on medical device (e.g., syringe, dropper) is inaccurate or inappropriate for administering the correct dose.
Names, a brand name/generic name of different products look alike—self-explanatory
Names, a brand name/generic name of different products sound alike—self explanatory
Names, brand names look alike—brand names of different products look alike.
Names, brand names sound alike—brand names of different products sound alike.
Names, generic names look alike—generic names of different products look alike.
Names, generic names sound alike—generic names of different products sound alike.
Nonmetric units of measurement (apothecary)—use of apothecary units of measurement results in misinterpretation (e.g., “cc” (cubic centimeter) written and misinterpreted as “u” (units).
Packaging/container design—the design of the package, bag, syringe, etc., caused or contributed to the error.
Similar packaging/labeling—example: packaging/labeling of two or more different products look similar, causing one product to be mistaken for the other.

on the label, which previously indicated the concentration as being 50 mg of PE per milliliter, was misinterpreted as the total number of PEs per vial. Also, health professionals were reportedly confused by the use of “phenytoin equivalents,” a prodrug concept introduced for this product. As a result, massive fosphenytoin overdoses were mistakenly administered.

Fosphenytoin is a prodrug, a compound that undergoes chemical conversion in the body to become the therapeutically active compound phenytoin. Cerebyx dosage will continue to be expressed in PEs. This terminology was adopted in an effort to simplify therapeutic conversions between phenytoin sodium and fosphenytoin sodium (i.e., 500 mg of phenytoin sodium injection is equal to 500 mg PE of fosphenytoin sodium injection). The manufacturer pointed out that by using PEs, prescribers will not have to make dosing adjustments when converting from phenytoin sodium to Cerebyx or vice versa. To reduce the risk of incorrect dosing, all healthcare providers should prescribe and dispense Cerebyx in PEs.

Parke-Davis has taken action to prevent future errors. The labeling for Cerebyx vials and packaging has been changed to further reinforce the total amount of drug in each vial. This is effective for both the 2- and 10-ml vials of the product. Although the new labeling further clarifies the total quantity of drug contained in the vial, the concentration of Cerebyx will remain 50 mg PE/ml.

Case 51832, 51845: Line extension creates confusion

Muro Pharmaceutical, Inc., introduced a new line extension, Prelone[®] Syrup 5 mg/5 ml (prednisolone, a steroid), to the existing product, Prelone 15 mg/5 ml. Because only one strength of Prelone had been available for many years, it was a general practice for prescribers to write for “Prelone Syrup” without indicating the strength.

Discussion: Manufacturers need to consider the transition time needed by practitioners to become familiar with the existence of a new strength. Confusion of this type is also seen when “long-acting” versions of a product are added to a product line, thereby changing the dosing regime to less frequent intervals. The product name is prescribed without the “long-acting” designation, and a medication error results. In similar cases, suffixes also cause errors when the product line extension adds a second strength and places a suffix such as “XL” after the product name to indicate long-acting release. Prescribers omit the suffix out of habit (for the initial formulation), and the patient receives the shorter-acting medication at the long-acting interval.

Muro Pharmaceutical, Inc., anticipated that a new concentration of an established product could indeed cause confusion and developed new packaging for both

concentrations. Muro also sent mailers that announced the availability of two concentrations of Prelone to 32,000 pediatricians and 65,000 pharmacies, wholesalers, HMOs, and PPOs.

Cases 50446, 50499, 50519, 50534, 50736, 50820, 50918: Poor contrast compromises readability

The unit-dose packaging of the quinolone Levaquin (levofloxacin, an antibacterial) is silver foil with black letters. The dose is reverse shaded. A reporter noted that the packages have to be held at just the right angle to be able to read the label. It was reported to be especially difficult to differentiate between the 250- and 500-mg strengths because the numbers were so difficult to read.

Discussion: Ortho-McNeil is redesigning the packaging for Levaquin to improve readability. Manufacturers should be aware that practitioners often operate in areas that have poor lighting. This makes double-checking the label to prevent errors even more difficult. For some products, there may not be adequate time to read the label carefully the first time without having to look again because of poor contrast. The use of embossed printing on plastic containers has also been reported to be difficult to read because there is no contrast and no paper label to aid in distinguishing the products visually or identifying them properly.

Total volume is the key

Eight reports received from pharmacists expressed concern about the labeling on the Bentlyl[®] (dicyclomine hydrochloride, an antispasmodic) 2-ml ampul. Practitioners reported that the label indicates only the drug concentration, 10 mg/ml, and not the total volume. Some practitioners believe the label information is incomplete. In one report, a 20-mg dose was ordered, but two ampuls were administered (4 ml total instead of 2 ml), leading to an overdose. This happened because the reporter mistook 10 mg as the total contents of one vial.

Discussion: Reports received by Hoechst Marion Roussel have prompted the company to return to the old-style labeling that includes the product's total volume data. According to the firm, this change will be implemented as quickly as possible.

Reports to the USP have identified the need for three items of information to appear on the vial or ampule: 1) the total volume, 2) the strength per milligram and 3) the total strength per total volume.

Although some manufacturers feel it would be unreasonable to include this amount of information on the container (especially containers of 1- and 2-ml sizes),

this information would assure little chance of misinterpreting the contents or strength.

Cases 040925, 050419, 041485: Wholesaler errs due to label similarity

A pharmacist reported that vials of Marsam's cefazolin sodium 1 g and 10 g appear identical in shape and have the same color flip-top closures. The pharmacy ordered the 1-g product from the wholesaler. Instead, the wholesaler sent 10-g bulk vials of cefazolin sodium along with stickers for the 1-g vial. The pharmacy, which does not normally stock the 10-g vials, interspersed the 10-g vials with the 1-g vials in their stock. Several vials were reconstituted in error. Fortunately, no patients received the wrong dose of cefazolin sodium.

In another reported incident, a pharmacist ordered the 10-g vials of cefazolin sodium but received the 1-g vials in error. Intending to reconstitute and then divide the 10-g vials into 1-g doses, a pharmacy technician inadvertently reconstituted the 1-g vials and proceeded to divide the total solution of each vial into ten 100-mg doses. Some of the prepared 100-mg doses of cefazolin sodium were administered to patients instead of their scheduled 1-g doses. No adverse effects to the patients were reported. The pharmacist felt the error occurred, in part, because the vials are identical in size and have similar labels.

Discussion: The pharmacist suggested that the color of the flip-top of the 10-g vial be changed. The company replied that although it is common practice to use color-coded labels and flip-tops to differentiate product lines or strengths, it tries to indicate the individual products in other ways, e.g., by varying the style and format of the label. Marsam revised the labeling of the cefazolin sodium 10-g bulk vial to help distinguish it from the 1-g, single-dose vial. The newly revised labeling included the following:

- the word "BULK" added in two places on the side panel
- screened color added to the box surrounding the product name
- "10 grams" printed in color
- the product name and strength on the back of the label printed in color

The use of color-differentiation is favored, whereas the use of color-coding is controversial because of the limited number of colors, color-blindness in our population, and inappropriate reliance on color in lieu of reading the label.

Case 042031: Packaged measuring devices

An order was written for 30 mg of Cyclosporine (an immunosuppressant) oral solution to be administered to a pediatric patient. However, for several days, the nurse

administered 300 mg, believing that the syringe was calibrated in milligrams, not in milliliters. The oral solution is available as 100 mg/ml. As the pharmacist reviewed the error, he noted that the syringes accompanying the medication were never designed with pediatric patients in mind. It is not possible to calculate any dose less than 50 mg. It is understandable how the nurse assumed that the "3" mark was for 30 mg—it is positioned between "2.5" and "3.5" (which are European expressions for the decimals 2.5 and 3.5). To harmonize products in the global market, the manufacturer chose to follow European convention for expressing numbers, which uses commas and decimals in the reverse manner as that as in the United States.

Discussion: This error is unusual because it involves a global trade issue. Manufacturers would prefer to harmonize products used in the United States with those available in other markets. If dose preparation was centralized in the pharmacy, this error might have been avoided.

Other medication errors involving medication-dispensing devices reported to the USP have included the interchange of devices supplied with specific products. Each device packaged with a medication is calibrated for that medication based on the viscosity and concentration of the specific liquid it delivers. These devices are not calibrated in any standardized way; some are measured in milligrams (mg), others in milliliters (ml), and others in cubic centimeters (cc). Still others have calibrations for the strength per drop or per teaspoonful. Policies should be in place so that the dispensing or use of droppers or calibrated cups provided with specific medications is restricted to those medications. Manufacturers that supply droppers with a stock bottle should supply enough droppers to enable breakdown of the liquid to usable volumes. For example, one company supplied only one dropper with its 8-oz bottle of morphine sulfate, even though the more common quantities dispensed are 2- and 4- oz. Alternatively, manufacturers should package medication in the volume expected to be dispensed per medication order.

Case 52348: Abbreviations

The USP received a medication error report involving the products Neumega® (oprelvekin) and Proleukin® (aldesleukin). Oprelvekin, a recombinant human interleukin-eleven product used to stimulate platelet production in selected patients undergoing chemotherapy, is sometimes abbreviated as IL-11. Aldesleukin, a recombinant human interleukin-two derivative indicated in designated patient populations for the treatment of metastatic renal cell carcinoma, is sometimes abbreviated as IL-2.

In the reported error, a physician used the abbreviation “IL-11” when ordering oprelvekin for a patient. Unfortunately, the order was misinterpreted to be interleukin-two (i.e., the number eleven was perceived to be the Roman numeral two). Five or more healthcare professionals, including pharmacists and nurses, mistook the order to be aldesleukin. The error went undetected for 4 days, until it was noted that the inventory of aldesleukin was nearly depleted.

Discussion: Practitioners should be especially vigilant when orders for these interleukin products are received. If abbreviations have been used in an order, the order should be clarified to ensure that patients receive the intended medication. This medication error exemplifies the value of implementing prescribing guidelines, such as the recommendations adopted by the NCC MERP. Specifically, when writing an order, prescribers should avoid the use of abbreviations, including those for drug names. Drug names should not have accepted abbreviations. Reference materials sometimes refer to these abbreviations as synonyms for the approved drug names. Manufacturers should discourage the use of abbreviations because of the potential to cause medication errors.

The following cases demonstrate how products can be misused because of the systems with which they interface.

Case 052718: Electronic drug reference products

A pharmacist asked one of the clinical pharmacists for information about Cartia®. Because an electronic drug reference listed the active ingredient as aspirin, the pharmacist was prepared to substitute an aspirin product for Cartia. The clinical pharmacist recognized the new product as Cartia XT® (diltiazem, a calcium channel blocker) and prevented the error.

Discussion: The manufacturer of Cartia XT shared the reporter’s concern and contacted the electronic reference source to investigate the matter. The publisher of the electronic reference stated that a salicylate product called “Cartia” is manufactured by Lusofarmaco in Portugal and Smithkline Beecham in Australia. Both Cartia products were verified as active current products by the publisher. The publisher said it has no way of excluding the foreign marketed Cartia because it is an active product imported from a master database that contains many foreign drug products. The electronic reference is published quarterly. “Cartia XT” was entered into the database that is currently being shipped to customers, who will now be able to choose between “Cartia” and “Cartia XT”. This should reduce confusion between the products.

As with many hard copy drug reference books, electronic drug references have lag time between production and the customer’s receipt of the reference databases. Unfortunately, this may result in inaccurate/outdated information and omission of current drug information, causing confusion and misinterpretation of drug information by the users. Healthcare providers should realize that reference sources, including electronic reference databases, are not infallible, and that they are only good as their contents of updated information. As a safeguard, the healthcare providers should make it a practice to check at least two different drug information sources to confirm information.

Case 52125: Computers and processing software

A pharmacist entered an order for Diflucan (an antifungal) for a patient who had been receiving Propulsid (a gastrointestinal emptying adjunct), which is a documented drug interaction. The pharmacy computer system had multiple drug interaction screens. The pharmacist passed these screens by pressing “next screen” without any resistance by the system for this dangerous drug interaction. The patient received two doses of Diflucan. On the second day, the patient coded and later died.

Case 51088

A patient died after 12 mg of I.V. Colchicine (an antidiarrheal medication) was given instead of 2 mg I.V. “until diarrhea,” as ordered. The physician was contacted by the pharmacist but the physician insisted on the dose. The computer program did not warn about the dangerous dose, and nurses had no idea they were giving an overdose.

Case 50908

Amoxicillin was prescribed and dispensed to a patient with a penicillin allergy. The front of the patient’s chart was not marked for an allergy, and the problem list indicating the allergy was covered with a misfiled document. The pharmacy software program does not screen for allergies, and the pharmacy profile was not marked with any allergies.

Digoxin pediatric elixir

Because the computer in one facility was limited to entering doses in milligrams, a neonate patient’s 20 microgram dose of digoxin first had to be converted to the equivalent milligram dose before it could be entered into the computer. A pharmacist incorrectly converted the 20 mcg dose and then entered it into the computer as 0.2 mg (instead of 0.02 mg). Consequently, the patient received four 200 microgram doses of digoxin instead of the 20 microgram

dose as ordered. The patient experienced digoxin toxicity before the error was discovered.

Teaspoonful versus mL

By default, a certain software program printed "teaspoonful" for any syrup preparation when a numerical figure was not followed by a specific measure, such as mL, for the dose. A prescription for $\frac{1}{2}$ mL albuterol syrup every 6 h for a 9-week-old infant was presented to the pharmacy, and the pharmacist entered " $\frac{1}{2}$ " into the computer but did not enter mL. Therefore, by default, the label printed $\frac{1}{2}$ teaspoonful every 6 h if needed for wheezing. The child was administered the overdose and was consequently admitted to the hospital emergency room for observation. Fortunately, the child was released with no permanent damage.

One versus one-half

New computer software was used to enter the directions for a cough medicine with a dose of "1–2 teaspoonsful." Instead, the new software printed the label as $\frac{1}{2}$ teaspoonful. The pharmacist did not check the label against the prescription and dispensed the product with the incorrect directions on the label.

Discussion: Computerized systems have become important tools in today's pharmacy settings. Computers have made prescription processing faster, easier, and more efficient. Computers have also provided for patient information to be readily available. However, as reliance on computers systems grows, care should be taken not to become totally dependent on these systems as the sole check in preventing medication errors.

Similar drug names

Confusion over similarity of drug names, either written or spoken, accounts for approximately one-quarter of all reports to the USP MER Program. Such confusion is compounded by illegible handwriting, incomplete knowledge of drug names, newly available products, similar packaging or labeling, and incorrect selection of a similar name from a computerized product list. The USP has produced a list of more than 1000 drug name pairs that have been reported as confusing. Manufacturers should refer to this list when selecting drug names. Recently, the USP voted to change Amrinone to Inamrinone when it was being confused with Amiodarone and caused fatal errors. This type of change is expensive to the industry and can be avoided by considering the potential for similarity in advance. Technologies and testing protocols, including voice and handwriting recognition, are available to help determine whether a drug name looks or sounds like another.

SUMMARY

The ability to predict error and thus avoid it is the focus of the science of human factors engineering. The adaptation of this science to the medication use process can help to predict the chances that a medication error will occur. Pharmaceutical manufacturers should design products including their names, labeling, and packaging so that errors can be avoided and safer systems and healthcare delivery result.

REFERENCES

1. Institute of Medicine. *To Err is Human: Building a Safer Healthcare System*; www.nas.org (accessed Dec 1999), November 1999.
2. Bates, D.W.; Spell, N.; Cullen, D.J.; Burdick, E.; Laird, N.; Peterson, L.A.; et al. The Costs of Adverse Drug Events in Hospitalized Patients. *J. Am. Med. Assoc.* **1997**, *277* (4), 307–311.
3. Phillips, D.P.; Christenfeld, N.; Glynn, L.M. Increase in U.S. Medication-Error Deaths between 1983 and 1993 [Research Letter]. *Lancet* **1998**, *351*, 643–644.
4. Bates, D.W. Frequency, Consequences and Prevention of Adverse Drug Events. *J. Qual. Clin. Pract.* **1999**, *19*, 13–17.
5. Flynn, E.A.; Barker, K.N. Medication Errors Research. *Medication Errors*; Cohen, M.R., Ed.; American Pharmaceutical Association: Washington, DC, 1999; 6.4–6.5.
6. Bates, D.W.; Cullen, D.J.; Laird, N.; Peterson, L.A.; Small, S.D.; Servi, D.S. et al. Incidence of Adverse Drug Events and Potential Adverse Drug Events. *J. Am. Med. Assoc.* **1995**, *274* (1), 29–34.
7. Bates, D.W. Preventing Medication Errors. *Medication Use: A Systems Approach to Reducing Errors*; Cousins, D.D., Ed.; Joint Commission on Accreditation of Health Care Organizations: 1999; 57–73.
8. Leape, L.L.; Bates, D.W.; Cullen, D.; Cooper, J.; Demonaco, H.J.; Gollivan, T. et al. Systems Analysis of Adverse Drug Events. *J. Am. Med. Assoc.* **1995**, *274* (1), 35–43.
9. Leape, L.L.; Woods, D.; Hatlie, M.; Kizer, K.; Schroeder, S.A.; Lundberg, G.D. Promoting Patient Safety by Preventing Medical Error [Editorial]. *J. Am. Med. Assoc.* **1998**, *280* (16), 1444–1445.
10. Phillips, D.L. "New Look" Reflects Changing Style of Patient Safety Enhancement [Medical News and Perspectives]. *J. Am. Med. Assoc.* **1999**, *281* (3), www.jama.ama-assn.org/issues/v281n3/full/jmn0120-1.html (accessed Jan 2000).
11. Bates, D.W.; Miller, E.B.; Cullen, D.J.; Burdick, L.; Williams, L.; Laird, N. Patient Risk Factors for Adverse Drug Events in Hospitalized Patients. *Arch. Intern. Med.* **1999**, *159*, 2553–2560.
12. Classen, D.C.; Pestotnik, S.L.; Evans, S. et al. Adverse Drug Events in Hospitalized Patients. *J. Am. Med. Assoc.* **1997**, *277*, 301–306.

MATHEMATICAL MODELING OF PHARMACEUTICAL DATA

David W.A. Bourne

University of Oklahoma, Oklahoma City, Oklahoma, U.S.A.

INTRODUCTION

The intent of this article is to give the reader an overview of mathematical modeling as it can be applied to pharmaceutical and especially pharmacokinetic data. The emphasis is on the application of nonlinear regression techniques for the determination of suitable models and best-fit parameter estimates. A number of topics are discussed, including simulation of models described as explicit, implicit, or differential equations; numerical integration methods; optimization algorithms; weighting schemes; evaluation of program output; and approaches for designing future experiments.

The discussion begins with a rationale for modeling and a general approach to the development of suitable models. A brief review of various pharmacokinetic and other pharmaceutical models are followed by a description of methods used to determine calculated values of the dependent variable. The next step in modeling the data involves parameter estimation using a suitable optimization procedure. Not all data points have the same error or uncertainty, thus a weighting scheme should be considered. A variety of schemes are commonly used and all have some advantages. An extremely important part of any modeling exercise is the evaluation of the computer output. Each computer program will provide various information such as plots of the data and residuals, tables of observed, calculated, and residual data, and a variety of statistical parameters. These all can be important in the overall evaluation of the results. The first experiment and the first model are usually not final. Determination of the best model to explain the available data may be required. More experiments may be necessary, thus optimal sampling times or sample sites may need to be explored.

REASONS FOR MODELS

Successful modeling of experimental data should result in a concise and, it is hoped, precise representation of the real world. Successful models allow the exploration of mechanisms, simplify the reporting of experimental results

by the condensation of the data collected allowing useful prediction of future results.

Mechanisms Exploration

Modeling allows the investigator to explore the underlying mechanism controlling various processes. The investigator may develop a hypothesis theoretically or based on preliminary observations. When fully developed, this hypothesis may result in one or more mathematical descriptions. On the basis of this hypothesis, experiments are planned and conducted and the results are analyzed. The most successful hypothesis or mathematical model will best explain the collected data. Variations between the predicted and measured observations may lead to refinement of the hypothesis (or the experimental techniques). An extension of the original experiments may be used to support and confirm the hypothesis. For example, a pharmacokinetic model developed after single-dose administration of a drug may be tested by changing the route of administration or by using multiple-dose administrations.

Data Condensation

Mathematical models allow for considerable data compression. Data that fill many notebooks and/or tables may be summarized by means of the few parameters of a well chosen model. Plasma concentration versus time data collected during a pharmacokinetic study in many patients may be summarized using a volume of distribution term and an elimination half-life. Other models may include more parameters, but there is always a considerable compression in the information required to describe the overall results of the study. An extensive drug stability study involving many samples stored at various elevated temperatures may be summarized as a single time for 10% decomposition at room temperature. Determination of appropriate models can be a very useful method of summarizing an experimental study.

Prediction

A thoroughly tested model with numerically determined parameter values should have very good prediction

potential. A successful drug regimen design relies on the development of suitable pharmacokinetic models. With a good pharmacokinetic model and parameter values determined in similar patients, it becomes possible to calculate successful dosage regimens for future patients, either directly or by using a Bayesian analysis with limited patient data.

GENERAL METHOD

The detailed design of an experimental study varies considerably depending on the area of research, the objectives of the study, the study material available, and available methods of data collection. As many variables as possible should be controlled or measured. Study material and assay methods should be well defined. The sample schedule should be optimized to collect the most informative data. (Optimal sampling strategies are discussed later.) Once the experiment is designed and the data collected, a mathematical model can be used to analyze the results. Typically, a number of models will be evaluated, as will the utility of the data and data schedule. The cycle may include revised analyses and/or revised experiments (Fig. 1). Methods for distinguishing between models and for selecting optimal sampling times are covered later in more detail.

This discussion of mathematical modeling is limited to methods based on the assumption of error in the y (dependent) term only. The objective or minimization function will generally be restricted to the sum of the weighted residuals between the observed and calculated data, weighted sum of squared residuals (WSS). The objective of the mathematical modeling approaches is to adjust the parameter values so that a minimum value of the WSS is achieved.

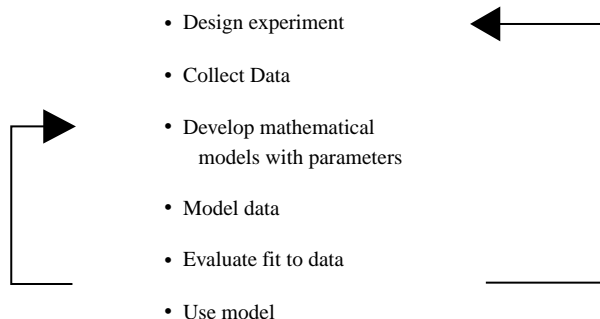


Fig. 1 Mathematical modeling: A general approach.

Error in y Term Alone

With any experiment, measurement or random error may be found in the independent variables, the dependent variables, or both. However, most common modeling programs assume that the independent variable can be set or measured without significant error. Thus, all the error is assumed to be with the dependent variable. Therefore, it is important that the independent variables are carefully measured or controlled. For example, when time is the independent variable, actual times should be recorded and used in the data analyses rather than scheduled times. Although most modeling of pharmaceutical data has confined the error to the y term, there have been attempts to include error in the x value during the fitting process (1).

Least-Squares Criteria

The objective of any modeling exercise is to place a calculated line (based on some relevant mathematical model) as close to the data collected as possible. The difference between individual data points and the calculated line (in a vertical direction—no error in the x terms) is called the residual. The sum of residuals could be zero even with very large residuals for individual points if the negative residuals canceled the positive values. An absolute residual might solve this problem, but more usefully, the squared residual will also achieve the desired result. This is the least-squares criterion. An extension of this is to weight each data point by the inverse of the estimated variance. This term is the objective function, WSS, shown in Equation (1), calculated for “ n ” data points:

$$\text{WSS} = \sum_{i=1}^{i=n} (\text{Calculated Value}_i - \text{Observed Value}_i)^2 \times \text{Weight}_i \quad (1)$$

With normal weighted nonlinear regression, the objective is to minimize this objective function. As it is described later additional terms may be added to the objective function when performing extended least-squares (Eq. 23) or Bayesian analyses (Eq. 24).

Parameter Adjustment

The minimization of the WSS value is achieved by systematic modification of the values of the adjustable parameters of the model chosen. Changing the parameters moves the calculated line and thus the WSS.

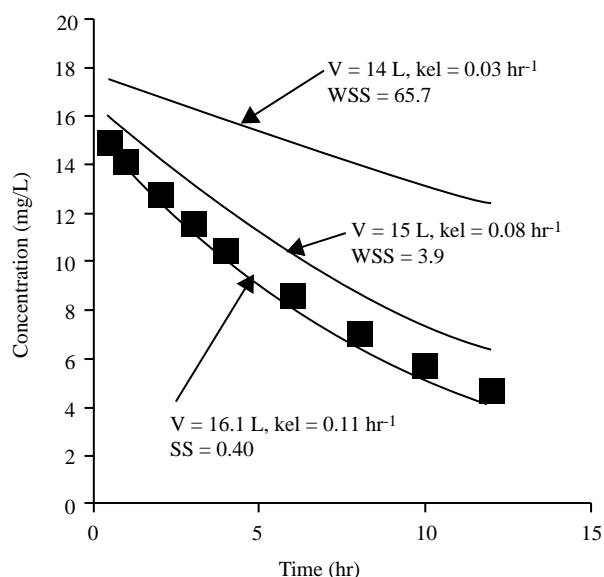


Fig. 2 Illustration of the effect of parameter values on WSS value. Calculated using a one-compartment IV bolus model with weighting by $1/\text{Val}^2$.

Achieving the minimum WSS most efficiently is the objective of the minimization algorithms incorporated into the various nonlinear regression programs. Fig. 2 illustrates how the WSS value changes with different parameter values.

MATHEMATICAL MODELS

Most pharmaceutical results can be quantitated and represented by a mathematical model. The model thus developed consists of the dependent variables measured during the experiments (y), independent variables (x), parameters (p ; adjustable or unknown), and constants (c). Thus, a mathematical model may have the general form shown in Eq. 2.

$$y = f(x, p, c) \quad (2)$$

A number of equations will be presented as examples of mathematical models in use in pharmacokinetics and pharmaceutics.

Pharmacokinetics

Pharmacokinetics is the mathematical description of drug absorption and disposition. As such, it lends itself to mathematical modeling and the use of the models developed previously. Within the field of pharmacokinetics,

a different types of models are used. These include compartmental, physiologically-based, and pharmacodynamic models.

Compartmental models: With this type of model, the subject is represented as a number of well mixed compartments. When all the rate processes are first-order, equations in the form of sums of exponential terms are commonly used. Thus a two-compartment model is illustrated by Eq. 3:

$$Cp = A \cdot \exp(-\alpha \cdot t) + B \cdot \exp(-\beta \cdot t) \quad (3)$$

where Cp is the dependent variable (concentration); t is the independent variable (time); and A , B , α , and β are parameters.

With more involved compartmental models, including, for example, Michaelis–Menten elimination kinetics, the model may be described more easily using differential equations. Thus, for a drug eliminated by a first-order excretion process and a Michaelis–Menten metabolic process, Eq. 4 holds:

$$\frac{dCp}{dt} = - \frac{Vm \cdot Cp}{Km + Cp} \quad (4)$$

$$\text{Initial condition} = \frac{\text{Dose}}{V}$$

where dCp/dt is rate of change of drug concentration; Cp is the dependent variable; ke , Vm , Km , and V are parameters; and Dose is a constant. The differential equation is solved numerically to provide Cp values versus time. More information concerning the various equations associated with compartmental pharmacokinetic models can be found in the text by Gibaldi and Perrier (2), Bourne, Triggs, and Eadie (3), Wagner (4, 5).

Physiologically based pharmacokinetic models (PBPK): In contrast to compartmental models where most of the parameters are abstract and empirically based, PBPK models include many physiologically relevant parameters. These parameters include tissue or organ volumes, blood or plasma flow rates, and partition coefficients between blood and tissue. In some studies, protein binding or tissue binding parameters may be included. Drug elimination is commonly expressed as renal, hepatic, or other clearance. These models are expressed as mass balance-based differential equations. Thus, the rate of change of drug concentration in any organ or tissue of interest is given by Eq. 5:

$$V \cdot \frac{dC}{dt} = [\text{Mass of drug flowing in} - \text{Mass of drug flowing out}] \quad (5)$$

For a single noneliminating organ, this becomes Eq. 6:

$$V \cdot \frac{dC_p}{dt} = \left[Q_i \cdot C_b - \frac{Q_i \cdot C_i}{R_i} \right] \quad (6)$$

where C_i and C_b (drug concentration in tissue “ i ” and blood) are the dependent variables; t is the independent variable (time); and Q_i , R_i , and V_i are tissue blood flow rate, tissue to blood partition coefficient, and tissue volume, respectively. More detailed descriptions of these types of models can be found in the text by Gibaldi and Perrier (2) or in the papers of Bischoff and Dedrick (4, 6).

Pharmacodynamic models: These models are designed to relate drug effect (dependent variable) to drug concentration or time (independent variable). Included in these models may be disposition parameters, such as volumes and rate constants, as well as drug effect parameters such E_{\max} (maximum effect) and $EC_{50\%}$ (concentration eliciting 50% of maximum effect). The Hill equation has been used successfully to describe the relationship between effect and drug concentration. More information about these types of models may be found in the text by Gibaldi and Perrier (2) or the paper by Sheiner et al (7). Jusko proposed an alternative model involving equilibrium with the hypothetical receptor site (8). Recently, Jusko et al. have described four basic models to describe indirect drug effects that include tolerance or rebound, which should be useful (9, 10).

Stability

Drug stability studies may require the use of zero-, first-, or second-order reaction models. Reaction rates may be measured as a function of pH and buffer concentrations to determine the influence of various catalysis possibilities. For example, the hydrolysis of a compound may be pseudo first-order, as shown in Eq. 7:

$$\frac{dC}{dt} = -k' \cdot [C] \quad (7)$$

where the first-order rate constant, k' , can be expressed as in Eq. 8:

$$k' = k_0 + k_H \cdot [H^+] + k_{OH} \cdot [OH^-] + k_{HB} \cdot [HB] + k_B \cdot [B^-] \quad (8)$$

Here the terms, k_0 , k_H , k_{OH} , k_{HB} , and k_B , refer to noncatalyzed and specific acid, specific base, buffer (acid), and buffer (base) catalysis. Additional details regarding the models required to describe drug stability studies can be found in textbooks such as the one by Carstensen (11).

SIMULATION OF DATA

Mathematical models involve many different types of equations. These may be explicit, implicit, or differential equations. Dependent variables expressed as explicit equations are easily calculated. For example, plasma concentrations after a single oral dose can be calculated as in Eq. 9:

$$C_p = \frac{F \cdot \text{Dose} \cdot ka}{V \cdot (ka - kel)} \cdot \{ \exp(-kel \cdot t) - \exp(-ka \cdot t) \} \quad (9)$$

Implicit equations include the dependent term in a form not readily separated from the other terms in the equation. One example in Equation (10) for drug concentrations after IV bolus administration, following Michaelis–Menten elimination kinetics as described by Wagner (12):

$$C_0 - C + K_m \cdot \ln\left(\frac{C_0}{C}\right) = V_m \cdot t \quad (10)$$

Here the dependent variable C is found in two parts of the equation and cannot be solved directly. Thus, an iterative method (13) may be necessary to determine the value of C at each value of t , the independent variable.

Integration Methods

Many processes in the pharmaceutical sciences are dynamic. Thus, models of these processes may commonly involve differential equations, which must be numerically integrated at each step in the optimization procedure. A variety of numerical integration methods can be used, and some of these are discussed later.

When a model is expressed as differential equations, they usually must be integrated before the optimization can be performed. Thus, the differential terms must be converted into values for the dependent variable. For example, the differential Eq. 11:

$$\frac{dC_p}{dt} = -kel \cdot C_p \quad (11)$$

could be converted into the integrated Eq. 12:

$$C_p = \frac{\text{Dose}}{V} \cdot \exp(-kel \cdot t) \quad (12)$$

Conversion from the differential equation form can be performed by analytical methods or numerically, using appropriate computer algorithms. Analytical integration may be mathematically intensive (or impossible), and numerical integration tends to be computationally intensive. Selection of the best approach depends on the equations involved.

Laplace transforms

Integration of differential equations can be quite involved and the subject of complete college courses (14). However, one useful technique is the method of Laplace transforms. An excellent tutorial is presented in the two papers by Mayersohn and Gibaldi (15, 16). Benet and Turi (17), and Benet (18) present more advanced techniques, such as the input and output disposition functions and the "fingerprint" technique for the solution of differential equations.

Another approach presented by Yamaoka et al. (19), uses a fast inverse Laplace transform to generate the integrated equation data. Thus the model is described in terms of the Laplace transform equations and solved numerically.

Numerical integration

Numerical integration also starts with a number of differential equations and initial conditions. However, now the computer program performs the integration. Using carefully defined algorithms, the computer program is able to start with the initial condition(s) and project forward with increasing values of the independent variable (often time) according to the slope given by the differential equation(s). With a careful selection of algorithm, accurate approximations to the required result can be calculated efficiently. The general approach can be understood more easily by looking at a simple point-slope method such as Euler's method. However, these methods are not

necessarily the most efficient. Runge–Kutta methods, including the Fehlberg modification, have advantages with the usual differential equations encountered in pharmacokinetics and pharmaceuticals. For systems where the rate constants are widely disparate, so-called "stiff" systems, other methods become more useful.

Point-slope methods: Euler's method follows directly from the initial condition as a starting point and the differential equation as the slope (Fig. 3). Consider the simple model of a single differential Eq. 13 with one first-order rate process:

$$\frac{dC_p}{dt} = -k_{el} \cdot C_p \quad (13)$$

with an initial condition of C_p at time = 0 of C_p^0 . Thus, the change in the value of C_p from time = 0 to time = Δt is shown by Eq. 14:

$$\Delta C_p = \frac{dC_p}{dt} \cdot \Delta t \quad (14)$$

and the value of C_p at time Δt by Eq. 15:

$$C_p^{\Delta t} = C_p^0 - k_{el} \cdot C_p \cdot \Delta t \quad (15)$$

If Δt is sufficiently small, relatively accurate calculations can be performed. Unfortunately, very small step sizes are often required for reasonable accuracy and truncation errors then can be significant. Thus, this method is not generally used with nonlinear regression programs.

Euler's methods can be derived from a more general Taylor's algorithm approach to numerical integration. Assuming a first-order differential equation with an initial value such as $[dy/dx] = y' = \text{function of } x$, and $y = f(x, y)$ with $y(x_0) = y_0$, if the $f(x, y)$ can be differentiated with respect to x and y , then the value of y at $x = (x_0 + h)$ can be found from the Taylor series expansion about the point $x = x_0$ with the help of Eq. 16:

$$y(x) = y_0 + hy'(x_0) + \frac{h^2}{2!} \cdot y''(x_0) + \dots \quad (16)$$

where h is an increment in x , that is Δt . It can be seen that including only the first two terms on the right side of Eq. 16 leads directly to Euler's method. This method gives increased accuracy with more terms on the right side of the equation. However, it is not readily applicable as a general-purpose procedure because of the requirement for higher-order derivatives. It has been useful theoretically in the development of more general-purpose methods such as the Runge–Kutta methods.

Runge–Kutta methods: The Runge–Kutta methods attempt to improve the accuracy of the calculation and avoid the evaluation of higher-order derivatives. The

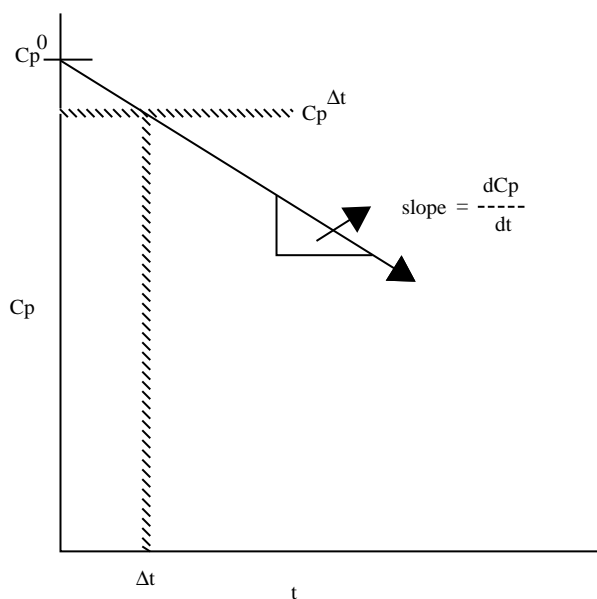


Fig. 3 Illustration of the point-slope method.

general approach is to determine values of the first-order differential equation at subintervals of the chosen step size. The way in which these subinterval results are combined can be derived from the Taylor series expansion. The most common of these Runge–Kutta methods is the classical fourth-order version. The values of the function are calculated four times for each step to improve the accuracy of the calculation. For complex functions, this may be a disadvantage. Another problem with this method is that it lacks any automatic step-size control. Additional details of these Runge–Kutta methods may be found in textbooks on numerical analysis (13).

The Runge–Kutta–Fehlberg is a further modification of the Runge–Kutta fourth-order method. It uses a fifth function evaluation to determine the appropriate step size. This method appears to be very efficient for nonstiff systems of differential equations. Additional details regarding this method and a computer listing can be found in a report by Fehlberg (20) and in the chapter by Watt and Shampine (21).

Multistep methods: Another class of methods, called multistep methods, involves the use of more than one previous value in the calculation of the “next” value. With the slope-point and Runge–Kutta methods, a single starting point (and the differential equation) is all that is required for the calculation of the next value at $(x + h)$. With a commonly used four-step method, values at $(x-3h)$, $(x-2h)$, $(x-h)$, and x are required in the calculation of y at $(x + h)$. One such method is that of Adams and Bashford (22).

The advantage of these techniques is that they require only one (additional) function evaluation for each step compared with four or five evaluations for the typical Runge–Kutta method. Thus, they should be faster. Their principle disadvantage is that they are not self-starting. Another method, Euler’s or Runge–Kutta, is required to calculate the first (three) values at x , $(x + h)$, and $(x + 2h)$ before the Adams–Bashford multistep equation can be used to continue the calculation. Additional starting values are also required whenever the step size is changed.

An extension of the multistep methods is the predictor-corrector approach. Here, the Adams–Bashford equation may be used to calculate a “predicted” value for y at $(x + h)$. Then a second, corrector, equation is used to refine the value of y . If the difference between the predictor and corrector values is within specified error limits, the calculation is continued to the next step, otherwise the step size will be adjusted to maintain the error limits specified. With fewer function evaluations per step, these methods can be faster than the Runge–Kutta methods; however, they are not self-starting.

Methods for stiff equations: Stiff systems of differential equations occur commonly with physiologically based models. That is, very fast processes and very slow processes occur in the same model. The ratio between the fastest and the slowest rate constant can be used as an index of stiffness. With ratios greater than 500, a number of the “nonstiff” numerical integration methods become very inefficient. Gear’s method for solving stiff systems has achieved wide acceptance (23, 24) and has been incorporated into a number of nonlinear regression programs such as SAAM II and Boomer. Adam and Gear’s methods are both included in Gear’s DIFSUB algorithm. The user has to specify which method they wish to use. The differential equation solver LSODA included in the ADAPT II program automatically switches between Adam and Gear’s methods as required.

FITTING MODELS TO DATA

Fitting the model to the observed data is an important task. Each mathematical model studied consists of independent and dependent variables, constants (possibly), and parameters that have to be estimated. The objective is to reduce the overall difference between the observed data and the calculated points by adjusting the values of the parameters. As mentioned earlier, the methods to be discussed assume that there is no error in the independent variable. Also, in general, the criteria of “best” fit will be the weighted sum of squared residuals between the observed and calculated data, the WSS. The chosen model can be validated in part by accurately describing the observed data.

Graphical Methods

It is often possible to manipulate the mathematical model into the form of a straight-line equation. For example, the one compartment pharmacokinetic model after an IV bolus can be expressed as a differential equation or as an exponential equation as shown earlier. By taking the natural log of the exponential equation, Eq. 17 in the form of a straight line can be derived:

$$\ln(Cp) = \ln(Cp^0) - kel \cdot t \quad (17)$$

Thus, by graphing the data on semilog graph paper and drawing a straight line through the data it is possible to read Cp^0 from the intercept and calculate kel from the natural log slope, as in Eq. 18:

$$kel = -\text{slope} = \frac{\ln(Cp^1) - \ln(Cp^2)}{t_2 - t_1} \quad (18)$$

Graphical methods have a number of advantages. They are usually fast and, when performed carefully, can give accurate results. A major plus for the method is the ability to “see” the data and visually determine whether there are systematic deviations from the chosen mathematical model. Graphical evaluation of the data may lead to consideration of alternative models. Graphical methods can be very useful in the estimation of initial parameter values. Unfortunately, these methods are not universally available, because it may not be possible to produce a straight-line equation. When manipulating the model to produce a straight line, the observed data is often “transformed,” which may involve taking the logarithm or the reciprocal of the observed data. With other models, more involved transforms are required. This not only distorts the data but also the error or variance of the data. When using linear regression by the graphical approach, it is assumed that the error in each x -value is similar. Once the data are transformed, this may not be true. Another problem that is especially relevant with the use of log transforms and semilog graphical data representation involves the visual distortion that can occur. Typically, much of the graph is taken up by the least accurate data points; i. e., the last few data points at the lowest (and thus, the least well defined) concentrations may dominate the graph. Thus, although graphical techniques such as semilog data plots may be useful in the determination of initial estimates, they are not as suitable for complete data analysis or presentation of the final results.

Other approaches have been used for more complex models. These include curve stripping or the method of residuals (2), either manually or using a computer program

such as CSTRIP and ESTRIP (25, 26). These techniques can separate a multiexponential curve into its component parts for initial estimates. Other techniques include deconvolution methods specific to the one and two compartment pharmacokinetic models (27, 28). The objective of the deconvolution method is to mathematically “subtract” the results obtained after IV administration from the oral or extravascular data. This results in information about the input or absorption process alone. More general methods have been presented by various researchers that do not rely on a particular compartmental model (29–31).

Nonlinear Regression Analysis

For all mathematical models that are not “naturally” straight lines, nonlinear regression analysis is often the best approach. The observed data and the corresponding dependent variable can be analyzed without transformation. Thus, the data and the error or variance are not distorted during the analysis. If necessary, clearly defined weighting schemes can be applied. Furthermore, multiple observation sets can be readily accommodated.

Nonlinear regression analyses involve relatively complex calculations and thus are well suited to computer assistance. However, the program must have a well developed sequence of steps or algorithm to follow. Some methods are better than others. The program is asked to find the minimum point on a weighted sum of squares (objective or minimized function) surface. For two parameters, this can be represented as a three-dimensional surface (Fig. 4).

With more parameters, it is difficult to represent the shape of the WSS surface, but the program still moves to a minimum value of the WSS. Depending on the shape of this surface, there may be a number of local minima. That is, regions that are “lower” than the surrounding surface but *not* the “global” minimum. The objective of the nonlinear regression analysis is to reach the global minimum. Helpful strategies include: 1) careful calculation of initial estimates; 2) repeated analyses with different sets of initial estimates; and 3) careful evaluation of the program output, especially the plot of calculated and observed data versus time (x -value).

Many basic algorithms, each with a number of refinements, are useful in the search for a global minimum. Some of these methods are described briefly. These are the grid search, steepest descent, Gauss–Newton, Marquardt, and simplex methods.

Grid search

This method is more informative, but it can be quite slow. Its major objective is to produce a weighted sum

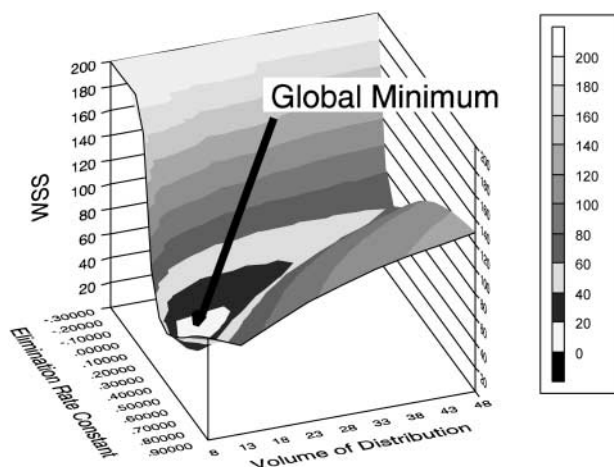


Fig. 4 A three-dimensional surface plot of WSS versus elimination rate constant and volume of distribution for a two-parameter model.

of squares surface diagram. In addition the minimum calculated WSS is estimated. The three-dimensional plot in Fig. 4 was calculated by this method. The calculation is set up by inputting the upper and lower limits of each parameter of interest. This range is split into a number of steps, and the program calculates the WSS at each point on the grid. The more steps, the smoother the surface but the longer the calculation. A three-parameter problem with only 10 steps per parameter (a relatively coarse grid) requires 1000 ($= 10 \times 10 \times 10$) calculations of the WSS. With many parameters and steps, these calculations can be very lengthy. This method is not usually used for nonlinear regression analyses, but its output may be educational.

Steepest descent

If the WSS surface is seen as similar to a geographical section with hills and valleys (at least for the two-parameter surface), the steepest-descent method would appear to follow the path of a round ball moving towards the minimum. When translated to a computer algorithm, some of the disadvantages become more apparent. The basic approach is to calculate the slope of the surface at the point of the initial (or current) parameter values. This can be calculated as $dWSS/dP$ over some small increment for each of the parameters P . By combining each of these partial derivatives over all the parameters, the direction of movement towards the minimum can be calculated. The second part of the problem is the distance h to move in the specified direction. This must be determined by finding the minimum WSS in the direction calculated from the slope. This means extra calculations of the WSS, which makes this process less efficient, especially as it approaches the minimum. The new parameter value is calculated with the help of Eq. 19:

$$P_{\text{new}} = P_{\text{old}} - h \bullet \frac{dWSS}{dP} \quad (19)$$

This new value is used as the P_{old} for the next iteration.

Gauss-Newton

If we can assume that the WSS surface between the initial estimate and the global minimum is convex, a Taylor series expansion leads to the Gauss–Newton approximation for a step closer to the minimum (32). Thus, the next point on the surface can be calculated as in Eq. 20:

$$P_{\text{new}} = P_{\text{old}} - \frac{dWSS}{dP} \bigg/ \frac{d^2WSS}{dP^2} \quad (20)$$

This method gives both a direction and a distance. Close to the global minimum, where the surface is

typically more regular, the convergence can be dramatic. Farther from the minimum, the surface may not be convex or as smooth, and the method may become slow or even move away from the minimum. Thus, good initial estimates are very helpful.

The Gauss–Newton method can be improved by using two modifications. The first is a damping process. Thus, the new WSS is calculated with the help of P_{new} calculated using Eq. 20. If this WSS is worse than the original WSS, the step size is halved, and the WSS is calculated again. This is repeated until a preset number of halvings (dampings) occur or a better WSS is achieved. Excessive damping can occur close to the minimum if the WSS surface is very flat, owing to numerical instability. In this case, the uncertainty in the final parameter values (expressed as coefficients of variation or confidence intervals) may be large.

A second method of improving the Gauss–Newton method is the Marquardt modification (33). In this case, the equation for P_{new} is modified by the addition of another term, μI , as in Eq. 21:

$$P_{\text{new}} = P_{\text{old}} - \frac{dWSS}{dP} \bigg/ \left(\frac{d^2WSS}{dP^2} + \mu I \right) \quad (21)$$

The analysis is started with a large value of μ , which has the effect of moving the equation in the direction of the steepest-descent method, which is better farther from the minimum. As the calculation approaches the minimum, μ is reduced progressively to automatically approach the Gauss–Newton method, which is better near the minimum.

Simplex

Another approach that is different from the previous methods is the simplex method of minimization (34). It involves the formation of a simplex, a geometric shape with $(m + 1)$ sides, where m is the number of parameters on the WSS surface. The WSS is calculated at each corner of the simplex and compared. The movement of the simplex across the WSS surface (toward the minimum) is controlled by a small number of rules. For example, the point with the highest WSS is reflected across the centroid (center of the simplex) to produce a new point. If this point has the lowest WSS, it is extended again. A point with a larger WSS causes the simplex to contract. By a series of such steps, the simplex moves across the WSS surface to approach the minimum value. Although the simplex method can be relatively slow, it has the advantage of computational simplicity that makes it useful for a variety of nonlinear regression problems.

Weighting Schemes

After selecting and describing an appropriate model, choosing a numerical integration method (if necessary) and a fitting algorithm, the mathematical modeler may need to select a weighting scheme for the data to be analyzed. Ideally, each data point is independent and has normally distributed error. Even when this is true, the magnitude of the variance or error in each data point may not be equal. Thus, a weight equal to the reciprocal of the variance may be applied to each data point in the calculation of the overall WSS. Information about the variance, standard deviation (SD), or coefficient of variation (CV) for each data point or how these values vary over the complete data set to be analyzed can be very useful in the selection of a “best” weighting scheme. The weighting scheme is best evaluated by looking at the weighted residual plots as described later.

If a single data set is to be analyzed (at one time) and the variance is similar for each data point, an unweighted regression may be appropriate; i.e., each data point may be given an equal weight. Unweighted regression may be appropriate when the variance is small. With physical systems, the weighting scheme may become less important. The weighting scheme becomes much more important when there is significant error or uncertainty in the data. With biological studies, such as pharmacokinetic determinations, the error in the measured data may be considerable, and the choice of weighting scheme becomes more important.

When the values of the dependent variable have considerable error and cover a relatively large range of values, a carefully selected weighting scheme may be useful or necessary. When the CV about a series of data points is constant, that is, there may be $\pm 5\%$ error in each data point, a weighting scheme can be estimated. Because the CV is equal to $SD/Value$:

$$SD = \text{Observed value} \times CV \text{ and Variance}$$

$$= \text{Observed value}^2 \times CV^2$$

With a single data set, the CV^2 can be ignored. However, when fitting more than one data set simultaneously, this value is important. Thus, the weight would be proportional to $1/\text{observed value}^2$. Other weighting schemes include a more general expression, with a and b as constants (35):

$$\text{Variance} = a \times \text{Observed value}^b$$

Provision may be made for an assay sensitivity term:

$$\text{Variance} = a + b \times \text{Observed value}^c$$

Another possibility, especially useful in a clinical setting with older concentration-versus-time data, is to apply a smaller weight to older data (36):

$$\text{Variance} = a + b \times \text{Observed value}^c \times d^{(\text{last} - \text{time})}$$

with d typically equal to 1.01 to 1.05.

Careful consideration of the data may suggest a number of equations for the estimated variance of the error in the data. The analyst should consider the overall results produced by the chosen weighting scheme and especially confirm that the weighted residual plots are satisfactory.

Selection of appropriate weighting schemes is even more important if the analyst is modeling more than one data set simultaneously. For example, drug concentration in plasma and cumulative amounts of drug eliminated into urine may have been collected. Consequently, plasma concentrations may range from 0 to 25 $\mu\text{g/ml}$ and will have quite different variance values to cumulative amount of drug in urine data ranging from 0 to 250 mg. If an unweighted analysis was undertaken, the plasma data (numerically lower values for concentrations) would tend to be ignored. Alternately, if the amount of drug excreted into urine was expressed in grams, not in milligrams, these data would be ignored during the fit because they are now much smaller numerically than the plasma concentration. Obviously, an appropriate weighting scheme must be developed. The same problem may occur if plasma concentrations after intravenous and oral administration is to be analyzed together. Again, consideration of the computer program output, especially the weighted residual plots, can help to confirm a chosen weighting scheme.

Iteratively Reweighted Least Squares

The variance formulas used in ordinary weighted least-squares analysis use the observed data values. This is easier to calculate because the observed values do not change during the fitting process. However, there can be advantages to using the calculated data values, instead of the observed values, in the variance formulas. For example, very low data points (which may have considerable error) are not given disproportionate weight. Thus, the technique of iteratively reweighted least squares may be useful (37).

Extended Least Squares

Another approach to weighting the data uses the data itself to develop the variance equation. This is the extended least squares (ELS) method (37, 38). Parameters for the variance equations are included in the fitting process as

well as parameters for the chosen model. Thus, extra data points are necessary for a successful analysis. Furthermore, the analyst must decide the form of the variance equation. An example of a typical variance equation is:

$$\text{Variance} = a \times \text{Calculated value}^b$$

where a and b are parameters to be fitted to the data. The objective function must also be altered with the ELS method because driving the calculated variance to a very large value would automatically make the WSS very small, regardless of the fit to the data. Thus, a $\ln(\text{variance})$ term is added to the WSS formula as a penalty, and the objective function to be minimized becomes Equation (22):

$$\text{Objective function} = \sum_{i=1}^{i=n} \frac{(C(\text{obs})_i - C(\text{calc})_i)^2}{\text{variance}(P, PV, t_i)} + \ln(\text{variance}(P, PV, t_i)) \quad (22)$$

where P values are the parameters of the model, PV values are the parameters of the variance equation, and n is the number of data points. This method is available in the nonlinear regression programs ADAPT II and NONMEM.

Bayesian Analysis

The analysis of clinical pharmacokinetic data offers additional challenges. Typically, the number of samples available from an individual patient can be limited. In some cases, only one or two samples may be available. If population-based pharmacokinetic values are available, it may still be possible to analyze this limited clinical information using a Bayesian approach (36). Using patient and population information, the objective function becomes a function of both the residual between the observed and calculated data (as in weighted least squares) and the residual between the population and the calculated values of the parameters, as shown in Eq. 23:

$$\text{Objective function} = \sum_{i=1}^{i=n} \frac{(C(\text{obs}) - C(\text{calc})_i)^2}{\text{Variance for data point}} + \sum_{j=1}^{j=m} \frac{(P(\text{pop})_j - P(\text{calc}))^2}{\text{Variance for parameter}_j} \quad (23)$$

Analysis of Population Data

A typical pharmacokinetic analysis of data from a single individual results in estimates for the values and variance (or SD) of each parameter. The calculated variance gives

an estimate of the closeness of fit, the adequacy of the number of data points, and the error in the data values. If this analysis is repeated in a number of individuals, it is also possible to obtain some information about the average population value and intersubject variance. With sufficient data available from each subject, this two-step approach can be efficient using any number of nonlinear regression programs.

A few programs are now available that allow the efficient simultaneous data analysis from a population of subjects. This approach has the significant advantage that the number of data points per subject can be small. However, using data from many subjects, it is possible to complete the analyses and obtain both between- and within-subject variance information. These programs include NONMEM and WinNONMIX for parametric (model dependent) analyses and NPEM when nonparametric (model independent) analyses (39) are required. This approach nicely complements the Bayesian approach. Once the population values for the pharmacokinetic parameters are obtained, it is possible to use the Bayesian estimation approach to obtain estimates of the individual patient's pharmacokinetics and optimize their drug therapy.

Evaluation of computer program output

Once the computer program completes the calculations, it is time to evaluate the output, which should be scrutinized very carefully. Each line is there for a reason. Obvious errors of data entry can be found by looking at the data tables and plots. Model misspecification may be more difficult to access. Each computer program will present users with an array of information. During the evaluation of this output, the fit to the data, the weighting scheme used, and parameter values can be assessed. Each of the items that may be output by a nonlinear regression program should be considered.

Observed and calculated data

The table of observed and calculated data includes the observed x and y values. Some programs are sensitive about how the data should be entered, and data entry errors may be common. This table should be checked to ensure that the program is working with the correct numbers. The calculated data can be compared with the observed data. If the numbers are not similar, there may be problems with parameter value limits or model specification. A residual or weighted residual is often included. These values should be of similar magnitude and randomly positive and negative. The most important use of this table is to ensure correct entry of the observed data.

Most programs provide plots of observed and calculated data versus the independent variable x . On both linear or semilog axes, these should be similar, with the calculated line running close to the observed data. If the observed and calculated data appear to be going in different directions, incorrect model specification should be suspected. If the two lines have a similar shape but do not coincide, may be a parameter value limit is causing a problem. Alternately, the program may have fallen into a local minimum. Rerunning the analysis with better (different) initial estimates may help. This plot can be useful in the identification of incorrectly entered data or outliers.

Another plot often presented is a plot of calculated data versus observed data (with a line representing identity). This can be a sensitive plot for the assessment of the model chosen. Significant systematic deviations from the identity line may indicate the need for additional parameters in the model.

Final Parameter Values

Most programs present the final “best-fit” values of the parameters with some indication of the uncertainty in these values. Parameter values far from the expected results may indicate an incorrectly specified model. Possibly a model constraint has been incorrectly specified. Incorrect units may be causing a problem. The uncertainty in the values may be expressed as an SD or CV. Alternately a CI may be calculated. A high CV, higher than 20% and especially higher than 50%, generally indicates a parameter that is not well described by the data. Maybe the model has too many parameters. High CV values may be caused by small number of data points or data with larger error. Samples may need to be taken from additional parts of the model to fully specify the parameters of interest. This becomes a problem of identifiability, as described below. Many programs also provide a correlation matrix of the estimated parameters. A high correlation (>0.95) may suggest that the model has too many parameters and that a simpler model may be satisfactory.

Weighted Residual Plots

The weighted residual plots can be very useful in the evaluation of the chosen model and weighting scheme. The basic approach is to look for a pattern in the plot. Two types of patterns may be evidence of a problem in the analysis. Residuals larger on one side of the plot compared with the other may indicate a problem with the weighting

scheme. A trend in the residuals either sloping up or down, or a “U” or inverted “U” shape, would suggest that one or two additional parameters should be added the model as seen in Fig. 5. The best result is a plot with no discernible pattern. This would support both the model and the weighting scheme selected for the analysis. A more complete discussion of residual plots can be found in the text by Draper and Smith (40).

Statistical Output

The WSS or other objective function is the major statistical parameter provided by any of the nonlinear regression programs. Typically, the analyst is looking for a minimum WSS value. However, this can be misleading if a variety of weighting schemes has been used throughout the analysis phase. For example, with observations greater than 1, the WSS when using an equal weight scheme will be much larger than with a reciprocal observed value or observed value-squared weighting scheme. Thus, comparisons of WSS (or WSS-related functions) must be made between analyses using identical weighting schemes. Some programs will normalize the sum of the weights to be equal to the number of data points. This may help in the comparison of results across the weighting scheme, but all

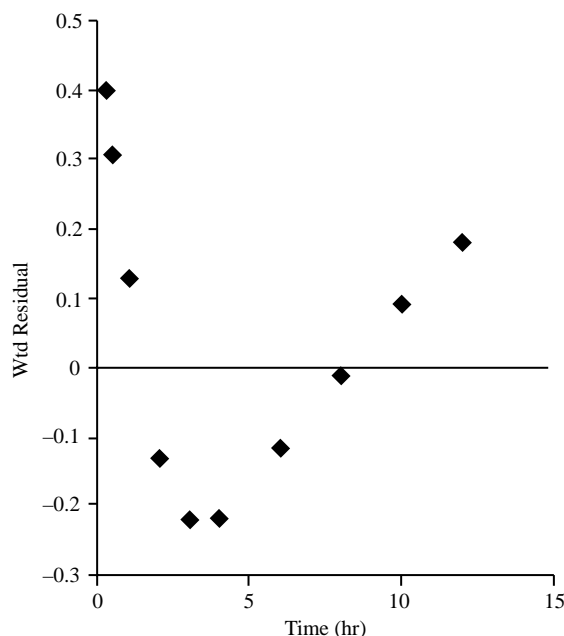


Fig. 5 Plot of weighted residual versus time after fitting data generated with a two-compartment model using a one-compartment model.

comparisons should be made with the same weighting scheme. Comparison of WSS values between similar data sets may help in the identification of sampling or analysis problems.

The correlation coefficient R and the coefficient of determination R^2 are often provided. Both terms are meant to indicate closeness of fit, with a value of 1 representing a perfect fit. With many data sets, a high value of R (greater than 0.9) seems relatively common even with a fit that visually looks poor. Thus, the analyst should not put too much emphasis on this parameter. Similarly, high values of R^2 are relatively easy to achieve. Although these parameters are widely used with unweighted linear least squares, they do not appear as useful in weighted nonlinear regression.

Comparison Between Different Models

A number of the output items already described are useful in the selection of a “best” model. Systematic deviations in the observed and calculated data plots may suggest the need for a model with additional parameters. Large values for the parameter CV values (or wide confidence intervals) may indicate a model with too many parameters. A number of patterns in the weighted residual plots may suggest the addition of parameters to the chosen model. The guiding rule is to start with the simplest model consistent with the data (and theory). As more parameters are added, it is expected that the WSS will decrease. The question becomes, has the inclusion of additional parameters produced a significant decrease in the WSS. Although developed for unweighted analyses, a number of approaches to this question have been used with weighted nonlinear regression. As suggested by Mandel (41) and Boxenbaum et al.(42), an F value can be calculated after the analysis using two different models (with different numbers of parameters), as shown by Eq. 24:

$$F(df_j - df_k, df_k) = \frac{WSS_j - WSS_k}{WSS_k} \times \frac{df_k}{df_j - df_k} \quad (24)$$

If the calculated F value is higher than the tabled F at the chosen level of significance (often 0.05), the use of the more detailed model is supported. Another commonly used parameter is the Akaike’s Information Criterion (AIC) value (43, 44). It is calculated for each model, and the model producing the lowest value (most negative value) is considered the better model. The AIC value is calculated using the number of data (n), WSS, and the number of parameters (m), as in Eq. 25:

$$AIC = n \ln(WSS) + 2 \times m \quad (25)$$

A variety of other criteria have been developed that may also be useful. It is hoped that each criterion (parameter variability, graphical output, AIC, F test, etc.) will complement each other, allowing the analyst to make an objective decision.

THE DESIGN OF FUTURE EXPERIMENTS

Once the first modeling phase is complete, the analyst may wish to consider revision of the experimental design to test the chosen model or to determine more reliable values for the parameters of the model. This leads to a number of approaches.

Extend the Range of the Current Model

Typically, a series of experiments is performed within a fixed framework. From these results, a “best” model may be selected and parameters determined. One test of this model is to make predictions beyond the range of the original experiments. These predictions can be tested by performing more experiments in the region of the new predictions and determine how well the model simulates the data. For example, a single-dose study may be conducted with data collected and analyzed to give a two-compartment pharmacokinetic model as most appropriate. This model could be tested by predicting and measuring drug concentrations after multiple-dose administration. Alternately a different route of administration may be used.

Identifiability and the Determination of Appropriate Sampling Sites

When working with complex models, it is easy to try to build a model more complicated than the data allow. For example, if a drug is metabolized and excreted unchanged, it is not possible to identify or determine a rate constant for excretion and for metabolism without measuring either drug or metabolite in urine. If the analyst attempted to model this overparameterized model with just drug-in-plasma data, the uncertainty (CV values, etc.) about the parameter values would be very high no matter how many data points were collected or how accurately they were measured. This is a problem in identifiability. The problem is to determine whether the minimum number of sampling sites for a particular model or a given number of sampling sites can adequately define a proposed model. There are a

number of approaches to this problem. For linear, time-invariant systems, the Laplace transform method is convenient, whereas the Taylor series method is a more general analytical method (45–47). Numerical approaches have been presented by Jacquez et al. (48, 49) with the development of the IDENT computer program. Another method is to simulate error-free data for the sample sites and the model chosen and “refit” these data. If the uncertainties in the parameters estimated are large (using numerous, error-free data), then an identifiability problem should be expected. The identifiability of a sample and model set should be determined during the initial experimental design stage. This can avoid considerable time and expense.

Determine Optimal Sampling Times to Reduce Parameter Uncertainty

Once the sampling sites have been identified, the next question is, at what times should the samples be collected. Without any previous knowledge of the model or system, sampling times are typically determined from intuition or from similar studies. Because of the uncertainty in the expected result, more samples are typically collected over as wide a range in time as possible. Once these first (pilot study) results have been analyzed, it is possible to make a more rational decision regarding the best sampling times. Some methods use the model selected and the parameter values first determined to obtain the best sampling times. The program SAMPLE, part of the ADAPT II package, uses the C- or D-optimality method. The result typically is one optimal sample time for each parameter to be estimated. Many investigators are uneasy about such a small sample number, thus additional samples may be included. By repeating the SAMPLE calculation with a range of parameter values, it would be possible to determine additional sample times. These times should give the “best” information about the parameters of interest.

Computer Programs

A large number of computer programs are available to perform the functions of mathematical modeling. Some of these are listed below. See <http://www.boomer.org/pkin/soft.html> for a more complete and up-to-date list.

1. ADAPT II is supplied as FORTRAN code for VAX VMS, MS DOS, and SUN UNIX systems. It performs simulations, fitting, and optimal sampling and includes extended least-squares and Bayesian optimization. Models can be expressed as integrated or differential equations using FORTRAN statements.

2. BOOMER is supplied as a compiled program for Macintosh and MS DOS systems. It performs simulation and fitting and includes Bayesian optimization. Models, integrated or differential equations, are expressed as a sequence of parameters.
3. IDENT2 and IDENT3 are provided as FORTRAN source code (IDENT2C is provided in C) for VAX VMS and other systems. It performs identifiability analysis.
4. MULTI programs by K. Yamaoka and colleagues (44) are provided as BASIC source code within the manuscripts. Different versions include fitting to integrated or differential equations, Bayesian analysis, and extended least squares.
5. NONMEM is provided as FORTRAN source code for UNIX, IBM, and other computers. The program performs nonlinear regression of individual or population data.
6. NPEM is a program for performing nonparametric expectation maximization.
7. SAAM II is provided as a compiled program for Windows or Macintosh systems and performs nonlinear regression analysis. It has a graphical user interface for model specification.
8. WinNonlin is provided as a compiled program for Windows systems and performs nonlinear regression analysis.

REFERENCES

1. Ko, H.C.; Jusko, W.J.; Ebling, W.F. Nonlinear Perpendicular Least-Squares Regression in Pharmacodynamics. *Biopharm. Drug Dis.* **1997**, *18*, 711–716.
2. Gibaldi, M.; Perrier, D. *Pharmacokinetics* 2nd Ed.; Marcel Dekker, Inc.: New York, 1982.
3. Bourne, D.W.A.; Triggs, E.J.; Eadie, M.J. *Pharmacokinetics for the Non-Mathematical*; MTP Press Ltd: Lancaster, UK, 1986.
4. Wagner, J.G. *Biopharmaceutics and Relevant Pharmacokinetics*; Drug Intelligence Publications: Hamilton, IL, 1971.
5. Wagner, J.G. *Pharmacokinetics for the Pharmaceutical Scientist*; Technomic Publishing: Lancaster, PA, 1993.
6. Bischoff, K.B. Physiological Pharmacokinetics. *Bull. Math. Biol.* **1986**, *48*, 309–322.
7. Sheiner, L.B.; Stanski, D.R.; Vozeh, S.; Miller, R.D.; Ham, J. Simultaneous Modeling of Pharmacokinetics and Pharmacodynamics: Application to *d*-tubocurarine. *Clin. Pharmacol. Ther.* **1979**, *2*, 358–371.
8. Jusko, W.J. Conceptualization of Drug Distribution to a Hypothetical Pharmacodynamic Effect Compartment. *Clin. Pharmacol. Ther.* **1993**, *54*, 112–113.
9. Sharma, A.; Jusko, W.J. Characterization of Four Basic Models of Indirect Pharmacodynamic Responses. *J. Pharmacokin. Biopharm.* **1996**, *24*, 611–635.

10. Sharma, A.; Ebling, W.F.; Jusko, W.J. Precursor-Dependent Indirect Pharmacodynamic Response Model for Tolerance and Rebound Phenomena. *J. Pharm. Sci.* **1998**, *87*, 1577–1584.
11. Carstensen, J.T. *Drug Stability: Principles and Practice*; Marcel Dekker, Inc.: New York, 1990.
12. Wagner, J.G. *Fundamentals of Clinical Pharmacokinetics*; Drug Intelligence Publications: Hamilton, IL, 1975; 35.
13. Gerald, C.P.; Wheatley, P.O. *Applied Numerical Analysis*, 4th Eds.; Addison-Wesley: Reading, MA, 1990.
14. Boas, M.L. *Mathematical Methods in the Physical Sciences*, 2nd Ed.; Wiley: New York, 1983.
15. Mayersohn, M.; Gibaldi, M. Mathematical Methods in Pharmacokinetics. I. Use of the Laplace Transform for Solving Differential Rate Equations. *Am. J. Pharm. Ed.* **1970**, *34*, 608–614.
16. Mayersohn, M.; Gibaldi, M. Mathematical Methods in Pharmacokinetics. II. Solution of the Two Compartment Open Model. *Am. J. Pharm. Ed.* **1971**, *35*, 19–28.
17. Benet, L.Z.; Turi, J.S. Use of General Partial Fraction Theorem for Obtaining Inverse Laplace Transforms in Pharmacokinetic Analysis. *J. Pharm. Sci.* **1971**, *60*, 1593–1594.
18. Benet, L.Z. General Treatment of Linear Mammillary Models with Elimination from Any Compartment as Used in Pharmacokinetics. *J. Pharm. Sci.* **1972**, *61*, 536–541.
19. Yano, Y.; Yamaoka, K.; Tanaka, H. A Nonlinear Least Squares Program, MULTI(FILT), Based on Fast Inverse Laplace Transform for Microcomputers. *Chem. Pharm. Bull.* **1989**, *37*, 1035–1038.
20. Fehlberg, E. *Low-Order Classical Runge-Kutta Formulas with Stepsize Control and Their Application to Some Heat Transfer Problems*; NASA TR R-315; NASA Tech. Report: 1969.
21. Watt, H.A.; Shampine, L.F. Subroutine RKF45. *Computer Methods for Mathematical Computation*; Forsythe, G.E., Malcolm, M.A., Moler, C.B., Eds.; Prentice-Hall: Englewood, NJ, 1977, 133–147.
22. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes in FORTRAN*; 2nd Ed. Cambridge University Press: Cambridge, UK, 1992; 741.
23. Gear, C.W. DIFSUB for Solution of Ordinary Differential Equations. Algorithm 407. Collected Algorithms. Communications of the Association for Computer Machinery (CACM): New York, 1971.
24. Gear, C.W. The Automatic Integration of Ordinary Differential Equations, Communications of the Association for Computer Machinery (CACM): New York, 1971; 14, 176–179.
25. Sedman, A.J.; Wagner, J.G. CSTRIP, A FORTRAN IV Computer Program for Obtaining Initial Polyexponential Parameter Estimates. *J. Pharm. Sci.* **1976**, *65*, 1006–1010.
26. Brown, R.D.; Manno, J.E. ESTRIP, A BASIC Computer Program for Obtaining Initial Polyexponential Parameter Estimates. *J. Pharm. Sci.* **1978**, *67*, 1687–1691.
27. Wagner, J.G.; Nelson, E. Kinetic Analysis of Blood Levels and Urinary Excretion in the Absorptive Phase after Single Doses of Drug. *J. Pharm. Sci.* **1964**, *53*, 1392–1403.
28. Loo, J.C.K.; Riegelman, S. New Method for Calculating the Intrinsic Absorption Rate of Drugs. *J. Pharm. Sci.* **1968**, *57*, 918–928.
29. Kiwada, H.; Morita, K.; Hayashi, M.; Awazu, S.; Hanano, M. A New Numerical Calculation Method for Deconvolution in Linear Compartmental Analysis of Pharmacokinetics. *Chem. Pharm. Bull.* **1977**, *25*, 1312–1318.
30. Vaughan, D.P.; Dennis, M. Mathematical Basis of Point-Area Deconvolution Method for Determining In Vivo Input Functions. *J. Pharm. Sci.* **1978**, *67*, 663–665.
31. Cutler, D.J. Numerical Deconvolution by Least Squares: Use of Prescribed Input Functions. *J. Pharmacokinet. Biopharm.* **1978**, *6*, 227–241.
32. Hartley, H.O. The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares. *Technometrics* **1961**, *3*, 269–280.
33. Marquardt, D.W. An Algorithm for Least-Squares Estimation of Non-Linear Parameters. *J. Soc. Indust. Appl. Math.* **1963**, *11*, 431–441.
34. Nelder, J.A.; Mead, R. A Simplex Method for Function Optimization. *Comput. J.* **1965**, *7*, 308–313.
35. Wagner, J.G. *Fundamentals of Clinical Pharmacokinetics*; Drug Intelligence Publications: Hamilton IL, 1975, 289.
36. Peck, C.C. Computer-Assisted Clinical Pharmacokinetics. *Pharmacokinetic Basis for Drug Treatment*; Benet, L.Z., Massoud, N., Gambertoglio, J.G. Eds.; Raven Press: New York, 1984, 353–354.
37. Peck, C.C.; Sheiner, L.B.; Nichols, A.I. The Problem of Choosing Weights in Nonlinear Regression Analysis of Pharmacokinetic Data. *Drug Metab. Rev.* **1984**, *15*, 133–148.
38. Peck, C.C.; Beal, S.L.; Sheiner, L.B.; Nichols, A.I. Extended Least Squares Nonlinear Regression: A Possible Solution to the Choice of Weights Problem in Analysis of Individual Pharmacokinetic Data. *J. Pharmacokinet. Biopharm.* **1984**, *12*, 545–558.
39. Schumitzky, A. Nonparametric EM Algorithms for Estimating Prior Distributions. *App. Math. Comput.* **1991**, *45*, 143–157.
40. Draper, N.R.; Smith, H. *Applied Regression*; Wiley: New York, 1966.
41. Mandel, J. *The Statistical Analysis of Experimental Data*; Interscience: New York, 1964; 164.
42. Boxenbaum, H.G.; Riegelman, S.; Elashoff, R.M. Statistical Estimation in Pharmacokinetics. *J. Pharmacokinet. Biopharm.* **1974**, *2*, 123–148.
43. Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Control.* **1973**, *19*, 716–723.
44. Yamaoka, K.; Nakagawa, T.; Uno, T. Application of Akaike's Information Criterion (AIC) in the Evaluation of Linear Pharmacokinetic Equations. *J. Pharmacokinet. Biopharm.* **1978**, *6*, 165–175.
45. Brown, R.F.; Godfrey, K.R. Problems of Determinacy in Compartmental Modeling with Application to Bilirubin Kinetics. *Math. Biosci.* **1978**, *40*, 205–224.
46. Godfrey, K.R.; Fitch, W.R. The Deterministic Identifiability of Nonlinear Pharmacokinetic Models. *J. Pharmacokinet. Biopharm.* **1984**, *12*, 177–191.
47. Wang, Y.-M.C.; Reuning, R.H. An Experimental Design Strategy for Quantitating Complex Experimental Models: Enterohepatic Circulation with Time-Varying Gallbladder Emptying as an Example. *Pharm. Res.* **1992**, *9*, 169–177.
48. Jacquez, J.A. Identifiability: The First Step in Parameter Estimation. *Fed. Proc.* **1987**, *46*, 2477–2480.

49. Jacquez, J.A.; Perry, T. Parameter Estimation: Local Identifiability of Parameters. *Am. J. Physiol.* **1990**, *258*, E727–E736.
50. See <http://www.boomer.org/pkin/book.html> for a more complete and up-to-date list.
51. Bourne, D.W.A. *Mathematical Modeling of Pharmacokinetic Data*; Technomic Publishing: Lancaster, PA, 1995.
52. Carstensen, J.T. *Modeling and Data Treatment in the Pharmaceutical Sciences*; Technomic Publishing: Lancaster, PA, 1996.
53. Draper, N.R.; Smith, H. *Applied Regression Analysis*; Wiley: New York, 1966.
54. Gerald, C.F.; Wheatley, P.O. *Applied Numerical Analysis*, 4th Ed.; Addison-Wesley: Reading, MA, 1990.
55. Gibaldi, M.; Perrier, D. *Pharmacokinetics*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1982.
56. Mandell, J. *The Statistical Analysis of Experimental Data*; Interscience: New York, 1964.
57. Sadler, D.R. *Numerical Methods for Nonlinear Regression*; Queensland Press: St. Lucia, Australia, 1975.
58. Wagner, J.G. *Biopharmaceutics and Relevant Pharmacokinetics*; Drug Intelligence Publications: Hamilton, IL, 1971.
59. Wagner, J.G. *Pharmacokinetics for the Pharmaceutical Scientist*; Technomic Publishing: Lancaster, PA, 1993.



Materials of Construction for Pharmaceutical Equipment

Michelle M. González

Amgen Inc., Thousand Oaks, California, U.S.A.

INTRODUCTION

Materials required in the fabrication of equipment and associated systems dedicated to the production of pharmaceutical products should be capable of withstanding the temperature, pressure, and chemical corrosiveness of the product to ensure its purity and integrity. They also must meet a high degree of cleanliness to convey sterile and nonsterile products or solutions, particularly when it applies to their contact surfaces. Contact surfaces are “all surfaces exposed to the product or from which liquids may drain, drop, or be drawn into the product,” and solution contact surfaces are “the interior surfaces of the circuit used exclusively for supply and recirculation of cleaning and/or sanitizing solutions.”^[1]

For many years, pharmaceutical equipment manufacturing has relied predominantly in stainless steel, particularly the 316L type to meet the sanitary requirements of equipment and interconnecting systems such as piping and/or tubing, and in-line monitoring sensors, etc. However, with increased frequency, plastic materials are being utilized to meet the same requirements. Among the most notable materials in this group are the fluorinated plastics including polyvinylidene fluoride (PVDF), polytetrafluoroethylene (PTFE), or perfluoroalkoxy (PFA) resin,^[2] and thermoplastics such as polypropylene (PP). Additionally, elastomers such as ethylene-propylene-diene rubber (EPDM), “steam resistant” fluororelastomers (FKM), platinum-cured silicon (pt-Si), and Kalrez[®] perfluoroelastomers (FFKM) parts using compounds KLR-6221 and KLR-6230^[3] have also played a very significant role by providing sealing and resilient applications that meet the requirements of Code of Federal Regulations (CFR) Title 21, Foods and Drugs current Good Manufacturing Practices (cGMPs).

The objectives of this article are to offer a comparison between the technologies involved in the production of those materials and their current uses, particularly in process and support systems, and to compare their microbiological impact in the manufacturing of pharmaceuticals.

STAINLESS STEEL

Selection of Stainless Steels

Stainless steels are uniquely qualified for bioprocessing applications not only because of their long service life, availability and fabricability, but also because they are noncorroding, noncontaminant, can be polished to very smooth finishes, are strong and rigid, can withstand heat and chemical sterilization treatments, and are easily welded. With more than 70 standard types of stainless steel produced, the industry's workhorse is the austenitic 18–8 (18% chromium, 8% nickel) group which includes types 304 and 316 and their L or low carbon content variations. With concerns for higher corrosion resistance, the new and better low carbon superaustenitic stainless steel AL-6XN (6% molybdenum) and the nickel-based alloys, Hastelloy B and Hastelloy C, are becoming notoriously important in the fabrication of vessels, piping, tubing, and fittings (see Table 1). Finally, the cast stainless alloys such as CF-8 (similar to 304), CF-8M (similar to 316), and CF-3M (similar to 316L) utilized in pumps, various types of valves (particularly ball type), and fittings also occupy a prominent position in the industry.

Standards and Codes

In some pharmaceutical industry circles, there is the belief that the stringent specifications that rule pharmaceutical processes have no comparison with the food and dairy industries.^[4] Actually, almost the reverse is true; the present standards and specifications for tubing fabrication and their product/solution contact surfaces, as well as the installation of systems, have their origin in the dairy industry through the well-recognized 3-A Sanitary Standards and Accepted Practices (Numbers 33-00, and 605-04). Published by the International Association for Food Protection (IAFP), formerly known as the International Association of Milk, Food and Environmental Sanitarians (IAMFES), these standards and practices cover: first, the sanitary aspects of polished metal tubing used to conduct dairy products in processing lines or systems that also may include sanitary fittings, and secondly, the cleaning of rigid solution lines and the clean-in-place (CIP) units that circulate the preinse, rinse,

**Table 1** Chemistry comparison

	%	304	304L	316	316L	AL-6XN
Cr	Chromium	18.0–20.0	18.0–20.0	16.0–18.0	16.0–18.0	20.0–22.0
Ni	Nickel	8.0–11.0	8.0–13.0	10.0–14.0 ^a	10.0–15.0	23.5–25.5
C	Carbon	0.08 max	0.035 max	0.08 max	0.035 max	0.03 max
Fe	Iron	Balance	Balance	Balance	Balance	Balance
Mo	Molybdenum	—	2.0–3.0	—	2.0–3.0	6.0–7.0
Mn	Manganese	2.0 max	2.0 max	2.0 max	2.0 max	2.0 max
Si	Silicon	0.75 max	0.75 max	0.75 max ^b	0.75 max ^b	1.0 max
P	Phosphorus	0.040 max	0.040 max	0.040 max	0.040 max	0.040 max
S	Sulfur	0.030 max	0.030 max	0.030 max	0.005–0.017 ^c	0.030 max
N	Nitrogen	—	—	—	—	0.18–0.25
Cu	Copper	—	—	—	—	0.75 max

^a 11.0–14.0 (A269).^b 0.030 (A269).^c Sulfur has greatest effect on weld quality. Controlling sulfur facilitates orbital field welds by minimizing stabilization problems. To comply with ASME BPE Standard, ASTM added A 270-98a “Supplementary Requirements,” S2. Pharmaceutical Quality Tubing—S2.1.1.

cleaning solutions, and postrinse liquids used for cleaning and sanitizing the product pipelines and process equipment.

As the pharmaceutical industry advanced, the development of new technologies to make or modify products by microbial and biochemical processes using living organisms or substances from those organisms (Biotechnology) increased the need for more stringent regulations. Many of these regulations were related to equipment cleanability and sterility, standardization of manufacturing methods, and integration of standards covering vessels, appurtenances, and equipment necessary for the industry. The American Association of Mechanical Engineers (ASME), through its Council on Codes and Standards (CSS) and a directive to the Board on Pressure Technology, initiated the creation of what is now known as the Bioprocessing Equipment (BPE-2002/BPE-a2000), an American National Standard. This Standard directly (or by reference) covers requirements for materials, design, fabrications, examinations, inspections, testing, certifications (for pressure systems), and pressure relief (for pressure systems) of vessels and piping for bioprocessing systems, including design for sterility and cleanability (Part SD), dimensions and tolerances for automatic welding and hygienic clamp tube fittings (Part DT), material joining (Part MJ), interior surface finishes (Part SF), and equipment seals (Part SG). Requirements of this Standard apply to all parts that contact the product, raw materials, or product intermediates during manufacturing, development, or scale-up, and all equipment or systems that are a critical part of product manufacturing, such as water-for-injection (WFI), clean/pure steam, ultrafiltration, and intermediate product storage.

While ASME BPE-2002 addresses requirements applicable to the design of equipment and components, other codes and standards define specific requirements for the manufacturing of components critical to the processes and support utilities, particularly piping, tubing, and fittings. The most notable are: American Society for Testing and Materials (ASTM) A269 “Standard Specification for Seamless and Welded Austenitic Stainless Steel Tubing for General Service,” ASTM A270 “Standard Specification for Seamless and Welded Austenitic Stainless Steel Sanitary Tubing,” and ASTM A312/ASME SA312 “Standard Specification for Seamless and Welded Austenitic Stainless Steel Pipe.”

Product and/or Solution Contact Surface Finishes

Surface finishes are all interior surface finishes accessible and inaccessible, that directly or indirectly come in contact with the designated product in BPE and distribution system components. Surface finishes have been qualified and quantified utilizing various names and measurement units such as Grit Numbers, USA Finish Numbers, Common Name, Ra ($\mu\text{in.}$), Ra (μ), Rmax (μ), Rmax (μ), RMS, ISO number, etc.^[5] Each of these roughness parameters have met their purpose but this variety of systems has also created broad and sometimes overlapping measurement ranges resulting in a high degree of confusion (see Table 2).

The need for a universally recognized and accepted surface roughness specification and measurement standard has become necessary and a final criterion has been determined by the standard Ra value. This standard known as Arithmetic Average Roughness, is defined by

Table 2 Surface measurements comparison

RMS ($\mu\text{in.}$)	RMS (μ)	Ra ($\mu\text{in.}$)	Ra (μ)	Grit size
80	2.03	71	1.80	80
58	1.47	52	1.32	120
47	1.20	42	1.06	150
34	0.86	30	0.76	180
17	0.43	15	0.38	240
14	0.36	12	0.30	320
		4–8		320 EP
		10		400

(From DCI, Inc., Stainless Steel Processing Equipment. Material/Weld FINISHES, pamphlet.)

ANSI/ASME B46.1 “Surface Texture—Surface Roughness, Waviness and Lay” as “the arithmetic average height of roughness component irregularities from the mean line measured within the sample length (L)” (see Fig. 1). The derivation Ra is usually expressed in $\mu\text{in.}$ and measured with profilometers and/or borescopes.

Electropolished (EP) Finishes

Because the pharmaceutical industry requires smoother product/solution contact surfaces to achieve a purer product with less danger of bacteria growing in surface defects or cross-contamination of one product with another,^[6] in addition to purely mechanical finishes, sanitary tubing and components are also available in highly polished surfaces.

These surfaces are achieved by electropolishing also known as “chemical machining” and/or “reverse plating,” an electrochemical process far superior to any available mechanical process for the removal of surface imperfections in stainless steels. Electropolishing levels and brightens the material surface by anodic dissolution in flowing mixed acid solutions, sometimes with organic additives, (electrolyte) and a cathode that is pulled through the inside of the tube or component (anode). When the proper combination of electrolyte current temperature is attained, the high points of surface irregularities (or high current density areas) are selectively removed at a greater

rate than the remainder of the surface resulting in improved surface measurements (see Figs. 2 and 3).

In addition to appearance, EP products offer other advantages such as an extremely smooth surface which minimizes the adherence of debris, an increased chromium to iron ratio which improves corrosion resistance, the creation of a passive layer that is free from iron contamination, improved ability to visually detect surface defects, and improved mechanical property performance through the minimization of stress risers.

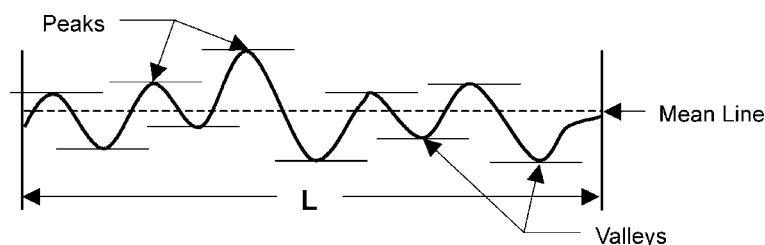
Joining Techniques

Requirements for the joining of BPE including vessels and tanks [built to ASME Boiler and Pressure Vessel Code (BPVC), Section VIII, Division 1], pumps, piping (built to ASME B31.3), tubing, and fittings, are described in Part MJ of the ASME BPE-2002 and are limited to process systems that contact bioprocessing products or product-process streams. Welding procedures for pressure vessels, tanks, piping, and hygienic tubing systems shall be qualified in accordance with ASME BPVC, Section IX.

Because tubing systems occupy a most notorious place in the construction of bioprocessing facilities, focus has been placed upon the prevalent method for joining these systems. Connections between tube and tube or tube and fitting and even tube/fitting to equipment can be accomplished by diverse means. However, the use of mechanical joints where impurities entrainment may occur, such as flanges, threaded joints, or even hygienic clamp type joints must be minimized. Thus, systems shall preferentially be joined using square butt-welding practices only because they present the lowest risk of contamination.

Automatic Orbital Welding

Typically, square butt-welded tubing joints are made using automatic orbital welding equipment which fuses the thin wall tubes and fittings together in a totally inert environment without the use of filler materials (auto-genous weld).^[7] Essentially, an arc established between a tungsten electrode (installed in a rotor within the weld

**Fig. 1** Pictorial display of surface texture.

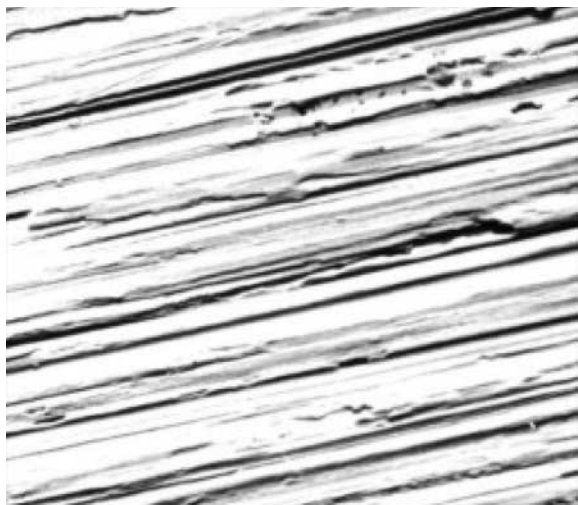


Fig. 2 Before electropolishing the mechanically polished surface of stainless steel appears rough as viewed under $1000\times$ magnification on Scanning Electron Microscope.

head) and the tubing accomplishes the fusion weld. This process consists of a series of spot welds in which the main welding current penetrates the material and the back-ground current chills the puddle. The quality of the fusion joint that is made by this equipment is predicated on the use of two pieces of material of the same thickness and grade or type; therefore, great care must be exercised in material and component selection.

The actual welding process requires the tube/tube or tube/fitting to be placed in a welding head (see Fig. 4) provided with clamping jaws to securely hold the parts in

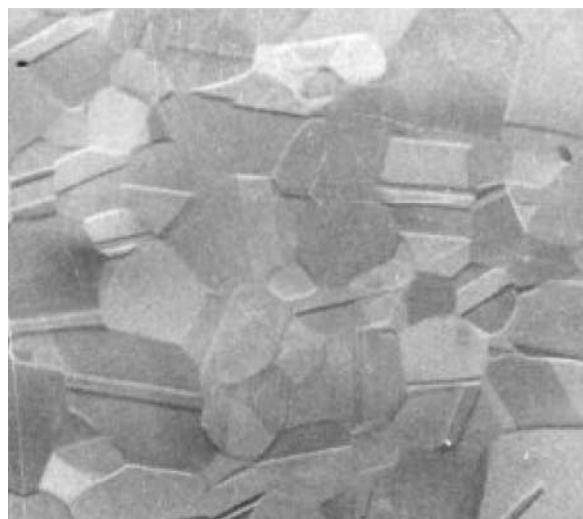


Fig. 3 After electropolishing the same surface appears smooth as viewed under $1000\times$ magnification on Scanning Electron Microscope.



Fig. 4 Automatic orbital welder in the process of joining two long tangent stainless steel fittings.

position and in alignment with the tungsten electrode contained within the head. The entire area to be welded is then enclosed in the welding head, forming a chamber that is filled with shielding gas, usually argon, to prevent oxidation of the welded material during the weld sequence. Meanwhile, the inside of the tube is purged free of oxygen allowing the entire area to be completely covered with argon gas. The weld cycle is preprogrammed and set in the machine, making the entire operation completely automatic. The electrode orbits around the weld seam on an internal gear while the head remains stationary and when the rotation is completed, the head can be opened and immediately removed from the welded section.

Stainless Steel Passivation

Passivation is a postweld treatment and/or cleaning process used to restore (by introduction of oxygen) the thin chromium oxide (with some oxides of iron and nickel) film or passive layer that forms on stainless steel naturally when in contact with air. Methods and tests for cleaning and passivation of critical water, product, and process systems are described in ASTM A380 "Standard Practice for Cleaning and Descaling Stainless Steel Parts, Equipment, and Systems."

Passivation removes/dissolves free iron and other anodic contaminants from the surface of corrosion-resistant parts leaving chromium and iron oxides as the primary metal components. Because welding disturbs



the passive layer by reducing the chromium and increasing the iron, thus altering the chrome/iron ratio (measure of corrosion resistance), upon completion and approval of the weld, the weld surface and adjacent boundary area must be brought back to a passive state. Also, certain normal operating conditions in typical WFI, reverse osmosis (RO), deionized water (DI), clean/pure steam, some process systems, and in some rare cases, CIP systems often lead to formation of the most prevalent form of self-catalyzing corrosion called “rouge” which is a colloidal form of iron oxide containing chromium and nickel in various forms. This problem may be accentuated by the use of high temperature, aggressive process chemicals, ultrapure water, and can be overcome only by restoring the surface to its passive state.

Some of the most recognized tests to detect, measure, and quantitate the chrome/iron ratio and to ensure that a passive layer has been established are the Ferroxyl Test for Free Iron, X-Ray Photoelectron Spectroscopy (XPS) or Electron Spectroscopy for Chemical Analysis (ESCA), and Auger Electron Spectroscopy (AES).

PLASTICS

ASTM defines a plastic as “a material that contains as an essential ingredient one or more organic polymeric substances of large molecular weight, is solid in its finished state, and, at some stage in its manufacture into finished articles, can be shaped by flow.”

As the pharmaceutical and biotechnology industries confront challenges such as increased competition, consolidation, globalization, high R&D costs, and demanding manufacturing standards and guidelines, the need for the use of alternative materials of construction becomes a more significant issue. Despite of the many advantages of stainless steel, the wide availability of thermoplastics and fluorinated plastics offering lower initial costs, less weight, complete resistance to corrosion, elimination of the passivation process, and extremely smooth internal surfaces have increased their application, particularly in systems such as purified water (PW) distribution loops and other processes.

Fluoropolymers or Fluorinated Plastics

Fluorinated plastics are thermoplastic paraffinic polymers where the hydrogen has been replaced by fluorine and, in some cases, chlorine.

PVDF—Approval FDA 21 CFR 177.2510

PVDF is a strong, abrasion resistant, and wide temperature service range (−40–284°F) fluorocarbon material. Similar

to PTFE, with the exception of not being fully fluorinated, it is a relatively inert material resistant to permeation of gases contributing little in the way of contamination to pharmaceutical water. It is chemically resistant to most acids, bases, and organic solvents and is ideally suited for handling wet or dry chlorine, bromine, and other halogens, but is unsuitable for handling strong alkalis, fuming acids, polar solvents, amines, ketones, and esters. It has a high tensile strength as well as a high heat deflection temperature. PVDF offers an extremely smooth surface finish (between 6 Ra and 8 Ra) and can be readily joined by the fusion process (free of internal beads or crevices), threading, or flanging.

The focal point of a sanitary system is its water, which has to be pure. A useful definition of high purity water is “water that has been purified to a degree not found in nature.” High purity water has to meet the drinking water standards of the U.S. Environmental Protection Agency (EPA) for feedwater of the compendial waters (those used as an ingredient in pharmaceutical products), United States Pharmacopeia (USP) PW, and USP WFI. Because of its high purity, low surface and joint extractables, and elevated temperature cleanability, natural unpigmented PVDF has become the most important alternative material for use in PW distribution systems.

PVDF, however, has been used primarily in water systems that operate at ambient temperatures. There is some skepticism in the use of PVDF in hot (176°F) WFI distribution loops, mainly because it has been observed that exposure to heat will discolor the material. The discoloration is the result of alterations to the chemical structure of PVDF on the atomic scale. Testing of piping samples revealed that leaching was insignificant and welds remained strong after 1 to 16 week exposures to hot ultrapure water.^[8]

PTFE—Approval FDA 21 CFR 177.1550

Introduced by DuPont as Teflon® in 1950, PTFE is a fluoroplastic resistant to practically every known chemical or solvent and possesses the highest useful temperature limit (−200–500°F) of commercially available plastics, has an exceptionally low coefficient of friction, high impact strength, and low surface energy. Usual processing techniques like injection molding are not possible with PTFE; it must be compression-molded into block form and then machined into a finished product. PTFE resin is pressed into shapes under high pressure at room temperature and then heated to 700°F to complete the molding (sintering process) and adjust the crystalline content.

Because of its elastic recovery properties, PTFE, a self-lubricating compound, is used in sealing and flexing applications such as sanitary seals, valve diaphragms,



o-rings in mechanical seals, and seals and seats of stem and rotary valves such as ball valves. It is also commonly used as a liner or coating material for valves and pumps for the pharmaceutical industry.

PFA resin—Approval FDA 21 CFR 177.1550

Of the melt-processible fluoropolymers, which are the most suitable for tubing, PFA provides the extreme thermal and chemical resistance required in the pharmaceutical processing. PFA, however, does not have the physical strength of PTFE at elevated temperatures and must be reinforced or designed with thickness to compensate for its softness. Tests have shown that among materials such as stainless steel, glass, silicon-coated glass, PP, and PVDF, even when proteins and micro-organisms were added to more efficiently wet test surfaces, PFA was the least affected by contaminants and biofilm growth.^[9] PFA has been proven in CIP–SIP operations, ultrapure water, WFI, and pharmaceutical applications.

Thermoplastics

Thermoplastics have a defined melting point; they can be melted, cooled, and remelted without destroying the physical or mechanical properties of the polymer.

PP—Approval FDA 21 CFR 177.1520

PP, a polyolefin polymer, is probably one of the most economical and widely used thermoplastics. A crystalline polymer with heat deflection temperature ranging from 195 to 240°F (higher than other common plastics) offers as its key properties, a high heat resistance (for piping, an upper limit of 212°F), a specific gravity of 0.91 if unmodified (the lightest of the most common thermoplastics), stiffness, and chemical resistance to caustics, solvents, acids, and other organic chemicals. Its use is not recommended with strong oxidizing type acids, detergents, low boiling hydrocarbons, alcohols, and some chlorinated organic materials. Because PP is a relatively inert material and contributes little in the way of contamination to pharmaceutical water, it is highly suitable for piping systems. PP also has proved to be an excellent material for laboratory and industrial drainage where mixtures of acids, bases, and solvents are involved. PP is available in schedules 40 or 80 for pressure service and can be joined by socket solvent welding, thermo-seal fusion process, threading (Sch. 80 only), or flanging. The plastic pipe industry recognizes that PP compounds containing more than 2.5% carbon black can have an outside service life of more than 25 yr.

Sealing Materials

Seals are those elements that create or maintain process boundaries between system components and/or subassemblies in order to ensure system integrity in validated process and utility systems. Seals must be biocompatible (able to be in contact with bacteria or mammalian cells without interfering with their metabolism or ability to live and multiply), must be corrosion and permeation resistant, their surface finishes must be free of molding imperfections and foreign matter on surfaces within the sealing area, and shall not generate particulate that may entrain the product.

Elastomers (materials that can be stretched or compressed repeatedly and, upon immediate release of stress, will return to their approximate original size) are long chain copolymers or terpolymers (two or three different monomers in one chain) that contain adequate cross-links among individual chains. Fluorinated elastomers are more stable than hydrocarbon or silicon elastomers because carbon–fluorine bonds are approximately 30% stronger than carbon–hydrogen bonds. There are five major FDA (21 CFR 177.2600) compliant elastomers used in the pharmaceutical industries: EPDM, FKM, platinum-cured silicon, and FFKM KLR-6221 and KLR-6230. Of the compounds used for medical seals, FFKM have the broadest range of chemical resistance of any other elastomeric material and are readily formable into many standard and custom seal configurations.^[10]

CONCLUSION

It is very clear to see that the biotechnology industry demands special care and attention in the selection of materials for the product contact surfaces as well as the solution contact surfaces. The emphasis has been placed principally in the cleanliness and corrosion resistance issues. To these purposes, a whole segment of the industry dedicates considerable time and effort in R&D for new materials and applications. Awareness and adherence to present codes and regulations as well as knowledge of new and advanced technologies are the keys to successful completion of design and engineering of systems utilizing these very valuable resources.

REFERENCES

1. International Association for Food Protection (IAFP), 3-A Sanitary Standards.
2. Wulf, B. Pristine Processing—Designing Sanitary Systems. *Chem. Eng.* **1996**, (November), 76.



Materials of Construction for Pharmaceutical Equipment

7

3. Duzick, T.C. The Impact of Elastomer Performance on Pharmaceutical Manufacturing Processes. *Pharm. Eng.* **2000**, (2), 22–34.
4. Kaupish, K.; Baker, E.; Sander, B.; Harmon, C. GMP Environments: Applications of Food and Dairy Industries Standards in Pharmaceuticals. *Pharm. Technol.* **1991**, (October), 12–14.
5. Tabenkin, A. To Each His Own Parameter? *Qual. Digest* **2001**, (June), 37–39.
6. Tverberg, J. Ensuring Quality in Electropolished Stainless Steel Tubing. *Tube Pipe J.* **1998**, (September/October).
7. Henon, B.K., Ph.D. Orbital Welding in Compliance with the New ASME Bioprocessing Equipment (BPE) 1997 Standard. *Pharm. Eng.* **1999**, (January/February), 44–58.
8. Burkhart, M.; Wermelinger, J.; Klaiber, F. Is Polyvinylidene Fluoride Piping Safe for Hot-Ultrapure-Water Applications? *MICRO* **1995**, (February), 27–31.
9. Fleming, J.R.; Kemkes, D.P. Using Ultrapure Plastic Tubing. *Chem. Eng.* **2000**, (February), 111–114.
10. <http://www.devicelink.com/mddi/archive/00/12/002> (accessed Dec 2001).

MANAGEMENT OF DRUG DEVELOPMENT

James E. Tingstad

Tingstad Associates, Green Valley, Arizona, U.S.A.

The most successful new drug-development programs require competent, caring, people-oriented leaders at all levels. Overwhelming social science data show that this approach will optimize productivity, efficiency, and creativity, while fostering employee growth, enthusiasm, cooperation, and loyalty.

THE DRUG DEVELOPMENT PROCESS

Before addressing the components of good management, here is a review of the basic new drug (new chemical entity or NCE) development process.

Brief Overview

The NCE works its way through the following groups before it can be marketed:

Synthetic Chemistry → Pharmacology → Toxicology → Pathology →
Regulatory Affairs [Investigative New Drug (IND) application] →
Product R&D → Clinical Research → Regulatory Affairs [New Drug Application (NDA)] → Approval by government regulatory agency

Additional Important Contributors

Other involved departments/disciplines that are equally important as the above groups include Analytical Chemistry, Biochemistry, Biopharmaceutics/Pharmacokinetics/Drug Metabolism, Chemical Pilot Plant, Experimental Engineering, Packaging Development, and Statistics. Additional significant contributors include Purchasing and Quality Assurance/Documentation.

Post-NDA Departments

After NDA approval, non-R&D functions—after being partially involved at various stages of NCE development—take over, including Chemical Manufacturing,

Engineering, Marketing, Packaging, Pharmaceutical Manufacturing, Quality Control, and Sales.

Departmental Responsibilities

Companies are organized differently; for example, the various engineering responsibilities may be in one or more departments. A brief description of typical functions follows:

- Synthetic chemistry synthesizes NCE candidates for pharmacological testing.
- Pharmacology examines the *in vivo* activity of the NCE in animals (1).
 1. For economic reasons, activity is determined in animals before safety, whereas in humans, preliminary safety studies need to come first.
 - Toxicology determines the “macro” negative effects of the NCE in animals.
 - Pathology examines the “micro” negative effects of the NCE in animals.
 - Product R&D designs a simple, preliminary dosage form for initial clinical trials (2).
 2. Additional dosage-form development is deferred until safety/activity experiments in humans show promise. Eventually, scale-up experiments are conducted in cooperation with Pharmaceutical Manufacturing.
- Clinical research determines the safety and efficacy of the NCE in humans.
 - Phase I examines small-dose tolerance and safety in a limited number of young adult, usually male, volunteers.
 - Phase II, involving hundreds of patients, investigates efficacy, dosage, and prominent side effects.
 - Phase III, utilizing thousands of patients, broadens Phase II experiments and determines safety, efficacy, and marketability.
 - Phase IV, conducted after regulatory approval and marketing, investigates additional medical uses.

- Regulatory Affairs works closely with all relevant organizational units and is the primary contact with government regulatory agencies.
- Analytical Chemistry develops stability-indicating assays for NCEs and identifies/quantifies impurities; cooperates with Product R&D on product stability studies.
- Biochemistry determines, among other experiments, the cell-level effects of NCEs.
- Biopharmaceutics, Pharmacokinetics, and Drug Metabolism examines the absorption, distribution, metabolism, and elimination of NCEs.
- Chemical Pilot Plant produces NCEs for R&D groups.
- Experimental Engineering helps design new processes and equipment.
- Packaging Development develops containers for new products with stability and consumer issues in mind.
- Statistics is involved in planning and interpreting many R&D experiments.
- Purchasing works with R&D to ensure consistent, high-quality raw materials from vendors.
- Quality Assurance/Documentation monitors procedures and records to comply with government-regulated Good Laboratory and Manufacturing Practices.
- Chemical Manufacturing supplies the bulk NCE to Pharmaceutical Manufacturing.
- Engineering offers process/equipment services to post-NDA groups.
- Marketing, in cooperation with Sales and Advertising, determines the overall strategy for supplying NCE products to primary customers (physicians and other healthcare professionals).
- Packaging packages and labels manufactured drug products for sale.
- Pharmaceutical Manufacturing produces finished drug products for consumer use.
- Quality Control monitors the quality and stability of manufactured/marketed drug lots.
- Sales supplies products to customers.

Departmental Interactions

It's important for managers and laboratory workers to meet with colleagues from other departments to learn about interrelations and interdependencies and to use that knowledge to ensure mutual understanding, respect, support, and cooperation. Here are examples:

Sales \longleftrightarrow All R&D groups

R&D managers and laboratory workers should be encouraged to spend a day or two (one-on-one) with a salesperson "on the road," for two primary reasons:

1. R&D individuals can observe first hand the results of their labors.
2. Salespersons can learn important, relevant scientific facts about what they are selling and can directly inform the R&D person about problems their customers experience.

Synthetic chemistry \longleftrightarrow Biopharmaceutics

Oral absorption data can help lead synthetic chemists in the most promising directions for NCE variations.

Synthetic chemistry \longleftrightarrow Product R&D

Formulators usually prefer the most water-soluble form of an NCE; this can conflict with Synthetic Chemistry and Chemical Manufacturing's interest in high yields.

Biopharmaceutics \longleftrightarrow Pharmacology

Interactional benefits go both ways, but Biopharmaceutics can help pharmacologists determine whether lackluster potency is attributable to inherent inactivity or poor oral absorption.

Product R&D \longleftrightarrow Toxicology

True event: Product R&D helped Toxicology determine that the apparent intestinal irritation of an orally administered NCE was caused not by the drug, but by the "innocuous" solvent (glycerin).

Analytical chemistry \longleftrightarrow Quality control

Analytical Chemistry develops stability-indicating assays for R&D, then transfers them to Quality Control when the product is marketed. Too often these groups are at loggerheads concerning what is an adequate, efficient assay procedure for the NCE. Management's mutual respect and cooperation, plus voluntary temporary interdepartmental transfers, can usually minimize these difficulties.

Pharmaceutical manufacturing \longleftrightarrow Product R&D

1. These two groups need to interact at the appropriate stages in new drug development (especially scale-up) to ensure that a dependable, high-quality product can be consistently and economically manufactured.
2. Hands-on Manufacturing employees may feel most comfortable interacting with R&D scientists at the B.S./Associate degree level rather than at the Ph.D. level.

3. Product R&D laboratory workers at all educational levels should spend their first month or two in Pharmaceutical Manufacturing—working hands-on rather than just observing.

Marketing ↔ All R&D groups

True event: A Marketing executive was touring R&D laboratories and asked a key synthetic chemist working on antibacterials what she thought was the most important quality in an antibiotic after therapeutic activity. She replied, “lack of bacterial resistance.” When the executive emphasized that Marketing was most concerned about lack of side effects, the scientist said, “No one ever told me that.”

Quality assurance and documentation ↔ All R&D groups

Quality Assurance and Documentation should be thought of—and think of itself and operate accordingly—as helpful and supportive, not punitive.

Experimental engineering/product R&D ↔ Pharmaceutical manufacturing

True events:

1. Experimental engineering and Product R&D developed a computer-controlled automatic lyophilization process that was then transferred to Pharmaceutical Manufacturing.
2. Air-suspension particle/tablet coating and electronic monitoring of Pharmaceutical Manufacturing’s tablet presses improved product quality and manufacturing efficiency.

DEVELOPING LINE-EXTENSION PRODUCTS

Once the first NCE product has been approved by the government regulatory agency, additional products (Product Line Extensions, or PLEs) are often developed. Requests for PLEs usually come from Marketing and Sales, but ideas can come from any employee/department, particularly from Product R&D. Regulatory approval of most PLEs is less demanding than for NCEs, e.g., there are fewer safety and clinical experiments.

SUCCESSFUL MANAGEMENT OF DRUG DEVELOPMENT

Successful management consists of a series of thoughts, attitudes, feelings, and skilled practices, not simply (and

counterproductively) an “I-am-the-boss-so-do-what-I-say” culture.

Characteristics of a Fertile R&D Work Environment

Psychiatrist Abraham Maslow (1) emphasizes that managers should not be sculptors (molding/forcing/shaping workers) but farmers who create a fertile work environment wherein all employees can learn, grow, and do their best. R&D managers can optimize group spirit, cohesiveness, innovation, and performance by looking at the team as a circle (in which everyone has a significant contribution to make), and not as the usual organizational pyramid with the boss at the top.

Here are characteristics of a fertile R&D work environment, starting with six cornerstones.

- **Ethics**—This is noted first for obvious reasons.
- **Empathy**—Empathy is the key to proper treatment of hands-on workers, e.g., managers should ask themselves how they would want their boss to treat them.
- **Respect**—Feeling respected (valued) as a unique person and as a contributor to group success is as important as anything in the workplace.
- **Trust**—This is an identical twin to respect. Without mutual trust, two or more people cannot survive a joint effort. The manager’s genuine trust of employees as well intentioned, responsible adults is crucial to an efficient, productive work atmosphere.
- **Caring**—Managers need to have a genuine concern for employees’ personal and professional well-being.
- **Communication**—Keeping employees informed fosters feelings of importance and “being in on things.” Managers who listen carefully and inform abundantly will soon be surrounded by good communicators.
- **Sense of Purpose**—The manager and group members need to know how their department fits into the R&D division.
- **Commitment**—Everyone has a strong commitment to the long-term health of the division and to harmonized company, R&D, group, and personal goals.
- **Competence and Dedication**—Group members are competent and confident in each other’s ability and dedication.
- **Urgent but Well-reasoned, Goal-Focused Activity**—Activity is goal-focused, with the general work pace being a healthy blend of urgency and contemplative, well reasoned, deliberate progress.
- **Involvement**—Productivity, personal growth, and morale will be high if everyone feels involved in the

planning, decision-making, and movement of the department toward its goals.

- Individuality—Everyone wants to be treated as an important, unique individual.
- Acceptance—Related to individuality, each person is valued just as they are; idiosyncrasies are tolerated, even appreciated, because variety is the “spice of life,” and no one is perfect.
- Civility—“Good management starts with good manners, society’s means of ensuring consideration of others as people” . . . Anonymous.
- Agreeableness, Amiability, Friendliness—These are potent antidotes for nervous tension and anxiety and create as much tranquility in a work environment as civility.
- Honesty, Candor, Openness—If trust is basic to all good things in human relationships, honesty, candor, and openness are basic to sustaining/strengthening those relationships.
- Stability, Security, Predictability—The manager needs to be ethical, fair, and consistent without being inflexible. This minimizes uncertainty and anxiety and makes employees feel safe and secure. More energy can then be focused on productive work.
- Cooperation—When co-workers respect and care for one another and when all people feel safe and secure about their “place in the sun,” the stage is set for cooperation rather than for competition, for mutual esteem and pride rather than for envy.
- Recognition—Everyone is recognized for his or her accomplishments and value to the organization. Consequently, they feel good about themselves, their colleagues, their boss, and the corporation.
- Thoughtfulness (Consideration for Others)—It has been said that nothing is more contagious than nervous tension, but surely thoughtfulness must run a close second. This can take many forms, but it generally involves getting outside oneself, being significantly oriented toward others rather than solely toward oneself. It means giving rather than taking.
- Genuineness, Realness—A quality work environment helps workers feel safe, secure, and accepted, thus encouraging them to be themselves, “warts and all.” This reduces facades, protectiveness, and defensiveness, and allows people to focus their energy on productive work.
- Independence and Interdependence—When employees are valued for their basic worth and respected for their individuality, they develop a strong sense of independence. Consequently, they will have little need to “throw their weight around” or assert their individuality at the expense of others. When people feel good about

themselves and receive emotional support from colleagues, they develop a strong sense of both independence and interdependence.

- Cohesive Group Spirit—Under good management, a group develops a close, family-type working relationship in which individual members care for, trust, and respect, and have confidence in one another.
- Deference—If the aforementioned characteristics are operative and the manager seeks advice from and defers to hands-on workers, group members will defer to one another’s expertise and judgment when appropriate.
- Pride (but not Arrogance)—Realistic pride in oneself and in the work group strengthens group cohesiveness and individual self-confidence and fosters a “can-do” attitude toward innovation and technical challenges.
- Loyalty and Enthusiasm—When colleagues work at being civil and respectful, loyalty and enthusiasm will arise spontaneously. As Napoleon said, “An army’s effectiveness depends on its size, training, experience, and morale . . . and morale is worth more than all the other factors combined.”

Honorable and Successful Management Practices

The following are specific suggestions for managing drug development. Two quotes say it all:

Honor is the quality of personal integrity. It is won slowly by a lifetime of small decisions where one places the virtues of compassion and justice ahead of his own advancement. — *Anonymous*

The only way to be a successful manager is to learn to behave like one. — *Jay Hall*

Behaving Ethically

This is listed first because everything must be based on consistent ethical principles and behavior. R&D managers work in an especially complex environment: they hold great power and interact in complicated ways with hands-on workers, peers, upper management, customers, local and national government agencies, and citizen groups. R&D managers also face a unique situation because often results must be taken on faith. Most projects involve extensive experimentation and complex interpretation; some involve equivocal results. Managers must:

1. Trust the competence and honesty of laboratory workers, and
2. Make honest use of the group’s experimental results.

The leader is primarily responsible for the ethical attitude, behavior, and performance of the department. Here are three ways to ensure success:

1. Set a good example—The manager sets the tone for any group, acting as a role model. When employees observe that the boss is consistently ethical, they will set similar high standards for themselves.
2. Minimize unethical conformity and group-think—Peer pressure dominates most teenagers (“All my friends do it ...”), but adults must resist the slightest unethical behavioral norms.
3. Create a safe, stable, predictable work environment—Unless they are pathological, people suffer ethical lapses out of confusion, fear, and insecurity, not out of evil. When workers are afraid of punishment or dismissal if schedules aren’t met, work tends to become sloppy. When R&D vice-presidents fear their jobs are in jeopardy, they might manipulate the information going to corporate management so that their organization will look good.

In summary, leaders need to have the personal qualities of candor, genuineness, and integrity. If not, they are in danger of becoming manipulators of people and situations.

Empathizing (The Golden Rule of Management)

As one of the six cornerstones of successful management, the importance of empathy cannot be overemphasized. Before managers make decisions affecting workers (read “almost all decisions”), they need to remember the Golden Rule of Management: How would I like to be treated? This safeguard on managerial behavior is nearly infallible; here are two examples:

Promotions

Case A: An experienced laboratory scientist believes that he deserves a promotion to the next grade. He raises the issue with his supervisor; she tells him that he has potential, but she is not yet convinced that he has sufficiently proven himself. Nineteen months later he is promoted.

Case B: An experienced research scientist’s manager comes into his laboratory and informs him that he has been promoted to the next grade. The scientist is delighted but surprised; he didn’t expect promotion for another year. The manager says that it may be a little early, but based on his performance and obvious potential, she believes that he deserves to be promoted now.

Reading this example, managers need to play the role of the laboratory scientist and ask themselves how would they would like to be treated by their boss; and would A or B elevate their spirit and incentive to work hard?

Treating employees as well-intentioned, responsible adults

Case A: Laboratory workers discover that, inadvertently, they are not in compliance with several Good Laboratory Practices regulations. When they inform the manager, he calls a department meeting, scolds all of them, orders them to shape up or else, and demands a full report within 3 days.

Case B: When the manager is informed of the GLP violations, he calls a department meeting, thanks them for alerting him to the problem, and asks them if he can be of any help; who would like to take charge of correcting the situation; and what deadline should be set so that he can reassure his boss.

Again, managers should play the role of laboratory workers and ask themselves how they would like to be treated, A or B.

Communicating

Here are two quite different definitions of communication:

1. The goal of communication is to persuade the listener to agree with the speaker. In this case, one gives little thought to the other’s position. Most time and energy are spent thinking about what to say next rather than listening.
2. The purpose of communication is to create understanding. Here the emphasis is on listening, accepting differences of opinion, and freely expressing feelings.

Many sounds (street noise, small talk at a party) are assimilated through a process of hearing. This is primarily a physical phenomenon, with only a small mental element and virtually no emotional component. But when someone tries to communicate something they consider important, hearing becomes inadequate; instead, the other person needs to listen to and understand what they have to say. Now the process requires three components—physical, mental, and emotional—and becomes much more complex.

Generalizations aside, managers need to look at how communication skills apply to their job. First, the leader needs to communicate to employees the kind of person he or she is. Good working relationships, the cornerstone of good performance, depend on people getting to know one

another. It is the manager's responsibility to initiate and encourage the communication process that brings about understanding. For example, he should meet at least once a year with each employee to find out:

- How things are going, personally and on-the-job;
- What is good and what is bad about the work environment;
- Always refer to the department as "ours," never "mine";
- Always say, when introducing a group member to someone else, "she works with me," never, "she works for me";
- Encourage being called by his first name, because hierarchy tends to disappear when people are on a first-name basis;
- Try to be the first one at department meetings; too many managers, for various reasons, wait until everyone else is assembled and then walk into the room. Making people wait is an inherent sign of disrespect, intended or not;
- Keep employees informed by posting all nonconfidential memos and notices on a large bulletin board under the categories of "urgent," "new," and "old." These actions are not gimmicks as long as they represent a sincere effort by the manager to communicate clearly and emphatically that everyone in the group is important.

Another valuable communication tool is seldom used—anonymous employee opinion surveys conducted by the manager, in addition to those of the Human Resources department. Annually, workers are asked to submit to the secretary their answers for three questions. The secretary then combines all responses:

1. What I like about my job;
2. What I don't like about my job; and
3. What my boss can do to make things better. Then (and most important), the manager should call a department meeting and ask the group to decide the priorities for improving the work environment.

Organizing

Hands-on laboratory workers

Experts agree that the number-one problem in any corporation is the underutilization of employees. This stems from management's failure to appreciate their intelligence and potential. Douglas McGregor (2) says that most people enjoy working and, given the chance, prefer to exercise self-direction and self-control. They are highly motivated, seek responsibility, and can be a major positive force in company operations if management allows them.

For example, too often Ph.D.s (scientists) are informal project leaders when B.S.-level laboratory workers (associate scientists) can do that job very well, allowing the Ph.D. to act as theoretical science advisors for several project teams. Then Laboratory Assistants (Technicians or Assistant Scientists) with high school diplomas or 2-year degrees can do most of the daily laboratory work with relatively little supervision rather than being treated as a "pair of hands." When employees feel respected and are encouraged to grow and take as much responsibility as possible, they will consult with higher-level scientists when necessary to minimize mistakes and to ensure group success.

First-line supervisors

Group Leaders/Section Heads should be responsible for approximately 15 laboratory workers and should think of themselves as supervisors, not superscientists. The nonsupervisory Ph.D.s should make most of the broad scientific decisions; the Group Leader/Section Head should concentrate on becoming a successful manager of people.

Management

Most organizations have too much management; the number of supervisors at all levels should be minimized.

Supervising

Douglas McGregor says that the supervisor should act as a helper, teacher, consultant, and colleague. Rarely should she assume the role of authoritative boss. Tom Peters agrees, saying "Leaders are servants." Hands-on workers produce results; the manager's primary job is to support them and remove obstacles that prevent them from doing their best. The best supervisor will ask workers what she can do to help; rather than making sure the jobs get done, she assists the workers in doing so.

Many leaders feel this approach reduces their organizational power (using "power" in the good sense), but the exact opposite is true.

1. The manager "sees over" more clearly because discussions with employees will sharpen the view of the situation for everyone.
2. Control and communication are enhanced because employees will interact frequently with a manager who is viewed as a source of help rather than as a giver of orders.
3. Treating employees as competent, responsible adults will greatly increase her influence with them.

Controlling

Originally, controlling meant running a tight ship; the boss watched over everyone and everything, told employees what to do and when to do it, and made all the important decisions himself. But this made the leader and hands-on workers antagonists, resulting in the manager knowing relatively little about what really went on in the trenches. In contrast, appropriate (nonauthoritarian) control means emphasizing employees' self-control simply because most people are trustworthy. This is especially true in R&D organizations in which the emphasis is on high technology, creativity, and innovation.

Peter Drucker, Tom Peters, and Douglas McGregor agree:

1. Drucker says that to be productive, workers need to have control over their work; control is a tool of employees and must never be their master.
2. Peters calls this the control paradox: less is more. Less central control and more genuinely delegated self-control for those closest to the action translates into tighter overall control.
3. McGregor believes that successful supervision is largely dependent on the manager's ability to predict and control human behavior, and the essence of control is selective adaptation. People control the physical world around them, not by expecting nature to do their bidding, but by adjusting their actions to natural laws. For example, humankind does not control surface water by commanding it to flow uphill; rather, people dig channels, adjusting to the fact that water obeys the law of gravity.

Similarly, effective management control consists of channeling workers' energies, interests, and capabilities into activities that meet organizational objectives. Management controls the work force by adjusting its decisions and actions to the realities of human nature, and not by telling people what to do and expecting blind obedience.

Delegating

Managers should delegate because employees like, want, and need to do things their own way. The wise leader assigns research projects only after consulting with laboratory workers, preferably at a department meeting, because bench scientists tend to have a much better feel than the manager concerning who has the time; who is most qualified; and how best to divide up responsibilities.

Also, employees feel more respected and involved when they are part of the work-assignment process.

Delegation, not relegation

Since the biggest waste in any organization is underutilization of employees, delegation makes sense. Unfortunately, too many leaders confuse delegation with relegation. Delegation means assigning responsibility and authority to a representative. Relegation, on the other hand, connotes consignment to an inferior position. When managers give employees narrow, menial tasks and expect them to do all "delegated" work exactly as they would, their action is relegation, not delegation.

Goals of delegation

Three primary goals of true delegation are to:

1. Relieve managers of some of their workload so that they have more time to think, meditate, plan, learn, and grow;
2. Move work and responsibility as far down the organizational ladder as possible, increasing efficiency; and
3. Offer all employees maximum challenge and opportunities for growth, even when formal promotions are not immediately available. This increases productivity and develops future leaders.

In harmony with these goals, proper delegation has two distinguishing characteristics:

1. Most of the delegated tasks are a meaningful part of the manager's job and not just drudgery to avoid;
2. Workers are allowed, even encouraged, to perform the delegated work in their own way, with the manager helping only when asked.

Case study: In the movie *Bullitt*, a police lieutenant is working on a case that is of intense interest to a powerful U.S. senator. The senator is not happy with the way the lieutenant is conducting the investigation, so he pressures the lieutenant's boss to force him to proceed differently. As it happens, the captain agrees with the senator but refuses to interfere, saying "It's his case, Senator."

Monitoring

Here managers can be either authoritarian (running a tight ship) or smart. They can foolishly spend much of their valuable time monitoring their operation or they can wisely delegate most of that function to employees, thus improving productivity and orderliness, fostering growth, and freeing up more time for broader, long-range tasks.

In a good work environment, the impetus for monitoring comes from below, not from above. This places a positive focus on the process. The manager trusts

group members and does not feel a need to monitor the operation closely. At the same time, employees are eager to keep managers informed because they sense that they are interested in their work and in departmental progress. They also recognize the need for management to know the general situation, to be informed of major progress or problems.

On the other hand, authoritative monitoring will yield negative results.

- Workers can't help but feel that managers don't trust them—else why would two people do one person's job? Most employees feel, correctly, that it is part of their job to monitor progress and to address problems.
- The situation is inherently inane, and perceived management inanities contribute to employee disrespect and alienation.
- Workers tend to lose interest in doing a careful job of monitoring when they see the boss repeating what they do.
- Management seldom performs lower-level tasks well. Hands-on employees, to whom such tasks are often challenging, are much more motivated and equipped to do them properly.
- Efficiency and productivity suffer, not only because of the redundancy, but because employees are forced to spend time educating the manager about the details.
- If managers get too involved with minutia, they tend to meddle in workers' jobs, and soup is not the only brew spoiled by too many cooks.

Advocating

This management task never appears in a job description, yet it is vital to enhancing the culture, environment, and performance of an R&D operation.

The world is not a fair place, and the world of work is even less fair. Why? In a free society, adults make most of their own decisions, but not in the workplace, where management and corporate rules reign supreme. Employees feel vulnerable, and "results only" oriented management practices can adversely affect their performance and well-being.

In unionized organizations, the agent or steward serves as employees' advocate (although R&D scientists are seldom part of a labor union); in nonunion companies, the Human Relations department usually fills that role. Seldom is management viewed as an advocate for workers, and rarely does management perceive itself that way. In fact, employees often consider the boss a powerful

adversary—the very reason they need an advocate! Surely productivity, not to mention loyalty and enthusiasm, suffers when workers and management consider themselves adversaries.

The manager as advocate

In a well managed, people-oriented organization, the primary advocate for each employee is the immediate supervisor. Not only is that individual in the best position to know and help workers, but she is the major beneficiary of the increased productivity, loyalty, and enthusiasm that follow. Experience has shown that if workers believe their manager is for them and wants them to succeed, and a helper and facilitator rather than an overseer; there is little they won't do for the manager. In fact, when she is under great stress or trouble, the roles become reversed and employees will rush to her aid.

Some managers believe that they should be neither advocates nor adversaries but impartial judges. However, social science professionals have shown that managers are no more rational or impartial than anyone else, that everyone labors under a cloud of personal bias. Furthermore, strict impartiality usually results in impersonal treatment, and no one likes being regarded as a "nonperson." How do managers go about becoming primary personal and organizational advocates for employees?

Personal advocacy guidelines

- Managers should get to know each person as an individual and develop a relationship based on mutual trust, respect, and caring. Then employees will feel comfortable bringing them their concerns, and the manager will be able to give employees optimum help.
- Managers need to be perceptive observers of employees, not to check up on them, but to pick up subtle signs of trouble. For example, if an ebullient person becomes very quiet on the job, and this persists for a week or two, the manager may want to say, "I don't want to be nosy, but is anything wrong? Anything I can do to help?" If the employee does not want to talk, the manager should not force the issue, but observe the situation. Showing genuine interest in individuals has a positive effect, even though they may not want to confide in their supervisor. When an employee's performance starts to decline and the manager is quite sure it is not because of any action or inaction on his part, it is best not to intrude on one's privacy; treat the employee as an adult and provide her every chance to work out the difficulties.

- If the employee's performance continues to deteriorate, there will come a time when the manager needs to meet with her and talk. But if he has built a good relationship with the employee she will most likely confide in him long before that point is reached. If the manager finds himself growing impatient with an employee during a difficulty, it's best that he wait a bit longer; it usually pays off. People appreciate a supervisor who shows patience and faith in them during times of trouble.
- When the worker does confide in the manager, he should refrain from giving advice unless asked. Any troubled individual coming to a supervisor needs, first of all, a sympathetic ear and then appropriate reassurance.
- Even when she presses him for advice, both individuals are best off if the manager simply outlines options and their advantages and disadvantages, leaving the ultimate decision to the employee.
- The manager should not try to do too much. If he senses that professional help may be appropriate, a suggestion to that effect or a referral to the company's employee assistance program may be in order.

Organizational advocacy guidelines

Not all problems are highly personal or greatly troubling. Then it is best for the manager to use his administrative ingenuity to deal directly with such problems as a temporarily stalled promotion, a continuing problem with another department, tension or a disagreement with a colleague, or resentment over an ill-defined or apparently unfair company policy. The no-advice rule softens considerably when the manager has some control over the situation. Employees do not expect management to solve all their problems, but they do expect their supervisor to try when it's important to them.

The advocating manager serves as a bridge between employees and the organization. A key girder in that bridge is "loyalty up and down." Workers need to feel that the manager is truly for them, whereas the manager's boss has to be confident that the subordinate is looking out for the company's interests as well. If the manager has a reputation as a strong advocate for employees, when irreconcilable conflicts arise (assuming no violations of ethics or the law are involved), he can come down on the side of the company without straining his relationship with workers.

One of managers' primary responsibilities is to remove impediments to employees' productivity. This can be viewed as advocacy as well, because such assistance improves workers' well-being.

Sheltering

Here is another managerial task that is missing from job descriptions, but it is very important, especially in these days of high-pressure, 60-h workweeks. What is sheltering? No matter what the weather outside, the roof, siding, and windows of a house provide a hospitable, safe environment for occupants. So, too, do managers need to create an optimum work atmosphere by sheltering hands-on workers.

Shelter them from what? Swirling about any organization are tensions, antagonisms, organizational red tape, and inanities—unpleasant and distracting. For example, perhaps the R&D vice-president, a tense, caustic individual who is uncomfortable with the deliberate pace of research, is constantly berating department managers to speed things up and increase productivity. The strong tendency in such a situation is to translate at least some of that unpleasantness and pressure down to hands-on workers—the "kick-the-dog" syndrome. With this response, not only are managers relieving some of their own frustrations and resentments, but they are convincing themselves, and trying to convince their boss, that they are team players who are bottom-line oriented. Unfortunately, such a reaction seriously damages the group's work environment and will likely reduce, not enhance, performance, especially in R&D, in which innovation and creativity require a positive atmosphere.

To avoid such debilitating problems, managers need to prevent disruption of the group's productive, relatively tranquil work environment by absorbing as much of the pressure and unpleasantness as possible. By taking most of that burden on their own shoulders, the managers protect the department from contamination by poor management practices elsewhere in the organization. This does not mean that legitimate pressures and emergencies should not filter down; employees can and will respond with vigor and enthusiasm to them, especially when the crisis is viewed as a challenge to the entire group. But no one can do their best when they are constantly pressured to hurry! hurry! hurry! Remember, "if you don't have time to do it right, where will you find the time to do it over?"

Successful sheltering is highly dependent on the level of mutual trust, respect, and confidence within the work group. When there is a strong, caring sense of family, intense sheltering is unnecessary because intrusions from the outside have minimal effect. Quality sheltering requires that managers get out of their offices and see what's going on in the department. If they have no sense of the group's day-to-day moods, detecting rising patterns

of tension or indifference becomes difficult. Also, employees are more apt to call problems to the manager's attention if they perceive her as interested, friendly, and accessible.

Five important personal qualities required for good sheltering are:

1. *Strength*. The manager needs to swim against a strong current (i.e., the system or an unreasonable, autocratic boss).
2. *Courage*. There is risk to the manager who stands up for her people against superiors or the system.
3. *Stamina*. Sheltering is a never-ending task.
4. *Ingenuity*. Deflecting a rushing stream (again, the system) usually works better than constructing a dam.
5. *Tact*. Diplomacy usually makes deflection acceptable to the system.

Managers who provide effective shelter for their employees will succeed, both with their workers and with their bosses.

Fostering Creativity

Managers contribute to creativity by:

1. Demonstrating enthusiasm and excitement for new ideas;
2. Managing with a light touch, allowing laboratory scientists the freedom to grow in their own way and at their own pace;
3. Encouraging employees to take risks and explore new territory—and being there with encouragement rather than criticism if they fail;
4. Being both flexible and secure and creating a work atmosphere with those same characteristics;
5. Being committed, not just to today's and this year's comfort and well-being, but to the long-term health of the organization and its employees; and
6. Hiring competent, innovative scientists.

Tips for encouraging creativity

1. "Two heads are better than one" is not a cliché. Ideas are often enriched through discussions with others.
2. Brain-storming sessions involving multidisciplinary groups often produce marketable ideas. The cardinal rule of brainstorming is that no expressed thought be evaluated during the session, because the threat of critique inhibits the free flow of ideas.
3. Managers should refrain from judging employees' ideas. It's best to ask workers to research their concepts with the help of their colleagues (e.g., R&D, Marketing, and Sales) and then evaluate it themselves.

4. When innovation involves replacement of old technology, management often entrusts development of the new technology to old-technology experts. At times this succeeds, but the effort often fails because these veterans have too much intellectual and emotional investment in the old way. It is usually best to assign development of new technology to competent but relevantly inexperienced scientists who will take a fresh, unencumbered approach to the problem.
5. Formal suggestion systems can be beneficial, but management must guard against calcification, by which the suggestions nourish the bureaucratic system instead of the other way around.

Motivating

When thousands of supervisors were asked to list, in descending order, what they thought motivated workers, they got it all wrong. They listed:

1. "Interesting work" as #5, but in the same questionnaire workers chose it as #1.
2. "Full appreciation for work done" as #8, whereas employees rated it #2.
3. "Feeling of being in on things" as #10 whereas workers assigned it #3.

It is clear that managers need to learn more about motivation.

Experts agree that human behavior is not random, but caused; internally motivated; and always directed toward some goal.

Strictly speaking, managers do not and cannot motivate employees. The best they can do is provide the stimuli to which workers react, driven by their own internal motivation.

Psychologist Abraham Maslow (3) hypothesized a hierarchy of needs that classifies human motivation, listed in descending priority:

1. Physiological (hunger, thirst);
2. Safety (security, stability, predictability);
3. Belongingness and social integration (companionship, being part of a group);
4. Esteem (self-respect and recognition/respect from others); and
5. Self-actualization or self-fulfillment (each person's drive to move toward being the very best of which he or she is inherently capable).

The basic needs (physiological) are most important, but once satisfied, they no longer motivate, and the next set becomes operative. Because most employees are not

starving nor threatened by anarchy, unless job security is a factor managers should be concerned with:

1. belongingness or social needs.
2. esteem, and
3. self-fulfillment.

Fostering Employee Growth

Human beings have innate pleasure in a sense of growth and improvement. — *Psychiatrist Willard Gaylin*

To foster workers' growth, managers must:

1. Understand people so that their efforts will harmonize with the realities of human nature;
2. Understand the learning process to ensure the optimal rate of growth; and
3. Apply that understanding on the job with diligence and patience.

Let's examine these three components.

Understanding people

The human characteristics most relevant to the learning/growing process are:

1. *Desire to grow.* Everyone has an inherent tendency to move toward psychological health and maturity.
2. *Personal freedom.* The more people have a say about what they do, how and when they do it, and the direction and pace of personal growth, the faster and surer they will progress.
3. *Uniqueness.* People are different from each other. Everyone is a unique individual, as varied as fingerprints. Each individual has different needs and distinct ways of satisfying those needs.

Understanding the learning process

What does an R&D manager need to know about the learning process?

- Learning can be cognitive, as in memorizing multiplication tables or reading books; or experiential, as when riding a bicycle or working effectively in groups. Most learning is a combination of the two, especially in technical organizations, in which cognitive scientific knowledge must be integrated with a wide variety of experiential skills.
- Personal growth is best achieved experientially. Carl Rogers (4) defines the elements involved:
 1. Primarily self-initiated, involving the entire person, physically, intellectually, and emotionally. A person learns best when he is ready and wants to learn;

2. Pervasive, making a difference in the attitudes and behavior of the learner;
3. Self-evaluated, in that the learner is the one who decides whether the learning experience is meeting his needs; and
4. Comprehensive, so that total (intellectual and emotional) meaning is experienced.

Application

Here are some general guidelines (keeping in mind the characteristics of a fertile work environment mentioned earlier). Leaders should:

- Manage with a light touch;
- Set a good example;
- Get to know workers as individuals;
- Involve employees as much as possible;
- Encourage risk-taking and creative thinking;
- Not underestimate employees' potential; and
- Consider overall personal growth.

Recruiting

When competent workers are hired, a manager is almost guaranteed good results. However with inadequate employees even a brilliant leader is in serious trouble. Therefore, good recruiting practices are basic to a successful R&D operation.

Effective recruiting depends first on good management of the people already on hand; the motivation, enthusiasm, and loyalty of group members will be obvious to visiting prospects.

True event. Under autocratic management, a division had 18 consecutive rejections; after changing to people-centered supervision, nine of 10 candidates accepted.

Two major steps in recruiting scientific personnel are:

1. Identifying quality candidates, and
2. Bringing them to the organization for in-depth interviews (including a seminar).

The best way to succeed with the first step is to visit universities for interviews with students. The ethical way to succeed with the second step is to have the candidate talk to as many hands-on workers as possible and to encourage interviewers to be honest and candid in answering questions about the organization. The more welcome and respected candidates feel, the better the chances they will accept an offer.

Recruiters should gather information about the candidate by talking with major advisors, other faculty members, former classmates, and formal references. The following personal characteristics are especially important:

competence, productivity, genuineness, growth potential, flexibility, open-mindedness, cooperativeness, deference, communication, self-confidence, motivation, thoughtfulness, ethics, independence, and commitment.

Conducting Formal Performance Reviews

Moral or diagnostic evaluations are always threatening.

—*Carl Rogers.*

The typical formal performance review is similar to a trip to the dentist: There is apprehension before, pain during, and a sense of relief afterward, when the session is over.

Douglas McGregor (2) explains why:

1. Evaluating an employee's performance is highly dependent on the manager's psychological makeup.
2. For various reasons (including managerial malfeasance), the appraisal often has little relation to reality.

True event. Roger, a first-line R&D supervisor, had always received above-average performance reviews. He made no waves, followed orders, did the required paperwork, and sleep-walked his way through corporate life for 10 years, totally ignoring employees' concerns and frustrations. When a new, highly competent department manager arrived, she quickly recognized Roger's incompetence and pressed for his demotion, transfer, or termination. This caused a great uproar because the "record"—a decade of innocuous, relatively positive performance appraisals by a variety of managers—painted an entirely different picture.

1. Experts agree that to a great extent, a worker's performance is a function of how he is managed.
2. Concerning criticism, the effectiveness in communication is inversely proportional to the employee's need to hear it. The more harsh the criticism, the less likely the person can/will accept it.
3. The manager may be able to convey negative judgments, but this will seriously damage the relationship.
4. Performance reviews accentuate the worker's dependence on the manager.
5. It's an open question if a troubled individual really wants to hear about his deficiencies.
6. Concerning amateur counselors (which managers are) Carl Rogers (5) says that the most they can accomplish is a temporary change, which then disappears, leaving the person more than ever convinced of his inadequacy.

There's a better, well-proven, successful approach (which assumes that all employees are performing

adequately; if that's not the case, then the manager needs to transfer or terminate unacceptable performers—see the next section).

1. Well before the actual reviews, employees are reminded that the "system" demands a formal performance appraisal and that certain rituals must, as in the past, be followed.
2. Then the manager reminds them that she has been in close contact with each of them throughout the year, and that consequently, there will be no surprises during the formal interview.
3. The manager assures everyone that their performance review will be a pleasant experience.
4. Then the manager holds formal reviews only when in a relaxed, reassuring mood.
5. When the person first comes into her office, she reiterates the first 3 points.
6. The manager emphasizes—genuinely—what a good job the employee is doing and how glad she is to have him in the department.
7. Then the manager goes into specifics concerning what she likes about the employee and his performance over the past year; he is encouraged to add accomplishments the manager has failed to mention.
8. Next, the manager reminds the employee that the system requires that she record some negatives and asks for suggestions (e.g., he often puts things off until the last minute).
9. The manager then reassures him—genuinely—concerning any weaknesses he brings up, e.g., "If you were perfect, you'd make the rest of us look bad." "Your intentions are always good, and your strengths far outweigh your weaknesses." Any additional discussion in this area should consist exclusively of the manager asking the employee how he feels about the negatives and, most important, where she—or the work situation—is deficient in helping him do his best.
10. The manager then records—honestly—the weaknesses the employee mentions, but puts them in as positive a light as possible (e.g., "Employee tends to complete some assignments at the last minute, but has a good sense of priorities and is always on schedule").
11. Then—and this is especially important—the manager asks the employee what she can do to help him improve even more. She then sits back and listens in a non defensive manner, thanking the employee—genuinely—for the candid feedback (and follows up on the suggestions).

12. The manager sums up by re-emphasizing—again, genuinely—an appreciation for the contributions to the group's accomplishments and her delight in having him as a member of the department.

In summary, criticism should be avoided here and in all management behavior, simply because it's counterproductive, and employees who feel accepted and valued by their supervisor will engage in self-criticism. This is, by far, more productive over the long term.

Transferring or Terminating Unproductive Employees

If, in spite of good management practices, a hands-on worker cannot do the job adequately, the manager is obliged to transfer the employee to a more suitable position elsewhere in the corporation. If that is not an option, then the employee should be terminated if legally possible and ethically acceptable.

Here's a recommended procedure to follow:

1. The manager should consult the Human Resources department personnel for official guidance. In fact, it is wise to work with them unofficially at a much earlier stage.
2. After ensuring that his actions have been consistent with company policy, the manager should bring the employee into his office for a private, candid, uninterrupted, and caring conversation.
3. He should begin by saying, "I'm sorry, but things are not working out and you are facing termination. You don't have to leave tomorrow, but if you were still here 6 months from now it would be a problem. Sometime between tomorrow and 6 months from now you need to find a job with a different company. I'll help you all I can, but in the end, that's your responsibility."
4. The person will almost certainly be upset and antagonistic toward the manager. At that point, he needs to remember that the employee is going through an ultratraumatic process and needs all the sympathy and support he can muster. The last thing the manager should do is try to defend his decision unless pressured by the employee, which usually does not occur.

Managers who recruit and manage well will seldom have to perform this unpleasant task.

Promoting

Proper selection of new supervisors is crucial to organizational success.

Promoting the wrong people

The most common mistakes made when choosing people for R&D management positions are:

1. Assuming the best laboratory worker will make the best supervisor (management is an entirely different world than laboratory research);
2. Choosing autocratic, results-oriented people rather than person-centered individuals; and
3. Going outside the company instead of promoting from within, which lowers employee morale and motivation.

Promoting the right people

A key element in promoting the right people is having the right people to promote; thus, recruiting the best candidates and creating a growth-fostering environment are crucial. It is also important to recognize the personal qualities (especially interpersonal orientation) needed to make a good supervisor. Managers should avoid promoting people with little talent for or interest in management. The best way to prevent this is to provide potential supervisors with temporary responsibilities and then evaluate their performance.

Organizing Project Teams

There are two common ways to organize project teams:

1. Informal (in which team members belong to scientific discipline-oriented departments with the leader's role being coordinator, not boss).
2. Formal (in which the project leader has organizational authority over team members).

In pharmaceutical R&D, it seems best to have informal project teams because:

1. Everyone needs an organizational "home" for emotional stability and security reasons, and formal project teams tend to make scientists migrant workers, e.g., when the project is completed, they are often assigned to another project with a new boss.
2. Formal project teams tend to inhibit intergroup cooperation and create unhealthy competition among teams (we should have the highest priority regarding limited joint resources and we want to make sure that our project comes out first in upper management's rating).

International R&D

Domestic R&D managers face special challenges in a multinational corporation; in general, they need to do the following:

1. Know the organizational specifics of their company's international operations;
2. Learn as much as possible about the cultural, economic, technical, and governmental differences among the various countries;
3. Establish close ties with their international R&D colleagues.

A multilingual R&D manager is especially valuable.

R&D Management and Corporate Management

R&D management needs to:

1. Understand and appreciate the views and concerns of corporate management;
2. Develop a broad vision for R&D that is harmonious with long-range corporate plans; and
3. Decide, with workers' help, what is in the best interests of the corporation.

Corporate management must learn enough about the world of R&D to:

1. Appreciate R&D's point of view;
2. Understand the R&D process so that funding is steady, not sporadic; and
3. Recognize the danger signals (confusion, aimlessness, consistently poor decisions, laboratory worker malaise) of poor R&D management to distinguish them from bad-luck cycles inherent to all R&D.

Industry and Academia

Interactions between industry and academia have mutual benefits.

Advantages for academia

1. Educational benefits:
 - Encourages the cross-fertilization of ideas;
 - Offers temporary, education-focused work in industry for faculty, undergraduates, and graduate students; and
 - Develops joint projects for increased knowledge.
2. Financial and other benefits:
 - Support for research
 - Possible employment opportunities for students after graduation
 - Consultantships for faculty
 - Rapid commercialization of academic research

Advantages for industry

1. An increased knowledge base for:
 - Cross-fertilization of ideas
 - More options for new and better products
 - More flexibility in R&D spending (academic support can be enlisted for an urgent but speculative project without making long-term internal commitments)
2. Greater professional development of employees through:
 - Teaching and lecturing opportunities in academia
 - Research sabbaticals
 - Internal short courses given by academic consultants
3. Successful recruiting of new personnel owing to:
 - More thorough evaluation of potential job candidates via summer employment of students
 - Improved corporate image among students and faculty.

BIBLIOGRAPHY

- Argyris, C. *Personality and Organization: The Conflict between the System and the Individual*; Harper & Row: New York, 1976.
- Bennis, W.; Nanus, B. *Leaders' Strategies for Taking Charge*; Harper Collins: New York, 1997.
- Drucker, P.F. *The Practice of Management*, 2nd Ed.; Harper Collins: New York, 1993.
- Goleman, D. *Emotional Intelligence: Why it can matter more than IQ*; Bantam Books: New York, 1995.
- Hall, J. *The Competence Process: Managing for Commitment and Creativity*; Teleometrics International: Woodlands, TX, 1980.
- Leavitt, H.J.; Bahrani, H. *Managerial Psychology: Managing Behavior in Organizations*, 5th Ed. University of Chicago Press: Chicago, 1988.
- Likert, R. *New Patterns of Management*; McGraw-Hill: New York, 1987.
- Maslow, A.H. *Eupsychian Management: A Journal*; Irwin-Dorsey: Homewood, IL, 1965.
- Maslow, A.H. *Motivation and Personality*, 2nd Ed.; Holtzman, W.G., Murphy, G., Eds.; Harper & Row: New York, 1970.
- McGregor, D. *The Human Side of Enterprise*; McGregor, C., Bennis, W.G., Eds.; McGraw-Hill: New York, 1985.
- Rogers, C.R. *Freedom to Learn for the 80s*; Charles, E. Merrill, Ed.; Columbus, OH, 1983.
- Rogers, C.R. *Carl Rogers on Personal Power: Inner Strength and its Revolutionary Impact*; Delacorte Press: New York, 1977.
- Sampson, R.C. *Managing the Managers*; McGraw-Hill: New York, 1965.
- Tingstad, J.E. *Good Technical Management Practices: A Complete Menu*; Interpharm Press: Buffalo Grove, IL, 1999.

LOZENGES

Robert W. Mendes

Massachusetts College of Pharmacy and Health Sciences (Retired), Dedham, Massachusetts, U.S.A.

Hridaya Bhargava

Massachusetts College of Pharmacy and Health Sciences, Boston, Massachusetts, U.S.A.

INTRODUCTION

Lozenges are solid preparations that contain one or more medicaments, usually in a flavored, sweetened base, and that are intended to dissolve or disintegrate slowly in the mouth. They can be prepared by molding (gelatin and/or fused sucrose and sorbitol base) or by compression of sugar-based tablets. Molded lozenges are sometimes referred to as pastilles, whereas compressed lozenges may be referred to as troches. They are intended to be allowed to dissolve on the back surface of the tongue to provide drug delivery locally to the mouth, tongue, throat, etc., to minimize systemic and maximize local drug activity.

The USP (1) currently recognizes Cetylpyridinium Chloride Lozenges and Nystatin Lozenges. However, more than five dozen over-the-counter (OTC) lozenge products are currently marketed (2). These contain a variety of active ingredients including antimicrobials and local anesthetics for throat pain; aromatics, herbals, zinc salts, decongestants, and cough suppressants for colds, allergy, cough, and congestion; and nicotine-like substances for smoking cessation.

HARD CANDY LOZENGES

Raw Materials

The types of raw materials used in medicated lozenges may vary according to a number of factors. Most medicated lozenges contain sugar, corn syrup, acidulant, colorant, flavor, and the medicament.

Sucrose, a disaccharide of glucose and fructose, is obtained from sugarcane or beet. The choice of beet or cane sugar is based on availability and geographical considerations. Sucrose and sucrose products are used in medicated lozenges because of their value as neutral sweeteners, their ready solubility, and their function as a “drier” to reduce the weight of the confection through crystallization. Invert sugar, derived from sucrose,

possesses the very desirable physical property of controlling the crystallization of concentrated sugar solutions and maintaining freshness of the finished product through its humectant properties.

Corn syrup is used in almost every type of confection to control sucrose and dextrose crystallization, which may lead to crumbling. Corn syrup in appropriate proportion with sucrose and dextrose allows the formation of an amorphous glass and produces a candy with the desirable appearance. The following physical properties of corn syrup are extremely important in the preparation of medicated candies: density, dextrose equivalent (DE), hygroscopicity, sugar crystallization, viscosity, freezing-point depression, and osmotic pressure.

Colorants are incorporated into medicated lozenges for appearance, product identification, and masking of physical degradation. Dyes and other organic colorants may degrade by heat or light via oxidation, hydrolysis, photooxidation, etc., and their compatibility with drug, excipients, and process conditions should be studied before selection. Suppliers of colors are excellent sources of information on current regulatory status of colorants.

Acidulants are generally added to medicated lozenges to fortify and strengthen their flavor profile. Organic acids such as citric, malic, fumaric, and tartaric acids are most commonly used. Citric acid alone or in combination with tartaric acid is the most common. Another use of acids in medicated lozenges is to alter the pH to maintain the integrity of the drug. Regular conversion corn syrup has a pH of 5.0–6.0. Addition of a weak organic acid to improve flavor lowers it to 2.5–3.0, a pH at which some medicaments exhibit maximum stability (3). If necessary, some drugs can be stabilized by adjusting the pH to 7.0–8.0 with a suitable weak base such as calcium carbonate. Some research has shown that excessive use of acidic lozenges could have the potential to enhance existing dental erosion (4), and that low pH (2.6–3.7) leads to dissolution of calcium and phosphorous from hydroxyapatite (5). Others have shown that excessive use of citric and tartaric acids may effect bioavailability of zinc in zinc lozenges (6, 7). Another report indicated that

the activity of cetylpyridinium chloride in candy-base lozenges is influenced by pH, with >5.5 being most desirable (8).

Acceptable taste is necessary to ensure patient acceptability, and this can be the determining factor between commercial success and failure of an OTC product. Flavors used in medicated lozenges must be compatible with the drug and excipients and capable of withstanding the rigors of the manufacturing conditions. Flavors consist of numerous chemicals that may interact with excipients or medicaments and that degrade by heat and light. Aldehydes, ketones, and esters may react with drugs. A classic example of flavor-drug interaction is that of a primary amine drug (benzocaine, phenylpropanolamine) with aldehyde-containing flavor components like cherry, banana, etc., resulting in the formation of a Schiff base, drug decomposition, and loss of efficacy. Adjustment of lozenge base pH to accentuate certain flavors (e.g., citrus) may also result in incompatibility with some medicaments (e.g., benzocaine).

The last major ingredient in lozenges is salvage obtained from lozenge batches rejected because of imperfect shape or size, presence of air bubbles, or unacceptable drug concentration. Salvage, if properly heated, can be reused in finished products without altering color, texture, lozenge base composition, or drug concentration (9). Before any salvage can be used as part of a medicated lozenge base, it should be adjusted to a pH of 4.5–7.5 to prevent excessive and uncontrolled formation of reducing sugars, and the stability of the drug at cooking cycles should be determined.

Processing Methods

There are three types of candy-base cookers: fire cookers, high-speed atmospheric cookers, and vacuum cookers. Vacuum cooking is the process of choice for manufacturing hard candy lozenges. It is based on the principle that water boils at a lower temperature under vacuum, and can thus be removed. Sugar solutions and corn syrup are boiled at 125–132°C, vacuum is applied, and, due to the heat of the batch, additional water is boiled off without further heating. The resulting vapor is condensed and removed by the cooling water of the vacuum pump (3).

The continuous batch process cooker installation consists of an automatic sugar dissolver, a sugar solution and corn syrup storage unit, metering pumps, precookers and a holding tank, a vacuum pump, and a collection kettle (10). Precookers are standard steam-jacketed kettles equipped with an additional heat exchanger to provide better circulation in addition to more exchange surfaces. Each component (water, sugar, corn syrup, and

salvage) is added by pumps and metering devices controlled by one gearing system, in order to bring the finished precooked syrup to the desired temperature (110–120°C) in 1 min or less (3). The short dwell time significantly reduces the Browning reaction and the amount (1–2%) of invert sugar developed (11). Cooking machines consist of a heating coil, intermediate chamber, vacuum chamber, flow metering valve, turning device that charges the receiving kettle, receiving kettle, kettle turning device, and rotary vacuum. In the standard procedure, a precooked sugar–corn syrup solution at 110–120°C is passed through a chrome–nickel steel cooking coil housed in a steam dome at a temperature of 135–150°C. The coil leads to an intermediate chamber that is vented to the atmosphere. From the intermediate chamber, the syrup is sprayed into the vacuum chamber, which is regulated by a vacuum-actuated metering valve. This affects reevaporation of the sugar mass (12), further increasing the content of dry substance. An adjustable timing device charges the receiving kettle by opening an air valve the moment the batch is cooked. The filled receiving kettle drops from the vacuum hood and is held in front of the cooker by a spring-activated timing device; it is replaced by an empty one (3). The kettle is charged by the syrup pump, resulting in uniform weight of all batches. By selecting an appropriate model, more than 3000 kg of candy can be produced per hour.

Replacement of the collection kettle by a continuously moving stainless steel belt to carry the candy base away from the cooker at a predetermined rate in an unbroken stream has resulted in the continuous process cooker, which has some advantages and some limitations. Its advantages are high-speed production, improved organoleptic characteristics, extended shelf life for physical properties, and cost effectiveness. The disadvantages are reduced controls, the addition of flavor at very high temperature, and increased possibility of nonuniform candy-base production. Preparation of candy base from gear metering of sugar solution and corn syrup to vacuum drying is identical to the batch process described above. However, candy base is continuously drawn off in a thin ribbon. As it leaves the cooker the flavor, preheated to 50–60°C, is injected. The flavored candy mass is dropped onto a variable-speed rotary cone head where it is mixed. Medicament as a solution or dispersion, preheated to 110–120°C, can be metered into the candy mass through dosing pumps. The candy mass then slides down the delivery chute onto the stainless-steel belt where it is mixed and sized by plows and rollers. The temperature of the conveyor belt is controlled by a spray of heated water that can be adjusted as necessary. The acidulant may be deposited onto the candy mass via a vibratory dosing

auger (9); the subsequent steps are identical to those of the batch process.

Candy-Base Manufacturing

The first step in the manufacture of medicated lozenges is the preparation of candy base, followed by the addition of medicament, flavor, acidulants, colors, etc., and finally by lozenge formation. Irrespective of process, the manufacture of medicated lozenges involves the cooking of candy base, mixing, batch forming, "rope" sizing, adjustment of weight, lozenge formation, cooling, and storage of lozenges.

Candy base is prepared from liquid sugar (67% sugar) and corn syrup (liquid glucose 43°Bé, 80% solids) in a ratio of about 60:40. Precooking is initiated under vacuum at controlled temperature. The precooked solution is transferred into the steam-heated coil, where it is boiled and from which it is moved to the intermediate chamber where final mixture is produced (3). The final moisture content of candy base should be about 1%, its temperature about 135°C, and its consistency plastic-like. The candy base is transferred into a kettle mounted on a suitable scale and, if necessary, batch weight is adjusted.

Heat-stable colors are added at this point as cubes or paste. The colored candy base is transferred to a cold stainless-steel cooling plate for the mixing operation. Mixing can be manual or mechanical, using a series of plows and rollers, or a mixer consisting of two arms, a plunger, and a slowly rotating table top.

The temperature of the mixing table is maintained at 40–50°C. Flavor, medicament, acidulant, and ground salvage are added to the colored candy mass when mixing is initiated. After completion, the medicated candy base is transferred to a warm slab and allowed to equilibrate to a uniform temperature.

The mass is cut into workable portions, properly tempered, and placed into batch-forming (holding) machines. For candy formers requiring a flat sheet, heated tables are used. For candies requiring cylindrical pieces to be fed into the batch formers, batch rollers are used. A plastic-like mass is formed into a sugar cone and transferred at a predetermined rate to the sizing roller. The operation of the batch former is synchronized with the rope sizer (12). In order to maintain a temperature at which the outer shell of the candy would not crack, the batch formers are kept at 80–90°C (3).

The height to which the batch formers are adjusted and the amount of material in them dictates the delivery rate of the candy flowing as a rope from the batch former to the sizing rollers. The sizers consist of sets of successively smaller forming rolls. The thickness of the rope is

determined by the diameter of the sizing rollers and governs the weight and size of the candy. The sizers are generally heated to maintain temperatures of 50–60°C, preventing the candy from cracking by rapid cooling (12).

For the formation of the final lozenge, the candy rope is discharged into a forming machine. The formed lozenges are then fed onto the distribution belt, which provides intensive cooling and shaking to prevent deformation of the still plastic lozenge (12).

The formed candy must be cooled as quickly as possible to prevent loss of shape. The candy is usually cooled on a conveyor belt made of chain or canvas. Multibelt coolers are designed in such a way that the first narrow belt (15–20 cm wide) runs as rapidly as the candy-forming machine. At the end of the first belt, a breaker separates the candy and distributes the lozenges uniformly across a wider, slower-moving second belt. The third belt is still wider and travels even more slowly, allowing enough time to cool the lozenge to the desired temperature (below 35°C) (3).

In the sizing operation, the medicated lozenges are collected as they leave the cooling belt and transferred to a series of counterrotating rollers separated via caliper adjustment (12). The sizing operation removes all oversized and undersized lozenges, ensuring uniformity.

The properly sized lozenges are collected and stored in a climate-controlled room at 15–20°C and a relative humidity of 25–35% until the product is cleared by the quality control unit for packaging.

Recent technological advances in lozenge manufacture include a patent (13) that teaches a continuous process for producing dextromethorphan lozenges. A new lozenge cutter apparatus consisting of a drum with an array of cutter elements is described in a 1997 patent (14). A distributing device for feeding lozenges (or other flat products) via a feed conveyor to a "user" machine, such as a packager, is described in a 1993 patent (15).

COMPRESSED TABLET LOZENGES

Commercially, the preparation of lozenges by tablet compression is less important than hard-candy manufacturing techniques. Essentially, lozenge tablets differ from conventional tablets only in their organoleptic and nondisintegrating properties and slower dissolution rate. The associated attributes of pleasant taste with or without matching color, smoothness, and mouthfeel during prolonged dissolution on the tongue, and the physical consideration of holding the tablet in the mouth while swallowing its dissolved components, present unusual

formulation requirements compared to those of tablets intended for swallowing or chewing. The commonly used drugs mentioned previously tend to be bitter, unpleasant tasting compounds. The desire to release these agents slowly in the mouth, in constant contact with the tongue, demands a formulation approach unlike that found in any other dosage form.

Processing and Excipients

Any of the common tablet-processing methods, such as wet granulation, dry granulation, or direct compaction, may be utilized in the production of lozenge tablets. However, because the tablets should dissolve very slowly without disintegration, wet granulation is preferable because it generally provides better control. Through the judicious use of wet binders that retard dissolution, it should be possible to design a formulation having the appropriate dissolution rate.

Formulating for slow dissolution, plus smoothness and good mouthfeel, requires careful excipient selection and appropriate process development to ensure that the controlling variables are dealt with correctly. Several important aspects of lozenge tablet manufacture are critical to all of the desired performance attributes of the finished product. These include assurance of necessary particle size and distribution, maintenance of correct moisture content, and achievement of proper tablet hardness. Process development and scale-up considerations must be thoroughly explored to ensure the establishment of proper specifications for these parameters. As always when the process involves wet granulation, the extent of wetting and the rate and extent of drying must be defined. Overwetting almost certainly produces harder granules that may have poor compressional characteristics, resulting in softer and more friable tablets unacceptable for lozenge application. Because of the lesser degree of particle deformation, such granulations produce tablets having a gritty mouthfeel. Overwetting also leads to longer drying times in order to achieve the desired moisture level or, to a higher moisture level due to failure to compensate through adjustment of the drying cycle.

Lozenge tablets are generally formulated as relatively large-diameter (>12.5 mm), flat-faced (beveled edge), heavy (>700 mg) tablets with high hardness (>15 kp). These physical attributes lead to ease of use in the mouth and contribute to the desired slow dissolution. The formulation factors primarily responsible for controlling dissolution, hardness, and mouthfeel are the presence of a high strength, dissolution-retarding binder and the absence of a disintegrant. Several commonly employed binders (16–18) meet these criteria and provide the

added benefit of delivering a demulcent-like action in the throat.

Gelatin (in the form of a warm 10% aqueous solution) and acacia gum (in the form of an aqueous mucilage) in wet granulation form very hard tablets with slow dissolution. Guar gum, used as an aqueous thixotropic mucilage, exhibits behavior similar to that of acacia gum.

The sugar bases frequently associated with lozenge tablets are sucrose or compressible sugar, dextrose, mannitol, and sorbitol, which are available in special tableting grades from a variety of excipient manufacturers. Generally intended for direct compaction applications, they may also be utilized with the above binders in wet-granulation systems. Lactose, because of its extremely low sweetness (15% of sucrose), is limited for use in lozenges because it would require the addition of an artificial sweetener of sufficient potency to overcome its blandness. Xylitol is relatively sweet and has an advantage in lozenge formulation with respect to its lack of caries production.

Artificial sweeteners are of significant importance to lozenge tablet formulations. As noted above, some sugars may not be sweet enough to mask the bitterness or sourness of many drugs. Presently, there is considerable regulatory disagreement worldwide concerning these materials; some are approved for use in certain countries but not others. Aspartame, asulfame-K, cyclamate, saccharin, and sucralose appear to be technically usable, but their selection requires care from the regulatory perspective. All have potency (sweetness) levels orders of magnitude higher than that of sucrose, permitting the use of very low concentrations (less than 1%) to cover most bitter drugs. Semisynthetic sweeteners, derived from glycyrrhiza, have enjoyed some degree of popularity over the years. They are much sweeter than sucrose, but less sweet than saccharin. It is strongly recommended that the formulator validate the current regulatory acceptance of the intended sweetener prior to its use for a particular product and market country. In addition, a recent report suggests that the common lubricant, magnesium stearate, reduced the effectiveness of cetylpyridinium chloride in compressed tablet lozenges, and that the concentration of this ingredient in such a formulation should be maintained at not more than 0.3% (19).

QUALITY CONTROL

Lozenges require the same quality assurance and control measures as any pharmaceutical dosage form. Because of their unique composition, however, certain additional methods are necessary.

In-Process Testing

In addition to all of the common in-process tests used for all dosage forms, certain specialized procedures are necessary for hard candy lozenges. These include checking the corn syrup and sugar delivery gears, temperature, steam pressure, and the cooking speed of precooker and cooker; analysis of the candy base and its moisture content, and determination of the sugar-to-corn syrup ratio using the dextrose equivalent method, percent-reducing sugar (by reacting candy with copper sulfate and alkaline cupric tartrate and titrating with dextrose solution), pH, cooked candy batch weight, lozenge weight, and lozenge size. The usual in-process tests for compressed tablets apply to lozenge tablets, including particle-size distribution, moisture content, flow, blend uniformity, tablet weight and thickness control, hardness, etc.

Batch-Release Testing

In addition to the usual quality control procedures and the above in-process tests, batch-release testing includes dosage uniformity and a test for grittiness, performed by partially dissolving lozenges under running tap water until one-third to one-half has been removed. No grittiness must be felt when rubbed between thumb and forefinger.

Test procedures that are ordinarily applied to compressed tablets are also employed for lozenge tablets. However, because the lozenge is intended to dissolve slowly in the mouth, typical disintegration and dissolution testing is inappropriate. Lozenges should be nondisintegrating; therefore, there is no need for disintegration testing. Dissolution specifications should be developed on the basis of a minimum and maximum time to physically dissolve, rather than on the basis of minimum percent drug released in the maximum time interval. As with hard candy lozenges, microbial testing may be appropriate, especially when wet granulation has been used in processing the materials and high concentrations of carbohydrates are present.

Stability

For both hard candy lozenges and compressed tablet lozenges, stability considerations extend to areas not usually of concern with other types of tablets. These products should not only conform to chemical and physical specifications, but should also exhibit satisfactory stability of organoleptic attributes. Because lozenges are flavored, flavor stability is important. There is, however, no objective method for measuring flavor stability in a finished dosage form, although GC may be used for chemical analysis of flavor compounds. Even subjective methods such as tasting are difficult because formal taste

panels are needed to acquire reliable data. Subtle changes in flavor with time, although not affecting product performance, may have a significant impact on product market acceptability. This could also be true of minor changes (increase or decrease) in tablet hardness, which could affect dissolution time and therefore acceptability.

PACKAGING

Medicated lozenges, especially those of hard candy base composition, are hygroscopic because of their unique ingredients. In order to be competitive in the marketplace, they need a shelf life longer than 3 years. For these reasons, they are usually packaged individually wrapped in polymeric moisture-barrier material. The wrapped lozenges are placed in a tight or moisture-resistant glass, polyvinyl chloride, or metal container that is overwrapped with cellophane or aluminum foil. This complex, multiple packaging is intended to provide the maximum possible protection from moisture in order to ensure the longest possible shelf life for the product.

SUMMARY

Lozenges are medicated confections designed to locally deliver drug to the mouth and throat. Very widely sold as OTC products, there are also prescription products marketed in the lozenge dosage form. Lozenges are used to deliver local anesthetics, cold and allergy treatments, antimicrobial drugs, antismoking drugs, and anticarcinogenic drugs. They are usually produced in the form of cooked, hard candies based on sugar and corn syrup or sorbitol. They are totally different from any other pharmaceutical dosage form in terms of ingredients and method of manufacture and, therefore, require specialized facilities. Compressed tablet lozenges differ from conventional tablets by requiring materials and methods that provide slow dissolution and drug release.

For these and other reasons, lozenges are produced by few pharmaceutical manufacturers, and represent a very small percentage of total pharmaceutical sales. No significant growth can be currently anticipated in these areas.

REFERENCES

1. *United States Pharmacopeia 24/National Formulary 19*; United States Pharmacopeial Convention, Inc.: Rockville, 1999.

2. *Non-Prescription Drug Products: Formulations & Features*; Knodel, L.C., Ed.; American Pharmaceutical Association: Washington, 1998.
3. Peters, D. Medicated Lozenges. *Pharmaceutical Dosage Forms: Tablets, 2nd Ed.*; Lieberman, H.A., Lachman, L., Schwartz, J.B., Eds.; Marcel Dekker, Inc.: New York, 1989; I, 419–463.
4. Lussi, A.; Portmann, P.; Burhop, B. Erosion on Abraded Dental Hard Tissues by Acid Lozenges: An In Situ Study. *Clin. Oral Investig.* **1997**, *1* (4), 191–194.
5. Grenby, T.H. Dental Properties of Antiseptic Throat Lozenges Formulated with Sugars or Lycasin. *J. Clin. Pharm. Ther.* **1995**, *20* (4), 235–241.
6. Marshall, S. Zinc Gluconate and the Common Cold. Review of Randomized Controlled Trials. *Can. Fam. Physician* **1998**, *44*, 1037–1042.
7. Garland, M.L.; Hagmeyer, K.O. The Role of Zinc Lozenges in Treatment of the Common Cold. *Ann. Pharmacother.* **1998**, *32* (1), 63–69.
8. Richards, R.M.; Xing, J.Z.; Weir, L.F. The Effect of Formulation on the Antimicrobial Activity of Cetylpyridinium Chloride in Candy Based Lozenges. *Pharm. Res.* **1996**, *13* (4), 583–587.
9. Clarke, W.T. The Re-use of Chocolate and Confectionary Rejects. *Manuf. Confec.* **1975**, *55* (6), 37–40.
10. Meiners, A. Incorporating Flavor and Acid in Continuous Manufacturing Systems. *Manuf. Confec.* **1973**, *53* (10), 58–59.
11. Hartel, R.W. Sugar Crystallization in Confectionary Products. *Manuf. Confec.* **1987**, *67* (10), 59–65.
12. Technical Bulletin. *Machines, Production Lines and Process Technologies for the Confectionery Industry*; Robert Bosch Corp.: Bridgman MI, 1998.
13. Dextromethorphan Continuous Lozenge Manufacturing Process, US Patent 5,302,394, April 12, 1994.
14. Lozenge Cutter Apparatus US Patent 5,676,982, Oct 14, 1997.
15. Distributing Device for Feeding Flat Products to a User Machine, US Patent 5,199,546, April 6, 1993.
16. Peck, G.; Bailey, G.E.; McCurdy, V.E.; Banker, G.S. Tablet Formulation and Design. *Pharmaceutical Dosage Forms: Tablets, 2nd Ed.*; Lieberman, H.A., Lachman, L., Schwartz, J.B., Eds.; Marcel Dekker, Inc.: New York, 1989; I, 98–107.
17. Bandelin, F.J. Compressed Tablets by Wet Granulation. *Pharmaceutical Dosage Forms: Tablets, 2nd Ed.*; Lieberman, H.A., Lachman, L., Schwartz, J.B., Eds.; Marcel Dekker, Inc.: New York, 1989; I, 160–164.
18. Mendes, R.W.; Anaebonam, A.O.; Daruwala, J.B. Chewable Tablets. *Pharmaceutical Dosage Forms: Tablets, 2nd Ed.*; Lieberman, H.A., Lachman, L., Schwartz, J.B., Eds.; Marcel Dekker, Inc.: New York, 1989; I, 362–387.
19. Richards, R.M.; Xing, J.Z.; Mackay, K.M. Excipient Interaction with Cetylpyridinium Chloride Activity in Tablet Based Lozenges. *Pharm. Res.* **1996**, *13* (8), 1258–1264.

LIQUID ORAL PREPARATIONS

Jagdish Parasrampuria

Galderma R&D, Cranbury, New Jersey, U.S.A.

INTRODUCTION

Liquid oral preparations are composed of many types of formulations, both aqueous and nonaqueous, including solutions, suspensions, and emulsions. Oral solutions are homogeneous mixtures of one or more solutes dissolved in a suitable solvent or mixture of mutually miscible solvents. In pharmaceutical terms, solutions are defined as "liquid preparations that contain one or more soluble chemical substances, usually dissolved in water and they do not, by reasons of their ingredients, method of preparation, or use, fall into another group of products" (1). Solutions are classified on the basis of physical properties, method of preparation, use, and type of ingredients (Table 1) (2). The distinction among different types of solutions is not always clear because of overlap in definitions. Therefore, in some instances, definitions are not important for commercial products.

APPLICATIONS

Liquid oral preparations are useful for a number of reasons. Patient compliance is often a problem with oral solid dosage forms, especially with young children and the elderly. This refusal to accept medication stems from the difficulties experienced by these age groups in swallowing tablets or capsules. Such difficulties can be overcome by administering the active compound in a palatable liquid form. An oral liquid can be readily administered to children and the elderly who are unable to swallow. Because solutions are homogeneous mixtures, the medication is uniformly distributed throughout the preparation. The dose can be easily adjusted as fractional doses by dilution to meet the needs of the patients. Extracts eliminate the need to isolate the drug in pure form, allowing several ingredients to be administered from a single source (e.g., pancreatic extract) and permit the preliminary study of drugs from natural sources. Some deliquescent and hygroscopic powders are more easily dispensed as liquids. Some drugs that are not tolerated in a concentrated form may be less irritating if dissolved in soothing liquid (3). Occasionally, solutions of drugs such

as potassium chloride are used to minimize adverse effects in the gastrointestinal tract. Because drugs are absorbed in their dissolved state, the rate of absorption of oral dosage forms usually decreases in the following order: aqueous solution > aqueous suspension > tablets or capsules. A drug administered in solution is immediately available for absorption from the gastrointestinal tract and is more rapidly and efficiently absorbed than the same amount of drug administered in a tablet or capsule.

LIMITATIONS

Drug substances in general are less stable in liquid media than in the solid dosage form. Special techniques are required to solubilize poorly soluble drugs. Masking the taste of inherently very bitter drugs is sometimes difficult. Extremely potent drugs with a low therapeutic index cannot be given in an oral liquid dosage form because dosage measurement errors could be made by the patients. As with other oral dosage forms, liquid oral preparations cannot be administered to the unconscious patient.

DESIGN AND FORMULATION

The design of liquid oral solutions involves the combination of ingredients with medicinal agents to enhance the acceptability or effectiveness of the product. The formulation of pharmaceutical liquids requires several considerations: concentration of the drug; solubility of the drug; selection of the liquid vehicle; physical and chemical stability; preservation of the preparation; and appropriate excipients such as buffers, solubilizers, sweetening agents, viscosity-controlling agents, colors, and flavors.

Pharmacists can handle liquid preparations in three ways (4): They may dispense the product in its original container; they may buy the product in bulk and repackage it at the time of dispensation; or they may compound a solution. Compounding may involve nothing more than mixing two marketed products in the manner indicated on the prescription or may require the incorporation of an active ingredient in a logical and pharmaceutically

Table 1 Classification of oral solutions

Type	Description
Syrup	Solutions containing high concentrations of sucrose or other sugars
Elixir	Sweetened solutions containing alcohol as a cosolvent
Spirit	Hydroalcoholic solutions of aromatic or volatile substances
Aromatic Water	Aqueous solutions of aromatic or volatile substances
Tincture	Hydroalcoholic solutions prepared from vegetable materials or chemical substances by dissolution or extraction
Fluid extract	Concentrated alcoholic solutions of animal or vegetable drugs obtained by removal of active constituent by extraction (maceration, percolation)

(From Ref. 2.)

acceptable manner into the aqueous or nonaqueous solvents forming the bulk of the product. Most prescriptions today are dispensed in their original containers. In these cases, the pharmacist depends on the manufacturer to provide a product that is effective, pharmaceutically acceptable, and stable when stored under recommended conditions. Most drug manufacturers attempt to guarantee efficacy by evaluating their products in a scientifically acceptable manner. However, in some instances, such efficacy is relative. For example, cough mixtures marketed by two different manufacturers may contain active ingredients of the same therapeutic class and concentration. It therefore becomes difficult to assess the relative merit of the two products. In such cases, the commercial advantages gained by one over another may be based on other product characteristics. Thus, color, odor, taste, pourability, and homogeneity are important pharmaceutical properties. Hence, the successful design and formulation of liquids, as well as other dosage forms, require both scientific and pharmaceutical acuity (5).

Solubility

Solubility is of prime importance in developing liquid oral solutions. The drug and other dissolved substances should remain solubilized throughout the shelf life of the product. Therefore, the drugs are present in solution at unsaturated concentrations; otherwise, the drug may crystallize as a result of changes in temperature or by “seeding” from other ingredients or particulate matter present. The taste of organic drugs has been shown to be a direct function of aqueous solubility (6). For example, increasing the chain length of clindamycin esters (thus reducing aqueous solubility) dramatically improves the taste (7).

It is also important to determine solubility at refrigeration temperature (2–8°C) to establish the usable concentration in the range of 2–25°C without the risk of

saturation and crystal growth. This procedure is necessary to cover the wide range of temperature conditions to which the product may be exposed during the normal distribution process. If the exact solubility has not been determined, general expressions of relative solubility may be used. These terms are defined in the *United States Pharmacopeia* (USP) and are listed in Table 2 (8).

The solubility of the drug substance is attributable in large part to the polarity of the solvent, often expressed in terms of dipole moment, which is related to the dielectric constant. Solvents with high dielectric constants dissolve ionic compounds (polar drugs) readily by virtue of ion–dipole interactions, whereas solvents with low dielectric constants dissolve hydrophobic substances (nonpolar drugs) as a result of dipole or induced dipole interactions (Van der Waals, London, or Debye forces). This principle is illustrated in Fig. 1 (9). The former is classified as polar solvents, with examples such as water and glycerin; the latter are nonpolar solvents, with example such as oils. Solvents with intermediate dielectric constants are classified as semipolar. The dielectric constants of some solvents are shown in Table 3 (8).

Table 2 USP terms of solubility

Terms	Part of solvent required to dissolve 1 part of solute
Very soluble	<1
Freely soluble	1–10
Soluble	10–30
Sparingly soluble	30–100
Slightly soluble	100–1000
Very slightly soluble	1000–10,000
Practically insoluble, or insoluble	>10,000

(From Ref. 2.)

Liquid Oral Preparations

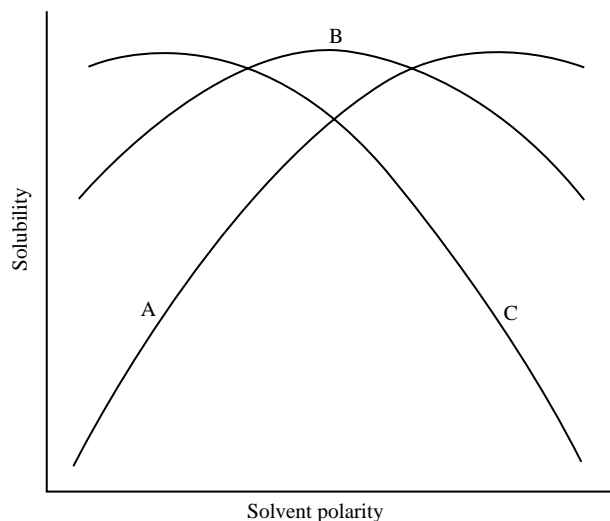


Fig. 1 Influence of solvent polarity on the solubility of drugs: A) Polar drug; B) semipolar drug; C) nonpolar drug. (From Ref. 9.)

A substance often exists in more than one crystalline form, such as chloramphenicol, dehydroepiandrosterone (DHEA), progesterone, sulfathiazole, cortisone or prednisolone, to name a few. Polymorphic transformations are structural differences resulting from different arrangements of molecules in the solid state.

Solubilization Techniques

Sometimes the desired drug concentration in a liquid dosage form cannot be reached because of the drug's low water solubility. Solubilization is the process by which the apparent solubility of a poorly water-soluble substance is increased. Solubilization techniques include addition of a cosolvent, salt formation, prodrug design, complexation,

Table 3 Dielectric constants of solvents at 25°C

Solvent	Dielectric constant
Water	78.5
Glycerin	40.1
Ethanol	24.3
<i>n</i> -Propanol	20.1
Benzyl alcohol	13.1
Polyethylene glycol 400	12.5
Cottonseed oil	3.0

(Modified from Ref. 8.)

Table 4 Comparison of drug solubilization techniques

Method	Approximate range of solubility increase	Reference
Cosolvency	1–1000×	(10)
Salt formation	1–1000×	(11)
Prodrug formation	1–1000×	(12)
Complexation	1–100×	(13)
Micellization	1–50×	(14)

(From Ref. 9.)

particle size reduction, and the use of surface-active agents (micellization). Table 4 shows a comparison of the magnitude of increase in solubility obtained in general with various solubilization techniques (9).

Cosolvency

Cosolvents are defined as water-miscible organic solvents that are used in liquid drug formulations to increase the solubility of poorly water-soluble substances. Cosolvency, then, refers to the technique of using cosolvents. The need to use cosolvents in the formulation of new drugs as solutions remains high, especially with the increasing structural complexity of new therapeutic agents. The importance of using cosolvents in the formulation of a peptide drug has been reported (14).

Cosolvency is highly effective in increasing drug solubility. Advantages include not only the large increases in drug solubility but also simplicity. In the past, ethanol was the most commonly used solvent in oral preparations because of its excellent solvent properties for many nonpolar drugs as well as its favorable taste. Its use is often undesirable, however, in preparations intended for pediatric patients. Ethanol may also accentuate the saline taste of ionic salutes (15). Sorbitol, glycerin, propylene glycol, and several polyethylene glycol polymers are cosolvents that are both useful and acceptable in the formulation of oral liquids. For example, Fig. 2 shows solubilization curves of alprazolam in cosolvent–water systems (16). Cosolvents are used not only to effect solubility of the drug but also to improve the solubility of volatile constituents used to impart a desirable flavor and odor to the product. The primary limitation of cosolvency is the toxicity of most water-miscible solvents that have a high potential for increasing drug solubility. The toxicological properties of a solvent that may limit or eliminate its use in drug formulations include its general toxicity, target organ toxicity, or tissue irritation. Even if found to be relatively nontoxic, a cosolvent can rarely be administered as a neat or 100%

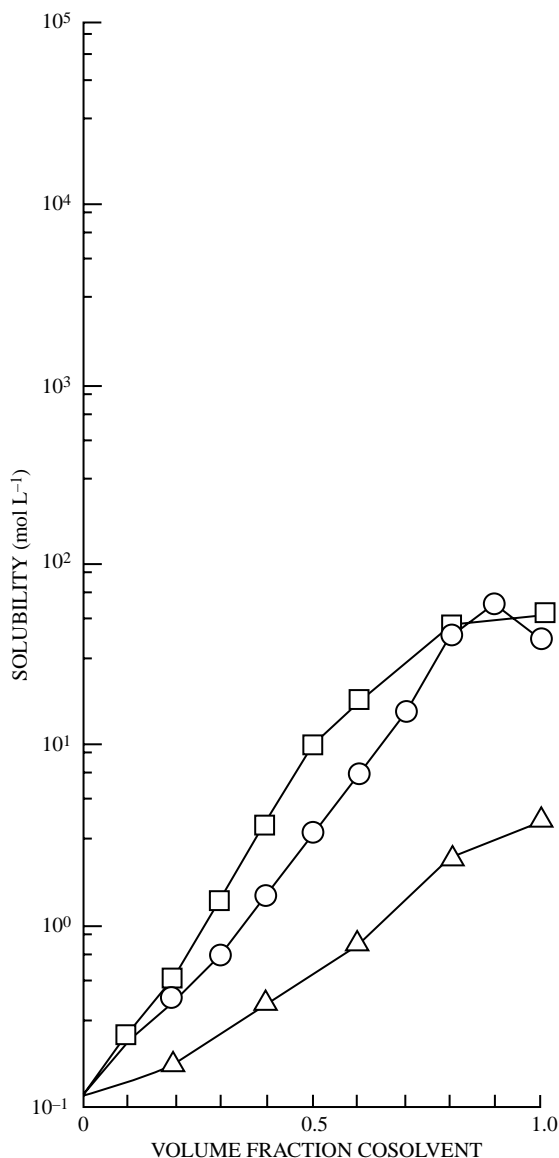


Fig. 2 Solubilization of alprazolam by ethanol (□), propylene glycol (○), and glycerin (Δ). (From Ref. 16.)

solvent because of its poor taste or objectionable odor. Although a cosolvent may increase the solubility of the drug, it may also affect the solubility of other polar or ionic components of the formulation such as buffer materials.

Early formulation work using cosolvency involved an empirical approach for choosing the type and amount of cosolvent for a liquid vehicle. An improvement in the purely empirical approach was the introduction of alligation methods that could be used to reformulate vehicles based on experimental formulation or solubility

data. Moore (17) reported a method using approximate dielectric constants of the pure and mixed solvents. The basic assumption is that all vehicles with the same dielectric constant solubilize a given drug to the same extent. However, the dielectric constant alone is a poor predictor of the degree of drug solubility. Rubino (9) has described various mechanisms and equations that were derived with the objective of establishing a basic understanding of and developing a mathematical approach to interpreting and predicting drug solubility behavior.

Salt formation

Many poorly soluble drugs can be solubilized in salt form (18). The compound (α -(2-Piperidyl) β -3,6-bis(trifluoromethyl)-9-Phenanthrenemethanol, an antimalarial agent, and its hydrochloride salt are both only slightly soluble in water. However, its lactate salt is approximately 200 times more soluble than the hydrochloride. This enhanced aqueous solubility is attributed in part to the decrease in crystal lattice energy, as indicated by a reduction in the melting point. If a particular salt form cannot be isolated because of its very highly solubility, the same end result (i.e., desired aqueous solubility) can be achieved by in situ salt formation. This is accomplished by using an appropriate acid or base to adjust the pH level while formulating the drug product solution.

Prodrug method

The solubility characteristics of a drug can be altered by chemical modification; this is referred to as the "prodrug" approach. The term was first used by Albert (19) for a compound that undergoes biotransformation before eliciting a pharmacological response. This method has been successful in the case of corticosteroids. The solubility of betamethasone in water, for example, is 5.8 mg/100 ml at 25°C. The solubility of its disodium phosphate ester is more than 10 g/100 ml, an increase in solubility greater than 1500-fold. Although methods such as salt and prodrug formation can result in high increases in solubility, they require synthesis of essentially new drug entities as well as additional animal studies to confirm their efficacy and safety. Thus, an undertaking of this magnitude can be justified only if no other reasonable approach is available.

Complexation

Complexation is another means of improving the aqueous solubility of insoluble compounds; it is described by Eq. 1:



where $[D]_s$ = concentration of drug in solution; $[L]_s$ = concentration of ligand in solution; and

$[D_n:L_m]_s$ = concentration of the drug-ligand complex solution.

A complex is an entity formed when two molecules, such as a drug and a solubilizing agent (ligand), are bound by weak forces (e.g., dipole–dipole interaction or hydrogen bonding). For complex formation to occur, drug and ligand molecules must be able to donate or accept a pair of electrons.

Complexation has several advantages, such as the reversibility of the interactions. Dissociation of the complex to the individual reactants occurs rapidly and spontaneously on dilution. Consequently, the biological effects of complexes can be predicted on the basis of knowledge of the pharmacological properties of each of the reactants. Another advantage is the predictability and physical stability of the systems. Because complex formation involves equilibrium attainment, once the necessary data defining the system parameters such as stability constants and solubility properties of the complexes have been gathered, the behavior of the system is totally reproducible and predictable. This is in contrast to the polymorphs and other crystal modifications, which can be thermodynamically unstable; they may undergo time-dependent changes, that may lead to changes in solubility behavior.

The disadvantage of complexation is the presence of the ligand in molar ratios equivalent to and often higher than those of the drug. However, its sensory or pharmacological effects may be unacceptable. The formulator must also take into account detrimental interactions of the ligand with the excipients. Furthermore, the apparent solubility increase realized by complexation compared with other techniques is usually less by an order of magnitude.

Micellization

Micellization has been defined by McBain as the spontaneous passage of poorly water-soluble solute molecules into an aqueous solution of a soap or a detergent in which a thermodynamically stable solution is formed (20). The mechanism for this phenomenon has been studied extensively and involves the property of surface-active agents forming colloidal aggregates known as micelles. When surfactants are added to a liquid at low concentrations, they tend to orient at the air–liquid interface. As additional surfactants are added, the interface becomes fully occupied, and the excess molecules are forced into the bulk of the liquid. At still higher concentrations, the molecules of surfactant in the bulk of the liquid begin to form oriented aggregates or micelles; this change in orientation occurs rather abruptly, and the concentration of surfactant in which it occurs is known as

the critical micelle concentration (CMC). Solubilization is thought to take place by virtue of the solute entrapped in or absorbed onto the micelle. Thus, the ability of surfactant solutions to dissolve or solubilize water-insoluble materials starts at the critical micelle concentration and increases with the concentration of the micelles.

It has been observed that lyophilic surface–active agents with hydrophilic–lipophilic balance (HLB) values above 15 are the best solubilizing agents. The commonly used solubilizing agents in pharmaceutical systems include polyoxyethylene sorbiton fatty acid esters (Tween series), polyoxyethylene monoalkyl ethers (BRIJ and MYRJ series), and sorbiton fatty acid esters (Span series). The choice of solubilizing agents is based on phase-solubility studies in which the solubility of the drug is determined as a function of surfactant concentration; several surfactants are included in these studies. The appropriate surfactant can then be selected on the basis of its efficacy as a solubilizer and its effect on other product characteristics and formulation adjuvants.

STABILITY

Drug substances in general are less stable in liquid media than in the solid dosage form. As a class of formulations, oral liquids are more complex in their composition than are parenterals, and more interactions can occur that might affect the stability of the product. Not only is it necessary to consider the solution stability of the drug, but also the effects on stability caused by excipients such as colorants, flavors, preservatives, solubilizers, thickening agents, and sweetening agents.

Chemical Stability

The techniques for predicting the chemical stability of homogeneous drug systems are well-defined (21, 22). The formulation chemist should consider both the pH-solubility profile and pH-stability profile to select the optimum pH for formulating the liquid oral dosage form. For example, Figs. 3 and 4 show that pH-stability and pH-solubility profiles (23, 24), respectively, for acetazolamide are approximately 4, which is the pH of optimum stability. However, the solubility above pH 7 is much higher because of sodium salt formation.

Physical Stability

Physical instability of liquid formulations involves the formation of precipitates, less soluble polymorphs,

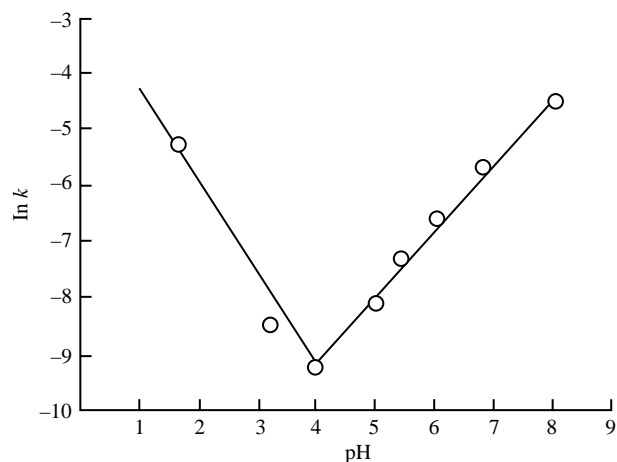


Fig. 3 pH-Stability profile of acetazolamide. (From Ref. 23.)

adsorption of the drug substances onto container surfaces, microbial growth, and product appearance (25). The acceptability of the product is a subjective evaluation and includes properties such as color, odor, taste, and clarity. Dye stability depends on the excipients used in the formulation. For example, FD&C Blue No. 2 has been

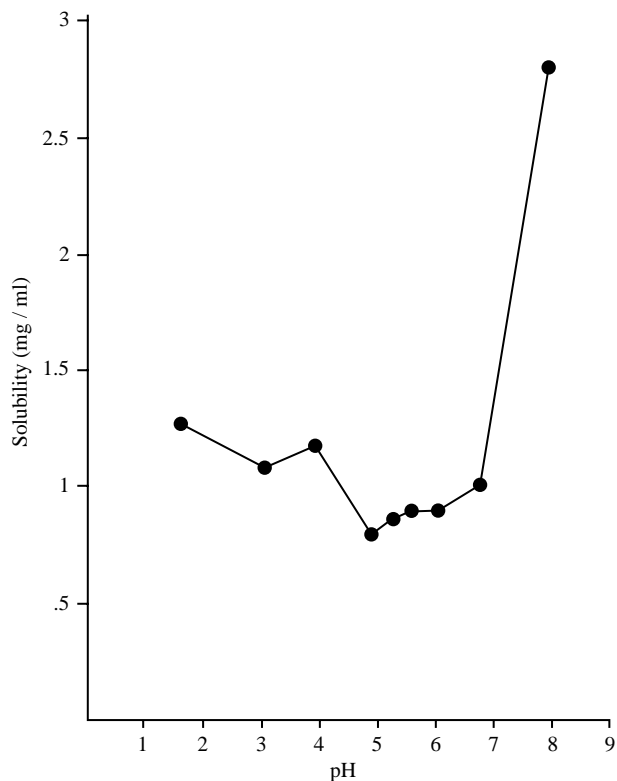


Fig. 4 pH-Solubility profile of acetazolamide. (From Ref. 24.)

found to fade more rapidly in the presence of several sugars (sorbitol, mannitol, dextrose, sucrose, lactose). In addition, trace amounts of impurities and nonionic surfactants such as pluronic F-68 contribute to the fading (26, 27).

Preservatives have been shown to bind to macromolecules (28); for example, the binding of methyl *p*-Hydroxybenzoate to nylon was found to be dependent on the size of the nylon membrane and the concentration of free methylparaben. The complexation of methyl *p*-Hydroxybenzoate with Tween 80 depends on the amount of free methylparaben and the concentration of Tween 80. Similar results have been reported for propyl *p*-Hydroxybenzoate.

COMPONENTS AND ADDITIVES

Sweetening Agents

Sweeteners are indispensable for liquid oral dosage forms. They are used to mask bitter or unacceptable tastes of constituents. Sweetening agents constitute a major portion of the solid content in most liquid oral dosage forms. The most commonly used sweeteners include sucrose, sorbitol, mannitol, liquid glucose, honey molasses, saccharin, and aspartame (Table 5). The types and concentrations of sweeteners for common prescription liquid medications are reported by Hill et al. (29).

Sucrose is the most widely used sweetener with a long history of use. It is a white crystalline powder, soluble in water and alcohol. It inhibits the growth of micro-organisms in solution at a sucrose concentration above 65 wt% by reducing the water-activity coefficient. Official Simple Syrup is an 85% w/v solution of sucrose in water. During the preparation of sucrose solution, care should be taken to avoid charring and caramalization caused by heat. Sucrose is chemically and physically stable in the pH range of 4.0–8.0. It is frequently used in conjunction with sorbitol, glycerin, and other polyols, which reduce its tendency to crystallize. One of the manifestations of the sucrose crystallization is “cap-locking” which occurs when sucrose crystallizes on the threads of the bottle cap and interferes with opening.

Liquid glucose is an extremely viscid substance that imparts both body and sweetness to liquid formulations. It is obtained by the incomplete hydrolysis of starch and consists chiefly of dextrose, dextrans, maltose, and water. It imparts a characteristic odor and flavor to the formulation in a manner similar to that of honey and molasses but to a lesser degree. Although liquid glucose is

Table 5 Sweetening agents

Sweetener	Sweetening power	Chemical structure	Comment
Sucrose	≈ 1		Most commonly used
Saccharin	≈ 500		Unpleasant aftertaste
Sodium cyclamate	≈ 30		Banned
Aspartame	≈ 200		Not very stable in solution

not a pure chemical entity, its method of manufacture can be well-controlled, and batch-to-batch variability is usually not a significant problem. The same is not true of honey and molasses, in which the quality depends on the sources, the time of the year they are produced, and natural factors that cannot be controlled.

Saccharin is a synthetic sweetening agent. It has approximately 500 times the sweetening power of sucrose, depending to some extent on the strength of the solution. The relative sweetening power of sucrose, depending to some extent on the strength of the solution. The relative sweetening power is greatest in dilute solution. Saccharin is a sucrose substitute for diabetics, the obese, and others who do not wish to ingest sucrose. It has no food value and is commonly used in the form of its sodium salt, which is more palatable than saccharin and comparatively free of unpleasant aftertaste. Sodium cyclamate is another synthetic sweetening agent that is approximately 30 times as sweet as sugar. However, its use as an artificial sweetener is no longer permitted in most countries because of the possible toxicity of its metabolite cyclohexylamine.

Aspartame, *N*-L- α -Aspartyl-L-Phenylalanine methyl ester, is 200 times sweeter than sucrose and, unlike saccharin, has no aftertaste. Its aqueous solubility is

adequate for formulation purposes. It is stable in the solid form, but its stability in solution depends on temperature and pH. It hydrolyzes to aspartylphenylalanine and diketopiperazine, with a loss in sweetness by aspartame synergistic with saccharin, sucrose, glucose, and cyclamate. In addition, its taste can be improved by adding sodium bicarbonate, gluconate salts, and lactose (30).

Flavoring Agents

The flavoring of pharmaceuticals is of great importance to liquid dosage forms intended for oral use in that it can mask the disagreeable taste of drugs (31).

Coloring Agents

Although the use of colorants in medicinal products affords no direct therapeutic benefit, the psychological effects have long been recognized. The appearance of clear liquid products depends primarily on the color and clarity of the solution. Many patients rely on color to recognize the prescribed drug and proper dosage. Unattractive medication can be made acceptable to the patient by careful selection of color. Color is usually

chosen according to the flavor of the product (for example, yellow for lemon or red for cherry).

The current list of Food, Drug, and Cosmetics (FD&C)-certified colorants contains both dyes and lakes. The latter are pigments that are insoluble in water and that impart color by dispersing and reflecting light; they are not used for aqueous solutions. In contrast, FD&C dyes are water-soluble and exhibit color by transmitting light. Dyes should be used at the lowest possible concentration required to produce the desired color; higher concentrations can result in a dull color. Most liquid drug products have dye concentrations of less than 0.001%. Because dyes are usually present in trace amounts, they should be dissolved before mixing with the bulk of the formulation. This ensures complete dissolution before further processing. If dye is added directly to the bulk mixing tank, the presence of small amounts of undissolved materials is difficult to determine and can cause problems later during the compounding. Factors influencing the shade and stability of dyes in liquid systems must be carefully considered. Among these are pH level, microbiological activity, exposure to light in the final package, and compatibility of the dye with other ingredients. Because color shades vary greatly at different pH levels, pH control is extremely important. All soluble dyes contain reactive sites, and some dyes may be incompatible with compounds containing polyvalent cations (such as

calcium, magnesium, or aluminum) and precipitate. Certain dyes, such as FD&C Blue No. 2 and FD&C Red No. 3, exhibit poor stability in aqueous solutions and should never be used for coloring aqueous liquid pharmaceutical products (32).

Preservatives

Liquid oral preparations are the most likely of all nonsterile pharmaceutical products to be contaminated by micro-organisms. Most of these preparations are marketed in a multidose form, enhancing the risk of exposure to microbes. The inclusion of sugars and other excipients enriches the preparation with growth-supporting substrates. The manufacturing process also contributes to possible microbiological contamination. In addition to the risks associated with even purified (USP) water, many raw materials of natural origin are used that may contain a large number of viable spores. Many over-the-counter pharmaceutical products have been prepared traditionally at low costs without sufficient care to prevent microbiological contamination. Therefore, it is essential that these preparations are protected against microbiological deterioration by adequate preservation.

Preservatives must fulfill certain criteria for acceptability. The major factors are those of safety and lack of toxicity after oral intake, particularly because liquid

Table 6 Preservatives used in pharmaceutical systems

Preservative	Usual concentration (%)
Acidic	
Phenol	0.2–0.5
Chlorocresol	0.05–0.1
α -Phenylphenol	0.005–0.01
Alkyl esters of <i>p</i> -Hydroxybenzoic acid	0.001–0.2
Benzoic acid and its salt	0.1–0.3
Boric acid and its salts	0.5–1.0
Sorbic acid and its salts	0.05–0.2
Neutral	
Chlorobutanol	0.5
Benzyl alcohol	1.0
β -Phenylethyl alcohol	0.2–1.0
Mercurial	
Thiomersal	0.001–0.1
Phenylmercuric acetate and nitrate	0.002–0.005
Nitromersol	0.001–0.1
Quaternary ammonium compounds	
Benzalkonium chloride	0.004–0.02
Cetylpyridinium chloride	0.01–0.02

(From Ref. 5.)

Table 7 Buffers commonly used in liquid pharmaceutical products

Buffer	pH	Usual concentration (%)
Acetic acid and a salt	3.5–5.7	1–2
Citric acid and a salt	2.5–6	1–3
Glutamic acid	8.2–10.2	1–2
Phosphoric acid salts	6–8.2	0.8–2

(From Ref. 5.)

medications are often administered to children and the elderly. Preservatives must be soluble, stable, microbiologically active, and compatible with the active ingredient as well as with other components of the formulation.

An ideal preservative that meets all the requirements does not exist. The choice must be made on an individual basis, balancing antimicrobial efficacy against safety. Frequently, a combination of two or more preservatives is needed to achieve the desired efficacy. The antimicrobial preservatives are classified into four major types: acidic, neutral, mercurial, and quaternary ammonium compounds (Table 6) (5).

Acidic preservatives are the most widely used for oral preparations, such as the *p*-Hydroxybenzoic acid esters and the salts of benzoic acid. These are adequately soluble in aqueous systems and possess both antifungal and antibacterial properties. Methyl and propyl *p*-hydroxybenzoic acid are often used together in a ratio of 10:1. The use of more than one ester makes possible a higher total preservative concentration owing to the independent solubilities of each and, according to some researchers, maximizes the antimicrobial effect.

The other three classes of preservatives have been widely used in ophthalmic, nasal, and parenteral products, but not frequently in oral liquid preparations. The neutral preservatives are volatile alcohols; their volatility introduces problems of odor and loss of preservative on aging in multidose preparations. The mercurials and quaternary ammonium compounds are excellent preservatives but are subject to incompatibilities (33). Mercurials are readily reduced to free mercury, and the quaternary compounds are inactivated by anionic substances.

Buffers

Changes in the pH level of a preparation may occur during storage because of degradation reactions within the product, interactions with container components, or dissolution of gases and vapors. To avoid these problems,

buffers are added to stabilize pH levels. A suitable buffer system should have adequate capacity to maintain the pH level of the product during storage. It can be based on the pH profile of the drug in solution. Commonly used buffer systems are acetates, citrates, phosphates, and glutamates (Table 7). Although buffers ensure pH stability, the buffer system can affect other properties such as solubility and kinetics. Buffers can act as general-acid or general-base catalysts and cause degradation of the drug substance. The ionic-strength contributions of the buffer systems can also affect stability. Therefore, the effect of buffer species should be studied before selecting any buffer system.

Antioxidants

Many drugs in solution are subject to oxidative degradation. Such reactions are mediated by free radicals or molecular oxygen and often involve the addition of oxygen or the removal of hydrogen. Drugs possessing favorable oxidation potential are especially vulnerable to degradation. Agents with an oxidation potential lower than that of the drug in question is called antioxidants. They are added to solutions alone or in combination with a chelating agent or other antioxidants and function by being preferentially oxidized and gradually consumed or by blocking an oxidative chain reaction where they are not consumed. Sulfites are the most common antioxidants in aqueous solutions. Irrespective of which sulfite salt is added, the antioxidant moiety depends on its final concentration and the final pH level of the formulation; metabisulfite is used at low pH. A single antioxidant may not provide complete protection. Certain compounds (e.g., ascorbic and citric acids) have been found to act as synergists, increasing the effectiveness of antioxidants, particularly those that block oxidative reactions. Frequently, chelating agents such as ethylenediaminetetraacetic acid derivatives (EDTA) are used in formulations containing trace amounts of heavy metals that would otherwise catalyze oxidative reactions.

Viscosity-Controlling Agents

It is sometimes desirable to increase the viscosity of a liquid to provide or to improve palatability or pourability. This can be achieved by viscosity-controlling agents such as polyvinylpyrrolidone or carboxymethylcelluloses. These compounds give aqueous solutions that are stable over a wide pH range. Methylcellulose and carboxymethylcellulose are available in a number of different

viscosity grades. The latter may be used in solutions containing up to 50% alcohol without precipitating. However, care should be taken to avoid precipitation, which occurs with the insoluble salts to a number of multivalent metal ions such as Al^{3+} , Fe^{3+} , and Ca^{2+} . Methylcellulose polymers do not form insoluble salts with metal ions but can be salted out when the concentration of electrolytes or other dissolved materials exceeds certain limits. These limits may vary from 2 to 40%, depending on the electrolyte and the type of methylcellulose involved. Viscosity-inducing polymers should be used with caution. They are known to form molecular complexes with a variety of organic and inorganic compounds. It is conceivable that highly viscid systems that resist dilution by gastrointestinal fluid may impede drug release and absorption.

MANUFACTURE

The preparation of liquid drug preparations involves choosing the ingredients (using principles already addressed) and the manufacturing equipment. The basic principles involved are the same, regardless of the materials and the quantities involved. Carstensen and Mehta report the scaling-up of solution dosage forms (34), including heating, agitation, and clarification along with basic equations and calculations.

Liquid processing lends itself to computer-controlled automation. A few pharmaceutical firms have already instituted automated or semiautomated processes for several large-selling liquid products (35).

Raw Materials

Although purified water (USP) is required in all operations, it is particularly important in liquid manufacturing. If de-ionized and other water-treatment equipment is used, special attention must be given to routine microbiological and chemical testing. Storage tanks for glycerin and propylene glycol should be constructed to facilitate examination as well as cleaning.

Equipment

Simple solutions are most straightforward to scaleup but require tanks of adequate size and suitable mixing capacity (35). Most equipment should have heating and cooling capabilities for rapid dissolution of formulation components. Adequate transfer systems and filtration equipment are required, but they must be monitored to

ensure that they can clarify the product without removing active or adjuvant ingredients. All equipment must be made of suitable, nonreactive sanitary materials and be designed and constructed to facilitate easy cleaning. Liquid pharmaceutical processing tanks, kettles, pipes, mills, filter housings, and so forth are most frequently fabricated from stainless steel. Of the three types commonly used in the industry (304, 308, and 316), type 316 is most often used because it is the least reactive. Stainless steel is virtually nonreactive but may react with some acidic pharmaceutical liquids (36). This problem can be minimized by treating the stainless steel with acetic acid or nitric acid solution to remove surface alkalinity. This process, known as passivation, may be needed periodically. For example, if an alkaline cleaning agent is used between batches of a reactive product, passivation may be required before the subsequent batch can be prepared.

Interaction with metallic surfaces can be minimized by using polytetrafluoroethylene (Teflon) liners. Although Teflon is inert, these liners have the potential disadvantages of cracking, breaking, flaking, and peeling, with resulting product contamination.

A valuable and practical discussion on the design of piping, valves, mixers, pumps, and controls to produce high-quality liquid products is given by FitzSimon (37).

Methods of Preparation

Dilute solutions of rapidly dissolving materials are prepared by adding the solute to the solvent and agitating until the solution is homogeneous. Heat may be required for more concentrated solutions or when the solute is slow to dissolve. Excipients are usually added in a specified order to increase the rate of dissolution and to facilitate a rapid approach to equilibrium. For this reason, menthol and flavors are charged as alcoholic solutions to the batch. Solutes present in small concentrations, particularly dyes and other intensely colored materials, should be dissolved before mixing with the main portion of the batch to ensure complete dissolution. If the solutes were charged directly to the bulk mixing tank, it would be difficult to detect small amounts of undissolved material at the bottom of the tank. As a rule, complete solution should be confirmed at every stage in the manufacture of a homogeneous liquid. In the laboratory, liquids are usually measured by volume. However, in large-scale production, gravimetric means of measurement are used. For this reason, all liquids components of the formulation are expressed in units of both volume and weight.

Solutions must be filtered and clarified; this stage of the process is called "polishing." A highly polished solution requires the removal of particulate matter down to at least 3 μm . Filters used in the manufacturing, processing, or packing of liquid drug products intended for human use should not release fibers. If a fiber-releasing filter needs to be used, it must be followed by a nonfiber-releasing filter of optimum porosity. Filter aids are commonly used to improve clarity and increase the flow rate, thus decreasing filtration time. The amount and type of filter aid must be determined during the development of the product; the amount usually does not exceed 0.5 g/L. Examples of filter aids are diatomaceous earth, carbon, expanded perlite, and cellulose.

Filling and Sealing

A liquid may be removed from a bulk container in portions to individual-dose containers more easily and uniformly than a solid. Certain fundamental features are found on all machines used for filling containers with liquid. A means is used for repeatedly forcing a measured volume of the liquid through the orifice of a delivery tube designed to enter the constricted opening of a container. The size of the delivery tube is governed by the opening in the container, the viscosity and density of the liquid, and the desired speed of delivery. The tube must freely enter the neck of the container and deliver the liquid deep enough to permit air to escape without sweeping the entering liquid onto the neck or out of the container; the tube should have the maximum possible diameter. Excessive delivery force causes splashing and troublesome foaming if the liquid has a low surface tension.

Small volumes of liquids (usually for pediatric use) are delivered by the stroke of the plunger of a syringe, which forces the liquid through a two-way valve that provides for an alternative filling of the syringe from a reservoir and delivery to a container. For heavy, viscous liquids, a sliding piston valve provides more positive action. A drop of liquid normally hangs at the tip of the tube after a delivery, which is removed by a retraction device. Filling machines should be designed so that the parts through which the liquid flows can be easily dismantled for cleaning. These parts should be constructed of nonreactive materials such as stainless steel. Small-volume filling machines are designed to deliver volume precisely. Syringes are typically also made of stainless steel. The stroke of the syringe can be repeated precisely; therefore, once a particular setting has been calibrated for a delivery, high precision is possible. The precision can be affected by certain operating factors such as the speed of delivery, the

uniformity of speed, the expansion of rubber tubing connecting the valve to the delivery tube, and the rapidity of the action of the valves.

Large-volume filling does not normally require the precision required for small volumes. Therefore, bottles of solution are usually filled by gravity, pressure, or vacuum devices. Gravity filling is relatively slow but simple. The liquid reservoir to a shut-off device at the filling line, which is usually hand-operated; the bottles are filled to a graduation mark. The pressure-pump filler is often operated semiautomatically and differs from the gravity filler principally in that the liquid is under pressure. It is usually equipped with an overflow tube connected to a receiver to prevent excess filling.

Vacuum filling is commonly used for large liquid volumes because it is easily adapted to automation. A vacuum is produced in a bottle when the nozzle gasket makes a seal against the lip of the bottle to be filled. The vacuum draws the liquid from a reservoir through the delivery tube into the bottle. When the liquid level reaches the level of an adjustable overflow tube, the seal is mechanically loosened, and the vacuum is released. Any liquid that has been drawn into the vacuum line is collected in a receiver and returned to the reservoir.

Accuracy and precision of filling with liquids vary with the method. Therefore, a method is selected to provide the degree of accuracy and precision required. A slight excess is required in each container to provide for the loss that occurs at the time of administration resulting from adherence to the wall of the container. The danger of overdose as well as economic factors limits the amount of excess desirable in a given container.

Highly viscous solutions require specially designed equipment. To obtain a reasonable flow rate, high pressure must be applied, or containers with large openings must be used to permit the entry of large delivery tubes. Sometimes jacketed reservoir tanks can be used to raise the temperature of the product and thereby lower its viscosity.

A problem common to all types of machines that fill containers with liquid but that is particularly bothersome in high-speed automatic equipment is excessive foaming. Foaming during the filling operation can be reduced by using filling equipment that minimizes product turbulence, closed-system filling to limit the introduction of air or other gases that cause foaming, mechanical defoaming devices, or reduction in the speed of the filling line. All these methods introduce considerable engineering and production difficulties. It would be preferable to formulate the product with careful consideration of the problems that might eventually be encountered in large-scale production and high-speed filling operations.

A microbial survey should be performed on all packaging materials that come in contact with the product to ensure the absence of microbial contamination. Attention must also be given to details during packaging operations. For example, on small-volume orders, bottle closures or tips for plastic squeeze-spray containers are often placed on the product by hand. This procedure can be a source of microbial contamination unless operators use gloves that are sterilized and disinfected periodically during use.

REFERENCES

1. Ansel, H.C. Popovich, N.G. *Pharmaceutical Dosage Forms and Delivery Systems*; Lea & Febiger: Philadelphia, 1990; 196–224.
2. *United States Pharmacopoeia XXIV-National Formulary XIX*; United States Pharmacopoeial Convention, Inc.: Rockville, MD, 2000; 2107–2118.
3. Ecanow, B. Liquid Medications. *Dispensing of Medication*; King, R.E., Ed.; Mack Publishing Co.: Easton PA, 1984; 100–139.
4. Nairn, J.G. Solutions, Emulsions, Suspensions, and Extractives. *Remington's Pharmaceutical Sciences*; Genaro, A., Ed.; Mack Publishing Co.: Easton, PA, 1990; 1492–1517.
5. Boylan, J.C. Liquids. *The Theory and Practice of Industrial Pharmacy*; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea & Febiger: Philadelphia, 1996; 457–458.
6. Valvani, S.C. Yalkowsky, S.H. Solubility and Partitioning in Drug Design. *Physical Chemical Properties of Drugs*; Yalkowsky, S.H., Sinkula, A.A., Valvani, S.C., Eds.; Marcel Dekker, Inc.: New York, 1980; 201–229.
7. Sinkula, A.A. Morozowich, W. Rowe, E.L. *J. Pharm. Sci.* **1973**, 62, 1106–1111.
8. DeLuca, P.P. Boylan, J.C. Formulation of Small Volume Parenterals. *Pharmaceutical Dosage Forms: Parenteral Medications*; Avis, K.E., Lachman, L., Lieberman, H.A., Eds.; Marcel Dekker, Inc.: New York, 1986; 1, 139–201.
9. Rubino, J.T. Cosolvents and Cosolvency. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1990; 3, 375–398.
10. Martin, A.; Swarbrick, J. Cammarata, A. *Physical Pharmacy*; Lea & Febiger: Philadelphia, 1983; 577–579.
11. Melby, J.C. Cyr, M., St. Metabolism **1961**, 10, 75–82.
12. Repta, A.J. Alternation of Apparent Solubility Through Complexation. *Techniques of Solubilization of Drugs*; Yalkowsky, S.H., Ed.; Marcel Dekker, Inc.: New York, 1981; 135–158.
13. Florence, A.T. Drug Solubilization in Surfactant Systems. *Techniques of Solubilization of Drugs*; Yalkowsky, S.H., Ed.; Marcel Dekker, Inc.: New York, 1981 15–90.
14. Sanders, L.M. Abstracts of American Pharmaceutical Associations 133rd Annual Meeting **1986**, 16, 14–16.
15. Pernarowski, M. Solutions, Emulsions, Suspensions and Extractives. *Remington's Pharmaceutical Sciences*; Osol, A., Hoover, J.E., Eds.; Mack Publishing Co.: Easton PA, 1970; 1488–1489.
16. Yalkowsky, S.H. Roseman, T.J. Solubilization of Drugs by Cosolvents. *Techniques of Solubilization of Drugs*; Yalkowsky, S.H., Ed.; Marcel Dekker, Inc.: New York, 1981; 91–134.
17. Moore, W.E. *J. Am. Pharm. Assoc. Sci. Ed.* **1958**, 47, 855–857.
18. Agharkar, S. Lindenbaum, S. Higuchi, T. *Am. Pharm. Sci.* **1976**, 65, 747–749.
19. Albert, A. *Nature* **1958**, 182, 421–423.
20. McBain, J.W. *Advances in Colloid Science*; Interscience: New York, 1942; 1.
21. Streng, W.H. *Drug Develop. Indus. Pharm.* **1985**, 11, 1869–1888.
22. Cartensen, J.T. *Drug Stability: Principles and Practices*; Marcel Dekker, Inc.: New York, 1990; 15–108.
23. Parasrampur, J. Gupta, V.D. *J. Pharm. Sci.* **1989**, 78, 855–857.
24. Parasrampur, J. University of Houston: Houston, 1989, 94.
25. Shami, E.G. Bernardo, P.D. Rattie, E.S. Ravin, L.J. *J. Pharm. Sci.* **1972**, 61, 1318–1321.
26. Kuramoto, R. Lachman, L. Cooper, J. *J. Am. Pharm. Assoc. Sci. Ed.* **1958**, 47, 175–180.
27. Scott, N.W. Goudie, A.J. Huettman J. *Am. Pharm. Assoc. Sci. Ed.* **1960**, 49, 467–472.
28. Patel, N.K. Kostenbauder, H.D. *J. Am. Pharm. Assoc. Sci. Ed.* **1958**, 47, 289–293.
29. Hill, E.M. Flaitz, C.M. Frost, G.R. *Am. J. Hosp. Pharm.* **1988**, 45, 135–142.
30. Beck, C.I. Application Potential for Aspartame in Low Calorie and Dietetic Food. *Low Calorie and Dietary Food*; Dwivedi, B.K., Ed.; CRC Press, Inc.: Boca Raton, FL, 1978; 68.
31. Adjei, A. Doyle, R. Reiland, T. Flavors and Flavor Modifiers. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1992; 6, 101–139.
32. Woznicki, E.J. Schoneker, D.R. Coloring Agents for Use in Pharmaceuticals. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1990; 3, 65–100.
33. Lachman, L. *Bull. Parent. Drug Assoc.* **1968**, 22, 127–144.
34. Carstensen, J.T. Mehta, A. *Pharm. Technol.* **1982**, 6, 64–77.
35. Yelvig, M. *Pharm. Technol.* **1984**, 8, 47–56.
36. Narurhar, A.N. Sheen, P.C. Corrosion in Pharmaceutical Processing. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1990; 3, 353–360.
37. FitzSimon, R. *Drug Dev. Commun.* **1976**, 2, 1–31.

MICROSPHERE TECHNOLOGY AND APPLICATIONS

Diane J. Burgess

University of Connecticut, Storrs, Connecticut, U.S.A.

Anthony J. Hickey

The University of North Carolina, Chapel Hill, North Carolina, U.S.A.

INTRODUCTION

The range of techniques for the preparation of microspheres offers a variety of opportunities to control aspects of drug administration. The term “control” includes phenomena such as protection and masking, reduced dissolution rate, facilitation of handling, and spatial targeting of the active ingredient. This approach facilitates accurate delivery of small quantities of potent drugs; reduced drug concentrations at sites other than the target organ or tissue; and protection of labile compounds before and after administration and prior to appearance at the site of action.

The characteristics of microspheres containing drug should be correlated with the required therapeutic action and are dictated by the materials and methods employed in the manufacture of the delivery systems.

The behavior of drugs *in vivo* can be manipulated by coupling the drug to a carrier particle. The clearance kinetics, tissue distribution, metabolism, and cellular interactions of the drug are strongly influenced by the behavior of the carrier. Exploitation of these changes in pharmacodynamic behavior may lead to enhanced therapeutic effect. However, an intelligent approach to therapeutics employing drug-carrier technology requires a detailed understanding of the carrier interaction with critical cellular and organ systems and of the limitations of the system with respect to formulation procedures and stability. A variety of agents have been used as drug carriers, including immunoglobulins, serum proteins, liposomes, microspheres, nanoparticles, microcapsules, and even cells such as erythrocytes.

Antineoplastic drugs (1–5), narcotic antagonists (6), steroid hormones (7, 8), luteinizing hormone releasing hormone analogs (9, 10), elastase (11), and other macromolecules (12) have been incorporated into microspheres. In addition, vaccines, living cells, and tissues have been encapsulated.

DEFINITION AND GENERAL DESCRIPTION

Microspheres can be defined as solid, approximately spherical particles ranging in size from 1 to 1000 μm . They are made of polymeric, waxy, or other protective materials, that is, biodegradable synthetic polymers and modified natural products such as starches, gums, proteins, fats, and waxes. The natural polymers include albumin and gelatin (13, 14); the synthetic polymers include polylactic acid and polyglycolic acid (15, 16).

The solvents used to dissolve the polymeric materials are chosen according to the polymer and drug solubilities and stabilities, process safety, and economic considerations. Substances can be incorporated within microspheres in the liquid or solid state during manufacture or subsequently by absorption. Fig. 1 shows two types of microspheres: Microcapsules, where the entrapped substance is completely surrounded by a distinct capsule wall, and micromatrices, where the entrapped substance is dispersed throughout the microsphere matrix.

Microspheres are small and have large surface-to-volume ratios. At the lower end of their size range they have colloidal properties (17). The interfacial properties of microspheres are extremely important, often dictating their activity. In fact, the principle of microsphere manufacture depends on the creation of an interfacial area, involving a polymeric material that will form an interfacial boundary and a method of cross-linking to impart permanency. The methods of manufacturing described later are by no means comprehensive and the reader should bear in mind that if the aforementioned criteria are adhered to, the only limitation to the manufacture of microspheres is the researcher's imagination.

HISTORICAL AND CONTEXTUAL PERSPECTIVE

The concept of packaging microscopic quantities of materials within microspheres dates back to the 1930s and

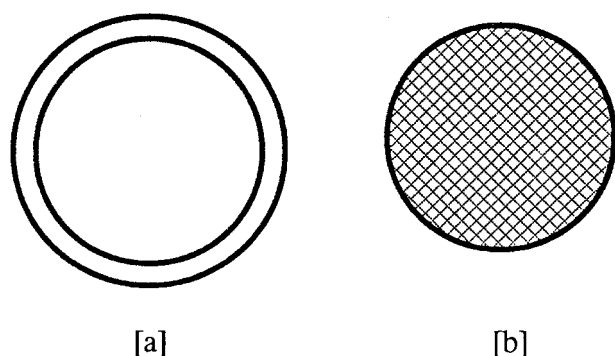


Fig. 1 Schematic diagram illustrating microspheres. (a) Microcapsule consisting of an encapsulated core particle, and (b) Micromatrix consisting of homogeneous dispersion of active ingredient in particle.

the work of Bungenberg de Jong and coworkers (18) on the entrapment of substances within coacervates. The first commercial application of encapsulation was by the National Cash Register Company for the manufacture of carbonless copying paper (19). The technology and applications have advanced over the last several decades. This technology is used by the agricultural, food, household products, medical, graphics, and cosmetics industries.

The potential use of microspheres in the pharmaceutical industry has been considered since the 1960s (20–22) for the following applications:

- Taste and odor masking
- Conversion of oils and other liquids to solids for ease of handling
- Protection of drugs against the environment (moisture, light, heat, and/or oxidation) and vice versa (prevention of pain on injection) (23)
- Delay of volatilization
- Separation of incompatible materials (other drugs or excipients such as buffers)
- Improvement of flow of powders
- Safe handling of toxic substances
- Aid in dispersion of water-insoluble substances in aqueous media, and
- Production of sustained-release, controlled-release, and targeted medications (24–27)
- Reduced dose dumping potential compared to large implantable devices

Microencapsulation has also been used medically for the encapsulation of live cells and vaccines. Biocompatibility can be improved by the encapsulation of artificial cells and biomolecules such as peptides, proteins, and

hormones (28, 29), which can prevent unwanted immunological reactions that would lead to inactivation or rejection. Microspheres are used for isolating materials until their activity is needed. The biotechnology industry employs microspheres to contain organisms and their recombinant products to aid in the isolation of these products (30).

PHARMACEUTICAL APPLICATIONS

A number of pharmaceutical microencapsulated products are currently on the market, such as aspirin, theophylline and its derivatives, vitamins, pancrelipase, antihypertensives, potassium chloride, progesterone, and contraceptive hormone combinations (31).

Microencapsulated KCl (Micro-K, R.H. Robins, Richmond, VA) is used to prevent gastrointestinal complications associated with potassium chloride. The dispersibility of the microcapsules and the controlled release of the ions minimize the possibility of local high salt concentrations, which could result in ulceration, hemorrhage, or perforation. Microspheres have also found potential applications as injection (32, 33) or inhalation (34–37) products. The number of commercially available products does not reflect the amount of research that has been carried out in this area, nor the benefits that can be achieved using this technology. Economic considerations have been a key factor in determining the number of pharmaceutical microencapsulated products. Most encapsulation processes are expensive and require significant capital investment for equipment. An exception is pan or spray coating and spray drying, since the necessary equipment may already be available within the company. An additional expense is due to the fact that most microencapsulation processes are patent protected.

OTHER APPLICATIONS

Applications of microencapsulation in other industries are numerous. The best known microencapsulated products are carbonless copying paper, photosensitive paper, microencapsulated fragrances, such as “scent-strips” (also known as “snap-n-burst”), and microencapsulated aromas (“scratch-n-sniff”). All of these products are usually prepared by gelatin–acacia complex coacervation. Scratch-n-sniff has been used in children’s books and food and cosmetic aroma advertising (38). Microcapsules are also extensively used as diagnostics, for example,

temperature-sensitive microcapsules for thermographic detection of tumors (39).

In the biotechnology industry microencapsulated microbial cells are being used for the production of recombinant proteins and peptides (30). The retention of the product within the microcapsule can be beneficial in the collection and isolation of the product. Encapsulation of microbial cells can also increase the cell-loading capacity and the rate of production in bioreactors. Smaller microcapsules are better for these purposes; they have a larger surface area that is important for the exchange of gases across the microcapsule membrane. Microcapsules with semipermeable membranes are being used in cell culture (40). A feline breast tumor line, which was difficult to grow in conventional culture, has been successfully grown in microcapsules (41). Microencapsulated activated charcoal has been used for hemoperfusion (29). Paramedical uses of microcapsules include bandages with microencapsulated anti-infective substances (25). A unique application of microencapsulation technology is for feeding organisms. Sea bass larvae have been fed with all-protein microcapsules or with microcapsules containing lipids to supplement their diet (42).

MICROSPHERE MANUFACTURE

The most important physicochemical characteristics that may be controlled in microsphere manufacture are:

- Particle size and distribution
- Polymer molecular weight
- Ratio of drug to polymer
- Total mass of drug and polymer

Each of these can be related to the manufacture and rate of drug release from the systems. The following discussion presents methods of manufacture of coated or encapsulated systems, referred to as microcapsules, and matrix systems containing homogeneously distributed drug, referred to as micromatrices.

Wax Coating and Hot Melt

Wax may be used to coat the core particles, encapsulating drug by dissolution or dispersion in the molten wax. The waxy solution or suspension is dispersed by high speed mixing into a cold solution, such as cold liquid paraffin. The mixture is agitated for at least one hour. The external phase (liquid paraffin) is then decanted and the microcapsules are suspended in a nonmiscible solvent, and allowed to air dry. Multiple emulsions may also be formed (43). For example,

a heated aqueous drug solution can be dispersed in molten wax to form a water-in-oil emulsion, which is emulsified in a heated external aqueous phase to form a water-in-oil-in-water emulsion. The system is cooled and the microcapsules collected. For highly aqueous soluble drugs, a nonaqueous phase can be used to prevent loss of drug to the external phase (44). Another alternative is to rapidly reduce the temperature when the primary emulsion is placed in the external aqueous phase.

Wax coated microcapsules, while inexpensive and often used, release drug more rapidly than polymeric microcapsules. Carnauba wax and beeswax can be used as the coating materials and these can be mixed in order to achieve desired characteristics (45). Wax-coated microcapsules have been successfully tableted. Small aerosol particles, 1–5 μm in diameter, have been condensation coated from a vapor of a fatty acid or paraffin wax (46, 47). These particles have been shown to exhibit reduced dissolution rates in vitro, corresponding to reduced absorption rates following deposition in the lungs of Beagle dogs.

Polyanhydrides (48) have been chosen for the preparation of microspheres because of their degradation by surface erosion into apparently nontoxic small molecules (49, 50). The mixture of polymer and active ingredient is suspended in a miscible solvent, heated 5°C above the melting point of the polymer and stirred continuously. The emulsion is stabilized by cooling below the melting point until the droplets solidify.

Spray Coating and Pan Coating

Spray coating and pan coating employ heat-jacketed coating pans in which the solid drug core particles are rotated and into which the coating material is sprayed. The core particles are in the size range of micrometers up to a few millimeters. The coating material is usually sprayed at an angle from the side into the pan. The process is continued until an even coating is completed. This is the process typically used to coat tablets and capsules.

Coating a large number of small particles may provide a safer and more consistent release pattern than coated tablets. In addition, several batches of microspheres can be prepared with different coating thicknesses and mixed to achieve specific controlled release patterns.

The Wurster process, a variation of the basic pan coating method, is an adaptation of the fluid-bed granulator (51, 52). The solid core particles are fluidized by air pressure and a spray of dissolved wall material is applied from the perforated bottom of the fluidization chamber parallel to the air stream and onto the solid core particles. Alternatively, the coating solution can be

sprayed from the top or the sides into an upstream of fluidized particles. This adaptation allows the coating of small particles (53). The fluidized-bed technique produces a more uniform coating thickness than the pan-coating methodology. Problems can arise with inflammable organic solvents because of the high risk of explosion in the enclosed fluidizer chamber. Explosion proof units have been designed; however, over the past two decades aqueous coating solutions are being used more and more.

Examples of aqueous coating solutions include water-soluble low molecular weight cellulose ethers (54, 55), emulsion polymerization latexes of polymethacrylates (56), and dispersions of water-insoluble polymers such as ethylcellulose in the form of pseudolatex (57). These solvent-free coating solutions provide a range of different coatings from fast disintegrating isolating layers to enteric and sustained-release coatings. Lehmann has reviewed different commercial methods, the conditions required for coating, and various coating formulas including illustrations of the types of equipment used (58).

Coacervation

Coacervation is the simple separation of a macromolecular solution into two immiscible liquid phases, a dense coacervate phase, which is relatively concentrated in macromolecules, and a dilute equilibrium phase (18). Coacervates may be described as liquid crystals and mesophases. In the presence of only one macromolecule, this process is referred to as simple coacervation. When two or more macromolecules of opposite charge are present, it is referred to as complex coacervation (18). Simple coacervation is induced by a change in conditions, which results in dehydration of the macromolecules. This may be achieved by the addition of a nonsolvent, the addition of microions, or a temperature change, all of which promote polymer-polymer interactions over polymer-solvent interactions. Complex coacervation is driven by electrostatic interactive forces between two or more macromolecules (59).

Bungenberg de Jong et al. (60) first showed that solid particles could also be entrapped in coacervate systems. On phase separation by simple or complex methods tiny coacervate droplets are formed, which sediment or coalesce to form a separate coacervate phase. The coacervate forms around any core material that may be present, such as drug particles (Fig. 2). Agitation of the coacervate system can prevent coalescence and sedimentation of the droplets, which can be cross-linked to form stable microcapsules by addition of an agent, such as

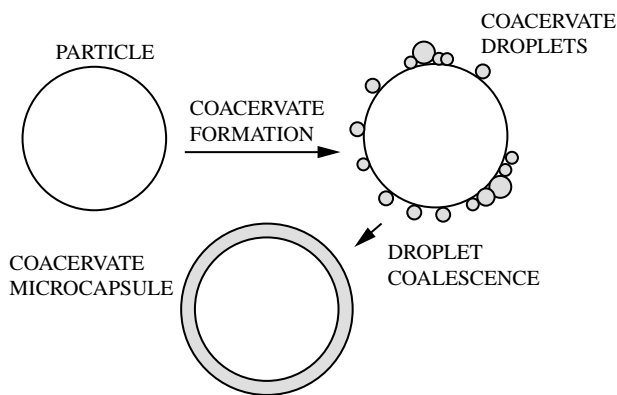


Fig. 2 Schematic diagram of the formation of a coacervate around a core material.

glutaraldehyde, or the application of heat (14, 24, 61, 62). Drug microencapsulation by coacervation has been reviewed by Madan (62) and by Nixon (24).

The large number of variables involved in complex coacervation (pH, ionic strength, macromolecule concentration, macromolecule ratio, and macromolecular weight) (18, 63) affect microcapsule production, resulting in a large number of controllable parameters. These can be manipulated to produce microcapsules with specific properties. Complex coacervate microcapsules have been formulated as suspensions or gels (64), and have been compounded within suppositories (65, 66) and tablets (66, 67).

Although many successful coacervate microencapsulation systems have been prepared, coacervate microcapsules have a number of limitations. They can be produced only at specific pH values, they require stabilization by cross-linking agents or heat, and the retention of the encapsulant depends on the extent of cross-linking. The pH limitation can be overcome to some extent by the addition of water-soluble nonionic polymers, such as polyethylene oxide or polyethylene glycol (68, 69). The presence of a small amount of these polymers allows microencapsulation to occur over an expanded pH range. For example, the pH range for coacervation of gelatin and acacia can be extended from pH 2.6–5.5 (63) to pH 2–9 (68). In addition, these polymers induce simple coacervation (68), as has been shown for macromolecules such as gelatin, carboxymethylcellulose, and ethylene-maleic anhydride copolymer. The pH range for simple coacervation is also expanded in the presence of these water-soluble nonionic polymers. For example, the pH range for simple coacervation of gelatin can be increased from only pH values close to the isoelectric point to the pH range of 5.5–9.5 (68).

Cross-linking of coacervates is necessary to stabilize coacervate emulsion droplets and hence form microcapsules. Both chemical cross-linking agents and the application of heat may be harmful to the encapsulant materials, such as thermolabile and chemically labile drugs and live cells. A stable coacervate system, formed without the use of chemical cross-linking agents or the application of heat, has been developed by Burgess and Singh (70, 71). This system is potentially useful for the delivery of protein and polypeptide drugs and other materials unable to withstand cross-linking procedures.

Calcium Alginate Microcapsules

Dropping or spraying a sodium alginate solution into a calcium chloride solution produces microcapsules. The divalent calcium ions cross-link the alginate, forming gelled droplets. These gel droplets can be permanently cross-linked by addition to a polylysine solution. Lim and Sun (72) developed this method for the encapsulation of live cells. Variations on this method with different polymers have been developed. Chitosan is a preferred polymer, because it has a better biocompatibility than alginate (73). Traditionally alginate beads were formed by dropping the alginate solution into the calcium chloride with a fine-bore pipette. However, the droplets were relatively large, because the drops do not fall until they reach a critical mass. Smaller droplets can be formed by using a pump to force the alginate through the pipette (72), a vibration system to help remove the drops from the end of the pipette (72), and an air atomization method (74).

Spray Drying

Spray drying is a single-step, closed-system process applicable to a wide variety of materials, including heat-sensitive materials. This process is often used commercially since the necessary equipment is frequently available at the manufacturing site. As a closed system, it is ideal for good manufacturing practice and the production of sterile materials. The drug and the polymer coating materials are dissolved in a suitable solvent (aqueous or nonaqueous) or the drug may be present as a suspension in the polymer solution. Alternatively, it may be dissolved or suspended within an emulsion or coacervate system. For example, biodegradable polylactide microcapsules can be prepared by dissolving the drug and polymer in methylene chloride (75). Methylcellulose and sodium carboxymethylcellulose spray-dried microspheres are prepared

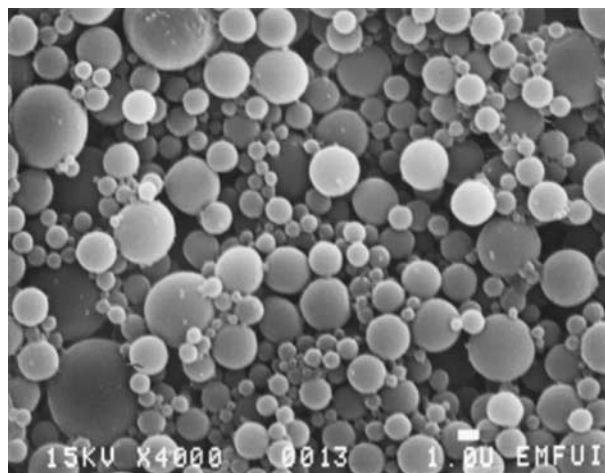


Fig. 3 Scanning electron micrograph of poly(lactic acid) microspheres containing phenolphthalein prepared by the solvent evaporation method (77). Magnification $\times 4000$.

by dissolving the polymers in aqueous systems. The microsphere size is controlled by the rate of spraying, the feed rate of the polymer drug solution, the nozzle size, the temperature in the drying and collecting chambers, and the size of these two chambers. The quality of spray-dried products is improved by the addition of plasticizers (76) that promote polymer coalescence and film formation and enhance the formation of spherical and smooth-surfaced microcapsules.

Solvent evaporation

This is one of the earliest methods of microsphere manufacture. The polymer and drug must be soluble in an organic solvent, frequently methylene chloride. The solution containing the polymer and the drug may be dispersed in an aqueous phase to form droplets. Continuous mixing and elevated temperatures may be employed to evaporate the more volatile organic solvent and leave the solid polymer–drug particles suspended in an aqueous medium. The particles are finally filtered from the suspension. Fig. 3 shows poly(lactic acid) particles prepared in this manner (77).

Precipitation

Precipitation is a variation on the evaporation method. The emulsion consists of polar droplets dispersed in a nonpolar medium (78). Solvent may be removed from the droplets by the use of a cosolvent. The resulting increase in the polymer drug concentration causes precipitation forming a suspension of microspheres.

Freeze Drying

This technique involves the freezing of the emulsion (78); the relative freezing points of the continuous and dispersed phases are important. The continuous-phase solvent is usually organic and is removed by sublimation at low temperature and pressure. Finally, the dispersed phase solvent of the droplets is removed by sublimation, leaving polymer-drug particles.

Chemical and Thermal Cross-Linking

Microspheres made from natural polymers are prepared by a cross-linking process; polymers include gelatin, albumin, starch, and dextran. A water–oil emulsion is prepared, where the water phase is a solution of the polymer that contains the drug to be incorporated. The oil phase is a suitable vegetable oil or oil-organic solvent mixture containing an oil-soluble emulsifier. Once the desired water–oil emulsion is formed, the water soluble polymer is solidified by some kind of cross-linking process. This may involve thermal treatment (79) or the addition of a chemical cross-linking agent such as glutaraldehyde to form a stable chemical cross-link as in albumin (80). If chemical or heat cross-linking is used, the amount of chemical and the period and intensity of heating are critical in determining the release rates and swelling properties of the microspheres (79). If glutaraldehyde is the cross-linking agent, residual amounts can have toxic effects.

NANOPARTICLES

Nanoparticles, 10–1000 nm polymeric particles, are prepared from the same natural and synthetic biodegradable polymers as microspheres (81–83). Albumin nanoparticles are prepared by the cross-linking processes mentioned previously. For the preparation of particles from synthetic polymers, heterogeneous bulk polymerization techniques of suspension, emulsion, and micelle polymerization are often used.

Suspension polymerization of water-insoluble liquid monomer and drug may be achieved by agitating a dispersion of droplets, 10–1000 nm in diameter, in a continuous aqueous phase (84). The temperature must be carefully controlled. An initiator is frequently employed to increase the reaction rate in the droplets, and the aqueous phase may contain stabilizers to prevent coalescence and thickening agents to increase the viscosity. The polymer is formed by reaction of the functional groups of the monomer. This process has the advantage that the

continuous phase absorbs the heat of the polymerization reaction and minimizes the temperature change within the droplets. However, aggregation of the particles may arise as the polymer molecules in suspended particles coalesce. Unfortunately, it is difficult to eliminate stabilizers and additives, used to prevent coalescence, from the final product.

Emulsion polymerization involves the dispersion of the monomer liquid in an aqueous phase to form droplets, 0.05–5 μm in diameter (85–88). An initiator and a surfactant, in a concentration higher than its critical micelle concentration, are present in the aqueous phase. Excess surfactant molecules form micelles whose hydrophobic interiors take up part of the available monomer, causing them to swell. Initiator radicals diffuse into these swollen micelles and begin the polymerization process. As the monomer is consumed, it is replaced by progressive diffusion of the remaining monomer from its location in the emulsified droplets to the interior of the micelles. The micelles continue to swell in size as polymerization proceeds. The enlarging surfaces compete for available surfactant, thus influencing the number of available micelles that can participate in the polymer formation. This method yields particles of very small size and predictable number at low temperatures. However, the particles usually have a high concentration of associated monomer, which may be toxic.

Micelle polymerization differs from emulsion polymerization in that all of the monomer and the drug is contained within micelles composed of surfactant. Diffusion of the monomer from the micelles is prevented by the nonsolvent properties of the outer phase. Therefore, the increase in particle size is negligible as the polymerization proceeds.

CHARACTERIZATION

Materials

The polymer employed to prepare microspheres must be characterized in terms of molecular weight and purity (89, 89–91), however this topic is beyond the scope of this article. Characterization of the materials may have implications for the formation of the microspheres. The viscosity and film-forming properties of the polymers used should be known. Viscosity can affect the tendency to form microspheres, their size, and even their shape. Burgess and coworkers (70, 92) have shown that albumin–acacia coacervates do not form microcapsules under certain conditions of pH and ionic strength, if the viscosity

of the coacervate phase is too high. Burgess and Carless (63) developed a method to predict the optimum conditions for complex coacervation based on the charge carried by the two polymers involved.

Microspheres

Size characterization may be conducted by various methods including light microscopy, resistance blockage techniques (Coulter analysis), light blockage techniques, light scattering, laser diffraction analysis, and for particles less than 1 μm , photon correlation spectroscopy. Electron microscopy, scanning electron microscopy, and scanning tunneling microscopy are used for surface characterization of microspheres. Fourier transform Raman spectroscopy or X-ray photoelectron spectroscopy may be used to determine if any of the material which should have been entrapped is present on the surface and if any other contaminants are present (93, 94). Other surface characterization techniques include surface charge analysis using microelectrophoresis. Surface charge can provide information regarding microsphere aggregation (14, 95). Surface charge is an important parameter with respect to the interaction of microspheres within the body (96).

Surface forces are important in the entrapment, wetting, and adhesion of core material by the coating material. The wettability of solids by different liquids is usually assessed by contact-angle measurement (97, 98). When wetting of the core material is poor, it is difficult or impossible to form microcapsules. For example, Eudragit RS dissolved in THF-cyclohexane failed to encapsulate charcoal particles but was successful in encapsulating potassium dichromate (99, 100). The surface hydrophobicity of oil droplets has been shown to affect their uptake into complex coacervate droplets (101). Oil uptake was directly related to the hydrophile-lipophile balance (HLB) value at the droplet interface. Additives, such as surfactants, which alter surface properties were shown to affect the uptake of core materials.

Biological Distribution

The disposition of microspheres upon entry to the body has been studied extensively. They are frequently labeled with a radionuclide to allow study by scintillation counting or scintigraphy (23). This work has largely focused on tissue distribution as a function of particle size, with the conclusion that particles below 7 μm tend to locate predominantly in the reticuloendothelial system of the liver, whereas particles 7–15 μm in size tend to be collected in the capillary system of the lung (77). The

acute toxicity (102), tissue interaction (103–105), cell interaction (106–108), and protein interaction (109) of various microspheres have been investigated. These studies demonstrated that the toxicity of microspheres is related to the number and size administered. Initial inflammatory responses to the administration of microspheres were consistent with observations that phagocytosis by neutrophils and macrophages may occur. The extent and nature of the cell and tissue effects of microspheres may be related to the surface characteristics of the polymer employed and to the particle size. For example, fibrinogen has been shown to associate with poly-DL-lactide microcapsules, affecting the surface-charge characteristics (104). This phenomenon is related to the hydrophobic nature of the surface. Indeed, hydrophilic coatings can reduce the uptake of microspheres by the liver and peritoneal macrophages (110–112).

THERAPEUTIC APPLICATIONS

Targeting

Drugs can be targeted to specific sites in the body using microspheres and other colloidal carrier systems. Targeting by colloid delivery systems is dealt with elsewhere in this encyclopedia (17). Targeting by microspheres may be passive, active, diversional, or physical. In passive targeting the microspheres follow their natural distribution in the body (which depends on particle size, shape, surface characteristics, particle deformation, and route of administration). In active targeting the natural distribution of the microspheres is altered (for example, by attachment of site-specific vectoring agents such as monoclonal antibodies and lectins). Diversional targeting means blocking the natural distribution of the microspheres, for example, partly or completely impairing the cells of the reticuloendothelial system, which would otherwise take up the microspheres. Physical targeting involves an external influence, such as a change in temperature or a magnetic field to direct the microspheres to the desired site.

Degrees of targeting can be achieved by localization of the drug to a specific area in the body, to a particular organ in the body (for example, the lungs), to a particular group of cells within the body (for example, the Kupffer cells), and even to intracellular structures (such as the lysosomes or the cell nucleus). The problems associated with directing microspheres to specific areas in the body following parenteral administration are discussed in the article Colloids and Colloid Drug Delivery in this encyclopedia (17).

Oral targeting can be achieved using microspheres; those less than 10 μm in diameter have been shown to target the Peyer's patch (113). Microspheres less than 5 μm were shown to be transported through the lymphatics within macrophages and those larger than 5 μm remained in the Peyer's patch (113). Toxoid vaccine microcapsules were effectively delivered and released in the gut-associated lymphoid tissue following oral administration (114).

Controlled Release

The rate of drug release from microspheres dictates their therapeutic action. Release is governed by the molecular structure of the drug and polymer, the resistance of the polymer to degradation, and the surface area and porosity of the microspheres (16, 115, 116). Reservoir delivery systems extend the residence time of drug within the systemic circulation and were originally focused on zero-order dissolution kinetics. This mathematical expression describes a linear relationship between rate of appearance in plasma and time. Ideally, the plasma drug concentration is independent of time for most of the dissolution period and is optimally maintained in the therapeutic window.

In nonporous polymeric systems the rate of drug release is dictated by the device surface area which is linked directly to its shape. Drug release from polymeric systems with a variety of geometries has been described (117). Zero-order release kinetics may be more easily achieved with slab or rod geometries than spheres. The rate of release from spheres may result from polymer diffusion or erosion (116, 118, 119). Diffusion-mediated release has been studied extensively and described mathematically (120, 121).

The internal structure of microspheres may vary as a function of the microencapsulation process employed (122). Reservoir microcapsules have a core of drug coated with a polymer. The drug is distributed homogeneously throughout the polymeric matrix in monolithic microspheres.

Controlled drug release from microspheres occurs by diffusion of drug through a polymeric excipient, diffusion of entrapped drug as the polymer erodes, and release of drug through pores in the polymeric microspheres. If the drug is released by diffusion through the polymer without erosion, the release depends on the surface area of the microspheres and the path length of the drug in transit to the surrounding environment. For example, increasing the surface area, by reducing particle size, results in an increased release rate. The path length of motion for the drug in the matrix can be controlled by manipulating the microsphere loading. Microspheres with a high drug

content release the active ingredient more rapidly than those with a low load. Physicochemical properties of the drug and excipient such as permeability of one in the other, identity of the polymer, degree of crystallinity, inclusion of plasticizers and fillers, and thickness of the polymer influence the drug release rate.

Release from reservoir microcapsules

The factors affecting drug release may be elucidated by a study of drug release from the simplest system, a reservoir microcapsule. Diffusion of drug through such a structure may involve transport not only through an isotropic medium, such as the drug in solution, but also through a polymeric membrane. Transport of drug through such a membrane involves dissolution of the drug in the polymer at the high-concentration side of the membrane interface and diffusion across the membrane in the direction of decreasing concentration. In addition, the concentration difference across the membrane, which is taken as the driving force for drug transport, tends to decrease as the solubility of the drug on the upstream side of the membrane decreases. Therefore, the dissolution rate of poorly soluble drugs can be an important factor in limiting drug release.

Consider a spherical reservoir device where the thermodynamic activity of the core material is maintained constant within the device, and the coating is inert, homogeneous, and of uniform thickness. The steady-state release rate derived from Fick's law is

$$\frac{dM}{dt} = 4DKC \frac{r_o, r_i}{r_o - r_i} \quad (1)$$

where r_o and r_i are the outside and inside radii, respectively, D is the diffusion coefficient of the drug molecule, K is the partition coefficient, and C is the concentration difference between either side of the coating. Assuming all parameters on the right side of Eq. 1 remain constant, consistent with no change in activity of the core material, C does not change. Integration of Eq. 1 over a finite period of the steady state would indicate that the drug release was zero-order. This is explained by the pathlength and surface area remaining constant since the membrane the drug has to traverse is of uniform thickness. If, however, the thermodynamic activity of the core material does not remain constant, then release is first-order.

It may be necessary to consider the effect of a boundary layer on the release rate. A boundary layer of appreciable drug concentration on the surface of the device would hinder drug release by diffusion. The effect of the layer is more marked with drugs of low solubility and with microparticles having irregular surfaces. From this simple

example, it can be seen that various factors affect the release rate from reservoir microcapsules.

Release from monolithic micromatrices

In a monolithic microsphere the path length does not remain constant, since the drug in the center has a longer path to travel than the drug near the surface, and therefore the rate of release decreases exponentially with time.

Nevertheless, monolithic microspheres can be made to release drug at an approximately constant rate (117, 123). The core loading of these microspheres may be increased to create structures similar to those of reservoir microcapsules. An optimum combination of particle sizes (a size distribution), may be prepared to achieve a constant rate of drug release. Preparing microspheres with an erodible polymer in such a way that maximum erosion occurs in conjunction with minimum diffusion may establish a constant release rate. Although the principles described here appear simple, they are difficult to utilize because of their dependence on a number of factors, each of which can complicate the process.

Live-Cell Encapsulation

Microcapsules have been investigated as potential artificial cells [first demonstrated by Chang in 1964 (124)] and as a means to immobilize live cells (72). Potential medical applications of artificial cells have been investigated such as artificial liver, artificial kidney, and red blood cell substitutes. The encapsulation of living cells has been investigated as a means to transplant tissues without immune rejection. The capsule membrane must be semipermeable in order to be impermeable to high molecular weight antibodies, which would cause rejection or destruction of the transplanted cells, but be permeable to lower molecular weight species such as oxygen, nutrients, and internally generated therapeutic agents (for example, hormones such as insulin). The encapsulation process must not be harmful to the cells and therefore must not involve harsh conditions, such as those caused by the use of organic solvents and heat during processing. The finished microcapsules must be sterile, stable, and biocompatible. The encapsulation of mammalian cells is more difficult than that of microbial cells since their membranes are more fragile, and these must be preserved during encapsulation. The temperature, pH, ionic strength, and toxicity of solvent and reagent must be carefully controlled. A common method of microencapsulating live cells is to entrap the cells in a protective gel and form a permanent membrane around the gel droplets. A calcium alginate gel method was developed by Lim and Sun (72), has been successfully

applied to mammalian cells. A microencapsulation system combining the easy setting characteristics of alginate and the stability and biocompatibility of hydrogel polymers has been devised (125, 126); it involves an alginate-HEMA graft copolymer.

Immobilized islets of Langerhans are able to respond to external glucose concentrations and release insulin into the systemic circulation (72). A number of animal studies have shown that alginate-polylysine encapsulated pancreatic islets were successful in correcting the diabetic state in rats for periods of two to three weeks (127).

Immobilized cells are also used in biotechnology in the production of protein molecules. For example, entrapped hybridoma cells have been used for the production of monoclonal antibodies (30) which are secreted into the microcapsules. This allows for easier collection of the antibodies compared to growing the hybridoma cells directly in the culture medium. The microcapsules are easily separated from the culture medium and broken to collect the antibodies. Isolation of the antibodies from the culture medium involves numerous purification steps, and product is lost during each of these steps to an extent which depends on the efficiency of the process. Live vaccines have been encapsulated. For example, *Bacillus Calmette Guerin* has been encapsulated in an alginate polylysine-alginate system (74).

STERILIZATION

Microspheres that are administered parenterally must be sterile. Sterilization is usually achieved by aseptic processing. The final product may not be able to undergo terminal sterilization, which may be detrimental to the delivery system, altering the release pattern or destroying the targeting properties. In addition, the entrapped drug or biological substance may not be able to withstand the heat of sterilization. Although the exterior of the microspheres can be investigated for sterility by conventional plating methodology, it is difficult to determine whether the interiors of the microspheres are free from contamination. The microspheres can be broken, although this introduces the possibility of false positive or false negative results. A method has been developed whereby the presence of viable organisms in the interior of microsphere systems can be determined without breaking the microcapsules using a detection method for organism metabolism (128).

Sterilization is one of many aspects that must be addressed when considering microspheres for commercial purposes (129). The development of a product is achieved most effectively by parallel approaches to formulation and

process design. Very little has been published with regard to industrial-scale manufacturing of microspheres for the delivery of pharmaceuticals.

REFERENCES

1. Fujimoto, S.; Miyazaki, M.; Endoh, F.; Takahashi, O.; Shrestha, R.D.; Okui, K.; Mori-moto, Y.; Terao, K. *Cancer* **1985**, *55*, 522–526.
2. Gupta, P.K.; Hung, C.T.; Perrier, D.G. *Int. J. Pharm.* **1986**, *33*, 137–146.
3. Gupta, P.K.; Hung, C.T.; Perrier, D.G. *Int. J. Pharm.* **1986**, *33*, 147–153.
4. Gupta, P.K.; Lam, F.C.; Hung, C.T. *Int. J. Pharm.* **1989**, *51*, 253–258.
5. Jones, C.; Burton, M.A.; Gray, B.N. *J. Pharm. Pharmacol.* **1989**, *41*, 813–816.
6. Schwowe, A.D.; Wise, D.L.; Howes, J.F. *Life Sci.* **1968**, *17*, 1877–1886.
7. Gardner, D.L.; Patanus, A.J.; Fink, D.J. Steroid Release from Microcapsules. *Drug Delivery Systems*; Gabelnick, H.L., Ed.; DHEW Publ. No. (NIH), 77–1238. Department of Health, Education, and Welfare: Washington, 1977; 265–278.
8. Anderson, L.C.; Wise, D.L.; Howes, J.F. *Contraception* **1976**, *13*, 375–384.
9. Kent J.S.; Lewis D.H.; Sanders L.M.; Tice T.S. U.S. Patent 4,675,189 June 23, 1987.
10. Sanders, L.M.; Kent, J.S.; McRae, G.I.; Vickery, B.H.; Tice, T.R.; Lewis, D.H. *J. Pharm. Sci.* **1984**, *73*, 1294–1297.
11. Martodam, R.R.; Twumasi, D.Y.; Liener, I.E.; Powers, J.C.; Nishino, N.; Krejcarek, G. *Proc. Natl. Acad. Sci.* **1979**, *76*, 2128–2132.
12. Rhine, W.D.; Hsieh, D.S.T.; Langer, R. *J. Pharm. Sci.* **1980**, *69*, 265–270.
13. Yapel A.P. U.S. Patent, 4,147,767 Apr. 3., 1979.
14. Burgess, D.J.; Carless, J.E. *Int. J. Pharm.* **1986**, *32*, 207–212.
15. Redmon, M.P.; Hickey, A.J.; DeLuca, P.P. *J. Contr. Rel.* **1989**, *9*, 99–109.
16. Lzumikawa, S.; Yoshioka, S.; Aso, Y.; Takeda, J. *J. Contr. Rel.* **1991**, *15*, 133–140.
17. Burgess, D.J. Colloids and Colloid Drug Delivery Systems. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1990; 3, 31–63.
18. Bungenberg de Jong, H.G. Reversible Systems. *Colloid Science*; Kruyt, H.G., Ed.; Elsevier: New York, 1949; II, 335–432.
19. Green B.K.; Schleicher L. U.S. Patent, 2,800,458 1957.
20. Khalil, S.A.H.; Nixon, J.R.; Carless, J.E. *J. Pharm. Pharmacol.* **1968**, *20*, 215–225.
21. Luzzi, L.A.; Gerraughty, R.J. *J. Pharm. Sci.* **1964**, *53*, 429–431.
22. Phares, R.E.; Sperandio, G.J. *J. Pharm. Sci.* **1964**, *53*, 515–518.
23. Zhou, S.; Hickey, A.J.; Jay, M.; Warren, S.M.; Lord, M.; DeLuca, P.P. *Pharm. Res.* **1988**, *5*, S76.
24. Nixon, J.R. *Microencapsulation*; Marcel Dekker, Inc.: New York, 1976.
25. Gutcho, M.H. *Microcapsules and Other Capsules, Advanced Since 1975*; Noyes Data Corp.: Park Ridge, NJ, 1979.
26. Lim, F. *Biomedical Applications of Microencapsulation*; CRC Press Inc.: Boca Raton, FL, 1984.
27. Kondo, T. *Microencapsulation: New Techniques and Applications*; Techno Books: Tokyo, 1979.
28. Chang, T.M.S. *Artificial Liver and Artificial Cells*; Plenum Press: New York, 1978.
29. Chang, T.M.S. *Biomedical Applications of Microencapsulation*; Lim, F., Ed.; CRC Press: Boca Raton, FL, 1984; 85.
30. Jarvis, A.P.; Spriggs, T.A.; Chigura, W.R. *In Vitro* **1982**, *18*, 276.
31. Donbrow, M. *Microcapsules and Nanoparticles in Medicine and Pharmacy*; Donbrow, M., Ed.; CRC Press: London, 1991; 1–14.
32. Kissel, T.; Demirdere, A. *Controlled Drug Delivery*; Müller, B.W., Ed.; Wissen-Schaftliche Verlagsgesellschaft mbH: Stuttgart, 1984; 103–131.
33. Gurney, R.; Peppas, N.A.; Harrington, D.D.; Banks, G.S. *Drug Dev. Ind. Pharm.* **1981**, *7*, 1–25.
34. Gupta, P.K.; Hickey, A.J. *J. Contr. Rel.* **1991**, *17*, 129–148.
35. Masinde, L.E.; Hickey, A.J. *Pharm. Res.* **1991**, *8*, S120.
36. Masinde, L.E.; Hickey, A.J. *Int. J. Pharm.* **in press**.
37. Gupta, P.K.; Hickey, A.J.; Mehta, R.; DeLuca, P.P. *Pharm. Res.* **1990**, *7*, S82.
38. Versic, R.J. *Drug Cosm. Ind.* **1989**, *144*(6), 30, 32, 34, 75.
39. Maggi, G.C.; Di Roberto, F.M. *Microencapsulation*; Nixon, J.R., Ed.; Marcel Dekker, Inc.: New York, 1976; 103–111.
40. Lim, F.; Buchler, R.J. *Methods in Enzymology*; Langome, J.J., Van Vunakis, H., Eds.; Academic Press: New York, 1981; 73.
41. Lim, F. *Adv. Biotechnol. Progr.* **1988**, *7*, 185–197.
42. Walford, J.; Lim, T.M.; Lam, T.J. *Aquaculture* **1991**, *92*, 225–235.
43. Bodmeier, R.; Wang, J.; Bhagwatwar, H. *J. Microencaps.* **1992**, *9*, 99–107.
44. Benita, S.; Zonai, O.; Benoit, J.-P. *J. Pharm. Sci.* **1986**, *75*, 847–851.
45. Bodmeier, R.; Wang, J.; Bhagwaywar, H. *J. Microencaps.* **1992**, *9*, 89–98.
46. Hickey, A.J.; Fults, K.; Pillai, R.S. *J. Biopharm. Sci.* **1992**, *3*, 107–113.
47. Pillai, R.S.; Yeates, D.B.; Miller, I.F.; Hickey, A.J. *Proceed. Intern. Symp. Contr. Rel. Bioact. Mater.* **1992**, *19*, 224–225.
48. Conix, A. *Macromol. Synth.* **1966**, *2*, 95–98.
49. Leong, K.W.; Brott, B.C.; Langer, R. *J. Biomed. Mater. Res.* **1985**, *19*, 945–955.
50. Mathiowitz, E.; Langer, R. *J. Contr. Rel.* **1987**, *5*, 13–22.
51. Wurster D.E. US Patent 2,648,609 1949.
52. Wurster, D.E. *J. Am. Pharm. Assoc. Sci. Ed.* **1959**, *48*, 451–454.
53. Robinson, N.J.; Gross, G.M.; Lantz, R.J. *J. Pharm. Sci.* **1968**, *57*, 1983–1988.
54. Porter, St.C. *Pharm. Technol.* **1979**, *3*, 54–59.
55. Porter, St.C. *Pharm. Technol.* **1980**, *4*, 66–75.

56. Lehmann, K.; Dreher, D. *Pharm. Ind.* **1972**, *34*, 894–903.
57. Banker, G.S. *Pharm. Technol.* **1981**, *5*, 54–62.
58. Lehmann, K. *Microcapsules and Nanoparticles in Medicine and Pharmacy*; Donbrow, M., Ed.; CRC Press: London, 1991; 73–97.
59. Burgess, D.J. *J. Coll. Interface Sci.* **1990**, *140*, 227–238.
60. Bungenberg de Jong, H.G.; Kruyt, H.R.; Lens, J. *Kolloidchem. Beih.* **1932**, *36*, 429–434.
61. Nixon, J.R.; Nouh, A. *J. Pharm. Pharmacol.* **1978**, *30*, 533–537.
62. Madan, P.L. *Drug Dev. Ind. Pharm.* **1978**, *4*(95).
63. Burgess, D.J.; Carless, J.E. *J. Coll. Interface Sci.* **1984**, *98*, 1–8.
64. Calanchi, M. *Microencapsulation*; Nixon, J.R., Ed.; Marcel Dekker, Inc.: New York, 1976; 93–101.
65. Urmeda, T.; Matsuzawa, A.; Yokoyama, T.; Kuroda, K.; Kuroda, T. *Chem., Pharm. Bull.* **1983**, *31*, 2793–2798.
66. Nakajima, T.; Takashima, Y.; Iida, K.; Mitsuta, H.; Koishi, M. *Chem. Pharm. Bull.* **1987**, *35*, 1201–1206.
67. de Sabata, V. *Microencapsulation*; Nixon, J.R., Ed.; Marcel Dekker, Inc.: New York, 1976; 143–162.
68. Jizomoto, H. *J. Pharm. Sci.* **1984**, *73*, 879–882.
69. Jizomoto, H. *J. Pharm. Sci.* **1985**, *74*, 469–472.
70. Burgess, D.J.; Singh, O.N. *J. Pharm. Pharmacol.* **1993**, *45*, 586–591.
71. Singh, O.N.; Burgess, D.J. *Pharm. Res.* **1972**, *9*(S8).
72. Lim, F.; Sun, A.M. *Science* **1980**, *210*, 908–910.
73. Rha C.K.; Rodrigues-Sanches D. *US Patent*. 4,749,620, 1980.
74. Kwok, K.K.; Groves, M.J.; Burgess, D.J. *Pharm. Res.* **1991**, *8*, 341–344.
75. Bodmeier, R.; Chen, H. *J. Pharm. Pharmacol.* **1988**, *40*, 754–757.
76. Wan, L.S.C.; Heng, P.W.S.; Chia, C.G.H. *J. Microencaps.* **1992**, *9*, 53–62.
77. Hickey, A.J.; Tian, Y.; Parasrampur, D.; Kanke, M. *Biopharmaceut. Drug Dispos.* **1933**, *14*, 181–186.
78. DeLuca, P.P.; Hickey, A.J.; Hazrati, A.M.; Wedlund, P.; Rypacek, F.; Kanke, M. *Topics in Pharmaceutical Sciences*; Breimer, D.D., Speiser, P., Eds.; Elsevier: New York, 1987; 429–442.
79. Burgess, D.J.; Davis, S.S.; Tomlinson, E. *Int. J. Pharm.* **1987**, *39*, 129–136.
80. Burgess, D.J.; Davis, S.S. *Int. J. Pharm.* **1988**, *46*, 69–76.
81. Couvreur, P.; Kante, B.; Lenaerts, V.; Scailteur, V.; Roland, M.; Speiser, P. *J. Pharm.Sci.* **1980**, *69*, 199–202.
82. Couvreur, P.; Tulkens, P.; Roland, M.; Trouet, A.; Speiser, P. *FEBS Letters* **1977**, *84*, 323–326.
83. Couvreur, P.; Kante, B.; Roland, M.; Guiot, P.; Bauduin, P.; Speiser, P. *J. Pharm. Pharmacol.* **1979**, *31*(331).
84. Seymour, R.B.; Carraher, C.E. *Polymer Chemistry*; Marcel Dekker, Inc.: New York, 1988; 332–342.
85. Seymour, R.B. *Introduction to Polymer Chemistry*; McGraw-Hill: New York, 1971.
86. Harkins, W.D. *J. Am. Chem. Soc.* **1947**, *69*, 1429–1444.
87. Harkins, W.D. *J. Polymer Sci.* **1950**, *5*, 217–251.
88. Smith, W.V.; Ewart, R.H. *J. Chem. Phys.* **1948**, *16*, 592–599.
89. Kotliar, A.M.; McDonnell, M.E.; Walsh, E.K. *A Guide to Materials Characterization and Chemical Analysis*; Sibilia, J.P., Ed.; VCH Publishers, Inc.: New York, 1988; 229–249.
90. Hanrahan, J.M.; Gabriel, M.K.; Williams, R.J.; McDonnell, M.E. *A Guide to Materials Characterization and Chemical Analysis*; Sibilia, J.P., Ed.; VCH Publishers, Inc.: New York, 1988; 81–84.
91. Seymour, R.B.; Carraher, C.E. *Polymer Chemistry, An Introduction*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1988; 156–169.
92. Burgess, D.J.; Kwok, K.K.; Megremis, P.T. *J. Pharm. Pharmacol.* **1991**, *43*, 232–236.
93. Binns, J.S.; Mella, C.D.; Davies, M.C. *Proceed. Intern. Symp. Contr. Rel. Bioact. Mater.* **1990**, *17*, 150–151.
94. Davies, M.C.; Lynn, R.A.P.; Khan, M.A.; Paul, A.; Domb, A.; Langer, R. *Proceed. Intern. Symp. Contr. Rel. Bioact. Mater.* **1990**, *17*, 232–233.
95. Labhasetwar, V.D.; Doric, A.K. *J. Microencaps.* **1991**, *8*, 83–85.
96. Tabata, Y.; Ikada, Y. *Biomaterials* **1988**, *9*, 356–362.
97. Wake, W.C. *Adhesion and the Formulation of Adhesives*; Applied Science: New York, 1982.
98. Kinlock, A.J. *Adhesion and Adhesives: Science and Technology*; Chapman and Hall: London, 1987.
99. Benita, S.; Hoffman, A.; Donbrow, M. *J. Microencaps.* **1985**, *2*, 207–215.
100. Donbrow, M.; Hoffman, A.; Benita, S. *J. Microencaps.* **1990**, *7*, 1–9.
101. Rabisková, M.; Opawale, F.O.; Burgess, D.J. *Pharm. Res.* **1991**, *9*, S-149.
102. Yokel, R.A.; Sabo, J.P.; Simmons, G.H.; DeLuca, P.P. *Toxicol. Letters* **1981**, *9*, 165–170.
103. Schoen, F.J.; Kintanar, E.B.; Osol, R.G.; Lee, E. *J. Biomed. Mater. Res.* **1986**, *20*, 709–721.
104. Vissler, G.E.; Robison, R.L.; Maulding, H.V.; Fong, J.W.; Pearson, J.E.; Argentieri, G.J. *J. Biomed. Mater. Res.* **1986**, *20*, 667–676.
105. Willmott, N.; Cummings, J. *Biochem. Pharmacol.* **1987**, *36*, 521–526.
106. Kanke, M.; Morlier, E.; Geissler, R.; Powell, D.; Kaplan, A.; DeLuca, P.P. *J. Parent.Sci. Technol.* **1986**, *40*, 114–18.
107. Scheffel, U.; Rhodes, B.A.; Natarajan, T.K.; Wagner, H.N. *J. Nud. Mod.* **1972**, *13*, 498–503.
108. Simon, S.I.; Schmid-Schonbein, G.W. *Biophys. J.* **1988**, *3*, 163–173.
109. Rolland, A.; Begue, J.-M.; Le Verge, R.; Guillouzo, A. *Int. J. Pharm.* **1989**, *53*, 67–73.
110. Makino, K.; Ohshima, H.; Kondo, T. *J. Coll. Interf. Sci.* **1987**, *115*, 65–72.
111. Illum, L.; Hunniball, I.M.; Davis, S.S. *Int. J. Pharm.* **1986**, *29*, 53–65.
112. Illum, L.; Davis, S.S. *J. Pharm. Sci.* **1983**, *72*, 1086–1089.
113. Jani, P.; Florence, A.T.; Halbert, G.W.; Langridge, J. *J. Pharm. Pharmacol.* **1989**, *41*, 47.
114. Ritger, P.L.; Peppas, N.A. *J. Contr. Rel.* **1987**, *5*, 23–36.
115. Pitt, C.G.; Schindler, A. *Controlled Drug Delivery*; Bruck, S.D., Ed.; CRC Press: Boca Raton, FL, 1983; I, 53–80.
116. Pitt, C.G.; Gu, Z.-W. *J. Contr. Rel.* **1987**, *4*, 282–292.
117. Cheung, W.K.; Yakobi, A.; Silber, B.M. *J. Contr. Rel.* **1988**, *6*, 263–270.
118. Carstensen, J.T. *Controlled Drug Delivery*; Müller, B.W., Ed.; Wissenschaftliche Verlagsgesellschaft mbH: Stuttgart, 1984; 132–145.

119. Crank, J. *The Mechanics of Diffusion*, 2nd Ed.; Oxford Science Publications: Oxford, 1975.
120. Singh, S.K.; Fox, R.O.; Fan, L.T. *Proceed. Intern. Symp. Contr. Rel. Bioact. Mater.* **1987**, *14*, 65–66.
121. Redmon, M.P. *Physicochemical Characteristics of a Porous Polymer Matrix for Drug Delivery*; Ph.D. Thesis University of Kentucky: Lexington, KY, 1989.
122. Brannon-Peppas, L. Design and Mathematical Analysis of Controlled Release from Microsphere-Containing Polymeric Implants. *J. Contr. Rel.* **1992**, *20*, 201–207.
123. Chang, T.M.S. *Science* **1964**, *146*, 525–525.
124. Sefton, M.V.; Dawson, R.M.; Broughton, R.L.; Biysnink, J.; Sugamori, M.E. *Biotechnol. Bioeng.* **1987**, *29*, 1135–1143.
125. Lamberti, F.V.; Sefton, M.V. *Biochim. Biophys. Acta.* **1983**, *759*, 81–91.
126. Sun, A.M.; O'Shea, G.M.; Leung, Y. *Artif. Organs* **1983**, *5* (Suppl.), 69.
127. Sun, A.M.; O'Shea, G.M.; Goosen, M.F.A. *Appl. Biochem. Biotechnol.* **1984**, *10*, 87–100.
128. Kwok, K.K.; Burgess, D.J. *Pharm. Res.* **1992**, *9*, 410–413.
129. Floy, B.J.; Visor, G.C.; Sanders, L.M. *Polymeric Delivery System: Properties and Applications. ACS Symposium Series*; EI-Nokali, M.A., Piatt, D.N., Sharpentier, B.A., Eds.; American Chemical Society: Washington, DC, 1993; 520, 154–167.
130. Davis, S.S.; Illum, L.; McVie, J.G.; Tomlinson, E. *Microspheres and Drug Therapy: Pharmaceutical and Medical Aspects*; Elsevier: Amsterdam, 1984.
131. Donbrow, M. *Microcapsules and Nanoparticles in Medicine and Pharmacy*; CRC Press: London, 1991.
132. Goosen, M.F.A. *Fundamentals of Animal Cell Encapsulation and Immobilization*; CRC Press: London, 1992.
133. Guiot, P.; Couvreur, P. *Polymeric Nanoparticles and Microspheres*; CRC Press: Boca Raton, FL, 1986.
134. Gutcho, M.H. *Microcapsules and Other Capsules, Advanced Since 1975*; Noyes Data Corp.: Park Ridge, NJ, 1979.
135. Illum, L.; Davis, S.S. *Polymers in Controlled Drug Delivery*; Wright: Bristol, UK, 1987.
136. Kondo, T. *Microencapsulation: New Techniques and Applications*; Techno Books: Tokyo, 1979.
137. Lim, F. *Biomedical Applications of Microencapsulation*; CRC Press, Inc.: Boca Raton, FL, 1984.
138. Müller, R.H. *Colloidal Carriers for Controlled Drug Delivery and Targeting*; CRC Press: London, 1979.
139. Nixon, J.R. *Microencapsulation*; Marcel Dekker, Inc.: New York, 1976.
140. Tomlinson, E.; Davis, S.S. *Site-Specific Drug Delivery*; John Wiley & Sons: Chichester, UK, 1986.
141. Kreuter, J. *Colloidal Drug Delivery Systems*; Marcel Dekker, Inc.: New York, 1994.
142. Benita, S. *Microencapsulation*; Marcel Dekker, Inc.: New York, 1996.
143. Diederichs, J.E.; Muller, R.E. *Future Strategies for Drug Delivery with Particulate Systems*; CRC Press: Boca Raton, 1998.
144. Burgess, D.J. Colloids and Colloid Drug Delivery Systems. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1990; 3, 31–63.
145. Davis, S.S. Colloids and Drug Delivery Systems. *Pharm. Technol.* **1987**, *11*, 110–117.
146. Kaye, B.H. Microencapsulation: The Creation of Synthetic Fine Particles with Specified Properties. *Kona* **1992**, *10*, 65–82.
147. Morimoto, Y.; Fujimoto, S. *CRC Crit. Rev. Ther. Drug Carrier Sys.* **1985**, *2*, 119–163.
148. Oppenheim, R.C. Solid Colloidal Drug Delivery Systems: Nanoparticles. *Int. J. Pharm.* **1981**, *8*, 217–234.
149. Thies, C. Microencapsulation. *Encyclopedia of Polymer Science and Engineering*; Kroschwitz, J., Ed.; John Wiley & Sons, Inc.: New York, 1987; 9, 724–745.
150. Tomlinson, E. Theory and Practice of Site-Specific Drug Delivery. *Adv. Drug Delivery Rev.* **1987**, *1*, 87–198.
151. Torchilin, V.P. *CRC Crit. Rev. Ther. Drug Carrier Systems* **1985**, *2*, 119–163.
152. Yappel, A.F. Albumin Microspheres. *Meth. Enzymol.* **1985**, *112*, 3–67.

MICROBIOLOGIC MONITORING OF CONTROLLED PROCESSES

Gregory F. Peters

Marghi R. McKeon

Lab Safety Corp., Des Plaines, Illinois, U.S.A.

INTRODUCTION

Microbiologic monitoring of controlled pharmaceutical and medical device manufacturing, and pharmacy compounding processes, is mandated in numerous standards and guidelines (1–3), although procedures, limits, and frequencies are not well defined (4). Because many characteristics of microbiologic sampling limit its value as a monitoring method (5), efforts to detect contamination in controlled environments require carefully developed and executed sampling plans to produce reliable data that confirm the acceptability of operating conditions.

Monitoring of any controlled process is a component of an outcome-producing, closed-loop system for assuring continued operation of critical processes in accordance with validated design conditions. To achieve this goal, a monitoring plan must be developed, conducted, and evaluated within the context of a Validation and Monitoring protocol. All results must be related to the original validated process, either as evidence that it continues to operate within acceptable limits, or as a means of detecting shifts in the process that might impinge on product quality. Ideally, monitoring results will also provide information that will be useful in determining the cause of such shifts.

The objectives of the monitoring plan within the validation and monitoring system for quality management must be clearly defined so that the information collected will be relevant to system goals. The limitations of sampling equipment and methods must be taken into consideration when developing the sampling plan and interpreting results. The underlying causes for shifts in various monitoring results must be understood in order to facilitate development of effective corrective action plans.

VALIDATION AND MONITORING RATIONALE

The regulatory requirements for validation of pharmaceutical aseptic processes are clear (6). Generally accepted

quality assurance principles require initial demonstration of the efficacy of any process (*validation*), followed by regular, periodic observation to demonstrate that the process continues to operate in accordance with validation conditions (*monitoring*).

Validation usually consists of a series of “worst-case” process simulations, wherein a sterile growth medium is substituted for product to demonstrate that processing consistently yields products of acceptable quality (6). During this Process Qualification (PQ) phase, variable conditions that might effect product quality are carefully defined, controlled, monitored, and documented, and the assumption is reasonably made that the process will then yield the same product quality achieved during the PQ, so long as all variable factors are controlled to duplicate validation conditions. This assumption is based upon the results of monitoring data obtained from a variety of sources. The validity of the assumption of acceptable quality is, therefore, dependent upon the reliability of the monitoring data as a measure of control of process variables.

Validation Protocol

The validation protocol should define the manufacturing or compounding process, its purpose in terms of the desired positive impact on product quality, and how that impact will be demonstrated. The protocol should include the following components:

1. A description of the product, and applicable release criteria including AOQL/ROQL;
2. The facility design rationale for maintaining process integrity, including identification and elimination of inaccessible areas that may be difficult to decontaminate, enumeration of the clean-space engineering controls, and how these controls will be applied, tested, and monitored;
3. A schematic description of the aseptic process and the critical work surfaces, work zones, and support areas, including the designation of particulate cleanliness class (7), microbial target values (8, 9) (Table 1), and

Table 1 Process monitoring targets/frequency

Function/Process	Cleanliness class ^{a,b}						Monitoring frequency				C.F.U. ^c Target															
	Class 100 ^a	Class 10,000 ^a	Class 100,000 ^a	M 3.5	M 5.5	M 6.5	Each shift	Daily	Twice weekly	Weekly																
											<0.1 Per Ft. ^d	<0.5 Per M ^d	<2.5 Per Ft. ^d	<3.0 Per M ^d	<20 Per M ^d	<100 Per Ft. ^d	3 Per Contact plate	5 Per Contact plate	10 Per Contact plate	20 Per Contact plate						
Critical worksurface ^e	•			•			•				A		A				S									
Support areas ^e		•			•		•						A			A		S								
Other support areas ^e			•			•			•					A		A										
Other potential prod./container contact ^e					•				•					A		A										
Other non-prod./container contact ^e			•			•				•				A		A										
Laminar airflow hoods	•			•			•				A				A											
Biological safety cabs.	•			•			•				A				A			S ^f								
Host-cell culture	•			•			•				A				A			S ^f								
Immediate Processing	• ^g		• ^h	• ⁱ		• ^h	• ^g			• ^h	A ⁱ		A ^h		A ^g			S ^{f,i}							S ^{h,i}	
Formulation					•		•				A		A		A			S								
Final Production	•			•			•								A			S ^f								
Equipment ^e	•			•			•											S ^f								
Equipment ^e					•		•			•								S								
Floor ^e	•			•			•																			
Floor ^e					•		•											S								
Floor ^e			•				•																		S ^j	
Personnel gloves ^e	•			•			•			•								S								
Personnel gloves ^k							• ^g		• ^k																S	
Personnel barriers ^e	•			•			•											S								
Personnel barriers ^e					•		• ^k		• ^g																S	
							• ^k		• ^g																	S

^aFederal Standard 209e U.S. Customary.

^bSI ^cColony-forming units.

^dCubic valve.

^eU.S.P. <1116>.

^fIncluding floor.

^gOther support areas.

^hClosed validated systems.

ⁱOpen systems.

^jRecommended.

^kImmediately adjacent to Class 100.

A. Aerobiologic; S. Surface.

Table courtesy of Lab Safety Corp., Des Plaines, IL.

- engineering control equipment validation methods^a (10);
4. The selection and justification of gowning and barrier techniques to ensure adequate isolation of personnel, based upon industry standards (3) and process requirements;
 5. A definition of the aseptic techniques and work practices of operative personnel, and a report of findings based upon videotaped observation of the actual work stream during prequalification runs for identification and elimination of personnel-generated contamination sources, identification of susceptible areas including critical sites and steps, and indicator sites;
 6. A description of sanitizing methods and sanitizing compound validation;
 7. A definition of the equipment and methods to be used in assuring reliable test data; and
 8. All test data, including instrument calibrations, testing and certification reports, and statistical justification.

Monitoring Plan

Following evaluation of all environmental monitoring data collected during the PQ, a monitoring plan (11) defining ongoing monitoring procedures, locations, and frequency should be implemented. The PQ data from product testing should be compared to environmental and process monitoring results to determine the monitoring sites and methods that best correlate with shifts in product quality. The plan should

1. Assure specified, periodic monitoring of critical manufacturing or compounding process parameters at critical points during periods of peak activity, and establish the circumstances and frequency with which monitoring is to be carried out to assure a reliable basis for claiming process control.
2. Provide for standardized, quantitative microbiologic sampling of process air, environmental surfaces, and personnel barriers, as well as sampling of other, related parameters.
3. Include sampling location maps, sample sizes, probe heights, methods, equipment, and frequency during

^aValidation testing of HEPA filters requires an exacting aerosol challenge of 100% of the filtration media, frame, and locking device in accordance with Secs. 40 and 50 F.S. 209b (18). Successful testing in this manner establishes control of the "first air" emanating directly from the filter, as it approaches the entrance plane within the unidirectional slipstream, to better-than-Class I conditions. Monitoring of HEPA filters in accordance with F.S. 209e (7) involves an average of DPC readings derived from a number of representative locations to assure Class 10 or 100 conditions at the entrance plane of the unidirectional slipstream (10)

manufacturing operations, and a method for statistical justification of results.

4. Include alert and action limit criteria for acting upon ongoing monitoring information.
5. Include a system for evaluating and modifying the monitoring plan to assure collection of reliable, useful data, and
6. Include a corrective action plan, and methods of verifying the efficacy of any corrective actions taken.

Limitations of Microbiologic Monitoring

The minimum media-fill validation requirement of not more than one sterility failure per thousand units, representing the minimum sterility assurance level of 10^{-3} (>99.9%) is the only microbiologic limit in the validation and monitoring scheme that is based upon demonstrated product quality. Achievement of this sterility assurance level represents the aggregate impact of all process design and control factors, including sampling and attendant laboratory procedures. [This limit, however, probably does not reflect the true integrity of a valid aseptic process (12).] All other limits are indices, which are used indirectly to demonstrate that the process is under control as validated. Because all environmental monitoring is necessarily performed at some point downstream and apart from the product, no absolute evaluation of product quality is obtainable through monitoring procedures, however intensive. In addition, testing and monitoring methods do not always parallel or identify the pathways through which contaminants are introduced into the product.

Difficulty in validating microbiologic monitoring methods results from a lack of comprehensive testing standards, reliable test equipment, and reliable methods for correlating sample data to predictions of product quality. Several characteristics and qualities of both contamination events and sampling methods limit the usefulness of microbiologic monitoring as a method of determining the acceptability of a specific product batch:

1. Microbiologic contamination events in controlled facilities are usually not randomly distributed in time, space, or by type of organism;
2. No single sampling method repeatedly recovers a known and consistent percentage of all types of organisms;
3. For most types of contamination detected, there are usually many possible sources, not the least of which are the sampling personnel, equipment, and lab processing; and
4. An extended interval is required for development of results.

Perspectives

These considerations underscore recent concerns that regulatory groups may require that unreliable environmental monitoring data be used as release criteria (15). Current industry standards and regulatory guidelines do not, and should not be interpreted to condone the rejection of batches on the basis of absolute environmental counts alone. Microbiologic monitoring is employed for practical reasons, not because it is ideal or unique in detecting shifts in process conditions.

Regulatory agencies and auditors understandably seek easy-to-interpret data as a basis for decisions regarding product acceptability, and are becoming increasingly hesitant to accept product release in the absence of demonstrable levels of microbiologic control. Conversely, industry is justifiably reluctant to set microbiologic monitoring limits because regulators may misinterpret their meaning in a quality assurance (QA) context. The failure to meet process control limits is quite different from the failure to meet product specifications. Failure to meet a monitoring limit means only that monitoring data can no longer demonstrate validation conditions, and product quality *may* be adversely affected. Enhanced product testing or other corrective actions may be indicated, but batch rejection should not be extrapolated from QA monitoring results, alone.

Setting Limits

In the QA context, limits are established to trigger specific actions, or outcomes. The alert (warning) limit is the point at which the operator should become alerted to the possibility of a deteriorating trend. When an action limit is exceeded, the operator must take action to identify and correct the condition(s) that are causing a verified trend before a "fail" limit is reached and the data fail to indicate process control and support continued production. In a well-designed and executed process, however, such a fail limit should never be exceeded, except in the event of a sudden and catastrophic breakdown of a critical process control component.

Akers noted that values presented in the current U.S.P. (1116) (8) are target values (13). Given this designation, it is reasonable to consider these values to be *operational target levels*, rather than *product quality control limits*. There are several models for setting alert, action and fail limits, although many only establish alert and action limits (14) (other terminology may be used). Extending one current model (14), the alert limit might be considered to be the 95th percentile. Analysis and trending of actual data allow the calculation of this limit, as well as the 97th percentile for the action limit, and the 99th percentile as

the fail limit. Regardless of the model used initially to set limits, they should be based upon both historical data, and an evaluation of correlations between monitoring results and product quality. Data analysis should include a mechanism for evaluation and modification of the monitoring program and limits.

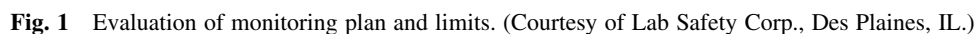
It is expected that results will fall within normally anticipated operating levels (8, 9) (Table 1) with 95% confidence, if randomness in critical environments and operations is sufficiently controlled. If data from successful PQ runs (when the process is demonstrated to be under control) do not meet this criterion, the monitoring methods may not measure a phenomenon that relates directly to process control, may not be sufficiently reproducible to provide useful information, or may have been incorrectly conducted. Every effort should be made to develop monitoring methods that comply with this performance expectation so that data will be useful.

Initial limits may be calculated and compared to results of any unsuccessful trials. These limits should eventually be adjusted based on historical data (see Fig. 1). When evaluating data to adjust limits, Wilson (14) noted, "Including data taken from a period of unusually high counts, where the process was out of control, will lead to inappropriately high alert/action limits."

Conduct of Sampling

Quality management and sampling personnel require both an in-depth understanding of the environmental sampling rationale, and a complete understanding of commonly available equipment, materials, sampling techniques, and development methods. Reporting forms should be carefully designed to convey all relevant information including identification of the technician, sample location (from a standardized sample map), date and time, media (including lot, expiration, and validation date), method, duration of sampling, and equipment (including calibration date and serial number). In addition, information such as the product batch, number and names of personnel, line throughput rate, number and nature of line interventions, and other available monitoring data such as room pressure and other engineering control status readings should be recorded. Any observed deviations from standard operating procedures (SOPs) should be noted and communicated to the individuals responsible for training and management of operative personnel. It is essential to repeat samples when such deviations occur in order to evaluate the impact they may have on results.

Sampling and laboratory personnel must be highly competent on both philosophical and functional levels, and



must develop and exercise *perfect* aseptic technique (13). A training program and operating procedures should be established defining all monitoring steps, including gowning, preparation of samplers, aseptic sampling techniques, sample recovery, handling and transport, and laboratory techniques for aseptic sample development. A laboratory QA program should assure that monitoring personnel conform to operating procedures and that technician skills are periodically tested and validated for high competence and flawless technique (13).

Sample Handling

Sampling, sample transport, and sample development should be conducted in a way that does not affect results. For example, if agar plates are improperly transported, condensate may form on the lid and drip onto the agar surface, redistributing microorganisms over the surface and around the edges of the plate, causing false readings. Agar plates should, therefore, be kept inverted and oriented horizontally during storage and transport. They should be handled gently, and transferred to the incubator as quickly as possible after exposure. With sieve impactors, false positives can usually be identified as colony forming units (CFUs) that fall outside the star pattern of jet indentations in the agar surface below the holes. Counts >20 CFUs may also be statistically corrected for increased accuracy by using the positive-hole correction table (15).

It is recommended that colony counts be made at several points in the incubation process, with separate tallies for bacterial and fungal colonies that tend to merge at a critical point during incubation, when fungal colonies may overgrow and obscure bacterial colonies. For this reason, any bacterial subcultures should be made prior to the onset of rapid fungal growth. Whenever possible, optical electronic colony counters with sufficient back-lighting and magnification to enhance contrast and enumeration should be employed to increase accuracy. In the presence of known or potentially high counts, the microscope enumeration method should be used to closely differentiate and count microcolonies in impact areas on sample plates following a short incubation period.

DEVELOPING A MONITORING PLAN

Site Selection

A critical site is a point at which the product is exposed to the environment, when something is added to the product

or product pathway, or a point at which unprotected product is manipulated. Any intervention into the process line increases the potential for contamination. (Examples of line interventions include the introduction, removal, or manipulation of materials and product, equipment adjustments, and sampling activities.) Particular attention should be given to these sites and events in the development of the monitoring plans (3).

Analysis of a videotape of repetitive prequalification should be studied for behavior and practices that may produce or harbor environmental contamination, leading to the refinement and optimization of work practices, and development of the formalized process to be instituted for the PQ validation run. The videotape may be used for identifying indicator sites, which should be incorporated into the monitoring plans, and intensively sampled during the validation run. These tapes should be retained and edited for both training and informational purposes.

For critical processes, it is important to select noninvasive sampling methods that have high collection efficiency for a broad range of organisms. To select the most suitable monitoring methods and equipment, the probable route of contamination for each critical site or process should be identified. For example, when the most likely route of potential contamination is touch, select surface sampling techniques for personnel barriers. When the most likely route is transfer from contaminated work surfaces, sampling of these surfaces is most useful. At sites where unprotected product is exposed to the environment, aerobiological monitoring is indicated, and, in unidirectional airflow, must be carried out isokinetically and isoaxially^b in the manner of nonviable particle-count testing. Some processing steps may require multiple sampling methods.

Controlled support areas adjacent to critical areas are the essential interfaces in the transition from the general environment to the aseptic processing core. These areas should be adequately pressurized, facilitating a gradient flow of contaminants from cleaner to dirtier areas (10). Controlled staging, support, material storage areas, and work practices should be examined and indicator sites identified. Controlled areas should be maintained and monitored in accordance with guidelines and industry standards (Table 1).

^bIsoaxial: A condition of sampling in which the direction (axis) of the airflow into the sampling probe inlet is the same as that of the unidirectional airflow being sampled (7). Isokinetic sampling: The condition of isoaxial sampling in which the mean velocity of the air entering the probe inlet is the same as that of the unidirectional airflow being sampled (7)

Personnel, Equipment, and Facility

Validation and monitoring of a process are normally divided into three main areas of concern: personnel, equipment, and facility.

The human factor is the greatest potential variable in any process. Uncontrolled variation in personal health and hygiene, barrier techniques, and aseptic technique may cause wide variation in contamination of controlled support areas and process materials during staging and preparation, as well as adventitious contamination of the aseptic process core and product. A suitable aseptic process, defining appropriate and standardized personal hygiene expectations, scrubbing and preparation techniques, barrier techniques, and operator techniques should be developed and challenged intensively during the PQ exercise. Personnel should periodically take both written and media-fill skill tests (3).

Ongoing monitoring for compliance with pertinent SOPs should then be conducted. Sampling of personnel barriers, such as gloves, shoe covers, hair cover, and gowns facilitates detection of potential “fallout” contaminants shed from personnel for evaluation of both barrier and aseptic techniques. This information may be useful in establishing required garb-change intervals, based upon measured garb-penetration times by endogenous contaminants. All accumulated data should be used periodically to develop a facility trend analysis which, in turn, modifies training and work practices as necessary.

All equipment used in controlled manufacturing or compounding processes should be designed, staged, and sanitized in a manner that facilitates unvarying routine operation, with minimal human intervention. This reduces the potential for random cross-contamination by operative personnel. Improperly sanitized or sterilized equipment or components are also a possible source of contamination.^c Monitoring of representative surfaces of process equipment should be carried out and documented.

Facility sampling should be carried out under both as-built and at-rest (7) conditions during initial installation qualification (IQ) and operational qualification (OQ) of the facility, in order to baseline and “bracket” performance of the engineering controls, and to identify the normal background flora present in the manufacturing environment. Sampling should then be conducted in-process

under operational conditions (7) during the PQ, to identify the impact of the process and personnel on the product and environment. It is important to monitor the validation process during all shifts and throughout the shift. Sites should be standardized and selected by statistical models or grid profiling (16), based upon testing and monitoring requirements appropriate to the specific process (Table 1).

Surface sampling is useful in verifying the effectiveness of housekeeping and sanitizing procedures. It may also provide an alert to poor materials preparation prior to introduction into the controlled environment, or to lapses in personnel technique or barrier use. Aerobiologic sampling is most useful when conducted in conjunction with a complete program for testing of the engineering control system (10). Recommended tests include the following:

1. Facility pressurization, which should be routinely monitored at recommended intervals (20);
2. High efficiency particulate air (HEPA) filter velocity and uniformity testing for laminar airflow (21), and volume in cubic ft/min (CFM) for conventional flow, including a determination of room installation air changes (10);
3. HEPA filter leak-integrity testing (18);
4. Nonviable particulate cleanliness testing (7); and
5. Smoke-tracer visualization for establishing the integrity of unidirectional-flow areas (10).

Periodic retesting of challenges 1–4 is required by some regulatory groups, with the interval determined by the nature of the process and product in a given area (17). Repeating Test 5 may be useful in evaluating failures and can be an extremely valuable training tool. Concomitant particle count testing may be useful in identifying contamination indicator sites.

Monitoring of laminar airflow workstations (LAFWs) requires a complete understanding of HEPA filtration system performance, and is frequently conducted in ways that do not yield useful information. When properly validated in accordance with Federal Standard 209b [Appendix A, para. 40 and 50^a], LAFWs provide air at the entrance plane which is far cleaner than Class 100 (10). Testing to this cleanliness level would permit particulate contamination levels two orders of magnitude greater than during filter OQ validation testing. More important, the use of any apparatus that samples discrete locations in a unidirectional slip stream is unlikely to detect filter leakage because isoaxial and isokinetic sampling at the exact point of leakage would be required. Therefore, placement of a sampling probe upstream from the product is unreliable and an unnecessary threat to sterility. The only practical, in-process use of these

^cThe sterilization process for any equipment or supplies that are sterilized prior to introduction into the controlled environment must be validated, with sterilization records and verifications included in all product batch histories. Validation of sterilization equipment, alone, is not sufficient to assure sterility. Because the types of materials being sterilized, and the arrangement of articles within the sterilizer can effect results, standardized load configurations must be developed and validated

instruments is to detect shifts in the amount of particles and microbiologic contaminants caused by the process at some point adjacent to or downstream from the product. Such a shift might signal a lapse in personnel technique, barrier use, or prestaging material preparation, or be caused by HEPA filter loading, which reduces airflow velocity.

Avoiding Sampling-Induced False Positives

Line interventions for sampling purposes must be balanced carefully against the total number of interventions necessary for production purposes. Sampling should present the minimum risk of contamination, which is theoretically the same for every line intervention. Because sampling-induced positives should not exceed 10% of total positives ($10^{-1}N_p$) (17), the number of sampling interventions should be significantly lower than the number of production line interventions. In isolators or other isolated critical processes, where no line interventions occur during production, not more than one, carefully controlled, aseptic sampling intervention is recommended.

Surface sampling the exterior of finished products, as indicator sites, assembled from purportedly sterile components as they exit the process while still under aseptic conditions, may be a more efficacious method of estimating microbiologic contamination potential than invading the critical production site. This method allows sampling the most critical site adjacent to the product, and more sites may be noninvasively sampled over a longer interval. In addition, this method may substantially reduce the incidence of sampling-induced contamination.

Monitoring Frequency

The frequency of monitoring should be determined by the maximum interval acceptable for an over-limit condition to remain undetected (19). This depends upon the critical nature of the process within the monitored area. In general, the minimum frequency should be consistent with applicable regulatory guidelines (Table 1). Although it has been suggested that monitoring frequency can be reduced if no over-limit condition is detected within a predetermined number of monitoring cycles, this practice is inconsistent with basic monitoring rationale. Monitoring is conducted to detect a breakdown in process controls, which may occur at any time. Even if no control component has failed for a prolonged period, it must be assumed that a failure will occur eventually and must be detected within the predetermined interval. In addition, lack of over-limit test results may be due to the fact that

monitoring method(s) are not sufficiently sensitive, or that limits are too high.

EVALUATION OF THE MONITORING PLAN AND LIMITS

Most discussions of microbiologic monitoring recommend that the monitoring plan and limits be based on historical data, but offer little guidance on how this can be accomplished. Figure 1 provides a guide for evaluation and revision of the monitoring plan and limits. An in-depth evaluation may be triggered by over-limit results from monitoring (Entry Point 1) or by adverse product testing results without detection of any over-limit condition through routine monitoring (Entry Point 2).

Entry Point 1:

1. Conduct routine monitoring. A counter (*C*) is used for Step 11. $C = 0$ at the beginning of the routine monitoring program.
2. If the results do not exceed any limit, then continue routine monitoring.
3. If the results exceed any limit, then perform retesting in triplicate to verify the accuracy of results. Retest under the same conditions noted on the sampling form (i.e., same time of day, same location and operator, same type of production).
4. If triplicate retest results are not over-limit, it is assumed that the original over-limit result was due to a nonassignable cause (*NAC*). Determine the probable cause of the over-limit count (i.e., unusual activities noted on test documentation, sampling, lab error, etc.). A record of positive *NAC*s should be kept and analyzed to determine ways to improve affected processes and sampling procedures. Return to routine monitoring.
5. If results are over an alert limit, but not over the action limit, then enhance monitoring frequency for *X* cycles. (*X* is determined by the critical level of the area and process where the over-limit event occurred, but should provide an adequate interval to assure detection of a continued deterioration of process control.)
 - a. If the alert limit is not exceeded again within *X* cycles, then return to Step 4.
 - b. If the alert limit is exceeded but the action limit is not, then proceed to Step 10.
 - c. If the action limit is exceeded, then go to the corrective action plan (*CAP*) (5).
6. If results following the triplicate retesting are over the action limit, but not the fail limit, then go to the *CAP*.

7. If the results following the triplicate resting are over the fail limit, traditional QA protocols usually require that operations cease. However, the appropriate action taken should depend on the critical nature of the monitored step and other conditions. An alternative to operation shut down may be to segregate and hold the product for enhanced testing for adverse effect; go to the CAP.
8. If implementation of the CAP results in the determination of the cause of the over-limit condition, then correct the condition, and retest in triplicate to verify that the problem was corrected. If no cause was found, then proceed to Step 10.
9. If test results following corrective action are within limits, then return to routine monitoring.
10. If test results following corrective action are still over-limit, or if no cause of the over-limit condition can be identified, then evaluate the product for adverse effects.
11. If no adverse impact on product quality can be detected, add 1 to the counter (*C*). The result may indicate that limits are too low, but one event is not sufficient to support a decision to increase limits.
 - a. If $C = 3$, then the limits are too sensitive, and should be increased.
 - b. If $C < 3$, return to routine monitoring. Because the results are over-limit at this point, a repeat investigation of the cause of over-limit results will be triggered. Limits should be increased judiciously, and it is important to be thorough in attempting to resolve any cause of over-limit testing with reasonable certainty before increasing limits. For example, if the cause of the over-limit result is sampling mistakes or lab error, there will be no detectable cause in the production facility, the process or engineering control evaluations, and probably no adverse effect on product quality. This should not, however, be interpreted to mean that limits are too sensitive.
12. If product is adversely affected, and no cause can be detected following implementation of the CAP, the monitoring plan and/or the process should be redesigned and revalidated.

Entry Point 2:

13. If product quality is below limits, but monitoring data did not detect the shift, then reevaluate monitoring data using lower limits to determine whether or not the process shift could have been detected. If the data have been graphically represented, this should be

quite simple; increasing the amplitude of the graph may be useful.

14. If lower limits would have detected the shift, then lower the limits and institute the CAP.
15. If lower limits would not have detected the shift, then evaluate the cause of the failure, and develop a new sampling strategy for the key step(s) where failure occurred. Institute the CAP and verify that corrective actions taken were effective in improving product quality.
16. If product quality improves, then add the new sampling method to the routine monitoring program.
17. If it does not, return to Step 12.

SELECTION OF MONITORING METHODS, MATERIALS, AND EQUIPMENT

Effective microbiologic monitoring of controlled processes usually includes sampling of process air for aerobiologic contamination, and facility, equipment, and operative personnel barriers for surface contamination. Equipment and methods used in monitoring procedures must be carefully considered for attributes and limitations and must be matched to sampling objectives to ensure that methods and techniques are noninvasive, and to facilitate development of well-organized sampling plans, techniques, data, and data trending analysis.

Surface Sampling

Surface sampling may be performed at the conclusion of critical operations to minimize disruption of these processes (8) and prior to sanitizing procedures (20) to estimate cumulative, inprocess contaminant burden (4). In addition, presanitization surface sampling is beneficial in detecting operations-induced bioburden and cross-contamination between environmental and equipment surfaces. Postsanitization surface sampling is useful for evaluating sanitizing methods and in retrieving sanitization-resistant isolates for identification and trend analysis in demonstrating sanitizing compound efficacy. The two most common types of surface sampling are swab-sampling, and surface contact sampling.

Swab-sampling

Swab-sampling is normally used for flat or irregular, nonabsorbent surfaces with qualitative development by inoculation of the swab matrix directly into nutrient broth, observed for growth/no growth. Quantitative development is also possible (5). The main advantage of the swab

method is accessibility to difficult-to-reach equipment surfaces and areas of the production environment. Limitations are excessive time consumption, increased potential for adventitious contamination due to the cumbersome nature of the procedure, and failure of enumeration processes to correlate to full recovery of organisms.

Contact plates

Surface contact plates are normally used for sampling flat or irregular, absorbent or nonabsorbent surfaces. The surface contact plate consists of a clear plastic base housing a convex protrusion of nutrient agar with a plastic cover. Sampling is accomplished by pressing the agar against the site.

The covered plate is then incubated for development, and the CFUs per square centimeter enumerated (21). Advantages of surface contact plates are reproducibility, speed, simplicity of collection mechanism, and minimized potential for adventitious contamination; collection and correlation to recovery of organisms are superior to swab-sampling.

Aerobiologic Sampling

Aerobiologic sampling is conducted in critical and controlled areas to detect airborne viable contaminants present during manufacturing operations. Aerobiologic sampling procedures, frequency, and limits should be established based upon environmental conditions required to maintain product quality, and established for each processing step (4) (Table 1). Aerobiologic sampling employs two basic methodologies:

1. The gravity settle plate, which provides passive measurement of microorganisms likely to deposit by sedimentation at critical and controlled sites within a given period, and
2. The volumetric air sampler, which provides active measurement of viable contaminants by mechanical aspiration and dynamic inoculation of process air.

Gravity settle plates

The gravity settle plate measures microorganisms settling from the air onto a known surface area in a known time. Settle plates may be positioned within the critical area at indicator sites where the product may become exposed to airborne contamination, and in controlled areas at locations identified as likely sources or areas of "fallout" aerobiologic contamination. Settle plates are not appropriate aerobiologic sampling method for monitoring the efficiency of unidirectional (laminar)

airflow or other air-cleaning devices. This is based upon studies (22, 23), and the general assumption that "... the settling velocity of contaminants (in unidirectional airflow) is negligible, which implies that gravitation plays an inferior role. With the assumption of a constant value of the diffusion coefficient, the diffusion equation in a velocity field within rectangular coordinates becomes

$$\frac{\partial c}{\partial t} + v_x \frac{\partial c}{\partial x} + v_y \frac{\partial c}{\partial y} + v_z \frac{\partial c}{\partial z} = D \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2} \right) \quad (1)$$

where c is concentration: v_x , v_y , v_z are velocities in the x , y , and z directions; and D is diffusion coefficient.

This gives the simplest possible mathematical model which describes a system with regard to transport of contaminants emitted in a source of an arbitrary position..." (23), demonstrating that particle dispersion in undisturbed streamlines is primarily a function of streamline uniformity and velocity. Disruptions of the parallel (laminar) airflow streamlines caused by equipment, personnel movement, and product result in turbulent flow, creating small and temporary vortices and eddies. It is only turbulent diffusion within the vortex that causes removal of entrained contaminants (23). Therefore settle plates, strategically placed, are reported to provide a superior method of predicting potential product contamination by mimicking the deposition of microbe-carrying particles (MCPs) into or onto the product^d (24). They are inexpensive, may be used to continuously monitor the entire production interval, are less invasive of aseptic operations, and may usually be placed closer to exposed products than volumetric air samplers.

Settle plates cannot be used for quantitative measurement of airborne microorganisms because the sample volume of sedimentation air samples cannot be measured. Air turbulence around an open plate may also effect collection results, and smaller particles may not settle at all (22). In addition, extended exposure times may result in

^dRegardless of placement of an aerobiological sampler in a laminar airflow work zone, it can at best measure the effect of the process at some point downstream from the product. For example, the mouth of a flask may be situated in "first air" issuing from the HEPA filter, while air impinging on the surface of a plate adjacent to it will be affected by disruptions of the airstream caused by the flask. Contamination found on the plate then results from a different set of conditions than those to which the product is subjected and does not exactly parallel the product contamination mechanism. Only a media-fill process simulation can fulfill this function. Aerobiologic sampling immediately downstream of the critical orifice can, however, detect downward shifts in the overall cleanliness of the critical process air, which in turn may indicate increased contamination potential near the product

some desiccation of the nutrient agar, resulting in poor microbial growth (25).

Volumetric air samplers

As an active sampling method, the volumetric air sampler aspirates a known volume of process air, capturing microorganisms into or onto a nutrient agar medium, a liquid, or a filter. Microorganisms are developed and quantified as an estimate of CFUs present in the sampled environment per cubic foot of air (or other volumetric measurement) (4). The quantitative principles of volumetric (active) air sampling may be expressed by

$$S(R_t)C = R_f \quad (2)$$

where S is source intensity, R_t is transport rate, C is correction factor, and R_f is failure rate.

Volumetric air sampling is accomplished by a number of different methodologies, including impingement, impaction through single or multiple orifices, centrifugal impaction, and filtration. Each method has inherent advantages and disadvantages that affect the value of the data collected relative to the specific application. Table 2 presents a comparison of popular samplers based upon relative cost, difficulty of use, appropriate applications, and other factors.

Impingement: In an impinger, a known volume of air is drawn through fluid in a glass vessel (20, 30). Particles separate from the airstream by impinging at the flask bottom, where they are stopped and retained by the liquid as the air continues to flow out through the pump system. High air velocities passing through the impinger effectively break up bacterial/particulate aggregates, resulting in microbial counts, which more closely reflect the actual number of microorganisms, leading to recommendations that impingers be used as the standard reference method for monitoring aerobiologic contamination (26, 28). However, impingers may require the addition of antifoam agents and replacement of fluid, due to agitation and evaporation loss during longer sampling procedures.

These additional steps increase the possibility of adventitious contamination. It has been demonstrated that the sampling efficiency of an impinger is dependent upon both system design and the particle sizes being sampled (29). Accuracy and reproducibility of results have been reported to be difficult, and particles of $<5.0 \mu\text{m}$ have been demonstrated to pass through the impingers tested (30).

Impaction: In slit-to-agar (STA) or sieve impactors, a known volume of air is aspirated through a single orifice

(STA), or multiple orifices (sieve), and viable particles, due to their inertia, are forced out of inlet airflow streamlines and impacted onto perpendicular, target nutrient agars as the streamlines abruptly change direction to bypass the target stage. In the centrifugal impaction sampler, high centrifugal forces created by “spinning” air through an impeller turbine at sufficient velocities to cause separation of microorganisms from sample air streamlines result in their impaction onto a nutrient agar strip placed at the inner periphery of the sampling chamber, parallel to the inlet airflow axis.

Sieve impactors are available in single-stage or multistage designs that facilitate both enumeration and sizing of aerobiological contaminants. As the sample air transits the device, sample velocities increase at each stage, resulting in gradient deposition and accurate sizing of microorganisms of smaller diameters and lower mass. Microorganisms aspirated by sieve samplers through a matrix of multiple-inlet orifices impact directly onto an agar medium for development from a single agar plate for each vertically stacked stage, with no further subculture steps required for enumeration. Advantages of sieve samplers are generally high particle deposition rates, the ability to size particles and vary sampling time and volume, and superior collection efficiencies when compared to other methods of aerobiological testing. Single- and six-stage configurations have been reported to be two of the three sampling methods of choice (27).

Use of STA samplers in isolators and critical process zones should be accomplished using a sterile sampling hose and probe, facilitating remote location of the sampler in a noncritical area. In monitoring a unidirectional slipstream, this hose/probe configuration should be both isoaxially oriented, and isokinetic^b, in order to minimize disruption of the slipstream. Advantages of the STA include the ability to revolve the plate at varying rates so that the samples may demonstrate changes in aerobiological concentrations directly over time, and the ability to obtain multiple samples with a single petri dish (31). STA samplers have historically been the standard against which other air samplers are assessed (17, 32). Agar plates are easily removed from the sampler for development, with contamination enumerated as CFUs per unit of air sampled.

The STA is reported to be both unsuitable for use in the presence of high concentrations of organisms (33) and cumbersome to use (17). In addition, it has been demonstrated that a significantly higher percentage of particles sized $0.5\text{--}0.8 \mu\text{m}$, and a significantly lower percentage of particles sized $3.0\text{--}25.0 \mu\text{m}$, were present in sample air, which had passed through the slit of an STA,

Table 2 Relative cost/difficulty comparisons

Sampling method/sampler	Acquisition cost ^a	Cost of Use/Sample 1–6 ^b	Ease of Use 1–6 ^b	Sampling Speed 1–6 ^b	Mobility 1–6 ^b	Contamination Potential to Sample	Contamination Potential to Environment	Reproducibility 1–6 ^b	Applications A-C	Isolators 1–6 ^b
Swab Sample Typical	●	5	2	2	2	4	1	2	C	4
Contact Plate Typical	●	2	1	1	1	1	1	1	B	1
Gravity Settle Plate Typical	●	1	1	1	1	1	1	3	A	1
SAS Super 90 Air Sampler	5	2	2	2	2	2	2	2	A	3
STA New Brunswick	5	1	3	2	4	2	3	2	A	1 ^c
Sieve Impactor Andersen 1-STAGE	2	1	3	2	2	2	3	1	A	3
Centrifugal Biotest RCS Plus	3	3	3	2	2	4	2	3	A	6
Sieve Impactor Anderson 6-STAGE	5	4	6	6	3	4	3	2	A	5
Gel Membrane Sartorius MD8	4	6	2	1	1	2	2	1	A	1 ^c
SMA P200 Impactor	6	1	2	2	2	2	3	2	A	4
Glass Impinger All Glass	6	2	4	3	4	4	3	2	A	4

Table 2 Relative cost/difficulty comparisons

Sampling method/sampler	Laminar Airflow 1–6 ^b	Critical Environments	Production Areas 1–6 ^b	General Areas 1–6 ^b	Flat Environmental Surfaces	Irregular Environ. Surfaces	Personnel Barriers 1–6 ^b	Volumetric (SP) Y/N	Remote Probe Possible	External Power	Sample
Swab Sample Typical	4	2	2	1	2	1	4	N	●	●	S
Contact Plate Typical	1	1	1	1	1	4	1	N	●	●	S
Gravity Settle Plate Typical	1	1	1	1	●	●	●	N	●	●	S
SAS Super 90 Air Sampler	2	3	2	2	●	●	●	Y	N	Y	S
STA New Brunswick	1 ^d	3	1	2	●	●	●	Y	Y	Y	S
Sieve Impactor Andersen 1-STAGE	2	2	1	1	●	●	●	Y	N	Y	S
Centrifugal Biotest RCS Plus	5	5	3	1	●	●	●	N	N	N	P
Sieve Impactor Anderson 6-STAGE	5	2	2	2	●	●	●	Y	N	Y	S
Gel Membrane Sartorius MD8	1 ^d	1	1	1	●	●	●	Y	Y	Y	P
SMA P200 Impactor	3	3	2	1	●	●	●	Y	N	Y	S
Glass Impinger All Glass	4	3	2	2	●	●	●	Y	N	Y	S

^aAcquisition cost in thousand dollars.^bDifficulty: 1–6 (easiest–hardest).^cWith Hose/probe attachment.^dWith Hose/isokinetic probe attachment.

A: Aerobiologic samples; B: Flat surface samples; C: Irregular surface samples; P: Proprietary media system; S: Standard Commercially-available system.

Table courtesy of Northview Biosciences Inc., Northbrook, IL.

than were found in ambient air (34). This was attributed to fragmentation of larger particles following passage through the slit of the STA.^c

Due to dehydration of the agar reported to occur over long sampling periods, continuous sampling exceeding 30 min using an impaction sampler is not recommended. Areas of loss have been reported for sieve samplers (40), including *inlet loss* (the effect of cross-wind at the sample inlet point), *interstage loss* (deposition of particles on internal surfaces other than the impaction agar), and *particle re-entrainment* (particles reintroduced into the airstream due to particle “bounce,” resulting from dehydration of the impaction agar).

Advantages of the centrifugal sampler are the capability of sampling large amounts of air (40 L/min) in a short time; it is quiet, lightweight, self-contained, and does not require cumbersome air pumps or external power for operation. Centrifugal samplers provide a good indication of environmental isolates (17).

Centrifugal sampling cannot be carried out isokinetically (23), and the accuracy of results is dependent upon the sizes of the particles being sampled. Since particulate sizes in the air volume being sampled are not routinely determined, the validity of the centrifugal sampler as a quantitative device has been called into question (36), especially for quantification of small particles (27, 37). Another recent study indicates that centrifugal sampling causes air to move in a turbulent, mixing manner, introducing heavily disturbed airflow patterns around the sampler which may, in turn, impart disturbances to any unidirectional airflow patterns being sampled (23). Reaspiration of sampled air is also a problem with earlier designs, creating difficulty in discriminating between incoming and outgoing airstreams, which is necessary to quantify microorganisms (38). Proprietary agar medium strips are specially designed and unique to this system, and require careful technique to insert and remove aseptically.

Membrane filtration: Membrane filtration (MF) sampling is accomplished by capturing aerobiological contamination as it passes through a cellulose membrane filter (CMF) or gelatin membrane filter (GMF). The mechanisms of MF particle removal are inertial impaction, diffusional interception, and direct interception. Following collection, the GMF may be plated aseptically onto an agar petri dish to dissolve, allowing microorganisms to grow directly on the nutrient medium. Dissolution of the membrane into a sterile solution is also possible (31).

^cInterestingly, this attribute was reported by investigators to be an advantage of the all-glass Impinger (31, 32).

While MF sampling has been demonstrated to be the most effective means of retaining aerobiological contamination, CMF sampling exhibits a lower recovery rate than an impinger when tested against stress-sensitive microorganisms, such as *Serratia marcescens* (39) or *Escherichia coli* (27) due to desiccation on the CMF surface. Studies have indicated that gelatin foam filters incorporated into GMF gave significantly higher recovery rates than CMF over the same sampling period (40, 41). Recent comparisons of sampling systems indicate that GMF is equally as effective as the STA sampler, irrespective of particle size, and is significantly more effective than centrifugal sampling in the collection of microorganisms with sizes <5.0 μm (31). A recent study comparing the GMF system with centrifugal, sieve, and STA systems in sampling the unidirectional airflow slipstream in the presence of visual tracers indicates the GMF sampler to be the only sampling method capable of isokinetic and isoaxial sampling with no visual disturbance to the laminar airflow pattern (31). However, in this study, the STA was tested without the remote hose-isokinetic probe device.

Limitations of the GMF are an additional aseptic subculture step, which increases the probability of adventitious contamination, and a proprietary membrane filter, which results in a per-sample cost currently exceeding 12 times that of the one-stage sieve, SAS, STA, SMA, and glass impinger systems, and four times that of the centrifugal sampler.

Growth Media

Growth and collection media used in microbiologic monitoring should be selected on the basis of the target organisms, areas and surfaces sampled, and inhibitory residues that may remain on the sampled surfaces. Media commonly used for environmental monitoring are listed in Table 3. Under certain circumstances (e.g., when obligate anaerobes are recovered from the product), additional, specific media and methods should be selected by a qualified microbiologist (17).

Comparison of Aerobiologic Samplers

The different characteristics and operating principles of aerobiological samplers do not facilitate direct and simple comparisons. The user should, therefore, carefully evaluate the numerous advantages and disadvantages of each method in selecting a sampler for the intended application (42) (Table 2). Two studies that provide basic comparisons of aerobiological sampling systems may offer useful information:

Table 3 Media commonly used for environmental monitoring

Medium	Selective for	Sample application
Tryptic soy agar (TSA) ^{a,b}	Aerobes and facultative anaerobes	Air and surface
Lethen agar ^c	Aerobes and facultative anaerobes	Surface
DE neutralizing agar ^d	Aerobes and facultative anaerobes	Surface
Sabouraud dextrose agar	Yeast and molds	Air and surface
Rose bengal agar	Yeast and molds	Air and surface
Buffer solution ^e		Surface

^aTryptic soy agar is also known as soybean casein digest agar.

^bUnmodified general purpose medium use for culturing bacteria and/or fungi.

^cContains additives used to neutralize residuals of halogen-based disinfectants, such as sodium hypochlorite (bleach).

^dContains additives used to neutralize residuals of halogen and quaternary ammonium chloride-based disinfectants.

^eSamples collected using sterile swabs and buffer solution must be transferred to media for culturing and enumeration.

(Courtesy of Northview Biosciences, Inc., Northbrook, IL.)

A study comparing eight bioaerosol samplers was carried out by Jensen et al. in 1992 (27). Results indicated that the Andersen 6-STG, I-STG, and Ace Glass AGI 30 samplers were the samplers of choice for recovering aerosols of free bacteria (i.e., mostly single cells of *E. coli* and *B. subtilis*, $d_{ae} \geq 2 \mu\text{m}$) under the controlled conditions of the study (43). Another study, comparing seven samplers commonly used in controlled environments, was conducted by Ljungqvist and Reinmüller in 1998 (44). This study indicated widely varying results for the impaction samplers tested. The limited number of parallel tests performed prevented an evaluation of comparative collection efficiencies based upon statistical considerations. The salient recommendations of this study are that results should be seen more "... as an indication of a [contamination] level and not be taken as a true absolute value," and that aerobiological samplers be selected carefully, based on practicalities of using different types for different locations or situations. Furthermore, this study recommends the simultaneous use of a discrete

particle counter (DPC) to measure the total number of airborne particles present in the area sampled.^f

ANALYSIS AND INTERPRETATION OF MONITORING RESULTS

Effective interpretation of data from microbiologic monitoring of the environment can be the most difficult aspect of the monitoring process. Several factors complicate this process, including the inherently nonrandom distribution of most microbial contamination events, errors in sample handling, variation of sampling technique from one monitoring event to the next, and seasonal shifts in the type and level of contaminants likely to be present in the general environment.

The purpose of statistical evaluation of sample data is to extrapolate from a collection of individual events (e.g., 30 min of process time) to the entire population of events (e.g.,

^fBecause it is impossible to derive instantaneous results from microbiologic testing, the authors agree that such data should, where possible, be correlated with a DPC as an instantaneous data source, in developing useful historical data. Although "... no universal relationship has been established between the total concentration of airborne particles and the concentration of viable airborne particles..." (7), such a correlation may be possible under controlled operational conditions within a specific area or facility (determined by Ljungqvist and Reinmüller in two facilities) to be approximately 10^{-4} (10,000:1). Such a correlation would facilitate a "viability index" as a rational means of correlating shifts in instantly available particle count values with probable corresponding shifts in aerobiological contamination. This technique would be very useful in the instantaneous identification of contamination indicator sites. Similar correlations have been established on a facility-specific basis by the Lab Safety Corp. during the course of regular, periodic aerobiological sampling of three bone marrow transplant

complexes over a period of several years. In all cases, a correlation of the total population of aerobiological contaminants to instantaneous DPC data (termed the "viability index") was used to trace the distribution and probable presence of the life-threatening organism *A. niger* in immunocompromised patient populations. Although it was found that in these highly controlled facilities (e.g., Class 1000 or better) the correlation was one to two orders of magnitude higher than that described by Ljungqvist and Reinmüller (42,44), the correlations were consistent, allowing facility managers to reliably detect possible life-threatening deteriorations of the critical patient environment through the use of instantaneous DPC data as an aerobiological contamination indicator. The data analyzed indicate that aerobiological contaminants appear to increase in proportion to nonviable contaminants as the cleanliness of a facility increases. This is probably due to the fact that, as general environmental contamination is eliminated, human activity becomes the principal source of contamination.

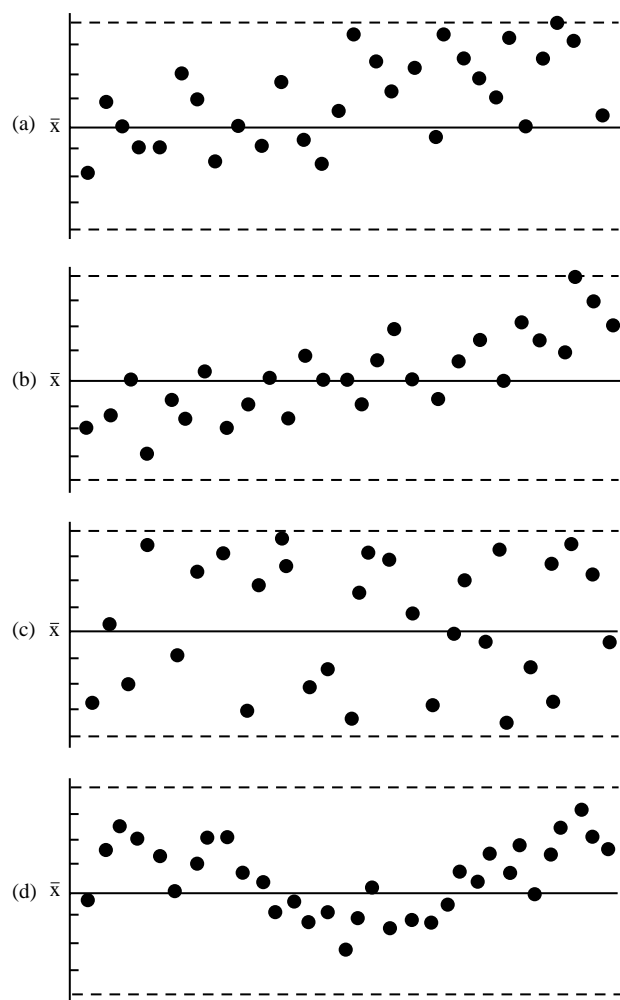


Fig. 2 Out-of-control pattern recognition. (a) Change or jump in level; (b) Trend or steady change in level; (c) Two populations; (d) Recurring cycles. (Chart courtesy of Prentice Hall, Inc., Upper Saddle River, NJ.)

8-h shift). Because microbial monitoring data usually measure the impact of human activity, which is not reproducible exactly from one event to the next, results usually do not fit standard statistical models for normal distributions. In spite of this limitation, it is necessary to summarize the data for comparison to limits. The best statistical methods of evaluation are determined by the nature of the data (14). Wilson suggests that microbial monitoring data histograms generally resemble Poisson or negative exponential distributions (14), whereas Akers points out that Poisson distributions may only be appropriate for systems with minimal human intervention (12). The formula for the Poisson distribution (45) is given by

$$P(C) = \frac{(np_0)^C}{C!} e^{-np_0} \quad (3)$$

where C is individual sample count, np_0 is average count, and $e = 2.718281$.

Trend analysis of results at individual sample locations may be more useful than statistical analysis of data summaries because each sampling location probably reflects a unique situation. Nontraditional groupings of data may also be valuable. For example, grouping all locations where a specific activity was noted on the sample collection form, grouping all data collected during a specific time frame (i.e., just after lunch, or near the end of a production cycle), or grouping all data for each operator may reveal specific problem areas.

The example control charts presented by Besterfield in Fig. 2 demonstrate four major types of out-of-control patterns (45). A fifth pattern is due to mistakes, which will usually show up as isolated, out-of-control points. All apply equally to production and sampling operations. All patterns may be observed on both range (R) charts and standard (or reference) process average charts, but are usually more common to charts.

Likely causes for each type of pattern can be identified, and a checklist of assignable causes applicable to the particular process should be developed through cause and effect (C&E) analysis (45). Examples of likely causes for these patterns are:

- A change or jump in pattern caused by an inexperienced operator, a change in raw materials, or a failure of an equipment part;
- A trend or steady change in level due to a gradual change in the production environment, a gradual change in equipment performance (e.g. HEPA filter loading), or a gradual tendency toward lax observation of SOPs;
- Two populations may be due to more than one process line or piece of critical equipment on the same chart, more than one operator on the same chart, or different samplers or sampling techniques; and
- Recurring cycles may be caused by periodic operator rotation, operator fatigue and rejuvenation cycles, sanitizing and cleaning cycles, and seasonal shifts.

Recurring cycles may be missed if sampling intervals happen to coincide with the cycle frequency, in which case only the low or high range of the cycle may be detected. Out-of-limit trends near the lower limits of the R chart represent superior performance and should be analyzed to identify methods of maintaining these process levels (45). Whatever statistical methods are employed for summarizing data, graphic representations, such as histograms and process control charts can be extremely useful for detecting trends or cyclic patterns in test results.

There are two types of over-limit results: *Random* results are due to chance (unassignable) causes, whereas *nonrandom* results are due to assignable causes. For a controlled process and facility, the objective is to differentiate between individual data points that are assignable and those that are not. If the individual over-limit event is not repeated during subsequent, multiple retests, it is not assignable and does not represent a deteriorating trend. All statistical evaluation methods include mechanisms for “discarding” spurious data. There is, however, a cause for any unassignable result, and efforts should be made to identify and understand it. All data have meaning, and may be useful for improving the process or testing procedures.

Speciation

Speciation of microorganisms is indicated when product testing results detect the presence of a specific organism, when evaluating the efficacy of sanitizing compounds and routines, and when monitoring results trigger the corrective action plan. Speciation should be carried out and analyzed by a qualified microbiologist familiar with the sampling equipment, sampling methods employed, and the origins of organisms commonly found in cleanrooms (17,46). Speciation should also be conducted periodically to identify isolates normally recovered when the process is operating within limits, and may be useful in identifying the probable cause(s) of any out-of-limit condition. During the initial phase of the corrective action plan, an analysis of probable contamination sources and routes should be made for all organisms identified (17). Information obtained by speciation may immediately indicate the most likely source. This information may also indicate less common contamination sources, such as perverted cleaning solutions.

Periodic re-evaluation of the monitoring plan should be carried out, and seasonal effects considered in trend analysis. Many sampling methods do not collect all organisms with equal efficiency, and organisms likely to be present may vary seasonally. Any seasonal shift (up or down) should be investigated by speciation, and sampler correction factors for the predominant organisms applied.

CORRECTIVE ACTION PLAN

The CAP should clearly define and document

1. The method of data analysis;
2. Alert, action, and fail limits;

3. Corrective actions to be employed in the event of detection of a deteriorating trend or an over-limit condition; and
4. A means of confirming the effectiveness of corrective action(s).

A verified trend above the action or fail limit should immediately trigger implementation of the CAP. Because human activity is the most likely source of process control failure, the investigative process normally begins with personnel, and proceeds through the various possible causes from most to least likely. An exception to this general plan is verification of room pressurization, which is a primary indication of engineering control equipment efficiency. Although routine monitoring of pressurization should detect any out-of-limit results, the simplicity of verifying proper pressurization suggests this as a first step.

In general, the cause of any deterioration in process or environmental control can be traced to one of three principle systems: a) personnel controls, b) process controls, or c) facility (engineering) controls. Increases in detected airborne microbiologic contamination levels may result from any of several conditions, and a simple set of logical challenges can be applied to the data to determine the most likely cause.

Challenge 1, *Is the increase real and reproducible?* If it is not reproducible, it may be due to sampling error, or NACs. If it is reproducible, it may be due to an actual increase in levels, or due to enhanced collection efficiency, due to changes in methods, materials, or seasonal or other shifts in the kinds of contaminants present (different organisms have different sampling efficiencies).; Challenge 2, *If the increase is real, is it due to an increase in source intensity, or to a decrease in the ability of engineering controls to maintain a clean air supply?* The easiest way to differentiate between these possibilities is to examine particle count data. There are several possible combinations of test results, each indicating a different cause for increased airborne contamination: a) If particle counts taken under operational conditions have not risen, but airborne microbiologic contamination has, it is most likely due to a breakdown in personnel discipline and/or gowning procedures. b) If operational particle counts have risen, but at-rest counts have not, it is again likely that the cause of elevated microbial contamination is personnel activity and that it represents an increase in source intensity (when human activity is eliminated, engineering controls are able to produce the same conditions that were present during the OQ validation phase). c) If at-rest particle counts have risen, the increase is probably due to a decrease in the efficiency of the engineering controls.

Similar logical tests can be applied to increases in surface contamination levels, which may be due to increases in source intensity, or decreases in the efficiency of barrier controls or cleaning and sanitizing procedures. Flow charts illustrating the logical evaluation of data, and investigation of out-of-limit results are useful as starting points in the development of corrective action plans (5).

REFERENCES

- Code of Federal Regulation. *The Current Good Manufacturing Practices*; Title 21, Part 211, U.S. Government Printing Office: Washington, DC, 1992.
- ASHP Technical Assistance Bulletin. Quality Assurance for Pharmacy-Prepared Sterile Products. *AJHP* **1993**, *50*, 2386–2398.
- United States Pharmacopeial Convention, Inc. Sterile Drug Products for home use. *USP* **1998**, *23*, 1963–1975, (1206).
- Roscioli, N. Environmental Monitoring Considerations for Biological Manufacturing. *BioPharm* **1996**, *9* (8), 32–38.
- Peters, G.F.; McKeon, M.R. Microbiologic Monitoring of Aseptic and Controlled Processes. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 2000; 19, 239–278.
- Center for Drugs and Biologics and Office of Regulatory Affairs. *Guideline on Sterile Drug Products produced by Aseptic Processing*; FDA: Rockville, MD, 1987.
- Institute of Environmental Science and Technology. In *Airborne Particulate Cleanliness Classes in Cleanrooms and Clean Zones*; Federal Standard U.S. Govt. Printing Office: Washington, DC, 1992.
- United States Pharmacopeial Convention, Inc. Microbial Evaluation of Clean Room and other Controlled Environments. *USP* **1998**, *23*, 4426–4433, (1116).
- National Aeronautics and Space Administration. In *NASA Standards for Clean Rooms and Work Stations for the Microbially Controlled Environment*; Publication NHB 5340.2; NASA: Washington, DC, 1967.
- Peters, G. Laminar Airflow Equipment: Engineering Control of Aseptic Processing. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1993; 8, 317–359.
- Satter, S. Biological and Physical Monitoring of a Controlled Environment. *J. Am. Contam. Contr.* **1998**, *1* (8), 27–31.
- Akers, J.; Agalloco, J. Sterility and Sterility Assurance. *PDA J. Pharm. Sci. Technol.* **1997**, *51* (2), 72–77.
- F.D.C. Reports, Inc. Quality Control Reports, The Gold Sheet September, 1998; 15–19.
- Wilson, J. Setting Alert/Action Limits for Environmental Monitoring Programs. *PDA J. Pharm. Sci. Technol.* **1997**, *51* (4), 161–162.
- Andersen Samplers, Inc.: Ed.; *Operating Manual TR# 76-900042*; Andersen Samplers, Inc. Atlanta, 1980; 23.
- Parenteral Drug Association, Inc. Fundamentals of a Microbiological Monitoring Program. *PDA J. Pharm. Sci. Technol.* **1990**, *44* (S1), S3–S14.
- BSI. *Guide to Operational Procedures and Disciplines Applicable to Clean Rooms and Clean Air Devices*; Part 1, British Standard 5295: London, 1989.
- General Services Administration. In *Clean Room and Work Station Requirements, Controlled Environments*; Federal Standard 209b, U.S. Govt. Printing Office: Washington, DC, 1976.
- Sanford, R. Cumulative Sum Control Charts for Admixture Quality Control. *Am. J. Hosp. Pharm.* **1980**, *37*, 655–659.
- Akers, J. Environmental Monitoring and Control: Proposed Standards, Current Practices, and Future Directions. *PDA J. Pharm. Sci. Technol.* **1997**, *51* (1), 36–47.
- Niskanen, A.; Pohja, M.S. Comparative Studies on the Sampling and Investigation of Microbial Contamination of Surfaces by the Contact Plate and Swab Method. *J. Appl. Bact.* **1977**, *42*, 53–63.
- Sayer, W.J.; MacKnight, N.M.; Wilson, H.W. Hospital Airborne Bacteria as estimated by the Andersen Sampler Versus the Gravity Settling Plate. *Am. J. Clin. Path.* **1972**, *58*, 558–562.
- Ljungqvist, B.; Reinmüller, B. Interaction between Air Movements and the Dispersion of Contaminants: Clean Zones with unidirectional Air Flow. *PDA J. Pharm. Sci. Technol.* **1993**, *47* (2), 60–69.
- Whyte, W. Support of Settle Plates. *PDA J. Pharm. Sci. Technol.* **50** (4), 201–204.
- Kingston, D. Selective Media in Air Sampling: A Review. *J. Appl. Bact.* **1971**, *34* (1), 221–232.
- Hering, S.V. Inertial and Gravitational Collectors. *Air Sampling Instruments for Evaluation of Atmospheric Contaminants*, 7th Ed.; Hering, S.V., Ed.; American Conference of Government Industrial Hygienists, Inc.: Cincinnati, 1989; 385.
- Jensen, P.A.; Todd, W.F.; Davis, G.N.; Scarpino, P.V. Evaluation of Eight Bioaerosol Samplers Challenged with Aerosols of Free Bacteria. *J. Am. Ind. Hyg. Assoc.* **1992**, *53* (10), 557–660.
- Brachman, P.S. Standard Sampler for Assay of Airborne Microorganisms. *Science* **1964**, *144*, 1295.
- Macher, J.M.; First, M.W. Personal Air Samplers for Measuring Occupational Exposures to Biological Hazards. *J. Am. Ind. Hyg. Assoc.* **1984**, *45* (2), 76–83.
- Lyons, C. Sampling Efficiencies of all-glass Midget Impingers. *J. Aerosol Sci.* **1992**, *23* (S1), S599–S602.
- Pickard, D.R.; Pendlebury, D.E. Examining Ways to Capture Airborne Microorganisms. *Clean Rooms* **1997**, *11* (6), 34–40.
- Benbough, J.E.; Bennett, A.M.; Parks, S.R. Determination of the Collection Efficiency of a Microbial Air Sampler. *J. Appl. Bacteri.* **1993**, *74*, 170–173.

33. Decker, H.M.; Wilson, M.E. A Slit Sampler for Collecting Airborne Microorganisms. *Appl. Environ. Microbiol.* **1954**, *2*, 267–269.
34. Fields, N.; Oxbarrow, G.; Puleo, J.R.; Herring, G. Evaluation of Membrane Filter Field Monitors for Microbiological Air Sampling. *Appl. Environ. Microbiol.* **1974**, *27* (3), 17–520.
35. Marple, V.A.; Willeke, K. Impactor Design. *Atmos. Environ.* **1976**, *10*, 891–896.
36. Kaye, S. Efficiency of Biotest Rcs as a Sampler of Airborne Bacteria. *PDA J. Parenter. Sci. Technol.* **1988**, *42* (5), 147–152.
37. Trudeau, W.L.; Fernandez-Caldas, E. Identifying and Measuring Indoor Biologic Agents. *J. Clin. Immunol.* **1994**, *94* (2:2), 393–400.
38. Macher, J.M.; First, M.W. Reuter Centrifugal Air Sampler: Measurement of Effective Air-Flow Rate and Collection Efficiency. *Appl. Environ. Microbiol.* **1983**, *45*, 1960–1962.
39. Goetz, A. Application of the Molecular Filter Membrane to the Analysis of Aerosols. *Am. J. Publ. Health.* **1953**, *43*, 150–159.
40. Noller, E.; Spendlove, J.D. An Appraisal of the Gelatin Foam Filter as a Sampler for Bacterial Aerosols. *Appl. Environ. Microbiol.* **1956**, *4*, 300–306.
41. Mitchell, R.B.; Fulton, J.D.; Ellingson, H.V. A Soluble Gelatin Foam Filter for Airborne Microorganisms at Surface Levels. *Am. J. Publ. Health* **1954**, *44*, 1334–1339.
42. Ljungqvist, B.; Reinmüller, B. Active Sampling of Airborne Viable Particles in Controlled Environments: A Comparative Study of Common Instruments. *Eur. J. Parenter. Sci.* **1998**, *3* (3), 59–62.
43. CMF Only. GMF was not included in this study.
44. Ljungqvist, B.; Reinmüller, B. Hazard Analyses of Airborne Contamination in Clean Rooms: Application of a Method for Limitation of Risks. *PDA J. Pharm. Sci. Technol.* **1995**, *49*, 239–243.
45. Besterfield, D. *Quality Control*; Prentice-Hall, Inc.: Upper Saddle River, NJ, 1998.
46. Hyde, W. Origin of Bacteria in the Clean Room and their Growth Requirements. *PDA J. Pharm. Sci. Technol.* **1988**, *52* (4), 154–159.

MICROBIAL CONTROL OF PHARMACEUTICALS

Nigel A. Halls

GlaxoSmithKline Global Manufacturing and Supply, Uxbridge, United Kingdom

INTRODUCTION

Pharmaceutically active products (drug products) are expected to be efficacious, however; the presence of microorganisms in these products may have adverse effects on their efficacy. The severity of the effects that microorganisms may have on any particular drug product is a function of the nature of the product, its intended use, and the nature of the microorganism concerned. At one end of the spectrum, microbial contamination of a sterile parenteral product may, on injection into a debilitated patient, result in fatality; at the other, patients may refuse to begin or continue a course of medication because of aromas, off-flavors, or discolorations of microbial origin. In either situation, or in any related situation, the presence of microorganisms ought to be avoided in drug products.

Microbiological standards for drug products are published in the pharmacopoeias and/or are required by regulatory agencies for their registration. Generally, these standards are concerned with the protection of the public from infection by limiting the numbers and types of microorganisms to levels that are unlikely to be harmful. In addition to standards applying to the products themselves, there are also microbiological standards applying to the conditions under which drug products are allowed to be manufactured. These manufacturing standards (Good Manufacturing Practices, or GMPs) are intended to ensure that finished product standards are being attained consistently.

Microbial control of pharmaceuticals is primarily concerned with minimizing the opportunities for drug products to be contaminated by microorganisms. It is secondarily concerned with minimizing the potential for any microorganisms that may have contaminated drug products to increase to levels that may risk the efficacy of the product. The testing of product samples for compliance with microbiological standards is only one small part of this. By and large, microbiological test methods are product-destructive and, as a consequence, it is unusual to find valid statistical sampling and testing being done at batch release. Finished product testing is at best confirmatory and in some cases may be dispensed with when manufacturing controls ensure that products are

highly unlikely to become microbiologically contaminated (parametric release in its broadest sense).

Minimizing the risk of microbiological contamination of drug products is assured by the application of microbiological and physical standards and controls to starting materials, product-contact packaging components, manufacturing facilities, manufacturing processes, and equipment. By and large, these assurances are obtained by applying controls that protect materials, equipment, and processes from sources of microbiological contamination. In recognition of the frailty of protective measures in all but the most extreme circumstances, microbiological contaminants are also routinely controlled by removal, inactivation, or destruction.

The microbiological standards applying to drug products are expected to be maintained until time of use by the patient (or healthcare professional) and throughout their shelf-lives. This presents two areas of concern relevant to microbiological control: first, that the product should be protected (usually by its packaging) from additional contamination after release to market, and second, that the product should be formulated to prevent proliferation of any microorganisms that may have been present at tolerable levels at the time of release.

Different types and levels of microbial control are applicable to different types of products. The single major division is between sterile and nonsterile products.

MICROBIAL CONTROL OF STERILE PRODUCTS

Sterility is defined as the total absence of all viable life forms. Parenteral products and ophthalmic products are expected to be sterile. Parenteral products must be sterile because their route of administration overrides the body's external physical barriers to infection. Ophthalmic products must be sterile because eye damage is often irreparable. No distinction can be made between microorganisms that are known to be specific causative agents of disease and those that are not. Any microorganism may be opportunistically pathogenic if administered

parenterally or if applied to susceptible tissues (the transparent parts of the eye have a particularly poor blood supply and, therefore, a less sensitive immunological response than do other parts of the body) or to debilitated or immunocompromised patients.

Sterility, the Sterility Test, and Sterility Assurance

Sterility has an absolute definition (absence of ALL viable life forms). As with all absolutes, it is difficult, if not impossible, to prove.

Sterility tests

The pharmacopoeial standard applying to sterile products is that they must be capable of passing a Test for Sterility. A Test for Sterility is described in *U.S. Pharmacopeia* (USP) under Section 71 and in the *European Pharmacopoeia* (PhEur) under Section 2.6.1. These were “harmonized” along with the *Japanese Pharmacopoeia* and the requirements of the Australian Therapeutic Goods Administration in 1999, but they still have some minor differences in detail.

The Test for Sterility relies on the detection of viable microorganisms within a sample that is directly or indirectly inoculated into broad-spectrum microbiological recovery media. Detectable microbial growth is confirmation of nonsterility, whereas sterility can be assumed from the absence of growth. In other words, the Test assumes sterility unless nonsterility can be demonstrated. The Test for Sterility is really a test for nonsterility, and even within that redefinition of its terms of reference, its indication is limited to those microorganisms capable of producing discernible growth under the specified test conditions. Many microorganisms are not recoverable under these conditions.

When the Test for Sterility first appeared in the USP and in the *British Pharmacopoeia* (BP) in the 1930s, it was described in the same terms as any other pharmacopoeial test method; that is, as a method whereby a microbiologist could determine whether a single article presented for analysis was sterile.

There was no indication in the pharmacopoeias until 1955 in the USP and until 1968 in the BP that the results of the Test should be or could be extended to apply to batches of product required to be certified as sterile. The USP initially required 10 units to be tested from each batch of autoclaved products and 20 units from each batch of other sterile products, the BP required 20 units. In the current pharmacopoeias, the sample size for the Test for Sterility is 20 units in all but a few exceptional circumstances.

However, with such a small sample size, successful results (no growth inferring with sterility or “passing” the Test for Sterility) provide little assurance that there is not a significant proportion of contaminated units in the batch. This has been recognized and debated for many years, even before the pharmacopoeial requirements were first published. In 1949, Knudsen (1) pointed out that when using a sample size of 20 units, batches containing 5% contaminated units would be “passed” as sterile on 35 of every 100 occasions. Brewer (2) elaborated further on the statistical limitations of sterility testing by showing that sample sizes of 20, 50, and even 100 units are hardly better than sample sizes of 10 units for detecting contamination in batches of product containing 0.1% contaminated units. The pharmacopoeial sampling requirements for the Test for Sterility are a compromise; they have been known to be, and shown to be, statistically unsound for decades. They are a compromise between the fact that the Test is destructive to the units sampled and the self-evident requirement for a very high level of assurance that batches of supposedly sterile product do not contain nonsterile units.

In practical terms a “pass” in the Test for Sterility should not be perceived to be of any more significance than any other successful measure of compliance with microbiological or physical standards or controls applicable to the manufacture of sterile products. A “pass” in the Test for Sterility must not be allowed to overrule any failure to comply with other environmental or control standard(s) because it is quite possible to “pass” the Test and still have a significant number of nonsterile units in the batch. On the other hand, a “failed” Test for Sterility is likely to be a good indicator of a genuine problem that has not been disclosed by some other microbiological or physical means.

Methods for sterility testing

The Test for Sterility may be performed in one of two ways, by direct inoculation (direct transfer) or by membrane filtration.

In direct inoculation, the product samples are put aseptically into the microbiological recovery medium and incubated. Clearly this approach is only suited to products that are not likely to be inhibitory to the growth of microorganisms in the recovery medium. An incubation period of 14 days is specified.

In the membrane filtration method, the product samples are put aseptically into a volume of noninhibitory diluent and then passed through a sterile 0.45- μm membrane filter (Fig. 1). The membrane is rinsed through with additional volumes of diluent, then aseptically cut in half. Half is

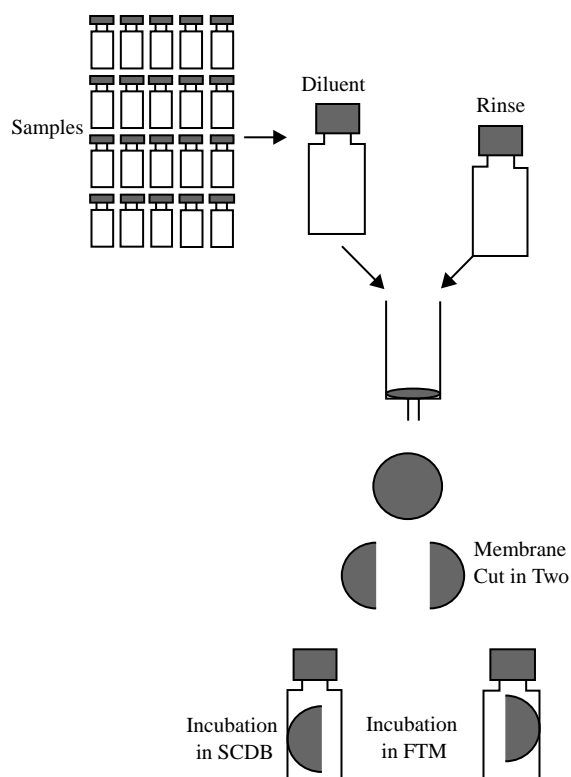


Fig. 1 Generalized scheme for the Test for Sterility by the membrane filtration method.

transferred to a container of soybean casein digest medium (SCDM) and the other to fluid thioglycollate medium (FTM).

A variation of membrane filtration is the use of “closed” systems such as the Millipore Steritest[®] system. At the heart of the Steritest system are two presterilized canisters, each with a membrane filter sealed into its base. The product under test is inoculated into its diluent and transferred via a peristaltic pump into the canisters and through the two membranes. During transfer the diluent is automatically split in two parts. After product filtration, the media are transferred into the canisters, SCDM to one canister and FTM to the other, by the same means. No manipulation of the membranes is required. The canisters are incubated at the temperatures specified in the pharmacopoeias.

Whichever variant of the membrane filtration method is used, SCDM is incubated at 20–25°C and the FTM at 30–35°C for 14 days.

Before 1999, the pharmacopoeias allowed incubation for only 7 days. Now the USP allows 7-day incubation only for terminally sterilized products. This extension of the incubation period from 7 to 14 days is a curious situation. The stimulus was an article published by the

Australian Regulatory Agency (3) indicating that it was possible for some drug products to be contaminated with microorganisms that would require more than 7 days of incubation to produce demonstrable growth under the conditions of the Test for Sterility. This was not, however, a remarkable new discovery. Brewer and Schmitt (4) observed (albeit with ethylene oxide sterilized medical devices) in 1966 that “slow growing” microorganisms (e.g., micrococci and diphtheroids) might take up to 4 weeks to produce demonstrable growth under the conditions of the USP Test for Sterility. There has been no actual evidence of patients having come to harm as a result of requiring only 7 days of incubation in the Test for Sterility. One can only suggest that some compromise was reached during the “harmonization” discussions, and the benefit must be for exporters to Australia where the “preharmonization” Test for Sterility (5) was far more demanding and restrictive.

Conditions for sterility testing

The Test for Sterility is vulnerable to false-positive and false-negative results. A false-positive means that a genuinely sterile product fails the test because of incidental contamination during the preparation of the sample or during the Test. A false-negative means that a nonsterile product passes the Test because viable contaminants fail to grow.

Both circumstances can have serious consequences, but false-positives are less likely to go unnoticed because high frequencies of false-positives have a commercial impact, at least delaying the release of product and at worst leading to suspension of manufacture.

The avoidance of false-negatives is addressed at length in the pharmacopoeias. Each batch of medium used in the Test for Sterility must have been shown to be capable of supporting the growth of low inocula of a specified array of microorganisms. Each Test for Sterility applicable to each specific product must be validated by repeated demonstration that the viabilities of low inocula of a specified array of microorganisms are not inhibited by product traces contained in the medium (direct inoculation) or on the membrane filter.

The pharmacopoeial attitude to false-positives changed in 1999, beginning in the 8th Supplement to USP 23 and in the 1999 Supplement of PhEur. An expectation of some level of false-positives had been tolerated in the Test for Sterility since its inception. This was in acknowledgement of there being a finite probability of the Test becoming contaminated as a result of the large number of aseptic manipulations required in its performance. However, times and technologies have changed.

When the Test for Sterility was first published in the 1930s, it was, at best, being done in glove boxes. The pharmacopoeias recognized the probability of incidental contamination and permitted, in the event of Test failure, an automatic right to retest without further ado. However, with the advent of HEPA filtration and laminar flow technology in the 1960s, the actual frequency of false-positive diminished, but the automatic right to retest remained. Isolation technology became available for doing the Test for Sterility in the last decade of the 20th century.

The automatic right to retest still remained in the pharmacopoeias until in the 1999 revisions, it was recognized that the testing technology was available whereby the probability of incidentally contaminating a Test for Sterility could be less than the probability of the sample actually being nonsterile. The current pharmacopoeial situation is therefore that retesting is only permitted when it can be unequivocally demonstrated that the contaminant in the Test arose incidentally during testing. Strictly, this is not a retest but a repeat Test because the first Test is at fault.

The consequence of this approach is that repeat testing is allowed only with a great deal of information about the microbiological conditions pertaining during the Test, and a large element of professional judgment by the functions held responsible for product release. Many companies have balanced the potential for the increased risk of commercial loss as a result of failures in the Test for Sterility with the potential for regulatory action in the event of disagreement over the judgment involved in repeat testing. They concluded that isolation technology is the only sensible future for the Test for Sterility.

Sterility assurance

Despite the attention, capital and resources given to the Test for Sterility, batches of product cannot be confirmed to be sterile by end-product testing. The Test for Sterility is in fact only a test for a specific broad range of microbial contaminants, and its sampling statistics are not capable of disclosing frequencies of contaminated units that would put patients at risk owing to nonsterility. This has been recognized by the pharmacopoeias, notably the PhEur, which for some years has carried a statement in Section 2.6.1 (formerly V.2.1.1):

... a satisfactory result only indicates that no contaminating microorganism has been found in the sample examined in the conditions of the test. ... The sterility test is ... the only analytical method available to the authorities who have to examine any product for sterility.

Then, how can sterility of batches of supposedly sterile products be confirmed by those functions within companies that are held responsible for product release? The answer is in the use of validated manufacturing processes based on sound scientific evidence that each product unit is most probably sterile.

This raises a second question of how much confidence must one have to claim sterility? The answer to this question, for terminally sterilized products, is that there must be no more than one chance in a million that viable contaminants survive in any one unit. This is called a sterility assurance level (SAL) of 10^{-6} . The answer for aseptically filled products is that the SAL must be as close to 10^{-6} as is technically possible, with the proviso that the degree of protection given to the process must afford no more than one chance in a thousand of any one unit becoming contaminated. This is called a contamination rate of 10^{-3} , and unlike the SAL it relates only to the protection given to the process and not to the potential for contaminants surviving or proliferating in actual products (6, 7).

The concept of the SAL is founded in academic studies of how microbial populations reduce in numbers in response to inimical treatments. In 1945 McCulloch (8) showed that for steam sterilization, a population of bacteria:

exposed to a lethal degree of heat ... decreases in a fairly orderly manner. Thus if 90% of the viable population is killed during the first minute of exposure, approximately 90% of the survivors will

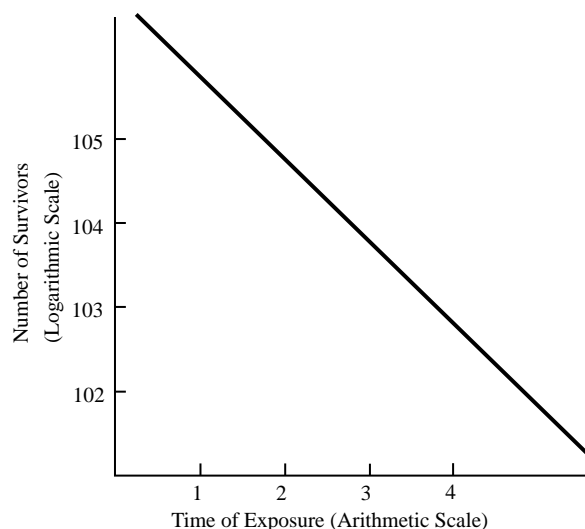


Fig. 2 Exponential inactivation of microorganisms (the survival curve).

be killed during each subsequent minute, which will continue until nearly all of the population is extinct.

This exponential order of inactivation of microbial populations (Fig. 2, the exponential survival curve) has subsequently been demonstrated to be a general characteristic of microorganisms in all processes of sterilization. The logarithmic axis of the exponential survival curve has no zero point. Thus, there can be no time of exposure at any temperature, no dose of radiation no matter how high, that can guarantee 100% inactivation of any microbial population.

Exponential inactivation is the basis of the concept of sterility assurance. If the behavior of microbial populations in response to a particular sterilizing procedure is regular and exponential over the region of the survival curve within which their response can be monitored, then the treatments required to achieve SALs of 10^{-6} can be extrapolated.

The aseptic filling process is not suited to the calculation of SALs in this way. This is because aseptic filling is not based on exponential inactivation and therefore provides no basis for extrapolation. The principle of aseptic manufacture is that, first of all, the drug product and its product-contact packaging components are sterilized to comply with SALs of 10^{-6} , and then they are brought together in ways that are intended to avoid further contamination (asepsis). Ideally, the finished product would still have an SAL of 10^{-6} . The short-fall from this ideal is primarily a function of the technology used in the aseptic filling process. Where sterile isolation technology is being used—this is a developing rather than an established technology at the time of writing in 2000—it is likely that there is very little shortfall. With standard technology, a level of process protection from microorganisms demonstrable through media fills offering a contamination rate of no more than 10^{-3} is nominally acceptable to regulatory bodies worldwide.

Microbiological GMP in the Manufacture of Sterile Drug Products

As far as manufacture is concerned, there are two broad categories of sterile products: those that are terminally sterilized (filled and hermetically sealed within their final containers before being exposed to a sterilization treatment), and those that are aseptically manufactured (the drug product and its product-contact packaging components are sterilized and then brought together in ways that are intended to avoid further contamination. Serious measures to protect the manufacture from microbiological contamination are necessary for both

categories but are clearly most critical in the poststerilization stages of aseptic manufacture.

There are several sources of standards and limits that may be applied to the manufacture of sterile products. They can be divided into sector-specific standards which are focused specifically on the manufacture of sterile pharmaceuticals and generic standards that are focused on particular technologies, irrespective of how the technologies are to be applied.

Sector-specific standards applying to manufacture of sterile drug products

There are five significant sources of sector-specific standards applying to the manufacture of sterile products. Two are mandatory, three are voluntary.

The two mandatory standards are Annex 1 of the European Union's GMPs (9) and the FDA's 1987 *Guideline on Sterile Drug Products Produced by Aseptic Processing* (10). The FDA document is under review at the time of writing this article.

The first of the two nonmandatory standards is contained in Section 1116 of the USP. It should be explained that this is a general chapter of the USP and is not therefore mandatory unless referenced in a USP monograph. At time of writing, Section 1116 is not referenced in any monographs.

The second of the voluntary standards is IS 13408 (11), applying to both medical devices and to pharmaceuticals according to the International Standards Organization (ISO) definition of "healthcare products." Here, it should be explained that the ISO is a voluntary body whose membership is drawn from national standards organizations. ISO standards have no legal authority in any country or territory unless the national standards body elects to adopt the ISO standard instead of publishing its own standards.

The third voluntary standard is The International Society of Pharmaceutical Engineer's (ISPE) *Baseline Pharmaceutical Engineering Guide* (12). This is one of a series of ISPE guidance documents that has limited endorsement by the FDA. It should be understood that only minimal standards are being endorsed and there are many reasons why specific sterile products and applications could be seen by the FDA to require compliance with higher standards.

Working with these five documents is not easy; they use different terminology and definitions and have different emphases that in part, may be a function of the 10-year span that encompasses publication dates. All of them are concerned with controlling microbiological contamination that may arise from the environment, from personnel, and from materials and equipment. Some individual standards within these documents may relate to microbial numbers,

others to physical characteristics that relate to contamination control.

All of these documents stress a requirement for room classifications according to the concentration of nonviable airborne particles at 0.5 μm and larger. The EU GMPs (9) subordinate classification of the various areas applicable to sterile manufacture according a broader based grading system. Grades A to D are defined in terms of a range of independent characteristics such as the concentration of nonviable airborne particles at two sizes in operational and nonoperational conditions, the concentration of airborne viable particles in operational conditions, and several other indices of microbial contamination on surfaces and hands, etc. The inclusion of microbiological limits within this grading system is important in that it emphasizes that it is not solely the concentration of nonviable airborne particles in clean-rooms (room classification) that determines their suitability for manufacture of sterile products.

Generic standards for technologies used in the manufacture of sterile drug products

The most important generic standards applying to technologies used in the manufacture of sterile drug products relate to the protection of manufacture from contamination from the air. The air supply to facilities for sterile manufacture must be filtered, in some specific applications, it must be unidirectional, its velocity may have to be controlled within specified limits, and there must be pressure differentials between adjacent areas.

In the late 1990s, there had been a great deal of ISO activity in the harmonization of various standards applying to air quality (13). The first three parts address classification of air cleanliness, methods for testing and monitoring, and metrology. Parts 1 and 2 cover much of the ground but are not identical to FS 209 (14) (see below). The significant question emerging as these standards are published has to do with the reaction of the pharmaceutical regulatory agencies. Whereas we can with confidence be sure that the standards applying to the specifications of HEPA filter media and to the in situ HEPA filter integrity test will be accepted, there still are some doubts about the standard for room classification.

Both of the two mandatory sector-specific standards (9, 10) currently require room classification according to Federal Standard 209. This is a U.S. government standard; its current revision is FS 209E (14), and it will not be automatically replaced by IS 14644 (13). Thus, we are dependent on the next revision of Annex 1 of the EU GMPs (9) and of the FDA's *Guideline on Sterile Drug Products Produced by Aseptic Processing* (10) to

determine whether they will be adopted for the manufacture of sterile drug products.

Validation and Control of Sterilization Processes

Although there are many sterilization processes used in association with the manufacture of sterile drug products, the three primary processes are steam sterilization, dry heat sterilization, and sterile filtration. Dry heat sterilization is, in the context of the manufacture of sterile parenteral products, a subset of dry heat depyrogenation (see below).

Steam sterilization

Steam sterilization is widely used as a terminal process for drug products in glass ampules, vials, syringes, and plastic containers. It is also used for sterilizing closures, filters, manufacturing equipment, and cleaning equipment, etc.

Steam sterilization in autoclaves has a long and strong scientific basis (see above). The essence of validation of steam sterilization processes is to demonstrate that temperature and time conditions are being achieved uniformly through every item included the autoclave load and that the lethality being achieved in practical situations corresponds to that which would be expected from sterilization theory.

Thus, compliance with limits for temperature uniformity throughout empty autoclaves (heat distribution studies) is an index of the way in which the autoclaves are engineered. Compliance with limits for temperature uniformity within items loaded into the autoclaves (heat penetration) is an index of the way in which items are wrapped for sterilization (where applicable) and how they are loaded into the autoclaves in relation to the positions of steam inlets, drains, racks, trays, and thermal sensors.

With modern autoclaves, the major impediment of concern to successful sterilization is air in porous loads. Air may be present as a contaminant of the steam supply such that the temperatures theoretically achievable at particular steam pressures are depressed, or air may be present within the load, insulating it from the contact with the steam required for predictable lethality.

Pure steam generators are designed to deliver steam of satisfactory quality—standards are published in a UK document, Health Technical Memorandum (HTM) 2010 (15)—but noncondensable gas like air and nitrogen can accumulate in the steam through poorly insulated and trapped distribution systems and therefore should be tested periodically close to point of use. Autoclaves suited to sterilization of porous loads require some form of air removal, most often by deep pulsed vacuums, before the introduction of steam. The effectiveness of air removal is

difficult to monitor by physical means; air entrapment may be a local phenomenon in particular locations in loads, in particular materials (e.g., in cartridge filters), or where items come into contact with one another (e.g., with rubber closures in bulk). Therefore, it is normal to use biological indicators to ensure that there is correspondence between theoretical and practical reality. Biovalidation of steam sterilization, although done extensively, is all very complex and, to an extent, a poorly understood. The basics are that a suitable biological indicator should be resistant to sterilization by steam (but not necessarily the most resistant microorganism known—spores of *Bacillus stearothermophilus*) are approved by the USP, PhEur, and others and are most often used. The biovalidation cycle must be of lower thermal lethality than the lowest thermal lethality allowable within the routine production-cycle specification. Some interesting reviews of biovalidation of steam sterilization have been published by Halls (16) and by Agalloco and colleagues (17).

Steam sterilization is not confined to autoclaves. Many items of manufacturing equipment, including some quite massive applications to vessels, pipe-work, filters, etc., are now sterilized in place with steam (SIP). It is essential that the equipment holds pressure, that air is removed, and that condensation does not accumulate in low points of the system. Biological validation is generally unsophisticated and is normally done by ensuring that the process inactivates several biological challenge test pieces, each carrying 10^6 spores of *Bacillus stearothermophilus* (D_{121} -values undefined but presumably meeting the pharmacopoeial criteria of equal to or greater than 1.5 min) placed at critical points. Academically, it is problematic to equate this approach with the sterility assurance level concept that applies to the sterilization of groups of items.

Steam sterilization processes are monitored for compliance with strict specifications of temperature, pressure, and time. Routine monitoring with biological indicators is not necessary. Indeed, any item of steam sterilizing equipment that is operating so erratically as to merit routine monitoring with biological indicators should be replaced.

Sterilization by filtration

Filtration is a means of sterilizing fluids by removing, rather than inactivating, microorganisms. The sterilization of liquids is used extensively in aseptic manufacture, sterilization of gases is used both in terminal sterilization and in aseptic manufacture. Most applications use cellulose esters, polyvinylidene fluoride, polytetrafluoroethylene, nylon, and other polymeric materials. The removal of microorganisms from fluids by passage through filters is very complex; sieving, or surface retention, is only one of a

series of mechanisms that depend on interactions among the chemistry and surface characteristics of the membrane, the microorganisms, and the suspending fluid (18).

The FDA (10) defines sterilizing filters as those that have pore-size ratings of 0.22 μm or smaller and relates this to a microbiological particle passage test using *Pseudomonas (Brevundimonas) diminuta*:

A sterilizing filter is one which, when challenged with the microorganism *Pseudomonas diminuta*, at a minimum concentration of 10^7 organisms per cm^2 of filter surface will produce a sterile effluent.

It is essential that the microbiological particle passage test is performed as part of the development of new sterile formulations. Because of its very specialized nature, the test is normally performed only by the filter manufacturers, who then provide limits for secondary physical tests (e.g., bubble point, pressure decay, forward flow, etc.), which can be applied to verify the pore size rating and integrity of the membrane filters.

Maintenance of Sterility after Product Release

The sterility of a sterile dosage form can only be guaranteed while it is protected from the surrounding nonsterile environment within a container made from materials impermeable to microbial penetration.

Container and closure systems for sterile products must be capable of withstanding process conditions, storage, and transport without compromising the sterility of the product. This has been recently emphasized by the FDA, which has omitted a former obligation to provide data from the Test for Sterility in stability program for new sterile products in favor of verifying the microbial integrity of the container-closure system (19).

There are no standard methods for verifying microbiological integrity of container-closure systems. Documents such as that published by the Parenteral Society in 1992 (20) and the PDA in 1998 (21) may be helpful in relating microbiological integrity to secondary physical tests, but they do not specify detailed microbiological test methods.

There are two general approaches, wet tests and aerosol challenge tests. Wet tests consider penetration of microorganisms in liquid suspension into sealed containers usually previously filled with sterile medium. The basic assumption is that the most vulnerable route for penetration of liquid filled containers by microorganisms is in the event of a continuous liquid film or “bridge” forming between the outside and the inside of a container. Aerosol challenge tests are less critical than wet tests and

should be applied only when total exclusion of moisture from the containment system can be ensured by secondary barriers.

No discussion of maintenance of sterility is complete without addressing multiple-dose ophthalmic presentations. The normal circumstances for sterile parenteral products are that they are unit dose or, if multiple dose, they are penetrated only by sterile transfer devices (syringes, giving sets, etc.) and used on one patient only. This is not the case for ophthalmic ointments and drops: unit-dose presentations are quite unusual and generally used only in hospital practice, for example, after eye surgery.

The multiple-dose presentation is the norm for ophthalmic products. The container is breached at first use and therefore has no further microbiological integrity over subsequent applications by the patient for an undefined period possibly up to the expiration date of the product. Under these circumstances, contamination of the drug products can and must occur; the proliferation of the contaminating microorganisms the products is controlled by use of antimicrobial and preserved formulations. After opening they are, effectively, non-sterile products. Why, therefore, do we go to the trouble to manufacture ophthalmic ointments and drops as sterile products? There is no clear answer to this question; it is an anomaly. It has been a worldwide regulatory requirement "since the beginning of time" for ophthalmic products to be manufactured as sterile, and there is undoubtedly little point in contesting its logic or consistency. Another possible anomaly concerning sterility may be arising from the FDA's 1997 proposal (25) to require all aqueous inhalation products to be sterile. This example is anomalous because only in very exceptional circumstances is the equipment used to administer inhalation products supplied and maintained sterile and, of course, neither are the nasopharyngeal and bronchial passages of the patient.

Pyrogens, Lipopolysaccharides, and Bacterial Endotoxins

Parenteral products are expected to be sterile because of the risk of infection. Another critical biological characteristic is that they should be free from pyrogens. Pyrogens are substances that, when injected in sufficient amounts into the human body, give rise to a variety of extremely unpleasant symptoms of which the most recognizable is a rise in body temperature. In extreme conditions, the rise in body temperature can be so rapid and to such an extent that the patient dies (endotoxic shock). In the pharmacopeias, pyrogenic products have

traditionally been defined in terms of the temperature rise induced in injected rabbits.

The causative agent of the pyrogenic response is lipopolysaccharide. This material is of microbiological origin, and although all bacteria appear to be capable of producing lipopolysaccharide, it is found primarily in the cell envelope of Gram-negative species. Naturally occurring lipopolysaccharide is referred to as bacterial endotoxin.

Bacterial endotoxin reacts with a high degree of specificity with a "clottable protein" contained in the amoebocyte cells of the horseshoe crab (*Limulus polyphemus*). This has allowed the development of in vitro testing of drug products and other substances for bacterial endotoxins. The pharmacopeias are progressively replacing their former requirements for in vivo rabbit pyrogen testing of parenteral products with the in vitro *Limulus* Amoebocyte Lysate (LAL) test. There are a variety of test systems now marketed (gel clot, turbidometric, chromogenic, etc.), but the pharmacopeias regard the gel clot method as the reference test. In the gel clot method for the LAL test, a quantity of LAL reagent of defined sensitivity (λ), e.g., 0.03 EU (endotoxin units) per milliliter, is mixed with an equal quantity of the product under test (or a dilution thereof). The mixture is incubated for a defined period and then inverted. The production of a gel is evidence of bacterial endotoxin in the product sample at a concentration equal to or greater than the sensitivity of the LAL reagent.

Product endotoxin limits are based on dosage regimes (22) derived from a formula K/M , where K is the threshold dose for any substance capable of giving a pyrogenic response in humans. With some exceptions K has been given a fixed value of 5 EU/kg body weight of the patient, and adult patient body weight is standardized to 70 kg for calculation. M is the maximum dose of endotoxin per kilogram of body weight of a patient that is permitted to be given in a single 1-hr period.

Endotoxin limits are therefore unique to each dosage form. Because K is fixed, it is usually only necessary to determine M from the maximum human dosage indicated in the product instructions.

The validation and performance of the endotoxin test require rather elaborate and detailed attention to use of controls. These details are specified in Section 85 of the USP harmonized with Section 2.6.14 of the PhEur.

Control of bacterial endotoxins

Although bacterial endotoxins are of microbiological origin, they are not lost with loss of viability. Of the sterilization processes commonly used in the manufacture

of sterile parenteral dosage forms (see above), only dry heat is capable of destroying bacterial endotoxins in a reasonable time frame (23). There is therefore no practical way of removing bacterial endotoxins from finished drug products; thus, they must be controlled at source.

The most likely source of bacterial endotoxins is water and product-contact packaging components that have been in contact with water. This is because bacterial endotoxins are most frequently found associated with Gram-negative bacteria, and Gram-negative bacteria have evolved to be primarily waterborne.

Water used in manufacture of sterile parenteral products must comply with pharmacopoeial limits for endotoxin of no more than 0.25 EU/ml (limits in the USP and PhEur for water for injection). In principle water complying with this limit can be produced by distillation, reverse osmosis, and ultrafiltration. PhEur allows only distillation to be used for the manufacture of ingredient water for parenteral products; the USP allows distillation and reverse osmosis (although reverse osmosis is rarely used in the United States, perhaps because of FDA pressure). Only the Japanese Pharmacopoeia allows water for injection to be manufactured using ultrafiltration.

Product-contact packaging components, such as glass vials, which are required to be sterile, are usually sterilized and depyrogenated by dry heat in ovens or tunnels. The standard required by the FDA for acceptable depyrogenation processes (10) is that they should be capable of reducing a bacterial endotoxin challenge by a factor of 10^{-3} (3 logs). Unfortunately, inactivation of bacterial endotoxins by dry heat is complex, and satisfactory process specifications are difficult to predict; inactivation of purified lipopolysaccharide may approximate to pseudo-second-order reaction kinetics (24). It appears that there may be a threshold temperature of approximately 160–170°C, below which a 3-log reduction of bacterial endotoxin cannot be achieved regardless of time of exposure.

MICROBIAL CONTROL OF NONSTERILE PRODUCTS

Nonsterile dosage forms are a diverse group of products. The microbiological risks that nonsterile products present to the patient are equally diverse. A single microbiological standard cannot sensibly be applied to all nonsterile products, and this is well reflected by pharmacopoeial and regulatory requirements. Similarly, there can be no single standard applied to the microbiological control and

prevention of contamination during the manufacture of nonsterile products.

The Microbial Limit Test

Microbial limit standards for some, but not all, nonsterile dosage forms are given in USP monographs. PhEur takes a slightly different approach. Section 5.1.4 lists microbiological criteria for two categories of nonsterile pharmaceutical products on the basis of usage. One of these categories is for products for topical and respiratory use, the other for products for oral and rectal administration. Guidance to methods suited to testing products for compliance with these standards is given in Section 61 of USP and Section V.2.1.8 of the PhEur.

In setting appropriate limits for particular nonsterile products, both pharmacopoeias take account of the significance of microorganisms to different types of product, to the way in which the product is used, and to the potential hazard to the patient. For instance, 47 of 48 monographs defining microbial limits listed in USP 24 for oral dosage forms restrict microorganisms (*E. coli* and/or *Salmonella*) that are pathogenic by gastrointestinal ingestion. These are, of course, not the only microorganisms that could endanger a patient taking an oral dosage form. They have been selected as indicators of the general type because there are well defined methods for their recovery and they are easily recognizable in culture. Similarly, the USP typically (but not universally) restricts from topical products only those microorganisms (*Staphylococcus aureus*, *Pseudomonas aeruginosa*) that have the potential to cause skin infections. Sixty-seven of 68 monographs in the USP 24 require the absence of at least these two species, and of these 67 only 11 have additional requirements. These limits applying to the absence of specific indicator microorganisms appropriate to the product's usage may be supplemented by general hygiene restrictions on total numbers of microorganisms per gram or milliliter (typically to no more than 100 cfu per gram or milliliter). A very similar overall approach is taken by the PhEur.

Requirements are not specified in the pharmacopoeias for sample sizes appropriate to testing batches of nonsterile products for compliance with microbial limits. The test is destructive to the product, and it is extremely unlikely that statistically valid sampling plans are in use anywhere; samples composed of 10 1-g amounts taken from 10 separate 15-g tubes of cream, or of three 3.3-ml amounts taken from three separate 500-ml bottles of syrup, may be typical.

The Microbial Limit Test is, like the Test for Sterility, only confirmatory. It is not generally mandatory to test

each batch of every nonsterile product for compliance with microbial limits. The exception is in the United States where it is mandatory to test every batch of nonsterile product. At the time of writing (early 2001), the FDA has proposed a relaxation of this rule for future registrations. However, in the case of tablets there is no requirement to register microbial limits, and therefore testing is not mandatory. The logic behind this is that the water content of tablets is too low to allow proliferation of microbiological contaminants.

It should also be kept in mind that the major regulatory agencies (FDA, MCA) would not want to have the pharmacopoeial limits on numbers of microorganisms perceived as tolerance of microorganisms contaminating and proliferating in nonsterile products. The acceptance of these limits is only an acknowledgment of the reality that when products are not manufactured as sterile, some microbiological contamination is inevitable. Well formulated nonsterile products manufactured under microbiologically controlled conditions are extremely unlikely to have bioburdens approaching the pharmacopoeial limits (typically 100 cfu per gram or milliliter). Most regulatory bodies now expect companies to set tighter microbial limits on their nonsterile products based on their typical results, these being seen as alert limits indicative of some loss of microbiological control if exceeded.

Microbiological GMP in the Manufacture of Nonsterile Drug Products

The extent to which the microbiological controls are required to be applied in the manufacture of nonsterile dosage forms is primarily a function of two factors: consequences of infection to the patients and probability of microorganisms proliferating in the product.

The seriousness of the risk to patients of infection from nonsterile dosage forms is primarily dictated by the route of administration of the dosage form. The probability of contaminating microorganisms surviving

and proliferating in a nonsterile dosage form is generally a function of its water content. Bearing these two concepts in mind, it is possible to classify the extent to which microbiological controls are required in the manufacture of nonsterile dosage forms according to a "good hotel guide" approach such as that given in Table 1. The precise allocation of "stars" is a matter of opinion, but at one extreme there is little doubt that aqueous inhalations merit "5-star" microbiological controls in manufacture, and possibly may soon be required to be sterile (25). Near the other extreme, the FDA, for example, does not require microbial limits to be registered for solid oral dosage forms.

There is little published guidance to microbiological controls for nonsterile manufacture. Equipment and facilities should be designed to minimize the opportunities for contact with air, personnel, and water.

Exposure to air is unavoidable, either as local atmospheric (environmental) air or as air from compressors used to operate manufacturing equipment and in fluid bed dryers, etc. Even for "1-star" nonsterile manufacture, the air supplied to areas used for manufacture and filling should be filtered. HEPA filters should be used in "5-star" nonsterile manufacture but are generally not necessary in "1-star," intermediate classifications of manufacture, requiring decisions on air filtration to be made from sensible risk assessments.

Personnel should be restricted from all areas of pharmaceutical manufacture to those who are necessary. Protection from contamination from personnel is by training in hygiene, enforcement of hygiene rules, and provision of protective clothing. Hair is the major source of contamination from personnel, and "street clothes" is the second most significant source. Hair-covers should be provided and worn properly in all areas of sterile manufacture; beards, moustaches, and other excessive facial hair should be covered. Overall sleeves should extend to the wrist and be elasticated or studded to provide a neat fit. Because dogs are still allowed to soil the streets,

Table 1 Classification of nonsterile manufacture according to the need for microbiological controls

Route of administration	Type of nonsterile dosage form		
	Aqueous liquids	Nonaqueous fluids	Solid dosage forms
Inhalation	☆☆☆☆☆	☆☆☆☆	☆☆☆☆☆
Intranasal, intrabuccal	☆☆☆☆	☆☆☆	☆☆☆
Topical	☆☆☆☆	☆☆	☆☆
Oral	☆☆☆	☆	☆
Anal	☆☆	☆	☆

protective footwear or shoe covers should be provided to all personnel allowed to enter pharmaceutical manufacturing areas.

Water is often the major ingredient in nonsterile products. When it is used as an ingredient, it must meet the pharmacopoeial limit for purified water of not more than 100 cfu/ml. Well designed and operated pharmaceutical purified water production and distribution set ups meet far tighter standards than these. Water is also often the principal cleaning fluid for equipment and facilities and is therefore unavoidable. It is a potent source of contamination because it usually contains sufficient nutrients to allow survival of metabolically versatile microorganisms, particularly *Pseudomonas spp.*

When water is left to stand, *Pseudomonas spp.* do not only survive, but increase in number. Water should not be permitted to stand on equipment (particularly in crannies and crevices), on floors, or in sinks and wash bays. Contamination spreads with water that forms films over surfaces and on the hands and clothing of personnel. Waterborne contaminants may be aerosolized by vibrations or when water falls more than a few centimeters.

To restrict the opportunity for contamination from water, there should be air breaks of approximately 5 cm installed between equipment drains and the tun dishes leading to foul drains.

The FDA has published guidances on the design and control of water systems (26).

Maintenance of Microbiological Quality of Nonsterile Products after Release

It is not desirable for those few microorganisms that may be present in nonsterile products at the time of release to increase to numbers to levels at which they may present a risk of infection to the patient or "spoil" the product.

Additional contamination in the period up to the first use by the patient or healthcare professional is provided by the packaging. It is neither necessary nor usual to find that nonsterile product packaging has been specifically designed or tested to be impermeable to microbiological contamination. On the other hand, many products are given good microbiological protection as a secondary consequence of the protection given to the stability of their active ingredients, etc. For instance, tablets packed individually in strips or foils are microbiologically protected until time of use. It is also unlikely for microorganisms to be able to contaminate pressurized metered-dose inhaler containers.

Microbiological protection of multiple-dose presentations such as liquid inhalations, nasal sprays, oral liquids, creams, and lotions is more complex. Once opened they are susceptible to microbiological contamination. If they are aqueous-based, they are in principle susceptible to proliferation of these "new" contaminants. To avoid this, they are formulated with antimicrobial agents or preservatives and are expected to be able to comply with preservative efficacy standards specified in the pharmacopoeias. Preservative efficacy tests (not harmonized) are described in Section 51 of the USP and Section VIII.14 of the PhEur (Fig. 3).

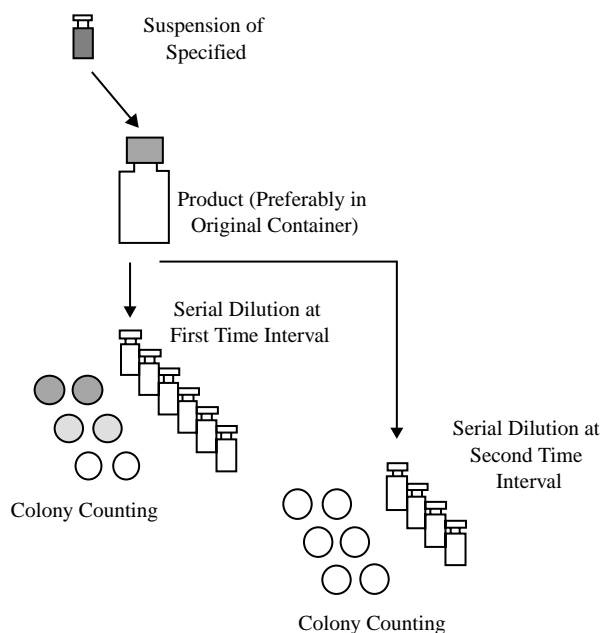


Fig. 3 Generalized scheme for preservative efficacy testing.

REFERENCES

1. Knudsen, L.F. Sample Size of Parenteral Solutions for Sterility Testing. *J. Am. Pharmaceutical Assoc.* **1949**, 38, 332–337.
2. Brewer, J.H. Sterility Tests and Methods for Assuring Sterility. *Antiseptics, Disinfectants, Fungicides and Chemical and Physical Sterilization*; Reddish, G.F., Ed.; Henry Kimpton: London, NC, 1957.
3. Bathgate, H.; Lazzari, D.; Cameron, H.; McKay, D. The Incubation Period in Sterility Testing. *PDA J. Pharm. Sci. Technol.* **1993**, 47 (5).
4. Brewer J.H. Schmitt R.F. Special Problems in the Sterility Testing of Disposable Medical Devices. Annual General Meeting of the Parenteral Drug Association: New York, November, 2, 1966.
5. Commonwealth Department of Health, *Standard for Sterile Therapeutic Goods*; Therapeutic Goods Order No. 11, Australian Government Publishing Service: Canberra, 1984.

6. Bernuzzi, M.; Halls, N.A.; Raggi, P. Application of Statistical Models to Action Limits for Media Fill Trials. *Eur. J. Parenteral Sci.* **1997**, 2 (1), 3–11.
7. Pharmaceutical Drug Association. *Technical Report No 22. Process Simulation Testing for Aseptically Filled Products*; Pharmaceutical Drug Association Inc.: Bethesda, MD, 1996.
8. McCulloch, E.C. *Disinfection and Sterilization*; Lea and Ferbig: New York, 1945.
9. Medicines Control Agency. *Rules and Guidance for Pharmaceutical Manufacturers and Distributors*; The Stationery Office: London, 1997.
10. Food and Drug Administration. *Guideline on Sterile Drug Products Produced by Aseptic Processing*; FDA: Rockville, MD, 1987.
11. *Aseptic Processing of Health Care Products. 1. General Requirements*; IS 13408-1, International Standards Organization: 1997.
12. ISPE Baseline Pharmaceutical Engineering Guide. *Sterile Manufacturing Facilities*; ISPE: 1998; 3.
13. Cleanrooms and Associated Controlled Environments I. Classification of Air Cleanliness; II. Specifications for Testing and Monitoring to Probe Continued Compliance with ISO 14644-1; III. Metrology and Test Methods. BS EN ISO 14644-1,2,3; , 1999.
14. *Airborne Particulate Cleanliness Classes in Cleanrooms and Clean Zones*; Federal Standard 209E. Federal Supply Service, General Services Administration: Washington. DC, 1998.
15. *Sterilization. III. Validation and Verification*; The Stationery Office: London, 1994; Health Technical Memorandum 2010.
16. Halls, N.A. Resistance “Creep” of Biological Indicators. *Sterilization of Medical Products*; Morrissey, R.F., Kowalski, J.B., Eds.; Polysciences Publications, Inc.: Champlain, NY, 1998.
17. Agalloco, J.P.; Akers, J.E.; Madsen, R.E. Moist Heat Sterilization—Myths and Realities. *PDA J. Pharm. Sci. Technol.* **1998**, 52 (6), 346–350.
18. Meltzer, T.H. *Filtration in the Pharmaceutical Industry*; Marcel Dekker, Inc.: New York, 1987.
19. Food and Drug Administration. *Guidance for Industry, Container-Closure Systems for Packaging Human Drugs and Biologics*; FDA Center for Drugs and Biologics: Rockville, MD, 1999.
20. *The Prevention and Detection of Leaks in Ampoules, Vials and other Parenteral Containers Parenteral Society Technical Monograph No. 3*; Parenteral Society: Swindon, UK, 1992.
21. *Pharmaceutical Package Integrity*; Technical Report No. 27 PDA: Bethesda, MD, 1998.
22. Weary, M.E. Understanding and Setting Endotoxin Limits. *J. Parenteral Sci. Technol.* **1990**, 44, 16–17.
23. Ludwig, A.D.; Avis, K.E. Dry Heat Inactivation of Endotoxin on the Surface of Glass. *J. Parenteral Sci. Technol.* **1990**, 44, 4–12.
24. Tsuji, K.; Harrison, S.J. Dry Heat Destruction of Lipopolysaccharide: Design and Construction of Dry Heat Destruction Apparatus. *Appl. Environ. Microbiol.* **1978**, 36, 710.
25. Federal Register. **September 23, 1997**, 62 (184).
26. *Food and Drug Administration Guide to Inspections of High Purity Water Systems*; FDA: Rockville, MI, 1993.

METERED DOSE INHALERS

Sandy J.M. Munro

Alan L. Cripps

GlaxoSmithKline, Ware, Hertfordshire, United Kingdom

INTRODUCTION

Metered dose inhalers (MDIs) are pharmaceutical delivery systems designed for oral or nasal use, which deliver discrete doses of aerosolized medicament to the respiratory tract. The MDI contains the active substance, dissolved or suspended in a liquefied propellant system held in a pressurized container that is sealed with a metering valve. Actuation of the valve discharges a metered dose of medicament as an aerosol spray through an actuator during oral or nasal inhalation.

The MDI may provide up to several hundred actuations, each containing typically from about 10 to 500 μg of drug dispersed in a 25 to 100 μl metered volume of liquid. The discharged liquid undergoes flash evaporation of the propellant to produce a finely dispersed aerosol spray. The deposition, and hence the clinical efficacy, are critically dependent on the mass of inhaled particles, which must have an appropriate aerodynamic size, typically less than 5 μm , to be deposited in the lungs (the respirable fraction) (1).

The first MDI products were developed by Riker Laboratories and marketed in 1956, using a newly patented design of metering valve. In most countries the MDI is now established as the principal dosage form of inhalation drug therapy for bronchial asthma and chronic obstructive pulmonary disease (COPD). Since its introduction, MDI technology has evolved steadily. However, with the phase-out in the commercial use of chlorofluorocarbon (CFC) propellants, which have been the mainstay of pharmaceutical MDIs, the pace of MDI technology development has accelerated with the transition to hydrofluorocarbon (HFC) propellants (2).

Despite their apparent simplicity in use, MDIs are complex devices involving the integration of formulation, container, metering valve and actuator (Fig. 1). Changes to any one of these components will affect the overall performance of the MDI, which is designed to ensure that the delivered dose and the particle size distribution of the drug in the aerosol spray are consistent over both the labelled number of actuations in the MDI and for the duration of the shelf-life (3).

The design and evaluation of MDIs are reviewed in references (4–7).

PROPELLANTS

The propellant or propellant mixture used in the MDI provides the energy necessary to generate a fine aerosol of drug particles suitable for delivery to the lungs or nasal cavity. Liquefied compressed gases are preferred over nonliquefied compressed gases such as nitrogen or carbon dioxide because they offer the following critical advantages for inhalation therapy:

- The discharge of defined aliquots of propellant from the MDI will undergo flash evaporation to give an aerosol of very small particles.
- The pressure inside the MDI remains consistent throughout the use of the entire contents, thus ensuring that the aerosol characteristics remain uniform during repeated discharges. At constant temperature, the vapor pressure remains constant while liquefied propellant remains. In contrast, aerosols generated using nonliquefied compressed gas coarsen during emptying of the MDI due to the decrease in gas pressure.

However, unlike nonliquefied compressed gases, the vapor pressure of liquefied propellants decreases significantly with decreases in temperature, such that below a certain temperature, the flash evaporation process is sufficiently retarded to give poor aerosol formation. For the propellants commonly used in MDIs, unacceptable aerosol formation is likely to occur below 0°C.

The ideal propellant for use in an MDI will exhibit the following properties (8):

- Nontoxic
- Inert and unreactive in the formulation
- Chemically stable under a range of conditions
- High purity
- Acceptable taste and odor
- Compatible with the packaging components (can, valve, actuator)

- Suitable vapor pressure
- Suitable density to facilitate suspension stability
- Suitable solvency properties
- Preferably nonflammable
- Acceptable cost

Until recently, only three chlorofluorocarbon (CFC) propellants, namely CFCs 11, 12 and 114 (Table 1), had been approved worldwide for use in medical MDIs. Their widespread acceptance was due to their ability to substantially meet the ideal propellant properties. All the CFC MDIs that are currently marketed employ CFC 12 as the major constituent mixed with either CFC 11 or with a mixture of CFC 11 and CFC 114. These mixtures of propellants closely obey Raoult's law and therefore the blend selected can be used to give a defined vapor pressure (Table 1). The inclusion of CFC 11 in the formulation also offered advantages in that it increased the solvency of most propellant systems, thereby facilitating the dissolution of surfactants in suspension formulations. By virtue of it being a liquid below 24°C, it was used as the primary dispersion medium for either suspending or dissolving the drug.

In 1974, Rowland and Molina (9) published their hypothesis that CFCs could lead to the depletion of stratospheric ozone. With confirmation of this theory through subsequent studies, an international agreement, The Montreal Protocol on Substances that Deplete the Ozone Layer (10), was drawn up and set in motion a timetable for the phase-out of both the manufacture and use of CFCs. Although the consumption of CFCs in inhalation products is probably insignificant in ozone depletion, representing approximately 0.4% of the worldwide CFC consumption in 1986 (11), the pharmaceutical industry has been working since 1987 to find alternative propellants with which to replace the CFCs used in MDIs. The establishment of the safety and suitability of such alternatives for use in medical aerosols is a lengthy and

complex process. As a consequence, MDIs have been identified as an essential use of CFCs and have remained exempt from the provisions of the Montreal Protocol during the transition process.

The search for propellants of low or zero ozone depletion potential (ODP) has led to the identification of a number of potential compounds. A number of chemical industry consortia were established to investigate the acute toxicity of the most promising candidates under The Programme for Alternative Fluorocarbon Toxicity Testing (PAFTT) (8).

The use of HCFCs has been considered, although they still have an appreciable ODP. For instance, HCFC 22 could provide a technically satisfactory replacement for CFC 12. However, revisions to the Montreal Protocol in 1990 require the phase-out of HCFCs by 2020.

The use of hydrocarbons such as isobutane is common in general consumer aerosols. However, their odor and flammability have deterred their use in medical aerosols although purer grades, which are odorless are now available. Dimethylether (DME) is used similarly because it combines zero ODP with superior solvency for various active components and appreciable miscibility with water, which may be important in the formulation. However, the high flammability of both the hydrocarbons and DME would require expensive modification of facilities for the manufacture, storage, and transportation of MDIs. Flame extension studies have shown that the flammability of these propellants is unlikely to present a significant risk during inhalation use due to the small metered volumes (12). Another challenge in the use of both hydrocarbons and DME as propellants in suspension MDI formulations may be their low density, compared with most drug substances, which would give rise to poor suspension stability leading to the potential for inconsistent dose delivery.

Of the alternatives identified, the hydrofluoroalkanes (HFAs) or HFCs were targeted for development as replacements for the CFCs in MDIs. Within this class, 134a and 227ea were adopted for inhalation toxicity testing by two consortia of pharmaceutical companies: IPACT 1 for 134a and IPACT 2 for 227ea (IPACT: International Pharmaceutical Aerosol Consortium for Toxicity Testing). These programs established the safety profile of both propellants, which has led to the recommendation by the Committee of Proprietary Medicinal Products (CPMP) of their suitability for use in MDIs (13).

The vapor pressure of 134a is higher than that of 227ea, but both are seen as alternatives for CFC 12. To date, no suitable replacement for CFC 11 has been identified. In some HFA MDI formulations, ethanol has been used as a

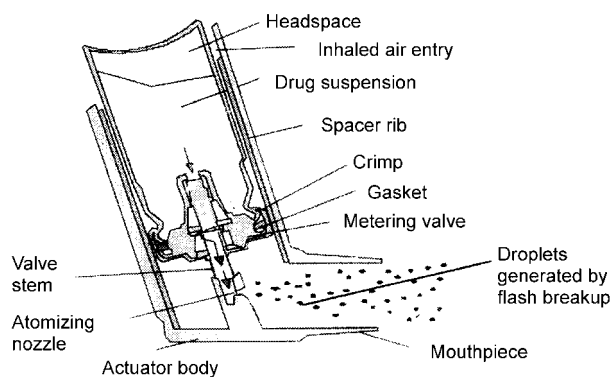


Fig. 1 Sectional view of a metered dose inhaler.

Table 1 Properties of fluorocarbon propellants and dimethyl ether^a

Propellant	Molecular formula	Molecular weight	Boiling point at 101.3 kPa (1 atm.) (°C)	Gauge vapor pressure (barr at 21 °C)	Liquid density (g/cm ³) at 21°C	Flammabil.b limits (vol. % in air)
CFC-11 ^b	CCl ₃ F	137.4	23.8	−0.1	1.49	NF ^c
CFC-12 ^b	CCl ₂ F ₂	120.9	−29.8	4.84	1.33	NF
CFC-114 ^b	CF ₂ ClCF ₂ Cl	170.9	3.6	0.89	1.47	NF
CFC-115 ^b	CClF ₂ CF ₃	154.5	−38.7	7.10	1.31	NF
HCFC-22	CHClF ₂	86.5	−40.8	8.37	1.21	NF
HCFC-123	CF ₃ CHCl ₂	152.9	27.1	0.79	1.47	NF
HCFC-124	CF ₃ CHClF	136.5	−0.11.1	2.28	—	NF
HCFC-141 ^b	CCl ₂ FCH ₃	117.0	32.1	0.69	1.24	7.6–17.7
HCFC-142 ^b	CH ₃ CClF ₂	100.5	−9.8	2.00	1.12	6.0–15.0
FC-C-318	C ₄ F ₃ (cyc.)	200.0	−5.8	1.75	1.51	NF
HFA-125 ^d	CF ₃ CHF ₂	120.0	−72.8	11.4	—	NF
HFA-134a ^c	CF ₃ CH ₂ F	102.0	−26.7	5.59 (20°C)	1.22	NF
HFA-152a ^c	CH ₃ CHF ₂	66.1	−23.9	4.35	0.91	3.71–18.0
HFA-227 ^c	CF ₄ CHFCF ₄	132.0	−17.0	3.99 (20°C)	1.41	NF
DME ^d	CH ₃ OCH ₃	46.1	−23.7	4.35	0.66	3.4–18.2

^aBased on data of E.I. du Pont de Nemours & Co Ltd., except for propellants HFA-134a and HFA-227.

^bPermitted used in most countries for specified MDI product.

^cNF = nonflammable Ozone depletion potential (ODP), relative to CFC-11 with ODP of 1.

^dHFA hydrofluoroalkane DME = dimethylether.

cosolvent to enhance the solubility of surfactants in place of CFC 11. The HFA propellants meet many of the criteria of the ideal propellant for use in an MDI, although the solvency properties of both 134a and 227ea are markedly different from those of the CFCs. Hence the conventional surfactants used in CFC MDIs were found to be incompatible. As a consequence, the development of HFA MDIs has required the resolution of a significant number of technical problems (3).

FORMULATION

In general, MDI formulations can take the form of either suspensions or solutions. Traditionally the preferred route has been to formulate a suspension of the micronized drug substance in the liquid propellant (CFC or HFA). In some cases, additional excipients (e.g., surfactants and/or cosolvents) have been added to improve the quality of the dispersion. The various MDI formulation options are described in detail later together with a description of some of the alternative options for the input drug substance.

Historical Overview of CFC-Based MDI Formulations

These typically comprised a suspension of the micronized drug substance in various ratios of propellants 12 and 11 and/or 114. A surfactant was usually added to improve the suspension behavior by decreasing the rate of flocculation and sedimentation or creaming and also by reducing the amount of drug deposition on the internal surfaces of the valve and container (4). The addition of a surfactant to moderate the suspension behavior and reduce internal drug deposition, can make the valve sampling and hence the dose delivery more reliable. Some CFC MDI products have also been formulated as solutions (14). Typically, the drug substance is solubilized by adding a small quantity of ethanol. The potential strengths and weaknesses of both suspension and solution formulations are discussed in detail in the following section, which covers formulation options for HFA-based MDIs (15).

HFA MDI Formulations

Input drug substance

The microfine drug powder used in suspension-based MDIs has typically been prepared using an air-driven fluid energy mill and this is still the method predominantly used for the currently marketed suspension-based MDIs.

In recent years, there has been a recognition that although the micronization process is a cost-effective method of producing a material with the correct particle size distribution for inhalation, the process, which relies on particle–particle collisions and attrition to gradually reduce the particle size, may also impart some undesirable characteristics to the particles produced.

As a consequence of the limitations of the micronization process, various particle engineering (16, 17) techniques are under investigation with a view to providing superior input microfine drug for MDI manufacture. Some of these techniques, for example, SEDS (Solution Enhanced Dispersion by Supercritical Fluids) (18) rely on alternative crystallization technologies to go straight from a solution of the required drug substance to crystals with exactly the required surface characteristics and particle size distribution without going through the high-energy micronization process. Other techniques extend spray drying technologies to produce particles with controlled density, porosity, particle size distribution, and shape as well as allow for coformulation with other solid excipients to produce particles with controlled drug release characteristics as well as potentially controlled regional deposition within the lung (19, 20). Particle engineering techniques may also offer enhanced performance to meet increasingly stringent regulatory requirements (21, 22).

HFA MDI solution formulations

In an HFA MDI solution formulation, the drug is completely dissolved using an HFA propellant (e.g., 134a) plus an appropriate cosolvent to produce a pure solution product. The cosolvent most commonly used is ethanol. This approach has been used successfully to produce a solution aerosol product for BDP (23). There are a number of advantages to a formulation of this type:

1. Potentially fewer issues around homogeneous valve sampling from the bulk.
2. As in the case of the BDP solution aerosol, enhanced efficiency of aerosolization, leading to high lung deposition compared with an equivalent suspension product.
3. Overcomes issues with suspension systems where the drug has measurable solubility in the propellant, i.e., formulation does not suffer from particle growth issues.
4. Provided there is always sufficient cosolvent present to preserve the true solution status of the product, no issues with drug deposition on the valve components and container.
5. May be a simpler filling process than for a suspension-based HF MDI.

There are also a number of disadvantages to true solution MDI products:

1. Solution products can be more susceptible to drug losses into the elastomeric components of the valve than for an equivalent suspension product. Drug losses of this kind can be an issue particularly for low dose products.
2. There are currently few options beyond ethanol in terms of cosolvents that would be Generally Regarded As Safe (GRAS). Thus, to develop a solution formulation using a cosolvent other than ethanol could necessitate an extensive toxicology testing program. Of the currently available respiratory drugs, there are few that are sufficiently soluble in ethanol or sufficiently low in strength to easily yield a solution product.
3. If too much ethanol is used to dissolve the drug, the vapor pressure may drop below that required to achieve efficient atomization.
4. If the product under development is transitioning from an existing CFC suspension-based product, it may be difficult to replicate the fine particle mass, the particle size distribution and/or the absorption characteristics. This can lead to problems in providing a seamless transition from the CFC product because of the potential to have to change the dose and because the product can smell and feel different to the original suspension formulation.

More recently, other (less volatile) organic modifiers, e.g., glycerol, have been added to solution-based HFA MDIs to modify the particle size distribution so that it more closely resembles that of the originator suspension product (24).

HFA MDI suspension formulations

There are a number of possible approaches to a suspension-based formulation. A micronized drug can simply be suspended in an HFA propellant or a mixture of HFA propellants. The principal advantage to a formulation of this type is that it is simple and contains no additional excipients with their inherent toxicological implications. The performance of a formulation of this simplicity will be dependent on the inherent properties of the drug substance and the propellants used. For example, if the drug substance is significantly more dense than the propellant(s) is, then rapid sedimentation of the suspension is likely to occur following agitation. This could create issues in terms of the valve sampling homogeneously from the bulk container contents. Further differences between the drug and propellant in terms of relative hydrophobicities and hydrophilicities can also result in rapid flocculation immediately postshaking or a tendency for the drug to

deposit on the MDI container walls and valve components. For the approach of creating a dispersion of drug in propellant to be successful, the drug should essentially be insoluble in the propellant(s) to provide good product stability. The formulation should also possess reasonable characteristics such that the suspension is easy to redisperse so that the valve still samples homogeneously from the bulk suspension contents in a time-scale consistent with the gaps a patient would typically leave between shaking and firing the inhaler. The tendency for drug to deposit on the inner surfaces of the container and valve can be controlled via sophisticated packaging technologies, e.g., can coating, making this kind of MDI formulation successful for a number of drug substances.

For CFC-based suspension formulations, a surfactant was typically included. A variety of surfactants were used in these systems, e.g., lecithin, oleic acid, sorbitan trioleate (14). All these surfactants were freely soluble in the CFC propellants and allowed for a degree of control over the suspension characteristics. Rates of flocculation, sedimentation, and creaming could be controlled and deposition on the internal container components was minimized. The transition to HFA-based MDIs has created significant issues in that none of the surfactants, previously used with the CFC products are soluble in HFA propellants alone. Some formulations have still used these surfactants, but the addition of a cosolvent (ethanol) has been required to solubilize the surfactant.

In addition to HFA formulations using the traditional surfactants plus cosolvents, there has also been some research work to identify novel surfactant molecules for use with HFA propellants. As yet none of these research programs have successfully yielded molecules with suitable properties to control the suspension characteristics whilst also retaining a suitable toxicology profile.

An interesting alternative approach to suspension formulations in HFA propellants is via engineered drug substance particles. Porous drug substance/excipient particles can be produced via spray drying techniques. It is theoretically possible to match the density of the particles with that of the liquid propellant to produce a formulation with desirable suspension characteristics (19).

CONTAINERS

The essential requirements of containers used for MDIs are that they are compatible with the formulation, have an ability to withstand internal pressures up to 1500 kPa, and can be manufactured with reproducible quality. The most widely used containers for MDIs are made from an

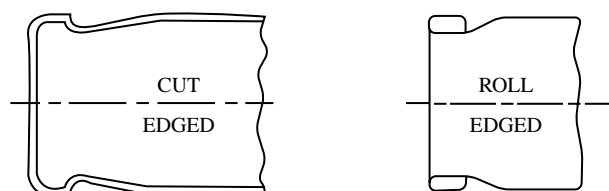


Fig. 2 Types of aluminum cans used with MDIs.

aluminum alloy, although glass bottles have also been used. Aluminum containers (cans) are preferred due to their light weight, strength, break resistance, compactness, and ability to provide light protection. There are two main types of aluminum cans—the cut-edge can and the rolled-edge can (Fig. 2). The cut-edge can, which is most commonly used in the manufacture of MDIs, is manufactured by a deep-drawing process to leave a sharp edge to the can rim (Fig. 3). After formation, the cans are washed with solvent and aqueous detergent to remove residues of oil, which is used as a lubricant in the drawing process. Rolled-edge cans are formed by impact extrusion (slugging) of a slug of high purity aluminum. The excess material is trimmed off and the can neck is externally rolled. The deep-drawing process is preferred because the cans have a more uniform weight and wall thickness, which facilitates fill weight control during the MDI manufacturing process.

Although aluminum cans generally show good compatibility with the propellant systems used in MDIs, drug degradation or drug deposition on the internal walls of the can may require the use of internal coatings. A suitable epoxy resin or phenolic vinyl type coating may be used (25). More recently, low-surface-energy coatings based upon perfluoropolymers have been proposed for use with HFA propellant systems (26).

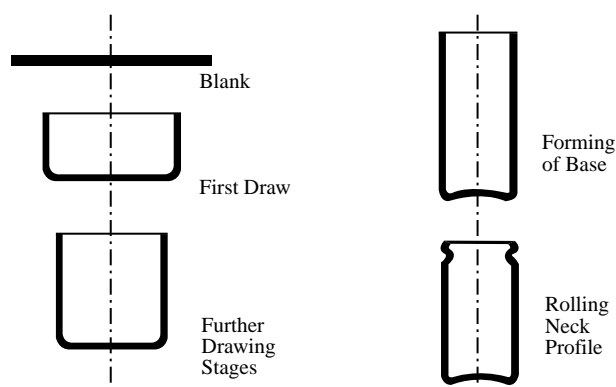


Fig. 3 Stages in the manufacture of aluminum cans by a deep-drawing process.

Glass bottles are not widely used because of their greater weight and potential fragility. Glass containers used with pressurized liquefied propellants are externally coated with a plasticized polyvinyl chloride (PVC) layer to retain glass fragments in the event of breakage. In the past, glass bottles may have been preferred for MDI formulations containing ethanol and CFC 11. Hydrogen chloride, formed via a free radical reaction, can result in corrosion of aluminum cans (27).

METERING VALVES: DESIGN AND FUNCTION

The primary function of the metering valve is to meter accurately and repeatably small volumes of the propellant-based formulation containing the drug. Secondly, it helps seal the pack to ensure minimal leakage of propellant. Most of the commercially available MDI products are fitted with metering valves designed to operate in the valve-down orientation, which eliminates the need for a dip tube. The typical MDI metering valve consists of two coupled valves placed on either side of a volumetric metering chamber (28). It is essential that during depression of the valve stem, the inner “valve” closes to isolate the container contents before the outer “valve” opens to allow discharge of the contents of the metering chamber. Reversal of this sequence would give rise to “continuous spraying,” placing the patient at risk of overdosage. The metering performance of the valve is a critical aspect of quality control. A diagrammatic representation of the sequence of operation of a metering valve is shown in Fig. 4.

A metering valve comprises a number of essential components (Fig. 5) having specific functions (Table 2). The most significant feature of the valve are the two seals or seats. These are separated by the metering chamber, which is created by the space between the stem, the body, and the seals. The cylindrical stem passes through the flat elastomeric seals to create a dynamic seal during movement of the stem along its axis. With the exception of the aluminum ferrule, all other components are either press-formed in stainless steel or injected-moulded in suitable plastics such as nylon, polyacetal, or polyester, which are dimensionally stable when in contact with the propellant. The stem return spring is invariably made in stainless steel. Good surface finish and careful dimensional control of all the components and their correct assembly are critical in ensuring satisfactory valve performance.

The valve is sealed against the container by a rubber flat cut or “O” ring gasket that is compressed by crimping the aluminum valve ferrule to the container neck.

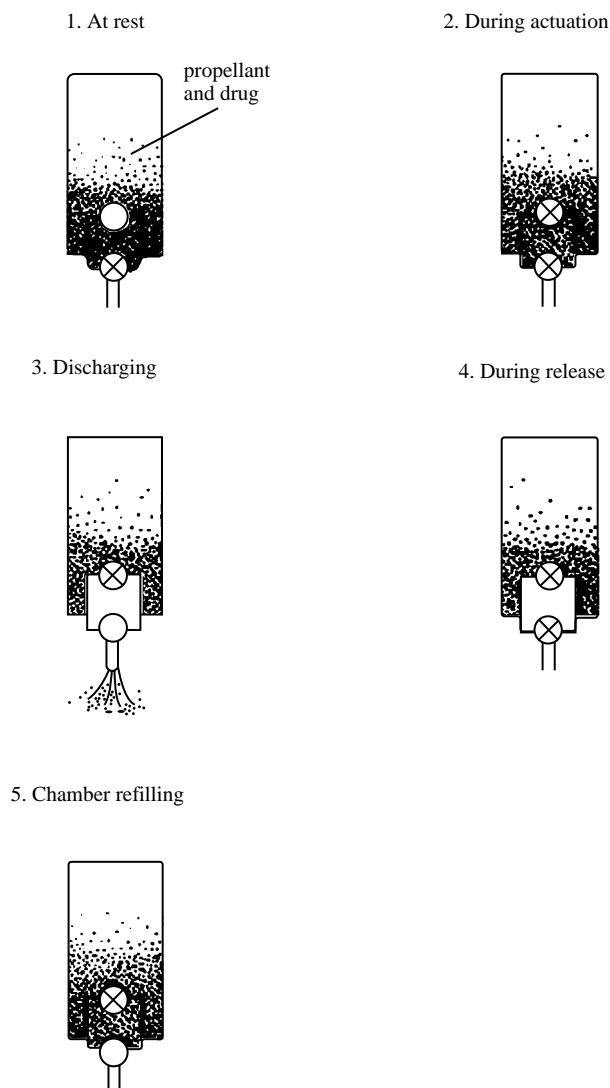


Fig. 4 Operation of an aerosol valve.

Crimping during the filling operation is controlled by monitoring the crimp height and diameter to ensure adequate compression of the gasket. The optimum crimp parameters will permit correct valve function and minimize propellant leakage.

ELASTOMER COMPONENTS

The elastomeric sealing components of the metering valve are particularly critical. In those valves used with CFC propellants, the elastomeric seals have typically been formed from an acrylonitrile/butadiene rubber, which has been cured with sulfur. These rubber seals may not be fully

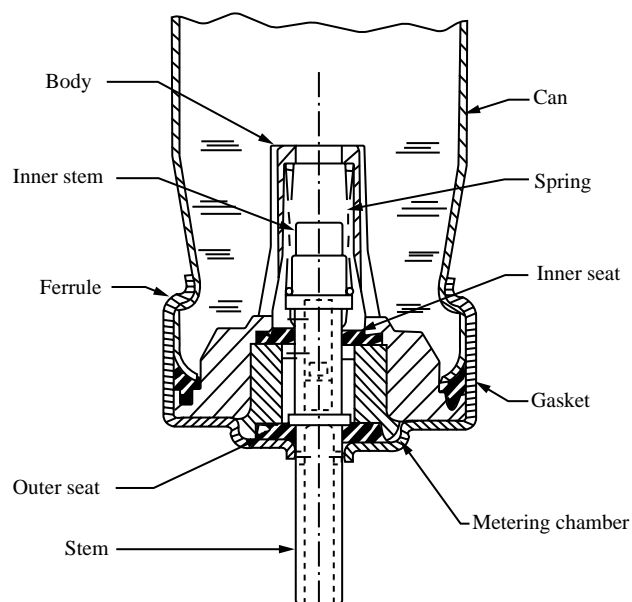


Fig. 5 Components of a metering valve.

compatible with HFA propellants; hence, alternative elastomeric materials have been used. These materials include peroxide-cured acrylonitrile/butadiene, ethylene-propylene diene monomer (EPDM), and chloroprene and thermoplastic elastomers (TPE). The elastomeric materials used to form the dynamic seals around the stem and the static gasket seal between the can and valve may differ based on the required properties of the rubber for the specific function of the seal (29). The most important characteristics of the elastomeric seals include their composition and degree of curing, which control the physical properties and swelling characteristics in situ, compatibility with the propellant and drug, their affinity for drug absorption, their dimensions and surface finish, and their potential to release foreign particulate matter and extractable material. The solvency properties of both CFC and HFA propellants is such that low levels of extractives from the elastomers can appear in the MDI formulation during storage. These may not present any safety concerns and their levels can be further minimized by using solvent-extracted elastomeric seals. The elastomeric seal material may also be selected, in part, to influence other changes that can occur during storage of an MDI. Under normal storage conditions, moisture will diffuse through the elastomeric seals into the MDI contents. The rate and quantity is dependent on the environmental conditions and the nature of the elastomeric material, the water vapor transmission properties of which can differ significantly (30).

Propellant leakage from an MDI is characteristic of this product and occurs by diffusion through both stem seals

Table 2 Function of the valve components

Valve component	Function
Gasket	Seal between the valve and can; usually made of rubber.
Ferrule	An aluminum cup that holds the valve components together and attaches them to the can in the crimping process.
Stem	Moving part of the valve that provides the metering action and connects the valve to the actuator. Its design provides entry and exit ports to the metering chamber to permit filling and discharge of the metered dose respectively.
Seat	Provides the main seal around the valve stem and is usually made from rubber. Most valves have two seats to provide the metering action.
Spring	Returns and holds the stem of the valve in the rest position after actuation. May be located inside or outside of the metering chamber.
Metering chamber	Defines the volumes of liquid discharged. Valves with 25-, 50-, 63-, and 100- μ l nominal metering volumes are available.

and gasket of the valve. The leakage rate increases with increase in environmental temperature, and tends to parallel the rubber swell for a particular propellant–rubber combination. Propellant leakage from an MDI may shorten the product shelf-life by raising the drug concentration of the liquid contents, thereby increasing the metered dose of medicament. Additionally, excessive leakage may prevent the labelled number of actuations being delivered by the MDI.

ACTUATORS AND SPACERS

Overview of the Components for the Basic MDI Actuator

For a simple MDI, the actuator is a one-piece plastic molding that performs a number of critical functions as a key packaging component in the overall system. The main plastic body of the actuator surrounds and protects the aerosol canister, and veins within the actuator help to locate the canister so that when it is depressed by the patient to release the dose, the canister moves straight down without flexing the valve stem. A high degree of flex could result in

**Fig. 6** Typical MDI actuators.

poor performance, e.g., as a result of continuous spraying of the valve. The veins also serve to centrally locate the canister within the actuator body to create airflow paths with minimum resistance so that air can be easily drawn through the device as the patient inhales (67). Some typical MDI actuators are shown in Fig. 6.

The actuator also incorporates the stem block containing the spray nozzle. The stem block contains a socket that the valve stem pushes into when the canister is placed inside the actuator. The socket is designed to provide a tight interference fit with the valve stem such that there is no tendency for the canister to fall out of the actuator on transportation and also so that the complete dose is pushed out of the spray orifice and does not leak back up the side of the valve stem to deposit on the valve ferrule. The final critical component of the actuator is the mouthpiece. The design of this component needs to be such that the patient is easily able to form a seal around it to draw air through the device and inhale the dose. The mouthpiece should also be designed to minimize the extent of drug deposition on the actuator during dose delivery. Typically, 10% to 20% of the dose delivered from the valve is deposited on the actuator and is therefore not available for delivery to the patient. Mouthpiece designs are usually a compromise of what is cheap and easy to mold, what is easy for the patient to use and form a seal around, and what will result in a small and portable device with acceptable drug delivery characteristics. A huge advantage to the MDI is that it is small, light, easy to carry, and discrete to use. A large and complicated design optimizing the delivery characteristics would probably be expensive for the manufacturer and unattractive to patients. A further component of most MDI actuator designs is a dust cap. This is most usually a separate plastic molding that

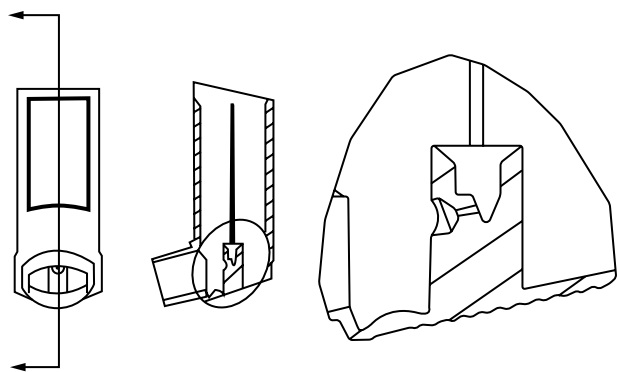


Fig. 7 Detailed diagram of MDI actuator stem block.

snap-fits onto the main actuator body. The main purpose of the dust cap is to prevent foreign matter from entering the mouthpiece and valve stem area while the device is being transported. Foreign matter in this area could either be inhaled by the patient during use or could potentially block the spray nozzle. A detailed cut-away drawing of a typical actuator stem block is shown in Fig. 7.

For a simple MDI actuator, the spray nozzle and valve stem socket are formed by metal pins that enter the stem block as part of the actuator molding process. It is essential to retain accurate dimensional control for these critical areas, to ensure absence of leakage and correct and consistent spray performance during discharge. Various combinations of nozzle diameter and length and expansion chamber volume (volume between the metering chamber and the actual spray nozzle, principally the internal volume of the valve stem), have been investigated. In practice, nozzle diameters range between 0.2 and 0.5 mm. Small nozzle diameters tend to give higher fine particle masses but also increase the length of time it takes to deliver the dose and can be more prone to blocking (31). The actuator parameters cannot be optimized in isolation but must be designed to work in harmony with a specific metering valve, container, and formulation combination (32).

Spacers

As mentioned earlier, the simplest MDI actuators are a compromise between portability, cost, and performance. When a simple MDI is used, a significant portion of the dose is deposited in the oropharyngeal cavity. This is because when the canister is pushed down for the dose to be delivered, the spray emerges from the actuator orifice at such a rate that there is insufficient space for the cloud to decelerate before it strikes the back of the throat or other parts of the oropharyngeal cavity. Simple low-volume spacer devices (70–150 ml) provide an extension to the mouthpiece to allow space for the deceleration of the

aerosol cloud to occur. These can either push-fit onto the simple classical type of actuator or can unfold as part of an integral actuator/spacer design (33–37). They significantly reduce the amount drug deposition in the oropharyngeal cavity (35) by allowing more time for the aerosol to decelerate. As a consequence, slightly more drug may be deposited in the lungs but the majority of the reduced throat deposition is accounted for by higher device/spacer deposition. The principal benefit in using this type of simple spacer is in reducing oropharyngeal irritation and hoarseness caused by the deposition of inhaled steroids (34, 38).

Spacer/Holding Chambers

In addition to the simple spacer devices that have been described, there are also much larger volume spacer devices (500–750 ml). These devices provide all the benefits of the simple devices but are capable of capturing and holding the dose prior to the patient inhaling. A one-way valve (or paired valves) may also be incorporated into the mouthpiece region to allow natural breathing cycles so that the patient breathes out into the atmosphere but breathes in through the spacer to inhale the dose held within the spacer. These larger spacers may be beneficial for asthmatic patients who show poor coordination of inhalation with inhaler actuation (33, 34) because actuating the canister to deliver the dose does not have to be coordinated with breathing in (39, 40). Drug deposition in a spacer increases with decreasing chamber size (as a function of width and length) (35, 36). Spacer deposition may be influenced by the size of the metering valve and the propellant composition (37). The delivery of a respirable drug from MDIs is generally similar with or without a larger spacer but with the advantages of lower oropharyngeal deposition and lower need for accurate coordination.

The materials used to manufacture spacers are also key to their performance. A build-up of static electricity on a spacer device can significantly affect the amount of drug deposited on it and, consequently, the dose the patient could receive (41). Similarly, the way in which spacer devices are washed or dried can also result in a build-up of static electricity (42).

BREATH OPERATED INHALERS AND OTHER DEVICE ENHANCEMENTS

The MDI has proved itself over many years of successful use as a highly effective method for delivering drug to the

lungs for the treatment of respiratory disease. The MDI is simple to use for most people, but there are difficulties that certain sectors of the patient population may have with the use of this device type.

Very young children and older people have significant difficulty coordinating their actuation of the device with breathing in (43). Patients who have this coordination difficulty can either be supplied with a dry powder inhaler (where the inspiration through the device is also responsible for the release and aerosolization of the powdered drug so that coordination is no longer an issue), or they may be prescribed a Breath Operated Inhaler (BOI).

Several of these device types are marketed; they are still MDIs but the device takes over the responsibility for actuating the can as the patient breathes in. Typically, the device is primed immediately before the patients takes their dose. This priming can be achieved either via the opening of the dust cap to reveal the mouthpiece or by operating a special priming lever. For the currently marketed devices, this results in a spring being compressed above the aerosol canister that has enough strength to actuate the can on release. The device contains a triggering mechanism that is released when the patients breathe in through the device to take their medicine. The release of the triggering mechanism allows the coiled spring to push the MDI canister down and deliver the medicine, negating the need for the patients to coordinate the activities (44). Although there is significant merit to devices of this type, they need to be carefully developed to ensure that the altered pattern of use in terms of priming, shaking, and taking the medicine works in harmony with the other device components to ensure that consistent doses of medicine are always delivered from the device.

More BOI devices are likely to reach the market. They could be either mechanical or electronic in operation. The incorporation of electronics into a device of this type opens up the possibility of including other patient features, e.g., counters and other data logging.

MDI COUNTERS

One significant issue with MDIs is that because of the aluminum canister used for most device types, it is impossible for the patient to know exactly how much medicine is left in the inhaler. Shaking the canister prior to taking the medicine can give the patient some confidence that formulation is still there within the device, but this is a very crude measure. MDIs are always filled with more medicine (termed the overfill) than the number of actuations mentioned on the label. Thus a 200 actuation

product might actually be filled with 240 actuations. There are a number of reasons for this:

1. There is always some leakage from devices of this type throughout their life. Including an overfill guarantees that the device is still capable of delivering the claimed number of actuations at the end of the shelf life.
2. There can be some variability in the dosing performance of an MDI as the formulation level becomes very low. Including an overfill guarantees that the MDI will perform consistently throughout the number of actuations claimed on the label.
3. There is a portion of the formulation that is not removable from the canister because of the positioning of the metering valve sampling port.

Using a glass container is one method of overcoming the difficulty by knowing how much medicine is left, but because of the overfill factor there will always appear to be medicine left even after the patient has exceeded the number of actuations claimed on the label. Also, glass containers may be more bulky than their aluminum counterparts and are certainly more expensive to produce.

The alternative to a transparent container is to fit the MDI with some kind of actuation counter or level indicator. The level indicator, much like the fuel gauge within a car, will give some idea of the amount of medicine that remains, but will not be an exact measure. A numerical counter is probably the preferred solution as it leaves the patient in absolutely no doubt about the amount of medicine remaining in the device. When the number of actuations claimed in the label have been taken from the device, the patient will know that a replacement is needed, thereby completely avoiding the uncertainty that would arise from continuing to use the device during the overfill phase between label claim and final device exhaustion. The counting mechanism could be either applied to the top of the aerosol canister or the valve area or be an integral part of the actuator and be mechanical or electronic in operation.

OTHER DEVICE ENHANCEMENTS

The incorporation of breath activation and/or counting mechanisms into the MDI opens the possibility of including other device features, particularly if the above-mentioned two features are achieved via electronic means. MDIs that feature data gathering technology allow the patient or physician to monitor how and when the medicine is being taken (compliance) or to monitor lung function and control the amount of medicine that the patient is taking are distinct possibilities.

Spray Atomization and Evaporation

The respirable fraction or Fine Particle Mass (FPM) of a finely aerosolized drug delivered from an MDI is highly dependent on the atomization of the formulation and the subsequent spray dynamics. The aerosol characteristics depend heavily on interactions between the propellant(s), the (typically) micronized drug particles, the formulation excipients, and the design and dimensions of the metering valve and the key actuator variables (stem block, atomization orifice, actuator airflow paths, and mouthpiece design). The importance of these inhaler variables has been demonstrated previously (45–47). MDIs give a much slower metered delivery and a much finer spray droplet size distribution than what would be expected for simple hydraulic atomization because a mixture of gas and liquid is discharged from the metering chamber.

Partial flash evaporation of the propellant occurs in the metering chamber and in the expansion spaces of the valve stem and actuator stem block (47). Further evaporation also occurs within the discharge nozzle, and alternating segments of liquid and gas pass through the nozzle in a process of effervescent atomization.

The spray droplets are ejected at a high velocity of about 25–30 m/s (48), and although the aerosol decelerates rapidly the velocity is still much higher than the inhaled air velocity, resulting in significant drug impaction in the oropharyngeal region. This effect has been mimicked crudely in simple throat models (49) and more accurately in throat models having carefully matched human dimensions (50).

The probability of oropharyngeal deposition is determined more by droplet size than by velocity and density because the particle inertia is proportional to the density, velocity, and the square of the diameter. It, therefore, follows that oropharyngeal drug deposition is reduced and the respirable drug delivery is increased when MDI sprays are finely atomized and evaporate rapidly. Such MDI sprays are generally promoted by increasing the propellant vapor pressure (45, 47, 51) and reducing the actuator spray nozzle diameter (45, 46, 52).

Spray evaporation is impeded by the inclusion of appreciable concentrations of nonvolatile miscible excipients in the formulation. These include ethanol as a cosolvent (53), and CFC 11, both having low volatility in the spray, which is also chilled by propellant evaporation, thereby further reducing the rate of evaporation. The particle size distribution of the resulting aerosol depends not only on the spray dynamics but also on the number and primary size distribution of the drug particles in liquid suspension and their degree of deagglomeration during

spray atomization. The evaporated aerosol particle size increases with the concentration of drug (45, 54) and nonvolatile excipients (45, 55) in suspension formulations and also in solution products.

MANUFACTURING

In the manufacture of MDIs, liquid filling procedures have been developed based on either a cold-filling method or a pressure-filling method. Both methods are suitable for either solution or suspension formulations and regardless of the process, it is important to maintain a low atmospheric relative humidity in the filling area to minimize condensation and possible absorption of water by the product.

The cold-filling process involves chilling the propellant and drug formulation to approximately -60°C in tanks to maintain the liquid state and metering it into the open cans in a single step. The cans are then sealed immediately by crimping on the valves.

The potential advantages of this procedure are that it is simple and can accommodate any valve, with minimal changes to the production line. The potential disadvantages include high energy consumption (refrigeration), inconsistent fill weights due to propellant evaporation, contamination from moisture condensation, and possible irreversible physical changes in the formulation at low temperature.

The pressure-filling process is now more commonly used and relies on the injection of at least part of the fill through the valve, which is crimped to the container. Pressure filling with CFC propellants generally involves preparing a concentrated suspension (or solution) of the drug in the high-boiling CFC 11, together with any excipients. This concentrate is metered into the open container in a cool area to minimize evaporative losses of propellant before crimping on the valve. The low-boiling propellant, CFC 12 or a CFC 12/CFC 114 blend, is then filled through the valve by high-pressure injection, a process commonly known as gassing. For this process, the potential advantages are that accurate fill weights are achievable and that filling can be undertaken at or near room temperature. The potential disadvantages are that the filling equipment is complex and expensive to install and that the valves must be suitable for pressure filling.

With the transition to HFA/HFC propellants, the lack of a direct replacement for CFC 11 has resulted in the development of a single-stage pressure-filling process. For the HFA/HFC propellants, a concentrated suspension (or solution) of drug is prepared in the propellant

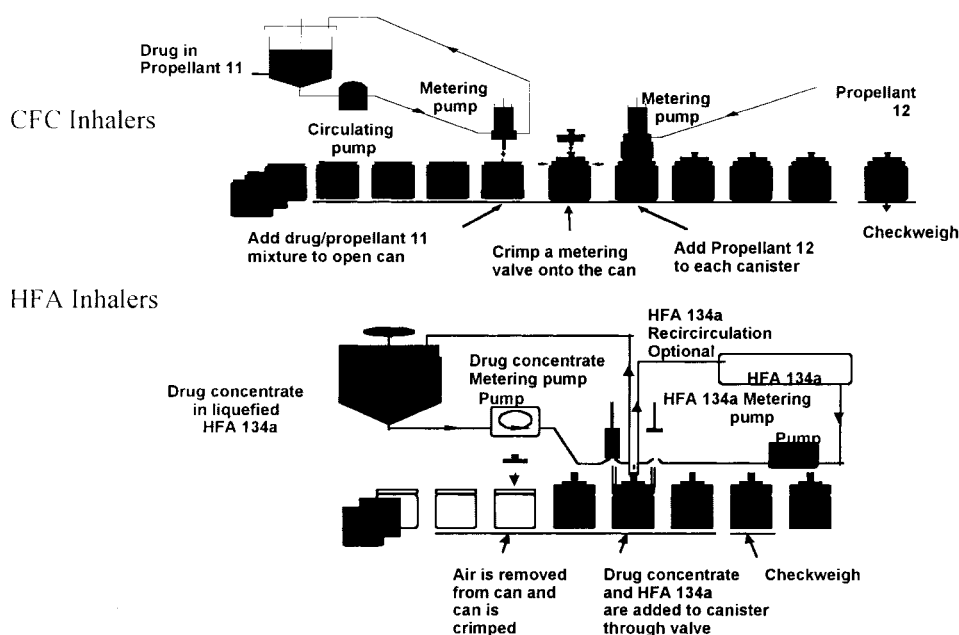


Fig. 8 Production pressure-filling sequence for MDI cans.

(134a or 227ea) contained in a pressurized system. A valve is crimped to a container and the concentrate is pressure-filled through the valve, followed by an aliquot of pure propellant (56). Such a process requires the development of a total pressurized filling system with compatible elastomeric seals and the use of pumps capable of recirculating, without cavitation, high-pressure liquids.

During the filling process, it is normal to displace the air from the container before valve crimping to avoid internal pressure changes in the MDI at the time of progressive discharge of the contents during use. For the cold filling and two-stage pressure-filling processes, air displacement occurs after filling the initial propellant liquid into the container. However, in the single-stage pressure-filling process, air is purged from the container in a separate operation prior to crimping the valve, using either a small injection of propellant or vacuum (Fig. 8).

High-speed rotary filling machines are widely used to provide concentrate filling, valve placement, and crimping and gassing in a continuous sequence. Similar set-ups are used for HFA/HFC MDIs with appropriate modifications to permit container purging. Intermittent indexing machines are often employed for small-scale filling. Filling lines may also be equipped with a checkweighing station to reject any over- or under-filled canisters. In addition, a heat-stressing facility may be employed after filling to raise the temperature of the MDIs to a predetermined value (e.g., 50–55°C) to increase the

internal pressure and thereby challenge the integrity of the crimp seal. After a defined quarantine period, further checkweighing, normally as part of the final packaging operation, can be used to reject canisters that have leaked.

QUALITY CONTROL

Quality control testing of MDI batches is applied to the individual inhaler components prior to manufacture, as in-process controls during the manufacturing, and to the finished product (Table 1). A number of publications describe quality control procedures for MDIs (21, 57, 58).

Component and In-Process Quality Control

Typical QC procedures for the aluminum canisters would include careful assessments of their dimensions, weight, strength, surface finish, and cleanliness. Important canister characteristics are neck finish and dimensions, regional can wall thickness, and external size of the canister, which influence valve sealing, canister strength and volume, and the fit into the actuator. Canister strength is typically monitored by the manufacturer, using a hydraulic pressure test. Glass bottles may be subjected to impact fracture drop tests.

The metering valve is generally the most critical inhaler component in terms of its effect on product performance;

Table 3 Quality control and analytical tests for the characterization of metered dose inhalers

Name of test/test area	Type of data/comments
<i>Individual component testing prior to manufacture of active product:</i>	
Containers	Dimensional accuracy and surface quality; strength testing by canister manufacturer
Metering valves	Dimensional accuracy and surface quality of individual components from dismantled valves
Mean weight per actuation and leakage rate	Samples filled on the production line and shot weight through use data generated to guarantee correct valve function; samples tested after a period of equilibration to ensure leak rates within specification
Input micronized drug	Chemical, physical, and particle size specifications
Propellant	Chemical and physical specifications
Surfactant	Chemical specification
Actuators	Dimensional accuracy and finish; accuracy and centralization of the atomization orifice
<i>In-process control testing that occurs during manufacture:</i>	
Atmospheric conditions	Monitoring of atmospheric temperature and humidity in manufacturing areas
Drug suspension concentration	Rapid analysis initially and during filling
Drug suspension	Periodic fill weight checks (if separate suspension followed by gassing)
Filled canisters	Automatically checkweigh all canisters
Gross leakage and safety	Containers heated to raise internal pressure; visible leakage and pack distortion examination
Control of leakage rate	Checkweigh all containers after a suitable quarantine period to meet specified limits
Metering-valve function	Each filled container is spray-tested after quarantine, following a number of priming actuations
<i>Analytical testing applied to completed product:</i>	
Appearance	Description of the suspension; absence of corrosion of canister; completeness of inhaler components
Identity	HPLC or infrared confirmation of identity and absence of polymorph or incorrect solvate
Microbial limit testing	e.g., to USP requirements
Spray pattern	Consistent, concentric pattern fired straight with respect to mouthpiece geometry
Water content	Karl Fischer or similar method
Foreign particulate matter	BP 1993
Leachables	Consistent profile between batches, careful consideration of levels versus specifications based on individual components
Pressure test	USP
Leak test	USP
Drug related impurities	Comparison with specifications based on preclinical/clinical batches
<i>Characterization of particle size distribution:</i>	
Multistage impactor	Careful consideration of particle size distribution profile data and a multipoint specification; individual stage data may be combined for comparison with the specification for QC purposes
Laser particle sizing	Tends to be used to characterize products more in development
Microscopy	Tends to be used to characterize products more in development
Characterization of drug delivered from the MDI	
Mean dose delivered and content uniformity	Careful consideration against various specifications covering mean and individual actuations through use assessing inter- and intracan variability; regulatory guidance varies from region to region (see Refs. 22, 58, 59)
Number of primes	Test to demonstrate the number of priming actuations the patient must fire to waste before performance is nominal
Number of actuations	To demonstrate that the product delivers nominal dose at the end of label claim (can be derived from CU data)
Actuator deposition	Generated during development to determine how much drug deposits on the actuator with each dose and thus what the ex-actuator dose is

therefore, it also requires detailed testing prior to use. The disassembled valve components are examined for dimensional accuracy and surface quality. Assembled valves are tested for correct metering function and formulation delivery. Test packs are stored until equilibration and the units are tested for propellant leakage (weight loss) and weight of metered delivery. The reproducibility of delivery has been reported for various valve propellant combinations (59). Where propellant mixtures are used in the product, there is some progressive fractionation during discharge of the pack contents because of greater loss of the more volatile component (59). The associated change in liquid propellant density can cause a small progressive change in metered delivery weights, but this does not influence the drug delivery, which is determined by the metered volume.

Common in-process control procedures for MDIs are provided in Table 3. In addition, stringent environmental controls are required for air cleanliness, humidity, and temperature during the manufacturing run. Rapid verification of the drug content of the fresh bulk suspension or solution may be important prior to commencing the filling run because reworking of aerosol products for recovery is not practical. It may be necessary to monitor the dispersion quality and drug concentration of suspension during filling. Various online monitoring methodologies are under development but still the principal method is via measuring the total drug per can for a sample taken from the manufacturing run. Inline checkweighing of the crimped units provides an indirect record of the suspension fill weight per canister.

MDI BATCH ACCEPTANCE

MDI products are subject to batch control and acceptance tests similar to those for other pharmaceutical dosage forms, that is, active drug identification, dose delivery, and dose uniformity. Additional special tests unique to inhalers, e.g., characterizing the particle size distribution of the delivered aerosol, are also applied. Typical tests are shown in Table 3.

It is now rare to specify determination of the metered drug delivery by discharging directly from the metering valve (without the actuator) into a solvent (60). Via this route, correction is subsequently made for the amount of drug that typically deposits on the actuator to leave the ex-device dose available to the patient. The preferred method now is to discharge the inhaler with the actuator fitted, into a trapping system with an airflow being drawn through it (e.g., the USP/NF method) (61). This method of

characterizing the ex-actuator dose is much more representative of the real patient case. Typically, the mean specified delivery ex-actuator is controlled to within 85–115% of the label claim, with additional limits also applied to the spread of individual doses.

PARTICLE SIZE ANALYSIS OF SPRAYS AND AEROSOLS

The characterization and quality control of the particle size distribution of the discharged aerosol has become one of the key tests applied to MDI and other inhaler products, and a wide variety of methods have been developed to make this possible. The available methods can be broadly split into two categories: optical (typically laser) methods or methods based on inertial impaction.

Optical methods typically have the advantage of producing detailed results very quickly on a small number of actuations, but the disadvantage is that they do not characterize the whole spray emerging from the device. They are also not able to distinguish drug particles from drug-free excipient particles or propellant alone droplets or even between two different drug substances as in the case of a combination product. The sizing results obtained via optical methods may give an incorrect or biased impression of the true particle size distribution for the pharmaceutically active substance. Nevertheless, these methods can prove to be extremely useful from a qualitative perspective.

SPRAY DROPLET SIZE ANALYSIS

The spray droplet size distribution has a major influence on the amount of oropharyngeal deposition and the amount of drug delivered to the lungs and the regions of lung deposition. Methods for spray droplet particle sizing from MDIs (4, 6) have been reviewed.

For the quality control of actuators, spray flume appearance and profile are often used (62). These methods potentially highlight any gross issues with the moulding of the actuator atomization orifice and stem block (e.g., firing off-centre or poor spray homogeneity). Flume profiles can also be obtained by timed flash photography (4, 63), but an even better picture of dynamic flume changes can be provided by high-speed video recording (62, 64).

Spray droplets have also been analyzed microscopically after impacting the spray on to specially coated slides. Microscopic laser holography has been used to measure droplet size distributions and concentrations for MDIs

(65). Laser video imaging has also been applied to MDI sprays (66).

Laser diffraction (LD) size analysis is a rapid and convenient noninvasive method used extensively for measuring the droplet size distribution of industrial sprays. LD analysis has been used for nonmetered dispenser sprays to study the effects of varying the propellants (67, 68) and valve orifices (69).

Oropharyngeal spray impaction depends on the droplet velocity and size. Both these parameters can be measured simultaneously and recorded for each droplet. With Phase Doppler Anemometer (PDA) instruments, useful results have been obtained on MDI sprays (70).

MDI AEROSOL PARTICLE SIZE DISTRIBUTION ANALYSIS

Methods for the size analysis of MDI aerosols have been reviewed (6, 71–73). Various sizing methods are available. Microscopic examination of particles based on discharging the spray against a slide (72, 73) can provide a limit test for unduly large single particles resulting from poor micronization or poor physical stability of the drug in liquid suspension, but this is only a very gross measure.

INERTIAL IMPACTORS

Inertial impactors fractionate aerosol particles aerodynamically according to their inertia, which increases with size. Inertial impaction is the most widely used method for sizing MDI aerosols, partly because a large sample is measured, consisting of the total aerosol generated from a number of actuations. Methods of this type have the advantage that the aerodynamic measurement accounts for the effects of particle shape and density and that the aerodynamic size distribution is measured specifically in terms of the drug mass independently of any other aerosolized constituents that may be present, giving a specific measure of the aerodynamic size distribution of the drug.

The most commonly used tool for the aerodynamic classification of particle size distributions is the Andersen cascade impactor. The Andersen impactor was originally developed for the environmental monitoring of air samples but is now commonly applied to characterizing medical aerosol size distributions. Particle size distribution profile data can be generated or an overall figure can be derived to loosely determine the mass of drug that falls into the respirable range. This is achieved by combining several impactor stages with particle size cut-offs roughly

corresponding to respirable particles (74). To gather the particle size distribution data, the MDI is actuated a number of times into the cascade impaction apparatus via a connecting elbow (termed the throat or induction port). On entering the impactor, the aerosol is passed through successive impaction plates with decreasing jet diameters. Velocities increase as the particles go down through the impactor to the point where they impact on the collection plate below. Cascade impactors (particularly the Andersen) have been used extensively to assess MDI aerosols (72). A schematic diagram of the Andersen impactor is given in Fig. 9.

To collect and deliver the MDI discharge to the impactor, two main types of collector system have been used: a large chamber to permit extensive or complete spray droplet evaporation before impaction (72, 75, 76), and a restrictive induction port (51, 73, 78) in which considerable droplet impaction occurs in a very crude analogy to the behavior for the human oropharynx. The extent of droplet evaporation before impaction, and therefore the size distribution measured, is a function of the lag time between spray discharge and impaction. The lag time depends on the size and shape of the particular inlet duct and upper impactor stages and also the volumetric airflow rate. The droplet evaporation rate depends on the volatility of the particular propellant/excipient system (75) and the effect of the airflow rate and air vapor mixing on vapor diffusion from the droplets. Clearly, it is essential in comparing inhaler systems to standardize the induction port, cascade impactor, and airflow rate (72, 77).

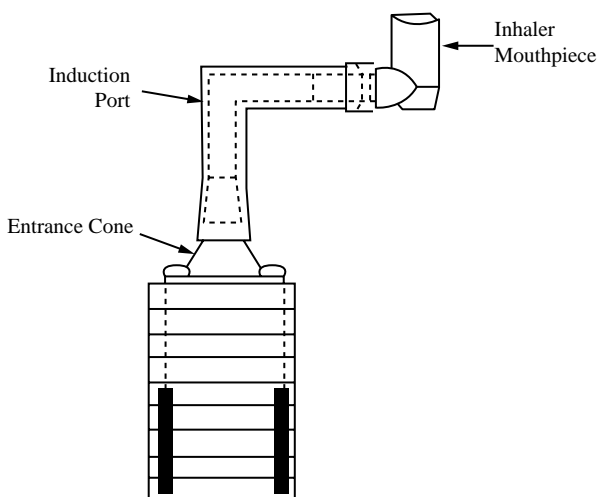


Fig. 9 Schematic representation of the Andersen cascade impactor.

The impacted drug fractions are normally measured via a specific chemical assay, although microweighing the plates has also been investigated (75, 78). A number of actuations are normally accumulated for each cascade impaction determination to collect sufficient drug for accurate assay. The United States Pharmacopoeia includes a test for MDIs, using a cascade impactor coupled to a throat of specified dimensions (74).

Simple two-stage impactors (or impingers) with a defined throat geometry have also been used to generate simple data more rapidly than for the multi-stage cascade impactors (or impingers) or for QC purposes (74, 79–81). The Twin Impinger (TI) has gradually become a much less widely used technique as it cannot provide the more detailed distribution data required for in-house product development or by the regulatory authorities, but it can be sensitive enough to reveal formulations that are unstable due to crystal growth or particle aggregation.

SINGLE PARTICLE OPTICAL SIZERS (SPOS)

Single particle optical sizers measure each aerosol particle passing through a small optical sensing zone to give rapid “real time” measurement of particle number diameter distribution and concentration. Low particle concentrations are necessary to avoid coincidence errors arising from multiple particle occupancy in the sensing zone. SPOS methods are rapid, but they have the following disadvantages for MDI aerosol evaluation:

- Drug containing particles are not distinguished from excipient or foreign particles.
- Aerosol sampling may not be representative due to inadequate measurement of the large particles and their possible partial loss in the sampling probe (82).
- Equivalent mass distribution is derived based on the assumption that the particles are spherical and of equal density.

Optical single-particle counters offer much more rapid aerosol size analysis than do impaction methods. They need to be applied with caution to MDIs for the reasons stated previously.

REFERENCES

1. Bisgaard, H. *Respir. Med.* **1997**, *91* (Suppl A), 20–21.
2. Tansey, I.P. *Pharm. J.* **1997**, *259*, 896–898.
3. Vervaet, C.; Byron, P.R. *Int. J. Pharm.* **1999**, *186*, 13–30.

4. Hallworth, G.W. The Formulation and Evaluation of Pressurised Metered Dose Inhalers. *Drug Delivery to the Respiratory Tract*; Ganderton, D., Jones, T., Eds.; Ellis Horwood: Chichester, UK, 1987; 87–118.
5. Atkins, P.J.; Barker, N.P.; Mathisen, D. The Design and Development of Inhalation Drug Delivery Systems. *Pharmaceutical Inhalation Pressurised Metered-Dose Inhaler Technology*; Hickey, A.J., Ed.; Marcel Dekker, Inc.: New York, 1992.
6. Moren, F. Aerosol Dosage Forms and Formulation. *Aerosols in Medicine—Principles Diagnosis and Therapy*; Moren, F., Newhouse, M.T., Dolvich, M.B., Newman, S.P., Eds.; Elsevier: Amsterdam, 1993; 2.
7. Purewal, T.S.; Grant, D.J.W. *Metered Dose Inhaler Technology*; Interpharm, Press Inc.: Illinois, 1998.
8. IPAC (International Pharmaceutical Aerosol Consortium). *Ensuring Patient Care—The Role of the HFC MDI*, 2nd Ed.; 1999.
9. Molina, M.J.; Rowland, F.S. *Nature* **1974**, *249*, 810.
10. Montreal Protocol on Substances that Deplete the Ozone Layer, Liaison Office of the United Nations Environmental Programme, New York, 1989.
11. Dalby, R.N.; Byron, P.R.; Sherpherd, H.R.; Papadopoulos, E. *Pharm. Technol.* **1990**, 26–33.
12. Dalby, R.N.; Byron, P.R. *Pharm. Technol. Int.* **1992**, *4* (1/2), 42–45.
13. CPMP (Committee for Proprietary Medicinal Products) Result of the Coordinated Review of 1, 1, 1, 2 - Tetrafluoroethane HFC-134a **13 July 1994**.
14. Harnor, K.J.; Perkins, A.C.; Wastie, M.; Wilson, C.G.; Sims, E.E.; Feely, L.C.; Farr, S.J. Effect of Vapor Pressure on the Deposition Pattern from Solution Phase Metered Dose Inhalers. *Int. J. Pharm.* **1993**, *95* (1–3), 111–116.
15. Byron, P.R. Towards the Rational Formulation of Metered Dose Inhalers. *J. Biopharm. Sci.* **1992**, 1–9.
16. Vanbever, R.; Mintzes, J.D.; Wang, J.; Nice, J.; Chen, D.; Batycky, R.; Langer, R.; Edwards, D.A. Formulation and Physical Characterization of Large Porous Particles for Inhalation. *Pharm. Res.* **1999**, *16* (11), 1735–1742.
17. Edwards, D.A.; Hanes, J.; Caponetti, G.; Hrkach, J.; Ben, J.A.; Eskew, M.L.; Mintzes, J.; Deaver, D.; Lotan, N.; Langer, R.S. Large Porous Particles for Pulmonary Drug Delivery. *Science* **1997**, *276* (5320), 1868–1871.
18. Shekunov, B.Yu. Crystallization Processes in Pharmaceutical Technology and Drug Delivery Design. *J. Cryst. Growth* **2000**, *211*, 122–136.
19. Dellamary, L.A.; Tarara, T.E.; Smith, D.J.; Woelk, C.H.; Adrastas, A.; Costello, M.L.; Weers, J.G. Hollow Porous Particles in Metered Dose Inhalers. *Pharm. Res.* **2000**, *17*, 168–174.
20. BenJebria, A.; Eskew, M.L.; Edwards, D.A. Inhalation System for Pulmonary Aerosol Drug Delivery in Rodents Using Large Porous Particles. *Aerosol Sci. Technol.* **2000**, *32*, 421–433.
21. FDA Centre for Drug Evaluation and Research (CDER), DRAFT Guidance for Industry, Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Products. <http://www.fda.gov/cder/guidance/index.htm>.

22. Adams, W.P.; Poochikian, G.; Taylor, A.S.; Patel, R.M.; Burke, G.P.; Williams, R.L. Regulatory Aspects of Modifications to Innovator Bronchodilator Metered Dose Inhalers and Development of Generic Substitutes. *J. Aerosol Med. Deposition, Clearance, and Effects in the Lung* **1994**, *7* (2), 119–134.
23. Ruffin, R.E.; Southcott, A.M. Chlorofluorocarbon-Free Inhalation Devices in Asthma: Are there Clinical Implications?. *Biodrugs* **1998**, *10* (2), 91–96.
24. Brambilla, G.; Ganderton, D.; Garzia, R.; Meakin, B.; Ventura, P. Proceedings of Drug Delivery to the Lungs IX. *Modulation of Aerosol Clouds Produced by HFA Solution Inhalers*; 1998, 155–159.
25. Williams, R.O.; Hu, C. *European J. Pharm. Biopharm.* **1997**, *44*, 195–203.
26. Novartis A.G. European Patent Application No. 94810478.1.
27. Downing R.C.; Parmalee H.M. Proc CSMA 37th Annual Meeting Freon Aerosol Report A19, E.I. Du Pont de Nemours and Co. Ltd. New York 1950.
28. Howlett, D. *Spray Technol. Market.* **1994**, 18–22.
29. Pischiak, A.H. *Proceedings of Drug Delivery to the Lungs X*; 1999, 171–174.
30. Williams, G.; Tcherevatchenkoff, A. *Proceedings of Drug Delivery to the Lungs VIII*; 1997, 91–94.
31. Dunbar, C.A. CA129 –153066/12 Journal 19981231. Part. Sci. Technol. **1997**, *15*, 253–271.
32. Tansey, I.P. The Challenges in the Development of Metered Dose Inhalation Aerosols using Ozone-Friendly Propellants. *Spray Technol. Mark.* **1994**, *7*, 26–29.
33. Weeke, E.R. *Eur. J. Dis.* **1982**, *63* (Suppl. 119), 105–109.
34. Konig, P. *Chest* **1985**, *82*, 276–284.
35. Moren, F. *Int. J. Pharmacol.* **1978**, *1*, 205–212.
36. Corr, D.; Dolovich, M.; McCormack, D.; Ruffin, R.; Obminski, G.; Newhouse, M.J. *Aerosol Sci.* **1982**, *13*, 1–7.
37. Moren, F. *Int. J. Pharmacol.* **1978**, *1*, 213–218.
38. Newman, S.P.; Newhouse, M.T. Effect of Add-On Devices for Aerosol Drug Delivery: Deposition Studies and Clinical Aspects. *J. Aerosol Med: Depositions, Clearance, and Effects in the Lung* **1996**, *9*, 55–70.
39. Zanen, P. Inhalation Anti-asthma Therapy with Spacers: Technical Aspects. *Monaldi Arch. Chest Dis.* **1994**, *49*, 258–264.
40. Barry, P.W.; O'Callaghan, C. Inhalational Drug Delivery from Seven Different Spacer Devices. *Thorax* **1996**, *51*, 835–840.
41. Wildhaber, J.H.; Devadason, S.G.; Eber, E.; Hayden, M.J.; Everard, M.L.; Summers, Q.A.; LeSouef, P.N. Effect of Electrostatic Charge, Flow, Delay and Multiple Actuations on the in Vitro Delivery of Salbutamol from Different Small Volume Spacers for Infants. *Thorax* **1996**, *51*, 985–988.
42. Dewsbury, N.J.; Kenyon, C.J.; Newman, S.P. The Effect of Handling Techniques on Electrostatic Charge on Spacer Devices: A Correlation with in Vitro Particle Size Analysis. *Int. J. Pharm.* **1996**, *137*, 261–264.
43. Crompton, G.K. Problems Patients have using Pressurised Aerosol Inhalers. *Eur. J. Respir. Dis.* **1982**, *63*, 101–104.
44. Newman, S.P.; Weisz, A.W.B.; Talee, N.; Clarke, S.W. Improvement of Drug Delivery with a Breath Actuated Pressurised Aerosol for Patients with Poor Inhaler Technique. *Thorax* **1991**, *46*, 712–716.
45. Polli, G.P.; Grim, W.M.; Bacher, F.A.; Yunker, M.H. *J. Pharm. Sci.* **1969**, *58*, 484–486.
46. Clark, A.R. *Metered Atomisation for Respiratory Drug Delivery*; Ph.D. Thesis Loughborough University of Technology: 1991.
47. Hickey, A.J. *Inhalation Aerosols—Physical and Biological Basis for Therapy, Lung Biology in Health and Disease*; Marcel Dekker, Inc.: 1996; 94.
48. Newman, S. *Deposition and Effects of Inhalation Aerosols*; AB Draco: Lund Sweden, 1983; 13–22.
49. Kim, C.S.; Eldridge, M.A.; Sackner, M.A. *Am. Rev. Respir. Dis.* **1987**, *135*, 157–164.
50. Swift, D.L. *J. Aerosol Sci.* **1992**, *23*, 495–498.
51. Backstrom, K.; Nilsson, P.G. *J. Aerosol Sci.* **1988**, *19*, 1097–1100.
52. Ranucci, J.A.; Cooper, D.; Sethachutkul, K. *Pharmacol. Technol.* **1992**, *4*, 84–92.
53. Miller, J.F.; Schatzel, J.F.; Vincent, B.J. *Colloid Int. Sci.* **1991**, *143*, 532–554.
54. Najafabadi, A.R.; Ganderton, D. *J. Pharm. Pharmacol.* **1991**, *43*(67P).
55. Najafabadi, A.R.; Ganderton, D. *J. Pharm. Pharmacol.* **1990**, *42* (Suppl. 12P).
56. Patent UK GB2236146B
57. Committee for Proprietary Medicinal Products. Replacement of Chlorofluorocarbons (CFCs) in Metered Dose Inhalation Products. 1993, MCA Eurodirect Publication No. 5378/93.
58. Inhalanda, Preparations for Inhalation, European Pharmacopoeia. Supplement: 1999, 671.
59. Contractor, A.M.; Richman, M.D.; Shangraw, R.F. *J. Pharm. Sci.* **1970**, *59*, 1488–1489.
60. British Pharmacopoeia. Her Majesty's Stationery Office: London, 1988; II, 875–876.
61. USP. United States Pharmacopoeia. 23rd Revision U.S. Pharmacopoeial Convention, Inc.: Rockville, MD, 1994; 1839.
62. Miszuk, S.; Gupta, B.M.; Chen, B.C.; Clawans, C.; Knapp, J.Z. *J. Pharm. Sci.* **1980**, *69*, 713–717.
63. Hallworth, G.W.; Kedgley, D. *J. Pharm. Pharmacol.* **1986**, *38* (Suppl), 27.
64. Dhand, R.; Malik, S.K.; Balakrishnan, M.; Verma, S.R. *J. Pharm. Pharmacol.* **1988**, *10*, 429–430.
65. Moren, F.; Hathaway, D. Proc. 3rd International Conference on Liquid Atomization and Spray Systems 1985. 1,111B.1 ICLASS–85: London, 1985; 1–6.
66. Hotham, G.A. Particle Size Analysis of Aerosols in Medicine. *Respiratory Aerosols*; Gale, A.E. Ed.; Gordon Harris: St. Agnes, South Australia, 1981; 47–71.
67. Johnsen, M.A. *The Aerosol Handbook*, 2nd Ed.; Dorland: Wayne NJ, 1982; 271, 273–354.
68. Hind, G. *Manf. Chem.* **1990**, *61*, 23–30.

69. Tsuda, S.; Dohara, K.; Shinjo, G. *Chem. Pharm. Bull.* **1988**, *37*, 2777–2781.
70. Burton G.R. Negus C.R. Proceedings of the Fourth Annual Conference of the Aerosol Society University of Surrey 1909; 211–216
71. Byron, P.R. *Respiratory Drug Delivery*. CRC Press, Inc.: Boca Raton, FL, 1973.
72. Hallworth, G.W.; Andrews, U.G. *J. Pharm. Pharmacol.* **1976**, *28*, 898–907.
73. Hallworth, G.W.; Hamilton, R.R. *J. Pharm. Pharmacol.* **1976**, *28*, 890–897.
74. The Pharmacopeial Convention. 7th Suppl.: Rockville, MD, 1992; 3122–3129, United States Pharmacopeia (USP/NF).
75. Kim, C.S.; Trujillo, D.; Sackner, M.A. *Am. Rev. Respir. Dis.* **1985**, *132*, 137–142.
76. Nilsson, G.; Nrunzell, A.; Helberg, H. *Acta Pharmac. Suec.* **1977**, *14*, 95–104.
77. Phillips, E.M.; Byron, P.B.; Fults, K.; Hickey, A.J. *Pharmacol. Res.* **1990**, *7*, 1228–1233.
78. Malton, C.A.; Hallworth, G.W.; Padfield, J.M. *J. Pharm. Pharmacol.* **1982**, *34* (Suppl), 65.
79. Hallworth, G.W.; Westmoreland, D.G. *J. Pharm. Pharmacol.* **1987**, *39*, 966–972.
80. Atkins, P.J. *Pharmacol. Technol.* **1992**, *4*, 26–32.
81. British Pharmacopoeia. Appendix XV11C Her Majesty's Stationery Office: London, 1988.
82. Bouchikhi, A.; Becquemin, M.H.; Bignon, J.; Roy, M.; Teillac, A. *Eur. Respir. J.* **1988**, *1*, 547–552.
83. CPMP (Committee for Proprietary Medicinal Products) Result of the Coordinated Review of 1, 1, 1, 2, 3, 3, 3 - Heptafluoropropane (HFC-227) **13 Sept 1995**.

Metabolite Identification in Drug Discovery

Kathleen Cox
Nigel Clarke
Diane Rindgen

Schering-Plough Research Institute, Kenilworth, New Jersey, U.S.A.

INTRODUCTION

The challenge in modern drug discovery is not only to rapidly screen through large number of potential lead drug candidates, but also to find out as much information about these compounds as possible. Since a high percentage of drugs fail in the preclinical and clinical phases of development after millions of dollars have been invested, this methodology is economical, and will hopefully move only lead drug candidates with the highest possibility for success forward for development. Many drug metabolism and pharmacokinetic assays have adapted readily to the discovery environment and are used as routine screens.^[1] In vitro methodologies such as Caco2 cell-based assays,^[2] CYP450 enzyme inhibition assays,^[3,4] and microsomal^[5] or hepatocyte incubations have been particularly useful in screening large numbers of compounds for absorption, inhibition of the common metabolizing enzymes, and metabolic stability, respectively. In vivo assays are typically less amenable to high throughput, but have come to the forefront in drug discovery because of the wealth of drug metabolism and pharmacokinetic information that can be obtained from them. Novel dosing procedures such as cassette, or N-in-one dosing,^[6,7] involve dosing large numbers of compounds in a single animal for screening purposes. The drawback of this type of experiment is the risk of drug–drug interactions. Alternatives to the cassette dosing procedure include sample pooling prior to analysis,^[8,9] which achieves the goal of higher throughput without compromising the pharmacokinetic evaluation of the drugs.

While metabolite identification studies provide critical information on drug candidates, these studies have typically been reserved for compounds late in the development phase. These studies are not amenable to high throughput as each compound will give a different metabolic profile, and evaluation of the data can be a lengthy and labor-intensive process. Traditional studies required radio labeled compounds, synthetic standards of potential metabolites, and sophisticated analytical instrumentation.

However, with the recent advances in analytical technology and software programs, metabolite identification studies are now playing a pivotal role in the discovery phase of new drug entities. Early identification of metabolic “hot spots” in a particular structural series provides valuable information to the medicinal chemists and can drive the progression of chemical structures in a particular therapeutic program. In addition, early characterization of potentially active or toxic metabolites can direct a program to more potent and safe recommendation candidates.^[10] This article provides an overview of the value that early metabolite identification studies can give to a discovery program. Analytical techniques, available in the discovery phase, are described for the early characterization of metabolites, focussing on the use of liquid chromatography-tandem mass spectrometry (LC-MS/MS), and the advances in software programs to aid the analyst in critically and rapidly evaluating the data produced. The focus is on small molecule applications.

METABOLISM AND DRUG DESIGN

Many drug candidates fail early in the development phase due to poor pharmacokinetics, such as their half-life being too long or too short, high clearance, or extensive first pass metabolism.^[11] An early evaluation of the metabolic fate of these compounds can be invaluable to the program. Typically, the compounds under investigation in the discovery phase of a specific program are very similar in structure. Medicinal chemists are searching for structure–activity relationships and a final compound is chosen by fine tuning the structure of interest for best selectivity, activity, and overall pharmacokinetic profile. Since structures are similar, it is often possible to determine specific portions of the molecule that are particularly susceptible to metabolic alteration. These structural characteristics can either be tailored into future molecules to shorten a lengthy half-life or tailored out of future molecules to lengthen a short half-life.

Another advantage of evaluating similar compounds in a structural series is the often predictable nature of the metabolic alterations. The analyst can quickly determine the most common metabolic alterations to a specific structure, and understand any novel or unexpected metabolic modifications. A fairly high throughput screening protocol can then be put into place to quickly evaluate all future compounds within the structural series for those specific metabolic alterations.^[12] This is particularly useful in screening for toxic metabolites such as acyl glucuronide or glutathione conjugates. In these cases, it is only necessary to know the mass of the metabolic alteration and the mass spectrometer can then rapidly evaluate samples generated from *in vivo* or *in vitro* studies for those metabolites. The generation of acyl glucuronide metabolites *in vitro* has been useful in evaluating the extent of formation, thus giving an idea of toxic liability.^[13,14] This assay is also amenable to a high throughput format. Such an *in vitro* approach has also been extended to investigate the stability of acyl glucuronides, again providing an idea of reactivity and potential toxicity.^[15] In this manner, early metabolic screening can play a pivotal role in driving the drug design of discovery compounds.

SPECIES DIFFERENCES IN METABOLISM (METABOLIC PROFILING)

All mammals exhibit differences in their biochemical make-up between species and sometimes even sex, particularly in the structures and activities of their cytochrome P450 metabolizing enzymes.^[16] Because of these differences, both the rate of drug metabolism and the metabolic profile may differ between animal species. It is possible in drug discovery to obtain an early look at the human metabolism of drug candidates through *in vitro* systems such as microsomes, hepatocytes,^[17] S9 fractions, and liver slices. It is, therefore, important to determine whether the human *in vitro* systems produce any unique metabolites, not expressed in other animal species. This is due to the critical importance during preclinical development that adequate exposure to all metabolites is achieved in the animal safety studies to ensure the validity of the clinical studies. Although *in vitro* systems are not always accurate predictors of *in vivo* metabolism, particularly in the case of humans, it is useful to see what metabolic pathways can be accessed in a human system and make sure that the metabolites thus produced are also generated in an animal system.

METABOLITE CHARACTERIZATION VS. IDENTIFICATION

Definitive metabolite identification studies require absolute structural identification of each metabolite produced. This is a labor-intensive process requiring separation of all metabolites produced in a particular biological matrix (i.e., bile, urine, plasma), analytical characterization of the structure of the metabolite, and confirmation with a synthetic standard. This methodology involves the evaluation of one metabolite after another in a serial fashion. Often the structure of a metabolite can be elucidated by mass spectrometry using exact mass measurements to obtain an empirical formula and/or obtaining a detailed fragmentation pathway for the metabolite. However, the most definitive technique for structural elucidation is NMR. This technique can not only determine the exact site of modification, but also the stereochemistry of the metabolite structure. The drawback of using NMR is that it typically requires a large amount of the metabolite in order to provide definitive structural data (micrograms) while LC-MS/MS typically requires nanograms or even picograms of material for structural characterization. While this kind of detailed analysis is required once a compound has progressed to the development stage, it is not required in drug discovery. Often, at this stage, it is sufficient to merely know the portion of the molecule that is metabolically labile. Furthermore, this approach does not require the systematic isolation and purification of each metabolite produced and, in fact, can be conducted in more of a parallel manner, evaluating all metabolites from a single sample within one chromatographic separation.

INSTRUMENTATION

High performance liquid chromatography coupled with tandem mass spectrometry (HPLC-MS/MS) is an analytical technique that is ideally suited for metabolite characterization.^[18,19] The chromatographic conditions are sufficient to separate the analyte and metabolites from the background, and the mass spectrometer is a selective and sensitive detector, capable of characterizing low levels of metabolites in complex biological matrices. There are a variety of mass spectrometers available and each possesses unique capabilities and advantages in the characterization of metabolites. The use of several complementary LC-MS/MS systems is the most stringent way to provide complete characterization of metabolites.

The triple quadrupole mass spectrometer is generally considered as the most versatile of all tandem MS instruments and is useful in the initial evaluation of metabolites

for a new compound.^[20–22] Several types of tandem MS experiments can be performed, requiring no previous knowledge of a compound's metabolites. They only require a synthetic standard of the starting drug compound. This compound, once ionized and passed into the collision chamber of the triple quadrupole mass spectrometer, will fragment into pieces upon energized collisions with gas molecules. This fragmentation is characteristic of the structure of the compound, and, as many of the metabolites typically have similar structural features to the parent compound, the metabolites will often dissociate into some of these same structural fragments. The triple quadrupole mass spectrometer is capable of performing a precursor ion scan experiment in which all ions are passed into the collision chamber and are dissociated into fragments. The fragments are then passed into a different chamber of the mass spectrometer and evaluated. A fragment, that is the same as that produced by the original drug compound, can be linked back to its originating ion (i.e., the metabolite). In this manner, the mass of the metabolite can be determined along with some information about the structure. This is a fairly selective scanning process so matrix or background ions will not contain structural features similar to the parent drug and will not be detected. A second, similar scanning technique can search for neutral portions of potential metabolites after dissociation in the collision chamber. This is a highly useful experiment since conjugated metabolites, such as glutathione, glucuronide, and sulfate conjugates, give characteristic neutral losses in the collision chamber. This experiment, called a constant neutral loss scan experiment, is so generic that it does not even require any prior knowledge of the parent drug. The precursor and constant neutral loss experiments can be performed within the LC timeframe required for metabolite identification experiments.^[23] Once a putative metabolite is discovered in these scanning experiments, it can be confirmed and further characterized by a third scanning technique called a product ion scan experiment. A product ion scan allows only the ion of the prospective metabolite to pass into the collision chamber. The metabolite dissociates into fragments characteristic of its structure, some fragments similar to the parent drug compound, and some different. A comparison of the product ion spectrum of the putative metabolite with the parent drug can provide a great deal of information about the metabolic alteration and structure of the metabolite.

Often these types of experiments can point to the portion of the drug that has been metabolically altered, rather than completely identify the structure of the metabolite itself. However, as mentioned previously, this is often enough in the discovery phase to alert the structural chemists to the metabolically labile portion of the

molecule, but sometimes further experiments are necessary to narrow the site and nature of the metabolic alteration.

A second type of MS instrument capable of performing a product ion scanning experiment is the ion trap mass spectrometer. The difference between this instrument and the triple quadrupole mass spectrometer is the ion trap is capable of performing numerous sequential product ion experiments on the same compound. For example, a potential metabolite is dissociated into several characteristic fragment ions in the first stage of MS/MS. Then, one of these fragment ions can be further dissociated into smaller ions. This is a second stage of MS/MS known as MS³. This process can continue (MSⁿ)^[24,25] until the site of metabolic alteration is determined or no more informative fragment ions are formed.

A high mass resolution/high mass accuracy tandem mass spectrometer is also useful in metabolite characterization experiments. These instruments can determine the exact mass of a potential metabolite or of one of its fragment ions. Although several accurate mass instruments are available, the most useful for metabolite ID experiments is a quadrupole time of flight (Q-ToF) instrument. Other instruments, such as a four sector mass spectrometer or a Fourier transform mass spectrometer, are capable of achieving even higher resolution and mass accuracies; however, they are cumbersome and expensive and require a specially trained operator. The types of molecules encountered in metabolite ID experiments are typically comprised of only organic atoms and are fairly low molecular weight (<1000 Da). Ultrahigh resolution and mass accuracy are not necessary for their characterization. In addition, the Q-ToF instruments have been proven to be fairly rugged systems, capable of evaluating low concentrations of metabolites in extremely dirty biological matrices such as bile. An accurate mass instrument is particularly effective when two nominally isobaric structures are possible for a potential metabolite. These are structures that have the same unit mass, but differ in their exact mass.^[26]

There have been great advances in mass spectrometric instrumentation over the past decade. All of the instruments described above are rugged and available for routine use. While these are not the only tandem mass spectrometric systems available, when used together, they provide a powerful complement for metabolite characterization experiments. A flow chart illustrating the complementary nature and effective utilization of these various MS/MS techniques is shown in Fig. 1. The most effective use of these instruments involves setting up the front end of the systems identically. All systems should be fitted with the same LC and injection system and the mass spectrometers should be set up in parallel with a radio flow

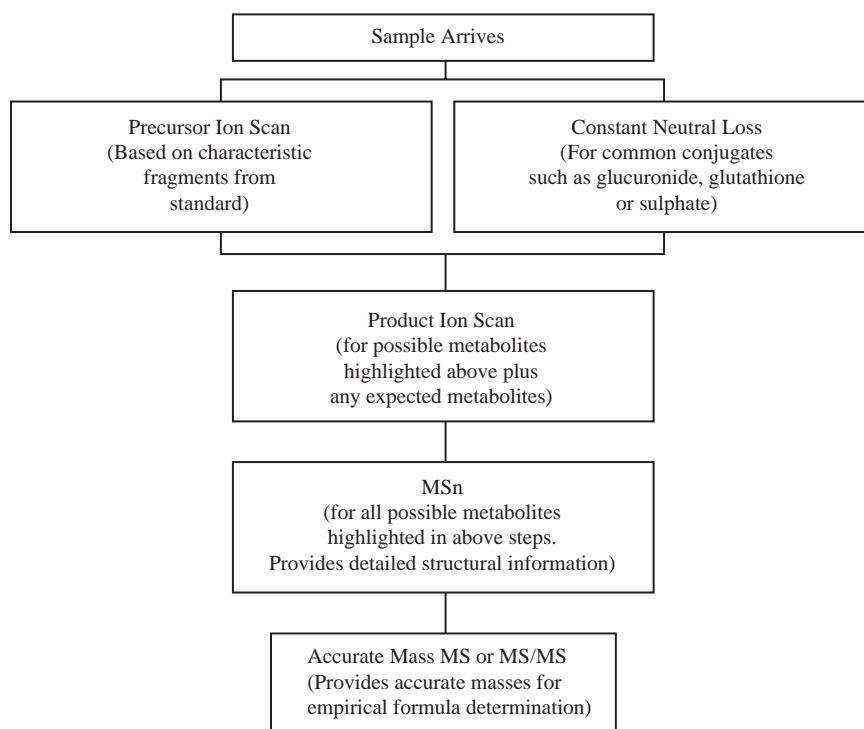


Fig. 1 Systematic approach to metabolite identification and characterization.

detector. In this manner, samples can be easily moved from one system to another for further characterization. If a sample is radioactive, the LC eluent will split after leaving the column and a portion of the sample will go into the mass spectrometer while another portion will pass into the radio flow detector. The systems can also be fitted with an on-line UV detector or other analog detection device. However, in discovery, the samples are typically extremely dirty and the metabolites are present at concentrations well below the matrix concentration, so detectors with some level of selectivity will be the most useful.

The metabolic profile of new drug entities can be evaluated in a variety of biological matrices. The advantage of *in vitro* matrices is that they provide a cleaner background, and the concentrations of metabolites can often be controlled by manipulating the incubation conditions. However, not all of the metabolic pathways are necessarily accessed through an *in vitro* system. The most comprehensive metabolic profile will be obtained from an *in vivo* system and the highest concentration of metabolites will be observed in the excreta (i.e., bile/urine/feces). Also, since these compounds are being evaluated early in discovery, these experiments often provide the first look at the metabolites of a new chemical entity. Due to this, there is a danger that routine sample cleanup methods, such as solid-phase extraction or

liquid–liquid extraction, will eliminate novel metabolites. In addition, as the new chemical entities are rarely radioactive, there is no way to trace whether critical metabolites are lost during cleanup. A more comprehensive evaluation of metabolites is obtained if the samples (bile/urine/plasma) can be analyzed directly. This again points to the need for a rugged, sensitive, and selective detection system.

SOFTWARE

It is easy to envisage how the tandem MS experiments described before could generate an enormous amount of data, leaving the daunting task of evaluating every chromatographic peak manually to the operator. However, significant advances in software technology have been made in recent years to allow a computer to critically evaluate these large datasets based on a series of preset criteria and rules. Not only does this provide a higher throughput for data evaluation, but it also leaves the operator free to design additional experiments, resulting in a more comprehensive evaluation of metabolites in a shorter time frame. As stated earlier, the most comprehensive and informative metabolite identification experiments come from the analysis of *in vivo* samples, typically bile

and urine. Bile, in particular, is an extremely dirty matrix, comprised primarily of numerous bile salts that can obscure the signals generated by potential metabolites. Since little or no sample cleanup is utilized in these samples, highly complex mass chromatograms can result. Computer algorithms now exist that will perform a detailed background subtraction on these samples, virtually eliminating contributions from the matrix. In order to perform this background subtraction, two separate experiments are performed on the LC-MS/MS system. The sample of interest is analyzed and then an injection is made of a control sample of the appropriate biological matrix. The analytical conditions are identical for each experiment, and the computer merely eliminates the peaks observed in the control biological matrix from the actual sample dataset, significantly simplifying the resulting mass chromatogram and in many cases revealing potential metabolites that were obscured by the chemical noise.

A second software feature capitalizes on the fact that many potential drug candidates contain atoms that are not naturally occurring in the body and which possess unique isotopic patterns. In particular, chlorine and bromine

atoms give unique and distinctive mass spectral signals. The computer can evaluate the sample dataset and find any chromatographic peaks that contain these characteristic isotopic patterns. An example of how these software routines can simplify a complex chromatogram is shown in Fig. 2 (reprinted with permission from American Pharmaceutical Review). This compound contained a tritium label, so the top trace (a) shows peaks corresponding to radioactive metabolites. The second trace (b) shows the total ion chromatogram (i.e., all masses from 300 to 1000 amu) with no software intervention. A background subtraction routine greatly simplifies this chromatogram in the third trace (c). Finally, this compound contains a chlorine atom, so an isotopic cluster algorithm further simplifies the resulting chromatogram in the fourth trace (d). Note how this last trace begins to resemble the radio chromatogram.

In addition, the computer can accept a list of potential or common metabolites such as hydroxylation, demethylation, or a combination of metabolic alterations. Most software packages will also accept custom alterations if the user knows of unusual metabolic processes that are specific

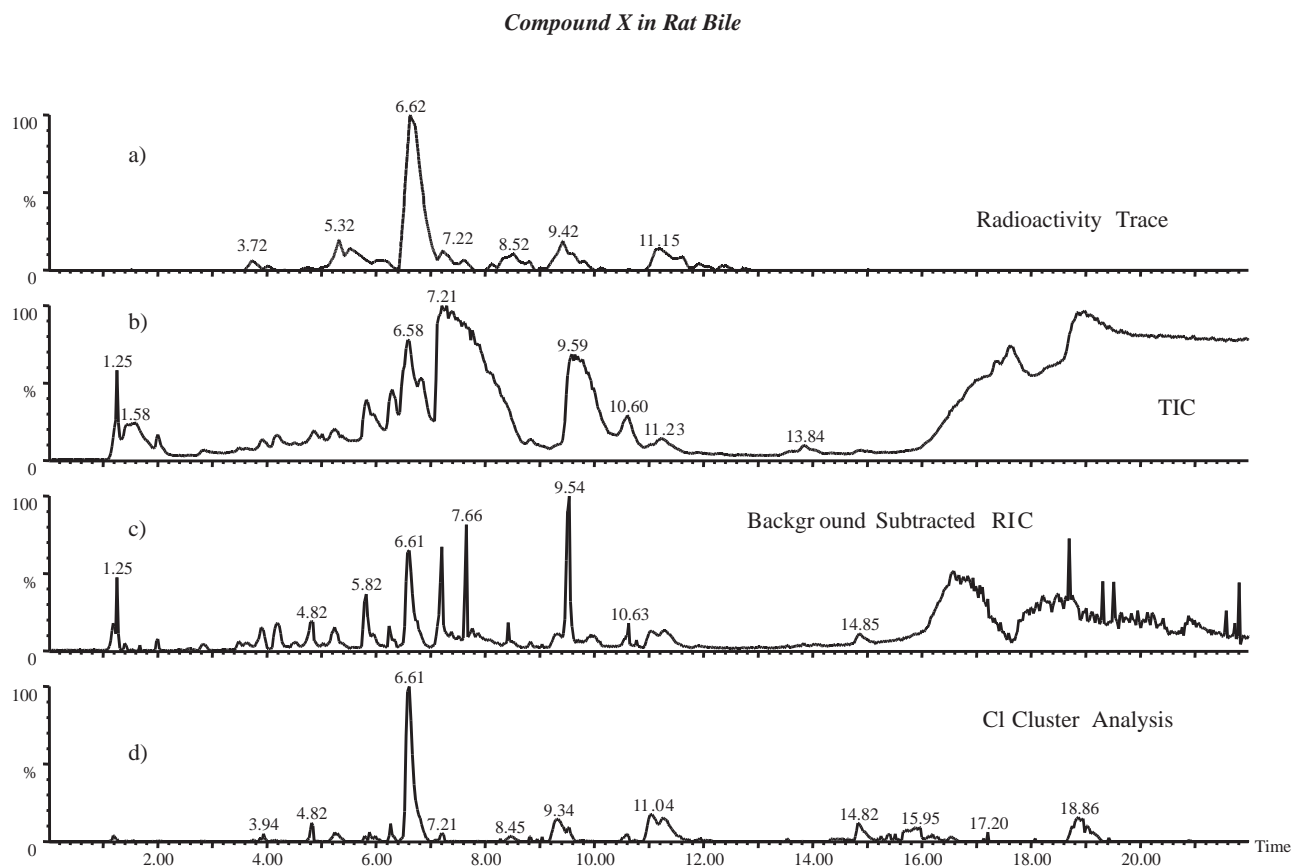


Fig. 2 Various software cleanup methods in the evaluation of metabolites of Compound X in rat bile. (From Ref. 26.)

for a given structural series of compounds. The software programs can calculate these masses, given only the mass of the parent drug, and search the resulting mass chromatograms for the presence of these metabolites. Furthermore, given a set of evaluation criteria, the computer can assess the integrity of the observed peaks to determine whether they represent real entities. Acceptance criteria can be dictated by the user and consist of factors such as how high the analytical signal is above the background noise and the width of the peak. These techniques are particularly powerful when used in conjunction with accurate mass data, as the masses of potential metabolites can be precisely determined and distinguished from other contaminants that may have the same nominal mass, but differ in their exact mass.

Most of the commercial software packages currently available offer a browser screen that allows the operator to view the peaks the computer picks out as potential metabolites and lets the operator have final say in whether these peaks should be evaluated further as potential metabolites. The software programs can then automatically set up additional MS/MS experiments for further characterization of these metabolites, significantly eliminating the amount of user intervention necessary. These experiments can be set up to run automatically overnight, leaving the results available for user evaluation in the morning. This type of data “mining” provides a rapid and efficient way of evaluating the tremendous amount of data generated in a mass chromatogram, significantly reducing the amount of manual data interpretation required by the user and increasing the throughput of metabolite identification experiments. A flow diagram of the metabolite software decision process is shown in Fig. 3 (reprinted with permission from American Pharmaceutical Review).

CONCLUSIONS AND FUTURE TRENDS

The contributions of metabolite identification studies to early drug discovery are often significant. It is of critical importance to a discovery program to know what metabolites are formed as this can drive the structural chemistry effort toward more effective drugs. It also provides an opportunity to obtain an early look at human metabolism from *in vitro* systems to make sure that there are no metabolites produced by human enzymes that are not adequately produced by other animal species. This can result in a higher success rate of drug candidates that make it into the clinical phase. However, only in recent years, with the considerable hardware and software advances in the use of LC-MS/MS technology, it has been realistic to implement these types of studies into the high throughput

environment of drug discovery. It is also important to redefine the goals of a metabolite identification experiment in the discovery environment. Many compounds in a structural series need to be evaluated in a fairly short time frame and often it is more important to point to metabolically labile portions of the molecule quickly rather than to spend months elucidating the exact stereospecific structure of a single metabolite. Furthermore, sample volumes are small and rarely radiolabeled. LC-MS/MS is a powerful analytical technique and can be used effectively to evaluate samples under these conditions. The mass spectrometer is a selective and sensitive detector, and is able to detect and characterize low levels of metabolites out of extremely complex biological matrices. The precursor and constant neutral loss scanning capabilities of an MS/MS instrument provide the means of searching through a complex mixture to find any entities that have structural similarities to the parent drug compound, and often very little prior information is necessary to make use of these experiments. Data from product ion scan experiments provide further structural characterization of potential metabolites and use of MSⁿ and accurate mass capabilities can narrow the site of modification even further and provide insight into the empirical formula of metabolites. The integration of these complementary MS/MS techniques provides an efficient and comprehensive approach to metabolite characterization.

The recent development of “smart” software programs allows the computer to quickly and critically evaluate large, complex datasets. This significantly reduces the amount of manual data interpretation required by the user and large datasets can be evaluated more rapidly. However, data interpretation is still crucial and continues to be the bottleneck to even higher throughput studies. To address this, the software programs are continuing to evolve in response to user needs. Data-dependent scanning capabilities are being developed to make the most use out of a single injection. Once the computer detects a potential metabolite in an LC run, it can automatically switch to a tandem mode to provide structural information on that metabolite “on the fly.” This type of experiment shows a great deal of potential; however, it is still problematic for extremely dirty samples such as bile.

Future software programs will also be able to utilize the input from other external analog detectors such as radio flow or UV detectors to further enhance the resulting dataset. In addition, there are software programs that are capable of interpreting MS/MS spectra and provide suggestions for the structures of the observed fragment ions. Mathematical programs such as CODA can further distinguish what is a real chemical signal and what is random noise in a dataset and correlation analysis can

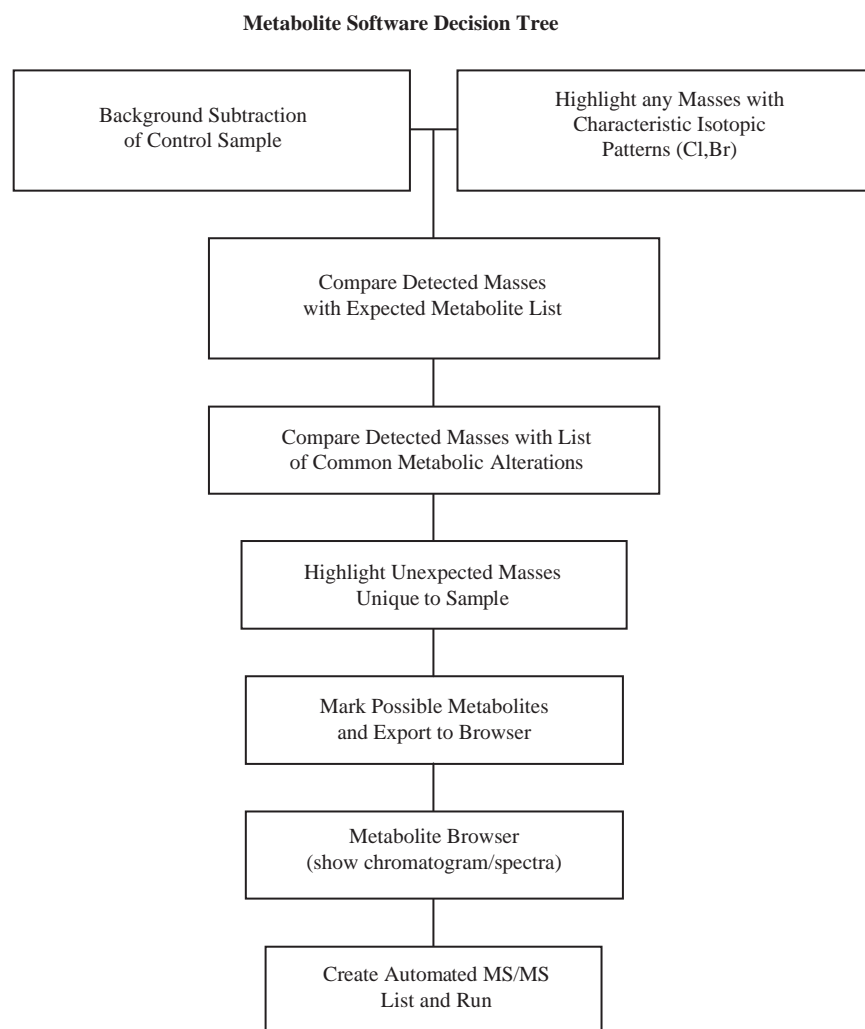


Fig. 3 Metabolite software decision tree. (From Ref. 26.)

evaluate complex matrices, pick out potential metabolites, and determine the degree of similarity between these observed compounds and the parent drug compound.^[27] Efforts are also being made to utilize the Internet and establish links to existing metabolite databases. Once a metabolite is discovered experimentally, the software will search the database to determine whether that type of metabolic alteration has been observed before and under what circumstances.

Implementation of metabolite identification studies into drug discovery is still very much an evolving process. Significant technological advances have been made over recent years and scientists are continuing to adapt and develop existing techniques to the rigorous demands of a discovery environment. While metabolite ID is by no means “high throughput,” experiments exist that can give timely and valuable information back to the discovery

project teams, driving the evolution of lead compounds and answering critical metabolic questions that will ultimately save the pharmaceutical company time, money, and resources in the future. Based on the value of these studies, the contributions from metabolite ID studies within drug discovery are only expected to increase in the future.

REFERENCES

1. White, R.E. High-throughput screening in drug metabolism and pharmacokinetic support of drug discovery. *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 133–157.
2. Caldwell, G.W.; Easlick, S.M.; Gunnet, J.; Masucci, J.A.; Demarest, K. In vitro permeability of eight beta blockers through Caco-2 monolayers utilizing liquid chromatography/electrospray ionization mass spectrometry. *J. Mass Spectrom.* **1998**, *33*, 607–614.

3. Palamanda, J.R.; Favreau, L.; Lin, C.C.; Nomeir, A.A. Validation of a rapid microtiter plate assay to conduct cytochrome P450 2D6 enzyme inhibition studies. *Drug Discov. Today* **1998**, *3*, 466–470.
4. Chu, I.; Favreau, L.; Soares, T.; Lin, C.C.; Nomeir, A.A. Validation of higher-throughput high-performance liquid chromatography/atmospheric pressure chemical ionization tandem mass spectrometry assays to conduct cytochrome P450s CYP2D6 and CYP3A4 enzyme inhibition studies in human liver microsomes. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 207–214.
5. Korfmacher, W.A.; Palmer, C.A.; Nardo, C.; Dunn-Meynell, D.; Grotz, D.; Cox, K.A.; Lin, C.C.; Elicone, C.; Liu, C.; Ducholsav, E. Development of an automated mass spectrometry system for the quantitative analysis of liver microsomal incubation samples: A tool for rapid screening of new compounds for metabolic stability. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 901–907.
6. Shaffer, J.E.; Adkison, K.K.; Halm, K.; Hedeon, K.; Berman, J. Use of “N-in-one” dosing to create an in vivo pharmacokinetic database for use in developing structure–pharmacokinetic relationships. *J. Pharmaceut. Sci.* **1999**, *88*, 313–318.
7. Olah, T.V.; McLoughlin, D.A.; Gilbert, J.D. The simultaneous determination of mixtures of drug candidates by liquid chromatography/atmospheric pressure chemical ionization mass spectrometry as an in vivo drug screening procedure. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 17–23.
8. Cox, K.A.; Dunn-Meynell, K.; Korfmacher, W.A.; Broske, L.; Nomeir, A.A.; Lin, C.C.; Cayen, M.; Barr, W.H. Novel in vivo procedure for rapid pharmacokinetic screening of discovery compounds in rats. *Drug Discov. Today* **1999**, *4*, 232–237.
9. Hop, C.E.; Wang, Z.; Chen, Q.; Kwei, G. Plasma-pooling methods to increase throughput for in vivo pharmacokinetic screening. *J. Pharm. Sci.* **1998**, *3*, 901–903.
10. Lin, J.H.; Lu, A.Y.H. Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol. Rev.* **1997**, *49*, 403–449.
11. Prentis, R.A.; Lis, Y.; Walker, S.R. Pharmaceutical innovation by seven UK-owned pharmaceutical companies (1964–1985). *Br. J. Clin. Pharmacol.* **1988**, *25*, 387–396.
12. Clarke, N.J.; Rindgen, D.; Korfmacher, W.A.; Cox, K.A. Systematic LC/MS metabolite identification in drug discovery. *Anal. Chem.* **2001**, *73*, 431A–439A.
13. Grotz, D.E.; Clarke, N.J.; Cox, K.A.; Korfmacher, W.A. *An In Vitro Method for the Assessment of Acyl Glucuronide Metabolite Formation Early in Drug Discovery Using HPLC-MS/MS*; Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, IL, May 27–31, 2001.
14. Rindgen, D.; Grotz, D.; Clarke, N.J.; Cox, K.A. The application of HPLC/tandem mass spectrometry for the assessment of acyl glucuronide formation in in vitro and in vivo systems in a drug discovery environment. *Am. Pharm. Rev.* **2001**, *4*, 52–58.
15. Hop, C.E.C.A.; Rabe, M.; Wang, Z.; Pang, J.; Leung, K.H.; Franklin, R.B. *Formation and Reactivity of Acyl Glucuronides Assessed by LC-MS/MS*; Proceedings of the 48th ASMS Conference on Mass Spectrometry and Allied Topics, Long Beach, CA, 2000.
16. Lindberg, L.P.; Negishi, M. Alteration of mouse cytochrome P450c₁ substrate specificity by mutation of a single amino acid residue. *Nature (Lond.)* **1989**, *339*, 632–634.
17. Stevens, J.C.; Fayer, J.L.; Cassidy, K.C. Characterization of 2-[[4-[[2-(1H-tetrazol-5-ylmethyl)phenoxy]methoxy]-methyl]quinoline *N*-glucuronidation by in vitro and in vivo approaches. *Drug Met. Dispos.* **2001**, *29*, 289–295.
18. Lee, M.S.; Kerns, E.H. LC/MS Applications in drug development. *Mass Spectrom. Rev.* **1999**, *18*, 187–279.
19. Korfmacher, W.A.; Cox, K.A.; Bryant, M.S.; Veals, J.; Ng, K.; Watkins, R.; Lin, C.C. HPLC-API/MS/MS: a powerful tool for integrating drug metabolism into the drug discovery process. *Drug Discov. Today* **1997**, *2*, 532–537.
20. Perchalski, R.J.; Yost, R.A.; Wilder, B.J. Structural elucidation of drug metabolites by triple quadrupole mass spectrometry. *Anal. Chem.* **1982**, *54*, 1466–1471.
21. Betowski, L.D.; Korfmacher, W.A.; Lay, J.O., Jr.; Potter, D.W.; Hinson, J.A. Direct analysis of rat bile for acetaminophen and two of its conjugated metabolites via thermospray liquid chromatography/mass spectrometry. *Biomed. Environ. Mass Spectrom.* **1987**, *14*, 705–709.
22. Kerns, E.H.; Rourick, R.A.; Volk, K.J.; Lee, M.S. Buspirone metabolite structure profile using a standard liquid chromatography-mass spectrometric protocol. *J. Chromatogr. B* **1997**, *698*, 133–145.
23. Busch, K.L.; Glish, G.L.; McLuckey, S.A. *Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry*; VCH Publishers Inc.: New York, 1988.
24. Stafford, G.C., Jr.; Kelley, P.E.; Syka, J.E.P.; Reynolds, W.E.; Todd, J.F.J. Recent improvements in and analytical applications of advanced ion-trap technology. *Int. J. Mass Spectrom. Ion Proc.* **1984**, *60*, 85–98.
25. Louris, J.N.; Cooks, R.G.; Syka, J.E.P.; Kelley, P.E.; Stafford, G.C., Jr.; Todd, J.F.J. Instrumentation, applications and energy deposition in quadrupole ion trap MS/MS spectrometry. *Anal. Chem.* **1987**, *59*, 1677–1685.
26. Cox, K.A.; Clarke, N.J.; Rindgen, D.; Korfmacher, W.A. Higher throughput metabolite identification in drug discovery: Current capabilities and future trends. *Am. Pharm. Rev.* **2001**, *4*, 45–52.
27. Fernandez-Metzler, C.L.; Owens, K.G.; Baillie, T.A.; King, R.C. Rapid liquid chromatography with tandem mass spectrometry-based screening procedures for studies on the biotransformation of drug candidates. *Drug Metab. Dispos.* **1999**, *27*, 32–40.

Melt Processes for Oral Solid Dosage Forms

Paul W.S. Heng

Tin Wui Wong

National University of Singapore, Singapore

INTRODUCTION

The design of pharmaceutical solid dosage form using meltable materials has become increasingly popular as alternative special processes to avoid aqueous or organic solvents. Its application has extended to the processing of solid dispersion, agglomerate, microsphere, and nanosphere for the administration of drugs via oral and parenteral routes.

The processing and drug release properties can be partly or wholly affected by the physicochemical characteristics of meltable materials. The judicious choice of the meltable material to be used has considerable influences on the property and quality of the resulting solid dosage form.

OVERVIEW

The manufacture of solid dosage forms using meltable materials has many advantages when compared with the conventional wet techniques. The avoidance of using aqueous solvent made the melt technique suitable for formulating moisture-sensitive drugs. The use of flammable organic solvents for processing effervescent and hygroscopic materials can be avoided, thus reducing the needed cost for flame-proof facilities and solvent-recovery equipment. Materials processed by the melt technique do not require solvent removal or a drying phase. The total processing time is shorter. By an appropriate selection of meltable materials, both immediate- and prolonged-release solid dosage forms can be prepared. Hence the release rate of a drug can be controlled by varying the composition of the meltable materials.

The most obvious disadvantage of the melt technique is that the process cannot be applied to heat-sensitive materials owing to the elevated temperatures involved. Nonetheless, *Lactobacillus acidophilus* bacteria, formulated by continuous melt technology, were found to have a higher survival rate than those prepared by conventional wet techniques, although the molten mass had a temperature of more than 100°C.^[1] Moisture often aggravates the thermolability of pharmaceutical materials. The absence of moisture minimizes the heat-related degradation.

Thus it can be concluded that the melt technique for processing heat-sensitive materials warrants further studies.

The melt processes require at least a meltable material in the manufacture of the solid dosage form. This article intends to address the influence of meltable materials on the processes of melt pelletization, melt granulation, and tumbling melt granulation. Approaches to meet the processing attributes of the meltable materials for these melt agglomeration processes, via spray congealing, are discussed.

GENERAL INFORMATION

The processors for both melt and wet techniques share similar features.^[2-7] Material processing by the melt technique essentially requires the heating jacket, thermal insulator, high-speed/high-shear agitation, and/or atomization equipment. Melt processing is accompanied by either a size enlargement or a reduction of the particles. Materials employed may be of single or multiple meltable components, as well as a mixture of meltable and non-meltable components. In the latter mixture, the weight proportion of nonmeltable component may vary from 10% to 90% depending on the design of the solid dosage form. The end products are usually granules or pellets produced by a one-step, one-run process.

In design and development of a formulation by melt processes, multiple-unit solid dosage form is popular with manufacturers as it allows a high degree of flexibility in modifying the dosage regimen. Multiple-unit solid dosage form can be designed with different proportions to meet the desired dose strength without altering the drug load in each entity by processing and/or formulation changes. Entities of different drugs that are not chemically compatible or containing the same drug but with different release rates may be blended in an appropriate ratio. Multiple-unit solid dosage form may effect drug release at the same site or at different sites along the route of administration and is less susceptible to dose dumping.

By the size enlargement processes, melt agglomeration may be employed to produce multiple-unit solid dosage form. The melt agglomerates are called granules when the end product has a wide size distribution spanning between 0.1 and 2 mm. The agglomerates are termed pellets if the



end product is spherical in shape and has a narrow size distribution of 0.5–2 mm. Melt pellets are often dispensed as a multiparticulate dosage form, whereas melt granules may be encapsulated or further subjected to tableting processes. The formative and growth processes of melt agglomerates are critically dependent on the physico-chemical properties of the meltable materials, which may be modified by spray congealing. Unlike melt agglomeration, which requires both meltable and nonmeltable components, spray congealing is able to process meltable materials with or without the nonmeltable component.

Melt Pelletization/Melt Granulation

Melt agglomeration is a process by which the solid fine particles are bound together into agglomerates, by agitation, kneading, and layering, in the presence of a molten binding liquid. Dry agglomerates are obtained as the molten binding liquid solidifies by cooling. Typical examples of melt agglomeration processes are melt pelletization and melt granulation. During the agglomeration process, a gradual change in the size and shape of the agglomerates would take place. It is usually not possible to clearly distinguish between granulation and pelletization. Thus granulation is considered a pelletization process when highly spherical agglomerates of narrow size distribution are produced. Conversely, an unsuccessful pelletization process may be classified as granulation.

The equipment for melt agglomeration include rotating drums or pans, fluid-bed granulators, low-shear mixers such as Z-blade and planetary mixers, and high-shear mixers. Presently, the more popular agglomeration equipment for industrial-scale production are high-shear mixers and fluid-bed granulators. In both methods, a gradual buildup of agglomerates occurs during the process. The marked difference between the methods is the absence of shearing forces in the fluid-bed process, whereas very high shearing forces are generated in high-shear mixing.

During a melt agglomeration process, the meltable binder may be added as molten liquid, or as dry powder or flakes. In the latter, the binder may be heated by hot air or by a heating jacket to above the melting point of the binder. Alternatively, the melt agglomeration process exploits an extremely high shear input, of a high-shear mixer, where the heat of friction alone raises the temperature of the binder and effects melting. Polyethylene glycols, fatty acids, fatty alcohols, waxes, and glycerides are some examples of meltable binders. Polyethylene glycols are commonly used as meltable binders when aqueous soluble binders are needed, while fatty acids, fatty alcohols, waxes, and glycerides are desirable meltable binders for products to be insoluble in the aqueous medium. Typically, the melting points of meltable binders range from 50 to 80°C. A lower-melting-point binder

risks situations where melting or softening of the binder occurs during handling and storage of the agglomerates. Higher-melting-point binders require high melting temperatures and can contribute to instability problems especially for heat-labile materials.

In assessing the influence of meltable materials on the formative and growth processes of melt agglomerates, it is imperative to have a thorough understanding of the melt agglomeration process. The mechanism of melt agglomeration is similar to that of wet agglomeration. An agglomerate growth is affected by the saturation state of the liquid in the agglomerate pore structure. However, the formation and growth of melt agglomerates are not complicated by binding liquid losses via evaporation during the agglomeration process. Agglomerative mechanisms can be subdivided as nucleation, coalescence, and layering (Fig. 1). These mechanisms are believed to constitute a complex set of elementary events that directly or indirectly influence the formative and growth processes of agglomerates. Nucleation is predominant with fine particles. It involves primary solid particles being bound together, on addition of a binding liquid, to form a three-phase air–solid–liquid loose cluster, which eventually lead to the formation of a primary nuclei. The particles are held by liquid bridges in a pendular state (Fig. 2). The formation of larger agglomerates followed successful random internuclei collisions by a process known as coalescence. Coalescence involves nuclei that have residual surface liquid to promote successful fusion of nuclei. The surface liquid imparts plasticity to the nuclei and is essential for enabling the deformation of nuclei surfaces for coalescence as well as promoting the rounding of agglomerates. Excess surface liquid may be increased by adding a binding liquid or by densification as a result of mechanical forces where liquid is squeezed out from the interior to the surfaces of agglomerates (Fig. 2). During coalescence, the liquid-bridging state of the nuclei changes from pendular through funicular to capillary state. Fine particles may be generated by random collisions during the size-enlargement processes. These

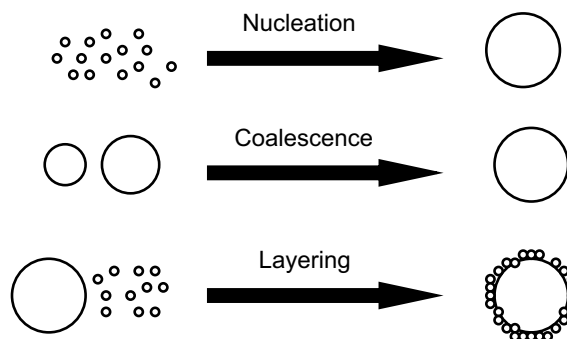


Fig. 1 Mechanisms of agglomeration.

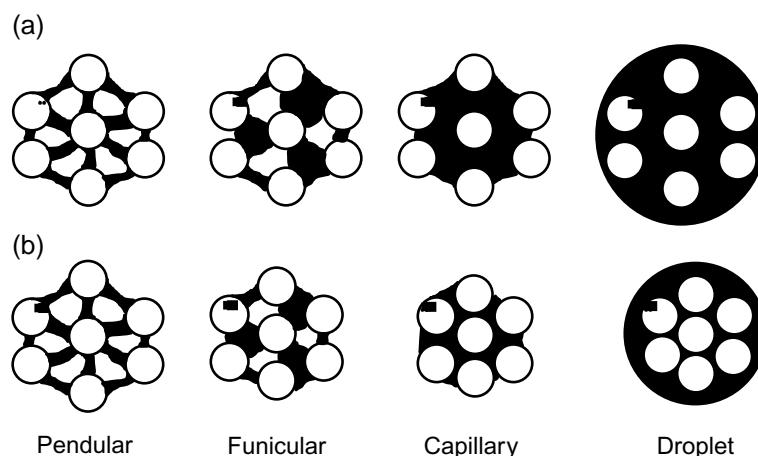


Fig. 2 Liquid-bridging state of agglomerates undergoing (a) binding liquid addition and (b) densification.

particles could be layered onto the preformed nuclei. Alternatively, fine particles may be supplemented by addition, either as a dry powder or molten mass, onto the preformed nuclei that are many orders of magnitude larger than the added material, to effect size enlargement by a layering process.

Two modes of melt agglomeration based on the elementary mechanisms have been proposed—distribution and immersion.^[8] In agglomeration by the distribution mode, a distribution of molten binding liquid on the surfaces of primary particles will occur, and agglomerates are formed via coalescence between the wetted nuclei (Fig. 3). In agglomeration by the immersion mode, nuclei are formed by immersion of the primary particles into the surface of a droplet of molten binding liquid (Fig. 3). The distribution of molten binding liquid to surfaces of nuclei has to be effected by densification prior to coalescence between the nuclei. Depending on the

relative size between the solid particles and the molten binding liquid droplets, the distribution will be a dominant mode when the molten binding liquid droplets are smaller than the solid particles or of a similar size. On the other hand, the immersion mode will dominate when the molten binding liquid droplets are larger than the solid particles. The distribution mode is promoted by a low molten binding liquid viscosity. In the case of immersion, it is more favorable for molten binding liquid of high viscosity, which could resist breakup by dispersive forces.

Melt agglomeration by the immersion mode produced slower initial agglomerate growth rate than that by the distribution method. At the initial phase of melt agglomeration, the surface plasticity of agglomerates bound by large molten binding liquid droplets of a high viscosity is lower. This translates to a reduced growth propensity of melt agglomerates by coalescence. However, with a greater extent of densification by prolonging the period

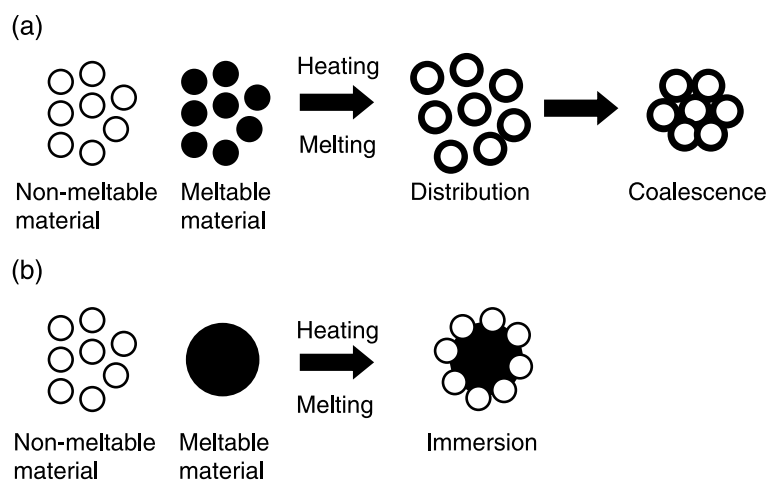


Fig. 3 Modes of melt agglomeration: (a) Distribution and (b) immersion.



of processing or by using high impeller speed, the agglomerates attain a comparable surface wetness and deformability. As the surfaces of these agglomerates are wetted by a molten binding liquid of higher viscous or tack forces, the growth propensity of agglomerates becomes increasingly greater and, at times, can be uncontrollable. Meltable binders are available as flakes, powders, or fine powders. Practically, the melt agglomeration process is less dependent on the particle size of a meltable binder except when a very high viscosity binder is used.^[8–10]

During melt agglomeration, the liquid saturation state has to be near the droplet state for highly viscous molten binding liquid to be sufficiently deformable for agglomerate growth by coalescence. There is a high risk of overwetting and uncontrollable agglomerate growth. A higher impeller speed may be applied to prevent an uncontrollable agglomerate growth by ensuring continuous comminution.^[11] Unlike the wet agglomeration process, the melt agglomeration process cannot be controlled through the removal of excess surface liquid by evaporation. However, evaporation of water of crystallization from the solid hydrate during a melt agglomeration process could render the growth process of agglomerates more controllable, probably by reducing the surface wetness through absorption of binding liquid into small pores formed in the agglomerate surfaces as a result of the loss of water of crystallization.^[12]

It is envisaged that melt agglomeration employing highly viscous molten binding liquid can also be controlled by using an impeller blade of appropriate geometry, which generates high relative swept volume or shear input,^[13] large solid particles as nonmeltable components that form agglomerates of low mechanical strength susceptible to comminution^[14] as well as keeping the off-bottom clearance, which critically affects the flow dynamics of the processing materials within a narrow range.^[15] Indirectly, melt agglomeration may be subjected to real-time monitoring of the process energy consumption, which is an indication of the work load for agglomeration.^[16,17]

Table 1 presents a list of meltable binders previously employed in melt agglomeration. These meltable binders vary in their melting range, hydrophilicity–hydrophobicity, and molecular size. The molten binding liquid viscosity has largely been identified as the core characteristic of meltable binders that has a strong bearing on the melt agglomeration process. In hydrophobic meltable binders, such as stearic acid, which has very low viscosity values below 50 mPa sec at the melting temperatures ranging from 70 to 90°C, the melt agglomeration process is more markedly dependent on the balance between the coalescence and the breakage of agglomerates.^[18] By using meltable binders with very low viscosities, it is

Table 1 Types of meltable binders in the melt agglomeration process

	Typical melting range (°C)
<i>Hydrophilic meltable binder</i>	
Gelucire 50/13	35–44
Poloxamer 188	~ 50.9
Polyethylene glycol	
2000	42–53
3000	48–63
6000	49–63
8000	54–63
10000	57–64
20000	53–66
Stearate 6000 WL1644	46–58
<i>Hydrophobic meltable binder</i>	
Beeswax	56–60
Carnauba wax	75–83
Cetyl palmitate	47–50
Glyceryl behenate	67–75
Glyceryl monostearate	47–63
Glyceryl palmitostearate	48–57
Glyceryl stearate	54–63
Hydrogenated castor oil	62–86
Microcrystalline wax	58–72
Paraffin wax	47–65
Stearic acid	46–69
Stearic alcohol	56–60

highly probable that crushing and layering are the dominant growth mechanisms occurring concurrently with coalescence. The melt agglomerates are more susceptible to breakage as a result of the low viscous strength. The formed agglomerates, compared to that prepared using higher-viscosity meltable binders, are usually smaller with wider size distribution, more porous, and less spherical in shape.

Tumbling Melt Granulation

A newer melt agglomeration technique, i.e., tumbling melt granulation, for preparing spherical beads has been reported.^[19–21] A powdered mixture of meltable and nonmeltable materials is fed onto the seeds in a fluid-bed granulator (Fig. 4). The mixture adheres onto the seeds with the binding forces of a melting solid to form the spherical beads. In preparing the spherical beads, both viscosity and particle size of the meltable materials should be kept at an optimum value. The particle size of a meltable material should be 1/6 or lower than the diameter of the seeds. High-viscosity meltable materials should not be employed to avoid agglomeration of seeds and producing beads of low sphericity.

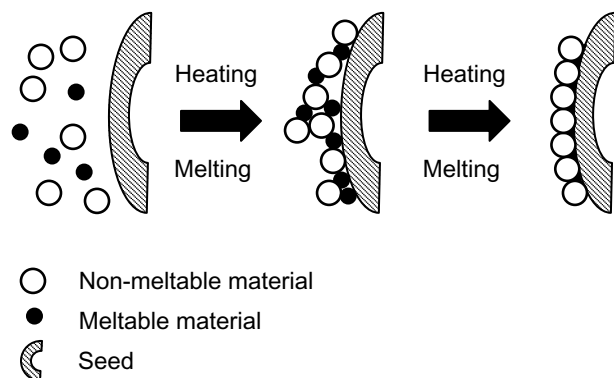


Fig. 4 Process of tumbling melt granulation.

Both particle size and viscosity of the meltable materials play a significant role in the melt agglomeration process. The control of the melt agglomeration process is best initiated by using meltable materials of controlled properties. For the melt pelletization and melt granulation processes, it is desirable that meltable materials have a high viscosity to improve the mechanical strength of the agglomerates, but a reduced particle size to prevent uncontrollable agglomerate growth. In tumbling melt granulation, small meltable particles with sufficient viscous binding forces are obligatory for the production of spherical beads.

Spray Congealing

Spray congealing is a melt technique of high versatility. In addition to manufacture multiparticulate delivery system,^[22,23] it can be applied to process the raw meltable materials into particles of defined size and viscosity values for the melt agglomeration process. Processing of meltable materials by spray congealing involves spraying a hot melt of wax, fatty acid, or glyceride into an air chamber below the melting point of the meltable materials or at cryogenic temperature. Spray-congealed particles (10–3000 μm in diameter) are obtained upon cooling. The congealed particles are strong and nonporous as there is an absence of solvent evaporation. Ideally, the meltable materials should have defined melting points or narrow melting ranges. Viscosity modifier, either meltable or non-meltable at the processing temperature, may be incorporated into the meltable matrix to change the consistency of the molten droplets.

CONCLUSION

The viscosity and particle size of a meltable material critically affect the formative process of the melt agglomerates. Thus their size, size distribution, shape, and

mechanical strength are also affected. This leads to an increasing interest to design meltable materials of required properties suitable for special melt processes. In the design and production of meltable materials with desired particle size and viscosity grade, spray congealing is a potential technique to be employed and developed for industrial application. Spray congealing technique is expected to gain widespread use with recent growing interest in melt pellets as controlled-release solid dosage form.

REFERENCES

1. Appelgren, C.; Eskilson, C. A novel method for the granulation and coating of pharmacologically active substances. *Drug Dev. Ind. Pharm.* **1990**, *16* (15), 2345–2351.
2. Chambliss, W.G. Conventional and Specialized Coating Pans. In *Pharmaceutical Pelletization Technology*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; Vol. 37, 15–38.
3. Olsen, K.W. Fluid Bed Equipment. In *Pharmaceutical Pelletization Technology*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; Vol. 37, 39–69.
4. Goodhart, F.W. Centrifugal Equipment. In *Pharmaceutical Pelletization Technology*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; Vol. 37, 101–122.
5. Jones, D.M. Solution and Suspension Layering. In *Pharmaceutical Pelletization Technology*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; Vol. 37, 145–164.
6. Goodhart, F.W.; Jan, S. Dry Powder Layering. In *Pharmaceutical Pelletization Technology*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; Vol. 37, 165–185.
7. Killeen, M.J. Spray Drying and Spray Congealing of Pharmaceuticals. In *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1996; Vol. 14, 207–221.
8. Schaefer, T.; Mathiesen, C. Melt pelletization in a high shear mixer. IX. Effects of binder particle size. *Int. J. Pharm.* **1996**, *139*, 139–148.
9. Wong, T.W.; Wan, L.S.C.; Heng, P.W.S. Effects of physical properties of PEG 6000 on pellets produced by melt pelletization. *Pharm. Dev. Technol.* **1999**, *4* (3), 449–456.
10. Schaefer, T.; Holm, P.; Kristensen, H.G. Melt granulation in a laboratory scale high shear mixer. *Drug Dev. Ind. Pharm.* **1990**, *16* (8), 1249–1277.
11. Schaefer, T.; Mathiesen, C. Melt pelletization in a high shear mixer. VIII. Effects of binder viscosity. *Int. J. Pharm.* **1996**, *139*, 125–138.
12. Schaefer, T.; Mathiesen, C. Melt pelletization in a high shear mixer. VII. Effects of product temperature. *Int. J. Pharm.* **1996**, *134*, 105–117.
13. Schaefer, T.; Taagegaard, B.; Thomsen, L.J.; Kristensen, H.G. Melt pelletization in a high shear mixer. V. Effects

- of apparatus variables. *Eur. J. Pharm. Sci.* **1993**, *1*, 133–141.
14. Johansen, A.; Schæfer, T. Effects of interactions between powder particle size and binder viscosity on agglomerate growth mechanisms in a high shear mixer. *Eur. J. Pharm. Sci.* **2001**, *12*, 297–309.
 15. Heng, P.W.S.; Wan, L.S.C.; Wong, T.W. Effects of off-bottom clearance on properties of pellets produced by melt pelletization. *Pharm. Dev. Technol.* **1999**, *4* (1), 27–33.
 16. Heng, P.W.S.; Wong, T.W.; Shu, J.J.; Wan, L.S.C. A new method for the control of size of pellets in the melt pelletization process with a high shear mixer. *Chem. Pharm. Bull.* **1999**, *47* (5), 633–638.
 17. Heng, P.W.S.; Wong, T.W.; Chan, L.W. Influence of production variables on the sphericity of melt pellets. *Chem. Pharm. Bull.* **2000**, *48* (3), 420–424.
 18. Eliassen, H.; Kristensen, H.G.; Schæfer, T. Growth mechanisms in melt agglomeration with a low viscosity binder. *Int. J. Pharm.* **1999**, *186*, 149–159.
 19. Maejima, T.; Osawa, T.; Nakajima, K.; Kobayashi, M. Preparation of spherical beads without any use of solvents by a novel tumbling melt granulation (TMG) method. *Chem. Pharm. Bull.* **1997**, *45* (3), 518–524.
 20. Maejima, T.; Osawa, T.; Nakajima, K.; Kobayashi, M. Application of tumbling melt granulation (TMG) method for preparing controlled release beads coated with hydrogenated castor oil. *Chem. Pharm. Bull.* **1997**, *45* (5), 904–910.
 21. Maejima, T.; Osawa, T.; Nakajima, K.; Kobayashi, M. Effects of species of non-meltable and meltable materials and their physical properties on granulatability in tumbling melt granulation method. *Chem. Pharm. Bull.* **1997**, *45* (11), 1833–1839.
 22. Schwendeman, S.P.; Tobfo, M.; Joworowicz, M.; Alonso, M.J.; Langer, R. New strategies for the microencapsulation of tetanus vaccine. *J. Microencapsul.* **1998**, *15* (3), 299–318.
 23. Passerini, N.; Perissutti, B.; Moneghini, M.; Voinovich, D.; Albertini, B.; Cavallari, C.; Rodriguez, L. Characterization of carbamazepine-gelucire 50/13 microparticles prepared by a spray-congealing process using ultrasounds. *J. Pharm. Sci.* **2002**, *91* (3), 699–707.

MUCOADHESIVE HYDROGELS IN DRUG DELIVERY

Hans E. Junginger

Maya Thanou

J. Coos Verhoef

Leiden/Amsterdam Center for Drug Research, Leiden University, The Netherlands

INTRODUCTION

This article focuses on defining the principles of bioadhesive delivery systems based on hydrogels to biological surfaces that are covered by mucus. An overview of the last decade's discoveries on mucoadhesion and applications of mucoadhesive hydrogels as drug carriers is given. Techniques that are frequently used to study the adhesion forces and physicochemical interactions between hydrogel, mucus, and the underlying mucosa are reviewed. Typical examples of applications of mucoadhesive hydrogels to mucosal routes of delivery are given. Finally, the perspectives of the application of these polymers in drug delivery are discussed.

Noninvasive drug delivery may require the administration of the drug delivery system (DDS) at an epithelium as a suitable site of absorption of the active compound. Such regions are usually called mucosae. In the human body several mucosal sites can be identified, the one mostly used for administration and absorption of therapeutics being the gastrointestinal route. In order to increase the residence time at these absorption sites, a so-called mucoadhesive delivery system has to be used. Generally, these systems consist of one or more types of hydrogels.

By definition, mucoadhesive hydrogels are a class of polymeric biomaterials that exhibit the basic characteristic of a hydrogel to swell by absorbing water and interacting by means of adhesion with the mucus that covers epithelia.

Bioadhesion has been defined as the attachment of synthetic or biological macromolecules to a biological tissue (1). The term mucoadhesion refers to the special case of bioadhesion where the biological tissue is an epithelium covered by mucus. Mucus is a thin blanket covering all epithelia that are in contact with the external environment in the gastrointestinal, respiratory, and urogenital tracts. The function of mucus is mainly the protection and lubrication of the underlying epithelium, but it may have additional functions dependent on the type of the covered epithelia. In each case of these mucosal routes, mucus characteristics (i.e., thickness) and functions are different. By this definition, the mucosal routes for drug delivery are:

- Buccal/oral route
- Nasal route
- Ocular route
- Vaginal route
- Gastrointestinal route

The concept of mucoadhesion in drug delivery was introduced in the field of controlled-release drug delivery systems in the early 1980s (2, 3). Thereafter, several researchers have focused on the investigations of the interfacial phenomena of mucoadhesive hydrogels (and of other type mucoadhesive compounds) with the mucus. Several techniques of studying these interactions were evaluated both in vitro and in vivo. These techniques have been recently reviewed by Harding et al. (4) and are given in Table 1.

A mucoadhesive hydrogel used as a drug delivery system should

1. Be loaded substantially by the active compound;
2. Not interact physicochemically with the active compound or create a hostile artificial environment that would lead to inactivation and degradation of the active compound;
3. Swell in the aqueous biological environment of the delivery-absorption site;
4. Interact with mucus or its components for adequate adhesion;
5. Allow, when swelled, controlled release of the active compound;
6. Be biocompatible (biomaterial) with the underlying epithelia by means of complete absence of cytotoxicity, ciliotoxicity, or other type of irreversible alterations of the cell membrane components;
7. Have the appropriate molecular size and conformation in order to escape systemic absorption from the administration site; and
8. Be excreted unaltered or biologically degraded to inactive, nontoxic oligomers/monomers that will be further subject of physical clearance.

Additionally, a mucoadhesive delivery system designed for controlled release of active compounds should be localized at specific sites of administration and absorption,

Table 1 Methods of studying mucoadhesion

Method	Comment
Direct assays	
Tensiometry	Force required to dislodge two surfaces, one coated with mucus, the other solid dosage form consisting of mucoadhesive hydrogel
Flow through	Flow rate dV/dt required to dislodge two surfaces; useful for microparticulate dosage forms
Colloidal gold staining	Measures the “adhesion number”
In vivo techniques	Endoscopy, gamma scintigraphy
Molecular mucin-based assays	
Viscometry and rheology	Intrinsic viscosity $[\eta]$ can be related to complex size via MHKS, ^a α coefficient
Dynamic light scattering	Diffusion coefficient, D , can be related to complex size via MHKS c. coefficient
Turbidity, light scattering	SEC MALLS ^b particularly useful for determining MW of mucin, turbidity, semiquantitative indicator
Analytical ultracentrifugation	Change in MW (sedimentation equilibrium), sedimentation coefficient ratio of complex to mucin
Surface plasmon resonance imaging methods	Needs mobile and immobile phase, atomic force microscopy (conventional and gold labeled), scanning tunneling microscopy

^aMark-Houwink-Kuhn-Sakurada.^bSize exclusion chromatography multiangle laser light scattering.

(Adapted from Ref. 4.)

and should prolong the residence time of the active compound at the site of administration to permit, if possible for one daily dosing.

The mucoadhesive delivery system, designed for the administration of macromolecular therapeutics like peptide or protein drugs, should have permeation-enhancing properties by means of alteration of the permeability properties of the underlying epithelium, and should protect the peptide drug from degradation by inhibiting the proteolytic enzymes usually present at the site of administration or by stabilizing the intrinsic environment of the delivery system by sustaining the suitable pH.

Solid dosage forms based on mucoadhesive polymers are used mainly for buccal delivery of drugs, whereas micro- or nanoparticulate formulations are preferred for the delivery of therapeutics in the nasal and intestinal tract (5).

During the last decade research was particularly focused on the delivery of mucoadhesive dosage forms in the gastrointestinal tract, which is the most favorable route of delivery with regard to patient compliance and ease of application. For this reason microparticulate formulations, consisting of mucoadhesive hydrogels, were designed and evaluated by different techniques (6, 7). However, solid dosage forms are more suitable for smaller cavities in the human body, like the oral cavity, either for systemic absorption of compounds or for local treatment of inflammatory diseases. In this case, monolithic devices

(tablets) made from mucoadhesive hydrogels were evaluated (8). Ocular and vaginal applications (for local or systemic absorption) of mucoadhesive hydrogels were also investigated.

The mucoadhesive properties of several classes of hydrogels have been identified, and two types of polymers have attracted special attention. Polyacrylates and their cross-linked modifications represent the anionic type, chitosan and its derivatives the cationic group. In addition, both types of polymers show a number of interesting characteristics beneficial for the administration of a wide range of therapeutics.

Specific type of mucoadhesive compounds, like lectins, have been evaluated to solve the difficulties presented by conventional mucoadhesive hydrogel systems, for instance in the gastrointestinal route. Since these compounds do not belong to the class of hydrogels, they are not extensively discussed here.

MUCUS

In higher organisms epithelia are covered by a protective gel layer defined as mucus. By weight, mucus consists mainly of water (95–99.5%) in which the mucous glycoprotein mucin (0.5–5%) is dispersed. Mucins are

the major components responsible for the gel-like structure of the mucus. They possess a linear protein heavily glycosylated by oligosaccharide side chains. This protein core consists of a single polypeptide chain. One in every three residues is L-serine or L-threonine, in which the O-3 atoms provide the sites for glycosidic linkage. A mucus glycoprotein is composed typically of about 80% of carbohydrates, which for humans are restricted to five monosaccharides:

- L-Fucose
- *N*-Acetylglucosamine
- D-Galactose
- *N*-Acetylgalactosamine
- Sialic acid

The sialic acid residues are usually in a terminal position on the carbohydrate chain whereas the ester sulfate residues are in a more internal position; both contribute to give the molecule a net negative charge. The molecular weights of mucus glycoproteins range from 0.5 to 16×10^6 Da (4).

Mucins can be divided into two classes, membrane bound and secretory forms. Membrane-bound mucins are attached to cell surfaces and may affect immune responses or inflammation. It has been suggested that the high expression of cell-surface mucins, such as sialomucin MUC1, may result in cell–cell and cell–matrix interactions (10). Secretory mucins emanate from mucosal absorptive cells and specialized goblet cells. They constitute the major component of mucous gels in the gastrointestinal, ocular, respiratory, and urogenital epithelia. This type of mucus gel layer is functioning mainly as a physical barrier and lubricant. Important constituents in the mucus include growth factors and trefoil peptides, both secreted by the specialized cells located next to ulcerated mucosal tissue. This appears to be an adaptive phenomenon, important in maintaining the barrier function of the mucosal tissue, enhancing cell migration, and healing after injury. Other substances present in mucus include secretory immunoglobulin A (IgA), lysozyme, lactoferrin, α_1 -Antitrypsin, salts, and *N*-Glycosylated glycoproteins (10).

At present, nine different human epithelial mucin genes have been identified, each of which contains distinct sequences repeated in tandem that encode (apo)mucin core polypeptides (MUC) (10).

Leung and Robinson (11) defined four characteristics of the mucus layer related to mucoadhesion:

1. Mucus is a network of linear, flexible, and random-coil macromolecules.
2. Mucin is negatively charged due to sialic acid and sulfate residues.

3. Mucus is a cross-linked network connected by disulfide bonds between mucin molecules.
4. Mucin is heavily hydrated.

Several techniques have been used to estimate the rate of mucus secretion, but their accuracy seems to be doubtful. Nevertheless, it has been concluded that a slow baseline secretion of mucus is maintained by exocytosis from goblet cells in the gastrointestinal tract, which appears to be under cholinergic control (12). Rubinstein and Tirosh used carbachol (cholinergic agonist) at different doses to increase the mucus thickness in different parts of the gastrointestinal tract (13).

Attempts were also made to estimate the rate of turnover of the mucus gel. Lehr et al. (14) measured the amount of mucus produced per time unit, using an in situ perfused intestinal loop model in the rat. They found that this turnover time varies between 0.8 and 4.5 h.

MECHANISM AND THEORIES OF MUCOADHESION

The mechanistic processes involved in mucoadhesion between hydrogels and mucosa can be described in three steps:

1. Wetting and swelling of the polymer to allow for intimate contact with the biological tissue.
2. Interpenetration of the bioadhesive polymer chains and entanglement of polymer and mucin chains, and
3. Formation of weak chemical bonds between entangled chains (Fig. 1).

Several theories have been proposed to explain the biomucoadhesive phenomena (1, 15–17).

The electronic theory is based on the assumption that the mucoadhesive hydrogel and the target biological tissue have different electronic structures. When two materials come into contact with each other, electron transfer occurs, causing the formation of a double layer of electrical charge at the bioadhesive–biological interface. The bioadhesive force is believed to be due to attractive forces across this electrical double layer.

The adsorption theory states that the bioadhesive bond formed between an adhesive substrate and tissue or mucosae is due to van der Waals interactions, hydrogen bonds, and related forces. Alternatively, when mucus or saliva are interacting with a solid dosage form, the molecules of the liquid are adsorbed on the solid surface. This is an exothermic process. The free energy of adsorption is given by

$$\Delta G_{AD} = \Delta H_{AD} - T \Delta S_{AD} \quad (1)$$

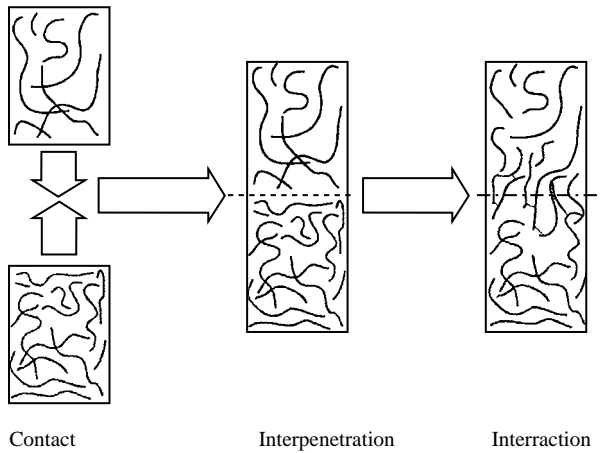


Fig. 1 Three stages in the interaction between a mucoadhesive polymer and mucin glycoprotein according to the interpenetration theory.

where ΔH_{AD} and ΔS_{AD} are the enthalpy and entropy changes, respectively. When adsorption takes place spontaneously, ΔG_{AD} is negative.

The contact angle θ of a liquid on a solid is the reflection of its wetting power. If $\theta = 0$, the liquid spreads freely on the solid surface and wets it. The relationship between the contact angle of a liquid on a solid and the surface tensions in the presence of saturated vapor of the liquid is given by the Young equation

$$\gamma_{sv} = \gamma_{sl} + \gamma_{lv} \cos \theta \quad (2)$$

where γ_{sv} is the solid/vapor surface tension, γ_{sl} the solid/liquid surface tension, and γ_{lv} the liquid/vapor surface tension.

Another way to relate the interfacial tension γ_{sl} to the individual surface tensions of a liquid and solid is given by the Good equation.

$$\gamma_{sl} = \gamma_{sv} + \gamma_{lv} - 2\phi(\gamma_{sv}\gamma_{lv})^{1/2} \quad (3)$$

Eq. 3 represents the reduction in interfacial tension resulting from molecular attraction between liquid and solid. The term ϕ is defined by

$$\phi = W_a/(W_{cl}W_{cs})^{1/2} \quad (4)$$

where W_{cl} and W_{cs} are the work of cohesion of the liquid and the solid, respectively, and W_a is the work of adhesion.

The diffusion theory states that interpenetration and entanglement of polymer chains are additionally responsible for bioadhesion. The intimate contact of the two substrates is essential for diffusion to occur; that is, the

driving force for the interdiffusion is the concentration gradient across the interface. The penetration of polymer chains into the mucus network, and vice versa, is dependent on concentration gradients and diffusion coefficients. It is believed that for an effective adhesion bond the interpenetration of the polymer chain should be in the range of 0.2–0.5 μm . It is possible to estimate the penetration depth (l) by

$$l = (tD_b)^{1/2} \quad (5)$$

where t is the time of contact and D_b is the diffusion coefficient of the bioadhesive material in the mucus.

The fracture theory is the most widely applied theory in studying mucoadhesion mechanisms. It accounts for the forces required to separate two surfaces after adhesion. The maximum tensile stress (σ) produced during detachment can be determined by Eq. 6 by dividing the maximum force of detachment F_m by the total surface area (A_0) involved in the adhesive interaction:

$$\sigma = F_m/A_0 \quad (6)$$

According to Duchêne and Ponchel (17), when tensiometry is used to measure the maximum detachment force as a function of the displacement of the upper support (function of the joint elongation), the work of bioadhesion can be defined as

$$W_b = F \times l \quad (7)$$

Additionally, the fracture energy for a zero extension rate can be defined as the bioadhesion work to the initial surface between the bioadhesive material (in the form of a tablet or disk) and the biological support of a surface A_0 , which allows for calculation of the fracture energy (ϵ) using

$$\epsilon = W_b/A_0 \quad (8)$$

Thermodynamically, the fracture energy is the sum of the reversible work (W_r , representing the reversible molecular interactions at the interface) and irreversible work (W_i representing the irreversible deformation of the interfacial joint); both W_r and W_i are expressed per unit area of the fracture energy by.

$$\epsilon = W_r + W_i \quad (9)$$

When a low extension rate is used for the measurements (for instance 1 mm/min), W_i can be considered as negligible and the fracture energy for zero extension rate (ϵ_0) is then equal to the reversible work of bioadhesion W_r , as shown by

$$\epsilon_0 = W_r = W_b/A_0 \quad (10)$$

The fracture theory does not take into account biological phenomena such as stress caused, for example, by movement of the tissue.

THE STUDY OF MUCOADHESION OF HYDROGELS

During the last decade several methods to study mucoadhesion phenomena or mucoadhesive properties of hydrogels were used (Table 1). Tensiometry has already been reported as a suitable method during late 1980s and is still the most frequently used technique to study mucoadhesiveness of hydrogels. Peppas and coworkers developed a tensile technique for measurements of the bioadhesive strength of tablets containing polyacrylic acid to bovine mucosae (16, 18). A tensiometry setup for the investigation of solid devices is shown in Fig. 2. The system is better mimicking the *in vivo* conditions when the experiment is performed in an aqueous environment and the mucosa of interest is originating from freshly prepared biological tissue. It has been suggested that this system is most suitable for buccal or vaginal application, where biological liquid is controllable. Tensiometry has been shown to be very useful in comparing the mucoadhesive properties of different hydrogels (19).

Mikos and Peppas (20) described the flow channel to measure the bioadhesion of polymer microparticles on mucin gels. Later Lehr et al. (21) used an *in situ* loop model in the rat for the investigation of mucoadhesive microspheres (Fig. 3). They concluded that this approach allowed the study of the transit of particles. Another technique to study the mucoadhesive properties of microspheres is the electrobalance method, as described by Chickering et al. (22, 23). Environmental conditions

like temperature and pH can be easily controlled and several parameters can be obtained from one single experiment. The authors suggested that mucoadhesion in the bioerodible materials used to conduct the studies is not attributable to chain entanglement but to hydrogen bonding between hydrophilic functional groups (COOH) and mucus glycoproteins.

Colloidal gold staining has been introduced by Park (24) for studying the "adhesion number." The polymer in a form of strips is incubated with colloidal gold–mucin conjugates, and after a rinsing procedure the absorbance of strips is measured. Colloidal gold staining has also been used to investigate mucin–chitosan interactions (25).

Rheology measurements were used by Mortazavi et al. (26, 27) to investigate mucus–Carbopol 934P interactions at different pH values and the role of water movement in mucoadhesion. Similar rheological techniques were applied for studying four different types of polymers, and it was concluded that molecular interpenetration is an important factor in mucoadhesion by strengthening the mucus in the mucoadhesive–mucosal interface (28). Rheological methods were used to investigate the interpenetration between ion-sensitive polymers (Carbopol and deacetylated gellan gum, Gerlite) with two commercially available mucins: submaxillary gland mucin and porcine gastric mucin (29). It was suggested that the increase in the elastic modulus of a polymer–mucin mixture (compared to the elastic modulus of polymer alone) indicates a positive interaction caused by mucoadhesion. However, the concentration of the polymer, the type of the mucin used, and the quantity of ions present appeared to have a strong influence on the interactions between mucins and ion-sensitive polymers, indicating that the explanation of mucoadhesion by means of interpenetration should be applied only to interpret special cases of mucoadhesion between polymers and mucins.

Rossi et al. (30) evaluated rheologically mucins of different origin with polyacrylic acid and sodium carboxymethyl cellulose. The same group also reported a novel rheological approach based on a stationary viscoelastic test (creep test) to describe the interaction between mucoadhesive polymers and mucins (31, 32). Jabbari et al. (33) used attenuated total-reflection infrared spectroscopy to investigate the chain interpenetration of polyacrylic acid in the mucin interface.

Other techniques used for studying molecular interactions between polymers and mucus include ultracentrifugation, surface plasmon resonance, and electromagnetic transduction (4, 34). Illum and coworkers (35) investigated the interaction of chitosan microspheres, using turbidimetric measurements and adsorption studies of mucin to the microspheres.

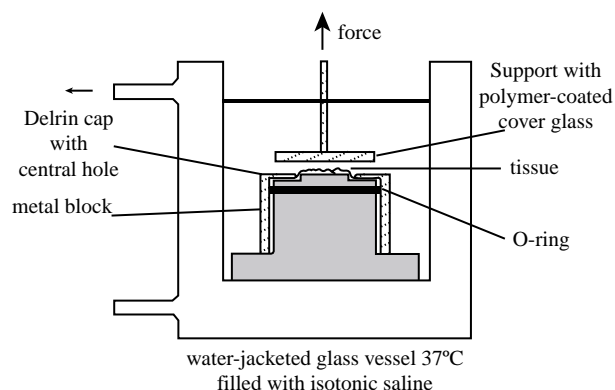


Fig. 2 Experimental setup to measure the force of detachment of mucoadhesive polymer films from mucosal tissue.

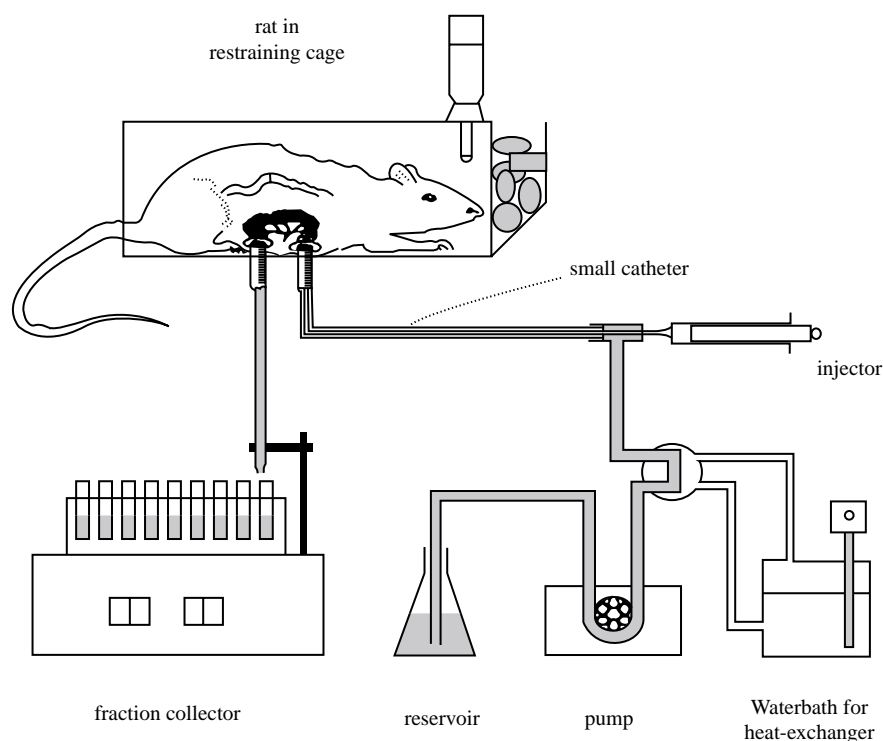


Fig. 3 Experimental setup to study the intestinal transit of mucoadhesive microspheres in a chronically isolated loop. (Adapted from Ref. 21.)

Microscopic visualization techniques have also been used to investigate mucus–polymer interactions (36–39). Transmission electron microscopy was used by Fiebrig et al. (36), whereas different microscopical techniques were used by Lehr et al. (37) for the visualization of mucoadhesive interfaces. Transmission electron microscopy in combination with near-field Fourier transform infrared microscopy (FTIR) has been shown to be suitable for investigating the adhesion-promoting effect of polyethyleneglycol added in a hydrogel (38). Moreover, scanning force microscopy may be a valuable approach to obtain information on mucoadhesion and specific adhesion phenomena (39).

THE MUCOSAL ROUTES OF DELIVERY

Buccal Route

The buccal route of drug administration is the most widely used method for application of mucoadhesive delivery systems. Both for local treatment of inflammation (i.e., aphthae) and for rapid absorption of compounds (nitroglycerin), formulation technologies have employed the buccal route for over two decades, and sublingual or gingival

dosage forms are already established in the market (8) (Table 2).

Oral mucosae are composed of multiple layers of cells, which show various patterns of differentiation dependent on the functions of different regions in the oral cavity (40). The oral mucosa is covered by a stratified, squamous epithelium, and three different types of mucosa can be distinguished: the masticatory, the lining, and the specialized mucosa. Blood supply to the oral cavity tissues is delivered via the external carotid artery, which branches to the maxillary lingual and facial artery. There are no mucus-secreting goblet cells in the oral mucosa, but mucins are found in human saliva. These mucins are water-soluble and form a gel of 10–200- μm thickness. Saliva, mainly composed of water (99%), is continuously secreted in the oral cavity and exists as a film with a thickness of 0.07–0.1 mm (40).

In the 1980s, Machida and Nagai (41) evaluated spray dosage forms based on hydroxypropyl cellulose (HPC) for the delivery of beclomethasone to treat recurrent and multiple aphthae. Previously a double-layered tablet of HPC and Carbopol 934P was introduced in the market for the treatment of aphthous stomatitis (41).

Bouckaert et al. (42) tested buccal tablets of miconazole based on modified starch–polyacrylic acid

Table 2 Drug products available for buccal and/or sublingual application using mucoadhesive polymers

Drug	Therapeutic area	Product names
Nitroglycerin	Angina pectoris	Suscard, Cardilate, Nitrobid, Nitromex, Nitrong
Isosorbide mononitrate	Angina pectoris	Imdur, Isordil, ISMO
Buprenorphine	Analgesia	Temgesic, Buprenex
Nicotine	Smoking cessation	Nicotinelle, Nicorette
Ergotamine	Migraine	Ergostat, Ergomar
Methyl testosterone	Hypogonadism, delayed puberty	Oroton Methyl, Testred, Virilon
Lorazepam	Anxiety, insomnia	Ativan

(Adapted from Ref. 8.)

mixtures. Although these tablets showed different mucoadhesion properties in vitro, no significant differences in the salivary content of miconazole could be observed in human volunteers.

Lee and Chien (43) evaluated mucoadhesive devices of a bilayer type, consisting of a fast-release layer containing polyvinylpyrrolidone (PVP) and a sustained-release layer of Carbopol 934P and PVP for prolonged delivery of luteinizing hormone-releasing hormone (LHRH) onto the porcine gingival and alveolar mucosa for 24 h. This device contained also an absorption enhancer (sodium cholate) and cetylpyridinium chloride to protect LHRH from degradation by microflora. The LHRH permeation appeared to increase by raising the loading of LHRH or enhancer in the fast-release layer. The formulation of the devices could be varied to achieve specific rates of transmucosal peptide drug permeation.

Nair and Chien (44) compared patches and tablets of different polymers (sodium carboxymethylcellulose, carbopol, polyethylene oxide, polymethyl vinyl ether–maleic anhydride, tragacanth) regarding their release characteristics of four drugs (chlorheximide, clotrimazole, benzocaine, and hydrocortisone). They observed sustained release of all four compounds from the mucoadhesive tablets, but only two of the active compounds, chlorheximide and clotrimazole, could be released in a controlled manner from the mucoadhesive patches.

Buccal bilayer devices (films and tablets) are comprised of a drug-containing mucoadhesive layer and a drug-free backing layer (45). The former consists of chitosan, free or cross-linked by an anionic polymer (polycarbophil, sodium alginate, gellan gum), and the latter of ethylcellulose. The in situ cross-linking of chitosan by polycarbophil gives tablets that exhibit controlled swelling, drug release, and adequate mucoadhesion to bovine sublingual mucosa.

The periodontal pocket is another site for drug delivery in the oral cavity. Needleman et al. (46) investigated three mucoadhesive polymers (cationic chitosan, anionic

xanthan gum, neutral polyethylene oxide) in vitro, using organ cultures, and in vivo in patients on their periodontal and oral mucosa. Of the polymers studied, chitosan displayed the longest adhesion in vitro and on the periodontal pockets, and the shortest adhesion on oral mucosa.

Nasal Route

The nasal route of drug administration is the most suitable alternative of delivery for poorly absorbable compounds such as peptide or protein drugs. The nasal epithelium exhibits relatively high permeability, and only two cell layers separate the nasal lumen from the dense blood-vessel network in the lamina propria. The respiratory epithelium is the major lining of the human nasal cavity and is essential in the clearance of mucus by the mucociliary system. This epithelium is composed of ciliated and nonciliated columnar cells, goblet cells, and basal cells. The respiratory epithelium is covered by a mucus layer, which can be divided into two distinctive layers (the periciliary layer and a more gel-like upper layer). The periciliary layer consists of a liquid of lower viscosity. Mucus is secreted from goblet cells as highly condensed granules by exocytosis. The mucus layer is propelled by the cilia toward the nasopharynx, and the function of the mucociliary clearance is to remove foreign substances and particles from the nasal cavity, preventing them to reach the upper airways (47).

Various structurally different mucoadhesive polymers were tested for their ability to retard the nasal mucociliary clearance in rats (48). Methylcellulose, sodium carboxymethyl cellulose, hydroxypropyl methylcellulose, chitosan glutamate, Carbopol 934P, polyethylene oxide 600K, and Pluronic F127 were applied in gel form, and their clearance was measured using microspheres labeled with a fluorescent marker incorporated into the formulation. The clearance rate of each polymer gel was found to be lower than that of a control microsphere suspension, resulting in

an increased residence time of the gel formulations in the nasal cavity. Methylcellulose (3%) gel gave the longest nasal residence time, whereas a Carbopol 934P (0.2%) aqueous gel was the least effective.

Illum et al. (49) evaluated chitosan solutions as delivery platforms for nasal administration of insulin to rats and sheep. They reported a concentration-dependent absorption-enhancing effect with minimal histological changes of the nasal mucosa in all concentrations applied.

Oechslein et al. (50) studied various powder formulations of mucoadhesive polymers for their efficacy to increase the nasal absorption of octreotide in rats. Although chitosan showed the highest water uptake (chitosan > microcrystalline cellulose > semicrystalline cellulose \gg pectin = hydroxyethyl starch = alginic acid = Sephadex G25), the highest peptide drug bioavailability was found after coadministration of alginic acid and Sephadex G25 powders (4.1 and 5.56%, respectively). The authors concluded that the calcium-binding properties of the polymers used correlated better with the increased octreotide bioavailability.

Nakamura et al. (51) studied the adhesion of water-soluble and neutral polymers, hydroxypropyl cellulose (HPC), xanthan gum (XG), tamarind gum (TG), and polyvinyl alcohol (PVA) to nasal mucosa in vitro and in vivo. The polymers, mixed with a dye, were applied as powders to the nasal cavity of rabbits, and the remaining dye residue was determined at 2, 4, and 6 h after nasal instillation with a thin fiberscope. The polymer XG showed the longest residence time of the dye in the cavity, followed by TG, HPC, and PVA in decreasing order. For the mixture XG and XG-PVA (2:8), some residue of dye could still be observed 6 h after administration. The order of adhesion of these polymers to agar plates in vitro agreed with that of their mucoadhesion in vivo. Illum et al. (52) introduced bioadhesive microspheres for nasal delivery of poorly absorbable drugs. Radiolabelled microspheres made from diethylaminoethyl (DEAE)-dextran, starch microspheres, and albumin microspheres were administered to human volunteers and appeared to be cleared significantly slower than solutions or nonmucoadhesive powder formulations. However, starch or hyaluronic acid microspheres significantly increased the absorption of peptide drugs from nasal mucosa (53).

Nakamura et al. (54) described a microparticulate dosage form of budesonide, consisting of novel bioadhesive and pH-dependent graft copolymers of polymethacrylic acid and polyethylene glycol, resulting in elevated and constant plasma levels of budesonide for 8 h after nasal administration in rabbits.

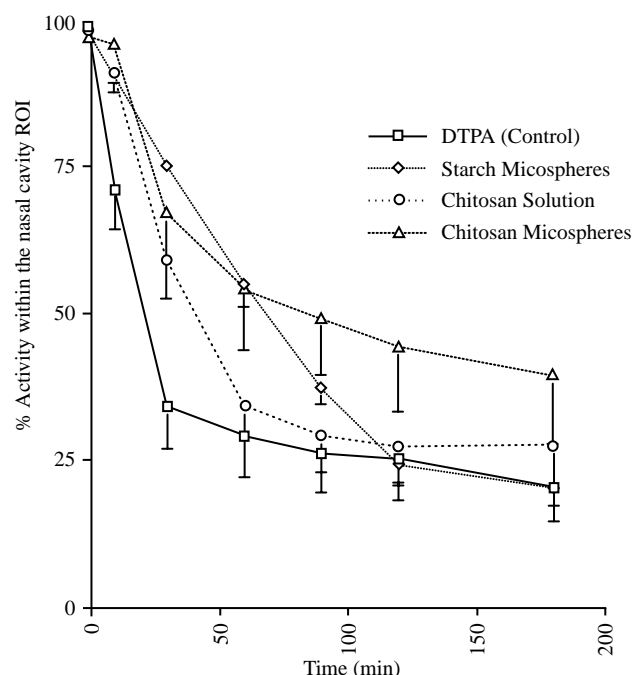


Fig. 4 The nasal clearance of bioadhesive formulations and a control in human volunteers. DTPA = diethylenetriaminepentaacetic acid. (From Ref. 55.)

Recently, starch and chitosan microspheres as well as chitosan solutions were tested for their clearance characteristics in human volunteers using gamma scintigraphy (55). The results revealed a 4-, 3-, and 2-times longer clearance half-life (compared to controls) for chitosan microspheres, starch microspheres, and chitosan solutions (Fig. 4). These observations support the hypothesis that chitosan delivery systems can reduce the rate of clearance from the nasal cavity, thereby increasing the contact time of the delivery system with the nasal mucosa and providing the potential for raising the bioavailability of drugs incorporated into these systems.

Ocular Route

The ocular route is used mainly for the local treatment of eye pathologies. Absorption of drugs administered by conventional eyedrops can result in poor ocular bioavailabilities (2–10%). This is due to the limited area of absorption, the lipophilic character of the corneal epithelium, and a series of elimination factors that reduce the contact time of the medication with the corneal surface, such as drainage of instilled solutions, lacrimation, and tear turnover and tear evaporation (56).

The first structure encountered by an ocular dosage form is the precorneal tear film, consisting of three layers:

Outer layer, oily and lipid, mainly prevents tear evaporation.

Middle layer, which is an aqueous salt solution layer, and Inner layer, a mucus layer secreted by the conjunctiva goblet cells and the lacrimal gland. This layer is important for wetting the corneal and conjunctival epithelia. The ocular membranes comprise the cornea (not vascularized) and the conjunctiva (vascularized). The corneal epithelium consists of five or six layers of nonkeratinized squamous cells, and it is considered to be the major pathway for ocular drug penetration (57).

The following types of mucoadhesive preparations have been evaluated for ocular drug delivery: hydrogels, viscous liquids, solids (inserts), and particulate formulations (57). Hui and Robinson (58) introduced hydrogels consisting of cross-linked polyacrylic acid for ocular delivery of progesterone in rabbits. These preparations increased progesterone concentrations in the aqueous humor four times over aqueous suspensions.

Davies et al. (59) compared the precorneal clearance of Carbopol 934P to that of an equiviscous nonmucoadhesive PVA solution and phosphate buffered saline (PBS) using lacrimal dacryoscintigraphy in the rabbit. The precorneal retention of the Carbopol 934P was shown to be significantly longer than that of PVA, which, in turn, was significantly longer than that of PBS. In the same study, Carbopol 934P solution produced a significant increase in bioavailability of pilocarpine as compared to PVA and PBS. The same authors (60) described phospholipid vesicles coated with Carbopol 934P or Carbopol 1342 (a hydrophobic modified Carbopol resin). The mucoadhesive polymer-coated vesicles demonstrated substantially enhanced precorneal retention compared to noncoated vesicles at pH 5. However, the polymer-coated vesicles did not increase the ocular bioavailability of entrapped tropicamide compared to noncoated vesicles and aqueous solutions.

Lehr et al. (61) investigated two gentamicin formulations of polycarbophil (neutralized vs. nonneutralized) to pigmented rabbit eye. Both polymeric formulations doubled the uptake of gentamicin by the bulbar conjunctiva.

Saettone et al. (62) evaluated low viscosity polymers (polygalacturonic acid, hyaluronic acid, carboxymethyl-amylose, carboxymethylchitin, chondroitin sulfate, heparan sulfate, and mesoglycan) as potential mucoadhesive carriers for cyclopentolate and pilocarpine in a study of their influence on miotic activity in rabbits. Small but significant increases in bioavailability were observed and a

correlation was found between the bioavailability of the two drugs and the mucoadhesive bond strength of the polymers investigated.

Calvo et al. (63) studied chitosan- and poly-L-Lysine (PLL)-coated poly- ϵ -Caprolactone (PECL) nanocapsules for ocular application. In comparison with commercial eyedrops, the systems investigated (uncoated, PLL-coated, and chitosan-coated nanocapsules) significantly increased the concentrations of indomethacin in the cornea and aqueous humor of rabbit eyes. The chitosan-coated formulation doubled the ocular bioavailability of indomethacin over the uncoated particles, whereas the PLL coating was ineffective. The authors concluded that the specific nature of chitosan was responsible for the enhanced indomethacin uptake and not the positive surface charge. Both the PLL- and chitosan-coated nanocapsules displayed good ocular tolerance (63).

A recent approach to ocular inserts was presented by Chetoni et al. (64) in a study of cylindrical devices for oxytetracycline, made from mixtures of silicone elastomer and grafted on the surface of the inserts with an interpenetrating mucoadhesive polymeric network of polyacrylic acid or polymethacrylic acid. The inserts were tested for drug release and retention at rabbit eyes. It was shown that some of the inserts are able to maintain prolonged oxytetracycline concentrations in the lacrimal fluid for 36 h.

Vaginal Route

The vaginal route is considered to be suitable for the local application and absorption of therapeutics like estrogens for hormone replacement therapy or contraception. Systemic absorption of peptide drugs such as LHRH agonists and calcitonin can also be achieved (65).

The vagina offers a substantial area for drug absorption because numerous folds in the epithelium increase the total surface area. A rich vascular network surrounds the vagina whereas the vaginal epithelium is covered by a film of moisture consisting mainly of cervical mucus and fluid secreted from the vaginal wall.

Conventional vaginal delivery systems include tablets, foam gels, suspensions, and pessaries. Mucoadhesive gel formulations based on polycarbophil have been reported to remain 3–4 days at the vaginal tissue, providing an excellent vehicle for the delivery of progesterone and nonoxynol-9 (66).

The benzyl ester of hyaluronic acid (HYAFF 11) is a highly mucoadhesive polymer which can be processed into microspheres. Such microspheres containing salmon calcitonin were intravaginally administered to rats as a dosage form for the prevention of ovariectomy

osteopenia (65). In recent studies, HYAFF 11–salmon calcitonin microspheres were formulated as single-dose pessaries, resulting in sustained plasma concentrations of calcitonin (67).

Gastrointestinal Route

The peroral route represents the most convenient route of drug administration, being characterized by high patient compliance. The mucosal epithelium along the gastrointestinal tract varies. In the stomach the surface epithelium consists of a single layer of columnar cells whose apical membrane is covered by a conspicuous glycocalyx. A thick layer of mucus covers the surface to protect against aggressive luminal content. This site of the tract is of minor interest for drug delivery since the low pH and the presence of proteolytic enzymes make the stomach a rather hostile environment. However, there are examples of dosage forms specially designed to be retained in the stomach such as some gastroretentive systems consisting of mucoadhesive hydrogels (68). Akiyama et al. (69) evaluated microspheres for prolonged residence time in the gastrointestinal tract of rats. They prepared two types of polyglycerol fatty acid ester (PGEF)-based microspheres, Carbopol 934P-coated microspheres, and Carbopol 934P-dispersion microspheres. Significantly longer residence times were observed after administration of the dispersion-microspheres than with the coated ones. Additionally, it was shown that the microspheres were retained in the stomach of the animals.

The small intestine is characterized by an enormous surface area available for the absorption of nutrients and drugs. This large area is formed by crypts and villi. The intestinal epithelium consists of a single layer of three types of columnar cells: enterocytes, goblet cells, and enteroendocrine cells. The enterocytes are linked to each other by tight junctions and desmosomes. The goblet cells are mucin-producing unicellular glands intercalated between the enterocytes. The enteroendocrine cells are scattered between the enterocytes and goblet cells and release hormones that can modify the local environment or influence the intestinal motility. At the terminal ileum, the Peyer's patches, a particular specialization of the gut-immune system, are located. This domain contains the M cells, which are specialized in endocytosis and processing luminal antigens. The large intestine (colon) has the same cell populations as the small intestine, and its main function is the absorption of water and electrolytes. The role of mucus in the intestine is to facilitate the passage of food along the intestinal tract and to protect the gut from bacterial infections (70).

In the past decade several difficulties have been encountered in the design of successful mucoadhesive delivery systems for peroral applications. The reasons may be due to shortcomings of the mucoadhesive properties of the polymers or to the peculiar physiological limits of the digestive tract, soluble mucins, and shed-off mucus, food, or other contents of the intestinal lumen which would inactivate the mucoadhesive properties of the delivery system before having reached the absorbing membrane. Furthermore, the adhesion of the delivery system can last as long as the gel-state mucus itself remains attached to the intestinal mucosal tissue. Mucus turnover is continuously removing the mucus gel layer attached to the epithelium by a steady-state process (71).

The failure in increasing residence time of mucoadhesive systems in the human intestinal tract has led scientists to the evaluation of multifunctional mucoadhesive polymers. Research in the area of mucoadhesive drug delivery systems has shed light on other properties of some of the mucoadhesive polymers. One important class of mucoadhesive polymers, poly(acrylic acid) derivatives, has been identified as potent inhibitors of proteolytic enzymes (72–74). The interaction between various types of mucoadhesive polymers, and epithelial cells has a direct influence on the permeability of mucosal epithelia by means of changing the gating properties of the tight junctions. More than being only adhesives, some mucoadhesive polymers can therefore be considered as a novel class of “multifunctional macromolecules” with a number of desirable properties for their use as delivery adjuvants (72, 75).

Lueßen et al. (73, 74) evaluated the mucoadhesive polyacrylates, polycarbophil and Carbopol 934P, for their potency to inhibit intestinal proteases. These polymers are able to inhibit the activities of trypsin, α -chymotrypsin, and carboxypeptidase A and B as well as of cytosolic leucine aminopeptidase. Carbopol 934P was found to be more efficient in reducing proteolytic activity than polycarbophil (74). The pronounced binding properties of polycarbophil and Carbopol 934P for bivalent cations, such as zinc and calcium, were demonstrated to be a major reason for the observed inhibitory effect. These polymers have been shown to remove Ca^{2+} and Zn^{2+} , respectively, from the enzyme structures, thereby inhibiting their activities. Carboxypeptidase A and α -chymotrypsin activities were observed to be reversible upon the addition of Zn^{2+} and Ca^{2+} ions, respectively. Therefore, it was concluded that polyacrylates are promising excipients to protect peptide drugs from intestinal degradation. In vitro studies, using the Caco-2 cell intestinal epithelium model, showed that Carbopol 934P was able to substantially increase the transport of a macromolecular paracellular

fluorescent marker (dextran) and the peptide drug 9-desglycinamide, 8-L-arginine vasopressin (76).

Carbopol 934P and chitosan gels were also tested in vivo for their ability to increase the absorption of the peptide analog buserelin when administered intraduodenally in rats (77). Both polymers increased the absorption of the peptide significantly, probably due to both permeation-enhancing and enzyme-inhibition properties; mucoadhesion played a secondary role. Chitosan was found to remarkably increase the peroral bioavailability of the peptide in comparison to Carbopol 934P (77), indicative of a more specific effect of chitosan with the tight junctions, than previously suggested by Artursson et al. (78).

Chitosan and chitosan salts, however, lack the advantage of good solubility at neutral pH values. They aggregate in solutions at pH values above 6.5, and recent studies have shown that only protonated chitosan (i.e., in its uncoiled configuration) can trigger the opening of the tight junctions, thereby facilitating the paracellular transport of hydrophilic compounds (79). This property implies that chitosan can be effective as an absorption enhancer only in a limited area of the intestinal lumen where the pH values are close to its pK_a . For this reason, chitosan and its salts may not be suitable carriers for targeted peptide drug delivery to specific sites of the intestine, for instance, the jejunum or ileum. To overcome this problem the chitosan derivative *N,N,N*-trimethylchitosan chloride (TMC) has been synthesized and characterized (80). This quaternized chitosan shows higher aqueous solubility than chitosan in a much broader pH range. Chitosan HCl and TMCs of different degrees of trimethylation were tested for enhancing the permeability of the radiolabelled marker ^{14}C -mannitol in Caco-2 intestinal epithelia at neutral pH values (for instance, pH 7.2). Chitosan HCl failed to increase the permeation of these monolayers and so did TMC with a degree of trimethylation of 12.8%. However, TMC with a degree of trimethylation of 60% significantly increased the permeability of the Caco-2 intestinal monolayers, indicating that a threshold value at the charge density of the polymer is necessary to trigger the opening of the tight junctions (81). Because of the absence of significant cyto- and ciliotoxicity, TMC polymers (particularly with a high degree of trimethylation) are expected to be safe absorption enhancers for improved transmucosal delivery of peptide drugs (82).

In recent studies, both in vitro (Caco-2 cells) and in vivo in rats, TMC with a degree of trimethylation of 60% was proven to be an excellent intestinal absorption enhancer of the peptide drugs buserelin and octreotide. The observed absolute bioavailability values were 13 and 16% for buserelin and octreotide, respectively (83)

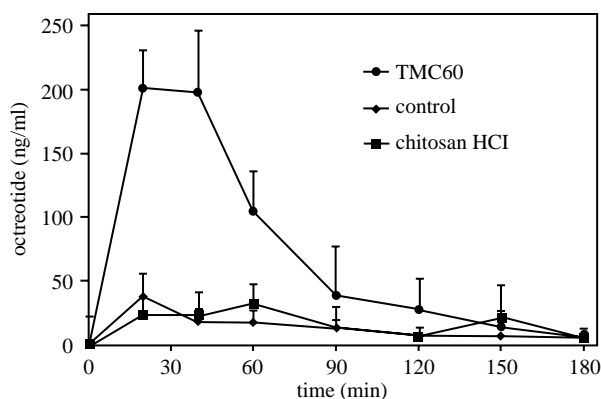


Fig. 5 Intestinal absorption of octreotide acetate in rats using mucoadhesive polymers. (From Thanou et al., unpublished data.)

(unpublished data; Fig. 5). Permeation-enhancing effects were more responsible for these increased bioavailabilities, rather than the mucoadhesive properties of the TMC polymers. Nevertheless, mucoadhesion is a prerequisite for these polymers in order to further act as absorption enhancers.

Mucus also appears to be a barrier to the permeation enhancing effect of polymeric or monomeric absorption enhancers. In the aforementioned TMC studies, the enhancement effect (enhancement ratio = permeation rate of the drug in the presence of polymer vs. permeation rate of the drug alone) was higher in vitro (Caco-2 cells; no mucus secretion) than the absorption enhancement in vivo. Meaney and O'Driscoll (84) studied the effect of mucus on the permeation properties of a micellar system consisting of sodium taurocholate in a coculture of Caco-2 and Ht29GlucH (mucin-secreting) cells. They found that the effect of bile salts on the permeation of hydrophilic paracellular markers was increased in the cocultures that were pretreated with the mucolytic compound *N*-acetylcysteine.

Bernkop-Schnürch (85) prepared a series of conjugates of protease inhibitors (pepstatin, Bowman-Birk, chymostatin, elastatinal, antipain bacitracin) and/or EDTA to three different types of polymers (carboxymethyl cellulose, polyacrylic acid, and chitosan). In addition to their mucoadhesive properties, most of these conjugates exhibited enzyme-inhibitory properties. Furthermore, the toxicity of these protease inhibitors was reduced by being covalently bound to the polymers. Chitosan-EDTA conjugates have proven to be potent inhibitors of the zinc-containing proteases as well as to be strong mucoadhesives (86).

In order to design highly mucoadhesive platforms for peroral drug delivery, Bernkop-Schnürch et al. (87) proposed thiol groups-bearing polycarboxophil modifications,

based on the mucolytic activity of thiols caused by disulfide exchange reactions between mucin glycoproteins and the mucolytic agent. Polycarbophil–cysteine conjugates appeared to exhibit superior mucoadhesiveness compared to polycarbophil itself, due to improved cohesion and rapid hydration of the gels. However, again mucus turnover is still the limiting factor of adhesion to the cell surfaces.

All the aforementioned polymers have been evaluated mainly for application in the intestine. Finally, the last part of the gastrointestinal tract, the rectum, should also be mentioned as a suitable site for delivery and fast absorption of therapeutics. Kim et al. (88) developed an in situ gelling and mucoadhesive acetaminophen liquid suppository prepared with poloxamers and sodium alginate. It was found that this particular formulation of acetaminophen in humans resulted in shorter T_{\max} and higher maximum plasma concentrations of drug (C_{\max}) than the conventional acetaminophen suppositories.

Suppositories are the preferable dosage form for patients that experience nausea. Yahagi et al. (89) evaluated a mucoadhesive suppository consisting of Witepsol H-15 and 2% Carbopol 934P for rectal delivery and absorption of the anti-emetic drug ramosclron hydrochloride (serotonin antagonist) in rabbits. These suppositories increased the $AUC_{(0-24h)}$ 2.5 times and prolonged the residence time compared to suppositories without mucoadhesive polymer. The anti-emetic effect of the formulation was tested in ferrets, and it was found that the Carbopol 934P-containing suppositories had the same effect as intravenous administration. This formulation was suggested as a once-a-day dosage form for the treatment of chemotherapy-induced nausea.

TRENDS AND PERSPECTIVES

In this article a number of polymer modifications have been described as novel drug delivery platforms, being second-generation mucoadhesive hydrogels. These polymers, both as safe absorption enhancers (75) or as improved mucoadhesive hydrogels, are the most recent developments in mucoadhesive delivery platforms for intestinal absorption of drugs.

Another trend observed during the past decade was the coating of liposomes with mucoadhesive polymers. Liposomes are coated with chitosan, long-chain PVA, and polyacrylates bearing a cholesteryl group (90). Chitosan-coated liposomes showed superior adhesion properties to rat intestine in vitro than the other polymer-coated liposomes. In vivo, chitosan-coated liposomes

containing insulin substantially reduced blood glucose levels after oral administration in rats, which were sustained up to 12 h after administration (90).

Another type of novel mucoadhesive formulations was suggested to be submicron emulsions (o/w), bearing droplets coated with Carbopol 940. These formulations have been shown to generate a 12-fold enhancement in rats in the oral bioavailability of the antidiuretic peptide drug desmopressin (91).

Specific adhesion approaches also show promise. The more specific bindings of plant lectins, mussel glue protein, and K99-fimbriae have been suggested as an alternative to the classical nonspecific mucoadhesive hydrogels. Tomato lectins were found to bind specifically onto both isolated porcine enterocytes and Caco-2 cells with the same affinity (92). However, lectin binding was inhibited in the presence of crude porcine gastric mucin, indicative of a marked cross-reactivity. Irache et al. (93) investigated three different plant lectins conjugated to latex, tomato lectin, *Asparagus pea* lectin, and *Mycoplasma gallisepticum* lectin. The extent of interactions of these three lectin–latex conjugates decreased from the duodenum to the ileum, when tested on rat intestinal mucosa without Peyer's patches. However, when mucosa containing Peyer's patches was used, a substantial increase in the interaction of the conjugates with the mucosa was found, which was more pronounced for the mycoplasma and asparagus lectins than for the tomato lectin (93).

A natural example of mucoadhesion can be the colonization of the small intestine by *Escherichia coli* strains mediated by cell-surface antigens called fimbriae (94). Fimbriae are long, thread-like protein polymers found on the surface of many bacterial strains. They enable bacteria to adhere to the brush border of epithelial cells. A special fimbriae antigen, K99-fimbriae, has been isolated from *E. coli* and bound to polyacrylic acid. The conjugate was tested by a hemagglutination assay for its ability to bind to equine erythrocytes, which have the same K-99-receptor structures as gastrointestinal epithelial cells. A 10-times stronger retention of erythrocytes was observed for the matrix-bound K-99 antigen than for the matrix-bound ovalbumin (94).

Mussel adhesive protein (MAP) is a 130-kDa protein produced by the blue mussel (*Mytilus edulis*), which provides strong adhesion to submerged surfaces. MAP films were prepared by drying and stored under nitrogen atmosphere. These films showed twice the adhesion strength of polycarbophil when tested on porcine duodenum in vitro (95).

All these examples of the applications of mucoadhesive polymers demonstrate that the use of mucoadhesive hydrogels is a powerful strategy to improve the absorption

of therapeutics across mucosal epithelia. With respect to buccal, nasal, or ocular delivery of drugs, the use of such carriers has already been successfully established. However, the application of the mucoadhesive polymers in the gastrointestinal tract is still waiting for a breakthrough. Specific-binding principles may be applied in the near future to design a third generation of mucoadhesive polymers for application in the gastrointestinal tract.

REFERENCES

1. Peppas, N.A.; Buri, P. Surface, Interfacial and Molecular Aspects of Polymer Bioadhesion on Soft Tissue. *J. Controlled Release* **1985**, *2*, 257–275.
2. Smart, J.D.; Kellaway, I.W.; Worthington, H.E.C. An In Vitro Investigation of Mucosa Adhesive Materials for Use in Controlled Drug Delivery. *J. Pharm. Pharmacol.* **1984**, *36*, 295–299.
3. Park, K.; Robinson, J.R. Bioadhesive Polymers as Platforms for Oral Controlled Drug Delivery: Method to Study Bioadhesion. *Int. J. Pharm.* **1984**, *19*, 107–127.
4. Harding, S.E.; Davis, S.S.; Deacon, M.P.; Fiebrig, I. Biopolymer Mucoadhesives. *Biotech. Gen. Engin.* **1999**, *16*, 41–86.
5. Duchêne, D.; Ponchel, G. Bioadhesion of Solid Oral Dosage Forms, Why and How? *Eur. J. Pharm. Biopharm.* **1997**, *44*, 15–23.
6. Ponchel, G.; Irache, J.M. Specific and Non-Specific Bioadhesive Particulate Systems for Oral Delivery to the Gastrointestinal Tract. *Adv. Drug Del. Rev.* **1998**, *34*, 191–219.
7. Chen, H.M.; Langer, R. Oral Particulate Delivery: Status and Future Trends. *Adv. Drug Del. Rev.* **1998**, *34*, 339–350.
8. Hoogstraate, J.A.J.; Wertz, P.W. Drug Delivery via the Buccal Mucosa. *Pharm. Sci. Tech. Today* **1998**, *1*, 309–316.
9. Ahuja, A.; Khar, R.K.; Ali, J. Mucoadhesive Drug Delivery Systems. *Drug Dev. Ind. Pharm.* **1997**, *23*, 489–515.
10. Campbell, B.J. Biochemical and Functional Aspects of Mucus and Mucin-type Glycoproteins. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 85–103.
11. Leung, S.H.; Robinson, J.R. The Contribution of Anionic Polymer Structural Features to Mucoadhesion. *J. Controlled Release* **1988**, *5*, 223–231.
12. MacAdam, A. The Effect of Gastrointestinal Mucus on Drug Absorption. *Adv. Drug Del. Rev.* **1993**, *11*, 201–220.
13. Rubinstein, A.; Tirosh, B. Mucus Gel Thickness and Turnover in the Gastrointestinal Tract of the Rat: Response to Cholinergic Stimulus and Implication for Mucoadhesion. *Pharm. Res.* **1994**, *11*, 794–799.
14. Lehr, C.-M.; Poelma, F.G.J.; Junginger, H.E.; Tukker, J.J. An Estimate of Turnover Time of Intestinal Mucus Gel Layer in the Rat In Situ Loop. *Int. J. Pharm.* **1991**, *70*, 235–240.
15. Baier, R.E.; Shafrin, E.G.; Zisman, W.A. Adhesion: Mechanisms that Assist or Impede It. *Science* **1968**, *162*, 1360–1368.
16. Ponchel, G.; Touchard, D.; Duchêne, D.; Peppas, N.A. Bioadhesive Analysis of Controlled-Release Systems. I. Fracture and Interpenetration Analysis in Poly(acrylic acid) Containing Systems. *J. Controlled Release* **1987**, *5*, 129–141.
17. Duchêne, D.; Ponchel, G. Principle and Investigation of the Bioadhesion Mechanism of Solid Dosage Forms. *Biomaterials* **1991**, *13*, 709–714.
18. Peppas, N.A.; Ponchel, G.; Duchêne, D. Bioadhesive Analysis of Controlled-Release Systems. II. Time-Dependent Bioadhesive Stress in Poly(acrylic acid)-Containing Systems. *J. Controlled Release* **1987**, *5*, 143–150.
19. Lehr, C.-M.; Bouwstra, J.A.; Schacht, E.H.; Junginger, H.E. In Vitro Evaluation of Mucoadhesive Properties of Chitosan and Some Other Natural Polymers. *Int. J. Pharm.* **1992**, *78*, 43–48.
20. Mikos, A.; Peppas, N.A. Bioadhesive Analysis of Controlled-Release Systems. IV. An Experimental Method for Testing the Adhesion of Microparticles with Mucus. *J. Controlled Release* **1990**, *12*, 31–37.
21. Lehr, C.-M.; Bouwstra, J.A.; Tukker, J.J.; Junginger, H.E. Intestinal Transit of Bioadhesive Microspheres in an In Situ Loop in the Rat—A Comparative Study with Copolymers and Blends based on Poly(acrylic acid). *J. Controlled Release* **1990**, *13*, 51–62.
22. Chickering, D.E., III; Mathiowitz, E. Bioadhesive Microspheres. I. A Novel Electrobalance-Based Method to Study Adhesive Interactions between Individual Microspheres and Intestinal Mucosa. *J. Controlled Release* **1995**, *34*, 251–262.
23. Chickering, D.E., III; Jacob, J.S.; Mathiowitz, E. Bioadhesive Microspheres. 2. Characterization and Evaluation of Bioadhesion Involving Hard, Bioerodible Polymers and Soft Tissue. *Reactive Polymers* **1995**, *25*, 189–206.
24. Park, K. A New Approach to Study Mucoadhesion: Colloidal Gold Staining. *Int. J. Pharm.* **1989**, *53*, 209–217.
25. Fiebrig, I.; Várum, K.M.; Harding, S.E.; Davis, S.S.; Stoke, B.T. Colloidal Gold and Colloidal Gold Labelled Wheat Germ Agglutinin as Molecular Probes for Identification in Mucin/Chitosan Complexes. *Carbohydr. Polym.* **1997**, *33*, 91–99.
26. Mortazavi, S.A.; Carpenter, B.G.; Smart, J.D. An Investigation of the Rheological Behavior of the Mucoadhesive/Mucosal Interface. *Int. J. Pharm.* **1992**, *83*, 221–225.
27. Mortazavi, S.A.; Smart, J.D. An Investigation into the Role of Water Movement and Mucus Gel Dehydration in Mucoadhesion. *J. Controlled Release* **1993**, *25*, 197–203.
28. Madsen, F.; Eberth, K.; Smart, J.D. A Rheological Assessment of the Nature of Interactions between Mucoadhesive Polymers and a Homogenised Mucus Gel. *Biomaterials* **1998**, *19*, 1083–1092.
29. Hägerström, H.; Paulsson, M.; Edsman, K. Evaluation of Mucoadhesion for Two Polyelectrolyte Gels in Simulated Physiological Conditions Using a Rheological Method. *Eur. J. Pharm. Biopharm.* **2000**, *9*, 301–309.
30. Rossi, S.; Bonferoni, M.C.; Lippoli, G.; Bertoni, M.; Ferrari, F.; Caramella, C.; Conte, U. Influence of Mucin Type on Polymer–Mucin Rheological Interactions. *Biomaterials* **1995**, *16*, 1073–1079.

31. Rossi, S.; Bonferoni, M.C.; Caramella, C.; Ironi, L.; Tentoni, S. Model-Based Interpretation of Creep Profiles for the Assessment of Polymer–Mucin Interaction. *Pharm. Res.* **1999**, *16*, 1456–1463.
32. Caramella, C.M.; Rossi, S.; Bonferoni, M.C. A Rheological Approach to Explain the Mucoadhesive Behavior of Polymer Hydrogels. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 25–65.
33. Jabbari, E.; Wisniewski, N.; Peppas, N.A. Evidence of Mucoadhesion by Chain Interpenetration at a Poly(acrylic acid)/Mucin Interface Using ATR-FTIR Spectroscopy. *J. Controlled Release* **1993**, *26*, 99–108.
34. Hertzog, B.A.; Mathiowitz, E. Novel Magnetic Technique to Measure Bioadhesion. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 147–173.
35. He, P.; Davis, S.S.; Illum, L. In Vitro Evaluation of the Mucoadhesive Properties of Chitosan Microspheres. *Int. J. Pharm.* **1998**, *166*, 75–88.
36. Fiebrig, I.; Harding, S.E.; Rowe, A.J.; Hyman, S.C.; Davis, S.S. Transmission Electron Microscopy Studies on Pig Gastric Mucin and its Interactions with Chitosan. *Carbohydr. Polym.* **1995**, *28*, 239–244.
37. Lehr, C.-M.; Bouwstra, J.A.; Spies, F.; Onderwater, J.; Van het Noordeinde, J.; Vermeij-Keers, C.; Van Munsteren, C.J.; Junginger, H.E. Visualization Studies of the Mucoadhesive Interface. *J. Controlled Release* **1992**, *18*, 249–260.
38. Sahlin, J.J.; Peppas, N.A. Enhanced Hydrogel Adhesion by Polymer Interdiffusion: Use of Linear Poly(ethylene glycol) as an Adhesion Promoter. *J. Biomater. Sci. Polym. Ed.* **1997**, *8*, 421–436.
39. Zur Mühlen, E.; Koschinski, P.; Gehring, S.; Ros, R.; Tiefenauer, L.; Haltner, E.; Lehr, C.-M.; Hartmann, U.; Schwesinger, F.; Plückthun, A. Force Microscopy of Cells to Measure Bioadhesion. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 197–221.
40. Squier, C.A.; Wertz, P.W. Structure and Function of the Oral Mucosa and Implications for Drug Delivery. *Oral Mucosal Drug Delivery*; Rathbone, M.J., Ed.; Marcel Dekker, Inc.: New York, 1996; 1–26.
41. Machida, Y.; Nagai, T. Bioadhesive Preparations as Topical Dosage Forms. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 641–657.
42. Bouckaert, S.; Lefebvre, R.; Remon, J.-P. In Vitro/In Vivo Correlation of the Bioadhesive Properties of a Buccal Bioadhesive Miconazole Tablet. *Pharm. Res.* **1993**, *10*, 853–856.
43. Lee, Y.L.; Chien, Y.W. Oral Mucosa Controlled Delivery of LHRH by Bilayer Mucoadhesive Polymer Systems. *J. Controlled Release* **1995**, *37*, 251–261.
44. Nair, M.K.; Chien, Y.W. Development of Anticandidal Delivery Systems. 2. Mucoadhesive Devices for Prolonged Drug Delivery in the Oral Cavity. *Drug Dev. Ind. Pharm.* **1996**, *22*, 243–253.
45. Remunan-Lopez, C.; Portero, A.; VilaJato, J.L.; Alonso, M.J. Design and Evaluation of Chitosan/Ethylcellulose Mucoadhesive Bilayered Devices for Buccal Drug Delivery. *J. Controlled Release* **1998**, *55*, 143–152.
46. Needleman, I.G.; Martin, G.P.; Smales, F.C. Characterisation of Bioadhesives for Periodontal and Oral Mucosal Drug Delivery. *J. Clin. Periodontol.* **1998**, *25*, 74–82.
47. Martin, E.; Schipper, N.G.M.; Verhoef, J.C.; Merkus, F.W.H.M. Nasal Mucociliary Clearance as a Factor in Nasal Drug Delivery. *Adv. Drug Del. Rev.* **1998**, *29*, 13–38.
48. Zhou, M.P.; Donovan, M.D. Intranasal Mucociliary Clearance of Putative Bioadhesive Polymer Gels. *Int. J. Pharm.* **1996**, *135*, 115–125.
49. Illum, L.; Farraj, N.F.; Davis, S.S. Chitosan as a Novel Nasal Delivery System for Peptide Drugs. *Pharm. Res.* **1994**, *11*, 1186–1189.
50. Oechslein, C.R.; Fricker, G.; Kissel, T. Nasal Delivery of Octreotide: Absorption Enhancement by Particulate Carrier Systems. *Int. J. Pharm.* **1996**, *139*, 25–32.
51. Nakamura, F.; Ohta, R.; Machida, Y.; Nagai, T. In Vitro and In Vivo Nasal Mucoadhesion of Some Water-Soluble Polymers. *Int. J. Pharm.* **1996**, *134*, 173–181.
52. Illum, L.; Jorgensen, H.; Bisgaard, H.; Krogsgaard, O.; Rossing, N. Bioadhesive Microspheres as a Potential Nasal Drug Delivery System. *Int. J. Pharm.* **1987**, *39*, 189–199.
53. Illum, L. Bioadhesive Formulations for Nasal Peptide Delivery. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 507–539.
54. Nakamura, K.; Maltani, Y.; Lowman, A.M.; Takayama, K.; Peppas, N.A.; Nagai, T. Uptake and Release of Budesonide from Mucoadhesive, pH-Sensitive Copolymers and their Application to Nasal Delivery. *J. Controlled Release* **1999**, *61*, 329–335.
55. Soane, R.J.; Frier, M.; Perkins, A.C.; Jones, N.S.; Davis, S.S.; Illum, L. Evaluation of the Clearance Characteristics of Bioadhesive Systems in Humans. *Int. J. Pharm.* **1999**, *178*, 55–65.
56. Saetstone, M.F.; Burgalassi, S.; Chetoni, P. Ocular Bioadhesive Drug Delivery Systems. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 601–640.
57. Joshi, A. Microparticulates for Ophthalmic Drug Delivery. *J. Ocul. Pharmacol.* **1994**, *10*, 29–45.
58. Hui, H.W.; Robinson, J.R. Ocular Delivery of Progesterone Using a Bioadhesive Polymer. *Int. J. Pharm.* **1985**, *26*, 203–213.
59. Davies, N.M.; Farr, S.J.; Hadgraft, J.; Kellaway, I.W. Evaluation of Mucoadhesive Polymers in Ocular Drug Delivery. *J. Viscous Solutions. Pharm. Res.* **1991**, *8*, 1039–1043.
60. Davies, N.M.; Farr, S.J.; Hadgraft, J.; Kellaway, I.W. Evaluation of Mucoadhesive Polymers in Ocular Drug Delivery. II. Polymer-Coated Vesicles. *Pharm. Res.* **1992**, *9*, 1137–1144.
61. Lehr, C.-M.; Lee, Y.H.; Lee, V.H. Improved Ocular Penetration of Gentamicin by Mucoadhesive Polymer Polycarbophil in the Pigmented Rabbit. *Invest. Ophthalmol. Vis. Sci.* **1994**, *35*, 2809–2814.

62. Saettone, M.F.; Monti, D.; Torracca, M.T.; Chetoni, P. Mucoadhesive Ophthalmic Vehicles: Evaluation of Polymeric Low-Viscosity Formulations. *J. Ocul. Pharmacol.* **1994**, *10*, 83–92.
63. Calvo, P.; VilaJato, J.L.; Alonso, M.J. Evaluation of Cationic Polymer-Coated Nanocapsules as Ocular Drug Carriers. *Int. J. Pharm.* **1997**, *153*, 41–50.
64. Chetoni, P.; Di-Colo, G.; Grandi, M.; Morelli, M.; Saettone, M.F.; Darougar, S. Silicone Rubber/Hydrogel Composite Ophthalmic Inserts: Preparation and Preliminary In Vitro/In Vivo Evaluation. *Eur. J. Pharm. Biopharm.* **1998**, *46*, 125–132.
65. Bonucci, E.; Ballanti, P.; Ramires, P.A.; Richardson, J.L.; Benedetti, L.M. Prevention of Ovariectomy Osteopenia in Rats after Vaginal Administration of Hyaff 11 Microspheres Containing Salmon Calcitonin. *Calcif. Tissue Int.* **1995**, *56*, 274–279.
66. Robinson, J.R.; Bologna, W.J. Vaginal and Reproductive-System Treatments Using Bioadhesive Polymer. *J. Controlled Release* **1994**, *28*, 87–94.
67. Richardson, J.L.; Armstrong, T.I. Vaginal Delivery of Calcitonin by Hyaluronic Acid Formulations. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 563–599.
68. Hwang, S.J.; Park, H.; Park, K. Gastric Retentive Drug-Delivery Systems. *Crit. Rev. Ther. Drug Carr. Syst.* **1998**, *15*, 243–284.
69. Akiyama, Y.; Nagahara, N.; Kashiara, T.; Hirai, S.; Toguchi, H. In Vitro and In Vivo Evaluation of Mucoadhesive Microspheres Prepared for the Gastrointestinal Tract Using Polyglycerol Esters of Fatty Acids and a Poly(acrylic acid) Derivative. *Pharm. Res.* **1995**, *12*, 397–405.
70. Schumacher, U.; Schumacher, D. Functional Histology of Epithelia Relevant for Drug Delivery. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.-M., Eds.; Marcel Dekker, Inc.: New York, 1999; 67–83.
71. Lehr, C.-M. From Sticky Stuff to Sweet Receptors—Achievements, Limits and Novel Approaches to Bioadhesion. *Eur. J. Drug Metab. Pharmacokinet.* **1996**, *21*, 139–148.
72. Lehr, C.-M. Bioadhesion technologies for the Delivery of Peptide and Protein Drugs to the Gastrointestinal Tract. *Crit. Rev. Ther. Drug Carr. Syst.* **1994**, *11*, 119–160.
73. Lueßen, H.L.; De Leeuw, B.J.; Pérard, D.; Lehr, C.-M.; De Boer, A.G.; Verhoef, J.C.; Junginger, H.E. Mucoadhesive Polymers in Peroral Peptide Drug Delivery. 1. Influence of Mucoadhesive Excipients on the Proteolytic Activity of Intestinal Enzymes. *Eur. J. Pharm. Sci.* **1996**, *4*, 117–128.
74. Lueßen, H.L.; Verhoef, J.C.; Borchard, G.; Lehr, C.-M.; De Boer, A.G.; Junginger, H.E. Mucoadhesive Polymers in Peroral Peptide Drug Delivery. II. Carbomer and Polycarbophil are Potent Inhibitors of the Intestinal Proteolytic Enzyme Trypsin. *Pharm. Res.* **1995**, *12*, 1293–1298.
75. Junginger, H.E.; Verhoef, J.C. Macromolecules as Safe Penetration Enhancers for Hydrophilic Drugs—A Fiction? *Pharm. Sci. Tech. Today* **1998**, *1*, 370–376.
76. Lueßen, H.L.; Rentel, C.O.; Kotze, A.F.; Lehr, C.-M.; De Boer, A.G.; Verhoef, J.C.; Junginger, H.E. Mucoadhesive Polymers in Peroral Peptide Drug Delivery. 4. Polycarbophil and Chitosan are Potent Enhancers of Peptide Transport across Intestinal Mucosae In Vitro. *J. Controlled Release* **1997**, *45*, 15–23.
77. Lueßen, H.L.; De Leeuw, B.J.; Langemeyer, M.W.; De Boer, A.G.; Verhoef, J.C.; Junginger, H.E. Mucoadhesive Polymers in Peroral Peptide Drug Delivery. VI. Carbomer and Chitosan Improve the Intestinal Absorption of the Peptide Drug Buserelin In Vivo. *Pharm. Res.* **1996**, *13*, 1668–1672.
78. Artursson, P.; Lindmark, T.; Davis, S.S.; Illum, L. Effect of Chitosan on the Permeability of Monolayers of Intestinal Epithelial Cells (Caco-2). *Pharm. Res.* **1994**, *11*, 1358–1361.
79. Kotzé, A.F.; Lueßen, H.L.; De Boer, A.G.; Verhoef, J.C.; Junginger, H.E. Chitosan for Enhanced Intestinal Permeability: Prospects for Derivatives Soluble in Neutral and Basic Environments. *Eur. J. Pharm. Sci.* **1999**, *7*, 145–151.
80. Sieval, A.B.; Thanou, M.; Kotzé, A.F.; Verhoef, J.C.; Brussee, J.; Junginger, H.E. Preparation and NMR Characterization of Highly Substituted *N*-Trimethylchitosan Chloride. *Carbohydr. Polym.* **1998**, *36*, 157–165.
81. Kotzé, A.F.; Thanou, M.M.; Lueßen, H.L.; De Boer, A.G.; Verhoef, J.C.; Junginger, H.E. Enhancement of Paracellular Drug Transport with Highly Quaternized *N*-trimethylchitosan Chloride in Neutral Environments: In Vitro Evaluation in Intestinal Epithelial Cells (Caco-2). *J. Pharm. Sci.* **1999**, *88*, 253–257.
82. Thanou, M.M.; Verhoef, J.C.; Romeijn, S.G.; Merkus, F.W.H.M.; Nagelkerke, J.F.; Junginger, H.E. Effects of *N*-trimethylchitosan Chloride, A Novel Absorption Enhancer, on Caco-2 Intestinal Epithelia and the Ciliary Beat Frequency of Chicken Embryo Trachea. *Int. J. Pharm.* **1999**, *185*, 73–82.
83. Thanou, M.; Florea, B.I.; Langermeijer, M.W.E.; Verhoef, J.C.; Junginger, H.E. *N*-Trimethylated Chitosan Chloride (TMC) Improves the Intestinal Permeation of the Peptide Drug Buserelin In Vitro (Caco-2 Cells) and In Vivo (Rats). *Pharm. Res.* **2000**, *17*, 27–31.
84. Meaney, C.; O'Driscoll, C. Mucus as a Barrier to the Permeability of Hydrophilic and Lipophilic Compounds in the Absence and Presence of Sodium Taurocholate Micellar Systems Using Cell Culture Models. *Eur. J. Pharm. Sci.* **1999**, *8*, 167–175.
85. Bernkop-Schnürch, A. The Use of Inhibitory Agents to Overcome the Enzymatic Barrier to Perorally Administered Therapeutic Peptides and Proteins. *J. Controlled Release* **1998**, *52*, 1–16.
86. Bernkop-Schnürch, A.; Krajicek, M.E. Mucoadhesive Polymers as Platforms for Peroral Peptide Delivery and Absorption: Synthesis and Evaluation of Different Chitosan-EDTA Conjugates. *J. Controlled Release* **1998**, *50*, 215–223.
87. Bernkop-Schnürch, S.A.; Schwarz, V.; Steininger, S. Polymers with Thiol Groups: A New Generation of Mucoadhesive Polymers? *Pharm. Res.* **1999**, *16*, 876–881.
88. Kim, C.K.; Lee, S.W.; Choi, H.G.; Lee, M.K.; Gao, Z.G.; Kim, I.S.; Park, K.M. Trials of In Situ Gelling and

- Mucoadhesive Acetaminophen Liquid Suppository in Human Subjects. *Int. J. Pharm.* **1998**, *174*, 201–207.
89. Yahagi, R.; Machida, Y.; Onishi, H. Mucoadhesive Suppositories of Ramosetron Hydrochloride Utilizing Carbopol. *Int. J. Pharm.* **2000**, *193*, 205–212.
90. Takeuchi, H.; Yamamoto, H.; Niwa, T.; Hino, T.; Kawashima, Y. Enteral Absorption of Insulin in Rats from Mucoadhesive Chitosan-Coated Liposomes. *Pharm. Res.* **1996**, *13*, 896–901.
91. Han, E.; Amselem, S.; Weisspapir, M.; Schwarz, J.; Yogev, A.; Zawoznik, E.; Friedman, D. Improved Oral Delivery of Desmopressin via a Novel Vehicle: Mucoadhesive Submicron Emulsion. *Pharm. Res.* **1996**, *13*, 1083–1087.
92. Lehr, C.-M.A.; Bouwstra, J.A.; Kok, W. Bioadhesion by Means of Specific Binding to Tomato Lectin. *Pharm. Res.* **1992**, *9*, 547–553.
93. Irache, J.M.; Durrer, C.; Duchêne, D.; Ponchel, G. Bioadhesion of Lectin–Latex Conjugates to Rat Intestinal Mucosa. *Pharm. Res.* **1996**, *13*, 1716–1719.
94. Bernkop-Schnürch, A.; Gabor, F.; Szostak, M.P.; Lubitz, W. An Adhesive Drug Delivery System based on K99-fimbriae. *Eur. J. Pharm. Sci.* **1995**, *3*, 293–299.
95. Schnurrer, J.; Lehr, C.-M. Mucoadhesive Properties of the Mussel Adhesive Protein. *Int. J. Pharm.* **1996**, *141*, 251–256.

MONOCLONAL ANTIBODIES FOR DRUG DELIVERY

John B. Cannon

Ho-Wah Hui

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

Pramod K. Gupta

TAP Pharmaceuticals, Deerfield, Illinois, U.S.A.

INTRODUCTION

The selective delivery of drugs to their site of action should increase their therapeutic effectiveness while minimizing unwanted side-effects. In the early 1900s, Paul Ehrlich proposed the potential use of antibodies as carriers of biological agents to the target sites, thus inventing the “magic bullet” concept. With the development of hybridoma technology, it is now possible to produce virtually unlimited quantities of homogenous antibodies having a defined specificity, that is, monoclonal antibodies (MoAbs), which have potential to fulfill Ehrlich’s vision. While MoAbs have found use in sensitive diagnostic tests, the therapeutic use of MoAbs and their conjugates are only beginning to realize the promise that was predicted with the advent of the core technology (1–5).

Basic Terminology

In simplistic terms, an antibody is an immunoglobulin synthesized by the body’s immune system in response to a foreign molecule (an antigen, i.e., an antibody generator), and is capable of binding the antigen with high specificity. In general, an antigen must have a relatively large molecular weight (>1000) to elicit an immune response; smaller molecules can be made to be antigenic by coupling it to a suitable macromolecule, e.g., albumin.

An antibody is a Y-shaped molecule (Fig. 1) (6), and contains two light chains and two heavy chains joined together by disulfide bonds. Each of the heavy chains also contains a carbohydrate residue. The bottom “trunk” portion of the antibody molecule is known as the constant (Fc) region because its amino acid sequence is often similar within a given animal species. The upper “arms”, the antigen binding regions (Fab), are known as the variable regions because its amino acid sequence is determined by the antigen responsible for its formation.

The variable region, in turn, has several “hypervariable” regions, also known as the “complementarity determining regions” (CDR), which show greater variability than the rest of the variable region.

Antibodies can be classified in the following:

Polyclonal Antibodies: After an antigen is injected into an animal by a regimen designed to induce an optimal immune response, serum can be collected from the animal and the immunoglobulin fraction isolated. This “antisera” is enriched with antibodies specific for the original antigen. Because a large number of lymphocytes are involved in the production of the antisera, antibodies produced by this classical method are called polyclonal.

Monoclonal Antibodies: An antibody is called “monoclonal” when each immunoglobulin is produced by a single clone of cells and hence is identical to every other molecule in the preparation, in terms of heavy as well as light chain structure. Thus they are highly specific and offer more consistent efficacy and predictable toxicity in vivo than the polyclonal counterparts (7).

Antibody Fragments: The earliest MoAbs examined in animal and clinical studies were murine antibodies. Because of their nonhuman origin, they are immunogenic in humans, i.e., they have a tendency to elicit a human antimouse antibody (HAMA) response. They also have been shown to have much shorter clearance rates than human MoAb’s. One approach to overcome these problems has been to cleave the antibody (e.g., by papain digestion) into its respective Fc and Fab fragments (Fig. 2a) (8). In general, the Fab fragments are less immunogenic than the corresponding intact antibodies, and their smaller molecular size may facilitate penetration into tumor tissue (9) and result in a longer half-life. However, they can lose some of their antigen binding capacity, and in some cases the

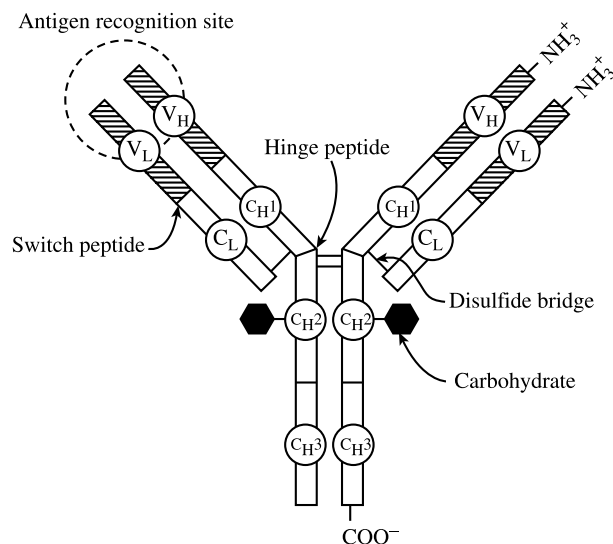


Fig. 1 Diagram of an monoclonal antibody molecule. (Adapted from Ref. 6.)

therapeutic effect may depend on the Fc portion of the antibody.

Chimeric Antibodies: The obvious solution to the problems encountered with murine antibodies would be to clone a fully human antibody. However, human hybridomas required for human MoAb production have been notoriously difficult to culture, and it may be impossible to obtain many of the appropriate antibodies. A strategy that has been devised to overcome the HAMA problem of murine MoAbs is by constructing a chimeric antibody (Fig. 2b), which contains the Fc region of human IgG, but the Fab regions are murine in origin. These can be made chemically by joining murine Fab fragments to the Human Fc fragment, but the preferred method is to use recombinant DNA technology, as detailed in a later section.

Humanized Antibodies: Although human studies have suggested that chimeric antibodies elicit less HAMA response than murine antibodies, they are still immunogenic due to their murine regions (generally about 30% of the total molecule). A major advance was achieved when it was recognized that only a small portion of an antibody molecule was actually responsible for antigen binding, in fact only the CDR regions. One can envision construction of a "humanized" antibody in which the majority of the antibody framework is human in origin, but the CDR's are murine (Fig. 2c). Synthesis of such humanized antibodies have been successfully achieved by recombinant DNA technology, and can have up to

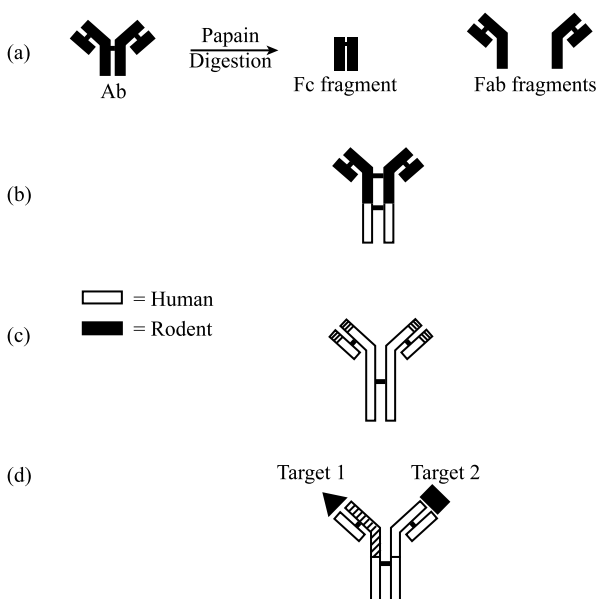


Fig. 2 Diagram of the most important MoAb constructs used clinically: a) antibody fragments; b) chimeric antibody; c) humanized antibody; d) bispecific antibody. (Adapted from Ref. 8.)

95% homology with human antibodies. (Some confusion in the literature exists regarding the terminology for humanized and chimeric antibodies, in that some authors use the terms interchangeably.)

Bispecific Antibodies: Antibodies can be constructed using recombinant DNA technology in which each of the two arms are specific for two different antigens (Fig. 2d). For example, bispecific MoAbs reactive with CD15 antigen, and composed of Fab fragments of anti-CD64 MoAb 32 and a whole IgM antimyeloid cell MoAb, PM-81, have been investigated for the therapy of patients with CD15 positive tumors, e.g., acute myelogenous leukemia, small cell carcinoma of the lung, colorectal cancer, and breast carcinoma (10).

Immunoconjugate: For MoAb targeted drug delivery, a drug is bound covalently to an antibody that is chosen to target it to the desired site of action. The resulting immunoconjugate may contain a spacer between the drug and antibody, or a polymer to increase the number of drug molecules that can be bound to each antibody. Another possibility is a radioimmunoconjugate, which is designed to be concentrated at the target site by the targeting antibody, allowing the radiation from the bound radioisotope to exert its cytotoxic affect. Alternatively, the drug can be incorporated noncovalently into a liposome or microsphere to which the targeting antibody

is bound to the surface, yielding an immunoliposome or immunomicrosphere, respectively.

PREPARATION AND MANUFACTURE OF ANTIBODIES AND ANTIBODY-BASED DELIVERY SYSTEMS

Manufacture

Every B-lymphocyte in an animal expresses an antibody of only one specificity. After it is triggered to differentiate, the B-cell turns into a plasma cell with the cytoplasmic machinery to synthesize and secrete large quantities of its own unique immunoglobulin. To prepare MoAbs, usually a mouse is immunized with the antigen of interest, e.g., human tumor cells. When an immune response ensues, B-lymphocytes from the spleen or lymph nodes of the animal are harvested in a single cell suspension. These cells are then fused with myeloma cells from the same species, using a fusogenic substance (e.g., polyethylene glycol) or electric current (11). Mutant myeloma cells are used which are deficient in an enzyme, hypoxanthine guanine phosphoribosyl transferase (HGPRT), which is needed for their survival in the presence of the folic acid antagonist, aminopterin. The resulting fused cells have the cytoplasmic machinery to promote cell division and produce large amounts of immunoglobulin. The cell suspension is then transferred to the wells of a microtiter plate in a medium such as hypoxanthine/aminopterin/thymidine. Only the hybridomas that have acquired the HGPRT from lymphocytes via cell fusion usually survive. The hybrids are cloned by limited dilution to one cell per well so that it is easy to identify an antibody of the desired titer, specificity, and avidity for propagation in mass culture. The cell supernatant is then purified by column or affinity chromatography to harvest the pure antibody.

For chimeric and humanized MoAb the above procedures are modified somewhat. The preferred method for the latter is to use recombinant DNA technology and construct a gene that expresses the chimeric or humanized antibody, by splicing the appropriate DNA sequences together in the plasmid of the hybridoma. Synthesis of such humanized antibodies have also been successfully achieved by recombinant DNA technology; the portions of gene encoding the murine CDR regions are spliced into the gene encoding the human antibody by transfection. Although these constructs proved to be much less immunogenic than murine or chimeric antibodies, early work indicated that they had lower antigen binding

capacity than the original murine MoAb. This was apparently because of the absence of certain residues in the human framework that while not directly involved in antigen binding, are required to retain the CDR regions in the correct conformation. Choice of human framework IgG, which is as homologous as possible with the murine antibody, will aid in this regard. In addition, sophisticated molecular modeling techniques based on X-ray crystallography and computer modeling has been used to identify these required residues, which has allowed introduction of the residues into the framework region by recombinant technology (12).

Due to the increased application of MoAbs in diagnostics and therapeutics, considerable effort has been made to develop technology for the large-scale production of MoAbs. Examples of the currently employed culture systems are hollow-fiber systems, suspensions, solid-phase cell immobilization, perfusion reactor, and encapsulation in semipermeable vesicles (13, 14). The system of choice is dependent on the cell line and on the desired characteristics and quantity of the final product. To increase the mixing efficiency of cell-culture equipments and to provide aeration, several different devices have been designed, e.g., vibromixer, marine propeller, turbine propeller, spinning magnetic bar, magnetic spinner, and airlift (14, 15).

Perfusion systems have also been used for successful scale-up of MoAb production. During the culture period, cell growth occurs exponentially until the cell density reaches a maximum. At that point, the medium needs a continuous supplementation of fresh nutrients and elimination of waste. In perfusion systems, fresh nutrients are supplied and wastes are removed continuously so that the medium meets the physiological needs of the cells. At steady state, the cell concentration is determined by space and other limitations. High cell densities have been achieved by immobilizing the cells in porous ceramic matrices or hollow fiber devices. Intermediate cell densities have been achieved by perfusion reactors with a spin filter, or in a fluidized bed reactor in which the cells are embedded in sponge-like microcarriers (14, 15).

Coupling Methods for Antibody Drug Conjugates

An important part of the design of an antibody-directed drug delivery system is the type of linkage and coupling method between antibody and drug. The drug can be covalently bound to the MoAb directly or through a short spacer, or the two can be conjugated through a linker such as a water-soluble polymer. Alternatively, a carrier such as

a liposome or a polymeric microsphere can be used, wherein the drug is entrapped in or bound to the carrier, and the MoAb is bound to the surface of the carrier. Characteristics that would comprise an ideal antibody directed delivery system could include, preparation by a method that has high efficiency and yield, and is capable of scaleup; high stability of the conjugate, both under shelf storage conditions and in the circulation after injection; and retention of antigen-binding ability of the antibody while it is carrying the drug to the target tissue. Finally, upon reaching the target, either the immunoconjugate itself should have the desired pharmacological effect equivalent to the free drug, or must release free drug or a derivative that is fully efficacious. Although such a system is probably impossible to achieve for most therapeutic applications, a variety of coupling reagents are fortunately available that aid in optimizing the properties of an immunoconjugate.

Amino, sulfhydryl, and carboxyl groups are the most common functional groups on the antibody, carrier, and drug molecules used for coupling. If the drug lacks the desired group, it may be possible to introduce it. For example, as shown in Fig. 3, succinic anhydride (Fig. 3a) can convert an alcohol or amino group to a carboxyl group; 2-iminothiolane (Traut's reagent, Fig. 3b) can convert an amino group to a sulfhydryl.

For linkage of drug to antibody, "classical" protein cross-linking reagents have been used to prepare immunoconjugates. For example, carbodiimide reagents (e.g., dicyclohexyl carbodiimide, DCC, Fig. 3c, and its water soluble analogs) link amino groups with carboxyls via amide bonds. In an "active ester" method, carboxyl groups of the drug are linked to *N*-hydroxy succinimide (NHS, Fig. 3d) in the presence of a carbodiimide to form an active ester derivative of the drug, which then reacts with the amino group of the antibody (16). Linkers such as dextran, allow conjugation of a much larger number of drug molecules with each antibody molecule. Thus, dextran and similar carbohydrate linkers are oxidized with periodic acid (Fig. 3e) to form aldehyde groups, which are then linked to amino groups of drug and antibody with formation of an imine. This product can be stabilized by reduction with sodium cyanoborohydride.

These simple reactions are often not specific enough for efficient immunoconjugate formation. More recently, a number of bifunctional reagents have been developed that are more specific in forming linkages of antibody to drug. Heterobifunctional reagents, which have two different reactive groups at the two ends of the molecule, have become the method of choice for preparation of immunoconjugates. Among the most widely used is

N-succinimidyl 3-(pyridyldithio) propionate (SPDP, Fig. 3f). Generally, the reagent is used to derivatize the drug with a pyridyl disulfide group; reaction of this species with the antibody containing free sulfhydryl groups yields the immunoconjugate. *N*-[6 maleimidocaproyl]oxy]succinimide (EMCS, Fig. 3g) is a reagent that reacts with amino groups at the succinimide end and sulfhydryl groups at the maleimide end. A similar reagent, *N*-[4 maleimidoethoxy succinyl]oxy]succinimide (MESS), has a metabolizable ester linker between the two active functionalities, thus providing a method to control release of free drug (17). Combinations of classical coupling methods with bifunctional reagents have also been used to advantage for preparation of immunoconjugates. For example, a 6 carbon spacer ending in a carboxyl group was introduced into dextran (MW 70,000), and then mitomycin C (MMC) was coupled to the spacer with a carbodiimide. The remaining carboxyl groups of the spacers were modified to amino groups, which were then coupled to MoAb A7 by means of SPDP, with a final MMC/MoAb ratio of 40. The antibody activity of the resulting conjugate was almost equivalent to native MoAb A7, and released free MMC by chemical hydrolysis to maintain cytotoxicity (18). Similarly, Zara et al. modified an IgM against human carcinoma by oxidation of its carbohydrate residues, which were then coupled to a bifunctional reagent, S-(2-thiopyridyl)-L-cysteine hydrazide (TPCH) via the hydrazide. After Ricin A was coupled to the other end of the reagent by a disulfide bond, the immunoconjugate retained full toxin and antibody activity with up to 16 TPCH molecules incorporated per antibody, suggesting that carbohydrate residues of the antibody were not involved in the antigen-binding process (19).

Plasma or intracellular enzymes such as esterases or proteases can potentially degrade immunoconjugate linkages. Glutathione reductase and related enzymes are instrumental in cleavage of disulfide bonds of immunoconjugates. The local pH of the target tissue or of its intracellular environment (e.g., lysosomes) may also increase the rate of release of drug from the immunoconjugate. Thus it is important to consider the possible physiological destinations after injection of the immunoconjugate, and to monitor its degradation under conditions mimicing the biological milieu, which it may encounter. Ideally, an immunoconjugate should be sufficiently stable in the circulation to allow targeting to take place; once the antibody binds to its antigen on the target cell surface, the entire immunoconjugate should be internalized into the cell and be degraded to release free drug or an active derivative. However, for many cell types internalization of the immunoconjugate

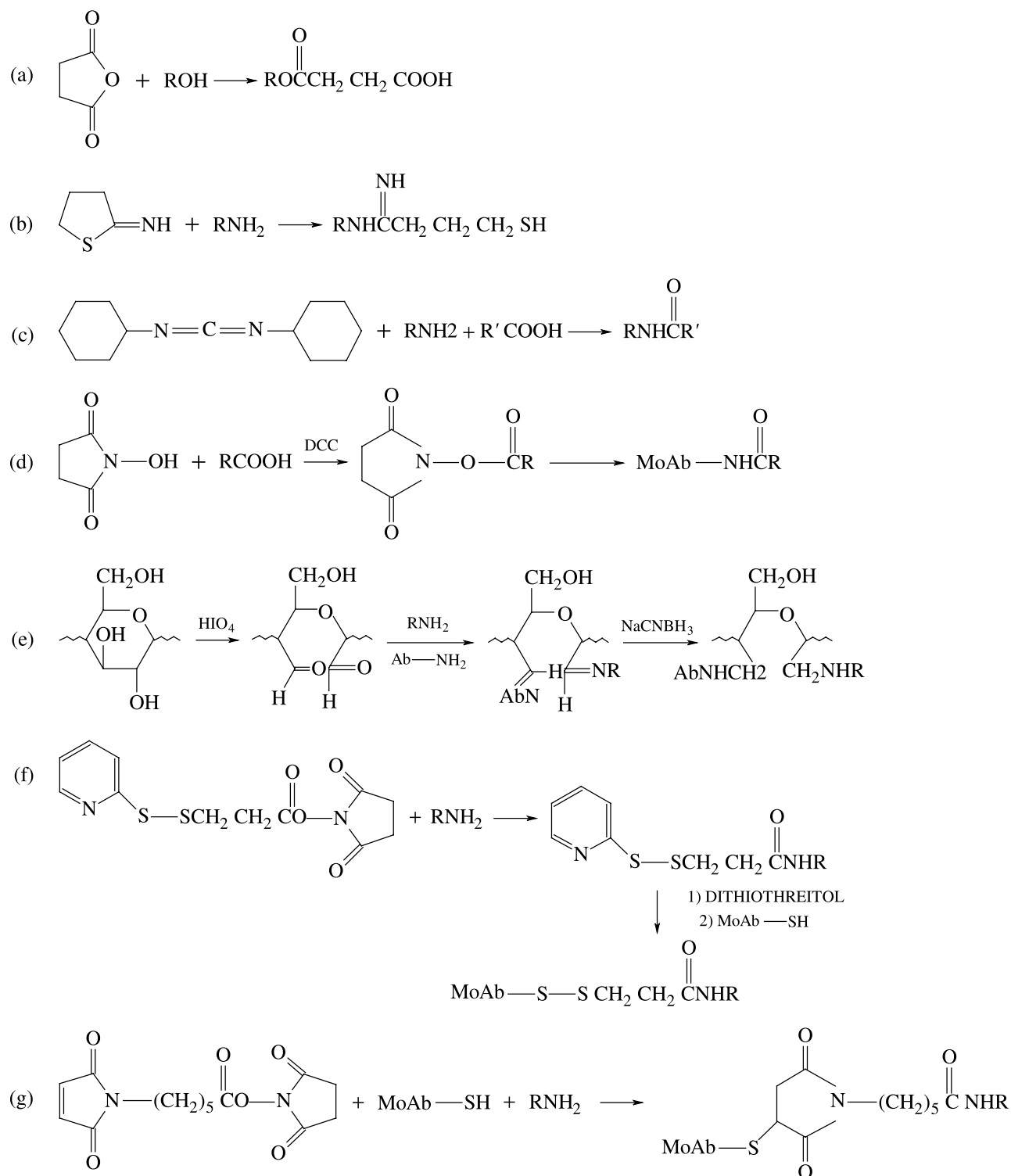


Fig. 3 Coupling reagents for conjugation of drugs to monoclonal antibodies, showing the reactions involved.

does not always occur in response to antibody binding. In that case, degradation of the immunoconjugate linkage should be sufficiently rapid to provide high local concentrations of free drug or active derivative for the desired pharmacological response. An advantage of bifunctional reagents is that they allow incorporation of a metabolizable linker into the immunoconjugate, thereby releasing free drug or an active derivative at a predictable rate.

A number of studies have explored coupling methods that allow control of the *in vivo* rate of release of active drug. Kaneko et al. (20) designed hydrazone linkers to release drug at lower pH: free doxorubicin was released from conjugated antibody within 6 h at 37°C at pH 4.5, conditions which mimic the environment of the lysosomes; the particular antibody used was known to be internalized. Similarly, Lavie et al. (21) constructed daunomycin conjugates linked to MoAb L6 via a polylysine and aconitate linkage, such that the conjugate releases free drug at pH 6. Although this antibody is not internalized, the lower pH of tumor tissue was suggested to lead to increased concentrations of free drug in tumors. New heterobifunctional reagents have been reported with greater versatility in release rate. The disulfide bond of SPDP conjugates has been shown to be labile in the circulation and release drug prematurely; an analog, NHS-ATMBA, is up to two orders of magnitude more stable than SPDP conjugates because of steric hindrance around the disulfide bond, and thus leads to more efficient targeting (22). Several new substituted 2-iminothiolane reagents, which exhibit increased stability of the disulfide bond of the resulting immunoconjugate, have also been synthesized (23).

Appropriate choice of spacer can also significantly improve the success of an immunoconjugate. For example, O'Neill and co-workers conjugated *N,N*-bis-(2-chloroethyl)-*p*-phenylenediamine (PDM) to the globulin fraction of rabbit anti-EL4 serum, which reduced the toxicity by as much as 20-fold relative to free drug. However, the conjugate was found to aggregate, making it difficult for clinical applications. Hence poly-L-glutamic acid (PGA) was tried as a spacer and it allowed preparation of a water-soluble PDM-PGA-Ig conjugate in molar ratio of 90:2:1 retaining 66% of the original antibody activity (24).

In the case of immunotoxins and other protein-antibody conjugates, a unique choice exists for their construction, viz. the conjugate can be made entirely by recombinant DNA techniques. This would require splicing the two genes together and expression of the chimeric gene in a monoclonal system. For example, the gene for angiogenin (a human toxin-like molecule) was

fused to the gene for an antitransferrin receptor, and the chimeric gene was introduced into a transfectoma to clone cell lines that secrete the hybrid antibody-angiogenin protein. The conjugate was active as a cytotoxic agent, and the activity was mediated by the transferrin receptors (25). Similar techniques were used to construct a conjugate of urokinase-type plasminogen activator and a humanized antifibrin antibody, resulting in a 12-fold enhancement of fibrinolytic activity of the conjugate relative to the parent (unconjugated) urokinase (26). Construction of an immunoconjugate by this approach has the advantage that once the clones are expressed, the immunoconjugate can be made in one step without chemical modification. Also, a single species is generally produced by this method. It has the disadvantage that generally only a 1:1 or perhaps 2:1 ratio of drug to antibody can be accommodated by the immunoconjugate.

IMMUNOTHERAPY WITH UNCONJUGATED MONOCLONAL ANTIBODIES AND RADIOIMMUNOCONJUGATES

Within the last decade, several unconjugated MoAbs have come to market for treatment of cancer, renal transplantation, and other applications, as shown in Table 1. Results from the clinical trials of these and other MoAbs in development have shown that unconjugated MoAbs are able to kill cancer and other cells. When the circulating MoAb binds to its target antigen, several mechanisms may be initiated that are responsible for the therapeutic effect. One is antibody-dependent cellular toxicity (ADCC), wherein neutrophils, mononuclear phagocytes, eosinophils, natural killer, and T-cells, which have receptors for IgG (Fc), are triggered to mediate cell destruction (27). Also important is complement dependent cytotoxicity (CDC) wherein complement binds to the Fc portion of the MoAb after antigen binding and initiates the complement cascade ending in cell death. MoAb binding to the target antigens on cell surfaces can also act as "blocking" antibodies, interfering with the binding of certain peptides or growth factors needed for cell growth, or elicit a regulatory effect on the metabolism of the cell, especially for B-cell lymphocytes and B-cell lymphomas. The interaction of MoAbs with growth factor receptors (such as transferrin receptors) may also have an antitumor effect via a regulatory mechanism because transferrin is essential for the growth of cells and its receptors are predominantly present on proliferating cells (28).

Table 1 Therapeutic MoAb drug currently marketed

Generic name	Trade name (company)	Type of MoAb	Application(s)
Rituximab	Rituxan (IDEC/Genentech)	Chimeric anti-CD20	Non-Hodgkins lymphoma
Trasuzumab	Herceptin (Genentech)	Humanized anti-HER2	Metastatic breast cancer
Palivizumab	Synagis (Medimmune)	Humanized anti-RSV epitope	Antiviral (Pediatric lower respiratory tract disease)
Muromonab-CD3	Orthoclone OKT3 (Ortho)	Murine anti-CD3	Immunosuppressant (renal transplantation)
Daclizumab	Zenapax (Roche)	Humanized anti-CD25	Immunosuppressant (renal transplantation)
Abciximab	ReoPro (Centocor)	Fab fragment of chimeric anti-7E3	Platelet aggregation inhibitor (Coronary intervention)
Basiliximab	Simulect (Novartis)	Chimeric anti-CD25	Immunosuppressant (renal transplantation)

Among the most successful has been Rituximab (Rituxan). This MoAb binds to the CD20 antigen of B-lymphocytes, which is expressed on >90% of non-Hodgkin's lymphoma B-cells. Upon binding, the Fc region of the MoAb recruits immune effector functions to mediate B-cell lysis, possibly by both CDC and ADCC mechanisms. In a multicenter clinical trial with 166 non-Hodgkin's lymphoma patients, who received 375 mg/m² over 4 doses, the overall response rate was 48% (6% complete, 42% partial). A second study with 37 patients gave similar response rates, and single doses of up to 500 mg/m² were well-tolerated (29).

Trasuzumab (Herceptin) binds to the extracellular domain of a transmembrane protein, human epidermal growth factor receptor 2 (HER2), which is overexpressed in 20–30% of primary breast cancer cells. It is thought to act primarily by ADCC. In a phase III trial, 222 breast cancer patients, who exhibited overexpressed HER2, were dosed weekly with 2 mg/kg trasuzumab after a 4 mg/kg loading dose. There was a 14% overall response (2% complete response and 12% partial response), which appeared to be correlated to the degree of HER2 overexpression. Overall response was much better when trasuzumab was combined with standard chemotherapy (viz., paclitaxel, doxorubicin + cyclophosphamide, or epirubicin + cyclophosphamide): 45% compared to chemotherapy alone (30). Similarly, trasuzumab combined with cisplatin, either in pegylated liposomes or in saline/mannitol solution, was significantly better than either treatment alone in retarding tumor growth in a mouse xenograft tumor model (31).

Some workers have proposed use of anti-idiotypic antibodies as type of "tumor vaccine." In this approach, a MoAb is prepared against a given tumor antigen.

Rather than using it for immunotherapy directly, it is used to inoculate mice, which produces a second antibody against the idiotype site of the original antibody (hence, the anti-idiotypic). After cloning and administration to patients, this anti-idiotypic MoAb would mobilize the patient's own immune system to produce a third antibody (i.e., an anti-anti-idiotypic), that would have the same idiotype of the first antibody and thus bind to the original antigen and lead to cytotoxicity (27). The perceived advantage of the approach is the multiplicative effect of the "vaccine," and it is believed to be more specific and safer than using the antigen itself as a vaccine. When 15 melanoma patients were treated with a mouse anti-idiotypic antibody homologous to a melanoma antigen, 7 patients developed the desired immune response, and there were 3 partial responses (32). Thus, although the approach may be promising, it has to be more fully evaluated to determine its utility.

Although MoAbs have many potential uses for tumor therapy, there are inherent problems associated with this approach: i) Cancer cells are heterogeneous, so those cells that are not recognized by the MoAb can escape and proliferate; ii) Some tumors contain semidead cores with poor circulation and thus cannot be reached by monoclonals; iii) MoAbs can interact with circulating target antigens before reaching their target; iv) Patients can experience possible immunogenic reactions. For these reasons, it has frequently proven more effective to combine MoAb treatment with standard chemotherapeutic agents.

Radioimmunoconjugates are MoAbs to which radio-nuclides have been conjugated, to provide cytotoxic radiation after the MoAb binds to its target antigen. The

isotopes most commonly used are Iodine-131 and Yttrium-90, both of which are β^- emitters having half-lives of 8 and 2.5 days, respectively. The former is covalently bound to tyrosine residues of the MoAb by standard chemical techniques, whereas the latter is chelated to a ligand that has been conjugated to the MoAb by techniques described in the previous section (e.g., diethylenetriaminepentaacetic acid ligand coupled with a mixed anhydride method) (33,34). Radionuclide emissions from both ^{131}I and ^{90}Y can extend to 1–5 mm of their final location, corresponding to several cell diameters. Thus, their chief advantage resides in their ability to kill tumor cells that are poorly accessible and/or antigen-negative. Unlike conventional radiation therapy, radioimmunoconjugates provide continuous radiation from the decay of the radionuclide, which allows less opportunity for the tumor cells to repair sublethal damage (35). Depending on the type of MoAb, the antibody itself may trigger CDC and ADCC mechanisms that supplement the effect of the radionuclide.

Although no radioimmunoconjugates have progressed to the market, a number have been examined in clinical trials. Bexxar (^{131}I -tositumomab) is an anti-CD20 MoAb examined in Phase III trials for non-Hodgkin's lymphoma (36). In an early trial of this radioimmunoconjugate, 19 patients with non-Hodgkin's lymphoma, who had been prescreened for favorable biodistribution of the MoAb, received 234–777 mCi of the ^{131}I -anti-CD20 MoAb. Because this was considered a myeloablative dose, the patients received autologous marrow reinfusion following the therapy. Although adverse effects were substantial due to the high dose of radiation, the regimen resulted in a complete response in 16 patients and a partial response in 2 patients; the MTD in terms of tissue exposure was determined to be ≤ 2700 cGy (37, 38). A Phase II trial with a similar regimen in 21 patients achieved 17 complete responses, with an 81% progression-free survival at 12 months (35).

APPLICATIONS OF MONOCLONAL ANTIBODIES IN DRUG DELIVERY

Principle of Targeting

Several classes of drugs lack specificity for diseased cells; for example, the cytotoxic action of chemotherapeutic agents is directed against any rapidly proliferating cell population. Due to this nonspecificity, many drugs have low therapeutic indices and often cause

serious side effects. One way of circumventing this problem is to deliver the drug in a manner such that it is preferentially localized at the desired site of action, or it predominantly attacks the diseased cells. This process is called targeting. Targeted drug delivery systems can be classified into three categories, viz. passive, physical, or active targeting (39). Passive targeting refers to the natural in vivo distribution pattern of the drug delivery system, which is determined by the inherent properties of the carrier (e.g., hydrophobic and hydrophilic surface characteristics, particle size and shape, surface charge, and particle number). For example, modulation of particle size makes it possible to passively target the lungs or reticuloendothelial system (RES) using particles $>7\text{ }\mu\text{m}$ or $0.2\text{--}7\text{ }\mu\text{m}$, respectively (40).

In physical targeting, some characteristics of the environment are utilized to guide the carrier to a specific site or to trigger selective release of its content at the site. Usually, it is accomplished via an external mechanism, such as induced local hyperthermia (e.g., using thermally sensitive liposomes) or a localized magnetic field (e.g., using magnetically responsive albumin microspheres). In active targeting, the natural disposition pattern of a carrier is modified to target it to specific organs, tissues, or cells. Although cell-specific ligands have been used to target carriers to specific cell types, this approach is probably limited to a small number of tumor types. MoAbs would thus appear to be the more generally applicable mode of active targeting. While the field is less advanced than unconjugated MoAb and radioimmunoconjugates, there has been some success in targeting toxins (i.e., immunotoxins) and drugs (i.e., drug immunoconjugates) using MoAbs as targeting agents.

Toxin Conjugates

Over the last two decades, several toxin proteins like diphtheria toxin and ricin have been conjugated to tumor specific antibodies, with moderate to high degree of success in tumor drug delivery. There are several toxins produced by plants (e.g., ricin, abrin, saporin, and gelonin) or bacteria (e.g., diphtheria toxin and pseudomonas exotoxin) used to construct immunotoxins. These toxins are highly potent, and generally a single toxin molecule is sufficient to lead to cell death (36). Most of the native toxins consist of two chains (e.g., Ricin A and B chains) one of which bind nonspecifically to cell surfaces and the other is responsible for the cytotoxicity. To construct an immunotoxin, the nonspecific binding chain (viz. Ricin B) must be removed or masked, and the

cytotoxic chain (viz. Ricin A) conjugated to a MoAb chemically or by recombinant methods. Most toxins of plant origin exert their cytotoxicity by deactivating the ribosomal protein synthesis, and thus require internalization. A few others do not require internalization and are membrane-acting by a cytolytic mechanism; these include bacterial α -hemolysin, streptolysin, and the equinatoxin of sea anemone (41). In one of the first Phase I trials of an immunotoxin, an antiCD22 MoAb Fab' fragment was conjugated to Ricin A and administered to 15 patients. The MTD was 75 mg/m² and there was a 38% partial response (42, 43). Of the total of 200 patients in 9 clinical trials, which examined a variety of ricin-based immunotoxins, there was only a 3% complete response and a 12% partial response (43). This mediocre success may be due in part to the high inherent immunogenicity of immunotoxins. For ricin-based conjugates, the dose limiting toxicity arises from the vascular leak syndrome, a condition characterized by extravasation of fluid into interstitial space. Clinical trials have also indicated that poorly vascularized tumors are not suitable for immunotoxin therapy (43), perhaps because of their high molecular weight. To be more penetrating and to be less immunogenic, immunotoxins and similar targeting molecules need to be made smaller (44).

Drug Immunoconjugates

Over the last several decades, a number of antitumor agents, including chlorambucil, methotrexate, daunomycin, and doxorubicin conjugated to tumor specific antibodies, have been investigated, with varying degrees of success in tumor drug delivery. The most extensively studied has been a doxorubicin-BR96 immunoconjugate (BMS-182248-1). BR96 is a chimeric MoAb specific for a Lewis antigen found on the surface of tumor cells. The immunoconjugate is formed using an acid-labile hydrazone linkage attached through the thiol groups of the MoAb, with 8 moles of doxorubicin/mole of MoAb. After rapid internalization into antigen-bearing cells, the conjugate is designed to release free doxorubicin from the MoAb hydrazone linkage in the acidic environment of the lysosome (45). When tested in mice with xenografted human lung, breast, and colon carcinomas, there was an 89% and 72% cure rate (tumor reduction to nondetectable levels) in the lung and colon models, respectively. In breast carcinoma xenograft, results were less spectacular, with 10% complete response and 60% partial response. Doxorubicin or MoAb alone gave <1% cures in any of the models (46, 47). In a rat lung carcinoma xenograft model, there was a 94% cure rate, even though, unlike

mice, the Lewis antigen is expressed in normal tissue of rats (46, 47).

Because of the encouraging *in vivo* results, clinical development of this immunoconjugate was initiated. The pH-rate profile was determined using a stability-indicating size-exclusion HPLC assay, and exhibited a maximum stability at pH 7.5. The predominant route of degradation was hydrolysis at the hydrazone linkage to release free doxorubicin, with aggregation being a secondary pathway. This was supported by ELISA assay that demonstrated no loss of conjugate after 1 week of storage at 2–8°C. However, because the stability even at this temperature was insufficient for clinical development, a lyophilized formulation was developed. Formulations that remained amorphous due to the use of sucrose or lactose as lyoprotectant showed the greatest stability, with a shelf life of the lyophilized product of more than 12 months at 2–8°C. (48). In a Phase II study, the BR96 Doxorubicin conjugate (BMS182248-1) was administered to 14 metastatic breast cancer patients. However, there was only 1 partial response (7%), in contrast to doxorubicin alone, which gave 44% response. The toxicity profile of the two regimens was markedly different, with the doxorubicin showing the usual cardiotoxicity and hematologic toxicity, whereas the immunoconjugate showed GI associated toxicity, similar to the Phase I studies. It was proposed that the lack of correlation of the Phase II trial with the preclinical *in vivo* data could be because of the presence of the Lewis antigen at sites in the GI tract. This may act as an "antigen sink," preventing targeting to the tumor tissues and instead exacerbate the GI toxicity (49). Colon and lung cancer models showed better clinical responses than the breast cancer models (46), suggesting that clinical trials in these cancers may be more promising than breast cancer.

Another promising immunoconjugate is CMA-676, which is a conjugate of an anti-CD33 MoAb and calicheamicin, an anticancer drug shown to be 1000-fold more potent than doxorubicin in animal models. A Phase II trial of 39 acute myeloid leukemia patients resulted in 2 patients with complete remission and 7 patients who showed temporary removal of leukemia cells from the blood. CMA-676 is now in pivotal clinical studies in a number of centers in North America and Europe (36).

Preliminary clinical studies with chlorambucil-anti-melanoma globulin conjugates in patients with disseminated diseases have also indicated improvement in patient survival (50). The evaluation of vindesine-anti-CEA antibody conjugates in patients with advanced metastatic cancer (and probably expressing CEA) has

demonstrated positive localization of the conjugate (51). In this study, 8 patients received escalating doses of antibody (1–42 mg) conjugated to 24–1800 μ g vindesine, and no toxicity or hypersensitivity was noticed in any patient (51). However in most instances, the tumor versus normal tissue distribution ratio of antibodies approximated 2:1, and it rarely demonstrated specificity leading to a more desirable ratio like 10:1 (52).

Some studies have demonstrated synergism in anti-tumor response with drug–antibody conjugate. For example, a clinical trial with bronchial carcinoma patients compared chemotherapy with immunochemotherapy. The latter group received chemotherapy immediately prior to the administration of antibodies against a resected portion of the primary tumor. The group receiving chemotherapy alone demonstrated a 60% recurrence rate along with a 41% death rate; however, the immunochemotherapy group demonstrated only a 25% recurrence rate along with a 16% death rate (53).

Bispecific MoAbs composed of anti-CD3 or anti-CD2 MoAb, chemically conjugated to antitumor antibody and coated on lymphokine-activated killer (LAK) cells, have been clinically investigated for the treatment of malignant glioma, lymphoma, and ovarian cancer with encouraging results. In a trial involving malignant glioma therapy, bispecific MoAb-coated LAK cells were injected intracranially following surgical removal of tumor and whole brain irradiation and/or chemotherapy. This resulted in 76% of the patients being tumor-free after 2 years, as opposed to 33% of the patients tumorfree with LAK cell treatment alone (54).

Because of the ability to achieve higher drug–antibody ratios, various water-soluble polymeric carriers have been examined as linking agents in immunoconjugates. For example, encouraging results have been demonstrated with daunomycin conjugated to MoAbs via dextran bridge in rats bearing AH66 hepatoma cells (55). Similarly, MMC was conjugated to an anti- α -fetoprotein MoAb via a human serum albumin carrier in a molar ratio of 30:1:1. Full antibody activity was retained, and the conjugate was 20-fold more cytotoxic than free MMC in vitro and was also more effective than free MMC in tumor-bearing mice (56). Poly(lysine) has also been successfully used as a carrier, e.g., for targeting methotrexate (57) and muramyl dipeptide (58). *N*-(2-hydroxypropyl)-methacrylamide copolymers have been extensively examined as carriers for tumor targeted drug delivery (59, 60). Enzymatically cleavable spacers (e.g., oligopeptides) have been incorporated into these conjugates to allow release of active chemotherapeutic agent (59, 61).

Immunoliposomes

Generally, the antigens expressed by tumor cells are not specific but are merely present in higher ratio than on the normal cells. Hence, systems such as immunoliposomes have been developed to exploit these opportunities, as they are expected to bind to a greater extent to high antigen density tumor cells than to low antigen density normal cells. In immunoliposomes, the number of antibody molecules per liposome can be varied by as much as two orders of magnitude (62). Using egg phosphatidylcholine, cholesterol, phosphatidylserine, and *N*-4-nitrobenzo-2-oxa-1-1,3-diazole phosphatidylethanolamine in molar ratio of 56:33:10:1, unilamellar liposomes with 12–55 antibody molecules per vesicle have been investigated for binding with RDM-4 lymphoma cells with varying antigen density. The increase in the valency of liposomes (i.e. number of antibody molecules per liposome) increased their binding with low as well as high antigen density cells, and thus the low valency immunoliposomes were found to allow better discrimination between target and normal cells (62). An additional advantage of immunoliposomes is that a relatively high drug loading can potentially be accommodated, with the result that a small number of antibody molecules conjugated to the surface of an immunoliposome can deliver many more drug molecules to the target than is otherwise possible. Once the drug is released into the target cell, no further transformation is needed, because the entrapment process does not involve any chemical modification of the drug.

Heath et al. have proposed the use of immunoliposomes for the intracellular delivery of compounds that intrinsically do not enter diseased cells. These compounds are cytotoxic if they are transported intracellularly. Methotrexate- γ -aspartate, a good example of this type of compound, has been encapsulated in liposomes composed of phosphatidylcholine, cholesterol, and 4-(*p*-maleimido-phenyl)-butyryl-phosphatidylethanolamine in a molar ratio of 10:10:1. The liposomes were conjugated to either specific (anti-K2Kk IgG2A) or nonspecific MoAbs (antisheep erythrocyte IgG2A). The binding of targeted liposomes was found to be six fold higher to L929 fibroblasts (which express H2Kk protein) than nontargeted liposomes, whereas their binding was comparable in a nonspecific BALB/c 3T6 cell lines. The growth inhibition studies using L929 fibroblasts demonstrated the IC50 of free drug, targeted immunoliposomes, and nontargeted immunoliposomes to be 0.68, 0.066, and 1.2 μ M, respectively (Fig. 4). Hence, the targeted immunoliposomes appeared to be 10 times more effective

than free drug and 18 times more effective than nontargeted immunoliposomes, whereas targeted liposomes actually had the least efficacy in the nonspecific BALB/c 3T6 fibroblasts (63).

Extensive work is being pursued to assess the potential of immunoliposomes for the targeted drug delivery to CD4 positive cells in patients with HIV infection. The HIV infected cells possess CD4, which can be targeted by conjugating anti-Leu3A (CD4) MoAbs onto the surface of drug-loaded liposomes. Preliminary studies showed that immunoliposomes possessing surface-bound anti-Leu3A may be used to target antiviral agents to cells at risk from HIV infection (64). Cell adhesion molecules are glycoproteins expressed on cell surfaces during pathological inflammatory states such as rheumatoid arthritis, atopic dermatitis, and asthma; thus they also provide an opportunity for targeting. An F10.2 antibody against the cell adhesion molecule ICAM-1 was conjugated to liposomes; the immunoliposomes bound to human bronchial epithelial cells in a specific, dose- and time-dependent manner, correlating to the degree of ICAM-1 expression. Immunoliposomes of this type therefore have potential for targeted drug delivery in inflammatory disease states (65).

Heat-sensitive immunoliposomes have also been evaluated for the feasibility of drug delivery (66). These liposomes release the entrapped drug at temperatures above the phase transition temperature of the lipid(s). In vitro cell culture studies based on dipalmitoyl phosphatidylcholine liposomes with entrapped ^3H -uridine have demonstrated enhanced intracellular delivery of drug as compared to that observed with free drug and liposomes without MoAbs. Similarly, selection of appropriate lipids can also allow synthesis of pH-sensitive liposomes. Inclusion of target-cell specific immunogenic moieties in these colloidal particles leads to preparation of pH-sensitive immunoliposomes. Huang and co-workers have used 8:2 molar ratios of dioleoylphosphatidyl ethanolamine and oleic acid to develop pH-sensitive liposomes (67). Arabinofuranosylcytosine (ara-C) and methotrexate were encapsulated in the liposomes that were rendered immunospecific against L-929 cells by homing specific MoAbs. Compared to free drug, drug encapsulated in antibody-free liposomes and pH-insensitive immunoliposomes, the drug-encapsulated pH-sensitive immunoliposomes were found to significantly enhance the cytotoxic activity. Pretreatment of target cells with excess free MoAbs or placebo immunoliposomes was found to block the cytotoxic effect of the drug-loaded pH-sensitive immunoliposomes. In addition, it was shown that the drug release from these specific carrier particles occurs in cell endosomes (67).

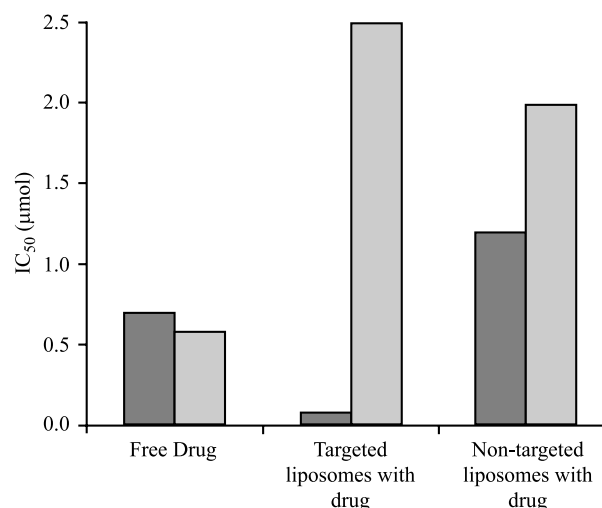


Fig. 4 Effect of treatment type on in vitro growth inhibition of (■) specific cell-line, L929 expressing H2Kk protein, and (□) nonspecific cell-line, BALB/c3T6. The targeted and nontargeted liposomes were conjugated with "anti-H2Kk IgG2A" and "anti-sheep erythrocyte IgG2A," respectively. (Adapted in part from Ref. 63.)

Because of the complexities involved with distribution, uptake, and pharmacological effects of targeted drug delivery systems, in vitro results do not always adequately predict the efficacy of a proposed MoAb-targeted system such as an immunoliposome. In fact, there are only a few in vivo studies of immunoliposomes that clearly demonstrate the promise shown by in vitro studies. One encouraging example is a study by Onuma et al. that compared the in vivo efficacy of doxorubicin loaded immunoliposomes against free drug in cows. MoAb c143 against the antigen expressed by bovine leukemia cells was conjugated to liposomes containing doxorubicin. Two groups of antigen-positive cows received four i.v. injections of either 0.4 mg/kg free drug or an equivalent dose of drug via immunoliposomes at an interval of 4 days. Whereas the two animals receiving free drug demonstrated only a slight decrease in their antigen positive cells, the three animals receiving drug-immunoliposomes gradually became free of antigen-positive cells, and 2 of these animals became antigen-negative in 6- and 14-week periods after treatment, respectively (68). In another study, MoAbs against tumor-associated antigens expressed on bovine leukemia cells were conjugated with liposomes containing doxorubicin, and the formulation was administered intravenously to BALB/c nude mice inoculated with BLSC-KU cells on day 0, 3, and 7 after the initiation of treatment. The results were

compared with untreated animals and the animals receiving doxorubicin liposomes conjugated to normal mouse IgG. The doxorubicin liposomes bearing target antigen specific antibodies significantly suppressed the tumor growth. The increase in tumor volume with this treatment, over 10 days after the initiation of therapy, was only 27% as opposed to 166% and 750% in the animals receiving nonspecific therapy and no therapy, respectively (see Fig. 5) (69). Histological screening of the tumors from animals receiving drug-loaded liposomes with specific antibody demonstrated scattered focal necrosis and marked proliferation of macrophages; however, in the untreated animals, active proliferation of tumor cells was observed with little involvement of macrophages (69).

A great deal of attention has been paid in recent years to long-circulating (also called sterically stabilized or “stealth”) liposomes, in which polyethylene glycol (PEG) molecules have been grafted to the surface of the liposomes by covalent attachment of PEG to liposomal phospholipids (specifically phosphatidylethanolamine). These long-circulating liposomes have been shown to avoid the rapid uptake by the reticuloendothelial system (RES), which normally plagues conventional liposomes without PEG; the circulating half-life can be increased by an order of magnitude (70, 71). In fact, the presence of a MoAb on the surface of a conventional immunoliposome may actually increase the uptake by the RES system (72), suggesting that modification to insure long-circulation may be especially important for immunoliposomes. When designing PEG-modified immunoliposomes, the composition must be optimized for both antigen binding and extended circulating lifetimes. Antigen recognition by the liposomal antibody can be sterically hindered by the presence of the PEG. This can be overcome by either reducing the polymer size to 2000, or by moving the antibody out to the terminus of the PEG rather than the liposome surface (72).

Despite the extensive *in vitro* and *in vivo* research on immunoliposomes, these MoAb targeted systems have apparently not yet reached clinical trials. This may be because of a variety of factors, including the difficulty of clearly demonstrating efficacy in suitable animal models, and the obstacles associated with scaleup and manufacture of system as complex as a sterically stabilized MoAb-targeted liposome. Other potential problems of immunoliposomes are that they may not adequately penetrate the vasculature of solid tumors; they may not adequately release the loaded drug into the target cells; and they may demonstrate immunogenicity (73). Use of humanized antibodies may alleviate the latter effect to some extent. Recently, sterically stabilized liposomes

conjugated to Fab' fragments of a humanized anti-HER2 MoAb (similar to Herceptin) were studied *in vitro* using confocal microscopy techniques. The immunoliposomes bound selectively and were internalized by HER2-over-expressing breast cancer cells, reaching 8000–23,000 vesicles/cell at saturating liposome concentrations, which was at least two orders of magnitude greater than cells with low HER2 expression (74). These immunoliposomes, containing doxorubicin and optimized with respect to Fab'/lipid/PEG composition for intracellular tumor delivery, were subsequently examined in an *in vivo* human xenograft breast cancer model; they demonstrated significantly increased antitumor efficacy compared to free doxorubicin or non targeted doxorubicin liposomes, and less systemic toxicity than free doxorubicin (75). The studies demonstrate the importance of optimizing a delivery system with respect to binding to the target epitope as well as uptake and/or release of available drug at the target site. Such considerations are necessary before successful demonstration of efficacy of immunoliposomes in the clinic. An ideal immunoliposome system should allow efficient encapsulation of intended compound so as to protect its degradation prior to and during endothelial transfer, and hence minimize inherent toxicity; it should also allow controlled release of drug

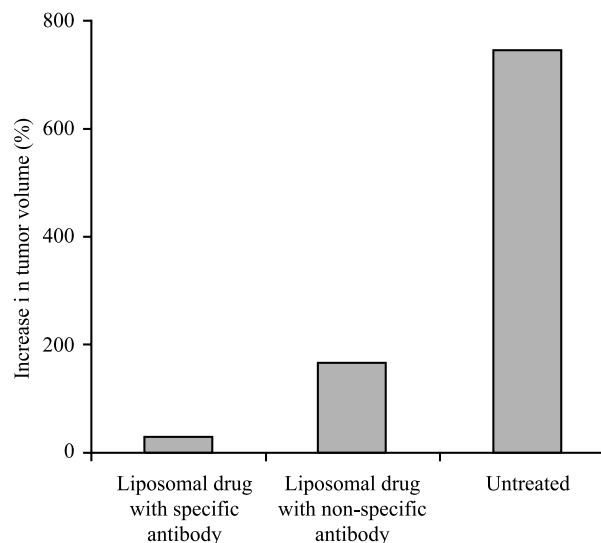


Fig. 5 Effect of treatment type on the increase in tumor volume in mice inoculated with BLSC-KU cells. The animals in treatment groups received *i.v.* doses of doxorubicin via tumor-specific or nonspecific liposomes on day 0, 3, and 7 after the initiation of treatment. The tumor-specific and nonspecific liposomes were conjugated with MoAbs against leukemia cells and normal mouse IgG, respectively. (Adapted in part from Ref. 69.)

in the extravascular compartment of target tissue. In this regard, it should be noted that the delivery systems based on particulate carriers may allow reversal of tumor cell drug resistance (Fig. 6) (76). Nevertheless, immunoliposomes fall short of meeting the above ideal criteria and much work remains to be done before they are clinically useful.

A related approach for lipophilic drugs is MoAb targeted emulsions. For example, a lung-targeted MoAb 34A was conjugated to the surface of a long-circulating emulsion composed of castor oil, phosphatidylcholine, and pegylated phosphatidylethanolamine. Upon intravenous injection into mice, 30% of the injected emulsion dose became preferentially associated with lung tissue within 30 min (77). A similar long-circulating emulsion conjugated to an anti-B-cell lymphoma MoAb LL2 was found to bind *in vitro* to three different Burkitt's lymphoma cell lines, and thus shows potential for delivery of anticancer drugs to B-cell malignancies (78).

Immunomicrospheres

In view of the availability of a wide variety of biocompatible and biodegradable polymers, and the ease of preparation of stable microparticles with predictable physicochemical characteristics, antibodies have been conjugated to polymeric microparticles for controlling their *in vivo* deposition.

Although a few *in vitro* studies have demonstrated promising results with immunomicrospheres (79, 80), limited information has been published on the *in vivo* efficacy of immunomicrospheres for drug delivery. In one case, following promising *in vitro* results, an *in vivo* study was conducted in mice bearing human tumor xenografts, using ^{14}C -polyhexylcyanoacrylate nanoparticles with adsorbed anti-osteogenic sarcoma MoAbs 971T/36. However, the particles were found to deposit predominantly in liver and spleen, and hence the study failed to demonstrate any appreciable improvements in drug delivery due to the immunocarrier (81). Lack of optimal particle size and/or tumor tissue permeability, lack of expression of sufficient Fab portions on the surface of particles, particle opsonization leading to a secondary non-interactive coating, distribution of specific antigens in the liver, and competitive displacement of the adsorbed MoAbs by serum components were suggested as possible reasons for this undesirable *in vivo* distribution of the immunoparticles. Another study has evaluated the *in vivo* drug delivery potential of albumin immunomicrospheres in mice (80). The microspheres bearing Lewis lung carcinoma MoAbs demonstrated slightly higher localization in lung carcinoma at 24 h after its administration.

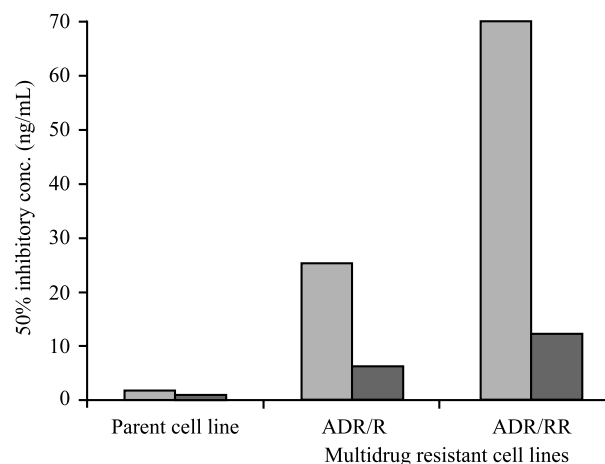


Fig. 6 A plot showing 50% inhibitory drug concentration *in vitro*, in normal and drug-resistant lines, following treatments with vinblastine solution (□), or vinblastine encapsulated in liposomes (■). The data suggest the possibility of the reversal of drug resistance by using carrier systems like liposomes. (Adapted in part from Ref. 76.)

REGULATORY CONCERNS

While a regulatory agency's prime concern remains the safety and efficacy of a new product in its proposed use, the technological issues concerning the manufacture of monoclonal-based therapeutics are not ignored. Indeed, because antibody-based systems involve specific immune reactions for their response, and they are comprised of components derived from biological origin, it is obvious that the proposed regulatory guidelines for the clinical use of these products are likely to be extremely strict. The technology for the development of antibodies and antibody-based delivery systems is relatively new and has been continuously expanding over the last several years. More and more practical changes in manufacturing process of these systems can therefore be expected over the next several years. Although most products are expected to be handled on a case-by-case basis, the following general guidelines should apply to all antibody-based systems: exhaustive characterization of the origin of cell lines, characterization of production procedures, product purification and characterization, quality control, and validation of processes involved during production and testings. For example, successful approval and use of a MoAb product will mandate declaration of the source, name and characteristics of the parent myeloma cell lines, and all pertinent details regarding the animal

species used for hybridization. The rationale for selecting a particular cell line along with criterion for its acceptance, the genotype and husbandry of animals used for in vivo production, and steps taken to control contamination is expected to play a critical role. Once the product is made, extensive purification to reduce the level of contamination (using techniques like affinity, size exclusion or ion-exchange chromatography and/or ultracentrifugation) is likely to increase the probability of its approval. Measures would need to be undertaken to insure that the product does not contain any biological contaminant transferred from the original malignant hybridoma cell lines.

Once a bulk lot is in hand, its characterization for immunoglobulin and subimmunoglobulin class, and testings for potential aggregation, denaturation, fragmentation of immunoglobulin and immunologic specificity would be required. If a MoAb fragment is used, its degree of homogeneity would need to be confirmed. Finally, information on the sterility and polynucleotide contamination of the lot is likely to be required. It would be expected that the process validation allows rejection of lots with viral or nucleic acid contaminants. Additional tests may include determination of the product stability with respect to fragmentation, aggregation, and loss of potency. The preclinical toxicity testing with the final product would be required in at least one species bearing relevant antigen. Following the identification of an appropriate animal model, GLP-compliant pharmacokinetic evaluations involving in vivo distribution, metabolism and excretion, would be desired.

As mentioned earlier, in most instances the antigens are preferentially associated with the target site rather than specifically present there, i.e., small amounts of the same antigens are present in one or more nontarget organs. Because the probability of unacceptable levels of this cross-reactivity is reasonably high, regulatory guidelines recommend screening for cross-reactivity. For in vitro screening, blood cells, cell culture lines, fluorescent antibody tests, radioaudiography, and/or similar other techniques would be useful. If possible, tissues from unrelated human donors could be used to screen phenotypic expressions of potentially cross-reactive tissue antigens. If these tests demonstrate positive cross-reactivity, extensive in vivo testing in animals sharing similar phenomenon would be required to determine its frequency as well as intensity. Alternatively, an isolated perfused human organ system could be used. If these choices are not available, limited clinical testings may be advisable with particular emphasis on the quantitation of the product's biodistribution over a period of time.

It should be realized that the above guidelines, generally meant to assess and regulate the antibody or antibody-component of an overall product, would need to be expanded and/or modified according to the characteristics of the final product. For example, in antibody-based drug delivery systems, the effect of drug and antibody on the potency and biological activity of the final system would need to be assessed.

CONCLUSIONS

Problems and Possible Solutions

Despite the promise of MoAb-directed drug delivery, there are still a multitude of problems that need to be worked out before the technology makes a large impact on therapy. A MoAb is often not as specific in vivo as would be predicted from in vitro studies; i.e., tumor antibodies may bind to normal cells as well as target cells. Despite the fact that antigens associated with tumor tissue have been identified, antigens are rarely specific enough to allow quantitative drug targeting. For example, CA 19-9, BW 494, and DU-PAN-2 have been identified as pancreatic tumor associated antigens. However, MoAb based therapy of pancreatic tumors has not been encouraging (82). In some cases, peak drug concentrations with MoAbs, in tumor tissue, have been found to be only 2–3 times higher than the surrounding normal tissues (83). Current literature suggests that the availability of high affinity MoAbs, which recognize specific antigens without cross-interaction with normal cells, is still scarce. The only exception to this observation is the surface immunoglobulin idiotype expressed by certain B-cell lymphomas (84).

The lack of genetic stability of antigens on tumor cell surfaces is another cause of low density of target antigen on the tumor cell. Antigenic modulation may result in nontumor specific antigen-antibody reaction, thus reversing the efficacy anticipated from the delivery system. Situations of low antigen density may readily saturate the MoAb-antigen binding. Quantitative evaluation of the localization of MoAbs in tumor tissue, at doses $<100\text{ }\mu\text{g}$, have demonstrated a direct correlation between tumor mass and quantity of antibody localized, and at 2–3 days after administration only 8% of the dose could be detected in the tumor. However, the administration of larger doses of MoAbs have been shown to reduce the fraction localized in the tumor, with 1–2 mg doses almost saturating the tumor (83). The presence of circulating tumor-associated antigens is

another factor that may decrease the overall efficacy of MoAb-directed delivery systems and complicate their evaluation. For example, the presence of circulating carcinoembryonic antigen has been shown to complicate the application of MoAbs against this antigen (85). In view of these problems, MoAb-directed delivery systems may ultimately be restricted to those few cases in which there are relatively high densities of known antigens in all cells of the target site.

The heterogeneity of tumor cells is another problem in targeting; i.e., a specific antigen may not be present in sufficient quantities in all cells of the target tissue to allow selection of suitable antibody. For example, it is now appreciated that multiple metastatic proliferation in a given host, and perhaps even in the same organ, can give rise to malignant tumors that contain heterogeneous subpopulation of cells with diverse biological characteristics, such as growth rate, antigenicity or immunogenicity, cell-surface receptors, response to individual and combined chemotherapeutic and immunological agents, invasiveness, and their overall metastatic potential (86). Trubetskoy et al. have proposed a method for MoAb-based drug delivery to target areas with heterogeneous antigens. The proposed method requires sequential administration of a mixture of modified antibodies against different antigens in the target area followed by administration of drug-carrier that recognizes and interacts with accumulated antibodies (Fig. 7) (87). The practical feasibility of this strategy was confirmed following administration of a mixture of biotinylated antibodies to target components followed by administration of biotinylated and avidin bearing liposomes. The binding of biotinylated liposomes via avidin was found to be higher than that achieved with liposomes bearing single antibody (87).

Solid tumors present special problems due to their frequent lack of vasculature; there is generally poor penetration of MoAbs, their fragments, and drug- or toxin-conjugates into solid tumor tissue. Because of the relatively intact microvascular barrier, and hence difficulties in carrier extravasation, the accumulation and uptake of immunoliposomes by solid tumor tissue is also generally low (88). On the other hand, the natural existence of increased transvascular permeability favors the use of MoAb-drug conjugates for the treatment of general lymphomas and leukemias.

The uptake of MoAb-based delivery systems by the reticulo-endothelial system (RES) is another drawback to these systems. It has been suggested that only 0.1–1% of the administered dose of antibody-based systems reaches nonRES sites, with ~8% dose reaching

non RES sites under optimal situations (89). However, a study comparing the *in vivo* tumor localization of anti-CEA MoAbs, and their F(ab')₂ and Fab' fragments, to human colon carcinoma grafts in nude mice has demonstrated greater tumor uptake of F(ab')₂ and Fab' fragments than the intact MoAbs (90), due to the smaller molecular size of the former. Often multiple intravenous injections, over a course of weeks, have been found to be more effective than single injections (91), and continuous infusion is more efficient than bolus regimen (43).

As mentioned earlier, the immunogenicity of "foreign" MoAbs has always been a major factor in the lack of success of MoAb-based therapeutic systems. It has been suggested that use of (Fab')₂ fragments may improve drug delivery without sacrificing the specificity of antigen-MoAb binding because elimination of Fc portion would likely reduce immunogenicity and nonspecific binding to normal cells (92). Humanized antibodies probably hold the greatest promise in decreasing the immune response of MoAb-based therapies. While immunotherapy with these entities is promising and indicates greatly decreased immune responses, little work has been reported on humanized immunoconjugates. These entities may still be immunogenic due to the non-antibody portion of the conjugate. For example, rats injected with an unconjugated murine MoAb failed to elicit an antibody response, whereas rats injected with a MoAb-Vinca alkaloid conjugate mounted a strong antibody response directed against the linker portion of the conjugate (93).

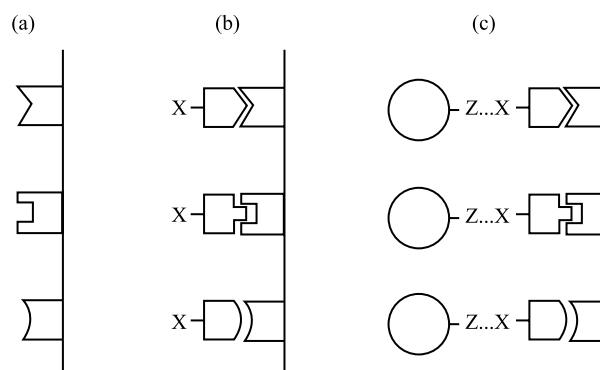


Fig. 7 A schematic representation of the unification of delivery systems to optimize therapeutic outcome with MoAbs: a) exposed target antigens; b) initial treatment with bridge molecules; c) specific binding of unified carrier systems. (Adapted in part from Ref. 87.)

Future Prospects

New applications in the field of antibody-directed drug delivery may be developed by combining the technology with another form of targeting or other means of optimization. For example, in photodynamic therapy a photosensitizing drug (e.g., a porphyrin, chlorin, purpurin, or phthalocyanine) is localized in a tumor, which is then irradiated to effect cell-killing (94). Several reports have described the use of MoAbs to further increase the localization. For example, Sn (IV) chlorin e6 was linked to the oligosaccharide moiety of an antimelanoma MoAb via a dextran carrier. In vitro studies indicated that phototoxicity was relatively specific for cells that exhibited the target antigen (95). Similarly, a chlorin derivative, meso chlorin e₆ mono(*N*-2-aminoethylamide), was linked via a tetrapeptide linker to an antibody directed against ovarian cancer cells. Targeted conjugates were taken up rapidly by cells and detected within the lysosomes, and the conjugate had higher photodynamic effects on ovarian carcinoma cells than nontargeted conjugates (IC₅₀ of 0.38, 290, and 0.34 μ M for MoAb-HPMA-e₆, HPMA-e₆, and free e₆, respectively) (96).

Another recent approach combines MoAb targeting with enzymatic prodrug activation. In this therapeutic method, called antibody-directed prodrug therapy (ADEPT), an enzyme-antibody conjugate is administered and allowed to accumulate in the target site (e.g., tumor). A latent, nontoxic prodrug is then injected, which on contact with the enzyme is converted into the active parent drug and subsequently kills the tumor cells (97). For example, a glutamic acid derivative of benzoic acid mustard was administered to choriocarcinoma-bearing mice, followed by a carboxypeptidase-antibody conjugate that cleaved glutamic acid from the active drug. Tumor contained the highest concentration of targeted enzyme conjugate, and was the only site in which all prodrug reaching the site was activated (98). The ADEPT technique has been tested clinically in colorectal cancer patients using para-*N*-(mono-2-chloroethyl monomethyl)-aminobenzoyl glutamic acid as the prodrug and an antibody conjugate of glutamate hydrolase as the activating enzyme, with temporary regression of disease in two out of five patients (99).

Despite the problems described in earlier sections, MoAbs should hold an important place in drug delivery and therapy in the future. Although the number of therapeutic applications that will eventually lend themselves to this technology may be small, the problems should not be insurmountable for these applications and may yield important advantages over

other therapies. Proper attention to detail must be taken in the choice of antibody, coupling method, drug, route of administration, dose, and other factors in order to design an effective therapy for a particular disease; possible mechanisms of distribution, uptake, metabolism, and pharmacological effect must be properly understood to develop a rationale for a particular MoAb directed therapy. Humanized antibodies hold great promise to alleviate the immune response encountered in past clinical trials using murine derived antibodies. Many of the problems in the scaleup of MoAb manufacture have been solved, primarily because of the rapid growth of diagnostic applications of MoAbs and the coming to market of therapeutic MoAb's. Much work remains to be done, however, in the scaleup of immunoconjugates and complex systems such as immunoliposomes and immunomicrospheres. Nevertheless, it is likely that the next decade will see a number of MoAb-directed therapies reach extended clinical trials and perhaps come to the market.

REFERENCES

1. Ram, B.P.; Tyle, P. Immunoconjugates: Applications in Targeted Drug Delivery for Cancer Therapy. *Pharm. Res.* **1987**, *4*, 181-188.
2. Bodey, B.; Siegel, S.E.; Kaiser, H.E. Human Cancer Detection and Immunotherapy with Conjugated and Non-Conjugated Monoclonal Antibodies. *Anticancer Res.* **1996**, *16*, 661-674.
3. Panchagula, R.; Dey, C.S. Monoclonal Antibodies in Drug Targeting. *J. Clin. Pharm. Ther.* **1997**, *22*, 7-19.
4. Buske, C.; Feuring-Buske, M.; Unterhalt, M.; Hiddeman, W. Monoclonal Antibody Therapy for B-Cell Non-Hodgkins's Lymphomas. *Eur. J. Cancer* **1999**, *35*, 549-557.
5. Multani, P.S.; Grossbard, M.L. Monoclonal Antibody-Based Therapies for Hematologic Malignancies. *J. Clin. Oncol.* **1998**, *16*, 3691-3710.
6. Devlin, T.M. *Textbook of Biochemistry with Clinical Correlations*; John Wiley & Sons: New York, 1982; 110.
7. Delmonico, F.L.; Cosimi, A.B. Monoclonal Antibody Treatment of Human Allograft Recipients. *Surg. Gynecol. Obstet.* **1988**, *166*, 89-98.
8. Myers, K.J.; Ron, Y. Current Approaches and Obstacles to Immunotherapy. *Pharm. Technol.* **1992**, *16*, 26-36.
9. Larson, S.M.; Carrasquillo, J.A.; Krohn, K.A.; Brown, J.P.; McGuffin, R.W.; Ferens, J.M.; Hill, L.D.; Beaumier, P.L.; Hellstrom, P.L.; Hellstrom, I. Localization of ¹³¹I-Labeled P97-Specific Fab Fragments in Human Melanoma as a Basis for Radiotherapy. *J. Clin. Invest.* **1983**, *72*, 2101-2107.

10. Ball, E.D.; Guyre, P.M.; Mills, L.; Fisher, J.; Dinces, N.B.; Fanger, M.W. Initial Trial of Bispecific Antibody-Mediated Immunotherapy of CD15-Bearing Tumors: Cytotoxicity of Human Tumor Cells Using a Bispecific Antibody Comprised of Anti-CD15 (MoAb PM81) and Anti-CD64/Fc Gamma RI (MoAb 32). *J. Hematother.* **1992**, *1*, 85–94.
11. Goding, J.W. *Monoclonal Antibodies: Principles and Practice*, 3rd Ed.; Academic Press: London, 1996; 153–160.
12. Kettleborough, C.A.; Saldanha, J.; Heath, V.J.; Morrison, C.J.; Bendig, M.M. Humanization of a Mouse Monoclonal Antibody by CDR Grafting. *Protein Eng.* **1991**, *4*, 773–783.
13. Randerson, D.H. Large-Scale Cultivation of Hybridoma Cells. *J. Biotechnol.* **1985**, *2*, 241–255.
14. Shepherd, P.; Dean, C. *Monoclonal Antibodies: A Practical Approach*; Oxford University Press: Oxford, 2000; 125–180.
15. King, D.J. *Applications and Engineering of Monoclonal Antibodies*; Taylor and Francis: London, 1998; 161–185.
16. Page, M.; Thibeault, D.; Noel, C.; Dumas, L. Coupling a Preactivated Daunorubicin Derivative to Antibody: A New Approach. *Anticancer Res.* **1990**, *10*, 353–357.
17. Arano, Y.; Matsushima, H.; Tagawa, M.; Koizumi, M.; Endo, K.; Konishi, J.; Yokoyama, A. A Novel Bifunctional Metabolizable Linker for the Conjugation of Antibodies with Radionuclides. *Bioconjugate Chem.* **1991**, *2*, 71–76.
18. Noguchi, A.; Takahashi, T.; Yamaguchi, T.; Kitamura, K.; Takakura, Y.; Hashida, M.; Sezaki, H. Preparation and Properties of the Immunoconjugate Composed of Anti-Human Colon Cancer Monoclonal Antibody and Mitomycin C-Dextran Conjugate. *Bioconjugate Chem.* **1992**, *3*, 132–137.
19. Zara, J.J.; Wood, R.D.; Boon, P.; Kim, C.-H.; Pomato, N.; Bredehorst, R.; Vogel, C.-W. A Carbohydrate-Directed Heterobifunctional Cross-Linking Reagent for the Synthesis of Immunoconjugates. *Anal. Biochem.* **1991**, *194*, 156–162.
20. Kaneko, T.; Willner, D.; Monkovic, I.; Knipe, J.O.; Braslawsky, G.R.; Greenfield, R.S.; Vyas, D.M. New Hydrazone Derivatives of Adriamycin and their Immunoconjugates—A Correlation Between Acid Stability and Cytotoxicity. *Bioconjugate Chem.* **1991**, *2*, 133–141.
21. Lavie, E.; Hirschberg, D.L.; Schreiber, G.; Thor, K.; Hill, L.; Hellstrom, I.; Hellstrom, K.-E. Monoclonal Antibody L6-Daunomycin Conjugates Constructed to Release Free Drug at the Lower pH of Tumor Tissue. *Cancer Immunol. Immunother.* **1991**, *33*, 223–230.
22. Greenfield, L.; Bloch, W.; Moreland, M. Thiol Containing Cross-Linking Agent with Enhanced Steric Hindrance. *Bioconjugate Chem.* **1990**, *1*, 400–410.
23. Goff, D.A.; Carroll, S.F. Substituted 2-Iminoethanolanes: Reagents for the Preparation of Disulfide Cross-Linked Conjugates with Increased Stability. *Bioconjugate Chem.* **1990**, *1*, 381–386.
24. Rowland, G.F.; O'Neill, G.J.; Davies, A.A. Suppression of Tumour Growth in Mice by a Drug-Antibody Conjugate Using a Novel Approach Linkage. *Nature* **1975**, *255* (5508), 487–488.
25. Rybak, S.M.; Hoogenboom, H.R.; Meade, H.M.; Raus, J.C.M.; Schwartz, D.; Youle, R.J. Humanization of Immunotoxins. *Proc. Nat. Acad. Sci. USA* **1992**, *89*, 3165–3169.
26. Vandamme, A.; Dewerchin, M.; Lijnen, R.; Bernar, H.; Bulens, F.; Nelles, L.; Collen, D. Characterization of a Recombinant Chimeric Plasminogen Activator Composed of a Fibrin Fragment-D-Dimer-Specific Humanized Monoclonal Antibody and a Truncated Single-Chain Urokinase. *Eur. J. Biochem.* **1992**, *205*, 139–146.
27. Weiner, L.M. An Overview of Monoclonal Antibody Therapy of Cancer. *Semin. Oncol.* **1999**, *26* (4 Suppl. 12), 41–50.
28. Taetle, R.; Honeysett, J.M.; Trowbridge, I. Effects of Anti-Transferrin Receptor Antibodies on Growth of Normal and Malignant Myeloid Cells. *Int. J. Cancer* **1983**, *32*, 343–349.
29. *Physician's Desk Reference*, 53rd Ed.; Rituxan, Medical Economics Co.: Montvale, NJ, 1999; 1070–1072–1384–1386.
30. Cobleigh, M.A.; Vogel, C.L.; Tripathy, D.; Robert, N.J.; Scholl, S.; Fehrenbacher, L.; Wolter, J.M.; Paton, V.; Shak, S.; Lieberman, G.; Slamon, D.J. Multinational Study of the Efficacy and Safety of Humanized Anti-HER2 Monoclonal Antibody in Women Who Have HER2-Over-Expressing Metastatic Breast Cancer That has Progressed after Chemotherapy for Metastatic Disease. *J. Clin. Oncol.* **1999**, *17*, 2639–2648.
31. Colbern, G.T.; Hiller, A.J.; Musterer, R.S.; Working, P.K.; Henderson, I.C. Antitumor Activity of Herceptin in Combination with Stealth Liposomal Cisplatin or Nonliposomal Cisplatin in a HER2 Positive Human Breast Cancer Model. *J. Inorg. Biochem.* **1999**, *77*, 117–120.
32. Mittelman, A.; Chen, Z.J.; Kageshita, T.; Yang, H.; Yamada, M.; Baskind, P.; Goldbexy, N.; Puccio, C.; Ahmed, T.; Arlin, Z. et al. Active Specific Immunotherapy in Patients with Melanoma. *J. Clin. Invest.* **1990**, *86*, 2136–2144.
33. Magerstadt, M. *Antibody Conjugates and Malignant Disease*; CRC Press: Boca Raton, FL, 1991; 17–36, 79–110.
34. Preparation and Use of DTPA-Coupled Antitumor Antibodies Radiolabelled with Yttrium-90. In *Antibody-Mediated Delivery Systems*; Hnatowich, D.J., Snook, D., Rowlinson, G., Stewart, S., Epentos, A.A., Eds.; Marcel Dekker, Inc.: New York, 1988; 353–374.
35. Corcoran, M.C.; Eary, J.; Bernstein, I.; Press, O.W. Radioimmunotherapy Strategies for Non-Hodgkin's Lymphomas. *An. Oncol.* **1997**, *8* (Suppl. 1), 133–138.

36. Panousis, C.; Pietersz, G.A. Monoclonal Antibody-Directed Cytotoxic Therapy: Potential in Malignant Diseases of Aging. *Drugs & Aging*. **1999**, *1*, 1–15.
37. Press, O.W.; Eary, J.F.; Appelbaum, F.R.; Martin, P.J.; Badger, C.C.; Nelp, W.B.; Glenn, S.; Butchko, G.; Fisher, D.; Porter, B. Radiolabeled-Antibody Therapy of B-Cell Lymphoma with Autologous Bone Marrow Support. *New Engl. J. Med.* **1993**, *329*, 1219–1224.
38. Press, O.W. Radiolabeled Antibody Therapy of B-Cell Lymphomas. *Semina. Oncol.* **1999**, *26* (5 Suppl 14), 58–65.
39. Poste, G.; Kirsh, R. Site-Specific (Targeted) Drug Delivery in Cancer Therapy. *Biotechnology* **1983**, *1*, 869–878.
40. Alving, C.R. Delivery of Liposome-Encapsulated Drugs to Macrophages. *Pharm. Ther.* **1983**, *22*, 407–412.
41. Panchal, R.G. Novel Therapeutic Strategies to Selectively Kill Cancer Cells. *Biochem. Pharmacol.* **1998**, *55*, 247–252.
42. Vitetta, E.S.; Stone, M.; Amlot, P.; Fay, J.; May, R.; Till, M.; Newman, J.; Clark, P.; Collins, R.; Cunningham, D. et al. A Phase I Immunotoxin Trial in Patients with B-Cell Lymphoma. *Cancer Res.* **1991**, *51*, 4052–4058.
43. Winkler, U.; Barth, S.; Schnell, R.; Diehl, V.; Engert, A. The Emerging Role of Immunotoxins in Leukemia and Lymphoma. *Ann. Oncol.* **1997**, *8* (Suppl 1), 139–146.
44. Kreitman, R.J.; Pastan, I. Immunotoxins for Targeted Cancer Therapy. *Adv. Drug Del. Rev.* **1998**, *31*, 53–88.
45. Willner, D.; Trail, P.A.; Hofstead, S.J.; King, H.D.; Lasch, S.J.; Braslawsky, G.R.; Greenfield, R.S.; Kaneko, T.; Firestone, R.A. (6-Maleimidocaproyl)Hydrazide of Doxorubicin—A New Derivative for the Preparation of Immunoconjugates of Doxorubicin. *Bioconjugate Chem.* **1993**, *4*, 521–524.
46. Trail, P.A.; Willner, D.; Lasch, S.J.; Henderson, A.J.; Hofstead, S.; Casazza, A.M.; Firestone, R.A.; Hellstrom, I.; Hellstrom, K.E. Cure of Xenografted Human Carcinomas by BR96-Doxorubicin Immunoconjugates. *Science* **1993**, *261* (5118), 212–215.
47. Sjogren, H.O.; Isaksson, M.; Willner, D.; Hellstrom, I.; Hellstrom, K.E.; Trail, P.A. Antitumor Activity of Carcinoma-Reactive BR96-Doxorubicin Conjugate against Human Carcinomas in Athymic Mice and Rats and Syngeneic Rat Carcinomas in Immunocompetent Rats. *Cancer Res.* **1997**, *57*, 4530–4536.
48. Barbour, N.P.; Paborji, M.; Alexander, T.C.; Coppola, W.P.; Bogardus, J.B. Stabilization of Chimeric BR96-Doxorubicin Immunoconjugate. *Pharm. Res.* **1995**, *12*, 215–222.
49. Tolcher, A.W.; Sugarman, S.; Gelmon, K.A.; Cohen, R.; Saleh, M.; Isaacs, C.; Young, L.; Healey, D.; Onetto, N.; Slichenmyer, W. Randomized Phase II Study of BR96-Doxorubicin Conjugate in Patients with Metastatic Breast Cancer. *J. Clin. Oncol.* **1999**, *17* (2), 478–484.
50. Ghose, T.; Norvell, S.T.; Guclu, A.; Bodurtha, A.; Tai, J.; MacDonald, A.S. Immunotherapy of Malignant Melanoma with Chlorambucil-Bound Antimelanoma Globulins: Preliminary Results in Patients with Disseminated Disease. *J. Natl. Cancer Inst.* **1977**, *58*, 845–852.
51. Ford, C.H.J.; Newman, C.E.; Johnson, J.R.; Woodhouse, C.S.; Reeder, T.A.; Rowland, G.F.; Simmonds, R.G. Localisation and Toxicity Study of a Vindesine-Anti-CEA Conjugate in Patients with Advanced Cancer. *Br. J. Cancer* **1983**, *47*, 35–42.
52. Sezaki, H.; Hashida, M. Macromolecule-Drug Conjugates in Targeted Cancer Chemotherapy. *CRC Crit. Rev. Ther. Drug Carrier Sys.* **1984**, *1*, 1–38.
53. Newman, C.E.; Ford, C.H.J.; Davies, D.A.L.; O'Neill, G.J. Antibody Drug Synergism an Assessment of Specific Passive Immuno Therapy in Bronchial Carcinoma. *Lancet* **1977**, *2*, 163–164.
54. Nitta, T.; Sato, K.; Yagita, H.; Okumura, K.; Ishii, S. Preliminary Trial of Specific Targeting Therapy Against Malignant Glioma. *Lancet* **1990**, *335*, 368–371.
55. Tsukada, Y.; Hurwitz, E.; Kashi, R.; Sela, M.; Hibi, N.; Hara, A.; Hirai, H. Chemotherapy by Intravenous Administration of Conjugates of Daunomycin with Monoclonal and Conventional Anti-Rat Alpha-Fetoprotein Antibodies. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 7896–7899.
56. Ohkawa, K.; Tsukada, Y.; Hibi, N.; Umemoto, N.; Hara, T. Selective In Vitro and In Vivo Growth Inhibition against Human Yolk Sac Tumor Cell Lines by Purified Antibody against Human Alpha-Fetoprotein Conjugated with Mitomycin C Via Human Serum Albumin. *Cancer Immunol. Immunother.* **1986**, *23*, 81–86.
57. Shen, W.-C.; Dug, X.; Feener, E.P.; Ryser, H.J. The Intracellular Release of Methotrexate from a Synthetic Drug Carrier System Targeted to Fc Receptor-Bearing Cells. *J. Contr. Rel.* **1989**, *10*, 89–96.
58. Midoux, P.; Martin, A.; Collet, B.; Monsigny, M.; Roche, A.-C.; Toujas, L. Activation of Mouse Macrophages by Muramyl Dipeptide Coupled with an Anti-Macrophage Monoclonal Antibody. *Bioconjugate Chem* **1992**, *3*, 194–199.
59. Seymour, L.W.; Flanagan, P.A.; Al-Shamkhani, A.; Subr, V.; Ulbrich, K.; Cassidy, J.; Duncan, R. Synthetic Polymers Conjugated to Monoclonal Antibodies: Vehicles for Tumor-Targeted Drug Delivery. *Sel. Cancer Thera.* **1991**, *7*, 59–73.
60. Rihova, B.; Kopecek, J. Biological Properties of Targetable Poly[N-(2-Hydroxypropyl)-Methacrylamide]-Antibody Conjugates. *J. Contr. Rel.* **1985**, *2*, 289–310.
61. Duncan, R.; Hume, I.C.; Kopeckova, P.; Ulbrich, K.; Strohalm, J.; Kopecek, J. Anticancer Agents Coupled To N-(2-Hydroxypropyl)-Methacrylamide Copolymers. *J. Contr. Rel.* **1989**, *10*, 51–63.
62. Houck, K.S.; Huang, L. The Role of Multivalency in Antibody Mediated Liposome Targeting. *Biochem. Biophys. Res. Commun.* **1987**, *145*, 1205–1210.
63. Heath, T.D.; Montgomery, J.A.; Piper, J.R.; Paphadjopoulos, D. Antibody-Targeted Liposomes: Increase in Specific

- Toxicity of Methotrexate-Gamma-Aspartate. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 1377–1381.
64. Phillips, N.C.; Tsoukas, C. Immunoliposome Targeting to CD4+ Cells in Human Blood. *Cancer Detect. Prev.* **1990**, *14*, 383–390.
 65. Bloemen, P.G.M. Adhesion Molecules: A New Target for Immunoliposome-Mediated Drug Delivery. *FEBS Letters* **1995**, *357*, 140–144.
 66. Sullivan, S.M.; Huang, L. Enhanced Delivery to Target Cells by Heat-Sensitive Immunoliposomes. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 6117–6121.
 67. Connor, J.; Huang, L. pH-Sensitive Immunoliposomes as an Efficient and Target-Specific Carrier for Antitumor Drugs. *Cancer Res.* **1986**, *46*, 3431–3435.
 68. Onuma, M.; Yasutomi, Y.; Yamamoto, M.; Watarai, S.; Yasuda, T.; Kawakami, Y. Antitumor Effect of Adriamycin Entrapped in Liposomes Conjugated with Anti-bovine Tumor Antigen Monoclonal Antibody in Leukemic Cows. *Zentralbl. Veterinarmed.* **1989**, *36*, 139–147.
 69. Onuma, M.; Odawara, T.; Watarai, S.; Aida, Y.; Ochiai, K.; Syuto, B.; Matsumoto, K.; Yashuda, T.; Fujimoto, Y.; Izawa, H.; Kawakami, Y. Antitumor Effect of Adriamycin Entrapped in Liposomes Conjugated with Monoclonal Antibody against Tumor-Associated Antigen of Bovine Leukemia Cells. *Jpn. J. Cancer Res.* **1986**, *77*, 1161–1167.
 70. Woodle, M.C. Sterically Stabilized Liposome Therapeutics. *Adv. Drug Del. Rev.* **1995**, *16*, 249–265.
 71. Allen, T.M. Long-Circulating (Sterically Stabilized) Liposomes for Targeted Drug Delivery. *Trends Pharmacol. Sci.* **1994**, *15*, 215–20.
 72. Maclean, A.L.; Symonds, G.; Ward, R. Immunoliposomes as Targeted Delivery Vehicles for Cancer Therapeutics. *Int. J. Oncol.* **1997**, *11*, 325–332.
 73. Mastrobattista, E.; Koning, G.A.; Storm, G. Immunoliposomes for the Targeted Delivery of Antitumor Drugs. *Adv. Drug Deliv. Reviews* **1999**, *40*, 103–127.
 74. Kirpotin, D.; Park, J.W.; Hong, K.; Zalipsky, S.; Li, W.L.; Carter, P.; Benz, C.C.; Papahadjopoulos, D. Sterically Stabilized Anti-HER2 Immunoliposomes: Design and Targeting to Human Breast Cancer Cells In Vitro. *Biochemistry* **1997**, *36* (1), 66–75.
 75. Park, J.W.; Hong, K.; Kirpotin, D.B.; Meyer, O.; Papahadjopoulos, D.; Benz, C.C. Anti-HER2 Immunoliposomes for Targeted Therapy of Human Tumors. *Cancer Lett.* **1997**, *118*, 153–160.
 76. Seid, C.A.; Fidler, I.J.; Clyne, R.K.; Earnest, L.E.; Fan, D. Overcoming Murine Tumor Cell Resistance to Vinblastine by Presentation of the Drug in Multilamellar Liposomes Consisting of Phosphatidylcholine and Phosphatidylserine. *Sel. Cancer Thera.* **1991**, *7*, 103–12.
 77. Young, K.; Liu, D.; Maruyama, K.; Takizawa, T. Antibody Mediated Lung Targeting of Long-Circulating Emulsions. *PDA J. of Pharm. Sci. Technol.* **1996**, *50*, 372–377.
 78. Lundberg, B.B.; Griffiths, G.; Hansen, H.J. Conjugation of an Anti-B-Cell Lymphoma Monoclonal Antibody, LL2, to Long-Circulating Drug-Carrier Lipid Emulsions. *J. Pharm. Pharmacol.* **1999**, *51*, 1099–1105.
 79. Lee, K.C.; Lee, J.Y.; Kim, W.B.; Cha, C.Y. Monoclonal Antibody-Based Targeting of Methotrexate-Loaded Microspheres. *Int. J. Pharm.* **1990**, *59*, 27–33.
 80. Akasaka, Y.; Ueda, H.; Takayama, K.; Machida, Y.; Nagai, T. Preparation and Evaluation of Bovine Serum Albumin Nanospheres Coated with Monoclonal Antibodies. *Drug Design Del.* **1988**, *3*, 85–97.
 81. Illum, L.; Jones, P.D.; Baldwin, R.W.; Davis, S.S. Tissue Distribution of Poly(Hexyl 2-Cyanoacrylate) Nanoparticles Coated with Monoclonal Antibodies in Mice Bearing Human Tumor Xenografts. *J. Pharmacol. Exp. Ther.* **1984**, *230*, 733–736.
 82. Bellet, D.; Bidart, J.M.; Rougier, P.; Bohuon, C. Use of Monoclonal Antibodies in the Treatment of Cancer of the Pancreas: Towards New Progress?. *Bull. Cancer (Paris)* **1990**, *77*, 283–238.
 83. Pimm, M.V.; Baldwin, R.W. Quantitative Evaluation of the Localization of a Monoclonal Antibody (791T/36) in Human Osteogenic Sarcoma Xenografts. *Eur. J. Clin. Oncol.* **1984**, *20*, 515–524.
 84. Thielemans, K.; Maloney, D.G.; Meeker, T.; Fujimoto, J.; Doss, C.; Warnke, R.A.; Bindl, J.; Gralow, J.; Miller, R.A.; Levy, R. Strategies for Production of Monoclonal Anti-Idiotypic Antibodies against Human B-Cell Lymphomas. *J. Immunol.* **1984**, *133*, 495–501.
 85. Mach, J.P.; Carrel, S.; Forni, M.; Ritschard, J.; Donath, A.; Alberto, P. Tumor Localization of Radiolabeled Antibodies against Carcinoembryonic Antigen in Patients with Carcinoma: A Critical Evaluation. *N. Eng. J. Med.* **1980**, *303*, 5–20.
 86. Fidler, I.J. The Evolution of the Metastatic Phenotype. *Proc. Am. Assoc. Cancer Res.* **1991**, *32*, 486–487.
 87. Trubetskoy, V.S.; Berdichevsky, V.R.; Efremov, E.E.; Torchilin, V.P. On the Possibility of the Unification of Drug Targeting Systems: Studies with Liposome Transport to the Mixtures of Target Antigens. *Biochem. Pharmacol.* **1987**, *36*, 839–842.
 88. Matzku, S.; Krempel, H.; Weckenmann, H.P.; Schirmacher, V.; Sinn, H.; Stricker, H. Tumour Targeting with Antibody-Coupled Liposomes: Failure to Achieve Accumulation in Xenografts and Spontaneous Liver Metastases. *Cancer Immunol. Immunother.* **1990**, *31*, 285–291.
 89. Ranney, D.F. Drug Targeting to the Lungs. *Biochem. Pharmacol.* **1986**, *35*, 1063–1069.
 90. Buchegger, F.; Haskell, C.M.; Schreyer, M.; Scazziga, B.R.; Randin, S.; Carrel, S.; Mach, J.-P. Radiolabeled Fragments of Monoclonal Antibodies against Carcinoembryonic Antigen for Localization of Human Colon Carcinoma Grafted into Nude Mice. *J. Exp. Med.* **1983**, *158*, 413–427.
 91. Sivam, G.; Pearson, J.W.; Bohn, W.; Oldham, R.K.; Sadoff, J.C.; Morgan, A.C. Immunoconjugates to a

- Human Melanoma-Associated Antigen. *Cancer Res.* **1987**, *47*, 3169.
92. Hurwitz, E.; Levy, R.; Maron, R.; Wilchek, M.; Arnon, R.; Sela, M. The Covalent Binding of Daunomycin and Adriamycin to Antibodies, with Retention of Both Drug and Antibody Activities. *Cancer Res.* **1975**, *35*, 1175–1181.
93. Johnson, D.A.; Barton, R.L.; Fix, D.V.; Scott, W.L.; Gutowski, M.C. Induction of Immunogenicity of Monoclonal Antibodies by Conjugation with Drugs. *Cancer Res.* **1991**, *51*, 5774–5776.
94. Dougherty, T.J. Photodynamic Therapy—New Approaches. *Semin. Surg. Oncol.* **1989**, *5*, 6–16.
95. Rakestraw, S.L.; Tompkins, R.G.; Yarmush, M.L. Antibody-Targeted Photolysis. In Vitro Studies with Sn(IV) Chlorine ⁶ Covalently Bound to Monoclonal Antibodies Using a Modified Dextran Carrier. *Proc. Nat. Acad. Sci. USA* **1990**, *87*, 4217–4221.
96. Omelyanenko, V.; Gentry, C.; Kopeckova, P.; Kopecek, J. HPMA Copolymer-Anticancer Drug-OV-TL16 Antibody Conjugates. II. Processing in Epithelial Ovarian Carcinoma Cells In Vitro. *Int. J. Cancer* **1998**, *75*, 600–608.
97. Bagshawe, K.D. Antibody-Directed Enzyme/Prodrug Therapy (ADEPT). *Biochem. Soc. Trans* **1990**, *18*, 750–752.
98. Antoniow, P.; Springer, C.J.; Bagshawe, K.D.; Searle, F.; Melton, R.G.; Rogers, G.T.; Burke, P.J.; Sherwood, R.F. Disposition of the Prodrug 4-(bis (2-Chloroethyl) Amino) Benzoyl-L-Glutamic Acid and its Active Parent Drug in Mice. *Br. J. Cancer* **1990**, *62*, 909–914.
99. Bagshawe, K.D.; Sharma, S.K.; Springer, C.J.; Antoniow, P.; Rogers, G.T.; Burke, P.J.; Melton, R.; Sherwood, R. Antibody-Enzyme Conjugates can Generate Cytotoxic Drugs from Inactive Precursors at Tumor Sites. *Antibody Immunoconj. Radiopharm.* **1991**, *4*, 915–922.

Molecular and Cellular Approaches to Lead Optimization in Pharmaceutical Development

C. K. Atterwill

M. G. Wing

Huntingdon Life Sciences, Cambridgeshire, United Kingdom

INTRODUCTION

Sensible lead candidate selection for preclinical development and registration should reduce the high rate of drug attrition, the cost of which increases with the distance that the failed drug candidate has proceeded down the research and development pipeline. It is estimated that “adverse” toxicokinetic and safety factors encountered during animal and human exposures in drug development account for approximately 60% of the reasons for termination during development (Fig. 1). Of this 60%, a combination of undesirable animal and human toxic events account for 21% of the total factors involved. Rates of attrition are high, especially during the early regulatory toxicology and clinical phases of testing. (Fig. 2). Not surprisingly, because of the large number of hits identified from primary high-throughput screens, lead candidate optimization has, therefore, emerged as a critical decision-making milestone from both ethical and commercial perspectives. The “models” or strategies for incorporation of a preclinical lead optimization stage are shown in Fig. 3. Essentially, the “old” or, in many cases, current, strategy is to take several lead candidate compounds through early small-scale clinical pharmacological/toxicokinetic/safety studies in vivo before a final selection is made following 1 month toxicology exposures on the lead development drug candidate for clinical exposure. Any later problems with the compound may then facilitate the design of a lead optimization “screen” at an earlier stage for possible follow-up compounds. A more modern and desirable approach would be (and is used in many cases now) to insert a mandatory in vitro ADME-Tox “screen” at the stage before the preliminary in vivo ADME-Tox exposures, which would then produce a “clean” lead development candidate (LDC) for full regulatory testing (Fig. 4). In short, there is a need for early rapid and robust screening assays which that will allow a lead series of compounds to be ranked for desirable or undesirable characteristics.

These preclinical lead optimization technologies (PLOT’S)^[1] must be sufficiently rapid to interface with high-throughput screens without creating a further pipe-

line bottleneck, be predictive of drug failure, and be highly cost-effective (Fig. 4). The assays which constituting the PLOT platform are typically in vitro systems, miniaturized and amenable to automation, thereby achieving the required throughput with minimal compound use, another crucial and limiting factor for facilitating the process of lead optimization (Fig. 5). Industry recognition of the Three Rs (3Rs) of *reduction, refinement, and replacement*,^[2] through the implementation of ethical review processes (ERP), have been partly instrumental in the increasing acceptance and incorporation of alternative in vitro models in early drug development. Additionally, pressure from other bodies such as the UK House of Lords Select Committee on Animal Experimentation, for researchers to pay increasing awareness to the 3Rs will hopefully add to the efforts made across all sectors, public and commercial, to integrate these new technologies into their processes.

PLOT IN PRACTICE

In common with any new paradigm shift, the PLOT strategy must overcome a number of hurdles. Hitherto, pharmaceutical drug discovery project teams have made decisions on compound development based largely on toxicokinetic data derived from short-term in vivo, range-finding compound exposures, which provide very limited evidence for organ-specific toxicity. This former reliance on in vivo data represents a significant obstacle to the PLOT strategy, because decisions on compound selection/rejection based on in vitro PLOT would be made before any animal exposure. Furthermore, as compounds that are selected for drug development following assessment by PLOT will still need to be evaluated in regulatory toxicology studies, it is essential that the PLOT screens are predictive of the corresponding regulatory equivalent, if it exists. The point is well illustrated by the popularity of the high-throughput SOS reporter gene mutagenicity assay, which gives results that correlate well with the regulatory gold-standard Ames assay,^[3] and requires one-tenth of the amount of compound compared with the micro-Ames assay.

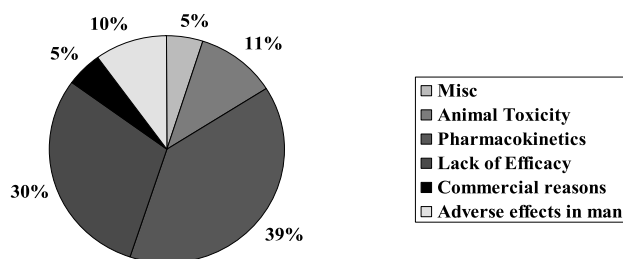


Fig. 1 Reasons for termination of research on novel chemical entities (NCEs). Data taken from information on new compounds taken into clinical trials by seven major UK pharmaceutical companies (1964–1985). (View this art in color at www.dekker.com.)

There are many individual reasons for preclinical/clinical drug failure, as listed in Table 1. These include mutagenicity, target organ toxicity, and poor bioavailability, the last being a result of a number of factors including poor absorption across the gastrointestinal tract, rapid clearance, or metabolism. A major factor for preclinical/clinical failure is the lack of suitable, practical, and/or sensitive biomarkers for target organ toxicity from which hazard and risk prediction can be extrapolated cross-species. It is anticipated that the advent of Molec-

ular Toxicology (i.e., toxicogenomics and proteomics—see later sections) will add considerable impetus to the toxicology biomarker area and incorporation into clinical pathology test batteries during drug development. Finally, unless the correct nonrodent species is selected for predicting human safety in regulatory toxicology studies, drug failure or serious adverse drug reactions (ADRs) may not occur until after preclinical development. In vitro PLOT screening can help select the correct nonrodent species very early in development.

PLOT TECHNOLOGIES

The final choice of assays, and the order in which they are performed, will depend on a number of factors, which include lead series number, compound availability (amount), the intended target, and the nature and extent of existing data on the lead series. Ideally, higher-throughput basic or level 1 screens that measure the ADME profile of the compounds should be used to analyze the large lead series, with lower-throughput toxicology assays reserved for a reduced lead series, based on compounds which have negotiated the level 1 filter (Table 2).

Because poor bioavailability represents one of the principal causes of compound failure, absorption and

Percentage of medicines dropped at each stage of development: 1997-2001

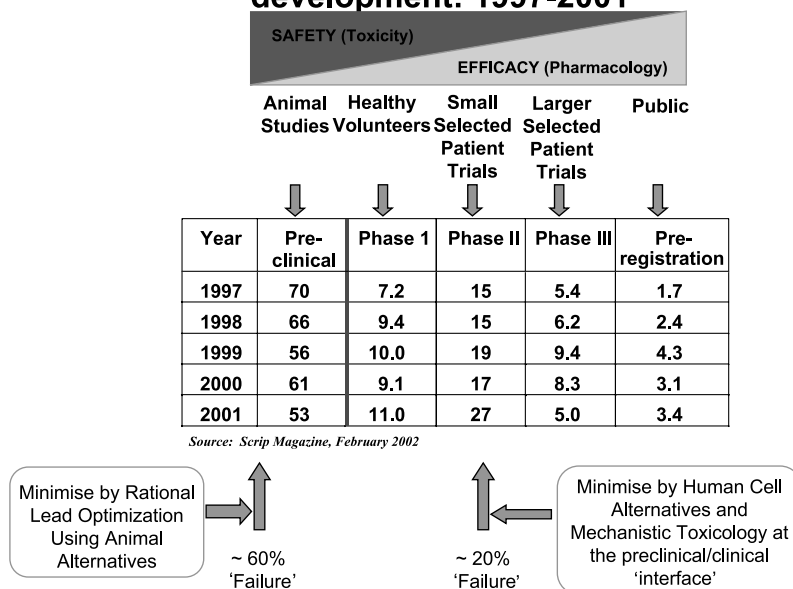


Fig. 2 Tabular representation of preclinical and clinical testing phases of drug development at which new compounds have been showed to fail for toxicokinetic and safety reasons. The data show profiles between 1997 and 2001, and also indicate the proportional contribution of safety vs. efficacy testing during drug development. (View this art in color at www.dekker.com.)

Preclinical Lead Optimisation Models

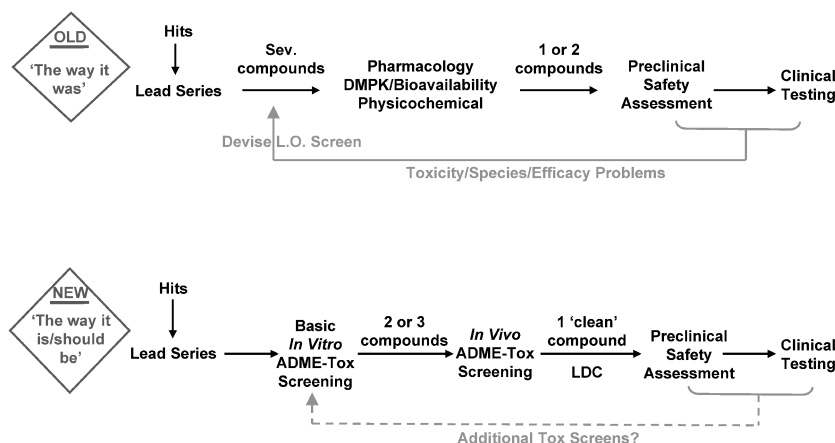


Fig. 3 Diagrammatic representation of Old and New strategies for lead optimization of new pharmaceutical molecules. The top section indicates the Old (and in some cases, current) way in which pharmaceutical companies take Hits to lead development candidate through discovery and early preclinical development testing. Essentially, no high-throughput formal lead optimization exists in this model but is developed for backup compounds by using biomarkers identified during lead compound failure in vivo. The bottom section indicates the New way forward (used now by most major pharmaceutical companies). In this model, a formal in vitro PLOT battery is inserted early in development before preliminary in vivo ADME-Tox screening and subsequent lead development candidate selection. (View this art in color at www.dekker.com.)

metabolism assays feature prominently in lead optimization screens. As oral administration represents the ideal route of drug delivery, models that predict low absorption across the gastrointestinal (GI) tract are commonly em-

ployed. Caco-2 cell assays have been widely used to predict drug uptake across the GI track, and a good correlation between in vitro and clinical data certainly exists for certain drugs (Fig. 6). This assay can also predict compounds that are substrates for the P-gp transporter, which leads to the undesirable efflux of an absorbed drug back into the lumen of the gut, resulting in its excretion. One drawback of the Caco-2 is its relatively low throughput, and so the parallel artificial membrane permeability assay (PAMPA) is increasingly used as a higher-throughput assay for the cost cost-effective assessment of passive drug absorption.^[4] Because these artificial membranes lack the cellular proteins associated with the active transport/efflux, or indeed metabolism, Caco-2 assays or cell lines transfected with specific drug transporters and human P450s may be used as a second and third tier for compounds with high permeability in the PAMPA assay (reviewed in Ref. [5]).

This principle of offering minimal core assays to screen large lead series applies well to the assessment of metabolism, which would be included as a level 1 assay (Table 2). In vitro metabolism assays might initially concentrate on compound stability in the presence of hepatocytes, microsomes, or an S9 extract, but extend to drug interactions on a more restricted series because apparent metabolic stability may be a result of inhibition of P450 function. As CYP3A4 accounts for $\times 40\%$ of the

Optimisation Technologies (PLOT) in Drug Discovery and Development

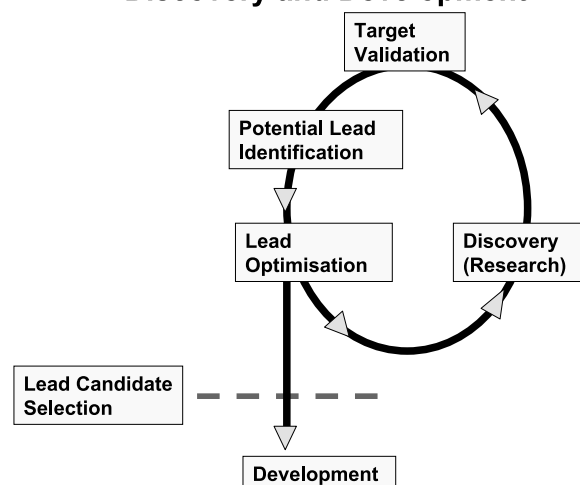


Fig. 4 The R&D ‘pipeline’ in modern drug development. (View this art in color at www.dekker.com.)



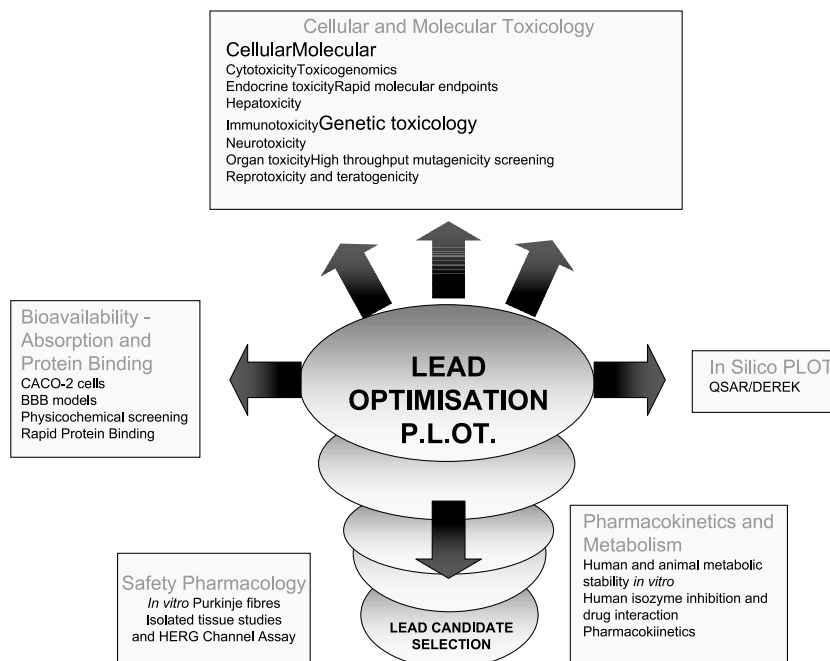


Fig. 5 Platforms comprising the preclinical lead optimization technology (PLOT) “filters” in the selection of lead development candidate molecules in drug development. (View this art in color at www.dekker.com.)

total cytochrome P450 in man, and is responsible for the metabolism of $\times 50\%$ of clinically used drugs, it makes sense to screen the inhibition or induction of this enzyme. Modulation of CYP3A4 activity is traditionally investigated by measuring enzyme function, i.e., testosterone metabolism or looking at enzyme levels by polymerase chain reaction (PCR) or Western blot, which may not be a cost-effective approach for large series lead optimization. However, recent advances in our understanding of the molecular basis for CYP3A4 induction have

identified the PXR (pregnane X receptor) as being important in drug-induced expression of this and other CYPs. Consequently, PXR reporter gene assays are being used to screen drugs with the potential to induce CYP3A4 (see, e.g., Ref. [6]), with positive results justifying a more intensive investigation. Other enzymes worth consideration are those where the existence of functional polymorphisms could limit the geographical market for a drug.

Mutagenicity testing is normally performed fairly late in compound development; however, as regulatory genetic toxicology testing in our laboratories consistently produces 20–25% *in vitro* positive results each year—which at best results in significant delays while further investigations are conducted or, at worst results, in compound failure—a faster and earlier screen is required. As described above, there now exist high-throughput reporter gene assays for mutagenicity, such as the SOS/umu assay. This has low compound requirements, and is suitable for inclusion in the PLOT platform. The basis for this microassay, originally developed by Oda et al.^[7] and adapted by Reifferscheid et al.,^[8] is that *Salmonella typhimurium*, carrying the SOS DNA repair gene fused to a reporter, respond to DNA damage by transcribing the fusion gene whose expression is correlated with the extent of DNA damage. A mammalian reporter assay, in which the human DNA repair

Table 1 Reasons for early preclinical/clinical failure

Insoluble	}	Poor bioavailability
Poorly G1 absorbed		
Fast cleared and/or metabolized		
Mutagenic/teratogenic		
Toxicity is pharmacologically/ target organ-related		
Cytotoxic/hepatotoxic		
Wrong choice of tox species to predict human safety		
Poor detection sensitivity of safety test system		
Drug–drug interactions via P450 isozyme competition/ induction/inhibition		



Table 2 Higher-throughput basic screens (level 1) and lower-throughput toxicology assays (level 2)

Level 1 assays

Physiochemical Screening—experimental and computational solubility, log *P*, log *D*, and *pK_a*

Rapid protein binding

Mutagenicity—bacterial reporter gene assay (SOS repair system) and DEREK (in Silico)

Cytotoxicity—MTT or neutral red uptake assay on mammalian cells

In vitro metabolism I—parent compound loss following hepatocyte or microsomal incubation

In vitro Metabolism II—drug–drug interactions, P450 characterization and P450 induction

Absorption I PAMPA

Absorption II—Caco-2 model

Absorption III—cell lines transfected with drug transporters

Level 2 assays

Immunotoxicity—inhibition of lymphocyte transformation and histamine release

Reprotoxicity and Teratogenicity—micromass and whole-embryo assays/stem cell assay

Hepatotoxicity—enzyme release and 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay on fresh or cryopreserved (human) hepatocytes

Endocrine disruption—yeast and mammalian reporter gene assay (oestrogenicity and androgenicity)

Neurotoxicity—brain spheroids (central nervous system (CNS)) and dorsal root ganglia (DRG) neurite outgrowth (peripheral nervous system (PNS)) assay

Cardiotoxicity—HERG patch clamp and/or HERG ligand assay

genes are fused with a reporter, is currently unavailable and eagerly awaited.

IN SILICO PLOT

Because compound availability will typically be limiting in terms of compound use, there is clearly an advantage in being able to perform computational physiochemical predictions of compound partition coefficients, isoelectric points, and solubility, based on structure. The computational prediction of drug bioavailability has been pioneered by Lipinski's "Rule of Five," which is based on compound molecular size, hydrogen bonding capacity, and lipophilicity.^[9] Increasingly sophisticated packages are being developed to analyze compound structure–activity relationships (SAR), i.e., ascribing toxicity to certain structures or substructural features. Compounds containing aromatic amines will be flagged as potential mutagens, or the existence of protein binding groups will be used to identify potential sensitizers.

PLOT AND TARGET ORGAN TOXICITY

The aforementioned basic ADME assays logically precede compound investigation in the lower-throughput, organ-specific toxicology assays, typically selected on the basis of single exposures in vivo, which may incorporate efficacy and toxicokinetic endpoints, subject to compound availability. The identity of organ-specific toxicity in vivo may stimulate the synthesis of a further lead series around the compound exhibiting the best ADME profile in the level 1 screens (Table 2). Because the liver represents a primary target organ for toxicity, level 2 screening for hepatotoxicity (e.g., using hepatocytes) may be selected by default, involving measuring enzyme release or metabolic activity by using primary hepatocytes in vitro. Although access to fresh human hepatocytes is always likely to remain problematic, it is anticipated that technical improvements will overcome current shortcomings in the quality of cryopreserved cells.

The increasing acceptance of in vitro assays by the regulatory bodies may also influence the choice of organ cultures. The ICH S7B Safety Pharmacology guideline,^[10] for instance, recommends cardiac toxicity to be evaluated on all new pharmaceuticals by investigating their potential to cause a delayed ventricular response in vitro in isolated cardiac Purkinje fibers and/or human cell lines transfected with HERG channels, alongside measurements in conscious animals (Fig. 7).

Caco-2 Model for permeability Caco-2 permeability v human data

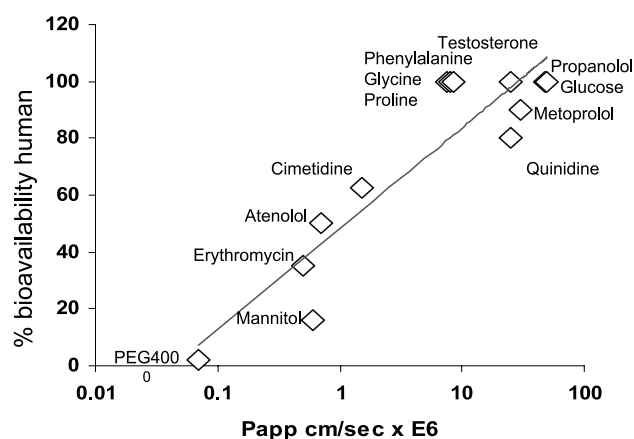


Fig. 6 Graphical representation of correlation between human bioavailability data for a series of known drugs and Caco-2 cellular permeability data (Papp coefficient). (This data is obtained with permission from AstraZeneca, 1998.) (View this art in color at www.dekker.com.)

Copyright © Marcel Dekker, Inc. All rights reserved.

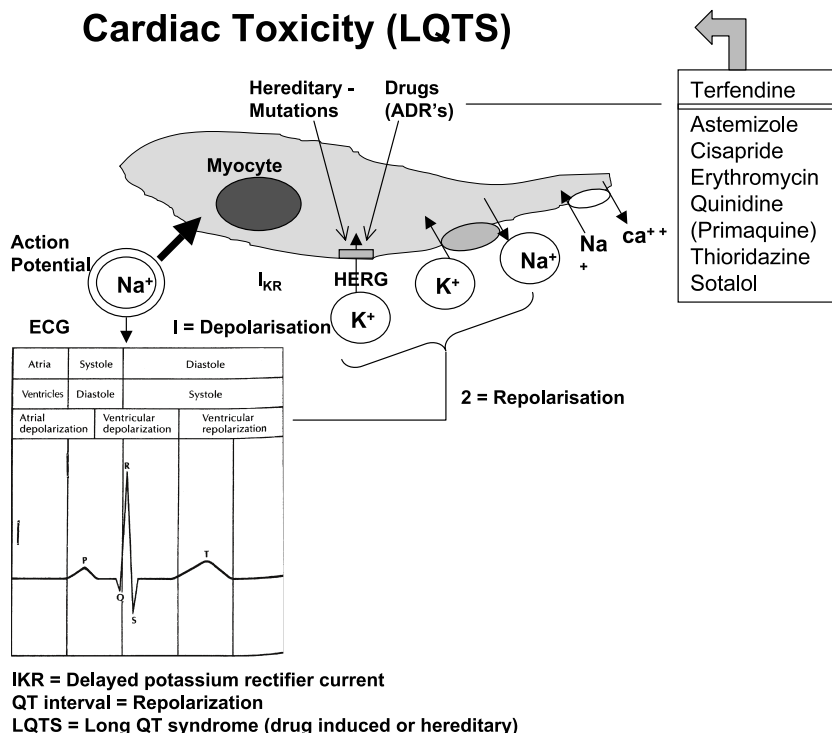


Fig. 7 Diagram highlighting the main cellular mechanisms/ion channels underlying depolarization and repolarization in cardiac myocytes. The loci of drug action in causing toxicity and ECG QT interval prolongation via block of HERG channels and leading to Torsades Des Pointes is shown, together with examples of known molecules producing such effects. (View this art in color at www.dekker.com.)

An alternative rationale for the selection of level 2 assays can be based around the intended target tissue; thus, drugs designed to target the central nervous system or immune system would warrant the use of level 2 assays to screen for, e.g., neurotoxicity or immunotoxicity, respectively. A second area where immunotoxicity screening would be especially important are clinical indications involving immunocompromized patients, where a functional assessment of the immune system is required as part of the preclinical development.^[11] These would include HIV and transplantation patients, who would be especially vulnerable to compounds with even minimal immunotoxic side effects. The in vitro assessment of reprotoxicity and teratogenicity have typically employed the low-throughput mouse whole-embryo, micromass, or xenopus assay. However, recent progress has been made in this area with the validation of an in vitro embryo-toxicity assay employing a stem cell line,^[12] which is amenable to higher throughput and represents a genuine level 2 in vitro assay (Table 2).

In summary, level 2 in vitro assays are employed to select, from lead compounds associated with the least target-organ toxicity, by using an appropriate in vitro screen. This process of in vitro screening, alternated with

in vivo exposure, may even extend beyond lead candidate selection in situations where lead candidate failure necessitates a lead follow-up series to be synthesized and screened in vitro by using PLOT platform (Fig. 3).

IRRITANCY, CORROSIVITY, SENSITIZATION, AND PHOTOTOXICITY

This is an area of significant advancement for industrial chemical notification and registration where several in vitro techniques have been legislated in favor of animal tests in certain scenarios or take their place in tier-testing strategies. The European Medicines Evaluation Agency (EMA) has also taken a positive stand on the replacement of animal studies by in vitro methods.

There is not a widespread use of primary in vitro screening assays by pharmaceutical companies to detect potential skin and eye toxicity in the lead optimization process. However, as some of these assays will be required as part of the worker safety/transportation/pre-clinical regulatory package, there are instances when these tests should perhaps be considered as part of the



selection process. In vitro assays exist that may be suitable for the assessment of skin and eye toxicity, some of which have received recent acceptance of by the EU regulatory authorities under the 27th Adaptation to Technical Progress following their successful validation by ECVAM. Accepted tests currently include measurement of skin corrosivity (i.e., EpiDerm™ or rat skin transcutaneous electrical resistance) and phototoxicity (3T3 cells). Other assays undergoing validation may already be acceptable to the regulatory authorities on a case by case basis (i.e., eye irritancy); however, in the absence of a positive result, an animal test would still be required. In addition to corrosivity and phototoxicity, it is anticipated that in vitro assays will ultimately allow irritancy and sensitization to be predicted.

In addition, when considering pharmaceutical compounds requiring intravenous, ocular or topical routes of administration as a strategic part of their development plan, then the same consideration for early in vitro screening should be made. In fact, if the final development plan for a particular therapeutic class indeed required such routes, then the lead optimization process (or some later preclinical selection step) could screen for these eventualities at an early stage.

MOLECULAR TOXICOLOGY AND PLOT

It has been stated that there are no toxicologically relevant outcomes in vitro or in vivo, with the possible

exception of rapid necrosis, that do not require differential gene expression arising through mRNA transcription or stabilization.^[13] These “stress” gene responses constitute an evolutionary developmental means by which an organism can protect itself from a hostile external environment (Fig. 8), and activate survival or “self-destruct” processes through apoptosis. Thus, the use of microarray technology [gene expression microarrays (GEMs or ToxCChips)] to establish gene expression “fingerprints” or expression patterns following compound exposure, and thus to develop screening systems for in vitro lead optimization processes, has become attractive to preclinical scientists interested in drug safety. For more in-depth information on the type and extent of stress genes and their expression patterns, the reader is referred to Refs. [13,14]. In addition to the measurement of gene expression, it is now possible to perform high-throughput proteomic profiling on cell extracts or body fluids by using platforms such as Ciphergen’s protein chips or specific protein profiles using antibody arrays. The analysis of protein vs. RNA may be preferable where the correlation between gene expression and protein is poor, where there is a low dynamic range for RNA upregulation, or in situations where protein function is post-translationally regulated (e.g., by phosphorylation) for instance. Furthermore, because biomarkers may be readily found in blood and urine, they provide a noninvasive source of starting material for proteomic analysis. The potential application of “omics” technologies to the modern research and

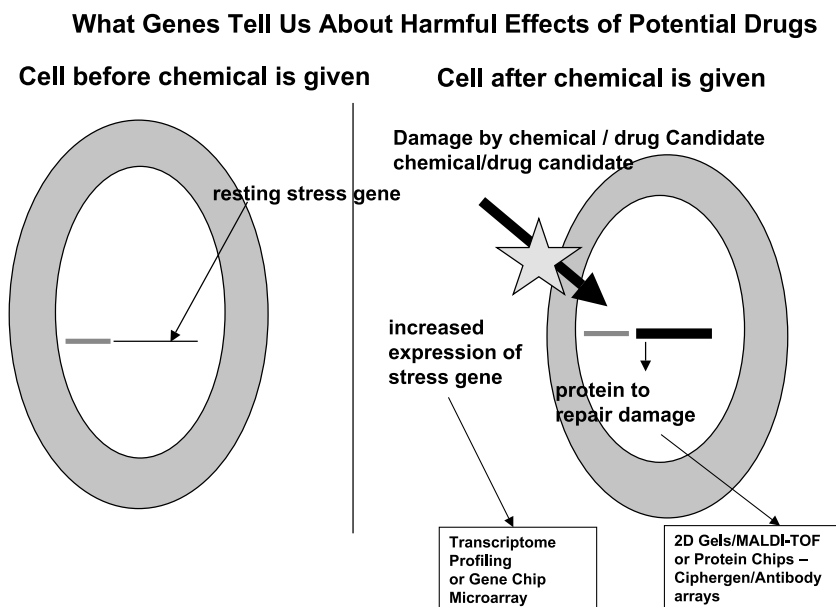


Fig. 8 Cellular and molecular events in toxicity showing also technologies used to measure effects. (View this art in color at www.dekker.com.)

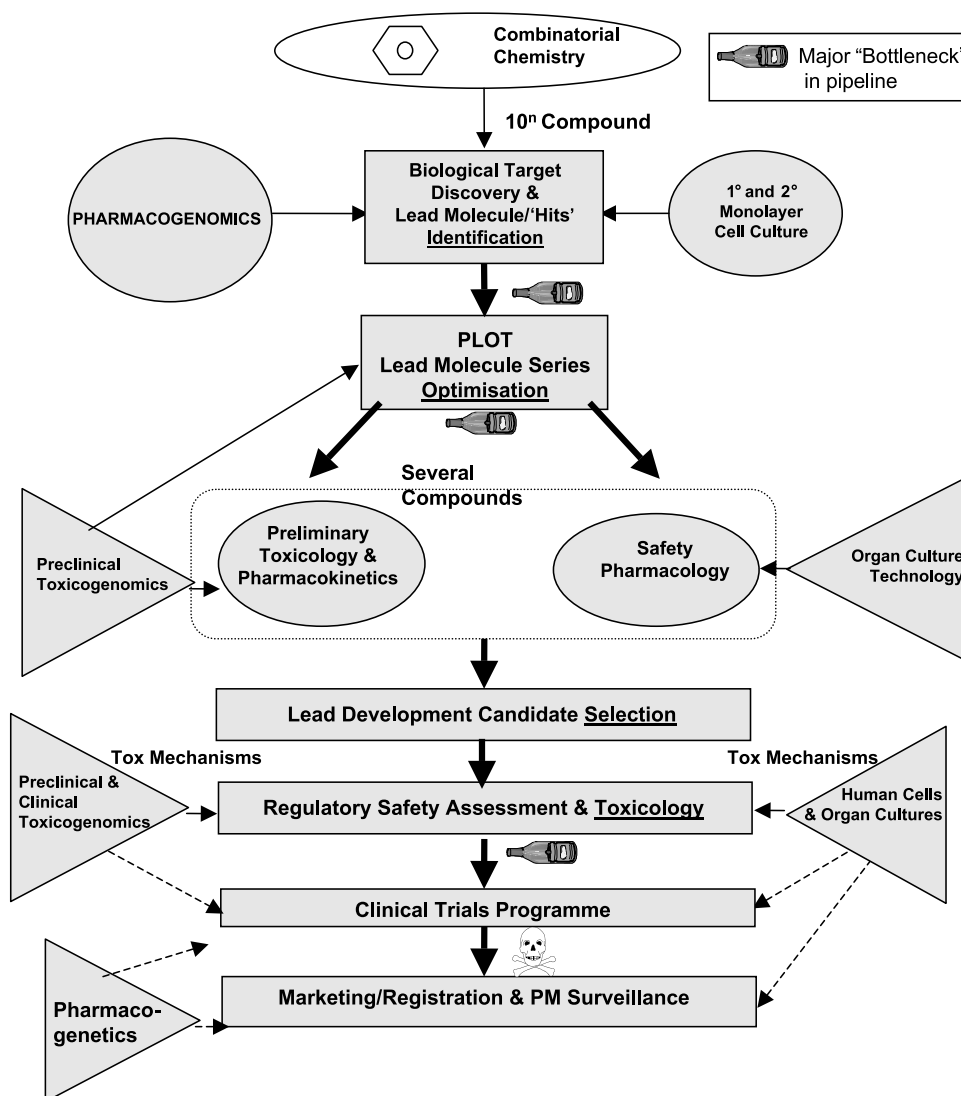


Fig. 9 Diagrammatic representation of the different stages in drug discovery and development showing where pharmacogenomics, toxicogenomics, and pharmacogenetics can be applied to a more “sophisticated” paradigm in tandem with the use of various cell culture technologies. The different potential bottlenecks in the process are also shown. (View this art in color at www.dekker.com.)

development process for new drug discovery and pre-clinical/clinical development is depicted in Fig. 9. Ultimately, this process will identify a lead development candidate molecule (lead candidate selection), by virtue of its failure to switch on “patterns” of stress genes or toxicity genes. The type of logistics potentially involved in preclinical toxicogenomics is depicted in Fig. 10, where lead candidates are tested in organotypic cultures or in vivo to look for evidence of toxicity using molecular endpoints in conjunction with traditional methodologies. Comparison with existing in vivo and emerging in vitro databases will increasingly allow the prediction of com-

pound toxicity, allowing further rounds of lead optimization to be performed in vitro to find cleaner drugs. This toxicological PLOT screen would, then, allow lead development candidate molecule selection by employing higher-throughput molecular technologies.

CELLULAR AND MOLECULAR “TOOLS” IN REGULATORY TOXICOLOGY

Following lead candidate selection and progression to the regulatory safety assessment program, toxicogenomics



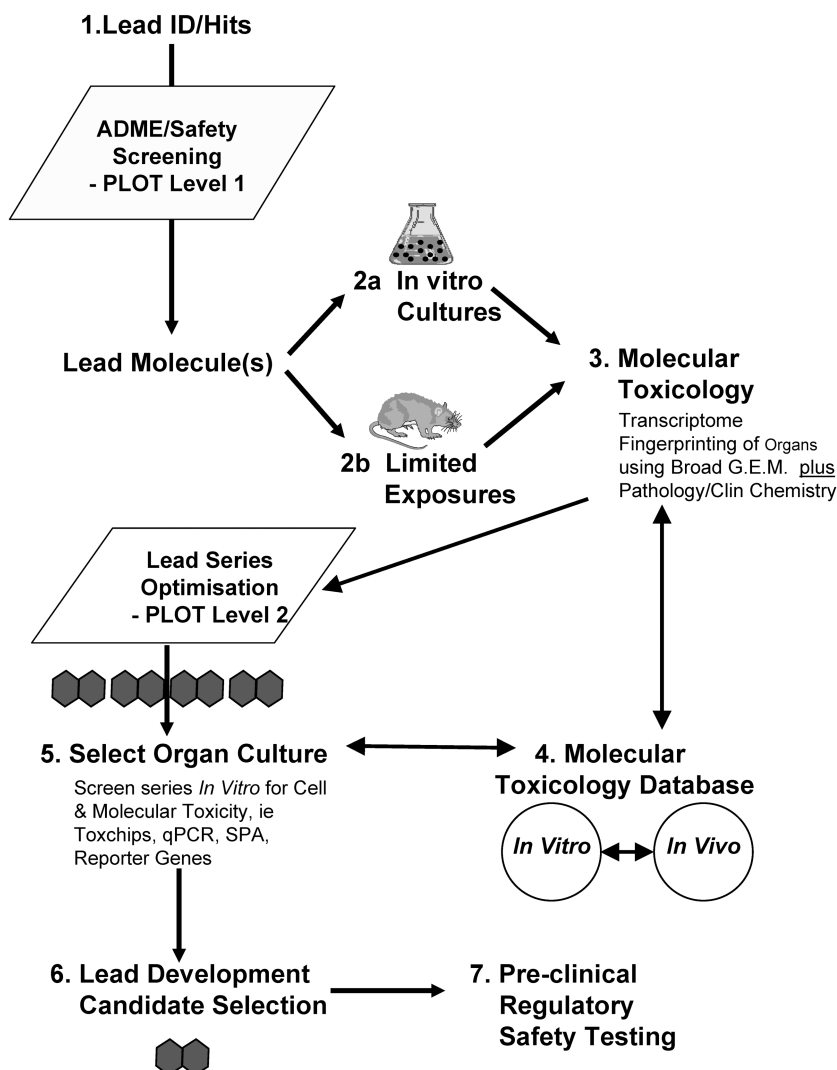


Fig. 10 A potential strategy for the selection of a lead development candidate molecule using toxicogenomic technology. Following lead optimization of a number of molecules using basic PLOT/ADMET screening (PLOT Level 1), one or several promising candidates might be administered (at high exposure levels) to a single animal whereby the organ pathology/biochemistry would be compared with tissue stress gene responses using gene chip/protein chip technology. Using this information, coupled with “knowledge” of toxicity-related stress gene/protein fingerprints from established databases, a cellular/in vitro screening system could be set up (PLOT Level 2) based on the principal gene/protein responders to screen out toxic molecules from further lead molecular series. The lead development candidate for further in vivo testing and full preclinical safety evaluation would then be selected based on a composite set of ADMET/PLOT and cellular target-organ toxicity data. (View this art in color at www.dekker.com.)

can be once more used (Fig. 9), to complete various aspects of the preclinical and clinical hazard and risk assessment process: this involves applying the same basic technologies as described above. However, there may be differences in interpretation of data, and more emphasis may be placed on the use of cells (in vitro or ex vivo) from nonrodent (e.g., dog or primate) species and humans. This would help to facilitate risk assessment as one moves from animal to human safety and Phase I clinical trials, as

well as the administrative and regulatory processes associated with new drug development.

Single nucleotide polymorphisms (SNPs) are single nucleotide differences present throughout the genome, which differ between unrelated individuals. Where they result in changes to the coding region of genes, the cSNPs may affect protein efficiency or specificity, which may ultimately result in adverse drug reactions (ADRs), as in the case of some of the P450 polymorphisms. This



problem is well illustrated by certain CYP2C9 polymorphisms and warfarin-associated haemorrhaging. The availability of SNP databases will increasingly permit pharmacogenetics to reduce the uncertainty in predicting an individual's response to new medicines by SNP profiling. Pharmacotoxicogenetic analysis of such SNPs will add information and enable a better selection to be made of volunteers or patients, and will ensure optimal safety/efficacy assessment for drug-tailored trials and therapies with the registered drug (Fig. 9). The importance of pharmacotoxicogenetics has not been lost on the regulatory bodies, and there is already an ongoing dialogue between the European Medicines Evaluation Agency and the pharmaceutical industry, to discuss the implications of this technology for the design of clinical trials.^[15]

HURDLES AND THE WAY FORWARD

It is of paramount importance to bear in mind that applying preclinical and molecular clinical technologies to regulatory drug development toxicology and human risk assessment, as described above, may preclude some pharmaceutical organizations from take-up on this strategy until the regulatory perspective is more clearly defined. Toxicology bioinformatics databases (containing gene expression data) are evolving that also contain conventional histopathology, hematology, and clinical chemistry data, in attempts to link these new molecular technologies with those biomarkers already established and familiar to regulatory toxicologists. Commercial organizations, (such as Memorec and Genlogic Inc.), as well as academic and governmental regulatory agencies, are now mindful of this need and are using both in vivo and in vitro exposure data to this end for both pharmaceuticals and industrial chemicals. In particular, human and rat hepatocytes have recently been extensively used for such validation studies and toxicogenomic data incorporated into databases alongside in vivo animal exposure data.^[16,17] In addition, (and from a pragmatic perspective), there may be the need to redefine the formal risk-assessment process (from a quantitative standpoint) in progressing from animal to first human exposure in terms of these new and relatively unknown "sets" of sensitive molecular biomarkers which may result in different no-observed effect levels/lowest observed effect levels, from the classical toxicopathological endpoints, and thus in potentially different levels of safety margin factoring in risk assessment.

In this paper, we have attempted to describe how basic ADME level 1 and organ toxicology level 2 PLOT platform assays can interface with high-throughput screens to help make decisions on selection of a lead development

candidate. It is envisaged that histopathology will increasingly be used in parallel with more-sensitive open gene array or proteomic analysis of tissues following in vivo exposures, as a means of identifying new "batteries" of toxicopathological biomarkers for preclinical toxicology and human safety assessment. It is not possible to pioneer new destinations in drug development and safety assessment until the journey has been undertaken. En route, much information has been collected and "casualties" have been inflicted, hence the need for comprehensive databases of molecular toxicological informatics to be established in different species vs. humans.

It is clear then that we are now entering a new era of both lead optimization technology and drug development. As we move from "old" to the "new" models for lead optimization of pharmaceutical molecules, (Fig. 3), harmonized preclinical strategies across the pharmaceutical industry are gradually being incorporated into research programs, mindful always of the need to apply the Three 3Rs principles of Russell and Burch^[12] to safety testing.

REFERENCES

1. *Approaches to High Throughput Toxicity Screening*; Atterwill, C.K., Goldfab, P., Purcell, W., Eds.; Taylor and Francis: London, 1999.
2. Russell, W.M.S.; Burch, R.L. *The Principles of Human Experimental Technique*; Methuen: London, 1959.
3. Reifferscheid, G.; Heil, J. Validation of the SOS/umu test using test results of 486 chemicals and comparison with Ames test and carcinogenicity data. *Mutat. Res.* **1996**, *369*, 129.
4. Zhu, C.; Jiang, L.; Chen, T.-M.; Hwang, K.-K. A comparative study of artificial membrane permeability assay for high throughput profiling of drug absorption potential. *Eur. J. Med. Chem.* **2002**, *37*, 399–407.
5. Mizuno, N.; Sugiyama, Y. Drug transporters: Their role and importance in the selection and development of new drugs. *Drug Metab. Pharmacokinet.* **2002**, *17* (2), 93–108.
6. Luo, G.; Cunningham, M.; Kim, S.; Burn, T.; Lin, J.; Sinz, M.; Hamilton, G.; Rizzo, C.; Jolley, S.; Gilbert, D.; Downey, A.; Mudra, D.; Graham, R.; Carroll, K.; Xie, J.; Madan, A.; Parkinson, A.; Christ, D.; Selling, B.; Lecluyse, E.; Gan, L.-S. CYP3A4 induction by drugs: Correlation between a pregnant X receptor reporter gene assay and CYP3A4 expression in human hepatocytes. *Drug Metab. Dispos.* **2002**, *30* (7), 795–804.
7. Oda, Y.; Nakamura, S.; Oki, I.; Kato, T.; Shinagawa, H. Evaluation of the new system (umu-test) for the detection of environmental mutagens and carcinogens. *Mutat. Res.* **1985**, *147* (5), 219–229.
8. Reifferscheid, G.; Heil, J.; Oda, Y.; Zahn, R.K. A microplate version of the SOS/umu-test for rapid detection of genotoxins and genotoxic potentials of environmental samples. *Mutat. Res.* **1991**, *253* (3), 215–222.



9. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3.
10. *ICH S7B Draft Guideline on Safety Pharmacology Studies for Assessing the Potential for Delayed Ventricular Repolarisation (QT Interval Prolongation) by Human Pharmaceuticals*; 2001. <http://www.emea.eu.int>.
11. *US FDA Guidance for Industry—Immunotoxicological Evaluation of Investigational New Drugs*; April, 2002. <http://www.fda.gov/cder/guidance/index.htm>.
12. Genschow, E.; Spielmann, H.; Scholz, G.; Seiler, A.; Brown, N.; Piersma, A.; Brady, M.; Clemann, N.; Huuskonen, H.; Paillard, F.; Bremer, S.; Becker, K. The ECVAM International validation study on in vitro embryotoxicity tests: Results of the definitive phase and evaluation of prediction models. *ATLA* **2002**, *30*, 151–176.
13. Farr, S.; Dunn, R.T. Review: Gene expression applied to toxicology. *Toxicol. Sci.* **1999**, *50*, 1–9.
14. Kind, C.; Hammond, T. Early Toxic Stress or Genes—Modulation and Screening Technologies. In *Approaches to High Throughput Toxicity Screening*; Atterwill, C.K., Goldfarb, P., Purcell, W., Eds.; Taylor and Francis: UK, 1999.
15. EMEA. *Report to CPMP on the EMEA Seminar on the Use of Pharmacogenetics in the Drug Development Process (CPMP/1483/00)*; 2002. Web site <http://www.eudra.org/humandocs/Humans/General.htm>.
16. Mendrick, P.; Higgs, B.W.; Porter, M.W.; Castle, A.L.; Orr, M.S. Using gene markers identified from a large database built with primary rat hepatocytes for prediction of human hepatotoxicity. *Toxicologist* **2003**, *72* (S1), 244.
17. Tang, L.; Komoroski, B.; Tieglar, K.; Tibbits, J.; Nolan, T.; Suizu, R.; Neft, R.; Kier, L.; Strom, S.; Li, A.P. A comparison of array hybridization, mRNA differential display and real-time PCR in the evaluation of the effects of Troglitazone on gene expression in primary human hepatocytes. *Toxicologist* **2003**, *72* (S1), 149.



Request Permission or Order Reprints Instantly!

Interested in copying and sharing this article? In most cases, U.S. Copyright Law requires that you get permission from the article's rightsholder before using copyrighted content.

All information and materials found in this article, including but not limited to text, trademarks, patents, logos, graphics and images (the "Materials"), are the copyrighted works and other forms of intellectual property of Marcel Dekker, Inc., or its licensors. All rights not expressly granted are reserved.

Get permission to lawfully reproduce and distribute the Materials or order reprints quickly and painlessly. Simply click on the "Request Permission/Order Reprints" link below and follow the instructions. Visit the [U.S. Copyright Office](#) for information on Fair Use limitations of U.S. copyright law. Please refer to The Association of American Publishers' (AAP) website for guidelines on [Fair Use in the Classroom](#).

The Materials are for your personal use only and cannot be reformatted, reposted, resold or distributed by electronic means or otherwise without permission from Marcel Dekker, Inc. Marcel Dekker, Inc. grants you the limited right to display the Materials only on your personal computer or personal wireless device, and to copy and download single copies of such Materials provided that any copyright, trademark or other notice appearing on such Materials is also retained by, displayed, copied or downloaded as part of the Materials and is not removed or obscured, and provided you do not edit, modify, alter or enhance the Materials. Please refer to our [Website User Agreement](#) for more details.

Request Permission/Order Reprints

Reprints of this article can also be ordered at

<http://www.dekker.com/servlet/product/DOI/101081EEPT120019307>

MOISTURE IN PHARMACEUTICAL PRODUCTS

R. Gary Hollenbeck

University of Maryland School of Pharmacy, Baltimore, Maryland, U.S.A.

INTRODUCTION

The earth contains 75% water (1), some of it in pharmaceutical products. With regard to solid dosage forms, pharmaceutical scientists have struggled for years to answer questions on how much is there, in what form is it, and if, when, and how might it influence a product or manufacturing process.

The term moisture, usually defined as wetness conferred by an unidentified liquid (2), is assumed here to be due to water. Thus, the scope of this article is the characterization of and consequences due to relatively small amounts of water associated with solids of pharmaceutical interest. Chemical stability, crystal structure, powder flow, compaction, lubricity, dissolution rate, and polymer film permeability are some properties of pharmaceutical interest that have been demonstrated to be influenced by the presence of moisture. Wet granulation, extrusion, spheronization, tray drying, freeze drying, spray drying, fluid-bed drying, tableting, and aqueous film coating are some unit operations that obviously depend on the amount and state of water present.

Moisture can and does influence the properties of individual active ingredients and excipients, and it is essential, as a first step, to characterize the effect of moisture on these individual components. Indeed, most of this article examines the sorption behavior of specific chemicals. However, the behavior of a pharmaceutical formulation is a complicated function of the individual component attributes. The following is an example of heterogeneous moisture distribution: A powder blend containing 2% moisture has a 10% content of an hydrophilic binder. If one assumes that all the moisture is associated with the binder, the moisture content of the binder itself would be 20%. Photomicrographic evidence demonstrates dramatically how the presence of a disintegrant, for example, can dramatically alter the environment in which a moisture-sensitive drug may exist at a particular relative humidity (3). Even though at least one model has been developed that allows the prediction of component moisture contents in a blend (4), the distribution of water in complicated formulations is largely uncharted territory.

Historical Perspective

In the area of moisture in pharmaceutical products, it is possible to identify three stages in the scientific and regulatory history. The first stage dealt more or less exclusively with the amount of water present in pharmaceuticals, most of which were products of natural origin, with regard to issues of potency and commerce.

In the second stage, there was a realization that water could affect the chemical and physical properties of drugs and dosage forms. The fact that water might exist in different states was exemplified by partitioning the water into "bound" or "free" moisture. The implication was that the free, or solvent-like, moisture was responsible for most stability and production problems (5).

The third stage came with the realization that even small amounts of "bound" moisture could have a dramatic impact on properties and processes of pharmaceutical interest. In the evolution of this scientific pursuit, it is now evident that the state of moisture is as important as the amount present. Although sophisticated thermodynamic characterization of adsorbates has been the subject of research since the pioneering work of Gibbs (6), this stage has been enhanced by the ability to examine behavior at the molecular level by using powerful new analytical tools.

Over the years, only few studies linked a thorough characterization of moisture with physical and chemical stability and production problems. Recent interest indicates that the subject is still important, there are still questions to be answered, and the gap between thermodynamic characterization and technological application will continue to narrow.

Compendial Standards

The methods for moisture determination in USP 24/NF 19 (7) are at best, classical, addressing only the determination of moisture content (8). The specifications for most official articles are arbitrary and may be excessively restrictive or broad. They are not supported by the type of data referred to previously that relate moisture content to the stability or performance of the article.

A USP Advisory Panel on Moisture Specifications initiated a revision process (9) with the inclusion of background material in the general chapter <1241> “Water-Solid Interactions in Pharmaceutical Systems.” However, this occurred over 10 years ago and it has not been followed by the inclusion of standard analytical tests that can be used to characterize the state of water.

The USP offers two methods for the determination of moisture content in solids: titrimetry (Karl Fisher titration) and gravimetry (e.g., thermal gravimetric analysis). Applications, advantages, and disadvantages of these methods are addressed later. Most articles listed in the official compendia contain specifications on “water” or “loss on Drying (LOD).” For chemicals, the gravimetric method is the same as the physical test in the general chapter <731> “Loss on Drying” (8). As volatile components other than water may be present, loss on drying is not a de facto moisture content determination. Inclusion of a “water” specification is an indication that the only volatile component present is water.

The nature of moisture content determinations is perhaps best exemplified by the fact that the *Handbook of Pharmaceutical Excipients* (10) lists 31 separate versions of laboratory tests to determine moisture content and one to determine equilibrium moisture content. These methods were used to assess the moisture content in the compilation of 148 monographs (10), many of which do not deal with solids.

BACKGROUND

Conventions, Definitions, and Terminology

The subjects of moisture and solids have not been distinguished by clear definitions and consistent usage. Table 1 clearly defines terms that are often misused.

Several conventions will be followed in this article. Moisture contents are expressed on a dry weight basis. When necessary, the subscript 1 is used for water, and the subscript 2 for the other component in the adsorbent; in the absence of a subscript, the property is to be attributed to water. Results of many routine tests conducted in my laboratory have been included here. Although the specific equipment used and all the experimental details are not provided, an effort has been made to include information that can substantially influence the results of the test (e.g., heating rate in thermal gravimetric analysis). These data should be taken as being representative of the substance tested and not necessarily as the results to be expected for a given lot of the substance examined with different instrumentation under different experimental conditions.

The sorption isotherm is the most widely used expression to quantify a substance's affinity for water. It is, as the term implies, a relationship at a constant temperature.

$$n = f(x) \quad (1)$$

where n , the number of moles of water sorbed, is a function of x , the partial pressure of water in the atmosphere at that temperature. The correspondence of n to moisture content and x to relative humidity, as defined in the Table 1, suggests that this functional relationship may be stated in numerous ways. Often, the graphical presentation of sorption isotherms is a plot of the dependent variable moisture content versus relative humidity, with both expressed on a percentage basis.

Idealized moisture isotherms are presented in this article for substances that sorb moisture in discrete stages (e.g., crystalline materials capable of forming a hydrates) and for substances that do not interact with water in discrete stages (11). These idealized isotherms form a basis for the discussion of deviations and unexpected effects of moisture sorption that can influence the physical or chemical properties of the solid.

Figure 1 is a sorption isotherm constructed for a hypothetical solid with a molecular weight of 180, which is capable of forming a monohydrate. In this case, the solubility of the substance in water is assumed to be 50 wt% and the critical humidity for the transition from anhydrous to monohydrate is assumed to be 60% RH. It is also assumed that a solution of the solid in water behaves ideally.

The stepwise character demonstrated in this profile is characteristic of a substance capable of forming a hydrate. At humidities below 60% RH, the moisture content of the solid remains virtually unchanged. When the activity of water reaches a critical relative humidity of 60%, water is sorbed by the solid as the anhydrous form converts to the monohydrate. For this example, the moisture content associated with the complete transition to the monohydrate is 10% (e.g., 1 mole of water = 18 g; 1 mole of the substance = 180 g).

Further increase of humidity has no appreciable effect on moisture content until a water activity is reached equal to that associated with a saturated solution of the substance in water at the temperature of the analysis. The relative humidity associated with this activity will be referred to as RH_s . In this case, $RH_s = 91\%$ relative humidity, as the solubility of 50 wt% is $x_2 = 0.09$. On a mole fraction basis, and, assuming ideal solution behavior, $x_1 = 1 - x_2 = P_1/P_1^0$. In this hypothetical case, the moisture content of the saturated solution when all the solid is dissolved is 100%; therefore, at 91% RH the monohydrate sorbs water from the atmosphere until dissolution is complete. The fact that constant humidity is produced by a saturated solution of a

Table 1 Definitions

Term	Symbol(s)	Definition
Sorption		The spontaneous acquisition of a component (water in this case) from the atmosphere by a system.
Adsorption		Sorption confined to the surface of the solid. The amount of water adsorbed is directly proportional to the surface area available.
Absorption		Sorption characterized by penetration of the sorbed component into the bulk structure of the solid. The amount of water absorbed does not depend on surface area.
Desorption		The spontaneous loss of the sorbed component (water) to the atmosphere.
Sorbent		The substance or system responsible for sorption.
Adsorbent		Ostensibly the substance responsible for adsorption. The term “absorbent” is not in common usage. Adsorbent is used without discriminating between adsorption or absorption.
Adsorbate		The substance being adsorbed (water in this case).
Moisture content	MC, W	The total amount of water present with the adsorbent.
Dry weight basis	W_D	An expression of moisture content related to the weight of dry solid.
	$\%MC_D$	$W_D = \text{wt. water/wt. of dry solid}$ Amount of moisture present with 100 g of dry solid.
		$\%MC_D = 100W_D$
Wet weight basis	W_W	An expression of moisture content referring to the total weight of the sample.
	$\%MC_W$	$W_W = \text{wt. water/wt. of sample}$ Amount of moisture present in 100 g of sample.
		$\%MC_W = 100W_W$
Equilibrium moisture content	EMC, $\%EMC$, $W_{D,eq}$, $W_{W,eq}$	The moisture content (sometimes expressed on a percentage basis) at equilibrium under specified conditions of temperature, pressure, and vapor composition.
Partial pressure of water in the atmosphere	P/P_0	The vapor pressure of water in the atmosphere (P) expressed as a fraction of the saturation vapor pressure of pure liquid water (P_0) at the same temperature.

(Continued)

Table 1 Definitions (*Continued*)

Term	Symbol(s)	Definition
	x	As the behavior of water vapor is generally assumed to be ideal (fugacity = pressure), x is the activity of water.
Relative humidity	%RH, $100P/P_o$, $100x$	The vapor pressure of water in the atmosphere (P) usually expressed as a percentage of the saturation vapor pressure of pure liquid water (P_o) at the same temperature.
Term	Symbol(s)	Definition
Bound moisture		Water associated with a solid exhibiting a vapor pressure less than P .
Free moisture, unbound moisture		Water present in a solid exhibiting a vapor pressure greater than P . This water is sometimes referred to as "freezable" water.
Hygroscopicity		A term synonymous with sorption, implying an acquired amount or state of water sufficient to affect the physical or chemical properties of the substance.
Water of hydration		Water present in regular positions within a crystal lattice (3). (There is a specific stoichiometry with the other molecule.)
Efflorescence		Spontaneous loss of water of hydration (not thermally induced).
Deliquescence		Sorption sufficient to produce dissolution of the substance.
Monolayer capacity	V_m	Volume of water (STP) adsorbed at STP when the surface is covered.
	Y_m	Amount of water adsorbed when the surface is covered.

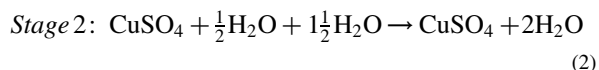
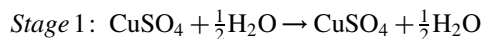
substance has been used as a basis for establishing controlled-humidity chambers.

At relative humidities above 91%, the colligative behavior of the solution results in absorption of water until the solution is in equilibrium with water vapor in the atmosphere. In Fig. 1, this region has been constructed based on Raoult's law.

Although some specific examples are given, several aspects of the sorption of water by crystalline substances deserve attention at this point. First, it is important to recognize that the large quantity of water that associates with these systems actually obscures small changes in moisture content that occur between hydrate transitions. It has been stated that lactose monohydrate, for example, "contains 5% water of crystallization and approximately

0.1% adsorbed water" (10), indicating that adsorption takes place on the hydrate.

Furthermore, this sorption isotherm is a thermodynamic relationship. Although it represents the equilibrium position, it does not provide any information about the rate at which the transitions take place. For example, the desiccant calcium sulfate goes through two stages of hydration, as shown in Eq. (2).



The first stage where the hemihydrate is formed takes place instantaneously in the presence of water vapor,

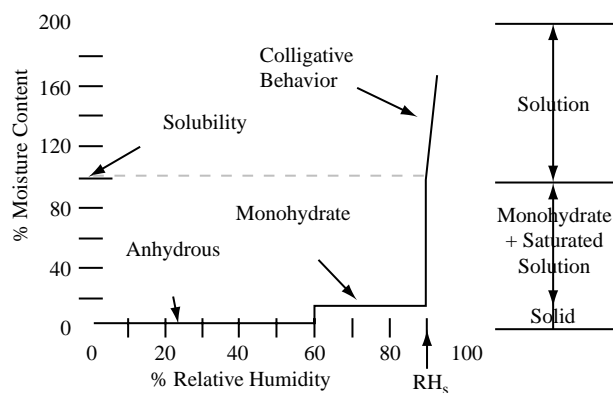


Fig. 1 Stepwise sorption isotherm for a hypothetical substance.

whereas the second stage requires a matter of weeks or months to reach completion, depending on the humidity at which the hemihydrate is stored (12).

Finally, the critical relative humidities are dependent on the nature of the solid. For example, the spontaneous dissolution process has been observed for many water-soluble substances at relative humidities significantly below that associated with a saturated solution of the substance in water (5). Van Campen et al. (13–15) have examined the moisture sorption kinetics of deliquescent solids at relative humidities above what they term the “critical relative humidity” (RH_o), where adsorbed water takes on the character of condensed water and serves as a solvent. It is important to recognize that a highly undesirable process such as deliquescence can occur when it may not be expected (e.g., when $RH_o < RH < RH_s$).

In all the previous examples, the process was absorption, and water ultimately had access to the entire solid; variations in particle size, and consequently, surface area, would not effect the profile. In many of the subsequent examples, not only is surface area a determinant of moisture content, but perturbation of the solid can occur during sorption to produce unexpected and dramatic effects.

Brunauer (16) organized the experimental isotherms observed for vapor sorption into the now classical five groups, Types I to V. Brunauer et al. (17) presented a unified theory of multilayer physical adsorption that was purported to describe all types of isotherms. Numerous texts present detailed discussion of the basis for this theory. In the context of this discussion, capillary condensation will not be considered and the focus will be Types I, II, and III behavior. Fig. 2 is a graphical representation of these three types, generated from Eqs. 3–5.

Type I isotherms, where sorption occurs on strong sorption sites, are characterized by a monotonic increase in the amount of vapor adsorbed, up to a maximum value assumed to represent a completed unimolecular layer. The Langmuir equation is generally a suitable functional relationship for $n = f(x)$, as given by

$$\frac{y}{1 + Bx} = y_m Bx \quad (3)$$

where y is the amount of water adsorbed at $x = P/P_o$, y_m is the monolayer capacity, and B is a constant, often referred to as the adsorption coefficient. Using the generally assumed value of 1250 nm^2 for the area occupied by a molecule of water on the adsorbent surface, the surface area can be estimated from y_m .

The characteristic inflection in Type II behavior occurs when multilayer sorption starts. The Brunauer, Emmett, Teller (BET) equation (17), shown as Eq. 4, can describe isotherms where multilayer sorption is evident:

$$\frac{y}{(1-x)(1+(C-1)x)} = y_m Cx \quad (4)$$

Based on the assumptions in the kinetic derivation of the BET equation, the constant C is related to the average heat of adsorption in the first layer. Type III isotherms occur when this value is low.

Application of the BET equation to experimental data has become the standard method for surface area determinations. The constants are determined from data

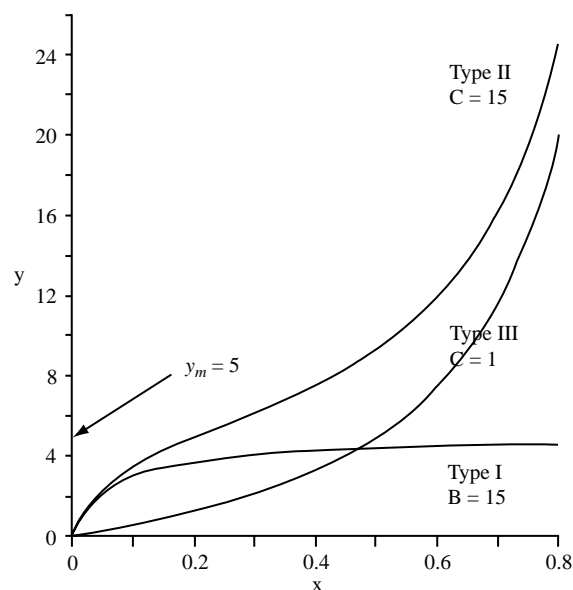


Fig. 2 Hypothetical Type I, II, and III sorption isotherms with a monolayer capacity of $y_m = 5$.

at low humidities (5–30% RH). Based on these parameters, the BET equation almost always overestimates the amount of water associated with the solid at high humidity.

As 100% RH is approached, the BET theory predicts that the amount of water adsorbed will be infinite. As the influence of the solid surface does not extend indefinitely, it should be expected that a finite number of molecular layers would be adsorbed (presuming that the solid is not soluble.) Many authors attribute the “infinity catastrophe” of the BET theory to the lack of consideration of lateral adsorbate interactions (18–20).

Ostensibly the constant C relates to the state of water associated with the solid; however, the simplistic assumptions (all molecules in the second and higher layers are assumed to have the same character as bulk water and horizontal interactions between molecules are ignored for all layers) used in the derivation preclude really useful thermodynamic information.

Equation (5), a modification of the BET equation which assumes a third thermodynamic state for the adsorbate, was developed independently by Guggenheim et al. (21).

$$\frac{y = y_m C K x}{(1 - Kx)(1 + (C - 1)Kx)} \quad (5)$$

where K is a constant that accounts for the intermediate state between tightly bound water in the first layer and bulk liquid. This model is also simplistic but it does fit data at high relative humidities better than the BET equation. Eq. 5 is termed the GAB equation (22).

All of the equations presented thus far are single-valued relationships between the amount adsorbed and the pressure of water in the atmosphere. However, hysteresis is frequently observed. When a sample that has reached a specific equilibrium moisture content is exposed to lower relative humidity, desorption occurs. In many cases, adsorption and desorption isotherms are not superimposable.

Fig. 3 is an example of a Type II sorption isotherm with hysteresis. Although the etiology of hysteresis has been the subject of many discussions (23–30), the phenomena usually can be attributed to

- Irreversible perturbation of the solid, that is, sorption can increase (swelling) or decrease (collapse of a freeze-dried cake) the surface area;
- Kinetic origin, that is, the new equilibrium position has actually not yet been reached; and
- An actual change in state of the adsorbate.

Most individuals recognize hysteresis in terms of a substance having different moisture contents at a specified

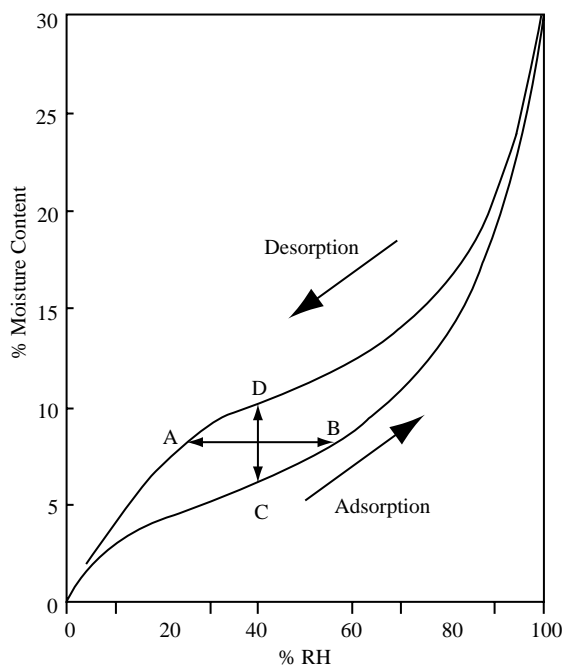


Fig. 3 Type II isotherm with hysteresis.

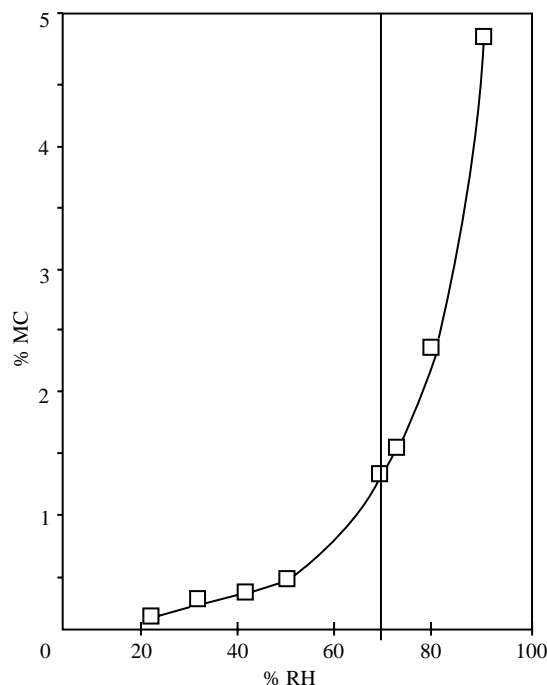


Fig. 4 Moisture sorption isotherm for lactose. (Adapted from Ref. 10.)

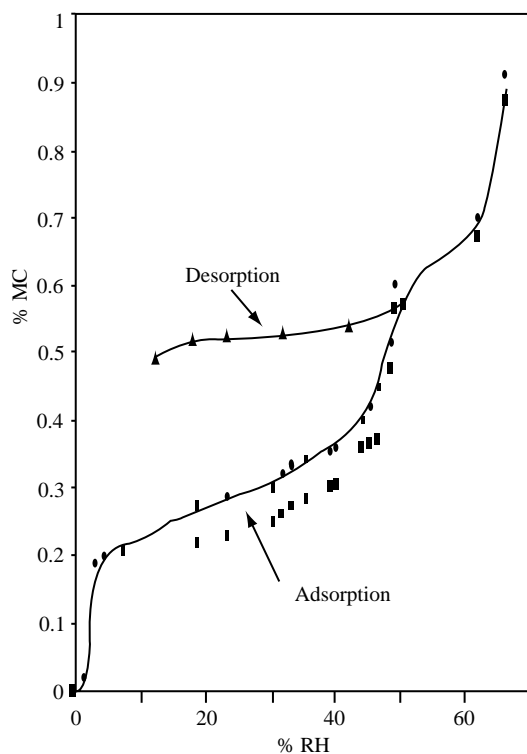


Fig. 5 Moisture sorption isotherm for compressible sugar. (Adapted from Ref. 31.)

relative humidity (CD in Fig. 3). The implications of this situation are obvious: the moisture content of a material depends on its history of exposure to different relative humidity conditions. Controlling relative humidity can only insure that the moisture content lies within the range specified by the isotherm. The other aspect of hysteresis is that substances with equal moisture contents can have adsorbed water with different chemical activity (AB in Fig. 3).

Although the sorption isotherm is fundamental to the characterization of moisture interaction with water, it is generally not possible to make any judgments about the effect of water on the substrate from the isotherm alone. For example, it is not possible to determine if an increase in moisture content is due to multilayer adsorption, swelling of the substrate, or some combination of the two.

Sorption isotherms for lactose are shown in Fig. 4, compressible sugar in Fig. 5 (31), microcrystalline cellulose in Fig. 6 (32), and aspirin in Fig. 7 (3). The isotherm for lactose has been adapted from data in the *Handbook of Pharmaceutical Excipients* (10); the step-wise isotherm expected is not evident. The dashed line in Fig. 4 is based on the statement that the monohydrate forms at 70% RH.

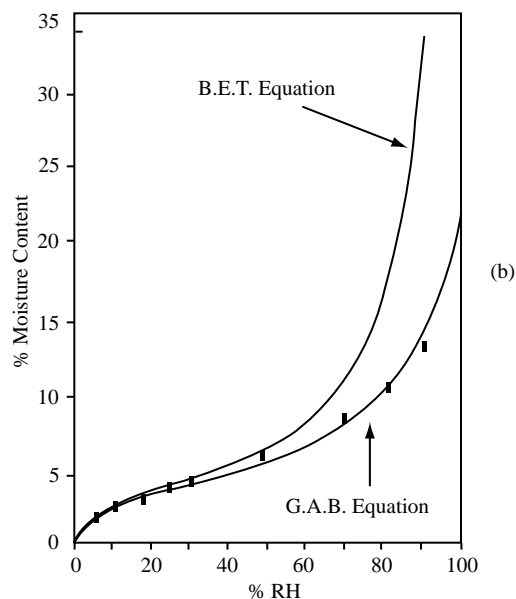
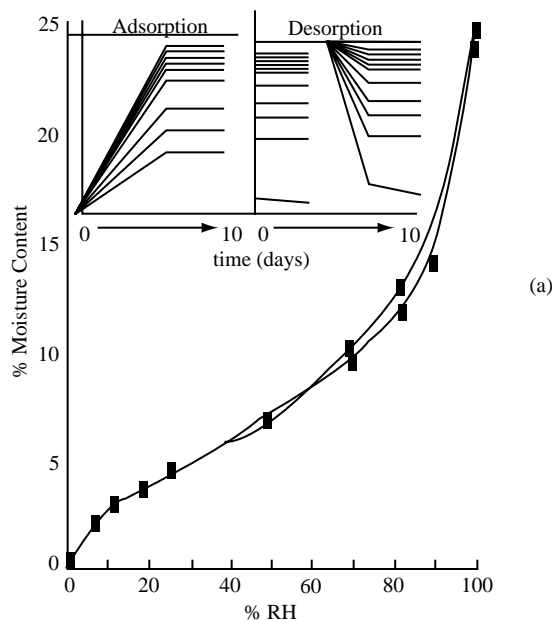


Fig. 6 (a) Moisture sorption isotherm for microcrystalline cellulose: adsorption and desorption. (Adapted from Ref. 22.) (b) BET and GAB equations describing the adsorption branch.

Compressible sugar NF is an example of a material that very slowly sorbs relatively small amounts of water, with a resulting isotherm that is not smooth. The adsorption isotherm was generated from a sample that had been carefully preconditioned to remove moisture. The dramatic difference between the desorption and adsorption branches of the isotherm indicates that water forms a

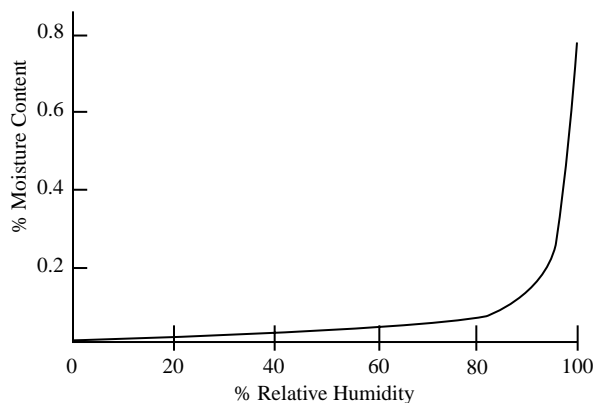


Fig. 7 Moisture sorption isotherm for aspirin. (Adapted from Ref. 3.)

hydrate-like phase (31). Zografi and Kontny (22) have demonstrated that almost all of the water taken up by compressible sugar is associated with the maltodextrin.

Moisture sorption of microcrystalline cellulose has been studied extensively (22, 32, 33). Fig. 6a includes the sorption and desorption studies for microcrystalline cellulose. The inserts are plots of moisture content versus time, which approximately represent the kinetics of sorption and desorption at each humidity. The equilibrium adsorption isotherm (Type II) has been fit to the BET equation ($C = 16.48$, $y_m = 0.033$ g/g solid) and this curve is presented in Fig. 6b along with a curve described by the GAB equation by using the same monolayer capacity.

The isotherm for aspirin would be classified as Type III, indicating a low affinity for water followed by multilayer sorption. These four isotherms cover a broad range of moisture interaction with solids of pharmaceutical interest.

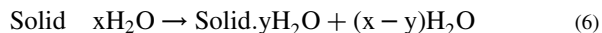
ANALYTICAL METHODS

There is no single analytical method that suffices in the characterization of moisture associated with solids. The best approach is a judicious combination of the following techniques.

In this era of automatic titrators, microprocessor-controlled thermal analysis, and definitive spectral techniques, one of the most powerful techniques, that is, optical microscopy, is frequently overlooked. The value of direct sample observation, preferably while it is exposed to different relative humidities, cannot be overstated. In the author's laboratory, a plexiglass chamber was constructed that can be placed on the stage of the microscope, through

which air of known humidity can be circulated. This simple technique has been very useful in examining the swelling (or lack) of disintegrants (34) and the influence of very hydrophilic excipients in combination with a moisture sensitive drug (3).

In a moisture determination by the physical separation of water from the solid, it is important to recognize that free and bound moisture must be dissociated from the solid by an applied stress, using



where x is the moles of water initially associated with the solid, y the moles of water still associated with the solid, and $(x - y)$ the amount of water released as a consequence of the stress. The stresses used differ according to the conditions (e.g., high temperature, low vapor pressure, anhydrous solvent systems). Hence, the moisture contents determined by different methods may very well be different. There is no guarantee that $y = 0$, and no reason to expect that y has the same value for different methods.

An equally important consideration is the fact that these stresses are different than those that may be responsible for moisture release during manufacturing or within the product after manufacture. One should approach the determination of moisture content with the disconcerting realization that different methods can produce different results and that none of these results may be relevant.

Thermal Methods

Thermal gravimetric analysis (TGA) is undoubtedly the most widely used method of moisture content determination. The sensitivity and sophistication of TGA instruments ranges from the classical moisture balance (LOD) to specially designed microbalances enclosed in chambers that may be evacuated. Microprocessor control of the temperature increase has led to more reproducible and discriminating information.

The result of a TGA of dicalcium phosphate dihydrate is shown in Fig. 8. The profile represents the weight loss with increasing temperature ($1.5^\circ\text{C}/\text{min}$) in an environment containing a desiccant. The total weight loss is the difference between the initial weight and the final constant weight of the dry solid. There are two regions in the profile, the first beginning at about 90°C and the second at about 170°C . Assuming that the only volatile component is water, calculations indicate $W_D = 0.228$ (22.8% MC), with W_D values of about 0.042 and 0.186 associated with Regions I and II, respectively.

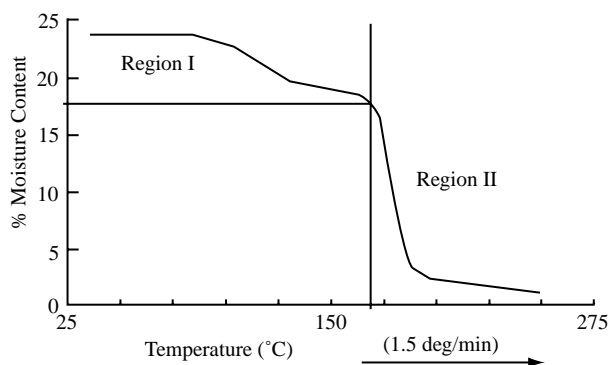
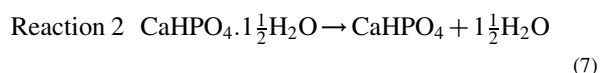
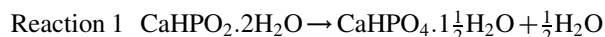


Fig. 8 Thermal gravimetric analysis of dicalcium phosphate dihydrate.

The profile in Fig. 8 is characteristic of a crystal hydrate, showing discrete regions of moisture loss occurring at relatively high temperature. This particular profile is consistent with the dehydration scheme shown in Eq. 7.



Differential scanning calorimetry (DSC) is a thermal method that measures the energy change accompanying a nonadiabatic process. A small sample of the moist solid contained in a metal sample container is exposed to a controlled increase in temperature. Water dissociated from the sample is swept out of the heating chamber by a nitrogen stream and the dynamic energy consumption (mcal/sec) required to keep the sample at the same temperature as an empty sample container is recorded. When the temperature is increased at a constant rate, the

area under the DSC curve reflects the energy, in the form of heat, associated with the phase change.

The DSC profile for dicalcium phosphate dihydrate (Fig. 9 curve A) includes two endotherms consistent with the dehydration scheme shown in Eq. 7. (In all DSC plots reported here, endotherms are positive deflections from the base line.) It should also be noted that the energy change measured is not specific for water; it includes, for example, energy consumed or released because of changes in crystal structure. Thus, it is generally not appropriate to estimate moisture content from the enthalpy change for dehydration and the specific heat of vaporization of water.

A common error is to look at temperature profiles (TGA or DSC) and conclude, for example, that no water is released from dicalcium phosphate dihydrate at 50°C. It is important to recognize that this is a dynamic test with a large temperature increase examined over a relatively short time periods. Exposure of dicalcium phosphate dihydrate at 50°C may indeed drive off water of hydration; the system is not held at any particular temperature and the relative rate of moisture loss is slow.

Curve B in Fig. 9 is the result of an interesting test on a sample of anhydrous dicalcium phosphate that had been exposed to 100% RH for three months. Clearly, some of the moisture acquired by the solid was converted into water of hydration. However, the broad endotherm from about 50 to 130°C represents the energy consumed in the dissociation of water that is not part of the crystal structure. The existence of two states of water in the same solid raises a number of questions pertaining to the wisdom of using anhydrous dicalcium phosphate as an excipient for moisture-sensitive drugs.

The TGA profile for lactose is shown in Fig. 10. This profile is consistent with a monohydrate that loses its water of hydration in one step. DSC data for lactose are shown in

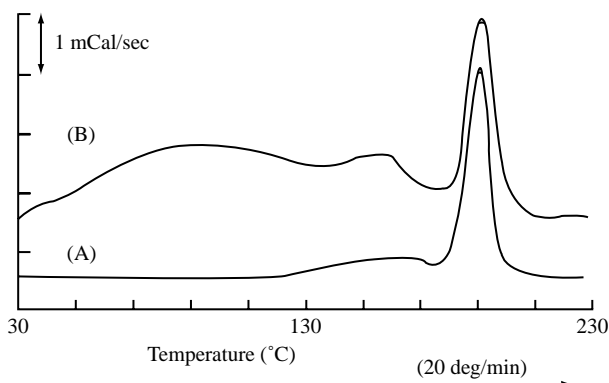


Fig. 9 Differential scanning calorimetry of dicalcium phosphate samples. (A) dicalcium phosphate dihydrate; (B) anhydrous dicalcium phosphate with adsorbed moisture.

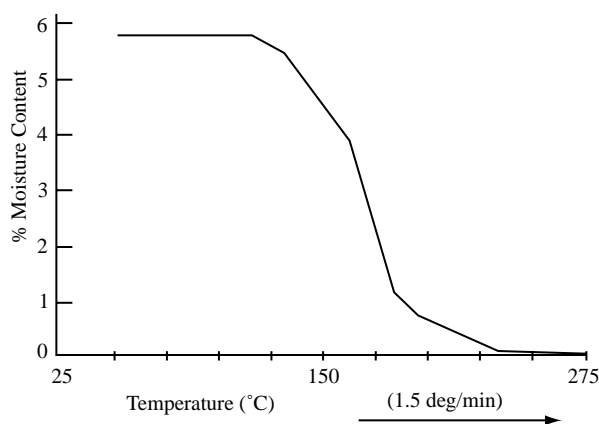


Fig. 10 Thermal gravimetric analysis of lactose monohydrate.

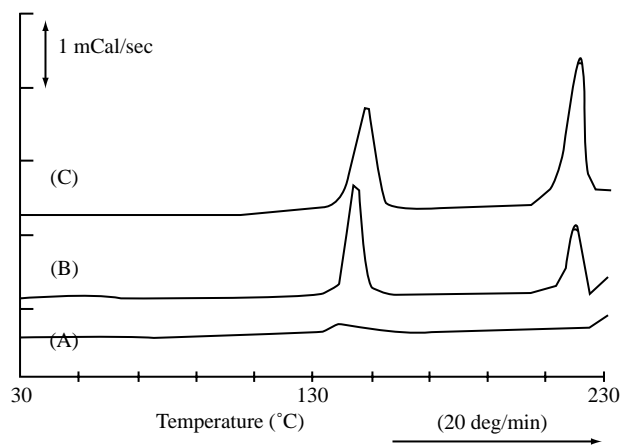


Fig. 11 Differential scanning calorimetry of lactose samples. (A) anhydrous lactose; (B) lactose monohydrate; (C) anhydrous lactose with adsorbed moisture.

Fig. 11. Scan A for anhydrous lactose is rather uneventful, although comparison with the other scans seems to indicate the presence of a small quantity of water. Scan B is for lactose monohydrate; the single water of hydration is lost at approximately 135°C, with decomposition occurring at about 200°C. Scan C is for an anhydrous lactose sample that had been exposed to 100% RH for three days. Clearly, the water sorbed by anhydrous lactose was quickly and completely converted to water of hydration.

Thermal analyses of microcrystalline cellulose reveal distinctly different results. The TGA (Fig. 12) shows a loss of moisture at low temperature and a reasonably monotonic decrease that does not indicate discrete stages of hydration. The DSC (Fig. 13) shows a broad region of moisture loss at temperatures well below the boiling point of water.

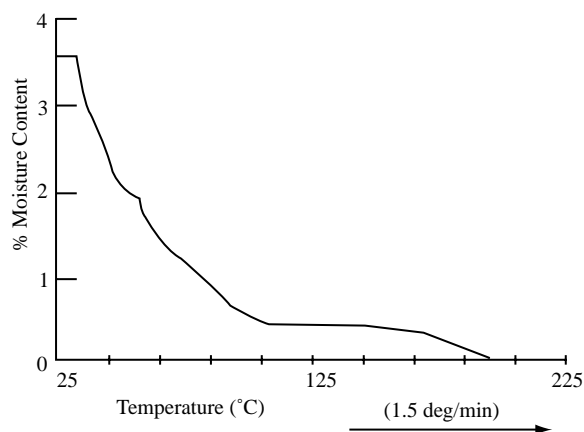


Fig. 12 Thermal gravimetric analysis of microcrystalline cellulose.

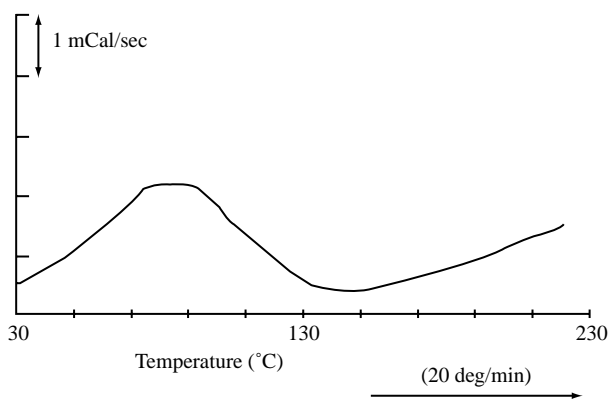


Fig. 13 Differential scanning calorimetry of microcrystalline cellulose.

The principal advantage of the thermal methods is convenience; however, these analyses are not specific for water, and exposure to high temperature may be an unrealistic stress or alter the sample. An acceptable determination of moisture content using a thermal method provides a result with minimal residual moisture and only minor alteration of the solid. Thermal methods can be combined with a specific titration of water by bubbling the evolved gas from TGA or DSC through the titration medium.

Karl Fisher Titration

For over 60 years, the specific titration of water has used a reagent developed by Karl Fisher, which consists of iodine, sulfur dioxide, and pyridine in methanol. The Karl Fisher titration of water is addressed in most analytical chemistry texts and is not presented here. A brief review with a pharmaceutical perspective has recently been published (35). However, the advantages and disadvantages of this method for the characterization of water associated with solids are discussed later.

In its simplest form, the Karl Fisher titration is a one-point determination of moisture content. The principal advantage is specificity for water. It is also a non-thermal method, which is very sensitive and can be easily automated. The main disadvantage is that the solid must dissolve in the titration medium to be sure that the total amount of moisture is released. If the analysis is carefully designed in such a way that moisture is "extracted" from the solid to the same degree each time, accurate and reproducible results can be obtained for solids that do not dissolve.

With automation, the Karl Fisher titration provides a titration kinetics profile (e.g., milliliters of titrant vs. time).

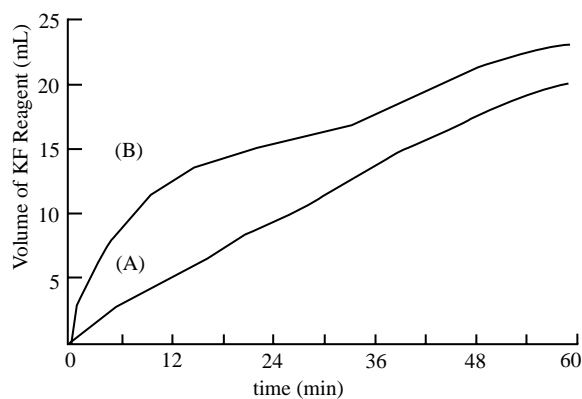


Fig. 14 Normalized Karl Fisher titrations of dicalcium phosphate dihydrate; (A) unmilled form; (B) milled form.

If the rate of water release from the solid is the rate-limiting step, the kinetics of the Karl Fisher titration profile can provide indirect information about the state of water associated with the solid. Fig. 14 shows the titrations of two different samples of dicalcium phosphate dihydrate. Here, and in the titration profiles in Fig. 15, the data have been normalized in such a way that each titration ultimately consumes 25 mL of reagent. This normalization is analogous to increasing the sample size in such a way that it contains an amount of available water which would require 25 mL of reagent to neutralize. Also, the direct titration of water can be finished in less than 90 s.

The titration profiles for dicalcium phosphate are not monotonic, but exhibit stages of dehydration consistent with the scheme shown in Eq. (7). It is also evident that the release of moisture is faster from the milled form (Fig. 14,

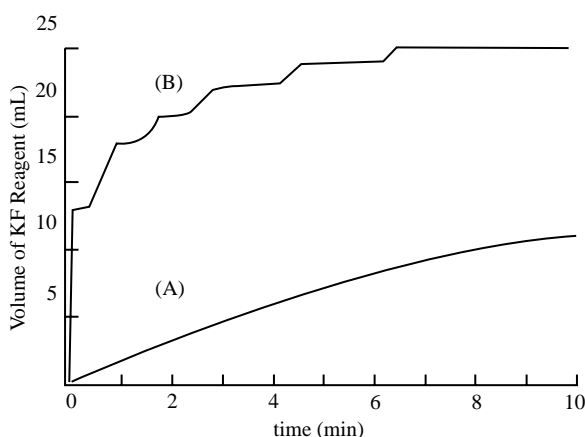


Fig. 15 Normalized Karl Fisher titration for dicalcium phosphate; (A) milled sample; (B) anhydrous sample with adsorbed moisture.

curve B) of the excipient. This is consistent with the increased total surface area and shorter diffusional path expected with smaller particle size.

In Fig. 15, Karl Fisher titration of anhydrous dicalcium phosphate, which had been exposed to 100% RH for 3 months (curve B), is compared to with that of milled dicalcium phosphate dihydrate (curve A). The adsorbed water, which was not incorporated into the crystal lattice, was released from the solid very rapidly.

Results of normalized Karl Fisher titrations for lactose monohydrate and microcrystalline cellulose both present profiles not significantly different from water, indicating a very rapid release of moisture.

Spectroscopic Methods

The most useful spectral methods for the characterization of water in solids are Fourier transform infrared spectroscopy (FTIR), nuclear magnetic resonance (NMR), and powder X-ray diffraction (XRD). A thorough treatment of these methods is not given here; instead the example of ampicillin presented by Brittan, et al. (36) is summarized.

Hydrates normally form crystal structures (pseudo-polymorphs) that differ from the anhydrous form. Different powder XRD patterns of ampicillin in the anhydrous and trihydrate forms are shown in Fig. 16.

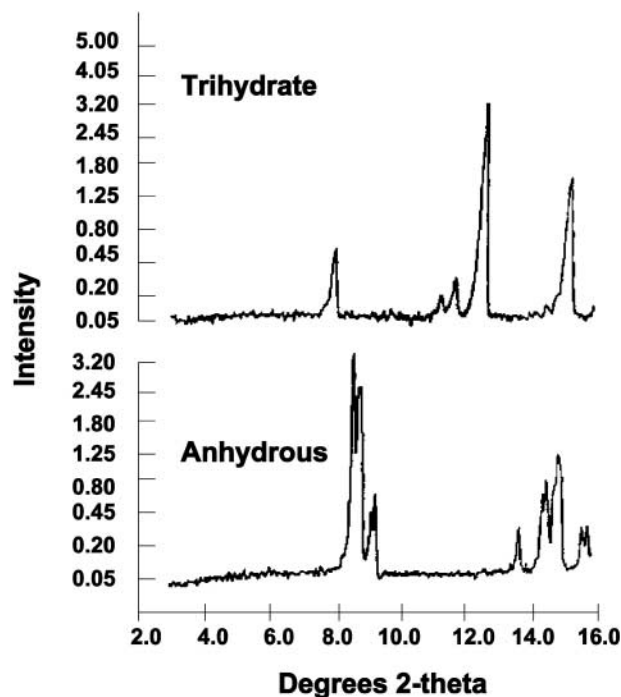


Fig. 16 Partial X-ray diffraction patterns of ampicillin samples. (Adapted from Ref. 36.)

Although moisture contents often present values that are close to being stoichiometric, X-ray confirmation of the differing crystal structure should be a requisite for designation as a hydrate.

The X-ray method is not just useful from the qualitative perspective. In samples that contain both anhydrous and hydrate forms, diagnostic regions of the pattern can be identified, and the relative areas of peaks in these regions may be used to establish the relative amounts of each phase (36).

Infrared analysis of water associated with a solid centers on an assessment of the degree to which the environment influences the stretching frequency associated with the —OH group. The —OH stretching mode for free water in the gaseous state has a characteristic energy of 3655 cm^{-1} . The frequency of this stretching is lowered when water is condensed and/or bound. Ice has a characteristic —OH stretching frequency of 3400 cm^{-1} . By comparison of the FTIR spectra for the anhydrous form with those of the sample with water, the —OH bands for water can be identified. The distinctive, sharp peak for crystalline water is shown in Fig. 17; when water is present in several states, multiple bands may be seen in the spectrum.

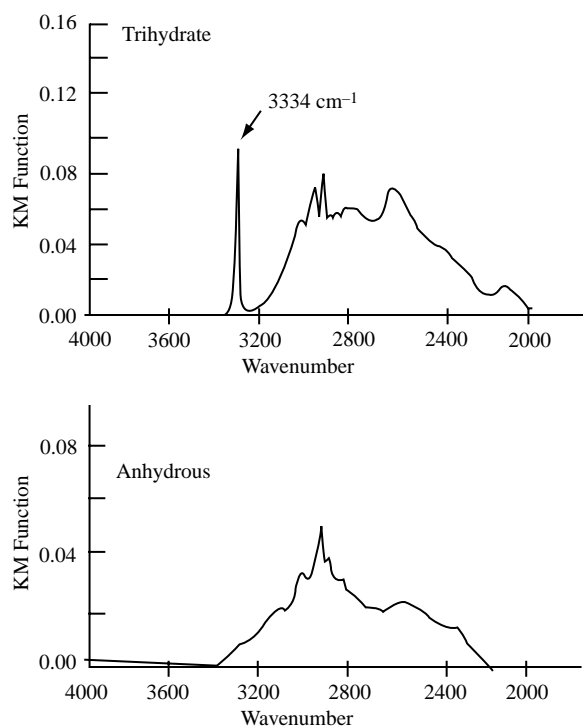


Fig. 17 Partial FTIR spectra of ampicillin. (Adapted from Ref. 36.)

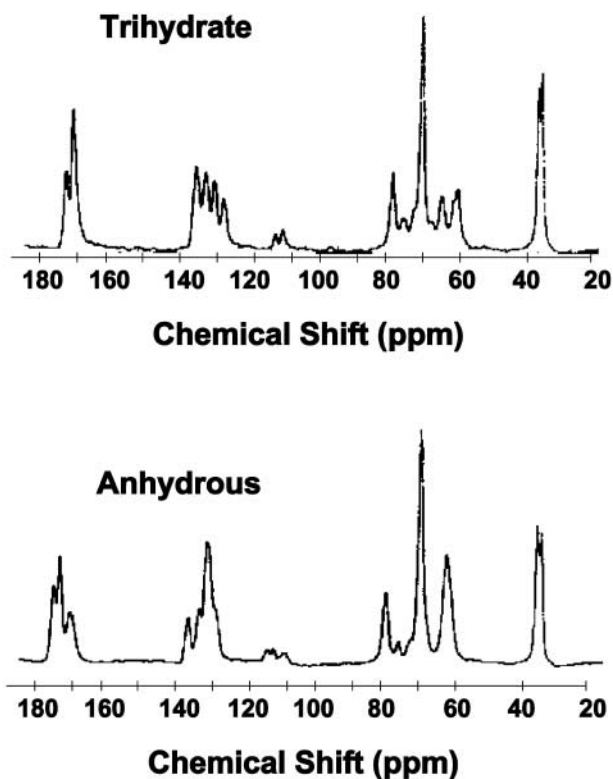


Fig. 18 Solid-state ^{13}C NMR spectra of ampicillin samples. (Adapted from Ref. 36.)

Solid-state NMR is a method that promises to increase the understanding of the state of water in solids and its specific influence on the chemicals of interest. In Fig. 18, for example, it is clear that the solid-state ^{13}C NMR spectra of ampicillin trihydrate is a very different form than that for the anhydrous form. Although three carbonyl resonances present in the anhydrous form (169.7, 172.7, and 174.6 ppm), it is interesting that only two (170.4 and 172.6 ppm) are fully resolved in the hydrate form (33). Thus, in the latter, two of the three carbonyls are equivalent. Evidence also exists for differences in carbons in the aromatic ring based on the presence of water of hydration.

INFLUENCE OF MOISTURE ON PRODUCT AND/OR PROCESS ATTRIBUTES

No attempt will be made to provide a comprehensive summary of the numerous studies that have been conducted and which show how water influences the processing and stability of pharmaceutical solids and dosage forms. For more information, the reader is referred

to some classical texts (10, 27, 37, 40, 41), recent general articles on the subject (5, 11, 42, 43), and those with specific emphasis on the role of water on stability (44–47) or compaction properties (48–52).

In an earlier section, the potential for a water-soluble substance to deliquesce was discussed. The emphasis here is on less obvious effects of moisture on solid dosage forms, and three associated areas that link to information presented earlier in this article are discussed: 1) moisture-induced changes in the state of the solid, 2) the effect of moisture on the performance of excipients in the manufacture of compressed tablets, and 3) the chemical stability of bioactive agents alone and in combination with excipients.

Moisture-Induced Changes

That moisture can have a dramatic effect on the physical character of a substance has been demonstrated recently by Carstensen and Van Scoik (51). Amorphous sucrose spheres were prepared by lyophilization and placed in an environment at 33% RH. Lyophilization characteristically produces a highly porous, amorphous solid cake. As shown in Fig. 19, its moisture content increases considerably in the first few days of the study. However, the moisture absorbed by the porous amorphous sucrose phase eventually caused a collapse of the structure and a commensurate reduction in moisture content because of the dramatic decrease in available “surface.” Not only did the physical structure change from a loose to a more dense

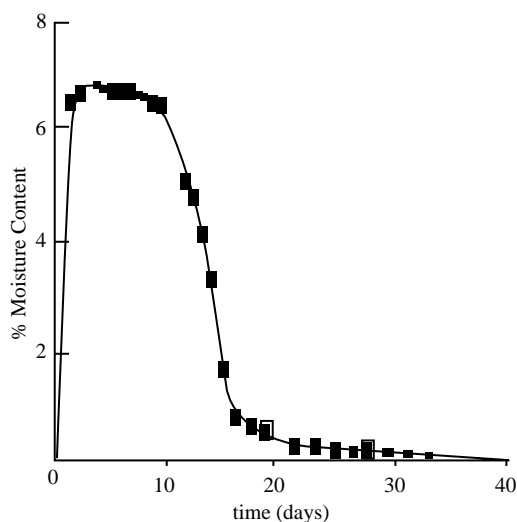


Fig. 19 Moisture uptake by amorphous sucrose at 23°C at 33%RH. (Adapted from Ref. 51.)

amorphous structure, but also the amorphous sucrose was shown to convert to the crystalline form at a rate that was dependent on relative humidity. This example is consistent with the increased emphasis being placed on changes in state of the solid as a result of sorption. One of the difficulties associated with a thermodynamic treatment of solids is the problem of dealing with changes in surface area as a continuous variable. An early thermodynamic treatment to address this problem was presented by Copeland and Young in 1961 (53). These authors considered a change in the number of moles of adsorbent as an addition or removal of particles with the same specific surface area, and, therefore, presented a basis for treating thermodynamic properties of powder systems as continuous functions. Wu and Copeland (54) used this approach in the characterization of barium sulfate. They found clear evidence to discount the “inert” adsorbent theory and stressed that even though thermodynamic variables of adsorbents are generally smaller than these for adsorbates, this can be misleading. Adsorbent properties are average properties of the respective component. When the adsorbed moisture is homogeneously distributed within the solid, this estimate is reasonable. However, if the process is adsorption and only the first few layers of the adsorbent are affected, the thermodynamic changes would be dramatically increased. Very few pharmaceutical studies involve the determination of thermodynamic properties of the adsorbent which is a promising area for further research.

The need to address changes in adsorbent is discussed by Zografi (5). He observes that water absorbed into the bulk structure of a solid can act as a plasticizer and depress the glass transition temperature. At temperatures above the glass transition point, the mobility of molecules or segments of molecules in the system increases (5). The change from the “glassy state” to the “rubbery state” can account for a number of physical chemical processes of pharmaceutical interest, including the collapse and subsequent crystallization of lyophilized cakes (see example previously), direct compaction properties, powder caking, permeability of coatings and packaging materials, and solid-state chemical stability (5). Recognition of this fact has been the single greatest recent advance in establishing a framework for understanding and predicting the impact of moisture.

Zografi and Kontny (22) corrected the experimentally determined monolayer capacities of microcrystalline cellulose for degree of crystallinity and found reasonably consistent values. This result supported the conclusion that water in microcrystalline cellulose is confined to the noncrystalline regions (22, 56).

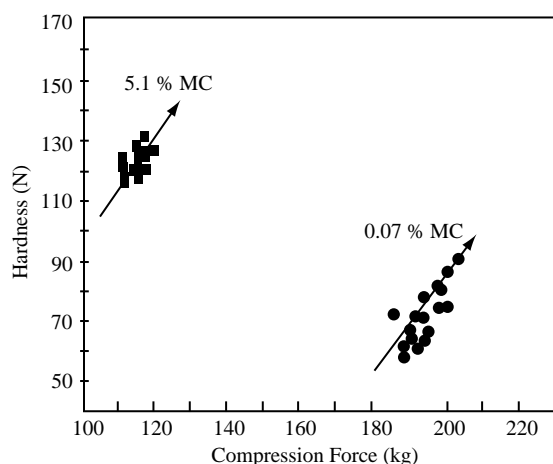


Fig. 20 Compression force vs. hardness plots for microcrystalline cellulose tablets with 0.07 and 5.1% MC. (The points represent individual tablets collected during manufacture.)

Effect of Moisture on Excipients and Tablet Manufacture

The bulk properties of celluloses are generally influenced by adsorbed moisture (58, 59). The effect of the change in bulk solid properties for microcrystalline cellulose has been demonstrated by a tableting operation in a very simplistic manner. Dry microcrystalline cellulose (%MC = 0.07) was compared with material with a moisture content above that associated with completion of the monolayer (%MC = 5.1). A thermodynamic picture of the character of water in these samples can be based on the adsorbate thermodynamic properties: $\Delta H > 3.5$ kcal/mole (14.65 kJ/mole), $\Delta G > 2.3$ kcal/mole (9.6 kJ/mole), $\Delta S < 4.12$ entropy units (e.u.) per mole with the dry solid, and

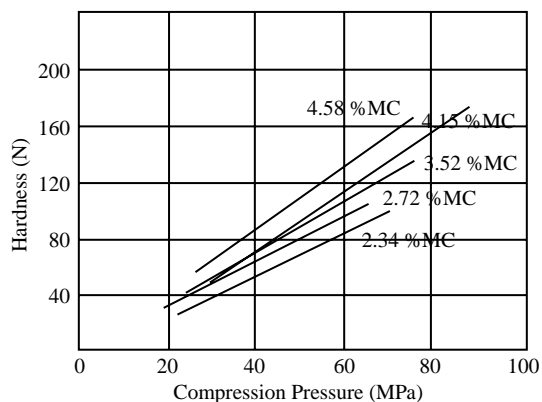


Fig. 21 Hardness vs. compression pressure for compressible sugar with different moisture contents. (Adapted from Ref. 31.)

$\Delta H = 1.5$ kcal/mole (6.27 kJ/mole, $\Delta G = 0.53$ kcal/mole (2.21 kJ/mole), $\Delta S = 3.35$ e.u./mole for moisture at the 5.1% level, or 1.5 times monolayer capacity. (The differential entropy goes through a maximum near the level where a monolayer is completed, and the large difference in free energy between the two states can be accounted for primarily in terms of the bonding of the water to the solid.)

In each case, preconditioned material was placed in the hopper of an instrumented tablet machine and the performance of these two materials was compared at constant machine settings. The sample with 5.1% MC produced tablets that weighed slightly less (an effect on flow and bulk density), and as a result were exposed to lower compression force. However, the moist microcrystalline cellulose resulted in harder tablets, even though it had been exposed to lower compression force. The loci of points on the compression force versus hardness profile (Fig. 20) for the tablets indicate a different fundamental behavior for the two materials.

In a similar, yet more extensive study, the compaction of compressible sugar was examined for materials preconditioned at different relative humidities. The hardness versus compression profiles (Fig. 21) for these samples show a group of lines whose slope appears to be a function of moisture content. This relationship is also demonstrated in Fig. 22, where the slope is used as a compressibility index. This index is a linear function of moisture content; samples with “desorbed” moisture did not differ from those with adsorbed moisture.

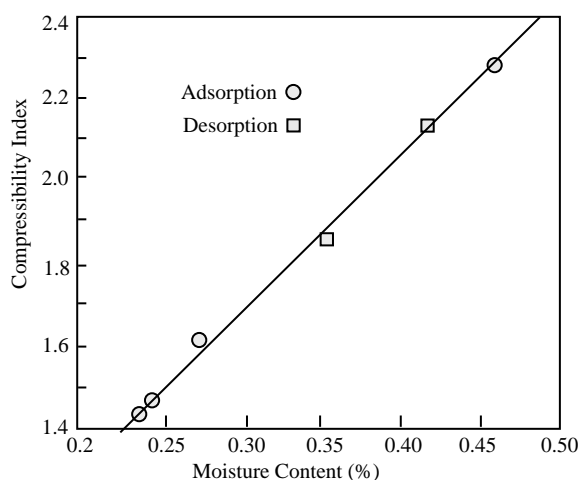


Fig. 22 Compressibility of compressible sugar as a function of moisture content. (Adapted from Ref. 31.)

Chemical Stability of Bioactive Agents

Despite recent advances in understanding the influence of moisture on the physical state of the solid, it is perhaps the effect of sorbed moisture on the chemical stability of moisture-sensitive drugs that is most important, particularly because many new bioactive agents are expensive moisture-sensitive proteins.

The literature related to the study of the influence of moisture on the properties and stability of proteins is extensive and growing rapidly. The examination of water associated with proteins and polymers is relevant to the pharmaceutical scientist dealing with the formulation and processing of small drug molecules as well because many excipients used belong in this large-molecule category. A review of protein stability from the pharmaceutical perspective has been presented by Hageman (44). He ascribes the effect of water content and activity on the solid-state stability of proteins to

- Changes in dynamic activity.
- Changes in conformational stability.
- Participation of water as a reactant or inhibitor, and
- Participation of water as a medium for mobilization of reactants.

Using an analysis of the hydration data of lysozyme and other proteins, Hageman (44) describes critical ranges of hydration based on certain properties of bound water (carboxylate absorbance at 1580 cm^{-1} , amide-I shift at 1660 cm^{-1} , OD stretching frequency at 2570 cm^{-1} , specific heat capacity, diamagnetic susceptibility). Below monolayer capacity, these physical properties of water do not change significantly, and this water has very little

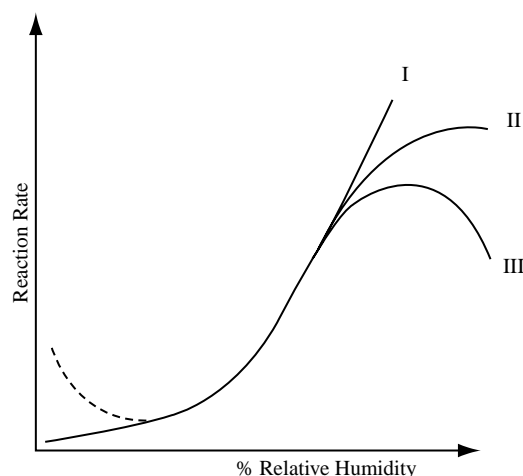


Fig. 23 Effect of sorbed water on the reaction rates of bimolecular reactions in the solid state. (Adapted from Ref. 43.)

mobility. Between 6% and 25% water, the properties of bound water change dramatically, above 25% water, the properties of bound water are similar to bulk water.

In Fig. 23, the role of moisture in bimolecular reactions is classified by Hageman into three cases. The increases in reaction rate are attributed to a change in state of the water associated with the solid as reflected by a lower effective viscosity. In Case I, there is a continual increase in reaction rate with increasing water content above the monolayer. When all the reactant has been solubilized and further water dilutes the medium, Case II results. If the dilution is extensive, or if water is a product inhibitor of decomposition, a rate reduction can be observed (Case III). Case III behavior is an example of the effect of moisture on the progress of the Malliard reaction for the glucose-containing formulations of α -N-acetyl-L-lysine, poly-L-lysine, insulin, casein, and plasma proteins (43). The fact that there can be a maximum degradation rate at a humidity other than 100% RH is observed in other situations as well.

The presence of excipients in a formulation can influence product stability. The conceptually appealing strategy of including a “moisture scavenger” in a formulation is based on this. In glucose-containing systems, it was demonstrated that liquid and solid humectants can influence the mobility of water in the system. The location of the maximum rate of reaction was found to vary from 40% to 80% RH, depending on the additives (43). The addition of liquids such as glycerol or propylene glycol lowered the mobilization point and facilitated the reaction at lower humidities. The addition of the solid humectant sorbitol reduced the reaction rate dramatically by decreasing free water for mobilization of reactants.

Carstensen (11) has presented a summary of the effects of moisture on the stability of smaller molecules by addressing three cases: decomposition with nondepleting moisture, simultaneous decomposition in the solid and dissolved states, and decomposition with limited amounts of water. In the context of this discussion, emphasis here is placed on hydrolytic degradation processes.

The first case with abundant moisture was initially modeled by Leeson and Mattocks (55) by assuming that the solid particle was surrounded by a layer of moisture sufficiently mobile to permit dissolution of the drug substance. If the layer remains saturated with drug, this approach yields an apparent zero-order degradation process given

$$-\frac{dM}{dt} = kS[\text{H}_2\text{O}] \quad (8)$$

where $-dM/dt$ is the rate of drug loss, k is a rate constant appropriate for the order of the reaction, S is the saturation

concentration of the drug in the layer, and $[H_2O]$ is the concentration of water. Although this simple approach often suffices in form to describe data observed in a stability study, it should not be expected to account for the effects of moisture present in the different states described previously.

The degradation of aspirin powder after 120 days at 100% RH and 25°C, for example, was found to be more than ten times greater than what would be expected based on suspension data (48). In addition, the actual rate of aspirin degradation has been found to increase with time (60,61). Carstensen presented a theory to describe this nonlinear behavior based on the fact that the formation of salicylic acid actually exposes moisture adsorption sites of higher energy on aspirin (61), thereby leading to the acquisition of more water. This explanation accounts for an increase in rate with time, but does not explain why the rate is higher for the powder than a suspension where the supply of water is in huge excess. This increased rate could be due to specific acid catalysis in a film of relatively small volume where the pH is dramatically affected by the hydrolysis itself (61).

Finally, the complicated situation faced when a moisture sensitive drug is combined with an hydrophilic excipient is shown in Fig. 24. In this study, aspirin was combined with 4% of the disintegrant and stored for 120 days at 25°C at the respective humidity. The moisture

contents stabilized within the first few days, and the degradation results are presented as moles of the degradation product salicylic acid per mole of water. Point A is the result for aspirin powder alone at 100% RH.

These results demonstrate many of the points presented here. At 100% RH, the reaction rate is lower than that expected for croscarmellose and sodium starch glycolate, because the disintegrants swelled and in doing so diminished the available moisture (crospovidone does not swell). Normalizing based on the total moisture present overstates the amount of water available to react and underestimates the actual reaction rate. At the other extreme (20% RH), the analysis is also biased by overstating the amount of water available, but for a different reason. The water associated with the system here is tightly bound to the disintegrant and not available. At intermediate humidities, depending on the excipient, the observed rate for degradation of the solid is higher than expected. This behavior may be a consequence of the activated sorption process and a specific acid catalysis described by Carstensen (61). Nevertheless, results of this sort are disconcerting for the product development pharmacist interested in making a stable tablet that disintegrates quickly.

SUMMARY

Real progress has been made in the last 10 years, providing pharmaceutical scientists with a solid basis for understanding the interaction of water with solids of pharmaceutical interest. Much of this progress has been the consequence of a paradigm shift: the model of the solid as an inert substrate is almost never valid. Further characterization of the state of water in solid-water systems may ultimately provide a basis for the design of stable formulations and permit the establishment of performance-based specifications for pharmaceutical excipients.

REFERENCES

1. *Encyclopedia Americana*; Grolier: Danbury, CT, 1991; 28, 432.
2. *Webster's Ninth New Collegiate Dictionary*; Merriam-Webster Inc.: Springfield, MA, 1990; 763.
3. Mitrevej, A.; Hollenbeck, R.G. *Int. J. Pharm.* **1983**, *14*, 243.
4. Kontny, M.J. *Drug Dev. Ind. Pharm.* **1988**, *14*, 1991.
5. Zografi, G. *Drug Dev. Ind. Pharm.* **1988**, *14*, 1905.
6. Gibbs, J.W. *The Collected Works of J. Willard Gibbs*; Yale University Press: New Haven, 1948; 1.

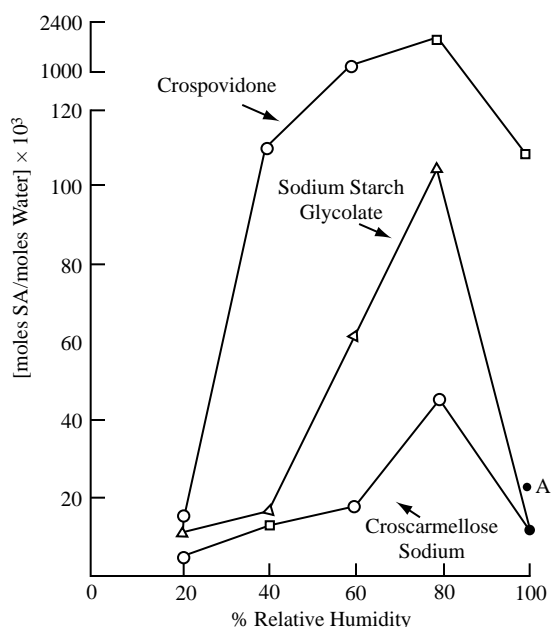


Fig. 24 Effect of humidity on the salicylic acid content in aspirin-disintegrant mixtures at 120 days and 25°C (A, aspirin powder). (Adapted from Ref. 60.)

7. *The United States Pharmacopeia 24th Rev., The National Formulary*, 19th Ed. United States Pharmacopeial Convention, Inc. Rockville MD, 2000.
8. Hollenbeck, R.G. Determination of Water: Compendial Viewpoint Proceedings of the Seventh Wisconsin Update Conference, Carstensen, J.T. Ed.; Extension Services in Pharmacy, University of Wisconsin, Madison, 1988; 77, 91.
9. Zografi, G. *Pharmaceutical Forum*, Nov-Dec: 3240 (1987).
10. *Handbook of Pharmaceutical Excipients*; American Pharmaceutical Association: Washington, DC, 1986.
11. Carstensen, J.T. *Drug Dev. Ind. Pharm.* **1988**, *14*, 1927.
12. Hammond, W.A. *Drierite and its Applications*; The Stoneman Press: Columbus, OH, 1958.
13. Van Campen, L.; Amidon, G.L.; Zografi, G.J. *Pharm. Sci.* **1983**, *72* (12), 1381.
14. Van Campen, L.; Amidon, G.L.; Zografi, G. J. *Pharm. Sci.* **1983**, *72* (12), 1388.
15. Van Campen, L.; Amidon, G.L.; Zografi, G. J. *Pharm. Sci.* **1983**, *72* (12), 1394.
16. Brunauer, S. *The Adsorption of Gases and Vapors*; Princeton University Press: Princeton NJ, 1943; 1.
17. Brunauer, S.; Emmett, P.H.; Teller, E. J. Am. Chem. Soc. **1938**, *60*, 309.
18. Hill, T.L. *Adv. Catal.* **1952**, *4*, 211.
19. Cassel, H.M. J. Chem. Phys. **1944**, *12*, 115.
20. Cassel, H.M. J. Phys. Chem. **1944**, *48*, 195.
21. Guggenheim, E.A. *Applications of Statistical Mechanics*; Clarendon Press: Oxford, 1986.
22. Zografi, G.; Kontny, M.J. *Pharm. Res.* **1986**, *3*, 187.
23. Dunford, H.B.; Morrison, J.L. *Can. J. Chem.* **1955**, *33*, 904.
24. Argue, G.H.; Maass, O. *Can. J. Res.* **1935**, *12*, 564.
25. Babbitt, J.D. *Can. J. Res.* **1942**, *20*, 143.
26. Morrison, J.L.; Dzieciuch, M.A. *Can. J. Chem.* **1959**, *37*, 1379.
27. Gregg, S.J. *The Surface Chemistry of Solids*; Reinhold Publishing Co.: New York, 1951.
28. Young, J.H.; Nelson, G.L. *Trans. Am. Agric. Eng.* **1967**, *10*, 260.
29. Young, J.H.; Nelson, G.L. *Trans. Am. Agric. Eng.* **1967**, *10*, 756.
30. York, P. J. *Pharm. Pharmacol.* **1981**, *33*, 269.
31. Tabibi, S.E.; Hollenbeck, R.G. *Int. J. Pharm.* **1984**, *18*, 169.
32. Hollenbeck, R.G.; Peck, G.E.; Kildsig, D.O. J. *Pharm. Sci.* **1978**, *67*, 1599.
33. Marshall, K.; Sixsmith, D. *Drug Dev. Commun.* **1974-75**, *1*, 51.
34. Mitrevej, A.; Hollenbeck, R.G. *Pharmaceut. Technol.* **1982**, *6* (10), 48.
35. Connors, K.A. *Drug Dev. Ind. Pharm.* **1988**, *14*, 1891.
36. Brittan, H.G.; Bugay, D.E.; Bogdanowich, S.J.; DeVincen-tis, J. *Drug Dev. Ind. Pharm.* **1988**, *14*, 2029.
37. Adamson, A.A. *Physical Chemistry of Surfaces*, 4th Ed.; John Wiley & Sons: NY, 1982.
38. Harkins, W.D.; Jura, G. J. Am. Chem. Soc. **1944**, *66*, 919.
39. Franks, F. *Water: A Comprehensive Treatment*; Plenum Press: New York, 1972; 1 and 5, 1975.
40. Carstensen, J.T. *Pharmaceutics of Solids and Solid Dosage Forms*; Wiley: New York NY, 1977.
41. Connors, K.A.; Amidon, G.L.; Stella, V.J. *Chemical Stability of Pharmaceuticals*; Wiley-Interscience: New York, 1986.
42. York, P. *Int. J. Pharm.* **1983**, *14*, 1.
43. Umprayn, K.; Mendes, R.W. *Drug Dev. Ind. Pharm.* **1989**, *13* (4-5), 653.
44. Hageman, M.J. *Drug Dev. Ind. Pharm.* **1988**, *14*, 2047.
45. Plotkowiak, Z. *Pharmazie* **1989**, *44* (2), 837.
46. Fassihi, A.R.; Persicaner, P.H.R. *Int. J. Pharm.* **1987**, *37* (1-2), 167.
47. Hageman, M.J.; Water Sorption and Solid-State Stability of Proteins. *Stability of Protein Pharmaceuticals*; Ahern, T.J., Manning, M.C., Eds.; Plenum Press: New York, 1992.
48. Shukla, J.; Price, J.C. *Pharm. Res.* **1991**, *8* (3), 336.
49. Ahlneck, C.; Alderborn, G. *Int. J. Pharm.* **1989**, *56* (2), 143.
50. Teng, C.D.; Alkan, M.H.; Groves, M.J. *Drug Dev. Ind. Pharm.* **1986**, *12* (11-13), 2325.
51. Carstensen, J.T.; Van Scoik, K. *Pharm. Res.* **1990**, *7*, 1278.
52. Groves, M.J.; Teng, C.D.; The Effect of Compaction and Moisture on Some Physical and Biological Properties of Proteins. *Stability of Protein Pharmaceuticals*; Ahern, T.J., Manning, M.C., Eds.; Plenum Press: New York, 1992.
53. Copeland, L.E.; Young, T.F. *Adv. Chem. Ser.* **1961**, *33*, 348.
54. Wu, Y.C.; Copeland, L.E. *Adv. Chem. Ser.* **1961**, *33*, 357.
55. Leeson, L.J.; Mattocks, A.M. J. Am. Pharm. Assoc. Sci. Ed. **1958**, *47* (5), 329.
56. Zografi, G.; Kontny, M.J.; Yang, A.Y.S.; Brenner, G.S. *Int. J. Pharm.* **1984**, *18* (1-2), 99.
57. Ben-Rayana, E.; Seddas, A.; Ruelle, P.; Ho, N.T.; Kesselring, U.W. *Mol. Cryst. Liq. Cryst.* **1990**, *187*, 617.
58. Blair, T.C.; Buchton, G.; Breezer, A.E.; Bloomfield, S.K. *Int. J. Pharm.* **1990**, *63* (3), 257.
59. Sadeghnejad, G.R.; York, P.; Stanley-Wood, N.G.; Water Vapour Interaction with Pharmaceutical Cellulose Powders. *Pharmaceutical Technology: Drug Stability*; Rubenstein, M.H., Ed.; Ellis Horwood Limited: West Sussex, UK, 1989.
60. Mitrevej, A. *A Comprehensive Study of Water Vapor Sorption by Aspirin-Disintegrant Blends and its Relationship to Aspirin Stability*; Ph.D. Thesis, 1982.
61. Carstensen, J.T.; Attarchi, F. J. *Pharm. Sci.* **1988**, *74* (4), 314.

MIXING AND SEGREGATION IN TUMBLING BLENDERS

Troy Shinbrot

Fernando J. Muzzio

Rutgers University, Piscataway, New Jersey, U.S.A.

MOTIVATION

Mixing of solids is essential to many industries, including pharmaceuticals, ceramics, metallurgy, chemicals, food, cosmetics, coal, and plastics. To give an idea of the magnitude of applications involving granular processes, U.S. worldwide production alone accounts for over a trillion kilograms of granular and powdered products annually, most of which must be uniformly blended to meet quality and performance goals. In this article, we present an example-oriented overview of current understanding of mixing and segregation mechanisms that are of importance to powder blending operations. We focus on industrial tumbler designs, which simultaneously comprise the bulk of solids blending operations and represent the systems for which predictive modeling appears to have the greatest potential. We direct the reader to existing literature sources (e.g., Harnby, 1997) for information on more specialized blending equipment. Numerous distinct mechanisms for both mixing and demixing of granular materials have been catalogued including convection, diffusion, shear, percolation, etc., and in most applications, several mechanisms act concurrently and interact in complex, and currently poorly understood, ways. Thus the mode by which powders are loaded into blenders of common design can alter the time needed to homogenize them by two orders of magnitude. Given that a certain blender can be designed to deliver acceptable performance in the laboratory, we have no consistent a priori mechanism to scale up the process and achieve the same performance in blenders of industrial size. Although comprehensive and predictive understanding of practical blending problems still remains a long-term goal, it has recently become possible to define models that generate respectable agreement with observations in practical granular devices (e.g., 3D tumblers), and methods have become sufficiently refined that systematic techniques to analyze new products and equipment are available, either off the shelf or as research-ready hardware.

FUNDAMENTALS

Research on granular flow and blending roughly can be divided chronologically: prior to about 1990, industrially usable results were mostly empirical (e.g., in experiments using a particular blend in a specific device), and fundamental research was largely analytic (e.g., using continuum approximations to the granular state applicable only to one phase of granular behavior). Although important progress was made into developing specialized engineering solutions and fundamental physical properties of granular systems, little generally applicable knowledge was attainable using either approach. Since that time, computational and methodological advances have permitted quantitative evaluations of granular flow, transport, and mixing at an unparalleled level of detail and accuracy. In this Section we have reviewed progress on tumbling flow and blending phenomenology that has led to the development of the best predictive models existing.

Definition of the “Granular State”

A chief limitation, and the principal area of opportunity for the future, in developing predictive understanding of granular flows is the coexistence of multiple, history-dependent, granular states. Within a device—be it a tumbler, a high shear intensifier, a mill, a fluidized bed, etc.—granular material can, and typically will, exhibit multiple rheologically different phases that vary nontrivially and often with profound consequences as a function of minor changes in material or environmental variables. This is a particular problem in the pharmaceutical industry, where products may be developed in dry, northern, latitudes and produced in wet, equatorial climates. Both hygroscopic excipients and actives behave very differently in these two environments, and blending regimens that work in one may fail in the other. Moreover, even within a single well-controlled bench-scale device, multiple phases are typically present. The tumbling blender is a case in point.

In Fig. 1, a deceptively common outcome of an attempt to blend dissimilar materials is shown, here of grains



Fig. 1 “Left-right” segregated state, here in a transparent V-blender, between larger (dark) and smaller (light) grains. This state occurs spontaneously at high fill levels and fast tumbling speeds in many tumbler designs.

differing only in size and color. In this transparent 4-L capacity^a V-blender, we have tumbled equal volumes of small light-gray and larger black grains at 6 rpm for 200 revolutions. The visibly segregated state is only one of several distinct segregated configurations that form spontaneously and reproducibly in all common blender geometries and scales, and persist despite the practitioner’s best efforts at modification of process parameters. Developing cures for this type of problem demands a systematic understanding of why segregated states occur in the first place, so that the cause of segregation can be addressed directly. This understanding in turn requires an analysis of the different granular behaviors seen during the tumbling operation.

A first step in the analysis of granular behaviors is the characterization of the different granular phases that are inevitably present during flow. Grains, unlike common fluids, must “dilate” in order to flow, that is, grains in the static state are interlocked, and cannot move without separating. The locations and timing of dilation can be quite complex, but once dilation occurs, the overt granular behavior seen remains static and solid-like far from the regions of dilation, and becomes respectively glassy (disordered), fluid-like, or gas-like near the shear interface as the shear rate increases. The modifier, “like,” is important to include, because a solid-like region is not truly elastic as it transmits stress along irregular compressive chains, undergoes slow creep and settling on time scales ranging from seconds to hours, and can solidify into a rigid “cake” over time scales of days to months. Likewise the fluid-like phase transmits shear discontinuously both in

space and time, and does not obey Navier–Stokes equations, and the gas-like phase is far from equilibrium and is not characterized by Maxwell–Boltzmann statistics. It is the differences between qualitative behaviors of different regions of a granular bed at different times and between any one of the behaviors and accepted models for flow and dispersion that make predictive understanding of even the simplest granular systems challenging.

Despite the intrinsic difficulties in developing an all-encompassing model for granular flow, important blending problems of practical interest have been effectively analyzed using model-based, computational, and semiempirical means. The current understanding of granular blending and demixing is summarized in subsequent sections.

Elementary Two-Dimensional Mixing Mechanisms

In tumbling applications, dilation and flow principally play out near the unconstrained upper surface of a granular bed, and the bulk of grains beneath are thought to remain nearly motionless during rotation of the blender. This simplified picture changes for some blender varieties (e.g., Fig. 2b), but predictive models for blending in most common blending geometries can be derived by disregarding all transport beneath the free surface. In the following sections, the best existing models and methods are summarized and their application to common tumbler designs is described.

Although it differs significantly from more complex practical blender geometries (see sections on Competitive Patterned Demixing and Examples), the horizontal drum tumbler has a simple geometry and therefore provides a useful environment for a first analysis of granular mixing and demixing. A horizontal drum mixer is the simplest form of a tumbling blender and it is used in many pharmaceutical, chemical, and metallurgical industries in the form of ball mills, dryers, rotary kilns, coating pans, and mixers.

Flow regimes

Flow in rotating drums has been described qualitatively in terms of regimes such as slipping, avalanching, rolling, cascading, cataracting, and centrifuging.

Slipping: The slipping regime occurs when the granular bed undergoes solid body rotation and then slides, usually intermittently, against the rotating tumbler walls. This most frequently occurs in simple drums that are only partially filled, and is often counteracted by including baffles of various designs along the inner walls of the tumbler. Although the slipping regime is not important for blending purposes per se, it is encountered even in effective blending systems, and an evaluation of the

^aThe reader should note that “capacity” customarily refers to a fraction (generally 60%) of the total interior volume of a blender.

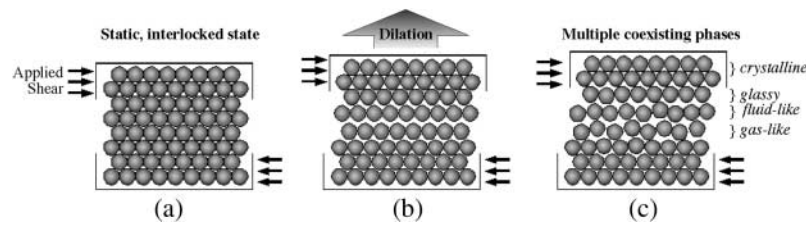


Fig. 2 Schematic of dilation mechanism that is a prerequisite for the flow of solids. (a) In undisturbed state, grains are interlocked and behave much like an ordinary solid. (b) A granular bed dilates in response to applied shear, and can then flow. (c) In the flowing state, the bed can form distinct crystalline, glassy, fluid-like and gas-like phases. The crystalline phase is regular and ordered, the glassy phase is disordered but static, the fluid-like state flows but exhibits enduring contacts, and the gas-like state is characterized by rapid and brief interparticle contacts.

number of times a bed is turned over per tumbler revolution will often reveal the presence of some slipping.

Avalanching: A second regime seen at slow tumbling speeds is the “avalanching” flow regime, also referred to as “slumping.” In this regime, flow consists of discrete avalanches that occur as a grouping of grains travels down the free surface and comes to rest before a new grouping is released from above. The avalanching regime is not seen in large tumblers (larger than a few tens of centimeters in diameter), but is an instructive case because flow and mixing can be solved in closed form for simplified drum geometries, and lessons from this regime can, with due caution, be carried over to more realistic systems.

To analyze this problem, one only need observe that if the angle of repose at the free surface immediately before an avalanche is ϑ_i , and after an avalanche is ϑ_f , then the effect of an avalanche is to carry a wedge of material in the angle $\vartheta_f - \vartheta_i$, downhill, as sketched in Fig. 3a for an idealized 2D disk blender. The same caricature applies at

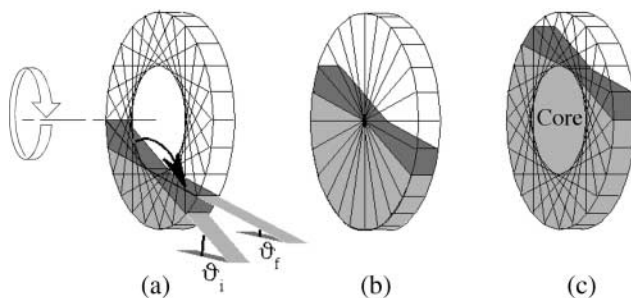


Fig. 3 (a) Avalanching flow in an idealized disk tumbler transports grains from an uphill wedge to a downhill wedge as the free surface relaxes from an initial angle ϑ_i to a final angle ϑ_f . This implies that global mixing occurs in quadrilateral regions where grains within one wedge intersect with a second wedge. (b) Consequently, tumblers mix more efficiently at low fill levels than at high, and global mixing nearly stops at 50% fill. (c) At fill levels above 50%, a core develops that does not visit the avalanching surface and, therefore, does not experience transport or mixing.

all fill levels, and one can readily use this model to make several concrete predictions. First, mixing occurs during avalanches through two distinct mechanisms: one, particles within a wedge rearrange during a single avalanche, and two, particles rearrange globally between wedges during successive avalanches. Second, at 50% fill (Fig. 3b) no two avalanching wedges intersect, so no global mixing between separated regions can exist, and mixing must be slow. Third, since flow occurs only near the avalanching surface, at high fill levels a nonmixing core necessarily develops (Fig. 3c). Although this model is oversimplified and neglects material variations, boundary effects, and other important physics, these conclusions carry over to more realistic tumbling systems.

Rolling: At higher tumbling speeds, discrete avalanches give way to continuous flow at the surface of the blend. Grains beneath the cascading layer rotate nearly as a solid body with the blender until they reach the surface. One can solve for flow and transport subject to certain simplifying assumptions in this regime as well. For this solution, one assumes that the grains are so small as to be regarded as a continuum and one takes the free surface to be nearly flat, as sketched in Fig. 4a. The interface between the flowing layer and the supporting bed beneath has been determined experimentally and through detailed computations to be roughly parabolic in shape (1) and by demanding mass conservation at this interface one can construct continuum flow equations for this system (2). If one simulates the mixing in an idealized disk blender of mechanically identical grains initially separated by color to left and right of a vertical central plane, one obtains the results (for a particular fill level and flowing layer depth) as shown in Fig. 4b. Corresponding experimental results are shown in Fig. 4c.

Cascading, cataracting, and centrifuging: For larger tumblers, or for tumblers rotated at higher speeds, the surface is manifestly not flat, as shown in Fig. 5 in a 1 meter diameter disk tumbler (3). This flow, termed cascading, differs qualitatively from the rolling flow solution: here the flowing layer is thin and is nearly

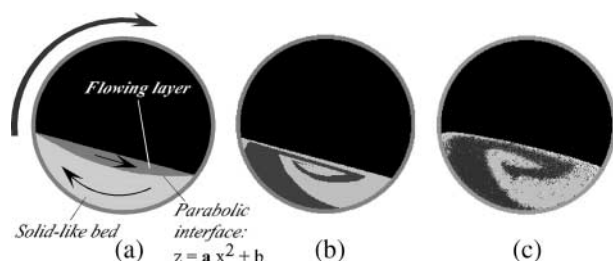


Fig. 4 (a) In the rolling regime, the blend separates into a flowing layer near the surface and a solid-like supporting bed. (b) By establishing simple conditions such as mass conservation, one can generate an analytic model for the flow, producing mixing patterns between initially separated and colored, but otherwise identical, grains. (c) Comparison with experimental mixing patterns using freely flowing grains in a small drum tumbler reveals substantial agreement. The snapshot in (c) is obtained from the interior of the blend using the solidification technique described in the section on Solidification in this article.

uniform in speed and thickness, and has been modeled as depth averaged, plug-like flow. As the rotation speed of the tumbler is increased, the surface becomes increasingly sigmoidal until grains become airborne, and at higher speeds yet, the grains centrifuge against the tumbler wall. These regimes are termed cataracting and centrifuging, respectively, and have not been well analyzed.

MIXING MECHANISMS IN THREE-DIMENSIONAL TUMBLERS

Although drum blenders represent a convenient paradigm for the purpose of categorizing granular behaviors, most

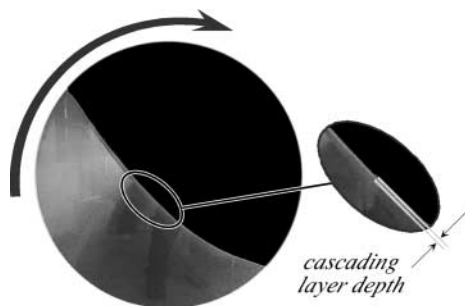


Fig. 5 Cascading flow occurs in large tumblers or during tumbling of fine, but freely flowing, grains. This snapshot shows a 1 m diameter transparent disk tumbler partially filled with colored $\sim 500\ \mu\text{m}$ irregular grains. The free surface is manifestly not flat, and the cascading layer is thin and nearly uniform with distance along the flowing surface.

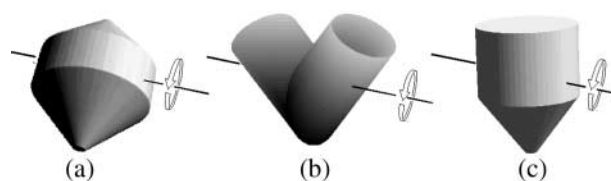


Fig. 6 Three common tumbler designs: (a) double-cone, (b) V-blender, and (c) tote, or bin blender.

blending operations occur in more complex tumbler geometries. Three of the most common geometries used in pharmaceutical operations are the double-cone, the V-blender, and the tote blender, sketched in Fig. 6. Each of these geometries have many variants, for example symmetry can be broken to introduce cross-flow by slanting the double-cone, by elongating one of the arms of the V-blender, or by inserting baffles in a tote.

To model flow and blending in complicated geometries, particle-dynamic simulations have been implemented. In these simulations, particles are treated as individual entities with physical properties (e.g., size, static and dynamic friction coefficients, coefficient of restitution, etc.) appropriate to the problem of interest, and Newton's laws of motion are integrated for each particle. Particle-dynamic simulations are similar in concept to molecular-dynamic simulations, but include features of importance to the flow of macroscopic particles (e.g., static and dynamic friction models) in place of microscopic properties (e.g., bond strengths and chemical potentials). Particle-dynamic simulations come in many different types depending on how they treat physical parameters such as rolling friction and particle shape or numerical issues such as search algorithms and routines to maintain computational stability. Results of distinct computational simulations can differ, sometimes significantly, and the importance of experimental validation of numerical results cannot be overemphasized.

Two of the most common classes of particle-dynamic simulations are termed "hard-particle" and "soft-particle" methods. Hard-particle methods calculate particle trajectories in response to instantaneous, binary collisions between particles and allow particles to travel ballistically between collisions. This class of simulation permits only instantaneous contacts, and is consequently often used in rapid flow situations as are found in chutes, fluidized beds, and energetically agitated systems. Soft-particle methods, on the other hand, allow each particle to deform elastically and compute responses using standard models from elasticity and tribology theory. This approach permits enduring particle contacts and is therefore the method of choice for tumbler applications. The simulations described in this article use soft-particle methods

that have been validated and found to agree in detail with comparison experiments (4, 5).

Convection

Mixing in all tumbling blenders consists of a fast convective stage, driven by the mean velocity of many particles, followed by a much slower dispersive stage, caused by velocity fluctuations leading to rearrangements of individual particles. Convection in grains (as in fluids) is by far the fastest and most efficient mixing mechanism, yet at the same time it suffers from the same mixing limitations known for fluids: convective flows can, and very often do, possess barriers to mixing (e.g., islands) that do not interact with surrounding material. Two pathologies are readily observed: overfilled mixers develop elliptic, nonchaotic, islands that rotate as a unit in the center of the granular bed (discussed earlier in Section on Flow Regimes), and symmetric blenders (seen in most standard designs) exhibit separatrices that divide the flow into noninteracting sectors. Beyond this, little is known currently of details of particle flow patterns and mixing barriers in practical, three-dimensional blender geometries. There is strong evidence, though, indicating that flow bifurcations analogous to those seen in fluids may be present in granular tumblers.

Convection in the context of granular blenders refers to transport associated with flow driven by gravity (in tumbling blenders) or impellers (in intensified, ribbon, or other blenders). Convection is observed in all functioning blender geometries, and can be visualized using particle-dynamic simulations. In Fig. 7 are shown successive front and side views taken a quarter revolution apart of 20,000

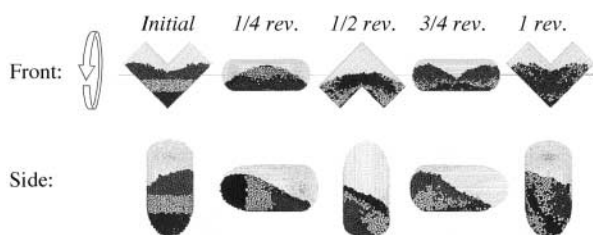


Fig. 7 Rapid, convective flow seen in particle-dynamic simulation of identical but colored spheres in V-blender. Top: view from front reveals that unlike in some designs, convection in this blender drives grains axially, alternately outward toward the tumbler arms and inward toward its center. This axial flow strongly influences mixing, as described in the section on Mixing Rates in this article. Bottom: view from side indicates that transport is dominated by a spiraling flow, seen also in drums and other blenders (cf. Fig. 4).

identical but colored spheres tumbled in a V-blender in the cascading regime. These snapshots illustrate the qualitative motion produced in this blender, which causes the bed to overturn from top to bottom. Mixing due to convective flow grows linearly with time insofar as the area of an interface between differently colored layers in these snapshots or in Fig. 4b and c, grows characteristically linearly with time. Although similar qualitative behaviors are seen in all tumbler geometries, the quantitative mixing seen can differ considerably between geometries (cf. Section. Mixing rates).

Dispersion

Dispersion, also referred to as diffusion, is contrasted with convection, which can effectively intersperse grains in a tumbler within tens to hundreds of revolutions. Dispersion also refers to the random relocation of individual grains due to collisions between adjacent particles and can take hundreds to thousands of revolutions to act. Thus particles cross a plane separating the two arms of the V-blender (or an equivalent symmetry plane in many other blender geometries) only as a result of occasional collisional happenstances and not as a result of an overall mean flow. Various stratagems, including the use of baffles, asymmetric cross-flow designs (referred to earlier), irregular rotation protocols, or axial rocking, have been introduced to mitigate this limitation. Notwithstanding these improvements, dispersion is the rate-limiting mechanism for mixing, and there is much potential for improvement of dispersive mixing.

Although convection is typically orders of magnitude more rapid than dispersion, as a practical matter the relative contribution of each mechanism to blending is strongly influenced by the initial distribution of species in the mixer. Thus ingredients loaded in horizontal layers (as in Fig. 7) can be mixed relatively rapidly, while ingredients layered side by side (either intentionally (as in Fig. 8) or inadvertently (as a result of careless loading of a tumbler) will typically mix enormously more slowly.

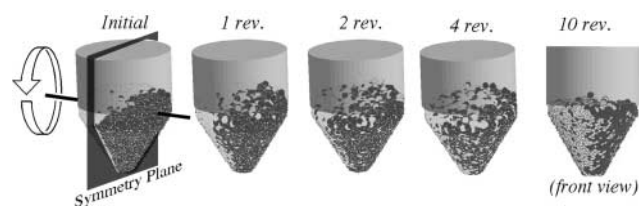


Fig. 8 Dispersive mixing is slow across the symmetry plane of a blender, here a tote design. After 10 revolutions, a front view reveals clear evidence of the initial left-right distribution of identical but colored spheres in this particle-dynamic simulation.

To visualize this effect, in Fig. 8 is shown the dispersive mixing of 8000 identical but colored grains loaded side by side, here in a tote blender. With each successive revolution, only a few particles cross the interface separating the two symmetric halves of the tumbler, and as a result, after 10 revolutions the original particle ordering is still unmistakable. Systematic assays obtained from experiments of blending of realistic pharmaceutical excipients and actives confirm that imperfectly loaded blends retain any initial asymmetry for many hundreds of tumbler revolutions.

Shear

A final class of mechanisms of granular and powder mixing is grouped under the category of shear. This includes a host of very different mechanisms that act in a shearing layer (cf. Fig. 2) or in a high-shear region as produced by an intensifier or related device. Grains rearrange during shearing and so disperse across a shear layer. In addition, large or irregular grains tend to be expelled from regions of high shear through a mechanism known in research on suspensions as “shear-induced migration.” Finally, arrangements of adjacent grains become distorted through the influence of external strain. This is the most classic form of shear mixing, which itself admits several distinct subcategories.

First, there is the simple effect of shear on deforming a granular bed. In simple tumblers, this can manifest itself in a nearly continuous distribution of strain, resulting in regular and predictable mixing behavior such as is displayed in Fig. 4c. This figure was obtained using freely flowing grains, of mean size nearly 1 mm.

Second, a granular bed consisting of weakly cohesive materials (e.g., nontacky grains in the size range 50–300 μm) exhibits stick-slip motion so that flow becomes intermittent rather than continuous. This is a situation of practical importance because pharmaceutical blends, for example, use particles across a broad range of sizes and materials. As the size of grains diminishes or as interparticle cohesion grows, stick-slip flow transforms mixing interfaces from smooth, regular, patterns as shown in Fig. 9 (500- or 700- μm cases) to a complex, irregular pattern, shown in Fig. 9 (300- or 100- μm cases). In simple geometries, this response to shear can be accurately modeled: if we assume that the flowing surface of a bed periodically sticks and slips, then the mechanism displayed in Fig. 4a can be embellished by allowing the shear band between flowing layer and bed to deform periodically (2). This produces patterns of mixing between initially separated but identical grains that are substantially similar to experiment, as shown at the bottom of Fig. 9.

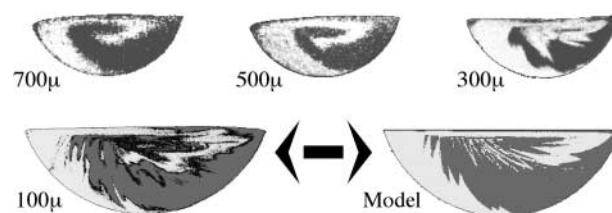


Fig. 9 Mixing patterns after 1 revolution in drum tumblers loaded with identical (except for color) grains in four experiments using successively finer grains, as well as in a model simulation of idealized stick-slip flow. At 700 and 500 μm , the mixing interface remains smooth and regular, while below about 300 μm , it becomes variegated due to intermittent slipping of the cascade. Each experimental snapshot shows a view from the interior of a blend using the solidification technique described in the section on Solidification in this article, and all cases began with light grains to the left of center and dark grains to the right.

This is important for blending because in smooth, regular flow, adjacent particles remain nearby for long periods of time, while in intermittent, stick-slip flow, particles can rapidly relocate across the blender. It is not difficult to show that periodic sticking and slipping results in an exponentially rapid growth of interfaces between separated regions of grains.

Third, for particles smaller than about 100 μm , cohesive forces (believed to be due to van der Waals interactions for intimate contacts, and to surface tension of adsorbed water layers for lubricated contacts) between particles becomes comparable to particle weights, and small particles can stick to one another in relatively rigid aggregates. Unless such aggregates are destroyed, the system will behave as if it had an effective particle size much larger than the primary particle size.

For strongly cohesive materials, it is typically necessary to fragment agglomerates through the introduction of high shear, “intensification,” devices such as impellers or mills that energetically deform grains on the finest scale. Intensification is commonly performed in an early preblending stage using a fraction of the total desired excipient to avoid overblending the final product. Results of intensification are discussed in the section on Assays.

DEMIXING

Processing the blends of dissimilar grains almost invariably promotes demixing, also referred to as segregation, characterized by the spontaneous emergence of regions of nonuniform composition. Segregation due to differences in particle size in a blend has drawn the

greatest attention in the literature, including studies of fluidized beds, chutes, hoppers, vibrated beds, and tumbling blender. Segregation due to particle density, shape, and triboelectric order have also been recorded. As a practical matter, segregation manifests itself in granular mixing that characteristically improves over a brief period (while convection generates large scale mixing) and then degrades, often dramatically (as slower segregational fluxes take over). Demixing should not be confused with the phenomenon of overblending, which is also frequently encountered in blending applications. Overblending is associated with physical degradation of material properties, as occurs for example when a waxy lubricant is excessively deformed causing it to coat pharmaceutical grains and reduce their bioavailability, or when coated granules are damaged through abrasion or fracture.

At the present time, mechanisms for segregation even in the simple tumbling drum remain obscure, and work on more complex and industrially common blender geometries is extremely limited. Three distinct types of demixing are moderately well characterized in tumblers. They are radial demixing, axial demixing, and competitive patterned demixing. We describe each of these in turn.

Radial Demixing

Segregation typically proceeds in two stages. First, large grains rapidly segregate radially, producing a central core of fine grains surrounded by larger grains, identified in Fig. 10 for a simple drum tumbler. Unlike the core seen in overfilled tumblers (Fig. 3c), this core appears at fill levels under 50% and is exclusively associated with migration of fine grains toward the center of an overturning blend. Radial segregation is seen in both quasi-2D and fully 3D blenders of various geometries. In simpler 3D geometries, such as the drum, double-cone or tote, the core is nearly always apparent when blending significantly dissimilar grains, while in more complicated geometries such as the V-blender or slant-cone the core becomes significantly distorted and may only be conspicuous for higher fill levels or in certain (e.g., upright) orientations of the blender. Even in the simplest case of the drum tumbler, however, the location and dynamics of the core remain somewhat enigmatic: as shown in Fig. 10, the core is actually located upstream of the geometric center of the granular cascade.

The core appears to form as a result of two cooperative influences. First, smaller grains percolate through the flowing layer to occupy successively lower strata each time the bed overturns. Second, once a sufficient volume of smaller grains has accumulated, the larger grains tend to roll increasingly freely over the (comparatively smooth)

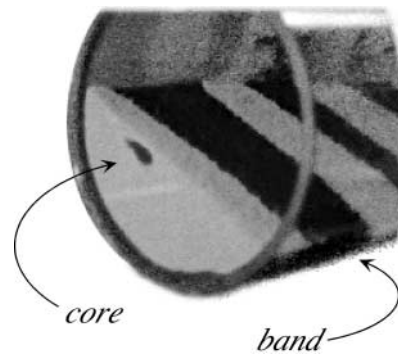


Fig. 10 Typical segregation pattern seen between fine (dark) and coarse (light) grains in small transparent drum tumbler. A core of fines extends along the entire length of the tumbler, connecting the bands that emerge at the surface in a single bulging tube. The coarse grains are constrained to flow within the confines defined by this tube. This constraint is important for understanding mechanisms of demixing in more complex geometries, as described in the section on Competitive Patterned Demixing in this article.

substrate of smaller grains. This higher speed surface flow reinforces the segregated state by expelling remaining slower small grains. These mechanisms are very robust, and cores are almost invariably found in tumbling of freely flowing grains with diameter ratios between about 1:1.5 and 1:7. As the diameter ratio approaches unity, the core becomes more diffused, while as the diameter ratio grows sufficiently large, fine grains can percolate increasingly freely through a matrix of larger grains or, if sufficiently fine, can coat the larger species.

Axial Demixing

A second stage of segregation occurs in drum tumblers as grains in the core migrate along the tumbling axis. Numerical and experimental investigations have attributed this migration to conflicting causes, e.g., a secondary flow within the core leading to a bulging of the core toward the surface versus different angles of repose of fine, mixed, and coarse grains. Whatever the ultimate cause, the result of this axial migration is the formation of a series of bands as shown in Fig. 10. In this final state, two pure phases of material are formed, divided by sharp boundaries with very little intermixing.

Competitive Patterned Demixing

In more complex, and more common, tumbler geometries, several distinct segregation patterns have been observed. These patterns are believed to arise from a competition between surface segregation of coarse grains flowing over

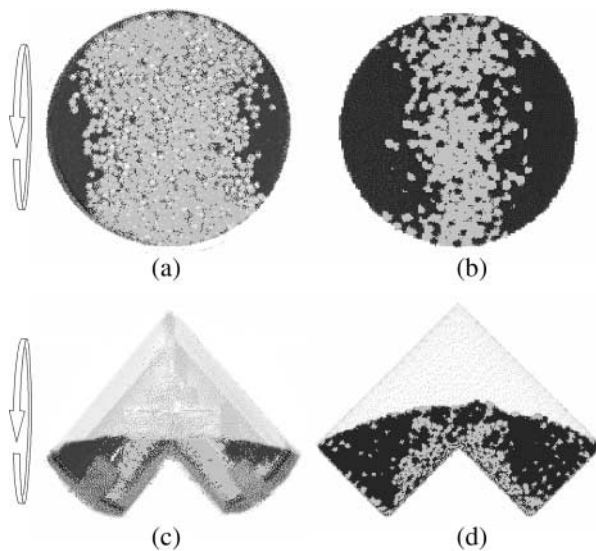


Fig. 11 Axial segregation in top views of double-cone blender from (a) experiment and (b) particle-dynamic simulation using large (light) and small (dark) spherical grains. Similar patterns are seen in other tumbler designs, for example in the V-blender in (c) experiment and (d) simulation.

a radially segregated core of fine grains and interactions with the boundaries of the tumbler. Despite significant differences between common blender geometries, there is substantial commonality in the ultimate patterns seen. For example, mixing of large, light-gray and small, dark-gray grains in a double-cone and a V-blender generates similar patterns in both experiments and particle-dynamic simulations, as shown in Fig. 11.

As parameters such as fill level, tumbler speed, and concentrations of the different particle species are varied, the patterns observed though change significantly. Importantly, there appear to be a few dominant and recurring patterns that are seen both in experiments and in simulations in all blender geometries. Notably at high fill levels and tumbling speeds, the “left-right state” shown in Fig. 1 appears to dominate. This pattern and two other common variants are shown at the top of Fig. 12 in top views of the surface of a double-cone blender. Each of these patterns appears reproducibly and spontaneously whenever different size grains are tumbled in any of several blender geometries. Simulations shown beneath the experimental figures in Fig. 12 use a continuum model in which large particles are convected on the surface of an idealized convex bed of smaller grains. Container geometry is included by assuming that large particles rebound specularly when they reach the downstream boundary of the idealized blender. Correspondence

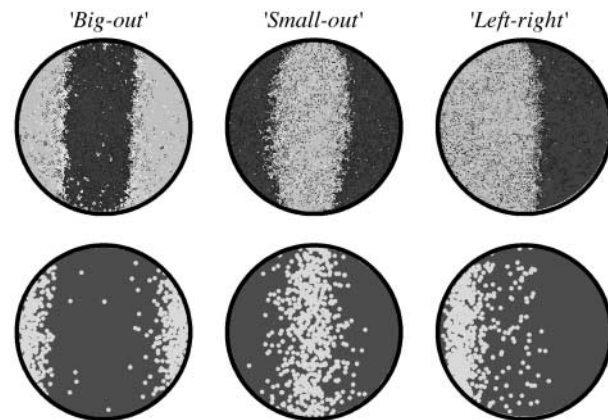


Fig. 12 Three common segregation patterns between large (light) and small (dark) grains seen in top views of double-cone blender. Top: experimental snapshots. Bottom: simplified continuum simulations.

between experimental data and this simulation indicates that ongoing improvements in modeling show promise for unveiling the underlying mechanisms of demixing and permitting eventual accurate modeling of practical granular processing systems.

Scaling

An ultimate goal of granular research is to enable processing to be scaled up from bench to pilot to full-scale operations. Scale-up methods are well developed in other, (e.g., fluid) operations, but remain a distant objective in granular systems. Several alternative scale-up approaches have been suggested. For example, the natural dimensionless group associated with granular flow at a known speed, v , down a specified height incline, h , under gravity, g , is the Froude number, defined to be the ratio of kinetic to potential energies: $Fr = v^2/gh$. Experimental analysis reveals, however, that neither mixing nor segregation scale with Fr . Reasons for this are several: the competing mixing mechanisms of convection, diffusion, and shear each act on different characteristic time scales, and their interactions are therefore complex in the extreme. Additionally grains accelerate down a cascading surface, and therefore the resulting behavior changes qualitatively depending on whether they do or do not reach an asymptotic speed in the cascade (for example, compare Figs. 4c and 5). These and other influences make first-principles analysis of granular scale-up problematic.

Nevertheless, reproducible empirical scaling relations do appear to be attainable for specific blenders. As an example, in the drum blender it has been determined that dynamically similar convective flow occurs provided that

$R \cdot \Omega^{2/3}$ is held constant for lower tumbling speeds (where particles reach a terminal speed in the cascading layer) and provided that $R \cdot \Omega^{1/2}$ is held fixed for higher tumbling speeds (where particles do not reach a terminal speed), where R is the drum radius and Ω is the tumbler rotation speed. Likewise the demixing patterns shown in Figs. 1, 11 and 12 are found at all blender scales, and dynamical similarity relations have been obtained for transitions between these patterns. These encouraging findings indicate that scaling of a particular granular blending process in a specified geometry may be possible, but that scaling relations can be expected to vary quantitatively as process parameters (e.g., material properties, fill levels, tumbler geometries, etc.) are changed. As in any rapidly evolving field, the current state of understanding is not yet fully adequate for practical needs, but research trends show promise for the intermediate and longer terms.

MIXING MEASURES

A prerequisite to meaningful evaluation and interpretation of mixing is the use and understanding of a reliable measure of mixing. Though this concept may seem straightforward, some care needs to be exercised in its implementation. Any mixing measure is obtained by first evaluating a relevant quantity, typically concentration, in specified sample regions. Ideally, in order for the samples to be representative, they should be taken uniformly from a flowing stream that is itself uniform both in space and time. In tumblers, this is not practical, and very often sampling consists of extracting small aliquots of grains from a static bed. We discuss techniques, and limitations, of extracting such aliquots shortly, but first it is worthwhile to review what one would do if provided with complete access to concentration data throughout a granular bed. Such an idealized scenario is sketched in Fig. 13, where we display sequential snapshots of a 2D bed. Suppose we can subdivide the blender into identical boxes—elements of area in 2D, elements of volume in 3D. In each box, let us define a concentration, C_a , of a species of interest. Thus C_a might range from 0 in the black regions of Fig. 13, to 1 in the light-gray regions.

Data Interpretation

The first thing to notice is that not all boxes are completely occupied by grains, and boxes near the periphery of the bed can bias results strongly. For example, in the leftmost snapshot, partially filled boxes consist of anomalously high concentrations C_a , and an unweighted evaluation of

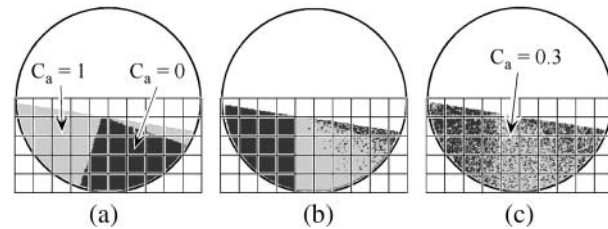


Fig. 13 (a) Subdivision of blend into sampling regions. By determining concentrations, C_a , of interest in each region, one can compute relevant statistics. (b) and (c) as time progresses, concentrations approach a mean value with standard error that diminishes as the number of sampled regions increases.

concentrations in all boxes would yield a mean concentration, for this particular case, of 60% light-gray grains. This is clearly wrong, and can be corrected either by reducing the box size (hence diminishing the fraction of boxes containing a boundary of the bed) and excluding boundary boxes or, preferably, by weighting each box based on its fractional fill level.

A second issue is that there is necessarily a trade-off between precision and resolution. That is, if the box sizes are reduced, one obtains more boxes and hence more values of C_a , and consequently any mixing measure will in principle have a lower standard error. On the other hand, this reduction has an obvious limit as the box size approaches the grain size, at which point any blend will statistically appear to be unmixed (because each box can only take on one of the two values, $C_a = 0$ or $C_a = 1$). In practice, a happy medium delivering both suitably low standard error and an adequate sample size^b is easily achieved; nevertheless it is important to understand what it is that one is seeking to obtain before deciding on a sampling protocol.

A third issue is illustrated in Fig. 14, where we plot three schematic mixing states, each containing the same fraction of light and dark regions. Plainly the leftmost state is fully segregated, but most simple measures would provide the same evaluation of mixedness for the two other states. Moreover, the rightmost schematic represents a common outcome of shearing flows (cf. striated state in Fig. 9). For this reason, it may be desirable to customize one's mixing measure to accurately reflect the desired endpoint of measurement, and correspondingly innumerable different mixing measures have appeared in the literature. These many different measures unfortunately do little to advance the comprehensive understanding of granular blending or to permit comparisons between different data sources, and it is

^bTo compute a lower bound on box size, one typically assumes Poissonian statistics for the presence or absence of particles within a given box, so that uncertainties in a box containing n particles go as.

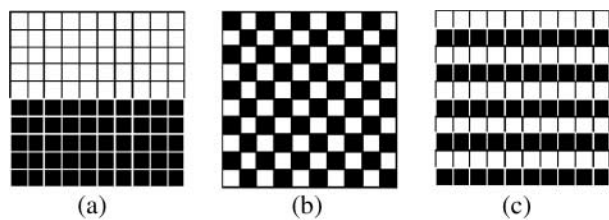


Fig. 14 Illustration of the importance of rational data interpretation. The state (b) intuitively looks well mixed, but has the same number of black and white boxes as states (a) and (c). Evidently a larger box size is required to make meaningful mixing measurements. States (b) and (c) have very different structures, but produce similar measures of mixing using common statistics. For this reason, many customized mixing measures have evolved in the literature.

consequently preferable, where feasible, to use a single set of mixing measures as suggested in the next section.

Intensity of Segregation

One of the most useful measures of mixedness is the intensity of segregation. This is little more than a normalized variance of concentration measurements. Intrinsic to the use of intensity of segregation, therefore, is the presumption that the mixing distribution is, at least to a first approximation, Gaussian. This raises two issues: first, it is not clear that granular mixing tends toward a Gaussian state, and second, in many practical applications a Gaussian is not the desired outcome. Indeed, in pharmaceutical processing, if a blend were Gaussian rather than uniform, then the unavoidable presence of exponential tails on a Gaussian distribution would guarantee that some small fraction of tablets made from the blend would be beyond any therapeutic range that one could specify. Fortunately, granular flows seem to scatter grains more uniformly than a simple Gaussian would predict, although the details and mechanisms for this behavior are not yet well understood.

With these caveats in mind, a definition of the intensity of segregation is:

$$I = \frac{\sigma^2 - \sigma_r^2}{\sigma_o^2 - \sigma_r^2}, \quad (1)$$

where σ^2 is the variance of sampled data, σ_r^2 is the variance of the same number of randomly chosen concentration data, and σ_o^2 is the variance of an initial, typically fully segregated, state, again consisting of the same number of data points. Several forms of I appear in the literature; this form is useful because it is normalized so that $I = 1$ and $I = 0$ correspond to completely segregated and randomly mixed states, respectively.

Two other mixing measures of importance to pharmaceutical processing are, first, the relative standard deviation (RSD), defined to be:

$$\text{RSD} = \frac{\sigma}{\langle C_a \rangle} \quad (2)$$

where σ is the standard deviation and C_a is the mean concentration over all samples taken, and the mixing rate, k . The mixing rate is defined according to the relation:

$$I = I_o e^{-kt} \quad (3)$$

where I_o is an initial intensity and t is time. Many, though not all, mixing mechanisms produce an exponential approach to uniformity as presupposed by Equation (3). Deviations from Equation (3) are discussed in the next Section; for now, if we assume that the intensity of segregation decays according to Equation (3), we can extract a mixing rate from any given experiment and thereby evaluate and compare mixing efficiencies of different blenders, of blenders operated under different conditions, and of effects of changes in material or other properties of interest.

Examples

Mixing rates

To examine the behavior of mixing measures, it is useful to begin by considering systems free of experimental uncertainties. Particle dynamic simulations such as those discussed in the sections on Mixing mechanism in Three Dimensional Tumblers; and Demixing, represent such ideal systems: the presence and locations of all particles are known and are free of sampling errors (discussed in the section on Sampling Techniques).

The simplest of these simulations from the point of view of mixing measurement is the double-cone. Raw variances taken from volume elements initially separated axially in the blender are shown in Fig. 15a. The variances are color coded according to the initial locations sketched in the inset. Evidently, after the first or second revolution of the blender, variances decay nearly exponentially (i.e., linearly on this semi-log plot). Mixing is slowest, however, between particles initially near the axial center of the blender (light gray in Fig. 15a). This is to be expected because, as commented earlier, the rate limiting mixing process is dispersion across the symmetry plane of the blender, or looked at another way, grains starting further from the symmetry plane experience a secondary axial convective flow, that grains closer to this plane do not.

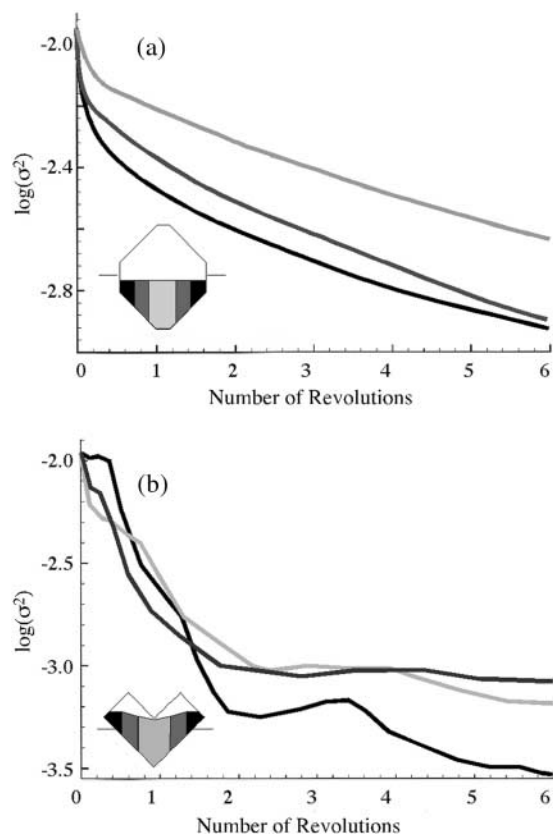


Fig. 15 Variance of concentrations of identical color-coded grains from particle-dynamic simulations of (a) a double-cone and (b) a V-blender. Notice that the smooth flow in the double-cone is reflected in a smooth, asymptotically exponential, reduction in variance with time, while the sloshing flow in the V-blender (cf. Fig. 7) produces periodic undulations in mixing response.

It is instructive to compare this simulation with its counterpart in the V-blender, shown in Fig. 15b, where mixing is much more irregular. This irregularity has been traced to periodic sloshing of grains to and from the arms of the blender. Again there is a barrier to mixing across the symmetry plane of the tumbler, but comparison of the scales of Figs. 15a and b reveals that the sloshing process accelerates mixing significantly. Thus as one would expect, the smooth flow seen in the double-cone is manifested in a smooth decay in variance shown in Fig. 15a and modeled by Equation (3), while the periodic sloshing of grains in the V-blender is poorly fit by such a simple relation. For all tumblers, it is notable that mixing evolves with each successive tumbler rotation and not with elapsed time, and all studies indicate that tumbler rotation rate (within a fixed regime) is nearly inconsequential.

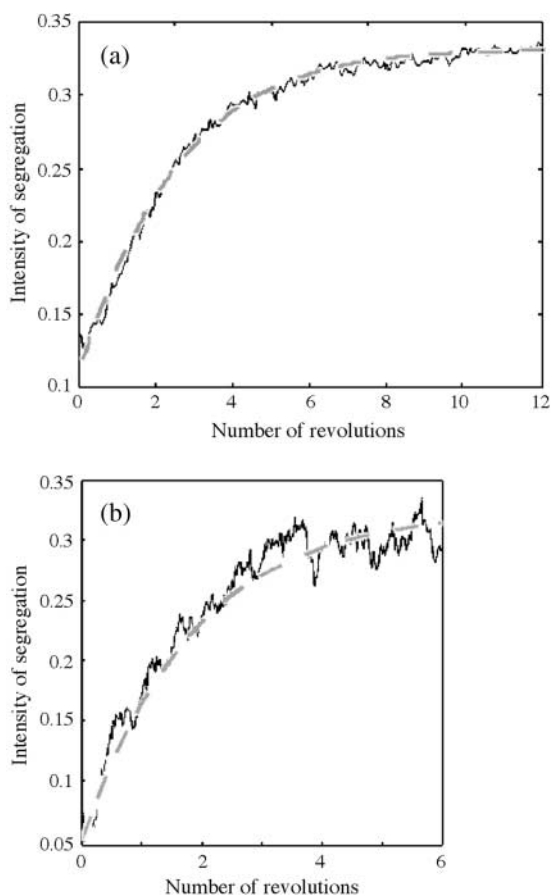


Fig. 16 Evaluation of segregation between different size particles in particle dynamic simulations in (a) double-cone and (b) V-blender.

Segregation rates

As its name suggests, the intensity of segregation is especially useful for analyzing segregating mixtures. For example, rather than evaluating I in a simulation of identical particles, as shown in Fig. 15, if we evaluate I for a simulation of the segregating mixture in the double-cone shown in Fig. 11b, we obtain the plot shown in Fig. 16a. Here the blend starts from an artificially mixed state, $I \ll 1$, and approaches an asymptotically segregated state, $I \cong 0.3$. The solid data points are obtained by dividing the total occupied volume into separate boxes and calculating the number of small grains within each box. The dashed gray curve is an exponential approach to an asymptote included for comparison, which indicates that segregation, as well as mixing, in simple geometry tumblers obeys a simple exponential relation. For more complex tumbler geometries, segregation approximately follows an exponential approach to an asymptotic

segregated state, but is more erratic, as shown in Fig. 16b for the V-blender.

SAMPLING TECHNIQUES

Thieves

In practical blending processes, one cannot obtain arbitrary quantities of pristine data as one can using particle-dynamic simulations, and one must settle for sampling a static bed, as mentioned previously. In such a case, it is especially important to understand sampling limitations and systematic biases. A common means of obtaining samples in a tumbler is by the use of a scoop or thief sampler. These samplers are inserted into the bed and extract samples from its interior. In such an application, it is necessary to obtain a representative spatial distribution of samples throughout the bed, and for this purpose it is recommended that one build a jig above the tumbler opening into which one can insert samplers at fixed locations in the plane of the opening and to specified depths perpendicular to the plane.

The behaviors of two popular types of thief samplers are shown in Figs. 17 and 18. In Fig. 17 is shown the result of inserting a “side-sampling” thief, consisting of a tube with a slot in its side that can be opened to allow grains to flow into a cavity and closed to extract the sample. In the photograph of Fig. 17a, an interior section of a solidified bed (see section on Solidification) of light gray 200- μm and dark 60- μm grains is shown. The grains are initially layered, and from this snapshot it is clear that the act of inserting the thief causes grains to be entrained along the insertion route. This entrainment causes local particle rearrangements that typically result in the bed appearing to be anomalously well mixed. It is also significant that side-sampling thieves rely on particle flow into the sampling cavity to obtain grains, and consequently freely-flowing or smaller grains can penetrate the sampling slot more readily than cohesive or larger grains. These conclusions are borne out in quantitative tests: in Fig. 17b, we plot the fraction of smaller beads in samples obtained using a side-sampling thief in separate experiments in which 60- μm grains are initially arranged in a single thick layer over a bed of 200- μm grains. Evidently the thief obtains samples almost entirely consisting of the smaller species, irrespective of the genuine concentration originally in the sampling location.

The problem that side-sampling probes do not allow larger or more cohesive grains to enter the sampling cavity can be mitigated through the use of “end-sampling” thieves such as the one shown in Fig. 18. In these thieves, the sampling tube is inserted to a desired depth in the bed, an aperture at the distal end of the probe is opened, and then

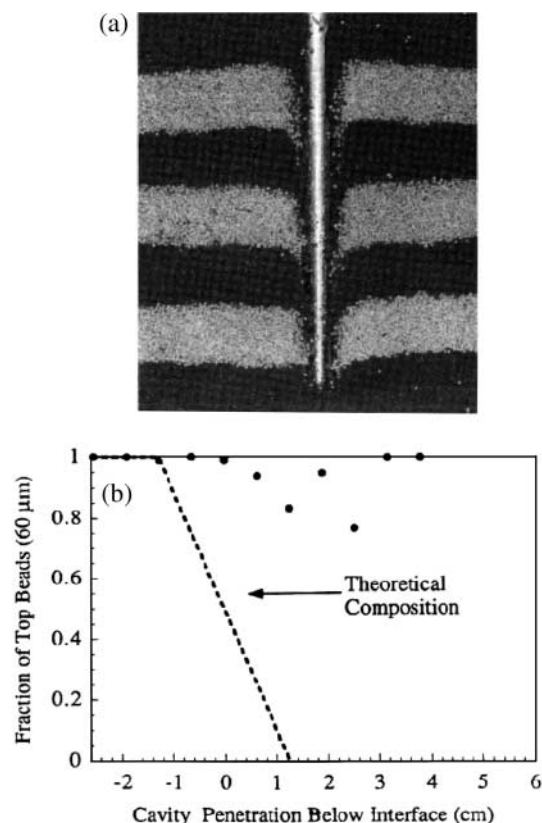


Fig. 17 Systematic sampling errors introduced by a side-sampling thief. (a) Initially layered configuration of large (light) and small (dark) grains are noticeably disturbed as this thief entrains grains during insertion. (b) This type of thief relies on free flow of grains to fill a cavity when a slot is opened in the side of the sampling tube. Consequently, fine and freely flowing grains are overrepresented by this probe, and fine grains are transported to regions where they were not originally placed.

the probe is pushed deeper into the bed and the aperture is closed again to allow extraction of the sample. Because grains are actively forced into the cavity, rather than passively flowing into it as in side-sampling thieves, this device is relatively free of differential sampling problems as can be caused by differences in particle flowability. However, as shown in Fig. 18a, these devices are typically bulky and consequently entrain considerable material during their insertion. The resulting sample concentration measurements (Fig. 18b) are therefore improved over those of the side-sampling thief, but remain very inaccurate and the data consistently overestimate a blend's mixedness.

Core Sampling

An alternative that is nearly free of either flow (e.g., Fig. 17b) or entrainment (e.g., Fig. 17a) anomalies is the

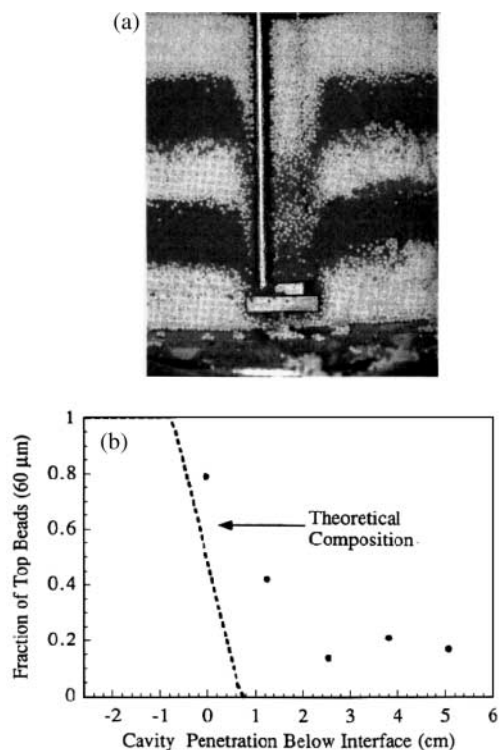


Fig. 18 Sampling errors introduced by an end-sampling thief differ from those of a side-sampling thief, but persist nonetheless. In this type of thief, a window is opened at the bottom of the sampling tube and grains are forced into a cavity by further insertion of the thief. This eliminates the bias toward grains that passively fill a cavity more easily than others, but on the other hand, (a) these thieves entrain more grains during insertion and (b) their performance again suffers from substantial systematic error.

“core-sampler.” This sampler extracts an entire contiguous core of grains throughout the depth of insertion. At its simplest, the probe consists of a thin walled tube that is inserted into a granular bed, together with a mechanized extrusion apparatus to permit samples to be extracted in a last-in, first-out manner after the tube has been removed from the bed. In the case of freely-flowing grains, which could otherwise flow out of the tube, the device is embellished by incorporating an end-cap that can be opened during insertion and then closed during extraction. Unlike the end-sampling thief, the end-cap mechanism here is internal to the sampling tube, and an entire core is extruded from the bed. The behavior of this device is demonstrated in Fig. 19. Using the end-cap (shown closed in Fig. 19a), the concentration data obtained compare favorably with other methods, as shown in Fig. 19b. Importantly, in the core sampler the core extends through the depth of the sampling tube and so sample sizes are large. Random uncertainties

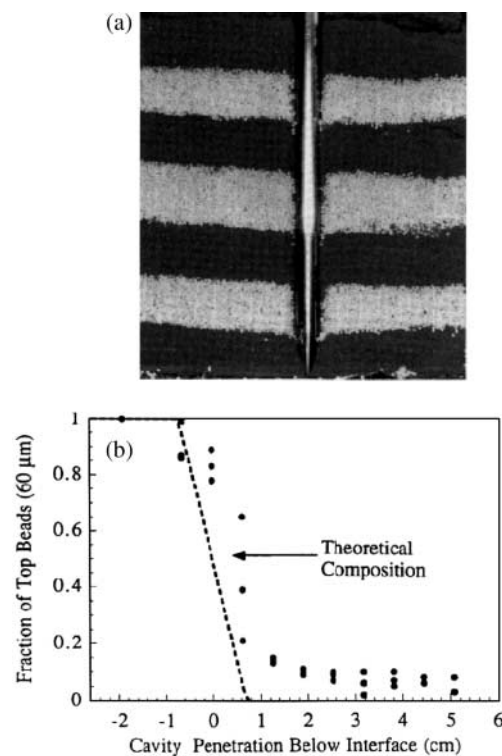


Fig. 19 Core sampler with end-cap can be used for freely-flowing (e.g., granulated) materials that would escape from the sampling tube during removal from the bed without the end-cap. (a) Very little entrainment is visible after insertion, and (b) systematic errors are reduced.

are therefore small and corrections for systematic errors that remain can be predictably made.

Without the end-cap, agreement between experiment and theory is further improved. In Fig. 20 we display core sampling results for three different inner diameter sampling tubes using a two-layer bed of common pharmaceutical excipient powders: microcrystalline cellulose and lactose. For all sampler diameters, the experimental data are indistinguishable from ideal expected concentrations. In practice, we note that it is important that the walls of the sampling tubes be polished (to prevent excessive entrainment and difficulty filling the tube during insertion) and that a well-regulated extrusion device be employed.

Assays

Once samples have been obtained, one can use a variety of available chemical, optical, spectroscopic, chromatographic, or other assays to determine concentrations of interest. For example, data in Fig. 20 were obtained using a calibrated densitometric technique in which one

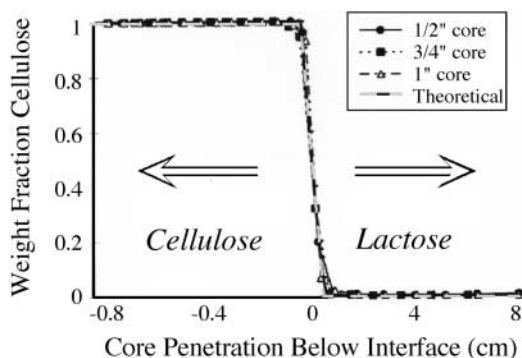


Fig. 20 Non-freely-flowing, powdered material can be extracted from a blend using a core sampler with an open end. In this case, measurement errors are virtually undetectable, here in a sampling experiment using a thick layer of microcrystalline cellulose above a bed of lactose.

of the two species was colored in advance; identical results have been obtained using other assay techniques, e.g., reflection near-infrared spectroscopy used to evaluate distributions of magnesium stearate (a common pharmaceutical lubricant). One particularly useful technique for evaluating the effectiveness of tumbler design or process modifications involves the use of conductivity assays. For this technique, standard nonionic excipients are blended with mock-actives consisting of simple soluble salts (e.g., KCl or NaCl) granulated in the size of interest. Samples extracted from the interior of the blend using core-samplers (inserted at controlled locations using a jig as described previously) are then extruded in fixed weight aliquots and dissolved in a known quantity of solvent (typically just deionized water). The concentration of the solution is then measured using commercially available probes, and by comparing concentration data with calibration curves obtained separately, one can rapidly and accurately establish mixing statistics of the mock-active as a functions of operational parameters of interest.

As an example, we have mentioned that it is typically necessary to fragment cohesive agglomerates through the use of an intensifier bar (a high speed impeller) or similar device. To investigate the effectiveness of intensification during preblending, in Fig. 21 we show plots of RSD of mock active vs. blending time in a V-blender with and without intensification. Each data point is obtained from a separate experiment in which the blender is loaded in predefined layers and then tumbled for the specified time, after which multiple samples are taken (again always from the same locations in the bed) and assayed. Because the samples are taken using entire intact cores, large numbers of data are obtained throughout the depth of the bed.

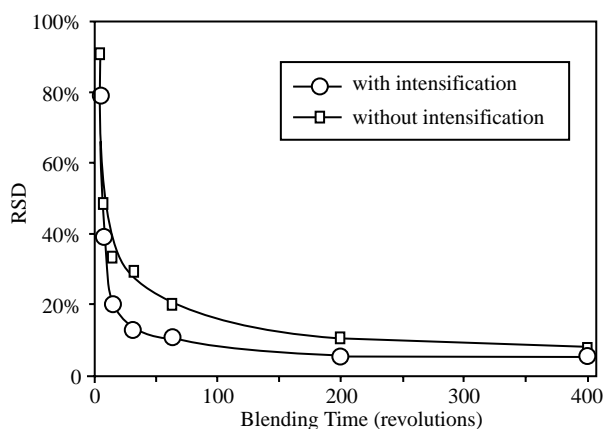


Fig. 21 Mixing rates in a V-blender tumbled at 24 rpm with and without shear induced by a high speed intensifier bar.

Typically nine cores are taken, distributed uniformly in a grid around the blender opening; since the V-blender has two arms, in this set of experiments five cores were taken through openings at the end of each arm.

By solidifying an entire blend, one can obtain data in enormous quantities and in exceedingly high detail. For example, in Fig. 23, we show the interior of a blend in a double-cone solidified using a methacrylate solution and then sliced open through the tumbler symmetry plane. Here $\sim 90\text{-}\mu\text{m}$ grains have been blended beginning with a segregated initial state, and the mixing structure, characteristic of stick-slip flow (described earlier), can be resolved down to the size of the grains themselves.

This type of sampling protocol permits the quantification of mixing under a variety of different conditions. As we have mentioned, because core sampling provides large quantities of bed material in a reproducible manner, one can assay some of the samples and retain others for later analysis—e.g., to perform dissolution or bioavailability tests or for archival purposes.

Solidification

Core sampling combined with calibrated assaying techniques represents the best available technology for quantification of mixing of industrial scale blenders. Other techniques are available, however, for laboratory scale studies. The gold standard for these is the solidification technique, described here. To solidify a granular bed, one performs a mixing experiment using colored or otherwise distinguishable grains and infiltrates the blend with a polymeric solution such as methacrylate copolymer, low viscosity epoxy, or heated gelatin. The polymer is allowed to set, and the solidified monolith is

sliced open using a saw (for large samples solidified with methacrylate or epoxy) or a knife (for samples set in comparatively soft gelatin) to reveal internal mixing patterns. Separate experiments using structured (e.g. intentionally layered) blends confirm that the blend is not disturbed by infiltration provided that the infiltration is carried out slowly and from one end of the tumbler to the other (so as to avoid trapping bubbles). This technique is appropriate only for laboratory scale apparatus, in which setting times are short and it is feasible to handle and slice the entire frozen blend. Additionally, the technique typically involves sacrificing the blending vessel (which is obviously not always practical), although blends solidified in gelatin can be released from the blender by heating it from outside.

Typical results are shown in Fig. 22 using a segregating mixture of fine, dark, and coarse, light, grains in a drum tumbler solidified using gelatin. The slices displayed show clear evidence both of radial and axial segregation. These slices incidentally also reveal an evident shortcoming of attempting to visualize mixing only from the outside of the blend. This has been done many times in the literature, but without validating the results by examining the interior of the bed, such studies can be grossly misleading, as the differences between the exterior and interior slices shown in Fig. 22 illustrate.

Other Techniques

Other, more technologically involved, techniques have also been developed for visualizing the interior of granular beds. These include:

- Diffusing wave spectroscopy, used to measure statistics of fluctuations in relatively thin, Hele-Shaw configurations
- Positron emission tomography, in which a single radioactive grain is tracked during flow within a granular bed using an array of external photomultipliers
- Magnetic resonance imaging, in which magnetic moments of hydrogenated grains are aligned in structured configurations (e.g., stripes) and these structures are tracked for short periods of time
- X-ray tomography, in which a population of radio-opaque grains are tracked in a flow of interest.

These techniques are typically expensive and cumbersome to implement; nevertheless they reveal flows within an optically opaque bed and provide valuable information not attainable otherwise. For example, in Fig. 24, we display results of x-ray tomography experiments that show the evolution of the interior mixing structure within a double-cone blender using molybdenum-doped tracer particles (dark in Fig. 24). These experiments represent a

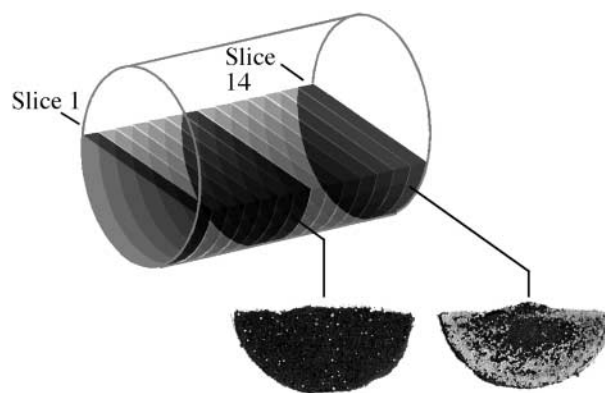


Fig. 22 Internal and external slices obtained by solidifying a blend, here of coarse (light) and fine (dark) grains in a drum tumbler. Segregation is evident, and the difference between internal and external slices highlights the importance of not relying on external appearances of tumbled blends.

scaled-up version of the solidification data shown in Fig. 23: the capacity is 8 times larger (~ 4.8 L vs. 0.6 L), and the particle diameter is 18 times larger (~ 1600 vs. 90 μm). Data of this kind reveal a complexity in flow and mixing evolution that simultaneously represents the cause

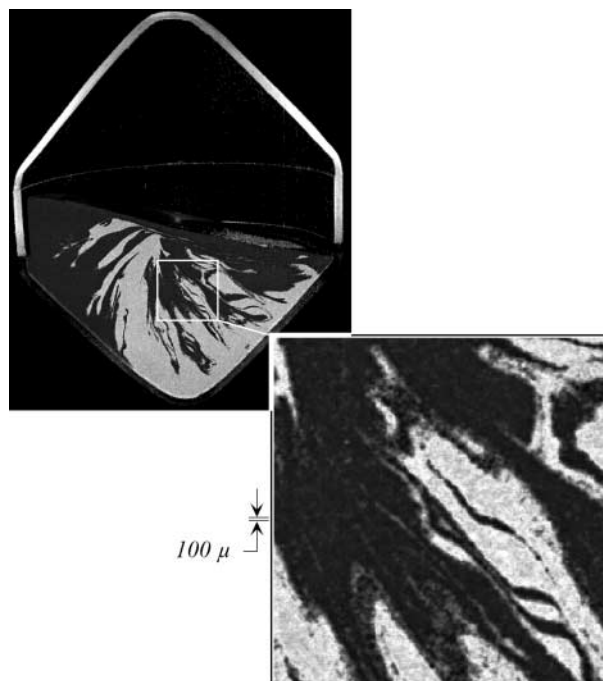


Fig. 23 High-resolution data are obtainable by solidifying and slicing an entire blend, here of identical but colored ~ 90 - μm grains in a double-cone blender.

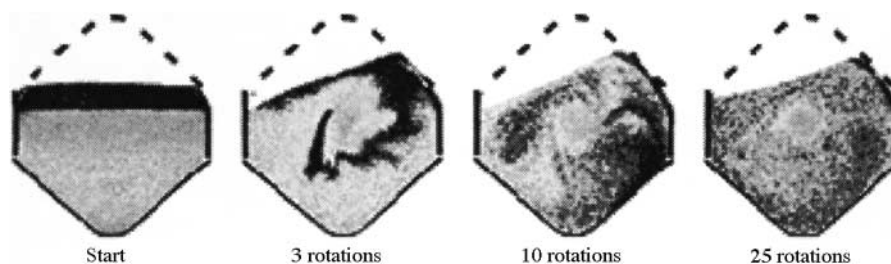


Fig. 24 X-ray tomographic time series of blending of radio-opaque grains in double-cone blender is representative of several new techniques available for on-line and in situ assays of blending mechanisms.

of historical difficulty in understanding the subject and the opportunity for future developments.

ACKNOWLEDGMENTS

We thank Albert W. Alexander, Paulo Arratia, Osama Sudah and Erinn Martin for vital contributions and the National Science Foundation, the New Jersey Commission on Science and Technology, the International Fine Particle Research Institute, Pfizer, Inc., Mobil, Technology, and Merck & Co., for financial support.

REFERENCES

1. Nakagawa, M.; Altobelli, S.A.; Caprihan, A.; Fukushima, E.; Jeong, E.K.; Non-Invasive Measurements of Granular Flows by Magnetic Resonance Imaging. *Expt's in Fluids* **1993**, *16*, 54–60.
2. Shinbrot, T.; Alexander, A.; Muzzio, F.J.; Spontaneous Chaotic Granular Mixing. *Nature* **1999**, *397*, 67, 5–8.
3. Lun, C.K.K.; Savage, S.B.; A Simple Kinetic Theory for Granular Flow of Rough, Inelastic, Spherical Particles. *J. Appl. Mech.* **1987**, *54*, 47–61.
4. Pöschel, T.; Buchholtz, V.; Static Friction Phenomena in Granular Materials: Coulomb Law vs. Particle Geometry. *Phys. Rev. Lett.* **1993**, *71*, 396, 3–6.
5. Müller, M.; Luding, S.; Herrmann, H.J.; Simulations of Vibrated Granular Media in 2D and 3D. *Friction, Arching and Contact Dynamics*; Wolf, D.E., Grassberger, P., Eds.; World Scientific: Singapore, 1997.
- Chester, A.W.; Kowalski, J.A.; Coles, M.E.; Muegge, E.L.; Muzzio, F.J.; Brone, D.; Mixing Dynamics in Catalyst Impregnation in Double-Cone Blenders. *Powder Technol.* **1999**, *102*, 85–94.
- Danckwerts, P.V.; The Definition and Measurement of Some Characteristics of Mixtures. *Appl. Sci. Res.* **1952**, *3*, 279–297.
- DasGupta, S.; Khakhar, D.V.; Bhatia, S.K.; Axial Segregation of Particles in a Horizontal Rotating Cylinder. *Chem. Eng. Sci.* **1991**, *46* (5/6), 1513–1517.
- Donald, M.B.; Roseman, B.; Mixing and Demixing of Solid Particles, Part I. Mechanisms in a Horizontal Drum Mixer. *British Chemical Engineering*, *7*, 749–753.
- Harnby, N.; The Selection of Powder Mixers. *Mixing in the Process Industries*, Butterworth Heinemann; Oxford UK, 1997; 42–61.
- Jaeger, H.M.; Nagel, S.R.; Behringer, R.P.; Granular Solids, Liquids, and Gases. *Rev. Mod. Phys.* **1996**, *68*, 1259–1273.
- Lacey, P.M.C.; Developments in the Theory of Particle Mixing. *J. Appl. Chem.* **1954**, May 4, 257–268.
- Moakher, M.; Shinbrot, T.; Muzzio, F.J.; Experimentally Validated Computations of Flow, Mixing and Segregation of Non-Cohesive Grains in 3D Tumbling Blenders. *Powder Technology* **2000**, *109*, 58–71.
- Muzzio, F.J.; Robinson, P.; Wightman, C.; Brone, D.; Sampling Practices in Powder Blending. *Int. J. Pharmaceutics* **1997**, *155*, 153–178.
- Muzzio, F.J.; Roddy, M.; Brone, D.; Alexander, A.W.; Sudah, O.; An Improved Powder-Sampling Tool. *Pharm. Tech.* **1999**, *23*, 92–110.
- Reynolds, O.; On the Dilatancy of Media Composed of Rigid Particles in Contact with Experimental Illustrations. *Philos. Mag.* **1885**, *20*, 46, 9–82.
- Ristow, G.H.; Simulating Granular Flow with Molecular Dynamics. *J. Phys. I France* **1992**, *2*, 64, 9–72.
- Robinson, P.; Muzzio, F.J.; Wightman, C.; Brone, D.; Gleason, E.K. End-Sampling Thief Probe. US Patent 5,996,426, 1999.
- Shinbrot, T.; Muzzio, F.J.; Nonequilibrium Patterns in Granular Mixing and Segregation. *Physics Today*, March, 2000; 25–30.
- Walton, O.R.; Numerical Simulation of Inelastic, Frictional Particle-Particle Interactions. *Particulate Two-Phase Flow*; Roco, M.C., Ed.; Butterworth-Heinemann: Oxford UK, 1993; 884–911.
- Williams, J.C.; Continuous Mixing of Solids—A Review. *Powder Technol.* **1976**, *15*, 237–243.

BIBLIOGRAPHY

Brone, D.; Wightman, C.; Connor, K.; Alexander, A.W.; Muzzio, F.J.; Robinson, P.; Using Flow Perturbations to Enhance Mixing of Dry Powders in V-Blenders. *Powder Technol.* **1997**, *91*, 165–172.

ORPHAN DRUGS

Carolyn H. Asbury

University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.

INTRODUCTION

As drug development costs began to rise in the late 1960s and 1970s with new Food and Drug Administration (FDA) requirements for demonstrating relative safety and efficacy, manufacturers of drugs, biologicals, and diagnostic agents faced a dilemma. How could they consider developing drugs that were important medically but not likely to be profitable? Drugs for diseases and conditions that were rare in the United States were one of several—and the most visible—of such drug types. Rarity meant few patients would be available to participate in clinical trials of safety and efficacy under the sponsor's Investigational Exemption of a New Drug (IND) application. Slow patient recruitment into trials increased development time, with the 17-year patent-protection clock ticking. And once the pharmaceutical sponsor received approval for the New Drug Application (NDA) required for interstate marketing, there would be few customers. This generally translated into a low return on investment (ROI). Moreover, the time devoted to testing and gaining approval for a drug for a rare disease was an opportunity cost. Manufacturers would otherwise have devoted the time and resources to products with more favorable market returns.

Six other drug categories were of limited commercial interest. They included products for (1):

1. Chronic diseases, requiring extended testing phases to assess long-term effects, using up critical patent-protection time;
2. Single administration, such as vaccines, requiring separate production facilities, used by many but on a one-time basis, and carrying high liability risks; and diagnostic agents that had relatively low sales volume;
3. Women of childbearing age, presenting unparalleled liability risks;
4. Children and the elderly, both of whom were excluded from clinical testing at the time. Children presented high liability risks and could be difficult to enroll in clinical trials. Elderly people, who often had multiple conditions, were excluded from trials by the FDA because the multiple conditions and drugs used to treat them would confound trial results. (Once a drug was

marketed, physicians could prescribe it for these two populations. However, optimal dosing had not been determined, and likely drug interactions in elderly patients had not been identified);

5. Diseases, rare or common, for which the intended drugs were not patentable including shelf chemicals, drugs known to exist, natural chemicals, and drugs for which the patent had expired;
6. Substance abuse or relapse prevention, intended for a population considered to be a high liability risk, uncooperative with treatment trials, and requiring extensive records once marketed; and
7. Developing countries, with the related problems of distributing and paying for products.

Observing that certain drugs could be in common use but were not potentially profitable enough to invite commercial introduction, Provost referred to them in the *American Journal of Hospital Pharmacy* in 1968 as homeless or “orphan” drugs (2). This chapter focuses on orphan drugs for rare diseases. It provides background on the issues and on attempts to deal with them—a description of the Orphan Drug Law, enacted in 1983, to address these issues—and an accounting of progress and problems to date emanating from the law.

BACKGROUND

The 1938 Food, Drug and Cosmetics Act and the 1962 Kefauver-Harris Amendments to it substantially altered drug development in this country. They strengthened the safety and efficacy determinations of drugs, but at cost, one of which was industry disinterest in drugs for rare diseases. The 1938 law required proof of safety after diethylene glycol, used as a sulfanilamide vehicle was found to form lethal quantities of oxalic acid in the body. This resulted in 100 fatalities, mostly children (3). The law required manufacturers to provide evidence that the drugs were relatively safe. For the FDA to keep a drug off the market, however, the onus was on the *agency* to prove that the drug was not safe. The thalidomide disaster changed that situation. Tragic birth defects were reported

in infants born to women in Europe and Canada who had taken the drug while pregnant. At the time the news broke, Congress was working on amendments to the 1938 law that required sponsors to provide proof of drug efficacy before receiving market approval. The thalidomide catastrophe propelled Congress to require sponsors also to show proof of safety. To implement these two requirements, the FDA created the IND process requiring sponsors to test for and establish relative safety before embarking on clinical trials to determine drug efficacy. The FDA set up the NDA approval process for ruling on safety and efficacy before the drug could be marketed interstate (4).

An essential interplay developed between patent protection and regulatory requirements. Patent protection became essential for emerging products developed by research-intensive, vertically integrated firms that looked to a few major market winners to survive in this era of increased development costs (5). Patent time used in the IND and NDA premarket stages reduced the time remaining for manufacturers to protect their products, once marketed, against less costly generic drug competition. Drug development also became more expensive. Development costs in the late 1960s and early 1970s, before the FDA amendments became operationally implemented, were estimated to range from a low of \$2.7 million to a high of \$16.9 million per new chemical entity (NCE) (6, 7). By the late 1970s, the estimate had risen to \$54 million (8). Since then, estimates have continued to leap upward: \$124 million in the late 1980s, \$231 million in 1991, and \$500 million today (9, 10).

By the late 1960s and early 1970s, market winners generated most of a company's profits and also helped encourage brand loyalties by prescribing physicians. Other drugs needed to at least break even. Although not blockbusters, these drugs generally were for large markets. Drugs for rare diseases usually did not break even and became therapeutic orphans. They became wards of government and university-sponsored development efforts.

Cancer treatment drugs were among the first wards of government. Even before the 1962 amendments to the Food, Drug and Cosmetics Act, the federal government had developed and maintained a role in stimulating the development of cancer drug treatment that was not being addressed by industry. The National Cancer Institute (NCI) created the Cancer Chemotherapy Program in 1955 with a \$5 million Congressional authorization. Congress decided that the NCI should take on the challenge. Impressed with industry's spectacular antibiotics development, Congress recognized that a low ROI was precluding industry's interest in exploiting the early successes in antitumor drugs (antifol aminopterin for acute childhood leukemia and

methotrexate for uterine choriocarcinoma). Beginning as a small grant-oriented program to develop antileukemia agents, the program was based on the use of transplantable tumors in syngeneic rodents as a system for testing new drugs. After the NCI drafted an agreement allowing for the trade secret status of data, industry submitted compounds for screening of bioactivity. Within three years, the Cancer Chemotherapy Program had grown into a \$35 million industrial contract effort. After a few missteps, when the FDA would not accept the cancer-funded research results as provided, the FDA and NCI developed a clear understanding of how data needed to be provided. From that point until at least the early 1980s, the NCI was involved in the development and/or clinical testing of every antineoplastic drug available in the United States (11–13).

In 1966, the National Institute of Neurological and Communicative Disorders and Stroke, as it was then named, established the Antiepileptic Drug Development Program to develop clinical trial methodologies and then to conduct clinical trials of antiepileptic agents already available in other countries. The program later developed a screening program similar to that of the NCI. The Institute filed INDs for drugs that entered clinical testing, and by 1981, four drugs had commercial sponsorship by the NDA stage (14, 15).

Although most of the federal funding for drug programs by the National Institutes of Health (NIH) was for the development of drugs for rare diseases—cancers and epilepsy—the NIH also devoted funding to development of other categories of orphan drugs. These included drugs to prevent and treat substance addiction and relapse, contraceptives for women of childbearing age, and some vaccines being developed by NIH and by the Centers for Disease Control (now the Centers for Disease Control and Prevention). Before 1983, when the Orphan Drug Act was signed into law, NIH drug development programs and grant-supported research had resulted in 13 drugs for the treatment of rare diseases being approved and on the market. In the 17 years before the Orphan Drug Law, the pharmaceutical industry had developed and marketed 34 drugs or biologicals for use in rare diseases or conditions (16). Ten of these marketed orphan drugs and biologicals had been developed solely by industry without government or university support (17).

The Gathering Storm

Nonetheless, several articles published in medical journals by academic researchers chronicled their plight in formulating new dosages for available drugs or

encapsulating chemical ingredients for clinical tests in patients with rare diseases, because no pharmaceutical sponsor had come to their rescue (18–20). One of these researchers, Cambridge University's John Walshe, proclaimed that this "do-it-yourself" problem had gone far enough and should be placed on a sound commercial basis (21). The FDA published a list of potentially promising therapeutic agents that were under development primarily by academic or government-based scientists who had failed to find commercial sponsors to undertake the costly phase III clinical trials for efficacy, file the NDA, and market the product (22). In 1979, the FDA convened a Task Force on Drugs of Little Commercial Value to determine how to find commercial sponsors. That same year, Louis Lasagna, who was then at the University of Rochester, invoked Provost's notion in an article in the journal *Regulation* by asking who will adopt the orphan drugs (23)?

NIH-funded academic researchers were reporting failed attempts to interest industry in taking on and securing NDA approval for drugs it had not developed. Industry had cited three major problems. First, NIH-funded scientists may not have gathered and analyzed data according to FDA requirements. Second, the trade secret status of data could not be preserved if the researcher had previously published information in the scientific literature. And third, many of the NIH-funded studies were of drugs that were not (or no longer) patentable. The FDA task force recommended that incentives be provided to industry to develop drugs of limited commercial value, but that any profits made from those incentives be returned, in whole or in part, to the government. (24). This arbitration-type approach was based on the assumption that industry would be willing to make trade-offs with the FDA on behalf of these drugs. But this assumption was made in the absence of any indication from industry that this was the case (25).

Although the task force was the first official policy response to the orphan drug situation (apart from the NIH-sponsored research responses), there were several private-sector efforts underway to promote the development and availability of orphan products. The Pharmaceutical Manufacturers Association (PMA, now called the Pharmaceutical Research and Manufacturers of America, or PhRMA) had been seeking sponsors among its member companies for promising orphan therapeutics. And a consumer group, the National Organization for Rare Disorders (NORD), was growing in number (now representing more than 20 million patients and their families) and undertaking efforts to raise public awareness and to help link patients to researchers conducting clinical studies of their diseases or conditions.

Lightning Strikes

A researcher at the Mount Sinai School of Medicine New York, who was seeking a pharmaceutical company sponsor for the drug L-5HTP for myoclonus, appealed to his Congressperson, Elizabeth Holtzman (D-New York) for a legislative remedy to the plight of orphan drug research. She introduced a bill in 1980 to establish the Office of Drugs of Limited Commercial Value to assist the NIH in developing drugs for the treatment of rare diseases (26). Although no action was taken on the bill, the committee heard testimony from a young California man suffering from Tourette's syndrome, a genetically determined neurological condition causing its victims to twitch, tic, and have uncontrollable verbal (and often abusive) outbursts. When haloperidol proved unsuccessful in controlling his symptoms, the patient had tried desperately to obtain and try pimozide, a drug available for the condition in Canada and Europe, but not in the United States.

A Los Angeles newspaper carried an account of the testimony, which caught the eye of the producer of the television series *Quincy*. Before long, a *Quincy* episode was devoted to dramatizing the conundrum of patients trying to cope with Tourette's syndrome and of industry trying to contend with the commercial disincentives to develop drugs to treat these patients. The episode demonstrated that there were no villains and no remedies. In a compelling scene before a Congressional committee, *Quincy* delivered an impassioned appeal to Congress to find a remedy. Shortly thereafter, *Quincy* star Jack Klugman was asked by Congressman Henry Waxman (D-California) to appear at a hearing on an orphan drug bill that was similar to Holtzman's, introduced by her colleague Representative Ted Weiss (D-New York). Klugman testified at the hearing, using the identical appeal he had delivered on the show (27). A *Wall Street Journal* editorial, entitled "Leave of Reality," likened Klugman's appearance before the Waxman subcommittee as an orphan drug expert to having Leonard Nimoy (Mr. Spock on *Star Trek*) testify as an expert on the nation's space program (28). By the end of 1981, Congressman Waxman had introduced H.R. 5238, the Orphan Drug Act.

The orphan drug survey

To find out the current status of development of drugs for the treatment of rare diseases, the subcommittee surveyed the industry, the NIH, and the FDA on three groups of drugs:

1. Products listed by PMA (now PhRMA) member companies as drugs for the treatment of rare diseases

- that industry sponsors had marketed or had made available on a compassionate basis to specialists;
2. Drugs and biologicals listed by the FDA as under development, but needing a commercial sponsor for FDA approval and marketing; and
 3. Drugs under development by NIH scientists or grantees.

From this survey, the Subcommittee learned that industry had marketed 34 drugs for the treatment of rare diseases extending back to 1965 and had made an additional 24 drugs (under development) available to physicians on a compassionate basis for treating patients with rare diseases. Most (82%) of the marketed drugs were for conditions affecting fewer than 100,000 people in the United States; 10% were for 100,000 to 500,000 people, and the remaining 8% were for up to 1 million people. Industry respondents indicated that substantial federal funding had been provided for research and/or development (R&D) for all but 10 of these marketed drugs.

The picture of disincentives confirmed claims by industry spokespeople sponsors indicated that the ROI for 83% of the drugs was lower than the sponsors' average return for marketed drugs, whereas development costs were higher than average for 12% of these drugs. Other issues cited by industry were the lack of clarity of FDA clinical testing guidelines when small numbers of patients were available and involved. This further eroded the sponsors' ability to estimate the length of clinical testing, and therefore the length of remaining patent protection time, once the drug was approved. Survey data indicated that industry-sponsored marketed orphan drugs took an average of 5.75 years for clinical testing (from filing the IND to filing the NDA). Unpatented orphan products were increasingly unlikely to be submitted for NDA approval, suggesting the importance of having at least some period of market protection. Whereas nearly two-fifths (39%) of industry-sponsored drugs for the treatment of rare diseases had been marketed in the 1960s, even though they were not protected by patent, this fell to 29% by the 1970s. Liability claims had been filed against the manufacturers of one-fifth of the marketed orphan drugs. The promising finding was that one in four industry-sponsored marketed orphan drugs also had a common indication. This suggested that orphan drugs were a relatively good market gamble (29).

Orphan Drug Act's Passage Provides Market Incentives and Regulatory Assistance

Based primarily on Congressional testimony marshaled by NORD revealing that millions of people and their families

were profoundly affected by rare diseases and conditions, and on survey data revealing market and regulatory disincentives to therapeutic progress, Congress passed the Orphan Drug Act during a lame duck session in December 1992. Called "the golden egg of the lame duck Congress" by the head of the Generic Pharmaceutical Industry Association (GPIA), the bill was signed into law the first week of January 1983 (30). The act (Public Law 97-414) was supported initially by the GPIA, which had begun seeking sponsors for developed orphan products, and eventually by the PMA after certain provisions were deleted or modified. Provisions in the law, and in subsequent amendments in 1984, 1985, and 1988, addressed market and regulatory issues and provided incentives for pharmaceutical industry development of drugs for the treatment of rare diseases.

The law, as amended, defines a rare disease or condition as one that affects fewer than 200,000 persons in the United States. Alternatively, the disease or condition can affect more than 200,000 persons in the United States, if there is no reasonable expectation that the cost of developing and making a drug available for such disease or condition in the United States will be recovered from U.S. sales of the drug. Therefore, a drug can be designated by the FDA as an orphan by demonstrating applicability of the law's financial criteria, regardless of the total number of people affected in the United States. Major provisions are the availability of two market incentives and the reduction of regulatory barriers. The FDA Office of Orphan Products Development administers nearly all the law's provisions.

One incentive is seven years of market exclusivity granted by the FDA for a specific indication of a product. Exclusivity begins on the date the FDA approves the marketing application for the designated orphan drug and applies only to the indication for which the drug has been designated and approved. An application for designation as an orphan drug for a specific indication must be made before submission of the NDA (or biologic product license application, PLA) for market approval (31). Other sponsors can receive approval for a different drug to treat the same rare disease or can receive approval to market an identical drug for some other orphan or common indication. Market exclusivity, therefore, only precludes a second sponsor from obtaining approval to provide an identical drug for the identical orphan indication for which the first sponsor received exclusive market approval. Initially, market exclusivity pertained only to unpatented products. A 1985 amendment allowed exclusive approval for all orphan drugs, whether patented or not. This change was designed to provide incentives for sponsors of products for the treatment of rare diseases whose patents

would expire before or soon after approval, or in cases in which prior publication (usually by an academic or government scientist) had precluded issuance of a patent.

A second incentive provides tax credits equal to 50% of the costs of human clinical testing undertaken in any given year by a sponsor to generate data required for obtaining FDA market approval through successful completion of the NDA process. The Internal Revenue Service administers the tax credit provisions.

The act provides for the FDA to award grants to support clinical studies of designated orphan products underdevelopment. FDA grant funding as of March 2000 from the FDA totaled \$126.3 million. From initial funding in 1983 of \$500,000 in grants, the grant program peaked in 1994 at \$12.3 million and has declined slightly but steadily thereafter, totaling \$11.1 million in 1999 (32). Applications are reviewed by outside experts and are funded according to a priority score. The FDA Office of Orphan Product Development provides information at its website (www.fda.gov/orphan/GRANTS/patients) on investigators seeking research subjects. Listed by the name of the disease or condition, information includes a description of the study, criteria for inclusion in clinical trials (such as age, stage of disease, etc.), and contact information on the clinical investigator seeking participants for clinical trials. Patients, their families, or their physicians are able to follow up with the clinical investigator.

To address regulatory barriers the FDA also provides formal protocol assistance when requested by the sponsor of an orphan drug. Although formal review of a request for protocol assistance is the direct responsibility of the FDA Centers for Evaluation and Research (one for drugs, the other for biologicals), the Office of Orphan Products Development is responsible for ensuring that the request qualifies for consideration. A sponsor need not have obtained orphan drug designation to receive protocol assistance.

Finally, the FDA is required to encourage sponsors to design open protocols for drug availability to patients not included in clinical trials.

The Current Status of Orphan Drugs

The FDA approved 201 orphan products between 1983 and March 2000 (33). In the 17 years since passage of the act, therefore, the number of approved orphan drugs has increased sixfold from the number approved in the 17 years before the act (Table 1). The drugs are classified within 16 therapeutic categories, primarily for the treatment of cancer, infectious disease, AIDS and related conditions, and central nervous system conditions (Table 2). Included in the 201 approved drugs are 24 (12%) that received FDA grant support for clinical trials (34). A list of these grant-supported products is available at the FDA Office of Orphan Products Development website (www.fda.gov/orphan/GRANTS).

The number of sponsors of approved orphan drugs (nearly all of them produced at pharmaceutical or biotechnology companies) has increased from 17 (for the 34 drugs marketed before the act) to 110 as of March 2000. The number of approved drugs per sponsor ranges from one to eight, with a preponderance of one-drug sponsors (Table 3). A total of 813 products have received orphan designations as of March 2000. Sponsors of approximately 25% of these designated products have filed INDs for the products, indicating that they are under active development (35).

Many of the designated and approved orphan products are developed at biotechnology companies. Beginning in the 1970s, molecular biology had begun to spur the creation of biotechnology research companies. These companies reportedly recognized early on that market exclusivity, and lack of competition to develop products for the treatment of rare diseases, provided protection essential to raising venture capital. As a result, orphan drugs are now among biotechnology's most prevalent, and, according to some, most lucrative products. Between 1988 and 1992, biotechnology product designations increased by 31%, from 8 to 39% of total orphan designations (36). In addition to industry, however, sponsors have included a university, an individual researcher, and a state public health unit, which in aggregate sponsored six orphan drugs at the time of approval. Lists of approved and designated orphan drugs can be obtained from the FDA Office of Orphan Products Development. These data suggest that orphan drug development has consistently risen over the years since the law was enacted. Aggregate sales of

Table 1 Designated and approved orphan drug products

	Pre-1983	1989	1991	3/2000
Cumulative total approved orphan products	34 ^a	36	54	235
Total sponsors of approved orphan drugs ^b	17			110

^aAt the time, these were called drugs for rare diseases, not orphan drugs.
^bSponsors identified only for the two endpoints.

Table 2 Approved orphan drugs: Number per therapeutic category^a

Category	Number of approved orphan products N = 201
Cancer	49
Infectious disease	23
Central nervous system	22
Hematopoietic	21
AIDS, AIDS-related	21
Endocrine	18
Inborn errors of metabolism	12
Renal	8
Cardiovascular disease	7
Respiratory	5
Gastrointestinal	5
Bone	3
Immunological	2
Dermatological	2
Antidote	2
Urinary tract	1

^aDoes not include drugs for rare diseases marketed before the 1983 Orphan Drug Law.

orphan drugs have been reported to be more than \$1 billion a year (37).

A Resulting Issue: High Pricing for Some Products

By the early 1990s, U.S. sales data collected for 41 of the approved orphan drugs that had been on the market for a year or more indicated that 75% of the products had generated earnings of less than \$10 million per drug. Three products had earnings between \$10 and \$25 million, six

drugs between \$26 and \$100 million, and two products more than \$100 million. Biotechnology firms produced four of the 11 drugs with relatively high sales.

Those orphan drugs that command high prices have generated intense controversy over whether the market exclusivity provision is creating an unnecessary monopoly, keeping prices artificially high. One example is recombinant human erythropoietin (r-EPO), intended for patients with chronic renal failure-related anemia. EPO eliminates the need for frequent blood transfusions by patients with end-stage renal disease who are undergoing kidney dialysis. These patients are covered under the federally financed Medicare program. Both Amgen Inc. and Genetics Institute applied to the FDA for market exclusivity for their r-EPO products. Amgen was the first to receive FDA approval and market exclusivity. Genetics Institute was the first to receive a patent. On appeal, the court ruled that Amgen had exclusive marketing rights. Sales of r-EPO exceeded \$100 million in the first six months of marketing, paid for by Medicare. By 1991, sales totaled \$400 million (38).

Human growth hormone (r-hGH), another example, generated sales of \$150 million in 1991. Intended to treat approximately 12,000 children in the United States with retarded growth caused by a lack of endogenous pituitary hormone, two companies provide r-hGH. Genentech received FDA market exclusivity in 1985. Eli Lilly received market approval two years later, based on the determination that the two products differed by one amino acid (39,40). But the shared market did not lead to price competition: each company was earning approximately \$20,000 per child annually, depending on the dosage needed.

A third example, aerosol pentamidine, generated sales of \$130 million in 1991. Helping to prevent *Pneumocystis carinii* pneumonia associated with the human immunodeficiency virus (HIV), the increasing number of users resulting from a rapidly escalating HIV prevalence rate was expected to soon exceed the 200,000 population figure specified in the law. This example, along with the other two, prompted efforts to seek a legislative remedy. A series of amendments were introduced and passed by Congress in 1990 that sought to eliminate orphan status for products intended for use in epidemics and to allow shared market access for identical products developed simultaneously for the same indication, in the hope that this would lead to price competition. The amendments were vetoed (41). As Arno, Bonuck, and Davis present in *Milbank Quarterly*, AIDS treatment drugs exemplify the policy dilemma of how to use the Act to meet the legislative intent of stimulating development of drugs for small patient populations without resulting in prices that make such drugs inaccessible (42).

Table 3 Approved orphan drugs per sponsor, March 2000^a

Drugs per sponsor	Sponsors
1	69
2	19
3	6
4	2
5	5
6	2
7	1
8	2

^aDoes not include sponsors of drugs for rare diseases approved before the Orphan Drug Act of 1983.

Another, more recent, example is enzyme therapy for Gaucher's disease, an inborn error of metabolism, treated with Ceredase. The therapy, which requires more than a ton of placenta annually to extract and make the drug, can cost as much as \$500,000 per year per person, depending on the dosage needed. A 1996 National Institutes of Health technology assessment panel addressed issues in diagnosis and treatment of the disease and concluded that despite the success of enzyme therapy, treatment is limited by the cost. The panel reported that it was imperative to define the appropriate clinical indications for treatment and to determine the lowest effective initial and maintenance doses (43).

Although the survey undertaken prior to the Act found that retail costs of drugs for the treatment of rare diseases were reportedly higher than the manufacturer's average for 60% of the products marketed before the act, the pricing for some orphan drugs after the act, as exemplified by these examples, is a critically important consequence of the law. This consequence merits continued public scrutiny. It is a sign of progress that for some orphan drugs, accessibility rather than availability is now the challenge requiring creative solutions.

REFERENCES

- Asbury, C.H. Medical Drugs of Limited Commercial Interest: Profit Alone is a Bitter Pill. *Intern. J. Health Serv.* **1981**, *11* (3), 451–462.
- Provost, G.P. Homeless or Orphan Drugs. *Am. J. Hosp. Pharm.* **1968**, *25*, 609.
- Silverman, M.; Lee, P.R. *Pills, Profits and Politics*; University of California Press: Berkeley, 1974; 2.
- Harris, R. *The Real Voice*; Macmillan: New York, 1964; 21–25.
- Temin, P. Technology, Regulation and Market Structure in the Modern Pharmaceutical Industry. *Bell J. Econ.* **1979**, *10* (2), 427–446.
- Report of the Panel on Chemicals and Health of the President's Science Advisory Committee, 73–500, NSF, U.S. Government Printing Office: Washington, DC, 1973.
- Schwartzman, D. *Innovation in the Pharmaceutical Industry*; Johns Hopkins University Press: Baltimore, 1976; 106–107.
- Hansen, R.W. The Pharmaceutical Development Process: Estimates of Development Costs and Times and Effects of Proposed Regulatory Changes. *Issues in Pharmaceutical Economics*; Chien, R., Ed.; Lexington Books: Lexington, MA, 1980; 151–181.
- DiMasi, J.; Hanson, R.; Grabowski, H.; Lasagna, L. The Cost of Innovation in the Pharmaceutical Industry and Health Economics. *J. Health Econ.* **1991**, *10*, 107–142.
- Rosenbaum, D.E. The Gathering Storm Over Prescription Drugs. *The New York Times* **Nov. 14, 1999**, *1*, Section 4 (Week in Review).
- Devita, V.; Oliverio, V.T.; Muggia, F.M.; Wiernick, P.W.; Ziegler, J.; Goldin, A.; Rubin, D.; Henney, J.; Schepartz, S. The Drug Development and Clinical Trials Programs of the Division of Cancer Treatment, National Cancer Institute. *Cancer Clin. Trials* **1979**, *2*, 195–216.
- Zubrod, C.G.; Schepartz, S.; Leiten, J.; Endicott, K.M.; Carrese, L.M.; Baker, C.E. *Cancer Chemotherapy Reports October 1996*, *50* (7); DHEW: Washington, DC, 1968.
- Rate of Development of Anticancer Drugs by the National Cancer Institute and the U.S. Pharmaceutical Industry and the Impact of Regulation. *Final Report for the National Cancer Institute*; University of Rochester Medical Center: New York, 1981; 40.
- Krall, R.A. Anti-Epileptic Drug Development. I. History and a Program for Progress. *Epilepsia* **1978**, *19* (4), 398–408.
- Program Performance Summary. *National Institute of Neurological and Communicative Disorders and Stroke Antiepileptic Drug Development Program*; NINCDS: Washington, DC, 1982.
- Orphan Drugs in Development, 1992 Annual Survey. Pharmaceutical Manufacturers Association: Washington, DC, 1992.
- Orphan Drug Act Report, 97th Congress, 2nd Session. Washington, DC, Sept 17; 1982.
- Van Woert, M. Profitable and Non-Profitable Drugs. *N. Engl. J. Med.* **1978**, *298* (16), 903–906.
- Rawlins, M. No Utopia Yet. *Br. Med. J.* **1977**, *2*, 1076.
- Rawlins, M. Editorial. *Lancet* **1976**, *2*(7970), 835–836.
- Walshe, J.M. Treatment of Wilson's Disease with Trientine (Triethylene Tetramine) Dihydrochloride. *Lancet* **1982**, *1*, 643–647.
- Asbury, C.H. *Medical Drugs of Limited Commercial Interest: The Development of Federal Policy*; Johns Hopkins School of Hygiene and Public Health Thesis, Baltimore, 1981; 157–175.
- Lasagna, L. Who Will Adopt the Orphan Drugs? *Regulation* **1979**, *3* (6), 27–32.
- U.S. Food and Drug Administration Report. *Interagency Task Force Report to the Secretary of DHEW*; FDA: Washington, DC, 1979; 1–82.
- Asbury, C.H.; Stolley, P. Orphan Drugs: Creating Federal Policy. *Ann. Intern. Med.* **1981**, *95* (2), 221–224.
- Holtzman, E. Rep. H.R. 7089, 96th Congress, 2nd Session. Washington, DC, April 17, 1980.
- Klugman, J.; Seligman, A. Testimony, Serial No. 97–17, U.S. Government Printing Office: Washington, DC, 1981; 10–16.
- Leave of Reality (Editorial). *The Wall Street Journal* **March 12, 1981**, .
- Asbury, C.H. *Orphan Drugs: Medical vs. Market Value*; D.C. Heath: Lexington, MA, 1985; 136–155.
- Waxman, H.R. Rep. H.R. 5328, 97th Congress, 1st Session. Washington, DC, 1981.
- FDA Office of Orphan Products Development, www.fed.gov/orphan/about/progovw.htm (accessed Feb 2000).
- Personal Communication, *FDA OOPD*, March 2000.
- FDA List of Approved Orphan Products through 02/24/2000*, FDA: Washington, DC, 2000; Provided on Request from the FDA OOPD.

34. Grant-Supported Products with Marketing Approval, *FDA OOPD*, Washington, DC, www.fda.gov/orphan/grants/magranants (accessed Feb 2000).
35. Personal Communication, *FDA OOPD*, March 2000.
36. Schuman, S.R. Implementation of the Orphan Drug Act. *Food Drug Law J.* **1992**, 47 (4), 363–403.
37. Grady, D. In Quest to Cure Rare Diseases, Some Get Left Out. *The New York Times*; Nov. 16, 1999.
38. Coster, J.M. Recombinant Erythropoietin: Orphan Product with a Silver Spoon. *Intern. J. Technol. Assess. Health Care* **1992**, 8 (4), 635–646.
39. Hilts, P.J. Seeking Limits to a Drug Monopoly. *The New York Times*; May 14, 1992, D1; 7.
40. Asbury, C.H. Evolution and Current Status of the Orphan Drug Act. *Intern. J. Technol. Assess. Health Care* **1992**, 8 (4), 573–582.
41. Asbury, C.H. The Orphan Drug Act: The first 7 years. *J. Am. Med. Assoc.* **1991**, 265, 893–897.
42. Arno, P.S.; Bonuck, K.; Davis, M. Rare Disease Drug Development and AIDS: The Impact of the Orphan Drug Act. *Milbank Q.* **1995**, 73 (2), 231–252.
43. NIH Technology Panel on Gaucher Disease. Gaucher Disease. Current Issues in Diagnosis and Treatment. *J. Am. Med. Assoc.* **1996**, 275 (7), 548–553.
44. Asbury, C. *Orphan Drugs: Medical vs Market Value*; D.C. Heath and Company: Lexington, MA, 1985.
45. Scheinberg, I.H., Walshe, J.M. Eds. *Orphan Diseases and Orphan Drugs*; Manchester University Press: Manchester, UK, 1986.
46. Wagner, J. Orphan Technologies. *Intern. J. Technol. Assessment* **1992**, 8 (4), www.fda.gov/orphan.
47. Arno, P.S.; Bonuck, K.; Davis, M. Rare Diseases, Drug Development, and AIDS: The Impact of the Orphan Drug Act. *Milbank Q.* **1995**, 73 (2), 241–252.
48. Henkel, J. Orphan Drug Law Matures into Medical Mainstay. *FDA Consumer* **1999**, 33 (3), 29–32.

NANOPARTICLES AS DRUG DELIVERY SYSTEMS

Elias Fattal

Christine Vauthier

University of Paris XI, Châtenay-Malabry, France

INTRODUCTION

Nanoparticles are small colloidal particles which are made of non biodegradable and biodegradable polymers. Their diameter is generally around 200 nm. One can distinguish two types of nanoparticles (Fig. 1): nanospheres, which are matrix systems; and nanocapsules, which are reservoir systems composed of a polymer membrane surrounding an oily or aqueous core. These systems were developed in the early 1970s. This approach was attractive because the methods of preparation of particles were simple and easy to scale-up. The particles formed were stable and easily freeze-dried. Due to these reasons, nanoparticles made of biodegradable polymers were developed for drug delivery. Indeed, nanoparticles were able to achieve with success tissue targeting of many drugs (antibiotics, cytostatics, peptides and proteins, nucleic acids, etc.). In addition, nanoparticles were able to protect drugs against chemical and enzymatic degradation and were also able to reduce side effects of some active drugs. This review focuses on the preparation and characterization methods of nanoparticles. The main applications of these systems are also described.

PREPARATION OF NANOPARTICLES

Polymer nanoparticles including nanospheres and nanocapsules (Fig. 1) can be prepared according to numerous methods that have been developed over the last 30 years. The development of these methods occurred in several steps. Historically, the first nanoparticles proposed as carriers for therapeutic applications were made of gelatin and cross-linked albumin (1, 2). Then, to avoid the use of proteins that may stimulate the immune system and to limit the toxicity of the cross-linking agents, nanoparticles made from synthetic polymers were developed. At first, the nanoparticles were made by emulsion polymerization of acrylamide and by dispersion polymerization of methyl-methacrylate (3, 4). These nanoparticles were proposed as adjuvants for vaccines. However, since they were made of nonbiodegradable polymers, these nanoparticles were

rapidly substituted by particles made of biodegradable synthetic polymers. Couvreur et al. (5) proposed to make nanoparticles by polymerization of monomers from the family of alkylcyanoacrylates already used in vivo as surgical glue. They succeeded in making nanoparticles by polymerization of the monomers in oil-in-water type emulsions prepared with an acidified aqueous phase. During the same period of time, Gurny et al. (6) proposed a method based on the use of another biodegradable polymer consisting of poly(lactic acid) used as surgical sutures in humans. In this method, nanoparticles were formed directly from the polymer. Based on these initial investigations, several groups improved and modified the original processes mainly by reducing the amount of surfactant and organic solvents. At that time, the methods developed were only able to produce nanospheres (Fig. 1A). A breakthrough in the development of nanoparticles occurred in 1986 with the development of methods allowing the preparation of nanocapsules corresponding to particles displaying a core-shell structure with a liquid core surrounded by a polymer shell (7–9) (Fig. 1B). From 1986, there was also an acceleration in the development of new methodologies for the preparation of all types of nanoparticles. The nanoprecipitation technique was proposed (10) as well as the first method of interfacial polymerization in inverse microemulsion (11). In the following years, the methods based on salting-out (12), emulsion–diffusion (13, 14), and double emulsion (15) were described. Finally, during the last decade, new approaches were considered to develop nanoparticles made from polysaccharides based on the gelation properties of these natural macromolecules (16). These nanoparticles were developed for peptides and nucleic acid delivery. Another goal was the development of surface modified nanoparticles to produce long circulating particles able to avoid the capture by the macrophages of the mononuclear phagocyte system after intravenous administration (17).

All the methods can be classified into two groups depending on whether the nanoparticles are formed at the same time than the polymer itself requiring a polymerization reaction or are directly obtained from a polymer. There are numerous valuable reviews on the subject (9, 18–21). The general principles of the methods leading

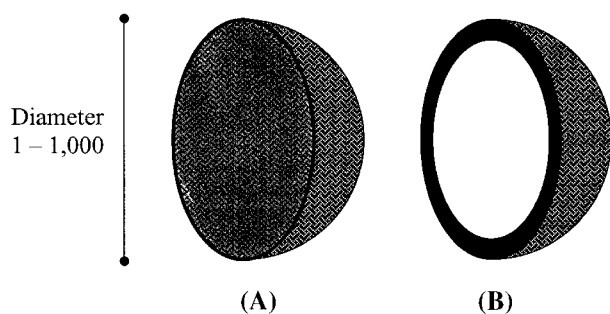


Fig. 1 Schematic representation of a nanosphere (A) and of a nanocapsule (B). In nanospheres, the whole particle consists of a continuous polymer network. Nanocapsules present a core-shell structure with a liquid core surrounded by a polymer shell.

to nanoparticle preparation are described and details of the most representative over are given.

Preparation of Nanoparticles by Polymerization

Nanospheres are mostly prepared by emulsion polymerization whereas nanocapsules are obtained by interfacial polymerization performed in emulsion or in microemulsion. In emulsion polymerization, the monomer itself, if liquid, is dispersed under agitation in a continuous phase in which it is nonmiscible. The polymerization is usually initiated by the reaction of the initiators with the monomer molecules that are dissolved in the continuous phase of the emulsion. The polymerization continues by further addition of monomer molecules that diffuse toward the growing polymer chain through the continuous phase. The growing polymer chain remains soluble until it reaches a certain molecular weight for which it becomes insoluble.

Therefore, phase separation occurs leading to the nucleation of the polymer particles and the formation of the tyndall scattering effect. Further growth of the nucleated particles occurs according to a mechanism that depends on the stability conditions of the whole system. This includes capture of new growing polymer chains, fusion or collision between nucleated particles (22). Throughout the polymerization, the monomer input in the continuous phase of the emulsion takes place by diffusion from the monomer droplets, which play the role of monomer reservoirs. When the reaction is completed, the particles formed contains a large number of polymer chains (19, 22). Emulsion polymerization can be performed in emulsifier free systems and in both oil-in-water and water-in-oil emulsions.

The poly(alkylcyanoacrylate) nanospheres, widely used as drug carriers, are prepared by emulsion polymerization according to a method initially introduced by Couvreur et al. (5). The monomers (isobutylcyanoacrylate, isohexylcyanoacrylate, *n*-butylcyanoacrylate) are dispersed in a continuous acidified aqueous phase under magnetic agitation. The anionic polymerization of the alkylcyanoacrylate is rapidly and spontaneously initiated by the remaining OH^- ions of the acidified water and is completed within 3–4 h depending on the monomer type (Fig. 2). The preparation is performed at low pH (pH ~ 2.5) to slow down the anionic polymerization of the alkylcyanoacrylate, therefore allowing the polymer to arrange as colloidal particles. Dextran 70 or Pluronic[®] F68 are usually dissolved in the aqueous phase to ensure the stability of the polymer particles. The size of the nanospheres can be controlled by the amount of Pluronic[®] F68 from a diameter of 40–250 nm for

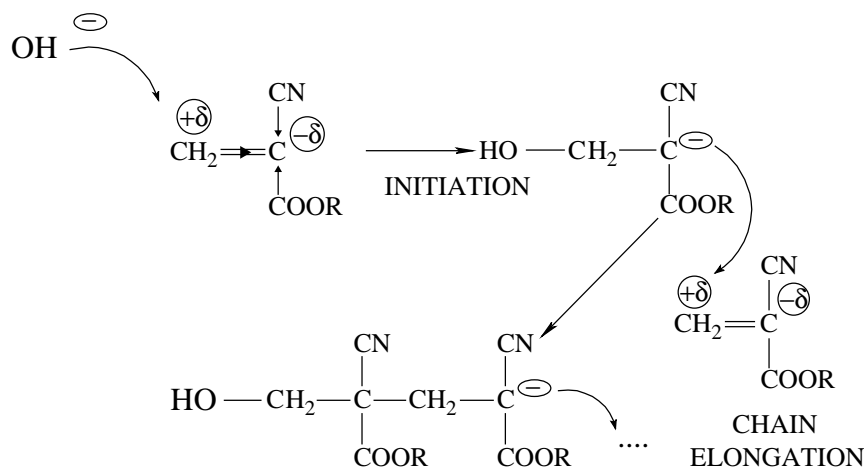


Fig. 2 Anionic polymerization of alkylcyanoacrylate initiated by the OH^- from the dissociation of the water molecule.

concentrations ranging from 3 to 0%, respectively (23). Numerous drugs have been associated with these nanospheres including doxorubicin, an anticancer agent, a peptide growth hormone, and several antibiotics. Antisense oligonucleotides were adsorbed on the nanosphere surface via the formation of an ion-pair with a cationic surfactant, cetyltrimethylammonium bromide. Finally, it should be mentioned that some drugs can lose their biological activity during the preparation of poly(alkylcyanoacrylate) nanospheres. Generally, these molecules contain chemical functions that are able to initiate the polymerization of alkylcyanoacrylates and be covalently attached to the polymer constituting the nanospheres. The mode of such an interaction has been elucidated for two molecules including phenylbutazone (an anti-inflammatory drug) and vidarabine (an antiviral molecule). In contrast, the side reactions can be used to achieve the covalent linkage of defined compounds to give specific properties to the nanospheres. This has been used with poly(ethylene glycol) that initiated the polymerization of isobutylcyanoacrylate to give poly(ethylene glycol)-coated nanospheres, therefore presenting a more hydrophilic surface than those prepared according to the original method (24).

Methylidene malonates are other monomers that give biodegradable polymers and polymerize according to a similar mechanism than alkylcyanoacrylates. These monomers were also used to make nanospheres by emulsion polymerization for drug delivery (25).

Nanocapsules can be prepared by interfacial polymerization of alkylcyanoacrylates (7). The main advantage of using these monomers is their very fast polymerization rate when they come into contact with water. Oil containing nanocapsules were prepared by the rapid dispersion of an ethanol phase including ethanol, the oil, the monomer, and the molecule to be encapsulated in an aqueous solution of surfactant. When the ethanol diffuses in the aqueous phase, tiny individual oil droplets form, and because of the contact with water at the oil/water interface, the polymerization of the alkylcyanoacrylate takes place on the droplet surface (26). This method is mainly adapted for the encapsulation of oily soluble substances (9). However, surprisingly, highly water-soluble molecules such as insulin could be entrapped in these nanocapsules with high encapsulation yields (up to 97%) (27).

Water containing nanocapsules were prepared by interfacial polymerization of the alkylcyanoacrylate in water-in-oil microemulsions. In these systems, water swollen micelles of surfactants of small and uniform sizes are dispersed in an organic phase. To prepare nanocapsules, the monomer is added to the oily phase of the

already prepared microemulsion. The anionic polymerization of the alkylcyanoacrylate is initiated at the surface of the water swollen micelles and the polymer formed locally to make the shells of the nanocapsules. In the method first introduced by Gasco and Trotta (11), the microemulsions were prepared with hexane as the organic phase and Aerosol-OT as the surfactant. Both of these constituents are not compatible for the development of an acceptable drug carrier system. Thus, the method was recently adapted to microemulsions formulated with more biocompatible compounds, but still quite high concentrations of surfactant (up to 14 wt%) are required for their preparation (28). The nanocapsules obtained by this method are dispersed in an organic medium and can mainly be used for oral administration. Indeed, for intravenous administrations, it is necessary to transfer the nanocapsules into an aqueous continuous phase. Such a transfer still remained a problem until recently, since the total elimination of the organic phase and the redispersion of the nanocapsules, in water is a difficult task to achieve avoiding the aggregation of the nanocapsules. Recently, Lambert et al. (29) proposed to perform this operation by ultracentrifugation of the nanocapsule dispersion over a layer of pure water. Using this simple approach, the nanocapsules transferred from the organic phase to the aqueous phase without any aggregation problem during the centrifugation. In addition, this technique allows the elimination of the excess surfactant, which remains in the organic phase. This nanoencapsulation method has special interest for the encapsulation of water soluble molecules such as peptides (28) and nucleic acids including antisense oligonucleotides (29).

Preparation of Nanoparticles Using a Polymer

In this group of methods, the nanoparticles are obtained from a polymer, which was prepared according to a totally independent method. This approach presents the major advantage that the polymers entering the composition of the nanoparticles are well characterized and their intrinsic physicochemical characteristics will not depend on the conditions encountered during the preparation of the nanoparticles as it can be the case with the previously described methods. Most methods starting from polymers have taken advantages of the physicochemical properties of the polymer used in terms of its solubility or its faculty to form a gel under certain conditions. Basically, two approaches are followed. One is based on the spontaneous formation of colloidal particles of the polymer that are then stabilized in a second step of the procedure. The second approach is based on the adaptation of methods initially developed to make microparticles. In this case, the

goal is to reduce the size of the particles formed with these methods. A third approach leading to the formation of very specific particles named SupraMolecular BioVectors by their authors (30) is described separately. The different ways to produce nanoparticles from a polymer is summarized in Fig. 3.

Methods based on the spontaneous formation of the nanoparticles

Spontaneous formation of nanoparticles can be achieved by taking advantage of the solubility and gelling properties of a dissolved polymer. Usually, the step allowing polymer colloidal particles to form is reversible, and it is necessary to complete the procedure by a second step required to stabilize the particles.

Based on the solubility properties of a polymer, the general principle is to prepare a solution of the polymer and to induce a phase separation by the addition of a nonsolvent of the polymer (10) or by a salting-out effect (2). The occurrence of the phase separation can be followed by turbidimetric measurements (2) or investigated using ternary phase diagrams (31). Phase separations leading to polymer colloid particles are usually obtained with diluted solutions of polymers. In a ternary phase

diagram, it corresponds to a small domain. Using higher polymer concentrations in the solvent, the colloidal particles formed at the limit of the phase separation as followed by turbidimetric measurements in the case of proteins. To facilitate the formation of colloidal dispersion of polymer particles, it is better to induce the phase separation in totally miscible solvent–nonsolvent systems. At that stage, the particles form spontaneously and quasi-instantaneously.

Once the proper conditions to obtain the polymer colloidal nanoparticles are identified, the particles must be stabilized. This is usually achieved either by the elimination of the polymer solvent by evaporation or by chemical cross-linking of the polymer as it is the case with proteins (32).

This method, also known as the nanoprecipitation method, can be applied to numerous synthetic polymers (10, 31). In general, the polymer is dissolved in acetone and the polymer solution is added into water. The acetone is then evaporated to complete the formation of the particles. Surface active agents are usually added to water to ensure the stability of the polymer particles. This easy technique of nanoparticle preparation was scaled up for large batch production. It leads to the formation of

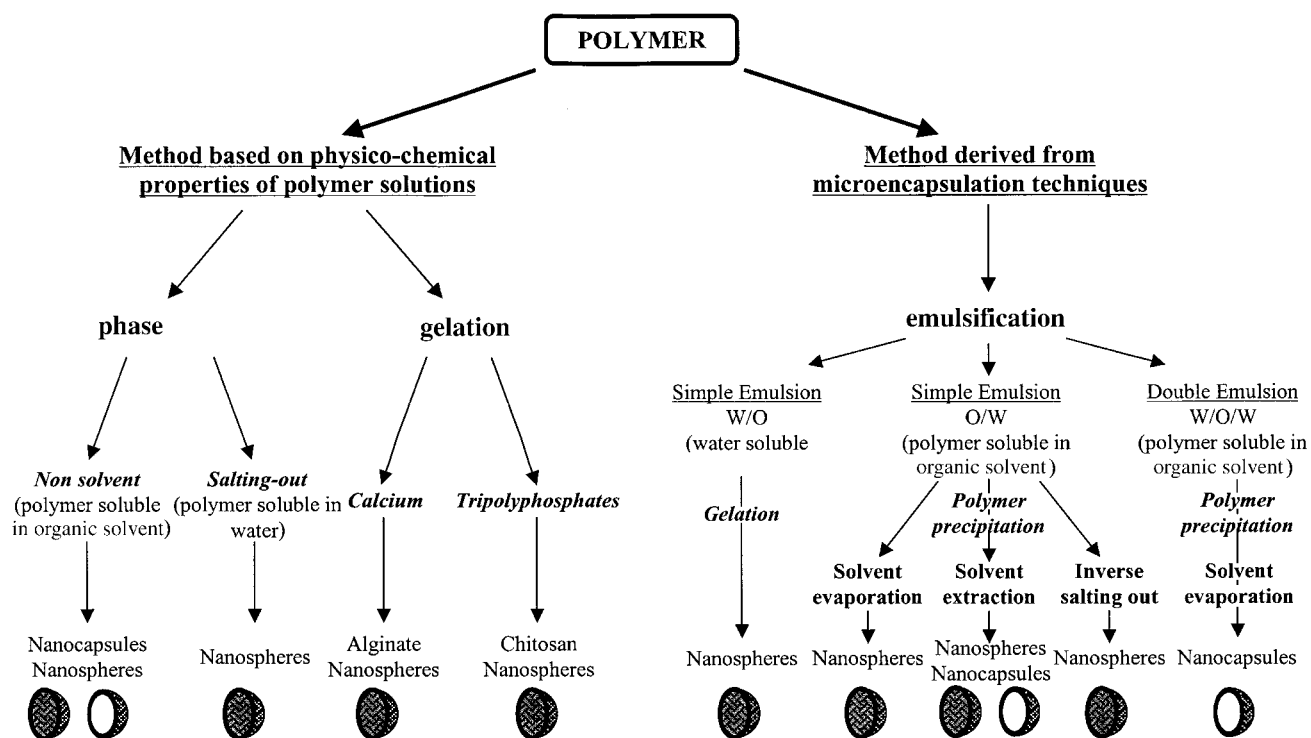


Fig. 3 Summary of the different methods to prepare nanospheres and nanocapsules from a polymer. W/O: water-in-oil, O/W: oil-in-water, W/O/W: water-in-oil-in-water.

nanospheres. Nanocapsules can easily be prepared by the same method just by adding a small amount of an organic oil in the polymer solution (8, 9). When the polymer solution is poured into the water phase, the oil is dispersed as tiny droplets in the solvent–nonsolvent mixture and the polymer precipitates on the oil droplet surface. This method leads to the preparation of oil-containing nanocapsules and can be valuably used for the encapsulation of liposoluble drugs.

Techniques based on the use of proteins are much more adapted to the encapsulation of hydrosoluble compounds and were recently developed to produce gelatin nanospheres as carrier systems for gene delivery (33, 34).

Based on the gelation properties displayed by certain polymers, nanoparticles are formed spontaneously by controlling the gelation process. This approach has been developed with alginate and chitosan that form highly water-swollen gels. The gelling agents for these two natural polysaccharides are respectively calcium and tripolyphosphate. Nanoparticles are formed in a certain domain of concentrations of the polysaccharide and of the gelling agent (35, 36). With the alginate, it has been shown that nanoparticles can be prepared when the respective concentrations of alginate and calcium are comprised in the domain of the pregel stage of the alginate gelling process (alginate 0.6 wt%, calcium chloride 0.9 mM). At this composition, tiny particles of gels form resulting from inter and intramolecular aggregations of alginate molecules caused by the interaction with calcium. These aggregates are stabilized by the formation of a polyelectrolyte complex with polylysine. Alginate and chitosan nanoparticles are interested for nucleic acid and protein delivery as recently reviewed by Vauthier and Couvreur (16).

Methods derived from microencapsulation techniques

Methods derived from microencapsulation techniques require the formation of an emulsion as a first step of the procedure. To produce nanometer-scale-sized particles, the size of the emulsion droplets must be small enough. This can be achieved by the use of special equipments such as high pressure homogenizers and microfluidizers. The energy input produced by these apparatus is important and is mainly due to high turbulence and cavitation forces. It allows an efficient dispersion of the polymer solution in the continuous phase. Once the desired emulsion is prepared, the formation of the nanoparticles can be induced according to two routes including the gelation of the polymer and the precipitation of the polymer either by solvent displacement or by solvent evaporation. Gelation can be induced by increasing the pH, adding calcium, or decreasing the temperature of emulsions containing

respectively, chitosan, alginate, and agarose (16). In the solvent displacement technique, the emulsion is formed with a solvent of the polymer, which is partially soluble in water and with an aqueous phase saturated with the solvent. Once the emulsion is formed with ethylacetate as an example, it is diluted by the further addition of water to displace the solvent from the dispersed phase inducing the polymer precipitation (14). This principle was also developed on the base of an inverse salting-out method (12). In this procedure, a solution of polymer in acetone is dispersed in an aqueous phase containing a high concentration of salt to keep the acetone nonmiscible with water. Just by diluting the emulsion with a large amount of pure water, the acetone is then extracted from the dispersed phase inducing the polymer precipitation. The total elimination of the acetone can be achieved by evaporation. Finally, precipitation of the polymer solubilized in the dispersed phase can also be induced by the removal of the solvent by evaporation (6, 37). In this method named emulsification-solvent evaporation, the polymer solvent diffuses through the aqueous continuous phase and evaporates at the air/water interface. The emulsion should remain under agitation during the time required for the total evaporation of the solvent. These methods produce nanospheres. To make nanocapsules, small amounts of oil can be added in the dispersed phase of the emulsion that will form the nanocapsules by solvent displacement (13). Nanocapsules can also be prepared with the solvent evaporation technique from a double water-in-oil-in-water emulsion (15). The removal of the organic solvent of the intermediate phase of this double emulsion by evaporation causes the polymer to precipitate at the surface of the inner aqueous phase. Whereas nanocapsules made by the solvent displacement method is more adapted for the encapsulation of lipophilic compounds, this last method allows the encapsulation of hydrosoluble compounds.

Preparation of SupraMolecular BioVector

SupraMolecular BioVector consists of polysaccharide nanospheres surrounded by a bilayer of phospholipids (30). At the origin, these systems were developed to mimic the low density lipoproteins that are natural colloidal structures encountered in the blood circulation and designed for the transport of cholesterol and cholesterol esters *in vivo*. SupraMolecular BioVector are prepared in several steps including the fonctionnalization and chemical cross-linking of a polysaccharide (starch or dextran), the purification and the drying of the modified polysaccharide followed by the fragmentation of the resulted powder under high pressure to produce small

polysaccharide nanoparticles. Finally, a lipid bilayer is adsorbed on the surface of the nanoparticles. The chemical modification of the polysaccharide forming the core of the SupraMolecular BioVector lead to various possibilities in terms of the type of molecules that can be associated with such a carrier system. Indeed, the polysaccharide core can be either negatively or positively charged or even neutral opening large application potential.

Preparation of Surface Modified Nanoparticles

Once intravenously administered, the body distribution of the nanoparticles is controlled by their surface properties. Indeed, despite their small size, nanoparticles display an enormous specific surface area that makes the interaction with the surrounding medium very important especially for their fate *in vivo*. Thus, the preparation of surface modified nanoparticles received much attention during the last decade to produce nanoparticles that are able to circulate for a long period of time in the blood stream at first and, more recently, to achieve an effective targeting of the device or to improve their bioadhesivity to mucosae.

Nanoparticles that are able to circulate for a long time in the blood stream should not be recognized by macrophages of the mononuclear phagocyte system. To achieve this goal, at least one of the two major known mechanisms involved in the recognition of foreign particles by macrophages should be avoided. These two mechanisms include the particle opsonization and the complement activation, which consists in protein adsorption and subsequent recognition by macrophages. A barrier to protein adsorption could be achieved by creating an efficient barrier of steric hindrance, therefore, by coating or adsorbing hydrophilic polymers to nanoparticle surface.

Nanospheres coated with poly(ethylene glycol) were first obtained by the simple adsorption of triblock copolymers of poly(ethylene glycol)–poly(propylene glycol)–poly(ethylene glycol) on the surface of already prepared nanospheres (38). To improve the stability of the poly(ethylene glycol) coating, nanospheres were prepared by nanoprecipitation or by emulsification-solvent evaporation using copolymers of poly(lactic acid)–co-poly(ethylene glycol) or of poly(alkylcyanoacrylate)–co-poly(ethylene glycol) (17, 39–41). Finally, poly(ethylene glycol) can initiate the polymerization of alkylcyanoacrylate to produce poly(ethylene glycol)-coated poly-(alkylcyanoacrylate) nanoparticles by emulsion polymerization (24, 41).

To make nanoparticles able to escape complement activation, Passirani et al. (42) proposed to coat

nanospheres with heparin. This compound, which is a polysaccharide, is a physiological inhibitor of complement activation *in vivo*. Heparin-coated poly(methylmethacrylate) nanoparticles were prepared by emulsion polymerization. In the method, the radical polymerization of methylmethacrylate was initiated by heparin according to an original method involving cerium ions and allowing heparin to covalently attach to poly(methylmethacrylate).

The next step now is the development of targeted nanoparticles toward a specific cell type. This has recently been investigated by Stella et al. (43) who prepared poly(alkylcyanoacrylate) nanoparticles showing residues of folic acid on their surface. These nanoparticles will be used to target cancer cells overexpressing a membrane receptor for the folic acid. The targeting moiety, consisting on the folic acid, was grafted on the surface of poly(aminopoly(ethylene glycol) cyano-co-hexadecylcyanoacrylate) nanoparticles that were obtained by nanoprecipitation.

In another way, chitosan was used as a coating agent for nanoparticles to improve their bioadhesive properties after oral and nasal administration (44). Indeed, chitosan is known to have bioadhesive properties as well as an interesting absorption enhancing capacity.

CHARACTERIZATION OF NANOPARTICULATE DRUG CARRIERS

Nanoparticles can be characterized by all the different physico chemical techniques that apply for polymer colloids (45). Concerning the development procedure of nanoparticles as drug carriers, the main physico chemical parameters that are investigated are the shape, the size, the surface properties, the density, and the concentration of the particles (19). The size as well as the size distribution are important parameters to be determined to achieve safe intravenous administration. Surface properties are also important to consider as nanoparticles display considerable specific surface area responsible for the interactions with the surrounding medium. Finally, the density and the concentration are required to deduce the specific surface area of the particles together with the size.

Nanoparticles can be visualized using different microscopy techniques. Transmission electron microscopy is usually applied to nanoparticles after negative staining with phosphotungstate acid or with uranyl acetate after it has been checked that the staining agents do not modify the particles. Recent progress in transmission electron microscopy now allows direct observations of the nanoparticles without the use of any staining agent that

may cause artefacts in some cases. In particular, direct observations of the nanoparticles after a sample of the nanoparticle dispersion has been freezing at very low temperature can be carried out by cryotransmission electron microscopy. Scanning electron microscopy is performed on samples coated with a thin layer of gold metal to produce the contrast. These techniques as well as those based on atomic force microscopy give useful images of the nanoparticles showing their shape. Measurements of the size and of the size distribution require determination of the diameters number of individual nanoparticles that may be assisted by the use of valuable image analysis softwares. The internal structure of the nanoparticles can be observed by freeze fracture and cryotransmission electron microscopy.

Generally, mean size and size distribution of nanoparticles are evaluated by quasi-elastic light scattering also named photocorrelation spectroscopy. This method is based on the evaluation of the translation diffusion coefficient, D , characterizing the Brownian motion of the nanoparticles. The nanoparticle hydrodynamic diameter, d_H is then deduced from this parameter from the Stokes Einstein law.

Other techniques can be used to determine the size and the size distribution of the nanoparticles. The field flow fractionation method is based on the separation of particles according to their size in a thin glass channel in which the flow carrying on the nanoparticles is submitted to an external perpendicular force produced either by a crossed flow or a sedimentation (46, 47). This technique, which can be applied for particles in a wide range of size (10 nm to several hundred μm), will gain more attention in the future for size determination and also for nanoparticle surface analysis (48). Size and size distribution of nanoparticles can also be determined by size exclusion chromatography performed on appropriate gels. This approach, requiring less equipment than the previous methods, presents the main limitation that only particles having a diameter lower than 120 nm can be characterized by this method (49).

As mentioned earlier, surface characteristics of the nanoparticles are of primary importance for the interaction of the nanoparticles with the surrounding medium. The main nanoparticle surface characteristics that are considered are the charge, the hydrophilicity, the chemical composition and the capacity to adsorb proteins and to induce complement activation.

The charge of the nanoparticle surface is usually evaluated by the measurement of their zeta potential, which gives information about the overall surface charge of the particles and how it is affected by changes in the environment (50). Zeta potential is affected by the surface

composition of the nanoparticles, the presence or the absence of adsorbed compounds, and the composition of the dispersing phase, mainly the ionic strength and the pH.

The hydrophilicity of the nanoparticle surface can be evaluated by hydrophobic interaction chromatography (51). This technique, based on affinity chromatography, allows a very rapid discrimination between hydrophilic and hydrophobic nanoparticles. The nanoparticles are passed through a column containing a hydrophobic interaction chromatography gel. The nanoparticles that are retained by the gel and only eluted after the addition of a surfactant are considered as hydrophobic, whereas the nanoparticles that do not interact with the gel and that are directly eluted from the column are considered as hydrophilic. Apart from the hydrophobic interaction chromatography, the field flow fractionation techniques recently appeared to present interesting potential for the characterization of nanoparticles with different surface characteristics (48).

X-ray photoelectron spectroscopy (ESCA) can be used to determine the chemical composition of the nanoparticle surface. This technique is a very useful tool for the development of surface modified nanoparticles providing a direct evidence of the presence of the components that are believed to be on the nanoparticle surface (38, 41).

The capacity of the nanoparticles to adsorb proteins and to activate the complement *in vivo* after intravenous administration will influence the fate of the carrier and its body distribution. To approach this aspect, *in vitro* tests have been developed to investigate the profile of the type of serum proteins that adsorbed onto the nanoparticle surface after incubation in serum and to evaluate the capacity of the nanoparticles to induce complement activation. The analysis of the protein adsorbed onto the nanoparticle surface can be performed by 2D-polyacrylamide gel electrophoresis. This technique allows the identification of the proteins that adsorbed onto the nanoparticle surface (52). To evaluate modifications of the composition of the adsorbed protein with time, a faster method based on capillary electrophoresis can also be used (53). Finally, the activation of the complement produced by nanoparticles can be evaluated either by a global technique or by a specific method measuring the specific activation of the component C3. In the global technique, nanoparticles are incubated with serum and, after the incubation, the remaining nonactivated complement in the serum is evaluated using a red blood cell lysis test (54, 55).

The concentration of nanoparticle in the dispersion can be deduced from gravimetric determination or by turbidimetric measurements based on the application of the Mie's law (48, 56). Density of the nanoparticles is evaluated

either by pycnometry (57) or by isopycnic centrifugation (13, 58).

PHARMACEUTICAL APPLICATIONS OF NANOPARTICLES

Nanoparticles were first developed in the mid-seventies by Birrenbach and Speiser (3). Later on, their application for the design of drug delivery systems was made available by the use of biodegradable polymers that were considered to be highly suitable for human applications (5). At that time, the research on colloidal carriers was mainly focusing on liposomes, but no one was able to produce stable lipid vesicles suitable for clinical applications. In some cases, nanoparticles have been shown to be more active than liposomes due to their better stability (59). This is the reason why in the last decades many drugs (e.g., antibiotics, antiviral and antiparasitic drugs, cytostatics, protein and peptides) were associated to nanoparticles.

Intravenous Administration

Fate of nanoparticles and their content after intravenous administration

The main interest of nanoparticles is their ability to achieve tissue targeting and enhance the intracellular penetration of drugs. After intravenous administration, nanoparticles are taken up by the liver, spleen and to a lower extent the bone marrow (60). Within these tissues, nanoparticles are mainly taken up by cells of the mononuclear phagocyte system (MPS) (61). The uptake occurs through an endocytosis process after which the particles end up in the lysosomal compartment (61) where they are degraded producing low molecular weight soluble compounds that are eliminated from the body by renal excretion (62). As a result of the MPS site specific targeting, avoidance of some organs was made possible, thus reducing the side effects and toxicity of some active compounds. Due to their strong lysosomal localization, one could imagine that nanoparticles are not suitable to target to the cytoplasm. To avoid their trapping within the lysosomal compartment, several compounds able to destabilize the lysosomal membrane were added to the nanoparticulate systems (e.g., cationic surfactant) (63) allowing some drugs to be delivered to the cytoplasm. Recently, to avoid MPS uptake, several groups have developed a strategy consisting of the linkage to the nanoparticles of poly(ethylene glycol) derivatives (39, 64, 65). This linkage results in a lower uptake of nanoparticles by the MPS and in a longer circulation

time. As a consequence, these so-called stealth[®] nanoparticles would be able to extravasate across endothelium that becomes permeable due to the presence of solid tumors.

Application to the treatment of intracellular infections

Intracellular infections were found to be a field of interest for drug delivery by means of nanospheres. Indeed, infected cells may constitute a "reservoir" for micro organisms, which are protected from antibiotics inside lysosomes. The resistance of intracellular infections to chemotherapy is often related to the low uptake of commonly used antibiotics or to their reduced activity at the acidic pH of lysosomes. To overcome these effects, the use of ampicillin, a β lactam antibiotic, bound to nanospheres was proposed as endocytosable formulation (66). The effectiveness of polyisohexylcyanoacrylate (PIHCA) nanospheres was tested in the treatment of two experimental intracellular infections.

Firstly, ampicillin-loaded nanospheres were tested in the treatment of experimental *Listeria monocytogenes* infection in congenitally athymic nude mice, a model involving a chronic infection of both liver and spleen macrophages (67). After adsorption of ampicillin onto nanospheres, the therapeutic activity of ampicillin was found to increase dramatically over that of the free drug. Bacterial counts in the liver were at least 20-fold reduced after linkage of ampicillin to PIHCA nanospheres. In addition, nanoparticulate ampicillin was capable of ensuring liver sterilization after two injections of 0.8 mg of nanospheres bound drug, whereas no such sterilization was ever observed with any of other regimens tested. Reappearance of living bacteria in the liver after the end of the treatment was probably due to a secondary infection derived from other organs such as the spleen, which was not completely sterilized by the treatment (67).

Secondly, nanosphere-bound ampicillin was tested in the treatment of experimental salmonellosis in C57/BL6 mice, a model involving an acute fatal infection (66). All mice treated with a single injection of nanoparticle-bound ampicillin survived, whereas all control mice and all those treated with unloaded nanospheres died within 10 days postinfection. With free ampicillin, an effective-curative effect required three doses of 32 mg each. Lower doses (3×0.8 mg and 3×16 mg) delayed but did not reduce mortality. Thus, the therapeutic index of ampicillin, calculated on the basis of mice mortality, was increased by 120-fold when the drug was bound to nanospheres.

In order to clarify the mechanism by which nanospheres improved the antimicrobial efficacy of ampicillin, Forestier et al. (68) have compared in vitro the efficacy of ampicillin bound to poly(isobutylcyanoacrylate) (PIBCA)

nanospheres with that of free ampicillin in terms of survival of *L. monocytogenes* in mouse peritoneal macrophages. After 30 h of incubation, nanosphere-bound ampicillin decreased the number of viable bacteria by 99% as compared to the controls whereas with free ampicillin, the number of bacteria was slightly lower than in the controls. Nanoparticle-ampicillin thus appeared to be much more effective than free ampicillin for inhibiting intracellular growth of *L. monocytogenes*. With in vitro *Salmonella typhimurium* infected macrophages, the situation was a little bit more complicated since the bactericidal effect of ampicillin-bound PIHCA nanospheres was poor although the intracellular capture of ampicillin was dramatically increased and its efflux in the extracellular medium reduced (69). In another study, confocal microscopy and transmission electron microscopy were used to establish the intracellular traffic of ampicillin-bound PIHCA nanospheres and its relation with the bacteria within the subcellular compartments (70). The data obtained clearly demonstrated the active uptake by phagocytosis of ampicillin-bound PIHCA nanospheres by murine macrophages and their localization in the same vacuoles as the infecting bacteria, but in a restrictive way (70). Thus, it was difficult to understand the limited bactericidal effect of ampicillin-bound nanospheres. The most probable explanation is to be found in the resistance mechanism of *S. typhimurium* involving the inhibition of the phagosome–lysosome fusion (71), which lets some bacteria in phagosomes free of nanospheres. If this proposed hypothesis (inhibition of phagosome–lysosome fusion) is correct, the dramatic efficiency observed in vivo should rather be due to the specific targeting of the infected tissues (rich in reticuloendothelial cells), than to an efficient intracellular targeting as it could be supposed.

In order to eliminate both dividing and nondividing bacteria, a fluoroquinolone antibiotic, ciprofloxacin, has been associated with PIBCA and PIHCA nanospheres. In an animal model of persisting *Salmonella* infection, although an effect on the early phase of the infection was observed, neither free nor nanosphere-bound ciprofloxacin was able to eradicate truly persisting bacteria (72).

Since they accumulate in the MPS, nanospheres hold promise as drug carriers for the treatment of visceral leishmaniasis (73). Thus, it has been shown that PIHCA nanospheres can be used as a carrier of primaquine whose activity was increased 21-fold against intracellular *Leishmania donovani* when associated with nanospheres (74). A part of the activity was attributed to the fact that phagocytosis of nanospheres led to the induction of a respiratory burst, which was more pronounced in infected than in noninfected macrophages (74). Dehydroemetine is

also one of the drug candidates for this treatment but has some side effects involving the heart, which were reduced after linkage with nanospheres (75).

Application to the treatment of cancer

When given intravenously, anticancer drugs are distributed throughout the body as a function of the physicochemical properties of the molecule. A pharmacologically active concentration is reached in the tumor tissue at the expense of massive contamination of the rest of the body. For cytostatic compounds, this poor specificity raises a toxicological problem, which presents a serious obstacle to effective therapy. The use of colloidal drug carriers could represent a more rational approach to specific cancer therapy. In addition, the possibility of overcoming multidrug resistance might be achieved by using cytostatics-loaded nanospheres.

The antitumor efficacy of doxorubicin-loaded nanospheres was first tested using the lymphoid leukemia L-1210 as a tumor model. In this study, one intravenous injection of doxorubicin-loaded PIBCA nanospheres was found to be more effective against L1210 leukemia than when the drug was administered in its free form following the same dosing schedule (76). Although the increased life span (ILS %) of mice injected with doxorubicin-loaded PIBCA nanospheres was twice as high as the ILS % for free doxorubicin, there were no long-term survivors.

The effectiveness of doxorubicin-loaded PIHCA nanospheres against L1210 leukemia was even more pronounced than that of doxorubicin loaded onto PIBCA nanospheres. The drug toxicity was markedly decreased when it was bound to this sort of nanospheres, so that impressive results were obtained with this formulation at doses for which the therapeutic efficiency of free doxorubicin was completely masked by the overpowering toxicity of the drug (76). Furthermore, preliminary experiments suggested that one i.v. bolus injection of doxorubicin-loaded nanospheres was more active, in L1210-bearing mice, than perfusion of the free drug for 24 h.

The superiority of doxorubicin targeted with the aid of poly(alkylcyanoacrylate) nanospheres was later confirmed in a murine hepatic metastases model (M5076 reticulosarcoma) (77). Irrespective of the dose and the administration schedule, the reduction in the number of metastases was much greater with doxorubicin-loaded nanospheres than with free doxorubicin, particularly if treatment was given only when the metastases were well established. The improved efficacy of the targeted drug, as clearly confirmed by histological examinations, shows that both the number and the size of the tumor nodules were lower when doxorubicin was administered in its

nanoparticulate form (77). Furthermore, liver biopsies of animals treated with the nanosphere-targeted drug showed a lower cancer cell density inside tumor tissue. Necrosis was often less widespread with the nanosphere-associated drug than in the control group and the group treated with free doxorubicin.

Studies performed on total homogenates of livers from both healthy and metastases-bearing mice showed extensive capture of nanoparticulate doxorubicin by the liver; no difference in hepatic concentrations was noted between healthy and tumor-bearing animals (77). In order to elucidate the mechanism behind the enhanced efficiency of doxorubicin-loaded nanospheres, doxorubicin measurements were made in both metastatic nodules and neighboring healthy hepatic tissue. This provided quantitative information concerning the drug distribution within these tissues (78). During the first 6 h after administration, the exposure of the liver to doxorubicin was 18 times greater for nanosphere-associated doxorubicin. However, no special affinity for the tumor tissue was detected and the nanospheres were seen by electron microscopy to be located within Kupffer cells (macrophages). However, at later time-points, the amount of drug in the tumor tissue increased in nanosphere-treated animals to 2.5 times the level found in animals given free doxorubicin. Since uptake of nanospheres by neoplastic tissue is unlikely, this increase in the doxorubicin concentration in tumor tissue probably resulted from doxorubicin released from healthy tissue, in particular Kupffer cells. Hepatic tissue could play the role of drug reservoir from which prolonged diffusion of the free drug (from nanospheres entrapped in Kupffer cell lysosomes) toward the neighboring malignant cells occurs.

This hypothesis raises the question of the long-term effect of an 18-fold increase of doxorubicin concentration in the liver. Although toxicological data have shown that doxorubicin-loaded nanospheres were not significantly or unexpectedly toxic to the liver in terms of survival rate at high doses, body weight loss, and histological appearance (79), this possibility should be borne in mind, especially since a temporary depletion in the number of Kupffer cells, and hence the ability to clear bacteria, was observed in rats treated with doxorubicin-loaded liposomes (80). A systematic study using unloaded poly(alkylcyanoacrylate) nanoparticles confirmed a reversible decline in the phagocytic capacity of the liver after repeated dosing, as well as a slight inflammatory response (81, 82). Nanoparticle-associated doxorubicin also accumulated in bone marrow, leading to a myelosuppressive effect (83). However, this tropism of carriers might be useful to deliver myelostimulating compounds such as granulocyte colony stimulating factor to reverse the suppressive effects

of intense chemotherapy (84). Nanospheres are also captured by splenic macrophages (85). In this study, the spleen architecture was shown to play a role in the localization of the nanospheres: in mice, uptake was mainly observed in metallophilic macrophages of the marginal zone whereas in rats, which have sinusoidal spleens similar to that of humans, particles were found in the red pulp macrophages.

On the other hand, alteration of the drug distribution profile by linkage to nanospheres can, in some cases, considerably reduce the toxicity of a drug because of reduced accumulation in organs where the most acute toxic effects are exerted. This concept was indeed illustrated with doxorubicin, which displays severe acute and chronic cardiomyopathy. After intravenous administration to mice, plasma levels of doxorubicin were higher when the drug was adsorbed onto nanospheres and at the same time the cardiac concentration of the drug was dramatically reduced (86). In accordance with the observed distribution profile, doxorubicin associated with nanospheres was found to be less toxic than free doxorubicin (78).

The ability of tumor cells to develop simultaneous resistance to multiple lipophilic compounds represents a major problem in cancer chemotherapy. Cellular resistance to anthracyclines has been attributed to an active drug efflux from resistant cells linked to the presence of transmembrane P-glycoprotein, which was not detectable in the parental drug-sensitive cell line. Drugs, such as doxorubicin, appear to enter the cell by passive diffusion through the lipid bilayer. Upon entering the cell, these drugs bind to P-glycoprotein, which forms transmembrane channels and uses energy from ATP hydrolysis to pump these compounds out of the cell (87). To solve this problem, many authors have proposed the use of competitive P-glycoprotein inhibitors, such as the calcium channel blocker verapamil, which are able to bind to P-glycoprotein and to overcome pleiotropic resistance. However, since the adverse effects of verapamil are serious, its clinical use to overcome multidrug resistance is limited.

During the past few years, many studies have been devoted to evaluating the antitumor potential of carrier-drug complexes (88). The effect of nanospheres loaded with doxorubicin, resistance to which is known to be related to the presence of P-glycoprotein, was evaluated. The cytotoxicity of free-doxorubicin, doxorubicin-loaded PIHCA nanospheres (NP-Doxorubicin) (mean diameter 300 nm), and nanospheres without drug (NP), against sensitive (MCF7) and multidrug resistant (Doxorubicin R MCF7) human breast cancer cell lines was compared (89). MCF7 cells were more sensitive to free-doxorubicin than

Doxorubicin R MCF7 cells with a 150-fold difference in the IC_{50} . No significant difference was observed in the survival rate of MCF7 treated with free-Doxorubicin or NP-doxorubicin. In contrast, for doxorubicin R MCF7, the IC_{50} for doxorubicin was 130-fold lower when NP-doxorubicin were used instead of free-doxorubicin (89). These results indicated that nanospheres provided an effective carrier for introducing a cytotoxic dose of doxorubicin into the pleiotropic resistant human cancer cell line Doxorubicin R MCF7.

Complementary experiments, conducted with other sensitive and resistant cell lines, have confirmed this efficacy of nanospheres (90, 91). Doxorubicin resistance was circumvented in the majority of the cell lines tested, and some encouraging results were obtained in vivo in a P388 model growing as ascites (90). Further studies were undertaken to elucidate the mechanism of action of polyalkylcyanoacrylate nanospheres. The incubation time and number of particles per cell were important factors (92) and, when PIBCA nanospheres were used, doxorubicin accumulation within P388/ADR resistant leukemic cells was increased compared with free drug, although no endocytosis of nanospheres occurred (93). On the other hand, when the less rapidly degradable PIHCA nanospheres were used, reversion was observed in the absence of increased intracellular drug (94). The degradation products of poly(alkylcyanoacrylate) nanospheres [mainly poly(cyanoacrylic acid)] were also able to increase both accumulation and cytotoxicity of doxorubicin, although they were soluble in the culture medium. Hence, the reversion of resistance seems to be due both to the adsorption of nanospheres on the cell surface and to the formation of a doxorubicin-poly(cyanoacrylic acid) ion pair, which facilitates the transport of the drug across the cell membrane (94).

In the light of the results obtained with doxorubicin-loaded nanospheres in the liver metastases model described earlier (77), the role of macrophages as a reservoir for doxorubicin was tested in a two-compartment coculture system in vitro with both resistant and sensitive P388 cells (95). Even after prior uptake by macrophages, doxorubicin-loaded PIBCA nanospheres were able to overcome resistance. However, this reversion was only partial. It was decided to take advantage of the particulate drug carrier offers to associate an anticancer drug and a compound capable of inhibiting the P-glycoprotein. This approach was tested with doxorubicin and cyclosporin A bound to the same nanospheres and was found to be extremely effective in reversing P388 resistance (95). The association of cyclosporin A with nanospheres would ensure that it reaches the same sites as the anticancer drug at the same time and would also reduce its toxic side effects.

As early as 1986, Al Khouri et al. (96) observed that like other colloidal carriers, nanocapsules, administered by the IV route in rabbits, were taken up rapidly by organs of the mononuclear phagocyte system. One application that takes advantage of this uptake concerns nanocapsules of (muramyl tripeptide cholesterol) (MTP-Chol). This immunostimulating agent is able to activate macrophages and induce toxicity toward tumor cells, and would therefore be a useful agent to treat metastatic cancer. The mechanisms by which activated macrophages arrest tumor proliferation include production of nitric oxide and $TNF-\alpha$. It was showed in in vitro models of rat alveolar macrophages and RAW 264.7 mouse monocyte macrophage line that nanocapsules based on poly(D,L-lactide) containing MTP-Chol are more efficient activators than the free drug (97, 98). This action could be due to an intracellular delivery of the immunomodulator encapsulated in nanocapsules after phagocytosis and to an intermediate transfer of the drug to serum proteins (99). This system has also demonstrated its efficiency in vivo; in fact, Barratt et al. (100) reported that antimetastatic effects of nanocapsules contained MTP-Chol in a model of liver metastases. Some antimetastatic activity was also seen after oral administration.

Nanospheres for oligonucleotide delivery

Oligodeoxynucleotides are potentially powerful new drugs because of their selectivity for particular gene products in both sense and antisense strategies. However, using antisense oligonucleotides in therapeutics is a challenge to pharmaceutical technology because of their susceptibility to enzymatic degradation and their poor penetration across biological membranes. Nanoparticulate preparations might be an interesting alternative because of better stability in the presence of biological fluids. In the case of nanospheres made of synthetic polymers [poly(alkylcyanoacrylate), poly(lactic acid)], since oligonucleotides have no affinity for the polymeric matrix, association with nanoparticles has been achieved by ion pairing with a cationic surfactant, cetyltrimethylammonium bromide (CTAB) adsorbed onto the nanoparticle surface (101). Oligonucleotides bound to poly(alkylcyanoacrylate) nanospheres in this way were protected from nucleases in vitro (63) and their intracellular uptake was increased (102). In addition, nanospheres were able to concentrate intact oligonucleotides in the liver and in the spleen (103). Antisense oligonucleotides formulated in this way were able to specifically inhibit mutated Ha-ras-mediated cell proliferation and tumorigenicity in nude mice (104).

This approach has recently been applied to the association of a phosphodiester antisense oligonucleotide directed against the 3' nontranslated region of the PKC α

gene with nanospheres prepared from PIBCA. These nanospheres were able to inhibit PKC α neo-expression in cultured Hep G6 cells (105).

Nanospheres containing oligonucleotides have also been formulated from a naturally occurring polysaccharide, alginate, which forms a gel in the presence of calcium ions. In this case, the oligonucleotides penetrate into the gel matrix by reptation, thus providing a high loading yield and good protection against nucleases (106).

Subcutaneous/Intramuscular Administration

Subcutaneous administration of nanoparticles was achieved mainly for the delivery of peptides and vaccines. It allows slow release of the entrapped drugs therefore reducing the number of administrations, increasing blood half-life of the active drug, and finally, in some cases, reducing side effects.

PIBCA nanospheres were injected subcutaneously to rats. Autoradiographic pictures obtained after using radiolabelled polymer have shown a progressive staining reduction in the muscular tissue suggesting that nanospheres were slowly biodegraded (107). In the same study, nanospheres were found to release a peptide (GRF) in a sustained manner. Comparison of the AUC of free GRF and GRF-loaded nanospheres showed that in addition to the slow release process nanospheres were able to improve the bioavailability of the peptide. This improvement could be attributed to the fact that free administered GRF is very quickly metabolized at the injection site, whereas it is partly protected from massive enzymatic degradation when it is administered associated with nanospheres (107).

A few examples of the use of nanospheres as adjuvant for antigens/allergens delivery were described in the literature. The main advantage of this approach is to design single shot vaccine. In this case the drug carrier has to remain at the site of administration and deliver either continuously or pulsatively the antigen. The use of slowly degradable polymers (PLA, PMMA) is suitable for this application, since peptide or protein release is more adequate. Poly(methyl methacrylate) were first investigated as adjuvants for injectable vaccines (108,109). These nanospheres were claimed to be biodegradable after subcutaneous or intramuscular injection and shown to exhibit very powerful adjuvant properties for a number of antigens. However, the adjuvant properties were shown to be better when the antigen was incorporated during the polymerization process than when adsorbed onto nanospheres (110). When comparing the effect of PMMA with other polymers (polystyrene and 2-hydroxyethyl methacrylate/methyl methacrylate copolymer, HEMA:

MMA), it was also demonstrated that a decrease in particle size and an increase in the hydrophobicity of nanospheres increased the antibody response after immunization against influenza whole and split virus, bovine serum albumin, and HIV2 split virus (111–113).

Oral Route

There are numerous reports showing that uptake and translocation of nanoparticles and microparticles take place after oral administration to animals (114–116). Different mechanisms have been proposed to explain the translocation of particulate material across the intestine: i) uptake via Peyer's patches or isolated lymphoid follicles; ii) intracellular uptake, and iii) intracellular/paracellular passage. The uptake of poly(alkylcyanoacrylate) nanocapsules by Peyer's patches has been shown by Damgé et al. (116). When administered in the lumen of an isolated ileal segment of the rat, polyalkylcyanoacrylate nanocapsules were found preferentially over Peyer's patches through which they passed massively and rapidly (116). Nanocapsules were clearly visible in M-cells and in intercellular spaces around the lymph cells. Intracellular uptake of nanospheres has been proposed by Kreuter et al. (117) based on electron-microscopic autoradiographic investigations showing radioactivity into epithelial and goblet cells after oral administration of poly(hexylcyanoacrylate) (PHCA) nanospheres labeled with ^{14}C . The translocation of particles by a paracellular pathway has been evidenced in a study done by Aprahamian et al. (118) using PIBCA nanocapsules. Nanocapsules were filled with an iodinated oil (lipiodol) in order to render them detectable using a scanning electron microscope equipped with an energy-dispersive X-ray spectrometer. When they were administered in an isolated segment of a dog jejunum, they appeared as vesicles associated with intraluminal mucus. Subsequently, they were observed in intravillus capillaries in close contact with red cells or adsorbed to the inner wall of endothelial cells. Among these three mechanisms and according to many studies involving nanoparticles made of other biodegradable and nondegradable polymers, translocation via the uptake in Peyer's patches seems to be the major pathway for a rapid and substantial passage after oral administration of nanoparticles. Although it might exist in certain situations, the passage of particles between the absorptive cells is rather less likely if the barrier of tight junctions has not been disrupted. Although there are abundant reports from various independent workers showing evidence of absorption of particulate systems by the gastrointestinal tract, the oral absorption of

nanoparticles remains a controversial issue. However, even if a more clear estimation of the quantity of absorbed particles is needed as well as a better understanding of the factors affecting particles uptake, it must be concluded that translocation of small sized particles like poly(alkylcyanoacrylate) nanoparticles is possible. The question remains if the extent of particle translocation is compatible with a strategy of drug administration with therapeutic perspectives. This will be discussed below.

Oral Delivery of Peptides and Proteins and Vaccines

Poly(isobutylcyanoacrylate) nanocapsules were shown 10 years ago to be able to encapsulate insulin and to increase its activity as assessed by a reduction of glycemia (119). Several aspects of this phenomenon are surprising: encapsulation of a hydrophilic drug in the oily core of nanocapsules; reduction of glycemia was only obtained with diabetic animals; hypoglycemia appeared two days after a single administration and was maintained for up to 20 days depending on the insulin doses, although the amplitude of the pharmacological effect (minimum level of blood glucose) did not depend on the insulin dose. Damgé et al. (116) and Lowe and Temple (120) suggested that nanocapsules could protect insulin from proteolytic degradation in intestinal fluids, based on the protection of encapsulated insulin, observed in the presence of different enzymes *in vitro*. Later studies showed that insulin did not react with the alkylcyanoacrylate monomer during the formation of nanocapsules and was located within the oily core rather than adsorbed on their surface (27, 121).

The capacity of insulin nanocapsules to reduce glycemia could be explained by their translocation through the intestinal barrier, as suggested by Damgé et al. (116); for example by paracellular pathway or via M cells in Peyer's patches (122). Recently, the use of Texas Red[®]-labeled insulin allowed this translocation to be visualized more readily (123). One hour after oral administration, nanocapsules reached the ileum. The presence of fluorescent areas within the mucosa and even in the lamina propria suggested that insulin-loaded nanocapsules could cross the intestinal epithelium. Although this passage is certainly an important factor, it does not explain the duration of the hypoglycemia. This prolonged action could be due to the retention of a part of the colloidal system in the gastrointestinal tract.

Interestingly, a prolonged hypoglycemic effect was also observed with insulin entrapped in poly(alkylcyanoacrylate) nanospheres when these were dispersed in an oily phase containing surfactant (124). This suggests that some

components of nanocapsules could act as promoters of absorption.

Recently, Damgé et al. (125) showed that the incorporation of octreotide, a somatostatin analogue, in poly(alkylcyanoacrylate) nanocapsules also improved and prolonged the therapeutic effect of this peptide, after administration by the oral route.

Even if the main limitation to oral administration of poly(alkylcyanoacrylate) nanoparticles is that their passage through the intestinal barrier is probably restricted and sometimes erratic, they represent an interesting tool for oral delivery of antigens. Indeed, M-cells appear to be the main site for the uptake of poly(alkylcyanoacrylate) nanoparticles after oral administration (116) and, furthermore, it is generally accepted that limited doses of antigen are sufficient for a mucous immunization. In fact, oral delivery of antigens may be considered as the most convenient means of producing an IgA antibody response. However, it is importantly limited by enzymatic degradation of antigens in the GI tract and, additionally by their poor absorption. Thus, it has been postulated that the use of micro- or nanoparticles for the oral delivery of antigens should be efficient if those systems are able to achieve the protection of the antigenic molecule. Poly(alkylcyanoacrylate) nanoparticles have been shown to enhance the secretory immune response after their oral administration in association with ovalbumin (126). This result was not fully reproduced in the case of poly(acrylamide) nanoparticles loaded with the same antigen. It was postulated that in the case of poly(acrylamide) nanospheres, much of the antigen was located at the surface of the polymer and could have been degraded during its passage through the gut. The relatively high surface concentration of ovalbumin adsorbed onto poly(butylcyanoacrylate) nanospheres may have reduced the ability of the proteolytic enzymes in the gut to gain access to and to degrade the antigen, resulting in a greater antigen availability.

Ocular Delivery

The anatomical structure and the protective physiological process of the eye exert a strong defense against ocular drug delivery. This is the reason why conventional ocular dosage forms exhibit extremely low bioavailability. Limited absorption of the drug through the lipophilic corneal barrier is mainly because of short precorneal residence time due to the tear turn-over, rapid nasolacrimal drainage of instilled drug from the tear fluid, and nonproductive absorption through the conjunctiva. Only a small proportion (1–3%) of the applied drug penetrates the cornea and reaches intraocular tissues. For these

reasons, it is necessary to develop efficient and more acceptable ocular therapeutic systems.

Different strategies can be carried out to improve the precorneal residence time and/or penetration ability of the active ingredient. Among them, one approach consists of using colloidal drug delivery systems such as nanoparticles. Initial studies carried out with nanocapsules, as ocular drug carriers, attempted to increase the penetration of lipophilic drugs into the eye by prolonging the precorneal residence time, as observed with other colloidal systems, liposomes and nanospheres. These studies, which concerned antiglaucomatous agents, such as betaxolol, carteolol, and metipranolol encapsulated in nanocapsules, only showed a reduction of the noncorneal absorption (systemic circulation) leading to reduced side-effects as compared with the free drug (127–129). These systemic side-effects are due to a poor ocular retention of drugs that are directly absorbed into the systemic circulation by conjunctival and nasal blood vessels. In two cases (carteolol and betaxolol), encapsulation in nanocapsules produced an improved pharmacological effect (reduction of intraocular pressure) than produced by free drug and nanospheres (although the penetration of nanocapsules was not tested) and reduced cardiovascular systemic side-effects (127,128). Metipranolol showed the same activity alone and associated with nanocapsules but, as in the case of carteolol and betaxolol, its side-effects were reduced. When betaxolol was used, the nature of the polymer making up the nanocapsule wall was found to be important, and poly(ϵ -caprolactone) was more efficient than PIBCA or poly(lactic-co-glycolic acid) (127).

Calvo et al. (130) explored the mechanisms of interaction of nanocapsules with ocular tissues to better understand the pharmacological responses obtained with antiglaucomatous agents. By confocal microscopy, they showed that poly(ϵ -caprolactone) nanocapsules could specifically penetrate the corneal epithelium by an endocytic process without causing any damage to the cells, in contrast with PIBCA nanoparticles, the uptake of which was associated with cellular lysis (131). These results explained the improved therapeutic effect and the reduction of systemic side-effects as a result of drug loss through the conjunctiva provided by poly(ϵ -caprolactone) nanocapsules by increasing corneal epithelium penetration of lipophilic drugs. Calvo et al. (132) also excluded the influence of the oily inner structure in the activity of the nanocapsules, in the light of the absence of differences in penetration between nanospheres and nanocapsules, in contrast with Marchal Heussler et al. (128) who observed a better therapeutic effect with nanocapsules than with nanospheres. Moreover, Calvo et al. (132) demonstrated with indomethacin-loaded nanocapsules that the colloidal

nature of the carrier was the main factor influencing its ocular bioavailability. The same authors were also interested in the influence of the nature and the charge of the surface of nanocapsules on their physical stability and on their ocular bioavailability (133, 134). They found that coating the negatively charged surface of poly(ϵ -caprolactone) nanocapsules with cationic polymers could prevent their degradation caused by the adsorption of lysozyme, a positively charged enzyme found in tear fluid (133). Moreover, they noticed that a cationic polymer, chitosan, adsorbed on the surface of nanocapsules was able to provide the best corneal drug penetration without any local intolerance as compared to another positively charged polymer. This was achieved by a combination of effects: Penetration of particles into the corneal epithelial cells, mucoadhesion of positively charged particles onto negatively charged membranes, and a specific effect on the tight junctions (134).

This effect of improvement of ocular absorption was also reported by Calvo et al. (135) with the immunosuppressive peptide cyclosporin A. The corneal level of the drug was increased fivefold as compared with an oily solution of the drug owing to a highly loaded nanocapsule preparation, also containing poly(ϵ -caprolactone). The efficacy of this topical formulation has also been observed on a penetrating keratoplasty rejection model in the rat (136). Le Boulrais et al. (137) also proposed an alternative preparation of cyclosporin nanocapsules based on poly(alkylcyanoacrylate) dispersed in poly(acrylic acid) gel able to drastically reduce toxicity of poly(alkylcyanoacrylates) on the cornea and to promote absorption of the drug.

REFERENCES

1. Scheffel, U.; Rhodes, B.A.; Natarajan, T.K.; Wagner, H.N. Albumin Microspheres for the Study of the Reticulo-Endothelial System. *J. Nucl. Med.* **1972**, *13*, 498–503.
2. Marty, J.J.; Oppenheim, R.C.; Speiser, P. Nanoparticles: A New Colloidal Drug Delivery Systems. *Pharm. Acta Helv.* **1978**, *53*, 17–23.
3. Birrenbach, G.; Speiser, P. Polymerized Micelles and their use as Adjuvants in Immunology. *J. Pharm. Sci.* **1976**, *65*, 1763–1766.
4. Kreuter, J.; Speiser, P. New Adjuvants on a Poly(methylmethacrylate) Base. *Infect. Immunol.* **1976**, *13*, 204–210.
5. Couvreur, P.; Kante, B.; Roland, M.; Guiot, P.; Baudhuin, P.; Speiser, P. Poly(cyanoacrylate) Nanoparticles as Potential Lysosomotropic Carriers: Preparation, Morphological and Sorptive Properties. *J. Pharm. Pharmacol.* **1979**, *31*, 331–332.
6. Gurny, R.; Peppas, N.A.; Harrington, D.D.; Banker, G.S. Development of Biodegradable and Injectable Lattices for

- Controlled Release Potent Drugs. *Drug Dev. Ind. Pharm.* **1981**, 7, 1–25.
7. Al Khoury-Fallouh, N.; Roblot-Treupel, L.; Fessi, H.; Devissaguet, J.P.; Puisieux, F. Development of a New Process for the Manufacture of Poly(isobutylcyanoacrylate) Nanocapsules. *Int. J. Pharm.* **1986**, 28, 125–136.
8. Fessi, H.; Devissaguet, J.P.; Puisieux, F. Procédé De Préparation Des Systèmes Colloïdaux Dispersibles Sous Forme De Nanocapsules, French Patent 86.18.444, 1986.
9. Legrand, P.; Barratt, G.; Mosqueira, V.; Fessi, H.; Devissaguet, J.P. Polymeric Nanocapsules as Drug Delivery Systems: A Review. *S.T.P. Pharma Sci.* **1999**, 9, 411–418.
10. Fessi, H.; Devissaguet, J.P.; Puisieux, F. Procédé De Préparation Des Systèmes Colloïdaux Dispersibles D'une Substance Sous Forme De Nanoparticules, French Patent 2.608.988, 1986.
11. Gasco, M.; Trotta, M. Nanoparticles from Microemulsions. *Int. J. Pharm.* **1986**, 29, 267–268.
12. Alléman, E.; Doelker, E.; Gurny, R. Preparation of Aqueous Polymeric Nanodispersions by a Reversible Salting-Out Process, Influence of Process Parameters on Particle Size. *Int. J. Pharm.* **1992**, 87, 247–253.
13. Quintanar-Guerrero, D.; Alléman, E.; Doelker, E.; Fessi, H. Preparation and Characterization of Nanocapsules from Preformed Polymers by a New Process Based on Emulsification–Diffusion Technique. *Pharm. Res.* **1998**, 15, 1056–1062.
14. Quintanar-Guerrero, D.; Alléman, E.; Fessi, H.; Doelker, E. Pseudolatex Preparation Using a Novel Emulsion–Diffusion Process Involving Direct Displacement of Partially Water-Miscible Solvents by Distillation. *Int. J. Pharm.* **1999**, 188, 155–164.
15. Zambaux, M.F.; Bonneauz, F.; Gref, R.; Maincent, P.; Dellacherie, E.; Alonso, M.J.; Labrude, P.; Vigneron, C. Influence of Experimental Parameters on the Characteristics of Poly(lactic acid) Nanoparticles Prepared by a Double Emulsion Method. *J. Controlled. Release* **1998**, 50, 31–40.
16. Vauthier, C.; Couvreur, P. Development of Polysaccharide Nanoparticles as Novel Drug Carrier Systems. *Handbook of Pharmaceutical Controlled Release Technology*; Wise, Trantolo, Cichon, Inyang, Stottmeister, Eds.; Chap. 21, Marcel Dekker, Inc.: New York, 2000; 413–429.
17. Gref, R.; Minamitake, Y.; Peracchia, M.T.; Domb, A.; Trubetskoy, V.; Torchilin, V.; Langer, R. Poly(ethylene Glycol)-Coated Nanospheres: Potential Carriers for Intravenous Drug Administration. *Pharm. Biotechnol.* **1997**, 10, 167–198.
18. Vauthier-Holtzscheler, C.; Benabbou, S.; Spenlehauer, G.; Veillard, M.; Couvreur, P. Methodology for the Preparation of Ultradispersed Polymer Systems. *S.T.P. Pharma Sci.* **1991**, 1, 109–116.
19. Kreuter, J. Nanoparticles. *Colloidal Drug Delivery Systems*; Kreuter, J., Ed.; Marcel Dekker, Inc.: New York, 1994; 219–342.
20. Quintanar-Guerrero, D.; Alléman, E.; Fessi, H.; Doelker, E. Preparation Techniques and Mechanisms of Formation of Biodegradable Nanoparticles from Preformed Polymers. *Drug Dev. Ind. Pharm.* **1998**, 24, 1113–1128.
21. De Jaeghere, F.; et al. Nanoparticles. *The Encyclopedia of Controlled Drug Delivery*; Mathiowitz, E., Ed.; Wiley and Sons, Inc.: New York, 1999; 641–664.
22. Lowell, P.A., El-Aasser, M.S. Eds.; *Emulsion Polymerization and Emulsion Polymers*, Wiley: New York, 1997; 801.
23. Seijo, B.; Fattal, E.; Roblot-Treupel, L.; Couvreur, P. Design of Nanoparticles of Less than 50 nm in Diameter. Preparation, Characterization and Drug Loading. *Int. J. Pharm.* **1990**, 62, 1–7.
24. Peracchia, M.T.; Vauthier, C.; Popa, M.; Puisieux, F.; Couvreur, P. Investigation of the Formation of Sterically Stabilized Poly(ethyleneglycol/isobutylcyanoacrylate) Nanoparticles by Chemical Grafting of Poly(ethylene glycol) During Polymerization of Isobutylcyanoacrylate. *S.T.P. Pharma Sci.* **1997**, 7, 513–520.
25. De Keyser, J.L.; Poupaert, J.H.; Dumont, P. Poly(diethyl methylidenemalonate) Nanoparticles as a Potential Drug Carrier: Preparation, Distribution and Elimination After Intravenous and Peroral Administration to Mice. *J. Pharm. Sci.* **1991**, 80, 67–70.
26. Gallardo, M.M.; Couarraze, G.; Denizot, B.; Treupel, L.; Couvreur, P.; Puisieux, F. Preparation and Purification of Isohexylcyanoacrylate Nanocapsules. *Int. J. Pharm.* **1993**, 100, 55–64.
27. Aboubakar, M.; Puisieux, F.; Couvreur, P.; Deyme, M.; Vauthier, C. Study of the Mechanism of Insulin Encapsulation in Poly(isobutylcyanoacrylate) Nanocapsules Obtained by Interfacial Polymerization. *J. Biomed. Mater. Res.* **1999**, 47, 568–576.
28. Watnasirichaikul, S.; Rades, R.; Tucker, I.G.; Davies, N.M. Manipulating the Release of Insulin from Nanocapsules Prepared by Interfacial Polymerization of Microemulsions, Proceedings of the 27th International Symposium on Controlled Release of Bioactive Materials of the Controlled Release Society, Inc.: Paris France July 9–13 2000, Abstract # 6138.
29. Lambert, G.; Fattal, E.; Pinto-Alphandary, H.; Gulik, A.; Couvreur, P. Polyisobutylcyanoacrylate Nanocapsules Containing an Aqueous Core as a Novel Colloidal Carrier for the Delivery of Oligonucleotides. *Biochimie* **1998**, 80, 969–976.
30. De Miguel, I.; Ioualalen, K.; Bonnefous, M.; Peyrot, M.; Nguyen, F.; Cervilla, M.; Soulet, N.; Dirson, R.; Rieumajou, V.; Imbertie, L.; Solers, C.; Cazes, S.; Favre, G.; Samain, D. Synthesis and Characterization of Supramolecular Biovector (SMBV) Specifically Designed for the Entrapment of Ionic Molecules. *Biochim. Biophys. Acta.* **1995**, 1237, 49–58.
31. Stainmesse, S. *Etude galénique D'un Nouveau Procédé D'obtention De Vecteurs Colloïdaux Submicroniques à Partir D'une Protéine Ou D'un Polymère Synthétique*; Thesis Université De Paris Sud, 1990.
32. Weber, C.; Coester, C.; Kreuter, J.; Langer, K. Desolvation Process and Surface Characterization of Protein Nanoparticles. *Int. J. Pharm.* **2000**, 194, 94–102.
33. Leong, K.W.; Mao, H.Q.; Truong-Le, V.L.; Roy, K.; Walsh, S.M.; August, J.T. DNA-Polycation Nanospheres as Non-Viral Gene Delivery Vehicles. *J. Control. Rel.* **1998**, 53, 183–193.
34. Truong-Le, V.L.; Walsh, S.M.; Schweibert, E.; Mao, H.Q.; Guggino, W.B.; August, J.T.; Leong, K.W. Gene

- Transfer by DNA-Gelatin Nanospheres. *Arch. Biochem. Biophys.* **1999**, *361*, 47–56.
35. Rajaonarivony, M.; Vauthier, C.; Couarraze, G.; Puisieux, F.; Couvreur, P. Development of a New Drug Carrier Made from Alginate. *J. Pharm. Sci.* **1993**, *82*, 912–918.
 36. Calvo, P.; Remunan-Lopez, C.; Vila-Jato, J.L.; Alonso, M.J. Novel Hydrophilic Chitosan-Polyethylene Oxide Nanoparticles as Protein Carriers. *J. Appl. Polym. Sci.* **1997**, *63*, 125–132.
 37. Vanderhoff, J.W.; El-Aasser, M.S.; Ugelstad, J. Polymer Emulsification Process, U.S. Patent 4.177.177, 1979.
 38. Illum, L.; Davis, S.S. The Organ Uptake of Intravenously Administered Colloidal Particles can be Altered Using Non-Ionic Surfactant (Poloxamer 338). *FEBS Lett.* **1984**, *167*, 79–82.
 39. Gref, R.; Minamitake, Y.; Peracchia, M.T.; Trubesky, V.; Torchilin, V.; Langer, R. Biodegradable Long-Circulating Nanospheres. *Science* **1994**, *263*, 1600–1603.
 40. Peracchia, M.T.; Desmaële, D.; Couvreur, P.; D'Angelo, J. Synthesis of a Novel Poly(PEG-Cyanoacrylate-co-Alkylcyanoacrylate) Amphiphilic Copolymer for the Development of "Stealth" PEG-Coated Nanoparticles. *Macromolecules* **1997**, *30*, 846–851.
 41. Peracchia, M.T.; Desmaële, D.; Vauthier, C.; Labarre, D.; Fattal, E.; D'Angelo, J.; Couvreur, P. Development of Novel Technologies for the Synthesis of Biodegradable Pegylated Nanoparticles. *Targeting of Drugs 6: Strategies for Stealth Therapeutic Systems*; Gregoriadis, B., McCormack, B., Eds.; Plenum Press: New York, 1998; 225–239.
 42. Passirani, C.; Ferrarini, L.; Barratt, G.; Devissaguet, J.P.; Labarre, D. Preparation and Characterization of Nanoparticles Bearing Heparin or Dextran Covalently-Linked to Poly(methylmethacrylate). *J. Biomater. Sci. Polym. Ed.* **1999**, *10*, 47–62.
 43. Stella, B.; Arpicco, S.; Peracchia, M.T.; Desmaële, D.; Hoebeke, J.; Renoi, M.; D'Angelo, J.; Cattel, L.; Couvreur, P. Characterization of Folic Acid-nanoparticles Conjugates for Tumoral Targeting Proceedings of the 27th International Symposium on Controlled Release of Bioactive Materials of the Controlled Release Society, Inc.: Paris France July 9–13, 2000; Abstract # 6307
 44. Kawashima, Y.; Yamamoto, H.; Takeuchi, H.; Kuno, Y. Mucoadhesive DL-Lactide/Glycolide Copolymer Nanospheres Coated with Chitosan to Improve Oral Delivery of Elcatonin. *Pharm. Dev. Technol.* **2000**, *5*, 77–85.
 45. Candau, F.; Ottewill, R.H., Eds.; *An Introduction to Polymer Colloids*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989; 240.
 46. Giddings, J.C. Field-Flow Fractionation: Analysis of Macromolecular, Colloidal and Particular Materials. *Science* **1993**, *260*, 1456–1465.
 47. Anger, S.; Caldwell, K.; Niehus, H.; Muller, R.H. High Resolution Size Determination of 20 nm Colloidal Gold Particles by SedFFF. *Pharm. Res.* **1999**, *16*, 1743–1747.
 48. Tan, J.S.; Butterfield, D.E.; Voycheck, C.L.; Caldwell, K.D.; Li, J.T. Surface Modification of Nanoparticles by PEO/PPO Block Copolymers to Minimize Interactions with Blood Components and Prolong Blood Circulation in Rats. *Biomaterials* **1993**, *14*, 823–833.
 49. Huve, P.; Verrecchia, T.; Bazile, D.; Vauthier, C.; Couvreur, P. Simultaneous Use of Size Exclusion Chromatography and Photon Correlation Spectroscopy for Poly(lactic Acid) Nanoparticle Characterization. *Chromatographia* **1994**, *675*, 129–139.
 50. Barratt, G. Characterization of Colloidal Drug Carrier Systems with Zeta Potential Measurements. *Pharmaceutical Technology Europe* **1999**, 25–32.
 51. Cartensen, H.; Muller, B.W.; Muller, R.H. Adsorption of Ethoxylated Surfactants on Nanoparticles. I. Characterization by Hydrophobic Interaction Chromatography. *Int. J. Pharm.* **1991**, *67*, 29–37.
 52. Blunk, T.; Hochstrasser, D.F.; Sanchez, J.C.; Muller, B.W.; Muller, R.H. Colloidal Carriers for IV Drug Targeting: Plasma Protein Adsorption Patterns on Surface Modified Latex Particles Evaluated by Two-Dimensional Polyacrylamide Gel Electrophoresis. *Electrophoresis* **1993**, *14*, 1382–1387.
 53. Olivier, J.C.; Vauthier, C.; Taverna, M.; Puisieux, F.; Ferrier, D.; Couvreur, P. Stability of Orosomucoid-Coated Poly(isobutylcyanoacrylate) Nanoparticles in the Presence of Serum. *J. Controlled Release* **1996**, *40*, 157–168.
 54. Vittaz, M.; Bazile, D.; Spenlehauer, G.; Verrecchia, T.; Veillard, M.; Puisieux, F.; Couvreur, P. Effect of PEO Surface Density on Long Circulating PLA-PEO Nanoparticles which Are Very Low Complement Activators. *Biomaterials* **1996**, *17*, 1575–1581.
 55. Passirani, C.; Barratt, G.; Devissaguet, J.P.; Labarre, D. Interactions of Nanoparticles Bearing Heparin or Dextran Covalently Bound to Poly(methylmethacrylate) with the Complement System. *Life Sci.* **1998**, *62*, 775–785.
 56. Irache, J.; Durrer, C.; Ponchel, G.; Duchene, D. Determination of Nanoparticle Concentration in Latexes by Turbidimetry. *Int. J. Pharm.* **1993**, *90*, R9–R12.
 57. Kreuter, J. Physico-Chemical Characterization of Polyacrylic Nanoparticles. *Int. J. Pharm.* **1983**, *14*, 43–58.
 58. Vauthier, C.; Schmidt, C.; Couvreur, P. Measurement of the Density of Polymeric Nanoparticulate Drug Carriers Made of Poly(alkylcyanoacrylate) and Poly(lactic Acid) Derivatives. *J. Nanoparticle Res.* **1999**, *1*, 411–418.
 59. Fattal, E.; Rojas, J.; Youssef, M.; Couvreur, P.; Andremon, A. Liposome-Entrapped Ampicillin in the Treatment of Experimental Murine Listeriosis and Salmonellosis. *Antimicrob Agents Chemother* **1991**, *35*, 770–772.
 60. Grislain, L.; Couvreur, P.; Lenaerts, V.; Roland, M.; Deprez-Decampeneere, D.; Speiser, P. Pharmacokinetics and Distribution of a Biodegradable Drug-Carrier. *Int. J. Pharm.* **1983**, *15*, 335–345.
 61. Lenaerts, V.; Nagelkerke, J.F.; Van Berkel, T.J.; Couvreur, P.; Grislain, L.; Roland, M.; Speiser, P. In Vivo Uptake of Polyisobutyl Cyanoacrylate Nanoparticles by Rat Liver Kupffer, Endothelial, and Parenchymal Cells. *J. Pharm. Sci.* **1984**, *73*, 980–982.
 62. Lenaerts, V.; Couvreur, P.; Christiaens-Leyh, D.; Joiris, E.; Roland, M.; Rollman, B.; Speiser, P. Degradation of Poly(Isobutyl cyanoacrylate) Nanoparticles. *Biomaterials* **1984**, *5*, 65–68.
 63. Chavany, C.; Ledoan, T.; Couvreur, P.; Puisieux, F.; Helene, C. Polyalkylcyanoacrylate Nanoparticles as Polymeric Carriers for Antisense Oligonucleotides. *Pharm. Res.* **1992**, *9*, 441–449.
 64. Peracchia, M.T.; Fattal, E.; Desmaele, D.; Besnard, M.; Noel, J.P.; Gomis, J.M.; Appel, M.; d'Angelo, J.;

- Couvreur, P. Stealth PEGylated Polycyanoacrylate Nanoparticles for Intravenous Administration and Splenic Targeting. *J. Controlled Release* **1999**, *60*, 121–128.
65. Bazile, D.; Prud'homme, C.; Bassoullet, M.T.; Marlard, M.; Spenlehauer, G.; Veillard, M. Stealth Me.PEG-PLA Nanoparticles Avoid Uptake by the Mononuclear Phagocytes System. *J. Pharm. Sci.* **1995**, *84*, 493–498.
 66. Fattal, E.; Youssef, M.; Couvreur, P.; Andreumont, A. Treatment of Experimental Salmonellosis in Mice with Ampicillin-Bound Nanoparticles. *Antimicrob. Agents Chemother.* **1989**, *33*, 1540–1543.
 67. Youssef, M.; Fattal, E.; Alonso, M.J.; Roblot-Treupel, L.; Sauzieres, J.; Tancrede, C.; Omnes, A.; Couvreur, P.; Andreumont, A. Effectiveness of Nanoparticle-Bound Ampicillin in the Treatment of *Listeria Monocytogenes* Infection in Athymic Nude Mice. *Antimicrob. Agents Chemother.* **1988**, *32*, 1204–1207.
 68. Forestier, F.; Gerrier, P.; Chaumard, C.; Quero, A.M.; Couvreur, P.; Labarre, C. Effect of Nanoparticle-Bound Ampicillin on the Survival of *Listeria Monocytogenes* in Mouse Peritoneal Macrophages. *J. Antimicrob. Chemother.* **1992**, *30*, 173–179.
 69. Balland, O.; Pinto-Alphandary, H.; Pecquet, S.; Andreumont, A.; Couvreur, P. The Uptake of Ampicillin-Loaded Nanoparticles by Murine Macrophages Infected with *Salmonella Typhimurium*. *J. Antimicrob. Chemother.* **1994**, *33*, 509–522.
 70. Pinto-Alphandary, H.; Balland, O.; Laurent, M.; Andreumont, A.; Puisieux, F.; Couvreur, P. Intracellular Visualization of Ampicillin-Loaded Nanoparticles in Peritoneal Macrophages Infected in Vitro with *Salmonella Typhimurium*. *Pharm. Res.* **1994**, *11*, 38–46.
 71. Buchmeier, N.A.; Heffron, F. Inhibition of Macrophage Phagosome–Lysosome Fusion by *Salmonella Typhimurium*. *Infect. Immun.* **1991**, *59*, 2232–2238.
 72. Page-Clisson, M.E.; Pinto-Alphandary, H.; Chachaty, E.; Couvreur, P.; Andreumont, A. Drug Targeting by Polyalkylcyanoacrylate Nanoparticles is Not Efficient Against Persistent *Salmonella*. *Pharm. Res.* **1998**, *15*, 544–549.
 73. Gaspar, R.; Oppendoes, F.R.; Preat, V.; Roland, M. Drug Targeting with Polyalkylcyanoacrylate Nanoparticles — In Vitro Activity of Primaquine-Loaded Nanoparticles Against Intracellular *Leishmania-Donovani*. *Ann. Trop. Med. Parasitol.* **1992**, *86*, 41–49.
 74. Gaspar, R.; Preat, V.; Oppendoes, F.R.; Roland, M. Macrophage Activation by Polymeric Nanoparticles of Polyalkylcyanoacrylates: Activity Against Intracellular *Leishmania-Donovani* Associated with Hydrogen Peroxide Production. *Pharm. Res.* **1992**, *9*, 782–787.
 75. Fouarge, M.; Dewulf, M.; Couvreur, P.; Roland, M.; Vranckx, H. Development of Dehydroemetine Nanoparticles for the Treatment of Visceral Leishmaniasis. *J. Microencapsul.* **1989**, *6*, 29–34.
 76. Brasseur, F.; Verdun, C.; Couvreur, P.; Deckers, C.; Roland, M. Evaluation Expérimentale De L'efficacité Thérapeutique De La Doxorubicine Associée Aux Nanoparticules De Polyalkylcyanoacrylate. *Proceedings of the 4th International Conference on Pharmaceutical Technology*, Paris, France, June, 3–5, 1986; 177–186.
 77. Chiannilkulchai, N.; Driouch, Z.; Benoit, J.P.; Parodi, A.L.; Couvreur, P. Doxorubicin-loaded Nanoparticles: Increased Efficiency in Murine Hepatic Metastases. *Sel. Cancer Ther.* **1989**, *5*, 1–11.
 78. Chiannilkulchai, N.; Ammoury, N.; Caillou, B.; Devissaguet, J.P.; Couvreur, P. Hepatic Tissue Distribution of Doxorubicin-Loaded Nanoparticles After I.V. Administration in Reticulosarcoma M 5076 Metastasis-Bearing Mice. *Cancer Chemother. Pharmacol.* **1990**, *26*, 122–126.
 79. Couvreur, P.; Grislain, L.; Lenaerts, V.; Brasseur, F.; Guiot, P.; Biornacki, A. Biodegradable Polymeric Nanoparticles as Drug Carrier for Antitumor Agents. *Polymeric Nanoparticles and Microparticles*; Guiot, P., Couvreur, P., Eds.; CRC Press, Inc.: Boca Raton, FL, 1986; 27–93.
 80. Daemen, T.; Hofstede, G.; Ten Kate, M.T.; Bakker-Woudenberg, I.A.; Scherphof, G.L. Liposomal Doxorubicin-Induced Toxicity: Depletion and Impairment of Phagocytic Activity of Liver Macrophages. *Int. J. Cancer.* **1995**, *61*, 716–721.
 81. Fernandez-Urrusuno, R.; Fattal, E.; Porquet, D.; Feger, J.; Couvreur, P. Evaluation of Liver Toxicological Effects Induced by Polyalkylcyanoacrylate Nanoparticles. *Toxicol. Appl. Pharmacol.* **1995**, *130*, 272–279.
 82. Fernandez-Urrusuno, R.; Fattal, E.; Rodrigues, J.M., Jr.; Feger, J.; Bedossa, P.; Couvreur, P. Effect of Polymeric Nanoparticle Administration on the Clearance Activity of the Mononuclear Phagocyte System in Mice. *J. Biomed. Mater. Res.* **1996**, *31*, 401–408.
 83. Gibaud, S.; Andreux, J.P.; Weingarten, C.; Renard, M.; Couvreur, P. Increased Bone Marrow Toxicity of Doxorubicin Bound to Nanoparticles. *Eur. J. Cancer* **1994**, *6*, 820–826.
 84. Gibaud, S.; Rousseau, C.; Weingarten, C.; Favier, R.; Douay, L.; Andreux, J.P.; Couvreur, P. Polyalkylcyanoacrylate Nanoparticles as Carriers for Granulocyte-Colony Stimulating Factor (G-CSF). *J. Controlled Release* **1998**, *52*, 131–139.
 85. Demoy, M.; Gibaud, S.; Andreux, J.P.; Weingarten, C.; Gourtin, B.; Couvreur, P. Splenic Trapping of Nanoparticles: Complementary Approaches for In Situ Studies. *Pharm. Res.* **1997**, *14*, 463–468.
 86. Verdun, C.; Brasseur, F.; Vranckx, H.; Couvreur, P.; Roland, M. Tissue Distribution of Doxorubicin Associated with PIHCA Nanoparticles. *Cancer Chemother. Pharmacol.* **1990**, *26*, 13–18.
 87. Kartner, N.; Evernden-Porelle, D.; Bradley, G.; Ling, V. Detection of P-glycoprotein in Multidrug-Resistant Cell Lines by Monoclonal Antibodies. *Nature* **1985**, *316*, 820–823.
 88. Brasseur, F.; Couvreur, P.; Kante, B.; Deckers-Passau, L.; Roland, M.; Deckers, C.; Speiser, P. Actinomycin D Absorbed on Polymethylcyanoacrylate Nanoparticles: Increased Efficiency Against an Experimental Tumor. *Eur. J. Cancer* **1980**, *16*, 1441–1445.
 89. Treupel, L.; Poupon, M.F.; Couvreur, P.; Puisieux, F. Vectorisation of Doxorubicin in Nanospheres and Reversion of Pleiotropic Resistance of Tumor Cells. *C. R. Acad. Sci. III* **1991**, *313*, 171–174.
 90. Cuvier, C.; Roblot-Treupel, L.; Millot, J.M.; Lizard, G.; Chevillard, S.; Manfait, M.; Couvreur, P.; Poupon, M.F. Doxorubicin-Loaded Nanospheres Bypass Tumor Cell Multidrug Resistance. *Biochem. Pharmacol.* **1992**, *44*, 509–517.

91. Bennis, S.; Chapey, C.; Couvreur, P.; Robert, J. Enhanced Cytotoxicity of Doxorubicin Encapsulated in Polyisohexylcyanoacrylate Nanospheres Against Multidrug-Resistant Tumour Cells in Culture. *Eur. J. Cancer* **1994**, *1*, 89–93.
92. Nemati, F.; Dubernet, C.; Colin de Verdière, A.; Poupon, M.F.; Treupel Acar, L.; Puisieux, F.; Couvreur, P. Some Parameters Influencing Cytotoxicity of Free Doxorubicin Loaded Nanoparticles in Sensitive and Multidrug Resistant Leukemic Murine Cells: Incubation Time, Number of Particles Per Cell. *Int. J. Pharm.* **1994**, *102*, 55–62.
93. Colin de Verdière, A.; Dubernet, C.; Nemati, F.; Poupon, M.F.; Puisieux, F.; Couvreur, P. Uptake of Doxorubicin from Loaded Nanoparticles in Multidrug-Resistant Leukemic Murine Cells. *Cancer Chemother. Pharmacol.* **1994**, *33*, 504–508.
94. Colin de Verdière, A.C.; Dubernet, C.; Nemati, F.; Soma, E.; Appel, M.; Ferte, J.; Bernard, S.; Puisieux, F.; Couvreur, P. Reversion of Multidrug Resistance with Polyalkylcyanoacrylate Nanoparticles: Towards a Mechanism of Action. *Br. J. Cancer* **1997**, *76*, 198–205.
95. Soma, C.E.; Dubernet, C.; Barratt, G.; Nemati, F.; Appel, M.; Benita, S.; Couvreur, P. Ability of Doxorubicin-Loaded Nanoparticles to Overcome Multidrug Resistance of Tumor Cells After Their Capture by Macrophages. *Pharm. Res.* **1999**, *16*, 1710–1716.
96. al Khouri, N.; Fessi, H.; Roblot-Treupel, L.; Devissaguet, J.P.; Puisieux, F. An Original Procedure for Preparing Nanocapsules of Polyalkylcyanoacrylates for Interfacial Polymerization. *Pharm. Acta Helv.* **1986**, *61*, 274–281.
97. Morin, C.; Barratt, G.; Fessi, H.; Devissaguet, J.P.; Puisieux, F. Improved Intracellular Delivery of a Muramyl Dipeptide Analog by Means of Nanocapsules. *Int. J. Immunopharmacol.* **1994**, *16*, 451–456.
98. Seyler, I.; Appel, M.; Devissaguet, J.P.; Legrand, P.; Barratt, G. Relationship Between NO-Synthase Activity and TNF-Alpha Secretion in Mouse Macrophage Lines Stimulated by a Muramyl Peptide Entrapped in Nanocapsules. *Int. J. Immunopharmacol.* **1996**, *18*, 385–392.
99. Seyler, I.; Appel, M.; Devissaguet, J.P.; Legrand, P.; Barratt, G. Macrophage Activation by a Lipophilic Derivative of Muramyl dipeptide. *J. Nanoparticle Res.* **1999**, *1*, 91–97.
100. Barratt, G.; Puisieux, F.; Yu, W.P.; Foucher, C.; Fessi, H.; Devissaguet, J.P. Anti-Metastatic Activity of MDP-L-Alanyl-Cholesterol Incorporated into Various Types of Nanocapsules. *Int. J. Immunopharmacol.* **1994**, *16*, 457–461.
101. Fattal, E.; Vauthier, C.; Aynie, I.; Nakada, Y.; Lambert, G.; Malvy, C.; Couvreur, P. Biodegradable Polyalkylcyanoacrylate Nanoparticles for the Delivery of Oligonucleotides. *J. Controlled. Rel.* **1998**, *53*, 137–143.
102. Chavany, C.; Saison-Behmoaras, T.; Le Doan, T.; Puisieux, F.; Couvreur, P.; Helene, C. Adsorption of Oligonucleotides Onto Polyisohexylcyanoacrylate Nanoparticles Protects them Against Nucleases and Increases Their Cellular Uptake. *Pharm. Res.* **1994**, *11*, 1370–1378.
103. Nakada, Y.; Fattal, E.; Foulquier, M.; Couvreur, P. Pharmacokinetics and Biodistribution of Oligonucleotide Adsorbed onto Poly(isobutylcyanoacrylate) Nanoparticles After Intravenous Administration in Mice. *Pharm. Res.* **1996**, *13*, 38–43.
104. Schwab, G.; Duroux, I.; Chavany, C.; Helene, C.; Saison-Behmoaras, E. An Approach for New Anticancer Drugs: Oncogene-Targeted Antisense DNA. *Ann. Oncol.* **1994**, *5*, 55–58.
105. Lambert, G.; Fattal, E.; Brehier, A.; Feger, J.; Couvreur, P. Effect of Polyisobutylcyanoacrylate Nanoparticles and Lipofectin Loaded with Oligonucleotides on Cell Viability and PKC α Neosynthesis in HepG2 Cells. *Biochimie*, *in press*.
106. Aynie, I.; Vauthier, C.; Chacun, H.; Fattal, E.; Couvreur, P. Sponglike Alginate Nanoparticles as a New Potential System for the Delivery of Antisense Oligonucleotides. *Antisense. Nucleic Acid Drug Dev.* **1999**, *9*, 301–312.
107. Gautier, J.C.; Grangier, J.L.; Barbier, A.; Dupont, P.; Dussosoy, D.; Pastor, G.; Couvreur, P. Biodegradable Nanoparticles for Subcutaneous Administration of Growth Hormone Releasing Factor (hGRF). *J. Controlled. Rel.* **1992**, *3*, 205–210.
108. Kreuter, J.; Speiser, P.P. In Vitro Studies of Poly(methyl methacrylate) Adjuvants. *J. Pharm. Sci.* **1976**, *65*, 1624–1627.
109. Kreuter, J. Possibilities of Using Nanoparticles as Carriers for Drugs and Vaccines. *J. Microencapsul* **1988**, *5*, 115–127.
110. Kreuter, J.; Liehl, E. Long-Term Studies of Microencapsulated and Adsorbed Influenza Vaccine Nanoparticles. *J. Pharm. Sci.* **1981**, *70*, 367–371.
111. Kreuter, J.; Haenzel, I. Mode of Action of Immunological Adjuvants: Some Physicochemical Factors Influencing the Effectivity of Polyacrylic Adjuvants. *Infect Immun* **1978**, *19*, 667–675.
112. Kreuter, J.; Berg, U.; Liehl, E.; Soliva, M.; Speiser, P.P. Influence of the Particle Size on the Adjuvant Effect of Particulate Polymeric Adjuvants. *Vaccine* **1986**, *4*, 125–129.
113. Kreuter, J.; Liehl, E.; Berg, U.; Soliva, M.; Speiser, P.P. Influence of Hydrophobicity on the Adjuvant Effect of Particulate Polymeric Adjuvants. *Vaccine* **1988**, *6*, 253–256.
114. Jani, P.; Halbert, G.W.; Langridge, J.; Florence, A.T. The Uptake and Translocation of Latex Nanospheres and Microspheres After Oral Administration to Rats. *J. Pharm. Pharmacol.* **1989**, *41*, 809–812.
115. Le Fevre, H.E.; Joel, D.D.; Shidlousky, G. Retention of Ingested Latex Particles in Peyer's Patches of Gerinfree and Conventional Mice. *Soc. Exp. Biol. Med.* **1985**, *179*, 522–528.
116. Damge, C.; Michel, C.; Aprahamian, M.; Couvreur, P.; Devissaguet, J.P. Nanocapsules as Carriers for Oral Peptide Delivery. *J. Control. Rel.* **1990**, *13*, 233–239.
117. Kreuter, J.; Muller, V.; Munz, K. Quantitative and Microautoradiographic Study on Mouse Intestinal Distribution of Polycyanoacrylate Nanoparticles. *Int. J. Pharm.* **1989**, *55*, 39–45.
118. Aprahamian, M.; Michel, C.; Humbert, W.; Devissaguet, J.P.; Damge, C. Transmucosal Passage of Polyalkylcyanoacrylate Nanocapsules as a New Drug Carrier in the Small Intestine. *Biol. Cell* **1987**, *61*, 69–76.
119. Damge, C.; Michel, C.; Aprahamian, M.; Couvreur, P. New Approach for Oral Administration of Insulin with

- Polyalkylcyanoacrylate Nanocapsules as Drug Carrier. *Diabetes* **1988**, *37*, 246–251.
120. Lowe, P.J.; Temple, C.S. Calcitonin and Insulin in Isobutylcyanoacrylate Nanocapsules: Protection Against Proteases and Effect on Intestinal Absorption in Rats. *J. Pharm. Pharmacol.* **1994**, *46*, 547–552.
 121. Aboubakar, M.; Puisieux, F.; Couvreur, P.; Vauthier, C. Physico-Chemical Characterization of Insulin-Loaded Poly(isobutylcyanoacrylate) Nanocapsules Obtained by Interfacial Polymerization. *Int. J. Pharm.* **1999**, *183*, 63–66.
 122. Michel, C.; Aprahamian, M.; Defontaine, L.; Couvreur, P.; Damge, C. The Effect of Site of Administration in the Gastrointestinal Tract on the Absorption of Insulin from Nanocapsules in Diabetic Rats. *J. Pharm. Pharmacol.* **1991**, *43*, 1–5.
 123. Aboubakar, M.; Couvreur, P.; Pinto-Alphandary, H.; Gouritin, B.; Lacour, B.; Farinotti, R.; Puisieux, F.; Vauthier, C. Insulin-Loaded Nanocapsules for Oral Administration: In Vitro and In Vivo Investigation. *Drug Devel. Res.* **2000**, *49*, 109–117.
 124. Damge, C.; Vranckx, H.; Balschmidt, P.; Couvreur, P. Poly(alkyl cyanoacrylate) Nanospheres for Oral Administration of Insulin. *J. Pharm. Sci.* **1997**, *86*, 1403–1409.
 125. Damge, C.; Vonderscher, J.; Marbach, P.; Pinget, M. Poly(alkyl cyanoacrylate) Nanocapsules as a Delivery System in the Rat for Octreotide, A Long-Acting Somatostatin Analogue. *J. Pharm. Pharmacol.* **1997**, *49*, 949–954.
 126. O'Hagan, D.T.; Palin, K.; Davis, S.S. Poly(butylcyanoacrylate) Particles as Adjuvant for Oral Immunization. *Vaccine* **1989**, *7*, 213–216.
 127. Marchal-Heussler, L.; Sirbat, D.; Hoffman, M.; Maincent, P. Poly(epsilon-caprolactone) Nanocapsules in Carteolol Ophthalmic Delivery. *Pharm. Res.* **1993**, *10*, 386–390.
 128. Marchal-Heussler, L.; Fessi, H.; Devissaguet, J.P.; Hoffman, M.; Maincent, P. Colloidal Drug Delivery Systems for the Eye. A Comparison of the Efficacy of Three Different Polymers: Polyisobutylcyanoacrylate, Poly(lactic-co-glycolic Acid), Poly-Epsilon Caprolactone. *S.T.P. Pharma* **1992**, *2*, 98–104.
 129. Losa, C.; Marchal-Heussler, L.; Orallo, F.; Vila Jato, J.L.; Alonso, M.J. Design of New Formulations for Topical Ocular Administration: Polymeric Nanocapsules Containing Metipranolol. *Pharm. Res.* **1993**, *10*, 80–87.
 130. Calvo, P.; Thomas, C.; Alonso, M.J.; Vila Jato, J.L.; Robinson, J.R. Study of the Mechanism of Interaction of Poly(epsilon caprolactone) Nanocapsules with the Cornea by Confocal Laser Scanning Microscopy. *Int. J. Pharm.* **1994**, *103*, 283–291.
 131. Zimmer, A.; Kreuter, J.; Robinson, J.R. Studies on the Transport Pathway of PBCA Nanoparticles in Ocular Tissues. *J. Microencapsul* **1991**, *8*, 497–504.
 132. Calvo, P.; Alonso, M.J.; Vila-Jato, J.L.; Robinson, J.R. Improved Ocular Bioavailability of Indomethacin by Novel Ocular Drug Carriers. *J. Pharm. Pharmacol.* **1996**, *48*, 1147–1152.
 133. Calvo, P.; Vila-Jato, J.L.; Alonso, M.J. Effect of Lysozyme on the Stability of Polyester Nanocapsules and Nanoparticles: Stabilization Approaches. *Biomaterials* **1997**, *18*, 1305–1310.
 134. Calvo, P.; Vila-Jato, J.L.; Alonso, M.J. Evaluation of Cationic Polymer-Coated Nanocapsules as Ocular Drug Carriers. *Int. J. Pharm.* **1997**, *153*, 41–50.
 135. Calvo, P.; Sanchez, A.; Martinez, J.; Lopez, M.I.; Calonge, M.; Pastor, J.C.; Alonso, M.J. Polyester Nanocapsules as New Topical Ocular Delivery Systems for Cyclosporin. *A. Pharm. Res.* **1996**, *13*, 311–315.
 136. Juberias, J.R.; Calonge, M.; Gomez, S.; Lopez, M.I.; Calvo, P.; Herreras, J.M.; Alonso, M.J. Efficacy of Topical Cyclosporine-Loaded Nanocapsules on Keratoplasty Rejection in the Rat. *Curr. Eye Res.* **1998**, *17*, 39–46.
 137. Le Burlais, C.A.; Chevanne, F.; Turlin, B.; Acar, L.; Zia, H.; Sado, P.A.; Needham, T.E.; Leverge, R. Effect of Cyclosporine a Formulations on Bovine Corneal Absorption: Ex-Vivo Study. *J. Microencapsul* **1997**, *14*, 457–467.

OPTIMIZATION METHODS

Gareth A. Lewis

Sanofi-Synthelabo, Chilly Mazarin, France

INTRODUCTION

What Is Optimization?

Optimization of a formulation or process is finding the best possible composition or operating conditions. Determining such a composition or set of conditions is an enormous task, probably impossible, certainly unnecessary, and in practice, optimization may be considered as the search for a result that is satisfactory and at the same time the best possible within a limited field of search. Thus, the type and components of a formulation may be selected, according to previous experience, by expert knowledge (possibly using an expert system), or by systematic screening as described later. Then the relative and/or total proportions of the excipients are varied to obtain the best endpoint, or a process is chosen, and a study is carried out to determine the best operating conditions to obtain the desired formulation properties. Both of these are optimization studies. This article concentrates on statistical experimental design-based optimization.

Screening, Factor Studies, and Optimization

Systematic screening and factor influence studies are closely related to optimization, being often sequential stages in the development process and involving statistical experimental design methods. Screening methods are used to identify important and critical effects, for example, in the manufacturing process. Factor studies are quantitative studies of the effects of changing potentially critical process and formulation parameters. They involve factorial design and are also quite often referred to as screening studies; however, the resulting relationships have just as often been used for optimization.

The type of study carried out will depend on the stage of the project. In particular, experimental design may be carried out in stages, and the experiments of a factor study may be augmented by further experiments to a design giving the detailed information needed for true optimization. It cannot be stressed too highly that the quality of a statistically designed experiment depends on the choice of experimental run with respect to an a priori model, and

this quality can and must be assessed before starting the experiments.

Brief Historical Review

Statistical methods for screening, factor studies, and optimization have been available for a long time: factorial designs since 1926 (1); screening designs since 1946 (2); and the central composite design for response surface optimization, was introduced by Box and Wilson, in 1951 (3). Their use started to be described in the pharmaceutical literature from the early 1970s, but it was only from approximately 1988 that there was a sudden increase in the number of published articles, and the numbers have continued to rise. A conception or presupposition of the difficulty or complexity of experimental design had to be overcome. The change has been attributed of course to a great extent to the availability of computing power and of relatively inexpensive high-performance software that allows previously difficult or advanced methods to be applied. In particular, much attention is now being given to robust processes and formulation, and there are developments in treating nonlinear and highly correlated responses (4).

Methods for Optimization

There are four primary methods. First, there is the statistically designed experiment, in which experiments are set up in a (normally regular) matrix to estimate the coefficients in a mathematical model that predicts responses within the limits of formulation or operating conditions being studied. This is generally the most powerful method, provided the experimentation zone has been correctly identified, and is the subject of most of this article.

Second, the direct optimization method, the best known being the sequential simplex, is a rapid and powerful method for determining an experimental domain, best combined with experimental design for the optimization itself.

Third, there is the one-factor-at-a-time method in which the experimenter varies first one factor to find the best value, then another. Its disadvantages are that it cannot be

used for multiple responses and that it will not work when there are strong interactions between factors.

Finally, the nonsystematic approach in which the knowledge and intuition of the developer allow him to improve results, changing a number of factors at the same time is often surprisingly successful in the hands of a skilled worker. Where he is less skilled or less lucky, he can waste a remarkable amount of time and resources.

The use of artificial intelligence and expert systems is treated elsewhere in this work.

SCREENING

Obtaining a Formulation Suitable for Optimization

Once the dosage form has been selected, the excipients must be identified, their choice often limited by practical considerations of time and resources determined by patents, company practice, or according to expert knowledge. However, it may be possible or necessary to test a number of different excipients for each function, for example, several diluents, lubricants, binders. This approach has proved useful in drug–excipient compatibility testing in which protoformulations are set up according to a statistical screening design to assess stability and compatibility.

Here the *factor* is the excipient's function. This can be set at different levels, the level being the excipient itself. So the factor may be "binder," and the levels are, for example, HPMC, povidone, polyvinylacetate, and no disintegrant present. A mathematical model relates the response (in this case, degradation) to composition. It includes variables corresponding to each factor with

(qualitative) levels corresponding to each excipient. Plackett and Burman (5) described designs suitable for treating this kind of problem. Designs with the factors at only two levels are widely used. However, there are other designs at 3, 4, and 5 levels as well as asymmetric designs derived from them in which the various factors take a different number of levels (5, 6).

It is assumed that there are no *interactions* between factors; that is to say, the effect of a given excipient on stability does not depend on what other excipients are found in the formulation. (The same reasoning applies to other kinds of factors or responses.) This can only be an approximation; however, if it should be necessary to take interactions into account, many more experiments would be needed, and it would probably be necessary to limit the number of levels for each factor to two for the number of experiments to be manageable.

The choice of excipients may be considered a qualitative optimization, their quantitative compositions not having yet been optimized. This and the fact that the process used will most likely be on a small laboratory scale may affect the choice of excipients. However, it is in most circumstances an unavoidable limitation.

An example of such a qualitative screening is shown in Table 1. This is an experimental design for testing the compatibilities of experimental drug (at two concentrations) with a number of number of excipients. The samples, which were wet granulated, were stored for 3 weeks at 50°C/50% relative humidity. The results are also given in Table 1. The mean degradation level was high, at 6.2%, indicating a fairly unstable drug. The effects of each excipient were calculated by linear regression, or, because the design is *orthogonal*, by linear combinations of the responses, and plotted in Fig. 1. There, the

Table 1 Experimental design and plan for granulated protoformulations

Number	Diluent	Disintegrant	Lubricant	Binder	Dose (%)	Degradation
1	Lactose	CCNa ^a	Mg stearate	Povidone	0.25	12.26
2	Cellulose ^b	CCNa	Mg stearate	HPMC	1.0	7.27
3	Phosphate ^c	CCNa	Glyceryl behenate	Povidone	1.0	11.43
4	Mannitol	CCNa	Glyceryl behenate	HPMC	0.25	4.94
5	Lactose	NaSG ^d	Glyceryl behenate	HPMC	1.0	1.63
6	Cellulose	NaSG	Glyceryl behenate	Povidone	0.25	4.56
7	Phosphate	NaSG	Mg stearate	HPMC	0.25	2.49
8	Mannitol	NaSG	Mg stearate	Povidone	1.0	4.79

^aCroscarmellose sodium.

^bMicrocrystalline cellulose.

^cCalcium hydrogen phosphate.

^dSodium starch glycolate.

If the factors are quantitative, they are set at their extreme values. Thus, if the factor is *granulation time*, and the possible range is 1.5–7 min, the normal values tested are 1.5 min and 7 min. They are expressed in terms of dimensionless *coded variables*, normally taking values -1 and $+1$. Thus, on transformation to the coded variable x_1 , 1.5 min corresponds to $x_1 = -1$, and 7 min corresponds to $x_1 = +1$.

If the factors are quantitative, they may take any number of levels. Only two-level designs are described here. Qualitative levels are set arbitrarily at the coded levels. If, for example, the screening method was one of the factors tested, wet screening could be set at -1 and dry screening at $+1$ (or vice versa).

Quite wide limits are generally chosen for screening quantitative factors. They are then often narrowed for more detailed quantitative study of the influence of factors where interactions between factors then are taken into account and for determining a predictive model for optimization.

The designs, proposed by Plackett and Burman in 1946 (2), comprise experiments in multiples of four. They will allow screening of up to one less factor than the number of experiments. Those with 2^n experiments (4, 8, 16, 32 H) are also fractional factorial designs. The nonfactorial designs have particular properties and complex *aliasing*, which has been held to make their interpretation difficult but also gives them certain advantages over the fractional factorial designs. The 12-experiment design, shown in coded variables (Table 2), is such a design, and is useful for about 7–11 factors.

The structure of the design is shown clearly in the table because the experiments are in their standard order. However, they should be carried out in a random order, as should all the designs described here, as much as is practicable.

The coefficient β_i is the *effect of the factor X_i* , and is equal to half the average change in the response y when the level of the factor is changed from $x_i = -1$ to $x_i = +1$. It is estimated (as b_i) in the Plackett–Burman design by subtracting the sum of the responses for experiments for which $x_i = -1$ from those for which $x_i = +1$ and dividing by the number of experiments. Important and unimportant effects can then be identified according to their absolute values. (Determining active factors from the results of a factorial design are shown later.)

Use of results of a screening design

Estimation of the effects allows influential or possibly influential factors to be identified. Noninfluential factors (small effects) will not require further study. They may be set at their midpoints, at their most economical values

(e.g., a short mixing time), or at their apparently best value even if the measured effect is apparently nonsignificant.

After elimination of these noninfluential factors, there may still be too many factors to optimize in terms of the resources available (time, raw material, operators, availability of equipment, etc.). Generally, these less influential factors are kept constant, equal to their best level and the remainder optimized. In more complex situations, it is advisable to carry out a more detailed study between the screening and optimization (response surface studies). This could be a completion of the screening study by means of a complementary foldover design (3, 7) or by a separate quantitative study to allow individual effects of the factors and/or their binary interactions to be calculated separately (shown in factorial designs, later).

All these studies on the process are generally done after the optimization of the formulation. However, because the effects of formulation and process changes are not generally independent, it may become necessary to carry out some sort of process study at the same time as the formulation optimization.

QUANTITATIVE PROCESS STUDIES USING FACTORIAL DESIGNS

Purpose

Whereas the purpose of a screening study is to determine which of a large number of factors have an influence on the formulation or process, that of a factor study is to determine quantitatively the influence of the different factors together on the response variables. The number of levels is usually again limited to two, but sufficient experiments are carried out to allow for *interactions* between factors.

Two-level full factorial designs

The simplest such designs are the 2^k full factorial designs, in which the experiments are all the 2^k possible combinations of two levels of k factors variables. Therefore, they consist of 4, 8, 16, 32, 64 H experiments for 2, 3, 4, 5, 6 H factors. Examples are given in Table 3 of the 2^2 , 2^3 , and 2^4 designs (each enclosed at the right and below by the solid lines). Thus, lines 1–4 of columns 1 and 2 show a 2^2 design for two factors, and lines 1–16 of columns 1–4 a 2^4 design for four factors.

The design is transformed into an experimental plan (with the natural or experimental values of the factor variable at each level $-/+1$). The mathematical model associated with the design consists of the *main effects* of each variable plus all the possible *interaction effects*,

Table 3 Some full and fractional factorial designs for two to five factors^a

	X_1	X_2	X_3	X_4	X_5	Response
1	-1	-1	-1	-1	1	189
2	1	-1	-1	-1	-1	56
3	-1	1	-1	-1	-1	94
4	1	1	-1	-1	1	80
5	-1	-1	1	-1	-1	212
6	1	-1	1	-1	1	212
7	-1	1	1	-1	1	76
8	1	1	1	-1	-1	125
9	-1	-1	-1	1	-1	351
10	1	-1	-1	1	1	534
11	-1	1	-1	1	1	275
12	1	1	-1	1	-1	219
13	-1	-1	1	1	1	154
14	1	-1	1	1	-1	752
15	-1	1	1	1	-1	374
16	1	1	1	1	1	478

^aThe response particle size (μm) in for the 2^{5-1} fractional factorial design. (From Ref. 8.)

interactions between two variables, but also between three and four factors and, in fact, between as many as there are in the model. However, although two-factor interactions are important, three-factor interactions are normally far less so. Higher-order interactions are invariably ignored and the values determined for them attributed to the random variation of the experimental system.

Determining Active Factors from the Results of a Factorial Design

We take the four-factor model as an example. The complete synergistic mathematical model consists of the constant term, four main variables (β_1x_1 H β_4x_4), six interactions between two factors ($\beta_{12}x_1x_2$, etc.), four interactions between three factors ($\beta_{123}x_1x_2x_3$, etc.) and one between four factors. The last five of these are not generally expected to be important. The model is thus:

$$y = \beta_0 + \beta_1x_1 \dots + \beta_{12}x_1x_2 \dots + \beta_{123}x_1x_2x_3 \dots + \beta_{1234}x_1x_2x_3x_4 + \varepsilon$$

The effects (coefficients) β_i in the model are estimated, usually by multilinear regression. The values obtained b_i are estimates because of the random experimental error (represented by ε in the equation). The next step is to decide which of the 15 effects calculated are *active* or important.

There are a number of ways of doing this. If the experiments have been replicated, ANOVA will reveal which effects are statistically significant. Otherwise, we rely on the fact that most of the effects are probably small and distributed randomly about zero. Thus, we look for the effects with the largest absolute values that stand out from the others (6). Making a normal probability plot of the distribution of their values is a widely used method.

The responses are usually treated separately; however, when there are a number of more or less correlated responses being studied, appropriate combinations (principal components) may be analyzed instead of the original responses (9).

Once the important effects have been identified, a simplified model can be written. If an interaction term has been identified, the corresponding main effects should also be included in the model even if they are not all found active. Thus, if the interactions between the factors X_1 and X_2 and the main effect of the factor X_1 are active, b_2x_2 should be included in the model as well as b_1x_1 and $b_{12}x_1x_2$.

Two-Level Fractional Factorial Designs

The number of experiments needed to study five or more factors in a full factorial design is large, and to determine the main effects and their interactions, a fraction of the full design is often sufficient. These are $2k-r$ fractional factorial designs, where $r = 1, 2, \text{H}$ for the half, quarter, etc. fractions. An example of a half-factorial design for five factors (2^{5-1}) is given in Table 2 (the entire table). Note that the first four columns are the same as the four factor, full-factorial design, and the column for the fifth factor is constructed by multiplying the first four columns together. Methods for constructing such designs and their limitations are described in many textbooks (5–7).

Evidently, for the 2^{5-1} design, the 16 triple and higher interactions are not determined. In fact, they are confounded with the calculated effects. Thus, the estimate of the interaction between factors one and two includes the triple interaction between the other three factors. Because the latter is assumed negligible, this does not usually matter.

Menon et al. studied the formation of pellets by fluid-bed granulation using this design (8). The five factors investigated were (X_1), the binder concentration; (X_2), the method of introducing it (dry or solution); (X_3), the atomization pressure; (X_4), the spray rate; and (X_5), the inlet temperature. Particle sizes of the resulting particles are shown in Table 3.

The coefficients of the model are calculated by linear regression (the logarithm of the particle size was used

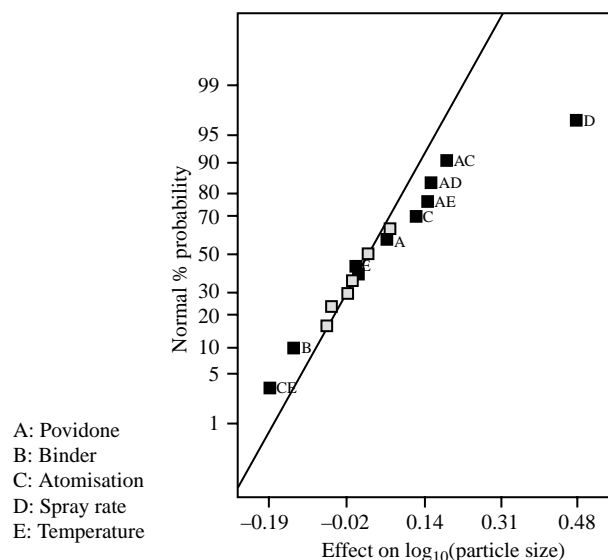


Fig. 2 Calculated effects form a two-level factorial design. Those to the left and right of the line are considered active. (From Ref. 8.)

here) and then plotted as a cumulative distribution of a normal plot (Fig. 2). The important coefficients are those that are strongly positive or negative, for example, the spray rate b_4 and the interaction between atomization pressure and inlet temperature b_{35} . Others not identified on the diagram are not considered significant and could well be representative mainly of experimental error. The equation can thus be simplified to include only the important terms. However, if interactions are included, their main effects should be included also, even if they are small. Here, we have:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 \\ + \beta_{13}x_1x_3 + \beta_{14}x_1x_4 + \beta_{15}x_1x_5 + \beta_{35}x_3x_5$$

Information That Can Be Obtained

The significant main effects are identified and also quantified. Thus, increasing the spray rate over the range studied will give an increase in the log(particle size of twice 0.24, representing a more than threefold increase. However, it can be seen that there is an interaction with the binder concentration; that is, the effect of spray rate depends on the amount of binder in the formulation. The effects of increasing spray rate are shown in Fig. 3 for both high and low levels of binder; the effect of spray rate is much greater at high levels of binder.

However, the effect of binder also interacts with two other factors, the atomization pressure and the inlet

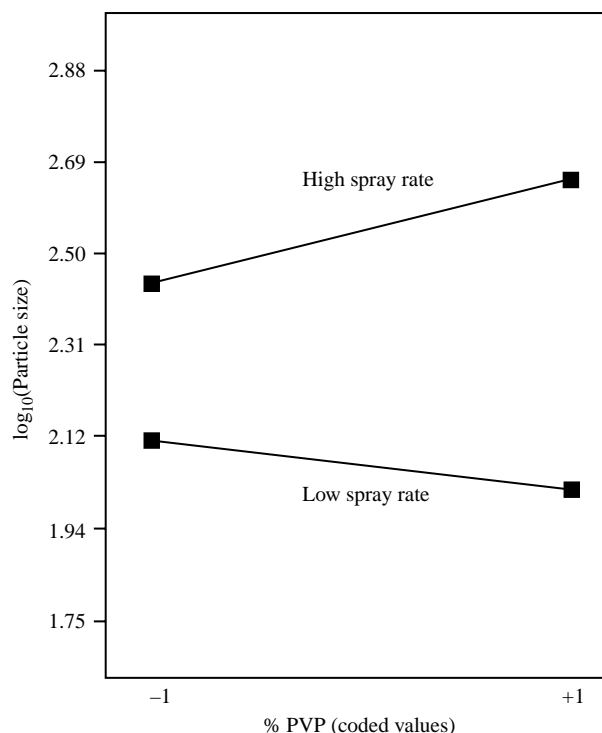


Fig. 3 Calculated effects from a two-level factorial design. Interaction diagram for spray rate and % povidone on particle size (logarithmic scale).

temperature. Thus, the individual variables cannot be considered separately.

Note also that there is a great deal of information often hidden in large designs (16 or more experiments), and, in particular, indications on factors affecting the robustness of a process may sometimes be extracted (see the last section).

Use of Center Points

In both screening and factor-influence studies in which the factor is quantitative, it is tempting to interpolate between the upper and lower limits. This is useful if only to find a more restricted zone for further study. However, in the case of a screening study, the limits studied are often so wide that it would be most unlikely for the estimated model to be accurate enough for prediction, and there is also likely to be curvature of the response surface over the experimental domain. Such attempts are less risky for the more detailed factorial studies, but even then, they should be used with caution.

Adding center points (experiments at the center of the domain, coded co-ordinates 0, 0, 0 is useful for factorial and screening experiments), even though they do not enter into the calculation of the model equation because:

1. They are often a priori at or near the most interesting conditions;
2. They allow identification of curvature in the responses (by comparing calculated with measured responses);
3. If they are replicated, the experimental reproducibility may be assessed; and
4. They may allow extension of the experiment at a subsequent stage to a central composite design for modeling of response surfaces (shown later).

EXPERIMENTAL DESIGNS FOR PROCESS OPTIMIZATION (INDEPENDENT VARIABLES)

In this section, we look at methods of obtaining a mathematical model that can be used for qualitative predictions of a response over the whole of the experimental domain. If the model depends on two factors, the response may be considered a topographical surface, drawn as contours or in 3D (Fig. 4). For more factors, we can visualize the surface by taking “slices” at constant values of all but two factors. These methods allow both process and formulation optimization.

Mathematical Models

The design used is a function of the model proposed. Thus, if it is expected that the important responses vary relatively

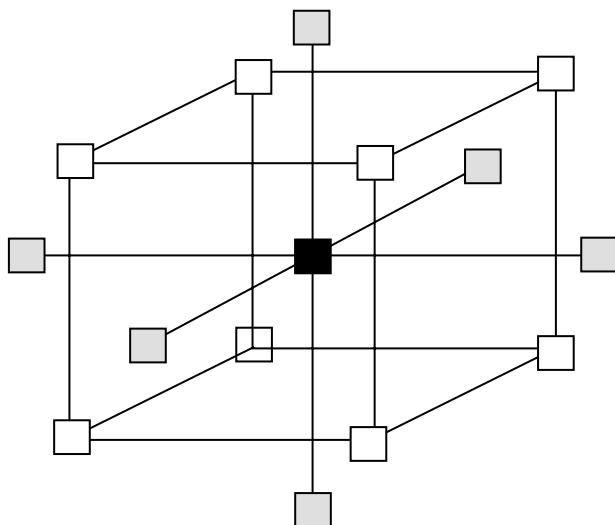


Fig. 4 Central composite design for three factors. The factorial points are shaded, the axial points unshaded, and the center point(s) filled.

little over the domain, a first-order polynomial will be selected. This will also be the case if the experimenter wishes to perform rather a few experiments at first to check initial assumptions. He may then change to a second-order (quadratic) polynomial model. Second-order polynomials are those most commonly used for response surface modeling and process optimization for up to five variables.

Examples of polynomial models are a first order model for five factors:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \varepsilon$$

and a second-order model for two factors:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \varepsilon$$

The coefficients in the models are estimated by multilinear least-squares regression of the data.

Third-order models are very rarely used in the case of process studies and, in any case, third-order terms are only added for those variables where they can be shown to be necessary (i.e., augmentation of a second-order model and the corresponding design). This does not mean that second-order designs are always sufficient, and other methods of constructing response surfaces may sometimes be useful.

Statistical Experimental Designs for First-Order Models

The design must enable estimation of the first-order effects, preferably free from interference by the interactions between factors other variables. It should also allow testing for the fit of the model and, in particular, for the existence of curvature of the response surface (center points). Two-level factorial designs may be used for this (shown earlier).

Important points to note when using a first-order model, with or without interactions, are that:

1. Maximum and minimum values of responses are of necessity predicted at the edge of the experimental domain;
2. The first-order model should normally be used only in the absence of curvature of the response surface. If the experimental values of the center points are different from the calculate values (i.e., there is lack of fit), then the response surface is curved and a second-order design and model should be used; and
3. The experimenter should test for interaction terms between two factors in the model. If interactions seem to be important he should make sure that they are properly identified.

Statistical Experimental Designs for Second-Order Models

The central composite design (Box–Wilson design)

This is the design most often used for response surfaces. It is a combination of a factorial with an axial design (3, 10) with experiments at a distance of $\pm\alpha$ along each axis (thus, the name). It requires a relatively large number of experimental runs, which can be a disadvantage if resources are limited. However, it can be carried out in two stages: the factorial design first then the axial design if the results are satisfactory.

If we wish to study the system by varying the parameters around a point of interest, the domain is a sphere, and the coordinates of axial experiments are outside those of the factorial ones. α is chosen to give the best statistical properties (e.g., constant prediction precision) and lies between 2 and approximately 2.4. The design for three factors, where α is set at 1.682, is shown in Fig. 5 and Table 4.

Center-point experiments must be done as part of both stages. Another advantage is that each factor is at five levels, thus allowing testing of lack of fit and for the possible need for cubic terms in the model.

Fig. 4 shows response surfaces calculated from the data of Senderak et al. (11) obtained using such a design at a constant value of the third factor.

Other standard designs

The central composite design is most often used, however, there are others whose particular properties make them particularly useful. One of these is the Doehlert design, which is part of a continuous hexagonal network (12). It requires slightly fewer experiments than does the central composite design but cannot be set up by augmenting a factorial design.

The design for three factors is shown in Table 5. It can be seen that the hexagonal design for two factors is the first seven rows and the first two columns. Thus, it is possible to add a factor to a design. Another advantage is that because it is part of a continuous network, it allows the experimental domain to be shifted in any direction by adding experiments at one side of the domain and eliminating them at the other (Fig. 6). Vojnovic et al. (13) give an example of its use in granulation.

Hybrid designs are saturated or almost saturated designs; that is, they have only enough experimental

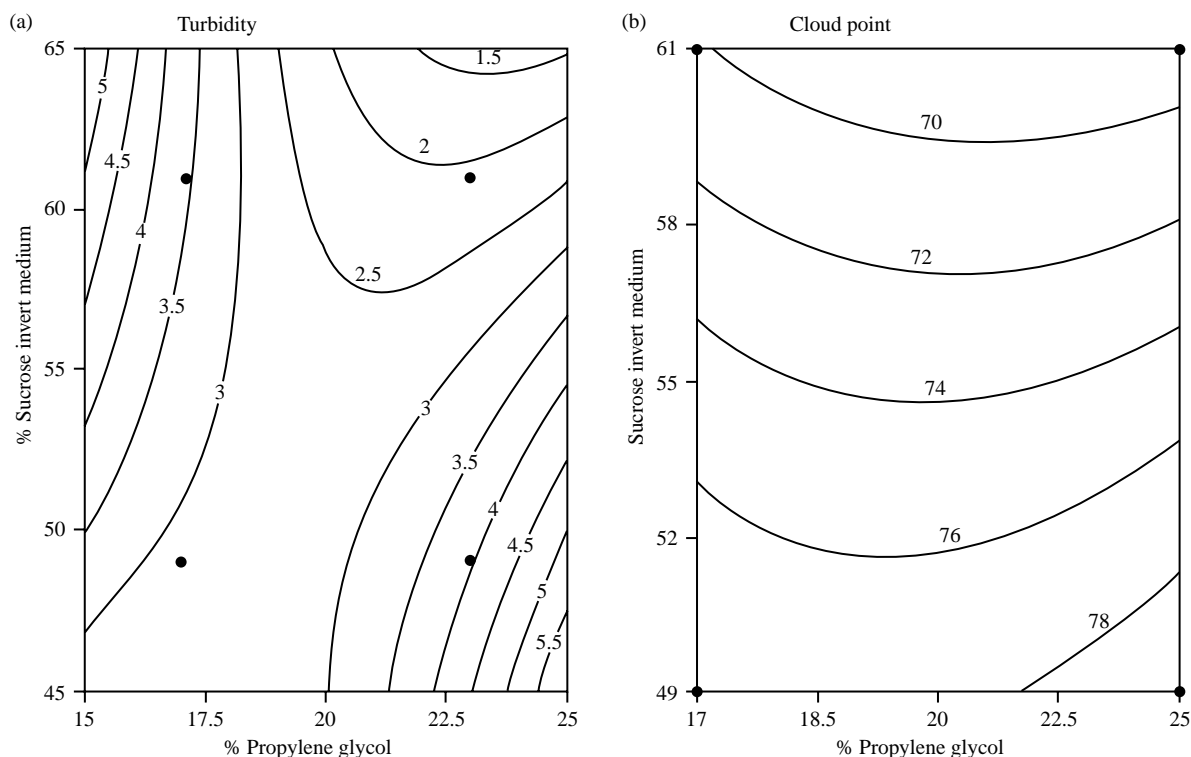


Fig. 5 Contour diagrams of (a) turbidity and (b) cloud point as function of % propylene glycol and sucrose invert medium. (slice taken at constant value of 4.3% polysorbate 80). (From Ref. 11.)

Table 4 A central composite design for three factors

Number	X_1	X_2	X_3	
1	-1	-1	-1	Factorial design 2^3
2	+1	-1	-1	
3	-1	+1	-1	
4	+1	+1	-1	
5	-1	-1	+1	
6	+1	-1	+1	
7	-1	+1	+1	
8	+1	+1	+1	
9	-1.682	0	0	Axial design
10	+1.682	0	0	
11	0	1.682	0	
12	0	+1.682	0	
13	0	0	-1.682	Center points ^a (number of replicates flexible)
14	0	0	+1.682	
15	0	0	0	
16	0	0	0	
17	0	0	0	

^aTo be included with both axial and factorial designs if carried out separately.

runs to calculate the coefficients of the quadratic model (10 runs for 3 factors 16 runs, for 4 factors, and 28 runs, for 6 factors). They are useful when the responses are not expected to vary enormously but where the quadratic model is esteemed necessary and resources (in possible numbers of experiments) are low (6, 14).

If the experimental region is defined by maximum and minimum values of each factor, then the domain is “cubic.” The central composite design can be applied to such a situation, the axial points being set then at ± 1 , coded values corresponding to the minimum and maximum allowed values. Other designs for the cubic domain are reviewed in Ref. 6.

Mixed and Irregular Domains—D-Optimal Designs

If the experimental domain is cubic and spherical or spherical, the standard experimental designs can normally be used. However, the domain may be irregular in shape as certain combinations of values variable may be excluded a priori for technical reasons or may even have been tried and failed to give a result, or certain factors may be forced to take either fixed discrete but numerical values or may even be qualitative in nature.

There are no classic experimental designs that exist for such circumstances, and a purely empirical approach is required: 1) to postulate a mathematical model that is expected to describe the response, and 2) to then select

from among the many possible experiments a design that will determine the model coefficients with maximum efficiency.

There are various ways of obtaining such a design, by far the most common being based on the exchange algorithm of Fedorov. There are also a number of criteria for describing how good the design is, the D-optimal criterion being the most usual, based on optimization of the overall precision of estimation of the coefficients of the model (6, 15, 16). This method and type of design is extremely flexible because:

Table 5 Doehlert design for three factors

k	X_1	X_2	X_3
1	0	0	0
2	1	0	0
3	0.5	0.866	0
4	-0.5	-0.866	0
5	0.5	-0.866	0
6	-0.5	0.866	0
7	-1	0	0
8	0.5	0.289	0.816
9	-0.5	0.289	0.816
10	0	0.577	0.816
11	0.5	-0.289	-0.816
12	-0.5	-0.289	-0.816
13	0	0.577	-0.816

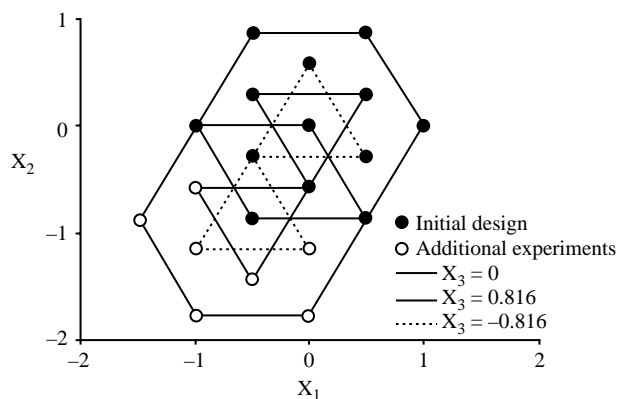


Fig. 6 Doehlert design in three dimensions (factors) showing extension to a new experimental domain.

1. It allows experiments within irregular experimental domains;
2. Previous experiments carried out within the experimental domain may be taken into account;
3. Classical designs in which experiments have failed to give a result may be repaired by redefining the domain and finding the best experiments (according to the D-optimal criterion) to replace the experiment(s) which failed;
4. The models may be polynomials with missing coefficients, or even nonpolynomials;
5. The experiments may be carried out in two or more stages, with models of increasing complexity;
6. Further experiments may be added to a D-optimal design to validate the model (lack of fit); and
7. They can be used for mixture models with constraints (see below).

In conclusion, a wide variety of experimental designs is available, allowing the design to be selected according to the problem in question, rather than adapting the experiment to the design.

EXPERIMENTAL DESIGNS FOR FORMULATION OPTIMIZATION (MIXTURE DESIGNS)

Formulations almost invariably consist of mixtures of a drug substance and excipients. Their properties usually depend not so much on the quantity of each substance present as on their proportions. The total comes to 100%, so the number of independent variables is one less than the number of components. This has the effect that the models and the designs have particular properties, and the designs

described above (screening, factor studies, and response surfaces) normally cannot be used. The entire topic of mixture designs is fully described by Cornell (17).

Mathematical Models for Mixtures

Because there is one less independent variables than the number of components, the polynomials take a particular form. For example, for three components, where the response y has a first-order dependence on the fractions x_1 , x_2 , x_3 , because $x_1 + x_2 + x_3 = 1$,

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \varepsilon$$

becomes

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

The variables cannot be varied independently. If there are no upper and lower restraints on the proportions of the components, the domain for three factors can be described as a equilateral triangle whose apices represent the pure components. A four-component mixture is described by a regular tetrahedron. For five components, the equivalent 4D figure must be imagined.

Just as the first-order mixture model has a different form from that for independent variables, so does the second-order design:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \varepsilon$$

The special cubic model describes a certain third-order curvature in the response surface:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \varepsilon$$

Mixture Designs and the Simplex Experimental Domain

The equilateral triangle and regular tetrahedron are described above as the domain of a mixture where all possible compositions of the components are allowed for are regular simplexes. (In the remainder of the section, they are referred to as simplexes.) Such circumstances in which there are no composition restraints are rare in formulation. However, if each component is present at a minimum level, and no other constraints are imposed, then the domain is also a simplex.

Designs in this case, primarily attributed to Scheffé (18), are derived very simply. That shown in Fig. 7 for three components is suitable for first-, second-, and partial

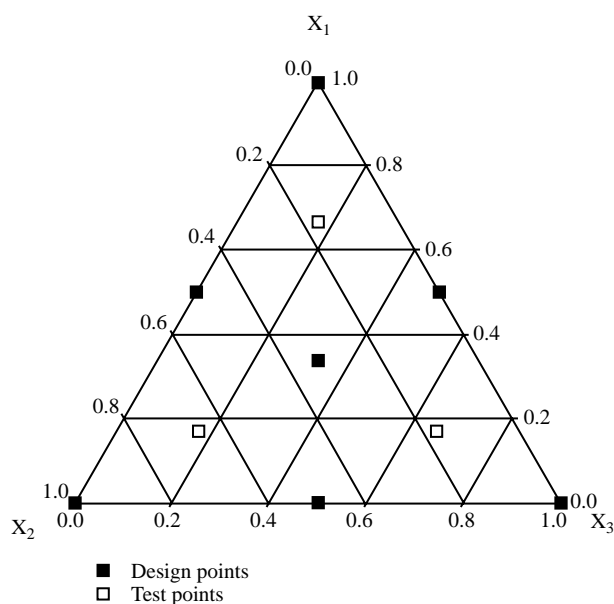


Fig. 7 Scheffé central composite design for three factors. Open squares are test points.

third-order models. The latter is the central composite design and is quite commonly used. Test points for checking model fit are also shown.

Constrained Systems and Pseudocomponents

Simplex designs are quite rarely used because such circumstances in which there are no composition restraints are rare in formulation. However, if each component is present at a minimum level, and no other constraints are imposed, then the domain is also a simplex. An example could be of the solubility of a drug being tested in ternary or quaternary mixtures of pharmaceutically acceptable solvents. The single constraint might be that a minimum percentage of water is required. In any case, the experimental domain would be a regular simplex, and standard designs may be used. In the case of solid dosage forms, simplex domains are rarer still. A possible example might be a study of the optimum composition of a diluent in a tablet formulation, the proportions of the active substance and other excipients being held constant. The diluent might consist of a mixture of lactose, microcrystalline cellulose, and starch, and its composition might then be adjusted to obtain optimum tableting properties as well as rapid disintegration and dissolution (for rapid action of the drug after the patient swallows the tablet). Again, standard experimental designs such as the simplex-centroid design may be used.

Constrained Systems and Nonsimplex Designs

Limits in the amounts of excipients present normally lead to the domain taking on an irregular shape. Each component must be present within a given concentration range to fulfill its function. For example, lactose or cellulose may make up most of the amount of a tablet or capsule, whereas magnesium stearate is limited to between 0.5 and 2%. In particular, when there are both upper and lower limits, the space is almost invariably nonsimplex.

Mixture models (such as those of Scheffé) are still useful, especially when there are three or more such excipients with fairly large ranges of variation. In solid formulations, this is often the case for diluents (or fillers) and also for the polymers or waxes incorporated into controlled-release tablets to form a matrix through which the drug diffuses slowly out when immersed in aqueous fluid, i.e., in the gastrointestinal tract.

The experimental designs of nonsimplex experimental regions are D-optimal for the selected model, obtained by an exchange algorithm (19).

Thus, we have the example of the optimization of a sustained-release tablet for which the release rate of a highly water-soluble drug was limited by its diffusion through a matrix. The matrix-forming substance is a cellulose derivative swelling in water (hydroxypropylmethylcellulose) but the diluents microcrystalline cellulose, lactose, and calcium phosphate also have a role. These four components were varied as well as the percentage of drug substance (to have two doses at constant tablet mass), and the experimental domain defined. A D-optimal design was then obtained for a second-order mixture model (using an exchange algorithm), the experiments performed, and the results analyzed by multilinear regression to give response surfaces as contour plots (Fig. 8). The formulation could thus be optimized to give the required drug release profile (6).

It is interesting to note that the work was done in two stages. Initially, experiments were chosen for a first-order mathematical model from the projected second-order design. These were carried out first, to check that there was no problem and that the experimental domain was adequate, before doing the remaining experiments for a predictive model that could be used for optimizing.

Conditions for Independent Variable Designs

If one of the components (for example, a diluent or solvent) is in considerable excess, and the limits for all other components are narrow in comparison, then it can be eliminated from the analysis because its concentration changes little. The concentrations of the remaining

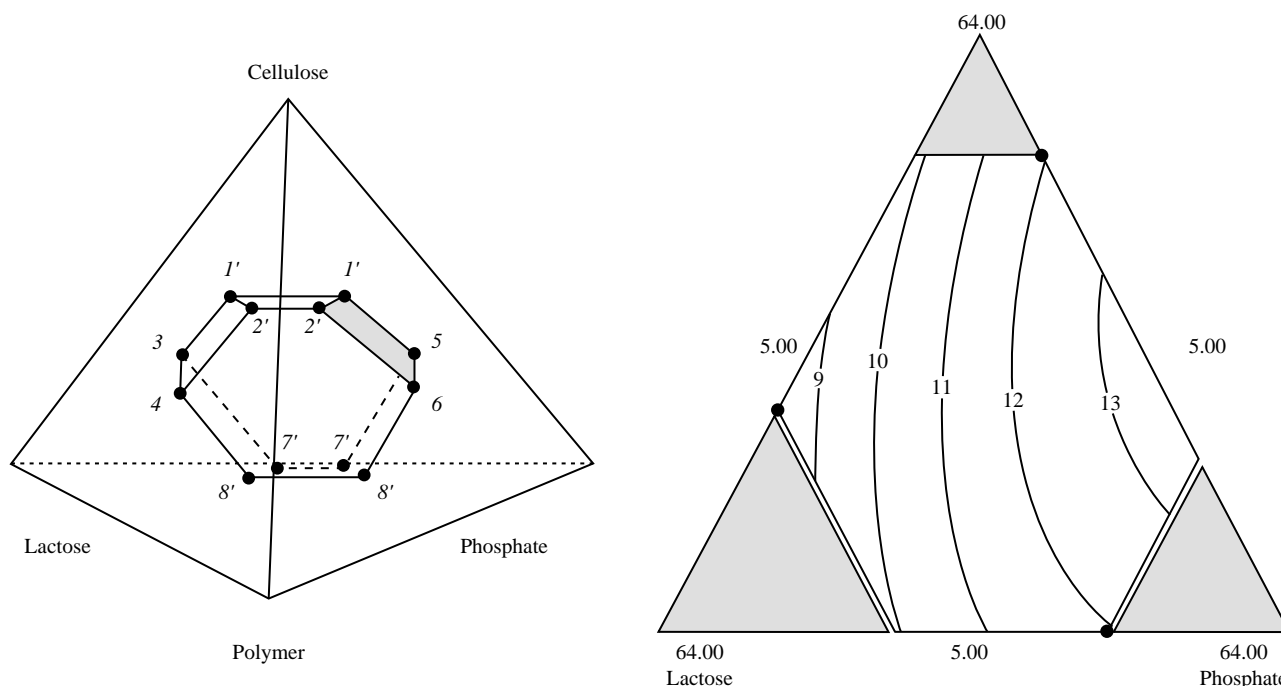


Fig. 8 D-optimal mixture design. (Left) definition of the design space. (Right) Contour plot of mean dissolution time at 25% polymer content. (From Ref. 6.)

components can then be treated as independent variables, and the methods described previously can be applied without using the special considerations for mixtures.

OPTIMIZATION METHODS USING RESPONSE SURFACE METHODOLOGY

Graphical Methods

It is usually relatively simple to find the optimum conditions for a single response that does not depend on more than four factors once the coefficients of the model equations have been estimated, provided, of course, that the model is correct. Real problems are usually more complex. In the case of pellet formation, it is not only the yield of pellets that is important but also their shape (how near to spherical), friability, smoothness, and ease of production. The optimum is a combination of all these.

One possible approach is to select the most important response, the one that should be optimized, such as the yield of pellets. For the remaining responses, we can choose acceptable upper and lower limits. Response surfaces are plotted with only these limits, with unacceptable values shaded. The unshaded area is the acceptable zone. Within that acceptable zone, we may

either select the center for maximum ruggedness of formulation or process or look for a maximum (or minimum or target value) of the key response.

Graphical Optimization of Two Opposing Responses

When there are only two independent factors (including the case of three mixture components), the responses may be plotted on a single graph. Graphics programs that allow plotting of upper and lower allowed limits of the responses, with portions of the diagram where there the responses are outside the limits shaded, are useful because they allow an acceptable zone to be identified very rapidly. An example of graphical analysis for formulation of an oral solution (11) is shown in Fig. 9. The objective was to reduce the turbidity as much as possible and to obtain a solution with a cloud point less than 70°C. A level of invert sucrose as high as possible was preferred (in spite of its deleterious effect on the cloud point). Slices were taken in the propylene glycol, sucrose plane (X_2 , X_3) at different levels of polysorbate, that at 4.3% being shown in the Fig. 9. This can be compared with the response surface in Fig. 5. An optimum compromise formulation is found at approximately 58% sucrose medium, 4.3% polysorbate 80, and 23% polyethylene glycol.

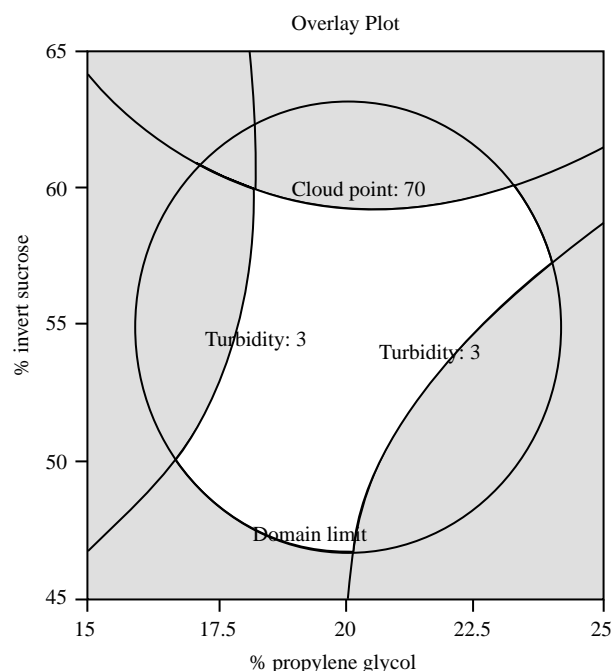


Fig. 9 Superposition of contour plots for turbidity < 3 ppm and cloud point $< 70^\circ$ to determine an optimum region ("slice" at 4.3% polysorbate 80). Compare with Fig. 5 (From Ref. 11.)

The method becomes difficult with four independent continuous factors, and for five or more variables, the method is totally impracticable despite its simplicity. The number of "slices" to be examined is simply too high—up to 125 diagrams to be displayed or plotted. Under such circumstances, the desirability method must be used.

Desirability

Derringer and Suich (6, 20) described a way of overcoming the difficulty of multiple, sometimes opposing, responses. Each response is associated with its own partial *desirability function*. If the value of the response is optimum, its desirability equals 1, and if it is totally unacceptable, its value is zero. Thus, the desirability for each response can be calculated at a given point in the experimental domain. An overall desirability function can be calculated by multiplying all of the r partial functions together and taking the r th root. Evidently, if the desirability for any response is zero at a point, the overall desirability there is also zero. The optimum is the point with the highest value for the desirability. The experimenter should study the contour plot of the desirability surface around the optimum and combine this with contour plots of the most important responses. A

large area or volume of high desirability will indicate a robust formulation or set of processing conditions.

A number of different forms, linear, convex, concave, unilateral, bilateral, are available for the dependence of the partial desirability on the value of the response. Weighting of responses is also possible. The method requires appropriate computer software, but it is a very powerful method of optimization, and with practice, it is relatively easy. It is especially appropriate for four or more factors. McLeod et al. give an example (21).

Limitations of Response Surface Methodology

The approach of using a mathematical model to map responses predictively and then to use these models to optimize is limited to cases in which the relatively simple, normally quadratic model describes the phenomenon in the optimum region with sufficient accuracy. When this is not the case, one possibility is to reduce the size of the domain. Another is to use a more complex model or a nonpolynomial model better suited to the phenomenon in question. The D-optimal designs and exchange algorithms are useful here as in all cases of change of experimental zone or mathematical model. In any case, response surface methodology in optimization is only applicable to continuous functions.

Lately, there has been a great deal of interest in the use of artificial neural networks in many fields, including that of prediction and expert systems, and they are of interest here for the description of response surfaces that have a nonlinear relation to the factor variables (22, 23). In such cases, the response surface may well fit the data better than that calculated from the model estimated by least-squares regression (24).

However, the choice of experiments is still important for the artificial neural network approach, and it is best selected in a regular pattern. The central composite design, in which each factor takes five levels, is a generally a good compromise (24). Great care must be taken not to "overfit," and, in general, more experiments are required than for the classic RSM approach.

SEARCHING FOR A NEW DOMAIN

The Steepest Ascent Method and Optimum Path Methods

Screening and factor studies will sometimes indicate whether, and if so, where we should search for an optimum within the domain being studied. However, if the optimum

(we are considering a single “key” variable here) lies outside the present experiment, then the steepest ascent method comes into its own. The direction of steepest increase of the response in terms of the coded variables is determined, and then experiments are carried out along this line. If a maximum or minimum value (according to the target) is found along this line, the point at which it is found could be the center of a new experimental design for optimization (7). The optimum path method (6, 13) is similar and is used for extrapolating from a second-order design along a curved trajectory.

Sequential Simplex Optimization

Introduction

Unlike the other optimization methods described here, the sequential simplex method for optimization neither assumes nor determines a mathematical model for the phenomena studied.

A simplex is a convex geometric figure of $k+1$ nonplanar vertices in k dimensional space, the number of dimensions corresponding to the number of independent factors. Thus, for two factors, it is a triangle, and for three factors, it is a tetrahedron. The method is sequential because the experiments are analyzed one by one as each is carried out. The basic method used a constant step size (25), allowing the region of experimentation to move at a constant rate toward the optimum. However, a modification that allows the simplex to expand and contract, proposed by Nelder and Mead (26) in 1965, is more generally used. It has been reviewed recently by Waters (27).

Optimization by the extended simplex method

Assume that we wish to optimize a response depending on three to five factors without assuming any model for the dependence other than the domain being continuous. We choose an initial domain and place a regular simplex in it. The experiments for the initial simplex are then carried out and the response measured. In the basic simplex method, an experiment is done outside the simplex in a direction directly opposite to the “worst” point of the simplex. The worst point is discarded, and a new simplex is obtained, the process being repeated. The simplex therefore moves away from the “poor” regions toward the optimum. In the extended simplex method, if the optimum is outside the initial experimental domain, we may leave it rapidly while expanding the simplex for a region with an improved response. As the simplex approaches the optimum, it is contracted rapidly.

Of the experiments of a given simplex let W , N , and B be the “worst” (W), “next worst” (N), and “best” (B) points

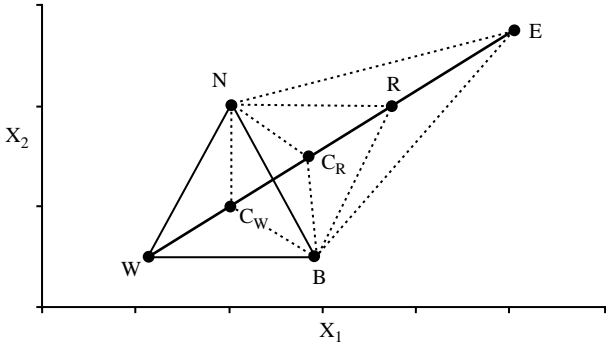


Fig. 10 Summary of the expanded simplex method of Nelder and Mead.

of the initial simplex. A new experiment R is carried out opposite point W to give a new simplex reflecting the original one. Depending on the value of the response at R relative to that at W , N , and B , the step size may be expanded to arrive quickly at the region of the optimum, and then be contracted around the optimum. The various possibilities are shown in Fig. 10 for two factors. “ $R > W$ ” means that point R is better than point W , etc.

R replaces W	if : $N \leq R \leq B$ or: $R > B$ and $E \leq B$	Reflection
E replaces W	if : $R > B$ and $E \leq B$	Expansion
C_R replaces W	if : $W < R \leq N$	Contraction (exterior)
C_W replaces W	if : $W > R$	Contraction (interior)

At the end of the sequential simplex, if more detailed information is needed, the experimenter may carry out a response surface study around the supposed optimum.

DESIGNING ROBUST PROCESSES AND FORMULATIONS

Until now, optimization and improvement have been taken as being equal or closer to what is considered most desirable with respect to the mean responses. However, it is also necessary that all units of all batches manufactured fall within those specifications. Apart from variation in the measurement method, all variation is attributed to the manufacturing process and the manufacturing and storage environment.

Taking the traditional quality control approach, any product that is within the specifications will pass and is

considered equally good. However, one might still normally consider that the nearer the response to the target, the better the product. Therefore, the key is to choose a formulation and/or condition that gives a product not only as close as possible to the target, but with as little variability as possible.

The basic concepts and seminal work in this field are from Taguchi (28), who stated that any product whose performance characteristics are different from the target values suffers a loss in quality, which he quantified by a parabolic function. He then classified factors as: 1) *control factors*, which can be controlled under normal operating conditions; and 2) *noise factors*, which are difficult, impossible, or very expensive to control.

The effects and interactions of control and noise factors could be measured by means of an experimental design, and then settings of the control factors would be determined that would minimize the effects of the noise factors.

One problem in such an approach, apart from the difficulty of controlling noise factors, is to know what they are. Examples of possible noise factors are the drug substance and excipient batches, the ambient temperature and humidity, the machine used, the exact granulation time, and the rate at which liquid is added.

The simplest approach in many cases would be to set up an experimental design in the control factors and repeat each experiment many times, hoping for enough natural variation in the noise factors to be able to find conditions to minimize variation. This requires a very large number of experiments, but it is sometimes the only possible way.

Taguchi's solution was to vary the noise factors artificially (28). A design is set up in the control factors, another (factorial) design in the noise factors, and the two multiplied together. The effect of changes in the noise factors can thus be assessed at each point and the variability minimized. This method is preferred to the previous method, but nonetheless, the number of experiments required using Taguchi's *orthogonal networks* is extremely high. Now it is more usual to set up designs in which the number of experiments, although still high, is minimized (29–31) and to find regions where the response is equal to the target value and is at the same time highly insensitive to the noise factors. The design must allow interactions among noise factors, and not only the control factors themselves, but preferably all the terms in the control factor model.

It should be noted that there is a great deal of information "hidden" in large factorial designs ($n = 16$). When analysis shows that only a few factors are significant, the residuals (differences between calculated

and measured values) may be analyzed (32). A small spread of residuals under certain conditions as opposed to others *may* indicate better reproducibility of the process or formulation under these conditions. An example of its pharmaceutical use is presented in the example of Menon et al. (8).

REFERENCES

1. Fisher, R.A. *The Design of Experiments*; Oliver and Boyd: London, 1926.
2. Plackett, R.L.; Burman, J.P. The Design of Optimum Multifactorial Experiments. *Biometrika* **1946**, *33*, 305–325.
3. Box, G.E.P.; Wilson, K.B. On the Experimental Attainment of Optimum Conditions. *J. Royal Stat. Soc. Ser. B* **1951**, *13*, 1–45.
4. Myers, R.H. Response Surface Methodology—Current Status and Future Directions. *J. Qual. Technol.* **1999**, *31* (1), 30–44.
5. Montgomery, D.C. *Design and Analysis of Experiments*; 2nd Ed.; Wiley: New York, 1984.
6. Lewis, G.A.; Mathieu, D.; Phan-Tan-Luu, R. *Pharmaceutical Experimental Design*; Marcel Dekker, Inc.: New York, 1999.
7. Box, G.E.P.; Hunter, W.G.; Hunter, J.S. *Statistics for Experimenters*; Wiley: New York, 1978.
8. Menon, A.; Dhodi, N.; Mandella, W.; Chakrabarti, S. Identifying Fluid-Bed Parameters Affecting Product Variability. *Int. J. Pharm.* **1996**, *140* (2), 207–218.
9. Carlson, R.; Nordahl, A.; Barth, T.; Myklebust, R. An Approach to Evaluating Screening Experiments when Several Responses are Measured. *Chemom. Intell. Lab. Syst.* **1992**, *12*, 237–255.
10. Box, G.E.P.; Draper, N.R. *Empirical Model-Building and Response Surface Analysis*; Wiley: New York, 1987.
11. Senderak, E.; Bonsignore, H.; Mungan, D. Response Surface Methodology as an Approach to the Optimization of an Oral Solution. *Drug Dev. Ind. Pharm.* **1993**, *19*, 405–424.
12. Doehlert, D.H. Uniform Shell Designs. *Appl. Stat.* **1970**, *19*, 231–239.
13. Vojnovic, D.; Rupena, P.; Moneghini, M.; Rubessa, F.; Coslovich, S.; Phan-Tan-Luu, R.; Sergeant, M. Experimental Research Methodology Applied to Wet Pelletization in a High-Shear Mixer. Part 1. *S.T.P. Pharma Sci.* **1993**, *3*, 130–135.
14. Roquemore, K.G. Hybrid Designs for Quadratic Response Surfaces. *Technometrics* **1976**, *18*, 419–424.
15. de Aguiar, P.F.; Bourguignon, B.; Khots, M.S.; Massart, D.L.; Phan-Tan-Luu, R. D-optimal Designs. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 199–210.
16. Atkinson, A.C. The Usefulness of Optimum Experimental Designs. *J. Royal Stat. Soc. Ser. B.* **1996**, *58* (1), 58–76.
17. Cornell, J.A. *Experiments with Mixtures*; 2nd Ed.; J. Wiley: New York, 1990.

18. Scheffé, H. Experiments with Mixtures. J. Royal Statist. Soc. Ser. B **1958**, *20*, 344–360.
19. Snee, R.D. Computer-Aided Design of Experiments—Some Practical Examples. J. Qual. Technol. **1985**, *17* (4), 222–236.
20. Derringer, G.; Suich, R. Simultaneous Optimization of Several Response Variables. J. Qual. Technol. **1980**, *12*, 214–219.
21. McLeod, A.D.; Lam, F.C.; Gupta, P.K.; Hung, C.T. Optimized Synthesis of Polygluteraldehyde Nanoparticles Using Central Composite Design. J. Pharm. Sci. **1988**, *77*, 704–710.
22. Bourquin, J.; Schmidli, H.; van Hoogevest, P.; Leuenberger, H. Basic Concepts of Artificial Neural Networks (ANN) Modeling in the Application to Pharmaceutical Development. Pharm. Dev. Technol. **1997**, *2* (2), 95–109.
23. Bourquin, J.; Schmidli, H.; van Hoogevest, P.; Leuenberger, H. Application of Artificial Neural Networks (ANN) in the Development of Solid Dosage Forms. Pharm. Dev. Technol. **1997**, *2* (2), 111–121.
24. Takahara, J.; Takayama, K.; Nagai, T. Multi-Objective Simultaneous Optimization Technique Based on an Artificial Neural Network in Sustained Release Formulations. J. Control. Rel. **1997**, *49* (1), 11–20.
25. Spendley, W.; Hext, G.R.; Himsworth, F.R. Sequential Application of Simplex Designs in Optimization and Evolutionary Operation. Technometrics **1962**, *4*, 441.
26. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. Comput. J. **1965**, *1*, 308.
27. Walters, F. Sequential Simplex Optimization—An Update. Anal. Lett. **1999**, *32* (2), 193.
28. Taguchi, G. *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*, UNIPUB/Fraus International White Plain, 1987.
29. Nair, V. Taguchi's Parameter Design: A Panel Discussion. Technometrics **1992**, *34*, 127–161.
30. Shoemaker, A.C.; Kwok, L.; Wu, C.F.J. Economical Experimentation Methods for Robust Design. Technometrics **1991**, *33* (4), 415–427.
31. Montgomery, D.C. Using Fractional Factorial Designs for Robust Process Development. Qual. Eng. **1990**, *3*, 193–205.
32. Box, G.E.P.; Meyer, R.D. Dispersion Effects from Fractional Designs. Technometrics **1986**, *28*, 19–27.
33. Anderson, V.L.; McLean, R.A. *Design of Experiments: A Realistic Approach*; Marcel Dekker, Inc.: New York, 1974.
34. Haaland, P.D. *Experimental Design in Biotechnology*; Marcel Dekker, Inc.: New York, 1989.
35. Myers, R.H.; Montgomery, D.C. *Response Surface Methodology*; Wiley-Interscience: New York, 1995.
36. Grove, D.M.; Davis, T.P. *Engineering Quality and Experimental Design*; Longman Scientific and Technical: Harlow, UK, 1992.
37. Schmidt, S.R.; Launsby, R.L. *Understanding Industrial Designed Experiments*, 3rd Ed.; Air Academic Press: Colorado Springs, 1993.

NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY IN PHARMACEUTICAL TECHNOLOGY

Kevin L. Facchine

Thomas M. O'Connell

GlaxoSmithKline, Research Triangle Park, North Carolina, U.S.A.

INTRODUCTION

Nuclear magnetic resonance spectroscopy plays a vital role in essentially all aspects of pharmaceutical development. In the early stages of lead development, NMR is used as a high throughput tool in the analysis of combinatorial libraries and in screening large numbers of compounds for receptor binding. In the synthesis and development of a lead compound, NMR is applied in all manner of structural analyses, including low-level synthetic impurities, degradation products, and metabolites. Advances in sensitivity and experimental design have enabled its use in the analysis of ever more complicated molecules at concentrations down to the nanomolar range. The coupling of NMR with HPLC systems has facilitated the characterization of trace compounds in complex mixtures. Further in-line coupling of NMR with MS systems has provided unprecedented amounts of structural information from a single analysis. In the final stage of drug development, solid-state NMR has been applied to the analysis of bulk drug properties such as polymorphic forms present in drug substances and formulations. The goal of this review is to introduce the reader to the basic principles of NMR and its application to the various stages of drug development. The references are not exhaustive, but have been selected to highlight the key principles and discoveries and point the interested reader toward further study.

PRINCIPLES OF MAGNETIC RESONANCE

The NMR Phenomenon

The nuclear magnetic resonance phenomenon occurs for certain nuclei that possess a magnetic moment μ . When placed in a magnetic field, nuclear moments will align themselves in a discrete number of orientations. The number of orientations is determined by the nuclear spin I , which is a fundamental property of certain nuclei and

has values of 0, 1/2, 1, 3/2, etc. The number of allowed orientations of a nucleus is given by $2I + 1$. The most commonly observed nuclei in NMR, for example, ^1H and ^{13}C , possess a spin equal to $\frac{1}{2}$ and therefore have only two allowed orientations. These two states can be thought of as having the nuclear dipole aligned either with the external magnetic field ($+\frac{1}{2}$, lower energy state) or against the field ($-\frac{1}{2}$, higher energy state). Nuclei such as ^{12}C and ^{16}O , have spin equal to 0 and therefore do not align in a magnetic field and cannot be observed using NMR.

The relationship between the magnetic moment and the nuclear spin is shown in Eq. 1. The proportionality constant γ is known as the gyromagnetic ratio.

$$\gamma = 2\pi\mu/hI \quad (1)$$

The gyromagnetic ratio is directly related to the sensitivity of detection of a particular nucleus. Nuclei with large γ values are more sensitive, that is, easier to detect than those with low values.

NMR experiments involve the application of pulses of radiofrequency (rf) radiation in order to excite nuclei from one energy state to another. The energy separation between these states is given by

$$\Delta E = h\nu = \gamma B_0/2\pi \quad (2)$$

where B_0 is the strength of the applied magnetic field. For the hydrogen nucleus (^1H), or proton as it is commonly referred to, the energy difference corresponds to a radio frequency of around 500 MHz with an external magnetic field of 11.7 T (1 T = 10^4 G). (Often the magnet strength is referred to in terms of the proton resonance frequency, hence an 11.7 T magnet is called a 500 MHz magnet.) Note also from Eq. 2 that γ for a given nucleus determines the relationship between the static magnetic field B_0 and the nuclear magnetic resonance frequency ν . NMR is typically considered to be one of the more insensitive spectroscopic methods due to the fact that the population difference between spin states is about 1 in 10,000 for protons in a 10 T magnetic field. This means that only about 0.01% of the

Table 1 Properties of nuclides important in the study of pharmaceuticals

Nuclide	Spin	% Natural abundance	Relative sensitivity ^a	Absolute sensitivity ^b	Gyromagnetic ratio (10 ⁷ rad s ⁻¹ T ⁻¹)	NMR frequency at 11.7436 T (MHz)
¹ H	1/2	99.985	1.00000	1.00	26.7515	500.000
² H	1	0.015	0.00965	1.45E-06	4.1065	76.753
³ H ^c	1/2	0	1.21354	0	28.5343	533.320
¹³ C	1/2	1.108	0.01591	1.76E-04	6.7281	125.752
¹⁵ N	1/2	0.366	0.00104	3.81E-06	-2.7126	50.699
¹⁹ F	1/2	100	0.83400	8.34E-01	25.1808	470.642
³¹ P	1/2	100	0.06652	6.65E-02	10.8391	202.589

^aExpressed relative to ¹H = 1 for constant field and equal number of nuclei.

^bAbsolute sensitivity = relative sensitivity × natural abundance.

^c³H is radioactive with a half life of 12.3 years.

(From Lide, D. R. Ed. *CRC Handbook of Chemistry and Physics*, 81st Ed.; CRC Press: New York, 2000, 2001; 9–92.)

sample gives rise to a signal in contrast to other spectroscopies such as IR and UV. Recent advances in commercially available superconducting magnets up to 21.1 T (900 MHz) have greatly increased the sensitivity of NMR experiments.

Table 1 lists some of the important properties of several commonly observed nuclides in the study of pharmaceuticals. Notice that some elements such as hydrogen, have several magnetically active isotopes with very different properties. Interestingly, ³H has the highest sensitivity to detection of any nucleus, but its use is limited by the added complexity of working with a radioactive isotope. The absolute sensitivity listed in the table takes into account the natural abundance of the isotope. Sensitivity can be improved in some studies by the chemical incorporation of magnetically active isotopes such as ¹³C and ¹⁵N.

The basic NMR experiment is most simply described using the vector model shown in Fig. 1. In this model, the bulk magnetization vector represents the slightly greater proportion of the nuclei aligned with the external magnetic field designated as the *z*-axis. The effect of an rf pulse generated perpendicular to the magnetic field is to tip this vector into the transverse (*x*, *y*) plane. In the transverse plane the nuclear spin imparts a precession frequency on each nucleus that is a function of its chemical environment, that is, each chemically distinct type of nucleus has a characteristic precession frequency. The detected signal is the sum of the cosine-modulated amplitudes of each of the precessing spins. Fourier transformation of this signal yields the familiar NMR spectrum. The spectrum contains valuable structural information in the form of chemical shifts, scalar couplings, relative intensities, and linewidths.

Each of these properties will be discussed in the next sections.

Chemical Shifts

When a molecule is placed in a magnetic field, electrons within the molecule shield the nuclei from the magnetic field. The actual magnetic field experienced by a given nucleus is therefore due to both the large external magnetic field, *B*₀, and the effects of nuclear shielding. The chemical shift is defined as the nuclear shielding divided by the applied field. It is always measured from a suitable reference, which is commonly a known compound added directly to the sample. The chemical shift is expressed in parts per million of the resonance frequency of the reference and calculated using Eq. 3

$$\delta(\text{ppm}) = [(\nu_{\text{sample}} - \nu_{\text{ref}}) \times 10^6] / \nu_{\text{ref}} \quad (3)$$

For proton and carbon NMR in organic solvents, the most common reference is tetramethylsilane (TMS), with four sets of equivalent methyl groups whose signal is assigned a value of 0 ppm. For aqueous solutions, trimethylsilylpropionate (TSP) is commonly used, and the trimethyl proton signal is also assigned a value of 0 ppm. The range of chemical shifts experienced by different nuclei is quite variable. The chemical shift range for ¹H is approximately 10 ppm. In contrast, several of the other nuclei listed in Table 1 have much great spectral dispersion. The ¹³C resonances typically cover a spectral window of approximately 250 ppm. For ¹⁵N and ¹⁹F, typical chemical shift ranges are greater than 400 ppm. The wide spectral dispersion experienced by the latter nuclei is very valuable in reducing spectral crowding.

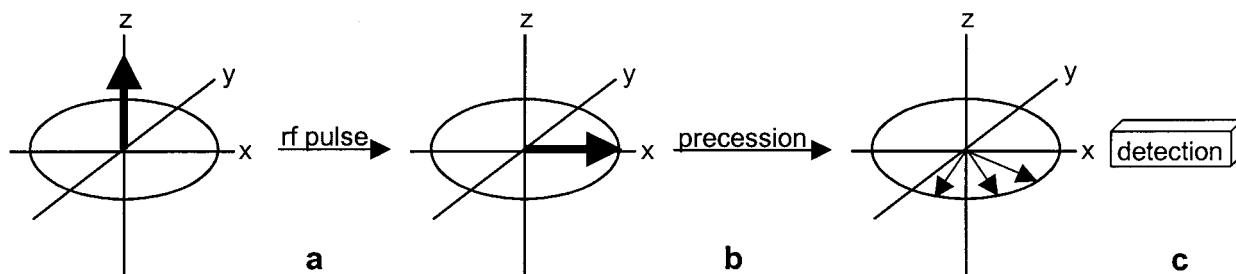


Fig. 1 The basic NMR experiment a) the rf pulse tips the bulk magnetization vector in the the x - y plane; b) the individual spins start to precess according to the frequency determined by their chemical environment; c) the precessing spins are detected along the x -axis as a set of cosine-modulated signals. Fourier transformation of these signals leads to the final spectrum.

Scalar Coupling

Signals in an NMR spectrum can present a characteristic pattern based on through bond interactions with neighboring nuclei. This phenomenon is known as scalar coupling. The proton spectrum of ibuprofen, shown in Fig. 2 has six chemically distinct sets of protons. At each chemical shift, the signals are split into multiple lines due to scalar coupling. This splitting arises from the interactions between neighboring magnetic dipoles (spins). For example, the methyl protons at C_3 of ibuprofen are next to the methine proton at C_2 . At any given time, some of the methine protons will be aligned either with or against the magnetic field. The two orientations of this nucleus result in two contributions to the effective field of the methyl group that leads to the presence of the doublet for the methyl group. The separation between the peaks is known as the coupling constant. The general rule for scalar coupling is that the number of peaks in a multiplet equals $2nI + 1$, where n is the number of neighboring equivalent nuclei and I is the nuclear spin. Thus, the methine proton at C_2 yields a quartet due to coupling to the three equivalent methyl protons. The intensities of each component of a multiplet follow the pattern of Pascal's triangle. This yields peak ratios of 1:1 for a doublet, 1:2:1 for a triplet, 1:3:3:1 for a quartet and so on. In highly crowded spectra the simple $2nI + 1$ rule can break down and more complicated patterns are seen. For more details on the interpretation of these patterns see Chapter 4 of the book by Friebolin (1). The effects of scalar coupling that bring about the multiplicity of a signal provides very useful information regarding the chemical structure.

Scalar coupling data can also yield valuable stereochemical information. The Karplus equation shown below provides a quantitative relationship between the three-bond scalar coupling constant J and the dihedral angle ϕ between two protons (2).

$$J_{\text{vicinal}} = 4.22 - 0.5 \cos \phi + 4.5 \cos^2 \phi \quad (4)$$

In this example, the coefficients above have been empirically optimized for H-C-C-H couplings. It should be noted that the form of Eq. 4 results in two possible dihedral angles for each coupling constant, but stereochemical arguments are often enough to determine a single conformation.

Integration

Another important characteristic of a signal in an NMR spectrum is its integral. Unlike other spectroscopic methods, the detected signal in NMR is directly proportional to the number of nuclei producing it. This

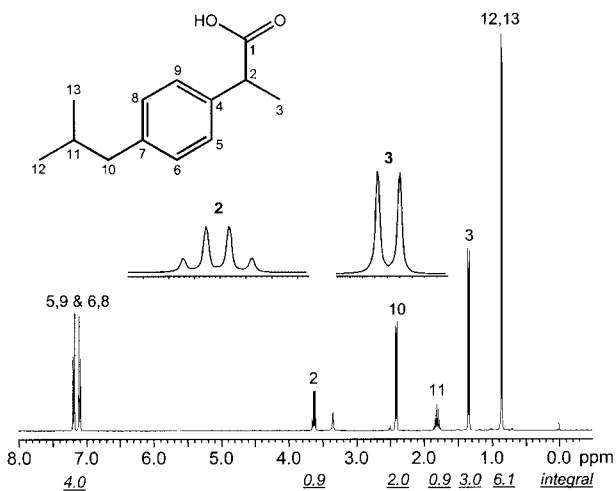


Fig. 2 One-dimensional ^1H spectrum of ibuprofen (5 mg in 600 μl DMSO- d_6 , 25°C, 400 MHz). Peak assignments are given above each set of signals. Expansions of the multiplets for the protons attached to C_2 and C_3 demonstrate the $2nI + 1$ rule of scalar coupling. The integral values are given below the chemical shift axis. Note that the peak for the hydroxyl proton at 11.1 ppm is not shown. Small peaks at 3.3, 2.5, and 0 ppm are from H_2O , DMSO- d_5 , and TMS, respectively.

feature is very useful in the interpretation of NMR spectra since the relative integrals of the signals directly correlate to the relative numbers of nuclei in the molecule giving rise to those signals. The integral values of the ibuprofen signals are displayed below the chemical shift axis in Fig. 2. A caveat in the quantitative interpretation of integrals is that different nuclei require different amounts of time to relax back to equilibrium between pulses. When a signal is not fully relaxed between pulses the amount of magnetization that will be tipped into the x - y plane by the subsequent pulse will be reduced thereby diminishing the integral value. These effects are seen in the slight deviations from the expected integer values in the ibuprofen spectrum. Precise and reliable quantitation of NMR signals therefore requires long delays between pulses. More details of relaxation follow.

Relaxation

After the rf pulse has tipped the magnetization into the transverse plane and the spins begin to precess, nuclear relaxation also begins to bring the signals back to equilibrium along the z -axis. The mechanisms by which energy is transferred to effect relaxation and the rates at which these mechanisms occur are very useful parameters and can be measured experimentally. There are several different relaxation mechanisms, which are more or less important under particular sets of conditions. In solution state studies, the dipole-dipole relaxation mechanism is by far the most important. Any given nucleus in a molecule is surrounded by other dipolar nuclei which are in motion. This motion leads to fluctuating magnetic fields, which can cause nuclear transitions and hence relaxation. Dipolar relaxation is therefore highly dependent upon the molecular motion. The three main forms of relaxation that are important for solution state structural studies are spin-lattice (T_1), spin-spin (T_2), and the nuclear Overhauser effect (nOe). These three mechanisms and their impact on molecular structure analysis are described below.

Spin-lattice relaxation is a process by which the excited spins give up energy to the surroundings (the lattice). This type of relaxation is most efficient when the molecule tumbles at a rate that is very close to the resonance frequency of the nucleus being studied. The rate of tumbling of a molecule is described by the correlation time τ_c . The correlation time can be approximated by

$$\tau_c = 4\pi\eta a^3 / 3kT \quad (5)$$

where η is the viscosity of the solution and a is the radius of the molecule. An approximation of the correlation time

in a typical organic solvent can be derived from the molecular weight (MW), using

$$t_c \approx MW \times 10^{-12} \quad (6)$$

For a proton in an 11.7 T magnet, T_1 relaxation will be fastest when the correlation time is on the order of 500×10^{-6} s. The dependence of T_1 upon correlation time is given in Fig. 3. In typical organic solvents the T_1 minimum is reached for molecules with a molecular weight in the neighborhood of 1–2 kDa.

The T_1 relaxation time is the critical relaxation parameter in obtaining quantitative integrations. For essentially complete relaxation to occur after application of a 90° pulse, a delay time of about $5 \times T_1$ between pulses is recommended, but for nonquantitative analyses, the time can be closer to $1 \times T_1$. In practice, pulses much less than 90° are often used, which allows for more quantitative spectra in less time. For protons on small molecules ($MW < 1000$), typical ^1H T_1 relaxation times range from less than one to several seconds. For ^{13}C , the relaxation times are much longer, in some cases greater than 60 s. The extremely long delays required for accurate ^{13}C integration limit the practical application of quantitative ^{13}C NMR. T_1 times can be readily measured using the inversion-recovery experiment in which the signal intensities are modulated by a series of pulses and delays (1). Curve fitting of the signal intensities as a function of the delays yields the T_1 values.

Spin-spin relaxation is a process in which there is no net loss or gain of energy, but the spins lose phase coherence. The basis of this type of relaxation is the

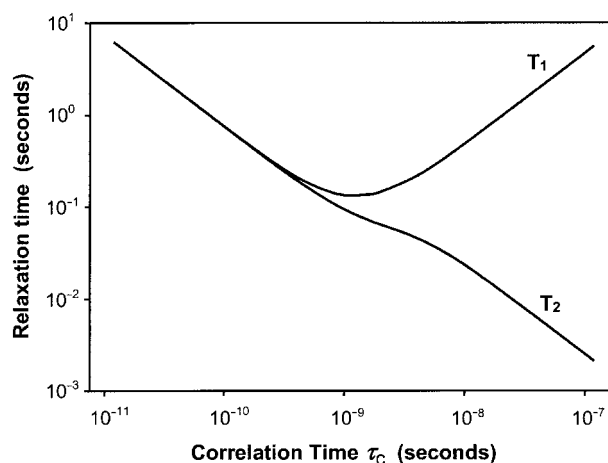


Fig. 3 Plot of T_1 and T_2 relaxation versus correlation time. The curves have been computed for the dipolar relaxation between two protons separated by 2.0 Å in a magnetic field of 11.7 T.

transfer of energy from one nucleus to another via the fluctuation magnetic fields. As one spin is excited to a higher state, another is relaxed back to the lower state. The net result of these transitions is that the phases of particular types of spins spread out or dephase. As the spins continue to dephase, they will eventually cancel one another resulting in the loss of the signal.

The T_2 time has a critical relationship to the linewidth of the NMR signal at half height ($\Delta\nu$) given by

$$\Delta\nu_{12} = 1\pi T_2^* \quad (7)$$

The term T_2^* is used to denote inclusion of the contributions to T_2 from magnetic field inhomogeneities. When the magnetic field is not totally homogeneous, the small differences in the magnetic field experienced by the same nucleus at different locations in the sample gives rise to slight differences in the precessional frequencies. This leads to the same type of phase coherence loss as the inherent T_2 processes. T_2 relaxation times can be measured using spin-echo experiments designed to eliminate the contributions of field inhomogeneities. These experiments are well described in the book by Friebolin (1).

As with T_1 relaxation, T_2 relaxation has a strong dependence upon the molecular correlation time. Unlike T_1 relaxation, T_2 relaxation does not reach a minimum and then increase, but continues to decrease, as shown in Fig. 3. Therefore large, slowly tumbling molecules have very short T_2 times. This poses a great challenge in the study of large molecules or molecules in the solid state since the lifetime of the signal is very short and the linewidths are very broad.

The nuclear Overhauser effect is the most widely measured of the relaxation phenomena in structural studies. The nOe experiments directly measure the dipole-dipole relaxation between nuclei. The great utility of the nOe is its potential to determine internuclear distances. The magnitude of the nOe is proportional to $(r_{IS})^{-6}$ where r_{IS} is the distance between spins I and S. The nOe effects an increase or decrease in the intensity of a particular signal, based on the spacial proximity of its neighbors and the dynamics of the molecule. For a more rigorous description of the physical basis of the nOe the reader is referred to the excellent text by Neuhaus and Williamson (3).

The dependence of the nOe upon molecular motion is shown in Fig. 4. Notice that both the sign and magnitude of the nOe are a function of the correlation time. An important feature of the nOe is that for molecules in a particular tumbling regime, the nOe passes through zero. This is important for many pharmaceutically relevant compounds since for some molecules with molecular

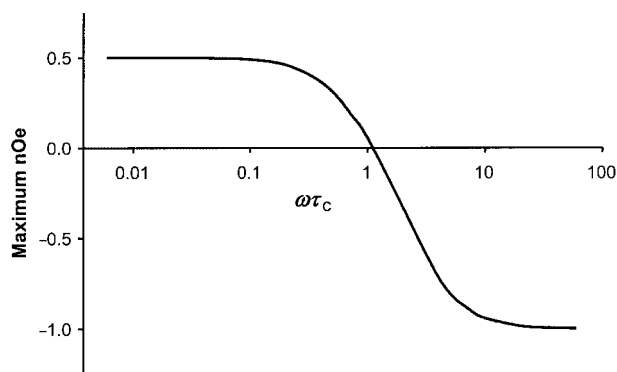


Fig. 4 Plot of maximum nOe versus $\omega\tau_c$ where $\omega (= 2\pi\nu)$ is the precessional frequency of the nuclei and τ_c is the correlation time. The curve has been computed for a pair of protons separated by 2.0 Å in a magnetic field of 11.7 T.

weights around 1–2 kDa in a viscous solvent like DMSO the tumbling rate can result in nOes very near zero. In this situation no nOe will be observed, irrespective of molecular structure. For these studies, another form of the nOe is more useful. The rotating frame Overhauser effect (rOe) is also function of dipolar relaxation, but occurs under a different set of experimental conditions (4, 5). The advantage of the rOe is that the effect does not pass through zero. The rOe experiments are somewhat more difficult to setup than nOe experiments and quantitative interpretation is more difficult, but they have proven very valuable in the structural analysis of many pharmaceuticals.

NMR TECHNIQUES

NMR has been an integral part of the chemist's analytical toolbox for decades. The most common and fundamental experiment is the one-dimensional (1D) ^1H experiment. The relatively high sensitivity of the ^1H nucleus makes this a very useful start, but for a complex molecule the 1D spectrum can be crowded and often uninterpretable. In these cases more advanced techniques can be used to provide increased resolution and specific types of structural information. In this section we will describe a range of NMR techniques from the simplest 1D experiments to complex multidimensional, multinuclear experiments. The focus of this section will be on the general principles underlying these experiments and their applications to molecules of pharmaceutical interest. Further details on these experiments can be found in the references.

One-Dimensional Experiments

A one-dimensional NMR spectrum provides a wealth of information, which for simple molecules may yield enough detail for complete structural characterization. The ^1H spectrum of a 5 mg sample of ibuprofen shown in Fig. 2 was taken in less than 1 min. Given the relative simplicity of this molecule, all of the ^1H resonances can be assigned by inspection. Full characterization of pharmaceutical molecules often requires that the ^{13}C chemical shifts be assigned as well. However, ^{13}C NMR spectra require much longer acquisition times than ^1H spectra due to the relative sensitivity and low natural abundance (see Table 1). Fig. 5 shows the 1D ^{13}C spectrum of the same 5-mg sample of ibuprofen taken in about 30 min. The broad chemical shift range for ^{13}C yields spectra with typically well-resolved peaks. This spectrum was acquired with the proton decoupler on throughout the experiment, which collapses all of the ^{13}C multiplets into singlets. Additionally the ^1H decoupling provides a signal enhancement of up to 200% due to the ^1H – ^{13}C nOe. [Further details on heteronuclear nOes can be found in Chapter 2 of Ref. (3)]. For relatively simple compounds in which there are ample amounts of sample, the 1D ^1H and ^{13}C spectra can yield unequivocal assignments.

For samples such as natural products, metabolites, and degradation products, the sample amounts are typically far less than the 5 mg used in the example above. In these cases the ^1H spectrum is still a necessity, but it may be impractical to acquire a ^{13}C spectrum. The signal to noise ratio (S/N) for NMR spectra is a linear function of sample concentration, but builds up as the square root of the number of scans. Therefore, if the sample concentration is reduced by a factor of 100, it will take 100^2 times the number of scans to achieve a spectrum with the same S/N .

Spin Polarization Transfer

Several methods of signal enhancement have been developed to help overcome the inherently low sensitivity of important nuclei such as ^{13}C and ^{15}N . These experiments rely on the principle of spin polarization transfer where magnetization on the more sensitive nuclei, usually ^1H , is transferred to the attached nuclei of interest through a specific sequence of pulses and delays. The signal enhancement achievable through polarization transfer is proportional to γ_I/γ_S , where I and S represent the abundant and rare spins, respectively. This corresponds to a factor of 4 for ^{13}C and a factor of 10 for ^{15}N (see Table 1) and can yield dramatic savings in

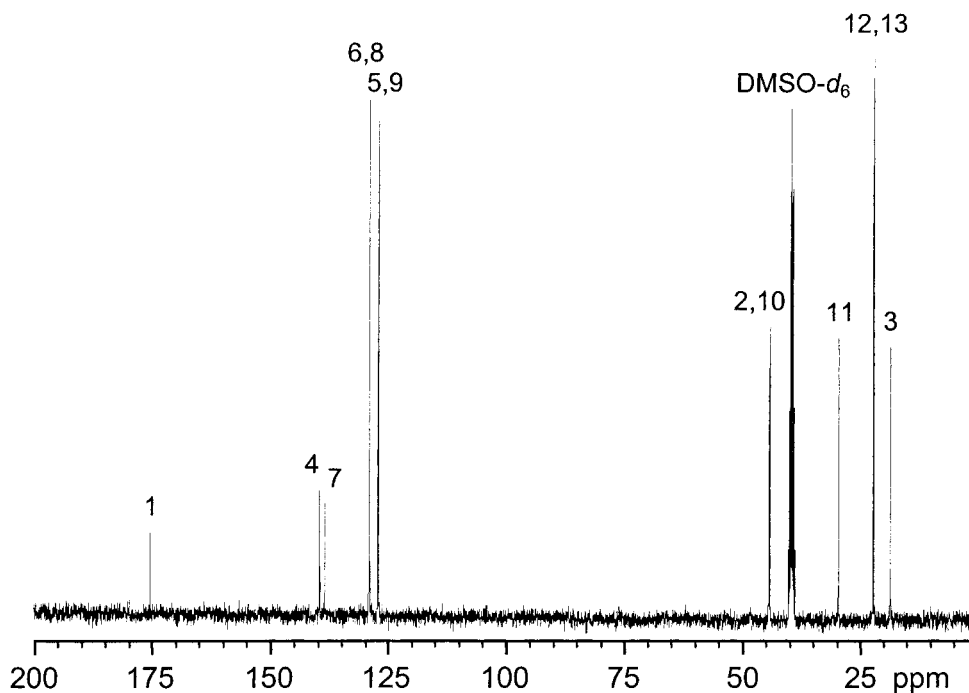


Fig. 5 ^{13}C NMR spectrum of ibuprofen (5 mg in 600 μl DMSO- d_6 , 25°C, 100 MHz). Peak assignments are given above the signals and refer to the chemical numbering shown in Fig. 2. Note the overlap between C_2 and C_{10} .

acquisition time. Polarization transfer experiments also enable use of faster repetition rates based on the typically shorter T_1 relaxation rates of the protons thereby providing an additional gain in overall sensitivity. Two methods of polarization transfer used in solution NMR are the INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) experiment (6) and the DEPT (Distortionless Enhancement by Polarization Transfer) experiment (7). The latter experiment provides the added feature of spectral editing such that subspectra containing different types of carbons can be obtained or a single spectrum generated where methyl and methine carbon resonances are phased positive and methylene carbon resonances are negative. Quaternary carbons are not observed in these experiments since the signals rely on the transfer of polarization from attached protons.

Two-Dimensional Experiments

As the size and complexity of a molecule increases, the proton NMR spectrum can easily become crowded and very difficult to interpret. Steroids are a common class of compounds of pharmaceutical interest and provide a very good example of spectral crowding. Steroid ^1H NMR spectra typically contain a region between approximately 1 and 4 ppm known as the methylene envelope in which a large proportion of the proton signals reside (8). In this situation, two-dimensional (2D) NMR can be used to spread the signals out into a second dimension providing an invaluable increase in resolution. The simplest of the 2D experiments is the COSY (CORrelation SpectroscopY) experiment. COSY spectra are characterized by two types of peaks, those along the diagonal and crosspeaks located off of the diagonal. The diagonal is effectively a 1D spectrum, and the crosspeaks indicate a scalar coupling relationship between the two resonances at coordinate chemical shifts of the crosspeak. Fig. 6 shows an expansion of the double quantum filtered COSY spectrum of the highly crowded region of dutasteride. The double quantum filtered COSY is a simple modification of the COSY experiment, which, among other benefits, yields narrower diagonal peaks allowing for crosspeaks close to the diagonal to be better resolved (9). Notice that the crosspeaks in the spectrum are characterized by a specific pattern, which contains the scalar coupling information between the correlated protons. COSY spectra can provide both connectivity and scalar coupling constants from highly crowded spectra, although the latter is not always straightforward.

TOCSY (TOtal Correlation Spectroscopy) is another important homonuclear 2D correlation experiment where

correlations arise due to the presence of homonuclear scalar coupling (10). In the standard COSY experiment, crosspeaks appear for spins in which the scalar coupling occurs over typically two to four bonds. In the TOCSY experiment crosspeaks can appear for spins separated by many more bonds as long as they are part of a contiguous network of coupled spins. The correlations are effected by the application of a series of low-power rf pulses termed the spin-lock. The duration of the spin-lock period determines the extent to which the correlations are propagated through the spin system. The TOCSY experiment is a useful complement to the COSY methods for the elucidation of complex structures.

The 2D nOe experiment NOESY (Nuclear Overhauser Enhancement SpectroscopY) provides correlations between nuclei that are close in space. While nOe experiments can be carried out using selective excitation of individual signals, one at a time, to determine the identity of protons proximal to the selected peak, the NOESY experiment enables the determination of nOe information between all spins in one experiment. Cross peaks in a NOESY experiment indicate which protons are close to one another. Typically nOe crosspeaks can be observed for protons less than 5 Å apart in the molecule. Fig. 7 shows the NOESY spectrum of dutasteride with several key correlations highlighted. The results of this experiment enable complete stereochemical assignment of all the protons in the molecule.

Two dimensional experiments can also show correlations between different types of nuclei. These heteronuclear experiments have the advantage that nuclei such as ^{13}C and ^{15}N have much wider chemical shift ranges, and therefore the 2D experiments achieve a tremendous reduction in spectral crowding. The HETCOR (HETeronuclear CORrelation) experiment was the first 2D experiment developed to provide correlations between ^1H and ^{13}C . In this experiment, the magnetization is transferred from a proton to its attached carbon via the one bond scalar coupling in a manner similar to the DEPT experiment. The acquisition parameters can be modified to allow for the correlation of protons to carbons several bonds away by optimizing for the smaller multiple bond heteronuclear scalar coupling constants. The multiple bond correlations can be combined with the single bond correlations to put together the pieces of a molecule for complete structure elucidation or spectral assignment. While it provides important C–H connectivity information, the HETCOR experiment suffers from relatively poor sensitivity resulting from the detection of the carbon signals.

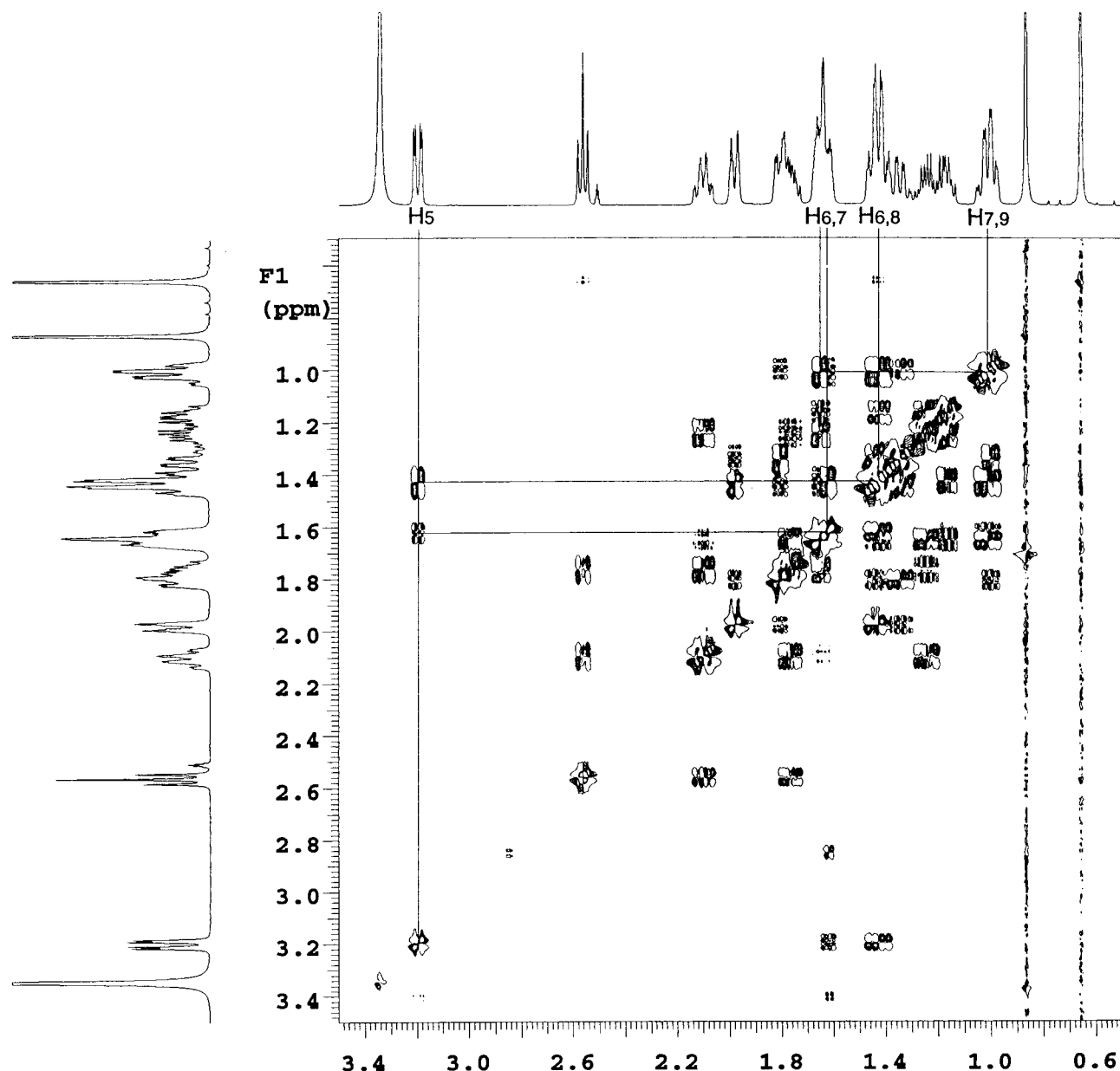


Fig. 6 Aliphatic region of a DQF-COSY spectrum of dutasteride in DMSO- d_6 . Correlations show scalar connectivities between neighboring protons and facilitate proton resonance assignments, which begin with H₅ and are traced throughout the spin system. (See Fig. 10 for structure and numbering scheme.) Complete analysis is complicated by overlapping resonances.

Inverse Detection Experiments

Inverse detection experiments consist of a sequence of pulses and delays that transfer the magnetization from ^1H to the attached X nuclei (typically ^{13}C or ^{15}N) and then back to ^1H for detection. In this way, the experiments maintain a sensitivity level much closer to

a standard ^1H experiment and represent a marked improvement over the HETCOR experiments. The first of these types of experiments was the HMQC (Heteronuclear Multiple Quantum Coherence) experiment (11). Fig. 8 shows the HMQC spectrum of dutasteride. The utility of this experiment is the ability to resolve overlapping proton resonances by spreading

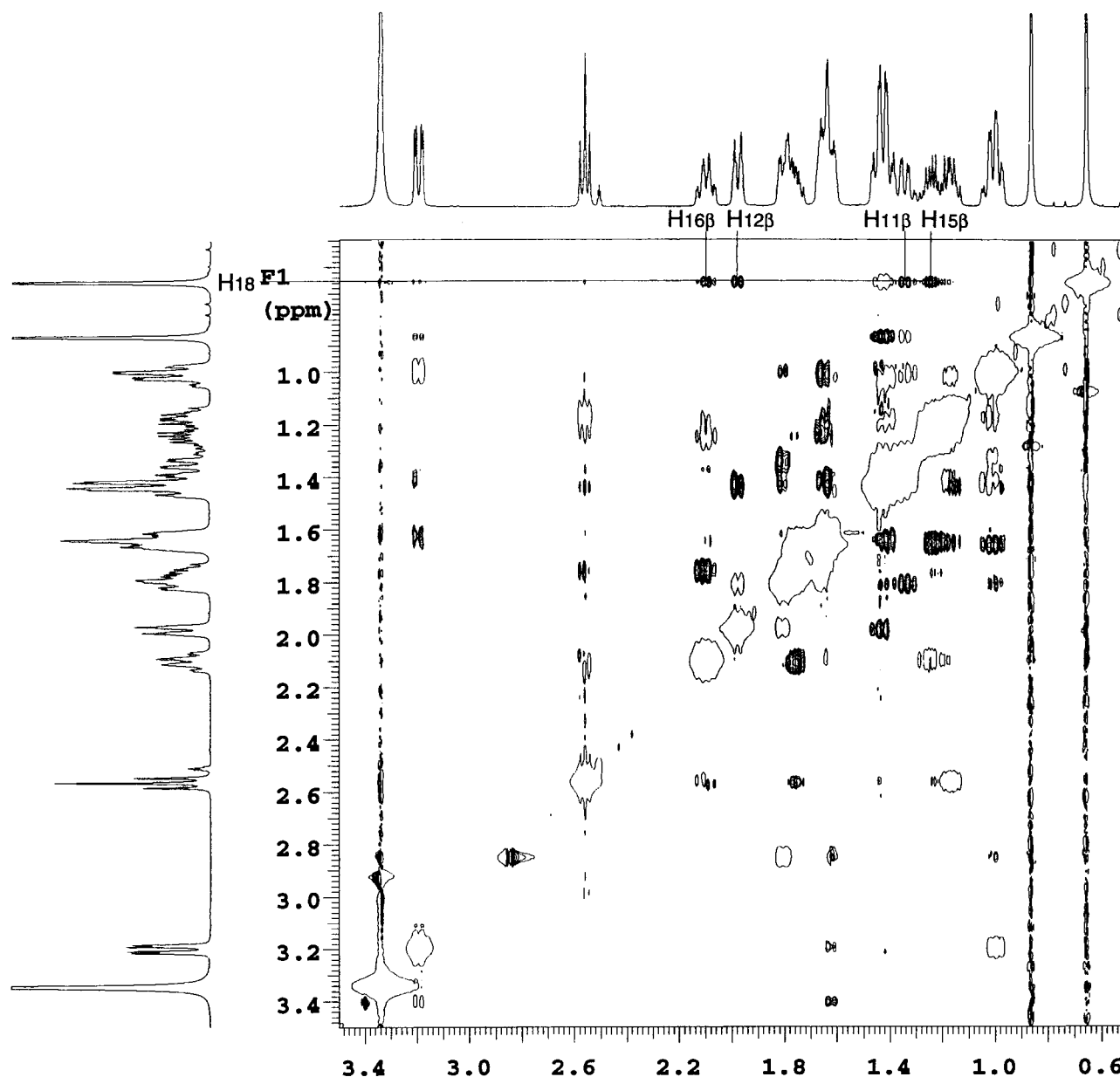


Fig. 7 Aliphatic region of a NOESY spectrum of dutasteride in DMSO- d_6 . Correlations show through space interactions between neighboring nuclei and are used to make relative stereochemical assignments for proton resonances. Correlations between the H₁₈ methyl group and nearby protons on the same side of the steroid ring system are highlighted. (See Fig. 10 for structure and numbering scheme.)

them out over the carbon chemical shift dimension and assign carbon resonances from known proton assignments or vice versa. Some key correlations are highlighted. Note that inequivalent methylene protons are readily identified as they show correlations to the same carbon. This experiment contains exactly the same type of information as the HETCOR, but spectra of

equivalent S/N can be acquired in about one-fourth the time for ^1H - ^{13}C correlations and one tenth the time for ^1H - ^{15}N correlations. Given the inherently low sensitivity of ^{15}N , direct observation of ^{15}N is rarely done, but inverse detection methods have made ^1H - ^{15}N correlations much more facile. A recent review by Martin describes the utility of the ^1H - ^{15}N inverse

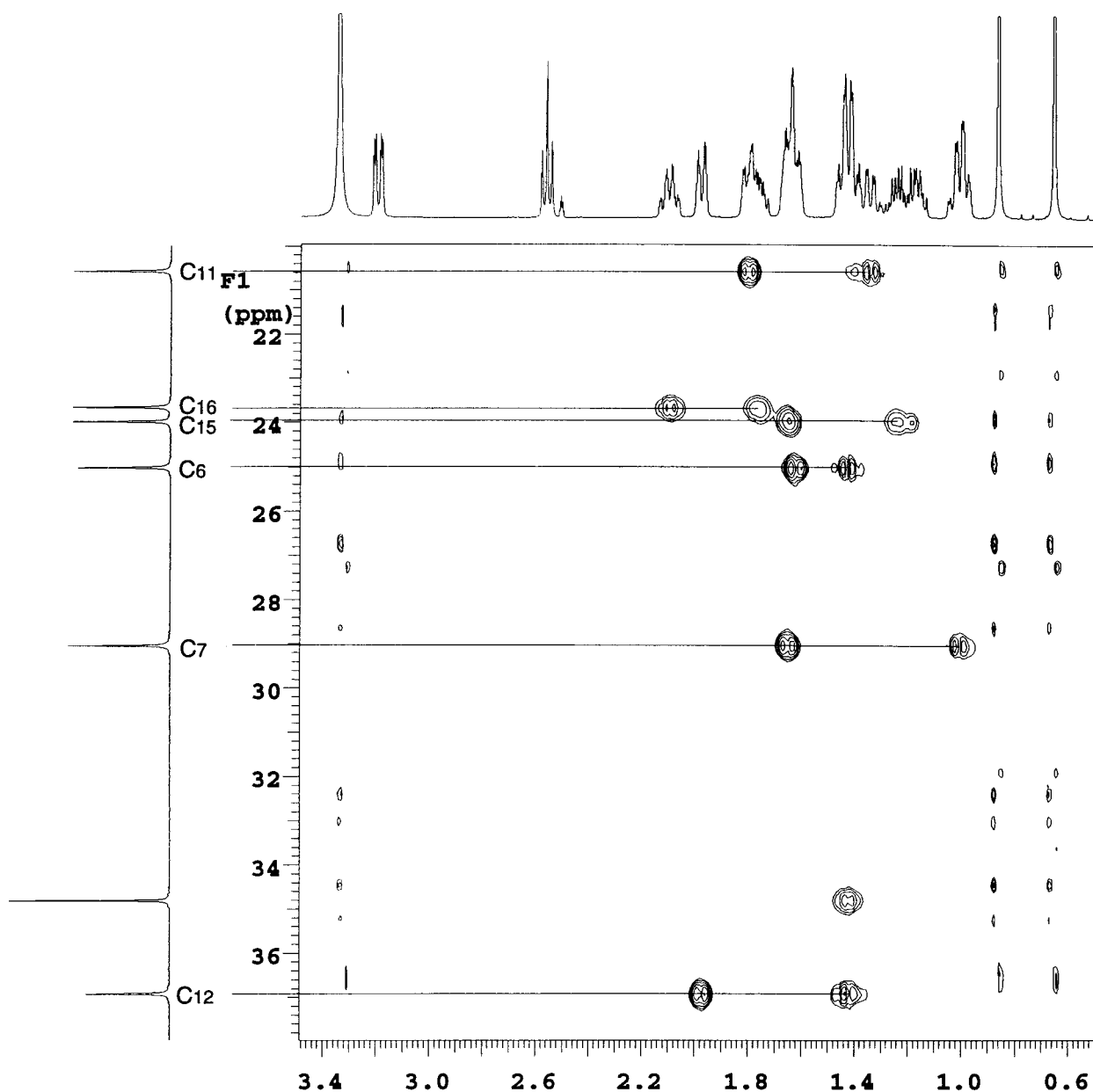


Fig. 8 Expansion of the HMQC spectrum of dutasteride in DMSO- d_6 showing heteronuclear correlations used to identify pairs of methylene protons and assign carbon resonances based on proton assignments. The six methylene proton pairs highlighted are readily identified since there are two protons with different chemical shifts correlated to each carbon. (See Fig. 10 for structure and numbering scheme.)

detection experiments in the structural determination of natural products (12).

The HSQC (Heteronuclear Single Quantum Coherence) experiment is another widely used inverse detection experiment (13, 14). It provides essentially the same information as HMQC, but relies on a different sequence

of pulses to effect the transfer of magnetization between ^1H and the heteronucleus. A direct comparison of HMQC and HSQC in the study of a natural product has indicated some advantages of the latter-sequence, which may provide improved sensitivity and narrower crosspeaks for improved resolution (15).

Long range proton connectivities to heteronuclei can also be obtained from inverse detection experiments. A modification of the HMQC referred to as the HMBC (Heteronuclear Multiple Bond Coherence) experiment can be optimized for transfer through multiple bonds based on the value of the multiple bond couplings (16), which is

typically about 8 Hz for a three-bond ^1H - ^{13}C coupling. The HMBC spectrum of dutasteride shown in Fig. 9 illustrates the utility of this experiment for assigning quaternary carbons, connecting isolated spin systems (e.g., H_{18} and H_{19}) to other spin systems in the molecule, and confirming assignments made from the COSY and HMQC spectra.

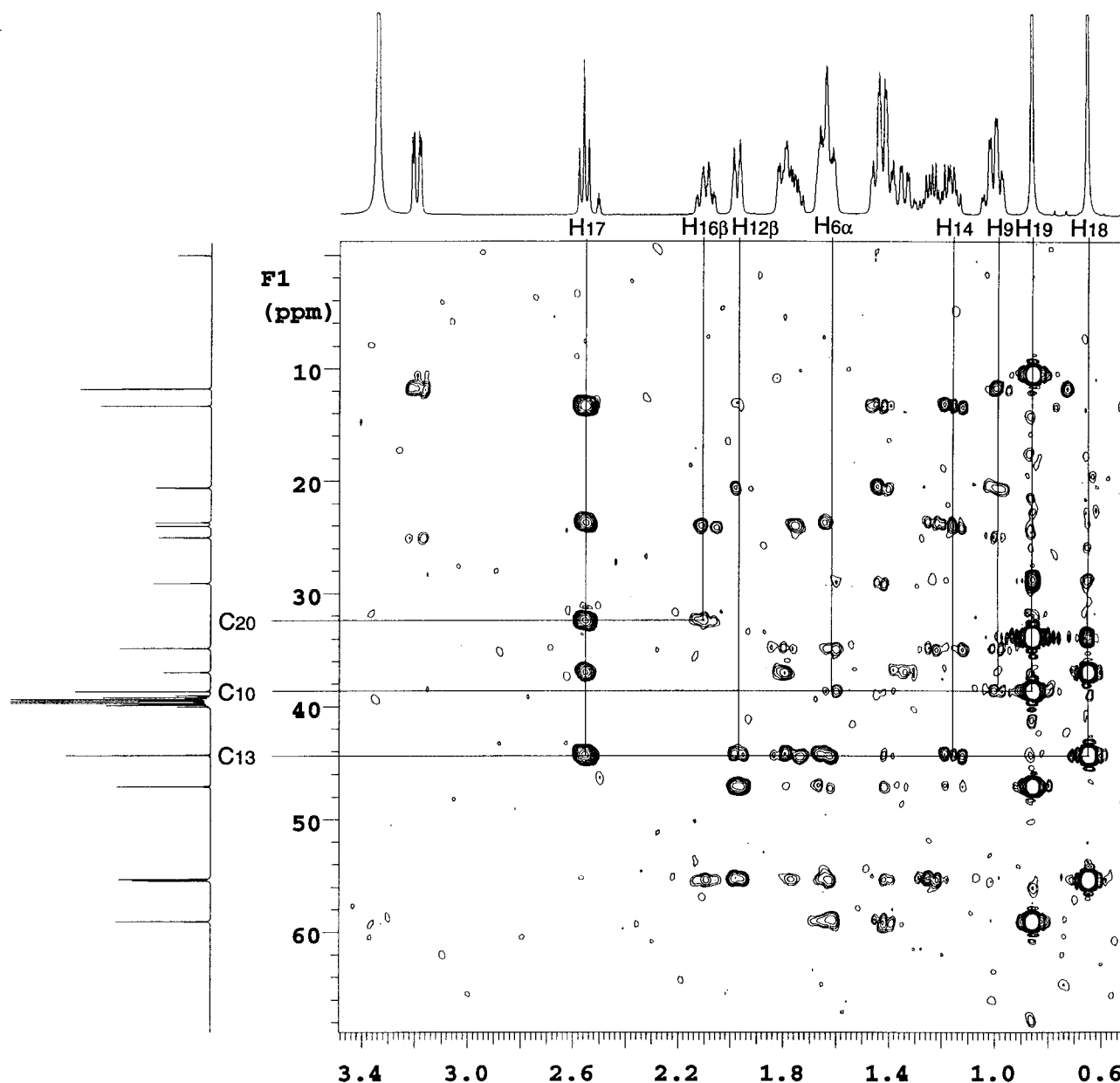


Fig. 9 Expansion of the HMBC spectrum of dutasteride in $\text{DMSO}-d_6$ showing multiple bond correlations used to assign quaternary carbon resonances. Highlighted are correlations used to assign C_{10} , C_{13} , and the carbonyl resonance C_{20} , which is folded over in the ^{13}C dimension. Other correlations in the spectrum serve to confirm resonance assignments made based on results from the other experiments. (See Fig. 10 for structure and numbering scheme.)

Recent years have seen a great deal of activity in the improvement of HMBC sequences including the accordion type experiments. These experiments are optimized for transfer through a wide range of heteronuclear couplings (e.g., 2–25 Hz) and can provide H–C correlations over two to four bonds. Accordion experiments have found great utility in the assignment of some complex natural products. The details and specific advantages of these experiments has been recently reviewed (17).

For small molecules of pharmaceutical interest, the foregoing array of NMR experiments yield most if not all of the information required to fully interpret the ^1H and ^{13}C NMR spectra and to completely characterize the structure of a molecule. The results of this type of analysis are summarized in Figs. 10 and 11, typical of what is often required for regulatory submissions.

Higher-Dimensional Experiments

There is no limit to the complexity of molecular structures and thus, it is clear that NMR spectra can become ever more complicated. In large molecules such as proteins, the second dimension is often not enough to provide the necessary resolution for facile interpretation. This has led to the development of three-dimensional (3D) experiments (18). In these experiments the resonances are spread out into a third dimension with the spectrum taking the form of a cube. Examples of a heteronuclear 3D experiment include HMQC–COSY and HMQC–NOESY in which the crosspeaks of a COSY or NOESY spectrum are further separated in the third dimension by their respective carbon chemical shifts (19). Four-dimensional (4D) experiments are also frequently used in the study of proteins in which correlations can be made between ^1H , ^{13}C , and ^{15}N in a single (triple-resonance) experiment. These spectra can be visualized as a cube in which each slice through the cube can be expanded into another cube. Protein studies often involve the incorporation of ^{13}C and ^{15}N isotopes into the protein, which greatly increases the sensitivity of the highly time consuming 3D and 4D experiments. An excellent description of higher dimensional experiments and their applications to protein structure elucidation is given in the review by Clore and Gronenborn (20).

Pulsed Field Gradients

One of the most important developments in NMR in recent years has been the application of pulsed field gradients (PFGs) (21). PFGs have long been used in magnetic resonance imaging methods, but only in the last decade has the technology been developed for routine use in high resolution NMR. PFGs consist of an extra magnetic field

applied across the sample in a spatially dependent manner. This imparts a phase shift to a given signal which is correlated to its position in the sample. PFGs have many utilities in high resolution NMR. They can be used to eliminate certain undesired signals by selectively dephasing them. Alternatively, signals can be selected using pairs of gradients in which the first gradient dephases all spins and a second gradient is optimized to refocus only selected spins. The judicious implementation of PFGs in pulse sequences has led to much cleaner spectra with fewer artifacts.

Another important use of PFGs is in the study of diffusion processes. A pair of PFGs of opposite sign separated by a delay time can be used to analyze the diffusion of molecules. If a molecule migrates significantly during the diffusion period, the second gradient will not effectively refocus those spins and the signals for that molecule will become attenuated. The use of PFG diffusion studies in high throughput screening studies will be discussed ahead.

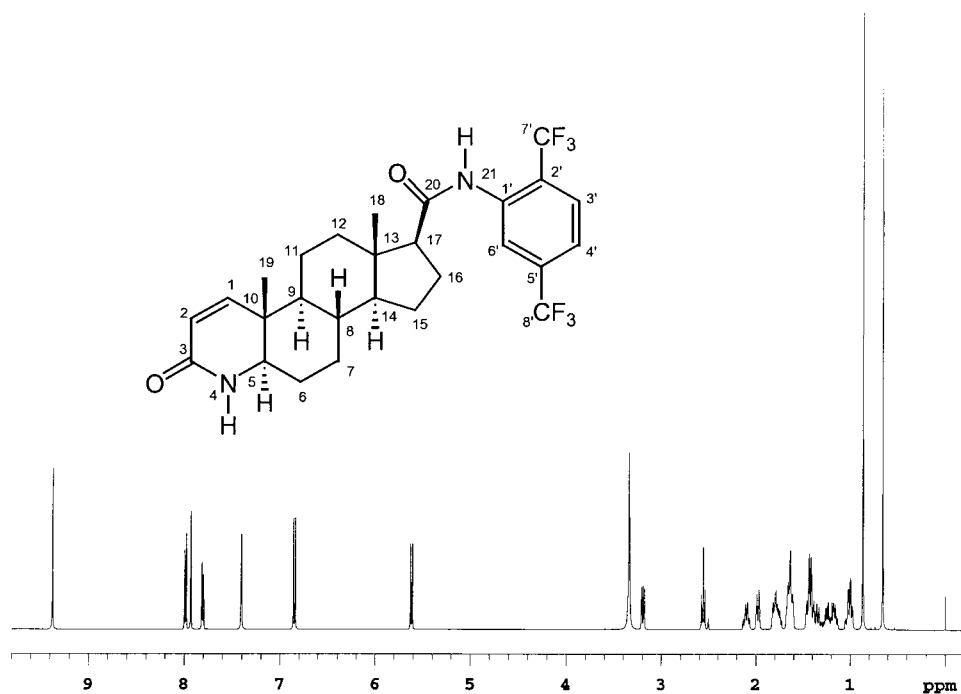
APPLICATIONS OF NMR

The NMR techniques described in the last section provide the foundation for many of the advanced applications of NMR in the pharmaceutical industry. The challenges of this industry have led to the optimization of hardware and experimental design to answer specific questions. Some of the most important questions and the NMR applications that have been developed to answer them will be described in the next sections.

Analysis of Small Sample Quantities

Small volume NMR probes

Samples such as metabolites, degradation products, and natural products may be very laborious to isolate and purify and are often available in only extremely small quantities. The sensitivity of NMR is considered low *vis a vis* other structural methods such as UV, IR, and MS, but the high information content has made increasing the limits of detection a very active and worthwhile pursuit. Sensitivity increases have traditionally been achieved by increasing the sample concentration or the magnetic field strength, but in recent years, optimization of the NMR probe has yielded significant gains. The NMR probe houses the hardware for the delivery of rf pulses and the detection of the NMR signal. Until recently, 5 mm sample tubes with a sample volume of 500–600 μl was the

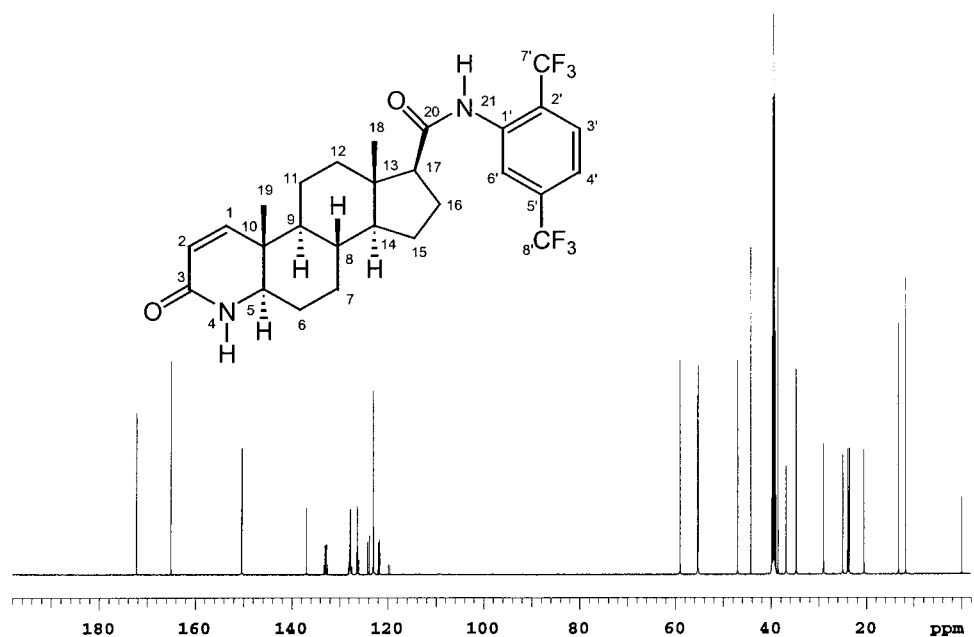


Chemical shift (δ) ^a	Multiplicity ^b	Integration	Assignment
9.38	s	1	21
7.99	d	1	3'
7.93	brs	1	6'
7.81	d	1	4'
7.40	d	1	4
6.85	d	1	1
5.62	dd	1	2
3.20	dd	1	5
2.56	t	1	17
2.10	m	1	16 β
1.98	m	1	12 β
1.80	m	1	11 α
1.76	m	1	16 α
1.65	m	2	7 β , 15 α
1.26	m	1	6 α
1.43	m	2	8, 12 α
1.42	m	1	6 β
1.34	m	1	11 β
1.24	m	1	15 β
1.17	m	1	14
1.01	m	1	7 α
0.99	m	1	9
0.87	s	3	19
0.66	s	3	18

^appm downfield from TMS =0.00 ppm: DMSO-*d*₅ and H₂O appear at 2.51 and 3.34 ppm, respectively.

^bs=singlet: d=doublet: t=triplet: q=quartet m=multiplet: br=broad.

Fig. 10 ¹H NMR spectrum and characterization of dutasteride in DMSO-*d*₆.



Chemical shift(δ) ^a	Multiplicity ^b	$J^{19}\text{F}-^{13}\text{C}(\text{Hz})$	Assignment
172.33	s	—	20
165.07	s	—	3
150.38	s	—	1
136.95	q	1.7	1'
132.89	q	32.6	5'
127.85	q	5.2	3'
127.83	q	29.6	2'
126.37	q	3.5	6'
123.09	q	273.1	8'
123.07	s	—	2
122.97	q	3.4	4'
122.77	q	273.9	7'
59.02	s	—	5
55.37	s	—	17
55.23	s	—	14
47.04	s	—	9
44.25	s	—	13
38.58	s	—	10
36.91	s	—	12
34.80	s	—	8
29.03	s	—	7
25.01	s	—	6
23.98	s	—	15
23.66	s	—	16
20.56	s	—	11
13.29	s	—	18
11.77	s	—	19

^appm downfield from TMS = 0.00 ppm: DMSO-*d*₆ appears at 39.47 ppm.

^bArising from ¹⁹F coupling: s = singlet; q = quartet.

Fig. 11 ¹³C NMR spectrum and characterization of dutasteride in DMSO-*d*₆.

standard for solution NMR studies. With very small sample amounts this results in a very dilute sample. Commercial NMR probes are now available for sample tubes as small as 1.7 mm with sample volumes down to 30 μ l. Efficient coupling of the NMR detector coil to physically smaller samples has resulted in a more than eightfold decrease in acquisition time over the standard 5-mm NMR sample tube configuration (22). The decrease in sample volumes has been taken even further by Sweedler and coworkers who have recently designed microcoil probes with a 200-nl sample volume (23). Reductions in sample volume provide an effective means to increase sensitivity, but further reductions may be less effective due to difficulties in routinely handling such small samples.

Low temperature probes

Increases in sensitivity can be also achieved by reducing the temperature of the detection coil. This was first demonstrated by Styles and coworkers who obtained substantial increases by operating the coil at cryogenic temperatures (24, 25). Recently an application of high resolution NMR on a natural product has been reported using a probe with rf coils made of high-temperature superconducting (HTS) materials operating at 25 K (26) where an enhancement in signal to noise of 3.5 was obtained over conventional probes. Currently, the HTS probes appear to have limited utility because the geometry of the coils cannot be optimized for standard cylindrical NMR tubes. Using the same idea of reduced temperature rf coil detection, cold probes are now being developed in which standard probe coil materials are cooled. This probe cooling combined with optimization for small sample volumes provides further sensitivity enhancements.

Flow NMR Methods

High-throughput NMR

Advances in parallel synthesis and combinatorial chemistry have given pharmaceutical companies the ability to generate an unprecedented number of compounds. In lead compound optimization efforts, libraries containing hundreds of compounds are often made for which structure validation is often necessary. For the synthetic chemist, high resolution NMR would be the tool of choice for this effort, but the standard method of preparing hundreds of individual NMR tubes for analysis would be prohibitively time consuming. This hurdle has been overcome by the development of flow NMR systems (27), which utilize a specialized flow cell in the probe so that samples can be injected into the NMR without the preparation of

individual sample tubes. HPLC type autosamplers connected to an injector system deliver the sample directly to the flow cell of the probe. In this way, samples in microtiter plates can be analyzed in an automated fashion. After the spectrum is taken, the flow can be reversed and the sample returned to its original location in the plate or diverted to waste. These probes have been optimized in terms of sensitivity to detect small sample quantities from micrograms down to hundreds of nanograms.

LC-NMR

Many of the most important chemical questions in the pharmaceutical industry involve the analysis of complex mixtures. Identification of low-level metabolites and drug substance impurities usually requires high-performance liquid chromatography for the separation of these mixtures or isolation of a compound of interest from a sample matrix. In these analyses, the structural information obtainable for the low-level compounds is limited by the type of detection used. The coupling of HPLC and mass spectrometry has become routine and provides useful molecular weight and fragmentation information, but this is often not enough for complete structure elucidation.

Recently, NMR spectrometers directly coupled with LC systems have become commercially available (28, 29). Spectra can be acquired in either of two modes, continuous or stopped flow. In continuous flow mode the spectrum is acquired as the analyte flows through the cell. This method suffers from low sensitivity since the analyte may be present in the cell for only a brief period of time, but it has the advantage of continuous monitoring of the LC peaks without interruption. Fig. 12a shows a contour plot of the continuous flow NMR analysis of a mixture of vitamin A acetate isomers (30). Fig. 12b shows the spectra taken from slices through the contour plot. These plots highlight the olefinic region of the spectra which provided ample information for the identification of each of the isomers. With very limited sample quantities, the more common method of LC-NMR analysis is stopped flow. Here the analyte peak is parked in the flow cell so any of the standard NMR experiments can be run.

A problem with the coupling of HPLC to NMR is that typical LC solvents give rise to very large background signals. The simplest way to avoid this is by using deuterated solvents, but given the relatively large volumes of solvent necessary, this can be very expensive. Recently developed solvent suppression methods have provided excellent means of reducing the signals from fully protonated solvents. These methods utilize a series of selective pulses along with pulsed field gradients and can simultaneously suppress several solvent signals in the spectrum (31). Another problem can arise when gradient elution methods

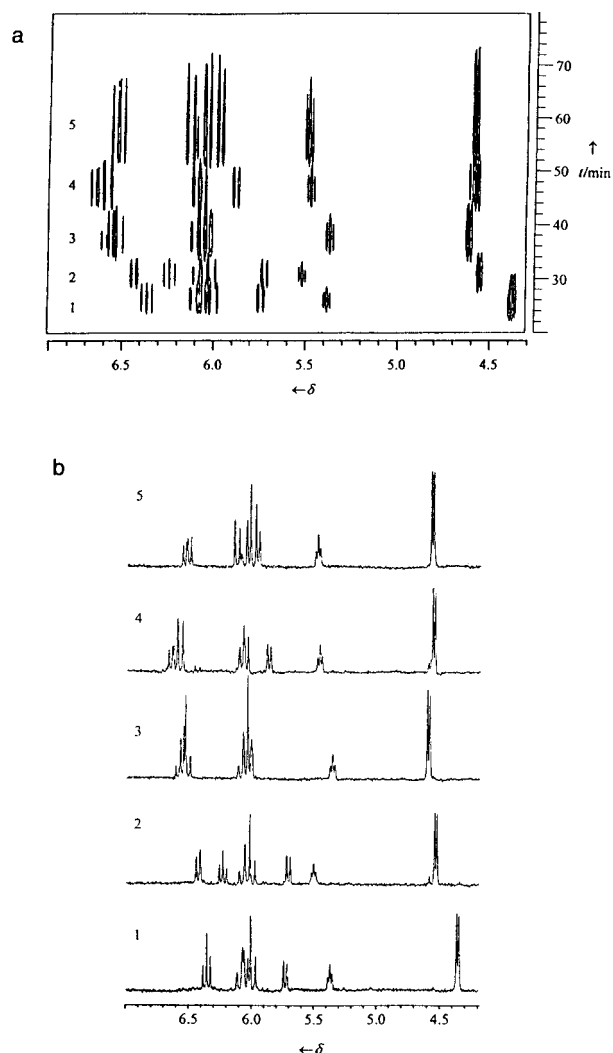


Fig. 12 a) Contour plot of the LC-NMR analysis of a mixture of vitamin A acetate isomers (400 MHz). b) Slices taken through the contour plot showing spectra of the individual components of the mixture. (From Ref. 30.)

are used in the LC separation because the chemical shifts of the solvent resonances can change as the gradient mixture changes. This is addressed by software that can dynamically modify the solvent suppression scheme to keep up with the chemical shift changes of the solvent resonances. Even with fully deuterated solvents for all mobile phase components solvent suppression methods are often still required for the residual protonated solvent signals.

LC-NMR-MS

The combination of LC with analytical methods has been taken a step further with in-line coupling of LC, NMR, and

MS systems to provide maximum information with minimal sample handling (32, 33). This setup yields both structural and molecular mass information from a single HPLC injection. The practical issues of running a mass spectrometer in close proximity to the magnetic field of the NMR provided somewhat of a hurdle to the technique, but this has been recently addressed by actively shielded magnets in which the stray magnetic fields are greatly attenuated by counter acting fields engineered into the magnet. The NMR and MS are typically setup in parallel with the flow split such that a small amount of the sample is sent to the MS and the rest to the NMR flow cell. This provides two sets of synchronized datasets, which facilitates interpretation. This also avoids potential issues of back pressure from the flow at the interface between the NMR and the MS.

NMR Based Screening Methods

Affinity NMR

One of the challenges that arises from the synthesis of combinatorial libraries is the screening of huge numbers of molecules for a desired biological activity. Mixtures of molecules are often screened in a single test which can lead to erroneous results. In some assays the sum of multiple small interactions can lead the screen to give a positive result when in fact no single molecule in the mixture is suitable for further study. A variety of experiments based on NMR have been developed recently that allow for the screening of complex mixtures with no need for physical separation and no risk of false positives from the additive effects of weak interactions. One of the methods for screening mixtures referred to as affinity NMR or diffusion edited NMR can be used to selectively observe only those ligands in a mixture that bind to a receptor molecule (34). The selection process is based on the changes in the diffusion rate of a ligand that occurs upon binding.

Affinity NMR experiments exploit the fact that the diffusion rate of molecules that bind to a receptor molecule will be greatly reduced. As described earlier, a PFG is applied to dephase all of the NMR signals. An equal but opposite PFG is applied after a delay time to refocus only those spins involved in binding to the receptor. Molecules not involved in binding will rapidly diffuse during the delay period between gradients, and their signals will be greatly attenuated. A qualitative ranking of ligand binding can be made based on the observed signal intensities. Although there are certain factors such as exchange phenomena that can complicate the interpretation of diffusion studies [see

Ref. (34)], the relative simplicity and nondestructive nature of diffusion based NMR screening make it a very valuable tool in drug discovery.

Transferred nOe screening

Another form of affinity NMR relies on the transferred nOe (35). As discussed earlier, small rapidly tumbling molecules display positive nOes and large slowly tumbling molecules display negative nOes. Upon binding to a large receptor, a small molecule assumes the correlation time of the receptor and large negative nOes develop. After the molecule dissociates from the receptor, the intensities of the signals remain perturbed by the large negative nOes, but the rapid small molecule tumbling rate yields narrow linewidths. The NOESY sequence can be used in these studies, and the negative nOe signals can be selectively displayed. The sensitivity of this method can be adjusted by varying the ratio of ligand to receptor. Typically, studies have observed the ligand in a 10–20-fold excess that allows these experiments to be carried out with very small amounts of receptor. This is a great advantage of transferred nOe based screening, especially with receptors that are difficult to produce and isolate.

Biomolecular NMR Studies

Protein structure studies

Due to their size and spectral complexity, proteins are among the most challenging molecules for NMR spectroscopy. As discussed earlier, a large number of sophisticated multidimensional, multinuclear NMR methods have been developed for the complete structural determination of proteins. The general procedure for protein structure determination can be summarized as four basic steps. The first step is sequential assignment of all amino acid resonances, using through-bond or through-space experiments. The second step is torsion angle determination using three-bond scalar coupling information. The third step is identification of through space interactions, using nOes. And the fourth step is calculation of the structure based on the structural restraints (nOes and torsion angles) using one or more computational refinement tools (20). This process requires large amounts of data to be acquired for a complete structure elucidation, which can be very time-consuming. Advances in molecular biology have made the incorporation of isotope labels in proteins using microbial systems relatively routine. With ^{15}N and ^{13}C labels the requisite 3D and 4D experiments can often be acquired in a matter of hours

rather than days. Currently it is possible to determine the structures of proteins in the 15–35 kDa range at a resolution comparable to ~ 2.5 Å resolution crystal structures.

Protein–ligand interactions

Solution state NMR experiments that probe the interactions between ligands and receptors have become an important part of the drug development process. When the structure of a protein has been completely determined by NMR, small changes in the chemical shifts of residues in the binding site will occur upon addition of an active ligand. These chemical shift changes are readily observed in $^1\text{H}\{^{15}\text{N}\}$ HSQC spectra acquired using ^{15}N labeled proteins. This method, known as chemical shift mapping, has been used to identify the binding sites of proteins, screen for active ligands, and design and optimize lead compounds.

High-throughput chemical shift mapping studies to screen for active ligands has recently been demonstrated using cryoprobe technology. The increased sensitivity of the cryoprobes enables protein $^1\text{H}\{^{15}\text{N}\}$ HSQC spectra to be acquired using very low protein concentrations. Since the total concentration of added molecules in these studies must be kept at a reasonable level (~ 5 – 10 mM), the use of lower protein concentrations allows for higher ligand concentrations. Hadjuk and coworkers analyzed mixtures of 100 compounds for the presence of tight binding molecules (36). The authors suggest that libraries of more than 200,000 compounds can be tested in less than one month using this strategy.

Structure–activity relationships (SAR) by NMR

A fundamental part of the drug design process is the development of structure–activity relationships. Recently an NMR method has been developed to produce SARs (37). In this method ligands are constructed from building blocks that have been optimized for binding to individual protein sub-sites. In the first step, a library of small organic molecules is screened for binding to a labeled protein. The binding event is detected by the observation of ^{15}N or ^1H chemical shift changes in the $^1\text{H}\{^{15}\text{N}\}$ HSQC spectra for resonances near the binding site. Once a lead molecule is identified, analogs are synthesized to optimize the binding. The process is then repeated to identify ligands that bind to a proximal binding site. Finally the separately optimized ligands are linked together to form a single high affinity ligand. This method was used to design a ligand to the immunosuppressant protein FK506 with an affinity of 19 nM.

SOLID-STATE NMR

A large proportion of pharmaceuticals end up in a solid formulation, so there is a clear need to characterize drugs in the solid state. Many different methods are available to study solid-state drug substances, including IR, Raman and X-ray diffraction, each with its own advantages and disadvantages (38). As in the solution state, solid-state NMR offers a potentially high level of structural information compared to other methods. In this section, a brief review of some of the fundamental differences between solid and liquid state NMR will be given along with some illustrative examples of the role of solid state NMR in pharmaceutical development.

Anisotropy in the Solid State

Simply placing a pellet of solid drug substance in an NMR tube and acquiring a ^1H or ^{13}C spectrum will yield very poor results. The resulting spectrum would typically appear as a collection of highly overlapping signals with linewidths in the thousands of Hertz. The basis for this extreme line broadening is the presence of anisotropic (orientation dependent) effects that are averaged out in solution, but become dominant in the solid state. The most important of these anisotropic effects are the dipole-dipole interactions. The local field, B_{loc} of any nucleus is influenced by the presence of neighboring dipoles. The effect on a given nucleus, I is described in Equation (8), where μ_s is the magnetic moment of a neighboring nucleus S, r_{IS} is the distance between nuclei I and S, and θ is the angle between the internuclear vector and the static B_0 field.

$$B_{\text{loc}} = \pm \mu_s r_{\text{IS}}^{-3} (3 \cos^2 \theta_{\text{IS}} - 1) \quad (8)$$

In solution, the rapid tumbling causes an averaging of the local field over all orientations that yields a net local field of zero. In the solid state, this averaging process does not occur, and the sum of all of the dipolar interactions yields linewidths on the order of kilohertz.

The other important contribution to broad overlapping lines in solid state NMR spectra is chemical shift anisotropy (CSA). As discussed earlier, the chemical shift is affected by electrons that shield the nucleus from the applied magnetic field. The magnitude of this shielding is a function of the particular orientation of the molecule in the magnetic field. Like the dipolar interactions, the effects of CSA averages out to zero in solution, but not in the solid state. The angular dependence of the CSA also takes the form $(3 \cos^2 \theta - 1)$

where θ is the angle between the B_0 field and the principle axis of the chemical shift (e.g., the C—O bond for a carbonyl carbon). The combined effects of dipolar coupling and CSA yield the tremendously broad lines observed in solid-state NMR.

Magic Angle Spinning and Cross-Polarization

The challenge of solid state NMR is to narrow the linewidths to an extent that the spectrum can be interpreted in a manner similar to a solution NMR spectrum. This challenge was met by noting that the $(3 \cos^2 \theta - 1)$ dependence of the dipolar interactions and CSA goes to zero when the angle θ is set to 54.7° , the magic angle. If a sample is spun about an axis oriented at the magic angle relative to B_0 , the average orientation of the crystal axes will be equal to the magic angle and the anisotropic effects will be greatly diminished. The very strong dipolar interactions between ^{13}C and the attached protons may still yield broadened lines, therefore ^1H decoupling is often used in addition to magic angle spinning to achieve narrow lines in ^{13}C spectra.

The most widely studied nucleus in solid state NMR of pharmaceuticals is ^{13}C . As in solution, observation of ^{13}C suffers from inherently low sensitivity. Additionally, ^{13}C nuclei in the solid state have very long relaxation times, which limits the number of scans that can be acquired in a given amount of time. To enhance the sensitivity and allow for faster repetition rate, cross-polarization (CP) methods have been developed. CP is very analogous to spin polarization transfer. A train of simultaneous rf pulses on both the ^1H and ^{13}C is used to transfer the ^1H magnetization to the ^{13}C . In this way, the intensity of the ^{13}C line is greatly increased and the pulse repetition rate is dependent upon the ^1H and not the ^{13}C relaxation. Cross-polarization magic angle spinning (CP-MAS) experiments constitute the majority of solid state NMR studies in the pharmaceutical industry and provide the foundation for many of the more advanced techniques.

Applications of Solid State NMR

Gel-phase NMR

The synthesis of combinatorial libraries relies heavily on solid phase synthetic methods. Previously the analysis of these libraries, using high throughput flow NMR methods was discussed. These analyses relied on the cleavage of the product from the solid phase support. In some cases, the cleavage reaction can be very harsh, and it is often desirable to analyze the solid phase reaction product with the compound still attached to the bead. Given the macromolecular

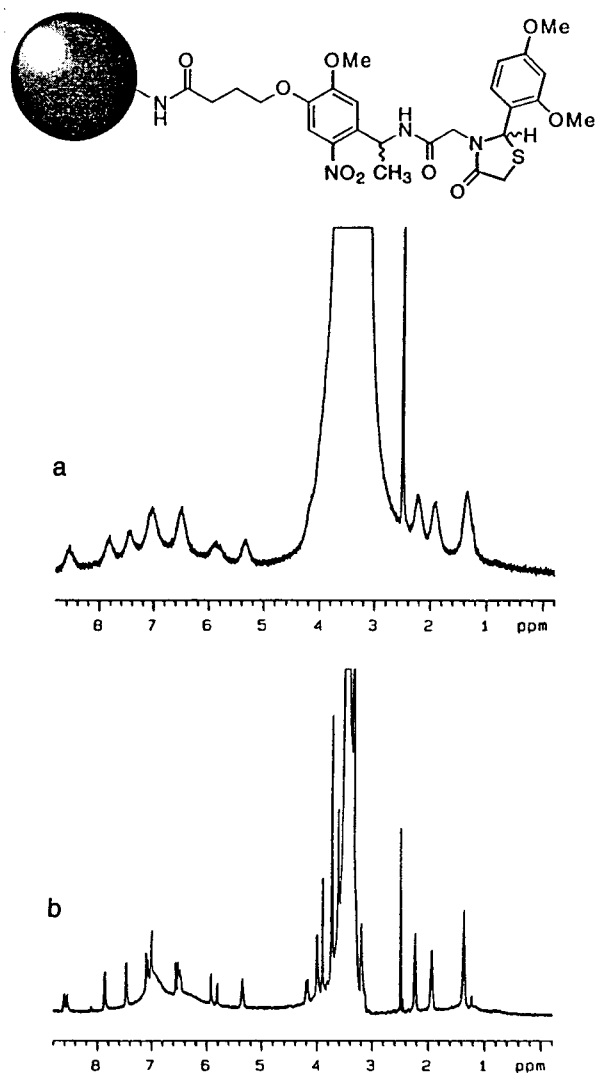


Fig. 13 a) ¹H spectrum of the “gel-phase” compound shown taken using a conventional 5-mm liquids probe. (100 mg sample suspended in 600 μ l DMSO-*d*₆, 500 MHz) b) ¹H spectrum of the same sample taken using the Varian Nano-probe (Varian NMR Instruments Palo Alto, CA, USA) probe with magic angle spinning (10-mg sample suspended in 30 μ l of DMSO-*d*₆, 500 MHz). (From Ref. 40.)

size of the bead, simply placing a sample of the solid phase material into an NMR tube would give spectra with broad overlapping lines. Magic angle spinning has been applied in order to narrow the lines of solid phase combinatorial samples (39). As the solid phase supports are swollen by solvents and therefore have properties that are neither distinctly solution phase nor solid phase, the term “gel-phase” NMR is used. This technique was used by Fitch and coworkers to generate 500 MHz spectra of organic

compounds bound to SPS resins with linewidths as narrow as 4 Hz (40). Fig. 13 shows the spectra of a “gel-phase” sample taken with conventional NMR methods and with magic angle spinning.

Polymorph studies

It has been estimated that approximately 30% of organic compounds crystallize in two or more forms that differ in the conformation and/or arrangement of molecules in the crystal lattice (41, 42). The crystalline form of a drug can have profound effects on the physicochemical properties. The density, melting point, dissolution rate, bioavailability, and ease of formulation are all influenced to some degree by the crystal form. Polymorphism is defined as the ability of a compound to exist in more than one solid-state form with identical chemical structures, but different crystal lattices. Solvates are an important form of polymorphism in which the crystal lattice changes by the inclusion of solvent. Given the influence of polymorphic forms on physicochemical properties, characterization of all polymorphic forms present in a drug substance is becoming a requirement of the relevant regulatory agencies.

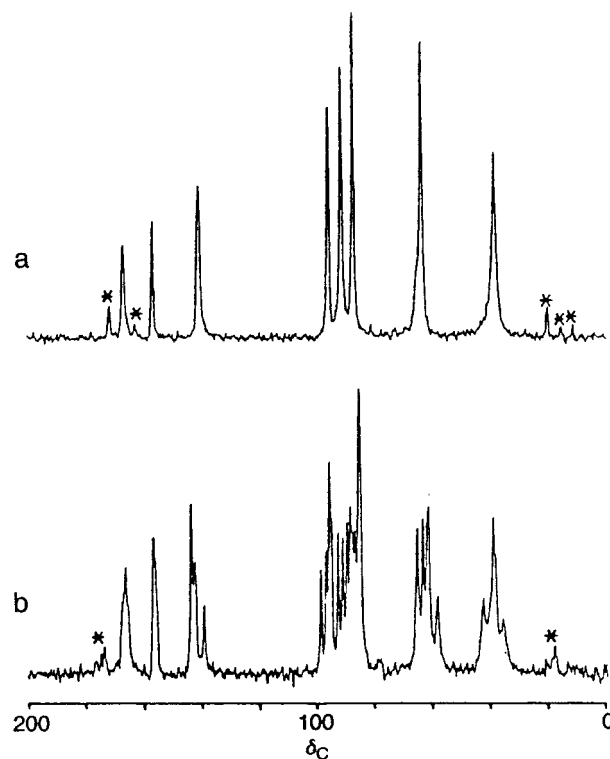


Fig. 14 Solid-state ¹³C CP-MAS spectra of two polymorphic forms of lamivudine recorded at 125 MHz a) Form II, b) Form I. Signals indicated by asterisks are spinning sidebands. (From Ref. 43.)

Solid-state NMR has emerged as a powerful tool in the analysis of polymorphic drug forms. ^{13}C CP-MAS spectroscopy can be used to identify the number of crystallographically inequivalent sites in a unit cell and to understand the molecular structure on the basis of the chemical shifts. Fig. 14 shows the solid-state ^{13}C spectra of two forms of lamivudine (43). Form II shows a relatively simple spectrum in which there is only one molecule in the crystallographic asymmetric unit. The spectrum of Form I is much more complex, especially in the region from 80 to 100 ppm. Analysis of this spectrum indicates that there are five distinct molecules in the unit cell. This rare phenomenon was subsequently confirmed by X-ray diffraction.

Solid-state structural studies

Standard 1D ^{13}C CP-MAS spectra provide highly valuable data in terms of characterizing a single polymorph or determining the number of polymorphs in a sample, but high resolution structural data on solid materials requires more sophisticated methods. These methods include CP-MAS variants of some of the 2D methods described earlier that can yield through-bond and through-space connectivities (44) and typically require the use of selective or uniform isotopic labeling. A detailed study of the polymorphic forms of cimetidine was carried out using selective ^{13}C labeling. The distance constraints obtained in this study were used in molecular modeling studies to determine the possible conformations of cimetidine present in the different polymorphs (45).

FUTURE DIRECTIONS—SMALLER, FASTER, EASIER

The challenge of NMR in the pharmaceutical industry is to provide the intrinsically high information content of NMR spectra on smaller sample amounts, more rapidly, and with less time-intensive interpretation. Currently, each of these areas are the focus of intensive research efforts. Sensitivity gains can be expected with further innovations in probe design and with higher magnetic fields. Advances in experimental design will continue to provide more detailed information for both the qualitative and quantitative analyses of pharmaceutical compounds. The rapid analysis and interpretation of NMR data is being addressed with advances in computational methods (46). Improvements in spectral prediction and automated structure elucidation will greatly reduce the interpretation bottleneck in NMR applications. These advances, along with those yet

uncovered, will insure that NMR assumes an even greater role in essentially all aspects of pharmaceutical research and development in the future.

REFERENCES

1. Friebolin, H. *Basic One- and Two-Dimensional NMR Spectroscopy*; VCH Publishers: New York, 1993.
2. Karplus, M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **1963**, *85*, 2870–2871.
3. Neuhaus, D.; Williamson, M. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; VCH Publishers: Cambridge, UK, 1989.
4. Bothner-By, A.A.; Stephens, R.L.; Lee, J.-M. Structure Determination of a Tetrasaccharide: Transient Nuclear Overhauser Effects in the Rotating Frame. *J. Am. Chem. Soc.* **1984**, *106*, 811–813.
5. Bax, A.; Davis, D.G. *J. Magn. Reson.* **1985**, *63*, 760–762.
6. Morris, G.A.; Freeman, R. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *J. Am. Chem. Soc.* **1979**, *101*, 760–762.
7. Doddrell, D.M.; Pegg, D.T.; Bendall, M.R. Distortionless Enhancement of NMR Signals by Polarization Transfer. *J. Magn. Reson.* **1982**, *48*, 323–327.
8. Kirk, D.N.; Toms, H.C.; Douglas, C.; White, K.A.; Smith, K.E.; Latif, S.; Hubbard, R.W.P. A Survey of the High Field ^1H NMR Spectra of the Steroid Hormones, Their Hydroxylated Derivatives and Related Compounds. *J. Chem. Soc., Perkin Trans. 2* **1990**, 1567–1594.
9. Piatini, U.; Sørensen, O.W.; Ernst, R.R. Multiple Quantum Filters for Elucidating NMR Coupling Networks. *J. Am. Chem. Soc.* **1982**, *104*, 6800–6801.
10. Davis, D.G.; Bax, A. Assignment of Complex ^1H NMR Spectra Via Two-Dimensional Homonuclear Hartmann-Hahn Spectroscopy. *J. Am. Chem. Soc.* **1985**, *107*, 2821–2822.
11. Bax, A.; Griffey, R.H.; Hawkins, B.L. Sensitivity-Enhanced Correlation of ^{15}N and ^1H Chemical Shifts in Natural-Abundance Samples Via Multiple Quantum Coherence. *J. Am. Chem. Soc.* **1983**, *105*, 7188–7190.
12. Martin, G.E.; Hadden, C.E. Long-Range ^1H – ^{15}N Heteronuclear Shift Correlation at Natural Abundance. *J. Nat. Prod.* **2000**, *63* (4), 543–585.
13. Bodenhausen, G.; Ruben, D. Nature Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy. *J. Chem. Phys. Lett.* **1980**, *69*, 185–189.
14. Brüchwiler, D.; Wagner, G. Selective Excitation of ^1H Resonances Coupled to ^{13}C . Hetero COSY And RELAY Experiments with ^1H Detection for a Protein. *J. Magn. Reson.* **1986**, *69*, 546–551.
15. Reynolds, W.F.; McLean, S.; Tay, L.-L.; Yu, M.; Enriquez, R.G.; Estwick, D.M.; Pascoe, K.O. Comparison of ^{13}C Resolution and Sensitivity of HSQC and HMQC Sequences and Application of HSQC-Based Sequences to the Total ^1H and ^{13}C Spectral Assignment of Clonasterol. *Magn. Reson. Chem.* **1997**, *35*, 45–462.
16. Summers, M.F.; Marzilli, L.G.; Bax, A. Complete ^1H and ^{13}C Assignments of Coenzyme B₁₂ Through the Use of

- New Two-Dimensional NMR Experiments. *J. Am. Chem. Soc.* **1986**, *108*, 4285–4294.
17. Martin, G.E.; Hadden, C.E.; Crouch, R.C.; Krishnamurthy, V.V. ACCORD-HMBC: Advantages and Disadvantages of Static Versus Accordion Excitation. *Magn. Reson. Chem.* **1999**, *37*, 517–528.
 18. Griesinger, C.; Sørensen, O.W.; Ernst, R.R. A Practical Approach to Three-Dimensional NMR Spectroscopy. *J. Magn. Reson.* **1987**, *73*, 574–579.
 19. Fesik, S.W.; Zuiderweg, E.R.P. Heteronuclear Three-Dimensional NMR Spectroscopy. A Strategy for the Simplification of Homonuclear Two-Dimensional NMR Spectra. *J. Magn. Reson.* **1988**, *78*, 588–593.
 20. Clore, G.M.; Gronenborn, A.M. NMR Structures of Proteins and Protein Complexes Beyond 20,000 Mr. *Nat. Struct. Biol.* **1997**, 849–853.
 21. Keeler, J.; Clowes, R.T.; Davis, A.L.; Laue, E.D. Pulsed Field Gradients: Theory and Practice. *Methods Enzymol.* **1994**, *239*, 145–207.
 22. Martin, G.E.; Hadden, C.E. Comparison of 1.7 mm Submicro and 3 mm Micro Gradient NMR Probes for the Acquisition of ^1H – ^{13}C and ^1H – ^{15}N Heteronuclear Shift Correlation Data. *Magn. Reson. Chem.* **1999**, *37*, 721–729.
 23. Olson, D.L.; Lacy, M.E.; Sweedler, J.V. High Resolution Microcoil NMR for Analysis of Mass-Limited, Nanoliter Samples. *Anal. Chem.* **1998**, *70*, 645–650.
 24. Styles, P.; Soffe, N.F.; Scott, C.A. An Improved Cryogenically Cooled Probe for High-Resolution NMR. *J. Magn. Reson.* **1989**, *84*, 376–378.
 25. Styles, P.; Soffe, N.F.; Scott, C.A.; Gragg, D.A.; Row, F.; White, D.J.; White, P.C.J. A High-Resolution NMR Probe in which the Coil and Preamplifier are Cooled with Liquid Helium. *J. Magn. Reson.* **1983**, *60*, 397–404.
 26. Logan, T.M.; Murali, N.; Wang, G.; Jolivet, C. Application of a High Resolution Superconducting NMR Probe in Natural Product Structure Determination. *Magn. Reson. Chem.* **1999**, *37*, 512–515.
 27. Keifer, P.A.; Smallcombe, S.H.; Williams, E.H.; Salomon, K.E.; Mendez, G.; Belletire, J.L.; Moore, C.D. Direct-Injection NMR (DI-NMR): A Flow NMR Technique for the Analysis of Combinatorial Libraries. *J. Comb. Chem.* **2000**, *2*, 151–171.
 28. Albert, K. On-Line Use of NMR Detection in Separation Chemistry. *J. Chromatogr. A* **1995**, *703*, 123–147.
 29. Lindon, J.C.; Nicholson, J.K.; Sidelmann, U.G.; Wilson, I.D. Directly Coupled HPLC-NMR and its Application to Drug Metabolism. *Drug Metab. Rev.* **1997**, *29*, 705–746.
 30. Albert, K.; Schlotterbeck, G.; Braumann, U.; Händel, H.; Spraul, M.; Krack, G. Structure Determination of Vitamin A Acetate Isomers through Coupled HPLC and ^1H NMR Spectroscopy. *Angew. Chem., Int. Ed.* **1995**, *34* (9), 1014–1016.
 31. Smallcombe, S.H.; Patt, S.L.; Keifer, P.A. WET Solvent Suppression and its Applications to LC-NMR and High-Resolution NMR Spectroscopy. *J. Magn. Reson., Ser. A* **1995**, *117*, 295–303.
 32. Pullen, F.S.; Swanson, A.G.; Newman, M.J.; Richards, D.S. Online Liquid Chromatography/Nuclear Magnetic Resonance/Mass Spectrometry—A Powerful Spectroscopic Tool for the Analysis of Mixtures of Pharmaceutical Interest. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1003–1006.
 33. Holt, R.M.; Newman, M.J.; Pullen, F.S.; Richards, D.S.; Swanson, A.G. High Performance Liquid Chromatography/NMR Spectrometry/Mass Spectrometry: Further Advances in Hyphenated Technology. *J. Mass Spectrom.* **1997**, *32*, 64–70.
 34. Chen, A.; Shapiro, M.J. Affinity NMR. *Anal. Chem.* **1999**, *669A*, 675A.
 35. Henrichsen, D.; Ernst, B.; Magnani, J.L.; Wang, W.-T.; Meyer, B.; Peters, T. Bioaffinity NMR Spectroscopy: Identification of an E-Selectin Antagonist in a Substance Mixture by Transfer NOE. *Angew. Chem., Int. Ed.* **1999**, *38*, 98–102.
 36. Hajduk, P.J.; Gerfin, T.; Boehlen, J.-M.; Häberli, M.; Marek, D.; Fesik, S.W. High-Throughput Nuclear Magnetic Resonance Based Screening. *J. Med. Chem.* **1999**, *42*, 2315–2317.
 37. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.
 38. Britton, H.G. *Physical Characterization of Pharmaceutical Solids*; Marcel Dekker, Inc.: New York, 1995.
 39. Keifer, P.A. High-Resolution NMR Techniques for Solid-Phase Synthesis and Combinatorial Chemistry. *Drug Discovery Today* **1997**, *11*, 468–478.
 40. Fitch, W.L.; Detre, G.; Holmes, C.P.; Shoolery, J.N.; Keifer, P.A. High-Resolution ^1H NMR In Solid-Phase Organic Synthesis. *J. Org. Chem.* **1994**, *59* (26), 7955–7956.
 41. Kuhnert-Brandstatter, M.; Riedmann, M. Thermal Analysis and Infrared Spectroscopy Investigation on Polymorphic Organic Compounds-I. *Mikrochim. Acta* **1987**, *2*, 107–120.
 42. Haleblan, J.; McCrone, W. Pharmaceutical Applications of Polymorphism. *J. Pharm. Sci.* **1969**, *58* (8), 911–929.
 43. Harris, R.K.; Yeung, R.R.; Lamont, R.B.; Lancaster, R.W.; Lynn, S.M.; Staniforth, S.E. 'Polymorphism' in a Novel Anti-Viral Agent: Lamivudine. *J. Chem. Soc., Perkin Trans. 2* **1997**, 2653–2659.
 44. Blümich, B.; Spiess, H.W. Two-Dimensional Solid State NMR Spectroscopy: New Possibilities for the Investigation of the Structure and Dynamics of Solid Polymers. *Angew. Chem. Int. Ed.* **1988**, *27* (12), 1655–1672.
 45. Middleton, D.A.; Le Duff, C.S.; Peng, X.; Reid, D.G.; Saunders, D. Molecular Conformations of the Polymorphic Forms of Cimetidine from ^{13}C Solid State NMR Distance and Angle Measurements. *J. Am. Chem. Soc.* **2000**, *122*, 1161–1170.
 46. Williams, A. Recent Advances in NMR Prediction and Automated Structure Elucidation Software. *Curr. Opin. Drug Discovery Dev.* **2000**, *3* (3), 298–305.



Nonprescription Drugs

Sujit S. Sansgiry

Lynn Simpson

Gauri Shringarpure

Amit Kulkarni

University of Houston, Houston, Texas, U.S.A.

INTRODUCTION

In the 1800s, manufacturers used extravagant promotional techniques, which were reflective of the nation's level of culture and taste, to sell their products. Exaggerated statements regarding the products and its effects were very common. Professional standards and ethics were at an all-time low and virtually nonexistent in most cases. Patients, who had to make a choice between a myriad of medication products, were usually confused and ignorant of the medication use process. It was not surprising that most of these patients turned to home remedies for minor illnesses and self-care during such episodes. Today, the self-care movement has progressed forward by leaps and bounds with consumers having autonomy in making most of their health care decisions with the vast amount of information and products currently available to them.

Although the selection and number of products sold from country to country may vary, the group of drugs or medications sold throughout the world are easily distinguishable into two general categories, namely, those obtained with a doctor's prescription and those that can be obtained without one. The latter collection of products are normally classified as nonprescription drugs or over-the-counter (OTC) medications and are discussed in this article. Both nonprescription and OTC medications are used synonymously throughout this article.

SELF-CARE—SELF-MEDICATION

When consumers purchase and use nonprescription medications they are involved in a process that is popularly known as self-medication. Hence, nonprescription medications are products that are lawfully available for self-medication, without prior approval from a healthcare professional. Nonprescription medications are sold primarily for the prevention, treatment, and symptomatic relief of certain acute and chronic diseases. These products are not used for treating serious and complex conditions that necessitate the attention and direction of a licensed health practitioner. However, the main presump-

tion of providing nonprescription medications is the easy access available to consumers when the need arises.

Currently, nonprescription medications play a significant role in America's healthcare system by providing effortless access to more than 100,000 OTC drug products marketed using over 800 significant active ingredients. There are various categories of OTC medications available to consumers for treating minor symptoms of headache to problems of excessive acid secretion in the stomach. A list of symptoms or conditions treated with OTC products can be seen in Fig. 1. The top 20 selling nonprescription medications in the United States during the year 2001^[1] can be seen in Table 1. These top selling OTCs correlate with the medical ailments, which consumers report being most commonly self-treated (Table 2).^[2]

Nonprescription drugs are potent chemical entities and just like prescription drugs, should be viewed relative to their pharmacology, toxicology, contraindications, use, adverse effects, drug interaction potential, effectiveness, and special considerations in dosage and administration. It is significant to note that many nonprescription medications were, at one time, available as prescription drugs before being converted to OTC status. Therefore, it is important that the same precautions be taken when promoting and using nonprescription medications as their prescriptive counterparts.

SELF-CARE/SELF-MEDICATION MOVEMENT

The history of self-medication movement began in the late 1970s or early 1980s. Within this time period, there has been a huge movement towards self-help or self-care manifested by a positive approach to wellness and a determination to be involved in the decision making process regarding illness and its treatment. Availability of advanced technology such as computers and the Internet have allowed consumers to have access to a wealth of medical knowledge at their fingertips. Consumers in the 21st century are taking charge of their own personal health and evaluating the need for medical intervention.^[3] In addition, they are managing their symptoms and physical

Abrasions	Gastritis
Acne	Gingivitis
Aches and pains (general, mild to moderate)	Hair loss
Albumin testing	Halitosis
Allergic reaction	Head lice
Allergic rhinitis	Headache
Anemia	Heartburn
Arthralgia	Haemorrhoids
Asthma	Impetigo
Athlete's foot	Indigestion
Bacterial infection (topical, superficial, uncomplicated)	Ingrown toenails
Blisters	Insect bites and stings
Blood pressure monitoring	Insomnia
Boils	Jet lag
Bowel preparation (diagnostic)	Jock itch
Burns (minor, thermal)	Motion sickness
Calluses	Myalgia
Candidal vaginitis	Nausea
Canker sores	Nutrition (infant)
Carbuncles	Obesity
Chapped skin	Occult blood in feces (detection)
Cold sores	Ostomy care
Colds (viral upper respiratory infection)	Ovulation prediction
Congestion (chest, nasal)	Periodontal disease
Conjunctivitis	Pharyngitis
Constipation	Pinworm infestation
Contact lens care	Premenstrual syndrome Pregnancy (diagnostic)
Contraception	Prickly heat
Corns	Psoriasis
Cough	Ringworm
Cuts (superficial)	Seborrhea
Dandruff	Sinusitis
Deficiency disorders (mineral, vitamin, enteral food supplements)	Smoking cessation
Dental care	Sprains
Dermatitis (contact)	Strains
Diabetes mellitus (insulin, monitoring equipment, supplies)	Stye (hordeolum)
Diaper rash	Sunburn
Diarrhea	Teething
Dry skin	Thrush
Dysmenorrhea	Toothache
Dyspepsia	Vomiting
Feminine hygiene	Warts (common and plantar)
Fever	Xerostomia
Flatulence	Wound care

Fig. 1 List of selected medical disorders amenable to nonprescription drug therapy.

well-being independently, using diet, exercise, OTC medications, herbal supplements, or other forms of complementary care. These acts define the Self-Care movement, when the consumers act as an advocate for their own well-being using preventative medicine and algorithms of care.^[2] Self-care is what an individual does for himself to establish and maintain health and to prevent and deal with illnesses, whereas self-medication is the selection and use of medicines by individuals to treat self-recognized symptoms and illnesses. Self-medication is only a part of self-care, which is a broadly defined term.^[4] Numerous research studies, utilizing survey techniques, have shown a definite trend towards the use of self-care behaviors among

American consumers and that consumers are becoming increasingly comfortable with self-medication. According to a survey done by the Consumers Healthcare Products Association (CHPA), previously known as Nonprescription Drug Manufacturers Association (NDMA), 59% of respondents say they are more likely to treat their own health conditions today than they were in the past.^[5] Additionally, 77% of consumers would take an OTC medication when they are sick with a common everyday ailment. This is approximately twice as many as those who would consult a physician or take a prescription medication. This illustrates the fact that informed, appropriate, and responsible use of nonprescription drugs is an important



Nonprescription Drugs

Table 1 Top 25 most popular OTC drugs

Rank	Product name
1	Private label internal analgesic tablets
2	Tylenol internal analgesic tablets
3	Private label cold/allergy/sinus tablets/packets
4	Advil internal analgesic tablets
5	Nicorette antismoking gum
6	Aleve internal analgesic tablets
7	Private label first-aid ointments/antiseptics
8	Benadryl cold/allergy/sinus tablets/packets
9	Nicoderm CQ antismoking patch
10	Private label antacid tablets
11	Bayer internal analgesic tablets
12	Tylenol PM internal analgesic tablets
13	Private label cold/allergy/sinus liquid/powder
14	Pepcid AC antacid tablets
15	Private label laxative tablets
16	Halls cough/sore throat drop
17	Vicks Nyquil cold/allergy/sinus liquid/powder
18	Zantac 75 antacid tablets
19	Excedrin internal analgesic tablets
20	Motrin IB internal analgesic tablets

(From Ref. [1].)

part of the self-care process. Problems with self-medication may arise when the consumers take OTC medications along with other prescription drugs, which may lead to overdose of a drug. A study was done in which consumers were given a personal education program and it has been shown that this program proved to be more effective in changing the consumer's self-medication behavior.^[6] It has been suggested that OTC medication use is a common self-care behavior and therefore the patient's OTC medication use data should be collected while recording medication history.^[7,8]

Table 2 Consumer use of OTC medications by ailment

OTC category	Use (%)
Headaches	81
Cold/cough/flu/sore throat	72
Skin problems	68
Heartburn/indigestion	66
Allergy/sinus	58
Premenstrual	53
Constipation/diarrhea	49
Muscle/joint/back pain	49
Upset stomach/nausea	46
Minor eye problems	46
Teeth/gum problems	30
Menopausal symptoms	19

(From Ref. [2].)

Why Do Consumers Use OTCs?

Access to self-medication options empowers consumers and enhances their desire to take control of their own health conditions. Some important reasons why consumers use nonprescription medications include convenience, cost, time, and their attitudes, values, and beliefs.^[5] Consumers turn to nonprescription medications for 38% of all health problems they experience.^[9] It is known that consumers in the United States use more nonprescription medications than anywhere else in the world, with the sale of these products exceeding \$30 billion dollars annually.^[10] Sales of OTCs have increased steadily over the years. Table 3 provides a 5-yr trend analysis of OTC sales by category.^[11] Yet for this vast volume, OTC medications still take up less than 2 cents of every healthcare dollar spent by an individual.^[12,13] In addition to the clinical and economic benefits of using nonprescription products, 92% of consumers report they are satisfied with the medications they have used for self-care.^[14]

These benefits not only accrue to the individual, but to society as well. The resource savings to the healthcare system through responsible self-medication allows better allocation of limited healthcare resources and physicians' time to important issues beyond the scope of self-care. It is inconceivable to consider the tremendous economic strain it could impose on the healthcare system if every consumer chose to visit a physician for every minor illness. Health centers would be overwhelmed and would not be able to work efficiently as illustrated by the fact that on an average each American experiences one potential self-treatable health problem every 3–4 days.^[5] Additionally, approximately 90% of the Americans consider themselves in poor health at least one or more times each month.^[5] Approximately 70% of consumers self-medicate on a regular basis, and an estimated 40% of the U.S. population uses at least one nonprescription drug within any given 48-hr period.^[5] Hence, easy access to nonprescription products is likely to be a priority for the American consumer.

OTC AVAILABILITY AND ABUSE

Because of these prevailing sentiments, unlike in other countries, nonprescription medications are available virtually everywhere in the United States. Even vending machine laws have been implemented that sanction the use of mechanical devices and vending machines for the sale of nonprescription drugs. These products are also widely available on the Internet and may have been the cause of various abuse cases involving teenagers and OTC cough and cold products. Recent statistics from the American Association of Poison Control Centers 2001 report that

**Table 3** OTC sales trends by major category

OTC category	1996 (\$)	1997 (\$)	1998 (\$)	1999 (\$)	2000 (\$)
Acne remedies	270	274	286	307	352
Analgesic and chest rubs	271	268	263	275	308
Antacids	1,883	1,220	1,224	1,232	1,189
Antidiarrheals	190	196	209	219	222
Antigas products	134	141	153	164	178
Antiperspirants	1,550	1,598	1,658	1,741	1,767
Antiseptics	121	125	128	135	136
Antisleep products	44	43	43	44	43
Antismoking products	286	562	564	623	664
Appetite suppressants	92	89	88	77	88
Baby care products	71	74	77	83	93
Cough/cold remedies	3,135	3,279	3,260	3,578	3,514
Ear drops	33	33	35	38	39
Eye care	510	528	545	556	522
Feminine hygiene	521	527	527	542	547
First aid	437	472	520	538	552
Foot preparations	357	396	443	488	504
Hair growth products	137	152	151	127	121
Jock itch	26	28	32	37	42
Laxatives	744	772	826	870	915
Lip remedies	217	227	230	269	301
Motion sickness prevention	34	35	36	40	43
Oral care	2,360	2,402	2,509	2,625	2,678
Pain remedies	2,889	2,985	3,027	3,207	3,185
Petroleum jelly	57	58	57	61	61
Rectal medications	177	171	176	178	185
Skin bleaching	26	26	26	27	39
Sleeping aids	98	99	99	112	124
Sunscreens/sunburn	281	321	332	375	385
Others	252	263	277	299	308
Total	15,390	16,504	17,364	18,867	19,105

(From Ref. [11].)

the number of cases involving abuse or self-harm from nonprescription cough and cold products increased to 13,393 in 2000 from 9,889 in 1999.^[15] Combination products containing ingredients like dextromethorphan and chlorpheniramine are being used to experience the pleasures of drug abuse such as those experienced by LSD.^[14] Unfortunately, the danger of ingesting high amounts of these products can cause deleterious effects on the cardiac system and the central nervous system. Hallucinations, delirium, high body temperatures, and dizziness can occur when these ingredients are abused. Additionally, many of these products also contain acetaminophen (APAP) that can cause liver toxicity when used in high doses. Pharmacists employ several methods to limit patient access to products of abuse. The most common technique is to keep the product out of sight.^[16] Other techniques include additional client questioning, providing advice, and limiting the quantity of product sold.^[16] Pharmacists need to be aware of the

misuse of these products and counsel consumers on the proper use and dangers of taking excessive doses.

Unlike the United States, many countries further divide nonprescription medications into different subclasses. Table 8 contains information regarding OTC products around the world. For example, in the United Kingdom, nonprescription medications known as P medicines may only be sold in registered pharmacies under the supervision of a registered pharmacist while other nonprescription medications, known as the general sale list medicines, may be sold at all other retail outlets. There has been a push in recent years to consider a third class of medications sold only under the supervision of the pharmacist in the United States.^[17] Many consumers and healthcare providers favor this move with more and more products available as OTCs. However, limiting the access to these products, especially in rural areas, has kept this issue still under debate.



INFORMATION SOURCES

Patients are exposed to volumes of information concerning OTC medications through a variety of sources.^[3] These include but are not limited to package labels, advertisements, advice from healthcare professionals, books, magazines, journals, and family members. It is essential, however, to recognize the value and significance of the key information made available to consumers through the proper labeling of nonprescription medications. The package label is one of the most important sources of information for consumers because it provides information at the time of purchase and at the time of product use. To ensure the safe and effective use of nonprescription products by consumers, the U.S. Food and Drug Administration (FDA) has developed specific guidelines that must be adhered to by all commercial manufacturers.^[18] These guidelines will be elucidated in depth in the following discussion.

OVER-THE-COUNTER HUMAN DRUGS: LABELING REQUIREMENTS—THE FDA GUIDELINES

The movement to reform OTC medication labels started in 1990, when the pharmacist planning service of Sausalito, California, together with various citizen groups petitioned the FDA to set standards for print size and style on the label of nonprescription medications.^[19] The FDA subsequently published a notice summarizing the contents of the citizen's petition and their comments on labeling of OTC medications in the Federal Register. In 1991, the NDMA convened a task force that conducted a comprehensive literature review and identified four factors that could affect the readability of OTC medication labels.^[19] These include medical, comprehension, illumination, and technical aspects necessary to improve OTC medication labels. These factors were addressed in the NDMA Label Readability Guidelines and included a set of recommendations that the drug manufacturers could adopt voluntarily.^[19] The FDA further requested comments from the public on font sizes, types, color, and various label design techniques that could enhance readability after the NDMA recommended their guidelines.

The American Pharmaceutical Association (APhA) suggested focusing particularly on four categories of information within a standardized format, which include primary use, dosing, cautions and major side effects, and active ingredients.^[19] A study conducted by Sansgiry and colleagues convinced the FDA that new guidelines were required for OTC labeling.^[20] This study concluded that the font size on OTC medication labels was very small and

consumers had difficulty reading the labels.^[20] Finally, in March 1999, the FDA published the new guidelines, entitled "Over-the-Counter Human Drugs: Labeling Requirements," mandating certain information on OTC drug labels.^[18]

The new label format should convey essential information to the consumer in a user-friendly format and provide consumers with improved information about the most suitable OTC medication for their condition, understanding its benefits and risks. The FDA has given 6-yr time for all the products to adopt the new labeling format with the first implementation of the rules to occur for most drugs during 2002.^[21] The time of labeling change is dependent on the date of the final monograph for each individual product. A drug that lacks the new label after its required implementation time will be considered misbranded and subjected to all the enforcement by the FDA. This labeling format would include a standardized drug facts panel, which includes information on active ingredients, purpose, use, and warnings.^[22] An example of a label with the new format can be seen in Fig. 2. Other proposed improvements include easier to read 6-point Helvetica type, distinct sections presented in standard order, and adequate spacing between letters, words, and lines of text. Fig. 3 includes an explanation on certain aspects of the new OTC labeling guidelines that were recommended.^[23]

SELECTING OTC MEDICATIONS

When it comes to selecting a nonprescription medication, it is important for the consumer to also understand the limitations and dangers of these drugs. Individuals must be educated that medications purchased OTC should never be taken for longer time periods or in higher doses than the label recommends. Symptoms that persist are a clear signal that it is time for triage to their physicians. It is also important to point out to consumers that they must read the label completely each time they purchase a product. Many times, two products from the same brand family are neither meant to treat the same condition nor may they contain the same active ingredient. This issue is further discussed in the section "OTC Brand Name Confusion."

The FDA felt the best way for consumers to select any OTC product is to cautiously read the label and talk to the pharmacist or physician in case of questions or doubts before product use. There are several things that the consumer should look for, when selecting an OTC medication, specifically the warnings. The important sections of the label that consumers need to review and evaluate before making purchase decisions can be seen in Table 4.

Drug Facts							
Active ingredient (in each tablet)	Purpose						
Chlorpheniramine maleate 2 mg	Antihistamine						
Uses temporarily relieves these symptoms due to hay fever or other upper respiratory allergies: ■ sneezing ■ runny nose ■ itchy, watery eyes ■ itchy throat							
Warnings Ask a doctor before use if you have ■ glaucoma ■ a breathing problem such as emphysema or chronic bronchitis ■ trouble urinating due to an enlarged prostate gland Ask a doctor or pharmacist before use if you are taking tranquilizers or sedatives							
When using this product ■ you may get drowsy ■ avoid alcoholic drinks ■ alcohol, sedatives, and tranquilizers may increase drowsiness ■ be careful when driving a motor vehicle or operating machinery ■ excitability may occur, especially in children If pregnant or breast-feeding, ask a health professional before use. Keep out of reach of children. In case of overdose, get medical help or contact a Poison Control Center right away.							
Directions <table border="1"> <tbody> <tr> <td>adults and children 12 years and over</td> <td>take 2 tablets every 4 to 6 hours; not more than 12 tablets in 24 hours</td> </tr> <tr> <td>children 6 years to under 12 years</td> <td>take 1 tablet every 4 to 6 hours; not more than 6 tablets in 24 hours</td> </tr> <tr> <td>children under 6 years</td> <td>ask a doctor</td> </tr> </tbody> </table>		adults and children 12 years and over	take 2 tablets every 4 to 6 hours; not more than 12 tablets in 24 hours	children 6 years to under 12 years	take 1 tablet every 4 to 6 hours; not more than 6 tablets in 24 hours	children under 6 years	ask a doctor
adults and children 12 years and over	take 2 tablets every 4 to 6 hours; not more than 12 tablets in 24 hours						
children 6 years to under 12 years	take 1 tablet every 4 to 6 hours; not more than 6 tablets in 24 hours						
children under 6 years	ask a doctor						
Other information store at 20-25° C (68-77° F) ■ protect from excessive moisture							
Inactive ingredients D&C yellow no. 10, lactose, magnesium stearate, microcrystalline cellulose, pregelatinized starch							

Fig. 2 Example of proposed OTC drug label format. (From Refs. [18], [23].)

Other ways to increase consumers' ability to read labels is to make sure that there is enough ambient light available at the point of display. It usually takes three times more light to read the same line at age 60 than at age 30.^[19] Elderly and the vision impaired should be advised to wear glasses or use magnifying glasses while reading labels to improve visibility.

U.S. FDA OTC REVIEW AND REGULATIONS

In the following, the acronym FDA will be used synonymously for U.S. FDA. The first legislation concerning the regulation of drugs in the United States was the Pure Food and Drug Act of 1906. This major federal law required that drugs only meet the standards of

Factor	Recommendation by FDA
Clear, easy to read type style	Helvetica or universe with no more than 39 characters per inch.
Bullets	Solid square or circle, 5 point type.
Text	Left justified
Hyphenation	Absent
Uppercase and lowercase letters only	No all caps
Font size	6 points (dark on light)

Fig. 3 New FDA guidelines for OTC medication labels. (From Refs. [18], [21], [22].)

**Table 4** Sections on OTC labels that consumers should consider before purchase

No.	Item	Comments
1	Product name	Is this the product that was recommended?
2	Active ingredients	Therapeutic substances in medicine
3	Purpose	Product category (such as antihistamine, antacid, or cough suppressant)
4	Uses	Symptoms or diseases that the product will treat or prevent
5	Warnings	When not to use the product, when to stop taking it, when to see a doctor, and possible side effects
6	Directions	How much to take, how to take, and how long to take it
7	Other information	Such as storage information
8	Inactive ingredients	Substances such as binders, colors, or flavoring agents. Some consumers may have allergies to these ingredients
9	Expiration date	Make sure the product has not expired or will expire during normal length of use
10	Tamper/child resistant packaging	Make sure all tamper resistant packaging is intact or do not purchase the product. Each product is allowed to have one dosage unit available that does not contain a child safe container in order to allow the elderly or infirmed patient to have the ease and accessibility to their medication

strength and purity claimed by the manufacturers. Drug safety was not mandated by law until the passage of the Federal Food, Drug, and Cosmetic (FDC) Act of 1938, which was prompted by negative reports on sulfanilamide. This act made no clear distinction between prescription and OTC medications but stated that all drugs must be safe for their intended use. No mention was made concerning the effectiveness of drugs in this legislation but labeling provisions were mandated which stated that products must be labeled with “adequate directions for use with conspicuousness and in terms such as to render it likely to be read and understood by the ordinary individual under customary conditions of purchase and use.”^[24] The 1951 Durham-Humphrey Amendment set up specific standards for classification that distinguished between a prescription and OTC medication. The prescription drug category included those medications that were deemed habit forming or toxic. Those medications were used for conditions that could not be self-diagnosed, and medications needing skills to administer or use them beyond the scope of the normal consumer were also included. This legislation eliminated consumer oriented labeling requirements for prescription products as these products could now be obtained only through prescriptive authority of a physician.

In 1962, the Kefauver-Harris amendments to the Food, Drug, and Cosmetic Act were passed, requiring that drug products be proved not only safe but also effective for their intended uses. Drugs on the market before 1938 were

grandfathered, but the 4500 new drug products, including 512 nonprescription products marketed after 1938, were to be reviewed by the FDA and removed from the marketplace if not found efficacious. The FDA, after contracting with the National Academy of Sciences/National Research Council (NAS/NRC) in 1966, began a procedure called Drug Efficacy Study Implementation (DESI) to assess the effectiveness of these products. It quickly became evident that there was a need to implement an extensive examination of the nonprescription medications in the marketplace. In 1972, the FDA began a large-scale scientific review of the more than 700 active ingredients in 300,000 different formulations to ensure that they were safe, effective, and properly labeled.^[24] This process commonly known as the “OTC Drug Review” was responsible for reclassifying many drugs from prescription to nonprescription status and for establishing regulations for proper labeling. It was made up of 17 expert panels that convened over a 10-yr period. Manufacturers submitted scientific literature to the group showing the effectiveness of their products. These expert panels developed monographs for each ingredient, developed antacid testing procedures, removed daytime sedatives from the market, created the SPF rating scale, implemented procedures for tamper resistant packaging, developed a tooth paste abrasion index, and put into action the aspirin-Reye’s syndrome warnings on aspirin containing products.



DRUG APPROVAL PROCESS

The 1938 FDC act required that all new drugs, including new drug products introduced after 1938, be proven safe for human use before being marketed and be cleared in advance through a new drug application (NDA). However, FDA's OTC drug review evaluated all nonprescription drugs for safety, effectiveness, and labeling, regardless of the date of marketing entry. The 1938 FDC act defines the market being divided into new drugs, which are defined by law as being recognized as safe and effective (RASE), and those that are generally recognized as safe and effective (GRASE). During phase I of the process, an Advisory Review Panel was set up, which gave advanced notice of proposed rulemaking. At this time a drug was given one of three different categories:

- Category I—GRASE.
- Category II—not recognized as safe and effective (NRASE).
- Category III—insufficient evidence to prove safety and efficacy.

Phase II deliberations then began, the outcomes of which concluded if the product can continue to be marketed pending further review or if the manufacturers need to reformulate their product in order to include it in Category I. During this phase, manufacturers were given a tentative final monograph that was printed in the Federal Register. The Final Rule, phase III, was also published in the Federal Register and explained if the product was safe and effective (Category I) and met monograph conditions or if the product was classified as not having met monograph conditions (Category II and III). This gave manufacturers the time necessary to prove their drugs safe and effective before marketing.

NEW DRUG APPLICATION PROCESS

One mechanism of gaining general recognition status for nonprescription drugs approved by the FDA's OTC review was through an NDA. NDA is necessary for a drug that is defined by law as NRASE until it has been precleared and approved by the FDA. Because of the OTC drug review, excess of 40 primary products have been reclassified from prescription to OTC. There are three ways in which such a change can be made outside the drug review process.^[24] The manufacturer may file an NDA if the clinical research data, which proves that the drug can be approved for an OTC indication or OTC dosage levels, is available. The FDA itself can file for reclassification if it is determined that prescription status is not necessary for the safe and effective use of the drug.

If the postmarketing safety experience of a drug can prove that the drug can be used safely without physician supervision, then a supplement to the NDA can be filed. For the drug to be even considered for the OTC status, the drug should meet certain criteria^[25]:

1. The indication or indications that the product is to be used as an OTC drug should be similar to the prescription indication or indications, and it must allow for easy self-diagnosis and monitoring by the patient.
2. The drug to be considered should not have unfavorable adverse event(s) and drug interaction(s). The drug should have relatively low toxicity and low potential for abuse.
3. The drug should not have a narrow therapeutic index nor require special monitoring or laboratory procedures to ascertain toxicity or therapeutic effect.
4. The drug should not have any other criteria that would make it impractical for OTC use.

Rx-TO-OTC SWITCH

In recent years, there has been a huge impetus for many drugs that were recognized as prescription only to be marketed as OTC medications. The reasons for this include patent expiration, increasing market share, and consumer demand. However, chronic conditions are a big hurdle for Rx-to-OTC switches and most of the drugs being switched are used for treating acute conditions (e.g., headache).^[26] This is because chronic OTC medication use raises questions about self-monitoring. Switches from prescription to nonprescription status have a huge impact on the health care economy. A list of some recently switched OTC medications can be viewed in Table 5. Studies have shown that these switched products can save Americans billions of dollars when taking into account prescription costs, doctor visits, insurance costs, and lost time from work.^[27] The overall effect of the Rx-to-OTC switch on healthcare costs is more than one can expect.^[28] The Wall Street Journal recently estimated that 30%–50% of healthcare costs for people over the age of 65 are for prescription drugs. A 1997 study showed that the American consumer could save approximately \$13 billion per year by using OTC medications switched from medications that had been available by prescription only.^[27] Additional information has suggested that a woman can save over \$80 by using switched OTC vaginal antifungal products for previously diagnosed recurring yeast infections.^[27] Furthermore, it has been recognized that consumers may save more than \$750 per year from cough/cold switches illustrated by the fact that doctor visits for the common cold fell by 110,000 per year

**Table 5** Examples of recent Rx-to-OTC switches

Product name	Manufacturer/distributor	Indication
Actron (ketoprofen)	Bayer Corp. Consumer Care	Internal analgesic
Axid AR (nizatidine)	Whitehall-Robins Healthcare	Stomach acid reducer
Lamisil AT (terbinafine hydrochloride 1%)	Novartis	Antifungal
Lotrimin Ultra (butenafine hydrochloride 1.0%)	Schering-Plough	Athlete's foot, jock itch, ringworm
Monistat 3 (miconazole nitrate 4.0%)	Advanced Care Products	Anticandidal
Orudis KT (ketoprofen)	Whitehall-Robins Healthcare	Internal analgesic
Pepcid AC (famotidine)	Johnson & Johnson-Merck	Stomach acid reducer
Rogaine (minoxidil 5.0%)	Pharmacia-Upjohn	Hair growth
Tagament HB (cimetidine)	SmithKline Beecham	Stomach acid reducer
Vagistat-1 (triconazole 6.5%)	Bristol-Myers	Anticandidal
Zantac 75 (ranitidine HCl)	Warner Wellcome	Stomach acid reducer

between 1976 and 1994, which resulted from switched products from prescription to OTC status.^[29] Rx-to-OTC switches account for over 30% of the total U.S. OTC market.^[27]

Types of Prescription to OTC Switches

There are two types of prescription to OTC switches: Complete Switch and Partial Switch. When a product becomes a nonprescription product with all dosage strengths, if more than one, switched with no prescription version of the product remaining, it is called a complete switch. An example of this type of switch is the changing of Nicorette gum from prescription to nonprescription status. Nicorette gum, 2 mg and 4 mg, is used OTC for smoking cessation. If a product has one or more doses that were formerly prescription only or the company manufactures lower strengths of the product that were previously unavailable on prescription but the higher doses still remain available only through prescriptions, then this is called a partial switch. An example of this switch is the Naproxen sodium preparations. Aleve, the OTC switched product, contains 220 mg Naproxen sodium per tablet, while Anaprox, 275 mg Naproxen sodium, is still available only by prescription. Benefits for a drug to be reclassified from prescription only to nonprescription status are numerous. The section "Benefits of Reclassification to OTC Status" will discuss benefits of Rx-to-OTC switches from manufacturers, consumers, and healthcare providers' prospective.

BENEFITS OF RECLASSIFICATION TO OTC STATUS

The potential for a product to be marketed for a longer period of time is quite an advantage for the manufacturer of a drug. By switching a drug that is threatened by patent

expiration from Rx-to-OTC status, it expands the market potential of the product and allows manufacturers to recoup the costs of bringing the drug to market originally. The switch from Rx-to-OTC also allows manufacturers to move into different competitive markets. The sale of an OTC switched product that was formerly prescription only has at times almost doubled or tripled.^[30,31] Because consumers tend to purchase more recognizable brand names when buying an OTC product, former prescription only products tend to capture a large market share when converted to OTC status. Approximately 50% of the drugs that were switched to OTC from prescription status are the top sellers in the U.S. marketplace.^[29]

Patients also benefit tremendously from the Rx-to-OTC switches. When a product that was formerly in prescription only status is classified as OTC, it allows patients to self-treat and gain more control over their health.^[31] This is especially true in cases where medical advice was once required, e.g., smoking cessation. The patient can now save time and money that was previously spent on visiting a physician. It has been shown that patients saved approximately \$150 million a year since hydrocortisone 1% was switched to OTC status.^[32]

The role of the pharmacist as a self-care advisor has been accentuated as more drugs are switched from Rx-to-OTC status and patients seek advice from accessible sources. Although manufacturers publish booklets on the use of their OTC products, counseling by the pharmacist is imperative. The pharmacist is the drug therapy expert and is in the best position to know what questions to expect from their patients and answer those queries as it pertains to their patient's individual health. As more and more OTC medications become available, the pharmacist will be consulted more often on drug interactions along with adverse effects and limitations of drug use for self-care. Finally, the pharmacist can advise the patient if they should be treating their condition with an OTC product or



seeking the advice of their physician. The topic regarding the role of pharmacist in self-care is discussed further in this article.

PROBLEMS WITH RECLASSIFICATION

A large issue of concern to the consumer is the cost of their medication. One issue with an Rx-to-OTC switch is that the drug, previously covered by an insurance company due to its prescription standing, is now no longer covered because of its change to OTC status. This may lead to patient confusion and frustration. Other factors that may contribute to these feelings include dual Rx and OTC marketing; in most of these cases, the OTC product is marketed in lower dose strength. However, in some cases the marketed OTC strength can be used to treat the same medical condition as the prescription strength of the product. Usually, the OTC strength should be used for a shorter duration of time or fewer times a day than the prescription strength, however this may lead to patient confusion over which dose of the medication is best to treat their particular symptoms. The second problem deals with the misperception regarding drug safety. Most patients feel that OTC medications are safe and totally free from side effects due to the OTC status. This perception may be due to the fact that consumers are allowed to buy OTC products without any consultation with a healthcare provider. This may lead consumers to make false assumptions about the safety and tolerability of a particular drug product, especially as it concerns other medications or health conditions they may have. Few consumers recognize that some OTC medications may have deleterious effects on their body including effects on blood pressure, blood glucose, and heart rate. A third problem concerns misleading advertisements by manufacturers. The limited medical knowledge of the patient makes them highly susceptible to such advertisements. These commercial advertisements help sway consumers to buy a particular product and can increase market share in the highly competitive OTC industry. Finally, an important concern of the healthcare practitioner concerning the reclassification is the loss of direct patient counseling over the medication. The FDA has recently attempted to rectify this by adding to the label of nonprescription medications the mandate to "ask your doctor or pharmacist before using this or any medications."^[18] With the advent of managed care and formulary management, insurance companies are also very interested in the Rx-to-OTC switch process. The petition brought forth to the FDA by WellPoint Health Networks, Inc. in 2001 exemplifies this change in attitude. This landmark case began as a campaign by California Blue Cross pharmacists. Blue Cross of California's parent

company, WellPoint Health Networks, Inc. petitioned the FDA to make three prescription antihistamines OTC: Claritin, Zyrtec, and Allegra. Schering-Plough, Pfizer, and Aventis, manufacturers of these products, opposed this change, claiming these drugs need to be monitored by doctors for safety reasons. Usually the manufacturer seeks the Rx-to-OTC change due to patent expiration, but in this case, the manufacturers were allied in the fact that these products were not ready to go OTC. Only once, in the last 18 yr, has the FDA approved nonprescription sales of a drug without a petition or support from the manufacturer. On May 11, 2001, the U.S. FDA advisory committee ruled that Claritin, Zyrtec, and Allegra should be sold OTC as allergy relievers. The advisory panel stated that manufacturers' own advertisements claimed that side effects of these drugs were no worse than placebo. The FDA said it agreed with that advice but issued no ruling. Surprisingly, seeing the potential for increased sales in the future, Schering-Plough Corporation decided to change tactics and filed an FDA application to switch Claritin voluntarily to nonprescription status, not only for nasal allergies but also for hives. The FDA went back to its advisors to ask if patients could correctly self-diagnose ordinary hives instead of a more serious rash, and if so, whether Claritin constituted an appropriate nonprescription treatment for this condition. Advisors to the FDA voted that enough evidence exists that Claritin can treat chronic hives without a sufferer needing a prescription but urged more studies for acute hives. The FDA is not bound by its advisers' recommendations but typically in the past has followed the advice of its advisory committees. It is evident that in the future we will see Claritin and many more products being marketed OTC.

OTC BRAND NAME CONFUSION

One major problem for the consumer brought about by the Rx-to-OTC switch process and the fact that manufacturers understand the draw of brand name recognition is that there is a great deal of confusion concerning ingredients used in a named drug product.^[33,34] Manufacturers retain the familiar trade names even though there are changes in the formula, because of immense brand loyalty. Pharmaceutical manufacturers spend a large percentage of their revenue on advertising their products to the general public through brand name recognition. A popular marketing tactic has been to use these established brand names to increase sales of other products in their line. This practice is known as creating line extensions. This sales tactic can be confusing to consumers especially if the product name carries over into entirely new and different product lines.^[33,34] For example, Tylenol[®], a product manufactured by McNeil could refer to analgesic, a cold



Nonprescription Drugs

11

medication, influenza medication, or a sinus product. The number of products with the Tylenol brand name marketed in the United States during 2000 can be viewed in Table 6. In other cases, a certain therapeutic outcome may be associated with a product's brand name that is not entirely accurate. The brand name product Fungi-nail[®], a solution manufactured by Kramer laboratories, implies that it will cure nail fungus. Although the pharmacist understands it is very difficult to achieve effective treatment of the nail fungus through topical treatment alone, consumers may not be aware. Brand loyalty is so important that manufactures often maintain a familiar trade name even though products in the line do not contain the same active ingredients. This is exemplified by Unisom sleep products: Unisom Sleep Gels[®] contain diphenhydramine hydrochloride as the active ingredient and Unisom SleepTabs[®] contain Doxylamine succinate as the sleep-causing agent. The pharmacist needs to be aware of and knowledgeable about these situations so they can counsel patients appropriately concerning patients' misconceptions.

THIRD CLASS OF DRUGS

Because of the confusion associated with OTC products and due to the numerous Rx-to-OTC switches, there are

many advocates of having a third class of drugs. Historically, the distribution of medicine in the United States is based on a two-class system: prescription and nonprescription. A nonprescription drug is one that the U.S. FDA has found to be safe and effective for direct consumer use based on the label instructions and warnings. A wider margin of safety is required for nonprescription drugs due to absence of physician and sometimes even pharmacist supervision. A third class of drugs would be available only through a pharmacist. Drugs in the third class would not be available in other convenient retail outlets, such as discount stores, grocery stores without a pharmacy, hotel gift shops, and convenience stores. The current two-class system serves two of the most fundamental demands of access and affordability. The FDA has repeatedly rejected a third class or transition class of drugs. As recently as March 9, 1994, the FDA stated that it would be inappropriate to restrict the sale of OTC drugs to pharmacies because such a restriction would reduce the number of outlets (from 750,000 to 65,000 nationwide) where consumers could purchase OTCs, limit competition, and raise some OTC drug prices, with no attendant public health benefit. Diminished competition will also have a disproportionate impact on traditionally underserved populations in urban and rural areas of the country. The U.S. Justice Department and the National Association of Attorneys General have opposed a third class of drugs, calling such proposals anticompetitive and anticonsumer because they create a monopoly in the distribution of nonprescription drugs.

Although the outlook for a third class of drugs is not bright, there are some groups that are proponents of this measure. The National Association of Boards of Pharmacy (NABP) passed a resolution in 1995 asking for legislation on this issue.^[35] The National Community Pharmacists Association (NCPA) passed a resolution in 1982 calling for a "pharmacist legend class" that would create a "buffer" period during which pharmacists could learn about and dispense newly switched medications before they are sold in locations with no healthcare intervention.^[35,36] The APhA also issued support for a third class of drugs.^[37] Finally, pharmacists themselves, understanding the ramifications of misuse of OTCs, backed the concept of a third class of medications that would only be available behind the counter.^[36] Unfortunately, this may be seen by opponents only as a special interest group trying to gain a monopoly on certain products to increase their own profits at the expense of other retailers. However, when more products are switched to OTC, consumer may find it confusing to remember all the information when making product choices, especially drug interactions. Hence, the issue of third class of drug may not be resolved yet.

Table 6 Example of potential confusion due to product line extensions

Brand name	Ingredients/dose
Tylenol	APAP 1000 mg
Tylenol PM	APAP 1000 mg Dextromethorphan hydrobromide 50 mg
Tylenol Cough	APAP 1000 mg Dextromethorphan hydrobromide 30 mg
Tylenol Sinus	APAP 1000 mg Pseudoephedrine hydrochloride 60 mg
Tylenol Allergy	APAP 1000 mg
Sinus	Pseudoephedrine hydrochloride 60 mg Chlorpheniramine maleate 4 mg
Tylenol Cold	APAP 650 mg Pseudoephedrine hydrochloride 60 mg Dextromethorphan hydrobromide 30 mg
Tylenol	APAP 650 mg
Multi-Symptom	Pseudoephedrine hydrochloride 60 mg Dextromethorphan hydrobromide 30 mg Chlorpheniramine maleate 4 mg



NONPRESCRIPTION DRUG ADVERTISEMENTS

Pharmaceutical companies spent an estimated \$1.8 billion on direct-to-consumer advertisements for prescription drugs in 1999.^[38] Although the FDA has the power to prohibit the sales of falsely advertised products, the FDA does not regulate or have authority over nonprescription drug advertising. Such an authority rests with the Federal Trade Commission (FTC). Consumers make product purchase decisions from the information that they receive through advertisements. If these advertisements are misleading, the consumers' drug knowledge is not adequate to detect this misinformation.^[39] This may lead to inappropriate purchases and possible adverse reactions. The Federal Trade Commission Act was amended in 1970 prohibiting the advertisers in using language that was not approved by the FDA, when describing the therapeutic benefit of nonprescription products. The standards imposed on prescription medications are more stringent than those for OTC medications. An advertisement for prescription medications not only points out the beneficial effects of the medication but also states contraindications and warnings. Advertisements for OTCs generally present only information on the beneficial effects of the medication, seldom mentioning contraindications or safety information. The FDA requires advertising to maintain a "fair balance," speaking to safety as well as efficacy. The FTC has no such requirement.^[40]

The National Association of broadcasters and the NDMA developed a code, which laid down the guidelines for manufacturers to follow in creating television advertisements for nonprescription drugs. According to these guidelines, the manufacturer among other things should also make sure that advertisements contain no claims of product effectiveness that are not supported by clinical or other scientific evidence and present no information suggesting that the product prevents or cures a serious condition that must be treated by a licensed practitioner.^[5] An in-depth discussion of OTC advertisements can be found in upcoming articles of the encyclopedia.

CONSIDERATION FOR SPECIAL AT-RISK POPULATIONS

Many segments of the population in the United States are considered high risk for any drug therapy because of differences in their ability to deal with side effects of medications, differences in their pharmacokinetic parameters as pertains to adsorption, distribution, and metabolism of drugs, or they have drug effects that are unique to their population due to age. These groups of

individuals may experience higher incidence of drug related problems than the typical patient. This is especially true in the nonprescription arena where these products are so accessible. Considering these issues the FDA had indicated a list of drug-drug interactions (Table 7).^[41] Products for pediatric individuals, pregnant or lactating women, and the geriatric consumers require special attention and monitoring when it comes to OTC medications.

While many OTC drugs can be used during pregnancy and lactation under a physician's supervision, some are known to be unsafe. As indicated on product labels, women who are pregnant, who may be pregnant, or who are nursing should consult a doctor before taking OTC medication.^[42,43] Women, who are pregnant or lactating, should be discouraged from using nonprescription medications without consulting their primary care provider. It is important to know what trimester a woman is in during her pregnancy because certain OTC medications can be used safely during different times of gestation. During breast-feeding, many medications can pass into the breast milk causing problems to the nursing infant. Nursing mothers who need to take OTC medications should take them right after breast-feeding or before an infant's nap period.^[43] Additionally, breast-feeding women should avoid products that are advertised as extra-strength, maximum strength, or long acting. Pregnant and lactating women should be advised to use nonpharmacological modalities whenever reasonably possible.

Children under 2 yr should not be given any OTC drug without doctor's approval. Before parents administer OTC drugs to their young children, they should consider whether they are truly necessary for the child.^[44] One of the major problems associated with recommending OTC medications to children is that it is difficult to extract an accurate medical history from them. Although most of the time a second hand history is available through their parents, it is difficult to recommend a nonprescription medication that will truly meet their needs for symptomatic relief. Additionally, most nonprescription medications do not include dosing information for those youngsters under the age of 2 yr, even those medications packaged for children. Several studies have indicated the inappropriateness of administering OTC cough and cold medications to a pediatric population.^[45-49] These OTC cough and cold medications are associated with several serious side effects in children.^[50-62]

Another potential population at risk that also consumes the most number of OTC products are the elderly. Nonprescription medications should be administered to the geriatric individual only after a complete assessment of comorbid disease states. Social, economic, and physiological factors all weigh into the decision of what OTC

**Table 7** Selected OTC drug interaction warnings for some OTC drug products

Product category	Drug interaction information
Acid reducers H2 receptor antagonists	For products containing cimetidine, ask a doctor or pharmacist before use if you are: taking theophylline (oral asthma drug), warfarin (blood thinning drug), or phenytoin (seizure drug)
Antacids	Ask a doctor or pharmacist before use if you are: allergic to milk or milk products if the product contains more than 5 g of lactose in a maximum daily dose taking a prescription drug Ask a doctor before use if you have: kidney disease
Antiemetics	Ask a doctor or pharmacist before use if you are: taking sedatives or tranquilizers Ask a doctor before use if you have: a breathing problem, such as emphysema or chronic bronchitis glaucoma or difficulty in urination due to an enlarged prostate gland When using this product: avoid alcoholic beverages
Antihistamines	Ask a doctor or pharmacist before use if you are taking: sedatives or tranquilizers a prescription drug for high blood pressure or depression Ask a doctor before use if you have: glaucoma or difficulty in urination due to an enlarged prostate gland breathing problems such as emphysema, chronic bronchitis, or asthma When using this product: alcohol, sedatives, and tranquilizers may increase drowsiness avoid alcoholic beverages
Antitussive	Ask a doctor or pharmacist before use if you are: taking sedatives or tranquilizers Ask a doctor before use if you have: glaucoma or difficulty in urination due to an enlarged prostate gland
Bronchodilators	Ask a doctor before use if you: have heart disease, high blood pressure, thyroid disease, diabetes, or difficulty in urination due to an enlarged prostate gland have ever been hospitalized for asthma or are taking a prescription drug for asthma
Laxatives	Ask a doctor before use if you have: kidney disease and the laxative contains phosphates, potassium, or magnesium stomach pain, nausea, or vomiting
Nasal decongestants	Ask a doctor before use if you: have heart disease, high blood pressure, thyroid disease, diabetes, or difficulty in urination due to an enlarged prostate gland
Nicotine replacement products	Ask a doctor before use if you: have high blood pressure not controlled by medication have heart disease or have had a recent heart attack or irregular heartbeat, since nicotine can increase your heart rate Ask a doctor or pharmacist before use if you are: taking a prescription drug for depression or asthma using a prescription nonnicotine stop smoking drug Do not use: if you continue to smoke, chew tobacco, use snuff, or use other nicotine-containing products
Nighttime sleep aids	Ask a doctor or pharmacist before use if you are: taking sedatives or tranquilizers

(Continued)

**Table 7** Selected OTC drug interaction warnings for some OTC drug products (*Continued*)

Product category	Drug interaction information
	Ask a doctor before use if you have: a breathing problem such as emphysema or chronic bronchitis glaucoma difficulty in urination due to an enlarged prostate gland When using this product: avoid alcoholic beverages
Pain relievers	Ask a doctor before taking if you: consume three or more alcohol-containing drinks per day
Stimulants	When using this product: limit the use of foods, beverages, and other drugs that have caffeine. Too much caffeine can cause nervousness, irritability, sleeplessness, and occasional rapid heartbeat be aware that the recommended dose of this product contains about as much caffeine as a cup of coffee
Topical acne	When using this product: increased dryness or irritation of the skin may occur immediately following use of this product or if you are using other topical acne drugs at the same time. If this occurs, only one drug should be used unless directed by your doctor

(From Ref. [41].)

medications should be suggested for the elderly individuals. Pre-existing medical conditions may affect the choice and use of nonprescription medications. For example, the use of nonprescription antihistamines should be avoided for those with glaucoma, urinary retention from benign prostatic hypertrophy, and emphysema. Many of the OTC sympathomimetics seen in the decongestants contained in nonprescription cough and cold medications may have effects on blood pressure and blood glucose in those patients with hypertension or diabetes, respectively. The elderly population consumes more drugs than any other group in our society. Because of this, polypharmacy, or the use of multiple medications plays a role in the use and choice of nonprescription medications. Many elders suffer from osteoarthritis or other diseases of pain and inflammation. Prescription and OTC medications may both contain ingredients that treat their pain but may contribute to stomach bleeding as a side effect of some pain medications. It is important for the geriatric individual to understand all the medications he is taking and the names of the active ingredients so they do not duplicate therapy. This group of patients is also prone to higher drug–drug interactions due to the high possibility of concurrent therapy. Pharmacokinetics parameters in the form of decreased renal function or liver function may also play a role in the elimination and possible toxicity of certain nonprescription medications. In summary, the special needs and risks of these groups of people need to be taken into account when recommending nonprescription drugs.

OTHER POTENTIAL DRUG THERAPY RISKS

There is a general misconception that since OTC medications are readily available, they are safe to be used by adults.^[63] It has been stressed throughout this article that because OTC medicines are available without a prescription does not mean that they should be taken without careful thought and consideration. When used by the wrong person, in the wrong dose, or in the wrong combinations, they can actually endanger the life they are meant to improve. An example of this is that Aspirin contributed to 33 intentional and accidental deaths in 1998, the most recent year for which data is available from the American Association of Poison Control Centers.^[64] APAP has recently come under attack for problems with liver toxicity. APAP is the most widely used pharmaceutical analgesic and antipyretic agent in the United States and the world; it is contained in more than 100 products. As such, APAP is one of the most common pharmaceuticals associated with both intentional and accidental poisoning.^[64] It is available OTC in 325 mg and 500 mg dosage forms. It is also found as a component of many prescription pain relievers including Darvocet, Percocet, and others. The maximum daily dose of APAP is 4 g in adults and 90 mg kg⁻¹ in children. The toxic dose of APAP after a single acute ingestion is 150 mg kg⁻¹ or approximately 7 g in adults, although the at-risk dose may be lower in persons with alcoholism and other susceptible individuals. However, when dosing recommendations are followed, the risk of liver toxicity is extremely small.^[64]



It is important that consumers understand the maximum dosages for OTC medications and read the labels to understand how often and how long they should self-treat their medical conditions so they do not overdose on these drugs. Overdose is an obvious pitfall to avoid, but even the right dose in combination with the wrong underlying condition can cause problems. The age of information technology has provided a lot of convenience to patients who have an access to the World Wide Web. There are several web sites available on the Internet (www.med-broadcast.com, with the Drug Check option) that actually allow users to enter the name of the drug, brand, or generic and provide all kinds of information ranging from how the medication works to contraindications. These sites would definitely help in decreasing the risk of wrong medications and would help the patient to know in advance the types of side effects they should expect. A list of websites that deal with the self-care industry around the world can be viewed in Table 8.

ROLE OF PHARMACIST IN NONPRESCRIPTION PRODUCT THERAPEUTICS

As one of the most trusted professions, the pharmacist exercises a lot of influence on the choice of OTC medication by patients. Research shows that 52% of the patients consult their pharmacist for OTC medications.^[65] A list of medications or OTC products most recommended by pharmacist^[66] can be viewed in Table 9. The pharmacist, who is the drug therapy expert, should and does play a large role in the counseling of patients on their OTC choices. Americans consume over \$50 billion nonprescription pain relief tablets each year to self-treat problems such as headaches, muscle aches, and arthritis. Many consumers do not bother to read the labels to prevent product misuse, they do not understand the risks or benefits of what they are taking, and many of them cannot

read the labels correctly. Patients may ask for recommendation to use an OTC from the pharmacist or even a nonprescription medicine by name. This familiarity with a product name does not imply that the patient has adequate knowledge about the drug. A survey done in 1998 by the APhA, which commissioned the Yankelovich study, to examine consumer attitudes and behavior regarding OTC medicines showed that among adult respondents 85% reported they have taken a nonprescription pain reliever at some point for pain and fevers, in fact one-third reported weekly use of these medicines.^[67] Forty-seven percent of all adults said that they do not always read medication labels; fewer than 40% consult pharmacists about these products; and over one-third said they were not aware of the risks associated with these medicines. Furthermore, 85% of the respondents said they had not heard of nonsteroidal anti-inflammatory drugs, or NSAIDs, which include aspirin and ibuprofen. When asked to name an NSAID, more adults mistakenly said APAP rather than aspirin. Of those surveyed, 74% did not associate stomach upset or bleeding with aspirin, and 43% were not aware of potential risks from taking an OTC pain reliever and prescription medication at the same time.^[67] These statistics point out the huge role for the pharmacist in the counseling and advising of OTC medications. As discussed previously, consumers have access to information about drug products through many avenues including the Internet, family members, and mass media. Some of this information may be very valuable but may also be erroneous. Consumers do not necessarily know about the side effects the drug may cause and the severity of the side effects nor are they aware of any contraindication to use, which can be pointed out during patient counseling. A study conducted by Sclar, Robinson, and Skaer^[68] has shown that consultation by final year pharmacy students resulted in almost 42% of the consumers changing their purchase decision. This resulted in a reduction in OTC medication expenditure by an average of U.S. \$1.53 per customer. This study also demonstrated that

Table 8 List of web sites for self-care industry around the world

Organization name	Web site
The Association of the European Self-Medication Industry (AESGP)	http://www.aesgp.be
Australian Self Medication Industry	http://www.asmi.com.au
China Nonprescription Medicines Association (CNMA)	http://www.cnma.org.cn
Consumer Healthcare Products Association (CPHA)—United States	http://www.chpa.org
Nonprescription Drug Manufacturers Association of Canada	http://www.ndmac.ca
Proprietary Association of Great Britain	http://www.pagb.co.uk
Organisation of Pharmaceutical Producers of India (OPPI)	http://www.indiaoppi.com
Proprietary Association of Japan	http://www.otc.gr.jp
World Self-Medication Industry (WSMI)	http://www.wsmi.org

**Table 9** List of some OTCs pharmacists recommend the most

Category	Most recommended product	Recommendation (%)	Number of recommendations/month for the category in general
Analgesics			
APAP	Tylenol	97	12
Aspirins	Bayer	80	9
Ibuprofens	Motrin	78	11
Ketoprofens	Orudis KT	86	4
Naproxen sodium	Aleve	96	6
Cough/cold/allergy			
Adult cold preparations	Comtrex	18	14
Children cold preparations	Dimetapp	44	11
Adult cough medications	Robitussin DM	47	14
Children cough medications	Triaminic	28	12
Allergy relief products	Chlor-Trimetom	44	13
Asthma relief products	Primatene	82	5
Flu remedies	Theraflu	51	8
Sinus remedies	Sudafed	45	11
Throat lozenges	Cepacol	20	8
Throat sprays	Chloraseptic	83	7
Gastrointestinal medications			
Antacids	Mylanta	55	7
Antidiarrheals	Imodium A-D	94	6
H2 antagonists	Pepcid AC	87	7
Hemorrhoidal preparations	Preparation H	69	4
Irritant or stimulant laxatives	Dulcolax	59	4
Topical dermatologicals			
Vaginal antifungals	Monistat	72	4

(From Ref. [66].)

the length of the consultation significantly influenced consumers to change their final purchase.

Pharmacist can help a patient know if there are any drug–drug or drug–food interactions with their OTC medications. A thorough medication history of both prescription and nonprescription medication including herbal supplements is a vital component to pharmaceutical care. By evaluating the medication history, the pharmacist can help the consumer make better-informed decisions concerning the purchase and use of nonprescription products. Additionally, the pharmacist has a responsibility to report to the patient's physician, the manufacturer, and the regulatory authorities for medicine, any relevant information about an adverse drug reaction encountered by the patient, which is associated with the purchase of a nonprescription medication.

In summary, the pharmacist has a significant role in assisting the consumer in their self-treatment. The pharmacist should assist the consumer in product selection, assess patient risk factors for OTC use, counsel the patient regarding proper drug use, maintain OTC medications on the patients medication profile, help the patient monitor themselves for toxicity and efficacy of the OTC medication, discourage the use of deceptive or

“quack” remedies, and prevent consumer delays in seeking appropriate treatment from other healthcare practitioners, if necessary. The pharmacist can make use of intervention strategies relating to patient/pharmacist interactions in the form of providing information, removing products from point of sale, sharing information with other local pharmacists, and making referrals to other members of the health care team.^[69]

FUTURE

The future of nonprescription medication clearly indicates growth. Americans will continue to rely on OTC medications as the most important aspect of self-care. Rx-to-OTC switches will continue to be a part of the drug industry's strategy to improve sales and increase consumer access to their products. With these continued changes there will arise the obvious problem of insurance coverage for OTCs. Most insurance companies do not cover OTC medications under their prescription benefits currently, but with increasing pressure from consumers and healthcare providers, this may change. Nonprescription medications, especially switched products, may become first line



Nonprescription Drugs

17

therapy in algorithms of patient care when dealing with managed care companies. The questions of reimbursement through medication copies are an issue that will certainly arise with insurance companies in time. With the new labeling requirements proposed by the U.S. FDA for OTC products, consumers should have a better knowledge regarding the medications they intend to take. However, the role of the pharmacist should continue to increase in the counseling of these products. With the vast amount of the population using these medications, new product availability, and the increasing interest in the American public in utilizing self-care, the pharmacist must focus their expertise on counseling and providing information to the public concerning these medications, which if used properly will have a large impact on our healthcare costs and satisfaction.

REFERENCES

1. Top 200 OTC/HBC Brands in 2001. *Drug Topics* **2002**, 146 (5), 53–54.
2. Self Care in the New Millennium: American Attitudes Toward Maintaining Personal Health and Treatment, Prepared by Roper Starch Worldwide, Inc., for CHPA, 2001.
3. Lowell, M. Consumer Trends in Self-Diagnosis, Home Testing and Over the Counter Medication Implications for the Registered Nurse. *Kans. Nurse* **1995**, 70 (5), 1–2.
4. The Role of Pharmacist in Self-Care and Self-Medication: <http://www.who.int/medicines/library/dap/who-dap-98-13/who-dap-98-13.pdf> (accessed Oct 7, 2002).
5. Covington, T.R. Self-Care and Non-prescription Pharmacotherapy. In *Handbook of Non-prescription Drugs*, 12th Ed.; APhA: Washington, D.C., 2000; 3–12.
6. Neafsey, P.J.; Strickler, Z.; Shellman, J.; Padula, A.T. Delivering Health Information About Self-Medication to Older Adults: Use of Touchscreen-Equipped Notebook Computers. *J. Gerontol. Nurs.* **2001**, 27 (11), 19–27.
7. Lowe, N.K.; Ryan-Wenger, N.M. Over the Counter Medications and Self-Care. *Nurse Pract.* **1999**, 24 (12), 34–44.
8. Barnett, N.L.; Denham, M.J.; Francis, S.A. Over-the-Counter Medicines and the Elderly. *J. R. Coll. Phys. Lond.* **2000**, 34 (5), 445–446.
9. CHPA. Aging and Health: The Role of Self-Medication, 1998. http://www.chpa-info.org/publications/Aging_and_Health.asp (accessed Nov 7, 2002).
10. OTC Facts and Figures: http://www.chpa-info.org/statistics/OTC_Facts_and_Figures.asp (accessed Oct 17, 2002).
11. OTC Sales by Category: http://www.chpa-info.org/statistics/OTC_Sales_by_Category.asp (accessed Oct 31, 2002).
12. ACNielsen Begins Tracking OTC Remedy Sales in Convenience Stores: <http://acnielsen.com/news/american/us/2002/20021007.htm> (accessed Nov 7, 2002).
13. National Health Expenditures: <http://healthchec.org/hottopics/nhe/home.html> (accessed Nov 7, 2002).
14. McFee, R.B.; Mofenson, H.C.; Caraccio, T.R. Dextromethorphan: Another Ecstasy? *Arch. Fam. Med.* **2000**, 9, 123.
15. OASAS. Over the Counter Danger: <http://www.oasas.state.ny.us/AdMed/otc.htm> (accessed Oct 27, 2002).
16. Hughes, G.F.; McElnay, J.C.; Hughes, C.M.; McKenna, P. Pharm Abuse/Misuse of Non-prescription Drugs. *Pharm. World Sci.* **1999**, 21 (6), 251–255.
17. Blank, C. Is This the Right Time for a Third Class of Drugs? *Drug Topics* **1997**, 141 (11), 106.
18. Over-the-Counter Human Drugs: Labeling Requirements. *Fed. Regist.* **1999**, 64 (51), 13254–13303.
19. Non-prescription Drug Manufacturers Association: Label Readability Guidelines; NMDA: Washington, D.C., 1991; 20036.
20. Sansgiry, S.; Cady, P.; Patil, S. Readability of Over-the-Counter Medication Labels. *J. Am. Pharm. Assoc.* **1997**, NS37, 522–528.
21. FDA. Better Medical Product Information, Status Report August 2000, OTC Labeling: <http://www.fda.gov/ope/nprstat98/infostat.htm> (accessed May 28, 2002).
22. FDA. OTC Labeling: Questions and Answers: <http://www.fda.gov/cder/otc/label/quesanswers.htm> (accessed Mar 4, 2002).
23. The New Drug Facts Label: <http://www.bemedwise.org/label/label.htm> (accessed Oct 17, 2002).
24. Gilbertson, W.E. FDA's Review of OTC Drugs. In *Handbook of Nonprescription Drugs*, 10th Ed.; Covington, T.R., Lawson, L.C., Eds.; APhA: Washington, D.C., 1993; 21–37.
25. Botstein, P. Criteria for Switching New Molecular Entities from Prescription to Over-the-Counter. *Drug Inf. J.* **1990**, 24, 29–31.
26. Juhl, R.P. Prescription to Over-the-Counter Switch: A Regulatory Perspective. *Clin. Ther.* **1998**, 20, C111–C117.
27. From Rx-to-OTC: The Switch Process, CHPA: http://www.chpa-info.org/issues/Switch_Process.asp (accessed Oct 29, 2002).
28. Brass, E.P. Changing the Status of Drugs from Prescription to Over-the-Counter Availability. *N. Engl. J. Med.* **2001**, 345 (11), 810–816.
29. Juhl, R. Switching Drug Products from Rx to OTC Status a Benefit to Pharmacy Practice. *US Pharm.* **2000**, 25 (4), 119–128.
30. Whitney, D.W. Product Liability Issues for the Expanding OTC Drug Category. *Food Drug Law J.* **1993**, 48 (3), 321–350.
31. Soller, R.W. Evolution of Self Care with Over the Counter Medications. *Clin. Ther.* **1998**, Suppl. C, C134–C140.
32. Bachrach, E.E. The FDA's New Over the Counter Drug Office and Advisory Committee: An Industry Perspective. *Food Drug Law J.* **1993**, 48, 563–566.
33. Gannon, K. Are All in the Family OTC Brands Misleading? *Drug Topics* **1992**, 136 (18), 50–52.
34. Pray, S.S. Confusion in the Market Place. *US Pharm.* **1995**, 20 (1), 16–18.



35. Ukens, C. Class Act—NABP Backs Third Class of Drugs. *Drug Topics* **1995**, 139 (10), 13.
36. Gannon, K. Rx to OTC Switches Raise New Call for Third Class. *Drug Topics* **1989**, 133 (20), 48.
37. Marshall, R.P. A Transition Category of Drugs: Win–Win–Win? *Am. Pharm.* **1992**, NS32 (1), 64–67.
38. Woloshin, S.; Schwartz, L.M.; Tremmel, J.; Welch, H.G. Direct-to-Consumer Advertisements for Prescription Drugs: What Are Americans Being Sold? *Lancet* **2001**, 358 (9288), 1141–1146.
39. Sansgiry, S.; Sharp, W.T.; Sansgiry, S.S. Accuracy of Information on Printed Over the Counter Drug Advertisements. *Health Mark Q* **1999**, 17 (2), 7–18.
40. Sheftell, F.; Dodick, D.; Cady, R. Direct-to-Consumer Advertising of OTC Agents Under Current FTC Regulations: Concerns and Comment. *Headache* **2001**, 41 (6), 534–536.
41. CDER. Drug Interactions You Should Know: <http://www.fda.gov/cder/consumerinfo/druginteractions.htm> (accessed Nov 7, 2002).
42. Mercill, A.W. Non-prescription Drugs. In *Principles of Pharmaceutical Marketing*, 3rd Ed.; Smith, M.C., Ed.; Lea and Febiger: Philadelphia, 1983; 206–227.
43. American Academy of Pediatrics Committee on Drugs. Transfer of Drugs and Other Chemicals into Human Milk. *Pediatrics* **1994**, 93, 137–150.
44. Williams, R. How to Give Medicines to Children. *FDA Consumer* **1996**, 30 (1), 6–9.
45. Smith, M.B.; Feldman, W. Over the Counter Cold Medications: A Critical Review of Clinical Trials Between 1950 and 1991. *J. Am. Med. Assoc.* **1993**, 269, 2258–2263.
46. Taylor, J.A.; Novack, A.H.; Almquist, J.R.; Rogers, J.E. Efficacy of Cough Suppressants in Children. *J. Pediatr.* **1993**, 122, 799–802.
47. Hutton, N.; Wilson, M.H.; Mellits, D. Effectiveness of an Antihistamine–Decongestant Combination for Young Children with the Common Cold: A Randomized, Controlled Trial. *J. Pediatr.* **1991**, 118, 125–130.
48. Clemens, C.J.; Taylor, J.A.; Almquist, J.R.; Quinn, H.C.; Mehta, A.; Naylor, G.S. Is an Antihistamine–Decongestant Combination Effective in Temporarily Relieving Symptoms of the Common Cold in Pre school Children? *J. Pediatr.* **1997**, 130, 463–466.
49. American Academy of Pediatrics. Committee on Drug Use of Codeine and Dextromethorphan Containing Cough Remedies in Children. *Pediatrics* **1997**, 99, 918–920.
50. Kernan, W.N.; Viscoli, C.M.; Brass, L.M. Phenylpropanolamine and the Risk of Hemorrhagic Stroke. *N. Engl. J. Med.* **2000**, 343, 1826–1832.
51. Burton, B.T.; Rice, M.; Shmertzler, L.E. Atrioventricular Block Following Overdose of Decongestant Cold Medication. *J. Emerg. Med.* **1985**, 2, 415–419.
52. Cetaruk, E.W.; Aaron, C.K. Hazards of Non prescription Medications. *Emerg. Med. Clin. N. Am.* **1994**, 12, 483–510.
53. Lake, C.R.; Gallant, S.; Masson, E.; Miller, P. Adverse Drug Effect Attributed to Phenylpropanolamine: A Review of 142 Case Reports. *Am. J. Med.* **1990**, 89, 195–208.
54. Pentel, P. Toxicity of Over-the-Counter Stimulants. *J. Am. Med. Assoc.* **1984**, 252, 1898–1903.
55. Glick, R.; Hoying, J.; Cerullo, L.; Perlman, S. Phenylpropanolamine: An Over the Counter Drug Causing Central Nervous System Vasculitis and Intra cerebral Hemorrhage. Case Reports and Review. *Neurosurgery* **1987**, 20, 969–974.
56. Kase, C.S.; Foster, T.E.; Reed, J.E.; Spartz, E.L.; Girgis, G.N. Intracerebral Hemorrhage and Phenylpropanolamine Use. *Neurology* **1987**, 37, 399–404.
57. Myer, P.; Bryant, B.; Cartwright, G.W. Overdose with a Combination Decongestant, Antihistamine and Antitussive Syrup. *Indiana Med.* **1985**, 78, 766–768.
58. Conway, E.E.; Walsh, C.A.; Palomba, A.L. Supraventricular Tachycardia Following the Administration of Phenylpropanolamine in an Infant. *Pediatr. Emerg. Care* **1989**, 5, 173–174.
59. Weesner, K.M.; Denison, M.; Roberts, R.J. Cardiac Arrhythmias in an Adolescent Following Ingestion of an Over-the-Counter Stimulant. *Clin. Pediatr.* **1982**, 21, 700–701.
60. Pender, S.E.; Parks, B.R. Toxicity with Dextromethorphan Containing Preparation: A Literature Review and Report of Two Additional Cases. *Pediatr. Emerg. Care* **1991**, 7, 163–165.
61. Gardin, J.M.; Shumacher, D.; Constantine, G.; Davis, K.D.; Leung, C.; Reid, C.L. Vulvular Abnormalities and Cardiovascular Status Following Exposure to Dexfenfluramine or Phenteramine/Fenfluramine. *J. Am. Med. Assoc.* **2000**, 283, 1703–1709.
62. Rostagno, C.; Cacioli, S.; Felici, M.; Gori, F.; Neri Sernieri, G.G. Dilated Cardiomyopathy Associated with Chronic Consumption of Phenidmetrazine. *Am. Heart J.* **1996**, 131, 407–409.
63. Kacew, S. Effect of Over the Counter Drugs on the Unborn Child: What Is Known and How Should This Influence Prescribing? *Pediatr. Drugs* **1999**, 1 (2), 75–80.
64. Susan, E.; Farrell, M.D. Program Director, Instructor, Department of Emergency Medicine, Harvard Medical School, Beth Israel Deaconess Medical Center: <http://www.emedicine.com/emerg/topic819.html> (accessed Oct 31, 2002).
65. Alkhawajah, A.M.; Eferakeya, A.E. The Role of Pharmacists in Patients' Education on Medication. *Public Health* **1992**, 106 (3), 231–237.
66. Fleming, H. OTCs Pharmacists Recommend Most. *Drug Topics* **2000**, 144 (14), 24–32.
67. Levy, S. APhA, McNeil, R.Ph.s Join Forces to Educate Consumers. *Drug Topics* **1998**, 142 (22), 48.
68. Sclar, D.A.; Robinson, L.M.; Skaer, T.L. Pharmacy Consultation and Over the Counter Medication Purchasing Outcomes. Over the Counter Medication Intervention Project Team. *J. Clin. Pharm. Ther.* **1996**, 21 (3), 177–184.
69. MacFadyen, L.; Eadie, D.; McGowan, T.J. Community Pharmacists' Experience of Over-the-Counter Medicine Misuse in Scotland. *R. Soc. Health* **2001**, 121 (3), 185–192.

NON-INVASIVE PEPTIDE AND PROTEIN DELIVERY

Patrick J. Sinko

Gerald Scucci

Rutgers University, Piscataway, New Jersey, U.S.A.

Yong-Hee Lee

Trega Biosciences, Inc., San Diego, California, U.S.A.

OVERVIEW

Peptide and protein drugs have been increasingly utilized in the treatment of various diseases since recent dramatic advances in recombinant DNA and modern synthetic technologies have allowed for the cost-effective production of considerable quantities of protein pharmaceuticals. For the past several decades, major efforts have been directed toward developing effective means for the nonparental (i.e., noninvasive) delivery of protein pharmaceuticals. However, low bioavailability has limited the success of noninvasive delivery attempts (1–5). The primary reasons for low bioavailability include unfavorable physicochemical characteristics such as large molecular weight, charge, hydrophilicity, and physicochemical instability, and biological limitations such as poor membrane permeability, and presystemic enzymatic metabolism. Administration via these routes often requires specialized delivery systems, absorption enhancers, and/or proteolytic enzyme inhibitors to improve bioavailability. Various noninvasive routes of administration have been investigated and reviewed to deliver protein pharmaceuticals, including oral, mucosal-membrane, pulmonary, and transdermal routes (6, 7). To date, nearly all therapeutic proteins are administered by intravenous (IV), subcutaneous (SC), or intramuscular (IM) injection. A partial listing of protein pharmaceuticals in clinical use is shown in Table 1.

The only notable success in the noninvasive delivery of peptides or protein pharmaceuticals has been the oral delivery of small peptide drug analogs such as the penicillins, cephalosporins, ACE inhibitors, and renin inhibitors (8, 9). The intestinal absorption pathways of small peptides differ significantly from those for larger peptides and proteins. Smaller peptide drugs undergo carrier-mediated absorption by means of the human intestinal peptide transporters (hPepT1 and hPT1) (8–10). Peptides analogs with greater than three amino acid residues are not typically transported via the peptide

carriers, although the exact mechanisms of absorption have not yet been fully elucidated. The focus of this chapter is on the noninvasive delivery of larger peptide and protein pharmaceuticals. Some success has been achieved in the noninvasive delivery of protein pharmaceuticals, most notably in nasal delivery. Several examples of systemic protein delivery via noninvasive routes are described in this chapter. The chapter concludes with a focused discussion on strategies for the oral delivery of protein pharmaceuticals using a model small protein drug, salmon calcitonin (sCT).

NONINVASIVE ROUTES

Oral Delivery

The mucosal surface of the digestive tract is a vast area covered by a monolayer of epithelial cells joined by tight junctions that provide an effective barrier to absorption. The potential convenience and improved patient compliance associated with oral delivery have led to considerable research in this area. Some protein drugs have surprisingly good absorption characteristics. For example, the cyclic 11-amino acid compound cyclosporine has a relatively high bioavailability when properly formulated. Compared with the standard formulation (Sandimmune, Novartis, NJ), the microemulsion formulation (Neoral) increases the AUC and C_{\max} , and reduces t_{\max} , inpatient variability, and the effect of bile on cyclosporine absorption (11, 12). Oral absorption of cyclosporine from the Sandimmune formulation is shown in Fig. 1. With the exception of cyclosporine, very limited success has been achieved in delivering protein pharmaceuticals orally (Table 1). Detectable quantities of growth hormone, insulin, and calcitonin can be found in the systemic circulation after oral administration. However, the absorption of these drugs is highly inefficient. Typically, upward of 1000 times more drug is required

Table 1 Therapeutic polypeptide drugs and their clinical application

Therapeutics	Molecular weight (in Daltons)	Indications	Route of administration ^a
Alglucerase	59,300	Gaucher's disease	IV infusion
α-1 Antitrypsin	52,000	Congenital α-1 antitrypsin deficiency	IV infusion
Calcitonin	4500	Paget's disease, postmenopausal osteoporosis, hypercalcemia	IM, SC, intranasal
Cyclosporine	1200	Prophylaxis for allogeneic organ rejection	Oral, IV infusion
Desmopressin	1183	Intranasal: primary nocturnal enuresis IV Infusion: hemophilia A and von Willebrand's disease	Intranasal, SC, IV infusion
Dornase alfa	37,000	Cystic fibrosis	Inhalation
Erythropoietin	30,400	Treatment of anemia	SC, IV infusion
Etanercept	150,000	Rheumatoid arthritis	SC
Factor IX	55,000	Christmas disease	IV infusion
Filgrastim	18,800	Severe chronic neutropenia	SC
Glatiramer acetate	Average: 4,700–11,000	Multiple sclerosis	SC
Imiglucerase	60,430	Gaucher's disease	IV infusion
Insulin	Varies: minimum; 6000	Diabetes mellitus	SC, IV infusion
Interferon alfacon-1	19,434	Hepatitis C infection	SC
α-Interferon	19,271	Hairy cell leukemia	IM, SC
β-Interferon	22,500	Multiple sclerosis	SC
γ-Interferon	16,000–25,000	Reduce frequency of infections associated with chronic granulomatous disease	SC
Oxytocin	1007.2	Labor induction	IV infusion
Proleukin	15,300	Carcinoma	IV infusion
Retepase	39,571	Management of acute myocardial infarction	IV infusion
Sargramostim	Varies: 19,500, 16,800; and 15,500	Myeloid reconstitution	SC, IV infusion
Somatrem	22,000	Growth hormone	SC, IM
Streptokinase	47,000	Fibrinolytic	IV infusion
Thymosin α1 (Thymalfasin)	3108	Chronic hepatitis B	SC (currently in phase 3 studies)
Tissue plasminogen activator	70,000	Fibrinolytic	IV infusion

^a SC, subcutaneous; IM, intramuscular; IV, intravenous.

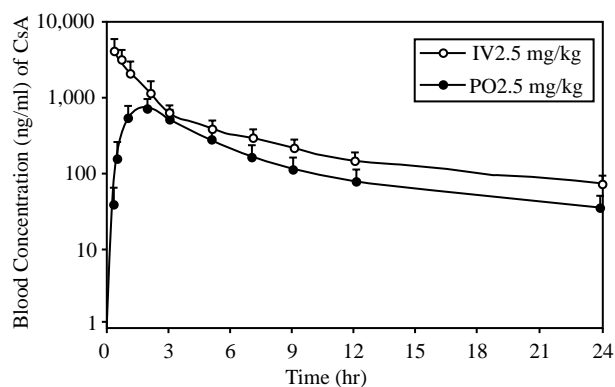


Fig. 1 Mean blood cyclosporine A (CsA) concentration (ng/ml)-time curves in 22 psoriatic patients (mean \pm SD) after a single IV administration of 2.5 mg/kg (IV infusion over 2 h with 0.5% solution of CsA in 0.9% NaCl) and after 1 week, oral administration of the same dose (Sandimmune, oral solution 100 mg/ml). The mean bioavailability was 44% (range, 22–63%). (From Galla, F.; Marzocchi, V.; Croattino, L.; Poz, D.; Baraldo, M.; Furlanut, M. Oral and Intravenous Disposition of Cyclosporine in Psoriatic Patients. *Ther. Drug Monitor.* **1995**, *17*, 302–304.)

to achieve the same effect after oral administration than when given by IV or SC injection. The oral delivery of protein drugs is critical for keeping health care costs low owing to enhanced compliance, reduced expenses compared with inpatient therapies, etc. The primary problems encountered with the oral delivery of protein pharmaceuticals include the poor intrinsic permeability owing to their hydrophilic nature and large molecular size, presystemic enzymatic metabolism by intestinal proteases and peptidases, chemical instability including tendencies to aggregate, and/or nonspecific binding to a variety of physical and biological surfaces. Because protein drugs are highly susceptible to all these factors, protein-delivery systems will have to simultaneously address all the issues to achieve higher bioavailability (13).

Although several approaches can be taken toward solving an oral bioavailability problem, the causes of incomplete bioavailability must first be understood. When poor membrane permeability is identified as a probable cause, approaches to formulate the drug with absorption enhancers that transiently modify biological membranes are used. Potential approaches to reduce presystemic metabolism include the transient modulation of the intestinal environment by limiting the activity of intestinal enzymes through the use of protease inhibitors (14); adjusting the local pH to values that correspond to the pH minima of specific enzymes present in the gut (15–17); maintaining high local drug concentrations to saturate

enzymes; regiospecific targeting because of regional differences in the activity of intestinal proteolytic enzymes, dilution, and spreading patterns; and surface area differences (18, 19). Other potential oral delivery approaches are chemical modification of proteins to produce prodrugs and analogs (20–22), substitution of D-amino acids to reduce hydrolysis (23, 24), and use of bioadhesive polymers that have a variety of mechanisms for enhancing protein drug absorption including inhibition of proteolytic enzymes or reducing the resistance of tight junctions (25).

Several attempts have been made to facilitate intestinal protein drug absorption by targeting specific absorptive transporter systems, such as the monosaccharide or bile acid transporters, or by modulating the secretory transport activity of *P*-glycoprotein (Pgp) (9). The coupling of unstable peptides with sugars has been demonstrated to improve hydrolytic stability and membrane permeability (26). Insulin has been modified with *p*-nitrophenyl- α -D-glucopyranoside and *p*-nitrophenyl- α -D-mannopyranoside. The coupling of these agents enhanced membrane permeation by an unknown mechanism and reduced the potential for enzymatic hydrolysis (27). In terms of modulating secretory transport, the intestinal absorption of some peptides that interact with Pgp, such as cyclosporine, tends to increase to some extent when efflux transporter inhibitors are used (28). Another delivery strategy includes the utilization of macromolecular ligands. Recently, the cellular and molecular mechanisms whereby macromolecules and particles are internalized have been studied (10). Although the use of endocytic pathways is currently highly inefficient, the mechanisms are being studied in cell culture models such as Caco-2 and Madin-Darby canine kidney (MDCK) cells in the hope of improving efficiency. By gaining a better understanding of how to target membrane transport mechanisms or modulate biological barriers, the oral bioavailability of protein drugs may be enhanced. Until then, the oral delivery of protein drugs will remain a considerable challenge for the foreseeable future, given the multiple biological obstacles that must be overcome to achieve therapeutic blood concentrations.

Mucosal Delivery

The high vascularity and accessibility of the mucosal membranes have made this tissue a potential route of protein delivery. Mucous membranes include the nasal, buccal, ocular, rectal, and vaginal membranes (29). Advantages of mucosal routes are that they bypass hepatic first-pass metabolism and are readily accessible, and locally acting agents such as penetration enhancers, enzyme inhibitors, and mucus-suppressing agents can be used.

Mucus is a highly viscous product that forms a protective coating over the lining of organs in contact with external media. Mucus is a mixture of large glycoproteins (mucins), water, electrolytes, sloughed epithelial cells, enzymes, bacteria and bacterial products, and various other materials, depending on the source and location of the mucus (30). Mucins are synthesized either by goblet cells lining the mucosal epithelium or by special exocrine glands with mucus cell acini. Mucoadhesive polymers have received considerable attention as platforms for protein drug delivery. Advantages include the ability to deliver drugs locally, prolong the residence time, and optimize contact with the absorbing surface to permit modification of tissue permeability to inhibit enzyme activity or to suppress mucus production (25–30). Bioadhesive bonds may be physical or mechanical bonds, secondary chemical bonds, or primary chemical bonds. Other approaches for mucosal protein delivery are formulating proteins in liposomes (7), microemulsions, and small particles (e.g., nanoparticles) (31, 32). The common rationale in all three cases is the protection of proteins from the local environment before to absorption and localization of the protein at or near the cellular membrane to optimize the driving force for passive permeation.

Nasal delivery

The nasal route has been intensively investigated because of its convenience (33). It presents a large surface area ($\sim 200 \text{ cm}^2$) that is lined by a single layer of columnar epithelial and goblet cells overlaying a rich blood supply. The extensive network of blood capillaries underneath the nasal mucosa provides an important part of the driving force required for the systemic delivery of proteins. Because the mucosal surface is covered by mucin that is cleared on a 15–30 min cycle, the residence time of solutions in the nasal mucosa is short. Therefore, mucoadhesive polymer solutions are typically required to increase the residence time. This route has been shown to be acceptable for peptides with 10 or fewer residues, whereas satisfactory bioavailability is obtained for proteins of 20 or more residues only when permeation enhancers are used (33).

The nasal mucosa contains enzymes capable of hydrolyzing peptides. The predominant enzyme appears to be aminopeptidase, among other exopeptidases and endopeptidases. The cytochrome P-450 activity in the olfactory region of the nasal epithelium is higher than in the liver (34). Phase II enzyme activity has also been found in the nasal epithelium (34). The metabolic cleavage of thyrotropin-releasing hormone (TRH) and met-enkephalin has been demonstrated in human nasal epithelial cell

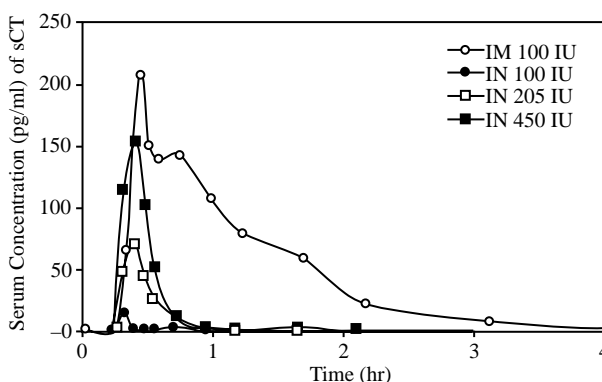


Fig. 2 Mean serum salmon calcitonin (sCT) concentration (pg/mL)-time curve after intramuscular (IM) administration of 100 IU of sCT and intranasal (IN) administration of 100, 205, and 450 IU of sCT with 0.5% sodium tauro-24,25-dihydrofusidate (STDHF) in 10 healthy subjects (7 males, 3 females). Relative to the i.m. dose, the bioavailabilities of the i.n. formulations containing 0.5% STDHF were 3.9, 7.9, and 7.4%, respectively. The nasal calcitonin doses were administered as one 100- μ l spray per nostril using two single-dose spray units. sCT-STDHF formulations were prepared by reconstitution of solid sCT with a sterile isotonic (NaCl) solution of 0.5% STDHF in 20 mM acetate buffer, pH 5.0, and used immediately. (From Lee, W.A.; Ennis, R.D.; Longenecker, J.P.; Bengtsson, P. The Bioavailability of Intranasal Salmon Calcitonin in Healthy Volunteers with and without a Permeation Enhancer. *Pharm. Res.* **1994**, *11*, 747–750.)

monolayers, suggesting that the nasal mucosa may be a significant metabolic barrier to the systemic delivery of protein drugs (35). Permeation enhancers and protease inhibitors are usually coadministered to achieve successful delivery of proteins by this route. Intranasal absorption of salmon calcitonin is demonstrated in Fig. 2. Relative to the IM dose, the bioavailability of intranasal (IN) administration of sCT was low, 3.9–7.9%. When eel calcitonin (eCT) was coadministered with nafamostat mesilate, a protease inhibitor, or with sodium decanoate, the sodium salt of fatty acid, the nasal absorption of eCT was significant, and serum calcium (Ca^{2+}) concentration was effectively decreased (36). Several nasal products have been available clinically including buserelin, desmopressin, oxytocin, and calcitonin (29).

Buccal delivery

Delivery of protein drugs through the buccal mucosa has also received considerable attention (37). Despite its apparent disadvantages, such as limited absorptive surface area and moderate mucosal permeability, the buccal route might be acceptable for small- or medium-sized peptide drugs. Potential advantages for buccal delivery are the

avoidance of gastrointestinal (GI) or hepatic first-pass metabolism, the feasibility of locally controlled absorption enhancement, the ease of administration and removal of an administered device, and the possibility of prolonged drug delivery and action. The buccal epithelium has an average thickness of 500–600 μm . The epithelial cells at the outer surface are continuously peeled off because of abrasion that occurs during mastication. Therefore, starting from nondifferentiated cells located above the basal membrane, fresh epithelial cells are produced by mitosis. During their 5- to 8-day passage from the basal membrane to the outer surface, the epithelial cells undergo maturation resulting in a change in form and size.

The buccal cavity exhibits greater proteolytic enzyme activity than does the nasal or vaginal mucosa. The metabolic activity is shown to reside primarily in the epithelium (38). A mucoadhesive buccal patch was evaluated in rabbits for transmucosal delivery of oxytocin (OT) after incorporation into custom coformulations of Carbopol 974P and silicone polymer (39). It was observed that plasma OT concentrations remained 20- to 28-fold greater from 0.5 to 3.0 h than concentrations in control animals administered placebo patches. Several transmucosal therapeutic systems (TmTs) were also investigated to study the enhanced/controlled delivery of leuteinizing hormone-releasing hormone (LHRH) in dogs (40). The TmTs is a track field-shaped bilayer mucoadhesive device consisting of fast- and sustained-release layers. A stream of 0.5–2 L of saliva constantly washes the oral cavity daily. The resulting salivary layer covering all oral epithelia can interfere with the penetration of drugs as well as with the adhesion of buccally administered delivery devices. Therefore, several factors must be considered in designing buccal dosage forms—the taste of drugs and excipients, the size of the dosage form, and the mechanism for fixing the dosage form onto the oral mucosa. Nevertheless, even under optimized conditions, the buccal delivery of proteins may not allow for bioavailabilities as high as that with other mucosal sites. Thus, the chances for buccal protein delivery, if any, will be restricted to special cases and for special proteins of high permeability. In this instance, however, buccal delivery might be a preferred route of administration, mainly owing to the undisputed acceptance and compliance of oral dosage forms and to the unmatched robustness of the epithelium.

Ocular delivery

The ocular route has been investigated for the systemic delivery of protein drugs (41, 42). Systemic absorption of proteins after topical administration to the eye results through contact with the conjunctival and nasal mucosa, the latter occurring as a result of the drainage through the

nasolacrimal duct. When systemic absorption is desired, absorption through the conjunctival and nasal mucosa needs to be maximized. The systemic delivery of proteins through the ocular route has several advantages including convenience, rapid absorption/onset, and avoidance of hepatic first-pass metabolism, and it is amenable to controlled-release delivery. However, this route typically yields low bioavailability, although it is well accepted by patients. As with other delivery routes, the size, charge, and hydrophilic nature of proteins are the primary determinants that limit their extent of absorption. Because the systemic delivery of proteins depends on their transport mechanisms and contact time with mucosal membrane of the conjunctiva and the nasolacrimal system (43), the following factors need to be considered: 1) precorneal factors such as tear drainage, instilled volume, viscosity, pH, and tonicity; 2) that the eye formulations used must avoid the potential for local irritation and/or side effects; and 3) that inclusion of permeation enhancers may be necessary when dealing with agents with a molecular weight higher than 10,000 Da.

Systemic delivery of insulin and glucagon through the ocular route was demonstrated as a feasible alternative to parental injection, particularly when permeation enhancers were added (44).

Rectal delivery

The rectal epithelium is columnar or cuboidal, with numerous goblet cells. However, unlike the small intestine, the rectal epithelium does not contain villi. The human rectum has a length of 5 inches and a surface area of only approximately 200–400 cm^2 , compared with 2,000,000 cm^2 for the small intestine. Consequently, absorption from the rectum could be much lower than from the remainder of the GI tract (45). The lower venous drainage system (inferior and middle hemorrhoidal veins) is connected directly to the systemic circulation by the ileac vein and vena cava, whereas the upper venous drainage system (superior hemorrhoidal vein) is connected to the portal vein system. Thus, an opportunity to reduce the extent of hepatic first-pass elimination exists in the rectum. The rectum also has a large number of lymphatic vessels that offer an opportunity to target drug delivery to the lymphatic circulation.

Extensive studies have been conducted regarding the rectal absorption of proteins, especially insulin. However, the rectal absorption of this drug is low, probably owing to a combination of poor membrane permeability and metabolism at the absorption site (46). Therefore, it is generally believed that absorption enhancers are required to achieve therapeutic plasma levels of rectally administered proteins. The rectal route can be an

extremely useful route for the delivery of drugs to infants and young children in whom difficulties can arise using per oral administration. Historically, the rectum has not been an accepted site of drug delivery. Its principal applications have been for local therapy, e.g., hemorrhoids, and for systemic delivery of drugs to populations presenting practical problems for parenteral or oral dosing (e.g., the elderly, infants, patients with epilepsy, etc.). Even though the systemic delivery of peptide can occur by this route, it is not widely accepted among the populations of the world.

Vaginal delivery

The vaginal wall consists of an epithelial layer (epithelial lamina and lamina propria), muscle layer, and tunica adventitia. It is regulated by cyclic alteration of the reproductive system, which is directly controlled by hormones such as estrogen, progesterone, leuteinizing hormone (LH), and FSH (45). Before puberty, the epithelium is very thin, but after puberty, it increases in thickness with estrogen activity. In the adult stage, the vaginal surface during the follicular phase appears homogeneous with large superficial polygonal cells with a high degree of proliferation caused by estrogen stimulation and the presence of cornification. This proliferation of cells concomitantly increases the epithelial thickness and number of layers. The lamina propria specialized supporting structure of the epithelial cells contains a blood supply, a lymphatic drainage system, and a network of nerve fibers. The vaginal epithelium is aglandular but is usually covered with a surface film of moisture. The pH of the vaginal lumen is controlled mainly by the lactic acid produced from cellular glycogen by the action of the normal microflora, Doderlein's bacilli. The arterial blood supply in the vagina is derived from the visceral branches of the internal iliac artery, and venous drainage occurs mainly via the uterine vein to the internal iliac vein.

It is known that several peptide hormones and antigenic proteins are absorbed intact through the vaginal membrane and that the bioavailability is greater than that by the oral route because of higher intercellular permeability and reduced first-pass metabolism. The amount and duration of the hypoglycemic and hypocalcemic effects induced by intrauterine delivery of insulin and calcitonin, respectively, were equivalent to those obtained after s.c. injections in intact and ovariectomized rats (47). Regarding the enzymatic barrier, few enzymes have been found in the vaginal epithelium and in the peptidase activity against enkephalins, substance P, insulin, and proinsulin in the absorptive mucous membranes in the rabbit. In fact, supernatants of homogenates of the vaginal, nasal, buccal, rectal, and ileal mucous membranes of

rabbit exhibit similar proteolytic activity (46). Vaginal delivery of sCT was demonstrated effectively in ovariectomized rats using the highly mucoadhesive polymer, the benzyl ester of hyaluronic acid (Hyaff 11), by closely adhering to the mucosal surface and by protecting the drug from enzymatic inactivation (48). Most of the vaginal preparations on the market are used for local action on the vaginal membrane, using antibacterial, antifungal, or antiviral agents.

Pulmonary Delivery

Pulmonary delivery provides an attractive route of administration for systemic protein delivery. Of all the noninvasive routes for protein delivery, the pulmonary route has provided the most encouraging data and has recently generated great interest in the biotechnology industry (49, 50). Advantages of the pulmonary route include the fact that the walls of the alveoli are thinner than are other epithelial/mucosal membranes, that the surface area of the lung is much greater, and that the lungs receive the entire blood supply from the heart. Of course, the lungs are rich in enzymes, and overcoming this barrier is no easy task. Peptide hydrolases, peptidases, and a wide variety of proteinases are present in lung cells. The respiratory tract has several unique features (51): 1) a large surface area that can be exposed to drug almost simultaneously as opposed to the intestine, which has a similar total surface area but does not allow for simultaneous exposure; 2) a high blood flow that does not directly expose absorbed drug to the clearance mechanisms present in the liver; and 3) relatively less metabolic activity.

The upper respiratory tract, including the trachea and large bronchi, has a relatively limited surface area for absorption compared with the alveolar region, which provides more than 95% of the surface area of the lung. The respiratory tract is lined on its luminal surface by a layer of columnar epithelial cells that become progressively less columnar in the smaller airways and alveolar region. The surface of the airway epithelial cells in the larger conducting airways is covered with cilia that aid in the clearance of material from the lung. Two types of epithelial cells are present in the alveolar region of the lung: the type I and the type II pneumocytes. Type I cells are flat cells with broad, thin extensions covering up to 95% of the alveolar surface, whereas type II cells are cuboidal cells without extensions that can differentiate into type I cells and participate in the repair of the epithelial cell surface after damage. The alveolar epithelium is assumed to be the site of protein delivery, based on surface area considerations. The passages leading to the lower lung from the nasal region are narrower than

the oral passages. This allows for much more efficient filtration and much less efficient delivery after nasal administration and has led to the development of a variety of devices that deliver aerosols via the mouth. Dry and liquid particles can be prepared and inhaled with the aid of dry powder dispensers, liquid aerosol generators, and nebulizers. These devices produce particles that range from 1 to 5 μm , which may penetrate deeply into the alveoli of the lung. Once deposited deep in the lung, the alveoli provide a large surface area (80–140 m^2) for rapid transfer into the pulmonary circulation.

In the literature, molecules ranging to greater than 100,000 Da have been demonstrated to be appropriate candidates for pulmonary delivery. Several proteins are currently under investigation for systemic delivery, including insulin, calcitonin, leuteinizing-hormone-releasing hormone (LHRH) analogs, granulocyte colony-stimulating factor (gCSF), and human growth hormone (hGH). Inhalation of regular insulin for meal time (i.e., postprandial) glucose control has been found to be safe, efficacious, and reliable in type I and type II diabetes patients (52). Compared directly with s.c. injection, inhaled insulin provides equivalent glucose control. A potential advantage of aerosol insulin is that it is more rapidly absorbed (serum peak at 5–60 min) and cleared than SC injection (peak at 60–150 min), which provides a more relevant and convenient therapy for mealtime glucose control. Deftos et al. (53) have evaluated the intrapulmonary (IP) delivery of sCT in normal subjects with a dry powder inhaler. Compared by dose, i.p. sCT had 66% of the bioactivity and 28% of the bioavailability of intramuscular (IM) sCT (Fig. 3).

Transdermal Delivery

Considerable effort has been given to the transdermal delivery of pharmaceutical proteins, but clinical applications have thus far been limited to nonprotein drugs (54). The skin is impermeable to molecules as large as proteins. The physical (stratum corneum) and proteolytic enzymatic barriers create a formidable barrier against any permeation under normal circumstances. A molecular size of approximately 1000 Da is generally believed the molecular size limit for transdermal delivery, but there are some reports on the permeation of macromolecules across the stratum corneum by passive diffusion. In addition to molecular size, the lipid-protein partition coefficient of the penetrant is very important. The skin also consists of an enzymatic barrier capable of metabolizing proteins. The composition of enzymes and the spectrum of metabolic reactions in the skin are similar to those in the liver, although the skin

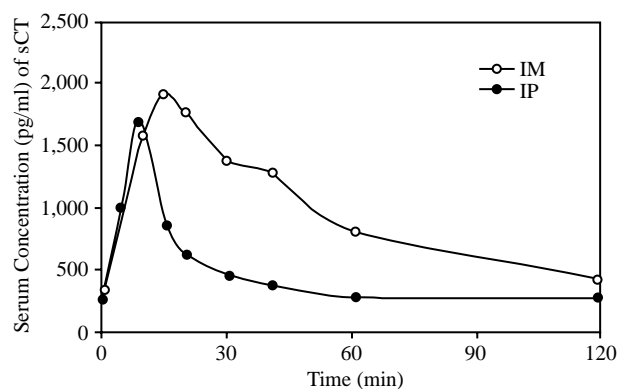


Fig. 3 Mean serum salmon calcitonin (sCT) concentration (pg/ml)-time curves after intramuscular (IM) administration of 100 IU of sCT and intrapulmonary (IP) administration of 160 or 320 IU of sCT in 10 normal males. sCT was administered with a dry powder-delivery inhaler. Dosage adjustments were made by dividing serum sCT concentrations by dose units. The apparent concentration of sCT at time zero represents assay blank in an RIA method. There were no additional substantial changes in sCT after 120 min. Compared by dose, IP sCT had 66% of the bioactivity and 28% of the bioavailability of IM sCT. (From Ref. 53.)

has significantly less metabolic activity than does the liver (55).

Various strategies have been attempted to surmount these barriers. These include the use of protease inhibitors to suppress enzymatic activity and the use of penetration enhancers to reversibly reduce the barrier resistance of the stratum corneum. Other alternatives include forced delivery under an electric field (iontophoresis) and ultrasonic energy (phonophoresis), but their efficacy has been limited by the large size and relatively low electrical charge of proteins. Green et al. (56) and Singh et al. (57) review iontophoresis.

Iontophoresis generates an electrical potential gradient that facilitates the movement of solute ions across the membrane and has been used with the greatest success in the treatment of hyperhidrosis. However, iontophoresis is capable of delivering large hydrophilic proteins in a continuous manner over a prolonged period because of their charged nature. Human studies have been performed to demonstrate the safe, effective, and reproducible delivery of a positively charged calcitonin analog with a molecular weight of approximately 3000 to the systemic circulation (Fig. 4). Some success has been reported in delivering peptides such as antinflamin 1 by iontophoresis (58). It has also been reported recently that proteins as large as insulin (~6000 Da), interferon γ (~17,000 Da), and erythropoietin (~30,400 Da) could be delivered

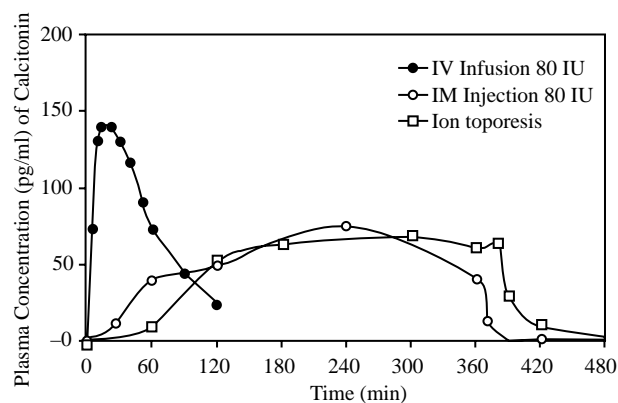


Fig. 4 Mean plasma calcitonin concentration (pg/ml)-time curve resulting from 6-h iontophoretic delivery ($n = 12$) of a calcitonin analog peptide in healthy male volunteers. Plasma levels obtained from IV infusion for 6 h (80 IU, $n = 8$) and IM injection (80 IU, $n = 11$) of the peptide are also shown. Plasma concentration time-profiles indicate that steady-state plasma levels are achieved rapidly, and the decline in plasma levels appears to be absorption-limited rather like an i.m. injection. However, short lag times are observed, which are consistent with diffusion across the outermost layers of the skin but shorter than those typically associated with passive transdermal diffusion. The active and placebo iontophoretic treatment patches were applied at a current density of $200 \mu\text{A}/\text{cm}^2$ for a 6-h period. Blood samples were removed for pharmacokinetic assessment at the start of dosing and at regular intervals during and after treatment. Plasma samples were assayed using an RIA method. (From Ref. 56.)

across the skin at therapeutic concentrations with the aid of low-frequency ultrasound (58).

ORAL DELIVERY STRATEGIES

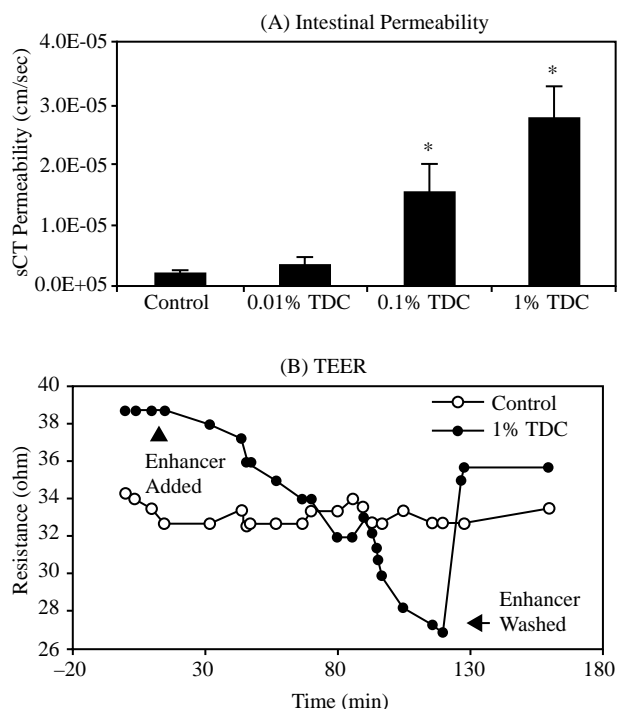
The remainder of this chapter focuses on strategies for improving the oral bioavailability of protein pharmaceuticals and, in particular, describing our experiences with salmon calcitonin. sCT is an endogenous polypeptide hormone composed of 32 amino acids that plays a crucial role in both calcium homeostasis and bone remodeling (59, 60). Four forms of CT are used clinically, namely synthetic human CT, synthetic salmon CT, natural porcine CT, and a synthetic analog of eel calcitonin. Currently, CT is administered parenterally or nasally (61, 62). To effectively inhibit the manifestations of metabolic bone disorders such as Paget's disease and osteoporosis, a frequent and relatively high dosage of CT is administered (63). The oral route is a preferred route of administration, considering the chronic nature of CT therapy. However,

because of extensive proteolytic degradation in the GI lumen and low intrinsic intestinal membrane permeability, insufficient oral bioavailability (BA) of CT necessitates the use of high doses (4000 to 6000 IU/mg) of sCT, even though sCT is 20–30 times more potent than hCT (150–200 IU/mg) (64–66). The unique structure of sCT protects it against its sequestration and metabolism in the liver (67). In this chapter, several approaches for enhancing the oral absorption of sCT are presented. sCT delivery systems for the treatment of osteoporosis are reviewed for many routes including nasal, transdermal, ocular, oral, bronchial, rectal, and vaginal administration (68).

Permeation Enhancement

The poor intrinsic permeability of proteins across biological membranes is well documented (1, 2, 4, 69) and can generally be attributed to their hydrophilic nature and large molecular size. Membrane carrier systems that facilitate the absorption of small peptides (di- and tripeptides) are not efficient at transporting larger peptides and proteins. Permeation enhancers have received considerable attention in attempts to modify the basic barrier properties of the intestinal epithelial cell membrane. Various classes of formulation additives including bile acids, salicylates, fatty acids, acylcarnitines, surfactants, medium chain glycerides, and chelating agents have been studied as absorption enhancers. Each of these agents exerts its enhancing effects by different mechanisms, and each has been associated with adverse effects (70).

Recently, we published in vitro and in vivo results of the evaluation of the performance of formulation additives to enhance the intestinal uptake of sCT (71). The effect of formulation additives on sCT effective permeability (P_{eff}) and transepithelial electrical resistance (TEER) was evaluated in side-by-side diffusion chambers using rat intestinal segments. Various additives such as sodium taurodeoxycholate (TDC), sodium taurocholate, sucrose stearate, sucrose ester-15, Tween 80, lauroyl carnitine chloride (LCC), myristoyl carnitine chloride, cetyl pyridinium chloride, and cetrimide were evaluated in rat jejunum at concentrations ranging from 0.01 to 1%. The effective permeability of sCT was greatest in the presence of TDC ($2.73 \pm 0.54 \text{ E-5 cm/s}$), increasing up to 14 times over the control ($1.91 \pm 0.45 \text{ E-6 cm/s}$) in a concentration dependent manner (Fig. 5A). The permeability enhancement relative to control was 5 times for LCC, 3.9 times for Tween 80, 3.2 times for sodium taurocholate, 2.6 times for myristoyl carnitine chloride, 2.3 times for sucrose stearate, 2.0 times for cetyl pyridinium chloride, 1.7 times for sucrose ester-15, and 1.3 times for cetrimide at 1% additives. The order of enhancement on the basis of



EC_{50} values (concentration of enhancer for 50% enhancement) was TDC, LCC > sodium taurocholate > tween 80 > myristoyl carnitine chloride > sucrose stearate > sucrose ester-15 > cetyl pyridinium chloride > cetrimide. After the exposure to various additives such as TDC, LCC, and cetrimide at concentration ranging from 0.1 to 1%, the TEER was reduced in a concentration- and contact time-dependent manner for all additives, whereas the TEER increased when the tissue was washed free of additives (Fig. 5B). When 1% formulation additives were washed after approximately 100 min exposure, the TEER returned to 92% of the initial value for TDC and approximately 80% of initial value for the others. As demonstrated during the in vitro and in vivo evaluation of formulation additives, the oral absorption of protein drugs can be maximized by balancing permeability increases and the toxic effects associated with the use of enhancers.

The effect of formulations on the intestinal bioavailability of sCT was studied in an intestinal and vascular access port (IVAP) dog model (71). A schematic representation of the IVAP dog with four ports in the duodenum (ID), ileum (IL), colon (IC), and portal vein (PV) is shown in Fig. 6. The regional (jejunal and ileal) bioavailabilities from formulations DDS1 and DDS2 [DDS1 and DDS2 contained sCT, citric acid (CA), and lauroyl carnitine chloride (LCC) or sodium taurodeoxycholate (TDC), respectively] were significantly

Fig. 5 (A) Plot of effective permeability (sm/s) of sCT in rat jejunum (mean \pm SD; $n = 3$). Effect of TDC was investigated at various concentrations ranging from 0.01 to 1%. * indicates the significant difference from control by $P < 0.05$. Side-by-side diffusion chambers were used. The exposed tissue surface area was 0.636 cm^2 and the volume of each half-chamber was 1.7 ml. Mixing in the chambers was controlled using a gas lift mechanism. The temperature was maintained at 37°C throughout the experiment. A 1.5- to 2-cm strip of rat intestinal tissue was excised from the animals, rinsed free of luminal contents using Ringer's buffer (pH 7.4), and mounted onto a diffusion half-chamber maintained at 37°C . Tissues were bathed in 15 mM Mes Ringer's buffer containing $50 \mu\text{M}$ of sCT with or without formulation additives on the mucosal side (pH 5, 290 mOsm/kg) and Ringer's buffer without additives on the serosal side (pH 7.4, 290 mOsm/kg). Small aliquots (0.5 ml) were taken from the serosal chamber at 30, 45, 60, 75, 90, and 105 min and analyzed by RIA. The effective permeabilities were calculated from the experimental data using the following equation based on Fick's First Law: $P_{\text{eff}} = (V_r/A \cdot C_0) \cdot dC/dt$ where V_r is the volume of the receiver chamber, A is the absorbing surface, C_0 is the initial drug concentration in the donor (mucosal) phase, and dC/dt is the change in drug concentration in the receptor (basolateral) phase per unit of time. In the presence of TDC, the effective permeability of sCT ($2.73 \pm 0.54 \text{ E-5 cm/s}$) was increasing up to 14 times over the control ($1.91 \pm 0.45 \text{ E-6 cm/s}$) in a concentration-dependent manner. (B) Transepithelial electrical resistance (TEER, ohm) versus time curves with or without 1% TDC in rat jejunum. Chambers connected to a voltage-current clamp for measuring transepithelial electrical resistance (TEER) of the intact tissue. The exposed tissue surface area was 1.5 cm^2 , and the volume of each half-chamber was 7 ml. Rat intestinal tissue was mounted in chambers, and then two sets of Ag/AgCl electrodes were connected to the voltage-clamp system to pass the current through the membrane. Tissues mounted in chambers usually required approximately 20 min for reaching temperature and TEER equilibrium. Experiments with formulation additives were not initiated until this time. The mucosal strips were then exposed to formulation additives at concentrations ranging from 0.01 to 1% at pH 4 Mes Ringer's buffer. After the initial equilibrium period, TEER was measured and subsequently recorded at regular time intervals. To determine the reversibility of additive effects on TEER, the buffer was replaced with fresh buffer without additives, and the TEER was monitored until a stable reading was observed. When 1% formulation additives were washed after an approximate 100-min exposure, the TEER was returned to 92% of initial value for TDC. (From Ref. 71.)

greater than their respective controls. Bioavailability enhancement ranged from 98 to 337% for all treatments and regions studied. Compared which sCT alone (without CA), bioavailability enhancement ranged from 1220 to 3070% for DDS1 or DDS2. The in vivo augmentation of sCT absorption by DDS1 or DDS2 correlated well to the in vitro permeation enhancement (Fig. 7).

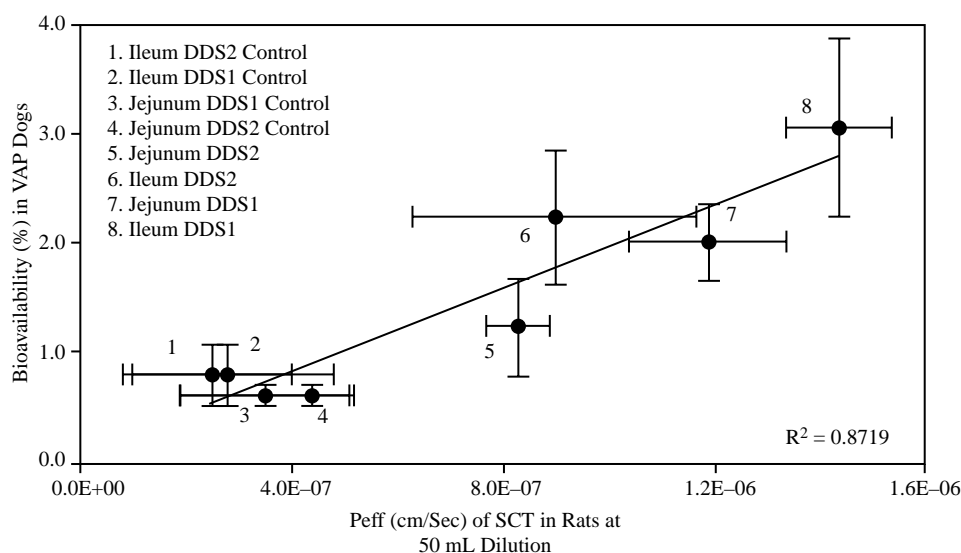


Fig. 7 The plot of in vitro effective permeability (P_{eff}) of sCT in rats (mean \pm SD; $n = 3$) versus in vivo absolute bioavailability of sCT formulation in intestinal and vascular access port (IVAP) dogs (mean \pm SD; $n = 5$ to 6). The in vitro diffusion study was carried out in side-by-side diffusion chambers at concentrations equivalent to the dilution of sCT control and DDS formulation ingredients in 50 ml of buffer solution. DDS1 contains sCT, CA, and LCC, and DDS2 contains sCT, CA, and TDC. The in vivo intestinal bioavailability experiment of formulations was carried out in an IVAP dog model at concentrations of sCT and DDS additives in 5 ml of buffer solution. The ports for IVAP infusions were accessed transcutaneously with a 22-G Huber needle. It was observed that formulations containing TDC or LCC had a significantly higher effective permeability compared with their respective controls. The enhancement in the ileal segment was significantly higher than in the jejunal segment for DDS1 and DDS2. Bioavailability enhancement ranged from 98 to 337% for all treatments and regions studied. The in vivo augmentation of sCT absorption by DDS1 or DDS2 correlated well to the in vitro permeation enhancement. (Reproduced from Ref. 71.)

Protection from Enzymatic Degradation

The susceptibility of proteins to enzymatic attack is well known and remains a major challenge of oral delivery (1, 3, 4, 72–73). Yamamoto et al. (74) reported that various protease inhibitors, including sodium glycocholate, camostat mesilate, and bacitracin, can increase the glucose-lowering activity of insulin administered to rat intestine and colon, apparently by inhibiting protease activity in the lumen and mucous layer of the intestinal tissue. The degradation of calcitonin is also decreased in the presence of protease inhibitors such as camostat and aprotinin (75). The mucoadhesive polymers polycarbophil and carbomer, approved by the U.S. FDA, may have potential for protecting peptides from tryptic degradation by immobilizing trypsin under the depletion of Ca^{2+} (76).

Another potential approach to minimize the activity of intestinal enzymes includes adjusting the pH of the intestinal contents to the corresponding pH minima for proteolytic enzyme activity. Several laboratories have shown that proteolytic activity against insulin, calcitonin, and insulin-like growth factor-I was completely inhibited by pH-lowering mechanisms using polyacrylic acid

polymer (16, 17). We recently used traditional pharmacokinetic techniques combined with radiotelemetric measurement of intestinal pH to elucidate the effect of pH modulation on the oral absorption properties of sCT (77). Studies were performed to characterize the disintegration of the formulation, intestinal pH changes, and the appearance of the peptide in the blood. Enteric-coated formulations containing sCT and various amounts of citric acid (CA) were tethered to a Heidelberg capsule (HC) and given orally to normal beagle dogs (Table 2).

Table 2 The composition of sCT enteric capsules tested in beagle dogs

Formulations	sCT (mg)	CA (mg)	LCC (mg)	Talc (mg)	Dextrose (mg)
97B	1.11	0	55.2	55.2	552.4
97C	1.20	145.7	54.7	54.7	400.7
97D	1.15	260.2	51.9	52.2	260.5
97E	1.19	565.1	56.3	56.3	0

CA; citric acid; LCC, lauroyl carnitine chloride.

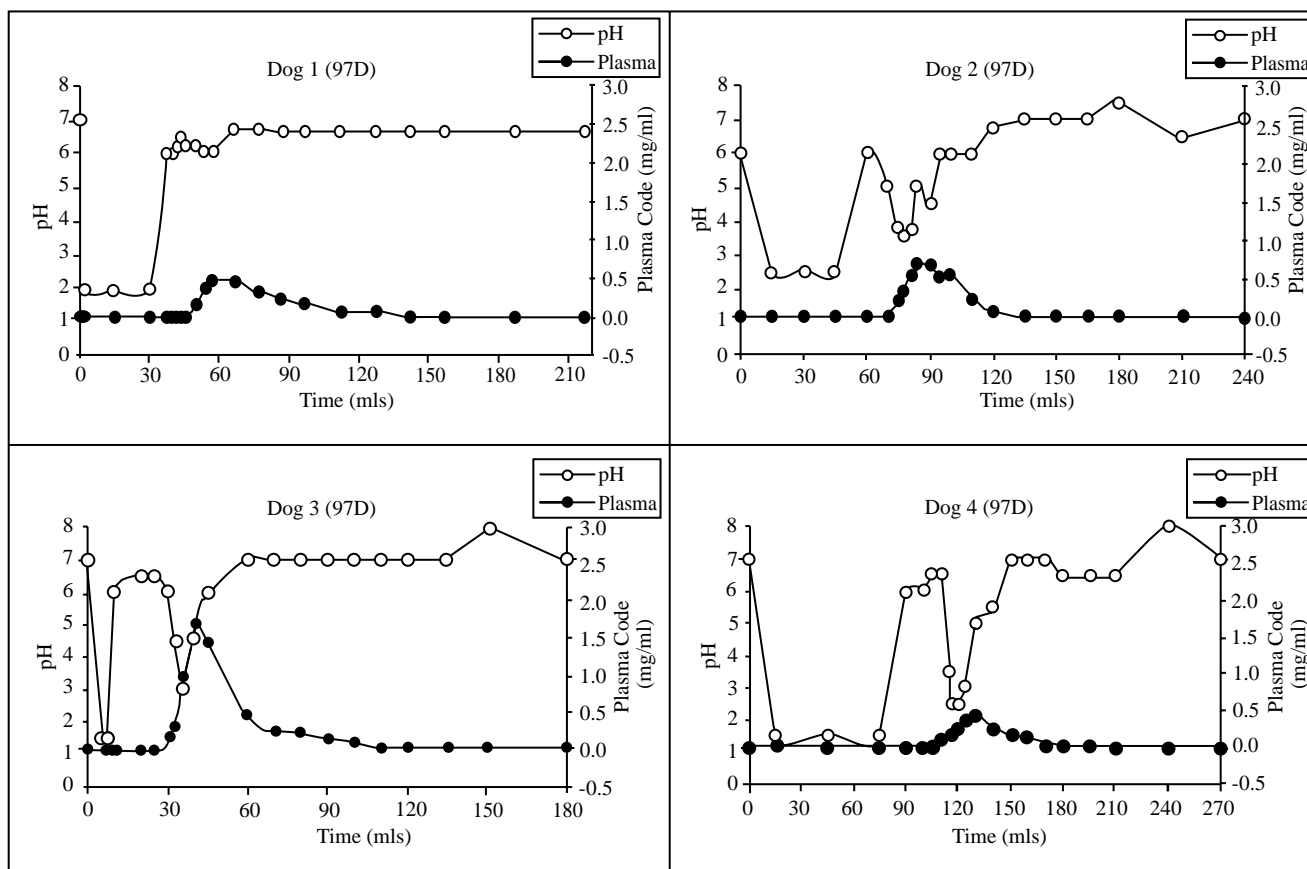


Fig. 8 The pH and plasma sCT concentration-time profiles after oral administration of formulation 97D (containing 1.15 mg of sCT, 260.2 mg of CA, 51.9 mg of LCC, 52.2 mg of Talc, and 260.5 mg of dextrose) (Table 2) tethered to a Heidelberg capsule in normal beagle dogs. Continuous determination of pH was accomplished using a radiotelemetric device, the Heidelberg capsule (HC). The device consists of a battery-operated, high-frequency radio transmitter and a pH electrode housed in a nondigestible acrylic capsule 7 mm in diameter and 20 mm in length. The dogs wore an antenna strapped around the body to receive the radio signal that was then recorded on a chart recorder. The capsule battery was activated with normal saline and calibrated in pH 1 and pH 7 buffer solutions maintained at 37°C. The HC was then tethered to the drug capsule using surgical thread (3-0 vicryl) and administered orally to dogs. Because pH values change with location within the gut and the drug capsule dissolution, alterations in pH were interpreted to be indicative of the movement of the HC-drug capsule through the different segments until the drug capsule dissolves. Generally, Heidelberg capsules provide readings with ± 0.5 pH unit accuracy and excellent in vivo reproducibility in the pH range of 1–8 for 22 h after activation. Four male beagle dogs were used to monitor the disintegration and oral absorption of sCT formulation 97D. An HC tethered to an enteric capsule was given orally with 10 ml of water to each dog. Two baseline blood samples were drawn before to dosing and another immediately after gastric emptying (GE). From the time the capsules enter the small intestine, blood samples were taken every 10 min until the HC showed a drop in pH, which signified the disintegration of the test capsule. More frequent blood sampling was performed from the time that disintegration was first detected, with blood samples taken at 3, 6, 9, 12, 15, 20, 30, 45, 60, 75, 90, 120, 150, and 180 min. The peak plasma concentrations of sCT were always observed when the intestinal pH declined. (Reproduced from Ref. 77.)

Blood samples were collected and analyzed using radioimmunoassay (RIA). Intestinal pH was continuously monitored using the HC system. The intraindividual variation in gastric emptying (GE) of the delivery system was large. There were large interindividual differences in the disintegration and absorption properties. However, the peak plasma concentrations of sCT were always observed

when the intestinal pH declined (Fig. 8). The intestinal pH decrease was not observed when CA was not included, but an intestinal pH decrease was obvious in all formulations that included CA. The intestinal pH was significantly affected by the amount of CA in the formulations. Plasma concentrations of sCT were observed in all formulations that included CA, but plasma concentrations of sCT were

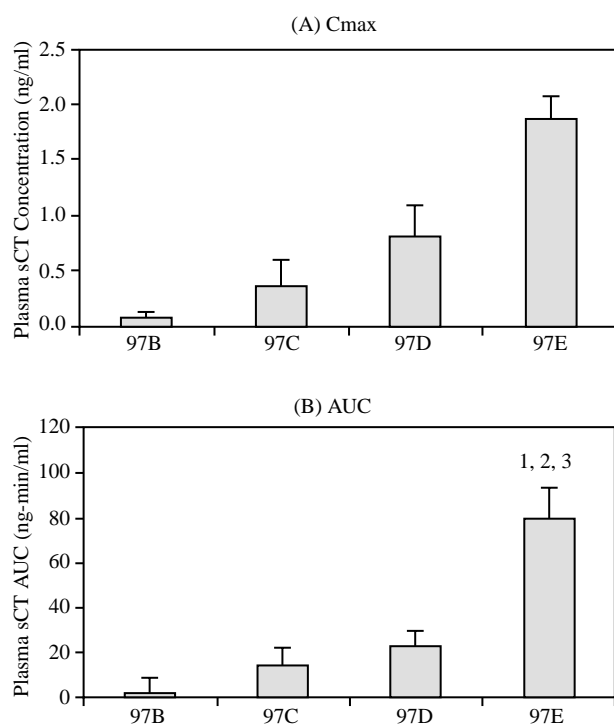


Fig. 9 The plasma C_{\max} (A) and AUC (B) of sCT after oral administration of various formulations (Table 2) in normal beagle dogs. Data are expressed mean \pm SEM ($n = 3$ to 4). Formulation 97E was significantly different from 97B(1), 97C(2), and 97D(3) by $P < 0.05$ using one-way ANOVA. By increasing the amount of CA in the formulation, the oral absorption of sCT increased gradually. (Reproduced from Ref. 77.)

not observed in formulation 97B in which CA was not included (Table 2). There was a good correlation between the time to reach the trough intestinal pH ($t_{\text{pH,min}}$) and time to reach the peak plasma concentration ($t_{\text{conc,max}}$) of sCT ($t_{\text{conc,max}} = 0.95 \times t_{\text{pH,min}} + 14.1$, $n = 11$, $r^2 = 0.91$). As a consequence, the intestinal pH decrease caused by CA appears to be critical for the oral absorption of sCT. By increasing the amount of CA in the formulation, the oral absorption (C_{\max} and AUC) of sCT increased gradually (Fig. 9). These results indicate that the oral absorption/or enhancement of sCT absorption is directly related to the stabilization of sCT by a reduction in intestinal pH. The pH stability of pancreatic trypsin (human) is optimal at pH 5–6. At pH 4.0, approximately 45% of the activity remained, whereas 15% of activity was retained at pH 3.5 (78). As evidenced in this study, reducing intestinal pH resulted in a significant improvement in sCT absorption.

Another approach to providing protection against proteolytic attack, rather than enzyme inhibition, has been to protect proteins in the physical environment

of the formulation itself. In recent years, significant efforts have been directed toward formulating proteins in microemulsions, small particles (e.g., nanoparticles), and bioadhesive particles (31, 32). The rationale in all three cases is often similar: protection of proteins from the intestinal environment before absorption and localization of the protein at or near the cellular membrane to optimize the driving force for passive permeation. Pegylation of a therapeutic protein may be a suitable form of the protein

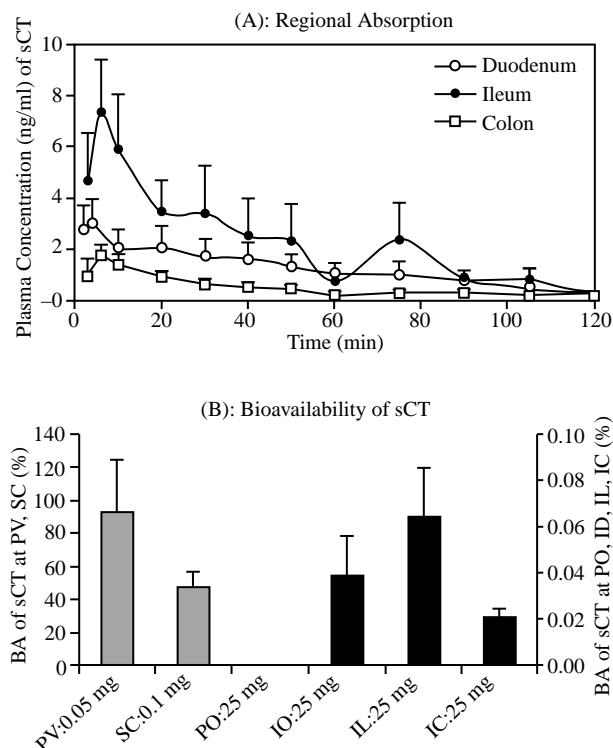


Fig. 10 (A) Plasma concentration (mean \pm SEM) of sCT versus time curves after bolus duodenal, ileal, and colonic administration in intestinal and vascular Access port (IVAP) dogs ($n = 5$ to 6). The ports for IVAP infusions were accessed transcutaneously with a 22-G Huber needle. The sCT formulation (25 mg/5 ml/dog) was infused rapidly (12 ml/min), and the port was cleared with a final flush of 1 ml of sterile water. Blood samples were drawn at 1, 3, 6, 9, 12, 15, 20, 30, 45, 60, 75, 90, 120, 150, 180, and 240 min. Absorption of sCT from the ileum was better than from the other regions studied. (Reproduced from Ref. 71.) (B) Bioavailability of sCT (mean \pm SEM) after portal venous (pv 50 μ g), subcutaneous (sc 100 μ g), and oral (po), duodenal (ID), ileal (IL), and colonic (IC) administration of 25 mg sCT in IVAP dogs ($n = 4$ to 6). The BA (mean \pm SD) of unformulated sCT was $92.8 \pm 32.0\%$ for portal vein, $47.5 \pm 9.3\%$ for subcutaneous, 0% for oral, $0.039 \pm 0.017\%$ for duodenal, $0.064 \pm 0.022\%$ for ileal, and $0.021 \pm 0.004\%$ for colon. The hepatic extraction of sCT is negligible in IVAP dogs.

in an oral-delivery formulation (79). Recombinant human granulocyte colony stimulating factor (PEG-gCSF) resulted in an increase in stability and in retention of in vivo bioactivity when administered by the intraduodenal route. Another reported approach uses mucoadhesion of nanoparticles having surface hydrophilic polymeric chains (80). In that report, it was observed that there was a good correlation between mucoadhesion and enhancement of sCT absorption in rats. The GI transit rates of nanoparticles having surface poly(*N*-isopropylacrylamide), poly(vinylamine), and poly(methacrylic acid) chains were reduced, and sCT absorption was improved. Temperature and pH-sensitive polymers for hCT delivery are also reported (81). Stimuli-sensitive polymers are suitable candidates for oral delivery vehicles because they will prevent gastric degradation in the stomach while providing a controlled release of a peptide drug. The beads made of the polymers with a high content of acrylic acid (most hydrophilic) provided better loading, stability, and release of hCT. In vivo biological activity of the released hCT was preserved.

Maintaining High Local Drug Concentrations

The extent of protein degradation by intestinal enzymatic attack is concentration-dependent following Michaelis–Menten kinetics. Because regional drug concentrations of orally administered proteins depend on intestinal spreading and dilution patterns (19), modulating regional drug concentrations may alter the oral absorption of proteins. Because sCT is an excellent substrate for the pancreatic serine protease trypsin, the rate of degradation of sCT in the GI lumen is dependent on the concentration of sCT in the intestinal lumen. If degradation was the controlling factor, sCT absorption would be much more affected at low sCT concentrations (i.e., the kinetics are in the first-order region of the Michaelis–Menten curve). To evaluate the intestinal dilution and spreading on oral sCT BA, slow infusion (2 ml/min) and high dilution (25 mg/20 ml/dog) treatments were evaluated in dogs in vivo (71). In IVAP studies, the slower and larger infusion volume resulted in significantly lower sCT absorption, demonstrating that intestinal dilution and spreading significantly affected the oral BA of sCT. Scott-Moncrieff et al. (82) reported increased insulin absorption after direct jejunal administration in dogs of a 30-mM sodium glycocholate and 40-mM linoleic acid mixed micelle formulation, although the apparent bioavailability was only 1.8%. This same formulation approach elicited 41% insulin bioavailability in a rat loop model, and the researchers proposed that the much-reduced effect in dogs was possibly attributable to

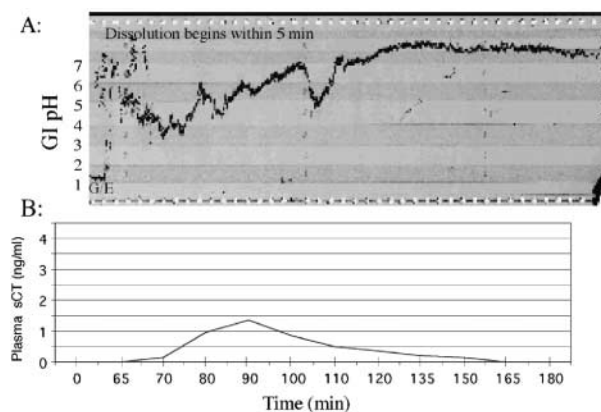


Fig. 11 Duodenal pH recovery (A) and sCT absorption (B) profiles when enteric capsule disintegrated at the duodenum after oral administration of an sCT formulation containing 565 mg of citric acid and 1.2 mg of sCT into the normal dogs. An enteric coated capsule was tethered to a HC with a 3-0 silk suture and given orally to normal dogs with 10 ml of water. Two baseline samples were drawn before dosing and another immediately after gastric emptying. From the time the capsules enter the small intestine (SI), blood samples were taken every 10 min. When the HC showed a drop in pH, indicating the disintegration of the test capsule, more frequent blood sampling was performed (3, 6, 9, 12, 15, 20, 30, 45, 60, 75, 90, 120, 150, and 180 min). Transit time was used to estimate the SI location being monitored. Capsule disintegration in the duodenum begins 5 min after gastric emptying (GE) at 65 min. The initially variable pH momentarily reached pH 3 and soon rose above pH 5. (Reproduced from Ref. 88.)

dilution and spreading of the formulation, resulting in a reduced concentration of insulin, which increased exposure of insulin to proteolytic enzymes.

Regiospecific Targeting

The regiospecific difference of intrinsic permeability of proteins and proteolytic enzymatic activity is well documented (18). Insulin absorption was greater in the ileum and large intestine than in the jejunum using an in situ loop method (83). Several studies have indicated reduced enzymatic activity in the distal intestine (72). It has been hypothesized that the colon could be an optimal site for the absorption of peptides and proteins from the GI tract. This hypothesis is based on the fact that the colon contains little or no digestive enzymes, therefore, the inherent stability of polypeptide materials should be higher. Although the distal small intestine has less luminal and apical proteolytic activity, it has high activities of some apical peptidases (84). Colonic contents showed high degradation of insulin and calcitonin with high chymotrypsin activity, which suggests that care should be

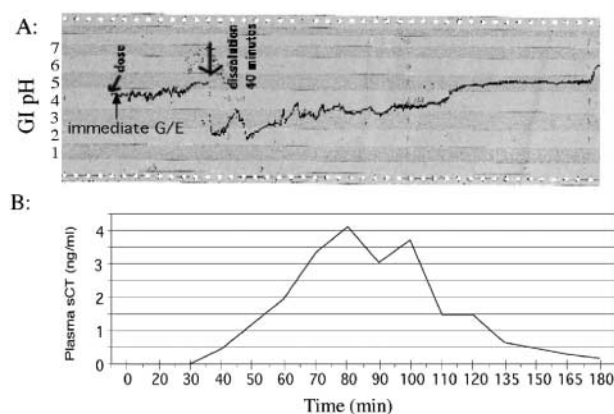


Fig. 12 Late jejunal/or ileal pH recovery (A) and sCT absorption (B) profiles when enteric capsule disintegrated at late jejunum after oral administration of an sCT formulation containing 565 mg of citric acid and 1.2 mg of sCT into the normal dogs. This figure shows a 40-min delay between GE (less than 5 min) and disintegration. The observed pH drop was sharper and steadier, with the pH remaining below 5 for nearly 90 min. The better absorption profile occurs when disintegration begins in the lower small intestine (LSI). (Reproduced from Ref. 88.)

taken when administering protein drugs to the large intestine for colon-specific drug delivery (75).

Recently, our group published baseline PK information on the regional dependence of unformulated sCT oral absorption evaluated by direct administration into the PO, ID, IL, and IC by means of surgically implanted, chronic catheters (71) (Fig. 6). The plasma concentration and bioavailability values of sCT in dogs are shown and compared with p.v. and s.c. routes in Fig. 10. Plasma sCT concentrations after a p.o. dose were below the LOQ (80 pg/ml), and this may be attributed to the extensive gastric and/or intestinal enzymatic degradation. Absorption of sCT after ID administration was rapid, with peak concentrations occurring within approximately 10 min. BA of sCT was $0.039 \pm 0.017\%$, $0.064 \pm 0.022\%$, and $0.021 \pm 0.004\%$ for ID, IL, and IC administration, respectively, in IVAP dogs. Compared with that in other regions, the low BA of sCT in the colon is probably related to the combined effects of poor membrane permeability and/or proteolytic degradation by microorganisms specifically residing in the colon. The colonic results in dogs suggest that additional formulation efforts may be required to successfully deliver sCT to the colon. Calcitonin was rapidly degraded in cecal supernatants by microorganisms (85). Interestingly, the rational design of colon-specific protein/peptide delivery systems has been based on two premises: 1) that the colon compartment has decreased pancreatic proteolytic activity, and/or 2) the colon

possesses bacterial activities with unique specificities for polymer targeting (86, 87).

The successful oral delivery of sCT also depends on intestinal spreading and dilution patterns (19). As a consequence, modulating regional sCT concentrations and/or pH using a formulation approach may alter the oral absorption of CT. To understand the oral absorption process, there is a need to define both the location of release of the delivery system's contents and its effective transit time through the gut. Recently, our group published results using a radiotelemetric measurement of intestinal pH, radiographic visualization of intestinal tract, and pharmacokinetics to investigate how upper small intestine (USI) and lower small intestine (LSI) react differently in intestinal spreading and pH recovery in conscious IVAP beagle dogs (88). Regional intestinal differences of spreading and pH recovery were studied. One port was placed in the duodenum and a second port in the ileum. Fluoroscopy and Heidelberg studies were performed to characterize the intestinal spreading and pH recovery. A radiopaque dye and CA were infused into the ports, and a radiopaque powder capsule containing CA was given orally. Fluoroscopy clearly showed that when the radiopaque dye was infused into the duodenum and capsule disintegration occurred early, there was significant dilution and spreading of the excipients throughout a large section of upper small intestine. Once mixed, the contents moved slowly down the GI tract. However, when the radiopaque dye was infused into the ileum and capsule disintegration occurred lower down, the excipients moved along as a bolus (i.e., a plug). Hence, the local exposure of the intestinal wall was more concentrated in the lower small intestine. The results of the pH monitoring concurred with those of the fluoroscopy studies. In the duodenum, the pH dropped only momentarily then rose quickly. However, steady pH lowering and slower recovery were recorded in the lower small intestine. To investigate the regional intestinal differences of spreading and pH recovery and their impact on sCT oral absorption, normal male beagle dogs were dosed with an enteric capsule containing 565 mg of CA and 1.2 mg of sCT (88). Regular blood samples were collected and analyzed using RIA to determine the absorption characteristics of sCT. Figs. 11 and 12 show two extreme cases demonstrating how the same formulation can produce different results depending on where it begins to disintegrate. Because sCT is protected by low pH conditions (78) and is readily absorbed when concentrations are high, the best absorption occurs when disintegration begins lower in the small intestine. Fig. 11 depicts the disintegration of the delivery system in the duodenum, beginning 5 min after GE (65 min). The initially variable pH momentarily reaches pH 3 and soon

risers above pH 5. Fig. 12 shows a 40-min delay between GE (less than 5 min) and disintegration. The pH drop was sharper and more steady. The pH remained below 5 for nearly 90 min. The C_{\max} and AUC were 1.3 ng/ml and 47 ng·min/ml, respectively, for duodenal disintegration (Fig. 11), whereas the values were 4.1 ng/ml and 234 ng·min/ml, respectively, for late jejunal/or ileal disintegration (Fig. 12). These results show the regiospecific intestinal absorption of sCT and how it relates to the effective transit time (i.e., disintegration time after GE) of the delivery system in the small intestine. Plasma levels of sCT were optimal when disintegration occurred in the mid to lower SI.

Chemical Modification

Peptide and proteins often possess physical and/or chemical properties that present significant stability problems not encountered with many small, organic drug molecules. Because of the complex nature of proteins, self-aggregation is always a concern in formulation efforts. The tendency of insulin to form hexamers is well documented, and the absorption of hexamers will most likely be very different from monomer absorption. Hovgaard et al. (89) reported the use of alkyl saccharide surfactants (e.g., dodecyl maltoside) to minimize insulin aggregate. The insulin–dodecyl maltoside complex also afforded some protection against enzymatic degradation. Human calcitonin is also known to self-organize into fibrillar structures with reduced biologic activity (90). The use of various surfactant approaches to maximize monomer concentration during protein release may afford advantages in minimizing the size of the complex that must cross epithelial cell layers. Another potential oral delivery approach is the chemical modification of proteins to produce prodrugs and analogs. It is plausible that this approach may protect protein against degradation by proteases and other enzymes present at the mucosal barrier and renders protein more lipophilic, resulting in increased BA. Using chemical modification with fatty acids (20, 21) or *N*-acylated α -amino acids (22), a significant increase in CT intestinal absorption was observed in comparison with the native CT. The stability and permeability of peptides were improved by acylation with fatty acids and the derivatized amino acids only weakly inhibited by trypsin or leucine aminopeptidase. The stability of peptides in gastrointestinal fluids and serum was improved by substituting D-amino acids (23, 24). As exemplified by cyclosporine, the cyclic 11-amino acid compound has a surprisingly high bioavailability by chemical modification to improve its stability within the GI tract and to make it more lipophilic to enhance membrane permeability. When

administered in the microemulsion formulation, 50% or more of this molecule can be absorbed from the GI tract (11, 12, 91). Chemical modification of small peptides has been successful in protecting certain peptide structures from enzymatic attack without significant loss of biological activity, but less success has been achieved with larger polypeptides.

CONCLUDING REMARKS

Establishing an oral delivery system for peptides and protein drugs is of great importance because parenteral administration results in poor patient compliance during chronic treatment, resulting in limited clinical utility. The advantages of oral delivery systems in terms of patient compliance and acceptability are further augmented by the potential cost savings because noninvasive delivery routes do not require sterile manufacturing, and administration can be effected without direct involvement of a healthcare provider. The clinical development of protein pharmaceuticals, however, has been impeded because of poor absorption across biological membranes and rapid proteolytic degradation that typically result in low bioavailabilities. Although only very limited success (e.g., cyclosporine) has been achieved in developing and marketing oral peptide delivery systems, interest remains extremely high. It is clear that a well controlled and rational formulation design process is necessary. The problems with protein delivery are not trivial and will not be overcome by trivial solutions. Because the barriers to protein absorption (permeability, enzymatic degradation, presystemic hepatic degradation, and chemical and physical stability) will likely exhibit significant protein specificity, delivery systems will have to be investigated and developed for efficacy and safety on a case-by-case basis. In this chapter, we focused on enhancing oral absorption using several approaches. Potential approaches to modulate the intestinal environment include modulating intestinal permeability, limiting the activity of intestinal enzymes, maintaining high local drug concentrations, and regiospecific targeting. Using these strategies, sCT oral delivery systems have been fabricated and successfully tested in humans.

REFERENCES

1. Zhou, X.H. Overcoming Enzymatic and Absorption Barriers to Non-Parenterally Administered Protein and Peptide Drugs. *J. Controlled Release* **1994**, 29, 239–252.

2. Amidon, G.L.; Lee, H.J. Absorption of Peptide and Peptidomimetic Drugs. *Annu. Rev. Pharmacol. Toxicol.* **1994**, *34*, 321–341.
3. Lagguth, P.; Bohner, V.; Heizmann, J.; Merckle, H.P.; Wolfram, S.; Amidon, G.L.; Yamashita, S. The Challenge of Proteolytic Enzymes in Intestinal Peptide Delivery. *J. Controlled Release* **1997**, *46*, 39–57.
4. Lee, V.H.L.; Yamamoto, A. Penetration and Enzymatic Barriers to Peptide and Protein Absorption. *Adv. Drug Delivery Rev.* **1990**, *4*, 171–207.
5. Bai, J.P.; Amidon, G.L. Structural Specificity of Mucosal-Cell Transport and Metabolism of Peptide Drugs: Implications for Oral Peptide Drug Delivery. *Pharm. Res.* **1992**, *9*, 969–978.
6. Pettit, D.K.; Gombotz, W.R. The Development of Site-Specific Drug-Delivery Systems for Protein and Peptide Biopharmaceuticals. *Trends Biotechnol.* **1998**, *16*, 343–349.
7. Wearley, L.L. Recent Progress in Protein and Peptide Delivery by Noninvasive Routes. *Crit. Rev. Ther. Carrier Syst.* **1991**, *8*, 331–394.
8. Dantzig, A.H. Oral Absorption of β -Lactams by Intestinal Peptide Transport Proteins. *Adv. Drug Delivery Rev.* **1997**, *23*, 63–76.
9. Tsuji, A.; Tamai, I. Carrier-Mediated Intestinal Transport of Drugs. *Pharm. Res.* **1996**, *13*, 963–977.
10. Lee, V.H.; Chu, C.; Mahlin, E.D.; Basu, S.K.; Ann, D.K.; Bolger, M.B.; Haworth, I.S.; Yeung, A.K.; Wu, S.K.; Hamm-Alvarez, S.; Okamoto, C.T.J. Biopharmaceutics of Transmucosal Peptide and Protein Drug Administration: Role of Transport Mechanisms with a Focus on the Involvement of PepT1. *J. Controlled Release* **1999**, *62*, 129–140.
11. Noble, S.; Markham, A. Cyclosporin. A Review of the Pharmacokinetic Properties, Clinical Efficacy and Tolerability of a Microemulsion-Based Formulation (Neoral). *Drugs* **1995**, *50*, 924–941.
12. Schroeder, T.J.; Cho, M.J.; Pollack, G.M.; Floc'h, R.; Moran, H.B.; Levy, R.; Moore, L.W.; Pouletty, P. Comparison of Two Cyclosporine Formulations in Healthy Volunteers: Bioequivalence of the New Sang-35 Formulation and Neoral. *J. Clin. Pharmacol.* **1998**, *38*, 807–814.
13. Neutra, M.R.; Kraehenbuhl, J.P. Transepithelial Transport of Proteins by Intestinal Epithelial Cells. *Biological Barriers to Protein Delivery*; Audus, K.L., Raub, T.J., Eds.; Plenum Press: New York, 1993; 107–129.
14. Bernkop-Schnurch, A. The Use of Inhibitory Agents to Overcome the Enzymatic Barrier to Preorally Administered Therapeutic Peptides and Proteins. *J. Controlled Release* **1998**, *52*, 1–16.
15. Friedman, D.I.; Amidon, G.L. Oral Absorption of Peptides: Influence of pH And Inhibitors on the Intestinal Hydrolysis of Leu-Enkephalin and Analogues. *Pharm. Res.* **1991**, *8*, 93–96.
16. Bai, J.P.; Chang, L.L.; Guo, J.H. Effects of Polyacrylic Polymers on the Luminal Proteolysis of Peptide Drugs in the Colon. *J. Pharm. Sci.* **1995**, *84*, 1291–1294.
17. Bai, J.P.; Chang, L.L.; Guo, J.H. Effects of Polyacrylic Polymers on the Degradation of Insulin and Peptide Drugs by Chymotrypsin and Trypsin. *J. Pharm. Pharmacol.* **1996**, *48*, 17–21.
18. Woodley, J.F. Enzymatic Barriers for GI Peptide and Protein Delivery. *Crit. Rev. Ther. Drug Carrier Syst.* **1994**, *11*, 61–95.
19. Grundy, D. *Gastrointestinal Motility*; MTD Press Limited: 1985.
20. Yamamoto, A. Improvement of Intestinal Absorption of Peptide and Protein Drugs by Chemical Modification with Fatty Acids. *Nippon Rinsho* **1998**, *56*, 601–607.
21. Muranishi, S. Delivery System Design for Improvement of Intestinal Absorption of Peptide Drugs. *Yakugaku Zasshi* **1997**, *117*, 394–414.
22. Leone-Bay, A.; Santiago, N.; Achan, D.; Chaudhary, K.; DeMorin, F.; Falzarano, L.; Haas, S.; Kalbag, S.; Kaplan, D.; Leipold, H.; Lercara, C.; O'Toole, D.; Rivera, C.; Rosado, C.; Sarubb, D.; Vuocolo, E.; Wana, N.; Milstein, S.; Baughman, R.A. *N*-acylated Alpha-Amino Acids as Novel Oral Delivery Agents for Proteins. *J. Med. Chem.* **1995**, *38*, 4263–4269.
23. Krongdahl, E.; Orzechowski, A.; Ekstrom, G.; Lennernas, H. Rat Jejunal Permeability and Metabolism of Mu-Selective Tetrapeptides in Gastrointestinal Fluids from Humans and Rats. *Pharm. Res.* **1997**, *14*, 1780–1785.
24. Hong, S.Y.; Oh, J.E.; Lee, K.H. Effect of D-Amino Acid Substitution on the Stability, the Secondary Structure, and the Activity of Membrane-Active Peptide. *Biochem. Pharmacol.* **1999**, *58*, 1775–1780.
25. Lehr, C.M. Bioadhesion Technologies for the Delivery of Peptide and Protein Drugs to the Gastrointestinal Tract. *Crit. Rev. Ther. Drug Carrier Syst.* **1994**, *11*, 119–160.
26. Mizuma, T.; Sakai, N.; Awazu, S. Na⁺-Dependent Transport of Aminopeptidase-Resistant Sugar-Coupled Tripeptides in Rat Intestine. *Biochem. Biophys. Res. Commun.* **1994**, *203*, 1412–1416.
27. Haga, M.; Saito, K.; Shimaya, T.; Maezawa, Y.; Kato, Y.; Kim, S.W. Hypoglycemic Effect of Intestinally Administered Monosaccharide-Modified Insulin Derivatives in Rats. *Chem. Pharm. Bull.* **1990**, *38*, 1983–1986.
28. Gan, L.-S.L.; Thakker, D.R. Applications of the Caco-2 Model in the Design and Development of Orally Active Drugs: Elucidation of Biochemical and Physical Barriers Posed by the Intestinal Epithelium. *Adv. Drug Delivery Rev.* **1997**, *23*, 77–98.
29. Sayani, A.P.; Chien, Y.W. Systemic Delivery of Peptides and Proteins across Absorptive Mucosa. *Crit. Rev. Ther. Drug Carrier Syst.* **1996**, *13*, 85–184.
30. Junginger, H.E. Bioadhesive Polymer Systems for Peptide Delivery. *Acta Pharm. Technol.* **1990**, *36*, 110–126.
31. Fix, J.A. Oral Controlled Release Technology for Peptides: Status and Future Prospects. *Pharm. Res.* **1996**, *13*, 1760–1764.
32. Bernkop-Schnurch, A.; Pasta, M. Intestinal Peptide and Protein Delivery: Novel Bioadhesive Drug-Carrier Matrix Shielding from Enzymatic Attack. *J. Pharm. Sci.* **1998**, *87*, 430–434.
33. Su, K.S.E. Nasal Route Delivery of Peptide and Protein Drug Delivery. *Peptide and Protein Drug Delivery*; Lee, V.H.L., Ed.; Marcel Dekker, Inc.: New York, Chap. 13.
34. Sarkar, M.A. Drug Metabolism in the Nasal Mucosa. *Pharm. Res.* **1992**, *9*, 1–9.
35. Kissel, T.; Werner, U. Nasal Delivery of Peptides: An in Vitro Cell Culture Model for the Investigation of Transport

- and Metabolism in Human Nasal Epithelium. *J. Controlled Release* **1998**, *53*, 195–203.
36. Watanabe, Y.; Mizufune, Y.; Utoguchi, N.; Endo, K.; Matsumoto, M. Studies of Drug Delivery Systems for a Therapeutic Agent Used in Osteoporosis. II. Enhanced Absorption of Elcatonin from Nasal Mucosa in Rabbits. *Biol. Pharm. Bull.* **1998**, *21*, 1191–1194.
37. Merkle, H.P.; Wolany, G.J.M. Intraoral Peptide Absorption. *Biological Barriers to Protein Delivery*; Audus, K.L., Raub, T.J., Eds.; Plenum Press: New York, 1993; 131–160.
38. Garren, K.W.; Repta, A.J. Buccal Absorption. III. Simultaneous Diffusion and Metabolism of an Amino-peptidase Substrate in the Hamster Cheek Pouch. *Pharm. Res.* **1989**, *6*, 966–970.
39. Li, C.; Bhatt, P.P.; Johnston, T.P. Transmucosal Delivery of Oxytocin to Rabbits Using a Mucoadhesive Buccal Patch. *Pharm. Dev. Technol.* **1997**, *2*, 265–274.
40. Nakane, S.; Kakumoto, M.; Yukimatsu, H.; Chien, Y.W. Oramucosal Delivery of LHRH: Pharmacokinetic Studies of Controlled and Enhanced Transmucosal Permeation. *Pharm. Dev. Technol.* **1996**, *1*, 251–259.
41. Krishnamoorthy, R.; Mitra, A.K. Ocular Delivery of Peptides and Proteins. *Ophthalmic Drug Delivery Systems*; Mitra, A.K., Ed.; Marcel Dekker, Inc.: New York, 1993; 455–469.
42. Chiou, G.C. Systemic Delivery of Polypeptide Drugs Through Ocular Route. *J. Ocul. Pharmacol.* **1994**, *10*, 93–99.
43. Lee, V.H.L. Precorneal, Corneal, and Postcorneal Factors. *Ophthalmic Drug Delivery Systems*; Mitra, A.K., Ed.; Marcel Dekker, Inc.: New York, 1993; 59–81.
44. Chiou, G.C.; Shen, Z.F.; Zheng, Y.Q. Adjustment of Blood Sugar Levels with Insulin and Glucagon Eyedrops in Normal and Diabetic Rabbits. *J. Ocul. Pharmacol.* **1990**, *6*, 233–241.
45. Muranishi, S.; Yamamoto, A.; Okada, H. Rectal and Vaginal Absorption of Peptides and Proteins. *Biological Barriers to Protein Delivery*; Audus, K.L., Raub, T.J., Eds.; Plenum Press: New York, 1993; 199–227.
46. Lee, V.H.L.; Yamamoto, A. Presentation and Enzymatic Barriers to Peptide and Protein Absorption. *Adv. Drug Delivery Rev.* **1990**, *4*, 171–207.
47. Golomb, G.; Avramoff, A.; Hoffman, A. A New Route of Drug Administration: Intrauterine Delivery of Insulin and Calcitonin. *Pharm. Res.* **1993**, *10*, 828–833.
48. Bonucci, E.; Ballanti, P.; Ramires, P.A.; Richardson, J.L.; Benedetti, L.M. Prevention of Ovariectomy Osteopenia in Rats After Vaginal Administration of Hyaff 11 Microspheres Containing Salmon Calcitonin. *Calcif. Tissue Int.* **1995**, *56*, 272–279.
49. Johnson, L.G.; Boucher, R.C. Macromolecular Transport Across Nasal and Respiratory Epithelia. *Biological Barriers to Protein Delivery*; Audus, K.L., Raub, T.J., Eds.; Plenum Press: New York, 1993; 161–178.
50. Yu, J.; Chien, Y.W. Pulmonary Drug Delivery: Physiologic and Mechanistic Aspects. *Crit. Rev. Ther. Drug Carrier Syst.* **1997**, *14*, 395–453.
51. Smith, P.L. Peptide Delivery Via the Pulmonary Route: A Valid Approach for Local and Systemic Delivery. *J. Control. Release* **1997**, *46*, 99–106.
52. Patton, J.S.; Bukar, J.; Nagarajan, S. Inhaled Insulin. *Adv. Drug Delivery Rev.* **1999**, *35*, 235–247.
53. Deftos, L.J.; Nolan, J.J.; Seely, B.L.; Clopton, P.L.; Cote, G.J.; Whitham, C.L.; Florek, L.J.; Christensen, T.A.; Hill, M.R. Intrapulmonary Drug Delivery of Salmon Calcitonin. *Calcif. Tissue Int.* **1997**, *61*, 345–347.
54. Banga, A.K.; Chien, Y.W. Dermal Absorption of Peptides and Proteins. *Biological Barriers to Protein Delivery*; Audus, K.L., Raub, T.J., Eds.; Plenum Press: New York, 1993; 179–197.
55. Merkle, H.P. Transdermal Delivery Systems. *Methods Funds Exp. Clinical Pharmacol.* **1989**, *11*, 135–153.
56. Green, P.G. Iontophoretic Delivery of Peptide Drugs. *J. Controlled Release* **1996**, *41*, 33–48.
57. Singh, P.; Maibach, H.I. Iontophoresis in Drug Delivery: Basic Principles and Applications. *Crit. Rev. Ther. Drug Carrier Syst.* **1994**, *11*, 161–213.
58. Mitragotn, S.; Blankschtein, D.; Langer, R. Ultrasound-Mediated Transdermal Protein Delivery. *Science*; Wash, DC, 1995; 269, 850–853.
59. McDermott, M.T.; Kidd, G.S. The Role of Calcitonin in the Development and Treatment of Osteoporosis. *Endocr. Rev.* **1987**, *8*, 377–390.
60. Patel, S.; Lyons, A.R.; Hosking, D.J. Drugs Used in the Treatment of Metabolic Bone Diseases. *Drugs* **1993**, *46*, 594–617.
61. Stevenson, J.C.; Evans, I.M. Pharmacology and Therapeutic Use of Calcitonin. *Drugs* **1981**, *21*, 257–272.
62. Reginster, J.Y.; Jeugmans-Huynen, A.M.; Sarlet, N.; McIntyre, H.D.; Franchimont, P. The Effect of Nasal hCT on Bone Turnover in Paget's Disease of Bone—Implications for the Treatment of Other Metabolic Bone Diseases. *Br. J. Rheumatol.* **1992**, *31*, 35–39.
63. Schneyer, C.R. Calcitonin and the Treatment of Osteoporosis. *MD Med. J.* **1991**, *40*, 469–473.
64. Sinko, P.J.; Hu, P.; Wagner, E.; Sturmer, A.; Gilligan, J.P.; Stern, W. Determination of the Intestinal Permeability of Recombinant Salmon Calcitonin. *Pharm. Res.* **1993**, *10*, (suppl.) S293.
65. Yu, H.; Jiang, F.; Stern, W.; Sinko, P.J. Intestinal Binding and Degradation of Recombinant Salmon Calcitonin. *Pharm. Res.* **1994**, *11*, (suppl.) S254.
66. Heintz, M.L.; Flanagan, E.; Orłowski, R.C.; Regnier, F.E. Correlation of Calcitonin Structure with Chromatographic Retention in High-Performance Liquid Chromatography. *J. Chromatogr.* **1988**, *443*, 229–245.
67. Singer, F.; Habener, J.F.; Greene, E.; Godin, P.; Pottisjun, J.T. Inactivation of Calcitonin by Specific Organs. *Nature New Biology* **1972**, *237*, 269–270.
68. Satoh, T.; Yoshida, G.; Orito, Y.; Koike, T. Drug Delivery System for the Treatment of Osteoporosis. *Nippon Rinsho* **1998**, *65*, 742–747.
69. Humphrey, M.J. The Oral Bioavailability of Peptides and Related Drugs. *Delivery Systems for Peptide Drugs*; Davis, S.S., Illum, L., Thomlinson, E., Eds.; Plenum Press: New York, 1986; 139–151.
70. LeCluyse, E.L.; Sutton, S.C. In Vitro Models for Selection of Development Candidates. Permeability Studies to Define Mechanisms of Absorption Enhancement. *Adv. Drug Delivery. Rev.* **1997**, *23*, 163–183.
71. Sinko, P.J.; Lee, Y.H.; Makhey, V.; Leesman, G.D.; Sutyak, J.P.; Yu, H.; Perry, B.; Smith, C.T.; Hu, P.;

- Wagner, E.J.; Falzone, L.M.; McWhorter, L.T.; Gilligan, J.P.; Stern, W. Biopharmaceutical Approaches for Developing and Assessing Oral Peptide Delivery Strategies and Systems: In Vitro Permeability and in Vivo Oral Absorption of Salmon Calcitonin (sCT). *Pharm. Res.* **1999**, *16*, 527–533.
72. Lee, V.H.L. Enzymatic Barriers to Peptide and Protein Absorption. *Crit. Rev. Ther. Drug Carrier Sys.* **1988**, *5*, 69–97.
 73. Bai, J.P.F.; Amidon, G.L. Degradation of Insulin by Trypsin and α -Chymotrypsin. *Pharm. Res.* **1992**, *9*, 969–978.
 74. Yamamoto, A.; Taniguchi, T.; Rikyuu, K.; Tsuji, T.; Fujita, T.; Murakami, M.; Muranishi, J. Effects of Various Protease Inhibitors on the Intestinal Absorption and Degradation of Insulin in Rats. *Pharm. Res.* **1994**, *11*, 1496–1500.
 75. Tozaki, H.; Emi, Y.; Horisaka, E.; Fujita, T.; Yamamoto, A.; Muranishi, S. Degradation of Insulin and Calcitonin and Their Protection by Various Protease Inhibitors in Rat Caecal Contents: Implications in Peptide Delivery to the Colon. *J. Pharm. Pharmacol.* **1997**, *49*, 164–168.
 76. Luessen, H.L.; Verhoef, J.C.; Borchard, G.; Lehr, C.M.; de Boer, A.G.; Junginger, H.E. Mucoadhesive Polymers in Peroral Peptide Drug Delivery. II. Carbomer and Polycarbophil are Potent Inhibitors of the Intestinal Proteolytic Enzyme Trypsin. *Pharm. Res.* **1995**, *12*, 1293–1298.
 77. Lee, Y.H.; Perry, B.A.; Labruno, S.; Lee, H.S.; Stern, W.; Falzone, L.M.; Sinko, P.J. Impact of Regional Intestinal pH Modulation on Absorption of Peptide Drugs: Oral Absorption Studies of Salmon Calcitonin in Beagle Dogs. *Pharm. Res.* **1999**, *16*, 1233–1239.
 78. Legg, E.F.; Spencer, A.M. Studies on the Stability of Pancreatic Enzymes in Duodenal Fluid to Storage Temperature and pH. *Clin. Chim. Acta* **1975**, *65*, 175–179.
 79. Jensen-Pippo, K.E.; Whitcomb, K.L.; DePrince, R.B.; Ralph, L.; Habberfield, A.D. Enteral Bioactivity of Human Granulocyte Colony Stimulating Factor Conjugated with Poly(ethylene Glycol). *Pharm. Res.* **1996**, *13*, 102–107.
 80. Sakuma, S.; Sudo, R.; Suzuki, N.; Kikuchi, H.; Akashi, M.; Hayashi, M. Mucoadhesion of Polystyrene Nanoparticles Having Surface Hydrophilic Polymeric Chains in the Gastrointestinal Tract. *Int. J. Pharm.* **1999**, *177*, 161–172.
 81. Serres, A.; Baudys, M.; Kim, S.W. Temperature and pH-Sensitive Polymers for Human Calcitonin Delivery. *Pharm. Res.* **1996**, *13*, 196–201.
 82. Scott-Moncrieff, J.C.; Shao, Z.; Mitra, A.K. Enhancement of Intestinal Insulin Absorption by Bile Salt-Fatty Acid Mixed Micelles in Dogs. *J. Pharm. Sci.* **1994**, *83*, 1465–1469.
 83. Kimura, T.; Sato, K.; Sugimoto, K.; Tao, R.; Murakami, T.; Kurosaki, Y.; Nakayama, T. Oral Administration of Insulin As Poly(vinyl Alcohol)-Gel Spheres in Diabetic Rats. *Biol. Pharm. Bull.* **1996**, *19*, 897–900.
 84. Bai, J.P.; Chang, L.L.; Guo, J.H. Targeting of Peptide and Protein Drugs to Specific Sites in the Oral Route. *Crit. Rev. Ther. Drug Carrier Syst.* **1995**, *12*, 339–371.
 85. Tozaki, H.; Emi, Y.; Horisaka, E.; Fujita, T.; Yamamoto, A.; Muranishi, S. Metabolism of Peptide Drugs by the Microorganisms in Rat Cecal Contents. *Biol. Pharm. Bull.* **1995**, *18*, 929–931.
 86. Saffran, M.; Kumar, G.S.; Savariar, C.; Burnham, J.C.; Williams, F.; Neckers, D.C. A New Approach to the Oral Administration of Insulin and Other Peptide Drugs. *Science Wash* **1986**, *233*, 1081–1084.
 87. Mersny, R.N. The Colon As a Site for Drug Delivery. *J. Controlled Release* **1992**, *22*, 15–34.
 88. Lee, Y.H.; Perry, B.A.; Sutyak, J.P.; Stern, W.; Sinko, P.J. Regional Differences in Intestinal Spreading and pH Recovery on the Effectiveness of Pharmaceutical Excipients and the Impact on Salmon Calcitonin Absorption in Beagle Dogs. *Pharm. Res.* **2000**, *17*, 284–290.
 89. Hovgaard, L.; Mack, E.J.; Kim, S.W. Insulin Stabilization and GI Absorption. *J. Controlled Release* **1992**, *19*, 99–108.
 90. Merkle, H.P. New Aspects of Pharmaceutical Dosage Forms for Controlled Drug Delivery of Peptides and Proteins. *Eur. J. Pharm. Sci.* **1994**, *2*, 19–21.
 91. Senel, F.M.; Yildirim, S.; Karakayali, H.; Moray, G.; Haberal, M. Comparison of Neoral and Sandimmune for Induction and Maintenance Immunosuppression After Kidney Transplantation. *Transpl. Int.* **1997**, *10*, 357–361.

OTIC PREPARATIONS

William H. Slattery III

House Ear Institute, Los Angeles, California, U.S.A.

INTRODUCTION

Otic preparations are commonly used to treat diseases of the external ear and occasionally of the middle ear. Diseases of the ear include cerumen impaction, dermatitis of the external ear canal, and infectious processes. External otitis (swimmers' ear) and chronic otitis media constitute the majority of infectious diseases of the ear. This article gives an overview of otic preparations, their uses, current availability, and the area of future development.

Anatomy and Physiology of the Ear

Figure 1 is a diagram of the normal ear, which comprises the external ear canal, middle ear space, and hearing canal, or cochlea. The outer two-thirds of the external ear canal is formed by a cartilage framework. The medial or inner one-third of the external ear canal has its framework composed of bone. There is a thin layer of skin covering this bone on the medial one-third. The outer two-thirds is also covered with skin but also has a thick soft tissue lining containing the apopilosebaceous unit. External sounds travel through the external ear canal to reach the tympanic membrane. Vibration of the tympanic membrane transmits a sound wave to the three middle ear ossicles. These, in turn, send the sound wave to the inner ear (cochlea), where it is transformed into a nerve impulse. This impulse is sent to the brain, where the perception of hearing occurs.

The normal external auditory canal has several mechanisms that protect it from infections. The S-shaped anatomy of the external auditory canal provides protection from foreign bodies under normal circumstances. The tragus provides protection anteriorly, and hair from follicles found just inside the meatus prevent airborne debris from entering.

The external auditory canal skin is normally acidic, with a pH level between 4 and 5. Keratin, which consists of desquamated epithelial cells, is produced by the epithelial (skin) lining of the external ear canal; it has an isoelectric point of pH 5. Any increase above this value causes hydration of the keratin layer, increasing susceptibility to pathogenic organisms. Because as most organisms responsible for otitis externa and chronic suppurative otitis

grow best at an alkaline pH level of 7.2–7.6, the acidic pH of the external ear canal is bactericidal or bacteriostatic to many of these pathogenic organisms. The natural low pH level is therefore helpful in preventing ear infections. Anything that alters the pH balance such as swimming may increase the risk of bacterial infections.

Enzymes produced by sweat and sebaceous glands provide antimicrobial activity. Muramidase, a lysozyme excreted by the sweat glands, may be effective in lysing *Staphylococcus epidermidis* and other gram-positive organisms found on the surface of the ear canal skin. Unsaturated fatty acids, resulting from the breakdown of lipids secreted from sebaceous glands, exert antimicrobial activity against gram-negative organisms and fungi (1).

An intact tympanic membrane protects the normal sterile middle ear space from bacterial pathogens. A ruptured tympanic membrane or tympanostomy tube allows external bacteria access to the middle ear space.

Cerumen, or earwax, is produced in the outer third of the external ear canal. It consists of keratin from the epithelial lining and the enzymes and unsaturated fatty acids produced by the sweat and sebaceous glands. It exerts a protective role by forming an oily, mechanical barrier, considered bacteriostatic and fungistatic, over the skin of the external ear canal. A cerumen plug consists mainly of sheets of keratin. It also contains hair and the secretions of both the sebaceous and ceruminous glands of the external ear canal. Contained within these secretions are glycopeptides, lipids, hyaluronic acid, sialic acid, heparin sulfate, lysosomal enzymes, and immunoglobulin (2). The overall chemical composition of cerumen consists of saturated and unsaturated long-chain fatty acids, alcohols, squalene, and cholesterol (3).

Pathology and Bacteriology of Otitis

Development of new drugs and devices to treat diseases of the external ear requires an understanding of pathophysiology and knowledge about the most common organisms responsible for ear inflammation and infections.

Acute otitis external (swimmers' ear) is an inflammatory condition of the external ear canal, most commonly precipitated by local trauma, that is, Q-tips, fingernails, or

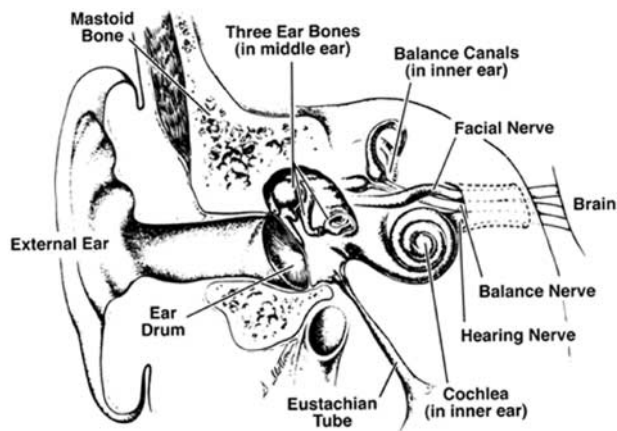


Fig. 1 Diagram of the ear.

other foreign objects that abrade the external ear canal skin. Other predisposing factors include (4):

- Maceration of the epithelial tissue in the ear canal from prolonged exposure to water or moisture;
- Plugging of sebaceous gland ducts, which lowers resistance to infection;
- Moisture absorption by the stratum corneum layer of the epithelium at humidity levels at approximately 80%;
- Elevated ambient temperature against a background of high relative humidity;
- Invasion of exogenous organisms through breaches in a damaged epithelial surface;
- An absence of cerumen; and
- The presence of an alkaline secretion.

Clinical manifestations of the preinflammatory stage of external otitis include itching of the external ear canal and congestion of the apophyseal unit. This is thought to result from the loss of lipids in the external auditory canal that, in turn, results in an increase in the aqueous content of the stratum corneum, causing intracellular edema. The acute inflammatory stage is seen with trauma induced by scratching. By this means, bacteria are allowed access to the dermis. A spectrum of clinical manifestations occurs, ranging from mild edema of the ear canal skin with a clear serous discharge to a severe form characterized by intense pain, a grossly edematous ear canal, and a purulent discharge. A chronic inflammatory stage has also been described. It is characterized by the thickening of the skin, eczematization, lichenification, and superficial skin ulceration (4).

Otomycosis is the result of a superficial fungal infection in the external ear canal. This may commonly result from an underlying bacterial infection or occur in the presence of moist cerumen. Fungi initially implant into the stratum

corneum, where they lie dormant for several days to weeks. They then grow in the superficial layer of the skin, causing inflammation. Initially, the patient usually complains of itching. In the earliest stages, mild edema may be the only clinical symptom. Later, the patient may present with a fungal mass consisting of waxy debris surrounded by a velvety gray membrane with small black spores (5).

Chronic suppurative otitis media is an inflammatory condition of the middle ear. The presence of a tympanic membrane perforation or a tympanostomy tube allows drainage into the external ear canal. Increased vascularity of the mucosa and submucosa, combined with acute and chronic inflammatory cells, is its hallmark. Granulation tissue, fibrosis, and osteoneogenesis are also commonly present. The granulation tissue contains neutrophils and plasma cells associated with small blood vessels and fibroblasts (6).

Pseudomonas aeruginosa and *Staphylococcus aureus* are the most common organisms responsible for acute otitis externa. *Proteus* species are thought to be responsible for the chronic inflammatory stage associated with it. *Aspergillus niger* and *Candida albicans* are the most common organisms responsible for otomycosis. *Mucormycosis*, yeast-like fungi, dermatophytes, and *Actinomyces* may also be seen (5).

Cultures from patients with chronic suppurative otitis media demonstrate that the most common responsible organisms are *P. aeruginosa* and *S. aureus*. These are usually mixed infections with a variety of organisms present (7, 8), and they exhibit a higher resistance to antibiotics.

In a study of 119 cases of chronic suppurative otitis media, Papstavros (9) noted a large number (81%) of gentamicin-resistant organisms in patients previously treated with topical gentamicin. Brook (7) cultured drainage from 54 children with chronic suppurative otitis media and showed that 70% of these patients harbored β -Lactamase-producing bacteria. Of 37 patients with β -Lactamase-producing bacteria, 21 had been treated previously with an oral penicillin or cephalosporin. No child in this study had received prior ototopical drops. In 39% of the children with bilateral-draining ears, a different organism predominated on each side. Both of these studies demonstrate the potential for encountering resistant organisms in chronic otitis media after previous oral or topical antibiotic therapy.

Cerumen Preparations

Normal desquamation of the epithelial layer of the external ear canal results in migration of cerumen toward

the meatus. Any interference with this normal self-cleaning mechanism results in cerumen impaction. An in vitro study has demonstrated that cerumenolysis occurs as the keratin cells, the major constituents of a cerumen plug, are hydrated, resulting in lysis of these cells (2). Consistent with this finding, aqueous preparations have been demonstrated to be better cerumenolytic products than organic, nonaqueous preparations (2, 10, 11). Before using any cerumenolytic agent, the presence of an intact tympanic membrane must be confirmed to prevent two possible complications. Introduction of these compounds into the middle ear is often painful, and the irrigation of cerumen with its keratin cells into the middle ear space may result in a cholesteatoma, a benign but destructive skin growth that requires surgical treatment.

Impacted cerumen can be simply removed by instillation of mineral oil or glycerin into the ear canal. These solutions soften the cerumen, facilitating the separation of the keratin plug from the epithelium, and thus assist the normal migration of cerumen toward the external meatus. Irrigation of the ear canal with hydrogen peroxide is another simple method. The release of oxygen provides a mechanical means for both softening the cerumen and separating it from canal skin. Hydrogen peroxide also increases the moisture within the ear canal, because the breakdown of hydrogen peroxide is water and oxygen. This may predispose to acute otitis externa by increasing the amount of moisture within the ear canal.

Unfortunately, these simple methods are not always successful, and cerumenolytics have been developed to help dissolve cerumen and facilitate its removal. In severe cases, flushing the ear canal with a rinsing solution or physical extraction is still required.

Over the years, numerous cerumenolytic agents composed of aqueous solutions or organic solvents have been tested (Table 1) (12). Currently, the most common over-the-counter products consist of carbamide peroxide in glycerin. The latter softens the wax, whereas the oxygen

released from the peroxide helps to loosen tissue debris. Usually, this agent has to be applied repeatedly over several days for the wax to soften. The softened wax may be removed with gentle irrigation or by blowing air into the external canal.

Triethanolamine polypeptide oleate condensate in propylene glycol (Cerumenex) requires a prescription. Its mechanism of action is thought to result from softening of the cerumen plug and lubrication of the ear canal (2), which is usually achieved in 15–20 min; however, mechanical removal of the cerumen plug may still be necessary. Application of this agent is best confined to a physician's office because severe allergic skin reactions occur occasionally.

Although not marketed as a cerumenolytic, docusate sodium has been found to be a very effective agent (13). Most commonly used as an aqueous fecal softener, its action on cerumen results in keratin cell expansion and lysis. The preparation is alkaline and in solution releases free hydroxyl ions (2). Drops placed in the ear canal for 10–15 min usually result in cerumen disimpaction. The plug may then be easily rinsed out of the ear canal or mechanically removed.

The ideal cerumenolytic preparation (a hypo-osmolar, alkaline, aqueous solution) has not yet been developed. It should be able to lyse the keratin cells of cerumen and allow for easy disimpaction (2).

In contrast to the use of cerumenolytic agents, cerumen replacement products for conditions such as dry skin, eczema, psoriasis, and chronic otitis of the external ear canal are occasionally required for some patients. Unfortunately, there is no product available currently to meet this need.

Antiseptics

Antiseptic agents are often used for the treatment of external ear canal disease. As with cerumenolytics, the

Table 1 Cerumenolytic products

Product	Cerumenolytic agent	Other ingredients
Cerumenex	Triethanolamine polypeptide oleate condensate	Propylene glycol Chlorbutanol (0.5%)
Debrox drops	Carbamide peroxide (6.5%)	Glycerin Propylene glycol Citric acid
Murine ear drops	Carbamide peroxide (6.5%)	Alcohol (6.3%) Glycerin

presence of an intact tympanic membrane must be confirmed before their use. Some antiseptics are commonly used for otologic surgical prophylaxis. Antiseptic otologic preparations are marketed only as the acetic acid solutions.

Acetic acid preparations (usually 2–5% solutions) have both antibacterial and antifungal activities. They are particularly useful against *P. aeruginosa*, *Staphylococci*, β -hemolytic *Streptococci*, *Candida* species, and *Aspergillus*. No organisms are resistant to these preparations (14). Acetic acid solutions placed in the external ear are generally well tolerated and nonsensitizing; however, instillation into the middle ear cavity is associated with pain. The primary drawback of these agents is the vinegar-like smell associated with the instillation. Acetic acid solutions may be combined with aluminum acetate or a steroid compound for anti-inflammatory and antipruritic properties (15). There is a tendency for acetic acid solutions to induce proliferation of the keratin layer, thus increasing the amount of debris within the ear canal associated with these infections. This may complicate the infection, making it slower to resolve.

General antiseptics such as povidine iodine (Betadine[®]), chlorhexidine gluconate (Hibiclens[®]), and hexachlorophene (pHisohex[®]) may be used ototopically for surgical prophylaxis. Povidine iodine is the most common preparation used because of its broad spectrum of activity against microflora, microzoa, and viruses. It must be prevented from entering the middle ear during surgical prophylaxis because it inhibits fibroblast migration during the healing process. Either chlorhexidine or hexachlorophene may be used for surgical prophylaxis in patients who are allergic to iodine. Chlorhexidine is the preferred agent in the iodine-allergic patient because it has a broad spectrum of antimicrobial activity against both gram-positive and gram-negative organisms. Hexachlorophene's bacteriostatic activity is more effective against gram-positive than gram-negative organisms, and supuration decreases its activity (15, 16).

Isopropyl alcohol is used to rinse the ear canal in patients prone to the development of external otitis. It is commonly applied after swimming as a prophylactic measure. Although isopropyl alcohol has broad bactericidal activity, it is widely used as a drying agent for the external ear canal. It displaces water left in the ear canal after swimming. Application into the middle ear space causes severe pain, and it should not be used in the presence of a perforated tympanic membrane.

Gentian violet and thimerosal (Merthiolate) are used for the treatment of fungal infections and are discussed later.

Antifungal Preparations

Most otomycotic infections are the consequence of treatment with antibiotics. Simple cleaning of the external ear canal and discontinuation of the medication will usually suffice to clear up the infection. However, primary and persistent infections require ototopical antifungal medications (Table 2).

Clotrimazole, as a 1% solution, is the most effective topical fungicide for the treatment of otomycosis. It is active against *Aspergillus* and *Candida* species, the most common pathogens responsible for these infections. It acts by interfering with the biosynthesis of ergosterol and is very effective for refractory or chronic cases caused by the dermatophytes or *Candida* species (17).

Amphotericin B is also an effective ototopical preparation. It may be used as a lotion or in a powder form (shown later). Its spectrum of activity covers a variety of fungi, including those responsible for otomycosis. Topical therapy is well tolerated with only rare minor side effects of local skin irritation reported. It is poorly absorbed through the skin (18).

Nystatin and miconazole have a spectrum of activity similar to that of amphotericin B against the common yeast and fungi responsible for otomycosis (17). Nystatin is occasionally used in solution as an ototopic drop. Miconazole is rarely used because it is readily available only as a cream. Application into the ear canal is difficult without impairing the hearing.

m-Cresyl acetate (Cresylate), a derivative of cresol, is marketed as an ototopical antifungal preparation. It is highly active against *Candida* and *Aspergillus*. The ear canal may be painted with *m*-Cresyl acetate on a cotton-tipped applicator, or the compound may be used in solution as an ototopical drop. It is considered the antiseptic of choice for the treatment of otomycosis because it is easily used in the outpatient setting, and, in contrast to gentian violet and thimerosal, no staining is associated with its use. However, eczematization may occur if applied to the concha, and therefore its application should be limited to the external ear canal (16, 19).

Gentian violet is used regularly in the office setting for the treatment of fungal infections. It also has bacteriostatic and bactericidal activity against most gram-positive organisms. The external ear canal is painted with the gentian violet on a cotton-tipped applicator under direct vision using an operating microscope. This compound is a strong dye, and blind application into the external ear canal is difficult without staining. For this reason, self-application by the patient is rarely performed (16).

Thimerosal (Merthiolate) may also be used topically for the treatment of otomycosis. Considered bacteriostat, it is

Table 2 Otomycotic preparations

Product	Antifungal agent	Other ingredients
Otic Domeboro solution	Acetic acid (2%)	Aluminum sulfate Boric acid
VoSol Otic solution	Acetic acid (2%)	Propylene glycol (3%) Benzethonium chloride
VoSol HC Otic solution	Acetic acid (2%)	Hydrocortisone (1%) Propylene glycol (3%) Benzethonium chloride
Fungizone lotion	Amphotericin B (3%) Thimerosal	Propylene glycol
Lotrimin solution	Clotrimazole	Polyethylene glycol
Mycelex solution	Clotrimazole	Polyethylene glycol
Cresylate solution	<i>m</i> -Cresyl acetate (25%)	Propylene glycol Isopropanol (25%) Ethanol (10%)
Gentian violet solution	Gentian violet (1%)	
Merthiolate	Thimerosal 1:1000	
Monistat-Derm lotion	Miconazole (2%)	Pegoxol 7 stearate Mineral oil Benzoic acid
Nystatin suspension	Nystatin (100,000 U/mL)	None

an excellent antiseptic agent in vitro against the common yeast and fungi responsible for otomycosis (16, 17).

Fungal cultures may be required in refractory or persistent cases of otomycosis, in which less common fungi may be responsible; the appropriate ototopical agent may be selected based on culture results. Systemic antifungal medications are rarely indicated for the treatment of otomycosis, unless associated with systemic fungal infections.

Topical preparations for the treatment of otomycosis should not be used in the presence of a perforated tympanic membrane.

Antimicrobial Drops

As a group, antimicrobial otic drops are the most commonly prescribed ototopical medication (Table 3). Most of the preparations listed in Table 3 contain a mixture of antibiotics in combination with a steroid agent. Acetic acid or an alcohol may be added for bactericidal activity. Some of these preparations contain acetic acid as the primary antibacterial agent. Most of these compounds have a low pH, between 3 and 5, similar to that of the normal external ear canal.

Antimicrobial otic drops should be used with caution in the presence of a tympanic membrane perforation because of the potential for ototoxicity. In the case of suppuration, the ear canal should be cleaned before drop instillation.

Purulence within the ear canal will not allow ototopical drops to penetrate the skin, and it is thus prevented from treating the ear infection. Ear cleaning is extremely important when suppuration occurs.

Neomycin and Polymyxin B are the two most common antibiotic agents found in ototopical preparations. Neomycin is bactericidal to many gram-positive and gram-negative organisms, including those responsible for external and chronic otitis such as *S. aureus*, *C. diphtheriae*, *Escherichia coli*, *Proteus*, *Enterobacter*, *Klebsiella*, and *Haemophilus influenzae*. It has no activity against anaerobes, and many strains of *Pseudomonas* are resistant. Cutaneous hypersensitivity is estimated to occur in 6–8% of patients who receive topical treatment (20).

Polymyxin B and Colistin (polymyxin E), first discovered in 1947, have similar antibiotic spectrums limited to gram-negative organisms. *P. aeruginosa* is particularly sensitive to these medications. Other sensitive gram-negative organisms include *Enterobacter*, *E. coli*, *Klebsiella*, and *Haemophilus*. These agents interact with the cell-membrane phospholipids to disrupt the bacterial cells; hypersensitivity is very rare. These antibiotics are poorly absorbed, even when applied to denuded skin (21).

Chloramphenicol may be used ototopically for selective cases of chronic otitis. Despite its relative lack of activity against *P. aeruginosa*, it is bacteriostatic against the other common organisms responsible for chronic otitis,

Table 3 Otic antimicrobial preparations

Product	Ingredients				
	Antimicrobial	Anti-inflammatory	Acid	Antiseptic	Others
Chloramphenicol Cipro HC Otic	Chloramphenicol Ciprofloxacin	Hydrocortisone (1%)	Hydrochloric acetic	Alcohol	Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Coly-Mycin S Otic	Colistin Neomycin	Hydrocortisone (1%)	Acetic	Thimerosal ^a	Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Cortisporin Otic solution	Polymyxin B Neomycin	Hydrocortisone (1%)	Hydrochloric		Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Cortisporin Otic suspension	Polymyxin B Neomycin	Hydrocortisone (1%)	Sulfuric ^b	Alcohol Thimerosal ^a	Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Cortisporin-TC Otic suspension	Colistin Neomycin	Hydrocortisone (1%)	Acetic	Thimerosal	Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Floxin Otic Lazersporin-C solution	Ofloxacin Polymyxin B Neomycin	Hydrocortisone (1%)	Hydrochloric		Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Otic Domeboro solution					Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Otobiotic	Polymyxin B	Hydrocortisone (0.5%)	Acetic Boric		Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Pedi-Otic suspension	Polymyxin B Neomycin	Hydrocortisone (1%)	Sulfuric ^b	Alcohol Thimerosal ^a	Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Pyocidin-Otic	Polymyxin B	Hydrocortisone (0.5%)	Hydrochloric ^b		Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Star-Otic solution			Acetic Boric		Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
Tridesilon Otic solution		Desonide (0.05%)	Acetic Citric		Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
VoSol Otic solution			Acetic		Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin
VoSol HC Otic solution		Hydrocortisone (1%)	Acetic		Propylene glycol Polyvinyl alcohol Sodium acetate Phospholipon 90HB Polysorbate Thonzonium Polysorbate 80 Sodium acetate Glycerin

^aPreservative.^bTo adjust pH.

including *E. coli*, *Clostridium* species, *S. aureus*, *H. influenzae*, *Bacteroides fragilis*, *Klebsiella* species, and certain strains of *Proteus* (15, 22). However, blood dyscrasias and death have been reported after local application, and local skin hypersensitivity may occur with topical therapy (12). Although rarely used for primary therapy, it is used for refractory chronic otitis, especially when a susceptible organism is cultured.

Despite similar antibacterial composition, ototopical preparations may differ in delivery vehicle and pH level. Cortisporin Otic Solution is the most acidic, whereas Coly-Mycin S Otic has a pH of 5, the highest of the group (12, 23). The low pH of these compounds and the alcohol used as an antiseptic agent cause a burning sensation when in contact with the middle ear.

Otobiotic and Pyocidin-Otic contain only the antibiotic polymyxin B. These compounds are useful for the patient allergic to neomycin but should be used only in certain cases because polymyxin B does not act against gram-positive organisms such as *Staphylococcus* and the gram-negative organisms *Proteus* and *B. fragilis*.

Two new fluoroquinolone otic preparations have recently been introduced. Floxic Otic and Cipro HC have recently been approved by the FDA for treatment of external ear disease. Ciprofloxacin and ofloxacin are the active ingredients in these two preparations. Both antimicrobials have a broad spectrum of activity, especially against organisms commonly responsible for otitis externa, including *Pseudomonas*, *S. aureus*, *P. mirabilis*, *Streptococcal* species, and various gram-negative enteric *Bacilli*. In addition, Cipro HC contains hydrocortisone, which adds to the anti-inflammatory properties of this agent. Cipro HC has a low pH level, which makes it ideal for treating external otitis. Floxin Otic has a buffered pH and is FDA-approved for installation into the middle ear space. One difficulty with these newly introduced medications is that many insurance health plans do not cover the cost. These medications are ideally suited for first-line therapy.

The ototopical antimicrobial preparations stated earlier suffice for most cases of otitis externa and selected cases of chronic suppurative otitis. However, these compounds have a limited effect in certain patients with resistant strains of bacteria, drug-induced allergies, or a tympanic membrane perforation that requires administration into the middle ear space. In the last case, ototopical preparations may cause pain because of the acidic pH or the presence of alcohol. Ototoxicity of neomycin, polymyxin B, and colistin is also of concern, and many otolaryngologists prefer topical ophthalmic preparations (23). Ophthalmic preparations are discussed in the article *Ocular Drug Formulation and Delivery* in this volume.

Ophthalmic compounds that may be used as ototopical remedies are given in Table 4 (12). Generally, these products differ from the otic preparations in a neutral pH and the absence of alcohol. For example, in contrast to its otic counterpart, Chloromycetin Ophthalmic has a buffered pH and offers a preparation with hydrocortisone. Likewise, the primary difference between Cortisporin Ophthalmic and the otic preparation is the neutral pH.

Compared with the otic antimicrobial preparations, ophthalmic antimicrobial preparations offer the physician a broader range of antibiotics to treat the difficult ear infection. Gentamicin, tobramycin, and sulfonamides are the most common antibiotics used.

Gentamicin and tobramycin ophthalmic preparations are commonly used to treat difficult ear infections. The latter is less toxic than gentamicin, and its activity against *Pseudomonas* is higher. These antibacterial agents also have a wide spectrum of activity including *Proteus*, *Klebsiella*, *E. coli*, and *Staphylococcus* (20). As stated earlier, resistance to gentamicin develops after topical use (9). Although both of these drugs are known to be ototoxic, the clinical significance is thought to be minimal.

Sulfonamides are bacteriostatic against a wide range of gram-positive and gram-negative organisms. In chronic otitis, ophthalmic preparations are used for their activity against *P. aeruginosa*, *Proteus* species, *Streptococcus*, *Corynebacterium*, *Diphtheriae*, and *H. influenzae*.

Ciprofloxacin is well known for its wide range of bactericidal activity, especially against *S. aureus*, *Staphylococcus epidermidis*, and *P. aeruginosa* (24). This preparation has the ideal antimicrobial spectrum for refractory otitis; however topical therapy should not be considered alone for more serious infections such as those with underlying osteomyelitis.

The ophthalmic preparations should only be used in refractory cases of otitis because organism resistance may develop with widespread use. Ophthalmic preparations are also not improved by the FDA for use in the ear canal, but treating physicians may use their discretion for refractory cases. The recent introduction of Floxin and Cipro Otic drops have decreased the use of ophthalmic preparations for the treatment of refractory ear canal conditions.

Powder Preparations

Powdered preparations have been used for many years in otology. These were originally applied as dusting powders for chronic otitis and were especially useful for a mastoid cavity. Before the advent of antibiotics, antiseptic and acid powders were insufflated into mastoid cavities. Unlike many other otic preparations, powders do not cause pain on administration.

Table 4 Ophthalmologic antimicrobial preparations commonly used ototopically

Product	Ingredients				
	Antimicrobial	Anti-Inflammatory	Acid	Antiseptic	Others
Chloromycetin ophthalmic solution	Chloramphenicol		Boric		Buffer
Chloromycetin hydrocortisone	Chloramphenicol	Hydrocortisone (2.5%)	Boric		Buffer Chloesterol Methylcellulose Benzethonium chloride ^a Benzethonium chloride ^a Sodium acetate Mannitol Edetate disodium Mineral oil Propylene glycol Polyoxyl 40 sterate Glyceryl monostearate Phenylmercuric nitrate Disodium phosphate Monosodium phosphate Benzalkonium chloride ^a Sodium phosphate Tyloxapol Edetate disodium Benzalkonium chloride ^a Propylene glycol Polyoxyethylene-polyoxypropylene compound
Ciloxan	Ciprofloxacin		Acetic Hydrochloric ^b		
Cortisporin ophthalmic solution	Polymyxin B Neomycin	Hydrocortisone (1%)	Sulfuric ^b	Cetyl alcohol	
Gantrisin ophthalmic solution Garamycin ophthalmic solution	Sulfisoxazole Gentamycin				
Metimyd ophthalmic solution	Sulfacetamide	Prednisolone		Thiosulfate Alcohol	
Neosporin ophthalmic solution	Polymyxin B Neomycin Gramicidin			Alcohol Thimerosal ^a	
Polytrim ophthalmic solution	Trimethoprim sulfate Polymyxin B Sulfacetamide		Sulfuric ^b		
Sulamyd ophthalmic solution				Thiosulfate	
Terra-Cortril ophthalmic solution	Oxytetracycline	Hydrocortisone (1.5%)			Methylcellulose Methylparaben Propylparaben Mineral oil Aluminum tristearate Benzalkonium chloride ^a Tyloxapol Edetate disodium Hydroxyethyl cellulose Sodium sulfate Tyloxapol
Tobradex ophthalmic solution	Tobramycin	Dexamethasone	Sulfuric ^b		
Tobrex ophthalmic solution	Tobramycin		Boric Sulfuric ^b		

^aTo adjust pH
^bPreservative

Table 5 Otological powder preparations^a

Ingredients	Amount per dosage (mg)
Chloromycetin	50
Sulfanilamide	50
Fungizone	5
Chloromycetin	50
Sulfanilamide	50
Fungizone	5
Hydrocortisone	1

^aFor patients who are allergic to sulfanilamide, these preparations are available without sulfanilamide.

A powder insufflator can be used for the instillation of antimicrobial agents into the external ear canal or mastoid cavity (25). Current antibiotic preparations suitable for the insufflator device are shown in Table 5 (26). They are packaged into capsules that fit into the insufflator. The patient can easily blow the powder into the ear canal without spreading it around. The activity of these antibiotics against organisms responsible for chronic otitis has already been stated. Other common antibiotics or antiseptics may be applied in powder form in an otolaryngologist's office; boric acid is the most common example (Fig. 2).

Anesthetic Preparations

Anesthetic agents (Table 6) are used to eliminate the pain associated with infections such as external otitis, otitis media, and bullous myringitis. They may also be used locally before surgical manipulation, most commonly during myringotomy. These agents are only recommended for patients with an intact tympanic membrane.

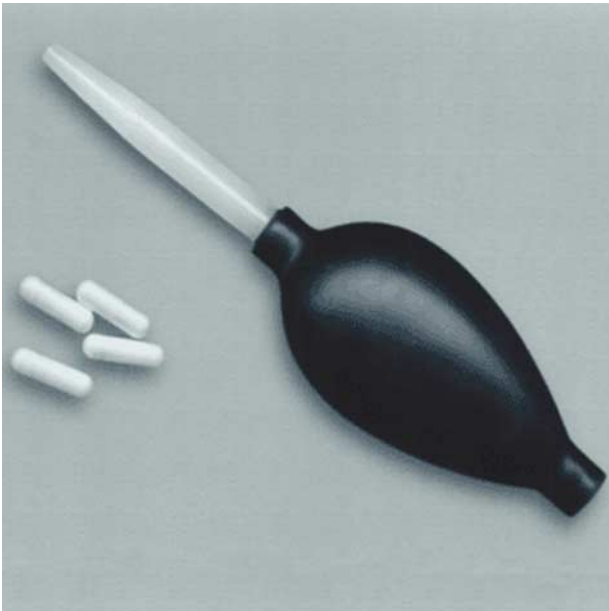


Fig. 2 Bulb insufflator used for antibiotic powders. The capsules contain the antibiotic powder and are placed inside the bulb insufflator.

Most local anesthetic preparations contain benzocaine. Because benzocaine is poorly absorbed through the skin, it remains localized for a long time; however, its effectiveness is unpredictable. Benzocaine has also been known to produce local hypersensitivity reactions (15).

EMLA is a new anesthetic ointment that anesthetizes the external ear canal and eardrum. After keeping the ointment in the external ear canal for 15–20 min it is removed. The surgical procedure may then be completed. This product has been used for myringotomy under local anesthesia.

Phenol is the common topical anesthetic for myringotomies. It is applied over the specific area of the tympanic membrane where the myringotomy is to be performed. It

Table 6 Otic anesthetic preparations

Product	Anesthetic agent	Other ingredients
Americaine Otic	Benzocaine (20%)	Glycerin Polyethylene glycol Benzethonium chloride
Auralgan Otic solution	Benzocaine Antipyrine	Glycerin
EMLA	Lidocaine Prilocaine	Polyoxyethylene Carboxypolyethylene Sodium hydroxide
Phenol	Carbolic acid	

acts by causing instant epidermal destruction. The section of the tympanic membrane in contact with the phenol turns white owing to the anticipated proteins, indicating an anesthetic effect. Healing occurs by hyperplasia of the epithelium and connective tissue but may take some time after applications (16).

Other Preparations

Propylene glycol is a good base for many of the combination antibiotic drops. It acts as a dehydrating agent to fungi and enhances the effectiveness of other antifungal medications. Occasionally, a patient may develop a contact dermatitis from this agent.

Corticosteroids are added to many ototopical combination drops to reduce the inflammation and puritis associated with the acutely infected ear. Corticosteroids may also be used primarily to treat dermatoses found in the external ear canal, primarily psoriasis and seborrheic dermatitis. These compounds may reduce the scaling, itching, and inflammation.

Silver nitrate, as a solution or as a powder on a stick applicator, is occasionally used in the external ear canal as a cauterizing agent. It may be applied to granulation tissue or to the site of a superficial infection. Generally, it is well tolerated; however, if it is excessively applied and bone is exposed, the area does not heal, and surgical correction may be required.

Ototoxicity

The subject of ototoxicity must be addressed whenever discussing the development of new ototopical preparations. It is an important topic from a clinical standpoint and a medicolegal point of view.

In the presence of an intact tympanic membrane, ototoxicity is less important, because the preparation has to be systemically absorbed for an ototoxic effect to occur. This issue is most relevant in cases of chronic suppurative otitis media with a perforated tympanic membrane for

which the ototopical medication has the potential to reach the inner ear via the middle ear. Placed within the middle ear, ototopical medication may diffuse across the oval or round window, resulting in inner ear absorption. These windows consist of a thin membrane separating the middle ear space from the inner ear fluids. There is controversy regarding the clinical relevance of ototoxicity in cases of chronic suppurative otitis media.

A comprehensive review of the ototoxicity of the various ototopical preparations has been published (28). Table 7 lists the agents in which ototoxicity has been demonstrated in animal models. There is no antiseptic that is thought to be free of ototoxicity. The antifungal medications nystatin, amphotericin B, clotrimazole, and tolnaftate have been tested in animal models and found to be free of ototoxicity. The antibiotic found in common otic compounds (Polymyxin B and Neomycin) has demonstrated ototoxicity. Chloramphenicol and Colistin are also ototoxic. The new quinolone otic preparations (Floxin and Cipro) have not demonstrated ototoxicity. The antibiotic Sulfacetamide present in ophthalmic preparations has not demonstrated ototoxicity. No evaluation of the ototoxic effects of topical Tobramycin has been published, although it is thought to be similar to gentamicin in this respect. Although hydrocortisone has demonstrated ototoxicity in animal models, other corticosteroids such as triamcinolone and dexamethasone have not. Desonide has not been tested.

Numerous animal studies have been undertaken to evaluate the ototoxic effects of these different medications; however, the results must be examined with caution because of the differences between the animal models and the human temporal bone. The small mammals used in these studies, usually chinchillas or guinea pigs, have a round window that is easily exposed and very thin compared with the human ear. The human round window is more deeply recessed in bone and is six times thicker than that in the chinchilla. Some temporal bones may actually demonstrate a thin shelf of bone or a thick fibrous plug covering the round window. (29) Further studies must be completed to determine the significance of these

Table 7 Ototoxic otic preparations

Solvent	Antifungal	Antiseptics	Antimicrobials	Anti-Inflammatory
Propylene glycol	Cresylate	Acetic acid Alcohol Benzalkonium chloride Iodochlorhydroxyquinolone Chlorhexidene acetate	Chloramphenicol Colistin Gentamycin Neomycin Polymyxin B	Hydrocortisone Povidone-iodine

differences with respect to ototoxicity and ototopical preparations.

A 1992 survey of otolaryngologists revealed that 80% believed that the risk of sensorineural hearing loss resulting from otitis media was higher than the risk of sensorineural hearing loss from using an ototopical agent known to be ototoxic (30). As McCabe noted in his editorial comment (31):

In 30 years of practice, I have not recognized a single ear damaged in hearing from any antibiotic ear drop, however long term. This total experience has not changed since then.

This comment is important not only for patient care, but also for medicolegal reasons.

SUMMARY

A variety of otic preparations have been reviewed here, including indications, side effects, and limitations. The ideal cerumenolytic compound has not yet been developed, nor has a cerumen replacement product. Antiseptics enjoy wide application but are limited when used in the presence of a perforated tympanic membrane. Available antifungal medications appear to be adequate for the treatment of otomycosis, although none have been approved as otic preparations. The limitations of the otic antimicrobial drops have been stated, including the potential both for organism resistance and for ototoxicity. The recent introduction of Floxin and Cipro HC products make available better otic antimicrobial drops for treatment of otitis externa and chronic otitis media. These drops have generated renewed interest in the development of ototopical preparations. Powder preparations are not widely available and must be specifically prepared by a pharmacist each time they are dispensed.

The current practice of some otolaryngologists using medications that have not been approved by the FDA for ototopical use demonstrates the need for new otic preparations. It is hoped that this article will lay the foundation that to enable the development of these new products.

REFERENCES

1. Cassisi, N.; Cohn, A.; Davidson, T.; Witten, B.R. Diffuse Otitis Externa, Clinical and Microbiologic Findings in the Course of a Multicenter Study on a

- New Otic Solution. *Ann. Otolaryngol. Rhinol. Laryngol.* **1977**, 86 (Suppl 39).
2. Robinson, A.C.; Hawke, M.; MacKay, A.; Ekm, J.K.; Stratis, M. The Mechanisms of Cerumenolysis. *J. Otolaryngol.* **1989**, 18 (6), 268–273.
3. Okuda, I.; Bingham, B.; Stoney, P.; Hawke, M. The Organic Composition of Earwax. *J. Otolaryngol.* **1991**, 20(3), 212–215.
4. Senturia, B.H.; Marcus, M.D.; Lucente, F.E. *Disease of the External Ear*; Grune & Stratton: Orlando, 1980; 31–36.
5. Lucente, F.E.; Smith, P.G.; Thomas, J.R. *Disease of the External Ear. Otolologic Medicine and Surgery*; Alberti, P.W., Rube, R.J., Eds.; Churchill Livingstone: New York, 1988; 1073–1092.
6. Meyerhoff, W.L. Pathology of Chronic Suppurative Otitis Media. *Ann. Otolaryngol. Rhinol. Laryngol.* **1998**, (Suppl. 131), 21.
7. Brook, I.; Yocum, P. Quantitative Bacterial Cultures and β -Lactamase Activity in Chronic Suppurative Otitis Media. *Ann. Otolaryngol. Rhinol. Laryngol.* **1989**, 98, 293–297.
8. Kenna, M.A.; Rosane, B.A.; Bluestone, C.D. Medical Management of Chronic Suppurative Otitis Media Without Cholesteatoma in Children—Update 1992. *Am. J. Otol.* **1993**, 14 (5), 469–473.
9. Papstavros, T.; Giamarellou, H.; Varlejides, S. Role of Aerobic and Anaerobic Microorganisms in Chronic Otitis Media. *Laryngoscope* **1986**, 96, 438–442.
10. Robinson, A.C.; Hawke, M. The Efficacy of Cerumenolytics: Everything Old is New Again. *J. Otolaryngol.* **1989**, 18(6), 263–267.
11. Bellini, M.F.; Terry, R.M.; Lewis, F.A. An Evaluation of Common Cerumenolytic Agents: An In-Vitro Study. *Clin. Otolaryngol.* **1989**, 14, 23–25.
12. *Physicians Desk Reference*; Medical Economics Company: Oradell, NJ, 1993.
13. Chen, D.A.; Caparosa, R.J. A Nonprescription Cerumenolytic. *Am. J. Otolaryngol.* **1991**, 12 (6), 475–476.
14. Jones, E.H. *External Otitis. Diagnosis and Treatment*; Charles C. Thomas, Publisher: Springfield, IL, 1965.
15. *Drug Evaluations Annual 1992*; American Medical Association: Washington, DC, 1992; 1479–1515.
16. Harvey, S.C. Antiseptic and Disinfectants: Fungicides; Ectoparasitocides. *The Pharmacological Basis of Therapeutics*, 5th Ed.; Goodman, L.S., Gilman, A., Eds.; Macmillan Publishing Co.: New York, 1975; 987–1017.
17. Stern, J.C.; Shah, M.K.; Lucente, F.E. In Vitro Effectiveness of 13 Agents in Otomycosis and Review of the Literature. *Laryngoscope* **1998**, 98, 1173–1177.
18. Lopez, L.; Evens, R.P. Drug Therapy of Aspergillus Otitis Externa. *Otolaryngol. Head Neck Surg.* **1980**, 88, 649–651.
19. *Personal communication*; The Recsei Laboratories: Goleta, CA, 1991.
20. Sande, M.A.; Mandell, G.L. Antimicrobial Agents: The Aminoglycosides. *Pharmacological Basis of Therapeutics*; 8th Ed.; Gilman, A.G., Rail, T.W., Nies, A.S., Taylor, P., Eds.; Pergamon Press: New York, 1991; 1098–1116.

21. Sande, M.A.; Mandell, G.L. Antimicrobial Agents: Tetracycline, Chloramphenicol, Erythromycin, and Miscellaneous Antibacterial Agents. *Pharmacological Basis of Therapeutics*, 8th Ed.; Gilman, A.G., Rall, T.W., Nies, A.S., Taylor, P., Eds.; Pergamon Press: New York, 1991; 1117–1145.
22. Fairbanks DNF Otic Topical Agents. *Otolaryngol. Head Neck Surg.* **1980**, *88*, 327–331.
23. Hoffman, R.A.; Goldofsky, E. Topical Ophthalmologics in Otology. *Ear Nose Throat J.* **1991**, *70* (4), 201–205.
24. Mandell, G.L.; Sande, M.A. Antimicrobial Agents: Sulfonamides, Trimethoprim, Sulfamethoxazole, Quinolones, and Agents for Urinary Tract Infections. *Pharmacological Basis of Therapeutics*, 8th Ed.; Gilman, A.G., Rall, T.W., Nies, A.S., Taylor, P., Pergamon Press: New York, 1991; 1047–1064.
25. House, J.W.; Sheehy, J.L. Powder Insufflator for the Ear. *Otolaryngol. Head Neck Surg.* **1983**, *91* (4), 461–462.
26. Personal communication, Medical Square Pharmacy: Los Angeles, CA, 1992.
27. Hickey, S.A.; Buckley, J.F.; O'Connon, A.F.F. Ventilation Tube Insertion Under Local Anesthesia. *Am. J. Otolaryngol.* **1991**, *12* (2), 142–143.
28. Rohn, G.N.; Meyerhoff, W.L.; Wright, C.G. Ototoxicity of Topical Agents. *Otolaryngol. Clin. North. Am.* **1993**, *26* (4), 747–758.
29. Alzamil, K.S.; Linthicum, F.H., Jr. Extranious Round Window Membranes and Plugs: Possible Effect on Intratympanic Therapy. *Ann. Otol. Rhinol. Laryngol.* **2000**, *109* (1), 30–32.
30. Lundy L.B.; Graham M.D. Ototoxicity and Ototopical Medications: A Survey of Otolaryngologists Presented at the Ninth Shambaugh-Shea Weekend of Otology Chicago, IL, March 6–8, 1992.
31. McCabe, B.F. Editorial Comment. *Ann. Otolaryngol. Rhinol. Laryngol.* **1990**, *99*, 41.

OUTSOURCING OF PHARMACEUTICALS

Duane B. Lakings

*Drug Safety Evaluation Consulting, Inc.,
Birmingham, Alabama, U.S.A.*

INTRODUCTION

Outsourcing of pharmaceutical development is big business and getting bigger. Of the over \$40 billion per year research and development budget of pharmaceutical companies, about one-fifth, or \$8 billion, is presently expended at contract service organizations (CSO), and the amount is projected to increase at a rate of about 10% per year in the future. The term CSO includes contract research organizations (CRO), contract manufacturing organizations (CMO), site management organizations (SMO), and any other organization that provides the pharmaceutical industry, or client, with a contract service. Most of this outsourced effort is being expended on the nonclinical drug development, clinical trials, and manufacturing aspects of the drug development process. However, many clients are now using CSOs for drug discovery and potential for development characterization support to generate novel compounds with biological activity against a given target system and then to identify those discovery leads that have the desired attributes with a minimum of detrimental effects needed to become successful preclinical drug candidates. The purposes of this article are to:

1. Describe the history and present status of CSOs;
2. Discuss how clients and CSOs interact;
3. Summarize the services CSOs provide clients;
4. Outline how virtual organizations and consultants interface with clients and CSOs; and
5. Present emerging trends for CSOs.

A complete listing of all the CSOs that offer clients support in one or more of the drug discovery and development processes is beyond the scope of this article. A number of reports are available that provide detailed information on CSOs and the services they provide. Some of these reports are listed in the References. Another common source of information on the services offered by CSOs is the Internet, where most CSOs maintain detailed, up-to-date websites.

HISTORY OF OUTSOURCING

CSOs have been around since the 1940s but have only been major participants in the drug development process for the past 30 years or so and in drug discovery for 10 years or less. The first firms to offer drug development services were not-for-profit research institutes such as Stanford Research Institute, now SRI International, and Midwest Research Institute. With the regulatory requirement that nonclinical safety studies, primarily in toxicology, be conducted according to Good Laboratory Practices (GLP) regulations, for-profit CSOs started to offer chronic toxicity and carcinogenicity study support. At about the same time, the clinical testing requirements needed to characterize the safety and efficacy of a novel chemical entity (NCE) increased dramatically, and Good Clinical Practices (GCP) regulations were issued. Monitoring and auditing of clinical trials to ensure that the data captured on Case Report Forms (CRFs) and other documents accurately reflected information in the raw data were initially accomplished by Clinical Research Associates (CRAs) of the clinical trial sponsor. First independent CRAs and then CSOs were employed by pharmaceutical firms to provide independent auditing of the clinical trial investigational sites.

In the 1980s, a new player came onto the field, biotechnology, with the intent to make macromolecules into therapeutic products. Although many of the new firms had excellent biology and pharmacology expertise, they knew little about the regulatory-driven nonclinical and clinical research and manufacturing requirements necessary to develop a discovery idea into a therapeutic product. Thus, these new companies turned to CSOs. However, most, if not all, of these service providers also did not know how to effectively characterize and develop macromolecules. Thus, the biotechnology companies, the regulatory agencies, and the CSOs rapidly learned how to solve the major problems associated with macromolecule research. After submitting an IND, the biotechnology companies turned to CSOs for assistance in conducting

clinical trials, which required the CSOs to expand services to include protocol development, CRF design, investigator and site selection, data management and statistical evaluation, and report writing. With the increased level of outsourcing, the number of clinical CSOs grew rapidly.

In the 1990s, a group of major events again changed the CSO playing field. The ability to generate libraries of compounds using combinatorial chemistry techniques, high-throughput screening (HTS) techniques to rapidly evaluate the biological activity of these libraries, and bioinformatics to manage and evaluate the large amount of data being generated changed the drug discovery process. The mapping of the human genome will produce a dramatic increase in the number of therapeutic targets available, with the number projected to grow from the few hundreds of targets available today to many thousands in the next few years. Drug discovery groups have been charged to expeditiously find new drug candidates to fill the drug development pipeline. Some companies have developed, or are developing, in-house groups to meet this challenge, whereas others formed partnerships with small biotechnology firms, and still others combined these approaches. In addition, a number of new CSOs were started to provide research services in these relatively novel areas of genomics, proteomics, combinatorial chemistry, HTS, and bioinformatics.

PRESENT STATUS OF OUTSOURCING

CSOs are rapidly adapting to meet the needs of their clients. These changes appear to take one of two avenues. Some CSOs, such as MDS and Quintiles Transnational, are attempting to become full-service, or almost full-service, CSOs by providing research services in drug discovery, potential for development characterization, nonclinical drug development, clinical trial research, and manufacturing. Other CSOs specialize in a given area, such as animal pharmacology studies or phase II to IV clinical trial support. Many pharmaceutical industry observers think that the midsize CSOs will disappear over the next few years, either through mergers with other CSOs or by acquisition by larger CSOs or companies that desire to enter this rapidly growing, but highly competitive, business.

Only a few years ago, drug discovery research was never, or only very infrequently, outsourced to CSOs. Grants to university groups were used to develop models and at times to test discovery leads in these newly defined models. Now, many pharmaceutical companies use CSOs in addition to university groups to support some aspects of

their drug discovery efforts. The transition from drug discovery to drug development has also undergone substantial change in the last decade. A number of CSOs now support this rapidly growing area to assist clients in determining whether a discovery lead(s) has the necessary attributes for further development. The greatest consolidation in the CSO industry has been in nonclinical and clinical drug development, in which contractors that had offered services in one of these areas are adding the other so that they can better serve clients. These additions are frequently accomplished by acquisition of other CSOs that are already offering the desired services, but some CSOs are developing and increasing their own capabilities. A similar trend is occurring in CSOs offering manufacturing support, in which capabilities to produce Good Manufacturing Practice (GMP)-quality drug substances and drug products are being augmented with services such as shelf-life stability studies and impurity profiling.

With consolidation and expansion, large CSOs are becoming bigger and are now being called mega-CSOs. Since 1996, the eight largest CSOs have made more than 40 acquisitions, suggesting that the market is shifting from small CSOs to larger firms. Some recent highly publicized acquisitions include Quintiles' purchase of Innovex and Medical Action Communications and Bulter Clinical Recruitment Service. For Phoenix International Life Sciences, which was acquired by MDS in 2000 and is now called MDS Pharm Services, expansion started in 1995 with the acquisition of I.T.E.M. Holding S.A., IBRD-Rostrum Global, Institute for Pharmacodynamic Research, Anawa Holding, McKnight, and Clinserve to provide multinational capability in clinical research and Chrysalis International to supplement services in nonclinical development. PPD, Inc. has purchased Wisconsin Analytical and Research Service, Gabbay Group, Ltd., Applied Bioscience International, Inc., Belmont Research, SARCO, Inc., GSX Technology, and ATP. Numerous other mergers and acquisitions have taken place, and the trend is expected to continue. However, for each CSO that combines with or is acquired by another CSO, a new player, usually offering a novel service such as genomics or in vitro metabolism, comes onto the field. Whether to use a mega-CSO or a number of smaller, but specialized, CSOs to support outsourcing needs is a question to be addressed by pharmaceutical clients. Another way of asking the question is whether "one-stop" shopping is better than looking at a number of "specialty houses" that offer only a limited line.

As pharmaceutical clients develop drug candidates for the global market, CSOs have expanded their services, primarily clinical trial support services, to include not only

North America, Europe, and Japan, but also a number of emerging markets. A major reason that pharmaceutical companies can enter the foreign markets and that CSOs can offer support services in these markets is the harmonization of the drug development processes through the efforts of the International Conference on Harmonisation (ICH) and the ICH guidelines. Even though ICH guidelines currently apply only to U.S., European, and Japanese research efforts, most other countries are adopting these guidelines. The presently issued ICH guidelines primarily deal with the quality of drug substances and drug products, nonclinical safety, and clinical trials. CSOs, although not a major participant in the definition of the ICH guidelines, are an influential group for their implementation.

CLIENT AND CONTRACTOR INTERACTIONS

Outsourcing is now a common practice of almost all pharmaceutical companies. For example, the use of CSOs to support some aspect of clinical trial research has grown from approximately 30% of the clinical studies conducted in 1993 to over 60% in 1997. The processes that a client uses to select a CSO and how a client interacts with the CSO ensure contracted research studies are monitored appropriately and completed on time and within budget.

A pharmaceutical company identifies a lead candidate that mediates a human disease. For a variety of reasons, corporate management decides to have some or all regulated nonclinical and clinical studies or the manufacturing aspects of drug development performed by a CSO(s). The drug development project team is commonly responsible for coordinating the outsourcing program and for ensuring that the development program stays on time, on track, and on budget. For a small company, this may be the responsibility of two or three researchers who need to have a good understanding of each of the scientific disciplines for which outsourced studies are being considered. A common practice for many firms is to use a consultant or a consulting firm to assist in outsourcing.

The first requirement for a successful research program at CSOs is to identify which aspects of the drug development program are to be conducted at a CSO and the projected timeline for when these studies need to be initiated and completed so that the results are available for decision making and regulatory agency submissions. A well constructed drug development logic plan provides much of this information. The client needs

to identify and then select the appropriate CSO(s) to conduct the desired research studies. The client also needs to monitor the CSO(s) to ensure that the studies are conducted as described in the study protocol and that the results generated are appropriately recorded first in the study records and then in a study report.

CSO Selection

The steps a client should take to identify and select CSOs include but are not be limited to:

- Preparing detailed study designs for each of the research projects to be contracted
- Determining which CSOs are to be considered
- Soliciting cost and time proposals for each study design from each CSO selected
- Evaluating the proposals and selecting those CSOs to be considered further
- Conducting site visits to ensure that the CSOs are qualified to conduct the research studies
- Negotiating time and cost for completion of the research studies
- Selecting the CSOs and awarding the contracts for each study to be outsourced.

The number of person-hours required for the identification and selection process depends on the size of the research program to be contracted. Normally, a minimum of 1 to 2 person-weeks is necessary to effectively evaluate three to four CSOs for each research study to be outsourced. Many clients use consultants to assist in the CSO selection process. However, these clients need to ensure that the consultant has the necessary expertise and knowledge of how CSOs operate. A common mistake is to hire a consultant with expertise in a disease area but not in the research process, such as toxicology or drug metabolism, or in regulatory compliance but not in the science necessary to characterize a drug candidate successfully.

Clients commonly use one of three strategies to identify CSOs. These strategies can be designated virtual, preselected, and special study. The virtual strategy is favored by clients who do not have the resources to conduct GLP-, GCP-, or GMP-regulated research studies. A primary benefit is that the various types of expertise and the infrastructure needed to support regulated studies can be devoted to completing research studies without the client having to build the in-house groups and facilities and thus experiencing costly time delays. A primary limitation is that the client can be vulnerable to poor CSO selection or to mismanagement by the CSO. In the

preselected strategy, a limited number of CSOs are prequalified to support a client's possible outsourcing needs. The qualification process usually includes a detailed site visit to determine which types of research studies can be placed at the CSO. This strategy can provide a synergistic working relationship between the client and the CSO. A major drawback is the unnecessary limitation of outsourcing. If a number of CSOs have the desired expertise but the client's prequalified list contains only a couple of those CSOs, other possibly better qualified CSOs are not even considered. The final strategy, special study, is used by some clients to place single or a few research studies with a contractor. This strategy allows a critical study to be completed to meet the timeline on a drug development plan. However, some companies use this strategy for all of their outsourcing needs and then attempt to integrate the results for the independently conducted studies into a drug development story. For a client with substantial drug development expertise, this strategy may work but requires considerable effort in identifying and selecting CSOs, monitoring the various CSOs, and synthesizing the results from the various research studies. CSOs are generally not in favor of this strategy because they become only "a pair of hands," and have little understanding of the overall development program and thus cannot provide the client with their considerable expertise. Whichever selection strategy is used, the client needs to select the CSO carefully. One poorly conducted study can effectively delay the drug development process until the study has been repeated and the results integrated into the overall story. If this delay is for a research study on the drug development critical path, the projected time for regulatory agency submission has to be changed, thus delaying approval for marketing and resulting in lost revenue for the client.

What are some of the items clients look for when evaluating and selecting CSOs? That list could be quite long, and only a few of the many criteria are presented here.

- *Project management skills:* The CSO project manager ensures that the contracted activities are completed in a timely and cost-effective manner. This individual needs to be identified early so that an acceptable working arrangement can be established.
- *Research area and therapeutic experience:* A CSO with expertise in a variety of scientific disciplines and therapeutic areas will have a pool of experts available as a knowledge resource as results become available for evaluation.
- *Flexibility and adaptability:* The ability to adapt to changes in research study parameters and timelines is

important to understanding the dynamic nature of the drug discovery and development process. A CSO needs to be able to interact effectively with a client to modify a study design or an ongoing research study.

- *Timeliness:* A CSO needs to be able to provide the contract service(s) agreed on in the desired time frame.
- *Integrity:* The results to be generated need to be accurate, complete, and unbiased. Previous performance or references from other clients can assist in ascertaining how the CSO will perform in providing "good" as well as "bad" news.
- *Quality assurance (QA) awareness:* The CSO's established procedures ensure both quality work and that current regulations are followed. A review of the CSO's standard operating procedure (SOP) manual and discussions with the QA manager will provide this information. Also, results from recently conducted regulatory agency audits should be reviewed.
- *Training:* The documented training process ensures that all research study personnel have the necessary skills for conducting their aspect of the contracted study.
- *Services offered:* The CSO offers a sufficiently broad scope of services for conducting the contracted research study in the shortest time possible. The services offered anticipate contract project demands and respond quickly to project dynamics.
- *Communication:* The CSO provides information in an acceptable format and on a timely basis. A constant open line of communication should exist between the client and the CSO.
- *Partnership:* The CSO considers itself a partner with the client and is service-oriented. The CSO creates a team environment to perform as an extension of the client's project team.

One item not listed above is cost. Most clients use cost as an evaluation criteria, at times the only criteria. Although important, cost needs to be evaluated in conjunction with other CSO selection items. A CSO offering the lowest price may not:

- Fully understand the research objectives;
- Be able to conduct the studies on the timeline desired;
- Include all the tests or evaluations necessary to support a regulatory submission; or
- Provide a complete evaluation of the study results.

If all other selection criteria among CSOs are equal, then cost can and should be used to select the contractor. Once a CSO has been selected, time and cost negotiating is justifiable to ensure that the client receives the best quality study for the best price.

CSO Monitoring

The identification and selection process is only the first step. The second aspect involves monitoring the CSOs to ensure that the research studies are conducted according to the research protocol, that the results are obtained using appropriate techniques and procedures, and that the data generated are correctly recorded and documented in the study report. Many clients hire firms, such as other CSOs or consultants, to monitor outsourced research studies, particularly clinical trials and manufacturing projects. In doing so, the clients should use the same techniques to select and evaluate the monitors as they use to identify a CSO. Monitoring studies at CSOs should include but are not be limited to:

- Reviewing and approving the protocols, which should provide information on all aspects of the study, and detailing the procedures to be followed to successfully complete the study
- Monitoring various aspects of each study to ensure that the data collected do not contain “surprises” that can prevent the results from being used to support submissions to regulatory agencies
- Assisting in the evaluation and interpretation of results to ensure that the data are analytically acceptable and correctly correlated to tell the story of the experimental results
- Reviewing study reports to ensure that they accurately reflect the generated results, document any deviation from the study protocols, and give appropriate conclusions.

The number of person-hours required to appropriately monitor a research study conducted by a CSO again depends on the size of the research program. Normally, a minimum of 1 person-week for each in-life or inpatient phase month of a research study is required and includes the time necessary to review and approve the study report.

OUTSOURCING OF DRUG DISCOVERY

Drug discovery strives to find new ways to discover novel compounds and to identify new targets for treating the various disorders affecting humans. After years of first synthesizing individual compounds, then purifying and obtaining physical and chemical characterization of the new chemicals, and finally testing the NCEs in pharmacological models of a particular human disease, the pharmaceutical industry has embraced combinatorial chemistry as a way to generate large numbers of structurally unique compounds, HTS to test these

chemicals for biological activity against a target, genomics to identify the genetic code for enzymes and proteins that might be new drug targets, proteomics to produce the newly identified natural proteins and to characterize their functions in the body, and bioinformatics to organize and evaluate the enormous amounts of data generated. Once a lead or, more likely, a group of leads is identified, the biological activity in other in vitro and in vivo pharmacology models is explored further, and structural activity relationship (SAR) techniques are used to further modify the molecular structures of the leads to identify NCEs with very high potency and specificity, which means the leads have minimal or acceptable adverse effects or interference with other physiological functions or organ systems.

The present cost and time estimate for the drug discovery and development process is approximately \$500 million and 12 to 14 years, respectively, for each new therapeutic agent that reaches the market. The time and cost profile is shown in Fig. 1. The drug discovery and preclinical phases, or the first few years, are relatively inexpensive relative to the clinical, nonclinical, and manufacturing phases, in which over 80% of the cost and more than half the time are expended. The \$500 million cost includes the money spent on the losers, those compounds that were identified as having biological activity against a target but that do not successfully complete the process. Because only approximately 2% of discovery leads selected for preclinical development enter into clinical studies and only an estimated 5% of these

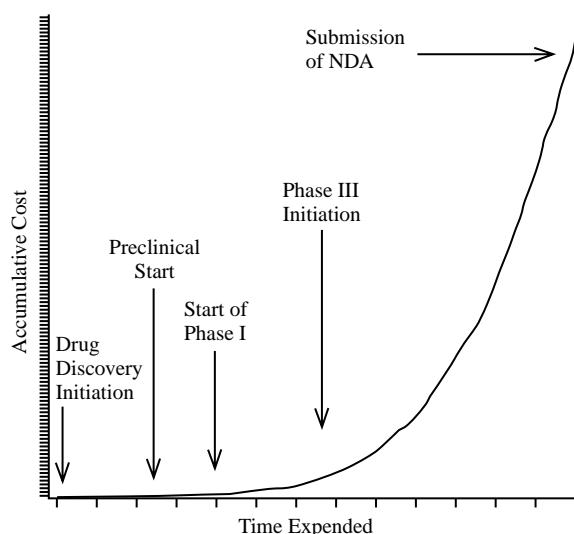


Fig. 1 Time and cost profile for drug discovery and development.

drug candidates result in New Drug Application (NDA) submissions (for an overall success rate of about 0.1%), finding and gleaning the losers probably uses more than 80% of the cost and half the time needed to put a novel therapeutic agent on the market. By identifying the losers early in the process and thereby reserving precious resources for discovering compounds that have a better chance of success, companies could substantially reduce the overall cost for each new therapeutic agent and use precious resources on discovering and developing compounds with a higher probability of being a winner and thus also shorten the time required for drug discovery and development.

All successful pharmaceutical firms understand this problem and are constantly searching for novel approaches to expedite the process and to reduce the cost. One approach used more and more is to outsource some or all of the drug discovery and pharmacology aspects to CSOs that have particular expertise that are not available in a pharmaceutical company and that would be costly in time and resources to establish. This section summarizes the present scientific disciplines involved in the drug discovery process. Because the techniques and procedures used in this rapidly evolving area are changing so quickly, only summary information can be presented. A few CSOs that provide services in more than one discovery area are listed in Table 1.

The entire human genome is projected to be sequenced by 2002 or shortly thereafter. The human genome contains 60,000 to 80,000 genes that code for proteins, many of which could be new targets of mediating human diseases, and many others that have little potential as targets for drug therapy. Present projections suggest that 5000 to 10,000 of the genes, or approximately 10%, may be involved in regulating some aspect of cellular function or in producing signals to simulate or inhibit extracellular events or other cells. In combination with chemical libraries and HTS techniques, these new targets are expected to identify many novel ways of treating diseases. A major problem is determining which of these genes are important in a disease process and which are not. Proteomics is the science that will help in these determinations. The newly discovered genes can be transfected into bacterial or mammalian cell lines to produce sufficient quantities of the protein for further characterization and study.

Combinatorial chemistry technology and the merging of automated library synthesis with the traditional medicinal chemistry principles to provide a more rational approach to library generation are providing pharmaceutical companies with millions of novel compounds. Advantages of generating these diverse libraries include acting as a synthetic substitute for natural samples in random discovery screening programs and providing a

Table 1 Various drug discovery services provided by CSOs

CSO	Gene ^a	Prot ^b	Chem ^c	Screen ^d	Bioinf ^e	Pharm ^f
Axiom Biotechnologies				X	X	X
Cellomics				X	X	X
Commonwealth Biotechnologies, Inc.	X	X	X			
Genespan	X	X		X		X
Jerini Bio Tools			X	X		
Marin Biologic	X	X	X		X	X
MDS		X	X	X		X
MRI	X	X	X			X
PPD, Inc.	X		X	X	X	X
SRA Life Sciences	X			X		X
SRI International			X	X		X
THETAGEN	X		X		X	
TNO Pharma Institute	X	X		X		X

^aGenomic research including sequencing, cloning, and analysis.

^bProteomic research including expression and analysis.

^cChemical synthesis of compounds including combinatorial chemistry to synthesize libraries of small organic molecules, peptides, and oligonucleotides.

^dScreening of drug discovery leads for pharmacological activity including HTS and assay development.

^eBioinformatic techniques for genomic, chemistry, and pharmacology research efforts.

^fPharmacology evaluation in either in vitro and in vivo models.

systematic approach for SAR studies to assist chemists and pharmacologists in defining the pharmacophore and other chemical structure attributes that can be important for metabolic stability, aqueous solubility, and lipid membrane penetration. A key aspect is molecular modeling, which is critical to understanding the chemical reactions and can be used to model structures, provide conformational analysis, and develop the pharmacophore or other structural attributes. To support these efforts, a number of CSOs offer combinatorial chemistry services and classical synthetic capabilities for small organic molecules and for macromolecules.

The library or other source of chemicals is screened for biological activity, usually employing an *in vitro* system in which a known biochemical process is agonized or antagonized. The identification and characterization of the target system can be a rate-limited step, especially for a newly discovered, genomic-defined target. The increased number of compounds to be tested has required pharmacologists to devise novel techniques to rapidly screen for biological activity. With a number of new targets becoming available, many pharmaceutical companies are turning to CSOs for assistance in developing and implementing HTS assay systems to evaluate the biological activity of the novel compounds in libraries.

Bioinformatics assists in understanding information content and flow in biological systems and processes and is becoming more and more important because of the unprecedented growth in quantity and diversity of these data. Applications for bioinformatic techniques include understanding biological processes and how they may malfunction, thus having an important place in drug discovery and development processes. At present, the role and function of most proteins are at best incomplete and often nonexistent. Bioinformatics can assist researchers in understanding what protein each gene produces and in determining the physiological role and function of each of these proteins. Bioinformatics can facilitate the selection of drug targets, biological pathways, receptor functions, gene regulation, and intercellular communication and the impact that each of these has on the others and on normal and disease states. A number of CSOs have developed bioinformatics tools and techniques and offer these services to clients. Because this area is growing so rapidly and new players are continuously entering the field, any list of CSOs involved in bioinformatics would be outdated before publication.

After a biologically active compound(s) has been identified, additional, experiments are conducted to further study the pharmacological activity. These studies are frequently conducted in *in vivo* models, and analogs of the

lead are also tested to further characterize the mechanism of action and to ensure that a compound with high potency and specificity is identified. This more classic *in vivo* approach to evaluating the biological potency is a more rigorous SAR assessment than one that uses *in vitro* systems because the discovery lead(s):

- Has to be administered to the animal model and be delivered to the site of action
- Is subjected to drug metabolizing enzymes
- May cause unexpected responses that describe a novel activity or an unwanted adverse effect
- Has to prevent the normal clearance processes from eliminating the lead before it reaches the target in sufficient concentration to elicit the desired biological response.

For these and other reasons, most pharmacologists still consider the use of *in vivo* animal models to evaluate the pharmacology of a discovery lead(s) to be the gold standard for assessing biological potency and specificity. An important requirement in SAR determination is having or developing an animal model that correlates with a disease in humans. Developing these animal models can be time-consuming and expensive, and many important human diseases do not yet have predictive animal models. Research to develop acceptable models for these diseases is a high priority at pharmaceutical companies. Many CSOs are also involved in the definition and characterization of novel *in vivo* models and, once available, offer these novel models to clients.

OUTSOURCING OF POTENTIAL FOR DEVELOPMENT CHARACTERIZATION

Completed drug discovery studies indicate that a lead(s) mediates a disease process and has potential as a therapeutic agent in humans. Is this lead now ready to be transferred to the preclinical drug development group? Or should additional, nondefinitive experiments be conducted to characterize more fully the properties of the lead(s)? If more studies are necessary, what experiments should be done? This section describes some of the experiments that could be conducted to evaluate the potential of a discovery lead to become a development candidate and provides information on CSOs that offer services in this relatively new and rapidly evolving field. These nondefinitive studies may also uncover problems that have to be resolved before starting the definitive preclinical development studies and designing the clinical trial protocols to evaluate safety and efficacy in humans. A number of questions should be

answered to effectively plan the research studies needed to assess the potential for development of a discovery lead. These questions include but are not limited to the following:

- What is the human disease indication for the potential drug candidate?
- What are the proposed route of delivery and frequency of dosing in human clinical studies?
- How long does the lead need to be in the body to elicit the desired pharmacological response?
- Will the physical and chemical properties of the lead allow delivery to the site of action?
- Can the lead be synthesized in sufficient quantity to support a drug development program?

Six scientific disciplines, shown in Fig. 2, are involved in the potential for development characterization of a lead or the assessment of which lead from a group has more development attributes and fewer drawbacks. The classic approach to discover lead optimization has been to turn the actives over to the pharmacology group. The lead with the highest potency is then selected for further development. During the past few years, a major strategic change has been the addition of screening for bioavailability or delivery, chemical and metabolic stability, and, in some cases, pharmacokinetics and toxicity to complement the pharmacology testing. A drug delivery group assesses which leads are effectively delivered by the proposed clinical route of administration. A solubility and stability group defines which leads are easily degraded and are thus not good prospects for formulation development or have insufficient aqueous solubility to first dissolve and then stay in solution after administration. The physiological distribution and disposition of these leads are studied by the pharmacokinetics group. The potential for the leads to be metabolized to inactive or active metabolites is ascertained by a drug metabolism group, which can also

conduct the drug–drug interaction studies recommended as part of the drug development process (1). An *in vitro* toxicology group evaluates the leads for cellular, genetic, and immunotoxicity. Finally, an acute toxicity study in animals may be conducted to select the lead with the optimal chance of successful development. This acute toxicity study is not an LD₅₀ study, which is not needed for overall risk assessment (2). If the pharmacologically defined actives do not meet the potential for development requirements, a pharmaceutical company has at least two choices. The first is to select the compound with the “best” overall profile and enter formal development, which will most likely result in the selected candidate being one of the preclinical or clinical losers. The other is to start the discovery process over, hoping that the next time an acceptable lead will be found and that the lost time does not result in a competitor discovering an acceptable compound that becomes first on the market.

Safety pharmacology evaluations are commonly considered to be part of the toxicology program but are in reality pharmacology studies at exaggerated doses designed to reveal the potential of an identified discovery lead(s) or drug candidate to produce adverse effects or secondary pharmacological effects on major organ systems. Safety pharmacology is commonly conducted shortly after a discovery lead has been selected for further development, and the results are used to assist in determining whether continued development is warranted, to support regulatory agency submissions, and to identify potential adverse effects. An ICH guideline (3) indicates that safety pharmacology studies include assessment of effects on vital functions such as the cardiovascular, central nervous, and respiratory systems and that these tests should be conducted before human exposure. In keeping with the current regulatory climate, most pharmaceutical companies conduct, or have a CSO conduct, a battery of tests and, at times, subdivide these tests into groups to first determine specific properties of compounds and then to more fully characterize those areas where a response is observed.

During the past few years, techniques to rapidly evaluate potential for development parameters of a large number of leads have been developed and are now offered to clients as a service by a number of CSOs. These screening methods use small amounts of each lead and, in some cases, can evaluate more than one lead simultaneously in a given sample. A few of the CSOs that offer clients more than one service in potential for development characterization and the scientific disciplines for which they have expertise are given in Table 2.

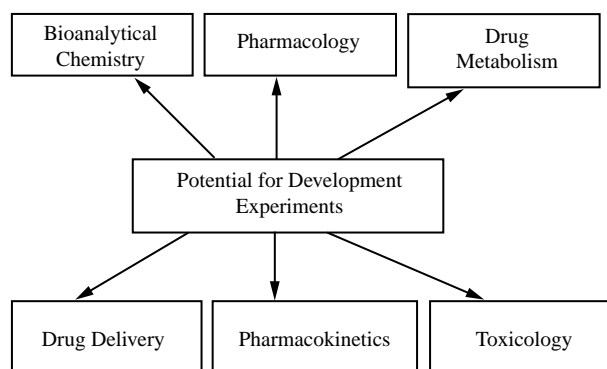


Fig. 2 Scientific disciplines involved in potential for development characterization evaluations.

Table 2 CSOs providing potential for development-assessment services

CSO	Tox ^a	SafPh ^b	DM ^c	PK ^d	DDel ^e	Bioan ^f
CEDRA			X			X
Genespan	X			X		
GENTEST	X		X	X	X	X
Huntingdon Life Sciences	X	X				
In Vitro Technologies	X		X	X	X	X
Marin Biologic	X					X
MetaXen			X		X	X
MRI	X			X	X	X
NaviCyte				X	X	X
MDS	X	X	X	X	X	X
Primedica		X		X		X
SRI–Serquest	X			X		X
SRA Life Sciences	X		X			X
SRI International	X	X	X			
TNO Pharma Institute	X	X	X	X	X	X

^aIn vitro and/or acute toxicology.^bSafety pharmacology.^cIn vitro drug metabolism including enzyme induction, metabolic stability.^dIn vitro and in vivo pharmacokinetics including protein binding and drug–drug interactions.^eDrug delivery including in vitro and in vivo absorption assessment.^fBioanalytical chemistry method development and support.

OUTSOURCING OF NONCLINICAL DRUG DEVELOPMENT

Before entering into human clinical trials, a drug candidate undergoes various preclinical studies to define and characterize its safety and disposition profiles in animal models. After entering the clinic, information on the pharmacokinetics and toxicology of the drug candidate in the relevant species (human) finally becomes available. The results from earlier drug development experiments are usually re-evaluated in consideration of this new information to ascertain whether the animal models were predictive of the efficacy and safety observed in humans. If the animal results are extrapolative, the remaining nonclinical animal studies are fairly straightforward and are conducted to provide supportive information on the safety of the drug candidate. However, if the early animal data do not extrapolate to humans, additional animal experiments may be needed to more fully understand the observed pharmacology and/or toxicology in humans. The following discussion presents an overview of the research studies conducted during nonclinical drug development, which can be divided into the broad categories of bioanalytical and analytical chemistry, formulation development, pharmacokinetics, drug metabolism, and toxicology. Some of the CSOs that offer services in a number of these disciplines are listed in Table 3.

A bioanalytical chemistry method is used to support definitive pharmacokinetic and toxicology studies after the assay has been appropriately validated. The validation experiments define acceptance and rejection criteria for the range of reliable results, the lower and upper limits of quantification, accuracy, precision, specificity, and recovery and should include appropriate stability studies (4, 5). The validated method needs to be documented in a test assay procedure, which lists the acceptance and rejection criteria for each parameter evaluated, and supported by appropriate SOPs.

Analytical chemistry and formulation development and characterization are aspects of manufacturing activities; however, these functions are critical for the successful completion of many nonclinical and clinical research studies. The section on outsourcing of manufacturing provides a summary discussion on analytical chemistry and formulation development.

Preclinical pharmacokinetic and bioavailability experiments are conducted to evaluate dose proportionality over the dose range used in toxicology studies and potential species-to-species differences in pharmacokinetic profiles. By incorporating one or two intravenous dose levels into the studies, absolute bioavailability can also be determined and information on the linearity of absorption, distribution, and disposition kinetics obtained. If more than one drug formulation is used in toxicology studies, relative

Table 3 CSOs providing various and multiple nonclinical services

CSO	Tox ^a	PK ^b	DM ^c	Bioan ^d	Form ^e
ABC Laboratories		X	X	X	
BIBRA International	X	X	X	X	
BioReliance	X		X	X	
Center de Recherches Biologiques	X	X	X	X	
ClinTrials Research	X		X	X	
Covance	X		X	X	X
Huntingdon Life Sciences	X	X	X	X	X
IIT Research Institute	X	X	X		
ILEX Oncology	X	X		X	X
Inveresk Research International Limited	X	X	X	X	X
ITR Laboratories Canada	X	X	X	X	X
LCG Bioscience	X	X	X	X	
Lovelace Respiratory Research Institute	X	X		X	X
MDS Pharmaceutical Services	X	X	X	X	
MRI	X	X	X	X	X
Northview Biosciences	X			X	X
Nucro-Technics	X			X	X
Oneida Research Services		X		X	X
Oread Inc.	X		X	X	X
PPD, Inc.		X	X	X	
Primedica	X	X	X	X	X
Quintiles Transnational	X	X	X	X	X
Ricerca	X	X	X	X	X
SGS Biopharma–Lab Simon	X	X		X	X
SRI–Serquest	X	X	X		X
SRI International	X	X	X	X	X
STS duoTEK	X		X	X	X
TNO Pharma Institute	X	X	X	X	X
WIL Research Laboratories	X	X	X	X	X

^aToxicology including subchronic and chronic, reproductive carcinogenicity.^bPharmacokinetics including toxicokinetics.^cDrug metabolism including radiolabel synthesis.^dBioanalytical chemistry.^eFormulation assessment, including solubility, stability, and delivery.

bioavailability experiments comparing the formulations can determine whether the extent of delivery is similar or different and thus can make extrapolation of pharmacology and toxicology results between animal species meaningful and useful in designing the later nonclinical studies and phase I safety and tolerance studies in humans. An important application of pharmacokinetics is determining the extent and duration of exposure, or toxicokinetics, in the test species. An ICH guideline (6) discusses the generation of toxicokinetic data to support the development of a drug candidate. Unless justified from pharmacokinetic results from humans or from toxicokinetic observations, additional animal pharmacokinetics are not usually conducted during nonclinical development. Types of animal pharmacokinetics that might be

performed include multiple-dose pharmacokinetics, bioavailability comparison when the formulation used in early toxicology studies is changed, newly identified drug candidate metabolites distribution and disposition evaluations, effect of food and time of feeding on the extent and duration of absorption, and drug–drug interactions if the animal model is considered predictive of humans.

Drug metabolism experiments in toxicology animal species are conducted to evaluate the protein binding, mass balance, and tissue distribution and to use an appropriately labeled compound. If the radiolabel is not nonmetabolically or metabolically stable, the results from the drug metabolism experiments or other studies using the labeled compound will have little if any meaning or usefulness in determining the metabolic fate of the drug

candidate. The two most common preclinical drug metabolism studies are for mass balance, which includes metabolite profiling and identification, and for tissue distribution. A routine aspect of most tissue distribution studies is whole-body autoradiography. Nonclinical drug metabolism studies that may be conducted include multiple-dose tissue distribution [as addressed in an ICH guideline (7)], additional characterization and evaluation of metabolites, and studies such as fetal-placental transfer and lacteal secretion, which provide toxicokinetic support for developmental and reproductive toxicology evaluations.

A number of preclinical toxicology studies need to be completed and documented in a regulatory agency submission before initiating human clinical trials. The number and types of toxicology studies depend on the disease type, the duration of treatment, and the country(ies) in which clinical trials will be conducted. Preclinical toxicology studies include genetic toxicology, local tolerance, immunotoxicity, and acute and subchronic tests. After human clinical trials have been initiated, nonclinical toxicology is continued with chronic toxicology studies, developmental and reproduction studies, and carcinogenicity studies. One of the most troublesome problems in interpreting toxicology results is determining whether these data are predictive of toxicity in humans. Quite often, animal toxicology may not correlate with human safety because the observed toxicities are species-specific. Species specificity is sometimes discovered early in development and can be used to design the early human trials to ascertain whether humans also manifest the observed toxicity.

An ICH guideline (3) indicates that local tolerance evaluations are to be conducted in animals using the route of administration proposed for humans. The genotoxicity (8,9) and immunotoxicity (3) potentials of a drug candidate are commonly studied shortly after a discovery lead has been selected for further development. The doses for definitive toxicology studies are defined in dose-range-finding studies, which have a primary goal of determining a maximum tolerated dose (MTD). The route of administration, frequency, and duration of dosing for this and other toxicology studies are determined from the expected clinical use of the drug candidate. When these parameters do not mimic those used in human clinical trials, unexpected findings surprises result, and the severity of the findings may be sufficient to stop development of an otherwise promising drug candidate. Most regulatory agencies require subchronic toxicity studies in two species, one of which must be a nonrodent, before the initiation of human clinical trials. ICH guidelines (3,10) suggest the

minimum duration of toxicity studies needed to support phase I, II, and III clinical trials. Regulatory agencies (2,3) require chronic toxicity studies in two species for drug candidates that are to be administered to humans for longer than 3 months. Both subchronic and chronic toxicology studies should include hematology, clinical chemistry, and histopathology evaluations. Developmental and reproduction toxicology studies are conducted to reveal any effect of a drug candidate or its metabolite on mammalian reproduction and to ascertain the potential risks to humans. These studies evaluate male and female fertility, embryo and fetal death, parturition and the newborn, the lactation process, care of the young, and the potential teratogenicity of the drug candidate. The combination of studies selected needs to allow exposure of mature adults and all stages of development from conception to sexual maturity to conception in the next generation (11,12). Carcinogenicity studies encompass most of the life span of the test species and are conducted to measure tumor induction in animals and to assess the relevant risk in humans (13–16). These studies are normally conducted concurrently with phase III human clinical trials and are required by regulatory agencies when human exposure to a drug candidate is anticipated to be longer than 6 months.

OUTSOURCING OF CLINICAL DRUG DEVELOPMENT

Drug development is a long, complicated, highly regulated process, with the clinical trial aspect of the overall endeavor being both the most time-consuming and expensive. The clinical development stage requires approximately 75% of that time and over 50% of the money expended. Pharmaceutical companies are constantly searching for ways to shorten the clinical trial stage. However, with time and cost of drug discovery and development both still increasing, these efforts have not been overly successful to date. One approach, through the ICH process, is to harmonize many of the clinical trial tasks so that research studies conducted in one country are acceptable to the regulatory authorities in other countries. An ICH guideline (17) provides assistance on some general considerations for clinical trials. This section addresses clinical drug development, which is subdivided into four phases:

- *Phase I:* Safety and tolerance evaluation in human volunteers or in a small number of patients and pharmacokinetic and pharmacodynamic studies (human pharmacology)

- *Phase II*: Efficacy in patients and dose-response and dosing regimen definition (therapeutic exploration)
- *Phase III*: Definitive safety and efficacy in patients to define benefit-to-risk relationship and to confirm the dose-response relationship (therapeutic confirmation)
- *Phase IV*: Studies performed after marketing approval, related to the approved indication, and used to obtain a better understanding of the benefit-to-risk relationship and to identify less common adverse events (therapeutic use)

For each clinical trial study conducted to support the development of a drug candidate, certain features are necessary to ensure that the study is appropriately designed and conducted; that the data generated are collected, evaluated, and interpreted; and that the results are documented in a clinical trial study report. These features can be designated clinical trial design, investigator and clinical trial site selection, clinical trial site monitoring, clinical trial supply management, data collection and management, data evaluation and interpretation, report

Table 4 CSOs that conduct phase I–IV clinical trials

CSO	Phase I		Phase		
	Unit ^a	Spec ^b	II	III	IV
Advanced Biomedical Research	X	X			
Anapharm	X	X			
Applied Analytical Industries	X	X	X		
ASTER•CEPHAC	X	X	X		
BIBRA International	X	X			
Clinical Pharmacology Associates	X	X			
Clinical Research Services Turku	X	X	X	X	X
ICSL Clinical Studies	X		X	X	X
Colorado Medical Research Center	X	X	X	X	X
Covalent Group			X	X	X
Covance	X	X	X		
CroMedica Global	X	X			
Drug Research and Analysis Corporation	X	X			
FARMOVS	X	X	X	X	
Hill Top Research			X	X	X
Inveresk Research International Limited	X	X			
Kendle	X	X	X		
LCG Bioscience	X	X			
Lovelace Respiratory Research Institute	X		X	X	
MDS	X	X	X		
Medeval Limited	X	X	X		
Northwest Kinetics	X	X	X		
PAREXEL	X		X		
Pharma Bio-Research International	X	X			
PPD, Inc.	X	X			
PharmaKinetics Laboratories	X	X			
Premier Research Worldwide	X	X			
Quintiles Transnational	X	X			
Research Testing Laboratory			X	X	X
Simbec Research	X	X			
TNO Pharma Institute	X	X			
Valorum	X	X			
West Pharmaceutical Services	X	X	X	X	X

^aPhase I unit with number of beds available listed, where known.

^bBioanalytical chemistry support and pharmacokinetic and special population studies conducted.

Table 5 CSOs providing multiple and various clinical trial support services

CSO	Phase	CTD ^a	SM ^b	DM/B ^c	RW ^d	CS ^e	A/RA ^f
Abt Associates Clinical Trials	I–IV	X	X	X	X		X
Advanced Biomedical Research	I–IV	X	X	X	X	X	
Analytical Sciences	I–IV		X	X	X		X
Anapharm	I–IV	X	X	X	X		X
Applied Analytical Industries	I–IV	X	X	X	X	X	X
Applied Logic Associates	I–IV	X	X	X	X		X
A.R. Kamm Associates	I–IV	X	X	X	X	X	X
Beardsworth Consult. Group	I–IV	X	X	X	X		X
Biomedical System	I–IV	X	X	X	X	X	
BZT	I–IV	X	X	X	X	X	X
Boston Biostatistics	II–IV	X	X	X	X		X
Cato Research	I–IV	X	X	X	X		X
Certus International	I–IV	X	X	X	X		X
Chiltern International	I–IV	X	X	X	X		X
Clinical Data Care	I–IV	X	X	X	X	X	X
Clinical Investigation Support	I–IV	X	X	X	X		X
Clinical R&D Services	I–IV	X	X	X	X		X
Clinical Research Services Turku	I–IV	X	X	X	X	X	
Clinimetrics Research Associates	I–IV	X	X	X	X		X
ClinPharm International	I–IV	X	X	X	X		X
ClinTrials Research	I–IV	X	X	X	X	X	X
Colorado Medical Research Center	I–IV	X	X	X	X		X
Covalent Group	II–IV	X	X	X			X
Covance	I–IV	X	X	X	X	X	X
CroMedica Global	I–IV	X	X	X	X	X	
Drug Research and Analysis Corporation	I–IV	X	X	X	X		X
FARMOVS	I–IV	X	X	X	X		X
Health Decisions	I–IV	X	X	X	X		X
IBAH	I–IV	X	X	X	X	X	X
ICON	II–IV		X	X	X	X	X
ILEX Oncology	I–IV	X	X	X	X		X
Innovus Research	II–IV	X	X	X	X		X
InSite Clinical Trials	II–IV	X	X	X	X		X
Integrated Research	II–IV	X	X	X	X		
Inveresk Research International Limited	I–III	X	X	X	X		X
IST Studien Therapeutica	II–IV	X	X	X	X	X	X
Kendle	I–IV	X	X	X	X	X	X
LCG Bioscience	I–IV	X	X	X	X	X	X
Lineberry Research Associates, L.L.C.	I–IV	X	X	X	X		X
MDS Pharmaceutical Services	I–IV	X	X	X	X	X	X
MEDDOC ApS	II–IV	X	X	X	X		X
Medical Industries Corporation	I–IV	X	X	X	X		X
MEDISEARCH International	I–IV	X	X	X	X		
MediTech International Company, Ltd.	I–III	X	X	X	X		X
METRONOMIA	I–IV	X	X	X	X		X
mimc-International Medical Consultants	II–IV	X	X	X	X	X	X
MTRA	II–IV	X	X	X	X		X
National Institute of Clinical Research	I–IV	X	X	X			X
OMEGA Contract Research Organization	III–IV	X	X	X	X		
Paragon Biomedical	II–IV	X	X	X	X	X	X
PAREXEL	I–IV	X	X	X	X	X	X

(Continued)

Table 5 CSOs providing multiple and various clinical trial support services (Continued)

CSO	Phase	CTD ^a	SM ^b	DM/B ^c	RW ^d	CS ^e	A/RA ^f
Pharma Bioresearch Int'l.	I–III	X	X	X	X		X
PPD, Inc.	I–IV	X	X	X	X	X	X
Pharmaceutical Research Associates	I–IV	X	X	X			X
PharmaKinetics Laboratories	I–IV	X	X	X	X		X
Pharmanet	II–IV	X	X	X	X		X
PharmaPart	I–IV	X	X	X	X	X	X
PharmaResearch	I–IV	X		X	X		X
Precision Research	I–IV	X	X	X	X	X	
Premier Research Worldwide	I–IV	X	X	X	X		X
ProTrials Research	I–IV	X	X	X	X	X	X
PSI Pharma Support	I–IV	X	X	X	X	X	X
Quintiles Transnational	I–IV	X	X	X	X	X	X
Research Services	I–IV	X	X	X	X	X	X
Research Testing Laboratories	II–IV	X	X	X	X		X
Schiff & Company	I–IV	X	X	X	X	X	X
SCIREX	I–IV	X	X	X	X	X	X
SGS Biopharm – Lab Simon	I–IV	X	X	X	X	X	X
Spadille ApS	I–IV	X	X	X	X	X	X
STATPROBE	I–IV	X	X	X	X		X
TOP Clinical Research	I–IV	X	X		X		X
Valorum	I–IV	X	X	X	X	X	X
Westat	I–IV	X	X	X	X	X	

^aClinical trial design and investigator and site selection.^bClinical trial site monitoring.^cData management and biostatistics.^dReport writing and other document preparation.^eClinical trial supply management.^fAuditing and regulatory affairs.

writing, auditing, and regulatory affairs. CSOs have become major players in the clinical trial process. Their support ranges from designing and conducting a phase I clinical trial at a CSO's investigational site to coordinating a multinational, multisite phase III clinical trial. CSOs offering clinical trial services can be subdivided into two major classifications: those that actually conduct clinical trials and those that provide various clinical trial services. Table 4 lists a few CSOs that conduct clinical trials, and Table 5 lists some of the many CSOs that offer clinical trial support services in a variety of areas.

Phase I clinical trial studies include the initial administration of an investigational new drug to humans, usually healthy male volunteers, because the objectives of the study are nontherapeutic. However, women are now frequently included in these studies of first-time-use in humans. Close attention is given to any adverse events (AEs) that occurs and to whether these events are related to the drug candidate and dose level, or are adverse drug

reactions (ADRs). An ICH guideline (18) provides harmonized definitions for various terms including AE, ADR, serious adverse events (SAEs), and unexpected ADRs. A common practice in the first phase I clinical trial is to collect blood and urine specimens to obtain preliminary pharmacokinetic information in humans. Additional phase I clinical trials are frequently conducted later to address specific concerns such as relative bioavailability comparison when the formulation is changed; effect of food and time of feeding; potential for drug-drug interactions; pharmacokinetics in subpopulations such as the elderly, children, and ethnic groups; and possible change in clearance in renally or hepatically impaired patients.

Phase II clinical trial studies are designed primarily to explore the relationship between the dose level and frequency of administration and the efficacy and safety observed in patients with a particular therapeutic indication or disorder. Normally, a primary endpoint is

selected to evaluate the efficacy; however, secondary endpoints are often included to establish criteria for monitoring patients in the more definitive phase III clinical trials. Phase II clinical trials are commonly placebo-controlled and blinded and may include a comparator drug. An ICH guideline (19) describes a number of possible study designs and also discusses the appropriateness of each design for the patient population and the disease indication and for obtaining the desired information. The study designs include parallel dose-response in which patients are randomized to several fixed dose groups, factorial parallel dose-response in which combination therapy is evaluated, cross-over dose-response in which each patient receives each dose level, forced titration in which all patients are administered a series of rising doses, and optional titration in which patients are titrated over a range of dose levels until they reach a predefined favorable or unfavorable response.

Phase III clinical trial studies are designed to confirm that the drug candidate is safe and effective in patients with the intended therapeutic indication and are conducted in a relatively large patient population and frequently at multiple sites. The size and duration of these studies make them, along with establishing the manufacturing facility, the most time-consuming and expensive aspects of drug development. Thus, phase III clinical trials need to be carefully planned and implemented so that the resulting database of generated safety and efficacy data is sufficient to demonstrate statistically that the drug candidate can effectively mediated the disease indication without causing substantial ADRs. For phase III clinical trial studies intended to show effective treatment for a chronic disorder, an ICH guideline (20) has been issued. Still another ICH guideline (21) discusses developmental approaches for drug candidates likely to be used in the elderly because the indication is a disease of aging, such as Alzheimer's disease, or the population with the disorder includes a significant number of geriatric patients. As with most pharmaceutical companies, CSOs do not have the necessary resources at their own facilities to actually conduct multisite, multinational phase III clinical trials.

Phase IV clinical trials start after a regulatory agency has granted a marketing authorization to a drug candidate, which is now a therapeutic product. These phase IV studies need to relate to the approved indication(s) and the dosage regimen(s) that were authorized and are intended to optimize the use of the drug. Phase IV clinical trials may include epidemiology, drug-drug interaction, and sub-population studies.

Not all clinical trials, especially large, multisite, multinational phase III studies and phase IV postmarketing surveillance, pharmacoeconomic, and quality-of-life

studies, can be conducted at a CSO facility. These types of studies, and many phase II efficacy studies, are conducted in research- or university-based hospitals or other investigational sites where a sufficient patient population with the disease or disorder to be tested is available. A number of CSOs offer services to support clinical trial studies that are implemented at one or more clinical trial sites. These services can be broken down into relatively broad categories, which are summarized later.

CSOs assisting clients with clinical trial design need to be an extension of the client's project team and have access to all the data to assist in developing a realistic clinical trial package to determine the safety and efficacy of a drug candidate. In addition to assisting clients with clinical development plans and preparing clinical trial protocols, many of these CSOs also offer clients services for the preparation of CRFs, investigator brochures, IRB and IEC applications, and other clinical trial-related documents. Unless a clinical trial is being conducted at a predetermined site, selection of the clinical site(s) and the investigator(s) who will be responsible for the clinical trial is the next step. The pharmaceutical company is ultimately responsible for all research studies conducted during the development of a drug candidate. However, the clinical trial investigator is a key player (22) and has to have certain qualifications to assume a number of responsibilities. The next step is training of site personnel in the specific aspects and requirements of the study protocol. This training can be critical, especially when the study uses specialized or novel techniques for evaluating safety and/or efficacy. Clinical trial monitoring assists in ensuring that difficulties or problems are detected early and that their occurrence or recurrence is minimized. Overseeing the quality of a clinical trial involves checks of various aspects that may include ascertaining whether the study protocol and SOPs are being followed, whether GCP regulations and other regulatory agency regulations are being followed, the acceptability of data being generated and listed on CRFs and other clinical trial documents, the success in reaching planned patient accrual targets, success in keeping patients in the trial, and tracking clinical trial supplies. An ICH guideline (22) offers guidance on the various aspects of monitoring.

The proper management of the drug candidate and other products, such as comparator drugs and placebos, is in many respects as important to the success of a clinical trial as the care and management of the patients. The clinical trial supplies need to be appropriately characterized, manufactured in compliance with GMP regulations, and, if applicable, coded and labeled to protect the blinding of the

clinical trial study (22). Outsourcing to CSOs or other companies of manufacturing and formulation development research is a common practice. This outsourcing does not relieve the clinical trial sponsor of responsibility for ensuring clinical trial supplies meet the requirements for the proper conduct of a clinical trial.

The primary focus of the data collection and management process is to provide a guaranteed, quality-assured, final database, which is accomplished by standardized strategies for handling data, verification of data, comparing and repairing data entries, and verifying the accuracy of the data. Inaccurate or incorrect clinical trial data can result in an otherwise successful clinical trial being erroneously interpreted for safety and/or efficacy parameters. Three ICH guidelines (18,23,24) address safety data management issues. The data collected from a well designed, well conducted clinical trial are evaluated to ascertain the benefit-to-risk relationship for the purposed therapy. When the patients who received a drug candidate have the disease manifestations completely eradicated and experience no other effects while patients treated with a placebo have a continuation of the disease process, the evaluation is not difficult. However, that is rarely, if ever, the case, and evaluation requires detailed statistical analysis of the collected data. An ICH guideline (25) covers statistical issues related to the scope of clinical trials, design techniques to minimize bias, types of clinical trial designs, conduct considerations, data analysis for efficacy, evaluation of safety and tolerance, and reporting.

Pharmaceutical companies are constantly bombarded with advertisements and brochures from CSOs and other groups that promise unrealistic speed in clinical trial design and management to expedite clinical development. What is missing from this picture? The compiling of the nonclinical and clinical findings in a detailed, well designed, believable story that results in the production of a high-quality regulatory agency submission for marketing authorization. This effort requires the expertise of scientific and medical writers who use clear, concise writing to share the science of drug development. All data and results, both positive and negative, generated during the development of a drug candidate need to be included. An ICH guideline (26) offers assistance in the generation and compilation of clinical trial study reports that are acceptable to all ICH region regulatory agencies. Many CSOs offer scientific and medical writing services, and some consider their medical writing capabilities to be more than just another service offered.

The audit of a clinical trial is independent and separate from routine clinical trial monitoring and quality-control functions. An audit is an aspect of quality assurance and evaluates clinical trial conduct and compliance with the

study protocol; applicable SOPs; GCP, GLP, and GMP regulations; and applicable regulatory agency regulations (22). A clinical trial auditor is to be independent from the clinical trial. Because a clinical trial auditor(s) cannot have been involved in the study, clients who use a CSO to support the various aspects of clinical drug development and then contract with the same CSO to audit clinical trials need to ensure that the auditor(s) is truly independent. To achieve this goal, clients may contract with one CSO to support the clinical trial and another CSO or a qualified independent consultant(s) to conduct the audit. An estimated 80% of the regulatory agency approval process focuses on the clinical trial sections of a marketing application submission. To ensure that these, and other, sections are complete and meet the perceived specifications of the regulatory agency reviewers, most pharmaceutical companies have established regulatory affairs groups. These groups commonly assist drug development project teams and company management in defining and implementing regulatory strategies for the various aspects of drug candidate development, offer advice and recommendations on regulatory issues applicable to the countries where marketing approval is being sought, and serve as the primary contact between the company and regulatory agencies. Some CSOs offer regulatory affairs capabilities.

OUTSOURCING OF CONTROL, MANUFACTURING, AND CHEMISTRY

After clinical trials, the next most time-consuming and expensive aspect of the drug development process is the synthesis and characterization of the drug substance and drug product, i.e., the manufacturing process. A manufacturing process that is inappropriately designed, implemented, and controlled can kill a promising drug candidate or active pharmaceutical ingredient (API), which is the term used by most manufacturing groups, as easily as can life-threatening adverse effects or insufficient biological activity. A drug candidate that is costly to effectively produce results in high prices to patient, that, unless the drug is the best or the only therapy for a disease indication, lowers the market share and thus the revenue to the pharmaceutical company. Similarly, a complex manufacturing process that is difficult to control may result in API variations, in both amount and impurity profiles, that can hinder regulatory agency approval unless the manufacturer can prove that the production processes are being following and are producing the same material in each production batch.

The manufacturing process is frequently changed to improve yield or decrease impurities. To ensure the drug substance is the same after processing changes, manufacturers develop and validate characterization procedures to demonstrate and document that the changes have not affected the material being produced and that no previously undetected impurities are present.

Many pharmaceutical companies outsource one or more of the control, manufacturing, and chemistry (CMC) processes, for which numerous regulatory agencies guidelines are available. Reasons for this outsourcing include but are not limited to:

- Avoiding capital investment in facilities, equipment, and personnel;
- Avoiding the need to establish and maintain a GMP manufacturing facility;
- Gaining access to additional capacity and technology;
- Accessing a contractor's expertise; and
- Expediting the time to market

Some CSOs that offer multiple CMC services to clients are listed in Table 6. These services have been subdivided into the major categories of manufacturing, raw materials, formulation development, method development, stability, packaging, and auditing.

The manufacturing process starts when the discovery scientists synthesizes a compound that has biological activity. The techniques used are usually sufficient to produce milligram to gram quantities for pharmacology studies but are rarely adaptable to the large-scale synthesis necessary for generating the much larger amounts of material needed for formulation development and preclinical and clinical studies. Shortly after a discovery lead is designated a preclinical drug candidate, the manufacturing group begins the research studies necessary to scale up the laboratory bench synthetic procedure to a larger, or manufacturing, process. Whether the drug substance is a small organic molecule or a macromolecule, the procedure may and usually does undergo substantial modification during scale up to commercial production. A few of the items that are evaluated during the establishment of a manufacturing process include:

- Identifying synthesis and formulation methods, including definition of equipment, processes, and scale of production;
- Defining and validating processes for fill, finish, and packaging;
- Determining procedures for waste disposal, including appropriate environmental assessments; and
- Determining the number of batches required to support development program and product launch.

One of the most important aspects of the manufacturing process is documentation and document control. Each step in the process, whether the results are positive or negative, needs to be evaluated and documented to show that the manufacturer understands what is happening and can control the synthesis, purification, formulation, and packaging of the drug product. Some of the documents needed include:

- Chemistry, manufacturing, and control documents;
- Appropriate standard operating procedures;
- Validation protocols for methods and processes;
- Progress and final reports for methods and processes; and
- Regulatory documents, such as IND, yearly updates, and NDA.

The raw materials used in the synthesis, purification, and formulation of an API are critical for the successful manufacture of a drug product. The supplier or vendor of each component and intermediate has to be identified and certified, commonly through a detailed site visit. The amount, timing, and cost of these supplies are negotiated so that the key ingredients are available when needed. Another important item is determining the shelf-life of each raw material to ensure that the ingredient has not deteriorated from the time of purchase to the time of use.

Once the drug candidate is available, formulation activities are initiated to develop a suitable dosage for administration to animals and, later, to humans. Depending on the route of administration, formulations may be solutions, suspensions, aerosols, creams, or solids. Because formulation ingredients can affect the delivery profile of a drug candidate, in some cases adversely, using a preclinical formulation that closely resembles and mimics the proposed clinical formulation is highly desirable. Thus, preformulation studies need to be conducted as early in the development process as possible. The formulation process usually continues throughout the development process, with changes made to improve stability, enhance delivery, or mask the taste or appearance of the drug product. Depending on the extent of the formulation change, additional nonclinical and clinical studies may be necessary to determine if the revised formulation has the same characteristics as the original formulation and, if not, what the extent and nature of the difference are.

Analytical chemistry methods are used for chemical characterization of the drug substance and drug product. This characterization includes qualitative tests for structural identification and for impurities, degradation products, and contaminants (27–29). Characterization relies on well characterized, validated, analytical

Table 6 CSOs that provide multiple and various CMC services

CSO	Manu ^a	RM ^b	Form ^c	MD ^d	Stab ^e	Pack ^f	Aud ^g
AAI	X		X	X	X		
Abbott Laboratories	X			X		X	
Alpharma USPD	X		X	X	X	X	
AMRESKO	X	X	X		X	X	
Atlantic Pharmaceutical Services	X	X	X	X			
Ben Venue Laboratories	X		X	X	X		
BioAnalytika Laboratories				X	X		
Boston Analytics				X	X	X	
Catalytical Pharmaceutical	X	X	X	X	X	X	X
Charles River Laboratories				X	X		
Chesapeake Biological Labs	X		X	X		X	
Chromak Research		X	X	X	X		
Circa Pharmaceutical	X	X		X	X	X	
Collaborative BioAlliance, Inc.	X	X	X				
Covance, Inc.		X		X	X	X	X
Dow Chemical Manufacturing Services	X	X	X	X	X	X	
DPT Laboratories	X		X	X	X	X	
Elemental Research		X		X	X		
Gibraltar Laboratories		X	X	X	X		
Global Pharm Inc.	X	X		X	X	X	
Hauser, Inc.	X		X	X	X	X	X
IBAH Pharmaceutical Services	X	X		X	X	X	
IDEC Pharmaceutical Corporation	X	X		X			
International Processing Corporation	X		X	X	X	X	
J.B. Laboratories	X			X	X	X	
Kansas City Analytical Services		X		X	X		
Lancaster Laboratories		X		X	X		
Magellan Laboratories		X		X	X		
MDS Pharmaceutical Services	X	X	X	X	X	X	
Metrics, Inc.	X	X	X	X	X		
Metuchen Analytical		X		X	X		
Midwest Research Institute	X	X	X	X	X		
New Life Resources	X	X		X		X	X
Nycomed Amersham	X			X	X		
Oneida Research Services		X		X	X		
Oread	X		X	X	X	X	X
Patheon, Inc.	X	X	X	X	X	X	
Performance Solutions							X
Pharm-ECO Laboratories	X	X		X	X		X
Pharmaceutical Development Center	X	X	X	X	X	X	
Pharmaceutics International	X		X	X	X	X	
Pisgah Labs	X	X		X			
PPD, Inc.		X		X	X		
Proceutics				X	X		X
ProClinical Pharm. Services	X			X	X	X	
Quantitative Technologies		X		X	X		
Quality Chemical Laboratories		X		X	X		
Ricerca	X	X		X	X		
Schwarz Pharma	X			X		X	
SGS U.S. Testing		X		X	X	X	
Sharp		X			X	X	

(Continued)

Table 6 CSOs that provide multiple and various CMC services (*Continued*)

CSO	Manu ^a	RM ^b	Form ^c	MD ^d	Stab ^e	Pack ^f	Aud ^g
Shuster Laboratories		X	X	X	X		X
Southern Testing & Research Labs		X		X	X		
SP Pharmaceuticals	X		X		X	X	
SRI International	X	X	X	X	X		
STAT-A-MATRIX				X	X		X
Steifel Research Institute		X	X	X	X		
Taylor Pharmaceuticals	X	X		X	X	X	
Tetrionics, Inc.	X			X	X		
Tower Laboratories	X		X		X	X	
Vital Pharma, Inc.	X		X	X	X	X	
West Coast Analytical Services		X	X	X	X		
West Pharmaceutical Services	X			X	X	X	

^aManufacturing including process development and validation.^bRaw material including procurement and testing.^cFormulation development and characterization.^dMethod development for drug substance and drug product characterization.^eStability testing of drug substance and drug product.^fPacking services for drug product.^gGMP audit of chemistry, manufacturing, and control processes.

chemistry methods, which may include USP methods such as residual solvents, moisture content, residue on ignition, and thermal determinations and drug candidate-specific methods such as identification, concentration, impurity profiles, solubility profiles, and stability under various temperatures and conditions. For proteins and polypeptides, additional characterization is necessary. Several ICH guidelines (30–37) have been issued on characterization and discuss stability testing, including photostability, impurities, residual solvents, and test procedure and acceptance criteria specifications (33, 34). Quantitative analytical chemistry methods provide information on drug substance concentration, impurity profiles, and stability profiles. ICH guidelines (35, 36) for validation of analytical chemistry procedures address experiments to ascertain and set acceptance specifications for specificity, detection limit, quantification limit, linearity, range, precision, accuracy, and robustness. Analytical methods need to be validated for each sample type to be analyzed.

Stability testing determines how a drug substance or drug product may change with time under various conditions, such as temperature, humidity, and light and defines storage conditions and shelf-life. A number of ICH guidelines (30, 32, 37) describe and discuss stability issues and testing requirement for NCEs and macromolecules. These guidelines provide manufacturers and CSOs with acceptance specifications for

accelerated and long-term stability studies. In addition to the drug substance and drug product, stability assessment is frequently conducted on raw materials, key intermediates, formulation excipients, and packaging materials.

After the drug product is prepared, the next step is packaging and labeling the material. Packaging requirements depend on the final drug product (solid, suspension, liquid) and the clinical indication (hospital, physician office, or home use). For liquids to be dosed parenterally, sterility of the drug product is important and validated, and aseptic filling procedures are necessary. For other dosage forms, specialized equipment may be needed to place the drug product into the desired package. Compatibility of the drug product with any packaging material, such as vials, inner liners, and closure systems, with which the product comes into contact needs to be evaluated. Proper labeling and package insert information is critical for the proper use of the drug product. The label and package insert provide physicians, pharmacists, and patients with information on:

- The drug substance, drug product, and formulation excipients;
- Clinical pharmacology and the biological action of the drug substance;
- Indications and usage for which the specific drug product has been approved;

- Contraindications for when the drug product should not be used;
- Warnings and precautions on the use of the drug product;
- Adverse reactions reported and attributable to the drug substance;
- Overdosage effects and treatment to counteract the effects;
- Dosage and administration specifications for the drug product; and
- How the drug product is supplied and appropriate storage conditions.

Audits are conducted to ensure that the manufacturer—either a pharmaceutical company or a CSO—has appropriately defined and documented the various processes and methods required to produce the drug substance and drug product and has the necessary personnel and facilities to conduct and control the manufacturing process and to determine whether GMP regulations are being followed. As with clinical trials, the pharmaceutical company developer is responsible for the manufacture of the drug substance and drug product, whether these activities take place at the company's facilities or are outsourced. Audits, whether conducted by regulatory agencies or by the developer, determine that the processes are under control and in compliance with GMP regulations. Problems discovered during audits can usually be resolved, and to prevent regulatory agencies from finding these areas of concern, pharmaceutical companies use their internal QA groups or their contract with the CSO or independent consultant to conduct GMP audits of the CMC aspects of drug development. As with its role in clinical trial auditing, the auditors have to be independent of the processes being audited, and at times, clients outsource auditing of a manufacturing CSO to another CSO that provides GMP auditing services.

VIRTUAL DRUG DEVELOPMENT ORGANIZATIONS AND CONSULTANTS

Another relatively new player in the drug discovery and development arena is the virtual drug development organization (VDO), which, like the CSO, offers pharmaceutical companies support in developing discovery ideas into therapeutic products. However, VDOs do not offer laboratory services, but provide clients with coordination of the various drug discovery and development activities necessary to characterize a drug candidate and produce the compound in sufficient quantity for

nonclinical, clinical, and marketing efforts. This coordination is accomplished by designing and outsourcing the research studies to CSOs and then compiling the results generated into regulatory agency submissions. VDOs operate in one of two ways. One is to contract with the pharmaceutical client and serve as an extension of the client's program. The other is to out-license a drug candidate from a pharmaceutical company or university technology transfer group. In both cases, VDOs interact closely with CSOs. Pharmaceutical firms need to evaluate VDOs using the techniques and requirements used to select and monitor CSOs.

Consultants have supported the pharmaceutical industry for many years and provide expert advice on the many aspects of the drug discovery and development processes. This advice includes assisting in the selection and monitoring of CSOs and in the interpretation of results from outsourced studies. Pharmaceutical companies select consultants in a variety of ways, the most common being the "good-old-boy" network, in which a pharmaceutical company executive knows, or has a friend who knows, someone who is now a consultant, commonly a retired pharmaceutical executive, a former regulatory agency reviewer, or a pharmacology professor at a university. Is this the optimal means of identifying an expert? Probably not, if the company wants an expert with knowledge and experience in a particular area(s) research, such as drug metabolism or drug product stability. The techniques outlined above for selecting a CSO should also be used for evaluating consultants, with research area experience, integrity, and communication skills probably being the most important aspects. Choosing a toxicology or clinical pharmacology consultant to assist with evaluating human pharmacokinetic results may prevent the data from being used effectively to design additional nonclinical toxicology or clinical trial studies.

EMERGING TRENDS

Major changes have occurred, and are occurring, in the pharmaceutical industry and in the drug discovery and development processes. They have caused numerous changes in the way CSOs provide support services to drug developers. What can be expected in the future? Are more changes on the horizon, and, if so, what might they be and how will they affect pharmaceutical companies and their service providers? Although predicting the future is usually quite inaccurate and often misleading, some trends, based on recent events, may be extended into the near future.

For pharmaceutical companies, present trends indicate that:

- Mergers and acquisitions will result in formation of more mega-pharmaceutical companies. Reasons include the desire to have a more global presence, to enter into a therapeutic area with previously established research staff and drug development pipeline, and to obtain novel technology to assist in accelerating drug discovery and development. Whether these mega-firms have a better chance than midsized or small pharmaceutical houses in obtaining marketing approval for major therapeutic products is unknown.
- Numerous new companies, primarily biotechnology firms that hope to turn a discovery idea into a therapeutic product, will be formed. Funding for these firms will be difficult to obtain, will come primarily from venture capitalists and government grants, and will be insufficient to take a discovery lead through complete drug development. Most of these start-up companies will attempt to outlicense or sell for much-needed cash and future royalties their discoveries to pharmaceutical companies or VDOs. The further these small companies are along in the drug development process will determine how much they will receive for their discoveries. A number of companies, approximately the same number that are formed, will go out of business, merge, or be acquired when their discovery does not result in drug candidates with the necessary attributes to become therapeutic products.
- The ICH process will provide more standardization to the requirements for the worldwide development of drug candidates. A global marketing application package that meets the requirements of most regulatory agencies and that includes nonclinical, clinical, and manufacturing aspects will be defined and accepted.
- Even the best efforts of drug developers will not reduce the time and cost of drug discovery and development. However, as novel drug discovery and potential for development characterization techniques are put into place and the ICH process is more fully implemented, the time and cost will not increase as rapidly as in the past.

For CSOs and other support groups, the trends noted previously offer nothing but opportunities. The prospects for CSOs are very bright, and CSOs that capitalize on the present and future needs of their clients will be highly successful. Trends for the CSO industry include that:

- Pharmaceutical companies will increase their outsourcing efforts;
- Drug discovery and potential for development characterization services will be high-growth areas because these fields are presently bottlenecks, and the situation is not expected to get any better;
- CSOs offering primarily or exclusively nonclinical drug development services will have to broaden their capabilities to attract and maintain clients,
- Clinical CSOs will have more studies than they can support effectively and will most likely expand to meet the needs of their clients,
- Clinical CSOs that provide services in a number of aspects of clinical trials and have a worldwide presence will grow and prosper,
- More mega-CSOs, such as Quintiles and MDS, will emerge. Likely candidates are these mega-CSOs will offer clients many, but probably not all, services needed for drug discovery and development. Some of these large CSOs may attempt to become pharmaceutical houses and use their considerable expertise to discover and develop compounds obtained through their own efforts or by inlicensing,
- The CSO field will not be dominated by the mega-CSOs. The present and future bottlenecks experienced by pharmaceutical companies will result in more CSOs, not fewer. These new CSOs will offer clients specialized services in drug discovery and potential for development characterization to evaluate multiple discovery leads simultaneously,
- As pharmaceutical companies merge and continue to "right-size," CSOs will become more a true partner and less a pair of hands to complete tasks. The clients will have to share more scientific information, both positive and negative, to the service providers that, in turn, will need to learn more about the entire drug discovery and development process for these partnerships to be effective and beneficial for both parties.
- CSOs will be a primary source of employment for new graduates and for down-sized pharmaceutical company employees in various scientific disciplines. The new graduates will bring in the latest academic technology to assist in expanding or defining novel services to be offered to clients. These young, highly motivated employees will learn the more applied science of pharmaceutical development from the former pharmaceutical company employees.

Whether many or all of these potential trends will occur in the near future is not known. One thing is certain, change will happen. Those pharmaceutical companies and their CSO partners that embrace change and adapt to novel

ideas and approaches to improve and possibly shorten the drug discovery and development processes will prosper. Companies that do not or are unwilling to embrace change will most likely become a memory. However, even a bad memory can offer guidance and be used to improve the process for finding novel therapeutic agents to treat one of the many human diseases or disorders that presently have only marginal or no effective treatment.

REFERENCES

1. *Drug Metabolism/Drug Interaction Studies in the Drug Development Process. Studies In Vitro*; US FDA: FDA Guidance for Industry: Washington, DC, 1997.
2. U.S. FDA's Proposed Implementation of ICH Safety Working Group Consensus Regarding New Drug Applications. Federal Register **1992**, 57 (73), 13105–13106.
3. *Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*; ICH Harmonized Tripartite Guideline (M3); 1997.
4. Karnes, H.T.; Shiu, G.; Shah, V.P. Validation of Bioanalytical Methods, *Pharm. Res.* **1991**, 8 (4), 421–426.
5. Shah, V.P.; Midha, K.K.; Dighe, S.; McGilveray, I.J.; Skelly, J.P.; Yacobi, A.; Layloff, T.; Viswanathan, C.T.; Cook, C.E.; McDowall, R.D.; Pittman, K.A.; Spector, S. Analytical Methods Validation, Bioavailability, Bioequivalence and Pharmacokinetic Studies. *Pharm. Res.* **1992**, 9 (4), 588–592.
6. *Note for Guidance on Toxicokinetics: The Assessment of Systemic Exposure in Toxicity Studies*, ICH Harmonized Tripartite Guideline (S3A); 1994.
7. *Pharmacokinetics: Guidance for Repeated Dose Tissue Distribution Studies*, ICH Harmonized Tripartite Guideline (S3B); 1994.
8. *Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals*, ICH Harmonized Tripartite Guideline (S2A); 1995.
9. *Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals*, ICH Harmonized Tripartite Guideline (S2B); 1997.
10. *Duration of Chronic Toxicity Testing in Animals (Rodent and Non Rodent Toxicity Testing)*, ICH Harmonized Tripartite Guideline (S4A); 1998.
11. *Detection of Toxicity to Reproduction for Medicinal Products*, ICH Harmonized Tripartite Guideline (S5A); 1993.
12. *Toxicity to Male Fertility: An Addendum to the ICH Tripartite Guideline on Detection of Toxicity to Reproduction for Medicinal Products*, ICH Harmonized Tripartite Guideline (S5B); 1995.
13. *Guidelines on the Need for Carcinogenicity Studies of Pharmaceuticals*, ICH Harmonized Tripartite Guideline (S1A); 1995.
14. *Testing for Carcinogenicity of Pharmaceuticals*, ICH Harmonized Tripartite Guideline (S1B); 1997.
15. *Dose Selection for Carcinogenicity Studies of Pharmaceuticals*, ICH Harmonized Tripartite Guideline (S1C); 1994.
16. *Addendum to Dose Selection for Carcinogenicity Studies of Pharmaceuticals: Addition of a Limit Dose and Related Notes*, ICH Harmonized Tripartite Guideline (S1CR); 1997.
17. *General Considerations for Clinical Trials*, ICH Harmonized Tripartite Guideline (E8); 1997.
18. *Clinical Safety Data Management: Definitions and Standards for Expedited Reporting*, ICH Harmonized Tripartite Guideline (E2A); 1994.
19. *Dose-Response Information to Support Drug Registration*, ICH Harmonized Tripartite Guideline (E4); 1994.
20. *The Extent of Population Exposure to Assess Clinical Safety for Drug Intended for Long-Term Treatment of Non-Life-Threatening Conditions*, ICH Harmonized Tripartite Guideline (E1); 1994.
21. *Studies in Support of Special Populations: Geriatrics*, ICH Harmonized Tripartite Guideline (E7); 1993.
22. *Guideline for Good Clinical Practice*, ICH Harmonized Tripartite Guideline (E6); 1996.
23. *Data Elements for Transmission of Individual Case Safety Reports*, ICH Harmonized Tripartite Guideline (E2B); 1997.
24. *Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs*, ICH Harmonized Tripartite Guideline (E2C); 1996.
25. *Statistical Principles for Clinical Trials*, ICH Harmonized Tripartite Guideline (E9); 1998.
26. *Structure and Content of Clinical Study Reports*, ICH Harmonized Tripartite Guideline (E3); 1995.
27. *Impurities in New Drug Substances*, ICH Harmonized Tripartite Guideline (Q3A); 1995.
28. *Impurities in New Drug Products*, ICH Harmonized Tripartite Guideline (Q3B); 1996.
29. *Impurities: Guideline for Residual Solvents*, ICH Harmonized Tripartite Guideline (Q3C); 1997.
30. *Stability Testing of New Drug Substances and Products*, ICH Harmonized Tripartite Guideline (Q1A); 1993.
31. *Stability Testing: Photostability Testing of New Drug Substances and Products*, ICH Harmonized Tripartite Guideline (Q1B); 1996.
32. *Stability Testing: Requirements for New Dosage Forms, Annex to the ICH Harmonized Tripartite Guideline on Stability Testing for New Drugs and Products*, ICH Harmonized Tripartite Guideline (Q1C); 1996.
33. *Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances*, ICH Harmonized Tripartite Guideline (Q6A); 1997.
34. *Specifications: Test Procedures and Acceptance Criteria for Biotechnology/Biological Products*, ICH Draft Consensus Guideline (Q6B); 1998.
35. *Text on Validation of Analytical Procedures*, ICH Harmonized Tripartite Guideline (Q2A); 1994.
36. *Validation of Analytical Procedures: Methodology*, ICH Harmonized Tripartite Guideline (Q2B); 1996.
37. *Quality of Biotechnological Products: Stability Testing of Biotechnological/Biological Products*, ICH Harmonized Tripartite Guideline (Q5C); 1995.
38. Corporate Capabilities. *Bio. Pharm* **1999**, 12 (12), 1–76.
39. The 2000 CSO Directory. Drug Information Association **2000**, 1–237.
40. *2000 Corporate Capabilities*, LC-GC; North America 2000, 1–138.

41. GEN Directory to Contract Research Organizations, Genetic Engineering News Directory to Biotechnology Companies.
42. 1999/2000 Buying Guide. Pharm. Process. **1999**, 16 (6), 1–224.
43. *Larka Directory of Contract Pharmaceutical Manufacturers (CPMs) Europe*, Larka Directory of Contract Research Organization (CROs) Europe. Larka Pharmaceutical Publications.
44. *The Directory of Contract Services; Vol. I. Dosage Form Manufacturing, Formulation, Packaging and Analytical Services; Vol. II.*, Biologics API Manufacturing and Process Development; PharmSource Information Services, Inc.
45. *The Technomark Registers 2000; Vol. 1. CROs in Europe Offering Clinical Research Services, Data Management & Regulatory Affairs; Vol. 2. CROs in Europe Offering Preclinical, Toxicology & Analytical Services; Vol. 3. CPMs in Europe Offering Packaging & Manufacturing Services; Vol. 4. North American CROs & CPMs.*
46. Lakings, D.B. Biological CROs: Enhancing the Opportunity for a Successful Drug Discovery and Development Program. D&MD Reports, 1999; 1-1–10-348.
47. *New Drug Approval Process, The Global Challenge. Drugs and the Pharmaceutical Sciences*, 3rd Ed.; Guarino, R.A., Ed.; Marcel Dekker, Inc.: New York, 2000; 1–463.

PAPERLESS DOCUMENTATION SYSTEMS

Ellen M. Williams

Michael McKenna

Pfizer, Inc., New York, New York, U.S.A.

INTRODUCTION

The fundamental purpose of documents is communication. Specifically, documents enhance the flexibility of communication. Documents improve the ability to transmit information in two ways. First, a document freezes a sender's message in time so that it may be accessed later. The wall of a cave marked with pictographs, a carved stone tablet, a photograph of the cave wall, a creased paper bag containing a shopping list, as well as a phonographic recording of a piano concerto all record information that persists through time beyond the act of its creation. The second way in which documents make our communications more versatile is by allowing the sender's message to travel through space independently of the sender. With the exception of the cave wall, each of the above examples is more or less portable. It is portability that gives documents the capacity to communicate across space as well as time.

By packaging a sender's message in such a way that it may be accessed by physically scattered recipients at any time after the original recording, documents enable marvelous things to happen. While proscribing real-time exchange between a message's sender and recipient, a physical record gives to ideas a life that they often would not have if merely spoken or performed once in time. Governments, religions, and much of society depend crucially on the existence of an ongoing exchange with certain key documents (1). On a smaller and more immediate scale, documents enable us to get things done in ways and on a scale that spoken words alone do not. Indeed, by allowing us to communicate across vast reaches of time and space, documents increase the coordination and scope of communication, thereby enabling tremendous feats of collaboration.

Perhaps more than other industries, pharmaceutical manufacturing and marketing depend on exquisite levels of collaboration among many different spheres of activity, ranging from analytical chemistry to clinical medicine, operations research, business, and the law. This complex coordination and a strict regulatory

landscape, which ultimately dictates the viability of a pharmaceutical product, demand that all scientific and business decisions from discovery through postmarketing research be rigorously and thoroughly documented. The sheer volume of written procedures, forms, analyses, reports, and correspondence that must be managed across a single product's lifecycle is enormous. When one considers enterprise-wide management of documentation supporting a portfolio of dozens of drugs in various stages of development, the numbers are truly staggering. Although regulatory authorities have recently issued comprehensive regulations (2–4) and guidelines (5–7) on electronic records management, these standards and the predicate rules on which they are based could rapidly become outdated in the evolving world of computing and wireless data exchange (8).

Until recently, document management in an enterprise has been organized around distinct efforts involving the control of three kinds of paper document (forms, reports, and manuals) and the administration of the folders and archiving procedures used to manage them (9). A number of economic and technological forces over the last quarter century have changed business in such a way as to strain the traditional methods of document management, if not render them wholly obsolete. For one thing, improvements in organizational, manufacturing, and communications technology that have compressed production cycles and decreased the time from concept to market in nearly all industries have also yielded a corresponding acceleration in demand and consumption. With improvements in communications and techniques of distribution come unavoidable increases in the volume of business transactions (10). Speeded up markets beget more business, which necessitates more efficient management of ever more documentation. As computer technology has been introduced into business environments over the last 25 years to adapt to this quickening pace, the system designed to manage paper has met with great difficulty storing electronic documents as well (11). Computing and communications technology is changing the fundamental status of documents.

EVOLUTION OF DOCUMENTATION TECHNOLOGIES

Initially, the implementation of a paperless documentation system in an enterprise is driven by a need or mandate to eliminate paper, economize the use of physical space, improve the efficiency of document retrieval, or otherwise improve the storage of information assets. Paper document management has referred to the storage, modification, and retrieval of documents that distinguish themselves from each other in terms of content, purpose, or format. The emphasis is on the introduction of a system that provides a storage solution. The true strength of most electronic document management systems, however, lies beyond the simple capacity to warehouse digital representations of paper documents.

During the 19th and 20th centuries document technologies have followed an evolution that has repeatedly and alternately focused on document production, reproduction, and distribution. The first long wave of innovation was centered around the means of producing text. From as early as the fourteenth century, at least one hundred writing machines were invented before the Sholes and Glidden typewriter was produced by the gunmakers Remington and Sons some sixty miles west of Albany in 1874 (12). The second major development in the history of modern document technology sought to enable the rapid dissemination and reproduction of paper documentation and is marked by the Haloid (later Xerox) company's introduction of the photocopier around 1960 (13). In the mid-1960s, by way of the IBM Selectric typewriter, which had magnetic tape storage, and the text editor, which handily managed unformatted text for machine processing, the third evolutionary step brought the management of all aspects of the composition and formatting of individual documents in the form of the WYSIWYG word processor (14). By the mid-1990s, the word processor had effectively satisfied all needs surrounding the efficient production and stylistic control of single documents within a given operating system or platform. Concurrent with the improvement of word processing technology, much attention was paid to enhancing the distribution of textual information to disparate locations across computer networks. The shift in attention took place on two fronts, looking first to the distribution of formatted content across multiple platform-independent systems and then to managing the storage and retrieval of, as well as the complex relationships among, large volumes of documents. 1990 saw Tim Berners-Lee create the first web browser and a simple markup

language that would become HTML (hypertext markup language) at the European Laboratory for Particle Physics (CERN) to enable collaboration among disparately located scientists (15). This innovation is, in fact, not so new. It was actually presaged nearly half a century earlier by Vannevar Bush in an article written for the *Atlantic Monthly* that described his "memex" machine, which bears an eerily prescient resemblance to today's World Wide Web and the practice of hypertext linking (16). By the mid-1990s innovations in document warehousing and distribution were built atop mature relational database technologies and the ability to efficiently link document objects stored on a server platform with their associated indexing attributes in a database (17).

The evolution detailed above has alternately expanded and narrowed its focus, shifting from the domain of the document itself to the interstices between multiple documents and their authors and consumers. The development emphasis has cycled back and forth between the production of documents and the reproduction and distribution of documents. As we move into the 21st century, the prevailing trend in document technology appears to be moving in the direction of fusing the cycles of document production, reproduction, and distribution. As much as business allows, documents will soon be designed to allow their transmission, reuse, and machine processing in a simple fashion. Indeed, more and more, business models will be forced to adapt to a landscape that operates on such "documents." The need for immediate access to "decisionable data" in drug development, the need to control development costs, the need to shorten development times to protect patent life, the need to retrieve and analyze data for rapid response to regulatory and legal inquiries are but a few scenarios making the case for adapting to the new landscape in the pharmaceutical industry (18).

DOCUMENT MANAGEMENT PRINCIPLES

The traditional records management model is based on cabinets, folders, and files (9). This physical model was given its logical extension in the first electronic document management systems, where files were placed into virtual cabinets and folders.

File Room Model and Security

Security models for documents are all based on controlling who can see documents, who can create or edit documents,

and who can delete documents. Securing these rights is implemented at numerous levels. It is illustrative to consider these in terms of a physical library or paper-based file room (9). First, you may need proper credentials simply to get in and browse the holdings. Second, once you've gained admittance to the filing area, your ability to view certain kinds of records may depend on your job title or departmental affiliation. Third, assuming you have rights to view a specific record, you may have permission only to view the final file (as opposed to a draft) under observation in the file room itself, and you may not be permitted to make a copy. Finally, if you are permitted to check the document out of the file room for a limited time, you will be required to sign your name to a dated logbook. Each of these constraints on your ability to interact with the holdings preserves the likelihood that files will remain unchanged for the next requestor and holds you accountable should they not.

Input–Output Model and Quality Control

Traditional document management rests on a very simple input–output model. An enterprise seeks to manage the storage of documents (input) in such a way that their retrieval (output) is simplified. The real goal is speedy retrieval of documents. Their intelligent storage is the means to achieve it. In most enterprises it will not do to sacrifice accuracy for speed, so the process must first attend to the security, integrity, and authenticity of the data being served. The enterprise must build in mechanisms to control and track input and output processes, which include versioning, searches, and collaborations or workflows. In the paperless environment, speed, intelligent storage, and accuracy of data and searching remain of paramount importance. As the paperless environment grows in volume and complexity, meeting these needs becomes increasingly difficult.

ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS (EDMS)

Goals and Objectives

Electronic document management concepts have provided the bridge between traditional paper-based information management and the threshold of a new information age. The putative goals of moving from a paper-based to a paperless documentation system are to improve business processes and ultimately to enhance enterprise

performance and the bottom line by exploiting the numerous advantages of leaving paper behind (18). The core objectives that support the attainment of those goals include the following:

- Increase efficiency and effectiveness by stewarding the generation of paper and electronic documents within the individual's scope of control;

- Increase productivity by optimally using, reusing, and recycling enterprise documents where possible;

- Increase the consistency of classification, indexing, and retrieval of documents;

- Increase the sharing of documents within an organizational unit, and between and among organizational units, while maintaining necessary access controls;

- Preserve decision and accountability trails for documents;

- Automate the retention and disposal procedures for document review and retirement. Given these goals and objectives for paperless systems (19), it is perhaps helpful to look at the specific business and functional requirements that drive and enable them.

System Requirements

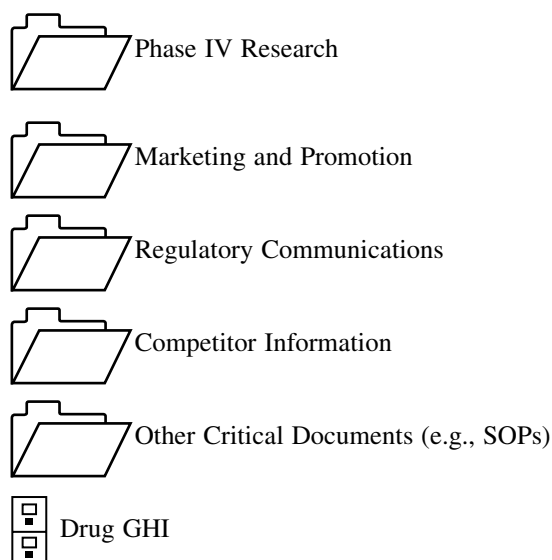
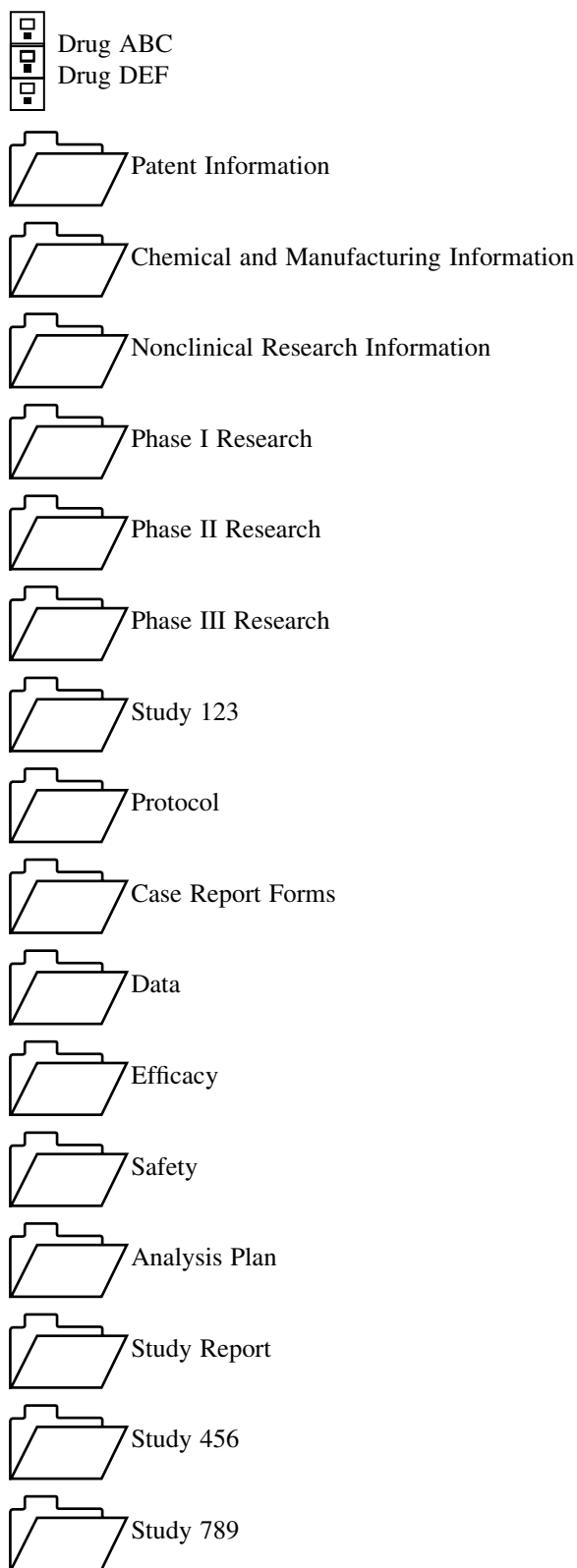
Before implementing any piece of technology in a business setting, it is necessary to define thoroughly both the business and functional requirements for the system. Business requirements describe, independently of technology, the high-level activities and processes that must be performed or enabled and the constraints that must be met by the system. Functional requirements outline in greater detail the lower level operations that must be performed and the constraints on the mechanisms for performing them.

Organization requirements: cabinet/folder structure

The most obvious business requirement for an EDMS is that it should enable the user to easily place and retrieve the items stored within. This means that files should be stored in the repository in an organized hierarchical file structure. Drug documents must be automatically stored in the appropriate location in the file structure based on their identifying information. Rather than having to manually place a file by navigating the entire hierarchy, which may be quite large in an enterprise with hundreds of compounds in various stages of development, a file could be placed following the entry of key content words or key file attributes (indexing information or metadata).

For a system managing documentation pertaining to the full life-cycle of drug development, including the conduct

of clinical trials, a useful cabinet and folder hierarchy might resemble the following:



Input requirements: Creating and importing documents

The true business value of an EDMS is realized to the extent that it serves as more than just the terminus for document-related activities in an enterprise. To fully realize its goals, an enterprise paperless documentation system will perform functions across the entire document life-cycle. In addition to the storage of existing documents, users with appropriate system privileges must be able to create new electronic documents from within the electronic document management system. The system must explicitly support the file formats used by the business, which typically include word processing, spreadsheet, graphics, and database files.

An enterprise EDMS must be flexible enough to accommodate documents that do not originate within the system. It must be possible to import documents whether they are electronic or paper in format. Technologically speaking, there are no constraints on the files that can be stored in modern electronic document management systems. It should be noted, however, that certain file types (e.g., databases) cannot be rendered meaningfully to certain standard immutable display formats (e.g., PDF) and will often be viewed by default as read-only in the authoring application.

Modification requirements: Version control and document lifecycle

The ability to track changes in a fully rationalized manner is central to achieving enterprise objectives related to efficiency, data integrity, and accountability. To facilitate control over the full development lifecycle of documents,

users with appropriate system privileges must be able to check out, modify, and check in documents.

In paper-based archives, questions about a document's authenticity or official status are usually answered by the surrounding policies and institutional mechanisms which dictate that only final and official documents are kept. In an EDMS environment, where document drafts are managed alongside their finalized counterparts, it is important to implement controls to ensure that the current version of a document is retrieved by default and that prior versions are retrieved only on explicit request. This is critical to both archiving and collaboration and workflow. The EDMS must implement a policy of nonreplication of documents and must also judiciously apply permissions to control changes in a document's life-cycle status. The system should distinguish between draft, review, approved, and final versions of a document and implement business rules defining security for the document based on its life-cycle status. For example, when a document is assigned a status of FINAL, users may view the document, but no one is permitted to modify it. The controls afforded by an EDMS should ensure confidence in answering questions about the official status or current version of a document.

Retrieving requirement: attributes and searching

In the digital age, ideas and information will be reprocessed and reused. Their portability and, consequently, their usefulness are dependent on the ability to find their particular data bit streams within a vast universe of binary number sequences. As the mass of documents in an enterprise EDMS grows and the filing structure that accommodates it both broadens and deepens, retrieving any specific item from the repository becomes more difficult. Speedy and accurate search and retrieval are achieved by allowing searches based on document content, attributes, or both. Users must also be able to save their queries for future reuse and store their queries in a centrally administered, single location. A simple FIND function must allow users to search either all versions of files or to restrict a search to only current versions of files. The system must validate and maintain a controlled list of category values and a controlled list of status values.

Searching for documents based on document content are "key word" searches using words or phrases. Document attributes, on the other hand, are summary information about the document, such as author's name, protocol number or drug name associated with the file, document title, modification date, business category, sponsoring business unit, document development milestones, project name, version status, version number, by which a document can be indexed. A document's attributes must uniquely identify it in the system. These

metadata allow for this intelligent searching through vast amounts of documentation. What metadata are captured at the time of document filing depends largely on the business requirements for searching through documents, though there are certain generic metadata that seem to apply across various settings.

As a company auditor, imagine we are confronted with the unwelcome but necessary chore of revisiting all approved contracts for the past five years undertaken by a particular division and involving a particular kind of endpoint analysis, "EP analysis," for compound XYZ. We need only briefly contemplate the prospect of searching through a paper-based archive for all occurrences of terms, say, "EP analysis" and "approved," to appreciate both immediately and fully the power of an electronic document management system. Even if we are able to physically separate all the paper contracts generated by the business unit in question, there is potentially still staggering work to be done in determining whether they involve the specified analysis or not. In a fully text-indexed EDMS with metadata indexing that anticipated the need for searching by business unit and the document type "contract," such a search would be trivial.

The richness of a business dialect probably varies directly with the specialization of the vertical market in which it competes. The ability to retrieve all relevant documents without having to specify every variant of equivalent terms can be of critical importance in performing effective and efficient searches. Such a method of searching through documents is the beginning of a movement away from the simple matching of string literals and towards the mining of data and meaning. The pharmaceutical industry is a striking example of an industry totally dependent on the ability to mine data and information for its economic viability.

In the end, EDMS systems allow for most effective searching when the filing hierarchy, indexing attributes, and some combination of full-text index, synonyms, or relevancy criteria are used in concert to specify the context around a particular search.

Output requirements: Viewing, printing, and exporting documents

Users with appropriate system privileges must be able to export documents in their native formats to any local or remote file systems to which they have access. Documents must be available for viewing in a standardized read-only format. Users must be able to view documents even if they do not have the native application. Users must also be able to match a printed or exported document to the original source file within the EDMS by means of unique system-generated identifiers referred to as watermarks.

Watermarks must not obscure any information on the document and must be in a standardized format. Watermarks must not be obscured when the document is bound.

Once a document is housed in an EDMS, it is very easy to publish it in many different places and still adhere to the principle of nonreplication. Enterprise intranets, local area networks (LANs) and wide area networks (WANs), make it possible to offer direct access to a document repository throughout an organization without requiring a physical replication of the database. Industrial strength document management systems are scalable and a single instance can support thousands of users. Also, even if the underlying filing hierarchy does not suit the needs of all customers, most EDMS packages allow for the repurposing of documents through a facility sometimes referred to as "virtual documents." This permits documents to be stored once in the repository but presented to various users in any number of ways. For instance, if both a clinical team and a biometrics team have their own standard document sets but each requires a copy of an investigator CV, two differently structured virtual documents can be created that each point to a single copy of the CV. Virtually all EDMS systems on the market today can serve as secure back-ends to web portals, enabling access to files stored in the system via ordinary HTML hyperlinks.

Electronic document management systems have the ability to distribute documents based on permissions or access control lists (ACLs). Document consumers can be notified automatically of a document's availability in the system. Once a file's status is updated to "final" or "approved," a message can be automatically sent to members of a distribution list alerting them that it has been published. Access to the document can be further streamlined by attaching to the email a pointer to the file in the EDMS.

Users with appropriate system privileges must be able to view documents. Typically, documents to be viewed in an EDMS are presented as read-only files, either in the authoring application or a standard immutable file format such as TIFF (Tagged Image File Format) or Adobe's portable document format (PDF). By virtue of its flexibility and platform-independence, PDF has become the storage standard in contemporary electronic document management systems (5). Most systems will automatically create PDF files and provide the PDF rendition in response to a user view request.

Workflow requirements: Collaboration and electronic approval

The pharmaceutical industry is a collaborative enterprise. An EDMS must provide the ability for users to electronically route documents with embedded messages

to other users electronically for review and approval. The system must enable routing of multiple collaborative document types, such as SAS logs, lists, tables, graphics, datasets, and program files. The inclusion of routing features that enable files to be sent among users for review, annotation, and editing obviates the need for the handoff of multiple paper copies via interoffice mail. Moreover, the integration of such routing utilities with standard office email packages makes it possible to notify collaborators that an item is ready for review in the EDMS environment.

The capstone to fully electronic workflow and, indeed, the key to eliminating paper handoffs from the workplace is some kind of electronic approval facility. Electronic approval can be implemented in several ways, each of which is meant to substitute for a handwritten signature. Broadly construed, electronic signatures are linked to their owners in by verifying one of three things: (1) what the owner knows; (2) what the owner has; and, most secure, (3) what the owner is.

Encryption and certificate authorities

With the growth of electronic systems for creating, maintaining, and storing an organization's critical business documents, and with growing numbers of those documents existing only electronically, our electronic identities are forced to grow. Indeed, with increased mobility within and across job sectors, these electronic identities may span more than a single organization. In this light, encryption and digital certificates present an appropriate final issue to consider when confronting the security and validity of electronic approval. Certificate authorities (CAs) are the digital world's equivalent of passport offices. These businesses issue digital certificates and validate the holder's identity and authority. Public Key Infrastructures (PKIs) and digital certificates are most trustworthy when they are vouched for by a trusted certificate authority. CAs embed an individual's or an organization's public key along with other identifying information into each digital certificate and then cryptographically "sign" it as a tamper-proof seal, verifying the integrity of the data within it and validating its use (20).

What the signatory knows: This method of approval is the simplest to administer, because it relies on the assignment of unique ID and password combinations to every member of an enterprise, a practice that is already quite common as a result of the widespread application of networked computing.

What the signatory has: Here, approvals are issued by the use of a smart card or other object containing encoded identifying information.

What the signatory is: Arguably the most secure of the three approval methods, approval here relies on some

sort of biometric identifier, some piece of physical and nonduplicable evidence of a person's identity. Such pieces of evidence include fingerprints, retinal vascular patterns, voice, and signature metrics, like speed, pressure, and stroke order (20).

Where electronic approval methods are implemented in the pharmaceutical industry, it will be of vital importance to comply with the FDA's Final Rule on Electronic Records and Electronic Signatures (21 CFR Part 11). In addition to broad guidance on the maintenance of the security and integrity of electronic records, the rule identifies specific requirements for the use of electronic signatures. First, signed records must indicate all of the following information: (1) the printed name of the signer; (2) the date and time of the signature; and (3) the meaning (e.g., review or approval). Second, the rule states that signatures executed to electronic records must be linked to the records in such a way that they cannot be excised, copied, or transferred to falsify an electronic record. Third, the rule establishes requirements around the uniqueness of electronic signatures to individuals within an organization and the controls that must be maintained to ensure that electronic signatures are used only by their genuine owners. Finally, and perhaps most importantly, 21 CFR Part 11 clarifies the necessity of notifying the users of electronic signatures that their electronic signatures are intended to be the legally binding equivalent of traditional handwritten signatures (2). The details of the rule surrounding electronic records and signatures are treated in greater detail later in this article

Security requirements: Protecting the corporate assets

Information and its knowledge base are the pharmaceutical company's most important asset after its people. The corporation's document management systems will either protect this asset or expose it to risk. Integrity of the information and assurance that the data were protected from fraudulent manipulation are at the heart of the new regulations (2) on controls for electronic records management systems. Technology has rationalized secure access to documents at each of the levels outlined in the schema of the paper-based file room. These levels range from broad to specific access privileges and are defined by (a) user login name (user security), (b) the user's affiliation with a group (group security), and (c) document status. The permissions associated with a given level of access entail those of all the more restrictive levels. For any user or group of users, the hierarchy of permissions is NONE, BROWSE, READ, WRITE, EDIT, DELETE, ALL.

Corporate network security: At the broadest level, access to the electronic file room must be monitored in the

same way as access to the physical file room. A firewall is the equivalent of security guards at the entrance to the building. Corporate firewalls are barriers that exist between an enterprise's computer networks, or intranet, and any external network. Generally, firewalls filter inbound and outbound data, provide or manage public access to requested locations, deny all services except those explicitly permitted by the system administrator, log traffic and activity through the firewall, and activate alarms as prowlers are detected (21).

Database security: At the level of the EDMS application, access is governed by unique user ID and password combinations. Here, database security serves as the guard at the door who requests to see your badge before allowing entry. The system administrator maintains an updated register of valid IDs and passwords for gaining access to the system linked, typically, to the current list of employees maintained by human resources (9).

Folder security: As with our brick and mortar analog, entrance to the file room does not grant you free access to all the holdings. A folder may be accessible to a specific class of user. User permissions over folders are typically limited to NONE, BROWSE, or READ. The ability to CREATE folders is typically the restricted province of the EDMS administrators.

File security: While it may be that a specific class of user has access to the entire contents of a folder, it may just as well occur that members of that class have access to all documents in the folder with some exception. For example, Human Resources employees may have access to documents containing the terms of employment for all employees except those in Human Resources itself. Security must be applied to these particular documents in the EDMS such that they cannot be viewed by Human Resources personnel.

Version security: In the paper world, archival documents held in a file room are in some sense final drafts. Security for a product document must be determined by its status. The electronic document management environment in which documents are managed across their entire life cycle by its very nature allows for the inclusion of draft materials. So, in addition to permitting or disallowing the viewing of the final version of a file, the electronic world must consider and secure access to draft versions. Usually, only authors, editors, and reviewers of files have access to the drafts before they are finalized. Once files are promoted to final status, this version is then promulgated to a broader readership. The draft versions, by contrast, may continue to be accessible by the authors, or may be locked or purged from the system.

Disaster recovery

Any treatment of security in the context of document management would be remiss if it did not mention audit trails and disaster recovery.

The system must maintain a log of document creation, modification, or deletion. The automated log includes information about the document that was affected, the time at which the event occurred, and the user who caused the event (2).

Disasters include the full range of unplanned catastrophes, from internal or external sabotage to the extreme forces of nature, like fires, tornadoes, and earthquakes that do real damage to a corporation's physical assets. Each of these occurrences presents a serious risk to an enterprise's profitability and viability. It is, therefore, essential that disaster recovery plans be well-conceived and fully supported by management.

Retention requirements: Storage and deletion

In keeping with the dictum that an EDMS should enable management of files across the entire document lifecycle, the system should automate retention schedules. The system should calculate scheduled document expiry based on business rules that map to system file categories. The system should then be programmed to notify document owners, usually at some predefined interval prior to actual destruction, as a final check on extenuating business needs for extending retention. Business rules based on regulatory and statutory requirements can be built into the system's automatic monitoring and retention/deletion functionality. For a comprehensive discussion of storage media, see Pollnow (22).

Scheduled retirement of documents is especially critical in the pharmaceutical industry, where corporations are routinely subject to audits by the regulatory authorities or to discovery in litigation. In such cases, retention of draft documents and other aged files often unnecessarily increases the company's liability exposure. Although administered centrally, retention programs ideally devise a mechanism for notifying document owners prior to destruction, alienation, or transfer, as the document's owner may know of compelling business or legal reasons to delay the destruction of the document (23, 24). An EDMS allows an enterprise to automate and rationalize all aspects of the retention program, including scheduling, notification, and retirement.

The EDMS Toolbox

An electronic document management system is a configuration of design tools, data capture tools,

messaging systems, repository databases, portals, and intra-/internetwork distribution and transfer options. There is no one magic bullet or single system that will provide solutions to the industry's document or knowledge management needs. With the almost daily introduction of new products, the best return on investment in technologies must be built on the principles of open architecture, best of breed products, and scalability. Designing, building, and constantly refining the corporate "digital nervous system" (18) is part of the cost of doing business in the paperless environment. In the next section we present an overview of key developments that undoubtedly will continue to redefine and reshape this complex and fluid corporate nervous system.

THE EMERGING PARADIGM

The transition from the paper environment to the early electronic environment was direct and intuitive. New computing models and technologies, however, are raising questions about the validity of this paradigm for complex information capture, access, and retention. In the current technology environment, forms, documents, data are not frozen, but dynamic and interactive. Communications and collaborations are as frequently machine to machine exchanges as direct human interactions.

It appears that in the rapidly evolving world of computing and communications (8, 25) the medium is indeed as important as the message. It is now possible to separate data and information from any document structure and technical platform. Whereas a document management system applies techniques of document locking, version control, and security to document objects, a component management system applies those same features to document fragments, or chunks, at a finer level of granularity. In other words, instead of checking out an entire document for editing, in a component management environment we may check out a single section, paragraph, or table. Much as the fundamental document objects that populate an EDMS can be combined in various ways as virtual documents, in a component management system document chunks form the basic elements and are "assembled" into whole documents. Component management is precisely the concept behind the regulatory search for a Common Technical Document (26).

The technology that enables both the creation of documents and the repurposing and redistribution of their content is XML, eXtensible Markup Language

(27). In terms of paperless documentation systems, XML signals a move from whole document management to component, or content, management. XML is a meta-language for describing languages that represent the content of a document. It is a subset of the Standard Generalized Markup Language (SGML), ISO Standard 8879, which has been around since 1986. SGML was devised as a means to create portable documents that are independent of any specific hardware or software.

HTML, which is also a subset of SGML, was developed as a means of transmitting hypertext documents over a network and specifies how a document should be displayed in a Web browser. HTML, however, describes neither the information content of a document nor its manner of organization. XML precisely fills this void by making use of tags and attributes to extend, validate, and unambiguously structure the content of a document.

Instead of viewing documents in an information system as passive objects to be served up to a human for reading, the SGML/XML paradigm regards them as active communications that can be parsed and manipulated by a computer. As such, they are data structures that can be used by computers rather than loose agglomerations of words, which can only be read by a human who comprehends the underlying grammar. XML's capacity to describe the content of a document as well as its organization makes it the ideal language for communication among both humans and machines. By encoding information in a format or tags that can be digested by both human and computational agents, XML is poised to become the true *lingua franca* of the Web in which both humans and silicon are initiators, brokers, and processors of transactions. To achieve this genuinely seamless communication, XML rethinks the very nature of documents themselves.

The usefulness of XML tags for structuring data presupposes the existence of a set of rules, or grammar, for their application. Such a grammar is embedded in either a Document Type Definition (DTD) or schema, both of which are defined by the XML author. A DTD is a set of syntax rules for tags. It establishes what tags may be used in a document, what attributes they have, and whether or not they can be nested inside other tags. Typically a DTD is maintained separately from a given XML document, though it can be part of it as well. Unlike HTML, which has a single DTD, XML supports as many user-defined DTDs as there are ways to structure information—thus, the extensibility. And also the danger. DTD's and, hence, XML are only as useful as the agreed upon standards that are used to implement them. The pharmaceutical industry stands to gain a great deal to the extent that it can establish and sustain extensive cross-corporation and FDA collab-

oration in the development of standard XML tags for its key documents. Should the industry fail to marshal a truly collaborative effort, it will have gained far less, being left instead with unique and peculiar DTDs and islands of XML that will exact a considerable cost to translate and maintain

Schemas are a newer development on the XML landscape. Like a DTD, a schema supplies the rules for building a document and indicates what tags may be used, what their attributes are, and how they relate to one another. Unlike a DTD, however, a schema has an additional level of specificity in that it can define data types. For example, a DTD might have a tag designated as `<DOSE>`, the content of which might be either numerical or a character string. A schema, by contrast, could ensure that the value entered was a number. This is clearly an appealing feature for the pharmaceutical industry, especially in the case of information exchange between databases or other applications making use of rigidly defined data types.

At the time of this writing, however, it still remains unclear whether schemas will overtake DTDs as the fundamental means of modeling XML documents.

XML structures document content but leaves presentation and formatting to other tools. The primary tool used for the job is now XSL, eXtensible Stylesheet Language. XSL consists of two parts: (1) a language for transforming XML documents and (2) an XML document for specifying formats. XSL allows users to process an XML document and dynamically render it to any number of formats. HTML, ASCII text, PDF, WML (wireless markup language) and XML itself are among the formats supported by the current technology.

XSL takes advantage of the intelligence built into the XML document to present its components in flexible ways. Thus, XSL does not merely apply style to the XML data; rather it evaluates, rearranges, and reassembles it. In fact, XSL permits multiple passes over an XML document so that information that appears once in the XML source can be presented multiple times in different formats within the same presentation. XSL, then, provides us with more than just an eye-catching rendition of XML data; it enables XML to realize its potential as flexible source information that can be created once but modified and presented infinitely many times. The application of this technology to the preparation and submission of information to multiple regulatory authorities is obvious.

Traditional paper models focused on discrete documents, file cabinets, and signature accountability. Setting aside the ability to commit fraud in any environment, the paper trail was clearly traceable from a physical file cabinet in the custody of a records manager responsible for

a uniform indexing scheme to the data owner through the signature on the document. The computer network trail, on the contrary, runs through multiple virtual file cabinets under multiple ownership, custody, and indexing schema with multiple signatures or access authorities applied at both the data and systems levels. Added to this is the complex multilayered architecture supporting the exchange and communication of the information (28). All of these elements are the document. The medium can no longer be separated from the message.

THE REGULATORY ENVIRONMENT: EFFECT OF LEGISLATION ON TRADITIONAL AND EMERGING MODELS

Developments in paperless documentation systems for the pharmaceutical industry are intimately linked to regulatory developments. That "there is no alternative to moving towards complete electronic record keeping" is acknowledged by industry and regulators (29,30). The most documented environment in pharmaceutical industry is unquestionably the manufacturing segment. As early as 1983, FDA issued guidelines on the inspection of computerized systems used in drug manufacturing (31). In this environment, where there is no evidentiary trail to support either the compliance or integrity of claims and, where there is no documentation of data, messaging, transactions, actions, criteria applied, historical experience, audit experience, or staff experience and training, the event, action, collaboration, or data itself are considered nonexistent. Repeatedly and often, FDA officials have stated, "If it is not documented, it did not happen." FDA modeled the requirements and standards of the Final Rule on Electronic Records and Signatures (Part 11) on the GCMR regulations (32). The agency acknowledges that most predicate regulations, however, were written without contemplating the use of any technology other than paper (2). (Preamble Comment XVI.A, p.13462)

Basic differences between the manufacturing environment and other segments of the pharmaceutical industry, particularly the clinical research environment, pose additional special difficulties for industry and regulators in complying with and applying standards to the electronic management of data and documents. In the manufacturing environment, documentation of processes and results is fairly self-contained within a single location or plant. There is typically a single cohesive business culture focused on common deliverables where individual business units are relatively seamlessly integrated. At the other extreme, in the clinical research segment of the

industry, documentation of processes and results is spread across multiple locations (e.g., investigational sites, CRO facilities, laboratories, sponsor offices). There are multiple business cultures, each with distinct deliverables (e.g., research facilities, consultant services, data management groups, marketing departments, finance). These multiple cultures and business units are not typically well integrated by a paper or any other technology platform.

There have been four pivotal legislative developments between 1995 and 2000 impacting the pharmaceutical industry:

1. *FDA Final Rule on Electronic Records and Electronic Signatures* (effective August 20, 1997): This is a permissive regulation allowing any FDA-regulated industry to maintain paperless documentation systems that will be considered the legal equivalent of paper records and record keeping systems.
2. *ICH Recommendations for Electronic Standards for the Transfer of Regulatory Information (ESTRI)* (effective March and July 1997): This agreement by member states defines open nonproprietary international standards for electronic communications and the transfer of data and documents between industry and regulatory authorities.
3. *Government Paperwork Elimination Act, Title XVII of Pub.L. 105-277* (effective October 21, 1998): This is a U.S. law requiring federal agencies to be prepared to accept by 2003 electronic information from persons required to maintain, disclose, or submit information to the U.S. government.
4. *Electronic Signatures in Global and National Commerce Act* (effective October 1, 2000): Again, this is a U.S. law giving electronic signatures and electronically signed documents in commercial transactions the same legal status as handwritten signatures applied to paper documents.

The latter two laws, although specifically United States federal regulations, will impact all e-commerce transactions because of the importance of the U.S. economy in the world marketplace. U.S. industries will seek the competitive edge advantage provided by this legislation.

Regulatory authorities are fully aware of the complexities that the new, rapidly evolving computer environment poses for their oversight responsibilities. Both FDA and ICH wish to avoid rapid obsolescence by crafting regulations and guidelines based on open nonproprietary standards and generic tools such as Adobe Acrobat's free portable document format (PDF) codes and reader application. The ICH ESTRI Gateway agreement is the first step in attempting to define a platform independent environment.

Final Rule on Electronic Records and Electronic Signatures

FDA's Final Rule on Electronic Records and Electronic Signatures (the Final Rule, Part 11) requires special discussion since it is the most comprehensive regulation to date on applying computer technologies to regulated industry (33). The rule itself is contained on two Space consumption pages. The most important part of the rule for the pharmaceutical industry is, arguably, the Preamble, which comprises 34 pages of discussion on how and why the FDA made its decisions on the requirements it would enforce. The preambles to both the proposed rule making and the final rule making detail current regulatory thinking on and compliance concerns with paperless documentation systems (34).

Scope

The initial announcement of the proposed rule referred to the regulations as an electronic signature rule. By the time of publication of the Final Rule in the Federal Register (62 FR 13430, March 20, 1997) (2), the title had been expanded to "Electronic Records; Electronic Signatures." FDA rejected comments that would have limited the scope of the rule to signature authority and manifestation issues or to those records only that are required to be signed, witnessed, or initialed. The agency's rationale was that the "reliability and trustworthiness of the electronic signature depends in large measure on the reliability of the underlying electronic record" and that electronic records need to be reliable, trustworthy, and compatible with FDA's mandate to protect the public health "regardless of whether they are signed" (2). (Preamble Comment 26, p.13438) Therefore, regulation, in the agency's opinion, was required for electronic record keeping per se. By March of 1997, the regulatory authorities were fully aware of the power of the new computing and networking technologies and the communications explosions fostered by the internet and wireless possibilities.

Part 11 defines records management as the "creation, signing, modification, storage, access, and retrieval" of records and the software and hardware platforms used in these processes. This regulation applies to any documentation required by an FDA "predicate rule." The Authorities cited for the rule are Secs. 201–903 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 321–393); Sec. 351 of the Public Health Service Act (42 U.S.C. 262)]. For example, 21 CFR 820.70 requires documentation of process and production controls in quality systems for medical devices. This is a predicate rule. If any of this documentation is created, signed, modified, stored, accessed, or retrieved electronically, the requirements of

Part 11 apply to the documentation and the hardware and software supporting it.

Applicability

There is a caveat to the application of Part 11. As noted, it is a permissive rule. Industry has the "option" of compliance. However, if a company currently uses any electronic systems to create, sign, modify, store, access, or retrieve records, it is presumed by the law that they have elected to follow the regulations and comply with the standards. Consequently, the company is subject to FDA enforcement of the rule under the food and drug regulations. If a company elects not to comply with the regulation, it must revert to all paper systems. No electronic systems are "grandfathered" under the Final Rule. Continued reliance on current electronic systems for any one of the electronic records management activities requires compliance with the requirements of Part 11 for security and controls (2). (Preamble Comment XVI.C.1, p. 13463)

Key issues

It is important for the industry to understand FDA's rationale in establishing these regulations, if appropriate and reasonable compliance policies and procedures are to be implemented. Any industry interpretation of the rule should be based on a thorough understanding of the preamble discussion of the final regulation (34). FDA's concern to protect the public health translates into four major issues addressed by the Final Rule: identification and authentication of data and source; system confidentiality and security; accountability for signing a document; and enforcement.

Identification and authentication of data and data source (audit trail): The ability to verify the integrity of records and to trace accountability for creation and modification of data are probably the most significant concerns of the FDA in relation to their ability to protect the public health and hold individuals and corporations accountable.

There must be a system-generated time-stamped audit trail, created independently of the operator of a system, for any access to, modification of, or deletion of records or the audit trail itself. This audit trail is intended to enable detection of record and signature falsifications and to provide an evidentiary trail in the event of falsification (2). (Preamble Comments 72–74)

Confidentiality and security (controls): FDA recognized that deliberate intent to commit fraud is very difficult to prevent (2). (Preamble Comment 7, p. 13433) Acknowledging this, the agency defined system control requirements at the operational, network, and device levels

“to ensure that representations of database information have been generated in a manner that does not distort data or hide non-compliant or otherwise bad information, and that database elements themselves have not been altered so as to distort truth or falsify a record.”

Requirements for controls are based on definition of “closed” or “open” system. Under these definitions there is no direct correlation between, for example, using a public phone line and an “open” system. Compliance with “closed” versus “open” standards is determined by how access rights to the data or documents are established and controlled by the owner(s) of this information. A system is defined as “closed” if access to the system containing the records or data is under the control of person(s) responsible for the content of the records or data in the system. A system is defined as “open” if access to the system is not under the control of the person(s) responsible for the content of the records therein. For example, dial-in retrieval over a public phone is “closed” where the records being accessed are under the control of the persons responsible for their content, whereas storage of records on a third party system is “open” because access to the records themselves is under the control of the third party. Sections 11.10 and 11.30 of the Final Rule list, respectively, the control measures required for establishing a “closed” or “open” system.

Access must be protected through use of unique biometric or digital identification technologies. Examples of biometric technologies include voice or fingerprint recognition. Nonbiometric or digital signatures must consist of two distinct identification codes such as a network user ID and password. All access rights and system supplied identification codes must be periodically reviewed, updated, and/or changed by the data/document owner(s) and all personal identification codes, such as passwords, periodically changed by the user to ensure that unauthorized parties do not gain access to the data/records. Open systems require an additional level of encryption on the data/records to prevent fraudulent access.

Accountability for signing (repudiation and links): Signatures can represent different intent and responsibility and the electronic signing must capture this metadata. Electronic signatures can be executed using biometric or digital technologies or a hybrid of these. Part 11 provides minimal standards and generic technical recommendations for ensuring that e-signatures can be unquestionably linked to their owners, cannot be repudiated, and are executed with clearly evident intent and understanding of the act of signing (35). Whatever final configuration of signature technologies is used, it must enable the FDA “to hold people (i.e., the individual) to the commitment they make under their electronic signature” (2). (Preamble Comment 19, p. 13456)

In fact, when it was suggested that FDA also hold business entities accountable as well as individuals, the agency rejected the idea because business entities do not sign records, individuals do (2). (Preamble Comment 90, p. 13450) (23)

Legal enforcement (authority, inspection, implementation): Part 11 cites the complete Federal Food, Drug, and Cosmetic Act as its authority for this regulation and for the scope of its enforcement jurisdiction. FDA wants “enforceable” baseline standards (2) (Preamble Comment 2, p. 13432) and intends to inspect any component of an electronic records and signature system that has bearing on the trustworthiness and reliability of the record or signature (2) (Preamble Comments 33, p. 13489). The agency reserves the right to conduct inspections, even of sensitive security systems, if deemed necessary “to enforce the provisions of the act and related statutes” (2). (Preamble Comments 32, p. 13439) Supplemental guidelines will provide additional detail on the standards against which FDA inspectors would judge compliance (7). FDA will consider the need for additional legislative initiatives or criminal law reform if their experience with the current rule warrants (2). (Preamble Comment 90 and 124, p. 13450 and p. 13458, respectively) The first FDA 483 inspection reports on compliance issues with Part 11 began to appear in 1999. Mr. Paul Motisse, widely considered FDA’s architect of Part 11, was transferred to the inspection branch in the same year. As the agency gains experience and confidence with the new technologies, enforcement activities related to Part 11 will significantly increase.

Economic Impact

FDA asserts that the benefits of electronic record keeping in reduced review time and related business costs, such as space and rapid information access, will offset the costs of compliance with the rule (2). (Preamble Comment XVI.C.1, p. 13463) As industry and regulators enter the brave new paperless world, this remains to be verified.

It might be beneficial to discuss some of the current wisdom about the relative merits of paper-based and paperless documentation systems. Some of the claims about the outright superiority of electronic-based systems are perhaps overstated and highly context-dependent. Others probably have as much merit as their face value would seem to indicate. Table 1 provides a summary inventory of the pros and cons of the two models for document management.

Although a paper-based model for managing enterprise documentation appears to fall short of the promise for complex document management requirements, the

Table 1 Pros and cons of paper-based and electronic document management

Paradigm	Pros	Cons
Paper-based	Familiar Durable Transparent security model	Costly Inefficient searches Space consumption Inhibits leveraging of information assets Environmental impact Emphasis on separate repository function Audit management overhead
Digital	Space efficient Efficient searches Promotes leveraging of information assets Saves paper Integrates active file management and repository function Automates audit management	Durability of storage media Ever-changing technology standards High initial IT investment IT management overhead Environmental impact

paperless model is not without its own downside. Let us consider some of the issues in greater detail.

Cost

Paper-based systems are said to be costly. At roughly \$0.10 per sheet of paper, printing and filing between 5 and 10 million pages—perhaps a conservative yearly estimate for an organization of 2,000 employees—would cost between \$0.5 million and \$1 million (19). Are digital documentation systems any less expensive given the costs of purchasing, servicing, and operating the component hardware and software systems? The answer depends on the level of digitization of the system. In a hybrid system, where some paper is printed and scanned into an electronic filing system, the average costs of implementing and maintaining a document indexing and scanning solution is \$0.50 per page and ranges between \$0.20 and \$1.20 per page. It is difficult to assess the cost efficiencies of digitally based documentation systems for production and filing of documents. The full costs must include the initial investment in electronic systems and the outlays required to maintain them. It is no less incumbent on the pharmaceutical industry than any other e-commerce business to find a new measure, as precise as per page costs, for measuring its cost of doing business in the paperless environment (18).

Efficiency

Paper-based systems are increasingly considered inefficient. Relative to the file room scenario used to support the paper-based documentation model, it is hard to argue with the very real efficiencies afforded by the ability to search

vast and disparately located document repositories electronically. In a paper-based environment, the effort required just to retrieve documents has been estimated to be on the order of 4–7 hours per week per person (19). This does not include the time required to travel physically from one's office to a file room, walk the aisles of the archives, peruse the shelves and folders, and finally to browse the file itself for the relevant passage. In a 2,000 person organization with 250 working days a year, this totals 50,000 to 87,500 person days per year across the organization.

By contrast, search and retrieval in an electronic environment is said to save 10 to 20 percent of the total individual work effort required, not just the effort required to retrieve documents. In one year, an electronic system would save this company 50,000 to 100,000 person-days of effort, potentially all the time spent searching in a paper-based system and more (19). If network downtime is equivalent to employee downtime, comparison between the two modes of search and retrieval appears absurd (36).

Space consumption

If the average enterprise professional creates approximately 500 documents a year, accumulates around 200–300 more from external sources, and an average document is from 5 to 10 pages in length, then an organization of 2,000 employees generates around 1 million documents containing some 5–10 million pages (19). Add the external documents and the organization contends with another 2 to 6 million pages. Continuing with the multiplication exercise, a ream of

paper is about 2 inches thick, so given the above volumes an enterprise must find space to accommodate anywhere from 2,333 to 5,333 stacked feet of $8\frac{1}{2}$ by 11 inch paper, or between 1515 and 3463 cubic feet. In raw terms, the upper end of this volume will fit into a room measuring 35' by 10' by 10'. When one considers that this does not take into account the folders, shelves, and aisle space required to intelligently store and navigate the documents, and that this is merely the volume of documentation encountered in a *single* year, entire floors of city-block-wide buildings can very rapidly disappear in an effort to retain it all.

Ten million pages requires about 500 gigabytes of disk storage space. The space occupied by the servers and disk drives for a decade worth of documents would probably not exceed the 35' by 10' room cited above as the minimum requirements for packing a year's worth of documents. Standard dimensions for 30-gigabyte hard drives on the market today are roughly 2.75" by 4" by less than half an inch in height. Simplifying somewhat, two hundred such drives, stacked one atop the other, would occupy less than 9 linear feet and only half a cubic foot. Even with the overhead required by server hardware, monitors, and the like, the space savings won by electronic storage of enterprise documentation are considerable. However, in the digital world, the cost of bandwidth is a major factor in "digitalizing" the workplace in both large and small corporate budgets (37).

Environmental impact

The high-tech digital alternative must be vastly more "green" by these measures. Yet, it is not immediately clear that the paperless documentation system is as environmentally friendly in practice. In general, the high-tech industry is pretty dirty. Chip manufacturing involves hundreds of chemicals (e.g., acids, cyanide compounds, silicon tetrachloride), many of them toxic or carcinogenic alone and, perhaps, more so in combination (39–42). The industry also wastes prodigious amounts of water, and pollutes both air and water (43). The reality of Moore's Law, which roughly states that the power of computer technology (i.e., the miniaturization and data density of transistors and microchips) doubles every 18 months, has landfills overflowing with obsolescent machines and components (44). Consider also the power required to operate electronic document management systems. The presumed environmental advantages, which *prima facie* seem obvious, are in fact not so clearcut.

Current wisdom maintains that paperless documentation systems are more environmentally friendly than paper-based systems. If an enterprise implements a system in which documents are generated, stored, and retrieved electronically, it can conceivably preserve the number of

trees corresponding to the number of sheets of paper it would otherwise generate. In a 2,000 person organization, which produces 10 million documents, that amounts to 850 trees a year (19). Typical offices produced 100 lbs of paper per head in 1975. In the still burgeoning paperless office, they produce more than 200 (1). One might argue that the paper industry exacts as in an even more terrible toll in terms of sulfur dioxide, resinous acid effluent (38).

Storage and durability

With respect to the most fundamental security issues (i.e., a document's persistence through time), it is rather easy to assume that digitized information, with its tight controls, rigid architecture and underlying ingenuity, must somehow offer greater long-term protection of stored document assets. To do so, however, is both to underestimate the durability of paper as a storage medium and overestimate that of digital media and the dominance of any single technology platform capable of reading digital media formats. Ineluctable technological obsolescence is perhaps the greatest threat to the preservation of electronic documents and data (45–48).

The above considerations have direct bearing on the presumed economic impact of regulations for the pharmaceutical industry on electronic record keeping systems. The pharmaceutical industry is, in fact, faced with a complex, costly development and implementation project once it heads down the road of electronic record management within the regulated parameters of Part 11. These costs and the attendant legal issues will increase within a fully digitalized, paperless environment. The key to a return on investment for the industry will be its ability to integrate and strategically plan technical implementations across multiple business divisions (49, 50).

CONCLUSIONS

Digital tools provide instantaneous distributed access to original experience. In a paperless world, one that is mobile, wireless, and extremely portable, customers seek data versus documents (37). Documents have become data embedded in inseparable multiple layered architectures (28). Technological and economic forces have essentially redefined document formats and purposes. The goal is more and more to preserve or re-create the immediacy of a communicative act, just-in-time delivery, real-time updates

"Document objects" have, in addition to text, come to include databases/spreadsheets, graphics, images, video, audio, web sites. What a document used to be is now a

small part of documentation—audit trails, metadata, etc.—where documentation of immediate experience is immediately available to third parties. Documentation is now an environment of dynamic interactive support as opposed to the static support of imaged paper.

The pharmaceutical industry is traditionally document driven. The technology exists currently, however, for conducting research, manufacturing product, and promoting prescription drugs in a totally paperless environment. Several years ago the CEO of Boeing Corporation threw down a gauntlet to the aerospace industry. He asked what would happen if they designed and built a plane without any paper, and succeeded in doing it. In 1999, Bill Gates laid out the business case (18) for conducting any business as e-commerce. Both individuals were visionaries in understanding the possibilities of the new information technologies and translating them into a competitive edge for their industries.

Within the clinical research arena, the CRF, for example, will become a metaphor for a complex, multidimensional process of data collection providing an on-line interactive profile of the subject's experience and data. New technologies such as voice recognition, digital medical imaging, targeted drug discovery techniques, revolutions in computing design and power (8, 25) and the genome project (51, 52) are redefining for both industry and regulators how we capture and manage subject-specific data in a meaningful profile of an individual subject's experience. These same technologies, especially as they enable the discovery, mapping, and repeated use of genetic information, will redefine the execution and need to monitor informed consent rights for subjects. The protocol itself will be redefined as a set of "edit checks," access rights, system administration rules, and metadata structures controlling the on-line collection, verification, and analysis of data.

How should the pharmaceutical industry apply these possibilities to drug development and marketing? Fully "digitalizing" clinical research, for example, would revolutionize the industry's concepts of document, data, and regulatory review. It is no less impossible to apply Boeing's challenge to the manufacturing sector. Of all the areas in the industry, marketing, which is built on messaging and interactive communications, has innumerable models and other industry experience with e-commerce. Within the next five years, the industry needs to challenge its traditional thinking on pharmaceutical documentation and thoroughly explore the possibilities and issues involved in abandoning traditional paper paradigms for a fully digital environment (49, 53).

How can the regulatory agencies respond to this new environment in keeping with their legal mandate to protect

the public health? The regulators have begun to tackle some of the issues in the Final Rule on Electronic Records (2) and the ICH ESTRI agreement (5). This legislation is merely the tip of a large, complex "iceberg" of law, public policy, and oversight techniques. Traditional documentation will become a metaphor for complex, integrated data capture and repository systems linked to enable dynamic "on-the-fly" profiling, analysis, and reporting of research and discovery results almost simultaneously with the direct research experience. The pharmaceutical industry, its regulators, and the consuming public have entered into a brand new collaborative world.

Glossary of Technical Terms

A comprehensive listing and definition of terms is available at <http://www.webopedia.com>

REFERENCES

1. Brown, J.S.; Duguid, P. *The Social Life of Information*; Harvard Business School Press: Boston, 2000.
2. *21 CFR Part 11: Electronic Records; Electronic Signatures; Final Rule*, Food and Drug Administration, 62 FR 13429, March 20, 1997.
3. *Government Paperwork Elimination Act*, Pub.L.105-277, Office of Management and Budget, Executive Office of the President, Proposed Implementation (64 FR 10895, March 5, 1999).
4. *Electronic Signatures in Global and National Commerce Act*, Pub.L.106-229, June 30, 2000.
5. *Electronic Standards for the Transfer of Regulatory Information (ESTRI)*, International Conference on Harmonization. M2 Recommendations, March 1997 (1.1-5.1) and July 1997 (5.2-5.3).
6. *Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients*, International Conference on Harmonization, July 19, 2000; Draft Consensus Guideline.
7. *Guidance for Industry: Computerized Systems Used in Clinical Trials*, Food and Drug Administration, April 1999.
8. Mann, C.C.; Rotman, D.; Waldrop, M.M.; Garfinkel, S.L.; Regalado, A. Beyond Silicon: The future of computing. *Technol. Rev.* **2000**, *103* (3), 42-84.
9. Sutton, M.J.D. Chapter 8: Developing the Physical Model. *Document Management for the Enterprise: Principles, Techniques, and Applications*; John Wiley & Sons: New York, 1996; 223-244.
10. Harvey, D. *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*; Blackwell: Cambridge, 1989.
11. Carlston, D. *Storing Knowledge*; www.longnow.com (accessed Oct 2000) Time & Bits.
12. Rehr, D. *The First Typewriter*; dcrehr@earthlink.net (accessed Oct 2000).
13. Carlson, C.F. *The Photocopier*; www.mit.edu (accessed Oct 2000).

14. *Word Processing; Information Processing*; www.britannica.com (accessed Oct 2000) Encyclopaedia Britannica.
15. Connolly, D. *A Little. History of the World Wide Web*; www.w3.org (accessed Nov 2000), 1945–1995.
16. Bush, V. As We May Think. *Atlantic Monthly* **1945**, 176, 101–108.
17. Sutton, M.J.D. Chapter 1: The Transition to Enterprise Document Management. *Document Management for the Enterprise: Principles, Techniques, and Applications*; John Wiley & Sons: New York, 1996; 1–22.
18. Gates, B. *Business @ The Speed of Thought, Using a Digital Nervous System*; Warner Books: New York, 1999.
19. Sutton, M.J.D. Chapter 3: Planning for Enterprise Document Engineering. *Document Management for the Enterprise: Principles, Techniques, and Applications*; John Wiley & Sons: New York, 1996; 53–92.
20. Yakal, K. Sign on the Digital Line. *PC Magazine*; Sept 19, 2000; 32–36.
21. Chapman, D.B.; Zwicky, E.D. *Building Internet Firewalls*; O'Reilly and Associates, Inc.: New York, 1995.
22. Pollnow, R.A. *Alphabet Soup, Acronyms Everywhere—Media Options in Data Storage*; Sept/Oct 2000; 44–47.
23. *Information Requirements Clearinghouse. Legal Responsibility for Records in a Corporation*; http://www.irch.com/irch_article2.htm (accessed Sept 2000).
24. *Information Requirements Clearinghouse. Legal Requirements for Filing Systems and Indexes*; http://www.irch.com/irch_article2.htm (accessed Sept 2000).
25. Williams, S. Computing After Silicon. *Technology Review* **2000**, 102 (5), 92–96.
26. *Organization of the Common Technical Document for the Registration of Pharmaceuticals for Human Use*, International Conference on Harmonization, July, Draft Consensus Guideline.
27. Bosak, J.; Bray, T. *XML and The Second-Generation WEB*; www.sciam.com (accessed Sept 2000): .
28. Horak, R.; Miller, M.A. Chapter 7: Fundamentals of Data Communications. *Communications Systems & Networks: Voice, Data, and Broadband Technologies*; M&Tbooks; Division of MIS: Press, Inc.; Subsidiary of Henry Holt and Company, Inc.: New York, 1997; 155–184.
29. *PhARMA (Pharmaceutical Manufacturers' Association) Position Paper on 21CFR Part 11*, Docket No. 92N-0251 Nov 30, 1999.
30. Raymond, R.; Galle, S.; Collom, W. Regulatory Submissions: From CANDA/CAPLA to 2002 and beyond. *Drug Information Journal* **2000**, 34 (3), 761–774.
31. Guide to Inspection of Computerized Systems Used in Drug Processing, U.S. Department of Health and Human Services, Public Health Service, Food and Drug Administration National Center for Drug and Biologics and Executive Director of Regional Operations, Feb 1983.
32. *Annual Report for 1999*; Food and Drug Administration, Office of Compliance.
33. Hoff, S. The Regulatory Environment for the New Millennium. *Drug Information Journal* **2000**, 34 (3), 659–672.
34. Woodrum, P. 21 CFR 11—more than meets the eye. *Applied Clinical Trials* **June 2000**, 86–94.
35. Bleicher, P. Sign Here, Please, Mr. Bondh. *Applied Clinical Trials* **August, 2000**, 28–31.
36. Brin, S.; Page, L. *The Anatomy of a Large-Scale Hypertextual WEB Search Engine*; <http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm> (accessed Sept 2000).
37. Gingrande, A. *Wireless Applications For a Brave New World*; Sept/Oct 2000; 66.
38. <http://www.greenpeace.org/toxics/reports/gopher-reports/papdam.txt> (accessed Sept 2000).
39. Ayers, J. Controlling the Dangers of High-Tech Pollution. *EPA Journal* **1984**, 10, 14–15.
40. Hayes, D. *Behind the Silicon Curtain: The Seductions of Work in a Lonely Era*; Black Rose Books: London, 1989.
41. Miller, M.W. Findings of Toxin Leakage in Silicon Valley Hurt Chip Makers' Reputation for Safety. *Wall Street Journal* **Aug 29, 1984**.
42. Seigel, L.; Markoff, J. *The High Cost of High Tech: The Dark Side of the Chip*; Harper & Row: New York, 1985.
43. Bills, B. *High Tech Toxins: Is the Computer Industry as Clean as it Seems*; www.canada.cnet.com (accessed Oct 2000).
44. Yang, D.J. *On Moore's Law and Fishing: Gordon Moore Speaks Out*; www.usnews.com (accessed Nov 2000).
45. Rothenberg, J. *Ensuring the Longevity of Digital Documents*; www.sciam.com (accessed Sept 2000).
46. Izarek, S. *Archiving Today's Digital Culture*; www.foxnews.com (accessed Oct 2000).
47. Lesk, M. Going Digital: Electronic Libraries will Make Today's Internet Pale by Comparison. But building them will not be easy. *Scientific American*; www.sciam.com (accessed Sept 2000).
48. Brand, S. Written on the Wind. *Civilization Magazine*; www.civmag.com (accessed Oct 2000).
49. Moore, G. *A Balanced Buy*; www.Darwinmag.com (accessed Nov 2000).
50. Heck, M. *Document Management Fuels e-Business*; www.infoworld.com (accessed Sept 2000).
51. Regalado, A. Mining the Genome Project. *Technology Review* **2000**, 102 (5), 57–63.
52. Philipkoski, K. *Genome Map Heralds Cheap Drugs*; www.wired.com (accessed Sept 2000).
53. Buder, R. The Corporate Logic. *Technology Review* **2000**, 103 (3), 88–90.

PARTICLE-SIZE CHARACTERIZATION

Brian H. Kaye

Laurentian University, Sudbury, Ontario, Canada

SAMPLING PROCEDURES

The subject of powder sampling has been addressed extensively in standard reference books. Several companies sell sampling equipment for installation in industrial systems (1–7). Even when the sample received by the laboratory is a representative sample, the analyst still faces the difficult task of taking a small subsample from the powder supplied. It has been shown that one of the most efficient devices for taking a subsample of the powder is known as the spinning riffler, shown in Fig. 1a. It consists of a ring of containers that rotates under the powder supply. The total powder supply is processed using this instrument. It has been shown that to obtain a representative sample, the time of powder flow through the apparatus divided by the time of rotation of the ring of containers should be a large number (4, 7). Difficulties arise using such devices with very fine powders because air currents caused by the rotation of the system can blow the fines away. Furthermore, if the powder is cohesive (sticky), the powder flow through the funnel can be impeded. In some situations, the flow properties of the powder can be modified by adding a silica flow agent, provided this does not interfere with the size characterization procedures used subsequent to the sampling procedure (8).

Another approach to the sampling of powders has been developed by Kaye and coworkers. In this procedure, the powder is thoroughly mixed in such a way that any sample taken at random is a representative sample (9, 10). The equipment used in this technique is shown in Fig. 1c. The mixing chamber is placed in a rotating drum lined with dimpled foam. The foam serves two purposes: first, it promotes quiet tumbling of the chamber, and second, it accentuates the lifting power of the rotating drum so that it lifts the partially filled chamber up the wall of the drum until it tumbles chaotically to a new position of equilibrium before it is again lifted up for the next tumble. The chaotic tumbling of the mixing chamber creates ideal conditions for powder mixing, and short mixing times have proved to be efficient in mixing the ingredients (11). At the end of the tumbling procedure, a small sampling cup attached to the lid of the mixing chamber is used to retrieve a representative sample. The tumbling chamber

can have various geometric shapes, and provided that dimpled foam is used, even a common laboratory jar can be tumbled. Efficiency of mixing is reduced if the jar is more than half-full, because this imposes restrictions on the free, random movement of the powder in the chamber. It is recommended that this type of device be used to homogenize any powder sample before using a subsample in an experimental investigation, if the sample has been kept for some time or has been poured in a laboratory environment. Powder segregation mechanisms are far more widespread in the laboratory than generally known (12, 13).

In Fig. 2, a new pneumatic sampler developed by Kaye and coworkers is shown. This instrument has the advantage that one can take a sample of any specified size by changing the position of the sintered frit in the central tube that is used when the sample is drawn into the equipment. Lubricating air provided through both concentric tubes enables the sampling device to be moved to a particular location within a powder. When the sampling location is reached, the airflow to the center tube is reversed to draw in the desired sample (14).

For sampling an aerosol system, a cascade impactor is used to fractionate the aerosol into various size ranges. A full discussion of these instruments has been reported by Kaye (1). Aerosols can also be sampled through various filter systems for subsequent examination by image analysis procedures. If the aim of the sampling process is to generate a deposit that can be examined through a microscope with an imaging system, surface filters, such as the Nucleopore filter and similar polycarbonate filters, are available in various pore sizes for filtering aerosols (15–17). The traditional paper filter is described technically as a depth filter. Although depth filters can sometimes be rendered transparent by using immersion oil, it is normally difficult to view the fine particles on depth filters because they penetrate into the pore structure of the filter (18).

The study of the size distributions of therapeutic aerosols creates a very difficult sampling task for the specialist. Wherever possible, size characterization of an aerosol system should be carried out *in situ*, using diffractometers as discussed later.

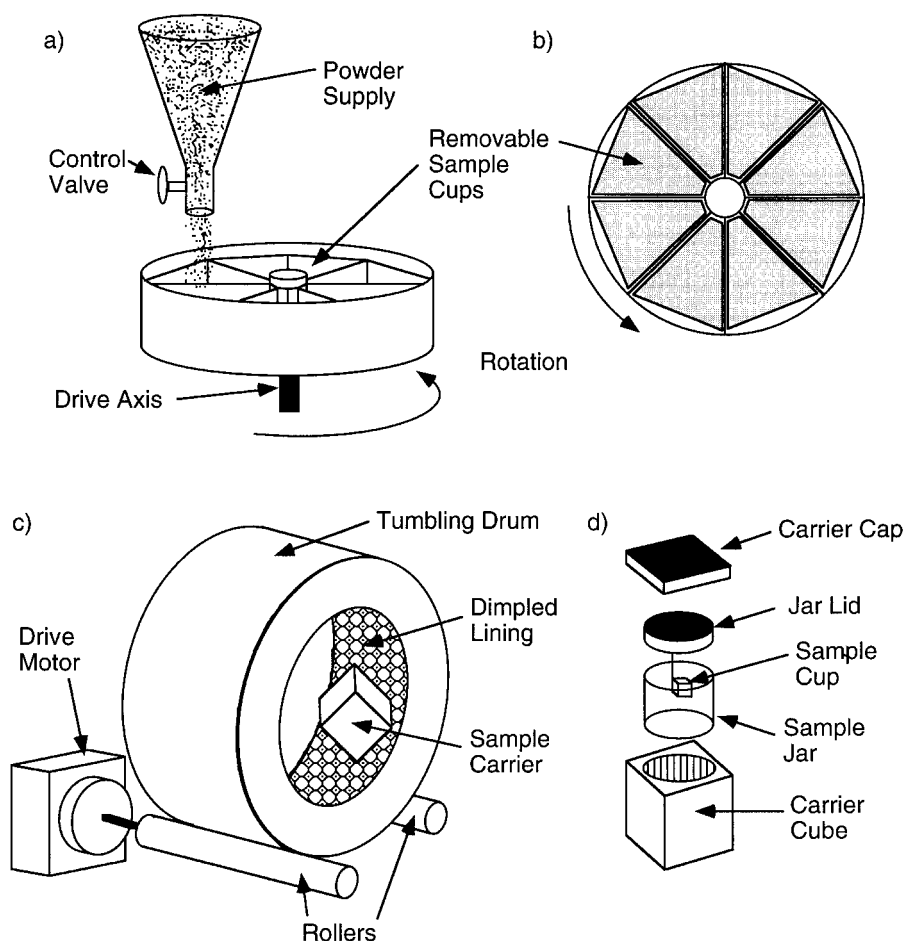


Fig. 1 Two devices that can be used to obtain a small representative sample from a larger supply. a) Systematic representative sampling of a powder can be achieved with the spinning riffler (7); b) the spinning riffler is composed of several removable sample cups; c) chaos-generating devices, such as the free-fall tumbling mixer, can be used for powder homogenization and sampling (11); d) exploded view of the sample carrier.

SAMPLE PREPARATION

In many situations, a powder sample to be characterized has to be prepared in a specified format for the characterization procedure. Thus, the powder may have to be spread out on a glass slide before microscopic examination, or a suspension of the material may have to be prepared in an appropriate liquid or gas. The act of dispersing the powder to be studied can change its size distribution radically, and the procedure used to prepare the sample for characterization should allow for what is known as the operational integrity of the fine particles. Thus, if the powder is to be dispersed in water, the use of ultrasonics to disperse it in the liquid can result in the shattering of agglomerates that would normally persist throughout the manufacturing and usage processes.

In general, the technology used to disperse a powder before a characterization study should match the severity of

dispersion forces that the powder will experience in use. Alternatively, if ultrasonics is used to generate a well dispersed powder, the analytical procedure protocol should be defined rigorously to avoid variations from operator to operator. Dispersing agents are frequently used, and great caution should be exercised because they can alter the structure of the system in a fundamental manner (19).

SIZE CHARACTERIZATION OF FINE PARTICLES AND POWDERS

Direct Examination with Microscopes and Other Imaging Devices

Extensive pioneering fine particle characterization studies were carried out by Heywood and Hausner, respectively,

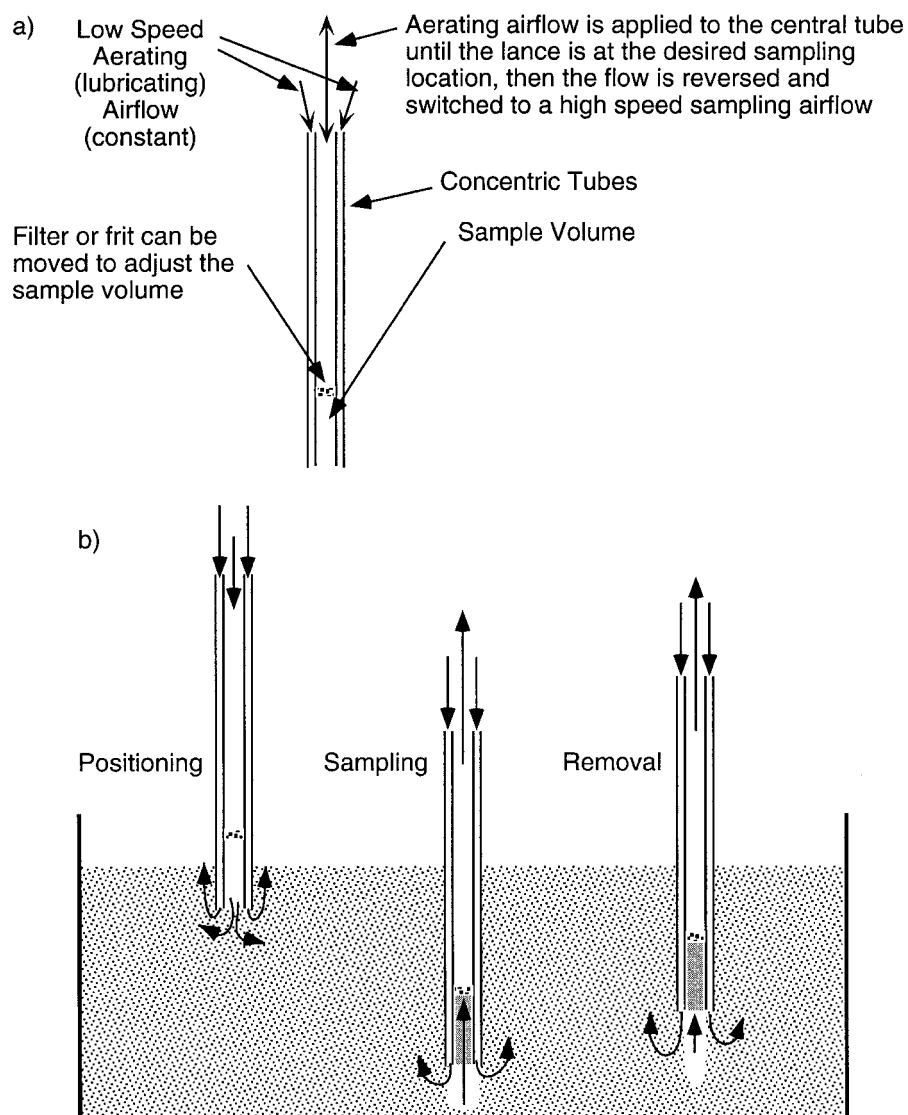


Fig. 2 The improved pneumatic lance is constructed from concentric tubes. Low-speed aeration from both tubes allows the lance to be inserted with minimum disturbance to the powder supply. Once the sampling location is reached, a high-speed flow is applied to the central tube to capture a sample below the filter or frit. a) Construction of the pneumatic sampler; b) process used to obtain a sample from a bulk of powder.

by examining images through microscopes and other imaging devices (20, 21). In these early works, the areas, or dimensions, of profiles were measured by direct comparison of the profile images with sets of reference circles engraved on what was known as an eye piece reticule (1). Recently, several sophisticated systems have been developed for computer-aided image analysis (1, 22).

The increased power of processing logic, available in modern computer-aided analysis systems, makes it possible to characterize the shape of fine particle profiles using Fourier analysis techniques and to describe

structures by means of fractal dimensions. The Fourier analysis techniques can be carried out in one or two dimensions (23). When exploring the structure of a profile in one-dimensional space by means of Fourier analysis, a reference point is located within the profile, and a geometric signature waveform is generated by rotating a vector at uniform angular velocity around the perimeter of the profile. The magnitude of the vector plotted against the angle of the vector generates the waveform. This waveform is subjected to Fourier analysis to generate a power spectrum of the various harmonics contributing to

the structure of the waveform. This technique is useful for rounded objects but generates complex information if there are deep convolutions or sharp edges on the profile. For such profiles, two-dimensional Fourier transforms can be generated by computer (23, 24).

A different procedure for describing the structure of rugged profiles has been developed from the theorems of fractal geometry (24–29). The basic concept used in fractal geometry is to add a fractional number to the topological dimension of a system to describe the space-filling ability of the system being described. Thus, in Fig. 3a, all the lines have a topological dimension of one. The fractional number added to this dimension creates the boundary fractal dimension of the line, a parameter that describes the ruggedness of the line. In Fig. 3b, the basic logic used to evaluate the fractal dimension of a profile of a powder grain by the equipaced method is shown. Polygons are constructed on a digitized form of the profile, a sequential set of x,y coordinates representing the profile, by pacing out a given number of steps around the profile. Polygons constructed in this way become perimeter estimates at the inspection resolution represented by the number of steps paced out along the profile. The perimeter estimate and the resolution are normalized with respect to the maximum projected length of the profile. To estimate the magnitude of the fractal dimension of the profile, the normalized values of the perimeter estimates are plotted against the normalized resolution, as indicated in Fig. 3c. This plot is known as a Richardson plot after a pioneer of the detailed studies of convoluted profiles such as these of islands (25, 26). The slope of the data lines on the Richardson plot represents the fractional number that has to be added to the topological dimension to describe a structure of the boundary. As with the profile studied in Fig. 3b, some fine particle profiles exhibit different fractal dimensions at different levels of inspection. Thus, in this profile, what is known as the structural boundary fractal dimension is revealed by the course resolution data, and what is known as the textural boundary fractal dimension is revealed at high-resolution inspection. The structural fractal dimension probably governs the packing and flow properties of a powder, whereas the texture governs the dissolution rate, adsorptive capacity, and chemical activity.

Because of the large amount of visual information imparted by an image of a fine particle, there has been a tendency to regard image analysis inspection as a fundamental method against which all others should be measured. The principal problem involved in the inspection of a system by image analysis is the difficulty of deciding what constitutes a separate and operationally functional fine particle. A failure to record the density of coverage of the surface used in a microscopic study of a

powder is a major source of uncertainty in the value of the reported data, because random juxtaposition of fine particles on a surface can make them appear as preexisting agglomerates (30).

Sieve Fractionation

Sieving is a widely used method for characterizing the range of grain sizes present in a powder. In this technique, a quantity of powder is separated into two fractions on a surface containing holes of a specified uniform size. The two main problems associated with sieve characterization are the difficulty of determining the point at which the fractionation process is complete and coping with the variation in sieve apertures present in new and worn sieves. For a discussion of see Refs. 1 and 22. The photomicrograph of a woven wire sieve surface in Fig. 4a illustrates the problems associated with the variations in sieve apertures. In Fig. 4b the data on the variation of the sieve apertures as characterized by various studies are plotted (31). The midpoint diameter of the trapezium created by the projected image of the woven wire surface provides the operational diameter of the sieve, and the size distribution of the apertures can be measured with the help of image-analysis techniques. The size distribution determined in this way can then be normalized by the nominal aperture of the sieve. It can be demonstrated from the data in Fig. 4b that the aperture distribution is Gaussian.

In an alternative technique, spherical glass beads are separated on the sieve, and when the remaining oversize beads are poured off of the sieve, the near mesh-sized beads trapped in the mesh are removed by inverting the sieve and giving it a sharp rap on a surface. The beads collected in this way are then sized. If an irregularly shaped powder is used in such an experiment, a typical set of grains trapped in the openings of the sieve allows one to determine the shape distribution of the profiles, as illustrated in Fig. 5. Electroformed sieves have a much narrower distribution of aperture sizes but are more fragile and expensive than woven wire sieves (1).

Sedimentation Techniques

Sedimentation procedures to evaluate particle size in terms of the equivalent spheres, which have the same settling speed in laminar flow conditions, are the basis of many techniques used to characterize fine particles. A suspension of fine particles is prepared, and their falling speed is determined with an immersed balance pan or by monitoring the settlement of the fine particles with the help of light or X-ray beams. The measured falling speeds of the fine

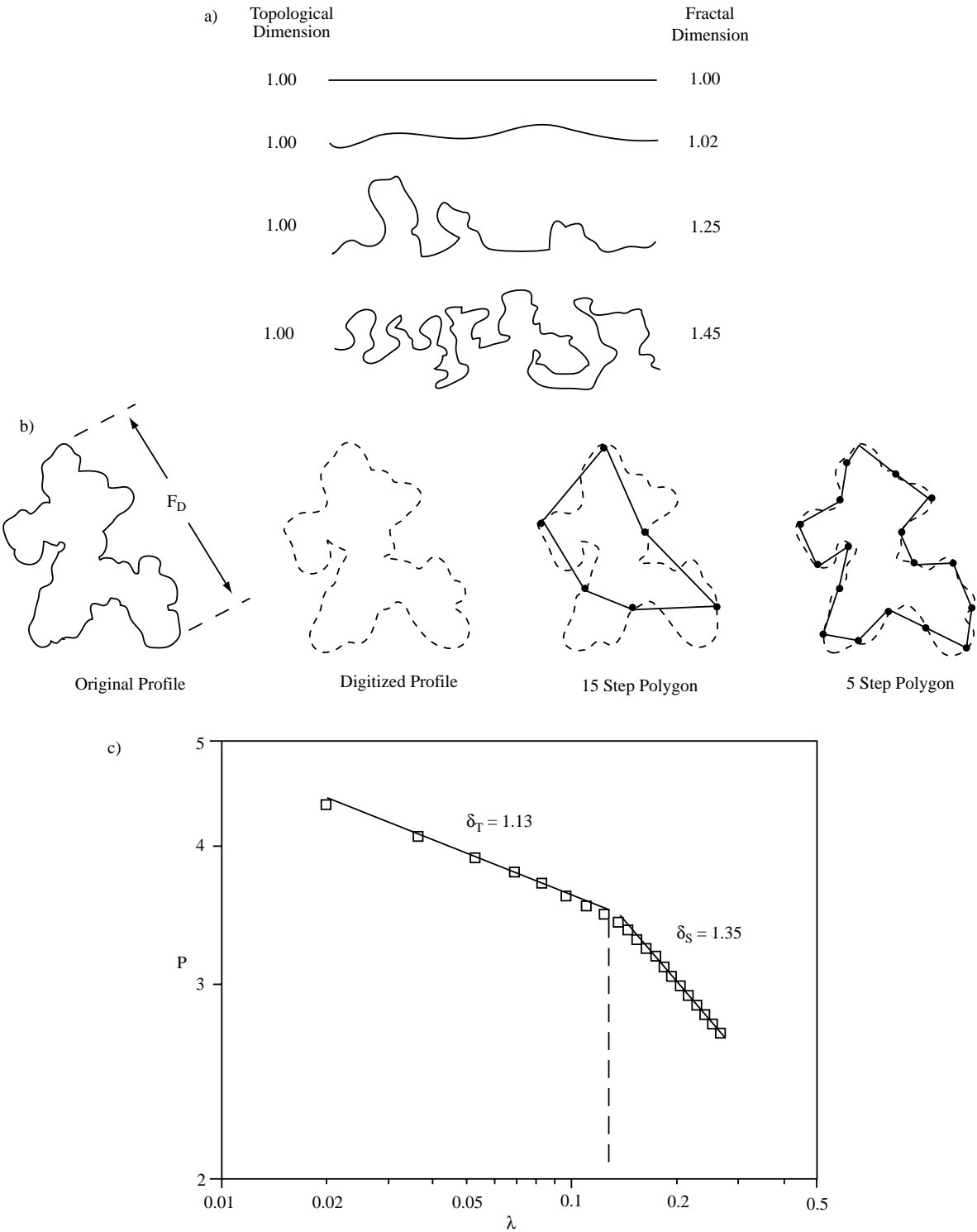
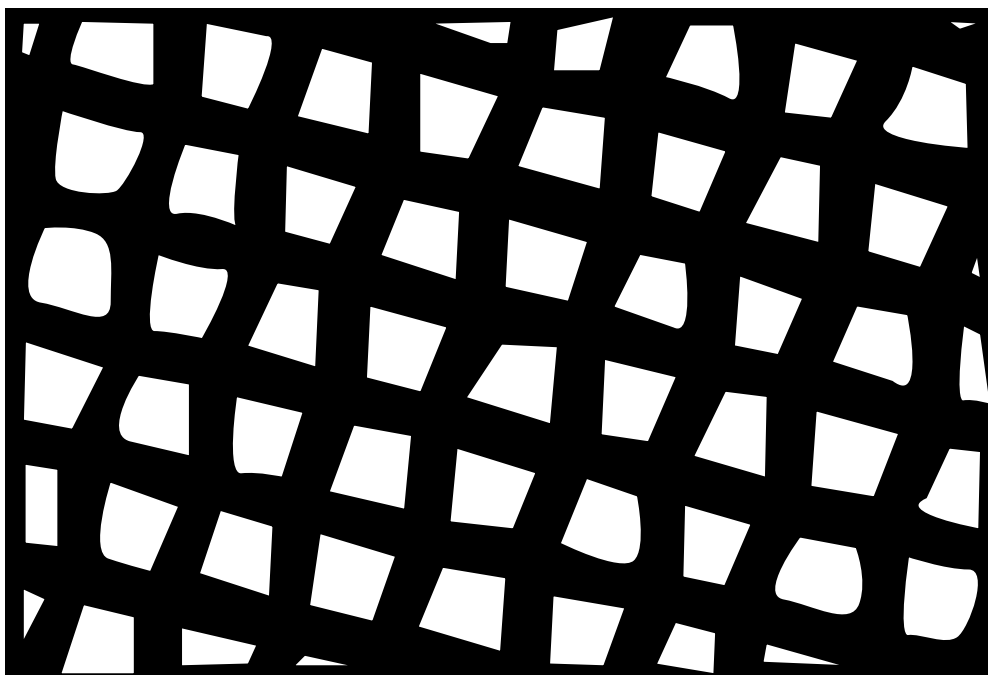


Fig. 3 Fractal dimensions can be used to evaluate the rugged structure of fine particles. a) Fractal dimensions used to describe the ruggedness of various lines; b) physical basis of the equipaced exploration technique for evaluating the fractal dimensions of rugged boundaries; c) data generated by the equipaced exploration technique for the profile of b; δ_S ; structural boundary fractal dimension; δ_T ; textural boundary fractal dimension.

a)



b)

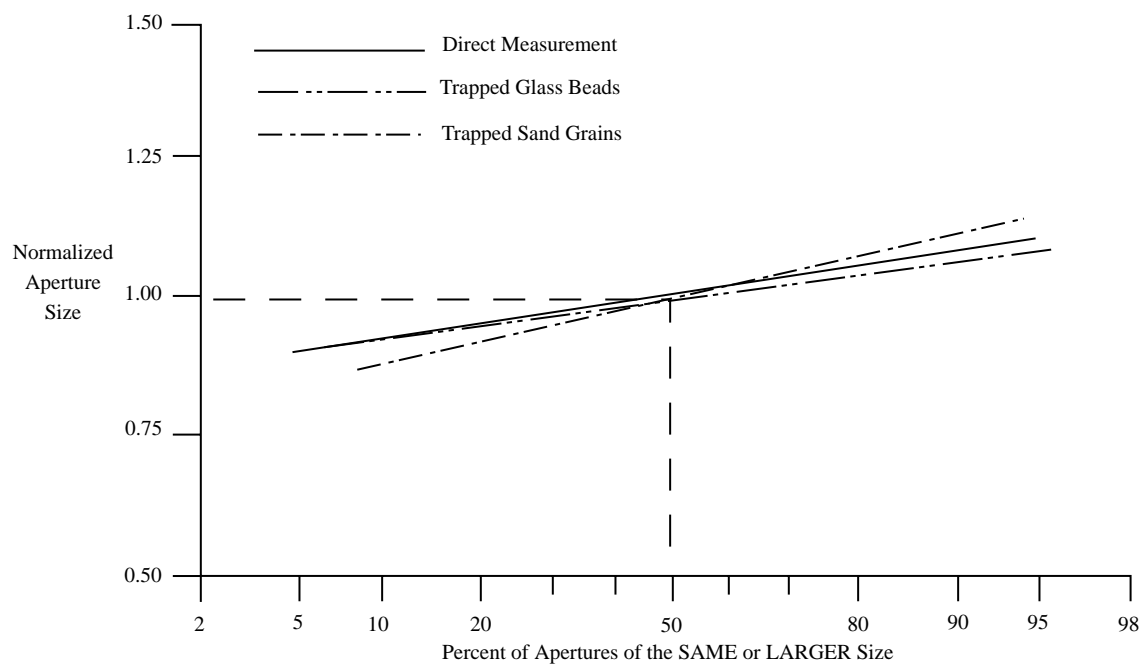


Fig. 4 Variations in mesh apertures of a woven wire sieve can be determined by several methods. a) Photograph of a woven wire sieve; b) the variations in sieve apertures can be determined either by direct inspection of the aperture or by examining near mesh size fine particles trapped in the apertures during the sieving process.

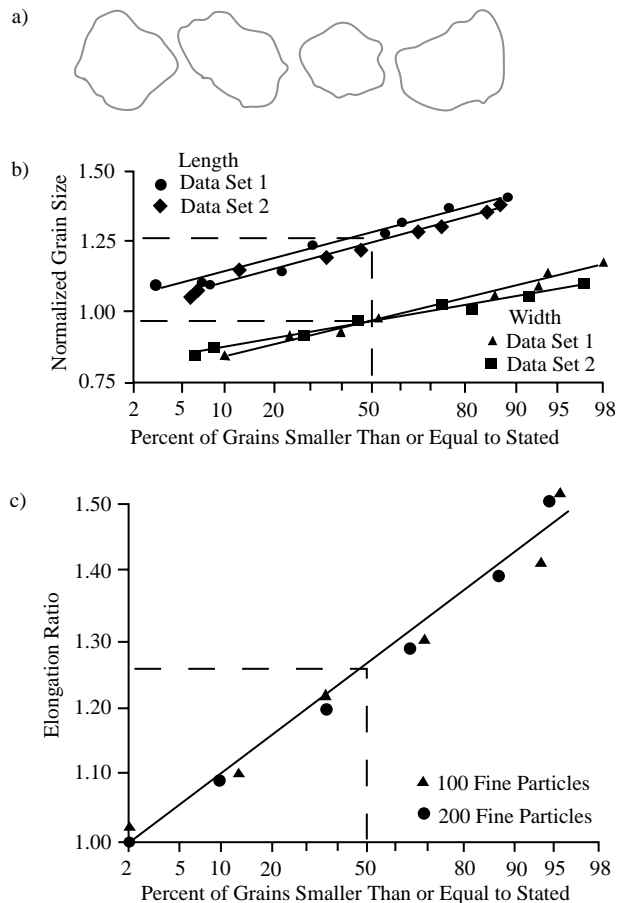


Fig. 5 When calibrating a sieve mesh with trapped irregularly shaped grains, a subset of powder grains is obtained that can be used to generate a shape description of the powder grains (31). a) Profiles of typical sand grain trapped in a sieve mesh; b) length and width distributions of sand grains trapped in a sieve mesh; c) elongation ratio (shape) distribution for sand grains trapped in a sieve mesh.

particles are inserted, along with the other appropriate parameters of the suspension, into Stokes' equation

$$d_s = \sqrt{\frac{18\eta t}{(\rho_p - \rho_L)gh}}$$

where d_s = Stokes diameter of the fine particle; η = viscosity of the liquid; g = acceleration due to gravity; h = distance through which the fall is timed; t = time required to fall the distance h ; ρ_p = density of the powder; and ρ_L = density of the liquid.

The configuration of the actual instrument used to measure the Stokes diameters of fine particles varies among instrument manufacturers (1, 22). Devices that

monitor the fine particle sedimentation with a light beam are known as photosedimentometers. Because of the difficulties of interpreting the concentration measurement of fine particles with diameters close to that of the wavelength of the light being used to monitor the dynamics of the suspension, some instruments use X-Ray beams to monitor the fine particle movements. Other instruments use centrifugal force to accelerate the settling dynamics of the suspended fine particles.

Sedimentation methods were the dominant size-characterization procedures in the 1950s and 1960s. In recent years, they have been replaced in the powder laboratory by the diffractometers. Diffractometers have the advantage of speed, but problems occur in the interpretation of the diffracted light signals. The X-ray-based sedimentometer manufactured by the Micromeritics Corporation, called the Sedigraph, is still widely used, partly because its use is written into some industrial standards governing size-characterization procedures (32, 33).

In recent years, disk centrifuges have been revived to characterize fine particles smaller than $1\ \mu\text{m}$, basically because of the work by Provder and coworkers in cooperation with Brookhaven Laboratories Ltd. (34).

Diffractometers

The advent of the laser has made the generation of diffraction patterns by a suspension of fine particles a relatively easy task. At the same time, the rapid development of computer-processing equipment and specialized photocells has made it possible to process the information in a group diffraction pattern to generate the particle size distribution of the fine particles in suspension. One of the first commercially available devices for generating size-distribution information from group diffraction patterns of a randomly dispersed array of fine particles was the CILAS device (Denver Autometrics, Inc., Boulder, CO) first developed in France to measure the size distribution of cement (35). Companies that have developed diffractometers do not divulge the structure of their software. The user of this equipment should be careful to obtain information on the data-processing protocol followed in any specific instrument to change the diffraction information into a size-distribution function. Some of the instruments assume a given distribution function and curve fit to accelerate the data processing. Sometimes such curve fitting can distort the data generated (36). Several manufacturers of these instruments provide extensive technical data on their performance, and the International Standard Organization is currently preparing a standard procedure for diffractometers. The laser

diffractometers are particularly useful for studying the size distributions of sprays and aerosol clouds (37–43).

Time-of-Flight Instruments

Another type of instrument that has been made possible by the availability of lasers is known as the time-of-flight instrument. Here, a narrow focused beam of laser light

explores an area of a suspension. The size of the particles in suspension is measured by the time it takes for a laser beam to pass across the profile of the fine particle. Sophisticated optical recording devices and electronic editors are used to generate the size distribution data from the information generated by the device.

An instrument developed and marketed by Galai Instruments (Haemek, Israel), sold in the United States for

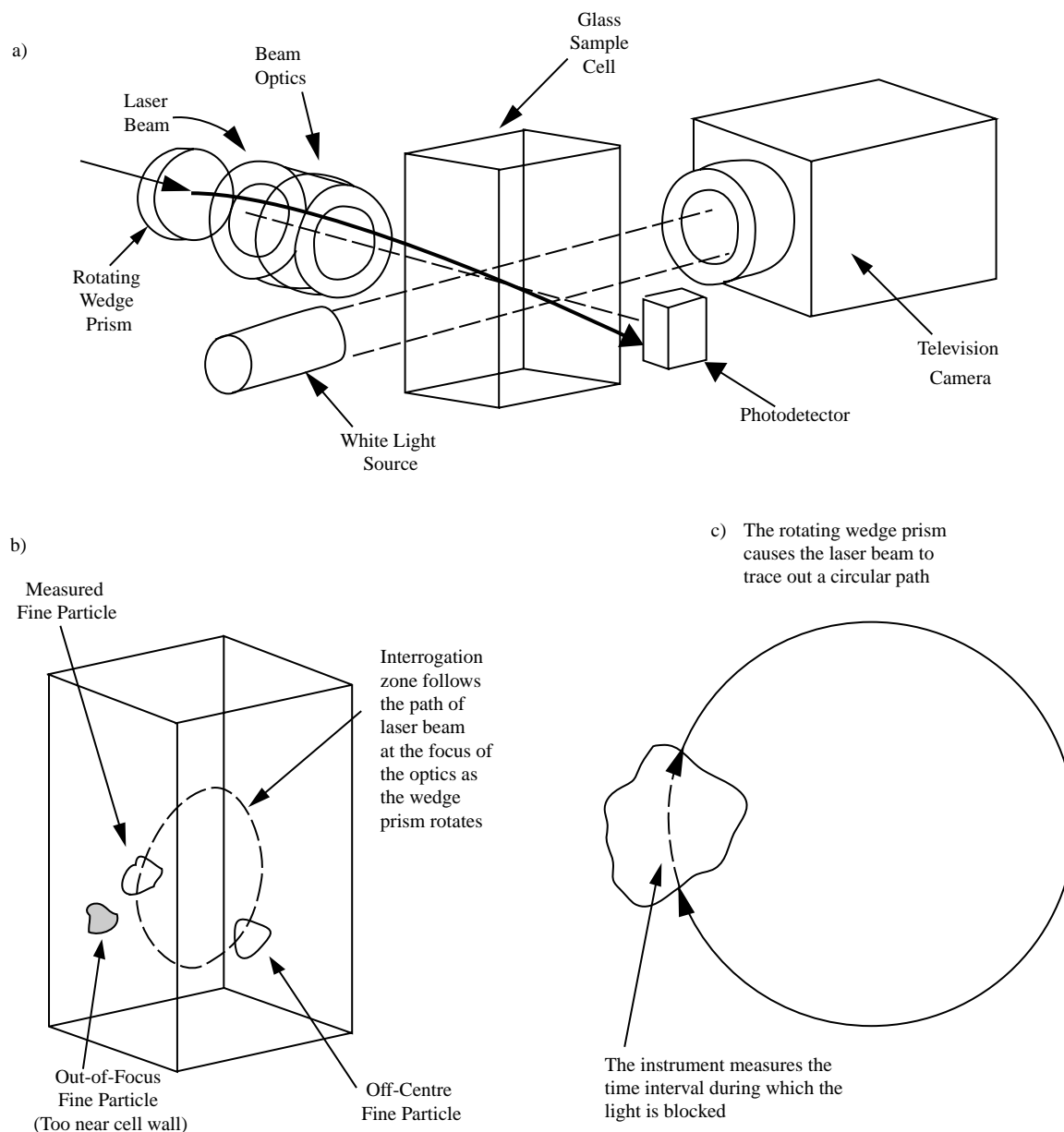


Fig. 6 The Galai particle-size analyzer uses a laser beam and a rotating wedge prism to measure the size of fine particles by the length of time the fine particle blocks the light reaching the photodetector. a) Basic layout of the Galai instrument; b) the laser beam traces a circular path within the sample cell, and the logic of the instrument rejects any particles that are off-center or out of focus; c) the length of time that the laser beam is blocked is related to the size of the fine particle. (From Ref. 44.)

several years by the Brinkmann organization but now marketed directly by Galai Instruments, is shown in Fig. 6. A useful feature of this instrument is that as the fine particles are being characterized by the scanning laser beam, they are also imaged on a television screen in such a way that any agglomeration can be detected during the analysis. The logic of the Galai system allows the fine particle shape to be measured concurrently with size (44).

Another time-of-flight size analyzer is known as the LASENTECH (Bellevue, WA). This system is portable and has been suggested for use as an online monitor for fine particles moving in a system as well as in the laboratory (45). A different type of time-of-flight instrument is known as the Aerosizer (TSI, St. Paul, MN) (46). Here, the stream of aerosol fine particles is accelerated across a gap defined by two laser beams. The time of flight across this gap is measured from the light signals scattered from the two light beams, and an electronic editor ensures single occupancy for the measurement series. The larger fine particles are slow to accelerate across the gap, whereas the smallest move with the speed of the feed air jet. The system is calibrated using standard fine particles.

Another similar instrument is also manufactured by TSI. The basic system for measuring the aerodynamic

diameter of aerosol fine particles is shown in Fig. 7. Using this instrument, the velocity of a moving fine particle being accelerated across the inspection zone is measured with the Doppler shift in two beams that have a different directional reference to the moving airstream. From one perspective, the two lasers beams can be regarded as creating interference fringes, whereas the aerosol fine particle moving across the fringe system creates an oscillating signal that can be related to the fine particle size via calibration measurements. The fact that the instruments with which aerodynamic sizes are measured by the movement of the fine particles across crossed laser beams involves laser Doppler shifts is not immediately obvious from reading the trade literature of companies marketing this type of instrument. Indeed, in this class of instruments, the interpretive theory is complex, and the user is generally provided with a calibrated instrument to carry out the characterization studies of interest (46, 47).

Photon-Correlation Spectroscopy

Another instrument, in which the physics of the measurements are not immediately obvious to the outside observer, is part of a group of instruments variously referred to as photon correlation, dynamic light scattering,

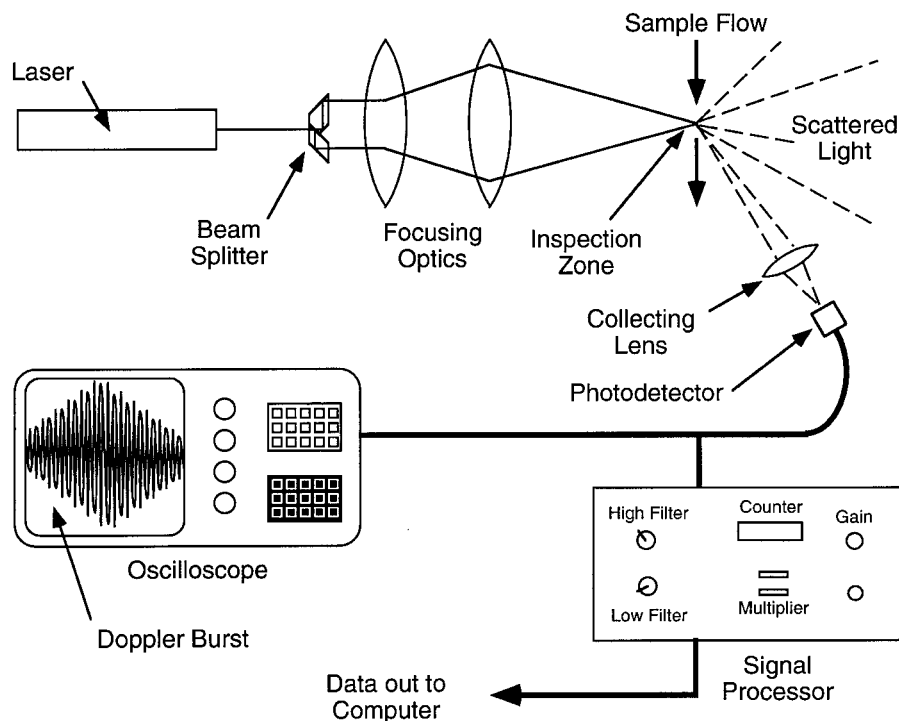


Fig. 7 A Doppler-shift procedure for measuring the aerodynamic size of aerosol fine particles is used by the TSI aerosol particle sizer. (Form Refs. 1 and 46.)

or quasielastic light scattering spectroscopes (often referred to as PCS, DLS, or QUELS). In this discussion, the term photon-correlation spectroscopy is used (48–51). Its physical basis is the monitoring of the Doppler shifts in reflected laser light created by the Brownian motion of submicron fine particles. In some cases, the technique can also be applied to fine particles of several microns in diameter. The equipment for actually measuring the Doppler shifts is relatively simple, but the overall expense is increased by the data-processing that is usually included. This instrument is useful for studying relatively simple size distributions such as latex suspensions. However, interpreting a wide range of sizes in suspension with this technique can involve complex data processing that, if carried out incorrectly, can generate confusing data (48, 50, 51).

Stream Counters

Another size characterization instrument group is known as stream counters, in which a stream of fine particles is passed through an inspection zone. The physical properties of the inspection zone are changed by the presence of the fine particles. The size of the fine particle is deduced from this change. In the Coulter counter, the fine particles to be characterized are placed in an electrolyte, and a stream of suspension is passed through an orifice between two electrodes as shown in Fig. 8 (52). The size of the fine particle is deduced from the measured resistance change between the electrodes.

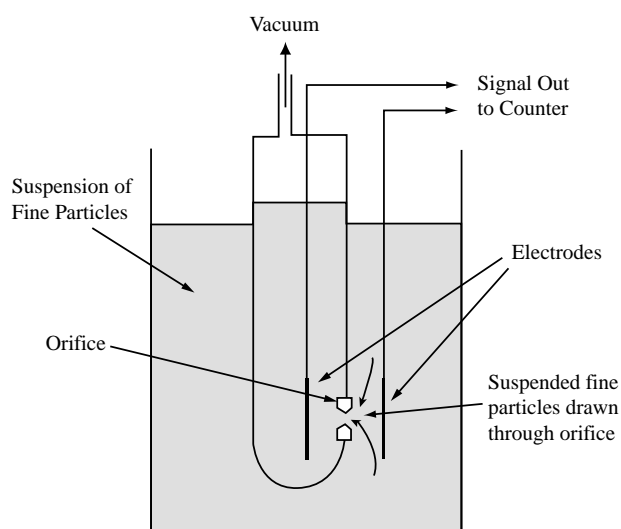


Fig. 8 Schematic representation of the operation of the Coulter counter. (Form Ref. 52.)

A major problem with stream counters is the single occupancy of the fine particles in the inspection zone. Should there inadvertently be two fine particles in the orifice, they register as one large fine particle, and the counting of the smaller particles has a deficit of two. This type of error is referred to as primary count loss (the undercounting of the smaller fine particles) and secondary count gain (the false registering of larger fine particles owing to multiple occupancy of the zone). Normally, the analysis with this type of instrument is carried out in a series of increasing dilutions until further dilution does not affect the measured size distribution. A difficulty sometimes encountered with this method is the availability of a conducting fluid that does not interact with the fine particles to be inspected. Over the years, various sophisticated data-processing techniques have been used to allow for problems associated with the Coulter counter, for example, when the fine particles are too close to the walls of the orifice or have extreme shape (1, 22).

In another group of stream counters, the fine particles in the inspection zone are monitored with a light beam. Various models of this type of instrument have been developed to count fine particles in liquids; others are specialized for the counting of dust fine particles in the air (1, 22, 53–55).

Elutriators for Size-Characterization Studies and Fine-Powder Fractionation

Elutriators are a class of instruments that fractionate fine particles according to their size by manipulating them in a moving fluid. They are among the first devices used for measuring size distributions of powders by fractionating them into various size groups and weighing the amount of powder in each group. The Roller elutriator was widely used in the powder metals industry (1). In recent years, elutriators have tended to be displaced from common use by diffractometers and other optically based instruments. They are still extremely useful, however, for fractionating powders into different sizes to study the physical variations of properties with size. Thus, a drug powder can be fractionated into various sizes to study the dependence of the bioavailability of a drug on its particle size. Fig. 9 shows three basic types of elutriators for fractionating powders and studying aerosol fine particles. In the gravity elutriator shown in Fig. 9a, air or another suitable fluid is passed upward through powder placed on a filter. As the air moves up through the column, the velocity of the moving fluid can be adjusted to move all fine particles below a certain size from the elutriator body to a fines collector, which may be a filter or a cyclone. The size limit, defining the size of the fine particles remaining

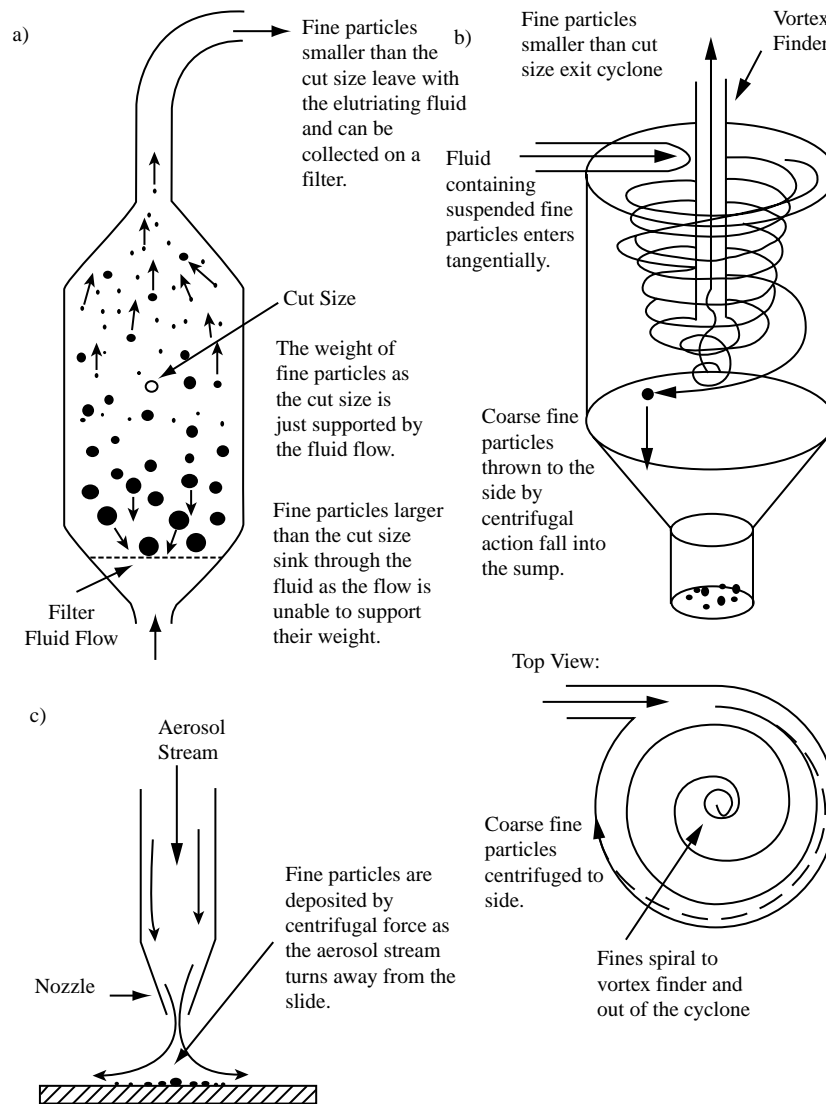


Fig. 9 Elutriators can use gravity or centrifugal force to separate fine particles into undersize and oversize fractions. a) Gravity elutriator; b) Cyclone (centrifugal elutriator); c) Impactor. (From Ref. 1.)

in the elutriator, is called the cut size of the elutriator. Because it is difficult to control the movement of the fluid and because of turbulence, the cut size and the fractionating power of an elutriator are not precise. A microscope or other suitable device is needed to investigate the actual fractionation of the powder in a given elutriator.

Cyclones are widely used in industry to fractionate powders. The operation of a cyclone, which is in fact a centrifugal elutriator, is shown in Fig. 9b (56). The stream of fluid suspension containing the fine particles to be fractionated enters the top of the cylindrical body

tangentially. The fluid stream is made to spiral downward through the body of the cyclone until it can reverse its flow and leave through a pipe known as the vortex finder. As the feed stream spins around the body of the cyclone, the coarse particles in suspension are thrown to the wall by centrifugal force. At the wall, the coarser particles fall down into the conical bottom of the cyclone. The cut size of the cyclone, which determines how small the fine particles leaving through the vortex finder are, is determined by the dimensions of the cyclone and the velocity of the fluid stream. Small cyclones are widely used in occupational hygiene studies (56).

Another device for depositing fine particles from an airstream is the jet impactor, shown in Fig. 9c. The airstream containing the suspended particles is impinged onto a glass slide. As the airstream is forced to turn because of the slide under the jet, centrifugal force pushes the suspended fine particles onto the slide. The smallest fine particle, which is just deposited on the slide, is determined by the jet-slide configuration and the speed of the airstream moving through the equipment (1).

Giddings and coworkers have developed a series of cross-flow classifiers known as field-flow fractionation devices. The procedures, often referred to as FFF, are defined by Giddings (57) as: "a family of high resolution techniques capable of separating and characterizing materials in the macromolecular and colloidal range and beyond." Applications of FFF span a ten- to fifteen-fold mass range, extending from molecules under 1000 molecular weight to particles 100 microns in diameter. Particles as diverse as cells, subcellular particles, viruses, liposomes, protein aggregates, fly ash, waterborne colloids, and industrial latexes and pigments have been separated.

Characterization of Powder Surface Areas

In powder studies, the surface area of the powder is an important parameter. It can be measured directly by means of gas adsorption studies, in which the amount of gas or another molecular item, such as dye molecules adsorbed onto the powder to form a monolayer, is determined. Several books have been written describing the theory and procedures for gas adsorption studies. Before 1977, it was believed that one of the basic problems with surface area estimates by gas adsorption was that uncertainties in the knowledge of the cross-sectional area of the adsorbed molecules made the estimates depend on the gas being used (58). In recent years, the gas adsorption studies of surface areas have been reinterpreted from the viewpoint of fractal geometry (59). It is now recognized that the surface area measured, using a given gas, depends on the accessibility of the rough surface to the adsorbed molecules, as illustrated in Fig. 10a. In a study of a series of adsorbent molecules of increasing size, Avnir and coworkers have shown that the surface area estimates can be plotted against the molecular size to obtain a Richardson plot from which the fractal roughness of the powder surface can be deduced (59, 60); a graph generated by Avnir and coworkers is shown in Fig. 10b. The slope of its data line can be used to deduce a fractal dimension of the rough surface. The fractal description of powder roughness is an important parameter for the bioavailability of a drug or the chemical reactivity of powder. Neimark has recently

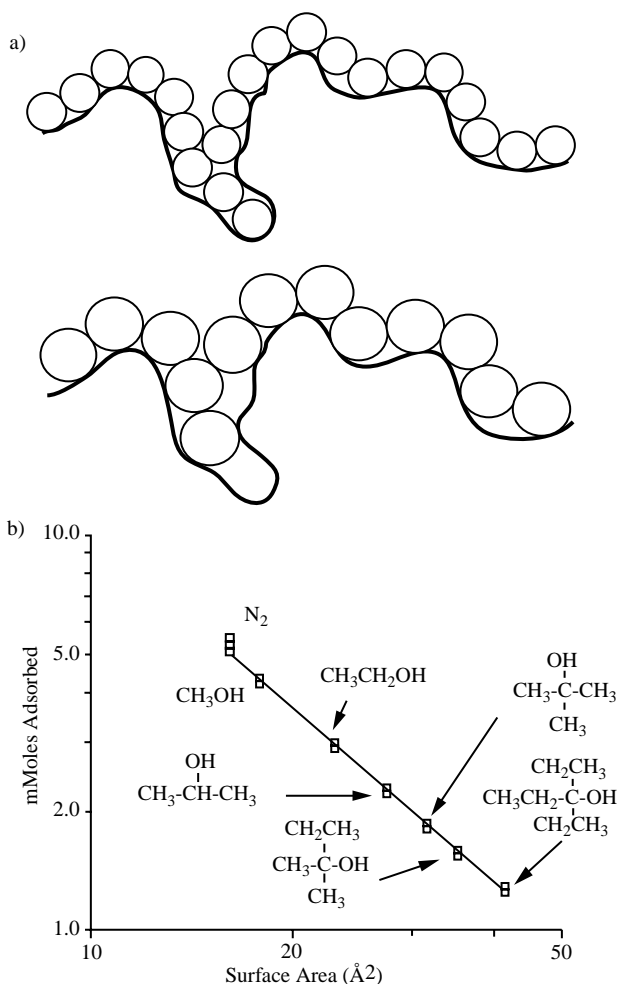


Fig. 10 Gas adsorption data permit the determination of the fractal dimension of a rough surface. a) The surface is estimated from the number of gas molecules that cover the surface. The estimate depends on the size of the gas molecule; b) the fractal dimension of a surface is derived from the results of gas adsorption studies with several different sized molecules (From Ref. 59.)

described a method for calculating the surface area and roughness of a powder by studying capillary condensation of a liquid on a powder (61).

The fineness of a powder can be studied with the help of permeability techniques, in which the resistance to fluid flow of a powder plug is measured and the fineness of the powder deduced from this measured resistance using various equations such as the Kozeny–Carmen equation. The interpretive equations used to calculate the surface area from permeability measurements make several assumptions concerning the pore structure of the packed powder bed, and the measured surface area from permeability studies should be regarded only as a measure

of fineness and not as an absolute measure of surface. Instruments such as the Fisher Subsieve Sizer and the Blaine Fineness Tester are permeability-based methods that, in the past, have been widely used in industry and are still often used for quality control in industry. A major advantage of permeability methods is that they use large amounts of powder, which minimizes the problems of sampling (62, 63).

PORE-SIZE DISTRIBUTION MEASUREMENTS

When the pore size of a packed powder bed or of the structure of porous powder grains is of interest, mercury-intrusion studies can be used to investigate the pore structure. Fig. 11 shows data of a study using the mercury-intrusion technique to examine the structure of a powder bed of porous grains (64). The amount of mercury entering a bed at different pressures is used to generate the data. Using the known contact angle of mercury with the material of the powder, the applied pressure can be interpreted in terms of the capillary tube through which mercury moves at that pressure. However, there has always been some controversy as to the physical significance of mercury-intrusion data because it only measures access pore diameter, not the volume of the pore behind the neck. (Theories interpreting mercury-intrusion data in terms of pore diameter are often referred to as ink-bottle interpretive models, the idea being that it is the neck of the ink bottle that represents the penetration diameter, not the diameter of the bottle behind the neck.) For the data in Fig. 11a, at low pressures the mercury is entering the voids between the grains of the powder, but when a pressure of approximately 2000 psi (13.8 MPa) is reached, the mercury starts to intrude into pores within the powder grains having access diameters of the order of $0.1\ \mu\text{m}$. It has recently been shown that the traditional way of presenting mercury-intrusion data can be revised to generate a fractal dimension in data space. The revised data in Fig. 11a are shown in Fig. 11c. The slopes α and β of these diagrams are fractal dimensions in data space (65, 66).

STANDARD REFERENCE POWDERS

Some manufacturers of instruments for characterizing fine particles claim that their instruments do not need calibration. Such claims should be treated with skepticism. In practice, many sizing instruments have to be calibrated using standard powders available from several vendors (67, 68). Because the various methods of exploring the size

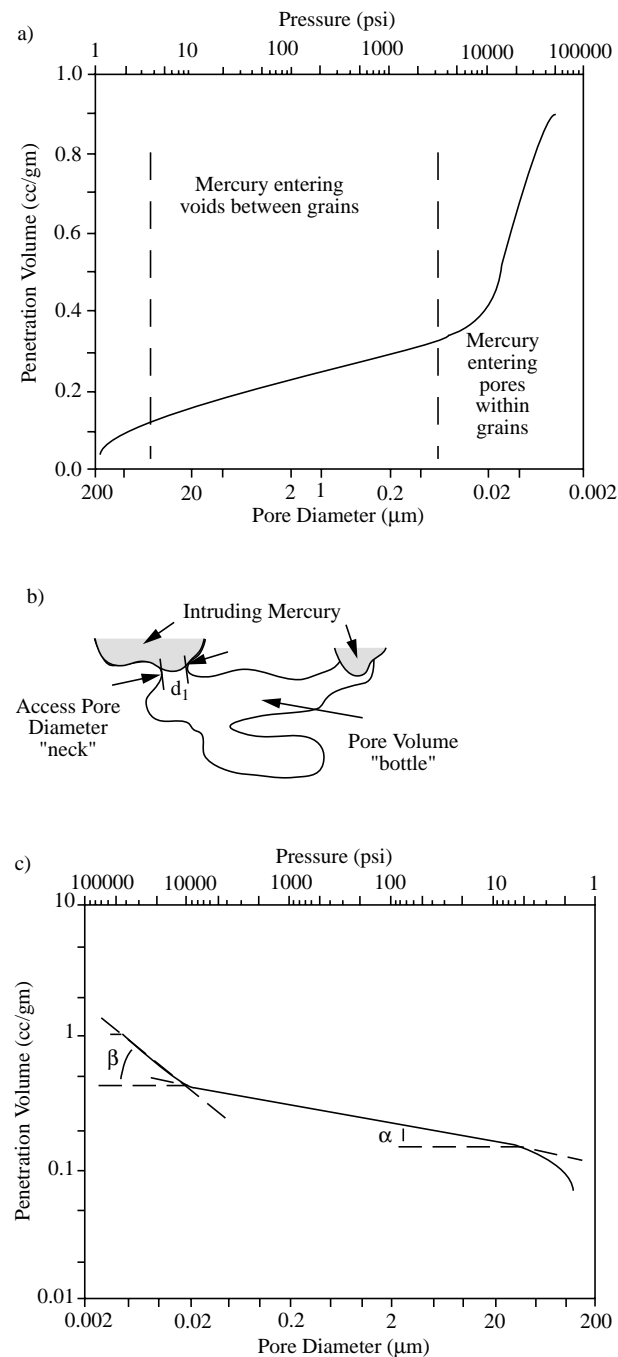


Fig. 11 Mercury intrusion porosimetry involves applying increasing pressure to a sample under study and noting the volume of mercury entering the pores within the sample. a) Traditional representation of mercury intrusion data; b) the physical significance of “ink bottle” theories of mercury intrusion have always been the subject of debate; c) a possible reinterpretation of the data in a) as fractal data. α and β are the slopes of particular regions of the resulting curve, which may be interpreted as fractal dimensions.

distribution of a powder evaluate different physical parameters, the size distributions of a powder generated by different methods do not always agree. The relationship between distribution functions, as evaluated by different methods, should be explored experimentally (69–72).

REFERENCES

- Kaye, B.H. *Characterization of Powders and Aerosols*; Wiley-VCH: Weinheim, 1999.
- Kaye, B.H. Efficient Sample Reduction of Powders by Means of a Riffler Sampler. *Soc. Chem. Ind. Monogr.* **1964**, 18, 159–163.
- British Standards Methods for the Determination of Particle Size Powders. Part I, Subdivision of Gross Sample Down to 0.2 ml*; BS3406, Part 1: London, 1961.
- Kaye, B.H. An Investigation into the Relative Efficiency of Different Sampling Procedures. *Powder Met.* **1962**, 9, 213–234.
- Sampling Equipment Literature is available from Gustafson, 6340 LBJ Freeway, Suite 180, Dallas, TX 75240.
- Sampling Equipment Literature is available from Gilson Company Inc., P.O. Box 200, Lewis Center, OH 43035.
- Information on the Spinning Riffler System is available from Microscal Ltd., 20 Mattock Lane, Ealing, London, W5 SBH.
- Kaye, B.H. Powder Rheology. *Powder Mixing*; Chapman & Hall: London, 1997; 103–161, Ch.3, Now available through Kluwer Academic Publishers: Dordrecht, The Netherlands.
- Kaye, B.H. *Chaos and Complexity*; VCH: Weinheim Germany, 1993; 148–154, See also Section 5.2, Poisson Tracking as a Technique for Studying Chaotic Conditions in a Mixer, in Ref. 8, 220–223.
- Kaye, B.H.; Clark, G.G.; Bohan, Monitoring Mixer Performance using the Size Distribution Information on Samples taken from a Mixing Process. In *Proceedings of Fine Powder Processing 99*, The Pennsylvania State University, University Park, PA, Sept, 20–22, 1999.
- Kaye, B.H.; Clark, G.G. In Evaluating the Performance of Chaotic Powder and Aerosol Sampling Devices Using Tracker Fineparticles Proceedings of PARTEC, Nürnberg Conference on Particle Size, May, 1989; Nürnberg Messer Centrum, Nürnberg, 1989.
- Kaye, B.H.; Turbitt-Daoust, C.; Clark, G.G. Segregation Dynamics of Powders Subjected to Vibration Proceedings of the Powder and Bulk Solids Conference, Rosemont, IL, May, 4–7, 1998; Reed Exhibition Companies, Norwalk, CT, 1998.
- Kaye, B.H.; Fairburn, M.; Clark, G.G. Techniques for Monitoring and Combating Segregation in Powder Systems, Proceedings of the Powder and Bulk Solids Conference, Rosemont, IL, May, 10–13, 1999; Reed Exhibition Companies, Norwalk, CT, 1999.
- Kaye, B.H.; Fairburn, M.; Hood, K.A. Efficient Sampling Protocols. In *Proceedings of the Powder and Bulk Solids Conference*, Rosemont, IL, May, 8–11, 2000; Proceeding published by Reed Exhibition Companies, Norwalk, CT.
- Pall Gelman Laboratory, 600 South Wagner Road, Ann Arbor, MI 48103–9019 USA.
- The Poretics Corporation, 151 Lindbergh Avenue, Livermore, CA.
- Collimated Holes Incorporated, 460 Division St., Campbell, CA 95008.
- Allen, T. Application of Precision Transparent Sieves to the Determination of Low Number Concentrations of Oversized Particles in Powders. *Part. Part. Syst. Characteristics*; Wiley-VCH: 1992; 9, 252–258.
- Parfitt, G.D. *Dispersion of Powders in Liquids*, 2nd Ed.; Wiley: New York, 1973.
- Heywood, H.H. Size and Shape Distribution of Lunar Fines, Sample, 12057, 72. In *Proceedings of Second Lunar Science Conference*, Vol. 13, 1971; 1989–2001.
- Hausner, H.H. Characterization of the Powder Particle Shape, Proceedings of the Symposium on Particle Size Analysis, Loughborough, England, 1967, Society for Analytical Chemistry, London, 20–27 (See also Ref. 1, 205–232.)
- Allen, T. *Particle Size Analysis*, 4th Ed.; Chapman and Hall: London, 1992.
- Kaye, B.H. *Characterization of Powders and Aerosols*; Wiley-VCH: Weinheim, 1999, The Use of Fourier Techniques to Characterize the Shape of Profiles in Ch. 2, Direct Measurement of Larger Fineparticles and the Use of Image Analysis Systems to Characterize Fineparticles, 21–58, and Ch. 7, Light Scattering Methods for Characterizing Fineparticles, 205–232.
- Kaye, B.H. Part I: Rugged Boundaries and Rough Surfaces Applied Fractal Geometry and the Fineparticle Specialist. *Part. Part. Syst. Charact.* **1993**, 10 (3), 99–110.
- Kaye, B.H. *Chaos and Complexity, Discovering the Surprising Patterns of Science and Technology*; VCH: Weinheim Germany, 1993.
- Kaye, B.H. *A Random Walk Through Fractal Dimensions*; VCH: Weinheim, Germany, 1989.
- Mandelbrot, B.B. *Fractals: Form, Chance, and Dimension*; Freeman: San Francisco, 1977.
- Schaeffer, D.W. Fractal Models and the Structure of Materials Matter. *Res. Soc. Bull.* **1988**, 13, 22–27.
- Kaye, B.H. Multi-Fractal Description of a Rugged Fineparticle Profiles. *Part. Part. Syst. Charact.* **1984**, 1, 14–21.
- Coincidences, Clusters and Catastrophes*; Ch. 13, 483–534 in Ref. 25.
- Kaye, B.H.; Yousufzai, M.A.K. Calibrating and Monitoring Woven Wire Sieving Surfaces. *Powder Bulk Eng.*, **1992**, (Jan.), 29–34.
- Sieves Sieving and Other Sizing Methods. In *Draft of Standard Determination of Particle Size Distribution, Laser Diffraction Methods*; International Organization for Standardization (ISO): Berlin, Germany.
- Some of the Classical Sedimentation Equipment as well as other Sizing Equipment is Available from Gilson Company Inc., P.O. Box 200, Lewis Center, OH 43035.
- Provder, T., Ed. *Particle Size Analysis*, ACS Symposium Series 332; American Chemical Society: Washington, DC, 1987.
- CILAS U.S.A. available through Denver Autometrics, Inc., 6235 Lookout Road, Bolder, CO 80301. Company Headquarters in France, Osi 47, Rue de Javel, 75015 Paris.

36. Kaye B.H. Dangers of Curve Fitting in the Deduction of Size Distribution from Diffraction Data. Proceedings of the Powder and Bulk Solids Conference, Rosemont, IL, May 6–9, 1991.
37. Sympatec, Inc. Princeton Service Center, 34890 U.S. Route 1, Princeton, NJ 08540-5706.
38. See Trade Literature of Shimadzu Scientific Instruments Incorporated, 7102 Riverwood Drive, Columbia, MD 21046.
39. Holve, D.J. Using Ensemble Diffraction to Measure Particle Size Distribution. Powder Bulk Eng. **1991**, 5 June (6), 15–19. See also trade literature of INSITEC, 2110 Omega Road, Suite D, San Ramon, CA 94583.
40. See Trade Literature of Coulter Counter Electronics, 590 West 20th St., Hialeah, FL 33010.
41. Malvern Instruments Incorporated, 10 Southview Rd., Southborough, MA 01772.
42. MICROTRAC System Manufactured by Leeds and Northrop Instruments, 3000 Old Roosevelt Blvd., St. Petersburg, FL 33702.
43. Fritsch GmbH, Industriestrasse 8, D-55743, Idar-Oberstein, Germany.
44. Technical Literature Available from Galai Production Limited, Industrial Zone, 10500, Migdal, Haemek, Israel.
45. LASENTECH is available from Laser Sensor Technology Inc., P.O. Box 3912, Bellevue, WA 98009.
46. TSI Particle Instruments/Amherst, 7 Pomeroy Lane, Amherst, MA 01002–2905, and TSI Inc., 2500 Cleveland Avenue, N, St. Paul, MN 55113.
47. Dantech Corporation, 777 Corporate Drive, Mahwah, NJ 07430.
48. Weiner, B.B. *Modern Methods of Particle Sizing*; Barth, H., Ed.; Wiley-Interscience: New York, 1984; Ch. 3.
49. See Technical Literature of Brookhaven Instruments Corporation, Brookhaven Corporate Park, 750 Bluepoint Road, Holtsville, NY 11742.
50. Nicoli, D.F.; Wu, J.S.; Chang, Y.J.; McKenzie, D.C.; Hasapidis, K.; Automatic High Resolution Particle Size Analysis by Single Particle Optical Sensing. Am. Lab. **1992**, July 24 (8), 39.
51. Mazumder, M.K.; E-Spart Analyzer: Its Performance and Applications to Powder and Particle Technology Processes. KONA **1993**, 11, 105–118. KONA is produced by the Hosokawa Micron International Inc., which Markets the E-SPART Analyzer; Literature is available from Micron Powder Systems (a member of the Hosokawa Micron group), 10 Chatham Rd., Summit, NJ 07901. (The term SPART Analyzer stands for: Single Particle Aerodynamic Relaxation Time analyzer).
52. Information on the Coulter Counter is available from Coulter Electronics, Inc., 590 West 20th St., Hialeah, FL 33010.
53. An Optical Stream Counter is available from the Climec Corporation, 1320 Colton Avenue, Redlands, CA 92373.
54. Royco Instruments for Studying Aerosols and Fineparticles in Liquids are available from Royco Instruments, Inc., 141 Jefferson Drive, Menlo Park, CA 94025.
55. The Widely used Stream Counter for Fineparticles in Fluid is the HIAC Counter, HIAC Instruments Division, P.O. Box 3007, 4719 West Brooke St., Monte Claire, CA 91763.
56. Humann, W.L. Cyclone Separators, a Family Affair. Chem. Eng. **1991**, June 8(6), 118–123.
57. Giddings, J.; Field Flow Fractionation. Chem. Eng. News **1988**, Oct. 95 (10), 34. (For Information on Field Flow Fractionation Research Development and Industrial Applications, Contact Field Flow Fractionation Research Center, Department of Chemistry, University of Utah, Salt Lake City, UT 84112.).
58. See Discussion of Gas Adsorption Methods in Ref. 22.
59. Avnir, D., Ed. *Fractal Approach to Heterogeneous Chemistry*; John Wiley & Sons: London, 1989.
60. Takayasu, H., Ed. *Fractals in the Physical Sciences*; John Wiley & Sons: New York, 1990.
61. Neimark, A.V. Percolation Theory of Capillary Hysteresis Phenomena and its Application for Characterization of Porous Solids. *Characterization of Solids II*; Rodrigues-Reinoso, F., et al. Eds.; Elsevier Science Publishers: Amsterdam, 1991; 67–75.
62. Kaye, B.H. Permeability Techniques for Characterizing Fine Powders. Powder Technol. **1967**, 1, 11–22.
63. Kaye, B.H.; Legault, P.E. Real-Time Permeability for the Monitoring of Fineparticle Systems. Powder Technol. **1973**, 23, 179–186.
64. Orr, C. Application of Mercury Penetration in Material Analysis. Powder Technol. **1969–1970**, 3, 117–123.
65. Kaye, B.H. Applied Fractal Geometry and the Fineparticle Specialist I. Rugged Boundaries and Rough Surfaces. Part. Part. Syst. Charact. **1993**, 10(3), 99–110.
66. Kaye, B.H. Fractal Dimensions in Data Space: New Descriptors for Fineparticle Systems. Part. Part. Syst. Charact. **1993**, 10, 191–200.
67. Standard powders are available from Duke Scientific Corporation, 135D San Antonio Rd. Palo Alto, CA 94303.
68. Calibration Fineparticles are available from Dyno Particles A. S., PO. Box 160N-2001, Lillestrom, Norway.
69. Merkus, H.G.; Bischof, O.; Drescher, S.; Scarlett, B. Precision and Accuracy in Particle Sizing. Round-Robin Results from Sedimentation, Laser Diffraction and Electrical Sensing Zone Using BCR 67 and 69. In Proceedings 6th European Symposium on Particle Characterization, PARTEC, Nurnberg, March 21–23, 1995.
70. Yamamoto, H.; Matsuyama, T. Comparative Study of Particle Size Analyses Using Common Samples. KONA **1995**, 13, 57–66.
71. Kaye, B.H.; Alliet, D.; Switzer, L.; Turbitt-Daoust, C. The Effect of Shape on Intermethod Correlation of Techniques for Characterizing the Size Distribution of a Powder I. Correlating the Size Distribution Measured by Sieving, Image Analysis, and Diffractometer Method. Part. Part. Syst. Charact. **1997**, 14, 219–224.
72. Kaye, B.H.; Alliet, D.; Switzer, L.; Turbitt-Daoust, C. The Effect of Shape on Intermethod Correlation of Techniques for Characterizing the Size Distribution of a Powder II. Correlating the Size Distribution as Measured by Diffractometer Methods, TSI-Amherst Aerosol Spectrometer, and Coulter Counter. Part. Part. Syst. Charact. **1999**, 16, 266–272.

PARTITION COEFFICIENTS

Eric J. Lien

Shijun Ren

University of Southern California, Los Angeles, California, U.S.A.

INTRODUCTION

A thorough understanding of partition coefficients is important to all research scientists and product development staff in various branches of the pharmaceutical field. The principle and applications are involved in several different areas of current pharmaceutical interest. These include the techniques of extraction, preservation of oil–water systems, penetration through packaging materials, absorption and distribution of drugs in vivo, protein binding and hemodialysis, drug metabolism, enzyme inhibition, drug–receptor interactions, drug-delivery systems, and drug targeting. Because the partition coefficient is a measure of hydrophobic-bonding tendency, and all proteins (enzyme, membrane, plasma, and receptor proteins) contain 20 to 45% of amino acids with nonpolar groups (e.g., leucine, isoleucine, phenylalanine, tyrosine, tryptophan, etc.), continuing interest in using partition coefficients in correlating biological activity with molecular structure is expected. Other areas of interest in the application of partition coefficient fall beyond the scope of this article. These include analytical chemistry, toxicology, forensic medicine, ecology, and environmental protection.

HISTORICAL BACKGROUND

The first observation that the ratio of concentrations of a solute (e.g., I_2 or Br_2), when distributed between an organic solvent (e.g., CS_2 or ether) and water, remained constant even when the volume ratio of the immiscible solvents changed widely, was first reported by Berthelot and Jungfleisch in 1872, as illustrated by Eq. 1:

$$P = C_o/C_{aq} = \text{equilibrium constant } K \quad (1)$$

In 1921, Smith suggested that partition coefficient P can be converted from one solvent system to another.

Thirty years later, Collander presented the standard linear free-energy relationship, shown in Eq. 2, using water and different alkanols:

$$\log P_2 = a \log P_1 + b \quad (2)$$

where a is a coefficient, and b is a constant.

At the turn of the century, Meyer and Overton discovered that most organic compounds (except nutrients) penetrate tissue cells as a lipid barrier and that their narcotic action parallels the oil–water partition coefficients of the compounds. In the early 1950s, Collander demonstrated that the penetration rate of plant cell membranes by various organic compounds was related to their oil–water partition coefficients. Cohen and Edsal studied the ratios of alcohol solubility to water solubility to define the relative lipophilic character of amino acids. The limited additivity of the partition coefficients of organic compounds was observed by Collander, Cohen, and Edsal in their studies. In the early 1960s, Salame and Pinsky derived the permachor method, given in Eq. 3, for calculation of the P factor for the prediction of chemicals permeation through a plastic membrane:

$$\log P_f = 16.55 - 3700/T - 0.22\pi \quad (3)$$

where π is the permachor constant, and T is the absolute temperature. A general equation was presented by these investigators, as in Eq. 4:

$$\log P_f = K - R\pi \quad (4)$$

where K is a temperature correction constant, and R is a polymer (e.g., plastic) correction term. Interestingly, at about the same time Hansch, Fujita, and co-workers made the most significant contributions toward the understanding and application of partition coefficients ($\log P$ and π) and greatly extended the linear free-energy-related (LFER) approach from organic chemistry to medicinal chemistry and biology. This renewed interest in the application of partition coefficients has stimulated many excellent reviews, books, and monographs.

THEORY AND EXPERIMENTAL METHODS OF MEASUREMENTS

Because the partition coefficient is measured when an equilibrium is reached, it is characterized by the equality of the chemical potentials μ_o and μ_{aq} of the solute in the two phases (organic and aqueous), as shown by Eqs. 5, 6, and 7:

$$\mu_o = \mu_o^0 + RT \ln C_o \quad (5)$$

$$\mu_{aq} = \mu_{aq}^0 + RT \ln C_{aq} \quad (6)$$

$$\text{If } \mu_o = \mu_{aq},$$

then:

$$\begin{aligned} P = C_o / C_{aq} &= e^{-(\mu_{aq}^0 - \mu_o^0) / RT} \\ &= e^{-\Delta\mu^0 / RT} \end{aligned} \quad (7)$$

where C_o and C_{aq} are the equilibrium concentrations of the solute in the organic and the aqueous phases, respectively; μ_o^0 and μ_{aq}^0 are the chemical potentials (in the organic and aqueous phases, respectively) at infinite dilution; and R is the gas constant, T the temperature, and P the partition coefficient. P is a constant for any compound in a given solvent system at a given temperature; this relationship is known as the Nernst law.

As with any equilibrium constant K , P is related linearly to the standard free-energy change when it is converted to the logarithmic scale, as in Eqs. 8, 9, and 10:

$$\Delta G^0 = -RT \ln K \quad (8)$$

For any equilibrium:

$$\Delta G^0 = -2.303 RT \log K \quad (9)$$

For partition processes:

$$\Delta G^0 = -2.303 RT \log P \quad (10)$$

Because of this linear free-energy relationship (LFER), $\log P$ is commonly used in most correlation studies instead of P .

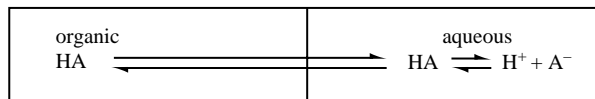
In the past 35 years or so, 1-octanol-water has been the most commonly used solvent system. It has been shown that for correlation with biological activity, organic solvents (such as 1-octanol) capable of forming hydrogen bonds usually give better correlation than those not capable (e.g., CCl_4 , cyclohexane, and other hydrocarbons). Extensive compilations on 1-octanol-water partition coefficient are available.

On the other hand, if one is interested in separating out thermodynamic properties such as enthalpy change (ΔH) and entropy change (ΔS), a solvent with minimum mutual

solubility with water (such as cyclohexane or heptane) is preferable.

Apparent Versus True Partition Coefficient ($\log P'$ Versus $\log P$)

If a solute is ionizable (either acidic or basic), two different species can exist in the aqueous phase, and therefore the apparent partition coefficient (P') or the true partition coefficients (P) can be measured, as shown below and in Eqs. 11, 12, 13, 14, 15:



where

$$P = [\text{HA}]_o / [\text{HA}]_{aq} \quad (11)$$

$$P' = [\text{HA}]_o / ([\text{HA}]_{aq} + [\text{A}^-]_{aq}) \quad (12)$$

and

$$P = P' / (1 - \alpha) \quad (13)$$

where α is the degree of ionization: for acids:

$$\alpha = 1 / [1 + \text{antilog}(\text{pK}_a - \text{pH})] \quad (14)$$

for bases:

$$\alpha = 1 / [1 + \text{antilog}(\text{pH} - \text{pK}_a)] \quad (15)$$

Because the degree of ionization is a function of the pH of the aqueous phase and the pK_a of the solute, the apparent partition coefficient P' fluctuates as the pH of the aqueous phase (usually a buffer solution) is changed, whereas the true (or corrected) partition coefficient (P) should remain constant. However, in reality the different buffer species may not only affect P' but also P because of different degrees of ion-pair formation and the different polar nature of the counterions used. Among the different buffer species, 1-octanol-phosphate buffer appears to give the most consistent results compared with 1-octanol-water. In some publications, the apparent partition coefficient P' is also described as the distribution coefficient D .

Because the separation of immiscible phases takes place only in the presence of gravity, it would not be possible to measure partition coefficients in outer space where the gravity is zero.

Shake-Flask Method

The shake-flask method is used most commonly in the measurement of partition coefficients. It is also the

standard procedure to validate other methods. A solute is simply shaken with two immiscible solvents (organic and aqueous), followed by analyzing the solute concentration in one or both phases. To avoid any volume changes in both phases, one phase is saturated with the other before the partitioning process. It is important to ensure that equilibrium is reached before the analysis. If a solute does not cause emulsification, vigorous shaking can reduce the time required to reach equilibrium, usually in a few minutes. However, if a solute with both polar and nonpolar groups present (such as saponin glycosides or surfactant-type compounds), a gentle and slow shaking procedure or even special devices such as a Doluisio and Swintosky Y-tube or a Schulman-type cell should be used.

Some true (undissociated, corrected) partition coefficients of representative drug molecules ranging from -2.26 to $+15.69$ are shown in Table 1.

For the quantitative analysis of the solute distributed in one or both phases, the most commonly used analytical methods include UV-visible spectrophotometric analysis for compounds with chromophore groups and gas-liquid chromatography (GLC). Colorimetric methods have also been used for specific compounds.

With the proper choice of solvent volume and sensitive analytical methods, $\log P$ values ranging from -5 to $+5$ can be measured. The temperature dependence of many partitioning systems is on the order of 0.01 log unit per degree in the 25°C range. Adequate temperature control is needed for high accuracy. This is more critical for volatile solvents like ether, chloroform, low-boiling hydrocarbons, and alcohols lower than 1-octanol.

Chromatographic Methods

In recent years, the availability of reproducible systems and precision instruments in high-performance liquid chromatography (HPLC) has prompted the application of chromatography in the rapid measurement of partition coefficients. In general, a linear relationship between $\log P$ and $\log K'$ from a set of compounds is required for the interpolation or extrapolation of $\log P$ values of additional compounds of congeneric nature, as shown in Eq. 16:

$$\log P = a \log K' + b \quad (16)$$

where the capacity factor K' is determined from the net retention time t_R relative to the nonadsorbed time t_0 , as defined by Eq. 17:

$$K' = (t_R - t_0)/t_0 \quad (17)$$

This is similar to the linear relationship R_m measured by thin-layer chromatography (TLC), given in Eq. 18:

Table 1 Selected true (corrected) partition coefficients of representative drug molecules measured in 1-octanol-water or 1-octanol-phosphate buffer

Drug	log P
L-Tyrosine	-2.26
Hydroxyurea	-1.80
Citric acid	-1.72
Ascorbic acid (Vit. C)	-1.64
Phenol red	-1.45
Streptozotocin	-1.45
Sulfanilamide	-0.73
Theobromine	-0.72
Chlortetracycline	-0.62
Ethanol	-0.31
Ellagic acid	0.27 ^a
Cimetidine	0.40
Atenolol	0.43
Metiamide	0.50
Procainamide	0.51
Morphine	0.76
Ephedrine	0.87
Aminopyrine	1.00
Colchicine	1.03
Chloramphenicol	1.14
Atropine	1.24
Digoxin	1.26
Nalidic acid	1.41
Phenobarbital	1.47
Daidzein	1.58 ^a
Hydrocortisone	1.61
Benzoic acid	1.72
Salicylic acid	1.73
Zileuton	1.81 ^a
Benzylpenicillin	1.83
Chloroform	1.97
Dexamethasone	1.99
Podophyllotoxin	2.01
Metoprolol	2.04
Naloxone	2.09
Mathapyrilene	2.81
Phenformine	2.94
Labetalol	3.18
Benadryl	3.20
Propranolol	3.29
Clobetasol-17-butyrate	3.63
Progesterone	3.87
Estradiol	4.01
Tamoxifen	4.03
Propoxyphene	4.18
Mefepristone (Ru486)	5.48 ^b
α-Tocopherol (vitamin E)	12.28 ^a
β-Carotene	15.69 ^a

^aCalculated value using CQSAR program of BioByte.

^bCalculated from the $\log P$ of progesterone and the π values of the substituents: $\log P_{\text{Ru 486}} = \log P_{\text{progesterone}} - \pi_{\text{CH}_3} + \pi_{\text{double bond}} + \pi_{(\text{CH}_3)_2\text{N}^+} - \pi_{\text{COCH}_3} + \pi_{\text{O}=\text{C}} + \pi_{\text{CH}_3} + \pi_{\text{OH}} = 3.87 - 0.5 - 0.2 + (-0.18) + 1.96 - (-0.71) + 0.48 + 0.5 + (-1.16) = 5.48$.

$$R_m = \log K' = \log[(1 - R_f)/R_f] = \log(1/R_f - 1) \quad (18)$$

Because of the limits inherent in the mathematical formula, R_m ranges only from -1.996 to $+1.996$ for all possible compounds. This makes it less sensitive than the direct measurement of $\log P$. Nevertheless, for many compounds, it is much easier and more economical to measure $\log K'$ or R_m than $\log P$.

Countercurrent and Filter Probe Methods

A countercurrent-based device, known as AKUFVE, is useful when a large amount of information resulting from varying T or pH is required on one or only a few compounds. This method has the disadvantages of difficulties in cleaning and operation as well as in the need of large quantities of materials. Another method is Tomlinson's filter-probe method, which samples the phase with the larger volume (generally the aqueous phase) and pumps it through a UV detector to monitor the state of equilibrium. A heavy metal probe is attached to the circulating stainless steel tubing with a special filter that prevents entrapment of the unwanted phase. This method is related to the shake-flask method at high phase-volume ratio. It may be useful for unstable compounds and, as a closed system, may be used over a wide temperature range. Kaufman et al. have reported a microelectrometric titration method for the direct measurement of the pKa and partition and distribution coefficients of narcotics and narcotic antagonists and their pH and temperature dependence. Based on this principle, an automated instrument is now available for the simultaneous measurement of pKa and $\log P$ (pION Inc., Cambridge, MA).

Calculation of Partition Coefficients

Hansch-Fujita π constant

In 1964, Fujita et al. proposed that $\log P$ was an additive-constitutive property and can be calculated by taking the sum of the $\log P$ of the parent molecule and the π of the substituent, as in Eqs. 19, 20, 21:

$$\log P = \Sigma \pi \quad (19)$$

$$\pi_{\text{substituent}} = \log P_{\text{substituted molecule}} - \log P_{\text{parent molecule}} \quad (20)$$

For example:

$$\begin{aligned} \pi_{\text{CH}_3} &= \log P_{\text{CH}_3\text{C}_6\text{H}_5} - \log P_{\text{C}_6\text{H}_6} = 2.69 - 2.13 \\ &= 0.56 \end{aligned} \quad (21)$$

$$\log P_{\text{Cl}(\text{C}_6\text{H}_5)\text{CH}_3} = \log P_{\text{C}_6\text{H}_6} + \pi_{\text{Cl}} + \pi_{\text{CH}_3} = 2.13 + 0.71 + 0.56 = 3.40 \text{ (calculated), measured value} = 3.33.$$

By definition, the π of hydrogen is zero. A scale of the π values of various groups from very hydrophilic to very hydrophobic is shown in Table 2. Extensive compilations of the π constants of various functional groups are available in the literature.

Rekker's fragmental constant f

The fragmental (reductionist) approach of calculating $\log P$ was initiated by Rekker and coworkers. Based on a collection of measured $\log P$ values, they applied statistical analysis to determine the average contribution of simple fragments such as C, CH, CH₂, CH₃, OH, NH₂, CONH₂, OCH₃COOH, etc (see Table 2). It was found that it was necessary to introduce corrections if two polar groups were separated by only one or two aliphatic carbons. Their postulation is given in Eq. 22:

$$\log P = \Sigma a_m f_m + \Sigma b_m F_m \quad (22)$$

where a = the number of occurrences of fragment f of type n , and b = the number of occurrences of correction factor F of type m .

Leo-Hansch f constant

Hansch and Leo used a constructionist (synthetic) approach by starting with a few carefully measured values of $\log P$ of simple structures like H₂ and CH₄ and derived a separate set of fragment constants. The two columns in Table 2 show the slightly different values obtained by the two groups. Different π and f values should be used for an aliphatic system.

It is worth noting that although π_{H} is zero, f_{H} according to the Rekker's scale is 0.18 and according to the Hansch–Leo scale is 0.23. The calculation of $\log P$ using the fragment method has been computerized by the Hansch–Leo group. In this CLOGP program, all known correction factors have been incorporated. The structure of any compound can be entered by a linear notation called SMILES, and by going through a substructure search algorithm GENIE, the $\log P$ value can be calculated according to Eq. 22. The SMILES program includes all isomerisms.

Other methods

Other published methods of estimating partition coefficients include the use of molecular surface and volume in predicting solubilities and free energies of desolvation and the application of principal-component analysis based on partition coefficient data. Suzuki and Kudo's CHEMICAL (Combined Handling of Estimation Methods Intended for Completed Automated $\log P$ Calculation) as well as a

Table 2 The π and f constants for some functional groups

Function x	π_x Aromatic system	π_x Aliphatic system	f_x Aromatic system	
			Rekker	Leo–Hansch
H—	0	0	0.18	0.23
F—	0.13	−0.17	0.42	0.37
Cl—	0.76	0.39	0.93	0.94
Br—	0.94	0.60	1.18	1.09
I—	1.15	1.00	1.47	1.35
CH ₃ —	0.50	0.50	0.70	0.89
HC≡C—		0.48		
CH ₂ =CH—		0.70		
C ₂ H ₅ —	1.00	1.00		
H ₂ =C(CH ₃)—		1.00		
CH ₂ =CHCH ₂ —		1.20		
<i>n</i> -C ₃ H ₇ —	1.50	1.50		
<i>i</i> -C ₃ H ₇ —	1.30	1.30		
<i>n</i> -C ₄ H ₉ —	2.00	2.00		
<i>s</i> -C ₄ H ₉ —	1.80	1.80		
<i>t</i> -C ₄ H ₉ —	1.68	1.68		
cyclo-C ₃ H ₅ —		1.21		
cyclo-C ₅ H ₉ —	2.14	2.14		
cyclo-C ₆ H ₁₁ —	2.51	2.51		
Adamantyl	3.30			
C ₆ H ₅ —	2.13	2.13	1.89	1.90
—(CH ₂) ₃	1.04			
—(CH ₂) ₄	1.39			
—(CH) ₄ —	1.24			
—CF ₃	1.07		1.25	1.11
—CH ₂ OH	−1.03	−0.66		
—CH ₂ COOH	−0.72	−0.76		
—COOH	−0.32	−1.26	0.00	−0.03
—COO [−]	−4.36			
—CONH ₂	−1.49	−1.71	−1.13	−1.26
—COOCH ₃	−0.01	−0.27		
—COCH ₃	−0.55	−0.71		
—CN	−0.57	−0.84	−0.23	−0.34
—OH	−0.67	−1.16	−0.36	−0.44
—OCH ₃	−0.02	−0.47		
—OCH ₂ COOH	−0.86			
—OCOCH ₃	−0.64	−0.91		
—CH=NNHCONH ₂	−0.85			
—CH=NNHCSNH ₂	−0.27			
—O- β -Glucose	−2.84			
—NH ₂	−1.23	−1.19	−0.90	−1.00
—N(CH ₃) ₂	−0.18	−0.32		
—NO	−0.12			
—NO ₂	−0.28	−0.82	−0.09	−0.03
—NHCOCH ₃	−0.97			
—NHCOC ₆ H ₅	0.72			
—N=NC ₆ H ₅	1.69			
—NHCONH ₂	−1.01			
—N(CH ₃) ₃ ⁺	−5.96			
—N ₃	0.46			

(Continued)

Table 2 The π and f constants for some functional groups (*Continued*)

Function x	π_x Aromatic system	π_x Aliphatic system	f_x Aromatic system	
			Rekker	Leo–Hansch
—SH	0.39	0.28	0.62	0.62
—SCH ₃	0.62			
—SCF ₃	1.44			
—SCCl ₃	1.65			
—SO ₂				−2.17
—SO ₂ F				0.30
—SO ₂ CH ₃	−1.26			
—SO ₂ CF ₃	0.55			
—SF ₅	1.55			
—SO ₂ NH ₂	−1.82			−1.59

multidescriptor highly nonlinear regression model proposed by Bodor et al. were applied, in which a 10-parameter (15-term) equation was used to correlate with the log P values of 118 compounds of varying complexity. (All the descriptors were derived from AM1 calculation and were related to the surface area, dipole moment value, and charge densities of the molecule.) Some of the terms were raised to 10^4 and 10^2 powers to give the best fit. It is difficult to explain the physical meaning of the highly complex polynomial equation.

Several experts in the field suggest that the ultimate goal of flawless calculation of log P has not yet been fully realized, especially when dealing with a highly complex structure with de novo functional groups.

Physical Factors Contributing to Log P or π : Bulk and van der Waals Forces, Dipolar Interactions, and Hydrogen Bonding

Over the past 25 years, considerable efforts have been devoted to delineate the fundamental nature of partition coefficients (log P or π). As a result, many correlations between log P (or π) with other structural descriptors or physicochemical parameters have been reported by various investigators. For example, Moriguchi et al. dissected log P into two intrinsic components, namely, molecular volume and polar effect, and showed that the partition coefficient of a nonpolar molecule is a linear function of the volume. For polar molecules, a hydrophilic group effect has to be added as a correction term in the evaluation of log P . Kamlet et al. correlated log P with the solvatochromic parameters π^* and β , which were derived to measure dipolar and hydrogen bond acceptor strengths of pure bulk solvents as well as the corresponding properties of solutes.

Franke et al. examined the dependence of hydrophobicity on solvent and structure and showed that log P values depend on solute bulk and polar and hydrogen-bonding effects. Ou et al. examined the quantitative relationship of log P with molecular weight (log MW), dipole moment (μ), and hydrogen-bond capability (HB_2) of various compounds. For 222 of 282 compounds, log P values were correlated with these three parameters with a correlation coefficient of 0.938 and standard deviation of 0.492, as shown in Eq. (23):

$$\log P = 5.84 \log MW - 0.36 \mu - 0.77 HB_2 - 8.86 \quad (23)$$

$n = 222$, $r = 0.938$, $s = 0.492$ where HB_2 is the sum total of energy decrement in a hydrogen-bond group.

In a similar manner, Yang et al. reported the general Eq. (24) to be applicable to a wide range of nonpolar and polar substituents (with only a few notable exceptions):

$$\pi = +a MW (\text{or } \nu W) - b HB - c \mu + d \quad (24)$$

where νW is the van der Waals volume, and HB can be the number of atoms in a group capable of forming H bonds (HB_1), or $HB_1 \times$ energy (HB_2).

This general model has been extended to the log P of disubstituted aromatic compounds and the solubilities of tetracycline derivatives.

Applications

Extraction

In both organic and analytical chemistry laboratories, it is a common procedure to extract a compound from one solvent to another. It is also a common knowledge that it is more efficient to use small volumes and multiple

extractions. This is shown by Eqs. 1 (given previously) and 25, 26, 27, assuming the two solvents are completely immiscible (e.g., $\text{H}_2\text{O}-\text{CCl}_4$).

$$P = (W_o/V_o)/[(W - W_o)V_{aq}] \quad (25)$$

or

$$\text{or } W_o = WPV_o/(PV_o + V_{aq}) \quad (26)$$

after n extractions:

$$W_n = W[PV_o/(PV_o + V_{aq})]^n \quad (27)$$

where W is the water phase, and V is the volume.

If two solvents are partially miscible (e.g., ether- H_2O), the equation provides only approximate values, which may still be useful for practical purposes.

Preservation of oil–water systems

Many pharmaceutical preparations containing oil–water systems (creams, ointments, or suspensions) are subject to microbial contamination. Bacteria in these heterogeneous systems are usually grown in the aqueous phase and at the oil–water interface. To preserve the shelf-life of these preparations, benzoic acid or other organic acids are added as preservatives. Because the microbial cell membrane is lipophilic in nature, the bacteriostatic actions of the acidic preservative are attributable almost entirely to the undissociated acid and not to the ionized form. A good understanding of the partition coefficient and the degree of ionization allows accurate calculation of the free un-ionized acid in the aqueous phase, which provides the bacteriostatic concentration.

Quantitative structure-activity relationships (QSAR)

Since the early work of Hansch et al., numerous examples of the quantitative correlation of biological activity with chemical structure have been reported. The success of QSAR relies heavily on the use of partition coefficients ($\log P$ or π) in extending the linear free-energy relationship (LFER) from homogenous organic chemical systems (i.e., the Hammett–Taft type approach) to compartmentalized heterogeneous biological systems.

Further analysis of the physical nature of the partition coefficient reveals that it is a composite property depending on size, shape, dipole moment, and hydrogen-bonding ability. Although many researchers have attempted to replace $\log P$ with other simple parameters, only limited success has been achieved for some, but not all, molecules. It appears that for entirely new complex molecules, it still would be necessary to measure the partition coefficient, preferably validated by the conventional shake-flask method.

Partition coefficients of peptides and ampholytes

Since the publication of the first edition of the *Encyclopedia of Pharmaceutical Technology*, significant progress has been made in the experimental measurements and calculation of the $\log P'$ values of small peptides ranging from di- to pentapeptides with and without ionizable side chains. Additional work will be needed to calculate the $\log P'$ values of large peptides and proteins. Testa's group has used a potential-pH representation to show ionic partition diagrams and analyzed the lipophilicity profiles of amphoteric compounds. They have also compiled a review on various computational approaches to lipophilicity, with 223 references cited.

Because an increasing number of new compounds are being synthesized every day, it is not feasible to have the partition coefficient of every new structure experimentally determined. According to the estimate of Leo and Hansch, the measured $\log P$ in octanol/water increases at the rate of approximately 1200/year. When a close congener with measured $\log P$ is available, it is easier to use the $\Sigma\pi$ method in calculating the $\log P$ of a structurally similar new derivative. On the other hand, for a large number of different structures, the CLOGP method based on the f constant of Hansch and Leo is suitable for obtaining calculated $\log P$ values fairly efficiently. An impressive correlation ($r^2 = 0.98$, $s = 0.21$) between the measured $\log P$ and calculated $\log P$ values of 9000 compounds has been obtained by this group. Regardless of the method used, it is necessary to measure the $\log P$ of a few model compounds for comparison. Many times, careful examination of the large deviation between the experimentally measured and the theoretically calculated values can uncover intra- or intermolecular interactions and thus lead to a better understanding of the phenomenon of partitioning of a solute between two immiscible phases saturated with each other.

The initial development of the fragmental constant f by Rekker has stimulated that of the systematic CLOGP method. Rekker has further compared the model of partition process with the passage of a ball through a “brick wall” to account for a frequently observed “magic number.” The model does not take into account the fact that neither 1-octanol nor the aqueous phase consists of homogeneous and ordered “bricks” as depicted in his diagram. At the present time, there is no rigorous theoretical method available for the calculation of $\log P$ of a complex de novo structure. Therefore, semiempirical methods and experimental methods will continue to be used.

For QSAR analysis, distribution coefficients (D) of ionizable compounds have been used by some investigators. From the mathematical formula of D , it can be

demonstrated that it is the same as the apparent partition coefficient (P') and can be easily converted to a true (or corrected) partition coefficient P , as shown in Eqs. 28, 29, 30:

$$D = [HA]_o / ([HA]_{aq} + [A^-]_{aq}) \quad (28)$$

$$P' = [HA]_o / ([HA]_{aq} + [A^-]_{aq}) \quad (29)$$

$$D / (1 - \alpha) = [HA]_o / [HA]_{aq} = P \quad (30)$$

If limited data points are available in QSAR analysis, $\log P'$ ($\log D$) can be used to account for different degrees of ionization as well as for different lipophilicities. If, on the other hand, a sufficient number of data points are available (more than 5 data points for each parameter being examined), it will be advantageous to use the log of the [undissociated] vs. [dissociated] ratio $\log U/D$ ($= pK_a - pH$ for acids, and $pH - pK_a$ for bases) as an independent variable as well as $\log P$. This will separate the effect of ionization from that of relative lipophilicity.

Avdeef has recently reported the refinement of partition coefficients and ionization constants of multiprotic substances based on a generalized, weighted, nonlinear least-squares procedure and pH titration curve. This method allows for the determination of pK_a and $\log P$ values of multiprotic substances with fairly close ionization constants.

A Prolog P program has been developed for the estimation of the distribution coefficients ($\log D$) of ionizable compounds, based on LFER-derived microscopic dissociation constants. Several estimation methods for calculating octanol-water partition coefficients have been reported (e.g., $\log KOW$; $X \log P$) using atom/fragment additive principles and correction factors. Clarke et al. have reported that for ionizable compounds, by using two different octanol volumes in a dual-phase potentiometric titration, both pK_a and partition coefficient values can be obtained by curve fitting.

BIBLIOGRAPHY

- Akamatsu, M.; Fujita, T. Quantitative Analyses of Hydrophobicity of Di- to Pentapeptides Having Unionizable Side Chains with Substituent and Structural Parameters. *J. Pharm. Sci.* **1992**, *81*, 164–174.
- Akamatsu, M.; Katayama, T.; Kishimoto, D.; Kurokawa, Y.; Shibata, H.; Ueno, T.; Fujita, T. Quantitative Analysis of the Structure-Hydrophobicity Relationship for N-Acetyl Di- and Tripeptide Amides. *J. Pharm. Sci.* **1994**, *83*, 1026–1033.
- Carrupt, P.A.; Testa, B.; Gaillard, P. Computational Approaches to Lipophilicity: Method and Applications. *Reviews in Computational Chemistry*; Lipkowitz, K.B., Boyd, D.B., Ed.; Wiley & Sons, Inc.: New York, 1997; 11, 241–315.
- Clarke, F.H.; Cahoon, N.M. Partition Coefficients by Curve Fitting: The Use of Two Different Octanol Volumes in a Dual-Phase Potentiometric Titration. *J. Pharm. Sci.* **1996**, *85*, 178–183.
- Csizmadia, F.; Tsantili-Kakoulidou, A.; Pander, I.; Darvas, F. Prediction of Distribution Coefficient from Structure. I. Estimation Method. *J. Pharm. Sci.* **1997**, *86*, 865–871.
- Dearden, J.C.; Bresnen, G.M. The Measurement of Partition Coefficients. *Quant. Struct.-Act. Relat.* **1988**, *7*, 133–144.
- Gao, H.; Wang, F.Z.; Lien, E.J. Hydrophobic Contribution Constants of Amino Acid Residues to the Hydrophobicities of Oligopeptides. *Pharm. Res.* **1995**, *12*, 1279–1283.
- Gao, H.; Lien, E.J.; Wang, F.Z. Hydrophobicity of Oligopeptides Having Un-Ionizable Side Chains. *J. Drug Target* **1993**, *1*, 59–66.
- Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, DC, 1995; 1–348.
- Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995; 1–557.
- Lien, E.J.; Ren, S.J. QSAR and Molecular Modeling of Bioactive Phytochemicals. *Phytochemicals as Bioactive Agents*; Bidlack, W.R., Omaye, S.T., Meskin, M.S., Topham, D.K.W., Ed.; Chapter 2, Technomic Publishing Co.: Lancaster, PA, 2000; 21–41.
- Lien, E.J. Partition Coefficients. *Encyclopedia of Pharmaceutical Technology*, 1st Ed. Swarbrick, J., Boylan, J.C., Ed.; Marcel Dekker, Inc.: New York, 1994; 11, 293–307.
- Meylan, W.M.; Howard, P.H. Atom/Fragment Contribution Method for Estimating Octanol-Water Partition Coefficients. *J. Pharm. Sci.* **1995**, *84*, 83–92.
- Pagliara, A.; Carrupt, P.A.; Caron, G.; Gaillard, P.; Testa, B. Lipophilicity Profiles of Ampholytes. *Chem. Rev.* **1997**, *97*, 3385–3400.
- Palekar, D.; Shiue, M.; Lien, E.J. Correlation of Physicochemical Parameters to the Hydrophobic Contribution Constants of Amino Acid Residues in Small Peptides. *Pharm. Res.* **1996**, *13*, 1191–1195.
- Reymond, F.; Steyaert, G.; Carrupt, P.A.; Testa, B.; Girault, H. Ionic Partition Diagrams: A Potential-pH Representation. *J. Am. Chem. Soc.* **1996**, *118*, 11951–11957.
- Tsantili-Kakoulidou, A.; Pander, I.; Csizmadia, F.; Darvas, F. Prediction of Distribution Coefficient from Structure. II. Validation of Prolog D, An Expert System. *J. Pharm. Sci.* **1997**, *86*, 1173–1179.
- Wang, R.; Fu, Y.; Lai, L.A. A New Atom-Additive Method for Calculating Partition Coefficients. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 615–621.
- Waterbeemd, H.V.D.; Mannhold, R. Programs and Methods for Calculation of Log P -Values. *Quant. Struct.-Act. Relat.* **1996**, *15*, 410–412.

PATENTS—INTERNATIONAL PERSPECTIVE

Stuart R. Suter

Suter Associates, Glenside, Pennsylvania, U.S.A.

Peter J. Giddings

SmithKline Beecham plc, Brentford, United Kingdom

INTRODUCTION

A strong patent system is important to the research-based pharmaceutical industry. The success of the pharmaceutical industry has been based on the discovery of new products that treat human disease states in new and unique ways. The patent system has provided protection for these innovative products for a period of time, allowing the industry to use the revenues gained to search for the next generation of products that will improve the health of society. It is important that persons in R&D of the industry understand the basic principles of the patent system.

WHAT IS A PATENT?

A patent is a grant of exclusive rights from a government to inventors for their invention in exchange for the inventors disclosing their inventions to society. This quid pro quo is valuable to society because the disclosure stimulates additional innovation and development. The alternative for inventors is to keep their inventions secret and thus deprive society of the opportunity for further advancement.

Patent rights are limited in time and are also limited to the sole right of excluding others from making, using, and selling the invention. Because governments grant these rights, they are effective only in the area controlled by that government. Thus, a United States patent provides protection only in the United States and its territories such as Puerto Rico. If inventors desire protection in Japan, Canada, or any European country, they must apply for a patent in each of those countries as well.

It is important to understand that this right granted by the patent is solely a right to exclude others. It does not carry with it the right to practice the invention by the inventors themselves. This right to use by the inventor may be limited by the existence of other patents that would be needed to practice the invention or by laws or regulations having nothing to do with patents. An example of the latter

is the health registration regulations in the countries where pharmaceutical products are being sold. An inventor not being free to use his invention because of other patents is also a common situation. For example, party X may have a patent claiming a broad genus of compounds useful to treat hypertension. Our inventor discovers a specific compound not disclosed specifically in the patent of party X, but through selection of the appropriate substituents in the defined genus this compound can be found. Its hypertensive properties are far superior to those of the compounds specifically disclosed in party X's patent, thus this inventor can obtain a patent for the compound but must obtain permission from party X before marketing the compound because of the patent rights of party X.

Patents belong to class of property that is referred to as "intellectual property." Intellectual property also includes trade secrets, trademarks, registered designs, and copyrights.

A trade secret is any information unknown to the public but gives the owner an economic or competitive advantage. Trade secrets have value only as long as they remain secret, which is difficult in this modern age. In the pharmaceutical industry, it is rare for any invention to be maintained as a trade secret because in the interests of public safety, the chemical composition of pharmaceutical products is always made public.

A trademark is a word, name, or symbol that identifies the source of the goods to which it applies. Two or more companies can sell the same product, but the public can identify each company's goods by the trademark. Pharmaceutical products available from multisources carry the same generic name and also the unique trademark of the specific seller; an example is the Tagamet brand of cimetidine.

A copyright protects the creations of artists and authors. In the pharmaceutical industry, copyrights are used to protect advertisements, product literature, and other copy used for product promotion.

A registered design right protects the aesthetic appearance of an article, i.e., features of a product's appearance rather than its technical features. In the

pharmaceutical industry, a registered design right might be obtained, for example, in respect to a particular tablet shape or to the appearance of the product's packaging.

The patent system traces its history to early Venice, where patents were granted at least as early as 1460. The Anglo-American systems have their origins in Great Britain, where Parliament enacted the Statute of Monopolies in 1624. This law provided the basis for the British patent system for many years. The American colonies, before achieving independence, had no power to grant patents for inventions under the British system. The basic U.S. patent system is founded on the U.S. Constitution. Article I, Section 8, gives Congress the power "to promote the progress of science and useful arts by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." The first Patent and Copyright Act in the United States was enacted on April 10, 1790.

PATENTS AND PHARMACEUTICALS

The patent system as applied to pharmaceuticals has been one of great controversy over the years. Many countries, especially those with few technology-based industries, have held that providing protection for drug products is against society's interests. Some countries include food in this category as well. Based on this theory, patents for pharmaceutical products have been limited or not allowed at all. The premise is usually that patents for chemical compounds can be claimed only for the process by which they are prepared. Thus, if someone develops a different process to make the compound, he or she can avoid the patent of the innovator.

Historically, as a country becomes more industrialized, it strengthens its patent system by providing protection for compounds per se, often referred to as "product protection." Examples are Germany and Japan, which had process protection for pharmaceutical products until 1968 and 1978, respectively, when they amended their laws to provide product protection. China had no patent law until June 1, 1985, when it introduced process-only protection for pharmaceuticals. Effective January 1, 1993, China amended its patent laws to allow product protection.

Since 1986, discussions have been ongoing at the governmental level to try to improve the standard of intellectual property laws around the world. These discussions, which have primarily been part of the General Agreement on Tariffs and Trade (GATT) negotiations among countries, culminated in 1994 with the signing of an agreement on Trade Related Aspects of Intellectual

Property Rights (TRIPS). This agreement established comprehensive standards for the protection and enforcement of intellectual property rights and became effective in January 1995. The agreement is complex and covers a wide range of patent-related issues including, for example, the provision of patent protection for compounds per se and the provision of a patent term of 20 years from filing. As a general rule, the date of application of the agreement for countries classified as "developed" was January 1, 1996, and for "developing" countries January 1, 2000. The "least-developed" countries have the option to defer application of the agreement until January 1, 2006.

In summary, patent protection for pharmaceutical products has improved significantly in recent years and will continue to improve.

New compounds are not the only inventions important to the pharmaceutical industry. Many products have been developed from sources in nature, such as extracts from plants or compounds isolated from fermentation broths from various microorganisms. Protecting these inventions has been difficult at times in various countries because the materials were present in nature and thus were viewed as natural products, not new or novel. These difficulties can be overcome in certain cases by claiming the compounds in their pure form.

Patenting of "living matter" changed dramatically with the arrival of biotechnology methods and procedures. In 1980, the U.S. Supreme Court decided the famous *Diamond vs. Chakrabarty* case. In this case, a new strain of bacteria, produced by a reproducible artificial procedure, had the ability to digest oil and thus was useful in dispersing oil slicks. It was important to have patent protection on the bacteria per se because they were used themselves and not in a process to produce another product. Certain groups and individuals in the public sector expressed great concern that this decision would lead to patenting other higher forms of life. In 1988, U.S. Patent No. 4,736,866 was issued, claiming the "Harvard Mouse." This transgenic mouse has the special property of having an oncogene sequence in the germ and somatic cells that makes the mouse useful in testing for carcinogenic materials or for compounds that confer protection against neoplasms. Since then, many patents for higher life forms (but not humans) have been issued.

In Europe, the issues surrounding the patentability of biotechnological inventions are the subject of vigorous ongoing legal and political debate. In July 1998, the European Parliament approved Directive 98/44/EC aimed at harmonizing laws across Europe with respect to the patenting of biotechnological inventions. Member states had until July 2000 to bring their laws into conformity

with the directive. However, the Dutch government has issued a challenge to the legality of the Directive at the European Court of Justice, and the decision of the court is awaited.

With respect to the specific case of the Harvard Mouse, an equivalent European patent has been allowed, but it is currently under opposition by a number of parties at the European Patent Office (EPO). Thus, the status of the patent is unclear at this time.

INTERNATIONAL TREATIES AND SYSTEMS

As noted above, patent protection is obtained on a country-by-country basis, and therefore patent applications must be filed in each country where protection is desired. International treaties and regional conventions have been set up to coordinate and make obtaining worldwide protection convenient and efficient. Several important treaties have been developed and are administered by the international organization known as the World Intellectual Property Organization (WIPO), centered in Geneva, Switzerland.

Paris Convention

This is the short title for the Paris Convention for the Protection of Industrial Property. This treaty was first signed in Paris in 1883 by 11 countries and has been revised several times over the years. Today 157 countries have ratified this treaty. A notable country of concern to the pharmaceutical industry that is not a member of the Paris Convention is Taiwan.

An important feature of the Paris Convention allows applicants to claim priority to their first-filed application in any member country, provided the applicant files a patent application in that country within 1 year. This is referred to as convention priority, and it is of great importance in the patent strategies developed in the pharmaceutical industry, as described below. If the first application meets the conditions of the Convention, the applications in all member countries are treated as if they were filed on the same day as the first application.

Budapest Treaty

When an invention involves a microorganism, most countries require the deposit of the biological material to complete the disclosure. WIPO established international uniformity under the Budapest Treaty of 1977, which became effective in 1980 and has been ratified by

48 countries. It provides for the establishment of a group of international deposit authorities. When a strain of microorganisms is deposited in any one of these authorities, this single deposit satisfies the necessary requirements for all signatory countries of this treaty. Under this convention, the formal requirements for making the deposit and maintaining the culture are set forth. Included in this is the possibility of a redeposit should the initial deposit become nonviable. The maintenance period of the deposit is a minimum of 30 years from the date of deposit.

Patent Cooperation Treaty (PCT)

The PCT first came into force in January 1978. It is in effect in 109 countries, including the United States, all the countries of the European Patent Convention, Japan, Canada, and China, as of January 1, 2001.

This treaty puts forth a process by which an applicant, through a simplified procedure, can file one application and designate that it be treated as an application in one or up to all of the PCT member countries. This application is filed in one of the official receiving offices in any of the official languages—English, German, French, Chinese, Japanese, Russian, and Spanish. The PCT has become a very important procedure in the pharmaceutical industry in providing an efficient means of obtaining maximum patent protection around the world in the most efficient and cost-effective manner.

The PCT application can be filed under the terms of the Paris Convention at the end of the 1-year period from the priority filing. The receiving office passes the application on to an International Searching Authority through which a search for the novelty of the invention is carried out. The application, together with the search report, is published 18 months from the priority date. After this publication, the applicant can choose two avenues for his application. First, if she or he wishes to proceed directly to each of the designated countries, the filing can be perfected in each of the designated countries within 20 months of the priority date. This is done by submitting the formal documents, translations, and fees required by each local patent office in each designated country. Alternatively, by the end of 19 months, a “demand” can be filed that the Preliminary Examining Authority of the PCT carry out a preliminary examination for patentability. This examination takes place between the 19th and 28th months of the priority date. During this period, the applicant receives the results of the examination and has the opportunity to present arguments in support of the patentability of his invention and/or amend his application. Again, the applicant must submit all formal documents, translations, and fees

required by each designated country to perfect the individual national filings by the end of the 30th month.

The main advantage of the PCT process is that it allows the applicant time to determine more clearly the commercial viability and importance of the invention. In the pharmaceutical industry, patents are filed on new potential products very early in their development. Many of these fall by the wayside during the development process. The PCT allows for an application to be filed and maintained with minimum expense up to the 30th month before the significant expense of national filing fees and translations is required. Thus, under a strategy using the PCT, the pharmaceutical company can maintain patent applications for the major economically important countries at minimum cost while having 30 months in which to study the invention and determine its commercial potential.

This treaty has been gaining in popularity and use in recent years. In 1999, the PCT was used to file 74,023 applications, compared with 14,874 in 1989 and 2,625 in 1979.

European Patent Convention

A regional system is one that arises from a regional treaty entered into by a number of countries within a geographic area. The European Patent Convention (EPC) is the most important of several regional patent systems. It was negotiated by a number of European countries and entered into force on October 7, 1977. Eight countries ratified it by the time the European Patent Office (EPO) began accepting applications on June 1, 1978. As of January 2001, 20 countries have ratified the EPC. In addition, six other countries, including Slovenia and Romania, have indicated that granted European patents can be validated to have effect in their countries as an alternative to applying for national applications in those countries.

Under the EPC, applicants can file one application in the EPO and designate in which of the 20 countries they desire the application to have effect. The application is examined by the EPO, and if the invention is found patentable, the application is granted, not as a single patent, but as a national patent for each country designated by the applicant at the time of filing. Under the EPC, it is possible to challenge the grant of the patent, provided the challenge is made within 9 months of the patent grant. This provides the opportunity potentially to have the patent declared invalid in one set of proceedings. This procedure is referred to as an opposition and still takes place in the EPO. If there is no opposition filed to the grant of the patent, then challenges to validity of the patent and enforcement need to take place in each country under the national patent laws of that country.

This convention provides an efficient means of obtaining patent protection in up to 26 countries through the filing and prosecution of only one application. When the EPC first went into effect, there was concern that if the prosecution was unfavorable, the applicant would not have any patent protection, whereas if the national system had been used, he would. With the passage of time, this concern has been greatly diminished, and now the convention is used by most, if not all, in the pharmaceutical industry on a routine basis.

Other Regional Systems

There are a number of other regional patent application systems in operation. In Africa, two regional systems operate: the African Regional Industrial Property Organization (ARIPO), formed in 1976, and the African Intellectual Property Organization (OAPI), formed in 1962. These systems have 11 and 15 member states, respectively. In the Middle East, the Gulf Co-operation Treaty operate to provide protection in six countries (for example, Saudi Arabia and The United Arab Emirates), and the Eurasian Patent System operates to provide protection in nine countries of the former Soviet Union, including Russia.

CLAIMABLE INVENTIONS IN THE PHARMACEUTICAL AREA

Section 101 of the U.S. Patent Law defines inventions and discoveries for which a patent can be obtained in very broad terms. Specifically, these are processes, machines, articles of manufacture, and compositions of matter, or any new and useful improvements of these. Article 52 of the EPC provides that a patent will be granted for any inventions that meet the three requirements for patentability. It then lists exclusions, which include “scientific theories and mathematic methods” in addition to medical treatments, as noted below. Many different inventions arising from research in the pharmaceutical field can be protected under these broad definitions.

Chemical Inventions

In the pharmaceutical field, chemical inventions have become of primary importance. The traditional inventions for which the pharmaceutical industry has sought patents are set forth in Table 1.

Specifically, a new composition of matter (NCM) or new chemical compound prepared by synthetic methods is the primary area of interest. They may be made using

Table 1 Potential chemical inventions

Compound per se
Pharmaceutical compositions of new compounds
New pharmaceutical compositions of existing compounds
Method of treatment, mechanistically or by disease state or both
Compound for use (broad first use)
New medical use for existing compound (second use)
Analogy processes for new compounds
Process per se (when novel and inventive)
Intermediates
Processes for preparing compositions
Different salt forms, hydrates, or polymorphs

synthetic methods or are isolated from natural sources such as plant or oceanic material or from fermentation broths. At times, a compound is known in an impure state that is unusable as a pharmaceutical product. If it is obtained in a purified state and meets the requirements of patentability, it can be claimed as a compound of a defined purity.

In addition to the compound per se, chemical processes for the preparation of the compound can be claimed. Analogous processes, which are a known chemical reactions that produce a new compound or use new starting materials, have various standards of patentability in different countries.

Pharmaceutical compositions of the new compound or new improved formulations of existing compounds are also patentable. For a new drug-delivery system, patent protection can be very important and useful, especially if developed for the delivery of an existing product.

A method of treating a particular disease or physiological condition is another important type of invention in the pharmaceutical field. It may be necessary to claim this type of invention by what is called “Swiss claims” to overcome the industrial applicability requirement in Europe. Such claims take the form of claiming the compound for use in the manufacture of a medicament for the treatment of a particular disease state. In the United States, the applicant may be required to provide proof to support the treatment claim. This is especially true if the invention is treating a disease that has been difficult or impossible to treat in the past, e.g., cancer or AIDS. A claim to “a method of treating cancer with compound X” will be challenged during examination. The most likely result will be amendment of the claim to a method of treating the specific cancer for which data can be provided to demonstrate utility.

Biotechnology Inventions

With the blossoming of biotechnology in the 1980s and, more recently, of genomics, a whole new specialty in

Table 2 Potential biotechnology inventions

Recombinant/genomics
The protein per se
Antibodies that react specifically with the protein and antiidiotypic antibodies
rDNA that encodes the protein
Expression systems
Recombinant host cells containing the DNA
Processes to make protein using recombinant host cells
Processes for purifying the protein
Processes to produce the antibody and antiidiotypic antibodies
Methods of using the DNA sequence
Pharmaceutical compositions of the protein
Method of treatment
Vectors
Promoters
Single nucleotide polymorphisms (SNPs)
Expressed sequence tags (ESTs)
Genomic DNA
Monoclonal antibodies
Monoclonal antibody (Mab)
Hybridoma that produces Mab
Process to prepare Mab
Method to use Mab
Pharmaceutical composition of the Mab
Novel antigen and related processes and methods

patent law has developed. In general, it is viewed that the principles of chemical patent practice apply equally to the biotechnology field. The Court of Appeals for the Federal Circuit (which is the U.S. Federal Court that hears all patent appeals from the Patent and Trademark Office and any Federal District Courts) has affirmed this in their decisions. Biotechnology inventions must satisfy the standard statutory requirements in the same manner as for any other invention. From the list of claimable inventions in Table 2, it is clear that biotechnology techniques have resulted in the production of inventions previously not obtainable by classic chemical methods.

REQUIREMENTS FOR PATENTABILITY

There are three basic requirements for patentability: novelty; nonobviousness or inventive step; and usefulness or industrial applicability. Each requirement may differ from country to country and is set forth by the statutes and regulations of each country.

Novelty

The first principle of patent law is that to obtain a patent, the invention must be new. The statutory requirements for novelty are set forth in Section 102 of the U.S. Patent Law and in Article 54 of the EPC.

In the United States, Section 102 has several stated requirements. One is that the invention must not have been known or used by others in the United States or patented or described in a printed publication in the United States or in any foreign country before the invention was made by the applicant. The second provision is that the invention must not have been patented or described in any printed publication anywhere in the world more than 1 year before the date of application in the United States. This is the 1-year grace period available in the United States but not in the rest of the world. These requirements are grouped under the term “anticipation” and mean that a single prior art reference must show the invention being claimed to nullify novelty. If an essential part of the invention is not present in a single publication, but is found in a second reference, anticipation does not exist (however, see the next section on obviousness).

Article 54 of the EPC defines an invention as “new if it does not form part of the state of the art.” It goes on to specify that the state of the art comprises “... everything made available to the public by means of a written or oral description, by use or in any other way before the date of filing....”

Nonobvious or Inventive Step

In many cases, the requirement of nonobviousness or inventive step as it is referred to in Europe, is the one that presents the most difficulty during the examination process. It is governed by Section 103 of the U.S. Patent Law and by Article 56 of the EPC. The general principle is that even though an invention is novel, a patent cannot be granted unless the inventor has done something more than one would expect any given person to have done in the art or field to which the invention pertains. Two or more references can be used to make the case of what would have been obvious to the given person skilled in the art. In the pharmaceutical field, chemical compounds are often known that are very close structurally and/or have similar utilities to the compounds claimed in the application. In these situations, the issue of proving nonobvious is the of most concern during the examination process.

Usefulness or Industrial Applicability

This requirement is covered by Section 101 of the U.S. Patent Law and by Article 57 of the EPC.

In the United States, usefulness is a very broad concept and does not mean a commercial utility is in hand. In the pharmaceutical field, a showing of in vitro test results that have some nexus with treating a physiological condition is sufficient to meet this requirement. If, however, the invention is useful only as a scientific curiosity, this requirement is not met. Chemical intermediates may or may not have the required usefulness. If they are useful to prepare products that are themselves useful then they are patentable. However, intermediates that are only useful to prepare compounds with no utility are not patentable.

In Europe, industrial applicability is required. This means, for example, that methods of medical treatment or diagnosis performed on a human or an animal are not susceptible to industrial application. Nevertheless, substances that are useful in these methods are patentable.

OBTAINING A PATENT

When an invention is made, consideration must be given as to whether to obtain patent protection. Companies have a variety of internal procedures to determine when and where to seek patent protection.

After it has been decided to obtain patent protection, a patent application is prepared and filed. This first filing is referred to as a priority filing and the filing date as the priority date. This is the date against which the invention will be judged for purposes of determining whether a patent should be granted, as discussed above.

Filing Applications

This first patent application is usually filed in the home country of the inventors. In many countries, there are laws and/or regulations that require this, or, if the inventor wants to file in another country first, certain requirements must be met. In the United States, there are regulations against exporting technology to other countries without government approval. An invention made in the United States must be filed first in the United States and not in any foreign country for 6 months after this filing without obtaining an export license from the U.S. Patent Office. If the applicant wishes to file first in another country, he or she must submit the application to the U.S. Patent Office and obtain the export license before filing in that country.

After the priority application is filed, the applicant must decide in what additional countries the application should be filed during the 12-month convention period available under the Paris Convention. This is often referred to as the foreign filing decision. The cost of protecting the invention

is directly proportional to the number of applications filed and the different language translations required by these filings. In times of cost consciousness, this foreign filing decision is not taken lightly.

Several options are available to applicants at this time. The first is to obtain a patent only in their home country and not pursue patent protection elsewhere. The second possibility is to proceed with a foreign filing in one or more foreign countries within the convention year. A final option is to abandon the home country application and not pursue patent protection.

The choice among these options depends on many factors. If the invention is completely or almost completely understood and its commercial potential known, then the decision as to in which countries to file can be made without much difficulty. Most pharmaceutical companies have listings of “filing groups” of countries that are used depending on the projected commercial potential for the invention. These filing groups may contain a few major countries up to a large number of countries, e.g., 50 to 100 countries or more. Each company develops its own set of filing groups based on numerous factors including costs, strength of protection in each country, the company’s markets, etc. Likewise, the selection of the filing group for each particular invention depends on factors such as type of invention, perceived commercial potential, projected market areas, and the like.

Alternatively, if the invention and its commercial potential are not fully understood, the applicant could abandon or abandon and refile. This process carries with it the danger that someone else may have filed a patent application between applicant’s priority date and the second priority date obtained through the refiling process. When this occurs, the applicant loses the rights to the invention in most countries. This can be especially dangerous in highly competitive areas in which many people are conducting research.

An alternate strategy is to file the foreign applications under the provisions of the PCT to obtain, an additional 18 months before the significant filing fees and cost of translations are incurred (see discussion under PCT, above).

Even after the 12-month priority period has passed, the applicant may still file patent applications, provided there has been no publication or public use of his invention. Applications filed in most countries including the EPO and PCT are published 18 months after the priority date. The United States will begin publishing applications filed on or after November 29, 2000. If the application has not been published and the applicant has not published in any scientific journals, he or she can still file in a country and obtain a patent. This is called a nonconvention filing

because the applicant does not claim rights back to his first priority date. Therefore, if new information about the commercial potential of the invention becomes available between the 12th and 18th months, one should always review it carefully and decide whether nonconvention applications should be filed.

Examination of the Application

Once an application has been filed, each patent office will examine it before granting a patent. The completeness of this examination differs from country to country. In some countries, it is a matter of merely verifying that the application has all the proper formal papers required under the laws and regulations of that country. Other countries carry out a rigorous examination as to the patentability of the invention in addition to the formal matters of proper documents. The United States, Japanese, and European patent offices examine patent applications for all requirements of patentability. In the United States, applications filed on or after May 29, 2000, must be granted within 3 years. If it is not, a patent term adjustment is available, provided the applicant has followed all procedures in a timely manner. Currently, the EPO starts the examination process approximately 2 to 3 years after the application is filed. Japan uses a deferred examination process. An application filed in the Japanese Patent Office remains dormant until the applicant requests the patent office to examine it. It can be deferred up to 7 years from the filing date. Once requested, the examination process in Japan normally takes 2 to 3 years to complete. On occasion, this process can take significantly longer because Japan has a pregrant opposition system. This means that once the patent office makes a determination to grant an application, it is published, and within 90 days of that publication date, anyone may file an opposition to the grant. If this occurs, a lengthy proceeding can ensue, possibly deferring the patent grant for many years.

Content of a Patent Application

Each country sets forth the requirements for a patent application through its laws and regulations. The United States does this through Section 112 of the Patent Law. This section specifies that the application must contain two parts: the specification and the claims. The requirements of the specification are set forth in the first paragraph of Section 112, as follows:

The specification shall contain a written description of the invention and the manner and process of

making and using it in such full, clear, concise and exact terms as to enable any person skilled in the art to which it pertains or with which it is most nearly connected to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention. To summarize, there are three requirements: a written description, an enabling disclosure, and the best mode. A written description and an enabling disclosure are universal requirements. However, the best mode requirement is distinctive to the U.S. patent system and is one that causes much debate among patent administrators. The requirement is based on the theory that applicants should not receive the privileges of patent protection if they have not disclosed the best methods of making and using the invention known to them at the time the patent application is filed. If applicants fail to disclose the best mode, even unintentionally, the patent can be held invalid.

Based on these disclosure requirements, the application normally contains the following:

1. Abstract
2. Summary of the background needed to understand the invention
3. Summary of the invention
4. Detailed description of how to practice the invention including appropriate examples and drawings.

The second paragraph of Section 112 requires that applicants point out and distinctly define the invention with a set of claims. Up to this point, a patent application has been a scientific article that teaches other people in the same field how to make and use the invention. This is no different than a scientific journal article. The claims, however, set forth what the applicant considers to be the invention for which patent protection is being sought. During the examination process, the claims may be modified to overcome objections. Therefore, one often finds, as a result of the examination process, the allowed claims of a granted patent often define less subject matter than is disclosed in the specification. Because the claims define the subject matter protected by the patent, when asked a question regarding infringement, a patent attorney will turn immediately to the claims to begin an analysis rather than to the description in the specification.

OWNERSHIP

Patents are property whose ownership may be governed law, contract, or by statutory provision. In many employer-

employee relationships, the employee has signed a contract of employment. This contract states that the employer owns all inventions made by the employee. In some countries, most notably Germany, there are elaborate statutory provisions for compensation of inventors by the employer. In all countries except the United States, the owner of the patent rights may file the patent application. Historically, the U.S. system has been based on rewarding the inventor and thus requires that the inventor apply for the patent. If the inventor is required by a contract of employment to assign the inventions to the employer, the U.S. application is made in the inventor's name, and an assignment noting the transfer of rights to the employer is recorded in the patent office.

INVENTORSHIP

In most of the world, inventorship of a patent may be relevant as to who owns the rights but it is not relevant to the validity of the patent. In fact, in some countries the inventor may never be mentioned or even appear on the patent. Again, the one important exception is the United States, where, as noted above, the inventor or inventors must file the patent application. If the wrong inventor or inventors intentionally apply for the patent, grounds for declaring the U.S. patent invalid are raised. Therefore, the proper inventors of the claimed invention are always determined according to the requirements of U.S. law.

It must be remembered that inventorship is different from authorship. Inventorship is based on legal requirements and must be strictly followed, whereas authorship is more arbitrary. The determination of inventorship is based on first inspecting the invention and determining what person or persons made an "inventive" contribution to the conception and reduction to practice of this invention. Conception is the mental steps taken to develop the invention. Reduction to practice is the physical process of taking the idea to the completed working invention. When two or more inventors (joint inventors) are involved, each must contribute to the claimed invention, but each is not required to have made a contribution to each claim of the patent.

Inventorship determination is not always straightforward and simple in today's research environment. However, to make these correct determinations, certain questions are asked. Did the person do only routine work or experiments as directed by another, or did he or she contribute something more? Was the invention completed because of the specific activity of this person? Did this person proceed beyond specific directions? In today's

modern pharmaceutical research atmosphere, in which teams are involved in the discovery and development processes, the patent attorney may find the determination of the correct inventorship a very difficult aspect in the preparation of a patent application. In a 1972 decision, U.S. District Court Judge Newcomer made the following observation:

The exact parameters of what constitutes joint inventorship are quite difficult to define. It is one of the muddiest concepts in the muddy metaphysics of patent law.

WHEN TWO OR MORE GROUPS MAKE THE SAME INVENTION

Sometimes two or more applicants make the same invention, but only one patent can be granted. The question is who receives the patent grant. In all countries except the United States, the first to file a patent application is granted the patent. In the United States, the patent is granted to the first to make the invention.

The first inventor is determined by a special administrative proceeding in the patent office called an interference. Evidence is presented by each applicant as to their earliest dates of conception and reduction to practice. Before January 1, 1996, only activities carried out in the United States could be used in these proofs. Now activities carried out in any country that is a member of the World Trade Organization can be used to prove when the invention was made.

In an interference, the proof requirements are very important. Activities or evidence given by the inventor must be collaborated in some manner. This is best accomplished through evidence from noninventors, although other forms of evidence have been used successfully in some cases. For this reason, research organizations have policies and procedures governing notebook recordkeeping in their laboratories.

Interferences are very complex and expensive proceedings that can delay the patent grant for many years. Fortunately, only less than 0.5% of all U.S. applications are involved in an interference.

LENGTH OF PATENT PROTECTION

The term of a patent grant is defined by the laws of each country, varying generally from 15 to 20 years. In some developing countries, patent terms are much shorter and

are of very little value to the pharmaceutical industry because the patent expires before the product can be marketed. Recently most countries, including the United States, have adopted a standard patent life of 20 years from the filing date. The United States, which until June, 1995, had a term of 17 years from the patent grant date, has recently enacted additional laws to restore any of the 20-year term lost because of delays in the patent office. Japanese patents have a 15-year term from the date of grant or 20 years from the date of filing, whichever is shorter.

In the pharmaceutical industry, the term of patent life is a very important factor. Because patents are filed very early in the life cycle of a new pharmaceutical product and much premarketing testing is needed before the health authorities will permit public sale of a product, a large portion of patent life is lost. Often, the term "effective patent life" for pharmaceutical products is used. Studies by the Pharmaceutical Research and Manufacturers of America have shown that effective patent life for pharmaceutical products averaged 15 years in the early 1960s and declined to 8 years in the early 1980s. Studies in Japan and Great Britain gave similar results. In recent years, the laws of many important countries have been changed to provide for the recapture of some of this lost patent life through patent term extension provisions.

Patent Term Extensions

Before 1980, in some countries, mainly those that were formerly part of the British Commonwealth, patent extensions were obtainable on petition at the end of the patent life. For a petition to be granted, the patent owner was required to show inability during the normal life of the patent to receive sufficient remuneration from the use of the patented invention. Under this system, patent extensions of 4 to 10 years could be obtained based on the evidence presented.

In 1984, the United States passed the first patent term restoration statute as part of the Drug Price Competition and Patent Term Restoration Act of 1984 (often referred to as the Waxman-Hatch Act). This was a significant law for the total pharmaceutical industry because it contained provisions important to both the research-based and the generic sections of the industry. Under this act, generic pharmaceutical companies were allowed to file abbreviated new drug applications (ANDA) and to do the testing required to submit an ANDA before the patent expired without being liable for patent infringement. The second part provided for the innovator of a new pharmaceutical product to receive up to 5 years of patent extension based on the time required to receive marketing approval from the Food and Drug Administration.

The extension can provide effective patent life for the product no longer than 14 years from the date of marketing approval.

Other countries have followed the U.S. lead regarding patent term extension. Japan enacted a provision effective January 1, 1986, whereby patents covering pharmaceutical products could be extended for up to 5 years. The Japanese provisions differ from the U.S. law in several ways. First, more than one patent can be extended for each product. Also, the Japanese law requires that the patent be granted 2 years before health authority approval of the product. Korea has enacted a patent term extension law very similar to that in the Japanese system. Australia has replaced its “lack of remuneration” system with a system based on regulatory delay.

Effective on January 1, 1993, Supplementary Protection Certificates (SPC), a system of providing extended protection for pharmaceutical products, was created by the European Community (EC). Such certificates do not actually extend the patent per se but confers protection for the product covered by the granted marketing approval (MA) and any use of the product as a pharmaceutical product for humans or animals. If more than one patent protects the product, the owner must select only one patent to be the basis for the SPC. SPCs are granted by individual national patent offices and come into effect only after the normal patent term expires. The duration of the SPC is equal to the time from the patent filing date to first MA in the EC, minus 5 years, subject to the limitations that the SPC cannot be effective for more than 5 years or provide protection for more than 15 years from the date of first MA in the EC.

The patent owner must request the SPC in the patent office of each country EC within 6 months of the first MA in that country or within 6 months of the patent grant date if it occurs after the first MA in that country. The SPC will be granted if, at the time the application is filed, 1) the patent has not expired; 2) an MA has been granted; 3) the product has not previously been the subject of an SPC; and 4) the MA is the first authorization to market the product.

The regulations allowed Spain and Greece to delay accepting the SPC system for 5 years. In addition, transitional provisions were adopted that allowed each country to choose to allow SPCs for products whose first MA was granted after January 1, 1982, June 1, 1985, or January 1, 1988.

MAINTAINING AND ENFORCING PATENTS

After a patent is granted, the owner must pay fees to keep it in force and may have to defend it from challenges to its validity and/or enforce it against infringers.

Maintenance Fees

In most countries, a patent does not automatically stay in force from the day it is granted until the end of its life. Most countries require payment of fees, referred to as renewal or maintenance fees, to keep the patent in force. In most countries, these fees are paid on an annual basis. Until 1980, the United States was a notable exception. However, in 1980 the United States modified its patent laws to require that fees be paid at three different time periods during the life of the patent. Specifically, renewal fees must be paid at 3.5, 7.5, and 11.5 years from the date the patent is granted. Although renewal fees vary from country to country, it is universal that the fees increase over the life of the patent. For example, the U.S. fees in 2000 are \$930, \$1870, and \$2820 for the three periods, whereas in Germany fees begin at approximately 150 DM and increase to 3500 DM for the 20th year. If the owner of the patent knows that it has no value to himself or others, he or she will generally stop paying the fees and allow the patent to lapse. Maintenance fees on a portfolio of patents can become very expensive, therefore, most companies have a program of regular review of their patent portfolio. Patents, that are no longer of value to them are lapsed.

Invalidation

Once a patent is granted, the patent owner is not guaranteed that it is valid. Challenges to the validity of the patent again vary from country to country, depending on the specific patent law of each country.

In the United States, the patent law specifically provides that a granted patent carries with it a presumption of validity. Until 1980, the validity of a patent could be challenged only in a federal court as part of an infringement action or by suit for declaratory judgment of invalidity. This latter action could only be initiated by a party who was threatened by the patent owner so that the requisite legal dispute actually existed between the parties. In 1980, the patent law was amended to allow any person to file a request in the patent office for re-examination of a U.S. patent based on new prior art. The patent office studies the request and determines whether to grant reexamination. This procedure is not used very often because the third party requestor has little opportunity to fully present its views. On November 29, 1999, a new reexamination law was passed that applies to patents originally filed after that date, which allows for greater participation by the third party.

In Europe, challenges to the validity of patents must be made in each individual country unless a central opposition has been made to a European patent (see above).

PATENT INFRINGEMENT

A patent is infringed when someone makes uses or sells the claimed invention without the permission of the patent owner. When this occurs, the patent owner can take legal action against the party.

In the United States, this is a civil action brought in the federal courts. The specific federal district court in which a patent owner can sue an infringer is governed by federal law. Patent litigation is very expensive and time-consuming. For this reason, it is not entered into without careful consideration of the consequences and analysis of all options available. In recent years, various alternative dispute-resolution proceedings have been used more often.

Because patents are limited in their effect in a specific country, a company with a pioneering drug may find itself suing patent infringers in a number of other countries. Usually these actions are in countries where the patent system is not as strong as that in the United States and the major European countries. The direct costs of patent litigation are usually lower outside the United States, but the time requirements can be just as extensive.

Patents to compounds or pharmaceutical compositions are infringed by their sale or use in the country of issuance. Patents to processes for the preparation of a compound cause a different problem. There is no direct infringement if the compound is made by the process in a country without patent protection and then imported into the country where the process is patented. However, the laws of many countries specify that the sale or use of compounds made directly by a patented process is an infringement of the process patent. In most cases, “directly” means the final step to prepare the compound. Such protection was not available in the United States until 1986, when the patent laws were amended. The amendment also has a provision whereby a party can ask the producer of a compound for a list of any U.S. process patents that are owned or licensed.

REGULATORY EXCLUSIVITY AND ORPHAN DRUGS

Although patents are the main defense against generic copying in the pharmaceutical industry, another important form of protection that has developed in recent years relates to the data generated by the originator of a product in support of the marketing approval application. Specifically, valuable periods of so-called regulatory exclusivity have become available in the United States,

Europe, and Japan. In addition, in the United States, the Orphan Drug Act has provided a special kind of regulatory exclusivity.

Regulatory Exclusivity

The interests of public safety and avoidance of unnecessary animal experimentation make it desirable that licensing authorities be able to cross-refer the originators file to establish safety and efficacy of generic versions of a product. Applications for marketing authorizations that rely on cross-referral to the originators full regulatory submission are generally known as abridged regulatory applications. Regulatory exclusivity is a temporary prohibition on cross-referral to the originator’s data without the originator’s consent. After a specified time, as detailed below, the originator cannot object to the cross-referral. Regulatory exclusivity is acquired rather than applied for by the originating drug manufacturer, and no certificate or other documentation is issued for the exclusivity period.

In 1980, Japan became the first country to have a regulatory exclusivity provision. The period is 6 years for new products, new combination products, or different routes of administration. For other modifications, such as a new indication or a new dosage regimen, the period of exclusivity is 4 years. If second applicants do their own complete safety and efficacy studies, these periods of exclusivity do not apply.

In the United States, these regulatory exclusivity rights are part of the Waxman-Hatch Act. The law provides that no abbreviated new drug application (ANDA) can be submitted for a generic equivalent for a new chemical entity (NCE) until 5 years after the approval date for the NCE. If the ANDA applicant certifies that the patent covering the NCE is invalid or not infringed, the period is 4 years. If a NCE is not involved or a pioneering supplemental NDA is approved,

Table 3 European regulatory exclusivity

10 Years	6 Years
United Kingdom	Denmark
Austria	Finland
France	Greece
Germany	Ireland
Belgium	Luxembourg
Holland	Portugal
Italy	Spain
Sweden	

a 3-year period exists before an ANDA can *become effective*. Thus, for non-NCEs, the regulatory period is reduced to 3 years. For a NCE the 5-year period is for *submission* of an ANDA, and so the FDA processing time would extend this exclusivity period.

Another interesting part of this law is that if a generic drug manufacturer challenges a patent and is successful, it is rewarded with a 6-month period during which no other generic products will be approved.

In Europe, regulatory exclusivity is more complex. More specifically, the countries may choose between a 10-year or 6-year period. According to European Community Directive 87/21, the national states can protect pharmaceutical products for 6 or 10 years from first marketing approval in the European Community. An exception to this is biotech products for which a 10-year period applies for all European countries. The countries that have chosen 10 or 6 years are presented in Table 3.

Orphan Drugs

Drugs that are used for treating rare diseases or conditions are called orphan drugs. The U.S. Orphan Drug Act provides that the FDA may not approve another application within 7 years of approval of the first unless the originator cannot assure availability to meet the needs of patients. This law was amended in 1985 to apply to patented as well as nonpatented drugs. The requirements to be recognized as an orphan drug that are 1) the disease affects fewer than 200,000 persons in the United States, or

2) recovery of the R&D costs from U.S. sales is unlikely. A request must be made to the FDA to obtain orphan drug status. The determination of eligibility is made as of the date of the request. This 7-year period of marketing exclusivity applies only to the individual uses of the compound or product and not to the compound itself.

In Japan, a law was introduced in October 1993 providing benefits for designated compounds including an accelerated examination, a 10-year rather than 5-year regulatory exclusivity period, and tax benefits. The conditions for orphan drug status are that the drug must be for the treatment of diseases with fewer than 50,000 patients in Japan and clinical trials essential to development for that use.

There are no specific orphan drug regulations in the EU. Other countries do not have exclusivity for orphan drugs at this time; however, provisions exist in some countries for speedier approval.

BIBLIOGRAPHY

- Grubb, P.W. *Patents in Chemistry and Biotechnology*; Clarendon Press: Oxford, 1986.
- Rosenstock, J. *The Law of Chemical and Pharmaceutical Invention: Patent and Nonpatent Protection*, 2nd Ed.; Aspen Law and Business: New York, 2000.
- Wallerstein, M.B. Magee, M.E.; Schoen, R.A. *Global Dimensions of Intellectual Property Rights in Science and Technology*; National Academy Press: Washington, DC, 1993.

PATENTS—UNITED STATES PERSPECTIVE

Lorie Ann Morgan

GlaxoSmithKline, Research Triangle Park, North Carolina, U.S.A.

Jeffrey Tidwell

University of North Carolina, Chapel Hill, North Carolina, U.S.A.

INTRODUCTION

The importance of pharmaceutical patents has increased with the dramatic growth of the generic pharmaceutical industry in the United States and the resulting competition between the research-based pharmaceutical industry and the generic pharmaceutical industry. The research-based pharmaceutical industry provides a continuing supply of new and better pharmaceuticals to treat illness and improve lifestyle. The generic pharmaceutical industry, on the other hand, strives to provide lower-cost generic alternatives to already approved drugs. Patents provide the research-based pharmaceutical industry with an opportunity to recoup the extensive investment costs involved in providing new and better pharmaceuticals to the public. At the same time, patents restrict the generic pharmaceutical industry's ability to obtain approval for competing generic drugs. While a patent claiming the drug is in force, the generic pharmaceutical manufacturer is precluded from conducting infringing activities. The balance between the need for extensive Food and Drug Administration (FDA) review and approval of generic drugs and the need to preserve the enforceability of patents protecting newly developed drugs and provide incentives for further research and development of new drugs prompted the Hatch–Waxman Amendments in 1984 (Drug Price Competition and Patent Term Restoration Act of 1984).

INNOVATOR PATENTS

Types of Patent Claims

Pharmaceutical companies spend billions of dollars on research and development programs each year in an effort to discover and develop innovative new medicines. As such, protection of these novel drugs and new technology developed during the research process from outside competitors is of paramount importance. The patent laws of the United States provide a mechanism by which

research-based pharmaceutical companies can protect their inventions and commercialize them exclusively for a term of 20 years. When applying for a patent, an inventor sets forth the subject matter to be protected at the end of the patent specification in the claims. The claims of the patent determine the scope of protection during the 20-year exclusivity period.

Patentable subject matter is "... any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof ..." (1). Patent laws provide that patents may be granted only for inventions that are deemed to be "new and useful." These requirements exclude pharmaceutical inventions that have not been shown to be reasonably safe and effective and chemical compounds that have no use except as intermediates for additional research. However, some special considerations have traditionally been applied in the case of chemical compounds, particularly those possessing therapeutic or pharmacological activity. A compound that is useful as an intermediate for the manufacture of a pharmacologically active compound may be claimed and afforded patent protection because it satisfies the utility requirement and is considered to be "useful." Further, a compound need not show therapeutic utility in humans to meet the utility requirement of patentability; pharmacological activity in animals or in vitro activity is considered sufficient in some cases. If the patent application purports to claim the use of a compound for the treatment of humans, the inventor must show that the compound is both safe and effective in humans.

Multiple patents may be issued that provide protection for the same new drug product because a single invention may be claimed in a number of different ways. For example, a patent may claim the compound itself, a novel use for the compound in treating disease, and a method for preparing and administering the compound.

Composition of matter

The most expansive protection is afforded by claims that present the particular novel product or drug compound

itself. Claims to specific chemical entities permit the innovator to exclude others from making, using, selling, offering for sale, or importing the drug compound into the United States regardless of the purpose for doing so and regardless of the formulation into which the drug is included.

Composition-of-matter claims can include claims to small organic compounds that may be useful as drug candidates as well as claims to more complex biomolecules. For example, a composition-of-matter claim may encompass claims to proteins, plasmids, pieces of DNA or RNA, and pharmacological receptors.

Another area of particular importance to the pharmaceutical industry of late is the patentability of living organisms. The last several years have seen a dramatic increase in the pace of biological research and technology. As a result, the courts have had to deal with the question of whether or not manmade living organisms qualify as patentable subject matter. Although the courts struggled with this issue, it is now fairly settled that manmade living organisms, such as genetically engineered animals, biologically pure cultures, and genetically engineered microorganisms, are patentable subject matter.

Natural products

An area of great importance to the pharmaceutical industry in particular relates to the patentability of natural products for use in treating disease. The last several decades have witnessed a dramatic increase in the search for compounds from natural sources, such as plants, for the treatment of human disease. It is established law that natural products are considered to exist in the “state of nature” and are not considered novel. Therefore, natural products generally cannot be claimed as a patentable composition of matter. This is true even if the natural product is isolated by novel means or is tested against novel drug targets. However, there are two ways in which natural products may be patentable. First, it is possible to obtain a method-of-use patent involving a natural product as it relates to novel biological or pharmacological activity. Second, it is possible to obtain composition-of-matter coverage for synthetic analogs of natural products because they are not considered to exist in the state of nature. This is of paramount importance because many new drugs are closely related, synthetic analogs of natural products.

Method claims

There are two types of method claims: method-of-making (or process) claims and method-of-using claims. An inventor may protect a new drug by claims directed toward the method of synthesizing the compound. Process claims provide the inventor with the right to exclude others

from using the claimed methods to prepare a chemical compound. In addition, the Process Patent Amendment Act (2), discussed below, provides the inventor with the power to prevent others from importing into the United States any compound or formulation thereof using the methods claimed in the patent. However, a process patent does not prevent others from making or using the compound or formulation when it is prepared by a method other than that claimed. Therefore, although method claims are valuable in serving to protect drug products, they are much more limited in scope than compound per se claims.

Another area of importance in terms of patenting methods of making relates to novel methods developed for the synthesis of chemical or biological target molecules. The last decade has witnessed an explosion of innovation in the areas of biochemistry and combinatorial chemistry, particularly as they apply to drug discovery and development. New methods of generating vast numbers of biological molecules, such as DNA and proteins, as well as new technology used to prepare chemical “libraries” for use in screening new drug candidates have exponentially increased the number of patent applications directed toward these applications. It is generally established that these new types of technology are patentable subject matter and as such can be patented.

The second type of method-of-use claims relates to methods of using the invention. In terms of pharmaceuticals, method-of-use claims typically relate to the use of a particular drug in the treatment of specific diseases or conditions. FDA approval of a new drug is linked to specific therapeutic uses, and therefore these types of method claims provide important protection for the innovator. These claims are important because they provide patent protection that is in addition to any composition-of-matter claims that may be applicable. Method-of-treatment claims are particularly important for compounds that are already subject to composition-of-matter claims or for which composition-of-matter claims are not available. For example, an inventor may discover that a previously known compound possesses unreported and unexpected anti-inflammatory properties. The inventor may not obtain composition-of-matter protection for the compound itself because the compound is contained in the literature and therefore does not meet the novelty requirement for patentability. However, the inventor may instead obtain a method-of-use claim based on the unexpected anti-inflammatory properties because a compound’s unexpected properties may be enough to support patentability.

Chemical intermediates

Generally, novel chemical compounds that are intermediates for the preparation of other chemical compounds of

unknown utility are not considered patentable subject matter. However, if the intermediate is used to produce a compound that is known to be useful, then the utility requirement is satisfied, and a patent may be obtained for that intermediate. Claims directed toward chemical intermediates are particularly valuable when they cover stable intermediates that are critical to the only commercially feasible synthetic route to the drug.

Formulations

Still further protection may be afforded by claims that cite a specific formulation of a drug product. Formulation claims allow the inventor to exclude others from making, using, selling, offering for sale, or importing the claimed formulation into the United States. They represent an important opportunity for expanding the patent portfolio protecting a drug, in part because FDA approval for a drug is specific to the formulation containing the drug. Formulation claims may not provide exclusivity to the innovator, but formulation claims expiring after the expiry of compound claims may be a useful tool in maintaining maximum market share for as long as possible.

Reach-through claims

Another recent development in pharmaceutical patents involves so-called reach-through claims. Reach-through claims seek to cover, for example, drugs used for the treatment of conditions that are identified through a particular drug-screening assay without actually specifying the drugs that are covered. These claims, once granted, are exploited through licensing to seek royalties based on sales of products that are developed in part through use of the patented research tool assay. This particular strategy has become a favorite among smaller biotechnology companies that have expended a large portion of their capital in developing new technologies or materials. Several well known examples of reach-through claims are licenses under the Cohen-Boyer patent on basic recombinant DNA techniques and the patents on the Harvard recombinant onco-mouse and Roche polymerase chain reaction. The status of reach-through claims has yet to be determined in the courts.

Compound claims directed toward the drug compound generally afford the broadest scope of patent protection against competitors. However, no single type of claim can afford the best patent protection. The best protection against competitors includes a portfolio of patents, each of which is directed toward a different aspect of the drug. For example, an inventor may obtain a patent for composition of matter that covers the chemical compound itself, another for its method of synthesis, several relating to different formulations, and still others to different therapeutic uses

of the drug. Typically, one seeks to protect new drugs with a broad portfolio of patents to make it difficult for competitors to design around any one patent. As such, it is essential that generic drug manufacturers identify the full range of patents surrounding a particular product to avoid liability for patent infringement.

EXCLUSIVITY FOR THE INNOVATOR

Patent Exclusivity for the Innovator

Patent term

Before June 8, 1995, U.S. patents expired 17 years from the date the patent was issued. However, changes were made in the U.S. patent laws in 1994, in accordance with the Uruguay Round Agreements Act (URAA). The URAA governs the specific date on which a granted U.S. patent expires, which is 20 years from the filing date of the application (3). The changes enacted in response to the URAA were not made to extend patent terms, but to harmonize the term provisions of U.S. patent law with those of other, leading trading partners. In the case of applications filed as continuations or divisionals of previously filed applications, the filing date of the first U.S. priority application is the date from which the 20-year term is measured. For international applications filed under the Patent Cooperation Treaty (PCT), the 20-year term is measured from the filing date of the international application designating the U.S. Foreign national priority (under 35 U.S.C. §119) is not considered in the calculation of the 20-year term. For example, an application filed on June 10, 1995, that does not assert any claim priority to an earlier U.S. application would have an expiration date of June 10, 2015. However, an application filed on January 23, 1997, that is a continuation of an application filed on June 10, 1995, and that claims priority to that date, would have an expiration date of June 10, 2015. The expiration date is the same if the application claims a priority to a previous foreign application.

For those patents currently in force and those that will issue from applications filed before June 8, 1995, the patent term is the longer of either 17 years from the issue date or 20 years from the first U.S. filing date. It is possible for the terms of patents that have already been issued to be extended if they fall within the transition period. The expiration date of an issued patent is extended if the date that is 20 years from the first U.S. filing date is later than the date that is 17 years from the patent issue date. Consequently, any patent that required less than 3 years of prosecution from the first U.S. filing date receives an

extension of term by virtue of the URAA 20-year term provisions.

A U.S. patent application filed on March 1, 1992, and issued on April 1, 1994, would have an expiration date calculated under the provisions for the transition period. The expiration date calculated 17 years from issue would be April 1, 2011. The expiration date calculated 20 years from the first U.S. filing would be March 1, 2012. Because the patent term as calculated 20 years from filing is longer, the expiration date of this patent would be extended to March 1, 2012; thus, the patent will receive an extension of term equal to 11 months. The extension is granted despite the fact that the patent was issued before the passage of the URAA provisions.

In addition to these term provisions, in 1999 Congress passed the Patent Term Guarantee Act as part of the American Inventors Protection Act. The Patent Term Guarantee Act, which will go into effect on May 29, 2000, provides that patent terms will be extended to compensate for certain processing delays in the U.S. Patent and Trademark Office and for delays in the prosecution of applications pending for more than three years. Also, extensions are available for delays in issuance of a patent attributable to interference proceedings, secrecy orders, and appellate review. Patent applicants that demonstrate due diligence in the prosecution of their application are guaranteed a minimum 17-year patent term. The act applies only to applications filed on or after the date of enactment, which is May 29, 2000.

Patent-term restoration

Patents claiming a pharmaceutical product, therapeutic uses for a pharmaceutical product, or methods of making a pharmaceutical product may also be eligible for an extension of patent term under the Patent Term Restoration Act (Hatch–Waxman Amendments) (4). Its purpose is to allow the innovator to recover valuable patent term lost during the regulatory review process. The innovator's patented product cannot be commercialized until regulatory approval is obtained. As a result, the innovator is not permitted to reap the benefits of its patent exclusivity until the lengthy approval process is complete. To be eligible for a patent term extension, the patent must claim a product or method of using or manufacturing the product, the patent must not have been previously extended, the application for patent term extension must be submitted within 60 days of approval, the product covered by the patent must have been subject to regulatory review before commercial marketing and use, the approved commercial marketing and use of the product must be the first commercial marketing or use under which the regulatory review period occurred, the patent must not have expired, and no other

patent term must have been extended based on the same regulatory review period (5). Although many patents may cover a single product, only one patent per product may be extended under these provisions, and the extension of the patent term applies to those claims that cover the product receiving regulatory approval.

The length of the patent term extension is directly related to the length of the regulatory review period. The period of extension is calculated by adding half of the length of the testing (IND) period to the length of the approval (NDA) period and subtracting any part of the period that occurred before the issuance of the patent and any part of the period in which the applicant did not act with due diligence (6). The maximum available term extension is 5 years (7). The total patent term, after the extension period, cannot exceed 14 years after the date of approval by the FDA (8). In other words, the patent expiry date, after the addition of the extension period, must be not later than 14 years after the date of FDA approval.

The passage of the URAA, adopting a 20-year patent term, raised questions in circumstances in which a single patent was eligible for a longer term under the URAA 20-year term provisions (hereinafter “URAA extension”) and had also received an extension under the Patent Term Restoration Act (hereinafter “Section 156 extension”). The most important question was whether a patentee could reap the benefit of both extensions by adding the URAA extension and the Section 156 extension to prolong the period of patent exclusivity. The Federal Circuit Court of Appeal resolved this issue in *Merck & Co. v. Kessler* (9). In that case, the Court held that the expiration date of a patent having both a URAA and a Section 156 extension is 20 years from filing plus the Section 156 extension, unless the patent is in force on July 8, 1995, solely by virtue of the Section 156 extension. In other words, the foregoing rule applies, except when the patent's 17-year term had expired before June 8, 1995, and the patent is only enforceable by virtue of the Patent Term Restoration Act. Unfortunately, the Federal Circuit Court did not provide a calculation for the expiration date of patents that are in force on June 8, 1995, only because of the Patent Term Restoration Act. However, it would seem that few patents are affected by this omission.

FDA Exclusivity

In addition to patent exclusivity, an innovator may help preserve its market share for a new drug through FDA exclusivity. FDA exclusivity does not guarantee the innovator an exclusive right to market the drug, but it does operate to block approval of ANDAs for generic versions of the drug. FDA exclusivity does not prevent

a manufacturer from developing and obtaining approval for a new formulation using the same active ingredient by filing a new drug application (NDA).

The length of the FDA exclusivity period depends on the nature of the new drug developed by the innovator. The statute provides that no ANDA may be submitted for a drug (i.e., new chemical entity) approved after September 24, 1984, until the expiration of 5 years from the date that the innovator received first approval, with one exception (10). An ANDA may be submitted before the expiration of 5 years if it contains a certification of patent invalidity or noninfringement (11). A new NDA for a previously approved drug, if supported by new clinical investigations, enjoys a 3-year exclusivity period during which no new ANDAs may be approved based on the new NDA (12). Supplemental NDAs supported by new clinical investigations also enjoy a 3-year period of exclusivity (13).

The Food and Drug Administration Modernization Act (FDAMA) was passed in 1998, providing for an additional period of exclusivity based on clinical trials conducted in pediatric patients (hereinafter “pediatric exclusivity”). The award of exclusivity is provided as an incentive to the industry to conduct pediatric studies that are requested, but not required, by the FDA. Pediatric exclusivity provides an additional 6 months of exclusivity at the end of any remaining exclusivity or patent life of any patents covering the approved product (14). To be eligible for pediatric exclusivity, the innovator must receive from the FDA a written request to conduct pediatric studies, submit study reports after receipt of the written request, and meet the conditions of the written request. The applicant may request that the FDA issue the necessary written request to qualify for pediatric exclusivity. The award of exclusivity attaches not only to the product tested in the pediatric studies but to any formulations, dosage forms, and indications for products having existing exclusivity or patent life that contain the same active ingredient.

Notice to Generic Manufacturers of Innovator’s Exclusivity

The FDA provides notice of the innovator’s FDA and patent exclusivity by publishing information regarding exclusivity in the FDA Approved Drugs Product List (the “Orange Book”). The Patent and Exclusivity Appendix to the list of approved products provides details regarding the expiry dates of all types of FDA exclusivity (including pediatric exclusivity) and the expiry dates of patents for each approved product.

The patent information provided in the Orange Book for an approved product is obtained from the innovator.

Innovators are charged with the duty to notify the FDA of the patent number and expiration date of any patent that claims the drug (i.e., the active ingredient or the formulation) or its approved therapeutic use (15). The innovator must update the patent information for an approved product in an appropriate and timely manner as new patents covering the approved product are granted. The innovator must also update the patent information to account for any patent term extension awarded. The submission of this patent information to the FDA ensures that generic drug manufacturers are on notice of the innovator’s patent rights and invokes certain innovator rights, discussed below, with respect to notification of an ANDA filing by a generic drug manufacturer and infringement action based on ANDA filing. For these reasons, it is in the innovator’s best interest to provide patent information in a timely manner, and, for purposes of competing with the innovator, it is in the generic drug manufacturer’s best interest to be knowledgeable of the patent information provided in the Orange Book.

GENERIC COMPETITION

The ability of a generic pharmaceutical manufacturer to commercialize a generic version of an innovator’s product centers around two closely related issues: FDA marketing approval and the innovator’s patent portfolio. In addition to the requirements designed to establish that the generic drug is in fact equivalent to the approved innovator drug in terms of composition, manufacture, biological activity and labeling (16), the ANDA must also include a patent certification, also known as a paragraph i, ii, iii, or iv certification, regarding the nature of any patent exclusivity for the approved drug and its approved indications (17). The ANDA applicant must certify that in the opinion of the applicant and to the best of the applicant’s knowledge with respect to each patent that claims the approved drug or its approved therapeutic use:

- i. Patent information has not been filed;
- ii. The patent(s) has expired;
- iii. The date on which the patent(s) covering the approved drug will expire; or
- iv. The patent(s) covering the drug is invalid or will not be infringed by the manufacture, use, or sale of the generic drug for which approval is sought.

The paragraph (i) certification states that patent information for the approved drug on which the ANDA is based has not been filed according to the requirements of the statute. This paragraph applies when there are no patents

covering the approved product or its use or if the innovator has failed to properly list the patents covering the drug. It is very rare that no patents will exist for an approved drug, with the possible exception of very old drug products. This is because the investment required for an innovator to obtain marketing approval for a new drug is so great that it would not be feasible for an innovator to commercialize a new drug that lacked a significant period of patent exclusivity during which the innovator could recover the substantial costs of research and development. However, a paragraph (i) certification also applies when the innovator has failed to comply with the rules requiring the listing of all patents covering the approved drug. If the innovator failed to list the patents covering the approved product, and the ANDA applicant could establish that it was not aware of any unlisted patents that claimed the approved drug, the ANDA applicant would likely choose to include a paragraph (i) certification in the ANDA. An ANDA filed with a paragraph (i) certification may be approved effective immediately, barring any FDA exclusivity held by the innovator.

Under a paragraph (ii) certification, the generic drug manufacturer certifies that the patents covering the approved product have already expired. Typically, paragraph (ii) certifications are used only for older drug products because it is common for the ANDA to be filed before the actual expiry of the innovator's patents. An ANDA filed with a paragraph (ii) certification may be approved effective immediately, barring any FDA exclusivity held by the innovator.

Under a paragraph (iii) certification, the generic drug manufacturer certifies that the innovator patents will expire on a certain date and requests marketing approval as of that date. In this case, the FDA may grant tentative approval of the generic drug manufacturer's ANDA before the patent expires, but the generic drug manufacturer does not receive full approval and therefore cannot commercialize the generic product until after the patent expires.

Under a paragraph (iv) certification, the generic drug manufacturer certifies that the innovator patent(s) is either not infringed by the generic product or that the patent(s) is invalid or unenforceable. Paragraph (iv) certifications are discussed in more detail in the following section on infringement.

In this manner, the approval of an ANDA is, in part, dependent on the generic drug manufacturer avoiding infringement of the innovator's patents. The first step to avoiding infringement of the innovator's patents is to identify the particular patents that cover a particular approved product.

Identifying Patents Relating to a Drug

The Orange Book

There are several ways in which a generic pharmaceutical manufacturer can identify the extent of patent exclusivity on a product it wishes to produce. An innovator must submit to the FDA a list of patents that it believes covers the drug product, formulation, or specific therapeutic use of the new drug (18). The patents are subsequently published in the Orange Book. The patents to be listed must be submitted in a timely manner. Any patents directed toward the drug that might issue after the approval of the drug must be disclosed to the FDA within 60 days of issuance. This ensures that the FDA listing of patent exclusivity for a given approved drug is current. Although the FDA does not have the resources to police innovators to ensure compliance with the patent identification regulation, the courts have occasionally imposed a disincentive for failing to comply with the regulation. In particular, one court has determined that the innovator could not sue the ANDA applicant for infringement on the basis of filing the ANDA because the innovator did not list the patent in the Orange Book before filing the law suit (19). The rationale given by the court is that Congress intended that an ANDA applicant consult only the Orange Book to determine the existence of an applicable patent claiming the listed drug or a use of the listed drug.

The Orange Book is an appropriate starting place for the identification of patents claiming an approved drug, formulations of the approved drug, and therapeutic uses of the drug. However, innovators are not required to specify process patents in the Orange Book. As a result, the search for patents that may prohibit the commercialization of a generic form of the drug cannot be complete by merely referring to the Orange Book. For purposes of identifying relevant process patents relating to an approved drug, the Process Patent Amendment Act of 1988 requires the innovator to identify patents it considers relevant to a particular drug on receipt of a written request to do so.

Requests under the Process Patent Amendment Act

The Process Patent Amendment Act of 1988 encourages the generic pharmaceutical manufacturer to submit a request for disclosure of process patents relating to a particular drug to the innovator or patent holder (20). Making such a request is considered evidence of good faith on the part of the generic drug manufacturer in the event that litigation later ensues over the patent(s) (21). The request for disclosure must be in writing and be made to a person engaged in the manufacture of a product, asking for

identification of all process patents owned by or licensed to that person as of the time of the request that the person reasonably believes could be asserted to be infringed if the product is imported into or sold, offered for sale, or used in the United States without prior authorization (22). The request for disclosure must be made by a person regularly engaged in the United States in the sale of the same type of products as those manufactured by the party to whom the request is submitted (23). It must also be made before the first importation, sale, or offer for sale of the product by the generic pharmaceutical manufacturer (24). Finally, it should include a statement that any patents identified by the innovator will be submitted to the party who will manufacture or supply the drug and that the generic drug manufacturer will request from the supplier a written statement that none of the processes claimed by the patents is used in the manufacture of the generic drug (25).

After the request for disclosure is received, the patent owner must reply within a reasonable period to be considered to have acted in good faith. However, no reply is required if the products are marked with the process patent numbers before the request is received by the innovator (26).

The Orange Book and the Process Patent Amendment Act are appropriate avenues for identifying the patent exclusivity of an innovator with respect to a specific drug. However, before proceeding with the investment required for the regulatory approval of a generic drug, the pharmaceutical manufacturer should conduct an independent search of the patent literature to identify additional relevant patents. For example, parties other than the innovator may hold patents that affect or even block the ability to commercialize a generic drug. Patent holders other than the innovator are not subject to the same statutory requirements and, furthermore, may not be readily identifiable without an independent search of the patent literature.

PATENT ENFORCEMENT BY THE INNOVATOR

General Principles of Patent Infringement

Infringement is defined as the making, using, selling, offering for sale, or importation into the United States of any patented invention without the authority of the patent owner (27).

Liability for patent infringement only extends throughout the life of the patent. Once the patent term expires, the invention is in the public domain and may be made, used, sold, and/or imported freely. Infringement liability attaches regardless of the quantity or amount of the patented invention that is made, used, sold, offered for

sale, or imported. However, the quantity or amount of infringing products may be relevant in determining the damages to be awarded to the patent owner.

Infringement liability is limited territorially. Patent infringement of a U.S. patent arises only when the patented invention is practiced in the United States. The manufacture, use, or sale of the patented invention outside the United States cannot be prohibited by the patent owner, except to the extent that the patent owner also holds a patent in each of the countries where the invention is practiced.

One caveat to the foregoing is that anyone who, without the authority of the patent owner, imports a product into the United States that is prepared abroad by a process claimed in a U.S. patent infringes the U.S. process patent under the 1988 Process Patent Amendments Act (28). In other words, if an unlicensed party uses a process claimed in an unexpired U.S. patent to produce a product outside the United States and then imports that product, the party is liable for the infringement of the U.S. process patent. Thus, even though the patented process is being practiced outside the United States, the patent owner may recover damages for infringement of the process patent once the product produced by the claimed process is imported into the United States. However, the act also provides that a product will not be deemed to be made by the patented process if: 1) it is materially changed by subsequent processes; or 2) it becomes a trivial and nonessential component of another product. For example, see *Eli Lilly and Company v. American Cyanamid Company* (29) for a case concerning what constitutes a “material change” to a drug product under the applicable statute.

Standards for Proving Infringement

The issuance of a valid U.S. patent gives the patent holder the right to exclude others from making, using, selling, offering for sale, or importing the patented invention during the patent term (30). Anyone who makes, uses, sells, offers for sale, or imports the patented invention without the patent owner’s consent infringes the claims of that patent. A product may infringe a patent literally or under the Doctrine of Equivalents.

Literal infringement

To define literal infringement, the plain language of the patent claims must first be properly interpreted. This is accomplished by consideration of the ordinary meaning of the language of the claim, the patent specification, the prosecution history of the patent, and the other claims in the patent. The individual claim elements are interpreted based on the ordinary meaning of the terms used, unless it is clear from the patent specification that the inventor intended to

use a special meaning. The product or process in question is evaluated for infringement with respect to the properly interpreted claims, not the preferred embodiment set forth in the specification or any commercialized embodiment of the patented invention. Thus, with respect to claim interpretation, the words of the claims are interpreted without consideration of the product in question.

Literal infringement focuses on individual claim elements rather than on the invention as a whole. Whether a product infringes the claims of the patent depends on whether the product literally embodies each and every element of those claims (31). Each element of a claim is material and essential to the definition of the invention. If the product or process does not use even one element of the patent claim, it will not literally infringe the claims. However, the accused infringer usually cannot escape liability for literal infringement merely by adding elements that are not found in the patent claims if each element cited in the claims is found in the product or process under investigation.

Infringement under the Doctrine of Equivalents

Although the requirements of literal infringement may not be satisfied, infringement may still be found under the Doctrine of Equivalents. This doctrine is satisfied when the product in question contains elements identical or equivalent to each claimed element of the patented invention (32).

To determine whether this product possesses elements identical or equivalent to elements of the patented product, many factors may be considered. For example, if this product performs substantially the same overall function in substantially the same way, to obtain the same overall result as the claimed product, the conclusion that the element in question is equivalent is supported (33). However, this is not the only factor evaluated to analyze the doctrine of equivalents. In addition, known interchangeability of the elements used in the accused product compared with the elements of the claimed product by those skilled in the relevant art supports the conclusion that the element of the accused product is equivalent to the claimed product. On the other hand, evidence of lengthy efforts to design around the claims of the patent supports the conclusion that the accused product is not equivalent to the patented product.

The Doctrine of Equivalents has recently been severely restricted by the Court of Appeals for the Federal Circuit. In *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., Ltd.* (33a) the Court held that amendments made to patent claims during prosecution of the patent application, for reasons related to patentability, foreclose the patent holder from subsequently proving infringement using the Doctrine

of Equivalents. Furthermore, patent applicants that amend claims during patent prosecution for unknown reasons are also precluded from relying on the Doctrine of Equivalents to prove infringement with respect to the amended claims.

Prosecution history estoppel

The basic effect of the doctrine of equivalents is to allow the patent owner to expand the scope of protection afforded by the literal language of the claims. However, the doctrine of equivalents does not allow the patent owner to expand the scope of the claims without restriction. The ability of the patent owner to expand the scope of the patent claims is restrained by the prior art and also by the doctrine of prosecution history estoppel.

The prior art limits the degree to which the claims may be interpreted because the claimed invention cannot be interpreted so broadly under the doctrine of equivalents as to encompass products that were known before the patent owner invented the claimed product. In other words, the claims may not be interpreted as broadly under the doctrine of equivalents as to read on the prior art.

The doctrine of prosecution history estoppel also limits the degree to which the claims may be interpreted under the doctrine of equivalents. The doctrine of prosecution history estoppel precludes the patent owner from interpreting the claims in a manner that would encompass, within the claim, subject matter that the patent owner surrendered during prosecution of the patent application to achieve issuance of the patent. Any subject matter that the patent owner surrendered during prosecution to obtain allowance of claims made in the application of the patent cannot be reclaimed under the doctrine of equivalents.

The fact that the accused infringer may have developed the accused product completely independently and without knowledge of the patent is irrelevant in the analysis of infringement liability. An accused infringer cannot escape liability by demonstrating that the accused product was developed wholly through independent research and development.

Invalidity as a Defense to Infringement

An important defense to infringement is that the patent in question is invalid. A patent, once issued, is presumed valid. However, occasionally patents are issued that are invalid for one or more reasons. The focus of the analysis of invalidity is on the mandates of 35 U.S.C. Sections 102 and 103, which set forth the conditions of patentability (although challenges under other sections can be raised as well). In particular, when a challenger presents evidence, such as a prior art reference, which is more pertinent than

that considered by the patent examiner during the prosecution of the patent application, the burden of proving invalidity is more easily met (34).

Under Section 102(a), a patent claim is invalid if the claimed product was known or used by others in the United States, or was patented or described in a printed publication in the United States or a foreign country before the product was invented by the patentee (35). To anticipate a claim under Section 102(a), the challenger must show that the claimed product was publicly available before it was invented by the innovator, thus establishing that the patent owner was not the first to invent the claimed invention.

According to Section 102(b), a patent claim is anticipated, and therefore is invalid, if the claimed invention has been patented or described in a printed publication, or has been on sale, more than 1 year before the effective filing date of the application from which the patent was issued (36). To anticipate the claim, a single prior art reference must show each element of the claimed invention arranged as set forth in the claim. However, in the appropriate setting, a secondary reference may be used to explain or demonstrate that a primary reference anticipates a patent claim inherently.

A patent claim may also be invalidated if the challenger shows that the patent does not cite the true inventor(s) (37). A patent must list only the names of those individuals who contributed to at least one element of one or more claims. If the patent includes additional individuals who are not inventors or if the patent fails to include one or more true inventors, and the error in the naming of inventors occurred with deceptive intent, the claim(s) may be invalidated.

Section 103 can be used to invalidate patent claims even when the claimed invention is not identically disclosed or described for purposes of Section 102, if the differences between the claimed invention and prior art are such that the claimed subject matter as a whole would have been obvious at the time the invention was made (38). A determination regarding validity under Section 103 must address the following factors (39):

1. The scope and content of the prior art.
2. Differences between the prior art and the claimed invention.
3. The level of ordinary skill in the pertinent art, and
4. Secondary considerations evidencing nonobviousness.

The secondary considerations include commercial success, long-felt need, failure of others, copying, praise by persons in the industry, departure from accepted principles, and widespread recognition in the art of the invention's significance.

Remedies for Patent Infringement

Several remedies are available for patent infringement. The patent owner may obtain an injunction preventing the accused infringer from continuing the unauthorized practice of the claimed invention (40). The preliminary injunction is an important remedy because it enables the patent owner to immediately stop the activities of the infringer without the necessity of first obtaining a judgment of infringement through a lengthy litigation process.

The patent owner may also obtain monetary damages from the infringer (41), being entitled to adequate compensation for the infringement. Monetary damages may be measured in different ways to determine the amount that the patent owner can recover. For example, monetary damages may be based on the profits lost by the patent owner as a result of the infringement. In any event, the damages for the patent owner will not be less than a reasonable royalty for the practice of the invention (42).

When the infringement is shown to be willful or intentional, the patent owner may recover an increased monetary award of up to three times the amount of monetary damages shown, plus attorney's fees (43).

Infringement Exemptions for Generic Pharmaceutical Manufacturers under Hatch–Waxman Provisions

The Hatch–Waxman provisions were enacted to strike a balance between the competing interests of protecting the innovation of new drugs and providing an incentive for continued research and development of new drugs versus providing lower cost generic alternatives. To permit generic pharmaceutical manufacturers to begin the testing involved in obtaining FDA approval for a generic product, the Hatch–Waxman Act exempts the generic pharmaceutical manufacturer from patent infringement liability for certain activities. Under the Hatch–Waxman Act, it is not an act of infringement for a generic pharmaceutical manufacturer to make, use, sell, offer for sale, or import into the United States a patented drug if the act of doing so is reasonably related to the development and filing of an application for federal regulatory approval (e.g., FDA approval) (44). This exemption from the definition of infringement allows the generic pharmaceutical manufacturer to make or import and use the patented drug before the expiration of the patent, thus enabling the generic pharmaceutical manufacturer to initiate the FDA regulatory review process before the expiration of the patent. The ability to begin these activities before patent expiration is essential to obtaining FDA approval for

commercialization as of the expiration date of the patent covering the drug. Without this provision, the patent owner would effectively receive the benefit of exclusivity not only during the patent term, but also during the regulatory review of the ANDA.

The purpose of the Hatch–Waxman Act is not, however, to insulate the generic pharmaceutical manufacturer from infringement liability to such an extent that the generic pharmaceutical manufacturer is permitted to start commercialization or even to obtain FDA approval before the expiration of the patent. As a corollary to the privileges granted to the generic pharmaceutical manufacturer, the Hatch–Waxman provision also specifies that the filing of an ANDA for the purpose of obtaining FDA approval for commercialization of a generic drug before the expiration of the patent covering the drug constitutes an act of infringement for which the generic pharmaceutical manufacturer may be liable (45).

The ANDAs that can be the subject of an infringement action under this provision include a paragraph (iv) patent certification, that is, a certification asserting either that the commercialization of the generic drug for which approval is being sought will not infringe the patent covering the innovator's drug or that the patent covering the drug is invalid. As noted above, an ANDA applicant including a paragraph (iv) certification is required by law to notify the patent owner that such an application has been filed. Thus, the patent owner is ensured of receiving notice of his right to commence an infringement action under Section 271(e). If after receiving such a notice the patent owner decides to file a lawsuit for infringement based on the filing of an ANDA including the paragraph (iv) certification, the owner must file the lawsuit within 45 days of receiving the notice of the ANDA filing (46). This action by the patent owner stops the FDA regulatory review of the ANDA, and approval would not be effective until 30 months after the date the patent owner received notification of the ANDA filing (47). The law does give some discretion to the court to shorten or lengthen the 30-month period, depending on the conduct of the parties during litigation (48). The act also provides exceptions to the 30-month period, depending on the outcome of litigation. For example, if the court finds that the patent is infringed and valid (contrary to the ANDA certification) before the expiration of the 30-month period, ANDA approval cannot become effective before the expiration date of the patent (49). This is true even if the 30-month period expires before the patent. If, on the other hand, the court finds that the patent is invalid or not infringed, the ANDA approval may be made effective as of the date of the court decision (50).

These provisions can be summarized as follows. A generic pharmaceutical manufacturer seeking to obtain

FDA approval for a generic drug before the expiration of the patent covering the drug must file an ANDA including a paragraph (iv) certification asserting that the patent will not be infringed by the commercialization of the generic drug or that the patent is unenforceable or that the patent is invalid. The ANDA applicant making this certification must also notify the patent owner of the filing of the ANDA including this certification and provide the details regarding the rationale for the applicant's belief that the patent will not be infringed or is invalid. If the patent owner does not challenge the veracity of the ANDA certification, the ANDA may be approved effective immediately, and with this approval, the ANDA applicant may proceed with the commercialization of the generic drug. If, however, the patent owner does file an infringement action against the ANDA applicant within 45 days after receiving the notification of the ANDA filing, the ANDA may not be approved until 30 months from the date on which the patent owner received notification of the ANDA or the patent expires or the court determines that the patent is invalid or not infringed.

Remedies for the innovator for infringement based on an ANDA filing

If the court finds that an ANDA applicant has infringed a patent by filing the ANDA with a paragraph (iv) certification, the question remains as to what remedy is available to the patent owner. The law explicitly states that the patent owner may not obtain an injunction that would prevent the generic pharmaceutical manufacturer from making, using, selling, offering for sale, or importing the patented drug if the purpose of doing so is solely for the preparation and filing of an application for FDA approval (51). This provision prevents the patent owner from circumventing the provisions of the Hatch–Waxman Act that explicitly exclude these activities from the definition of infringement. However, an injunction can prevent the generic pharmaceutical manufacturer from commercializing the generic drug.

Monetary damages against the infringing ANDA applicant are also an available remedy. However, because the ANDA was not approved, commercial use or sales on which monetary damages could be based are highly unlikely. For the same reason, treble damages and attorney's fees based on willful infringement are, in reality, also unavailable. Perhaps the most important practical remedy for the patent owner is the fact that FDA approval of the ANDA is blocked by an infringement action. Without FDA approval, the generic drug cannot be commercialized, thus preserving the patent owner's exclusivity.

Case study: Bayer AG versus Elan Pharmaceutical Research Corporation

The Hatch–Waxman provisions are demonstrated by the case of *Bayer AG v. Elan Pharmaceutical Research Corporation* (52). Bayer AG and Bayer Corporation (“Bayer”) filed an application and was granted U.S. Patent No. 5,264,446 (the ‘446 patent) in November 1993 covering nifedipine, a drug used for the treatment of hypertension. The ‘446 patent covered nifedipine itself as a composition-of-matter claim, a method for making the drug, as well as a method of treatment using the drug. The purpose of the patent was to provide patent coverage for a sustained-release version of the drug. The patent contained 12 independent claims, and each claim specified a specific surface area (SSA) for the nifedipine crystals used in making tablets of the drug. Specifically, the ‘446 patent covered nifedipine crystals with an SSA of 1.0 to 4 m²/gram in admixture with a solid diluent to result in a sustained-release formulation of the compound.

In April 1997, Elan submitted an ANDA to the FDA seeking approval for a product they claimed was bioequivalent to Bayer’s product containing nifedipine as the active ingredient. Elan’s ANDA covered a once-daily formulation containing 30 mg of active ingredient with an SSA of no less than 5 m²/gram. Pursuant to federal regulations (53), Elan sent notice of its ANDA filing and paragraph IV certification, noting that the nifedipine it intended to manufacture and market did not infringe the claims of Bayer’s ‘446 patent. Bayer filed suit against Elan, claiming both literal infringement and infringement under the doctrine of equivalents. The District Court ruled in favor of Elan on both infringement issues, holding that Elan’s proposed sale of a nifedipine product having an SSA greater than that claimed in the ‘446 patent did not constitute literal infringement and that prosecution history estoppel prevented Bayer from contending infringement under the doctrine of equivalents (54). Bayer appealed the decision of the District Court.

The United States Court of Appeals for the Federal Circuit subsequently upheld the decision of the District Court with respect to both issues of literal infringement and infringement under the doctrine of equivalents. With regard to the issue of literal infringement, the Court of Appeals reasoned that for a product to literally infringe a patent claim, the product in question must contain each limitation of the asserted claim. In other words, the accused product must contain each element of the patented composition to constitute direct infringement of the patent. If any claimed limitation is missing in the accused product, there is no literal infringement as a matter of law. Elan presented evidence at trial that the nifedipine crystals it

intended to use in its product had an SSA equal to or greater than 5 m²/gram. The Court of Appeals held that since the ‘446 patent covered only nifedipine crystals with an SSA between 1.0 and 4.0 m²/gram, the manufacture and sale of Elan’s nifedipine product did not constitute literal infringement.

Bayer alternatively argued that the manufacture and sale of Elan’s nifedipine product constituted infringement under the doctrine of equivalents. As addressed previously, infringement under the doctrine of equivalents can occur if there is not a substantial difference between the limitations contained in the patent and the accused product. However, one caveat of the doctrine of equivalents is the theory of prosecution history estoppel under which a patentee is prevented from claiming infringement for subject matter that was clearly and unmistakably surrendered during prosecution of the patent application. In this case, there was clear evidence that the original application which eventually issued as the ‘446 patent contained claims for nifedipine with an SSA range of 0.5 to 6 m²/gram. During prosecution of the application, the patent examiner rejected Bayer’s claims as obvious over the prior art. In response to the examiner’s rejection, Bayer amended its claims to restrict the scope to only nifedipine crystals with an SSA of 1.0 to 4 m²/gram, claiming that crystals in this range provided unexpected and advantageous bioavailability and sustained-release properties. Elan argued that because Bayer had unequivocally surrendered any claims to nifedipine crystals having an SSA of greater than 4 m²/gram, Elan’s ANDA specifying crystals with an SSA of no less than 5 m²/gram did not constitute infringement under the doctrine of equivalents. The Court of Appeals agreed with Elan’s argument, reasoning that a competitor examining the entire prosecution history of the ‘446 patent would conclude that Bayer had surrendered any claim to a nifedipine drug product having an SSA greater than 4 m²/gram. As a result, the Court of Appeals held that Elan’s ANDA for generic nifedipine did not infringe the claims of Bayer’s ‘446 patent either literally or under the doctrine of equivalents. Consequently, the lawsuit instituted by Bayer was lost, and Elan’s ANDA proceeded to approval by the FDA.

Process Patents and the Hatch–Waxman Act

Process patents (i.e., patents claiming processes for making a drug compound or formulation) are not listed by the FDA in connection with approved drug formulations. Thus, process patents are not subject to the patent certification requirements of ANDAs and do not invoke the Hatch–Waxman procedures relating to

infringement outlined above. If the patent owner learns of the generic pharmaceutical manufacturer's use, sale, offer for sale, or importation of the drug or formulation prepared by the patented process, the patent owner has an action for infringement under the Patent Process Amendment Act, even if the process of making the drug or formulation is conducted outside the United States (55). The same is true if the generic pharmaceutical manufacturer uses, sells, offers for sale, or imports an intermediate useful for making the drug that is prepared by a process claimed in a U.S. patent. Thus, even though process patents do not invoke the important Hatch–Waxman provisions pertaining to infringement, the protection they provide may be invaluable particularly if the only commercially viable means of obtaining the drug is through the patented process. If this is the case, the patent owner may block the commercialization of the generic drug, not by blocking the FDA approval process, but by blocking the generic pharmaceutical manufacturer's ability to obtain the drug compound. Consequently, the protection afforded by the Patent Process Amendment Act, if used correctly, can be an important complement to the protection afforded by the Hatch–Waxman Act. As noted above, there are two important caveats to the Process Patent Amendment Act. A product imported into the United States will not be deemed to infringe on a valid process patent if: 1) the product is materially changed by subsequent processes, or 2) it becomes a trivial and nonessential component of another product. Of course, the terms “material change” and “trivial and nonessential” are vague and open to broad interpretation. The federal courts are still struggling with the meaning of this broad language and with what constitutes infringement.

MOTIVATIONS FOR CHALLENGING A PATENT

Irrespective of the potential liability for patent infringement, the motivation still remains for a generic pharmaceutical manufacturer to challenge the patent(s) covering a drug in an attempt to obtain FDA approval to market a generic version of the drug before the expiration of the innovator's patent(s). There are several incentives for challenging an innovator's patent. For example, although an issued patent is presumed to be valid, occasionally new evidence comes to light after grant that may invalidate one or more of the patent claims. If the generic pharmaceutical manufacturer obtains this evidence and discloses this to the patent owner in the notification required when filing an ANDA with a paragraph (iv) certification, the patent owner may prefer to negotiate a settlement with the generic

pharmaceutical manufacturer rather than risk the possibility of one or more of the claims being declared invalid by the court. Obviously, if the generic pharmaceutical manufacturer hopes to persuade the patent owner to accept such a compromise, the validity of the patent should be fully investigated before filing the ANDA.

If instead of reaching a settlement with the patent owner, the generic pharmaceutical manufacturer successfully challenges the patent, this company has the privilege of keeping the first generic drug on the market and has a head-start on would-be competitors. As a further incentive, the FDA regulations provide an exclusivity period for the first generic drug to enter the marketplace (56). The generic pharmaceutical manufacturer that successfully challenges the patent has an exclusivity period that extends for 180 days after either the first commercial marketing of the generic drug or the date of the first court decision holding the patent invalid or not infringed (57). During this period, no other ANDAs for the same drug may be approved. This period of exclusivity for the generic pharmaceutical manufacturer, coupled with the status of being the first generic drug on the market, can lead to the establishment of a strong market position and a return on the manufacturer's investment in challenging the patent.

CONCLUSION

Growth in the pharmaceutical industry has contributed significantly to the importance of pharmaceutical patents and the Hatch–Waxman Amendments in the United States. The importance of these issues will likely continue to increase as more and more foreign pharmaceutical companies reach toward the lucrative U.S. market for the sale of pharmaceutical products and formulations developed abroad. In the past, most small or midsize foreign pharmaceutical companies have avoided the U.S. markets because of stringent FDA regulations. As the demand for lower-cost pharmaceutical products increases, the profitable U.S. market can no longer be ignored. The balance between the rights of the patent owners and the desire for lower-cost generic alternatives to approved drugs will continue to elevate the importance of the Hatch–Waxman Amendments and will no doubt lead to significant debate as efforts begin for the revision of these provisions.

REFERENCES

1. 35 U.S.C. §101 and §100.
2. 35 U.S.C. §271(g).

3. 35 U.S.C. §154(a)(2).
4. 35 U.S.C. §156.
5. 37 C.F.R. §1.720.
6. 35 U.S.C. §156; 37 C.F.R. §1.775.
7. 35 U.S.C. §156(g)(6)(A).
8. 35 U.S.C. §156(c)(3).
9. 80F.3d 1543, 38 USPQ2d 1347 (Fed. Cir. 1996).
10. 21 U.S.C. §355(j)(5)(D)(ii). Other Time Periods Apply to Products Approved Before September 24, 1984. See 21 U.S.C. §355(j)(5)(D).
11. 21 U.S.C. §355(j)(5)(D)(ii).
12. 21 U.S.C. §355(j)(5)(D)(iii).
13. 21 U.S.C. §355(j)(5)(D)(iv).
14. 21 U.S.C. §355a.
15. 21 C.F.R. 314.53 and 21 U.S.C. §355(b)(1).
16. 21 U.S.C. §355(j)(2)(A).
17. 21 U.S.C. §355(j)(2)(A)(vii).
18. 21 C.F.R. 314.53 and 21 U.S.C. §355(b)(1).
19. *Abbott Labs r. Zenith Labs. Inc.*, 934 F.Supp 925, 36 USPQ2d 1801 (1995).
20. 35 U.S.C. §287(3).
21. 35 U.S.C. §287(3)(B)(i).
22. 35 U.S.C. §287(4).
23. 35 U.S.C. §287(4)(A)(i).
24. 35 U.S.C. §287(4)(A)(ii).
25. 35 U.S.C. §287(4)(A)(iii).
26. 35 U.S.C. §287(4)(C).
27. 35 U.S.C. §271(a).
28. 35 U.S.C. §271(g).
29. *Eli Lilly and Company v. American Cyanamid*, 82 F.3d 1568 (Fed. Cir. 1996).
30. 35 U.S.C. § 154(a)(1).
31. *Consolidated Aluminum Corp. v. Foseco Intl. Ltd.*, 716 F. Supp 316, 10 USPQ2d 1143 (N.D. Ill. 1988).
32. *Hilton Davis Chem. Co. v. Warner-Jenkinson Co., Inc.*, 62 F.3d 1512, 35 USPQ2d 1641 (Fed. Cir. 1995) (*En Banc*), *Rev'd. and Remanded*, No. 95–758 (U.S. March 3, 1997).
33. This Test was First Set Forth in *Pennwalt Corp. v. Durand-Wayland, Inc.*, 833 F.2d 931 (Fed. Cir. 1987).
- 33a. *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., Ltd.*, 234 F.3d 558 (Fed. Cir. 2000).
34. *EWP Corp. v. Reliance Universal Inc.*, 755 F.2d 898, 225 USPQ2d 20 (Fed. Cir. 1985).
35. 35 U.S.C. §102(a).
36. 35 U.S.C. §102(b).
37. 35 U.S.C. §102(f).
38. 35 U.S.C. §103.
39. *Graham v. John Deere Co.*, 383 U.S. 1 (1966).
40. 35 U.S.C. §283.
41. 35 U.S.C. §284.
42. 35 U.S.C. §284.
43. 35 U.S.C. §285.
44. 35 U.S.C. §271(e)(1).
45. 35 U.S.C. §271(e)(2).
46. 21 U.S.C. §355(j)(4)(B)(iii).
47. 21 U.S.C. §355(j)(5)(B)(iii).
48. 21 U.S.C. §355(j)(5)(B)(iii).
49. 21 U.S.C. §355(j)(5)(B)(iii)(II).
50. 21 U.S.C. §355(j)(5)(B)(iii)(I).
51. 35 U.S.C. §271(e)(3).
52. *Bayer AG r. Elan Pharmaceutical Research Corp.*, 2000 WL 572705 (Fed. Cir. 2000).
53. 21 U.S.C. §505(j)(2)(B)(ii).
54. *Bayer AG v. Elan Pharmaceutical Research Corp.*, 64 F.Supp.2d 1295 (N.D. Ga. 1999).
55. 35 U.S.C. §271(g).
56. 21 U.S.C. §355(j)(5)(B)(iv).
57. 21 U.S.C. §355(j)(5)(B)(iv) and *Mylan Pharmaceuticals Inc. v. Shalala*, 53 USPQ2d 1449 (DC DC 2000).

PEPTIDES AND PROTEINS—BUCCAL ABSORPTION

Ashim K. Mitra
Hemant H. Alur
Thomas P. Johnston

University of Missouri—Kansas City, Kansas City, Missouri, U.S.A.

INTRODUCTION

Background and Rationale

In recent years, proteins and peptides are emerging as a major class of therapeutic agents. Pharmaceutical scientists are faced with the challenges of a) selection of a suitable route of drug delivery and b) formulation of these bioengineered drugs. The most common route of protein and peptide drug delivery has been parenteral. However, this route is associated with pain on administration, resulting in poor patient compliance, and the formulation needs to be sterile. Drugs administered by the gastrointestinal route are subjected to acid hydrolysis and extensive gut and/or hepatic first-pass metabolism. Thus, these protein drugs may exhibit poor oral bioavailability. Noninvasive mucosal and transdermal routes offer effective alternative routes for systemic drug delivery. The transdermal delivery route is limited to potent, lipophilic compounds, does not provide rapid blood levels, and is less permeable than oral mucosa (1). Various absorptive mucosae have been identified and investigated for systemic drug delivery. These include nasal, ocular, pulmonary, rectal, vaginal, buccal, and sublingual. Certain associated drawbacks limit extensive use of the nasal, ocular, pulmonary, rectal, and vaginal mucosae. The buccal and sublingual routes do not have many of these limitations; hence, both routes seem attractive alternative routes for systemic drug delivery. In this article, only buccal drug delivery is addressed.

Advantages

Because of rich buccal vascularity, drugs delivered by the buccal route gain direct access to the systemic circulation and are not subject to first-pass metabolism. Also, therapeutic agents do not come in contact with the acidic digestive fluids secreted by the gastrointestinal tract.

Relative to the nasal and rectal routes, the buccal mucosa has low enzymatic activity, and drug inactivation owing to biochemical degradation is not as rapid and extensive (2).

Excellent accessibility to the buccal mucosa makes application of the dosage form painless, precisely located, and easily removable without discomfort at the end of the application period. The oral cavity consists of a pair of buccal mucosae. Thus, a drug-delivery system can be applied at various sites either on the same mucosa or, alternatively, on the left or right buccal mucosa on different applications. This is particularly advantageous if the delivery system contains a drug or excipient that mildly and reversibly damages or irritates the mucosa.

A buccal drug-delivery system is applied to a specific area on the buccal membrane. Moreover, the delivery system can be designed to be unidirectional in drug release so that it can be protected from the local environment of the oral cavity. It also permits the inclusion of a permeation enhancer/protease inhibitor or pH modifier in the formulation to modulate the membrane or the tablet-mucosal environment at or near that particular application site. Although the irritation is limited to a well-defined area, the systemic toxicity of these enhancers/inhibitors and modifiers can be reduced. The buccal mucosa is well-suited for this type of modification because it is less prone to irreversible damage (3). In the event of drug toxicity, the delivery of drugs can be terminated promptly by removal of the dosage form.

In addition, the buccal route may be useful for unconscious patients and in patients who have recently undergone surgery or have experienced upper gastrointestinal tract disease that would affect oral drug absorption.

Disadvantages

The surface area available for absorption in the buccal mucosa is much smaller than the gastrointestinal, nasal, rectal, and vaginal mucosae. The buccal mucosa is continuously bathed by saliva, and the secreted saliva lowers the drug concentration at the absorbing membrane. These two factors, along with the permeability coefficient of the drug, affect the overall absorption rate by this route. In addition, the buccal mucosa is less permeable than any of the mucosae noted above.

Involuntary swallowing of saliva containing dissolved drug or swallowing the delivery system itself would lead to a major loss of drug from the site of absorption. Talking, eating, and drinking affect the retention of the delivery system and therefore may constitute limitations associated with this route of drug administration (1). In addition, there is a risk of choking on the dislodged drug-delivery device.

Taste, irritancy, and allergenicity also may limit the number of drugs that can be delivered by the buccal route.

ANATOMY AND PHYSIOLOGY OF THE ORAL MUCOSA

Structure

The oral mucosa is anatomically divided into three tissue layers (Fig. 1) (4). These three layers are the 1) epithelium, 2) basement membrane, and 3) connective tissues.

Epithelium

The epithelium consists of approximately 40–50 layers of stratified squamous epithelial cells. The epithelial cells originate from a layer of basal cells, which are cuboidal in shape, undergo continuous mitosis, and move to the surface. As the cells migrate to the surface through the intermediate layers, they differentiate and become larger, flattened, and surrounded by an external lipid matrix (membrane-coating granules). This external lipid matrix

determines the drug permeability of the tissue. Although gingiva (gum) and the hard palate are keratinized, areas such as buccal, sublingual, and the soft palate are nonkeratinized. The thickness of buccal epithelium varies with location and typically ranges from 500 to 800 μm in humans, dogs, and rabbits. The estimated cell turnover time is 5–6 days (5). In addition, the buccal epithelium is also characterized by the presence of intercellular gap junctions.

Basement membrane

The basement membrane (BM) is a continuous layer of extracellular materials and forms a boundary between the basal layer of epithelium and the connective tissues of the lamina propria and the submucosa. The BM can be subdivided into the a) lamina lucida, b) lamina densa, and c) a sublayer of fibrous material. The functions of the BM include providing 1) adherence between epithelium and underlying connective tissues, 2) mechanical support for epithelium, and 3) a barrier to the passage of cells and some large molecules.

Connective tissues

Connective tissues consist of lamina propria and submucosa, if present. The lamina propria is a continuous sheet of connective tissue composed of blood capillaries and nerve fibers serving the oral mucosa.

Vascular drainage from the oral mucosa is principally by way of the lingual, facial, and retromandibular veins. These veins open into the internal jugular vein and thus avoid first-pass metabolism.

The buccal mucosae from monkeys, apes, dogs, pigs, and rabbits possess physiology very similar to that of human buccal mucosa (5).

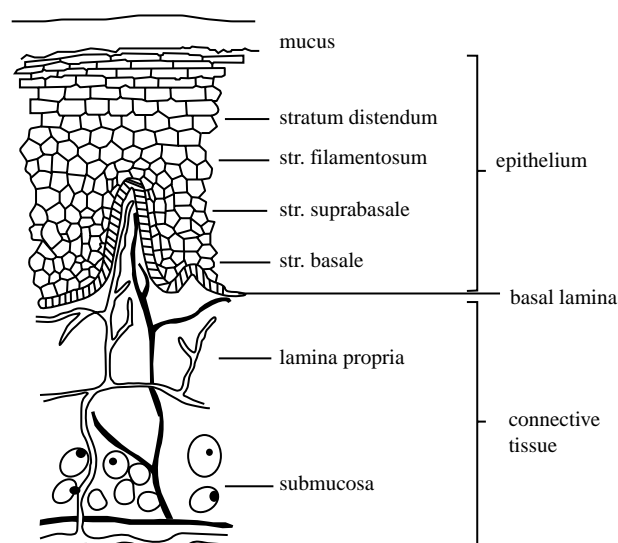


Fig. 1 Schematic diagram showing the principal components of oral mucosa. (From Ref. 4.)

Permeability

Permeability barriers

The permeability of buccal mucosa lies somewhat between the skin epidermis and intestinal mucosa.

Epithelium: The predominant barrier to drug diffusion resides approximately within the outermost one-third of the epithelium. This is true of both keratinized and nonkeratinized epithelia. Therefore, keratinization is unlikely to offer major resistance to buccal permeation.

Membrane coating granules (MCG): MCGs are spherical or oval organelles (100–300 nm in diameter) found both in keratinized as well as in nonkeratinized epithelia but are different with regard to composition in both epithelia. MCGs discharge their contents into the intercellular space and thus form the permeability barrier.

Permeant factors: The permeation of a drug molecule across the buccal mucosa is dependent on the following.

1. Molecular size—for hydrophilic substances, as molecular weight and molecular size/radius ascends, permeability typically diminishes. Small molecular weight permeants (MW < 100 Da) are rapidly transported through buccal mucosa.
2. Lipid solubility—for non-ionizable compounds, as the lipophilicity rises, the drug permeability typically increases. To maximize the absorption rate, a drug should be available in the salivary film at its solubility limit.
3. Ionization—for ionizable drugs, maximal permeation occurs at the pH at which ionization is least, i.e., where the drug is predominantly in the unionized form (6, 7).

The rate of drug absorption for the transcellular route is pH-dependent. Such dependency results from the fact that the membrane/aqueous partition coefficient for an ionizable drug is pH-dependent.

Basement membrane (BM): The BM has an enormous surface area compared with the epithelium owing to connective tissue papillae, which may affect the effective diffusional pathlength.

Mechanism of Drug Transport

The major pathway of drug transport across buccal mucosa seems to follow simple Fickian diffusion (8, 9). Passive diffusion occurs in accordance with the pH-partition theory. Considerable evidence also exists in the literature regarding the presence of carrier-mediated transport in the buccal mucosa (10, 11). Examination of the equation for drug flux is shown by Eq. 1:

$$J = \frac{DK_p}{h} \Delta C_e \quad (1)$$

where J = drug flux, D = diffusivity, K_p = partition coefficient, ΔC_e = concentration gradient, and h = diffusional pathlength shows that the flux may be increased by decreasing the diffusional resistance of the membrane by making it more fluid, increasing the solubility of the drug in the saliva immediately adjacent to the epithelium, or enhancing lipophilicity through prodrug modification. Because of the barrier properties of the tight buccal mucosa, the rate-limiting step is the movement of drug molecules across the epithelium.

Two pathways of permeation across buccal mucosa are transcellular, in which the passage of drug occurs through the individual cells of the mucosa, and

paracellular, in which the passage of drug occurs through intercellular junctions of the mucosa. Permeability coefficients typically range from 1×10^{-5} to 2×10^{-9} cm/s for oral mucosa (12). The pathway of drug transport across oral mucosa may be studied using: 1) microscopic techniques using fluorescent dyes (12), 2) autoradiography (5), and 3) confocal laser scanning microscopic procedures (13).

FACTORS AFFECTING SYSTEMIC ORAL MUCOSAL DELIVERY

Membrane Factors

Regional differences in both permeability and thickness affect both the rate and extent of drug reaching the systemic circulation (14). Keratinization and composition, although not major factors, of the various oral mucosae affect systemic mucosal drug delivery. Additional factors such as absorptive membrane thickness, blood supply, blood/lymph drainage, cell renewal rate, and enzyme content will also govern the rate and extent of drug absorption into the systemic circulation.

Environmental Factors

Saliva

A major portion of saliva is composed of water (99%) and has a pH of 6.5–7.5 depending on the flow rate and location (15). An increase in the salivary flow rate leads to the secretion of watery saliva. Stimulated salivary secretion affects the film thickness and aids in easy migration of test compounds from one region of the mouth to another. Salivary pH is also important because passive diffusion of unionized drug is the major mechanism of oral absorption (16, 17).

Salivary glands

Drug-delivery systems, therefore, should not be placed either over a duct or adjacent to a salivary duct because this may dislodge the retentive system or may result in excessive washout of the drug or rapid dissolution/erosion of the delivery system, making it difficult to achieve high local drug concentrations. Also, if a retentive system is placed over salivary ducts, the reduced salivary flow rate may produce less/no mucus that is required for proper attachment of a mucoadhesive delivery device.

Movement of the oral tissues

Talking, eating, and swallowing may cause some mouth movement leading to dislodgment of the delivery device (1). The movement of the tongue may also influence the delivery of drugs from a mucoadhesive, retentive system owing to the tongue swiping across the dosage form and adjacent tissues as well as to induction of suction pressures from the tongue compressing against the hard palate.

Dosage Form Design Considerations

Overview

A mucoadhesive buccal drug delivery system should

1. be convenient to apply and unobtrusive when in place,
2. not incorporate a bitter-tasting drug,
3. have a smooth surface rather than a textured surface,
4. preferably achieve unidirectional release of the drug, and
5. use excipients (both diluents and the mucoadhesive polymers) that do not irritate or damage the mucosa, that are nontoxic, and that do not stimulate salivary secretion.

The *size* of the delivery system varies with the type of formulation, i.e., a buccal tablet may be approximately 5–8 mm in diameter, whereas a flexible buccal patch may be as large as 10–15 cm² in area. Mucoadhesive buccal patches with a surface area of 1–3 cm² are most acceptable (18). It has been estimated that the total amount of drug that can be delivered across the buccal mucosa from a 2-cm² system in 1 day is approximately 10–20 mg (12). The *shape* of the delivery system may also vary, although for buccal drug administration, an ellipsoid shape appears to be most acceptable (18). The *thickness* of the delivery device is usually restricted to only a few millimeters. The *location* of the delivery device also needs to be considered. A mucoadhesive retentive system is preferred over a conventional dosage form. A bioadhesive buccal patch would appear to be the most appropriate delivery system because of its flexibility and the area of the buccal mucosa available for its application. The maximal duration of buccal drug retention and absorption is approximately 4–6 h because food and/or liquid intake may require removal of the delivery device.

BIOADHESION

The word *bioadhesion* can be defined as the ability of a material (synthetic or natural) to “stick” (adhere) to

a biological tissue for extended periods (18). The phenomenon of bioadhesion can be visualized as a two-step process. The first step involves the initial contact between polymer and the biological tissue. The second step is the formation of secondary bonds owing to noncovalent interactions.

Biomembrane Characteristics

Oral mucosae are covered with mucus that serves as a link between the adhesive and the membrane. Mucin is a polyelectrolyte under neutral or slightly acidic conditions because of the terminal sialic acid residues having a pK_a value of 2.6 (19). At physiological pH (7.4), the mucin molecule is polyanionic, which contributes to bioadhesion.

Adhesive Characteristics

A variety of polymers including water-soluble and water-insoluble and ionic and non-ionic hydrocolloids and water-insoluble hydrogels can be used in bioadhesive systems (20). Factors affecting the bioadhesive properties of the polymer include the following.

Molecular weight and polymer conformation

In general, the adhesive strength of a polymer increases with molecular weights greater than 100,000 (21). The molecule must have adequate length to allow chain interpenetration into the mucus layer. However, the size and conformation of the polymer molecule play an important role as well (22).

Cross-linking density of the polymer

The strength of mucoadhesion decreases with an increase in cross-linking as this leads to a decrease in the polymer's diffusion coefficient (23) and chain segment flexibility and mobility (which in turn reduces interpenetration).

Charge and ionization of the polymer

Anionic polymers provide better efficiency than do cationic or uncharged polymers with respect to both adhesiveness and toxicity (24). Also, polymeric adhesives with carboxyl groups are preferred over those with sulfate groups (25).

Concentration of the polymeric adhesive

In general, the more concentrated the polymeric adhesive becomes, the more the bioadhesive strength diminishes.

The coiled molecules become solvent-poor in a concentrated solution, which, in turn, decreases the available chain length for interpenetration into the mucus layer. Therefore, a critical concentration of the polymeric adhesive is required for optimum bioadhesion (26).

pH of the medium

pH influences the charge on the surface of the mucus and the polymer (27). Charge density on the surface of mucus will vary with pH owing to the differences in dissociation of the functional groups on the carbohydrate and amino acid moieties.

Hydration of the polymer

Swelling affects bioadhesion (28), although an increase in the degree of swelling does not always result in an increase in the bioadhesive strength. However, a high water activity is required to hydrate the mucoadhesive component to expose the bioadhesive site(s) for secondary bond formation, to expand the gel to create pores of sufficient size, and to mobilize all flexible polymer chains for interpenetration. A critical degree of hydration of the mucoadhesive polymer is needed for optimum bioadhesion (20). A greater degree of hydration lowers the adhesive strength owing to the formation of a slippery mucilage.

Theories of Bioadhesion

In general, both physical and weak chemical bonds are responsible for mucoadhesion. Physical/mechanical bond formation can be explained as the entanglement of the adhesive polymer and the extended mucin chains. When this diffusion is mutual, it leads to maximum bioadhesive strength.

Attaching bonds may be either primary owing to covalent bonding or secondary owing to electrostatic, hydrogen, or hydrophobic bonding. Electrostatic or hydrogen bonding results primarily because of hydroxyl ($-\text{OH}$), carboxyl ($-\text{COOH}$), sulfonate ($-\text{SO}_3\text{H}$), and amino ($-\text{NH}_2$) groups. Several theories of bioadhesion have been proposed, i.e., wetting, diffusion, electronic, fracture, and adsorption. The mechanism of bioadhesion appears to be best explained by a combination of the wetting, diffusion, and electronic theories.

Measurement of Bioadhesion

Measurement of bioadhesion not only helps in screening the candidate polymer but also assists in studying the mechanism of bioadhesion. However, performance of

the final dosage form containing the polymer and the drug is the best test for bioadhesion.

In vitro measurements

Measurement of either tensile or shear stress is the most commonly used in vitro method to measure bioadhesion. All in vitro measurements provide a rank order of bioadhesive strength for a series of candidate polymers. Measurement of tensile strength involves quantitating the force required to break the adhesive bond between the test polymer and a model membrane. This method typically uses a modified balance or tensile tester. A section of freshly excised rabbit stomach tissue with the mucosal side exposed is secured on a weighed glass vial and placed in a beaker containing USP-simulated gastric fluid. Another section of the same tissue is secured onto a rubber stopper with a vial cap with the mucus side exposed (Fig. 2). A small quantity of the test polymer is placed between the two mucosal tissues. The force required to detach the polymer from the tissue is then recorded (29). Measurement of shear strength involves quantitating the force that causes the polymer to slide in a direction parallel to the plane of contact between the polymer and the mucus. This method uses a glass plate suspended from a microbalance on which the test polymer is coated (Fig. 3). This plate is then dipped in a temperature-controlled mucus sample. The force required to pull the plate out of the mucus sample is determined under constant experimental conditions (22).

Additional in vitro methods include adhesion weight (30), fluorescent probe (24), flow channel (31), mechanical spectroscopic (32), falling film (33), colloidal gold staining

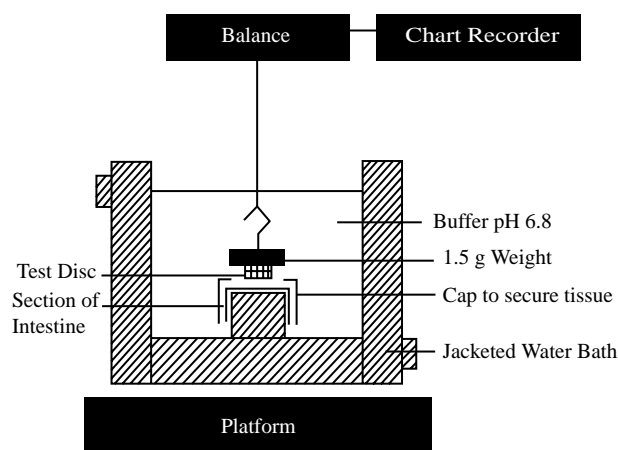


Fig. 2 Schematic diagram showing the apparatus and the setup for assessing the tensile strength. (From Ref. 93.)

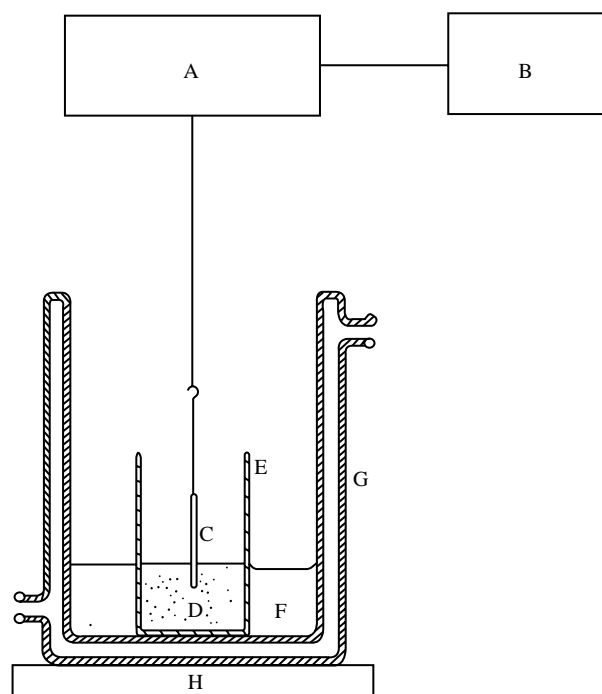


Fig. 3 Schematic diagram showing the apparatus and the set up for assessing the shear strength. Key: (A) Microforce balance; (B) Chart recorder; (C) Glass plate (side on); (D) 1 ml Homogenized mucus; (E) Glass vial; (F) Water; (G) Water jacket at 20°C; (H) Platform moving in vertical direction. (From Ref. 22.)

(34), viscometric method (35), thumb test (36), adhesion number (36), and electrical conductance (36).

In vivo measurements

In vivo methods are relatively few and measure the residence time of bioadhesives at the application site (36). Techniques such as γ -scintigraphy, perfused intestinal loop and radiolabeled transit studies using ^{55}Cr -labeled bioadhesive polymer (37), and $^{99\text{m}}\text{Tc}$ -labeled polycarbo-phil (38) have been used for this purpose.

Bioadhesive Polymers in Buccal Drug Delivery

A variety of water-soluble and water-insoluble polymers of both synthetic and natural origin (20) have been studied as bioadhesives.

Overview

Bioadhesive polymers are used mainly to overcome the short residence time of drug and the dosage form to improve localization of the drug and to achieve controlled or sustained release of the drug. Bioadhesive polymers can be divided into three broad categories: 1) "wet" adhesives,

i.e., polymers that become sticky on hydration, 2) polymers that are electrostatic in nature and adhere primarily owing to nonspecific and noncovalent interactions, and 3) polymers that can bind to a specific site on the cell surface (39).

An ideal bioadhesive should

- be nontoxic, nonabsorbable, and nonirritating to the mucus membrane,
- form a strong noncovalent bond with the mucin-epithelial cell surfaces,
- allow easy incorporation of drug and should not offer hindrance to drug release, and
- not decompose on storage or during the shelf life of the dosage form.

Some of the other desirable characteristics of the polymer have been presented previously in *Bioadhesion*.

Drug release from soluble polymers is accompanied by the gradual erosion-type dissolution of the polymer. Therefore, polymer dissolution/drug diffusion may be the overall hybrid mechanism of release. Drug release from nonsoluble hydrogels generally follows Fickian or non-Fickian diffusion kinetics (40). The mechanism of release may be determined by modeling the drug release (first 60%) to the empirical equation (Eq. 2),

$$\frac{M_t}{M_\infty} = kt^n \quad (2)$$

where M_t/M_∞ = fraction of drug released, k = kinetic constant, t = time, and n = diffusional exponent. The mechanism of drug release may be Fickian diffusion when the value of $n = 0.5$; anomalous (non-Fickian) transport when $n = 0.5 < n < 1.0$, and case-II transport when $n = 1.0$. A value of n greater than 1 signifies super case-II transport as the mechanism of drug release (41).

DOSAGE FORMS

Several bio/mucoadhesive dosage forms have been developed, and mechanism(s) of bioadhesion have been delineated.

Buccal Dosage Forms

Buccal mucosa presents a relatively smooth and immobile surface for the placement of a bioadhesive dosage form. The amount of drug that can be incorporated is limited because of the size limitation of the buccal dosage form. In general, a drug with a daily requirement of 25 mg or less is suitable for buccal delivery. Drugs with a short half-life, requiring sustained/controlled delivery, or exhibiting poor

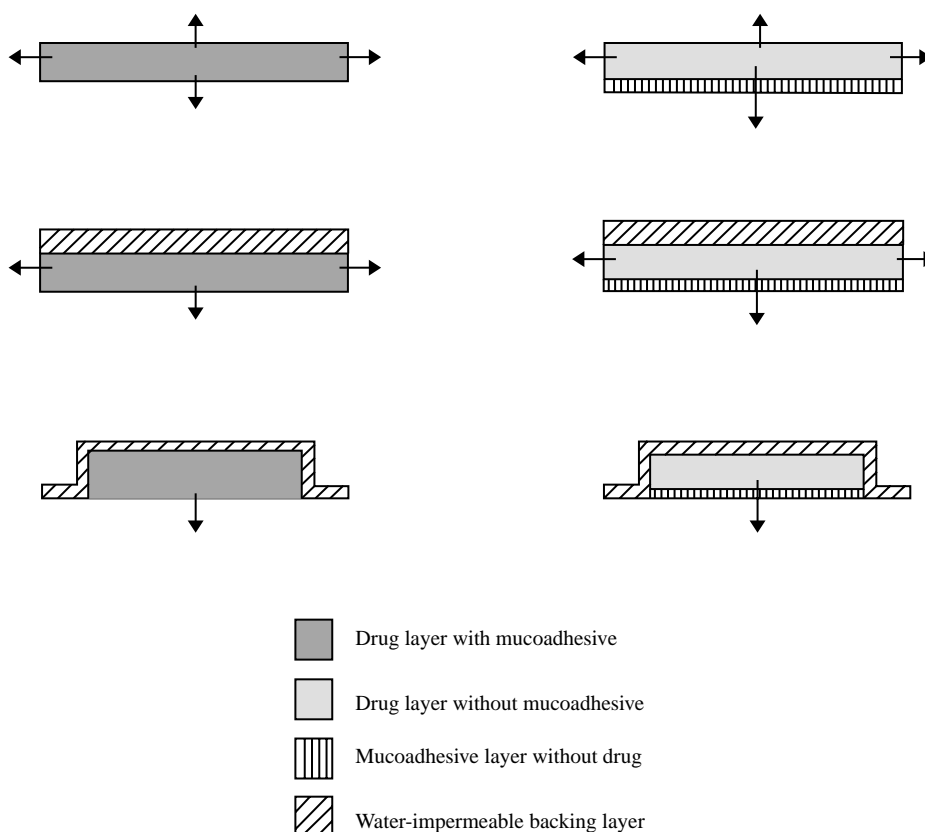


Fig. 4 Schematic diagram showing the geometric designs of buccal delivery devices. (From Ref. 91.)

aqueous solubility and drugs that are sensitive to enzymatic degradation may be delivered successfully across the buccal mucosa. The dosage forms developed for this purpose include tablets, adhesive patches, adhesive gels, and adhesive ointment. Adhesive tablets and patches can be formulated to release the drug either unidirectionally or multidirectionally by varying the extent and permeability of the backing (Fig. 4).

Formulation Development of Buccal Dosage Forms

Novel dosage forms such as adhesive tablets, patches, gels, and ointments have been developed primarily for systemic delivery of therapeutic agents. These dosage forms are also capable of providing sustained drug delivery.

Buccal dosage forms can be of 1) reservoir type and 2) matrix type.

Reservoir type

Drug formulations of the reservoir type are surrounded by a polymeric membrane, which controls the release

rate. Reservoir systems present a constant release profile, provided 1) the polymeric membrane is rate limiting and 2) an excess amount of drug is present in the reservoir.

Matrix type

Drug is uniformly dispersed in the polymer in matrix-type systems, and drug release is controlled by the matrix. Drug molecules dispersed in the polymer have to dissolve in the medium and then diffuse through the polymer network. Therefore, a drug dispersion and drug-depletion zone always exists in the matrix. A thin hydrodynamic diffusion layer also exists at the interface of the drug and the matrix. A matrix system may result in a constant release profile only at early times when the drug-depletion zone is rather insignificant.

The parameters that determine the release rate of a drug from a delivery device include polymer solubility, polymer diffusivity, and thickness of the polymer diffusional path, and the drug's aqueous solubility, partition coefficient, and aqueous diffusivity. Finally, the thickness of the hydrodynamic diffusion layer, the amount

of drug loaded into the matrix, and the surface area of the device all affect the drug's release rate.

Buccal adhesive tablets

Adhesive tablets may be either monolithic or multilayered devices. Monolithic tablets can be prepared by conventional techniques of either direct compression or wet granulation. These tablets provide the possibility of holding large amounts of drug. Using either compression or spray coating, a partial coating of every face except one that is in contact with the mucosa with a water-impermeable material such as cellophane, hydrogenated castor oil, Teflon, ethyl cellulose, etc., may cause unidirectional drug release. Multilayered tablets may be prepared by adding each formulation ingredient layer by layer into a die and by compressing it on a tablet press (Fig. 4). These tablets can be designed to deliver drugs either systemically or locally. For multilayered tablets, incorporation of the drug into the adhesive layer, which is immediately adjacent to the mucosal surface, may aid in optimizing bioadhesion.

Buccal adhesive patches

Adhesive patches may also be monolithic or multilayered devices of the reservoir or matrix type for either systemic or local drug delivery. Two primary types of manufacturing processes are usually used to prepare adhesive patches. These include solvent casting and direct milling (with or without a solvent). The intermediate product is a sheet from which patches are punched. A backing is then applied to control the direction of drug release and to minimize deformation and disintegration of the device during residence in the mouth. Preparation of adhesive patches by the solvent-casting method involves casting of appropriately prepared aqueous solutions of either polymer (for drug-free patches) or a drug/polymer mixture onto a backing layer sheet mounted on a stainless steel plate by means of a frame. Drying may then be performed by perfusing with a thermostated stream of water or by air drying. The temperature is typically selected based on the excipients used in the formulation. On complete drying, the laminate may be cut into the desired shape and size

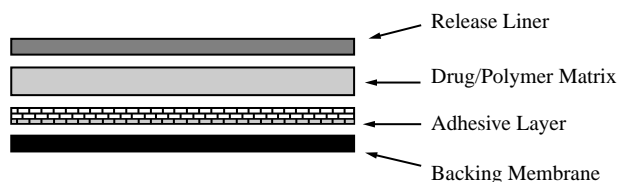


Fig. 5 Schematic diagram showing the design of mucoadhesive buccal patch. (From Ref. 81.)

using a suitable punch and a die set. Preparation of adhesive patches by direct milling is done by homogeneously mixing the drug and the bioadhesive, with or without the aid of a solvent, using a two-roll mill. The polymer/drug mixture may then be compressed to its desired thickness, and patches of appropriate size may be cut or punched out. The polymer/drug mixture prepared with a solvent may require an additional drying step afforded by air or oven drying (Fig. 5).

Design of Experiments

A formulation may be evaluated for both in vitro and in vivo release and mucosal permeation by designing appropriate experiments.

In vitro release/permeation studies

In vitro release studies may be designed depending on the shape and application of the dosage form because no standard method is available. A survey of the literature indicates that the apparatus used for the release study varied from a typical USP type I/II apparatus to a rotating basket immersed in a beaker with agitation conditions being rotational (50–250 rpm) to magnetic stirring or mechanical shaking (39). A variety of testing media have been used such as distilled water, chloroform, phosphate buffer, saline, a mixture of methanol, and water, etc. (39). In vitro release studies provide valuable information regarding the behavior of the delivery device and the mechanism of drug release from the delivery device.

In vitro permeation studies can be conducted using a glass diffusion cell with the buccal tissue mounted between the two halves of the cell, which may be filled with constantly stirred buffer solutions. Buccal mucosa may then be excised from the canine or porcine or rabbit cheek immediately after sacrifice. Such permeation studies may provide meaningful results on the simultaneous processes of drug transport and metabolism in the tissue. Attention must be given to the viability of the excised tissue under in vitro conditions. Electrophysiologic characterization appears to be a valuable tool to indicate the viability of the tissue after excision. To retain tissue viability, continuous bubbling with O_2/CO_2 mixtures and the addition of glucose to the buffer media may be useful. Viability can be determined using MTT assay and confocal imaging of vital staining (42). Alternatively, cell culture technique can be used, where by several parameters can be varied to arrive at an optimum delivery device. This technique also helps in understanding the possible mechanism of drug transport and therefore may provide strategies for modification of either the drug molecule or the drug-delivery device (43).

In vivo absorption studies

Animal studies: Only a small number of absorption studies have been conducted in animals. This may be because buccal administration of drugs to animals is difficult and often produces artifactual results (44). Animals selected as models should be representative of human absorption, tissue, enzymes, degree of keratinization, etc. Conscious or anesthetized animals may be used for absorption studies. Conscious animals should be preconditioned by administering placebos for several consecutive days. The experimental dosage form may then be administered/applied to the buccal mucosa. Absorption may be followed by monitoring plasma and/or urinary drug concentrations as well as the amount of drug remaining in the dosage form after the experiment.

Human studies: Human studies are conducted after the dosage form has been optimized with respect to release of the drug, shape of the dosage form, amount of the drug to be incorporated, and completion of preclinical toxicity studies in animals. Human volunteers may be asked to place the dosage form on the buccal mucosa, and the absorption process may be followed by monitoring the amount of drug remaining in the dosage form after certain time intervals and/or by sampling biological matrices (blood, urine, etc.) with appropriate pharmacokinetic/pharmacodynamic analyses. The amount of drug remaining in the dosage form may reflect absorption only if the absorption step is not rate limiting, and an attempt to prevent swallowing a drug/saliva solution has been made (45).

Evaluation of toxicity and irritation

Irritation is very subjective and may differ widely from treatment to control subjects. Most irritation occurs as a result of penetration enhancers. Evaluation of toxicity and irritation should be concerned with: 1) mucosal tissue irritation, 2) extent of damage to the mucosal cells, and 3) rate of recovery.

Mucosal tissue irritation: Irritation is a complex phenomenon involving interaction among the solution properties of the vehicle, mucosal transport, biological transport, and local drug disposition. To date, no definite relationship has been established between the structure of a penetration enhancer, for example, and the degree of irritation it may cause following buccal application. However, a relationship between the pK_a value of an ionizable compound (benzoic acid derivatives) and irritation as measured by the degree of erythema has been reported (46, 47). Azacycloalkanone enhancers demonstrated more irritancy with alkyl than those with alkenyl chains (48). In general, it would appear that the

most effective penetration enhancers induce the greatest degree of irritation to mucosal tissues.

Extent of Damage to Mucosal Cells: Permeation enhancement implies possible alteration of the protective permeability barrier either by: 1) an increase in the fluidity of intercellular lipids (relatively nontoxic) and/or 2) extraction of intercellular lipids or denaturation of cellular proteins (much more damaging/toxic). Therefore, it is imperative that the permeation enhancer: 1) exert a reversible effect, 2) not be systemically absorbed, and 3) not cause cumulative toxicity or permanent changes in the barrier properties. Application of up to 1% sodium lauryl sulfate or cetylpyridinium chloride to the ventral surface of the tongue of dogs resulted in desquamation, widening and separation of keratin (49). The buccal mucosa of rabbits treated with 0.5% sodium deoxycholate or 0.1% sodium lauryl sulfate demonstrated loss of surface epithelial cells (50). Sodium taurocholate and lysophosphatidyl chloride increased buccal insulin absorption in dogs with no mucosal irritation (51).

Methods used to assess membrane damage: Several methods are commonly used to estimate the degree of damage to biological membranes induced by various permeation enhancers. The following methods are a partial listing:

1. morphological examination by scanning or transmission electron microscopy (52),
2. morphological examination using light microscopy and appropriate staining, e.g., hematoxylin and eosin (H&E) (53),
3. determination of the extent of hemolysis caused by a permeation enhancer (54),
4. determination of the release of cellular constituents, e.g., lactate dehydrogenase (LDH) (55),
5. measurements of the changes in the electrical resistance of the membrane,
6. measurements of the changes in the permeability to various markers, e.g., insulin, mannitol, and FITC-dextran, and
7. measurements of changes in cilia movement, e.g., with nasal mucosa.

Rate of recovery of mucosal membranes: The rate of recovery is generally inversely related to the extent of membrane damage, i.e., a greater and more rapid recovery from permeation enhancers that induce minimal damage such as acylcarnitines (56) and sodium glycocholate compared with enhancers such as sodium deoxycholate (57) and polyoxyethylene-9-lauryl ether. The permeability of the tight junction is sensitive to the extracellular calcium concentration. Resealing of tight junctions has been shown to be accelerated if there is a

high extracellular calcium concentration rather than an elevated cytoplasmic calcium concentration (58).

Miscellaneous toxicity concerns: Additional toxicity concerns include interference with normal metabolism and function of mucosal cells, e.g., water absorption by these cells (59). The unconjugated bile acids are known to block amino acid metabolism and glucose transport (60). There is a possibility of biotransformation of these enhancers to toxic or carcinogenic substances by hepatic monooxygenases. Absorption of permeation enhancers into the systemic circulation can also cause toxicity, e.g., azone and hexamethylene lauramide (61), which are absorbed across skin. Moreover, changes in membrane fluidity may alter the activity of membrane-bound transport proteins and enzymes.

BUCCAL DRUG DELIVERY OF PEPTIDES AND PROTEINS

Buccal drug delivery avoids acid- and enzyme-mediated degradation and hepatic first-pass metabolism. However, the bioavailability of therapeutic polypeptides and proteins is generally very low (<5%) owing to low lipid solubility

and their inherent larger molecular weight compared with conventional small molecules. Degradation of proteins and peptides by enzymes such as aminopeptidases, carboxypeptidases, several endopeptidases, and esterases is also a reason for their low bioavailability. The premise of passive transport as the mechanism of peptide and protein absorption across the buccal mucosa is widely accepted. Endocytotic processes are not apparent in buccal epithelium. No active or carrier-mediated peptide transport systems are present in the buccal epithelium except those responsible for the absorption of a few amino acids such as glutamic acid (anionic) and lysine (cationic) (62). In general, the various aspects addressed thus far can be applied for buccal drug delivery of peptides and proteins (Table 1). It can be noted from the data presented in Table 2 that in general as the molecular weight of the peptide increases, the bioavailability decreases. This suggests that the peptides and proteins, being hydrophilic and globular in nature, are transported by paracellular route.

Biological Activity

Biological activity is the most important concern with the delivery of therapeutic peptides and proteins. Suscepti-

Table 1 Chronological survey of in vivo experiments on buccal delivery of peptides and proteins

Author(s)	Year (REF)	Peptide (in vivo model)
Dillon et al.	1960 (63)	Pitocin (human)
Miller	1973 (64)	Oxytocin (human)
Dawood et al.	1980 (65)	Oxytocin (human)
Ishida et al.	1981 (66)	Insulin (dog)
Anders et al.	1983 (67)	Protirelin (human)
Schurr et al.	1985 (68)	Protirelin (human)
Aungst and Rogers	1988 (69)	Insulin (rat)
Aungst et al.	1988 (70)	Insulin (rat)
Nakada et al.	1988 (71)	Calcitonin (rat)
Oh and Ritschel	1988 (51)	Insulin (rabbit)
Ritschel et al.	1988 (72)	Insulin (dog)
Ho and Barsuhn	1989 (73)	Protirelin, oxytocin (dog)
Wolany et al.	1990 (74)	Octreotide (dog)
al-Achi and Greenwood	1993 (75)	Insulin (rat)
Heiber et al.	1994 (76)	Calcitonin (dog)
Bayley et al.	1995 (77)	Recombinant human interferon- α B/D hybrid (rat, rabbit)
Gutnaik et al.	1996 (78)	Glucagon-like peptide I (human)
Nakane et al.	1996 (79)	Leuteinizing hormone-releasing hormone (dog)
Hoogstraate et al.	1996 (80)	Buserelin (pig)
Li et al.	1997 (81)	Thyrotropin-releasing hormone (rat)
Li et al.	1997 (54)	Oxytocin (rabbit)
Alur et al.	1999 (82)	Calcitonin (rabbit)

Table 2 Data from in vivo experiments on buccal delivery of peptides and proteins

Peptide/protein	MW (Da)	Species				
		Rat	Rabbit	Dog	Pig	Human
TRH	362	—	—	—	—	4%
Oxytocin	1007	—	0.1%	—	—	<10%
LHRH	1182	—	—	0.41% (alone) 0.34–1.62% (with enhancers)	—	—
Buserelin (analog of LHRH)	≈1182	—	—	—	1% (alone) 5% (with enhancers)	—
Calcitonin	3432	—	16 and 37%	550 IU over 6 h	—	—
GLP-I	4169	—	—	—	—	7% (7–36 amide fragment)
Insulin	5808	0% (alone) 25% (with enhancers)	0% (alone) 5% (with enhancers)	0% (alone) 0.5% (with enhancers)	—	0–4%
IFN α -B/D hybrid	19,000	<1%	<1%	—	—	—

bility of these molecules to denaturation by various manufacturing processes may seriously limit the number of methods that can be used in the fabrication of delivery systems. Important process variables such as temperature, pressure, and exposure to organic solvents, etc., during manufacturing need to be considered. The formulation strategies presented above can be applied to the development of peptide and protein formulation in general, or they can be modified according to special needs.

Temperature

High temperatures can break native S—S bonds and form new S—S bonds that can “lock” the protein into a denatured configuration. Low pH, sodium dodecyl sulfate, Tween 80®, chaotropic salts, and exogenous proteins have been used to protect proteins from thermal inactivation (83). Ethylene glycol at 30–50% was used as a protectant of antiviral activity of β -Interferon preparations (84). Human serum albumin was used in recombinant human interferon- $\beta_{\text{ser-17}}$, which resulted in increased thermal stability (47). Water-soluble polysaccharides such as

dextrans and amylose and point-specific (site-directed) mutagenesis (85) have also been used to increase thermal stability of therapeutic proteins and peptides.

Pressure

Proteins are not very sensitive to pressure changes, and only at large values of pressure do they exhibit conformational changes observed when denatured by heat or changes in solution pH (86). A model enzyme (protein), namely, urease, did not lose much of its activity until the compaction pressure exceeded 474 mPa, above which 50% of the relative activity was lost (87).

Pharmacokinetics and Pharmacodynamic Responses

Pharmacokinetics and pharmacodynamic responses have to be evaluated separately. Initial consideration of a drug candidate for buccal delivery may be its low biological half-life ($t_{1/2}$) possibly owing to high first-pass metabolism and gastrointestinal degradation. It should be noted that both the t_{max} and C_{max} increase (88) with an increase in

$t_{1/2}$. The fraction of the therapeutic peptide absorbed via the oral mucosa should not be calculated from pharmacodynamic response data alone because the efficiency of peptide absorption with respect to its pharmacodynamic response depends not only on the total dose absorbed but also on the rate at which the peptide is taken up by the target organ. Pharmacokinetics after buccal dosing can also be performed using moment analysis (89).

Enzymatic Degradation

The proteolytic activity of the buccal mucosa presents a significant barrier to the delivery of proteins and peptides. Buccal homogenate studies may provide initial data concerning the rate and extent of biochemical degradation of peptides when delivered by the buccal route (90). The disadvantage of homogenate studies includes the inability to distinguish among cytosolic, membrane-bound, and intercellular proteolytic activity. Because protein and peptide transport can be either trans- or paracellular in nature, the exact location of these proteolytic enzymes is important.

A novel concept of using bioadhesive polymers such as derivatives of poly(acrylic acid), polycarbophil, and carbomer to protect therapeutically important proteins and peptides from proteolytic activity of enzymes, endopeptidases (trypsin and α -chymotrypsin), exopeptidases (carboxypeptidases A and B), and microsomal and cytosolic leucine aminopeptidase (91) has been developed. However, cysteine protease (pyroglutamyl aminopeptidase) may not be inhibited by polycarbophil and carbomer (91).

Table 3 Exemplary absorption enhancers under investigation for buccal administration

Class	Exemplary compounds
Chelators	EDTA
Surfactants	Benzalkonium chloride Brij 35 Laureth-9 Sodium dodecylsulfate
Bile salts	Sodium deoxycholate Sodium glycocholate
Fatty acids	Sodium myristate
Peptidase inhibitors	Aprotinin
Miscellaneous	Chondroitinase ABC 1-Dodecylazacycloheptan-2-one Cyclodextrin Quillajasaponine Sodium salicylate

Enzyme Inhibitors

Protease inhibitors are generally used as enzyme inhibitors. These include aprotinin (79), bestatin (92), chondroitinase, and hyaluronidase. According to several published reports, these inhibitors appear to lack adequate effectiveness when administered simultaneously with various peptides in vivo (79).

Permeation Enhancers

Generally, these enhancers fall into six categories: 1) chelators (e.g., EDTA, EGTA); 2) surfactants, which are subdivided into non-ionic, e.g., laureth-9, polysorbate 80, sucrose esters and dodecylmaltoside, cationic; (e.g., cetylmethylammonium bromide), and anionic (e.g., sodium dodecyl glycocholate, sodium lauryl sulfate); 3) Bile salts and other steroidal detergents (e.g., sodium glycocholate, sodium taurocholate, saponins, sodium taurodihydrofusidate, and sodium glycodihydrofusidate); 4) fatty acids (e.g., caprylic acid); 5) nonsurfactants (e.g., 1-dodecylazacycloheptan-2-one (Azone), salicylates, and sulfoxides); and 6) enzymes (e.g., phospholipases, hyaluronidases, neuraminidase, and chondroitinase ABC).

Different mechanisms of absorption enhancement for these permeation enhancers have been proposed. Most permeation enhancers are thought to disrupt the lipid bilayer, which increases membrane fluidity. Some enhancers (nonionic and presumably ionic surfactants) may solubilize and extract lipids (92). Bile salts enhance absorption, stabilize enzyme-labile drugs, and potentially inhibit proteolytic degradation and aggregation of therapeutic proteins by formation of micelles (92). Permeation enhancers, which open tight junctions, are of little benefit in oral mucosal drug delivery because tight junctions are uncommon in these tissues. Structure-/absorption enhancement activity relationships have not been completely characterized for permeation enhancers (Table 3). However, for surfactants, the structure of the polar head groups strongly influences the permeability. For example, it has been reported that for surfactants, absorption enhancement was greatest for ether-based surfactants rather than for esters with similar structure (92).

SUMMARY

In recent years, there has been explosive growth in our understanding of the mechanisms associated with the absorption of drugs, especially of therapeutic peptides and

proteins. Scientists from a variety of disciplines continue to elucidate the variables associated with the optimal formulation and delivery of drugs via the oral mucosa. A greater understanding of the para- and transcellular route of drug absorption, proteolytic enzyme activity that may potentially degrade therapeutic peptides, and simultaneous degradation of compounds during the mucosal transport process is essential to the development of buccal delivery systems. Moreover, methods to increase drug flux (e.g., use of permeation enhancers) without associated toxicity, strategies to inactivate proteolytic enzymes, and innovative approaches with regard to controlled drug-delivery and mucoadhesive dosage forms will all improve the delivery of drug substances via the oral cavity.

ACKNOWLEDGMENT

This work was supported in part by NIH grants (AI 36624, EY 09171, and EY 10659; A.K.M.).

REFERENCES

- Rathbone, M.J.; Drummond, B.K.; Tucker, I.G. The Oral Cavity as Site for Systemic Drug Delivery. *Adv. Drug Del. Rev.* **1994**, *13*, 1–22.
- de Varies, M.E.; Bodde, H.E.; Verhoef, J.C.; Junginger, H.E. Developments in Buccal Drug Delivery. *Crit. Rev. Ther. Drug. Carr. Syst.* **1991**, *8*, 271–303.
- Merkle, H.P.; Anders, R.; Wermerskirchen, A. Mucoadhesive Buccal Patches for Peptide Delivery. *Bioadhesive Drug Delivery Systems*; Lenaerts, V., Gurney, R., Eds.; CRC Press: Boca Raton, FL, 1990; 105.
- Chen, S.Y.; Squier, C.A. The Ultrastructure of the Oral Epithelium. *The Structure and Function of Oral Mucosa*; Meyer, J., Squier, C.A., Gerson, S.J., Eds.; Pergamon, NY, 1984; 7.
- Harris, D.; Robinson, J.R. Drug Delivery Via the Mucous Membranes of the Oral Cavity. *J. Pharm. Sci.* **1992**, *81*, 1–10.
- Moffat, A.C. Absorption of Drugs Through the Oral Mucosa. *Top. Med. Chem.* **1972**, *4*, 1.
- Beckett, A.H.; Hossie, R.D. Buccal Absorption of Drugs. *Handbook of Pharmacology*; Brodie, B.B., Gillette, J.R., Eds.; Springer: NY, 1971; 25.
- Siegel, I.A.; Hall, S.H.; Stambaugh, R. Permeability of the Oral Mucosa. *Current Concepts of the Histology of Oral Mucosa*; Squier, C.A., Meyer, J., Eds.; Charles C. Thomas: IL, 1971; 274.
- Dali M.M.; Heran C.L.; Kaeppli M.; Stetsko P.I.; Smith R.L. Intra-Oral (Buccal/Sublingual) Bioavailability of Propranolol in Conscious Rabbits as a Function of PH and Dosing Variables Proceedings of the American Association of Pharmaceutical Scientists Annual Meeting New Orleans LA Nov 19–23 1999, 1394, Abstract Number 2559.
- Evered, D.F.; Sadoogh-Abasian, F.; Patel, P.D. Absorption of Nicotinic Acid and Nicotinamide Across Human Buccal Mucosa In Vivo. *Life Sci.* **1983**, *27*, 1649–1661.
- Evered, D.F.; Mallet, C. Thiamine Absorption Across Human Buccal Mucosa In Vivo. *Life Sci.* **1983**, *32*, 1355–1358.
- Gandhi, R.B.; Robinson, J.R. Oral Cavity as a Site for Bioadhesive Drug Delivery. *Adv. Drug Del. Rev.* **1994**, *13*, 43–74.
- Junginger, H.E.; Hoogstraate, J.A.; Verhoef, J.C. Recent Advances in Buccal Drug Delivery and Absorption—In Vitro and In Vivo Studies. *J. Control. Rel.* **1999**, *62*, 149–159.
- Squier, C.A.; Johnson, N.W.; Hopps, R.M. *Human Oral Mucosa Development, Structure and Function*; Blackwell Scientific: Oxford, 1976; 7.
- Edgar, W.M. Saliva: Its Secretion, Composition and Functions. *Br. Dent. J.* **1992**, *172*, 305–312.
- Beckett, A.H.; Triggs, E.J. Buccal Absorption of Basic Drugs and Its Application as an In Vivo Model of Passive Drug Transfer Through Lipid Membranes. *J. Pharm. Pharmacol.* **1967**, *19*, 31S–41S.
- Al-Sayed-Omar, O.; Johnston, A.; Turner, P. Influence of PH on the Buccal Absorption of Morphine Sulphate and Its Major Metabolite, Morphine-3-Glucuronide. *J. Pharm. Pharmacol.* **1987**, *39*, 934–935.
- Anders, R.; Merkle, H.P. Evaluation of Laminated Muco-Adhesive Patches for Buccal Drug Delivery. *Int. J. Pharm.* **1989**, *49*, 231–240.
- Kornfeld, R.; Kornfeld, S. Comparative Aspects of Glycoprotein Structure. *Annu. Rev. Biochem.* **1976**, *45*, 217.
- Peppas, N.A.; Buri, P.A. Surface, Interfacial and Molecular Aspects of Polymer Bioadhesion on Soft Tissues. *Advances in Drug Delivery Systems Controlled Release Series*; Anderson, J.M., Kim, S.W., Eds.; Elsevier: Amsterdam, 1986; 1, 257.
- Chen, J.L.; Cyr, G.N. Compositions Producing Adhesion Through Adhesion. *Adhesive Biological Systems*; Manly, R.S., Ed.; Academic Press: NY, 1970; 163, Chap. 10.
- Smart, J.D.; Kellaway, I.W.; Worthington, H.E.C. An In Vitro Investigation of Mucosa-Adhesive Materials for Use in Controlled Drug Delivery. *J. Pharm. Pharmacol.* **1984**, *36*, 295–299.
- Barrer, R.M.; Barrie, J.A.; Wong, P.S.L. The Diffusion and Solution of Gases in Highly Crosslinked Copolymers. *Polymers* **1968**, *9*, 609–627.
- Park, K.; Ch'ng, H.S.; Robinson, J.R. Alternative Approaches to Oral Controlled Drug Delivery Bioadhesives and In Situ Systems. *Recent Advances in Drug Delivery Systems*; Anderson, J.M., Kim, S.W., Eds.; Plenum Press: NY, 1984; 163.
- Park, K.; Robinson, J.R. Bioadhesive Polymers as Platforms for Oral Controlled Drug Delivery. *Int. J. Pharm.* **1984**, *19*, 107–127.
- Gurny, R.; Meyer, J.M.; Peppas, N.A. Bioadhesive Intraoral Release Systems: Design, Testing, and Analysis. *Biomaterials* **1984**, *5*, 336–340.
- Ch'ng, H.S.; Park, H.; Kelley, P.; Robinson, J.R. Bioadhesive Polymers as Platforms for Oral Controlled Drug Delivery II: Synthesis and Evaluation of Some

- Swelling, Water-Insoluble Bioadhesive Polymers. *J. Pharm. Sci.* **1985**, *74*, 399–405.
28. Flory, P.J. *Principles of Polymer Chemistry*; Cornell University Press: NY, 1953; 541.
29. Park, H.; Robinson, J.R. Physicochemical Properties of Water Insoluble Polymers Important to Mucin/Epithelial Adhesion. *J. Control. Rel.* **1985**, *2*, 47–57.
30. Smart, J.D.; Kellaway, I.W. In Vitro Techniques for Measuring Mucoadhesion. *J. Pharm. Pharmacol.* **1982**, *34*(Suppl.), 70.
31. Mikos, S.A.; Smart, J.D. Scaling Concepts and Molecular Theories of Synthetic Polymers to Glycoprotein Networks. *Bioadhesive Drug Delivery Systems*; Lenaerts, V., Gurney, R., Eds.; CRC Press: Boca Raton FL, 1990; 25.
32. Mortazavi, S.A.; Smart, J.D. Factors Influencing Gel-Strengthening at the Mucoadhesive-Mucus Interface. *J. Pharm. Pharmacol.* **1994**, *46*, 86–90.
33. Teng, C.L.C.; Ho, N.F.L. Mechanistic Studies in the Simultaneous Flow and Adsorption of Polymer Coated Latex Particles on Intestinal Mucus. *J. Control. Rel.* **1987**, *6*, 133–149.
34. Park, K.A. A New Approach to Study Mucoadhesion: Colloidal Gold Staining. *Int. J. Pharm.* **1989**, *53*, 209–217.
35. Hassan, E.E.; Gallo, J.M. A Simple Rheological Method for the In Vitro Assessment of Mucin-Polymer Bioadhesive Bond Strength. *Pharm. Res.* **1990**, *7*, 491–495.
36. Kamath, K.R.; Park, K. Mucosal Adhesive Preparations. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1994; 10, 133.
37. Ch'ng, H.S.; Park, H.; Kelly, P.; Robinson, J.R. Bioadhesive Polymers as Platforms for Oral Controlled Drug Delivery II: Synthesis and Evaluation of Some Swelling, Water-Insoluble Bioadhesive Polymers. *J. Pharm. Sci.* **1985**, *74*, 399–405.
38. Khosla, R.; Davis, S.S. The Effect of Polycarbophil on the Gastric Emptying of Pellets. *J. Pharm. Pharmacol.* **1987**, *39*, 47–49.
39. Ahuja, A.; Khar, R.K.; Ali, J. Mucoadhesive Drug Delivery Systems. *Drug Dev. Ind. Phar.* **1997**, *23*, 489–515.
40. Lee, P.I. Kinetics of Drug Release from Hydrogel Matrices. *Advances in Drug Delivery Systems*; Anderson, J.M., Kim, S.W., Eds.; (Controlled Release Series) Elsevier: Amsterdam, 1986; 1, 277.
41. Peppas, N.A.; Sahlin, J.J. A Simple Equation for the Description of Solute Release. III. Coupling of Diffusion and Relaxation. *Int. J. Pharm.* **1989**, *57*, 169–172.
42. Imbert, D.; Cullander, C. Buccal Mucosa In Vitro Experiments. I. Confocal Imaging of Vital Staining and MTT Assays for the Determination of Tissue Viability. *J. Control. Rel.* **1999**, *58*, 39–50.
43. Audus, K.L.; Bartel, R.L.; Hidalgo, I.J.; Borchardt, R.T. The Use of Cultured Epithelial and Endothelial Cells for Drug Transport and Metabolism Studies. *Pharm. Res.* **1990**, *7*, 434–451.
44. Pitha, J.; Harman, S.M.; Michel, M.E. Hydrophilic Cyclodextrin Derivatives Enable Effective Oral Administration of Steroidal Hormones. *J. Pharm. Sci.* **1986**, *75*, 165–167.
45. Merkle, H.P.; Anders, R.; Wermerskirchen, A.; Raechs, S. Buccal Route of Peptide and Protein Drug Delivery. *Peptide and Protein Drug Delivery*; Lee, V.H.L., Ed.; Marcel Dekker, Inc.: New York, 1990; 545.
46. Lee, V.H.L. *Peptide and Protein Drug Delivery*; Lee, V.H.L., Ed.; Marcel Dekker, Inc.: New York, 1990; 1.
47. Berner, B.; Wilson, D.R.; Guy, R.H.; Mazzenfa, G.C.; Clarke, F.H.; Maibach, H.I. The Relationship of pKa and Acute Skin Irritation in Man. *Phar. Res.* **1988**, *5*, 660–663.
48. Okamoto, H.; Hashida, M.; Sezaki, H. Structure-Activity Relationship of 1-Alkyl- or 1-Alkenylazacycloalkane Derivatives as Percutaneous Penetration Enhancers. *J. Pharm. Sci.* **1988**, *77*, 418–424.
49. Siegel, I.A.; Gordon, H.P. Effects of Surfactants on the Permeability of Canine Oral Mucosa In Vitro. *Toxicol. Lett.* **1985**, *26*, 153–158.
50. Oh, C.K.; Ritschel, W.A. Biopharmaceutical Aspects of Buccal Absorption of Insulin. *Methods Find. Exp. Clin. Pharmacol.* **1990**, *12*, 205–212.
51. Ritschel, W.A.; Ritschel, G.B.; Forusz, H.; Kraeling, M. Buccal Absorption of Insulin in the Dog. *Res. Commun. Chem. Pathol. Pharmacol.* **1989**, *63*, 53–67.
52. Sithigornugul, P.; Burton, P.; Nishihata, T.; Caldwell, L. Effects of Sodium Salicylate on Epithelial Cells of the Rectal Mucosa of the Rat: A Light and Electron Microscopic Study. *Life Sci.* **1983**, *33*, 1025–1032.
53. Li, C.; Bhatt, P.P.; Johnston, T.P. Transmucosal Delivery of Oxytocin to Rabbits Using a Mucoadhesive Buccal Patch. *Pharm. Dev. Technol.* **1997**, *2*, 265–274.
54. Longenecker, J.P.; Moses, A.C.; Flier, J.S.; Silver, R.D.; Carey, M.; Dubovi, E.J. Effects of Sodium Taurodihydrofusidate on Nasal Absorption of Insulin in Sheep. *J. Pharm. Sci.* **1987**, *76*, 351–355.
55. Sakai, K.; Kutsuna, T.M.; Nishino, T.; Fujihara, Y.; Yata, N. Contribution of Calcium Ion Sequestration by Polyoxyethylated Nonionic Surfactants to the Enhanced Colonic Absorption of *p*-Aminobenzoic Acid. *J. Pharm. Sci.* **1986**, *75*, 387–390.
56. Fix, J.A.; Engle, K.; Porter, P.A.; Leppert, P.S.; Selk, S.J.; Gardner, C.R.; Alexander, J. Acylcarnitines: Drug Absorption-Enhancing Agents in the Gastrointestinal Tract. *Am. J. Physiol.* **1986**, *251*, G332–G340.
57. Hersey, S.J.; Jackson, R.T. Effect of Bile Salts on Nasal Permeability In Vitro. *J. Pharm. Sci.* **1987**, *76*, 876–879.
58. Pitelka, D.R.; Taggart, B.N.; Hamamoto, S.T. Effects of Extracellular Calcium Depletion on Membrane Topography and Occluding Junctions of Mammary Epithelial Cells in Culture. *J. Cell Biol.* **1983**, *96*, 613–624.
59. Teem, M.V.; Phillips, S.F. Perfusion of the Hamster Jejunum with Conjugated and Unconjugated Bile Acids: Inhibition of Water Absorption and Effects on Morphology. *Gastroenterology* **1972**, *62*, 261–267.
60. Clark, M.L.; Lanz, H.C.; Senior, J.R. Bile Salt Regulation of Fatty Acid Absorption and Esterification in Rat Everted Jejunal Sacs In Vitro and into Thoracic Duct Lymph In Vivo. *J. Clin. Invest.* **1969**, *48*, 1587–1599.
61. Tang-Liu, D.D.S.; Neff, J.; Zolezio, H.; Sandri, R. Percutaneous and Systemic Disposition of Hexamethylene Lauramide and Its Penetration Enhancement Effect on Hydrocortisone in a Rat Sandwich Skin-Flap Model. *Pharm. Res.* **1988**, *5*, 477–481.
62. Gandhi, R.B. *Some Permselectivity and Permeability Characteristics of Rabbit Buccal Mucosa*; Ph.D. Thesis, University of Wisconsin, Madison, 1990.

63. Dillon, T.F.; Douglas, R.G.; du Vigneaud, V.; Barber, M.L. Transbuccal Administration of Pitocin for Induction of Labor. *Obstet. Gynecol.* **1960**, *15*, 587–592.
64. Miller, G.W. Induction of Labor by Buccal Administration of Oxytocin: Review of 50 Cases. *J. Am. Osteopath. Assoc.* **1973**, *72*, 1110–1113.
65. Dawood, M.Y.; Ylikorkala, O.; Fuchs, F. Plasma Oxytocin Levels and Disappearance Rate After Buccal Pitocin. *Am. J. Obstet. Gynecol.* **1980**, *138*, 20–24.
66. Ishida, M.; Machida, Y.; Nambu, N.; Nagai, T. New Mucosal Dosage Form of Insulin. *Chem. Pharm. Bull.* **1981**, *29*, 810–816.
67. Anders, R.; Merkle, H.P.; Schurr, W.; Ziegler, R. Buccal Absorption of Protirelin: An Effective Way to Stimulate Thyrotropin and Prolactin. *J. Pharm. Sci.* **1983**, *72*, 1481–1483.
68. Schurr, W.; Knoll, B.; Ziegler, R.; Anders, R.; Merkle, H.P. Comparative Study of Intravenous, Nasal, Oral, and Buccal TRH Administration Among Healthy Subjects. *J. Endocrinol. Invest.* **1985**, *8*, 41–44.
69. Aungst, B.J.; Rogers, N.J. Site Dependence of Absorption-Promoting Actions of Laureth-9, Na Salicylate, Na₂EDTA, and Aprotinin on Rectal, Nasal, and Buccal Insulin Delivery. *Pharm. Res.* **1988**, *5*, 305–308.
70. Aungst, B.J.; Rogers, N.J.; Shefter, E. Comparison of Nasal, Rectal, Buccal, Sublingual, and Intramuscular Insulin Efficacy and the Effects of a Bile Salt Absorption Promoter. *J. Pharmacol. Exp. Ther.* **1988**, *244*, 23–27.
71. Nakada, Y.; Awata, N.; Nakamichi, C.; Sugimoto, I. The Effect of Additives on the Oral Mucosal Absorption of Human Calcitonin. *J. Pharmacobiol. Dyn.* **1988**, *11*, 395–401.
72. Ritschel, W.A.; Ritschel, G.B.; Forusz, H.; Kraeling, M. Buccal Absorption of Insulin in the Dog. *Res. Commun. Chem. Pathol. Pharmacol.* **1989**, *63*, 53–67.
73. Ho N.F.H.; Barsuhn C.L. Buccal Delivery of Drugs Proceeding of the International Symposium on Controlled Release Bioactive Materials CH-Basel 1989; *16*, 24.
74. Wolany G.J.M.; Munzer J.; Rummelt A.; Merkle H.P. Buccal Absorption of Sandostatin (Octreotide) in Conscious Beagle Dogs Proceeding of the International Symposium on Controlled Release Bioactive Materials CH-Basel 1990; *17*, Abstract D341.
75. al-Achi, A.; Greenwood, R. *Buccal Administration of Human Insulin in Streptozocin-Diabetic Rats*, *Res. Commun. Chem. Pathol. Pharmacol.* **1993**, *82*, 297–306.
76. Heiber, S.J.; Ebert, C.D.; Dave, S.C.; Smith, K.; Kim, S.W.; Mix, D. In Vivo Buccal Delivery of Calcitonin. *J. Control. Rel.* **1994**, *28*, 269–270.
77. Bayley, D.; Temple, C.; Clay, V.; Steward, A.; Lowther, N. The Transmucosal Absorption of Recombinant Human Interferon-Alpha B/D Hybrid in the Rat and Rabbit. *J. Pharm. Pharmacol.* **1995**, *47*, 721–724.
78. Gutniak, M.K.; Larsson, H.; Heiber, S.J.; Juneskans, O.T.; Holst, J.J.; Ahren, B. Potential Therapeutic Levels of Glucagon-Like Peptide I Achieved in Humans by a Buccal Tablet. *Diabetes Care* **1996**, *19*, 843–848.
79. Nakane, S.; Kakumoto, M.; Yukimatsu, K.; Chien, Y.W. Oramucosal Delivery of LHRH: Pharmacokinetic Studies of Controlled and Enhanced Transmucosal Permeation. *Pharm. Dev. Technol.* **1996**, *1*, 251–259.
80. Hoogstraate, A.J.; Coos Verhoef, J.; Pijpers, A.; van Leengoed, L.A.; Verheijden, J.H.; Junginger, H.E.; Bodde, H.E. In Vivo Buccal Delivery of the Peptide Drug Buserelin with Glycodeoxycholate as an Absorption Enhancer in Pigs. *Pharm. Res.* **1996**, *13*, 1233–1237.
81. Li, C.; Koch, R.L.; Raul, V.A.; Bhatt, P.P.; Johnston, T.P. Absorption of Thyrotropin Releasing Hormone in Rats Using a Mucoadhesive Buccal Patch. *Drug Dev. Ind. Pharm.* **1997**, *23*, 239–246.
82. Alur, H.H.; Beal, J.D.; Pather, S.I.; Mitra, A.K.; Johnston, T.P. Evaluation of a Novel, Natural Oligosaccharide Gum as a Sustained-Release and Mucoadhesive Component of Calcitonin Buccal Tablets. *J. Pharm. Sci.* **1999**, *88*, 1313–1319.
83. Sedmak, J.J.; Grossberg, S.E. Interferon Stabilization and Enhancement by Rare Earth Salts. *J. Gen. Virol.* **1981**, *52*, 195–197.
84. Knight, E., Jr.; Fathey, D. Human Fibroblast Interferon: An Improved Purification. *J. Biol. Chem.* **1981**, *256*, 3609–3611.
85. Wells, J.A.; Powers, D.B. In Vivo Formation and Stability of Engineered Disulfide Bonds in Subtilisin. *J. Biol. Chem.* **1986**, *261*, 6564–6570.
86. Kornblatt, M.J.; Hoa, G.H.B. The Pressure-Induced Inactivation of Mammalian Enolases is Accompanied by Dissociation of the Dimeric Enzyme. *Arch. Biochem. Biophys.* **1987**, *252*, 277–283.
87. Teng, C.L.D.; Groves, M.J. The Effect of Compactional Pressure on Urease Activity. *Pharm. Res.* **1988**, *5*, 776–780.
88. Alur, H.H.; Paher, S.I.; Mitra, A.K.; Johnston, T.P. Transmucosal Sustained Delivery of Chlorpheniramine Maleate in Rabbits Using a Novel, Natural Mucoadhesive Gum as an Excipient in Buccal Tablets. *Int. J. Pharm.* **1999**, *188*, 1–10.
89. Kondo, S.; Sugimoto, I. Moment Analysis of Intravenous, Intraduodenal, Buccal, Rectal, and Percutaneous Nifedipine in Rats. *J. Pharmacobiol. Dyn.* **1987**, *10*, 462–469.
90. Yamamoto, A.; Hayakawa, E.; Lee, V.H. Insulin and Proinsulin Proteolysis in Mucosal Homogenates of the Albino Rabbit: Implications in Peptide Delivery from Nonoral Routes. *Life Sci.* **1990**, *26*, 2465–2474.
91. Luessen, H.L.; Verhoef, J.C.; de Boer, G.; Junginger, H.E.; de Leeuw, B.J.; Borchard, G.; Lehr, C.M. Multifunctional Polymers for the Peroral Delivery of Peptide Drugs. *Bioadhesive Drug Delivery Systems*; Mathiowitz, E., Chickering, D.E., III, Lehr, C.M., Eds.; Marcel Dekker, Inc.: New York, 1999; 299.
92. Lee, V.H.L. Enzymatic Barriers to Peptide and Protein Absorption. *CRC Crit. Rev. Ther. Drug. Carrier Syst.* **1988**, *5*, 69–97.
93. Smart, J.D. An In Vitro Assessment of Some Mucosa Adhesive Dosage Forms. *Int. J. Pharm.* **1991**, *73*, 69–74.

PEDIATRIC DOSING AND DOSAGE FORMS

Rosalie Sagraves

University of Illinois at Chicago, Chicago, Illinois, U.S.A.

INTRODUCTION

The administration of medications to pediatric patients is in many ways difficult because health care providers and parents are faced with many challenges not experienced, or experienced to a lesser degree, than when medications are prescribed for and taken by adults. First, less information is available about the use of most medications for pediatric patients. In fact only about 20% of drugs marketed in the United States have labeling for pediatric use (1). Milap Nahata, in a 1999 article on pediatric drug formulations, stated that "only five of the 80 drugs most commonly used in newborns, and infants are approved for pediatric use" (1). Second, many drugs that are used for some pediatric patients are not in appropriate dosage forms for use by children. This includes even some medications approved for use in pediatric patients. These issues have resulted in many questions that need to be answered about drug administration to pediatric patients. For example, is the drug approved for use in pediatric patients and in what age groups? If not approved, is there scientific information that enables us to determine whether the drug is safe and effective for pediatric patients of various ages? If the drug is available commercially for pediatric use, what dose should be administered and how frequently? What route should be used for administration, and what dosage form selected? If the drug is not available in an appropriate dosage form for childhood use, can it be prepared extemporaneously? Is there stability studies, palatability tests, clinical data in children, etc. that pertain to the extemporaneous formulation? How should the drug be monitored for effectiveness as well as for adverse effects? Information determined in adult medication studies may not be applicable to pediatric patients because of pharmacokinetic and pharmacodynamic differences as well as differences in disease states for which a particular drug might be used. Many questions about the use of particular drugs in various age groups of pediatric patients can only be answered through well-designed, randomized controlled studies in pediatric patients who need certain medications for particular health problems.

In 1997, the Food and Drug Administration (FDA) proposed new regulations for how pharmaceutical manufacturers would access safety and efficacy of certain new drugs that could have pediatric indications (1, 2). Thereafter, the FDA and the American Association of Pharmaceutical Scientists (AAPS) held a conference with academicians, pharmaceutical industry representatives, and U.S. Pharmacopeia (USP) representatives to discuss these proposed FDA regulations.

The FDA Modernization Act (FDAMA) of 1997 contains within it financial incentives for the development and marketing of drugs that could be used for pediatric patients (3). Some of these incentives include an extension of 6 months on market exclusivity and waiving fees for supplemental applications needed for receiving the approval of drugs for pediatric use that are already approved for adult use. In addition, the FDA published a list of drugs approved in adults for which additional pediatric data may produce health benefits for pediatric patients (4). For drugs on this list, FDA may ask a pharmaceutical manufacturer why it has not sought approval of a particular drug for pediatric use. So far there has not been much advancement in this area. This may be due to the wait for final approval of FDAMA.

Various medical and pharmacy organizations have worked hard throughout the years in their efforts to better educate children, parents, educators, and health care providers about the medications and their appropriate use. Indeed, individuals who help care for children may not be adequately trained to educate children about medications that they need to use. Therefore in June 2000, the USP started the development of three target initiatives: principles for educating children about their medications, guidelines for developing and evaluating information for children, and developing specific curricular information in a modular format (5). The USP position about educating children about medications may be found on their website (www.usp.org). The following information pieces have been developed by the USP (5):

- *Guide to Developing and Evaluating Medicine Education Programs and Materials for Children and*

Adolescents (joint publication of the American Health Association and USP)

- *A Kid's Guide to Asking Questions about Medicines*
- *Teaching Kids about Medicines*
- *Talking to Children about Their Medicines* (pamphlet developed jointly by Pfizer and USP to be disseminated to pediatricians and children's families)
- An Annotated Bibliography of Research and Programs Relating to Children and Medications

The USP has started working with the National Center for Health Education in New York to develop educational materials that can help school systems nationwide to know more about medications that students may need to take. The USP also adopted the following resolution to address the work that needs to be done in the area of health education (5):

Facilitate and contribute to the development of a rational school medicines policy, including guidelines for student, faculty, and staff medicine education, for acquisition, transport, storage, administration, use, and disposal of medicines; for protection of privacy; and for record-keeping in primary and secondary schools. Initiatives should be undertaken in collaboration with appropriate partners.

The USP has been working with the National Institute of Child Health and Human Development (NICHD) to develop a list of drugs for which more pediatric information is needed to insure proper use in children. The USP is also evaluating similarities and differences among neonates, children, and adults that may affect medication dosing and which might help in the appropriate

labeling of medicines for pediatric use. The USP is reviewing the literature and developing tables for drugs, using evidence-based information. USP members can find this information by contacting Joyce Weaver (jpw@usp.org).

This overview of pediatric dosing and dosage forms covers issues that pediatric health care providers face daily, such as age-related drug pharmacokinetic and pharmacodynamic changes that occur secondarily to physiologic changes in maturing neonates, infants, children, and adolescents that can affect drug absorption from various routes of administration as well as drug distribution, metabolism, and elimination. To be more knowledgeable about pharmacokinetic changes, therapeutic drug monitoring (TDM) must be undertaken for drugs with narrow therapeutic indexes and for those for which pharmacodynamic data (i.e., pharmacologic response that correlates to the drug concentration at the receptor site) correlates with pharmacokinetic information. Also addressed will be drug administration by various routes including intravenous (i.v.), oral (p.o.), intramuscular (i.m.), subcutaneous (s.c.), percutaneous, rectal, otic, nasal, ophthalmic, and inhalation. Another issue discussed is product selection for pediatric patients.

To better understand changes in drug disposition, the pediatric population needs to be categorized into various groups (Table 1) because children vary markedly in their absorption, distribution, metabolism, and elimination of medications. This occurs because neonates, infants, children, adolescents, and adults have different body compositions (i.e., as to their percentages of body water and fat) and have their body organs in different stages of development.

Table 1 Pediatric age groups terminology

Terms	Definition
Gestational age	Time from the mother's last menstrual period to the time the baby is born; at birth, a Dubowitz score in weeks gestational age is assigned, based on the physical examination of the newborn
Postnatal age	Age since birth
Postconceptional age	Age since conception, i.e., gestational plus postnatal age
Neonate	First 4 weeks or first month of life
Premature neonates	Born at less than 37-weeks gestation
Fullterm neonates	Born between 37- and 42-weeks gestation
Postterm neonates	Born after 42-weeks gestation
Infant	1 month to 1 year of age
Child	1–12 years of age
Adolescent	12–18 years of age

PEDIATRIC PHARMACOKINETICS AND PHARMACODYNAMICS

Effect of Developmental Physiologic Changes on Pharmacokinetics and Pharmacodynamics of Drugs

Rational pediatric pharmacotherapy is primarily based on the knowledge about a particular drug, including its pharmacokinetics and pharmacodynamics, that may be modified by physiologic maturation of the child from birth through adolescence. Physiologic changes that occur can affect drug absorption, distribution, metabolism, and elimination. The most dramatic changes occur during the neonatal period.

Oral Absorption

Drug absorption from the gastrointestinal (GI) tract is dependent on patient factors, physicochemical properties of the orally administered drug, and the drug formulation. Patient factors that affect GI absorption include absorptive surface area, maturation of the mucosal membrane, gastric and duodenal pH, gastric emptying time, GI motility, enzyme activity, bacterial colonization of the GI tract, and dietary intake, including the specific gastric content status at the time when a medication is ingested (6–8). Patient factors are influenced by rapid maturational changes that occur throughout early childhood, but which occur primarily during the first few months of life.

Most drugs are absorbed across the GI tract by passive diffusion, but a variety of drug physicochemical factors influence the extent of absorption. These factors include molecular weight, lipid solubility, ionization as well as disintegration and dissolution rates (7). In addition, drug absorption may be dependent on the dosage form selected (e.g., a liquid, a tablet that may need to be crushed, or a sustained-release product), and the particular brand selected. For timed-release preparations, the release characteristics must also be taken into consideration.

Gastric pH

When examining patient-specific factors such as gastric pH, which affect oral absorption, it should be noted that infants born vaginally who are at least 32-weeks gestation, usually have gastric pHs between 6 and 8 at birth (7, 8). Gastric pH then falls rapidly within a few hours after delivery to a pH of less than 3 (7, 8). The initial gastric pH is alkaline compared to that of adults and results from the presence of amniotic fluid in the infant's stomach (9, 10). Thereafter, gastric pH remains acidic until approximately day 10, then a nadir in acid production occurs between

days 10 and 30 of life. Then gastric acid production begins to increase, but gastric pH and maximal gastric output may not mirror that of adults on a per kilogram basis until after the neonatal period (7).

Gastric emptying and gastrointestinal motility

Gastric emptying time in neonates, especially those less than 24 h of age, may be variable (7). It may not reach adult levels until 6–8 months of age and may be associated with diet (11, 12). Gastrointestinal transit time may be prolonged and peristaltic activity unpredictable in young infants (8, 13); both appear affected by the feeding (13). Lebenthal and colleagues noted that breast-fed infants, older than 45 days of age, had gastric transit times longer than 10 h while formula-fed infants had transit times less than 10 h (14). It should also be noted that young infants have a propensity to reflux their gastric contents because of GI immaturity. All these factors affect the extent to which a drug may be absorbed.

Enzyme activity and microflora in the gastrointestinal tract

Pancreatic enzyme activity may be low at birth, but enzymes such as amylase, lipase, and trypsin develop to adult levels within the first year of life (15). Premature infants appear to have lower amylase levels than do full-term infants. Low concentrations of pancreatic enzymes may be the reason why newborns have a decreased ability to cleave prodrug esters such as chloramphenicol palmitate (7). Lipid-soluble drugs may not be well absorbed by neonates because of low lipase concentrations and bile acid pool (8).

More information is needed about the microflora of the GI tract and its effect on drug absorption. In addition, the effects of various diets and antibiotic use can alter the microflora of the GI tract (7).

Absorptive surface area

The surface area of the small intestine in young infants is proportionately greater than in adults. This physiologic difference may allow for increased drug absorption from the GI tract.

Intramuscular Absorption

When a child is unable to take a medication orally or the drug is unavailable for oral use, there may be a need to administer a drug parenterally by either the i.v. or i.m. route. Of these, the latter may be less desirable because of pain, irritation, and decreased drug delivery as compared to i.v. administration. Drug absorption after i.m. administration depends on various physicochemical and

patient factors. Physicochemical factors to be considered include lipid or water solubility, drug concentration, and surface area. When addressing drug solubility, it should be noted that lipophilic drugs readily diffuse through the capillary walls of endothelial cells whereas water-soluble drugs diffuse at fairly rapid rates from interstitial fluid to plasma via pores in capillary membranes (16). A lipid-soluble drug may be more rapidly absorbed i.m., but a water-soluble drug may be more desirable because the drug must be stable in an aqueous solution until administered. After administration, the drug must then be water soluble at physiologic pH until absorption occurs (16).

Drug absorption may be dependent on concentration, but available data do not allow us to determine whether an increased or decreased drug concentration results in better absorption. An increase in the osmolality of a pharmaceutical preparation secondary to the addition of another substance such as an excipient may decrease or slow down i.m. adsorption (16). Absorption occurs more rapidly when diffusion involves a large area of muscle or the drug spreads over a large muscle mass. The massaging of an injection site after i.m. administration increases the rate of absorption (16).

A physiologic determination of i.m. drug absorption is dependent on the adequacy of blood flow to muscle groups used for drug administration. Absorption rates differ at injection sites because blood flow varies among different muscle groups. For example, the absorption of a drug-administered i.m. in the deltoid muscle is faster than from the vastus lateralis that, in turn, is more rapid than from the gluteus (16, 17). This occurs because blood flow to the deltoid muscle is 7% higher than to vastus lateralis and 17% higher than to gluteal muscle groups (18). Physiologic conditions that reduce blood flow to a muscle group may adversely alter the rate and/or extent of a drug-administered i.m. Decreased perfusion or hemostatic decompensation, frequently observed in ill neonates and young infants, may reduce i.m. drug absorption. Drug absorption may be adversely affected in neonates who receive a skeletal muscle-paralyzing agent such as pancuronium (16) because of decreased muscle contraction. A small muscle mass in neonates and young infants may also reduce the ability of a drug to be adequately absorbed.

The injection technique used may alter i.m. absorption. This was noted when needles of different lengths were used. The use of a longer needle (38 vs. 31 mm, 1 1/2 vs. 1 1/4 in.) for i.m. administration in adult patients resulted in higher diazepam serum concentrations (18). This probably occurred because the drug administered with the shorter needle was actually administered s.c. rather than i.m.

Some drugs are absorbed more slowly after i.m. than oral administration; examples include diazepam, digoxin,

and phenytoin. This probably occurs because these drugs require a mixture of alcohol, propylene glycol, and water for solubility, and they are insoluble in the muscle after i.m. administration (18).

Complications associated with i.m. administration include nerve injury, muscle contracture, and abscess formation (19). Less common problems include intramuscular hemorrhage, cellulitis, skin pigmentation, tissue necrosis, muscle atrophy, gangrene, and cyst or scar formation. In addition, injury may occur from broken needles and inadvertent injection into a joint or vein (19).

Subcutaneous Absorption

The s.c. route is used for the administration of drugs such as insulin that require slow absorption. Injection technique and patient factors, such as fluid status and physical build, are important (18). Exercise, elevation or warming of the injection site, or inadvertently administering a drug i.m. rather than s.c. can increase absorption and be dangerous in some situations, such as hypoglycemia occurring in a diabetic patient from excessive insulin absorption (18). Adverse effects that can occur secondarily to s.c. administration include tissue ischemia, sterile and nonsterile abscesses, lipodystrophy, cysts, and granulomatous formation.

Intraosseous Drug Absorption

If an i.v. line cannot be placed, the intraosseous drug administration route can be used for pediatric patients during, for example, cardiopulmonary resuscitation (CPR) because drug delivery by this route is similar to that for i.v. administration (20). If drug or fluid deliver by this route is sluggish, a saline flush can be used to clear the needle. Intraosseous administration is used to deliver medications such as epinephrine, atropine, sodium bicarbonate, dopamine, diazepam, isoproterenol, phenytoin, phenobarbital, dexamethasone, and various antibiotics (20).

Percutaneous or Transdermal Absorption

The percutaneous (transdermal or topical) route for systemic drug delivery is used infrequently for pediatric patients. Medications are typically applied to the skin for their local effect. In the future, this route may be used more frequently for systemic effects as more transdermal systems are developed for drug delivery.

The percutaneous absorption or the transdermal delivery of a drug occurs in the following manner. Initially a topically applied drug is absorbed into the stratum corneum and diffuses through that layer of skin

into the epidermis and then into the dermis where drug molecules reach capillaries and enter the circulatory system. Diffusion through the stratum corneum is the rate-determining step unless skin perfusion is decreased. If the latter case, diffusion is controlled by the transfer of drug molecules into capillaries rather than by the diffusion process previously explained. Percutaneous or transdermal absorption (21, 22) is affected by

- Patient age.
- Application site.
- State of hydration of the stratum corneum.
- Thickness and intactness of the stratum corneum.
- Physical characteristics of the solute, and
- Physical characteristics of the vehicle or solvent.

Drug diffusion may be explained by Eq. 1:

$$J = \frac{K_m \times D_m \times C_s}{\ell} \quad (1)$$

where J is flux, K_m is the partition coefficient, D_m is the diffusion constant under specific conditions such as temperature and hydration, C_s is the concentration gradient, and ℓ is the length or thickness of stratum corneum (21).

Lipid-soluble drugs are better absorbed into the stratum corneum than are water-soluble drugs, but the latter do not easily traverse the stratum corneum. Thus, lipid-soluble drugs are more likely to be stored in the stratum corneum, whereas water-soluble drugs are more likely to diffuse across the stratum corneum to the epidermis and dermis (21).

Patient age

Drug absorption transdermally is not appreciably different in various age groups of patients except for neonates less than 32-week gestation at birth (23). Drug absorption is increased in premature neonates, because the stratum corneum is not completely formed at birth. An example of increased drug absorption occurred in two premature neonates who were repeatedly washed with 3% hexachlorophene and developed encephalopathy secondary to drug absorption (21). The absorption of the corticosteroid betamethasone valerate after topical application in children resulted in hypothalamus–pituitary–adrenal axis suppression. Children may have increased drug absorption from the percutaneous application of drugs not because of higher absorption rate but because of a greater topical application or a larger dose per kilogram. Examples of deaths in children from percutaneous drug absorption include those caused by salicylic acid and phenol absorption (21). Toxicity has also been noted with

the topical application of iodine and alcohol-containing products (23).

Application site

The ability of a drug to be absorbed transdermally depends on the thickness of the stratum corneum. For example, absorption occurs more readily through abdominal skin than through skin on the plantar surface of the foot. Topical absorption may be enhanced from a particular site by the application of an occlusive dressing.

Status of the stratum corneum

Percutaneous absorption of a drug is enhanced by the hydration of the stratum corneum. Such hydration affects the absorption of hydrophilic drugs more than lipophilic drugs. Drugs will penetrate damaged skin more than intact skin. Skin damaged because of dryness will allow for increased drug penetration through areas where the skin is cracked or broken.

Solute

The penetration of the solute (or drug) depends on its polarity and on the polarity of the delivery vehicle.

Vehicle or solvent

Drug-delivery vehicles typically used for topical application include lotions, ointments, creams, emulsions, and gels. Substances such as emulsifiers may be added to the drug and vehicle to improve the texture of an emulsion, a stabilizer to preserve drug stability, the vehicle, or both, a thickening agent to increase viscosity, or a humectant to draw moisture into the skin (21). It is particularly important to consider the vehicle and other additives when selecting a topical drug preparation for a neonate, especially when premature, because of the greater possibility of absorption of not only the drug but also other product ingredients. Toxic reactions have occurred in neonates from ingredients considered “inactive.”

Transdermal Drug-Delivery Systems

Drugs chosen for delivery via a transdermal drug-delivery system must adequately penetrate the skin in such a way that the system determines the delivery rate that should be fairly constant (21). In addition, the drug must not irritate or sensitize the skin. It is hoped that in the future more drugs will be developed for transdermal delivery. This could become an alternative route for drug delivery to children who have difficulty with oral administration.

Endotracheal Absorption

The endotracheal (ET) route has been used to administer medications during CPR when other routes, such as the i.v. route, are unavailable. It provides rapid access as well as rapid drug absorption and distribution (24). Some studies have shown that the time to reach peak absorption is similar to that for the i.v. route, but serum concentrations achieved were 10–33% of that achieved with i.v. administration, resulting in a weaker response. A depot effect has also been demonstrated for drugs such as epinephrine. Work needs to be done to determine the optimal dose by this route, drug-delivery vehicle, and the most effective delivery technique.

Rectal Absorption

The rectal route is used for local and systemic therapy for the following reasons (25):

- Nausea or vomiting.
- Rejection of oral medication because of its taste, texture, etc.
- Upper GI disease that might affect absorption.
- Medication absorption affected by food or gastric emptying.
- Medication is readily decomposed in gastric fluid but may be stable in rectal fluid, and
- First-pass effect of high-clearance drug may be partially avoided.

Absorption from the rectum depends on various physiological factors such as surface area, blood supply, pH, fluid volume, and possible metabolism by microorganisms in the rectum. The rectum is perfused by the inferior and middle rectal arteries, whereas the superior, the middle, and the inferior rectal veins drain the rectum (25). The latter two are directly connected to the systemic circulation; the superior rectal vein drains into the portal system. Drugs absorbed from the lower rectum are carried directly into the systemic circulation, whereas drugs absorbed from the upper rectum are subjected to hepatic first-pass effect (25). Therefore, a high-clearance drug should be more bioavailable after rectal than oral administration. The volume of fluid in the rectum, the pH of that fluid, and the presence of stool in the rectal vault may affect drug absorption. Because the fluid volume is usually low compared to that in other areas of the GI tract, a drug may not be completely soluble. In addition, a variety of organisms colonize the rectum, and it is debated whether these organisms are involved in drug metabolism (25). Absorption is also influenced by the dosage form used. For example, drugs are rapidly absorbed rectally from aqueous or alcoholic solutions, whereas absorption

from a suppository depends on its base, the presence of a surfactant, particle size of the active ingredient(s), and drug concentration (25). The following problems may be associated with the rectal route for drug administration (25):

- Decreased absorption secondary to defecation of the rectally administered pharmaceutical product.
- Less absorption rectally than orally because the absorbing surface area of the rectum is smaller.
- Dissolution problems for rectally administered medications because of lower fluid volume in the rectum than in the stomach, duodenum, etc..
- Microorganisms in rectum may cause degradation of some medications, and
- Patient or parent acceptance.

Distribution

A drug is distributed by moving from a patient's systemic circulation to various compartments, tissues, and cells. Distribution depends on patient factors, drug physiochemical properties, and the route of drug administration. Patient factors that influence drug distribution or the volume of drug distribution (V_d) include body composition, perfusion, protein- and tissue-binding characteristics, and permeability (7, 8). Many of these characteristics are age dependent. Drug physiochemical properties that may influence distribution include molecular weight, pK_a , and partition coefficient.

Differences in body composition

Age-related changes in body composition can alter the V_d of a drug. At birth, 85% of the weight of a premature infant may be water, compared to approximately 75% as total body water (TBW) in a full-term infant (13). Neonates have the highest percentage of extracellular water (65% of TBW in premature infants as compared to 35–44% in full-term neonates and 20% in adults). The intracellular water (ICW) is more stable throughout life (i.e., 25% in premature neonates, 33% in full-term neonates, and 40% in adults) (7). An infant's percentage of TBW approaches that of an adult male by 1 year of age (60% TBW); it reaches the same about the time of puberty or 12 years of age (8). Women have a lower percentage of TBW (50%) than men do because they have a higher concentration of body fat. Thus, neonates, because of their high TBW, have a higher V_d for water-soluble drugs such as aminoglycosides than older children or adults. For example, the V_d for an aminoglycoside such as gentamicin approximates that of extracellular cellular fluid volume, 0.5–1.2 L/kg for a neonate, but only 0.2–0.3 L/kg for an older child or an adult (8).

Adipose tissue increases from as little as 0.5% in a premature infant to approximately 16% of body weight for a full-term infant (26, 27). Boys experience a spurt in body fat between the ages of 5 and 10 years, and then a gradual decrease in fat content until about 17 years of age; girls usually have a rapid increase in adipose tissue at puberty (9). Thus, one would expect neonates and young infants to have a decreased V_d for lipid-soluble drugs. This has been noted for diazepam in neonates who have exhibited an apparent V_d of 1.4–1.8 L/kg compared to 2.2–2.6 L/kg in adults (28).

Protein binding

Neonates have lower concentrations of various plasma proteins (e.g., albumin concentrations about 80% of those in adults) for drug binding, but the albumin present may also have a lower affinity for binding drugs than noted for adults who are receiving the same medications. This lower affinity for binding drugs may result in a competition for various albumin-binding sites with substances such as bilirubin. Plasma protein binding noted in adults is usually achieved in children by the age of 1 year (13).

In neonates drugs such as various penicillins, phenobarbital, phenytoin, and theophylline have lower protein-binding affinity than in adults. This may increase the concentration of free or pharmacological active drug in neonates, and may also change the apparent volume of distribution. Thus, neonates may require different doses on a mg/kg basis compared to that for adults for these drugs to achieve appropriate therapeutic serum concentrations.

In addition to binding to plasma proteins in the neonate, some drugs such as sulfonamides may displace plasma bilirubin from binding sites. This may increase an infant's risk for developing kernicterus. The significance of drugs displacing bilirubin is controversial because bilirubin may have a greater affinity for albumin than drugs have (29).

Tissue binding

The binding of drugs to various body tissues appears to vary with age; for example, digoxin binding to erythrocytes is higher in neonates than in adults. This may be due to the increased number of binding sites on neonatal erythrocytes (30).

Drug penetration into the central nervous system

A drug is more likely to cross into the central nervous system (CNS) of a neonate rather than an older child or an adult. This most likely occurs because its CNS is less mature and the blood–brain barrier is less formed. This is an important consideration when antimicrobial therapy is

needed for the treatment of bacterial meningitis or anticonvulsant for seizures.

Metabolism

Although drug metabolism can occur in various body organs including the lungs, GI tract, liver, and kidneys as well as in the blood, the liver is the primary organ for metabolism. Most drugs are metabolized from lipid-soluble parent compounds to more polar, less lipophilic metabolites that are more readily eliminated renally (9).

Hepatic metabolism

Most drug metabolism occurs in the liver by phase I or phase II metabolic processes. Phase I reactions primarily biotransform an active drug to a more water-soluble compound that typically is inactive or has less activity than the parent compound. Oxidation, reduction, hydrolysis, and hydroxylation are examples of phase I reactions (6, 8, 16). Oxidation is primarily catalyzed by the cytochrome (CYP) P450 system that has a multitude of isozymes (at least 13 primary enzymes) with a multitude of isozymes of specific gene families (6). It appears that isozymes CYP450 1A2, 2D6, 2C19, and 3A3/4 are involved in drug metabolism in humans (6). Oxidizing enzyme systems appear to mature after birth so that by the age of 6 months, activity is similar to or even exceeds adult levels. More information about drugs affected by phase I reactions may be found in (6).

Phase II reactions (glucuronidation, sulfation, acetylation, and glutathione conjugation) usually involve the conjugation of active drugs with endogenous molecules to form metabolites that are more water soluble (16); glucuronidation is the most thoroughly studied reaction. It is postulated that maternal glucocorticoids inhibit the development of glucuronyltransferase, the enzyme involved in glucuronidation in utero. After birth this metabolic system matures rapidly and reaches adult levels by the age of 2 years (29).

Sulfate conjugation appears to be fully developed immediately prior to or at the time of birth. Infants and young children readily sulfate acetaminophen; in adults the major metabolic route is glucuronidation (31). Little is known about acetylation in neonates or infants. It is believed that neonates have an extremely low capacity for acetylation at birth, but this pathway matures at approximately 20 days of age (29).

Theophylline is an example of a drug that is readily metabolized in neonates by N-methylation to caffeine (process not relevant clinically in older infants, children, and adults). It is also a compound that has pharmacologic

activity versus apnea (like theophylline), but which may have toxicity when it is not readily metabolized by the liver, and its elimination is slowed by immature kidneys (6).

Neonates require close monitoring if their mothers received enzyme inducers such as phenytoin, phenobarbital, carbamazepine, or rifampin during pregnancy or if they need one of these drugs themselves (16). Examples of drugs that inhibit the metabolism of other medications include cimetidine, erythromycin, and ketoconazole (16).

Renal Elimination

The kidneys are the major route for drug elimination, especially for water-soluble compounds or the metabolites of lipid-soluble drugs. Renal drug elimination is dependent on renal blood flow, glomerular filtration, and tubular secretion and reabsorption. These functions appear to mature at different rates in the neonate and infant. Full-term infants achieve renal blood flow similar to that of adults by the age of 5–12 months; glomerular filtration approaches adult values by the age of 3–5 months (6). Premature neonates exhibit lower rates for glomerular filtration at birth than do full-term neonates, and more time is required of them postnatally to develop filtration ability (32). This is probably due to their lack of as many functional nephrons at birth. Tubular function is less mature in the neonate at birth than is glomerular filtration, and it matures at a slower rate. Tubular function begins to approach adult values by 7 months of age. Renal function is equal to that of adults by 1 year of age.

Aminoglycosides (e.g., gentamicin, tobramycin, amikacin) and digoxin are drugs whose eliminations are affected by renal maturation. The renal elimination of aminoglycosides in neonates and young infants parallels the maturation of glomerular function and correlates with creatinine clearance (33). The renal elimination of digoxin parallels kidney maturation. Dosage adjustment for this drug is necessary as renal function matures in neonates and young infants. In addition, older infants and children require higher mg/kg doses of digoxin than do adults to achieve the same serum concentrations. This may be due to decreased digoxin absorption or increased renal elimination (8).

THERAPEUTIC DRUG MONITORING

Therapeutic drug monitoring should encompass the entire drug-use process including drug selection, product selection, administration route, patient age, appropriate

dosing on a mg/kg or mg/m² basis, and monitoring serum concentrations when appropriate and observing the patient for optimal drug effect(s) and possible adverse drug events.

Important Differences in Pediatric Serum Drug Concentrations

For many drugs, especially for those with narrow therapeutic indexes, serum concentration ranges have been determined that correlate to minimum and maximum therapeutic effects as well as to the development of toxicity. Therapeutic serum concentration ranges for various drugs have been developed for adults, and these data have been applied to pediatric patients including neonates. Such data may be appropriate to monitor drug therapy in children, but possibly not in children of all ages or possibly not in children at all. For example, Painter et al. (34) noted that neonates need higher serum phenobarbital concentrations than do older children and adults to terminate seizures. Gilman et al. (35) observed that higher phenobarbital loading doses were needed to achieve serum concentrations in neonates that would reduce the occurrence of seizures. Thus, there may be a need for different serum concentration ranges for various drugs needed by different age groups of patients for a similar pharmacodynamic or therapeutic outcome.

Free serum concentrations, rather than total concentrations, of some drugs such as phenytoin may need to be monitored in some patients, including neonates, who have low serum albumin. Gilman has advocated the possibility of using individualized dosing and serum concentration range for pediatric patients because children, especially neonates, have rapidly maturing functions of various organs and changes in albumin for drug binding (36).

Serum concentration monitoring of various drugs administered to pediatric patients may appropriately give information about the drug but not its metabolites. This may be a problem when children metabolize specific drugs differently than adults with resulting differences in metabolite concentrations or the presence of different metabolites. This has been noted when premature infants have been administered theophylline for central apnea. A major metabolite of theophylline in neonates is caffeine, although only small concentrations of this metabolite are noted in older children and adults (37, 38). Caffeine is effective in treating apnea, and thus may add to the effectiveness of theophylline. This may help explain why lower theophylline serum concentrations may be needed for apnea rather than asthma. In addition, the presence of the 4-en metabolite of valproic acid noted in the serum of infants and young children, but not adults, receiving this

medication for seizures may be responsible for the hepatotoxicity of this drug in young pediatric patients (36, 39).

Serum Drug Concentrations

Because of the cost associated with therapeutic drug monitoring, serum drug concentrations must be drawn appropriately to provide useful information. Drugs typically followed pharmacokinetically are those with narrow therapeutic indexes for which there is an association between pharmacokinetic and pharmacodynamic data or toxicity. For many drugs, especially for those administered orally, the determination of trough concentrations (serum concentrations obtained prior to administration) may be most appropriate. This eliminates differences in absorption rates that could influence peak concentrations (e.g., orally administered phenobarbital, phenytoin, carbamazepine, or valproic acid). Trough concentrations may be important for drugs such as digoxin that take time to distribute to tissue receptors in such a way that serum concentrations reflect pharmacodynamic effects. Peak concentrations are best used for determining toxicity and therapeutic effects of drugs with short half-lives.

Table 2 gives therapeutic serum concentrations and pharmacokinetic information for some drugs administered to pediatric patients.

Technical Factors

Sample size and timing of blood drawing for serum concentration determination

Because of the small blood volume and the small size of veins, it is technically difficult to draw blood from neonates, infants, and young children for therapeutic drug monitoring, and it is therefore important to determine the best drawing schedule. For example, when are peak and trough data needed compared to trough data only? For anticonvulsants administered orally or i.v., trough concentrations are needed, whereas for aminoglycosides it may be important to obtain both peaks and troughs.

DOSING REGIMENS

Drugs for pediatric patients should be dosed on a mg/kg or a mg/m² basis using information available for the patient's age group. In addition, the patient's renal and hepatic functions must be considered. The route for administration must be determined based on the severity of the illness, the availability of the medication for a particular route of

administration, and whether the patient is able to take a medication orally.

The Bibliography succeeding the References at the end of this chapter contains a list of handbooks and other references that are useful sources of dosing information for neonatal and/or pediatric pediatrics. In addition, drug information centers in pediatric hospitals or university settings are another excellent resource for pediatric drug information.

EXCIPIENTS OR ADDITIVES IN MEDICATIONS

Pharmaceutical products may contain, in addition to the active or therapeutic agent(s), a variety of other ingredients that are termed inactive or inert that are categorized as excipients or additives (flavorings, sweeteners, preservatives, stabilizers, diluents, lubricants, etc.). The words inert or inactive are misnomers for some excipients because some have been shown to cause adverse effects. Neonates and young children are at risk for such adverse effects, because they may not be able to metabolize or eliminate an ingredient in a pharmaceutical product in the same manner as an adult. In addition, patients of various ages have experienced allergic reactions to excipients such as tartrazine dyes.

Benzyl alcohol is a preservative that may be present in multidose vials of bacteriostatic sodium chloride and bacteriostatic water for injection and pharmaceuticals available in multidose vials for parenteral use. An association between the presence of benzyl alcohol in solutions used for flushing intravascular catheters and to reconstitute medications and a gasping syndrome and deaths in neonates was first reported in the early 1980s (40, 41). The neonates also displayed clinical findings such as an elevated anion gap, metabolic acidosis, CNS depression, seizures, respiratory failure, renal and hepatic failures, cardiovascular collapse, and death. Those at highest risk were premature infants who weighed less than 1250 g at birth (40–42). In a study by Benda et al., premature neonates who survived benzyl alcohol administration were compared to neonates born after the use of benzyl alcohol-containing flush solutions was discontinued (43). They noted that survivors had a higher incidence of cerebral palsy (50%) compared to infants who did not receive benzyl alcohol flushes (2.4%) ($P < 0.001$). In addition, the incidence of cerebral palsy and developmental delay was 53.9 versus 11.9% in the two populations ($P < 0.001$). The cause is probably associated with benzyl alcohol use and the inability of neonates, especially those who are premature, to adequately metabolize benzyl

Table 2 Pediatric pharmacokinetic data of some medications^a

Drug	Therapeutic serum concentration (μg/ml)	Bioavailability (for oral drugs) (%)	Plasma protein binding (%)	V _d (L/kg)	t _{1/2} (h)
Carbamazepine	4–12	>70	40–90	1.5 (neonate) 0.8–1.9 (child)	8–25 (child) t _{1/2} varies with multiple dosing
Clonazepam	20–80 ng/mL	>85	47–80	3.2 (child)	20–40 (child)
Ethosuximide	40–100	~100	0	0.6–0.7 (child)	24–36 (child)
Gentamicin	trough ≤ 2 peak 4–10	Not available	<30	0.4–0.6 (neonate) 0.3–0.35 (child)	3–11.5 (<1 wk) 3–6 (1 wk–6 mo) 1.2 (child)
Phenobarbital	15–40	80–100	40–60	0.6–1.2 (neonate) 0.7–1 (child)	45–173 (neonate) 37–72 (child)
Phenytoin	10–20	85–95	>90	1–1.2 (premature neonate) 0.8–0.9 (full-term neonate) 0.7–0.8 (child)	6–140 (<8 days) ^b 5–80 (9–21 days) ^b 2–20 (21–36 days) ^b 5–18 (child) ^b
Theophylline	5–15	up to 100%, depending on the formulation	32–40 (neonate) 55–60 (child)	0.4–1 (premature neonate) 0.3–0.7 (child)	19.9–35 (neonate) 3.4 ± 1.1 (1–4 yrs)
Valproic acid	40–100 (150) ^c	100	>90 ^d	0.2 (child)	23–35 (neonate) 4–14 (child)

^aAge or stage of life in parentheses.
^bMichaelis–Menton pharmacokinetics; T_{1/2} varies with serum concentration.
^cUpper end of the serum concentration range is not definitely established.
^dMay vary with serum concentration.
(Adapted from Sagraves, R. Epilepsy and other Convulsive Disorders. In *Pediatric Pharmacotherapy*, 2nd Edn: Kuhn, R.J., Ed.) University of Kentucky: Lexington, 1993; Taketomo, C.K.; Hodding, J.H.; Kraus, D.M. *Pediatric Dosage Handbook*; LEXI-COMP, Inc.: Hudson, OH, 2000; Kauffman, R.E. Drug Therapeutics in the Infant and Child. In *Pediatric Pharmacology: Therapeutic Principles in Practice*; Yaffe, S.J., Aranda, J.V., Eds.; W.B. Saunders Co.: Philadelphia, 1992; 212–219; Rane, A. Drug Disposition and Action in Infants and Children. In *Pediatric Pharmacology: Therapeutic Principles in Practice*; Yaffe, S.J., Aranda, J.V., Eds.; W.B. Saunders Co. Philadelphia, 1992; 10–19.)

alcohol (44). The American Academy of Pediatrics (45), the Centers for Disease Control (46), and the FDA (47) recommend that the administration of products containing benzyl alcohol be avoided in infants. Preservative-free i.v. flush solutions are recommended (45–47).

Initially, it was believed that benzyl alcohol was only toxic in neonates who received doses greater than 99 mg/kg (42), but it has been suggested that lower doses may be toxic, resulting in kernicterus and intraventricular hemorrhages (48, 49). Therefore, pharmaceutical preparations and fluids containing benzyl alcohol should be avoided in premature neonates.

Benzoic acid and sodium benzoate are added in low concentrations to various pharmaceutical preparations as bacteriostatic and fungistatic agents. Hypersensitivity reactions to benzoates have occurred when administered to allergic patients, such as those with asthma, those who do not tolerate aspirin, and those with a history of urticaria (44). Hyperbilirubinemia and systemic effects attributed to benzyl alcohol may occur in premature neonates because benzyl alcohol is metabolized to benzoic acid (44).

Propylene glycol is found as a solvent in some i.v. multiple vitamin preparations and a variety of pharmaceutical preparations for parenteral administration including phenytoin, digoxin, and diazepam. MacDonald et al. (50) noted that neonates who received MVI-12 (propylene glycol dose of approximately 3 g/day) versus those who received MVI concentrate (propylene glycol dose of approximately 300 mg/day) had a significant increase in seizures. In addition, infants in the first group suffered from hyperbilirubinemia and renal failure. (Although MVI concentrate is no longer on the market in the United States, it was used in the early 1980s.)

Serum hyperosmolality has been reported in infants who received vitamin preparations (51), and in burn patients due to the topical absorption of propylene glycol-containing products (52, 53). In addition, burn patients have experienced metabolic acidosis with a high anion gap, decreased ionized calcium concentrations, acute renal failure, and death from topical propylene glycol absorption (44, 54). Problems associated with the oral ingestion of propylene glycol-containing products by children include CNS depression, seizures, and cardiac dysrhythmias (55). Hypotension, cardiac dysrhythmias, respiratory depression, and seizures have occurred after the rapid administration of phenytoin that may be associated with the propylene glycol (56).

The American Academy of Pediatrics Committee on Drugs recommends that medications intended for pediatric use be ethanol free (57). If, because of stability or solubility problems with the active ingredient(s), liquid medications need ethanol as an ingredient, but they should

not contain more than 5% v/v ethanol (57). The Academy also recommends that the ingestion of a single dose of an ethanol-containing product by a pediatric patient should not result in blood ethanol concentrations greater than 25 mg/100 mL, the volume of a packaged liquid medication should be of a minimal amount so that its entire ingestion would not result in a lethal dose and safety closures should be on all medicinals containing greater than 5% v/v ethanol. In addition, the Academy suggests that children under 6 years of age who need an ethanol-containing OTC preparation be under medical supervision and that doses of any ethanol-containing product be spaced at intervals to avoid ethanol accumulation (57).

The Academy of Pediatrics made their recommendations concerning ethanol exposure from medications based on potential acute and chronic ethanol-related problems. Acutely, the coadministration of ethanol may alter drug adsorption or metabolism, and may result in drug interactions (e.g., increased sedation when taken with sedatives). Disulfiram-like reactions have occurred after the ingestion of an alcohol-containing medication or when an ethanol-containing product is used in conjunction with medications such as metronidazole, sulfonamides, chloramphenicol, or cefamandole (57). The CNS effects (muscle incoordination, a longer reaction time, behavioral changes) are the most commonly reported acute adverse reactions associated with ethanol ingestion. Such reactions have occurred with blood ethanol concentrations in the range of 1–100 mg/100 mL (57). Lethal ethanol doses in children occur at approximately 3 gm/kg although deaths due to ethanol-induced hypoglycemia have occurred at lower doses or because of interactions with other medications (57, 58). Chronic ethanol exposure may induce hepatic enzymes, and may thus alter the clearance of drugs such as phenytoin, phenobarbital, and warfarin (59). Examples of other additives that have been problematic in pediatric patients include lactose (55), tartrazine dyes (44, 55) and sulfites (55).

It is therefore important for health care professionals, and especially those who are responsible for selecting and administering medications to premature neonates, to examine pharmaceutical preparations for the presence of inactive ingredients as well as for the active drug. The provision of medications should be based on choosing the safest preparations possible. Various brands of medications should be compared to ensure that products without hazardous excipients. In the hospital setting, pharmacy and therapeutics committees and the pharmacy department play important roles in this process because they compare pharmaceutical preparations for formulary selection. In the outpatient setting, physicians and pharmacists must responsibly select the most appropriate

brand of a particular medication. Kumar et al. (60) recommend that labeling for pharmaceutical products should include the names and the amounts of excipients as well as active ingredients to help health care professionals select appropriate drug products for neonates.

INTRAVENOUS ADMINISTRATION

Without being properly instructed about methods used for administering i.v. medications to pediatric patients, health care personnel may give a medication incorrectly, resulting in an inappropriate or unexpected therapeutic response. Therefore, it is important that health care personnel (nurses, physicians, pharmacists) understand how medications are administered by this route.

The i.v. route is most frequently chosen for medication delivery when a patient's clinical condition requires that a medication be administered by the most expeditious and complete method possible. In addition, some drugs are only available for i.v. administration. Although this route is the most reliable for drug delivery to the systemic circulation, problems can occur that reduce and/or delay medication delivery because of the product selected, dosage volume needed, or frequency of administration, but problems can also be associated with the i.v. delivery system used. The latter occur most frequently when a small medication volume is administered at a slow rate as is often needed for a neonate or young infant. A brief discussion of problems associated with i.v. drug delivery to pediatric patients is given here. A thorough overview of i.v. drug administration to pediatric patients is provided in Refs. (61 and 62).

Disposable IV Equipment, Effects on Drug Delivery

Infusion rates and location of injection sites

The first article to explore problems that can occur with i.v. drug delivery to pediatric patients was published in 1979 by Gould and Roberts (63). They demonstrated in their study using an in vitro system for drug administration (Fig. 1) that infusion rates as well as the location of the injection sites in the i.v. infusion system influence the infusion profile of i.v. administered medications. Figure 1 shows the effects of different i.v. fluid rates on the length of time to infuse 95% of a gentamicin dose administered at various sites in the infusion system (63). It was reported that at a slow infusion rate of 3 ml/h, a drug takes longer to be infused and that the time for infusion time depends on the site of administration (i.e., the further the drug

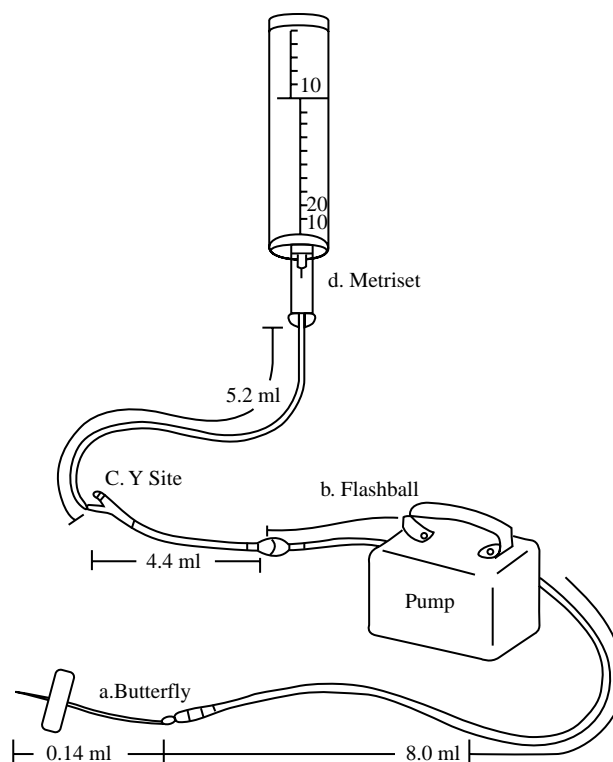


Fig. 1 Intravenous administration system used by Gould and Roberts. (From Ref. 63.)

injection from a patient, the longer to administer 95% of the medication, see Figs. 2,3). Thus, it took approximately 400 min to infuse gentamicin at an infusion rate of 3 ml/h via a Y-site in the administration system. However, the same drug administered at a butterfly injection site at the same fluid flow rate reduced the length of time to administer 95% of the drug to less than 20 min. Gould and Roberts also stated that the time needed for drug administration in their i.v. system was longer than what had been expected (63).

Type of injection site

Leff and Roberts (61) demonstrated that the amount of drug received by a pediatric patient and the drug-delivery rate are influenced by the type of injection site (Y-site, T-type, T-connector, stopcock, etc.) and the volume (dead space) contained in the particular site. For the delivery of small dosage volumes (less than 1 ml) i.v. tubing should have microinjection sites that prevent a drug from being sequestered in the injection site. In addition, the amount of i.v. fluid needed to adequately flush microinjection sites to clear the medication would be less than needed to flush injection sites found on tubing used to administer drugs to adults.

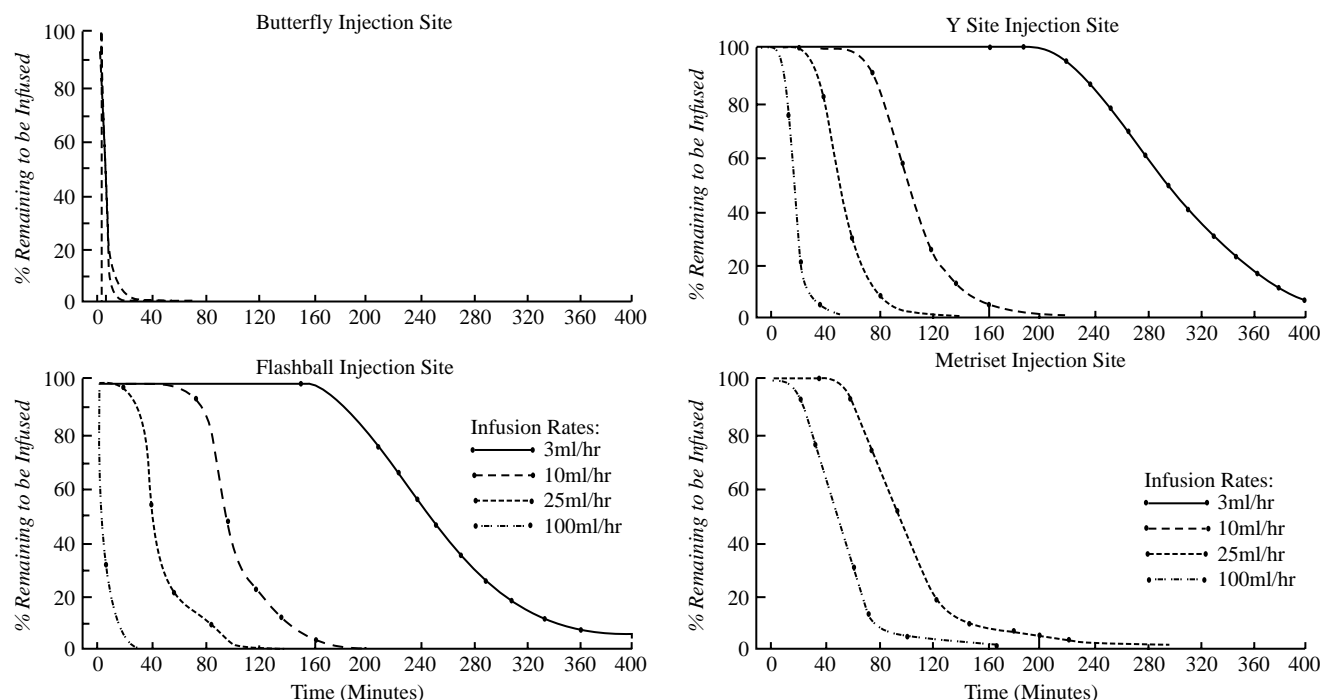


Fig. 2 Influence of i.v. flow rate on the infusion profile of gentamicin. (From Ref. 61.)

Fluid flow dynamics

Another characteristic of i.v. tubing that affects drug delivery is fluid flow dynamics. It appears that flow in i.v. tubing is best characterized by laminar flow, and the radius of the tubing. Poiseuille's law describes flow in i.v. tubing as

$$\frac{q_v}{4nL} = \frac{Pr^4}{4nL} \quad (2)$$

where q_v is the volumetric flow rate, P is the pressure change in the i.v. tubing, r is the radius of the i.v. tubing, n is the viscosity of the fluid, and L is the length of the i.v. tubing.

Thus, for i.v. delivery to pediatric patients, microbore tubing with an intraluminal diameter of <0.06 in. should be used rather than macrobore tubing. The use of microbore tubing allows the use of longer tubing lengths without significantly increasing delivery time.

Filters

A filter, especially one with a large reservoir volume, may prolong and/or reduce drug delivery. This occurs if the drug and its diluent are of different densities and there is a layering out of the drug in the filter (64). Therefore, a filter with a smaller reservoir volume should be selected.

Drug and Fluid Considerations for Intravenous Drug Administration

Characteristics of the drug and the fluid such as drug volume, osmolality, pH, and density may affect i.v. drug delivery. The frequency and duration of drug administration is also important as is the need for the infusion system to handle multiple drugs. This may lead to drug incompatibilities and problems in medication scheduling.

Osmolality and pH

Osmolality and pH must be considered when preparing a drug solution for i.v. administration to pediatric patients. Problems such as tissue irritation, pain on injection, phlebitis, electrolyte shifts, and even intraventricular hemorrhages in neonates have been associated with the administration of drug solutions with high osmolalities (65). Drug solutions should have osmolalities similar to serum osmolality, if possible. To control the osmolality, a drug can be diluted with a vehicle selected for i.v. infusion via a syringe infusion system (65) or the i.v. flow rate can be adjusted to achieve a particular drug-vehicle osmolality (61).

Density

If the density of a drug is significantly different from that of the diluent, the drug may layer out on the filter or in the i.v. tubing. The latter occurs more frequently if macrobore

tubing, a low flow rate, the i.v. system, or if the tubing is in a particular position. A density problem can be avoided by using microbore tubing which promotes mixing; this is especially important when i.v. flow rates are low, as are needed for neonates or young infants.

Frequency and Duration of Drug Administration; Multiple Drugs

To ensure that frequent doses are administered at appropriate intervals or that multiple drugs are administered to avoid drug incompatibilities, a syringe infusion pump can be used to administer drug volumes over a specific length of time. This helps avoid a situation where part of a drug dose is left in the tubing when the i.v. set is changed, as has been reported for manual administration techniques. More than one syringe pump can be used to simultaneously administer compatible drugs in a parallel system into a micro-Y-site or stopcock.

Types of Intravenous Administration

Drugs may require i.v. administration as continuous infusions or at intervals (q4h, q6h, q12h, etc.). Manual methods require the administration of the drug into the i.v. system at an injection site (Y-site, T-connector, stopcock, etc.), added to the i.v. solution in a mixing chamber, or added to an i.v. bag to be administered via gravity. A syringe pump or another mechanical device may be used for drug administration.

Manual administration

Manual administration is not as accurate as using a syringe pump for drug administration. It has been used for small volumes of medication (< 3 ml), a low flow rate (< 20 ml/h), or if the antegrade (forward toward the patient) injection of a drug bolus is safe (61). If a medication is to be administered antegrade, it should be administered slowly into a microinjection site toward the patient; microbore tubing should be placed between the injection site and the patient to reduce the time to get the drug to the patient. Leff and Roberts (61) recommended that the volume of the drug to be injected by the antegrade technique should be a smaller volume than the tubing fluid volume between the injection site and the patient. If the medication volume is too large to be safely given by antegrade administration, but the i.v. fluid flow rate is low (< 20 ml/h), the drug may be administered by a retrograde technique (Fig. 3).

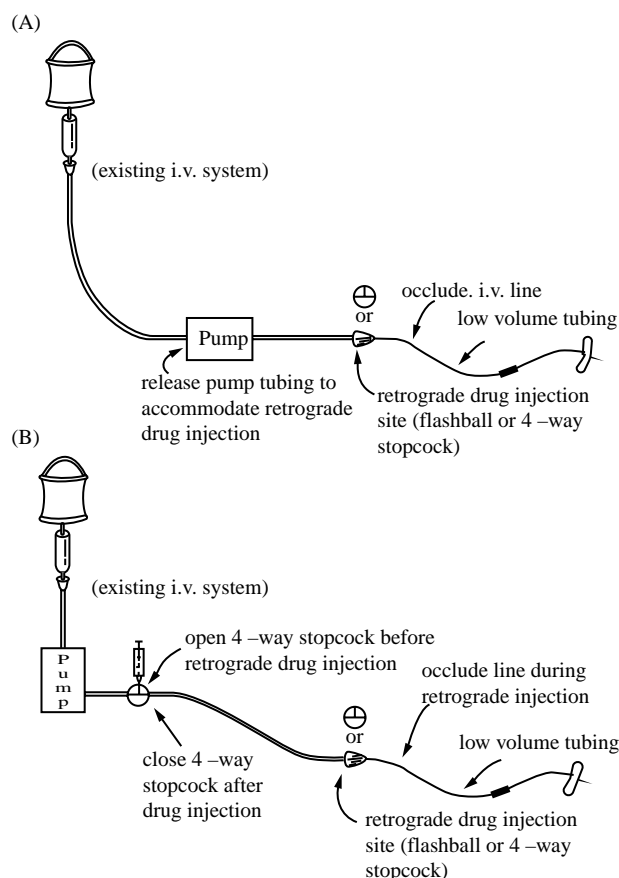


Fig. 3 Examples of retrograde system setups. (From Leff, R.D.; Roberts, R.J. Methods for intravenous drug administration in the pediatric patient. *J. Pediatr.* 1981, 98, 631–635.)

Mechanical system for drug administration

If a mechanical system is chosen for drug administration, the appropriate infusion device must be selected based on its operating mechanism, flow accuracy, flow continuity, and ability to detect occlusions. Other important factors include an alarm system, ease of operation, ability to be cleaned easily, and safety from children inadvertently trying to change pump settings. A syringe pump is best for delivering small dosage volumes and when intermittent intervals are needed for medications. It is the mechanical device most often selected for medication administration because it can be used for intermittent administration of small and large doses, or for the continuous infusion of medications at low rates. A drug can be administered separately from the primary i.v. fluid flow rate, with the drug and the fluid mixing for a short distance therefore in microbore tubing (see Fig. 4) before reaching the patient. In addition to being able to more accurately deliver medications than by manual methods,

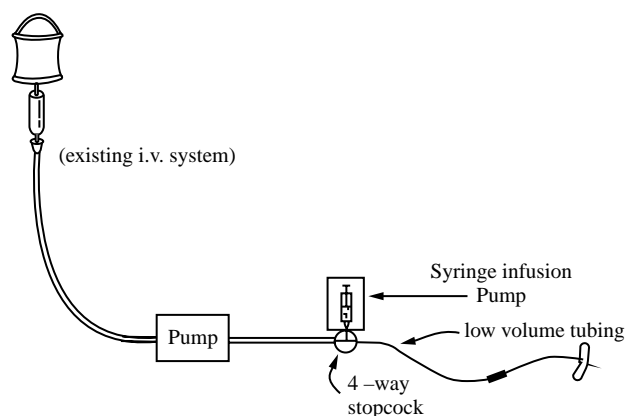


Fig. 4 Syringe pump setup with drug administered separately of the primary i.v. fluid flow rate. Mixing of the drug and i.v. fluid occurs at a stopcock and for a short distance in microbore tubing. (From Leff, R.D.; Roberts, R.J. *Methods for intravenous drug administration in the pediatric patient.* J. Pediatr. 1981, 98, 631–635.)

syringe pump systems have the advantage of being able to separate the administration of incompatible drugs, reduce difficulties associated with the administration of multiple doses, and shorten the time required to administer medications.

Additional Comments about IV Drug Administration

Reed and Gal (6) recommended the following steps to decrease problems associated with i.v. drug administration to pediatric patients:

1. Standardize and document total time for drug administration.
2. Document the volume of any solution used to flush an i.v. dose.
3. Standardize infusion techniques for drugs administration, especially for those with a narrow therapeutic index.
4. Use the largest gauge cannula that can be used.
5. Standardize dilution and infusion volumes for drugs given by intermittent i.v. injection, and avoid attaching lines for drug infusion to a central hub with solutions infused at widely disparate rates, and
6. Use low-volume i.v. tubing and use the most distal sites for drug administration.

In addition, one must remember that for infants the amount of fluid required for drug administration may take away from the amount of fluid available for nutrition. Thus, with medication administration, the fluid volume

must be as restrictive as possible so that the bulk of the daily fluid intake can be saved for nutrition. Health care providers must closely monitor daily fluid intake from all sources to prevent fluid overload and must also watch the osmolality of medications with diluents.

ADMINISTRATION OF ORAL MEDICATIONS

The oral route is typically the preferred route for medication administration to pediatric patients. Other routes may be used, if the patient cannot take a medication orally because of vomiting, being unable to swallow, or the medication is unavailable for oral use. In addition, for specific problems it may be better to deliver the medication directly to the area being treated, for example, inhalation, ophthalmic administration, or otic administration.

Dosage Forms

Oral liquids

Liquid medications are the most commonly administered oral medications to pediatric patients because of the ease of swallowing by infants and young children who cannot swallow solid dosage forms. However, availability of some medications as liquid formulations may be limited. If not available in liquid form, a solid dosage form may need to be modified by the pharmacist, other health care provider, or by the parent. If a solid dosage form is modified, for example a suspension is prepared, will the drug be stable and for how long, and will it be absorbed differently than the original dosage form? These are just a few questions that must be answered about the extemporaneous preparation of a drug product for a pediatric patient.

Alcohol-free products should be selected for pediatric patients whenever possible. Furthermore, the inactive ingredients or excipients contained in an oral preparation should be identified. this is especially important if the patient is known to have had an adverse reaction to a particular excipient or there is another reason to avoid a particular additive in a medication. The Committee on Drugs of the American Academy of Pediatrics recommended that pharmaceutical products contain a qualitative listing of inactive ingredients in order that products containing these substance could be avoided in patients who had problems with specific adjuvants (55). Kumar et al. (60) contains lists of inactive ingredients (sweeteners, flavorings, dyes, and preservatives) found in

Table 3 Osmolalities of liquid pharmaceuticals for oral administration

Liquid preparation	Osmolality (Mean \pm SEM)
Propylene glycol	8326 \pm 1467
Saccharin-containing drugs	
Albuterol	65
Haloperidol	47
Suspensions	2500 \pm 246
Sugar-containing ^a drugs	5574 \pm 594

^aSucrose, mannitol, glucose, and others.
(From Ref. 66.)

many liquid medications such as analgesics, antipyretics, antihistamine decongestants, cough and cold remedies, antidiarrheal agents, and theophylline preparations. The authors of the previous article and Golightly et al. have reviewed adverse effects associated with many inactive ingredients (44, 60).

Liquid medications, taken orally, can cause diarrhea and other GI symptoms, or they may aggravate GI distress that a patient is already experiencing. These GI effects can be associated with the high osmolality of some oral liquids. Osmolalities have been determined for various oral liquids (66). For preparations containing propylene glycol, or various sugars (e.g., sucrose, mannitol, glucose), osmolalities were noted to be high (see Table 3). It is important to compare various brands of liquid medications because they may contain different excipients and may have different osmolalities.

Sustained-Release Preparations

Most medications have shorter half-lives in children than in adults, and therefore children may need sustained-release products to maintain serum concentrations in the therapeutic range. For example, a sustained-release theophylline product may be needed for a child with asthma. It may need to be administered every 8 h to the child as compared to every 12 h for a healthy, nonsmoking adult to maintain therapeutic serum concentrations. When choosing a sustained-release theophylline preparation for a child, it must be remembered that because of differences in release properties, theophylline sustained-release products are not interchangeable. A product selected for the pediatric asthma patient should be reliably absorbed with a minimal serum concentration variation and not a preparation that has exhibited a difference in bioavailability when administered with or without food (67–69).

Extemporaneous liquid preparations

Because many medications are not available as liquid preparations, there are times when powder papers or suspensions must be prepared. An excellent information source about the preparation of liquid dosage forms for pediatric patients has been published by Nahata and Hipple (Nahata, M. C., Hipple, T. F. *Pediatric Drug Formulations*, 4th Ed. Harvey Whitney Books: Cincinnati, 2000).

Product selection

Products for oral administration should be in a dosage form most readily taken by the child. If the child is old enough to participate in the decision-making process, he or she may state a preference for a liquid, chewable tablet, tablet, or capsule, if the needed drug is available in a variety of dosage forms and appropriate dosage. If a liquid medication is needed, a product should be chosen based on texture, taste, and ease of administration. Other factors that must be considered are the absence of alcohol and dyes, and an osmolality that is close to physiologic (280–290 mOsm/kg). Are there excipients or adjuvants in the product, and if so, what are they and what is their concentration? Is there bioavailability information or pharmacokinetic information for the oral medication in pediatric patients, and if so, in what age groups? Is there information about the extemporaneous product that is to be prepared?

Rebecca Chater, a North Carolina pharmacist, recommends that pediatric patients be involved in medication counseling in order to improve their understanding of why a medication is needed. In the counseling process, the word medication should be used and not drug because of the connotation associated with the latter in today's society (70, 71). Wheeler recommends that, when possible, a product be selected that requires the fewest number of doses administered per day, for example, every 12 h dosing rather than every 8 h, so the medication does not need to be taken to school or day care for administration. If a medication must be given outside of the home, she recommends that two small labeled bottles be dispensed or one large bottle with a small empty bottle labeled to be used for medication administration at day care or school.

Health care providers including nurses, pharmacists, and physicians should demonstrate to parents and older children how medications should be administered and offer appropriate dosing devices (oral syringe, dropper, cylindrical medication spoon, or a small-volume doser with attachable nipple) to enable parents to accurately measure liquid products. A household teaspoon or tablespoon should not be used for medication administration because they are inaccurate. Kraus and Stohlmeyer

(72) explain the use of a new oral liquid medication delivery system that can be used for infants and young children who still use a bottle for feeding.

Administration techniques

The following information is presented to help health care providers counsel parents and older children on how medication should be administered by various routes.

Oral liquids

An oral liquid medication needed for an infant or young child should be shaken well, if required, accurately using an appropriate device. If a dropper or an oral syringe is used, the liquid should be administered toward the inner cheek. Administration in the front of the mouth may allow the child to spit out the medication, whereas administration toward the back of the mouth may result in gagging or choking. The oral syringe should be of an appropriate size to allow for administration into the inner cheek.

Oral solid dosage forms

A medication available only as a solid dosage form, may be prepared as an extemporaneous liquid (e.g., suspension) or it may be modified for oral use, for example, by crushing. As mentioned previously, a sustained-release product should not be crushed or chewed. For a solid, nonsustained-release medication, the product can be crushed and mixed with a small amount of food just prior to administration. Examples of foods that may be used for mixing include applesauce, yogurt, or instant pudding, but the medication should not be added to an entire dish of food or to infant formula, because the infant or child may not eat/drink the entire portion and thus not receive the total amount of medication.

OTHER ROUTES

Intramuscular Administration

Absorption of i.m. administered medications depends on the injection site because perfusion of individual muscle groups differs. For example, drug absorption from the deltoid muscle is faster than that from the vastus lateralis that is more rapid than from the gluteus (16, 17). In addition, lower perfusion or hemostatic decompensation, frequently observed in ill neonates and young infants, may reduce i.m. absorption. It may also be decreased in neonates who receive a skeletal muscle-paralyzing agent such as pancuronium because of decreased muscle contraction.

In addition, the smaller muscle mass of neonates and young infants provides a small absorptive area.

The injection technique and the length of the needle used may affect drug absorption, and thus serum concentrations. For example, using a longer needle (1½ vs. 1¼ in. or 3.8 vs. 3.1 cm.) for i.m. administration resulted in higher diazepam serum concentrations in adults (18). Therefore, it is important to select the appropriate site for drug administration as well as the appropriate length and needle bore. Sites that can be used for i.m. administration include anterior thigh and vastus lateralis, gluteal area and deltoid.

The midanterior thigh (rectus femoris) and the middle third of the vastus lateralis are used for i.m. administration to young infants as well as to older children (19). These sites are better developed and larger than other muscle groups that are used for drug administration to older children or adults. The technique is shown in Fig. 5 (19). With the patient lying supine, the “needle should be inserted in the upper lateral quadrant of the thigh, directed inferiorly at an angle of 45° with the long axis of the leg and posteriorly at a 45° angle” (19) to the surface on which the patient is lying. The person administering the injection should compress the tissues of the injection site to help stabilize the extremity. A 1-in. (2.5-cm) needle has been recommended for pediatric patients by Bergeson et al. (19) while Newton et al. (18) recommend a 23–26 gauge 1½-in. (3.8-cm) needle. The volume of drug that can be administered in this manner is 0.1–1 ml in infants and 0.1–5 ml in older children and adults (18).

The gluteal musculature develops as the infant or child increases his or her mobility; it becomes a more suitable injection site in children who are walking (18, 19). Damage to the sciatic nerve is the major problem associated with this injection site, and it occurs more commonly in infants because of their lack of gluteal muscle mass (73). Injury to the gluteal nerve, resulting in muscle atrophy, has occurred even when the injection technique was appropriately performed (19). Other nerves including the pudendal, posterior femoral cutaneous, and the inferior cluneal nerves have been damaged because of poor injection technique (19). Additional adverse effects associated with this drug administration route are discussed by Bergeson et al. (19).

A technique for gluteal administration is shown in Fig. 6 although other techniques are also used (19). All techniques involve the determination of the upper outer quadrant (see Fig. 6 for anatomical landmarks). After the location of the upper outer quadrant is determined, the needle should be inserted at a 90° angle to the surface on which the patient is lying, not to the patient’s skin (19). This site can be used for older children. A 1-in. (2.5-cm)

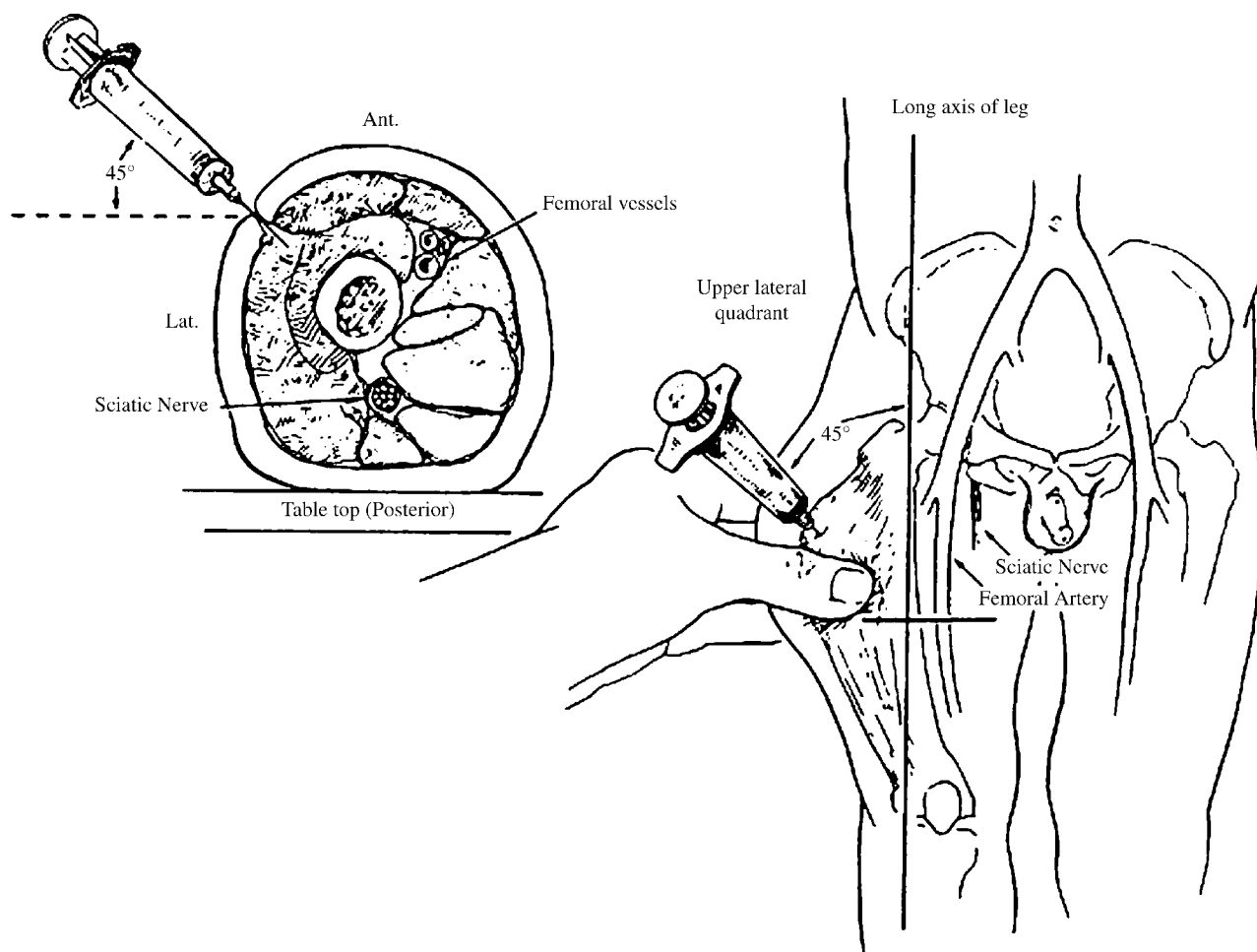


Fig. 5 A technique for anterior lateral-thigh intramuscular injection. (From Ref. 19.)

needle has been recommended (19). The volume of drug that can be administered in this manner is 0.1–5 ml for older children and adults (18).

The ventrogluteal (gluteus medius and minimus) site may be less hazardous for i.m. administration than the dorsogluteal (gluteus maximus) site (19). The technique is shown in Fig. 7 (19). The person administering a drug i.m. ventrogluteally should first note the anatomical landmarks (anterior superior iliac spine, tubercle of the iliac crest, and upper border of the greater trochanter). The needle is inserted into a triangular area bounded by these landmarks while the patient is in the supine position. The location for this injection can be determined “by placing the palm over the greater trochanter, the index finger over the anterior superior iliac spine, and spreading the index and middle fingers as far as possible” (19).

The deltoid muscle can be used for i.m. injections in older children, but it is not an option for young infants and children because of their limited muscle mass. Although

there are few complications associated with this administration route, nerve injury can occur (21). The technique is shown in Fig. 8 (19). The area for deltoid administration should be fully visible so that the anatomical landmarks can be visualized. Then the needle for deltoid injection should enter the muscle halfway between the acromion process and the deltoid tuberosity to avoid hitting the underlying nerves (19). The drug volume that can be administered by this route to older children and adults is 0.1–2 ml (18). The recommended needle length for older children is 1 in. (2.5 cm).

Subcutaneous Administration

The s.c. route is used for drug administration, such as insulin, that requires slow absorption. It is not commonly employed for medication administration for pediatric patients but is used for specific drugs. Typically a 1/2- or 1-in. (1.25- or 2.5-cm) needle is used with the volume of

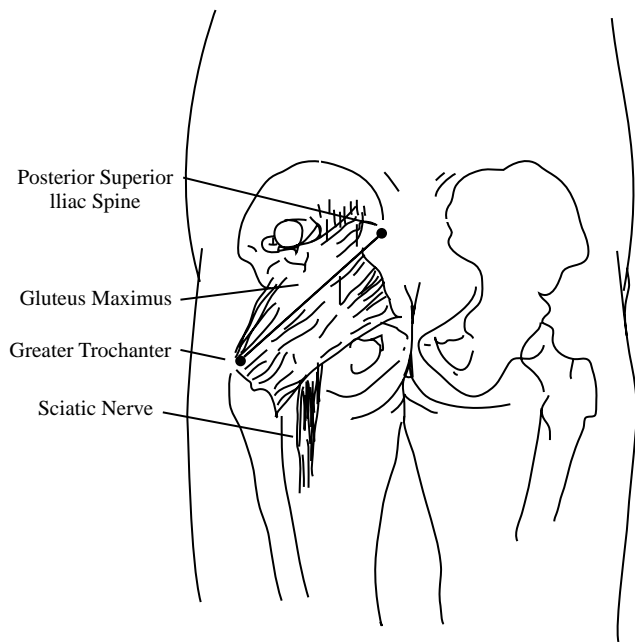


Fig. 6 A technique for gluteal-area intramuscular injection. (From Ref. 19.)

drug that can be administered by this route ranging from 0.1 to 1 ml (drug volume administered depends on patient size).

Percutaneous Administration

The skin should be thoroughly cleaned prior to applying a topical ointment, cream, etc. A thin layer of ointment or

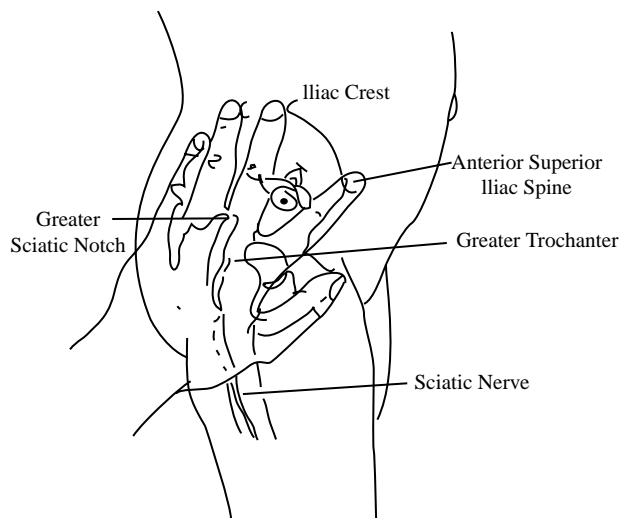


Fig. 7 von Hochstetter technique for ventrogluteal intramuscular injection. (From Ref. 19.)

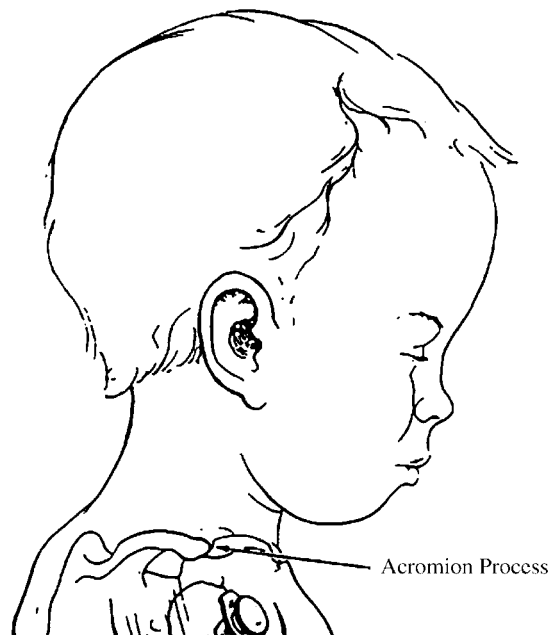


Fig. 8 A technique for deltoid intramuscular injection. (From Ref. 19.)

cream should be applied to the prescribed area to reduce the possibility of a toxic reaction. The area of the skin where the medication is applied should not be covered or occluded unless instructed to do so by the physician because this procedure may increase drug absorption. Specific information should be given on how to cover the area.

Rectal Administration

Before the administration of a rectal suppository, the child's rectal area should be thoroughly cleaned. The infant or child should be placed on his or her side or stomach. The wrapper should be removed from the suppository and its pointed end should be inserted into the rectum above the anal sphincter. (If only half a suppository is prescribed, the suppository should be cut lengthwise before administration.) A finger cot or finger wrapped in plastic can be used for administering the suppository (74). Because an infant or small child cannot adequately retain a suppository in the rectum, the buttocks can be held together firmly for a few minutes after rectal administration to hold the suppository in place (74).

Otic Administration

Otic preparations should be at room temperature prior to administration. If the otic product is a suspension, it should be gently shaken for approximately 10 sec. before

administration (74). The child should be lying on his or her side, and the earlobe should be gently pulled down and back to straighten the outer ear canal (for adults the earlobe is pulled up and back). Then the prescribed number of drops should be instilled into the ear without placing the dropper in the ear canal. The patient should be kept in a position with the ear tilted for approximately 2 min to help keep the drops in the ear (75). This procedure may be repeated for the treatment of the other ear, if needed. The tip of the dropper should be wiped with a clean tissue after use.

Nasal Administration

For adults, the first step in administering nose drops or a nasal spray is blowing the nose to clear the nasal passages of mucus and other secretions, but infants and young children are unable to do this. Therefore, the nasal passages may need be cleared with a bulb syringe prior to medication administration. A child should lie down on his or her back, or a young infant or child should be placed in a lying position, and the head should be tilted slightly backwards. An appropriate amount of medication should then be placed in each nostril. Thereafter, the infant or child should remain quiet for a few minutes to allow the medication to be absorbed. The dropper should be rinsed with hot water before it is returned to the medication container.

Ophthalmic Administration

An ophthalmic medication should be at room temperature prior to administration. If the eye drops are in a suspension, the container should be gently shaken before administration. A child old enough to follow directions should tilt his or her head slightly backward and to the side so that the eye drops will not drain into the tear ducts near the nose. The eyelids should be separated and the patient should be asked to look up. The appropriate amount of medication is instilled into the lower eyelid, using the medication dropper, which should be accomplished without touching the eyelids. The patient should look downward for a few seconds after drug administration. The eye(s) should then be closed for several minutes in order to spread the medication across the eyeball and be absorbed if the effect is to be systemic (76). In addition, it has been recommended to gently put pressure on the inside corner of the eye for at least a minute to retard drainage of the medication (74). If a squeeze bottle is used, the appropriate amount of medication should be gently squeezed into the eye(s). For each of these methods, the dropper or the tip of the squeeze bottle should be kept away from the eye or skin to avoid contamination of the administration device (76). The

dropper should not be rinsed after use because this could lead to contamination of the dropper and the medication. The package insert should be reviewed for specific information.

Another method for administering eyedrops to children recommends that the drops be applied to the inner canthus of the eye while the patient keeps his/her eyes closed until told to open them after medication administration (77). Approximately 66% of the medication administered in this fashion was absorbed. In addition, this method may increase compliance and make children more cooperative.

For the administration of an ophthalmic ointment to a child who can cooperate, the child should tilt his or her head backwards and look up. After the hands have been washed, the person to administer the medication should gently pull down the lower eyelid(s) for drug administration. A thin layer of ointment should then be placed in the lower eyelid(s). Afterwards, the eyelid(s) should be closed for 1 to 2 min to allow for the spreading of the medication and absorption. During this process, the tip of the ophthalmic applicator should not touch the eye. After administration is completed, the tip of the applicator tube should be cleaned with a clean tissue and be tightly capped (76). The package insert should also be reviewed for specific information.

Inhalers

For the use of an inhaled medication (e.g., β_2 -agonists, corticosteroids, antivirals, cromolyn, etc.), it is crucial for the child and parents to understand the mechanism of the metered dose inhaler (MDI) or nebulizer, if used. The package insert should also be reviewed for information about the specific drug product. A decision may also need to be made as to whether a spacer may be needed for use with the medication canister.

REFERENCES

1. Nahata, M.C. Pediatric Drug Formulations: Challenges and Potential Solutions. *Ann. Pharmacother.* **1999**, *33*, 247–249.
2. Food and Drug Administration. Regulations Requiring Manufacturers to Assess the Safety and Effectiveness of New Drugs and Biological Products in Pediatric Patients, 21 CFR Parts 201, 312, 314, and 601, Docket No. 97N-0165, U.S. Department of Health and Human Services; Washington, DC, August 17, 1997.
3. FDA Modernization Act of 1997 Public Law 105–115 105th Congress, Washington, DC, 1997.
4. Food and Drug Administration. List of Approved Drugs for Which Additional Pediatric Information May Produce Health Benefits in the Pediatric Population, Docket No. 98N-0056, U.S. Department of Health and Human Services; Washington, DC, May 20, 1998.

5. USP Member Memorandum, U.S. Pharmacopoeia, Rockville, MD, June 22, 2000.
6. Reed, M.D.; Gal, P. Principles of Drug Therapy. *Nelson Textbook of Pediatrics*, 16th Ed.; Behrman, R.E., Kliegman, R.M., Jenson, H.B., Eds.; W.B. Saunders Company: Philadelphia, 2000; 2229–2234.
7. Reed, M.D. The Ontogeny of Drug Disposition: Focus on Drug Absorption, Distribution, and Excretion. *Drug Inform. J.* **1996**, *30*, 1129–1134.
8. Milsap, R.L.; Hill, M.R.; Szeffler, S.J. Special Pharmacokinetic Considerations in Children. *Applied Pharmacokinetics: Principles of Therapeutic Drug Monitoring*, 3rd Ed.; Evans, W.E., Schentag, J.J., Jusko, W.J., Eds.; Applied Therapeutics: Vancouver, WA, 1992; 10.1–10.32.
9. Stewart, C.F.; Hampton, E.M. Effect of Maturation on Drug Disposition in Pediatric Patients. *Clin. Pharm.* **1987**, *6*, 548–564.
10. Christie, D.L. Development of Gastric Function during the First Month of Life. *Infancy*; Leblenthal, E., Ed.; Raven Press: New York, 1981; 109–120.
11. Cavell, B. Gastric Emptying in Preterm Infants. *Acta Paediatr. Scand.* **1979**, *68*, 725–730.
12. Cavell, B. Gastric Emptying in Infants Fed Human Milk or Infant Formula. *Acta Paediatr. Scand.* **1981**, *70*, 639–641.
13. Blumer, J.L.; Reed, M.D. Principles of Neonatal Pharmacology. *Pediatric Pharmacology: Therapeutic Principles in Practice*; Yaffe, S.J., Aranda, J.V., Eds.; W.B. Saunders Co.: Philadelphia, 1992; 164–177.
14. Leblenthal, E.; Lee, P.C.; Heitlinger, L.A. Impact of the Development of the GI Tract on Infant Feeding. *J. Pediatr.* **1983**, *101*, 1–9.
15. Kearns, G.L.; Reed, M.D. Clinical Pharmacokinetics in Infants and Children: A Reappraisal. *Clin. Pharmacokinet.* **1989**, *17* (Suppl. 1), 29–67.
16. Greenblatt, D.J.; Koch-Weser, J. Intramuscular Injection of Drugs. *N. Engl. J. Med.* **1976**, *295*, 542–546.
17. Evans, E.F.; Proctor, J.D.; Frantkin, M.J.; Velandia, J.; Wasserman, A.J. Blood Flow in Muscle Groups and Drug Absorption. *Clin. Pharmacol. Ther.* **1975**, *17*, 44–47.
18. Newton, M.; Newton, D.; Fudin, J. Reviewing the Big Three Injection Routes. *Nursing* **Feb 1992**, *92*, 34–42.
19. Bergeson, P.S.; Singer, S.A.; Kaplan, A.M. Intramuscular Injections in Children. *Pediatrics* **1982**, *70*, 944–948.
20. Sagraves, R.; Kamper, C. Controversies in Cardiopulmonary Resuscitation: Pediatric Considerations. *DICP, Ann. Pharmacother.* **1991**, *25*, 760–772.
21. Ghadially, R.; Shear, N.H. Topical Therapy and Percutaneous Absorption. *Pediatric Pharmacology: Therapeutic Principles in Practice*; Yaffe, S.J., Aranda, J.V., Eds.; W.B. Saunders Co.: Philadelphia, 1992; 72–77.
22. American Academy of Pediatrics Committee on Drugs. Alternative Routes of Drug Administration—Advantages and Disadvantages (Subject Review). *Pediatrics* **1998**, *100*, 143–153.
23. McRorie, T. Quality Drug Therapy in Children: Formulations and Delivery. *Drug Inform. J.* **1996**, *30*, 1173–1177.
24. Ward, J.T., Jr. Endotracheal Drug Therapy. *Am. J. Emerg. Med.* **1983**, *1*, 71–82.
25. de Boer, A.G.; Moolenaar, F.; de Leede, L.G.J.; Breimer, D.D. Rectal Drug Administration: Clinical Pharmacokinetic Considerations. *Clin. Pharmacokinet.* **1982**, *7*, 285–311.
26. Iob, V.; Swanson, W.W. Mineral Growth of the Human Fetus. *Am. J. Dis. Child* **1934**, *47*, 302–306.
27. Widdowson, E.; Spray, C.M. Chemical Development in Utero. *Arch. Dis. Child.* **1951**, *26*, 205–214.
28. Morselli, P.L. Clinical Pharmacokinetics in Neonates. *Clin. Pharmacokinet.* **1976**, *1*, 81–98.
29. Reed, M.D.; Besunder, J.B. Developmental Pharmacology: Ontogenic Basis of Drug Disposition. *Pediatr. Clin. North Am.* **1989**, *36*, 1053–1074.
30. Kearin, M.; Kelly, J.G.; O'Malley, K. Digoxin “Receptors” in Neonates: An Explanation of Less Sensitivity to Digoxin in Adults. *Clin. Pharmacol. Ther.* **1980**, *28*, 346–349.
31. Levy, G.; Khanna, N.N.; Soda, D.M.; Tsuzuki, O.; Stern, L. Pharmacokinetics of Acetaminophen in the Human Neonate: Formation of Acetaminophen Glucuronide and Sulfate in Relation to Plasma Bilirubin Concentration and D-Glucuronic Acid Excretion. *Pediatrics* **1975**, *55*, 818–825.
32. Siegel, S.R.; Oh, W. Renal Function as a Marker of Human Fetal Maturation. *Acta Paediatr. Scand.* **1976**, *65*, 481–485.
33. Siber, G.R.; Smith, A.L.; Levin, M.J. Predictability of Peak Serum Gentamicin Concentration with Dosage Based on Body Surface Area. *J. Pediatr.* **1978**, *94*, 135–138.
34. Painter, M.J.; Pippenger, C.E.; Wasterlein, C.; Barmada, M.; Pitlick, W.; Carter, G.; Abern, S. Phenobarbital and Phenytoin in Neonatal Seizures: Metabolism and Tissue Distribution. *Neurology* **1981**, *31*, 1107–1112.
35. Gilman, J.T.; Gal, P.; Duchowny, M.S.; Weaver, R.L.; Ransom, J.L. Rapid Sequential Phenobarbital Treatment of Neonatal Seizures. *Pediatrics* **1989**, *83*, 674–678.
36. Gilman, J.T. Therapeutic Drug Monitoring in Neonate and Paediatric Age Group. Problems and Clinical Pharmacokinetic Implications. *Clin. Pharmacokinet.* **1990**, *19*, 1–10.
37. Tserng, K.Y.; Takieddine, F.N.; King, K.C. Developmental Aspects of Theophylline Metabolism in Premature Infants. *Clin. Pharmacol. Ther.* **1983**, *33*, 522–528.
38. Lonnerholm, G.; Lindstrom, B.; Paalzow, L.; Sedin, G. Plasma Theophylline and Caffeine and Plasma Clearance of Theophylline During Theophylline Treatment in the First Year of Life. *Eur. J. Clin. Pharmacol.* **1983**, *24*, 371–374.
39. Dreifuss, F.E.; Santilli, N.; Langer, D.H.; Sweeney, K.P.; Moline, K.A.; Menander, K.B. Valproic Acid Hepatic Fatalities: A Retrospective Review. *Neurology* **1987**, *37*, 379–385.
40. Gershanik, J.J.; Boecler, B.; George, W.; Sola, A.; Leitner, M.; Kapadia, C. The Gasping Syndrome: Benzyl Alcohol (BA) Poisoning. *Clin. Res.* **1981**, *29*, 895A, Abstract.
41. Brown, W.J.; Buist, N.R.M.; Gipson, H.T.C.; Huston, R.K.; Kennaway, N.G. Fatal Benzyl Alcohol Poisoning in a Neonatal Intensive Care Unit. *Lancet* **1982**, *1*, 1250, Letter.
42. Gershanik, J.J.; Boecler, B.; Ensley, H.; McCloskey, S.; George, W. The Gasping Syndrome and Benzyl Alcohol Poisoning. *N. Engl. J. Med.* **1982**, *307*, 1384–1388.
43. Benda, G.I.; Hiller, J.L.; Reynolds, J.W. Benzyl Alcohol Toxicity: Impact on Neurologic Handicaps Among Surviving Very Low Birth Weight Infants. *Pediatrics* **1986**, *77*, 507–512.
44. Golightly, L.K.; Smolinske, S.S.; Bennett, M.L.; Sutherland, E.W., III; Rumack, B.H. Pharmaceutical Excipients. Adverse Effects Associated with Inactive Ingredients in Drugs Products (Part I). *Med. Toxicol.* **1988**, *3*, 128–165.

45. American Academy of Pediatrics Committee on Fetus and Newborn and Committee on Drugs. Benzyl Alcohol: Toxic Agent in Neonatal Units. *Pediatrics* **1983**, 72, 356–358.
46. Centers for Disease Control. Neonatal Death Associated with Use of Benzyl Alcohol—United States. *MMWR* **1982**, 31, 290–291.
47. Benzyl Alcohol May be Toxic to Newborns. *FDA Drug Bull.* **1982**, 12, 10–11.
48. Hiller, J.L.; Benda, G.I.; Rahatzad, M.; Alen, J.R.; Culver, D.H.; Carlson, C.V.; Reynolds, J.W. Benzyl Alcohol Toxicity: Impact on Mortality and Intraventricular Hemorrhage Among Very Low Birth Weight Infants. *Pediatrics* **1986**, 77, 500–506.
49. Jadine, D.S.; Rogers, K. Relationship of Benzyl Alcohol to Kernicterus, Intraventricular Hemorrhage, and Mortality in Preterm Infants. *Pediatrics* **1989**, 83, 153–160.
50. MacDonald, M.G.; Getson, P.R.; Glasgow, A.M.; Miller, M.K.; Boeckx, R.L.; Johnson, E.L. Propylene Glycol: Increased Incidence of Seizures in Low Birth Weight Infants. *Pediatrics* **1987**, 79, 622–625.
51. Glasgow, A.M.; Boeckx, R.L.; Miller, M.K.; MacDonald, M.G.; August, G.P. Hyperosmolality in Small Infants Due to Propylene Glycol. *Pediatrics* **1983**, 72, 353–355.
52. Berkeris, L.; Baker, C.; Fenton, J.; Propylene Glycol as a Cause of an Elevated Serum Osmolality. *Am. J. Clin. Pathol.* **1979**, 72, 633–636.
53. Fligner, C.L.; Jack, R.; Twiggs, G.A.; Raisys, V.A. Hyperosmolality Induced by Propylene Glycol. *JAMA* **1985**, 253, 1606–1609.
54. Kulick, M.I.; Lewis, N.S.; Bansal, V.; Warpeha, R. Hyperosmolality in the Burn Patient: Analysis of an Osmolal Discrepancy. *J. Trauma* **1980**, 20, 223–228.
55. American Academy of Pediatrics Committee on Drugs. “Inactive” Ingredients in Pharmaceutical Products. *Pediatrics* **1985**, 76, 635–643.
56. Louis, S.; Jutt, H.; McDowell, F. The Cardiocirculatory Changes Caused by Intravenous Dilantin and Its Solvent. *Am. Heart J.* **1967**, 74, 523–529.
57. American Academy of Pediatrics Committee on Drugs. Ethanol in Liquid Preparations Intended for Children. *Pediatrics* **1984**, 73, 405–407.
58. Rumack, B.H.; Spoerke, D.G.; *Poisindex Information System*; Micromedex, Inc.: Denver, 2000.
59. Hoyumpa, A.M.; Schenker, S. Major Drug Interactions. Effect of Liver Disease, Alcohol, and Malnutrition. *Ann. Rev. Med.* **1982**, 33, 113–149.
60. Kumar, A.; Rawlings, R.D.; Beaman, D.C. The Mystery Ingredients: Sweeteners, Flavorings, Dyes, and Preservative in Analgesic/Antipyretic, Antihistamine/Decongestant, Cough and Cold, Antidiarrheal, and Liquid Theophylline Preparations. *Pediatrics* **1993**, 91, 927–933.
61. Leff, R.D.; Roberts, R.J.; *Practical Aspects of Intravenous Drug Administration*; American Society of Hospital Pharmacists: Bethesda, MD, 1992.
62. Nahata, M.C. Intravenous Infusion Conditions: Implications for Pharmacokinetic Monitoring. *Clin. Pharmacokinet.* **1993**, 24, 221–229.
63. Gould, T.; Roberts, R.J. Therapeutic Problems Arising from the Use of the Intravenous Route for Drug Administration. *J. Pediatr.* **1979**, 95, 465–471.
64. Rajchgot, P.; Radde, I.C.; MacLeod, S.M. Influence of Specific Gravity on Intravenous Drug Delivery. *J. Pediatr.* **1981**, 99, 658–661.
65. Santeiro, M.L.; Sagraves, R.; Allen, L.V. Osmolality of Small-Volume I.V. Admixtures for Pediatric Patients. *Am. J. Hosp. Pharm.* **1990**, 47, 1359–1364.
66. Bloss, C.S.; Sybert, K. Osmolality of Commercially Available Oral Liquid Drug Preparations. In *ASHP Annual Meeting June 1991*; 48, P-35 Abstract.
67. Maish, W.; Sagraves, R. Childhood Asthma. *U.S. Pharmacist* **1993**, Jan, 36–58, 105.
68. Hendeles, L.; Weinberger, M.; Szeffler, S. Safety and Efficacy of Theophylline in Children with Asthma. *J. Pediatr.* **1992**, 120, 177–183.
69. Hendeles, L.; Weinberger, M. Selection of a Slow-Release Theophylline Product. *J. Allergy Clin. Immunol.* **1986**, 78, 743–751.
70. Martin, S. Catering to Pediatric Patients. *Am. Pharm.* **1992**, 32, 47–50.
71. Chater, R.W. Pediatric Dosing: Tips for Tots. *Am. Pharm.* **1993**, 33, 55–56.
72. Kraus, D.M.; Stohlmeyer, L.A.; Hannon, P.R. Infant Acceptance and Effectiveness of a New Oral Liquid Medication Delivery System. *Am. J. Health-Syst. Pharm.* **1999**, 56, 1094–1101.
73. Gilles, F.H.; Matson, D.D. Sciatic Nerve Injury Following Misplaced Gluteal Injection. *J. Pediatr.* **1970**, 76, 247–254.
74. Administration Guides. *Drug Store News* **1993**, Dec 13, 7–14.
75. Otic Preparations Monograph. *Facts and Comparisons*; Olin, B.R., Ed.; J.B. Lippincott Co.: St. Louis, November 1992; 517.
76. Topical Ophthalmic Monograph. *Facts and Comparisons*; Olin, B.R., Ed.; J.B. Lippincott Co.: St. Louis, May 1993; 477b–477c.
77. Smith, S.E. Eyedrop Instillation for Reluctant Children. *Br. J. Ophthalmol.* **1991**, 75, 480–481.

BIBLIOGRAPHY

- Benitz, W.E.; Tatro, D.S.; Davis, S., Eds. *The Pediatric Drug Handbook*, 3rd Ed.; Harcourt Health Sciences Group: St. Louis, 1995.
- Leff, R.D.; Roberts, R.J. *Practical Aspects of Intravenous Drug Administration*; American Society of Health-System Pharmacists: Bethesda, 1992.
- Nahata, M.C.; Hipple, T.F. *Pediatric Drug Formulations*, 4th Ed.; Harvey Whitney Books: Cincinnati, 2000.
- Nelson, J.; Bradley, J. *Nelson's 2000–2001 Pocket Book of Pediatric Antimicrobial Therapy*, 14th Ed.; Lippincott Williams & Wilkins: Philadelphia, 2000.
- Phelps, S.; Hak, E. *Guidelines for Administration of Intravenous Medications to Pediatric Patients*, 6th Ed.; American Society of Health-System Pharmacists: Bethesda, 1999.
- Siberry, G.K.; Iannone, Eds.; *Harriett Lane Handbook*, 15th Ed.; Mosby: St. Louis, 2000.
- Taketomo, C.K.; Hodding, J.H.; Kraus, D.M.; *Pediatric Dosage Handbook*, 5th Ed.; Lexi-Comp., Inc.: Hudson, OH, 2000.
- Young, T.; Mangum, B.; *Neofax 2000*, 13th Ed.; Blackwell Science: Oxford, 2000.

PELLETIZATION TECHNIQUES

Isaac Ghebre-Sellassie

Pfizer Global Research and Development, Morris Plains, New Jersey, U.S.A.

Axel Knoch

Pfizer Global Research and Development, Freiburg, Germany

INTRODUCTION

Historically, the term *pellet* has been used by a number of industries to describe a variety of agglomerates produced from diverse raw materials, using different pieces of manufacturing equipment. These agglomerates include fertilizers, animal feeds, iron ores, and pharmaceutical dosage units and thus do not only differ in composition but also encompass different sizes and shapes. As a result, pellets meant different things for different industries. In the pharmaceutical industry, pellets can be defined as small, free-flowing, spherical particulates manufactured by the agglomeration of fine powders or granules of drug substances and excipients using appropriate processing equipment. The term also has been used to describe small rods with aspect ratios of close to unity. Although pellets have been used in the pharmaceutical industry for more than 4 decades, it has only been since the late 1970s, with the advent of controlled-release technology, that the advantages of pellets over single-unit dosage forms have been realized.

Pellets offer a high degree of flexibility in the design and development of oral dosage forms. They can be divided into desired dose strengths without formulation or process changes and also can be blended to deliver incompatible bioactive agents simultaneously and/or to provide different release profiles at the same or different sites in the gastrointestinal (GI) tract. In addition, pellets, taken orally, disperse freely in the GI tract, maximize drug absorption, minimize local irritation of the mucosa by certain irritant drugs, and reduce inter- and inpatient variability (1).

Given the enormous advantages of multiparticulate systems over single-unit oral dosage forms, extensive research has focused recently on refining and optimizing existing pelletization techniques as well as on the development of novel manufacturing approaches that use innovative formulations and processing equipment. The most commonly used and intensely investigated

pelletization processes are powder layering, solution/suspension layering, and extrusion-spheronization and are defined first and then addressed in detail.

Powder layering involves the deposition of successive layers of dry powder of drug or excipients or both on preformed nuclei or cores with the help of a binding liquid. Because powder layering involves the simultaneous application of the binding liquid and dry powder, it generally requires specialized equipment. The primary equipment-related requirement in a powder-layering process is that the product container should have solid walls with no perforations to avoid powder loss beneath the product chamber before the powder is picked up by the wet mass of pellets that is being layered on.

Solution/suspension layering involves the deposition of successive layers of solutions and/or suspensions of drug substances and binders on starter seeds, which may be inert materials or crystals/granules of the same drug. In principle, the factors that control coating processes apply to solution or suspension layering and, as a result, require basically the same processing equipment. Consequently, conventional coating pans, fluid-bed centrifugal granulators, and Wurster coaters have been used successfully to manufacture pellets. The efficiency of the process and the quality of pellets produced are in part related to the type of equipment used.

Extrusion-spheronization is a multistep process involving dry mixing, wet granulation, extrusion, spheronization, drying, and screening. The first step is dry mixing of the drug and excipients in suitable mixers followed by wet granulation, in which the powder is converted into a plastic mass that can be easily extruded. The extruded strands are transferred into a spheronizer, where they are instantaneously broken into short cylindrical rods on contact with the rotating friction plate and are pushed outward and up the stationary wall of the processing chamber by centrifugal force. Finally, owing to gravity, the particles fall back to the friction plate, and the cycle is repeated until the desired sphericity is achieved.

Other pelletization processes that either have limited application or are still at the development stage include

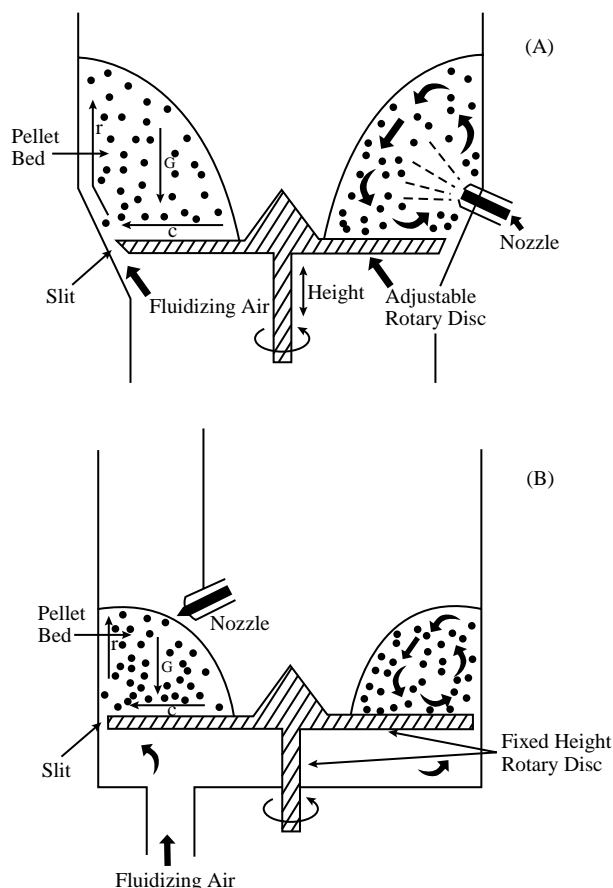


Fig. 1 Schematic representation of centrifugal fluid-bed equipment and process with a single-walled product chamber. (A) Glatt GPCG and GRG Granulators. (B) Freund CF-Granulators. (Adapted from Ref. 5).

spherical agglomeration or balling, spray congealing/drying, and emerging technologies such as cryopelletization and melt spheronization. These processes are addressed briefly.

POWDER LAYERING

The first equipment used to manufacture pellets on a commercial scale was the conventional coating pan, a machine that has been used by pharmaceutical firms, primarily for sugar coating, for a long time. During the 1950s, the industry, in an attempt to prolong the release of drugs from solid oral dosage forms, explored new processes that allowed the manufacture of multiparticulate drug-delivery systems. The turning point came when candy seeds, which had been used for topping decorations

in foodstuffs such as pastries, were used as starter seeds to develop and manufacture sustained-release pellets using conventional coating pans (2). The process was an extension of a procedure that was used to manufacture the candy seeds themselves and involved the successive layering of powder and binder solution on sugar crystals (3). Consequently, the conventional coating pan became the first pharmaceutical equipment used not only to manufacture nonpareils but also to develop sustained-release products of a number of prescription drugs using nonpareils as starter seeds.

Conventional coating pans, however, have had significant limitations as pelletization equipment. The degree of mixing is very poor, and the drying process is not efficient. Mixing is a function of the pan shape, the tilt angle, the baffle arrangement, and the rotational speed of the pan itself. These parameters must be optimized to provide uniform drying and sufficient particle movement to eliminate the potential formation of dead spots during the operation and to maximize yield. For instance, during pelletization, elliptical pans tend to have fewer stagnant spots than do cylindrical pans and, consequently, are the equipment of choice (4). Reducing the tilt angle can also minimize formation of dead spots. If the rotational speed of the pan is too slow, segregation may occur owing to percolation and induce the preferential layering of drug onto larger particles. In addition, prolonged contact time among the particles could favor particle agglomeration if the liquid feed rate leads to surface wetness and stickiness that induce coalescence.

During powder layering, a binding solution and a finely milled powder are added simultaneously to a bed of starter seeds at a controlled rate. In the initial stages, the drug particles are bound to the starter seeds and subsequently to the forming pellets with the help of liquid bridges originated from the sprayed liquid. These liquid bridges are eventually replaced by solid bridges derived either from a binder in the application medium or from any material, including the drug substance, that is soluble in the liquid. Successive layering of the drug and binder solution continues until the desired pellet size is reached. Throughout the process, it is extremely important to deliver the powder accurately at a predetermined rate and in a manner that maintains equilibrium between the binder liquid application rate and the powder delivery rate. If the powder delivery rate is not maintained at predetermined equilibrium levels, overwetting or dust generation may occur, and neither the quality nor the yield of the product can be maximized. Toward the end of the layering process, it is likely that fines may be generated owing to potential interparticle and wall-to-particle friction and appear in the final product, thereby lowering the yield. The problem can

be overcome if the application medium is sprayed on the cascading pellets at the end of the layering process to increase the moisture level at the pellet surface and facilitate layering of the fines onto the pellets. Caution must be exercised, however, not to overwet the product bed because the powder-liquid equilibrium that has been established will change and will need to be adjusted accordingly. In an ideal process, no agglomeration occurs, and the particle population at the end of the process remains the same as that of the starter seeds or cores, with the only difference being an increase in the size of the pellets and thus in the total mass in the pan.

Pieces of equipment that overcame the limitations of coating pans and revolutionized powder-layering processing as a pelletization technique are tangential spray or centrifugal fluid-bed granulators. Although tangential spray equipment was originally developed to perform granulation processes, its application was later expanded

to cover other unit operations including the manufacture and coating of pellets.

Although there are variations in the design of centrifugal or rotary granulators, the basic operational principle that determines the degree of mixing and thus the efficiency of the process remains the same and includes centrifugal force, fluidization air velocity, and gravitational force (Fig. 1). During a layering process, these three forces act in concert to generate a spiral, rope-like motion of the particles in the product bed. The rotating disk, which may have fixed or variable speeds, creates a centrifugal force that pushes the particles outward to the vertical wall of the product chamber or stator. The fluidization air, which is directed toward the slit between the periphery of the disk and the stator, generates a force that carries the particles vertically along the wall of the product container into the expansion chamber. The particles lose their momentum and cascade down toward the center of the rotating disk owing to gravitational force. The cycle repeats itself, bringing about a thorough mixing unparalleled by any other powder-layering equipment. The degree of mixing depends on the fluidization air volume and velocity, the slit width, the bed size, and the disk speed. These variables, as well as liquid and powder application rates, atomization air pressure, fluidization air temperature, and degree of moisture saturation determine the yield and quality of pellets.

As mentioned earlier, the key process variable that determines the success of any pelletization process in centrifugal fluid-bed granulators is the degree of mixing, which is partly dictated by the radial velocity of the disk. At low radial velocities, the extent of mixing becomes inadequate, as indicated by the loss in the spiral rope-like motion of the particles. Thus, the rate at which the particles traverse the spray zone is prolonged and could potentially lead to excessive agglomeration. Caking is another serious problem encountered during powder layering at low radial velocities. At high radial velocities, the particle-to-particle and particle-to-wall frictional forces become intense with a very rapid pellet turnover. Generally, this condition leads to an uncontrollable, wobbly bed, resulting in severe particle attrition; that is, the breaking forces overcome the forces that contribute to particle growth through layering, and the pelletization process cannot proceed as intended. In such a situation, not only is a large amount of fines generated, but some of the attrited particles may agglomerate to form nuclei that are subsequently layered upon. Because the drug content of the pellets that contain inert starter seeds and those that use the newly formed nuclei are different, the process may create content uniformity problems. The optimum radial velocities of the disk are generally 3–8 m/s irrespective

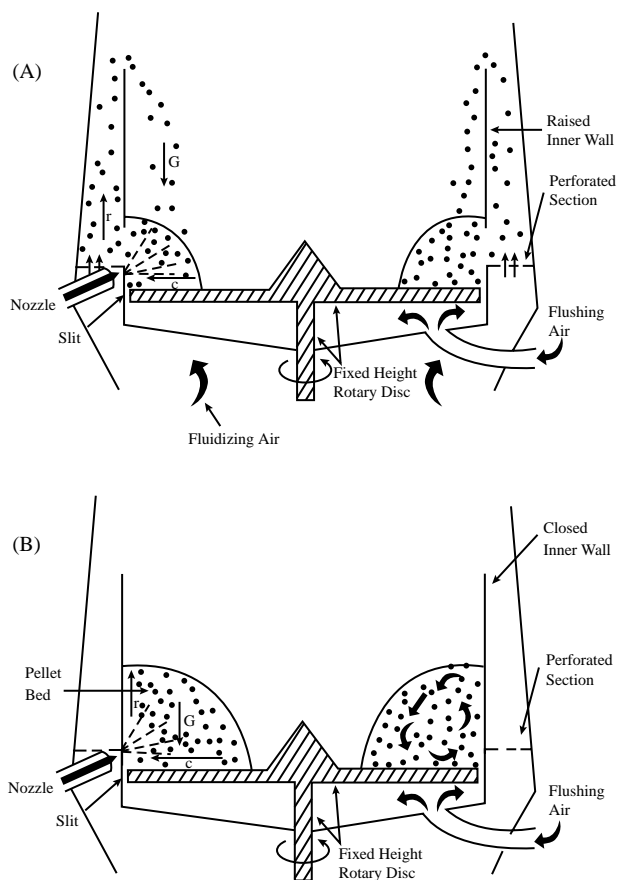


Fig. 2 Schematic representation of centrifugal fluid-bed equipment and process with a double-walled product chamber (Niro-Aeromatic). (A) open position. (B) closed position. (Adapted from Ref. 5.)

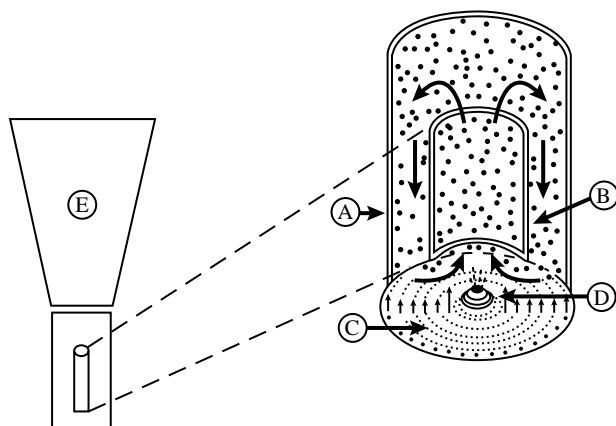


Fig. 3 Schematic representation of the Wurster product chamber and process. (A) product chamber. (B) partition. (C) orifice plate. (D) nozzle. (E) expansion chamber. (Adapted from Ref. 6.)

of the granulator size (6). At the initial stages of the pelletization process, the radial velocity may be kept low and increased as the mass of the batch size increases. After the layering process is completed, the radial disk velocity is usually reduced to avoid particle attrition during the drying step.

Another parameter that plays a key role during layering is the disk clearance or slit width. In some centrifugal granulators, the slit width is fixed, and the fluidization air velocity can be varied only with a change in the fluidization air volume. In others, the slit width is variable and is adjusted to meet the processing requirements. As the disk

clearance is increased, the air velocity decreases, and pellet turnover is reduced. Conversely, as disk clearance decreases, the air velocity increases, resulting in a rapid pellet turnover. Therefore, for maximum process efficiency, the disk clearance is chosen not only to avoid loss of pellets through the gap between the disk and the vertical wall into the plenum and to minimize the loss of powder into the exhaust system but also to generate an air velocity that provides the desired degree of pellet turnover.

Aside from the rotary disk, the other unique feature of the centrifugal equipment is the spray method. During layering, the liquid is sprayed tangentially to and concurrent with particle movement. This feature is primarily responsible for the high yield that is typically obtained from a process involving this type of equipment. The distance the droplets travel before they impinge on the particles is short, and consequently, the droplets are picked up by the particles almost completely with little, if any, loss to the wall of the product chamber or because of spray drying, assuming that the other process variables are optimized.

In addition, powder and binder liquid delivery rates are critical parameters that must be carefully evaluated. Because the surface moisture/solvent determines the extent of binding between the forming pellets and the powder being introduced, the powder delivery and the liquid application rates should be adjusted to establish an equilibrium conducive to layering. It is essential that the powder delivery remains well-controlled during processing, whether airflow is mediated by suction

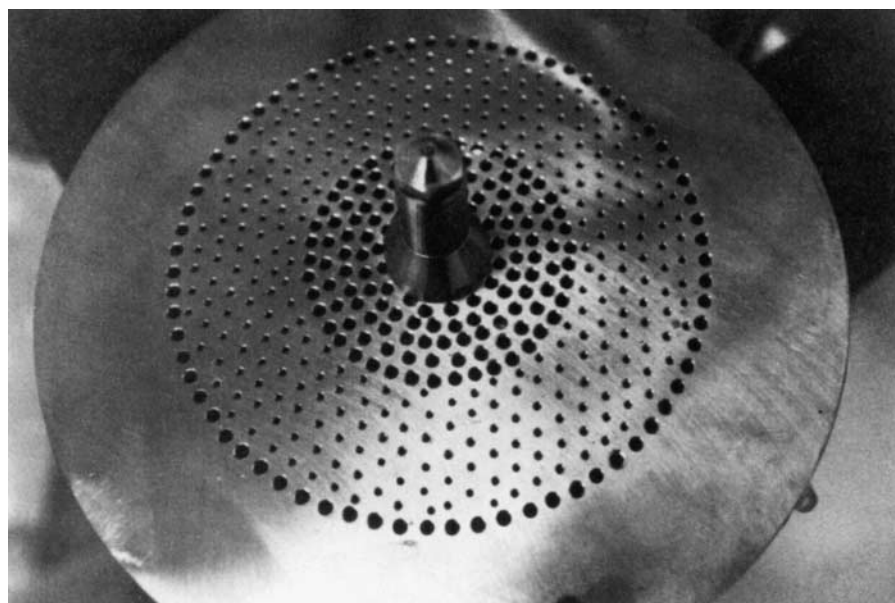


Fig. 4 Air distributor or orifice plate of a Wurster coater. (From Ref. 5.)

(negative pressure), as in the Glatt centrifugal granulators, or by positive air flow, as with the Freund CF granulators.

With a double-walled centrifugal granulator, the process is carried out with the inner wall in the open or closed position (Fig. 2) (7). With powder layering, the inner wall is closed so that simultaneous application of liquid and powder could proceed until the pellets have reached the desired size. The inner wall is then raised, and the spheres enter the drying zone. The pellets are lifted by the fluidization air up and over the inner wall back into the forming zone. The cycle is repeated until the desired residual moisture level in the pellets is achieved.

Scaling up of a layering process from the laboratory to production equipment is straightforward. As with all fluid-bed equipment, the most critical parameters that need to be met, in addition to the spiral rope-like motion of the particles, are the density of the particles and the droplet size within the spray zone. Because more than one spray gun is used in production equipment, it is imperative that the guns are positioned in such a way that optimum drying of the layered drug is attained before the pellets entering the next spray zone. Generally, the rate-limiting step in the process is saturation of the fluidization air by the application medium. Sometimes, it is likely that the spiral rope-like motion of the particles could be lost during scale up, and baffles must be inserted to enhance particle motion) (5).

As for formulation components, they also must meet certain requirements. Binder solutions must have a high binding capacity. Micronizing or finely milling the drug before layering improves the efficiency of the layering process significantly and provides morphologically smooth pellets that are suitable for film coating. However, in the majority of cases, micronization tends to impact flow and thus the delivery rate, a critical process parameter. Therefore, it is likely that during processing, powders may adhere to the sides of the hopper or the feed screw and may even form rat holes within the hopper. To improve the flow properties of the drug substance, glidants are incorporated into the powder before processing. Chemically, glidants could be hydrophobic or hydrophilic and are chosen based on the type of formulation selected.

Finally, the rheological properties of the binding liquid, the liquid application rate, and drying air temperature should be optimized to produce the desired product temperature. In addition, the powder should be delivered at a rate that maintains a balance between the surface wetness of the cores and powder adhesion. If the product bed temperature is high, powder is lost to the exhaust system. If it is too low, it leads to agglomeration and/or the formation of new nuclei that serve as cores for further layering. If that happens, content uniformity of the product

could be severely impacted. Therefore, the product temperature should be optimized to keep the particle population the same throughout the layering and subsequent drying steps.

SOLUTION/SUSPENSION LAYERING

The Wurster coating process, which was invented about 30 years ago, had evolved through elaborate design modifications and refinement into ideal equipment for the manufacture of pellets by solution/suspension layering. The high drying efficiency inherent in fluid-bed equipment, coupled with the innovative and efficient design features of the Wurster process, has allowed the machines to hold center stage in pharmaceutical processing technology. Not only have the manufacture and coating of pellets become routine and efficient, but scaling up of the process, which is key to the viability of any processing technique, has proved to be predictable and economically feasible.

The primary features that distinguish Wurster equipment from other fluid-bed equipment are the cylindrical partition located in the product chamber and the configuration of the air distributor plate, also known as the orifice plate (Figs. 3 and 4). The latter is configured to allow most of the fluidization or drying air to pass at high velocity around the nozzle and through the partition, carrying with it the particles that are being layered on. Once the particles exit the partition, they enter the expansion chamber, where the velocity of the air is reduced below the entrainment velocity, and the particles fall back to the area surrounding the partition (referred to as the down bed). The down bed is kept aerated by the small fraction of air that passes through the small holes on the periphery of the orifice plate. The particles in the down bed are transported horizontally through the gap between the air distributor plate and the partition by suction generated by the high air velocity that prevails around the nozzle and immediately below the partition. The volume of air that passes through the down bed outside the partition is just enough to generate modest particle movement. Because the spray direction is concurrent with particle movement, and particle motion is well-organized under optimum conditions, uniform layering of drug substances is consistently achieved. Because the partition height, that is, the gap between the partition and the orifice plate, controls the rate at which the particles enter the spray zone, it is an important variable that needs to be optimized for every batch size. For instance, at a given load size and fluidization air

volume, the partition height can be reduced or increased to provide either a well-controlled particle motion that produces the desired pellet movement or a bubbling down bed that leads to disorganized particle movement and inefficient process.

The disadvantage of the Wurster process is the inaccessibility of the nozzles. If the nozzles are clogged at any time during the layering process, the operation has to be interrupted, and the spray guns must be removed for cleaning. The problem can be alleviated by screening the formulation or by using a spray gun with a bigger nozzle. Another aspect of the process that is challenging when multiple nozzles are used is the potential overlap of adjacent spray zones. Although the position of the nozzle is fixed, the spray zone overlap can be minimized using the air cap at the end of the spray gun.

During scale up, the number of partitions in the product container is increased with an increase in batch size; the diameter of the partitions, however, is kept the same. The intent is to maintain the particle movement and processing dynamics that were established during the development phase when a single partition was used. The fluidization air volume is increased to compensate for the increase in the number of spray zones and consequently to create well-organized particle movement in all partitions. If the fluidization air volume is too high,

it is likely to produce a disorganized bed that in turn adversely affects the layering process. The fluidization air volume is also critical because it, rather than the size of the batch, dictates the liquid application rate.

Another piece of equipment that has been used to manufacture pellets by solution/suspension layering is the tangential spray or centrifugal granulator. An important parameter that needs to be established early in the process is quantity of starter seeds or cores charged into the machine that should, at the minimum, cover the nozzles during start-up; otherwise, the sprayed liquid droplets will either be sprayed onto the wall of the product container or become entrained in the fluidization air. The latter could lead to clogging of the filters, which would then not only increase the pressure differential across the filters and compromise the batch but also may complicate equipment cleanup. In either case, the yield, and probably the quality of the pellets, will be reduced dramatically. As the size of the forming pellets increases, the mass in the bed increases, and the fluidization air volume is continuously increased to provide optimum expansion and mixing of the product bed, as indicated by the spiral rope-like motion of the bed. If such particle movement is not achieved, the bed load may be excessive, and the process may have to be terminated. During layering, the product bed expands both vertically and horizontally, and

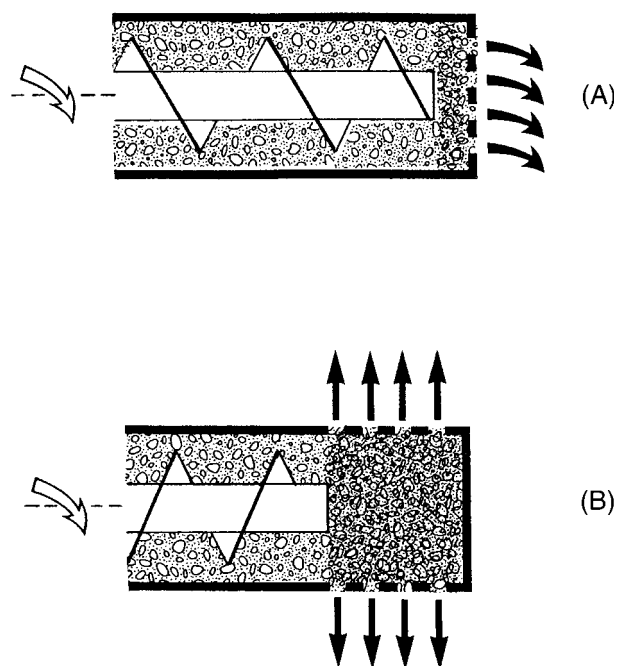


Fig. 5 Schematic representation of screw-fed extruders: (A) axial extruder. (B) radial extruder.

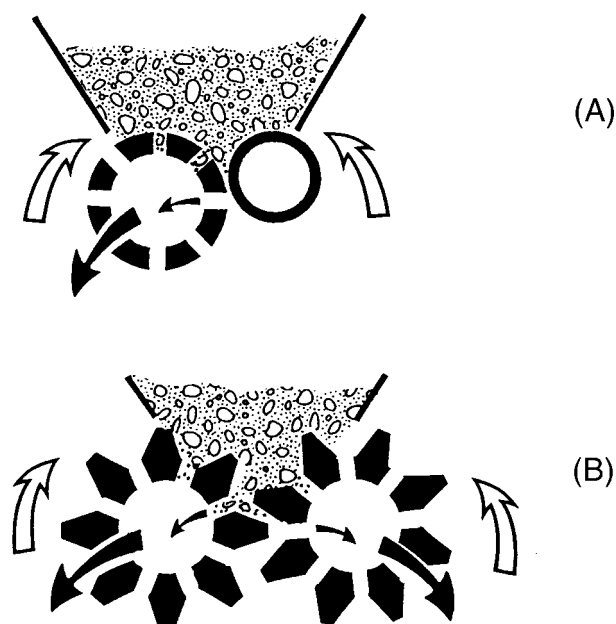


Fig. 6 Schematic representation of gravity-fed extruders: (A) rotary-cylinder extruder. (B) rotary-gear extruder.

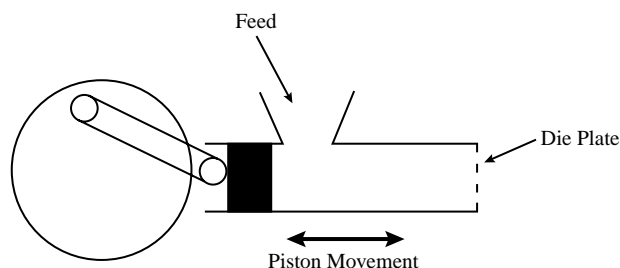


Fig. 7 Schematic representation of a ram extruder. (From Ref. 9.)

consequently a several-fold increase in batch weight can be realized in a single step.

During processing, all the components of the formulation are first dissolved or suspended in an appropriate quantity of application medium to provide a formulation with the desired viscosity and is then sprayed onto the product bed. The sprayed droplets immediately impinge on the starter seeds and spread evenly on the surface, provided the drying conditions and fluid dynamics are favorable. This is followed by a drying phase that renders dissolved materials to precipitate and form solid bridges that would hold the formulation components together as successive layers on the starter seeds. The process continues until the desired quantity of drug substance and thus the target potency of the pellets are achieved. The rate of particle growth is rather slow and is limited by the rate of solvent removal. Furthermore, the sprayed droplets must have the necessary rheological properties to spread evenly over the cores, as with coating processes. This must be followed immediately by a very rapid evaporation of the application medium. The evaporation rate should neither impair binder effectiveness nor lead to overwetting and subsequent agglomeration. Ideally, no new nuclei are formed, and the particle population remains the same; however, the sizes of the pellets increase as a function of time, and as a result, the total mass of the system also increases.

Although optimization of process variables is critical for the successful development of a pelletized product, pellets can only be manufactured routinely on a large scale if the formulation is not sensitive to slight variations in processing parameters. Therefore, it is imperative that the formulation characteristics are carefully identified and optimized, both qualitatively and quantitatively, during the development phase. These include drug solubility, type and concentration of binder, and viscosity of the solution/suspension. The working viscosity range usually determines the solid content of the formulation that can be sprayed successfully onto the starter seeds and forming pellets.

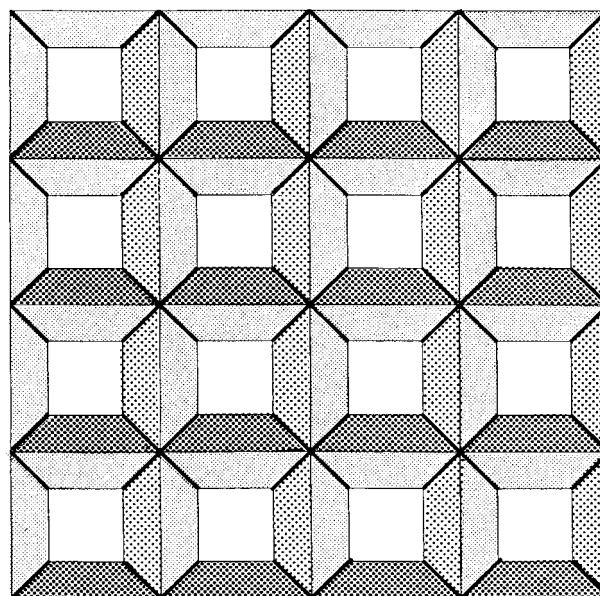


Fig. 8 Schematic representation of a spheronizer friction plate with a cross-hatch pattern.

Solution/suspension layering is usually used when the desired drug loading of the pellets is low because production of high-potency pellets from a low solids content formulation is not economically feasible. An important factor that needs to be considered when suspensions are used as opposed to solutions is the particle size of the drug. Micronized drug particles tend to provide pellets that are smooth in appearance, a property that is extremely desirable during subsequent film coating, particularly for controlled-release applications. If the particle size of the drug in the suspension is large, the amount of binder required to immobilize the particles onto the cores will be high, and, consequently, pellets of low potency are produced. The morphology of the finished pellets also tends to be rough and may adversely affect the coating process and the coated product. Moreover, because particles detach easily from the core they are being layered on owing to frictional forces, yield is usually low.

Although it is possible to manufacture pellets from a formulation that does not contain binders, almost invariably, the layers of drug applied tend to delaminate or break off from the cores in the later stages of the layering process or in the subsequent drying step. Therefore, binders are consistently used during solution/suspension layering to impart strength to the pellets. They are usually low-molecular-weight polymers that are compatible with the drug substance. They should not increase the viscosities of the formulations appreciably and should not, unless

intended to do so, modify the release characteristics of the pellets.

EXTRUSION–SPHERONIZATION

Extrusion–spheronization as a pelletization technique was developed in the early 1960s and since then has been researched and discussed extensively. Interest in the technology is still strong, as witnessed by the extent of coverage of the topic in scientific meetings and symposium proceedings, as well as in the scientific literature. The technology is unique in that it is not only suitable for the manufacture of pellets with a high drug loading but it also can be used to produce extended-release pellets in certain situations in a single step and thus can obviate the need for subsequent film coating.

Extrusion–spheronization is a multistep process involving a number of unit operations and equipment. However, the most critical pieces of processing equipment that, in effect, dictate the outcome of the overall process are the extruders and the spheronizers (8).

A variety of extruders, which differ in design features and operational principles, are currently on the market and can be classified as screw-fed extruders, gravity-fed extruders, and ram extruders.

Screw-fed extruders have screws that rotate along the horizontal axis and hence transport the material horizontally; they may be axial or radial screw extruders (Fig. 5). Axial extruders, which have a die plate that is positioned axially, consist of a feeding zone, a compression zone, and an extrusion zone. The product temperature is controlled during extrusion by jacketed barrels. In radial extruders, the transport zone is short, and the material is extruded radially through screens mounted around the horizontal axis of the screws.

Gravity-fed extruders include the rotary cylinder and rotary gear extruders, which differ primarily in the design of the two counter-rotating cylinders (Fig. 6). In the rotary-cylinder extruder, one of the two counter-rotating cylinders is hollow and perforated, whereas the other cylinder is solid and acts as a pressure roller. In the so-called rotary-gear extruder, there are two hollow counter-rotating gear cylinders with counterbored holes.

In ram extruders, a piston displaces and forces the material through a die at the end (Fig. 7). Ram extruders are preferred during formulation development because they are designed to allow for measurement of the rheological properties of formulations (9–11).

Because extruders were initially developed to serve industries other than the pharmaceutical industry, they were

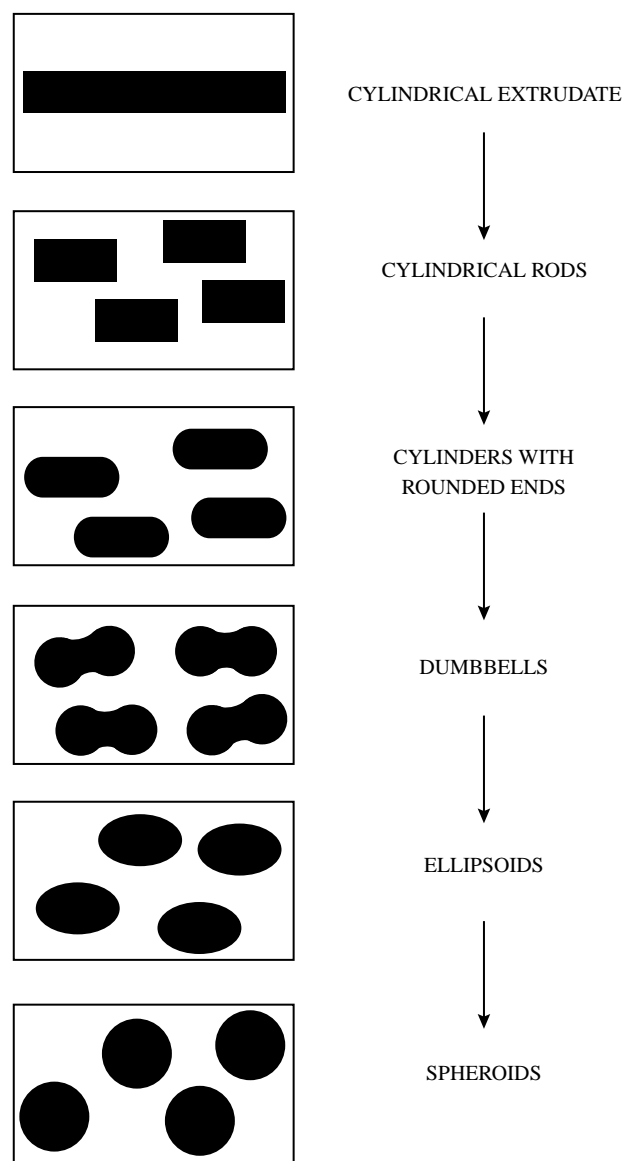


Fig. 9 Shape transitions during a spheronization process.

designed without GMP considerations. Therefore, changes are made to ensure pharmaceutical regulatory compliance, including qualification requirements. At the minimum, all the parts of the extruder that come in contact with the product must be of high-quality stainless steel, and the construction must be such that cleaning requirements are easily met. Whenever possible, the machine should permit easy monitoring and documentation of all critical process parameters such as pressure at the die plate, product temperature, power consumption or torque of the drive unit, and rotational speed of the screw. Some of the GMP requirements surrounding extruders have been

highlighted in an article that described a qualification procedure for an extruder that is suitable for the manufacture of clinical samples (12).

The extrusion/spheronization process was first introduced to the pharmaceutical industry in 1964 with the invention of the marumerizer (13). Since then, significant improvements have been made to the machine, and currently well-designed marumerizers of different sizes are commercially available. A marumerizer or spheronizer consists of a static cylinder or stator and a rotating friction plate at the base. The stator can be jacketed for temperature control. The friction plate, which has a grooved surface, is the most critical part of the equipment that initiates the spheronization process. A typical friction plate has a crosshatch pattern, where the grooves intersect at a 90° angle, as shown in Fig. 8. The groove width is selected based on the desired pellet diameter. Usually, groove diameters 1.5–2 times the target pellet diameter are used. The diameter of the friction plate is approximately 20 cm for laboratory-scale equipment or up to 1.0 m for production-scale units. The rotational speed of the friction plate is variable and ranges from 100 to 2000 rpm, depending on the diameter of the unit.

A new variation of spheronizers that was introduced into the market are the so-called air-assisted spheronizers. Basically, they are similar to the standard spheronizers except that they are designed to permit a conditioned air stream to pass from beneath the rotating disk through the gap or slit between the cylindrical wall and the

rotating friction plate into the product bed. The addition of such a feature presumably improves pellet turnover and brings about a spiral rope-like motion that facilitates spheronization.

Axial-type extruders tend to produce extrudates with slightly higher densities. Twin-screw extruders have better material transport characteristics and higher capacity or throughput than do single screw extruders. Radial-type extruders have higher throughput but produce less dense extrudates than those obtained from axial-type extruders. The product temperature in radial extruders increases very little during the extrusion process, probably because of a shorter compression zone and shorter die opening depth. Absence of heat build-up during extrusion leads not only to well-controlled spheronization process but also allows for the processing of thermolabile drug substances.

Although the extrusion step is a continuous process with a very high throughput, the subsequent steps (spheronization, drying, and sizing) are batch processes, and thus are rate limiting. As a result, the overall extrusion–spheronization process is a batch rather than a continuous process. However, a semicontinuous process can be implemented by having an extruder feed, alternatively, into two spheronizers. As one of the spheronizers is spheronizing the extrudates, the other discharges the formed pellets and is recharged with fresh extrudates. The process is then repeated as the roles of the spheronizers are reversed alternately.

In a batch process, a defined quantity of extrudate is fed into the spheronizer from the top, and the spheronized

Table 1 Critical factors in extrusion–spheronization

	Characteristic	Significance ^a	References
Drug substance	Particle size	+++	
	Particle size distribution	++	
	Particle shape	+++	
	Solubility	++	(26)
Formulation	Water content	+++	(23–25, 27, 28)
	Water temperature	+	(25)
	Excipients type	+++	(13, 18, 19, 24, 19, 26)
	Excipients concentration	++	(24, 26)
	Excipients particle size	++	(12, 13, 28)
Extrusion	Extruder type	++	(13)
	Extruder speed	++	(14, 15, 22–23, 25)
	Extrusion screen size	+++	(22, 24)
	Thickness of the die plate	+	(14, 15, 30)
Spheronization	Spheronizer speed (rpm)	+++	(23–25, 27, 25, 29)
	Spheronizer load	+	(23, 24)
	Spheronization time	+++	(23–25)
	Friction plate design	+	

^aRelative significance: +, low; ++, medium; +++, high.

particles are discharged by centrifugal force via a discharge chute located at the base of the stator or vertical wall of the cylinder. The extrudates are spheronized by interparticle collisions and particle-to-wall frictional forces. The various stages of the spheronization process are shown in Fig. 9. The spheronizing time is usually 2–15 min, depending on the formulation characteristics. Processing time remains constant, provided the composition of the extrudate, including the water content, is kept constant. Because relatively large amounts of water or solvent are incorporated into the formulation, the final pellets contain significant quantities of residual moisture or solvent and are oven-dried or dried in a fluid-bed dryer before further processing. A sizing step might be necessary to separate the fractions if the particle size distribution is wider than intended. In general, the pellets are spherical and have a narrow particle size distribution. The most critical process parameters in the spheronization step that influence the yield and quality are the design of the grooves and rotational speed of the friction plate and the residence time in the spheronizer.

In an extrusion–spheronization process, formulation components such as fillers, lubricants, and pH modifiers play a critical role in producing pellets with the desired attributes. The granulated mass must be plastic and sufficiently cohesive and self-lubricating during extrusion. During the spheronization step, it is essential that the extrudates break at appropriate length and have sufficient surface moisture to enhance formation of uniform spherical pellets.

The degree of liquid saturation of the granulation is another critical factor that needs to be optimized. Granulations containing low moisture content may generate extrudates that produce large quantities of fines during the spheronization step. If the moisture level is too high, the extrudates may adhere to each other and form bundles of strands that cannot be processed further. Therefore, the extrudates must have sufficient mechanical strength to form strands during the extrusion but must also be easily broken into uniform rods during spheronization to provide pellets with a narrow particle size distribution. Generally, the liquid content of the wet powder mixture is approximately 20–30% (w/w). Solvents such as ethanol or mixtures of water and ethanol may be used as granulating liquids when pure water cannot be used for stability or solubility reasons.

Excipients play a more significant role during extrusion–spheronization than during with any other pelletization process (14). They facilitate extrusion and determine the sphericity of the wet pellets; they also impart strength and integrity to the pellets. Microcrystalline cellulose is one of the most widely investigated

excipients. It is used as filler and a spheronization aid, regulating the water content and distribution in the granulation. In effect, it modifies the rheological properties of the formulation and imparts plasticity to the pellets. Lactose is another excipient that has been studied extensively and used to evaluate the factors that govern the mechanism of pellet formation by extrusion–spheronization.

In contrast to layering processes, extrusion–spheronization can be used to manufacture pellets with sustained-release characteristics without the application of functional membranes to control release. For instance, matrix-type pellets can be produced with the help of mixture of microcrystalline cellulose and sodium carboxymethylcellulose (15, 16). Organic acids can be incorporated into the pellet matrix to stabilize sensitive drug substances or modify the release characteristics, especially if the solubility of the drug substance being formulated is pH-dependent (17). Although water or other granulation media act as lubricating agents during extrusion, lubricants are sometimes incorporated to improve processing.

Finally, it is not surprising that the drug substance itself plays an important role in the pelletization process, particularly at high drug loading. Physicochemical properties such as particle size, polymorphism, wettability, and solubility determine not only the amount of active ingredient that can be incorporated in the formulation but also the quality and thus the shape and surface smoothness of the final pellets.

Because extrusion–spheronization is a very complex manufacturing process that depends on a number of formulation and processing factors (Table 1), it has been studied extensively. Some of these studies used multifactorial statistical designs to determine the significance of the various factors identified above (12, 18–20). Despite the extensive work that had been done, additional research is still ongoing as demonstrated by the number of publications in various scientific journals (21–28).

SPHERICAL AGGLOMERATION

Spherical agglomeration, or balling, is a pelletization process in which powders, on addition of an appropriate quantity of liquid or when subjected to high temperatures, are converted to spherical particles by a continuous rolling or tumbling action. Spherical agglomeration can be divided into two categories—liquid-induced and melt-induced agglomerations. Over the years, spherical agglomeration has been carried out in horizontal drum pelletizers, inclined

dish pelletizers, and tumbling blenders; more recent technologies use rotary fluid-bed granulators and high-shear mixers. Although spherical agglomeration has been practiced routinely in the iron ore and fertilizer industries, its application in the pharmaceutical industry is marginal at best. Nevertheless, the process is one of the most thoroughly investigated pelletization processes, and as a result, a number of mechanisms describing the various phases of pellet formation and growth during spherical agglomeration have been proposed (29).

During liquid-induced agglomeration, liquid is added to the powder before or during the agitation step. As powders come in contact with a liquid phase, they form agglomerates or nuclei, which initially are bound together by liquid bridges. These are subsequently replaced by solid bridges, which are derived from the hardening binder or any other dissolved material within the liquid phase. The nuclei formed collide with other adjacent nuclei and coalesce to form larger nuclei or pellets. The coalescence process continues until a condition arises in which bonding forces are overcome by breaking forces. At this point, coalescence is replaced by layering, whereby small particles adhere on much larger particles and increase the size of the latter until pelletization is completed (30, 31). If the surface moisture is not optimum, some particles may undergo nucleation and coalescence at different rates and form different sizes of nuclei admixed with the larger pellets. As a result, spherical agglomeration tends to produce pellets with a wide particle size distribution.

The rate and extent of agglomerate formation depend, in part, on formulation variables such as particle size and solubility of the powder, the degree of liquid saturation, and the viscosity of the liquid phase. The moisture content is particularly critical because it determines whether nucleation occurs to initiate pelletization or whether the nuclei formed have the necessary plasticity to bring about coalescence after collisions between two nuclei. Furthermore, layering of fine particles on the larger nuclei or pellets occurs only if the surface moisture is above the critical level. The rate and extent of agglomerate formation also depend on processing variables, which are specific to a given piece of equipment. In the case of drum pelletizers, drum speed, residence time, load size, and angle of inclination, relative to the horizontal, are critical process variables that need to be optimized. Variables critical to a spherical agglomeration process using centrifugal fluid-bed granulators are disk speed, load size, residence time, fluidization air volume and temperature, and disk clearance (32–43).

Melt-induced agglomeration processes are similar to liquid-induced processes except that the binding material is a melt. Therefore, the pellets are formed with the help of

congealed material without having to go through the formation of solvent-based liquid bridges (44–55).

SPRAY DRYING AND SPRAY CONGEALING

Spray drying and spray congealing, known as globulation processes, involve atomization of hot melts, solutions, or suspensions to generate spherical particles or pellets. The droplet size in both processes is kept small to maximize the rate of evaporation or congealing, and consequently the particle size of the pellets produced is usually very small. During spray drying, drug entities in solution or suspension are sprayed, with or without excipients, into a hot air stream to generate dry and highly spherical particles. As the atomized droplets come in contact with hot air, evaporation of the application medium is initiated. This drying process continues through a series of stages whereby the viscosity of the droplets constantly increases until finally almost the entire application medium is driven off and solid particles are formed. Generally, spray-dried pellets tend to be porous. For a thorough discussion on spray drying technology, see Masters (56).

During spray congealing, a drug substance is allowed to melt, disperse, or dissolve in hot melts of waxes, fatty acids, etc., and sprayed into an air chamber, where the temperature is below the melting temperatures of the formulation components, to provide spherical congealed pellets under appropriate processing conditions. A critical requirement in a spray-congealing process is that the formulation components have well-defined, sharp melting points or narrow melting zones. Because the process does not involve evaporation of solvents, the pellets produced are dense and nonporous. See Atilla and Suheylo (57) for a detailed description of the formulation and processing requirements of spray congealing as a pelletization technique.

CRYOPELLETIZATION

Cryopelletization is a process whereby droplets of a liquid formulation are converted into solid spherical particles or pellets by using liquid nitrogen as the fixing medium. The technology, which was initially developed for lyophilization of viscous bacterial suspensions, can be used to produce drug-loaded pellets in liquid nitrogen at -160°C . The procedure permits instantaneous and uniform freezing of the processed material owing to the rapid heat transfer that occurs between the droplets and liquid nitrogen. The pellets are dried in conventional freeze dryers. The small

size of the droplets and thus the large surface area facilitate the drying process. The amount of liquid nitrogen required for manufacturing a given quantity depends on the solids content and temperature of the solution or suspension being processed. It is usually between 3 and 5 kg per kilogram of finished pellets.

The equipment consists of a container equipped with perforated plates at the bottom. Immediately below the plates at a predetermined distance is a reservoir of liquid nitrogen, in which a conveyor belt with transport baffles is immersed. The conveyor belt has a variable speed and can be adjusted to provide the residence time required for freezing the pellets. The perforated plates generate droplets that fall and freeze instantaneously as they come in contact with the liquid nitrogen below. The frozen pellets are transported out of the nitrogen bath into a storage container at -60°C before drying. Equipment of different size, ranging from laboratory scale to production size, is available commercially. (See Refs. (58, 59) for a detailed description.)

The most critical step in cryopelletization is droplet formation, which is influenced not only by formulations-related variables such as viscosity, surface tension, and solids content but also by equipment design and the corresponding processing variables. The diameter and design of the shearing edge of the holes on the container plates are critical. For instance, the diameter of the holes determines the flow rate, which, in turn, is governed by the viscosity of the formulation. The diameter of the holes also influences the size and shape of the pellets. The smaller the holes, the smaller the pellets produced.

The shape of the droplets depends in part on the distance the droplets travel before contacting the liquid nitrogen. The distance has to allow sufficient time for the formulation of spherical droplets as they contact the liquid nitrogen. When all processing parameters are carefully characterized and optimized, smooth, spherical pellets can be routinely manufactured. In cases, in which the desired pellet diameter is less than 2 mm, the liquid nitrogen is stirred to prevent agglomeration.

Solutions or suspensions suitable for cryopelletization have high solids content and low viscosities. Another important property is the surface tension of the liquid formulation, which partly determines the pellet size. The addition of a surfactant to the formulation reduces the surface tension and results in smaller particle size. Pellet size also depends on the properties of the drug substance.

Immediate-release formulations typically consist of the drug substance, fillers such as mannitol and lactose and binders such as gelatin, gelatin hydrolysates, and polyvinylpyrrolidone. Cross-linked biopolymers based on collagen derivatives are used for sustained-release pellets.

See Refs. (59, 60) for a detailed discussion on the formulation and processing variables.

MELT SPHERONIZATION

Melt spheronization is a process whereby a drug substance and excipients are converted into a molten or semimolten state and subsequently shaped using appropriate equipment to provide solid spheres or pellets. The process requires several pieces of equipment such as blenders, extruders, cutters (known as pelletizers in the plastics industry), and spheronizers. The drug substance is first blended with the appropriate pharmaceutical excipients, such as polymers and waxes, and extruded at a predetermined temperature. The extrusion temperature must be high enough to melt at least one or more of the formulation components. The extrudate is cut into uniform cylindrical segments with a cutter. The segments are spheronized in a jacketed spheronizer to generate uniformly sized pellets. The spheronization temperature needs to be high to partially soften the extrudate and facilitate deformation and eventual spheronization (61). Depending on the characteristics of the formulation ingredients, pellets that exhibit immediate- or sustained-release characteristics can be manufactured in a single step. The pellets produced are unique in that they are monosize, a property unmatched by any other pelletization technique. However, the process is still in the development stage, and additional work is needed before the process becomes a viable pelletization technique.

REFERENCES

1. Beachgaard, H.; Nielson, G.H. Controlled Release Multiple Units and Single Unit Doses. *Drug Dev. Ind. Pharm.* **1978**, *4*, 53–67.
2. Special Delivery: Advances in Drug Therapy. *The Research News*; University of Michigan, 1986; 1.
3. Cimitata, L.E. How to Manufacture and Polish Smallest Pan Goods-Nonpareil Seeds. *Confectioners J.* **1951**, 41–43.
4. Chambliss, W.C. Conventional and Specialized Coating Pans. *Pharmaceutical Pelletization Technology*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; 16–17.
5. Jan, S.; Goodhart, F.W. Dry Powder Layering. *Pharmaceutical Pelletization Technology*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; 182–183.
6. Jones, D.M. Solution and Suspension Layering. *Pharmaceutical Pelletization Technology*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; 158–159.
7. *Niro-Aeromatic Product Manual* Niro-Aeromatic, Inc.: Columbia, MD, 1992.

8. Hicks, D.C.; Freese, H.L. Extrusion Spheronization Equipment. *Pharmaceutical Pelletization Technology*; Ghebresellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; 71–100.
9. Fielden, K.E.; Newton, J.M.; Rowe, R.C. A Comparison of the Extrusion and Spheronization Behavior of Wet Powder Masses Processed by a Ram Extruder and a Cylinder Extruder. *Int. J. Pharm.* **1992**, *81*, 225–233.
10. Fielden, K.E.; Newton, J.M.; Rowe, R.C. The Effect of Lactose Particle Size on the Extrusion Properties of Microcrystalline Cellulose-Lactose Mixtures. *J. Pharm. Pharmacol.* **1989**, *41*, 217–221.
11. Dietrich, R.; Bräusse, R. Erste Erfahrungen Und Validierungsversuche an Einem Neu Entwickelten GMP-Gerechten Und Instrumentierten Pharma-Extruder. *Pharm. Ind.* **1988**, *50* (10), 1179–1186.
12. Dietrich, R. Food Technology Transfers to Pellet Production. *Manuf. Chem.* **1989**, *8*, 29–33.
13. Nakahara N. Method and Apparatus for Making Spherical Granules U.S. Patent. 3,277,520, 1964.
14. Harris, M.R.; Ghebresellassie, I. Formulation Variables. *Pharmaceutical Pelletization Technology*; Ghebresellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; 217–239.
15. O'Conner, R.E.; Schwartz, J.B. Spheronization II. Drug Release from Drug Diluent Mixtures. *Drug Dev. Ind. Pharm.* **1985**, *11* (9, 10), 1837–1857.
16. Ghali, E.S.; Klinger, G.H.; Schwartz, J.B. Modified Drug Release from Beads Prepared with Combinations of Two Grades of Microcrystalline Cellulose. *Drug Dev. Ind. Pharm.* **1989**, *15* (9), 1455–1473.
17. Bianchini, R.; Bruni, R.; Gazzaniga, A.; Vecchio, C. Influence of Extrusion/Spheronization Processing on the Physical Properties of D-Indobufen Pellets Containing pH Adjusters. *Drug Dev. Ind. Pharm.* **1992**, *18* (14), 1485–1503.
18. Mesiha, M.S.; Valles, J. A Screening Study of Lubricants in Wet Powder Masses Suitable for Extrusion-Spheronization. *Drug Dev. Ind. Pharm.* **1993**, *19* (8), 943–959.
19. Hasznos, L.; Langer, I.; Gyarmathy, M. Some Factors Influencing Pellet Characteristics Made by an Extrusion/Spheronization Process. I. Effects on Size Characteristics and Moisture Content Decrease of Pellets. *Drug Dev. Ind. Pharm.* **1992**, *18* (4), 409–439.
20. Hileman, G.A.; Goskonda, S.R.; Spalitto, A.J.; Upadrashta, S.M. A Factorial Approach to High Dose Product Development by an Extrusion/Spheronization Process. *Drug Dev. Ind. Pharm.* **1993**, *19* (4), 483–491.
21. Ku, C.C.; Joshi, Y.M.; Bergum, J.S.; Jain, N.B. Bead Manufacture by Extrusion/Spheronization. A Statistical Design for Process Optimization. *Drug Dev. Ind. Pharm.* **1993**, *19* (13), 1505–1519.
22. Baert, L.; Fanara, D.; Remon, J.P.; Massart, D. Correlation of Extrusion Forces, Raw Materials and Sphere Characteristics. *J. Pharm. Pharmacol.* **1992**, *44*, 676–678.
23. Battaille, B.; Rahman, L.; Jacob, M. Etude Des Parametres de Formulation Sur Les Caracteristiques Physico-Techniques de Granules de Theophylline Obtenus Par Extrusion-Spheronisation. *Pharma. Acta Helv.* **1991**, *66* (8), 223–236.
24. Newton, J.M.; Chow, A.K.; Jeewa, K.B. The Effect of Excipient Source on Spherical Granules Made by Extrusion/Spheronization. *Pharm. Tech.* **1993**, *3*, 166–174.
25. Battaille, B.; Ligarski, K.; Jacob, M.; Thomas, C.; Duru, C. Study of the Influence of Spheronization and Drying Conditions on the Physico-Mechanical Properties of Neutral Spheroids Containing Avicel pH 101 and Lactose. *Drug Dev. Ind. Pharm.* **1993**, *19* (6), 653–671.
26. Helen, L.; Yliruusi, J.; Muttonen, E. Process Variables of the Radial Screen Extruder, II, Size and Size Distribution of Pellets. *Pharm. Tech. Int.* **1993**, *1*, 44–53.
27. Lindner, H.; Kleinebudde, P. Use of Powdered Cellulose for the Production of Pellets by Extrusion/Spheronization. *J. Pharm. Pharmacol.* **1994**, *46*, 2–7.
28. Kleinebudde, P. Use of a Power-Consumption-Controlled Extruder in the Development of Pellet Formulations. *J. Pharm. Sci.* **1997**, *84* (10), 1259–1264.
29. Kleinebudde, P. Crystallite-Gel-Model for Microcrystalline Cellulose in Wet-Granulation, Extrusion, and Spheronization. *Pharm. Res.* **1997**, *14*, 804–809.
30. Ghebresellassie, I. Mechanism of Pellet Formation and Growth. *Pharmaceutical Pelletization Technology*; Ghebresellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; 123–143.
31. Wan, L.S.C. Manufacture of Core Pellet Formation and Growth. *Pharmaceutical Pelletization Technology*; Ghebresellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1989; 123–143.
32. Vecchio, C.; Bruni, G.; Gazzaniga, A. Preparation of Indobufen Pellets by Using Centrifugal Rotary Fluidized Bed Equipment Without Starting Seeds. *Drug Dev. Ind. Pharm.* **1994**, *20* (12), 1943–1956.
33. Holm, P.; Bonde, M.; Wigmore, T. Pelletization by Granulation in a Roto-Processor RP-2. I. Effects of Process and Product Variables on Granule Growth. *Pharm. Technol. Eur. Sep.* **1996**, *8*, 22, 24, 26, 28, 30, 32, 34, 36.
34. Holm, P. Pelletization by Granulation in a Roto-Processor RP-2. Effects of Process and Product Variables on Agglomerates' Shape and Porosity. *Pharm. Technol. Eur. Oct.* **1996**, *8*, 38–45.
35. Vertommen, J.; Kinget, R. Influence of Five Selected Processing and Formulation Variables on the Particle Size Distribution, and Friability of Pellets Produced in a Rotary Processor. *Drug Dev. Ind. Pharm.* **1997**, *23* (1), 39–46.
36. Vertommen, J.; Rombaut, P.; Kinget, R. Shape and Surface Smoothness of Pellets Made in a Rotary Processor. *Int. J. Pharm.* **Jan 1, 1997**, *146*, 21–29.
37. Vertommen, J.; Rombaut, P.; Kinget, R. Internal and External Structure of Pellets Made in a Rotary Processor. *Int. J. Pharm.* **Feb 23, 1998**, *161*, 225–236.
38. Vertommen, J.; Rombaut, P.; Michoel, A.; Kinget, R. Estimation of the Amount of Water Removed by Gap and Atomization Air Streams During Pelletization in a Rotary Processor. *Pharm. Dev. Technol.* **1998**, *3* (1), 63–72.
39. Vojnovic, D.; Rupena, P.; Moneghini, M.; Rubessa, F.; Sergeant, M.; Experimental Research Methodology Applied to Wet Granulation in a High-Shear Mixer. I. S. T. P. II *Pharma Sci.* **1993**, *3* (2), 130–135.
40. Holm, P.; Shaefer, T.; Kristensen, H.G. Pelletization by Controlled Wet Granulation in a High-Shear Mixer. S. T. P. *Pharma Sci.* **1993**, *3* (2), 130–135.
41. Vertommen, J.; Michoel, A.; Rombaut, P.; Kinget, R. Production of Pseudoephedrine HCl Pellets in a High Shear Mixer-Granulator. *Eur. J. Pharm. Biopharm.* **1994**, *40* (1), 32–35.

42. Vojnovic, D.; Moneghini, M.; Masiello, S. Design and Optimization of Theophylline Pellets Obtained by Wet Spheronization in a High-Shear Mixer. *Drug Dev. Ind. Pharm.* **1995**, *21* (18), 2129–2137.
43. Maggi, L.; Bonfanti, A.; Santi, P.; Massimo, G.; Catellani, P.L.; Bellotti, A.; Colombo, P.; Zanchetta, A.; Suitability of a Small Scale High-Shear Mixer for Powder Pelletization. *Pharm. Technol. Eur.* **Oct**, **1996**, *8*, 82, 84, 86, 88–90.
44. Schaefer, T.; Holm, P.; Kristensen, H.G. Melt Granulation in a Laboratory Scale High Shear Mixer. *Drug Dev. Ind. Pharm.* **1990**, *16* (8), 1249–1277.
45. Schaefer, T.; Holm, P.; Kristensen, H.G. Melt Pelletization in a High Shear Mixer. I. Effects of Process Variables and Binder. *Acta Pharm. Nordica* **1992**, *4* (3), 133–140.
46. Schaefer, T.; Holm, P.; Kristensen, H.G. Melt Pelletization in a High Shear Mixer. II. Power Consumption and Granule Growth. *Acta Pharm. Nordica* **1992**, *4* (3), 141–148.
47. Schaefer, T.; Holm, P.; Kristensen, H.G. Melt Pelletization in a High Shear Mixer. III. Effects of Lactose Quality. *Acta Pharm. Nordica* **1992**, *4* (4), 245–252.
48. Schaefer, T.; Taagegaard, B.; Thomsen, L.J.; Kristensen, H.G. Melt Pelletization in a High Shear Mixer. IV. Effects of Process Variables in a Laboratory Scale Mixer. *Eur. J. Pharm. Sci.* **1993**, *1* (3), 125–131.
49. Schaefer, T.; Taagegaard, B.; Thomsen, L.J.; Kristensen, H.G. Melt Pelletization in a High Shear Mixer. V. Effects of Apparatus Variables. *Eur. J. Pharm. Sci.* **1993**, *1* (3), 133–141.
50. Schaefer, T. Melt Pelletization in a High Shear Mixer. X. Agglomeration of Binary Mixtures. *Int. J. Pharm.* **Aug 9**, **1996**, 149–159.
51. Schaefer, T.; Mathiesen, C. Melt Pelletization in a High Shear Mixer. VII. Effects of Product Temperature. *Int. J. Pharm.* **May 28**, **1996**, *134*, 105–117.
52. Schaefer, T.; Mathiesen, C. Melt Pelletization in a High Shear Mixer. VIII Effects of Binder Viscosity. *Int. J. Pharm.* **Aug 9**, **1996**, *139*, 125–138.
53. Schaefer, T.; Mathiesen, C. Melt Pelletization in a High Shear Mixer. IX. Effects of Binder Particle Size. *Int. J. Pharm.* **Aug 9**, **1996**, *139*, 139–148.
54. Maejima, T.; Osawa, T.; Nakajima, K.; Kobayashi, M. Preparation of Spherical Beads without Any Use of Solvents by a Novel Tumbling Melt Granulation (TMG) Method. *Chem. Pharm. Bull.* **1997**, *45*, 518–524.
55. Maejima, T.; Osawa, T.; Nakajima, K.; Kobayashi, M. Effects of Species of Non-Meltable and Meltable Materials and Their Physical Properties on Granulatability in Tumbling Melt Granulation Method. *Chem. Pharm. Bull.* **1997**, *45*, 1833–1839.
56. Masters, K. *Spray Drying Handbook*, 4th Ed.; John Wiley & Sons: New York, 1985.
57. Atilla, H.A.; Suheyli, K.H. Preparation of Micropellets by Spray Congealing. *Multiparticulate Oral Drug Delivery*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1994; 17–34.
58. European Patent EP 0, 081, 913, 1985.
59. German Patent DE 37 11 169, 1988.
60. Knoch, A. Cryopelletization. *Multiparticulate Oral Drug Delivery*; Ghebre-Sellassie, I. Ed.; Marcel Dekker, Inc. New York, 1994; 35–50.
61. Ghebre-Sellassie, I.; Nesbitt, R.; Fawzi M. Novel Pharmaceutical Pellets and Process for Their Production PCT Application, WO 93/07859, April 29, 1993.
62. Harrison, P.J.; Newton, J.M.; Rowe, R.C. The Characterization of Wet Powder Masses Suitable for Extrusion/Spheronization. *J. Pharm. Pharmacol.* **1985**, *37*, 686–691.

PEPTIDES AND PROTEINS—ORAL ABSORPTION

Eugene J. McNally

Jung Y. Park

Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, Connecticut, U.S.A.

INTRODUCTION

Rapid developments in biotechnology have posed new challenges for pharmaceutical research scientists to develop peptide (>3 amino acids) and protein drugs. Although these peptide and protein drugs are highly potent and specific in their physiological actions, they are in general difficult to administer orally. The reason is that most peptide and protein drugs are very unstable in the gastrointestinal (GI) tract and show poor oral absorption because of their size and hydrophilic nature (there are specific exceptions, e.g., cyclosporine). Although each nonparenteral delivery route has its own advantages and disadvantages, based on the available published data and the known permeability of mucosal membranes, it may be possible to rank them as the preferred route for the delivery of peptide and protein drugs. It appears that the nasal cavity is the preferred route for the delivery of peptide and protein drugs, followed by the vaginal, pulmonary, oral, and transdermal routes, respectively. It is common knowledge that peptide and protein drugs are easily hydrolyzed and digested by acids and enzymes in the GI tract. In addition, these drugs usually show low bioavailability because of their poor membrane permeability in the GI tract.

To deliver peptide and protein drugs orally, it is necessary to protect them from the hostile GI environment. Many strategies, such as coadministration with nonspecific protease inhibitors as well as the use of chemical modification approaches to improve resistance to enzymatic degradation, have been reported in the scientific and patent literature. In recent years, a number of drug-delivery companies specializing in protein delivery have emerged. A number of these companies are attempting to apply formulation approaches such as liposomes or microspheres to protect protein drugs from enzymatic degradation in the GI tract or using site-specific delivery to the colon or rectum to bypass the harsh GI environment. In addition, to increase the GI absorption of the peptide and protein drugs, absorption enhancers as well as carriers or prodrug approaches are being explored.

The purposes of this article are sixfold: 1) to review the GI tract as it relates to the absorption of drugs; 2) to review briefly the structures of peptides and proteins; 3) to review the enzymatic and physical barriers for intestinal permeability of peptides and proteins; 4) to review the absorption of amino acids, peptides, and proteins across the GI tract; 5) to review the kinds of absorption enhancers and how they work; and 6) to review possible ways to improve the absorption of poorly bioavailable peptide and protein drugs.

GASTROINTESTINAL PHYSIOLOGY RELEVANT TO ABSORPTION

A rational approach to understanding the GI absorption of peptides and proteins requires some knowledge of the GI tract such as morphology, function of different components of the cells, geometry, ultrastructure, biochemical processes, hydrodynamics in the intestinal lumen, and transport mechanisms.

The Digestive System

The digestive system consists of the canal from the mouth to the anus (oropharynx, esophagus, stomach, small intestine, and large intestine) and associated organs (salivary glands, liver, gallbladder, and pancreas) (Fig. 1). The small intestine is divided into three parts, the duodenum, jejunum, and ileum, and is the primary site of absorption for most drugs. The large intestine, which is also called the colon, is divided into the caecum, ascending colon, transverse colon, descending colon, sigmoid colon, rectum, and anus. The salivary glands, liver, gallbladder, and pancreas deliver digestive secretions and help digestion and absorption of peptide and protein drugs.

Stomach

The inner surface of the stomach consists of well-defined tissue layers: the muscle and the submucosal and mucosal layers. The absorption function of stomach is minimal

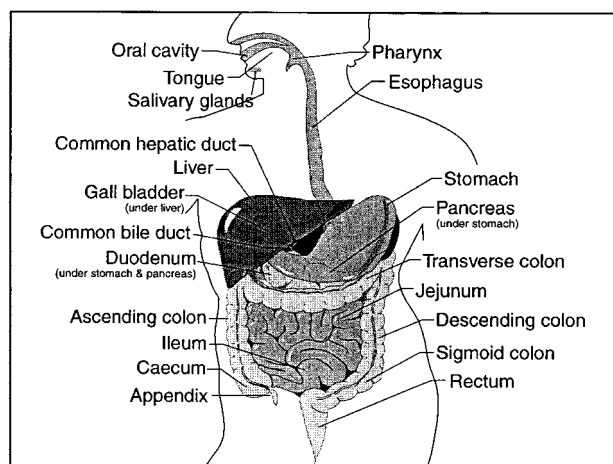


Fig. 1 The human digestive system. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

owing to the limited surface area, lack of villi, thick mucosal layer, and short residence time. The epithelium of the gastric mucosa secretes hydrochloric acid, pepsin, intrinsic factor, and bicarbonate (1). The low pH of gastric juice generated by hydrochloric acid causes protein denaturation (unfolding), which results in increased exposure of peptide bonds to pepsin, which breaks down the proteins into peptides. However, hydrolysis in the stomach is incomplete because pepsin can only break peptide bonds between specific amino acids.

Small intestine

Most enzymatic hydrolysis of the macromolecules in food occurs in the small intestine. The small intestine is approximately 5 m long with a radius of approximately 2 cm, and is divided into three anatomical regions: the duodenum (the first 25 cm), the jejunum (the second 2.0 m), and the ileum (the last 2.75 m) (1). The intestinal membrane wall consists of four basic layers: the mucosa, submucosa, muscularis, and serosa.

The mucosal layer, comprising the luminal surface of the small intestine, is responsible for the digestive and absorptive functions of the small intestine. The mucosal surface area is much larger than predicted for a simple cylinder. Circular folds account for this amplification. The mucosal surface area is extended further by fingerlike projections called villi and depressions called crypts. The villi are 0.5–1.0 mm in height. Each villus and crypt is lined by epithelial cells that are covered with many closely packed microvilli that project into the intestinal lumen. If the small intestine is viewed as a simple cylinder, its mucosal surface area would be on the order of half of a square meter. However, in reality, the mucosal surface

area of the small intestine is approximately 250 square meters, comparable with size of a tennis court.

Mucous surface: The mucosa of the small intestine consists of three layers (Fig. 2): an absorptive layer, a continuous single sheet of columnar epithelium; the lamina propria, a layer heterogeneous in composition and cell type; and the muscularis mucosa, a muscular layer separating the mucosa and submucosa.

The lamina propria consists primarily of connective tissue and supports the epithelium lining. There is ample evidence that the lymphoid cells and associated structures of the lamina propria play an important role immunologically. Small lymphoid nodules are present in the upper small intestine, whereas large organized aggregates of lymphoid tissue (Peyer's patches) are present in the ileum. The lymphoid cells, nodules, and Peyer's patches of the lamina propria, along with the intraepithelial lymphocytes, establish the so-called gut-associated lymphoid tissue (GALT), which is a major subgroup of the immune system and makes up as much as 25% of the gastrointestinal mucosa.

A one-cell thick sheet of epithelial cells covers the surfaces of the villi and lines of crypts. Some of the cell types identified in the epithelial lining of the small intestine are enterocytes (digestion and absorption), goblet cells (mucus secretion), endocrine cells (hormone secretion), and M cells (absorption of food and antigens).

Absorptive cells: Absorptive cells (commonly called enterocytes or brush border cells) are the most prevalent type of cells on the tips of villi and are the most important for absorption (Fig. 3). In humans, these cells are 20–30 μm in height and 8–10 μm in width. The mucosal surface of this cell is characterized by the presence of microvilli (the brush border, approximately 1.0 μm in height and 0.1–0.2 in width). The microvilli look something like a brush, as shown in Fig. 3. For this reason, the microvillus border of intestinal epithelial absorptive cells is called the brush border. The microvillus plasma membrane has a trilaminar structure (70–90° A thick) composed of proteins, neutral lipids, phospholipids, and glycolipids. An integral and dynamic part of the plasma membrane is the glycocalyx, which is a uniform layer of filamentous glycoproteins (Fig. 4) (2). This glycocalyx layer has a negative charge at physiological pH primarily owing to the presence of sialic acids at the terminal position in the carbohydrate chain. The microvillus and its glycocalyx are viewed as the digestive-absorptive unit of the enterocyte. The plasma membrane of microvilli has an unusually high protein-to-lipid ratio, owing to the presence of specialized proteins possessing enzymatic, receptor, and transport properties. This microvillus plasma membrane contains a number of

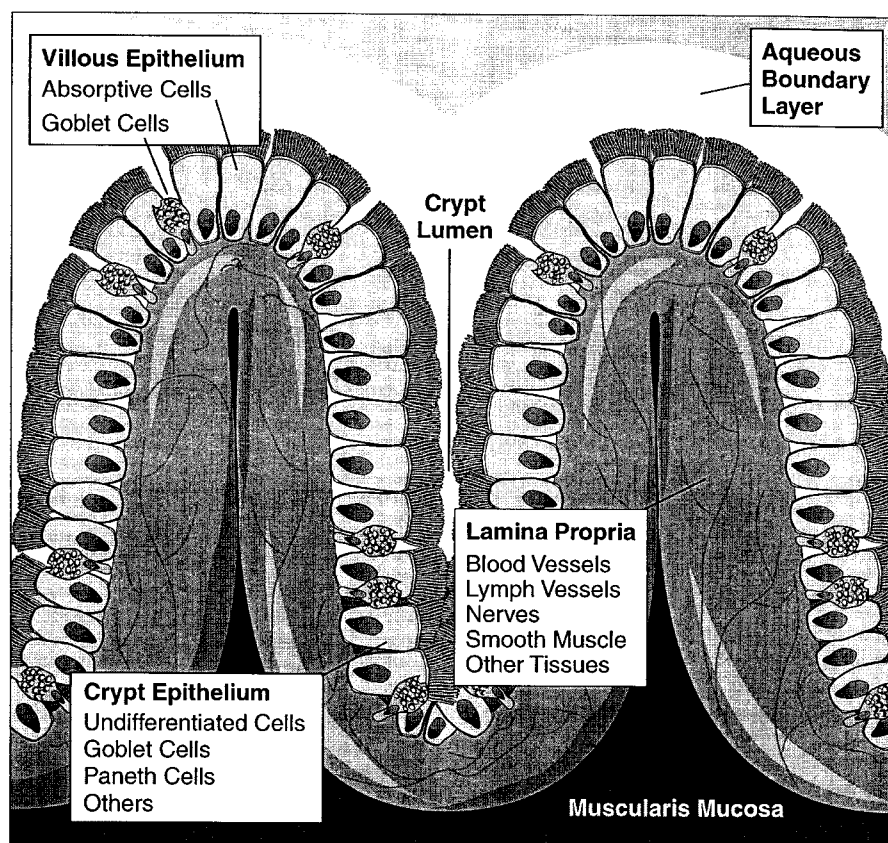


Fig. 2 Schematic depiction of two-section villi and a crypt to illustrate the small intestine mucosa. Also shown is an aqueous boundary layer located at the intestinal lumen and membrane interface. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

enzymes (peptidases) involved in the hydrolysis of peptides. In addition, proteins responsible for the cotransport of sodium and amino acids have been found in the microvillus plasma membrane. Receptor proteins specific for certain substances have been found on the microvilli of enterocytes in different regions of the small intestine. For example, the receptors for the intrinsic factor vitamin B₁₂ complex are present in the microvilli of ileal enterocytes but are not found on jejunal cells. This is why vitamin B₁₂ is absorbed exclusively from the ileum. Vitamin B₁₂ has been explored as a delivery system for peptide and proteins by covalently conjugating vitamin B₁₂ to peptides (leuteinizing hormone-releasing hormone, LHRH) or proteins (bovine serum albumin, BSA) (3).

The absorptive cell basolateral membrane rests on the lamina propria. The basolateral membrane is different from the apical membrane. It has a low protein-lipid ratio and is thinner and more permeable than the apical membrane. In addition, different enzymes are present at the basolateral membrane compared with the apical membrane. Adjacent

cells are connected by a junctional complex called the tight junction. This tight junction is important because it represents one possible route of intestinal absorption, known as the paracellular route. The other major route of absorption is across the cell, also known as the transcellular route. These routes will be discussed later.

Aqueous boundary (diffusion) layer: The aqueous boundary layer (often referred to as the stagnant, unstirred, or aqueous diffusion layer) is an important hydrodynamic barrier that a drug must traverse before reaching the surface of the mucosal membrane (4). Before a molecule in the intestinal lumen passes through the membrane, it must first cross the aqueous boundary layer located at the intestinal lumen and membrane interface (Fig. 2). The liquid in this layer, in reality, is not static, as the term “unstirred” implies, but represents a film at the surface where diffuse and natural convective mixing occurs. This unstirred layer can be a rate-limiting step for the absorption of hydrophobic molecules. However, hydrophilic molecules such as peptides will diffuse through the

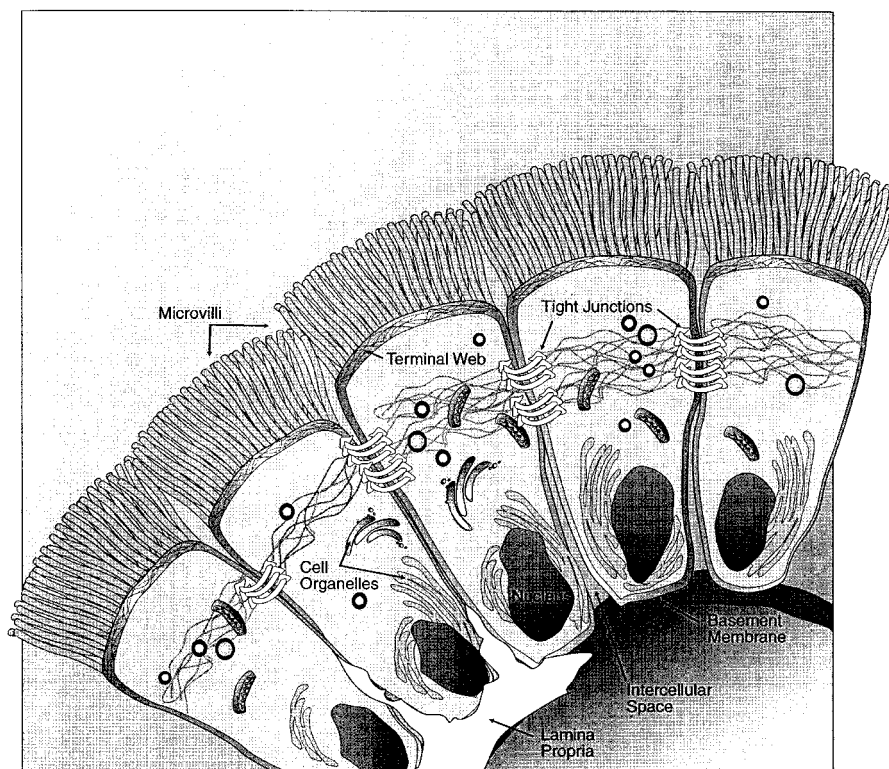


Fig. 3 Schematic depiction of intestinal epithelial absorptive cells. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

aqueous boundary layer with less resistance. The effective thickness of the aqueous boundary layer is thinner around the tips of the villi and thicker and less stirred in the valleys of the villi. The existence of the aqueous boundary layer is physically sound and experimentally demonstrable (5).

Luminal and membrane metabolism of peptides and proteins: In meaningful studies on peptide and protein drug absorption in the small intestine, it is prerequisite to distinguish among cavital, membrane contact, and intracellular drug metabolism (4). Cavital metabolism takes place in the lumen of the small intestine by enzymes such as trypsin, chymotrypsin, carboxypeptidase, and elastase, which are secreted by the pancreas. Membrane contact metabolism is carried out by aminopeptidases localized on the brush border membrane. Intracellular metabolism occurs inside of the cells. The known intracellular enzymes are cytoplasmic peptidases, prolidase, dipeptidase, and tripeptidase (1). A more detailed discussion of this topic is presented in "Intestinal Absorption Barriers," later.

Intestinal blood flow: The mechanistic relationship among intestinal blood flow and absorption, secretion, and metabolic activity of the intestinal mucosa is unclear.

However, there is evidence that impaired intestinal blood flow rate correlates with a decrease in drug absorption rate. It has been postulated that reduced blood flow slows down the absorption rate by: 1) decreasing the effective concentration gradient across the epithelial layer for passively absorbed molecules, by not rapidly carrying the molecule away, and 2) by lowering oxygen supply to the absorption cells needed to maintain the active transport mechanism for absorption.

In general, when molecules transport through the epithelium cell layer directly into the mesenteric blood draining the small intestine, the total mass transfer resistance may be described by the sum of resistance for barriers (aqueous boundary layer in front of the membrane, the membrane itself, and the aqueous boundary layer in blood side) in series (6):

$$\frac{1}{P_{app}} = \frac{1}{P_{aq}} + \frac{1}{P_{membrane}} + \frac{1}{P_{blood}} \quad (1)$$

where P_{app} is apparent permeability coefficient; P_{aq} is permeability coefficient of the aqueous boundary layer; $P_{membrane}$ is permeability coefficient of the membrane; and

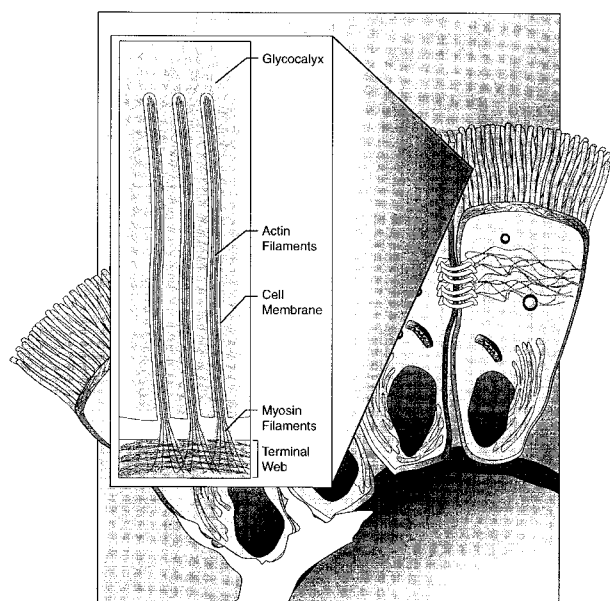


Fig. 4 Schematic depiction of glycocalyx. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

P_{blood} is permeability coefficient of the blood boundary layer.

When the permeabilities of the aqueous boundary layer and membrane are relatively large, reduced blood flow becomes the rate-limiting step. For example, the apparent permeability coefficient is approximately 85% controlled by blood flow when P_{aq} and P_{membrane} are each approximately 10 times greater than P_{blood} .

Large intestine (colon)

The large intestine (or colon) is divided into five parts: 1) the caecum, or opening into the colon; 2) the ascending colon, or second section of the large intestine; 3) the transverse colon; 4) the descending colon; and 5) the sigmoid colon, which is at the end of the colon, ending in the rectum (Fig. 1). The large intestine is approximately 1.5 m long, and its proximal diameter is approximately 2 inches. The wall of the large intestine is divided into four layers, exactly the same as the small intestine: the serosa, muscularis externa, submucosa, and mucosa. However, the mucosa does not contain villi, and the microvilli and the cells are much less dense than those in the small intestine. It also contains a large number of bacteria, which digest food residues into caloric substance that are subsequently absorbed. The total bacteria count in the large intestine is higher than that in the stomach and small intestine. Attempts have been made to deliver peptide and protein drugs to the colon by coating the drug particles with copolymers that are sensitive to bacterial azo-reductases (7).

In addition, the colon has recently received considerable attention as a possible delivery site for peptide and protein drugs because of low proteolytic activity in this region of the GI tract.

Rectum

The rectum is the terminal 15 to 19 cm of the large intestine. The mucous membrane of the rectum consists of a layer of cylindrical epithelial cells, without villi. The pH in the rectum is between 7.2 and 7.4. The surface area is only 200–400 cm². Drug absorption takes place through three veins of the submucous plexus. Superior rectal veins enter into the portal veins and deliver drugs to the liver. However, the inferior and middle rectal veins connect with the inferior vena cava and thus bypass the liver. This may provide an advantage for the delivery of some peptide and protein drugs through the rectal membrane, thus bypassing the liver.

STRUCTURAL AND FUNCTIONAL ASPECTS OF PEPTIDES AND PROTEINS

In addition to understanding the morphology and function of the GI tract, it is also important to have a thorough understanding of the physical/chemical properties of peptides and proteins to rationally formulate them for successful oral delivery. An extensive review is beyond the scope of this article [see other sections in this text and a recent text on formulation and delivery of proteins (8)]. Instead, a brief description of peptides and proteins is provided that focuses on those characteristics that give rise to the unique biological activity of these molecules. Peptides are formed by loss of water from the NH₂ and COOH groups of adjacent amino acids; they are referred to as di-, tri-, tetra- (etc.) peptides depending on the number of amino acids composing them. The term oligopeptides refers to peptides that have fewer than eight amino acids, whereas polypeptide refers to those peptides that contain approximately eight or more amino acids. An amino acid unit in a polypeptide is called a residue. Polypeptides that contain from approximately 50 (molecular weight ≈6000) to more than 8000 amino acid residues (molecular weight ≈1,000,000) are called proteins (9). Each protein molecule is a polymer of α-amino acids linked together in a sequential manner by peptide bonds. Although more than 100 amino acids occur in nature, particularly in plants, only 20 are commonly found in most proteins.

Proteins are one of the major naturally occurring organic compounds, and they constitute much of animal and plant tissue. The word protein comes from the Greek

word *proteios*, meaning first place, because proteins are thought to be the most important part of living matter. They are instrumental in almost everything cells do. Proteins are the most abundant components of cells. They account for more than 50% of the dry weight of most cells (9). They serve as antibodies, enzymes, hormones, transport mediators, and structural elements.

Unlike traditional small molecular weight chemical drugs, peptide and especially protein drugs are highly complex molecules that possess primary, secondary, tertiary, and quaternary structures. The primary structure refers to the linear amino acid sequence of a peptide or protein. The secondary structure refers to the way in which segments of the peptide backbone orient into regular patterns such as a α -helix or β -sheet. The tertiary structure refers to the native conformation that is formed by the folding of the secondary structures to a compact, tightly folded structure to reach the most thermodynamically stable state. Quaternary structure occurs when a protein molecule consists of two or more polypeptide chains (10).

A protein's biological function depends on its unique conformation. This conformation is a consequence of the specific linear sequence of the amino acids that makes up the polypeptide chain (primary structure) and the specific three-dimensional structure (conformation) the protein molecule adopts. In general, nonpolar residues tend to fold into the center of the structure to get away from water, whereas the polar residues tend to stay on the surface in contact with water. To maintain the biological activity of the protein drug, this unique higher-order physical structure must be preserved during passage through the GI tract and upon absorption. Subtle changes in the chemical and physical structure can lead to a loss of biological activity.

INTESTINAL ABSORPTION BARRIERS

Enzymatic Barriers

Most dietary proteins are known not to be absorbed in humans as intact forms. Instead, they are usually broken down into amino acids or di- and tripeptides first in the GI tract. The stomach secretes pepsinogen, which is converted to the active protease pepsin by the action of acid. Pepsins, which are most active at pH 2 to 3, hydrolyze partially digested dietary proteins. The partially digested dietary proteins are further broken down by proteolytic enzymes (peptidases) produced by the pancreas and secreted in the duodenum of the small

intestine. The peptidases that break the internal peptide linkages are known as endopeptidases, whereas those that attack the terminal, or end, groups of amino acids are called exopeptidases. The endopeptidases are trypsin, chymotrypsin, and elastase, and the only exopeptidase enzymes are carboxypeptidases (1). All four enzymes (trypsin, chymotrypsin, carboxypeptidases, and elastase) are secreted by the pancreas as inactive, proenzyme forms of trypsinogen, chymotrypsinogen, procarboxypeptidases, and proelastase. The enzyme enterokinase (also called enteropeptidase) within the lumen of the small intestine converts trypsinogen to trypsin. Then, trypsin converts chymotrypsinogen, procarboxypeptidases, and proelastase to chymotrypsin, carboxypeptidases, and elastase, respectively (1). These enzymes, whose function in vivo is to break down food proteins into amino acids that can be then absorbed and turned into energy, will also degrade a protein drug. It is important to understand the function and sites of secretion of these enzymes so that strategies to protect protein drugs from degradation can be developed.

These four peptidases secreted from the pancreas convert proteins and polypeptides to oligopeptides. The luminal degradation is up to 20% of the total degradation in a given small intestinal segment (11). The rest of the degradation occurs on contact with the brush border membrane or after entry into the cell. Brush border peptidases such as amino oligopeptidase, amino peptidase, and dipeptidyl aminopeptidase then break down the oligopeptides to amino acids (up to 70%) and di- and tripeptides (up to 30%) (1). The di- and tripeptides that cross the brush border membrane are converted to single amino acids by the intracellular enzymes known as cytoplasmic peptidases, prolidase, dipeptidase, and tripeptidase (1).

The above-mentioned enzymatic barriers must be overcome to improve oral absorption of peptide and protein drugs from the GI tract. This may be possible to achieve to some extent by the coadministration of proteolytic enzyme inhibitor or by chemical modification of peptides or proteins and by other formulation approaches (see "Intestinal Absorption of Amino Acids, Peptides, and Proteins," later).

Physical Barriers

Absorption barriers are related to the permeability of drug molecules across the gastrointestinal membrane including the colonic membrane. There are two distinct mechanisms for molecules to cross the membrane: via paracellular transport and transcellular transport (Fig. 5) (12). Paracellular transport involves only passive diffusion where the molecules pass through the tight junctions between the epithelial cells. In contrast, transcellular transport can occur

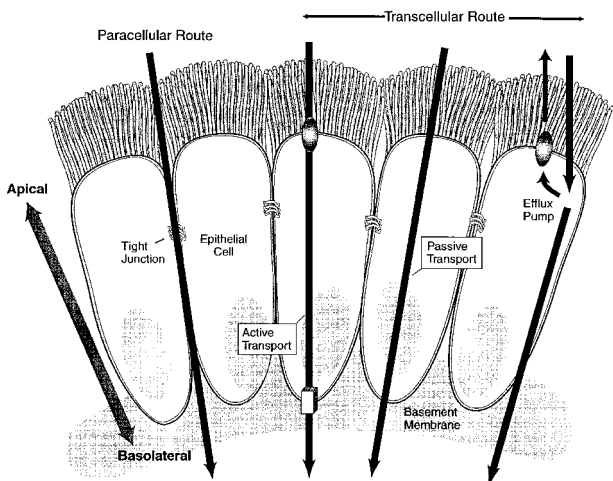


Fig. 5 Routes for the transport of drugs across the GI epithelial cells. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

by passive diffusion as well as by active transport, or endocytosis. In general, the hydrophilic molecules diffuse predominantly through the paracellular route, whereas the lipophilic molecules traverse predominantly through the epithelial cells.

Transcellular transport

Three processes are involved in transcellular transport across the intestinal epithelial cells: simple passive trans-

port, passive diffusion together with an efflux pump, and active transport and endocytosis. Simple passive transport is the diffusion of molecules across the membrane by thermodynamic driving forces and does not require direct expenditure of metabolic energy. In contrast, active transport is the movement of molecules across the membrane resulting directly from the expenditure of metabolic energy and transport against a concentration gradient. Endocytosis processes include three mechanisms: fluid-phase endocytosis (pinocytosis), receptor-mediated endocytosis, and transcytosis (Fig. 6). Endocytosis processes are covered in detail in “Absorption of Polypeptides and Proteins,” later.

The mechanism whereby drugs are absorbed from the GI tract is complex. Understanding the intestinal transport mechanism is crucial to the prediction of oral drug absorption. The physical model (13–15) utilizes the basic principles of thermodynamics and mass transport. The physical model for the simultaneous passive and active membrane transport of drugs in the intestinal lumen is depicted in Fig. 7. The bulk aqueous solution with an aqueous boundary layer on the mucosal side is followed by a series of heterogeneous membranes consisting of parallel lipoidal and aqueous channel pathways for passive and active transport. Thereafter, a sink on the serosal side follows.

The rate of disappearance of a drug from the intestinal lumen and appearance in the blood is given by:

$$\frac{dC_b}{dt} = -\frac{A}{V}P_{app}(C_b - C_{blood}) \tag{2}$$

when there is no accumulation of drug in the blood side, i.e., sink conditions

$$\frac{dC_b}{dt} = -\frac{A}{V}P_{app}C_b = -K_uC_b \tag{3}$$

where C_b is the total drug concentration in the lumen; $K_u = (\frac{A}{V})$; P_{app} , the apparent first-order absorption rate

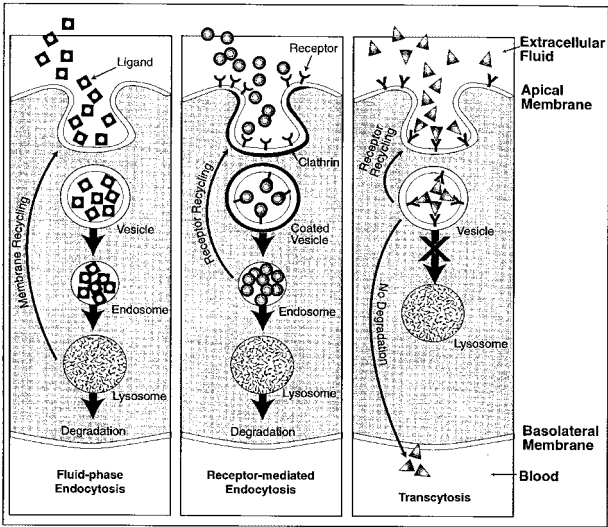


Fig. 6 Schematic depiction of cellular uptake mechanisms: fluid-phase endocytosis, receptor-mediated endocytosis, and transcytosis. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

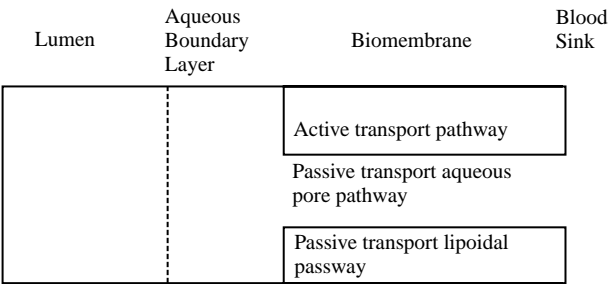


Fig. 7 The physical model for the simultaneous passive and active membrane transport of drugs in the intestinal lumen.

constant; A is the surface area; and V is the luminal solution volume.

Assuming no significant aqueous boundary layer on the blood side, the apparent permeability coefficient (P_{app}) in Eq. 1 is expressed by:

$$P_{app} = \frac{1}{\frac{1}{P_{aq}} + \frac{1}{P_{membrane}}} \quad (4)$$

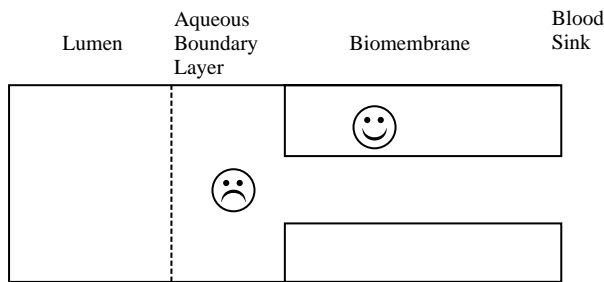
With the use of Eq. 4, the absorption rate constant can be expressed as:

$$K_u = \frac{A}{V} \cdot \frac{1}{\frac{1}{P_{aq}} + \frac{1}{P_{membrane}}} \quad (5)$$

For highly lipophilic drugs ($\sim \log PC > 3.0$), the absorption rate constant will be dependent on the diffusion rate across the aqueous boundary layer in front of the membrane, and is expressed by:

$$K_u = \left(\frac{A}{V}\right) \cdot P_{aq} \quad (6)$$

and is also shown in illustration:



For the hydrophilic drugs such as peptides and proteins, the absorption rate constant will be dependent on the diffusion across the membrane ($P_{aq} \gg P_{membrane}$).

$$K_u = \left(\frac{A}{V}\right) \cdot P_{membrane} \quad (7)$$

Table 1 includes additional mathematical definitions of effective membrane permeability coefficient for various transport mechanisms.

During the past 10 years, it has been reported that the 170-kDa P-glycoprotein, which is known as the principal component of pleiotropic (multidrug) resistance (MDR) in tumor cells, works as an apically polarized efflux transporter (also known as an efflux pump). This efflux transporter opposes the transcellular movement of drugs in the epithelial cells. P-glycoprotein is known to be present in the apical region of epithelial cells in the kidney, liver, and GI tract. Many drugs are known to be substrates for this efflux transporter including cyclosporin A (16) and other peptides (17). The existence of efflux pumps for peptides is demonstrated by measuring the apparent permeability coefficient (P_{app}) values of peptides (e.g., cyclosporin A) using the Caco-2 cell culture model with and without efflux pump inhibitors. With efflux pump inhibitors (chlorpromazine and progesterone), the P_{app} value of apical to basolateral transport for both cyclosporin A increases significantly compared with the control (without inhibitors). In contrast, the basolateral to apical transport is decreased with an inhibitor. Cyclosporin A is a substrate for the efflux transporter P-glycoprotein, and the efflux pump is located on the apical side of the intestinal epithelial cells. Whenever unexplainable poor absorption of peptides is seen, the role of these efflux pumps should be examined.

Table 1 Effective membrane permeability coefficients for various transport mechanisms

Mechanism	$P_{membrane} =$
Passive diffusion of small electrolytes	$P_L X_s + P_p$
Passive diffusion of small nonelectrolytes	$P_L + P_p$
Passive diffusion of large nonelectrolytes	P_L
Passive transport of ampholytes	$P_L^{\pm} X_s^{\pm} + P_p^+ X_s^+ + P_p^- X_s^-$
Passive diffusion coupled with active transport	$P_L X_s + P_L^* (1 - X_s)$ where $P_L^* = J_{max}^* / K_L^*$

P_L is the permeability coefficient of the lipoidal membrane; P_p is the permeability coefficient of the aqueous pore; X_s is the fraction of unionized species; P_L^{\pm} is the permeability coefficient of the lipoidal membrane of neutrally charged species; P_p^+ , P_p^- is the permeability coefficient of aqueous pore of positively and negatively charged species, respectively; X_s^{\pm} , X_s^+ , X_s^- is the fraction of neutrally, positively, and negatively charged species at the membrane surface, respectively; P_L^* is the permeability coefficient of the membrane for active transport; J_{max}^* is the maximum flux for active transport; K_L^* is the Michaelis constant for active transport.

Paracellular transport

The primary barrier to paracellular transport is the tight junction (also known as *zonula occludens*), which plays a central role in sealing the intercellular space in epithelial cells (18). These tight junctions are located just under the brush border and form a seal between adjacent epithelial cells. Tight junctions act as a gate and fence to control the intercellular movement of molecules. In addition, tight junctions prevent free diffusion and intermixing of certain apical and basolateral plasma membrane proteins, resulting in polarization of the epithelial cell layers. These junctions are composed of paired intramembrane strands, which consist of lipid molecules and several protein components such as occludin, claudin-1, and claudin-2. Based on coexpression studies, it appears that claudin-1 and claudin-2 are more likely involved in sealing the tight junctions, and occludin plays a supporting role for coordinating and modulating functions of tight junctions (19). The strands are thought to contain multiple and discrete aqueous pores (12). The tight junctions are lined with fixed negative charges such as COO^- , SO_4^{2-} , and PO_4^{3-} ions present on the surface of glycoproteins and proteoglycans on neighboring cell membranes (20). Because the tight junctions are negatively charged, this barrier has cation selectivity. In general, cations diffuse through the tight junctions faster than anions (21). The radius of the tight junctions of the intestinal mucosa can be a critical factor for the paracellular transport of peptide and protein drugs. The radius of the tight junction is about 12 Å (22). Because of the size limit of the tight junction, only small molecules and ions are capable of diffusing through the paracellular route. Amino acids, dipeptides, and tripeptides may be small enough to be absorbed across the intestinal wall through the paracellular route, but polypeptides and proteins are restricted because of their size.

One approach for overcoming this diffusion restriction is to change the aqueous pore radius of tight junctions by coadministering drugs with enhancers. Numerous enhancers have been investigated to increase the permeability of the intestinal membrane. They can be categorized into two groups: anionic surfactants such as long-chain acylcarnitines, bile acids, disodium dodecyl sulfate, and sodium caprate, and calcium-chelating agents such as EDTA and citrate (17). The calcium-chelating agents reduce the levels of calcium and induce opening of tight junctions. The anionic surfactants open up the tight junctions by interacting with the cell membrane. It is very important to weigh the benefits versus the toxic risks of enhancers with respect to causing long-term, irreversible damage to the membrane. Very little work has been reported on the toxicological effects of enhancers. It is generally known

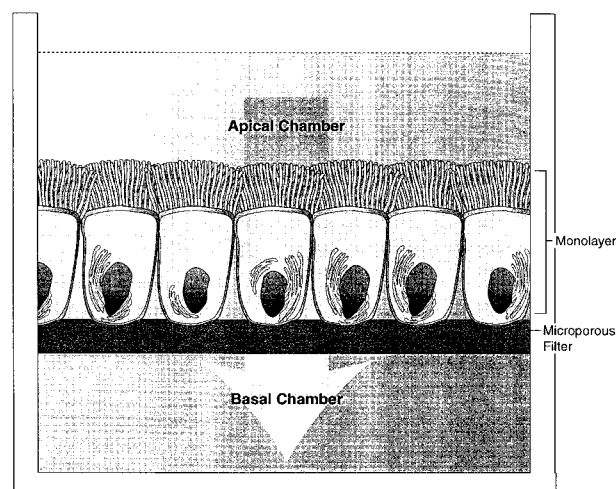


Fig. 8 Schematic representation of Caco-2 cell. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

that nonionic surfactants exhibit lower levels of toxicity than do ionic surfactants; however, the ionic surfactants are considered much more potent enhancers.

TECHNIQUES FOR ASSESSING THE INTESTINAL MEMBRANE PERMEABILITY

Various techniques have been used for assessing the permeability of drug molecules across the intestinal membrane including as everted sacs (23), brush border membrane vesicles (24), intestinal rings (24), recirculation method (25), modified in situ Doluisio technique (4), through-and-through in situ method (26), Caco-2 cell culture model (17), and human intubation method (27). During the last 10 years, the Caco-2 cell culture model (Fig. 8) has gained the most attention in the pharmaceutical industry. This human colon carcinoma cell line model has been shown to mimic the intestinal epithelial cells and is used for assessing the permeability of molecule (17). This Caco-2 cell model has been used as a routine screening tool for compounds produced by combinatorial chemistry.

In the Caco-2 cell experiment, the apparent permeability coefficient (P_{app}) is determined by the following equation:

$$P_{app} = \frac{1}{AC_{d,0}} \cdot \frac{dM}{dt} \quad (8)$$

where $C_{d,0}$ is the initial concentration of drug in the donor side, and M is the mass of the drug in the receiver side at the time t .

The apparent permeability coefficient is further defined in Eq. 9:

$$\frac{1}{P_{\text{app}}} = \frac{1}{P_{\text{aq}}} + \frac{1}{P_{\text{mono}}} + \frac{1}{P_{\text{F}}} \quad (9)$$

where P_{F} is the permeability coefficient of the filter support and P_{mono} is the permeability coefficient of the Caco-2 cell monolayer.

INTESTINAL ABSORPTION OF AMINO ACIDS, PEPTIDES, AND PROTEINS

Much is known regarding the uptake of food-derived amino acids and peptides from the GI tract. This is useful information to review because it gives insight into how peptide and protein drugs are likely to be absorbed.

Absorption of Amino Acids

The uptake of single amino acids is more active in the ileum than in the jejunum. The intestinal cell membrane contains many different amino acid transporters. Amino acids are absorbed across the brush border plasma membrane into the epithelial cell by certain specific amino acid transporters. Transporter types B, B^{0,+}, IMINO, β, and X_{AG}⁻ are sodium-dependent, and types b^{0,+} and y⁺ are sodium-independent transporters (1). The sodium-dependent transporters bind amino acids only after binding sodium. The amino acid-bonded transporter then undergoes a conformational change that dumps sodium and the amino acid inside the cell, followed by its reorientation back to the original form. The basolateral membrane of the absorptive cell contains additional transporters that carry amino acids from the cell into the blood. The amino acid transporters of the basolateral membrane are categorized as sodium-dependent types A and ASC and sodium-independent types asc, L, and y⁺ (1).

Transporter type X_{AG}⁻ is for the transport of acidic amino acids, and basic amino acids are preferred substrates for transporter type y⁺. The transporter types A and IMINO help the transport of amino acids such as proline and hydroxy-proline. The remaining transporters (B, B^{0,+}, β, b^{0,+}, ASC, asc, L) are for the transport of neutral amino acids. The intestinal absorption of amino acids is a stereochemically specific mechanism. The rate of absorption of the L-isomer is greater than that of the corresponding D-isomer when racemic mixtures of the amino acid are introduced in the small intestine (28).

Absorption of Small Peptides

After the break down of proteins by proteolytic enzymes, the pancreas, and brush border peptidases, the di- and tripeptides are absorbed through the epithelial cell membrane. Many studies have shown that intact di- and tripeptides are absorbed across the epithelial cell membrane by active transport via specific carrier systems. The absorption process is mediated by the hydrogen-coupled peptide transporter (PEPT1) located in the intestinal apical cell membrane (29). Because there are 20 amino acids, there may be 400 dipeptides and 8000 tripeptides with different molecular sizes and charges. This means that a single membrane transport system has a high affinity for di- and tripeptides but very low affinities for tetra- or higher oligopeptides. Peptide-like drugs such as angiotensin-converting enzyme (ACE) inhibitors, bestatin, cephalosporins, beta-lactam antibiotics, and renin inhibitors are recognized by peptide transporter because of their structural similarities with small peptides. Once inside the enterocyte, the vast majority of di- and tripeptides are digested into amino acids by intracellular peptidases and are exported from the cell into the blood. Only a very small number of these small peptides enter the blood intact. Peptides that are resistant to hydrolysis by cytoplasmic peptidases exit across the basolateral cell membrane by less well-studied basolateral peptide transporters. The sodium/hydrogen exchanger maintains the incoming proton gradient on the apical membrane side, whereas the sodium/potassium-ATPase in the basolateral

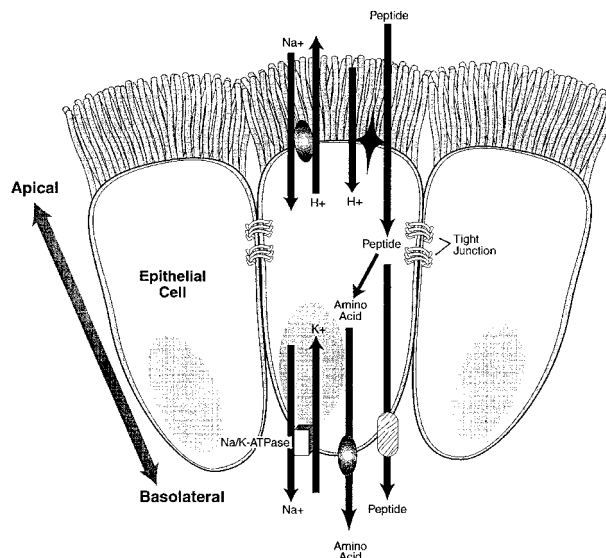


Fig. 9 Model of peptide transport across intestinal epithelial cells. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

Table 2 GI Absorption of peptides and peptidelike drugs

Compound	MW ^a	Extent of absorption (%)	Reference
Dietary di- and tripeptides	200–300	5–50	(30)
Aminocephalosporins	350	>50	(31)
Enalapril	377	>50	(32)
Dietary tetrapeptides	≈400	≈5	(28)
Thyrotropin-releasing hormone (TRH) analogs	≈400	≈5	(28)
Talampicillin	482	>50	(33)
Enkephalins	600	<2	(28)
Pepstatinyl glycine	740	<2	(28)
Cyclic somatostatin	806	<5	(34)
Bradykinin	1060	<2	(28)
Vasopressin	1200	<2	(28)
Cyclosporine	1203	>50	(35)
Leuprolide	1208	<5	(36)

^a Molecular weight.

membrane maintains a low intracellular sodium concentration. Thus, the sodium/hydrogen exchanger coupled with the sodium/potassium-ATPase drives the transport of di- and tripeptides across the epithelial cells (Fig. 9). The intestinal peptide transporter PEPT1 can transport di- and tripeptides but not free amino acids or tetrapeptides.

The intestinal absorption of dipeptides is a stereochemically specific mechanism. Dipeptides of L-L forms are transported more easily than are their isomers with a D-Amino acid. Dipeptide isomers of the L-L form have the highest absorption rates, followed by L-D and D-L isomers, and then D-D isomers.

The intestinal absorption mechanisms of amino acids and di- and tripeptides are characterized relatively well, as described above. However, only a limited amount of

work has been published for the absorption of tetra- and higher peptides, and the results are somewhat mixed. In general, it appears that tetra- or higher peptides are not well absorbed from the GI tract. Table 2 shows the oral bioavailability of peptides and peptide-like drugs (molecular weight <1500).

Absorption of Polypeptides and Proteins

Enterocytes of the intestinal membrane do not have transporters to carry polypeptides and proteins across the intestinal membrane, and they certainly cannot permeate through tight junctions because of their size. Also, polypeptides and proteins are substrates for luminal, brush border, and cytosolic enzymes. Therefore, as

Table 3 GI Absorption of polypeptide and protein drugs

Compound	MW ^a	Extent of absorption (%)	Reference
Beta-endorphin	≈3500	<2	(28)
Calcitonin	≈3500	<2	(28)
Corticotropin (ACTH)	≈4700	<2	(28)
Insulin	5700	0.5	(36)
Growth hormone	22,600	<2	(28)
Horseradish peroxidase (HRP)	40,000	3	(37)
Bovine serum albumin (BSA)	50,000	4.5	(38)

^a Molecular weight.

illustrated in Table 3, peptide/protein drugs are poorly absorbed across the GI tract.

One exception to the poor absorption of intact proteins is that neonates have the ability to absorb intact proteins for a few days after birth. Absorption cells of the neonatal mammalian intestine are functionally and morphologically specialized for the uptake and transport of milk macromolecules (39). The newborn has low levels of proteolytic enzymes in the intestinal lumen, and this highly unique endocytic mechanism allows intact proteins to adhere to the epithelial cell membrane and transport across the cell membrane. This function is very important for newborns to acquire passive immunity by absorbing immunoglobulins in colostral milk. In contrast to humans and rodents, many animals such as cattle, sheep, horses, and pigs do not provide enough antibodies across the placenta, and the young are born without circulating antibodies. Therefore, for these animals, the acquisition of passive immunity after birth depends primarily or entirely on the uptake of immunoglobulins in the small intestine by the endocytic mechanism. In addition, the highly endocytic nature of the small intestine provides no protection to invasion of bacterial proteins and antigens. At maturation, there is an abrupt change in the epithelial morphology of the small intestine that results in the cessation of the highly endocytic nature of the epithelium and in the corresponding loss in the capacity to absorb intact proteins.

Although the adult intestine provides a more effective epithelial barrier than that in newborns, evidence indicates that small quantities of intact proteins are transported across intestinal epithelial cells. There are two possible mechanisms associated with the absorption of intact proteins: endocytosis and transport through the epithelium of Peyer's patches.

Endocytosis

The epithelial membrane of the GI tract consists of a continuous barrier of cells, which allows the transport of low-molecular-weight molecules by simple diffusion or various carrier processes. Macromolecules such as proteins may be absorbed from the intestinal lumen by cellular vesicular processes, through fluid-phase endocytosis (pinocytosis), or by receptor-mediated endocytosis or transcytosis (Fig. 6). In pinocytosis, extracellular fluid is captured within an epithelial membrane vesicle. It begins with the formation of a pocket when a localized region of the epithelial membrane sinks inwardly. As the pocket deepens, it pinches into the cytoplasm, forming a vesicle containing the macromolecule that had been outside the cell. As a consequence, the majority of macromolecules taken into the cell by the vesicle do not find their way into the underlying tissues and thence to the blood circulation.

Rather, the vesicle shuttles the macromolecule to an endosome, which fuses with a lysosome where the membrane is recycled back to the epithelial membrane and the macromolecule is degraded. For example, HRP is taken up by the cell through the endocytosis process in the rabbit jejunum, and approximately 97% of these molecules are degraded during passage through the lysosomal system (37). Pinocytosis is unspecific in the substances it transports. Certain viruses and micro-organisms get into the cells by the endocytic mechanism, but they are not broken down within the lysosomal system. Some parasites and micro-organisms actually live within the lysosomal system, whereas others can find their way into the cytoplasm through mechanisms that may involve the presence of certain structures on membrane surfaces or changes in lysosomal pH.

In contrast, receptor-mediated endocytosis is very specific. Embedded in the membrane are proteins with specific receptor sites exposed to the extracellular fluid. The receptor proteins are usually clustered in regions of the membrane called coated pits, which are lined on their cytoplasmic side by a fuzzy layer consisting of a protein called clathrin. When ligands bind to the receptor sites, they are carried into the cell by the inward budding of a coated pit to form a coated vesicle. Clathrin-coated vesicles become uncoated and fuse to form an endosome. Ligand and receptor dissociate within the endosome, and the receptor shuttles back to the cell surface. The endosome fuses with the lysosome on which ligand degradation occurs.

Transcytosis is a process by which an endocytic vesicle (endosome) carries its contents across the epithelial cell without fusion with a lysosome (Fig. 6). The majority of endocytosed proteins are degraded, indicating that the transcytotic pathway is a minor one, with most endocytosed protein being routed to lysosomes. According to careful measurement (37), only 3% of horseradish peroxidase (HRP) avoids fusion with the lysosomal compartments to reach the blood circulation. However, it is interesting to note that immunoglobulin A (molecular weight ≈ 160 kDa) and immunoglobulin M (molecular weight ≈ 970 kDa) internalize into the cell by receptor-mediated endocytosis and are conveyed from one side of the cell to the other without fusing with the lysosome. It would be very interesting to explore whether the transcytosis process of immunoglobulins and the above-mentioned endocytosis processes of microorganisms and parasites, in which no degradation by the lysosomes occurs, could be harnessed for the delivery of peptides and proteins.

Transport through the epithelium of Peyer's patches

Although the GI tract has effective barrier properties, the mucous membrane is one of the major sites of entry for

most pathogens. The defense of these vulnerable membrane surfaces is provided by organized structures of lymphoid follicles known as Peyer's patches. Peyer's patches were discovered by Johanni Peyeri in 1677 and consist of 30 to 40 lymphoid nodules on the outer wall of the intestines (40). They are most prominent in the ileum of the small intestine in humans and are characterized by the presence of specialized epithelial cells called M (microfold or membranous) cells. Morphologically, M cells are quite different from absorptive cells. Their apical luminal surface contains numerous microfolds, truncated microvilli, and low levels of membrane-bound enzymes. In addition, M cells possess a relatively sparse glycocalyx and almost no lysosomes. Furthermore, the basolateral membrane surfaces of M cells are deeply invaginated toward the luminal side to form a pocket that is filled with a cluster of B cells, T cells, and macrophages (Fig. 10). All these unique structural characteristics of M cells are specialized for endocytosis of macromolecules or particles and transport to the lymphatic system. Several proteins such as native ferritin, HRP, and lectins are taken up by a pinocytosis process and traverse across M cells from the lumen to the extracellular space. This is in contrast to immunoglobulin A, which is internalized by a receptor-mediated endocytosis process, as described in the previous section. In addition to macromolecules, viruses and bacteria have been shown to gain entry via the Peyer's patches. The HRP concentration affects the transport process from the lumen to the lymph system; low levels of HRP are predominantly taken up by M cells, but higher levels are simultaneously taken up by M cells as well as by

absorptive cells (39). HRP is transported through the intestinal segments containing Peyer's patches more quickly in the intact form than through neighboring patch-free segments. The apparent permeability coefficient (P_{app}) of HRP in the segment containing Peyer's patches is approximately seven-fold larger than in the segment without Peyer's patches (41).

APPROACHES TO IMPROVE THE ORAL ABSORPTION OF PEPTIDES AND PROTEINS

As noted in the introduction to this article, the oral absorption of peptides (>3 amino acids) and proteins is very poor because of their potential degradation by the strong acidic environment and enzymes in the GI tract. Also, peptides and proteins have a very low permeability across the GI membrane.

Various formulation concepts have been introduced as potential ways to protect peptide and protein drugs from the hostile GI environment to increase their oral absorption such as use particulate drug carriers (microspheres, lipo-somes, and lectins), coadministration of enzyme inhibitors and absorption enhancers, use of chemical modification (prodrug), and site-specific delivery to the colon or rectum. Some of these approaches are discussed later.

Proteinoid Microspheres

Protenoids are thermally condensed amino acids and spontaneously form microspheres when exposed to an acidic medium (42). The microsphere size is approximately 1–5 μm in diameter. The proteinoid microspheres are very stable at lower pH conditions of 1 to 3 but unstable at the pH range of 6 to 7. Therefore, the proteinoid microspheres are able to protect the peptide and protein drugs from the gastric acids and enzymes while in the stomach and to release the encapsulated drug in the small intestine to be available for absorption. The proteinoids are also able to inhibit the activity of peptidases such as trypsin and chymotrypsin. Oral administration of proteinoid encapsulated insulin in diabetic rats has shown a significant hypoglycemic effect and provided a longer duration of action than when administered subcutaneously. Also, proteinoid-encapsulated calcitonin has been shown to significantly decrease serum calcium levels in treated rats compared with rats in a placebo control group. Both results demonstrate that proteinoid encapsulation can enhance the oral absorption of peptide and protein drugs (42).

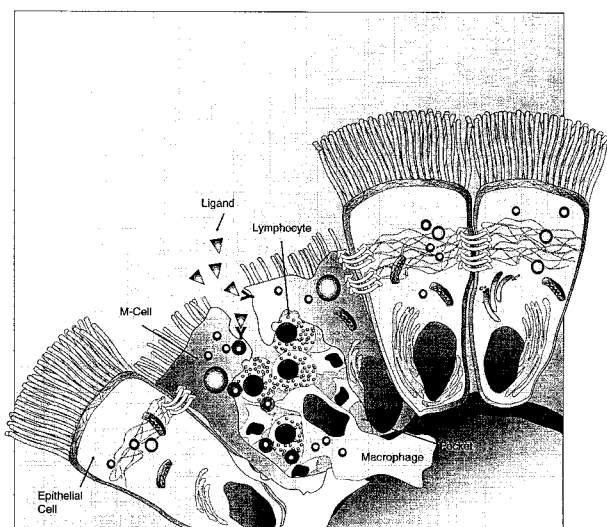


Fig. 10 Schematic depiction of the M cell. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

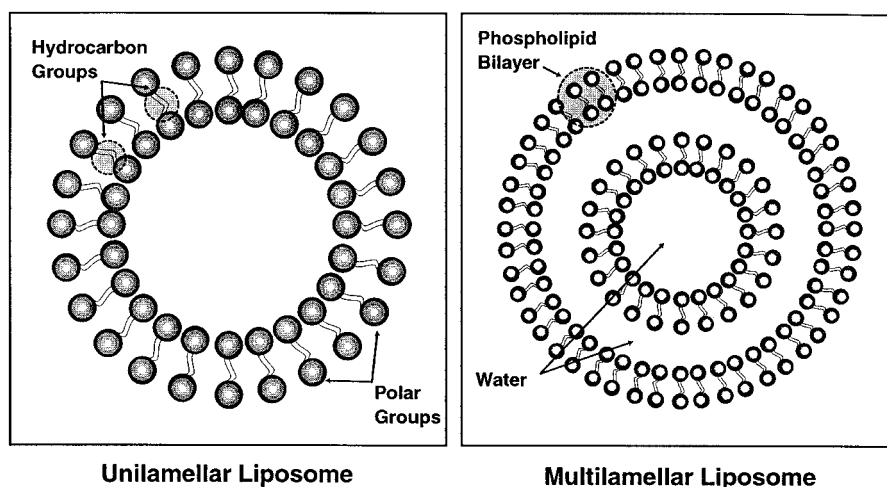


Fig. 11 Types of liposome: unilamellar liposome and multilamellar liposome. (Illustration by Leigh A. Rondano, Boehringer Ingelheim Pharmaceuticals, Inc.)

Liposomes

Liposomes have received considerable attention as a possible delivery tool for peptide and protein drugs by protecting labile compounds from degradation or by enhancing the uptake of poorly absorbed compounds. Liposomes have been studied extensively as a potential oral delivery system for proteins, especially insulin, and the oral administration of liposome-entrapped insulin into diabetic rats has produced a significant fall in blood glucose levels (43). There are three types of liposomes: multilamellar liposomes (MLV), 0.05–10 μm small unilamellar liposomes (SUV), 0.025–0.05 μm and large unilamellar liposomes (LUV), 0.2–2 μm (Fig. 11).

Liposomes are defined as vesicular lipid bilayers that enclose a volume of aqueous solution. Liposomes can be formed from a variety of phospholipids. The most widely used lipid is phosphatidylcholine. The liposome's outer membrane is semipermeable, and its permeability can be altered by varying the types of lipids used in preparing the liposomes. Liposomes can be made with a high permeability by using unsaturated lipids that will make a more fluid membrane. However, liposome with low permeability can be obtained by using a mix of phospholipids and cholesterol. Cholesterol is known to condense the packing of the phospholipids, thereby reducing their permeability and increasing the stability of the phospholipid bilayers. Negatively charged liposomes can be prepared by using phosphatidylserine or phosphatidylglycerol, and positively charged liposomes are made using stearylamine. Three possible mechanisms in the membrane transport of liposome-entrapped drugs

have been proposed (44): free molecule mechanism, simultaneous free molecule and direct liposome/membrane mechanism, and direct liposome/membrane mechanism. The free molecule model is for water-soluble drugs and includes the diffusion of a drug molecule from the liposome into the lumen first and then by membrane transport of the free molecule. The simultaneous free molecule and direct liposome/membrane mechanism is the sum of the permeations attributed to the drug molecule in the lumen and the direct transfer of molecule between liposome and membrane. The last mechanism, the direct liposome/membrane mechanism, is for water-insoluble drugs. This mechanism likely occurs because the release kinetics from a liposome system into the lumen is negligible.

Since 1970, there have been numerous reports detailing the advantages of liposomes as drug-delivery systems. However, in general these systems have not been successful for oral delivery. Some of the reasons may be related to: 1) their lack of stability in the GI tract and 2) their susceptibility to changes in pH, bile salts, and lipases. In addition, it has been shown that liposome systems are immunogenic.

Lectins

Lectins are known to have specific binding properties to the epithelial cell surface and have been tested as a possible oral delivery system for peptide and protein drugs. The lectin-coated nanoparticles that contain peptide or protein drugs can protect against degradation in the lumen of the small intestine and facilitate the uptake of

peptide or protein drugs across M cells by acting as a specific targeting ligand.

Coadministration of Enzyme Inhibitors

To promote the oral absorption of polypeptide and protein drugs from the GI tract, the enzyme barrier must be overcome. Various protease/peptidase inhibitors have been tested to enhance the oral bioavailability of peptide and protein drugs. To date, the known inhibitors are aprotinin, bacitracin, Bowman–Birk inhibitor, camostat mesilate, soybean trypsin inhibitor, sodium glycocholate, and chymotrypsin inhibitor (FK-448). Some of these inhibitors have absorption-enhancing activity in addition to enzymatic inhibition activity. As an example, insulin has often been used as a model protein drug in studies on enzymatic inhibition activity. Trypsin inhibitor and aprotinin have shown a marginal effect on increasing insulin absorption in rats. However, a significant hypoglycemic effect has been observed after administration of insulin with sodium glycocholate, camostat mesilate, and bacitracin (45), which may be related to the fact that these compounds act not only as enzymatic inhibitors but also as absorption enhancers. Although various potential candidates for enzyme inhibition and absorption enhancers have been identified, the long-term safety of these compounds in humans must be evaluated further.

Coadministration of Absorption Enhancers

Most polypeptide and protein drugs show low permeability across the intestinal membrane because of their polarity and size. Therefore, one approach to increase the permeability of these drugs is to coadminister with absorption enhancers. Thus far, the known enhancers tested for the oral delivery of peptide and protein drugs are bile salts, nonionic surfactants, anionic surfactants, lysolecithin, amines, medium chain glycerides, and salicylates. Although their mechanisms of action are not well understood, several possible mechanisms have been proposed, especially for insulin absorption, which include: 1) that some enhancers, such as bile salts, act not only as an absorption enhancer but also as an enzyme inhibitor; 2) that the enhancers act as a kind of dispersing agent to prevent aggregation of peptide and protein molecules in solution, resulting in increased solubility of the drugs; 3) that the enhancers reduce the viscosity of the membrane mucous layer and increase the membrane fluidity, resulting in increased absorption by opening up the aqueous channel on the cell membrane; and 4) that the positively charged enhancers may interact with the negatively charged epithelial cell membrane and neutralize the membrane

surface, resulting in increased absorption of the protein. Most enhancers are known to cause membrane irritation, and their long-term toxicity has not been well characterized and must be established (46).

Chemical Modifications

The reasons for poor absorption of most natural peptides and proteins from the GI tract are that most natural peptides and proteins are hydrophilic compounds and have a low partition coefficient (log of octanol/water partition coefficient, log P) on the order of -1 . For example, the log P values of TRH and vasopressin are -1.43 and -2.15 , respectively. There are some exceptions such as cyclosporine. Cyclosporine is a lipophilic peptide (log P = 3.0) and exposes very few polar groups on the surface. Medicinal chemists have routinely applied three types of chemical modification to improve the absorption of peptide and protein drugs including analogs, irreversible derivatives, and prodrugs. The irreversible derivative and the analog approaches are usually applied when compounds exhibit poor absorption because of in vivo metabolism. Peptide drug examples include enkephalins, TRH, and vasopressin, and insulin can be a model for protein drugs. The chemical modifications by the irreversible derivative and the analog approaches can prevent hydrolysis of peptides or proteins; however, these modifications may also reduce biological activity. For example, chymotrypsin, one of the major peptidases secreted by the pancreas into the intestinal lumen, is known to cleave at five bonds within the insulin molecule in a very short time. Four of these five residues play an important role in maintaining the biological activity of insulin and are essential for receptor binding ability (47). A number of chemical modifications at these sites could produce greater chemical stability; however, such changes in the insulin molecule could result in a loss in activity because of possible alterations of the three-dimensional structure of the protein.

The prodrug approach is the most widely applied chemical modification for improving the absorption of peptides. The unique feature of the prodrug approach is that the optimum physicochemical properties required for the drug with respect to lipophilicity and degradation can be achieved without altering the intrinsic biological activity of the parent drug. The prodrug itself is inactive. However, once the peptide prodrug is absorbed, it is converted to an active peptide, usually by an enzyme. A good example is an ACE inhibitor, enalapril, which was found to be orally well absorbed and metabolized to the active form, enalaprilat, in the liver. In contrast, the parent drug, enalaprilat, is very poorly absorbed via the oral route.

Colon Delivery

The colon has received considerable attention as a possible delivery site for peptide and protein drugs because enzymatic activities, especially peptidases, are significantly lower in the colon compared with the small intestine, and the residence time in the colon is longer than in the small intestine. However, there are some drawbacks for site-specific delivery of peptide and protein drugs in the colon as follows:

1. The colonic mucosa does not have the villi and microvilli of the small intestine. Therefore, the surface area available for absorption in the colon is considerably less than that in the small intestine.
2. Peptidases such as aminopeptidases and diaminopeptidases are at lower concentrations in the colon than in the small intestine. However, it has been shown that prolyl endoprotease and collagenase activities are five to six times higher in the colon than in the small intestine. TRH is more readily hydrolyzed to deaminated TRH in colonic homogenates compared with small intestine and rectum. Therefore, peptide drugs that are substrates for prolyl endoprotease and collagenase will likely be degraded in the colon and may not be suitable for colon delivery.
3. The concentration of bacteria in the colon, largely anaerobic species, is much higher than that in the small intestine. This high concentration of bacteria may lead to faster degradation of certain drugs such as digoxin, which is degraded to dihydrodigoxin by the microflora in the colon.

The microflora degradation mechanism has been exploited as a possible tool for the site-specific delivery of peptide and protein drugs. Peptide and protein drugs are coated with azoaromatic groups to form an impermeable film to protect them from digestion in the stomach and small intestine. When the polymer-coated peptide and protein drugs reach the colon, the colonic bacteria cleaves the azo bonds and breaks the polymer film, releasing the drugs into the lumen of the colon for absorption. This polymeric system was demonstrated to protect and deliver orally administered insulin and vasopressin (7) in rats.

Rectum Delivery

In general, oral delivery of polypeptide and protein drugs is limited because of degradation in the GI tract and poor absorption through the membrane. Rectal administration offers some advantages compared with oral delivery, such as low enzymatic activity, neutral pH, and partial avoidance

of first-pass metabolism. Even though the rectal route has long been known as a specific absorption site for the delivery of small lipophilic molecules, rectal delivery of large hydrophilic molecules is problematic because of poor absorption, and absorption enhancers may be required. Increased absorption of peptides, such as des-enkephalin-gamma-endorphin and desglycinamide arginine vasopressin, and proteins, e.g., albumin, insulin, and a somastatin analog, has been demonstrated by coadministration with enhancers. The enhancers used for the rectal absorption of insulin are surface active agents; bile acids; EDTA; and phospholipids such as lecithin, saponins, sodium salicylate, organic alcohols, acids, amines, and fats. Their mechanisms of action are generally not fully understood. Some of the possible enhancing mechanisms are described in "Coadministration of Enzyme Inhibitors," previously.

CONCLUSION

Rapid developments in biotechnology have led to the production of large quantities of pure potent and highly specific polypeptide and protein drugs. To date, they have been administered primarily by the parenteral route, even though the oral route is the patient's preferred choice. The oral route offers many advantages over parenteral administration, such as ease of administration and increased patient compliance. However, in general the peptide and protein drugs are very difficult to administer orally because they are of large molecular weight, hydrophilic in nature, and substrates for peptidases and proteases in the GI tract. There are many different locations where peptide and protein drugs can be hydrolyzed after oral administration. Their exposure to the strong hostile acidic environment of the stomach may cause partial degradation. Further degradation occurs in the lumen of the small intestine by the peptidases secreted by the pancreas, and then additional degradation by enzymes on the brush border and inside the epithelial cells follows. Even if some of the peptide and protein drugs survive and enter the cell, as they diffuse across the cell, they are taken up by lysosome where most of them are hydrolyzed by the lysosomal enzymes. Finally, those peptide and protein drugs that get through the epithelial cell into the portal vein could still be metabolized by the liver. In addition, they usually show poor permeability across the intestinal membrane because of their molecular size and polar nature.

To minimize degradation and improve absorption of peptide and protein drugs administered orally, several factors are significant. First, peptide and protein drugs

have to be protected from the acidic environment in the stomach and from enzyme degradation in the lumen and on the brush border of the small intestine. This may be accomplished by coadministration of peptidase inhibitors or by chemical modifications, such as analogs or prodrug approaches, or formulation approaches, such as microspheres and liposomes.

Second, because of their molecular size and hydrophilic nature, peptide and protein drugs are poorly absorbed through the intestinal membrane, and absorption has to be improved. This may be accomplished by coadministration of penetration enhancers or by chemical modifications such as increasing lipophilicity.

Finally, all these approaches may be possible; however, there is still a major hurdle to overcome: how intact peptide and protein drugs may pass across the inside of the epithelial cell without being taken up by the lysosome and also bypass first-pass metabolism in the liver. The best chance of avoiding lysosomal uptake and first-pass metabolism is by diffusion through the M cells on the Peyer's patches, which contain almost no lysosomes, and unloading the diffusates directly into the lymphatic system. This may be accomplished by exploring whether the transcytosis mechanism of antigens or microparticulates across the intestinal wall could be applied for the delivery of peptide and protein drugs.

In conclusion, although oral delivery is the preferred route for peptide and protein drugs, there are many drawbacks, as addressed previously. Successful oral delivery of peptide and protein drugs is likely to remain a formidable challenge for some time.

ACKNOWLEDGMENTS

We gratefully acknowledge Leigh Rondano of Boehringer Ingelheim Pharmaceuticals Inc., for her preparation of all the figures in this article. In addition, we thank Dr. Martha Brown, Mayur Dudhedia, and Mary Tanenbaum of Boehringer Ingelheim Pharmaceuticals, Inc. for their kind comments.

REFERENCES

- Kutchai, H.C. The Gastrointestinal System. *Physiology*, 4th Ed.; Berne, R.M., Levy, M.N., Eds.; Mosby: St. Louis, MO, 1998.
- Egberts, H.J.; Koninkx, J.F.; Dijk, J.; Mouwen, J.M. Biological and Pathobiological Aspects of the Glycocalyx of the Small Intestinal Epithelium. A review. *Vet. Q.* **1984**, *6* (4), 186–199.
- Russell-Jones, G.J.; Aizpurua, H.J. Vitamin B₁₂: A Novel Carrier for Orally Presented Antigens. *Proceedings of the International Symposium on Controlled Release of Bioactive Materials*, 1988; *15*, 142–143.
- Park, J.Y. Biophysical Model Approach to the Mechanistic Studies of Intestinal Absorption of Drugs. The University of Michigan: Ann Arbor, 1977; Thesis.
- Sallee, Y.L.; Wilson, F.A.; Dietschy, J.M. Determination of Unidirectional Uptake Rates for Lipids Across the Intestinal Brush Border. *J. Lipid Res.* **1972**, *13* (2), 184–192.
- Ugolev, A.M. *Physiology and Pathology of Membrane Digestion*. Plenum Press: New York, 1968.
- Saffran, M.; Kumar, G.S.; Savariar, C.; Burnham, J.C.; Williams, F.; Neckers, D.C. A New Approach to the Oral Administration of Insulin and Other Peptide Drugs. *Science* **1986**, *233* (4768), 1081–1084.
- McNally, E.J. *Protein Formulation and Delivery. Drugs and the Pharmaceutical Sciences*; Marcel Dekker, Inc.: New York, 2000; 99.
- Campbell, N.A. *Structure and Function of Macromolecules. Biology*, 2nd Ed.; The Benjamin/Cummings Publishing Co.: Redwood City, CA, 1990.
- Lehninger, A.L.; Nelson, D.L.; Cox, M.M. *Amino Acid Oxidation and the Production of Urea. Principle of Biochemistry*, 2nd Ed.; Worth Publishers: New York, 1993; 510.
- Adibi, S.A.; Moore, E.L. The Number of Glycine Residues Which Limit Intact Absorption of Glycine Oligopeptides in Human Jejunum. *J. Clin. Invest.* **1977**, *60* (5), 1008–1016.
- Tsukita, S.; Furuse, M. Pores in the Wall: Claudins Constitute Tight Junction Strands Containing Aqueous Pores. *J. Cell. Biol.* **2000**, *149* (1), 13–16.
- Ho, N.F.H.; Higuchi, W.I.; Turi, J. Theoretical Model Studies of Drug Absorption and Transport in the GI Tract III. *J. Pharm. Sci.* **1972**, *61* (2), 192–197.
- Ho, N.F.H.; Higuchi, W.I. Theoretical Model Studies of Intestinal Drug Absorption. IV. Bile Acid Transport at Premicellar Concentrations Across Diffusion Layer-Membrane Barrier. *J. Pharm. Sci.* **1974**, *63* (5), 686–690.
- Ho, N.F.H.; Park, J.Y.; Morozowich, W.; Higuchi, W.I. Physical Model Approach to the Design of Drugs with Improved Intestinal Absorption. *Design of Biopharmaceutical Properties Through Prodrugs and Analogs*; Roche, E.B., Ed.; American Pharmaceutical Association: Washington, DC, 1977; 136–227.
- Augustijns, P.F.; Bradshaw, T.P.; Gan, L.-S.L.; Hendren, R.W.; Thakker, D.R. Evidence for a Polarized Efflux System in Caco-2 Cells Capable of Modulating Cyclosporin a Transport. *Biochem. Biophys. Res. Commun.* **1993**, *197* (2), 360–365.
- Burton, P.S.; Conradi, R.A.; Hilgers, A.R.; Ho, N.F.H. Evidence for a Polarized Efflux System for Peptides in the Apical Membrane of Caco-2 cells. *Biochem. Biophys. Res. Commun.* **1993**, *190* (3), 760–766.
- Anderson, J.M.; van Itallie, C.M. Tight Junctions and the Molecular Basis for Regulation of Paracellular Permeability. *Am. J. Physiol.* **1995**, *269* (32), G467–G475.
- Daugherty, A.L.; Mrsny, R.J. Regulation of the Intestinal Epithelial Paracellular Barrier. *Pharmaceutical Science & Technology Today* **1999**, *2* (7), 281–287.
- Lutz, K.L.; Siahaan, T.J. Molecular Structure of the Apical Junction Complex and Its Contribution to the Paracellular Barrier. *J. Pharm. Sci.* **1997**, *86* (9), 977–984.

21. Knipp, G.T.; Ho, N.F.; Barsuhn, C.L.; Borchardt, R.T. Paracellular Diffusion in Caco-2 Cell Monolayers: Effect of Perturbation on the Transport of Hydrophilic Compounds that Vary in Charge and Size. *J. Pharm. Sci.* **1997**, *86* (10), 1105–1110.
22. Adson, A.; Raub, T.J.; Burton, P.S.; Barsuhn, C.L.; Hilgers, A.R.; Audus, K.L.; Ho, N.F.H. Quantitative Approaches to Delineate Paracellular Diffusion in Cultured Epithelial Cell Monolayers. *J. Pharm. Sci.* **1994**, *83* (11), 1529–1536.
23. Lukie, B.E.; Westergaard, H.; Dietschy, J.M. Validation of a Chamber that Allows Measurement of Both Tissue Uptake Rates and Unstirred Layer Thicknesses in the Intestine Under Conditions of Controlled Stirring. *Gastroenterology* **1974**, *67* (4), 652–661.
24. Osiecka, I.; Porter, P.A.; Borchardt, R.T.; Fix, J.A.; Gardner, C.R. In Vitro Drug Absorption Models. I. Brush Border Membrane Vesicles, Isolated Mucosal Cells and Everted Intestinal Rings: Characterization and Salicylate Accumulation. *Pharm. Res.* **1985**, *2* (6), 284–292.
25. Kakemi, K.; Arita, T.; Muranishi, S. Absorption and Excretion of Drugs. XXV. On the Mechanism of Rectal Absorption of Sulfonamides. *Chem. Pharm. Bull.* **1965**, *13* (7), 861–869.
26. Ho, N.F.H.; Park, J.Y.; Ni, P.; Higuchi, W.I. Advancing Quantitative and Mechanistic Approaches in Interfacing Gastrointestinal Drug Absorption Studies in Animals and Humans. *Animal Models for Oral Drug Delivery in Man: In Situ and In Vivo Approaches*; Crouthamel, W., Sarapu, A.C., Eds.; American Pharmaceutical Association: Washington, DC, 1983; 27–106.
27. Schedl, H.P.; Clifton, J.A. Cortisol Absorption in Man. *Gastroenterology* **1963**, *44* (2), 134–145.
28. TenHoor, C.N.; Dressman, J.B. Oral Absorption of Peptides and Proteins. *Sciences Technologies Pratigues Pharma. Sci.* **1992**, *2* (4), 301–312.
29. Liang, R.; Fei, Y.-J.; Prasad, P.D.; Ramamoorthy, S.; Han, H.; Yang-Feng, T.L.; Hediger, M.A.; Ganapathy, V.; Leibach, F.H. Human Intestinal H⁺/Peptide Cotransporter. Cloning, Functional Expression, and Chromosomal Localization. *J. Biol. Chem.* **1995**, *270* (12), 6456–6463.
30. Nightingale, C.H.; Greene, D.S.; Quintiliani, R. Pharmacokinetics and Clinical Use of Cephalosporin Antibiotics. *J. Pharm. Sci.* **1975**, *64* (12), 1899–1927.
31. Ganapathy, V.; Leibach, F.H. Is Intestinal Peptide Transport Energized by a Proton Gradient. *Am. J. Physiol.* **1985**, *249* (12), G153–G160.
32. Yokohama, S.; Yamashita, K.; Toguchi, H.; Takeuchi, J.; Kitamori, N. Absorption of Thyrotropin-Releasing Hormone After Oral Administration of TRH Tartrate Monohydrate in the Rat, Dog and Human. *J. Pharm. Dyn.* **1984**, *7* (2), 101–111.
33. Bell, J.; Peters, G.E.; McMartin, C.; Thomas, N.W.; Wilson, C.G. Estimation of Gut Absorption of Peptides by Biliary Sampling. *J. Pharm. Pharmacol.* **1984**, *36* (Suppl.), 88.
34. Wood, A.J.; Lemaire, M. Pharmacologic Aspects of Cyclosporine Therapy: Pharmacokinetics. *Transplant. Proc.* **1985**, *17* (4 Suppl 1), 27–32, Review.
35. Okada, H.; Yamazaki, I.; Ogawa, Y.; Hirai, S.; Yashiki, T.; Mima, H. Vaginal Absorption of a Potent Luteinizing Hormone-Releasing Hormone Analog (Leuprolide) in Rats I: Absorption by Various Routes and Absorption Enhancement. *J. Pharm. Sci.* **1982**, *71* (12), 1367–1371.
36. Crane, C.W.; Luntz, G.R. Absorption of Insulin from the Human Small Intestine. *Diabetes* **1968**, *17* (10), 625–627.
37. Heyman, M.; Ducroc, R.; Desjeux, J.F.; Morgat, J.L. Horseradish Peroxidase Transport Across Adult Rabbit Jejunum in Vitro. *Am.J.Physiol.* **1982**, *242* (5), G558–G564.
38. Kimm, M.H.; Curtis, G.H.; Hardin, J.A.; Gall, D.G. Transport of Bovine Serum Albumin Across Rat Jejunum: Role of Enteric Nervous System. *Am. J. Physiol.* **1994**, *266* (29), G186–G193.
39. Gornellia, P.A.; Walker, W.A. Macromolecular Absorption in the Gastrointestinal Tract. *Adv. Drug Del. Rev.* **1987**, *1*, 235–248.
40. Kuby, J. *Immunology*, 2nd Ed.; Feeman, W.H., Ed.; W.H. and Company: New York, 1994; 76–77.
41. Ho, N.F.H.; Day, J.S.; Barsuhn, C.L.; Raub, T.J. Biophysical Model Approaches to Mechanistic Trans epithelial Studies of Peptides. *J. Controlled Rel.* **1990**, *11*, 3–24.
42. Steiner, S.; Rosen, R. Delivery Systems for Pharmacological Agents Encapsulated with Proteinoids. US. Patent 4925673, 1990.
43. Patel, H.M.; Ryman, B.E. Oral Administration of Insulin by Encapsulation Within Liposomes. *FEBS Lett.* **1976**, *62* (1), 60–63.
44. Ganesan, M.G.; Weiner, N.D.; Flynn, G.L.; Ho, N.F.H. Influence of Liposomal Drug Entrapment on Percutaneous Absorption. *Int. J. Pharm.* **1984**, *20*, 139–154.
45. Bernkop-Schnürch, A.; Marschütz, M.K. Development and In Vitro Evaluation of Systems to Protect Peptide Drugs from Aminopeptidase. *N. Pharm. Res.* **1997**, *14* (2), 181–185.
46. Zhou, X.H. Overcoming Enzymatic and Absorption Barriers to Non-Parenterally Administered Protein and Peptide Drugs. *J. Controlled Rel.* **1994**, *29*, 239–252.
47. Pullen, R.A.; Linsay, D.G.; Wood, S.P.; Tickle, I.J.; Blundell, T.L.; Wormer, A.; Krail, G.; Brandenburg, D.; Zahn, H.; Glieman, J.; Gammeltoft, S. Receptor-Binding Region of Insulin. *Nature* **1976**, *259* (5542), 369–373.

FURTHER READING

- Amidon, G.L.; Wolfgang, S. Membrane Transporters as Drug Targets. *Pharmaceutical Biotechnology*; Kluwer/Plenum Publishers: New York, 1999; 12.
- Bai, J.P.F.; Chang, L.-L.; Guo, J.-H. Targeting of Peptide and Protein Drugs to Specific Sites in the Oral Route. *Crit. Rev. Ther. Drug Carrier Syst.* **1995**, *12* (4), 339–371.
- Davis, S.S.; Illum, L.; Tomlinson, E. NATO ASI Series. Series A: Life Sciences. *Delivery Systems for Peptide Drugs*; Plenum Press: New York, 1986; 125.
- Lee, V.H.L. Peptide and Protein Drug Delivery. *Advances in Parenteral Sciences*; Marcel Dekker, Inc.: New York, 1991; 4.
- Woodley, J.F. Enzymatic Barriers for GI Peptide and Protein Delivery. *Crit. Rev. Ther. Drug Carrier Syst.* **1994**, *11* (2,3), 61–95.

PEPTIDES AND PROTEINS—PULMONARY ABSORPTION

Igor Gonda

Aradigm Corporation, Hayward, California, U.S.A.

INTRODUCTION

The human respiratory tract has the potential to provide the means for noninvasive drug delivery of molecules that could not be efficiently and reproducibly, or rapidly, delivered without injecting them into the body. From the early parts of the 20th century, there have been many attempts to use inhalation to deliver insulin for the treatment of diabetes (1). These early approaches failed because the aerosol generators (jet nebulizers and metered dose inhalers) had low and variable delivery to the lung. Inhalation therapy for the treatment of diseases of the respiratory tract, especially asthma, underwent spectacular growth in the last three decades of the 20th century (2). The discovery of potent drugs that could be administered in one or two puffs from a low cost hand-held device, such as a metered dose inhaler or dry powder inhaler, led to the acceptance of aerosols as the preferred mode of therapy of asthma. One of the main advantages of targeting the lung was the minimization of systemic side effects of these drugs. Nevertheless, it was observed already in the early development of these asthma products that systemic absorption of the inhaled drugs did occur (3).

The biotechnology revolution resulted in the production of many potentially valuable therapeutic proteins. However, it was recognized that these molecules had properties that necessitated the use of injections for their administration. Intensive research in noninvasive drug delivery accompanied the dawn of recombinant biologics, but it met with somewhat limited success. Early work on peptides such as insulin (1) and some proteins (4) certainly indicated that such molecules were absorbed from the lung. Mackay et al. (5) reported in 1994 a summary of extensive studies in which the absolute bioavailabilities of several proteins and peptides were measured in rat models of pulmonary, nasal, and colonic absorption. The compounds were human and salmon calcitonins, the human parathyroid hormone and its 34-peptide fragment, hirudin, and a hybrid alpha interferon. All routes of administration exhibited general reduction of bioavailability with increasing molecular

weight. However, the pulmonary route (via intratracheal administration) resulted in much higher values than the other two routes. Thus, pulmonary delivery was known to be exceptionally promising among the noninvasive routes of delivery of peptides and proteins, but the key reasons for its poor delivery efficiency and variability needed to be investigated and removed. Industrial and academic research in the latter part of the last century demonstrated the need for targeting of the drug into “deep lung” to increase the efficiency of delivery and in particular to avoid the highly variable deposition in the oropharyngeal cavity (6). The conventional therapeutic aerosol generators used for delivery of asthma drugs to the airways were not designed with these requirements in mind. New types of delivery systems had to be developed that could generate very fine particles with aerodynamic properties suitable for drug delivery to the distal part of the respiratory tract with large highly absorptive surfaces.

SAFETY OF PROTEINS AND PEPTIDES DELIVERED TO THE LUNG

Satisfactory efficiency of delivery is only meaningful when it goes hand-in-hand with adequate safety. The inhalation route in general has been associated with a very good safety record for the delivery of asthma drugs evidenced both during the development stages and by their wide acceptance postapproval. The experience with inhaled protein and peptide drugs is so far relatively limited. Overall, the pulmonary safety of inhaled proteins and peptides has been good (7). The most extensive preclinical and clinical experience thus far is available for the recombinant human deoxyribonuclease (rhDNase) approved by inhalation for the treatment of cystic fibrosis in 1993 (8, 9). Inhaled leuprolide, a small peptide, was extensively studied with “clean” preclinical and clinical safety findings (10). A new large safety data base is being generated in conjunction with late stage development of insulin by several companies (11, 12).

ABSORPTION AND OTHER CLEARANCE MECHANISMS IN THE RESPIRATORY TRACT

Absorption of the protein and peptide molecules delivered to the respiratory tract competes against various other clearance mechanisms. When a particle containing a therapeutic substance is deposited in the respiratory tract, the following mechanisms of the drug-containing particle clearance can take place:

Removal of particulates from the respiratory tract:

- Mucociliary clearance
- Phagocytosis
- Clearance of the phagocytosed particles

After the drug is released from the particle (e.g., by dissolution, diffusion, or erosion), the three above listed mechanisms can also clear the free drug. Additional clearance of the released drug can take place as follows:

Removal of molecules deposited in the respiratory tract:

- Metabolism
- Chemical (non-enzymatic) decomposition
- Binding to fluid and tissue components in the respiratory tract
- Lymphatic uptake
- Absorption into blood stream

Absorption, Phagocytosis, and Mucociliary Clearance

These clearance pathways are not distributed uniformly throughout the lung. The gas-exchange areas, for example, do not have the mucociliary clearance but they have alveolar macrophages that can phagocytose “foreign” materials. These alveolated regions also have a much greater and more permeable environment than the upper and conducting airways. The main barrier in the deep lung to transport into blood stream appears to be the alveolar epithelium rather than the endothelial cells. While the majority of the cells lining the alveoli are Type II, most of the surface area (~95%) is covered by Type I cells. Alveolar cell monolayers resembling Type I cells are therefore thought to be representative of the barrier for pulmonary absorption.

Several authors reported on the biphasic nature of peptide and protein absorption from the lung (13, 14). The rapid component was proposed to be representing paracellular absorption through water-filled pores. The origins of the slower absorption phase into the blood circulation are not well understood. The respiratory tract has a rich lymphatic system, and, in addition to transcellular

and paracellular transport into the blood circulation, it is possible that a significant portion of macromolecules ends up first in the lymphatic system and is only then slowly released into the blood circulation. The molecular weight dependence of transfer of materials into the intrapulmonary lymph nodes following intratracheal administration in the rat was investigated with a series of labeled dextrans. It was found that the threshold for increased ratio of blood/lymph transfer was between 10 and 20 kDa, similar to the large intestine. Certain absorption enhancers were found to increase this ratio, suggesting that the use of such materials could be exploited in the future to target therapeutics into the pulmonary lymph nodes (15).

The clearance mechanisms have been modeled quantitatively to get a better understanding of the effects of interplay of the complex kinetics on the magnitude and duration of residence in the lung (Fig. 1) (16). This model can be significantly simplified if only a few of the pathways dominate the transport and a only low-resolution compartmental analysis is experimentally feasible (17).

Colthorpe et al. (17) showed that the mucociliary clearance is the elimination pathway that contributes most to the reduction of bioavailability of insulin, at least in the rabbit model they studied. Thus, the less insulin is deposited on the conducting airways and the more that gets into the nonciliated gas exchange spaces, the higher is the bioavailability. Similar conclusions were reached studying the pulmonary absorption of growth hormone in the same model (18). The pulmonary to blood transfer rate constants were found to be also somewhat higher for the aerosol delivery compared with intratracheal administration. This is consistent with the pioneering work of Schanker's group (19), who showed in experiments with inhibited mucociliary clearance that the aerosol technique results in faster absorption rates presumably associated with the coverage of a bigger surface area of the respiratory tract and better penetration into the permeable lung periphery. This was confirmed more recently also in the work of Niven et al. (20) in which they compared instillation of liquid formulations, powder insufflation, and aerosol delivery of recombinant human granulocyte colony stimulating factor (rhG-CSF) and its pegylated derivatives in several species. They highlighted the fact that the nature of the distribution of the drug in the respiratory tract has a far greater impact on the rate and extent of absorption than even the molecular modifications. Similarly, it was shown (21) that the bioavailability of leuprolide administered to dogs increased with the downstream distance from epiglottis. In humans, Newman et al. (22) compared the bioavailability of a peptide molecule delivered to human volunteers in the form of inhaled “coarse” and “fine” aerosols. The ratio of

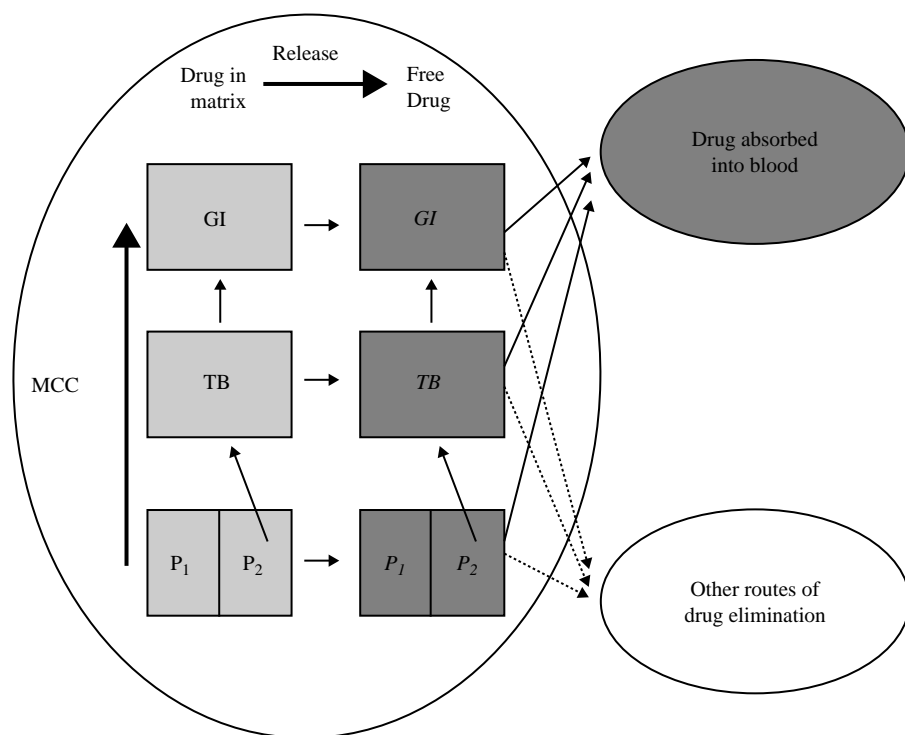


Fig. 1 Inhaled drug in a carrier can exist in the slow clearing pulmonary compartment (P_1) without mucociliary clearance or in the faster clearing pulmonary compartment with mucociliary clearance (P_2), tracheobronchial compartment (TB) or gastrointestinal tract (GI). When the drug is released into these compartment (the released drug is represented by italicized letters P_1 , P_2 , TB, GI), it can be absorbed into the bloodstream, or it can be removed by non-productive pathways such as mucociliary clearance represented by the vertical arrows, or chemical and enzymatic decomposition. Large molecules may also enter lymphatics before they appear in the blood stream. (The detailed model was described in Ref. 16 while the simplified three-compartment model represented by the ellipses is from Ref. 17.)

the bioavailabilities was almost identical to the ratio of the mass of the drug delivered to alveoli (as assessed by external gamma scintigraphic measurement). Getting protein deeper into the lung to access greater and more permeable surface areas that are not cleared by the mucociliary escalator thus results in better bioavailability.

Recent work with insulin provides evidence that the total lung volume at the end of the delivery impacts the kinetics of absorption of this peptide: delivery of fine particle insulin aerosol resulted in faster absorption with a higher plasma peak level in humans when the inhalation was done with a deep breath (close to vital capacity), as compared with a more shallow breath (about 50% of the vital capacity) (14, 23). The kinetics following the latter was similar to subcutaneous absorption of insulin. The exact reasons for this observation are unknown. However, the lung does have the above-described water channels that could expand during breathing. If the size of the peptide or protein molecule approaches the diameter of these channels, it would be expected that the channel expansion would lead to faster absorption. For molecules whose size

exceeds the channel diameter, the lung volume does not play a role in their pulmonary absorption rate (24).

Age, Disease, and Smoking Effects on Protein and Peptide Absorption

The effect of age on pulmonary absorption of three macromolecular markers, bovine IgG (BigG, MW = 150 kDa), bovine serum albumin (BSA, MW = 67 kDa), and 1-deamino-cysteine-8-D-arginine vasopressin (dDAVP, MW=1.067 kDa) was studied following intratracheal instillation in young and adult rats. The bioavailabilities for the three compounds were approximately 1.5, 5, and 20% in the adult rats. Low bioavailabilities were also found in the young rats for BigG and BSA but the absorption of dDAVP was significantly increased to 45% (25). Smoking profoundly increases the rate of absorption of insulin (26).

In experiments with aerosolized BSA and dDAVP in the rat (27), acute inflammation markedly increased the bioavailabilities.

Absorption Enhancers

The use of absorption enhancers for pulmonary delivery of peptides and proteins has been investigated in animal models but has not found commercial applications yet. This is presumably out of concern for potential safety impact from the use of such excipients. For example, Yamamoto et al. (28) showed that bacitracin and *N*-lauryl-beta-D-maltopyranoside in particular enhanced pulmonary absorption of insulin in a rat model but the toxicity of these enhancers was not investigated. The mechanism of “absorption enhancement” by bacitracin could have been in fact suppression of enzymatic degradation. Protease inhibitors were also found to improve the pulmonary absorption of a calcitonin analog (29) in a rat model but the mechanisms of this enhancement, or the potential pulmonary toxicity, were not studied. Changes to the barrier properties ranging from transient effects to complete stripping of epithelial cells have been among proposed mechanisms of pulmonary absorption enhancement.

Metabolism of Proteins and Peptides

There does not appear to be any significant proteolytic degradation by enzymes in the extracellular fluid in normal human airways and alveoli. However, the presence of peptidases makes absorption of small peptides variable and difficult to predict. Peptidase-resistant peptides generally show better bioavailabilities than other peptides with comparable molecular weights (30). Difference in metabolism and absorption of D and L forms of peptides glycyl-D-phenylalanine and glycyl-L-phenylalanine was demonstrated by Morimoto et al. (31). The L peptide was subject to metabolism, and it had significant paracellular transport with a smaller transcellular component. In contrast, the D peptide was not metabolized and it was only transported by passive diffusion via the paracellular route.

Given the complexity of the clearance mechanisms and the differences in the anatomy and physiology of the respiratory tract among various animal species, it is not surprising to find that the absorption rate constants for proteins and peptides are difficult to predict with any degree of accuracy. Fig. 2 shows the dependence of the apparent absorption rate constants as a function of molecular weight in dogs, rabbits, rats, and humans. While it is clear that generally the rate of pulmonary absorption is reduced for bigger molecules, the actual rate for a macromolecule in humans can be reliably determined at present only through direct experimentation.

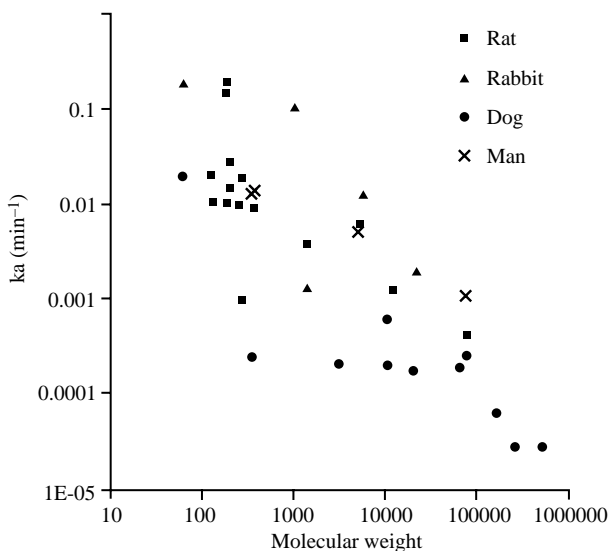


Fig. 2 Apparent pulmonary absorption rate constants as a function of molecular weight calculated from data in Refs. 17, 18, 32–34. (From Dr. Glyn Taylor, University of Wales, Cardiff, UK, and Dr. Stephen Farr, Aradigm Corporation, Hayward, CA.)

REGIONAL DEPOSITION OF INHALED PARTICLES AND DROPLETS IN THE HUMAN RESPIRATORY TRACT

Since the human respiratory tract is anatomically and physiologically a very heterogeneous system, the rate and extent of absorption of macromolecules as well as their potential adverse reactions depend on the regional doses. The most convenient method to deliver drugs to the respiratory tract is by inhalation. Other methods of delivery such as intratracheal instillation are used in experimental settings but are generally unsuitable for real-life therapeutic products. This section describes the basics underlying the deposition dosimetry following inhalation of particles.

Aerodynamic Particle Size

The probability of deposition of particles in the various parts of the human respiratory tract depends on their aerodynamic diameter, D_{ae} . This parameter in turn is a function of the physical dimensions, shape, and density of the particles. For spherical particles, the aerodynamic diameter is simply

$$D_{ae} = D \times \sqrt{(\rho/\rho_0)}$$

where D is the physical diameter, ρ is the density of the particle, and ρ_0 is unit density (in the old cgs units, it was

1 g/cm³). For droplets containing dilute aqueous solutions, the aerodynamic diameter is therefore equal to the geometric diameter of the droplet. For porous particles, the ratio of densities can be less than one, in which case the geometric diameter is larger than the effective aerodynamic diameter.

Mechanisms of Deposition

Particle deposition in the human respiratory tract takes place primarily by three mechanisms. Inertial impaction causes filtering of particles according to their aerodynamic size and velocity. This mechanism of deposition is especially important in the upper and central airways. In particular, oropharyngeal deposition increases approximately in proportion to the particle velocity and the square of its aerodynamic size. Therefore, particles with large aerodynamic diameters, especially if they are inhaled rapidly, will not be delivered to the lung for absorption. Gravitational sedimentation and diffusion are important for deposition of particles in airways and the alveolated regions. The extent of deposition by both of these mechanisms increases with breath-holding. Particles with aerodynamic diameters around 0.5 μm are neither small enough to deposit rapidly by diffusion, nor large enough to be deposited without prolonged breath-holding by sedimentation. It would be also practically quite difficult to produce submicron size aerosols containing proteins and peptides. The aim is therefore to produce a slow, moving aerosol cloud with drug containing particles in the size range 1–3 μm to minimize oropharyngeal deposition since peptides or protein are poorly absorbed from this region. Indeed, deposition of inhaled particles in this part of the human respiratory tract has been the primary cause of inefficiency and poor reproducibility of pulmonary delivery in the past (6).

Importance of the Mode of Inhalation on Deposition

The nose is a much better particle filter than the mouth, and oral inhalation is therefore preferred for more efficient and reproducible delivery to the lung.

We have already discussed the impact of inspiratory flow rate on deposition: the faster the subject inhales, the more material is deposited at bends and bifurcations in the upper and central airways, and the less material reaches the deep lung. Inspiratory flow rate may also affect the performance of inhalation systems; this is particularly important in passive dry powder inhalers in which the energy of breathing is utilized to deagglomerate the formulation (35). Thus, for such systems, there is a

contradictory requirement for the need to use high inspiratory flow rate to achieve fine particle size for deep lung delivery and yet to have sufficiently low inspiratory flow rate to avoid impaction before the drug entry to the absorptive surfaces devoid of mucociliary clearance.

It would be also expected that to get to the deep lung, the drug containing aerosol needs to be inhaled at the beginning of the inspiration, starting from “empty” lung since there is relatively little mixing of an inhaled “bolus” with the rest of the inhaled air (36). Therefore, the preferable maneuver is a full exhalation followed by a slow, deep inspiration, and finally a breath-holding period. In practice, a few seconds appears to be adequate for practically complete deposition of particles in the optimum size range for “deep lung” delivery.

Recurrent training of patients taking inhalation therapy is required to achieve reproducible delivery of inhalation therapy. Electronic systems with visual feedback to the patients have been developed to assist the patients to use their drug product correctly (36–38).

DETERMINATION OF PULMONARY ABSORPTION

The structural and functional heterogeneity of the respiratory tract also leads to the finding that generalization of results of pulmonary absorption experiments can only be made if the regional deposition of the materials in the respiratory tract is also determined. Intratracheal administration facilitates quantitative deposition of the material whose absorption properties are being studied. Several variants of this technique exist: some of them require surgical procedure (e.g., incision that exposes the rodent trachea through which a needle with the drug formulation is injected), some utilize the insertion of a cannula or a microspray nozzle into the animal's trachea. However, the material thus deposited is not distributed to the same extent throughout the respiratory tract as when given by inhalation of aerosols (17–20).

Lung cell cultures can provide mechanistic insights (39) but they do not represent the complexity in the delivery and disposition of drugs in the human respiratory tract. Perfused lung organ studies provide the next level in complexity (40). Various animal models have been used with the view to predict quantitatively absorption of peptides and proteins from the human lungs. However, due to the major differences in the anatomy and physiology of respiration in primates, the predictive power of these models is quite limited as evidenced by the data in Fig. 2. (Animal models are, of course, essential in the assessment

of safety and they can provide valuable mechanistic information.)

The use of drugs in conjunction with gamma radiation emitting radiolabels that facilitate noninvasive measurement of regional doses is the favorite way to investigate mechanistic aspects of pulmonary absorption. The drug can be radiolabeled directly, or, more often, the radiolabel and drug are physically mixed. It is important to prove that the radiolabel does not affect the performance of the drug product either chemically or in terms of its emitted dose (ED) and particle size distribution (PSD). Further, the functional performance, ED and PSD, measured *in vitro* with the drug assay and the radiolabel assay should be the same. This method can thus determine both the total dose to the lung as well as its distribution in various parts of the respiratory tract, both in animal (17, 18, 41–42) and humans (22, 43).

STABILITY ISSUES

A key issue for effective protein and peptide absorption is the preservation of biochemical and structural integrity during the preparation, storage, and aerosolization of the drug molecule. Many proteins and peptides are available as stable aqueous formulations for delivery by injection. However, the process of aerosolization can cause damage to the active molecule. The generation of small droplets provides a vast increase in the air-liquid interfacial area, which may cause unfolding of proteins followed by aggregation. This is particularly likely to happen for hydrophobic proteins that undergo multiple recirculation in jet nebulizers (44) or during spray-drying (45) and may be prevented by addition of suitable surfactants (42). Single-pass systems that form aerosol by extrusion of the solution through a fine nozzle (46) do not appear to cause protein denaturation, in contrast to conventional nebulizers that involve multiple recirculation of the solutions. Thermal denaturation can occur also during the high temperature spray-drying or with some ultrasonic nebulizers that warm the solution for nebulization during their operation (47). Freeze-dried parenteral protein preparations are typically unsuitable for delivery by inhalation for at least two reasons: they often contain excipients such as citrate that could cause irritation if inhaled in sufficient quantities. The freeze-dried materials form cohesive powders that do not lend themselves to be dispersed into respirable particles. Spray-drying has been employed to make respirable protein powders (48, 49). Judicious choice of the quality and quantity of excipients used in dry powders is required to minimize the potential

for adverse reactions, especially in subjects with compromised airways (50, 51). Avoidance of excessive temperature and the use of excipients such as certain sugars in spray-dried formulations may prevent the formation of potentially immunogenic protein aggregates and loss of activity. The nature of the formulation can also affect the quality and physical stability of the particles carrying the peptide or protein. These particles need to flow well during the filling operation and to provide high emitted doses, and they also have to disperse readily into particles with the right aerodynamic diameters. The appropriate balance thus needs to be struck between the chemical stability of the biologic and the physical stability on storage that preserves the solid state form and protects the particles against aggregation and loss of dispersibility into respirable particles (48, 51).

CASE STUDIES WITH SPECIFIC MOLECULES

Insulin

Pulmonary delivery of insulin for systemic absorption in the treatment of diabetes has been studied extensively since the early days of insulin discovery almost a century ago (1). Colthorpe et al. (17) and Pillai et al. (42) demonstrated in rabbit and monkey models, respectively, that the deeper into lung the dose of insulin was delivered, the higher was the bioavailability. The work of Laube et al. (43) showed the need to achieve deep pulmonary deposition of this molecule for efficient absorption in humans. Hand-held liquid and dry powder delivery systems have been developed to generate insulin-containing aerosols with the majority of the particles in the aerodynamic size range 1–3 μm . The relative bioavailability compared with subcutaneous injection based on the insulin contained in the dosage form was $\sim 11\%$ (52) for the dry powder system and $\sim 16\%$ (53) for the aqueous-based bolus delivery system. The reproducibility of pharmacokinetic and pharmacodynamic parameters following pulmonary administration was reported to be similar to subcutaneous delivery. The total lung volume at the end of inspiration was reported to have a major effect on insulin absorption, with deep breath leading to a significantly faster absorption than a more shallow inhalation (14, 23). This suggests that in addition to the inspiratory flow rate and inspired volume at the time of insulin delivery by inhalation, total lung volume may need to be controlled, too, for efficient and reproducible delivery of this molecule that has a rather narrow therapeutic index.

Human Growth Hormone

Wall and Smith (54) reviewed the literature data on pulmonary absorption of recombinant human growth hormone (rhGH) in animal models. A wide range of absolute bioavailability ranging from 3–36% was found. Bioavailability of hGH delivered to the lung increases dramatically the deeper into the lung the formulation reaches (18). Since rhGH is a growth factor, at least theoretically there is a possibility that its delivery to the lung could cause unwanted growth of some lung tissues. Aerosolized delivery of rhGH for 11 days induced body growth of hypophysectomized (“hypox”) rats but caused no abnormal lung growth in these animals. A mild immune reaction was seen in the lung, which was not present, when aerosolized bovine growth hormone (bGH) was used instead of hGH. bGH has a much closer amino acid sequence to rat growth hormone than hGH and is thus not significantly immunogenic in this species. Interestingly, IgG titers following subcutaneous administration of hGH were higher than those in the inhalation group and no IgE antibodies were found in the course of inhalation treatment in rats. While total body weight increased in the hypox rats treated with aerosolized hGH, the ratio of lung weight to total body weight did not change. GH receptors are present in many types of cells, but no GH receptor message was found on the “air side” of rabbit lung (55).

Luteinizing Hormone Releasing Hormone (LH-RH)

LH-RH analogs were extensively researched (10). The nonapeptide leuprolide in particular underwent extensive development. Absorption was very significantly dependent upon the depth of deposition of the material in animal models: essentially quantitative absorption was found when the intratracheal administration in beagle dogs was done 20–25 cm downstream from epiglottis. Leuprolide acetate suspension aerosols were prepared in chlorofluorocarbon propellants. No adverse reactions were found in association with multiple dosing of these formulations in beagle dogs. Bioavailabilities ranging from ~7–26% found in human studies with these metered dose inhalers, in good agreement with the general fraction of the nominal metered dose that reaches the lung from this type of aerosol delivery system (56).

Soluble Recombinant Interleukine-4 Receptor

(IL-4R) was delivered by nebulization to Cynomolgus primates (41). The peak serum levels were reached after

9.5 h. Absolute bioavailabilities close to 30% were measured.

Interferons

Niven et al. (13) studied the absorption of consensus interferon in rodent models. Absolute bioavailabilities approaching 70% were found for the pulmonary route, with evidence of biologically active molecules being absorbed. Mackay et al. (5) compared the absolute bioavailabilities of a hybrid BDBB alpha-interferon in the rat following different routes of administration and found 0.75% for colon, 0.014% for buccal, 0.5% for nasal, and 4.5% for the lung (via intratracheal instillation). Pharmacokinetics of inhaled alpha-interferon was followed in several human studies: Kinnula et al. (57) reported detectable levels with high doses and systemic adverse reactions indicative of biological activity of the absorbed cytokine. In a related study (58), it was observed that the recombinant version of the molecule seems to be less well absorbed than the natural interferon. Differences in metabolism between the natural and recombinant alpha-interferon were also found in a perfused rabbit lung model (59).

Calcitonin

The salmon form of this 32-amino acid non-glycosylated polypeptide is available as a nasal spray with a bioavailability in humans of ~3% (60). Mackay et al. (5) reported that the absolute bioavailabilities of human calcitonin across the colon of rat and man, and the lung (via intratracheal instillation) of rat, were 0.9, 0.15 and 36%.

Human Recombinant Granulocyte Colony Stimulating Factor

The systemic bioavailability of recombinant human granulocyte colony stimulating factor (rhG-CSF) based on the amount reaching the lung lobes following intratracheal instillation to the hamster was estimated to be 62% relative to intracardiac administration. Remarkably, about 20% of the absorbable dose of this 18.8-kDa protein was present in serum within 6 min postadministration. The biological activity of the absorbed material was confirmed by the increase of the circulating white blood cells (20). No apparent toxicity was found even on repeated dosing in animals with this protein. Pegylated forms of this molecule were also found to be well absorbed. The rate of and extent of absorption were very much affected by the mode of administration (instillation, powder insufflation, and aerosol administration), reflecting

Table 1

Physiological parameter	Breathing maneuver for optimum peptide and protein absorption
Lung volume prior to inhalation	Full exhalation
Volume inhaled at the time of actuation of aerosol delivery	Minimum volume inhaled prior to actuation of drug delivery
Inspiratory flow rate	The minimum flow rate consistent with acceptable performance of the aerosol delivery device
Total volume inhaled during delivery	Close to vital capacity (shown to be important for insulin)
Respiratory pause between inspiration and exhalation	Several seconds (minimum time to be determined by experimentation)

the differences in the penetration of the material into the deep lung.

PULMONARY ABSORPTION OF PEPTIDES AND PROTEINS INTENDED FOR LOCAL EFFECTS IN THE LUNG

Numerous proteins have been investigated for administration to the respiratory tract to treat local disease (61). In some of these investigations, absorption into the systemic circulation was also followed. These results have their utility when absorption of similar molecules is being considered for the purpose of systemic delivery or to estimate the systemic exposure of locally delivered macromolecules for safety purposes.

Cyclosporine

Cyclosporine is an endecapeptide isolated from fungi. It is an immunosuppressive used to prevent rejection in organ transplantation. Studies have been carried out to investigate the effectiveness of inhalation of this compound in patients receiving lung transplants. Very rapid peak plasma levels were observed followed by a slow phase but the interpretation of this behavior is compromised by the low aqueous solubility of this compound (62).

Alpha-1 Antitrypsin

(A1AT) has been used in injectable form for the treatment of hereditary A1AT deficiency that markedly increases the risk of development of emphysema. Absorption of aerosolized human plasma and recombinant A1AT into the lymph and blood was studied in sheep and humans (63). Human plasma A1AT was found in the sheep blood and interstitial lymph at concentrations ~1/1000 of that in the

alveolar epithelial lining fluid (ELF). The recombinant human A1AT is nonglycosylated and has a terminal methionine residue. In the sheep model, this molecule disappears from the alveolar fluid faster than the human plasma A1AT, with the lymph levels around 10% of ELF and blood levels about 10% of the lymph. The recombinant form was also detected in the blood 24 h after administration to humans.

Recombinant Human Deoxyribonuclease

(rhDNase) has been approved for administration by inhalation for the treatment of cystic fibrosis (9). Bioavailability <15 and <2% in rodents and monkeys respectively was found with this molecule in single-dose studies. In humans, the serum levels of rhDNase following nebulization did not lead to a significant increase above the baseline levels of the endogenous rhDNase.

Anti-IgE

Elevated levels of immunoglobulin E (IgE) in the respiratory lumen and blood are associated with allergic asthma. Sweeney et al. measured concentrations of a monoclonal humanized antibody against IgE (E25) in the blood and bronchoalveolar lavage of different animal species and humans following administration into the respiratory tract. Only small quantities of the antibody were absorbed over a period of several days. The authors suggested that the mechanism of uptake was nonspecific (64).

SUMMARY

While there are no currently approved therapeutic protein or peptide products for delivery via the pulmonary route

into the systemic circulation, the data on the efficiency, reproducibility, and safety of molecules such as insulin are particularly encouraging. From a molecular perspective, there is a general trend in the reduction of the rate and extent of absorption with increasing molecular weight, but predictive theories are lacking due to our poor understanding of the complexity and multitude of channels of entry of drugs from the respiratory tract into the systemic circulation. The aerodynamic size distribution is a key parameter affecting the regional distribution, and hence the absorption, of macromolecular drugs from the lung. The synchronization of optimum breathing and aerosol delivery appears to be a prerequisite for efficient and reproducible delivery; the key parameters to control are listed in Table 1.

Relatively little work has been done so far on the control of the rate of absorption of macromolecules from the lung with modified release formulations. Slow pulmonary absorption of insulin in the rat was reported when the drug was encapsulated in polylactic glycolic acid spheres (65). The long-term safety of such long-acting excipients needs to be investigated: in contrast to small molecules, it would be expected that significant accumulation of these materials could occur on multiple dosing (16).

REFERENCES

- Farr, S.J.; Taylor, G. Insulin Inhalation. Its Potential as a Nonparenteral Method of Administration. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, L.A., Gupta, P.K., Eds.; Ch. 13; New York, 1997; 371–378.
- Gonda, I. The Ascent of Pulmonary Drug Delivery. *J. Pharm. Sci.* **2000**, *89*, 940–945.
- Moss, G.F.; Jones, K.M.; Ritchie, J.T.; Cox, J.S.G. Plasma Levels and Urinary Excretion of Disodium Cromoglycate After Inhalation by Human Volunteers. *Toxicol. Appl. Pharmacol.* **1971**, *20*, 147–156.
- Takada, K.; Yamamoto, M.; Asada, S. Evidence for the Pulmonary Absorption of Fluorescent Labelled Macromolecular Compounds. *J. Pharm. Dyn.* **1978**, *1*, 281–287.
- Mackay, M.; Phillips, J.; Steward, A.; Hastewell, J. Pulmonary Absorption of Therapeutic Peptides and Proteins. *Respiratory Drug Delivery IV*; Byron, P.R., Dalby, R.N., Farr, S.J., Eds.; Interpharm Press, Inc.: Buffalo Grove, IL, 1994; 31–37.
- Gonda, I. Targeting by Deposition. *Pharmaceutical Inhalation Aerosol Technology*; Hickey, A.J., Ed.; Marcel Dekker, Inc.: New York, 1992; 61–82.
- Adjei, A.L.; Gupta, P.R., Eds.; *Inhalation Delivery of Therapeutic Peptides and Proteins* Marcel Dekker, Inc.: New York, 1997.
- Green, J.D. Pharmacotoxicological Expert Report Pulmozyme™ RhDNase Genentech, Inc.: Human Exptl. Toxicol. **1994**, *13*(suppl. 1), S1–S42, 7.
- Gonda, I. Deoxyribonuclease Inhalation. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.R., Eds.; Ch. 12; Marcel Dekker, Inc.: New York, 1997; 355–365.
- Adjei, L.A.; Lu, F.M.-Y. LH-RH Analogs. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.R., Eds.; Ch. 14; Marcel Dekker, Inc.: New York, 1997; 389–412.
- Berelowitz, M.; Becker, G. Inhaled Insulin—Clinical Pharmacology and Clinical Study Results. *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J., Eds.; Serentec Press, Inc.: Raleigh, NC, 2000; 151–154.
- Clauson, P.G.; Balent, B.; Brunner, G.A.; Sendhofer, G.; Jendle, J.H.; Hatorp, V.; Dahl, U.L.; Okikawa, J.; Pieber, T.R. PK-PD of Four Different Doses of Pulmonary Insulin Delivered with the AERx® Diabetes Management System. *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J., Eds.; Serentec Press, Inc.: Raleigh NC, 2000, 155–161.
- Niven, R.W.; Whitcomb, K.L.; Woodward, M.; Liu, J.; Jornacion, C. Systemic Absorption and Activity of Recombinant Consensus Interferon After Intratracheal and Aerosol Administration. *Pharm. Res.* **1995**, *12*, 1889–1995.
- Farr, S.J.; McElduff, A.; Mather, L.E.; Okikawa, J.; Ward, M.E.; Gonda, I.; Licko, V.; Rubsamen, R.M. Pulmonary Insulin Administration Using the AERx® System: Physiological and Physicochemical Factors Influencing Insulin Effectiveness in Healthy Fasting Volunteers. *Diabetes Technol. Ther.* **2000**, *2*, 185–197.
- Hanatani, K.; Takada, K.; Yoshida, N.; Nakasuji, M.; Morishita, Y.; Yasako, K.; Fujita, T.; Yamamoto, A.; Muranishi, S. Molecular Weight-Dependent Lymphatic Transfer of Fluorescein Isothiocyanate-Labeled Dextran After Intrapulmonary Administration and Effects of Various Absorption Enhancers on the Lymphatic Transfer of Drugs in Rats. *J. Drug Targeting* **1995**, *3*, 263–271.
- Gonda, I. Drugs Administered Directly into the Respiratory Tract: Modeling of the Duration of Effective Drug Levels. *J. Pharm. Sci.* **1988**, *77*, 340–346.
- Colthorpe, P.; Farr, S.J.; Taylor, G.; Smith, I.J.; Wyatt, D. The Pharmacokinetics of Pulmonary-Delivered Insulin: A Comparison of Intratracheal and Aerosol Administration to the Rabbit. *Pharm. Res.* **1992**, *9*, 764–768.
- Colthorpe, P.; Farr, S.J.; Smith, I.J.; Wyatt, D.; Taylor, G. The Influence of Regional Deposition on the Pharmacokinetics of Pulmonary-Delivered Human Growth Hormone in Rabbits. *Pharm. Res.* **1995**, *12*, 356–359.
- Brown, R.A., Jr.; Schanker, L.W. Absorption of Aerosolized Drugs from the Rat Lung. *Drug Metab. Disp.* **1983**, *11*, 355–360.
- Niven, R.W. Feasibility Studies with Recombinant Human Granulocyte Colony-Stimulating Factor. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.R., Eds.; Ch. 15; Marcel Dekker, Inc.: New York, 1997; 413–452.
- Qui, Y.; Gupta, P.K.; Adjei, A.L. Absorption and Bioavailability of Inhaled Peptides and Proteins. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.R., Eds.; Ch. 4; Marcel Dekker, Inc.: New York, 1997; 89–131.

22. Newman, S.P.; Hirst, P.H.; Pitcairn, G.R.; Clark, A.R. Understanding Regional Lung Deposition Data in Gamma Scintigraphy, Pulmonary Absorption of Therapeutic Peptides and Proteins. *Respiratory Drug Delivery VI*; Dalby, R.N., Byron, P.R., Farr, S.J., Eds.; Interpharm Press: Buffalo Grove, 1998; 9–16.
23. Farr, S.J.; Gonda, I.; Licko, V. Physicochemical and Physiological Factors Influencing the Effectiveness of Inhaled Insulin, Pulmonary Absorption of Therapeutic Peptides and Proteins. *Respiratory Drug Delivery VI*; Dalby, R.N., Byron, P.R., Farr, S.J., Eds.; Interpharm Press, Inc.: Buffalo Grove, IL, 1998; 25–33.
24. Egan, E. Lung Inflation, Lung Solute Permeability, and Alveolar Edema. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* **1982**, *53*, 121–125.
25. Folkesson, H.G.; Westrom, B.R.; Karlsson, B.W. Permeability of the Respiratory Tract to Different-Sized Macromolecules After Intratracheal Instillation in Young and Adult Rats. *Acta. Physiol. Scand.* **1990**, *139*, 347–354.
26. Kohler, D. Aerosols for Systemic Treatment. *Lung* **1990**, *168* (suppl.), 677–684.
27. Folkesson, H.G.; Westrom, B.R.; Dahlback, M.; Lundin, S.; Karlsson, B.W. Passage of Aerosolized BSA and the Nonapeptide DDAVP via the Respiratory Tract in Young and Adult Rats. *Exp. Lung Res.* **1992**, *18*, 595–614.
28. Yamamoto, A.; Umemori, S.; Muranishi, S. Absorption Enhancement of Intrapulmonary Administered Insulin by Various Absorption Enhancers and Protease Inhibitors in Rats. *J. Pharm. Pharmacol.* **1994**, *46*, 14–18.
29. Morita, T.; Yamamoto, A.; Takakura, T.; Hashida, M.; Sezaki, H. Improvement of the Pulmonary Absorption of (Asu 1, 7)-eel Calcitonin by Various Protease Inhibitors in Rats. *Pharm. Res.* **1994**, *11*, 909–913.
30. Patton, J.S.; Nagarajan, S.; Clark, A.R. Pulmonary Absorption and Metabolism of Peptides and Proteins. *Respiratory Drug Delivery VI*; Dalby, R.N., Byron, P.R., Farr, S.J., Eds.; Interpharm Press, Inc.: Buffalo Grove, IL, 1998; 17–24.
31. Morimoto, K.; Yamahara, H.; Lee, V.H.L.; Kim, K.J. Dipeptide Transport Across Rat Alveolar Epithelial Cell Monolayers. *Pharm. Res.* **1993**, *10*, 1668–1674.
32. Taylor, G.; Colthorpe, P.; Farr, S.J. Pulmonary Absorption of Proteins: Influence of Deposition Site and Competitive Elimination Processes. *Respiratory Drug Delivery IV*; Byron, P.R., Dalby, R.N., Farr, S.J., Eds.; Interpharm Press, Inc.: Buffalo Grove, IL, 1994; 25–30.
33. Huchon, G.J.; Montgomery, A.B.; Lipavsky, A.; Hoeffel, J.M.; Murray, J.F. Respiratory Clearance of Aerosolized Radioactive Solutes of Varying Molecular Weight. *J. Nucl. Med.* **1987**, *28*, 894–902.
34. Effros, R.M.; Mason, G.R. Measurements of Pulmonary Epithelial Permeability In Vivo. *Am. Rev. Resp. Dis.* **1984**, *129* (Suppl.), S59–S65.
35. Gonda, I. Physico-Chemical Principles in Aerosol Delivery. *Topics in Pharmaceutical Sciences 1991*; Crommelin, D.J.A., Middha, K.K., Eds.; Ch. 7; Medpharm Scientific Publishers: Stuttgart, 1992; 95–115.
36. Farr, S.J.; Rowe, A.M.; Rubsamen, R.; Taylor, G. Aerosol Deposition in the Human Lung Following Administration from a Microprocessor Controlled Pressurised Metered Dose Inhaler. *Thorax* **1995**, *50*, 639–644.
37. Phipps, P.R.; Gonda, I.; Anderson, S.D. Apparatus for the Control of Breathing Patterns During Aerosol Inhalation. *J. Aerosol Med.* **1992**, *5*, 155–170.
38. Gonda, I.; Schuster, J.A.; Rubsamen, R.M.; Lloyd, P.; Cipolla, D.; Farr, S.J. Inhalation Delivery Systems with Compliance and Disease Management Capabilities. *J. Contr. Rel.* **1998**, *53*, 269–274.
39. Mathias, N.R.; Yamashita, F.; Lee, V.H.L. Respiratory Epithelial Cell Culture Models for Evaluation of Ion and Drug Transport. *Adv. Drug Del. Rev.* **1996**, *22*, 215–249.
40. Byron, P.R.; Roberts, N.S.; Clark, A.R. An Isolated Perfused Rat Lung Preparation for the Study of Aerosolized Drug Deposition and Absorption. *J. Pharm. Sci.* **1986**, *75*, 168–171.
41. Pettit, D.K.; Moutvic, R.; Maliszewski, C.F.; Abbott, N.M.; Gombotz, W.R.; Smith, P.B.; Zhusti, A.; Slauter, R.W. Pharmacokinetics of Interleukin-4 Receptor Delivered to Primates by Nebulization and Intravenous Injection. *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J., Eds.; Serentec Press Inc.: Raleigh, NC, 2000; 405–406.
42. Pillai, R.S.; Hughes, B.L.; Wolff, R.K.; Heisserman, J.A.; Dorato, M.A. The Effect of Pulmonary-Delivered Insulin on Blood Glucose Levels Using Two Nebulizer Systems. *J. Aerosol Med.* **1996**, *9*, 227–240.
43. Laube, B.L.; Benedict, G.W.; Dobs, A.S. Time to Peak Insulin Level, Relative Bioavailability, and Effect of Site of Deposition of Nebulized Insulin in Patients with Noninsulin-Dependent Diabetes Mellitus. *J. Aerosol Med.* **1998**, *11*, 153–173.
44. Niven, R.W.; Ip, A.Y.; Mittelman, S.D.; Farrar, C.; Arakawa, T.; Prestrelski, S.J. Protein Nebulization: I. Stability of Lactate Dehydrogenase and Recombinant Granulocyte-Colony Stimulating Factor to Air-Jet Nebulization. *Int. J. Pharm.* **1994**, *109*, 17–26.
45. Mumenthaler, M.; Hsu, C.C.; Pearlman, R. Feasibility Study on Spray-Drying Protein Pharmaceuticals: Recombinant Human Growth Hormone and Tissue-Type Plasminogen Activator. *Pharm. Res.* **1994**, *11*, 12–20.
46. Schuster, J.A.; Rubsamen, R.M.; Lloyd, P.; Lloyd, J. The AERxTM Aerosol Delivery System. *Pharm. Res.* **1997**, *14*, 354–357.
47. Cipolla, D.C.; Clark, A.R.; Chan, H.-K.; Gonda, I.; Shire, S.J. Assessment of Aerosol Delivery Systems for the Recombinant Human Deoxyribonuclease I (rhDNase). *STP Pharma Sciences* **1994**, *4*, 50–62.
48. Clark, A.R.; Dasovich, N.; Gonda, I.; Chan, H.-K. The Balance between Biochemical and Physical Stability for Inhalation Protein Powders. *Respiratory Drug Delivery V*; Byron, P.R., Dalby, R.N., Farr, S.J., Eds.; Interpharm Press, Inc.: Buffalo IL, 1996; 167–174.
49. Chan, H.K.; Gonda, I. Solid State Characterization of Spray-Dried Powders of Recombinant Human Deoxyribonuclease (rhDNase). *J. Pharm. Sci.* **1998**, *87*, 647–654.
50. Anderson, S.D.; Brannan, J.; Spring, J.; Spalding, N.; Rodwell, L.T.; Chan, K.; Gonda, I.; Walsh, A.; Clark, A.R. A New Method for Bronchial-Provocation Testing in Asthmatic Subjects Using a Dry Powder of Mannitol. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 758–765.
51. Clark, A.R.; Shire, S.J. Formulation of Proteins for Pulmonary Delivery. *Protein Formulation and Delivery*;

- McNally, E.J., Ed.; Ch. 7; Marcel Dekker, Inc.: New York, 2000; 201–234.
52. Gelfand, R.A.; Schwartz, S.L.; Horton, M.; Law, C.G.; Pun, P.F. Pharmacological Reproducibility of Inhaled Human Insulin Pre-Meal Dosing in Patients with Type 2 Diabetes Mellitus. *Diabetes* **1998**, *47*, A99.
 53. Kipnes, M.; Otulana, B.; Okikawa, J.; Farr, S.; Jendle, J.; Thippawong, J.; Schwartz, S. Pharmacokinetics and Pharmacodynamics of Pulmonary Insulin Delivered Via the AERx Diabetes Management System in Type 1 Diabetics. Abstr. EASD, Jerusalem, Sept. 2000.
 54. Wall, D.A.; Smith, P.L. Inhalation Therapy for Growth Hormone Deficiency. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.R., Eds.; Ch. 16, Marcel Dekker, Inc.: New York, 1997; 453–469.
 55. Patton, J.S.; Platz, R.M. Routes of Delivery: Case Studies (2): Pulmonary Delivery of Peptides and Proteins for Systemic Action. *Adv. Drug Deliv. Rev.* **1992**, *8*, 179–186.
 56. Laube, B.L. In Vivo Measurements of Aerosol Dose and Distribution: Clinical Relevance. *J. Aerosol. Med.* **1996**, *9*(Suppl. 1), S77–S91.
 57. Kinnula, V.; Mattson, K.; Cantell, K. Pharmacokinetics and Toxicity of Inhaled Human Interferon-Alpha in Patients with Lung Cancer. *J. Interferon Res.* **1989**, *9*, 419–423.
 58. Maasilta, P.; Halme, J.; Mattson, K.; Cantell, K. Pharmacokinetics of Inhaled Recombinant and Natural Alpha Interferon. *Lancet* **1991**, *337*, 371.
 59. Bocci, V.; Pessina, G.P.; Pacini, A.; Paulesu, L.; Muscettola, M.; Mogensen, K.E. Pulmonary Catabolism of Interferons: Alveolar Absorption of 125 I-Labeled Human Interferon Alpha is Accompanied by Partial Loss of Biological Activity. *Antiviral Res.* **1984**, *4*, 211–220.
 60. Physicians' Desk Reference, PDR Electronic Library, Release 2000; Medical Economics Company.
 61. Adjei, A.L. Part Three: Localized Delivery of Peptides and Proteins to the Lung: Case Studies. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.R., Eds.; Marcel Dekker, Inc.: New York, 1997; 277–279.
 62. Burckart, G.J.; Keenan, R.; Griffith, B.P.; Iacono, A.T. Cyclosporine. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.R., Eds.; Ch. 8, Marcel Dekker, Inc.: New York, 1997; 281–299.
 63. Hubbard, R.C. Alpha-1 Antitrypsin. *Inhalation Delivery of Therapeutic Peptides and Proteins*; Adjei, A.L., Gupta, P.R., Eds.; Ch. 10, Marcel Dekker, Inc.: New York, 1997; 315–330.
 64. Sweeney, T.D.; Marian, M.; Achilles, K.; Bussiere, J.; Ruppel, J.; Shoenhoff, M.; Mrsny, R. Biopharmaceutics of Immunoglobulin Transport Across Lung Epithelium. *Respiratory Drug Delivery VII*; Dalby, R.N., Byron, P.R., Farr, S.J., Peart, J., Eds.; Serentec Press, Inc.: Raleigh, NC, 2000; 59–66.
 65. Edwards, D.A.; Hanes, J.; Caponnetti, G.; Hrkac, J.; Lotan, N.; Ben-Jebria, A.; Langer, R. Large Porous Biodegradable Particles for Pulmonary Drug Delivery. *Science* **1997**, *276*, 1868–1871.

PEPTIDES AND PROTEINS—TRANSDERMAL ABSORPTION

Richard H. Guy

M. Begoña Delgado-Charro

Diego Marro

Centre Interuniversitaire de Recherche et d'Enseignement, Archamps, France

University of Geneva, Switzerland

INTEREST AND RATIONALE

An increasing number of peptides and proteins will be introduced into therapeutics in the forthcoming years. The gradual maturation of the biotechnology industry has led to many advances, not the least of which is that it is now possible, for the first time, to produce large quantities of highly pure peptides and proteins. However, their potential usefulness as therapeutic agents will be much enhanced if constraints imposed by the parenteral route of delivery, the only one currently available for most of these agents, can be overcome. The design of successful dosage forms for the delivery of these complex biotechnology products by alternative routes is a challenging objective for the pharmaceutical scientist.

Possible nonparenteral routes for the delivery of peptide and protein drugs include the nasal, oral, transdermal, buccal, ocular, rectal, and vaginal pathways (1–3). In general, the primary advantages of these strategies over parenteral delivery are the relatively noninvasive nature and simplicity of administration. However, it goes without saying that these alternative routes are generally much less efficient than an injection (whether intravenous, intramuscular, or subcutaneous) with frequently very low bioavailabilities.

Transdermal delivery has attracted considerable interest as a route for administering peptides and proteins. Among its appealing features, the transdermal path a) avoids the hepatic first-pass effect and gastrointestinal breakdown, a very important factor for these metabolically very fragile drugs; b) provides controlled and sustained administration, particularly suitable for the treatment of chronic disease; c) reduces side-effects, often related to the peak concentrations of the circulating agent; d) enables self-administration and improves patient compliance, due to its convenience and ease of use; and e) permits abrupt termination of drug effect by simply removing the delivery system from the skin surface (3).

Because the skin has a relatively low proteolytic activity compared with other tissues, the poor inherent skin

permeability of the peptide drugs is the prime limiting factor for their delivery by the transdermal route (1). Generally speaking, peptide and protein drugs are polar and compared with “normal” drugs, of high molecular weight. Such molecules diffuse poorly, as a result, across the skin, making the use of an appropriate enhancement strategy (as we shall see) obligatory. Several approaches being investigated to enhance the transdermal absorption of peptide and protein drugs are reviewed in this chapter. These include the use of chemical penetration enhancers, iontophoresis, the application of transient high-voltage pulses (electroporation), the use of ultrasound (sonophoresis), and various, so-called minimally invasive strategies.

SKIN BARRIER FUNCTION

Stratum Corneum and Routes of Passive Permeation

The remarkable barrier function of the skin is primarily located in the stratum corneum (SC), the thin, outermost layer of the epidermis (4). The SC consists of several layers of protein-filled corneocytes (i.e., terminally differentiated keratinocytes) embedded in an extracellular lipid matrix. Attached to the outer corneocyte envelope are long-chain covalently bound ceramides that interact with the lipids of the extracellular space. These lipids are composed primarily of free fatty acids, ceramides, and cholesterol arranged in multiple lamellae (5). Passive permeation across the SC is believed to occur primarily via the intercellular lipid pathway (Fig. 1a), which constitutes the only continuous phase through the SC (6). Evidence for this deduction is provided by the temperature dependence of SC water permeability (7) and the remarkable correlation between water transport and the degree of disorder of the intercellular lipids (4). In addition, several imaging techniques, using various different tracers have been used to directly visualize permeation via the intercellular path (6, 8). Further, it has recently been concluded that the

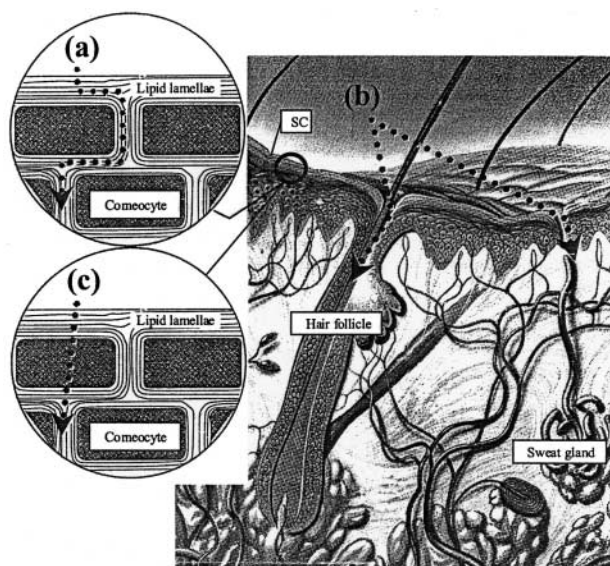


Fig. 1 Routes of passive permeation across the stratum corneum (SC): (a) intercellular lipid pathway; (b) appendageal transport; and (c) transcellular path.

barrier function of the SC is rather uniformly distributed across its entire thickness (9). Appendageal transport through hair follicles and sweat glands is another potential route, these structures offering “shunt” pathways across the continuity of the SC (Fig. 1b). Visualization of appendageal transport has been accomplished both for passive diffusion and for percutaneous transport enhanced by one means or another (e.g., iontophoresis) (6, 10). Passively, while it has been postulated that the appendageal structures dominate initially, their contribution to total steady-state transport is believed to be rather small due to the small area available relative to the entire surface of the skin (10, 11). A third possible route across the SC is the transcellular path (Fig. 1c). However, despite its favorable surface area and short pathlength (compared with the tortuous intercellular route), no evidence exists in support of molecular permeation by this way; indeed, the implausibility of the route has been demonstrated by certain authors (8, 12).

Structure–Permeation Relationships: Application to Peptide Absorption

Most structure–permeation evaluations of transdermal transport indicate the strong relationship between a chemical's permeability coefficient across the skin (K_p) and its lipophilicity (13–15). Potts and Guy (14) analyzed a diverse database of skin penetration data, comprising 91 compounds (15) and developed a simple relation indicating that K_p depended only on lipophilicity, as

measured by the chemical's octanol–water partition coefficient (P), and molecular weight (MW):

$$\log K_p(\text{cm/sec}) = -6.3 + 0.71 \log P - 0.0061 \times \text{MW}$$

However, for compounds of high lipophilicity, the algorithm needs to be modified (16) to account for the fact that these substances may become transport-limited not by their diffusion through the SC, but rather by their ability to partition out of this membrane into the underlying aqueous, viable epidermis. At the opposite end of the lipophilicity scale, the very small $\log P$ values of polar compounds, including peptide and proteins, of course, many of which retain a charge (or multiple charges) over a broad pH range, mean that their K_p values are extremely small and, hence, their inherent transdermal fluxes, without a suitable enhancement technique, are not therapeutically useful. As a general point, it must be emphasized that no matter what one might do to enhance transdermal delivery, the route remains useful exclusively for compounds of high potency (including peptides).

STRATEGIES FOR THE TRANSDERMAL DELIVERY OF PEPTIDES AND PROTEINS

Passive Delivery and Chemical Enhancement

It is widely accepted that the two most important physicochemical parameters that determine a molecule's skin permeability are a) its lipophilicity, with a \log (octanol/water partition coefficient) value of 2 being quite favorable, and b) its molecular size—smaller compounds permeating better than big ones (14). Thus, it is no surprise that the passive transdermal delivery of peptides and proteins, which are typically either very polar (or charged) and/or of high molecular weight (>1000 Da), is extremely inefficient and rarely results in fluxes, which would elicit significant therapeutic effect.

Consequently, a number of enhancer formulations have been considered in an attempt to improve peptide transport across the skin. These vehicles include, as we shall briefly discuss, either well-known chemical promoters or combinations thereof, peptide metabolism inhibitors, or colloidal (principally liposomal) structures. Table 1 summarizes much of the available literature. Before addressing specifically certain examples of the results obtained, it is worth noting that, although increased peptide transport has been possible, little attention has been focused upon the effects of the enhancer formulations on the skin (such as irritation and, in consequence, long-term tolerability) nor

Table 1 Effect of chemical enhancers and vehicle composition on the transdermal passive delivery of peptide drugs

Peptide	Vehicle	Skin model	Comments	Ref.
A. Passive + chemical enhancer Ocreotide [NL ⁴ , D-Phe ⁷]- α -MSH (melanotropic peptide analog)	40% EtOH; 1% Decylmethyl sulfoxide in 40% EtOH; 50% Dimethyl sulfoxide 26% PEG 400; 74% PEG 3350	Hairless mouse skin; heat-separated human epidermis Mice shaved skin	In vitro; clinically relevant delivery In vivo; follicular melanogenesis induction	(17) (18)
	Azone; PEG	Mouse skin; rat skin Dermatomed human skin Human SC and de-glycerinized dermatomed skins (Dutch Burns Society)	In vitro; delivery through mouse but not through rat skin In vitro; delivery measured In vitro; enhancers required for relevant delivery	(19) (20) (21)
	<i>n</i> -Decylmethyl sulfoxide (10 mM) + metabolism inhibitors (piromycin, amastatin) Sodium laurylsulphate (0.9 to 19.2 $\times 10^{-3}M$)	Hairless mouse skin Rat shaved skin	In vitro; study of the metabolism; coadministration of inhibitors to increase flux In vitro; effect of pH and peptide concentration; no effect of enhancer found	(22) (23)
	Azone (3–25%); dimethyl sulfoxide; DMAC; NNDMT; transcutoil; labrafil	In vitro hairless mouse/rats; in vivo rats; “clipping” hair shaved	Best results with Azone (clinical effect in vivo: reduction of urine volume and increase in urine osmolality)	(24)
	Carbopol gel + different enhancers and protease inhibitors: e.g., bile salts, gabexate, bestaline, OG, OTG.	Shaved rat	In vivo; effect of protease inhibitors; observation of the hypocalcemic effect	(25)
	Carbopol gel; vitamin D3 + Estradiol + OTG + aurocholate	Rats (suffering experimental osteoporosis); shaved skin Nude mouse; dermatomed human skin; snake skin	In vivo; efficiency in treating experimental osteoporosis in rats In vitro	(26) (27)
	2% HPMC; EtOH:water (4:1) + enhancer: menthol (1-2%); camphor (2%); methyl salicylate; (2%); lauric acid (2%); decanoic acid (2%); 20–80% ethylalcohol; 10% urea			
	PG, azone; glyceryl mono-oleate; EtOH; β -cyclodextrin In PBS solution	Dermatomed human and monkey skin Rat shaved skin	In vitro; 80 h permeation time Prodrug approach: Permeation of parent compound and acylated derivatives.	(28) (29)
	B. Passive + colloidal vehicle Oil/water emulsion liposomes	Hairless mouse skin	In vitro; stripping procedure; cyclosporin accumulated; high cyclosporin into the skin	(30)
	Liposomes Liposomes	Human skin grafted in nude mice Dermatomed human skin	In vivo; measurement of ICAM-1 induction In vitro; measurement of bio-activity of the transported peptide (10% active)	(31) (32)
Nafarelin Tetragastrin	Liposomes + iontophoresis Microemulsion liquid crystals	Human skin Stripped rat skin	In vitro In vitro	(33) (34)

Abbreviations: EtOH (ethanol); PEG (polyethylene glycol); SC (stratum corneum); DMAC (dimethyl acetamide); NNDMT (*N,N*-dimethyl-*m*-toluamide); OG (*n*-Octyl- β -D-glucoside); OTG (*n*-Octyl- β -D-thiogluconide); HPMC (hydroxypropylmethylcellulose); PG (propylene glycol); PBS (phosphate buffer solution); ICAM-1 (intercellular adhesion molecule-1).

has the fact that animal models of questionable relevance (e.g., the shaved skin of hairy rats, notorious for their higher skin permeability relative to man even in the absence of penetration enhancer) have been used. To illustrate the “gap” between laboratory and reality, with respect to the feasibility of some enhancer vehicles proposed, one can cite the delivery of octreotide from a formulation containing 40% ethanol and 50% dimethylsulphoxide (17) and the delivery of leuprolide from a 4:1 v/v ethanol–water solution to which various known chemical skin permeation enhancers had been added (27). It is not foolhardy to predict that both these approaches would cause local irritation of the skin *in vivo* in man and that their chronic use would be precluded.

Among the more interesting and reasonable results published are a series of studies demonstrating the delivery of the melanotropic peptide [Nle⁴, D-Phe⁷]- α -MSH across mouse and human skin from a relatively mild formulation consisting of 1:3 v/v PEG 400-PEG 3350 (18–20). *In vivo* across shaved mouse skin, in fact, it was possible to show induction of melanogenesis (18). Unfortunately, however, these data remain at best semiquantitative: no actual flux values were measured precluding comparison with other work in the field.

Positive pharmacological results have also been observed following the delivery of the relatively high molecular weight peptide, elcatonin (3363 Da, a calcitonin analog) across shaved rat skin *in vivo* (25, 26). Hypocalcemia was documented together with a beneficial effect on this animal model of osteoporosis. Successful formulations included taurocholate and peptidase inhibitors, such as gabexate and bestatine. Another approach has involved the use of lipophilic prodrugs of the peptide tetragastrin but, in this case, one has to ask whether the improvement in delivery is sufficient to warrant the additional regulatory complexities concomitant with the prodrug strategy (29).

The largest body of recent work in this area has involved the use of colloidal carriers, in particular liposomes (see Table 1); comprehensive reviews of this research can be found elsewhere (35–37). A general observation on the information published to-date is that the amount of lipid formulation applied is often very high when compared with the quantity of intercellular SC lipid in the treated area of the skin. That such liposomal formulations can significantly perturb (or, perhaps, overwhelm is a better word) normal barrier function is not too surprising, especially if the exogenous lipids employed can mix efficiently with their endogenous counterparts (30, 32, 38); this may explain why liposomes based upon lipid mixtures that mimic the composition of those in the SC are sometimes useful.

These lipid-based systems have been used to deliver peptides for both dermatological and systemic therapy. In the former case, the administration of cyclosporin (CSA) to treat psoriasis has been considered, with the idealistic objective of somehow localizing the drug in the skin while minimizing its entry into the systemic circulation (30). *In vitro*, from the lipid-based vehicles, almost no CSA permeated into the receptor phase in conventional diffusion cell experiments. However, this result was not different from the “control” formulation, which simply contained CSA in 40% ethanol. Much of the drug (>60%), and the lipids from the liposomal-based formulation, were actually found in the SC at the end of the experiment. In the same category, a particularly well-conducted study considered the delivery of γ -interferon from liposomes across human skin *in vitro* (32); topical application of γ -interferon has been proposed for the treatment of atopic dermatitis and keloidal scarring. Initial investigation with γ -interferon suggested a quite short lag-time (on the order of an hour) but that the liposomally delivered peptide had lost most of its biological activity (32). In a subsequent series of experiments, the work was extended to an *in vivo* model of human-skin grafted nude (athymic) mice (31). It was found that the γ -interferon delivered in this case retained its biological activity and promoted epidermal ICAM-1 induction (Fig. 2).

With respect to systemic delivery of peptide and protein drugs encapsulated in lipid-based carriers, Cevc et al. (39) have investigated the transdermal delivery of insulin encapsulated in “elastic” liposomes (TransfersomesTM), which differ from conventional liposomes in that a surfactant is added to the formulation (e.g., sodium cholate, sodium dodecyl sulfate, or bile salts). A dose-dependent systemic hypoglycemia has been reported in human volunteers following topical application of such an insulin formulation, over an area of 45–90 cm²; the effect was typically 25–45% of that following a subcutaneous injection of the same formulation. The authors claim that this outcome is not the result of the disrupting effect of the ethanolic (10%), surfactant-rich formulation on the integrity of the skin barrier, since “rigid” liposomes (ethanolic liposomes without bile salt) and insulin/mixed micelles (which contain the highest bile salt concentration of all formulations studied) had no significant hypoglycemic effect. On the contrary, they argue that the specifically elastic properties of the transfersomes are the crucial features that allow the lipid vesicle to carry the encapsulated agent across the skin (39).

Finally, the topical application of proteins to intact skin as a noninvasive method for immunization and vaccination has been described recently. The cholera toxin (86 kDa) has been shown to induce significant antibody production when

topically applied to shaved, previously hydrated, mouse skin (40). The cholera toxin has also been shown to act as an immunizing adjuvant, that is, it enhances immune responses against less potent antigens, such as diphtheria toxoid, and tetanus toxoid, when coadministered topically (40). A subsequent report identified several other bacterial products that also elicit adjuvant properties, but having less risk of toxicity and reduced secondary effects than cholera toxin (41). Transdermal immunization has also been obtained in mice with large proteins, such as the gap junction protein, encapsulated in lipid vesicles (42, 43). A more sophisticated approach, the transdermal delivery of an antigen-encoding plasmid, which induces *in vivo* the production of the antigen protein, has also been investigated and will be discussed later in this chapter. Overall, the future development of noninvasive, easily administered vaccines is of much interest, particularly for massive immunization programs in developing countries, where the lack of trained personnel capable of performing accurately repeated injections (which are far from welcomed) poses a significant barrier to prophylactic health care.

Iontophoresis

Iontophoresis is a noninvasive technique that uses a mild electric current ($<0.5 \text{ mA/cm}^2$) to facilitate the transfer of molecules across the skin (Fig. 3). The frequently polar and/or charged nature of biotechnology drugs makes them

potentially well-suited candidates for electrically controlled drug delivery across the skin (44, 45). Indeed, as we shall see, human studies using iontophoresis have already demonstrated the safe, effective, and reproducible delivery of intact peptides to the systemic circulation. Specific accomplishments for certain peptide iontophoretic applications of particular note to date are: a) that steady-state plasma levels can be achieved relatively rapidly and then maintained for reasonably prolonged periods of time, until the current is stopped and b) that the current profile can be adjusted to change the delivery profile, and hence to preprogram perhaps the more complex dosing regimes required. In addition, for two small nonpeptide drugs, at least, small easily portable and integrated iontophoretic devices are in advanced stages (Phase III) of development.

Iontophoresis enhances drug delivery across the skin by two principal mechanisms: electrorepulsion and electroosmosis (46) (Fig. 4). Electrorepulsion is the direct effect of the applied electric field on a charged permeant. Practically speaking, this means, for example, that a positively-charged (cationic) drug is formulated at the anode and, upon applying the electric field, is repelled toward and through the skin. The drug ion acts, therefore, as a charge carrier in the electric circuit. However, the fraction of the total charge delivered by the power supply that is carried by the drug may be quite small and, for larger peptides, rarely approaches even a few percent. This is because there may be competing (e.g., buffer, electrolyte) cations in the formulation and there are always competing anions (especially Cl^-) carrying a significant amount of charge out of the skin toward the anode (Fig. 3). Formulations must therefore be optimized to minimize the presence of competing ions, at least in the direction from the skin surface into the body. The second mechanism, electroosmosis, results from the fact that the skin supports a net negative charge at physiological pH (47, 48) [the pI of the human skin has been shown to be within the range 4.5–5 (47)]. Imposing an electrical potential gradient across a charged membrane produces a convective solvent flow in the direction of counterion transport (i.e., from anode-to-cathode in the case of skin). This solvent flow therefore augments the electrotransport of cationic, and very polar, yet neutral, compounds, while acting against the electromigration of anions (Fig. 4).

The relative importance of electrorepulsion and electroosmosis has been the subject of much investigation (49–58). It is generally agreed that small, highly mobile ions (e.g., Na^+ , Cl^- , small charged amino acids) are principally moved across the skin by electrorepulsion, whereas large, bulky species carrying only a fraction of the charge passing across the skin can only be transported by electroosmosis (Fig. 5). As molecular size increases for

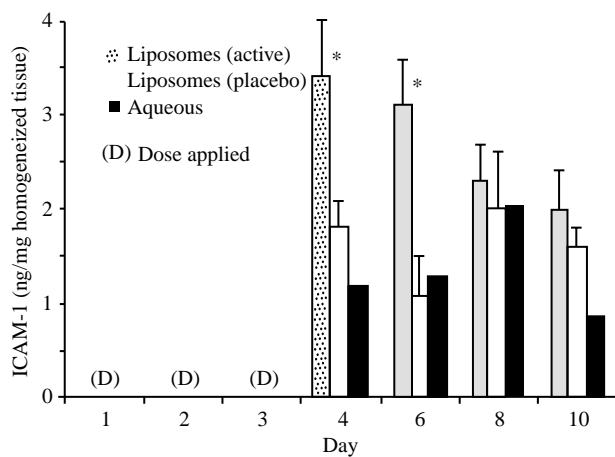


Fig. 2 ICAM-1 production after topical application of gamma interferon to a human skin graft on nude mice. One dose ($100 \mu\text{l}$) of a liposomal (20 mg/ml lipid) or aqueous (phosphate buffered saline) formulation was applied daily for three days. Active formulation contained 0.51 mg/ml gamma interferon. Placebo liposomal formulations contained no added protein. Significant differences (active versus placebo) are indicated by asterisks (*). (Redrawn from data in Ref. 31.)

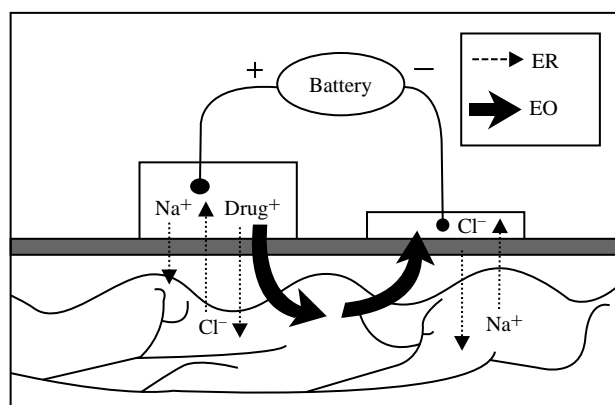


Fig. 3 Schematic representation of iontophoresis. Two electrode chambers, connected to a power source, are placed in contact with the skin. Upon application of the electric field, drug ions are repelled from the electrode of similar polarity (in this case, cations are repelled from the anode). This electrorepulsion (ER) also imposes “inward” motion on i) other cations present in the anode formulation, and ii) the “outward” transport of anions (e.g., Cl^-) from within the skin. At the ‘non-working’ electrode (in this case, the cathode), negative anions from the electrolyte are driven into and through the skin, while cations (e.g., Na^+) are “extracted” from the tissue. The direction of the electroosmotic flow (EO) is also shown (see text for details).

cations, therefore, there will be a transition in the dominant mechanism from electrorepulsion to electroosmosis (50) (however, it should be said that where this transition occurs has not been fully defined); for anions, on the other hand, as molecular size increases, it is clear that, at the transition, the electrorepulsive contribution will be cancelled by electroosmotic convective flow going in the opposite direction (a fact that explains, at least in part, why insulin has proved a futile candidate for iontophoresis, see later). It should also be noted that there are examples of lipophilic cations (including peptides such as the LHRH analogs, nafarelin, and leuprolide) that can apparently associate strongly with the net negatively charged lipophilic skin, thereby neutralizing the anionic membrane, reducing its permselectivity and “turning off” electroosmotic flow, that is, the principal mechanism for the transport of such species (52–54, 58, 59).

The principal parameters controlling the iontophoresis of peptides/proteins can be summarized as follows:

- a. *Charge and pH:* As discussed above, in general, cationic species are delivered more efficiently than anions because they can take advantage of both electrorepulsive and electroosmotic mechanisms. Neutral species can also be enhanced if formulated appropriately at the anode. Increasing charge does not necessarily lead to

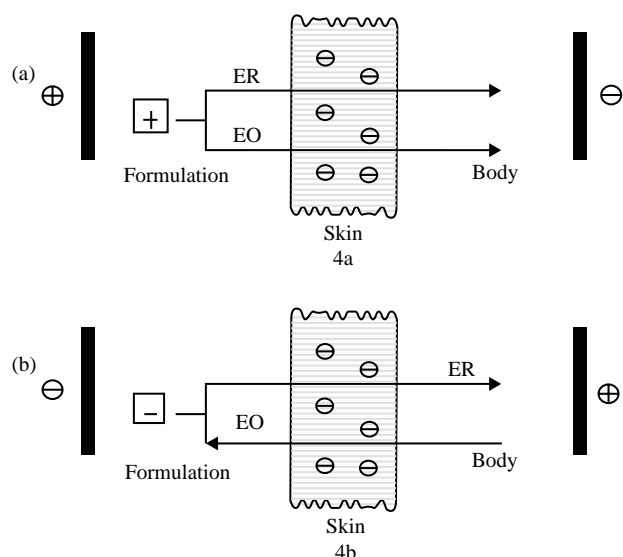


Fig. 4 Imposing an electrical potential gradient across a charged membrane produces a convective solvent flow in the direction of counter-ion transport (i.e., from anode-to-cathode in the case of skin). This electroosmotic effect (EO) adds to electrorepulsion (ER) to enhance the transport of cationic compounds during iontophoresis (4a) while acting against the electromigration of anions (4b).

better delivery (all other parameters remaining equal) primarily due, it is believed, to increased association with the membrane and, once again, reduced electroosmosis (48). The pH can be adjusted to optimize peptide stability, and sometimes to elicit better transport (because, e.g., the peptide shifts from being zwitterionic to cationic). However, pH cannot be adjusted too far i) because of concerns about irritation; ii) as this alters the inherent charge on the skin and, in consequence, its permselectivity properties.

- b. *Molecular size:* Simplistically, small charged molecules are better delivered than larger ones for reasons already articulated (Fig. 5). The absolute relationship is probably complex and may be difficult to deduce. For molecules delivered only by electroosmosis, transport should be somewhat independent of size (assuming that the dimensions of the drug never approach that of the pathway). Few studies have addressed this issue with respect to peptides/proteins specifically, although one investigation with poly-L-lysines of increasing molecular weight clearly indicated that a 4-kDa oligomer was more easily transported than the higher molecular weight (7- and 26-kDa) species (59).
- c. *Formulation:* The basic principles that apply generally to iontophoresis are especially relevant to peptides/

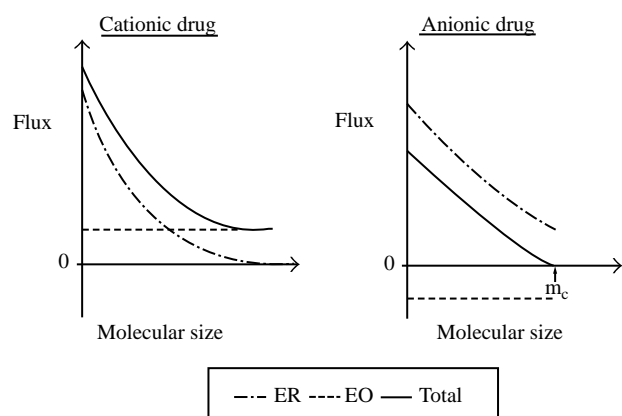


Fig. 5 Effect of molecular size on the relative importance of electrorepulsion (ER) and electroosmosis (EO) to the overall iontophoretic transport of cations and anions. Small, highly mobile cations are principally moved across the skin by ER. But, as molecular size increases, the fraction of charge carried by a cationic drug decreases, and the principal mechanism of transport becomes EO. For anions, on the other hand, EO is a negative contribution to the total flux and, once the molecular size reaches a critical value (m_c , completely cancels out the ER contribution to electrotransport (resulting in no net flux).

proteins. If the drug carries a measurable fraction of the charge being passed, then it makes sense to minimize the presence of competing ions in the formulation. Generally speaking, lower electrolyte levels also mean that electroosmosis is slightly higher too (49, 60). Moreover, stability issues may demand at least some level of background electrolyte and/or buffer; thus, polymeric buffers, for example, which are not necessarily competitive for charge-carrying, have been used to improve peptide delivery (61). Peptide stability during storage of an iontophoresis device has also been addressed (44, 62). Because iontophoretic formulations employ aqueous-based gels, there are, in particular, potential problems of hydrolysis. This issue has been attacked recently in the case of calcitonin by making a dry reservoir disc for iontophoresis (via compression of a mixture of freeze-dried peptide and gelatin), which is hydrated at the moment of use (62). Proof-of-concept of the idea has been achieved via the observation of a hypocalcemic response in rabbits (62). The other crucial component of the formulation, of course, is the electrode. In general, reversible, and especially Ag/AgCl, electrodes are preferred (63). These very stable and reproducible electrodes are well suited to iontophoresis and avoid the problems inherent in bare metal electrodes such as those made of platinum (in particular, the hydrolysis of water that leads to very

large changes in pH in the electrode formulations and/or demands the presence of a strong buffer) (64–67). A potential problem with the use of Ag/AgCl electrodes for peptide/protein delivery is that these drugs may be inactivated/degraded at the silver chloride surface (52, 68). It may be necessary, therefore, to devise a means to keep the drug away from the electrode (e.g., use of a semipermeable membrane or a salt bridge) (52, 68–70).

Now, we can summarize the principal teachings of the field with respect to peptide and protein delivery by iontophoresis. First, we consider the major results reported in vivo in different animal models. The tripeptide Thr-Lys-Pro has been iontophoresed into hairless rats (71). The apparent urinary excretion rate of the peptide following in vivo delivery was shown to be completely consistent with the measured flux in vitro. Pretreatment of rats with an iontophoretic current followed by passive application of the peptide, also resulted in enhanced delivery. No histological changes were detected following current passage in vivo, only slight reddening at the skin area under the electrodes. Notable delivery rates of octreotide (a somatostatin analog) were obtained following application of mild current densities ($50\text{--}150\text{ }\mu\text{A}/\text{cm}^2$) in vivo in rabbits (72). Increasing the intensity of the applied current elicited a proportional increase in peptide plasma levels, and drug input declined quickly upon current termination. On the other hand, another study, in which human calcitonin was delivered into hairless rats, showed that the lowering of serum calcium was not linearly dependent upon either current density or time of current application (73). Iontophoresis of LHRH has been investigated in vivo in pigs (74). Elevated LHRH concentrations were measured in the blood and concomitant increases in LH (luteinizing hormone) and FSH (follicle stimulating hormone) levels were observed, demonstrating that the hormone was delivered as the pharmacologically active species. Circulating levels of LHRH fell rapidly upon termination of iontophoresis. A larger peptide, growth hormone releasing factor, GRF (1–44) (MW 5040), was delivered by iontophoresis into hairless guinea pigs, resulting in steady-state plasma levels of $\sim 0.2\text{ ng/ml}$, which, in terms of flux, signifies that an input rate of $\sim 3.16\text{ }\mu\text{g/h}$ was achieved (75).

Secondly, we survey briefly the published findings on peptide/protein delivery by iontophoresis in man. Delivery of calcitonin gene-related peptide (CGRP) and vasoactive intestinal polypeptide (VIP) has proven useful in the clinical treatment of venous stasis ulcers (76). In 66 patients, 40 cm^2 iontophoretic patches were applied to intact skin in the proximity of the ulcer. Pulsed electric

current was delivered for 20 min. It was found that CGRP and VIP delivery was enhanced, as deduced from the clinical results, and that the electric current passing close to the ulcer area had a positive influence on the healing process. The iontophoretic delivery of leuprolide, a LHRH analog, has also been investigated in human subjects (77, 78). It was shown, following iontophoresis, that the observed increases in LH levels were comparable with those obtained after subcutaneous injection (77). The result was particularly remarkable considering the low current density delivered: that is, 0.2 mA over 70 cm² ($\sim 3.1 \mu\text{A}/\text{cm}^2$). Only 2/13 volunteers reported a tingling sensation during current passage while some erythema (which, nevertheless, resolved quickly after current termination) was observed at the electrode sites in 6/13 subjects. In a further report in human volunteers, the effect of formulation variables on the iontophoretic delivery of leuprolide (78) was examined. Highest transport was observed with the lowest leuprolide concentration investigated (Fig. 6). Changes in the ionic strength of the donor formulation had a greater effect on circulating leuprolide than on LH and testosterone levels, an observation that can be partially explained in terms of interindividual differences in the pharmacological response.

Lastly, we address insulin delivery to which much effort has been dedicated. While the insulin “spikes” required postprandially may be difficult to achieve with iontophoresis, the method may have the potential to mimic the physiological, nearly constant, basal secretion (1 IU/h) of the hormone, which is observed in the nondiabetic adult (i.e., a requirement that is not provided by intermittent subcutaneous injections) (79). A recent, excellent review (80) of the work performed (almost exclusively in animal models) concludes that while insulin iontophoresis can be sufficient to treat a small diabetic animal, the best deliveries achieved are still 1–2 orders of magnitude below that necessary to meet the basal secretion level in humans. It seems unlikely, therefore, that we will see an insulin iontophoretic delivery system on the market in the foreseeable future.

Electroporation

The application of high-voltage electrical pulses to the skin (so-called electroporation) is another approach that has also been used to increase peptide delivery across the epidermal barrier. Several reviews of the application of electroporation to increase transdermal delivery have been published within the last few years (81–85). Unlike iontophoresis, which employs small currents (0.5 mA/cm²) for relatively long periods of time (many minutes to hours), electroporation involves exposure of the skin to

relatively high voltages (on the order of 30–100 V imposed across the skin) for rather short times, typically one to several hundred milliseconds (85–87).

Mechanistically, as well, electroporation almost certainly differs from iontophoresis and an important effect involves the creation of new, low-resistance pathways through the stratum corneum. However, a detailed discussion of the physical impact of electroporation on skin barrier function is beyond the scope of this chapter, and the reader is referred to the literature for specific details (84, 86, 88).

In terms of peptide/protein delivery using electroporation, the most systematic work has been performed with the decapeptide LHRH (89, 90), and the ability of electroporation to induce its rapid delivery relative to “conventional” iontophoresis, in particular, has been clearly demonstrated (89). Specifically, using human skin *in vitro*, it was shown that a single electroporative pulse applied for 5 ms, followed by 30 min of iontophoresis at “normal” current densities, induced significantly higher delivery of LHRH than iontophoresis alone (89) (Fig. 7). An obvious allusion to an intravenous bolus followed by an infusion can be drawn from this work. Subsequently, in a sophisticated *ex vivo* model (the isolated perfused porcine skin flap), the pulsatile input of LHRH by electroporation was again examined, with the

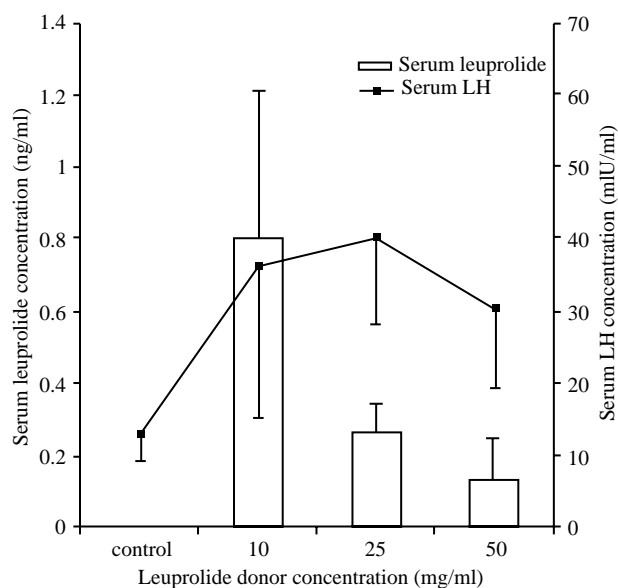


Fig. 6 Serum leuprolide concentration and resulting circulating levels of luteinizing hormone (LH) in human volunteers following iontophoresis, as a function of drug donor concentration. Constant current (0.2 mA) was applied for 10–12 h. (Redrawn from data in Ref. 78.)

reproducibility of input following a second pulse of particular interest (90).

Another recent pair of investigations have examined the application of electroporation to enhance the delivery of cyclosporin for the treatment of psoriasis (91, 92). Compared to passive transport, single electroporative pulses resulted in up to a 60-fold enhancement in hairless rat skin permeability depending upon the vehicle used.

It must be said, however, that practical questions pertaining to electroporation are important barriers to its further development. Use of the high voltages required raises significant questions about safety, and how exactly electroporation would be practised in reality—only in hospital, or a doctor's office? What are the effects of electroporation on the skin? Acutely, within certain limits, the results are cautiously optimistic but, chronically, no one knows. And, from a marketing standpoint, electroporation must be grouped with other so-called “minimally invasive” technologies, which effectively remove the SC from the delivery equation. Under these circumstances, will electroporation compete with microneedles or “painless” injections? In short, electroporation has a long way to go before it can ever be perceived as a serious challenge to a simple injection.

Sonophoresis

The use of ultrasound (US) to enhance percutaneous absorption (so-called sonophoresis or phonophoresis) has been studied over many years, and is the basis of US propagation and US effects on tissue, and the use of US in transdermal delivery have been reviewed in detail (93–96). The proposed mechanisms by which US enhances skin penetration include cavitation, thermal effects and mechanical perturbation of the SC; that is, US acts on the barrier function of the membrane (96).

Sonophoresis has employed three distinct categories of US: “high-frequency” or diagnostic US (2–10 MHz), “mid-frequency” or therapeutic US (0.7–3 MHz), and “low-frequency” US (5–100 kHz). It appears, from a general overview of the literature, that the efficiency of US-mediated drug delivery depends on several factors, including US frequency, intensity (i.e., power per unit area), continuous versus pulsed mode, duty cycle, duration, coupling medium, and so on. The fact that very few studies have used common values for some or any of these parameters almost certainly accounts for the different and sometimes contradictory results in the public domain.

It should be noted that the US beam is made up of two components: the field closest to the transducer, and the field further away (the final, diverging conical part) (93). The relative size of these two zones, and their separation,

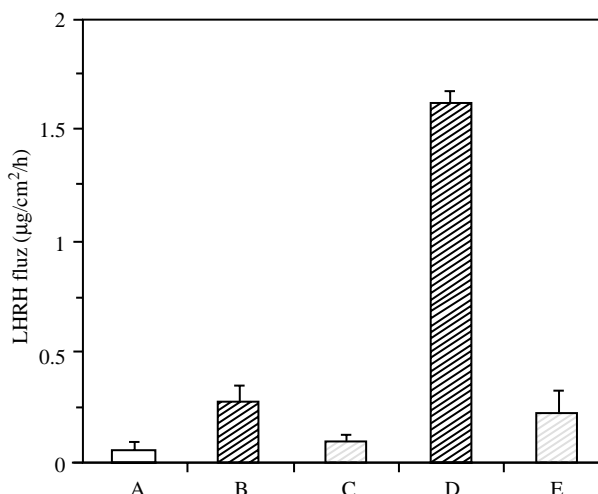


Fig. 7 Transport of LHRH (mean \pm SD) across human epidermis in vitro. A) passive flux, B) transport after 30 min of iontophoresis at 0.5 mA/cm²; D) transport after a single electroporative pulse (1000 V, 5 msec) followed by 30 min of iontophoresis at 0.5 mA/cm². The passive fluxes 2 h after termination of treatments B and D (columns C and E, respectively) are also shown. (Redrawn from data in Ref. 89.)

is a function of a) the US wavelength (i.e., frequency) and b) the transducer radius. The distribution of energy across these two regions is not uniform; indeed, near the transducer, it becomes quite complex, emphasizing again how small changes in experimental protocol between studies can lead to significantly different results.

From a practical standpoint, unlike the situation with iontophoresis, the transition from laboratory-size US equipment to a small, user friendly, and compatible device remains a challenging “scale-down” problem. Furthermore, as mentioned above, it appears that US acts on the skin barrier per se (while iontophoresis seems to exert its effects more specifically on the substance targeted for delivery) implying, therefore, that control of the enhancement technology continues to be an issue for the long-term, safe, and effective application of this method.

With respect to US-enhanced peptide and protein delivery, the most impressive results have been seen in the “low-frequency” domain. Tachibana et al. were the first to report the increased delivery of insulin across the skin of hairless mice (97) and rabbits (98) in vivo. Using US at 105 kHz (90 min, “pulsed”: 5 s on, 5 s off), significant insulin delivery and concomitant lowering of blood sugar were seen in diabetic rabbits (98) (Fig. 8). At a slightly lower frequency (48 kHz), a similar glucose-lowering effect was observed in the hairless mouse (97). Subsequently, Mitragotri et al. (99) described similar results in

hairless rats; in this case, the US frequency was 20 kHz, with a duration of application of up to 4 h, using 100 ms pulses at US intensities between 12.5 and 225 mW/cm² [lately, this intensity range has been corrected to 1.6–14 W/cm² (100)]. The same authors also reported that this low frequency US approach could be used to significantly enhance the delivery of interferon- γ (MW ~17 kDa) and erythropoietin (MW ~48 kDa) in vitro, that is, that macromolecules of much greater size could now be deliverable across the skin. It should be said, however, that no further data on these proteins has been reported since the initial publication in 1995 (99).

Additional work, however, has addressed mechanistic aspects of the effects of low-frequency US. Cavitation and thermal effects have been postulated and, to a certain extent, characterized (101–104), but further work is clearly needed to define exactly how US interacts with the skin barrier to increase its permeability.

Only one publication has presented data to support the idea that higher frequency (1 MHz) US can be used to increase peptide delivery across the skin (105). The enhanced transport of poly-L-lysine (MW = 4-kDa) across human epidermis has been reported, a finding completely at odds with the relatively modest action of US at this frequency on the percutaneous penetration of several “small” molecular weight substances (96).

Finally, in a related approach, Lee et al. have used photomechanical waves to enhance the skin delivery of 5-aminolevulinic acid (ALA), a δ -amino acid used in photodynamic therapy (106). The broadband, compressive waves appear to interact with tissues uniquely via mechanical forces. ALA delivery was shown to be proportional to the peak stress (388–503 bar) of the photomechanical wave applied.

In conclusion, despite the exciting low-frequency US results obtained, there remain important questions about mechanism, local skin effects and tolerability, reversibility, and ultimate practicability. These issues must be addressed before sonophoresis can move from research to development mode.

Gene and DNA Delivery to the Skin

The goal of gene therapy is to modify in some way the synthetic capability of a target cell, or population of cells, to provoke in situ the desired “therapeutic response” (107). While conceptually attractive, gene therapy is presently limited by problems associated with delivery; that is, how does one get the gene to the place where it can elicit its effect, and then how does one appropriately regulate the expression of the gene introduced?

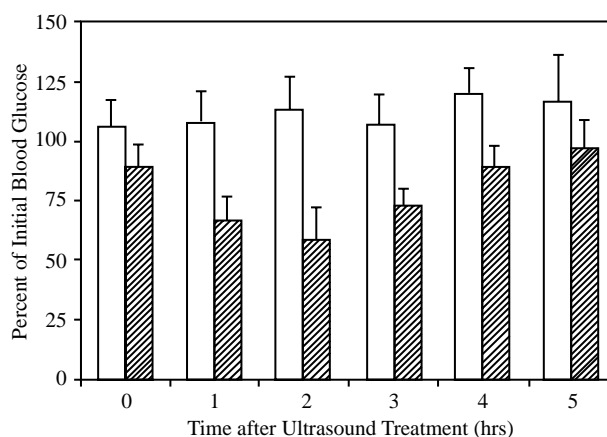


Fig. 8 Change in blood glucose level following insulin delivery by ultrasound exposure to alloxan-diabetic rabbits. Filled bars (mean \pm SD) show the relative blood glucose level following ultrasound treatment for 1.5 h. The area of skin exposed was 7 cm²; the ultrasound frequency used was 105 kHz. Open bars indicate the corresponding control values. (Redrawn from Ref. 98.)

Gene delivery to, or via, the skin has three principal objectives (108): a) the treatment of primary cutaneous diseases and wounds, b) inducing skin cells to encode a protein which is subsequently released into the systemic circulation in a controlled fashion, and c) immunization. The first two instances involve a chronic delivery challenge, whereas immunization requires only acute exposure to the DNA involved.

Chronic delivery can be achieved, at least theoretically, either ex vivo or in vivo. In the former approach, cells are harvested from the skin by biopsy and cultured in vitro. Gene transfer is then effected in this in vitro setting (e.g., by electroporation) and the transfected cells are finally grafted back into the host (108). Clearly, this is a labor-intensive, complex, and expensive procedure. In the latter case, genetic material has been administered directly to intact skin, using either viral or nonviral vectors (e.g., adenovirus or modified herpes virus and DNA complexed with liposomes, respectively). In addition, a variety of physical methods have been investigated for DNA transfer into skin including direct injection (of both the naked DNA and plasmid associated with different vectors), particle bombardment (the “gene gun”), ultrasound, electroporation, and other so-called minimally invasive techniques (108–110). However, direct gene transfer in vivo, while attractive conceptually and technically easier to perform, has proven to-date to be rather inefficient and to result in only low and transient levels of gene expression (108).

DNA immunization consists of encoding an antigen in a plasmid, which is then transfected into cells in vivo. This results in expression of the antigen and the subsequent

induction of an immune response to the expressed protein (111). As the protein antigen is produced “naturally” in vivo, both cellular and humoral responses are developed. Skin has been proposed as an attractive site of immunization, and several skin cell types can become involved in DNA uptake, antigen expression and presentation. Quite recently, for example, in a mouse model, immunization has been reported following topical application of an aqueous solution of naked DNA (112). Given what one knows about the skin’s passive permeability to macromolecules, it is clear that only a very small quantity of DNA vaccine is required for immunization.

In summary, while the field of gene therapy in general is attracting widespread interest at this time, real successes at the level of the skin (beyond those positive results obtained in various model systems) are few and far between. It is clear, then, that considerably more work needs to be done. Detailed, and current, overviews of the state-of-the-art in cutaneous gene delivery can be found in Refs. (108) and (111).

Minimally Invasive Delivery Technologies

Finally, we consider the delivery of peptides and proteins across the skin by methods that physically circumvent the stratum corneum barrier in ways designed to be less invasive than the use of a classic needle and syringe. In these approaches, therefore, the excellent barrier function of the SC is acknowledged, and one turns one’s attention to bypassing this membrane in as least offensive a fashion as possible. This strategy for peptide delivery into the systemic circulation has been known for some time, however. The Matse Indians from Peru use one variation of this idea in shamanic hunting practices (113). Having obtained an extract from the skin of the frog *Phyllomedusa bicolor* (which contains ~7 wt% active peptides), a formulation is prepared and then applied to a small area of skin the epidermal barrier of which is first burned away with a smouldering twig. Within minutes of application, the treated individual experiences violent peripheral, gastrointestinal, and cardiovascular effects, followed by further remarkable central responses (increase in physical strength, resistance to hunger, thirst, stress, and so on).

Back in the world of intellectual property and patent protection, a number of sophisticated technologies have evolved in recent years and have often been applied to peptide and protein delivery objectives. Among the approaches used, one can cite ablation of the SC by laser radiation, by heat, and by erosion, as well as the use of particle bombardment and microneedles of various configuration. Some approaches are even suggested for

use in combination with other transdermal technologies, such as iontophoresis. As yet, none of these methods are approved clinically for peptide/protein delivery and all have yet to show real superiority over the gold standard injection via a conventional needle and syringe. An excellent review of the field has recently been written by Down et al. (114).

Mid-infrared laser ablation of the SC has been used to allow the transdermal delivery of interferon- γ (115). The mechanism by which this laser beam “explodes” away the SC is complex, yet quite reproducible and controllable, in that the number of laser pulses is correlated to the degree of SC damage (and hence to the peptide’s permeability through the tissue). Tissue destruction below the SC appeared to be minimal.

The MicroPorsTM technology (114, 116) consists of directing tightly focused thermal energy into the SC to create micropores. The skin is contacted by a wire mesh through which a current is passed that causes local heating sufficient to burn small holes in the barrier. Delivery of insulin by this approach has been suggested (116). A more macroscopic method to create an erosion in the SC is via “suction de-epithelialization” (117). Using a vacuum, a small blister (6 mm diameter) is formed on the skin; the tissue separates at the dermal–epidermal junction. The roof of the blister is then removed (“guillotined,” as it were) exposing a small area of dermis to which a drug solution can be directly applied. Feasibility, in vivo in man, was demonstrated using the antidiuretic peptide 1-deamino-8-D-arginine vasopressin (dDAVP) and stable, therapeutically active plasma levels were achieved rather quickly (117). The successful delivery of an oxytocin analog has also been shown (118). The authors claim that the dermal microcirculation remains intact and functional following creation of the erosion. Mild inflammation is observed and the erosion self-heals over time with, apparently, minimal scarring. However, practical applications of this approach, especially for chronic disease treatment, appears unlikely, and issues related to local infections will have to be addressed.

High-velocity particle delivery across the skin is the technology of a major drug delivery operation based in the UK and USA (PowderJect, Inc.) (119). Once a drug has been formulated as an appropriate and well-characterized powder, it is then introduced into a compact hand-held device in which a supersonic flow of gas accelerates the particles to a speed high enough that they can collide with the skin having enough energy to penetrate the outer layers and effect drug delivery. The depth and extent of delivery depends on the speed, diameter, and density of the drug particles. Again, insulin has been a peptide of choice for study by this method (120).

Finally, with the explosion of interest in micromachining and microfabrication, it is not surprising that drug delivery aficionados have turned their attention to the production of arrays of micron-sized microneedles capable of creating transport pathways across the SC without eliciting disagreeable perception on the part of the patient (121, 122). Various configurations and forms of these devices have been evolved, the majority of which involve extremely sharp tips and needle lengths on the order of 100 μm . At the time of writing, no specific publications detailing peptide/protein delivery using this technology have appeared.

Overall, then, the minimally invasive field is an active and imaginative one, of much perceived promise. However, it remains to be seen, as this technology matures and as it moves closer to real clinical testing against known effective methodologies, whether the current level of excitement (and investment) will be rewarded by the realization of successful and marketable products.

CONCLUSIONS

The relatively noninvasive nature of transdermal drug delivery, and the fact that this route can simultaneously avoid problems associated with presystemic metabolism and mimic (at least, to some extent) parenteral input profiles, are significant advantages. There have been, therefore, diverse attempts to exploit the skin for peptide and protein delivery. As we have noted before, transdermal administration, with or without one or more enhancement technologies, will always be limited to potent drugs and this accounts, once more, for the effort devoted to peptide and protein (i.e., typically very active substances) administration via this route.

Even the simplest structure-permeation relationships for percutaneous transport (no matter their limitations) indicate that the bigger a molecule, the poorer its diffusivity across the skin. Thus, larger molecules cannot be considered as transdermal candidates unless they are extremely potent. Vaccines are one such example and, given the skin's role as an immunological organ, represent a potentially very interesting opportunity for the transdermal route. For other peptides and proteins, at least to-date, an enhancement strategy appears to be obligatory in order that the skin's barrier function can be overcome.

The different approaches used have all yielded *some* effect, in that increased peptide/protein delivery has been

achieved. However, with the possible exception of iontophoresis, few of the various technologies employed have achieved an advanced state of maturity. Even in the case of iontophoresis, despite considerable activity and many publications in the literature, there are no peptide delivery systems on the immediate horizon (although there are products with "normal" drugs in the later steps of development). Typically, in the biotech era of the last 20 years, the newer technologies, including the "minimally invasive" methods (i.e., those about which we know the least and whose associated problems are poorly, or have yet to be, characterized!) are now attracting most attention and the highest level of investment.

It remains to be seen, therefore, whether the transdermal route will eventually yield a marketable delivery system for one or more peptides and/or proteins. For some of the enhancement technologies available, considerably more work needs to be done to establish that there is a distinct advantage over a needle and syringe (e.g., ultrasound, electroporation, and all "minimally invasive" methods); for iontophoresis, much of the groundwork has been done and the basic, mechanistic principles are becoming clearer—the challenge now is to identify the right peptide candidates and to initiate a serious development effort. Vaccine delivery appears to be an intriguing opportunity and of much interest; time will tell if the transdermal route will prove compatible with this important goal.

REFERENCES

1. Audus, K.L.; Raub, T.J. Biological Barriers to Protein Delivery. *Pharmaceutical Biotechnology*; Plenum Press: New York, 1993; 4.
2. Wearley, L.L.; Banga, A.K. Peptide and Protein Drug Delivery. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1995; 11, 395–411.
3. Hadgraft, J.; Guy, R.H. *Transdermal Drug Delivery—Development Issues and Research Initiatives*; Marcel Dekker, Inc.: New York, 1989; 35.
4. Potts, R.O.; Francoeur, M.L. The Influence of Stratum Corneum Morphology on Water Permeability. *J. Invest. Dermatol.* **1991**, 96, 495–499.
5. Wertz, P.W. The Nature of the Epidermal Barrier: Biochemical Aspects. *Adv. Drug Deliv. Rev.* **1996**, 18, 283–294.
6. Turner, N.G.; Nonato, L.B. Visualization of Stratum Corneum and Transdermal Permeation Pathways. *Mechanisms of Transdermal Drug Delivery*; Potts, R.O., Guy, R.H., Eds.; Marcel Dekker, Inc.: New York, 1997; 83, 1–40.

7. Potts, R.O.; Francoeur, M.L. Lipid Biophysics of Water Loss Through the Skin. *Proc. Natl. Acad. Sci.* **1990**, *87*, 3871–3873.
8. Boddé, H.E.; Van den Brink, I.; Koerten, H.K.; Haan, F.H.N. Visualization of In Vitro Percutaneous Penetration of Mercuric Chloride; Transport Through Intercellular Space Versus Uptake Through Desmosomes. *J. Controlled Release* **1991**, *15*, 227–236.
9. Kalia, Y.N.; Pirot, F.; Guy, R.H. Homogeneous Transport in a Heterogeneous Membrane: Water Diffusion Across Human Stratum Corneum in Vivo. *Biophys. J.* **1996**, *71*, 2692–2700.
10. Grimnes, S. Pathways of Ionic Flow Through Human Skin in Vivo. *Acta Derm. Venereol. (Stockh.)* **1984**, *64*, 93–98.
11. Keister, J.C.; Kasting, G.B. The Use of Transient Diffusion to Investigate Transport Pathways Through Skin. *J. Controlled Release* **1986**, *4*, 111–117.
12. Albery, W.J.; Hadgraft, J. Percutaneous Absorption: In Vivo Experiments. *J. Pharm. Pharmacol.* **1979**, *31*, 140–147.
13. El Tayar, N.; Tsai, R.S.; Testa, B.; Carrupt, P.A.; Hansch, C.; Leo, A. Percutaneous Penetration of Drugs: A Quantitative Structure–Permeability Relationship Study. *J. Pharm. Sci.* **1991**, *80*, 744–749.
14. Potts, R.O.; Guy, R.H. Predicting Skin Permeability. *Pharm. Res.* **1992**, *9*, 663–669.
15. Flynn, G.L. Physicochemical Determinants of Skin Absorption. *Principles of Route-to-Route Extrapolation for Risk Assessment*; Gerrity, T.R., Henry, C.J. Eds.; Elsevier: New York, 1990; 93–127.
16. Cleek, R.L.; Bunge, A.L. A New Method for Estimating Dermal Absorption from Chemical Exposure. I. General Approach. *Pharm. Res.* **1993**, *10*, 497–506.
17. Weber, C.J.; Jicha, D.; Matz, S.; Siverly, J.; O'Dorisio, T.; Strausberg, L.; Laurencot, J.; McLarty, A.; Norton, J.; Kazim, M.; Reemtsma, K. Passage of Somatostatin Analogue Across Human and Mouse Skin. *Surgery* **1987**, *102*, 974–981.
18. Hadley, M.E.; Wood, S.H.; Lemus-Wilson, A.M.; Dawson, B.V.; Levine, N.; Dorr, R.T.; Hruby, V.J. Topical Application of a Melanotropic Peptide Induces Systemic Follicular Melanogenesis. *Life Sciences* **1987**, *40*, 1889–1895.
19. Dawson, B.V.; Hadley, M.E.; Kreutzfeld, K.; Dorr, R.T.; Hruby, V.J.; Al-Obeidi, F.; Don, S. Transdermal Delivery of a Melanotropic Peptide Hormone Analogue. *Life Sciences* **1988**, *43*, 1111–1117.
20. Dawson, B.V.; Hadley, M.E.; Levine, N.; Kreutzfeld, K.L.; Don, S.; Eytan, T.; Hruby, V.J. In Vitro Transdermal Delivery of a Melanotropic Peptide Through Human Skin. *J. Invest. Dermatol.* **1990**, *94*, 432–435.
21. Boddé, H.E.; Verhoef, J.C.; Poncet, M. Transdermal Peptide Delivery. *Biochem. Soc. Trans.* **1989**, *17*, 943–945.
22. Choi, H.K.; Flynn, G.L.; Amidon, G.L. Transdermal Delivery of Bioactive Peptides: The Effect of *n*-Decylmethyl Sulfoxide, pH, and Inhibitors on Enkephalin Metabolism and Transport. *Pharm. Res.* **1990**, *7*, 1099–1106.
23. Banarjee, P.S.; Ritschel, W.A. Transdermal Permeation of Vasopressin. I. Influence of pH, Concentration, Shaving and Surfactant on In Vitro Permeation. *Int. J. Pharm.* **1989**, *49*, 189–197.
24. Banarjee, P.S.; Ritschel, W.A. Transdermal Permeation of Vasopressin. II. Influence of Azone on In Vitro and In Vivo Permeation. *Int. J. Pharm.* **1989**, *49*, 199–204.
25. Ogiso, T.; Iwaki, M.; Yoneda, I.; Horinouchi, M.; Yamashita, K. Percutaneous Absorption of Elcatonin and Hypocalcemic Effect in Rat. *Chem. Pharm. Bull.* **1991**, *39*, 449–453.
26. Ogiso, T.; Iwaki, M.; Tanino, T.; Paku, T. Effectiveness of the Elcatonin Transdermal System for the Treatment of Osteoporosis and the Effect of the Combination of Elcatonin and Active Vitamin d₃ In Rat. *Biol. Pharm. Bull.* **1993**, *16*, 895–898.
27. Lu, M.F.; Lee, D.; Rao, G.S. Percutaneous Absorption Enhancement of Leuprolide. *Pharm. Res.* **1992**, *9*, 1575–1579.
28. Roy, S.D.; de Groot, J.S. Percutaneous Absorption of Nafarelin Acetate, an LHRH Analog, Through Human Cadaver Skin and Monkey Skin. *Int. J. Pharm.* **1994**, *110*, 137–145.
29. Setoh, K.; Murakami, M.; Araki, N.; Fujita, R.; Yamamoto, A.; Muranishi, S. Improvement of Transdermal Delivery of Tetragastrin by Lipophilic Modification with Fatty Acids. *J. Pharm. Pharmacol.* **1995**, *47*, 808–811.
30. Egbaria, K.; Ramachandran, C.; Weiner, N. Topical Delivery of Ciclosporin: Evaluation of Various Formulations Using In Vitro Diffusion Studies in Hairless Mouse Skin. *Skin Pharmacol.* **1990**, *3*, 21–28.
31. Short, S.M.; Paasch, B.D.; Turner, J.H.; Weiner, N.; Daugherty, A.L.; Mrsny, R.J. Percutaneous Absorption of Biologically-Active Interferon-Gamma in a Human Skin Graft-Nude Mouse Model. *Pharm. Res.* **1996**, *13*, 1020–1027.
32. Short, M.S.; Rubas, W.; Paasch, B.D.; Mrsny, R.J. Transport of Biologically Active Interferon-Gamma Across Human Skin In Vitro. *Pharm. Res.* **1995**, *12*, 1140–1145.
33. Vutla, N.B.; Betageri, G.V.; Banga, A.K. Transdermal Iontophoretic Delivery of Enkephalin Formulated in Liposomes. *J. Pharm. Sci.* **1996**, *85*, 5–8.
34. Février, F.; Bobin, M.F.; Lafforgue, C.; Martini, M.C. Advances in Microemulsions and Transepidermal Penetration of Tyrosine. *S.T.P. Pharma. Sci.* **1991**, *1*, 60–63.
35. Rolland, A.; *Pharmaceutical Particulate Carriers: Therapeutic Applications*; Rolland, A., Ed.; Marcel Dekker, Inc.: New York, 1993; 367–421.
36. Schreier, H.; Bouwstra, J. Liposomes and Niosomes as Topical Drug Carriers, Dermal and Transdermal Drug Delivery. *J. Controlled Release* **1994**, *30*, 1–15.
37. Cevc, G.; Blume, G.; Schätzlein, A.; Gebauer, D.; Paul, A. The Skin: A Pathway for Systemic Treatment with Patches and Lipid-Based Agent Carriers. *Adv. Drug Deliv. Rev.* **1996**, *18*, 349–378.
38. Weiner, N.; Williams, N.; Birch, G.; Ramachandran, C.; Shipman, C., Jr.; Flynn, G. Topical Delivery of Liposomally Encapsulated Interferon Evaluated in a Cutaneous Herpes Guinea Pig Model. *Antimicrob. Agents Chemother.* **1989**, *33*, 1217–1221.

39. Cevc, G.; Gebauer, D.; Stieber, J.; Schätzlein, A.; Blume, G. Ultraflexible Vesicles, Transfersomes, Have an Extremely Low Pore Penetration Resistance and Transport Therapeutic Amounts of Insulin Across the Intact Mammalian Skin. *Biochim. Biophys. Acta* **1998**, *1368*, 201–215.
40. Glenn, G.M.; Rao, M.; Matyas, G.R.; Alving, C.R. Skin Immunization Made Possible by Cholera Toxin. *Nature* **1998**, *391*, 851.
41. Scharton-Kersten, T.; Yu, J.M.; Vassell, R.; O'Hagan, D.; Alving, C.R.; Glenn, G.M. Transcutaneous Immunization with Bacterial ADP-Ribosylating Exotoxins, Subunits, and Unrelated Adjuvants. *Infect. Immun.* **2000**, *68*, 5306–5313.
42. Paul, A.; Cevc, G.; Bachhawat, B.K. Transdermal Immunization with Large Proteins by Means of Ultradeformable Drug Carriers. *Eur. J. Immunol.* **1995**, *25*, 3521–3524.
43. Paul, A.; Cevc, G.; Bachhawat, B.K. Transdermal Immunisation with an Integral Membrane Component, Gap Junction Protein, by Means of Ultradeformable Drug Carriers, Transfersomes. *Vaccine* **1998**, *16*, 188–195.
44. Green, P.G. Iontophoretic Delivery of Peptide Drugs. *J. Controlled Release* **1996**, *41*, 33–48.
45. Delgado-Charro, M.B.; Guy, R.H. Iontophoresis of Peptides. *Electronically Controlled Drug Delivery*; Berner, B., Dinh, S.M., Eds.; CRC Press: Boca Raton, 1998; 129–157.
46. Kim, A.; Green, P.G.; Rao, G.; Guy, R.H. Convective Solvent Flow Across the Skin During Iontophoresis. *Pharm. Res.* **1993**, *10*, 1315–1320.
47. Marro, D.; Guy, R.H.; Delgado-Charro, M.B. Characterization of the Iontophoretic Perselectivity Properties of Human and Pig Skin. *J. Controlled Release* **2000**, *70*, 213–217.
48. Burnette, R.R.; Ongpipattanakul, B. Characterization of the Permselective Properties of Excised Human Skin During Iontophoresis. *J. Pharm. Sci.* **1987**, *76*, 765–773.
49. Pikal, M.J. The Role of Electroosmotic Flow in Transdermal Iontophoresis. *Adv. Drug Deliv. Rev.* **1992**, *9*, 201–237.
50. Guy, R.H.; Kalia, Y.N.; Delgado-Charro, M.B.; Merino, V.; López, A.; Marro, D. Iontophoresis: Electorepulsion and Electroosmosis. *J. Controlled Release* **2000**, *64*, 129–132.
51. Delgado-Charro, M.B.; Kalia, Y.N.; Guy, R.H. The Relative Contributions of Electorepulsion and Electroosmosis to Iontophoretic Transport. *Proc. Int. Symp. Controlled Rel. Bioact. Mater.* **1998**, *25*, 599–600.
52. Delgado-Charro, M.B.; Guy, R.H. Characterization of Convective Solvent Flow During Iontophoresis. *Pharm. Res.* **1994**, *11*, 929–935.
53. Hoogstraate, A.J.; Srinivasan, V.; Sims, S.M.; Higuchi, W.I. Iontophoretic Enhancement of Peptides: Behaviour of Leuprolide Versus Model Permeants. *J. Controlled Release* **1994**, *31*, 41–47.
54. Hirvonen, J.; Kalia, Y.N.; Guy, R.H. Transdermal Delivery of Peptides by Iontophoresis. *Nature Biotech.* **1996**, *14*, 1710–1713.
55. Green, P.G.; Hinz, R.S.; Cullander, C.; Yamane, G.; Guy, R.H. Iontophoretic Delivery of Amino Acids and Amino Acid Derivatives Across the Skin “In Vitro”. *Pharm. Res.* **1991**, *8*, 1113–1120.
56. Green, P.G.; Hinz, R.S.; Kim, A.; Szoka, F.C.; Guy, R.H. Iontophoretic Delivery of a Series of Tripeptides Across the Skin In-Vitro. *Pharm. Res.* **1991**, *8*, 1121–1127.
57. DelTerzo, S.; Behl, C.R.; Nash, R.A. Iontophoretic Transport of a Homologous Series of Ionized and Non-ionized Model Compounds: Influence of Hydrophobicity and Mechanistic Interpretation. *Pharm. Res.* **1989**, *6*, 85–90.
58. Hirvonen, J.; Guy, R.H. Transdermal Iontophoresis: Modulation of Electroosmosis by Polypeptides. *J. Controlled Release* **1998**, *50*, 283–289.
59. Turner, N.G.; Ferry, L.; Price, M.; Cullander, C.; Guy, R.H. Iontophoresis of Poly-L-Lysines: The Role of Molecular Weight. *Pharm. Res.* **1997**, *14*, 1322–1331.
60. Santi, P.; Guy, R.H. Reverse Iontophoresis—Parameters Determining Electroosmotic Flow. I. PH and Ionic Strength. *J. Controlled Release* **1996**, *38*, 159–165.
61. Sarpotdar, P.P.; Daniels, C.R. Use of Polymeric Buffers to Facilitate Iontophoretic Transport of Drugs. *Pharm. Res.* **1990**, *7* (Suppl.), S-185.
62. Santi, P.; Colombo, P.; Bettini, R.; Catellani, P.L.; Minutello, A.; Volpato, N.M. Drug Reservoir Composition and Transport of Salmon Calcitonin in Transdermal Iontophoresis. *Pharm. Res.* **1997**, *14*, 63–66.
63. Cullander, C.; Rao, G.; Guy, R.H. Why Silver/Silver Chloride? Criteria for Iontophoresis Electrodes. *Prediction of Percutaneous Penetration*; Brain, K.R., James, V.J., Walters, K.A., Eds.; STS Publishing: Cardiff, U.K., 1993; 3b, 381–390.
64. Knoblauch, P.; Moll, F. In Vitro Pulsatile and Continuous Transdermal Delivery of Buserelin by Iontophoresis. *J. Controlled Release* **1993**, *26*, 203–212.
65. Gupta, S.K.; Kumar, S.; Bolton, S.; Behl, C.R.; Malick, A.W. Optimization of Iontophoretic Transdermal Delivery of a Peptide and a Non-Peptide Drug. *J. Controlled Release* **1994**, *30*, 253–261.
66. Kumar, S.; Char, H.; Patel, S.; Piemontese, D.; Iqbal, K.; Malick, A.W.; Neugroschel, E.; Behl, C.R. Effect of Iontophoresis on In Vitro Skin Permeation of an Analogue of Growth Hormone Releasing Hormone Factor in the Hairless Guinea Pig Model. *J. Pharm. Sci.* **1992**, *81*, 635–639.
67. Brady, A.B.; Corish, J.; Corrigan, O.I. Passive and Electrically Assisted Transdermal Delivery of Desamino-8-D-Arginine Vasopressin In Vitro from a Gel Matrix. *Prediction of Percutaneous Penetration*; Scott, R.C., Guy, R.H., Hadgraft, J., Boddé, H.E. Eds.; STS Publishing: Cardiff, U.K., 1991; 2, 401–409.
68. Hager, D.F.; Mancuso, F.A.; Nazareno, J.P.; Sharkey, J.W.; Siverly, J.R. Evaluation of a Cultured Skin Equivalent as a Model Membrane for Iontophoretic Transport. *J. Controlled Release* **1994**, *30*, 117–123.
69. Delgado-Charro, M.B.; Guy, R.H. Iontophoretic Delivery of Nafarelin Across the Skin. *Int. J. Pharm.* **1995**, *117*, 165–172.
70. Delgado-Charro, M.B.; Rodriguez-Bayón, A.M.; Guy, R.H. Iontophoresis of Nafarelin: Effects of Current Density and Concentration on Electrotransport In Vitro. *J. Controlled Release* **1995**, *35*, 35–40.

71. Green, P.G.; Shroet, B.; Bernerd, F.; Pilgrim, W.R.; Guy, R.H. In Vitro and In Vivo Iontophoresis of a Tripeptide Across Nude Rat Skin. *J. Controlled Release* **1992**, *20*, 209.
72. Lau, D.T.W.; Sharkey, J.W.; Petryk, L.; Mancuso, F.A.; Yu, Z.; Tse, F.L.S. Effect of Current Magnitude and Drug Concentration on Iontophoretic Delivery of Octreotide Acetate (Sandostatin®) in the Rabbit. *Pharm. Res.* **1994**, *11*, 1742–1746.
73. Thysman, S.; Hanchard, C.; Pr  at, V. Human Calcitonin Delivery in Rats by Iontophoresis. *J. Pharm. Pharmacol.* **1994**, *46*, 725–730.
74. Heit, M.C.; Williams, P.L.; Jayes, F.L.; Chang, S.K.; Riviere, J.E. Transdermal Iontophoretic Peptide Delivery: In Vitro and In Vivo Studies with Luteinizing Hormone Releasing Hormone. *J. Pharm. Sci.* **1993**, *83*, 240–243.
75. Kumar, S.; Char, H.; Patel, S.; Piemontese, D.; Malick, A.W.; Iqbal, K.; Neugroschel, E.; Behl, C.R. In Vivo Transdermal Iontophoretic Delivery of Growth Hormone Releasing Factor GRF (1–44) in Hairless Guinea Pigs. *J. Controlled Release* **1992**, *18*, 213–220.
76. Gherardini, G.; G  rlek, A.; Evans, G.R.D.; Milner, S.M.; Matarasso, A.; Wassler, M.; Jernbeck, J.; Lundberg, T. Venous Ulcers: Improved Healing by Iontophoretic Administration of Calcitonin Gene-Related Peptide and Vasoactive Intestinal Polypeptide. *Plast. Reconstr. Surg.* **1998**, *101*, 90–93.
77. Meyer, B.R.; Kreis, W.; Eschbah, J.; O'Mara, V.; Rosen, S.; Sibal  s, D. Transdermal Versus Subcutaneous Leuprolide: A Comparison of Acute Pharmacodynamic Effect. *Clin. Pharmacol. Ther.* **1990**, *48*, 340–345.
78. Lu, M.F.; Lee, D.; Carlson, R.; Rao, G.S.; Hui, H.W.; Adjei, L.; Herrin, M.; Sundberg, D.; Hsu, L. The Effects of Formulation Variables on Iontophoretic Transdermal Delivery of Leuprolide to Humans. *Drug. Dev. Ind. Pharm.* **1993**, *19*, 1557–1571.
79. Guy, R.H. Current Status and Future Prospects of Transdermal Drug Delivery. *Pharm. Res.* **1996**, *13*, 1765–1769.
80. Sage, B.H., Jr. Insulin Iontophoresis. *Protein Delivery—Physical Systems*; Sanders, L.M., Hendren, R.W., Eds.; Plenum Publishing Corp.: New York, 1997; 319–341.
81. Vanbever, R.; Pr  at, V. In Vivo Efficacy and Safety of Skin Electroporation. *Adv. Drug Deliv. Rev.* **1999**, *35*, 77–88.
82. Weaver, J.C.; Vaughan, T.E.; Chizmadzhev, Y. Theory of Electrical Creation of Aqueous Pathways Across Skin Transport Barriers. *Adv. Drug Deliv. Rev.* **1999**, *35*, 21–39.
83. Prausnitz, M.R. A Practical Assessment of Transdermal Drug Delivery by Skin Electroporation. *Adv. Drug Deliv. Rev.* **1999**, *35*, 61–76.
84. Jadoul, A.; Bouwstra, J.; Pr  at, V. Effects of Iontophoresis and Electroporation on the Stratum Corneum: Review of the Biophysical Studies. *Adv. Drug Deliv. Rev.* **1999**, *35*, 89–105.
85. Pliquett, U. Mechanistic Studies of Molecular Transdermal Transport Due to Skin Electroporation. *Adv. Drug Deliv. Rev.* **1999**, *35*, 41–60.
86. Vanbever, R.; Pliquett, U.F.; Pr  at, V.; Weaver, J.C. Comparison of the Effects of Short, High-Voltage and Long, Medium-Voltage Pulses on Skin Electrical and Transport Properties. *J. Controlled Release* **1999**, *69*, 35–47.
87. Chen, T.; Segall, E.M.; Langer, R.; Weaver, J.C. Skin Electroporation: Rapid Measurements of the Transdermal Voltage and Flux of Four Fluorescent Molecules Show a Transition to Large Fluxes Near 50 V. *J. Pharm. Sci.* **1998**, *87*, 1368–1374.
88. Vanbever, R.; Pr  at, V. In Vivo Efficacy and Safety of Skin Electroporation. *Adv. Drug Deliv. Rev.* **1999**, *35*, 77–88.
89. Bommannan, D.B.; Tamada, J.; Leung, L.; Potts, R.O. Effect of Electroporation on Transdermal Iontophoretic Delivery of Luteinizing Hormone Releasing Hormone (LHRH) in Vitro. *Pharm. Res.* **1994**, *11*, 1809–1814.
90. Riviere, J.E.; Monteiro-Riviere, N.A.; Rogers, R.A.; Bommannan, D.; Tamada, J.A.; Potts, R.O. Pulsatile Transdermal Delivery of LHRH Using Electroporation: Drug Delivery and Skin Toxicology. *J. Controlled Release* **1995**, *36*, 229–233.
91. Wang, S.; Kara, M.; Krishnan, T.R. Transdermal Delivery of Cyclosporin-A Using Electroporation. *J. Controlled Release* **1998**, *50*, 61–70.
92. Wang, S.; Kara, M.; Krishnan, T.R. Topical Delivery of Cyclosporin a Coevaporate Using Electroporation Technique. *Drug Dev. Ind. Pharm.* **1997**, *23*, 657–663.
93. Weiman, A.E. Physical Principles of Ultrasound. *Principles and Practice of Echocardiography*; Weiman, A.E. Ed.; Lea & Febiger: Philadelphia, 1994; 3–28.
94. Meidan, V.M.; Walmsley, A.D.; Irwin, W.J. Phonophoresis —Is it a Reality. *Int. J. Pharm.* **1995**, *118*, 129–149.
95. Camel, E. Ultrasound. *Percutaneous Penetration Enhancers*; Smith, E.W., Maibach, H.I. Eds.; CRC Press: Boca Raton, 1995; 369–382.
96. Mitragotri, S.; Blankschtein, D.; Langer, R. Sonophoresis: Enhanced Transdermal Drug Delivery by Application of Ultrasound. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C. Eds.; Marcel Dekker, Inc.: New York, 1996; 14, 103–122.
97. Tachibana, K.; Tachibana, S. Transdermal Delivery of Insulin by Ultrasonic Vibration. *J. Pharm. Pharmacol.* **1991**, *43*, 270–271.
98. Tachibana, K. Transdermal Delivery of Insulin to Alloxan-Diabetic Rabbits by Ultrasound Exposure. *Pharm. Res.* **1992**, *9*, 952–954.
99. Mitragotri, S.; Blankschtein, D.; Langer, R. Ultrasound-Mediated Transdermal Protein Delivery. *Science* **1995**, *269*, 850–853.
100. Mitragotri, S.; Farrell, J.; Tang, H.; Terahara, T.; Kost, J.; Langer, R. Determination of Threshold Energy Dose for Ultrasound-Induced Transdermal Drug Transport. *J. Controlled Release* **2000**, *63*, 41–52.
101. Mitragotri, S.; Edwards, D.A.; Blankschtein, D.; Langer, R. A Mechanistic Study of Ultrasonically-Enhanced Transdermal Drug Delivery. *J. Pharm. Sci.* **1995**, *84*, 697–706.
102. Menon, G.K.; Bommannan, D.B.; Elias, P.M. High-Frequency Sonophoresis: Permeation Pathways and Structural Basis for Enhanced Permeability. *Skin Pharmacol.* **1994**, *7*, 130–139.
103. Meidan, V.M.; Docker, M.; Walmsley, A.D.; Irwin, W.J. Low Intensity Ultrasound as a Probe to Elucidate the

- Relative Follicular Contribution to Total Transdermal Absorption. *Pharm. Res.* **1998**, *15*, 85–92.
104. Merino, G.; Delgado-Charro, M.B.; Kalia, Y.N.; Guy, R.H. Mechanism of Enhanced Transdermal Transport by Low-Frequency Ultrasound: The Role of Local Heating. *Proc. Int. Symp. Controlled Release Bioact. Mater.* **2000**, *27*, 942–943.
 105. Weimann, L.J.; Wu, J. Enhancement of Skin Penetration of Poly-L-Lisine by Ultrasound. *Proc. Int. Symp. Controlled Release Bioact. Mater.* **1999**, *26*, 417–418.
 106. Lee, S.; Kollias, N.; McAuliffe, D.J.; Flotte, T.J.; Doukas, A.G. Topical Drug Delivery in Humans with a Single Photomechanical Wave. *Pharm. Res.* **1999**, *16*, 1717–1721.
 107. Blau, H.M.; Springer, M.L. Gene Therapy—A Novel Form of Drug Delivery. *N. Engl. J. Med.* **1995**, *333*, 1204–1207.
 108. Khavari, P.A. Therapeutic Gene Delivery to the Skin. *Mol. Med. Today* **1997**, December, 533–538.
 109. Jaroszeski, M.J.; Gilbert, R.; Nicolau, C.; Heller, R. In Vivo Gene Delivery by Electroporation. *Adv. Drug Deliv. Rev.* **1999**, *35*, 131–137.
 110. Lin, M.T.; Pulkkinen, L.; Vitto, J.; Yoon, K. The Gene Gun: Current Applications in Cutaneous Gene Therapy. *Int. J. Dermatol.* **2000**, *39*, 161–176.
 111. Babiuk, S.; Baca-Estrada, M.; Babiuk, L.A.; Ewen, C.; Foldvari, M. Cutaneous Vaccination: The Skin as an Immunologically Active Tissue and the Challenge of Antigen Delivery. *J. Controlled Release* **2000**, *66*, 199–214.
 112. Fan, H.; Lin, Q.; Morrissey, G.R.; Khavari, P.A. Immunization Via Hair Follicles by Topical Application of Naked DNA to Normal Skin. *Nature Biotech.* **1999**, *17*, 870–872.
 113. Erspamer, V.; Falconieri-Erspamer, G.; Severini, C.; Potenza, R.L.; Barra, D.; Mignogna, G.; Bianchi, A. Pharmacological Studies of ‘Sapo’ From the Frog *Phyllomedusa Bicolor* Skin: A Drug Used by the Peruvian Matses Indians in Shamanic Hunting Practices. *Toxicol.* **1993**, *31*, 1099–1111.
 114. Down, J.A.; Harvey, N.G. Minimally Invasive Systems for Transdermal Drug Delivery. in press.
 115. Nelson, J.S.; McCullough, J.L.; Glenn, T.; Wright, W.H.; Liaw, L.L.; Jacques, S.L. Mid-Infrared Laser Ablation of Stratum Corneum Enhances In Vitro Percutaneous Transport of Drugs. *J. Invest. Dermatol.* **1991**, *97*, 874–879.
 116. Eppstein, J.A.; Delcher, H.K.; Hatch, M.R.; McRae, M.S.; Papp, J.; Woods, T.J. Insulin Infusion Via Thermal Micropores. *Proc. Int. Symp. Controlled Release Bioact. Mater.* **2000**, *27*, 1010–1011.
 117. Svedman, P.; Lundin, S.; Svedman, C. Administration of Antidiuretic Peptide (DDAVP) by Way of Suction De-epithelialised Skin. *Lancet* **1991**, *337*, 1506–1509.
 118. Lundin, S.; Svedman, P.; Höglund, P.; Jönsson, K.; Broeders, A.; Melin, P. Absorption of an Oxytocin Antagonist (Antocin) and a Vasopressin Analogue (dDAVP) Through a Standardized Skin Erosion in Volunteers. *Pharm. Res.* **1995**, *12*, 2024–2029.
 119. Sarphie, D.F.; Johnson, B.; Cormier, M.; Burkoth, T.L.; Bellhouse, B.J. Bioavailability Following Transdermal Powdered Delivery (TPD) of Radiolabeled Insulin to Hairless Guinea Pigs. *J. Controlled Release* **1997**, *47*, 61–69.
 120. Sarphie, D.F.; Varadi, A.; Bellhouse, B.J.; Ashcroft, S. Transdermal Delivery of Powdered Bovine Insulin to Diabetic and Non-Diabetic Wistar Rats. *Diabetes* **1995**, *44* (Suppl. 1), 129a.
 121. Henry, S.; McAllister, D.V.; Allen, M.G.; Prausnitz, M.R. Microfabricated Microneedles: A Novel Approach to Transdermal Drug Delivery. *J. Pharm. Sci.* **1998**, *87*, 922–925.
 122. McAllister, D.V.; Kaushik, S.; Patel, P.N.; Mayberry, J.L.; Allen, M.G.; Prausnitz, M.R. Solid and Hollow Microneedles for Transdermal Protein Delivery. *Proc. Intl. Symp. Controlled Release Bioact. Mater.* **1999**, *26*, 192–193.

PHARMACEUTICAL EXCIPIENT TESTING—A REGULATORY AND PRECLINICAL PERSPECTIVE^a

Paul Baldrick

Regulatory Affairs-Pharmaceuticals, Covance Laboratories Ltd., Harrogate, England

INTRODUCTION

Pharmaceutical excipients are additives used in the formulation of pharmacologically active drugs and can be viewed as any ingredient of a medicinal product other than the active ingredient. Excipients include diluents, fillers and bulking agents, binders and adhesives, propellants, disintegrants, lubricants and glidants, colors, flavors, coating agents, polishing agents, fragrances, sweetening agents, polymers, and waxes. A review of the literature indicates approximately 1300 excipients used in the pharmaceutical industry. Chemical and manufacturing (quality) data can be found for many of these excipients in European, U.S., and Japanese pharmacopoeias (1–3) or in a range of other publications (4–6), but safety (preclinical) data are limited. General safety data can be found in some publications (5,7). Reference to excipients included in drug submissions is given in the FDA *Inactive Ingredients Guide* (8), the *National Formulary* (9), and the U.K. *ABPI Data Sheets* (10). It is assumed that reference to an excipient in these publications is a guarantee of acceptance by a regulatory body of the excipient in a new drug formulation. Although this is often the case, there may be gaps in the available data that may necessitate additional preclinical evaluation. The use of a novel excipient in a drug formulation necessitates preclinical testing and evaluation. This article addresses strategies for preclinical assessment of completely novel, essentially similar, and established excipients in view of current difficulties posed by the lack of a concerted international guideline relating directly to pharmaceutical excipients. In addition, the active role of excipients in producing adverse patient effects and in modulating active drug release is presented.

REGULATIONS AND GUIDELINES

From an international regulatory point of view, there is currently little help to address the question of registration of an excipient as a separate entity. Furthermore, there is no such thing as an approved excipient (11). In Europe, it is assumed that novel excipients need to be evaluated as new chemical entities (12–14). Established excipients are included in marketing authorization applications (MAAs) for new drugs, with the assumption that their presence and characterization in pharmacopoeias will not cause concern with European regulators. Information on excipients is expected in MAA applications (15). In the United States, the Food and Drug Administration (FDA) assesses and permits use of excipients as part of a new drug application (NDA). As in Europe, it is assumed that the use of an “approved” excipient ensures its acceptance in the new drug formulation. Thus, the FDA favors commercially established excipients (16). Interestingly, in the U.S. guideline relating to preclinical data for NDA submission, excipients are not mentioned (17). Although excipients per se are not mentioned in the Japanese guidelines (18), the situation with regard to such materials is similar to that in Europe and the United States. Under the new Japanese Ministry of Health and Welfare (JMHW) evaluation system (operational since 1997), the assessment of pharmaceutical products containing excipients with prior use in Japan is performed at the Pharmaceuticals and Medical Devices Evaluation Center (PMDE). Approval of a product that contains an excipient with no prior use in Japan has to be evaluated by the Subcommittee on Pharmaceutical Excipients of the Central Pharmaceutical Affairs Council (CPAC) concurrently with the approval process undertaken through the PMDE Center (19). Reasons for inclusion of the excipient; precedents of use; and quality, stability, and safety data are all needed.

The lack of international regulatory guidelines led to the formation of the International Pharmaceutical Excipients Council (IPEC) in 1991. This industry association, with European, U.S., and Japanese membership, has championed the international standardization of excipients,

^aUpdated from Pharmaceutical Excipient Testing—A Regulatory and Preclinical Perspective. In *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J.; Boylan, J.C. (Eds.), 1st Ed., Marcel Dekker, Inc., New York, 2000, Vol. 19, 289–310.

the introduction of useful new excipients, and the development of safety evaluation guidelines. These guidelines, addressing all the primary routes of drug administration, have been published (16). These proposed safety guidelines has been given to the FDA by the IPEC, with a request for an evaluation procedure, independent of the NDA process, for new pharmaceutical excipients (20).

In Europe, the need for companies to fully consider the types of excipients in new drug formulations has been the focus of a guideline released recently by the European commission on drug label and package leaflet excipient information for all MAAs (14). This guideline indicates that all excipients must be listed on the drug label for parenteral, topical, inhalation, and ophthalmic products. For other medicinal products, only those excipients with a defined and recognized action of effect, as listed in the guidance document (over 40 materials are mentioned), need to be declared on the label. In drug package leaflets, all excipients must be included. European regulations and guidelines (12, 13, 21) also require that a full statement of the excipients used be given in the Summary of Product Characteristics (SPC) document for new drug formulations. In Japan, the identity and quantity of the inactive components must be indicated on package inserts and similar materials for medicines for oral use, injection, and applications on the skin and mucosal surfaces (22).

The situation is different in the United States, where labeling of inactive ingredients remains voluntary (23), although such labeling has been adopted by two major pharmaceutical industry trade associations (the Proprietary Association and the Pharmaceutical Manufacturers Association). It may be that companies are reluctant to give details on their products because the excipients will not have patent protection. Thus, full disclosure of excipients is still not a general practice and can lead to opposition (24). The FDA indicated, however, that inactive ingredients (that have the same meaning as excipients) that present an increased risk of toxic effects to neonates or other pediatric subgroups need to be noted in the contraindications, warning, or precautions section of labels on prescription drugs for humans (25).

In Europe, in the preclinical area stage, reference is made (but without any details) to the suitable demonstration of safety of any new propellant (26, 27). Vaccine guidelines group excipients (identified as inactive components including stabilizers) with adjuvants and preservatives as additives and indicate that safety appraisal is needed, again without details (28). A recent European regulatory guidance on the development of pharmaceuticals has stated that, "A new substance introduced as a constituent will be regarded in the same way as a new active ingredient H unless it is already approved for use in

food for orally administered products, or in cosmetics for topical administration. Additional data may still be required where an excipient is administered via an unconventional route, or in high doses" (29). However, although this wording may allow for some reduced data for commonly used nonpharmaceutical materials, it is unlikely that in the meantime in Europe, novel excipients will be treated as other than new chemical entities.

CATEGORIES

New (Novel) Recipients

A new excipient is a compound that has not been used previously or permitted for use in a pharmaceutical preparation (16). The need to develop an excipient as a new chemical entity, leading to expense and potential time delay (the excipient would be of low regulatory priority compared with active new drug formulations), has resulted in a reluctance in the pharmaceutical industry to introduce new materials. Indeed, it is likely that if there is doubt about the safety of an excipient, pharmaceutical manufacturers would rather reformulate than incur registration delays (11). Furthermore, despite the possibility that greater effectiveness of pharmaceutical products containing new excipients exists, a desire to avoid regulatory delays has resulted in a lack of research in this area (19). A review of the 12 most common excipients (water, magnesium stearate, starch, lactose, microcrystalline cellulose, stearic acid, sucrose, talc, silicon dioxide, gelatin, acacia, and dibasic calcium phosphate, respectively) in NDAs from 1964 to 1984 shows that six were being used in 1904, increasing to the majority of these materials by 1949 (30). A review of the most common excipients in U.K.-licensed medicines showed the following materials: water, magnesium stearate, povidone, sodium chloride, stearic acid, and dextrose (31). In 1992, Strattan reported that no truly new excipients have become available for general use in the past 50 years (32). The appearance of the cyclodextrins in the 1990s and the recent approval of the inhalation propellant hydrofluoroalkane (HFA) have only slightly changed this fact. The U.S. pharmaceutical companies have primarily used "grandfathering" for their new substances (using old "unapproved" drug formulations), thus preventing advancement in the use of excipients.

As noted above, the IPEC has proposed guidelines for the safety assessment of new excipients. It is hoped that implementation of these guidelines would expedite the review of a proposed new excipient by regulatory agencies

Table 1 Summary of excipient guideline^a

Tests	Routes of exposure for humans						
	Oral	Mucosal	Transdermal	Topical	Parenteral	Inhalation/intranasal	Ocular
Appendix 1 (Base set):							
Acute oral toxicity	R	R	R	R	R	R	R
Acute dermal toxicity	R	R	R	R	R	R	R
Acute inhalation toxicity	C	C	C	C	C	R	C
Eye irritation	R	R	R	R	R	R	R
Skin irritation	R	R	R	R	R	R	R
Skin sensitization	R	R	R	R	R	R	R
Acute parenteral toxicity	—	—	—	—	R	—	—
Application site evaluation	—	R	R	R	R	R	—
Pulmonary sensitization	—	—	—	—	—	R	—
Phototoxicity/photoallergy	—	—	R	R	—	—	—
Bacterial gene mutation	R	R	R	R	R	R	R
Chromosomal damage	R	R	R	R	R	R	R
ADME-intended route	R	R	R	R	R	R	R
28-day toxicity (two species) intended route	R	R	R	R	R	R	R
Appendix 2							
90-day toxicity (most appropriate species)	R	R	R	R	R	R	R
Teratology (rat and/or rabbit)	R	R	R	R	R	R	R
Additional assays	C	C	C	C	C	C	C
Genotoxicity assays	R	R	R	R	R	R	R
Appendix 3							
Chronic toxicity (rodent, nonrodent)	C	C	C	C	C	C	C
One-generation reproduction	R	R	R	R	R	R	R
Photocarcinogenicity	—	—	C	C	—	—	—
Carcinogenicity	C	C	C	C	C	C	—

^a Extent of testing is dependent on conditions and duration of exposure: Appendix 1 for exposures of less than 2 weeks, Appendix 2 for exposures of 2 to 6 weeks, and Appendix 3 for exposures of greater than 6 weeks.
(From Ref. 16.)
R, required; C, Conditional.

(16). A summary of the IPEC guidelines is given in Table 1. This organization has defined a testing strategy for human exposure to single or limited dosing (less than 2 weeks), limited and repeated dosing (2–6 weeks), and long-term dosing (longer than 6 weeks). Different tests for oral, mucosal, transdermal, topical, parenteral, inhalation/intranasal, and ocular routes in humans are described. Common recommended toxicity studies for longer term use in humans include acute oral and dermal toxicity studies, eye and skin irritation studies, a skin sensitization study, genotoxicity and ADME studies, subacute and chronic studies, and reproduction studies. These guidelines are useful as a basis for the development of a new excipient, although they raise some questions. The basis for the guidance is that because of the defined inert nature of excipients, the safety assessment procedure should not be as complex or as extensive as that for an active drug. However, the program described is not dissimilar to the strategy used for an active drug. Furthermore, acute dermal, irritation, and skin sensitization studies are not routinely performed in new drug assessment unless there are specific issues or questions about the drug class. Finally, and most important, if a drug company was developing a drug with a completely new excipient in the formulation, it is possible to include the excipient alone as an extra dose group in the preclinical studies. These studies would presumably use the clinical formulation (containing the excipient). It is vital that kinetics studies identify any drug-excipient interaction. As for active drug, a sensitive analytical method is needed to measure levels of excipient (if possible) in plasma samples from toxicology studies. If the impurity profile of the new excipient causes concern, additional toxicological evaluation of each impurity has to be carried out as well. It is crucial to clarify the excipient impurity profile because even established pharmacopoeia-listed materials such as magnesium stearate can raise questions as to the safety and toxicity of its impurities (33).

In Japan, Uchiyama has recently published requirements for the safety evaluation of new excipients (19). These requirements include studies on acute, subacute, and chronic toxicity, mutagenicity, effects on reproduction, dependency, antigenicity, carcinogenicity, and local irritation (human patch test). The first five of these tests are mandatory. With the exception of the local irritation test, for which a domestic trial is required, non-Japanese data are acceptable for these studies. Even if a material has been used in a pharmaceutical product outside Japan, the material is treated as a new excipient if there has been no prior use in Japan, although relevant overseas data for the material are acceptable for regulatory submission. A material is treated as a new excipient

when the route of administration differs or the dose level exceeds that of prior use even after approval for the Japanese market (19).

Although not viewed as a typical excipient, the inhalation propellant HFA has been approved in both the United States and Europe in recent years. Because of concerns about ozone layer depletion defined in the Montreal protocol, the need for a replacement for chlorofluorocarbons (CFCs) in metered-dose inhalers (MDIs) used by patients with asthmas led to the development of two types of HFA (134a and 227) in Europe (34, 35). Both compounds underwent extensive preclinical testing in the early to mid-1990s. The change from established CFC-containing drugs to those with HFA requires toxicity-bridging studies. From a regulatory point of view, these studies can vary from 1 month to lifetime in duration. A recent U.S. publication has revealed a number of preclinical studies, including those for carcinogenicity, using hydrochlorofluorocarbons (HCFCs) and hydrofluorocarbons (HFCs) as replacements for CFCs (36). The data show that the new HFA-134a and -227 excipients are as safe for use in humans as the CFCs they are replacing (37, 38). Using the example of HFA, it can be seen that even with government pressure to change a formulation, considerable time and effort are needed for the development of a new excipient.

Finally, with regard to the development of new materials for inhalation products, Wolff and Dorato suggest acute or short-term toxicology studies of a number of possible excipients in different formulations (34). This process will allow the narrowing to one or two acceptable possibilities that can then undergo more complete toxicological testing. The testing of new inhalation excipients is also addressed in a guidance from the FDA's Center for Drug Evaluation and Research (CDER) (39). A complete toxicological evaluation is recommended for a new excipient with unknown inhalation toxicology potential (such as a new propellant) alone as well as bridging studies in animals with the new complete formulation. Such studies are generally sufficient for regulatory approval.

Essentially New Excipients

Essentially new excipients can be thought of as substances resulting from a structural modification of an "approved" excipient, a recognized food additive, a structurally modified food additive, or a constituent of an over-the-counter (OTC) medicine. Food additives are similar to excipients in that they are used to impart specific functional characteristics and are usually added

in low levels (16). International organizations including the Joint Food and Agriculture Organisation (FAO) and the World Health Organization (WHO) Expert Committee (JECFA) have established safety standards such as an acceptable daily intake (ADI) or “generally recognized as safe” (GRAS) for a number of food additives, including materials that have applications as excipients. Safety evaluation of these additives involves a detailed risk-assessment process, with ADI calculations having factors such as 100-fold-above-animal data as safety margins.

Although not covered by any regulatory guidance, the safety of many food additives determined through JECFA evaluation and long-term human exposure allows for approval of these materials as new excipients in drug formulations without the need for extensive preclinical testing. Indeed, the FDA favors commercially established food additives and GRAS substances (16). However, limitations and possible significant financial investment exist because of the possibility of long waiting periods for approval and the fact that food additive data relate to oral administration and are of little use for nonoral products (11).

Additionally, problems may arise if the preclinical data on which the assessment is made are old and unreliable and raise specific toxicological concerns. Considering all these factors, it is probable that a robust expert review of available preclinical and human exposure data for a food additive may satisfy a regulatory body. As noted above, there is an indication in Europe that new pharmaceutical excipients already used in food for orally administered products or for topical cosmetics may not have to be considered new chemical entities, and therefore full evaluation may be bypassed (29).

Other problems may arise if the level of the food additive to be used as a proposed new excipient is higher than the ADI value or if a food additive or an established excipient are structurally (albeit simply) modified. The new material may have a different or more toxic profile. There are no easy solutions to these

situations for pharmaceutical companies. The regulatory answer will be that each material is assessed on a case-by-case basis. However, some form of preclinical testing will probably be required, and a possible minimal program is given in Fig. 1. In this program, an Ames study determines that the new excipient has no genotoxicity potential, and a single-dose toxicity study shows that there are no adverse effects after administration of the material at the limit (high) dose generally used in the pharmaceutical industry. An investigative mass, balance, whole-body autoradiography study provides information on absorption, distribution, metabolism, and excretion. This study also involves an investigation of suitable labeling of the material. An *in vitro* metabolism study (e.g., in rat versus human hepatocytes) may also be useful for modified food additives and excipients to compare the break-down process and to show possible differences between rat and human in these processes.

A 1-month toxicity study will establish whether the limit dose is causing any form of high-dose toxicity or metabolic overload. It may be useful to include toxicokinetic satellite animals in this study to demonstrate what level of exposure to the excipient occurs in the blood. Favorable findings from these studies can be presented to the regulatory bodies and advice sought as to whether additional testing is recommended. As with other preclinical studies in drug development, all these investigations should be performed according to good laboratory practice (GLP).

In the United States, OTC products do not need regulatory approval of submission of an NDA. Thus, if an excipient has a record of safe use in an OTC medicine, the FDA may accept the material, provided the product manufacturer can demonstrate evidence that the excipient is safe to use (11). However, this manner of approval is not guaranteed to be successful and may not be useful for nonoral administration routes.

In Japan, all these pathways to reduce ease the burden of introducing new excipients are more complex. If not already approved as a pharmaceutical excipient in Japan, food additives for oral administration or cosmetic substances for external application in drug formulations requires that they be treated as new excipients (19). For excipients already used in orally or intravenously administered products, a change to an externally applied product necessitates additional safety testing. Studies include acute and subacute toxicity (including adsorption through the skin) and local irritation investigations (19).

Many excipients used by the pharmaceutical industry in the last 15 years in sugar-free medical preparations

Test
Ames test
Acute oral toxicity study
Mass/balance/whole body autoradiography (WBA) study and/or <i>in-vitro</i> metabolism study
One-month repeat dose toxicity study

Fig. 1 Minimal preclinical testing strategy for an essentially new excipient.

probably come under the category of essentially new excipients. Pressure for their introduction has been encouraged by the definite relationship between the dietary consumption of sucrose and the incidence of dental caries (40, 41). These materials include intense sweeteners such as saccharin and cyclamate plus bulk sweeteners such as the polyols sorbitol, xylitol, and lactitol. These materials are all either approved for food use or have pharmacopoeia monographs in existence or in draft (42). Literature reviews show number of preclinical safety studies for these excipients. Many of these studies are related to specific toxicological concerns such as bladder tumor formation with saccharin use and adrenal glands and testes proliferation with polyol use (Table 2).

The cyclodextrins (CDs) fall into the categories of new and essentially new excipients. CDs are enzymatically modified starches with many favorable properties as excipients. The use of CDs in food products in the late 1970s and 1980s, together with extensive preclinical testing, has resulted in the inclusion of these materials in marketed drugs (32, 62). Various companies now market forms of CDs as safe and efficient delivery systems for active drugs.

Established (“Approved”) Excipients

The presence of well known established excipients in a new drug formulation does not necessarily mean that regulatory authorities question will not their inclusion. However, issues can be avoided if the excipient is clearly characterized in the quality section of the documentation and addressed sufficiently in the preclinical section. The preclinical assessment can take the form of an expert review of available safety data that defines safety margins between values reported in the literature and the level(s) used in the new drug formulation. This process is becoming easier for toxicologists as more data on established excipients become available through published literature reviews and/or publication of in-house studies. Examples include lactose (58), povidone or polyvinylpyrrolidone (PVP) (50, 63), saccharin (64), polyethylene glycols (PEGs) (65), HFA-134a (66), and vaccine adjuvants/excipients (67). Similarly, the establishment of global standardized monographs in major pharmacopoeias is helpful. However, it is surprising how little useful toxicology data are available for many well known materials such as maltodextrin, peppermint oil, and menthol.

Problems can arise when the currently available published data suggest that there may be potential toxicity concerns, especially when an excipient approved for one dose route is applied to another route with a different systemic exposure and target site. Scientific discussions of the data on the relevance of animal findings to humans, together with the establishment of safety margins, are necessary. The absence of toxicity findings in the new drug preclinical program that includes the excipient can also be used in these discussions. Examples of established excipients with reported toxicological concerns and the relevance of these findings to humans are given in Table 2. Another potential issue for discussion arises if the level of excipient in the new drug formulation exceeds that already known to the regulatory agency for prior use.

A less important but interesting point is that many of the established and pharmacopoeia-listed excipients are not accompanied by strict standardized assay methods. In this situation, a “new” assay needs to be established by the company with the new drug formulation for characterization and/or plasma measurement of the excipient.

The discussion above indicates that little or no extra preclinical studies are required for well known excipients. From an unofficial FDA regulatory perspective, this view is supported for inhalation excipients by De George et al. (39), who state that no additional toxicology information is generally needed for qualitative or quantitative changes of well characterized, nontoxic excipients used in approved inhalation drug products. A case-by-case assessment, however, will be used for reformulations because excipient toxicity may change. In addition, for excipients with previous use in humans but limited inhalation toxicity information, evaluation of inhalation toxic potential of the excipient after repeated dosing is recommended.

ADVERSE EFFECTS OF EXCIPIENTS

By traditional definition, excipients should be inert and display no pharmacological activity. However, a number of clinically relevant adverse reactions are known for various established excipients, albeit of low occurrence and uncommon compared with the overall prevalence of adverse drug reactions. This area is covered extensively in the literature (e.g., 7, 14, 23, 40, 68–71). Some examples are given in Table 3. Groups susceptible to excipient toxicity include low-birth-weight infants (particularly during the first 2 weeks of life) and patients

Table 2 Toxicological issues with various excipients

Excipient	Toxicological finding	Relevance to humans	Reference
Lactose and polyols	Adrenal medullary proliferative changes/tumors in rats	Considered to be rat-specific and caused by altered calcium homeostasis attributable to high dose levels given and not relevant to humans	43–45
Lactose and polyols	Hyperplastic/neoplastic changes in rat testes	See above	46
Saccharin	Proliferative changes in male rat bladder epithelium	Considered to be species-specific with large doses used. Mechanism of action may involve urinary proteins not normally found in humans	23, 40, 47–49
PVP	Accumulation in the reticuloendothelial system in rodents, with the occurrence of “foam cells”	Toxicological significance is unclear from the literature, but findings may be related to high-dose regimens used. No adverse effects from long-term storage in humans. There is also over 50 years of safe use for PVP in humans	50–52
Menthol	Limited sensitization reaction in guinea pigs	Toxicological significance is not clear from the literature, although very infrequent human allergy reactions are reported in the presence of menthol (8)	53
Limonene	Hyaline droplet formation in male rat kidney	Considered to be male rat-specific and related to the presence/accumulation of alpha 2 μ -globulin	54–56
Talc	Lung tumors in female rats	Related to high dose level and resulting chronic toxicity	57
Talc	Adrenal gland neoplasms (pheochromocytomas) in rats	Review of data indicates that tumors were not treatment-related	57
Maltodextrin	Minimal reversible laryngeal irritation (squamous metaplasia) with 4% maltodextrin in chronic rat inhalation study	Considered to be a background finding of no consequence to humans. May be related to the presence of peppermint oil (1%) in dosing preparation. ^{a,b} Local irritation of the mouth and oesophagus can occur in patients taking peppermint oil preparations (7, 10)	Personal observation
CFCs	Cardiac arrhythmias, notably in the dog	Relatively good safety margins and long/wide use of CFC propellants in MDIs has shown these compounds to be safe	34

^a Minimal laryngeal squamous metaplasia has also been reported in a chronic rat study with lactose (58) and may be related to a physiological adaptive response attributable to the inhalation of particulate dust at high levels over long periods. Certain histological changes in the respiratory epithelium have been defined as adaptive and not toxic (59, 60).

^b Peppermint oil contains a large percentage of menthol (30–55%) (1). Any question that the laryngeal irritation may be related to the presence of menthol within peppermint oil is doubtful because a recent chronic rat inhalation study has shown that 5000 ppm of menthol had no substantial effect on the histopathological changes in the respiratory tract normally associated with inhalation of mainstream cigarette smoke (61).

Table 3 Adverse effects encountered with various pharmaceutical excipients

Excipient (purpose)	Major adverse event(s)	Reference
Benzalkonium chloride (bacterial preservative)	Irritant known to discolor soft contact lens; rare hypersensitivity, bronchoconstriction	7, 14, 23
CFCs (propellant)	Bronchospasm	23
Aspartame (artificial sweetener)	Headache, various neuropsychiatric disorders, seizures, rare hypersensitivity reactions. Contraindicated for patients with phenylketonuria	7, 23, 40, 70
Saccharin (artificial sweetener)	Rare hypersensitivity, photosensitivity (with cross-sensitivity in patients with sulfonamide allergy)	7, 23, 40, 70
Benzyl alcohol (preservative)	Rare hypersensitivity reactions including contact dermatitis. Contraindicated for infants and young children because of high dose level toxicity	7, 14, 23, 70
Lactose (filler/bulking agent)	Contraindicated for patients with lactose insufficiency (diarrhea, abdominal pain)	14, 58
Propylene glycol (solubilizer)	CNS toxicity, hyperosmolality, lactic acidosis, hypersensitivity including contact dermatitis	7, 23, 70
Benzoic acid (preservative)	Irritant; may have role in asthma, urticaria, anaphylaxis, vasculitis	7, 14, 70
Ethanol (vehicle)	Irritation/dry skin. Contraindicated at high dose levels with liver disease, epilepsy, alcoholism, pregnant women, children	14
Polyols (artificial sweeteners)	Diarrhea	14, 58
Glycerol (vehicle)	Headache, stomach upset, diarrhea, increased serum osmolality, cardiac arrhythmias	7, 14
Sodium and potassium (supplement)	Adverse in cases of low sodium/potassium diet, stomach upset, diarrhea. Phlebitis and injection pain with potassium	7, 14, 23, 40, 70
Sulphites (antioxidants)	Allergic hypereactions, e.g., anaphylaxis, bronchospasm, gastrointestinal irritation/upset	7, 70
Gums, e.g., arabic, tragacanth (emulsifying agents)	Rare hypersensitivity reactions	7, 14, 70
Cremophor EL–Polyoxyethylated castor oil (solvent)	Anaphylactic reactions, nausea, vomiting, colic, hyperlipidaemias	7, 23, 70
Azo dyes e.g., tartrazine (dyes)	Anaphylactoid reaction. Contraindicated for aspirin-intolerant individuals	7, 70
Chlorobutanol (preservative)	Anaphylactic shock/hypersensitivity reactions, CNS toxicity	7, 14, 70
Formaldehyde (preservative)	Allergic reactions, stomach upset, diarrhea, irritant, role in asthma	7, 70
Aluminium hydroxide (adjuvant/adsorbant)	Aluminium allergy, skin nodules	7
Methylcellulose (stabilizer/suspending agent)	Intestinal upset	7
Hydroxyethylcellulose (stabilizer/suspending agent)	Ocular irritation	14, 40
Gluten/wheat starches (binders)	Intolerance for patients with coliac disease	

with asthma (69). It is interesting that, as for many adverse effects, most of the reported excipient-related findings in patients were not or could not be predicted by the preclinical data.

In Europe, it is now a requirement to include adverse excipient effects in the product leaflet in a consumer-friendly manner (14); information on ingredient content is also required in Japan (22). The fact that the labeling of inactive ingredients remains voluntary in the United States means that excipient-related adverse reactions will continue to be reported. Wong (24) highlights a worst-case scenario by showing that a drug can be marketed in various formulated dosage forms such as tablets, injections, and oral solutions, with each of these forms containing a different set of excipients that are not disclosed on the package label. Because the choice of excipients for any pharmaceutical preparation is decided by each company, different manufacturers producing the same drug may use different excipients in different proportions. Another potential issue is the quality of the excipients used in the drug formulation. Inferior grades of excipients may seriously jeopardize the quality and thus the potential safety and efficacy of the final product (24). Fortunately, standardization of the quality of excipients in pharmacopoeias reduces the significance of efficacy as an issue, although the example of the use of counterfeit excipients resulting in 90 pediatric fatalities in Haiti in the mid-1990s should not be forgotten (71).

Chowhan (33) points out the increasing awareness of excipient manufacturers and users of the importance and consequences of process changes. From a medical perspective for drugs with technically inert ingredients that are not necessarily pharmacologically inactive, full information is needed to facilitate the assessment and treatment of patient symptoms. Most excipient-related toxicity is preventable when the formulation is known (69,72). However, difficulty in obtaining such information from U.S. manufacturers continues to be a problem (73). Literature on the occurrence of well known excipients in pharmaceutical products is available as an aid to physicians in selecting preparations containing different excipients when an adverse reaction occurs (74, 75).

Overall, even in early drug development, a consideration of the excipient profile of the final product is important. If use of the established excipients is planned, an awareness of potential adverse effects in humans is necessary. Development should use pharmacopoeia-listed materials. For new excipients, this information is not available, and safety will be assessed in the preclinical evaluation. Extrapolation may only be possible if the new excipient is in the same class or is structurally similar to

established materials, but such a comparison would not necessarily be reliable.

DRUG-EXCIPIENT INTERACTIONS AND AN ACTIVE ROLE FOR EXCIPIENTS

That all excipients are neither truly inert nor inactive is shown by the fact that drug-excipient interactions can considerably affect the physiological availability of many drug products (76). By either accelerating or retarding the release of the active ingredients, excipients can affect the therapeutic performance of the drug by increased or reduced bioavailability (24,69). Indeed, relatively small variations in the physical properties of an excipient can produce a significant difference in the behavior of formulated products (5). In turn, the modified bioavailability may also result in adverse reactions. For modern drug-delivery systems, certain excipients can have a well defined function associated with the need to achieve a specific drug bioavailability profile or therapeutic effect (11). Liposomes (phospholipid-based vesicles) have been increasingly explored as novel drug-delivery systems, and there are examples of active excipient involvement, such as that with PVP- and PEG-liposomes, in development (77,78). Review of the literature shows a surge in interest in using excipients in sustained-release formulations in the 1990s. Various selective functional regions are identified, such as colonic targeting. Currently, many of these new formulations with an active role for the excipient have only been tested using in vitro pharmaceutical systems to establish release patterns, although the role of materials such as methylcellulose in tablet formulations for sustained release has been well established (5). It is beyond the scope of this article to review the progress of these developments, but Naidoo (79) has discussed the effect of excipients on controlled drug-release properties of dosage forms. Reports on controlled delivery of peptide drugs and vaccines using active excipient properties are also available (80,81). Such roles may reflect the definition of an excipient in the U.K. legislation on medicines as “any substance which does not contribute directly to the pharmacological action of the medicinal product otherwise than by regulation of the release of the active ingredients” (82).

Preclinical and clinical pharmacokinetics studies with a new drug indicate systemic exposure after administration of the therapeutic dose formulation. Thus, even if there is an interaction from an established or new

excipient, adequate safety evaluation can be assessed. However, it would be useful to demonstrate to a regulator that possible interference of the active drug by the rest of the formulation had been considered. If an excipient is added to the new drug form for an active role, the proposed mechanism of action will need to be discussed carefully.

OVERALL SUMMARY AND CONCLUSION

Pharmaceutical excipients have a vital role in drug formulations, a role that has tended to be neglected as evidenced by the lack of mechanisms to assess excipient safety outside a new drug application process. Currently, it is assumed that an excipient is “approved” when the new drug formulation of which it is a constituent receives regulatory acceptance. The existing system works for well known excipients that are listed in international pharmacopoeias and for which there are published safety data. However, drug companies are faced with a lack of information on the testing needed for completely or essentially new excipients (the latter category comprising recognized food additives, structurally altered approved excipients or food additives, or constituents of OTC medicines). Existing regulations and guidelines indicate that new excipients be treated as new chemical entities and therefore, by inference, undergo full toxicological evaluation. Other drug requirements in Europe and Japan require that information on excipients appear in patient literature, although there is no similar legal requirement in the United States.

For new (novel) excipients, drug companies need to carefully assess the benefits of using the new substance in consideration of the extra workload, costs, and possible regulatory delays or rejection. However, now that an active role for excipients is recognized in the form of sustained-release or drug-delivery systems, there is more need for companies to use new excipients. The replacement of CFCs with HFAs in inhalation therapy, although driven by necessary government demands, still involved extensive preclinical evaluation of the new excipient propellants and the usual regulatory approval times. Old CFC formulations with either HFA-134a or -227 as the new excipient are still being submitted and/or are under regulatory review and need to include bridging toxicity studies. The change from CFC has also meant modifications in MDI structure that, in turn, can result in the need for other formulation changes (and additional excipients). The publication of

the IPEC recommendations for safety evaluation of new excipients is helpful, although there is still no official response from the FDA. The suggested evaluation is not dissimilar to the full preclinical program necessary for a new chemical entity. If a drug company decides to introduce a new excipient, extensive preclinical data are probably necessary. However, if preliminary investigations, such as those given in Fig. 1, show that the material is safe at high dose levels, the excipient can enter the full drug development program as part of the formulation and/or in the form of an extra dose group. Expert evaluation of the data should identify any potential problems.

The situation for developing essentially new excipients is different inasmuch as some data on the material or a related material are available. Provided that these data are expertly evaluated and/or supported by some preliminary studies (as given in Fig. 1), along with the usual new drug studies, regulators should not raise disapproval. Each excipient, however, will be assessed on a case-by-case basis, with issues such as toxicological findings in the “old” data and duration of use of the excipient carefully assessed. As with new excipients, the increased workload and costs, along with regulatory approval time, will need to be considered. The cyclodextrins have entered pharmaceutical use by means of being present in food products. However, this process has taken many years and has involved extensive preclinical testing. Other materials such as saccharin have been established from food use, but they have had to undergo severe regulatory challenge over whether preclinical findings of bladder tumors are relevant in human use.

Established excipients pose some issues for regulatory consideration. Such issues include the use of a well known excipient by an “unapproved” route, for example, inhalation of a material that has been used in oral formulations. In these cases, some preliminary toxicity studies may be necessary. Additionally, a well designed scientific review of safety data on the excipient will be necessary, with old findings fully evaluated.

The acceptance that excipients are neither inert nor inactive substance and may affect drug bioavailability and cause adverse reactions (commonly hypersensitivity) means that even more consideration of these materials is needed when formulating new drugs. With these facts in mind, regulators will expend greater efforts to examine the ingredients in drug formulations. Furthermore, because there is no international requirement to provide details of drug formulations to health professionals (notably in the United States), adverse reactions will continue to be reported, putting extra pressure both on drug companies and

on regulators. Overall, the lack of regulatory guidance for excipient development leads to confusion for drug companies. In view of international agreement in other areas of drug development, an excipient-testing strategy would be an excellent topic for International Conference on Harmonization (ICH) consideration. Furthermore, with more standardization of pharmacopoeia materials, it may be possible to have excipients reviewed by a committee of an international pharmacopoeia. Safety data would be assessed by elected experts and published. Hopefully, these data would be acceptable to international regulatory bodies.

REFERENCES

1. *European Pharmacopoeia*, 3rd Ed.; European Pharmacopoeia Commission, Council of Europe: Strasbourg, France, 1997; 1299–1300.
2. *United States Pharmacopoeia*, 23rd Edn.; United States Pharmacopoeial Convention, Inc.: Rockville MD, 1995.
3. *Japanese Pharmacopoeia*, 13th Ed. The Society of Japanese Pharmacopoeia, Yakuji Nippo Ltd.: Tokyo, 1996.
4. *U.S. Physicians Desk Reference*, 50th Ed.; Medical Economics Company, Inc.: Montvale, NJ, 1996.
5. Kibbe, A.H. *Handbook of Pharmaceutical Excipients*, 3rd.; The American Pharmaceutical Association, Washington DC; The Pharmaceutical Press: London, 2000; 336.
6. Smolinske, S.C. *Handbook of Food, Drug and Cosmetic Excipients*, CRC Press: Boca Raton, FL, 1992.
7. Reynolds, J.E.F. *Martindale. The Extra Pharmacopoeia*, 31st Ed.; Royal Pharmaceutical Society: London, 1996.
8. *Inactive Ingredients Guide*; U.S. Food and Drug Administration, Division of Drug Information Resources: Rockville, MD, 1996.
9. *The National Formulary*, United States Pharmacopoeial Convention, Inc.: Rockville, MD, 1995.
10. *ABPI Compendium of Data Sheets and Summaries of Product Characteristics 1998–1999*; (For Irrigation with Peppermint Oil, See p.981) Datapharm Publications Ltd: Whitehall, London, 1998.
11. Wotton, P.K.; Khosla, R. Acceptability of Excipients in the European Community and USA. *BIRA J.* **1991**, 10 (7), 11–13.
12. *Rules Relating to Marketing Authorisation of Medicinal Products for Human Use and Related Council Directives 81/852/EEC, 91/507/EEC and 92/18/EEC plus Council Recommendations 83/571/EEC and 87/176/EEC*; Office for Official Publications of the European Communities: Luxembourg, 1998.
13. Medicinal Products for Human Use, Presentation and Content of the Dossier. *The Rules Governing Medicinal Products in the European Union Notice to Applicants*; Office for Official Publications of the European Communities: Luxembourg, 1998; 2B.
14. Medicinal Products for Human Use, Safety, Environment and Information. *The Rules Governing Medicinal Products in the European Union, Guidelines*; Office for Official Publications of the European Communities: Luxembourg, 1998; 3B.
15. *Note for Guidance: EC Application Format*; CPMP and DGIII Guideline III/3038/91, Commission of the European Communities: Brussels, adopted March 1992.
16. Steinberg, M.; Borzelleca, J.F.; Enters, E.K.; Kinoshita, F.K.; Loper, A.; Mitchell, D.B.; Tamulinas, C.B.; Weiner, M.L. A New Approach to the Safety Assessment of Pharmaceutical Excipients. *Regul. Toxicol. Pharm.* **1996**, 24, 149–154.
17. Center for Drug Evaluation and Research (CDER). Guidance for Industry, Guidelines for the Format and Content of the Nonclinical Pharmacology/Toxicology Section of an Application. FDA: Rockville, MD, 1987.
18. *Japanese Technical Requirements for New Drug Registration*; Yakuji Nippo Ltd.: Tokyo, 1997.
19. Uchiyama, M. Regulatory Status of Excipients in Japan. *Drug Inform. J.* **1999**, 33, 27–32.
20. *British Institute of Regulatory Affairs BIRA News*; Institute Update 84, 1992.
21. *Note for Guidance: Summary of the Product Characteristics*; Adopted October 1991 Commission of the European Communities, Brussels, adopted October 1991.
22. Japanese Ministry of Health and Welfare. *Notice on the Indication of the Inactive Ingredients in Prescription Drugs*; Pharmaceutical Affairs Bureau, Ministry of Health and Welfare, Yakuji Nippo Ltd: Tokyo, 1988.
23. American Academy of Pediatrics, Committee on Drugs. “Inactive” Ingredients in Pharmaceutical Products: Update. *Pediatrics* **1997**, 99 (2), 268–278.
24. Wong, Y.L. Adverse Effects of Pharmaceutical Excipients in Drug Therapy. *Ann. Acad. Med. Singapore* **1993**, 22 (1), 99–102.
25. Specific Requirements on Content and Format of Labelling for Human Prescription Drugs; Revision of “Pediatric” Subsection in the Labelling. Federal Register: , 1994.
26. *Note for Guidance: Replacement of Chlorofluorocarbons (CFCs) in Metered Dose Inhalation Products*, CPMP and DGIII Guideline III/537/93, Commission of the European Communities, Brussels, adopted December 1993.
27. *Matters Relating to the Replacement of CFCs in Medicinal Products*, CPMP and DGIII Guideline III/5462/93, Commission of the European Communities, Brussels, adopted December 1993.
28. *Note for Guidance on Preclinical Pharmacological and Toxicological Testing of Vaccines*, CPMP/SWP/465/95, EMEA: Canary Wharf, London, adopted December 1997.
29. *Note for Guidance on Development Pharmaceuticals*, CPMP/QWP/155/96, EMEA, Canary Wharf, London, adopted January 1998.
30. Shangraw, R. Developments in Tablet Excipients Since 1960. *Manufact. Chem.* **1986**, (December), 22–23.
31. Robertson, M.I. Regulatory Issues with Excipients. *Int. J. Pharm.* **1999**, 187, 273–276.
32. Strattan, C.E. 2-Hydroxypropyl- β -Cyclodextrin. I. Patents and Regulatory Issues. *Pharm. Technol. Int.* **1992**, 4 (4), 45–49.

33. Chowhan, Z.T. A Rational Approach to Setting Limit Tests and Standards on Impurities in Excipients. *Pharm. Tech. Europe* **1995**, 7 (10), 66–75.
34. Wolff, R.K.; Dorato, M.A. Toxicologic Testing of Inhaled Pharmaceutical Aerosols. *Crit. Rev. Toxicol.* **1993**, 23 (4), 343–369.
35. Maginley, R. The Transition to CFC-Free Metered Dose Inhalers (MDIs): Regulatory Process, Implications and Conundrum. *BIRA, Regul. Rev.* **1998**, 7, 11–18.
36. Bakshi, K.S. Toxicity of Alternatives to Chlorofluorocarbons: HFC-134a and HCFC-123. *Inhal. Toxicol.* **1998**, 10, 963–967.
37. Harrison, L.I.; Donnell, D.; Simmons, J.L.; Ekholm, B.P.; Cooper, K.M.; Wyld, P.J. Twenty-Eight-Day Double-Blind Safety Study of an HFA-134a Inhalation Aerosol System in Healthy Subjects. *J. Pharm. Pharmacol.* **1996**, 48 (6), 595–600.
38. Blumenthal, M.N.; Casale, T.B.; Fink, J.N.; Uryniak, T.; Casty, F.E. Evaluation of a Nonchlorofluorocarbon Formulation of Cromolyn Sodium (Intal) Metered-Dose Inhaler Versus the Chlorofluorocarbon Formulation in the Treatment of Adult Patients with Asthma: A Controlled Trial. *J. Allergy Clin. Immunol.* **1998**, 101 (1), 7–13.
39. De George; et al. Considerations for Toxicology Studies of Respiratory Drug Products. *Reg. Toxicol. Pharmacol.* **1997**, 25 (2), 189–193.
40. Golightly, L.K.; Smolinske, S.S.; Bennett, M.L.; Sutherland, E.W., III; Rumack, B.H. Pharmaceutical Excipients. II. Adverse Effects Associated with “Inactive” Ingredients in Drug Products. *Med. Toxicol.* **1998**, 3, 209–240.
41. Kinghorn, A.D.; Kaneda, N.; Baek, N.I.; Kennelly, E.J.; Soejarto, D.D. Noncariogenic Intense Natural Sweeteners. *Med. Res. Rev.* **1998**, 18 (5), 347–360.
42. Herbert, K. Novel Excipients for Sugar-Free Preparations—Functionality and Performance. *European Pharmaceutical Contractor*, Ballantyne Ross Ltd: London, 1998; 132–139.
43. Bär, A. Sugars and Adrenomedullary Proliferative Lesions: The Effects of Lactose and Various Polyols. *J. Am. Coll. Toxicol.* **1988**, 7 (1), 71–81.
44. Roe, F.J.C. Relevance for Man of the Effects of Lactose, Polyols and Other Carbohydrates on Calcium Metabolism Seen in Rats: A Review. *Hum. Toxicol.* **1989**, 8, 87–98.
45. Lynch, B.S.; Tischler, A.S.; Capen, C.; Munro, I.C.; McGirr, L.M.; McClain, R.M. Low Digestible Carbohydrates (Polyols and Lactose): Significance of Adrenal Medullary Proliferative Lesions in the Rat. *Regul. Toxicol. Pharmacol.* **1996**, 23, 256–297.
46. Bär, A. Significance of Leydig Cell Neoplasia in Rats Fed Lactitol or Lactose. *J. Am. Coll. Toxicol.* **1992**, 11 (2), 189–207.
47. Cohen, S.M. Human Relevance of Animal Carcinogenicity Studies. *Regul. Toxicol. Pharmacol.* **1995**, 21 (1), 75–80.
48. Cohen, S.M. Cell Proliferation and Carcinogenesis. *Drug Metab. Rev.* **1998**, 30 (2), 339–357.
49. Wysner, J.; Williams, G.M. Saccharin Mechanistic Data and Risk Assessment: Urine Composition, Enhanced Cell Proliferation, and Tumour Promotion. *Pharmacol. Ther.* **1996**, 71 (1–2), 225–252.
50. Robinson, B.V.; Sullivan, F.M.; Borzelleca, J.F.; Schwartz, S.L. *PVP: A Critical Review of the Kinetics and Toxicology of Polyvinylpyrrolidone (Povidone)*, Lewis Publishers, Inc.: Chelsea MI, 1990.
51. Frommer, J. The Pathogenesis of Reticulo-Endothelial Foam Cells. Effect of Polyvinylpyrrolidone on the Liver of the Mouse. *Am. J. Path.* **1956**, 32, 433–453.
52. Nelson, A.A.; Lusky, L.M. Pathological Changes in Rabbits from Repeated Intravenous Injections of Periston (Polyvinyl Pyrrolidone) or Dextran. *Proc. Soc. Exp. Biol. Med.* **1951**, 76, 765–767.
53. Sharp, D.W. The Sensitisation Potential of Some Perfume Ingredients Tested Using a Modified Draize Procedure. *Toxicology* **1978**, 9, 261–271.
54. Alden, C.L. A Review of Unique Male Rat Hydrocarbon Nephropathy. *Toxicol. Pathol.* **1986**, 14 (1), 109–111.
55. Lehman-McKeeman, L.D.; Rodriguez, P.A.; Takigiku, R.; Caudill, D.; Fey, M.L. D-Limonene-Induced Male Rat-Specific Nephrotoxicity: Evaluation of the Association Between D-Limonene And $\alpha_2\mu$ -Globulin. *Toxicol. Appl. Pharmacol.* **1989**, 99, 250–259.
56. Flamm, W.G.; Lehman-McKeeman, L.D. The Human Relevance of the Renal Tumor-Inducing Potential of D-Limonene in Male Rats: Implications for Risk Assessment. *Regul. Toxicol. Pharmacol.* **1991**, 13, 70–86.
57. Goodman, J.I. An Analysis of the National Toxicology Program's (NTP) Technical Report (NTP TR 421) on the Toxicology and Carcinogenesis Studies of Talc. *Regul. Toxicol. Pharmacol.* **1995**, 21 (2), 244–249.
58. Baldrick, P.; Bamford, D.G. A Toxicological Review of Lactose to Support Clinical Administration by Inhalation. *Food Chem. Toxicol.* **1997**, 35, 719–733.
59. Gopinath, C.; Prentice, D.E.; Lewis, D.L. *Atlas of Experimental Toxicological Pathology*; Gresham, G., Ed.; MTP Press Ltd.: Lancaster, 1987; 24–42.
60. Burger, G.T.; Renne, R.A.; Sagartz, J.W.; Ayres, P.H.; Coggins, C.R.E.; Mosberg, A.T.; Hayes, A.W. Histologic Changes in the Respiratory Tract Induced by Inhalation of Xenobiotics: Physiologic Adaptation or Toxicity? *Toxicol. Appl. Pharmacol.* **1989**, 101, 521–542.
61. Gaworski, C.L.; Dozier, M.M.; Gerhart, J.M.; Rajendran, N.; Brennecke, L.H.; Aranyi, C.; Heck, J.D. Thirteenth-Week Inhalation Toxicity Study of Menthol Cigarette Smoke. *Food Chem. Toxicol.* **1997**, 35 (7), 683–92.
62. Thompson, D.O. Cyclodextrins-Enabling Excipients: Their Present and Future Use in Pharmaceuticals in Drug Formulations. *Crit. Rev. Ther. Drug Carrier Syst.* **1997**, 14 (1), 1–104.
63. Adeyeye, C.M.; Barabas, E. Povidone. *Analytical Profiles of Drug Substances and Excipients*; Brittain, H.G., Ed.; Academic Press: San Diego, 1993; 22, 555–683.
64. Renwick, A.G. Saccharin: A Toxicological Evaluation. *Comments Toxicol.* **1998**, 3 (4), 289–305.
65. Rowe, V.K.; Wolf, M.A. *Patty's Industrial Hygiene and Toxicology*; Clayton, G.D., Clayton, F.E., Eds.; Interscience: New York, 1982; 3844–3852.

66. Alexander, D.J.; Libretto, S.E. An Overview of the Toxicology of HFA-134a (1,1,1,2-Tetrafluoroethane). *Hum. Exp. Toxicol.* **1995**, *14* (9), 715–720.
67. Vogel, F.R.; Powell, M.F. A Compendium of Vaccine Adjuvant and Excipients. *Pharm. Biotechnol.* **1995**, *6*, 141–228.
68. Napke, E.; Stevens, D.G.H. Excipients and Additives: Hidden Hazards in Drug Products and in Product Substitution. *Can. Med. Assoc. J.* **1984**, *131*, 1449–1452.
69. Golightly, L.K.; Smolinske, S.S.; Bennett, M.L.; Sutherland, E.W., III; Rumack, B.H. Pharmaceutical Excipients. I. Adverse Effects Associated with Inactive Ingredients in Drug Products. *Med. Toxicol.* **1988**, *3*, 128–165.
70. Barband, A. Place of Excipients in Drug-Related Allergy. *Clin. Rev. Allergy. Immunol.* **1995**, *13*, 253–263.
71. Gebhart, F. Drugs' Inactive Ingredients Drawing Government Attention. *Drug Top.* **1997**, *141*, 18S–20S.
72. Scott, A.W. Non-Medicinal Ingredients. *Drug Safety* **1990**, *5* (1), 95–100.
73. Johnston, K.R.; Govel, L.A.; Andritz, M. Gastrointestinal Effects of Sorbitol as an Additive in Liquid Medications. *Am. J. Med.* **1994**, *97* (2), 185–191.
74. Kumar, A.; Aitas, A.T.; Hunter, A.G.; Beaman, D.C. Sweeteners, Dyes and Other Excipients in Vitamin and Mineral Preparations. *Clin. Pediatr.* **1996**, *35* (9), 443–450.
75. Altimaras, J.; Nieto-Hernandez, T.; Buitrago, F. Presence of Excipients in Pharmaceutical Products in Three Sources of Therapeutic Information. *Aten. Primaria.* **1996**, *18* (4), 190–193.
76. Monkhouse, D.C.; Lach, J.L. Drug-Excipient Interactions. *Can. J. Pharm. Sci.* **1972**, *7* (2), 29–46.
77. Reimer, K.; Fleischer, W.; Brogmann, B.; Schreier, H.; Burkhard, P.; Lanzendorfer, A.; Gumbel, H.; Hoekstra, H. Povidone-Iodine Liposomes—An Overview. *Dermatology* **1997**, *195* (Suppl 2), 93–99.
78. Ishida, O.; Maruyama, K. Transferring Conjugated PEG-Liposomes as Intracellular Targeting Carrier for Tumor Therapy. *Nippon Rinsho* **1998**, *56* (3), 657–662.
79. Naidoo, R. Examining Excipients. *Pharm. Cosmet. Rev.* **1997**, *24*, 23 25–27.
80. Zhao, Z.; Leong, K.W. Controlled Delivery of Antigens and Adjuvants in Vaccine Development. *J. Pharm. Sci.* **1996**, *85*, 1261–1270.
81. Rubinstein, A.; Tirosh, B.; Baluom, M.; Nassar, T.; Friedman, M. Rationale for Peptide Drug Delivery to the Colon and the Potential of Polymeric Carriers as Effective Tools. *J. Controlled Rel.* **1997**, *46*, 59–73.
82. Medicines Order. *The Medicines (General Sale List) Order*, The Stationery Office Ltd.: London, 1997.

PHARMACEUTICAL QUALITY ASSURANCE MICROBIOLOGY

Anthony M. Cundell

Wyeth-Ayerst Pharmaceuticals, Pearl River, New York, U.S.A.

INTRODUCTION

The pharmaceutical microbiologist has an important role in product development, manufacturing process development, ensuring control of microorganisms in the manufacturing environment and routine raw material, inprocess material, and product testing. The involvement of experienced microbiologists in each stage of the product life cycle is important to maintain product quality. Typically, microbiologists are involved in formulation, manufacturing process development, and specification-setting decisions that can prevent microbial contamination of pharmaceutical products. In this chapter, the appropriate level of involvement of microbiologists in establishing the quality, purity, efficacy, and safety of pharmaceutical and over-the-counter drug products is discussed, and areas of future challenge to the pharmaceutical microbiologist are explored.

A review of the regulations governing the pharmaceutical industry outlines some of the formal responsibilities of the microbiologist. The U.S. Federal Regulations that govern the pharmaceutical industry are called Current Good Manufacturing Practices (cGMPs). The U.S. Federal Food and Drug Administration (FDA) was mandated by the 1962 New Drug Amendments to the U.S. Federal Food, Drug, and Cosmetics Act to promulgate regulations that had the force of law to ensure that drug manufacturers maintain the safety, identity, strength, quality, and purity of their products. These official regulations were published in the Federal Register, September 29, 1978, as 21 C.F.R. Parts 211 through 226 (1) as they pertain to drugs. The regulations are considered minimum requirements, and failure to comply with any part of the regulations during the manufacture, processing, packaging, or holding of a drug renders that product to be adulterated under section 501(a) (2) (B) of the Federal Food, Drug, and Cosmetics Act. The drug product as well as the persons who are responsible for the failure to comply with the regulations may be subject to regulatory action.

Sections of the cGMPs most pertinent to the pharmaceutical microbiologist include Subpart B Organization and Personnel: 211.22 Responsibilities of the

Quality Control Unit and 211.25 Personnel Qualifications; 211.113 Control of Microbiological Contamination; and Subpart I Laboratory Controls: 211.167 Special Testing.

The cGMP regulations require that a pharmaceutical manufacturer have an independent quality control unit responsible for approval or rejection of all components, drug product containers, closures, in-process materials, packaging materials, labels and drug products, review of production records for possible error, and investigation of manufacturing deviations. The quality control unit also reviews and approves all specifications and procedures impacting on the products and leads the investigation of manufacturing deviations and product failures.

The quality control unit must have access to an adequate testing laboratory to aid in the approval of the materials under its control. One of these laboratory facilities would be suitably equipped and staffed to conduct microbiological testing. The quality control unit need not manage the microbiology laboratory. The laboratory could be run by quality control, research and development, even manufacturing or could be a contract testing laboratory, provided it meets cGMP requirements and is responsive to the needs of the quality control unit. Although not a regulatory requirement, it is industry practice to use the audit process as a tool to ensure that the microbiology laboratory meets all regulatory requirements and internal company policies.

The 211.25 Personnel Qualifications requirement is that all persons engaged in all phases of pharmaceutical manufacture have the education, training, and experience to enable them to do their job and have a working knowledge of the cGMP regulations that applies equally to laboratory personnel. For example, bench microbiologists should have a bachelor of microbiology or allied life sciences degree and be adequately trained in the laboratory procedures and testing documentation conducted in the microbiology laboratory. They need to be assigned responsibilities in keeping with their level of skill and experience. Microbiologists with supervisory or managerial responsibilities need training in supervisory skills, scheduling, budgeting, laboratory investigations, technical report writing, pharmaceutical products, and the manufacturing processes. They need to understand the requirements of the quality control unit and ensure that

the unit is supplied with quality test results in a timely and cost-effective manner. Since the pressures to manufacture, test, and release products in a timely manner can be considerable, they need to work well under pressure and enjoy team work.

The educational background of bench microbiologists, supervisors, and managers is now even more important, given, the current transition from classic to nucleic acid-based testing methods.

The demands of microbiological testing require that the core educational background of the staff, supervisors, and managers be in microbiology. Training and experience in aseptic techniques are necessary. According to the author, the skill sets of chemists, pharmacists, and even biologists do not allow them to readily act effectively as microbiologists without extensive training. Course work invaluable to the pharmaceutical microbiologist includes:

- Isolation, enumeration, and identification of bacteria and fungi;
- Pathogenic microbiology;
- Microbial physiology and biochemistry;
- Introductory chemistry including organic, inorganic, and physical chemistry and quantitative analysis;
- Introductory physics;
- Introductory mathematics;
- Introductory pharmaceutical manufacturing;
- Introductory statistics and probability; and
- Written and oral expression with emphasis on technical report writing.

According to 211.113, Control of Microbiological Contamination, pharmaceutical manufacturers need written procedures describing the systems designed to prevent objectionable microorganisms in both nonsterile and sterile drug products. All sterilization processes used to manufacture parenteral drugs need to be validated.

There needs to be a laboratory test for each batch of drug product to determine that the product conforms to specification, including the identity and strength of each active ingredient, before release. Where sterility and/or pyrogen testing are conducted on specific batches of short-lived radiopharmaceuticals, such batches may be released before completion of this testing, provided such testing is completed as soon as possible. The 211.165 Testing and Release for Distribution regulation states that there be appropriate laboratory testing, as necessary, of each batch of drug product required to be free of objectionable microorganisms. This implies that each and every batch of product need not undergo microbial evaluation.

The pharmaceutical microbiologist in product development plays a major role in bring safe products to the market. Typically, the Research & Development (R&D)

microbiologist is found in the analytical development group. The role of the R&D microbiologist is to develop and validate the microbial tests that may be applied to the new pharmaceutical products to confirm that they are not contaminated with an excessive number of microorganisms or objectionable microorganisms that may infect patients or degrade the quality of the product during its shelf life.

These tests include compendial tests for microbial limits, sterility, bacterial endotoxins, and antimicrobial effectiveness.

These compendial tests need to be developed and qualified for each new products before the phase II clinical trials, which involve 100–200 subjects. Considering the large investment to bring a new pharmaceutical drug product to market, i.e., an estimated U.S. \$ 250–500 million, it is important not to jeopardize the future of a new product or the subjects in the clinical trial by administering a drug product that may be contaminated with objectionable microorganisms. However, for the results of microbial testing to have any meaning, the microbial test needs to be qualified as suitable for use with each product.

Many dosage forms have active pharmaceutical ingredients or contain preservative systems that may inhibit the recovery of any bioburden associated with the product. This inhibition may be overcome by inactivating the active ingredient or preservative system with neutralizing agents, dilution to overcome the inhibition, or a combination of both strategies. U.S. Pharmacopoeia chapters 51, Antimicrobial Effectiveness Test; 61 Microbial Limit Tests; and 71 Sterility Testing (2) contain specific instructions on how to qualify the test for use with specific pharmaceutical drug products and the microorganisms to use during this process.

A major concern for Pharmaceutical Operations, i.e., Materials Management, Manufacturing, and QA, is whether the R&D microbiologists develop microbial tests that meet the QA requirements of robustness, simplicity, and standard for ease of testing for the routine release testing of pharmaceutical products. Good communications between the two microbiology groups will ensure the smooth technology transfer of the most appropriate tests from R&D to QA.

During development of the manufacturing process, an experienced microbiologist should be consulted as to the potential for microbial contamination of the product. Issues may include the selection of appropriate pharmaceutical ingredients, the ability of the manufacturing steps to control microbial contamination, the validation of sterilization processes, the cleaning and sanitization of process equipment, the adequacy of the water system, the

holding times for intermediates, the training of personnel, and the design of the packaging.

CURRENT MICROBIOLOGICAL TESTING PRACTICES

The compendial microbial methods currently used for the routine testing of pharmaceutical products are generally conservative and may be used to referee disputes concerning the microbial contamination of pharmaceutical products. The USP is recognized as an official compendium by the U.S. Federal Food, Drug and Cosmetic (FDC) Act. USP standards are used to determine the identity, strength, quality, and purity of pharmaceutical articles. The “General Chapters” section of the USP includes requirements for tests and assays and is numbered from 1 to 999, whereas general information include chapters 1000 and above. Of particular interest to the pharmaceutical microbiologist are USP 24 Informational Chapters 1116, Microbiological Evaluation of Clean Rooms and other Controlled Environments; 1111 Microbiological Attributes of Pharmaceutical Articles; 1225, Validation of Compendial Methods; and 1231, Water for Pharmaceutical Purposes (3). Testing chapters pertinent to pharmaceutical microbiology and their JP and Ph. Eur. counterparts are as follows:

USP Chapter 51, Antimicrobial Effectiveness Test

Antimicrobial preservatives are substances added to multiuse nonsterile liquids, ointments, and creams and sterile injectable products to protect them from microbial contamination that may be introduced inadvertently during use of the product (postmanufacturing).

The test for antimicrobial effectiveness is used to demonstrate the effectiveness of any added antimicrobial preservative(s). Compendial references include USP 24 Chapter 51, Antimicrobial Effectiveness Test; JP XIII, General Information 3, Preservatives-Effectiveness Tests; and Ph. Eur. 3rd Ed., Biological Tests, 5.1.3., Efficacy of Antimicrobial Preservation.

USP Chapter 61, Microbial Limits Tests

The tests for microbial limits and recommendations for microbial quality criteria of raw materials, excipients, drug substances, and pharmaceutical products have been established in pharmacopoeial compendia for over 30 years. These tests are listed in USP 24, Chapter 61, Microbial Limits Tests; Ph. Eur. 3rd Ed., Biological Tests

2.6.12 and 2.6.13, Microbial Contamination of Products not Required to Comply with the Test for Sterility (Total Viable Count, Tests for Specified Micro-Organisms); and JP XIII 30, Microbial Limit Test.

USP Chapter 71, Sterility Test

The sterility test is applicable for determining whether drug substances, preparations, or other pharmacopeial articles are sterile as defined by the compendial method. A satisfactory result indicates only that no contaminating microorganisms have been found in the sample examined according to the conditions of the test. Therefore, the result is a function of the efficiency of the adopted sampling plan. Compendial references to sterility testing include USP 24, Chapter 71, Sterility Tests; Ph. Eur. 3rd Ed., Biological Tests 2.6.1., Sterility; and JP XIII 45, Sterility Test.

General Informational Chapter 1116, Microbiological Evaluation of Clean Rooms and Other Controlled Environments

The microbiological monitoring of air, surfaces, and personnel in facilities used for sterile pharmaceutical manufacturing is discussed in the USP 24 Informational Chapter 1116, Microbiological Evaluation of Clean Rooms and Other Controlled Environments. The chapter also covers the design and implementation of a microbiological monitoring program and suggests monitoring frequencies and microbiological acceptance criteria.

USP Informational Chapter 1231, Water for Pharmaceutical Purposes

Types of water and methods and specifications for testing them are listed in USP Informational Chapter 1231, Water for Pharmaceutical Purposes. The USP also references Standard Methods for the Examination of Water and Waste Water (APHA), 19th Ed., for information on specific test methods (4).

Other Testing Methods

New microbial testing methods are being introduced in the marketplace based on advanced technologies. These new tests represent improvements in the timeliness and quality of testing.

The USP 24 General Notices states that alternative methods may be used to determine that products comply with the pharmacopeial standards for advantages in

accuracy, sensitivity, precision, selectivity, and adaptability to automation or computerized data reduction or for any other special circumstances. Such alternative or automated methods must be validated; however, when disputed, the compendial method is conclusive because it is the official or referee test. In addition, USP Chapter 61, Microbial Limit Tests, states that automated methods may be substituted, provided they are validated and give equivalent or better results, whereas USP Chapter 71, Sterility Tests, states that alternative procedures may be used to demonstrate that an article is sterile, provided the results obtained are at least of equivalent reliability.

It is not required to have prior FDA approval to use an alternate method to a compendial test. According to 21 CFR 314.70, "Supplements and Other Changes to an Approved Application," the addition or deletion of an alternate analytical method does not require prior approval and may be filed in the Annual Product Report. However, we would need to document the equivalency of the alternate method to the regulatory or compendial test method and the validation report must be available for an FDA investigator to inspect at our manufacturing site. Where the test method is particularly novel it may be advisable to include the test in an NDA supplement so the FDA can review the new method and your company can get prior FDA approval before the new test method is implemented.

ORGANIZATION OF THE PHARMACEUTICAL QUALITY ASSURANCE MICROBIOLOGICAL TESTING LABORATORY

The microbiology manager has three major roles in the QA organization. They are: 1) establishing, staffing, and running the microbiological testing laboratory; 2) monitoring pharmaceutical ingredients, water for pharmaceutical purposes, the manufacturing environment, and finished products submitted to the laboratory to demonstrate control of microbial contamination of the pharmaceutical products manufacturing at the site; and 3) providing microbiological expertise to the QA organization to prevent microbial contamination.

The microbiology manager is responsible for the establishment of a suitably constructed and equipped laboratory; recruiting and retaining an appropriately educated, skilled, and experienced staff; and operating the laboratory in compliance with all company policies and cGMP regulations. The needs of business require that microbial testing be conducted in a timely manner so that products can be released to the market.

The microbiological monitoring program established at a pharmaceutical manufacturing site will depend on the range of products manufactured there. Typical microbiological monitoring programs for release testing and environmental monitoring for nonsterile and sterile product manufacturing sites are given in Table 1.

The general procedures to be followed when selecting a microbial testing strategy a Marketed Product Stability Program for pharmaceutical drug products based on cGMP and compendial requirements and commitments made in regulatory filings are tabulated.

These approaches are in general accord with C.F.R. 21, Parts 211.113 Control of Microbiological Contamination, Section (b); 211.137, Expiration dating; 211.167, Special Testing Requirements Section (a); U.S.P. 24 51, Antimicrobial Effectiveness Testing, 61, Microbial Limit Tests, 71, Sterility Tests, 85, Bacterial Endotoxin Tests, General Informational Chapter 1151, Pharmaceutical Dosage Forms and the June, 1998, Draft FDA Stability Testing of Drug Substances and Drug Products guidance document (5). The tests that may be included in the program include: 1) antimicrobial effectiveness testing, 2) Microbial Limit Testing, 3) Sterility Testing, 4) Bacterial Endotoxin Testing, and 5) Container-closure Integrity Testing.

Table 2 outlines a possible testing policy.

Antimicrobial Effectiveness Testing

The following principles apply to preservative effectiveness testing in a pre- and postmarketed product stability program.

1. The selection of the preservative system for multiuse new products is the responsibility of the R&D formulation group. Typical shelf specifications are 80 to 120% LS. The appropriate preservative system for the particular formulation should be demonstrated to be effective by microbial challenge down to at least 75% and preferably to 50% of the target concentration. It is recommended that during development, the product be formulated with preservative concentrations of 100, 75, and 50% of the labeled amount and be subjected to Antimicrobial Effectiveness Testing to determine the lowest effective preservative concentration.
2. The release and shelf-life specifications are established based on both the premarketed stability data for the preservative system concentration and the Antimicrobial Effectiveness Test results.
3. If the study outlined in item 1 was not conducted during product development, it is recommended that QA and/or the Technical Services groups undertake a study

Table 1 Sample microbiological monitoring programs

Dosage form	Monitoring	Frequency
Tablets and capsules	Pharmaceutical ingredients	Periodic after history is established; rule accept on supplier certificate of analysis
	Purified water Manufacturing environment Products	Loop daily and taps weekly Quarterly Periodic after history is established owing to low water activities of tablets and capsules
Topicals, otics, vaginal and rectal products	Pharmaceutical ingredients	As above
	Purified water Manufacturing environment Products	Loop daily and taps weekly Weekly or monthly Routine for products with high water activity; periodic after history is established for product with low water activity
Nasal sprays and inhalants	Pharmaceutical ingredients	As above
	Purified water Manufacturing environment Products	Loop daily and taps weekly Daily or weekly Routine for products with high water activity; periodic after history is established for product with low water activity
Injectable products, ophthalmic products, and inhalation solutions	Pharmaceutical ingredients	As above
	Purified water Manufacturing environment Products	Loop and taps daily Every shift in critical aseptic processing areas Every batch with the exception of terminally sterilized products approved for parametric release

Table 2 Sample testing policy

Dosage form	Microbial test	Testing plan	Test intervals^a
Tablets, powder-, and liquid-filled capsules	Microbial Limit Test (TAMC and TCYMC only)	Test development, scale-up, and validation batches only	0, 6, 12, 24, and 36 months.
Topical liquids, ointment, and creams	Microbial Limit Test (TAMC and TCYMC only); USP Antimicrobial Effectiveness Test (AET) for multi use products	Aw <0.75 test development, scale-up and validation batches only; Aw >0.75 all batches on stability for Microbial Limit and first three batches for Antimicrobial Effectiveness (AE) ^b	0, 6, 12, 24, and 36 months; 0, middle of stability period and expiry
Vaginal creams and suppositories	Microbial Limit Test (TAMC and TCYMC only) and AET	As above	0, 6, 12, 24, and 36 months
Rectal creams and suppositories	As above	As above	0, 6, 12, 24, and 36 months
Nasal sprays	As above	As above	0, 6, 12, 24, and 36 months
Inhalation sprays and aerosols	As above	As above	0, 6, 12, 24, and 36 months
Ophthalmic ointments and solutions	Sterility Test, Container-Closure Integrity (CCI) and AET	Test all batches on stability with the exception of the first three batches for AE ^b	0, 12, 24, and 36 months
Injectables and Inhalation solutions	Sterility Test; CCI and AET; Bacterial Endotoxin Test	Test all batches on stability with the exception of the first three batches for AE ^b	0 only; 0, 12, 24, and 36 months; 0 and expiry only

^aTime intervals suggested in the 1998 Draft FDA Stability Guide. Add additional annual test intervals if the expiration dating exceeds 36 months.^bJustify using a stability-indicating preservative assay only at all time intervals as a substitute for the USP Antimicrobial Effectiveness Test by confirming preservative efficacy at 50, 75, and 100% of label claim.

to justify the current specifications and the elimination of routine Antimicrobial Effectiveness Testing in the Stability Program.

4. All preservative systems for both parenteral and nonsterile dosage forms should meet the 3 log reduction at 14 days for bacteria, i.e., USP category I requirements. EP/BP Antimicrobial Effectiveness Testing would be run only if requested by the Marketing Group.
5. Preservative effectiveness testing should be included at the 3-, 12-, 24- and 36-month intervals for pilot validation batches or of the first three commercial batches of a new product only.
6. With subsequent batches, chemical assays would be used only to confirm the preservative level, because the effectiveness during shelf life is being demonstrated. If the formulation is changed, the preservative effectiveness must be verified with at least one batch throughout the shelf life.
7. The choice of additional challenge organism used in formulation development will be determined by 1) the range of activity of the preservative system, i.e., if a preservative system has reduced activity against *Pseudomonas* spp., additional organisms from these genera or related genera could be added to the challenge organisms; 2) the organisms considered objectionable for that product and dosage form; and 3) the frequency of isolation of organisms from the manufacturing environmental and product monitoring.
8. Repeat challenges should be limited to the evaluation of preservative systems that cannot be improved because of formulation difficulties and limitation owing to the intended site of use and products that may be misused during multiple consumer use.

Microbial Limit Test

The following principles apply to microbial limit testing within a pre- and post marketed product stability program.

1. The inclusion of a routine Microbial Limit Test in a marketed product stability protocol will depend on the pharmaceutical dosage form. Typically, the test would be used for only nonsterile products, particularly oral liquids, nasal sprays; and topical liquids, lotions and creams that have a sufficient water activity to support the growth of microorganisms. In contrast, tablets, powder and liquid-filled capsules, topical ointments, vaginal and rectal suppositories, nonaqueous liquids, and inhalation aerosols with a water activity too low to allow for the product to support the growth of microorganisms would not be routinely tested.

2. To establish a Microbial Limit Testing history, all development, clinical, scale-up, and process validation batches of new nonsterile dosage forms would be tested to verify that the pharmaceutical ingredients, manufacturing process, and packaging do not contribute the bioburden of the product. After the testing history has been established, products with a water activity below 0.75 should not include Microbial Limit Testing in the stability protocol.

Sterility Testing

The following principles apply to sterility testing within a pre- and postmarketed product stability program.

1. All injectable and ophthalmic products with the exception of terminally sterilized product subject to parametric release should undergo Sterility Testing at release.
2. Because the sterility assurance of an injectable or ophthalmic product is established through media fill or sterilization validation for aseptically filled and terminally sterilized products, respectively, Sterility Testing has been included in past stability protocols as a measure of container-closure integrity of the product throughout its shelf life. If there is a continued need because of previous regulatory commitments to include Sterility Testing in a protocol, then testing at release and expiry is recommended.
3. Whenever possible, Container-Closure Integrity Testing should be substituted for Sterility Testing as recommended in the draft FDA Stability Guide.

Bacterial Endotoxin Test

The following principles apply to Endotoxin Testing in a pre- and postmarketed product stability program.

1. All injectable products should be tested for endotoxin at release.
2. Because, in the absence of bacterial growth in the product, the endotoxin level will not increase on storage during shelf life Bacterial Endotoxin Testing is not indicated in the stability protocol. If there is a continued need to include Endotoxin Testing in a protocol, then testing at release and at expiry is recommended

Container-Closure Integrity Test

The following principles apply to integrity testing a pre- and postmarketed product stability program.

1. The integrity of the container-closure system as a microbial barrier should be assessed using an appropriately sensitive and adequately validated Container-closure Integrity Test.
2. One of a number of physical Container-closure Integrity tests may be selected and validated against the Bacterial Liquid Immersion Test. The Physical Leak Test should be correlated to bacterial ingress.
3. The selection of the physical Container-closure Integrity Test method should be made after consideration of the container-closure type, the performance criteria, and the available validated test methods.
4. Test methods described in the literature include bubble, helium mass spectrometry, liquid trace (dye), head space analysis, vacuum/pressure decay, weight loss/gain, and high-voltage leak detection (6).
5. The number of samples tested should reflect the sampling requirements provided in the USP 71 Sterility Tests.
6. The testing should be performed annually and at expiry.

The third role of supplying microbiological expertise to the Manufacturing and Quality Assurance is important because depth of experience in microbiology may be lacking in these organizations in some companies.

Typically, the management of these organizations is trained in chemistry, engineering, business, or pharmacy. Microbiologists should assist management to exercise the best judgment on microbiological issues. Given that the ultimate objective is to prevent microbial contamination of pharmaceutical and OTC drug products, it is important that pharmaceutical microbiologists be knowledgeable in the areas of microbiological testing, infectious diseases, compendial changes, regulatory issues, product formulation, and manufacturing processes so they can give credible advice.

A common practice in microbiological testing is that pharmaceutical ingredients and products are tested without full consideration of their significance. Sometimes, all raw materials purchased and product manufactured are submitted to the microbiology laboratory and tested, or if materials for testing are selected, insufficient judgment is made with respect to where materials are tested, i.e., materials with a low risk of microbial contamination are tested whereas materials with a high risk are not tested. Testing should always reflect the risk of microbial contamination. An important managerial tool to rationalize the microbial testing is the reduced testing program. An important aspect of a reduced pharmaceutical ingredient monitoring program, after supplier audit and an evaluation of the equivalency

of results from the supplier's certificate of analysis and the manufacturing site microbiological testing laboratory, is an understanding of the potential risk of the microbial contamination of a pharmaceutical ingredient, manufacturing environment, or pharmaceutical product. The microbiologist needs to make a judgment based on the source of the ingredient, how it is processed, its water activity and testing history, and how it is used in formulations to determine whether periodic microbial monitoring to ensure that the testing laboratory confirms the microbial results reported on the supplier certificate of analysis is justified.

IMPLEMENTATION OF NEW MICROBIOLOGICAL TESTING METHODS

Opportunities exist to implement new microbiological testing methods as alternatives to the compendial methods to improve the quality of the test results and reduce the product-release cycle time. Selection of candidate test methods, proof of concept studies, assay development and validation, regulatory approval, and implementation of the new microbiological testing methods are major issues that need to be addressed to take advantage of the new technologies. These new methodologies offer significant improvements in terms of the speed, accuracy, precision, specificity, etc., with which testing can be performed.

The majority of testing performed today relies on century-old methods based on the recovery and growth of microorganisms using solid or liquid microbiological growth media. This is true in part because these methods can be very effective and have a long history of application. However, they are often limited by slow microbial growth rates, unintended selectivity for microorganisms that grow in nutrient-rich culture media, and the inherent variability of microorganisms in their response to culture methods. Despite the limitations of current methods, acceptance of new and potentially superior methods is often slow because of the understandable conservative tendency of microbiologists.

This may be in part attributable to a lack of clear guidance regarding the demonstration of their equivalence to existing methods acceptable to regulatory agencies and validation of the equipment associated with the new methods.

Considerable guidance can be found regarding the validation of chemical methods that is applicable to microbial testing. Examples include USP Chapter 1225, "Validation of Compendial Methods," and a recent

publication by the International Conference on Harmonization (ICH), "Validation of Analytical Methods." These publications provide very specific instructions regarding the demonstration of new analytical chemistry methods and their equivalence to existing methods.

When instrumentation is developed for existing microbiological methods to automate sample handling, result reading, or data management, it is not difficult to demonstrate the equivalency of the alternate method using guidelines developed for chemical assays because the test remains essentially the same. In a similar manner, when a new technology continues to rely on the measurement of microbial growth (e.g., impedance, ATP bioluminescence, or other metabolic changes in a microbial culture), equivalence can be readily demonstrated. However, when a new method is based on novel technology without direct ties to the existing method (e.g., microbial identification by rRNA amplification versus patterns of biochemical reactions, or counting fluorescent-labeled bacterial cells instead of colony-forming units on an agar plate), demonstration of equivalency may require a new application of the validation principles, although the method provides higher quality results. The principles that can be applied to the validation of new microbiological testing methods are found in the recently published PDA Technical Report (7).

For convenience, the technologies are divided into growth-based technologies, viability-based technologies, cellular component or artifact-based technologies, and nucleic acid-based technologies as shown in Table 3.

IMPLEMENTATION OF A RISK-BASED MICROBIOLOGICAL TESTING PROGRAM

A testing program to be both cost-effective and to control microbial contamination must reflect the potential risks of microbial contamination of pharmaceutical drug products. A knowledge of product formulation, manufacturing processes, packaging, and ability of product to support microbial growth can be applied to develop rational specifications and a monitoring program that reflects the potential risk to the consumer of each dosage form. This emphasis on potential risk will require that the pharmaceutical microbiologist most closely addresses products having a higher potential for microbial contamination to best serve the needs of pharmaceutical companies and end-user of products.

A recall is a removal or correction of a marketed product by the pharmaceutical manufacturer when that product violates the laws enforced by the FDA. Unlike the FDA's other methods for achieving compliance such as seizures and court-ordered injunctions, recalls are almost always voluntary. The FDA cannot order a company to recall a product, except in some cases involving infant formulas, biological products, and devices that present a serious health hazard. A class I recall occurs when there is a reasonable probability that the use of or exposure to a violative product will cause serious adverse health consequences or death. A class II recall occurs when use of or exposure to a violative

Table 3 Classification of new biological testing methods

Testing method	Example technologies
Growth-based technologies	ATP bioluminescence Impedance/conductivity Hydrophobic grid membrane filter methods
Viability-based technologies	Direct epifluorescent filter Microscopy membrane laser Scanning fluorescence cytometry Fluorescence flow cytometry
Cellular component or artifact-based technologies	GL chromatographic fatty acid profiles MALDI-TOF mass spectrometry Fluorescence antibody techniques Enzyme-linked immunosorbent assay <i>Limulus</i> amoebocyte lysate-endotoxin assay
Nucleic acid-based technologies	Nucleic acid probe Polymerase chain reaction–DNA amplification 16S rRNA sequencing techniques Automated riboprinting

Table 4 Summary of the nonsterile pharmaceutical and OTC products recalled, 1991–1998, by the FDA because of microbial contamination problems as to class of recall ($n = 46$)

Year	Recalls	Class I	Class II	Class III	% <i>Pseudomonads</i>
1998	7	2	3	2	57
1997	5	1	2	2	60
1996	4	1	2	1	75
1995	4	0	3	1	0
1994	8	1	4	3	38
1993	9	0	6	3	44
1992	6	1	5	0	33
1991	3	0	0	3	0
1991–98	46	6	25	15	38

(From Ref. 8.)

product may cause temporary or medically reversible adverse health consequences or in which the probability of serious adverse health consequences is remote. A class III recall occurs when use of or exposure to a violative product is not likely to cause adverse health consequences. When the center receives the Recall Report from the FDA district office, it evaluates the health hazard presented by the product and categorizes it as class I, II, or III. The classification is determined by an ad hoc Health Hazard Evaluation Committee made up of FDA scientists chosen for their expertise. Classification is done on a case-by-case basis, considering the potential consequences of a violation.

The average number of recalls per annum for microbial contamination of nonsterile pharmaceutical and OTC drug products is six (Table 4). The emphasis on waterborne Gram-negative bacteria of the species *Burkholderia* (*Pseudomonas*) *cepacia* (nine recalls), *P. putida* (three recalls), *P. aeruginosa* (three recalls), *Pseudomonas* spp. (two recalls), and *Ralstonia* (*P.*) *pickettii* (one recall) is notable and reflects the concern for bacteria capable of growth in liquid oral dosage forms that overwhelm the preservative system.

Analyses of the underlying probable cause of the microbial contamination of nonsterile products suggest to this researcher that they are the result of: 1) microbial contamination of water for pharmaceutical purposes, 2) the use of pharmaceutical ingredients with higher microbial counts, 3) failure of preservative systems to protect liquid products, 4) microbial contamination during the manufacturing process, or 5) improper use and/or storage of the products during shelf life.

The suitable water systems, the appropriate management of the equipment, and the appropriate monitoring programs were emphasized during past regulatory

inspections so that manufacturers have no excuse for using unsuitable ingredient water during the manufacture of products. In most cases, pharmaceutical ingredients of high microbial quality can be selected for pharmaceutical manufacturing. More emphasis must be placed on preservative systems during formulation development, especially on opportunities to optimize a preservative system by manipulating the pH, surfactant properties, and water activity of formulations to make them unsuitable for microbial growth. A comprehensive discussion of the application of water activity determination in product formulation and the development of microbial monitoring programs has appeared recently in the literature (9).

A greater appreciation of the ability of different manufacturing steps to affect the microbial content of a formulation by formulators, microbiologists, and manufacturing personnel would be helpful. Exposure of pharmaceutical products to high humidity during shelf life may increase the water activity of a liquid, ointment, cream, or tablet, allowing for the growth of microorganisms on the surface of the product. Thus suitable packaging and appropriate patient handling of the drug product are important.

FUTURE TRENDS IN THE MICROBIOLOGICAL TESTING LABORATORY

The four major trends in the pharmaceutical microbiological testing laboratory are: 1) the drive to introduce new microbiological testing technologies, 2) the organization of the laboratory based on work stations, 3) the use of computerized information management systems, and

4) the change of emphasis from testing to prevention of microbial contamination.

If the pioneering German bacteriologist Robert Koch visited a routine microbiological testing laboratory in the pharmaceutical, biotechnology, or medical device industry, he would recognize that most of our techniques were first developed or used in his laboratory during the last three decades of the 19th century.

These methods include the fixing and staining of bacterial cells on glass slides for microscopic examination and photomicroscopy; growth of colonies in solid media; streaking for isolation of pure cultures in solid media; the use of agar-agar as a support for microbiological media in Petri dishes; serial dilution and plating in solid media to enumerate the microbial population in water; monitoring bacteria in the air; the classification of bacteria by their cellular morphology and differential staining; sterilization of microbiological media by filtration or steam sterilization; disinfectant testing; and aerobic and anaerobic incubation. A major trend is under way in the pharmaceutical QA microbiological testing laboratory in which the classic microbiological cultural methods developed in the late 19th century will be replaced for routine testing by biochemical, fluorescent cytometric, and nucleic acid-based techniques. Although many companies have developed instruments to automate the running or miniaturizing of existing test methods, technological improvements are progressing rapidly, with new methods based on fluorescent laser detection and nucleic acid-based detection. The future of microbiological testing is in the commercialization of automated specific detection methods that will reduce our reliance on cultural methods. This should result in routine testing with significantly shorter test cycle times and higher quality results.

Organization of the microbiology laboratory is more and more frequently based on self-directed work teams, with the widespread use of workstations based on the newer testing technologies. The workstations will reflect the product mix manufactured by the pharmaceutical company. This will rationalize the deployment of lab personnel and the flow of materials and information through the laboratory. Common workstations include: 1) sample receipt and distribution, 2) water for pharmaceutical monitoring, 3) microbial limits of pharmaceutical ingredients and nonsterile products, 4) antimicrobial effectiveness testing, 5) microbial identification, 6) environmental monitoring, 7) growth promotion, 8) sterility testing, 9) microbial assay of antibiotics and vitamins, and 10) information management.

Each testing workstation manned by trained laboratory personnel who will rotate through the laboratory contain dedicated testing equipment interfaced to LIMS, media, reagents, and supplies, with accompanying SOPs, training documents, and calibration and preventative maintenance logs. Test specimens would be received, inspected, and entered into the information management system via keyboard or by bar code scanning and distributed to the appropriate workstation. The results would be generated at the workstations, and test results will be reviewed and transferred from the workstation into LIMS.

In conclusion, the important role of the microbiologists in the pharmaceutical industry must be reinforced. The need to involve experienced microbiologists in each stage of the product life cycle to maintain product quality, safety, and efficacy is highlighted in this chapter.

REFERENCES

1. Current Good Manufacturing Practices in Manufacturing, Processing, Packaging of Drugs. 21 C.F.R. §211 through §226.
2. Testing Chapters: Antimicrobial Effectiveness Testing, Ch. 51; Microbial Limit Tests; Ch. 61; Sterility Tests, Ch. 71. U.S.P. 24.
3. Informational Chapters: Microbiological Evaluation of Clean Rooms and Other Controlled Environments, Ch. 1116; Microbiological Attributes of Pharmaceutical Articles, Ch. 1111; Pharmaceutical Dosage Forms, Ch. 1151; Validation of Compendial Methods, Ch. 1225; Water for Pharmaceutical Purposes, Ch. 1231. U.S.P. 24.
4. Eaton, A.; Clesceri, L.S.; Greenberg, A.E. *Standard Methods for the Examination of Water and Wastewater*, 19th Ed.; American Public Health Association, 1998.
5. FDA Draft Guidance for Industry. *Stability Testing of Drug Substances and Drug Products*; FDA: Washington, DC, June 1998.
6. *Pharmaceutical Packaging Integrity*. PDA Technical Report No. 27, April 1998.
7. *The Evaluation, Validation and Implementation of New Microbiological Testing Methods*. PDA Technical Report No. 33, April 2000.
8. *The Gold Sheet*; FDC Reports, Inc.: Chevy Chase, MD, Jan. 1991.
9. Friedel, R.R.; Cundell, A.M. The Application of Water Activity Measurement to the Microbiological Attributes Testing of Nonsterile OTC Drug Products. *Pharm. Forum* 1998, 25 (2), 6087–6090.

BIBLIOGRAPHY

Baird, R.M. *Microbial Quality Assurance in Cosmetics, Toiletries and Nonsterile Pharmaceuticals*; Taylor and Francis: London, 1996.

- Clontz, L. *Microbial Limit and Bioburden Tests*; Interpharm Press: Buffalo Grove, IL, 1998.
- Denyer, S.; Baird, R. *Guide to Microbiological Control in Pharmaceuticals*; Ellis Horwood: London, 1990.
- Difco Laboratories. *Difco Manual*, 11th Ed.; Division of Becton Dickinson & Company: Sparks, MD, 1998.
- FDA *Bacteriological Analytical Manual*, 8th Ed.; AOAC International: 1998.
- FDA *Guide to Inspections of Microbiological Quality Control Laboratories*, July 1993.
- Murray, P.R. *Manual of Clinical Microbiology*; ASM Press: Washington, DC, 1995.
- Olson, W.P. *Automated Microbial Identification and Quantitation*; Interpharm Press: Buffalo Grove, IL, 1998.
- Russell, A.D.; Hugo, W.B.; Ayliffe, G.A.J. *Principles and Practices of Disinfection, Preservation and Sterilization*, 3rd Ed.; Blackwell Science Ltd.: Oxford, U.K., 1999.
- Willig, S.H.; Tuckerman, M.M.; Hitchings, W.S. *Good Manufacturing Practices for Pharmaceutiucals*; Marcel Dekker, Inc.: New York, 1982.

Pharmacokinetics: Effects of Food and Fasting

Rajesh Krishna

Bradford K. Jensen

Aventis Pharmaceuticals, Inc., Bridgewater, New Jersey, U.S.A.

INTRODUCTION

Patient compliance may be improved when drug administration is tied to meal timing and intake. However, depending on the drug, the composition and the size of the meal, and the time of food intake relative to dose administration, the pharmacokinetic properties of the drug may be altered. These perturbations in the pharmacokinetics of a drug as mediated by food may cause a wide range of effects including increased drug absorption, prolonged time to peak concentrations, increased variability in drug exposure, decreased drug absorption, or inhibition of metabolism and elimination processes. Many times, these changes may adversely influence the safety and efficacy of the drug substance or product and they are highly undesirable. For new chemical entities that have been approved for marketing, the label alerts the patient and the physician by including a statement under Dosage and Administration on the extent of alterations in peak concentrations (C_{\max}) and area under the plasma concentration vs. time curve (AUC), potential clinical significance and directions on time of dose administration relative to time of meal intake. The Food and Drug Administration (FDA) has issued a draft biopharmaceutics guidance for industry entitled "Food-Effect Bioavailability and Fed Bioequivalence Studies: Study Design, Data Analysis, and Labeling" in October 2001. However, a final guidance has not yet been issued to date.

A number of intrinsic and extrinsic factors influence the pharmacokinetics of a drug in both the fasted and the fed states. These include anatomical, pathological, and physiological factors as well as the physical and chemical properties of the drug, formulation, or excipients. Because of the inherent complexities in perturbations that food may cause in the pharmacokinetics, safety, and efficacy of a drug, food-effect assessments in early drug development are typically performed proactively by pharmaceutical companies to facilitate design of pivotal clinical efficacy studies and, eventually, the drug product label. This article will highlight the factors influencing the pharmacokinetics of the drug in the fasting and the fed conditions as well as the importance of performing food-effect studies in early drug development. This article will also provide a pharmaceutical industry perspective on the

models that are presently available to predict the presence and the extent of food-mediated alterations in absorption, metabolism, and clearance as they relate to a drug development program. Although food intake can affect the pharmacokinetics of a drug for a number of routes of administration, the emphasis of this article will be on drugs administered via the oral route of administration.

EFFECTS OF FOOD ON THE PHARMACOKINETICS OF A DRUG

Food has a myriad of effects on the pharmacokinetic behavior of a drug substance regardless of therapeutic class (Table 1). Key variables include the meal itself, the composition of the meal, the drug, the formulation, the concomitant medications, the time of day, and the disease pathology. Such differences are often as a result of altered physiological influences mediated by food. Examples of these physiological effects include the degree to which the gastric emptying rate is reduced, the variations in the degree to which digestive enzymes are secreted, the dynamics of an altered pH and the consequences for ionized drugs in the gastrointestinal (GI) tract, the biliary excretion, and the stimulation of hepatic and/or splanchnic blood flow. Other important influences include physical interaction between the elements of food and the drug itself, which can interfere with absorption of a drug or merely act as a barrier to absorption from gut mucosa. Fig. 1 is a representative illustration of how a plasma concentration vs. time profile appears for a drug given in the fasted state and its effects (increased or decreased absorption) in the presence of food. For a comprehensive review of the effects of food on the pharmacokinetic behavior of drugs, the reader is referred to authoritative reviews published by Welling, Fleisher et al., and Singh.^[1-5] The discussion below is primarily centered on how this knowledge can be integrated into rational drug development. Examples are drawn from case studies of drugs in drug development, the issues of identifying food effects early in the development, timing, design, and conduct of such studies as well as the issues and opportunities in this specific area of drug development.



Table 1 Partial list of drugs whose pharmacokinetics are influenced by intake of food

Abacavir
Adinazolam
5-Aminosalicylic acid
Atorvastatin
Avitriptan
Bromazepam
Bumetanide
Celecoxib
CGP 43371
Clodronate
Cyclosporin
Danazol
Didanosine
Erythromycin
Fexofenadine
Furosemide
Ganciclovir
Halofantrine
Inidnavir
Itraconazole
Levofloxacin
Methotrexate
Nifedipine
Pravastatin
Rifabutin
Stavudine
Tacrine
UFT/leucovorin
Venlafaxine
Zolmitriptan

HUMAN FOOD-EFFECT STUDIES AND AVAILABLE PHARMACOKINETIC METHODOLOGIES

The variation in the size and intake of food, the type of food, ethnicity, the caloric value, the fat content, the time gap between meals, and the intergastric volume can influence the physiological environment wherein the drug may find itself upon administration. All of these factors may result in alterations in the bioavailability of a given drug such as stimulation of biliary secretion and splanchnic blood flow, changes in GI pH, or delay in gastric emptying. In order to capture these perturbations in bioavailability, the FDA recommends a high-fat and high-calorie meal for testing the effect of a meal on drug pharmacokinetics.^[8] Historically, this meal has been considered a “worst case scenario” for a drug–food interaction study. For new chemical entities submitted for approval in the United States, well-controlled pharmacokinetic studies are performed using this “standard high-

fat meal” as the fed condition. One may argue that the above definition of standard high-fat meal is probably not representative of a standard breakfast in the United States or other parts of the world. Exceptions include vegetarian or vegan meals, low-fat/low-calorie meals for diabetics as recommended by the American Diabetic Association (ADA), and meals that are specific for each ethnic group. All of these differences compounded by genetic predisposition to pharmacokinetics may profoundly impact the pharmacokinetics of a drug and its prescribing habit. Nonetheless, keeping the type (or lack) of meal as a constant, one can elucidate the effect of other variables (formulation, time of meal intake relative to dose, etc.) on pharmacokinetics keeping in mind that the meal should be reflective of one that is used (recommended) during patient therapy.

There are two types of food-effect studies that are often performed in humans. These studies are designed to provide dosing recommendation with meals or to provide a mechanistic investigation on the effect of food on the pharmacokinetics of drugs. Examples of the former include studies done in conformity with the FDA guidance on food-effect studies or called “standard assessments.” Examples of the latter include exploratory mechanistic studies to investigate gastric emptying using tools, such as gamma-scintigraphy, or modulating gastric pH using ranitidine coadministration, with the aim to provide/elucidate underlying reasons for observed food effects.

Standard Food-Effect Assessments—Regulatory and Product-Labeling Considerations

The draft FDA guidance entitled “Food-Effect Bioavailability and Fed Bioequivalence Studies: Study Design, Data Analysis, and Labeling” provides sponsors with information on the design of appropriate studies to assess impact of food-mediated alterations in pharmacokinetics and to provide suitable dosing guidelines on drug product label.^[8] The guidance provides useful strategies for bio-waivers if the drug in question belongs to a particular Biopharmaceutics Classification System (BCS) category.^[6,7] For example, for highly soluble and highly permeable drugs (i.e., BCS Class I), food effects should be inconsequential (between test and reference products) due to the inherent properties of the drug. The premise is that the compound is minimally influenced by food as a result of pH- or site-independent absorption and lack of dissolution differences.

A suitable study design for definitively assessing the influence of food on the pharmacokinetics of a drug or drug product includes performing a single-dose, randomized, balanced, two-period, two-treatment (fed vs. fasted)

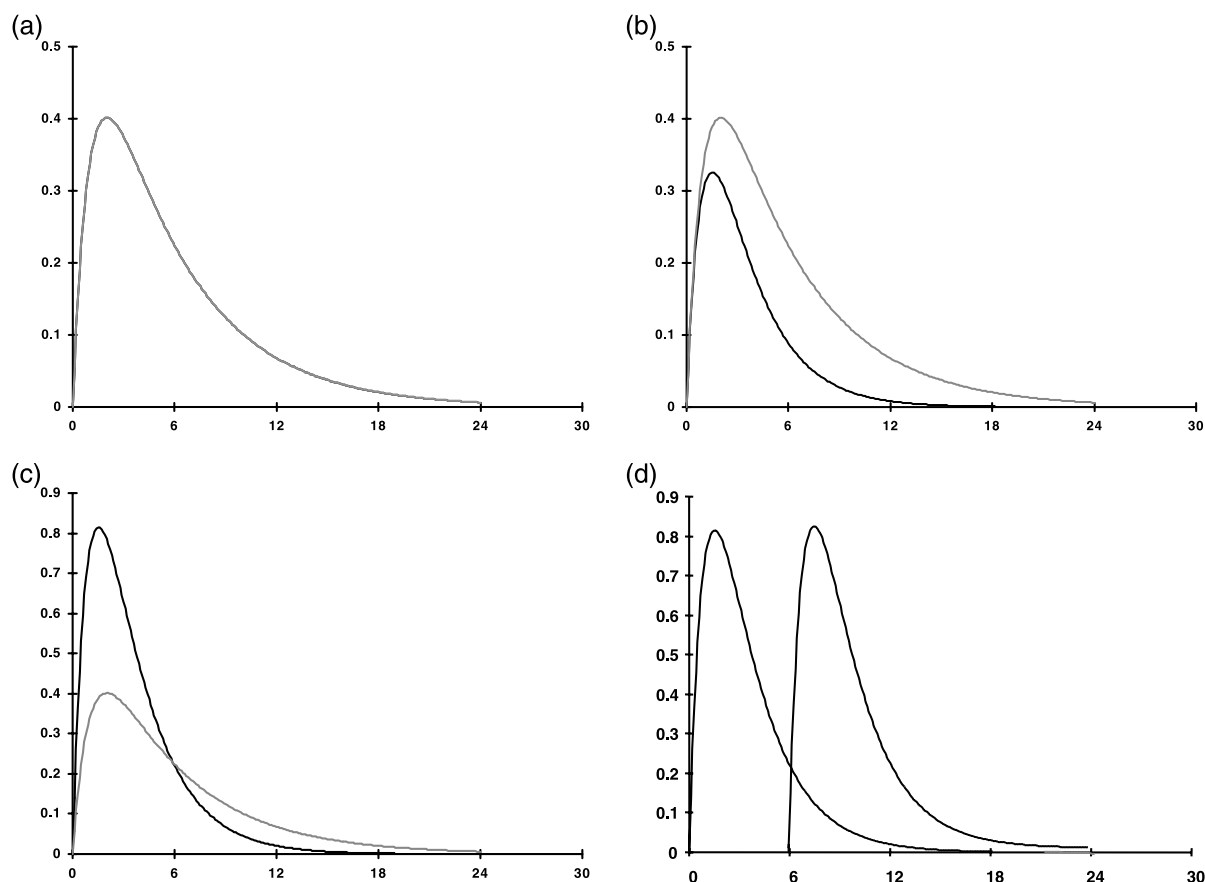


Fig. 1 a: Representative plasma concentration vs. time profile of a drug A after oral dosing in the fasted state. b: Food decreases exposure (C_{\max} and AUC) of drug A. c: Food increases exposure (C_{\max}) of drug A. d: Food delays exposure but does not decrease or increase exposure (C_{\max} or AUC) of drug A.

crossover study, with an adequate washout period between treatments. The number of subjects in the study should provide adequate statistical power for capturing small differences in exposure variables [peak concentrations (C_{\max}), T_{\max} , and area under the plasma concentration vs. time curve (AUC)]. The guidance also asks for using the highest strength of the formulation to be evaluated in such studies. For the fasting treatment, at least a 10-hr fast prior to dosing is recommended: “Following an overnight fast of at least 10 hours, subject should take the drug product. No food should be allowed for at least 4 hours post-dose. Water can be allowed at libitum after 2 hours. Scheduled standardized meals should be served throughout the remaining study period.” The issue is not the “number” of hours of predose fast time, but the stomach volume. If it contains < 50 mL (fasting state depends on the volume present in the upper region of the stomach), the stomach is in an interdigestive state. It is generally recommended to have a postdose fast time of about 4 h. For the fed treatment, the regulatory

guidance indicates a standard high-fat meal composed of two eggs fried in butter, two strips of bacon, two slices of toast with butter, 4 oz of hash brown potatoes, and 8 oz of whole milk. The caloric breakdown of this meal translates into approximately 150 cal from protein, 250 cal from carbohydrate, and 600 cal from fat. Standard pharmacokinetic calculations are performed using noncompartmental methods, e.g., C_{\max} , AUC, T_{\max} , etc. The guidance provides directions on how the data are applied to claim label attributes; for example, the degree to which 90% confidence intervals for the ratio of geometric means between fasted and fed legs are contained relative to the bounds of 80–125% for AUC or C_{\max} .^[8]

One issue that routinely arises is how to provide the FDA-recommended high-fat meal at non-U.S. investigative sites, where this meal is not readily accessible. Simple substitution of locally available food (e.g., sausage for bacon) that maintains the caloric breakdown of the FDA-recommended high-fat meal has been an acceptable approach.



A second type of study involves applying food as a covariate in population pharmacokinetic models.^[9–11] This is an emerging area of interest that is gaining regulatory concurrence. The advantages of using multilevel population pharmacokinetic models include the determination of the extent of both interindividual and intra-individual variability, the use of a limited sparse sampling approach, and the use of multiple covariates within a large Phase III trial setting. This strategy allows the simultaneous analysis of relevant covariate relationships (negative vs. positive correlation, proportional correlation, etc.) influencing the key pharmacokinetic parameters as a function of body surface area, age, weight, sex, hematological or biochemical indices, creatinine clearance, and, more importantly, any delays in drug absorption as mediated by food intake. Such analyses can be performed using tools, such as NONMEM, with standard conditional estimation methods (first order, first-order conditional estimate, or first-order conditional estimate with interaction, etc.). This allows for estimation of typical values for key pharmacokinetic parameters, V/F , CL/F , absorption rate constant, or lag time. Shepard et al. used a sequential approach with an indicator variable to assess the effect of food on the parameter of interest.^[11] Their model included $K_a = \theta_1(1 - Q) + \theta_2Q$, where Q was an indicator variable with a value of 0 (fasted) or 1 (fed), and θ_1 and θ_2 were the parameters of interest, with $K_a = \theta_1$ for fasted treatment and $K_a = \theta_2$ for fed treatment.

A number of recent examples in the literature describe the powerful use of population pharmacokinetic strategies to maximize information pertaining to food-effect behavior. These include a combination of noncompartmental and population pharmacokinetic analyses to answer a scientific question. For example, Zhou et al. used noncompartmental analysis to determine the extent of decrease in bioavailability of a drug mediated by food and then showed, using deconvolution and subsequent population pharmacokinetic analysis to assess the impact of meal timing on absorption, that the reduction in bioavailability under fed conditions was related to a reduction in the extent of absorption and not to a reduction in the rate of absorption.^[10] The use of the population pharmacokinetic approach in a multicenter trial in patients to assess food effect using sparse sampling and several categorical and continuous covariates is a more common application of this strategy. However, it is not clear that a population pharmacokinetic approach definitively addresses a food interaction without a traditional Phase I study.

Scintigraphic Assessments

The technique of gamma scintigraphy has been used to evaluate biopharmaceutical properties of drug substances

administered orally in the fed and the fasted states to evaluate food effects. Specifically, the procedure allows for studying GI transit of radiolabeled drug formulations in the fed state. With this method, one can understand the physiological processes by which the GI tract manages food and how the stomach empties meals of varying chemical composition. Such type of studies may also reveal the potential sites of absorption of an investigational product throughout the GI tract and how this is impacted in the fed state. This information is particularly useful when designing modified-release formulations. Technetium (^{99m}Tc) is a widely used nuclide in radiopharmaceuticals because of its optimal energy, short half-life, and small radiation dose needed. For a basic review on pharmacoscintigraphy tools available, the reader is referred to an excellent review by Wilding et al.^[12]

MODELS FOR ASSESSING IMPACT OF FOOD ON THE PHARMACOKINETICS OF A DRUG SUBSTANCE

In Vitro Models

Biorelevant dissolution models have been described in the literature as a means to predict food-effect behavior on the absorption of drugs.^[13–17] Reppas et al. have suggested several dissolution media that simulate fasted and fed conditions in the human body.^[14–17] They used milk (3.5% fat) and United States Pharmacopeia (USP)-simulated gastric fluid with and without pepsin as dissolution media representing the fed and the fasted states in the stomach, respectively. Dressman and Reppas have showed that such biorelevant media have distinct advantages over compendial media in the prediction of food effects for poorly water-soluble drugs.^[15] An example of how dissolution tests may be used to predict the in vivo effect of food on drug absorption is highlighted in Fig. 2 for a BCS class II (highly permeable, poorly water-soluble drug) drug, danazol. Fig. 2a shows the dissolution profile of danazol in different media, namely, water, simulated intestinal fluid (SIF), FaSSIF, and FeSSIF.^[17] FaSSIF represents the fasted state in small intestine and FeSSIF represents fed state in small intestine. From these dissolution tests, it is clear that there is enhanced dissolution in the simulated fed state relative to the simulated fasted condition. Not surprisingly, the extent of this in vitro increase is similar to the extent of increase in danazol exposure in human subjects observed in the fed state as shown in Fig. 2b.^[18] The mechanism of food effect is related to increased solubilization mediated by bile salts for this steroid drug.^[15] This strategy appears to be predictive of drugs that belong to

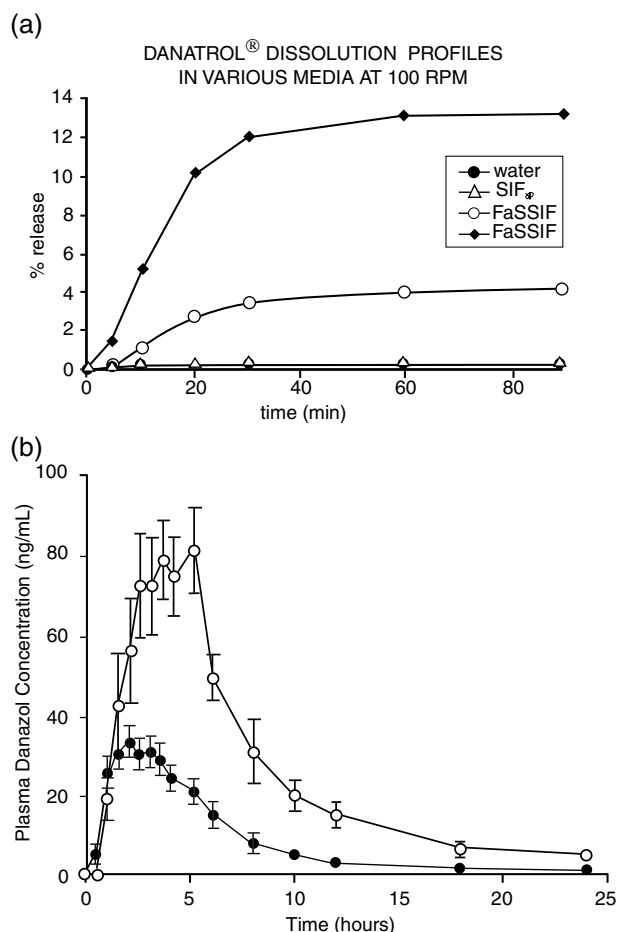


Fig. 2 a: In vitro dissolution profile for Danazol. (From Ref. [17].) b: In vivo plasma concentration vs. time profile for Danazol (closed circles represent fasted state and open circles represent fed state). (From Ref. [18], Sage Publications/Corwin Press Inc.)

the BCS Class II category and may not be generally applicable to drugs belonging to other categories.

In Vivo Models

A number of animal models have been used to describe the fasting and the fed pharmacokinetics of drugs with the aim of predicting human performance. These include mice, rats, rabbits, guinea pigs, dogs, pigs, monkeys, to name a common few. The discussion below will address the relative merits and demerits of rodents, dogs, and pigs as predictive models for humans. It should be kept in mind that animal models may be helpful in assessing large and overt food interactions in oral bioavailability. However, relatively small differences that may be ultimately important in human studies and therapeutics will probably not be detected.

Rodents

Rodents (mice, rats, guinea pigs) are generally not preferred as models for assessing food effects for various reasons, namely, differences in GI anatomy and physiology, difficulties in administering “human” food, and lack of implementing a crossover design assessment. Other potential caveats include differences in the pharmacokinetic and the metabolism behavior in rodents as compared to humans and nonhuman primates that may negate assessment in this species. Metabolic profiling of the investigational product in a panel of animal models will reveal the suitability of a particular species as a predictor of human pharmacokinetic behavior. Furthermore, the extent of serial sampling is also constrained in smaller rodents such as mice. However, rodents are one of the two species commonly and typically used as a toxicology species in drug development; it is of importance to assess the impact of food not necessarily from the point of view of predicting human food effect but more so in ensuring adequate exposure multiples are obtained for toxicokinetic purposes. Few authors have explored rat as a model for predicting food-effect behavior.^[19–21]

Dogs

Pentagastrin-pretreated dogs have been shown to be useful in screening biopharmaceutical properties of new formulations. Coupled with the features that a crossover study design can be easily applied and human food easily administered to these animals, the dog remains one of the commonly used models in assessing the pharmacokinetics of drugs.^[22,23] The disadvantages of using a dog model include dissimilar GI physiology and, hence, differences in pharmacokinetic behavior (for the pentagastrin-untreated dogs), need for modulation of GI physiology using pentagastrin and other physiological agents, and dissimilar drug metabolism/transport between dogs and humans for some compounds. Additionally, dogs are relatively more sensitive to physiological effects of certain pharmaceutical excipients commonly employed in solubilization of poorly water-soluble drugs (e.g., Polysorbate 80 and resultant orthostatic hypotension).

Although gastric emptying time is relatively similar between dogs and humans in the fasted state, the intestinal transit time is almost twice as long in humans.^[24] These apparent differences can be offset by administering pentagastrin pretreatment which allows the gastric acidity, the emptying time, and the small intestinal transit time to closely approximate human conditions, i.e., approximately pH 2, 0.7 and 4 hr, respectively.^[25] Another approach to synchronize the gastric environment in dogs to closely mimic humans includes a combined treatment of intramuscular pentagastrin with intravenous atropine sulfate.^[26]

Pentagastrin is an analog of gastrin that reproducibly stimulates gastric acid secretion and is usually administered intramuscularly prior to biopharmaceutical experimentation in doses of approximately 6–10 µg/kg. Conflicting results in the literature of drugs studied in these pretreated dogs have negated the extensive use of this model. For example, pentagastrin-pretreated dogs were more predictive of the human bioavailability of didanosine than the untreated dogs.^[22] In contrast, data from Kanerva et al.^[27] suggest that the pentagastrin-pretreated dog was unable to reliably predict the absorption of dexamethasone, an acid-labile drug.^[27]

A typical design of a food-effect study in dogs is as follows. Beagle dogs (usually $n = 6-8$) are administered a pharmacological dose of a drug formulation (hard gelatin capsule or tablet) in a randomized or nonrandomized crossover design. Treatment legs may include dosing after an overnight fast (fasted state) and dosing with a low-fat meal, a high-fat meal, or other specialized meals (fed state). The composition of meals that have been administered to dogs can be highly variable. For example, Paulson et al.^[23] administered low-, medium-, and high-fat diets.^[23] The low-fat diet was composed of one slice of toasted white bread spread with 0.5 oz of jelly, 8 oz of skim milk, and 6 oz of orange juice. The medium-fat diet was composed of one slice of toasted white bread with 0.5 oz each of peanut butter and jelly, 1 oz of dry cereal (cornflakes), 8 oz of skim milk, 6 oz of orange juice, and one banana. The high-fat diet was composed of two slices of toasted white bread spread with 1.2 oz of butter, two eggs fried in butter, two slices of cooked bacon, 2 oz of hash brown potatoes fried in butter, and 8 oz of whole milk. These diets were homogenized then administered either in a bowl or with a syringe to the dogs. A suitable washout period separated the treatment legs. Blood samples are collected at predetermined intervals predrug and postdrug administration and pharmacokinetic behavior assessed with and without food.

Pigs and minipigs

The digestive tract of a minipig is anatomically and physiologically similar to that of the human adult digestive tract. In fact, the weanling minipig model has a GI tract that is closely similar to children between the age of 2 and 5 years.^[28,29] Specifically, the similarities exist for the gastric villi, the regulation and the composition of gastric secretions, the transit times as well as in the digestive and the absorptive behaviors.^[30–34] Regardless of these advantages, the use of the minipig in pharmacokinetic assessments has been relatively uncommon. One example of an application to the assessment of food effects in minipigs is the study on the effect of various fat-content diets on the bioavailability of theophylline following a

400-mg single-dose Theo-24.^[35] The observed similarities in the pharmacokinetics of theophylline in pigs and humans led Koritz et al. to suggest swine as a predictive model for the assessment of theophylline bioequivalence in humans.^[36]

FOOD-EFFECT ASSESSMENT IN DRUG DEVELOPMENT: KEY CONSIDERATIONS

Timing of Food-Effect Studies

There are generally two types of clinical pharmacokinetic/pharmacodynamic (PK/PD) studies that may be performed during drug development (Table 2). One is a “pilot” study to obtain preliminary exploratory assessment used for facilitating internal decision making for the progression of a compound and the other is a more formal statistically powered “pivotal” study to provide dosing recommendation for purposes of the drug product label. The key difference is that the pilot study is often powered for large differences and the pivotal study is powered for small differences for *a priori* defined endpoint. The pilot study, which usually uses a developmental formulation and a guess of the therapeutic dose, can also provide valuable information for developing a well-designed pivotal study. Label-driven food-effect assessments are performed in later stages of drug development, particularly when the “market” formulation is available and the therapeutic dose is known from Phase III efficacy studies. Appropriate recommendations for performing such studies are available in the Guidance for Industry issued by the FDA in the United States.

The exploratory assessment of food on the pharmacokinetics of new chemical entities in the drug development chain of events varies largely within companies. This is

Table 2 Typical clinical pharmacology and PK/PD studies in early drug development

I	Clinical pharmacology studies
	Single-ascending-dose safety, PK, and PD
	Multiple-ascending-dose safety, PK, and PD
	Proof of mechanism study
	Proof of principle study
II	Clinical PK/PD studies
	Age and gender PK
	Drug–drug interaction
	Renal and hepatic impairment PK
	BA/BE food and fasted PK
	Dose proportionality and diurnal variation
	Single- and multiple-dose PK/PD
	14C ADME study
	Tissue distribution (PET, gamma scintigraphy, etc.) PK

also driven by marketing considerations, competitive advantage, flexibility for dosing options, and the therapeutic index of the drug if known. Several authors have suggested inclusion of an exploratory assessment early in the drug development program because appropriate formulation development may be initiated for drugs exhibiting undesirable food effects.^[37,38] A caveat for this approach is the large number of food-effect studies that may need to be performed if multiple changes are made for the formulation and/or the excipients within them. If multiple-formulation switches occur in a program, it may be worthwhile to delay the food-effect assessment to discount the formulation as a variable.

A convenient method to examine food effects is the inclusion of a food-effect assessment as part of the first-in-man study once the maximum tolerated dose (MTD) is identified. Depending on the drug bioavailability characteristics in animals, physicochemical properties of the drug and preliminary data on whether the drug exhibits dissolution rate-limited absorption behavior, a dose that is 1/2 to 1/4 of the MTD, may initially be used for the preliminary food-effect assessment. This method potentially avoids undesirably high exposures of the drug when adequate safety data are not available. Dose selection will also depend on the predicted therapeutically relevant dose range, the linearity in pharmacokinetic behavior, and the sensitivity of the bioanalytical assay. Typically, the FIM study design includes nonrandomized, noncrossover, single-ascending doses in eight subjects (six active, two placebo), where doses are escalated until an MTD is reached and dosing typically performed after an overnight fast. For example, in the food-effect leg, 0.25–0.5 × MTD cohorts return for a second dose after a wash-out period wherein dose is administered with a standard high-fat meal. This allows within-subject comparison of exposure with and without food, but does not account for period effects. Hence, it is not unusual to obtain inconclusive results with this design. To lend rigor in the scientific assessment of food effect, it may be more desirable to have the food-effect assessment done as part of the first-in-man (FIM) study using a randomized crossover design in eight active subjects.

In general, however, it is recommended to perform specific well-designed food-effect studies with adequate power for large differences to capture meaningful deviations in exposure.

Single vs. Multiple Dosing—Dose Selection

The selection of the appropriate dose in assessing the effect of food on the pharmacokinetics of a drug is an important consideration. Dose selection may be based on several criteria, namely, therapeutically relevant concentrations (if known), safety margin, linear phar-

macokinetics, etc. It is generally required to evaluate a single dose of a drug for such an assessment. However, depending on nonlinear pharmacokinetic behavior and the degree of accumulation (or induction) upon multiple dosing, a multiple dose may be a better option. A multiple-dose study may also be more meaningful if the primary or secondary objective is to evaluate the effect of food on the safety or the pharmacodynamics of a drug, where single-dose administration may not be sufficient to induce meaningful perturbations in pharmacodynamics. An example of this approach is in the multiple-dose administration of the “statin” group of antihyperlipidemic agents for assessing the effect of food on the lipid-lowering efficacy [i.e., LDL-cholesterol (LDL-C) levels] given the slow turnover rate of LDL-C.

Dose selection based on pharmacokinetic profile

The oral absorption of certain highly permeable, poorly water-soluble drugs with low oral bioavailability in the

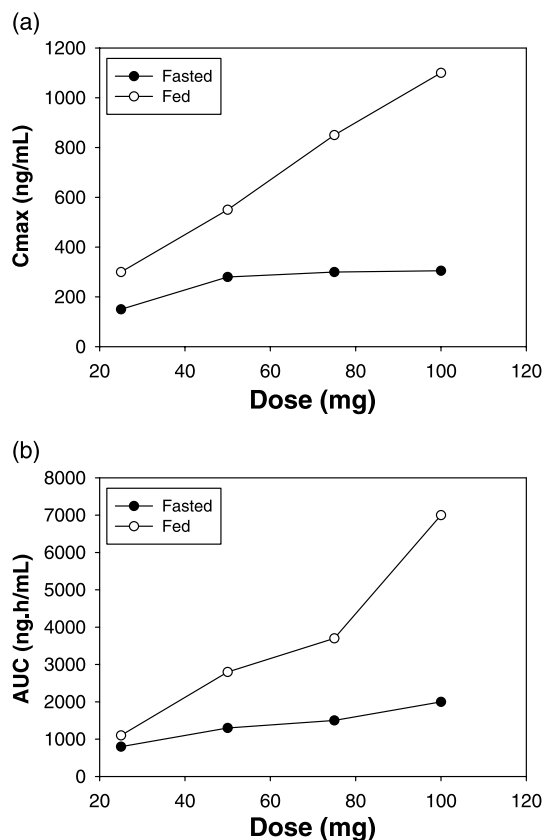


Fig. 3 a: Relationship between dose and C_{\max} for acitretin. (From Ref. [39].) b: Relationship between dose and AUC for acitretin. (From Ref. [39].)

fasted state may be profoundly influenced by the intake of fatty meal. Dose selection for such drugs can be challenging and important. For example, acitretin is a retinoid indicated for the treatment of psoriasis. Depending on the dose of acitretin used, food appears to increase the absorption in a varying manner. Fig. 3a and b highlights the relationship between dose and treatment (fed vs. fasted) on the drug's C_{\max} and AUC. In the fasted state, both the C_{\max} and the AUC reach a plateau at 50-mg dose with minor, if any, increases in exposure at doses higher than 50 mg. In contrast, the drug exhibits almost dose-proportional increases in C_{\max} and AUC over the entire tested dose range up to 100 mg, when the drug is administered with a high-fat meal.^[39] When acitretin is given with food, there is also a decrease in intersubject variability in exposure parameters. These results have important implications for investigational agents exhibiting dissolution rate-limited absorption during drug development as this presents an opportunity for key formulation development to increase bioavailability if safety is not an issue and if formulation options are available. This example also illustrates the need for an early assessment of a food interaction study. In acitretin, the patients in the Phase III pivotal studies were instructed to take their daily acitretin dose with the main meal of the day based on this information.

Dose selection based on PK/PD relationship

A second more germane consideration is dose selection based on PK/PD relationships. Depending on the reference point on a PK/PD chart (Fig. 4), food may have little or profound effect on either PK, PD, or both. For example, drug A shows profound reductions in C_{\max} and AUC, but blood pressure reduction is relatively unaffected by food; it is conceivable that the drug PD is at the maximal effect where large differences in PK leads to minor changes in PD. On the other hand, if the dose of drug B was selected such that its PD is at 50% of maximal effect, then there is much larger PD differences for observed differences in drug exposure.

An example of the former case (drug A behavior) is illustrated in the effect of food on the PK and the PD of pravastatin, a representative statin in the treatment of hyperlipidemia. Although a low-fat, low-cholesterol meal altered the pharmacokinetics of 20-mg pravastatin significantly, no adverse impact on pharmacodynamics (lipid-lowering efficacy) was observed.^[40] Concomitant intake of meals reduced pravastatin C_{\max} by 49% and AUC by 31%, the change in LDL-C-lowering efficacy was minimal (–37% with meal vs. –36% 1-hr premeal). The authors recommend administering pravastatin without regard to meals. It is worth noting here that the LDL-C-lowering efficacy of pravastatin plateaus at ca. –32%

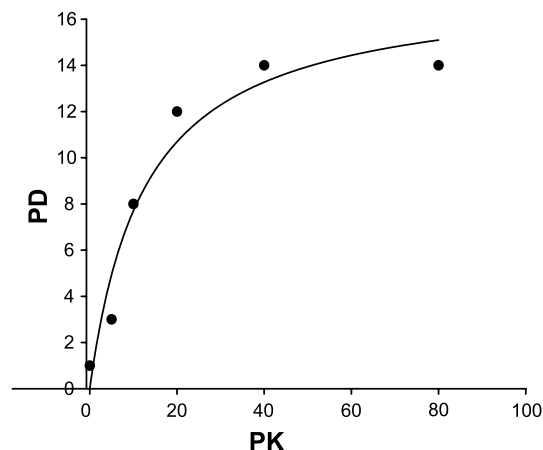


Fig. 4 Representative illustration of PK/PD relationship for a hypothetical drug.

(20 mg) to ca. –37% (80 mg) changes from baseline; at –37%, the PD is already at the flat portion of the sigmoidal PK–PD curve (Fig. 4).^[41] Consequently, even with a change in PK approximating 50%, there is minimal impact on PD. It is unlikely that the PD would be influenced significantly by food for compounds whose PD is at this flat portion of the PK–PD curve than would be the case if the dose selected represented a PD of 0.5 times maximal effect, where there is a much larger variability for PD differences with variability in PK.

Type of Meals

Another key consideration when performing food-effect studies is the type of meal utilized for assessing effect of food on the pharmacokinetics of a drug. Because of differing ethnicities, geographical regions, and variations in the eating behavior of human beings across the world, it is conceivable that the “standard” breakfast, lunch, and dinner will vary extensively. Then it becomes increasingly difficult to assess the true significance of food and its impact on PK/PD and to develop dosing recommendations that are systematically applied to a drug that is prescribed globally.

In order to provide a systematic basis for capturing food-mediated alterations in the pharmacokinetics of drugs, the FDA draft guidance recommends a standard high-fat meal as being composed of “2 eggs fried in butter, 2 strips of bacon, 2 slices of toast with butter, 4 ounces of hash brown potatoes, 8 ounces of whole milk.” Differences in pharmacokinetic behavior may be expected depending on the type of breakfast utilized; for example, an English breakfast is different in composition and caloric value to an American breakfast or a special diet in adherence with ADA recommendations. To derive

meaningful data across these varying types of meals, meals are often categorized in terms of carbohydrate, protein, and fat content as “high-fat,” “low-fat,” “high-protein,” or “high-carbohydrate” diet. An alternate categorization is based on the caloric value of the meal itself, as “low calorie” or “high calorie.” The influence of differing breakfast compositions and the influence of a solid vs. a liquid (whole milk) meal were investigated by Colburn et al. for etretinate.^[42] The drug was administered in the fasted state or in the fed state with a standard high-fat breakfast, a standard high-carbohydrate breakfast, and whole milk (16 oz). Etretinate concentrations in plasma were greater when given with a high-fat meal and whole milk compared to the fasted state or when given with a high-carbohydrate meal. Meals with differing fat contents can also influence drug absorption and disposition by influencing physiology. It is known that higher levels of fat may inhibit gastric emptying. Moreover, the GI transit times are different for solid and liquid meals and this may impact the pharmacokinetics of a drug. Table 3 is a partial list of some of the components in meals that have an effect on drug absorption, metabolism, and clearance. The lack of consistency in the scientific literature on the type of meal used in the investigation prohibits data comparisons across populations unless such differing compositions are used in a single study.

Of more importance in relation to the influence of food on absorption is the rate of blood flow in the capillary system. Food may cause an increase in splanchnic blood flow leading to alterations in drug absorption, depending on the ingredients present in the food that is ingested. Closely tied into this principle is the extent to which a given drug is extracted by first-pass metabolism when traversing through the liver via the portal circulation. Both the type and the size of meal can influence splanchnic blood flow. Using a noninvasive technique for the assessment of hepatic haemodynamics using ra-

diocolloid imaging, Walmsley et al. assessed the influence of meal on hepatic, mesenteric, and splenic arterial effective blood flows.^[43] They showed that both the total effective liver blood flow and the mesenteric arterial effective flow increased by 28–69% after eating. It is known that long-chain fatty acids stimulate human intestinal blood flow.^[44]

It has become increasingly evident that several ingredients in food or components of a meal can influence drug metabolism and clearance via inhibition of cytochrome P450 isozymes or modulation of drug transporters such as p-glycoprotein.^[45–53] These ingredients include vegetables and fruit products and components such as St. John’s wort, garlic, grapefruit juice, certain type of orange juices, apple juice, red, black, or white pepper.^[45–53] Red wine and green tea also have ingredients that can alter the metabolism/clearance of a drug. In order to differentiate these influences from those arising from the altered GI physiology, the drug development program may need to include separate food and metabolism–transporter interaction studies to assess the overall impact on dosing recommendations.

Time of Meal Consumption Relative to Dose

The time of meal consumption relative to dose has important implications for compounds whose pharmacokinetics are severely impacted by food. The extent of food-mediated alterations in pharmacokinetics are influenced by the time interval between ingestion of food and intake of drug. Notably, as the time elapsed between meal ingestion and drug intake becomes longer, the less severe the effect of food on the pharmacokinetics of a drug would presumably be. Consider the following example. If one were to fix the meal intake as a constant and to vary the time of intake of a meal, then differences in the rate and the extent of drug absorption may be observed. Fig. 5 relates AUC as a function of time of meal intake for tacrolimus (FK506, Prograf) following a single 5-mg oral dose.^[54] The treatments were: A) 10-hr fast; B) drug intake 1 hr before breakfast; C) drug intake immediately following consumption of the breakfast; and D) drug intake 1.5 hr after beginning consumption of the breakfast. As expected, tacrolimus absorption was the highest in the fasted state.^[54] The lowest impact on bioavailability was when taking the drug 1 hr prior to a meal; the highest impact was when tacrolimus was ingested immediately after a meal (treatments C and D).

Marathe et al. have investigated the effect of food on the pharmacokinetics of avitriptan in a systematic manner as part of the avitriptan drug development program. In the first study,^[55] they showed that concomitant ingestion of food significantly reduced avitriptan bioavailability. Specifically, mean C_{max} and AUC were reduced 70% and

Table 3 Partial list of ingredients of meals that influence drug absorption, metabolism, and elimination

High-fat-containing products
Juices
Orange juice
Grapefruit juice
Apple juice
Red wine
Black pepper
Tea
Garlic
Broccoli



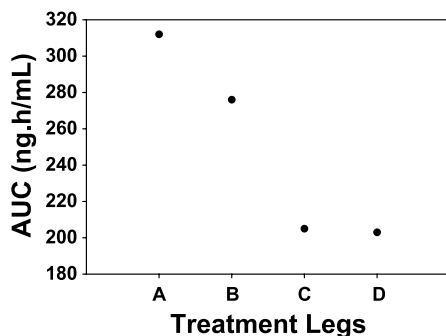


Fig. 5 Relationship of time of meal and drug AUC. (From Ref. [54].)

45%, respectively, while T_{\max} was delayed from 45 min to 2 hr when avitriptan was administered as a 50 mg capsule 5 min after a standard high-fat meal. Because avitriptan was developed as a treatment for migraine, this dramatic food effect has important implications for the developability of the drug for this indication. In the second study, Marathe et al. designed a meal-timing study where avitriptan was administered after an overnight fast and at 0.5, 1, 2, and 4 hr after intake of a standard high-fat meal.^[56] There was a meaningful relationship between avitriptan bioavailability as a function of time interval between dose and meal intake with the impact of food decreasing over time. Mean C_{\max} decreased by 61%, 58%, 50%, and 35% after 0.5-, 1, 2-, and 4-hr postmeal, respectively; mean AUC decreased by 35%, 31%, 34%, and 19% after 0.5-, 1, 2-, and 4-hr postmeal, respectively. Median T_{\max} values were 0.5 hr (fasted), 1.5 hr (0.5- and 1-hr postmeal), 1.25 hr (2-hr postmeal), and 0.75 hr (4-hr postmeal). In a later study by the same authors, scintigraphic analysis showed that avitriptan was rapidly absorbed from the upper small intestine after emptying from the stomach.^[57] Gastric emptying was slow and continuous in the fed state leading to extended absorption. The authors also attributed the high intrasubject variability in C_{\max} and AUC to variability in gastric emptying under fasted and fed conditions.

A similar study was reported by Laitinen et al.^[58] on the effect of time of meal on the bioavailability of the bisphosphonate drug clodronate.^[58] Exposure (AUC) was the highest when clodronate was administered 2-hr premeal, followed by a 1-hr premeal (91% of 2-hr premeal AUC), a 0.5-hr premeal (69% of 2-hr premeal AUC), with meal (10% of 2-hr premeal AUC), and a 2-hr postmeal (34% of 2-hr premeal AUC). Although the 2-hr premeal does not present a “true” fasting state, there was a measurable difference between these short time intervals premeal and postmeal leading the investigators to conclude that clodronate be administered at least 0.5 hr

prior to breakfast. The closeness of this prescribed time of drug intake to the meal ensures some balance of patient compliance with minimally acceptable compromise on the impact on drug absorption.

For a drug with a wide therapeutic margin, such variations in drug absorption relative to timing of a meal may be inconsequential in terms of clinical relevance; however, the same cannot be said for a drug with narrow therapeutic margin.

Type of Formulation

The type of formulation used, the excipients contained in the formulation as well as the particle size of the drug substance all may influence drug absorption and, thus, confound the observed food-effect behavior for a given drug.^[59–63] For example, a highly water-soluble drug may be formulated in a tablet that contains a poorly water-soluble polymer excipient that influences release of the drug from the matrix and subsequent dissolution properties of the tablet. Certain hydrophilic polymers, such as hydroxypropylmethylcellulose, have been implicated in dose dumping of modified- or controlled-release formulations in the presence of food. The use of varying excipients within formulations are often the cause of conflicting data for observed food effects. Whereas some excipients, such as mannitol, influence drug absorption by altering transit times, others, such as PEG-300, Cremophor EL, and Polysorbate 80, inhibit p-glycoprotein and other drug transporters^[59,60,63].

In a drug development program, it is sometimes necessary to change the formulation depending on drug development timelines and availability of drug substances to meet the needs of clinical pharmacology and PK/PD studies in a timely manner. For example, the formulation may change from an aqueous suspension in the first-in-man study to a capsule formulation in proof-of-concept study to, eventually, a market formulation that would be a tablet formulation. Each time such formulation changes are made, a new food-effect study may need to be performed to assess the influence of the altered formulation on the observed food effect to design the Phase III study in an optimal manner. To minimize the number of food-effect studies performed due to formulation changes, it is essential to maintain consistent formulation strategy across the drug development phase. Inclusion of excipients should occur based on informed input from a variety of functional areas including pharmaceutical sciences, clinical pharmacology, clinical PK/PD, and drug safety. A pivotal food-effect study needs to be performed using the same formulation used in Phase III pivotal efficacy trials or the market formulation so that appropriate dosing recommendations with regard to meals are provided.

Site of Drug Absorption

Absorption of drugs from the GI tract differs extensively depending on the intrinsic properties of the drug. For example, a drug may be absorbed from the upper GI tract or throughout the GI tract. This will have important implications for food effects as well as for other applications such as developing modified-release formulations or demonstrating bioequivalence of a test formulation vs. the reference formulation. For a drug that has an absorption window of the upper GI tract, a delay in T_{\max} as mediated by food may lead to suboptimal absorption.

CONCLUSION

It is very clear that food can have significant pharmacokinetic and pharmacodynamic consequences for drug substances. Although not discussed in this article, food can also have deleterious effects on efficacy and safety of a product. A growing body of literature now points at many pharmaceutical excipients that are typically used in drug development cycle to influence drug disposition by modulating metabolism and efflux properties of drugs. Whether the information on food effects is derived for internal company decision making and developability considerations or for dosing recommendations with the final market formulation, understanding the myriad and often complex actions mediated by food is critical in drug development. A rationale selection of available choices within the context of a well-designed and adequately powered clinical study will provide important clues for the drug development scientist to: 1) enable well-designed late-stage clinical efficacy trials; 2) optimize the formulations for early- and midstage developmental compounds; or 3) provide prescribing options to the physician for optimal patient compliance on soon-to-be-marketed products.

REFERENCES

1. Welling, P.G. Influence of food and diet on gastrointestinal drug absorption: A review. *J. Pharmacokinet. Biopharm.* **1977**, *5* (4), 291–334.
2. Welling, P.G. Interactions affecting drug absorption. *Clin. Pharmacokinet.* **1984**, *9* (5), 404–434.
3. Welling, P.G. Effects of food on drug absorption. *Pharmacol. Ther.* **1989**, *43* (3), 425–441.
4. Fleisher, D.; Li, C.; Zhou, Y.; Pao, L.H.; Karim, A. Drug, meal and formulation interactions influencing drug absorption after oral administration. Clinical implications. *Clin. Pharmacokinet.* **1999**, *36* (3), 233–254.
5. Singh, B.N. Effects of food on clinical pharmacokinetics. *Clin. Pharmacokinet.* **1999**, *37* (3), 213–255.
6. Lobenberg, R.; Amidon, G.L. Modern bioavailability, bioequivalence and biopharmaceutics classification system. New scientific approaches to international regulatory standards. *Eur. J. Pharm. Biopharm.* **2000**, *50* (1), 3–12.
7. Amidon, G.L.; Lennernas, H.; Shah, V.P.; Crison, J.R. A theoretical basis for a biopharmaceutic drug classification: The correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm. Res.* **1995**, *12* (3), 413–420.
8. Food-Effect Bioavailability and Fed Bioequivalence Studies: Study Design, Data Analysis, and Labeling. In *Draft Guidance for Industry*; U.S. Department of Health and Human Services, Food and Drug Administration, October 2001.
9. Jorga, K.; Fotteler, B.; Banken, L.; Snell, P.; Steimer, J.L. Population pharmacokinetics of tolcapone in parkinsonian patients in dose finding studies. *Br. J. Clin. Pharmacol.* **2000**, *49* (1), 39–48.
10. Zhou, H.; Khalilieh, S.; Lau, H.; Guerret, M.; Osborne, S.; Alladina, L.; Laurent, A.L.; McLeod, J.F. Effect of meal timing not critical for the pharmacokinetics of tegaserod (HTF 919). *J. Clin. Pharmacol.* **1999**, *39* (9), 911–919.
11. Shepard, D.R.; Mani, S.; Kastrissios, H.; Learned-Coughlin, S.; Smith, D.; Ertel, P.; Magnum, S.; Janisch, L.; Fleming, G.F.; Schilsky, R.L.; Ratain, M.J. Estimation of the effect of food on the disposition of oral 5-fluorouracil in combination with eniluracil. *Cancer Chemother. Pharmacol.* **2002**, *49* (5), 398–402.
12. Wilding, I.R.; Coupe, A.J.; Davis, S.S. The role of gamma-scintigraphy in oral drug delivery. *Adv. Drug Deliv. Rev.* **2001**, *46* (1–3), 103–124.
13. Charman, W.N.; Porter, C.J.; Mithani, S.; Dressman, J.B. Physiochemical and physiological mechanisms for the effects of food on drug absorption: The role of lipids and pH. *J. Pharm. Sci.* **1997**, *86* (3), 269–282.
14. Nicolaidis, E.; Svmillides, M.; Dressman, J.B.; Reppas, C. Biorelevant dissolution testing to predict the plasma profile of lipophilic drugs after oral administration. *Pharm. Res.* **2001**, *18* (3), 380–388.
15. Dressman, J.B.; Reppas, C. In vitro–in vivo correlations for lipophilic, poorly water-soluble drugs. *Eur. J. Pharm. Sci.* **2000**, (Suppl 2), S73–S80.
16. Nicolaidis, E.; Galia, E.; Ethymiopoulos, C.; Dressman, J.B.; Reppas, C. Forecasting the in vivo performance of four low solubility drugs from their in vitro dissolution data. *Pharm. Res.* **1999**, *16* (12), 1876–1882.
17. Galia, E.; Nicolaidis, E.; Horter, D.; Lobenberg, R.; Reppas, C.; Dressman, J.B. Evaluation of various dissolution media for predicting in vivo performance of class I and II drugs. *Pharm. Res.* **1998**, *15* (5), 698–705.
18. Charman, W.N.; Rogge, M.C.; Boddy, A.W.; Barr, W.H.; Berger, B.M. Absorption of danazol after administration to different sites of the gastrointestinal tract and the relationship to single and double peak phenomena in the plasma profiles. *J. Clin. Pharmacol.* **1993**, *33*, 1207–1213.
19. Lee, K.H.; Xu, G.X.; Schoenhard, G.L.; Cook, C.S. Mechanisms of food effects of structurally related antiarrhythmic drugs, disopyramide and bidisomide in the rat. *Pharm. Res.* **1997**, *14* (8), 1030–1038.



20. Cook, C.S.; Zhang, L.; Osis, J.; Schoenhard, G.L.; Karim, A. Mechanism of compound-and species-specific food effects of structurally related antiarrhythmic drugs, disopyramide and bidisomide. *Pharm. Res.* **1998**, *15* (3), 429–433.
21. Pao, L.H.; Zhou, S.Y.; Cook, C.; Kararli, T.; Kirchhoff, C.; Truelove, J.; Karim, A.; Fleisher, D. Reduced systemic availability of an antiarrhythmic drug, bidisomide, with meal co-administration: Relationship with region-dependent intestinal absorption. *Pharm. Res.* **1998**, *15* (2), 221–227.
22. Knupp, C.A.; Shyu, W.C.; Morgenthien, E.A.; Lee, J.S.; Barbhuiya, R.H. Biopharmaceutics of didanosine in humans and in a model for acid-labile drugs, the pentagastrin-pretreated dog. *Pharm. Res.* **1993**, *10* (8), 1157–1164.
23. Paulson, S.K.; Vaughn, M.B.; Jessen, S.M.; Lawal, Y.; Gresk, C.J.; Yan, B.; Maziasz, T.J.; Cook, C.S.; Karim, A. Pharmacokinetics of celecoxib after oral administration in dogs and humans: Effect of food and site of absorption. *J. Pharmacol. Exp. Ther.* **2001**, *297* (2), 638–645.
24. Lui, C.Y.; Amidon, G.L.; Berardi, R.R.; Fleisher, D.; Youngberg, C.; Dressman, J.B. Comparison of gastrointestinal pH in dogs and humans: implications on the use of the beagle dog as a model for oral absorption in humans. *J. Pharm. Sci.* **1986**, *75* (3), 271–274.
25. Sagara, K.; Nagamatsu, Y.; Yamada, I.; Kavata, M.; Mizuta, H.; Ogawa, K. Bioavailability study of commercial sustained-release preparations of diclofenac sodium in gastrointestinal physiology regulated-dogs. *Chem. Pharm. Bull. (Tokyo)* **1992**, *40* (12), 3303–3306.
26. Sagara, K.; Kawata, M.; Mizuta, H.; Shibata, M. Utility of gastrointestinal physiology regulated-dogs: Bioavailability study of a commercial sustained-release dosage form of theophylline. *Biol. Pharm. Bull.* **1994**, *17* (7), 931–934.
27. Kanerva, H.; Klebovich, I.; Drabant, S.; Urtti, A.; Nevalainen, T. Different absorption profiles of deramciclane in man and in dog. *J. Pharm. Pharmacol.* **1998**, *50* (10), 1087–1093.
28. Moughan, P.J.; Birtles, M.J.; Cranwell, P.D.; Smith, W.C.; Pedraza, M. The Piglet as a Model Animal for Studying Aspects of Digestion and Absorption in Milk-Fed Humans. In *Nutritional Triggers for Health and in Disease*; Simopoulos, A.P., Ed.; Karger: Basel, 1992; 40–113.
29. Daher, G.C.; Lawson, K.D.; Tallmadge, D.H.; Vanderploeg, P.; Miller, K.W. Disposition of ingested olestra in weanling mini-pigs. *Food Chem. Toxicol.* **1996**, *34* (8), 693–699.
30. Bjorkman, O.; Crump, M.; Phillips, R.W. Intestinal metabolism of orally administered glucose and fructose in Yucatan miniature swine. *J. Nutr.* **1984**, *114* (8), 1413–1420.
31. Forsum, E.; Goranzon, H.; Rundgren, M.; Thilen, M.; Hambraeus, L. Protein evaluation of mixed diets. Comparative study in man and in the pig and rat of vegetable-animal and vegetable protein diets. *Ann. Nutr. Metab.* **1981**, *25* (3), 137–150.
32. Graham, H.; Aman, P. The pig as a model in dietary fibre digestion studies. *Scand. J. Gastroenterol., Suppl.* **1987**, *129*, 55–61.
33. Kurihara-Bergstorm, T.; Woodworth, M.; Feisullin, S.; Beall, P. Characterization of the Yucatan miniature pig skin and small intestine for pharmaceutical applications. *Lab. Anim. Sci.* **1986**, *36* (4), 396–399.
34. Miller, E.R.; Ullrey, D.E. The pig as a model for human nutrition. *Annu. Rev. Nutr.* **1987**, *7*, 361–382.
35. Shiu, G.K.; Sager, A.O.; Velagapudi, R.B.; Prasad, V.K.; Skelly, J.P. The effect of food on the absorption of controlled-release theophylline in mini-swine. *Pharm. Res.* **1988**, *5* (1), 48–52.
36. Koritz, G.D.; Bourne, D.W.; Hunt, J.P.; Prasad, V.I.; Bevell, R.F.; Gautam, S.R. Pharmacokinetics of theophylline in swine: A potential model for human drug bioavailability studies. *J. Vet. Pharmacol. Ther.* **1981**, *4* (3), 233–239.
37. Karim, A. Importance of Food Effect Studies Early in Drug Development. In *Bioavailability, Bioequivalence, and Pharmacokinetic Studies*; Midha, K.K., Nagai, T., Eds.; Business Center for Academic Societies Japan: Tokyo, 1996; 221–229.
38. Li, Z.; Vachharajani, N.N.; Krishna, R. On the assessment of effects of food on the pharmacokinetics of drugs in early development. *Biopharm. Drug Dispos.* **2002**, *23* (4), 165–171.
39. Wiegand, U.W.; Jensen, B.K. Pharmacokinetics of Acitretin in Humans. In *Retinoids: 10 Years On*; Saurat, J.H., Ed.; Karger: Basel, 1991; 192–203.
40. Pan, H.Y.; DeVault, A.R.; Brescia, D.; Willard, D.A.; McGovern, M.E.; Whigan, D.B.; Ivashkiv, E. Effect of food on pravastatin pharmacokinetics and pharmacodynamics. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1993**, *31* (6), 291–294.
41. Pravachol tablets. Physician's Desk Reference. Thomson Micromedex. <http://pdrel.thomsonhc.com/pdrel/librarian/>.
42. Colburn, W.A.; Gibson, D.M.; Rodriguez, L.C.; Bugge, C.J.; Blumenthal, H.P. Effect of meals on the kinetics of etretinate. *J. Clin. Pharmacol.* **1985**, *25* (8), 583–589.
43. Walmsley, B.H.; Fleming, J.S.; Ackery, D.M.; Karran, S.J. Non-invasive assessment of absolute values of hepatic haemodynamics using radiocolloid scintigraphy. *Nucl. Med. Commun.* **1987**, *8* (8), 613–621.
44. Moneta, G.L.; Taylor, D.C.; Helton, W.S.; Mulholland, M.W.; Strandness, D.E. Duplex ultrasound measurement of postprandial intestinal blood flow: Effect of meal composition. *Gastroenterology* **1988**, *95* (5), 1294–1301.
45. Foster, B.C.; Foster, M.S.; Vandenhoek, S.; Krantis, A.; Budzinski, J.W.; Arnason, J.T.; Gallicano, K.D.; Choudri, S. An in vitro evaluation of human cytochrome P450 3A4 and P-glycoprotein inhibition by garlic. *J. Pharm. Pharmacol. Sci.* **2001**, *4* (2), 176–184.
46. Di Marco, M.P.; Edwards, D.J.; Wainer, I.W.; Ducharme, M.P. The effect of grapefruit juice and seville orange juice on the pharmacokinetics of dextromethorphan: The role of gut CYP3A and P-glycoprotein. *Life Sci.* **2002**, *71* (10), 1149–1160.

47. Dresser, G.K.; Bailey, D.G.; Leake, B.F.; Schwarz, U.I.; Dawson, P.A.; Freeman, D.J.; Kim, R.B. Fruit juices inhibit organic anion transporting polypeptide-mediated drug uptake to decrease the oral availability of fexofenadine. *Clin. Pharmacol. Ther.* **2002**, *71* (1), 11–20.
48. Takanaga, H.; Ohnishi, A.; Yamada, S.; Matsuo, H.; Morimoto, S.; Shoyama, Y.; Ohtani, H.; Sawada, Y. Polymethoxylated flavones in orange juice are inhibitors of P-glycoprotein but not cytochrome P450 3A4. *J. Pharmacol. Exp. Ther.* **2000**, *293* (1), 230–236.
49. Jodoin, J.; Demeule, M.; Beliveau, R. Inhibition of the multidrug resistance P-glycoprotein activity by green tea polyphenols. *Biochim. Biophys. Acta* **2002**, *1542* (1–3), 149–159.
50. Piver, B.; Berthou, F.; Dreano, Y.; Lucas, D. Inhibition of CYP3A, CYP1A and CYP2E1 activities by resveratrol and other non volatile red wine components. *Toxicol. Lett.* **2001**, *125* (1–3), 83–91.
51. Neuman, M. Metabolic effects and drug interactions provoked by certain vegetables: Grapefruit, St. John's wort and garlic. *Presse Med.* **2002**, *31* (30), 1416–1422.
52. Vang, O.; Frandsen, H.; Hansen, K.T.; Sorensen, J.N.; Sorensen, H.; Andersen, O. Biochemical effects of dietary intakes of different broccoli samples: I. Differential modulation of cytochrome P-450 activities in rat liver, kidney, and colon. *Metabolism* **2001**, *50* (10), 1123–1129.
53. Wilkinson, G.R. The effects of diet, aging and disease-states on presystemic elimination and oral drug bioavailability in humans. *Adv. Drug Deliv. Rev.* **1997**, *27* (2–3), 129–159.
54. Bekersky, I.; Dressler, D.; Mekki, O. Effect of time of meal consumption on bioavailability of a single oral 5 mg tacrolimus dose. *J. Clin. Pharmacol.* **2001**, *41* (3), 289–297.
55. Marathe, P.H.; Greene, D.S.; Lee, J.S.; Barbhuiya, R.H. Assessment of effect of food, gender, and intra-subject variability in the pharmacokinetics of avitriptan. *Biopharm. Drug Dispos.* **1998**, *19* (3), 153–157.
56. Marathe, P.H.; Greene, D.S.; Kollia, G.D.; Barbhuiya, R.H. Evaluation of the effect of food on the pharmacokinetics of avitriptan. *Biopharm. Drug Dispos.* **1998**, *19* (6), 381–394.
57. Marathe, P.H.; Sandefer, E.P.; Kollia, G.E.; Greene, D.S.; Barbhuiya, R.H.; Lipper, R.A.; Page, R.C.; Doll, W.J.; Ryo, U.Y.; Digenis, G.A. In vivo evaluation of the absorption and gastrointestinal transit of avitriptan in fed and fasted subjects using gamma scintigraphy. *J. Pharmacokin. Biopharm.* **1998**, *26* (1), 1–20.
58. Laitinen, K.; Patronen, A.; Harju, P.; Lovtyniemi, E.; Pylkkanen, L.; Kleimola, T.; Perttunen, K. Timing of food intake has a marked effect on the bioavailability of clodronate. *Bone* **2000**, *27* (2), 293–296.
59. Martinez, M.N.; Amidon, G.L. A mechanistic approach to understanding the factors affecting drug absorption: A review of fundamentals. *J. Clin. Pharmacol.* **2002**, *42* (6), 620–643.
60. Adkin, D.A.; Davis, S.S.; Sparrow, R.A.; Huckle, P.D.; Phillips, A.J.; Wilding, I.R. The effect of different concentrations of mannitol in solution on small intestinal transit: Implications for drug absorption. *Pharm. Res.* **1995**, *12* (3), 393–396.
61. Adkin, D.A.; Davis, S.S.; Sparrow, R.A.; Huckle, P.D.; Phillips, A.J.; Wilding, I.R. The effects of pharmaceutical excipients on small intestinal transit. *Br. J. Clin. Pharmacol.* **1995**, *39* (4), 381–387.
62. Basit, A.W.; Newton, J.M.; Short, M.D.; Waddington, W.A.; Ell, P.J.; Lacey, L.F. The effect of polyethylene glycol 400 on gastrointestinal transit: Implications for the formulation of poorly-water soluble drugs. *Pharm. Res.* **2001**, *18* (8), 1146–1150.
63. Hugger, E.D.; Novak, B.L.; Burton, P.S.; Audus, K.L.; Borchardt, R.T. A comparison of commonly used polyethoxylated pharmaceutical excipients on their ability to inhibit P-glycoprotein activity in vitro. *J. Pharm. Sci.* **2002**, *91* (9), 1991–2002.



PHARMACOPEIA STANDARDS: *EUROPEAN PHARMACOPOEIA*^a

Agnès Artiges

European Directorate for the Quality of Medicines (EDQM), Council of Europe, Strasbourg, France

INTRODUCTION

The purpose of a pharmacopoeia and particularly of the *European Pharmacopoeia* is to promote public health by providing common standards recognized by health authorities and all those concerned with the quality of medicines. Such standards are to be of appropriate quality as a basis for the safe use of medicines by patients and consumers. Their existence facilitates the free movement of medicinal products in Europe and ensures the quality of medicinal products exported from Europe.

As noted previously (1), the status of the *European Pharmacopoeia* is based on the existence of an international convention created under the aegis of the Council of Europe (2). The primary characteristics and goals of the *European Pharmacopoeia* are described in this convention. The following chapters present its evolution in the area of the quality of medicines to fulfill the needs of European and International Harmonisation for both regulatory authorities and industries.

THE THIRD EDITION AND ITS CHARACTERISTICS

The *European Pharmacopoeia* plays an important role, not only for well known products manufactured for many years but also for new types of medicines by:

1. elaborating unified specifications for substances from different sources;
2. publishing validated methods;
3. providing common reference substances;
4. producing monographs that are clear (description of impurities controlled by the monograph); and
5. providing a forum for users and other organizations.

The 3rd Edition was published in 1996, and an annual supplement is published in midyear (supplement 1998 in 1997, supplement 1999 in 1998, supplement 2000 in 1999).

^aThroughout the text of this article, the European spelling *pharmacopoeia* is used

All new editions of the *European Pharmacopoeia* now include the year their texts entered into force in their titles; this should make them easier to use. Each supplement is cumulative and replaces the previous one and includes on average approximately 100 new monographs and more than 150 revised texts.

It is available in both print and electronic versions that are published by the European Directorate for the Quality of Medicines (EDQM) of the Council of Europe (3).

Content

The 3rd Edition includes an introductory chapter (General Notices), a chapter on general methods and texts and reagents, followed by monographs in alphabetical order on all type of substances used in the preparation of drug products including excipients (4) and herbal drugs (5). A few monographs on preparations such as radiopharmaceuticals, vaccines, and some hormone preparations (e.g., insulin) are included. There is also a monograph on homeopathic preparations.

The last chapter of the book is focused on general monographs for dosage forms.

We comment here on the main categories of monographs.

General notices

This introductory chapter summarizes the characteristics of the legalities of the *European Pharmacopoeia* and gives the principal definitions. It explains the legal significance of each part of the *Pharmacopoeia* and the role of each section of a monograph. This chapter is essential reading when reference must be made to a monograph of the *European Pharmacopoeia*.

Monographs on chemical substances

All such monographs are presented in the same format, which is described in the "Technical Guide for the Elaboration of Monographs of the European Pharmacopoeia" (6). This guide contains the same concepts as those defined in Community guidelines, which themselves contain the guidelines adopted jointly at the international level, via the International Conference on Harmonisation (ICH). Hence, the two guidelines on analytical validation

are integrated into this technical guide and are even supplemented by specific chapters on the principal methods of analysis (such as spectrophotometry and liquid chromatography, etc.).

The ICH guideline on impurities is also integrated into the technical guide, and the guideline on residual solvents has been integrated into a general chapter of the *European Pharmacopoeia*.

European Pharmacopoeia monographs on chemical substances have therefore been modified in connection with these changes for better control of the impurity profile of substances produced by numerous manufacturers using diverse methods of synthesis. Each revised or new monograph now contains an impurities section at the end that describes the list of impurities known to be detectable by the monograph. Whenever necessary, this impurities section consists of two parts: the list of qualified impurities and the list of impurities that can be detected analytically by the monograph but that are not qualified according to the ICH guideline. The list of impurities includes both the chemical nomenclature and the graphic formula, which makes the section easier to use.

In addition, the presentation of monographs has now been supplemented by the establishment of a procedure for Certification of Suitability of Monographs (see later), thus fully satisfying the requirements of Directives 75/318/EEC for medicines for human use (7) and 81/852/EEC for medicines for veterinary use (8), and the EU note for guidance, "Summary of Requirements on Active Substances in Part II of the Dossier" (9).

Finally, the identification section of these monographs has also been modified to clarify the presentation of monographs that give alternative series of identification tests. The status of this section is clearly defined in the general notices. They specify that when there are two identification series, the first, more complete series will be fully implemented by manufacturers of active substances; the second series is an alternative with less sophisticated analytical methods, but its use requires that the product be traceable from the manufacturer (which carried out all the methods described in the first series) to the user.

Monographs on biological substances

All types of biological medicines such as hormones (10), vaccines for human use (11–13), vaccines for veterinary use (14), and blood products (15) are covered in specific monographs.

Important conceptual changes have also been made to the establishment of this type of monograph to satisfy the needs of Community licensing and to keep up with progress in that field. The following changes merit special attention:

1. the introduction of a production section;
2. the replacement, whenever possible, of tests involving the use of laboratory animals; and
3. the elimination of the test for abnormal toxicity and its replacement by the test for endotoxins (LAL).

The role of the production section is described in the General Notices. The requirement described in this section applies primarily to the manufacturers of the substance in question and to the body of inspectors responsible for checking compliance with the prescriptions of the *European Pharmacopoeia* or with the information given in the licensing dossier.

The tests described in this section cannot necessarily be carried out on the finished product by outside analysts, as with the tests described in the sections on identification, tests, and assay. Nevertheless, they play a major role to guarantee the quality of the substances in question.

In recent years, the European Pharmacopoeia Commission has elaborated a policy of replacing the use of animals in quality control testing of medicines in parallel with the application of the corresponding Convention of the Council of Europe. A sizeable program has been set up to apply the 3-R concept (refine, reduce, replace). To this end, the Council of Europe, represented by the EDQM, and the Commission of the European Communities are now working on an extensive standardization programme (16) to set up collaborative studies to:

1. evaluate, develop, and improve the standardization of test methods for biologicals;
2. prepare European working standards;
3. apply the 3-R concept to replace the use of laboratory animals; and
4. continue the harmonization of test methods for biologicals in Europe and, if possible, the world, in collaboration with the World Health Organization (WHO).

These collaborative studies have led to the establishment of European working standards. Consequently, the titers and potencies of biological products will be expressed with respect to the same reference standard. The existence of reference standards recognized throughout Europe enables national control agencies and manufacturers to avoid costly duplications of work on secondary standards that could otherwise lead to disagreements.

Collaborative studies are also aimed at the validation of alternative reference methods. Comparative tests of various analytical or operating procedures can be used to validate a method of choice or even to establish a close correlation among a method involving tests on

animals, an in vitro biological method, and a method based on physicochemical analysis, thus facilitating the replacement of one method by another in the future.

To facilitate communication and understanding among partners, the results of collaborative studies are published in special issues of *Pharmeuropa* (17).

Monographs on dosage forms

A new chapter has been introduced that brings together all the monographs describing dosage forms, and whenever necessary, the monographs have been supplemented by technological tests and harmonized so that the chapter constitutes a coherent whole.

It should be noted that this chapter and the previous EEC guideline published in 1991 on "authorised terms for dosage forms, routes of administration and containers" were revised together by the European Pharmacopoeia Commission at the request of the Commission of the European Communities.

Thus, both tasks were carried out in parallel and coherently: on the one hand, the revision of all the monographs of the *European Pharmacopoeia* and on the other hand, the revision of the Community guideline on authorized terms.

This guideline has been replaced by a revised version elaborated in the *European Pharmacopoeia* called "Standard Terms" (18); it has been translated into all the languages of the Community. Indeed, not only does this document give the terms in the languages of the Community, but it also includes terms in the national languages of several delegations to the *European Pharmacopoeia* that are not members of the European Union but that also wish to provide a translation in their language. The revised document will therefore list terms in 21 European languages (Bulgarian, Croatian, Czech, Danish, Dutch, English, Finnish, French, German, Greek, Hungarian, Icelandic, Norwegian, Italian, Polish, Portuguese, Slovak, Slovenian, Spanish, Swedish, and Turkish).

This relatively large document, which covers both human and veterinary medicines, is published as a special issue of *Pharmeuropa*. The first version was published in November 1996, then revised yearly; the next revised version was published in February 2000. This will produce a harmonious and coherent whole that can be used throughout Europe. The terms are mandatory for applications and summaries of product characteristics for EU centrally and decentrally authorized products.

The list can be extended on request if justification is provided. A specific procedure has been set up (18). Forms are also available on the EDQM Web site.

General monograph on methods of manufacture

Another way to address the quality of the monograph is to apply the concept developed in the 1980s and 1990s, which is that one cannot control the quality of a product simply by testing the finished product. One is controlling only what is being sought. The quality of a product has to be included from the very beginning and maintained throughout the manufacturing process.

To avoid repeating key points of policy in its texts, the European Pharmacopoeia Commission's current approach is to prepare general monographs that cover all the specific monographs but that can also be referred to for substances that have no monograph in the *European Pharmacopoeia*. Thus, new general monographs on the method of production have been prepared, such as the monograph "Products of Fermentation" (19), "Products of r-DNA Technology" (20), and "Products with Risk of Transmitting Spongiform Encephalopathy" (21).

In addition, compilers of the *European Pharmacopoeia* have been asked to add a production section to the monographs for specific substances. The production section gives key points but not all the details that refer to a specific manufacturer. In addition to the production section in specific monographs, general concepts in line with new ICH guidelines are also identified. Such details are described in the application for a Certificate of Suitability by each manufacturer (see above).

Revision of Monographs

The European Pharmacopoeia Commission is very attentive to updating its monographs. Revisions can be made at any time if a pharmacopoeial or licensing authority requests a particular change. The European Pharmacopoeia Commission has defined its criteria for revision of monographs when:

1. poor quality products appear on the market;
2. a request is made in connection with the certification procedure;
3. there is a public health risk (presence of nonqualified impurities, risk of falsification, etc.);
4. the analytical methods are no longer adequate;
5. certain tests in a monograph are no longer applicable;
6. reagents are unavailable; and
7. the patent is close to the expiry date.

In addition, a more extensive program of revisions corresponding to new developments in methods is carried out every 5 years. Such revisions take account of new guidelines from the licensing authorities and the need for harmonization within families of substances.

How and by Whom Are Texts Elaborated?

The *European Pharmacopoeia* is elaborated by a Commission made up of national delegations from regulatory authorities, with decisions requiring a unanimous vote by the delegations (2). (Today, the Convention has been signed by 27 European countries^b including all EU members and by the EU Commission itself. Nine other European^c and eight Non-European countries are also observers.)

The monographs are elaborated by the groups of experts appointed by the European Pharmacopoeia Commission based on proposals by the national delegations. The experts participating in this work are from industry, universities, and national control laboratories.

The *European Pharmacopoeia* has progressively been replacing the national pharmacopoeias. Nevertheless, we need to emphasize the major role played by the national pharmacopoeia secretariats in the elaboration of the common European work. Indeed, they are in the best position to identify national needs and to organize:

1. consultation with the pharmaceutical, chemical, and biological industries that manufacture in their territories (consolidated comments sent to the secretariats of the *European Pharmacopoeia*);
2. consultation with other governmental organizations impacted by the work (e.g., ministries of industry or agriculture); and
3. collaboration with other departments in medicine-related agencies such as licensing authorities, national control laboratories, and inspection.

Finally, the national secretariats play an important role in providing information at the national level on how European rules are elaborated, and any national text is prepared and revised in compliance with the *European Pharmacopoeia*.

Before final adoption, all texts and monographs are published in *Pharmeuropa* for 4 months of public inquiry. For European countries, comments should be sent through the national secretariat; for non-European companies, comments should be sent directly to the EDQM-Council of Europe (B.P. 907, F-67029 Strasbourg Cedex, France). *Pharmeuropa*, the users's forum is a quarterly publication prepared and published by the EDQM.

^bAustria, Belgium, Bosnia-Herzegovina, Croatia, Cyprus, Czech Republic, Denmark, Finland, France, The Former Yugoslav Republic of Macedonia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom of Great Britain and Northern Ireland and the European Union.

^cAlbania, Bulgaria, Estonia, Latvia, Lithuania, Malta, Poland, Rumania, Ukraine.

Implementation by the Parties

The *European Pharmacopoeia* is a common supranational document that supersedes any national texts for the signatory parties. A common implementation date is adopted by the European Pharmacopoeia Commission for each supplement/edition (or individual text for rapid revision) that enters into force by means of a resolution of the Council of Europe/Public Health Committee.

Although the signatory parties are bound by the texts and specifications of the *European Pharmacopoeia* published by the Council of Europe, they are free to implement them in ways that are compatible with their technical, legal, and administrative regulations. Thus, different procedures are used in different countries.

A majority of members implement directly the volume published by the Council of Europe; others continue to issue a national pharmacopoeia that republishes all or some of the harmonized European texts translated if necessary into the national language or style (Austria, Bulgaria, the Czech Republic, Germany, Greece, Hungary, Portugal, Spain, Switzerland, and the United Kingdom). In all cases, it is the European text that is implemented and made legally binding, superseding any translation in cases of doubt.

CERTIFICATION OF SUITABILITY OF MONOGRAPHS

This unprecedented procedure for Certification of Suitability of Monographs of the *European Pharmacopoeia* was established a few years ago (22). Why? For centuries, pharmacopoeias were designed to be references that were complete and obligatory in themselves. However, in the 20th century, with the growth of world trade, the EU licensing authorities found that they increasingly had to ask manufacturers to provide in their licensing applications complete details on the synthesis of their product to demonstrate that the product was suitably controlled by the *European Pharmacopoeia* monograph. Thus, it then became apparent that it was absolutely necessary to reconsider how this essential information could be made available to those who needed it without duplication of work by the *Pharmacopoeia* compilers and licensing authorities.

To solve this problem, it was necessary not only to adapt the content of monographs to new needs but also to set up a procedure for Certification of Suitability that would establish a link between licensing and the *Pharmacopoeia* in this area.

As noted above, the content of monographs was supplemented where relevant by production and impurities sections to make these monographs more complete and clear (23).

The certification procedure is a complement and bridge between the public standards described in the *European Pharmacopoeia* and the need to prepare a file for licensing. This procedure is a result of much common discussion and agreement among the partners concerned. It was in fact made to measure its collaboration not only with the European regulatory authorities so that they could rely on it totally and recognize unreservedly its validity but also with the industries so that they could be absolutely sure of the protection of industrial property.

On the basis of the data collected during the elaboration of the monograph and the specific data provided by a specific manufacturer on a specific substance, the Certificate of Suitability certifies that both types of data make it possible to conclude that the quality of the substance corresponds to the quality defined in the *European Pharmacopoeia* monograph.

In principle, a certificate can be granted for any substance (active substances, excipients) such as organic or inorganic substances, substances produced by fermentation as indirect gene products, and products with risk of TSE for which a monograph published in the *European Pharmacopoeia* exists. Excluded, however, are biological substances such as proteins, products obtained from human tissues, vaccines, blood products, and preparations.

The certificate, granted for 5 years, may include additional specifications (methods and limits) when monograph specifications do not fully control the purity of the substance (e.g., control of residual solvents, specific impurities, etc.).

The procedure is described in Resolution AP-CSP (99) 4, together with the content of the file to be submitted (24). This procedure, lists of certificates granted (420 at the end of January 2000), and lists of assessors are published regularly in *Pharmeuropa* and on the EDQM Web site (<http://www.pheur.org>, direct link with the Certification Unit: cert@pheur.org).

Certificate of Suitability Versus European Drug Master File

Although both procedures have the same aim and require dossiers with exactly the same contents, the certification procedure is especially designed to cover substances for which there is a monograph in the *European Pharmacopoeia*, whereas the European Drug Master File procedure is aimed at substances for which there is no *European Pharmacopoeia* monograph.

In conclusion, a procedure makes it possible to avoid duplication of work not only by manufacturers of raw materials and manufacturers of medicines (finished products) when they prepare licensing dossiers but also by the licensing authorities and pharmacopoeia authorities when they assess these dossiers. Differences among the various European licensing authorities in approach and assessment of compliance with *European Pharmacopoeia* monographs are also avoided, and clearer communications are facilitated. Finally, the procedure allows *European Pharmacopoeia* monographs to be constantly updated to keep up with new developments in the world market.

TOWARD THE 4TH EDITION

European Pharmacopoeia authorities have decided to publish a fourth edition. For this purpose, both the technical guide and the style guide have been revised, and more general monographs have been or will be elaborated.

Technical Guide

This guide (6) is intended for the experts who participate in the elaboration of *European Pharmacopoeia* monographs; it establishes the general rules to be followed. This document specifies the philosophy behind the choice of techniques for identification testing of a substance and for the determination of the limit contents of impurities and the methods used to detect them.

The 2nd Edition had been revised particularly to be in line with the ICH guideline. The 3rd Edition was published in February 2000.

Style Guide

This is an internal guide for professionals of the *European Pharmacopoeia* and for the secretariat to make the monograph style more uniform in English and French, the two official languages of the Council of Europe. The aim is to provide the means of drafting clear, unambiguous texts. The style, which will appear in the 4th Edition, is more telegraphic than that used in the 3rd Edition.

General Monographs

The fundamental goal of the *European Pharmacopoeia* is to promote the harmonization of standards for medicinal products in the member states and in view of this, the European Pharmacopoeia Commission considers that it is desirable and possible to extend the scope of general

monographs to encourage a convergence of approach by licensing authorities and thus avoid future difficulties in harmonization.

The *European Pharmacopoeia* contains a number of general monographs that cover categories of products defined by the:

1. presentation of the medicinal product (dosage form monographs such as "Tablets," "Eyedrops," etc.);
2. nature of the product (radiopharmaceutical preparations, vaccines for human use, etc.); and
3. methods of production (products of fermentation, products of rDNA technology, products with TSE risks).

New general monograph on substances for pharmaceutical use (active substances and excipients) (25) had been prepared to include in the *European Pharmacopoeia* the ICH guideline for residual solvents and the ICH qualification threshold for new impurities without revising each individual monograph concerned. In addition, the monographs are used to simplify the *Pharmacopoeia* requirements concerning sterility, bacterial endotoxins, and pyrogens tests contained in individual monographs. However, because it is the policy of the Commission to add these requirements whenever it is considered to be appropriate, a general regulation will now replace the regulations in individual monographs.

Furthermore, the monograph gives a definition of the terms "active substance" and "excipients"; it explains the policy of the Commission concerning polymorphic forms, active substances of special grade, processing of active substances with and without the addition of excipients, and the two sets of identification tests that may be contained in a monograph.

The status of these general monographs will be defined in the chapter, "General Notices."

International Harmonization with USP and JP

This review will not be completed without highlighting the close relationship that has developed since 1990 among the European, Japanese, and U.S. pharmacopoeias. They cofounded the Pharmacopoeial Discussion Group, which is working diligently for harmonization at the world level and which participates in the ICH program (26–28). This group meets regularly (twice a year) in Europe, Japan, and the United States. Approximately 50 monographs on excipients and 20 general methods of analysis proposed by national associations of manufacturers of pharmaceutical products have been selected for convergence and harmonization in the three pharmacopoeias.

A special section of the *European Pharmacopoeia* quarterly journal *Pharmeuropa* is now dedicated to this activity. Joint open conferences organized by the five pharmacopoeias, in Verona, Italy (on biotechnology products in April 1993); St. Petersburg, FL (on excipients in 1994); Barcelona, Spain (on microbiological tests in 1996); Seville, Spain (on dosage-form pharmacotechnological tests in 1998); and Strasbourg, France (on new trends in biologicals in 1999) regularly brought together specialists from all over the world.

REFERENCES

1. Artiges, A. Pharmacopoeial Standards: European Pharmacopoeia. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J., Eds.; Marcel Dekker, Inc.: New York, 1995; 12, 53–71.
2. Convention on the Elaboration of a European Pharmacopoeia. *European Treaty*; Series No. 50, Council of Europe: Strasbourg, France.
3. *European Pharmacopoeia*, 3rd Ed., Council of Europe: Strasbourg, France.
4. Artiges, A. Excipients: Aspects Réglementaires. *Thérapie* **1999**, *54*, 15–19.
5. Artiges, A. Pharmacopoeial Standards for Herbal Medicinal Products in Europe, Proceedings of the 5th ESCOP International Symposium.
6. *Technical Guide for the Elaboration of Monographs*, 3rd Ed.; Council of Europe: Strasbourg, France, December 1999; *Pharmeuropa* Special Issue.
7. *The Rules Governing Medicinal Products in the European Union*, 5, Council Directive 75/318/EEC, Amended.
8. *The Rules Governing Medicinal Products in the European Union*, 5, Council Directive 81/852/EEC, Amended.
9. *The Rules Governing Medicinal Products in the European Union*, Guidelines on Medicinal Products for Human Use, 3.
10. Charton, E. Hormones: The Role of the European Pharmacopoeia. Alternatives to Animals in the Development and Control of Biological Products for Human and Veterinary Use. *Dev. Biol. Standardisation* (Basel, Karger) **1999**, *101*, 159–167.
11. Castle, P. Policy and Progress of the European Pharmacopoeia in the Use of Alternatives of Animal Testing in Vaccine Production and Quality Control. *Alternatives to Animal Testing in the Production and Quality Control of Vaccines: Present Practice and Prospectives*; RIVM: Bilthoven, 1992; 41–52.
12. Castle, P. Alternatives to Animal Testing: Achievements and Recent Developments in the European Pharmacopoeia. *Development in Biological Standardization*; PEI: Langen, 1994; 86, 21–29.
13. Castle, P. The European Pharmacopoeia and Humane End Points, International Conference on the Use of Humane Endpoints in Animal Experiments for Biomedical Research, Zeist, NL, November, 23–25, 1998; 86.

14. Artiges, A. The Role of Pharmacopoeias. *Veterinary Vaccinology*; Part. 9, Pastoret, et al. Eds.; 708–711.
15. Artiges, A. Quality and Safety of Plasma Products: Control Authority Batch Release Within the Countries of the European Union and Economic Area, Proceedings of EPFA-EAPPI, 5th Annual European Regulatory Affairs Symposium, Vienna, Austria, September, 29–30, 1998, 108–110.
16. The Biological Standardisation Programme. *Pharmeuropa Bio*; Council of Europe: Strasbourg France, July 1996; Special Issue 96–1.
17. *Pharmeuropa Bio* Council of Europe: Strasbourg, France, Special Issues 96-1, 96.2; Special Issues 97-1, 97.2; Special Issue 98.1; Special Issues 99–1, 99.2.
18. Standard Terms: Pharmaceutical Dosage Forms, Routes of Administration, Containers. *Pharmeuropa*; Council of Europe: Strasbourg, France, January 2000; Special Issue.
19. *European Pharmacopoeia*, 3rd Ed.; Monograph No. 1468, 1111–1112, Supplement 2000.
20. *European Pharmacopoeia*, 3rd Ed.; Monograph No. 0784, 1435–1438.
21. *European Pharmacopoeia*, 3rd Ed.; Rapid Implementation Monograph No. 1483, Resolution AP-CSP (99) 5, *Pharmeuropa* 12 (2).
22. Helboe, P. Certification Procedure: Purposes and Procedure The Feed-Back Mechanism, Proceedings of the International Conference on the Vision of the European Pharmacopoeia in the 21st Century, July, 1997, *Pharmeuropa*, Special Issue, Workshop Session II.
23. Artiges, A. Certification of Suitability of Monographs of the European Pharmacopoeia: Summary of Objectives and Scope of the Procedure, Proceedings of the International Conference, Berlin, Germany, November 11–12, 1999, *Pharmeuropa* Special Issue 2000.
24. Certification of Suitability of Monographs of the European Pharmacopoeia. *Pharmeuropa* **2000**, 12 (2), Resolution AP-CSP (99) 4.
25. Monographs on “Active Substances” And “Excipients.” *Pharmeuropa* **1999**, 11 (3).
26. Artiges, A. International Harmonisation of Pharmacopoeias ICH, Proceedings of the First International Conference on Harmonisation, Brussels, 1991, 143–152, Topic 3: Pharmacopoeias.
27. Halperin, J.A. Session 5, Harmonisation of Pharmacopoeial Monographs and Methods, Proceedings of the Second International Conference on Harmonization (ICH), Orlando, 195–217.
28. ICH Guideline on Specification on Drug Substances and Drug Products, (Q6A-paragraph 2.8).

PHARMACOPEIAL STANDARD: JAPANESE PHARMACOPOEIA^a

Mitsuru Uchiyama

Japan Pharmacists Education Center, Tokyo, Japan

INTRODUCTION

The *Pharmacopoeia of Japan* (JP), which dates back to 1886, provides the official standards and test methods for regulating the properties and quality of drugs important for medical treatment. In general, the standards set forth in the JP affect not only those articles included in the JP but all drugs and drug products in circulation in Japan. The JP has played an important role in ensuring and improving drug quality through all phases of development, application, evaluation, distribution, inspection, and consumption before and after manufacture. Thus, revision of the Pharmacopoeia must occur within a structure and system that are unaffected by bias and undue influence of any party for its standards to be recognized as credible and reliable for ensuring drug quality.

FRAMEWORK

The JP is currently published by the Ministry of Health and Welfare through the Committee on the JP of the Central Pharmaceutical Affairs Council (CPAC), pursuant to the Pharmaceutical Affairs Law. The Committee on the JP includes a subcommittee consisting of 12 advisory panels, each concentrating on one of the following subject areas: principles of revision; selection of articles; medicinal chemicals; biologics and biologicals; general test methods; physical test methods; biological test methods; preparations; crude drugs; nomenclature; and pharmaceutical excipients.

The panel members are scientists from national institutes, universities, and prefectural laboratories, who are appointed every 2 years by the Minister of Health and Welfare. As a rule, a panel member is not appointed for more than 8 consecutive years. A certain number of liaison members from industry serve on 10 of the 12 panels. In contrast, the panels responsible for nomenclature and selection of articles do not include liaison members from industry.

^aThroughout the text of this article, the European spelling *pharmacopoeia* is used.

The JP secretariat is part of the Evaluation and Licensing Division of the Pharmaceutical and Medical Safety Bureau of the Ministry. The secretariat works under the scientific and technical support of the National Institute of Health Sciences (NIHS). The Pharmaceutical and Medical Safety Bureau is a new organization, established in July 1998, as a result of a reorganization of the former Pharmaceutical Affairs Bureau. The CPAC itself will be terminated in January 2001 in response to further reorganization of government offices and will resume operation as the council responsible for both pharmaceutical and food sanitation affairs.

REVISION PROCESS

Revision Cycle

The Pharmaceutical Affairs Law stipulates that the JP be revised at least once every 10 years. However, beginning in 1967, because of the rapid progress and changes in medicinal and pharmaceutical research and development, the revision cycle of the JP has been reduced to once every 5 years. Thus, the 11th and 12th editions were published in 1986 and 1991, respectively, and the most recent edition, the 13th, was published in April 1996. An English version has subsequently been released. In the event of a discrepancy between the Japanese original and its English translation, the former is considered official. The English version of the JP is available through Yakuji Nippo, Ltd. (Kanda Izumicho 1, Chiyoda, Tokyo 101-8648, Japan) (1).

A supplement, the first in JP history, was published in October 1988, 2 years after publication of the 11th edition. The Ministry published two supplements to the 12th edition in October 1993 and December 1994, respectively. The first and second supplements to the 13th edition became available in December 1997 and December 1999, respectively. The 14th edition will be published in April 2001.

Principle of Revision

The current fundamental goal and most important consideration of the JP revision, to which the evaluation

of new drug quality also conforms, is to obtain pharmaceuticals of consistent quality, but not necessarily of higher purity, to provide maximum benefit to the consumer. The specification should be sufficient to at least identify and ensure adequate quality, and the analytical procedures should be accurate and easy to perform. However, a proper balance between accuracy and ease of use needs to be maintained to allow for this analysis. Although extremely high accuracy is not necessary, the limit of detection or the limit of quantitation, as well as the recovery in purity tests, should be validated by sufficient data. Hazardous chemicals should be avoided, and experimental testing on animals should be minimized.

Revision Process

Proposals for revision of the JP monographs or test methods can be submitted through the JP secretariat in the Pharmaceutical and Medical Safety Bureau by any concerned individual or organization. The proposed monograph, either for a new entry or for a revision, is then drafted by a panel member or an industry professional. New or revised test methods are handled in a similar manner. To accomplish this, panels are assisted by the Japanese Pharmaceutical Manufacturers' Associations of Tokyo and of Osaka, the Japan Pharmaceutical Excipients Council, the Crude Drugs Association of Tokyo, the Federation of Crude Drugs Association of Japan, and the Japanese Society of Hospital Pharmacists.

The draft monograph is reviewed by the two panels on Medicinal Chemicals. The Panel on Medicinal Chemicals I adopts new entries, whereas the Panel on Medicinal Chemicals II revises existing monographs. The drafts for traditional medicines and pharmaceutical excipients are reviewed by the Panel on Crude Drugs and the Panel on Pharmaceutical Excipients, respectively. Nomenclature and chemical structures are examined by the Panel on Nomenclature once the draft monograph has been finalized. Additions or changes to the three general test methods (chemical, physical, and biological) and reagents are addressed by the corresponding panel for each test method.

The establishment of reference standards is adopted at the suggestion of the related panels after due consideration of the opinion of the NIHS, which is directly responsible for establishing those reference standards in cooperation with the Society of Japanese Pharmacopoeia. The Society of Japanese Pharmacopoeia (Shibuya 2-12-19, Shibuya, Tokyo 150-0002, Japan) is a nonprofit private organization that carries out activities in support of MHW administration and regulation of pharmaceuticals. The Society

plays several roles and, in particular, makes reference standards available including a part of the pharmacopoeial reference standards. It also distributes *Pharmacopoeia* and related informative documents, such as the *JP Forum*, and convenes public meetings and symposia.

JP Forum

After finalization by each panel, all drafts for revision are opened for public comment through the *JP Forum*, a vehicle for notification and commentary (2). The *JP Forum* is published quarterly by the Society of Japanese Pharmacopoeia under the auspices of the JP secretariat. The *JP Forum* was first published in January 1992 as a medium for both local and international communication. It will continue to provide a more open revision process for the JP and announce revisions and future directions of the JP committees. The *JP Forum* is issued in Japanese, primarily for domestic users, but articles related to the international community and commentary will be issued in English concurrently with the Japanese version. Therefore, we anticipate that the *JP Forum* will aid in the process of international harmonization and will promote a better understanding of and increased trust in the JP.

After comments are reviewed and changes are made, the revised draft is reviewed by the Committee on the Japanese Pharmacopoeia and then by the Executive Committee of the Central Pharmaceutical Affairs Council before it is submitted for publication in the JP.

Time Frame

Regarding the interval between new drug approval and the adoption of a pharmacopoeial monograph in Japan, the first monograph generally does not appear in the JP until publication of the outcome of its reevaluation. The term of current reevaluation for a new active ingredient is generally 6 years after approval. The reevaluation period is needed to ensure the safety and efficacy of new drugs. For a particular drug, such as an orphan drug, the reevaluation period is 10 years after approval.

In addition, not all new drugs are listed in the JP, and the Panel on the Selection of Articles decides which drugs should be listed based on their importance in terms of medical treatment. The innovators sometimes do not want their drugs to be listed in the JP, concerned that such listings will trigger the introduction of generic brands. As a result, the JP contains fewer monographs than does the *U.S. Pharmacopoeia–National Formulary* (USP–NF). The JP 13 contains 1292 monographs, which is approximately one-third the number included in the USP–NF.

SIZE, SCOPE, AND PRESENTATION OF JP 13

The JP 13 is a single volume comprising two parts. Part 1 includes 824 monographs of widely used drug substances and their preparations, for 532 organic and inorganic chemical ingredients, 187 preparations (single-ingredient dosage forms), 91 antibiotics, 11 radiopharmaceuticals, and 3 medicinal gases. Part 2 includes 468 articles, for 106 mixed preparations, 132 pharmaceutical excipients, 172 traditional (crude) drugs, 30 biologics, 21 miscellaneous substances of plant or animal origin, and 7 surgical dressings. In addition to the monographs, the JP contains sections on notices of importance, rules, standards, test methods, and apparatus commonly applicable to any of its articles. There are 131 JP Reference Standards.

The General Notices in the JP provide specific definitions. Some examples are:

1. definitions of standard temperature, ordinary temperature, room temperature, and lukewarm as 20, 15–25, 1–30, and 30–40°C, respectively;
2. the term “in vacuum” indicating, unless otherwise specified, a pressure not exceeding 15 mm Hg;
3. the tabulation of the degree of coarseness or fineness of a powdered medicine; and
4. that in the monograph, if the upper limit of the content of an ingredient determined by assay is not specified but expressed simply as not less than a certain percentage, 101.0% should be understood as the upper limit.

The dosage form monograph in the JP has a section on methods of preparation that refers to the General Rules for Preparations, which gives definitions, methods of preparation, storage, and other information on 28 different dosage forms. The insoluble particulate matter test, included as a subsection in the section Injection, is currently under discussion among pharmacopoeia compilers for the purpose of establishing international harmonization.

The General Tests, Processes, and Apparatus section contains 60 test methods, such as fluorometry and electrometric titration, dissolution and disintegration tests, content uniformity, and a bacterial endotoxin test. Qualitative tests, reagents and test solutions, and standard solutions are also included. The Infrared Reference Spectra of 124 chemical entities are included in an appendix and will be updated in subsequent editions.

The JP monographs consist of the drug name, description, identification, rational values, and purity, as well as special tests if any, tests for preparation, and assay. No specifications are given in the JP 13 under monographs

for antibiotics, biologics, and radiopharmaceuticals. However, they may be found in Requirements for Antibiotic Products of Japan 1993 (formerly the Japanese Minimum Requirements of Antibiotic Products) and other corresponding compendia.

NONPHARMACOPOEIAL STANDARDS FOR PHARMACEUTICALS IN JAPAN

In addition to the *Pharmacopoeia of Japan*, there are several compendia, standards, and guides in which Japanese standards for pharmaceuticals are published. The following are standards determined for drugs under the provisions noted above, although the specifications and standards included in these compendia are mandatory. As presented above, some of the monographs in the following compendia are quoted in the JP:

- Requirements for Antibiotic Products of Japan 1993 (to be combined with the JP; antibiotic substance will be included in JP 14, due for release in 2001, and antibiotic products will be compiled in the Japanese Pharmaceutical Codex)
- Minimum Requirements for Biological Products
- Minimum Requirements for Blood Grouping Sera
- Radiopharmaceutical Standards

Several standards are published as advisements from the Pharmaceutical Affairs Bureau that set voluntary standards for various pharmaceuticals to ensure and improve their quality. Because the standards and specifications in these guides are voluntary, there is no overlap with the JP monographs.

- Japanese Pharmaceutical Codex, which contains monographs for 682 active ingredients and 175 pharmaceutical preparations (3)
- Japanese Pharmaceutical Excipients, which has 206 monographs for pharmaceutical excipient
- Standards for Crude Drugs
- Standards for Raw Materials for Clinical Diagnostics
- Guideline for Radiopharmaceuticals for In Vitro Diagnostics
- Insecticide Standards

IMPLICATIONS OF PHARMACOPOEIA IN THE REGULATORY PROCEDURES

Given that ensuring drug quality, efficacy, and consumer safety is a common objective for both

pharmacopoeial standards and drug quality regulation, the basic policy of one is not inconsistent in principle with that of the other. The JP Committee and regulatory agencies will enjoy a more harmonious relationship as they continue to develop closer contacts, both legal and otherwise. The JP secretariat belongs to the same Pharmaceutical and Medical Safety Bureau of the Ministry that oversees new drug approvals. Furthermore, several members of the JP Committee are also members of the NIHS, which shares some responsibility for reviewing new drug applications (NDAs) and for establishing guidelines for the technical requirements governing NDAs. These guidelines are updated whenever the JP is revised. History suggests that the revision of the JP and the revision of guidelines can be easily synchronized by the regulatory agencies that work closely together (4, 5).

Conformity of NDA Dossier with Pharmacopoeia

In the NDA procedures, specifications and methodologies should be based, as much as possible, on the entries in the JP, and careful attention must be given to the terminology used in the JP. In principle, reagents, test solutions, and testing apparatus used in proposed methodologies should be the same as those in the JP. When those used are not in the JP, their quality, formulas, schematics, and dimensions, etc. are to be entered on an attached sheet in accordance with the entries for such items in the JP. If the reference standards used are not specified in the JP or by the NIHS, details of the specifications and test methods must still be attached.

How Revisions of the JP Affect Drugs in a Market

As noted above, pharmacopoeial descriptions are mandatory for all drugs, regardless of whether they are listed in the JP or are in the process of being evaluated as a new drug. It follows, therefore, that revisions of the JP will greatly influence the standards for those pharmaceuticals. Revisions of the JP affects three areas of products already on the market:

1. In cases in which a new pharmacopoeial monograph is established, the NDA holder of a particular drug is required to apply for permission to change approved items, such as specifications and test methods, to ensure that the drug satisfies the pharmacopoeial monograph. A grace period of 1.5 years is allowed under current regulations for those products already

on the market to comply with new pharmacopoeial standards.

2. Whenever the standards in a monograph governing content, properties, testing, and purity limits for purity are revised, all related drug products currently in circulation must be changed to comply with the new standards in the JP. Manufacturers are responsible for ensuring that their products comply with the new standards. If these products do not comply, manufacturers are required to standardize the quality of their products to the level of the new standards within the grace period of 1.5 years.
3. When there are changes in the general notice, general rules for preparations, and specified test apparatus and test conditions in the General Tests, Processes, and Apparatus, previously approved drug products in some cases no longer conform to the specifications for which they were originally approved. For example, if the pH value of the medium in a disintegration test for a solid preparation changes, some products may yield different results under the new pH value requirement. In such a case, changing part of the formulation is recommended so that the product can comply with the specifications under the new test conditions. These situations occasionally present obstacles for later international harmonization of general test methods. It is necessary to overcome such obstacles by taking regulatory measures.

INTERNATIONAL HARMONIZATION

Role of Pharmacopoeia in the International Conference on Harmonization (ICH) Activities

The JP compilers are now part of an international harmonization effort with compilers of the *U.S. Pharmacopoeia* and the *European Pharmacopoeia*. Pharmacopoeias were taken up as an Expert Working Group (EWG) subject at an earlier stage of the ICH. However, pharmacopoeial harmonization should attempt to facilitate the international circulation of drugs and to improve drug quality rather than to simply facilitate the approval of new drugs. Furthermore, the Pharmacopoeial Discussion Group (PDG), which consists of European, U.S. and Japanese pharmacopoeia compilers and which was established in 1989 before the ICH, has held meetings periodically in parallel with the ICH to investigate a wide range of problems related to pharmacopoeia and to work on the harmonization of all

the monograph specifications and general test methods according to a long-term schedule. The PDG agreed with the ICH Steering Committee in that it felt it more appropriate to present a regular progress report at each ICH meeting from the PDG rather than to retain it as an EWG topic.

PDG Policy

To date, in its action plan for harmonization, the PDG has placed a high priority on the monographs of major excipients that can be used for a number of new drug products as well as on important general test methods. The PDG has established its general policy for harmonization, in which it is shown that the goal of harmonization is to bring the policies, standards, monograph specifications, analytical methods, and acceptance criteria of pharmacopoeias into agreement. Nonetheless, the PDG recognizes that such unity may not always be obtained. Where unity cannot be achieved, harmonization means agreement based on objective comparability and a clear statement of any differences. The goal, therefore, is harmony, not unison (6).

In harmonization of analytical methods, the ideal approach is to establish a single method that satisfies the criteria for validation of all pharmacopoeias. However, this is not always possible because of a number of unavoidable differences in the technical circumstances of each nation. When different tests or methods yield the same results, provisions have already been made in the three pharmacopoeias to allow for alternative methods. In such cases, alternative methods should be subjected to validation via a comparison with the standard analytical procedure.

It is undoubtedly impossible to achieve harmonization of quality for new drugs without harmonization of compendial standards and methodology. The PDG will continue to proceed with what it believes to be the correct approach to pharmacopoeial harmonization and to contribute to advancing effective harmonization for the quality of new drugs and products.

Mutual Agreement

In the licensing process, it is essential for each regulatory authority to recognize as equivalent those test procedures that have been harmonized and adopted as validated methods by the PDG. This concept is already included in the Japanese *Guidelines for Preparation of Section B of the Documents Accompanying New Drug Applications* as of September 1995; in its assertion that, "The analytical

procedures in the Japanese Pharmacopoeia and other compendia and those which are accepted through international harmonization are considered to be validated methods."

The ICH step 4 guidelines on specifications, Q6A, October 6, 1999, notes in the item concerning pharmacopoeia that:

References to certain procedures are found in pharmacopoeias in each region. Wherever they are appropriate, pharmacopoeial procedures should be utilized. Whereas differences in pharmacopoeial procedures and/or acceptance criteria have existed among the regions, a harmonized specification is possible only if the procedures and acceptance criteria defined are acceptable to regulatory authorities in all regions. The full utility of this guideline is dependent on the successful completion of harmonization of pharmacopoeial procedures for several attributes commonly considered in the specification for new drug substances or new drug products (7).

In this manner, key objectives of harmonization will be attained only when the PDG is able to achieve mutually agreeable standards and test methods, which provide the same conclusions when performed on the same specimens, even if they use different specifications, procedures, or reagents. We should continue to take necessary steps to deepen international cooperation and to obtain harmonized compendial standards and methodology, using the PDG as the forum for harmonization of drug quality.

CONCLUSION

This article describes the development of JP standards and their international harmonization and the relationship of the JP with drug regulatory procedures in Japan. It also notes that the JP has been playing a basic and leading role in ensuring and improving the quality of drugs in Japan. The pharmacopoeias of every country must constitute the basic technical foundation for the nation's drug quality standards and contribute to the promotion of public health by providing better pharmaceuticals to the public in the most efficient manner. The role of any pharmacopoeia is not only to define objectives for drug quality control based on the establishment of standards, test methods, and acceptance criteria, but also to provide means to implement the international harmonization of pharmaceutical regulation.

REFERENCES

1. Society of Japanese Pharmacopoeia. *The Japanese Pharmacopoeia*, 13th Ed.; Yakuji Nippo, Ltd., Tokyo, 1996; <http://www.yakuji.co.jp>.
2. Society of Japanese Pharmacopoeia. *Japanese Pharmacopoeial Forum*; Society of Japanese Pharmacopoeia: Tokyo, <http://www.sjp.or.jp>.
3. Society of Japanese Pharmacopoeia. *Japanese Pharmaceutical Codex 1997*; Jiho, Inc.: Tokyo, <http://www.jiho.co.jp>.
4. Pharmaceutical Review System Study Group, Ministry of Health and Welfare. *Pharmaceutical Administration in Japan*, 9th Ed.; Yakuji Nippo, Ltd.: Tokyo, 2000; <http://www.yakuji.co.jp>.
5. English Regulatory-Information Working Group. *Pharmaceutical Administration and Regulation in Japan (1999-12)*; Japanese Pharmaceutical Manufacturers Association: Tokyo, 1999; <http://www.nihs.go.jp/mhw/koukai/1999/yakuzi/version5e.pdf>.
6. Pharmacopeial Discussion Group. Policies for Harmonization of the Three Pharmacopoeias. *Japanese Pharmacopoeial Forum* **1995**, 4 (4), 65–67.
7. ICH Expert Working Group, *Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances*; ICH Steering Committee: October 6, 1999; <http://www.nihs.go.jp/dig/ich/qindex-e.html>.

PHARMACOPOEIAL STANDARDS: THE *UNITED STATES PHARMACOPEIA* AND THE *NATIONAL FORMULARY*^a

Lee T. Grady

United States Pharmacopoeia, Rockville, Maryland, U.S.A.

INTRODUCTION

United States Pharmacopoeia (USP) and *National Formulary* (NF) standards and specifications relate to the quality, purity and strength, packaging, and labeling of medicines and related articles. The standards are public standards, and new and revised standards are published regularly. Official drug names and definitions are established, as well as tests or assay procedures that allow determination of compliance with the standards. But do the compendia reflect the state of pharmaceutical technology?

Resolutions adopted by the 1995 USP Convention are indicative of the interest and professional expertise shown by delegates. The General Committee of Revision elected by the Convention included categories of expertise critical to a modern pharmacopoeia. From the 138 standards-related members, 20 Division of Standards Development subcommittees were organized, along with a joint standards-information Nomenclature Committee, to carry out the continuous revision process.

The USP24-NF19 contains 3777 monographs and 164 general chapters (1). There are 543 new monographs, 504 in the USP and 39 in the NF. Continuous revision led to many new or improved requirements; 3941 individual revisions were processed through *Pharmacopoeial Forum* (PF) during the 5-year cycle. Obsolete material deleted during the preparation of this volume included 130 USP and 12 NF monographs, and 4 general chapters.

USP and Legal Recognition

Unlike all other pharmacopoeias, the USP and NF are not produced by government. The USP and NF are published by the U.S. Pharmacopoeial Convention, Inc., a voluntary,

not-for-profit institution that holds the public trust. Standards established in the USP and the NF are recognized by law and can be enforced by federal and state authorities. References to the USP and NF occur in numerous statutes regulating articles used in medical and pharmacy practice. The most significant is recognition of the official compendia in the Federal Food, Drug, and Cosmetic Act. These statutes usually empower the governmental agency to enforce the law using certain defined aspects of the compendia. Most commonly recognized are USP and NF standards for determining the identity, strength, quality, and purity of the articles and specifications for packaging and labeling. The Pure Food and Drug Law enacted by Congress in 1906 is a landmark in U.S. history. At that time, the USP-NF standards were given legal status, and the federal government was empowered, now resident in the FDA, to enforce USP requirements.

Congress reaffirmed this authority in 1938, and even the sweeping 1962 amendments did not alter this essential fact. During the 1995–2000 cycle, two statutes, the Dietary Supplements for Health and Education Act of 1994 (DSHEA) and the Food and Drug Administration Modernization Act of 1997 (FDAMA), extended the use of the USP and NF by amending the Federal Food, Drug, and Cosmetic Act. DSHEA specifically provided that a dietary supplement represented as conforming to the specifications of an official compendium will be deemed misbranded if it fails to do so. FDAMA requires pharmacists compounding drug products to use bulk drug substances that comply with the standards contained in a USP or NF monograph and the General Chapter on Pharmacy Compounding. Thus, the USP and NF are not published in response to any statute, but statutory recognition places burdens on the USP and NF with regard to clarity and precision of presentation.

Separate from recognition by drug laws of the various nations and use by registration or related authorities, pharmacopoeial requirements are used by commercial codes in that a request for a standardized article should be satisfied only by the article that meets those standards.

^aThroughout this article, the abbreviation USP, when used alone, signifies the U.S. Pharmacopoeial Convention, Inc. The abbreviation USP followed by Roman numerals signifies a particular revision of the *Pharmacopoeia*. The abbreviation USP-NF is used to signify the *U.S. Pharmacopoeia-National Formulary* and are taken to be USP24 and NF19, unless otherwise specified.

Pharmacopeias historically came before government regulations because rules or specifications had to be established before they could be supervised and enforced by the government. In recent decades, however, governments will not permit marketing of drugs unless these are duly registered or scrutinized for quality. Approval of new drugs is most widely appreciated, but international commerce also places such demands on a continuing basis. Compendial standards are a regular feature of all these governmental processes, usually by way of reference in one document or another.

Mission

The mission of the USP is to promote the public health through establishing and disseminating legally recognized standards of quality and information for the use of medicines and related articles by healthcare professionals, patients, and consumers.

U.S. Pharmacopeial Convention

The U.S. Pharmacopeial Convention (USPC) consists of delegates of nearly 400 organizations, including colleges of pharmacy and medicine, professional associations, and some federal agencies. The Convention arose from a national convention called by Dr. Lyman Spalding in 1817 to develop national drug standards for the polyglot, rapidly expanding developing nation. It met in 1820 and published that year the first U.S. Pharmacopeia (Fig. 1). The present revision is USP24 and NF19 (1). Between 1888 and 1975, the NF had been published separately by the American Pharmaceutical Association. It was acquired by the USPC

in 1975, along with the assets of the Drug Standards Laboratory that was a joint body funded by USP, the American Pharmaceutical Association, and the American Medical Association.

Essential aspects of compendial standards are intrinsic in the history and composition of the Convention. USP standards are meant to describe an acceptable article from the point of view of the physician–pharmacist–patient interfaces; and they are inherently time-of-use (that is, shelf life) requirements. Another outcome of this focus is that practical, medically significant aspects are dominant in assigning requirements and the limits therein. Compendial standards are always established from the viewpoints of the medical and pharmaceutical professions, which in the United States are represented by the USPC.

General Committee of Revision

The Quinquennial USP Convention in 1995 elected a Committee of Revision of 138 outstanding scientists and practitioners in the various disciplines relating to quality standards and drug information. Approximately 1000 prospective experts made themselves available when both standards and information programs are considered. These experts are volunteers from industry, academia, and government. Of these, 100 were experts in public standards. The other 38 accepted assignments in the Drug Information Division. Each elected standards-setting expert was assigned to one or more of 20 subcommittees (Table 1). The 20 subcommittee chairpersons and the chair of the Committee of Revision make up the Division of Standards Development Executive Committee. The entire Committee adopts “Rules and Procedures” (2, 3) by which

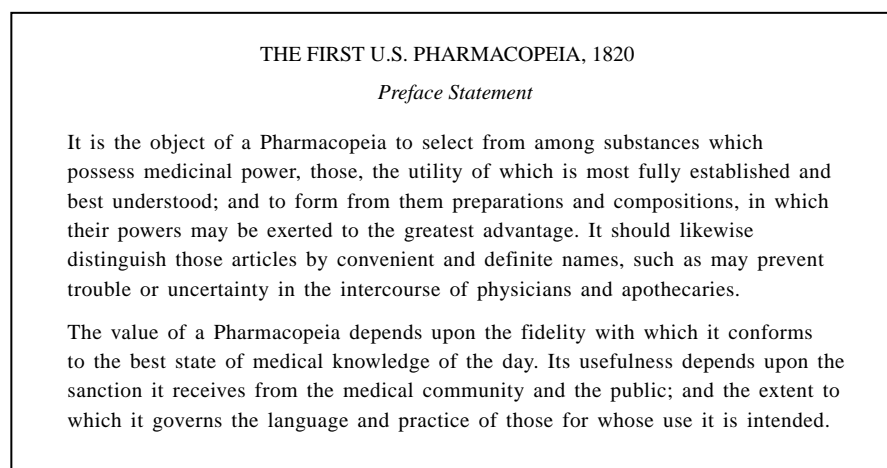


Fig. 1 *U.S. Pharmacopeia* Preface Statement, 1820.

Table 1 USP Subcommittees and chairs, 1995–2000

Subcommittee	Chairperson
Antibiotics	Henry S.I. Tan, Ph.D.
Biopolymers, Bioproducts, and Vaccines	Everett Flanigan, Ph.D.
Biotechnology and Gene Therapy	Robert L. Garnick, Ph.D.
Chemistry 1	Stanley L. Hem, Ph.D.
Chemistry 2	Dennis K.J. Gorecki, Ph.D.
Chemistry 3	Judy P. Boehlert, Ph.D.
Chemistry 4	Elliott T. Weisman
Chemistry 5	Edward G. Lovering, Ph.D.
Dissolution and Bioavailability	Thomas S. Foster, Pharm.D.
Excipients 1 (Monographs)	Zak T. Chowhan, Ph.D.
Excipients 2 (Methods)	Gregory E. Amidon, Ph.D.
General Chapters	Thomas P. Layloff, Ph.D.
Microbiology	Joseph E. Knapp, Ph.D.
Multisource Products Issues	James T. Stewart, Ph.D.
Natural Products	Paul Kucera, Ph.D.
Nonprescription Drugs and Nutritional Supplements	David B. Roll, Ph.D.
Packaging, Storage, and Distribution	Thomas Medwick, Ph.D.
Radiopharmaceuticals	Dennis P. Swanson, M.S., R.Ph.
Toxicity, Biocompatibility, and Cell Culture	Sharon J. Northup, Ph.D.
Water and Parenterals	James C. Boylan, Ph.D.

revision proceeds. A portion of the compendia is delegated to each subcommittee and, in agreement with its chairman, to each subcommittee member. Responsibilities are distributed for the more than 3700 monographs and the more than 160 general tests and information chapters. Every sentence in the USP and NF and every standard and its recognition are reexamined during each revision cycle. And every addition or change is given advance public notice and opportunity for comment.

The listing of subcommittee titles in Table 1 indicates the present scope of Committee interest and activity. The simple fact that a subcommittee has been created to deal with a subject area is a statement of its importance and implies that there are priorities in that area on the agenda of the Committee. As these interests change, the alterations are reflected in the USP periodical, *Pharmacopeial Forum*, discussed below. As outlined in the “Rules and Procedures” and noted above, any actions of subcommittees, or considered revisions to or adoptions of standards, are published for public evaluation and comment and are subject to mechanisms for appeal or request for postponement.

Council of Experts

The year 2000 USP Convention has adopted new structures and processes for setting public standards. The

Council of Experts replaces the historic General Committee of Revision. The resultant array of expert bodies and responsibilities is not available at this writing but may be found in *Pharmacopeial Forum* in 2000.

Headquarters

The USPC occupies several headquarters buildings complete with laboratories in Rockville, MD. It employs a staff of approximately 300 scientists, editors, and other employees to support the work of the Committee of Revision (Council of Experts). November 1998 saw the opening of the *Reference Standards Center*, a facility housing the *Reference Standards Laboratory* and the *Reference Standards Operations* unit. The USP Research and Development Laboratory remains in the 1989 headquarters building. As a publisher, the USPC performs its own phototypesetting, and its extensive in-house editing and production capability have been important to the timely publication of the standards. A modernization program is in progress to take advantage of the rapidly changing world of information-based products and technologies.

Financial Aspects

Independent standards setting is possible only because support for the program is undistorted by the impact of

standards on the governing bodies that ensure adequate funds. Funding is not dependent on those elements on which the standards impinge. Direct and narrow-based economic support from either government or industry would place standards setting in a less free or dampened environment. Opposed priorities or response to adverse impacts cannot be isolated from direct support. Financial independence is the component on which the USP program turns. Neither government nor industry gives direct supporting funds to the USPC for standards-setting activities or programs. Because these two are the major sources of funds for most nonprofit institutions, the corollary is that the USPC lacks immense reserves on which to draw and must be prudent in its expenditures. The USPC generates adequate income to meet its challenge as a public trust and to support growth and some ancillary programs noted here. Revenues for the USPC come from the following significant sources: sale of Reference Standards and publications. The committee members, advisory panelists, trustees, and convention delegates serve without monetary compensation, out of their sense of public service and professional responsibility.

PROGRAMS AND PUBLICATIONS

Pharmacopeial Forum and Continuous Revision

Pharmacopeial Forum, with the ability to publish official supplements twice yearly, has brought to fruition a dominant feature of the USP continuous revision. Only in this way can the USP keep pace with the progress of pharmaceutical technology and thereby escape the doom of continuing obsolescence. As the art of pharmacy, whether in design, manufacturing, or testing, became ever more the science of pharmacy, the compendia kept pace. An electronic, web-based product became available in 2000.

The bimonthly *Pharmacopeial Forum*, called “the journal of drug standards development and official compendia revision,” presents proposed new or revised USP and NF standards for public review and comment. *Pharmacopeial Forum* enables the reviewer to see at a glance both the text that is proposed for deletion and the text that is proposed for addition or modification. *Pharmacopeial Forum* is offered by subscription to all interested parties. Thus, interested scientists and practitioners from the general public other than the Committee of Revision and its advisory panels have access to the latest proposed revisions in the official standards and tests and can readily transmit their comments, suggestions, and data to USP headquarters for consideration by the Revision Committee.

In addition to presenting proposals recommended by the Committee of Revision in a section entitled “In-process Revision,” the *Pharmacopeial Forum* contains several other sections. The “Headquarters Column” gives information on publication deadlines, news, or statements of the Pharmacopeial Convention and the Committee of Revision, summaries of issues discussed by the Drug Standards Division Executive Committee, and various tabulations or lists that aid in keeping track of the multifaceted revision program. A section called “Stimuli to the Revision Process” publishes reports or statements of authoritative committees, scientific articles relevant to compendial issues, general commentaries by interested parties, and collations of comments received in response to policy initiatives. A recent addition is a section on International Harmonization.

Pharmacopeial Forum is intended to promote public comment at the earliest possible stage in standards development. Industry is the largest single participant in that activity. This periodical has been shown to stimulate comment and therefore can be credited with increasing the pace of revision. What comes to mind is an analogy of USP revision to a thixotropic gel—as the pressure is increased, so is the flow. This periodical operates in tandem with the program of Open Conferences to ensure vigorous participation in standards setting.

Related Publications

Efforts are under way to make the USP-NF more useful worldwide. Incorporation of monographs for multivitamin products, biotechnology-derived products, veterinary drugs, and botanicals and other dietary supplements help. The *USAN* and *USP Dictionary of Drug Names* has wide international applicability, thereby making it a repository of International Names & Nomenclature (INN) and British Adopted Names (BAN), as well as U.S. Adopted Names (USAN). *Pharmacopeial Forum* now contains announcements of proposals from the *Japanese Pharmacopoeia* and *European Pharmacopoeia* for revision of standards for international harmonization.

Open Conferences and Meetings

Consistent with the USP policy of emphasis on public participation in standards setting, open conferences and meetings are held to allow interactive examination of selected topics. Twenty-five of these conferences were held between 1980 and 1999, with an average attendance of approximately 150. Roundtable format is used to get maximum participation policy initiatives as well as proposed standards revisions on the table.

International Training and Outreach Programs

USP standards and information are recognized and used by many countries (19, 21), and the USP seeks to encourage, enhance, and facilitate their use. Examples of USP programs to train non-U.S. scientists and to seek their advice on how USP programs and products can be made more useful to other countries include sponsorship of visiting scientists and scholars, supporting doctoral and postdoctoral fellowships, cosponsorship of meetings and conferences, and formation of the International Health Advisory Panel.

Visiting Scientists and Scholars Program

Since 1990, the USP has substantially expanded its program for visiting scientists and scholars. Originally conceived as a program to train scientists from pharmacopoeias and official control laboratories, the program has been expanded to include scientists who spend up to 4 months in the USP Drug Research and Testing Laboratory or in the secretariat for the USP-NF within the standards development divisions to gain experience in laboratory techniques or pharmacopoeial revision procedures, as well as scholars from regulatory agencies and drug information centers around the world interested in the work of the information-development divisions in compiling and updating the USP DI database and USP information products.

Since 1990, the USP has hosted visiting scientists and scholars from China (13), Indonesia (2), Argentina (5), Nigeria (3), Poland (2), Romania (2), Turkey (2), Russia (2), and 1 each from Germany, Japan, Thailand, Korea, Kenya, Italy, and Kyrgyzstan.

Fellowship Program

The Fellowship Program is directed primarily at pre- and postdoctoral U.S. students. The USP has awarded 160 fellowships since 1981. The USP Fellowship Program also grants awards to non-U.S. students and postdoctoral fellows at U.S. universities and at universities outside the United States, provided the criteria for award can be satisfied. Currently the USP grants 10 fellowships annually: 6 in drug standards and 4 in drug information. Stipends are \$15,000 per year, and a fellow may compete for support for a second year. Eligibility criteria require, among other factors, that the fellowship application be signed by a member of the USP Committee of Revision, an advisory panel, or the board of trustees at the institution where the applicant is studying. Applications are reviewed for technical merit and relevance to USP standards and

information programs by committees of members of the Committee of Revision from the standards and information divisions.

Asian edition

In response to globalization and increasing use of the USP-NF abroad and to the foreign exchange picture, USP23-NF18 and USP24-NF19 both were published in Mumbai (Bombay) India simultaneous with the main publication in the United States.

Spanish language version of the USP-NF

The Spanish language edition of USP 23-NF 18 was published in 1995 as a DOS-based electronic product. In 1998, the product was converted to a Windows-based electronic product. Work on the translation of USP-NF text began in 1993 with the exploration of machine translation software suitable for use in the translation of the characteristic text of the USP-NF. The software selected was ENG-SPAN-AMSM from the Pan American Health Organization. This was fortified by a microdictionary and software macros specific to the compendia. Since the initial release in 1995, the USP has released Supplements for the Spanish language edition concurrent with the release of the English language Supplements. The database continues to be kept abreast of revisions to the USP and NF, but it was not marketed after 1998.

NATURE OF COMPENDIAL STANDARDS

Scope

The first USP begins, "It is the object of a Pharmacopeia to select from among substances H those, the utility of which is most fully established and best understood" (Fig. 1). Subsequent USP Conventions have broadened that to encompass all safe and effective medicines and some related articles. Recognition by the USP is not based on the identity of the manufacturer, the environment in which used, or the existence of borders or restrictions on commerce. Articles are adopted into the USP and receive standards scientifically appropriate to the articles, irrespective of whether the article is over-the-counter or prescription only and whether the article is used at home, in the practitioner's facility, in the hospital, or even in a licensed facility. The same modern standards are applied, and the public is offered the same parameters of strength, quality, and purity. Also, the same opportunities exist for analytical challenge or regulatory compliance.

USP and NF Standards Are Public Standards

This distinction was expressed well by the late C.A. Johnson of the *British Pharmacopoeia* (4):

The first point to be underlined is that the pharmacopoeia provides a collection of publicly available standards that are open for inspection and challenge by all—a fundamental right in any civilized society. The standards are established only after wide public consultation—a procedure that deserves to be maintained and supported. Because of the public availability of standards for medicines any authority or, indeed, any individual, may cause a random sample of a medicine on the market to be challenged.

As noted previously, compendial standards arose from the professions on behalf of the public. There are no attempts at regulation by the USP of the daily application of pharmaceutical technology. USP standards are not to be confused with such concepts as the “product description.” As addressed below, compendial standards are not manufacturing directions as such and do not constitute the manufacturers’ release criteria. They define the acceptable article as and when used.

Functions and Responsibilities of Compendial Standards (5)

What are the functions of USP standards? First, remember the nature of any standard. A standard is a rule, a principle of orderliness, that implies: 1) an element of agreement or acquiescence of most concerned, 2) some authority, and 3) some benefit to be gained by all concerned. To obtain this agreement in a reasonably democratic way, as opposed by way of authoritarian edict, there must be an element of compromise. However, one fact cannot be ignored: adoption of one rule implies rejection of alternative rules. In a real sense, this is loss of freedom of choice. The question comes—what do we get in trade? Why standardize? Why go through the occasionally wrenching experience of compromise? We get orderliness instead of chaos, and where there is orderliness, and only then, we get predictability. The essence of a USP standard is predictable drug product quality. The benefits of that predictability are settled on all concerned: the public, the professionals, the manufacturers of quality products. Without predictability in use, drugs would not have won the fullness of their position in society today. Physical and chemical tests and specifications are given for measurable quality parameters. Official drug names and definitions are established, which give a common vocabulary and a clear

understanding to communications within and among the professions.

Why do we persist in saying “public” standards? Because compendial standards, proposed and adopted, are published and circulated in public. The standards state to the public, manufacturers, and professions what constitutes predictable drug product quality from lot to lot and from manufacturer to manufacturer. The parameters of that predictability are published for all to see, for all to discuss, for all to grant their element of agreement, or for that matter, for all to contest publicly. Other agreements and other compromises, however reasonable and neatly filed away and for whatever reason not available, cannot fill that simple, rightful need of public information.

Codifying an element of predictability is not the only function of public standards. The USP-NF standards are enforceable by law, and are enforced strongly. This is the second function of USP standards: providing a basis for active protection of the public from unacceptable articles, not merely establishing ideals to be ignored when convenient. The USP and NF give detailed tests or assay procedures that allow determination of compliance with the standards; indeed, to call for a quality level without supplying a proof test would be unenforceable. Third, the quality manufacturer knows that other manufacturers of the same article must pass muster as well. Two less obvious purposes should be noted. The participants in the revision process also focus timely attention of industry and government on important problems or new technologies. Also, to a large measure the compendia define the state of the art of pharmaceutical analysis and therefore have impact as well on general drug registration and nonofficial articles both here and abroad.

Beyond defining acceptable articles, what are the responsibilities of USP and NF standards? The USP must strike a balance between the right of all concerned parties to act independently and efficiently and the need to preserve the rights of the patient to safety and quality. The USP and NF are responsible first to the public. These compendia can encourage the experts to make the quality judgments that the public cannot. The USP and NF are responsible to the practitioners of medicine and pharmacy who understand and are bound by the need for standards. The USP and NF are responsible to put forth standards that are meaningful to government and that can be enforced. The USP and NF are responsible to industry to put forth standards that are meaningful, practical, and affordable. The goal of the USP and NF is to set standards that are equal to the need, that can be enforced, that can be met by capable manufacturers, and that, at least for the moment, satisfy pharmaceutical scientists and practitioners.

Drug product quality in the United States is more reliable, more predictable than other aspects of the system of healthcare delivery as a whole. Adding tests or tightening specifications at the manufacturer level that does not help patients or professionals derive further benefits from drugs does little except increase patient or taxpayer drug bills. There is something to be said for knowing when to stop. We must be able to recognize diminishing return!

Drug product quality stands out in sharpest detail and is most highly visible to those of us in this technical community. Our concerns are those of the specialist for his or her field. In contrast, consider the heavy mist through which we gaze at problems of predictable therapy—not of differences in predictability resulting from variations in lots or brands, but of differences in pharmacogenetics; dietary habits such as preferences in beverages or entrees; metabolism as a function of disease, age, or sex; or, on a larger scale, failures in patient compliance or drug interactions.

General Approaches to Standards (5)

We can recognize three distinct but interdependent, effective standards: 1) USP-NF standards, 2) current good manufacturing practices, and 3) in-house quality assurance or process validation protocols. The three programs are different but have the same objective: drug product quality. The official standards are developed to be meaningful in this multiple context. In recent years, a similar array has occurred in most developed nations, and this is rightfully seen as a measure of success.

Currently, public standards tend to be performance rather than design standards. To illustrate: Are only automobile exhaust emissions measured as proof of success in reducing pollution or is an “official engine,” complete with engineering drawings, promulgated? In performance tests, desired attributes are tested for postmanufacture. This is end-product testing. These answer the fundamental question: How did it turn out? Older monographs, with prescription-like formulas but few tests, were essentially design standards. Only a few such appear in USP24. The resurgence of pharmacy dispensing now will result in more design-type standards. The trend toward performance tests has been the case in all standards except, notably, plumbing codes. One can imagine the difficulty of converting a visit of the plumbing inspector from an examination of allowable joints, angles, and materials to an experiment in which recovery of a test specimen of a hamburger slurry is a pass/fail criterion. Pharmaceutical technology now takes this one step further

and recasts the design standard as a “validated process” when the demonstration by end-product testing can lead to sufficient confidence in both the design and the process of carrying out the design.

Design standards intrinsically must be more detailed, more complex, more arbitrary, and more authoritarian than performance standards. Specific formulas can be specified by USP-NF, as can methods of compounding or processes of manufacture if necessary. However, end-product testing is believed by most pharmaceutical scientists to be adequate public standardization. That is what the patient does—end product testing!

USP-NF standards are established with the specific intent of serving as instruments of enforcement. Because these standards may be the common ground in a legal contest, the compendia publish both the experimental details and the limits within which the results must fall. Pass or fail—not how well, or by how much. Is it or is it not an acceptable article? Claims that a product “exceeds” USP specifications distort standards concepts. The compendial standard is that which any specimen of that article must meet at any time during its valid life. Although not a criterion aimed at use as a release standard, it is clearly the ultimate element in assigning expiration dates. It is the door at the end of a long hallway, not an immediate opening to the marketplace.

Limitations of Standards

Let us start with a truism. Quality is built into a product; it cannot be tested into it. It is also true that failure to build quality into a product can be tested for, as can variations from a good norm. The last century coped with a problem and a paradox: a redefinition of what was good and, most significantly, what was good enough (6). In this way, statistical quality control developed. We accepted the idea that even if absolute precision in manufacturing were possible (and it is not), it would be superfluous. When we strive for product uniformity, we still remember that quality will vary, so that what we really aim for is standardization of quality within limits. That is why “quality” requires control. The object of control is to enable us to do what we want to do, with economic limits (6).

We accept the inevitability of imperfection, thus we must accept the inevitable need for quality standards. Thus must we also accept the imperfection of those same quality standards. Compendial standards must be what can be done; failure to publish a standard because it is imperfect can be failure to serve the public. I submit that a willingness to move forward in the face of imperfections is

consistent with the second maxim stated by Descartes (7), which applies to any uncertain situation, to move resolutely and unswervingly in the most promising direction.

TECHNICAL FEATURES OF COMPENDIAL STANDARDS

Pharmacopoeias always have been as much statements of current technology as of current therapy. Examination of an older pharmacopoeia from any time or place gives a fair measure of the contemporary drugs available, therapeutic strategies, and medical knowledge. No less does such examination give a fair measure of the contemporary pharmaceutical technology. Pharmacopoeial standards arise from the contemporary, controlling concepts of what constitutes good quality and, further, how one achieves or demonstrates that quality. However, selection requirements can come only from processes and equipment in general use. Available technology determines what is to be required or standardized, just as it determines what can be made.

A rough correlation existed between pharmacopoeias and pharmaceutical technology. At least that was so until the modern drug era, when new drug discoveries proceeded at rates that, in one generation, rendered many pharmacopoeias obsolete, especially those that could not revise as frequently as is done by the USP. Increased international trade and acceptance undercut the need for large numbers of national pharmacopoeias. Technology forged ahead on many fronts in manufacturing processes and equipment, dosage form design, new kinds of excipients, and new analytical capabilities. Only a few pharmacopoeias continue to publish specifications of practical import consistent with the state of pharmaceutical technology, and these are primarily regional or international in scope. Different adaptations to rapidly changing therapy and technology account for some of the unevenness in and among pharmacopoeias and a consequent need for harmonization. We report here on current compendial specifications, what they are, by whom and where they are established, and how and why they are created.

Official Names

Pharmacopoeias are an accumulation of responses to problems in medicine, pharmacy, or pharmaceutical technology as these presented themselves over the years, even over centuries. From the beginning of pharmacopoeias, a fundamental function has been the establishment

of names “such as may prevent trouble or uncertainty in the intercourse of physicians and apothecaries” (Fig. 1). Different things should have different names; that is a fundamental principle of compendial standards. This applies to dosage forms as well as to molecules. This deeply rooted function seldom gets much attention until someone wants to change something. A recent case is the desire to change the capability of consumers to identify the ingredients of products through product labeling of all inactive ingredients. The need for concise, unambiguous names for these ingredients as well is inescapable of notice.

A cooperative effort since 1961 between the American Medical Association and the USPC was augmented in 1964 by the addition of the American Pharmaceutical Association, then the publisher of the NF, to form what has been known since as the U.S. Adopted Names (USAN) Council. The U.S. Food and Drug Administration (FDA) was invited to join the Council in 1967, and the FDA refers to USAN names in its regulations. The Council consists of persons conversant with the needs and problems of naming drugs and an appeals body. The names determined by the Council are incorporated, along with other names for drugs (including public, proprietary, chemical, and code-designated names), in an annual book, *USAN and the USP Dictionary of Drug Names*, published by the USPC.

Method and Specification Selection

The selection of USP methods and specifications can be concluded to be practical consequences of what has been successful in detecting drug product quality variability and what can ensure demonstration that variations in quality remain within acceptable limits. USP methods are selected for their fitness for use for that monograph application, just as USP Reference Standards are adopted on the basis of suitability for intended use in that monograph.

Tests and specifications, and reagents as such, were introduced into the USP in the 1890s. The first truly instrumental determination used was the Wild polaris-trobometer for natural oils in 1890. Trends were established in USP XV (1955) toward decreased monograph prescriptions, or design standards, and, separately, toward the increasing range of General Tests. The USP now has more than 160 general chapters. There are few analytical methods used in modern pharmaceutical analysis that are not used by the compendia.

An interesting historical change in the orientation of the USP is implicit in this discussion. Until the age of instrumentation, there was an expectation that compendial

methods could be performed in a community pharmacy. What became possible decided where it could be done, and the focus moved out of the pharmacy and into the central analytical laboratory.

Tremendous changes in the selection of compendial methods have occurred in recent decades. Classic physical and chemical measurements had long been established as objective criteria for quality assurance and for compliance testing, but these had significant limitations for dosage forms, both as assays and as limit tests. As recent decades have witnessed a torrent of instrumental developments, it should be no surprise that the compendia have experienced a resultant flood of applications that use those instrumental methods popular in the industry. The previous limiting factors, the availability of instruments their and reliability (or certainty of repair) ceased to be a problem even before the revolution in microelectronics.

The USP placed emphasis in the last 20 years on stability-consciousness and impurity-consciousness in preparing monographs. See 1086, *Impurities in Official Articles* (1). These major trends explain the prominence of chromatography. It is obvious that the major portion of recent revision revolves around chromatography: high-pressure liquid chromatography (HPLC), gas-liquid chromatography (GLC), and thin-layer chromatography (TLC) for identity tests and reference standard evaluation; HPLC, GLC, and TLC for purity or limit tests; and HPLC and GLC for assays. Much effort is spent in evaluating chromatographic systems and identifying System Suitability Tests (see discussion below) to make these methods more reliable. The amazing speed with which HPLC was adopted as mature scientific measurement is nowhere more evident than in drug analysis and, therefore, in the USP. The USP 24/NF 19 contains 1800 HPLC/GLC and 740 TLC initial references. Because of cross-referencing of substance methods in dosage form monographs, the total number of chromatography-based requirements is much, much higher.

To a large extent, modern separation science has solved a previous significant problem in compendial method selection, that of interfering substances. The USP does not lock all manufacturers into the same formulation and manufacturing procedure, and the USP allows considerable freedom as long as the substances added in dosage forms do not interfere, which these must not do (8), with the official tests and assays. Because the manufacturer can change the formula or introduce new products, it is unreasonable to assume that the USP would have tested every conceived and about-to-be conceived product. Thus, in exchange for the freedom of formulating, the individual manufacturers have the responsibility of keeping up with

proposals in the *Pharmacopeial Forum* and assessing the applicability of proposed methods to their specific existing products and their contemplated new products, i.e., validation for their formulations. Separation science now available makes it likely that a method workable for all can be identified.

Automated Methods

In considering the introduction of new test methods, it has been necessary for the USP to recognize the appropriateness of automating compendial assays and tests. The Committee of Revision prefers to adopt automatable procedures, especially for multiple-unit specimens such as those arising from dissolution or content uniformity requirements. Thus, since the publication of the USP XVIII in 1970, the General Notices to the USP has included automated procedures in discussions on the use of suitable alternative methods. The General Notices to the USP24 state (9):

Automated procedures employing the same basic chemistry as those assay and test procedures given in the monograph are recognized as being equivalent in their suitability for determining compliance. Compliance may be shown also by the use of alternative methods, chosen for advantages in accuracy, sensitivity, precision, selectivity, or adaptability to automation or computerized data reduction or in other special circumstances. Since Pharmacopeial standards and procedures are interrelated, only the result obtained by the procedure given in this Pharmacopeia is conclusive where a difference appears or in the event of dispute.

Thus, the USP states explicitly what is to be done and in what sequence but not by whom or by what.

Stability

Monographs do not include a specific section that deals directly with this major aspect of drug product quality, primarily because stability is related to a number of factors including chemical structure, formulation ingredients and processing, bioburden, packaging materials and storage, repackaging, and geography. Concern for stability is manifest in the Pharmacopeia in at least seven ways: 1) the general requirement for expiration dates, 2) monograph packaging and storage requirements, 3) standards for packaging materials and containers, 4) standards for repackaging and storage, 5) limit tests for decomposition products, 6) stability-indicating tests or assays, and 7) the informational General Chapter, "Stability Considerations

in Dispensing Practice.” Underlying these standards are standard definitions of storage conditions. An emerging establishment of requirements for Labile preparations will add an eighth stability-resultant category. New informational chapters are likely to expand this list. Stability is not causal—it is the effect of choices during formulation development, packaging, and storage instructions.

Test Results, Statistics, and Standards (10)

Confusion of compendial standards with release tests and with statistical sampling plans occasionally occurs. Interpretation of results from official tests and assays requires an understanding of the nature and style of compendial standards. Tests and assays given in the USP prescribe operation on a single specimen; that is, the singlet determination. This is the minimum sample on which the attributes of a compendial article should be measured. Some tests, such as those for dissolution and uniformity of dosage units, require multiple-dosage units in conjunction with a decision scheme. It is to be understood that these tests, albeit using a number of dosage units, are in fact the singlet determinations of those particular attributes of the specimen. These quantities should not be confused with statistical sampling plans. The compendial procedures demonstrate compliance of the attributes of an article with compendial standards only for that specimen (of one or more dosage units) that is subjected to analysis. Repeats, replications, and extrapolation of results to larger populations are neither specified nor proscribed by the compendia; such decisions are dependent on the objectives of the testing.

Commercial or regulatory compliance testing, or manufacturer’s release testing, may or may not require examination of additional specimens, in accordance with predetermined guidelines or sampling strategies. Moreover, a statistical plan may be necessary to relate the shelf life of the product, test results, and release specifications to the ultimate compendial specifications.

System Suitability Tests

System Suitability Tests are now characteristic of USP methods (10, 11). They are based on the concept proposed by this author in early 1971 that the instrument, reagents, packings, conditions, procedural details, detectors, electronic accessories, and even the analyst constitute a single system that is therefore amenable to an overall test of system function. Reliable chromatographic performance, for example, may require specifications for resolution,

column efficiency, peak tailing, precision of replications, or extremes of conditions. Such tests obviate the necessity of specifying a multitude of instrumental settings, model numbers, names of manufacturers, packings and lot numbers, and other physical-chemical and engineering characteristics. Also avoided is the distribution of official lots of chromatographic packings through the compendial headquarters. System suitability tests were first made official in the USP XIX and have come into general application.

New Technology and USP Contents

During the revision cycle, the USP introduced a Web site that contains up-to-date information about the USP and its program as well as ordering instructions. The USP home page is located at www.usp.org.

Law and regulations ensure the adequacy of the premises, practices, and documentation of the manufacture of drug substances and preparations expected to result in articles that comply with pharmacopoeial standards. It was the practice of previous USP revisions to publish and update the FDA Good Manufacturing Practices (GMPs) in a general information chapter *Good Manufacturing Practices* (1077). This was perceived to be a service to those in manufacturing establishments and pharmacists. Rapid expansion of the Internet and almost universal establishment of Web sites by all institutions abrogates, in general, the perceived need that necessitates this service. Therefore, this revision cycle culminated in the deletion of federal and other texts now reliably available at no cost through the new technology. Federal documents can be found on the U.S. Food and Drug Administration Web site at www.fda.gov. Along with *Good Manufacturing Practices* (1077), two other general information chapters based on federal regulations are deleted: *Federal Food Drug and Cosmetic Act Requirements Relating to Drugs for Human Use* (1076) and the frequently revised *Controlled Substance Act* (1071).

Drug Release

Nowhere have the value and failures (and political pressure) of pharmaceutical technology been more apparent than in the complex of processes that amount to release of active ingredient from its pharmaceutical presentation. A major concern, now a generation old, was reliable release of drug from solid oral dosage forms (11). The USP and NF each introduced dissolution tests in six monographs in 1970. For some drugs, absorption is dissolution-rate controlled, reduction in bioavailability

results. This is precisely a reduction in strength and has no additional resulting in medical significance. The problem arises when two formulations of such a drug are compared where dissolution rates are substantially different—this is known as bioinequivalence. This led to much commercial mischief and distortion of scientific facts to achieve protection of existing products from competition. Some scientists emphasized drug chemical properties as the central factor—particle size or solubility, for example. Others pointed to the inactive ingredients used in the processing of compressed tablets—lubricants and disintegrants, for example. Or there was emphasis on the process itself, as in overcompression. Pragmatists noted the continued use of hydrophobic tablet coatings—shellac, for example.

A drug release test, dissolution, was a satisfactory standard every time there was a problem of any practical consequence. Of equal significance was the recognition of the strength of dissolution testing as a tool for quality control. Thus, equivalence in dissolution behavior was sought in light of both bioavailability and quality control considerations. The USP 24 has more than 600 monographs for tablets and capsules: 500 have an official dissolution requirement; 100 others have a different performance test considered appropriate for acceptable drug product quality.

For USP tablet and capsule monographs, we know of none where two articles, fully in conformance, have clinically significant inequivalence. Experience has demonstrated that when a medically significant difference in bioavailability has been found among supposedly identical articles, a dissolution test has been efficacious in discriminating among these articles. Because the USP sets forth attributes of an acceptable article, such a discriminating test is satisfactory because the dissolution standard can exclude definitively any unacceptable article. Therefore, no compendial requirements for animal or human tests of bioavailability were necessary. The practical problem has been the obverse; that is, dissolution tests are so discriminating of formulation factors that may only sometimes affect bioavailability not uncommon for a clinically acceptable article to perform poorly in a typical dissolution test. In such cases, the Committee of Revision has been mindful to include as many acceptable articles as possible but, at the same time, to exclude dissolution specifications so generous as to raise reasonable scientific concern for bioinequivalence.

There is no known medically significant bioinequivalence problem with articles of which 75% is dissolved in water at 37°C for 45 minutes with the use of either official apparatus at usual speed. A majority of monographs have

that as the requirement, and this is called “First Case” in the *Pharmacopoeial Forum*. Other articles for which there are no known or likely medically significant bioavailability problems have required some adjustment of medium or apparatus.

Details of tests and specifications can be found in the USP. For ordinary tablets and capsules, the apparatus used (basket and paddle) affords low levels of agitation appropriate for challenging drug products with a significant potential for diminished physiologic availability. Similarly, only the most discriminating solvents are selected. See 1088, *In Vitro and In Vivo Evaluation of Dosage Forms*. Extended-release formulations present different standards problems. Pharmacologically sound judgment that a drug is acceptable as an extended-release product obviates the necessarily conservative choice of apparatus and conditions applied to ordinary articles. Either the drug does not have a significant absorption problem or its metabolites are active, in which case a higher agitation apparatus may be necessary. A new apparatus is presently under consideration to handle those cases in which the current apparatus is unworkable.

Dietary Supplements

Dietary supplements are now covered by federal legislation (see DSHEA under Legal Status of the Official Compendia, above). They are a rapidly growing aspect of consumer choice. The USP and NF made great progress in this field in the years after the 1995 Convention. The Convention debated vigorously on the proper placement in the compendia, if any, necessitated by the resurgence in the United States of the use of botanicals, amino acids, and other substances. These had fallen into disuse earlier in the century (see 1995–2000 *Resolutions* in the *Proceedings*). The feasibility of standards was established quickly, and there developed a broad support for USP-NF monographs. The area of information for consumers and practitioners has been the more contentious. The first task was to identify the most widely used botanicals having no counteracting safety concerns.

Dietary supplements are on a different legal footing from drugs, being regulated as foods or under food statutes. In contrast to drugs, supplements are not required to comply with compendial requirements unless labeled to be the official article, in which case it is necessary to use compendial methods. Labels that specify USP or NF in supplements are gaining momentum. Although there is extensive analytical experience with vitamins and minerals, problems in the analysis of botanicals are likely to arise because a tradition of testing is not widespread.

Very complex mixtures are the rule, and there is uncertainty as to the critical components.

In response to a resolution adopted at the USPC's 1995 Quinquennial Meeting, the Committee of Revision addressed issues concerning natural products of plant origin used as dietary supplements. The Subcommittee on Natural Products, assisted by the Advisory Panel on Analytical Methods for Identification and Characterization of Natural Products, set priorities for 21 botanicals for standards development. The criteria underlying the selection included absence of safety risk; extent of use as reported by trade sources; positive assessment by recognized pharmacognosists, usually on a presumption of beneficial pharmacological action and history of use in traditional medicine; and the ability of the article to be able to meet typical USP-NF monograph requirements.

A new admission policy addresses botanicals used as dietary supplements recommended for their official adoption by the USP Committee of Revision. When the FDA has approved a use or a USP DID advisory panel has accepted a use, it is to the USP. If neither condition is met, and there is no safety concern, the article is admitted into the NF. If there is a safety concern for an article, then no NF monograph is published, and USP DID will publish a monograph that is negative on the use of this article. The first official monograph resulting from the 1995 USPC appeared in the Seventh Supplement to the USP in 1997. That was Ginger, USP. A total of 18 monographs were published through the First Supplement to the USP24-NF19.

Compounding Pharmacy Monographs

A corresponding resurgence in compounding practice has arisen from the unavailability of strengths or forms suitable for special populations, especially pediatric patients, and for short-life preparations such as a buffered, diluted solution of sodium hydrochlorite. The Advisory Panel on Pharmacy Compounding was installed. Pharmacists from this Panel advised the Subcommittee on Packaging, Storage, and Distribution where most scientific questions on prescription stability, packaging, and storage could best be resolved.

The USP laboratory evaluated all the monographs, and its findings were published in *PF*. Ten official monographs were published through the USP24-NF19 Second Supplement. The International Association of Compounding Pharmacists is playing a key role in selecting among the many known formulas to identify those of more medical merit and wider usage. Additional monographs are expected to appear in subsequent supplements. If

a commercial article is available, no compounding USP monograph is needed. As noted above (see Legal Status of the Official Compendia), recent legislation supports and guides pharmacy compounding.

Excipients

Standards for more than 350 excipients are published by the USPC, primarily in the NF (approximate 270 versus 80 in the USP). Before 1975, the USP and NF were published by different organizations; each had monographs for drug substances, dosage forms, and excipients. The USP adopted the medically best, and the NF adopted those widely used. With the 1975 merger, it was decided to publish the so-called inactive ingredients in the NF and, in addition, that both the USP and NF publications would be bound together in a single volume and have joint supplements.

The advance of pharmaceutical technology ever forces forward new or refined excipients, some with heretofore unexploited properties. Polymers are a case in point and have been central to many of the technological advances of recent years. New and different challenges for compendial standards are offered by materials used in new wave formulations. Modern analytical chemistry allows rather thorough evaluation of materials.

Excipients were a major, and the initial, component of programs for international harmonization. The USP focuses on functionality tests, such as for that compressibility. These are available for many excipients. Why incorporate functionality requirements into the NF that assess parameters that are critical to a minority of purchasers, thus restricting the channels of commerce and probably raising everyone's costs? If viscosity range or degree of crystallinity is critical to one formulation, isn't that the problem for that formulation? One solution would be to adopt a standard test but to require that the product labeling state the specification to be applied to the contents. The product label would state the definitive parameter(s). The NF is satisfied at present with an array of standards of identity, purity, assay, moisture, packaging, storage and labeling, and so forth that supports the name and labeling of materials.

Reference Standards

In support of its program for public standards for drugs, the USPC supplies the Reference Standards required by the monograph standards, tests, and assays of both the USP and the NF. Biologic assays in the USP were the first to require use of reference materials. Authoritative sources of

these were needed, and the first were adopted in 1926. As an index of the growth of analytical science, 45 were needed in 1950 for the USP and NF and 438 in 1980. As of 2000, there are 1400.

One cannot fail to note the vast expansion of the collection in the last few decades. Surely this was not fueled by additional biologic assays. Underlying the initial growth phase was the widespread utilization of spectrophotometry for identification and assay. Separation science was the second phase in pharmaceutical industry control laboratories. As a corollary, USP and NF method selection moved in the same direction. Spectrophotometric identity tests and assays are more reliable, especially for compliance testing, when performed in the relative mode, which uses a reference standard, rather than the absolute mode, which is the norm in titrimetry. There is some residual difference of opinion in other countries on this point, but that is rendered moot by the widespread adoption of separation science by the pharmaceutical industry and, thus, by the compendia. It is a characteristic of chromatographic methods that a reference standard be required, sometimes more than one for a procedure. The accumulation of modern tests and assays results in 5 to 10 uses for many reference standards.

Two critical features of this program should be remembered: all USP Reference Standards are subjected to collaborative testing to generate a purity profile and then must be approved by the USP Reference Standards Committee, which is composed of members from the Division of Standards Development Executive Committee. Three or more laboratories test each one. For this purpose, reference standard purity is defined as known composition with respect to intended use. Adoption of a chemical batch demands a known purity profile, which is obtained by meaningful experiments. Although the USP or NF monograph requirements are met, testing usually ranges beyond the monograph tests to construct a purity profile, particularly in the accumulation of more extensive chromatographic characterization. In contrast to the highly purified synthetic organic medicinals, heterogeneous substances of natural origin continue to be used in addition to the highly purified products of chemical syntheses. Biotechnology-based products in particular are chemically complex. These usually have counterparts in international standards, but the inherent heterogeneity makes direct interchange unworkable. In these cases, careful characterization of the compositional and biologic profile allows for lot-to-lot consistency.

Existence of a reference standard used in a compendial procedure can in effect transfer the quality standard from

print to the contents of the reference standard vial. Specifically, the content of the drug product is assayed relative to the USP Reference Standard, which is taken to be 100%. The late C.A. Johnson was fond of stating that the printed standard turning yellow was not the same as a yellowing reference standard.

The Committee of Revision takes steps relative to the continuing acceptability of any specific lot of a USP Reference Standard. Publication of a list of current items in each *Pharmacopoeial Forum* keeps the collection up to date in this regard, as well as indicates which are new, have been deleted, or are unavailable.

PARTICIPATION

Food and Drug Administration

Close working relations with the FDA continued as a hallmark of the revision program, both with individual scientists and with FDA laboratories and centers. Formal liaison efforts were conducted primarily through the Compendial Operations Staff in the Center for Drug Evaluation and Research of the FDA. Staff are responsible for obtaining and coordinating agency comments on proposals appearing in the *PF* and for serving as an official point of contact with the Center. They have been effective in contributing many suggestions for improvement as well as a greater degree of consistency between FDA and compendial requirements. Each subcommittee has an ad hoc reviewer assigned by the FDA to ensure communications.

The National Center for Drug Analysis in St. Louis, MO, was a constant, valued, and cooperative participant in the revision process for nearly 3 decades. This laboratory continued from past cycles to do extensive development and review of tests and assays. During this cycle, it continued as the primary FDA participant in the ongoing evaluation of established and proposed new USP Reference Standards. Moreover, careful review was given to many General Chapters and issues in harmonization. The FDA is the single most productive outside source of scientific data and information.

Industrial Cooperation

The *U.S. Pharmacopeia* and scientists in industry interact in a number of important ways (13). The pharmaceutical industry is highly quality conscious; thus the program of a standards-setting body is of pervasive interest. Many interfaces with industrial scientists can be discerned: they

use USP standards; serve as members on the General Committee of Revision; serve as members on the Board of Trustees and on appeals bodies and as delegates to the USP Convention; they propose revision and review revision proposals; they serve on USP Advisory Panels, cooperating organization committees, and panels; they help develop USP Reference Standards; and they purchase USP publications and reference standards. Of these, the most important is that they observe the standards, consistent with the original intent (Fig. 1) of the *U.S. Pharmacopeia*.

Participation by individual

Nearly equal numbers of industrial and academic scientists are involved with USP and NF standards determinations, and 10 of 20 DSD Executive Committee members are industrial. An average of half of the DSD Executive Committee consisted of industrial scientists during the last three cycles. The DSD Executive Committee is the policy-setting body and addresses all broad-interest topics. Pharmaceutical technology is the predominant category of information discussed. Other participants include government scientists and medical and pharmacy practitioners.

Participation by manufacturers

Industrial participation on advisory panels, or the General Committee of Revision itself, in addition to industrial interaction on scientific proposals to and from the USP, represents a substantial time demand for any single company. The enormous amount of laboratory work that supports the pharmaceutical and analytical technology incorporated in USP and NF monographs should be emphasized. In terms of economic value, all other participation does not match the cost of laboratory experiments performed by the industry. Indeed, it is what can be done at the bench and on the manufacturing floor that determines what can or cannot appear in the *U.S. Pharmacopeia*.

It should not be concluded that the USP relies entirely on comments from a single manufacturer. There are other interested parties. Testing may be performed as deemed necessary to supplement or corroborate results reported by or disputed among manufacturers. USP laboratories develop data to enable the Committee of Revision to decide on standards establishment and revision. Note that commercial products are not tested by the USP to determine compliance with official standards. Rather, USP testing is called on as necessary to 1) test methods that are considered for new or revised standards in monographs, 2) test products to generate data for the USP Committee of Revision when information needed for

standards development and improvement is not forthcoming from other sources, 3) participate in collaborative studies, 4) test materials considered for use as USP Reference Standards, and 5) test USP Reference Standards periodically for degradation.

Industrial Organizations

Revision of the *U.S. Pharmacopeia*, particularly in development of new General Chapters and tests and assays for specialized products, continues to progress through the strong and proactive participation by industry.

No recent pharmacopoeia was issued without notable contributions by the Pharmaceutical Research Manufacturers of America through its standing technical committees in quality control, statistics, and biologicals. Careful reference to the *PF* also will reveal many contributions by scientists from individual member companies.

The Council for Responsible Nutrition (CRN) and the Consumer Health Products Association (CHPA) contribute to new standards for over-the-counter articles. Growth continued on the multiple vitamin-mineral combinations described by class monographs for cough and cold remedies, which include a very large number of products.

Parenteral Drug Association (PDA) members and committees focus on standards for sterile products, testing, and environments.

Self-Regulation

A continuing element of the USP seldom remarked on (13) is the fact that our program of public standards represents the substantial realization of self-regulation by the pharmaceutical industry. Self-regulation is an ideal. It is a natural ideal for free peoples and a historic impulse in our national life. However, it is not directly achievable as a uniform or practical principle. Pressures of money and competition and differing relative values get in the way. What is an obvious improvement to one is an inconvenience or an unreasonable expense to another. Unpopular causes are not written into most corporate job descriptions, especially when the employer suffers the costs of that employee's self-regulatory campaign and, more starkly, sees competition that gets a free ride. In this regard, one should recall the extensive industrial cooperation described above. This fact is remarkable, and it is so important that all understand this characteristic of the *U.S. Pharmacopeia* repeated here.

What is necessary to achieve the ideal of self-regulation is an instrument to transform individual efforts.

Table 2 1990 to 2000 Committees of Revision

870	New monographs adopted
23	New general chapters adopted
270	Deleted monographs
9240	Individual revision actions proposed in <i>Pharmacopoeial Forum</i>

That instrument is a neutral but authoritative institution. It can insulate individuals and diffuse responsibility while maintaining necessary accountability. The truly knowledgeable can act on their natural inclinations to self-regulate but remain insulated from all negative counterforces. That is one primary reason the USP is so productive, even though it is a small organization.

The ideal of self-regulation realized through institutional transformation has another consequence. It explains why USP standards are not limited in practice to the “lowest common denominator” but are the “acceptable common denominator.” Although our subcommittees strive for consensus when a number of parties are involved, or acquiescence of one party, USP standards are held to the more idealistic level of knowledgeable self-regulation. Consensus is not prerequisite.

Often quoted is de Tocqueville’s observation that Americans like to form associations to accomplish mutual purposes without the burden of additional government. This is fueled by our spirit of voluntarism. The USP effort, starting in 1820, certainly fits that description (14). However, why can’t trade associations be the vehicles of effective self-regulation? All who have worked with these know it cannot be done readily when any member’s vested interest is seriously threatened. Consensus in the strictest sense applies. The same type of experts, if elected to the USP Committee of Revision, can take that last essential step in self-regulation: they can set meaningful end-product standards free of strict consensus. It is the majority vote that decides. It is the unique nature of the USPC that effects that transformation. It does what deserves to be done, but without all the baggage that comes with more government.

To appreciate the impact of this element, consider the sheer magnitude of the USP revision effort. Table 2 gives summary facts about the magnitude of recent USP activity. Each of these pharmacopoeial revisions involved industrial participation. Ask yourself: How could so many things get done unless participants of all types were willing or at least favorably disposed? They are so disposed because in this quality-conscious industry, people want standards and because most would be willing to regulate themselves.

INTERNATIONAL RELATIONSHIPS

International Harmonization

The phenomenon of globalization of pharmaceutical manufacturing and distribution, already evident in the excipient monographs of the NF 18, continued throughout the revision cycle leading to the USP 24-NF 19.

Although originally founded as an organization to standardize medicines in the United States, the USP and its products and services are now known and used throughout the world. In today’s transnational and multinational economy for pharmaceuticals, the USP needs a strong international presence and influence for its survival as well as for continued growth and recognized leadership. Economic forces are driving major trading parties to affiliate to reduce trade barriers. Integral to this process is harmonization of requirements, regulations, and standards governing the approval and marketing of drugs, devices, etc., by governments. (See also *Harmonization of Pharmacopoeial Standards* in the second edition of this Encyclopedia.) In view of the contemporaneous publication of that chapter, also prepared here, only formal publication agreements are discussed herein.

Two organizations account for harmonization of new and revised standards: the Pharmacopoeial Discussion Group (PDG) and the International Conference on Harmonisation (ICH). Primary impact is attributed in this cycle again to the PDG consisting of the USP with the Japanese and European pharmacopoeias. Joint open conferences held on an international basis were crucial to any progress here because these ensured direct contact among experts from the three pharmacopoeias (see “Open Conferences and Meetings,” earlier).

Harmonization is not a one-time event, and it does not imply superimposability of texts. Residual disharmony can exist at the end of any specific attempt to harmonize standards. Iterations of harmonization, already evident, will be the norm in the future. Refer to the *PF* as the essential source of any future understanding or tracking of harmonization topics or sequences.

One serious discrepancy exists between requirements in this Pharmacopoeia and ICH guidelines beyond the relative mandatory status. USP-NF requirements are enforceable by U.S. federal and state laws and regulations. The standards give the attributes of acceptable articles already used in healthcare. A key contrast can be seen in the strict focus of the ICH, which is explicit in treating new drugs, that is, those not yet in use. The situations differ dramatically in practical consequences and breadth of application. Therefore, the goal of harmonization in some cases is not currently

achievable because justification does not exist to disturb many established products in favor of the most recent new drug approval practices.

Foreign Use of the USP and NF

Evidence of the utility of the USP beyond North America is found in the many orders received for USPC publications (USP, Supplements, *Pharmacopoeial Forum*) and reference standards. In return, the contributions of foreign scientists in questioning existing standards and in commenting on *Pharmacopoeial Forum* proposals frequently reveal insights not otherwise focused in domestic queries and comments. Although many of these insights that originate in foreign countries are passed to the USP through domestic affiliates, committee members usually are aware of the originator's contribution. Foreign scientific input is extensive and receives the same consideration on merit as does domestic comment.

Limitations on foreign use of USP standards arise from that same high technological level of the standards. Instruments cost, availability, and maintenance prospects set high barriers where foreign exchange is a problem for a country. Arranging for training of pharmaceutical scientists, and then keeping them, is also a limiting factor.

The Drug Information Division selects the person from a large pool of candidates for nomination in the particular class of service. For example, the nominating committees for the 1990–1995 Committee of Revision screened a pool of over 1200 candidates before they agreed on 228 nominees for 114 positions, representing 45 classes of service.

The advisory panels have an even more active program to identify highly qualified persons from other countries. As the USP DI database becomes more widely known and used outside the United States, the USP seeks qualified persons from countries using USP drug information to provide advice on the utility of the information in those countries and advice on how the drugs in the database are used in those countries. In this way, the database can be enriched and made more useful to a greater number of countries.

OTHER PHARMACOPEIAS

International Pharmacopoeia

The USP has established formal arrangements with individual countries or pharmacopoeias that are not direct participants in harmonization, although the practical

effects may be the same. As noted above, harmonization per se is the subject of a separate chapter in the second edition of this Encyclopedia.

Canada

Canada has a special relationship with the USP. The USP-NF has been officially recognized by statute as a legal compendium in Canada since 1954. Approximately 95% of drugs claiming compliance with compendial standards in Canada follow the USP-NF. The USP is proud of its status in Canada and values the contributions of its Canadian members. Representatives of the Therapeutic Products Programme (TPP) of Health Canada and the Canadian Pharmaceutical Association are members-at-large of the U.S. Pharmacopoeial Convention. Relations between the Health Protection Branch and the USP are cordial; the USP seeks to work closely with the TPP, as it does with the FDA. Six Canadians were elected to the USP Committee of Revision for the 1995–2000 period, and 45 Canadians have been appointed to USP advisory panels. Canadian pharmaceutical scientists, physicians, pharmacists, and other healthcare professionals take an active interest in the USP. In addition to the USP-NF being official in Canada, the USP drug information database *USP DI* covers all drugs on the Canadian market as well as on the U.S. market.

Argentina, Mexico, and Brazil

The USP initiated discussions with compilers of the three other active pharmacopoeias in the Western Hemisphere. USP scientists, practitioners, and staff have been intimately involved in meetings and programs, both in standards and in information. These successful efforts put down the groundwork for the first agreement with Argentina and subsequently with Brazil and Mexico between the USP and other Western Hemisphere pharmacopoeias that allows use of any part of the USP or NF, in English or Spanish, as an aid to elaboration of a national pharmacopoeia. Establishment of the same standards obviates the less desirable effort of harmonization. Adoption of the NAFTA pact within Canada, Mexico, and the United States stimulated progress on this bilateral basis.

The *Farmacopea de los Estados Unidos Mexicanos* (FEUM) is the Mexican Pharmacopoeia. Stimulated by the signing of the North American Free Trade Agreement (NAFTA), the USP and the FEUM initiated a dialogue in January 1993 to explore opportunities for harmonization. Originally conceived as a semiannual meeting after

the Pharmacopoeial Discussion Group (PDG) model, the USP has hosted five visiting scientists from Latin America.

China

The *Pharmacopoeia of China* contains volumes dealing with both Western and traditional Chinese medicines. A series of informal visits has taken place at the initiation of the Chinese Pharmacopoeia Commission (ChPC). The USP has been host to visiting scientists from China and has had seven scientists from the Chinese Pharmacopoeia in residence for periods up to 4 months each.

International Pharmacopoeia

The World Health Organization (WHO) took over the programs begun by the League of Nations, based on the work of the Brussels conferences in the early 1900s. Special emphasis was directed at the establishment of the *International Pharmacopoeia* (IP). The IP differs from national and regional pharmacopoeias in that it is not directly mandatory in few countries. Focus is on the needs of developing countries where acquisition and use of modern, automated analytical technologies may be difficult. The IP provides analytical methodologies that can be used where more advanced technologies are not readily available. The WHO program for International Non-proprietary Names is coordinated with the U.S. Adopted Names Council. Over the years, members of the Committee of Revision and members of the USP staff have served on the Expert Committees on Specifications for Pharmaceutical Preparations and Antibiotics, continuing the USP's long-standing commitment to WHO drug quality programs.

In accordance with Article 41 of its Constitution, the WHO published a first edition of the *International Pharmacopoeia* in 1952. Because no official national representatives were part of the Commission, the *International Pharmacopoeia* was considered not to have legal force. That legal characteristic has continued. Nevertheless, the *International Pharmacopoeia* continues to attract the energies of pharmaceutical scientists from many countries. The *International Pharmacopoeia* has been positioned to address national needs not satisfied by the large regional pharmacopoeias such as the USP. The USP must be in accord with well-equipped analytical laboratories, technologically advanced manufacturers, highly varied and competitive markets, and therapeutic strategies and expectations characteristic of our intensive, personalized medical care. In contrast, the *International Pharmacopoeia* seeks to publish standards

using technology appropriate to some small- and medium-sized quality control laboratories (20). Basic tests to check drug identity and to exclude substantial decomposition are notable and extensive achievements by WHO experts. A scheme for certification of suppliers is also published.

Color tests have fallen into disuse in North America as convenient but analytically powerful instruments have come in the budgetary range for control laboratories. This condition does not yet pertain to much of the world. Attention to the provisions of the *International Pharmacopoeia* in those countries can make important contributions to the drug product quality received by their populations. To help realize this, the WHO has supplied detailed guidance for staffing and equipping small, medium, and large analytical laboratories with increasing degrees of technology and instrumentation (20).

The present scope of the *International Pharmacopoeia* continues to be that defined by Wieniawski (21), that of furnishing quality specifications for drug substances and general requirements for dosage forms. Thus, no monographs for individual dosage forms are published in the *International Pharmacopoeia*. WHO leadership may intend to include dosage forms in the future, but a more definitive statement cannot be made at this writing. The three volumes of the third edition give standards for "essential" drug substances (22).

European Pharmacopoeia

European Pharmacopoeia provisions now take precedence over all European national pharmacopoeias.

British Pharmacopoeia

Many similarities existed between the *British Pharmacopoeia* and the USP. Common elements of origin, continuing commerce, and a long history of cooperation have resulted in close correspondence in critical details. Both are legal instruments, both are recognized in the laws of countries other than those of origin, both establish standards for dosage forms as well as for bulk pharmaceutical chemicals. Both use experts from industry, academia, and government at all stages of standards development. Also, for most of its history, the *British Pharmacopoeia* was not a direct government effort.

Other National Pharmacopoeias

There are national pharmacopoeias established by governments to meet perceived needs for compendial standards for locally produced articles. A number of these

have published recent editions as evidence of continuing commitment, but it is often difficult to locate a U.S. bookseller for these. Cities of publication are Berlin, Paris, Prague, Rome, Tokyo, Peking, Seoul, Stuttgart, and Taipei. Most countries require imported articles to conform to one or another widely recognized pharmacopoeia, such as the USP or the *British Pharmacopoeia*.

USP Products and Services in Other Countries

Since 1982, the USP has developed distributorships in other countries to meet the needs of customers more directly. The Zentrallaboratorium Deutscher Apotheker [Central Laboratory of the German Pharmacists Association (ZL)] in Eschborn, Germany, was the first distributor and now serves clients in Europe and in the Middle East. The largest distributor is Promochem GmbH, with headquarters in Wesel, Germany, and branches throughout Europe. The British Pharmaceutical Society distributes Reference Standards throughout the United Kingdom. The Society for the Japanese Pharmacopoeia became a distributor in 1991. In Canada, at least two firms distribute Reference Standards. Distributor information appears in each bimonthly Reference Standard Catalog published in *Pharmacopoeial Forum*.

Agency for International Development

The USP had a cooperative agreement with the U.S. Agency for International Development (USAID) since 1992 to assist developing countries in establishing programs of pharmaceutical management (including procurement, storage, and distribution) and control (including country-specific formularies and drug information resources). This program is jointly funded by the USAID and the USP.

U.S. Department of Commerce

In 1992, the Department of Commerce (DOC) established the Special American Business Internship and Training (SABIT) program to train scientists from the newly independent states (NIS). The USP secured an agreement whereby the DOC sponsors travel expenses for six Russian scientists under the USP Visiting Scientists Program to learn techniques in drug analysis at the USP Drug Research and Testing Laboratory. To date, two scientists have each completed 4-month training programs. The laboratory in which these scientists will work will be

established in Moscow and funded by the World Bank. The International Foundation for Drug Efficacy and Safety (IFDES), a not-for-profit group chartered under Russian, Swiss, German, and U.S. laws, is a voluntary group working with the NIS to assist in restoring their programs of drug supply regulation and quality control.

REFERENCES

1. *The United States Pharmacopoeia*, 24th Rev. The National Formulary, 19th Ed.; The United States Pharmacopoeial Convention, Inc.: Rockville, MD, 1999; Both Official from January 1, 2000.
2. Rules and Procedures of the General and Executive Committees of Revision for 1995–2000. *Pharmacopoeial Forum* **1995**, 21, 1619.
3. USP, 24th Rev., xxxvii–xiii.
4. Johnson, C.A. The Purposes and Limitations of a Pharmacopoeia. International Symposium on the Compendia and Industry, Montreal 1985; *Pharmazeut. Ind.*: 1986; 48, 760.
5. Grady, L.T.; J. A. Pharm. Assoc. **1976**, 16, 603, NS.
6. Boorstin, Daniel J. *The Americans—The Democratic Experience*; Chap. 22, Random House: New York, 1973.
7. Descartes, R. *Discourse on the Method*; [L. J. Lafleur transl. 1637.
8. Added Substances. *Ingredients and Processes*; USP, 24th Rev., General Notices.
9. *Tests and Assays—Procedures*; USP, 24th Rev., 7 General Notices.
10. USP, 24th Rev., General Notices.
11. Grady, L.T.; *Pharmazeut. Ind.* **1983**, 45, 640.
12. King, R.H.; Grady, L.T.; Reamer, J.T. *J. Pharm. Sci.* **1974**, 63, 1571.
13. Grady, L.T. Cooperation between USP and Industry [see Ref. 5].
14. Heller, W.M. Private Communication.
15. Florey, K. Reflections of a Quality Control Watcher. *Progress in the Quality Control of Medicines*; Deasey and Timonesy, Ed.; Elsevier Biomedical Press: Amsterdam, 1981; 15.
16. Stainer, C. *Ceremony of the Presentation of the European Pharmacopoeia*; Council of Europe: Strasbourg, France, 1969; 1, 5–71.
17. Schorn, P.J. International Harmonization of Pharmacopoeial Standards [see Ref 5].
18. *European Pharmacopoeia*, 2nd Ed.; II. 10th Fascicle; Council of Europe: Strasbourg, France, 1986.
19. Sagath, J. Proceedings of the International Meeting on Pharmacopoeias and Quality Control of Drugs, Cingolani, E., Ed.; 228 *Annali Istit. Super. Sanita* **1975**, 11, 228.
20. World Health Organization. *WHO Technical Report Series*; World Health Organization: Geneva, 1984.
21. Wieniawski, W. Proceedings of the International Meeting on Pharmacopoeias and Quality Control of Drugs, 204–210.
22. World Health Organization. *The International Pharmacopoeia*, 3rd Ed.; World Health Organization: Geneva, 1988.

PHOTODECOMPOSITION OF DRUGS

Hanne Hjorth Tønnesen

University of Oslo, Oslo, Norway

INTRODUCTION

Numerous compounds degrade when exposed to light. Some light-sensitive drugs are rapidly affected, either by natural light (particularly ultraviolet) or by artificial light (e.g., fluorescent light). This may lead to a change in the physicochemical properties of the product, e.g., a precipitate is formed, the product becomes discolored or cloudy, a loss in viscosity is observed, or the active ingredient undergoes photodegradation, which may not be visually detectable. The most obvious result of drug photodecomposition is a loss of potency of the product. Consequently, this can result in a drug preparation that is therapeutically inactive. Although this does not occur often, the study of photodegradation is important because the decomposition products that might occur in the formulation during storage may be toxic. Such products may also develop by the action of sunlight on the epidermal layers of the skin or in the eye of patients receiving the drug and may thereby cause photosensitivity reactions. Photostability testing is therefore an essential part of product development and is needed to ensure that satisfactory product quality is maintained during practical usage.

In this chapter, various aspects of drug photoreactivity are presented. Common photochemical reactions by which drugs have been found to decompose are described. The influence of formulation factors, e.g., excipients, on product stability is discussed. Finally, variables that have to be considered in photostability studies are noted.

PHOTOREACTIVITY OF DRUG SUBSTANCES

The photochemical reaction is a complex process. During the last decade, a body of data relating to drug degradation pathways has been accumulated and recently reported (1–3). The light-sensitive drug molecules may be affected directly or indirectly by irradiation, depending on how the radiant energy is transferred to the drug molecules. Primary photochemical reactions occur when the drug molecule itself absorbs energy, i.e., when there is a certain overlap between the absorption spectrum of the molecule and

the incident radiation. Many drug substances are white, and thus the degradation depends primarily on the amount of UV radiation absorbed by the material. Colored substances absorb light in the visible region of the spectrum. Degradation products formed during shelf life can be colored and thereby change the overall absorption characteristics of the formulation. Any overlap of the product absorption spectrum with the photon source impinging on it has the potential to cause a photochemical change. One consequence may be that a drug substance with an absorption spectrum that does not overlap with the photon source still photodecomposes in a formulated product. This takes place in a process called photosensitization. The energy absorbed by the nondrug molecule is imparted to the active ingredient, which subsequently degrades. The absorbing component is called a photosensitizer. The sensitizer may transfer the absorbed energy completely and not be altered itself in the process, but in many cases it will undergo some degradation. Photochemical stability of a drug compound in a formulation therefore cannot be predicted from only the absorption spectrum/stability studies of the drug in a pure solvent. Stability studies of the drug substance in the final product must also be considered.

Light acts as a reactant and never as a catalyst in a photochemical reaction. Light energy (a photon) is absorbed by the promotion of an electron from an initially occupied, low-energy orbital to a high-energy, previously unoccupied orbital. The result is an electronically excited molecule. In the singlet state, the electron spins are paired (antiparallel), whereas in the triplet state, the electron spins are unpaired (parallel). Excited states are both better donors and better acceptors of electrons than are ground states. They are unstable and often return to the ground state by dissipating their energy as heat and/or light. In many cases, the excited state leads to the formation of high-energy products such as free radicals or radical ions. Photochemical processes usually occur in two stages. The primary reaction is the reaction directly attributable to the absorption of a photon, i.e., it involves the excited state. This reaction does not depend on temperature for activation of the molecules. The primary photochemical reaction, however, will often be followed by secondary (thermal)

reactions occurring from the intermediates produced by the primary photochemical process (e.g., radicals, radical ions). These intermediates can eventually react through "dark" reactions to form the final, stable products. In some cases, the final products may resemble the products of the purely thermal reaction (dark reaction from the molecular ground state), but this similarity is coincidental.

Drugs are usually stored in contact with atmospheric oxygen and are always in contact with dissolved oxygen in circulating blood and living tissues. Energy transfer from an excited drug molecule will often lead to the formation of reactive oxygen species (ROS)(4). A type I photosensitized reaction proceeds through the transfer of electrons or protons. The resulting cation or neutral radical is likely to undergo further reactions. In the absence of oxygen, dimerization or recombination can result. When oxygen is available in sufficient concentrations, molecular oxygen will rapidly add to the radical, leading to the formation of peroxy radicals. These radicals are highly reactive and tend to abstract a proton from neighboring molecules. This sequence is often described as a chain reaction. In another photosensitization mechanism, an electron is transferred from an excited drug molecule to ground-state oxygen to give an anion radical known as superoxide. Superoxide is a powerful oxidizing agent and very toxic to biological systems.

Most molecules are singlet in the ground state. The oxygen molecule is unusual in that it is a triplet. In a type II photosensitized reaction, energy transfer from an excited drug molecule to ground state oxygen gives excited oxygen in the singlet state. Singlet oxygen is very reactive toward organic molecules and can lead to the formation of hydroperoxides by addition to olefinic bonds.

A relationship between structure and photoreactivity can be difficult to predict, but certain structural types are

known to have a high possibility of photodecomposition. Examples are shown in Table 1. Olefinic carbon-carbon double bonds can undergo *cis-trans* isomerization and may cyclize as a result; arylacetic acids are likely to decarboxylate; and some amines *N*-Dealkylate and haloaromatic compounds tend to dehalogenate. A drug can also sensitize its own degradation or drugs in a mixture can undergo cross-sensitization (5). Unfortunately, several pathways are reported for many drugs, complicating the elucidation of mechanisms. An extensive list of common photoreactions of drug substances was compiled by Greenhill and McLelland (1).

A drug that displays photochemical reactivity in vivo may cause adverse photosensitivity effects in patients. Sunlight penetrates the skin to a sufficient depth to reach molecules circulating in the surface capillaries, or it can react with compounds accumulated in the eye. In both cases, sunlight may convert the drug to a toxic decomposition product or induce the formation of ROS. Singlet oxygen and superoxide are both toxic to human tissues. Cutaneous photosensitization can be classified as phototoxicity, in which the skin reactions derive directly from photosensitized damage to the cellular components of the skin, and photoallergy, in which the mechanisms of the immune system are activated. Phototoxic reactions may be oxygen-dependent (photodynamic) or oxygen-independent (nonphotodynamic). In some cases, photodegradation products can circulate in the blood stream and cause damage to deep-seated organs. However, it should be noted that a combination of drugs and light can be beneficial, such as in the treatment of vitiligo, psoriasis, and skin cancer and in the development of site-specific drug-delivery systems.

The rate of a photochemical reaction is in general dependent on the rate at which light is absorbed by the system (i.e., the number of photons absorbed per second) and the efficiency of the photochemical process (i.e., the quantum yield for the reaction). The quantum yield is usually independent of wavelength. A primary photochemical reaction follows first-order kinetics in a formulation that contains the drug substance in a low concentration (6). The kinetics is more complicated at higher concentrations. Most of the light will then be absorbed close to the sample surface. The drug molecules inside the volume become protected from irradiation and do not participate in the primary reaction (inner filter effect). If the concentration of the active compound is high (absorbance approaches 2), the degradation kinetics may follow a pseudo-zero-order rate. In this case, essentially all the light is absorbed by the drug, and the rate-limiting factor is the light intensity. It is, however, the overlap integral that determines the rate of the reaction. Because of a small overlap integral, first-order kinetics can be

Table 1 Chemical functions that are expected to introduce photoreactivity

Carbonyl group
Nitroaromatic group
<i>N</i> -oxide function
Carbon-carbon double bond
Aryl halide
C—H bond α to an amine nitrogen
C—H bond at a benzylic position
Reactions with singlet oxygen
Alkenes
Polyenes
Phenol
Sulfides

observed even if the product contains a large amount of the absorbing species. The rate of disappearance of the drug in a sensitized reaction may be dependent on both the drug and the sensitizer concentrations, i.e., the degradation process follows second-order kinetics. It is difficult to use reaction order to characterize photodegradation in the solid state. The photochemical processes take place on the product surface, and the change in total concentration measured as a function of irradiation time does not necessarily follow any particular reaction order model (7).

INFLUENCE OF PRODUCT FORMULATION ON DRUG PHOTOSTABILITY

Most photochemical reactions are affected by the medium, i.e., both by excipients and by the type of formulation (8). Examples of formulation factors influencing product photostability are given in Table 2. A drug substance in an intravenous infusion product is likely to be presented as a dilute solution, whereas the active compound of a topical preparation represents a situation part way between the solution phase and the solid state. Parenteral solutions ensure high light impingement on the drug molecules owing to a large surface-to-volume ratio and usually a low drug concentration. In the solid state; the depth of light penetration is determined by the characteristics of the sample surface. For example, a capsule and a tablet have different light-scattering characteristics and different surface area-to-volume ratios.

Cosolvents, such as ethanol or higher alcohols, and surfactants are often used in a formulation to enhance the solubility of sparingly soluble compounds or to modify the stability of the drug molecule. Cosolvents will change the polarity and, in some cases, the viscosity of the medium. This may influence the photochemical reactivity of a compound. An increase in solubility will change the sample absorbance. Dissolution of particle aggregates leads to an increase in light impingement on the separate drug molecules. Surfactants will form micelles that can interact with drug molecules. This can lead to a change in drug molar absorptivity or in reactivity owing to a micro-environmental effect. Cosolvents and surfactants can have a photostabilizing or photodestabilizing effect on the product, as demonstrated for phenobarbital and nitrofurazone, respectively (9, 10).

For many drug substances, the photodegradation process is highly dependent on the ionization form of the molecule (e.g., ciprofloxacin, midazolam, and chloroquine) (11–13). The dicationic form of the antimalarial drug mefloquine is almost not photolyzed, whereas the monocationic and neutral forms of the molecule readily undergo photodecomposition (14). The oxygen concentration plays an important role in the process. Flushing the mefloquine samples with helium during irradiation leads to a substantial increase in the degradation rate. Destabilization by removal of oxygen is also demonstrated for other drugs such as nitrazepam (15). For oxygen-sensitive compounds, a stabilizing effect would ordinarily be obtained by purging the solution and headspace with an inert gas. However, it should be recalled that for reactions

Table 2 Formulation factors that may influence the photostability of drugs

Solutions	Solid preparations	Miscellaneous
Drug concentration	Crystal modification	Cyclodextrins (type)
Solvent system	Particle size	Liposomes (type, charge)
Cosolvents	Particle surface (porous, smooth)	Micelles (type, charge)
Surfactants		Emulsifying agent
pH	Color	Partition coefficient
Oxygen concentration	Coating	
Buffer salt (type, concentration)	Thickness of powder bed	
Ionic strength	Container	
Metal ions		
Chelating agents		
Antioxidants		
Preservatives		
Sweetening agents		
Tonicity adjustors		
Colors		
Container		

in which oxygen participates only catalytically, even trace amounts of residual oxygen may render the product unstable. An increase in ionic strength is reported to have a photostabilizing effect on certain drugs by providing a protective film of solvated ions around the reacting molecule (16). This effect is not observed in mefloquine (14). On the other hand, an increase in degradation rate was measured as a function of phosphate ion concentration in the buffer system used. The phosphate ion is known to influence the photochemical properties of compounds (e.g., tyrosine) by facilitating proton transfer from the excited state of the reacting species (17). Various types of buffer salts exert different effects on the photodegradation process, as demonstrated for the drug daunorubicin (18). Buffer salts such as citrate can also change the absorption characteristics of the formulation by forming complexes with other components present, leading to products that absorb in the visible part of the spectrum.

In many cases, the degradation rate is highly dependent on the presence of trace metals. A variety of chelating agents and antioxidants is available for use in pharmaceutical preparations. The most commonly used in aqueous systems are bisulfite and EDTA. Many drugs have shown reactivity toward bisulfite. A significant photodestabilizing effect on epinephrine has been demonstrated (19). Fe(III)-EDTA chelates are reduced by superoxide quite rapidly. EDTA therefore will not inhibit degradation of drugs in systems in which the iron-catalyzed Haber-Weiss reaction plays an important role. The effect of antioxidants and chelating agents on drug instability must be carefully evaluated before use. Other excipients such as tonicity adjusters, colors (see below), and sweetening agents could further interact with commonly used stabilizers or influence the photoreactivity of the drug substance (20–23).

Photoprotection by spectral overlay with suitable excipients can stabilize various drugs and preparations. Incorporation of the yellow compound curcumin in the shell of soft gelatin capsules was demonstrated to have a protective effect on drugs such as nifedipine, chloramfenicol, frusemide, and clonazepam (24). The addition of colors is further shown to stabilize drugs in tablets, solution, and topical preparations (25). A mixture of colors or pigments, however, can undergo catalytic fading (26) or induce degradation of other components in the formulations by radical formation (27, 28).

Photodegradation in the solid state takes place only at the sample surface. The degradation rate is therefore dependent on factors that will influence the depth of light penetration, i.e., change the absorption and reflection at the surface (e.g., particle size, crystal modification, color, thickness of powder bed, and coating of the individual

particles or the dosage form). Mefloquine, chloroquine, carbamazepine, and furosemide are examples of drug substances that show different decomposition rates dependent on their polymorphous modification (29–32).

A change in drug photoreactivity by complexation with suitable carriers can be observed both in solution and in the solid state. Metronidazole was found to be less sensitive to irradiation after complex formation with sodium urate (33). The extent of photodegradation of various drugs including phenothiazine has been reduced by inclusion complexation with cyclodextrins (34–36). There are marked differences in stabilizing effect among various types of cyclodextrin. Liposomes or a combination of liposomes and cyclodextrins are also demonstrated to improve drug photostability (37, 38).

The method used most commonly to protect photosensitive drugs is to place the preparation in a protective market pack or in a colored or amber immediate container. During storage and use, the protective market pack may be removed. In cases in which the immediate container is made of transparent glass or plastic material, little protection against radiation is obtained. Amber containers offer better protection but vary in thickness, chemical nature (plastic), and degree of coloration, which may influence their photoprotective properties. The stabilizing effect of amber glass as the only means of photoprotection is not satisfactory for highly photolabile drugs such as molsidomine (39). Even brown glass can offer inadequate protection, as demonstrated for drugs like epinephrine, isoprenaline, and levarterenol (40). The destabilizing effect of the container was attributed to release of alkali and traces of heavy metal ions that impart color to the glass.

PHOTOSTABILITY STUDIES

It is essential to obtain information about the photoreactivity of a drug molecule as early as possible in the formulation process (41). Knowledge about the photochemical and photophysical properties of a drug substance is important for the handling, packaging, and labeling of the product. One approach to evaluate overall photosensitivity is to design a photoassay, as noted in an earlier report (42). A basic protocol for testing new drug substances and products for first submissions is described in the ICH Guideline for photostability testing, which has been implemented since January 1998 (43). The Guideline notes that photostability testing should be an integral part of stress testing. Although the proposed test is reasonably simple to conduct, there are some variables that have to be controlled carefully during photostability studies.

Problems related to the application of the current guideline are discussed in a recent review (44). Important factors to consider are irradiation source, irradiation level, calibration, and presentation of the samples (45–48).

Various irradiation sources can be used in the stability studies of drugs and drug products. The source(s) selected should be comparable in spectral distribution with those to which products are exposed in practical use. The portion of radiant energy from the sun reaching the Earth includes the UVB (290–320 nm), UVA (320–400 nm), and visible (VIS) light (400–700 nm) ranges. It may be difficult, however, to predict the exact amount of UV and VIS irradiation to which the product is exposed during shelf life. The irradiation conditions therefore should provide a “worst-case” exposure. The ICH Guideline gives two options for the selection of irradiation source. Option 1 addresses exposure to outdoor daylight or window glass-filtered daylight. In most cases, a source providing glass-filtered daylight (ID65 according to ISO 10977) would be appropriate. Option 1 can be achieved by use of a fluorescent lamp that combines UV and VIS outputs or by use of a xenon or metal halide lamp. To simulate indoor conditions, it is necessary to use a window glass filter in combination with sources producing significant radiation below 320 nm (e.g., xenon and metal halide lamps, near-UV fluorescent tubes). Option 2 in the ICH Guideline addresses exposure to indoor fluorescent light (cool white). During shelf life the product is likely to be exposed to a mixture of natural light through window glass and artificial light (e.g., light from a fluorescent tube). A UV source is therefore added in Option 2 to cover the spectral region of natural light that may reach the sample. According to this option, the test is more severe than that for indoor fluorescent light alone, but it does not adequately simulate daylight through window glass.

The ICH Guideline recommends a total exposure of 200 Wh/m^{-2} in the UV range of 320 to 400 nm and 1.2 million lux hours in the VIS range (400–700 nm). For an average source simulating window glass-filtered daylight, a total irradiance of 200 Wh/m^{-2} in the UV region corresponds to approximately 0.45 million lux hours in the visible region. A test run with the end criterion of 1.2 million lux hours is therefore likely to exceed the requirement of 200 Wh/m^{-2} with a factor of up to 2.5 to 3 when a lamp according to Option 1 is selected, unless precautions are taken (44). In this context, Option 2 may be easier to handle because the UV lamp can be turned off when the desired UV exposure is obtained.

The ICH Guideline offers no guidance for the choice of Option 1 or 2. Although the options are not scientifically equivalent, they could be regarded as equivalent for purposes of a confirmatory study. It is important, however,

to combine the results with knowledge already obtained from other tests before labeling decisions are made.

The overall illumination is specified in the ICH Guideline, whereas the irradiance level is not indicated. An irradiance level that is high enough to accelerate the test without causing unwanted temperature effects must be selected. Tests conducted at significantly different irradiance levels could not be compared unless correlation has been established. The distance between the source and the sample surface should be defined to keep the radiant intensity constant. The irradiance at the specimen surface may also change as a function of the sample location inside the test chamber. The UV and visible levels should therefore be measured across the test chamber to ensure that the samples are placed at points of equal irradiance.

Even though the lamps are changed at defined intervals as specified by the producer, it is essential to calibrate the light source and periodically monitor its irradiance to obtain the predetermined exposure value. For instrumentation without a built-in sensor, calibration can be performed manually. The ICH Guideline recommends the use of a calibrated radiometer or a validated actinometric system to monitor the exposure in the UV region and a calibrated luxmeter to determine the overall illumination in the visible range. Neither will provide any information about the spectral distribution of the irradiation source. A UV filter radiometer is a broad-band meter designed to measure incident radiation in the UV region. An optical filter limits the spectral responsivity to a certain band (e.g., UVB and UVA). Unfortunately, there are no international standards for the filters and radiometers from different manufacturers, and therefore they may measure different fractions of radiant energy. A luxmeter will measure light as perceived by the human eye, i.e., it has a photopic response curve. The photopic sensitivity range of the human eye is approximately 380 to 780 nm, with the highest sensitivity at 555 nm. Neither the radiometer nor the luxmeter have a constant spectral responsivity over the actual wavelength range (i.e., the relative weighting given to the different wavelengths is not constant). The devices should be calibrated by the manufacturer at regular intervals. If the meters are used as received, they are well suited for measuring evenness of irradiance across the sample area and changes in total output with time. However, they cannot be used to give an absolute measurement of irradiance or to compare irradiance between sources unless they are calibrated specifically for each source (44). The ICH Guideline proposes quinine actinometry as an alternative to the radiometer for calibration in the UV range. The total irradiance is then determined by using a reaction of known photochemical efficiency (47, 49).

Samples may be placed in clear or amber glass, in plastic, in marketed containers, or in a Petri dish during exposure. The spectral transmittance characteristics of the container will influence the results. For example, the illuminance inside clear glass is higher than that inside amber, blue, or green glass, and the spectral distribution is different. Other factors to consider are the orientation of the sample container (upright, inverted, or sideways), thickness of the container material, and surface area/volume (weight) ratio. It is recommended in the ICH Guideline that the sample thickness should not exceed 3 mm for drug substances in the solid form. Preparations such as tablets or capsules should be spread in a single layer. The primary disadvantage in using a protective container is that a significant increase in temperature can be expected. Therefore, dark controls should be placed alongside the authentic sample. Stirring or shaking of samples during exposure is not recommended for the purpose of photostability testing, according to the ICH Guideline.

REFERENCES

- Greenhill, J.V.; McLelland, M.A. Photodecomposition of Drugs. *Prog. Med. Chem.* **1990**, *27*, 51–121.
- Greenhill, J.V. Is the Photodecomposition of Drugs Predictable? *Photostability of Drugs and Drug Formulations*; Tønnesen, H.H., Ed.; Taylor & Francis: London, 1996; 83–110.
- Albini, A.; Fasani, E. Photochemistry of Drugs: An Overview and Practical Problems. *Drugs. Photochemistry and Photostability*; Albini, A., Fasani, E., Eds.; The Royal Society of Chemistry: Cambridge, 1998; 1–74.
- Moore, D.E. Photophysical and Photochemical Aspects of Drug Stability. *Photostability of Drugs and Drug Formulations*; Tønnesen, H.H., Ed.; Taylor & Francis: London, 1996; 9–38.
- Moore, D.E.; Mallesch, J.L. Photochemical Interaction between Triamterene and Hydrochlorthiazide. *Int. J. Pharm.* **1991**, *76*, 187–190.
- Tønnesen, H.H. Photochemical Degradation of Components in Drug Formulations. I. An Approach to the Standardization of Degradation Studies. *Pharmazie* **1991**, *46*, 263–265.
- Sande, S.A. Mathematical Models for Studies of Photochemical Reactions. *Photostability of Drugs and Drug Formulations*; Tønnesen, H.H., Ed.; Taylor & Francis: London, 1996; 323–340.
- Kerker, R. Untersuchungen zur Photostabilität von Nifedipin, Glucocorticoiden, Mosidomin und Ihren Zubereitungen, München, 1991, Dissertation.
- Shahjahan, M.; Enever, R.P. Photolability of Nitrofurazone in Aqueous Solution. II. Kinetic Studies. *Int. J. Pharm.* **1996**, *143*, 83–92.
- Asker, A.F.; Islam, M.S. Effect of Sodium Thiosulphate on the Photolysis of Phenobarbital: Evidence of Complex Formation. *PDA J. Pharm. Sci. Technol.* **1994**, *48*, 205–210.
- Torniaainen, K.; Tammilehto, S.; Ulvi, V. The Effect of pH, Buffer-Type and Drug Concentration on the Photodegradation of Ciprofloxacin. *Int. J. Pharm.* **1996**, *132*, 53–61.
- Andresin, R.; Tammilehto, S. Photochemical Decomposition of Midazolam. IV. Study of pH-Dependent Stability by High-Performance Liquid Chromatography. *Int. J. Pharm.* **1995**, *123*, 229–235.
- Nord, K.; Orsteen, A.-L.; Karlsen, J.; Tønnesen, H.H. Photoreactivity of Biologically Active Compounds. X. Photoreactivity of Chloroquine in Aqueous Solution. *Pharmazie* **1997**, *52*, 598–603.
- Tønnesen, H.H. Photoreactivity of Biologically Active Compounds. XV. Photochemical Behaviour of Mefloquine in Aqueous Solution. *Pharmazie* **1999**, *54*, 590–594.
- Cornelissen, P.J.G. Beijersbergen van Henegouwen, G.M.J. Photochemical Decomposition of 1,4-Benzodiazepines. Nitrazepam. *Photochem. Photobiol.* **1979**, *30*, 337–341.
- Chinnian, D.; Asker, A.F. Photostability Profiles of Minoxidil Solutions. *PDA J. Pharm. Sci. Technol.* **1996**, *50*, 94–98.
- Lakowicz, J.R. *Principles of Fluorescence Spectroscopy*; Plenum Press: New York, 1983; 360.
- Islam, M.S.; Asker, A.F. Photoprotection of Daunorubicin Hydrochloride with Sodium Sulfite. *PDA J. Pharm. Sci. Technol.* **1995**, *49*, 122–126.
- Brustugun, J.; Tønnesen, H.H.; Klem, W.; Kjønneksen, I. Photodegradation of Epinephrine by Sodium Metabisulfite. *PDA J. Pharm. Sci. Technol.* **2000**, *50*, 136–143.
- Asker, A.F.; Canady, D. Influence of Certain Additives on the Photostabilizing Effect of Dimethylsulfoxide for Sodium Nitroprusside Solutions. *Drug Dev. Ind. Pharm.* **1984**, *10*, 1025–1039.
- Asker, A.F.; Harris, C.W. Influence of Certain Additives on the Photostability of Physostigmine Sulfate Solutions. *Drug Dev. Ind. Pharm.* **1988**, *14*, 733–746.
- Ho, A.H.L.; Puri, A.; Sugden, J.K. Effect of Sweetening Agents on the Light Stability of Aqueous Solutions of L-Ascorbic Acid. *Int. J. Pharm.* **1994**, *107*, 199–203.
- Asker, A.F.; Colbert, D.Y. Influence of Certain Additives on the Photostabilizing Effect of Uric Acid for Solutions of FD&C Blue No. 2. *Drug Dev. Ind. Pharm.* **1982**, *8*, 759–774.
- Tønnesen, H.H.; Karlsen, J. Studies on Curcumin and Curcuminoids. X. The Use of Curcumin as a Formulation Aid to Protect Light-Sensitive Drugs in Soft Gelatin Capsules. *Int. J. Pharm.* **1987**, *38*, 247–249.
- Thoma, K. Photodecomposition and Stabilization of Compounds in Dosage Forms. *Photostability of Drugs and Drug Formulations*; Tønnesen, H.H., Ed.; Taylor & Francis: London, 1996; 111–140.
- Kuramoto, N.; Kitao, T. Mechanism of the Photofading of Dye. Contribution of Singlet Oxygen in the Catalytic Fading of Anthraquinone Dye Mixtures. *J. Chem. Technol. Biotechnol.* **1980**, *30*, 129–135.
- Skowronski, T.A.; Rabek, J.F.; Rånby, B. The Role of Commercial Pigments in the Photodegradation of Poly (Vinyl Chloride) (PVC). *Polym. Degrad. Stab.* **1984**, *8*, 37–53.
- Sidhu, D.S.; Sugden, J.K. Effect of Food Dyes on the Photostability of Aqueous Solutions of L-Ascorbic Acid. *Int. J. Pharm.* **1992**, *83*, 263–266.

29. Nord, K.; Andersen, H.; Tønnesen, H.H. Photoreactivity of Biologically Active Compounds. XII. Photostability of Polymorphic Modifications of Chloroquine Diphosphate. *Drug Stab.* **1997**, *1*, 243–248.
30. Tønnesen, H.H.; Skrede, G.; Martinsen, B.K. Photoreactivity of Biologically Active Compounds. XIII. Photostability of Mefloquine Hydrochloride in the Solid State. *Drug Stab.* **1997**, *1*, 249–253.
31. Matsuda, Y.; Akazawa, R.; Teraoka, R.; Otsuka, M. Pharmaceutical Evaluation of Carbamazepine Modifications: Comparative Study for Photostability of Carbamazepine Polymorphs by Using Fourier-Transformed Reflection-Absorption Infrared Spectroscopy and Colorimetric Measurement. *J. Pharm. Pharmacol.* **1994**, *46*, 162–167.
32. de Villiers, M.M.; van der Watt, J.G.; Lötter, A.P. Kinetic Study of the Solid State Photolytic Degradation of Two Polymorphic Forms of Furosemide. *Int. J. Pharm.* **1992**, *88*, 275–283.
33. Habib, M.J.; Asker, A.F. Complex Formation Between Metronidazole and Sodium Urate: Effect on Photodegradation of Metronidazole. *Pharm. Res.* **1989**, *6*, 58–61.
34. Lutka, A. Effect of Cyclodextrin Complexation on Aqueous Solubility and Photostability of Phenothiazine. *Pharmazie* **2000**, *55*, 120–123.
35. Mielcarek, J. Photochemical Stability of the Inclusion Complexes of Nicardipine with α -, γ -Cyclodextrin, Methyl- β -Cyclodextrin and Hydroxypropyl- β -Cyclodextrin in the Solid State and in Solution. *Pharmazie* **1996**, *51*, 477–479.
36. Thoma, K.; Kübler, N. Einfluss von Hilfsstoffen auf die Photozersetzung von Arzneistoffen. *Pharmazie* **1997**, *52*, 122–129.
37. Habib, M.J.; Asker, A.F. Photostabilization of Riboflavin by Incorporation into Liposomes. *J. Parent. Sci. Technol.* **1991**, *45*, 124–127.
38. Loukas, Y.L.; Jayasekera, P.; Gregoriadis, G. Novel Liposome-Based Multicomponent System for the Protection of Photolabile Agents. *Int. J. Pharm.* **1995**, *117*, 85–94.
39. Thoma, K.; Kübler, N. Einfluss der Wellenlänge auf die Photozersetzung von Arzneistoffen. *Pharmazie* **1996**, *51*, 660–664.
40. Wollmann, H.; Grünert, R. Einfluss des Sichtbaren Lichtes auf die Haltbarkeit von Isoprenalin-, Epinephrin- und Levaterenollösungen in Unterschiedlichen Behältnissen. *Pharmazie* **1984**, *39*, 161–163.
41. Merrifield, D.R.; Carter, P.L.; Clapham, D.; Sanderson, F.D. Addressing the Problem of Light Instability During Formulation Development. *Photostability of Drugs and Drug Formulations*; Tønnesen, H.H., Ed.; Taylor & Francis: London, 1996; 141–154.
42. Tønnesen, H.H.; Kristensen, S.; Nord, K. In Vitro Screening of the Photoreactivity of Antimalarials: A Test Case. *Photostability of Drugs and Drug Formulations*; Tønnesen, H.H. Ed.; Taylor & Francis: London, 1996; 267–285.
43. ICH Harmonised Tripartite Guideline. Q1B: Photostability Testing of New Drug Substances and Products. Federal Register **1997**, *62*, 27115–27122.
44. Tønnesen, H.H. The ICH Photostability Guideline. A Discussion of Experimental Conditions. *Pharmaceutical Photostability and Stabilization Technology*; Thoma, K., Piechocki, J.T., Eds.; Marcel Dekker, Inc.: New York, In press.
45. Piechocki, J.T. Selecting the Right Source for Pharmaceutical Photostability Testing. *Drugs. Photochemistry and Photostability*; Albini, A., Fasani, E., Eds.; The Royal Society of Chemistry: Cambridge, UK, 1998; 247–271.
46. Boxhammer, J.; Willwoldt, C. Design and Validation Characteristics of Environmental Chambers for Photostability Testing. *Drugs. Photochemistry and Photostability*; Albini, A., Fasani, E., Eds.; The Royal Society of Chemistry: Cambridge, UK, 1998; 272–287.
47. Favaro, G. Actinometry: Concepts and Experiments. *Drugs. Photochemistry and Photostability*; Albini, A., Fasani, E., Eds.; The Royal Society of Chemistry: Cambridge, UK, 1998; 295–304.
48. Moore, D.E. Standardization of Photodegradation Studies and Kinetic Treatment of Photochemical Reactions. *Photostability of Drugs and Drug Formulations*; Tønnesen, H.H. Ed.; Taylor & Francis: London, 1996; 63–82.
49. Baertschi, S.W. Commentary on the Quinine Actinometry System Described in the ICH Draft Guideline on Photostability Testing of New Drug Substances and Products. *Drug Stab.* **1997**, *1*, 193–195.

PHYSIOLOGICAL FACTORS AFFECTING ORAL DRUG DELIVERY

Clive G. Wilson
Bridget O'Mahony
Blythe Lindsay

University of Strathclyde, Glasgow, United Kingdom

INTRODUCTION

Thirty years on from the explosion of commercially successful applications of targeted and controlled release pharmaceutical formulations, it is evident that there remains a need for further refinements and innovations in the field of drug delivery. Many of the larger corporations focus on the use of novel technology for extension of a product life cycle; however, in many cases solubility and permeability issues limit the application of such technologies. Additional complications arise due to inter- and intrasubject variability, compliance, and chronobiological variation in disease incidence.

Variability has many causes and an understanding of the factors that predict that certain therapeutic approaches will work better for some individuals than in others is key. This article will consider a few such examples from the literature and our own studies, which we hope will illustrate how physiological factors impact on disease treatment.

ORAL DRUG DELIVERY

In most cases, oral drug delivery is the cheapest and most convenient method of dosing. Unfortunately, it is difficult to achieve a precise control of the plasma-concentration-time profile by this route due to marked intra- and intersubject variation in gastrointestinal transit even under the rigidly controlled conditions of the clinical trial. Daily patterns of food intake, activity, and posture are large contributors to this variation. Drugs that are only absorbed from specific areas of the gastrointestinal tract, i.e., have a narrow "window of absorption," will be most affected by alterations in transit. The major determinants of this variation will be the amount of food and drink consumed.

In the western world, the average adult consumes and excretes between 1.5 and 2.5 L of fluid per day. The liquid is consumed as beverages whose intake is closely linked to the level of salt in the diet or as water in the fluid component of food. In addition, metabolic processes

generate approximately 350 ml of water per day. Water turnover in the epithelial tissues of the gut (secretion and absorption) is estimated at 9 L per day and in view of this prodigious flux, we might expect that most of the internal cavities of the body have water in excess.

This is however, not true and certain environments of the gut might be regarded as moist rather than wet: for example, the oropharynx, the esophagus, and most of the large bowel. This lack of water increases the variability of drug delivery to these areas.

ESOPHAGEAL TRANSIT

For dosage forms to reach the stomach they must first pass through the oesophagus, a 25–30 cm long moist hollow tube. It is commonly assumed that swallowed dosage forms pass without hindrance into the stomach unless an underlying oesophageal condition is present. Up to 60% of healthy subjects over the age of 60 years report a problem in swallowing intact tablets (1). To assist swallowing of a tablet, patients are instructed to take a dosage form with plenty of water. It might be expected that simple encouragement and education could encourage compliance. In practice, when patients are presented with a 240-ml glass of water and instructed to swallow a tablet "according to normal practice," they imbibe only two to three mouthfuls (between 50 and 100 ml). Fig. 1 shows the variation in water volume imbibed with a placebo tablet under these conditions. These data suggest that patients are unlikely to be concordant with physician's or pharmacist's instructions.

Water intake and posture are recognized as important factors in avoiding "pill-erosion" of the oesophagus. This condition is usually caused by drugs that are irritant, highly acidic or basic, e.g., NSAIDs, potassium chloride, ferrous salts, tetracycline, and rifampicin (2–4). If tablets are taken without water or when semirecumbent, there is a risk that they may remain lodged in the lower third of the oesophagus. This may result in mucosal damage as the formulation starts to dissolve, causing a high local

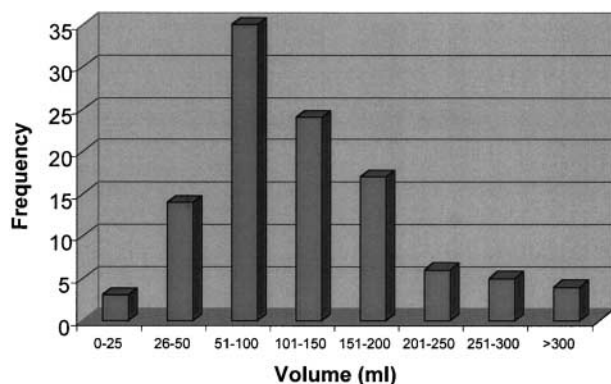


Fig. 1 Volume of water swallowed with a capsule when allowed water ad lib. (Data courtesy of Dr. Richard Dansereau, to be published.)

concentration of drug and dehydrating the surface epithelium. In addition, components of the formulation may help to form a strong bioadhesive bond between the unit and the tissue surface, aiding retention.

Esophageal retention frequently occurs even in clinical trials where correct dosing is carefully monitored. Figure 2 illustrates esophageal retention of a radiolabelled fast-dissolving analgesic hard gel formulation administered to a young volunteer. Subjects received two doses of the radiolabelled formulation. In this subject, the first dose cleared normally while the second adhered to the lower third of the oesophagus leading to significant reduction in peak plasma paracetamol concentration compared with other volunteers (Fig. 3).

As we get older, the ability to swallow certain kinds of formulations becomes problematic. In a series of studies conducted at the Queen's Medical Center, Nottingham, it was found that elderly patients had problems swallowing

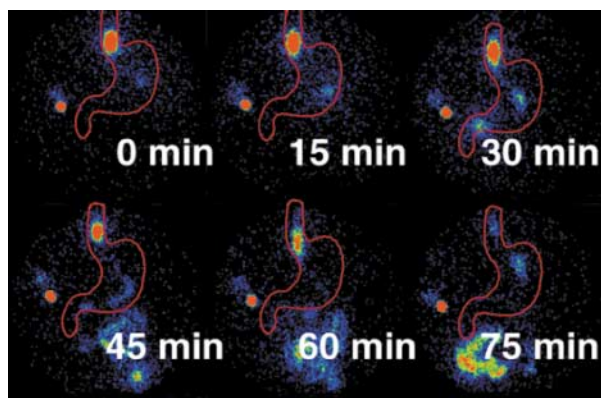


Fig. 2 Radiolabelled hard gelatin capsule lodged in the esophagus.

hard gelatin capsules. This appears to be due to an inability to coordinate the cricopharyngeal reflex to both water and capsule bolus (5, 6).

GASTRIC EMPTYING

The process of gastric emptying is extremely complex and is influenced by many factors such as presence of food, food content, pH, and posture. Most drugs are not absorbed from the stomach and are therefore dependent on the gastric emptying process to deliver them to their site of absorption. Consequently, the process and factors that affect it have been extensively studied with many conflicting views reported.

The stomach can exist in two states: fed and fasted. The empty stomach has a volume of approximately 50 ml, that increases to about 1 L when full. The fate of a dosage form is dependant on the state of the stomach at time of administration.

In the fed state, the environment within the stomach is infinitely variable in terms of food content. The nature of food intake is not only specific to race and geographic location but unique for each person and can vary significantly on a day-to-day basis. An average daily intake of 3–4 kg of food and drink is typical and some 5 L of fluids such as saliva, gastric juice, pancreatic juice, and other body liquids are added to the stomach contents during the day. The liquid–solid content within the stomach is a key factor in determining the fate of ingested material. Liquids and solids have separate and distinct emptying patterns.

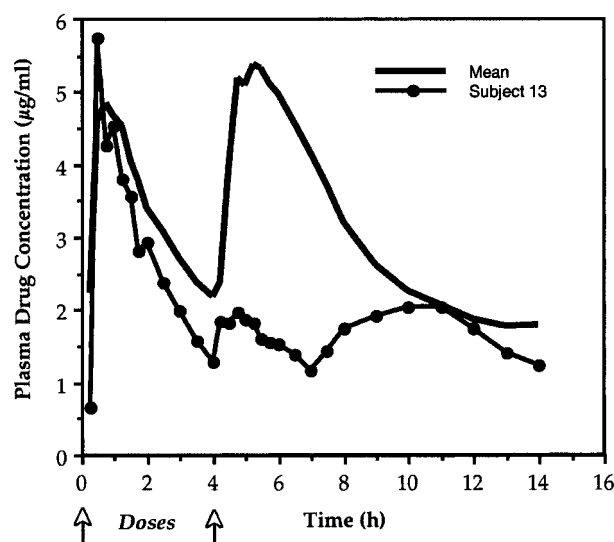


Fig. 3 Reduction in peak plasma-concentration on second dose.

Following a meal, liquids empty first in a monoexponential pattern, before the gastric emptying of solids is initiated.

The stomach acts as sieve and a calorie regulator for the small intestine: As a consequence, meals that contain large fragments of food or are nutrient-dense will take longer to empty and hence will delay the passage of dose forms to the small intestine. Small tablets and pellets mix and empty from the stomach with food, but large nondisintegrating tablets are reliant upon the migrating myoelectric complex (MMC, the patterns of motor activity that act to move food through the gastrointestinal tract) in the fasted pattern of motility to empty them from the stomach. If the stomach is maintained in the fed state by continuous feeding then large dosage forms are retained in the stomach for that period.

Often patients are instructed to take their medication "with a meal," but instructions are never precise and this can be interpreted by the patient as taking the medication immediately before, during the meal, or just after the food. O'Reilly and co-workers (7) demonstrated that the initial-gastric emptying rate of multiparticulates dosed before a meal was faster than for those dosed during and after food.

Gastric emptying follows a circadian rhythm with slower emptying occurring in the afternoon compared with the morning. Such variation can be very marked: In a study by Ghoo and colleagues emptying of the solid phase of the meal decreased by over 50% when it was consumed in the evening (8). The fasted pattern of motility establishes itself overnight, as this is generally the longest period in which no food intake occurs. MMC occur with the greatest frequency in the early hours of the morning. Changes in MMCs probably produce the largest physiological variability in oral drug delivery.

FAT AND GASTRIC HOMOGENEITY

The fat content of food is one of, if not the, most important influence on the rate of gastric emptying. Within the stomach, fat will separate and form layers exposing the pyloric-duodenal region to different amounts of fat according to posture of the individual: erect, supine, prone, or lying on one side. Since digested fat has a marked inhibitory effect on gastric emptying, retardation will occur to varying degrees. Brown and colleagues found that fat quickly layered in the proximal stomach away from the pylorus, thus delaying gastric emptying (9). The delaying effect of lipid on gastric emptying is increased in the elderly (10).

If a dispersing phase such as minced meat is present, the fat will spread more uniformly. Edelbroek and colleagues labelled the fat phase of a meal with Tc-99m-thiocyanate,

to compare the gastric emptying and intragastric distribution of oil in a soup, based meal with and without minced beef (11). The emptying rate of oil in the oil/soupmeal was about twice that for oil consumed in the oil-soup-minced beef meal. These results show that major differences in the intra-gastric distribution of oil occur following incorporation of predispersed solids into the meal.

The gastric residence of a meal with identical composition will be prolonged if the fat content is used to fry the food rather than be ingested as the cold oil. The behavior of oils within the stomach is also affected by other constituents of the meal that are present. The effect of fat on gastric emptying and absorption of nutrients depends on the relation to the other components of the meal when the fat is consumed.

CARBOHYDRATE AND GASTRIC HOMOGENEITY

A study conducted in our laboratories looking at the absorption of a drug from low-volume, oil-filled soft gelatin capsules resulted in a variable bioavailability profile. Labelling the capsule with Tc-99m showed that formulation emptied immediately to the duodenum when given with fluid after a carbohydrate-based meal (sandwich). This unusual behavior was also investigated using magnetic resonance imaging (MRI). In this case, it was not necessary to label the formulation as the oil can be discriminated from the gastric contents using a T1, weighted FLASH sequence.

In the clinical trial, oil filled gelatin capsules were given immediately following a sandwich meal and the subjects were imaged in the prone position. Fig. 4 shows a cross-section of the body at the level of the upper abdomen: the agglomeration of the sandwich by the pyloric antrum into a "doughball" can be clearly seen. Following ingestion the meal is softened by mastication to aid swallowing. Although the maximum rate of salivary secretion is less than 5 ml min^{-1} , the bread is sufficiently softened. This moistened bolus is swallowed and the subsequent action of the stomach consolidates the mass. Eventually, the material is hydrated and the ball disperses. A small, almost buoyant gelatin capsule taken after the meal virtually bounces off the carbohydrate mass and exits from the stomach. This was confirmed by the gamma scintigraphic imaging study.

EFFECT OF POSTURE AND GRAVITY ON GASTRIC EMPTYING

Many studies have provided evidence that changes in posture can affect both the gastric emptying rates of

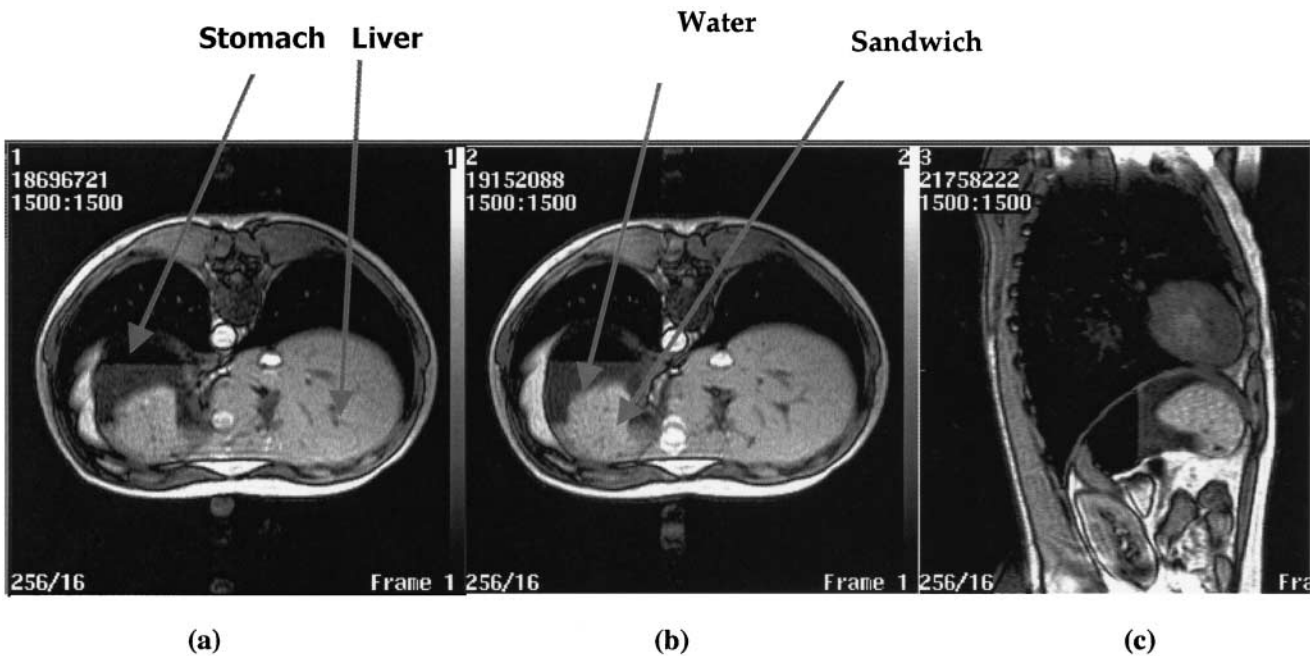


Fig. 4 MR images of the upper GI tract that show the compression of swallowed bread into a dough ball, viewed in (a) and (b) horizontal and (c) sagittal section.

materials and the absorption characteristics of drugs. Lying down decreases the rate of gastric emptying when compared with sitting and a combination of sitting and standing produces the most rapid gastric emptying. Bland, unbuffered liquids, and pellet formulations empty more slowly from the stomach when the subject is lying down compared with when he is upright or sitting erect. For floating formulations, such as raft-forming alginates, the buoyant raft empties faster than food in subjects lying on their left side or their backs and slower in subjects lying on their right side with the raft positioned in the greater curvature. When the subjects lay on their left side the raft was presented to the pylorus ahead of the meal and so emptied first.

The time to maximum plasma-concentration (T_{max}) of coadministered soluble paracetamol and nifedipine is significantly decreased when subjects are standing or lying on their right side compared with when they lie on their left side (12). These postures also resulted in a significantly higher peak plasma-concentration and area under the plasma-concentration–time curve of nifedipine.

GENDER AND PREGNANCY

Studies have examined the effect of gender and the menstrual cycle on gastrointestinal transit. Gastric

emptying of both solids and liquids is reported to be slower in women than men (13). Premenopausal women and postmenopausal women taking oestrogen and progesterone replacement were shown to have slower

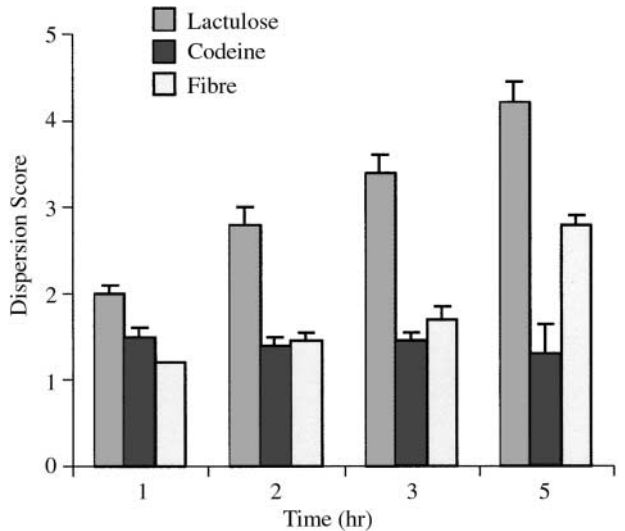


Fig. 5 Mean plasma quinine concentrations following dosing with 50 mg of the dihydrochloride salt in the Pulsincap after treatment with codeine or lactulose. Error bar shows standard deviation.

gastric emptying of liquids than men. However, while premenopausal women and postmenopausal women taking hormone replacement therapy showed slower gastric emptying of liquid, postmenopausal women not taking hormones had a similar rate to men (14). It has been shown that females have faster emptying times when in the ovular stage of the menstrual cycle.

In women, pregnancy can drastically alter transit of drugs in certain regions of the gastrointestinal tract, although gastric emptying is not delayed (15). Nausea and vomiting occurs in 50–90% of pregnant women and may be so severe that hospitalization is required to prevent dehydration. Often pregnant women suffer from heartburn and constipation, which is attributed to decreased oesophageal sphincter pressure and impaired colonic motility (16).

SOCIAL EFFECTS—SMOKING AND ALCOHOL

Scintigraphic studies do not clarify the effect of smoking on transit. One study reported that smoking or chewing nicotine gum did not appear to affect oesophageal transit or the rate of liquid emptying; however, solid emptying was delayed (17). Other studies reported delayed emptying on both liquids and solids, which was associated with increased periods of retrograde intragastric movement of solids from distal to proximal stomach (18, 19). This suggests an initial impairment of antral function and/or a lack of antropyloroduodenal coordination once a contraction is initiated.

Strong alcoholic drinks such as whisky delay gastric emptying. A study showed that administration of beer or white wine significantly accelerated gastric emptying compared with ethanol of the same concentration (20). This suggests that the observed effect is due to compounds in the beer and white wine other than alcohol. Gastric emptying is delayed in 24% of chronic alcoholics (21). Finally, smoking delays the gastric emptying of alcohol indicating the possibilities of complex interactions due to social activities (22).

SMALL INTESTINAL ABSORPTION, FOOD, AND EXCIPIENTS

The small intestine is the main organ of digestion with specialized sites of nutrient uptake and drug absorption. The small intestine is a convoluted tube about 5 m in length with three regions: duodenum, jejunum, and ileum. During the fed phase the contractions serve to mix chyme with enzymes and digestive secretions, circulate the contents to

facilitate contact with the intestinal mucosa, and finally propel the contents towards the large bowel. The major areas in which digestion occurs are the jejunum and ileum, and digestion of proteins and fats is largely complete as the chyme enters the caecum. Small intestinal transit in man has been measured by a wide variety of techniques and is generally accepted to be around 4 h.

Components of food can alter drug absorption by affecting drug solubility. Recent findings suggest that drinks such as grapefruit juice can increase the bioavailability of certain drugs, by reducing presystemic intestinal metabolism (23). The solubility of the drug in the GI tract can also be enhanced when the drug is highly soluble in a coadministered food component. The solubility of dicumarol is five times higher in defatted milk than in buffer at 37°C. This increase in solubility and bioavailability is due to the main milk protein casein. Griseofulvin absorption is also enhanced by concomitant food intake, especially after ingestion of heavy fat meals. This enhanced bioavailability is probably due to the increased bile output and the prolonged gastric emptying.

Excipients such as mannitol can affect small intestinal transit, which in turn can affect the absorption of certain drugs. Oral solutions are rarely likely to fall short of bioequivalence relative to solid oral formulations, although during the development of a ranitidine effervescent oral solution dosage form containing sodium acid pyrophosphate (SAPP), a marked decrease in absorption was observed in the extent of ranitidine absorption from the liquid formulation relative to the conventional oral tablet. The formulation contained 150 mg ranitidine with 1132 mg SAPP together with 1.5 MBq ¹¹¹indium chloride solutions. Small intestinal transit time was decreased to 56% in the presence of the excipient. The rapid small intestinal transit associated with an excipient of a solution dosage form resulted in a decreased extent of ranitidine absorption (24).

Intestinal transit rate is highly dependent on the motility state of the GI tract either fasted or fed partly due to the higher viscosity of chyme in the fed state. Blair and co-workers conducted a study in 20 men (energy intake 1272–5342 kcal/day) and found that higher calorific intake was associated with faster transit (25). Exercise in moderation appears to have no effect on transit.

THE COLONIC ENVIRONMENT

By the end of the small intestine, deposition is almost complete and there is no need for intestinal secretions to aid assimilation. The principal role of the colon is to resorb

water and reclaim sodium; however, complex carbohydrate components of vegetable origin have nutritional value but are relatively resistant to attack from intestinal secretion. In the caecum, a complex bacterial environment digests the soluble, fermentable carbohydrates to yield short-chain fatty acids, which are assimilated into the systemic circulation by the colon, together with vitamin K released from the plant material.

The presence of the complex carbohydrates provides the environment with a viscous hydrogel structure; water removal gradually yields a mass of bacteria bound by undigested carbohydrates (celluloses) to form the stool. The presence of a hydrogel softens the mass and also provides water for dissolution. Carbon dioxide release is also a fermentation product, and if the redox potential is sufficiently low, bacteria can produce methane and hydrogen that can be detected in the breath particularly after the ingestion of pulses. In the upright position, the gas will rise to the transverse colon: It is estimated that an adult produces approximately 2–3 L per day on 20 g fermentable fibre (most of which is eliminated in the breath) (26). The presence of large volumes of gas appears to restrict the availability of water past the hepatic flexure, and the consequences of this must be carefully considered.

Residence of dosage forms is approximately 10–12 h in the proximal colon and 12 h in the descending and rectosigmoid colon regions, giving a whole colon transit time of 24–48 h. The relative residence time in the segments of the large bowel is highly dependent upon feeding pattern. Administration of a midday meal to a previously fasted subject may produce dramatic movement, shifting dosage forms from the ascending to the descending colon. Dietary fibre in the form of bran, whole-meal bread, and fruit and vegetables reduces colonic transit time but increases faecal weight by acting as a substrate for colonic bacterial metabolism. For example, an additional 20 g/day of bran increases faecal weight by 127% and decreases mean transit time from 73 ± 24 to 43 ± 7 h (27). However, the same quantity of fibre given in the form of cabbage, carrot, or apple produced a smaller effect.

EFFECTS OF CHANGES TO THE COLONIC ENVIRONMENT

Administration of osmotic laxatives such as lactulose to healthy subjects increases defaecation frequency, producing a reversible syndrome that mimics irritable bowel syndrome. Administration of 20 ml lactulose three times per day creates a more fluid environment as indicated by increased stool water content (28). To examine the effect

of such an environment on drug delivery we administered pulse release units (Pulsincap) containing quinine dihydrochloride (50 mg) following treatment with lactulose for 3 days (29).

Plasma-concentration-time profiles following lactulose treatment showed a faster T_{\max} , a higher C_{\max} , and an increased AUC indicating significantly improved absorption. These data correlated with a higher dispersion score for the marker indicating that the water content of the colon plays an important part in determining the extent of absorption in man. This will be especially important for pulsed release systems in which the exterior is covered by an impermeable coat. The sluggish mixing and the low availability of water past the ascending colon could compromise efficient release of the drug encapsulated in the system.

EFFECTS OF DISEASE ON GASTROINTESTINAL TRANSIT

Esophageal transit time is adversely affected by various disease states, for example, achalasia (30) scleroderma, oesophageal carcinoma, Barrett's oesophagus, cervical vertebropathy (31), left side heart enlargement, and reflux oesophagitis. Abnormalities in oesophageal function can occur as a result of a variety of disease such as chronic alcoholism and diabetes mellitus (32).

Diseases that affect gastric motility and emptying are predominantly diseases of the gastrointestinal tract itself, although diseases such as diabetes that produce neuropathy can also alter transit. Diseases associated with accelerated transit can reduce the bioavailability of drugs delivered in a controlled-release formulation and conversely diseases that slow transit can enhance absorption. Stasis or stagnation at a particular site is always of concern since high local concentrations of drug may result. Inhibition of motility, especially of the stomach, is widely recognized as a consequence of many acute illnesses such as severe pain of any origin, trauma, major infections, and metabolic disturbances.

Some diseases only affect one of the phases of gastric emptying. Generally, duodenal ulcers produce accelerated emptying while gastric ulcers reduce antral motility, producing normal emptying of liquids but delayed emptying of solids. Emptying of a solid meal is slowed in patients with pernicious anaemia and atrophic gastritis but in achlorhydric patients liquids empty rapidly.

Gastro-oesophageal reflux is an extremely common disease affecting between 10 and 20% of the general population, although it is speculated that the incidence is

considerably higher since a large proportion of sufferers self-medicate. The effect of this disease on gastric emptying is unclear since some studies report no effect on the emptying of liquids or mixed meals whilst others demonstrate a delay. It is possible that the emptying of solids in a mixed meal is selectively delayed, suggesting impaired antral motility. This would lead to a greater difference in emptying among liquids, pellets, and tablets than in normal subjects.

There are conflicting reports as to the effect of obesity on gastrointestinal transit. Some studies show no effect, whilst others report delayed emptying of solids particularly in men. This phenomenon is not reversed after significant weight loss. The eating disorders anorexia nervosa produces both delayed solid and liquid emptying (33). Voluntary suppression of defecation also decreases gastric emptying rate (34). The effects of gastrointestinal disease on dosage form performance have recently been reviewed by Milovic and Stein; the reader is referred to this review for further information (35).

Diarrhea and constipation are associated with many illnesses and always produce concern to the effectiveness of therapeutic agents. Like gastro-oesophageal reflux, the aetiology of these conditions is diverse and hence the effect on transit will vary. For example, in a group of constipated patients, a scintigraphic study using a water-soluble marker demonstrated a rapid and diffuse spread through the colon, 5/37 showed normal transit, 26/37 showed that the major site of hold-up was the transverse colon, and the splenic flexure and 6/37 the hold-up was in the descending and rectosigmoid colon (36). Prodrugs such as sulphasalazine, which rely on colonic bacteria to release the active moiety, will be affected by diarrhea (37).

Diseases such as diabetes, which have diverse complications, can produce different changes in gastrointestinal motility in different patients. Delayed oesophageal transit is common, although it does not tend to produce clinically significant problems. However, diabetic gastroparesis is a problem. These patients demonstrate reduced postprandial antral times compared with diabetic-without this complication, although overall transit is not different. There was no correlation among disturbed gastric clearance, impaired gall bladder contraction, and prolonged colonic transit time in the patients with cardiovascular autonomic neuropathy nor was there a correlation between any disturbed motor function and age or duration of diabetes (38).

Patients with cystic fibrosis require pancreatic enzyme supplements to aid food digestion. A recent investigation in which enteric coated enzyme pellets were labelled with ¹¹¹indium and given to patients with a pancake meal

labelled with Tc-99m tin colloid showed that the pellets passed through the intestine ahead of food (39). These data suggest that the incorrect delivery of this enzyme might be responsible for malabsorption and the development of strictures in the proximal colon caused by the high-dose supplements reaching this region before the food.

pH IN HEALTH AND DISEASE

Food affects bulk pH in the stomach by dilution, buffering, and stimulating acid secretion. In the colon, pH is affected by meals containing a high proportion of fermentable fibre. The greatest secretory activity occurs in the stomach within the first hour of eating and the volume of gastric juices produced may be up to twice that of the meal. Studies using a triple pH electrode show that the peak buffering effects of food occurs within 30 min of eating with recorded variations in the stomach varying from pH 2 to around 4.5 (40). Even small volumes of water taken with medications can cause a temporary rise in pH due to neutralization effects on the residual acid.

Content of the meal also affects gastric pH. For example, a pure carbohydrate meal has no detectable effect on acidity (41), whereas a high protein meal matched for calorific content confirming that it is not the protein but its digestion products, the peptides, and amino acids, which are the potent stimulators of acid secretion. A liquid meal, rather than a mixed phase meal, with a balance of carbohydrate and protein has a strong buffering effect but the pH falls rapidly as the liquid is emptied.

Some research workers have found that the basal gastric pH can be surprisingly high. Spontaneous achlorhydria has been observed in healthy subjects during which they have temporary periods of profound reduction in acid secretion and may have complete anacidity (42, 43).

The pH in the small intestine is less variable than the stomach or colon and is in the region of 5.0–6.5, rising slowly along its length. The mean pH in the right colon is 6.4 rising to 6.6 and 7.0 in the transverse, and descending colon, respectively. The presence of fermentable dietary fibre produces a reduction in both viscosity and pH in the colon (44). Recent studies in Japan suggest that patients with Crohn's disease, whether active or in remission, have wider fluctuations in colonic pH, and values in the right and left colon are much lower than would be expected on the basis of control subjects (45).

A circadian rhythm of basal gastric acidity is known to occur with acid output being highest in the evening and lowest in the morning. The daytime patterns of gastric pH vary greatly between individuals, in part due to

the differences in the composition of meals, the physiological responses of acid secretion and gastric emptying provoked postprandially; however, it has been found that the nocturnal patterns of gastric acidity are very similar with very low pH between midnight and early morning (41). The later in the day the evening meal is taken, the later the nocturnal peak of acidity occurs (46). It is, therefore, important to standardize the time for the evening meal when comparing the nocturnal effects of antisecretory drugs.

pH AND GENDER

There is a sex-related difference in human gastric acid secretion. Healthy women secrete significantly less basal and pentagastrin-stimulated acid than men with a median 24 h integrated acidity of 485 mmol h^{-1} versus 842 mmol h^{-1} . In a sample of 365 healthy subjects, the average basal pH was 2.16 ± 0.09 for men and 2.79 ± 0.18 in women (47).

pH AND AGE

It has always been assumed that gastric acid secretion decreased with age, however, a recent study has demonstrated that a group of healthy subjects with a mean age of 51 years (range 44–71 years) had a higher basal acid production than a group with a mean age of 33 years (range 23–42 years) (48). The age, related increase in secretion was greater in men than women but it was not correlated with height, weight, body surface area, or fat-free body mass or by the increased incidence of *Helicobacter pylori* infection.

Very few data are available on gastrointestinal pH in children. In 12 healthy subjects aged 8–14 years, the mean gastric pH was 1.5 and duodenal pH was 6.4, but this gradually rose in the small intestine reaching a peak value of 7.4 in the distal ileum (49). The pH dropped to 5.9 as the pH radiotelemetry capsule entered the caecum but increased to 6.5 in the rectum. The median gastric residence time of the telemetry capsule was 1.1 h, small intestinal transit time was 7.5 h, and colonic transit time was 17.2 h.

CONCLUDING REMARKS

It is clear that numerous physiological factors and disease states alter the gastrointestinal environment and affect

the transit and absorption of a formulation through the gastrointestinal tract. An appreciation of these physiological factors is essential for the optimization of clinical studies. However, high individual variability will remain the major challenge in clinical trial design.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the permission to use data generated in clinical trials in association with industry and clinical services. In particular, we thank Dr. M Frier and Professor A.C. Perkins (Queen's Medical Centre Nottingham), Dr. Tim Grattan (SB Healthcare), Dr. Richard Dansereau (Proctor & Gamble Pharmaceuticals Inc.), and Dr. J. Foster (Western Infirmary, Glasgow).

REFERENCES

1. Hey, H.; Jorgensen, F.; Sorensen, K.; Hasselbalch, H.; Wamberg, T. Esophageal Transit of Six Commonly Used Tablets and Capsules. *Br. Med. J.* **1982**, *285*, 1717–1719.
2. Pemberton, J. Esophageal Obstruction and Ulceration Caused by Oral Potassium Therapy. *Br. Heart J.* **1970**, *32*, 267–268.
3. Eng, J.; Sabanathan, S. Drug-Induced Esophagitis. *Am. J. Gastroenterol.* **1991**, *86*, 1127–1133.
4. Smith, S.J.; Lee, A.J.; Maddix, D.S.; Chow, A.W. Pill-Induced Esophagitis Caused by Oral Rifampin. *Ann. Pharmacother.* **1999**, *33* (1), 27–31.
5. Perkins, A.C.; Wilson, C.G.; Blackshaw, P.E.; Vincent, R.M.; Dansereau, R.J.; Juhlin, K.D.; Bekker, P.J.; Spiller, R. Impaired Esophageal Transit of Capsule Versus Tablet Formulation in the Elderly. *Gut.* **1994**, *35*, 1363–1367.
6. Perkins, A.C.; Wilson, C.G.; Frier, M.; Vincent, R.M.; Blackshaw, P.E.; Dansereau, R.J.; Juhlin, K.D.; Bekker, P.J.; Spiller, R.C. Esophageal Transit of Risedronate Cellulose-Coated Tablet and Gelatin Capsule Formulations. *Int. J. Pharm.* **1999**, *186*, 169–175.
7. O'Reilly, S.; Hardy, J.G.; Wilson, C.G. The Influence of Food on the Gastric Emptying of Multiparticulate Dosage Forms. *Int. J. Pharm.* **1987**, *34*, 213–216.
8. Ghoo, R.H.; Moore, J.G.; Greenberg, E.; Alazraki, N.P. Circadian Variation in Gastric Emptying of Meals in Humans. *Gastroenterology* **1987**, *93*, 515–518.
9. Brown, B.P.; Gross, J.; Schulzedelrieu, K.; Carmichael, D.C.; Barloon, T.J.; Abuyousef, M.M. Layering of Fat and Intermittent Cessation of Antral Contractions Delay Liquid Fat Emptying in the Human Stomach. *Gastroenterology* **1994**, *106* (4 SS), A472.
10. Nakae, Y.; Onouchi, H.; Kagaya, M.; Kondo, T. Effects of Aging and Gastric Lipolysis on Gastric Emptying of Lipid in Liquid Meal. *J. Gastroenterol.* **1999**, *34* (4), 445–449.
11. Edelbroek, M.; Horowitz, M.; Maddox, A.; Bellen, J. Gastric Emptying and Intra-gastric Distribution of Oil in

- the Presence of a Liquid or a Solid Meal. *J. Nuc. Med.* **1992**, *33* (7), 1283–1290.
12. Renwick, A.G.; Ahsan, C.H.; Challenor, V.F.; Daniels, R.; Macklin, B.S.; Waller, D.G.; George, C. The Influence of Posture on the Pharmacokinetics of Orally Administered Nifedipine. *Br. J. Clin. Pharmacol.* **1992**, *34* (4), 332–336.
 13. Datz, F.L. Considerations for Accurately Measuring Gastric-Emptying. *J. Nuc. Med.* **1991**, *32* (5), 881–884.
 14. Hutson, W.R.; Roehrkasse, R.L.; Wald, A. Influence of Gender and Menopause on Gastric-Emptying and Motility. *Gastroenterol.* **1989**, *96* (1), 11–17.
 15. Magides, A.D.; Macfie, A.G.; Richmond, M.N.; Reilly, C.S. Gastric-Emptying in Pregnancy. *Brit. J. Anaesth.* **1990**, *65* (4), 580P–581P.
 16. Wald, A.; Van Thiel, D.H.; Hoechstetter, L.; Gavalier, J.S.; Egler, K.M.; Verm, R.; Scott, L.; Lester, R. Gastrointestinal Transit: The Effect of the Menstrual Cycle. *Gastroenterol.* **1980**, *80*, 1497–1500.
 17. Miller, G.; Palmer, K.R.; Smith, B.; Ferrington, C.; Merrick, M.V. Smoking Delays Gastric Emptying of Solids. *Gut* **1989**, *30* (1), 50–53.
 18. Scott, A.M.; Kellow, J.E.; Eckersley, G.M.; Nolan, J.M.; Jones, M.P. Cigarette Smoking and Nicotine Delay Postprandial Mouth-Cecum Transit Time. *Dig. Dis. Sci.* **1992**, *37*, 1544–1547.
 19. Scott, A.M.; Kellow, J.E.; Shuter, B.; Nolan, J.M.; Hoschl, R.; Jones, M.P. Effects of Cigarette Smoking on Solid and Liquid Intragastic Distribution and Gastric Emptying. *Gastroenterol.* **1993**, *104*, 410–416.
 20. Pfeiffer, A.; Hohl, B.; Kaess, H. Effect of Ethanol and Commonly Ingested Alcoholic Beverages on Gastric Emptying and Gastrointestinal Transit. *Clin. Investig.* **1992**, *70* (6), 487–491.
 21. Wegener, M.; Schaffstein, J.; Dilger, U.; Coenen, C.; Wedmann, B.; Schmidt, G. Gastrointestinal Transit of Solid Liquid Meal in Chronic-Alcoholics. *Dig. Dis. Sci.* **1991**, *36* (7), 917–923.
 22. Johnson, R.D.; Horowitz, M.; Maddox, A.F.; Wishart, J.M.; Shearman, D.J.C. Cigarette Smoking and Rate of Gastric Emptying: Effect on Alcohol Absorption. *Br. Med. J.* **1991**, *302* (67), 20–23.
 23. Evans, A.M. Influence of Dietary Components on the Gastrointestinal Metabolism and Transport of Drugs. *Ther. Drug Monitoring* **2000**, *22* (1), 131–136.
 24. Koch, K.M.; Parr, A.F.; Tomlinson, J.J.; Sandefer, E.P.; Digenis, G.A.; Donn, K.H.; Powell, J.R. Effect of Sodium Acid Pyrophosphate on Ranitidine Bioavailability and Gastrointestinal Transit Time. *Pharm. Res.* **1993**, *10* (7), 1027–1030.
 25. Blair, E.H.; Wing, R.R.; Wald, A. Rapid Orocecal Transit in Chronically Active Persons with High Energy Intake. *J. App. Physiol.* **1991**, *70* (4), 1550–1553.
 26. Campbell, J.M.; Fahey, G.C. Psyllium and Methylcellulose Fermentation Properties in Relation to Insoluble and Soluble Fiber Standards. *Nutr. Res.* **1997**, *17* (4), 619–629.
 27. Cummings, J.H.; Branch, W.; Jenkins, D.J.A.; Southgate, D.A.T.; Houston, H.; James, W.P.T. Colonic Response to Dietary Fibre from Carrot, Cabbage, Apple, Bran and Guar Gum. *Lancet* **1978**, *i*, 5–9.
 28. Barrow, L.; Spiller, R.C.; Wilson, C.G. Pathological Influences on Colonic Motility—Implications for Drug Delivery. *Adv. Drug Delivery Rev.* **1991**, *7* (1), 201–218.
 29. Hebden, J.M.; Gilchrist, P.J.; Perkins, A.C.; Wilson, C.G.; Spiller, R.C. Stool Water Content and Colonic Drug Absorption: Contrasting Effects of Lactulose and Codiene. *Pharm. Res.* **1999**, *16* (8), 1254–1259.
 30. Mearin, F.; Papo, M.; Malagelada, J.R. Impaired Gastric Relaxation in Patients With Achalasia. *Gut* **1995**, *36* (3), 363–368.
 31. Hep, A.; Vanaskova, E.; Tosnerova, V.; Prasek, J.; Vizda, J.; Dite, P.; Ondrousek, L.; Dolina, J. Radionuclide Esophageal Transit Scintigraphy—A Useful Method for Verification of Esophageal Dysmotility by Cervical Vertebroplasty. *Dis. Esophagus* **1999**, *12* (1), 47–50.
 32. Holloway, R.H.; Tippet, M.D.; Horowitz, M.; Maddox, A.F.; Moten, J.; Russo, A. Relationship Between Esophageal Motility and Transit in Patients with Type I Diabetes Mellitus. *Am. J. Gastroenterol.* **1999**, *94* (11), 3150–3157.
 33. Hutson, W.R.; Wald, A. Gastric-Emptying in Patients with Bulimia Nervosa and Anorexia-Nervosa. *Am. J. Gastroenterol.* **1990**, *85* (1), 41–46.
 34. Tjeerdsmas, H.C.; Smout, A.; Akkermans, L. Voluntary Suppression of Defecation Delays Gastric-Emptying. *Dig. Dis. Sci.* **1993**, *38* (5), 832–836.
 35. Milovic, V.; Stein, J. Gastrointestinal Disease and Osage Form Performance. *Oral Drug Absorption: Prediction and Assessment*; Dressman, J.B., Lennernas, H., Eds.; Marcel Dekker, Inc.: New York, 2000; 17–30.
 36. Roberts, J.P.; Newell, M.S.; Deeks, J.J.; Waldron, D.W.; Garvie, N.W.; Williams, N.S. Oral [in-111] Dtpa Scintigraphic Assessment of Colonic Transit in Constipated Subjects. *Dig. Dis. Sci.* **1993**, *38* (6), 1032–1039.
 37. Rijk, M.C.M.; Vanhogezaand, R.A.; Vanschaik, A.; Vantongerren, J.H.M. Disposition of 5-Aminosalicylic Acid from 5-Aminosalicylic Acid-Delivering Drugs During Accelerated Intestinal Transit in Healthy Volunteers. *Scand. J. Gastroenterol.* **1989**, *24* (10), 1179–1185.
 38. Werth, B.; MeyerWyss, B.; Spinass, G.A.; Drewe, J.; Beglinger, C. Non-Invasive Assessment of Gastrointestinal Motility Disorders in Diabetic Patients with and without Cardiovascular Signs of Autonomic Neuropathy. *Gut* **1992**, *33* (9), 1199–1203.
 39. Hillel, P.G.; Tindale, W.B.; Taylor, C.J.; Frier, M.; Senior, S.; Ghosal, S. The Use of Dual-Isotope Imaging to Compare the Gastrointestinal Transit of Food and Pancreatic Enzyme Pellets in Cystic Fibrosis Patients. *Nuc. Med. Communi.* **1998**, *19* (8), 761–769.
 40. Washington, N.; Washington, C.; Wilson, C.G. *Physiological Pharmaceutics*; Rubenstein, M., Wilson, C.G., Eds.; Taylor Francis Series in Pharmaceutical Technology, Taylor & Francis: London, 2001.
 41. Bumm, R.; Blum, A.L. Lessons from Prolonged Gastric pH Monitoring. *Aliment. Pharmacol. Therap.* **1987**, *1*, 518S–526S.
 42. Desai, H.G.; Zaveri, M.P.; Anita, F.P. Spontaneous and Persisting Decrease in Maximal Acid Output. *Br. Med. J.* **1971**, *ii*, 313–315.

43. Waterfall, W.E. Spontaneous Decrease in Gastric Secretory Response to Humoral Stimuli. *Br. Med. J.* **1969**, *4*, 459–461.
44. Tomlin, J.; Read, N.W. Comparison of the Effects on Colonic Function Caused by Feeding Rice Bran and Wheat Bran. *Eur. J. Clin. Nutr.* **1988**, *42* (10), 857–861.
45. Sasaki, Y.; Hada, R.; Nakajima, H.; Fukuda, S.; Munakata, A. Improved Localizing Method of Radiopill in Measurement of Entire Gastrointestinal pH Profiles: Colonic Luminal pH in Normal Subjects and Patients with Crohn's Disease. *Am. J. Gastroenterol.* **1997**, *92* (1), 114–118.
46. Lanzon-Miller, S.; Pounder, R.E.; McIsaac, R.I.; Wood, J.R. Does the Timing of the Evening meal Affect the Pattern of 24 Hour Intragastric Acidity. *Gut* **1988**, *29*, A1472.
47. Feldman, M.; Barnett, C. Fasting Gastric pH and Its Relationship to True Hypochlorhydria in Humans. *Dig. Dis. Sci.* **1991**, *36*, 866–869.
48. Goldschmiedt, M.; Barnett, C.C.; Schwarz, B.E. Effect of Age on Gastric Acid Secretion and Serum Gastric Concentrations in Healthy Men and Women. *Gastroenterology* **1991**, *101*, 977–990.
49. Dantas, R.O.; Dodds, W.J. Measurement of Gastrointestinal pH and Regional Transit Times in Normal Children. *J. Pediatr. Gastroenterol. Nutr.* **1990**, *11* (2), 211–214.

Pilot Plant Design

Mickey L. Wells
Samuel B. Balik
Ralph B. Caricofe
Charles W. Crew
Ronald A. Sanftleben
A. Wayne Wood

GlaxoSmithKline, Inc., Research Triangle Park, North Carolina, U.S.A.

INTRODUCTION

A pilot plant design should support three key strategic objectives:

- Formulation and process development.
- Clinical supply manufacture.
- Technology evaluation, scale-up, and transfer.

Integration of these objectives serves to maximize product development, speed, and efficiency. The pilot plant is a “hybrid” clinical manufacturing and development facility, which integrates early development activities, clinical supply manufacture, technology evaluation, scale-up, and transfer to production sites. The manufacture of clinical supplies is an integral part of the development process. As the clinical program requirements increase, batch size increases, and critical formulation and process performance information can be obtained during manufacturing scale-up. This information and experience contribute to smooth and efficient production transfer.

Specific design layout depends on a number of factors. The design may be constrained if the construction entails renovation or conversion of an existing building compared to designing a pilot plant on a new site. The various technologies and systems, such as fluid-bed granulators, microwave processors, lyophilizers, water systems, and dust collectors, have specific design requirements and constraints that require special consideration. Location of a new facility must be evaluated early in the program. This evaluation is essential in determining site–location factors, such as proximity to development, production, clinical packaging, and warehouse, as well as support staff considerations, such as quality assurance (QA), quality control (QC), validation, maintenance, calibration, microbiology, and other functions.

The final design should result in a facility that supports the key strategic objectives, is functional, and has

low maintenance and operating costs. The interaction between the design firm, corporate engineering, and the client is central to an operational design and end user satisfaction. In addition to the architects and engineers, the core multidisciplinary team should include personnel from QA, QC, validation, maintenance, calibration, pilot plant operations, formulation development, information technology, purchasing, environmental health, and safety. Security, telecommunications, purchasing, and the insurance carrier should assign members to the project who can serve as resources for specific needs.

A design firm should be chosen according to the range of services offered and experience in designing good manufacturing practices (GMPs) facilities. Selecting a firm that offers architectural, engineering, and construction consulting, as well as life-safety services is recommended. A pilot plant is a “hybrid” clinical manufacturing and development facility, and overdesigning for waste treatment, safety, and fire codes is highly likely if the proper experience and interpretation of codes and operational aspects are lacking. Visiting facilities the firm has designed is useful to examine the quality of past designs and to find out how well the firm fulfilled its obligations and dealt with problems. Including key design firm members in the contract is highly recommended. Experienced design engineers and a staff with a proven track record are considered to be most critical.

Early involvement of the Food and Drug Administration (FDA) in reviewing the facility design is advised. This allows the most current FDA interpretation of GMPs to be incorporated into the facility design. As the design is being finalized, it should be presented to the local, district, and Washington FDA offices for review and critique. Discussions concerning material flow, personnel flow, building surfaces, air classification, and the validation master plan may be a useful means of presenting the design and opening a dialogue with the FDA. Extending an invitation to the local and district FDA personnel to

visit the facility during various phases of construction is also recommended.

STRATEGIC OBJECTIVES

Formulation and Process Development

Strategic decisions regarding the dosage forms to be developed in the pilot plant and the technologies employed significantly affect the design.^[1] These strategic decisions are often based on current production technology, dosage form development, and assessment of potential delivery system requirements to support future line extensions and the pipeline of new chemical entities (NCEs). An assessment of the proportional utilization of the various technologies and anticipated processes is important in deciding how much space to allocate to each in the layout.^[2] During the transition from the development scale to pilot scale to full-scale manufacturing, identification and control of critical component, formulation, and process variables are essential for final production. The facility and equipment should be able to capture critical process information. The more sophisticated the technology, the greater the need to control and monitor critical parameters. It may also be desirable to monitor and control these parameters from remote locations.

Clinical Supply Manufacture

To manufacture clinical supplies, GMP-compliance considerations must be the basis of the design features of the pilot plant.^[3] Integrated validation plans for the equipment and facility should be available before the basis of design is finalized. These design features must meet safety and environmental facility requirements to handle compounds for which only marginal information is available regarding potential toxicity, potency, and physical–chemical properties. Highly potent compounds need to be handled and manufactured in an isolated and contained environment requiring special design controls.^[4]

Clinical supply manufacture should be viewed as an integral part of the product development process. As the program requirements grow, the increased clinical trial demand needs to be met by increased manufacturing capacity. Critical formulation and process performance information is gained during the scale-up of clinical batches. Often there is a shortage of drug substance, and the first opportunity for the formulation scientist to scale-up a product is when clinical supply requirements justify the need.

Pharmaco-economic studies that evaluate quality of life and other medical outcome criteria enroll thousands of patients and require millions of dosage units. These larger batch requirements can be manufactured in production or can be outsourced to a contract resource organization (CRO). In either case, a certain degree of control and flexibility is given up, which may slow clinical programs. Having direct access and control to a clinical manufacturing facility with large scale capability allows flexibility and short turnaround times to meet clinical demands. During facility design, strategic decisions regarding manufacturing capacity based on clinical needs significantly impact space and equipment capacity.^[1] Table 1 shows batch sizes and production capacity based on typical unit weights that could be used in determining design objectives.

Technology Evaluation, Scale-Up, and Transfer

Flexibility in facility design can better accommodate new technology evaluation. The technology may be new to the company or novel to the industry and offers advantages that current production and pilot plant facilities lack. New technologies may increase production speed, and provide continuous processing, maintenance, scale-up, cleaning, and containment improvement. Novel technologies may include evaluation of the manufacture of novel delivery systems or processes. In any case, a flexible pilot plant design (utilities and space) gives the development scientist

Table 1 Maximum batch manufacturing capabilities based on typical unit weights

Technology	Batch capacity (kg)	Units per batch
Solids, 500 mg/unit	200	400,000
	20	40,000
	5	10,000
Semisolids, 50 g/unit	500	10,000
	100	2,000
	10	200
Liquids, 100 mL/unit	1000	10,000
	100	2,000
	10	100
Parenteral, 10 mL/unit	250	25,000
	50	5,000
	10	1,000
Aerosol, 20 g/unit	250	25,000
	100	5,000
	10	500

and technical operations personnel the opportunity to collaborate in technology evaluation.^[5]

Critical formulation and process performance information can also be obtained in the pilot plant.^[6] As the clinical supply quantity requirements increase and larger batches are needed, operations can be scaled into larger equipment of the same or similar operating principles. This makes it possible to identify critical process variables and ranges and to study the effect of scale. Fig. 1 shows fluid-bed granulator–dryers, and Fig. 2 shows coating pans of different sizes. The closer the pilot plant equipment resembles production equipment with respect to design and scale, the fewer trials, often with costly drug substance, are necessary when transferring the process to production. Large-scale equipment in the pilot plant should be at least 10% of the capacity used in production, but ideally should approach 25–100% of the capacity of full-scale equipment.

Typically, a combination of engineering models, statistically designed experiments, and regression models can be used to define the manufacturing process with

respect to process parameters and ranges to ensure adequate control of the full scale process.^[7] The scaling experience improves the understanding of the product and process scale parameters. This knowledge contributes to a smooth and expedient production transfer, providing a strong scientific rationale, which leads to credible new drug applications (NDAs) and FDA pre-approval inspections (PAIs).

DESIGN AND GMP-COMPLIANCE CONSIDERATIONS

Layout

In designing the facility and its layout, the impact on GMPs must be kept in mind. The layout should address the need for flexibility, restricted access, personnel flow, and material flow.

Product development starts with many unknowns that eventually, through well-designed development programs,



Fig. 1 Fluid-bed granulator driers of 1, 5, and 50 kg capacity, which provide flexible clinical supplies manufacturing capacity and can be used to understand critical process variables and scaling parameters.



Fig. 2 Coating pans of 1, 15, and 150 kg capacity.

result in a value-added product. Because of these unknowns, it is difficult to predict exactly what kind of physical facilities will be required. Thus, there is a need to incorporate as much flexibility into the design as possible. The desired flexibility can be achieved in a number of ways. For example, fixed process equipment is kept to a minimum, while portable equipment is installed whenever possible. Movement of portable equipment is made easier by wide corridors and 10 ft (3 m) door heights. In addition, the process rooms can be designed for multipurpose use, which adds flexibility. The concept of multipurpose process rooms determines the degree of flexibility desired for utilities, such as the range of power supplies and plant systems including dust collection, vacuum cleaning, water, static grounding, and compressed air.

Restricted access and personnel flow affect the layout. The size of the facility and anticipated nominal daily personnel level determine the size of restroom, locker, gowning, and shower areas. The types of garments worn by personnel in the pilot plant should be specified and provisions made for their cleaning, disposal, and storage. Gowning and degowning procedures should be investigated and formalized.

Material receipt, sampling, and quarantine processes should be well planned. The material and components used in manufacturing clinical supplies should enter the facility through a controlled access area. Sampling and quarantine procedures affect the facility layout and design. Special precautions for sampling potent (highly active) compounds should be considered. This may require the use

of specially designed rooms that include air locks, gowning areas, and isolation sampling methods. Delivery and shipping personnel should have access restricted to the receiving and dock area. A receiving office should be located immediately adjacent to the actual receiving area. A restroom, telephone, and seating area for the transportation personnel are other design options. The type of lift trucks used requires certain clearance heights and widths. Battery-operated lift trucks require space for charging the equipment, and ventilation requirements must be evaluated. Receipt, sampling, and quarantine storage of flammable solvents require special consideration such as location, explosion-proof lighting, and electrical connections. Solvent-dispensing practices and spill containment procedures are important parameters. Space is required for the temporary holding of flammable waste products as well as other waste.

The warehouse layout depends on the manufacturing scale, type of products, and storage requirements. Temperature and relative humidity set points and ranges should be defined, as well as special storage conditions, which require refrigerators, freezers, and controlled access. Storage capacity and space constraints may dictate high-density shelving and determine pallet-stacking height. Storage within the facility is best, but off-site storage for medium-to long-term supplies or large quantity purchases may be a more cost-effective alternative. The method of inventory identification, status, and the selection of lift truck size determine the design and capacity of the receiving, shipping, and warehouse areas. Systems that meet GMP iden-

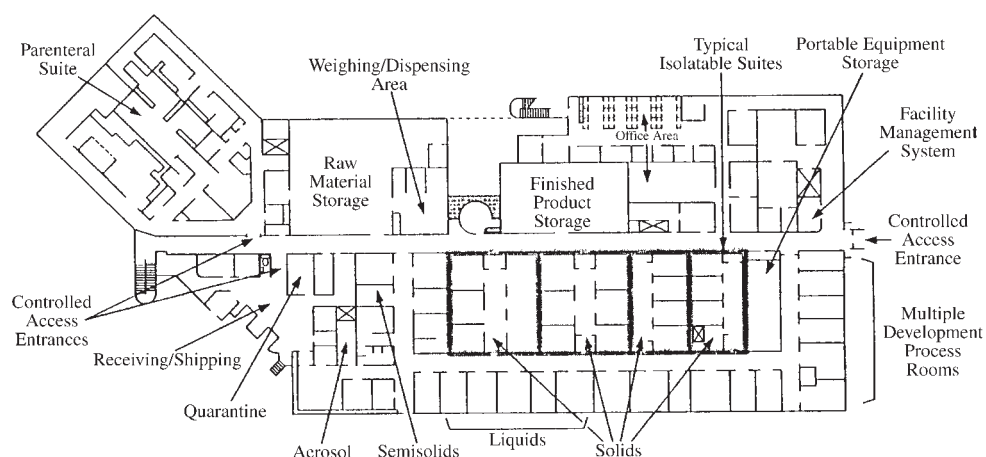


Fig. 3 A pilot plant layout.

tification and status requirements without requiring storage of material of the same status in the same location maximize space utilization.

Fig. 3 shows a pilot plant layout with restricted personnel access. This design would support the development and clinical manufacture of solid dosage forms, liquids, semisolids, aerosols, and sterile products. Multipurpose rooms are incorporated in each area to maximize the use of portable equipment, and scale factors similar to those shown in Table 1 are employed. Isolation suites are indicated in the manufacturing area; their purpose and design are discussed in more detail later. The sterile area is isolated from the main corridor by the interior corridor design. At the far left, the main corridor provides access for future facility expansion, if necessary.

Construction

As a multipurpose, multiproduct facility, the walls, ceilings, and floors need to be constructed of readily

cleanable surfaces. Radial joints at the wall, ceiling, and floor junction facilitate cleaning and wash-down. Cabinetry should be of nonporous construction and have slanted tops to minimize dust collection and aid in cleaning.

Table 2 lists a number of available, acceptable, wall and ceiling material compositions and their relative attributes of purchase and installation costs, cleanability, durability, resilience, and ease of repair. The connections between the wall and window frames, doorframes, and the floor need to be carefully constructed to prevent moisture damage and enhance cleanability. Dry-wall fiberglass mats painted with epoxy provide durable, cleanable surfaces at a lower cost than many other options.

Floor selection criteria should include chemical resistance, stain resistance, scuff and marring resistance, ability to radius cove, durability, esthetics, and ease of maintenance and repair. Table 3 lists several floor compositions and selection criteria. Special areas may require static dissipative or conductive flooring in order to minimize explosion potential with flammable material.

Table 2 Ceiling and Wall Material Compositions^a

Material	Cost of Purchase and Installation	Cleanability	Durability	Resiliency	Ease of Repair
Dry-wall with epoxy paint	□	□	■	■	□
Dry-wall, fiberglass, and epoxy	□	□	▣	▣	▣
Stelvatite	▣	▣	□	▣	■
Vinyl (Kydex)	■	□	□	□	■
Masonry and epoxy	□	■	□	■	▣
Masonry, plaster, and epoxy	▣	□	□	■	▣

^a Key: □ better; ▣ intermediate; ■ worse

Table 3 Floor Composition^a

Material	Cost of Purchase and Installation	Cleanability	Durability	Resiliency	Ease of Repair	Stain Resistance	Chemical Resistance
Concrete (sealed)	□	□	▣	■	□	▣	▣
Epoxy terazzo	■	□	□	■	■	□	□
Quarry tile	■	▣	▣	■	■	□	▣
Concrete, epoxy coated	▣	▣	▣	▣	▣	■	■
Sheet vinyl	▣	■	■	□	▣	■	▣

^a Key: □ better; ▣ intermediate; ■ worse

Metal or reinforced fiberglass doors and doorframes are most desirable. Doorframes that terminate at the top of the cove base allow easy cleaning. Top locking of the inactive door is preferred for dual-panel doors because this eliminates the need for floor penetrations and cleaning concerns. The door closures should be readily cleanable.

Light fixtures should be washable and allow for easy maintenance. Frames are often sealed at the ceiling surface to prevent moisture intrusion and microbiologic growth. Lighting levels in the process areas should be specified in the design. Switches in the hallway or automatic switches facilitate cleaning operations and eliminate the need for expensive explosion-proof devices.

The need for floor drains should be considered, as well as the ability to collect wastewater for treatment or incineration. If floor drains are installed, the floors should slope toward the drain from all directions. Drain cover materials should also be evaluated with regard to quality, ease of sanitation, and cleaning. Fig. 4 shows a multi-purpose room with an easily cleaned floor, dry-wall fiberglass mat with epoxy-painted walls, floor drain, and utility panel. This room also has self-contained hot- and cold-water wash-down capability.

Building Systems and Utilities

The following building systems and utilities need to be addressed at the design phase: HVAC (heating, ventilation, and air conditioning), pharmaceutical water systems, compressed gases (nitrogen, air, respiratory air), electrical outlets with various voltages, dust collection, vacuum cleaning, communication systems, emergency back-up power generation, environmental monitoring, and control of temperature, relative humidity, and differential pressure. The selection and extent to which these systems and utilities are included in the pilot plant design depend on the pharmaceutical product produced, e.g., oral solid vs. sterile dosage forms, and the technologies chosen.^[8,9]

The pilot plant HVAC system must be suitable to consistently produce a quality product and flexible enough to accommodate different process areas and dosage forms. Although not intended for universal application, typical HVAC requirements have been developed for the process areas in a pilot plant.^[10] Different process areas might include manufacturing areas such as dry, liquid, and sterile processing, general support areas of pharmacy, in-process and final storage, and general laboratories including QC and inspection. Other HVAC design considerations might include noise level in occupied spaces, determination of allowable drift before recalibration of sensors and controllers, level of acceptable airborne contamination, and maintenance.

If the most critical utilities used in the pharmaceutical pilot plant are those that come in contact with the product, then water is most likely at the top of the list. Water is commonly used during processing of the product and also for cleaning of equipment and process areas. Water quality can be divided into two subcategories. The first is compendial, of which there are two types, USP purified water and USP water for injection (WFI). There are many different approaches for designing compendial water systems. These systems are, in general, among the most difficult building utilities to commission and maintain. Critical assessment of the volume demand and number of use points define the system size. Smaller systems are easier to commission and maintain and minimize the number of use points. The quality of feed water should be assessed to confirm that the water treatment could handle seasonal variations. Excellent resources are available regarding the design, construction, and operation of compendial water systems, including clean steam.^[11]

The second category of water is noncompendial. Noncompendial water, often referred to as potable water, is generally used for initial rinsing and cleaning of process rooms and equipment. This water should meet the Public Health Services Drinking Water Standards.^[10]

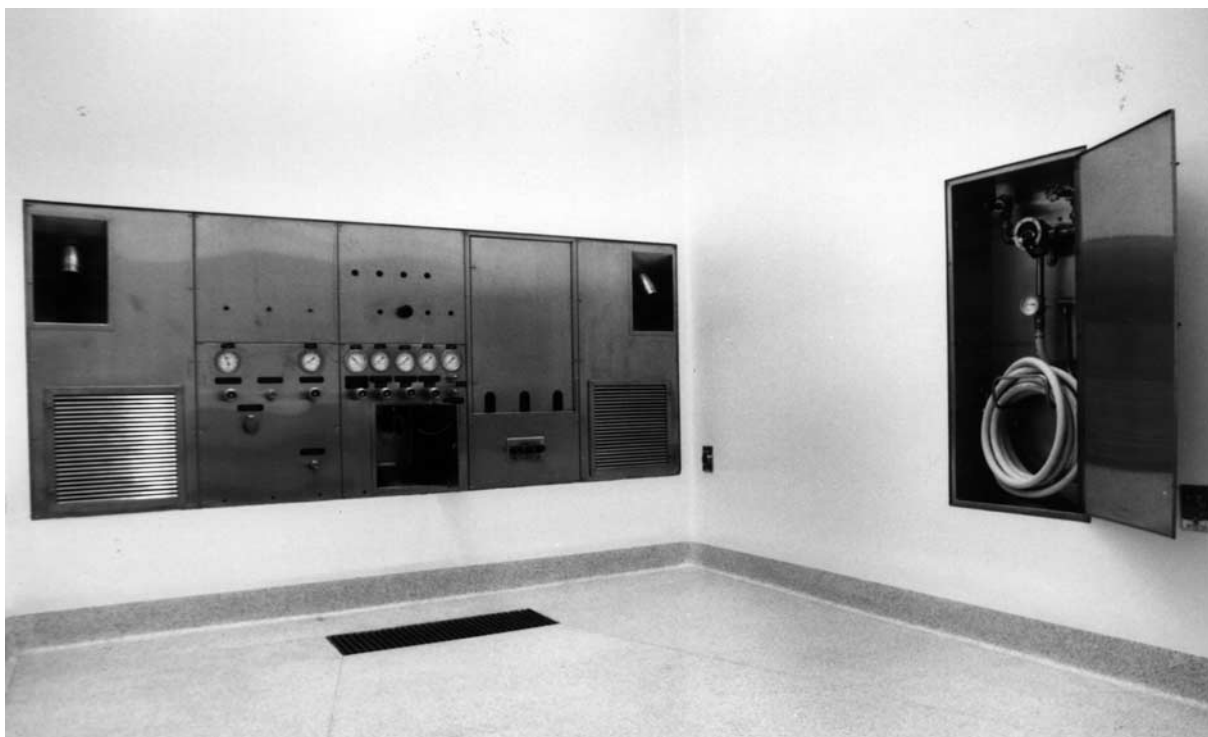


Fig. 4 Multipurpose room with an easily cleanable covered terrazzo floor, dry-wall fiberglass mat with epoxy-painted walls, floor drain, and utility panel. The room has a self-contained hot and cold water wash-down capability.

Noncompensial water is also used in the laboratory environment and is usually referred to by the final treatment step, e.g., reverse osmosis (RO) water and deionized (DI) water.

Provisions should be considered for compressed gases, such as nitrogen, which can be used in the presence of highly explosive dust or solvents as an inert gas blanket or as the energy source in fluid-air micronizers. Compressed air is often used in explosion-proof environments to operate pneumatic switches and as the power source for mixers. Respiratory air may be needed for work with highly active compounds. Compressed gases and respiratory air should be free of oil or other contaminants and meet stringent particulate requirements. Piping for gas utilities may require special consideration during fabrication to control particulate matter such as welding under inert atmosphere. Point of use filters can be used for other process gases such as nitrogen and carbon dioxide.^[10] Respiratory airdrops should be placed at locations that provide easy access to the user. The system can be designed with “quick disconnections” that are keyed and color-coded for each type of compressed gas to prevent inadvertent or inappropriate use.

The service and cleaning of dust-collection and vacuum-cleaning units deserves special attention. Isolated “bag houses” designed with breathing air, decontamination showers, and floor drains with shunt valves that allow

waste water to be pretreated should be considered for dealing with highly active compounds.

A building monitoring and control system can range from simple to complex, depending on the degree of control desired. The simplest level can be used for monitoring and controlling air temperature and relative humidity. Other areas where monitoring and control may be desirable include exhaust duct monitoring for lower explosives limits in solvent processes, differential room pressures, differential pressure across high efficiency particulate air (HEPA) filters, conductivity in USP-WFI water systems, compressed air–nitrogen pressures to detect leaks, oxygen monitoring in areas using nitrogen, chilled process water and clean steam pressure, a synchronized clock system, and environment particulate monitoring in parenteral areas. A system for monitoring and archiving critical environmental and utilities data is shown in Fig. 5. This system allows routine monitoring of critical instrument alarms and provides continuous updates of environmental conditions within and outside the facility. The system identifies the specific process room number and the temperature, percent relative humidity, and differential pressure status. The information can be archived for retrieval during an FDA inspection. Custom report generation allows data arrangement for easy review.



Aseptic Area Partial View

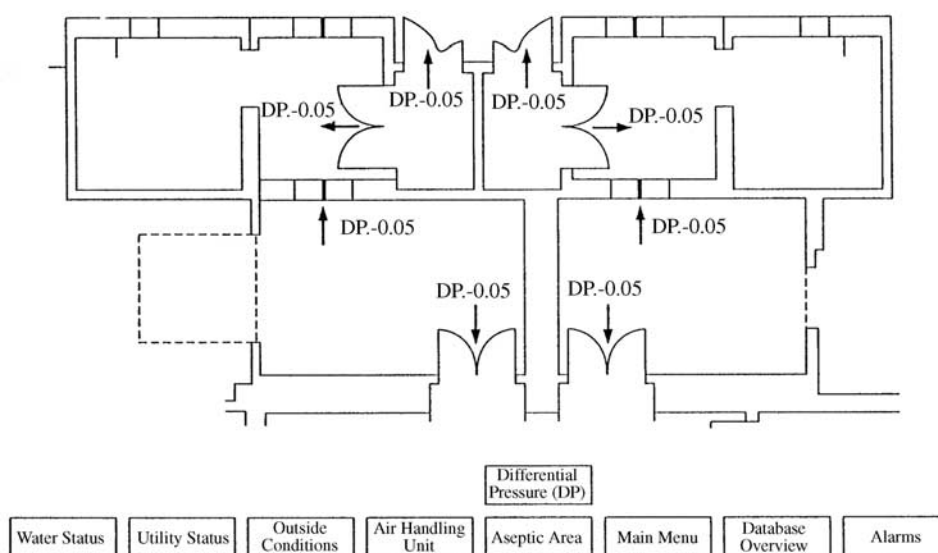


Fig. 5 Facility monitoring and control room (top) and operator interface screen with aseptic area differential pressure readings (bottom). Features that can be monitored via touch screen from the control room include water status, utility status, outside conditions, air-handling units, and process areas in nonsterile and sterile locations.

Service access to system utilities is an important design consideration. Routine calibration, maintenance, and repair are integral parts of maintaining a GMP facility. Areas to store critical spare parts and supplies should be included.

Equipment

Equipment can be viewed as fixed or portable; both types should be easy to clean and maintain. Fixed equipment has

the largest impact on the building design. Special considerations such as explosion proofing, explosion relief, explosion containment, and the need for uninterrupted power supply should be identified early in the design process. Dedicated process air handlers may be required for fluid-bed granulator-dryers and tablet coaters. Clean-in-place (CIP) systems may require special attention. Storage of portable equipment, operating supplies, and spare parts is important. The storage area for portable equipment should be placed strategically to complement the material and personnel flow plans.

SAFETY AND ENVIRONMENTAL CONSIDERATIONS

Emergencies

Emergency communication notification systems, telephones, smoke and gas detectors, and sprinkler systems need to meet state and corporate fire and emergency codes, as well as insurer's requirements.^[12] Specific processes or technologies may require a wet or pre-action sprinkler system. Local code, corporate standards, or site preference determine the specific type chosen. Consideration should be given to the manufacturing processes performed in the area. In a multipurpose facility, the issues should be addressed using a "worst case" scenario. Pre-action (dry piping) systems may be desirable in a sterile product area where accidental leakage could lead to microbiological contamination and area shutdown. System drainage is important and draining procedures should be developed for accidental loss of air pressure within the piping. Alternative methods of fire prevention, such as fire-resistant construction, may be specified to eliminate the need for sprinkler systems.^[13] Windows in the process rooms and hallways can be aligned in such a way as to allow observation of operators from several different vantage points. Safety glass can be used to meet fire codes. Phones and intercom systems can be used to enhance communications between operators, support staff, and management. The selection of communication systems should be based on ease of use and cleanliness.

An emergency notification system should be designed to attract the attention of operators who may be wearing noise protection or breathing-air suits. Visual flashing lights and a loud siren intercom system are simple yet effective means of notifying employees of an emergency situation. Pressurization of emergency exits may be needed when an emergency alarm is actuated.^[12] "Man-down" type of emergency systems may also warrant considerations.

Environmental Discharges

Pilot plant discharges into the environment must comply with local, state, and federal regulations. Of particular concern are facilities that use organic solvents for granulating or coating processes. Atmospheric discharges from equipment such as incinerators, scrubbers, and dust collectors require permits from state authorities and must comply with the federal Clean Air Act.^[14] Permits are required for the generation, treatment, and recycling of hazardous wastes and emissions.^[14] Liquid discharges

into the local water system must meet limits set by the federal Clean Water Act, as well as any other state and local regulations.^[15] Isolation and pretreatment or an alternate means of disposal of wastewater will likely be required if highly potent or toxic compounds are to be processed.

Handling and Prevention of Deflagrations and Explosions

The potential for explosions is an ongoing concern in pharmaceutical operations. Many pharmaceutical dusts, airborne in sufficient concentration, are capable of propagating a flame (deflagration) in the presence of oxygen and an ignition source.^[16] Hybrid mixtures of dust and flammable vapors, such as may be encountered with solvent-based granulating or coating, can create exceptionally explosive conditions. Explosion prevention of solvent storage and dispensing areas requires special attention to ventilation, grounding, and dispensing methods. Static grounding and electrical outlets rated for hazardous environments should be included in all processing areas to eliminate potential ignition sources. In many circumstances, the potential for generating a deflagration is unavoidable, so the pilot plant design must include methods for suppressing, containing, isolating, or venting an explosion in such a way as to protect operators and minimize damage to the equipment and facility. The most commonly used method, explosion venting, incorporates a blowoff panel in the equipment or ductwork to safely vent the explosion to the exterior of the plant. An explosion suppression system should be used if any dust collectors are to be placed within the interior of the building or if toxic materials, which cannot be vented outside the building, are to be used.^a For additional information refer to the National Fire Protection Association (NFPA) Codes 68, 69, 318, 497, and 654.^[17-21]

Handling of Highly Potent or Toxic Compounds

Building design, equipment design, and the use of physical barriers and personal protective equipment (PPE) are the primary methods of maintaining a safe environment, inside and outside the facility, while handling highly potent or toxic compounds. When designing a pilot plant facility to be used for handling highly potent and/or toxic compounds, numerous criteria must be considered:

^a<http://www.fike.com/epvent.htm> (accessed November 2001).

- Unit operations to be performed.
- Range of batch sizes to be processed.
- Capital expenditures and operational costs.
- Anticipated occupational exposure limits (OELs) to be encountered.
- Employee safety and ergonomics.
- Material, equipment, and personnel flow.
- GMP requirements.
- Cross-contamination risks.
- Flexibility of the facility.

Containment of highly potent or toxic compounds is most effectively accomplished by using a physical barrier such as a glove box. A product or process that can be isolated within a physical barrier can significantly reduce the need for air handling capacity, minimize cross-contamination potential, and reduce employee exposure.^[22]

In situations where it is not feasible to depend only on physical barriers, an isolation suite can be designed for working with these materials. Air locks and rooms

that have negative differential pressure to the main hallways are required, as well as HEPA-filtered inlet air and double-HEPA-filtered exhaust air. An uninterrupted power supply should be connected to the exhaust air handler to maintain a negative differential pressure if the suite is in operation during a power outage. If the exhaust air handler fails, all plant air handlers should be programmed to automatically turn off to minimize the potential for a positive differential pressure to develop. Each isolation suite may require a dedicated HVAC system.^[23]

The isolation suites can be designed with respiratory air capability. The air suits should be cleanable, safe, comfortable, and available in a variety of sizes. The respiratory air compressor should be placed on an uninterruptable power system.

Provisions should be made for water decontamination showers and wastewater diversion. The decontamination showers and rooms can have drains and shunts to divert wastewater and treat it, if necessary, before it is sent to the general waste stream. Fig. 6 shows an operator gowned in



Fig. 6 Operator gowned in a breathing-air suit and standing in the doorway of decontamination shower.



Fig. 7 Aerosol isolation suite where manufacture takes place in a totally enclosed system with clean-in-place capability. The operator is gowned in a washable suit and is wearing a powered-air-purifying respirator.

a breathing-air suit and standing in the doorway of a water decontamination shower, which can be programmed to deliver a water wash followed by drying air for a predetermined cycle time.

Fig. 7 illustrates the use of facility design, equipment design, and personnel protection. The picture is taken from an aerosol isolation suite where the manufacturing process is performed in a totally enclosed system with CIP capability. The operator is shown in a washable Tyvek[®] suit, and is wearing a battery-powered, air-purifying respirator.

The weighing, dispensing, sampling, and vacuum dust-collection areas for highly active materials require special design considerations. Fig. 8 shows a weighing and dispensing room. The airflow has been validated by video smoke and tracer gas tests. In addition, local point containment is achieved by local exhaust units that fit directly over the variously sized mauser drums. Operators have access to breathing air suits if desired. Fig. 9 shows a purpose-built containment unit used for

dispensing small quantities of highly active compounds. The extracted dust is collected in a specially designed bag house. The dust is remotely augered into drums in a self-contained recovery room equipped with air locks, respiratory air, and water decontamination showers. The bag house exhaust air is filtered through a bag-in/bag-out HEPA filter system.

OPERATIONAL COSTS

The operational costs vary greatly depending on the types and extent of utilities, as well as the technology incorporated. A sterile product facility is much more expensive to operate than a pilot plant manufacturing only solids. NCEs, which may present bioavailability and/or stability challenges, may require the use of newer, more expensive technologies. Operational costs can be minimized if the pilot plant is on the same site with a production facility where power-generation capability and support staff



Fig. 8 Weighing and dispensing room. Each mauser drum has special extraction units that fit directly over the variously sized drums.

functions, such as maintenance, calibration, engineering, housekeeping, security, validation, QA, microbiology, QC, shipping, receiving, and training, can be shared. If the facility is isolated, space requirements increase significantly and the support overhead increases the operational costs.

The building and operating costs can often be justified based on the direct costs of product development and clinical manufacture under contract. In addition, delays in clinical programs can cost the company millions of dollars in lost revenue. However, if building or operational costs cannot be justified, it may be more effective to outsource work to a CRO.^[24] This may be especially true for new or smaller companies.^[25–27]

In the past, CROs have been used by pharmaceutical companies to provide extra capacity in the later stages of development. More recently, CROs have expanded their role to include early stage development and specialized technologies.^[28]

Outsourcing to a CRO is not fail-safe. If it is to be a successful partnership, both companies must explicitly state the needs and expectations of each other. This should include considerations of confidentiality and ownership of intellectual property.^[29,30]

SUMMARY

The design and construction of pilot plant facilities must be well thought out in order to accomplish the strategic objectives desired. Flexibility in design is critical to allow product and process development, clinical supplies manufacturing, and process technology scale-up and evaluation to be carried out in an efficient, scientific manner. In today's business environment, the well-designed pilot plant is necessary to shorten the product development cycle, with the ultimate goal of getting new products and medicines to the market rapidly and efficiently.



Fig. 9 Containment unit built for dispensing small quantities of highly active compounds.

ACKNOWLEDGMENTS

The second edition authors would like to sincerely acknowledge the work of the first edition authors—Robert M. Franz, Robert D. Copeland, Lawrence D. Lewis, and William C. Stagner. Most of the first edition's text is retained. The basic goal in revising the first edition chapter was to add current references. This was easily done with the excellent foundation and framework the first edition authors provided.

Next, the authors would like to acknowledge Alan F. Parr for reviewing the revised chapter.

Finally, the authors would like to acknowledge the administrative help of Cindy Rozier in preparing this revision.

REFERENCES

1. Cole, Graham C. *Pharmaceutical Production Facilities: Design and Applications*, 2nd Ed.; Taylor & Francis Inc.: Bristol, PA, 1998; 7–23.
2. Cole, Graham C. *Pharmaceutical Production Facilities: Design and Applications*, 2nd Ed.; Taylor & Francis Inc.: Bristol, PA, 1998; 29–45.
3. Willig, Sidney H. *Good Manufacturing Practices for Pharmaceuticals*, 5th Ed.; Marcel Dekker, Inc.: New York, 2001; 367–371.
4. Cole, Graham C. *Pharmaceutical Production Facilities: Design and Applications*, 2nd Ed.; Taylor & Francis Inc.: Bristol, PA, 1998; 257–261.
5. Cole, Graham C. *Pharmaceutical Production Facilities: Design and Applications*, 2nd Ed.; Taylor & Francis Inc.: Bristol, PA, 1998; 245–291.
6. Chao, Allen Y.; Forbes, F. St. John; Johnson, Reginald F.; von Doehren, Paul. Prospective process validation. In *Pharmaceutical Process Validation*, 2nd Ed.; Berry, I.R., Nash, R.A., Eds.; Marcel Dekker, Inc.: New York, 1993; 234–238.
7. Chao, Allen Y.; Forbes, F. St. John; Johnson, Reginald F.; von Doehren, Paul. Prospective process validation. In *Pharmaceutical Process Validation*, 2nd Ed.; Berry, I.R., Nash, R.A., Eds.; Marcel Dekker, Inc.: New York, 1993; 236–248.
8. Lorenzo, P.; Wheeler, W.P.; Ciepiak, G.S.; Newberger, S.F.; Van Vessum, M.; Hsu, R.T.; Chu, J.C.; Koller, N.C.;

- Linder, J.; Yoakam, D.; Mann, B. Oral solid dosage forms. *ISPE Baseline Pharmaceutical Engineering Guide*, 1st Ed.; ISBE, 1998; Vol. 2, Tampa, FL.
9. Davis, B.; Durkin, J.; Xilogiannis, D.; Fortune, D.; Shelley, S.; Huibregste, W.; Stamford-Smith, A.; Wood, C.; Mulhall, M.; Farquharson, G.; McLucas, D.; Walker, N. Sterile manufacturing facilities. *ISPE Baseline Pharmaceutical Engineering Guide*, 1st Ed.; ISPE, 1999; Vol 3, Tampa, FL.
10. Ricigliano, J.V.; Rowland, P.E. HVAC and utilities guidelines for pilot plants. *Pharm. Eng.* **1994**, *14* (5), 56–59.
11. Geisler, G.L.; Myers, R.; Biskup, J.; Bader, B. Water and steam systems. *ISPE Baseline Pharmaceutical Engineering Guide*, 1st Ed.; ISPE, 2001; Vol. 4, Tampa, FL.
12. Lorenzo, P.; Wheeler, W.P.; Ciepiak, G.S.; Newberger, S.F.; Van Vesslem, M.; Hsu, R.T.; Chu, J.C.; Koller, N.C.; Linder, J.; Yoakam, D.; Mann, B. Oral solid dosage forms. *ISPE Baseline Pharmaceutical Engineering Guide*, 1st Ed.; ISPE, 1998; Vol. 2, 48, Tampa, FL.
13. Davis, B.; Durkin, J.; Xilogiannis, D.; Fortune, D.; Shelley, S.; Huibregste, W.; Stamford-Smith, A.; Wood, C.; Mulhall, M.; Farquharson, G.; McLucas, D.; Walker, N. Sterile manufacturing facilities. *ISPE Baseline Pharmaceutical Engineering Guide*, 1st Ed.; ISPE, 1999; Vol. 3, 114, Tampa, FL.
14. Conaway, J. Plant design and engineering. In *Automation and Validation of Information in Pharmaceutical Processing*, DeSpautz, J.F. Ed.; Marcel Dekker, Inc.: New York, 1998; 367–386.
15. Lorenzo, P.; Wheeler, W.P.; Ciepiak, G.S.; Newberger, S.F.; Van Vesslem, M.; Hsu, R.T.; Chu, J.C.; Koller, N.C.; Linder, J.; Yoakam, D.; Mann, B. Oral solid dosage forms. *ISPE Baseline Pharmaceutical Engineering Guide*, 1st Ed.; ISPE, 1998; Vol. 2, 65, Tampa, FL.
16. DeBellis, Michael P. Dust explosion protection in pharmaceutical processing. *Pharm. Eng.* **2000**, *20* (2), 44–53.
17. NFPA Code 68, Guide for Venting Deflagrations, National Fire Protection Association, 1998 Edition.
18. NFPA Code 69, Standard on Explosion Prevention Systems, National Fire Protection Association, 1997 Edition.
19. NFPA Code 318, Standard for the Protection of Cleanrooms, National Fire Protection Association, 2000 Edition.
20. NFPA Code 497, Recommended Practice for the Classification of Flammable Liquids, Gases, or Vapors and of Hazardous (Classified) Locations for Electrical Installations in Chemical Process Areas, 1997 Edition.
21. NFPA Code 654, Standard for the Prevention of Fire and Dust Explosions from the Manufacturing, Processing, and Handling of Combustible Particulate Solids, 2000 Edition.
22. Lorenzo, P.; Wheeler, W.P.; Ciepiak, G.S.; Newberger, S.F.; Van Vesslem, M.; Hsu, R.T.; Chu, J.C.; Koller, N.C.; Linder, J.; Yoakam, D.; Mann, B. Oral solid dosage forms. *ISPE Baseline Pharmaceutical Engineering Guide*, 1st Ed.; ISPE, 1998; Vol. 2, 45, Tampa, FL.
23. Lorenzo, P.; Wheeler, W.P.; Ciepiak, G.S.; Newberger, S.F.; Van Vesslem, M.; Hsu, R.T.; Chu, J.C.; Koller, N.C.; Linder, J.; Yoakam, D.; Mann, B. Oral solid dosage forms. *ISPE Baseline Pharmaceutical Engineering Guide*, 1st Ed.; ISPE, 1998; Vol. 2, 43, Tampa, FL.
24. Lightfoot, G.D.; Vogel, J.R. Contracting CROs into your organization: New strategies for new challenges. In *The Drug Development Process—Increasing Efficiency and Cost-Effectiveness*, Welling, P.G., Lasagna, L., Banakar, U.V., Eds.; Marcel Dekker, Inc.: New York, 1996; 317–335.
25. Katz, M. Planning ahead for manufacturing facility changes (outsourcing). *Biopharm* **2001**, *14* (1), 46–47.
26. Beddia, F.; Bertrand, S.; Dion, S.B.; Brown, J.G.; Palado, T.; Rolfe, B. Project management of small R&D companies part I. *Pharm. Technol.* **2000**, *24* (11), 94.
27. Beddia, F.; Bertrand, S.; Dion, S.B.; Brown, J.G.; Palado, T.; Rolfe, B. Project management of small R&D companies Part II. *Pharm. Technol.* **2000**, *24* (12), 68.
28. Soltero, R.A.; Stagner, W.C.; Kenley, R.A.; Caudill, W.L. The changing role of pharmaceutical development organizations. Contract services 2000 for the pharmaceutical industry. *Pharm. Technol.* August **2000**, *2* (Suppl.), 36–42.
29. Woodward, R.S. Keys to developing an effective outsourcing strategy. *Biopharm* **2000**, *13* (2), 22–26.
30. Messing, S.H. Getting the most out of outsourcing. *Pharm. Technol. Eur.* **2000**, *12* (5), 72–74.

Pilot Plant Operation

Mickey L. Wells
Samuel B. Balik
Ralph B. Caricofe
Charles W. Crew
Ronald A. Sanftleben
A. Wayne Wood

GlaxoSmithKline, Inc., Research Triangle Park, North Carolina, U.S.A.

INTRODUCTION

Several attributes of the pilot plant allow it to serve as a vital function in achieving the key strategic objectives of:

- Formulation and process development.
- Clinical supply manufacture.
- Technology evaluation, scale-up, and transfer.

These attributes include: 1) a current good manufacturing practices (cGMPs) environment; 2) a flexible highly trained staff; 3) equipment to support multiple dosage form development; and 4) equipment at multiple scales based on similar operating principles to those in production. These key attributes help define current trends in pilot plant operation. Most of the development and clinical manufacturing activities should be condensed and performed under cGMPs. The application of cGMPs early in the development process is mandated by the current regulatory environment, while at the same time market competition is demanding shorter development times. These aspects, and the often limited availability of drug substance, result in the necessity to obtain as much development information as possible from each and every batch manufactured in the pilot plant. Even batches targeted solely for clinical supplies often incorporate some aspect of experimental design, process optimization, or scale-up in an effort to maximize the information obtained during new product development.

In order to improve efficiencies within the pilot plant, many firms are abandoning the practice of having separate clinical supplies manufacturing staff and formulation process development staff in favor of a more flexible, responsive group capable of supporting product development from beginning to end. The staff must be trained in formulation development, process development, scale-up, technology transfer, cGMP compliance, and clinical supplies manufacturing. In order to maximize development

efficiency, it is crucial to incorporate clinical supplies manufacture, process development, scale-up, and technology transfer activities within the same departmental organization. To this end, a matrix-management approach is often utilized for project teams responsible for product development activities.

Although the pilot plant must simulate the manufacturing environment in which the new product will ultimately be produced, there are many differences in operation because of the specific objectives of the two types of facilities. The pilot plant facilitates product development activities, whereas the manufacturing plant routinely fabricates products for the marketplace. The pilot plant must be flexible in operation in order to accommodate the very nature of product development, which is often at odds with the routine required in a true manufacturing facility.

Recognizing and managing the difference in objectives between the pilot plant and production facility form the basis for success in pilot plant operation. This chapter highlights the unique operational aspects of the pilot plant environment.

VALIDATION

Most basic validation activities within the cGMP pilot plant are identical to those expected in practice in a manufacturing facility.^[1–5] A validation master plan should be developed that addresses the design specifications and qualification, installation qualification (IQ), operational qualification (OQ), and performance qualification (PQ) of all major utility systems, process equipment, and computer control systems. Installation and commissioning data should be retained as part of an engineering documentation package. A workable change control system should be established. Process areas should meet or exceed current Food and Drug Administration (FDA) standards.

A fully validated pilot plant ensures compliance with cGMPs, but more importantly, guarantees data integrity during product development. Data generated during this period in the pilot plant are typically used to set product specifications and in-process controls, as well as process variable control limits. Validation of the utilities and equipment in the pilot plant ensures that the facility adequately simulates the manufacturing environment. This is especially true for automated equipment and equipment requiring validation of computer control systems.^[6] The information collected during product development in a validated facility provides the appropriate database for subsequent scale-up and technology transfer to the manufacturing site.

There are several areas within the pilot plant environment that routinely differ with respect to validation when compared to a production facility; these include manufacturing process validation, cleaning validation, and revalidation activities. Manufacturing processes within the pilot plant environment are often not repeated. A process may be implemented only once before a change in scale, equipment, or process variable is implemented. The usual manufacturing scenario of performing process validation on three or more identical sequential batches is often not relevant in the pilot plant environment. In order to ensure that each batch is acceptable, especially with respect to clinical use, it is important to collect sufficient in-process data to guarantee the quality of the batch.^[7] Data can be generated during the actual experimental development work showing that the process achieves what it purports to do, or extensive samples can be collected from each batch as manufactured. A combination of experimental data and batch manufacturing data may also serve to prove process integrity. Specifications should be set with respect to critical product characteristics, such as content uniformity and potency. The lack of a sufficient database and minimal experience with a process early in development often require that the specifications be set at the widest limits allowed by current regulatory guidelines, e.g., USP specifications. As experience with the product grows, and assuming the database supports it, specifications may become tighter as technology transfer proceeds.

Cleaning validation is different in the pilot plant environment than in manufacturing.^[8–10] Again, the lack of process repetition often creates the need to verify that equipment is clean before and/or after the manufacture of batches used in clinical studies on a per batch basis. This can be accomplished by swab testing critical product contact surfaces before and/or after equipment use to ensure that residual drug is absent or at an acceptably low level. Cleaning techniques for new chemical entities (NCEs) must be verified on materials of construction or

based on solubility before new compounds enter the pilot plant environment. This information, combined with a routine cleaning procedure, i.e., standard operating procedure (SOP), for a particular piece of equipment, may serve as an acceptable alternative to swabbing after every batch.

Finally, utility and equipment revalidation or validation review may be structured differently than techniques typically used in the manufacturing environment. Unlike the production facility, the pilot plant may not use every piece of equipment or system on a routine basis. In fact, some equipment may remain idle for significant periods of time. Revalidation may need to be scheduled based on hours in use rather than a program designed for periodic review. It should be pointed out that some equipment or systems might need more frequent revalidation activities because of lack of use. No matter what basis is chosen, it is imperative that a formal revalidation program be established in the pilot plant environment.

TRAINING

The development of a training program for pilot plant staff is complicated by the inherent diversity of operations and personnel. Unlike the production facility, which is highly specialized and subject to infrequent operation changes, the pilot plant must be prepared to make a variety of dosage forms in response to a wide range of product development programs. The diversity of pilot plant operations, equipment, and personnel requires a flexible entry, reinforcement, and remedial training program. An extensive tracking system and flexible scheduling system are also required. Most importantly, the training program must be developed to meet the changing priorities of the product development cycle while maintaining cGMP compliance in the manufacture of clinical supplies.

Training within the pilot plant can be broken down into four major areas:

1. Compliance with quality standards, such as cGMPs.
2. Safety and environmental responsibilities.
3. Compliance with SOPs.
4. Technical skills and knowledge.

The order in which these objectives are presented is referred to as the “top-down approach” to training. It has been used successfully in the pharmaceutical industry.^[11] A systematic approach to develop a cGMP training program should include formal training plans, written training manuals, effective training methods, assignment of responsibilities, written documentation and periodic

review, and written certification of instructors.^[12] A dynamic training process is extremely important. Initial training must be reinforced with current, up-to-date programs that address changes in job requirements, performance expectations, and technology.

Pilot Plant Staff

The staff operating the pilot plant is often diverse in educational background, work experience, and function. The multidisciplines can be divided into the broad categories of scientific groups, support groups, and contractors. The last two groups can be further subdivided into technical and nontechnical personnel. Regardless of the category, a basic training program should be defined for each. Access to specific areas of the pilot plant may be based on the level of training applied to a particular category of staff, e.g., sterile environment.

Training for the scientific group, which includes formulation and process development scientists, represents the majority of the training requirements. Personnel with diverse educational backgrounds may be challenged to operate in different areas of the pilot plant on a multitude of equipment in different dosage form areas. The mindset of the scientist is experimentation. The evaluation of formulations and processes is the focus of their work. It is extremely important that training be well defined to attain an adequate level of competency, and that all required training be completed prior to initiation of any pilot plant activity. This ensures the safety of personnel and equipment throughout the process and guarantees the manufacture of a quality product.

Cross-training scientific staff in the use of many types of pilot plant equipment is vital to carrying out efficient product development programs, but creates significant difficulties with respect to training. Reinforcement training becomes paramount in order to maintain the level of competency necessary to ensure proper operation of equipment and compliance with procedures. A well-trained individual who regularly performs an operation obtains the type of reinforcement necessary to successfully and correctly complete a task. For this reason, it may be necessary to devote highly trained scientific staff to operations in specific areas of the pilot plant, e.g., the sterile process areas. Other scientific staff with less expertise in a particular operational area can assist in development activities based on a lower level of basic training, but may be relegated to less technical operations compared to those with routine experience. It may sometimes be necessary to assign “experts” in a defined operational area of the pilot plant based on the

development activities required. On-the-job training with an experienced guide or “expert” is a simple and effective approach to entry and reinforcement training. Training provided by the equipment manufacturer can be useful in the completion of entry-level training and development of reinforcement training programs for the scientific and support staff.

The technical support groups must also be appropriately trained. They may represent equipment and facility engineering, maintenance, and calibration groups, as well as quality groups, i.e., QA/QC. Although their educational backgrounds may be very diverse, the focus of these groups is confined to pilot plant operation rather than the experimental techniques of the scientific staff. The greatest challenge with respect to training equipment engineering, maintenance, and calibration groups is to impart knowledge of how their responsibilities and actions in the pilot plant affect the ultimate quality of the products manufactured for clinical use. The quality groups typically have a good grasp on cGMPs and how their activities within the pilot plant can affect product quality. Training of quality groups is often very specific with respect to their responsibilities, but it is important that they also receive appropriate basic training in cGMPs.^[11]

The nontechnical support groups represent individuals responsible for cleaning and sanitation. Training requirements are usually limited to access specific areas of the pilot plant.

The importance of training temporary and contract labor should not be overlooked. The time period and frequency with which the contractor or temporary employee will operate within the facility and the required level of access should be used to define requirements.

Training Program

A pilot plant training program has several key aspects that may require a different approach compared to programs typically utilized in a production environment. These include identification of educational and instructional training programs and development of training manuals, individualized training syllabuses, and a documentation and tracking system.

Educational and instructional training programs should be defined for the facility based on the scope of operation of each functional area. For both educational and instructional training, manuals should be developed. The instructional training manuals can be divided into entry, reinforcement, and remedial training. Entry-level training is targeted for a new or prospective employee and should provide assurance that a certain skill level has been attained.

Reinforcement-level training is intended for an employee who has already been trained in a certain area. This training is intended to reinforce skills already acquired and to prevent performance problems. Remedial training is intended to correct identified skill deficiencies. The training required for equipment may range from reading the SOP and completing an evaluation form, e.g., test, for simple equipment to more extensive supervised and documented practical experience with a qualified trainer.

Pilot plant educational courses are usually tailored to a general audience since they impact all personnel. Instructional courses, however, may be very technical and assume a certain level of competency. In all cases, the evaluations developed as part of the training manual should effectively assess the student's level of competency. Any deficiencies identified in the assessment should be immediately corrected with additional training.

New technologies, frequently encountered during the development process, affect pilot plant training programs. New technologies pose a special problem because training must be conducted by a qualified individual. Some new technologies can be studied at the equipment supplier's facility and subsequent initial training can occur in the pilot plant with a representative of the equipment manufacturer as trainer. For technologies invented in the development group, it is recommended that the inventor become the initial trainer.

The diversity of pilot plant operations to be performed and the variety of personnel involved require an individualized training syllabus for each employee. The "top-down" approach mentioned previously works well in this regard. The focus is again to complete the higher education courses before proceeding to the instructional or area-specific procedures. Access to and operation within the pilot plant areas can be controlled by successful completion of basic requirements defined for each functional area. A simple way to identify and document each individual's training requirements is a checklist of courses and procedural training required based on the pilot plant area and function within the area. The employee reviews the checklist with management and identifies the training required to work on a particular project within the pilot plant. The result is a documented individualized training syllabus. Once defined, the syllabus can be used to train personnel to meet the resource needs for the project.

All training must be documented and a tracking system established by using the individualized syllabus to identify initial and reinforcement training requirements. The syllabus can also be used for entry into a database or one of the commercially available computer tracking systems.

A master schedule of courses should be developed for distribution to the pilot plant staff. Scheduling of courses

can be very resource demanding, depending on the types of courses. Resources can be optimized with the help of videos, computer-based training (CBT), and well-designed self-study courses. However, the benefits of group "on-the-job" training are important in reinforcing the learning gained by self-study courses.^[13]

Training Media

Self-study programs have been used to successfully meet the challenge of training a diverse pilot plant staff. The computer industry has made significant strides in learning retention with the aid of visualization technologies based on increased information storage capabilities on compact discs.^[14] The techniques include interactive CBTs and interactive audio and video compact discs (CD-I).^[15] The former is enhanced by interactive courseware, which stimulates interest and documents completion of the course work. Distribution of multimedia over local and wide area networks (LANS and WANS) offers additional flexibility in that the operator can access the training program from work and operational areas. The advantages of these types of technology to pilot plant staff include:

- Immediate availability of material for entry or reinforcement training from various work locations.
- Immediate retrieval of specified information.
- Visual representation of procedures and operations.
- Consistency of instruction.
- Personalized learning rate.
- Practice by performing simulated operations.
- A more conducive environment for cross-training.

Although some of these advantages are limited by current technologies, more sophisticated training media based on enhanced visualization systems will be available in the future.

ENGINEERING, MAINTENANCE, AND CALIBRATION SUPPORT

As pilot plant facilities and operations become more complex and technically diverse, the need for support groups, such as engineering, maintenance, and calibration, becomes critical to perform planned activities in a cGMP-compliant manner.

Engineering

Engineering and architectural support is obviously required during the design, construction, commissioning,

and validation of the pilot plant facility. However, some larger facilities may require dedicated engineering support staff for the coordination, scheduling, and direction of ongoing operations. The primary objective of the engineering staff is to provide timely support and direction in facility and equipment requirements necessary to ensure the efficient implementation of drug development projects. Support provided by the engineering staff may include:

- Equipment acquisition, installation, and repair.
- Engineering documentation and control for new systems and equipment.
- Management of the facility to ensure that critical systems, e.g., environmental controls, utilities, etc., are operational for compliance with cGMPs.
- Coordinating and scheduling equipment set-up and related activities for development projects.
- Ensuring a controlled inventory of critical spare parts, e.g., tablet tooling.
- Budgeting for equipment, technology upgrades, and operating supplies.
- Directing housekeeping activities.

Engineering support staff also play a pivotal role in pilot plant operation by contributing to the selection and evaluation of new process and material-handling equipment, recommending containment techniques to minimize exposure to potent drugs, and providing support for new technologies which may be critical in the development of an NCE. Equipment evaluation, procurement, installation, and qualification can be enhanced by the formation of project teams, which include engineering and scientific staff, as well as members from purchasing, validation, calibration, maintenance, safety, environmental, quality assurance (QA), and other necessary functions. A written procedure outlining the roles and responsibilities of members of the team can be critical to the successful implementation of a new facility and process technologies necessary for product development.

For large facilities, internally developed or commercially available computer software has been found useful for the engineering support staff in scheduling process equipment and space for development, maintenance, calibration, and validation activities.^a

Maintenance

A maintenance program is necessary within the pilot plant to meet the requirements of cGMPs and to ensure data

integrity and equipment reliability during the development process. The maintenance program must be documented and written procedures established. All maintenance activities on critical systems and equipment must be documented, approved, and archived for retrieval during FDA inspections. Changes not considered routine must be documented and approved through an established control procedure. It is important to document critical changes to computer systems or code necessary for maintenance activities.

Although the basis for a pilot plant maintenance program is similar to that of a production facility, there are several differences worth mentioning. The pilot plant maintenance program may be more diverse with respect to systems and equipment because of the nature of product development. Furthermore, equipment utilization within the pilot plant may be less routine than in the manufacturing environment, requiring a different approach to preventative maintenance. A preventative maintenance program based on hours of utilization may be warranted for some equipment in the pilot plant rather than the typical scheduled approach, e.g., annual, semiannual, etc. Another approach is to routinely review the utilization of equipment in the pilot plant and reestablish the preventative maintenance program based on historical data. This requires close interaction between engineering support and maintenance staff. The results are significant improvements in resource utilization and efficiency during preventative maintenance activities.

It is critical that the maintenance staff operating within the pilot plant be responsive, well trained, and has expert knowledge of equipment and systems in the facility. Staff members should be keenly aware of how changes to critical systems and equipment can affect product quality. An effective written procedure is recommended for communicating critical variances identified during routine and nonroutine maintenance activities. Changes and variances must be communicated in a rapid manner to the engineering support staff and ultimately to the scientists in order to ensure the integrity of ongoing development work. Quality staff should be rapidly notified if clinical batches are affected in any way so that appropriate actions can be implemented.

A computer-based system for facility management, control, and data archiving is useful for identifying critical systems that may require maintenance, as well as trending alarms and environmental data for routine review.^{[16]b} Vibration analysis systems have been found useful in predicting maintenance and repair needs of motors and other systems.^[17]

^a<http://www.sap.com/company/> (accessed November 2001).

^b<http://www.wonderware.com/home.htm> (accessed November 2001).

Calibration

Routine calibration of critical instruments on systems and equipment is required for compliance with cGMPs, as well as for maintaining the integrity of data generated during the development process. The pilot plant environment is often more complex with respect to instrumentation than that found in manufacturing. A preponderance of data is often collected through extensive instrumentation during product and process development in order to determine critical and noncritical process variables and control levels. Once this information has been collected and analyzed in the pilot plant, only process variables deemed critical and their related instrumentation may be transferred to the manufacturing environment. For these reasons, pilot plant calibration staff must be well trained and have expertise in a wide range of instrument technologies. As with maintenance personnel, they must be fully aware of the effect of calibration-related activities on product quality, and variances should be relayed to appropriate engineering, scientific, and quality staff via written procedures. Because of the extensive instrumentation on equipment and systems, it is important that engineering, calibration, maintenance, and scientific staff determine those that are critical to product quality. Critical instruments should be routinely calibrated and identified as such to the user. Noncritical instruments may be used to collect additional data. They are often calibrated separately as part of the routine preventative maintenance program.

MATERIAL CONTROL, INVENTORY, AND DISPENSING

Material flow and control in the pilot plant require flexibility, precision, and broad scope. There are major differences between production and pilot plant facilities with regard to these activities. For example, long-range planning for the manufacture of development and clinical batches is often limited, inventory is highly varied and often not used repetitively as a result of the experimental nature of development activities, and material use and quality status must be considered as dispensing is planned. Conventional paper tracking systems can be used and are highly effective in many manufacturing and pilot facilities, but it is desirable to implement a more flexible and efficient computerized system in the development-oriented environment of the pilot plant. The system should distinguish between various function-based user groups. Membership in a particular group determines what system capabilities are available to the user. The detailed

description that follows addresses how experimental and clinical supply requirements can be controlled and separated within the pilot plant, allowing maximum flexibility of activities.

Inventory

Inventory can be maintained in a validated computer-based inventory–ordering–dispensing system. Inventory may consist of stock articles, which are kept on hand for general use, and specialty articles, which are ordered for a particular purpose. The former should be automatically tested for approval, whereas the latter can be received into inventory as either clinical or experimental material. Clinical materials must be fully tested and formally approved prior to use. Experimental items are available for nonclinical work only. This approach allows tracking of all lots received, while minimizing resource requirements in the QA and quality control (QC) areas.

Articles must be identified in the inventory database with a unique code before they can be received into the pilot plant.^[18] Upon receipt, the materials should be checked, verified, and labeled with inventory and quarantine labels. A receipt report is electronically received by the QA group, which initiates appropriate sampling. The sampled material, which to this point is kept in a secure holding area, is moved to a warehouse upon release by QA.^[19] Any change of location is recorded in the computer system.

Quantity, status, and location of inventory should be ascertained using a statistically based method, which can separate the inventory according to the frequency of usage. Articles with high activity are sampled at a higher rate. Inventory accuracy, in terms of quantity, must be maintained within predetermined levels, for example, between 95% and 105% of the actual level, or an investigation will be required to identify the cause of the discrepancy.^[20] This approach differs from the cycle count methodology typically employed in a manufacturing operation.^[21]

Inventory quantity, status, and location should be tracked at the container level. The system should allow users to review inventories on-line, or in printed form, via a series of predefined queries and reports or via flexible queries, which may be generated on an ad hoc basis. Access to the system should be available via any number of remote terminals.

Orders

All orders must be placed through the computer system. A batch or general order can be placed for any article defined

in the inventory system directly from a desktop computer. Placement of a batch order automatically generates a batch number. General orders are placed for all articles not used in a batch formulation.

The ordering mechanism should differentiate between order types (experimental vs. clinical), which will allow the system to check the intended use against the material status. Should the material status not be suitable for the intended use of the requested lot, e.g., quarantined or experimental material has been ordered for clinical use, the user should be alerted immediately to resolve the issue. A hold can be placed on the order, pending attainment of the proper material status. During this time, dispensing of the article in question is blocked. The ordering procedure should be flexible to the point that an individual placing an order may select a dispensing weight variance, e.g., $\pm 1\%$, identify a specific lot and container, or allow the system to pick the lot and container automatically based on first in, first out (FIFO) criteria. The order may also be employed to describe use, reference a notebook number, or conduct other activities.

The dispensing of designated materials, typically drug substances, should require specific authorization before the order can be completed. Materials that require this degree of control are identified during the definition stage, prior to formal receipt into the facility. An order submitted for dispensing can be printed in the dispensing area for scheduling by dispensing personnel. Receipt of the order, projected dispensing date, and completion for dispensing can be confirmed and communicated to the requester via electronic mail.

Labeling

The system must provide the ability to print various inventory, sample, and pre- and post-dispensing labels. All labels should include a bar code and printed text with the pertinent batch and/or lot information. Labels must comply with GMP–GLP (good laboratory practices) requirements to meet the needs of both laboratories and clinical manufacturing. Labels should be available in multiple sizes to accommodate laboratory sample containers as well as production drums.

Future Systems

The electronic inventory, ordering, and dispensing system is the first step in creating a computer-based system to support the areas requiring GMP documentation, and/or manually intensive operations, which provide a

disproportionately small return on the effort invested. Future improvements to the system include:

1. An electronic interface between the user, the inventory system, QA, and purchasing to support a paperless purchase order.
2. Electronic logbooks.
3. Electronic batch records.

Implementation will improve efficiency as well as quality compliance by consistent application of regulatory guidelines. Ancillary benefits include the ability to retrieve documentation quickly for an FDA inspection and the development of accessible product histories, which can be electronically passed to operations as products move from development to commercial sites.

PROCESS AND MANUFACTURING ACTIVITIES

Pilot plant processes and manufacturing activities include formulation and process development studies, clinical supply manufacture, and technology evaluation, scale-up, and transfer. Packaging for stability and clinical studies may also occur in the pilot plant, but these activities are often performed in separate, well-defined facilities. It is beyond the scope of this article to discuss all of these areas in detail. Since pilot plant construction is often driven by the strategic necessities defined earlier, only these areas are addressed.

Formulation and Process Development

Scientific staff must be able to access the pilot plant and operate in an experimental mode using procedures and processes without compromising the cGMP integrity of the facility.^[22] Procedures should be flexible enough to allow development batches to be manufactured in a fashion considered efficient by the development scientist. As previously discussed, it is important to have a detailed material control, inventory, and dispensing system in place to address differences in materials for experimental and clinical manufacturing. In-process and finished products should be labeled and recognizable with respect to experimental vs. clinical materials. For some early development work or evaluation of new technologies, equipment may be utilized before being fully validated. Research notebooks, rather than formalized batch records, are a necessity in the development environment. Quality group involvement with the manufacture of experimental batches may be limited. Typically, procedures that address

the operational aspects of the facility are equivalent no matter what the manufacturing activity. For example, equipment and room logs, cleaning procedures, training in equipment operation for safety purposes, and environmental monitoring programs should be transparent with respect to manufacturing activities.

Clinical Supplies

A key activity in a well-designed pilot plant is the manufacture of clinical supplies. A pilot plant should be flexible in design in order to accommodate a wide range of dosage forms. Flexibility provides the following advantages, which are directly applicable to clinical supply and “comparator” (positive clinical control) manufacture:

1. The optimum dosage form is chosen for development based on physical, chemical, or biological attributes of a new compound.
2. Different dosage forms can be developed in response to different patient demographics.
3. Comparator dosage forms can be developed which are different than forms routinely manufactured.

Efficient clinical supply manufacturing requires a wide range of equipment sizes, preferably of identical design, to handle requirements based on bulk drug supply. Clinical supply needs can quickly move from a few hundred dosage units in Phase I–II studies to millions of dosage units for Phase III–IV studies. The ability to rapidly scale processes is directly linked to the availability of similar, if not identical, equipment at the large scale. As clinical supply needs continue to grow, production of a few large batches compared to several smaller batches results in a substantial savings in time and resources for processing, testing, and documentation. The generic regulatory requirement that biobatches of solid, oral dosage forms one-tenth of the planned commercial batch size, or at least a minimum of 100,000 dosage units, highlights the advantage of producing large batches of clinical supplies in the pilot plant.

The quickest route of entry into a clinical trial is often through the use of a sterile parenteral dosage form. An ampule may be desirable early in development because only one material has to be considered for product compatibility, allowing rapid development and entry into clinical studies. The sterile process areas within a pilot plant should be operated with the same degree of control as those found in a routine manufacturing facility. Because of the unique environment, training required, and specialized operating procedures, sterile process areas within the pilot plant should strictly adhere to defined regulatory

requirements and be operated under cGMPs.^[23] The microbiological integrity of the area must be maintained by regular cleaning and sanitation. The sanitation program may revolve around the utilization of the area. Periods of heavy use may require more frequent sanitation. Ultimately, the environmental monitoring program determines the frequency required.

Because the scientific staff may utilize the sterile process area on an infrequent basis, training is critical.^[24] Staff expecting to work in the area should be thoroughly trained and participate in regularly scheduled bacteria–growth–media fills. Since batch sizes may be smaller than those typically found in production, the number of units filled during bacteria–growth–media simulations may be based on the actual scale and process time of the product manufactured in the pilot plant environment. It is recommended that at least 3000 units be filled to meet current regulatory guidelines if the scale of equipment warrants.^[23]

A range of process equipment should be available for the manufacture of sterile clinical supplies:

1. Glass and stopper washing and siliconization equipment.
2. Dry heat and steam sterilizers.
3. Ampule and vial fillers.
4. Freeze dryers.
5. Capping equipment.

The process equipment should cover a wide range of sizes, units, volumes, and operating conditions, e.g., speeds, for maximum flexibility in relationship to batch size and sterile dosage form requirements. Examples of pilot plant equipment utilized for the manufacture of sterile products are shown in Fig. 1.

As in a routine production facility, programs for environmental monitoring of viable and nonviable particulates, bioburden, and pyrogen loads of incoming materials should be established. Prewashed and sterilized components may be useful in improving operating efficiency.

The ability to manufacture a wide range of solid and nonsterile liquid dosage forms within the pilot plant is required for efficient clinical supply development. Areas within the pilot plant dedicated to nonsterile manufacturing should be of flexible design. Operation in these areas should meet the same standards of cGMP required in a routine manufacturing facility. Flexibility in equipment size is required to meet demands based on drug availability. Since many clinical supplies are double-blinded with the help of encapsulation technology, equipment should be capable of filling and overfilling

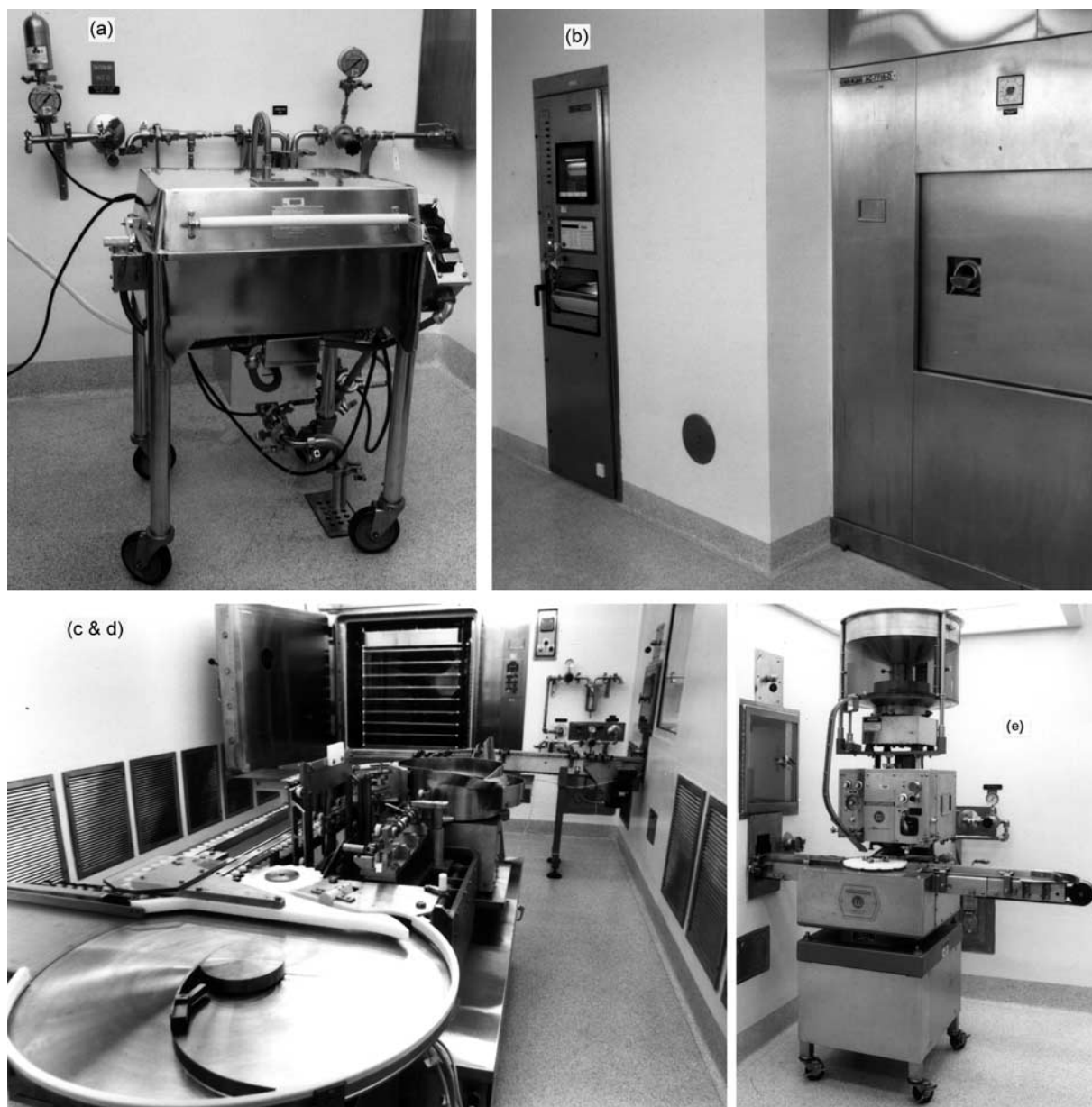


Fig. 1 Sterile dosage form processing equipment useful for clinical supply manufacture in the pilot plant. (a) Vial and ampule washer (Metromatic, Oyster Bay, NY). (b) Steam sterilizer (Amsco Finn-Aqua, Apex, NC). (c) Vial filler (TL Systems Corp., Minneapolis). (d) Lyophilizer (Edwards High Vacuum International, Tonawanda, NY). (e) Vial capper (The West Co., Phoenixville, PA).

powders, granules, beads, and tablets into hard gelatin capsules. Fig. 2 shows typical encapsulation equipment for the manufacture of clinical supplies.

Aerosol and semisolid dosage forms may be required in some clinical trials. Aerosol manufacturing needs semiautomatic and automatic equipment (Fig. 3) with the capability to produce from 100 to 10,000 units in order to meet clinical supply demands.

All of the above mentioned dosage forms may be utilized in clinical trials as the developed product or as a clinical comparator. Clinical “comparators” (positive clinical controls) are an outgrowth of the regulatory environment in which Phase III clinical studies must include the comparison of the generally recognized therapeutic treatment with the NCE being developed and a placebo. To avoid bias in the clinical program, the

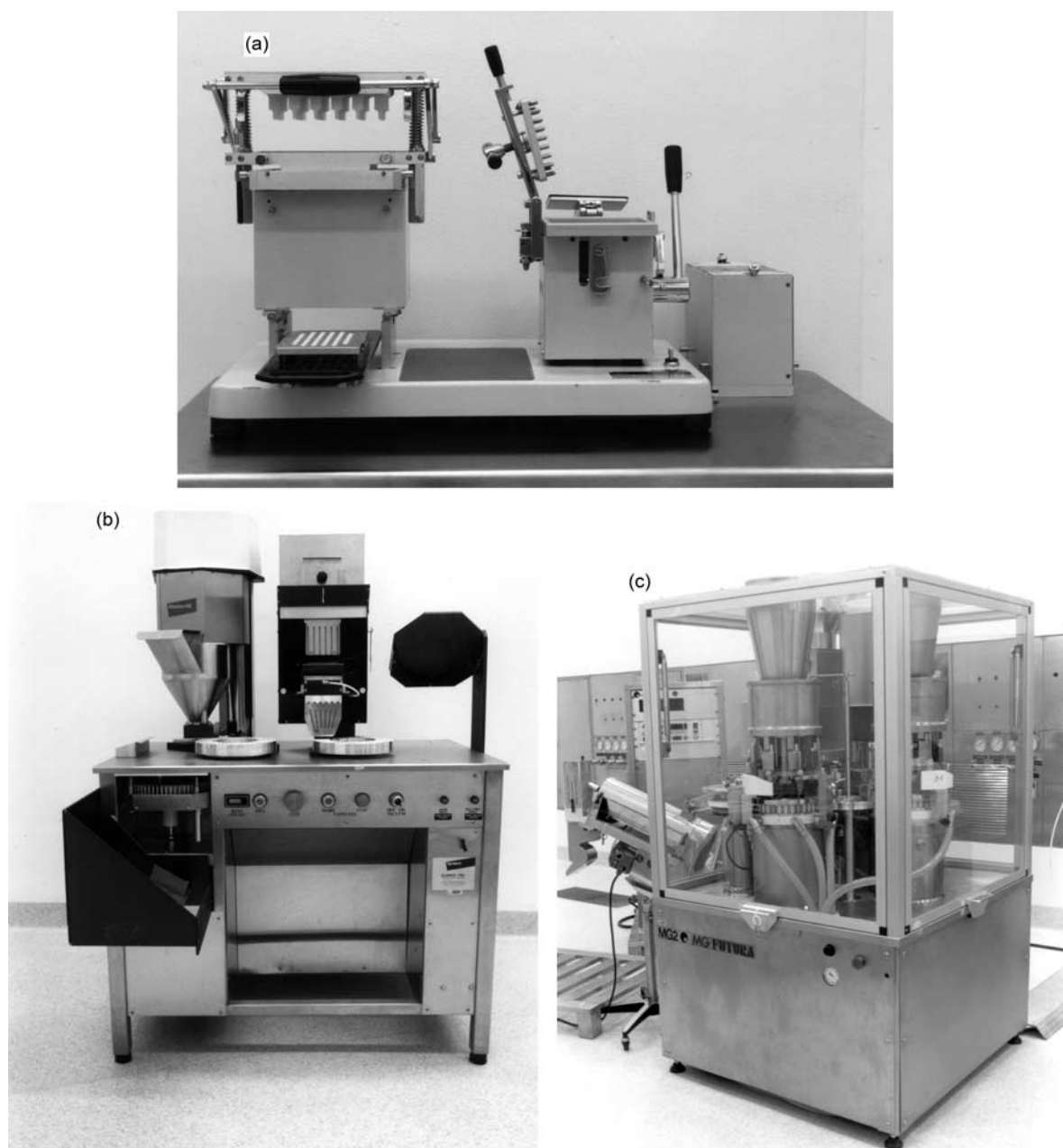


Fig. 2 Encapsulation equipment useful for clinical supply manufacture in the pilot plant. (a) Manual encapsulation equipment (Dott. Bonapace & C., Milano, Italy). (b) Semiautomated encapsulation equipment (Elanco Qualicaps, Indianapolis). (c) Automated encapsulation equipment (MG America, Inc., Fairfield, NJ).

comparator and placebo products are blinded, that is, manufactured in such a way that they are virtually indistinguishable from each other. For products that differ in dosage form, a double-dummy design is often employed. For example, if an NCE is being developed as a tablet and the desired comparator is marketed as a metered-dose inhaler (MDI), identical placebos of each

product would be produced. The patient in the clinic would receive an active MDI along with placebo tablets or vice versa. The blinding of comparator agents often mimics the development program for an NCE, but with greatly reduced timing. The blinding process typically requires limited preformulation studies, formulation studies, process development and optimization, scale-up and

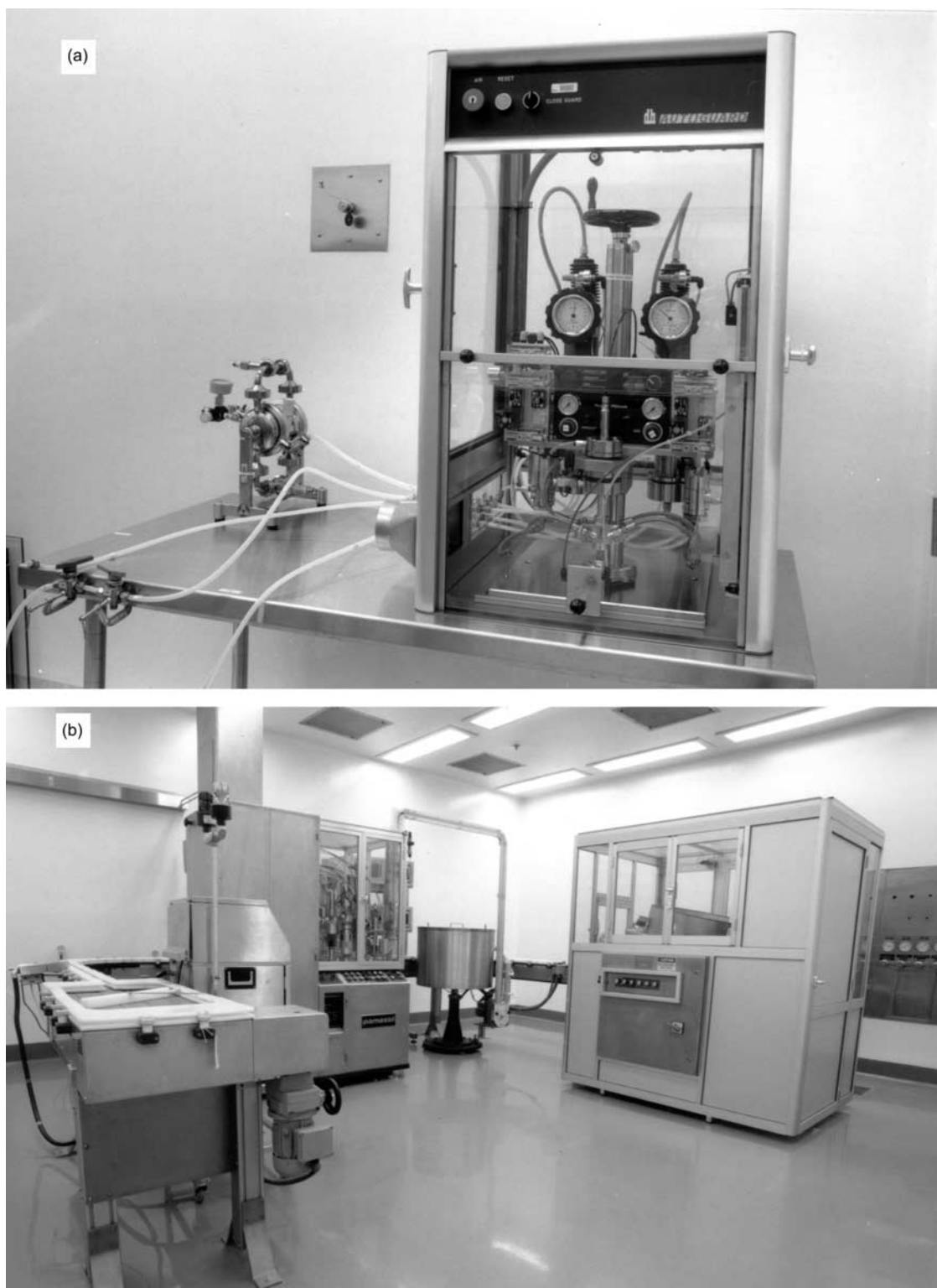


Fig. 3 Aerosol filling equipment useful for clinical supply manufacture in the pilot plant. (a) Manual filling equipment. (b) Automated rotary filling equipment (both D. H. Industries Limited, Barking Essex, UK).

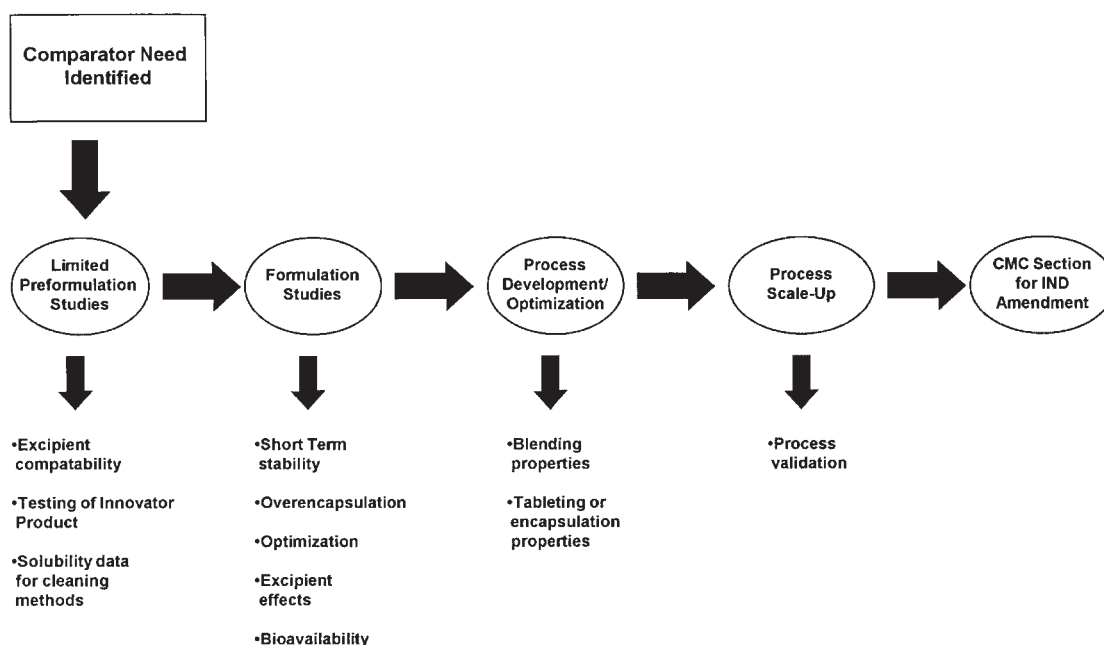


Fig. 4 Typical activities in the development of blinded clinical comparators.

regulatory documentation. Depending on the process, in vivo bioequivalence studies may also be required to ensure that blinding does not alter drug availability. Fig. 4 gives the details of the activities involved with the development of blinded clinical comparators. Blinding techniques for common dosage forms are shown in Table 1.

A matrix team approach has been utilized successfully by the authors to optimize the interfacing between development groups, support groups, and customers (medical or marketing groups) when preparing to manufacture new clinical supplies or comparators. This ensures that the finished products meet the customer's needs. Fig. 5 is a process map showing key matrix groups and activities typically involved with supplying clinical dosage forms. The breadth and scope of clinical supplies manufacturing, packaging, and shipping activities can become remarkably complex as illustrated by the following detailed description of a single block activity taken from Fig. 5 (preparation for manufacturing and packaging):

- Finalization of dosage formulas.
- Development, validation, and transfer of analytical methods.
- Final product specifications.
- Schedule of manufacturing.
- Secrecy agreements and statement of work and service contracts for contractors.
- In-process tests and specifications.
- Cleaning of manufacturing facility.

- Cleaning of equipment.
- Swabbing of equipment for residuals.
- Testing of equipment by microbiological means.
- Testing of facility by microbiological means.
- Sampling requirements.
- Swab results.
- GMP validation audit (contractor).
- Product–item code identification.
- Raw materials order.
- Certificates of analysis (COAs) for raw materials.
- Raw materials receiving and sampling.
- Testing of raw materials and recommendation for release.
- Raw materials release.
- Packaging materials order.
- COAs for packaging components.
- Release of packaging components.

Once a specific clinical product or comparator is developed, it may become more efficient and cost effective to manufacture and stock bulk supplies from the pilot plant when it is known that they will be routinely used in future clinical studies.

Technology Evaluation, Scale-Up, and Transfers

The pilot plant and its staff play a critical role in technology evaluation, scale-up, and transfer activities of new pro-

Table 1 Blinding Techniques for Common Clinical Dosage Forms Manufactured in the Pilot Plant

Techniques	Purpose
Solids^a	
Overencapsulation of intact tablet or capsule and filler	Stability testing and comparative dissolution
Film coating to obscure logos and proprietary color	Stability testing and comparative dissolution
Milling of dosage form and recompression on standard tooling	Stability testing and bioequivalency test
Milling of dosage form and encapsulation of powder	Stability testing and bioequivalency test
Encapsulation of tablet granulation	Stability testing and comparative dissolution
Sugar coating of tablets to obscure engravings	Stability testing and comparative dissolution
Liquids^b	
Repackage product and produce matching placebo	Stability testing
Obliterate all product labeling and packaging matching placebo in identical container dosage system	Stability testing of placebo
Metered-Dose Inhalers^c	
Obliterate all product labeling and produce matching placebo utilizing identical valve and canister, (utilize active product actuators for active and placebo products)	Stability testing of placebo only
Powder (for reconstitution)^d	
Obliterate all labeling and product packaging for a matching placebo powder	Stability testing for placebo only
Repackage product into standard packaging and produce packaging for a matching placebo powder	Stability testing for active placebo

^a For critical studies, bioequivalency testing should be considered in all cases. All techniques typically require the development of a matching placebo dosage form. Overencapsulation of a bead-filled capsule may necessitate a bead-filled placebo to match the rattling sound of the active product.

^b Color, clarity, viscosity, smell, taste, and identical packaging are critical blinding parameters.

^c Canister pressure, fill weight, actuation sound, and taste are critical blinding parameters.

^d Powder color, texture, particle size, fill weight, smell, ease of reconstitution, reconstitution volume, viscosity, taste, color stability after reconstitution, and physical stability of suspension are critical blinding parameters.

ducts. These activities begin early in the development cycle and include technical aspects of process development and scale-up, organization and responsibilities of technology-transfer teams, documentation of the transfer process, and often preparation for an FDA pre-approval inspection (PAI). A properly designed and operated pilot plant can enhance the collection of scientific data necessary to support internal transfer activities, as well as regulatory submissions and FDA PAIs.^[25]

Following are four key technical aspects that must be addressed during scale-up in the pilot plant:

1. Identification and control of critical component and formulation variables early in development.
2. Pilot plant equipment that simulates as closely as possible equipment used at the manufacturing site.
3. Identification of critical process parameters and operating ranges with pilot plant equipment through the use of engineering and regression models, i.e., statistically designed experiments.

4. Collection of product and process data to adequately characterize each unit operation.

The facility design plays a critical role in addressing each of these technical aspects. However, scientific and pilot plant staff involved in manufacturing operations within the pilot facility also play a key role in ensuring the smooth and timely transfer of process technology to the manufacturing site.

In the past, the transfer of formulation and manufacturing technology was sometimes discretely passed from development staff to manufacturing staff with little interaction or foresight. Nowadays, however, it is commonly recognized that interaction of these groups at an early development stage is critical in obtaining an efficient and successful transfer. Scientific and pilot plant staff play a key role in demonstrating new product manufacturing techniques to production personnel in the pilot plant environment. A team-oriented approach to the manufacture of pilot or large-scale batches in the pilot plant will

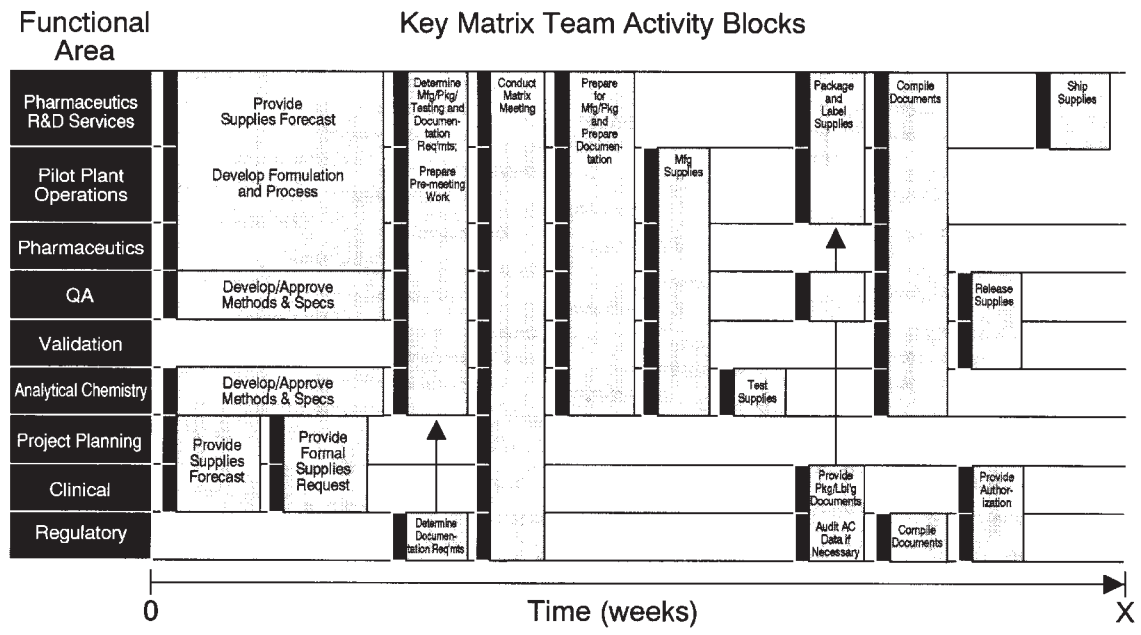


Fig. 5 Process map depicting key functional areas and matrix-related activities involved with supplying clinical dosage forms from the pilot plant.

allow key production site personnel to view and comment on the process and make specific recommendations for improvement based on knowledge of the manufacturing site. An example of the matrix team approach to the

transfer of manufacturing technology, key development, and production milestones, and the activities related to these milestones, is presented in Fig. 6. Interactions between pilot plant staff that begin during product

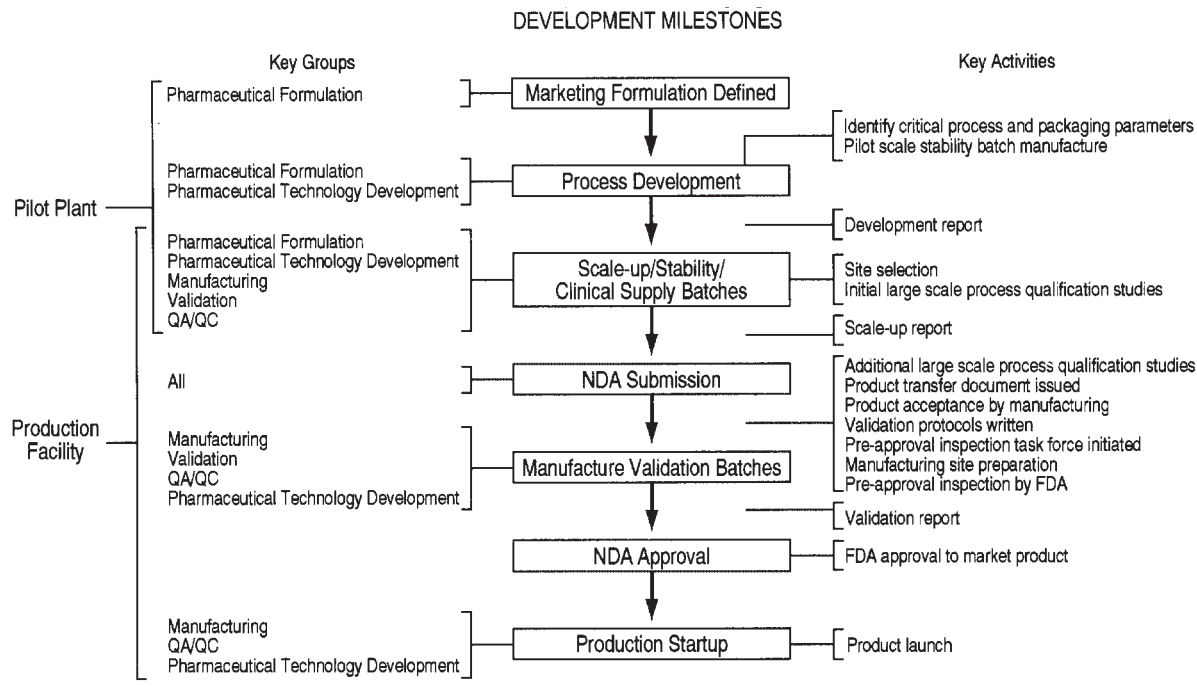


Fig. 6 The relationship of key groups, activities, and development milestones typically experienced during the transfer of formulation and process technology from the pilot plant to the production facility.

development and carry through production start-up ensure a smooth transition of the product to the marketplace.^[26]

Precise documentation of manufacturing trials at both the pilot plant and the production facility, with comparison of the results, assists in the development of a rationale relating biobatches and stability batches manufactured at the pilot scale to full scale product manufactured at the production site. Establishing the link between pilot and production scale batches is often critical in obtaining a successful FDA PAI.

QA–QC

As the size and complexity of the pilot plant increases, it may become necessary to have dedicated staff in the QC and QA areas to ensure compliance with cGMP and other regulatory requirements.

QA

The FDA promulgated that when the drug development process reaches the stage where product is manufactured for human use, compliance with cGMP is required.^[27] The facility should be validated with respect to utilities, equipment, and procedures and meet necessary regulatory requirements. Activities of the pilot plant QA scientist may differ from those performed at the production facility. The QA scientist should establish the procedural systems required to operate the pilot plant in a cGMP-compliant manner during the drug development process, especially with respect to clinical supply manufacturing. The QA scientist responsible for pilot plant activities should have a firm grasp on the breadth and scope of new product development activities and how regulatory requirements should be applied in the development atmosphere of the pilot plant. It is recommended that QA personnel be included as part of the team responsible for new product development in order to maintain an awareness of specific project activities in the pilot plant and provide expertise when necessary. Other QA activities similar to those in the production environment include documentation control of the production process and QC laboratory, raw material control, in-process material control, addressing deviations, overseeing production cleaning and processing, assessing of product quality, and establishing training programs for the pilot plant.^[28] It is particularly important for smooth operation of the pilot plant that systems, procedures, and communication links be established in all these areas and that pilot plant staff be fully aware of and trained in each.

QC

The QC function responsible for activities in the pilot plant may be structured somewhat differently than that found in a production operation. Release testing of finished products and materials may be undertaken by the analytical chemistry group responsible for the development of methods for an NCE. Release testing of materials and products for clinical studies in which methods have been formally developed may be undertaken by a separate QC group. For larger, more complex pilot plants, it becomes necessary to have a pilot plant QC function dedicated to the facility and responsible for physical, chemical, and microbiological testing of finished clinical products, components, and raw materials; physical, chemical and microbiological environmental sampling, testing, and trending within the pilot plant, e.g., water, compressed gases, air; and testing for validation and revalidation programs. It is also advantageous to have the QC team perform in-process testing during development, scale-up, and technology transfer activities. The QC personnel may become involved with project teams responsible for new product development and lend their expertise when required.

SUMMARY

The pilot plant is designed with basic strategic objectives in mind. As the facility becomes larger and more complex, organizational support is required from a number of technical disciplines. Although many activities are similar to those required in a production facility, it is imperative to recognize that basic differences exist between the research and development (R&D) environment and the manufacturing environment. Organization of personnel and definition of activities revolve around the development of new compounds and products rather than routine manufacturing. Successful operation of the pilot plant requires that these differences be taken into account, structured, and managed toward efficient new product development while meeting the rigorous scientific and regulatory standards required for the pharmaceutical industry.

ACKNOWLEDGMENTS

The second edition authors would like to sincerely acknowledge the work of the first edition authors—Leonard A. Amico, Ralph B. Caricofe, James D. English, Gary W. Goodson, Lawrence D. Lewis, and Robert

M. Franz—and those who provided critical reviews and helpful input for the first edition of this chapter—Bob Copeland, Bill Stagner, Phil Keicher, Ray Starrett, Mike Jozwiakowski, Linda Parks, Ken Korndoerfer, June Mullaney, Dave Moyer, and Anh Nguyen. Most of the first edition's text is retained. The basic goal in revising the first edition chapter was to add current references. This was easily done with the excellent foundation and framework the first edition authors provided.

Next, the authors would like to acknowledge Alan F. Parr for reviewing the revised chapter.

Finally, the authors would like to acknowledge the administrative help of Cindy Rozier in preparing this revision.

REFERENCES

1. U.S. Food and Drug Administration, *Guideline on General Principles of Process Validation*; U.S. FDA: Rockville, MD, May 1987.
2. Berry, I.R.; Nash, R.A. *Pharmaceutical Process Validation*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1993.
3. Cloud, P. *Pharmaceutical Equipment Validation—The Ultimate Qualification Guidebook*; Interpharm Press, Inc.: Buffalo Grove, IL, 1998.
4. Carleton, F.J.; Agalloco, J.P. *Validation of Pharmaceutical Processes, Sterile Products*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1999.
5. Sofer, G.; Zabriskie, D.W. *Biopharmaceutical Process Validation*; Marcel Dekker, Inc.: New York, 2000.
6. deSpautz, J.F. *Automation and Validation of Information in Pharmaceutical Processing*; Marcel Dekker, Inc.: New York, 1998.
7. U.S. Food and Drug Administration, *Guideline for the Preparation of Investigational New Drug Products (Human and Animal)*; U.S. FDA: Rockville, MD, March 1991.
8. U.S. Food and Drug Administration, *Guide to Inspection of Validation of Cleaning Processes*; U.S. FDA: Rockville, MD, July 1993.
9. LeBlanc, D.A. *Validated Cleaning Technologies for Pharmaceutical Manufacturing*; Interpharm Press, Inc.: Englewood, CO, 2000.
10. PDA Biotechnology Cleaning Validation Subcommittee, *Cleaning and Cleaning Validation: A Biotechnology Perspective*; PDA: Bethesda, MD, 1996.
11. Akers, M.; Bishara, R.H.; Barber, P.; Hatgas, S.G.; Meyer, D.K.; Wachter, R.O. CGMP education and instruction: a corporate approach to employee training worldwide. *Pharm. Technol.* **1993**, 17, 51–60.
12. Tetzlaff, R.J. In *FDA Inspection of GMP Training Programs*, GMP Training and Education Association's Biannual Conference, Baltimore, MD, Oct 2–4, 1991.
13. Moreland, R.L.; Myaskovsky, L. Exploring the performance benefits of group training: transactive memory or improved communication? *Org. Behav. Hum.* **Dec 2000**, 82 (1), 117–133.
14. Wright, E.E. Making the multimedia decision on strategies for success. *J. Inst. Del. Syst.* **1993**, 7, 15–22.
15. Moss, S.H.; Redfern, P.H.; Brown, K.N. Opportunities for the use of computer-based media in pharmaceutical training and education. *Pharm. Sci. Technol.* **1998**, 1 (2), 54–61.
16. *InTouch Man–Machine Interface Application General or Development System User Guide and Reference Manual*; Rev. B Wonderware Corporation: Irvine, CA, September, 1993.
17. Jackson, C. *The Practical Vibration Primer*; Gulf Publishing Co.: Houston, 1979.
18. Willig, S.H. *Good Manufacturing Practices for Pharmaceuticals*; Marcel Dekker, Inc.: New York, 2001; 225–226.
19. Willig, S.H. *Good Manufacturing Practices for Pharmaceuticals*; Marcel Dekker, Inc.: New York, 2001; 83.
20. Willig, S.H. *Good Manufacturing Practices for Pharmaceuticals*; Marcel Dekker, Inc.: New York, 2001; 214–215.
21. Plossl, G.W. The Production and Inventory Control Environment. In *Production and Inventory Control Handbook*, 3rd Ed.; Green, J.H. Ed.; McGraw-Hill: New York, 1997; 2.3–2.33.
22. Willig, Sidney H. *Good Manufacturing Practices for Pharmaceuticals*, 5th Ed.; Marcel Dekker, Inc.: New York, 2001; 99–138.
23. U.S. Food and Drug Administration, *Guideline on Sterile Drug Products Produced by Aseptic Processing*; U.S. FDA: Rockville, MD, June 1987.
24. De Vecchi, F. Validation of Environmental Control Systems Used in Parenteral Facilities. In *Validation of Pharmaceutical Processes*; Carleton, F.J., Agalloco, J.P. Eds.; Marcel Dekker, Inc. New York, 1999; 228–229.
25. Willig, Sidney H. *Good Manufacturing Practices for Pharmaceuticals*, 5th Ed.; Marcel Dekker, Inc.: New York, 2001; 367–371.
26. Goldberger, Francis. *Pharmaceutical Manufacturing, Quality Management in the Industry*; Herissey Press: Evreux, France, 1991; 429–445.
27. U.S. Food and Drug Administration, *Guideline for the Preparation of Investigational New Drug Products (Human and Animal)*; U.S. FDA: Rockville, MD, March 1991.
28. Peine, I.C. *Quality Assurance Compliance Procedures for Pharmaceutical and Biotechnology Manufacturers*; Interpharm Press, Inc.: Buffalo Grove, IL, 1994.

PLANTS AS DRUGS

Christine K. O'Neil

Duquesne University, Pittsburgh, Pennsylvania, U.S.A.

Charles W. Fetrow

St. Francis Medical Center, Pittsburgh, Pennsylvania, U.S.A.

HISTORY OF PLANTS AS DRUGS

Phytomedicine, the use of plants or plant parts to evoke a therapeutic cure or to treat an ailment, has been part of humankind's attempt to free itself of disease for several thousand years. Some references suggest that Neanderthals may have been one of the first phytomedicinal practitioners. Archeologists exhuming relics from lake beds in the Middle East have found evidence suggesting that early human's carried plant parts for more than just food or clothing. Some of the earliest writings found in Babylonian clay tablets from 3000 B.C. are about plants used for ceremonies, magic, and medicine. During the next thousand years, parallel cultures in China, India, and Egypt developed written records of medicinal herbs. Among other early historical documentation, the ancient Middle Easterners appear to have been the one of the first to rigorously document the use of plants for various diseases, compiling the first known pharmacopoeia, entitled the *Materia Medica*. The Greek historian Herodotus recounts how the Egyptians worshiped certain plants, believing that the some plants held the secret to a healthy life and longevity. Not to be outdone by the Egyptian peoples, the Greeks incorporated various plants and flowers into various aspects of Greek mythology. One of the largest compilations, dating approximately 600 B.C., is credited to a series of Chinese emperors and provides detailed instructions about the use, benefits, and preparation of herbs.

Scholars throughout the centuries have made valiant attempts to demystify herbs; however, people continued to hold deep beliefs about the significance of plants. The Doctrine of Signatures, dating from the first century A.D., suggests that some aspect of the plant's appearance provided clues to its medicinal properties. This belief remained popular for 15 centuries and is noted in many Asian and Western cultures, including that of the Native Americans.

In the beginning, these primitive medicinals were used primarily in their natural form and incorporated into compresses and poultices. Much of the knowledge obtained from these first pharmaceuticals was put to use

by Native Americans and early American colonists. Not surprisingly, plants were manipulated into somewhat more complex formulations known as decoctions and infusions, the fundamental herbal preparations resembling modern-day teas. Typically, these botanical remedies were given to folks suffering from illness without regard to prior investigation. Formulas for various ailments were passed down from household to household much like recipes in a cookbook.

As science emerged after the 17th century, plants were classified and demystified. New technical skills permitted analysis of the plant's components so that standardized tinctures and extract could be prepared. Extraction of the relevant chemicals from these plants became popular around the turn of the 19th century. Active principal components such as opium and digitalis were isolated and applied therapeutically, although still lacking much in the way of formal prior investigation. As science advanced, medicines were synthesized and herbalism declined. Newly developed principles of organic chemistry made it possible to replicate plant-produced chemicals, paving the way for creative manipulation of these molecular entities, leading to the synthesis of new compounds that preserved the beneficial properties of the natural chemical, but minimized its toxic effects. The conception of local anesthetic agents from the naturally occurring alkaloid cocaine and the creation of the aspirin from natural salicylic acid in willow bark are important hallmarks that characterize the beginning of the era of "allopathic" medicines.

Many medicines that we use today were isolated from plant sources (Table 1). Examination of today's allopathic medications reveals that approximately 25–33% of currently available modern medicines in the United States have their origin in plants, animal, or mineral systems. These include aspirin, digoxin, quinine, colchicine, and vinca alkaloids, to name a few.

The focus on synthesized and biotechnolgy medicines has continued. However, in the latter part of the 20th century, there has been an intense renewed interest in

Table 1 Some contemporary pharmaceuticals of plant origin

Modern drug name	Therapeutic indications	Botanical name	Common plant name
Capsaicin	Topical analgesic	<i>Capsicum annum</i>	Red pepper plant
Cascara sagrada	Laxative	<i>Rhamnus purshiana</i>	Cascara sagrada
Colchicine	Gout; anti-inflammatory	<i>Colchicum autumnale</i>	Autumn crocus
Digoxin	Cardiac inotroph for heart failure/arrhythmia	<i>Digitalis purpurea</i>	Foxglove
Ephedrine and pseudoephedrine	Sympathomimetic/decongestant	<i>Ephedra sinica</i>	Mahuang
Methysergide	Vasoconstrictor for headaches	<i>Claviceps purpurea</i>	Ergot
Opiates (i.e., morphine, codeine)	Narcotic analgesics	<i>Papaver somniferum</i>	Poppy
Pilocarpine	Cholinergic agent used for glaucoma	<i>Pilocarpus jaborandi</i>	Jaborandi tree
Podophyllum	Antimitotic for venereal warts	<i>Podophyllum peltatum</i>	Mayapple
Quinidine and quinine	Antiarrhythmic/antimalarial agent	<i>Cinchona pubescens</i>	Quinine tree
Reserpine	Antidepressant	<i>Rauvolfia serpentina</i>	Rauwolfia
Senna concentrate	Laxative	<i>Cassia senna</i>	Senna
Taxol	Chemotherapeutic drug	<i>Taxus brevifolia</i>	Pacific yew
Vincristine and vinblastine	Chemotherapeutic drug	<i>Catharanthus roseus</i>	Madagascar periwinkle
Warfarin	Blood thinner for clots	<i>Melilotus officinalis</i>	Sweet clover
Yohimbine	Treatment of impotence	<i>Pausinystalia johimbe</i>	Yohimbe

herbalism. The United States, however, appears to be one of the last countries to embrace this practice of phytomedicine. The practice of phytomedicine has remained primarily outside the mainstream of contemporary American medical practice. Germany, the Orient, and several European countries (Italy, Spain, the Netherlands, Belgium, and others) have taken a more aggressive approach. Eighty percent of the world's population reportedly uses herbs for medicinal purposes (1). One of the most famous countries for this is Germany, which, by recent estimates, has allowed some 600–700 herbal products to be marketed in that country. Approximately 70% of German physicians now prescribe phytopharmaceuticals to their patients, which serves to maintain one of the world's largest markets for herbal drugs (2). Furthermore, the Commission E, a branch of German government somewhat similar to the U.S. Food and Drug Administration (FDA), has compiled therapeutic monographs for several hundred herbal medicines that discuss their general safety and efficacy.

Despite a sluggish start, the commercialized herbal industry is now blossoming in the United States. More than 500 different herbs are currently marketed in the United States, responsible for over 3.2 billion in sales in 1997(3). Limited regulatory guidelines and direct-to-consumer advertising have created a booming herbal market. Products are readily available in health food stores, supermarkets, and pharmacies, through mail-order catalogs, and via the Internet. Safety and efficacy data for these products are extremely limited. Varro E. Tyler, a well known author and researcher once said, "More

misinformation regarding the efficacy of herbs is currently being placed before consumers than at any previous time, including the turn-of-the century heyday of patent medicines." With the exceptions of the German Commission E monographs and some emerging studies, a critical evaluation of these entities is lacking. This presents both opportunities for research and challenges to modern science, medicine, and pharmacy.

TERMINOLOGY

Herbal products are considered a type of alternative medicine (e.g., herbal medicines, Chinese herbs, homeopathy, acupuncture, biofeedback, color therapy, music therapy, hypnotherapy, aromatherapy, Ayurvedic medicine, massage, therapeutic touch, Bach flower remedies, chiropractic, reflexology, naturopathy, and more). According to the Office of Alternative Medicine of the National Institutes of Health (NIH), characteristics of alternatives medicine include treatments that lack sufficient documentation in the United States for safety and effectiveness against specific disease and conditions and are not generally taught in U.S. medical schools or reimbursable by health insurance providers. Although uncertainty exists about the safety and efficacy of herbal products, there is even confusion about the terms used to describe such products. Some products are not herbs (i.e., saw palmetto is a tree) or botanicals (melatonin, glucosamine). Thus, further clarification is necessary.

Herbs are specifically defined as nonwoody, low-growing plants such as basil and parsley. Herbal medicine is considered to be the use of crude drugs of plant origin to treat illness or to promote health. A more correct term for this would be botanical medicine. Phytomedicinals are those common preparations, including capsules, tablets, tinctures, and fluidextracts that have been prepared from plant sources. This should be distinguished from plant-derived drugs that have been isolated, purified, and standardized from plant sources.

The words “natural” and “organic” are quite appealing to the consumer. Synthetic is often considered less desirable. Many believe that natural is better, safer, or not foreign to the body, but quite the opposite may exist. Native is identical to what is produced by or present in the body. Natural products refers to substances that are used to promote health or treat illness derived from plant, mineral, or animal sources. Organic refers to the level of pesticides or chemicals used in the growing process. For example, insulin from pork or beef sources is natural, but not native. Recombinant insulin is synthetic, but native. Thus, synthetic, as in synthetic insulin or estrogen, does not necessarily mean foreign or less desirable.

Two other broad terms that are used quite frequently are nutraceuticals and dietary supplements. Nutraceuticals include food, dietary supplements, and medical or functional foods that have a health or medical benefit including the prevention or treatment of disease. The newest term, introduced by the Dietary Supplement Health and Education Act of 1994 (DSHEA) is dietary supplement. A dietary supplement is neither a food nor a drug, according to the FDA. This term encompasses vitamins, minerals, herbs or other botanicals, amino acids, and any other dietary substance for use by humans to supplement the diet and promote health.

TRENDS IN HERBAL USE

The prevalence of alternative medicine use in the United States is steadily increasing. One may even describe the phenomenon as an explosion. A landmark survey published in 1993 estimated that 33.8% of Americans used one type of alternative therapy (4). A follow-up survey released in 1997 reported that frequency to be 42.1%, with 12.1% taking herbal medicine (5). In over 50% of the cases, use of alternative medicine was not supervised by the primary medical physician, and only 38.5% of consumers reported such use to their physicians.

Why do individuals seek alternative medicine? Many individuals are seeking health promotion and disease

prevention and believe that the “natural” way may be the best. For many people, conventional therapies may not be available or ineffective or may carry significant risk that the user may not be willing to accept. Some have tried several conventional therapies without relief and look to alternative medicine as the only remaining option. Additionally, extensive direct to consumer advertising and limited regulatory oversight have fueled the expansion of the alternative medicine industry.

Recent surveys have revealed some characteristics of alternative medicine consumers (5). Approximately 50% of consumers are between the ages of 35 and 49 and have a college-level education. Income levels exceeded \$50,000 for 48.1% of alternative medicine consumers. This may not be surprising because most insurance companies do not reimburse for such products. There appears to be a regional difference with use of alternative medicine, with over 50% of consumers living in the western portion of the United States. An earlier survey revealed that alternative medicine consumers were more likely to have one or more health conditions and less likely to be enrolled in an HMO. Furthermore, they had twice as many visits to traditional medicine providers and had a higher level of unmet medical need.

REGULATORY ISSUES REGARDING HERBALS

A major contention of the herbal medicine advocacy is the notion that because these entities are natural products, they are somehow safer and better for human consumption. However, the vast majority of alternative medicine products are essentially unregulated and not yet required to demonstrate efficacy, safety, or quality before becoming commercially available (6, 7). Currently, in the United States, there exists continued debate on what role the FDA should have in regulating and approving alternative medicines (7, 8).

The regulatory status of herbal medicine has changed over the past century. At one time, the United States Pharmacopeia, (USP), 1st Edition contained mostly herbal medicines. The first attempt by the U.S. government to regulate any “medicine” was the Food and Drugs Act of 1906, which simply prohibited the adulteration and/or misbranding of drugs. The act focused primarily on the quality of products being marketed but neglected the safety and efficacy of the medicines themselves. The Food and Drugs Act itself arose from public pressure imposed on the government after a series of fraudulent incidents involving patent medicine manufacturers and meat-packing firms were exposed and widely publicized.

The FDA was established by Congress in 1928, but it had been granted little authority and even less guidance with regard to how to proceed. The issue of product safety was finally addressed in the late 1930s after Elixir of Sulfanilamide contributed to the deaths of more than 100 people. With the 1938 Food, Drug and Cosmetic Act (FDCA), drugs were required to demonstrate safety before marketing. The FDCA defined drugs as substances, other than foods, that are recognized in *USP/National Formulary* and are intended to treat or prevent disease or affect body structure. In effect, herbs not included in the *USP/NF* were now considered food substances. Herbs that had been in the *USP/NF* were viewed as exempt and held official drug status until more recently.

As a result of Kefauver–Harris Amendment in 1962, drugs were required to demonstrate efficacy before marketing. The FDA presently regulates the pharmaceutical industry by requiring new product manufacturers to file a New Drug Application (NDA) for each new entity, which must include scientifically sound laboratory and clinical trials that demonstrate a drug product's safety and efficacy. Herbals and other products that lacked safety and efficacy data were considered over-the-counter agents. In 1972, panels were formed to evaluate the active components of these over-the-counter agents. The results of this effort were released in 1990. Products were classified as category I, II, or III. Category I products were generally recognized as safe and effective and were not misbranded. Category II and III substances were considered to be unsafe, ineffective, misbranded, or lacking sufficient data (7). Only a few herbals, such as senna, earned category I status. Because it was not cost-effective for most companies to present data, most substances now were considered foods or food additives that could not present labeling claims.

Although most herbals packaging could not be labeled with therapeutic claims, products continued to be sold with readily available literature, pamphlets, and advertisements touting the benefits of these substances. The Nutrition Labeling and Education Act of 1990 sought to improve labeling and education about food products and dietary supplements. At one point, there was even some consideration given to removing such products from the market. Concern over the ability to market herbal products led to extensive lobbying in Congress, which resulted in the Dietary Supplement, Health and Education Act of 1994 (DSHEA).

The DSHEA classified herbal products as dietary supplements intended to supplement the diet. Also included in the definition of dietary supplement were vitamins, minerals, botanicals, amino acids, and other

substances intended to supplement the diet. Dietary supplements were not considered foods or drugs and therefore were exempt for FDA oversight and the pre-market approval process. For all products introduced before October 1994, the burden of proof to demonstrate safety was now in the hands of the FDA and not in the hands of the manufacturer. Products introduced after October 1994 must be proven safe by the manufacturer. Manufacturers are currently not required to submit safety and efficacy data, and there are no good manufacturing standards (GMP) in place. However, the FDA does have the authority to establish GMP standards, and they are currently under development. Preliminary proposals indicate these GMP standards will more likely reflect GMP for food rather than for drugs.

The Office of Alternative Medicines (OAM) was also established at this time for the study and compilation of data on alternative medicines. The OAM functions within the NIH, and it is hoped that the OAM can further clarify the role of alternative medicines such as herbal remedies in this country. In other countries, similar committees (such as Commission E in Germany) have reviewed the safety and efficacy of herbs and published the results so that product debates can be resolved.

Although there is limited information regarding the therapeutic efficacy and standards for herbals some resources do exist. The most widely recognized source is the German E Commission E monographs, which contain information on more than 300 herbal products. Other well referenced resources include the *American Herbal Pharmacopeia*, the *British Herbal Pharmacopeia* (BHP), the *British Herbal Compendium*, the European Scientific Cooperative for Phytomedicines (ESCOP), the *U.S. Pharmacopeia*, and the World Health Organization (WHO).

At this time, consumers cannot be assured of purchasing a product that meets any regulatory standard, despite what may be listed on the label. Because they are derived from plant sources, the ability to standardize herbal content from lot to lot is confounded by many variables. Growing conditions, storage, harvesting, preparation, and processing all may affect the quality of the final product. Additionally, the actual active ingredient of many of these plant products may be unknown.

In review, alternative medicines have been exempted from demonstrating efficacy (Kefauver-Harris Amendment) by being classified as dietary supplements (Dietary Supplement Health and Education Act). This means that herbal products cannot make therapeutic claims on the label or package but may distribute third-party information regarding therapeutic claims, as long as this information is not misleading or product-specific.

LABELING ISSUES REGARDING HERBALS

Although the FDA lists many herbs as safe, herbal manufacturers cannot legally claim therapeutic efficacy of their product for a disease state without the evidence to support this claim. Dietary supplements may claim effects on structure and function of the body and may have authorized health claims. They cannot be labeled as intending to treat, prevent, mitigate, cure, or diagnose any disease. Dietary supplements may not be represented as food and cannot use the term “significant scientific agreement” when making a claim. Supplement manufacturers are required to have data on file to substantiate any “structure/function” claims that the manufacturer makes about the product. Examples of structure/function claims might include statements such as “helps you relax,” “supports prostate health,” “for liver maintenance,” or “helps promote cardiovascular health.” Claims that would not be allowed include such statements as “lowers cholesterol,” “improves benign prostatic hypertrophy,” or “treats immunodeficiency.” The FDA may ban dietary supplements if they pose an imminent hazard and may ban claims that are inherently misleading. Although the FDA regulates labeling, advertising of dietary supplements fall under the purview of the Federal Trade Commission. The FDA has suggested that manufacturers not make any claims with respect to their products and use in pregnancy. Additional rulings from the FDA are expected in the near future with respect to dietary supplements and their use in pregnancy.

THE MAKING OF A CRUDE DRUG

Approximately 250,000 species of flowering plants exist in the world today. Only a small percentage has been studied adequately for pharmacological activity. Based on past performance, it is reasonable to suggest that many valuable agents await being discovered in plants that we have not yet screened for therapeutic potential. Anecdotal reports of therapeutic efficacy and local folk medicine lore are a means to identify potentially valuable plant entities.

The term “pharmacognosy” refers to the study of chemicals from “natural” sources for medicinal application. Although this frequently refers to the study of chemical entities in higher plants, lower plants (fungi, molds, and yeast); animals; marine animals; fish; insects; and minerals are also fair game for evaluation of potential medicinal agents. In most industrialized nations, crude drugs (natural substances that are only collected and dried before manufacturing) are seldom used as the chief

therapeutic agent. Usually, important chemical constituents of the plant are identified, removed, derived, or modified and applied therapeutically in a consistent pharmaceutical vehicle in an effort to enhance the “natural” benefits and reduce any inherent risk.

Many plants can be grown in similar climates that resemble the plant’s native land. Compatibility of the plant to a particular region and the cost of harvesting the plant in that same area are two criteria that determine the availability of crude drugs in a particular country. Additionally, national and international restrictions on the collection of wild plants tend to limit availability of plant resources and escalate cost of production. These factors also force countries to specialize in producing only certain types of phytomedicinal resources.

Despite the tendency for a country to produce only certain phytosubstrates, the quality of crude drugs is often suspect. By the time a crude drug arrives at the manufacturing process, it has been exposed to various opportunities for adulteration, deterioration, and contamination.

During the collection or harvesting phase, skilled labor can be an important factor that influences quality of the harvest of wild plants. Deliberate adulteration most frequently occurs with expensive natural substances or natural substances in short supply. Mechanical devices may be more economical in the collection of some plants but may be of little use when only particular parts of another plant are desired. Cultivated plants ensure a more reliable source of the desired plant with less risk of substitution. Environmental conditions (temperature, rainfall, day length and sunlight, altitude, atmosphere, and soil) can dramatically affect quality, concentration, and presence of active constituents in plants. The age of the plant can also be an important determinant of the quality of active constituents in a crude drug.

At this stage, before drying, any insect-infested or disease-infested part of the plant should be removed. “Garbling” is the term for this semiskilled process of removing unwanted material (dirt, debris, and unnecessary plant parts) from the plant before drying and again before packaging and storage.

Drying can take from a few hours to a few weeks. This depends on the relative humidity of the local climate and the physical nature of the plant constituents. Drying by artificial heat (hot water pipes, stoves, and continuous belt dryers) carries the advantage of shorter drying times. Veterans of the process have learned when to stop the drying phase to prevent plant parts from becoming too brittle and overdried.

Deterioration of the dried product can occur as moisture from the surrounding air (usually approximately 10–15%)

and light return to the plant after drying. Return of moisture to the product provides a more favorable environment for contamination from molds, bacteria, and insects. Some processes introduce sterilization as a way to minimize microbial contamination. Proper storage and preservation must take place to ensure that quality of the product is maintained until delivery to the manufacturing facility where the crude drug undergoes a variety of grinding, crushing, extraction, and distillation procedures before being formed into an herbal pharmaceutical.

STANDARDIZATION OF DIETARY SUPPLEMENTS

Official standards are absolutely necessary to ensure the quality, reliability, and homogeneity of herbal products for consumers. Standardized products are paramount to those in healthcare planning to conduct clinical research with these products. Independent laboratories and university-affiliated research reports have documented the considerable variation that exists in terms of quality and reliability in these products. Abroad, the ESCOP, composed of manufacturers of herbal medicines and herbal associations, is working with European research groups to develop quality-control standards for the production of natural products. This committee is developing monographs for incorporation into such references as the *British Herbal Pharmacopoeia* and the *British Herbal Compendium*.

Likewise, in 1995, the USP commissioned an advisory panel on natural products whose mission was to establish standards and develop information concerning herbal or “dietary” supplements. Supplement monographs created by this endeavor address various issues associated with the standardization of individual herbals. The following list of section headings outlines the information found in each of the monographs:

1. Title—identifies the most commonly accepted name of the entity.
2. Definition—describes plant parts used, genus, species, authority, and family of the botanical.
3. Packaging and storage—cites appropriate packaging and storing conditions designed to promote integrity of the product.
4. Labeling—states requirements for label nomenclature.
5. Reference standards—identifies appropriate reference standards.
6. Botanic characteristics—describes visible and microscopic shape and structure characteristics of the whole plant or plant parts.

7. Identification—describes pharmacognostic tests useful in the identification of the entity.
8. Total ash—sets limits for the amount of inorganic residue remaining after incineration.
9. Acid-insoluble ash—sets limits for the amount of foreign inorganic residue remaining after boiling the total ash with 3N hydrochloric acid (an indication of how much dirt and soil remain in sample).
10. Water-soluble ash—sets limits for the residue remaining after boiling the total ash with water.
11. Foreign organic matter—limits the amount of nondrug-containing matter.
12. Loss on drying—sets criteria for loss limits of water, volatile oils, or other volatile chemical compounds.
13. Water content—limits variation in the water content of dried botanicals.
14. Alcohol-soluble/water-soluble extractives—sets thresholds for minimum acceptable amount of aqueous-, alcohol-, or aqueous alcohol-soluble extractives.
15. Volatile oil—describes the quantity of volatile oil present in the botanical.
16. Heavy metals—limits heavy metals present in the botanical.
17. Pesticide residue—sets strict limits of pesticide content.
18. Microbial limits—sets limits of total bacteria and mold count.
19. Marker substances and content tests—establish standards for quantitative chemical analysis of botanical for the presence of certain marker substances that aid in proper identification.

RISKS AND REALITIES OF HERBAL MEDICINES

Herbal drugs have many unknown and undocumented risks, side effects, and drug interactions. Like contemporary, rigorously tested pharmaceuticals, herbal medicines have some risk associated with their consumption. The fact that a plant is completely “natural” does not necessarily make that plant entirely risk-free. Several plants, when consumed in their most natural form, can cause grave illness or even death to humans and animals. A partial list of some of these natural herbal agents with the potential to harm is listed in Table 2. Herbalists, scientists, and the general public routinely avoid many of the plants listed in Table 2 because of their impending risks. However, hundreds of additional herbs and alternative medicines exist, much of their toxicologic profiles untested and undescribed. Adverse reactions with

Table 2 Examples of plants with the potential to harm

Common name	Botanical name	Reported toxic events
American mistletoe	<i>Phoradendron flavescens</i>	Hypertension and hypertensive crisis
American yew	<i>Taxus canadensis</i>	Cytotoxic
Arnica	<i>Arnica montana</i>	Violent gastroenteritis, nervous disorders, muscle weakness, collapse
Autumn crocus	<i>Colchicum autumnale</i>	GI toxicity, vomiting, neurologic toxicity, kidney failure
Belladonna	<i>Atropa beladonna</i>	Anticholinergic toxicity
Betal palm	<i>Areca catechu</i>	Teratogen
Bird's foot trefoil	<i>Lotus corniculatus</i>	Cyanide poisoning, convulsions, paralysis, coma, death
Bittersweet nightshade	<i>Solanum dulcamara</i>	Cardiac toxicity
Black nightshade	<i>Solanum americanum</i>	Cardiac toxicity
Black locust	<i>Robinia pseudoacacia</i>	Bradycardia, nausea, vomiting, dizziness
Bloodroot	<i>Sanquinaria canadensis</i>	Destroys tissue on application
Blue flag	<i>Iris versicolor</i>	Nausea, vomiting, and diarrhea
Broom	<i>Cytisus scoparius</i>	Diarrhea, GI toxicity/dehydration
Calabar bean	<i>Physostigma venenosum</i>	Cholinergic toxicity
Castor oil plant	<i>Ricinus communis</i>	GI toxicity/dehydration
Celandine	<i>Chelidonium majus</i>	
Chapparral	<i>Larrea tridentata</i>	Fulminant hepatic failure
Chinese lantern	<i>Physalis alkekengi</i>	
Comfrey	<i>Symphytum officinale</i>	Hepatotoxicity
Cotton	<i>Gossypium hirsutum</i>	Hypokalemia, male sterility, heart failure at high doses
Daffodil	<i>Narcissus pseudonarcissus</i>	CNS depression, miotic, coma, salivation, vomiting, death
Death camas	<i>Zigadenus elegans</i>	
Desert plume	<i>Stanleya pinnata</i>	
Ergot	<i>Claviceps purpurea</i>	Hallucinations, hypertension, tissue ischemia, St. Anthony's fire
Figwort	<i>Scrophularia nodosa</i>	
Foxglove	<i>Digitalis purpurea</i>	Bradycardia, heart block, arrhythmia
Goldenseal	<i>Hydrastis canadensis</i>	Hyperreflexia, hypertension, convulsions, respiratory failure
Green false hellebore	<i>Veratrum viride</i>	
Heliotrope	<i>Heliotropium europaeum</i>	Hepatotoxicity
Hedge mustard	<i>Sisymbrium officinale</i>	Cardiac toxicity, heart failure
Hemp dogbane	<i>Apocynum cannabinum</i>	Cardiac stimulant, arrhythmias
Henbane	<i>Hyoscyamus niger</i>	Anticholinergic toxicity
Horse chestnut	<i>Aesculus hippocastanum</i>	Bleeding
Indian pink	<i>Spigelia marilandica</i>	Overdoses have caused fatalities
Indian tobacco	<i>Lobelia inflata</i>	Vomiting, paralysis, hypothermia, collapse, coma, death
Jalap root	<i>Exagonium purga</i>	Dramatic purgative cathartic
Jimsonweed	<i>Datura stramonium</i>	Anticholinergic toxicity, hallucinations
Larkspur	<i>Delphinium ajacis</i>	
Life root	<i>Senecio longilobus</i>	Hepatic failure from hepatic veno-occlusive disease
Lily of the valley	<i>Convallaria majalis</i>	Cardiac toxicity
Marsh marigold	<i>Caltha palustris</i>	Second most common killer of livestock (accidental ingestion)
Mayapple	<i>Podophyllum peltatum</i>	Severe GI irritation
Monkshood, wolfsbane	<i>Aconitum</i> spp.	Cardiotoxicity, neurotoxicity, hypotension, arrhythmias
Moonseed	<i>Menispermum canadense</i>	Tachycardia, severe vomiting, purging
Morning glory	<i>Ipomoea purpurea</i>	Potential for hallucinations/psychosis
Mountain laurel	<i>Kalmia latifolia</i>	
Periwinkle	<i>Vinca major, vinca minor</i>	Cytotoxicity, renal failure, hepatic failure, neurological damage
Poison hemlock	<i>Conium maculatum</i>	Birth defects, "crooked calf disease"
Pokeweed	<i>Phytolacca americana</i>	
Queen's delight	<i>Stillingia sylvatica</i>	GI toxicity, mutagenic
Red baneberry	<i>Actea rubra</i>	

(Continued)

Table 2 Examples of plants with the potential to harm (*Continued*)

Common name	Botanical name	Reported toxic events
Rosebay rhododendron	<i>Rhododendron maximum</i>	
Strychnine tree	<i>Strychnos nux-vomica</i>	CNS stimulation leading to seizures and cardiac arrest
Tall buttercup	<i>Ranunculus acris</i>	
Wallflower	<i>Cheiranthus cheiri</i>	Cardiac toxicity, heart failure, bradycardia
White false hellebore	<i>Veratrum album</i>	
Wild cherry	<i>Prunus virginiana</i>	Dyspnea, vertigo, convulsions
Wild licorice	<i>Glycyrrhiza lepidota</i>	Hypotension, hypernatremia, hypertension, muscle weakness
Wintercress	<i>Barbarea vulgaris</i>	Renal damage in animals
Wormseed	<i>Chenopodium ambrosioides</i>	Nausea, dizziness, convulsions, paralysis
Wormwood	<i>Artemisia absinthium</i>	Rhabdomyolysis, renal failure, seizures, "absinthism," mental doze
Yellow jessamine	<i>Gelsemium sempervirens</i>	Paralysis and death

herbal medicines can be directly related to exposure of chemical components of the plant (intrinsic) or to inappropriate or incorrect manufacturing/production procedures during preparation of a dietary or herbal supplement (extrinsic)(9). Drug interactions, toxicity from high concentrations, or undesirable effects from simple ingestion of the chemicals of the herbal supplement are considered typical intrinsic misadventures.

Although laws do not require reporting of adverse events of dietary supplements to the FDA at this time, many untoward events have been documented. The FDA has created MEDWATCH [(800) FDA-1088] to receive and compile reports of adverse reactions from pharmaceuticals. Healthcare providers are encouraged to report to this organization any and all adverse events related to dietary supplements and their consumption.

A review of adverse events related to herbal medicines reported in the medical literature from 1992 to 1996 has been compiled (10). This report highlights cases of hypersensitivity reactions, hepatotoxic reactions, and various types of renal damage associated with various herbal products. Some Chinese herbal preparations appear to be notorious for causing nephropathy (11–13). One of the more infamous adverse events related to the consumption of a dietary supplement was associated with the amino-acid L-tryptophan, touted for its ability to reduce pain and promote sleep. During the late 1980s, it was discovered that excipients or tablet "fillers" contained in a few L-tryptophan products caused a rare, but reportedly fatal, syndrome (eosinophilia-myalgia syndrome). Additionally, popular dietary supplements and weight-loss products containing ephedra alkaloids have become a recent cause of adverse effects (e.g., increased blood pressure, tremor, arrhythmia, seizure, stroke, heart attack, and death), as supported

by several hundred incident reports to the FDA between 1993 and 1997 (14). Furthermore, a study designed to assess the prevalence of herbal product use and its associated morbidity in a population of adults with asthma found an increase in the number of hospitalizations attributed to patients who self-medicated with herbal products and black coffee or tea (15). In a large retrospective study of admissions to a Taiwan hospital, 4% were found to be drug-related. Herbal medications ranked third as the drug category most responsible for adverse effects (16).

Herbal-Drug Interactions

Potential drug interactions are usually either unknown or unconsidered. Many herbs contain naturally occurring coumarins and cardiac glycosides, chemicals that potentiate hypoglycemia or hyperglycemia and promote sedation, or various other actions that can endanger the lives of patients taking prescription medications. Recently, two reports strongly suggest that a potential drug interaction exists between some drugs when used in combination with St. John's Wort (17, 18). It appears that St. John's Wort may act as an enzyme inducer of the cytochrome P-450 system. Subsequently, drugs that undergo metabolism by the CYP-450 3A isoenzyme must be used more cautiously when taken in combination with St. John's Wort. Not surprisingly, metabolism-oriented interactions are not the only type of drug interactions reported with herbal medicines. A case report of malabsorption of levothyroxine has been reported because of the ingestion of a combination herbal remedy (19). Some recent reviews cite potential risks related to herb-anesthesia interactions and other preoperative concerns with herbal medicines (20).

Propagation of Herbal Medicine Misinformation

Inadequate or inappropriate dissemination of information to the lay public, combined with less-than-stringent regulation, leads unwary consumers down the path of dangerous adverse events related to the consumption of herbal medicines. Successful dietary supplement and herbal manufacturers spend several million dollars each year on advertising. Existing regulations and product labeling of herbal supplements fail to provide ample warning of risks to consumers. In many cases, product advertising could be considered misleading, or at least questionable, despite FDA restrictions to limit manufacturer claims that relate only to proper health maintenance (10). The awesome marketing power of the Internet has by far the greatest potential to magnify and express the unknown dangers of herbs, if information regarding commercialization and marketing of alternative medicines is left unreviewed. Herbal products otherwise unobtainable in the United States have been purchased over the Internet and have been associated with substantial morbidity in uninformed consumers (21).

Misbranding, Substitution, Contamination, and Adulteration

Generally, dietary supplement and herbal medicine manufacturers are not required to meet the same standards and regulations that apply to the pharmaceutical superpowers.

In the spring of 1998, the FDA issued a warning regarding an herbal medicine called *Sleeping Buddha*. Apparently, the product contained an unlabeled sedative, the benzodiazepine estazolam. Fortunately, no misadventures with this product were reported. At about the same time, the FDA issued another warning regarding an herb called plantain. This plant contains digitalis-like glycosides and, therefore, consumption could be problematic for elderly patients or patients with cardiac disease. Potential dangers of this sort fall in the category of extrinsic misadventures. These are usually related to a lack of standardization, contamination, adulteration, or substitution of the products, in addition to either misidentification or misbranding of the supplement. Additionally, adulteration of Chinese herbs with mefenamic acid has been responsible for cases of herbal nephropathy (12). A more detailed discussion of misbranding, adulteration, and the like has already been published (9).

Many plants can vary considerably as to their quantities of active chemical constituents. This variability may depend on the time of year of harvest; age of the plant; method of pollination; and/or conditions of watering,

wind, weather, and soil. These variables can lead to considerable differences in the finished product after manufacturing.

Additionally, consumers are often exposed to a variety of compounds when they ingest herbal supplements. Supplement manufacturers do not always separate out the active chemical ingredients of the plant; subsequently consumers ingest many different chemicals that occur “naturally” in the plant along with the reportedly active constituent.

Additional concerns arise from the notable lack of stability testing of the product, because these data are usually not available on the label or otherwise. Various contaminants (i.e., heavy metals, aspirin, caffeine, theophylline, diuretics, corticosteroids, benzodiazepines, atropine, and others) have been discovered in reportedly “pure” herbal products (10).

Lack of Consensus on Usage and Monitoring

There exists an overwhelming lack of consensus as to how an herbal medicine should be ingested or applied therapeutically. Recommended doses vary considerably from source to source, even among noted advocates in the field. No standards exist for monitoring adverse events or effectiveness.

Delay in Time to Proven Therapy

The time an individual spends evaluating a desired response from an herbal supplement could be spent seeking professional medical advice and proven pharmacotherapy. In cases such as severe depression, extreme mental illness, or other life-threatening diseases, the delay in seeking professional help could be the difference between life and death.

Abuse of Herbal Medicines

Unlike their prescription-only counterparts, herbal medicines can be purchased and consumed freely by virtually anyone. This can take place without forethought or advice, without restriction or limitation, without even something as insignificant as confirming a person's age or identification. Yet, in stark contrast, we require that individuals be of a certain age before purchasing cigarettes or alcohol. Individuals wishing to purchase these chemicals must be able to demonstrate proof of the minimum age requirement by displaying a valid identification card. This is likely required of minors because society believes they are not mature enough and perhaps

lack “sufficient education” before the designated age to handle the potential dangers and responsibilities that accompany use of chemicals (cigarettes and alcohol) that can intoxicate, addict, or even cause physiological damage. Herbal medicines, like cigarettes and alcohol, are not without the potential for abuse, the potential to harm, and in some cases, the potential to addict or intoxicate. A similar corollary to the minor purchasing cigarettes or alcohol can be envisioned, analogous to a uninformed or misinformed consumer purchasing herbal medicines.

One such example is the opium poppy plant, which has been used to treat pain for centuries. Its derivatives continue to be used in modern medicine, yet its potential for abuse is also well documented. No one would argue that this plant should be regulated because of its potential for abuse. The same can be said of the leaves of the coca shrub, which have been chewed by the Incas and their descendants for countless generations to help them work in the high altitudes of the Andes Mountains. In many cultures, traditional healers and religious leaders have used psychoactive plants as part of their rituals, yet the perception remains that plants are safe and pose no danger for abuse or dependence. Many plants may have been safely used for hundreds of years in controlled settings such as religious rituals, but when used indiscriminately by the general public, they can lead to problems. A good example is the herbal anxiolytic kava, which is currently being promoted heavily. It has been used traditionally in Polynesia for hundreds of years without any reported adverse consequences, yet it is now known to increase the effects of alcohol and even by itself can lead to intoxicating effects (22–24). Its users continue to consume kava even after they develop adverse effects such as the skin reaction, which can occur from chronic use. Like many drugs with abuse potential, not everyone abuses a substance to the same degree. However, as its use becomes more prevalent, more and more cases of misuse will continue to appear. The abuse potential of many drugs is not identified immediately but rather after controlled clinical trials. Unfortunately, as noted above, herbal products do not have to undergo these rigorous trials before being marketed and thus problems can be identified only after many people experience an adverse effect such as dependence.

Cost

Although herbal medicines tend to be less expensive than their FDA-labeled counterparts, the cost can still be substantial if examined over several weeks to months. Costs related to delayed effective therapy, side effects, drug interactions, and hospital admissions might also be considered in this regard. As herbal medicine becomes

more popular and as herbal medicine manufacturers spend more and more on advertising, the margin of cost difference between the two types of drugs will diminish.

SPEAKING THE LANGUAGE—A BRIEF LOOK AT SOME “ALTERNATIVE” TERMINOLOGY

In this section, we define some terminology commonly used when discussing alternative medications. Throughout the Encyclopedia, we have attempted to replace this terminology with more readily recognizable medical terms. However, in the event that the reader comes across an unfamiliar term, we hope to have included the definition for that term here in this section. We have made some assumptions as to what we thought might be unknown by the average healthcare practitioner not skilled in alternative medicine. *The information contained in this section is purely informative and not meant to be instructional or considered factual in any way.*

Alternative Remedy Formulations

Compresses and poultices

A compress is clean linen soaked in an infusion or decoction, whereas a poultice is a solid material (powdered or fresh dried herb, often fashioned into a paste), that is then typically placed over a wound in an effort to accelerate the healing process. In some cases, a covering of wax paper and warmth from a hot water bottle is thought to be beneficial.

Decoctions

A decoction is a water-based preparation, much like a tea. Its use is thought to be more potent with hard, woody plants that require higher heat to release their active constituents. Dried herb is placed into boiling water for 10–15 min and then strained before consuming.

Extracts

Extracts are concentrated solutions—occasionally concentrated solid or powdered extracts are produced—of chemical constituents of the plant. These are obtained using alcohol-based (organic) or water-based (aqueous) solvent extraction of the plant, after which the solvent is partially or completely removed from the solution.

Infusions

Making an infusion is essentially synonymous with making a tea. Infusions can be hot or cold and are particularly favored for chemical compounds that are thought to be heat-labile. Leaves, flowers, or tender stalks

are often prepared for consumption using this process. The solution is typically strained free of debris before consumption. This type of infusion is not to be confused with a sterile intravenous infusion administered in contemporary medicine.

Tinctures

These are a form of weakly concentrated alcoholic extract of a plant. Occasionally, vinegar or glycerin extracts have been used. Tinctures typically do not need heat to help extract active constituents of the plant, but they do need the solvent and a longer period to extract the active ingredients.

Alternative remedy therapeutic effects (not necessarily proven) include:

1. Alteratives—otherwise known as “blood cleansers,” which promote health and vitality.
2. Adaptogens—enable the body to deal with stress more effectively.
3. Anticathartics—help the body remove edema, congestion, and mucous, typically associated with sinus and respiratory infections/conditions.
4. Astringents—shrink and toughen skin cells, reduce secretions and discharges.
5. Bitters—Supposedly stimulate digestion through “taste sensation.”
6. Carminatives—serve to promote normal bowel function by stimulating peristalsis and “relaxing” the gastrointestinal tract.
7. Cholagogues—stimulate secretion of bile from gallbladder, also produce a laxative effect.
8. Demulcents—internal use soothes and protects inflamed/irritated tissue.
9. Emmenagogues—promote normal menstrual flow.
10. Emollients—external application soothes and protects inflamed/irritated skin.
11. Galactagogues—increase lactation in breastfeeding women.
12. Sialagogues—stimulate flow of saliva.
13. Styptics—stops external bleeding through “astringent” properties.
14. Tonics—agents that “strengthen” an organ or organ system.

Chemical Constituents of Plants

Alkaloids

These naturally occurring nitrogen-containing chemicals (amines) have very diverse pharmacological effects. Examples include codeine, morphine, caffeine, and emetine.

Anthraquinones

Constituents that typically exhibit laxative effects. Examples are aloin, emodin, barbaloin, rhein, and chrysophanol.

Bitter principles

Diverse in structure, these include iridoids and terpenes thought to promote normal bowel function.

Coumarins

These are naturally occurring anticoagulants.

Flavonoids

Examples include vitamin P and hesperidin. They perform a wide range of activities.

Glycosides

Sugar-containing compounds, glycosides perform various activities. Examples include digitalis glycosides, sennosides, and cascariosides.

Phenols

Phenols are weak acids that exhibit analgesic and antiseptic properties. Examples are salicylic acid and eugenol.

Saponins

These are constituents that form foams when agitated in water-based solutions. They are thought to have antiinflammatory, hemolytic, and expectorant effects. Saponins are found in extracts of yucca, sarsaparilla, alfalfa, fenugreek, licorice, and ginseng.

Tannins

Tannins have astringent properties but have the potential to produce liver damage at high concentrations. They are thought to be useful in the treatment of burns and wounds.

Volatile oils (essential oils)

Composed of complex organic compounds (phenols, acids, alcohols, ethers, ketones, and aldehydes), volatile oils evaporate when exposed to air. Volatile oils are found in many plants and may produce the aroma of the plant. Volatile oils exhibit various properties; but some common oils are antiseptic or local irritants, or sedative. Oils of peppermint, clove, cinnamon, garlic, and thyme are volatile oils.

SUMMARY

Precedent suggests it may be foolish to ignore promising chemical constituents in the herbal compendium. Experts

in pharmacognosy (the study of drugs that originate from plant and animal systems) suggest that only a small percentage of plants have been thoroughly investigated for pharmacological activity. Today's major pharmaceutical companies appear to have forgotten this fertile reservoir of unique chemical entities in favor of spin-offs of existing successful agents and new compounds in biotechnology. Our previous practice of developing new medicines from plants has helped many people and, in some cases, brought about the synthesis of even more valuable compounds than those taken directly from the plant itself. With global public interest in the area of alternative and complementary medicine at an all-time high, it is imperative for the safety of patients that we investigate these compounds thoroughly and provide relevant clinical data with respect to herbal medicines and their derivatives. Pharmaceutical companies may wish to redirect some attention toward these botanical chemicals. Patients can only make informed decisions about their health care if indeed healthcare providers themselves are truly informed.

REFERENCES

1. Atherton, D. Towards the Safer Use of Traditional Remedies. *Br. Med. J.* **1994**, *308*, 673–674.
2. Alschuler, L.; et al. Herbal Medicine: What Works, What's Safe. *Patient Care* **1997**, (October), 49–68.
3. Bartels, C.L.; Miller, S.J. Herbal and Related Remedies. *Nut. Clin. Pract.* **1998**, *17*, 5–19.
4. Eisenberg, D.M.; Kessler, R.C.; Foster, C.; et al. Unconventional Medicine in the United States: Prevalence, Costs, and Patterns of Use. *N. Engl. J. Med.* **1993**, *328*, 246–252.
5. Eisenberg, D.M.; Davis, R.B.; Ettner, S.L.; et al. Trends in Alternative Medicine Use in the United States, 1990–1997: Results of a Follow-Up National Survey. *J. Am. Med. Assoc.* **1998**, *280*, 1569–1575.
6. Tyler, V.E. *The Honest Herbal. A Sensible Guide to the Use of Herbs and Related Remedies*, 3rd Ed.; Pharmaceutical Product Press: New York, 1993.
7. Castleman, M. *The Healing Herbs: The Ultimate Guide to the Curative Power of Nature's Medicine*; Rodale Press: Emmau, PA, 1991.
8. Sale, J.D. *Overview of Legislative Developments Concerning Alternative Health Care in the United States*; John E. Fetzer Institutes: Kalamazoo, MI, 1994.
9. Drew, A.K.; Myers, S.P. Safety Issues in Herbal Medicine: Implications for the Health Professions. *Med. J. Aust.* **1997**, *166*, 538–541.
10. Ernst, E. Harmless Herbs. A Review of the Recent Literature. *Am. J. Med.* **1998**, *104*, 170–178.
11. Lord, G.M.; Tagore, R.; Cook, T.; et al. Nephropathy Caused by Chinese Herbs in the UK. *Lancet* **1999**, *354*, 481–482.
12. Abt, A.B.; Oh, J.Y.; Huntington, R.A.; et al. Chinese Herbal Medicine Induced Acute Renal Failure. *Arch. Intern. Med.* **1995**, *155*, 211–212.
13. Vanherweghem, J.L. Aristolochia Nephropathy in Humans: An Outbreak in Belgium. *J. Am. Soc. Nephrol.* **1993**, *4*, 327.
14. Nightingale, S.L. New Safety Measures are Proposed for Dietary Supplements Containing Ephedrine Alkaloids. *J. Am. Med. Assoc.* **1997**, *278*, 15.
15. Blanc, P.D.; et al. Use of Herbal Products, Coffee or Black Tea and Over the Counter Medications as Self-Treatments Among Adults with Asthma. *J. Allergy. Clin. Immunol.* **1997**, *100*, 789–791.
16. Lin, S.H.; Lin, M.S. A Survey on Drug-Related Hospitalization in a Community Teaching Hospital. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **1993**, *31*, 66–69.
17. Piscatelli, S.C.; Burnstein, A.H.; Chaitt, D.; et al. Indinavir Concentration and St. John's Wort. *Lancet* **2000**, *355*, 547–548.
18. Ruschitzka, F.; Meier, P.; Turina, M.; et al. Acute Heart Transplant Rejection Due to St. John's Wort. *Lancet* **2000**, *355*, 548–549.
19. Geatti, O.; Barkan, A.; Turrin, D.; et al. L-Thyroxine Malabsorption Due to the Ingestion of Herbal Remedies. *Thyroidol. Clin. Exp.* **1993**, *5*, 97–102.
20. Murphy, J.M. Preoperative Considerations with Herbal Medicines. *AORNJ* **1999**, *69* (1), 173–183.
21. Weisbord, S.D.; et al. Poison on Line: Acute Renal Failure Caused by Oil of Wormwood Purchased Through the Internet. *N. Engl. J. Med.* **1997**, *337*, 825–827.
22. Cantor, C. Kava and Alcohol. *Med. J. Aust.* **1997**, *167* (10), 560.
23. Spillane, P.K.; et al. Neurological Manifestations of Kava Intoxication. *Med. J. Aust.* **1997**, *167* (3), 172–173.
24. Heiligenstein, E.; et al. Over-the-Counter Psychotropics: A Review of Melatonin, St. John's Wort, Valerian and Kava-Kava. *J. Am. Coll. Health* **1998**, *46* (6), 271–276.

POLAROGRAPHY AND VOLTAMMETRY

A. David Woolfson

The Queen's University of Belfast, Belfast, United Kingdom

INTRODUCTION

Voltammetry is a term that encompasses all measurements based on controlled electrolysis at a microelectrode. Polarography, first introduced by the Czech electrochemist Jaroslav Heyrovsky in 1922, is voltammetry at a special form of mercury microelectrode, the dropping mercury electrode (DME). Mercury electrodes can only be driven to negative potentials because otherwise, the metal dissolves in aqueous solutions as Hg^{2+} . Consequently, polarography is an electroanalytical method based on the cathodic reduction of electroactive species, either metal cations or electroreducible organic species, in an electrically conducting solution. By contrast, voltammetry is based on electroanalysis involving anodic oxidation, preferably in a flowing system in which a self-cleaning action prevents fouling of the solid electrode surface by the products of the electrochemical reaction, thereby leading to nonreproducible current/voltage curves.

HISTORICAL BACKGROUND OF POLAROGRAPHY AND VOLTAMMETRY

The determination of electrocapillary curves for mercury, a phenomenon attributable to changes in the surface tension of the liquid metal as a function of applied potential in an electrolyte solution, was known from the beginning of the 20th century. An electrode consisting of mercury dropping from a fine glass capillary was devised for this purpose. However, secondary maxima appeared at certain points on the electrocapillary curves. The origin of these distortions to the electrocapillary curve was unknown at that time, and it was this phenomenon that Heyrovsky originally sought to investigate and explain. He noted that when certain cations were added to an electrolyte solution, kinks appeared on the electrocapillary curves at potentials close to the values for known electrochemical processes at the DME. The applied voltage between the DME and a mercury pool formed at the bottom of a cell containing an aqueous metal ion solution was gradually increased. Under these conditions,

the resulting current had a small initial value that began to rise rapidly in a reproducible manner as the voltage scan progressed. The point at which this rapid rise in current occurred, the threshold potential, depended only on the species of metal ion present in the test solution. The rapid current rise increased linearly with an increasing applied voltage until the current eventually became constant again. This constant current value, which was found to be proportional to the concentration of metal ion in solution, forms the quantitative basis of polarographic/voltammetric analysis. The method by which Heyrovsky obtained the current voltage curve at the DME involved gradually increasing the applied external voltage (the voltage scan), measuring the corresponding mean current, and plotting this current against the applied voltage. This method constitutes analytical polarography, sometimes referred to as classic or direct current (d.c.) polarography to distinguish it from more modern variants. Heyrovsky was awarded the Nobel Prize in 1959 for his work on the discovery and development of polarographic analysis.

Classic d.c. polarography was limited in its development by the unreliability of early polarographs (instruments used to record current versus applied voltage curves), and the difficulties inherent in operating what were, at that time, relatively complex instruments. Although the method was suitable for metal cations, analytical applications in the field of electroactive organic compounds were restricted by the problem of the higher electrical resistance of nonaqueous solvent systems. Thus, only a water or water-alcohol system could be used with a conventional polarographic cell making up a two-electrode system of DME and reference electrode. Furthermore, determination of metals could be achieved at great sensitivity with spectroscopic methods such as flame emission and atomic absorption spectroscopy. This situation generally remained unchanged until the late 1950s, when modified polarographic techniques, such as square-wave, pulse, and a.c. polarography, began to appear, facilitated by the development of solid-state electronics. The development of a three-electrode system, linked to an electrical circuit known as a potentiostat, overcame the high resistance associated with the use of nonaqueous systems and widened the analytical applicability of the polarography. These developments,

which form the basis of modern polarographic and voltammetric analysis, including their applications to the pharmaceutical sciences, have been referred to as the “renaissance” in polarography (1).

THEORY OF POLAROGRAPHY AND VOLTAMMETRY

In a simple electrolysis experiment in which an electrolyte solution is electrolyzed between two platinum plate electrodes, an anode and a cathode, efficient stirring of the solution is necessary to drive ions toward the electrodes. When the potential difference between the two electrodes, which is a function of the applied external voltage, is sufficiently high, ions are discharged at the electrode surfaces. Such electrodes are working electrodes. The current flowing caused by ion discharge is an electron transfer or faradaic current, i.e., a current attributable to transfer of electrons to the electrode (anodic oxidation) or gain of electrons from the electrode (cathodic reduction). This faradaic current continues to flow until all of the electroactive species is consumed by electrolysis (Fig. 1A). Replacement of a platinum plate electrode with a platinum wire microelectrode and use of a quiet (unstirred) solution result in the electrolysis current reaching a limiting value (Fig. 1B).

The current interest in polarography is attributable to electron transfer between the electrode and the electroactive species. Therefore, the limiting current must be proportional to the concentration of the electroactive species in the bulk solution. This remains true provided the electroactive species can only reach the electrode surface by diffusion along a concentration gradient. The concept of a limiting diffusion current is central to quantitative analysis by polarography or voltammetry. However, the magnitude of this limiting current is nonreproducible because the electrode surface becomes fouled easily by the products of the electrochemical reaction. To overcome this problem, a microelectrode with a renewable surface is required. Use of a microelectrode ensures that an infinitesimally small proportion of the bulk electroactive species is consumed in a single polarographic run, making the technique essentially nondestructive of the analyte.

Suitable working electrodes for voltammetry are those that can be driven to take up a new potential in response to an applied external voltage, a process known as electrode polarization. Reduction or oxidation of an electrochemically active species at a working electrode results in depolarization of the electrode. The word depolarizer is therefore sometimes used to describe an electrochemically active species. In voltammetry, the working electrode may

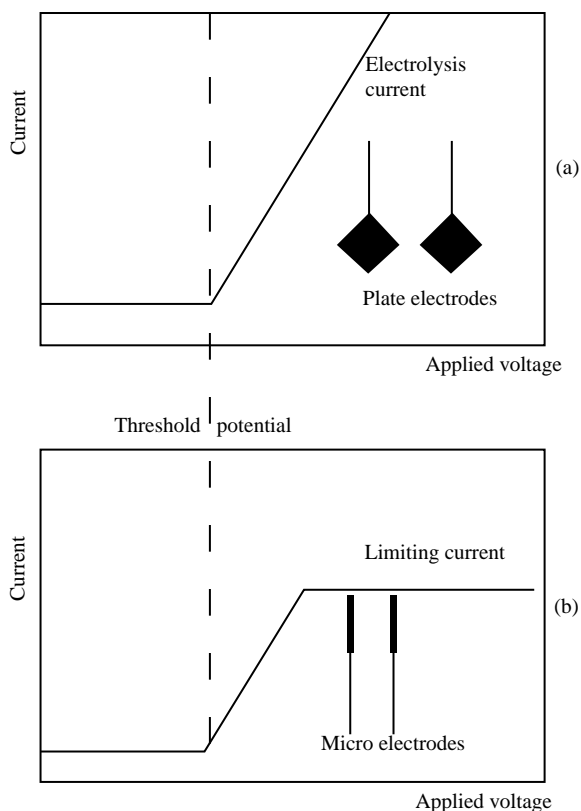


Fig. 1 Comparison between electrolysis and limiting currents. (A) A faradaic electrolysis current between platinum plate electrodes at an applied voltage beyond the threshold potential of the electroactive species. (B) Effect of replacing the platinum plate electrodes with platinum wire microelectrodes. In the absence of stirring, the current reaches a limiting value rather than increasing linearly with an increasing applied voltage.

be fabricated from metals such as gold or platinum, various forms of carbon, or metallic mercury.

The use of mercury dropping from the tip of a fine glass capillary (the DME) as the polarizable electrode has certain specific advantages. Most important, however, the electrode surface is reproducibly renewed as each succeeding drop is formed at the capillary tip. Electrical connection to the DME can be made using a brass post projecting through the glass capillary and contacting the liquid mercury column. The primary disadvantages of mercury as an electrode relate to environmental and safety concerns, difficulties in using mercury in flowing systems (although this is now possible to some extent), and its restriction to electroactive species amenable to cathodic reduction rather than to anodic oxidation. Problems of safety and practicality can be solved through the use of a multimode electrode polarographic/voltammetric electrode assembly. This replaces the classic DME and can

also generate a hanging mercury drop electrode (HMDE.) for use in stripping analysis, in addition to the intermediate static mercury drop electrode (SMDE). Multimode electrodes also offer replacement electrode assemblies for non-DME voltammetric applications in both quiet and stirred solutions. Rotating electrode designs can also be accommodated. For mercury electrodes, the multimode electrode is compact and does not require the gravitational force of the mercury column to extrude the drop. Rather, the drop is formed pneumatically, using nitrogen gas pressure. The mercury is hermetically sealed, an important safety consideration, and only a few milliliters are required for up to 200,000 drops without the need for refilling.

Classic d.c. Polarography

In classic polarography at the DME, as with classical electrolysis, the electrochemical reduction occurs when the applied potential becomes sufficiently negative, i.e., when the applied potential exceeds the threshold value for a given depolarizer. The applied potential is in the form of a linearly increasing voltage ramp with a typical slope of between 2 and 10 mV s⁻¹. Unlike electrolysis, however, the current resulting from application of the ramp voltage does not continue to increase indefinitely until all the electroactive material is consumed. Rather, the current reaches a limiting value represented by the plateau in Fig. 2. This current is limited because when the applied potential is sufficiently negative, the rate of electron transfer becomes instantaneous and exceeds the rate of supply of the depolarizer to the electrode surface. Because the depolarizer can reach the electrode surface only by diffusion along a concentration gradient, the process is said to be diffusion-limited, and the resulting electron-transfer current is the limiting diffusion current, i_d (Fig. 2). The limiting diffusion current is directly proportional to the analyte concentration in the bulk solution.

The electroactive species can also reach the electrode surface by migration under the influence of the electrical field between the electrodes. This gives rise to a migration current that is not diffusion- and, therefore, concentration-dependent. This migration current must be eliminated by providing an excess of charge carriers in the solution that are not discharged within the working potential range of the experiment. Various salt solutions may be used for this purpose. More conveniently, because electrochemical reactions are often pH-dependent, a buffer solution may be used. This solution is variously referred to as the base, inert, or supporting electrolyte. Therefore, although the ions of the supporting electrolyte will move through the solution and carry charge, no current will flow in the external circuit because no faradaic process will occur in the absence of a depolarizer.

The electroactive species can also reach the electrode surface by convection, giving rise to a convection current that is, again, nonconcentration-dependent. Convection effects are attributable to stirring of the solution or, less frequently, to thermal currents. Thus, polarography and voltammetry are carried out in quiet (unstirred) solutions.

The concentration-dependent mass transport process, when other mass transport processes have been eliminated, is diffusion of the electroactive species toward the electrode surface along a concentration gradient. As the electroactive species approaches the surface of the DME, it will be electrochemically reduced. Thus, in a narrow solution layer, the diffusion layer, immediately adjacent to the drop surface, there will be a lower concentration of the electroactive species than that present in the bulk solution, giving rise to the concentration gradient. It may be shown from Fick's law of diffusion that for an electroactive species diffusing across a thin diffusion layer of thickness d , the diffusion current, i_d , will be given by Eq. 1.

$$i_d = n \cdot F \cdot A \cdot D \cdot (C - C_i)/d \quad (1)$$

where D is the diffusion coefficient of the electroactive species, F is Faraday's Constant, A is the drop surface area, n is the number of electrons transferred per molecule of depolarizer, C is the bulk concentration of the depolarizer, and C_i is its concentration in the diffusion layer.

As C_i approaches zero, the rate of diffusion becomes proportional to the concentration of depolarizer in the bulk solution. Beyond the threshold potential, the electron transfer reaction will be initiated, and, as the potential is gradually increased, the rate of this reaction will continue to increase until it exceeds the rate of supply of the depolarizer to the electrode surface by diffusion, with all other mass transport processes having been suppressed. Under these conditions, the diffusion process becomes the rate-limiting step, and the resulting faradaic current, i_d , is now said to be the limiting diffusion current. Because i_d is measured over many individual drop lifetimes, it is properly described as the average limiting diffusion current and is given by Eq. 2, where i_d and C are the only variables.

$$i_d = n \cdot F \cdot A \cdot D \cdot C/d \quad (2)$$

The limiting diffusion current is described quantitatively by the Ilkovic equation (Eq. 3).

$$i_d = 708 \cdot n \cdot D^{1/2} \cdot m^{2/3} \cdot t^{1/6} \cdot C \quad (3)$$

where m is the rate of flow of mercury from the DME in mg s⁻¹ and t is the drop lifetime in seconds. D has units of cm² s⁻¹, C is expressed as mM l⁻¹, and i_d is expressed in mA.

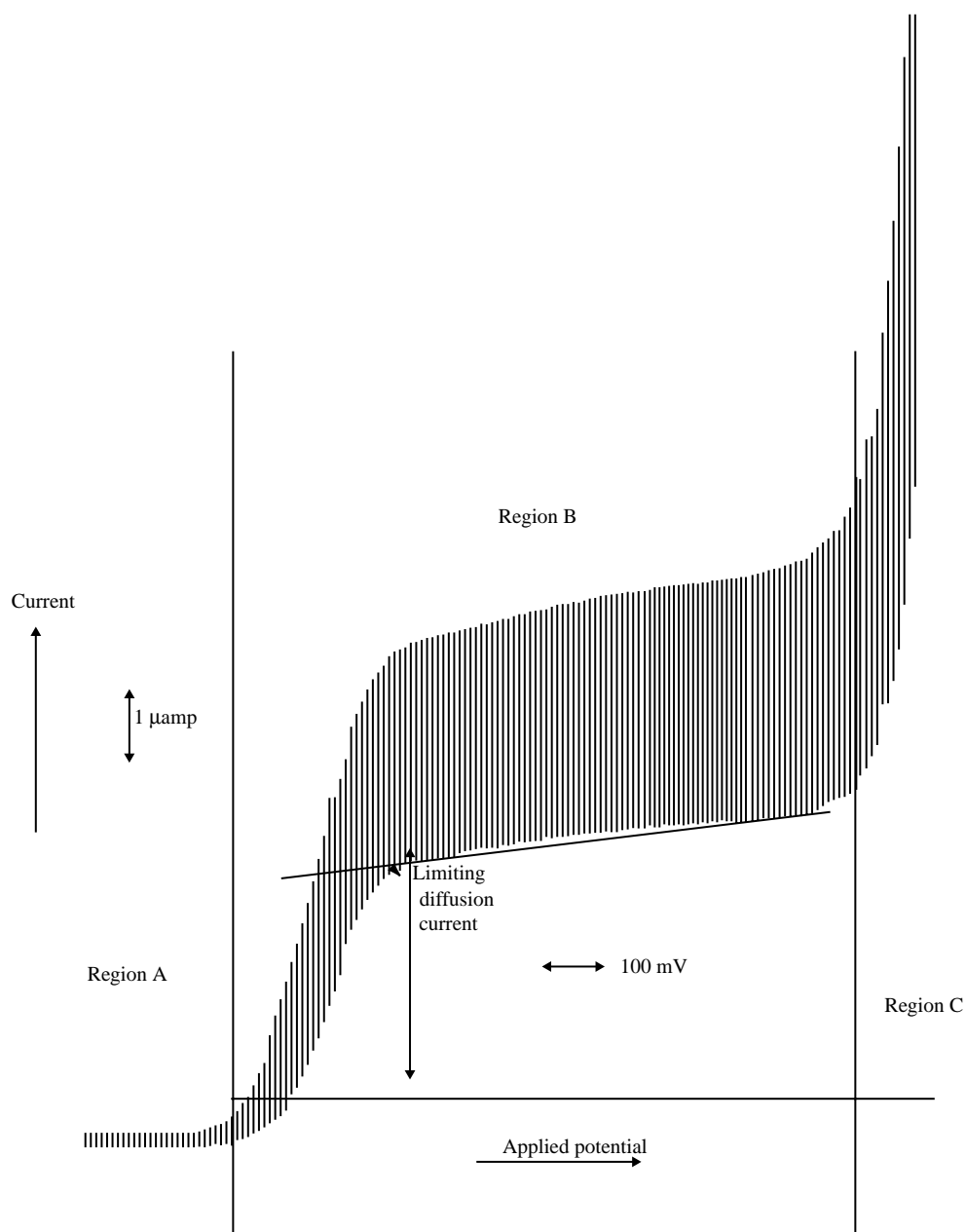


Fig. 2 Classic d.c. polarogram of diazepam (20 mg ml^{-1}) in 0.1 M sulfuric acid as the supporting electrolyte. Region A is the background current attributable to the supporting electrolyte, region B shows the rising faradaic current as it reaches a limiting value, and region C is the cut-off point at which the current goes off scale caused by the reduction of hydrogen ions.

Electrical Double Layer

When only the inert electrolyte is present in the polarographic cell a residual current will still flow. This current, which is nonfaradaic, is attributable to the formation of an electrical double layer in the solution adjacent to the electrode surface (Fig. 3). At all applied potentials, a

current flows to develop this double layer, and the process may be considered analogous to the charging of a parallel plate capacitor. Therefore, the charging current is a capacitance current and varies during the drop lifetime, i.e., with the size of the mercury drop. When the drop surface area is increasing rapidly from the start of the drop lifetime, the capacitance current is a maximum, falling to a

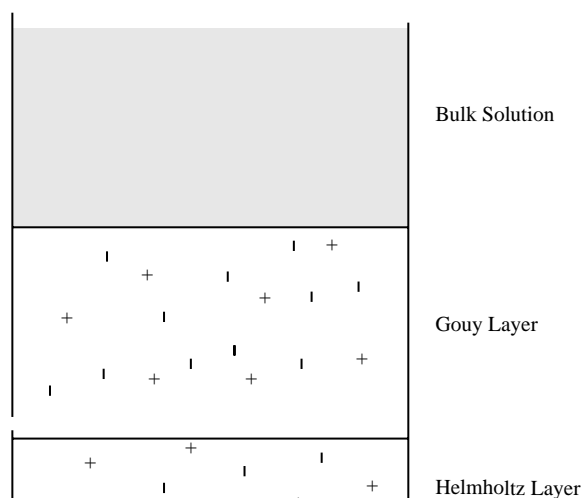


Fig. 3 The electrical double layer at a mercury drop shorted to a reference electrode and with no external applied potential.

minimum near the end of the drop lifetime when the drop size is at a maximum and the surface area of the drop is momentarily constant. The magnitude and direction of the capacitance current vary with the applied potential because of the variation in the surface tension of mercury with electrode potential. When the mercury drop is at its maximum surface tension, there is effectively no electrical double layer at the drop surface and, therefore, no capacitance current, a point known as the electrocapillary maximum. Beyond this potential, the capacitance current changes direction as the double layer is reversed, with the mercury drop now possessing a negative charge. The practical consequence of this is to impose the familiar serrated pattern on the polarographic wave (Fig. 2).

Mechanisms of Electrode Processes

The shape of the polarographic wave is further influenced by the nature of the electrode process occurring at the drop surface. Polarographic waves may be reversible, irreversible, or quasireversible. The overall electrode process comprises the diffusion, electron transfer, and electrochemical reaction steps.

Reversible processes are those that attain thermodynamic equilibrium at every instant of the drop life owing to rapid electron transfer. Reversible processes give rise to well-defined d.c. polarograms, and diffusion control is always the determining factor. Irreversible processes are so slow that equilibrium is not attained during the drop lifetime, and d.c. polarograms dependent on such processes often show poor definition. The rate-controlling step may be either the electron transfer process or the

subsequent chemical reaction. Many organic reductions at the DME, however, fall into an intermediate category, quasireversible processes. Whereas the rate constant for the reverse reaction will be negligible for a wholly irreversible reaction, it has an intermediate value for quasireversible reactions. Such reactions are normally seen only with longer drop times of at least 3 s.

The reversibility, or otherwise, of an electrode process is best investigated using the technique of cyclic voltammetry (2), in which a rapid forward and reverse voltage ramp is applied in triangular form (Fig. 4A) to interact with both the electroactive substance and its reduction product that, for quasi- or fully reversible processes, may be oxidized back to the starting material, giving the characteristic waveform shown in Fig. 4B. The separation between anodic and cathodic peaks indicates whether the electrode process is quasi- or fully reversible. Additional mechanistic investigations can also be made in respect to the number of electrons involved per molecule

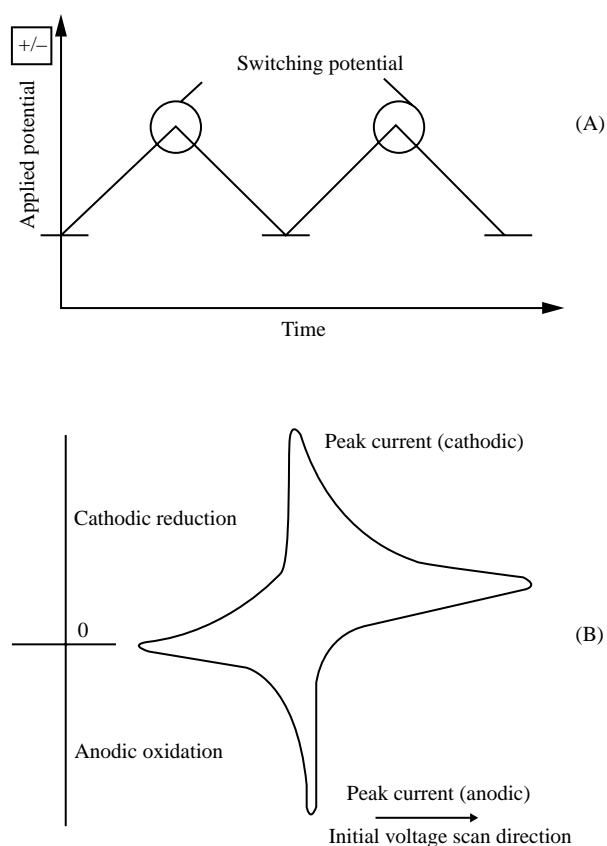


Fig. 4 Cyclic voltammetry. (A) Voltage waveform showing the rapid forward and reverse voltage sweeps. (B) Typical cyclic voltammogram for completely reversible system.

in the electron transfer process, a factor that can be determined by controlled-potential coulometry (3).

Effect of Oxygen in Polarography

In polarography, but not in anodic voltammetry, it is necessary to provide a facility for removing oxygen from the electrolyte solution in the polarographic cell. This is normally achieved by bubbling oxygen-free nitrogen through the solution for 10 min before starting the voltage scan. The surface of the solution is then blanketed by oxygen-free nitrogen during the polarographic run to prevent ingress of additional oxygen from the atmosphere. Removal of oxygen is necessary because the dissolved gas is polarographically active and can mask the analytical signal of interest in certain potential regions.

MODERN POLAROGRAPHIC AND VOLTAMMETRIC METHODS

In classic d.c. polarography, the actual current measured comprises the limiting diffusion current, together with current components, because of background electrical signals and, more important, the current charging the double layer capacitor at the electrode surface. Classic d.c. polarography is in many ways best suited to the elucidation of electrode processes. It lacks sensitivity for modern analytical purposes, and the sigmoidal current/voltage curves are difficult to measure. Modern polarographic and voltammetric methods are now available (Fig. 5) in which these problems have generally been resolved, producing an analytical technique with much enhanced sensitivity and a more easily interpretable current–voltage waveform.

Current-Sampled d.c. Polarography

Current-sampled polarography is a modern variant of the original *tast* (from the German “to touch”) polarography, a method that involved the measurement of the polarographic current only at a fixed time interval during each drop lifetime. This was originally accomplished by the use of a mechanical touching contact (4) but is achieved in modern instruments using a digital approach in which the current is electronically sampled at a precise moment, typically 20 ms, near the end of the drop lifetime when the area of the drop is effectively constant. For this purpose, the drop lifetime is precisely and mechanically controlled by the instrument rather than being gravity-dependent, a feature shared by other modern methods. Current-sampled polarography produces a smooth polarogram by

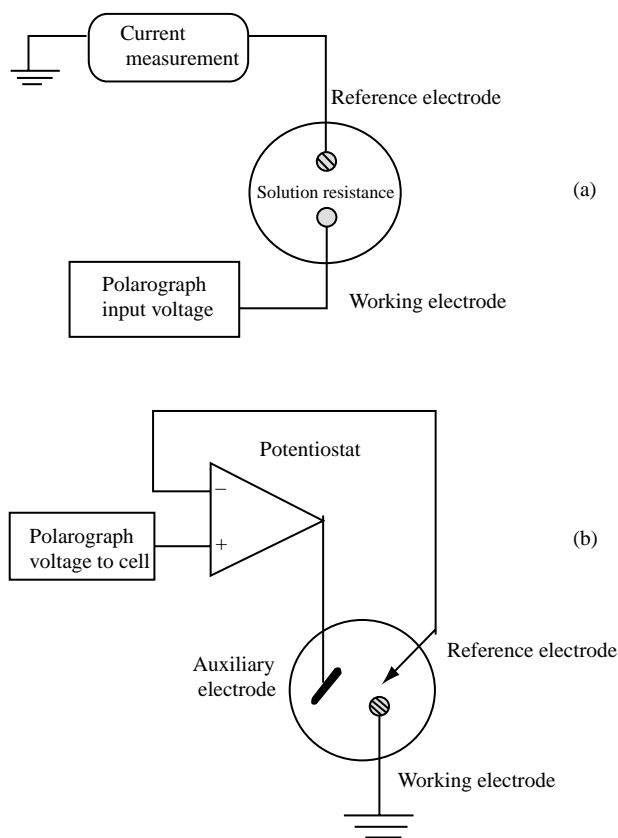


Fig. 5 Electrical circuitry for polarography/voltammetry. (A) A simple two-electrode system. (B) Illustrates a modern three-electrode system incorporating a potentiostat circuit.

elimination of the current variation during the drop lifetime (Fig. 6). However, although the typical serrations of the classic polarogram are gone, a slight staircase pattern can still be discerned on the current-sampled polarogram, a feature shared with many other modern voltammetric methods. The staircase pattern is attributable to the sampled current being held in the memory of the instrument and its value fed out continuously to the recording device until the next sampling period.

Current-sampled polarography offers only a marginal improvement on the sensitivity of the classic method because there is a more favorable faradaic-to-charging-current ratio at the end of the drop time when the drop is almost stationary. Thus, its only real benefit over the classic method is the clearer polarogram obtained.

Pulse Polarography

For routine quantitative analysis, pulse polarography and voltammetry are perhaps the most useful of the modern

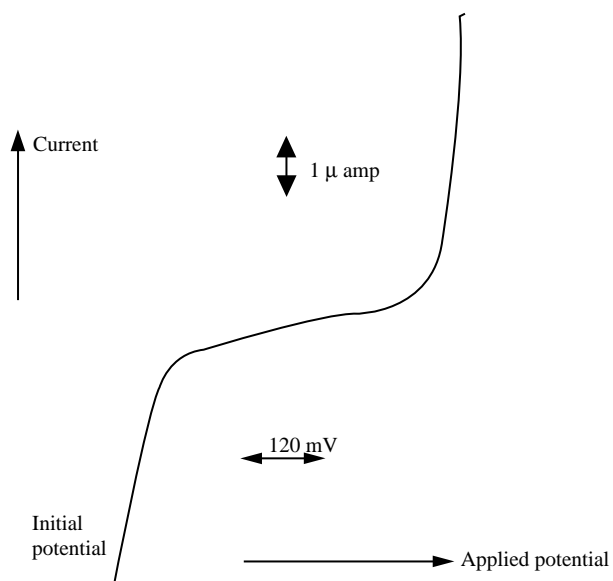


Fig. 6 A typical current-sampled polarogram. Note the staircase pattern indicating the individual current sampling periods.

variants on the classic method. Unlike d.c. and current-sampled polarography, the applied voltage is not a simple d.c. ramp but has a more complex format involving periodic application of the potential during short time intervals. There are two major pulse methods: normal (integral) pulse and differential pulse polarography/voltammetry, although many variations of these are now available.

The primary advantage of pulse methods is its significantly more favorable faradaic-to-charging-current ratio and, in the case of differential pulse methods, a more conventional Gaussian-shaped waveform for the current potential plot. In pulse methods, the initial charging current is increased when the pulse is applied, effectively giving a pulse charging current. If the pulse falls on the rising (faradaic) portion of the polarogram/voltammogram, there will be a large increase in current, over and above the value of the pulse charging current. Provided the pulse has been applied at the end of the drop lifetime, when the drop surface area of the DME is briefly constant, or if a solid electrode is used, then both currents will decay from the point of the initial pulse application, the current decay being a function of time. However, the charging current component of the total measured current decays much more rapidly than does the faradaic component. Thus, if the current is measured (sampled) at the end of the pulse application period, it will consist primarily of the faradaic component, thus yielding a marked sensitivity increase over d.c. methods. Sensitivity is also greater because the boundary diffusion layer at the electrode-solution interface

is narrower than is that when the potential is applied continuously. Therefore, the rate of diffusion of the electroactive species toward the electrode is increased, with a concomitant increase in the diffusion current.

Normal (Integral) Pulse Polarography/Voltammetry (NPP/NPV)

In NPP/NPV (5), the voltage is applied in a series of increasing voltage pulses (Fig. 7) from a baseline voltage selected by the analyst. Between pulses, the baseline voltage is restored. The pulse amplitude increases linearly with time, depending on the conditions set by the operator. The voltage pulse is applied at the end of the drop lifetime in polarography. Although precise timings vary with different instruments, typically the voltage pulse is applied for the last 60 ms of the drop. As with current-sampled polarography, the resulting current is sampled over only the final 20 ms of the drop lifetime, producing a polarogram/voltammogram identical in appearance to that of the current-sampled technique but with a greater current yield. The appearance of the polarogram is perceived to be the primary disadvantage of NPP/NPV because analysts tend to prefer a more conventional Gaussian-shaped graphic output of data, the sigmoidal shape being difficult to quantify.

Differential Pulse Polarography/Voltammetry (DPP/DPV)

DPP and DPV are probably the most analytically useful of all the voltammetric methods. DPP and DPV produce current-potential plots in the typical peak form familiar to,

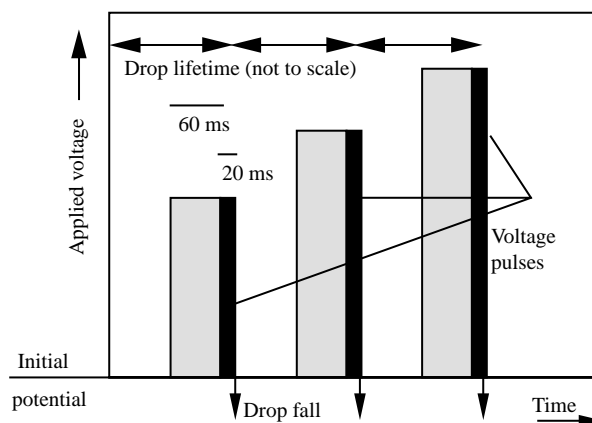


Fig. 7 Applied voltage waveform for normal pulse polarography.

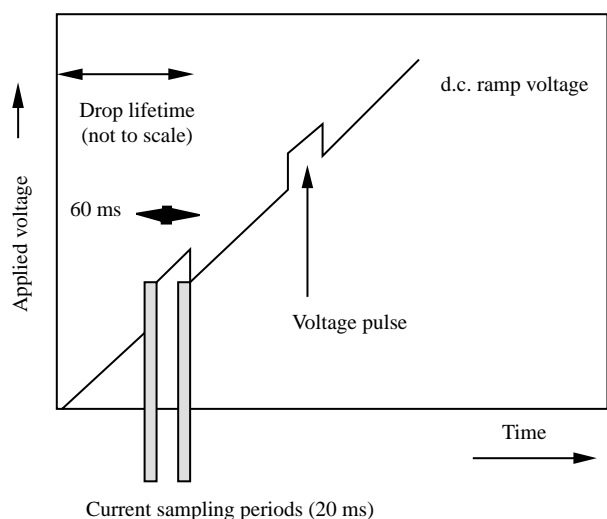


Fig. 8 Applied voltage waveform for differential pulse polarography.

and readily interpretable by the analyst. DPP and DPV result from a variation in the pattern of the applied voltage.

In DPP/DPV, a small fixed voltage pulse of between 5 and 100 mV is superimposed on a slow linear voltage ramp (Fig. 8). When the ramp voltage coincides with the faradaic process, a faradaic current occurs continuously, along with the charging current. When the small pulse voltage is applied, in addition to the ramp voltage, at a given time, then both a new faradaic and a new charging current will be generated. As with NPP/NPV, the faradaic component decays during the pulse application, which, for polarography, is at the end of the drop lifetime. The charging current owing to the pulse also decays, but much more rapidly than the faradaic component so that if the current is sampled at the end of the pulse application, there is maximum separation between faradaic and charging currents and, thus, maximum sensitivity. In fact, in DPP/DPV the current is sampled twice, typically for 20 ms immediately before application of the pulse and again for the last 20 ms of the pulse application. The differential signal thus recorded results from the small increase in current (di) because of the small increase in the applied voltage pulse (dE). This further reduces the charging current contribution before the pulse and leads to an additional sensitivity increase over NPP/NPV. Effectively, it is the derivative (di/dE) of the classic polarographic wave that is produced (Fig. 13). This quantity, when plotted against the applied potential, yields a peak-shaped output because the change in the faradaic component when the constant voltage pulse is applied reaches a maximum on the steepest part of the polarogram and falls to almost zero in the baseline and plateau regions.

The sensitivity of DPP/DPV exceeds even that of the normal pulse method by approximately a single order of magnitude, being approximately equivalent to that of gas chromatography with F.I.D. (approximately 10^{-8} M). It may, therefore, be used for the determination of drugs in biological matrices. In addition, DPP and DPV have excellent resolving power, being able to differentiate peaks, which are no more than 50 mV apart, owing to different electroactive species in the same solution. Both normal and differential pulse techniques are suitable for use with solid working electrodes such as glassy carbon. The various instrumental timings remain the same, even though the constraint imposed by the variable area of the DME has been removed. The $E_{1/2}$ value is, of course, not discernible from DPP/DPV and is replaced by the near-identical quantity E_p , the peak potential.

The magnitude of the peak current (i_p) in DPP/DPV is given by Eq. 4 (6).

$$i_p = \frac{n^2 F^2}{RT} \cdot E \cdot A \cdot C \cdot \frac{D}{\sqrt{\pi t}} \cdot \frac{\exp(E - E_{1/2} + 0.5 E) \cdot CnF/RT}{\{1 + [\exp(E - E_{1/2} + 0.5 E) \cdot nF/RT]\}^2} \quad (4)$$

where n is the number of electrons transferred in the electrode reaction, F is Faraday's constant, R is the universal gas constant, T is the absolute temperature, DE is the pulse amplitude or modulation, A is the electrode surface area, C is the concentration of the electroactive species, D is the diffusion coefficient of the electroactive species, t is the time elapsed from pulse application to current measurement, E is the ramp potential just before application of the pulse, and $E_{1/2}$ is the half-wave potential.

Equation 4 indicates that a linear relationship exists between peak current in DPP/DPV and peak potential and that the peak current will increase with pulse amplitude. However, the charging current also increases with pulse amplitude so that the value for DE must be chosen to maximize i_p but must not be too large; otherwise, resolution of the peak will be diminished. Typical values for DE are 50 or 100 mV.

Linear Sweep Voltammetry (LSV)

Linear sweep voltammetry involves the application of a rapid voltage scan, 100 mV s^{-1} or higher, to a stationary electrode such as the HMDE or a solid electrode such as glassy carbon. The theoretical treatment is based on the Randles-Sevcik equation (2, 7). At slow scan rates, the magnitude of the concentration gradient across the diffusion layer is governed by the rate of depletion of the electroactive species across this layer. With the fast

scan rates used in LSV, the diffusion layer is narrower, the concentration gradient is consequently larger, and the resulting diffusion current is greater. As the depolarizer is used up by reaction at the electrode surface, the diffusion layer widens as it extends further into the bulk solution, and, unlike the DME, equilibrium conditions are not periodically restored by the stirring effect of the falling drop. Thus, there is a gradual decay in the diffusion current, giving a peak-like appearance to the LSV output (Fig. 9). The rapidly increasing potential in LSV results in nonequilibrium conditions at the electrode surface throughout the period of the voltage scan.

Alternating Current Polarography/Voltammetry (ACP/ACV)

Alternating current polarography and voltammetry encompass a wide range of polarographic and voltammetric modes characterized by a periodic applied voltage waveform, such as a square-wave, pulsed, or saw-tooth pattern. The production of such waveforms may require a voltage function generator in addition to the normal polarograph, although some instruments have the function generator built in and can therefore perform ACP/ACV as a standard technique.

The most common applications of ACP/ACV involve the application of a small-amplitude sinusoidal alternating

potential superimposed onto a ramp voltage. The resulting current is an alternating current, the d.c. component being filtered out by use of a phase-sensitive current detector. This is possible because the faradaic and charging components of the current, respectively, have phase angles of 45 and 90° with reference to the applied sinusoidal potential. The detector rejects the 90° component and measures only the faradaic current.

The applied sinusoidal voltage has a typical amplitude range of ± 50 mV, and the polarogram is a plot of the fundamental harmonic alternating current (a.c.) against the ramp voltage. Because the d.c. current component is filtered out, the current values before and after the rising portion of the d.c. polarogram are close to zero in the a.c. mode, and therefore, the resulting polarogram has a peak shape that approximately follows the rising portion of the d.c. wave. The use of a fast voltage scan, analogous to LSV, is possible with ACV at a solid (stationary) electrode.

The special case of square-wave voltammetry (SWV) is worth noting separately from other alternating current techniques because it is both more rapid and more sensitive than DPP/DPV. In SWV, the applied potential waveform is a staircase with constant step height on which is superimposed an asymmetrical forward and reverse voltage pulse of constant amplitude and very short duration, typically less than 10 ms. Thus, the entire polarogram may be run in about approximately 1 s, with the enhanced sensitivity of the method owing to sampling of the current at the end of both the forward and reverse directions of the pulse.

Stripping Voltammetry

This is an ultrasensitive technique most widely used in the trace determination of metals and, increasingly, for organic compounds, including pharmaceuticals. The outstanding sensitivity of the method is due to an initial preconcentration (accumulation) step that can result in a 1000-fold increase in the available analyte concentration compared with the bulk solution. The most commonly used working electrode design for stripping analysis has a single mercury drop hanging from the electrode tip during the course of the stripping experiment. Such an electrode is the HMDE, although other designs such as the mercury film electrode (MFE), in which the mercury is supported as a thin film on a carbon electrode support (8), and SMDE have become increasingly important (9). The SMDE allows faster stirring rates during the deposition step than are possible with the HMDE, the limiting factor in the latter case being the dislodgement of the mercury drop by vigorous stirring. In addition to mercury, nonplated solid

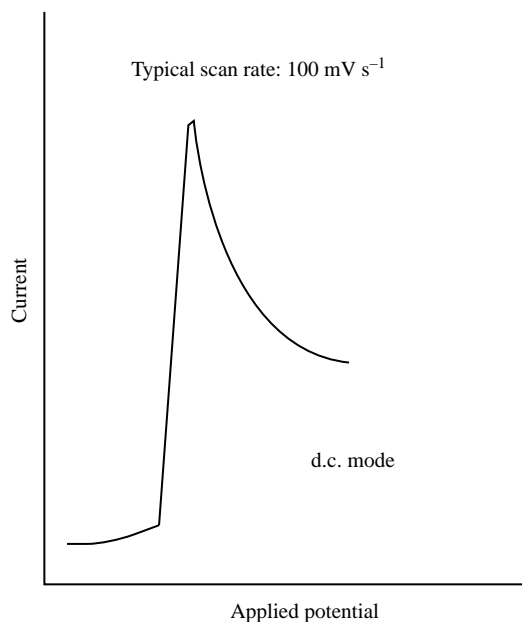
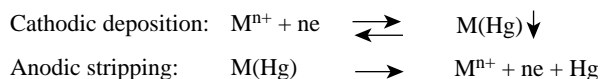


Fig. 9 Typical appearance of a linear sweep voltammogram at a fast-scan rate of 100 mV s^{-1} .

electrodes fabricated from a noble metal (gold or platinum) or from carbon (glassy carbon, carbon paste) have also been used for stripping analysis, typically for the determination of metals that are insoluble in mercury that have very positive redox potentials. For most pharmaceutical applications of stripping analysis that involve organic compounds, mercury electrodes are used. Stripping analysis can then be performed using a conventional modern polarograph linked to a specific working electrode suitable for the chosen analysis.

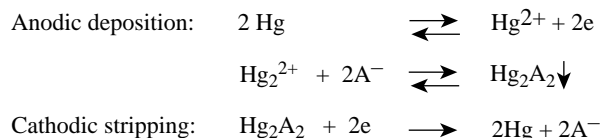
Stripping analysis for compounds of pharmaceutical interest consists of two steps. First, the electroactive material is deposited onto a mercury electrode, thus concentrating the analate by extracting it from the bulk solution. This controlled deposition, which may be electrolytic, is carried out for a defined time period, with constant stirring of the bulk solution because, in this case, it is necessary to drive the analyte toward the electrode surface as efficiently as possible. Second, there is the stripping step. This involves stripping (removing) the analyte from the electrode surface back into the solution by application of a suitable potential. This second step is the measurement step, the resulting faradaic current being quantitative in respect to the amount of analate present in the bulk solution. The applied potential can be a simple ramp voltage, pulse, or periodic waveform. Thus, most of the modern polarographic modes can be coupled to the stripping step to give additional increases in both sensitivity and selectivity. Fast stripping steps such as the use of the semidifferential mode (10) have become increasingly popular, although differential pulse and linear sweep modes remain prevalent.

Anodic stripping voltammetry (ASV) has been the most widely used stripping variant, typically for the trace analysis of metals in solution. Thus, an electrolytic deposition step onto a mercury cathode, possibly lasting several minutes, is followed by an anodic stripping step in which the potential scan goes toward positive values. The deposition step itself results in the formation of a metal amalgam. As with all stripping variants, hydrodynamic parameters (stirring rate, deposition time, solution composition, electrode location) must be carefully controlled and reproducible to obtain a quantitative response to changing bulk concentrations. The process may be summarized as:



Cathodic stripping voltammetry (CSV) may be considered the reverse of ASV in that the electrolytic

deposition step is carried out at a positive (anodic) applied potential, the deposited analate then being stripped by application of a cathodic voltage scan. CSV has been used for the determination of various anions (11) and for certain drug molecules such as organosulfur compounds (12). CSV may be summarized as:



Adsorptive stripping voltammetry (AdSV) is of increasing importance in trace determinations of pharmaceutical compounds. In this method, the preconcentration step is adsorptive rather than electrolytic, resulting in an adsorbed film of the analate on the electrode surface (13). The stripping step typically uses LSV or the differential pulse mode in either the cathodic or anodic direction, as required. The HMDE is typically used for cathodic reductive stripping, whereas carbon or noble metal electrodes are used in the adsorptive mode.

Factors affecting the adsorptive process in AdSV include the solvent, solution pH, mass-transport processes, stirring rate, deposition time, and applied potential. The sensitivity advantage of AdSV over conventional polarography/voltammetry using an equivalent mode may be up to 100-fold. However, the primary advantage of AdSV is its ability to simplify sample preparation when the analyte is in a complex matrix, such as the determination of a drug or its metabolite in body fluids. Having adsorbed the analyte to the electrode directly from the complex medium, the electrode with the adsorbed analate film may then be transferred to a blank electrolyte solution before the stripping step. This method, known as medium exchange (14), has considerable potential, particularly in a flow analysis mode (15) for both pharmaceutical and clinical analyses. Its value is perhaps not yet fully realized because of a preference for, and greater familiarity with, chromatographic methods.

Electrochemical Detection for High-Performance Liquid Chromatography (ELCD) and Flow-Injection Analysis (ED-FIA)

Pharmaceutical analysts often have no experience in direct polarographic or voltammetric methods, but almost all will have used high-performance liquid chromatography (HPLC) for the determination of drugs and/or metabolites in biological matrices or drugs and/or their degradation products in pharmaceutical formulations. Spectroscopy

(ultraviolet and fluorescence) is the most common detection method in HPLC, but for molecules that do not possess a suitable chromophore or when increased sensitivity or specificity is required, electrochemical detection offers a suitable alternative. ELCD is applicable to any molecular species capable of electrochemical oxidation or reduction at an electrode. Most detectors are based on solid electrodes, notably carbon paste and, particularly, glassy carbon. These detectors are best operated in the anodic mode, but they can also be used for reductive processes. Alternatively, mercury films on a noble metal electrode can be used for cathodic reduction.

The principle of ELCD is described in Fig. 10 and has been addressed in detail by Stulik and Pacakova (16). The detector is set to an applied voltage large enough to cause the electrochemical reaction of interest to occur, i.e., the applied voltage is on the plateau of the polarographic wave. Thus, when the species of interest is eluted from the column, a faradaic current is produced in the detector and results in the usual chromatographic signal. This is an example of amperometric detection because an electrical current is responsible for the analytical signal. With a solid electrode, the detector may be described as a voltammetric detector. Various detector cell designs are available. The most common configuration is the wall-jet cell (Fig. 11), in which the eluent is sprayed against an internal wall of the detector cell. The wall is fabricated from glassy carbon, a highly polished and impermeable material. The flowing eluent spray ensures that any products from the electrochemical oxidation reaction are

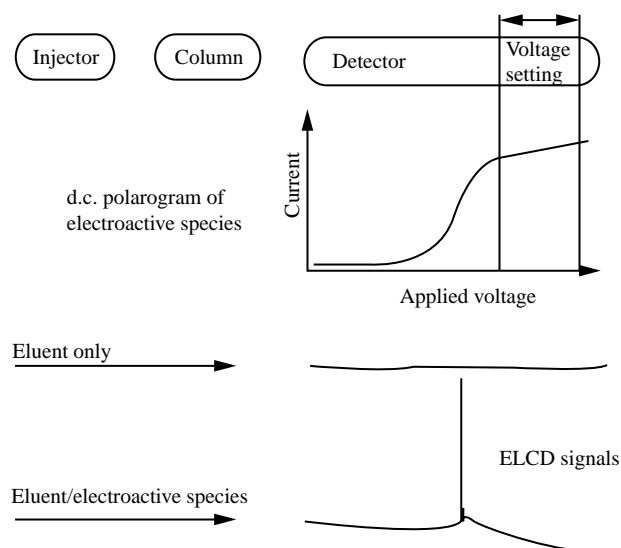


Fig. 10 Principle of electrochemical detection for HPLC.

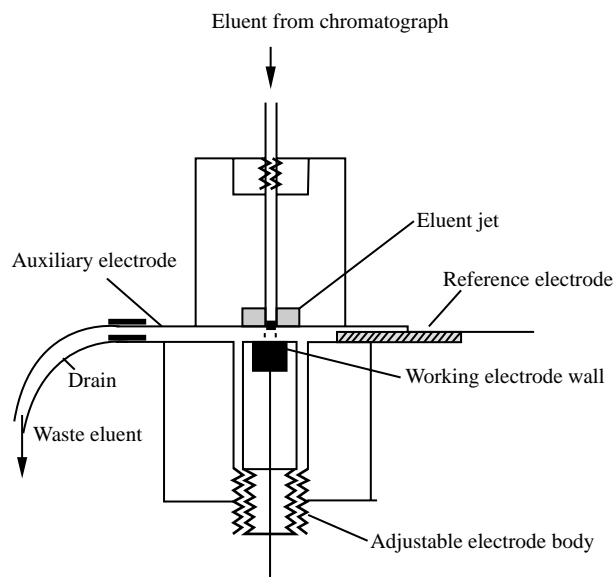


Fig. 11 Wall-jet electrochemical detector cell for HPLC.

removed from the electrode surface, thus ensuring signal reproducibility.

In conjunction with the various designs of electrochemical detector cells, a range of applied voltage waveforms may be used for ELCD (Fig. 12). The

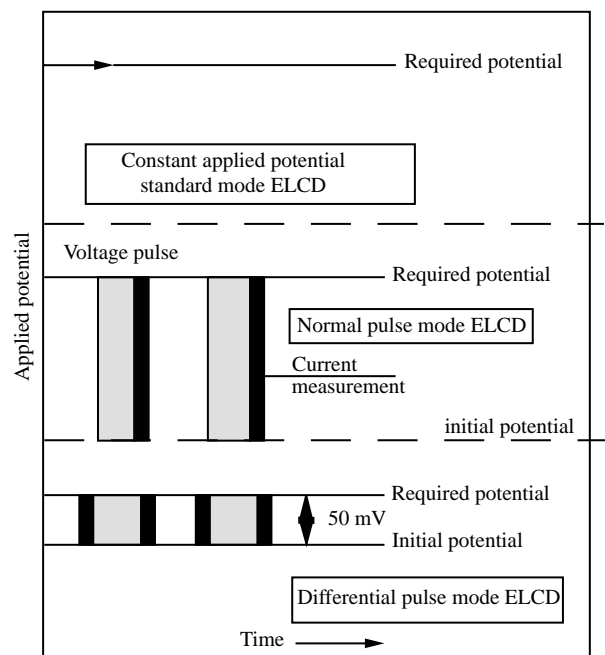


Fig. 12 Applied voltage waveforms for various modes of electrochemical detection in HPLC.

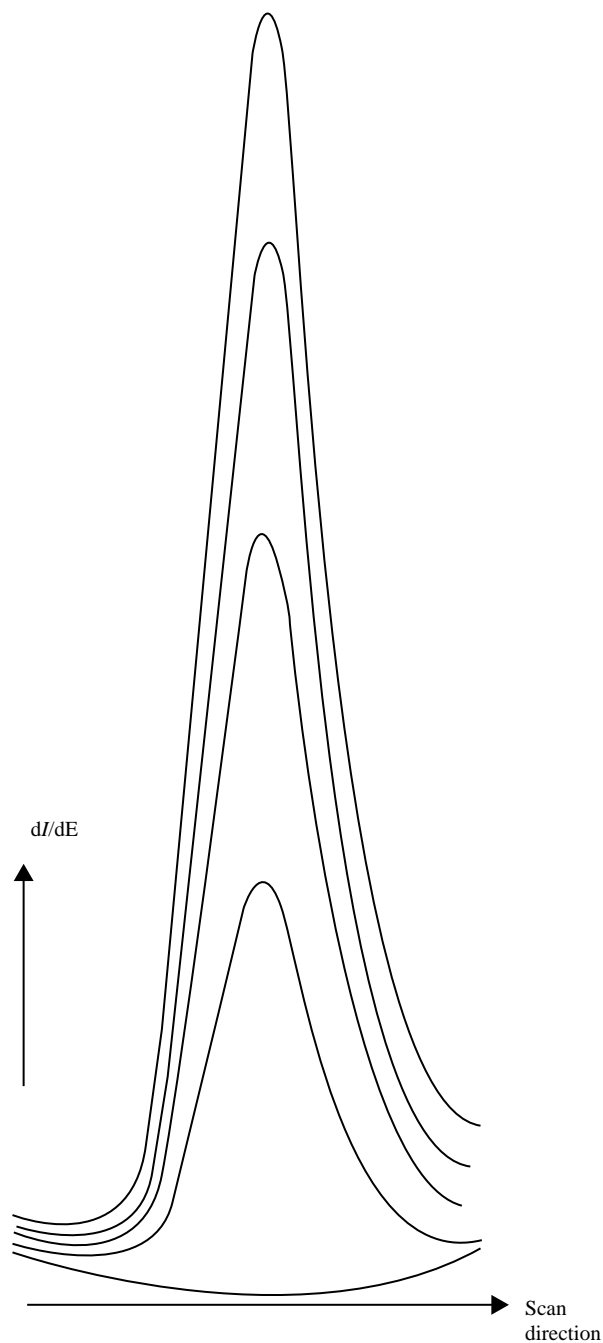


Fig. 13 Differential pulse voltammograms of pyridoxine hydrochloride at the glassy carbon electrode obtained by adding successive aliquots (0.1 ml) of an aqueous solution (0.1M) of the vitamin to pH 4 citric acid as the supporting electrolyte.

differential pulse mode (Fig. 13) is particularly suited to ELCD because it can further enhance the separability of components by selectively setting the detector to the peak potential of an individual component in a mixture.

Additional information can also be obtained using a rapid-scan square-wave detector. As the name implies, this mode uses very fast voltage scans so that a three-dimensional output can be obtained. In effect, many voltammograms are run of a component as it elutes from the column, giving an output of current versus applied potential versus time. Thus, peak purity can be checked, and the device is analogous in its applications to the better known diode array spectrophotometric detector. Voltammetric detectors may also be used in a flow-injection mode (17), effectively omitting the chromatographic step, for drug analysis. Again, a wide range of detector cell designs and polarographic/voltammetric modes are available.

APPLICATIONS OF POLAROGRAPHY AND VOLTAMMETRY TO PHARMACEUTICAL ANALYSIS

A comprehensive review of the use of polarography and voltammetry for drug analysis is beyond the scope of this article. (See Bibliography for reviews on the subject.) The analytical range of the methods is also well illustrated in the references (18–28).

CONCLUSIONS

The availability of a wide range of modern polarographic and voltammetric modes, together with the use of ELCD and stripping analysis, offers the analyst powerful analytical tools for performing drug assays. The decision as to whether an electroanalytical method should be used for a given assay and the selection of a particular polarographic/voltammetric technique will depend on the drug of interest, the analytical matrix, and the type of data that are required. Given the advances in analytical instrumentation, polarography/voltammetry may now be used with confidence, when necessary, by all pharmaceutical analysts.

REFERENCES

1. Flato, J.B. Renaissance in Polarographic and Voltammetric Analysis. *Anal. Chem.* **1972**, *44*, 75a–85a.
2. Sevcik, A. Oscillographic Polarography with Periodical Triangular Voltage. *Coll. Czech. Chem. Commun.* **1948**, *13*, 349–354.
3. Hickling, A. Automatic Control of the Potential of a Working Electrode. *Trans. Faraday Soc.* **1942**, *38*, 27–33.

4. Kronenberger, K.; Strehlow, H.; Elbel, A.W. Intermittent Polarography. *Leybolds Polarogr. Berlin* **1957**, *5*, 62–77.
5. Barker, G.C.; Gardner, A.W. Pulse Polarography. *Z. Anal. Chem.* **1960**, *173*, 79–83.
6. Vire, J.C.; Patriarche, G.J.; *Pharm. Int.* **1981**, (May), 119–123.
7. Randles, J.E.B. Cathode Ray Polarograph. *Trans. Faraday Soc.* **1948**, *44*, 322–327.
8. Florence, T.M. Anodic Stripping Voltammetry with a Glassy Carbon Electrode Mercury-Plated In Situ. *J. Electroanal. Chem.* **1970**, *27*, 273–281.
9. Peterson, W.R. Static Mercury Drop Electrode. *Am. Lab.* **1979**, *12*, 69–74.
10. Goto, M.; Ikenoya, K.; Kajihara, M.; Ishii, D. Application of Semi-Differential Analysis to Anodic Stripping Voltammetry. *Anal. Chim. Acta* **1978**, *101*, 131–138.
11. Brainina, K.Z. Film Stripping Voltammetry. *Talanta* **1971**, *18*, 513–539.
12. Davidson, I.E.; Smyth, W.F. Cathodic Stripping Voltammetry of Some Sulfur-Containing Organic Compounds in Biological Fluids. *Anal. Chem.* **1977**, *49*, 1195–1198.
13. Wang, J. Adsorptive Stripping Voltammetry. *Int. Lab.* **1985**, *15*, 68–70.
14. Neeb, R.; Kiehna, I.Z. Effect of Salts on the Anodic Peak Height in Inverse Voltammetry. *Anal. Chem.* **1967**, *241*, 142–155.
15. Wang, J.; Dewald, H.D.; Greene, B. Anodic-Stripping Voltammetry of Heavy-Metals with a Flow-Injection System. *Anal. Chim. Acta* **1983**, *146*, 45–50.
16. Stulik, K.; Pacakova, V. *Electroanalytical Measurements in Flowing Liquids*; Ellis Horwood: Chichester, UK, 1987.
17. Tougas, T.P. Recent Developments in Electrochemical Detection for Flow Injection Analysis. *Int. Lab.* **1988**, *18*, 17–28.
18. Smith, W.F.; Ivaska, A. A Study of the Electrochemical Oxidation of Some 1,4-Benzodiazepines. *Analyst* **1985**, *110*, 1377–1379.
19. Forsman, U. Cathodic Stripping Voltammetric Determination of Trace Amounts of Penicillins. *Anal. Chim. Acta* **1983**, *146*, 71–86.
20. Sithole, B.B.; Guy, R.D. Separation of Nitroimidazoles by Reversed-Phase High-Pressure Liquid Chromatography. *Talanta* **1986**, *33*, 95–97.
21. Wang, J.; Peng, T.; Lin, M.S. Trace Measurements of Tetracyclines Using Adsorptive Stripping Voltammetry. *Bioelectrochem. Bioenerg.* **1986**, *15*, 147–156.
22. Wang, J.; Bonakdar, M.; Morgan, C. Voltammetric Measurement of Tricyclic Antidepressants Following Interfacial Accumulation at Carbon Electrodes. *Anal. Chem.* **1986**, *58*, 1024–1028.
23. Wang, J.; Freiha, B.A.; Deshmukh, B.K. Adsorptive Extractive Stripping Voltammetry of Phenothiazine Compounds at Carbon Paste Electrodes. *Bioelectrochem. Bioenerg.* **1985**, *14*, 457–467.
24. Ballantine, J.; Woolfson, A.D. The Application of Differential Pulse Voltammetry at the Glassy Carbon Electrode to Multivitamin Analysis. *J. Pharm. Pharmacol.* **1980**, *32*, 353–356.
25. Kontoyannis, C.G.; Antimisialis, S.G.; Douroumis, D. Simultaneous Quantitative Determination of Diazepam and Liposomes Using Differential Pulse Polarography. *Anal. Chim. Acta* **1999**, *391*, 83–88.
26. Zapardiel, A.; Bermejo, E.; Lopez, J.; Hernandez, L.; Gil, E. Electroanalytical Determinations of Halazepam—Study of Interaction with Human, Serum Albumin. *Microchem. J.* **1995**, *52*, 41–52.
27. Fernandez-Marcote, M.S.M.; Mochon, M.C.; Sanchez, J.C.J.; Perez, A.G. Electrochemical Reduction of Prilocaine as Its *N*-Nitrosamine Derivative At the Mercury Electrode. *Electroanalysis* **1998**, *10*, 492–496.
28. Reddy, G.V.S.; Reddy, S.J. Estimation of Cephalosporin Antibiotics by Differential Pulse Polarography. *Talanta* **1997**, *44*, 627–631.

BIBLIOGRAPHY

- Bersier, P.M.; Bersier, J. Polarographic, Voltammetric, and HPLC-EC of Pharmaceutically Relevant Cyclic Compounds. *Electroanalysis* **1994**, *6*, 171–191.
- Bond, A.M. *Modern Polarographic Methods in Analytical Chemistry*; Marcel Dekker, Inc.: New York, 1980.
- Hart, J.P. *Electroanalysis of Biologically Important Compounds*; Ellis Horwood: Chichester, UK, 1990.
- Hoffmann, H.; Volke, J. Polarographic Analysis in Pharmacy. In *Electroanalytical Chemistry*; Nurnberger, H.W., Ed.; John Wiley & Sons: New York, 1974.
- Smyth, W.F. *Polarography of Molecules of Biological Significance*; Academic Press: London, 1979.
- Smyth, W.F.; Woolfson, A.D. Drug Assays: The Role of Modern Voltammetric Techniques. *J. Clin. Pharm. Ther.* **1987**, *12*, 117–134.
- Wang, J. *Stripping Analysis*; Verlag Chemie: Deerfield Beach, 1985.

POLYMORPHISM: PHARMACEUTICAL ASPECTS

Harry G. Brittain

Center for Pharmaceutical Physics, Milford, New Jersey, U.S.A.

INTRODUCTION

It had been known since the middle of the 18th century that many substances could be obtained in more than one crystal form, and so the properties of these solids were studied to the fullest extent possible with the characterization tools (e.g., crystal morphology and melting phenomena) available at that time (1). Eventually the work of von Laue and Bragg on the diffraction of X-rays by crystalline solids led to the development of technology that could be used to directly study the structures of such materials and to provide the structural justification for the phenomenon that became known as polymorphism.

Among other things, it became established that the nature of the structure adopted by a given compound on crystallization would then exert a profound effect on the solid-state properties of that system. For a given material, the heat capacity, conductivity, volume, density, viscosity, surface tension, diffusivity, crystal hardness, crystal shape and color, refractive index, electrolytic conductivity, melting or sublimation properties, latent heat of fusion, heat of solution, solubility, dissolution rate, enthalpy of transitions, phase diagrams, stability, hygroscopicity, and rates of reactions were all affected by the nature of the crystal structure.

Subsequently, workers in pharmaceutically related fields realized that the solid-state property differences derived from the existence of alternate crystal forms could translate into measurable differences in properties of pharmaceutical importance (2). For instance, it was found that various polymorphs could exhibit different solubilities and dissolution rates, and these differences sometimes led to the existence of nonequivalent bioavailabilities for the different forms. Since then, it has become recognized that an evaluation of the possible polymorphism available to a drug substance must be thoroughly investigated early during the stages of development. In various compilations, it has been reported that polymorphic species are known for most drug substances (3, 4) and that one should be surprised to encounter a compound for which only one structural type can be formed.

THEORETICAL CONCEPTS

The full specification of a polymorphic system is specified by the thermodynamic properties of the phases involved. A solid phase has a uniform structure and composition throughout, is separated by other phases by defined boundaries, and undergoes a phase transition when a particular solid phase becomes unstable under a given set of environmental conditions. The course of these phase changes is dictated by differences in free energy at the transition that are associated with structural or compositional changes. Classical thermodynamics provides a basis for understanding the nature of these transitions.

During a phase transition, the free energy of the system remains continuous, while the entropy, volume, and heat capacity undergo discontinuous changes. Phase transitions are classified as being of the same order as the derivative of the Gibbs free energy that exhibits a discontinuous change at the transition. Gibbs free energy (G) is defined from enthalpy (H) and entropy (S) by

$$\begin{aligned} G &= H - TS \\ &= E + PV - TS \end{aligned}$$

where T is the absolute temperature, P is the pressure, V is the volume, and E is the energy. It can easily be shown that transformations in which a discontinuous change occurs in volume or entropy (i.e., requiring a latent heat of transformation) will belong to the first order, while those involving a discontinuous change in heat capacity, thermal expansivity, or compressibility will belong to the second order.

The classical Clapeyron equation adequately predicts the features of first-order phase transitions, and this has been established for a number of examples of first-order transitions effected by the deliberate variation of temperature or pressure (5). Second- or higher-order transitions are not readily explained by classical thermodynamics. Unlike the case of first-order transitions, where the free-energy surfaces of the two phases intersect sharply at the transition temperatures, it is difficult to visualize the nature of the free-energy

surfaces in second- or higher-order transitions. In second-order transitions, changes in heat capacity as well as compressibility and thermal expansivity can be detected at the transition temperature, complicating the analysis.

Under a given set of environmental conditions, the most stable polymorph will be the one having the lowest free energy ($G'_{\text{metastable}} > G_{\text{stable}}$), and all metastable forms must eventually transform to the most stable form. If some combination of P and T exists so that $G'_{\text{metastable}} = G_{\text{stable}}$, then a reversible phase transition may take place and one terms this situation *enantiotropy*. If however, $G'_{\text{metastable}} > G_{\text{stable}}$ at all values of P and T , then any process that converts the metastable form into the stable form must be irreversible. This situation is termed *monotropy*.

One generally finds, therefore, that absolute values for thermodynamic parameters are less important than are relationships that predict the relative stability of the various phases of a polymorphic system. Although it is possible to calculate such energy differences from considerations of the lattice energies of the different structures (6), most workers instead employ the time-honored empirical rules that have been developed over time (7). For instance, since $G'_{\text{metastable}} > G_{\text{stable}}$, then the vapor pressure of the stable form must be less than the vapor pressure of the metastable form.

A number of empirical rules have been proposed to deduce the relative order of stability of polymorphs and the nature of the process that interconverts these (i.e., enantiotropy vs. monotropy). Among the better known are the *Heat of Transition Rule*, which states that if an endothermic transition is observed at some temperature, it may be assumed that there must be a transition point located at a lower temperature where the two forms bear an enantiotropic relationship. Conversely, if an exothermic transition is noted at some temperature, it may be assumed that there is no transition point located at a lower temperature. This in turn implies that either the two forms bear a monotropic relationship to each other or that the transition temperature is higher than the temperature of the exotherm.

Another empirical rule is the *Heat of Fusion Rule*, which states that if the higher melting form has a lower heat of fusion relative to the lower melting form, then the two forms bear an enantiotropic relationship. Less well obeyed is the *Density Rule*, which states that the most dense form will be the most stable at absolute zero. Strictly speaking, the Density Rule is only properly applied to polymorphs of molecular solids where intramolecular hydrogen bonding is not a significant factor.

STRUCTURAL ASPECTS

An ideal crystal is constructed by the regular spatial repetition of identical structural units. One defines the symmetry properties of crystals in terms of a periodic *lattice*, or a 3D grid of lines connecting points in a given structure. The repetitive motif is termed the *unit cell* (which will typically contain a group of molecules), each of which is located at a lattice point. The unit cell is defined by the magnitude of its projections (a , b , and c) along the crystal axes and the angles between the cell axes (α , β , and γ). The symmetry of a crystal is ultimately summed up in its crystallographic *space group*, which is the entire set of symmetry operations that define the periodic structure of the crystal.

When considering the structures of organic molecules, one finds that different modifications can arise in two main distinguishable ways. Should the molecule be constrained to exist as a rigid grouping of atoms, these may be stacked in different motifs to occupy the points of different lattices. This type of polymorphism is then attributable to packing phenomena and so is termed *packing polymorphism*. On the other hand, if the molecule in question is not rigidly constructed and can exist in distinct conformational states, then it can happen that each of these conformationally distinct modifications may crystallize in its own lattice structure. This latter behavior has been termed *conformational polymorphism* (8).

Packing Polymorphism

During the very first series of studies using single-crystal X-ray crystallography to determine the structures of organic molecules, Robertson reported the structure of resorcinol (1,3-dihydroxybenzene) (9). This crystalline material corresponded to that ordinarily obtained at room temperature, and was later termed the α -form. Shortly thereafter, it was found that the α -form underwent a transformation into a denser crystalline modification (denoted as the β -form) when heated at about 74°C and that the structure of this newer form was completely different (10). The crystal structures of the α - and β -forms (viewed down the c -axis, or (001) crystal plane) are found in Fig. 1.

By its nature, resorcinol is locked into a single conformation, and it is immediately evident from a comparison of the structures in Fig. 1 that each form is characterized by a different motif of hydrogen bonding. In particular, the α -form features a relative open architecture that is maintained by a spiraling array of hydrogen bonding that ascends through the various planes of

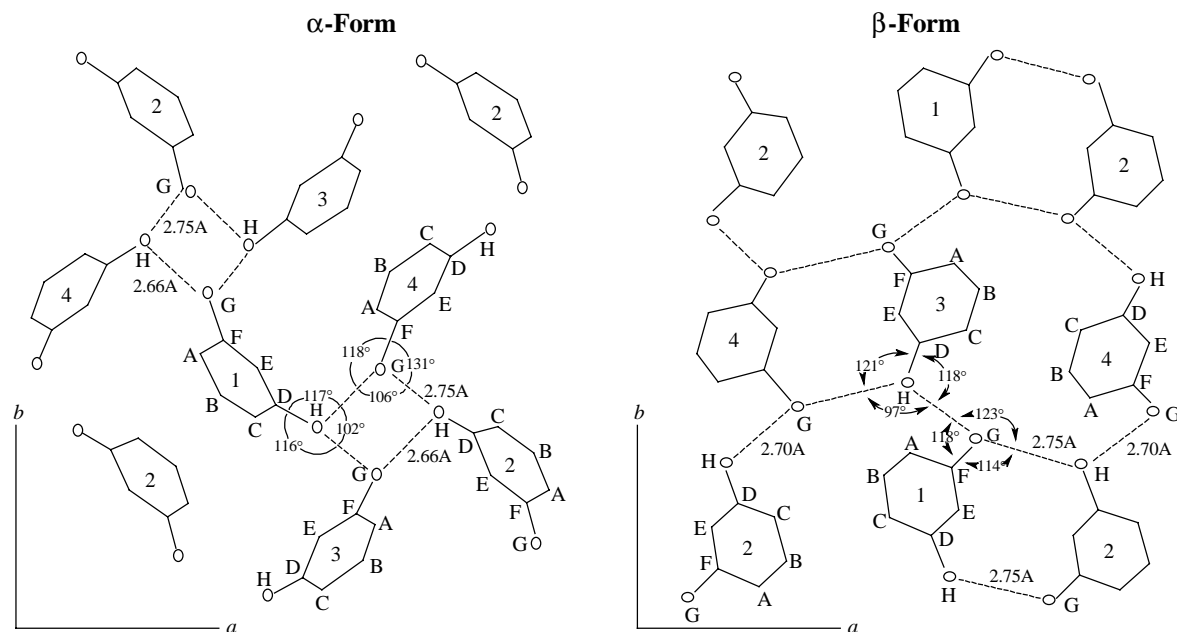


Fig. 1 Crystal structures of the α - and β -forms of resorcinol, as viewed down the c -axis (001 plane). (Adapted from Refs. 9 and 10.)

the crystal. In the view illustrated (defined by the ab -plane), the apparent closed tetramic grouping of hydroxyl groups is actually a slice through the ascending spiral. The effect of the thermally induced phase transformation is to collapse the open arrangement of the α -form by a more compact and parallel arrangement of the molecules in the β -form. This structural change causes an increase in crystal density on passing from the α -form (1.278 g/cm³) to the β -form (1.327 g/cm³). In fact, the molecular packing existing in the β -form was described as being more typical of hydrocarbons than of a hydroxylic compound (10).

Conformational Polymorphism

Probucol(4,4'-[(1-methylethylidene)bis(thio)]-bis-[2,6-bis(1,1-dimethylethyl)phenol]) is a cholesterol-lowering drug that has been reported to exist in two forms (11). Form II has been found to exhibit a lower melting point onset relative to Form I, and samples of Form II spontaneously transform to Form I upon long-term storage. The structures of these two polymorphic forms have been reported, and detailed views of the crystal structures are given in Fig. 2.

The conformations of the probucol molecule in the two forms were found to be quite different. In Form II, the CSCSC chain is extended, and the molecular symmetry

approximates C_{2v} . This molecular symmetry is lost in the structure of Form I, where now the torsional angles around the two CS bonds deviate significantly from 180°. Steric crowding of the phenolic groups by the t -butyl groups was evident from deviations from trigonal geometry at two phenolic carbons in both forms. Using a computational model, the authors found that the energy of Form II was 26.4 kJ/mol higher than the energy of Form I, indicating the less symmetrical conformer to be more stable. The crystal density of Form I was found to be approximately 5% higher than that of Form II, indicating that the conformational state of the Probucol molecules in Form I yielded more efficient space filling.

Solvatomorphism

One may define a solvatomorph as a crystalline solid in which solvent molecules have become included in the structure through the existence of positional substitution at positions that are site specific and that are related to other solvent molecules through translational symmetry. Other types of structural solvation exist (12) but will not be discussed here. Since water is such a ubiquitous substance, it is not surprising that the most important type of solvatomorphism involves the incorporation of water into a crystal lattice.

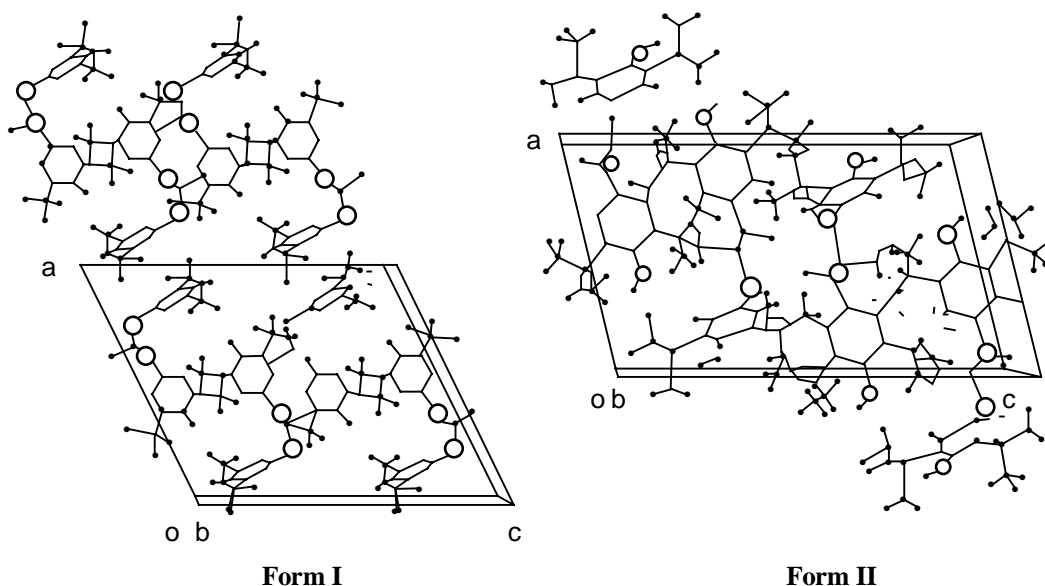


Fig. 2 Crystal structures of Forms I and II of Probucol. (Adapted from Ref. 11.)

Ampicillin (4-thia-1-azobicyclo[3.2.0]heptane-2-carboxylic acid) is an antibacterial agent that has been found to crystallize in one trihydrate and at least two anhydrate forms, and the structures of these have been critically compared (13). The transition temperature for the two forms in the presence of water has been found to be 42°C, where the trihydrate forms when crystallization is conducted below this value, and the anhydrate forms when the crystallization is effected at temperatures exceeding 42°C. Structures for the two solvatomorphs are shown in Fig. 3.

The ampicillin molecule exists as a zwitterion in both forms, with the overall molecular configuration being fairly similar as well. The structural differences induced by the presence of the water molecules in the trihydrate phase are evident in the differing configurations of the respective thiazolidine rings, where more planarity is found for the trihydrate phase than is found for the anhydrate phase. It was deduced that the three water molecules were extensively involved in the hydrogen bonding, and that they were located in a channel that lay parallel to one of the molecular screw axes. The intricate network of hydrogen bonding was invoked to explain the relative difficulty associated with dehydration of the trihydrate phase. Finally, the molecular packing of the two forms was judged to be so completely different that the authors concluded that there would be no way for the trihydrate phase to convert to the anhydrate phase as a pure solid–solid transition.

GENERATION OF POLYMORPHS

It is essential to determine the range of crystalline forms that are accessible to a potential drug substance and to determine which of the various forms will be the one used in products used in pivotal trials. To answer this question, investigators must conduct whatever studies might be required to evaluate the full range of possible polymorphs and solvatomorphs. The situation can be further complicated by the phenomenon of disappearing polymorphs, where metastable crystal forms become impossible to produce once more stable forms are uncovered (14).

Ideally, development programs devise a screening protocol for the discovery and preparation of any and all solid-state forms of chemical entities that may exist. Such protocols do not absolutely exclude the discovery of additional forms at later stages of development (i.e., during scale-up), but such approaches provide a comfort level regarding the level of knowledge and awareness regarding the scope of crystalline forms that may exist. A detailed exposition of the means available for the generation of polymorphs and solvates is available (15).

The first and primary method for production of polymorphs entails slow solvent evaporation of saturated solutions, with the rate of evaporation being adjusted by empirical means. Examples of solvents routinely used for such work are listed in Table 1 together with their boiling

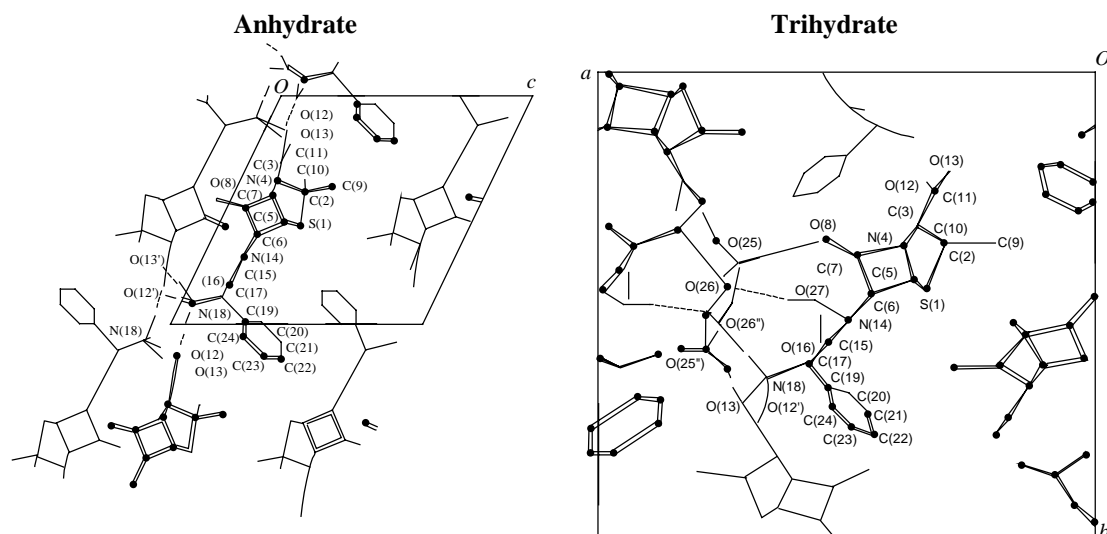


Fig. 3 Crystal structures of the anhydrate (viewed down the *b*-axis) and trihydrate (viewed down the *c*-axis) solvatomorphs of ampicillin. (Adapted from Ref. 13.)

Table 1 Solvents routinely used to isolate compound polymorphs and solvates

Solvent	Boiling point (°C)
Dipolar aprotic	
Dimethyl formamide	153
Acetonitrile	81
Dimethyl sulfoxide	189
<i>N</i> -methyl pyrrolidinone	80
Protic	
Water (various pH values)	100
Methanol	65
Acetic acid	115
Ethanol	78
<i>i</i> -Propanol	82
<i>n</i> -Propanol	97
<i>n</i> -Butanol	118
Lewis acidic	
Dichloromethane	40
Chloroform	61
Lewis basic	
Acetone	56
2-Butanone	80
Tetrahydrofuran	66
Ethyl acetate	77
Methyl butyl ether	56
Aromatic	
Toluene	111
Xylene	140
Pyridine	115
Nonpolar	
Hexane	69
Cyclohexane	81

points. The process of solution-mediated transformation can be considered the result of two separate events, beginning with dissolution of the initial phase, and completing with nucleation and growth of the final, stable phase. If two polymorphs differ in their melting points by 25–50°C, for monotropic polymorphs the lower melting, more soluble, form will be difficult to crystallize. The smaller the difference between the two melting points, the easier can it be to obtain the unstable or metastable forms.

Another commonly used crystallization method involves controlled changes in temperature. Slow cooling of a hot, saturated solution can be effective in producing crystals if the compound is more soluble at higher temperatures, while slow warming can be used if the compound is less soluble at higher temperatures. Sometimes it is preferable to heat the solution to boiling, filter to remove excess solute, then quench cool using an ice bath or even a dry ice-acetone bath.

There are situations where kinetics determines the course of crystallization, and thermodynamics becomes of secondary consideration. For example, Ostwald's Law of Stages states that, "when leaving an unstable state, a system does not seek out the most stable state, rather the nearest metastable state which can be reached with loss of free energy." This form then transforms to the next most soluble form through a process of dissolution and crystallization. For crystals whose formation is dominated by kinetic factors, it is essential to isolate the metastable form from the crystallization solvent by rapid filtration so that subsequent phase transformation would not occur.

During the characterization of solids obtained from solvent crystallization studies, one finds that thermal

treatment may be a means to produce new crystal forms. For instance, when using differential scanning calorimetry as an analysis technique, one can observe an endothermic peak corresponding to a phase transition, followed by a second endothermic peak corresponding to melting. Sometimes there is an exothermic peak between the two endotherms, representing a crystallization step. In these cases it is often possible to prepare the higher melting polymorph by thermal treatment.

In accordance with Ostwald's rule, the cooling of melts of polymorphic substances ordinarily yields the least stable modification, which subsequently rearranges into the stable modification in steps. Since the metastable form will have the lower melting point, it follows that supercooling is necessary to crystallize it from the melt. After melting, the system must be supercooled below the melting point of the metastable form, while at the same time the crystallization of the more stable form or forms must be prevented. Quench cooling a melt can sometimes result in formation of an amorphous solid that on subsequent heating undergoes a glass transition followed by crystallization.

Substances often crystallize containing water or solvent molecules located at specific sites in the crystal lattice, defining new crystalline forms known as solvatomorphs. Since water is a pharmaceutically acceptable solvent, hydrate species are of primary importance to drug development. The variety of hydrates that can exist has been summarized (12). Most solvatomorphs form with an integral number for the solvent/molecule ratio, but this is not always the case.

In the simplest type, water is bound to inorganic cations as part of a coordination complex. This type of water is denoted as water of crystallization and is common for inorganic compounds. For example, nickel sulfate forms a well-defined hexahydrate, where the waters of hydration are bound directly to the Ni(II) ion. Well-defined multiple hydrate species can also form with organic molecules, where the water molecules bridge unit cells in the overall structure. Finally, water molecules can exist in a semispecific manner, lining cavities within the crystal structure. This last hydrate type is often termed a channel hydrate.

Typically, hydrates are obtained by recrystallization from water. For example, trazodone hydrochloride tetrahydrate was prepared by dissolving the anhydrate in hot distilled water, allowing the solution to remain at room temperature overnight, and storing the collected crystals at 75% relative humidity and 25°C until they reached constant weight (16). Hydrates can sometimes be obtained by suspending the anhydrous material in water, a process that is analogous to Ostwald ripening. For instance,

aqueous suspensions of anhydrous metronidazole benzoate are metastable, and storage at temperatures lower than 38°C leads to monohydrate formation accompanied by crystal growth (17). The exposure of an anhydrous powder to high relative humidity often yields the formation of new hydrate forms. For example, the experimental anticholesterol compound SQ-33600 was found to form a multitude of hydrate forms on exposure to various relative humidity environments (18).

METHODS OF CHARACTERIZATION

Once a variety of crystalline solids have been produced using a suitable polymorph protocol, it is very important to characterize these by proper techniques so that the system can become better defined. Fortunately, extensive discussions of the techniques suitable for the characterization of pharmaceutical solids are available (19, 20). The fruits of the most important characterization technique, single-crystal X-ray diffraction, have already been discussed in connection with the phenomenon of polymorphism in an earlier section. Certainly the determination of the crystal structures of all possible polymorphs and solvatomorphs would constitute the ultimate in characterization, especially if one could also conduct spectroscopic and thermal investigations on the same crystals used for the structural determination. This situation is rarely realized, so a series of additional techniques are ordinarily brought into use.

Of all the methods available for the physical characterization of solid materials, it is generally agreed that crystallography, microscopy, thermal analysis, solubility studies, vibrational spectroscopy, and nuclear magnetic resonance are the most useful for characterization of polymorphs and solvates. However, it cannot be overemphasized that the defining criterion for the existence of polymorphic types must always be a nonequivalence of crystal structures. For compounds of pharmaceutical interest, this ordinarily implies that a nonequivalent X-ray powder diffraction pattern is observed for each suspected polymorphic variation. All other methodologies must be considered as sources of supporting and ancillary information, but cannot be taken as definitive proof for the existence of polymorphism by themselves.

A correctly prepared sample of a powdered solid will present an entirely random selection of all possible crystal faces at the powder interface, and the diffraction off this surface provides information on all possible atomic spacings in the crystal lattice (21). The relationship between observed scattering angles and the spacings

between planes of molecules in the lattice consists of Bragg's Law:

$$n\lambda = 2d \sin \theta$$

where n is the order of the diffraction line, λ is the wavelength of the incident X-ray beam, d is the distance between the planes in the crystal, and θ is the observed angle of beam diffraction. To measure a powder pattern, a randomly oriented sample is prepared so as to expose all the planes of a sample, irradiated with monochromatic X-ray radiation, and the angles measured at which coherent scattering of X-rays is observed.

An extremely important tool for the characterization of polymorphs and solvates is that of microscopy, since the observable habits of differing crystal structures must necessarily be different and therefore useful for the characterization of such systems (22). Clearly, visual observation of materials suspected of being polymorphs or solvatomorphs would immediately follow their crystallographic study, which in turn would make the science of optical crystallography (23) an essential aspect of any program of study. Both optical and electron microscopies have found widespread use for the characterization of polymorphs and solvates. Although optical microscopy is more limited in the range of magnification suitable for routine work (working beyond 600 \times being difficult when observing microcrystalline materials), the use of polarizing optics introduces enormous power into the technique not available with other methods. Electron microscopy work can be performed at extraordinarily high magnification levels (up to 90,000 \times on most units), and the images that can be obtained contain a considerable degree of 3D information.

Often referred to as fusion microscopy or hot-stage microscopy, thermal microscopy can be an extremely valuable tool for the characterization of polymorphic or solvate systems. The technique requires that one make observations during the heating and cooling of a few milligrams of substance on a microscope slide, as well as observations on the crystallized material (24). It is therefore possible to conduct a very rapid analysis using only small quantities of material, and the entire phase diagram of a drug material can be deduced upon the conduct of suitably designed experiments. The most widely used device in the conduct of thermal microscopic studies is the hot state of Kofler, which has facilitated the conduct of an extraordinary number of studies (25).

Thermal analysis methods are defined as those techniques in which a property of the analyte is determined as a function of an externally applied temperature. Regardless of the observable parameter

measured, the usual practice requires that the physical property and the sample temperature are recorded continually and automatically and that the sample temperature is altered at a predetermined rate. Thermal reactions can be endothermic (melting, boiling, sublimation, vaporization, desolvation, solid-solid phase transitions, chemical degradation, etc.) or exothermic (crystallization, oxidative decomposition, etc.) in nature. Such methodology has found widespread use in the pharmaceutical industry for the characterization of compound purity, polymorphism, solvation, degradation, and excipient compatibility (26).

Although a large number of thermal analysis techniques have been developed, the most commonly applied are those of thermogravimetry (TG, the measure of thermally induced weight loss of a material as a function of applied temperature), differential thermal analysis (DTA, the difference in temperature existing between a sample and a reference as a function of temperature), and differential scanning calorimetry (DSC, the difference in heat capacity between the sample and a reference as a function of temperature). The primary applicability of DTA and DSC analysis to the study of polymorphs and solvatomorphs has been to obtain information about any phase transformations that take place as a function of temperature.

The simplest and most straightforward application of thermal analysis is concerned with studies of the relative stability of polymorphic forms. For example, DTA thermograms enabled the deduction that one commercially available form of chloroquine diphosphate was phase pure, while another consisted of a mixture of two polymorphs (27). DTA analysis was used to demonstrate that in spite of the fact that different crystal habits of sulfamethazine could be obtained, these in fact consisted of the same anhydrous polymorph (28). In a study aimed at profiling the dissolution behavior of the three polymorphs and five solvates of spironlactone, DTA analysis was used in conjunction with powder X-ray diffraction to establish the character of the various materials (29).

DSC analysis can also be used to obtain temperatures of compound melting, and such information can be of value in establishing the relative orders of stability in polymorphic systems. In addition, for suitable systems the technique can be used to study any phase interconversion that takes place during the DSC study. For instance, Form I of iopanoic acid yielded a single melting endotherm at 154°C, but the thermogram obtained on Form II was much more complicated (30). Form II was found to exhibit one endotherm at 133°C (the melting transition of Form I), an exotherm at 141°C (crystallization to Form II), and another endotherm at 153°C (melting of the recrystallized Form II).

DSC analysis represents a superior method of thermal analysis, in that the area under a DSC peak is directly proportional to the heat absorbed or evolved by the thermal event, and integration of these peak areas yields the enthalpy of reaction (in units of calories/gram or Joules/gram). Even though conclusions reached on the basis of enthalpies of fusion are possibly compromised by their omission of the entropy contribution, an indication of the thermodynamic trends inherent in the system is often possible. For instance, the same polymorphic form of moricizine hydrochloride was deduced on the basis of thermal analysis and equilibrium solubility measurements (31). On the other hand, auranofin represents a compound for which one anhydrous polymorphic form is predicted to be the most stable by virtue of its melting point and heat of fusion but for which solubility measurements demonstrate that the other polymorph was in fact the thermodynamically stable form (32).

The energies associated with the vibrational modes of a chemical compound can be observed directly through their absorbance in the infrared region of the spectrum or through the observation of the low-energy scattered bands that accompany the passage of an intense beam of light through the sample (the Raman effect). When the vibrational modes associated with the molecules in a polymorphic system are perturbed by features of the different crystal structures, these methods can be used in the spectroscopic investigation of polymorphs and solvates (33).

When the FTIR spectra of polymorph systems differ substantially, the results may readily permit the identification of a particular form. For instance, the two forms of ranitidine hydrochloride yielded spectra that differed in the region above 3000 cm^{-1} and in the regions spanning $2300\text{--}2700\text{ cm}^{-1}$ and $1570\text{--}1620\text{ cm}^{-1}$ (34). Zanolone has been found to crystallize in a number of different forms, each of which yields a characteristic infrared spectrum (35). When solvent molecules are incorporated in a crystal lattice, the new structure is often sufficiently different from that of the anhydrous phase so that many of the molecular vibrational modes are altered.

The vibrational modes of a compound may also be studied using Raman spectroscopy, where one measures the inelastic scattering of radiation by a nonabsorbing medium (36). Although both infrared absorption and Raman scattering yield information on the energies of the same vibrational bands, the different selection rules governing the band intensities for each type of spectroscopy can be exploited by the skillful worker. For instance, both types of vibrational spectroscopy were used to investigate the polymorphism of nimodipine, and it was evident from the intensity relations that, although each

technique yielded a summary of the vibrational transitions, substantial differences in band intensity were readily discernible (37).

One technique that is becoming increasingly important for the characterization of materials is that of solid-state nuclear magnetic resonance (NMR) spectroscopy, and the application of this methodology to topics of pharmaceutical interest has been amply demonstrated (33, 38). The NMR spectra of polymorphs or solvatomorphs often contains nonequivalent resonance peaks for analogous nuclei since the intimate details of the molecular environments associated with differing crystal structures can lead to perturbations in nuclear resonance energies.

In its simplest application, solid-state NMR spectra can be used to qualitatively differentiate between polymorphs or solvates, much in the manner described for vibrational spectroscopy. When detailed assignments of solid-state spectra have been made, the technique can be used to deduce differences in molecular conformation, which cause crystallographic variations to exist. During the development of fosinopril sodium, a crystal structure was solved for the most stable phase, but no such structure could be obtained for its metastable phase (39). Studies of the solid-state vibrational and NMR spectroscopies permitted the deduction that the solid-state polymorphism was associated with different conformations of an acetal sidechain. The NMR data also suggested that additional conformational differences between the two polymorphs were associated with *cis-trans* isomerization along the peptide bond, which in turn resulted in the presence of non-equivalent molecules existing in the unit cell.

The solid-state NMR technique can be used to deduce quantitative measurements of phase composition, as has been reported for the anhydrate and dihydrate phases of carbamazepine (40). The applications of solid-state ^{13}C -NMR spectra for the study of polymorphs and solvates can go beyond evaluations of resonance band positions, making use of additional spectral characteristics. For instance, studies of $T_{1\rho}$ relaxation times of furosemide polymorphs were used to show the presence of more molecular mobility and disorder in Form II, while the structure of Form I was judged to be more rigid and uniformly ordered (41).

POLYMORPHISM AND SOLUBILITY

Since the different lattice energies (and entropies) associated with different polymorphs or solvatomorphs give rise to measurable differences in a large variety of physical properties, it is hardly surprising that a family of

different forms should exhibit different solubilities and dissolution rates (42). These varying solubilities can in turn be very important during the processing of drug substances into drug products (43) and may have implications for the adsorption of the active drug from its dosage form (44). A solid having a higher lattice free energy (i.e., a less stable polymorph) will tend to dissolve faster, since the release of a higher amount of stored lattice free energy will increase the solubility and hence the driving force for dissolution. At the same time, each species would liberate (or consume) the same amount of solvation energy, since all dissolved species (of the same chemical identity) must be thermodynamically equivalent. The varying dissolution rates possible for different structures of the same drug entity can in turn lead to varying degrees of bioavailability for different polymorphs or solvates.

Solubility determinations were used to characterize the polymorphism of 3-(((3-(2-(7-chloro-2-quinolinyl)-(E)-ethenyl)-phenyl)-((3-dimethylamino-3-oxopropyl)-thio)-methyl)-thio)-propanoic acid (45). The solubility of Form II was found to be higher than that of Form I in both isopropyl alcohol (IPA, solubility ratio equal to 1.7 over the range of 5–55°C) and methyl ethyl ketone (MEK, solubility ratio equal to 1.9 over the range of 5–55°C), indicating that Form I is the thermodynamically stable form over this temperature range. An analysis of the entropy contributions to the free energy of solution from the solubility results implied that the saturated IPA solutions were more disordered than were the corresponding MEK solutions, in turn indicating the existence of stronger solute–solvent interactions in the MEK solution.

Phenylbutazone has been found to be capable of existing in five different polymorphic structures, characterized by different X-ray powder diffraction patterns and melting points (46). While Form I exhibited the highest melting point (suggesting the least energetic structure at the elevated temperature), its equilibrium solubility was the lowest in each of the three solvent systems studied (demonstrating the lowest free energy). These findings indicate that Form I is the thermodynamically most stable polymorph both at room temperature and at the melting point (105°C). Identification of the sequence of stability for the other forms at any particular temperature was not straightforward, and it was concluded that the order of stability at room temperature was not the same as that at 100°C. This clearly indicates that some of the forms are enantiotropically related and that others are related by monotropism.

When the hydrates or solvates of a given compound are stable with respect to phase conversion in a solvent, the equilibrium solubility of these species can be used to

characterize these systems. For instance, the equilibrium solubility of the trihydrate phase of ampicillin at 50°C is approximately 1.3 times that of the more stable anhydrate phase at room temperature (47). However, below the transition temperature of 42°C, the anhydrate phase is more soluble and is therefore less stable.

Solution calorimetry can be used on one level to merely obtain the enthalpy of solution for a given solute, or can be used in a deeper sense to obtain a full thermodynamic description of a system. The determination of solubility data over a defined temperature range can be used to calculate the differential heat of solution of a given polymorphic form. One can subtract the differential heats of solution obtained for the two polymorphs to deduce the heat of transition (ΔH_{Trans}) between the two forms:

$$\Delta H_{\text{Trans}} = \Delta H_{\text{S}}^{\text{B}} - \Delta H_{\text{S}}^{\text{A}}$$

where $\Delta H_{\text{S}}^{\text{A}}$ and $\Delta H_{\text{S}}^{\text{B}}$ denote the differential heats of solution for polymorphs A and B, respectively. The validity of the assumption regarding constancy in the heats of solution for a given substance with respect to temperature can be made by determining the enthalpy of fusion (ΔH_{F}) for the two forms, and then taking the difference between these:

$$\Delta H_{\text{Trans}} = \Delta H_{\text{F}}^{\text{B}} - \Delta H_{\text{F}}^{\text{A}}$$

where ΔH_{Trans} represents the heat of transition between forms A and B at the melting point. When a sufficient number of assumptions are made, one deduces that ΔH_{Trans} and ΔH_{Trans} should be equal, and thus DSC results can be used to verify the solution calorimetry results. For example, the heats of fusion and solution have been reported for the polymorphs of auranofin (32). The similarity of the heats of transition deduced in 95% ethanol (2.90 kcal/mol) and dimethylformamide (2.85 kcal/mol) with the heat of transition calculated at the melting point (3.20 kcal/mol) provides a fair estimation of the thermodynamics associated with this polymorphic system.

To illustrate the importance of free energy changes, consider the solvate system formed by paroxetine hydrochloride, which can exist as a nonhygroscopic hemihydrate or as a hygroscopic anhydrate (48). The heat of transition between these two forms was evaluated both by DSC ($\Delta H_{\text{Trans}} = 0.0$ kJ/mol) and by solution calorimetry ($\Delta H_{\text{Trans}} = 0.1$ kJ/mol), which indicates that both forms are isoenthalpic. However, the free energy of transition (–1.25 kJ/mol) favors conversion of the anhydrate to the hemihydrate, and such phase conversion can be initiated by crystal compression or by seeding techniques. Since the two forms are essentially isoenthalpic, the entropy increase that accompanies the phase

transformation is responsible for the decrease in free energy and may therefore be viewed as the driving force for the transition.

A basic thermodynamic understanding of a polymorphic system requires a determination of the free energy difference between the various forms. The two polymorphs of 3-amino-1-(*m*-trifluoromethylphenyl)-6-methyl-1*H*-pyridazin-4-one have been characterized by a variety of methods, among which solubility studies were used to evaluate the thermodynamics of the transition from Form I to Form II (49). At a temperature of 30°C, the enthalpy change for the phase transformation was determined to be -5.64 kJ/mol. From the solubility ratio of the two polymorphs, the free energy change was then calculated as -3.67 kJ/mol, which implies that the entropy change accompanying the transformation was -6.48 cal/Kmol. In this system, one encounters a phase change that is favored by the enthalpy term but not favored by the entropy term. However, since the overall free energy change (ΔG_{Trans}) is negative, the process takes place spontaneously, provided that the molecules can overcome the activation energy barrier at a significant rate.

In other cases, an unfavorable enthalpy term was found to be compensated by a favorable entropy term, thus rendering negative the free energy change associated with a particular phase transformation. Lamivudine can be obtained in two forms, one of which is a 0.2-hydrate obtained from water or from methanol that contains water and the other of which is nonsolvated and is obtained from many nonaqueous solvents (50). Form II was determined to be thermodynamically favored in the solid state. Solubility studies of both forms as a function of solvent and temperature were used to determine whether entropy of enthalpy was the driving force for solubility. Solution calorimetric data indicated that Form I would be favored in all solvents studied on the basis of enthalpy alone. In higher alcohols and other organic solvents, Form I exhibited a larger entropy of solution than did Form II, compensating for the unfavorable enthalpic factors and yielding an overall negative free energy for the phase change.

It is generally recognized that studies of dissolution rate are best conducted on compacted materials, where the process of forming the compact regulates the particle size and surface area of the solid. Such work yields the intrinsic dissolution rate, the trends of which usually parallel those deduced from studies of equilibrium solubility. Since under constant hydrodynamic conditions the intrinsic dissolution rate is proportional to the solubility of the dissolving solid, the most stable polymorphic form will exhibit the slowest intrinsic dissolution rate.

REFERENCES

1. Verma, A.R.; Krishna, P. *Polymorphism and Polytypism in Crystals*; John Wiley & Sons: New York, 1966; 1-7.
2. Brittain, H.G. *Polymorphism in Pharmaceutical Solids*; Marcel Dekker, Inc.: New York, 1999; 331-361.
3. Borka, L.; Haleblan, J.K. Crystal Polymorphism of Pharmaceuticals. *Acta Pharm. Jugosl.* **1990**, *40*, 71-94.
4. Borka, L. Review on Crystal Polymorphism of Substances in the European Pharmacopoeia. *Pharm. Acta Helv.* **1991**, *66*, 16-22.
5. Ubbelohde, A.R. *Reactivity of Solids*; de Boer, J.H., Ed.; Elsevier: Amsterdam, 1961; 141-168.
6. Mooij, W.T.M.; van Eijck, B.P.; Kroon, J. Ab Initio Crystal Structure Predictions for Flexible Hydrogen-Bonded Molecules. *J. Am. Chem. Soc.* **2000**, *122*, 3500-3505.
7. Burger, A.; Ramberger, R. On the Polymorphism of Pharmaceuticals and Other Molecular Crystals. I. Theory of Thermodynamic Rules. *Mikrochim. Acta (Wien)* **1979**, 259-271, II. Applicability of Thermodynamic Rules. *Ibid.* **1979**, 273-316.
8. Bernstein, J. Conformational Polymorphism. *Organic Solid State Chemistry*; Desiraju, G.R., Ed.; Elsevier: Amsterdam, 1987; 471-518.
9. Robertson, J.M. The Structure of Resorcinol: A Quantitative X-Ray Investigation. *Proc. R. Soc. London* **1936**, *A157*, 79-99.
10. Robertson, J.M.; Ubbelohde, A.R. A New Form of Resorcinol. I. Structure Determination by X-Rays. *Proc. R. Soc. London* **1938**, 122-135.
11. Gerber, J.J.; Caira, M.R.; Lötter, A.P. Structures of Two Conformational Polymorphs of the Cholesterol-Lowering Drug Probucol. *J. Cryst. Spect. Res.* **1993**, *23*, 863-869.
12. Morris, K.R.; Rodriguez-Hornado; Hydrates, N. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J., Eds.; Marcel Dekker, Inc.: New York, 1993; 7, 393-440.
13. Boles, M.O.; Girven, R.J. The Structures of Ampicillin: A Comparison of the Anhydrate and Trihydrate Forms. *Acta Cryst.* **1976**, *B32*, 2279-2284.
14. Dunitz, J.D.; Bernstein, J. Disappearing Polymorphs. *Acc. Chem. Res.* **1995**, *28*, 193-200.
15. Guillory, J.K. Generation of Polymorphs, Hydrates, Solvates, and Amorphous Solids. *Polymorphism in Pharmaceutical Solids*; Brittain, H.G., Ed.; Marcel Dekker, Inc.: New York, 1999; 183-226.
16. Sasaki, K.; Suzuki, H.; Nakagawa, H. Physicochemical Characterization of Trazodone Hydrochloride Tetrahydrate. *Chem. Pharm. Bull.* **1993**, *41*, 325-328.
17. Caira, M.R.; Nassimbeni, L.R.; van Oudtshoorn, B. X-Ray Structural Characterization of Anhydrous Metronidazole Benzoate and Metronidazole Benzoate Monohydrate. *J. Pharm. Sci.* **1993**, *82*, 1006-1009.
18. Morris, K.R.; Newman, A.W.; Bugay, D.E.; Ranadive, S.A.; Singh, A.K.; Szyper, M.; Varia, S.A.; Brittain, H.G.; Serajuddin, A.T.M. Characterization of Humidity-Dependent Changes in Crystal Properties of a New HMG-CoA Reductase Inhibitor in Support of Its Dosage Form Development. *Int. J. Pharm.* **1994**, *108*, 195-206.
19. Brittain, H.G. Methods for the Characterization of Polymorphs and Solvates. *Polymorphism in Pharmaceutical*

- Solids*; Brittain, H.G., Ed.; Marcel Dekker, Inc.: New York, 1999; 227–278.
20. Threlfall, T.L. Analysis of Organic Polymorphs. *Analyst* **1995**, *120*, 2435–2460.
 21. Klug, H.P.; Alexander, L.E. *X-ray Diffraction Procedures*, 2nd Ed.; Wiley-Interscience: New York, 1974; 271–311.
 22. Halebian, J.K. Characterization of Habits and Crystalline Modifications of Solids and Their Pharmaceutical Applications. *J. Pharm. Sci.* **1975**, *64*, 1269–1288.
 23. Wahlstrom, E.E. *Optical Crystallography*, 4th Ed.; John Wiley & Sons: New York, 1969.
 24. McCrone, W.C. *Fusion Methods in Chemical Microscopy*; Interscience: New York, 1957.
 25. Kuhnert-Brandstätter, M. *Thermomicroscopy in the Analysis of Pharmaceuticals*; Pergamon Press: Oxford, 1971.
 26. Giron, D. Applications of Thermal Analysis in the Pharmaceutical Industry. *J. Pharm. Biomed. Anal.* **1986**, *4*, 755–770.
 27. Van Aerde, Ph.; Remon, J.P.; De Rudder, D.; Van Severen, R.; Braeckman, P. Polymorphic Behavior of Chloroquine Diphosphate. *J. Pharm. Pharmacol.* **1984**, *36*, 190–191.
 28. Maury, L.; Rambaud, J.; Pauvert, B.; Lasserre, Y.; Bergé, G. Physicochemical and Structural Study of Sulfamethazine. *J. Pharm. Sci.* **1985**, *74*, 422–426.
 29. Salole, E.G.; Al-Sarraj, H. Spironolactone Crystal Forms. *Drug Dev. Indust. Pharm.* **1985**, *11*, 855–864.
 30. Stagner, W.C.; Guillory, J.K. Physical Characterization of Solid Iopanoic Acid Forms. *J. Pharm. Sci.* **1979**, *68*, 1005–1009.
 31. Wu, L.-S.; Torosian, G.; Sigvardson, K.; Gerard, C.; Hussain, M.A. Investigation of Moricizine Hydrochloride Polymorphs. *J. Pharm. Sci.* **1994**, *83*, 1404–1406.
 32. Lindenbaum, S.; Rattie, E.S.; Zuber, G.E.; Miller, M.E.; Ravin, L.J. Polymorphism of Auranofin. *Int. J. Pharm.* **1985**, *26*, 123–132.
 33. Brittain, H.G. Spectral Methods for the Characterization of Polymorphs and Solvates. *J. Pharm. Sci.* **1997**, *86*, 405–412.
 34. Cholerton, T.J.; Hunt, J.H.; Klinkert, G.; Martin-Smith, M. Spectroscopic Studies on Ranitidine—Its Structure and the Influence of Temperature and pH. *J. Chem. Soc. Perkin Trans. 2* **1984**, 1761–1766.
 35. Rocco, W.L.; Morphet, C.; Laughlin, S.M. Solid-state Characterization of Zanolone. *Int. J. Pharm.* **1995**, *122*, 17–25.
 36. Grasselli, J.G.; Snavely, M.K.; Bulkin, B.J. *Chemical Applications of Raman Spectroscopy*; John Wiley & Sons: New York, 1981.
 37. Grunenber, A.; Keil, B.; Henck, J.-O. Polymorphism in Binary Mixtures, as Exemplified by Nimodipine. *Int. J. Pharm.* **1995**, *118*, 11–21.
 38. Bugay, D.E. Solid-State Nuclear Magnetic Resonance Spectroscopy: Theory and Pharmaceutical Applications. *Pharm. Res.* **1993**, *10*, 317–327.
 39. Brittain, H.G.; Morris, K.R.; Bugay, D.E.; Thakur, A.B.; Serajuddin, A.T.M. Solid State Characterization of Fosinopril Sodium Polymorphs. *J. Pharm. Biomed. Anal.* **1993**, *11*, 1063–1069.
 40. Suryanarayanan, R.; Wiedmann, T.S. Quantitation of the Relative Amounts of Anhydrous Carbamazepine and Carbamazepine Dihydrate in a Mixture by Solid-State Nuclear Magnetic Resonance. *Pharm. Res.* **1990**, *7*, 184–187.
 41. Doherty, C.; York, P. Frusemide Crystal Forms; Solid State and Physicochemical Analyses. *Int. J. Pharm.* **1988**, *47*, 141–155.
 42. Grant, D.J.W.; Higuchi, T. *Techniques of Chemistry. Solubility Behavior of Organic Compounds*; Saunders, W.H., Jr. Ed.; Doctoral Thesis John Wiley and Sons: New York, 1947; 21.
 43. Halebian, J.K.; McCrone, W.C. Pharmaceutical Applications of Polymorphism. *J. Pharm. Sci.* **1969**, *58*, 911–929.
 44. Higuchi, W.I.; Lau, P.K.; Higuchi, T.; Shell, J.W. Polymorphism and Drug Availability: Solubility Relationships in the Methylprednisolone System. *J. Pharm. Sci.* **1963**, *52*, 150–153.
 45. Ghodbane, S.; McCauley, J.A. Study of the Polymorphism of 3-((3-(2-(7-chloro-2-quinolinyl)-(E)-ethenyl)-phenyl)-((3-dimethylamino-3-oxopropyl)-thio)-methyl)-thio)-propanoic Acid by DSC, TG, XRPD, And Solubility Measurements. *Int. J. Pharm.* **1990**, *59*, 281–286.
 46. Tuladhar, M.D.; Carless, J.E.; Summers, M.P. Thermal Behavior and Dissolution Properties of Phenylbutazone Polymorphs. *J. Pharm. Pharmacol.* **1983**, *35*, 208–214.
 47. Poole, J.W.; Bahal, C.K. Dissolution Behavior and Solubility of Anhydrous and Trihydrate Forms of Ampicillin. *J. Pharm. Sci.* **1968**, *57*, 1945–1948.
 48. Buxton, P.C.; Lynch, I.R.; Roe, J.A. Solid-state Forms of Paroxetine Hydrochloride. *Int. J. Pharm.* **1988**, *42*, 135–143.
 49. Chauvet, A.; Masse, J.; Ribet, J.P.; Bigg, D.; Autin, J.M.; Maurel, J.M.; Patoiseau, J.F.; Jaud, J. Characterization of Polymorphs and Solvates of 3-Amino-1-(*m*-Trifluoromethylphenyl)-6-Methyl-LH-Pyridazin-4-One. *J. Pharm. Sci.* **1992**, *81*, 836–841.
 50. Jozwiakowski, M.J.; Nguyen, N.A.; Sisco, J.M.; Spancake, C.W. Solubility Behavior of Lamivudine Crystal Forms in Recrystallization Solvents. *J. Pharm. Sci.* **1996**, *85*, 193–199.

POTENTIOMETRY

Karel Vytras

University of Pardubice, Pardubice, Czech Republic

INTRODUCTION

Potentiometry is a method of electroanalytical measurement in which the equilibrium voltage of the cell consisting of an indicator electrode and a proper reference electrode is measured using a high-impedance voltmeter, i.e., effective at zero current. The potential of the indicator electrode is a function of particular species present in solutions and their concentration. By judicious choice of electrode material, the selectivity of the response to one of the species can be increased, and thus, interferences from other ions can be minimized. The method allows the determination of concentrations with detection limits of the order of 0.1 μmol per liter, although in some cases, as little as 10 pmol differences in concentration can be measured.

EQUILIBRIA AT INTERFACES

Electric potential is the electric work necessary to transfer the unit charge in vacuum from the infinite distance to a position, the potential of which is to be established. If this position is situated inside of a phase (metal, solution, etc.), it is called the inner electric potential and is denoted by ϕ . The chemical potential of an ion in the presence of an electric potential is called its electrochemical potential, $\tilde{\mu}$, expressed as in Eq. 1:

$$\tilde{\mu} = \mu + zF\phi \quad (1)$$

where μ is the chemical potential, z is the charge of the particle, and F is Faraday's constant ($96.487 \text{ C mol}^{-1}$).

When two phases containing electrically charged particles come into contact, an electrical potential difference develops at their interface. A description of the interface is therefore essential when investigating the charge transfer. If the system is in equilibrium, the appropriate electrochemical potentials must be equal. Thus, for the charged particle i present in phases 1 and 2, Eq. 2 must be valid.

$$\tilde{\mu}_i(1) = \tilde{\mu}_i(2) \quad (2)$$

This equation can be rewritten as in Eq. 3 in terms of the standard chemical potentials, μ_i^0 , and the activities, a_i ,

$$\begin{aligned} \mu_i^0(1) + RT \ln a_i(1) + z_i F \phi(1) \\ = \mu_i^0(2) + RT \ln a_i(2) + z_i F \phi(2) \end{aligned} \quad (3)$$

where R is the gas constant ($8.313 \text{ J K}^{-1} \text{ mol}^{-1}$) and T the absolute temperature (measured in K). Thus, Eq. 4 gives the potential difference at the interface,

$$\begin{aligned} \Delta\phi &= \phi(2) - \phi(1) \\ &= [\mu_i^0(1) - \mu_i^0(2)]/z_i F \\ &\quad + RT/z_i F \ln[a_i(1)/a_i(2)] \end{aligned} \quad (4)$$

which is the so-called Galvani potential difference, $\Delta\phi$.

ELECTRODE POTENTIALS

The potential difference $\Delta\phi$ given by Eq. 4 cannot be measured directly. An electrochemical cell is required, consisting of two electrode compartments (both first and second half-cells). Only in such a cell is the potential difference E measurable, as given by Eq. 5:

$$E = \Delta\phi_{\text{st}} - \Delta\phi_{\text{nd}} \quad (5)$$

If the second half-cell is taken as reference and combined with others, the same value $\Delta\phi_{2\text{nd}}$ occurs in all the differences in Eq. 5. This value can be chosen conventionally and set equal to zero (this amounts to a shift of the coordinate). The electrode for which the condition $\Delta\phi_{2\text{nd}} = 0$ is valid is the standard hydrogen electrode, in which hydrogen under the standard pressure [$p(\text{H}_2) = p^0 = 1.013 \times 10^5 \text{ Pa}$] is bubbled over the surface of the platinum black through a solution of hydrogen ions of unit activity, schematically written as $\text{Pt(s)}|\text{H}_2(\text{g})|\text{H}^+(\text{aq})$, where the vertical lines indicate the interphase. The reaction that determines the electrode potential is the so-called half-cell reaction, $2\text{H}^+(\text{aq}) + 2e = \text{H}_2(\text{g})$. Owing to this convention, the first term of Eq. 5, $\Delta\phi_{1\text{st}}$, can be substituted by E and is called the electrode potential. In the same manner, the first

expression on the right in Eq. 4 can be replaced by Eq. 6:

$$E^0 = [\mu_i^0(1) - \mu_i^0(2)]/z_i F \quad (6)$$

and is called the standard electrode potential (values of selected standard electrode potentials are listed in Table 1).

The simplest situation occurs if a metal (phase 1) is immersed in an electrolyte solution (phase 2), for example, $\text{Zn(s)}|\text{Zn}^{2+}(\text{aq})$. An electrode and its electrolyte make up a half-cell compartment or a so-called first-order electrode, the potential of which is given in Eq. 7:

$$E = E^0(\text{Zn}^{2+}/\text{Zn}) + RT/2F \ln a(\text{Zn}^{2+}) \quad (7)$$

known as the Nernst equation. The expression Zn^{2+}/Zn is used to shorten the half-cell reaction, which is $\text{Zn}^{2+}(\text{aq}) + 2e = \text{Zn(s)}$; a zinc activity in pure zinc solid phase is equal to a unit. It follows from the text above that for a hydrogen electrode, the electrode potential can be written as in Eq. 8:

$$E = RT/2F \ln [a(\text{H}^+)]^2 / p(\text{H}_{2,\text{rel}}) \quad (8)$$

where $p(\text{H}_{2,\text{rel}}) = p(\text{H}_2)/p^0$ is a partial pressure of hydrogen related to the standard pressure, and, in view of the convention, $E^0(\text{H}^+/\text{H}_2) = 0$. Thus, Eq. 8 can be rewritten as Eq. 9:

$$E = -0.05916 \text{ pH} - 0.05916/2 \log p(\text{H}_{2,\text{rel}}) \quad (9)$$

because the term $RT/F \ln 10 = 0.05916 \text{ V}$ at 25°C . Thus, the hydrogen electrode can be considered a pH electrode (analytically, however, it has certain disadvantages compared with the glass electrode that now dominates pH measurements; see below).

Other half-cells that can be used as pH electrodes are metal-metal oxide electrodes. The most widely used is the antimony-antimony oxide electrode, which can be written as $\text{Sb(s)}|\text{Sb}_2\text{O}_3(\text{s})|\text{H}^+$ or $\text{Sb(s)}|\text{Sb}_2\text{O}_3(\text{s})|\text{OH}^-$ and with half-cell reactions assumed to be either $\text{Sb}_2\text{O}_3(\text{s}) + 6\text{H}^+ + 6e = 2\text{Sb(s)} + 3\text{H}_2\text{O}$ or $\text{Sb}_2\text{O}_3(\text{s}) + 3\text{H}_2\text{O} + 6e = 2\text{Sb(s)} + 6\text{OH}^-$. For example, the potential of the electrode can be expressed as Eq. 10, assuming unit activities of Sb, Sb_2O_3 , and H_2O :

$$E = E^0(\text{SbO}/\text{Sb}) - 0.05916 \text{ pH} \quad (10)$$

The electrode is favored for measurements in situations in which other electrodes are easily fouled, but because of its poor precision (0.1 pH unit when properly calibrated), it is mostly used as the pH indicator in titrations. Nevertheless, similar electrodes based on noble metals (for example, $\text{Pd}|\text{PdO}$) are still subjected to further development.

All electrodes depend on oxidation and reduction, but the term oxidation-reduction electrode, or redox electrode, is usually reserved for the case in which a species

exists in solution in two oxidation stages. This electrode is denoted $\text{M(s)}|\text{Ox, Red}$, where M is an inert metal (usually platinum) serving as an electron carrier and making electrical contact with the solution. The half-cell equilibrium can either be simple (e.g., $\text{Fe}^{3+} + e = \text{Fe}^{2+}$) or be affected by other ions (e.g., $\text{MnO}_4^- + 8\text{H}^+ + 5e = \text{Mn}^{2+} + 4\text{H}_2\text{O}$). Corresponding electrode potentials are then expressed by Eq. (11):

$$E = E^0(\text{Fe}^{3+}/\text{Fe}^{2+}) + RT/F \ln a(\text{Fe}^{3+})/a(\text{Fe}^{2+}) \quad (11)$$

or Eq. 12, respectively:

$$E = E^0(\text{MnO}_4^-/\text{Mn}^{2+}) + RT/5F \ln a(\text{MnO}_4^-) \cdot [a(\text{H}^+)]^8 / a(\text{Mn}^{2+}) \quad (12)$$

In the second case, the term of Eq. 13 can be isolated:

$$E^{\text{of}} = E^0(\text{MnO}_4^-/\text{Mn}^{2+}) + 8RT/5F \ln a(\text{H}^+) \quad (13)$$

It expresses the formal redox potential and its pH dependence.

REFERENCE ELECTRODES

An insoluble salt electrode (also called a second-order electrode) consists of a metal covered by a porous layer of its insoluble salt. The whole assembly is immersed in a solution containing a corresponding anion. For example, a silver-silver chloride electrode is denoted $\text{Ag(s)}|\text{AgCl(s)}|\text{Cl}^-$; the electrode potential is a combination of the equation analogous to Eq. 7, and the solubility product of a sparingly soluble salt, $K_s(\text{AgCl}) = a(\text{Ag}^+) \cdot a(\text{Cl}^-)$, is shown in Eq. 14:

$$E = E^0(\text{AgCl}/\text{Ag}) - RT/F \ln a(\text{Cl}^-) \quad (14)$$

where a standard electrode potential for a half-cell reaction $\text{AgCl(s)} + e = \text{Ag(s)} + \text{Cl}^-$ is expressed by Eq. 15:

$$E^0(\text{AgCl}/\text{Ag}) = E^0(\text{Ag}^+/\text{Ag}) + RT/F \ln K_s(\text{AgCl}) \quad (15)$$

In potentiometric measurements, these electrodes are used as reference half-cells. For this purpose, their potential stability must be first guaranteed by the constant anion activity in the solution in contact. The usual arrangement is $\text{metal}|\text{insoluble salt}|\text{inner solution} : \text{test solution}$, where the vertical dashed line marks the diffusive barrier. The most frequently used reference electrodes are the silver-silver chloride electrode (preferred for its stable and reproducible potential, low temperature hysteresis, a wide useful temperature range, and easy preparation; when

Table 1 Selected standard electrode potentials

Half-cell reaction	E^0 , V
$\text{Ag}^+ + \text{e} = \text{Ag(s)}$	+0.799
$\text{AgBr(s)} + \text{e} = \text{Ag(s)} + \text{Br}^-$	+0.073
$\text{AgCl(s)} + \text{e} = \text{Ag(s)} + \text{Cl}^-$	+0.222
$\text{AgI(s)} + \text{e} = \text{Ag(s)} + \text{I}^-$	-0.151
$\text{Bi}_2\text{O}_3(\text{s}) + 3 \text{H}_2\text{O} + 6\text{e} = 2 \text{Bi(s)} + 6 \text{OH}^-$	-0.44
$\text{Br}_2(\text{l}) + 2\text{e} = 2 \text{Br}^-$	+1.065
$\text{BrO}_3^- + 6 \text{H}^+ + 6\text{e} = \text{Br}^- + 3 \text{H}_2\text{O}$	+1.44
$\text{C}_6\text{H}_4\text{O}_2$ [quinone] + $2 \text{H}^+ + 2\text{e} = \text{C}_6\text{H}_4(\text{OH})_2$ [hydroquinone]	+0.699
$\text{Ce}^{4+} + \text{e} = \text{Ce}^{3+}$	+1.70
$\text{Cl}_2(\text{g}) + 2\text{e} = 2 \text{Cl}^-$	+1.359
$\text{ClO}_3^- + 6 \text{H}^+ + 6\text{e} = \text{Cl}^- + 3 \text{H}_2\text{O}$	+1.45
$\text{ClO}_4^- + 2 \text{H}^+ + 2\text{e} = \text{ClO}_3^- + \text{H}_2\text{O}$	+1.19
$\text{Cr}_2\text{O}_7^{2-} + 14 \text{H}^+ + 6\text{e} = 2 \text{Cr}^{3+} + 7 \text{H}_2\text{O}$	+1.33
$\text{Cu}^{2+} + 2\text{e} = \text{Cu(s)}$	+0.337
$\text{Cu}^{2+} + \text{e} = \text{Cu}^+$	+0.153
$\text{Cu}^+ + \text{e} = \text{Cu(s)}$	+0.521
$\text{CuI(s)} + \text{e} = \text{Cu(s)} + \text{I}^-$	-0.185
$\text{Fe}^{3+} + \text{e} = \text{Fe}^{2+}$	+0.771
$\text{Fe(CN)}_6^{3-} + \text{e} = \text{Fe(CN)}_6^{4-}$	+0.36
$2 \text{H}^+ + 2\text{e} = \text{H}_2(\text{g})$	0.000
$\text{H}_2\text{O}_2 + 2 \text{H}^+ + 2\text{e} = 2 \text{H}_2\text{O}$	+1.776
$\text{Hg}_2^{2+} + 2\text{e} = \text{Hg(l)}$	+0.788
$2 \text{Hg}^{2+} + 2\text{e} = \text{Hg}_2^{2+}$	+0.920
$\text{Hg}_2\text{Cl}_2(\text{s}) + 2\text{e} = 2 \text{Hg(l)} + 2 \text{Cl}^-$	+0.268
$\text{Hg}_2\text{SO}_4(\text{s}) + 2\text{e} = 2 \text{Hg(l)} + \text{SO}_4^{2-}$	+0.615
$\text{HNO}_2 + \text{H}^+ + \text{e} = \text{NO(g)} + \text{H}_2\text{O}$	+1.00
$\text{I}_3^- + 2\text{e} = 3 \text{I}^-$	+0.536
$\text{IO}_3^- + 6 \text{H}^+ + 6\text{e} = \text{I}^- + 3 \text{H}_2\text{O}$	+1.085
$\text{K}^+ + \text{e} = \text{K(s)}$	-2.925
$\text{MnO}_2(\text{s}) + 4 \text{H}^+ + 2\text{e} = \text{Mn}^{2+} + 2 \text{H}_2\text{O}$	+1.23
$\text{MnO}_4^- + 8 \text{H}^+ + 5\text{e} = \text{Mn}^{2+} + 4 \text{H}_2\text{O}$	+1.51
$\text{MnO}_4^- + 4 \text{H}^+ + 3\text{e} = \text{MnO}_2(\text{s}) + 2 \text{H}_2\text{O}$	+1.695
$\text{MnO}_4^- + \text{e} = \text{MnO}_4^{2-}$	+0.564
$\text{Na}^+ + \text{e} = \text{Na(s)}$	-2.714
$\text{NO}_3^- + 3 \text{H}^+ + 2\text{e} = \text{HNO}_2 + \text{H}_2\text{O}$	+0.94
$\text{O}_2(\text{g}) + 4 \text{H}^+ + 4\text{e} = 2 \text{H}_2\text{O}$	+1.229
$\text{O}_2(\text{g}) + 2 \text{H}^+ + 2\text{e} = \text{H}_2\text{O}_2$	+0.682
$\text{O}_2(\text{g}) + \text{H}_2\text{O} + 4\text{e} = 4 \text{OH}^-$	+0.401
$\text{O}_3(\text{g}) + 2 \text{H}^+ + 2\text{e} = \text{O}_2(\text{g}) + \text{H}_2\text{O}$	+2.07
$\text{Pb}^{2+} + 2\text{e} = \text{Pb(s)}$	-0.126
$\text{PbO}_2(\text{s}) + 4 \text{H}^+ + 2\text{e} = \text{Pb}^{2+} + 2 \text{H}_2\text{O}$	+1.455
$\text{PbO}_2(\text{s}) + \text{SO}_4^{2-} + 4 \text{H}^+ + 2\text{e} = \text{PbSO}_4(\text{s}) + 2 \text{H}_2\text{O}$	+1.685
$\text{PbSO}_4(\text{s}) + 2\text{e} = \text{Pb(s)} + \text{SO}_4^{2-}$	-0.350
$\text{S(s)} + 2 \text{H}^+ + 2\text{e} = \text{H}_2\text{S(g)}$	+0.141
$\text{SO}_4^{2-} + 4 \text{H}^+ + 2\text{e} = \text{H}_2\text{SO}_3 + \text{H}_2\text{O}$	+0.172
$\text{S}_2\text{O}_8^{2-} + 2\text{e} = 2 \text{SO}_4^{2-}$	+2.01
$\text{S}_4\text{O}_6^{2-} + 2\text{e} = 2 \text{S}_2\text{O}_3^{2-}$	+0.08
$\text{Sb}_2\text{O}_3(\text{s}) + 6 \text{H}^+ + 6\text{e} = 2 \text{Sb(s)} + 3 \text{H}_2\text{O}$	+0.152
$\text{Sn}^{2+} + 2\text{e} = \text{Sn(s)}$	-0.136
$\text{Sn}^{4+} + 2\text{e} = \text{Sn}^{2+}$	+0.154
$\text{Ti}^{3+} + \text{e} = \text{Ti}^{2+}$	-0.369

(Continued)

Table 1 Selected standard electrode potentials (*Continued*)

Half-cell reaction	E^0, V
$\text{Ti}^{3+} + 2\text{e} = \text{Ti}^+$	+1.25
$\text{Ti}^+ + \text{e} = \text{Ti}(s)$	−0.336
$\text{TiCl}(s) + \text{e} = \text{Ti}(s) + \text{Cl}^-$	−0.557
$\text{UO}_2^{2+} + 4\text{H}^+ + 2\text{e} = \text{U}^{4+} + 2\text{H}_2\text{O}$	+0.334
$\text{Zn}^{2+} + 2\text{e} = \text{Zn}(s)$	−0.763

s = solid; *l* = liquid; *g* = gas.

filled with saturated KCl solution, its potential is 0.198 V at 25°C) and the calomel (or mercury–mercurous chloride) electrodes (consisting of mercury covered with a paste of Hg and Hg_2Cl_2 , which, when in contact with saturated KCl solution, the potential is 0.244 V at 25°C; a disadvantage of classic calomel electrodes is their considerable temperature hysteresis). Other reference electrodes are not often used.

The Ross reference electrode differs from the others and consists of a platinum wire immersed in a solution containing tri-iodide and iodide ions. The Pt electrode responds to the redox potential established by the iodine(tri-iodide)-iodide couple. This solution is separated from the sample by a bridge electrolyte, which is 3 M KCl.

LIQUID JUNCTION

In a cell with two different electrolyte solutions in contact, as in the application of a reference electrode, there is an additional source of potential difference across the interface of the two miscible electrolytes, inner reference electrode solution : test solution. As noted above, the vertical dashed line is used in a cell scheme to denote such an interface and indicates the source of so-called liquid-junction potential, $\Delta\phi_L$. For example, if two HCl solutions of different concentrations are in such contact, the mobile H^+ ions diffuse into the more dilute solution. The bulkier Cl^- ions follow, but initially more slowly, which results in a potential difference at the junction. However, after the brief initial period, the potential reaches a value that the ions diffuse at the same rate.

Generally, the liquid-junction potential increases with increasing difference of the cation and anion mobilities at the interface. In an ideal situation, both ions are of the same mobility. Such salts are called equitransferent salts, or equitransferent mixtures. For that reason, KCl is commonly used as the inner electrolyte of reference electrodes; a mixture of 1.8 M KCl + 1.8 M KNO_3 serves

best. The liquid-junction potential, although not known, remains constant within limits, appears as a constant offset voltage, and is contained in the apparent (or formal) standard potential, E^{of} . If both H^+ and OH^- do not participate at the border solutions, it usually does not exceed ± 10 mV. In direct potentiometric measurements, the liquid-junction potential can partially be reduced by the use of the reference electrode with a limited diffusion (the boundary between the two miscible electrolytes being a porous diaphragm) or by incorporating a salt bridge filled with a proper electrolyte solution that is equivalent to the use of a reference electrode with two bridge electrolytes (the so-called double-junction reference electrode) containing two porous diaphragms with limited diffusion, e.g., $\text{Ag}(s)|\text{AgCl}(s)|\text{KCl} : \text{a bridge electrolyte} : \text{a test solution}$. In the notation for cells, a double vertical line denotes an interface for which it is assumed that the junction potential has practically been eliminated, e.g., $\text{Ag}(s)|\text{AgCl}(s)|\text{KCl} :: \text{a test solution}$, which is equivalent to the previous scheme.

Some electroanalytical texts that contain sections with emphasis discussed in previous articles are given in Refs. 1 and 2.

MEMBRANE ELECTRODES

The difficulty that can arise with the cells comprising the first- and second-order electrodes noted above is whether oxidized and reduced species of more than one redox couple are present in solution so that they can contribute to the overall equilibrium potential that is thus a mixed potential. Such a measurement can have low selectivity in some real situations. This can be overcome by measuring the difference of potential across a membrane composed from a material that can selectively participate at ionic exchange equilibria.

A membrane separating medium 1 and medium 2 results in a three-phase system, medium 1|membrane|medium 2 because the membrane behaves as isolated phase (M).

Interfaces at which an ionic exchange occurs are geometric barriers between two phases. Membranes for potentiometric electrodes are as immiscible as conveniently possible with respect to the bathing solutions and solid contacts. They are usually constructed of hydrophobic organic liquids and solids or inorganic solids of low water solubility. Nevertheless, useful membranes are not electrical insulators. Porous membranes (organic liquids and solids, synthetic ion exchangers) absorb and become saturated by an external solvent, usually water. They also permit water from two bathing solutions with nonidentical ionic strengths to pass slowly from one side of the membrane to the other. However, many membranes are nonporous, and solvent transport is not an important process when considering membrane potential responses. Useful membranes are often solid or liquid electrolytes because they are composed of partially or completely ionized acids, bases, or salts or contain potentially ionizable species. A characteristic of these membranes is the presence of charged sites. A result of the ionic exchange occurring at both membrane interfaces is the membrane potential $\Delta\phi_M$, expressed as a difference of the inner electric potential of the two phases, which are separated by the membrane, as shown in Eq. 16:

$$\Delta\phi_M = \phi(2) - \phi(1) \quad (16)$$

Generally, the membrane potential is completed by three factors: the two potential differences at both inner and outer interfaces and a liquid-junction potential that can be formed across the membrane (especially in liquid membranes). Nevertheless, a galvanic half-cell, represented by a membrane electrode immersed in the sample test solution, usually consists of an ion-selective membrane, an internal electrolyte, and an internal reference electrode, e.g., $\text{Ag}|\text{AgCl}|\text{internal electrolyte } \text{Cl}^-$, $\text{Q}^+|\text{membrane selective to } \text{Q}^+|\text{test solution}$. This represents a conventional construction of the sensing half-cell; the other half-cell is represented by an external reference electrode. There are many factors to the experimentally observed voltage of a cell containing a membrane electrode that can be considered constant: the inner and outer reference electrode potentials, the potential difference at the inner membrane interface, and, at least in one instant, both liquid-junction potentials formed across the membrane and at the external reference electrode : test solution interface.

Thus, ion-selective membrane electrodes can be defined as electrochemical sensors that allow potentiometric measurements of the activity of particular species in aqueous and mixed solvents or partial pressures of dissolved gases in water. However, these sensors may

respond to certain other ions in the sample in addition to the selected i ion; interferences by such j ions are usually expressed by the Nikolskii-Eisenman (Eq. 17):

$$E = \text{constant} + f(2.303 RT/z_i F) \log\{a_i + \sum k_{ij}^{\text{pot}} a_j^{z_i/z_j}\} \quad (17)$$

where E is the voltage of the potentiometric cell in which an ion-selective membrane electrode participates, and the constant term includes all the constant potential contributions noted above; f is a correction for non-Nernstian response, terms z_i and z_j denote the charge of the ions i and j , and k_{ij}^{pot} is the potentiometric selectivity coefficient.

It should be noted that over the past 30 years, analytical methods utilizing ion-selective technology have been developed at an ever-increasing rate, and much additional information can be found in texts addressing the subject. Some are noted in the references (3–10).

Based on the membrane material, the ion-selective electrodes may be divided into the following groups.

Electrodes with Solid Membranes

These include membranes composed of solid salts that may be single crystals but are more often polycrystalline pellets pressed from powdered starting materials. These solid exchangers respond to species that exchange directly and rapidly and influence the activity of ions that exchange directly. For example, silver salt electrodes are believed to be rapid ion exchangers of both Ag^+ and component anions. In mixtures, Ag_2S holds the more soluble halides in a matrix and produces both electronic and ionic conductivities throughout the membrane. The electrode is also sensitive to other ions that can form sparingly soluble precipitates on the membrane surface, particularly if they have lower solubility products than those of the material of the membrane. As an example, the AgCl -based electrode responds to bromide and iodide; mercury(II) is always a serious interferent because HgS is less soluble than Ag_2S . For the same reason, membranes containing HgS instead of Ag_2S have the advantage of lower detection limits. Some examples of the membrane materials are LaF_3 for fluoride ions, $\text{AgX} + \text{Ag}_2\text{S}$ or $\text{Hg}_2\text{X}_2 + \text{HgS}$ for halide or pseudohalide (CN^- , SCN^-) ions, Ag_2S for both silver and sulfide ions, metal sulfide + Ag_2S for some metal ions (Cd^{2+} , Cu^{2+} , Pb^{2+}), and $\text{Cu}_2\text{Se} + \text{CuSe}$ for copper(II) ions.

Glass Electrodes

These are the oldest and best investigated ion-selective electrodes, made of various multicomponent glasses (11).

The surface of a glass membrane must be hydrated to respond to pH. Hydration is accompanied by a reaction in which singly charged cations of the glass are exchanged for hydrogen ions of the solution $\text{H}^+(\text{aq}) + \text{Na}^+(\text{glass}) = \text{Na}^+(\text{aq}) + \text{H}^+(\text{glass})$. The silicate structure $\text{Si}-\text{O}^-$ of a glass provides cation-bonding sites. A well soaked membrane is covered by a layer of silicic acid gel approximately 10^{-5} to 10^{-4} mm thick.

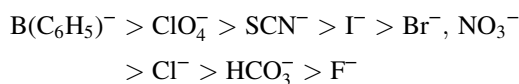
Glass electrodes respond to the activity of both the hydrogen ion and alkali metal ions in basic solution in which the former is small relative to the latter. At high pH, a negative pH error occurs that suggests that the electrode is responding to alkali metal ions as well as to hydrogen ions. The magnitude of the error varies according to the kind of singly charged cation and the glass composition. Because of this error in early glass electrodes, glasses were developed for which the alkaline error is negligible below a pH level of approximately 12. Other glasses have been developed that allow the determination of cations other than hydrogen ions such as Li^+ , Na^+ , K^+ , and NH_4^+ . Glass electrodes for H^+ , Na^+ , and K^+ are available commercially.

Electrodes with Liquid Membranes

These include membranes made of liquid electroactive substances or with electroactive substances dissolved in a suitable nonvolatile, water-immiscible solvent (mediator). In early designs, the organic phase was placed between two aqueous phases in bulk or with the support of a thin, porous cellulose sheet, sintered glass, or the like. As work with these sensors proceeded, more durable polymer supports were developed, most often poly(vinyl chloride)(PVC). An electroactive compound is dissolved in a solvent (usually tetrahydrofuran or cyclohexanone) together with the PVC and a suitable plasticizer. The solvent evaporates, leaving a plasticized PVC ion-selective membrane. Because the plasticized polymer behaves like a viscous liquid, the properties of the electrode are very similar to those of the original wet membrane. The electroactive materials used in these membranes fall into three main classes: ion exchangers, neutral carriers, and charged carriers.

In liquid ion-exchange electrodes, the electroactive material is usually the salt (ion pair) with a highly lipophilic cation or anion to guarantee oil solubility. For example, nitrate-, fluoroborate-, and perchlorate-selective electrodes are commercially available, based on corresponding ion pairs of an anion with tris(substituted 1,10-Phenanthroline)-metal(II) cations or a quaternary

ammonium salt containing at least one long-chain alkyl group. Selectivity over interferences is limited to ions less oil-soluble than the primary ion. In the case of the above-noted electrodes, the Hofmeister series is usually obeyed:



and can serve as a rule for deciding possible interferences. Similarly, an organic anion such as tetraphenylborate is a suitable membrane counterion for cations (e.g., K^+ , Rb^+ , Cs^+ , Tl^+ , and univalent organic cations). A similar effect is seen because the electrodes are most selective for organic cations such as long-chain quaternary ammonium ions. These cation electrodes have frequently been used as sensors for clinically important molecules that are cationic at low pH levels (drugs).

Another category of liquid-membrane electrodes is based on neutral carriers. These are lipophilic, multi-functional compounds with active groups that are primarily alternating ether and/or keto oxygens that can form a cage for the positively charged ion. Cyclic polyethers and similar macrocyclic compounds, such as valinomycin and nonactin, are believed to discriminate among cations on the basis of size; cations that fit well in the complexing site are most strongly complexed. Potassium-selective electrodes based on valinomycin are the best examples of marketed sensors of this type. Various crown ethers, polyoxyethylene chain-containing compounds, or special selective carriers have been prepared and used in sodium-, barium-, and calcium-selective commercial electrodes.

A third class of ion-selective carriers consists of so-called charged carriers or associated ion exchangers. Unlike the simple ion exchangers, however, the selectivity of charged carriers is dictated by the degree of association of the analyte ion with the carrier as well as the partitioning of the analyte into the membrane solvent. The most notable examples of such agents are the alkyl phosphates originally used in Ca^{2+} -selective electrodes; calcium selectivity is further enhanced by using alkyl phosphonates as membrane mediators.

Gas-Sensing Electrodes

Gas-sensing electrodes are examples of multiple membrane sensors; these contain a gas-permeable membrane separating the test solution from an internal thin electrolyte film in which an ion-selective electrode is immersed. For example, for the ammonia sensor,

the pH of the recipient layer is determined by the Henderson-Hasselbach equation (Eq. 18), derived from the chemical equilibrium between solvated ammonia and ammonium ions:

$$\text{pH} = \text{pK}_a(\text{NH}^+) + \log[\text{NH}_3]/[\text{NH}_4^+] \quad (18)$$

If the solution layer contains a large background concentration of ammonium salt (NH_4^+ picrate is often used), the pH of the immersed glass electrode is proportional to only one variable, $[\text{NH}_3]$, which is in turn dependent on the amount of NH_3 that diffuses across the gas-permeable membrane, as shown by Eq. 19:

$$E = \text{constant} + 0.05916 \log p(\text{NH}_3) \quad (19)$$

where $p(\text{NH}_3)$ is the partial pressure of NH_3 in the sample measured. Similarly, when the glass pH electrode is replaced by a polymeric membrane responsive to NH_4^+ (a nonactin-based NH_4^+ -selective electrode), the signal is not dependent on the pH but on the equilibrium concentration of ammonium acquired from the sample. The ammonia gas-sensing electrode is primarily used for the determination of ammonium salts after the addition of NaOH, which releases ammonia from a sample under test. The total nitrogen can be determined after Kjeldahl decomposition of a sample.

The first gas-sensing electrode based on a similar principle was the carbon dioxide electrode, developed to determine CO_2 in blood. Later, sensors for other gases (e.g., SO_2 , NO_x , and HCN, etc.) appeared on the market.

Electrodes with Biocatalytic Membranes

The gas-sensing configuration described above forms a very useful basic unit for potentiometric measurements of biologically important species (12–17). In principle, the immobilized or insolubilized biocatalyst is placed on a conventional ion-selective electrode used to measure the decrease in the reactants or the increase in products of the biochemical reaction. The biocatalyst include isolated enzymes, subcellular fractions, intact bacterial cells, and whole sections of mammalian or plant tissues. Because ions are usually formed during these reactions, it is possible to determine the substrate by monitoring the ion activity. For example, amygdalin can be degraded by β -Glucosidase to benzaldehyde, glucose, and HCN; the CN^- -selective electrode is used as an inner sensor. Similarly, the NH_4^+ -selective electrode can respond to NH_4^+ ions produced in the reaction of urea with urease.

In recent years, gas-sensing probes have been used most frequently because of their high selectivity over

common cations and anions. Thus, the biocatalytic urea sensor noted above can be constructed by immobilizing urease onto the gas-permeable surface of an ammonia gas sensor; when the probe is inserted into a buffered sample containing urea, the enzyme catalyzes its conversion to NH_3 , which is measured by the gas-sensing electrode described previously. By coupling the efficient and selective catalyzing powers of an enzyme to the selective detection of a gas-sensing electrode, it is possible to construct a sensitive tool for the measurement of many pharmaceutically important compounds. However, in general for biosensors, it should be noted that amperometric principles predominate in a majority of their recent constructions.

SELECTIVITY COEFFICIENTS

In measurements with ion-selective electrodes, interference by other ions is expressed by selectivity coefficients k_{ij}^{pot} , as in Eq. 17. If the nature of the ion-selective membrane is known, these interferences may easily be estimated. For example, in the determination of chloride with a Cl^- -selective electrode containing AgCl as the electroactive component in its membrane, concentrations of bromides or iodides (generally X^-) must be controlled because they form less soluble silver salts than AgCl; the solubility products of corresponding silver halides are used in Eq. 20 to estimate the selectivity coefficient:

$$k_{\text{Cl},\text{X}}^{\text{pot}} \cong K_s(\text{AgCl})/K_s(\text{AgX}) \quad (20)$$

However, this estimation is approximate only. For exact determinations, two methods are primarily used.

In the separate-solution method, the cell voltage E_i is measured first in a solution of free determinant i , followed by E_j measured in a solution of interferant j . By applying Eq. 17 for these two solutions, Eq. 21 is obtained:

$$\log k_{ij}^{\text{pot}} = (E_j - E_i)z_iF/2.303RT + \log a_i - z_i/z_j \log a_j \quad (21)$$

This equation can be simplified further, assuming the main interferants are the ions of the same charge and considering solutions of the same activities.

In the simplest mixed-solution method, the cell voltage is measured in a series of solutions containing a range of activities of the primary ion and fixed activity of the interferant (Fig. 1). When a graph of E versus $\log a_i$ is constructed, the usual Nernstian slope changes

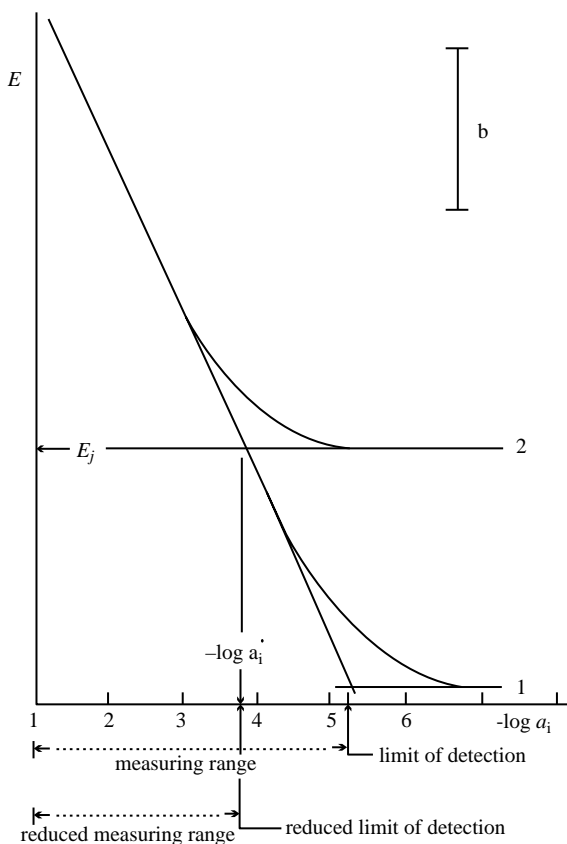


Fig. 1 Calibration graphs of ion-selective electrodes and evaluation of selectivity coefficients. 1) Calibration response against $-\log a_i$ in solution of free determinand i . The practical limit of detection may be taken as the activity (or the concentration) at the point of intersection of the extrapolated lines as shown. 2) Calibration against $-\log a_i$ in the presence of the interferant j , the activity of which is of a known and constant value. Response E_j is obtained for $-\log a_i^* = -\log k_{ij}^{\text{pot}} a_j^{z_i/z_j}$. The resulting interference can restrict the measuring range. b) The abscissa approximately equal to $60/z_i$ mV.

to an invariant E_j voltage at low a_i values, corresponding to the response to a constant a_j activity. The intersection of asymptots to this dependence gives a $\log a_i^*$ value, which is used to calculate k_{ij}^{pot} according to Eq. 22:

$$k_{ij}^{\text{pot}} = a_i^* / a_j^{z_i/z_j} \quad (22)$$

The weakness of all the approaches is the assumption that the slope $2.303RT/z_iF$ is Nernstian or at least unaffected by the presence of interferants. In addition, for many electrodes, especially liquid-membrane systems, selectivity coefficients are not highly reproducible or precise quantities because they are

time-dependent. This is why selectivity coefficients established by various methods can differ; nevertheless, they never differ by order of magnitude, and therefore they serve well to estimate measurement errors in the presence of interfering species.

EQUIPMENT

The selection of equipment for potentiometric measurements is guided primarily by the precision required and the application intended. An instrument must draw essentially no electricity from the cell being used. Historically, measurements (highly precise but tedious) were performed by compensation of the measured voltage with the aid of a reference cell, so that virtually no current passed through the measuring cell. This kind of potentiometer has by now been almost completely displaced by electronic voltmeters with very high internal resistances. This is particularly significant in measurements with membrane electrodes, which may have resistances of 1 to 100 M Ω or even more. These high-resistance instruments are usually called pH meters or ion-meters. A new generation of instruments is represented by microprocessor-controlled meters, making automatic evaluation of the measurements possible. They are produced by many companies (e.g., Orion Research, Radiometer, and Metrohm). Equipment for automated potentiometric titrations has also been developed; the apparatus is generally of the module type, and its individual components (ion-meter, auto-burette, recorder, and computer unit) may be used separately.

EXPERIMENTAL TECHNIQUES

Direct Potentiometry

Direct potentiometric measurements are used to complete chemical analyses of species for which an indicator electrode is available. The technique is simple, requiring only a comparison of the voltage developed by the measuring cell in the test solution with its voltage when immersed in a standard solution of the analyte. If the electrode response is specific for the analyte and independent of the matrix, no preliminary steps are required. In addition, although discontinuous measurements are mainly carried out, direct potentiometry is readily adapted to continuous and automatic monitoring.

Table 2 Standard buffer solutions

Composition	pH Values at °C					
	0	20	25	30	38	60
0.1 m hydrochloric acid	1.187	1.194	1.197	1.200	1.202	1.213
0.05 m potassium tetraoxalate	1.666	1.675	1.679	1.683	1.691	1.723
Saturated (25°C) potassium hydrogen tartrate			3.557	3.552	3.548	3.560
0.05 m potassium dihydrogen citrate	3.863	3.788	3.776	3.766		
0.05 m potassium hydrogen phthalate	4.003	4.002	4.008	4.015	4.030	4.091
1+1 phosphate buffer (0.025 m KH_2PO_4 + 0.025 m Na_2HPO_4)	6.984	6.881	6.865	6.853	6.840	6.836
1+3.5 phosphate buffer (0.008695 m KH_2PO_4 + 0.03043 m Na_2HPO_4)	7.534	7.429	7.413	7.400	7.384	
0.01 m borax	9.464	9.225	9.180	9.139	9.081	8.962
1+1 carbonate buffer (0.025 m NaHCO_3 + 0.025 m Na_2CO_3)	10.317	10.062	10.012	9.966		
Saturated (25°C) calcium hydroxide	13.423	12.627	12.454	12.289	12.043	11.499

Measurement of pH

The electrochemical pH cell consists essentially of a measuring (or indicator) pH electrode, together with a reference electrode, both being in contact with the solution under investigation. Frequently, a pH glass and calomel electrodes are combined; then the pH meter measures the cell voltage E , given by Eq. 23, for example:

$$E = E_{\text{ref}} - E_{\text{ind}} + \Delta\phi \quad (23)$$

where E_{ref} is the reference electrode potential, E_{ind} the indicator electrode potential, and $\Delta\phi_L$ is the liquid-junction potential arising at the boundary between the dissimilar liquids. For a cell with a pH-measuring electrode, Eq. 23 can be rewritten as Eq. 24:

$$E = \text{constant} - 2.303 RT/F \log a(\text{H}^+) + \Delta\phi_L \quad (24)$$

where the constant term (including E_{ref} and the constant substituting E^0 of the pH electrode) and $\Delta\phi_L$ are not available. These unknowns are eliminated by subtracting the voltage $E(X)$ of the cell immersed into a test solution of pH(X) and that of $E(S)$ measured with the same cell immersed into a standard pH(S) solution (Table 2), as in Eq. 25:

$$E(X) - E(S) = -2.303 RT/F [\log a(\text{H}^+)_x - \log a(\text{H}^+)_s] \quad (25)$$

and, subsequently, as in Eq. 26:

$$\text{pH}(X) = [E(X) - E(S)]F/2.303 RT + \text{pH}(S) \quad (26)$$

assuming that the difference $\Delta\phi_L(X) - \Delta\phi_L(S)$ is eliminated. It follows that the accuracy of the pH(X) value is partially dependent on how close in character

the tested solution is to the standardizing solution and how strictly Nernstian the cell response is; clearly, it is desirable to standardize as closely in pH as possible. For more accuracy measurements, it is suggested that two standard buffers be used, pH(S_1) and pH(S_2), which straddle the pH(X) value. The pH(X) is then given by Eq. 27:

$$\begin{aligned} & [\text{pH}(X) - \text{pH}(S)] / [\text{pH}(S) - \text{pH}(S)] \\ & = [E(X) - E(S)] / [E(S) - E(S)] \end{aligned} \quad (27)$$

Eq. 27 represents the so-called practical definition of the pH scale (18). Evidently, in practice the voltage differences expressed in Eqs. 26 and 27 are not measured inasmuch as the meters provide direct pH readings. Hence, instead of the $E(S)$ values, the pH(S) values are adjusted directly on the instrument scale. In the preparation of standard buffer solutions (see Table 2), it is essential to use high-purity materials and carbon dioxide-free freshly distilled water, the specific conductance of which should not exceed 2 $\mu\text{S}/\text{cm}$.

Frequently, it is necessary to measure pH at nonambient temperatures. Biological samples are often stored just above 0°C, clinical samples are measured at 38°C, buffers and gels used in media are measured at approximately 60°C, and many industrial processes take place at higher temperatures. The Nernst equation contains a temperature term that can be corrected by automatic temperature compensation; detailed instructions are attached to each pH meter. The glass Ross pH electrode, containing a Pt wire as a reference immersed in a solution based on iodine (tri-iodide) and potassium iodide, gives the best performance in terms of a fixed, reproducible contact potential that is thermodynamically reversible (the Pt wire potential does change with

temperature, but the redox solution consists of a buffer with an equal and opposite temperature coefficient, which results in a reference system showing almost no potential change with temperature).

Measurement of pX

This includes direct activity (concentration) measurements using other than H^+ -selective electrodes (19). Different methods are available, although none is as well organized as the pH measurement. The simplest procedure is to measure the voltage of the cell containing an ion-selective electrode in solutions of graduated concentrations, usually between 10^{-6} and 10^{-1} mol/L (or similarly on the pH scale, between pX 6 and 1). A typical calibration graph is linear between pX 1 and 5, defined for $pX = -\log a_i$ (Fig. 1). In practice, however, the determination of concentration is more frequently requested. In this case, the cell voltage values are plotted against the logarithms of the concentration of the ion determined, i.e., $pX = -\log c_i$. Such a calibration graph, however, differs from that obtained by measuring the activity at higher concentrations when the activity coefficients y_i are less than 1 (just to remember the relation between activity and concentration, which is $a_i = c_i y_i$). A series of standard solutions with a composition as close as possible to that of the sample is used, and the conditions are maintained identical to those used for the measurement on the sample (pH and ionic strength adjustments, screening of interferants, etc.). The best results are obtained with simulated standards, in which the effects of the other components of the sample solution are included in the calibration curve. As reported recently (20), dramatic improvement of the lower detection limit to phenomenal picomolar concentrations may be obtained by modifying the composition of the inner electrolyte of the ion-selective electrode.

The most frequently used mode (because it is the simplest mode) is calibration with two standard solutions. It is appropriate for any analysis by direct potentiometry in the Nernstian range of an ion-selective electrode, particularly for analyses carried out at varying temperatures. The calibration is performed with two standards in each sample batch. The first gives the value of the cell constant (or E^0 of the sensing electrode) and the second the calibration slope. The two standards should span the concentration range expected in the samples because any error is magnified by extrapolation. In the non-Nernstian region, the concentrations are obtained from a calibration graph rather than by calculation, and because the graph is curved, more (at least four) standard solutions are needed to define it.

Addition (as well as subtraction) methods may also be used. Both require a knowledge of the calibration slope but not of the cell constant. The simplest method includes two voltage readings, E_1 before and E_2 after the addition of a volume V_s of a standard solution to V_x volume of the sample, as shown in Eqs. 28 and 29:

$$E_1 = \text{constant} + \text{slope} \log c_x \quad (28)$$

$$E_2 = \text{constant} + \text{slope} \log (c_x V_x + c_s V_s) / (V_x + V_s) \quad (29)$$

where c_x denotes the concentration of the sample and c_s is the concentration of the standard solution. From the cell voltage change ($\Delta E = E_2 - E_1$), the unknown c_x concentration can be calculated as shown in Eq. 30:

$$c_x = c_s V_s / [(V_x + V_s) 10^{\Delta E / \text{slope}} - V_x] \quad (30)$$

In the well-known addition method, the slope factor is determined simultaneously with the concentration by iterative calculation. If a multiaddition method is used, the unknown c_x concentration can be evaluated graphically (Fig. 2). For $(V_x + V_s) 10^{E/\text{slope}} = 0$, the negative value of the volume equivalent to the unknown concentration is read from an intercept of the dependence with x -axis, $-V_e$, and the concentration calculated as shown in Eq. 31:

$$c_x = -c_s V_e / V_x \quad (31)$$

Flow measurements

A useful and rapid method of automated analysis is the technique of flow-injection analysis (FIA). The sample or a reagent is injected into the stream of a solution of constant composition. Calibration of FIA systems requires the injection of standard solutions, equal in volume to that of the sample, into the carrier stream. The background chemical composition of the standards should be equal, as nearly as possible, to that of the samples. Frequent standardization is not necessary because the measurement of peak height, albeit on a sloping base line, is relatively unaffected by cell voltage drift. Some difficulties can appear with peristaltic pumps, owing to extraneous potentials caused by pulsation of the stream. Cells with a small volume ($<20 \mu\text{l}$) or the cells of the wall-jet type are the most acceptable for continuous measurements (21, 22).

Potentiometric Titrations

In contrast to direct potentiometry, the potentiometric titration technique offers the advantage of high accuracy

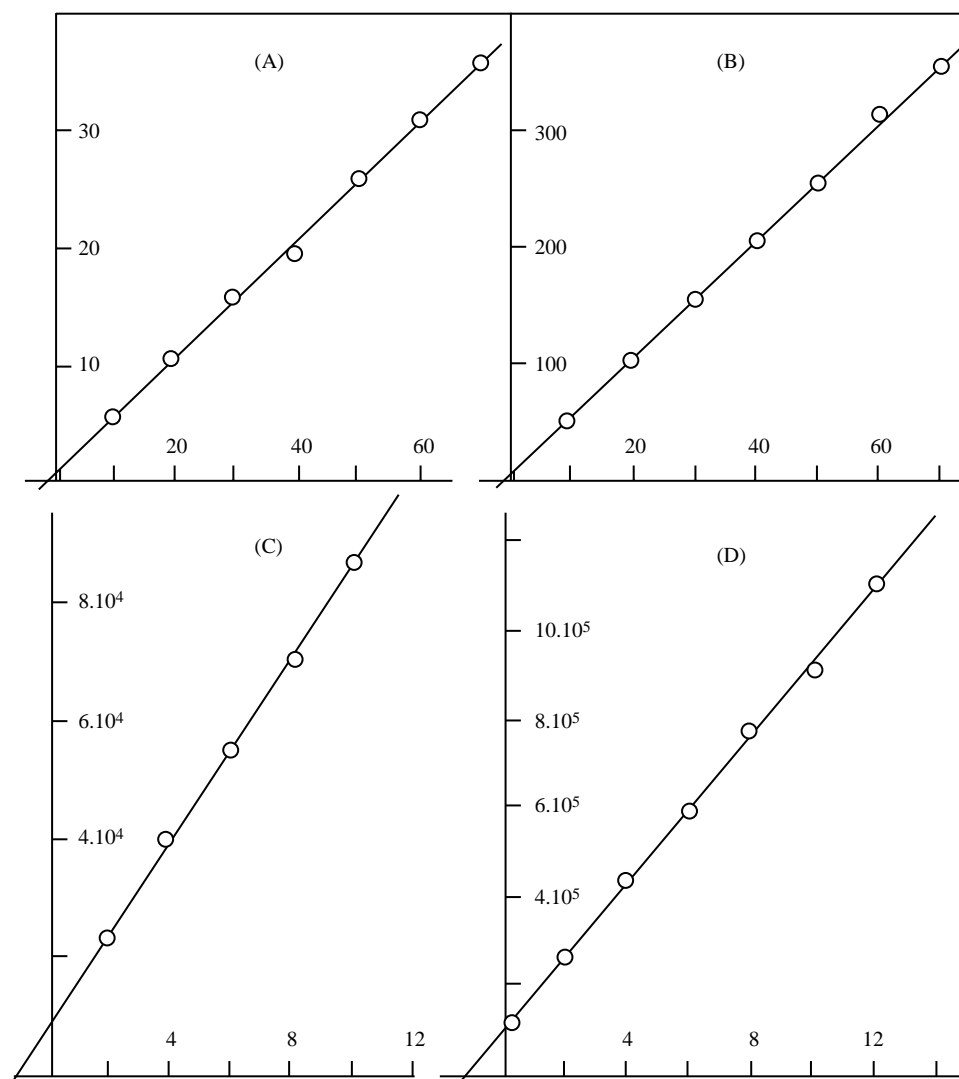


Fig. 2 Multiaddition method with graphic evaluation of the known ion concentration in a sample. Experimental data for the determination of tetrafluoroborate using a BF_4^- -selective electrode. (A) $c_x = 10^{-4}$ mol/L, $c_s = 10^{-2}$ mol/L; (B) $c_x = 10^{-5}$ mol/L, $c_s = 10^{-3}$ mol/L; (C) $c_x = 10^{-3}$ mol/L, $c_s = 10^{-1}$ mol/L; (D) $c_x = 10^{-4}$ mol/L, $c_s = 10^{-2}$ mol/L; $V_x = 100$ mL in all cases. Values of $(V_x + V_s)10^{EF/2.303RT}$ are plotted on the y-axis against volume of the standard solution added, V_s , on x-axis. (Adapted from Sb. Ved. Pr., Vys. Sk. Chemickotechnol. Pardubice **1986**, 49, 149.)

and precision, although at the cost of increased time and increased consumption of titrants. Another advantage is that the potential break at the titration endpoint must be well defined, but the slope of the sensing electrode response need be neither reproducible nor Nernstian, and the actual potential values at the endpoint are of secondary interest. In many cases, this allows for the use of simplified sensors.

For titrations to a fixed potential, the calibration slope is not needed, but the correct potential must be chosen by some means of calibration. In pH titrations, for example,

this can be by means of a buffer solution. The errors involved are usually much smaller than those in direct potentiometry.

PHARMACEUTICAL APPLICATIONS

For analytical control of pharmaceuticals, most pharmacopoeias describe accurate methods, which, however, in some cases are lengthy and difficult. The ion-selective

electrode techniques offer several advantages in terms of simplicity and rapidity over official methods (23–29). Generally, electrodes of all the types noted above can be used to analyze compounds of pharmaceutical significance. For example, halide ion-selective electrodes are often used in the determination of cationic compounds containing quaternary nitrogen and a counter-halide ion. Traditional titration techniques, such as potentiometric pH or redox titrations, are used to determine acids and bases or compounds containing groups that may be oxidized or reduced.

Most ion-selective electrodes, however, cannot be used for the direct determination of functional groups in organic compounds unless they are converted into ionic species. Thus, direct potentiometry is performed rarely with commercial ion-selective electrodes. For example, the CN^- -selective electrode gives an almost Nernstian response in the determination of substituted phenylacetone nitriles and benzonitriles. Thiols can be determined by direct measurement of the voltage of a cell with an Ag_2S -based electrode. Such examples are, however, not typical for applications of ion-selective electrodes.

Typical Drug Substance-Selective Electrode

The typical drug substance-selective electrodes are usually of the liquid-membrane type. It should be noted that virtually any ionic species can be detected and measured by liquid ion-exchange electrodes. The principle for design is as follows: To build a membrane responsive to anion X^- , for example, the salt Q^+X^- is dissolved in a nonvolatile solvent; the Q^+ cation must be highly lipophilic. Similarly, for an electrode responsive to cation Q^+ , an oil-soluble salt Q^+X^- is used, where the X^- anion is lipophilic. Thus, the quaternary long-chain alkyl and aryl ammonium salts and high-molecular-weight cationic dyes etc. are known to behave as liquid anion exchangers suitable for the preparation of anion-selective liquid-membrane electrodes. Tetraphenylborate [or tetrakis(substituted phenyl)borates] and high-molecular anionic surfactants such as dodecylsulfate show good selectivity for heavier univalent inorganic cations and are also often used in membrane electrodes for other “onium” ions. These rules are observed in the construction of drug substance electrodes; a few examples are given in Table 3.

Titration Based on Ion-Pair Formation

In these titrations (28–31), a cationic (Q^+) or anionic (X^-) species is titrated with an oppositely charged

titrant (X^- or Q^+), respectively. If the substance determined or the titrant (or both) has adequate lipophilic character, the poorly soluble ionpair precipitates, $\text{Q}^+ + \text{X}^- = \text{QX}$. Further, the same ion pairs can be used as active substances for liquid membrane-type electrodes because of their good extractability into organic water-immiscible solvents. For that reason, simplified sensors can be used to monitor such titrations. One strategy used with some success is that of coated-wire electrodes. Here, the membrane material is applied directly onto a metal wire (this assembly has no internal reference electrode, which is replaced by a direct contact). In addition, the membranes can be prepared from polymer solutions containing no active material. The organic phase (plasticizer) of a sensor, when immersed in a stirred aqueous suspension of the QX ion pair, becomes gradually saturated, and the concentration is given by Eq. 32:

$$\begin{aligned} [\text{QX}]_{\text{org}} &= K_{\text{ex}}(\text{QX})[\text{Q}^+]_{\text{aq}}[\text{X}^-]_{\text{aq}} \\ &= K_{\text{ex}}(\text{QX})K_{\text{s}}(\text{QX}) \end{aligned} \quad (32)$$

where $K_{\text{ex}}(\text{QX})$ and $K_{\text{s}}(\text{QX})$ are the stoichiometric extraction constant and the solubility product of QX , respectively. The $[\text{QX}]_{\text{org}}$ concentration indicates the number of ion-exchanging places in the membrane, although surface adsorption can also participate. This is why all liquid and/or plastic membrane electrodes sensitive to ions other than those to be determined can also be used to monitor titrations based on ion pair formation. Simple and inexpensive sensors of the coated-wire type, prepared by dipping the central conductor (Pt, carbon rod, but also aluminium wire can be used) into a solution containing dissolved polymer and plasticizer and allowing the solvent to evaporate, are sufficiently suitable.

Regarding titrants, cationic substances (for example, protonized alkaloids, compounds containing quaternary nitrogen, etc.) are usually titrated with sodium tetraphenylborate, the exact concentration of which is determined titrimetrically against a standard substance such as thallium(I) nitrate or pure copper(II) or nickel(II) salts in the presence of 1,10-Phenanthroline. For titrations of anionics, substituted quaternary ammonium or pyridinium salts are applied (30). The procedures are simple and represent an ecologic alternative to so-called two-phase titrations (32).

Titration Based on Complex Formation

Organic molecules containing atoms such as nitrogen, oxygen, and sulfur can donate pairs of electrons.

Table 3 Examples of drug substance-selective electrodes

Substance	Type	Electroactive compound in membrane	Mediator
Acetylcholine	PVC	Acridine orange reineckate	DOP
Amitriptyline	Liq	Eosin, tetraphenyl- or tetrakis(3-Chlorophenyl)-borates as counterions	NC, NT
Amphetamine	Liq	Amphetamine octadecylsulfate	NB
Atropine	Liq or PVC	Ion-pairs of atropine with tetraphenyl- or tetra-kis(3-chlorophenyl)borates, reineckate, dipicrylamine, or tetraiodomercurate(II)	BA, NB, NT, OA or BEHP, DNP, DOP
Bamethan	PVC	Bamethan tetraphenylborate	Various
Brucine	PVC	Potassium tetraphenylborate	DBP
Bupivacaine	PVC	Bupivacaine dinonylnaphthalenesulfonate	
Butylscopolamine	PVC	<i>N</i> -Butylscopolamine tetraphenylborate	DBP
Cholic acid	Liq or PVC	Benzyltrimethylammonium cholate or tributylcetylphosphonium benzoate	NB
Codeine	Liq	Codeine dipicrylamine	NB
	PVC	Potassium tetraphenylborate	DBP
Ephedrine	Liq	Ephedrine 5-Nitrobarbiturate or flavianate	NB or OA
	PVC	Ephedrine tetraphenylborate	DOP
Glutamates	Liq	Methyltricaprylammonium glutamate	DA
Lidocaine	Liq	Lidocaine dipicrylamine or reineckate	NB
	PVC	Tetrakis(3-Chlorophenyl)borate or dinonyl-naphthalenesulfonate as counter ions	DNP or
Nicotine	Liq or PVC	Nicotine tetraphenyl- or tetrakis(3-Chloro-phenyl)borates	NT, NB
Novocaine	Liq	Novocaine tetraphenylborate or dipicrylamine	NB
Oxalate	Liq	Tricaprylmethylammonium oxalate	DA
Papaverine	Liq	Ion-pairs with alkylsulfates, arenesulfonates, or tetraphenylborate	NB
Phencyclidine	PVC	Phencyclidine dinonylnaphthalenesulfonate	DOP
Pilocarpine	PVC	Pilocarpine reineckate or tetraphenylborate	DBP
Quinine	PVC	Quinine tetraphenylborate	DBP, DBS, NB, NPOE
Salicylate	Liq	Tricaprylmethylammonium or tetrahexyl-ammonium salicylate	DA
	PVC	Ethyl violet as a counterion	
	Epoxy	Trioctylmethylammonium salicylate	
Strychnine	Liq	Strychnine picrolonate or tetrakis(3-Methyl-phenyl)borate	
Sulfamerazine	Liq	Tris(bathophenanthroline) iron(II) as a counterion	NB
Vitamins B ₁ , B ₆	Liq	As for papaverine	

Liq = liquid membrane with ion-exchange solution into the porous diaphragm; PVC = plasticized poly(vinyl chloride); Epoxy = conductive epoxy resin membrane. BA = benzyl alcohol; BEHP = bis(2-Ethylhexyl) phthalate; DA = 1-Decanol; DBP = dibutyl phthalate; DBS = dibutyl sebacate; DNP = dinonyl phthalate; DOP = dioctyl phthalate; NB = nitrobenzene; NPOE = 2-Nitrophenyl octyl ether; NT = 2-Nitrotoluene or 4-Nitrotoluene; OA = 1-octanol.

As electron donors, they can form complexes with metal ions (M), which are electron acceptors and are bound as unidentate or multidentate ligands (L). Depending on other groups of the organic molecule, these complexes can be charged or neutral and soluble or insoluble in water. Thus, the chemical reaction can be considered to be $nL + mM = M_mL_n$ (charges are omitted). Experience has shown that the reaction noted above is affected by the pH value and by

competitive side reactions caused by various substances usually present in the solution (such as other ligands from buffer constituents or counterions). Thus, if the conditional stability constant of the complex $\beta(M_mL_n)$ is sufficiently high, the ligand compound can be determined by titration with solutions of metal salts. Appropriate metal ion-selective electrodes are used to monitor these titrations. Regarding pharmaceuticals, many of them can be titrated

with solutions of copper, lead, mercury, silver, and other metal salts (28, 29).

Titration in Nonaqueous Media

Potentiometric measurements in nonaqueous media (33, 34) are performed in a similar manner and with similar apparatus as in aqueous media. There is a difference in the composition of the salt bridge established between the electrode and the solution being tested. In titrations in inert solvents, moreover, there are differences in the shielding and grounding of the titration vessel. Cell voltage is measured using a potentiometer of high internal resistance. Nonaqueous media are used because the substance has low water solubility or it acts as a weak acid or a weak base in aqueous solution. In solvents similar to water, the interactions between solute and solvent resemble the processes that take place in water, but differences owing to the smaller dielectric constants, the formation of different ion pairs by further associations, and the change in the electrostatic field as a function of concentration can be observed. Potentiometric titrations in nonaqueous media can be applied over a wide area; the endpoint varies with the strength of the acid or base to be determined, the solvent, and the titrant. For the determination of water-insoluble weak bases, proton-donating acidic solvents (such as acetic or propionic acids) are used. A titrant is perchloric acid that, dissolved in the medium, gives acetacidium ions $\text{CH}_3\text{COOH}_2^+$. Proton-acceptor basic solvents, such as ethylenediamine, pyridine, butylamine, and dimethylformamide, are used for the solutions of acids and compounds with similar behavior (phenols, imides, sulfonamides, etc.). In this case, the conjugate base of the dissolved acid and the solvated proton are formed; tetrabutyl ammonium hydroxide or other basic titrants are used. In inert solvents such as hexane, benzene, carbon tetrachloride, etc., the highly variable strengths inherent in acids and bases are displayed. The solvent does not take part in the neutralization process, and the product formed is of an additive or associative character.

Titration with Sodium Nitrite

This is a traditional titration that is still often used to determine compounds containing primary amino groups on the aromatic ring, with the endpoint being monitored with a Pt indicator electrode. This method, applied to the determination of pharmaceuticals (sulfonamide-based compounds, benzocaine, procaine, etc.), is rapid, and the results are in good agreement with official methods of analysis (35).

Titration Based on Azo-Coupling Reactions

Aromatic diazonium salts, obtained in titrations with sodium nitrite, can be used in azo-coupling reactions (36). Various aromatic amines, phenols, and compounds containing active methylene groups can be titrated with arenediazonium salts, from which 4-Bromo-1-Naphthalenediazonium chloride seems to be the most widely applicable titrant. Compounds that react slowly with arenediazonium salts can be determined by back-titration when the excess of arenediazonium salt is back-titrated with either sodium tetraphenylborate or 2,4-Diaminotoluene. Indirect determination is useful for secondary amines, which react with arenediazonium ions to form triazenes. The determination of diazonium salts of ampholytic character is based on the reaction of these salts with 1-Phenyl-3-Methyl-5-Pyrazolone, the excess of which is titrated with 4-Bromo-1-naphthalenediazonium chloride solution.

Kinetic Methods

These have also been applied to the analysis of drug substances because they have advantages over equilibrium techniques, especially when mixtures of closely related compounds, compounds that react slowly, or catalytically acting compounds are to be analyzed. The selectivity and sensitivity of kinetic methods of analysis combined with the selectivity and sensitivity of ion-selective electrodes provide a versatile combination that may lead to new analytical schemes (37).

REFERENCES

1. Compton, R.G.; Sanders, H.W. *Electrode Potentials*; Oxford Univ. Press: Oxford, 1996.
2. Brett, M.A.; Oliveira Brett, A.M. *Electroanalysis*; Oxford Univ. Press: Oxford, 1998.
3. Bailey, P.L. *Analysis with Ion-Selective Electrodes*; Heyden: London, 1976.
4. Lakshminarayanaiah, N. *Membrane Electrodes*; Academic Press: New York, 1976.
5. Veselý, J.; Weiss, D.; Štulík, K. *Analysis with Ion-Selective Electrodes*; Horwood: Chichester, 1978.
6. Koryta, J.; Štulík, K. *Ion-Selective Electrodes*; 2nd Ed.; Cambridge Univ. Press: Cambridge, 1983.
7. Buck, R.P. *Electrochemical Methods: Ion-Selective Electrodes*. *Water Analysis*; Minear, R.P., Keith, L.H. Eds.; Academic Press: Orlando, 1984; II, 250–321.
8. Janata, J. *Principles of Chemical Sensors*; Plenum Press: New York, 1989.
9. Midgley, D.; Torrance, K. *Potentiometric Water Analysis*, 2nd Ed.; Wiley: Chichester, 1991.

10. Cattrall, R.W. *Chemical Sensors*; Oxford Univ. Press: Oxford, 1997.
11. Eisenman, G. *Glass Electrodes for Hydrogen and Other Cations*; Marcel Dekker, Inc. New York, 1967.
12. Arnold, M.A. *An Introduction to Biocatalytic Membrane Electrodes*; Int. Lab., 1983; 13 (6), 24–32.
13. Guilbault, G.G.; Schmidt, R.D. Biosensors for the Determination of Drug Substances. *Biotechnol. Appl. Biochem.* **1991**, 14, 133–145.
14. Kauffmann, J.M.; Guilbault, G.G. Enzyme Electrode Biosensors: Theory and Applications. *Bioanalytical Application of Enzymes*; Suetler, C.H., Ed.; Wiley: New York, 1992; 36, 63–113.
15. Pratinis, D.M.; Telting-Diaz, M.; Meyerhoff, M.E. Potentiometric Ion-, Gas-, and Bioselective Membrane Electrodes. *Crit. Rev. Anal. Chem.* **1992**, 23, 163–186.
16. Buerk, D.G. *Biosensors, Theory and Applications*; Technomic Publ. Co. Lancaster, 1993.
17. Diamond, D. *Principles of Chemical and Biological Sensors*; Wiley: New York, 1998.
18. Bates, R.G. *Determination of pH, Theory and Practice*; Wiley: New York, 1964.
19. Buck, R.P.; Cosofret, V.V. Recommended Procedures for Calibration of Ion-Selective Electrodes (Technical Report). *Pure Appl. Chem.* **1993**, 65, 1849–1858.
20. Bakker, E.; Bühlmann, P.; Pretsch, E. Polymer Membrane Ion-Selective Electrodes—What are the Limits? *Electroanalysis* **1999**, 11, 915–933.
21. Ríoka, J.; Hansen, E.H. *Flow-Injection Analysis*; Wiley: New York, 1981.
22. Štulík, K.; Pacáková, V. *Electroanalytical Measurements in Flowing Liquids*; Horwood: Chichester, 1987.
23. Baiulescu, G.E.; Cosofret, V.V. *Applications of Ion-Selective Membrane Electrodes in Organic Analysis*; Horwood: Chichester, 1977.
24. Ma, T.S.; Hassan, S.S.M. *Organic Analysis Using Ion-Selective Electrodes*; Academic Press: London, 1982; 1/2.
25. Cosofret, V.V. *Membrane Electrodes in Drug-Substance Analysis*; Pergamon Press: Oxford, 1982.
26. Cunningham, L.; Freiser, H. Ion-Selective Electrodes for Basic Drugs. *Anal. Chim. Acta* **1982**, 139, 97–103.
27. Cosofret, V.V.; Buck, R.P. Drug-Substances Analysis with Membrane Electrodes, Ion-Sel. *Electrode Rev.* **1984**, 6, 59–121.
28. Vytas, K. The Use of Ion-Selective Electrodes in the Determination of Drug Substances. *J. Pharm. Biomed. Anal.* **1989**, 7, 789–812.
29. Vytas, K. Contemporary Trends in the Use of Ion-Selective Electrodes in the Analysis of Organic Substances. *Advanced Instrumental Methods of Chemical Analysis*; Churáek, J., Ed.; Academia, Praha, and Horwood: Chichester, 1993, 142–164.
30. Vytas, K. Potentiometric Titrations Based on Ion-Pair Formations. *Ion-Sel. Electrode Rev.* **1985**, 7, 77–164.
31. Schulz, R. *Titration von Tensiden und Pharmaka, Moderne Methoden für den Praktiker*; Verlag für chemische Industrie H. Ziolkowski: Augsburg, 1996.
32. Vytas, K.; Kalous, J.; Jeková, J. Automated Potentiometry as an Ecologic Alternative to Two-Phase Titrations of Surfactants. *Egypt. J. Anal. Chem.* **1997**, 6, 107–123.
33. Gyenes, I. *Titration in Non-Aqueous Media*; Akademiai Kiado: Budapest, 1967.
34. Šafařík, L.; Stránský, Z. *Titrimetric Analysis in Organic Solvents*; Elsevier: Amsterdam, 1986.
35. Šubert, J. Drug Analysis. VI. Studies of Conditions for Diazotizing Titrations of Drugs with Potentiometric Indication and Standardization of a Volumetric Sodium Nitrite Solution. *Farm. Obz.* **1981**, 50, 273–278.
36. Vytas, K. Ion-Selective Electrodes in Titrations Involving Azo-Coupling Reactions, Parts 1–5. *Anal. Chim. Acta* **1982**, 141, 163–171, (1984); 162, 141–151; (1984); 162, 373–377 (1985); 175, 309–312 (1985); 175, 313–317.
37. Efstathiou, C.E.; Kouparis, M.A.; Hadjiioannou, T.P. Application of Ion-Selective Electrodes in Reaction Kinetics and Kinetic Analysis. *Ion-Sel. Electrode Rev.* **1985**, 7, 203–259.

Powder Sampling

Helena J. Venables

Liverpool John Moores University, Liverpool, United Kingdom

INTRODUCTION

Taking a large amount of powder for analysis is expensive and wasteful, and so it must be reduced.

Sampling in its strict sense is therefore a simple mass reduction.^[1]

The purpose is to collect a manageable amount of the powder which is representative of the batch as a whole. Many small samples are taken from all parts of the whole, which, together, are representative.^[2] Theoretically, every particle should have an equal chance to be selected for examination,^[3] and this should be attained with minimum disturbance to the system.^[4] The composition of the original powder must therefore be retained during sampling.^[1] Powder characteristics change under an applied load and attrition and segregation may occur in transfer. Representative sampling is essential for the relevance of any subsequent testing^[2] and is defined by the selection method, which must be accurate and reproducible. To produce reliable samples, a correctly designed sampler is required.^[1] A review of sampling techniques is shown in Table 1.

Important rules should be followed to optimize sampling:

1. Sample from a moving stream of powder.
2. Sample the whole stream for equal periods of time, rather than part of the stream for all of the time.^[2]

OVERVIEW

All analysts have a personal preference for sampling patterns, and random number tables can eliminate this problem.^[6]

Sampling can be undertaken in two ways:

1. By taking increments on flowing streams of powder.
2. By splitting, where the whole is handled.^[1]

The number, the size of each sample, and the method should all be considered. The acceptable error dictates the number of samples to be taken. The more samples taken,

the smaller the error.^[7] Despite extreme care, error causes a difference in composition of a sample from that of the batch.^[8]

It is extraordinarily difficult to carry out any selection process without introducing individual preference or bias into the final selection.^[6]

When determining appropriate sample size, the Scale of Scrutiny must be considered.

SCALE OF SCRUTINY^a

When a mixture is examined closely, regions of segregation are often found. The smallest region that can measure imperfections in a mix is the "Scale of Scrutiny" and can be a length, area, volume, or weight.^[9] The Scale of Scrutiny is directly related to the finished product. If the product is a tablet, then the Scale of Scrutiny is the weight of a tablet. The lower limit is set by particle size.^[10] When the Scale of Scrutiny is large, or when the particle size is small, a large number of particles are present in a sample and the mix appears uniform.^[9] Segregation can never be fully avoided, and sampling the final product is therefore preferred.^[7]

MANUAL SAMPLING^a

Sample Thief

This tool can be useful for free-flowing powders, with two concentric tubes, one enclosing the other. The outer tube is pointed, with holes cut in corresponding positions in inner and outer tubes. The holes are opened or closed through the rotation of the inner tube to capture material (Fig. 1).^[12]

Another variation is a concentric, inner slotted tube with a shoulder at the lower, pointed end. The outer tube has no holes and sits on the shoulder of the inner tube in the closed position. The thief is pushed in the powder diagonally and the outer tube is raised to expose the slot,

^aReprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.



Table 1 Powder sampling

Apparatus	Type of sampler	Method of sampling	Scale	Advantages	Disadvantages
1. Sample thief	Manual	Tubular steel retains a core sample when inserted into powder	Large and small	<ul style="list-style-type: none"> • Good for free flowing powders 	<ul style="list-style-type: none"> • Additional weight at the bottom of bag may vary sample size • Can be hard to push into powder • Fines may lodge between tubes • Particles can fracture • Fines compact, impeding flow • Segregation may occur as fines percolate into sample more easily than coarse particles • A plug of powder can be pushed ahead of thief and surface material contaminates sample • Personal preference introduces bias for the area sampled
2. Hand scoop	Manual	Cross-sectional sample from moving stream, bags, or barrels	Large	<ul style="list-style-type: none"> • Simple • Cheap 	<ul style="list-style-type: none"> • Thin layer may remain on belt, leading to bias • Overfilling can lead to an excess of fines
3. Shovel	Manual	Pits are dug in the powder bed and a shovelful was taken from bottoms and sides	Large, up to several tons	<ul style="list-style-type: none"> • Simple • Cheap 	<ul style="list-style-type: none"> • Cannot be used with particles of more than 5 cm in diameter
4. Cross-cut sampling	Manual/semiautomatic/automatic	Material is shoveled from the conveyor belt	Large	<ul style="list-style-type: none"> • Simple 	<ul style="list-style-type: none"> • Can leave a layer on the belt and result in bias
5. Pneumatic lance	Semiautomatic	Air flow used on entry and exit of lance from powder bed	Large	<ul style="list-style-type: none"> • Disturbing of powder minimized over sample thief • Porous plate prevents too many fines because of strong air current 	<ul style="list-style-type: none"> • Personal preference may bias sample
6. Vacuum probe sampler	Semiautomatic	Powder extracted by vacuum	Large	<ul style="list-style-type: none"> • Simple 	<ul style="list-style-type: none"> • Difficult to sample below surface without contamination • Personal preference leads to bias • Fines more easily extracted than coarse particles

(Continued)

Table 1 Powder sampling (*Continued*)

Apparatus	Type of sampler	Method of sampling	Scale	Advantages	Disadvantages
7. Gravity-flow auger sampler	Semiautomatic	Slotted tube in flowing powder rotates and worm screw carries out material	Large and small	<ul style="list-style-type: none"> • Easy to use 	<ul style="list-style-type: none"> • Bias is still a problem here • Difficult to sample all of powder stream, therefore bias
8. Sampling from a moving stream	Manual/automatic	Powder is sampled as it falls off the conveyor	Large	<ul style="list-style-type: none"> • If carried out properly, it can be a very good sampling technique • Easily designed into a new plant 	<ul style="list-style-type: none"> • Segregation can occur on the conveyor belt • If overfilled, a greater number of fines than coarse particles are collected • Difficult and expensive to fit into an existing plant • Obtaining a fixed sample is difficult • Difficult to prevent dust escaping
9. Full-stream trough sampler	Automatic	Samples powder as it falls off conveyor	Large	<ul style="list-style-type: none"> • Can be used to sample dusty material 	<ul style="list-style-type: none"> • Must not overfill • Difficult and expensive to install into existing plant
10. Arc path cutter	Automatic	Chute moves through powder stream and collects sample	Large	<ul style="list-style-type: none"> • No operator bias 	<ul style="list-style-type: none"> • Difficult to take more than one sample size • Must cover the whole of the stream to avoid bias
11. Straight path cutter	Automatic	Rectangular chute moves through powder stream	Large	<ul style="list-style-type: none"> • Different sample sizes can be taken easily 	<ul style="list-style-type: none"> • Must cover the whole of the stream to avoid bias
12. Moving-flap sample divider	Automatic	A flap in the stream samples powder or allows it to be stored depending on its position	Large	<ul style="list-style-type: none"> • Efficient as it covers the whole of the stream when sampling 	<ul style="list-style-type: none"> • Resultant sample is large and subsampling is needed. Bias may be introduced here • Bias is present because of one side being sampled more than the other
13. Integrated automatic sampling plant	Automatic	Primary sample is selected and repeatedly screened, resulting in the final sample	Large	<ul style="list-style-type: none"> • Quick • Convenient 	<ul style="list-style-type: none"> • Variations in materials can cause problems • Bias increases as the plant is worn
14. Chute splitter	Subsampler	A series of chutes split sample repeatedly	Large	<ul style="list-style-type: none"> • Can be repeated until the desired sample size is achieved 	<ul style="list-style-type: none"> • If segregation occurs, the result can be misleading • Prone to operator bias

(Continued)

Table 1 Powder sampling (*Continued*)

Apparatus	Type of sampler	Method of sampling	Scale	Advantages	Disadvantages
15. Cone and quartering	Subsampler	Powder poured through cone and divided into 4 equal parts. This is repeated until the desired sample size is reached	Small	<ul style="list-style-type: none"> • Simple 	<ul style="list-style-type: none"> • Prone to operator bias as fine particles remain in the center of the cone • Symmetry is difficult to achieve but essential for accuracy
16. Spinning riffler	Subsampler	A steady stream of powder flows into a rotating basket of containers	Large/small	<ul style="list-style-type: none"> • Good for subsampling large samples • Good for powders with good flow properties • Minimal bias • More efficient than other samplers tested 	<ul style="list-style-type: none"> • Air currents may displace fines but can be avoided with a slower rotation speed • Expensive • Time consuming • Segregation may be a problem
17. Free-fall tumbler mixer	Subsampler	A ladle in the lid of the mixer collects a good representative sample	Large/small	<ul style="list-style-type: none"> • Can be used with fines present • Representative sample produced in a short time period 	
18. Hopper sample divider	Subsampler	Hopper oscillates and powder falls into two containers. Only one of the contents is kept	Small	<ul style="list-style-type: none"> • Sample size can be controlled by monitoring time over each container 	<ul style="list-style-type: none"> • Large number of increments needed for accuracy • Very low accuracy
19. Table sampler	Subsampler	Powder flows down inclined plane and prisms and holes split the powder	Small	<ul style="list-style-type: none"> • Simple 	<ul style="list-style-type: none"> • After each separation should be a complete mix to avoid errors

Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

which must face upward. The material falls into the container and the tube was removed, slowly opening the tube to collect a representative sample from all parts of the powder bed.^[12]

The thief can sometimes be fairly hard to push into the powder bed, and it is not a good idea to have too many close-fitting, moving surfaces when handling powders, because fines clog and bind between them.^[13] Particles may fracture or fine material may compact by the motion of the sampler, impeding flow. Bias is introduced because of the personal preference for the area sampled. Segregation may occur because of fines percolating into the sample more easily than coarse particles.^[11] A plug of powder can be pushed ahead of the thief and the surface material will contaminate the sample.^[6]

An alternative technique, which reduces segregation, is a unit-dose compacting sample thief (Fig. 2). Bias is still likely with this method because of the invasive nature of any sample thief.^[14]

Hand Scoop

A scoop can take a cross-sectional sample from a bag, barrel, or flowing stream. From the powder stream, a single movement of the scoop completely across the stream collects the sample. Opposite directions for each collection should be used.^[12] The scoop width should be at least 2.5–3 times the largest dimension of the particles to prevent overfilling leading to an excess of fines,^[15] and care must be taken not to leave a thin layer of powder on the belt.^[11] With fines, a larger scoop minimizes moisture

loss, but the volume may be too large and using a shovel would be more convenient.^[15] This method is widely used in manufacturing industries as it is simple and cheap.^[11]

Sampling with a Shovel

In this method, an imaginary grid is applied to a pile of material. At the intersections, pits are dug, roughly 30 cm deep, and a shovelful is taken from the bottoms and sides of each hole.^[12] This method is only applicable to large amounts (up to several tons) of material, but it cannot be used with particles larger than 5 cm in diameter.^[15]

Cross-Cut Sampling

Using a conveyor belt, the belt is stopped and a sample is taken by hand or an automatic or semiautomatic device. A suitable size for sampling can be designated by markings on the conveyor belt and all material within markings was collected. If an automatic system is used, a mechanically operated head moves across the material at a preset interval and moves a sample into the collection point.^[12]

SEMI-AUTOMATIC SAMPLING^a

Pneumatic Lance

This method is used for bulk powder. A gentle flow of air out of the nozzle allows the probe to move through the powder bed. When at the site, air is very slowly reversed

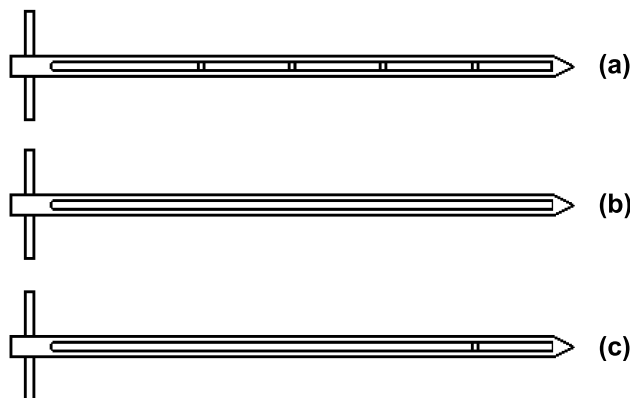


Fig. 1 Sample thieves. (a) Separate holes along the tube, useful for segregation determination. (b) Full length chamber, useful for spot sampling. (c) Chamber at pointed end only to gain an average from many samples.^[11] Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.



Fig. 2 The unit-dose compacting sample thief. The sample is collected in the “S” section and the outer rod rotates to open and close the thief. Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

to draw up a sample, which is collected against a porous plate at the end of the probe.^[13] The porous plate prevents great numbers of fines being present in the sample if the air current is too strong (Fig. 3). The pneumatic lance minimizes powder disturbance and therefore has an advantage over a sample thief, although bias is still a problem.^[13]

Vacuum Probe Samplers

Large samples can be extracted from bins or holds using a vacuum cleaner principle. Contamination is a problem when sampling below the surface, fines are extracted preferentially to coarse particles, and bias is a problem.^[12]

Gravity-Flow Auger Sampler

A slotted tube is rotated in a flowing mass, and the material collected is carried out of the tube by a worm screw. It is difficult to sample the entire stream, and bias remains to be a problem.^[12]

SAMPLING FROM A MOVING STREAM^a

With these methods, great care should be taken to avoid the effect of segregation. The powder is taken as it falls from the conveyor and can display two types of segregation:

1. If the powder has been charged on the conveyor, fines concentrate at the center of the belt while coarse particles roll to the outer edges such as the segregation seen when sampling from a heap.
2. If the powder is exposed to vibration on the belt, percolation causes larger particles to rise to the top of the powder bed and fines remain at the bottom.

The whole of the powder must be collected for a short period of time, and care must be taken moving the sampler in and out of the stream.^[11]

Various methods exist for sampling from a moving stream and three of these are described (Fig. 4).



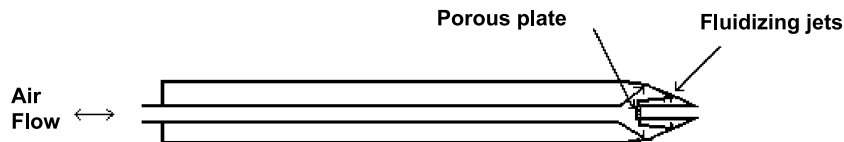


Fig. 3 The pneumatic lance. Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

With these methods and all others that sample from a moving stream of powder, the dimensions of the sample collector are important. The width of the receiver should be greater for smaller particles than for larger ones because of the tendency for fines to drift during the movement through air. To minimize error, the ratio of box width to particle diameter should be at least 20:1, larger when sampling fines. The depth should be large enough to prevent the receiver from becoming full while sampling. If it overfills, fines will percolate through the heap that forms, while coarse particles are lost. The receiver length should be sufficient to sample the full length of the powder stream.^[11] Disadvantages of this method include:

1. It is difficult and expensive to add to an existing plant but is easily designed into a new one.
2. The sample quantity is proportional to plant rate, which can vary greatly. It can therefore be difficult to obtain a fixed sample, and sample quantity may also be too large.
3. Enclosing the sampler to prevent dust escaping can be difficult.^[11]

Full-Stream Trough Sampler

This can be used for sampling dusty material. Sampling is only carried out on the return stroke (Fig. 5). The trough must not overflow.^[11]

BULK SAMPLING^a

Bag Sampling

When many tons of material need sampling from sacks, several sacks should be selected systematically, i.e., 100th, 200th, 300th, etc. or through use of a random number table or $\sqrt{n}-1$ (9 of 100). Each sample is individually assessed to determine the variation and if it is acceptable, or combined to give an average. The thief sampler is recommended, but this can lead to bias.^[16] It is preferable to use a spinning riffler; however, this is expensive to install. It does avoid rejecting good material and accepting poor material which would occur with the above, inexpensive technique.^[11]

Sampling from Wagons and Containers

This is almost impossible to do satisfactorily because of the major segregation in the filling and vibrations in transport. Sampling should not be carried out in the top 30 cm to avoid the segregation in the surface layer that will have occurred as a result of vibration. When removing the samples, no surfaces on which particles can slide should be introduced. This can be achieved using a sample thief.^[11]

Extraction should be carried out at eight points in the bed (Fig. 6). This method of sampling is not satisfactory but may have to be used in certain circumstances.^[11]

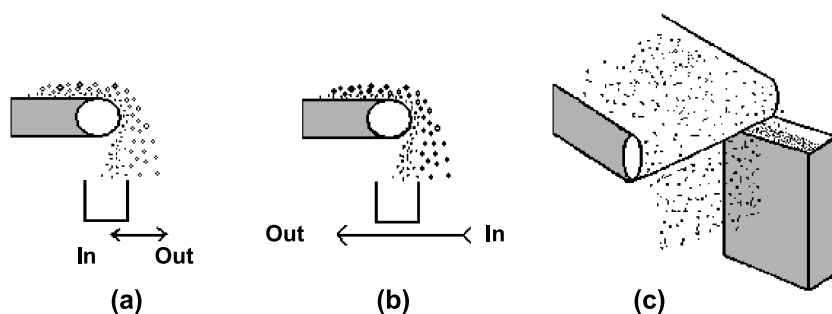


Fig. 4 Methods of sampling from a moving stream. (a) This leads to an excess of coarse particles, if a longer time is taken while stationary than inserting the receiver and removing it, because of the surface usually being rich in coarse particles. This region of the stream is sampled for longer time. (b) This shows good sampling technique. One movement samples the entire stream, but this is not always possible as a result of obstructions. The ratio of stationary time to moving time for the receiver should be as large as possible. (c) This method should be used only when it is not possible to collect the whole of the stream because of too large a sample. A sample collector is passed through the powder stream. Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

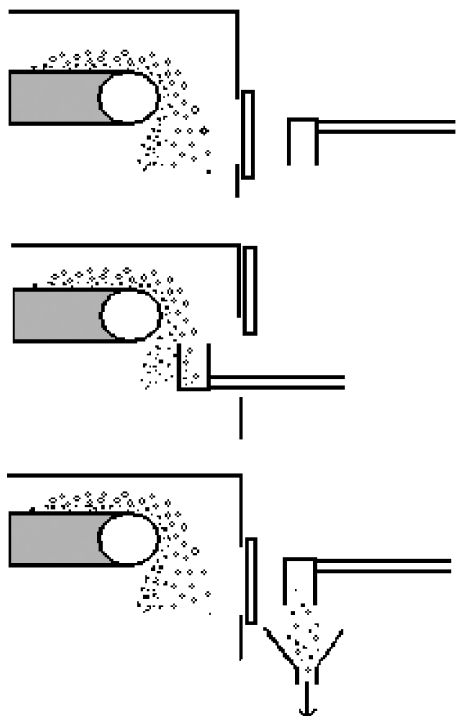


Fig. 5 The full-stream trough sampler. Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

SAMPLING FROM HEAPS^a

This should not be carried out because marked segregation occurs, with fine particles concentrating in the center of the cone (Fig. 7). Powder should preferentially be sampled from a moving stream of powder instead.^[11]

AUTOMATIC SAMPLING^a

A predetermined amount of powder is sampled at regular intervals or continuously. Because of the inherent bias, regular inspections are needed. All the material should be sampled at a constant rate, and clogging of the material should be avoided.

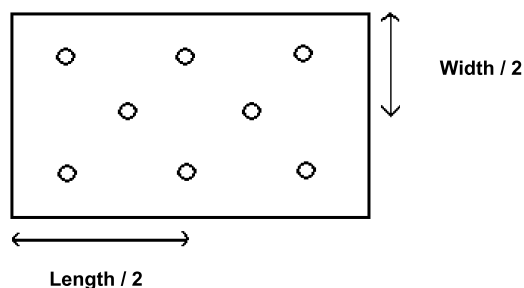


Fig. 6 Locations for sampling from wagons and containers. Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

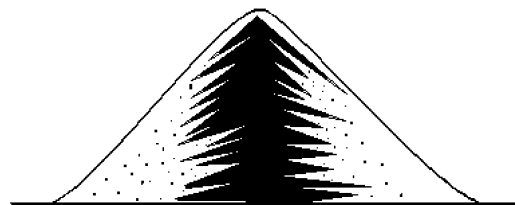


Fig. 7 Segregation in powder heaps. Finer particles are illustrated in black and coarser particles are in white. Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

Arc Path Cutter

A chute at an angle is mounted on a vertical shaft, which rotates to periodically move the chute through the stream of powder for a set time. By increasing the speed of rotation, an increase in the number of increments is taken, but this does not alter sample size.

Straight Path Cutters

A rectangular chute moves through the powder stream and rests outside it for a variable time period. This allows different sample sizes to be taken relatively easily. The whole of the stream must be covered to avoid bias.^[15]

Moving-Flap Sample Divider

A flap pivoted about a horizontal axis can rest in either of two positions. The powder stream flows to storage or to be sampled and the time for the flap to be in either position can be controlled automatically (Fig. 8). If the sample is too large, a table sampler is used and an extra powder is returned to the store. This method would be efficient if it was not for the use of the table sampler which causes error.^[11] A slight bias is involved as one side of the stream is sampled more than the other.

Integrated Automatic Sampling Plant

This apparatus selects the primary sample, which is then screened repeatedly to produce the final sample. The method is quick and convenient; however, variations in materials can cause problems and bias in sampling increases as the plant becomes worn.^[15]

SUBSAMPLING^a

This process results in a representative small sample taken from a large one.^[15]



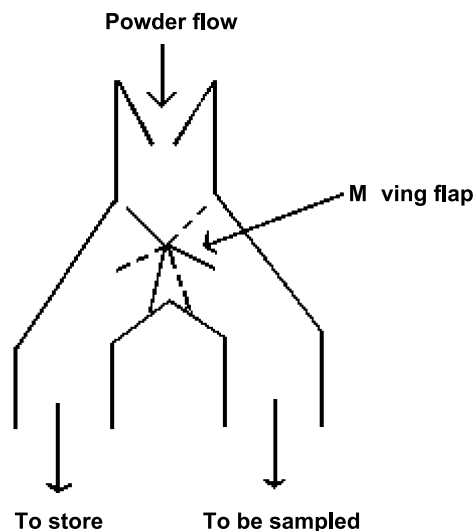


Fig. 8 A moving-flap sample divider. Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

Cone and Quartering

Powdered material is poured into a cone and divided into four equal parts. One quarter is recovered and the process was repeated until the required weight of sample is obtained. This method is prone to operator bias, as finer particles tend to remain in the center of the cone because of size segregation.^[17] Symmetry of the cone is also quite difficult to achieve but is necessary for accuracy, so this method is not recommended.^[11]

Chute Splitting

This apparatus is a V-shaped trough, at the bottom of which is a series of chutes feeding two trays on either side

of the trough. The sample enters the chute and is halved repeatedly until the desired sample size is achieved.^[17] The distance between the slots should be at least three times the size of the largest particle diameter.^[15] This method is prone to segregation and operator bias because of the unequal splitting of the sample. It has been found that a greater number of chutes can increase the overall efficiency.^[11]

The Spinning Riffler

A spinning riffler works by running a steady stream of powder into a rotating basket of containers (Fig. 9).^[13] Two or more portions can be collected and it is a continuous dividing process.^[15] The sample of powder contained in each basket can be made more representative by making the baskets rotate faster or increasing the duration of sampling. Rotation speed is limited by air currents disturbing the powder.^[13] Large samples are more difficult to subsample and riffles are very useful in this situation.^[15]

If a powder has good flow properties, then a spinning riffler works very well; however, there should be only a small percentage of fines because air currents displace them and the sample is not representative. The riffler is useful for a single powder ingredient, but segregation can be a disadvantage. It is also fairly expensive and quite time-consuming;^[13] however, the spinning riffler has been found to be more efficient than cone and quartering and the chute splitter and is less prone to operator bias with free flowing powders.^[17]

The efficiency of one riffler was found to be dependent on the relative proportions of the mixture using a mixture of quartz and copper sulfate crystals. Efficiency increased with increasing particle size, was reproducible under similar experimental conditions, and was not affected by the number of volume units (where a volume unit was the

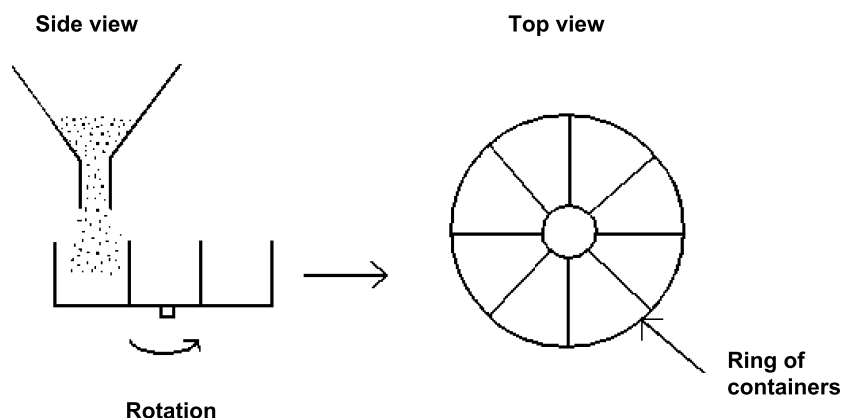


Fig. 9 The spinning riffler. Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

single presentation of a sample to the feed. The minimum was 100 in this experiment set, so this applies only to larger values). These conclusions may not apply where size segregation occurs.^[11]

In another set of experiments, the relative efficiencies of different samplers were compared, and it was concluded that very little confidence could be placed in cone and quartering, scoop sampling, or table sampling. The spinning riffler emerged as so superior to the others that it should be used as often as possible. With the riffler, a minimum of 35 presentations are needed for optimum results to be obtained, and if the speed of rotation is too great, the efficiency falls as a result of air currents setup.^[18]

Free-Fall Tumbling Mixer

This can be used if fines are present in a sample. A sample container is mounted with a ladle in its lid and after a short period of tumbling, it contains a representative sample of the powder as a result of the chaotic conditions created inside the mixer.^[13]

Hopper Sample Divider

The hopper feed oscillates on a horizontal axis, and the powder falls into two collectors. Only one of the contents is kept. Monitoring the amount of time over each container can control sample size. A large number of increments are needed for accuracy.^[11]

Table Sampling

The material is placed on the top of an inclined plane containing a series of holes. Prisms in the path of the stream fraction the powder: some falls into the holes and is discarded while the rest carries onto the next set of prisms and holes and the process is repeated. The powder that remains at the bottom of the plane is the sample. The initial feed must be uniformly distributed and after each separation must be a complete mix because of the errors occurring. There is a very low accuracy with this technique.^[11]

SAMPLING A SAMPLE^a

Having taken great care to extract a sample from a batch of powder, it is pointless to then take a small amount of this sample for analysis with little care of the technique used. By taking a sample from the bulk equivalent to the Scale of Scrutiny, this can minimize problems associated with powder handling. Analytical chemists tend not to realize that what they analyze is the end product of an intricate sampling plan.^[8]

Table 2 Errors involved in the sampling process

Sample	Bias (%)
Primary	≤ 1000
Secondary	≤ 50
Analysis	0.1–1

Reprinted from Ref. [5] by courtesy of Marcel Dekker, Inc.

They (chemists) are content to be mere mechanics, material grabbers. One wonders how many analysts there are that do not realise that taking the analysis sample from a bottle of pulp is a sampling process whose extraction error may be a magnitude greater than the analytical error.^[19]

Sampling is a stepwise process. Errors involved in the sampling process as a whole are illustrated by Table 2. The whole batch is represented by the first sample, the first by the second sample, and so on until the portion for analysis is produced. This is summarized in two main stages:

1. Primary sampling. This involves all steps outside the analytical laboratory resulting in a laboratory sample.
2. Secondary sampling. This includes all steps carried out within the analytical laboratory.^[8]

In the analytical laboratory, seemingly small issues can have large effects on the accuracy of sample analysis. When weighing a sample:

1. Powder may be left on the tray, but weighing by difference does not solve the problem fully. The powder remaining on the tray will most likely be cohesive fines, leading to bias as a result of exaggerated concentrations of coarse particles analyzed.
2. The powder may be “washed in,” leaving none behind. This is preferable, but the choice of solvent is critical to ensure extraction and delivery.^[20]

BLEND UNIFORMITY: A HISTORY

In 1992, the United States filed a court action against Barr Laboratories Inc. alleging that they had violated the Federal Food, Drug, and Cosmetic Act. Barr Laboratories then brought action against the United States, seeking relief from “ad hoc drug regulations.”^[21] Paracetamol (acetaminophen) and Codeine Phosphate tablets were recalled by the company because of doubts as to their potency.^[22] In judgment, Judge Wolin set new standards for the industry. Consequent discussions throughout the



pharmaceutical industry have led, to date, to the Product Quality Research Institute (PQRI) submitting a recommendation to the Food and Drug Administration (FDA) for new industrial guidelines. The ensuing debate is summarized in the following section.

United States vs. Barr Laboratories (1993)

Judge Wolin stated that end-product testing cannot be used to replace blend uniformity analysis (BUA) as he considered it to be limited. Sample size for BUA in validation and production batches should be up to 3 times the active ingredient dosage size, but could be up to 10 times if large errors could be identified and proven to be responsible for batch failures. In the event of failure, the same sample should be retested rather than a newly taken sample as the latter may disguise poor homogeneity. Resampling should not be used to release failed products, unless USP guidelines (content uniformity and dissolution) are followed or the original sample was prepared improperly. Samples should also be taken from weak or hot spots in the blend to demonstrate sampling as representative. Remixing was considered an invalid process which casts doubt on other passing batches; therefore frequent remixing was unacceptable.^[23]

Reactions to the Ruling

Following the “Barr decision,” the FDA suggested that blend uniformity tests should be more stringent than end-product tests as disorder increases as the process progresses.^[24] In response to the recommended sample size of no greater than three times the dosage size (now commonly referred to as the unit-dose sampling requirement), problems with sampling bias were raised. These lead to frequent batch failures and expense within the industry as results are unrepresentative of the blend uniformity.^[25] The entire sample should be analyzed and a duplicate sample retained in case retesting is required.^[26]

The unit-dose samples taken by four different commercially available thieves were compared, and the composition of the samples was found to vary greatly. One thief appeared to produce representative samples and limited bias, although only a single blend type was used.^[27]

Food and Drug Administration Blend Uniformity Draft Guidance (1999)

This document stated that BUA is not usually needed for more than 50 mg active per dosage form unit or more than 50% w/w active in a dosage form. End-product testing alone will normally suffice in these circumstances because of the large proportion of drug present in the blends. Under these limits, BUA is recommended as a routine test.

A sample size of less than or equal to three times the individual dosage weight is also recommended. If the sample bias is an issue and can be proven, up to 10 times the dosage weight can be collected. Six to ten samples should be obtained from drums or blenders, and the weight of the sample tested should be equivalent to the dosage weight used.^[28]

The response from the PhRMA Statistics Working Group was that sampling a blend with a sample thief can segregate the blend, thereby resulting in a sample that is unrepresentative and can be highly variable. Subsampling should not be recommended as it can be unrepresentative and highly variable. Judge Wolin had assumed that sampling using a thief consistently obtained representative samples throughout a powder bed. This is not the case and therefore, in process testing, is not needed.^[29]

Product Quality Research Institute Response

A Blend Uniformity Working Group (BUWG) workshop was held in September 2000, sponsored by PQRI, with attendants from the FDA, industry, and academia. The overwhelming response was that BUA on every batch was not a value-added test as there has been no statistical correlation between blend uniformity and end-product content uniformity. In preference, extended finished product testing was favored. The most common objection to BUA was limited sampling technology.^[30]

The BUWG called for data from pharmaceutical companies to verify the acceptance criteria stated in the FDA's draft guidance. Computer simulations were used, alongside the acquired data to test the hypothesis “blend uniformity testing in routine manufacture is not predictive of the uniformity of dosage units.”^[31] Using this input alongside the research carried out, a recommendation was drafted to the FDA. It stated that both blenders and Intermediate Bulk Containers (IBCs) should be sampled extensively during BUA to determine “dead spots” in blenders, segregation in IBCs, and sampling error presence. Blend sampling designs should then be developed for each product, including variables such as sample thief design and sample size, to try to minimize error and bias. Stratified sampling should be used. This technique deliberately targets weak spots throughout the blending and compression/filling operation to obtain samples. A minimum of 20 evenly spaced sample points for the dosage units should also be defined. These stringent techniques should reveal otherwise hidden errors occurring during process development and eliminate weighing errors in assay.^[32]

The FDA's Pharmaceutical Science Advisory Committee concluded at its May 8th, 2002 meeting that the PQRI recommendation for stratified sampling could be applied to routine production as well as bioequivalence

and validation batches. The FDA will internally review the PQRI recommendations and send them any concerns or further suggestions. The results will form an FDA guidance on content uniformity.^[32]

CONCLUSION

The Scale of Scrutiny must be carefully considered when sampling as both small and large samples can give very misleading results.

Sampling is integral to analytical techniques at all stages of manufacture especially in scale up. Undertaken correctly, it can prevent the losses that the company may suffer from otherwise. Although it is not revenue-producing, it is important economically and heavy losses can easily be avoided through the use of correct techniques.

Overall analytical reliability depends on sampling accuracy (negligible bias) and reproducibility (acceptable variance).^[8]

Guidelines are to sample from a moving stream of powder and to sample the whole of the stream for many equal periods of time. This ensures that all parts of the powder are equally accessible and this maximizes accuracy.^[2] The sample thief does not meet these criteria and therefore should not be used. Despite being popular, it suffers from many disadvantages, which lead to bias and error. Most other methods are immediately ruled out as suitable if these guidelines are followed.

The automatic sampling methods discussed are subject to bias, sampling some parts of the stream more than others. Sampling directly from a moving stream involves great care to avoid segregation, but method "b" in "Sampling from a Moving Stream" can avoid this problem.

Two important methods are the spinning riffler and the free-fall tumbling mixer. The first has been proven to be the most superior method of sampling in experiments and should be used whenever possible. The second has not been investigated fully but deserves further study. Methods that rely on chaotic conditions created in the mixing process should be perfect for sampling. The powder is in motion and all is sampled, following the recommended guidelines. The problem of losing fines into the surrounding atmosphere is also avoided.^[13]

If the FDA chooses to follow the stratified sampling recommendation of the PQRI, blend uniformity testing and the large problem of sampling may soon become history. Until then, however, guidelines should be followed at all times, awareness should be raised to minimize errors, and the importance of sampling should not be underestimated. Testing must be carried out frequently to ensure that each blend reaches the standards set.

ACKNOWLEDGMENT

I would like to thank my Director of Studies, Dr. James Wells, for all his help with my research. I would also like to thank my sponsors Pfizer Global R&D for their support and Marcel Dekker, Inc. for allowing the reproduction of material.

REFERENCES

1. Gy, P. *Sampling for Analytical Purposes*; Wiley: Chichester, 1998.
2. Svarovsky, L. Characterization of Powders. In *Principles of Powder Technology*; Rhodes, M.J., Ed.; John Wiley & Sons: Chichester, 1990; 35–69.
3. Hawkins, A.E. Characterizing the Single Particle. In *Principles of Powder Technology*; Rhodes, M.J., Ed.; John Wiley & Sons: Chichester, 1990; 9–34.
4. Ashton, M.D.; Schofield, C.; Valentin, F.H.H. The use of a light probe for assessing the homogeneity of powder mixtures. *Chem. Eng. Sci.* **1966**, *21*, 843–849.
5. Venables, H.J.; Wells, J.I. *Powder Sampling*; Drug Dev. Ind. Pharm., Marcel Dekker Inc.: New York, 2002; Vol. 28 (2), 107–117.
6. Harnby, N. The Sampling of Particulate Mixtures. In *Powtech '75. Proceedings of the 3rd International Powder Technology and Bulk Solids Conference*; A.S. Goldberg, Heydon & Son Ltd.: London, 1976.
7. Hersey, J.A. Sampling and assessment of powder mixtures for cosmetics and pharmaceuticals. *J. Soc. Cosmetic Chemists* **1970**, *21*, 259–269.
8. Gy, P. Sampling—Diving into the unknown. *LC-GC* **1994**, *12* (11), 808–817.
9. Orr, C. *Particulate Technology*, 1st Ed.; Collier-Macmillan Ltd.: London, 1966; 87–123.
10. Shotton, E.; Ridgway, K.C. *Physical Pharmaceutics*; Clarendon Press: Oxford, 1974; 236–240.
11. Allen, T. *Powder Technology Theory*, 3rd Ed.; Chapman and Hall, 1981.
12. Bicking, C.A. Sampling. In *Encyclopedia of Chemical Technology*, 2nd Ed.; Interscience Publishers: Kirk-Othmer, 1964; Vol. 17, 744–762.
13. Kaye, B.H. Powder Mixing. In *Powder and Powder Mixture Characterization Technology*; Chapman & Hall: London, 1997; 77–95.
14. Chang, R.K.; Shukla, J.; Buehler, J. An evaluation of a unit-dose compacting sample thief and a discussion of content uniformity testing and blending validation issues. *Drug Dev. Ind. Pharm.* **1996**, *22* (9 & 10), 1031–1035.
15. Crosby, N.T.; Patel, I. *General Principles of Good Sampling Practice*; Royal Society of Chemistry: Cambridge, 1995.
16. Kaye, B.H.; Naylor, A.G. Powder Sampling Procedures. In *Particle Technology. Proc., Seminar, Madras, 1971*; Indian Institute of Technology, 1972; 47–66.
17. Allen, T. Sampling and size analysis. *Silica. Ind.* **1964**, *29* (12), 509–515.

18. Khan, A.A. MSc Thesis; Bradford University, 1968.
19. Royle, A.G. Minerals Industry International, 1993.
20. Wells, J.I. *Pharmaceutical Preformulation*; Ellis Horwood: Chichester, 1988; 192–208.
21. *Barr Court Decision*; 1993. <http://www.gmp1st.com/barr2.htm> (accessed August 2002).
22. FDA. 1993. <http://www.fda.gov/bbs/toppics/ENFORCE/ENF00242.html> (accessed August 2002).
23. United States v. Barr Laboratories. District Court of New Jersey. 1993, 812F, Supp. 458.
24. Guentensberger, J.; Lameiro, P.; Nyhuis, A.; O'Connell, B.; Tigner, S. A statistical approach to blend uniformity acceptance criteria. *Drug Dev. Ind. Pharm.* **1996**, 22 (11), 1055–1061.
25. Berman, J.; Schoeneman, A.; Shelton, J. Unit dose sampling: A tale of two thieves. *Drug Dev. Ind. Pharm.* **1996**, 22 (11), 1121–1132.
26. PDA technical report No. 25. Blend uniformity analysis: Validation and in-process testing. *PDA J. Sci. Technol.* **1997**, 51 (53), 1–44.
27. Pastor, R.; Gómez, R.; Ramirez, J.L.; Rodriguez, D.; Mortensen, N.; Bellamy, M.K. A novel sample thief designed to avoid biased data. *Pharm. Tech. Yearbook* **1999**, 47–54.
28. Guidance for Industry. In *ANDAs: Blend Uniformity Analysis Draft Guidance*; FDA Center for Drug Evaluation and Research, 1999; 1–7.
29. PhRMA Statistics Working Group. *Comments on ANDAs: Blend Uniformity Analysis—Draft Guidance*; 1999.
30. PQRI. <http://www.pqri.org> (accessed November 2000).
31. PQRI. *Results of Statistical Analysis of Blend and Dosage Unit Content Uniformity Data Obtained from the PQRI BUWG Data-Mining Effort*; <http://www.pqri.org> (accessed August 2002).
32. FDA Advisory Committee. http://www.fdaadvisorycommittee.com/FDC/AdvisoryCommittee/Committees/Pharmaceutical+Science/050802_ANDA+blend/050802_ANDAblendR.htm (accessed August 2002).
33. Smith, R.; James, R.V. *The Sampling of Bulk Materials*; Royal Society of Chemistry: London, 1981.

POWDERS AS DOSAGE FORMS

Jean-Marc Aiache

Erick Beyssac

Faculty of Pharmacy, Clermont-Ferrand, France

INTRODUCTION

Powders are both the simplest dosage forms and the basis of many other solid dosage forms, such as tablets, capsules, etc. Many drugs or ingredients are also in powder form before processing.

Powders were originally designed as a convenient mode of administering hard vegetable drugs such as roots, barks, and woods; powders were also found to be convenient for dispensing insoluble chemicals such as calomel, bismuth salts, mercury, and chalk.

Presentation in powder form permits drugs to be reduced to a very fine state of division, which often enhances their therapeutic activity or their efficacy by an increase of dissolution rate and/or absorption (1, 3–5). Divided powders are also found to be convenient for administering drugs that are excessively bitter, nauseous, or otherwise offensive to the taste.

This article describes the preparation of powders from bulk materials to finished products according to the administration routes, packaging, and storage conditions.

HISTORICAL BACKGROUND

In all the old books of Pharmacy were included powdered dosage forms resulting from the fine division of animal, vegetable, mineral, and synthetic solid substances (1).

Such powders can be administered singly (simple powders) or as a mixture of different medicinal powders (compound powders) (5). One of the most ancient compound powders was “Hiera Picra” (sacred bitters), a mixture of Aloe and Canella introduced in about 500 B.C. as a laxative (3). (A large number of bitter powders containing Aloe as the principal ingredient bore the name of Hiera.) This powder was listed in various Pharmacopoeias, and it was recognized up to the 4th edition of the *National Formulary* (official until 1926).

Other famous compound powders of the past (1–8) include:

- compound powder of Glycyrrhiza
- compound Senna powder

- Dover’s powders introduced by the English physician Thomas Dover as Pulvis Diaphoreticus in the early Eighteenth century (and until recently in some Pharmacopoeias)
- aromatic powder of chalk (a simplified version of a complex confection devised by Sir Walter Raleigh during his imprisonment) known as The CONFECTIO RALEGHIANA (1721)
- Dr. James’ antimonial powder (fever powder) patented in 1747
- SEIDLITZ powders (a saline cathartic originated and patented by Thomas Savory in 1815). They owed their value to the mineral properties of the Seidlitz springs in Germany (which contain magnesium sulfate).

DRUG POWDER FORMULATION

According to the European Pharmacopeia, powders as dosage forms are made of solid, dry, free, and more or less fine particles. They contain one or more active ingredients with or without excipients and if necessary, coloring or flavoring substances.

So, the formulation of powders as dosage forms is performed according to the following steps:

- obtention of powder as raw material from an original drug (animal, vegetable drugs, or animal or synthetic chemical entities) by different methods of division:
- mixing of various powders with or without excipients as a function of the powders’ characteristics (e.g., flow properties)
- modification of their density (e.g., by granulation) if necessary
- packaging of the finished product for an easy patient’s use

OBTENTION OF POWDERS AS RAW MATERIAL

Introduction

The main process is the mechanical (5) division that reduces lump drugs into fragment of different sizes

(coarse division). To reduce the size, comminution is then used.

Coarse division includes various operations such as cutting, chopping, crushing, grinding, milling, micronizing, and trituration, which depend on the type of equipment used, on the raw material to be treated (vegetable, synthetic, or mineral), and on the convenient particle size. These operations produce heat during processing, which makes it necessary to know certain physical and chemical properties of the drug itself, such as:

- liquefaction temperature
- melting point
- sticking properties
- thermolability
- hardness
- moisture content
- brittleness

Some physical aspects of fracture mechanisms are existing: Crushing or comminution produces small-sized particles that present a new free surface area. This requires energy. With current knowledge, it is not possible to determine accurately the necessary energy for good comminution, according to the particle size before and after treatment and their physical properties (e.g., brittle, elastic, viscous, plastic) (9).

Several fundamental laws have been suggested and they have been described previously (10–13).

Methods and Equipment

First, to decrease risks of contamination and deterioration, the equipment used in pharmaceuticals should be made of stainless steel, be easy to disassemble for cleaning, and be used in a closed room equipped with a dust removal system.

Coarse division

The devices are described in Table 1.

Comminution

Comminution gives particles smaller than that coarse division. The following factors may influence the choice of the appropriate device:

Physical properties of the drug (13)

Hardness: Hard substances must be subjected to compression, impact, and abrasion or milling, but equipment wear is severe. In some instances, mill wear may be so extensive that it leads to highly contaminated products.

Abrasiveness: This is measured on Moh's scale: 1–3 = soft substances, 8–10 = hard substances.

Elasticity

Friability

Friability: Plant products require a cutting or a chopping action and cannot be subjected to pressure or impact techniques (9, 10).

Moisture content: Several Pharmacopoeias recommend drying (moisture content < 5%) the drug substances before comminution (oven at 40–45°C) to avoid liquefaction or agglomeration. Hydrates that may release their water during the process require cooling or low-speed processing.

Initial feed particle size and degree of size reduction desired: Each device has a reduction ratio: it can accept materials of sizes in a certain range and will provide particle of a set size. The particle size distribution of the obtained powder must be as even as possible to obviate further treatment of the larger particles. It is therefore useful to combine, in the same equipment, a milling device with some type of classifier, such as a screen. The oversize particles are returned to the mill on a continuous basis while the particles of the desired size pass through the screen. This kind of equipment is very convenient for sensitive substances or to prevent overmilling and overproduction of fines.

Table 1 Different operations of a coarse division (with definition and devices used)

Operation	Definition	Devices used
Cutting	Obtention of coarse fragments (e.g., vegetable drugs)	Blades
Crushing	Division of coarse fragments	Mortar and pestles (Wedgwood ware, porcelain, glass, iron)
Attrition	Breaking down a substance by rubbing two surfaces together	Non-or stainless steel worn

Table 2 Different operations of comminution (with their devices)

Operation	Laboratory scale devices	Industrial scale devices ^a
Attrition	Mortar and pestles	Jaw crushing device
	Grinding stones	Mills (vertical or horizontal)
	Cutter mills	
	Sieves and screens	
Rolling	Roller mills	Roller mills (smooth or saw-toothed cylinders)
		Attrition mills (diameter 1–2 mm)
		Cutter mills (diameter > 0.2 mm)
		Hammer mills (can be used in liquid
		N ₂ or CO ₂ to decrease the milling temperature
		and to protect the drug product)
		Pebble or ball mills (diameter
		< 2 mm)(planetary pot or jarr
		mills or vibration mills)
		Micronizers or jet mills (fluid-energy mills)
		(diameter 1–10 μm)
		Centrifugal impact pulverizer (diameter 0.05–2 mm)

^aAll these devices have been described (see Ref. 13).

Particle shape desired (13): Different kinds of equipment produce different shapes (Table 2).

Quantity of material to be treated: The selected equipment must have the required capacity. The operation can be continuous or not.

All these operations are achieved by sieving to separate the coarse particles for grinding again if necessary.

It is to be noted that safety must be taken into account:

Noise: All equipment must be installed in a special sound-insulated room.

Toxicity: Due to particle air diffusion, protection of workers is compulsory. Cross-contamination must also be prevented.

Powder Characterization

Particle size

All these data must be introduced in the drug product registration file as well as for the raw material due to their impact on the drug bioavailability as for the powder dosage form as finished product.

A powder is characterized by its particle size, which is of importance in achieving optimum production of medicines (9). First of all, this parameter influences the dissolution rate of the drug in vivo, which in turn influences absorption rate and the onset of therapeutic activity. Particle size is important during the production of solid dosage forms in the manufacture of tablets and capsules. Appropriate equipment controls the mass of drug

and other particles by volumetric filling. Any interference with the uniformity of fill volumes may therefore alter the mass of the drug incorporated into the dosage form (tablets or capsules) and so reduce the content uniformity. Powders with different particle sizes have different flow and packing properties, which alter the volume of powders during each event. To avoid such problems, the particle size of drugs should be defined during formulation and must be as uniform as possible (9).

However, it is not easy to evaluate the particle size of a powder. For a large lump, it is possible to measure it in three dimensions. But if the substance is milled, the resulting particles are irregular with different numbers of faces and it would be difficult or impracticable to determine more than a single dimension (14). For this reason, a solid particle is often considered to approximate to a sphere characterized by a diameter. The measurement is thus based on a hypothetical sphere that represents only an approximation to the true shape of the particle. The dimension is thus referred to as the equivalent diameter of the particle (14).

It is possible to generate more than one sphere that is equivalent to a given irregular particle shape. It is useful to evaluate Feret’s or Martin’s statistical diameter (12), which depends on both the orientation and shape of particle. Feret’s diameter is determined from the mean distance between two parallel tangents to the projected particles perimeter, and the Martin’s diameter is the mean chord length of the projected particle perimeter (Fig. 1). Instead of sphere equivalent diameter, it is possible to use equivalent

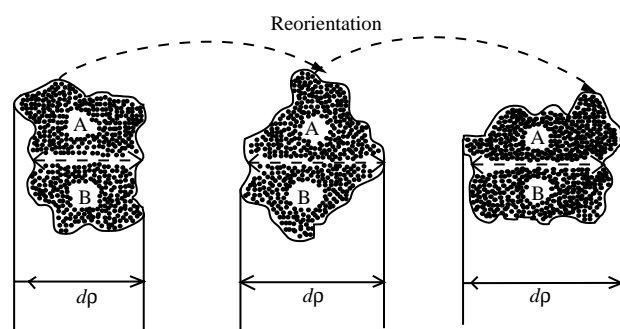


Fig. 1 Influence of particle orientation on statistical diameters. The change in Feret's diameter is shown by the distances, d_F ; Martin's diameter (d_m) corresponds to the dashed lines.

volume, sedimentation volume, mass, or sieve mass of a given particle to determine the particle size of a powder.

Finally, as powders contain particles of different diameters, it is necessary to present the results of size distribution in the form of histogram corresponding to each equivalent diameter evaluated. Such a histogram presents an interpretation of the particle size distribution, enabling the percentage of particles having a given equivalent diameter to be determined and allowing different particles size to be compared (Fig. 2A). An alternative to the histogram representation is obtained by sequentially adding the percent frequency values (Fig. 2B) to produce a cumulative percent frequency distribution (14). Finally, to summarize the data obtained in determining particle size, it is possible to use statistical methods (14).

Recently, some recommendations to include the following items in a particle sizing validation report have been published (15):

- the procedure (or reference to it)
- precision
- range (suitability assessment including microscopic comparison)
- robustness

Detection limit, quantification limit, accuracy, and specificity are not normally considered appropriate for validation of particle sizing methods.

The author of these recommendations added that the validation of particle sizing technique cannot be completed with the first batch of a sample, as data are insufficient to fully validate the procedure at this stage. It is not always important to know the absolute particle size of the first batch manufactured, but it is important to know how the second and subsequent batches compare with it. The validation of a particle size method should be in phases, with parts completed when the first sample is analyzed and

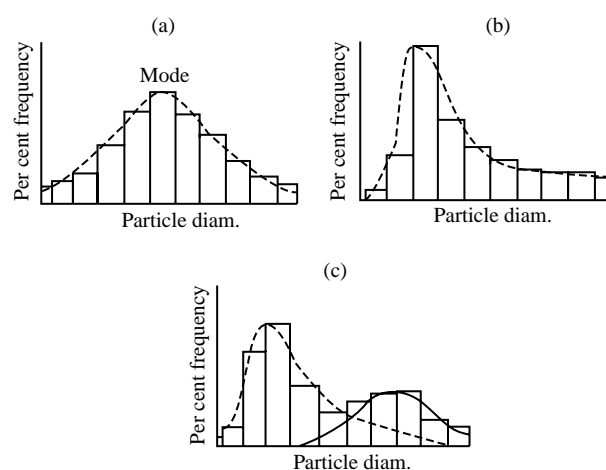


Fig. 2 Frequency distribution curves corresponding to (A) a normal distribution, (B) a positively skewed distribution, and (C) a bimodal distribution.

other data completed or added later, for example, prior to the regulatory filing. Calibration or verification of an instrument being used is assumed to be normal part of GMP and may be used as a suitability assessment (15).

The different methods of particle size analysis can be grouped into several categories: size range analyses (sieve methods), wet or dry methods, and manual or automatic methods (laser-light). All these methods are described elsewhere (16, 17).

Flow properties (18, 19)

Flow properties of powders are important parameters in mixing and segregation phenomena, essentially during storage.

Depending on manufacturing and packaging processes, different flow characteristics are required that are essential for quality control and for the optimization formulation.

In fact a powder, as it has been said, can be made of one or many active powder drugs with or without excipients, and this final mixture will represent the dosage form to be packed and dispensed to patients. So, it is necessary in most of the cases to add excipients for diluting the active powder drug, to improve the packaging properties, and/or to insure a good compliance from the patient (e.g., taste masking). So, the excipients must present the following properties:

- optimal particle size distribution
- high flowability
- high compressibility
- optimal capacity for a drug

- sufficient reworkability
- physiologically inert
- resistant to heat, humidity, and oxidation
- tasteless and odorless

The bulk mixture obtained must present flow properties that are determined by

1. flow through an orifice
2. angle of repose
3. bulk density or bulk tapped density (18)
4. shear cells

These methods are described quite in all the pharmacopeias and in an article elsewhere (20).

OBTENTION OF POWDERS AS DOSAGE FORMS

A powder as dosage form, as it has been said, is made of one or more drug powders, generally, with excipients. So, the mixing of all the components and the validation of this operation are very important.

Mixing of Powders

The aim of mixing is to obtain a homogeneous association of several solid products. Each fraction or dose randomly sampled must contain all the components in the same proportion as in the whole preparation (21). This is achieved because the different components “are treated so as to lie as nearly as possible in contact with a particle of each of the other components” (22).

Mixing has been described as a stochastic process by means of stationary and nonstationary MARKOV chains (10) in which the probabilities of particle movement from place to place in the bed are determined.

To obtain a good mixture, it is necessary to take into account the characteristics of the raw material, the finished mixture, the equipment used, and the operating conditions. The mixing process must be validated (discussed later).

Factors influencing the mixing of powders

Ideally, to avoid segregation, it is useful to mix powders with very close characteristics. Particle size is the most important factor. Normally, all the powders should have the same particle size. It is therefore necessary to mill and sieve them before mixing so that they present the same size. However, every substance has its own properties, and the same size does not mean the same shape and the same other properties.

The density of each component affects the stability of the mixture. The heavier particles have a tendency to fall

to the bottom while the lighter ones rise to the top of the powder bed.

The amount (proportion) of each component is also another factor. Good homogeneity is more difficult to obtain if the proportion of one component is small. It is very important that this small amount is perfectly distributed throughout the whole mixture, for example, when a very active substance is diluted in a large amount of excipient to facilitate dosage form preparation and administration (same problem for coloring). So, it is recommended, first to mix together small quantities of constituents and then continue progressive additions. It is also possible to introduce small amounts of drug substances into the mixture, as a solution in a volatile solvent. After mixing, the solvent is evaporated.

Some mixtures need the addition of an adjuvant; thus, to mix magnesium oxide and charcoal, it is necessary to use alcohol or etheroxide, which are evaporated after mixing.

If these factors are not under control, segregation (demixing) can occur. Powder particles can be separated in a mixing device because their pathways depend on their size and density: this happens in rotating shell mixers. Segregation can also be seen when mixer is emptied, when powder is discharged, or during transport and storage. Due to vibrations, there are free spaces in which denser particles can slip and fall to the container's bottom. This separation is facilitated by the size and shape of the particles, which can slip more or less over each other. For very fine powders, density differences are more important than particle adhesion and friction. No segregation is seen for particles under 40 μm diameter. Three types of segregation have been described:

1. Percolation segregation is the movement of the small particles through the voids in the static powder bed.
2. Trajectory segregation occurs during mixing: particles are set in motion and kinetic energy is imparted to them. Larger particles have larger energies and tend to move a greater distance into a powder mass before they are brought to rest. This may result in preferential separation and can occur in both horizontal and vertical planes.
3. Densification segregation: density differences between particles may cause segregation. Large particles move upward through the mass. It seems that the small particles beneath dense large particles are slightly compacted, supporting the large particles. On vibration, the small particles move beneath the large particles and raise them upward in the bed.

The effect of mixing time on segregation is that nonsegregating mixtures continue to improve with an

extended mixing time, but the reverse may be true for segregating mixes.

The factors promoting segregation require a longer time to establish a segregation mix that is needed to produce a reasonable degree of mixing. It is therefore counterproductive to prolong the mixing time beyond an optimum point.

Equipment

The three primary mechanisms responsible for mixing are (10) convective movement of relatively large portions of the bed, shear failure that primarily reduces the scale of segregation, and diffusive movement of individual particles.

Most efficient mixers operate by all three mechanisms. Thus, mixing can be considered a random shuffling-type operation, involving both large and small particle groups and even individual particles.

The ideal mixer should rapidly produce a complete blend with as gentle as possible a mixing action to avoid damage. It should be easily cleaned and discharged, be dust-tight, and require low maintenance and low power consumption.

Laboratory equipment: The pharmacist most often employs the mortar and pestle for the small-scale mixing usually required for prescription compounding. However, spatulas and sieves may also be used occasionally. The mortar and pestle method combines comminution and mixing in a single operation.

Industrial equipment: Rotating-shell mixers or tumbling mixers are shown in Fig. 3. Drum-type, cubical-shaped, double-cone, and twin-shell blenders with their axis of rotation horizontal to the center of the drum on its axis increase crossflow and improve the mixing action. Cubical and polyhedron-shaped blenders with a rotating axis set at various angles are also available. Double-cone blenders or tumbling mixers have been developed; the Y-cone mixer (Fig. 3) is a good example

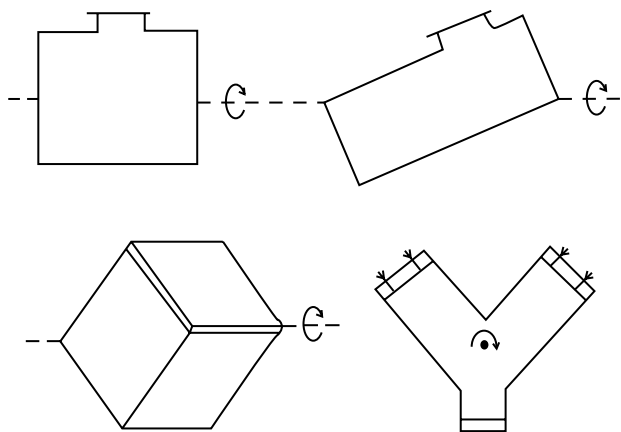


Fig. 3 Tumbling mixers.

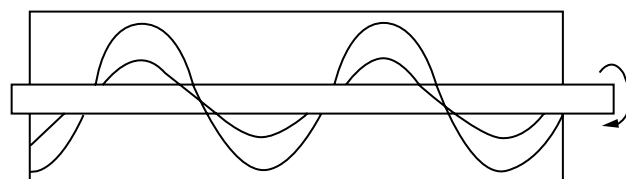


Fig. 4 Ribbon mixer.

(13). On rotation the charge flows into the top two arms of the Y and then back into the third arm. Mixing by shear and diffusion takes place when the streams mingle. Time must be allowed for the mix to flow into the arms, and there is an optimum speed rotation. The zigzag blender is an extension of the twin-shell blender.

A fixed-shell mixer or agitator mixer is shown in Fig. 4. Also called ribbon mixer, it consists of a relatively long throughlike shell with a semicircular bottom. The shell is fitted with a shaft on which are mounted spiral ribbons, paddles, or helical screws alone or in combination. These mixing blades produce a continuous cutting and shuffling of the charge by circulation the powder from end to end of the trough as well as rotationally. The shearing action that develops between the moving blade and the trough serves to break down powder agglomerates. A more recent type of agitator mixer is the Nautamixer (Fig. 5) that consists of a conical vessel fitted at the base with a rotating screw fastened to the end of a rotating arm at the top. The screw conveys material toward the top where it cascades back into the mass.

Sigma-blade and planetary-paddle mixers are used for solid blending and as a step prior to the introduction of liquids (Figs. 6 and 7). This type of mixer rapidly breaks down agglomerates.

Vertical impeller mixers employ a screw-type impeller, which constantly overturns the batch.

Motionless mixers are part of continuous-processing devices with no moving parts. They consist of a series of fixed flow-twisting or flow-splitting elements.

Validation of the Mixing Process

The mixing process is validated by the verification of the homogeneity of mixed powders, for instance, by taking three samples at three different levels of the mixer (top, middle, and bottom) and determining the amount of different active substances in these samples. Each sample must be of the same size as in the kind of dosage form to be used. A mixing index can also be used.

Other characteristics of the mixed powder such as particle size, flavor, flow rate, and bulk density before and after tapping have to be verified.

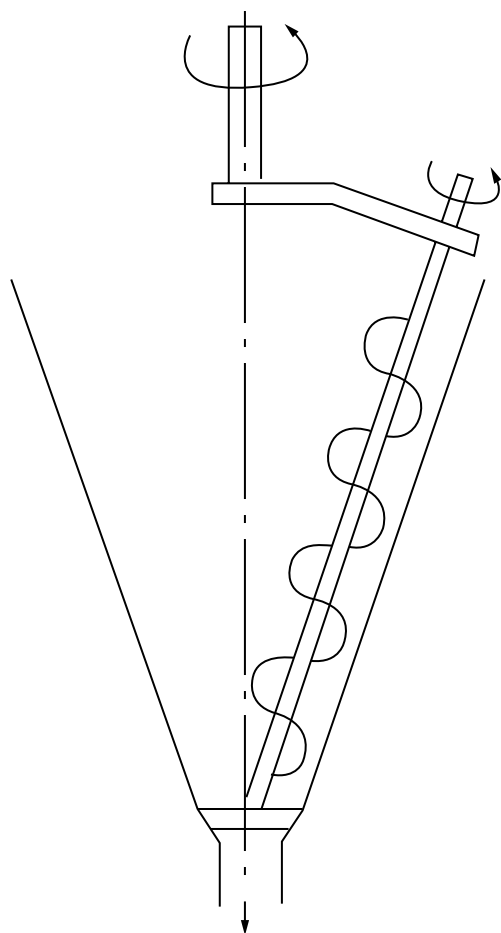


Fig. 5 Nautamixer.

The process reproducibility is assessed by performing operation and verification three times. The validation result is considered correct if the amount of each active substance found in each sample and all the other characteristics evaluated are within the set limits (22).

CLASSIFICATION AND EXAMPLES OF POWDERS AS DOSAGE FORMS

They can be classified as a function of their route of administration.

Oral Administration

Mode of administration and packaging

They are generally administered in or with water or with other appropriate liquids. In a few cases, they can be

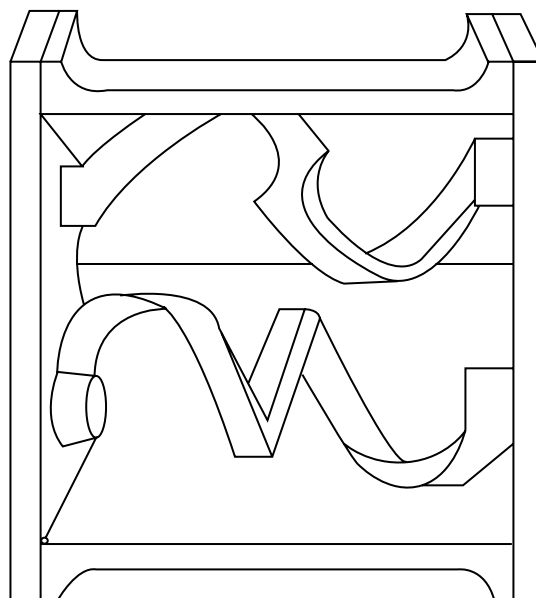


Fig. 6 Sigma-blade mixer.

swallowed as they are. Both single and multiple doses are available.

Multiple-dose powders, presented in band or metal box, require a measure to deliver the prescribed dose. Initially, it was an ordinary spoon, but given the different sizes and shapes of the spoons used worldwide, powder density, humidity, degree of settling, fluffiness due to agitation, and personal judgment of patients (selecting a full or half spoon according to be patient's mood), it has been decided to use other devices. They are generally special measuring plastic spoons on which the volume is indicated and corresponds to a weight of powder as a function of its density (this volume corresponds to an adult or child dosage).

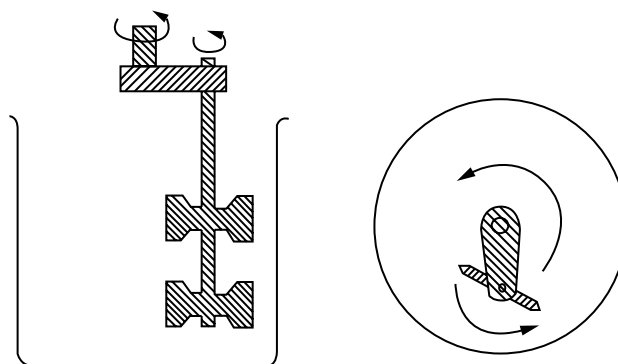


Fig. 7 Planetary-paddle mixer.

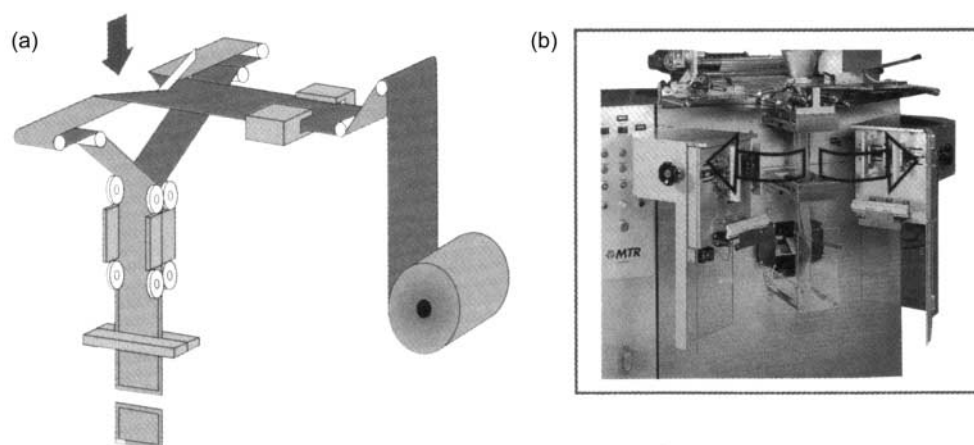


Fig. 8 (a) Foil preparation. (b) Multilane vertical sachet-filling machine. The hinged panels can be opened by 90° to allow complete access to the product dosing area, the film needed, and the sealing and cutting station.

Currently, the best way is to use single dose contained in a folded paper (in community pharmacy) or sachets filled with the same accuracy as a tablet or capsule by a fully automatic machine. The sachet is made of paper, aluminum, and/or complex mixtures that are a combination of aluminum and plastic substances. An aluminum or paper foil is covered on one or both sides (internal and/or external) with polyethylene, PVC, or PVDC (e.g., paper + polyethylene + aluminum + polyethylene). Generally, the foils are presented in rolls, printed on their outer face with the drug name, and assembled in automatic machines (Fig. 8) that use one or two rolls on which sachets are predrawn. First, the foils are cut longitudinally in a by-pass station and folded longitudinally. These folded paper webs are drawn off a roll, crosscut, and guided into presealed sachets. Powders are added by various devices (Fig. 9) such as

micropiston, vacuum feeding, volumetric drawer slide filler, dosing plates (or telescopic cups), suction devices, and discharge and filling funnels for non-free-flowing powders. After filling, the sachets are sealed (three or four side sealed sachets) at 100–145°C for 0.5 s. During filling, the weight must be controlled. After filling, the sachets are submitted to different controls such as sealing, integrity of foils, opening ease as well as permeability to gases, and to humidity. These controls are generally performed on the raw foils and on the finished products.

Excipients and/or modification of powder density or flow characteristics

To obtain a good filling, the powders must be homogeneous during the whole filling time and must flow regularly to obtain similar volume. Homogeneity is

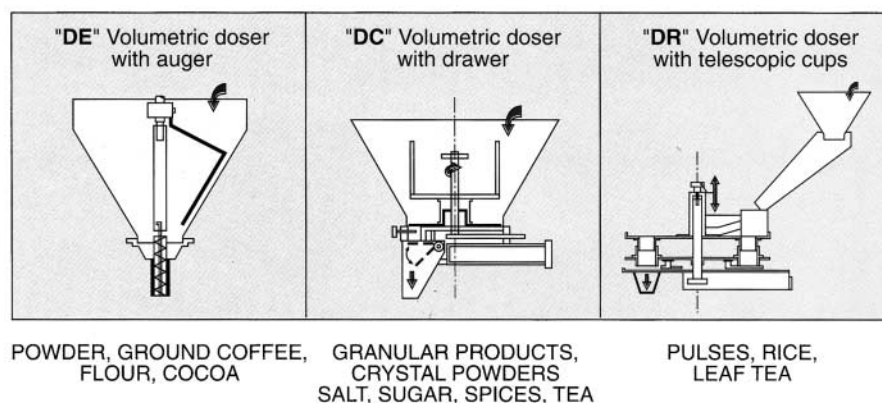


Fig. 9 Different types of volumetric dosers.

obtained by the help of the mixing devices described previously. The mixers are connected to the filling machine. If the duration of mixing is increased, it is absolutely necessary to prevent uncontrolled demixing during this last step.

For the second point, it is necessary to know the flow properties of the powders, and generally it is compulsory to add some excipients to modify these properties and adapt them to the filling device.

The excipients used are fillers, but the most important are lubricants, such as magnesium stearate, PEG 6000, calcium stearate, glycerol palmito stearate, and calcium behenate, which increase the flowability but at the same time modify the hydrophobic characteristics of powders. Their choice depends on compatibility with active ingredient and the flow properties of powders. The characteristics of the excipient have been described previously.

Sometimes, it is necessary to modify the density of the powder by granulation made by the ordinary way (23) as for tablet preparation or using new techniques such as rotogranulation (24). This transformation is sometimes quite necessary because it is easier to modify the taste using flavoring substances or flavor modifiers (25). In this last case, a recent article describes the taste masking as a consequence of the organization of powder mixture (26).

Examples of powders as dosage forms

Effervescent powders: Initially these powders, described in the Pharmacopeia, consisted of sodium bicarbonate, tartaric acid, and potassium and sodium tartrate (wrapped in white paper). Each was separately dissolved in water and then mixed before drinking: they were called “English Seidlitz powders” (the Seltz or the English Soda powders are quite similar) (2).

Nowadays, effervescent powders are available in single or multiple units that contain acid substances and carbonates or bicarbonates, which quickly react in water by releasing carbon dioxide. They are dissolved or dispersed in water before being taken. This is a great advantage because the drug is then in solution, the pH of which is close to 7, which allows rapid passage through the pylorus. The drug absorption from the gut wall as well as the onset of therapeutic activity may thereby be hastened. Effervescent powders contain (27):

1. *Acid materials*

Acids: citric (monohydrate or anhydrate), tartaric, ascorbic (drug or excipient), fumaric, nicotinic, acetylsalicylic (as drug or excipient), malic, and adipic acids (seldom used).

Anhydrides: glutaric, citric.

Salts: sodium dihydrogeno-citrate, sodium acid phosphate, sodium fumarate.

2. *Sources of carbon dioxide*

Salts: sodium bicarbonate (the most widely used), sodium carbonate, potassium carbonate, calcium carbonate, sodium glycine carbonate (28).

3. *Other excipients* (main characteristic, solubility in water):

Lubricants: PEG 6000 is most frequently used, alone or with sodium stearyl fumarate, sodium benzoate, sodium chloride, sodium acetate, or D,L-leucine.

Binders: PVP, hydrogenated maltodextrin, maltodextrins, PEG 6000.

Others: sweeteners, flavors, colors, surfactants, anti-foaming agents (polydimethylsiloxane).

The manufacturing process requires a very low and controlled relative humidity (about 20% or less, if necessary) as well as a controlled ambient temperature close to 25°C. Generally, effervescent powders are made by classical wet granulation of the acid and carbonate part separately or mixed with water–ethanol or isopropanol. Special care must be taken in the case of total granulation to limit the effervescent reaction, except in case of use of the technique of granulation with a rotor (29).

The stability and shelf-life of effervescent powders are essentially assured by protection against external moisture. Single-dose sachets are manufactured with suitable complexes to prevent premature reaction caused by external moisture or excess internal moisture from the granulation process. Special stability studies under different conditions are required for these powders. The dissolution time must be less than 3 min; this is an index of stability.

Many drugs and drug compositions are used in effervescent products including aspirin, acetaminophen, ibuprofen, nonsteroid-antiinflammatory drugs, cimetidine, antibiotics, mucolytic drugs, vitamins, and others. It has been demonstrated many times that drugs are rapidly, and sometimes better, absorbed in this form than in conventional dosage forms (30, 31).

Powders for Parenteral Use

For parenteral use, solid sterile substances are distributed in their final packages. A clear solution nearly free of particles or a uniform one is obtained after shaking with the prescribed volume of an appropriate sterile liquid. Freeze-dried substances for parenteral use are also used.

After dissolution or dispersion, preparations must comply with assay requirements for injectable preparations or injectable preparations for perfusion. Their preparation requires the same care as parenteral solutions, i.e. sterilization of raw material or finished product sterilization (32).

Powders for Cutaneous Applications

These are single- or multiple-unit powders free of agglomerated particles. They must be sterile for application on open wounds or damaged skin.

Multiple-unit powders for local application are preferably packaged in a dredger or a pressurized container (for skin, teeth, or vaginal douche use). These preparations consist of a dispersion of a solid phase (drug) in a liquid propellant (liquid phase). By action on the actuator, the suspension is released by gas pressure. The propellant in contact with ambient air is evaporated and the powder remains on the treated area. The particle size obtained depends on the powder particle size before the preparation of the suspension.

Many powders are presented in this special dosage form (e.g., antiseptics, antifungal drugs) as well as cosmetic preparations (dry shampoo, etc.). Generally, the drug is mixed with an inert diluent such as starch, magnesium carbonate, magnesium stearate, or silicon dioxide of particle size close to 50–60 μm . To increase the suspension stability and to prevent closure of the actuator,

lubricants and/or surfactants are added (mineral oils, isopropyl myristate, lecithin, sorbitane fatty acid esters). These last substances will also prevent agglomeration or crystallization of drug substances (percentage about 0.1–1%).

The propellants used were some years ago chlorofluorocarbons (CFCs) (11 and 12) in various proportions (65:35 or 50:50). The manufacturing process includes the suspension formation in liquid propellant 11 in a cooled and hermetically closed mixer containing drug and lubricants. This mixture is introduced in the cans that are closed by a special powder valve, and propellant 12 was then injected through the valve. The upper part of the valve has often a special aperture through which a gaseous phase cleans the valve as well as the actuator.

A major current issue is the suppression of the use of CFCs due to ozone depletion. Other liquid propellants (134a, 227, or dimethyl ether) are not used as easily as the CFCs and formulation is very difficult. So, many of these dosage forms disappeared or were substituted by “saltcellar” systems. The use of very volatile solvents, such as volatile silicone, can be proposed if their boiling point is close to room temperature (33), the use of butane is not recommended due to the explosion risks.

Powders for Pulmonary Application

As a consequence of the suppression or reduced use of CFC propellants, a new kind of dosage form is under

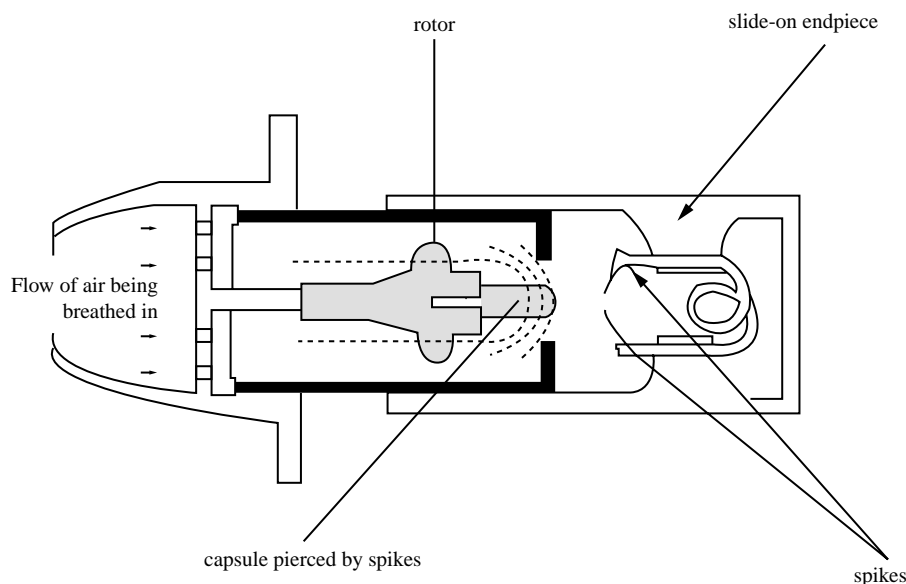


Fig. 10 Spinhaler.

worldwide development: drug powder inhalers (DPIs). The metered dose inhalers (MDIs) will probably be replaced by these. The MDIs were formulated several years ago. A drug in powder form (with a particle size close to $5\text{ }\mu\text{m}$) was suspended in propellant 11 with a surfactant. Propellant 12 was added to the can after closure.

The metered valve delivers a fixed volume of suspension that after propellant evaporation, releases the active drug in the upper respiratory tract. Owing to its small diameter size, about 10% of the drug reaches the bronchopulmonary tract and provides therapeutic activity.

Currently, propellant issues have led companies to develop DPI. These consist of a very fine active drug (particle size $5\text{ }\mu\text{m}$ or less) mixed with an inert excipient such as lactose. The mixture is administered through a special device.

Spinhaler (Fisons) (Fig. 10) and Rotahaler (Allen and Hanbury) (Fig. 11) were introduced more than 10 years ago for delivery of single metered doses of sodium cromoglycate, salbutamol, or beclomethasone dipropionate in powder form, with the energy required to disperse the powder being derived from the patient's own inspiratory effort. Both the Spinhaler and Rotahaler (Ciba systems) are rather inconvenient to use because it is necessary to load a gelatin capsule containing the drug powder into the device immediately prior to use. However, a lot of new multidose dry powder inhalers have been recently introduced. The Diskhaler (Allen and Hanbury) contains eight doses of either salbutamol or beclomethasone dipropionate (2-day supply at normal dosing levels) in blisters around the periphery of a small disk. The Diskus is more interesting: it contains 60 doses of antiasthmatic drug

with a special counter, which indicates the number of delivered doses. The Turbuhaler (Astra) (Fig. 12) contains 200 metered doses of the bronchodilator terbutaline sulfate in the manner of a pressurized MDI, but without additives of any kind. The corticosteroid budesonide is available in a Turbuhaler in some countries.

The clickhaler is based on the same principle: a reservoir containing a number of doses releases one after an actuation.

It seems that all these devices are easy to use and are readily accepted by patients. According to some authors, they deliver a similar percentage of the drug dose to the lungs as a correctly used MDI and has an equivalent efficacy (34).

However, patients with low inspiratory flow rate (essentially children) have many difficulties to use them. Furthermore, when the particles are inhaled quickly, their inertia increases and there is an impaction in the throat. So, new devices are developed to improve the patient's compliance. They consist of a pump that compresses air in a special chamber so that the powder is expelled as a cloud, owing to the pressure delivered by the compressed air (35). These systems are close to the old MDI.

CONCLUSION

The preparation of powders as dosage forms comprises many steps that are the same as for the preparation of powders used to manufacture other solid dosage forms such as tablets or hard gelatin capsules.

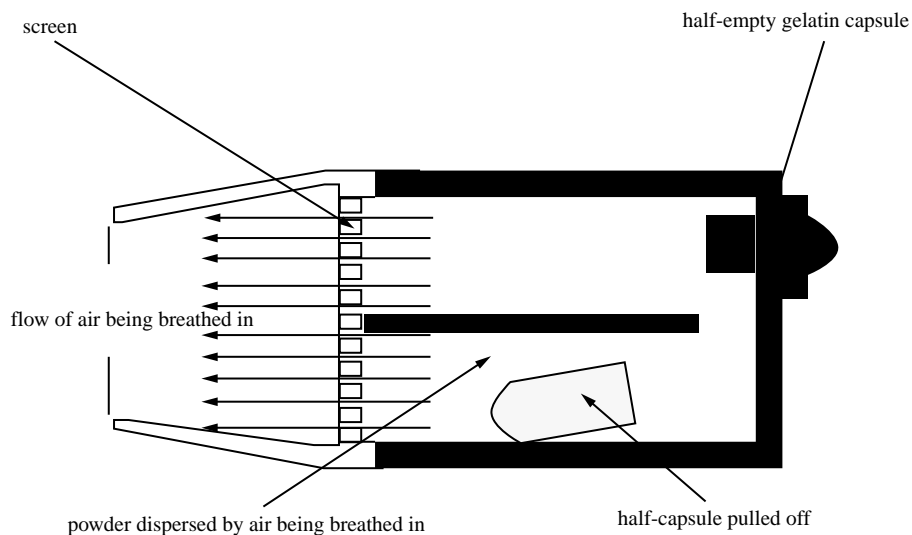


Fig. 11 Rotahaler.

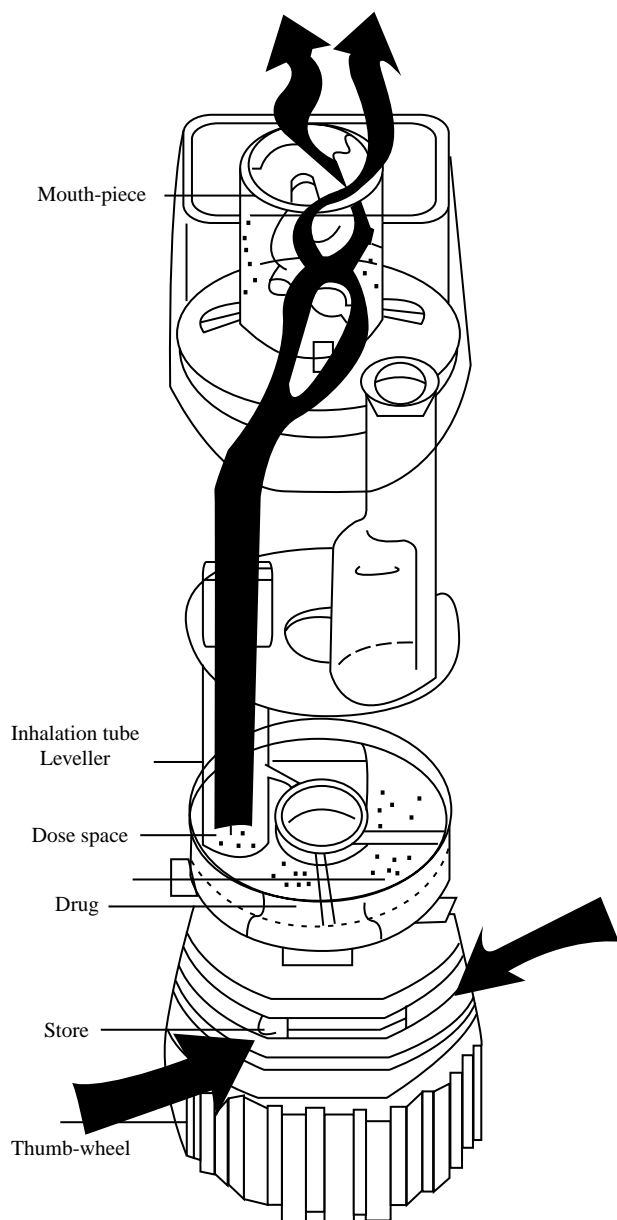


Fig. 12 Turbuhaler.

Years ago powders were used as dosage forms more frequently before the advent of more modern forms. However, biopharmaceutical problems arose with these new drug dosage forms, and companies have returned to powders and developed these further. Industrial sachet manufacturing considerably reduces handling and improves storage. Pressurized systems allow easy dispensation of disinfectants, which increases patients' compliance. It can be expected that powders will continue to be used in the future.

REFERENCES

1. Astruc, A. *Traité de Pharmacie Galénique*, 4th Ed.; Librairie Maloine: Paris, 1946.
2. Buerki, R.A.; Higby, G.J. History of Dosage Forms and Basic Preparations. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1993; 7, 299–338.
3. *Codex Medicamentarius, Pharmacopée Française*; Bechet Jeune: Paris, 1837.
4. *Codex Medicamentarius, Pharmacopée Française*; J.B. Bailliére & Fils: Paris, 1866.
5. Henry, N.E.; Guibourt, C.B. *Traité de Pharmacie*; J.B. Bailliére: Paris, 1847.
6. *Palestra Farmaceutica Chimico Galenica*; Madrid, 1706.
7. *Oficina Medicamentorum Valentiae MDCI*; Valence, 1601.
8. *Farmacopea Cathalana Sive Antidotarium Barcinonense*; Barcelona, 1686.
9. Le Hir, A. *Pharmacie Galénique*, 6th Ed.; Masson: Paris, 1992.
10. Ripple, E.G. Powders. *Remington's Pharmaceutical Sciences*; Gennaro, A.R., Ed.; Mack Pub. Co.: Easton, PA, 1980.
11. Parrot, E.L.; Milling, I.N.; Lachman, L.; Lieberman, H.A.; Kanig, J.L. *The Theory and Practice of Industrial Pharmacy*; Milling, I.N., Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea & Febiger: Philadelphia, 1970; 100–119.
12. Perry, R.H. *Chemical Engineers' Handbook*, 4th Ed.; McGraw-Hill: New York, 1963.
13. Aiache, J.-M.; Beyssac, E. Powders as Dosage Forms. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1994; 12, 389–419.
14. Staniforth, J.N. Particle Size Analysis. *Pharmaceutics: The Science of Dosage Form Design*; Aulton, M.E., Ed.; Churchill Livingstone: Edinburgh, UK, 1988.
15. Bell, R.; Dennis, A.; Hendriksen, B.; North, N.; Sherwood, J. Position Paper on Particle Sizing: Sample Preparation, Method Validation and Data Presentation. *Pharm. Technol. Eur.* **1999**, 11 (11), 36–42.
16. *Pharmeuropa* **1993**, 5, 89–91.
17. Staniforth, J.N. Particle Size Analysis. *Pharmaceutics: The Science of Dosage Form Design*; Aulton, M.E., Ed.; Churchill Livingstone: Edinburgh, UK, 1988.
18. Neumann, B.S. The Flow Properties of Powders. *Advances in Pharmaceutical Sciences*; Bean, H.S., Carless, J.E., Beckett, A.H., Eds.; Academic Press: London, UK, 1967; 2, 181–221.
19. Carr, R.L. *Chem. Eng.* **1965**, 7, 163–167.
20. Neumann, B.S. The Flow Properties of Powders. *Advances in Pharmaceutical Sciences*; Bean, H.S., Carless, J.E., Beckett, A.H., Eds.; Academic Press: London, 1967; 2, 181–221.
21. *Génie Pharmaceutique. Galenica, Technique et Documentation*; Paris, 1982; 3.
22. Travers, D.N. Mixing. *Pharmaceutics: The Science of Dosage Form Design*; Churchill Livingstone: Edinburgh, UK, 1988.

23. Kristensen, H.G. Granulations. *Encyclopedia of Pharmaceutical Technology*; 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1993; 109–120.
24. *European Pharmacopeia*, 1999.
25. Adjei, A.L.; Doyle, R.; Reiland, T. Flavors and Flavor Modifiers. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1992; 6, 101–139.
26. Barra, J.; Lescure, F.; Doelker, E. Taste Masking as a Consequence of the Organisation of Powder Mixes. *Pharm. Acta Helv.* **1999**, *74*, 37–42.
27. Lindberg, N.O.; Engfors, H.; Ericsson, T. Effervescent Pharmaceuticals. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1992; 5, 45–71.
28. Aiache, J.-M. Les Comprimés Effervescents. *Pharm. Acta Helv.* **1974**, *49*, 169.
29. Gauthier, P.; Aiache, J.-M. Rotary Fluidized Bed Process and Its Possibilities. *Symposium Glatt*; Marseille: Octobre 1997.
30. Febvre, P.; Aiache, J.-M.; Bex, J.F. Etude En Continu In Vivo De La Salicylémie Du Lapin Après Ingestion D'aspirine Soluble Sous Différentes Formes Galéniques. *C.R. Soc. Biol.* **1969**, *163*, 1160.
31. Vatie, J.; Slama, A.; Vitre, M.T.; Mignon, M. Ranitidine 300 Mg Effervescente. Activité Antiacide Et Pouvoir Tampon Liés Au Complexe Effervescent. *Med. Chir. Dig.* **1992**, *21* (28).
32. Turco, S.J. Dosage Forms: Parenteral. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1991; 4, 231–247.
33. Beyssac, E.; Aiache, J.-M.; Grumel, J.-M. *Modification of the Impaction Device to Assess Particle Size Distribution of Nasal Sprays*, 3th European Congress of Pharmaceutical Sciences; Edinburgh, UK, September 1996.
34. Newman, S.P. Metered-Dose Pressurized Aerosols and the Ozone Layer. *Eur. Resp. J.* **1990**, *3*, 495–497.
35. Aiache, J.-M. The Ideal Drug Delivery System: A Look into the Future. *J. Aer. Med.* **1991**, *4*, 323–334.

PRESERVATION OF PHARMACEUTICAL PRODUCTS

Peter Gilbert

David G. Allison

University of Manchester, Manchester, United Kingdom

PRESERVATION

Microbial spoilage of pharmaceutical products has been known for many years. Spoilage may result in the deterioration of the product due to loss of potency or to the initiation of an infection in the user. Sterile pharmaceutical products (single dose or multiple dose forms) require the addition of an antimicrobial preservative when they have been manufactured under aseptic conditions from presterilized ingredients. Where the products are subject to a terminal sterilization process, only the multiple dose category requires the addition of an antimicrobial agent. In the latter instance, the preservative is added to protect the product and end user against the consequences of microbial entry during use. Chemical antimicrobial agents are thus added to all multidose sterile formulations and to aqueous and aqueous-based nonsterile pharmaceuticals. Their function is to reduce the microbial load to a level, which is safe for the designated use of the product, and to maintain the numbers of viable microorganisms at or below that value for the storage and use life of the product. Preservatives must, therefore, be stable within the formulation for the shelf life of the product and be capable of dealing with all the abuses made to it by the consumer and user (i.e. contamination during use, incorrect storage *etc*). Table 1 presents, in alphabetical order by chemical grouping, the agents most often employed for preservation of pharmaceutical products.

For multidose sterile products, the preservative must be capable of reestablishing sterility between each use, whereas for a nonsterile topical cosmetic the function of the preservative might simply be to prevent growth. The associated toxicity of preservatives often limits the concentrations at which they can be employed; thus, lower concentrations are generally employed for ophthalmic products and injectables. In choosing a preservative the likely capacity required, the rate of killing desired, and the ingredients and pH of the formulation must be borne in mind.

The BP (1988) test for the "Efficacy of Preservatives in Pharmaceuticals" makes it quite clear that a preservative

must in the first instance *kill* microorganisms and only accepts later that it will prevent growth. However, it does insist that there must be no outgrowth of the test organism, that is, the preservative possesses sufficient residual capacity to inhibit the growth of any survivors.

DEFINITION OF TERMS

Disinfectants, antiseptics, and preservatives are chemicals that have the ability to destroy or inhibit the growth of microorganisms, and are used for this purpose. These and other terms commonly employed are defined as follows:

- *Disinfectants*: Chemical agents or formulations that are too irritant or toxic on body surfaces, but are used to reduce the level of microorganisms from the surface of inanimate objects to one that is safe for a defined purpose.
- *Antiseptics*: Chemical agents or formulations that can be used as an antimicrobial agents on body surfaces.
- *Preservatives*: Chemical agents or formulations that are capable of reducing the number of viable microorganisms within an object or field to a level that is safe for its designated use and will maintain the numbers of viable microorganisms at or below a level for the use/shelf-life of the product.
- *Bacteriostasis*: A state in which the growth of microorganisms is halted or inhibited.
- *Bactericide*: A chemical antimicrobial agent that reduces the viability of a population of microorganisms exposed to it. This term is meaningless without specifying the concentration range over which this effect is obtained; such concentration ranges will vary between different species of microorganisms.
- *Bacteriostat*: A chemical antimicrobial agent that can prevent the growth of microorganisms within an otherwise nutritious environment. This term is meaningless without specifying the concentration at which this effect is achieved. Bacteriostatic concentrations do vary between different species of microorganisms.

Table 1 General properties of some widely used preservatives

Preservative	Advantages	Disadvantages	Application
<i>Acids (organic):</i> benzoic acid, parabens, sorbic acid	Active against bacteria and fungi	Highly pH dependent	Oral and topical formulations; gums and syrups
<i>Alcohols:</i> ethyl or isopropyl, chlorbutol, bronopol	Broad spectrum, including that against acid-fast bacteria	Volatile; poor penetration of organic matter	Solvent; eye drops and injections; synergistic properties
<i>Aldehydes:</i> formaldehyde, gluteraldehyde	Broad spectrum antibacterial, antifungal, and sporicidal activity	Acid solutions inactivated with temperature; toxic and carcinogenic	Chemical sterilization and storage of surgical instruments (e.g., endoscopes)
<i>Biguanides:</i> chlorhexidine, polyhexamethylene biguanide	Mainly active in cationic form against gram-positive bacteria	Water insoluble; inactivated by organic matter; limited antifungal activity	Solution for hard contact lenses and other ophthalmic products
<i>Halogens:</i> hypo-chlorite, povidone-iodine, chloroform	Broad spectrum of antibacterial, fungal, and viral activity	Unstable; corrosive; inactivated by organic matter	Limited use nowadays
<i>Organic mercurials:</i> mercury, silver, thiomersal, phenylmercuric acetate	Broad spectrum of antibacterial activity	Low capacity to organic matter, ionic and some nonionic surfactants; toxicity	Eye drops; contact lens solutions
<i>Phenolics:</i> cresol, chlorocresol, bisphenol	Cheap, rapid activity against gram-positive bacteria and fungi	Low water solubility; adsorbed by rubber; volatile, irritant, pH dependent	Creams
<i>Quaternary ammonium compounds:</i> cetrimide, benzalkonium chloride	Narrow spectrum (gram-positive bacteria) of activity; surfactant properties	Low capacity to organic matter; low activity at acidic pH; incompatible with soaps, ionic and nonionic surfactants	Eye drops; surgical creams; ointments

It should be noted that terms such as bactericide and bacteriostat should be discouraged; in the USP and EP, the term "antimicrobial agent" has replaced these terms.

PRESERVATIVE IDEALS

At present there is no perfect preservative, and all materials are a compromise of a number of often contrary properties. The following are the properties of an ideal preservative compound and need to be considered when choosing a preservative.

1. *Definable in chemical terms*: Many of the existing preservatives, such as the quaternary ammonium compounds, are mixtures of various homologues. Often the activity obtained is a function of the mixture composition. Unless it is possible to define and control mixture composition, the performance of the agents will be variable, even if they conform to a pharmacopoeial specification.
2. *Broad spectrum of activity*: The compounds must possess a broad spectrum of antimicrobial activity against all species of microorganisms and also toward bacterial endospores. In practice, the only compounds that meet this requirement are formaldehyde, glutaraldehyde, hypochlorite, and ethylene oxide. All these compounds are highly irritant at sterilizing concentrations to be used in pharmaceutical products. Formaldehyde is, however, used at low concentrations in some shampoos; in these cases contact with the skin is short-lived and irritancy minimal. Agents such as quaternary ammonium compounds, phenolics, and the parabens group possess good activity against gram-positive bacteria but little or no activity toward spores. Certain gram-negative organisms such as *Pseudomonas aeruginosa* are virtually resistant to these agents. Generally, antifungal activity is difficult to obtain. Combinations of preservatives are sometimes employed to widen the spectrum of activity to include molds, bacteria, yeasts, and endospores.
3. *Effectiveness*: The compounds must be effective over a wide range of pH in order to be effective in all formulations. In practice, compounds are generally more active at either acid or alkaline pH. Thus, the pH of a formulation determines the types of preservative suitable for inclusion.
4. *Stability*: The compounds must be stable to light and elevated temperatures for the expected shelf life of the product. The effects of pH upon stability should be minimal. In this respect it is worth noting that the

preservative Bronopol is stable only in the dark and at an acid pH. Under alkaline conditions or in the light it rapidly decomposes to give formaldehyde at concentrations that would be ineffective as a preservative. Instability to light can be protected against the packaging in a light-proof container. Storage tests must be performed on all formulations to ensure that adequate levels of preservative remain at the end of then expected shelf life of the product.

5. *Solubility*: Preservatives should ideally be used at concentrations much lower than that of the main constituents of the formulation. Their solubility ought to be such that it is possible to add them as a concentrated solution and where there is no danger of creating a saturated solution.
6. *Aesthetics*: Preservatives should have no perceptible odor, color, or taste, which might affect the aesthetic qualities of the final product. This can be of crucial importance for a cosmetic product but is less important for medical ones.
7. *Volatility*: Preservatives should be nonvolatile. Thus, chloroform is not an ideal preservative as it is lost from the formulation each time it is exposed to air.
8. *Product incompatibility*: Preservatives should not be incompatible with any of the likely excipients within the product formulation. This would include incompatibilities with the container material and also the active ingredients. In practice this is very difficult to achieve.
9. *Toxicity*: At the concentrations employed, the preservative should be nonirritant, not cause hypersensitivity reactions, and be nontoxic. In this respect, the site of application is critical. Relatively few compounds are approved for use in ophthalmic products due to their high sensitivity towards xenobiotics. Also, compounds safe for use on intact skin might be hazardous for inclusion in parenteral products.
10. *Solubility in oil*: Preservatives must not be too oil soluble as this can produce problems in two- and three-phase systems where the preservative accumulates in the oil and micellar phases and is unavailable for antimicrobial action in the biological (aqueous) phase. It is worth noting that the oil:water partition coefficient can alter as a function of pH and also as a function of the nature of the oil.
11. *Cost*: The preservative must be cost-effective in the context of the overall product positioning (i.e., cost of goods ratio).

Although a very large number of antimicrobial compounds have been examined for their suitability for preservation of pharmaceuticals, only a few (<20) are

currently used in the majority of pharmaceutical products. Moreover, in many cases, the “least unsuitable” rather than the “optimum” preservative is selected for a particular product.

DYNAMICS OF PRESERVATION

The critical lethal parameters of the effectiveness of antimicrobial agents are the concentrations of the agents employed, the time of exposure, the types and numbers of organisms exposed, the pH of the environment or formulation, and the temperature of application. An understanding of the effects of these conditions upon the killing process and of the terms that describe such dependence is important to rationalize the use of preservatives in different situations and formulations.

Time of Contact

When exposed to lethal concentrations of a chemical antimicrobial agent, the viability of a bacterial population decreases exponentially with time (Fig. 1). Such pseudo-first-order kinetics can be described by the equation

$$\frac{S}{S_0} = Ae^{-kt} \quad (1)$$

$$\log \frac{S}{S_0} = kt + \log A \quad (2)$$

where k is the rate constant, S is the number of surviving organisms at time t , and S_0 is the initial number of viable organisms within the population.

D -value is the time for 90% kill. That is,

$$\log S - \log S_0 = -1 \quad (3)$$

and

$$D\text{-value} = 1/k \quad (4)$$

In practice, such obedience to first-order kinetics is rare for chemical inactivation. More commonly, deviations from first-order kinetics are obtained, when there is either an initial lag in the rate of killing (Fig. 2A) or when the rate of killing decreases with time of exposure (Fig. 2B). The former is commonly observed when the concentrations of preservative are minimally bactericidal and reflects the time required for injuries caused to the bacterial cells to assume lethal proportions. Such plateauing therefore decreases as concentrations of agent employed become greater (Fig. 2A) and are inapparent at

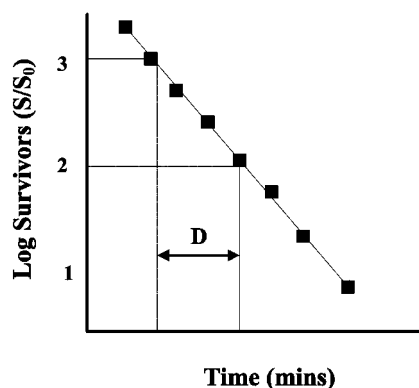


Fig. 1 Time survival kinetics for the inactivation of a bacterial population by treatment with a chemical agent. The D -value represents the time taken, at a given concentration of biocide, to reduce the surviving population by 90%.

very high concentrations. The second effect reflects the presence of resistant organisms within the population, capacity effects on the action of the preservative, or failure of the neutralizer. In such cases commonly observed with quaternary ammonium compounds and biguanides, the levels of kill at the plateau value increase with increasing concentration of preservative. From these relationships, an estimate of the initial rate of killing (k) or the D -value (actual time to effect a 90% reduction in viability) can be made, but their predictive value must be questioned.

Approximation to first-order kinetics implies that the sensitivity toward preservatives of a population of micro-organisms is normally distributed. Decreases in the rate of killing with time often results from adsorptive loss of the preservative onto and into the killed cells, causing a

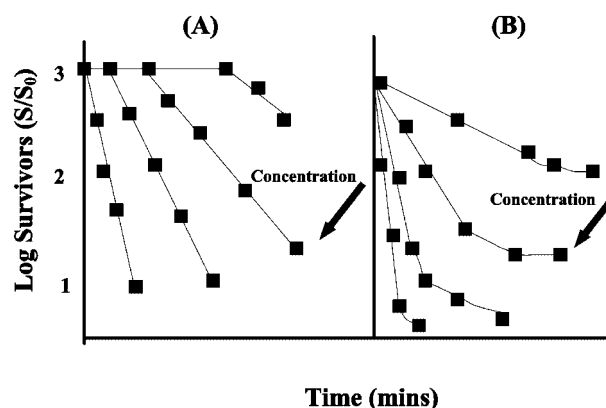


Fig. 2 Deviations from first-order kinetics of the inactivation of microbial populations by treatment with chemical agents due to (A) an initial lag in the rate of killing or (B) a decrease in the rate of killing with time of exposure.

decrease in the available concentration of the preservative for killing the surviving organisms. The susceptibility of preservatives to such adsorptive losses is described as its capacity.

Exponential decreases in viability with time mean that total elimination of all living cells cannot be guaranteed. Rather, knowing the initial size of the population and its susceptibility, the probability of having achieved sterility might be assigned to particular inactivation processes. Such assignments assume, often wrongly, that first-order kinetics are universally applicable.

Concentration of Preservative

The efficacy of an antimicrobial agent varies as a direct function of its concentration. At very low concentrations the antimicrobial effect might be negligible, at intermediate concentrations the effects might be growth-inhibitory, whereas only at relatively high concentrations might they be bactericidal. Often different mechanisms of action are involved to bring about these different effects. For example, chlorhexidine inhibits growth through direct effects upon respiratory enzymes and ATP synthesis, but kills through causing gross permeability changes to the bacterial cell envelope. The minimum growth inhibitory concentration (MIC) of an agent gives little or no information about its bactericidal activity, but might indicate weaknesses in the overall antimicrobial spectrum of activity. Minimum lethal concentrations can range from $2 \times \text{MIC}$ – $20 \times \text{MIC}$, dependent upon the antimicrobial substance and target organism. In the bactericidal range, the concentration exponent (η) describes the dependence of activity on concentration. This can be obtained as the slope of a graph of $\log D$ or k versus \log concentration. Although these values vary slightly among organisms, they are fairly characteristic of compound groups and also often relate to mammalian toxicity. Thus, compounds with high concentration exponents more readily lose their biological activity when diluted upon parenteral administration than those with low exponents (Table 2).

The concentration exponent gives no indication as to the level of activity of a particular agent. It only indicates the degree of dependence of activity upon concentration. If the \log_{10} of a death time is plotted against the \log_{10} of the concentration, a straight line is usually obtained, the slope of which is the concentration exponent. Given the D -value for a single concentration and concentration exponent, it is possible to calculate (from Eq. 5) the activity at any other concentration. As different antimicrobial mechanisms are often involved in growth inhibition and killing, the use of concentration

Table 2 Some examples of preservative concentration exponents

Compound	Exponent
Organic mercurials	0.9–1.0
Iodine	0.9–1.0
Formaldehyde	1.0–1.1
Benzalkonium chloride	0.8–2.5
Bronopol	0.7–0.9
Parabens	2.5–3.0
Phenolics	4.0–9.9
Aliphatic alcohols	6.0–12.7

exponents to extrapolate MIC data to bactericidal situations should be avoided.

$$(\eta) = \frac{(\log \text{Dat}C_2 - \log \text{Dat}C_1)}{\log C_1 - \log C_2} \quad (5)$$

The value of the concentration exponent also gives an indication of the susceptibility of the agent to inactivation by dilution and/or losses in activity through adsorption of the agent on organic debris or killed bacterial cells. The loss of an antimicrobial agent from solution in this manner relates to concentration in solution and absorptivity rather than to the inherent antimicrobial activity of the agent. The activity and concentration exponents of four hypothetical agents are described (Fig. 3). Agents A & B and C & D have similar concentration exponents

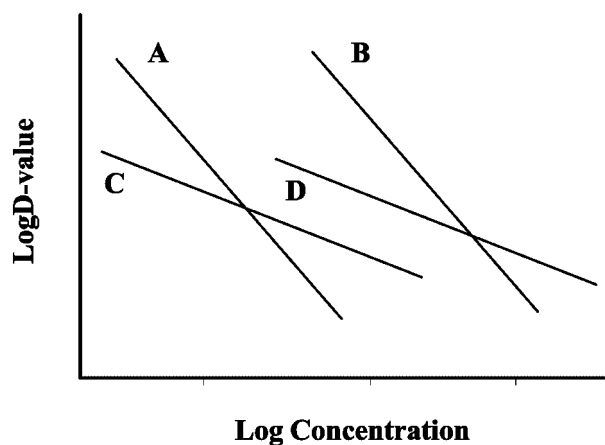


Fig. 3 Comparisons of the activity and concentration exponents of four hypothetical preservatives. Agents A & B and C & D have similar concentration exponents but different activities, whereas A & C and B & D have similar activities yet different concentration exponents.

but different activities, whereas A & C and B & D have similar activities yet different concentration exponents.

If isotoxic concentrations of each agent (equivalent *D*-value) are employed, and undergo a uniform drop in concentration through adsorption, the compounds with the higher activity will undergo the greatest loss of activity (*A* & *C* > *B* & *D*). For compounds of similar activity, the one with the higher concentration exponent will have the greatest activity loss (*A* > *C* and *B* > *D*). This can be generalized in the statement that capacity is inversely related to both the concentration exponent and activity. Preservatives used in sterile products generally have high concentration exponents and high activity. Such considerations are also helpful in determining inactivation conditions for the preservatives prior to sterility and preservative effectiveness tests.

Temperature Effects

The rate of killing of an antimicrobial agent is directly dependent on the temperature of the interaction. This dependence is described by the temperature coefficient (*Q* or *θ*), which can be determined either from the slope of the graph relating the *D*-value or *K* and temperature, or from Eq. 6.

Temp. Coefficient [*θ*_{*t*₁–*t*₂}]

$$= \frac{D\text{-value } t_2(D_2)}{D\text{-value } t_1(D_1)} = \frac{kt_1}{kt_2} \tag{6}$$

The subscripts denote the temperature range over which the change in activity is defined. This is commonly given as the coefficient per 10°C rise (i.e., *θ*₁₀ or *Q*₁₀), but is sometimes quoted for 1°C and 25°C changes.

Although the temperature coefficients vary a little between organisms, they are characteristic for particular preservative groups (Table 3). A high *θ*₁₀ value renders a compound suitable for application at elevated temperatures but unsuitable for the preservation of formulations that might be stored over a wide temperature range.

It should be noted that the temperature will also affect the degree of ionization in some circumstances; therefore,

Table 3 Temperature coefficients (10°C) for some commonly used preservatives

Compound	<i>θ</i> ₁₀
Aliphatic alcohols	30–50
Phenolics	3–5
Formaldehyde	1.5
Chlorocresol	6.0
Organic mercurials	1.0

the concentration exponent for some compounds will be apparently affected by temperature.

Effect of pH

The degree of ionization of acidic and basic antimicrobial agents depends on pH. Some compounds are active only in the unionized state (e.g., phenolics) whereas others are preferentially active as either the anion or cation. It therefore follows that the activity of a particular concentration of an agent will be enhanced at a pH that favors the formation of the active species. Thus, cationic antibacterials such as acridines and quaternary ammonium compounds are more active under alkaline conditions. Conversely, phenols and benzoic acid are more active in an acid medium. Chlorbutol is less active above pH 5 and unstable above pH 6. Phenylmercuric nitrate is only active at above pH 6 whereas thiomersal is more active under acid conditions. The sporicidal activity of glutaraldehyde is considerably enhanced under alkaline conditions whereas hypochlorites are virtually ineffective at above pH 8.

Knowledge of the *pK_a* of a preservative allows the calculation of the fraction undissociated at any particular pH and hence (via the concentration exponent) the level of activity at that pH.

Fraction of undissociated biocide

$$= \frac{1}{1 + \text{antilog } (pH - pK_a)} \tag{7}$$

The pH also affects activity by altering the net surface charge on the microorganisms. Thus, as pH is decreased, the surface becomes less negative and adsorption and thereby activity of cationic agents is reduced. These effects are greater for gram-negative than gram-positive bacteria.

Presence of Interfering Substances

In most instances, before an antibacterial agent can act on a cell, they must first combine. The presence of other materials, commonly referred to as organic matter, may reduce the effect of such an agent by adsorbing or inactivating it and thus reducing the amount available for combining with target cells. Extraneous matter may be able to form a protective coat around the cell, thereby preventing the penetration of the active agent to its site of action.

PRESERVATIVE INTERACTIONS

Until the late 1930s, cosmetic and pharmaceutical formulations were stabilized with soaps. Although preservatives

were often required, spoilage was only rarely a problem. With the advent of new materials for emulsion stabilization, preservation of the systems, or the lack of it, became a major problem. Nowadays we are aware of apparent incompatibilities between the new excipients and the established preservative agents and of the many interactions that can lead to preservative failure. No all-embracing set of rules have emerged to enable the "correct" preservative to be chosen for each formulation; only general guidelines exist.

The simplest form of interaction is one of adsorption/adsorption of the preservative from the formulation into and onto container and excipients. In this respect, it is convenient to divide these into colloidal excipients, suspensions, metal ions, and polymers.

Colloids

- **Gums:** Tragacanth, acacia, and plant polysaccharides such as agar and alginates are generally anionic in nature and will adsorb cationic preservatives such as chlorbutol, benzalkonium chloride, substituted phenols, phenylmercuric acetate, and merthiolates. Generally the adsorptive process increases with the increase in concentration of the gum or polysaccharide. Colloidal solutions of tragacanth and acacia, but not agar or alginates, can often be preserved with cationic agents by increasing the concentration of the preservative to saturate the colloid. For preservatives such as parabens, the problem is exacerbated at alkaline pH when the preservative becomes fully ionized.
- **Starches:** These bind substituted phenols but this can be overcome by increasing the concentration of the phenolic. Gelatin shows a slight interaction with the parabens group.
- **Proteins:** These interact both with cationic and anionic preservatives depending on to the pH of the formulation and the net charge of the protein.

- **Nonionic surfactants:** Interactions with a range of preservatives are unpredictable. Cationic and anionic surfactants interact and are contraindicated with anionic and cationic preservatives respectively.
- **Polyethylene glycols:** These interact particularly with the phenolic group of preservatives. The interaction increases with increase in molecular weight of the polymer. With increasing temperature, the extent of binding increases due to thermal desolving of the polymer and the creation of further binding sites.

Suspending Agents

These include clays, bentonite, kaolin, and talcs. They are generally anionic in nature and interact with most cationic agents. Inactivation depends on the pH of the system. Preservatives are only loosely associated with clays and therefore, desorb on dilution. The extent of these interactions are difficult to predict.

Container Materials

Nylon strongly adsorbs phenols and sorbic acid by covalent linkage. The degree of interaction depends on the pH, nature of the solvent, temperature, contact time, surface area, preservative concentration, and thickness of the nylon.

Acrylates and polyethylene affect most preservatives, including organic mercurials, phenolics, and benzoic acids. Leaching of hydroxyl ions from glass raises the pH and indirectly affects preservative activity. In this respect, glass must pass a "limit-test" for alkalinity.

Changing the container of a product from glass to plastic can alter its preservation. The container is as much a part of the formulation as is the product that is placed in it. Equally, adsorption and absorption into rubber closures can be significant, as exemplified by the phenolics (Table 4).

Table 4 Loss of preservative through absorption into rubber caps

Preservative	Initial concentration (% w/v)	Final concentration (% w/v)	Percentage loss
Phenol	0.5	0.39	22
Cresol	0.3	0.21	30
Chlorocresol	0.1	0.04	60
Phenyl mercuric nitrate	0.001	0.00005	95

PRESERVATION OF TWO-PHASE SYSTEMS

When preservatives are added to two-phase oil–water systems, depending upon the hydrophilicity of the agent, they partition between the two phases. Bacteria will only grow within the aqueous phase or at the oil–water interface. Thus, preservative lost to the oil is unavailable for antibacterial action. Hence the concentration in the water phase is reduced, affecting the rate of kill, but the capacity of the system is unaffected as preservative lost by adsorption onto organic debris and bacteria is replaced from that held in the oil.

The distribution between the oil and water phases is easily predicted from the oil–water partition coefficient, K_w^0 (Eq. 8)

$$K_w^0 = \frac{\text{Concentration in oil}}{\text{Concentration in water}} \quad (8)$$

By knowing the concentration of preservative required for the necessary biological effect in the aqueous phase (C_w), the oil: water phase ratio (θ) and the K_w^0 for the system, C , the concentration needed in the formulation, may be calculated from the expression

$$C_w = \frac{C(\theta + 1)}{K_w^0 \theta + 1} \quad (9)$$

It is also worth noting that altering the nature of the oil can affect partition coefficients quite dramatically, for example, the differences between vegetable and mineral oils (Table 5).

If the preservative favors the oil, increasing the oil: water ratio dramatically decreases available concentration; if it favors the aqueous phase, the reverse occurs and can cause solubility problems. Unfortunately, both temperature and pH also affect K_w^0 , in addition to the effects of pH directly on the activity of the preservative.

Table 5 Partition data for various preservatives

Preservative	K_w^0	
	Mineral oil	Vegetable oil
Chlorocresol	0.5	11.7
Methyl parabens	0.02	7.5
Propyl parabens	0.5	80.0
Butyl parabens	3.0	280.0
Cetrimide USP	<1.0	<1.0
Bronopol	0.043	0.11
Phenonip	>1.0	>1.0
Phenyl mercuric nutrate	<1.0	<1.0

PRESERVATION OF THREE-PHASE SYSTEMS

Most oil–water systems contain surfactants as emulgens in order to provide a stable emulsion. Some preservatives may associate directly with the surfactants whereas others may partition into the micellar phase of the surfactant, treating it as a third phase for partitioning.

These effects can be calculated to some extent, where C_w depends not only on K_w^0 but also on surfactant concentration. Initially it is assumed that the surfactant micelles form part of the aqueous phase. Thus, if K_w^0 is the oil:aqueous phase partition coefficient and C_A the total aqueous concentration, then, as in Eq. 9,

$$C_A = \frac{C(\theta + 1)}{K_A^0 \theta + 1} \quad (10)$$

Now

$$C_w = \frac{C_A}{R}$$

where

$$R = C_A/C_w \text{ that is } C_A = RC_w.$$

As

$$K_w^0 = C_A/C_w \text{ and } C_0 = K_w^0 C_w,$$

substitution in Eq. 9 gives

$$C_A^w = \frac{C(\theta + 1)}{C_0/C_A^{\theta+1}} \quad (11)$$

and substituting once again for C_0 ,

$$C_A = \frac{C(\theta + 1)}{(K_w^0 \frac{C_w}{C_A \theta}) + 1} \quad (12)$$

Substituting in Eq. 12 for $C_A = RC_w$ gives

$$RC_w = \frac{C(\theta + 1)}{(K_w^0 C_w / RC_w \theta) + 1} \quad (13)$$

which then reduces to

$$C_w = \frac{C(\theta + 1)}{[K_w^0 \theta + R]} \quad (14)$$

Eq. 14 allows the active concentration in the water phase to be calculated from C , the phase ratio (θ), K_w^0 and R . This is useful as K_A^0 varies with the concentration of the surfactant. R can be found from a graph relating

C_A/C_W and surfactant concentration, where the slope of the line is k .

$$R = 1 + k[\text{surfactant concentration}]$$

All preservative manufacturers provide values of k for a range of surfactants.

TOXICITY AND SAFETY CONSIDERATIONS

There are certain limitations on the use of antimicrobial preservatives based on their toxicity and potential side effects. The British Pharmacopoeia recommends that the maximum quantity of an injection containing an antimicrobial preservative to be administered at a single occasion is 15 ml. The USP, on the other hand, recommends a maximum of 5 ml. These maximum volumes must be considered when toxicity data and concentrations of the agents included in the formulations are considered. The final concentration of any substance employed is a balance between its potential toxicity and antimicrobial activity. Additionally, because meninges are readily irritated, the inclusion of antimicrobial preservatives in any formulation that might gain access to the cerebrospinal fluid is precluded. Intracardiac and intraarterial injections must also not contain an antimicrobial preservative.

It is not surprising that preservatives sometimes prove toxic to humans. To assess the toxicity of any substance, it is necessary to know its acute and cumulative toxicity together with its "no effects levels" and also to be aware of its mutagenic, teratogenic, and carcinogenic potential. In addition, its local actions as well as its liability to cause hypersensitization might be important. Any new preservative must be subjected to a battery of such mandatory tests to provide these data as must a new application of an established biocide.

ACUTE TOXICITY

Preservatives, for the most part, are used at very low concentrations and the question of their acute toxicity rarely arises. The margin of safety is not always very high, however.

SUBACUTE AND CHRONIC TOXICITY

When substances are given repeatedly, manifestations of toxicity may differ, particularly if the compound is given

at a rate faster than what the individual can detoxify or excrete. Some preservatives contain mercury and this may accumulate in the body.

REPRODUCTIVE TOXICITY

As part of the general safety testing of all preservatives, consideration is given to possible effects in higher animals on the fertility of either sex, the gestation or postnatal care, and development of the embryo. No direct causal relationships have been shown between any of the commonly used preservatives and reproductive toxicity. However, indirect reproductive toxicity has been suggested, for example, benzoic acid as a preservative, increases blood salicylate levels, and may cause a significant increase in the teratogenic potential of aspirin. Organic mercury has been shown to be responsible for a congenital form of Minamata disease, but not at levels likely to be taken into the body through mercury-containing preservatives.

CARCINOGENICITY AND MUTAGENICITY

Several preservatives have been scrutinized for being possible carcinogens, and their use regulated (e.g., chloroform and formaldehyde). Another possibility is that the preservative interacts with amines or amides to form carcinogenic nitrosamines. Published data on the mutagenicity of preservatives is limited as tests such as the Ames test can only be carried out at concentrations far less than those employed.

PRESERVATIVE COMBINATIONS

The ideal properties expected from preservatives are not met by any of the current, established agents. All chemical agents have their limitations in terms of their antimicrobial activity, resistance to organic matter, stability, incompatibility, irritancy, or toxicity. It is also unlikely that new preservatives will be developed because of high costs of putting a new compound through toxicity tests and the likelihood of a new compound being impossible to cover by patent. More likely preservative combinations will be employed in the future, possibly to give synergistic combinations (e.g., hydrogen peroxide and peroxygen compounds, chlorhexidine and cetrimide). In this manner,

Table 6 Examples of interactions between preservatives

Preservative A	Preservative B		
	Synergistic	Additive	Antagonistic
Benzalkonium	Phenylmercurics	3-Cresol	Hexachlorophane
Benzoic acid	Dehydroacetic acid	Parabens	Boric acids
Bronopol	Benzalkonium	Sorbic acid	—
Chlorbutanol	Phenylethanol	—	EDTA, EGTA
Chlorocresol	Phenylethanol	—	—
Phenylmercurics	3-Cresol	—	—
Parabens	Germall	Benzoic acid	Boric acid
Sorbic acid	Dehydroacetic acid	Benzoic Acid	Parabens

preservative combinations may be used to extend the range and spectrum of preservation (Table 6); for example, by combining a series of alkyl esters of 4-hydroxybenzoic acid, water solubility is decreased to such an extent that both the aqueous and oil phase of an emulsion are protected.

When acting simultaneously on a microbial population, combinations of preservatives may cause an increased, decreased, or unchanged antimicrobial response when compared with their summed individual effects. A minimum requirement of any combination should be that it achieves at least the same level of protection overall as the individual components do. Ill-considered preservative combinations may lead to inclusion of irrelevant agents. In general, agents from the same chemical group or those having the same mode of action are likely to produce merely additive effects whereas those exhibiting different mechanisms or sites of action may serve either to reinforce (synergize) or reduce (antagonize) their individual activities. Such effects may be evaluated by preparing mixtures of the two preservatives being investigated and determining their growth inhibitory power by an MIC determination. Results may then be plotted in the form of a graph, termed an isobologram, an example of which is shown in Fig. 4.

In general, synergistic effects can occur through three different mechanisms. The first mechanism is by inhibition of inactivation. Here, one compound increases the effective concentration of the other by inhibiting the microbial system responsible for its inactivation. To date, however, there are no reported observations of this mechanism with preservatives. The second synergistic mechanism is very common amongst preservatives and occurs when one compound increases the accessibility of targets to another. This usually occurs when one agent

exerts its action by increasing the permeability of the cell wall or cell membrane. Besides damaging the cell in its own right, this can permit access (or increased access) of the second compound to targets that were previously concealed. An example of this type of synergy is seen in the cooperative action between benzalkonium chloride and thiomersal which disrupts the cytoplasmic membrane, thereby facilitating access of organomercurial agents that react with sulphhydryl-containing enzymes in the cytoplasm. This type of synergy does not necessary require

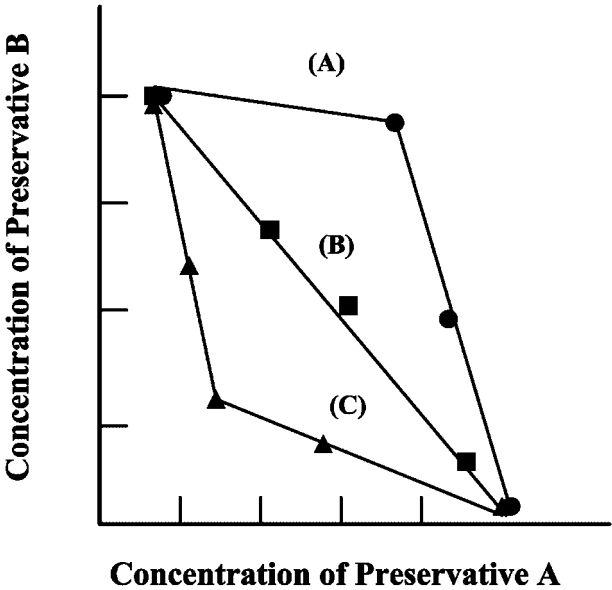


Fig. 4 Isobologram showing three possible outcomes of interaction between preservatives A and B, namely (A) antagonism, (B) additivity, and (C) synergy.

both compounds to possess significant antimicrobial activity. Finally, perhaps the broadest category of synergistic interactions is that in which two compounds act simultaneously at different targets, thereby significantly influencing the biochemistry of the cell. Although these targets must be different in order to produce synergy, they can be closely related. For example, the synergy between chlorocresol and 2-phenylethanol is thought to be through their respective effects on the generation and coupling of a proton gradient to active transport.

There are many significant practical benefits and advantages of preservative combinations. These include an increased spectrum of activity; the need for lower

concentrations of each of the individual components, thereby possibly reducing potential toxicity; the prevention of resistance to individual preservatives; possible enhancement of antimicrobial activity beyond that expected by simple addition; an extended time course of preservation achieved by combining a labile, markedly biocidal preservative with a stable longer-acting ingredient. Furthermore, careful selection of individual agents for combination, based on their physicochemical properties, may serve to overcome microbiological problems created by the physical limitations of individual preservatives. In this respect, factors affecting preservative efficacy must be carefully considered.

PROCESS CHEMISTRY IN THE PHARMACEUTICAL INDUSTRY

Kumar G. Gadamasetti

ChemRx Advanced Technologies, Inc., San Francisco, California, U.S.A.

Ambarish K. Singh

The Bristol-Myers Squibb Pharmaceutical Research Institute, New Brunswick, New Jersey, U.S.A.

INTRODUCTION

In the pharmaceutical industry, identification of a development drug candidate (preclinical lead profile, PLP, or early candidate notification, ECN), filing of investigational new drug application (IND), and new drug application (NDA) are important milestones before the launch of a new drug. The IND and the NDA are the events where the industry interacts with the Food and Drug Administration (FDA) prior to launching. Various governmental departments play specific roles in furthering drug development programs. The medicinal or discovery chemists identify the new drug candidates to treat or prevent a particular medical indication, whereas the process chemists are responsible for devising a synthesis and supplying the active pharmaceutical ingredient (API) or bulk drug substance (BDS) in multigram quantities for various studies needed to file the IND and support other drug development programs. On approval of the IND, the compound can be administered to humans for the first time as part of the phase-I clinical studies, also known as first-in-man (FIM) trials. Studies that are reported in an IND include synthesis, animal toxicology, pharmaceuticals and formulation, drug substance and drug product stability and safety, and metabolic and pharmacokinetics.

The clinical development is the most expensive and resource-intensive segment of the process. Process research and development play a key role in shortening the overall timeline from candidate identification to drug launch. On an average, it takes about 10–15 years from discovery to launch a drug into the market at a cost in excess of \$400 million. The paradigm shift (in the 1990s) of increasing the number of compounds entering development has made a tremendous impact on chemists responsible for preparing supplies of these new drug candidates. Early and effective interaction of process research personnel with medicinal chemists and early innovations in process development are believed to shorten the IND and NDA timelines, respectively. These overall reduced timelines would allow for an early launch of the drug into the market (1).

PHARMACEUTICAL DRUG DEVELOPMENT EVENTS: THE CHEMIST'S VIEWPOINT

A simplified view of the pharmaceutical drug development events (not to the timescale) is given in Fig. 1.

Genomics to Lead Development Candidate (PLP)

During the past decade, the pharmaceutical industry has seen a paradigm shift in the drug discovery process. The driving force for such change, arguably, was propelled by a variety of factors, including

- Generation of new targets (for human diseases) from advances in genomics and functional genomics.
- Advances in combinatorial chemistry methods to increase the number of compounds for high throughput screening (HTS), and
- Advances in automation and high-throughput screening techniques for rapid identification of lead compounds.

Although these approaches promise to provide an increased number of novel drug candidates for evaluation in the treatment of a greater number of diverse diseases, successful realization of the potential benefit of these compounds is very much dependent on the ability of the pharmaceutical industry to develop suitable manufacturing processes.

PLP Development Candidate to IND and the Kilo Lab

The IND studies require relatively large amounts of the drug substance, substantially more than what was prepared during the course of the medicinal chemistry programs. Generally, it becomes the responsibility of the process research group to produce this material within the shortest feasible time frame (2). Most of the major pharmaceutical companies have kilo-lab facilities as an interface between the process research labs and the pilot plants. Pharmaceutical process groups are organized in different ways. In some companies, process research is responsible for

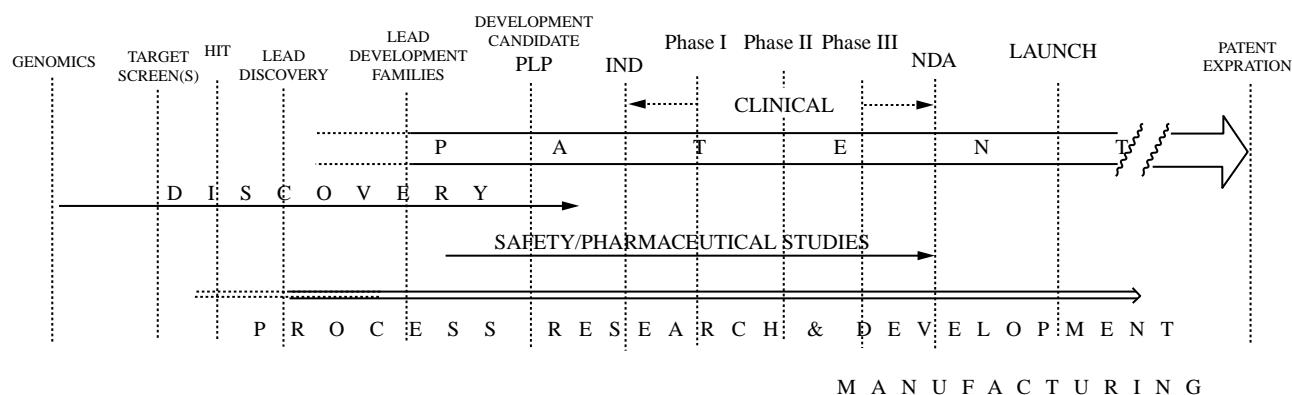


Fig. 1 Pharmaceutical drug development events, a simplified view. *Preclinical*: Medicinal chemistry, combinatorial chemistry, process research. *Clinical*: Process R&D & manufacturing (operations).

providing the supplies of BDS required for an IND filing, including the initial clinical batch (or batches). Typically, the initial supplies are prepared by the combination of modified medicinal chemistry synthesis and new, alternative synthetic processes more suitable for scale-up. In some companies, the chemists in process research labs and kilo labs share this responsibility. Kilo-lab scale-up work typically is performed in 22-L glassware and small reactors (50–100 L). The process development group, on the other hand, is responsible for further optimization and scale-up in the pilot plant and interfaces with the manufacturing.

Clinical Phase, NDA, and Launching of the Drug

The clinical phase (phase I) of the drug development begins after the IND where humans are subjected to the clinical trials. After a successful phase-I trial, efficacy of the drug is tested on a large number of patients as part of the phase-II/III clinical trials. Once the phase-II/III clinical trials are completed, the NDA is filed with the FDA. Clinical trials and NDA are prerequisites for launching the drug.

PROCESS RESEARCH AND DEVELOPMENT

The goal of the medicinal or drug discovery chemist is to identify a new lead to treat or prevent some particular medical indication and define a synthetic strategy to allow for the preparation of as many analogs as possible. The focus of process chemistry differs from routine organic chemistry. It emphasizes optimization and defines the controls to make the sequence of chemical reactions amenable to scale-up. A viable process should reliably yield

a high-purity product made by a process unencumbered by a patent. In short, the overall thrust of the scientists engaged in process chemistry is to develop the shortest, least expensive, safest, and most environmentally friendly processes to produce the API in multi-kilogram quantities.

The term “process” is, in general, misinterpreted as scale-up work by the overwhelming majority of the scientists and technologists involved with drug discovery and development programs in the pharmaceutical industry. Scientists and engineers engaged in the various aspects of pharmaceutical process research and development have a highly refined appreciation for the challenges of large-scale synthesis and the purification of the API. An algorithm for process research and development is shown in Fig. 2.

The mission of process chemistry in the pharmaceutical industry is to provide documented, controlled synthetic processes for the manufacture of the supplies to support the development programs and future commercial requirements of the API. The science and technology associated in accomplishing this mission provide a tremendous challenge to the individuals or the group of individuals for the drug supply progress from milligram to metric ton quantities.

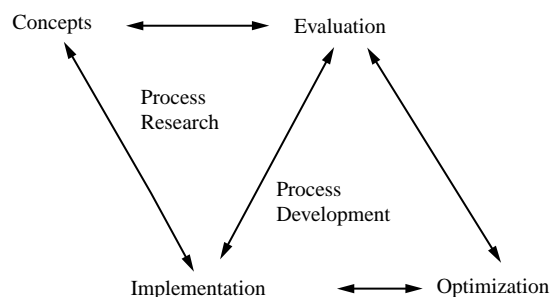


Fig. 2 Algorithm for process research and development.

Route Scouting

As the development drug candidate moves from discovery to process, a workable synthesis is available, which, however, may not be scalable. Route scouting plays a key role in identifying synthetic transformations that are safe, practical, scalable, cost effective, and environmentally friendly, thus setting the stage for eventually delivering a manufacturing synthetic process. An integral part of route scouting is to identify key intermediates that could be easily outsourced.

An example of route scouting and process improvements

Scheme 1 depicts the medicinal chemistry approach to the synthesis of BMS-180291 (Ifetroban).

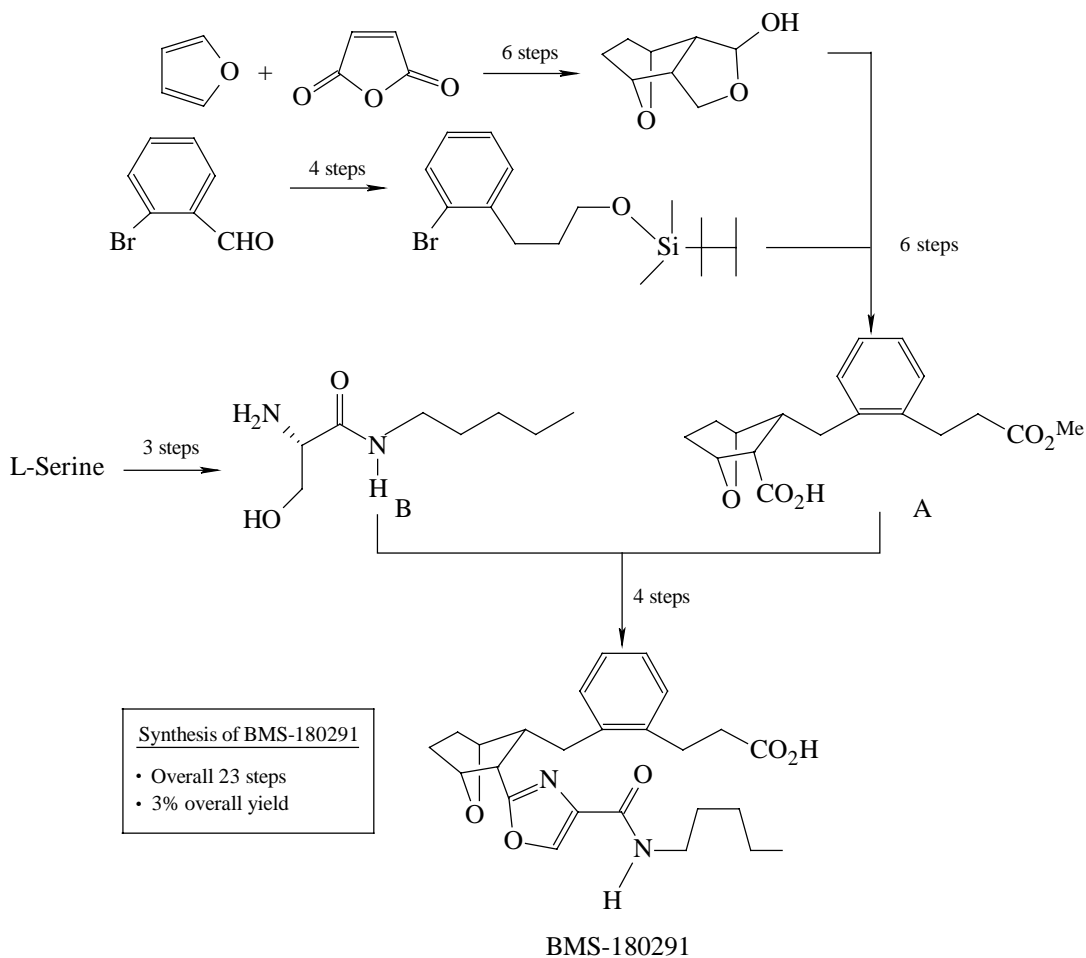
A condensed schematic of the overall alternative synthesis of BMS-180291 (Ifetroban) is provided in Scheme 2.

The synthesis of the key acid ester intermediate A has been reduced dramatically from 16 to 4 steps, and the yield has been increased 10-fold from 5 to 52%. The overall synthesis of Ifetroban has been cut virtually in half (from 23 to 12 steps) and the yield (via the longest linear sequence) improved from 3 to 28%.

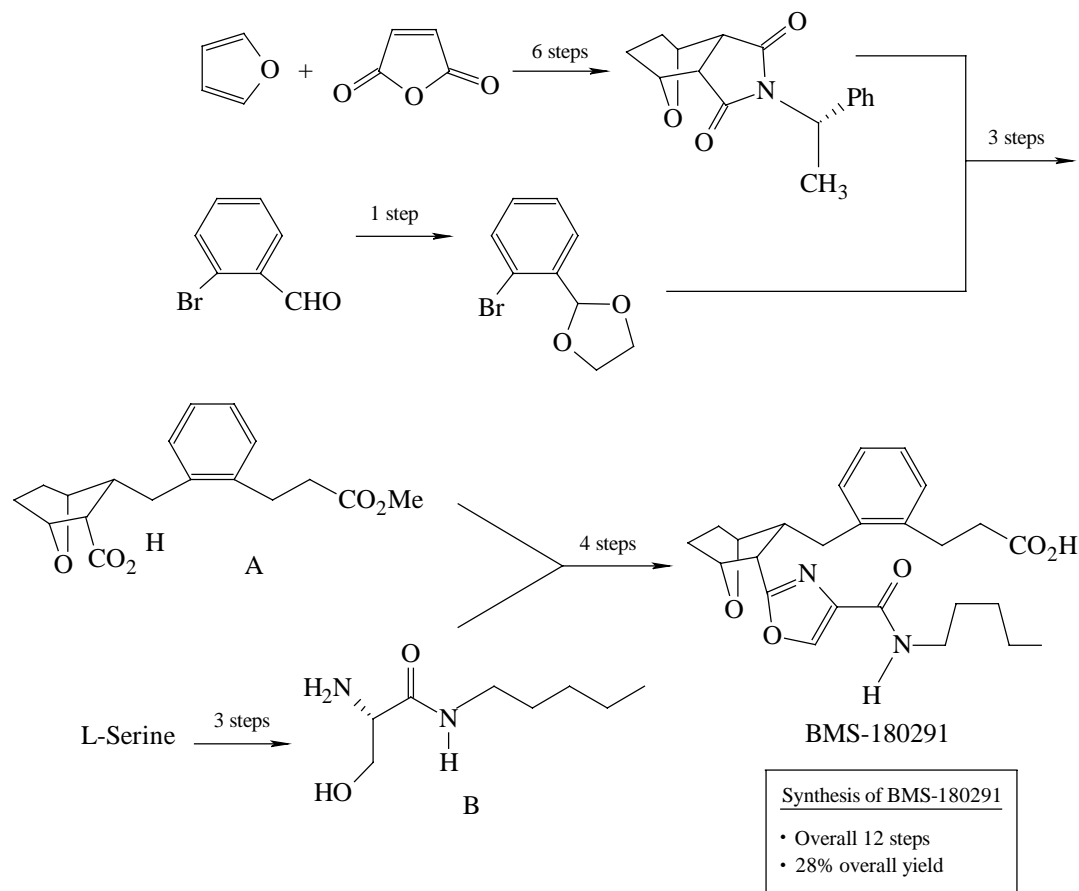
Process Development and Optimization

Human safety is the first priority when developing a process, followed by quality (purity, crystallinity of the product, etc.) and product yield. In addition, the process chemist takes into account the following factors:

- Hazards associated with the chemical step (human and equipment safety).
- Waste generated (disposal cost, environmental concerns).



Scheme 1 Original synthesis of BMS-180291.



Scheme 2 Alternative synthesis of BMS-180291.

- Cost and availability of raw materials, reagents, and solvents (cost of goods, sourcing issues).
- Ease of performing the reaction on-scale (savings in capital and labor).
- Ease of product isolation (crystallization), and
- Opportunity to combine two or more chemical steps with no intermediate isolation (telescoping to increase throughput).

Once a route has been selected, the next step is to understand the effect of interactions of variables with each other on the reaction yield and quality for each and every step of the synthetic scheme and to define the limits for these process variables. This requires designing appropriate experiments including controls. This is important because as the scale of the reaction increases, every unit operation, such as addition of reagents, solvents, distillation, phase splits, etc., takes a considerably longer time. Mixing begins to play a major role as the size of the vessel increases. Due to uneven mixing, localized differences in variables such as temperature and

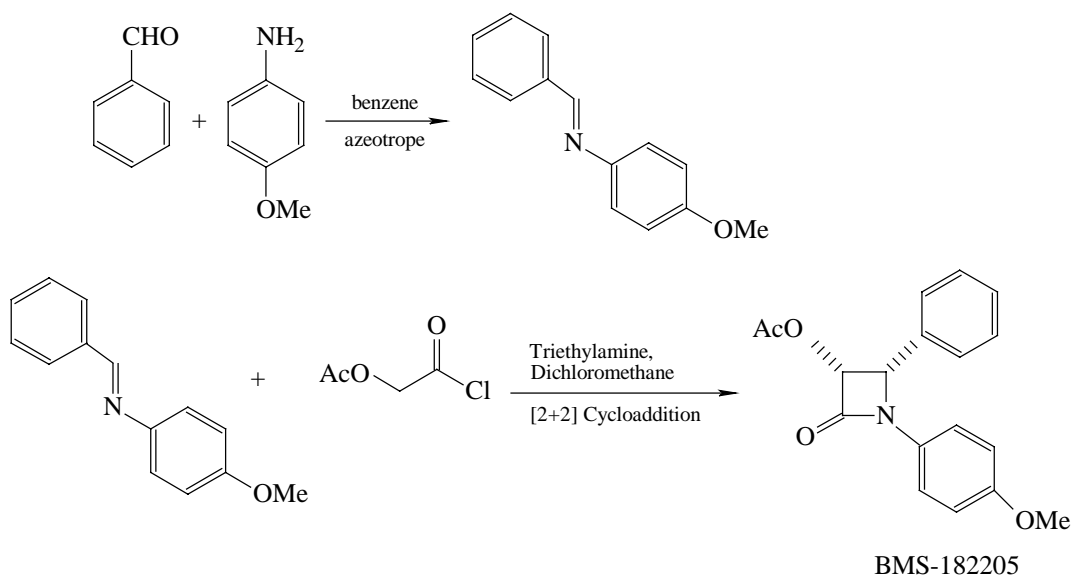
concentration are expected, which may have an adverse effect on the reaction.

Example 1: Process development of BMS-182205 (Scheme 3)

Issues: environmental concerns, human safety, waste management, process inefficiency.

The first step in the synthesis of the Paclitaxel[®] side chain has been reported in the literature by Holton (4). The drawbacks of this procedure from a process development standpoint are

1. The use of benzene and dichloromethane (human and environmental concerns).
2. No opportunity for telescoping (since the first reaction is conducted in benzene and the second in dichloromethane).
3. Low yield (68% after crystallization from ethylacetate and hexane), and
4. Brown-black product (color).



Scheme 3 Synthesis of BMS-182205.

Process improvements: *Phase I.* Benzene was replaced with a safer solvent, toluene, which was also used for the [2 + 2] cycloaddition reaction, thus facilitating the telescoping of the two reactions. However, dichloromethane was added during the work-up to prevent product precipitation. The product was crystallized from 2-propanol in 80% yield. The phase I of development eliminated the use of benzene and allowed for telescoping of the two chemical steps. However, a solvent exchange from toluene–dichloromethane to 2-propanol was still needed to crystallize the product. The color of the product after this modification was dark brown.

Phase II. Additional crystallization studies revealed that the product could be crystallized from toluene–heptane and acetone–water. The product also exhibited limited solubility in toluene. With this information at hand, both reactions were conducted in toluene as before; however, during the work-up, dichloromethane was not added. Instead, aqueous hydrochloric acid was added to neutralize excess triethylamine. Heptane was then added to precipitate the product directly from the reaction mixture. The crude wet product was crystallized from acetone–water with an overall yield of 80%. The color of the product was brown. The advantages of this modification are the elimination of several extraction and back-extraction steps and a reduction in processing time and in solvent consumption.

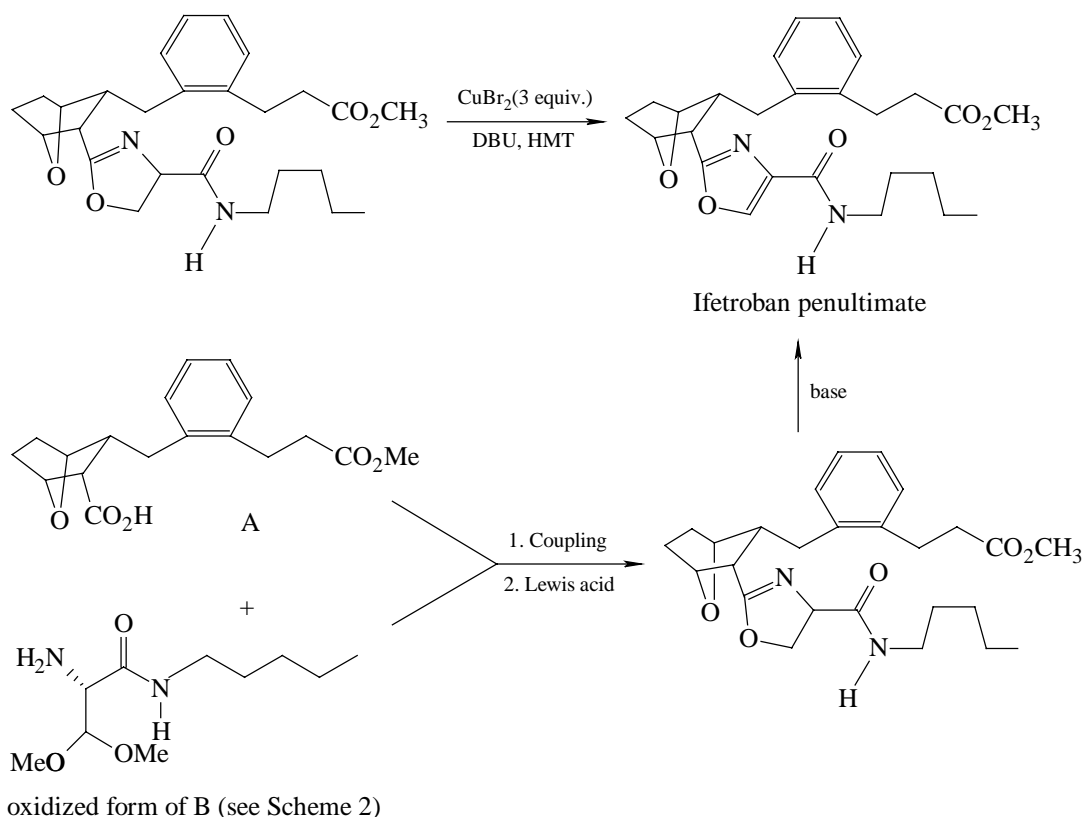
In an attempt to further improve the yield and address the product color issue, the following process variables were studied: the rate of addition of acetoxyacetyl chloride, the reaction temperature, and the effect of other amine bases.

The rate of addition of acetoxyacetyl chloride was optimized to be 2–3 h, at 3, 10, and 17°C. Reactions conducted at 3°C afforded the product with the highest quality. Two amines, *N*-methylmorpholine (NMM) and diisopropylethylamine (DIPEA), were evaluated in addition to triethylamine. In the presence of NMM, the reaction stopped after 65% conversion. In triethylamine the reaction took an additional 3 h for completion, after the addition of acetoxyacetyl chloride. However, in DIPEA, the reaction was instantaneous, the yield was further improved to 87%, and the product was free of brown-colored impurities.

Example 2: Process development of Ifetroban (Scheme 4)

Issues: environmental concerns, human safety, cost of goods (5).

In the improved synthesis of Ifetroban described previously, environmental concerns due to special handling of copper bromide waste and hazards associated with hexamethylene tetramine (HMT) on manufacturing scale led to further perfection of the synthesis. Mechanistic considerations suggested that an oxidized form of aminoamide *B* (Scheme 4) would eliminate the necessity for a late-stage copper-mediated oxidation. This was indeed accomplished. The cyclization–elimination sequence was initiated by a Lewis acid and completed by base-mediated elimination to afford the Ifetroban penultimate. In addition to eliminating the need for copper bromide and HMT, this modification helped to reduce the cost of the product by an additional 15%.



Scheme 4 Process development of Ifetroban.

Manufacturing process, route from medicinal chemistry to multi-kilos for clinical study supplies:

A highly optimized and concise large-scale synthesis of a purine bronchodilator was developed by the Astra Production Chemical company from Sweden (6). Supplies for the initial biological studies were generated by the medicinal chemistry route shown in Scheme 5. The overall yield was about 14%, which was improved in the environmentally friendly manufacturing process to about 51% (Schemes 6 and 7).

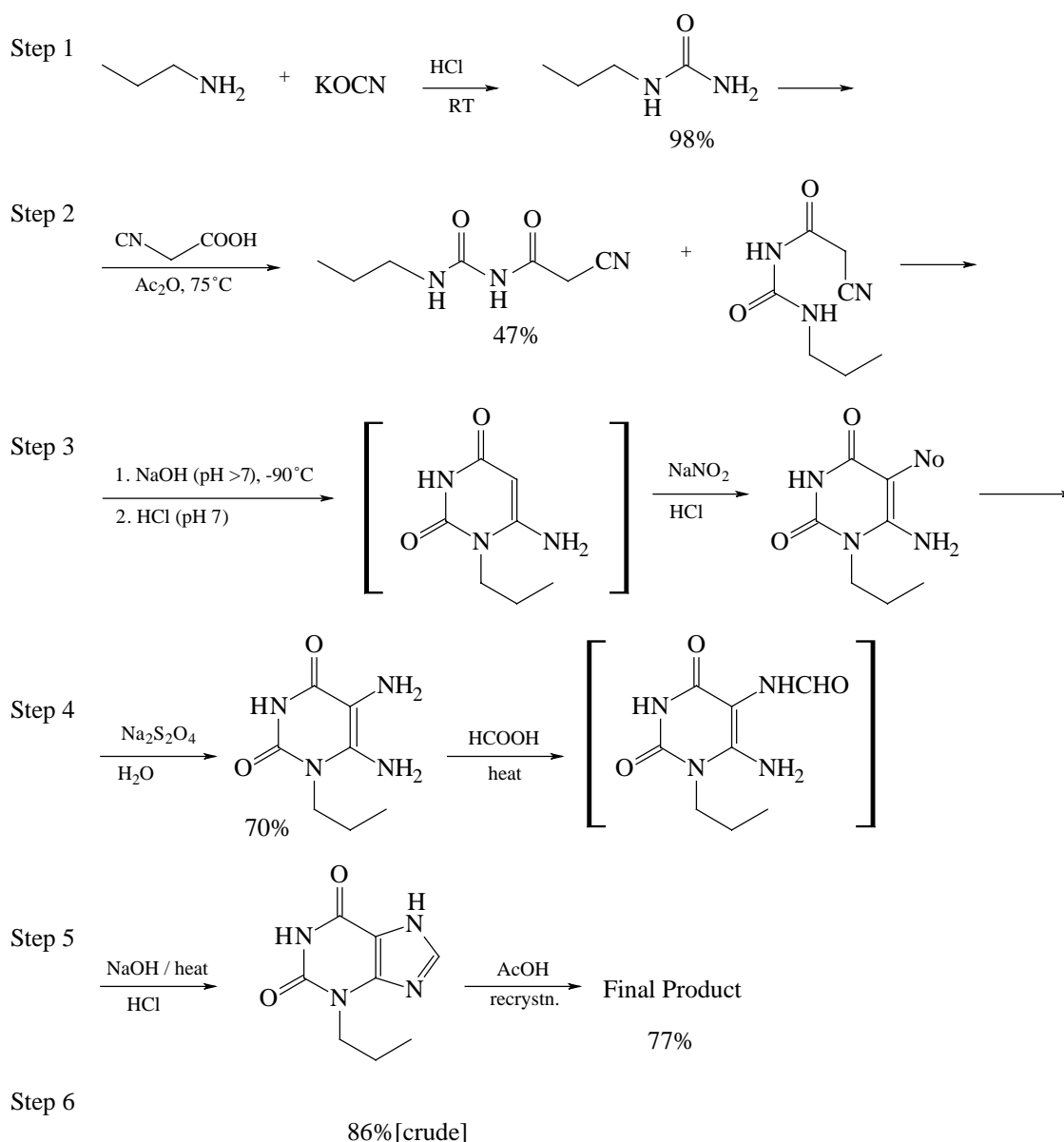
Process Hazards and Safety

Process safety has become an integral part of the development of new synthetic routes, and more and more relevant information is gathered in typical process development labs (7). Many development facilities have now established special process safety departments with state-of-the-art equipment and well-trained personnel. It is important for process chemists considering alternative process routes to know the potential hazards from the main reaction and from the unwanted side reactions in each case so that the hazards of reactivity are included in the factors

reviewed in developing and selecting the final process route; three main parameters determine the design of safe chemical processes:

1. The potential energy of the chemicals involved and understanding of the inherent energy (exothermic release or endothermic absorption) during a chemical reaction.
2. The rates of reaction (energy release in the form of heat or pressure) that depend on the temperature, pressure, and concentrations. In any hazard evaluation process, the rates of reaction during normal and abnormal operations (including the worst credible case) must be considered in order to design an inherently safe process.
3. An equipment train must adequately remove any heat or pressure generated in a reaction. The effects and requirements of scale-up must be considered.

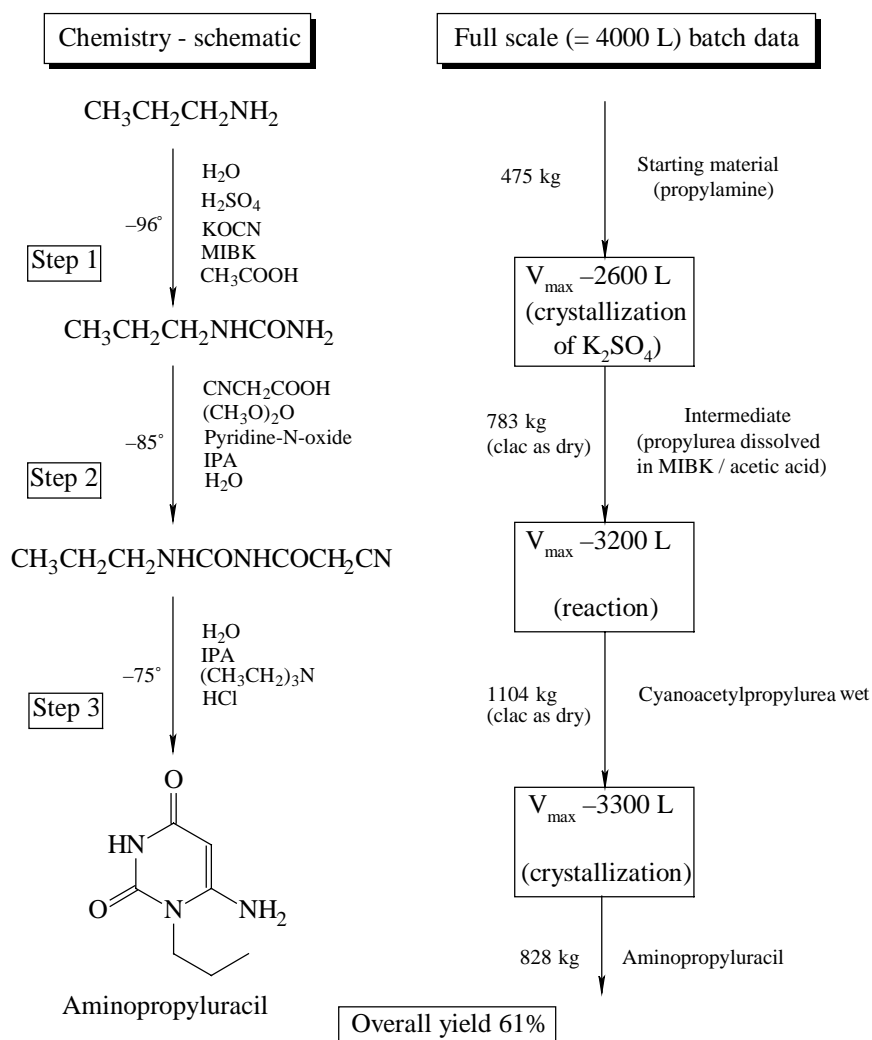
In most cases, a team of development chemists, engineers, and safety personnel evaluate and assess hazards associated with each and every step of a process, from performing a reaction to storage of waste streams in drums. The team recommends a set of safety experiments to be conducted, and data are collected for reaction



Scheme 5 Original synthesis of enprophylline by the medicinal chemistry department at Astra.

exotherms, powder explosivity, gas evolution, and compatibility of reaction mixtures or reaction waste with pilot-plant equipment and storage drums. Experiments are designed and conducted to determine the potential for initiation of a runaway reaction and the effect of decompositions that may occur on runaway. Based on the safety assessment of a process, appropriate measures are taken in the pilot plant to eliminate or minimize any hazard. During the early stages of process research (route

scouting), only small amounts of materials are available. In many cases, only theoretical information from the literature or from calculations is readily available. Screening tests can be run to identify the reaction hazards. As the route-scouting efforts enter the process development phase, additional material becomes available so that the reaction hazards can be studied more extensively to test “what-if” scenarios. During full-scale production, the chemical hazards may be reevaluated to address changed



Scheme 6 Astra commercial process of 6-amino-1-propyluracil.

production requirements or other process changes such as the use of a different source of raw materials.

Numerous test methods are available using a variety of sample sizes and conditions. The tests provide qualitative or quantitative data on onset of temperature, reaction enthalpy, instantaneous heat production as a function of temperature, maximum temperature, and/or pressure excursions as a consequence of a runaway, and additional data useful for process design and operation.

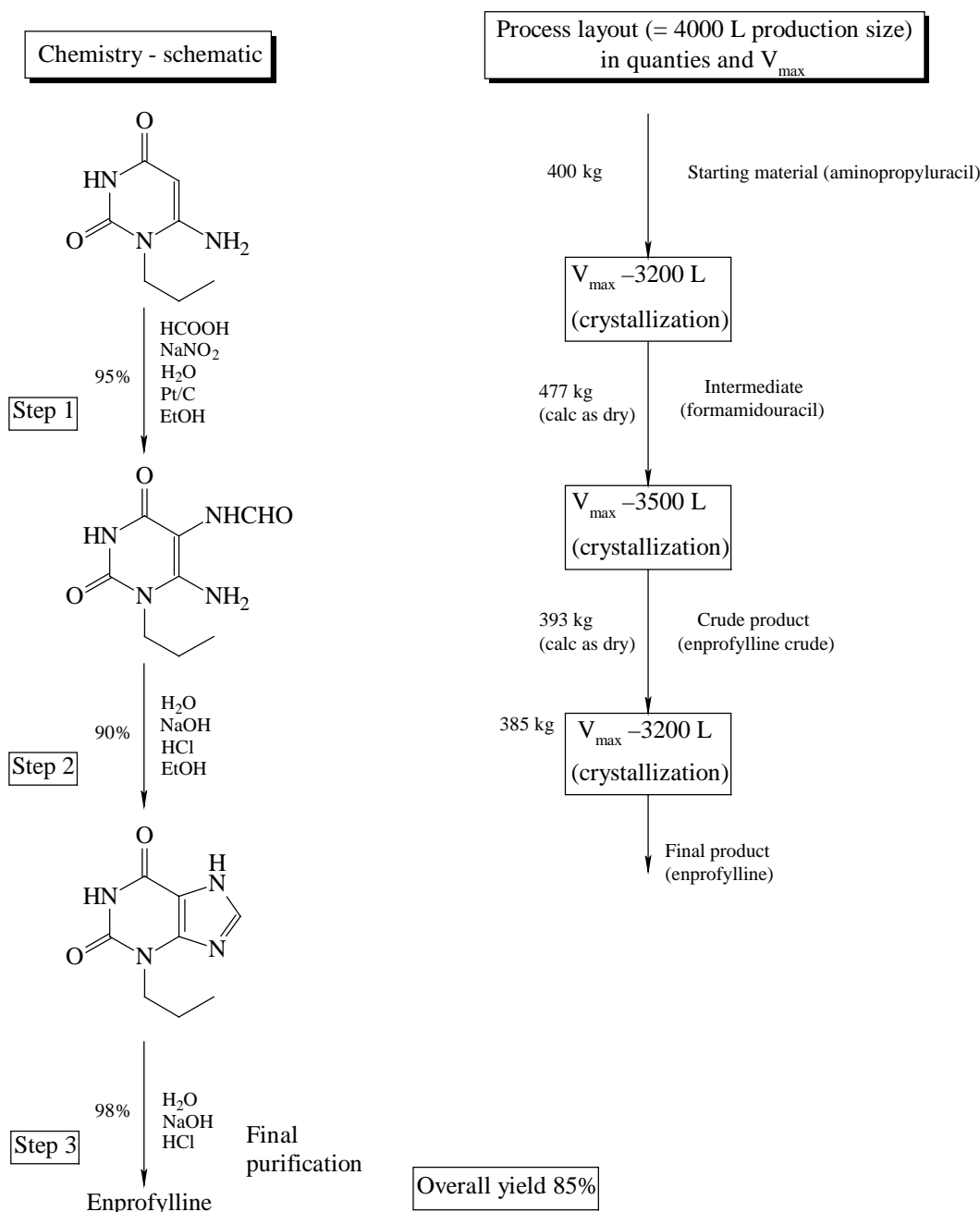
“Green Chemistry”: Environmental Concerns

The research that involves the end-of-process treatments to eliminate pollutants is termed “green chemistry.” As Ronald Breslow (Columbia University) pointed out (8), concern for the environment is as old as the biblical injunction, “hurt not the earth, neither the sea, nor

the trees.” The following example indicates approaches to the environmentally benign chemistry. The process described is high yielding with water as the by-product. Sato et al. (9) have developed an efficient, environmentally friendly method for oxidizing primary and secondary alcohols (Scheme 8). The Japanese scientists use hydrogen peroxide as an oxidant, a quaternary ammonium hydrogen sulfate as a phase-transfer agent, and tungsten as catalyst.

Enzymatic Intervention

The recent trend in pharmaceutical industries is to incorporate a microbial technology (MT) or enzymatic technology division in the process R&D and manufacturing (need basis) departments. The role of the MT division is usually to provide pharmaceutical drugs



Scheme 7 Astra full-scale synthesis of enprofylline from 6-amino-1-propyluracil.

(e.g., pravastatin, β -lactam drugs, etc.), and chiral (optically pure) building blocks or synthons for the ongoing synthetic programs.

A number of cholesterol-lowering drugs (Pravachol, Zocor, and Mevacor) are prepared by enzymatic processes. Penicillin antibiotics that have been in the market for decades are produced in large quantities by the enzymatic process.

Enzymatic catalysis

The extensive use of enzymatic catalysts in organic synthesis has been documented (10–13). Enzymes represent a broad range of efficient chemical catalysts. They are classified mainly into six categories:

1. Hydrolases (hydrolysis of amides, esters, glycosides, and lactones)

2. Isomerases (C–C bond migration, E/Z isomerization, and racemization)
3. Lyases (addition to π -bonds)
4. Oxidoreductases (reversible oxidations and reductions)
5. Synthetase formation and breaking of C–C, C–N, C–O, C–S, and phosphate ester bonds
6. Transferases (transfer of acyl, glycosyl, or phosphoryl groups from one molecule to another)

Scheme 9 gives three examples.

Design of Experiments (DOE)

An understanding of how various process variables affect the chemistry is necessary for the design of a chemical process (14) that can reliably provide the product in high yield and quality. This understanding can be obtained from experimentally determined rate laws for the main and side reactions that relate temperature, concentration, pressure, solvent effects, and equivalents of each reagent to yield and impurity levels. The process chemist may take the approach of evaluating the importance of process variables by changing one variable at a time. This method can help generate chemical knowledge but is not efficient and does not easily provide the quantitative information needed to rank the importance of process variables. Furthermore, the process chemist is often faced with a task of quickly scaling up the synthesis without the benefit of completely understanding the mechanism of the chemistry involved. With the aid of statistical design approach, a great deal of

useful information can be obtained with relatively few experiments. Statistical approaches involving factorial designs are ideal for studying processes where the underlying principles are not well developed or are extremely complicated.

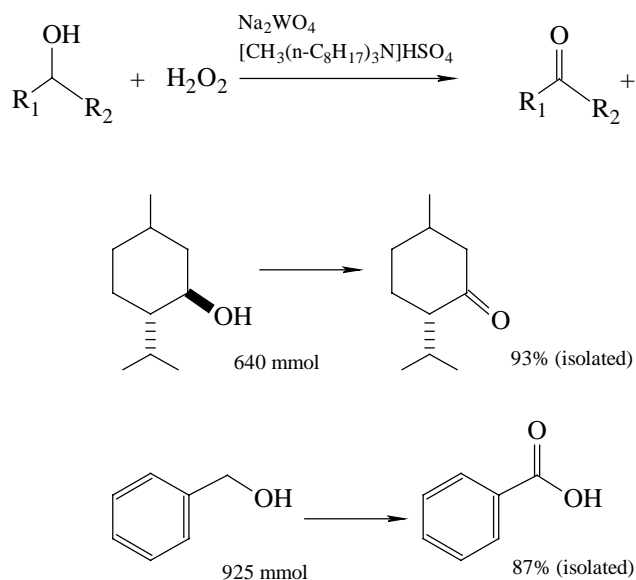
Factorial experiments consists of a systematic variation of two or more process variables at a time. For a two-level experiment, each variable is set to a high or a low value according to a standard pattern. An experimental run is conducted for each possible combination of variable settings. Selection of the low and high levels for each variable is important for obtaining meaningful results. If the levels selected are too close together, the calculated effect could be no larger than the experimental noise. Selecting widely separated levels could result in running the reaction under unrealistic conditions, for example, above the solvent boiling point, or above the decomposition temperature for reagents and reactive intermediates.

Data analysis of factorial designs involves a comparison of the experimental responses at the high and low settings of each variable. The results can be plotted in several different ways to develop an understanding of the effect of changing two or more process variables at a time with regard to reaction yield and quality of the product.

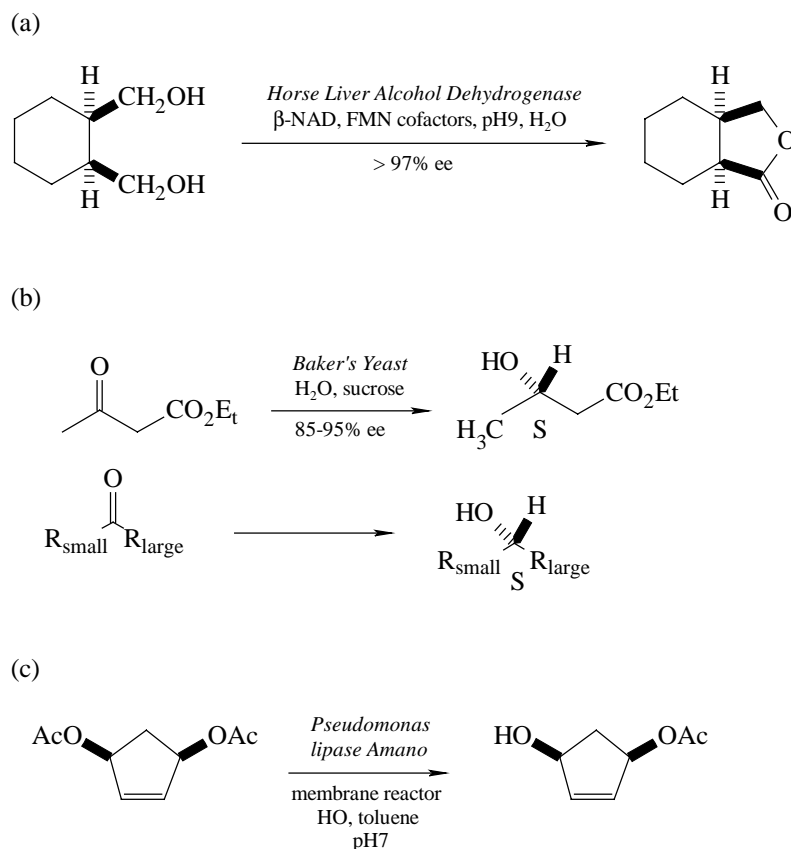
Drug Substance Crystal Form (Final Form)

The crystal form of the drug substance (15) is important for pharmaceutical industries, where products are specified not by chemical composition but by their performance. Good crystallinity, good bioavailability, satisfactory aqueous solubility and dissolution rate, and satisfactory physicochemical properties, such as stability, hygroscopicity, and flowability are required. Choosing the appropriate crystal form (final form) often involves crystallizing the drug substance from various solvent systems to search for polymorphism and screening of various salts of the drug substance (if chemically possible).

The chemical, biological, and physical characteristics of the drug substance can be manipulated and hence optimized by conversion to a salt form. Every compound that exhibits acid or base characteristics can participate in salt formation. Various salts of the same compound often behave quite differently because of the physical, chemical, and thermodynamic properties they impart to the parent compound. Table 1 lists the top 10 FDA-approved commercially marketed final drug forms, and Table 2 lists the top 10 salts that are not approved by the FDA but that are in use in other countries. Only salts of organic compounds have been considered here because most drugs are organic substances. The relative frequency with which each salt



Scheme 8 Noyori's oxidation of alcohols.



Scheme 9 Enzyme-mediated chemical transformations. (a) Enantioselective enzymatic oxidation and lactonization; (b) enzyme reduction with baker's yeast and enantioselective rule; and (c) enzymatic hydrolytic desymmetrization.

type has been used is calculated as a percentage, based on the total number of anionic or cationic salts in use.

The salt form is known to influence a number of physicochemical properties of the parent compound, such as dissolution rate, solubility, stability, and hygroscopicity. These properties, in turn, influence the absorption, distribution, metabolism, and excretion of the drug. This knowledge is essential for a complete understanding of the onset and duration of action, the relative toxicity, and the possible routes of administration. For example, certain

salts of the strong base choline have proved to be considerably less toxic than their parent compound. This observation led to the preparation of choline salicylate as an attempt to reduce the gastrointestinal (GI) disturbances associated with salicylate administration. Clinical studies indicated that choline salicylate elicited a lower incidence of GI distress, was tolerated in higher doses, and was of greater benefit to the patient than was acetylsalicylic acid (aspirin).

The chemical and physical stability of a drug can influence the choice of dosage form, the manufacturing and packaging, and the therapeutic efficacy of the final preparation. Systematic determination of the thermal stability, solution stability (at various pHs), and light sensitivity of a drug and its derivatives, both alone and in the presence of additives (excipients), provides essential information toward selecting the most suitable salt and dosage forms.

Depending on the mechanism of degradation, different salt forms impart different stability characteristics to the parent drug. Sparingly soluble salts used in the formulation

Table 1 Salts approved by the FDA

Anion	Percent	Cation	Percent
Hydrochloride	43	Sodium	62
Sulfate	7.5	Potassium	11
Bromide	5	Calcium	10.5
Chloride	4	Zinc	3
Tartrate	3.5	Meglumine	2.3

Table 2 Salts not approved by the FDA

Anion	Percent	Cation	Percent
Glycerophosphate	0.88	Piperazine	0.98
Aminosalicylate	0.25	Bismuth	0.98
Aspartate	0.25	Diethylamine	0.33
Bisulfate	0.25	Tromethamine	0.33
Hydroiodide	0.25	Barium	0.33

of suspensions reduce the amount of drug in solution and hence its degradation. Differences in hygroscopicity of several salts influence the stability of the drug in the dry state. For example, the stability of penicillin G and its salts has been widely studied because of the drug's therapeutic importance and its characteristic instability.

A solution of penicillin is not stable beyond two weeks even at refrigerator temperatures. However, the use of suspensions of sparingly soluble amine salts (procaine and hydrabamine salts) in aqueous vehicles allowed marketing of a "readymade" penicillin product.

Crystallization of the drug substance and its salt forms

In academia, a synthetic organic chemist rarely thinks about isolating a compound by crystallization, unless a single crystal structure is required. Most of the time, the chemist depends upon column chromatography to purify the compound. In contrast, in the pharmaceutical industry much depends upon crystallization. For example, crystal morphology and particle size have a direct impact on the filtration of a crystal slurry, cake compressibility, bulk stability and dissolution, bulk density, and flow characteristics. Crystallization is the preferred way of isolating the product from a reaction mixture. Knowledge of various crystallization systems from which the drug substance could be crystallized can provide information on polymorphism, thus expediting the selection of final crystal form. In polymorphism, two crystal forms with the same molecular structure are distinguished by the way in which the molecules are packed within the crystal lattice; each form has distinct physical and thermodynamic properties. In 1998, Abbott Laboratories withdrew its HIV drug, Ritonavir, because of the unexpected appearance of a new crystal form that had different dissolution and absorption characteristics than the standard product (16).

Crystallization is generally preceded by two types of nucleation. The primary nucleation occurs with the formation of clusters of molecules at the submicron level. When the concentration exceeds saturation to afford supersaturation, the clusters become nuclei. The secondary

nucleation is caused by particles due to primary nucleation or seeds. There are many strategies to achieve supersaturation to initiate crystallization such as cooling, evaporation, and antisolvent addition.

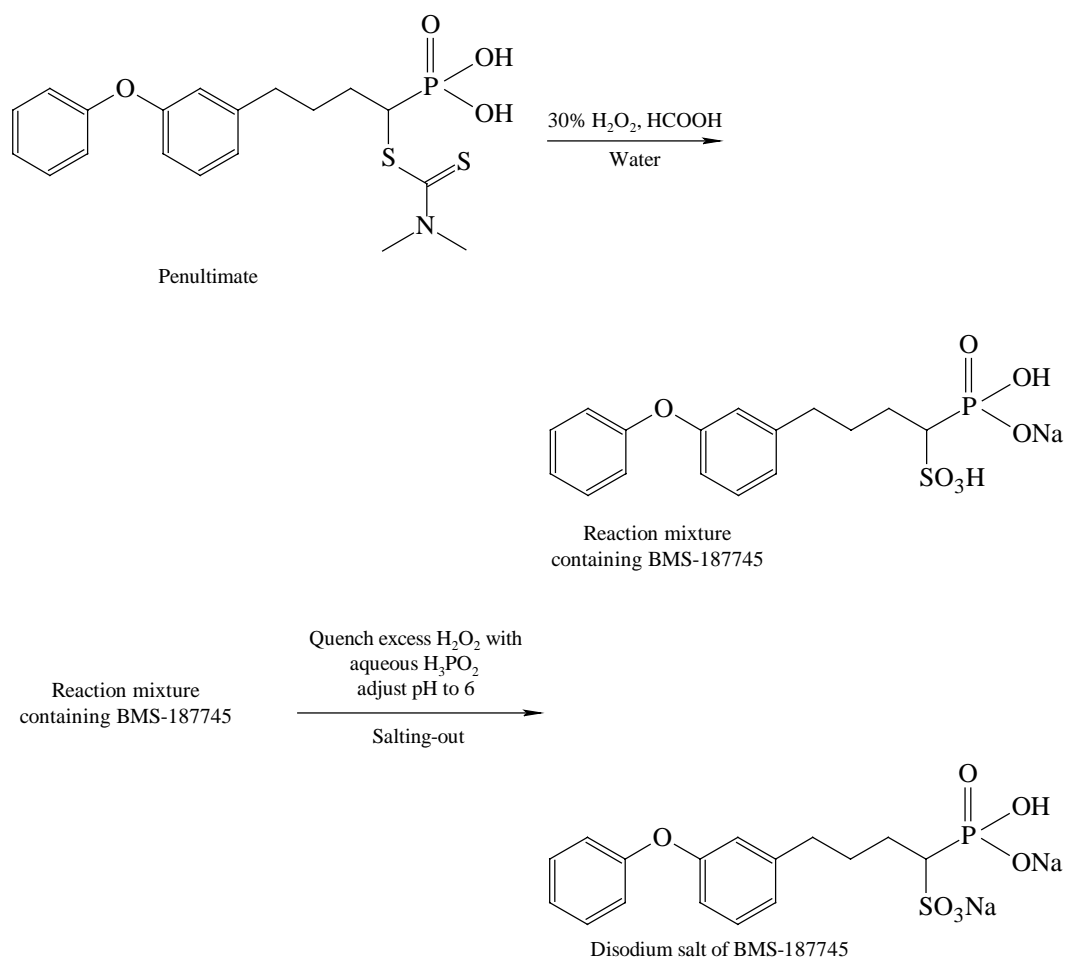
Early on in the drug development program, only small amounts of material are available for crystallization studies. Parallel crystallization technique in test tubes allows for the identification of many solvent systems using small amounts of material. On a small scale, it is not easy to control the rate of cooling or the rate of evaporation to achieve supersaturation. However, the antisolvent addition strategy to achieve supersaturation in combination with seeding allows rapid identification of several crystallization systems using a minimum amount of compound.

Salts of a substrate having an acid or an amine functionality are prepared by dissolving the substrate and the counter-base (or acid) separately in appropriate solvents, mixing the two solutions in equimolar ratios, and removing the solvents under vacuum to afford a solid. This method has two distinct advantages: First, the composition of the substrate remains the same during various crystallization attempts, and second, this portion of the work can be automated with the use of liquid handling systems.

The next step is to determine the solubility of the substrate (or its salts) in different solvents. This can also be performed by an automated liquid handling system. Depending upon the solubility of the substrate in water-miscible solvents (alcohols, acetone, tetrahydrofuran, etc.) and water-immiscible solvents (ethyl acetate, methyl-*tert*-butyl ether, heptane, etc.), the process chemist can identify one or many solvent systems from which the substrate (or its salts) could be crystallized using the antisolvent addition strategy.

Another crystallization technique is used when the isolation of a highly water-soluble compound in its salt form is required from aqueous reaction mixtures. This technique takes advantage of the common-ion effect and is based on the le Chatelier's principle, which states that, "if, to a system in equilibrium, a stress is applied, the system will react so as to relieve the stress." Thus, in aqueous solutions, the solubility of the compound in salt form can be reduced by adding large amounts of a common ion that is more soluble than the salt of the compound.

Example: BMS-187745 is a potent inhibitor of squalene synthase and an efficient cholesterol-lowering agent in orally dosed animals. The final step of the synthesis involved oxidation of the penultimate in water with hydrogen peroxide and formic acid. The disodium salt was chosen as the final crystal form. Because of



Scheme 10 BMS-187745 crystallization by salting-out.

extremely high solubility of the disodium salt in water, it cannot be extracted from water by common organic solvents. However, by employing the concept of the common-ion effect, the disodium salt can be easily crystallized from water. To accomplish this, the pH of the reaction mixture, after quench with hypophosphorous acid, is adjusted to the desired range of 6.05–6.25 (pKa of disodium salt of BMS-187745 is 6). In this pH range, sodium salts of formic, hypophosphorous, and phosphoric acids are present in high enough concentration to help lower the solubility of the disodium salt of BMS-187745 and effect crystallization (Scheme 10).

Automation in Process Chemistry Laboratories

Since the 1960s, automation has been the major tool for dramatically improving productivity (17–21). Automation techniques were introduced on the plant floor to

improve quality, increase safety, and streamline the work. Significant savings were realized initially from shorter delivery times, lower unit costs, smaller inventories, and fewer product failures. Due to an escalating cost of bringing a drug to market, companies are under immense pressure to shorten development time. The pharmaceutical industry has embraced new technologies in its discovery programs to increase throughput, generate precise data, and ensure accuracy.

Drug discovery scientists have already adopted tools such as combinatorial chemistry synthesizers; robotic systems for HTS; and software packages for computational chemistry, molecular modeling, and design of experiments to identify lead compounds (“hits”). As the number of “hits” grows, there is a potential for process R&D to become the “bottleneck.” The number of compounds that can enter process development is limited by the number of process chemists available to work on these compounds. Without some type of automated process development

machines to increase throughput, it will not be feasible to evaluate all the potential "hits." Created by the results from the HTS, demand for high-throughput development (HTD) tools is growing. Various companies, who in the past may have only addressed the needs of drug discovery scientists, are in the process of inventing and developing instruments that can be used by process chemists to perform many experiments in parallel. Examples of commercially available HTD tools are described below.

Parallel reactions

Charybdis Technologies: In the Calypso Reaction BlockTM, multi-well reaction arrays are designed for both solution and solid-phase synthesis applications. Well volumes range from 2 to 10 ml, pressures up to 207 kPa (30 psi), and temperatures from -80 to +180°C.

Quest Synthesis Technology: Quest 210 is designed for up to 20 reactions in 5- or 10-ml reaction vessels. Each of the 20 vessels can be efficiently stirred, heated, or cooled, and maintained under an inert atmosphere. Quest 205 is designed for synthesis on a large scale, with two banks of five 100-ml reaction vessels.

Argonaut Technologies: EndeavorTM allows parallel reactions under pressures up to 3.3 MPa (33 atm) and temperatures up to 200°C. Each of the eight vessels (working volume 15 ml) can be independently controlled for temperature, pressure, and gas delivery.

SurveyorTM: Suitable for parallel process development and optimization with online sampling and integrated HPLC analysis, it employs 10 reaction vessels (working volume 15–45 ml), with individually controlled reaction temperatures from -40° to +150°C, with the ability to reflux. Reagent addition, reaction parameter control, sampling, and HPLC injection are controlled by built-in software.

Bohdan Automation, Inc: The process development workstation can run up to 12 independent reactions with working volumes of 25 ml. The following operations are automatically carried out: reagent preparation and addition, individual heating, cooling, and mixing, reaction sampling and quenching, and transfer to optional HPLC or FTIR modules.

Parallel purification

Isco, Inc: The CombiFlashTM Sg 100c System is suitable for purifying 10–35 g of material. It provides time- or peak-based fraction collection with on-line UV detection and a linear and/or step gradient with two solvents at a flow rate of 10–100 ml/min.

The CombiFlash Si 10x System simultaneously purifies up to 10 samples, with a linear and/or step gradient with two solvents with a total flow rate of 10–100 ml.

Automated sample handling and analysis

Bohdan Automation, Inc: The Balance AutomatorTM offers a cost-effective alternative to manual weighing, reduces errors and operator tedium, and processes up to 120 samples per hour. This system can accommodate various container sizes and balances with the help of interchangeable parts.

Gilson: The Gilson 215 Liquid Handler is a versatile, large-capacity, septum-piercing liquid handler for safe and efficient transfer. It helps in transferring samples to other analytical systems or inject samples directly to an on-line HPLC system.

OUTSOURCING

Outsourcing covers a broad range of services (22). It can eliminate the need for additional staff, facilities, and/or equipment. Outsourcing in drug discovery and development is expected to continue its remarkable growth over the next decade as drug synthesis becomes more complex. This dynamic growth is due to a number of factors, including the ongoing consolidation in the pharmaceutical industry and a tight labor market. As larger companies consolidate, they seek synergies resulting in release of R&D personnel. Retrospectively, companies need to get more done with fewer internal resources and this promotes outsourcing. The tight labor market has made it increasingly difficult to recruit and retain staff, and many companies are turning to outsourcing to gain access to highly trained scientists who are available. Pharmaceutical companies are also looking into long-term relationships with outsourcing partners.

FDA GUIDELINES AND REGULATORY ISSUES

The guidelines on drug preparation (GLP and cGMP compliance) regulatory issues are described in details in the Center for Drug Evaluation and Research (CDER) by the FDA (23). A complete section is dedicated to API and the GMP issues. The readers are encouraged to seek the reference for further details.

REFERENCES

1. Picano, G.P.; Wheelwright, S.C. The New Logic of High-Tech R&D. *Harvard Business Review* (Sept./Oct. 1995), 93.
2. Gadamasetti, K.G., Ed.; *Process Chemistry in the Pharmaceutical Industry*; Marcel Dekker, Inc.: New York, 1999.

3. Mueller, R.H. A Practical Synthesis of Ifetroban Sodium. *Process Chemistry in the Pharmaceutical Industry*; Gadamasetti, K.G., Ed.; Chap. 3; Marcel Dekker, Inc.: New York, 1999; 37–55.
4. Holton, R.A. Method for Preparation of Taxol. Eur. Pat. Appl. **1990**.
5. Swaminathan, S.; Singh, A.K.; et al. Tetrahedron Lett. **1998**, 39, 4769.
6. Federsel, H.-J.; Jakupovic, E. A Highly Optimized and Concise Large Scale Synthesis of Purive Bronchodilator. *Process Chemistry in the Pharmaceutical Industry*; Gadamasetti, K.G., Ed.; Chap. 6; Marcel Dekker, Inc.: New York, 1999; 91–106.
7. *Guidelines for Chemical Reactivity Evaluation and Application to Process Design*; Center for Chemical Process Safety of the American Institute of Chemical Engineers: New York, 1995.
8. Breslow, R. The Greening Chemistry. Chem. & Eng. News **Aug.**, **1996**.
9. Sato, K.; Akoki, M.; Takagi, J.; Noyori, R. J. Am. Chem. Soc. **1997**, 119, 12386.
10. Azerad, R. Bull. Soc. Chim. Fr. **1995**, 132, 17.
11. Theil, F. Chem. Rev. **1995**, 95, 2203.
12. Wong, C.-H.; Whitesides, G.M. *Enzymes in Synthetic Organic Chemistry*; Pergamon: Oxford, 1994.
13. Collins, A.N.; Sheldrake, G.N.; Crosby, J., Eds. *Chirality in Industry*; John Wiley & Sons: New York, 1992, 1997; I and II.
14. Pilipauskas, D.R. Using Factorial Experiments in the Development of Process Chemistry. *Process Chemistry in the Pharmaceutical Industry*; Gadamasetti, K.G., Ed.; Chap. 22; Marcel Dekker, Inc.: New York, 1999; 411–428.
15. Berge, S.M.; Bighley, L.D.; Monkhouse, D.C. Pharmaceutical Salts. J. Pharm. Sci. **1977**, 66 (1), 1.
16. Blagden, N.; Davey, R. Polymorphs Take Shape. Chem. in Brit. **March 1999**, 44.
17. Owen, M.; Dewitt, S. Laboratory Automation in Chemical Development. *Process Chemistry in the Pharmaceutical Industry*; Gadamasetti, K.G., Ed.; Marcel Dekker, Inc.: New York, 1999; 429–455.
18. Conner, K. The Drive to Improve the Bottom Line. Today's Chemist at Work **Nov 1999**, 29.
19. Sullivan, M. Automation Accelerates Synthesis. Today's Chemist at Work **Sept 1999**, 48.
20. Studt, T. Raising the Bar on Combinatorial Discovery. Drug Discovery Development **Jan/Feb 2000**, 24.
21. Harness, J.R. Automated Sample Handling Supports Synthesis and Screening. Drug Discovery & Development **Jan 1999**, 69.
22. Waring, J. Strategic Outsourcing Fuels Research Growth. Drug Discovery & Development **Jan/Feb 2000**, 40.
23. CDER web site: <http://www.fda.gov/cder>. A Complete section is Dedicated to the Internationally Harmonised Guide for Active Pharmaceutical Ingredients (API) and Good Manufacturing Practice (GMP); Washington, Sept 1997.

PRODRUG DESIGN

Bernard Testa
Joachim Mayer

University of Lausanne, Lausanne, Switzerland

A CASE FOR PRODRUG DESIGN

Drug disposition and metabolism are of essential significance in pharmaceutical research because of the interdependence of pharmacokinetic and pharmacodynamic processes. Limited intestinal absorption, inadequate distribution, fast metabolism, and toxic metabolites are some of the causes of failure of drug candidates during development. To reduce the rate of attrition resulting from such pharmacokinetic defects, disposition and metabolic studies should be initiated as early as possible in the screening of lead candidates.

Avoidance of the foreseeable or proven pharmacokinetic defects thus assumes considerable significance in drug research. However, pharmacokinetic (PK) and pharmacodynamic (PD) optimization may not be compatible, meaning that efficacy at the target may be decreased or lost during PK optimization (1). A telling example of such a situation is provided by the novel drug class of neuraminidase inhibitors of therapeutic value against type A and B influenza in humans. Here, target-oriented rational design has led to highly hydrophilic, poorly absorbed agents such as Ro-64-0802, which shows very high in vitro inhibitory efficacy toward the enzyme but low oral bioavailability because of its high polarity (2). To circumvent this problem, Ro-64-0802 is marketed as Oseltamivir, its ethyl ester prodrug (Fig. 1). After intestinal absorption, the prodrug undergoes rapid enzymatic hydrolysis and produces high and sustained plasma levels of the active agent. As demonstrated by this example, the prodrug concept may thus prove to be a valuable alternative to disentangle PK and PD optimization. In other words, rather than attempting to improve lead candidates within a unitary rational design process, it may be that PK optimization can be achieved by the application of the prodrug concept to research compounds with high in vitro activity. Some of the concepts of interest in prodrug design are examined here.

OBJECTIVES AND PRINCIPLES OF PRODRUG DESIGN

Prodrugs are defined as therapeutic agents that are inactive per se but are predictably transformed into active metabolites (3, 4). As such, prodrugs must be contrasted with soft drugs, which are active per se and yield inactive metabolites.

Prodrug design aims at overcoming a number of barriers to a drug's usefulness (Table 1). The major objectives of prodrug design derive from these considerations and are also listed in Table 1.

Many successes have been recorded in prodrug design, and a large variety of such compounds have proven their therapeutic value. Several complementary viewpoints can be adopted when addressing prodrugs, namely, their chemical classification, their mechanism of activation (i.e., enzymatic and/or nonenzymatic), their tissue selectivity, the possible production of toxic metabolites, and the gain in therapeutic benefit (Table 2).

The chemical classification distinguishes between carrier-linked prodrugs (drugs linked to a carrier moiety by a labile bridge) and bioprecursors, which do not contain a carrier group and are activated by the metabolic creation of a functional group (4). In the former, the carrier moiety is often and conveniently linked to polar groups such as —OH, —NHR, and —COOH. Relevant examples of bioprecursors are provided by chemotherapeutic agents whose activation occurs by reduction in oxygen-deprived cells. Thus, the one-electron reduction of 3-amino-1,2,4-benzotriazine 1,4-dioxide to a cytotoxic nitroxide is believed to account for the antitumor activity of this bioprecursor (5). Such bioprecursors appear as a viable class of prodrugs because they avoid potential toxicity problems caused by the carrier moiety. However, attention must be given here to metabolic intermediates. A special group of carrier-linked prodrugs are the site-specific chemical delivery systems (6). Macromolecular prodrugs are synthetic conjugates of drugs covalently bound (either directly or via a spacer) to proteins, polypeptides, polysaccharides, and other biodegradable

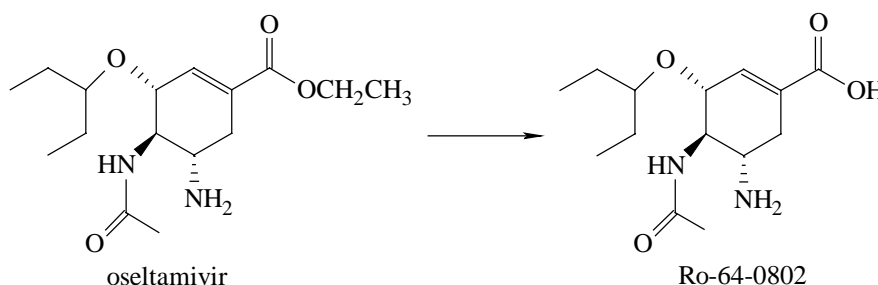


Fig. 1 The structure of the neuraminidase inhibitor Ro-64-0802 and its ethyl ester prodrug Oseltamivir.

polymers (7). A special case is provided by drugs coupled to monoclonal antibodies.

Prodrug activation occurs enzymatically, nonenzymatically, or sequentially (enzymatic step followed by nonenzymatic rearrangement). As much as possible, it is desirable to reduce biological variability, hence, the particular interest currently received by nonenzymatic reactions of intramolecular cyclization–elimination (8) presented at length below. The problem of tissue or organ targeting is another important aspect of prodrug design. Various attempts have been made to achieve organ-selective activation of prodrugs, such as dermal delivery (9) and brain penetration (6). For example, the selective presence of cysteine conjugate β -lyase in the kidney suggests that this enzyme might be exploited for delivery of sulfhydryl drugs to this organ (10).

The toxic potential of metabolic intermediates, of the carrier moiety, or of a fragment thereof must also be kept in mind. This is illustrated by formaldehyde-releasing prodrugs such as *N*- and *O*-acyloxymethyl derivatives or Mannich bases. Similarly, arylacetylenes assayed as potential bioprecursors of antiinflammatory arylacetic

acids proved many years ago to be highly toxic because of the formation of an intermediate ketene. The gain in therapeutic benefit provided by prodrugs is an issue with no general conclusion. Depending on both the drug and its prodrug, the therapeutic gain may be modest, marked, or even significant. In the case of marketed drugs endowed with useful qualities but displaying some unwanted property, the expected therapeutic gain is usually modest to marked. In the case of difficult candidates showing excellent target properties but suffering from some severe physicochemical and/or pharmacokinetic drawback, a marked to significant benefit can be obtained.

CHEMICALLY ACTIVATED PRODRUGS: REACTIONS OF CYCLIZATION–ELIMINATION

So many biological factors may affect enzymatic reactions that the resulting interspecies and interindividual variability renders prodrug design unreliable and sometimes even problematic. To circumvent such difficulties, an increasing

Table 1 Barriers to drug usefulness and corresponding objectives of prodrug design

Pharmaceutical barriers	Pharmaceutical objectives
Insufficient chemical stability	Improved formulation (e.g., increased hydrosolubility)
Poor solubility	Improved chemical stability
Offensive taste or odor	Improved patient acceptance and compliance
Irritation or pain	
Pharmacokinetic barriers	Pharmacokinetic objectives
Low oral absorption	Improved bioavailability
Marked presystemic metabolism	Prolonged duration of action
Short duration of action	Improved organ selectivity
Unfavourable distribution in the body	
Pharmacodynamic barriers	Pharmacodynamic objectives
Toxicity	Decreased side effects

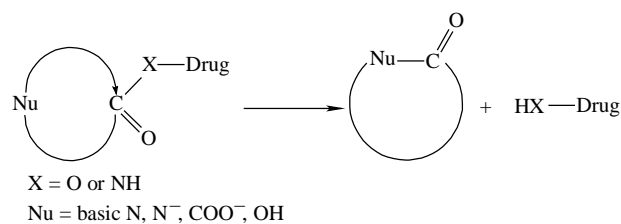
(From Ref. 3.)

Table 2 Complementary viewpoints in prodrug design

Chemical classification
Bioprecursors
Classical carrier-linked prodrugs
Site-specific chemical delivery systems
Macromolecular prodrugs
Drug–antibody conjugates
Mechanisms of activation
Enzymatic
⇒ biological variability
Nonenzymatic
⇒ no biological variability
Mechanisms of tissue/organ selectivity
Tissue-selective activation of classic prodrugs
Site-specific delivery of ad hoc chemical systems
Potential toxicity
Of a metabolic intermediate
Of the carrier moiety or a metabolite thereof?
Gain in therapeutic benefit
Prodrugs of marketed drugs (post hoc design); modest to marked benefit
Prodrugs of difficult candidates (ad hoc design); marked to significant benefit

number of studies have proposed and investigated prodrugs activated by a purely or predominantly nonenzymatic mechanism. Prodrug of this type include:

- (2-oxo-1,3-dioxol-4-yl)methyl esters
- Mannich bases
- oxazolidines
- esters with a basic side chain sterically able to catalyze intramolecular hydrolysis

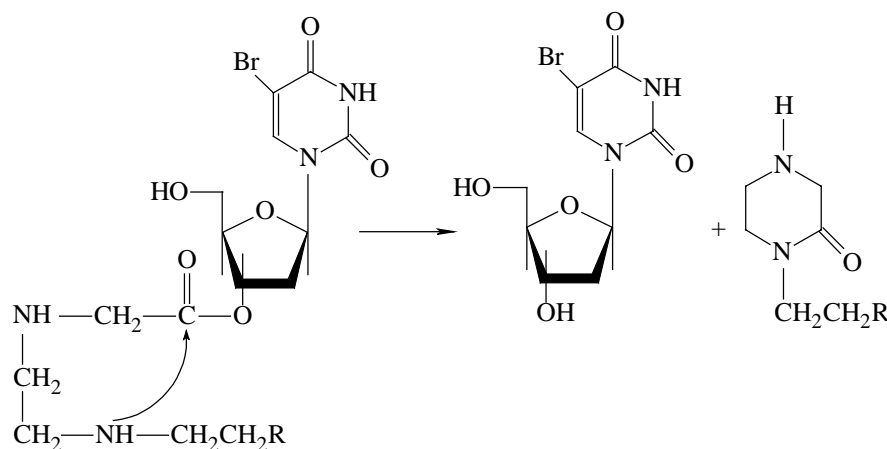
**Fig. 2** General reaction scheme for the intramolecular activation of prodrugs by cyclization–elimination. (Modified from Ref. 11.)

- esters and amides undergoing intramolecular nucleophilic cyclization–elimination.

Here, we focus on prodrug activation by intramolecular cyclization–elimination (8, 11). These reactions occur in specifically designed prodrugs of phenols, alcohols, and amines. A number of design strategies exist to achieve such mechanisms, as illustrated below with selected examples. The general chemical principle of these reactions is shown in Fig. 2. In such a schematic representation, the carrier moiety is a side chain attached to the drug by a carbonyl group (i.e., by an ester or amide function) and containing a nucleophilic group symbolized by “Nu.” The latter directly attacks the carbonyl in a reaction of nucleophilic substitution whose outcome is cyclization of the carrier moiety and elimination of the drug molecule.

Cyclization–Elimination Due to a Basic Amino Group

Nucleophilic attack by a basic amino group has proven to be a useful strategy to achieve and modulate cyclization–elimination. The radiation sensitizer

**Fig. 3** Activation of basic ester prodrugs of 5-bromo-2'-deoxyuridine by cyclization of the promoiety. (Modified from Ref. 12.)

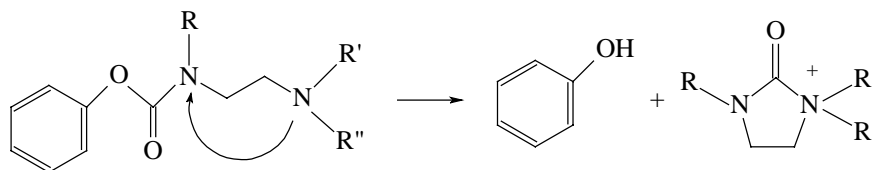


Fig. 4 Activation of basic carbamates of phenols by cyclization–elimination of the promoity. (From Ref. 13.)

5-bromo-2'-deoxyuridine was derivatized with diamino acids to obtain the prodrugs shown in Fig. 3, with $R = H$ or cyclohexyl (12). The reaction of cyclization proceeded cleanly to yield the drug and a piperazinone derivative without any other detectable product being formed. No hydrolysis was seen in acidic solutions. At pH 7.4 and 37°C in a buffer solution, the half-life for the two prodrugs was 23 and 30 min, respectively. In human plasma under the same conditions, the values were 70 and 47 min, respectively, suggesting protection from breakdown by binding to proteins. In rat plasma, the half-life of the first compound was 47 min, again suggesting protection, but it was 5 min for the second compound. This is an indication that enzymatic hydrolysis by rat plasma hydrolases is possible in some cases.

Thus, intramolecular activation by cyclization–elimination was modulated by steric factors. In addition, this example shows that hydrolysis may be catalyzed by enzymes depending on substrates and biological conditions.

Valuable insights can be found in an informative study on the reactivity of phenyl carbamates of ethylenediamines (Fig. 4) (13). Their half-lives of chemical activation at pH 7.4 and 37°C showed that N,N' -dimethyl and N,N',N' -trimethyl substitution gave the fastest rates of cyclization, whereas N - H , N' - H , and/or N -ethyl groups markedly decelerated the reaction. The other major interest of this study is the proof it reported that no enzymatic hydrolysis occurred. Indeed, these compounds were incubated with human plasma, pig liver homogenates, and rat liver homogenates, the half-lives of release of

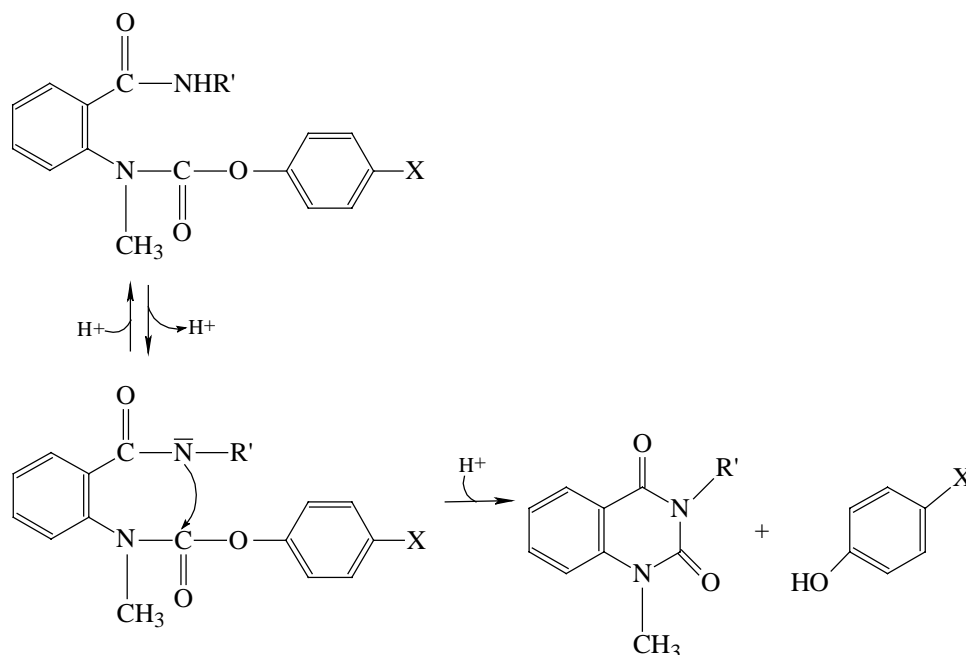


Fig. 5 Simplified reaction mechanism of intramolecular cyclization–elimination of phenyl carbamates of anthranilamides. (From Ref. 14.)

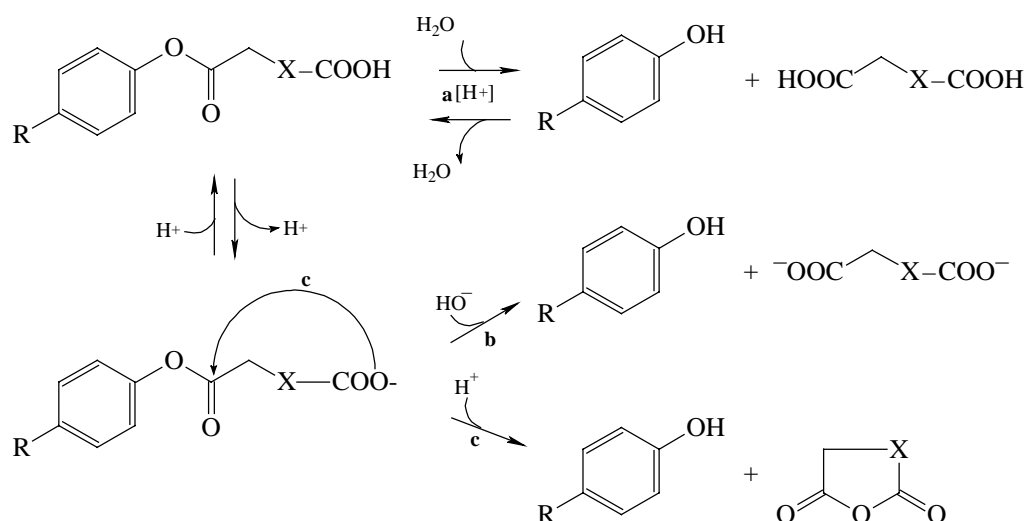


Fig. 6 Activation of hemiester prodrugs of phenols by proton-catalyzed hydrolysis (reaction a), hydroxyl-catalyzed hydrolysis (reaction b), or cyclization–elimination (reaction c). Enzymatic hydrolysis is not represented. (Adapted from Ref. 15.)

phenol being either identical to those in buffer, or slightly larger, because of protein binding.

Cyclization–Elimination Due to an Acidic Amido Group

In addition to attack by a basic nitrogen, there exists also the possibility of an intramolecular attack by an anionic nitrogen, i.e., a deprotonated amido nitrogen. This is exemplified by *N*-(2-carbamoylphenyl)carbamates of model phenols (Fig. 5; X = H, Cl, or OCH₃) (14). In such promoieties, the deprotonated carboxamido group attacks the carbamate carbonyl to form a quinazolinone and release the phenol. The stability of a large series of *N*-(2-carbamoylphenyl)carbamates was explored in buffer solutions and in diluted human plasma. The rate of nonenzymatic cyclization–elimination was highly sensitive to the nature of the carboxamido substituent (R' in Fig. 5). An alkyl substituent larger than methyl strongly decreased reactivity, presumably by steric hindrance. In contrast, an electron-withdrawing

substituent increased reactivity by facilitating deprotonation. Globally, these effects were significant because the rates span four orders of magnitude.

In human plasma, two groups of compounds were seen. In most cases, the reaction in plasma (i.e., chemical plus enzymatic activation) was approximately two-fold faster than in buffer. This indicates that enzymatic hydrolysis in human plasma, if any, was modest at best. Only for the three prodrugs with an unsubstituted carboxamido group (R' = H) was the enzymatic reaction several-fold faster than was intramolecular catalysis, suggesting these compounds to be substrates of plasmatic hydrolases. Thus, the *N*-(2-carbamoylphenyl)carbamate promoiety allowed a highly modulatable intramolecular activation with little enzymatic activation.

Cyclization–Elimination Owing to a Carboxylate or Hydroxyl Group

Intramolecular cyclizations are not restricted to attack by a nucleophilic nitrogen (basic amino or acidic amido group).

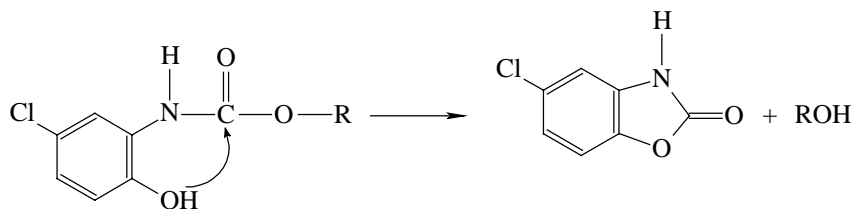


Fig. 7 Cyclization of *N*-2-hydroxyphenyl carbamates as potential prodrugs of benzoxazoles and phenols. (From Ref. 16.)

Table 3 Specific difficulties in prodrug design and development

Objectives	Strategy	Restrictive conditions
Minimize the number of proposed candidates	Careful prodrug design based on prediction of target phys-chem and PK properties	Available
Maximize the explored space of phys-chem and PK properties		Local quantitative models Global qualitative models (e.g., “rule of 5”)Dubious Global quantitative models
Minimize the additional synthetic work		
Reach target phys-chem profile	Careful weighing of costs and benefits	Limited acceptance of additional costs and efforts
Reach target PK profile	HTP phys-chem profiling Virtual screening	Limited or dubious relevance of some HTP techniques Dubious predictive capacity of some models (see above)
Reach target metabolic behavior, in particular target rate of activation	HTP PK profiling Virtual screening	Limited or dubious in vivo relevance of some HTP screens Dubious predictive capacity of some models (see above)
Lack of toxicity of prodrug, promoiety or fragment thereof	Fast in vitro metabolic assessment	Dubious extrapolation to in vivo situation
	Knowledge-based design Toxicity screens	Incomplete knowledge Limited or dubious in vivo relevance of screens

HTP = high throughput; phys-chem = physicochemical; PK = pharmacokinetic.

They can also be catalyzed by a nucleophilic oxygen as found in a carboxylate, phenolic, or alcoholic group. Illustration of the catalytic role of a carboxylate group can be found in hemiester prodrugs of phenol (taken as model compound) or paracetamol (Fig. 6; R = H or NHCOCH_3 , respectively) (15). In addition to enzymatic hydrolysis, three mechanisms of chemical hydrolysis were seen, namely, acid-catalyzed, base-catalyzed, and an intramolecular nucleophilic attack, resulting in cyclization–elimination (Fig. 6, reactions a, b, and c, respectively). In buffer solutions, the relative importance of these three pathways was clearly pH-dependent. At physiological pH, cyclization–elimination was the predominant reaction, with half-lives ranging from 1 to 350 min at 37°C. Reactivity at this pH was markedly influenced by the length and degree of substitution of the promoiety and by the pK_a value of the phenol. Thus, succinate esters (Fig. 6; X = CH_2CH_2) were approximately 150 times more reactive than were glutarate esters (X = $\text{CH}_2\text{CH}_2\text{CH}_2$); C-methylation of the promoiety (X = $\text{CH}_2\text{CH}(\text{CH}_3)\text{CH}_2$) also increased reactivity. Esters of paracetamol (R = NHCOCH_3) were degraded about twice as fast as were esters of phenol (R = H).

In most cases, the hemiesters in Fig. 6 underwent no or little enzymatic degradation in human plasma, in agreement with the known inertness of hemiesters toward cholinesterase. In contrast, very rapid hydrolysis was usually seen in pig and rat liver preparations, indicating the involvement of carboxylesterases. The only inert compound was the 3,3-dimethylglutarate hemiester of paracetamol (Fig. 6; X = $\text{CH}_2\text{C}(\text{CH}_3)_2\text{CH}_2$). It certainly would be interesting to have data on the hydrolysis of such prodrugs by human hepatic enzymes.

The hydroxyl functionality in phenols and alcohols is also of potential interest as an intramolecular nucleophile. One example of this type is presented here (16). *N*-2-hydroxyphenyl carbamates (Fig. 7) undergo intramolecular cyclization with quantitative liberation of a phenol or an alcohol. The interest of this approach is that the product of cyclization is itself a drug, namely, the skeletal muscle relaxant chlorzoxazone. A large variety of phenols and alcohols were investigated in this work, most of them model compounds, but paracetamol was among them. In this case, a prodrug of this type liberates not one but two drugs and can be called a mutual prodrug.

The mechanism of activation in Fig. 7 was demonstrated to be an intramolecular nucleophilic attack by the phenolate ion on the carbonyl group. As a consequence, the rate of reaction increased linearly with pH, up to a plateau at 8–9 and beyond. The major factor influencing the rate of reaction was the acidity of the leaving ROH molecule, with $t_{1/2}$ values at pH 10 and 25°C ranging from

290 days for ROH = $\text{CH}_3\text{CH}_2\text{OH}$ (pK_a 16.0) to 3–12 s for ROH = phenols of pK_a 9–10. The $t_{1/2}$ value for ROH = paracetamol was 7.1 s at pH 7.4 and 37°C (16). Some prodrugs and mutual prodrugs in Fig. 7 were also examined for their stability in human and rat plasma (16). These *N*-2-hydroxyphenyl carbamates showed two- to threefold increases in $t_{1/2}$ values in human and rat plasma compared with buffer. This indicates the absence of an enzymatic hydrolysis and a modest stabilization owing to binding to plasma proteins.

SOME PROBLEMS IN PRODRUG RESEARCH

The benefit brought forth by a prodrug relative to the active agent is worth considering. Schematically, it appears from innumerable data in the literature that the gain will be considerable when the development of an innovative and very promising agent is blocked by a major pharmacokinetic or pharmaceutical defect that a prodrug strategy can overcome. In contrast, the gain will be negligible when the drug defect is tolerable or barely improved by transformation to a prodrug. But what of the specific difficulties encountered in designing and developing prodrugs? Table 3 schematically presents some of these difficulties (17), to which one can add possible complications in registration.

Thus, it's no wonder so many medicinal chemists are critical of prodrugs. However, and this is our conclusion, a lucid view cannot ignore the potential benefit, in this case the mere existence of a number of successful prodrugs. Nabumetone, oseltamivir, and pivampicilline are just a few examples that come to mind. They demonstrate that in a number of cases, prodrug design may indeed allow the separate optimization of PK and PD properties.

REFERENCES

1. Testa, B.; Caldwell, J. Prodrugs Revisited—The Ad Hoc Approach as a Complement to Ligand Design. *Med. Res. Rev.* **1996**, *16*, 233–241.
2. Oxford, J.S.; Lambkin, R. Targeting Influenza Virus Neuraminidase—a New Strategy for Antiviral Therapy. *Drug Discov. Today* **1998**, *3*, 448–456.
3. Stella, V.J.; Charman, W.N.A.; Naringrekar, V.H. Prodrugs: Do They Have Advantages in Clinical Practice? *Drugs* **1985**, *29*, 455–473.
4. Wermuth, C.G. Designing Prodrugs and Bioprecursors. *Drug Design: Fact or Fantasy?*; Jolles, G., Woolridge, K.R.H., Eds.; Academic Press: London, 1984; 47–72.
5. Riley, R.J.; Workman, P. Enzymology of the Reduction of the Potent Benzotriazine-di-N-oxide Hypoxic Cell

- Cytotoxin SR 4233 (WIN 59075) by NAD(P)H: (Quinone Acceptor) Oxidoreductase (EC 1.6.99.2) Purified from Walker 256 Rat Tumour Cells. *Biochem. Pharmacol.* **1992**, *43*, 167–174.
6. Bodor, N. Redox Drug Delivery Systems for Targeting Drugs to the Brain. *Ann. N.Y. Acad. Sci.* **1987**, *507*, 289–306.
 7. Duncan, R. Drug-Polymer Conjugates: Potential for Improved Chemotherapy. *Anti-Cancer Drugs* **1992**, *3*, 175–210.
 8. Testa, B.; Mayer, J.M. Design of Intramolecularly Activated Prodrugs. *Drug Metab. Rev.* **1998**, *30*, 787–807.
 9. Chan, S.Y.; Li Wan, PoA. Prodrugs for Dermal Delivery. *Int. J. Pharmaceut.* **1989**, *55*, 1–16.
 10. Hwang, I.Y.; Elfarra, A.A. Cysteine S-Conjugates May Act as Kidney-Selective Prodrugs: Formation of 6-Mercaptopurine by the Renal Metabolism of S-(6-puriny)-L-Cysteine. *J. Pharmacol. Exp. Ther.* **1989**, *251*, 448–454.
 11. Shan, D.; Nicolaou, M.G.; Borchardt, R.T.; Wang, B. Prodrug Strategies Based on Intramolecular Cyclization Reactions. *J. Pharm. Sci.* **1997**, *86*, 765–767.
 12. Saari, W.S.; Schwering, J.E.; Lyle, P.A.; Smith, S.J.; Engelhardt, E.L. Cyclization-Activated Prodrugs. Basic Esters of 5-Bromo-2'-Deoxyuridine. *J. Med. Chem.* **1990**, *33*, 2590–2595.
 13. Thomsen, K.F.; Strøm, F.; Sforzini, B.V.; Begtrup, M.; Mørk, N. Evaluation of Phenyl Carbamates of Ethyl Diamines as Cyclization-Activated Prodrug Forms for Protecting Phenols Against First-Pass Metabolism. *Int. J. Pharma.* **1994**, *112*, 143–152.
 14. Thomsen, K.F.; Bundgaard, H. Cyclization-Activated Phenyl Carbamate Prodrug Forms for Protecting Phenols Against First-Pass Metabolism. *Int. J. Pharma.* **1993**, *91*, 39–49.
 15. Fredholt, K.; Mørk, N.; Begtrup, M. Hemiesters of Aliphatic Dicarboxylic Acids as Cyclization-Activated Prodrug Forms for Protecting Phenols Against First-Pass Metabolism. *Int. J. Pharma.* **1995**, *123*, 209–216.
 16. Vigroux, A.; Bergon, M.; Zedde, C. Cyclization-Activated Prodrugs: N-(Substituted 2-Hydroxyphenyl and 2-Hydroxypropyl)Carbamates Based on Ring-Opened Derivatives of Active Benzoxazolones and Oxazolidinones as Mutual Prodrugs of Acetaminophen. *J. Med. Chem.* **1995**, *38*, 3983–3994.
 17. Testa, B.; Mayer, J.M. Concepts in Prodrug Design to Overcome Pharmacokinetic Problems. *Pharmacokinetic Optimization in Drug Research—Biological, Physicochemical and Computational Strategies*; Testa, B., van de Waterbeemd, H., Folkers, G., Guy, R., Eds.; Wiley—VCH: Weinheim, in press.
 18. Balant, L.P.; Doelker, E. Metabolic Considerations in Prodrug Design. *Burger's Medicinal Chemistry and Drug Discovery*; 5th Ed., Wolff, M.E. Ed.; Wiley: New York, 1995; 1, 949–982.
 19. Borchardt, R.T.; Repta, A.J., Stella, V.J., Eds. *Directed Drug Delivery—A Multidisciplinary Approach*, Humana Press: Clifton NJ, 1985.
 20. Bundgaard, H.; Hansen, A.B., Kofod, H., Eds. *Optimization of Drug Delivery*, Munkgaard: Copenhagen, 1982.
 21. Bundgaard, H. *A Textbook of Drug Design and Development*; Krogsgaard-Larsen, P., Bundgaard, H., Eds.; Harwood: Reading, 1991; 113–191.
 22. Harper, N.J. Drug Latentiation. *Progress in Drug Research*; Jucker, E., Ed.; Birkhäuser: Basel, 1962; 4, 221–294.
 23. Testa, B.; Jenner, P.; *Drug Metabolism: Chemical and Biochemical Aspects*; Marcel Dekker, Inc.: New York, 1976.
 24. Testa, B.; *The Metabolism of Drugs and Other Xenobiotics—Biochemistry of Redox Reactions*; Academic Press: London, 1995.
 25. Testa, B. Drug Metabolism. *Burger's Medicinal Chemistry and Drug Discovery*; 5th Ed.; Wolff, M.E., Ed.; Wiley: New York, 1995; 1, 129–180..
 26. Testa, B.; Mayer, J.M.; *Hydrolysis in Drug and Prodrug Metabolism—The Biochemistry and Enzymology of Hydrolases*; Wiley—Verlag Helvetica Chimica Acta: Zurich, in preparation.
 27. Widder, K.J., Green, R., Eds. *Drug and Enzyme Targeting (Methods in Enzymology)*; Academic Press: Orlando, FL, 1985; 112.

PROJECT MANAGEMENT

Jerry J. Groens

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

Cara R. Frosch

Covance Central Laboratory Services, Inc., Indianapolis, Indiana, U.S.A.

INTRODUCTION

New products are developed and introduced in organizations primarily through the use of projects and teams. Research of new technologies and innovation is essentially accomplished through the use of projects as well. The processes for developing and introducing new products are certainly more structured and well understood when compared with research; however, the essence of each is that they are projects.

A project is a *temporary endeavor undertaken to create a unique product or service* (1). Project management provides the processes, tools, techniques, skills, and knowledge needed to manage a project such that the results meet or exceed the stakeholder needs and expectations. Stakeholders are the individuals or organizations whose interests may be positively or negatively impacted by the project.

Searching for sources of project management in the literature, an individual finds hundreds of books, articles, and cases covering the topic in all industries, environments, and situations. These sources emphasize that every project is different and that projects exist in an environment of change, uncertainty, and inconsistency. However, some of these sources then proceed to state that “project management is a science” and that the processes of project management are consistent across all industries and environments. Any successful practitioner in project management in the fields of product development and, more specifically pharmaceutical development, knows that this is not correct. Project management is more, much more of an art than a science, and the application of project management is not consistent across industries, environments, and situations.

For example, many of the elements of project management applied in the construction and telecommunications industries do not work in pharmaceutical development. This includes earned value, configuration management, and detailed representations of the work breakdown structures. These concepts actually provide too much structure when applied to pharmaceutical or any product development and

could potentially have a negative impact on innovation, creativity, and effective teaming. Therefore, the creation of a development framework, based upon the principles and practices of project management, is required to address the uncertain world of pharmaceutical development. This chapter will focus on that framework.

A PROJECT MANAGEMENT FRAMEWORK

A project management framework has six elements. These are:

- Project definition,
- Project team and organization,
- Project planning, scheduling, and control,
- Problem solving and decision making using prototypes,
- Senior management review and control, and
- Proactive, real-time change management (2).

Each of these items is key to the successful management of a new product development project. For example, the project definition processes develop:

- The project scope and objectives,
- The business case and the technologies utilized,
- The project strategy or direction being pursued,
- The sponsorship and championship for the new product
- Concept, and
- The final product performance targets and deliverables.

This project management chapter will cover each of these elements of a project management framework and provide sources for further study.

PROJECT DEFINITION

This first element of a project management framework is used to define project scope and objectives. It outlines what is within and not within the project boundaries,

the business case and market need, the technologies to be used to solve that business case, and the initiation, selling, and sponsoring of the project.

The scope of the project includes not only the definition of the final product deliverable but the level of work effort required to complete this project. For example, a new pharmaceutical drug will be used to reduce long-term intractable pain, will be in an injectable form, and will use existing packaging configurations in defined concentrations and formulations. This is an initial, general definition of the product scope. The project scope, however, is much more time consuming and complex to develop and would include:

- The number, type, and cost of clinical studies or trials (including number of patients and sites) required in each phase of development to complete the project;
- The formulation, analytical chemistry, packaging and sterilization/microbiological work needed; and
- The manufacturing processes, equipment, and other resources required to produce the product.

Resource estimates (hours and dollars) for people's time, expenses, and capital are projected along with an approximation of the timing of the need for these resources. This becomes the basis for a project approval document or *charter*, which also includes the product scope, project objectives, business case, and financial evaluation.

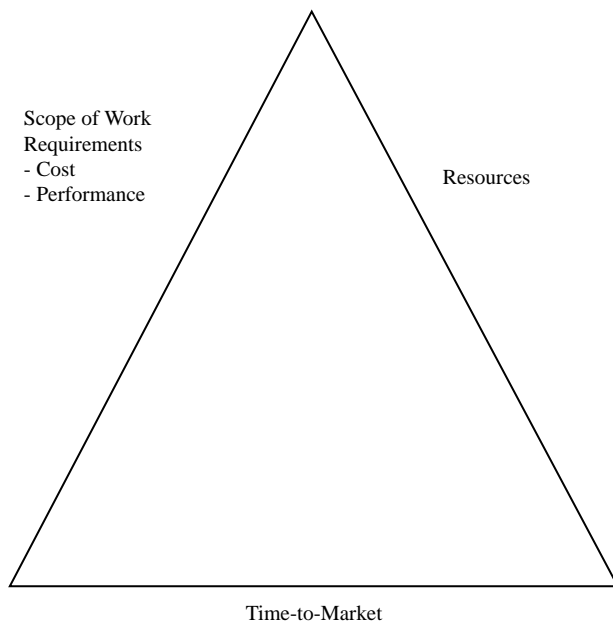


Fig. 1 New product projects objectives. (Adapted from Ref. 3.)

Project objectives are also a key element of project definition and set the direction for a product development project. These objectives operate in an environment where trade-offs are necessary, and are often referred to as the *triple constraint* (see Fig. 1). Looking at these three constraints, a few questions may come to mind. For example:

- How often have you been on a project where the scope did not change? Probably never, and if you have, this is very unusual.
- How often have you been able to estimate the original resources required on a new project as closely as $\pm 5\%$? Probably never, and if you have, this is very unusual.
- How often are you allowed as much time as needed to complete the project? Probably never and, in fact, cycle time reductions are the norm.

Given these experiences, and assuming that these are typical and true, a project manager recognizes the magnitude of the challenges in being successful because meeting these objectives is the measurement of success. This emphasizes the relevance of an old phrase that many project managers quote from time to time: “fast, cheap, or good; pick two.” This may be a catchy phrase, but it highlights a crucial dilemma in project management. To help solve this dilemma, the project manager needs to understand the expectations of the sponsor and key stakeholders and to periodically assess them throughout the project life cycle. The project manager may identify some “give” or slack on one and possibly more of these expectations depending on the project, with its unique uncertainties and challenges. In the pain management example, most likely a specific target level of pain reduction is crucial to successfully achieve the product (product performance) requirements. This will probably represent the area where changes in project scope and direction will most likely occur. Since getting the product to market quickly is a goal (time-to-market), providing sufficient funding (resources) to support these objectives is critical. So, where is the “give” in the objectives based upon the needs of the management stakeholders? Managing the triple constraint and the expectations of management with regards to these objectives requires the development and management of the project's business case with financials.

The business case for a new product project should include the following:

- An evaluation of the market and customer needs including appropriate market research studies,
- An analysis of the competition including their potential products and strengths and weaknesses,

- A technical assessment of the capability of the organization as it relates to the product competing in the market,
- A statement of alignment to the organization's business strategy, and
- A preliminary evaluation of the financial opportunity of the new product, which may include use of return on investment, discounted cash flow and internal rate of return, net present value, and break-even time (4).

Due to the uncertainty involved in the evaluation of new products, financial analysis tools that consider risks and opportunities are more appropriate and valuable than deterministic approaches. These new approaches to project financial evaluation that consider uncertainty include options analysis and Monte Carlo simulation. Due to their proactive handling of uncertainty, these tools can more accurately calculate the risks and opportunities of a new product concept (5). With the use of a financial analysis model, basic trade-off statements can be developed by the project manager to assist in understanding the importance of each objective. In the pain management product example, a statement emphasizing the value of time would be "a week delay in the project costs \$1 million in today's money" (6).

Of course, all of this project definition process and effort, hopefully, will lead to an approved and ultimately successful new pharmaceutical product. The need for a concurrent process of project selling and support is emphasized due to the nature of new product development, the uncertainty of the projects, high cost, and limited resources. To accomplish this task, a new product project requires two types of individuals early in its life cycle:

champions and sponsors. Project champions are individuals within the organization who support the project and sell it to management. Project sponsors are senior-level executives whose support is needed to fund and provide resources for the project. Of course, the more champions that a new product has selling the concept, the greater the potential for funding and support from senior-level executives. The level of support needed and number of senior executives increases as the funding requirements of the project increase. This is probably obvious, and yet not recognizing them can result in missing a key new product concept.

Early in the life of a new product, its future is tenuous and requires an approach that not only puts it through a rigorous business, technical, and financial evaluation but allows it to nurture and grow into a development project where appropriate. An organization should have an early new product process that: 1) captures and provides an opportunity for all the valuable new concepts that an organization is researching to be evaluated on a level playing field, and 2) focuses the organization on those key few projects that allow for fast product development. This concept, called the "funnel" (Fig. 2) in some sources (2), shows the need for the following elements of successful initial project definition:

- A proactive process of mining both externally and internally for new product concepts,
- Research for these new product concepts and funding for the same to assure that the concepts are developed to an appropriate level prior to evaluation, and
- A process to evaluate new product concepts to assure that the best concepts are fully funded based upon the criteria identified above under business case.

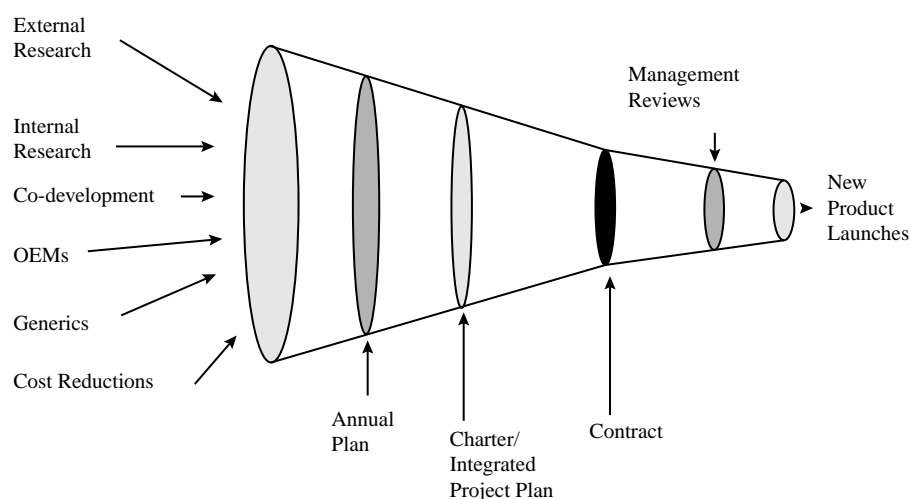


Fig. 2 New product "funnel."

The evaluation process, sometimes known as “stage gates” and “phased reviews”(7), needs to be accomplished early in the development project life and continue on throughout the life of a project as needed and defined in a *contract or integrated project plan*. The project plan will be reviewed later in this chapter along with its use in project reviews and stage gates.

PROJECT TEAM AND ORGANIZATION

A team is a small number of people with complementary skills who are committed to a common purpose, performance goals, and approach for which they hold themselves mutually accountable (8). Teams, however, are not necessarily the answer to all business endeavors, but the challenges of new pharmaceutical drug development lends itself very well to this organizational approach.

“Core” or primary team and support teams are used often to keep the number of team members small. A graphical illustration of this type of core and support team is shown in Fig. 3. Core team members represent the different functions of an organization, have significant authority and responsibility in the organization (especially for high priority projects), and have both specialist/technical and generalist/business skills. In the pain management example, the core team may include team members from:

- Program management,
- Marketing management,
- Clinical development,
- Formulation development,
- Analytical chemistry,
- Process development,
- Operations management, and
- Regulatory management.

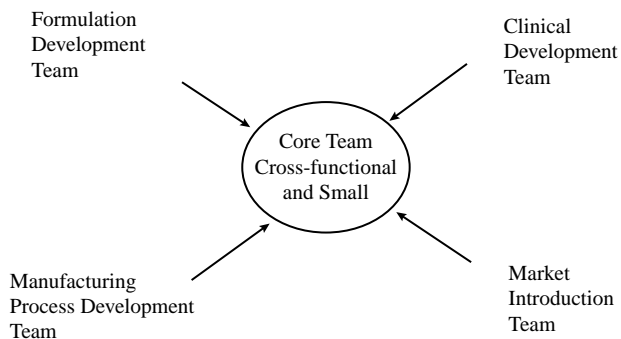


Fig. 3 Design of a project team.

This team example meets the definition criteria in that it is small in number and contains team members that have complementary skills and represent all or most of the functions in the project. Support teams and support team members would then be set up to work on the major elements of the project. In the graphical illustration shown in Fig. 3, these support teams would work on formulation development, clinical development, market introduction, and manufacturing process development.

To be small and effective, a core team should have around twelve members. For major, high-priority projects each team member must have considerable authority and responsibility within their respective functional organization. Each team member

- Ensures functional expertise on the project,
- Represents functional perspective on the project,
- Ensures functional deliverables are met, and
- Proactively raises functional issues that impact the team (2).

This “heavyweight” team member has the authority to either make the decisions and solve the problems on their own or involve the right functional support for decision making and problem solving.

Team members also represent the team. “Heavyweight” team members

- Share responsibility for team results,
- Revise project tasks and content,
- Establish project status reporting and other organizational responsibilities,
- Participate in monitoring and improving team performance,
- Share responsibility for ensuring effective team processes,
- examine issues from an executive point of view, and
- Understand, recognize, and responsibly challenge the boundaries of the project and team processes (2).

Team members must realize that they are part of a team. For the team to be most productive and effective, team members must collaborate with other team members, put aside their own personal or functional agenda, and use dialogue (10, 11) to develop a common understanding and create the best cross-functional approach for the project and the organization.

The term “heavyweight” has been used in this section to designate team members with significant authority and responsibility. Essentially, a heavyweight team is one where more decision making and problem-solving authority and responsibility reside in the team rather than in the functions. As shown in Fig. 4, teams need both a functional or specialist perspective and a project or

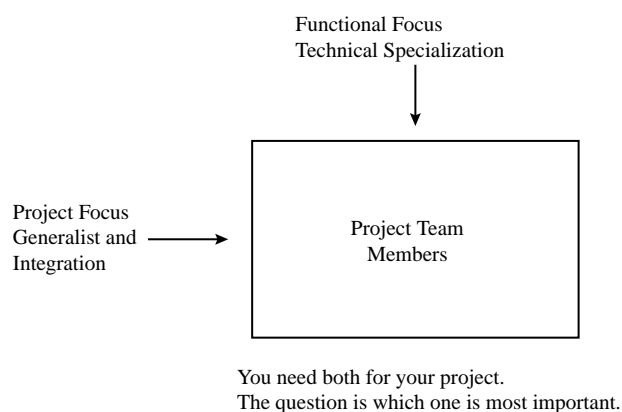


Fig. 4 Project organization structures.

generalist perspective to be effective. Depending on the needs of the project, either the functional or project perspective will be more important to its success. For example, a research project requires more of a technical specialization than a generalist perspective to be successful. Consequently, a functional approach would be more appropriate. Also, a small, generic drug project may not require the team clout and generalist perspective that a larger, proprietary drug would require. Consequently, a “lightweight” team, one where the functional perspective is more important and decisions are made mainly within the functions, may be more appropriate. A “heavyweight” team would have a stronger team oriented, cross-functional perspective rather than a functional perspective. A project or venture team, one where the team members are removed from their functions to report directly to the project leader, may be appropriate in some proprietary drug projects. Such a “heavyweight” team should be created in situations where a new business or market is being entered (12).

Leading a “heavyweight” team requires a project manager who has significant authority, responsibility, and skills, both technical and generalist, to be successful. A “heavyweight” project manager is one who manages, leads, and evaluates other members of the core team, champions the core concept, manages “in motion,” is a multilingual translator (being fluent in different functional languages), and provides direct interpretation of market and customer needs. These individuals “earn the respect and right to carry out these roles based on prior experience, carefully developed skills, and status earned over time. A qualified leader who can play those roles as a heavyweight project manager is a prerequisite to an effective heavyweight team structure (2).

For an effective and productive team-oriented structure, organizations must support cross-functional

teams and use these teams in a way that makes them more effective. Support of teams requires teams be recognized formally as the approach for developing and introducing new pharmaceutical drugs to market. This may sound obvious, but in practice this requires new approaches to

- Performance appraisal—including team-member and total team results with functional evaluations,
- Promotion and pay—creating incentives based upon team results and involvement in successful new product teams, and
- Career development—providing lateral career paths across the functions that develop heavyweight team members.

Anything less than changing the core approaches to these items will result in a team environment that is less than optimal (13).

PROJECT PLANNING, SCHEDULING, AND CONTROL

Although project management was certainly applied in practice prior to the mid-1950’s, the development of the critical path method and the Project Evaluation and Review Technique (PERT) is considered to be the initiation of the modern practice of project management. The critical path method is a network analysis technique used to predict project duration by analyzing which sequence of activities or path most likely has the least amount of scheduling flexibility or the least amount of float. This critical path determines the earliest completion of the project (1). The critical path method relies on one estimate for the duration of a task, whereas PERT uses three estimates. These are most likely (i.e., the critical path method), worst case, and best case estimates. In practice, PERT becomes a very cumbersome approach to scheduling and tracking time.

Network logic determines the relationship between the tasks or team member deliverables, either sequential (finish to start) or overlapping (start to start and/or finish to finish). There are numerous software packages available at low cost to support calculation of the critical path and float. However, the real challenge of project management tools is the creation of the project network based upon historical experience and team involvement. By definition projects are unique and, in the pharmaceutical arena, each project is different. Therefore, the importance of team member involvement in the development, tracking, and adjustment of the project schedule is crucial to successful application of project management basics.

A popular approach for team members to develop a project network diagram is a concept known as the “yellow sticky” method. This method uses post-it notes, roll-paper, and markers and the team to develop a network diagram. The suggested steps of the “yellow sticky” method are as follows:

1. Identify the milestones that the project team will track with management along with target completion dates.
2. Identify the team member deliverables or responsibilities that need to be completed prior to the milestone dates (this provides ending linkages for deliverables).
3. Create the network diagram as follows:
 - Identify which tasks can be started today, based on availability of information not upon availability of resources.
 - Identify which tasks can be started after the initial tasks are completed, which tasks can be completed after that and so forth to the end of the project or the next major milestone.
4. Obtain team member commitments to deliverables—usually best accomplished after the first planning meeting.
5. Negotiate between project milestone dates and team member deliverable dates.

Step 2 uses a very valuable tool known in project management as the work breakdown structure. The work breakdown structure is a “deliverable-oriented grouping of project elements which organizes and defines the total scope of the project” (1). In this example, recommended for pharmaceutical development use, the project is organized at the first level by the milestones that will be tracked. These are the milestones that senior management wants to see in the project. Examples are FDA submission, IND submission, product first lot to stock, first clinical supplies produced, etc. Under the milestones, each team member has deliverables, which are “any measurable, tangible, verifiable outcome, result or item that must be produced to complete a project or part of project” (1). Example of team member deliverables would be to develop the analytical method, develop the product formulation, develop the sterilization process, develop the manufacturing process, develop the specifications, produce clinical supplies, develop the clinical program, and prepare regulatory submissions. Project planning at a level of detail below this can certainly be done. However, a good rule of thumb is to plan down to the level of detail where a team member accepts primary responsibility for the deliverable and no further, unless two things happen:

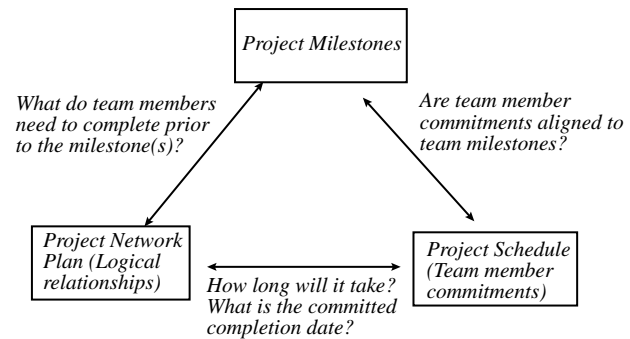


Fig. 5 Team-based project planning and scheduling approach.

- A team member does not understand the job and/or
- A team member is not motivated to do the job.

Standard operating procedures should handle the first issue and having the right team members should handle the second issue. Planning in more detail than this will result in unmotivated team members (i.e., micromanagement) and a schedule that is too cumbersome and detailed for tracking.

Steps 1–3 of this planning method are best accomplished in a core team meeting, along with support team members. Steps 4 and 5 are best accomplished after the meeting when the team members have time to evaluate their workload and commit to the deliverables. Negotiating the deliverable dates with project milestones is also best accomplished after the meeting. Project team members should be advised of the status of this process, but the project manager needs to accomplish this with individual team members, parts of the project team and management, and/or external stakeholder groups. This part of the process is graphically shown in Fig. 5. If there is a misalignment between management’s expectations for completion of major milestones, there are a number of options:

- Change the milestone date,
- “Fast track” the project schedule, or
- “Crash” the project.

Changing the milestone date is straightforward. The target date is rescheduled because the original assumptions for this date were proven to be incorrect when the team came together to schedule the project. “Fast tracking” the project means to compress the project schedule by overlapping activities that would normally be done in sequence. This requires some risk taking, which needs to be clarified. “Crashing” the project means to decrease the total project duration after analyzing a number of alternatives to determine how to get the maximum duration compression for the least cost.

“Crashing” usually occurs by adding resources to reduce time. This happens when additional resources will not significantly reduce the time because the additional resources are not knowledgeable about the technology or part of the team process or a particular task requires a set period of time to complete.

The result of the project planning and scheduling process should be a planning document often called the project or integrated project plan. This serves to emphasize that it represents the integrated plan of all the team members. An integrated project plan should include the following information:

- The project scope and objectives,
- The business case and financial analysis and model of the project,
- The project risks and issues with contingency plans,
- The project plan and schedule (in milestone form for management review and in detailed team member deliverable form for the team members),
- The roles and responsibilities of the team members (best represented through a work breakdown structure tied to project milestones),
- Functional and technical strategies (e.g., regulatory strategy, plans, and challenges), and
- Performance measurements and incentives.

This is meant as an example for illustrative purposes, and the project manager and team should develop their own format and outline based upon their needs.

The Achilles’ heel of project management is the time and resource estimate. In the majority of cases these are actually wrong. The only question is the level of inaccuracy in the estimate. There is no real scientific approach to better estimates and, similarly with all forecasts, it is more of an art and a “feel” to estimating than a science. Consequently, due to the uncertainty of projects

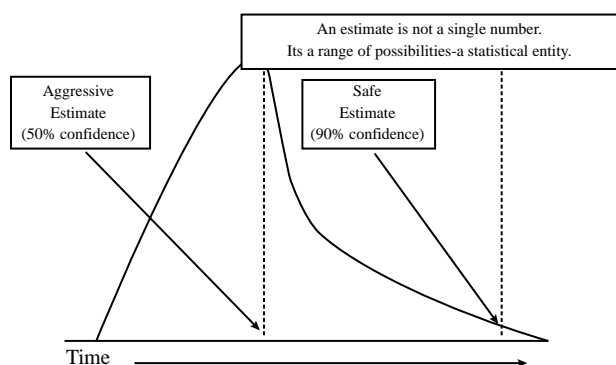


Fig. 6 Time and resource estimates: The Achilles’ heel of project management.

and estimates, it is best to try to work with two numbers: the most likely and the high number (see Fig. 6). New, workable (as opposed to PERT) approaches to project management (i.e. “critical chain” project management) are available to schedule around uncertainty (14).

Once a project schedule is complete, it is used to track and review project progress towards completion. Some basic concepts on this are as follows:

- Project controls need to start with the team members.
- Team members need to take ownership for their tasks and part of this ownership is communicating status on a timely basis.
- A hierarchy of project reviews is needed in support of a pharmaceutical development project. Starting with project core and support team reviews on a weekly or more often basis, project reviews with management and other stakeholders are important to having a common understanding of project status.

Each of these elements would be part of a communication plan to be reviewed in a future section.

PROBLEM SOLVING AND DECISION MAKING USING PROTOTYPES

Pharmaceutical development projects consist of phases, technology transfer processes, milestones, and problem-solving cycles. Understanding these basic building blocks of pharmaceutical development projects in general and more specifically for your project is important towards understanding how your project will progress and how to structure that progress for status and communication purposes. For example,

- The phases of clinical development for a pharmaceutical development project: phase I, II, III, IV and post-marketing;
- The technology transfer from research to development and from development to manufacturing;
- The major milestones of a pharmaceutical project: FDA submission and first lot to stock;
- The use of different product prototyping cycles and their value towards solving problems, communicating results, and making progress.

An example of a pharmaceutical development life cycle is shown graphically in Fig. 7.

The transfer of technology from one organization to another is a key phase of the project itself. As shown in Fig. 8, there are two different approaches that can be used in the transfer of technology:

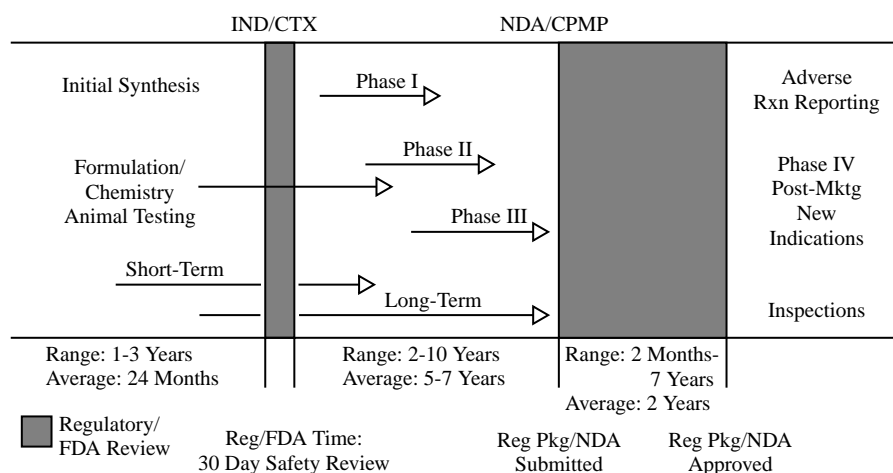


Fig. 7 Drug development phases.

- The traditional, sequential approach or
- The overlapping, concurrent approach.

The sequential approach is often known as the “throw it over the wall” approach. In this approach, the upstream group (e.g., the development team) waits until the product development effort is complete and creates specifications for the new product before communicating this information to the downstream group (e.g., the manufacturing team). Traditionally, this approach has been popular because it gives the perception of being efficient, that is, no time is wasted on incomplete designs and development work. Most often, however, this results in significant rework and inefficient use of time.

The overlapping approach is the more effective and efficient approach in practice and results in a truly integrated transfer of technology. In this approach, both

the upstream and downstream groups work together to understand and solve the problems that need to be resolved to transfer the technology. Examples of these problems may include

- Development and transfer of the analytical methods.
- Development and transfer of the solution process.
- Development of product packaging, and
- Development of sterilization steps and methodology.

In an overlapping approach, team members from both the upstream and downstream groups work together to accomplish these tasks, resolve the problems, and make the necessary decisions. This requires early and intense involvement of the downstream group, even at times when it appears that progress is not occurring. This also requires early sharing of incomplete and, at times, inaccurate information by the upstream group. Sharing of this information may be considered risky due to the uncertainty of response from the downstream group. This approach requires collaboration and trust between the two groups—two basic elements to successful product development and teams (2).

Another basic building block of any product development project are design–build–test cycles. These cycles use prototypes to serve as a focal point for problem solving, testing, communication, and conflict resolution. Using these cycles provides feedback on decisions made so far and identifies issues that need to be resolved. Examples of the prototyping cycles for a clinical development project are

- Identify the bulk drug vendor,
- Determine formulation feasibility,
- Complete engineering runs,

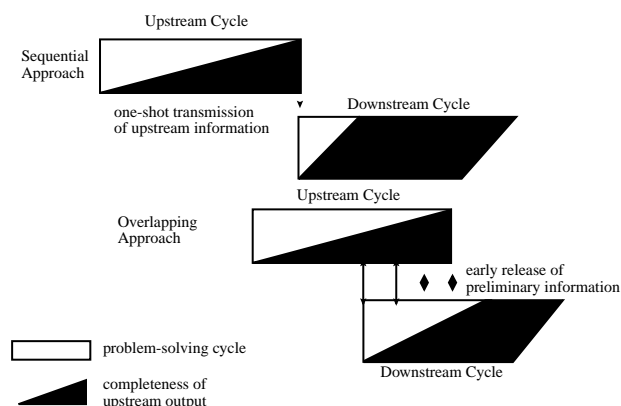


Fig. 8 Cross-functional integration requires overlapping problem solving. (Adapted from Ref. 9.)

- Produce clinical and/or stability samples,
- Ramp-up manufacturing process, and
- Produce first lot to stock product for sale.

These examples are shown to illustrate the importance of the design–build–test cycle process. Cycles will vary from project to project. Completing each of these cycles is a very effective approach to tracking and demonstrating project progress. Prototypes can provide very focused communication about the progress to date and the remaining tasks required to complete the project (2).

SENIOR MANAGEMENT REVIEW AND CONTROL

Senior management provides the sponsorship, championship, and funding for the pharmaceutical development projects that new product teams are working on. They need to be involved and understand the importance of the project, its link to business strategy and growth, the risks and issues that the project is facing, and the status of the project. This understanding is important throughout the life of a project particularly at the start of the project. As shown in Fig. 9, the opportunity to effectively impact the project direction and scope is greater at the start of the project. However, as the project progresses, it becomes more difficult and costly to impact the project direction and scope. Typically, senior management involvement increases as the project gets closer to providing return on investment. This is usually late in the project life cycle and due primarily to an increase in project visibility as sales and marketing activities intensify.

Senior management involvement, especially at the start of the project, is key to successful project outcomes. Some

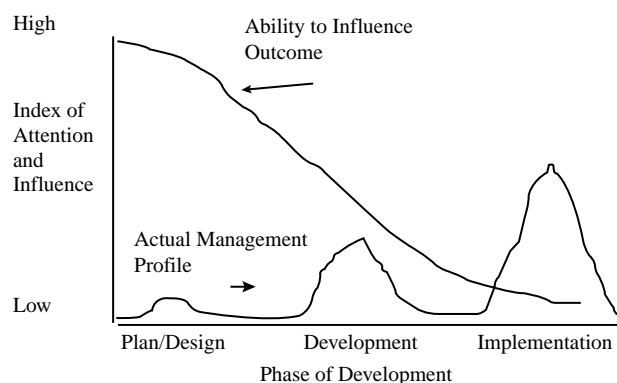


Fig. 9 Obtain senior management involvement early in setting projects direction. (Adapted from Ref. 15.)

process guidelines to assure management involvement are as follows:

1. Senior management should define strategic direction that will be used to drive new product development.
2. Senior management should define the new product project portfolio and the aggregate set of projects that will achieve their strategic objectives.
3. Senior management should be actively involved in assigning the type of project team (see earlier section) and type of project, i.e., where this project falls in the portfolio.
4. Senior management should develop and support a project review process that encourages early involvement in the project to validate the business assumptions and decrease project surprises (12).

The project manager of a pharmaceutical development project should proactively drive these process elements to assure active and early involvement of senior management in their project. The project must have the right support and visibility. Additionally, these process mechanisms assist in timely and appropriate project discontinuations.

Project reviews with senior management are a key element of the new product process. The project manager should have the ability to interface directly with senior management. As previously mentioned, a hierarchy of reviews is important to assure timely, continual communication of project status and issues. Also, stage gate reviews are a very popular process for clarifying project definition and evaluating the project's business case and technical issues. Senior management should get involved in these reviews to assure that the project continues to be on target and meet the overall goals for new product development. Senior management reviews should evaluate

- The business issues and goals of the project,
- Issues related to the customer's needs and the marketplace;
- Resource management;
- Corporate and business unit direction and product line fit;
- Sales and payback expectations;
- Project schedule;
- Timely introduction to market, and
- The availability and maturity of the technology to meet quality and cost targets;

These reviews can be scheduled to occur on a regular basis (e.g., after a phase is completed or every quarter, as stage or phase gates propose) or on an as-needed basis by senior management or the project team. The timing will be dependent upon the organization's requirements and the

specifics of the project. These reviews can be mandatory (i.e., project can not proceed until this review is completed) or flexible. This is dependent upon organization's requirements and the specifics of the project. In all instances, these reviews need to be used to determine whether or not to continue with the project (e.g., make go/no go decisions) or change project direction (e.g., scope change). The concept of scope change will be the topic of the last section of this article.

PROACTIVE, REAL TIME CHANGE MANAGEMENT

As mentioned in this chapter previously, scope changes are to be expected in a pharmaceutical development project. The only question is the number and severity of these scope changes. Consequently, projects need a process for managing scope changes. Mandating the same, centralized controls for all projects, however, would probably be too constricting and result in unneeded bureaucracy—perhaps even impacting the ability for the project to be successful. The project team needs to self-regulate; manage scope changes depending on the unique needs of each project; and have the discipline to ensure appropriate process controls and management review.

Managing scope changes requires the proactive installation of a scope change process for projects that includes a method for identification and management of project risks. A suggested process for managing scope changes is the following:

1. *State the real problem*—Why is this scope change needed? It has been suggested to categorize scope changes into types to assist in determining the relative necessity and importance. Again, this may be overkill.
2. *Gather the relevant facts*—What is the impact if the scope change is implemented? What is the impact on the time to market and effect on development and final product costs? What will happen if the scope change is not implemented? These facts should be obtained from the appropriate people (i.e., the stakeholders that will be impacted one way or another by this scope change).
3. *Develop several alternative solutions*—Should nothing be done? Implement the total scope change as requested? Is there a compromise solution? Which stakeholders are interested in each solution and what are their expectations?
4. *Analyze and review impact for alternatives*—What are the impacts on time to market, investment cost, and product cost for each alternative? Financial analysis is a key element of this step because senior management

uses financials as a critical business indicator. Therefore, each alternative needs to be evaluated versus a financial model. What are the risks and trade-offs to consider for each alternative?

5. *Adopt the best alternative*—This is the connection to the senior management review process. For major scope changes, the project manager should provide an analysis of the major solutions. Based upon financial analysis and qualitative factors, the project manager should present the project team's recommendation to key stakeholders and senior management for their agreement and support.
6. *Tell everyone*—Obviously, the project team should be advised of the change in project scope because they will need to implement it. Project stakeholders also need to be advised so that they can support the scope change with the appropriate resources.
7. *Audit the outcome*—This step is undervalued. The success or failure of a project often relies on how well a scope change was or was not implemented. There are lessons to be learned from capturing the events that led to and caused scope changes and whether the change had a positive or negative impact on project success (17).

As mentioned previously, financial analysis of scope changes and the development of alternative solutions are critical steps towards the successful implementation of a scope change. Senior management needs this information to understand the financial impact on the business and the project manager must provide it. A suggested approach is illustrated graphically in Fig. 10. This approach requires that the project manager use a financial model for the project (which all projects should have) and understand

- The cost/benefit of time delays or improvements;
- The financial impact of increased or reduced product cost and/or project investment; and
- The benefit of improved product use, indications, functionality, or features.

Outlining the relationship of each of these variables to the scope change is an important first step. For example, delaying the project by one week costs approximately \$5 million in division margin. A second step is to understand the specific impact of the scope change on each variable. For example, a scope change to add an additional product use and indication could delay the project three months (a 50% probability), cost an additional \$5 million to implement, and result in additional division margin of \$50 million over the life of the product. Is this worth the investment? Of course, the intangibles also need to be considered.

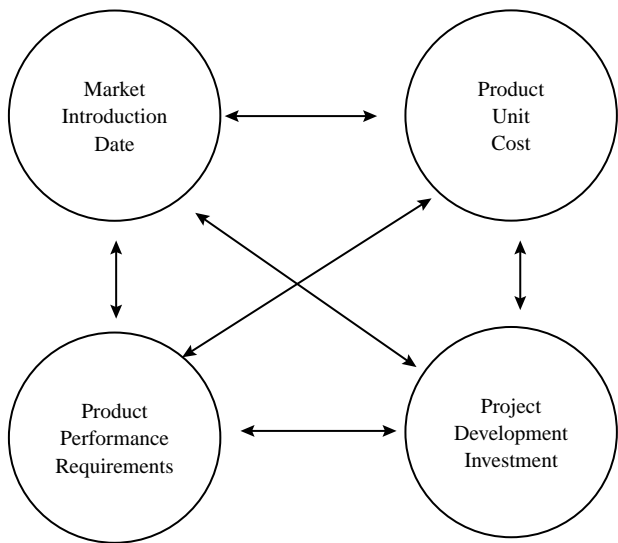


Fig. 10 Financial evaluation for justification of a scope change. (Adapted from Ref. 16.)

This “broad brush strokes” type of financial analysis is fundamental towards making informed decisions regarding trade-offs around the project objectives and scope changes (6). Given the magnitude of uncertainty of and complexity associated with pharmaceutical development, scope changes in this environment require more sophisticated analysis approaches (e.g., Monte Carlo simulation, options analysis, etc.). These tools are beyond the scope of this chapter, and the reader is asked to review the sources mentioned in the earlier sections of this chapter.

Uncertainty is reality in pharmaceutical development projects. Project uncertainty can be found in

- *The marketplace and the competitors:* What other companies will enter this market with a new drug and indication and what will be the impact?
- *The technology and its maturity:* Will our pharmaceutical and medical technology successfully solve the customer’s problem or, at the least, produce better customer outcomes? Will the customer understand and use this technology?
- *The regulatory and governmental environment:* Will the regulatory bodies and government agencies modify the rules? If so, will it affect the basic assumptions of the project? Is there impact on the development, approval, or marketing of the product?
- *The suppliers of bulk, unique commodities and services—*Can the suppliers deliver on the sub-objectives of the project? Are there risks related to their technology, capabilities, or governing environment?

- *The project internal risks (e.g., funding, schedule, team components):* Are the project team members capable of doing their tasks? Will the funding be available to support the project? Can the schedule be met?

These are general categories of risks and not meant to be all inclusive. Using these risk categories as a starting point, the project manager should expand on and flesh out specific details in order to define the most complete picture.

Risk identification is the first step of a risk management process. After risk identification, the project team needs to evaluate each risk in terms of two elements: 1) the probability of occurrence and 2) the financial impact. A popular tool in this process is a risk matrix shown graphically in Fig. 11. The matrix is used to identify the most important risks—those for which a contingency plan needs to be developed. The most important risks would fall into Quadrant IV of the matrix—those with a high probability of occurrence and high financial impact. As the example matrix indicates, the lines drawn between high and low is not quantitatively set but based upon the input of the project team. For risks in Quadrants II and III, continual monitoring is needed to determine whether the situation has changed significantly to warrant a contingency plan even though such a plan was not developed initially.

The basic concepts for risk management in a new product project environment are as follows:

1. Risk identification, quantification, and contingency planning is best accomplished with the full core team being involved.
2. Contingency plans will not be implemented for all risks that have been identified. However, when identifying risks, a common approach is to suggest a contingency plan for this risk. This can serve to encourage a proactive positive approach.

		Probability of Occurrence	
		Low	High
Financial Impact	High	Quadrant II Low Probability High Impact No Contingency Plan Monitor	Quadrant IV High Probability High Impact Develop Contingency Plan
	Low	Quadrant I Low Probability Low Impact No Contingency Plan	Quadrant III High Probability Low Impact No Contingency Plan Monitor

Fig. 11 Risk management plan.

3. Risk management is an ongoing process. You can't complete it and leave it. Many a project manager learned this hard way. Continual monitoring of risks, especially those in Quadrants II and III, is essential.
4. In general, risks resulting from project uncertainty are potentially the most damaging scope changes that could impact the project. It behooves the project manager and the team to proactively manage scope changes through a risk management process.

CONCLUSION

Traditional project management methodology (i.e., for the construction industry) developed 40 years ago has metamorphosed into an approach for managing new product development projects. This more flexible form of the application works well within a decentralized, team-oriented environment—an environment that is conducive to successful pharmaceutical new product development. Project management within the pharmaceutical arena plays a key role in bringing teams together, integrating solutions across functions, and providing a generalist view of projects to balance the more scientific/technical perspective. The future successful project manager will have a generalist orientation, hold team effectiveness and cross-functional integration as the top priority, and strive to overcome the barriers that would undermine these core tenets.

It is not enough to understand the concepts or utilize some of the tools of project management. For the most optimal results, organizations need to recognize the importance and fully support the role of project managers. Often organizations not fully committed to project management opt for interim or temporary solutions such as dual project leadership—a project having both a technical leader and a project manager. A similar tactic would be to have “dual hat” project managers—those who are project managers as a secondary role to their primary role as a functional or technical manager. These need to be replaced with a project manager performing solely that role and who will possess enough skills, knowledge, and

clout to get the project completed. This is not an easy solution, but organizations will find that the right solutions are not often the easy ones (18).

REFERENCES

1. Project Management Institute (PMI) Standards Committee. *Project Management Body of Knowledge*; PMI, 1995.
2. Wheelwright, S.C.; Clark, K.B. *Revolutionizing Product Development*; Free Press: New York, 1992.
3. Reinertsen, D.G. *Managing the Design Factory*; Free Press: New York, 1997.
4. House, C.; Price, R. The Return Map: Tracking Product Teams. *Harvard Bus. Rev.* **1991**, Jan/Feb.
5. Amram, M.; Kulatilaka, N. *Real Options*; Harvard Business School Press: 1999.
6. Smith, Preston; Reinertsen, Donald. *Developing Products in Half the Time*; Van Nostrand Reinhold: New York, 1998.
7. Cooper, R.G. *Winning at New Products*; Addison-Wesley: Reading, MA, 1993.
8. Katzenbach, J.R.; Smith, D.K. *The Wisdom of Teams*; Harper Business, 1994.
9. Hayes, R.H.; Wheelwright, S.C.; Clark, K.B. *Dynamic Manufacturing*; Free Press: New York, 1988.
10. Issacs, W. *Dialogue, The Art of Thinking Together*; Currency Doubleday, 1999.
11. Senge, Peter. *The Fifth Discipline*; Currency Doubleday, 1990.
12. Wheelwright, S.C.; Clark, KimB. *Leading Product Development*; Free Press: New York, 1994.
13. Holahan, P.; Markham, S. Factors Affecting Multifunctional Team Effectiveness. *The PDMA Handbook of Product Development*; Rosenau, M.D., Jr., Ed.; John Wiley & Sons, Inc.: New York, 1996.
14. Newbold, Robert C. *Project Management in the Fast Lane*; APICS Series on Constraints Management. The St. Lucie Press, 1998.
15. Gluck, F.W.; Foster, R.N. Managing Technological Change: A Box of Cigars for Brad. *Harvard Bus. Rev.* **1985**, 141, Sept–Oct.
16. Smith, P.; Reinertsen, D.G. *Developing Products in Half the Time*; VNR: New York, 1998.
17. Rosenau, M. *Successful Project Management*; John Wiley & Sons, Inc.: New York, 1998.
18. Bowen, H.K.; Clark, K.B.; Holloway, C.A.; Wheelwright, S.C. Make Projects the School for Leaders. *Harvard Bus. Rev.* **1994**, Sept/Oct.

PROTEIN BINDING OF DRUGS

Sylvie Laganière

Origenix Technologies, Inc., Quebec, Canada

Iain J. McGilveray

McGilveray Pharmacon, Inc., Ottawa, Ontario, Canada

INTRODUCTION

When a drug reaches the systemic circulation, either after intravenous administration or after absorption following extravascular administration, it can be distributed in the elements of blood (erythrocytes, etc.) or bind to plasma proteins. Blood transports the drug to different organs where it diffuses at different rates. The drug not bound to plasma proteins will diffuse in the extravascular compartments and tissues where it can then bind to other proteins or other tissue components.

The free drug concentrations are more closely correlated to the desirable or undesirable (toxic) pharmacological responses of a drug, because it is the unbound drug that is available to reach tissue receptors (1–3). However, it is the total (bound + unbound) drug concentration, measured in plasma or blood that is most often used in therapeutic drug monitoring to warn of ineffective or toxic levels and to adjust doses if needed. It is easier to measure total drug concentration after a simple blood draw, and usually the ratio of free to total drug concentration is constant. This chapter reviews different aspects of drug-protein binding.

METHODS FOR THE DETERMINATION OF DRUG PROTEIN BINDING

The most commonly used methods to determine drug binding to plasma proteins are equilibrium dialysis, ultrafiltration, and microdialysis. All have advantages and disadvantages, and results are method- and condition-specific.

As well as measuring the binding level of drugs in plasma, procedures can examine the number of binding sites, the affinity constants, and the nature of the protein involved (4) that allow interpretation of the impact of binding on drug pharmacokinetics and response. The general procedures involve separation of the free ligands,

usually small drug molecules, from bound species attached to the protein.

Equilibrium Dialysis (ED)

Introduced in 1943 (5), ED remains the most frequently used procedure (6). A membrane separates two compartments, and at equilibrium, one compartment contains the plasma or serum with protein and bound ligand, whereas free drug is sequestered to the buffer solution compartment. The unbound fraction is determined by the ratio of drug concentration on the buffer side $[D]$ divided by that in the plasma $[D] + [DP]$. Results are influenced by drug properties, proteins (content and concentration), volume of compartments, buffer strength, and ionic composition as well as by the thickness and physicochemical characteristics of the membrane (4). Time and temperature are major environmental factors, and dialysis for 4 h (or less) at 37°C has been found optimal for acidic and basic drugs (6).

In general, natural cellulose membranes are applied (7). Adsorption is low with drugs having acidic or hydrophilic characteristics. Nonspecific drug adsorption should be investigated when studying basic and lipophilic drugs (4).

In experiments, dilution or volume shifts from osmotic gradients between plasma and buffer compartments may be considerable and require correction (8). Poorly water-soluble drugs are difficult to study because use of organic solvents interferes with equilibrium distribution, and such drugs may also aggregate and adhere to membrane surfaces (7).

Ultrafiltration

Although introduced at the same time as ED (9), routine application of ultrafiltration has only recently become feasible because of improvements in membranes and equipment. Separation of the protein and bound ligand from the free drug in solution occurs using a suitable membrane that retains the proteins and is assisted either by positive pressure or more commonly by centrifugation.

The key advantage of ultrafiltration is speed (as little as 15 min), which is an important criterion for clinical monitoring situations (6) and for unstable drugs. As with ED, membrane properties and volume are important. The membranes (e.g., Amicon®) are permiselective and the cut-off can be directed at protein molecular weight sizes e.g., separation of albumin and globulins such as α 1-acid glycoprotein (AAG). However, in some cases, high-molecular-weight drugs can be retained (7). Adsorption of drugs on the membrane is problematic with low concentrations, especially of lipid-soluble ligands. Saturation of the adsorbing sites with study drug or coating with silicone has been applied to overcome such problems.

Theoretically, the free ligand concentration will be constant in the ultrafiltrate and retentate (10). Nonetheless, a decrease in concentration of free drug in the retentate will result in the dissociation of bound ligand. However, provided the ultrafiltrate is not greater than 40% of the initial volume, the free ligand concentration is not affected (11). A disadvantage of both ED and ultrafiltration for low-concentration or highly bound ($\geq 90\%$) drugs is the need for radioisotopes to provide the sensitivity for quantitation.

Microdialysis

Microdialysis perfusion is proposed as an alternative to traditional techniques for studying drug-protein interaction. Small molecules in the sample, such as a drug, diffuse in the fiber and are transported to collection vials for analysis. The dialysis membrane excludes larger molecules such as protein and drug-protein-bound drugs. The dialysate could be analyzed using standard techniques. Microdialysis perfusion is as rapid as ultrafiltration, but the samples do not suffer from re-equilibration during separation of free drug from bound drug.

Because the technique requires more chromatographic analysis, it has a longer analytical time compared with the ultrafiltration technique. This technique could be used for in vitro determination of protein binding, also, because microdialysis probes can be implanted intravenously, binding of drug could be determined in vivo under physiological conditions after dosing (12). There has been some validation of the technique used for in vitro determination of the unbound fraction of drugs in plasma. Although there is good agreement between equilibrium dialysis and microdialysis for phenytoin and racemic aminoglutethimide, the unbound fractions are significantly higher for furosemide and disopyramide using microdialysis (13). Good correspondence was demonstrated

between ultrafiltration and microdialysis for acebutolol, acetaminophen, cephalothin, chloramphenicol, isoniazid, phenytoin, salicylic acid, theophylline, and warfarin (14). It appears that the microdialysis technique may not be suitable for all compounds because of their physicochemical properties and analytical limitations (15, 16).

Pacifici and Viani (6), commenting on comparisons among results from different methods applied to protein binding, warn that methods for drug binding need to be standardized. Currently, results depend greatly on technique and interpretation of results; thus, experimental details such as anticoagulant used, type of dialysis or filtration membrane, buffer characteristics, duration of experiment, and temperature should be taken into account.

PROTEIN BINDING PARAMETERS

Drugs bind to different proteins, such as albumin, globulins (e.g., α 1-acid glycoprotein AAG and transcortin), and lipoprotein in plasma, and to tissue proteins. The reversible binding of a drug to macromolecules such as proteins is an equilibrium process. The relationship is described by the law of mass action:



where $[D]$, $[P]$, and $[DP]$ are the molar concentration of unbound (free) drug, unoccupied proteins, and the drug-protein complex, respectively, and k_1 and k_2 are rate constants for the reactions.

The equilibrium association constant for this reaction K_A is defined as k_1/k_2 and provides an index of the affinity between the binding sites and the ligand. The inverse of the association constant ($1/K_A$) is the equilibrium dissociation constant K_D for the drug-protein complex:

$$K_A = \frac{k_1}{k_2} = \frac{[DP]}{[D] \cdot [P]} \quad (2)$$

At equilibrium, the unbound fraction of drug (f_u) can be calculated from the following relationship:

$$f_u = \frac{[D]}{[D] + [DP]} \quad (3)$$

A protein can have several independent binding sites such that they may exhibit either the same or different affinities for the drug. The binding to one site may not affect binding to another site. The capacity constant (N_{TOT}) is the number of sites per mole of protein times the molar concentration of protein and has units of sites/L.

The concentration of bound drug can be expressed as the following:

$$[DP] = \sum_{i=1}^{n=i} \frac{N_{TOT,i}[D]}{K_{D,i} + [D]} \quad (4)$$

where i is the number of classes of binding sites, $N_{TOT,i}$ the capacity constant, $[D]$ the unbound drug concentration, and $K_{D,i}$ the dissociation constant for the i site. Therefore, the extent to which a drug binds depends on the affinity of the drug, the number of binding sites per molecule, and the concentrations of both drug and binding proteins.

Fig. 1 shows a plot of $[DP]$ versus $[D]$ for a single class of binding sites. As the unbound drug concentration increases, the concentration of bound sites increases. When $[D]$ is much larger than K_D , the equation predicts that $[DP]$ reaches a maximum where it is equal to the maximum number of sites N_{TOT} . The dissociation constant K_D represents the concentration of unbound drug when half the sites are occupied, $[DP] = N_{TOT}/2$. The plots of Eq. 3 after linear transformations such as the Scatchard plot, the Woolf plot, and the double reciprocal plot can be used to determine the dissociation and capacity constants. The plots will be linear when there is one class of binding sites and will be curvilinear in presence of two or more classes of binding sites on the protein or if it binds to sites in more than one protein (17). For example, the Scatchard equation for a single class of binding site is:

$$\frac{[DP]/P_T}{[D]} = N \cdot K_A - \{(K_A \cdot [DP])/P_T\} \quad (5)$$

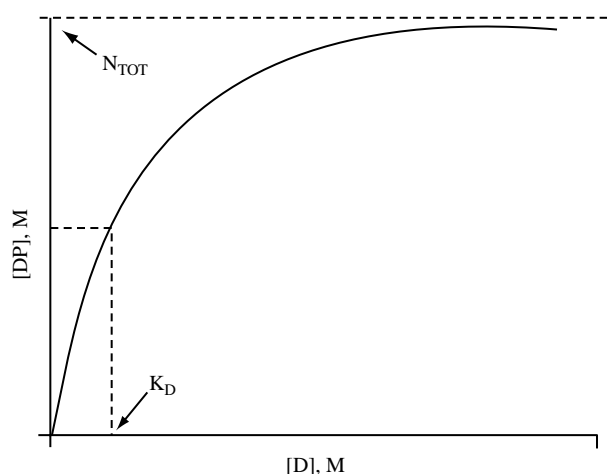


Fig. 1 A plot of bound drug $[DP]$ versus unbound drug concentration $[D]$ for a single class of binding sites.

where P_T is the molar concentration of total protein (occupied + unoccupied), and N the number of sites per mole of protein. A plot of $([DP]/P_T)/[D]$ versus $[DP]/P_T$ gives a linear slope of $-K_A$ and a y-intercept of $N \cdot K_A$. Refined estimates of the capacity and affinity constant can be generated using a nonlinear least-squares computer program such as PCNONLIN (18).

Once the capacity and affinity constants have been obtained, the unbound plasma fraction of drug that binds to a single class of binding sites can be calculated for any value of free drug concentration:

$$f_u = \frac{[D]}{[D] + [DP]} = \frac{K_D + [D]}{N_{TOT} + [D] + K_D} \quad (6)$$

When a drug binds to two classes of binding sites such as albumin and AAG, the equation is:

$$f_u = \frac{1}{1 + \frac{N_{TOT1}}{K_{D1} + [D]} + \frac{N_{TOT2}}{K_{D2} + [D]}} \quad (7)$$

These equations indicate that the unbound fraction of the drug is dependent on the binding capacity N_{TOT} , the dissociation constant K_D , and the unbound drug concentration at equilibrium. At low free drug concentrations (i.e., $[D] \ll K_D$), the unbound fraction will be independent of change in drug concentration.

DRUG BINDING PROTEINS

Human plasma contains more than 60 proteins (19). There are three main proteins associated with the binding of drugs: albumin, AAG, and lipoproteins. This article does not review drug transport proteins of recent interest such as p-glycoproteins, organic anion transporters, and dipeptide transporters, which have been reviewed recently (20).

Albumin

zAlbumin (MW 66,300), an important binding protein present in plasma (40%) and interstitial fluid (60%) (21), which binds mostly to acidic (anionic) drugs (22) but also to cationic drugs, accounts for approximately 60% of total plasma protein (23). This water soluble protein has an isoelectric point of approximately pH 5. Therefore, at physiological pH, this protein is negatively charged, and acidic drugs bind usually to the N -terminal group. Normal serum concentrations are 38–48 g/L. Albumin is synthesized in the liver, and its half-life is approximately 19 days (23).

Two primary high-affinity drug binding sites for human albumin have been described. Binding site I, initially described for warfarin, was also shown to be involved in the binding of other drugs such as sulfonamides, phenytoin, valproic acid and phenylbutazone. Site II (the benzodiazepine binding site) binds semisynthetic penicillins and probenecid. Naproxen, indomethacin, and tolbutamide have been shown to bind to both sites.

AAG

AAG (orosomucoid) is synthesized and metabolized by the liver with a half-life of approximately 5.5 days (24) and is an α_1 -globulin protein that is smaller (MW 40,000) than albumin. It is an acute phase reactant, and concentrations increase in stress situations including diseases. At least four polymorphic patterns and four genetic variants in human plasma have been reported (25, 26).

Many basic drugs (cationic) such as propranolol, lidocaine, and neutral drugs bind significantly to AAG (normal concentration 0.07–1.1 g/L) (27) and/or to lipoproteins in addition to albumin. Generally, the affinity of a drug that binds to both albumin and AAG is higher for AAG. Because the concentration of AAG in plasma is lower than that of albumin, the AAG is referred to as a low-capacity, high-affinity protein. AAG also binds to some acidic (28, 29) and neutral drugs such as carbamazepine (17) and prednisolone (30). Interpatient variability in the plasma free fraction of several drugs such as propranolol (31), carbamazepine (32), disopyramide (33), alprenolol (34), imipramine (34), lidocaine (35), and methadone (36), is related to AAG concentration in plasma.

Lipoproteins

The lipoproteins are a heterogeneous group of proteins (MW 2,500,000 for β -lipoprotein). They are classified into four groups: chylomicrons, very low-density lipoproteins (VLDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL) (37, 38). Lipoproteins can account for 95% of total probucol binding in plasma (37). Neutral and basic lipophilic drugs most commonly bind to lipoproteins; some acid drugs may bind to a lesser extent (37). Drugs that bind to lipoproteins include probucol, cyclosporin, propranolol, lidocaine, pindolol, digoxin, digitoxin and tetracycline (37, 39–41). Quinidine and some antidepressants drugs such as amitriptyline and nortriptyline exhibit saturable binding to lipoproteins. An inverse relationship was reported between the unbound fraction of drugs and concentrations of various lipoproteins

in serum for amitriptyline, imipramine, nortriptyline, and cyclosporin (37, 39, 42).

A list of some acidic, basic, and neutral drugs highly bound (>70%) to the different plasma proteins after therapeutic doses is given in Table 1 (43).

INFLUENCE OF BINDING ON DRUG DISPOSITION

Volume of Distribution

The apparent volume of distribution at steady state (V_{ss}) is influenced by plasma and tissue protein binding according to equation:

$$V_{ss} = V_P + (f_P/f_T)V_T \quad (8)$$

where f_P and f_T are the unbound fraction of drug in plasma and tissue, respectively, and V_P and V_T are the volumes of distribution in plasma and tissue (44). The importance of binding in the plasma versus the tissue compartment is the primary determinant of the apparent volume of distribution for a drug. The distribution equilibrium of the drug in the body varies among tissues.

Tissue binding

Tissue binding can be a major determinant of pharmacokinetic characteristics of drugs, including volume of distribution and half-life. At present, no entirely satisfactory method is available for the determination of tissue binding. Sampling is invasive, and experimental animal values are difficult to validate for human correlations (45).

Examples of changes in tissue binding are the tissue displacement of digoxin by quinidine (46) and the decrease in tissue binding of digoxin in uremia (47). As a result in both cases, the volume of distribution of digoxin is decreased.

The binding of drug to tissue is usually reversible. In some cases, however, there is covalent binding, which by definition is not reversible. This applies to drug or metabolite and could be important because it could be related to toxicity (48). A good correlation has been reported in animals between the degree of covalent binding to hepatic protein and the severity of hepatic necrosis of paracetamol, isoniazid, adriamycin, and furosemide (48).

Characteristics of drugs with a large volume of distribution include weak binding to plasma proteins, high affinity to tissue protein, and high liposolubility. Increased volume of distribution can be related to an increased unbound fraction in plasma as, for example, for propranolol in chronic stable liver disease (47). Increased

Table 1 Predominant binding proteins of drugs >70% bound to plasma proteins

Albumin (% bound)	Albumin and AAG (% bound)	Albumin and lipoproteins (% bound)	Albumin, AAG, and lipoproteins (% bound)
Ceftriaxone (A)	Alprenolol (B)	Cyclosporine (N) ^a	Amitriptylline (B)
Clindamycin (A)	Carbamazepine (N)	Probucol (N) ^a	Bupivacaine (B)
Clofibrate (A)	Disopyramide (B) ^b		Chlorpromazine (B)
Dexamethasone (N)	Erythromycin (B)		Diltiazem (B)
Diazepam (B)	Lidocaine (B)		Imipramine (B)
Diazoxide (A)	Meperidine (B)		Nortriptyline (B)
Dicloxacillin (N)	Methadone (B)		Perazine (B)
Digitoxin (N)	Verapamil (B)		Propranolol (B)
Etoposide (N)			Quinidine (B)
Ibuprofen (A)			
Indomethacin (A)			
Nafcillin (A)			
Naproxen (A)			
Oxacillin (A)			
Phenylbutazone (A)			
Phenytoin (A)			
Probenecid (A)			
Salicylic acid (A)			
Sulfisoxazole (A)			
Teniposide (N)			
Thiopental (A)			
Tolbutamide (A)			
Valproic acid (A)			
Warfarin (A)			

^aAlbumin is minor binding protein.

A, indicates acid; B, base; N, neutral.

(From Ref. 43.)

volume of distribution could also result from shift of albumin-bound drug from the vascular to the interstitial space, as reported for ceftriaxone during open-heart surgery (49). For an AAG-bound drug, a decrease in volume of distribution would be related to an increase in binding in plasma because of stress-related conditions, increasing available AAG. Distribution equilibrium is achieved when unbound drug concentrations are the same in plasma and tissue.

Clearance

Hepatic clearance

The simple venous equilibrium model of organ clearance developed by Rowland (50) and Wilkinson (51) can be used to predict the effect of protein binding on hepatic plasma clearance:

$$Cl_H = \frac{Q(Cl_{I,u}f_u)}{Q + (Cl_{I,u}f_u)} \quad (9)$$

where $Cl_{I,u}$ is the intrinsic unbound hepatic clearance, Q the hepatic plasma flow, and Cl_H the hepatic clearance.

Other models of hepatic clearance include the sinusoidal model and the dispersion model (52). All these models assume that only unbound drug can move freely through cell membranes and that the unbound fraction of a drug measured in the systemic circulation is the same as the fraction available for hepatic metabolism. The $Cl_{I,u}$ value is measured by dividing the intrinsic clearance of total drug by the measured unbound fraction.

There is, however, evidence of nonrestrictive clearance by the liver when the fraction of drug available for metabolism may not be restricted to the unbound fraction of drug (53). Drugs that bind to AAG or lipoproteins could be taken up directly by the hepatocytes by endocytosis (54, 55).

Restrictive drug clearance

A drug is said to be restrictively cleared if the extraction efficiency by an eliminating organ (E) is less than or

equal to the unbound fraction of drug measured in the venous circulation. For these drugs, only the unbound fraction is available for distribution in the tissues, biotransformation, and excretion. For restrictive drugs with low extraction efficiency ($Cl_{l,u} \cdot f_u < Q$), a change in unbound fraction will affect the blood clearance as Eq. 9 simplifies to Eq. 10:

$$Cl_H = Cl_{l,u} \cdot f_u \quad (10)$$

The average steady-state drug concentration in plasma for a restrictively cleared and poorly extracted drug is determined by:

$$C_u = \frac{FD/\tau}{Cl_{l,u}} \quad (11)$$

where τ is the dosage interval. Therefore, a change in the unbound fraction will not affect the unbound concentration. It will then not be necessary to alter drug dosage after a change in binding in this situation.

It is important to note that an increase in the f_p value (fraction unbound in the plasma) caused by a displacement from drug during drug-drug interaction will cause only a transient rise in the unbound drug concentration, followed by a return to the predisplacement value when steady state has been reached (56, 57).

The influence of altered plasma or tissue binding on the elimination half-life depends on the relative influence of binding on the volume of distribution and the clearance. For a restrictively cleared and poorly extracted drug with moderate to large volume of distribution (i.e., >0.4 L/kg), the equation to reflect the half-life could be represented as:

$$t_{1/2} = 0.693 (V_T / Cl_{l,u} \cdot f_T) \quad (12)$$

and a change in plasma binding would have little influence on the half-lives of restrictively cleared drugs (e.g., diazepam, phenytoin) (56). For a drug with a small volume of distribution (<0.4 L/kg) such as warfarin, the $t_{1/2}$ value will be affected by a change in plasma or tissue binding.

Examples of significant drug interactions in which protein binding displacement is the major factor are phenytoin displacement by salicylic acid and tolbutamide and warfarin displacement by trichloroacetic acid, a metabolite of chloral hydrate (57). A change in dosage regimen is not necessary for phenytoin because, as described above, a displacement of restrictively cleared drugs from protein will decrease the total drug concentration but at steady state will not affect the average unbound concentrations. For warfarin, it might be necessary to temporarily adjust the dosage regimen because

this is a narrow therapeutic index drug with a long half-life and small volume for distribution.

Renal clearance

In the kidney, only the unbound drug is filtered. If the drug is not bound and there is absence of drug secretion and reabsorption, the renal clearance is a measure of the glomerular filtration rate (GFR). If the drug is protein-bound, the renal clearance of total drug in the plasma is less than the GFR, but the clearance of the free drug is equal to the GFR, as described by the relationship $Cl_R = f_u \cdot GFR$. For a drug that is either reabsorbed or secreted, the effect of altered protein binding is less predictable. Renal clearance expressed in terms of the renal processes is as follow:

$$Cl_R = (Cl_{RF} + Cl_{RS})(1 - FR) \quad (13)$$

where Cl_{RF} is the renal filtration clearance, Cl_{RS} the renal secretion clearance, and FR the reabsorbed fraction, and

$$Cl_{RS} = RBF f_u Cl_l / (RBF + f_u Cl_l) \quad (14)$$

where RBF is the renal blood flow. The clearance of drugs that undergo extensive tubular secretion (e.g. penicillins) is independent of plasma protein binding. The lack of reported side effects for the interaction of ceftriaxone with probenecid may be explained by the wide therapeutic index of the ceftriaxone (58).

Nonrestrictively cleared drugs

For nonrestrictively cleared drugs ($E > f_B$), which are highly cleared by an organ ($Cl_l > Q$), the unbound fraction is not rate-limiting because the clearance is determined primarily by the organ blood flow, and a change in unbound fraction will not affect the blood clearance. For example, the bioavailability of an oral dose will be independent of changes in blood or plasma protein binding. Changes in the plasma binding of nonrestrictively cleared drugs, given orally or intravenously, should not affect the average total drug concentration (56). Examples of nonrestrictively cleared drugs include morphine, meperidine, lidocaine (59), verapamil (60), and tricyclic antidepressants (61).

Drugs, that are nonrestrictively cleared by renal tubular secretion include penicillin and acetazolamide (57). Although significant displacement of penicillins by aspirin and sulfonamides occurs, as shown by an increase of up to 88% in unbound fractions of penicillins and no change in total drug concentration, this interaction is not clinically important because of the wide therapeutic index of these drugs (62). It is worth noting that although the clearance is

not affected by a change in binding, a change in the f_u value may affect the drug response. The elimination half-lives of nonrestrictively cleared drugs is affected in relation to the equation:

$$t_{1/2} = \frac{0.693(V_P + (V_T \cdot f_B/f_T))}{Q} \quad (15)$$

For a drug with a large volume of distribution, such as propranolol, a smaller unbound fraction in the plasma will result in a shorter half-life, a greater fluctuation during the dosage interval, and no change in the average total drug concentration. For drugs with small volumes of distribution, such as penicillins, a change in plasma or tissue binding will have little influence on their half-lives. A change in drug binding to tissues alters the apparent volume of distribution. Therefore, an increase in tissue binding will produce an increase in volume of distribution and half-life.

Concentration Dependent Binding

The degree of binding of a drug is independent of drug concentration when the molar unbound concentration is well below the dissociation constant of the drug-protein complex. Nonlinear binding during therapeutic doses may occur when the K_D value and the therapeutic concentrations are close (see Fig. 1). For valproic acid (63), the K_D value for the high-affinity site is $5 \times 10^{-5} M$, and the therapeutic range of free concentrations is 0.7 to $2 \times 10^{-4} M$. For salicylic acid (22), the K_D value is $10^{-5} M$ and the therapeutic range of free concentrations is 3 to $7 \times 10^{-5} M$. Nonlinear binding is also reported for diflunisal, an anti-inflammatory drug (64), and for thiopental (65) and diazoxide (66) all of which bind predominantly to albumin. Important examples of concentration dependent binding are shown for disopyramide (2, 67) over the therapeutic concentration range and for lidocaine at the upper boundary of the therapeutic range (68, 69). The concentration-dependent binding of prednisolone, observed at very low unbound concentrations, is explained by its high-affinity binding to transcortin (70).

For a restrictively cleared drug with concentration-dependent binding, such as disopyramide, the unbound drug concentration increases proportionally with the increase in dosage regimen. Also, with the increase in the fraction unbound, the clearance also increases. Although the total concentration also increases, it is less than proportional to the increase in dosage rate. In this case, the total concentration can be misleading, and dose adjustment should be based on the unbound concentration

of disopyramide or, if unavailable, on the patient's response (2, 71).

Blood Versus Plasma Clearance

Pharmacokinetic parameters should preferably be determined from blood, but usually drug concentration is measured in plasma or serum. Although the free fraction will be different in tissue versus plasma, the unbound concentration will be the same at equilibrium as $(f_P C_P) = (f_B \cdot C_P)$.

The determination of the blood/plasma ratio will help estimate the importance of the distribution of the drug in the erythrocytes. Drug binding to erythrocytes is usually rapidly reversible.

Drugs that bind strongly to red blood cells may exhibit concentration-dependent uptake from plasma. This is reported for acetazolamide (72), chlorthalidone (73), and cyclosporine (74). The determination of the blood to plasma (B/P) ratios can be method specific. For example, the determination of the B/P ratio of cyclosporine A increases from 1.5 to 3.0 when estimated at body temperature ($37^\circ C$) compared with a ratio of 2.5–10.0 at room temperature ($20^\circ C$) (75). Drugs reported to bind strongly to the erythrocytes (B/P ratio > 1) are promazine, chlorpromazine, propranolol, salicylate, phenobarbital, pentazocine, and phenytoin (48).

Usually, the pharmacokinetic parameters are calculated from plasma rather than from whole blood data. The use of plasma-related parameters to measure organ extraction ratios and intrinsic clearance could, however, be misleading. When drug binding is similar in plasma and red blood cells, $C_B \approx C_P$ because $f_B \approx f_P$. However, if the binding in plasma is greater than that in red blood cells, $C_B < C_P$ and $AUC_B < AUC_P$ and use of plasma clearance will provide an underestimate of the blood clearance. Maximum errors are of the order of 40%. If the B/P ratio is > 1 , the clearance determined from plasma concentrations would significantly overestimate blood clearance and could exceed hepatic blood flow.

Usually, total plasma concentration of drug is measured in plasma. The total drug concentration in blood can be estimated by:

$$C_B = C_{RBC} \cdot HCT + C_P(1 - HCT) \quad (16)$$

where HCT is hematocrit, and C_{RBC} is the drug concentration in red blood cells. Although drug binding to plasma proteins is usually rapidly reversible, some studies reported the presence of irreversible binding to albumin (e.g., the acyl glucuronide metabolite of zomepirac (76) and tolmetin (77)).

Stereoselectivity

For acidic drugs, stereoselectivity was demonstrated for both albumin binding sites I and II (78). Stereoselective binding was reported for ibuprofen enantiomers, with the mean averaged unbound fraction of the R(−) enantiomer being 0.419%, significantly less than that of the S(+) enantiomer of 0.643%. The percentage unbound of each enantiomer was concentration-dependent over the therapeutic range and was influenced by the presence of its optical antipode (79). Stereoselective binding in humans was reported for acidic drugs such as etodolac, flurbiprofen, ibuprofen, moxalactam, pentobarbital, phenprocoumon, and warfarin and for basic drugs such as chloroquine, disopyramide, methadone, propranolol, mexiletine, and verapamil. A list of human plasma protein binding of drug enantiomers was summarized by Jamali et al. (80).

Stereoselectivity in protein binding of enantiomers can also differ among species. For propranolol, a basic drug bound to AAG, the R-enantiomer binds less than the S-isomer in humans and dogs, and the reverse is observed in rat. Also, although the difference in binding is small in humans and dog, it is significant in rat (81).

Although the enantiomers of disopyramide displaced each other from AAG binding sites, resulting in a 2- to 2.6-fold change in f_u value in vitro (82), no significant enantiomer–enantiomer interaction in serum was found in vivo, presumably because of the buffering effect of serum albumin (60). The free fraction of verapamil enantiomers, determined from ex vivo volunteer samples after intravenous therapy (R-verapamil 0.06 ± 0.01 , and S-verapamil 0.12 ± 0.02), was similar to that observed in vitro in the subjects' predose serum. The free fraction of both enantiomers was higher after oral drug therapy, the R- and S-isomers being 0.13 ± 0.02 and 0.23 ± 0.03 , respectively. The difference in binding might be attributable to competition for serum binding sites from verapamil metabolites, which reach higher concentrations after oral administration (60).

A difference in the binding of the two isomers, quinine and quinidine, was reported (1, 2), and the f_u value of quinine and quinidine was $7.5 \pm 2.2\%$ and $12.3 \pm 2.3\%$, respectively, in normal human plasma.

INFLUENCE AND CONSEQUENCES OF BINDING ON DRUG PHARMACODYNAMIC EFFECTS

The free fraction of drug is most likely to relate to pharmacological effect. None-the-less, because it is

methodologically easier to determine total drug concentrations in blood or serum and the free/total drug concentration is usually constant, total concentrations are commonly applied in therapeutic drug monitoring. However, for some drugs the free/total ratio is variable, and disease conditions and drug interactions may alter it. In general, for monitoring of drug concentrations, the plasma or serum is used as the liquid of reference for practical reasons, but blood concentration is preferable because it reflects total drug concentration available for distribution.

Monitoring of Free Drug Concentrations

The monitoring of free drug concentrations may be necessary for drugs with concentration-dependent binding over the therapeutic range or in patients with diseases or conditions for which the unbound fraction of a drug could be significantly altered from normal (83). Compared with drugs that are mostly free in plasma, drugs that are highly bound ($f_u < 30\%$) would be candidates for monitoring because they are most likely to show significant differences in the unbound fraction under certain conditions.

Drugs for which total drug concentrations are usually measured and for which there would be advantage in monitoring free drug concentrations include carbamazepine, phenytoin, valproic acid, disopyramide, lidocaine, and quinidine (84–86). For disopyramide, monitoring of free drug concentration should be applied to the active enantiomer (87). However, monitoring of free concentration of these drugs in all patients is not recommended unless one of the following criteria is exhibited: 1) patients with disease likely to be associated with altered unbound fraction, 2) patients under treatment with a drug combination with potential protein binding interaction, and 3) patients showing unexpected drug response at a determined total drug concentration.

FACTORS AFFECTING PROTEIN BINDING AND CLINICAL CONSEQUENCES

Changes in protein binding may be related to factors affecting the concentration of proteins, to the variation in conformation of the binding protein such as for albumin, and to displacement from a protein binding site. The consequences of alteration in protein binding may be more important clinically for drugs highly bound ($<90\%$) to plasma proteins at therapeutic concentrations and drugs with small a volume of distribution and narrow therapeutic index (88). A list of pathologic/physiologic conditions

Table 2 Pathological/physiological conditions associated with altered protein concentrations

	Albumin		AAG	Lipoproteins
↓ Plasma protein concentration	Acute febrile infections	Malnutrition	Advanced age	Familial deficiencies ^a
	Acute viral hepatitis	Neonates/young infants	Cirrhosis	Hyperthyroidism
	Acute pancreatitis	Nephrotic syndrome ^a	Neonates/young infants	Low cholesterol diet
	Advanced age ^a	Pregnancy ^a	Oral contraceptives	
	Analbuminemia ^a	Prolonged bedrest	Pregnancy	
	Burn injury ^a	Protein-losing nephropathy	Severe liver disease	
	Cancer ^a	Renal failure		
	Cirrhosis ^a	Rheumatoid arthritis		
	Cystic fibrosis	Stress		
	Hyperthyroidism	Surgery ^a		
	Malabsorption	Trauma injury ^a		
	Dehydration		Acute myocardial infarction ^a	Alcoholism
	Gynecological syndrome		Administration of some enzyme inducers ^a	Antihypertensive drugs
	Optic neuritis/retinitis		Advanced age	Biliary obstruction
	Psychosis		Burn injury ^a	Diabetes mellitus
↑ Plasma protein concentration	Unspecified neuroses		Cancer	Familial-hyperlipoproteinemia ^a
			Chronic pain syndrome	Gout
			Inflammatory disease ^a	High-cholesterol diet
			Pneumonia	Hypothyroidism
			Renal transplant	Liver disease
			Surgery ^a	Nephrotic syndrome
			Trauma injury ^a	Pancreatitis
				Phenytoin or cyclosporine administration.
				Pregnancy
				Renal failure

^aConditions likely to be associated with major changes.
(From Ref. 43.)

associated with altered protein concentrations is given in Table 2.

Physiologic Conditions

Age

The plasma protein binding of drugs bound to albumin and AAG in neonates is significantly lower than that in healthy adults. AAG is low at birth and increases gradually over 12 months to reach a normal concentration (89). Factors associated with the decrease in binding are lower albumin and AAG concentrations and the presence of fetal albumin. Also, high serum concentrations of bilirubin and free fatty acids occur in the neonate, and these can compete for the albumin binding sites (90–93).

In the elderly, a small decrease in albumin concentration is associated with a small decrease in binding for salicylate (94), diazepam (95, 96), phenylbutazone (94) and valproic acid (97). Also in the elderly, an increase in concentration of AAG is associated with an increase in binding of AAG-bound drugs (98). Such an increase in AAG concentrations could be related to the greater incidence of inflammatory diseases in this population.

Pregnancy

Significant decreases in protein binding of albumin-bound drugs are reported particularly during the third trimester of pregnancy, and are related to decreased albumin concentrations and higher, concentration of fatty acids, especially during labor. As a result, increases in the unbound fraction of salicylate (70–80%) (64–69), sulfisoxazoles (70–80%) (99), phenytoin (30%) (99–101), diazepam (40–60%) (99, 100), and valproic acid (50%) (100) occur.

For AAG-bound drugs, a 35–80% increase of the unbound fraction in the third trimester is reported for propranolol (92), lidocaine (92), and bupivacaine (102). After pregnancy, whereas albumin takes 1 month to return to normal levels, AAG and free fatty acid levels return to normal within as few days of delivery (101).

Ethnicity

Formely, much of the information obtained on drugs was acquired in Caucasian subjects. A report on inter-ethnic differences in drug response showed differences in drug protein binding between Chinese and Caucasian subjects. No difference was shown for warfarin, which is an albumin-bound drug, whereas the binding of lidocaine was significantly lower in Chinese subjects and related to lower concentrations of the binding protein AAG (103). Because lidocaine toxicity is related to free drug, this difference could have therapeutic significance. A similar reduction in

the binding of propranolol and disopyramide was also reported previously (104, 105). This ethnic difference does not seem to extend to drugs bound to albumin. Studies, that examined diazepam and salicylic acid, describe findings similar to those observed for warfarin (106).

Gender, smoking, obesity, nutritional status, surgery

The unbound concentration of diazepam is 14% greater in females than in males, whereas there are no significant differences in the binding of lidocaine (107). Although McNamara (69) reported a small increase in plasma binding of lidocaine, resulting in a decrease in free fraction from 0.31 to 0.26 in smokers, another study reported no effect of smoking on the plasma binding of lidocaine and diazepam (96).

In obese women, the AAG concentrations are doubled, and the unbound plasma fraction of propranolol is 30% lower than that in women with normal body weight (108). In the obese, the albumin concentrations, and phenytoin binding are normal, whereas diazepam binding is slightly reduced (108). In undernourished or hospitalized patients, the AAG concentrations are 30–40% higher than those in unhospitalized patients, and the propranolol free fraction is reduced by approximately, 30% (109). A marked decrease in albumin concentration has been reported in malnutrition, and it seems associated with a reduction in protein synthesis (81).

In postoperative situations, with lower albumin and increased AAG concentrations, the unbound free fractions of propranolol (101) and quinidine (111) are 30–40% reduced and 100–150% higher for the albumin-bound phenytoin (112, 113). For drugs such as quinidine, which are restrictively cleared and bound to AAG, the concentration of which would rise after surgery, no effect on unbound drug concentration should be anticipated (111). In patients undergoing open-heart surgery, the unbound free fraction of ceftriaxone is up to four-fold higher compared with that in healthy volunteers because of low plasma albumin and high free fatty acid concentrations (49, 114).

Disease States

Renal disease

The most common cause of alterations in plasma albumin concentrations is hypoalbuminemia, which is associated with a variety of physiological and pathological conditions (see Table 2). This decrease in binding capacity results in an increase in unbound concentrations in plasma. Gugler et al. (115) reported a significant inverse correlation between albumin concentration and the

Table 3 Altered plasma protein binding of drugs associated with renal dysfunction

↓ Binding		↑ Binding	
Cephalosporins	Furosemide	Phenytoin	Chlorpromazine
Chloramphenicol	Midazolam	Prazosin	Lidocaine
Clofibrate	Morphine	Salicylic acid	
Diazepam	Naproxen	Sulfonamides	
Dicloxacillin	Penicillin G	Triamterene	
Diazoxide	Pentobarbital	Warfarin	
Diflunisal	Phenobarbital		
Digitoxin	Phenylbutazone		

(From Ref. 43.)

unbound plasma fraction of phenytoin in patients with nephrotic syndrome in chronic renal failure and uremia, a decrease in the binding of albumin-bound drugs was reported and related to the presence of endogenous inhibitors that accumulate (116, 117). A list of drugs associated with altered plasma protein binding in renal dysfunction is given in Table 3 (43).

Liver disease

In liver disease, the changes in drug binding may be related to a decrease in albumin and AAG concentrations attributable to either a decreased rate of synthesis or a loss of plasma proteins to the interstitial compartments. Changes may also be caused by an accumulation of endogenous inhibitors of drug binding such as bilirubin (59). For albumin-bound drugs such as diazepam and tolbutamide, the mean unbound fractions are 65–75% higher in chronic alcoholics (118) and correlate with albumin concentrations. For the AAG-bound erythromycin, the unbound fraction is two to four times higher in patients with cirrhosis compared with control subjects (119) and correlates with AAG concentrations. Increases in unbound fraction have been reported for prazosin, propranolol, morphine, diazepam, tolbutamide, phenylbutazone, phenytoin, quinidine, triamterene, theophylline, and lidocaine (48).

Pathologies associated with inflammatory conditions

AAG is a protein known as an acute phase reactant. As a result of inflammation, injury during physiological trauma, and stress (21, 39, 84, 88, 114), an increase in the binding capacity of AAG is commonly observed in patients. For a drug that is highly bound, an increase in AAG plasma levels will result in a significant decrease in unbound plasma fraction. Examples include propranolol (120), lidocaine (121), and disopyramide (34) after myocardial

infarction and propranolol and chlorpromazine (122) in Crohn's disease.

An increase in AAG concentrations after an acute myocardial infarction causes a decrease in the unbound plasma fraction of propranolol and lidocaine (nonrestrictively cleared drugs). With intravenous or oral propranolol, an increase in dosage regimen may be required (120). For lidocaine, it is recommended that normal infusion rates be used because total drug concentrations increase during long-term infusion, and the increase in binding indicates that the unbound fraction does not change significantly (121).

Cancer and burn injury

Cancer and burn injury are associated with increased concentrations of AAG and decreased concentrations of albumin in plasma. For AAG-bound drugs, a 20–30 % decrease in the unbound fractions of lidocaine (123), methadone (36), propranolol (31), and imipramine (124) is observed, whereas the unbound fractions of albumin-bound drugs increase for tolbutamide (30%) (123), diazepam (180%) (124), and phenytoin (150%) (125).

Diabetes mellitus

The decreased plasma binding of sulfisoxazole in patients with diabetes is related to in vivo glycosylation of albumin, whereas the decreased plasma binding of diazepam may be caused by high concentrations of free fatty acid displacers (126, 127).

Thyroid disease

A decrease in concentration of proteins and an increase in concentration of free fatty acids in plasma are observed in hyperthyroidism. For propranolol and warfarin, an increase of up to 20–30 % in the unbound plasma fractions has been reported (128).

Table 4 In vivo drug interactions caused by protein displacement

Drug	Displacer	Drug	Displacer
Bupivacaine	Lidocaine	Phenytoin	Salicylic acid
Carbamazepine	Valproic acid ^a		Tolbutamide
Ceftriaxone	Probenecid ^a		Phenylbutazone ^a
Diazepam	Valproic acid ^a		Valproic acid ^a
Methotrexate	Salicylic acid ^a	Valproic acid	Salicylic acid ^a
	Probenecid ^a	Warfarin	Diflunisal
	Sulfisoxazole ^a		Phenylbutazone ^a
Penicillin	Aspirin		
	Sulfamethoxy pyridazine		

^aDrugs also known to inhibit renal tubular secretion or hepatic metabolism of displaced drugs.
(From Ref. 43.)

Cystic fibrosis

Cystic fibrosis is associated with hypoalbuminemia (129). For theophylline, the unbound fraction in plasma is 30% higher in these patients compared with control subjects (129).

Drug Interactions

A change in the binding affinity of a drug may be related to the displacement by other drugs. Competitive displacement may result in an increase in the free fraction of drug in plasma, as the apparent dissociation constant (K_{Dapp}) for the displaced drug is then expressed by:

$$K_{Dapp} = \frac{K_D(1 + [I])}{K_I} \quad (17)$$

where [I] is the unbound concentration of the inhibitor in plasma, and K_I is the dissociation constant for the binding of the inhibitor.

There appear to be more cases of drug displacement involving drugs that are highly bound to albumin compared with drugs that bind to AAG and albumin. For example, the unbound free fraction of lidocaine can be increased significantly by therapeutic concentrations of bupivacaine of 3–4 mg/L (130, 131). No significant binding interaction has been reported for disopyramide, which binds almost exclusively to AAG (2). Other examples of in vivo displacement resulting in increased unbound fraction of drugs in plasma are listed in Table 4 (43).

In Vitro Artifacts

The use of vacutainer tubes and heparin was shown to alter the determination of protein binding. Heparin was shown

to decrease the plasma binding of certain drugs including phenytoin, propranolol, lidocaine, diazepam, quinidine, and verapamil (132–134). This is also an in vitro artifact attributable to continued ex vivo activity of the lipoprotein lipase enzyme and accumulation of fatty acids in the blood collection tube.

High unbound fraction of basic drugs was attributed to displacement from the AAG binding sites by a plasticizer in the cap stopper (135, 136). Since then, the stoppers have been reformulated. Storage containers and other anticoagulants have also been responsible for altering binding measurements. Binding inhibitors such as plasticizers in bags could be responsible for the alteration in the binding of disopyramide (2). The unbound fraction of phenytoin and meperidine is 80% higher in citrated plasma compared with serum or heparinized plasma (137).

Drug Regulatory Concerns

A New Drug Submission should provide information on the quality, safety, and efficacy of the drug product under the conditions of proposed use. Therefore, the chemical, pharmacokinetic, biopharmaceutical, pharmacological, toxicological, and clinical aspects should be investigated and reported. This includes detailed information on the protein binding, with an emphasis on populations at risk, such as with renal or hepatic insufficiency or when AAG may be altered. The ICH Common Technical Document on efficacy indicates the sections where proteins binding of a new active substance should be documented. The submission should include the preclinical and human pharmacokinetic information necessary to fully understand the biodisposition of the drug and its relation to the pharmacodynamic response and therapeutic and/or toxic effect. In addition, for drug–drug or other interactions, the

clinical importance of alterations in protein binding and the mechanism of the interaction should be determined, for example, if total or free concentration is affected only under initial displacement and whether these persist during chronic dosage.

It is also essential to include clinically significant protein binding information in the product monograph. In the clinical setting, the information should help explain the inter and inpatient variations in pharmacokinetics that would be reflected in variable pharmacodynamic effects and would certainly indicate when a dose or dosage adjustment is necessary. As noted above, changes in protein binding do not always require that adjustment in dosage regimens is needed, and there must be sufficient information for proper inferences to provide safe and effective therapy to patients.

REFERENCES

- McDevitt, D.G.; Frisk-Holmberg, M.; Hollifield, J.W.; Shand, D.G. *Clin. Pharmacol. Ther.* **1976**, *20*, 152–157.
- Lima, J.J. *J. Pharmacol. Exp. Ther.* **1981**, *219*, 741–747.
- Evans, W.E.; Rodman, J.H.; Relling, M.V.; Petros, W.P.; Stewart, C.F.; Pui, C.; Rivera, G.K. *J. Pharmacol. Exp. Ther.* **1992**, *260*, 71–77.
- Sebille, B. *Fund. Clin. Pharmacol.* **1990**, *4* (Suppl. 2), 151s–161s.
- Davis, B.D. *J. Clin. Invest.* **1943**, *22*, 753.
- Pacifici, G.M.; Viani, A. *Clin Pharmacokinet.* **1992**, *23*, 449–468.
- Sebille, B.; Zini, R.; Madjar, C.; Thuaud, N.; Tillement, J.P. *J. Chromatog. Biomed.* **1990**, *531*, 1–77.
- Tozer, T.; Gambertoglio, G.; Furst, D.; Avery, D.; Holford, N. *J. Pharm. Sci.* **1983**, *72*, 1442–1446.
- Rehberg, P.B. *Acta Physiologica Scandinavica* **1943**, *5*, 115–126.
- Bowers, W.F.; Fulton, S.; Thompson, J. *Clin. Pharmacokinet.* **1984**, *9* (Suppl 1), 49–60.
- Whitlam, J.B.; Brown, K.F. *J. Pharm. Sci.* **1981**, *70*, 146.
- Scott, D.O.; Bell, M.A.; Lunte, C.E. *J. Chrom.* **1990**, *506*, 461–469.
- Eklblom, M.; Hammarlund-Udenaes, M.; Lundqvist, T.; Sjöberg, P. *Pharm Res.* **1992**, *9* (1), 155–158.
- Herrera, A.M.; Scott, D.O.; Lunte, C.E. *Pharm. Res.* **1990**, *7*, 1077–1082.
- Sjöberg, P.; Olofsson, I.M.; Lundqvist, T. *Pharm. Res.* **1992**, *9*, 1592–1598.
- Wong, S.L.; Wang, Y.; Sawchuk, R.J. *Pharm. Res.* **1992**, *9*, 332–338.
- MacKichan, J.J.; Zola, E.M. *Br. J. Clin. Pharmacol.* **1984**, *18*, 487–493.
- Metzler, G.M. *A User's Manual for NONLIN and Associated Programs*; The Upjohn Co.: Kalamazoo, 1974.
- Putnam, F.W. *The Roster of the Plasma Proteins. The Plasma Proteins*; Putnam, F.W., Ed.; Academic Press: New York, 1975; 57–131.
- Sadee, W.; Gravel, R.C.; Lee, A.Y. *Pharm. Biotechnol.* **1999**, *12*, 29–58.
- Tillement, J.P.; Lhoste, F.; Giudielli, J.F. *Clin. Pharmacokinet.* **1978**, *3*, 44–154.
- Jusko, W.J.; Gretch, M. *Drug Metab. Rev.* **1983**, *14*, 427.
- Peters, T., Jr. *Serum Albumin. The Plasma Proteins*; Putnam, F.W., Ed.; Academic Press: New York, 1975; 133–181.
- Schmidt, K. α 1-Acid Glycoprotein. *The Plasma Proteins*; Putnam, F.W., Ed.; Academic Press: New York, 1975; 183–228.
- Lunde, P.M.K. *Inflammation and α 1-Acid Glycoprotein: Effect on Drug Binding. Drug-Protein Binding*; Reidenberg, M.M., Erill, S., Eds.; Praeger Publishers: New York, 1986; 201–219.
- Tinghly, D.; Baumann, P.; Conti, M.; Jonzier-Perey, M.; Schopf, J. *Eur. J. Clin. Pharmacol.* **1985**, *27*, 661–666.
- Wagner, J.G. *Protein Binding. Pharmacokinetics for the Pharmaceutical Scientist*; Technomic Publishing Co. Inc.: Lancaster, PA, 1993; 243–258.
- Urien, S.; Albengres, E.; Zini, R.; Tillement, J.P. *Biochem. Pharmacol.* **1982**, *31*, 3687–3689.
- Urien, S.; Albengres, E.; Pinquier, J.L.; Tillement, J.P. *Clin. Pharmacol Ther.* **1986**, *39*, 683–689.
- Milsap, R.L.; Jusko, W.J. *J. Steroid. Biochem.* **1983**, *18*, 191.
- Abramson, F.P.; Jenkins, R.N.; Ostchega, Y. *Clin. Pharmacol. Ther.* **1982**, *32*, 659–663.
- Baruzzi, A.; Contin, M.; Perucca, E.; Albani, F.; Riva, R. *Eur. J. Clin. Pharmacol.* **1986**, *31*, 85–89.
- David, B.M. *Br. J. Clin. Pharmacol.* **1983**, *15*, 435–441.
- Piafsky, K.M.; Borga, O. *Clin. Pharmacol. Ther.* **1977**, *22*, 545–549.
- Edwards, D.J.; Lalka, D.; Cerra, F.; Slaughter, R.L. *Clin. Pharmacol. Ther.* **1982**, *31*, 62–67.
- Abramson, F.P. *Clin. Pharmacol. Ther.* **1982**, *32*, 652–658.
- Lemaire, M.; Urien, S.; Albengres, E. *Lipoprotein Binding of Drugs. Drug-Protein Binding*; Reidenberg, M.M., Erill, S., Eds.; Praeger Publishers: New York, 1986; 201–219.
- Scanu, A.M.; Edelstein, C.; Keim, P. *Serum Lipoproteins. The Plasma Proteins*; Putnam, F.W., Ed.; Academic Press: New York, 1975; 317–391.
- Piafsky, K.M. *Clin. Pharmacokinet.* **1980**, *5*, 246–262.
- Lemaire, M.; Tillement, J.P. *J. Pharm. Pharmacol.* **1982**, *34*, 715–718.
- Hughes, T.A.; Gaber, A.O.; Montgomery, C.E. *Ther. Drug. Monit.* **1991**, *13*, 289–295.
- Lindholm, A.; Henricsson, S. *Ther. Drug Monit.* **1989**, *11*, 623–630.
- MacKichan, J.J. *Influence of Protein Binding and Use of Unbound Drug Concentrations. Applied Pharmacokinetics*; Evans, W.E., Schentag, J.J., Jusko, W.J., Eds.; Applied Therapeutics Inc.: Vancouver, WA, USA, 1992; 33–37.
- Gibaldi, M.; McNamara, P.J. *Eur. J. Clin. Pharmacol.* **1978**, *13*, 373–378.
- Fichtl, B. *Tissue Binding of Drugs: Methods of Determination and Pharmacokinetic Consequences. Plasma Binding of Drugs*; Belpaire, F., Bogaert, M.,

- Tillement, J.P., Verbeeck, R., Eds.; Academia Press: Gheny, 1991; 149–158.
46. D'Arcy, P.F.; McElnay, J.C. *Pharmacol. Ther.* **1982**, *17*, 211–220.
47. Klotz, U. *Clin. Pharmacokinet.* **1976**, *1*, 204.
48. Labaune, J.P. Fixation au Niveau Des Fractions Sanguines. *Pharmacocinetique*; Labaune, J.P., Ed.; Masson Publisher: Paris, 1984; 116–130.
49. Jungbluth, G.L.; Pasko, M.T.; Beam, T.R.; Jusko, W.J. *Antimicrob. Agents Chemother.* **1989**, *33*, 850–856.
50. Rowland, M.; Benet, L.Z.; Graham, G.G. *J. Pharmacokin. Biopharm.* **1973**, *1*, 123–136.
51. Wilkinson, G.; Shand, D. *Clin. Pharmacol. Ther.* **1975**, *18*, 377–390.
52. Morgan, D.J.; Smallwood, R.A. *Clin. Pharmacokinet.* **1990**, *8*, 61–76.
53. Wilkinson, G.R. Plasma Binding and Hepatic Drug Elimination. *Drug-Protein Binding*; Reidenberg, M.M., Erill, S., Eds.; Praeger Publishers: New York, 1986; 220–232.
54. Gupta, S.K.; Benet, L.Z. *Pharm. Res.* **1990**, *7*, 46–48.
55. Meijer, D.K.F.; van der Sluijs, P. *Pharm. Res.* **1989**, *6*, 105–118.
56. MacKichan, J.J. *Clin. Pharmacokinet.* **1984**, *9* (Suppl 1), 32–41.
57. MacKichan, J.J. *Clin. Pharmacokinet.* **1989**, *16*, 65–73.
58. Stoeckel, K.; Trueb, V.; Buback, U.C.; McNamara, P.J. *Eur. J. Clin. Pharmacol.* **1988**, *34*, 151–156.
59. Blaschke, T.F. *Clin. Pharmacokinet.* **1977**, *2*, 32–44.
60. Gross, A.S.; Heuer, B.; Eichelbaum, M. *Biochem. Pharmacol.* **1988**, *37*, 4623–4627.
61. Tozer, T.N. Implications of Altered Plasma Protein Binding in Disease States. *Pharmacokinetic Basis for Drug Treatment*; Benet, L.Z., Ed.; Raven Press: New York, 1984; 173–93.
62. Kunin, C.M. *Clin. Pharmacol. Ther.* **1966**, *7*, 180–188.
63. Patel, I.H.; Levy, R.H. *Epilepsia.* **1979**, *20*, 85–90.
64. Lin, J.H. *Clin. Pharmacokinet.* **1987**, *12*, 402–432.
65. Morgan, D.J.; Blackman, G.L.; Paull, J.D.; Wolf, L.J. *Anesthesiology* **1981**, *54*, 468–473.
66. Pearson, R.M. *Clin. Pharmacokinet.* **1977**, *2*, 198–204.
67. Meffin, P.J.; Robert, E.W.; Winkle, R.A.; Harapat, S.; Peters, F.A.; Harrison, D.C. *J. Pharmacokin. Biopharm.* **1979**, *7*, 29–48.
68. Tucker, G.T.; Boyes, R.N.; Bridenbaugh, P.O.; Moore, D.C. *Anesthesiology* **1970**, *33*, 287–314.
69. McNamara, P.J.; Slaughter, R.L.; Visco, J.P.; Elwood, C.M.; Siegel, J.H.; Lalka, D. *J. Pharm. Sci.* **1980**, *69*, 749–751.
70. Jusko, W.J.; Rose, J.Q. *Ther. Drug Monit.* **1980**, *2*, 169–176.
71. Lima, J.J. Disopyramide. *Applied Pharmacokinetics*; Evans, W.E., Ed.; Applied Therapeutics, Inc.: Spokane WA, 1986; 1210–1253.
72. Wallace, S.M.; Riegelman, S. *J. Pharm. Sci.* **1977**, *66*, 729–731.
73. Dieterle, W.; Wagner, J.; Faigle, W. *Eur. J. Clin. Pharmacol.* **1976**, *10*, 37–42.
74. MacKichan, J.J.; Zola, E.M. *Br. J. Clin. Pharmacol.* **1984**, *18*, 487–493.
75. Van der Berg, J.W.O.; Werhoef, M.L.; De Boer, A.S.H.; Schalm, S.W. *Clin. Chem. Acta.* **1985**, *147*, 291–297.
76. Smith, P.C.; McDonagh, A.F.; Benet, L.Z. *J. Clin. Invest.* **1986**, *77*, 934–939.
77. Hyneck, M.L.; Smith, P.C.; Munafo, A.; McDonagh, A.F.; Benet, L.Z. *Clin. Pharmacol. Ther.* **1988**, *44*, 107–113.
78. Simonyi, M.; Fitos, I.; Visy, J. *Trends Pharmacol. Sci.* **1986**, *7*, 112–116.
79. Evans, A.M.; Nation, R.T.; Sansom, L.N.; Bochner, F.; Somogyi, A.A. *Eur. J. Clin. Pharmacol.* **1989**, *36*, 283–290.
80. Jamali, F.; Mehvar, R.; Pasutto, F.M. *J. Pharm. Sci.* **1989**, *78*, 695–715.
81. Belpaire, F.M.; Vermeulen, A.M.; Bogaert, M.G. Stereoselectivity of Plasma Protein Binding and Its Pharmacokinetic Consequences. *Plasma Binding of Drugs*; Belpaire, F., Bogaert, M., Tillement, J.P., Verbeeck, R., Eds.; Academia Press: Gheny, 1991; 137–147.
82. Lima, J.J. *Life Sci.* **1987**, *41*, 2807–2813.
83. Levy, R.H.; Moreland, T.A. *Clin. Pharmacokinet.* **1984**, *9* (Suppl 1), 1–9.
84. Svensson, C.K.; Woodruff, M.; Baxter, J.; Lalka, D. *Clin. Pharmacokinet.* **1986**, *11*, 450–469.
85. Perucca, E. *Clin. Pharmacokinet.* **1984**, *9* (Suppl 1), 71–78.
86. Woosley, R.L.; Siddoway, L.A.; Thompson, K.; Cerkus, I.; Roden, D.M. *Clin. Pharmacokinet.* **1984**, *9* (Suppl 1), 79–83.
87. Lima, J.J.; Wenzke, S.C.; Boudoulas, H.; Schaal, S.F. *Ther. Drug Monit.* **1990**, *12*, 23–28.
88. Wilkinson, G.R. *Drug Metab. Rev.* **1983**, *14*, 427–465.
89. Bienvendu, J.; Sann, L. *Clin. Chem.* **1981**, *27*, 721–726.
90. Besunder, J.B.; Reed, M.D.; Blumer, J.L. *Clin. Pharmacokinet.* **1988**, *14*, 189–216.
91. Kurz, H.; Mauser-Ganshorn, A.; Stickel, H.H. *Eur. J. Clin. Pharmacol.* **1977**, *11*, 463–467.
92. Wood, M.; Wood, A.J. *J. Clin. Pharmacol. Ther.* **1980**, *29*, 522–526.
93. Morselli, P.L. *Clin. Pharmacokinet.* **1976**, *1*, 81–98.
94. Wallace, S.; Whiting, B. *Br. J. Clin. Pharmacol.* **1976**, *3*, 327–330.
95. Greenblatt, D.J. *Clin. Pharmacol. Ther.* **1980**, *27*, 301–312.
96. Davis, D.; Grossman, S.H.; Kitchell, B.B.; Shand, D.G.; Routledge, P.A. *Br. J. Clin. Pharmacol.* **1985**, *19*, 261–265.
97. Perucca, E.; Grimaldi, R.; Gatti, G.; Pirrachio, S.; Crema, F.; Frigo, G.M. *Br. J. Clin. Pharmacol.* **1984**, *17*, 665–669.
98. Kelly, J.G.; O'Malley, K. Drug-Protein Binding in Old Age. *Drug-Protein Binding*; Reidenberg, M.M., Erill, S. Eds.; Praeger Publishers: New York, 1986; 163–171.
99. Dean, M.; Stock, B.; Patterson, R.J.; Levy, G. *Clin. Pharmacol. Ther.* **1980**, *28*, 253–261.
100. Perucca, E. *J. R. Soc. Med.* **1981**, *74*, 422.
101. Bardy, A.H.; Hilesmaa, V.K.; Teramo, K.; Neuvonen, P. *J. Ther. Drug. Monit.* **1990**, *12*, 40–46.
102. Denson, D.D.; Coyle, D.E.; Thompson, G.A.; Santos, D.; Turner, P.A.; Myers, J.A.; Knapp, R. *Clin. Pharmacol. Ther.* **1984**, *35*, 702–709.
103. Feely, J.; Grimm, T. *Br. J. Clin. Pharm.* **1991**, *31*, 551–552.

104. Zhou, H.J.; Roshaki, R.P. *New Engl. J. Med.* **1989**, 565–570.
105. Zhou, H.J.; Adedoyin, A.; Wilkinson, G.R. *Clin. Pharm. Ther.* **1990**, 48, 10–17.
106. Ghoneim, M.M.; Kortila, K.; Chiang, C.K.; Jacobs, L.; Schoenwald, R.; Mewaldt, S.; Kayaba, K.-O. *Clin. Pharmacol. Ther.* **1981**, 29, 749–756.
107. Routledge, P.A.; Shand, D.; Barchowsky, A.; Wagner, G.; Stargel, W. *Clin. Pharmacol. Ther.* **1981**, 30, 154–157.
108. Benedek, I.H.; Fiske, W.D., III; Griffen, W.O.; Bell, R.M.; Blouin, R.A.; McNamara, P.J. *Br. J. Clin. Pharmacol.* **1983**, 16, 751–754.
109. Jagadeesan, V.; Krishnaswamy, K. *Eur. J. Clin. Pharmacol.* **1985**, 27, 657–659.
110. Feely, J.; Forrest, M.B.; Gunn, A.; Hamilton, W.; Stevenson, I.; Crooks, J. *Clin. Pharmacol. Ther.* **1980**, 28, 759–764.
111. Fremstad, D. *Eur. J. Clin. Pharmacol.* **1976**, 10, 441–444.
112. Elfstrom, J. *Clin. Pharmacokinet.* **1979**, 4, 16–22.
113. Elfstrom, J. *Acta. Neurol. Scand.* **1977**, 55, 455.
114. Jungbluth, G.L.; Pasko, M.T.; Jusko, W.J. *J. Pharm. Sci.* **1989**, 78, 807–811.
115. Gugler, R.; Azarnoff, D.L. *Clin. Pharmacokinet.* **1976**, 1, 25–35.
116. Reidenberg, M.M.; Drayer, D.E. *Clin. Pharmacokinet.* **1984**, 9 (Suppl 1), 18–26.
117. Niwa, T.; Takeda, N.; Maeda, K.; Shibata, M.; Tatematsu, A. *Clin. Chem. Acta* **1988**, 173, 127–138.
118. Thiessen, J.J.; Sellers, E.M.; Denbeigh, P.; Dolman, L. *J. Clin. Pharmacol.* **1976**, 16, 345–351.
119. Barre, J.; Houin, G.; Rosenbaum, J.; Zini, R.; Dhumeaux, D.; Tillement, J.P. *Br. J. Clin. Pharmacol.* **1984**, 18, 652–653.
120. Routledge, P.A.; Stargel, W.; Wagner, G.; Shand, D. *Br. J. Clin. Pharmacol.* **1980**, 9, 438–439.
121. Routledge, P.A.; Stargel, W.; Kitchell, B.B.; Barchowsky, A.; Shand, D. *Br. J. Clin. Pharmacol.* **1981**, 11, 245–250.
122. Piafsky, K.M.; Borga, O.; Odar-Cederlof, I.; Johansson, C.; Sjoqvist, F. *New Eng. J. Med.* **1978**, 299, 1435–1439.
123. Jackson, P.R.; Tucker, G.T.; Woods, H.F. *Clin. Pharmacol. Ther.* **1982**, 32, 295–302.
124. Martyn, J.A.; Abernethy, D.R.; Greenblatt, D.J. *J. Clin. Pharmacol. Ther.* **1984**, 5, 535–539.
125. Bowdle, T.A.; Neal, G.D.; Levy, R.H.; Heimbach, D.M. *J. Pharmacol. Exp. Ther.* **1980**, 213, 97–99.
126. Ruiz-Cabello, F.; Erill, S. *Clin. Pharmacol. Ther.* **1984**, 6, 691–695.
127. Erill, S.; Calva, R. *Post-Translational Changes of Albumin as a Cause of Altered Drug-Plasma Protein Binding. Drug-Protein Binding*; Reidenberg, M.M., Erill, S., Eds.; Praeger Publishers: New York, 1986; 220–232.
128. Feely, J.; Stevenson, I.H.; Crooks, J. *Clin. Pharmacokinet.* **1981**, 6, 298–305.
129. Prandota, J. *Drugs* **1988**, 35, 542–578.
130. McNamara, P.J.; Slaughter, R.L.; Pieper, J.A.; Wyman, M.G.; Lalka, D. *Anesth. Analg.* **1981**, 60, 395–400.
131. Goolkasian, D.L.; Slaughter, R.L.; Edwards, D.J.; Lalka, D. *Eur. J. Clin. Pharmacol.* **1983**, 25, 413–417.
132. Wood, M.; Shand, D.; Wood, A.J. *Clin. Pharmacol. Ther.* **1979**, 25, 103–107.
133. Brown, J.E.; Kitchell, B.B.; Bjornsson, T.D.; Shand, D.G. *Clin. Pharmacol. Ther.* **1981**, 30, 636–643.
134. Dube, L.M.; Davies, R.F.; Beanlands, D.S.; Mousseau, N.; Beaudoin, N.; Chan, B.; Ho-Ngoc, A.; McGilveray, I.J. *Biopharm. Drug Dispos.* **1989**, 10, 55–68.
135. Borga, O.; Piafsky, K.M.; Nilsen, O.G. *Clin. Pharmacol. Ther.* **1977**, 22, 539–544.
136. Midha, K.; Loo, J.C.K.; Rowe, M. *Res. Commun. Psychol. Psychiatry Behav.* **1979**, 4, 193–203.
137. Jackson, A.J.; Miller, A.K.; Narang, P.K. *J. Pharm. Sci.* **1981**, 70, 1168–1169.

PYROGENS AND ENDOTOXIN DETECTION

James F. Cooper

Charles River Endosafe, Charleston, South Carolina, U.S.A.

INTRODUCTION

A pyrogen is defined as a fever-producing agent. Pyrogens are substances that cause febrile reactions when sufficient amounts enter the circulatory system. Bacterial endotoxin is the most significant pyrogen because of its potency and ubiquity. The Bacterial Endotoxins Test has generally replaced the rabbit pyrogen test in the pharmaceutical applications.

Early Research Efforts Involving Intravenous Therapy and Pyrogens

Our awareness of pyrogens began with the advent of intravenous therapy by Sir Christopher Wren in 1656 (1). During the development of experimental medicine in the late 18th and early 19th century, many documented reports describe intravenous infusion therapy to humans that were frequently accompanied by febrile episodes. The mechanism of fever induction by infusion therapy was unknown. Since septic fever or wound fever occurred frequently when tissue from open wounds or surgical sites decomposed, physicians speculated that pyrogens might be formed in these tissues from processes such as fermentation or putrefaction. After Louis Pasteur (1822–1895) discovered that bacteria were the infectious agents that caused fermentation, there was additional thought that pyrogens were associated with bacteria. However, it was unclear whether the pyrogens were produced by bacteria or were inherently a part of bacteria. The history relevant to bacterial pyrogens in medicine has been reviewed elsewhere (2–4).

Pyrogen research beginning more than a century ago concentrated on the chemical, physiological, and pharmacological nature of bacterial pyrogens. In 1894, Centanni extracted pyrogenic substances from a large variety of bacteria, including *Escherichia coli* and typhoid (5). He also showed that these bacterial pyrogens were not proteins and that they were heat stable. Hort and Penfold's studies, published in 1912, provided the best understanding of the nature of injection fevers to that date (6–8). They used rabbits to standardize an assay for fever. By utilizing their

fever assay and the staining procedure of Gram, they demonstrated that the pyrogenic bacteria were predominately gram-negative, whereas the nonpyrogenic types were gram-positive. They were usually able to correlate the pyrogenic activity of the distilled water used in their studies with its bacterial count and showed that dead Gram-negative bacteria (GNB) were also capable of inducing pyrogenicity. They concluded that the cause of all injection fevers was a filterable, heat-stable bacterial substance.

The classic investigations of Seibert during 1923–1925 established conclusively that injection fevers were caused by heat-stable, filterable components of GNB (4, 9–11). She also successfully developed a process to consistently produce infusion fluids that could pass rabbit pyrogen assays.

Advent of Large Volume Parenteral (LVP) Solutions

Seibert's methods for producing nonpyrogenic intravenous (IV) fluids enabled hospital pharmacies to produce solutions that were safe for patient therapy (10, 11). Her manufacturing methods also launched the commercial LVP solutions industry in the decade preceding World War II. Production of today's commercial LVP solutions relies on the availability of large volumes of nonpyrogenic water, known as Water for Injection (WFI), which is typically generated by distillation. Raw material ingredients are screened for the absence of pyrogen and added to WFI. After nonpyrogenic ingredients are verified, they are mixed with the WFI in depyrogenated tanks and dispensed through filling equipment into depyrogenated containers. These filled containers are immediately autoclaved to sterilize the solution and prevent the growth of any GNB, pyrogen-producing organisms that might be present in the mixture.

NATURE OF PYROGENS

The term "pyrogen" is frequently used to describe the pyrogen most significant to the pharmaceutical industry, GNB endotoxin. Since pyrogen is a general term for any substance that causes fever after IV administration or

inhalation, it is important to differentiate nonmicrobial and microbial pyrogens. Nonmicrobial pyrogens include pharmacological agents like bleomycin, colchicine, and polynucleotide poly-I:C. For sensitized hosts, there can be antigens of human blood products, such as human serum albumin, penicillin drugs, or other therapeutic agents (12, 13).

Microbial pyrogens are by far the most significant problem for pharmaceutical manufacturers. Bacteria (killed and live), fungi, plasmodia (malarial parasites), and viruses (live) can all act as pyrogens. Additionally, bacterial products, including streptococcal exotoxins, staphylococcal enterotoxin and bacterial endotoxin lipopolysaccharide (LPS), as well as fungal products, can also act as pyrogens (13). Because the LPS component in the endotoxin is toxic in very small quantities, GNB endotoxin is the microbial pyrogen most significant for pharmaceutical producers, as mentioned above. The amount of USP Reference Standard Endotoxin needed to initiate pyrogenicity in humans and rabbits is about 1 ng/kg (14). Whole cells induce pyrogenicity when large numbers are phagocytized by the macrophages they encounter as they enter the bloodstream. Therefore, although it requires the administration of at least 10,000 organisms/kg of the most pyrogenic GNB bacteria to cause a pyrogenic reaction in rabbits (15), the number of gram-positive or fungal organisms required to induce the same effect is orders of magnitude higher, $10\text{--}10^8/\text{kg}$ (12, 16, 17).

Mechanism of Fever Induction

Endotoxin pyrogen induces fever by an indirect process. On entry into the circulatory system, endotoxin is bound to LPS-binding protein (LPB) that transports it to receptor cells in the reticuloendothelial system. The main target cells are circulating mononuclear cells, which produce proinflammatory cytokines such as interleukin-1 (IL-1), interleukin-6 (IL-6), and tumor necrosis factor- α (TNF α) (18,19). These cytokines are involved in acute and chronic inflammation, induce fever, and modulate the host's response to bacterial infection (20).

Bacterial Endotoxin Pyrogen

Bacterial endotoxin is a high molecular weight complex that constitutes an integral component of the outer cell wall membrane of GNB. These bacteria constantly shed endotoxin into the environment as they grow and multiply, as well as when they die and disintegrate (21). Consequently, bacterial endotoxin can either exist in a cell-associated state or in a free state. Endotoxin that

remains cell-associated can be removed from a solution by microporous sterilizing filters, but endotoxin in a free state easily passes through most sterilizing filters. Since bacterial endotoxin is heat stable, it is not fully destroyed by a routine autoclave process. Endotoxin is ubiquitous and can be found wherever GNBs exist. Endotoxin is found in the soil, in food, in ground waters, and on any surface touched by these substances. Since the vast majority of bacteria that grow and multiply in water systems are gram-negative, bacterial endotoxin is a normal constituent of ordinary water. For this reason, it is a common contaminant in laboratories and in wet manufacturing areas, including laboratory glassware, research equipment, and water baths. It can contaminate sampling equipment, storage containers, and any reusable materials that retain moisture.

Chemical Nature and Structure of Endotoxins

Naturally occurring bacterial endotoxins contain the lipid, carbohydrate, and protein makeup of the outer cell membrane of GNB (see Fig. 1). However, most of the commercial endotoxin preparations have been purified by various extraction procedures and are generally free of nucleic acids, proteins, phospholipids, and other bacterial cell components (21). The primary chemical configuration that remains after purification is a polysaccharide structure that is covalently bound to a lipid component called Lipid A. Based on its chemical nature, which is common to various bacterial families, this substance is referred to as lipopolysaccharide (LPS). Although the terms endotoxin and LPS are often used interchangeably, most reference "endotoxin" standards are purified preparations that are more correctly described as LPS.

The Lipid A component of LPS is embedded in the outer membrane of the bacterial cell, whereas the polysaccharide protrudes into the environment (21). The polysaccharide component is composed of two parts, the core oligosaccharide that is connected to Lipid A and a longer oligosaccharide O-specific chain that is attached to the core (Fig. 2). The O-specific chain is the most variable component of the complex. It consists of 20–40 repeating units that include up to eight sugars, and it is responsible for the specific immune reaction that each type of GNB is able to evoke in a host. The core oligosaccharide is much less variable than the O-specific chain, and its influence on the host is less profound, although it can trigger antibody production in response to mutant endotoxins that lack an O-specific chain. The core

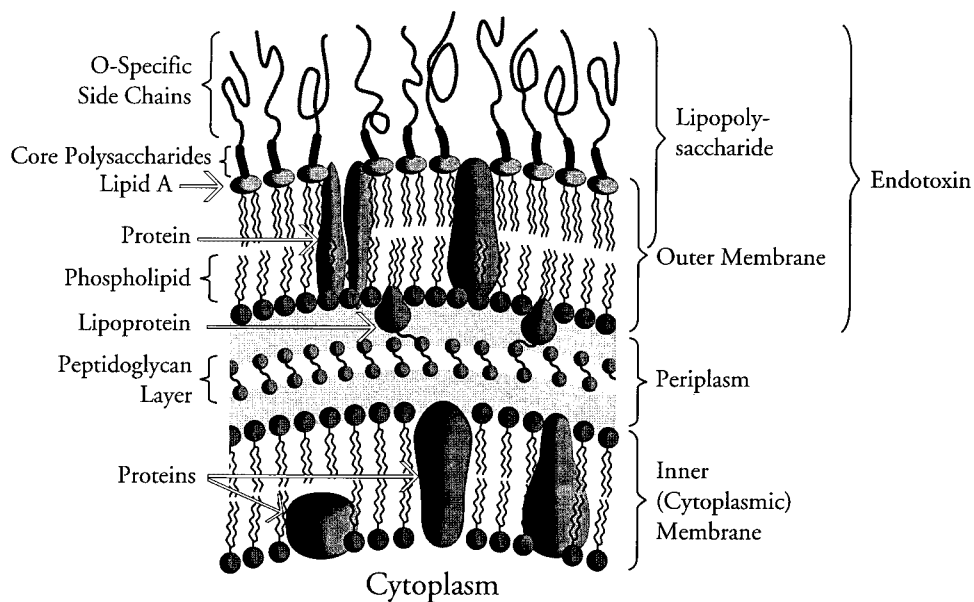


Fig. 1 Cell membrane of a gram-negative bacteria, including the lipopolysaccharide, protein, and phospholipid of the outer membrane (O-specific = oligosaccharide-specific).

is divided into the inner core that attaches to Lipid A and the outer core that links up to the O-specific chain. The inner core is the more interesting of the two segments because it bears two unusual sugars, a seven-carbon

heptose and 3-deoxy-D-manno-2-octulosonic acid KDO. KDO links the polysaccharide core to Lipid A and is found in all endotoxins, but it occurs nowhere else in nature, except in certain plants and algae (22).

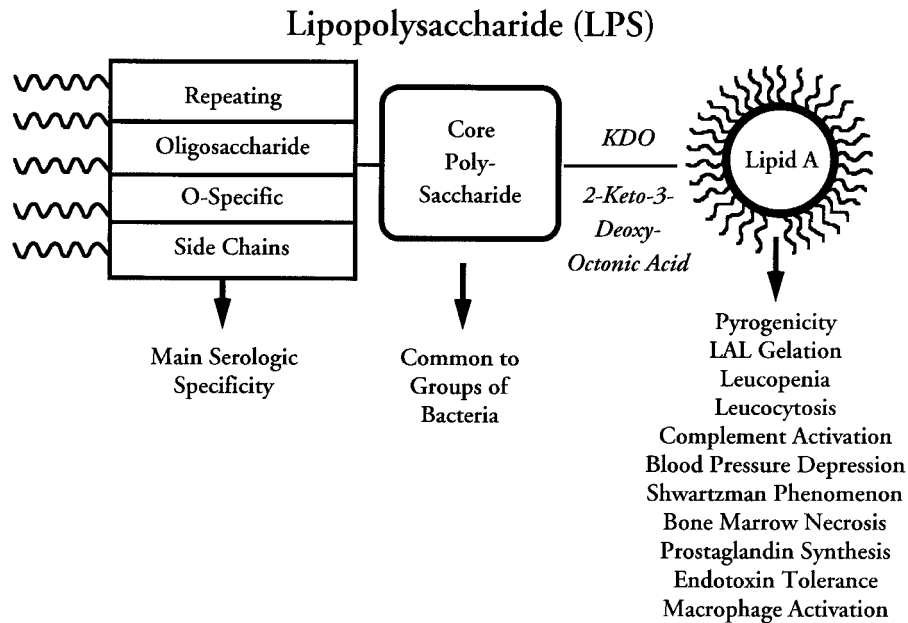


Fig. 2 The major components of lipopolysaccharide (LPS).

Lipid A is the least variable component of LPS. It consists of a disaccharide of glucosamine, which is highly substituted with amide-linked and ester-linked long-chain fatty acids, each with a backbone of about 14 carbon atoms (21). Lipid A is responsible for the vast majority of both the harmful and the beneficial biological activities that have been attributed to endotoxin. A partial list of these activities is shown in Fig. 2.

Biological Properties of Endotoxins

Bacterial endotoxins have fascinated investigators for years, in part because of their extensive and diverse biological properties and activities. Endotoxins are potent substances, which elicit a broad spectrum of the harmful physiologic responses that are produced in hosts by pathogenic GNB (21). At the extreme, they can produce profound alterations in organ function, such as hypotension and disseminated intravascular coagulation, which can lead to severe morbidity or death. On the other hand, endotoxins are also active stimulators of the mammalian defense system, which can enhance the body's capacity to cope with both microbial infections and malignant tumors. This dichotomy of biological activity continues to intrigue research scientists who are attempting to modify the Lipid A component of endotoxin in such a way that its toxic effects can be reduced or eliminated while its beneficial effects are retained (23).

Fever

Fever is one of the dramatic biological effects produced by endotoxin and one of the easiest to measure. For this

reason, it is one of the most studied and best understood of the many physiological activities that are initiated by endotoxin. The current understanding of the pathogenesis of fever is shown in Fig. 3 (21). When exogenous pyrogens enter the blood stream, they are removed from circulation by phagocytosis. The host's phagocytic cells (primarily peripheral monocytes) are thereby stimulated to synthesize a family of cytokines, or endogenous pyrogens, that are released into the circulation. These endogenous pyrogens travel to the hypothalamus, the thermoregulatory center of the body, via arterial blood supply. There the endogenous pyrogens induce various cells to increase the level of arachidonic acid metabolites (primarily the cyclooxygenase-derived prostaglandins, prostacyclins, and thromboxanes). There is considerable evidence that prostaglandin E₂ (PGE₂) is the major arachidonic acid metabolite associated with increasing the hypothalamic thermostat to febrile levels. During fever, levels of PGE₂ are elevated in the cerebral spinal fluid; the ability of aspirin and other antipyretics to reduce fever is directly

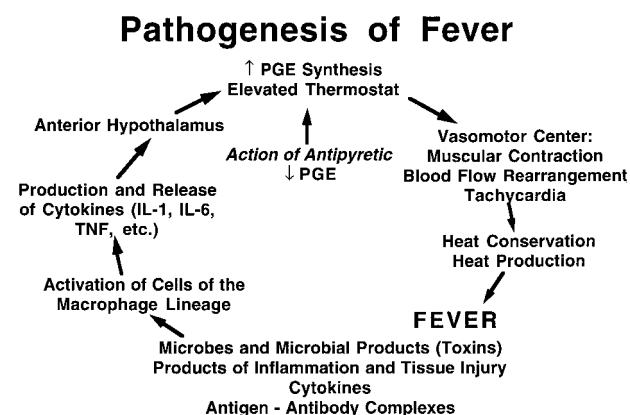


Fig. 3 The sequential events that produce fever in humans and animals.



Fig. 4 Horseshoe crabs are placed in a restraining rack and passively bled for a few minutes. They are promptly returned to the sea unharmed. Amebocyte blood cells are concentrated, washed, and lysed to produce LAL reagent. (Photo provided by J. Cooper, Endosafe, Charleston, SC.)

related to their ability to block brain cyclooxygenase. The new thermostatic setting signals the nerves that stimulate peripheral blood vessels to constrict those vessels and conserve body heat. This causes the body to chill, and the discomfort that results stimulates other physiologic mechanisms within the body to generate more heat in order to return the body to a homeostatic condition. The end result is fever (18, 24).

The systemic response to threshold pyrogenic doses of endotoxin in humans is described in studies that were conducted by the National Institutes of Health. The USP Reference Standard Endotoxin (RSE), *Escherichia coli* 0113:H10:K0, was injected into healthy volunteers at dosages ranging from 2 to 4 ng/kg (25). About 1 h after administration, a monophasic rise in temperature was observed, accompanied by chills and rigors. The chills lasted for 15 to 30 min and were followed by varying degrees of myalgias, arthralgias, headache, and nausea over the next 2 h. During this period, the subjects appeared sallow and exhibited general malaise. The peak core temperature (38.5–40°C) occurred at 3 h, followed by defervescence and symptomatic improvement over the next 3–5 h.

Septic shock

As mentioned earlier, bacterial endotoxin is a concern to the pharmaceutical industry because of its ability to induce pyrogenic reactions at intravenous dosages as low as 1 ng/kg. In higher dosages, endotoxin can be lethal. In one report, a self-administered intravenous dose of 1 mg of *Salmonella minnesota* endotoxin initiated the full clinical manifestations of septic-shock syndrome in a middle-aged laboratory worker (26). Septic-shock syndrome includes a high-cardiac-output form of hypotension, disseminated intravascular coagulation, abnormalities of hepatic and renal function, and noncardiogenic pulmonary edema. This patient was diagnosed and treated promptly and survived the episode.

Beneficial biological effects of endotoxins

Tumor reduction was attributed to endotoxin (27). Other beneficial effects of endotoxins on the host defense system include enhanced immune protection, protection from lethal irradiation, enhanced nonspecific resistance to infection, and resistance to toxic doses of endotoxin (tolerance). A hypothesis has been proposed (28) to explain the seemingly contradictory findings of the toxic vs. beneficial effects of endotoxins. It suggests that they are the primary signals that animals use to detect potentially harmful GNB and that the host defense system employs many of the body's defenses to both detect and to react against the threat of bacterial infection. This hypothesis also proposes

that the body possesses feedback mechanisms that, in most cases, attempt to match the bacterial endotoxin threat with an appropriate response. Thus, the intensity of the responses to endotoxins has evolved to maximize protection while minimizing the biological cost and self-damaging effects to the host. In general, the benefits of these responses far outweigh the risks, and when harmful events do occur, they are examples of inappropriate activity of overly protective systems (28).

Physical Properties of Endotoxins

It is important to be aware of the various physical properties of endotoxins in order to understand why they change their behavior when placed into different environments. An understanding of these physical properties is also essential for designing effective processes for depyrogenation.

Heat stability

It has long been known that endotoxins are thermostable in the presence of moist heat and that they are not appreciably destroyed by routine autoclaving processes. Early research showed that boiling was not completely effective (9). However, endotoxins can be destroyed by dry heat at temperatures above 180°C. In fact, dry heat is the method of choice for depyrogenating heat-resistant materials, such as glass and equipment.

Size

The size of endotoxin is dependent on its aggregation state, which, in turn, is dependent on its surrounding environment. In an aqueous environment, LPS is arranged in a bilayer. Its hydrophilic Lipid A components are clustered in the center of the bilayer, whereas its hydrophilic components are exposed to the surrounding solution. Bacterial LPS is stabilized by divalent cations, and in the presence of magnesium and calcium, it forms bilayer sheets and vesicles with a diameter of about 100 nm (21). However, if the divalent cations are removed from the environment, the bilayer breaks down into micellar forms that are 20–70 nm long and about 3–7 nm thick. In the presence of detergents, these micelles can be even further broken down into subunits of 0.8–1.2 nm in diameter and 10–60 nm in length.

The tendency for endotoxin to aggregate into larger and larger entities in an aqueous environment is due to the attraction that the hydrophobic groups of the LPS molecules have for one another. Aggregated endotoxin vesicles of 0.1 μm in diameter have been visualized by using electron microscopes. Even in its largest aggregation state, endotoxin passes through 0.22 μm sterilizing

filters. Conventional reverse-osmosis membranes that are nominally rated at pore sizes on the order of 1.0 nm are able to remove endotoxin from water (29). This process also removes any salts that may be present.

Molecular weight

In a typical aqueous environment that contains small amounts of divalent cations, endotoxin has a molecular weight of about 10^6 Da (21). If divalent cations are removed from the aqueous environment by chelators, the endotoxin bilayers break down into micelles of 300,000 to 1,000,000 molecular weight. When these micelles are further broken down in the presence of surface-active agents, their molecular weight drops again to about 10,000–20,000. However, these steps are completely reversible; if the detergents are dialyzed out, and divalent cations are added back to the endotoxin, the micelles and then the membranous structures reassemble themselves (30).

As the salt concentrations in water increase, endotoxin forms larger and larger molecular weight aggregates. The aggregation state of endotoxin affects both its solubility and biological reactivity. As the molecular weight of endotoxin increases, its solubility in water decreases. Additionally, as the molecular weight of endotoxin increases, toxicity in rats, rate of clearance from blood, interaction with complement, and affinity for cells also increase. However, endotoxin lethality in mice decreases as its molecular weight increases, as does its pyrogenicity in rabbits (31).

Electrostatic properties

At pH levels above 2, endotoxin aggregates are negatively charged and behave as anions (21). This property accounts for the attraction that endotoxin aggregates have for divalent cations in solutions. It is also the characteristic that provides the mechanism of action for endotoxin removal by cationically charged adsorbents.

PYROGEN TESTING

The development of the large volume parenteral (LVP) drug industry prompted the need for pyrogen testing to assure their safety. Pyrogen contamination is a greater problem for manufacturers of LVPs than it is for producers of small volume injectables because the initiation of patient fevers by parenteral solutions is dose dependent rather than concentration dependent. In other words, the onset and extent of injection fever depends on the total amount of pyrogen delivered to a patient and not on the concentration of pyrogen per milliliter of drug. Therefore,

an LVP must meet a more stringent standard for nonpyrogenicity than lesser volume drugs.

The heavy demand for LVP drug therapy prior to and during World War II and the need to ensure that commercial infusion fluids were free from pyrogens caused the United States Pharmacopeia (USP) to undertake the development of a compendial test for pyrogens (32). In 1941, the Committee of Revision of the USP authorized Subcommittee 3 on Biological Assays to begin the first USP collaborative study of pyrogen. Using *Pseudomonas aeruginosa* filtrates, prepared by the Division of Bacteriology of the FDA, the collaborative study was undertaken by the FDA, the NIH, and 14 pharmaceutical manufacturers (33,34). The study utilized the rabbit pyrogen test used earlier by Hort and Penfold and Seibert and her co-workers (7–9). Large numbers of rabbits were challenged with both pyrogenic material and physiologic saline solution. The results of the study, which were published in 1943, led to the inclusion of the first compendial pyrogen test in the 12th edition of the USP in 1942. Although refinements have been added from time to time, the USP rabbit pyrogen test remains relatively unchanged from the original format.

Basically, the pyrogen test involves measuring the rectal temperature of rabbits, both prior to and after the intravenous injection of a test solution in the ear veins. If the animals exhibit febrile responses that exceed established limits, the test solution is judged to be pyrogenic. Rabbits became the animal of choice because they are relatively inexpensive, are easy to handle, and have a labile thermoregulatory mechanism. Rabbits frequently produce false-positive pyrogen tests. For this reason, a negative result is more significant than a positive test, which makes the rabbit a good choice for assuring the absence of pyrogen in a test solution (32).

In 1969, Greisman and Hornick compared three purified endotoxin preparations on a dose-per-weight basis in rabbits and healthy adult males (35). Their results showed that rabbits and humans require approximately the same amount of endotoxin on a weight basis to induce threshold pyrogenic responses.

USP Rabbit Pyrogen Test

The USP <151> Pyrogen Test (36) is designed for solutions that can be tolerated by the rabbit in doses that do not exceed 10 ml/kg body weight and can be delivered within a time frame that does not exceed 10 min. Exceptions to these requirements are given in individual USP product monographs or federal regulations for biologics. Specifications for pyrogen test material handling,

calibration limits for recording equipment, and requirements for housing and conditioning the rabbits are also given in the <151> test chapter.

For the initial pyrogen tests, groups of three healthy, mature rabbits are chosen. Accurate temperature-sensing devices, such as clinical thermometers or thermistor probes, are inserted into the rectum of the rabbits to record their body temperature. If these probes remain inserted throughout the test period, the rabbits must be restrained with light-fitting neck stocks that permit them to move about. During the test, food but not water is withheld from the animals.

A control temperature is determined not more than 30 min prior to injection of the test dose. This is the base for determining any temperature increase resulting from the injection. Test solutions are warmed to $37 \pm 2^\circ\text{C}$ prior to injection. After injection, rabbit temperatures are recorded at 30-min intervals between 1 and 3 h. Temperature decreases are considered as zero rise. If no rabbit shows an individual temperature rise of 0.5°C or more above its control temperature, the product meets the requirements for the absence of pyrogens. If any rabbit shows an individual temperature rise of 0.5°C or more, the test is continued with an additional five rabbits. If not more than three of the eight rabbits show individual temperature rises of 0.5°C or more, and if the sum of the eight individual maximum increases does not exceed 3.3°C , the material under examination meets the USP <151> requirements for the absence of pyrogens (36).

Human Cell-Based Pyrogen Test

Pyrogens induce human monocytes to release proinflammatory cytokines such as IL-1, IL-6, IL-8, and TNF- α , as previously discussed. Test methods have been designed that include incubation of a test sample with monocytes in whole blood or in cultured cell lines and analysis of a specific cytokine after a suitable time. This cell-based methodology may provide an alternative to rabbit pyrogen testing that is required for human blood products. More development is needed to determine the optimum cytokine for analysis and to assess the nature of interference conditions (37, 38).

BACTERIAL ENDOTOXINS TEST (BET)

A *Limulus* amebocyte lysate (LAL) reagent is the basis for an in vitro pyrogen test method that is specific for bacterial endotoxin pyrogen. For this reason, it is now referred to as the bacterial endotoxins test (BET), although BET and

LAL testing are used interchangeably. When it was first introduced, there was concern by the industry and regulators that its specificity would limit its application in the parenteral industry (39). However, experience gained in this industry over the past 25 years confirms that endotoxin is the principal pyrogen of concern to pharmaceutical and medical device manufacturers. The BET has steadily gained acceptance globally as a replacement for the rabbit pyrogen test.

Discovery of *Limulus* Amebocyte Lysate (LAL) Reagent

The LAL test reagent is prepared by lysing amebocyte blood cells obtained from the American horseshoe crab, *Limulus polyphemus*. The origin of LAL reagent is traced to Frederick Bang, who first recognized the association between bacterial endotoxin and *Limulus* blood coagulation in studies of marine organisms at the Marine Biological Laboratory in Woods Hole, MA. An interest in immunity and infectious disease led Bang to study how *Limulus* would respond to an injection of bacteria. He observed that gram-negative *Vibrio* bacteria and its extracted endotoxin caused death in horseshoe crabs, not from infection but from intravascular coagulation (40).

Jack Levin, who had an interest in the effects of endotoxin on platelets and human blood coagulation, joined Bang at Woods Hole, MA, for several summers to do research on endotoxin-initiated *Limulus* blood coagulation. This collaboration produced several publications that explained a simple mechanism of enzyme-mediated interaction of LAL with endotoxin (41, 42). Levin discovered a way to harvest the intracellular fluid by osmotic lysis of stabilized amebocytes. Modern-day preparation of LAL reagent still follows Levin's basic methods.

Levin's original interest in the LAL test was for its potential as a clinical diagnostic tool for endotoxemia. However, a collaboration between Levin and Cooper produced the test's most celebrated application as a pyrogen test for the parenteral industry. The advent of short-lived radioactive drugs in the late 1960s prompted a need for a pyrogen test that was quick and required only a small volume of test material. As Cooper searched for an alternative endotoxin test for these drugs, he compared LAL sensitivity with rabbit response. The threshold pyrogenic level in rabbits was determined for LPS derived from *Escherichia coli* and *Klebsiella*. An excellent correlation was found between rabbit febrile response (fever index) and LAL reactivity (gel time), which

indicated that the *in vitro* test was indeed an indicator of a biological response (43). The LAL test yielded unmatched sensitivity and reproducibility.

Cooper and Mills set up the first commercial LAL production facility in 1971 in Chincoteague, VA, for Mallinckrodt, Inc. Production sites for LAL reagent are located on the eastern coast of the United States where horseshoe crabs are found abundantly in shallow coastal waters. The LAL test succeeded as the first “alternative to an animal test” because of amazing similarity between *Limulus* and mammalian response to endotoxin (44). The activation of human monocytes in blood by endotoxin to produce proinflammatory cytokines compares strongly to the activation of amebocytes in *Limulus* to produce clotting enzymes. As LAL reagents became commercially available, the pharmaceutical industry began to investigate the use of LAL testing as an alternative to the rabbit pyrogen test.

LAL clotting mechanism

The LAL test is based on the primitive blood-clotting defense mechanism that protects the horseshoe crab from the hostile sea of GNB that surrounds it. When a crab is wounded and invading bacteria enter its blood stream, its amebocyte blood cells respond by releasing granules that contain a coagulogen protein substance. Intracellular serine protease zymogens in the crab’s blood are triggered by the presence of endotoxin to initiate a series of activations that subsequently produce a coagulin gel clot from this coagulogen protein. The sequential activation of Factor C, Factor B, and the clotting enzyme in this coagulation cascade results in an enormous amplification in sensitivity of LAL for endotoxin.

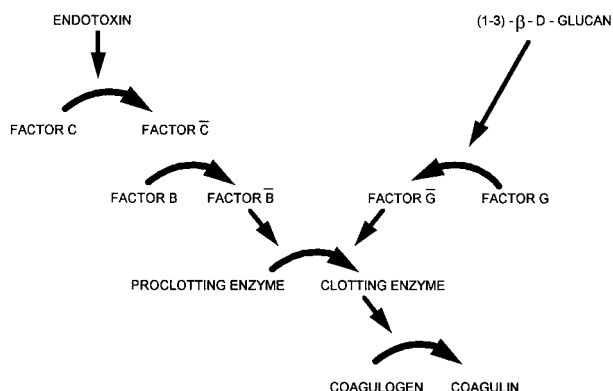


Fig. 5 Stages in the blood coagulation system of horseshoe crab amebocytes.

Figure 5 shows a coagulation scheme proposed by Nakamura (45).

The horseshoe crab is also exposed to fungus and yeast in its environment, which may explain why Factor G responds to a (1→3)-β-D-glucan, a component from the cell wall of these organisms. This pathway gives rise to a nonendotoxin activation of the LAL coagulation system (46).

Bacterial Endotoxins Test Methods

The observation of gel formation in a test tube as an endpoint for an endotoxin assay provides the means for a very simple test. The need for more objective quantitative methods has led scientists to develop a variety of automated methods for endotoxin measurement. This discussion is limited to the three methods that were accepted by the Food and Drug Administration (FDA).

Gel-clot BET

The gel-clot end point is the most commonly used endotoxin test. It is simple and requires minimal laboratory equipment and facilities. Equal volumes of test solution and LAL reagent (usually 0.1 ml each) are mixed in 10 × 75 mm glass test tubes. After incubation at 37°C for 1 h, the tubes are observed for clot formation after inverting them. Formation of a solid gel clot that withstands inversion of the tube constitutes a positive test. Each lot of gel-clot reagent licensed by the FDA must be labeled with its sensitivity (λ) to the reference standard endotoxin (RSE). The test can be used qualitatively to judge samples as positive or negative at the reagent’s sensitivity; this method is called a Limit test. The test may be also used as a semi-quantitative assay by titrating positive samples to an endpoint and multiplying the last positive sample dilution by the labeled sensitivity.

Kinetic turbidimetric assay

During the LAL-endotoxin reaction, the solution mixture becomes increasingly turbid. The kinetic turbidimetric assay (KTA) requires an incubating microplate or tube reader driven by an endotoxin-specific software. The reaction mixtures in a KTA system are continually monitored for changes in optical density in each sample that are caused by scattering and absorption of light. Generally, the KTA method measures the onset time needed to reach a predetermined absorbance by each reaction mixture. The onset times of samples are compared with those of endotoxin standards to yield quantitative values for each sample or control that contains endotoxin.

Kinetic and endpoint chromogenic assays

Unlike previously described tests, chromogenic LAL tests do not utilize the coagulogen protein from LAL reagent to produce an endpoint. Although endotoxin activates the same enzymatic cascade from the reagent, as previously described, the clotting enzyme reacts with a synthetic substrate that has been added to the reaction mixture. The substrate consists of a colorless amino acid chain attached to a chromophore. The activated clotting enzyme cleaves the bond that holds the chromophore to the amino acid; the amount released is proportional to the concentration of endotoxin (47). The chromophore that is released changes the color of the reaction mixture, thereby increasing the optical density. In the kinetic chromogenic assay (KCA), the reaction times of assay mixtures are determined with the same methods and readers used for KTA determinations. Endpoint chromogenic reactions may be done with nonincubating readers, where LAL, substrate, and quenching agent are added sequentially. After the reaction is terminated, the absorbency is read over a one-log range. In contrast, the kinetic turbidimetric and chromogenic assays may be conducted over a 2-to-4 log range.

Compendial and Regulatory Status of the BET

As confidence grew in the capability of the LAL test, as a screening tool for endotoxin pyrogen, pharmaceutical and medical device industries began to replace the rabbit test with the new *in vitro* test. The USP has continually revised the BET to keep abreast of advancements in LAL methodology.

The USP <85> Bacterial Endotoxins Test

The <85> BET first appeared in 1980 as an informational chapter in the *USP XX*. Significantly, the USP adopted the FDA *E. coli 0113* endotoxin standard as its reference standard and assigned units of potency to it (48). It also included information about calibrating a control standard endotoxin (CSE) to the RSE.

The first large-scale conversion of the rabbit test to the BET occurred in Supplement 5 to *USP XXII* when the BET became the official endotoxin test for 185 USP articles. The *USP 24* has endotoxin limits for over 650 USP articles.

Endotoxin limit

Endotoxin limits were introduced by the FDA to ensure the absence of unsafe levels of endotoxin in parenteral

products. Because endotoxin is ubiquitous, it was necessary to assign an allowable, safe limit that was below the threshold dose of endotoxin for pyrogenicity. In 1983, the USP officially replaced the <151> Pyrogen Test in USP monographs for 29 radiopharmaceuticals and five pharmaceutical waters. The USP subcommittee, responsible for revising the general chapters of the compendium, announced in 1987 its intent to replace the <151> Pyrogen Test with the BET for all USP articles for which the BET could be validated (48). Endotoxin limits were applied to USP articles using the *K/M* formula that was devised by the FDA. The endotoxin tolerance limit for humans, *K*, was set at 5 USP endotoxin units (EU) per kilogram of body weight (5 EU/kg), and *M* was defined as the maximum dose of the substance administered to an individual per kilogram per hour. The maximum dose is usually taken from the labeling of the parenteral manufacturer. The endotoxin limit for a substance is expressed in EU/mg or unit of product if it is administered on a weight basis, such as a small volume parenteral. Infusion solutions and medical device extracts have an endotoxin limit of 0.5 EU/ml. Drugs that are dosed on a volume basis also have an endotoxin limit in EU/ml.

Endotoxin limit for cerebrospinal fluid administration

Intrathecal administration applies to parenteral drugs that are infused directly into cerebral spinal fluid spaces. The tolerance limit, *K*, for these drugs is 0.2 EU/kg because intrathecal administration is the most toxic route for endotoxin pyrogen. An outbreak of aseptic meningitis followed intraspinal administration of nuclear imaging agents that were contaminated with endotoxin, but had passed the USP Pyrogen Test (49). One-milliliter volumes of the agents that had less than 10 EU per dose produced serious patient reactions and prompted the tighter endotoxin limit.

Harmonized BET

A harmonized text for <85> bacterial endotoxins test (BET) became effective in January of 2001 (50). It was a product of the International Conference on Harmonization (ICH). It was drafted by the Japanese Pharmacopeia and agreed on by the European Pharmacopeia and the USP. In adopting the harmonized BET, the USP chapter introduced radical changes. The text was simplified to make it easier to understand and follow. Whereas the previous version was a gel-clot test only, the new BET provides standardization for both gel-clot and photometric LAL methods. Any validated method may be used for a USP article however, the gel-clot method is the referee test in the unlikely case of a dispute.

FDA Validation Guideline for the LAL Test

The FDA was actively involved from the outset in developing the BET as an alternative to pyrogen testing. In 1972, a collaboration of Cooper with the FDA Bureau of Biologics (now Center for Biologics Evaluation and Research) established LAL methodology within the FDA. A study that compared LAL and rabbit tests on a group of biological and radiopharmaceutical drugs indicated that the LAL test was a rapid, sensitive, and reproducible way to detect pyrogen in these products (51). In 1973, the *Federal Register* announced the intention of the FDA to license LAL reagents as a biological product. That announcement also proclaimed the usefulness of the reagent for detecting endotoxin (52).

A *Federal Register* announcement in 1977 gave conditional approval for the BET as a release test for medical devices and biological products, provided that product manufacturers submitted appropriate test-validation data to each respective agency to amend their product's registration documents or license (53). The Bureau of Medical Devices established its own endotoxin standard and supported an industry collaborative study to develop an endotoxin limit and uniform test method for parenteral devices. As acceptance of LAL testing grew, the FDA decided to prepare a single guideline that would apply to all Agency-regulated products that were subject to screening for endotoxin. A FDA Task Force was formed with representatives from the Agency's various Centers to standardize test-validation criteria. A guideline for end-product testing by LAL methods was published in 1987 (54).

The FDA LAL test-validation guideline describes how to use all LAL methods and identifies three basic requirements:

1. The LAL reagent used in all validation, in-process, and end-product tests must be licensed by CBER.
2. The product manufacturer must perform an initial qualification of their laboratory personnel and facility.
3. Inhibition and enhancement tests must be conducted on test products to ensure that the products do not interfere with the detection of endotoxin.

This guideline has been the most influential document in LAL testing to date. Annual updates of the Appendix E, endotoxin limits for established parenteral products, were published by the FDA until 1994. With the upgrades in test procedures and endotoxin limits published in individual product monographs, the FDA no longer has a need to revise the 1987 guideline.

Overcoming Inhibition and Enhancement Conditions

Conditions for the LAL reaction with endotoxin require pH neutrality and optimum levels of sodium and divalent cations. A uniform temperature of 37°C optimizes the rate of reaction. Most therapeutic drug products require dilution with LAL reagent water (LRW) before testing to avoid interference, which is recognized by improper recovery of the positive product control (PPC). Inhibition is a failure to recover the PPC, whereas enhancement is high recovery of the PPC. Cooper (55) has described ways to identify the concentration of a substance, which is chemically compatible with the BET, and to validate the BET with drug products and excipients.

Inhibition mechanisms

There are three principal causes of inhibitory effects in gel-clot and kinetic LAL testing (55). One is an adverse chemical condition, such as a nonneutral pH, suboptimum levels of sodium ions and divalent cations (magnesium and calcium), or inhibition by trivalent cations such as soluble forms of aluminum and gadolinium. The LAL manufacturer can alleviate many of these problems by incorporating optimum concentrations of all necessary components into the LAL formula, in particular, significant amount of an organic buffer. The second important type of inhibition is loss of purified endotoxin used for the PPC. This invalidity is somewhat of an artifact because purified endotoxin, LPS, is more unstable in water than in native endotoxin, which has a stabilizing protein attached. This instability leads to container adsorption and molecular aggregation that makes the positive control unavailable for LAL detection, and this loss is not always overcome by vortex mixing. Optimum formulation of the CSE by the supplier is critical to the recovery of positive controls. Finally, interference results from inadequately controlled test parameters including test accessories, reagents, and analyst's proficiency.

Enhancement

The LAL reaction is specific for endotoxin with the exception of a glucose polymer from cellulose, yeast, and certain other microbial sources, known as (1→3)-β-D-glucan (46). This glucan activates LAL the same as endotoxin and produces a synergistic enhancement of the PPC in kinetic LAL studies. The BET allows glucan-containing products to be tested by LAL reagents that are treated to make them specific for endotoxin and avoid a false-positive result.

Preparation of a positive product control (PPC)

The PPC, required by the FDA Guideline, is a sample of the test material that contains a concentration of endotoxin that is double the labeled sensitivity of the LAL reagent (54). The PPC must be tested with each sample in a Limits test or kinetic LAL assay to ensure that a result is valid and free of interference. The most accurate and reliable technique is the “hot spike” method that requires adding 10 µl of endotoxin standard to the reaction vessel (tube or well) before addition of LAL. The gel clot method requires the addition of 10 µl of 20λ to the reaction mixture. Kinetic methods require the addition of 10 µl of a standard, 10 times greater than the spike concentration, to the wells or tubes designated as PPCs. Only inhibition is seen in gel-clot Limits tests, whereas both inhibition and enhancement are seen in kinetic LAL methods.

Sensitivity of a BET

The results of endotoxin tests for in-process solutions, bulk materials, and finished parenteral products should be reported in the same units as those assigned to the product. Two factors determine the sensitivity of a BET. For infusion solutions and device extracts, the gel-clot sensitivity or the lowest point on the standard curve (lambda for kinetic LAL) and the amount of dilution determine test sensitivity (55). For products that have an endotoxin limit in EU/mg, the choice of lambda and the concentration of the test material determine sensitivity. The formula for product-specific sensitivity (PSS) is a convenient way to calculate the sensitivity of a BET for this type of product, where:

$$\text{PSS} = \frac{\text{Lambda (EU/mL)}}{\text{Test concentration (mg/mL)}}$$

DEPYROGENATION

A process that removes or destroys endotoxin in a solution or on a material is “termed”? depyrogenation. Endotoxin is difficult to eliminate because it is ubiquitous in nature, stable, and pervious to sterilizing filters. Chemical destruction requires treatment with strong base or oxidants, which is usually too corrosive for practical use. Aseptic processing of parenteral products requires that all components, excipients, and active ingredients be made endotoxin-free (depyrogenated) before filling or assembly (55, 56). The BET is the test of choice to monitor the effectiveness of depyrogenation.

Water

Municipal water systems usually contain 5-to-50 EU/mL of endotoxin. Pharmaceutical waters are treated by ultrafiltration, distillation, or reverse osmosis to separate endotoxin from water (55). The principal source of endotoxin is bacteria within a water system in the form of biofilm or colonies entrapped in resin beds, etc. A water system will be contaminated unless there is an ongoing sanitization program. The endotoxin limit for Water for Injection was set at 0.25 EU/ml because it is a critical vehicle and a major source of pyrogens.

Glassware

Dry heat sterilization is used to depyrogenate glassware and other heat-stable materials (55, 57). Temperatures in the range of 250°–325°C rapidly inactivate endotoxin by thermal incineration. The size and mass of a load influence the time required to reach equilibrium within the oven; therefore, empirical data are required to verify that desired conditions were achieved. Overkill cycles in the range of 4–6 log reduction values (LRV) are common. A heat exposure of 250°C for at least 30 min provides greater than a 3-log reduction.

Oven cycles are most accurately challenged with an endotoxin indicator that contains 1000–10,000 EU per container in a 5- to 10-ml vial. This configuration enables complete recovery of the endotoxin challenge because it is small enough to vortex mix. Indicators are placed inside a glass pack of vials or large vessels, exposed to the oven cycle, and analyzed for log reduction.

Elastomeric Closures

Closures are traditionally sterilized and depyrogenated by a combination of washing, rinsing, and steam sterilization. All three of these steps are effective in reducing endotoxin and should be considered in the validation process. Steam sterilization as a depyrogenation process is underappreciated (58). The combined steps consistently give >3-log reduction values. The recovery of endotoxin challenge levels from closures is less efficient than recovery from glass because of the porous nature of the closure and inability to apply vortex mixing efficiently. An acceptable recovery for challenge stoppers is >10% (55).

Bulk Products

Outbreaks of pyrogenic reactions underscore the need to assign endotoxin alert limits (EAL) and to screen blood and fermentation products for endotoxin. Two clusters of

pyrogenic reactions were traced to a bulk producer of gentamicin (59). Patients reacted to endotoxin levels similar to the threshold pyrogenic dose of 4.1 EU/kg determined in a study of reference endotoxin in a population of healthy male volunteers (60).

An EAL for a bulk product is a fraction of the endotoxin limit (EL) for the finished product. An appropriate range for an EAL is at least four to five times less than the EL, depending on solubility and interference properties of the material. A validated method for a bulk substance should include a specific method to dissolve the drug and dilute it to a compatible concentration (55).

Biotherapeutic products

The elimination of endotoxin from a recombinant or fermentation product is challenging because the host organism often contributes enormous amounts of endotoxin to the bulk material. The product is usually separated from endotoxin and feedstream impurities by affinity columns.

Future of the BET

The trend among pharmaceutical and medical device manufacturers is to move from gel-clot to kinetic LAL methods to quantify results, enhance test efficiency, and improve management of BET data. Powerful endotoxin-specific software for kinetic methods addresses the needs for simplicity, comprehensive reporting, trend analysis, and compliance with software validation requirements. In the past 25 years, the BET has enabled the parenteral industry to greatly reduce the levels of endotoxin pyrogen in its products. Competition between bait fishermen and the LAL industry for the horseshoe crab has brought about coast-wide conservation measures to maintain *Limulus* resources. The BET will continue to be an essential tool for assuring the safety of parenteral products.

REFERENCES

- Annen, G.I.I. An Exhibition of Books on the Growth of Our Knowledge of Blood Transfusion. *Bull. N.Y. Acad. Med.* **1939**, *15*, 622–632.
- Dudrick, S.J.; Rhoades, J.E. *JAMA* **1971**, *215*, 923–949.
- Macht, S.D. Three Hundred Years of Parenteral Nutrition: The History of Intravenous Nutritional Therapy. *Conn. Med.* **1980**, *44* (1), 27–30.
- Westphal, O.; Westphal, U.; Sommer, T. The History of Pyrogen Research. *Microbiology*. **1977**, 221–238.
- Seibert, F.B. Pyrogens from an Historical Viewpoint. *Transfusion*. **1963**, *3*, 245–249.
- Centanni, E. Untersuchungen über Das Infektionsfieber, Das Fiebergift Der Bakterien. *Deutsch. Med. Wochenschr.* **1894**, *10*, 148.
- Hort, E.; Penfold, W.J. Microorganisms and Their Relation to Fever. *J. Hyg. Camb.* **1912**, *12*, 361–390.
- Hort, E.; Penfold, W.J. A Critical Study of Experimental Fever. *Proc. Roy. Soc. Med., London, Ser. B.* **1912**, *85*, 174–186.
- Hort, E.; Penfold, W.J. The Reaction of Salvarsan Fever to Other Forms of Injection Fever. *Proc. Roy. Soc. Med., Part III, Pathology.* **1912**, *5*, 131–139.
- Seibert, F.B. Fever-Producing Substances Found in Some Distilled Waters. *Am. J. Physiol.* **1923**, *67*, 90–104.
- Seibert, F.B. The Cause of Many Febrile Reactions Following Intravenous Injections. *Am. J. Physiol.* **1925**, *71*, 621–651.
- Seibert, F.B.; Mendel, L.B. Temperature Variations in Rabbits. *Am. J. Physiol.* **1923**, *67*, 83–89.
- Pearson, F.C., III Pyrogens Other than Endotoxin. *Pyrogens*; Marcel Dekker, Inc.: New York, 1985; 64–76.
- Dinareello, C.A. Production of Endogenous Pyrogen. *FASEB.* **1979**, *38*, 52–56.
- Elin, R.J.; Wolff, S.M.; McAdam, K.P.W.J.; Chedid, L.; Audibert, F.; Bernard, C.; Oberling, F. Properties of Reference *Escherichia Coli* Endotoxin and Its Phthalylated Derivative in Humans. *J. Infect. Dis.* **1981**, *144*, 329–336.
- Marcus, S.; Anselmo, C.; Luke, J. Studies on Bacterial Pyrogenicity, II. A Bacteriological Test for Pyrogens in Parenteral Solutions. *J. Am. Pharm. Assoc.* **1960**, *9*, 616–619.
- Atkins, A.; Freedman, L.R. Studies in Staphylococcal Fever, I. Responses to Bacterial Cells. *Yale J. Bio. Med.* **1963**, *35*, 451–471.
- Braude, A.I.; McConnell, J.; Douglas, H. Fever from Pathogenic Fungi. *J. Clin. Invest.* **1960**, *39*, 1266–1276.
- Atkins, E. Pathogenesis of Fever. *Physiol. Rev.* **1960**, *40*, 580–605.
- Dinareello, C.A. Interleukin-1 and Its Biologically Related Cytokines. *Adv. Immunol.* **1989**, *44*, 155–205.
- Bendtsen, K. Interleukin-1, Interleukin-6 and Tumor Necrosis Factor in Infection, Inflammation and Immunity. *Immunol. Lett.* **1988**, *19*, 183–192.
- Pearson, F.C., III Endotoxin. *Pyrogens*; Marcel Dekker, Inc.: New York, 1985; 23–56.
- Rietschel, E.T.; Brade, H. Bacterial Endotoxins. *Sci. Am.* **1992** (August), 55–61.
- Rietschel, E.T. Foreword. *Immunobiol.* **1993**, *187*, 167–168.
- Hellon, R.; Townsend, Y. Mechanisms of Fever. *Pharm. Ther.* **1983**, *19*, 211–244.
- Martich, G.D.; Boujoukos, A.J.; Suffredini, A.F. Response of Man to Endotoxin. *Immunobiology.* **1993**, *187*, 403–416.
- Taveira da Silva, A.M.; Kaulbach, H.C.; Chuidian, F.S.; Lambert, D.R.; Suffredini, A.F.; Danner, R.L. Brief Report: Shock and Multiple-Organ Dysfunction After Self-Administration Of *Salmonella* Endotoxin. *New Eng. J. Med.* **1993**, *328*, 1457–1460.
- Naunts, H.C. Bacterial Pyrogens: Beneficial Effects on Cancer Patients. *Biomedical Thermology*; Alan R. Liss, Inc.: New York, 1982; 687–696.
- Legrand, E.K. An Evolutionary Perspective of Endotoxin: A Signal for a Well-Adapted Defense System. *Med. Hypoth.* **1990**, *33*, 49–56.

29. Nelson, L. Application of Reverse Osmosis in Pyrogen Removal. *Depyrogenation, Technical Report No. 7*; Parenteral Drug Association: Philadelphia, 1985; 28–36.
30. Sweadner, K.J.; Forte, M.; Nelson, L. Filtration Removal of Endotoxin (Pyrogens) in Solution in Different States of Aggregation. *Appl. Environ. Microbiol.* **1977**, *34*, 382–385.
31. Galanos, C.; Freudenberg, M.A.; Luderitz, O.; Rietschel, E.T.; Westphal, O. Chemical, Physicochemical and Biological Properties of Bacterial Lipopolysaccharides. *Biomedical Applications of the Horseshoe Crab (Limulidae)*; Cohen, E., Ed.; Alan R. Liss: New York, 1979; 32–332.
32. Weary, M. The Rabbit Pyrogen Test. *Pyrogens*; Marcel Dekker, Inc.: New York, 1985; 104–118.
33. Welch, H.; Calvery, H.D.; McClosky, W.T.; Price, C.W. Method of Preparation and Test for Bacterial Pyrogens. *J. Am. Pharm. Assoc.* **1943**, *32*, 65–69.
34. McClosky, W.T.; Price, C.W.; Van Winkle, W.J.; Welch, H.; Calvery, H.O. Results of First USP Collaborative Study of Pyrogens. *J. Am. Pharm. Assoc.* **1943**, *32*, 69–73.
35. Greisman, S.E.; Hornick, R.B. Comparative Pyrogenic Reactivity of Rabbit and Man to Bacterial Endotoxin. *Proc. Soc. Exp. Biol. Med.* **1969**, *131*, 1154–1158.
36. <151> Pyrogen Test. *U.S. Pharmacopeia 24*; United States Pharmacopeial Convention, Inc.: Rockville, MD, 2000; 1850.
37. Taktak, Y.S.; Selkirk, S.; Bristow, A.F.; Carpenter, A.; Ball, C.; Rafferty, B.; Poole, S. Assay of Pyrogens by Interleukin-6 Release from Monocytic Cell Lines. *Pharm. Pharmacol.* **1991**, *43*, 578–582.
38. Dinarello, C.A.; Gatti, S.; Barfai, T. Fever: Links with an Ancient Receptor. *Curr. Biol.* **1999**, *9*, R147–150.
39. Pearson, F.C.; Weary, M. The Significance of Limulus Amebocyte Lysate Test Specificity on the Pyrogen Evaluation of Parenteral Drugs. *J. Parenter. Drug Assoc.* **1980**, *34*, 103–108.
40. Bang, F.B. A Bacterial Disease Of *Limulus Polyphemus*. *Bull. Johns Hopkins Hosp.* **1956**, *98* (3), 325–351.
41. Levin, J.; Bang, F.B. The Role of Endotoxin in the Extracellular Coagulation Of *Limulus* Blood. *Bull. Johns Hopkins Hosp.* **1964**, *115* (3), 265–274.
42. Levin, J.; Bang, F.B. Clottable Protein In *Limulus*: Its Localization and Kinetics of Its Coagulation by Endotoxin. *Thromb. Diathes. Haemorrh.* **1968**, *19*, 1186–1197.
43. Cooper, J.F.; Levin, J.; Wagner, H.N., Jr. Quantitative Comparison of In Vitro (*Limulus*) and In Vivo (Rabbit) Methods for the Detection of Endotoxin. *J. Lab. Clin. Med.* **1971**, *78*, 138–148.
44. Flint, O. A Timetable for Replacing, Reducing and Refining Animal Use with the Help of In Vitro Tests: The Limulus Amebocyte Lysate Test (LAL) as an Example. *Alternatives to Animal Testing. New Ways in Biomedical Sciences*; Reinhardt, C.A., Ed.; Weinheim: Germany, 1994; 27–43.
45. Nakamura, T.; Morita, T.; Iwanaga, S. Lipopolysaccharide-Sensitive Serine-Protease Zymogen (Factor C) Found in *Limulus* Hemocytes. *Eur. J. Biochem.* **1986**, *153*, 511–521.
46. Cooper, J.F.; Weary, M.E.; Jordan, F.T. The Impact of Non-Endotoxin LAL-Reactive Materials on Limulus Amebocyte Lysate Analyses. *PDA J. of Pharm. Sci. Technol.* **1997**, *51*, 2–6.
47. Harada, T.; Morita, T.; Iwanaga, S.; Nakamura, S.; Niwa, M. A New Chromogenic Substrate Method for the Assay of Bacterial Endotoxins Using *Limulus* Hemocyte Lysate. *Biomedical Applications of the Horseshoe Crab (Limulidae)*; Cohen, E., Ed.; Alan R. Liss, Inc.: New York, 1979; 209–220.
48. Weary, M.E. Understanding and Setting Endotoxin Limits. *J. Parenter. Sci. Technol.* **1990**, *44*, 16–18.
49. Cooper, J.F.; Harbert, J.C. Endotoxin as a Cause of Aseptic Meningitis After Radionuclide Cisternography. *J. Nucl. Med.* **1975**, *16*, 809–813.
50. <85>Bacterial Endotoxins Test. *USP 24-NF 19*; The United States Pharmacopeial Convention, Inc.: Rockville, MD, 2000; 2875–2879, Supplement 2.
51. Cooper, J.F.; Hochstein, D.H.; Seligmann, E.B. The Limulus Test for Endotoxin (Pyrogen) in Radiopharmaceuticals and Biologicals. *Bull. Parent. Drug Assoc.* **1972**, *26*, 153–162.
52. U.S. Public Health Service. Limulus Amebocyte Lysate: Additional Standards. *Fed. Reg.* **1973**, *38*, 26130.
53. Dabbah, R.; Ferrry, E.; Gunther, D.A.; Hahn, R.; Mazur, P.; Neely, M.; Nicholas, P.; Pierce, J.; Slade, J.; Watson, S.W.; Weary, M. Pyrogenicity of E. Coli 055:B5 by the USP Rabbit Pyrogen Test—HIMA Collaborative Study. *J. Parenter. Drug Assoc.* **1980**, *34*, 212–216.
54. US Food and Drug Administration. *A Guideline on Validation of the Limulus Amebocyte Lysate Test as an End-Product Endotoxin Test for Human and Animal Parenteral Drugs, Biological Products and Medical Devices*; Food and Drug Administration: Rockville, MD, 1987.
55. Cooper, J.F. Bacterial Endotoxins Test. *Microbiology in Pharmaceutical Manufacturing*; Prince, R., Ed.; Parenteral Drug Assoc., Bethesda, Davis Horwood International Publication, Ltd.: Modalmeny, UK, 2001; p. 537–567.
56. Weary, M. Depyrogenation. *Pyrogens*; Marcel Dekker, Inc.: New York, 1985; 203–218.
57. Tsuji, K.; Harrison, S. Dry Heat Destruction of Lypopolysaccharide: Dry Heat Destruction Kinetics. *Appl. Environ. Microbiol.* **1978**, *36*, 710–714.
58. Bamba, R.; Matsui, R.; Watabe, I. Effect of Steam-Heat Treatment with/without Divalent Cations of the Inactivation of Lipopolysaccharides from Several Bacterial Species. *PDA J. Parenter. Sci. Technol.* **1996**, *50*, 129–135.
59. Fanning, M.M.; Wassel, R.; Piazza-Hepp, T. Pyrogenic Reactions Associated with Gentamicin Therapy. *N. Eng. J. Med.* (2000) *343*, 1658–1659.
60. Hochstein, H.D.; Fitzgerald, E.A.; McMahon, F.G.; Vargas, R. Properties of US Standard Endotoxin (EC-5) in Human Male Volunteers. *J. Endotoxin Research.* **1994**, *1*, 52–56.

Quality Assurance of Pharmaceuticals

Barbara K. Immel

Immel Resources LLC, Petaluma, California, U.S.A.

Whatever you can do or dream you can, begin it. Boldness has genius, power, and magic in it.

— Goethe

INTRODUCTION

The essence of pharmaceutical quality assurance (QA) is to protect the patient and to ensure his or her safety in taking a company's drug product. One definition of a quality product is that it meets its specifications, that it is fit for its intended use, and that it satisfies the customer. Having an effective QAU is required by both United States and worldwide regulations governing the industry. Quality assurance must be independent of other departments and fully able to discharge its responsibilities and exercise its authority within a company without interference or conflict of interest. Quality assurance achieves its mission by putting necessary systems, procedures, and controls in place and continually improving them.

RESPONSIBILITIES

The major responsibilities of QA include auditing and evaluating operations, the approval of product and document quality (approval here meaning either "approve" or "reject"), and managing change control, problem investigations, and product recalls, among other tasks. Specific tasks vary depending upon whether the QA group is working with research and development on good laboratory practice (GLP) studies, in a manufacturing environment ensuring compliance with good manufacturing practice (GMP) regulations while producing clinical or commercial product, or working with medical monitors, principal investigators, institutional review boards, or contract research organizations (CROs) in complying with good clinical practice (GCP) regulations.

Quality assurance is primarily a preventive function, the goal being to put good systems, standard operating procedures (SOPs), and other needed documents in place

to consistently produce quality product, prevent problems or mistakes from occurring, and release only product, raw materials, studies, and reports that meet specifications. Products that meet specifications may still be rejected due to unsatisfactory conditions during manufacture (such as environmental monitoring excursions or other events that may affect the product's safety and purity).

Quality assurance can also be a reactive function, responding to a product quality problem or regulatory action. This group must have a written recall SOP and system (which closely follows 21 CFR 7) such that it can quickly consult with a cross-section of employees, including regulatory, research and development, clinical trial, manufacturing, quality control (QC), public or investor affairs, and legal counsel, in determining the known facts and the best course of action (and corrective action) to take, including rapidly initiating a product recall if necessary. Before discussing specific QA roles and responsibilities under the regulations, let us look at the overall process or life cycle of pharmaceutical product development.

PRODUCT DEVELOPMENT IN A NUTSHELL

In a recent PhRMA study, the average cost and length of time to develop a single pharmaceutical product from idea to marketing was estimated at \$500 million and 15 yr.^[1] A few investment-banking firms have estimated the cost to be even higher. The pharmaceutical industry has several unique features. One is that the industry exists to create products to treat disease, reduce pain and suffering, improve the quality of life, and find cures for disease. The second is that the industry is highly regulated. Obviously, to remain in business, a company must be profitable or have sustaining revenue of some sort. James Tingstad, the well-respected, now retired, columnist for *Pharmaceutical Technology*, says in his recent book, *Good Technical Management Practices*, that in our industry, the patient should always be number one because he or she is ill, followed by employees, and then by stockholders. Obviously, this involves a very difficult balancing act—balancing the needs of the patients with the needs of employees, shareholders, and investors.

Beginning with the birth of an idea, a potential new product is evaluated in research and development. Certain studies (animal toxicology or safety studies) must be performed following GLPs. If the safety data look good and the decision is made to further develop the product, an investigational new drug application (IND) is filed in order to initiate human clinical trials within 30 days of filing.

During clinical trials, all material produced to be given to patients must be manufactured according to GMPs. The stringency with which the GMPs are applied can be viewed as a continuum, with less stringent GMP requirements in clinical trials and tighter requirements in later phases. The sterility of clinical material, as well as the need to prevent any risk of contamination or cross-contamination, must be satisfactorily addressed and assured before venturing into human clinical trials.

Clinicians or others involved in conducting the trials must follow another set of regulations known as GCPs. Typically, there are three to four major phases of clinical trials for a drug or biologic product: Phase I, II, and III (Phase IV being postmarketing studies). If patient's data prove promising, and if the product has been demonstrated to be both safe and effective, a new drug application (NDA) is filed.

Before a new product gains approval to be marketed in the United States, two major steps must be taken. One is the technical or scientific review, in which reviewers from the U.S. Food and Drug Administration (FDA), and sometimes an outside advisory committee of experts, review the scientific and clinical data to determine the product's safety and efficacy. As part of that review, an FDA investigator is sent to the company's manufacturing and research facilities to perform a pre-approval inspection (PAI) or prelicensing (PLI) inspection, to see if the data are trustworthy, if applicable regulations were followed in developing and manufacturing clinical product, and if the manufacturing facility appears capable of following GMPs during commercial manufacture. The site that will be manufacturing commercial material must pass the PAI or PLI for approval to be given.

If the scientific or medical review as well as the inspection results are acceptable, the second step is approval to market. Adherence to GMPs is then required throughout the lifetime of the product. Frequently, the company will opt to begin another clinical trial to study the use of the compound for another indication, or to complete additional testing required by the FDA. Those trials must be performed following GCPs (Fig. 1).

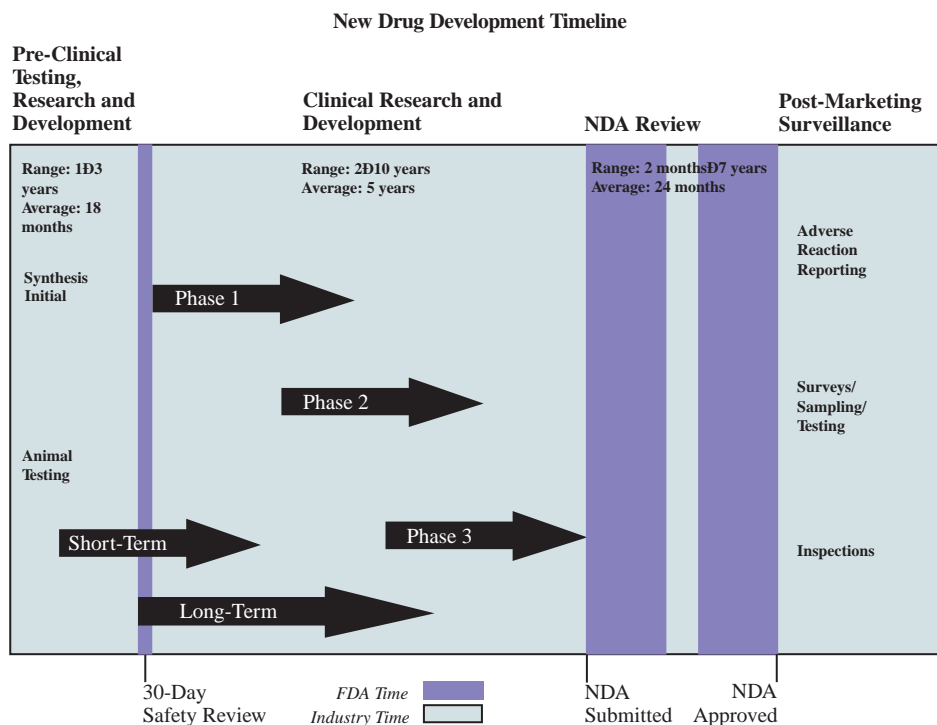


Fig. 1 New drug development timeline. (From *FDA Consumer Special Report on New Drug Development in the United States*, U.S. Food and Drug Administration, Rockville, MD, January 1995.)

GLPs

In the United States, GLP regulations (21 CFR 58) apply to safety or animal toxicology studies performed during research and development. These regulations define GLPs for conducting nonclinical laboratory studies that support or are intended to support applications for research or marketing permits for products regulated by the FDA, including food and color additives, animal food additives, human and animal drugs, medical devices for human use, biological products, and electronic products.

Compliance with this part is intended to assure the quality and integrity of the safety data filed.^[2]

Good laboratory practices require that management establish a QAU to monitor each study to assure conformance with the regulations. The GLPs also state that the QAU must be entirely separate from and independent of personnel engaged in the direction and conduct of the study.^[2]

Regulatory requirements of the QAU include:

- Maintaining a copy of the master schedule sheet of all nonclinical laboratory studies, indexed by test article and containing the test system, nature of study, date study was initiated, current status of each study, identity of sponsor, and name of study director.
- Maintaining copies of all protocols pertaining to all nonclinical laboratory studies for which the unit is responsible.
- Inspecting each nonclinical laboratory study at adequate intervals to ensure the integrity of the study.
- Maintaining written and properly signed records of each periodic inspection showing date of inspection, study inspected, phase or segment of study inspected, person performing the inspection, findings and problems, recommended actions and actions taken, and any scheduled date for reinspection.
- Immediately bringing to the attention of the study director and management any problems found that are likely to affect the study's integrity.
- Periodically submitting to management and the study director written status reports on each study, noting any problems and corrective actions taken.
- Determining that no deviations from approved protocols or SOPs were made without proper authorization and documentation.
- Reviewing the final study report to assure that the report accurately describes both the methods and SOPs and that the reported results accurately reflect the raw data.
- Preparing and signing a statement to be included with the final study report that specifies the date's inspections were made and the date's findings were reported to management and the study director.^[2]

In addition, per U.S. GLPs, the responsibilities and procedures of the QAU, all records it maintains, and the method of indexing such records must be in writing and continually maintained. Required records include internal audit (inspection) dates, the studies inspected, the phase of each study inspected, and the name of the individual performing the audit. That information as well as copies of all written procedures must be made available upon request to FDA investigators. Upon request, facility-testing management must also supply a written certification to the FDA that the internal audits have been implemented, performed, documented, and followed up.^[2]

Outside the United States, the Organization for Economic Cooperation and Development (OECD) publishes GLPs that have been adopted by its members and the European Union. These GLPs apply to the testing of chemicals to obtain information about their properties and/or their safety with respect to human health and the environment. The Japanese Ministry of Health and Welfare (MHW) has also published GLPs that apply to nonclinical laboratory safety studies that support applications to manufacture or import drugs or applications to re-examine new drugs in accordance with the country's Pharmaceutical Affairs Law.^[3]

The good news is that harmonization efforts are under way, and basic GLP principles are similar. Nevertheless, differences do exist. Major differences include the requirements for the specific contents of protocols and final reports and the retention times for records, samples, and specimens.^[3] Manufacturers must become familiar with the GLPs that concern their products, learn to which worldwide regulatory agencies their data will be submitted, and know which legislation covers the material being produced and evaluated.^[4]

Pitfalls to Avoid

Recurring, problematic areas in GLPs seem to be in documentation, training, working with and monitoring contract laboratories, and obtaining necessary management support. A noted FDA investigator and national GLP expert recently stated that the most common problems one sees may result from a general lack of support for the QAU, including:

- Assigning responsibility for the QAU to an inexperienced employee (e.g., someone who holds a business degree but lacks laboratory experience).
- Becoming complacent about training (e.g., assigning responsibility to employees who can perform the procedures but who do not know how to train).

- Misusing standardized copy to put together protocols and reports and not proofreading the documents.
- Failing to validate computerized operations.
- Management's belief that required reviews of final reports and protocols can be shortened when computers are used.
- Reduced, or nonexistent, funding support for QA.

The investigator asks which is less expensive: funding an additional one or two QA employees or not being able to use a study or include it in a regulatory submission because it was not conducted following GLPs.

Other GLP problems reported by an experienced FDA investigator include:

- GLP studies and "basic research" being done in the same laboratory without a clear division of research.
- Lack of coordination between contract facilities and sponsor of study.
- Outdated SOPs that do not reflect current practice.
- employees not following or not knowing SOPs, "maverick" SOPs.
- Inadequate review of studies, protocols, and SOPs by QA.
- Lack of documentation of installation and/or validation of equipment, computers, and software.
- Expired reagents and test articles.
- Failure to separate species (e.g., rats and pigs housed together).
- Improper archiving of records and specimens—lack of labeling or identification.
- No documented GLP training (especially for support staff).^[5]

It is critical to train and offer refresher training in not only good laboratory notebook practices but also in all job skills, test methods, and instrument systems used, and to review data rapidly. It is important that a well-respected individual provide the training. Good notebook practices are also essential to successfully file for and defend patents. A summary of laboratory notebook tips is outlined in the following section.

Twenty-Five Laboratory Notebook Tips

1. Use permanent, indelible ink when making entries. Write clearly and legibly. Remember that your audience includes not only you or your supervisor, but also the company's regulatory and QA employees, patent attorneys, and FDA and other regulatory agency investigators who may be reviewing your data years after you record it.

- Describe and provide quantities for all materials used.
2. For developmental or patentable work, state the object, purpose, and results of each experiment clearly and concisely.
3. Record all operating details and conditions, including yields, product or compound names, lot numbers of standards and reference materials, suppliers, and any expiration dates.
4. Use good scientific methods. Narrow scientific variables to one; strive for reproducibility, accuracy, and precision.
5. Develop and use approved methods. Use compendial methods whenever available. Validate non-compendial methods. Make every effort to put these suggestions in place so that you can rely upon and consistently reproduce your results.
6. Reference any methods, specifications, chromatographic data, and chart books. Include document numbers, effective dates, and revision numbers for any controlled documentation.
7. Each page must be signed and dated by the individual who makes the entry and who does the work.
8. Attach graphs, charts, etc. to notebook pages with permanent adhesive. When unfolded, attached documents should be within the confines of the opened notebook.
9. Development data should be checked, signed, and dated, or witnessed (signed and dated, with the notation, "read and understood") as soon as possible.
10. Initials are acceptable in place of a signature if the initials are on file and if this practice is written and approved in the company notebook SOPs.
11. Avoid negative or extraneous comments. Do not state opinion. Be factual.
12. For products under development, a witness must have the technical ability to witness the work but must neither have conceived the idea nor have taken part in any way in the work performed by another employee.
13. Each day's work should be started on a separate page with lines drawn diagonally across the unused portion of the previous page.
14. The supervisor or checker should sign under any diagonal lines across unused portions of a page.
15. If work is continued on a second page, start the page with the phrase "Continued from page" Each page must show the date of entry.

16. No entries should be made beneath attached sheets, and nothing should be obscured.
17. Inserts should be signed and dated by the person making the entry and checked and dated, or witnessed, by another.
18. If data are not kept in the notebook, they must be checked, signed, dated, and identified to provide a reference to the pertinent page of the laboratory notebook.
19. Lab notebooks must not be copied without the area supervisor's or QA manager's knowledge and approval.
20. Every notebook must be assigned a unique number when checked out, and reconciled when it is returned to the company archives.
21. Lab notebooks are confidential and are the property of the company. Multiple, completed notebooks should not be kept in the researcher's or analyst's work area. Filled notebooks should be returned to company archives for proper storage. Intellectual property is the lifeblood of an organization. Every company should have a system set up for the proper archival and retrieval of laboratory notebooks. A central location should be chosen, and it should be both fire- and waterproof. Proper filing should be enforced, including requiring that departing employees return their laboratory notebooks prior to their leaving the company.
22. Every company should establish a laboratory notebook policy and a record retention policy.
23. All employees must have documented training in notebook procedures and in any revisions to those procedures.
24. Take the time to organize your information with a table of contents—a table listing all experiments and the pages in which they are documented—so that the information can be easily retrieved when the company puts together a submission or prepares for an FDA inspection.
25. To reference the notebook, include the analyst's or researcher's initials, notebook number, and notebook page.^[4]

GMPs

The U.S. GMPs for drugs (21 CFR 210/211), made final in 1978, were among the first GMPs to be published by any nation. The GMPs were intended to help ensure the safety and efficacy of all products.

The regulations contain the minimum CGMP for methods to be used in, and the facilities or controls to be

used for, the manufacture, processing, packing, or holding of a drug to assure that such drug meets the requirements of the act as to safety, and has the identity, strength, quality, and purity characteristics that it purports or is represented to possess.^[10]

Responsibilities given the QA function, called the “QC unit” in GMP regulations, are as follows. The regulations state that there must be a QC unit with the following responsibilities. Besides applying to the manufacture of drugs or finished pharmaceuticals, 21 CFR Part 210 and 211 serve as a guide for the manufacture of active pharmaceutical ingredients (APIs) or biologic APIs.

- Approving or rejecting all components, drug product containers, closures, in-process materials, packaging material, labeling, and drug products.
- Reviewing production records to assure that no errors have occurred or, if errors have occurred, that they have been fully investigated.
- Approving or rejecting drug products manufactured, processed, packed, or held under contract by another company.
- Having or providing adequate laboratory facilities for the testing and approval (or rejection) of components, drug product containers, closures, packaging materials, in-process materials, and drug products.
- Approving or rejecting all procedures or specifications impacting the identity, strength, quality, and purity of the drug product.

As with the GLPs, responsibilities and procedures applicable to the QC unit must be in writing and must be followed.^[10] The proposed revision to the GMPs for finished pharmaceuticals (not yet final at the time this article was published) codify good practices already in use at many companies, including reviewing and approving validation protocols and reviewing changes in product, process, equipment, or other changes to determine if and when revalidation is warranted.^[11]

Major responsibilities include the auditing and evaluation of operations. Quality assurance must independently assess and inform senior management of the quality and compliance status of their operations, suppliers, and contractors relative to their SOPs, material specifications, company standards, and regulatory requirements. Quality assurance usually manages or supervises internal, corporate, and regulatory agency audits and accompanies auditors, unless the company has a separate compliance group to perform this function (whichever department is responsible, the responsibilities should be defined in an SOP and followed).

Quality assurance is also responsible for approving or rejecting product quality or “dispositioning” a lot. Making disposition decisions is one of the toughest jobs a QA professional will ever perform (see “Making QA Decisions”). Dispositions (whether approval or rejection) should only be made after a thorough review of all applicable records. The disposition decision should always be made by an experienced QA professional or executive, never by a junior employee or consultant, even though nothing in the U.S. regulations prohibits the latter. (In the European Union, only a “qualified person,” with the requisite experience and so designated on the company’s manufacturing or marketing authorizations, may release a lot. This requirement applies to any lots to be released or sold in the European Union.)

In discharging their review and approval responsibilities, QA must review the adequacy of procedures, methods, sample plans, test methods, and internal and external specifications and approve SOPs, specifications, methods, validation plans and results, and related documents. The QC unit must also approve raw materials, components, facilities (including both internal facilities and those of suppliers and contractors) and in-process and finished goods.

Where there is a program allowing Production or another department to perform required inspections and tests, the program and results of its activities shall be subject to QA periodic audits or monitoring. This monitoring may include QA inspections and tests performed in parallel or in place of Production.

Quality assurance must approve each batch to be marketed by the unit, manufactured by the unit for use in clinical trials, submitted to regulatory agencies in support of registrations, or used to support validation of a process.

Purely experimental batches do not require QA approval. All approval must be based on a thorough QA review of the adequacy of the facility, its procedures, methods, records, inspections, and test results, prior to batch release. This applies to both internally manufactured batches and those made by contractors.

Finally, QA is also responsible for managing change control, problem investigations, and product recalls. QA must review and approve change, control SOPs, and ensure that QA activities are completed by approving proposed changes based on the impact of the change on quality and compliance with the regulations.

Quality assurance must approve written problem investigation SOPs, evaluate investigations, approve actions, and ensure that corrective and preventive follow-up is done. QA must identify and lead to conclusion any situation that requires a recall of a distributed product or similar action. Recalls are one of the toughest situations a

QA professional will face. Items to consider in a recall meeting are given later in this article.

Common sense dictates that the head of QA may not be at a lower organizational reporting level than the head of Production. To minimize conflicts of interest, the head of QA should not have routine accountability for processing, packaging, purchasing, production planning, engineering, or inventory control. Groups that develop data used for QA decisions, such as laboratories and inspection departments, usually report to QA but may report to another non-Production area such as R&D. The QC laboratory, which routinely performs tests of raw materials, in-process materials, and finished product, is not a manufacturing support operation.^[18]

Major Tasks

Common tasks of a QA group involved in manufacturing clinical or commercial product include the following:

- Controlled documentation and record-keeping.
- Batch record (the “recipe” for manufacturing each lot) issuance and review.
- Batch or lot disposition (approve or reject).
- Internal audits.
- Review of (may help write) regulatory submissions).
- Change control.
- Problem investigations (including deviations, nonconformances, yield discrepancies, out-of-tolerance results, and so on).
- Customer complaints.
- Annual product reviews (review of all lots of each commercial product).
- Environmental monitoring and control.
- Training.
- Stability program.
- Retention of retains and samples.
- Sampling.
- Validation.
- Handling quality crisis.
- Making recall decisions.
- Escorting outside investigators, responding to regulatory agency inspection reports or communications.
- QC laboratory—performs all testing.^[18]

Making QA Decisions

When reviewing the seriousness of a product or quality problem and deciding whether to release a lot, consider the following questions.^[15] Please note that even a lot that meets its specifications may need to be rejected,

based upon environmental monitoring or other serious excursions.

- Does the product meet its specifications?
- What is its indication? What is the product used for: heart patients? immunocompromised patients? pediatric or elderly patients? (Please note that the critical item in a recall situation is whether there is a flaw in the product, not the indication for the product.)
- What is the dosage form? How is the product delivered? (Injectables are among the most demanding of dosage forms because their contents are immediately delivered to the bloodstream.)
- How serious is the problem? What is the risk involved? Is contamination a possibility? If so, what is the organism and its probable source? What does the contaminant do to people when delivered by your dosage form?
- What do the environmental monitoring data look like? Did the product pass a sterility test? (The results of environmental monitoring—including operator touch plates and EM data for the facility, air, water, and equipment—are more important than sterility tests, which do not guarantee sterility. Please note that the environmental monitoring data are also extremely important for the bulk drug substance for a biologic product.)
- Should this product be released? Would you take medication from this lot? Would you give it to your child?
- Finally, what other products or lots are affected? Were the same equipment, air, room, materials, or stoppers used for another product or lot? Look for the “ripple effect”: Other products might be similarly affected. If they to are affected or suspect, they too must be thoroughly investigated.

Items to Consider in a Recall Meeting

This information^[6] is not all-inclusive. Please consult your company’s QA executive, legal counsel, quality advisory team, recall procedure, or the specific regulatory requirements for handling a recall in the Code of Federal Regulations (21 CFR 7).

Identifying scope of the problem:

- Which product? Indication?
- How many units or lots?
- When and where distributed?
- Other products or dosage forms affected?
- When and how did you learn of the problem?
- Internal discovery or test result? If yes, give details.

- External discovery? If so, give nature and details of original complaint, with name, phone number, and address of source.
- Has corrective action been taken?
- Has there been any reply to the complaint?

Determining the cause of the problem:

- If unknown, what are probable causes?
- How is it being investigated?
- How soon will cause be identified?
- Did product fail to meet specifications?
- Is product or device being used for treatment or diagnosis?
- Relationship, if any, of the product or device to reported incident or adverse event?

Classifying the recall:

- Risk to patient?
- Life-threatening problem?
- Product indication and patient population?
- Seriousness of defect or recall issue?

Conducting the recall:

- Notify distributors—get confirmation of receipt.
- If life threatening, get all units back and destroy them, document their destruction, and ensure traceability.
- If a few units affected, notify by phone and follow-up by letter.
- Write a script for the phone work, and insist that individuals follow that script, so they do not babble.
- If sales technical representatives installed the product (such as with a device or instruments), have them retrieved.
- Before distribution, have the recall letter first reviewed by legal counsel; the letter should succinctly tell the recipients only the necessary information. Clearly identify the letter as a recall letter.
- Follow your recall SOP. Plan to conduct recall effectiveness checks.
- Notify the FDA local office (or the appropriate world regulatory agencies if your product is available outside the United States).
- Identify the product(s) and scope of the problem (all lots or units).
- What is the cause if known (if not, working on it)?
- What are the risks and the recommended classification of the recall?
- Is a 72-h field alert needed? A medical device report (MDR)? Biologic product errors and accident (E/A) report?

- How do you plan to conduct the recall? Have all affected lots (and/or products) been identified?

This last point is very important to avoid doing one recall and then having to do another a few days or weeks later. Such behavior also does not make regulatory agencies feel comfortable about either the state of control or the ability to handle a recall.

Pitfalls to Avoid

According to FDA compliance executives with the Center for Drug Evaluation and Research (the division that regulates drugs), recent GMP compliance problems include:^[16]

- Lack of process and computer validation.
- Out-of-specification results inadequately investigated.
- Failure to investigate injury complaints.
- Cross-contamination due to improper equipment cleaning, particularly with APIs.
- Lack of complete validation for aseptic processing.
- Poor practices by operators in aseptic areas.
- Inadequate environmental monitoring.
- No validation of analytical methods.
- Inadequate employee training.
- Failure to demonstrate that analytical methods used for stability testing indicated stability.

Common reasons for GMP warning letters and deficiencies observed during pre-approval inspections include laboratory controls, production and process controls, records and reports, and validation (cleaning, process, assay, and computer).^[17]

Every major world nation has its own version of the GMPs, although the basic principles are the same. European Union GMPs define pharmaceutical QA as a “wide-ranging” concept that covers all matters which individually or collectively influence the quality of a product. It is the sum total of the organized arrangements made with the object of ensuring that medicinal products are of the quality required for their intended use”.^[13] The U.S. FDA is widely viewed as one of the toughest regulatory agencies worldwide. Other world agencies such as the Medicines Control Agency (MCA) and European Medicines Evaluation Agency (EMA) based in London, BfArM (the German agency), the Health Protection Branch (HPB) of Canada, and Koseisho or the Japanese Ministry of Health are also viewed as being quite stringent. A general consensus (among Americans, in any event!) is that if you are fully in compliance with all International Conference

on Harmonization (ICH) guidances and the U.S. regulations, you will most likely be in compliance with other world regulations.

GCPs

Good clinical practices are regulations governing the conduct of human clinical trials. In the United States, regulations that govern this discipline include the following:

- 21 CFR 50: protection of human subjects.
- 21 CFR 56: institutional review boards.
- 21 CFR 312: IND.
- 21 CFR 314: NDA.
- 21 CFR 320: bioavailability and bioequivalence requirements.

Members of the European Union follow *Good Clinical Practice for Trials in Medicinal Products in the European Community* and *The Manufacture of Investigational Medicinal Products*, an annex to the *EC Guide to Good Manufacturing Practice*. The Japanese MHW has also published GCP regulations. However, since the publication of guidelines by the ICH, a consortium of individuals working to set a unified standard for the European Union, Japan, and the United States, the ICH GCP consolidated guidance has become the de facto standard for conducting clinical trials. The ICH guidance was developed taking into account current GCPs of the European Union, Japan, United States, Australia, Canada, the Nordic countries, and the World Health Organization (WHO).^[6]

The ICH guidance defines QA as “All those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with GCP and the applicable regulatory requirement(s).” Quality control, on the other hand, is defined as “The operational techniques and activities undertaken with the QA system to verify that the requirements for quality of the trial-related activities have been fulfilled”.^[6]

Specific responsibilities assigned to the QA/QC group of the company sponsoring the trial (the “sponsor”) are as follows:

- The sponsor is responsible for implementing and maintaining QA and QC systems with written SOPs to ensure that trials are conducted and data are generated, documented (recorded), and reported in compliance with the protocol, GCP, and the applicable regulatory requirement(s).
- The sponsor is responsible for securing agreement from all involved parties to ensure direct access to all

trial-related sites, source data/documents, and reports for the purpose of monitoring and auditing by the sponsor, and inspection by domestic and foreign regulatory authorities.

- Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly.
- Agreements made by the sponsor with the investigator/institution and/or with any other parties involved with the clinical trial must be in writing either as part of the protocol or in a separate agreement.

The guidance also requires that the sponsor ensure that all clinical trial materials are manufactured in accordance with applicable GMPs.^[6] In the clinical setting, duties that normally would be performed by QA in a manufacturing or GLP environment (such as selecting, monitoring, or auditing of the site, training, and so on) may be performed by clinical research associates (CRAs), the medical monitor's staff, or the CRO's staff. Who performs the work or what the department is called is not as important as ensuring that the required work is being thoroughly performed, documented, and monitored.

Pitfalls to Avoid

The clinical field is changing rapidly, and clinical professionals must stay abreast of the changes. There are major differences between the European, United States, Japanese, and ICH requirements, particularly in their organization and detail, although all four versions address the key issues. Any company wishing to conduct human clinical trials in any country must become familiar with and comply with all applicable regulations. At the time of the publication of this article, it was still a requirement that any company wishing to market a product in Japan, conduct clinical trials in Japan (despite the adoption of the ICH guidelines). And since the recent death of the young man participating in a gene therapy trial in the United States, clinical trials have been under increased scrutiny and enforcement. In the United States, inspections done to enforce GCP and GLP compliance are termed "bioresearch monitoring" inspections.

Problem areas reported by an experienced FDA investigator during inspections of clinical investigators include protocol violations, insufficient recordkeeping, failure to have consent forms completed prior to subject's beginning the trial, reporting of adverse events, and drug accountability (inventory recordkeeping).

Problems noted during inspections of the Institutional Review Board (also known as the Ethics Committee in Europe; the group that reviews and approves the research to be done in that facility, the clinical protocol, investigator's brochure, and so on), again by an experienced FDA investigator, include:

- Lack of written procedures.
- Lack of membership requirements.
- Not requiring an investigator's brochure.
- Inadequate informed consent.
- Approving research while lacking a quorum.
- Inadequate recordkeeping, particularly minutes.
- Inappropriate use of emergency approval.
- Inappropriate use of expedited approval.
- Inadequate continuing review of research.^[5]

Sponsors of clinical trials, their monitors, and/or the CROs they delegate to, are responsible for:

- Obtaining FDA or other world regulatory agency approval, where necessary, before study begins (some studies require the filing of an IND; simpler studies may sometimes be initiated after the approval of the Institutional Review Board or Ethics Committee).
- Manufacturing and labeling investigational products appropriately.
- Initiating, withholding, or discontinuing clinical trials as required.
- Refraining from commercialization of investigational products.
- Controlling the distribution and return of investigational products.
- Selecting qualified investigators to conduct studies.
- Disseminating appropriate information to investigators.
- Selecting qualified persons to monitor the conduct of studies.
- Evaluating and reporting adverse experiences.
- Maintaining adequate records of studies.
- Adequately monitoring clinical investigators.^[5]

Sponsors may transfer responsibility for any of their obligations to an experienced, reputable CRO, but under the regulations, such transfers are permitted only by written agreement. Responsibilities not specified in the written agreement are not transferred. Effective monitoring by the sponsor or their CRO is considered critical in preventing fraud at clinical sites. The monitoring or auditing is usually done by a QA employee, or someone performing that role for the sponsor.^[5]

INTO THE FUTURE

A recent PhRMA survey shows that from 1980 to 1997, the number of employees in QA/QC and Production decreased by 10,000—a reduction of 11.7%—while the number of medical R&D and marketing employees nearly doubled.^[2] Could there be any correlation between these statistics and the spate of GLP, GMP, and GCP warning letters and other regulatory actions taken against companies by the U.S. FDA?

There may be fewer QA/QC employees because work is being outsourced, there is more automation, and there is a reduction in force and plant consolidations as a result of mergers. Perhaps, in the push to speed products to market and cut costs, manufacturers are harming themselves by insufficiently staffing some areas, not always providing adequate training or supervision, making do with temporary employees, and operating by triage.^[4]

The consequences, however, for regulatory violations and severe quality problems in the United States depend upon the severity and duration of the issue at hand. The worst possible ramification, of course, is that a patient would be injured.

U.S. regulatory ramifications for noncompliance include invoking the application integrity policy (for fraud), serious inspection observations, a warning letter, import alerts, withheld research and/or product approvals, cancellation of government contracts, product recalls, seizure, consent decree of permanent injunction, civil money penalties, suspension or revocation of licenses, and prosecution (including indictments and temporary or permanent debarment, if found guilty).^[5]

Clearly, it is in every company's self interest to not only put the patient first, but also to follow all applicable regulations. In the recent past, we have seen extremely expensive civil money penalties, with the most expensive one to date being \$100 million for repeated GMP violations over a 6-yr period.^[13] As we enter the 21st century, let us commit to doing the right thing, putting patients first, investing in necessary resources, systems, and innovations to not only ensure patient safety but also continually improve our operations, and commit to remaining current with changes in worldwide regulatory expectations and meeting at least the minimum required regulations. Any company can have an effective QA system by:

- Putting meaningful, useful systems in place
- Continuously implementing improvements
- Having good and/or seasoned QA and QC professionals
- Having sufficient number of employees
- Giving employees a voice

- Having the support of senior management
- Conducting thorough problem investigations that determine and address root causes
- Remaining current with all applicable regulations and proposed regulations
- Networking with peers to remain current with industry practice
- Funding both compliance and business improvements
- Training and developing its people
- Staying current with what regulatory agencies, investigators, and inspectors are asking about.^[6]

Not having an effective quality system and continuing to be in noncompliance with regulatory agency requirements or expectations will most likely lead to serious consequences. We all have a responsibility to minimize risk—to the patients we serve as well as to our companies.^[6]

Regulatory Timeline

Pure Food and Drug Act. Creates one of the first government regulatory agencies (now known as the FDA); the culmination of 25 yr of lobbying, this act makes it illegal to sell “adulterated” or “misbranded” food or drugs.

Federal Food, Drug, and Cosmetic (FD&C) Act. *Tragedy:* Sulfanilamide “elixir” made with poisonous solvent causes 107 deaths. *Result:* Requires manufacturers to prove the safety of products before marketing.

Two Unrelated Events. Insulin amendment requires FDA to test and certify purity and potency of insulin. *Tragedy:* Nearly 300 deaths and injuries from distribution of sulfathiazole tablets tainted with phenobarbital. *Result:* FDA revises manufacturing and quality controls drastically, the beginning of what will later be called GMPs.

1962: Kefauver–Harris drug amendments. *Tragedy:* Thalidomide, a sleeping pill and treatment for morning sickness, causes birth defects in thousands of European babies. *Result:* Manufacturers must prove efficacy of products before marketing them and ensure stricter control over drug testing.

1978: CGMPs final rule for drugs (21 CFR 210/211). Establishes minimum CGMPs for manufacturing, processing, packing, or holding drug products.

1979: GLPs (21 CFR 58) final rule. Establishes GLPs for conducting nonclinical laboratory studies that support applications for research or marketing permits for human and animal drugs, medical devices for human use, and biological products.

1982: Tamper-resistant packaging regulations issued for OTC products. *Tragedy:* Acetaminophen-capsule poisoning by cyanide causes seven deaths. *Result:* Revision of GMPs to require tamper-resistant packaging.

1983: Two unrelated regulatory events. The Guide to the Inspection of Computerized Systems in Drug Processing initiates tighter controls on computers and computer validation. Federal Anti-Tampering Act makes it a federal crime to tamper with packaged consumer products.

1987: Guideline on general principles of process validation. FDA expectations regarding the need for process validation are outlined.

1992: Generic drug enforcement act. Precipitated by illegal acts involving abbreviated NDAs. *Result:* Created debarment penalty and initiated PAIs for all drug applications.

1996: Two unrelated events. Proposed revision to U.S. CGMPs for drugs and biologics (21 CFR 210/211) adds detail for validation, blend uniformity, prevention of cross-contamination, and handling out-of-specification results.

ICH Guidance for Industry: E6, GCP: consolidated guidance from ICH becomes the de facto standard for conducting human clinical trials.

1997: Electronic records final rule (21 CFR 11). Requires controls that ensure security and integrity of all electronic data.

1998: Draft guidances. Manufacturing, processing, or holding APIs and investigating out-of-specification (OOS) test results for pharmaceutical production.

Definitions and Other Useful Information

The following definitions are abstracted from the requirements of laws and regulations enforced by the U.S. FDA.^[9]

The FD&C Act defines drugs as “articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man or other animals” and “articles (other than food) intended to affect the structure or any function of the body of man or other animals.” It is the intended use that determines whether something is a drug. Thus, foods and cosmetics may be subject to the drug requirements of the law if therapeutic claims are made for them. The FD&C Act prohibits adulteration or misbranding of any drug and requires the “new drugs” be reviewed and approved by FDA before they go to market.

Adulteration means contaminated with filth or putrid. Over the years, the definition has been expanded to include products that are manufactured without following GMPs.

Misbranding applies to statements, designs, or pictures in labeling that are false or misleading as well as to the failure to provide required information in labeling.

Drug applications typically fall into three categories: an NDA, a new animal drug application (NADA), or an abbreviated new drug application (ANDA) for generic products.

Any individual found guilty of breaking the required regulations or performing a prohibited act per the U.S. FD&C Act (such as adulterating or misbranding a product) can be fined up to \$250,000. Any corporation or partnership found guilty of committing a prohibited act can be fined up to \$1 million.

ACKNOWLEDGMENTS

The author gratefully acknowledges Carolyn Hall of Dekker for all her unbelievable patience, hard work, and effort in seeing this chapter into print, as well as the entire Dekker editorial and production staff. Special thanks to John Parker, Ph.D., for recommending the author. The author also wishes to thank everyone who served as a source for this article, as well as Anne Montgomery of BioPharm Magazine and Stefan Schuber of Pharmaceutical Technology. Last but not least, the author gives lifelong thanks to Linda Smith of Genetics Institute, who graciously reviewed this chapter, to her clients, and to Maureen O'Connor and the former employees of Syntex Corporation (now Roche Bioscience) for teaching the author and many others about good quality assurance. Thank you all very much.

REFERENCES

1. *PhRMA Industry Profile 1998*; Pharmaceutical Research and Manufacturers of America: Washington, DC, 1998.
2. General: Good laboratory practice for nonclinical laboratory studies. *Code of Federal Regulations, Food and Drugs*; revised April 2000; Title 21, Part 58 U.S. Printing Office: Washington, DC.
3. DeWoskin, Robert S.; Taulbee, Stephanie M. *International GLPs*; Interpharm Press: Buffalo Grove, IL, 1993.
4. Immel, Barbara K. The forgotten GLPs. *Pharmaceutical Technology Analytical Laboratory Supplement*, February 1999; Advanstar Communications: Eugene, OR, 1999.
5. Anderson, Carl. One FDA investigator's view. *UC Berkeley Extension Drug Development Course*, San Francisco, CA, Fall 1999.
6. Immel Barbara K., When things go wrong; Part 1: The consequences of noncompliance. *BioPharm*, August 2000; Advanstar Communications: Eugene, OR, 2000.

7. International conference on harmonization, guide for industry: E6, good clinical practice—Consolidated, Guidance, Geneva, Switzerland, April 1996. Published in the Federal Register **9 May 1997**, 62(90), 25691–25709, www.fda.gov/cder/guidance/959fnl.pdf.
8. Maynard, Donald E.; Dent, Nigel J.; Yoshida, Hideo; Dewoskin, Robert S. *International Good Clinical Practice: The GCP Compliance Handbook*; Interpharm Press: Buffalo Grove, IL, 1996.
9. *Requirements of Laws and Regulations Enforced by the U.S. Food and Drug Administration*; FDA: Rockville, MD, revised 1997. Available at www.fda.gov/opacom/morechoices/smallbusiness/blubook.htm.
10. Current good manufacturing practice in manufacturing, processing, packing, or holding of drugs. *Code of Federal Regulations, Food and Drugs*; revised April 2000; Title 21, Part 210/211. U.S. Printing Office: Washington, DC.
11. 21 CFR parts 210 and 211 current good manufacturing practice: Amendment of certain requirements for finished pharmaceuticals; proposed rule. *Federal Register*; 3 May U.S. Food and Drug Administration, 1996.
12. Guide to good manufacturing practice for medicinal products for human and veterinary use and annexes. *The Rules Governing Medical Products in the European Community*; June European Commission: Brussels, Belgium, 1991.
13. Lewis, C. Investigator's reports: Manufacturing misdeeds cost abbot record-breaking payments. *FDA Consumer* 2000, 34 (3), May–June Available at www.fda.gov/fdac/departs/2000/300_irs.html.
14. Immel, Barbara K. A brief history of the GMPs: The power of storytelling. *BioPharm*, August 2000; Advanstar Communications: Eugene, OR, 2000.
15. Barbara K., Immel. Step up to the responsibility of QA and QC. *BioPharm February 2000*; Advanstar Communications: Eugene, OR, 2000.
16. Famulare, Joseph. In *Compliance Update*, 24th Annual GMP Conference, Athens, GA, March 2000.
17. Falcone, Marie. *FDA Inspection Trends*, WorldPharm Conference, Philadelphia, PA, September 1998.
18. Immel, Barbara K. The essence of quality assurance. *UC Berkeley Extension Course on Drug Development*, San Francisco, CA, Spring 2000.



Quality Systems Management

Gary C. Harbour

Pharmacia Enterprises S.A., Luxembourg-G.D., Luxembourg

Robert G. Kieffer

RGK Consulting, Palm Desert, California, U.S.A.

INTRODUCTION

Managing for quality is and always has been a critical task of management no matter what the industry. In the pharmaceutical industry today, achieving quality of products, services, and processes is as important as it ever was; however, today we need to achieve it with reduced resources. We need to do more with less. We need to be as concerned with the quality system's efficiency as well as its effectiveness. Rising regulatory and customer requirements make this ever more challenging.

In this article, we will define the components of the quality system. We will present the pros and cons of various models of a quality system—drug GMPs, the medical device quality system regulations, ISO 9000, and the Malcolm Baldrige award criteria. In addition to guidance on the design of the quality system, we will describe some techniques that we have found to be effective in managing it.

THE QUALITY SYSTEM

Traditionally in the pharmaceutical industry, the quality function is divided into two parts, quality control and quality assurance. Quality control centers on testing products to assure their compliance to specification. It is in general concerned with evaluating events from the past. FDA has recognized this deficiency with the often-quoted philosophy that “you cannot test quality into a product.” Quality assurance is focused on building quality into a product through activities like validation, process and environmental control, and documentation. Quality assurance is in general concerned with events in the present. Only relatively recently has the pharmaceutical industry begun to emphasize a quality management accountability for continuous improvement of processes, people, and culture. Quality management is about the future, about prevention, and management's role in improving the quality system.

A system is a group of parts, components, or processes that work together to achieve a common goal. For the

quality system, the common goal is understanding and meeting customer's needs. The components of the quality system are processes, people, and culture (see Fig. 1). There are four types of processes that comprise the system:

- **Management**—the management processes are the most important. Some of the management processes that are critical to the quality system are: the communication of policy, standards, directions, objectives, and priorities to the organization; periodic reviews of the quality system's health and need for improvement; creation of the quality culture; and ensuring the availability of adequate resources.
- **Core**—the core processes are those that produce product for sale to a customer. For each step in the core process, the critical control points or variables and the acceptable ranges of these variables must be identified. The controls may be done in process by production or off-line in the QC labs. The quality of the core process depends principally on the design of the process and its validation; i.e., on its intrinsic capability.
- **Supporting**—some of the key supporting processes in a GMP plant are: audits, change control, failure investigations, labeling (artwork generation, labeling use and control), maintenance, materials management, complaint handling, product release, stability/expiration dating, training, trend analysis, and validation. Supporting processes add value in proportion to their contribution to the core process.
- **Supplier**—the supplier process is essential for the proper functioning of the core process. The quality of the core process depends to a large degree on the quality of incoming raw and packaging materials. The goal is supplier partnership. The quality of the supplier's own processes directly impacts the quality of your core process.

The focus on process quality is important because the work gets done through processes. The quality of the output depends on process quality and efficiency gains come from process improvements. It is also necessary to remember, “over the long haul, strong people cannot

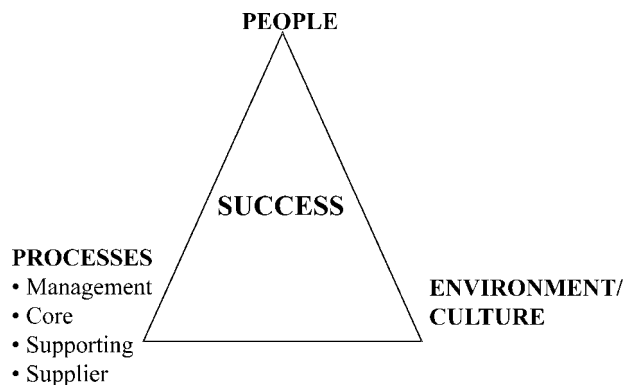


Fig. 1 Quality system components.

compensate for a weak process. All too often, management relies on individual or team heroics to overcome fundamentally flawed processes.”^[1] The principal components of a process are shown in Fig. 2. The most important component is ownership; i.e., the process is actively managed. One generally finds that departments are managed but processes are not. For internal processes, the supporting processes, the inputs, and outputs are generally information; information that is supplied by internal suppliers and used by internal customers. The activities in the process should be value added. It is not unusual to find that nonvalue added activities in the pharmaceutical manufacturing plant exceed 25%.

The second component of the quality system is people. Given that we have capable processes, we then need people with the requisite knowledge and skills. We have to hire and promote people with the potential for the task and importantly with the correct attitude. We need to develop them to their full potential. If we have high performance standards, all employees will have training needs. These must be prioritized. A tool for the determination of training needs and their prioritization is described in Ref. 2. Adequate time and resources must be allocated to training. World-class companies allocate about 100 hr yr⁻¹ per person for training. If improved

performance is the goal of training, its effectiveness must be evaluated by observation of performance or by improvement in business metrics.

The third component of the quality system is culture. Given capable processes and skilled employees, management needs to create a culture that fosters the desired behavior and that supports the natural motivation of employees to take pride in a job well done. Some aspects of the quality culture are:

- There is a clear vision of the desired future and how to get there.
- Management leads by example—if quality is number one, it is number one on management’s agenda and decisions are made accordingly.
- Performance standards are high, employees are held accountable for their actions and performance feedback is frequent, with emphasis on positive feedback.
- There is a lack of fear in the workplace, innovation/improvements are encouraged with the understanding that sometimes mistakes will be made.
- Individuals are respected and participate in the management of the organization to the extent of their capabilities.
- Learning, individual and organizational, is valued.
- Teamwork is fostered.
- Adequate resources are provided.

MODELS OF A QUALITY SYSTEM

The Food and Drug Administration’s (FDA) manufacturing regulations for drug products were published as Current Good Manufacturing Practice for Finished Pharmaceuticals in the Federal Register, Vol. 43, No. 190, September 29, 1978. These cGMPs along with various guidances published since 1978 provide the rules by which drug products to be sold in the United States must be manufactured. The cGMPs are not a good model of a quality system. They are not structured as a system, but are more a set of specific rules. In this limited sense, they define compliance but miss the mark in defining an adequate quality system. There are some significant flaws, vagaries, and omissions in them. The basic philosophy of these cGMPs is the philosophy of quality of the 1950s and 1960s—control, inspect, do not trust production, and the QC/QA department is responsible for quality.^[3]

A good model of a quality system is provided by FDA’s Quality System Regulations for medical devices.^[4] It is based on the ISO 9000 standard. These regulations are based on a modern, up-to-date philosophy of quality. They view quality as a total system, hold management responsible for quality, and emphasize the importance

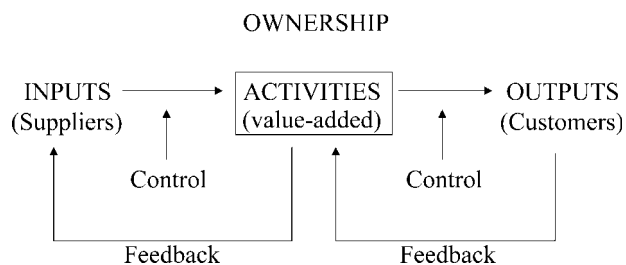


Fig. 2 Model of a process.

Table 1 Quality management principles

Customer focus
Leadership
Involvement of people
Process approach
Systems approach to management
Continual improvement
Factual approach to decision making
Mutual beneficial supplier relationships

Table 2 Malcolm Baldrige categories and point values

Categories	Point values (points)
1. Leadership	120
2. Strategic planning	85
3. Customer and market focus	85
4. Information and analysis	90
5. Human resource focus	85
6. Process management	85
7. Business results	450

of design and quality planning.^[3] Furthermore, FDA's inspections of medical device plants are system based.^[5]

Our recommended model of a quality system for a plant producing drug products is ISO 9000:2000, with the addition of the rules specific to drug products taken from the cGMPs. ISO gives the requirements for a quality management system, while the cGMPs gives the compliance requirements. The year 2000 version of ISO 9000 represents the current best thinking about quality and how to achieve it.^[6] The quality management principles as listed in Table 1 represent the philosophy upon which the quality system is built.

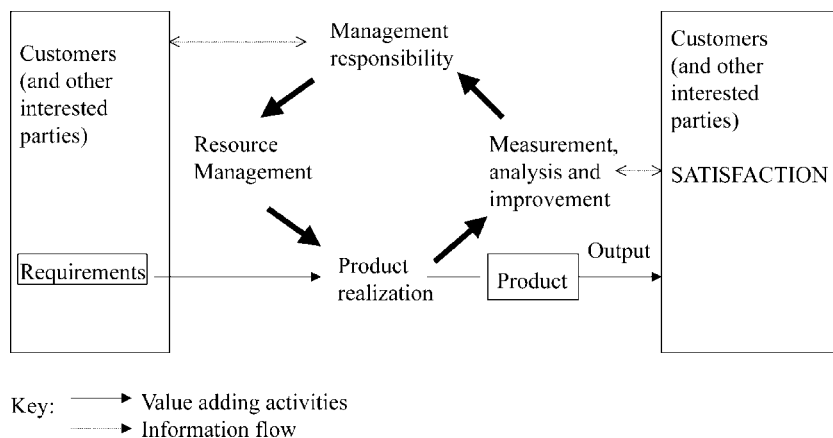
Fig. 3 represents ISO's view of the quality system. ISO 9000 emphasizes the criticality of the role of top management for the achievement of quality products and processes. This role is described in detail in the document. Besides the emphasis on system and process, ISO highlights the importance of culture/environment. "Through leadership and actions, top management can create an environment where people are fully involved and in which a quality management system can operate effectively."

The Malcolm Baldrige quality award represents a still higher standard.^[7] It is very comprehensive, involving all

aspects of the business. It's core values and concepts are similar to those of ISO 9000. The categories and point values are given in Table 2.

Note the heavy weighting given to results. The winners of this award generally score in the 700's. It provides a useful tool for a self-assessment, a benchmarking tool, even if your organization does not want to go to the expense of applying for the award.

After reviewing model quality systems like ISO, Malcolm Baldrige, or the European Quality Award, it is evident that cGMPs fall short of being a true quality system. They lack many elements common to these quality models such as leadership, planning, process orientation, and continuous improvement. A corollary concern is sometimes expressed that the pharmaceutical business is so regulated that a best practice quality system like ISO is not sufficient. However, a central tenant in all good quality systems is a customer focus, continuously evaluating customer needs/requirements and satisfaction. Companies that truly understand and put in practice these quality systems will have no trouble meeting regulatory customer requirements.

**Fig. 3** Continual improvement of the quality management system.

QUALITY PLANNING PROCESS

In order to continuously improve the quality system, there must be in place a quality planning process. The purpose of this process is:

- To evaluate the current quality system relative to the desired quality system, both short and long term.
- To develop action plans and allocate resources so that the desired improvements in the quality system are achieved.
- To monitor and measure performance against the plan.
- To adjust the plan as needed.

A diagram of the quality planning process is shown in Fig. 4. This planning process represents the first four steps in the continuous improvement cycle (Fig. 5), discussed in the “The Quality Improvement Process.”

The quality planning process begins with setting competitive standards, with creating the vision of excellence that ensures success and that focuses and motivates all employees. An important part of creating this picture of the desired future is benchmarking; both business and quality targets and the processes that enable the achievement of these targets. One can benchmark best internal performance and even other pharmaceutical plants/companies. However, some of the best quality practices are not found in the pharmaceutical industry, but are found in other industries. The books and magazines published by the American Society for Quality are a good source of best quality practices. One can also learn much from the Malcolm Baldrige winners.

The second step is to honestly and accurately assess one's current state. The performance of process assessments has proven to be an effective tool for doing this.^[8] This method differs significantly from the traditional compliance audit. It focuses on efficiency as well as effectiveness in achieving standards. It looks beyond observations or symptoms for root causes. It focuses on the quality of the processes and their outputs. It borrows from the Malcolm Baldrige award the evaluation of:

1. Approach (process design and process ownership).
2. Deployment (training, resource allocation, and integration among departments and with other processes).
3. Results (continuous improvement of performance metrics).

Once the desired end-point is determined and the current situation is assessed, a gap analysis is completed and objectives are set to close the gaps. An important step in the process is the prioritization of objectives, based on balancing available resources with risks. The development of the plan is ideally a participative one, involving the maximum number of people. After the plan is developed, personal and team performance objectives need to be aligned with the goals of the plan. Finally, it is necessary to integrate the quality plan with the overall business plan.

“What gets measured gets done.” The health of the quality system and the effectiveness of the quality plan need to be monitored. Therefore selecting the right set of quality metrics is important.^[9,10] Traditionally quality costs are broken down into preventive, appraisal, internal failure, and external failure costs. A commonly used

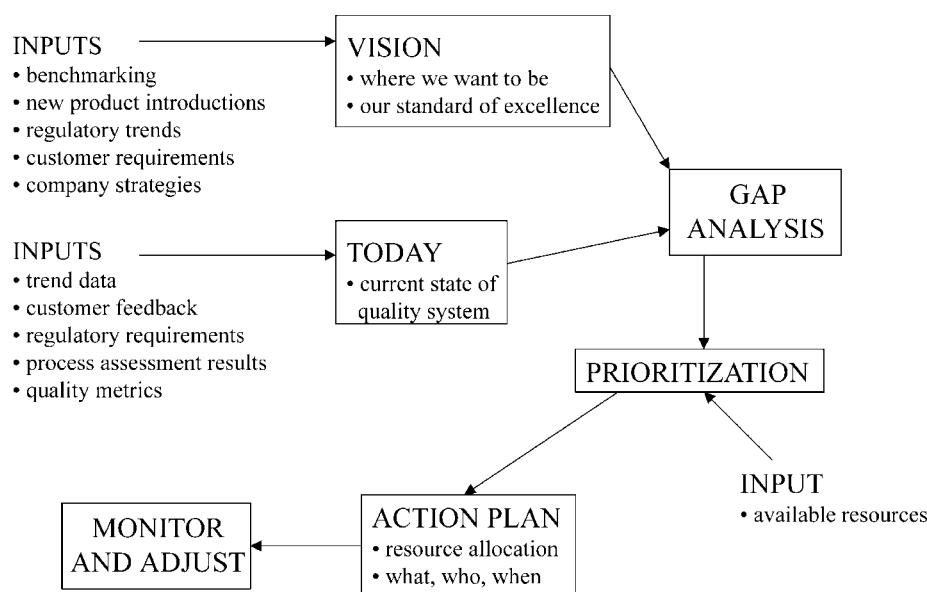


Fig. 4 The planning process.

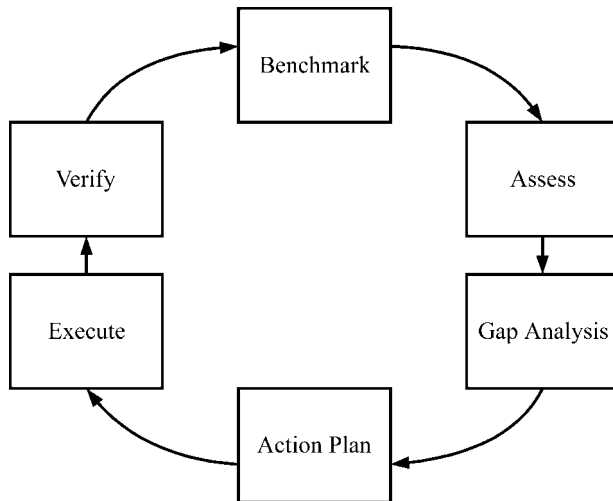


Fig. 5 The continuous improvement process.

indirect measure of internal failure costs is not-right-first time. Cycle time is another useful indirect measure of quality. Process quality can be measured.^[8] Processes can be rated on a 1–5 scale where:

1. Process and ownership not defined, personnel are not trained, results show process is ineffective.
2. Process and ownership are defined, personnel are trained, process metrics are in place; but process is not always followed and results need improvement to meet minimum standards.
3. Process is consistently followed and results are satisfactory.
4. Process has been continuously improved in effectiveness, efficiency, and cycle time over the past 2 yr. It is appropriately fail-safe and mistake proof.
5. Based on benchmarking process is considered world class.

For each key process there should be a set of metrics. These process metrics could consist of process cost, effectiveness, right-first-time, and customer satisfaction. Examples of metrics for the change control process are:

- Cycle time.
- Changes made without an adverse product impact (%).
- Number of changes successfully implemented per year.
- Customer/employee satisfaction with change control process.
- Completeness of documentation.

The quality metrics need to be reported to management so that they can determine the health of the quality system

and take whatever action is necessary to improve it. The importance of metrics and their selection cannot be over emphasized. “People need an assessment of some kind to give them a sense of where they really are. They need to know the starting point from which they will proceed in improving the quality of their contribution.”^[11]

THE QUALITY IMPROVEMENT PROCESS

The link between quality planning and quality improvement is execution. The continuous improvement cycle (Fig. 5) starts with benchmarking and internal assessment to arrive at a gap analysis. From the gap analysis an action plan with assigned leadership, resources, timelines, and metrics is developed. From this quality plan, the plant’s human resources must successfully execute the plan for improvement to occur. It is the leadership’s responsibility to assure that the plan is “do-able,” appropriate and adequate resources are assigned, and that the culture of the plant supports change. This starts with the project leaders who must have not only the technical skills, but also the project management and people skills to accomplish the project. Training in the proper skills to lead a quality improvement project is available from many consulting firms promoting the concept of “Six-Sigma.” People trained in these skills are often called “Black Belts.”

Once the project has been completed and determined to be successful by appropriate metrics, it is time for management to reinforce the culture of continuous improvement through appropriate team rewards and recognition. It is best practice for the plant to reinvest some of the saving from any improvement in other improvement opportunities.

RESPONSIBILITIES IN THE QUALITY SYSTEM

Quality must be every employee’s responsibility. The quality department will never succeed in forcing quality into products and processes. Instead, the plant’s management team must, by their actions, hold each employee accountable for the quality of their work. Management must, in turn, assure that all employees understand the requirements, have the necessary skills, and have adequate time and resources to perform their job.

The specific responsibilities of some key management personnel are:

1. Plant manager, the champion of quality
 - Establish the expectation that quality is everyone’s responsibility by his leadership actions.



- Establish in the total organization a quality culture of customer focus, continuous improvement, and accountability for quality.
- Assure appropriate resources are provided for quality and compliance.
- Make quality planning a prime component of the plant's planning process.
- Recognize and reward quality performance.

2. Quality manager

- Own the system/process assessment process.
- Facilitate system/process design and improvement teams.
- Assure the organization is adapting to changes in the external environment.
- Develop the needed quality knowledge and skills within the organization.
- Interpret the compliance and customer requirements.
- Assure quality/compliance accountabilities are fulfilled by organization.

3. Production manager

- Assure processes are performed, measured, controlled, and documented as designed.
- Provide opportunity for the workforce to identify and implement quality and efficiency improvements of processes.
- Identify root causes of failures and implement preventive actions.

4. Engineering manager

- Provide an infrastructure that is appropriate, calibrated, validated, controlled, documented, and maintained.
- Provide technical support to the improvement process.

5. All employees

- Focus on the customer, internal, and external.
- Continuously improve one's process.
- Be accountable for the quality of one's work.

- Sets direction.
- Creates customer focus.
- Establishes clear and visible values.
- Sets high expectations.
- Creates strategies and systems to achieve excellence, stimulate innovation, build knowledge and capabilities.
- Inspires and motivates the workforce.
- Encourages all employees to contribute, to develop and learn, to be innovative, and to be creative.

To put these leadership goals into practice, we have established Quality Management Teams (QMT) in each of our plants.^[12] The QMT consists of the complete plant management team—quality, production, engineering, technical support, human resources, and finance. The plant managers chair the QMT to symbolize their responsibility for quality and because they have the power to assure all employees follow the quality principles. The purpose of the QMT is:

- To provide plant management with a forum to exercise their responsibility for the establishment and maintenance of an adequate and effective quality system.
- To foster a partnership to manage continuous improvement in quality by focusing on processes, people, and culture.
- To jointly plan for improvement of the quality dimensions of business performance and to own the quality plan.

The QMTs meet monthly for at least an hour and quarterly for a more extended time to perform a complete review of the quality system and quality plan.

SUMMARY

Quality is fundamentally about customer focus. Both the pharmaceutical industry and the FDA have this in common; their customer is the patient. While quality control and quality assurance are important, best quality practice today emphasizes the management of the quality system. The quality system consists of important processes, highly skilled people, and an empowering/motivating culture. Training of personnel at all levels must be a priority for management. The recommended model of a quality system is ISO 9000 with the specific requirements of GMPs added. Continuous improvement of the quality system is assured by having an effective quality planning process. All employees must take responsibility for the quality of their work. A critical but difficult process is communication. Management must provide the direction (top–down), listen (bottom–up), and

LEADERSHIP AND MANAGEMENT

We have seen already that both ISO 9000 and the Malcolm Baldrige award criteria give prominence to the role of top management in establishing, maintaining, and improving the quality system. Visionary leadership according to the Malcolm Baldrige award:



encourage cross-functional cooperation. It is clear that quality depends on the leadership of the organization.

REFERENCES

1. Rummler, G.A.; Brache, A.P. *Improving Performance—How to Manage the White Space on the Organization Chart*; Jossey-Bass: San Francisco, CA, 1990; 46–47.
2. Kieffer, R.G.; Stoker, J.R.; Nally, J.D. The Knowledge and Skills of the Successful QA/QC Manager. *Pharm. Technol.* **1998**, *Jan*, 70–78.
3. Kieffer, R.G. Global Trends, Needs, Issues. *PDA J. Pharm. Sci. Technol.* **1998**, *July–August*, 151–153.
4. Federal Register, Vol. 61, No. 195, October 7, 1996, 52655–52662.
5. Guide to Inspections of Quality Systems, Food and Drug Administration: Washington, D.C., August 1999.
6. ISO 9000 International Standard, American Society for Quality: Milwaukee, WI.
7. Baldrige National Quality Program, National Institute of Standards and Technology: Gaithersburg, MD.
8. Nally, J.; Kieffer, R.G.; Stoker, J.R. From Audit to Process Assessment—The More Effective Approach. *Pharm. Technol.* **1995**, *Sept*, 128–140.
9. Kieffer, R.G.; Stoker, J.R. Quality Performance Measurement and Reporting. *Pharm. Technol.* **1986**, *June*, 54–58.
10. Stoker, J.R.; Kieffer, R.G. QA/QC Effectiveness and Efficiency. *Pharm. Technol.* **1997**, *Mar*, 134–144.
11. Karl Albrecht, Chairman Karl Albrecht International, personal communication.
12. Harbour, G.C.; Kieffer, R.G. Quality Management Team (QMT)—Managing the Quality System. *Pharm. Technol.* **2000**, *Oct*, 152–158.

RADIOCHEMICAL METHOD OF ANALYSIS

R. Raghavan

Jose C. Joseph

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

INTRODUCTION

In nuclear medicine, drugs containing radioactive metals, metal complexes, and metal conjugates are used for diagnosis and therapy of various diseases. Radioactive materials used as pharmaceuticals are not only small organic and inorganic molecules but are also macromolecules such as monoclonal antibodies and antibody fragments that are attached to radioactive metals. Nuclear medicine has become a \$12 billion medical industry, and more than one-third of the hospitals in the United States currently use radioisotopes for such procedures. It is anticipated that diagnostic procedures in the United States are likely to exceed 20 million by the end of 2000. The successful use of radiochemicals needs a basic understanding of radiation, radioactivity, and the nature and characteristics of instruments to detect and quantitate radiation. This article addresses these applications related to radioactivity and radiochemical methods of measurement.

ATOMIC STRUCTURE, NUCLEAR STABILITY, AND RADIOACTIVITY

Atomic physics describes the structure of atoms in complex mathematical terms of quantum mechanics. However, the model of the atom as described by Niels Bohr in 1913 is very simple, pictorial, and more than adequate for a basic understanding of the phenomenon of radioactivity. Bohr's planetary model of the atom consists of a dense positively charged nucleus surrounded by negatively charged electrons (e) in orbits of well-defined energy states. The nucleus consists of positively charged protons and neutral particles called neutrons. The protons and neutrons are held together by very a strong nuclear force of attraction, effective at very close distances (approximately 10^{-13} cm). These strong forces for each nucleus are computed in terms of binding energy. The electroneutrality of the atom is maintained by the orbital electrons, which are equal in number to that of the protons. This number is called the atomic number, Z . The masses of the atoms (A) and other particles are described in terms of

atomic mass units (amu). The amu is defined as 1/12th the mass of a carbon atom with atomic mass of 12.0000. The properties of these nuclear particles known as nucleons and electron are summarized in Table 1.

Any configuration of protons and neutrons is called a nuclide. There are three nuclides of the element hydrogen with atomic number 1. Some characteristics of the nuclides are listed in Table 2.

The notation A_ZX is used to indicate the nuclide of an element. The three nuclides of hydrogen are called isotopes of hydrogen. Tritium with an N/Z ratio of two is unstable. When the N/Z ratio becomes higher, the nucleus become unstable and results in the disintegration of the nucleus so as to achieve a stable N/Z ratio and therefore a stable nucleus. This process is called radioactive decay. This radioactive process can be spontaneous in some naturally occurring nuclides; then these elements are said to be naturally radioactive. When such instability is brought about by bombarding stable nuclides with high-energy particles, it is called artificial radioactivity. Of nearly 3000 known nuclides of elements, which are either man-made or natural, 287 nuclides of 83 elements are stable; the rest are unstable to varying degrees. The unstable nuclides disintegrate to form stable nuclides with release of energy and nuclear particles. Binding energy and nuclear stability are found to be dependent on the ratio of number of neutrons to protons (N/Z or n/p ratio). To be stable, at least one proton is required. The most stable heavy nuclide is ${}^{209}_{83}X$, with 83 protons and 126 neutrons ($N/Z = 1.5$). In general, when the N/Z ratio is greater than 1.6, the radioactive nuclide readjusts to a stable ratio of N/Z with the release of energy and particles of matter. The three nuclides of hydrogen are called isotopes. Other members of the nuclide family are isobars and isotones. The characteristics of these nuclides are given in Table 3.

Radioactive Decay

Different radioactive species undergo disintegration at different rates. The rate of this decay or activity is characteristic of the individual nuclide and is proportional to the number of radioactive nuclides present at the beginning of this time interval. The proportionality constant

Table 1 Mass, charge, and energies of nucleons and electrons

Particle	Mass (Kg)	Mass (amu)	Charge	Energy ^a (MeV)	Comments
Electron (e)	0.9108×10^{-30}	0.000549	-1	0.511	—
Proton (p)	1.6721×10^{-27}	1.00728	+1	938.8	—
Neutron (n)	1.6744×10^{-27}	1.00867	0	939.9	—
Alpha (α)	6.6465×10^{-27}	4.003874	+2	3726.7	Binding energy of the α -Particle is 28.29 MeV
Beta (β^-)	0.9108×10^{-30}	0.000549	-1	0.511	Variable kinetic energy depending on how ejected
Positron (β^+)	0.9108×10^{-30}	0.000549	+1	0.511	Variable kinetic energy depending on how ejected
Neutrino (ν)	0	0	0	—	—
Photon	0	0	0	—	Varying energy

^aEnergy is based on the rest mass of the particle.

is called the decay constant and is denoted by λ . The decay constant is a measure of the probability that a certain radioactive nucleus will disintegrate within a specified time interval. These disintegrations are characteristic of the nuclide and are unaffected by pressure, temperature, concentration, and other physical or chemical properties of the radionuclide. This rate constant is conveniently denoted in terms of $t_{1/2}$, or half-life. The half-life of a radionuclide is the time required for the sample activity to decrease to half its initial value. $t_{1/2}$ is related to rate constant (λ) as follows:

$$\lambda = 0.6932/t_{1/2}$$

In fact, less than 1% will be radioactive in seven half-lives, and after 10 half-lives, greater than 99.9% of the radioactive nuclide will have lost its activity. The half-life refers to that of a pure nuclide. In a sample containing mixtures of disintegrating radionuclides, the total activity is the sum of the separate activities. From a plot of relative activity against time, the individual half-lives can be computed. However, in practice, this can be realized for mixtures containing 3 or <3 nuclides. Other commonly used terms in nuclear medicine and pharmacy are average (mean) half-life, biological half-life, and effective half-life. Average half-life is the mean lifetime of a nuclide, and it is equal to $1.44 \times t_{1/2}$. Biologic half-life, t_b , is the time required

for the body to eliminate half the administered dose by normal biological process of elimination. Effective half-life (t_{eff}) is a measure of how fast the body eliminates the radioactive material by the combination of biological elimination and radioactive decay:

$$1/t_{\text{eff}} = 1/t_b + 1/t_{1/2}$$

Unit of Activity

The fundamental SI unit of activity is the Becquerel (Bq). One Bq is equal to one disintegration per second (dps). Because this is a very small unit, it is more often expressed in kilobecquerels or kBq. However, the older historical unit of activity C_i is normally used for radiopharmaceuticals. The Curie was defined in terms of the number of disintegrations per second of 1 g of ^{226}Ra and is equal to 3.7×10^{10} dps. Other commonly used units are millicurie and microcurie (mCi and μCi). The unit of C_i represents absolute activity (A). However, relative activity R is proportional to the efficiency of the counting device. The device reports in counts per minute.

$R = qA$; q = efficiency quotient. Sometimes specific activity, in terms of radioactivity per unit mass of an element or radiolabeled compound or unit volume of solution is also specified. In these case, the mass or volume should be clearly specified.

Table 2 Configuration of nuclides of hydrogen

Nuclide	No. of protons (Z)	No. of neutrons (N)	N/Z (n/p) ratio	Notation	Nuclear stability
Hydrogen	1	0	0	^1_1H	Stable
Deuterium	1	1	1	^2_1H	Stable
Tritium	1	2	2	^3_1H	Unstable

Table 3 Characteristics of nuclides

Nuclide	Atomic number (Z)	Mass number	Neutron number	Chemical property	Examples
Isotopes	Same	Different	Different	Same	$^{15}_8\text{O}$, $^{16}_8\text{O}$, $^{17}_8\text{O}$, $^{18}_8\text{O}$
Isobars	Different	Same	Different	Different	$^{67}_{29}\text{Cu}$, $^{67}_{30}\text{Zn}$, $^{67}_{31}\text{Ga}$,
Isotones	Different	Different	Same	Different	$^{59}_{26}\text{Fe}$, $^{60}_{27}\text{Co}$, $^{62}_{29}\text{Cu}$, (33 neutrons each)
Isomers	Same	Same	Same	Same	$^{99\text{m}}_{43}\text{Tc}$, and $^{99}_{43}\text{Tc}$. ($^{\text{m}}$) meta stable isomer

Decay Processes

The radioactive decay process involves the emission of radiation, which is dependent on the mode of decay of the particular radionuclide. Radiation resulting from any decay process can be classified as alpha (α), beta (β), gamma rays (γ), and/or other emissions.

α -Particles

Alpha (α)-particles are doubly charged, highly energetic helium nucleus. α -Particles originate in the nuclei of heavier atoms. The emission involving α -Particles is the most efficient process for a radionuclide to attain stability because the nuclide loses both charge and mass. Because α -Particles are very heavy (7400 times of that of an electron) and doubly charged, they attract electrons from the surrounding medium when they pass through a medium. This results in the ionization of the medium. An α -Particle loses energy, slows down, and finally becomes a helium nucleus. In each ionization process, it loses 34 eV per event. For example, a loss of 3.4 MeV of energy causes 100,000 ionization events. These particles travel very short distances, called the range. Because they are extremely efficient in ionizing, they lose energy very rapidly. This range is approximate 4 cm in air and a few thousandths of a centimeter in biological tissues. Because of this tremendous amount of energy transfer, α -Particles can cause extensive damage to organs and tissues when they are exposed to this radiation. Generally, α -Particles arise during the natural decay of elements with a Z value greater than 83. A typical alpha decay process is represented below:

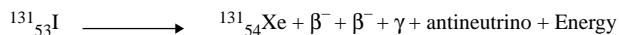


β -Particles

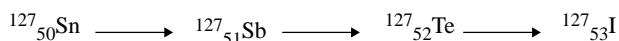
A β -Particle is a high-velocity nucleon ejected out of a decaying nucleus. β -Particles have the rest mass (0.000548 amu) of an electron. If it is negatively charged, it is called a negatron and if positively charged, it is called a positron. In common usage, β -Particle emission refers to

the negatron (β^-) and β^+ -Emission is called positron emission. For example, the decay of $^{32}_{15}\text{P}$ to $^{32}_{15}\text{S}$ results in the emission of β^- , antineutrino, and release of energy. The difference in the masses of the two nuclides, 0.001836 amu (31.965675 for P – 31.9638390 for S), results in the release of energy equivalent to 1.70 MeV. Part of this energy is used up in the ejection of particles, and the remaining is used in the release of antineutrino. Because no other energy is emitted, ^{32}P is called a pure β -Emitter. Sometimes some portion of the energy difference may be retained in the nucleus, and consequently the nuclide resides in a nuclear excited state. The excited nucleus may lose the excess energy and return to the stable ground state in the form of electromagnetic radiation (gamma-rays). β -Particles are emitted with variable energy from zero to the maximum of the difference in energy between the parent and daughter nucleus. The variability in energy arises from the distribution of energy between the β -Particle and the antineutrino.

Negatively charged β -Particles are emitted when the radionuclide has more neutrons than required by the number of protons for stability. Sometimes more than one β -Particle may be emitted. For example, ^{131}I decays with the emission of six β -Particles and 14 γ -Rays of different energies:



Nearly 100% of all these decay processes correspond to β -Particles of energy 0.61 MeV and γ -Radiations of energy 0.364 MeV. The decay to form stable nuclides may involve the formation of many unstable intermediate radionuclides. For example, the decay of $^{127}_{50}\text{Sn}$ to $^{127}_{53}\text{I}$, shown below, involves the formation of a number of intermediate nuclides:



In general, β^- - and β^+ -Particles penetrate deep into the medium; however, they do not cause damage to tissues and

organs. Radionuclides that decay by β -Particle emissions are used very extensively in nuclear medicine for diagnostic and therapeutic applications. Positron-emitting nuclides are used in nuclear medicine for diagnostic purposes. β^+ -Emitting radionuclides are under active study for use in radiotherapy. An example of a decay process involving emissions of positrons follows:



γ -Rays (γ)

During the disintegration of the nucleus, part of the energy is used in the creation of excited-state radionuclide. In the excited state, the nuclide is unstable. The release of energy while the excited nucleus returns to the ground state appears as electromagnetic radiation. These radiations are called γ -Rays. γ -Emission is common when the difference between the excited state and the lowest energy ground state nucleus is greater than 100 keV. Because γ -Emission is electromagnetic radiation, there is no change in the neutron number or mass number or atomic number. Therefore, invariably, γ -Radiation is always preceded by a nuclear decay reaction involving emission of α , β^- , β^+ particles.

γ -Rays are high-energy electromagnetic radiation such as X-Rays with no electrical charge. γ -Rays are different from X-Rays in that they differ with respect to their origin. X-Rays originate from orbital electrons, whereas γ -Rays originate from the decay of a nuclide. The γ -Rays also ionize the medium by striking orbital electrons with high energy and by knocking the electrons out of the atom. These ejected electrons cause secondary ionization referred to as indirect ionization. The degree of penetration of γ -Rays is extremely high. There are at least seven different processes by which γ -Rays can interact with matter. However, the processes associated with γ -Ray interaction or production, which are pharmaceutically relevant, are electron capture, isomeric transition, and internal conversion. These are briefly discussed below.

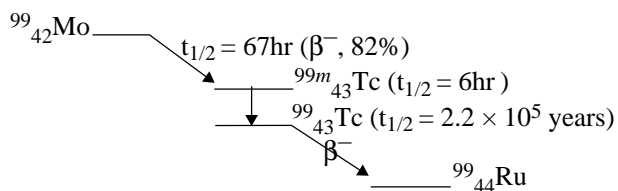
Electron capture is a decay process in which an orbital electron loses energy and becomes absorbed into the nucleus. As a result, outer-shell electrons jump to fill the inner-shell vacancy resulting in an electron orbit reshuffling. This process results in the release of X-Rays, and these electrons are subsequently absorbed and result in the release of weakly bound orbital electrons. These are called Auger electrons. When the decay process involves positron emission, it is almost always followed by electron capture. As a result of nuclear reaction, an excited radionuclide (with energy >100 keV) loses energy by de-excitation:



More often, these occur in multiple steps with the release of γ -Ray photons of multiple energy. Thus, the resulting γ -Ray spectrum is unique to the decaying radionuclide; this uniqueness, therefore, is used to identify the unknown nuclide. γ -Rays with no mass and charge can penetrate into matter and bring about other chemical reactions in the system. γ -Rays are widely used in nuclear medicine.

An excited radionuclide may remain in several excited states before reaching ground state. However, transitions can occur within these excited states with the emission of γ -Rays. These transitions are called isomeric transitions. When these isomeric transitions are significantly long-lived, these are called metastable states.

An example of the decay process is shown below. (\nearrow — increase in Z, \nwarrow — (left arrow) decrease in Z)



A process alternative to isomeric transition is called internal conversion (IC). In some cases, the γ -energy is absorbed by a K-shell (inner) electron. This electron is ejected out with lower energy. This ejected electron is the internal conversion electron, and the process is called internal conversion. Some diagnostic and therapeutic radionuclides with corresponding half-lives are given below.

Diagnostic radionuclides: ($t_{1/2}$ in units as indicated are given in parentheses.) $t_{1/2}$ hours: ^{67}Cu (62.0); ^{67}Ga (78.3); ^{90}Y (64.1); ^{117}In (67.9); ^{99m}Tc (6); ^{201}Tl (72) { β^- -particles}- t min; ^{11}C (20.4); ^{13}N (10.1); ^{15}O (2.0); ^{18}F (1.9); ^{68}Ga (68.0); and ^{82}Rb (1.25) { β^+ }.

Therapeutic radionuclides: $t_{1/2}$ days: ^{32}P (14.3); ^{47}Sc (3.4); ^{67}Cu (2.6); ^{64}Cu (0.5); ^{90}Y (2.7); ^{105}Rh (1.5); ^{111}Ag (7.5); ^{117m}Sn (13.6); ^{131}I (8.0); ^{149}Pm (2.2); ^{153}Sm (1.9); ^{166}Ho (1.1); ^{177}Lu (6.7); ^{186}Re (3.8); ^{188}Re (0.7).

RADIATION DETECTION AND MEASUREMENT

Interaction of Radiation with Matter, Ionization Chamber, and GM Counters

The detection and quantitation of nuclear radiation are based on its interaction with material contained in the

detector. Ionization of the gas particles in the medium and scattering are the two most common types of interaction of radiation with matter. Radiation causes darkening of photographic emulsion, ionization of a gas or a mixture of gases, or fluorescent scintillation. In radiology, exposure and observation of the photographic film are most commonly used. When α -Particles with high kinetic energies impinge on gases enclosed in a chamber, they produce approximately 40,000 ($\pm 10,000$) ion pairs per cm. The range of α -particles is short (approximately 6 cm in air and 2 μM in lead); β -Particles have longer and irregular paths and produce approximately 100–700 ion pairs per cm. A typical ionization chamber, shown in Fig. 1, consists of a sealed tube containing helium, neon, or gas–air mixtures placed in a space between two electrodes. The incoming particles create ion pairs; the ions are separated and collected at the electrodes of opposite charge. Electrons are collected at the anode. The number of such ions collected at each electrode is also a function of the applied voltage across the electrodes. If the response is measured in terms of pulse height (proportional to the number of electrons collected), a plot of this value against applied voltage will exhibit a characteristic response as given in Fig. 2. The different regions in Fig. 2 are explained below. Region I is called the recombination region. In this region, ion pairs produced increase linearly as a function of applied voltage (<100 V) as recombination is proportionately decreased. This is not a useful region for measurement. In region II (100–400 V), pulse height attains a plateau because all ions formed are collected. This region is used in ionization chambers to measure energies of γ -Rays and X-Rays. This region is also used for identification of α - and β -Particles because the height at which plateau occurs is characteristic of the nature of the particle. Dose meters and dose rate meters, which measure high-intensity radiation fields,

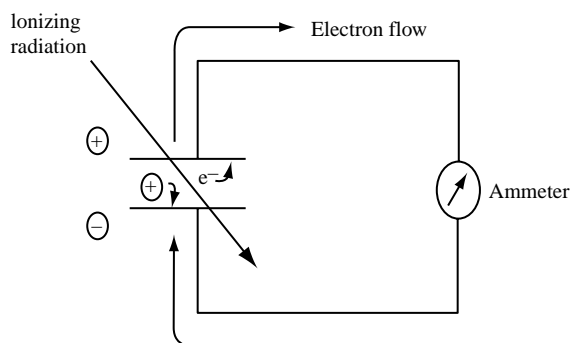


Fig. 1 Block diagram of a simple air ionization chamber. Electrons are attracted to the positive plate, and positive ions go to the negative plate.

operate in this region. In region III (400–800 V), the number of pulses produced is proportional to the intensity of radiation. The size of the pulses counted gives a measure of the primary ionization produced. In this region, the primary electrons are accelerated with high gain in energy. This high energy causes additional secondary electrons to be generated as a result of interaction with gases in the ionization chamber. These secondary electrons cause further ionization. The net effect is pulse amplification; the size of the amplification is of the order of 100–10,000 times that of the size of the ions initially produced in the ionization chamber.

In Fig. 2, the two response plots correspond to two particles of different energy. Ionization chambers operate in this region.

Region IV (applied voltage 800–1000 V) is limited proportional region and is not important for purposes of detection. When the applied voltage is between 1000 and 1500 V, the size of the pulse is no longer proportional to the initiating event. However, each pulse corresponds to a single event. The counters that operate in this region are called Geiger–Mueller counters or GM counters. The GM counters are very sensitive. Because of the high gain involved, these are used primarily in regions where the radioactivity is of very low intensity, (e.g., radioactive contamination). The readings are given in milli-, micro-, or Roentgens/h or in counts per minute. In the GM counter, ionization produced spreads to the entire gas. Therefore, the counter may not respond to a succeeding second incoming pulse before a recovery time of approximately 100–300 μs . This dead time for recovery is unusually high compared with the dead time for the proportional counter, which is approximately 1 μs . Secondly, because a fixed amount of gas is present in the chamber, the purity of the gas in the chamber decreases. Therefore, the GM counter has to be calibrated frequently, using standard ^{226}Ra or ^{137}Cs sources

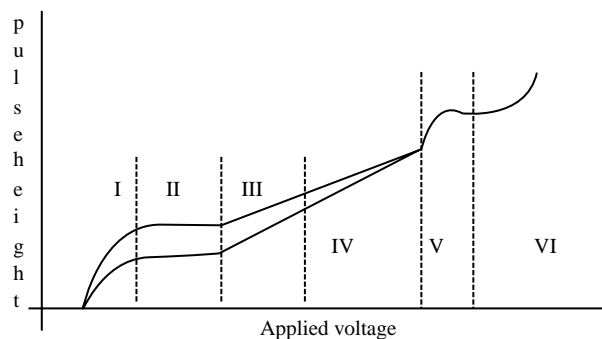


Fig. 2 Pulse height versus applied voltage response.

per Nuclear Regulatory Commission (NRC) requirements. Above an applied voltage of 1500 V, continuous discharge occurs, and this is not useful for any measurement.

Scintillation Detectors

When radiation interacts with certain substances called fluors or phosphors, it produces a flash of light called scintillation. The scintillation is then detected using a sensing element, amplified, sorted, and recorded by counting. The scintillation-detecting instruments include well counters, scanners, thyroid probes, and scintillation cameras called Auger cameras. In addition, all these instruments consist of a collimator (excluding well counters), photomultipliers, a high-voltage power supply, an amplifier, a gain control unit, a pulse height analyzer, and instruments or computers for appropriate display modes. The scintillation cameras also contain coordinate-positioning (x,y) circuits.

Solid-State Detectors

These detectors are made of semiconducting materials. In these detectors, solid-state electrodes are made from Li doped with Si or Ge. The resolution is approximately 1–2 keV for 1 MeV γ -Rays and sometimes provides a greater than 10-fold improvement over NaI (Tl) scintillation detectors, described below. These are commercially available and more often used in research-grade instruments.

A brief review of scintillation phosphors and their uses are given in Table 4.

Liquid Scintillators

The sample radionuclide is dissolved in a liquid scintillator called the scintillation cocktail. It consists of two principal components. The first is a primary solvent such as toluene, xylene, or 1,2,4-Trimethylbenzene (pseudocumene). The second component is the fluor solute, 2,5-Diphenyloxazole (PPO) that emits UV light at ~ 380 nm. The cocktail may also contain one or more the following:

- A secondary solvent such as dioxane to improve solubility of aqueous samples or surfactants such as sodium dodecylbenzenesulfonate as emulsifier
- A secondary scintillator to shift the wavelength of photons emitted (~ 380 nm) to the wavelength response of some photomultiplier tubes (PMT, ~ 420 nm)
- One or more adjuvants for purposes of suspending or solubilizing biological tissues.

A brief description of other components of scintillation detectors (generally applicable to liquid scintillators also) is given below.

Photomultiplier (PM) tubes

The PM tube has a light-sensitive electrode called the photocathode. It emits electrons when photons strike it. The electrons are then accelerated from the photocathode to the anode of PM tube by the application of approximately 1000 V in steps of approximately 100 V by a series of electrodes called the dynodes. In the PM tube, secondary electrons are produced, resulting in pulses of 10^5 to 10^8 electrons. Typically a phototube with 10 dynodes delivers approximately 4^{10} electrons. This gain or amplification is dependent on the dynode voltages.

Preamplifier

Even though such a large number of secondary electrons are generated, these are not adequate to generate enough current. The voltage pulse is amplified by a factor of 4–5 by the preamplifier without loss of power. The preamplifier also provides the driving force necessary to prevent loss in the several feet of connecting cables.

Linear amplifier

The pulses received from preamplifier have wide variation in energies of the particles. The gain in the amplifier is of the order of 8000 such that a 1-mV signal is amplified to approximately 8 V while still maintaining the proportionality of the energy delivered by the particle (more often γ -Ray) to the detectors. The amplified pulse is then delivered to pulse height analyzer.

Pulse height analyzers (PHA)

The pulses that emerge from the amplifier usually have different amplitudes owing to differences in energies. These analyzers are essentially energy sorters. Single-channel PHAs count pulses of a given amplitude, whereas multichannel analyzers (MCAs) scan whole energy range and record the pulses in each channel. For example, by using an MCA, γ -Ray spectrum can be recorded and, thus, these instruments that use MCAs are called γ -Ray spectrometers. MCAs may have as many as 4000 channels.

X-Y-positioning circuits

These are unique to scintillation cameras, known as Auger cameras, used in nuclear medicine studies. Approximately 19–91 PMTs are mounted on a Na(Tl) crystal used in the camera. These crystals are typically thick. The number of PMTs, which are optically coupled to the back of the crystal, is determined by the size and shape of the crystal. A maximum amount of light will be received by the PMT nearest to the point of interaction compared with the other PMTs, which are positioned differently. The amount of light received in these PMTs is proportional to the solid

Table 4 Scintillation phosphors and their uses

Particle detected	Phosphors used	Characteristics/comments
Alpha (α)	Thin sheet of plastic scintillator or zinc sulfide embedded in a transparent tape	Phosphors are wrapped in very thin aluminized mylar foil to exclude light. α -Particles, because of low penetrating power, lose all their energy to scintillators. β - and γ -Particles produce only smaller pulses and by adjustment of appropriate setting of the discriminator, α -Particles alone can be counted.
Beta (β)	Single crystals of anthracene, <i>trans</i> -stilbene are commonly used. Sometimes naphthalene doped with anthracene can also be used.	For β -Particle counting, especially low-energy particles from ^{14}C and tritium, liquid scintillators are commonly used (see discussion below on liquid scintillators).
Gamma (γ)	Sodium iodide doped with thallium iodide are used [NaI(Tl)]; thallium, which is present at 0.1–0.4% of sodium emits 420 nm scintillations. This phosphor is very efficient in absorbing γ radiation because of high atomic number of iodine and high density of sodium iodide.	Sodium iodide is hermetically sealed. It is normally used in well counters. A typical well counter used in laboratory and that used in cameras in nuclear medicine are show in Figs. 3 and 4.

angle subtended by the PMT. Therefore, X – Y -positioning of the camera has to be controlled and known so that X – Y -coordinate of the γ -Ray interaction can be assessed accurately. These data are stored in a computer and then processed or recorded on Polaroid or X-Ray films.

Efficiency of detection in scintillation detectors

In gas-filled as well as scintillation detectors, the observed count rate is typically less than the actual decay rate of the radionuclide. The efficiency of detection may differ from particle to particle under identical conditions using the same type of detector. The factors that affect the efficiency of detection are operating voltage, resolving time, geometry of the instrument used in relation to the position of the sample with respect to the detector, scaler, energy resolution, absorption by cells, and sometimes constituents of the sample itself. For example, the scintillation cocktail sometimes reduces counts considerably. This effect is known as quenching. For accurate measurements of radioactivity, appropriate correction for quenching is required.

Frequent calibration of the instruments with the use of appropriate standards is required to make suitable allowances for decreases in the efficiency of the instruments. Such calibration standards are available from the National Institute of Standards and Technology (NIST). Other sources traceable to NIST standards through active program of participation in comparison measurements also provide such standards. The United States Pharmacopeia (USP) also provides nuclear decay data for new calibration standard. USP 24 lists $t_{1/2}$, energy of photons, and number of

photons per disintegrations, for the following radionuclide standards: ^{137}Cs , $^{137\text{m}}\text{Ba}$, ^{22}Na , ^{60}Co , ^{57}Co , ^{54}Mn , ^{109}Cd , ^{109}Ag , and ^{129}I .

Tomographic Imagers

Tomography is a process in which three-dimensional images are constructed using a large number of two-dimensional slices of images from an object. Computed tomography uses rigorous mathematical algorithms to reconstruct these images. When radionuclide emissions are used, it is called emission tomography. Two common techniques are now used to obtain images using emission tomography, namely, single photon emission computed tomography (SPECT) and positron emission tomography (PET). In SPECT, γ -Emitting nuclides are used, whereas in PET, positron-emitting nuclides are used. In the SPECT system, an object is photographed using many Auger cameras at a number of small angles (between 3 and 10°) around the object. The total span may be between 0 and 180° or between 0 and 360°. Because the rotating cameras provide two-dimensional digital images, these are stored as 64 × 64 matrix (for 180° span) or as 128 × 128 matrix (for 360° span) digital information. The Auger cameras use NaI (Tl) crystals in the detector heads. In PET, positron-emitting radionuclides are used. Each positron emitted from these radionuclides travels through tissues, deposits energy, and finally is annihilated by interaction with an electron. The annihilation results in the formation of two photons traveling in opposite directions (180° apart) with energy of 0.511 MeV. Two Auger cameras are placed 180°

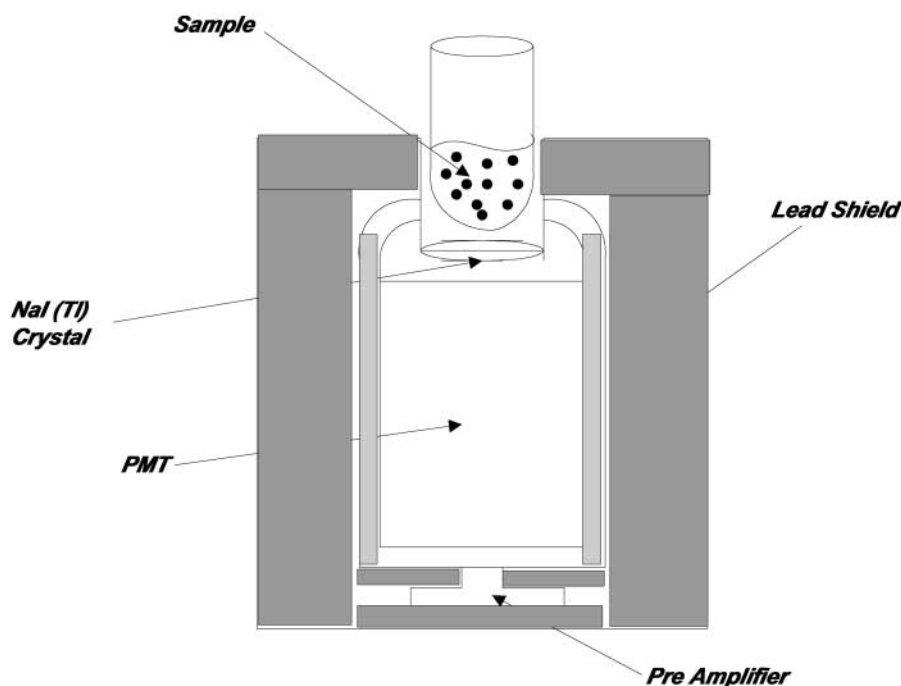


Fig. 3 A typical well counter used for measurement of gamma rays.

apart but at the same distance from the object. Each camera detects a photon at the same time. By appropriately moving cameras in pairs, the data are collected over many angles and stored in 64×64 or 128×128 matrix. Thus, this electronic collimation brought about by simultaneous detection increases sensitivity and reduces the need for use of collimators as in SPECT. Additional advantages of PET include easy availability of radionuclides with short half-lives of isotopes of elements commonly found in organic molecules (such as F, N, C, and O). (See USP 24 for additional details of requirements for radiopharmaceuticals for PET.)

ANALYSIS OF RADIOCHEMICALS

Radiochemical methods of analysis are considerably more sensitive than other chemical methods. Most spectral methods can quantitate at the parts-per-million (ppm) level, whereas atomic absorption and some HPLC methods with UV, fluorescence, and electrochemical methods can quantitate at the parts-per-billion (ppb) levels. By controlling the specific activity levels, it is possible to attain quantitation levels lower than ppb levels of elements by radiochemical analyses. Radiochemical

analysis, in most cases, can be done without separation of the analyte. Radionuclides are identified based on the characteristic decay and the energy of the particles as

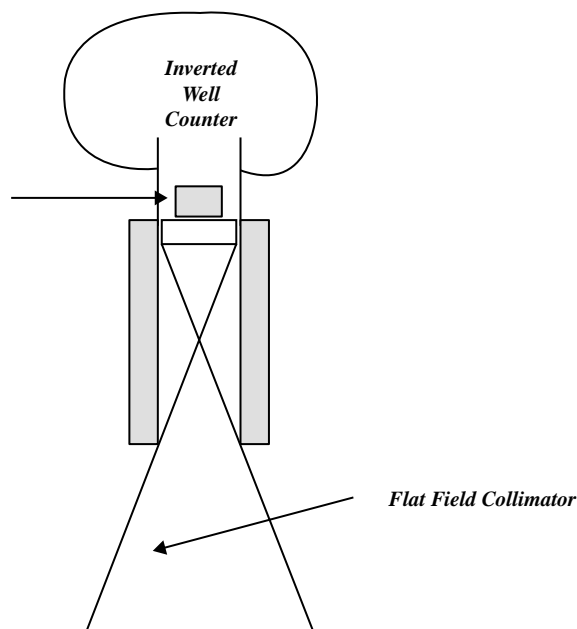


Fig. 4 A typical well counter used in nuclear medicine.

described in detection procedures presented above. Radiochemical methods of analysis include tracer methods, activation analysis, and radioimmunoassay techniques.

Tracers and Tracer Methods of Analysis

Radiochemical tracers or radiotracers are compounds labeled with radioisotopes. For tracer methods, the compound to be measured or a suitable reagent is radiolabeled. A measurement of the redistribution of tracer within such a sample–reagent reaction system provides the required quantitative analytical information. Major advantages of tracer methods are high sensitivity, simplicity, and speed. Radiotracers are more commonly used for following mechanisms of biological and/or chemical processes or if there is need to eliminate complicated separation procedures, especially in biological processes.

Isotopic dilution analysis

In isotope dilution analysis, a known amount of radiolabeled compound with known specific activity is spiked to a known amount of an unknown mixture containing the same compound made up of stable isotopes. Then, the components of the radiotracer-diluted samples are mixed thoroughly to form a homogeneous mixture. This isotopically diluted mixture, with known levels of dilution, is then suitably treated to isolate a small amount of the desired constituent. The radioactive isotope content of the isolated portion is determined by measuring its specific activity. From the specific activities of the tracer before and after dilution, the concentration of the component in the mixture can then be calculated. The major advantage is that these isolation procedures need not be quantitative; however, it is necessary that the compound isolated should be pure enough for an activity determination. Isotope dilution analyses are used for determination of inorganic trace elements and for the determination of organic compounds in biological systems. If separations are required and the radioisotope concentration is not adequate to bring about separation, dilutions can be accomplished by using a nonradioactive compound (or carrier) with similar chemical behavior. For example, if radio strontium is to be precipitated and it is in low amounts such that it cannot be quantitatively precipitated, it can be coprecipitated along with a calcium salt by the addition of calcium before precipitation. An alternative procedure to isotope dilution, reverse or inverse isotope dilution, can be used to determine the quantity of radioactive compound by dilution with an inactive compound. This procedure is applicable when a

system contains an unknown amount of isotopically labeled substance of known specific activity.

Radioisotope exchange

For understanding mechanisms and kinetics of organic or biological reactions, nonradioactive atoms in molecules or ions are allowed to exchange with appropriate radio-labeled compound. After chemical exchange between a labeled compound and the test sample (the chemical form of the element being different in the two solutions, e.g., iodine in CH_3I vs $^{131}\text{I}^-$ in labeled sodium iodide), the specific activity of the element becomes the same in the sample and reagent. A measured decrease in the activity of the reagent or increase in the activity of the sample can then be related to the amount of element present in the sample. Isotopic exchange methods of analysis are very sensitive, rapid, and specific. Isotopic exchange plays a very important role in pharmacokinetic studies. For example, a drug molecule may be labeled with tritium to follow the drug distribution. Exchange of tritium with an unlabeled compound or with a water molecule in the vicinity may lead to erroneous interpretation of the drug distribution if suitable precautions are not taken to avoid potential isotopic exchange reactions.

Radiotracers as radiopharmaceuticals:

Methods for radiopharmaceutical analysis

When radiolabeled compounds (radiotracers) are used for diagnostic and therapeutic purposes, it is called a radiopharmaceutical. A radiopharmaceutical should be easily produced, inexpensive, readily available, have relatively short halflife, and preferably should be a γ -Emitter with an energy between 30 and 300 keV. Such a γ -Emitting nuclide should invariably decay by electron capture or isomeric transition. When positron-emitting nuclides are used, they must have specific localization in the desired organ or tissue and should not result in undue radiation exposure. Radiotracers, used as pharmaceuticals, are produced in one of two ways: either as a product of nuclear fission reactions or as a product of nuclear reactions induced by high-energy accelerator particles or neutrons. Particle accelerators are instruments that cause nuclear reactions by bombarding target nuclides with highly accelerated and energized particles such as protons, deuterons, and electrons. For production of radiopharmaceuticals, on-site accelerators called cyclotrons are used. For production of positron-emitting radionuclides linear accelerators are used. Readily transportable instruments, which serve as sources for short-lived radionuclides, are called generators. Typically, in a generator, a long-lived parent nuclide is allowed to decay to a short-lived daughter

radionuclide. Using differences in chemical properties, the daughter nuclide is separated from the parent. Typical pharmaceutically relevant radionuclides and their sources are summarized in Table 5.

A radiopharmaceutical is a radioactive chemical used as a pharmaceutical. Therefore, as pharmaceuticals, they should have proper ionic strength, pH, isotonicity, and osmolarity. In addition to chemical purity, radionuclide purity and radiochemical purity also have to be demonstrated. The radionuclide purity refers to the ratio of the amount of radioactivity corresponding to that of desired radionuclide to the total amount of radioactivity owing to other isotopes and other isotopic impurities. Radionuclide purity is determined by measuring half-lives and other characteristics of radiation of the radionuclide. Radiochemical purity refers to the fraction of total radioactivity in the desired form. Radiochemical impurities are general chemical impurities formed by the chemical as a result of decomposition of the chemical entity, whereas all the resulting impurities may still be radioactive. Potential decomposition pathways are the same as for any other pharmaceutical. Examples include acid and base hydrolysis products, oxidized and reduced species, additional radiolysis products, and photochemical and thermal decomposition products of the chemical. For example in many ^{99m}Tc -labeled complexes, free unreacted $^{99m}\text{TcO}_4^-$ and hydrolyzed ^{99m}Tc are radiochemical impurities. A number of analytical methods can be used to detect and determine radiochemical impurities. For example, when separating the impurities from radiochemical compound of interest using HPLC, if a radiochemical detector is used, all the radioactive impurities that are separated can be quantitated, and radiochemical purity can be ascertained. These include but are not limited to distillation, precipitation, paper, thin layer, gel chromatography, HPLC, ion exchange, solvent extraction, and other separation and purification techniques. The techniques and principles of these

techniques are the same as for any other pharmaceutical. Thus, these techniques are not addressed here.

Activation Analysis

Activation analysis is a process in which a target trace element in a sample matrix is irradiated with particles in a nuclear reactor. As a result, an activated radionuclide is formed. The characteristic particles or γ -Rays emitted are used for qualitative identification and, more often, for quantitative measurement. The most common activation analysis is neutron activation analysis (NAA). In this technique, a sample containing the element is irradiated with neutrons in a reactor. After irradiation, γ -Emissions ensue from the decaying radionuclide. These are quantitated by using appropriate semiconductor radiation detectors. Detecting γ -Rays of a specific energy identifies the radionuclide. These particular energy values correspond to unique energies characteristic of the decaying radionuclides. For example, when ^{24}Na decays to ^{24}Mg , the γ -Rays released have unique energies of 1.268 and 2.754 MeV. A plot of γ -Ray counts versus energy yields a γ -Ray spectrum; the area under the curve is proportional to the radioactivity of the sample. The rate at which γ -Rays are emitted is proportional to the concentration of the radionuclide. When a large number of elements in a single sample matrix are analyzed using appropriate instruments for quantitation, without separation of the elements, it is called instrumental neutron activation analysis. When there is spectral interference in the sample matrix and if chemical separation is carried out to isolate the element, it is called radiochemical neutron activation analysis. When other charged particles are generated for analysis by activation, it is called charged particle activation analysis (CPAA). CPAA is applied to the determination of elemental concentration in surface layers.

All methods of activation analysis are very accurate and sensitive, and a precision of approximately 2% RSD is easily attainable. Detection limits in parts per billion or lower, depending on the element and sample matrix, are

Table 5 Examples of radionuclides and their production mode

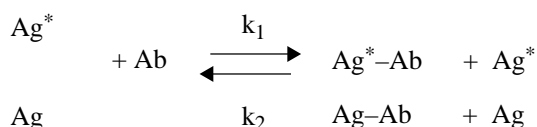
Production mode →	Reactor	Accelerator	Generator
Radionuclides Produced → (some examples)	^{99}Mo , ^{131}I , ^{137}Cs , ^{133}Xe , ^{32}P , ^{51}Cr , ^{125}I , ^{135m}Ba , ^{195m}Pt .	^{123}I , ^{57}Co , ^{67}Ga , ^{201}Tl , ^{111}In , ^{127}Xe , ^{81}Rb . The following positron generating nuclides are produced on site: ^{11}C , ^{13}N , ^{15}O , ^{18}F ;	^{99}Mo to ^{99m}Tc , to $^{99}\text{Tc}^a$, ^{113}Sn to ^{113m}In , to ^{113}In , ^{87}Y to ^{87m}Sr , ^{81}Rb to ^{81m}Kr , to ^{81}Kr , ^{195m}Hg to ^{195m}Au , to ^{195}Au ; other elements used for generator reactions include ^{68}Ge , ^{62}Zn , ^{137}CS , and ^{82}Sr .

^aDecay as indicated below: parent to daughter to granddaughter.

easily attainable. As many as 60 different elements that can form radionuclide can be analyzed using NAA. NAA finds wide application in a number of other fields, and these are summarized in Table 6.

Radioimmunoassay

Radioimmunoassay (RIA) is a technique based on the formation of antigen–antibody complex. This technique essentially involves the application of isotope dilution analysis. An antigen is typically a protein of molecular weight greater than 10,000 that stimulates the production of antibody in an animal body. The antigen subsequently binds with the antibody. Antigen is usually measured in the patient's sample, and the antigen becomes the analyte. To an antibody, a mixture of labeled and unlabeled antigen is added in excess such that the quantity of antibody needed to bind is allowed to be insufficient. As a result, both types of antigen compete with the limited amount of antibody in the sample. The reaction in an RIA mixture can be described as follows.



k_1 and k_2 are rate constants; equilibrium constant $K = k_1/k_2$.

To a constant amount of labeled antigen and antibody, increasing amounts of unlabeled antibody are added. The initial amount added is still in excess of the antibody needed for binding. As a result of competing reactions of the labeled and unlabeled antigen, the greater the concentration of the unlabeled antigen added, the less is the amount of bound labeled complex (Ag^*-Ab complex) and hence greater is the free (unbound) antigen. After

incubation to equilibrium at a specified temperature and time, unique to the system, separation of the free labeled Ag^* , the fraction of the bound labeled antigen is determined by measuring the activity of the radioactive nuclide. By plotting the percent of bound labeled antigen versus the concentration of antigen added, the concentration of the unknown antigen can be determined. Several radionuclides such as ^{14}C , ^3H , ^{131}I , ^{32}P , ^{75}Se , ^{59}Fe , ^{99}Mo , and ^{57}Co have been used for RIA. However, ^{125}I is the most commonly used radionuclide for RIA. The earliest method of separation included electrophoresis and chromatography. However, today RIA kits are marketed by commercial manufacturers with detailed description of the principles of the method, methods of use, sensitivity, precision, and limitations of use for specific uses. The RIA methods are very rapid, sensitive, specific, and inexpensive, especially for large biological samples in complex sample matrices.

The RIA technique is applied in assays of hormones, steroids, peptides, aminoglycosides such as tobramycin and gentamycin, insulin, many immunoglobulins, different types of viral hepatitis, plasma catecholamines, angiotension-converting enzymes, many vitamins including vitamin B_{12} , human growth hormones, many folate derivatives, and others. Many commercially available RIA kits, unique to each kit, contain series of standards with known concentrations of unlabeled antigen, a vial of suitable labeled antigen, a vial of antibody solution, and appropriate precipitants or other analytical aides.

RADIATION SAFETY

When radiation energy is absorbed by tissues, depending on the dose received, rupture of chemical bonds occurs that causes damage to cells and tissues. Such damage may not be clinically apparent even for years if the absorbed

Table 6 Application of neutron activation analysis

Field of application	Typical application
Geology	Characterization of elemental deposition or formation
Environmental science	Atmosphere, soil, and water pollution
Industrial engineering	Corrosional and frictional effects
Agriculture	Nutrient deficiency
Chemical studies	Contamination, characterization of reference standards, surface effects
Forensic science	Sample identification, correlation of samples to crime scenes

dose is small. Radiation damage is measured in terms of absorbed dose. Dose is defined as the energy imparted to a material per unit mass. The SI unit of dose is called Gray, or Gy (joules/Kg). The traditional unit is rad. One Gy = 100 rads. However, the effectiveness of all radiations is not the same. The effectiveness is dependent on the nature of the particle, the thickness of the tissue or organ, and the characteristics of the material or organ to which such a dose is administered. Thus, to account for such differences in effectiveness, another term, called dose equivalent, is used. The SI unit of dose equivalent is Sievert, which is numerically equivalent to Gy multiplied by the appropriate weighting factor corresponding to biological tissue or organ. For example, the weighting factor (W_R) for many tissues is 0.3, whereas it is 0.03 for bone surfaces. The radiation weighting factor is also dependent on the energy of the particles. For example, for neutrons with less than 10 keV, $W_R = 5$, whereas it is equal to 20 for neutrons greater than 100 keV.

For purposes of administered dose of a radiopharmaceutical, the package insert contains a table that lists dosage for the "average" patient as a function of administered activity. This information may be used for calculation of radiation dose. If needed, a medical physicist will calculate administered dose based on Medical Internal Radiation Dose (MIRD) committee recommendations.

External exposure to ionizing radiation as a result of occupation is measured using dosimeters. Three different types of dose monitors are commercially available. They are pocket dosimeters, film badges, and thermoluminescent detectors (TLDs). Pocket dosimeters are GM counting-based digital dosimeters, which provide immediate reading. Film badges are the least expensive and most popular. Using appropriate filters in the film holder, accurate readings to exposure to different particles can be measured. The major disadvantage with film badges is the waiting period required for developing and processing films. In TLDs, the radiation received is stored in holders containing crystals of LiF or manganese-activated calcium fluoride. Subsequently, during measurement, the crystals are heated to required temperatures such that they emit light. When measured, the emitted light provides a measure of absorbed dose.

REGULATORY REQUIREMENTS

The regulations of various advisory groups and government organizations have to be met in handling, use, and disposal of radioactive material and radiopharmaceuticals. These include the FDA, NRC, Department of Transpor-

tation (DOT), Environmental Protection Agency (EPA), and Occupational Safety and Health Administration (OSHA). Of these, the FDA regulates safety, stability, efficacy, and toxicity of the radiopharmaceuticals; NRC regulates all reactor-produced by-products regarding use, handling, disposal, and radiation safety and protection of workers and the public using the facilities. Title 10 of the Code of Federal Regulations, 10 CFR 20, contains the regulations for radiation protection, whereas 10 CFR 35 addresses medical uses of radioactive materials. Title 49, 49 CFR, addresses packaging and transportation of radioactive materials. In addition, all USP requirements have to be met for a radiopharmaceutical. The International Committee on Radiation Protection (ICRP) and the National Committee on Radiation Protection (NCRP) provide radiation dose recommendations for adoption to other regulatory agencies. The NRC requires that licensees follow the ALARA philosophy regarding radiation exposure. ALARA is an acronym for as low as reasonably achievable. Under the ALARA concept, when radiation exposure of a worker exceeds 10% of the allowed occupational limit, an investigation is required. If it exceeds 30%, investigation and corrective action are necessary per NRC requirements.

As with any other pharmaceutical, radiotracers used as pharmaceuticals are subject to all regulatory requirements. The radiopharmaceutical manufacturer must, at a minimum, satisfy the requirements of the NRC and the FDA.

BIBLIOGRAPHY

1. Alfassi, Z.B. *Chemical Analysis by Nuclear Methods*; John Wiley & Sons: Chichester, UK, 1994.
2. Anderson, C.J.; Welch, M.J. *Chem. Revs.* **1999**, 99, 2219.
3. Bernier, D.R.; Christian, P.E.; Lanagan, J.K. *Nuclear Medicine Technology and Techniques*, 4th Ed.; Mosby: St. Louis, MO, 1997.
4. Bhargava, K.; Du, J.; Raghavan, R.; Joseph, J. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J.S., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1996; 13, 255–281, and References Contained Therein.
5. Britton, K.E.; Granowska, M. *Nuclear Medicine in Clinical Diagnostics and Treatment*; Murray, P.C., Ell, P.J., Eds.; Churchill Livingstone: New York, NY, 1994; 2, 779–800.
6. Chatter, J.F.; Mahe, M. *Nuclear Medicine in Clinical Diagnostics and Treatment*; Murray, P.C., Ell, P.J., Eds.; Churchill Livingstone: New York, NY, 1994; 2, 865–876.
7. Collier, B.D.; Krasnow, A.Z.; Hellmann, R.S. *Skeletal Nuclear Medicine*; Collier, B.D., Fogelmann, I., Rosenthal, L., Eds.; Mosby: St. Louis, MO, 1996.
8. Early, P.J.; Sodee, D.B. *Principles and Practice of Nuclear Medicine*; Mosby: St. Louis, MO, 1995.

9. Ehmann, W.D.; Vance, D.E. *Radiochemistry and Nuclear Methods of Analysis, Vol. 116: Chemical Analysis*; John Wiley & Sons: New York, NY, 1991.
10. Helmer, R.G. *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th Ed.; Kroschwitz, J.I., Howe-Grant, M., Eds.; John Wiley & Sons Inc.: New York, NY, 1996; 20, 871–906.
11. Kist, A.A. J. Anal. Chem. **1996**, *51*, 80, (Transl. of Zh. Anal. Khim).
12. Kolesov, G.M. J. Anal. Chem. **1996**, *51*, 71, (Transl. of Zh. Anal. Khim).
13. Lazewatsky, J.L.; Crane, P.D.; Scott, Edwards, D. *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th Ed.; Kroschwitz, J.I., Howe-Grant, M., Eds.; John Wiley & Sons: New York NY, 1996; 20, 930–962.
14. Marek, M.J.; Lambert, C.R.; Rhodes, B.A. *Techniques for Direct Radiolabeling of Monoclonal Antibodies*; RhoMed Incorporated: Albuquerque, NM, 1994.
15. Mettler, F.A.; Guiberteau, M.J. *Essentials of Nuclear Medicine*, 3rd Ed.; W.B. Saunders Co.: Philadelphia, PA, 1991.
16. Muller, S. *Nuclear Medicine in Clinical Diagnostics and Treatment*; Murray, P.C., Ell, P.J., Eds.; Churchill Livingstone: New York, NY, 1994; 1, 195–212.
17. Perkins, A.C. *Nuclear Medicine in Pharmaceutical Research*; Perkins, A.C., Frier Eds.; Taylor and Francis: Philadelphia, PA, 1999; 15–30.
18. Pimm, M.V. *Nuclear Medicine in Pharmaceutical Research*; Perkins and Frier Ed.; Taylor and Francis: Philadelphia, PA, 1999; 133–162.
19. Powsner, E.R.; Widman, J.C. *Tietz Textbook of Clinical Chemistry*, 3rd Ed.; Burtis, C.A., Ashwood, E.R., Eds.; W.B. Saunders Co.: Philadelphia, PA, 1999; 113–132.
20. Powsner, R.A.; Powsner, E.R. *Essentials of Nuclear Medicine Physics*; Blackwell Science Inc.: Malden, MA, 1998.
21. Saha, G.B. *Fundamentals of Nuclear Pharmacy*, 3rd Ed.; Springer-Verlag: New York, NY, 1992.
22. Skoog, D.A.; Holler, F.J.; Nieman, T.A. *Principles of Instrumental Analysis*, 5th Ed.; Harcourt Grace College Publishing: Chicago, 1988; 810–828.
23. Stroebel, H.A.; Heinemann, W.R. *Instrumental Analysis: A Systematic Approach*; John Wiley & Sons Inc.: New York, 1989.
24. Spencer, R.P. *Nuclear Medicine in Clinical Diagnostics and Treatment*; Murray, P.C., Ell, P.J. Eds.; Churchill Livingstone: New York, NY, 1994; 2, 647–665.
25. Taillefer, R.; Tamaki, N. *New Radiotracers in Cardiac Imaging, Principles and Applications*; Appleton and Lance: Stamford, CT, 1999.
26. *United States Pharmacopial Convention, United States Pharmacopeia*; 24th Revision: New York, NY, 2000, 1981–1990; USP Convention.
27. Volkart, W.A.; Hoffmann, T.J. Chem. Rev. **1999**, *99*, 2269.
28. Weber, D.A.; Eckermann, K.F.; Dillman, L.T.; Rayman, J.C. *MIRD Radionuclide Data and Decay Schemes*; New York Society of Nuclear Medicine, 1989.
29. Yalow, R.S. J. Chem. Ed. **1999**, *76*, 767.
30. Ekin, R.P. J. Chem. Ed. **1999**, *76*, 769.
31. Ullmann, E.F. J. Chem. Ed. **1999**, *76*, 781.
32. <http://www.psich./www.lrp.hn/lrp>.
33. <http://www.triumf.ca/safety/rpt/html>.
34. <http://www.brighamrad.harvard.edu/education/online/BrainSPECT/>.
35. <http://www.research.atlantech.fr/Topic>.
36. <http://www.fz-rossendorf.de/FWB/fwbc/handzet/>.
37. <http://www.anl.gov/labDB/current/Ext/>.
38. <http://www.catcmb.cua.edu/receptor.htm>.
39. <http://www.tardis.jyu.fi/ECRIS/rip>.
40. http://www.fkpi.com/Release_Comp/Releases/International/.
41. <http://www.radio-isotopes.com/nucmed.htm>.

RADIOLABELING OF PHARMACEUTICAL AEROSOLS AND γ -SCINTIGRAPHIC IMAGING FOR LUNG DEPOSITION

Hak-Kim Chan

University of Sydney, Sydney, New South Wales, Australia

OVERVIEW

Inhalation aerosols have been used successfully to deliver drugs to the lung for local and systemic therapeutic effects. In vivo evaluation of pharmaceutical inhalation products is achieved by γ -Scintigraphic imaging of the aerosol deposited in the lung. Imaging provides direct information on the amount and location of the drug deposited in the lung after inhalation (1). This local bioavailability, rather than the systemic bioavailability after absorption, is pertinent to drugs that act directly on the lung. For drugs that act systemically, deposition site will affect the rate and extent of absorption in the lung. As a result, lung deposition using γ -Scintigraphy has been proposed for bioequivalence studies of aerosol products (2). The deposition data can be linked further to the clinical response and in vitro particle size distribution, adding a new dimension to the inter-relationship between them.

To measure lung deposition by imaging, the aerosol must be first labeled or tagged with a suitable radionuclide. Radiolabeling techniques have been developed for current inhalation products including nebulizers, propellant-driven metered-dose inhalers (MDIs), and dry powder inhalers (DPIs).

Lung imaging is achieved using a γ -Camera. The camera creates an image of the γ -Rays emitted by the radionuclide in the lung (3). In the past, numerous lung deposition studies on radiolabeled aerosol products have been carried out using planar imaging by which only the anterior or posterior 2-Dimensional view of the lung is collected. However, because the lung is a 3-Dimensional object, spatial distribution of the aerosol in the lung can be best obtained using tomographic (rather than planar) imaging such as single photon emission computed tomography (SPECT). SPECT is a technique for producing cross-sectional images of radionuclide distribution in the body. This is achieved by imaging the lung at different angles (e.g., 64 or 128 images/180° or 360°) around the thorax using a rotating γ -Camera, followed by computational image reconstruction.

BACKGROUND OF RADIOLABELING PHARMACEUTICAL AEROSOLS

It is a prerequisite that the pharmaceutical aerosols can be suitably radiolabeled before any lung scintigraphic imaging can begin. Ideally, the drug can be directly radiolabeled, i.e., chemically by substitution of an atom in the drug molecule with a radioactive isotope. This can be achieved using positron emitters, e.g., C-11, N-13, and O-15 atoms. Positrons are positively charged electrons that when combined with an electron, produce two γ -Rays with equal energy (511 keV) emitted at 180° to each other. Theoretically, because C, N, and O atoms are present in all organic molecules, they can be used to label virtually any drug. In reality, the use of these atoms is limited by their short half-lives (e.g., 20, 10, and 2 min for C-11, N-13, and O-15 atoms, respectively), relative to the time taken for manipulating the drug (including organic synthesis for radiolabeling and successive processing for product characteristic assurance such as particle size distribution of the aerosol particles). Further, to produce these positron emitters, it is necessary to have a nearby cyclotron facility, which means extra cost. So far, only three antiasthmatic compounds have been successfully radiolabeled by positron emitters: ipratropium bromide (Br-77) (4), triamcinolone acetonide (C-11), and fluticasone propionate (F-18) (5). This is indicative of the difficulties for direct radiolabeling.

Because of the forementioned limitation, γ -Emitters have been used to radiolabel the drug indirectly for γ -Scintigraphy. In this case, the radiolabel associates with the drug by physical means instead of by chemically incorporating into the drug molecule via covalent bonds. Thus, instead of being a direct chemical approach, it is an indirect radiolabeling. Technetium-99m is the most commonly used pure γ -Emitter for indirect radiolabeling of pharmaceutical aerosols. The γ -Ray of ^{99m}Tc has sufficient energy (140 keV) to penetrate body tissues without significant absorption or scattering, but when it reaches the detector of the γ -Camera, it will be absorbed and converted into light photons, thus optimal for

γ -Camera imaging. The half-life of ^{99m}Tc is 6 h, which is long enough for handling and imaging but not too long as to increase the radiation dose to the subject unnecessarily. As a result, ^{99m}Tc is used for the majority of nuclear medicine imaging studies. Once inhaled into the lung, ^{99m}Tc can have a much shorter biological half-life depending on the physical form, e.g., as diethylenetriamine penta-acetic acid (DTPA) or DTPA complex, it will be absorbed rapidly from the lung into systemic circulation, followed by glomerular filtration in the kidney to the bladder where it is excreted in the urine. The whole process could take less than 2 h if the subject drinks plenty of water.

Regardless of the type of aerosol products to be radiolabeled, a fundamental requirement in the radiolabeling of aerosol products is that the radiolabel must associate with the drug such that not only the radiolabel distribution matches the drug distribution but also that the radiolabel distribution matches that of the unlabeled commercial product.

RADIOLABELING OF NEBULIZER SOLUTIONS

Radiolabeling of nebulizer solutions is by far the simplest among the aerosol products and is achieved by simply mixing the radionuclide with the drug solution. Because the radionuclide and the drug are uniformly distributed in the solution and provided that there is no precipitation of the ingredients occurred, each nebulized aerosol droplet would contain both radioactivity and drug in proportion to the droplet size. ^{99m}Tc in complexing with DTPA or human serum albumin is widely used as the radionuclide. Sodium pertechnetate is not suitable because the free anion, as with iodide, has a high affinity for the thyroid. Lung deposition of nebulized salines and drug solutions, including nedocromil sodium (6), salbutamol, fenoterol, ipratropium bromide, carbenicillin, pentamidine isethionate (7), flunisolide (8), and liposomes containing beclomethasone dipropionate (9), has been studied using this radiolabeling technique of mixing the radionuclide with the drug solution.

RADIOLABELING OF PROPELLANT-DRIVEN MDIs

Historically, there are three major approaches of radiolabeling suspension MDIs, developed by Few et al. (10), Newman et al. (11), and Kohler et al. (12). The first two methods were initially developed for polymeric

Scheme 1 MDI radiolabeling (10)

Elution of ^{99m}Tc as sodium pertechnetate from a ^{99}Mo - ^{99m}Tc generator
 Adjustment of pH to 7–9 with concentrated NH_4OH
 Addition of a drop of 5% tetraphenylarsonium (TPA) chloride aqueous solution
 Extraction of ^{99m}Tc -TPA into chloroform^a
 Separation by passing through phase-separation paper^b
 Collection of the filtrate (the CHCl_3 phase containing the ^{99m}Tc -TPA)
 Evaporation of the CHCl_3 to dryness
 Reconstitution in the appropriate medium containing the compound for radiolabeling

^aFor example, by mechanical shaking or sonication the solution with CHCl_3 .

^bFor example, Whatman 1^P silicone-treated phase-separation filter paper.

particles but were later modified for the radiolabeling of drugs.

Early in 1970, Few et al. (10) radiolabeled polystyrene particles for a mucociliary clearance study. The radiolabeled aerosols were produced by a spinning disk generator. The radiolabeling technique involves the key steps of extracting sodium pertechnetate ($\text{Na}^{99m}\text{TcO}_4$) into chloroform as tetraphenylarsonium pertechnetate followed by evaporation of the chloroform. A solution of polystyrene is then added to the radioactive residue and dispersed (Scheme 1). This technique has subsequently been adopted by other workers for radiolabeling pharmaceutical MDIs of sodium cromoglycate (13, 14). It is important to note that in this technique, the complexing agent tetraphenylarsonium chloride is classified as poisonous (15). Although the actual amount of the compound inhaled is in the nanogram range, safety to the researchers and the subjects inhaling the aerosols has to be carefully ensured.

Another way to radiolabel the pharmaceutical MDI is to use Teflon particles with a size distribution similar to that of the drug of interest (Scheme 2). This was carried out in 1981 by Newman et al. (11) who used Teflon particles of mean size $2 \pm 0.4 \mu\text{m}$ to mimic the MDI aerosols of bronchodilators. Thus, the Teflon particles were used as a surrogate for the drug. The Teflon particles were almost monodispersed and were produced using a spinning disk aerosol generator. However, this approach is limited by the physicochemical characteristics of the Teflon particles being different from those of the drug particles. Also, the aerosol particle size distribution of the Teflon may not match that of the drug. By co-administering a physical mixture of the radiolabeled Teflon particles with the drug

Scheme 2 MDI radiolabeling (11)

Elution of ^{99m}Tc as sodium pertechnetate from a ^{99}Mo - ^{99m}Tc generator
 Removal of sodium by passing the eluate through a cation exchange column
 Collection of the filtrate
 Evaporation of the filtrate to dryness
 Addition of Teflon particles suspended in 40% alcohol
 Generation of radiolabeled Teflon particles by spinning disc technique^a
 Collection of radiolabeled Teflon particles
 Transfer of the radiolabeled particles to an empty canister
 Addition of the content from a commercial MDI^b into the canister^c containing the radiolabeled particles, followed by recrimping of the MDI
 Mixing by sonication

^aFor example, a spinning speed of 62,000 RPM had been used.

^bFor example, the MDI is chilled at -60°C before the transfer is carried out.

^cFor example, the canister is chilled at -60°C before the transfer is carried out.

salbutamol, the lung deposition and clinical response have been monitored simultaneously (16, 17).

The first attempt to radiolabel drug particles (instead of polymers such as polystyrene or Teflon particles) for pharmaceutical aerosols was carried out on fenoterol and salbutamol by Kohler et al. (12) (Scheme 3). However, it was later found that the method would change the particle

Scheme 3 MDI radiolabeling

Elution of ^{99m}Tc as sodium pertechnetate from a ^{99}Mo - ^{99m}Tc generator
 Extraction of ^{99m}Tc -TPA into methyl ethyl ketone (MEK).^a
 Repeat the extraction if necessary.
 Separation of the aqueous and MEK phases^b
 Collection of the MEK phase (containing the pertechnetate)
 Evaporation of the MEK to dryness in a glass beaker
 Addition of propellant^c and surfactant^d to the dry pertechnetate
 Concentration by evaporating the propellant to 0.2 mL
 Transfer of the propellant containing the pertechnetate and surfactant into an MDI canister to be radiolabelled, followed by recrimping of the MDI
 Mixing by shaking

^aFor example, by shaking the pertechnetate solution with approximately equal volume of MEK.

^bFor example, in a separating funnel.

^cFor example, Freon 11, 5–10 ml had been used.

^dFor example, 1% sorbitan trioleate had been used.

(From Ref. 12.)

Scheme 4 MDI radiolabeling

First 4 steps are the same as in Scheme 3.

Evaporation of the MEK (containing the pertechnetate) to dryness in an empty canister

Transfer of the content from a commercial MDI into the canister containing the pertechnetate for radiolabeling, followed by recrimping of the MDI

Mixing by sonication

This method is simpler than Scheme 3 as it does not involve the propellant concentration step and the subsequent transfer of the concentrate. (From Ref. 18)

size distribution of the labeled aerosol, resulting in a coarser aerosol than the unlabeled product. The method, after subsequent improvement by Summers et al. (18) (Scheme 4), has become widely used for the radiolabeling of MDIs. This method is preferred over other methods because it does not involve extraction with tetraphenylarsonium chloride and chloroform.

It is worth noting that each of the radiolabeling examples noted previously can be further modified for the study need. For example, the drug particles can be suspended in the organic phase containing the radiolabel and spray-dried, followed by reconstitution in the propellants, as carried out on salbutamol sulfate (19).

Although the radiolabeling methods have been widely used, the mechanism of association between the radiolabel and the drug particles has been studied only recently. The study by Farr (20) on the MDI systems indicated that in the CFC formulation, the radiolabel $^{99m}\text{TcO}_4^-$, being hydrophilic, would associate with hydrophilic domains including the surface of hydrophilic drug particles and the interior of surfactant reverse micelles. There is a need to extend the study to other systems on hydrophobic drugs and on non-CFC propellants. See Table 1 for methods of radiolabeling MDIs.

RADIOLABELING OF DPIs

A method of wide application to radiolabeling of dry powders is by adsorbing the radiolabel on the particles in a suitable liquid (Scheme 5). This is achieved by wetting the drug particles with a nonsolvent containing the radiolabel, followed by evaporation of the solvent, leaving the radiolabel on the surface of the drug particles. The method has been applied to radiolabel terbutaline sulfate and budesonide (32, 33). Factors affecting the radiolabeling of dry powder formulations include the physicochemical nature of the drug, choice of nonsolvent

Table 1 Examples of radiolabeling pharmaceutical MDIs

Drug MDI	Study objective	Reference
Salbutamol sulfate	Development of radiolabeling method ^a	19
Salbutamol	Radiolabeling method development; simultaneous measurement of lung deposition and bronchodilator response ^b	16
Salbutamol	Comparison of lung deposition and bronchodilator response among MDI, DPI and nebulizer ^b	17
Sodium cromoglycate	Development of radiolabeling method ^c	30
Sodium cromoglycate	Comparison of MDI with DPI	31
Sodium cromoglycate	Effect of spacer	13
Sodium cromoglycate	Effect of higher dose (5 mg/puff); effect of spacer	14
Nedocromil sodium	Development of radiolabeling method	18
Nedocromil sodium	Ventilator suitability	21
Terbutaline sulfate, salbutamol	Development of radiolabeling method	22
Nacystelyn	Lung deposition	23
Salbutamol and fenoterol	Development of radiolabeling method and lung deposition	24
Flunisolide and fenoterol	Lung deposition and effect of a spacer	25
Salbutamol	Development of radiolabeling method and comparison of MDI with DPI	26
Beclomethasone dipropionate in hydro fluorocarbons	Correlation of in vitro particle sizing with in vivo deposition	27
Fenoterol	Comparison of nebulizer with MDI plus a holding chamber	28
Salbutamol	Aerosol delivery by microprocessor control (SmartMist TM)	29

^aDrug particles were suspended in the chloroform phase and spray-dried.

^bRadiolabeled Teflon was used (the drug itself was not radiolabeled).

^cSpray-drying of the drug was followed by reconstitution in an MDI.

Scheme 5 DPI radiolabeling

First 4 steps are the same as in Scheme 3
 Evaporation of the MEK (containing the pertechnetate) to dryness
 Redissolution of the pertechnetate in a suitable liquid^a
 Addition of the pertechnetate solution to the drug powder to be radiolabeled^b
 Evaporation of the liquid^c
 Filling of the radiolabelled powder to the DPI^d

^aThe liquid must be a nonsolvent for the drug powder to be radiolabeled; some suitable ones include water for budesonide, chlorofluorocarbon 11 for salbutamol

^bMixing, e.g., by sonication for 20 min, if required

^cFreeze drying has been used for water removal.

^dBlend with lactose carrier, if required.

for the drug, solubility of radiolabel in the nonsolvent, moisture level, electrostatic charge, and the number of processing steps (30). Unfortunately, the details still generally remain as proprietary information and are not available in the literature. For example, exactly how the spherical agglomerates of budesonide powder are to be wetted with the ^{99m}Tc solution to ensure reproducible radiolabeling has not been reported.

Alternatively, a method resembling Scheme 1 can be used, provided the drug to be radiolabeled will not dissolve in CHCl₃. This method involves the same first six steps as in Scheme 1. After the filtrate (the CHCl₃ phase containing the ^{99m}Tc-TPA) is collected, it is added to the drug powder. The CHCl₃ is then evaporated, e.g., at 70°C, leaving the radiolabel with the powder. The powder is ready for filling into the DPI. Bronchodilators (salbutamol, terbutaline sulfate) and prophylactics (nedocromil sodium)

Scheme 6 DPI radiolabeling

- Preparation of solution of the drug to be radiolabeled^a
- Addition of sodium pertechnetate to the drug solution^b
- Spray drying of the solution^c
- Collection of dry powder
- Filling of the radiolabelled powder to the DPI^d

^aFor example, 6% w/w for sodium cromoglycate
^bFor example, 1 ml of pertechnetate in normal saline
^cFor example, Buchi Minispray dryer, model 190, liquid feedrate 60 ml/min, inlet air temp 180°C, air throughput 2.4 m³/min and nozzle air pressure 800 kPa.
^dFor example, Blend with lactose carrier, if required

for asthma have been successfully labeled using this technique for lung deposition studies (31, 34). There was also an earlier method for the radiolabeling of dry powders (primarily sodium cromoglycate) by spray-drying (35, 36). The basic principle can be considered the same as that in Scheme 1 for MDI in that radiolabeled particles were produced by evaporation of radiolabel-containing atomized droplets. The method is very straightforward (Scheme 6) but suffers the limitation that spray-dried particles may not be physicochemically the same as those in the commercial products. This is because milling rather than spray-drying is normally used for micronization of the drug particles. Spray-drying, following ^{99m}Tc adsorption to the surface of the particles, has also been used to prepare radiolabeled cromoglycic acid and nedrocromil powders (37). The exact mechanism of association between the radiolabel and the drug particles is unknown and is generally regarded as a surface-coating phenomenon. However, the surface is proportional to the square of the particle size, whereas the volume (drug mass) the cube of the particle size. It follows that to have a match between

the radiolabel and drug mass in the aerosol, the radiolabeled particles must exist as agglomerates rather than as single particles. See Table 2 for examples of radiolabeling DPIs.

γ-SCINTIGRAPHIC IMAGING FOR LUNG DEPOSITION

Aerosol deposition in the lung is measured by γ-Scintigraphy routinely used in nuclear medicine. A γ-Camera consists of a collimator located in front of a detector, behind which is an array of photomultiplier tubes (3). The collimator allows only γ-Rays with a defined angle to reach the detector. The detector, which is a scintillation crystal made of sodium iodide, converts the γ-Rays to light photons. The photomultiplier tubes then transform the light photons into an electrical signal that can be displayed in the X- and Y-position on a monitor such as a cathode-ray tube or a computer. Planar imaging, which provides 2-Dimensional anterior and/or posterior views of the lung, has been commonly used in the past to study deposition. This involves inhalation of the radiolabeled aerosols with a collimated γ-Camera placed in front of or behind the chest of the subject. The image is formed as the γ-Ray photons emitted from the radiolabeled aerosol in the lung fall on the camera detector. The primary advantage of planar imaging is that the image acquisition and data processing are less effort-demanding. Also, compared with tomographic imaging (see SPECT, below), it can utilize a lower radioactivity dose. However, a severe limitation of planar imaging is that it compresses the 3-Dimensional lung into a 2-Dimensional view, and thus it cannot provide 3-Dimensional spatial information about the aerosol distribution in the lung. This may become more critical if

Table 2 Examples of radiolabeling pharmaceutical DPIs

Drug DPI	Study objective	Reference
Terbutaline sulfate	Development of radiolabeling method for Turbuhaler in lung deposition study	33
Nedocromil sodium	Lung deposition study and comparison with an MDI	34
Budesonide	Nasal distribution study	32
Budesonide	Lung deposition in children with cystic fibrosis	38
Salbutamol	Lung deposition study using a new DPI	30
Disodium cromoglycate	Lung deposition comparison between a conventional MDI and a new DPI	34
Salbutamol sulfate	Development of radiolabeling method and comparison with an MDI	26
Cromoglycic acid and Nedocromil	Development of radiolabeling method	37

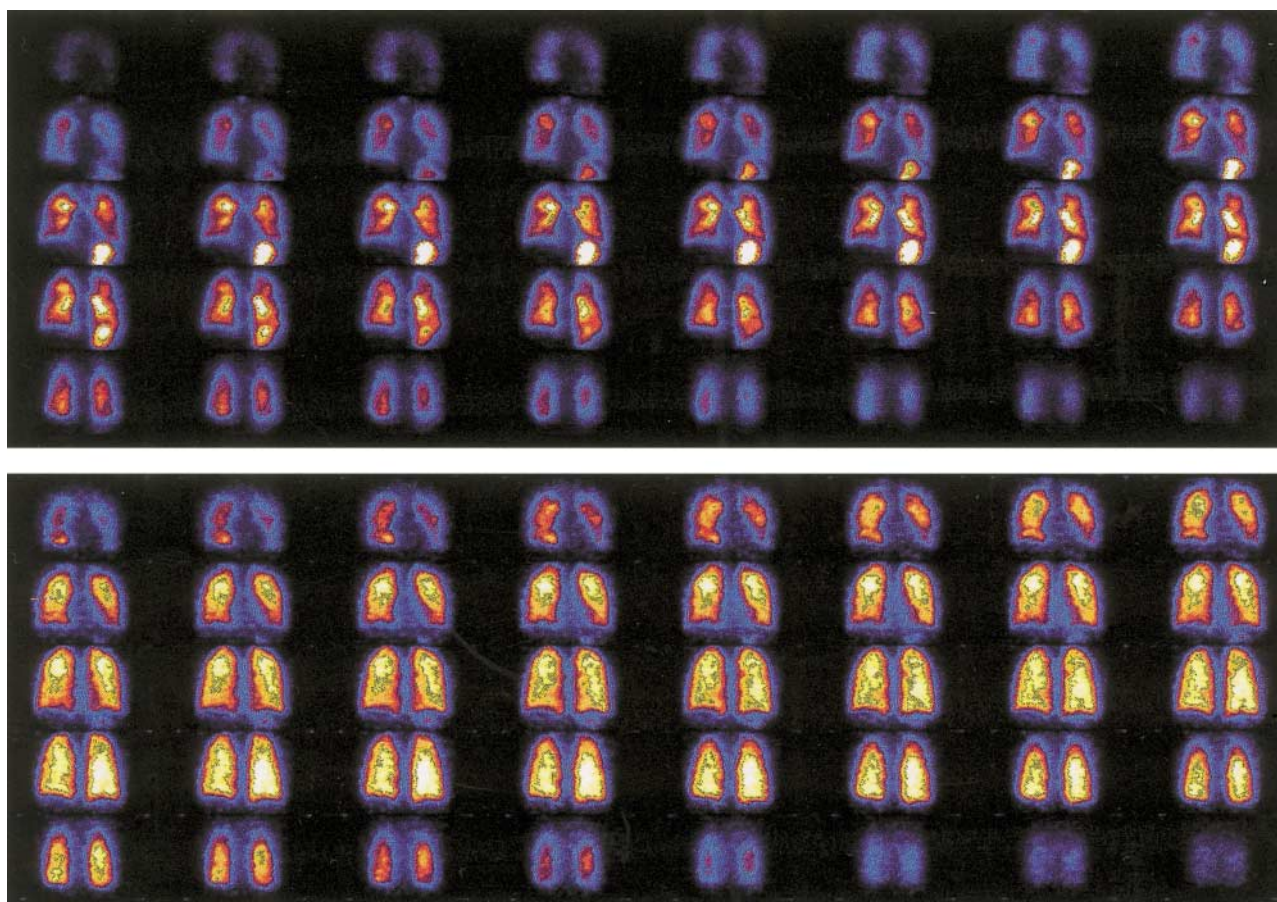


Fig. 1 Fast 1-min SPECT coronal section lung images of a healthy subject after inhalation of normal saline aerosols containing ^{99m}Tc -DTPA in large and small droplets (7 μm , top 3 μm , bottom) generated by air-jet nebulizers. The difference in deposition between the two aerosols is clearly shown in these images.

thelung deposition data are to be used for bioequivalence comparison (1,2).

Three-dimensional image data can be achieved by SPECT. SPECT has been used to measure lung deposition of aerosols in a number of studies (6, 39–44, 49). This involves imaging the lung at different angles using a SPECT camera rotating round the chest of the subject after radioaerosol inhalation. The acquired raw data are then processed by high-speed computers to reconstruct the lung images in the coronal, sagittal, or transverse section (e.g., see Fig. 1 for the coronal section images). The question remains as to whether SPECT offers any advantage over planar γ -Scintigraphy for quantitation of total (as opposed to regional) lung deposition. This can be answered by comparing the measurement of a known amount of radioactivity in a lung phantom (e.g., a perspex container with size and shape similar to the human lung) using both techniques. The results of this comparison are not yet published.

Misconceptions of SPECT

High-dose radiation has often been quoted as a disadvantage of SPECT (1). Obviously, the radioactive dose should never be administered more than required, regardless of whether SPECT or planar imaging is used. It is interesting to note that a number of planar imaging studies were carried out in the past using radiation doses comparable with those for SPECT (43). In practice, the dose consideration must be balanced by the benefits of the deposition details gained using SPECT. If planar imaging cannot provide the required deposition details, then SPECT should be considered. Otherwise, carrying out a suboptimal study using planar imaging without obtaining the deposition data required is not justified because it exposes the subjects to unwanted radiation exposure. The amount of radioactivity delivered to the lung for SPECT is approximately 60 MBq, which is equivalent to an effective whole-body dose of 0.4 mSv. In comparison, the

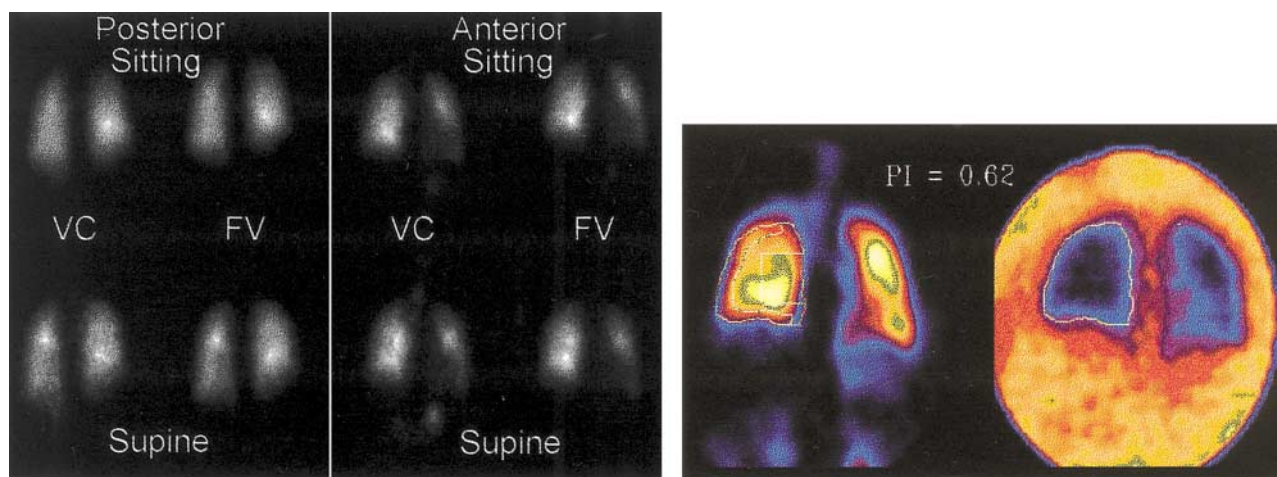


Fig. 2 Lung images of a healthy subject after inhalation of an aqueous aerosol containing ^{99m}Tc -DTPA generated by a novel aerosol inhaler, the AERx system of Aradigm Inc. (Ref. 42). *Left*: planar images from aerosols inhaled using two different breathing maneuvers [vital capacity (VC) or fixed volume (FV) of 1 L above the functional residual capacity] and two different postures as stated. *Right*: a midcoronal image acquired using SPECT. Superposition of the lung outline from the transmission SPECT image (right) on the emission SPECT image confirmed the excellent peripheral deposition of the aerosol.

average radiation dose from background radiation is approximately 2–3 mSv in a year, depending on the geographic location.

Another common misconception about SPECT is the long image-acquisition time and the associated problem of relocation of deposited aerosol particles in the lung (1). Historically, this was true when a single-head SPECT γ -Camera was used in the 1980s for lung deposition because the imaging process took approximately 15–20 min (43). During this length of time, relocation of the radiolabel occurred as confirmed by comparing lung images immediately before and after the SPECT acquisition. Radiolabel movement would create problems of inconsistent projection during image reconstruction, leading to artifacts in the reconstructed images. Despite these limitations, SPECT has been shown to be superior to planar imaging in differentiating lung deposition among aerosols (39). In a more recent study on a novel inhaler using different breathing maneuvers (42), SPECT was able to detect a subtle difference in the aerosol deposition that was not observed using planar imaging (Fig. 2). The problem of long acquisition time has recently been resolved by fast dynamic SPECT (44, 49). Using a triple-head

γ -Camera, the SPECT acquisition time has been reduced to only 1 min (Fig. 1). Thus, 20–30 complete individual 1-min SPECT image acquisitions can be collected in 20–30 min, making dynamic study feasible. The technique can also be applied to study clearance of an aerosol from the lung after deposition (Fig. 3).

Attenuation and Scatter Correction

Attenuation occurs when the γ -Ray photons emitted from the radionuclide in the lung are absorbed by the body tissues before reaching the camera detector. Attenuation causes a reduction in measured radioactivity counts. Because of the heterogeneous nature of the thorax tissues (lung, muscles, bones), the attenuation cannot be assumed to be uniform and must be measured. Experimentally, the attenuation information can be obtained from a

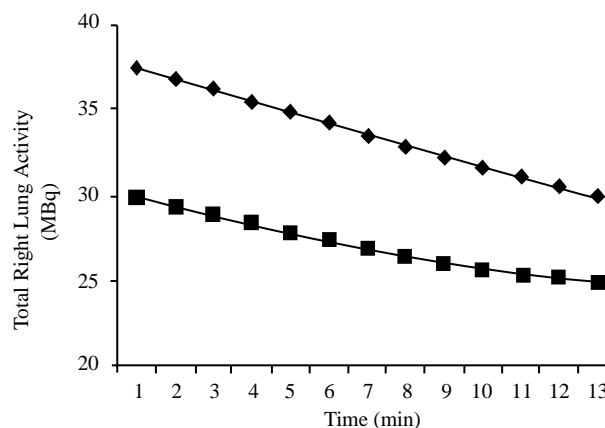


Fig. 3 The time course of activity clearance from the lung measured by fast 1-min SPECT on a healthy subject after inhalation of normal saline aerosols containing ^{99m}Tc -DTPA in large and small droplets (7 μm , \blacklozenge ; 3 μm , \blacksquare). In the plot, 2-min frames are used for clearer illustration.

transmission scan that uses an external γ -Radiation source for imaging of the thorax. To avoid uncertainty in image re-alignment between studies, transmission imaging is best carried out simultaneously with the aerosol (emission) study (45). This can be achieved by attaching the radiation source to the γ -Camera during the aerosol lung image acquisition. The use of the same radionuclide for the transmission as the emission imaging would eliminate uncertainties attributable to differences in the photon energy. When simultaneous emission–transmission imaging is not feasible, the attenuation information can also be obtained from a separate transmission study before the aerosol inhalation. Alternatively, x-ray computerized tomography (CT) scanning can be carried out, but it involves a higher radiation dose to the subject. Also, the separate transmission scan would require extra time, and image re-alignment may be more difficult.

Scatter occurs when the γ -Ray photons emitted from the radionuclide in the lung encounter the thorax tissues and change the traveling direction from their original path. Therefore, scatter causes mispositioning of the detected counts, leading to a loss of image contrast and quantitative accuracy. Quantitative assessment of aerosol deposition would require both attenuation and scatter correction. The latter is also important for proper image reconstruction in SPECT. For additional reading, texts of SPECT imaging are recommended (46, 47).

Data Analysis

Image processing

Planar images can be analyzed on the anterior or posterior view. Because the lung has a thickness, the radiolabel may occur anywhere along the thickness. The effect owing to deposition at varying thicknesses can be minimized by taking the geometric mean of the anterior and posterior images.

Compared with planar images, SPECT data require more rigorous treatment involving mathematical image reconstruction algorithms. Because of simplicity and speed, the filtered back-projection algorithm has been used in clinical SPECT studies including aerosol deposition in the lung. Filtered back-projection uses summation techniques using projections (count rate profiles) of the acquired image data. Artifacts in the back-projected image are corrected by filtering (using mathematical operations) the data before or after reconstruction. Because of the nonuniform attenuation regions of the thorax, other iterative reconstruction techniques involving attenuation and scatter correction have also been used (46, 47).

Drawing of the lung region

The lung outline is obtained from the transmission image or from ventilation image using radioactive gases such as $^{81\text{m}}\text{Kr}$, when available. After superimposing the lung outline on the aerosol emission image, the lung is then divided into different regions of interest, and the amount of activity in each region is measured. Regions such as apex and basal or central and middle and peripheral are most commonly used. A conventional way to analyze aerosol distribution (or deposition pattern) is to express it as the ratio of the activity in one region to another, e.g., peripheral to central, which is conventionally defined as the penetration index. However, there is no consensus on how the size and shape of each region should be drawn. For example, both rectangular regions and lung shape regions have been used for the planar images in the past, but the regions can vary from laboratory to laboratory. Thus, there is a need to standardize the methods for drawing lung regions for comparing studies among different laboratories. The rationale of drawing the lung regions is that the peripheral region should contain more small airways than the central region. In a planar image, there will inevitably be extensive overlapping of structures in the lung. In relative measurement or comparative studies using the same subject as his or her own control, the region size and shape may not be as crucial as in the absolute measurement when the aim is to assign the deposition to certain anatomical structures. The latter is a formidable task because the whole lung contains approximately 1.6 million airway branches. It has recently been attempted on SPECT images with the aid of computer modeling (48). Success will provide valuable information for interpreting the γ -Camera images in relation to the aerosol distribution in the airways.

CONCLUDING REMARKS

Radiolabeling of nebulizer solutions, propellant-driven MDIs, and DPIs is generally well documented. For some DPIs, the procedure may still be lacking in sufficient detail to ensure reproducible radiolabeling. Deposition of radiolabeled aerosols in the lung has been measured primarily by planar imaging. Tomographic imaging using SPECT can provide 3-Dimensional information about the spatial distribution of the aerosol in the lung. The recent development of fast dynamic SPECT has made it the method of choice for aerosol imaging in the new millennium.

ACKNOWLEDGMENTS

I am grateful to Karen Smith for technical assistance and to Stefan Eberl for the availability of Figs. 1, 2 and 3.

REFERENCES

1. Snell, N.J.C.; Ganderton, D. Assessing Lung Deposition of Inhaled Medications. *Resp. Med.* **1999**, *93*, 123–133.
2. Newman, S.P.; Wilding, I.R. Gamma Scintigraphy: An In Vivo Technique for Assessing the Equivalence of Inhaled Products. *Int. J. Pharm.* **1998**, *170*, 1–9.
3. Eberl, S.; Zimmerman, R.E. Nuclear Medicine Imaging Instrumentation. In *Nuclear Medicine in Clinical Diagnosis and Treatment*; Murray, I.P.C., Ell, P.J., Eds.; Churchill Livingstone: Edinburgh, 1998; 2, 1559–1570.
4. Short, M.D.; Singh, C.A.; Few, J.D.; Studdy, P.R.; Heaf, P.J.D.; Spiro, M.D. The Labelling and Monitoring of Lung Deposition of an Inhaled Synthetic Anticholinergic Bronchodilating Agent. *Chest* **1981**, *80 Suppl*, 918–921.
5. Berridge, M.S.; Lee, Z.; Zheng, L.; Heald, D.L. Nasal Distribution of Inhaled Drugs Determined by Positron Tomography. *J. Aerosol Med.* **1999**, *12*, 117.
6. Chan, H.K.; Phipps, P.R.; Gonda, I.; Cook, P.; Fulton, R.; Young, I.; Bautovich, G. Regional Deposition of Nebulised Hypodense Non-Isotonic Solutions in the Human Respiratory Tract. *Eur. Resp. J.* **1994**, *17*, 1483–1489.
7. Newman, S.P. Scintigraphic Assessment of Therapeutic Aerosols. *Crit. Rev. Ther. Drug Carrier Sys.* **1993**, *10*, 65–109.
8. Newman, S.P.; Steed, K.; Reader, S.J.; Hooper, G.; Zierenberg, B. Efficient Delivery to the Lungs of Flunisolide Aerosol from a New Portable Hand-Held Multidose Nebulizer. *J. Pharm. Sci.* **1996**, *85*, 960–964.
9. Vidgren, M.T.; Waldrep, J.C.; Arppe, J.; Black, M.; Rodarte, J.A.; Cole, W.; Knight, V. A Study of ^{99m}Tc-Labelled Beclomethasone Dipropionate Dilauroylphosphatidylcholine Liposome Aerosol in Normal Volunteers. *Int. J. Pharm.* **1995**, *115*, 209–216.
10. Few, J.D.; Short, M.D.; Thomson, M.L. Preparation of ^{99m}Tc Labelled Particles for Aerosol Studies. *Radiochem. Radioanal. Lett.* **1970**, *5*, 275–277.
11. Newman, S.P.; Pavia, D.; Moren, F.; Sheahan, N.F.; Clarke, S.W. Deposition of Pressurised Aerosols in the Human Respiratory Tract. *Thorax* **1981**, *36*, 52–55.
12. Kohler, D.; Fleischer, W.; Matthys, H. New Method for Easy Labeling of Beta-2-Agonists in the Metered Dose Inhaler with Technetium ^{99m}. *Respiration* **1988**, *53*, 65–73.
13. Newman, S.P.; Clark, A.R.; Talaei, N.; Clarke, S.W. Pressurised Aerosol Deposition in the Human Lung with and without an Open Spacer Device. *Thorax* **1989**, *44*, 706–710.
14. Newman, S.P.; Clark, A.R.; Talaei, N.; Clarke, S.W. Lung Deposition of 5mg Intal from a Pressurised Metered Dose Inhaler Assessed by Radiotracer Technique. *Int. J. Pharm.* **1991**, *74*, 203–208.
15. In *The Merck Index*; Merck & Co. Whitehouse Station: NJ, 1996.
16. Zainudin, B.M.Z.; Tolfree, S.E.J.; Biddiscombe, M.; Whitaker, M.; Short, M.; Spiro, S.G. An Alternative to Direct Labelling of Pressurised Bronchodilator Aerosol. *Int. J. Pharm.* **1989**, *51*, 67–71.
17. Zainudin, B.M.Z.; Biddiscombe, M.; Tolfree, S.E.J.; Short, M.; Spiro, S.G. Comparison of Bronchodilator Responses and Deposition Patterns of Salbutamol Inhaled from a Pressurised Metered Dose Inhaler, as a Dry Powder, and as a Nebulised Solution. *Thorax* **1990**, *45*, 469–473.
18. Summer, Q.A.; Clark, A.R.; Hollingworth, A.; Fleming, J.; Holgate, S.T. The Preparation of a Radiolabelled Aerosol of Nedocromil Sodium for Administration by Metered-Dose Inhaler that Accurately Preserves Particle Size Distribution of the Drug. *Drug Invest.* **1990**, *2*, 90–98.
19. Arppe, J.; Vidgren, M. Practical Gamma Labelling Method for Metered-Dose Inhalers and Inhalation Powders. *STP Pharm. Sci.* **1994**, *4*, 19–22.
20. Farr, S.J. The Physico-Chemical Basis of Radiolabelling Metered Dose Inhalers with ^{99m}Tc. *J. Aerosol. Med.* **1996**, *9* (Suppl 1), S27–S36.
21. Everard, M.L.; Stammers, J.; Hardy, J.G.; Milner, A.D. New Aerosol Delivery System from Neonatal Ventilator Circuits. *Arch. Dis. Child.* **1992**, *67*, 826–830.
22. Aug, C.; Perry, R.J.; Smaldone, G.C. Technetium ^{99m} Radiolabeling of Aerosolized Drug Particles from Metered Dose Inhalers. *J. Aerosol. Med.* **1991**, *2*, 127–135.
23. Hardy, J.G.; Everard, M.L.; Coffiner, M.; Fossion, J. Lung Deposition of a Nacystelyn Metered Dose Inhaler Formulation. *J. Aerosol. Med.* **1993**, *6*, 37–44.
24. Ballinger, J.R.; Calcutt, L.E.; Hodder, R.V.; Proulx, A.; Gulenchyn, K. Improved Method to Label Beta-2 Agonists in Metered-Dose Inhalers with Technetium-99m. *J. Aerosol. Med.* **1993**, *24*, 787–794.
25. Hammermaier, A.; Bidlingmaier, A.; Waitzinger, J.; Stechert, R.; Wenske, H.; Jaeger, H. Radio Labelling of Drugs in a Metered Dose Inhaler (MD) and Lung Deposition in Human Subjects. *J. Aerosol. Med.* **1994**, *7*, 173–176.
26. Biddiscombe, M.F.; Melchor, R.; Mak, V.H.F.; Marriott, R.J.; Taylor, A.J.; Short, M.D.; Spiro, S.G. The Lung Deposition of Salbutamol, Directly Labelled with Technetium-99m, Delivered by Pressurised Metered Dose and Drug Powder Inhalers. *Int. J. Pharm.* **1993**, *91*, 111–121.
27. Leach, C.L.; Davidson, P.J.; Boudreau, R.J. Improved Airway Targeting with the CFC-Free HFA-Beclomethasone Metered-Dose Inhaler Compared CFC-Beclomethasone. *Eur. Resp. J.* **1998**, *12*, 1346–1353.
28. Fuller, H.D.; Dolovich, M.B.; Posmituck, G.; Wong, P.W.; Newhouse, M.T. Pressurised Aerosol Versus Jet Aerosol Delivery to Mechanically Ventilated Patients. *Am. Rev. Resp. Dis.* **1990**, *142*, 440–444.
29. Farr, S.J.; Rowe, A.M.; Rubsamen, R.; Taylor, G. Aerosol Deposition in the Human Lung Following Administration from a Microprocessor Controlled Pressurised Metered Dose Inhaler. *Thorax* **1995**, *50*, 639–644.
30. Pitcairn, G.R.; Newman, S.P. Radiolabelling of Dry Powder Formulations. In *Respiratory Drug Delivery VI*;

- Dalby, R.N., Byron, P.R., Farr, S.J. Eds.; InterPharm Press Inc. Buffalo Grove IL, 1998, 397–399.
31. Pitcairn, G.; Lunghetti, G.; Ventura, P.; Newman, S. A Comparison of the Lung Deposition of Salbutamol Inhaled from a New Powder Inhaler at Two Inhaled Flow Rates. *Int. J. Pharm.* **1994**, *102*, 11–18.
 32. Thorsson, L.; Newman, S.P.; Weisz, A.; Trofast, E.; Moren, F. Nasal Distribution of Budesonide Inhaled via a Powder Inhaler. *Rhinology* **1993**, *31*, 7–10.
 33. Newman, S.P.; Moren, F.; Trofast, E.; Talaei, N.; Clarke, S.W. Deposition and Clinical Efficacy of Terbutaline Sulphate from Turbuhaler, A New Multi-Dose Powder Inhaler. *Eur. Resp. J.* **1989**, *2*, 247–252.
 34. Pitcairn, G.; Lim, J.; Hollingworth, A.; Newman, S.P. Scintigraphic Assessment of Drug Delivery from the Ultrahaler Dry Powder Inhaler. *J. Aerosol. Med.* **1997**, *10*, 295–306.
 35. Vidgren, M.T.; Karkkainen, A.; Karjalainen, P.; Paronen, T.P. A Novel Labelling Method for Measuring the Deposition of Drug Particles in the Respiratory Tract. *Int. J. Pharm.* **1987**, *37*, 239–244.
 36. Vidgren, M.; Paronen, P.; Vidgren, P.; Vainio, P.; Nuutinen, J. Radiotracer Evaluation of the Deposition of Drug Particles Inhaled from a New Powder Inhaler. *Int. J. Pharm.* **1990**, *64*, 1–6.
 37. Chan, H.-K.; Gonda, I. Preparation of Radiolabeled Materials for Studies of Deposition of Fibers in Human Respiratory Tract. *J. Aerosol. Med.* **1993**, *6*, 241–249.
 38. Devadason, S.G.; Everard, M.L.; MacErlan, C.; Roller, C.; Summers, Q.A.; Swift, P.; Borgstrom, L.; Le Souef, P.N. Lung Deposition from the Turbuhaler in Children with Cystic Fibrosis. *Eur. Resp. J.* **1997**, *10*, 2023–2028.
 39. Phipps, P.R.; Gonda, I.; Bailey, D.L.; Borham, P.; Bautovich, G.; Anderson, S.D. Comparisons of Planar and Tomographic Gamma Scintigraphy to Measure the Penetration Index of Inhaled Aerosols. *Am. Rev. Resp. Dis.* **1989**, *139*, 1516–1523.
 40. Chua, H.L.; Collis, G.G.; Newbury, A.M.; Chan, H.-K.; Bower, P.N.; Sly, P.D.; Le Souef, P. The Influence of Age on Aerosol Deposition in Children with Cystic Fibrosis. *Eur. Resp. J.* **1994**, *7*, 2185–2191.
 41. Phipps, P.R.; Gonda, I.; Anderson, S.D.; Bailey, D.; Bautovich, G. Regional Deposition of Saline Aerosols of Different Tonicities in Normal and Asthmatic Subjects. *Eur. Resp. J.* **1994**, *7*, 1474–1482.
 42. Chan, H.-K.; Daviskas, E.; Eberl, S.; Robinson, M.; Bautovich, G.; Young, I.H. Deposition of Aqueous Aerosol of Technetium-99m Diethylene Triamine Penta-Acetic Acid Generated and Delivered by a Novel System (AER_x) in Healthy Subjects. *Eur. J. Nucl. Med.* **1999**, *26*, 320–327.
 43. Chan, H.-K. Use of Single Photon Emission Computed Tomography in Aerosol Studies. *J. Aerosol. Med.* **1993**, *6*, 23–36.
 44. Chan, H.K.; Eberl, S.; Daviskas, E.; Constable, C.; Young, I.H. Dynamic SPECT of Aerosol Deposition and Clearance in Healthy Subjects. *J. Aerosol. Med.* **1999**, *12*, 135.
 45. Bailey, D.L. Transmission Scanning in Emission Tomography. *Eur. J. Nucl. Med.* **1998**, *25*, 774–787.
 46. Bailey, D.L.; Parker, F.A. Single Photon Emission Computed Tomography. In *Nuclear Medicine in Clinical Diagnosis and Treatment*; Murray, I.P.C., Ell, P.J. Eds.; Churchill Livingstone: Edinburgh, 1998; 2, 1589–1602.
 47. Meikel, S.R.; Hutton, B.F.; Bailey, D.L. A Transmission-Dependent Method for Scatter Correction in SPECT. *J. Nucl. Med.* **1994**, *35*, 360–367.
 48. Fleming, J.S.; Hashish, A.H.; Conway, J.H.; Hartley-Davies, R.; Nsaaim, M.A.; Guy, M.J.; Coupe, J.; Holgate, S.T.; Moore, E.; Bailey, A.G.; Martonen, T.B. A Technique for Simulating Radionuclide Images from the Aerosol Deposition Pattern in the Airway Tree. *J. Aerosol. Med.* **1997**, *10*, 199–212.
 49. Eberl, S.; Chan, H.-K.; Daviskas, E.; Constable, C.; Young, I.H. Aerosol Deposition and Clearance Measurement: A Novel Technique using Dynamic SPET. *Eur. J. Nucl. Med.* **2001**, *28*, 1365–1372.

RECEPTORS FOR DRUGS: DISCOVERY IN THE POST-GENOMIC ERA

Jeffrey M. Herz

Applied Receptor Sciences, Mill Creek, Washington, U.S.A.

William J. Thomsen

Arena Pharmaceuticals, San Diego, California, U.S.A.

INTRODUCTION

A paradigm shift has occurred in receptor-based drug discovery based upon the elucidation of the complete human genome and advances in genomics research, which have revealed a plethora of potential new receptors on which to focus drug discovery efforts. Based on sequence, structural, and functional similarities of a great number of receptors, these receptor targets can be grouped into relatively few multigene superfamilies. Molecular biological approaches have also revealed that additional receptor subtypes and isoforms are generated through multiple mechanisms. The current challenge lies in the development and application of new strategies for rapidly mining the receptor genome sequences to establishing and understanding their function, which has been referred to as “functional genomics.” Practical problems and approaches to assays for the characterization of the ligand binding and elucidating signaling properties of expressed orphan receptors are critically evaluated. The ensuing characterization of orphan receptors is expected to reveal therapeutically important receptor targets that will serve as a platform for much of the drug discovery enterprise. These findings will spur the discovery and development of a new generation of receptor-subtype specific drugs with enhanced therapeutic specificity.

CONCEPT AND DEFINITION OF RECEPTORS

To paraphrase Voltaire, if receptors did not exist it would be necessary for pharmacologists to invent them. In fact, this is precisely what J.N. Langley did around the turn of the century when he suggested the existence of “receptive substances” through which tissues were able to selectively recognize agonist and antagonist chemicals (i.e., ligands) and, importantly, also provide a mechanism for the translation of this recognition event into a physiological response. Thus, a basic and enduring operational definition of a receptor is that it “must recognize a distinct chemical

entity and translate information from that entity into a form that the cell can read to alter its state accordingly, for example, by a change in membrane permeability, activation of a guanine nucleotide regulatory protein, or an alteration in the transcription of DNA” (1). Hence, the attributes of *ligand recognition* and *signal transduction* are both fundamentally necessary to define receptors. The transduction process may be mediated through an integral part of the receptor structure and may involve receptor interactions with additional nonreceptor proteins. A signal transduction feature of many ligand–receptor interactions is the activation of a kinase domain that initiates phosphorylation in a signal transduction pathway.

This definition has been the basis for an enormous body of scientific investigation into the function and regulation of receptors and mechanisms of drug action. However, final proof of the existence of receptors did not occur until relatively recent applications of modern biochemistry and molecular biology to purify, sequence, clone, and express pure receptor proteins. This lack of proof notwithstanding, the therapeutic basis of many modern, and not so modern, drugs resides in their specific interactions with receptor molecules located in the plasma membrane or cytosol of target cells. In fact, these specific interactions have provided the experimental basis for their discovery and development.

CRITERIA FOR RECEPTOR CLASSIFICATION

Primacy of Molecular Structure

Historically, pharmacologists classified receptors based on the concept that a single receptor mediated a pharmacological response to a single endogenous agonist and that receptor subtypes could be defined primarily by pharmacological properties. Receptor subclassification systems were dependent upon the availability of selective and potent natural substances (toxins and alkaloids) or synthetic ligands, which could selectively elicit or inhibit biochemical and physiological responses from receptors. In some

cases, subtypes could be distinguished by different signal transduction mechanisms associated with each receptor subtype. While this approach led to understanding the molecular and cellular mechanisms of many drugs and to significant therapeutic advances, it did not provide evidence of the numerous and diverse subtypes that are now known to comprise all receptor families.

The enormous molecular diversity and multiplicity of receptor subtypes for a given neurotransmitter or hormone were not fully appreciated until the application of modern molecular biology techniques. The GPCRs (G-protein coupled receptor) currently comprise the single largest receptor superfamily, with estimates of over 1000 receptors in the human genome. Within each receptor family, the multiplicity of receptor subtypes has greatly exceeded the numbers that were predicted on the basis of pharmacological data alone. Now frequent cloning of many receptor genes and the study of the encoded recombinant proteins have both revolutionized the criteria necessary for classification of receptors, which otherwise could not be distinguished pharmacologically and provided fundamental insight into mechanisms of drug action at the molecular level.

Current classification criteria for receptors developed by the Committee for Receptor Nomenclature and Drug Classification of the International Union of Pharmacology (NC-IUPHAR) are based on a combination of information derived from molecular structure (structural), pharmacological characteristics of the receptor (operational or recognitory) and on signal transduction mechanisms used by the receptor (transductional) (1). Of these three essential criteria, the amino acid sequence of the receptor provides a definitive identification of a distinct protein and thus serves as an unambiguous primary basis for classification. Identification of the endogenous ligand or ligands provides a secondary means to group receptors into families. Integration of pharmacological evidence obtained from radioligand binding studies of selective agonists, antagonists, and allosterics (modulatory ligands), and their characterization in functional assays (changes in intracellular cAMP levels, calcium, or reporter gene assays) enables a comprehensive assessment for classification based upon all three essential criteria.

Receptors can be most reliably subclassified and defined on the basis of antagonist affinities, whereas data obtained with agonists are considered much less useful since intrinsic activity and potency have been found to be cell and tissue dependent. Information on receptor-effector mechanisms that reflect the molecular signaling properties of the receptor provides another tier for receptor classification, although transduction mechanisms for novel receptors may be unclear and subject to controversy.

Furthermore, the use of heterologous cellular expression systems for the study of recombinant receptors has revealed the potential for receptor subtypes to initiate signaling events not normally associated with their physiological role. These studies have also demonstrated that the ultimate response to receptor activation is cell-type specific. Hence, the potential ambiguity that may arise in defining receptor transductional characteristics has led to the recognition that this characteristic has more limited value among classification criteria. The question of how to integrate receptor structural information, pharmacological characteristics, and transductional properties into a rational scheme of receptor classification and nomenclature is a subject of continuing discussion among pharmacologists (2).

Receptor Superfamilies, Families, and Subtypes

From the vast number of amino acid sequences deduced from fully sequenced genomes and from cloned receptor proteins, primary amino acid and structural similarities have been identified that are common to numerous receptor types that possess distinct pharmacology. Putative receptor genes usually contain recognizable sequence motifs to suggest a superfamily of receptor they might encode. This homology method has been widely used to find common sequence motifs that typically occur in the extracellular ligand binding domains, transmembrane domains, or within intracellular domains of receptors. This has enabled receptor classification into superfamilies based on sequence similarities that encompass many receptor proteins that differ pharmacologically but are structurally and functionally similar since they share a single molecular signal transductional mechanism. The "common structure–common function" concept is not unique to receptors but originates from the observation that families of proteins are derived from a common ancestral gene. Receptor subtypes within a superfamily are generally considered to have arisen through evolutionary mechanisms of gene duplication and genetic drift leading to divergence from a common progenitor receptor gene rather than reconstruction of new genes *de novo*. Indeed, members of a family are often clustered in a relatively small region of the human chromosome, and the genomic organization of highly homologous receptors (and receptor subunits) often reveals identical structures consisting of the same number of protein-encoding exons. For many members of the GPCRs superfamily, the entire protein is encoded by only a single exon.

Despite the enormous multiplicity of receptor subtypes, a limited number of basic superfamilies have been recognized that currently suffice to accommodate all of the signal transducing receptors. As an example, in the GPCRs

human somatostatin receptors are shown, using the single-letter amino acid code. The seven transmembrane regions are indicated, which show the highest level of sequence identity from the overall receptor sequences. Biophysical and biochemical studies have shown that these sequences

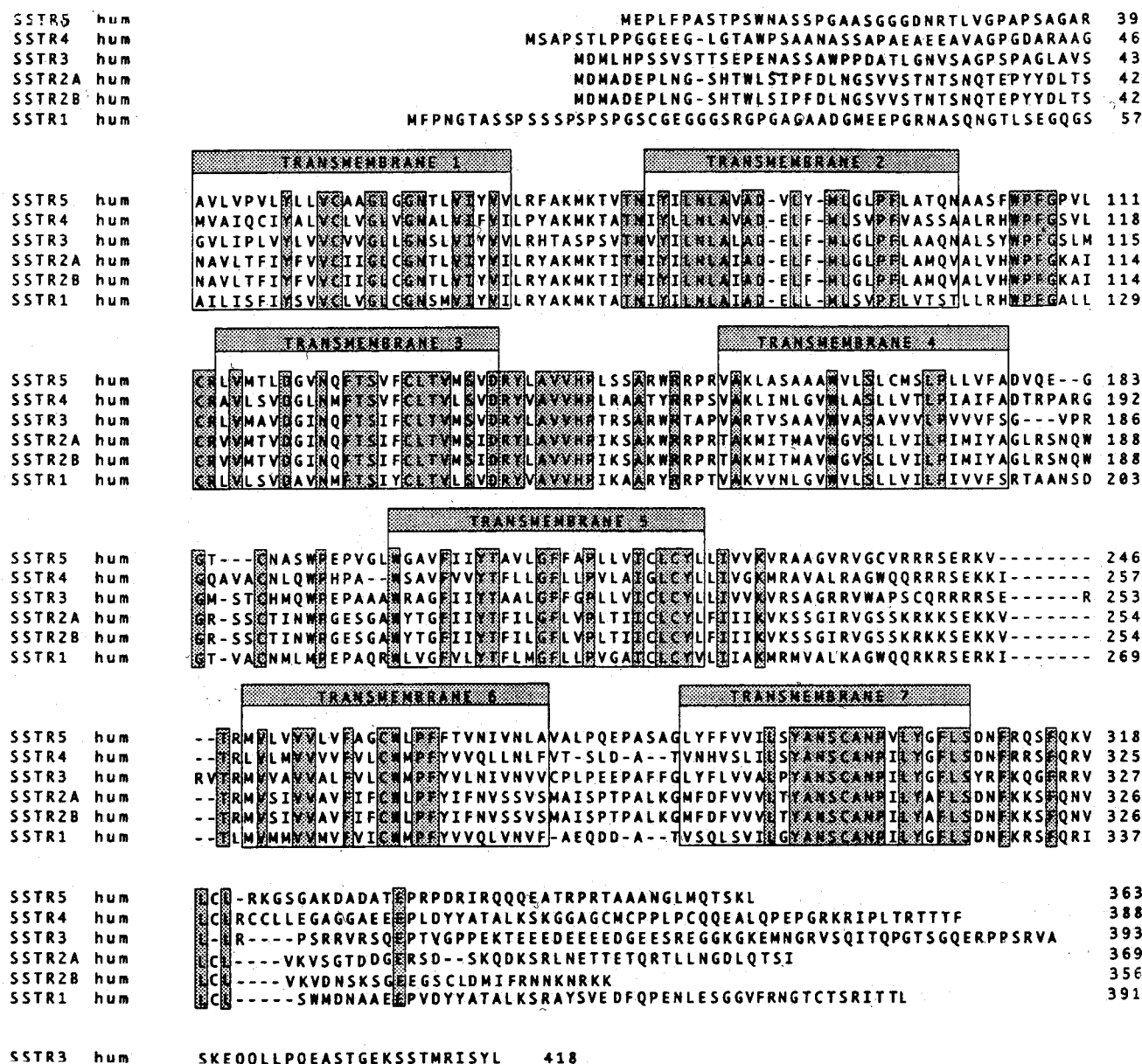


Fig. 1 Primary amino acid sequence comparison between five members of the human somatostatin receptor family. The amino acid sequences are aligned, using the single-letter amino acid code. Gaps introduced in the sequences to optimize the alignments are represented with dashes (–). The boxes indicate amino acids that are identical in all five receptor subtypes. The highest level of homology occurs within the seven canonical transmembrane domains of the human SST receptors (indicated above the sequences) and these domains have been derived by a combination of hydropathy analysis and comparison with the transmembrane domains of the other G-protein coupled receptors. The high level of homology among the members of the SST receptor family in the seven membrane spanning domains is shared by other G-protein coupled receptors that are coupled to inhibition of adenyllyl cyclase.

form a bundle of transmembrane α -helices in the plasma membrane and that the amino terminus is present on the extracellular surface and the carboxyl terminus is present on the cytoplasmic surface of the membrane. Other smaller structural motifs act as functional microdomains in GPCRs. These domains have been mapped by site-directed and deletion mutagenesis studies, generation of receptor chimeras, and through the use of antibodies and include sequences critical for ligand binding, coupling to G-proteins and phosphorylation (Fig. 2). Functionally, all agonist-activated GPCRs signal through interaction with heterotrimeric G proteins, triggering a diverse array of cellular responses. Other major recognized superfamilies include the ligand gated-ion channel receptor (LGCR), receptor tyrosine kinase (RTK), receptor protein tyrosine phosphatase, tumor-necrosis factor (TNF)/nerve growth factor (NGF) receptor, TGF- β serine/threonine kinase receptor, and nuclear receptor superfamilies. Undoubtedly, as many novel sequences of receptors become known that cannot be accommodated within the existing superfamily framework, and their signal transduction properties are defined, new superfamilies of receptors will be proposed. Receptors are further subclassified into receptor families and subfamily groups on the basis of relative

sequence homologies, operational and signal transduction mechanisms. The nomenclature system frequently employed has been to name families with reference to their endogenous ligands, for example, glutamate, serotonergic (5-HT), dopaminergic, epidermal growth factor. Each family typically consists of multiple receptors subtypes that share similarities in their molecular biological, pharmacological, biochemical, and physiological properties. Within a receptor family, members of one subfamily are more closely structurally homologous to one another (on average between 50 and 80%) than to members of other subfamilies (in the range of 25%). However, the current nomenclature system carries forward several notable idiosyncrasies with regard to ligand-receptor systems in which a single endogenous ligand serves as the neurotransmitter for receptors belonging to two distinct superfamilies. Hence, acetylcholine is the transmitter for muscarinic acetylcholine receptors that are GPCRs with seven transmembrane domains (3) and the unrelated nicotinic acetylcholine receptors that are multi-subunit ligand-gated ion channel receptors (LGCRs) (4). Similarly, several neurotransmitters interact with receptors that belong to these two different superfamilies. Glutamate receptors are divided into metabotropic (GPCRs) and the

GPCR Domains

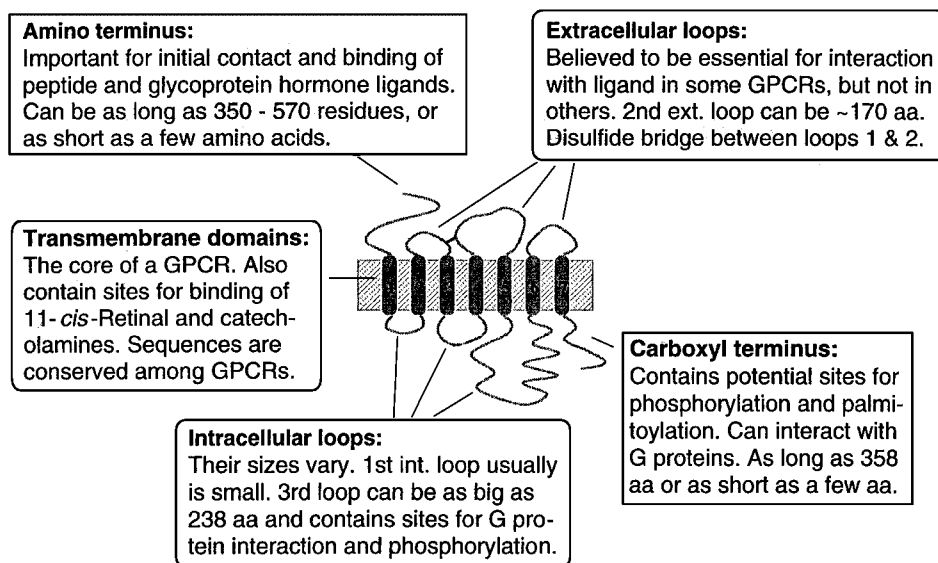


Fig. 2 Seven transmembrane spanning model of the a typical G-protein coupled receptor showing functional domains. The core of about 175 amino acid residues that forms the transmembrane domains (defined by hydrophobicity analysis) are highly conserved among superfamily members. The extra- and intracellular loop regions and C-terminus are much more divergent, even among closely related receptors. Sites for glycosylation are found on the amino terminus and a palmitoylation of a Cys in the C-terminal domain are shown. Intracellular residues involved in G-protein coupling and interaction with receptor-specific kinases that mediate desensitization occur on the intracellular loops.

ionotropic, which include the NMDA, AMPA, and kainate families (LGCRs). GABA receptors are classified as GABA_B (GPCRs) and GABA_A (LGCRs), and purinergic P2 (P2X and P2Y) receptors also have subtypes in both superfamilies. All serotonergic receptors (5HT₁–5HT₇) belong to the GPCR superfamily with the exception of the 5HT₃ receptor, which is a LGCR 5,6).

Multiligand/Multireceptor Families

It is well recognized that all superfamilies contain examples of receptor subtypes that interact with more than one species of endogenous ligand. Receptor binding and tissue-culture experiments have shown that a structurally related family of ligands may bind and activate a single receptor. The existence of families of endogenous ligands that activate either one or more closely related receptor subtypes presents a significant challenge of “receptor promiscuity” to both receptor classification and drug–receptor design. The newly recognized prevalence of such multi-ligand receptor families present among diverse receptor superfamilies suggests that the classical pharmacological dictum of one ligand per one receptor is no longer tenable.

Within the GPCR superfamily, the individual subtypes of CC-chemokine receptors bind and are activated by multiple CC-chemokine ligands. Determining chemokine structure and function is extremely difficult especially when cells contain an ensemble of receptors on their membrane. For example, most chemokine receptors have more than one high-affinity endogenous agonist, and most chemokines bind to more than one receptor subtype. Within the 11 member CC-chemokine receptor family, there are at least five subtypes of chemokine receptors involved in binding the CC-chemokines (also known as β -chemokines): MCP-1, MIP-1 α , and RANTES. Based on studies to date, one receptor (CCR-1) binds all three ligands, another (CCR2) binds only MCP-1, and RANTES binds to CCR1, CCR3, and CCR5. Similar observations of complex ligand–receptor promiscuity are seen with the CXC- or α -chemokine family and five receptor subtypes. Two closely related neutrophil-derived CXC (IL-8) receptors, CXCR-1(IL8R_A) and CXCR-2 (IL8_B), have been cloned whose binding characteristics account for the binding observed with neutrophils. CXCR-1 receptors interact with three ligands (IL-8, GCP-2, and NAP-2) that are also known to bind CXCR-2 receptors.

A second example of multiligand/multireceptor interactions is illustrated by the bone morphogenetic proteins (BMPs), which are members of the TGF- β superfamily. The BMP family of protein ligands has diverse functional roles, including bone and cartilage formation, cell proliferation,

apoptosis, differentiation, and morphogenesis. A number of BMPs have been discovered and, based upon sequence homology, most of the GDFs (growth/differentiation factors) have been added to the BMP family, bringing the number of ligands in this family to around 25. TGF- β , BMP-2, and BMP-7 interact with at least six receptor subtypes of the potentially multimeric Type I and Type II receptors in the TGF- β receptor serine/threonine kinase superfamily. The complexities in devising a uniform, rational nomenclature scheme that incorporates endogenous multiligand-/multireceptor interactions and also is consistent with instances in which a cell surface receptor can simultaneously function as both a ligand for a receptor on another cell and as an independent receptor entity are substantial.

Receptor Subtypes Composed of Combinatorial Subunits

The central tenet of the current classification system is to define a unique receptor subtype for each receptor that is composed of only single polypeptide encoded by distinct gene. As an example, the dopamine receptor family is known to contain at least five distinct subtypes, dopamine D₁–D₅ receptors, each receptor consisting of one subunit encoded by separate genes. A more complex system arises in the case of multisubunit oligomeric receptors, as in the case of the LGCR superfamily. These multisubunit receptors are known to exist as hetero-oligomers with many subunit combinations that comprise distinct subtypes. For several families, homologous genes encoding highly related receptor subunits have been designated by a combination of greek letters and numbers. Molecular cloning studies of GABA_A receptors have identified five different subunit types named α , β , γ , δ , and ρ , which share significant amino acid sequence identity with each other (30–40%) and with other members of the ligand-gated ion channel superfamily. Nearly all of these subunits also have a number of different isoforms (α 1–6, β 1–4, γ 1–3, δ , and ρ 1–2) and all sequences within each isoform family are highly homologous (70–80% identity). Individual GABA_A receptor subtypes are thus defined by the distinct combination of subunits and stoichiometry that compose the oligomeric complexes, such as α 2 α 3 β 3 γ 2, α 1 α 3 β 2 γ 2, and α 1 α 3 β 2 γ 2. However, subunits of the NMDA, AMPA, and kainate receptor-ion channel complexes have been classified using an alternate system; NMDA1 and NMDA2A–D, and glu1-7 and ka1-2 . The combinatorial mechanism of receptor assembly has the potential for producing a large number of receptor subtypes and is a common feature of the multigene families that comprise the LGCR superfamily.

Molecular Mechanisms Producing a Diversity of Receptor Isoforms

Additional variation in receptor sequences occurs through naturally arising mutations, naturally occurring allelic variants, RNA splicing, and RNA editing. For example, the human dopamine D₄ receptor subtype contains a direct repeat of a 16 amino acid segment in the putative third intracellular loop of the receptor. Naturally occurring allelic variants of this receptor subtype contain variable numbers of direct repeat units (e.g., 2, 4, 7 repeats; dopamine D_{4.2}, D_{4.4}, D_{4.7}) producing a substantial number of receptor variants in the human with different primary amino acid sequence but as yet indistinguishable receptor binding and signal transduction pharmacology for a wide range of ligands (7). Other subtype variants occur at the level of RNA splicing giving rise to length variants of the receptor subtype. For example, differential splicing results in the long and short isoforms of the dopamine D₂ receptor, D_{2L} (D_{2A}) and D_{2S} (D_{2B}) (7), and produces two proteins differing by 29 amino acids. RNA editing is yet another mechanism producing sequence and functional diversity and has been found to produce seven different edited forms of the kainate GluR6 subunit, all encoded by only one gene. While many receptors are subject to posttranslational modifications, such as phosphorylation, which modulate receptor properties, these differences are not considered to define a receptor subtype.

Species homologs of the same receptor subtype display a high degree of amino acid sequence homology, typically showing 85–95% sequence identity between species. This is a greater sequence similarity than is usually found between distinct receptor subtypes in the same species. The introduction of cloning has enabled identification of species homologs of receptors that otherwise might be defined as distinct receptor subtypes due to their species-specific pharmacology, as was found in the case of the rat 5-HT_{1B} and human 5-HT_{1B} (5HT_{1DB}) receptors (8). Since a change in as little as a single amino acid within the binding site domain can markedly alter ligand affinity, some receptor species homologs exhibit distinctive pharmacology. A notable example is the comparative binding of the nonpeptide antagonist ligands, CP96345 and RP67580, to the human and rat neurokinin NK₁ receptors. While CP96345 binds with nanomolar affinity to the human NK₁ receptor, it exhibits two orders of magnitude lower affinity for the rat receptor, whereas RP67580 shows selectivity for the rat receptor. The magnitude of the selectivity difference is noteworthy given that the rat and human receptors are 95% identical. Construction of single residue substitutions in recombinant NK₁ receptors demonstrated that an exchange of two

residues was sufficient to reverse the species selectivity of the two ligands. Due to the number of cases in which species homologs differ from human receptor pharmacology, the use of cloned human receptors for drug discovery screening programs is increasingly favored.

Orphan Receptors

Putative receptors identified by gene cloning that exhibit homology to known receptors in the existing superfamilies but for which no known ligands have been identified are referred to as *orphan receptors*. Initially, these novel receptors may be classified based upon the level of sequence homology. When the endogenous ligand for an orphan receptor remains unknown, the receptor may be provisionally named based on the binding of a synthetic ligand to the receptor until the endogenous ligand is identified. As the functions of an orphan receptor are uncovered, these novel receptors frequently become targets for the development of new therapeutics.

LIGAND–RECEPTOR INTERACTIONS DEPEND UPON RECEPTOR CONFORMATIONAL STATES

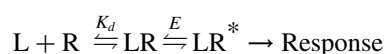
Resting, Activated, and Desensitized Receptor States Differ in Ligand Affinity

One of the major criteria in receptor subtype classification is the pharmacological selectivity profile as defined by ligand affinities (equilibrium dissociation constant, K_d) and activation constants. Since binding data overcome many of the limitations inherent in studies of biological responses, ligand affinities have provided a practical method to define and classify related subtypes. Nevertheless, accurate comparisons of agonist affinities of cloned receptors with binding affinities characteristic of the native cell type or tissue may be difficult to obtain. Agonist affinities determined for recombinant receptors expressed in distinct heterologous cell types have been found to vary substantially and differ from value characteristic of the native cell type (9). These findings underscore the need to compare ligand affinities and function of an expressed recombinant receptor in several host cell lines and for related receptor subtypes to be expressed in a common host cell line in order to allow accurate pharmacological comparisons (10). Receptor classification schemes will need to take into account the host cell environment when evaluating ligand affinity and functional data.

In contrast to traditional receptor theory, it is now evident from molecular pharmacological analysis that

transitions between receptor states can be induced by interaction of a receptor with a full spectrum of ligands that have very different capacities to activate the receptor. Ligand-induced changes in the population distribution of receptor molecules between these states govern the binding and functional properties of the receptor. As discussed in detail ahead, measurements *in vitro* and *in situ* have revealed the importance of temporal responses in characterizing multiphasic ligand–receptor interactions for the ligand-gated ion channel and GPCRs superfamilies.

An agonist, by virtue of the molecular “information” it contains (e.g., size, 3D configuration, charge distribution, hydrogen- or ionic-bonding residues, chirality), selectively binds to a complementary 3D surface or binding site domain formed by amino acid residues of a receptor protein and initiates a cellular response. Drugs classified as full agonists elicit the maximal response while compounds that only elicit a fractional response are referred to as partial agonists. Regardless of whether the agonist is an endogenous physiological ligand, a natural product, or a synthetic compound, agonist binding is coupled to conformational changes in the receptor protein that involve a molecular transition to a common, active receptor state. As an example, binding of either full, strong, or weak partial agonists to the nicotinic acetylcholine receptor induced a transition to an open channel state that is characterized by the same unitary conductance. Although the frequency and duration of the receptor open state differ, this is indicative of a single, active conformation of the receptor. The simplest schemes for receptor activation take into account that occupation of a receptor (R) by an agonist (L) results in a conformational change in the receptor to create an agonist-activated state (LR*), which can bring about an effect or response, and can be represented as:



where L is the agonist, R represents the unoccupied (inactive) receptor, R* is the active state of the receptor, K_d is the dissociation constant for agonist-receptor binding, and E is the constant describing the equilibrium between the LR and the LR* states of the receptor.

Contemporary models of receptor activation attempt to incorporate increasingly detailed knowledge of multiple conformational states, including desensitized receptor states. A fundamental component of these molecular activation schemes for receptors is that receptors exist in a dynamic equilibrium between three states, an inactive (R), an active conformation (R*) and a desensitized state (R'), and that the biological response to a given ligand is governed by its ability to change the equilibrium (or its relative preference for binding) between the three receptor

states. Continued exposure of LGCRs to agonists promotes a multiphasic conversion to desensitized receptor states that develop on time scales ranging from milliseconds to minutes. Generally, GPCR activation is also followed rapidly by a loss of responsiveness, also termed desensitization, which is then followed by a period of recovery or resensitization. These changes in signaling are tightly regulated, primarily via mechanisms that involve rapid GPCR phosphorylation, internalization of receptors that remove them from the cell surface, and either degradation or recycling back to membrane surface (11).

Many receptors that function as ligand-gated ion channels can be described by a two-state cyclic scheme for receptor desensitization (including only one desensitized state), which was initially deduced from extensive studies of the kinetics of nicotinic acetylcholine receptor (AChR) state transitions induced by ligand binding (12). This model was expanded to a general coupled-equilibria model that described the molecular species and influence of agonists and noncompetitive inhibitors (noncompetitive antagonists) on receptor states (13). As shown in Fig. 3, the model incorporated the fact that the AChR contains two interacting agonist binding sites and a topographically distinct binding site for an allosteric, noncompetitive inhibitors.

For many receptors, the agonist binding site of the receptor protein can interconvert between high and low affinity binding states or receptor conformations. Moreover, differences in affinity of the agonists for the same site in different receptor states may be dramatic. In the case of LGCRs, the desensitized states display higher agonist binding affinity and no receptor-mediated ion permeability. It is the preference for ligand binding to the desensitized receptor that is the driving force underlying conversion to the desensitized state. The binding affinity (K_d) of acetylcholine to the muscle subtype of the nicotinic acetylcholine receptor measured at equilibrium is about 10 nM (desensitized state), 4–5 orders of magnitude below the apparent dissociation constant for the permeability response (activatable state) mediated through the same binding site (14).

Multiple agonist-affinity receptor states have also been described for many GPCR receptor families. GPCRs may adopt either a high-affinity or low-affinity state for agonists (K_H and K_L), which may differ by 1000-fold when characterized in native membranes purified from either brain tissue or derived from cell lines. Low-affinity agonist states for receptors, which typically cannot be determined by direct binding of a radiolabeled agonist, have been measured by employing radiolabeled antagonists and examining the pattern of competition over a wide range of agonist concentrations. For recombinant GPCRs, agonist-induced

changes in affinity depend upon efficient coupling of the expressed recombinant receptor to endogenous G-proteins present in the host cell line, and coupling is highly dependent on the combination of receptor and clonal cell line. For example, agonist-stimulated D₃ dopamine receptor-mediated effects were lacking, and an inability to demonstrate significant affinity shifts for agonist ligands was found in the presence of guanine nucleotides in COS or CHO cells (14). However, receptor coupling to inhibition of adenylyl cyclase was readily obtained for closely related subtypes in the D₂ receptor subfamily, recombinant D₂ and D₄ receptors, expressed in a variety of cell lines (CHO-K1, HEK-293, C6-glioma, Ltk⁻) (15,16).

Antagonist Modulation of Receptor Conformational States

According to classical models for drug–receptor interaction, full competitive antagonists and agonists

share the ability to bind to a common site on the receptor molecule but differ in that antagonists are devoid of intrinsic activity. While a competitive antagonist occupies the binding site of the receptor, it does not promote conversion of the receptor to the active conformation. Indeed, at the biochemical level, nicotinic acetylcholine receptor agonists and antagonists that differ in size and structure (acetylcholine, tubocurarine, and snake venom α -toxins) have been mapped to the same molecular site on the receptor α -subunit and their binding is mutually exclusive. However, recent data suggest that no overlap in the binding site for competitive ligands is required if they bind in a mutually exclusive manner to different binding sites that are present only in different receptor conformations. Competitive antagonists may be subclassified into at least two categories: 1) neutral antagonists that do not exhibit a preference for an inactive or active receptor conformation and have no effect on basal receptor

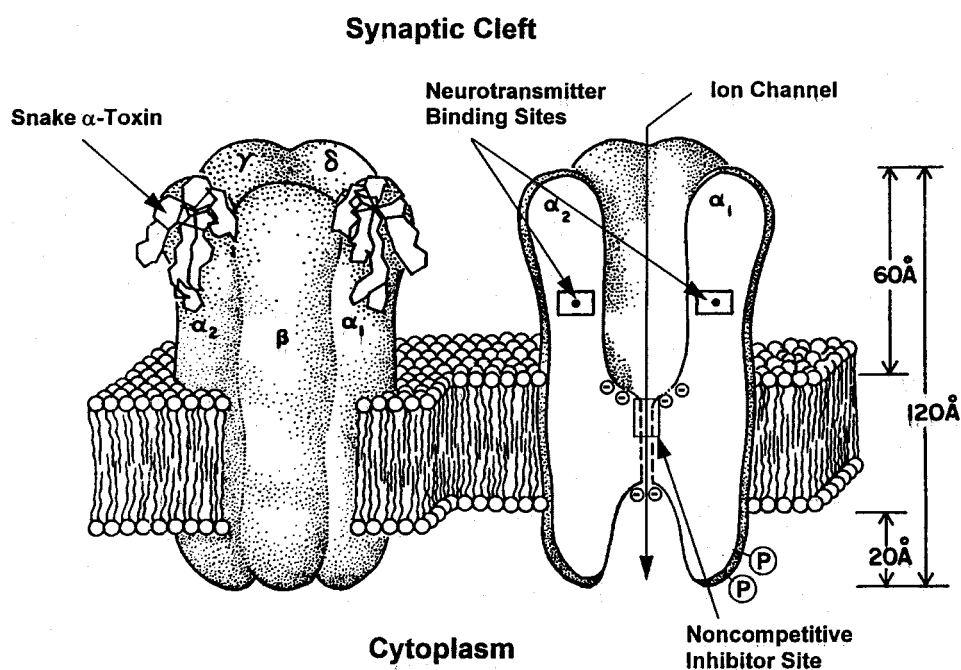


Fig. 3 Structure and ligand binding sites of the muscle subtype of the nicotinic acetylcholine receptor. The structure is derived from electron microscopy of two-dimensional crystals of receptors in *Torpedo* membranes. The five subunits together form a cylindrical shape approximately 120 Å long with a 70 Å diameter, which spans the membrane. The receptor on the right shows a section along the axis of the receptor and the cation conducting pathway in profile. A 60 Å diameter central hydrophilic tube is present on the synaptic side which is the entrance to the transmembrane ion channel which narrows at the level of the membrane. Current resolution of the receptor cannot resolve dimensions of the ion channel within the membrane bilayer which permeability studies have defined as an aqueous pore of about 7 Å diameter. Locations of the ligand binding sites have been mapped within the receptor structure by fluorescence energy transfer techniques. The two agonist binding sites, one on each α -subunit, are 20–30 Å from the membrane surface and the single allosterically-coupled noncompetitive inhibitor site is located in the channel. Positions of two snake venom α -toxin molecules (peptides of 7500 kD, peptide backbone shown) bound to the synaptic surfaces of the α -subunits are shown for the receptor on the left.

activity and 2) negative antagonists, also termed inverse agonists, which exhibit the defining property of inhibiting agonist-independent receptor activity and possess negative intrinsic activity. Furthermore, negative antagonists stabilize a different receptor conformation (state) than neutral antagonists.

For both LGCRs and GPCRs, whose activity is governed by distributions between inactive and active states, negative antagonists promote conversion of the receptor to an inactive conformation and stabilize a conformation of the receptor that is distinct from the unliganded receptor (resting state). For G-protein coupled receptors, studies of receptors expressed at high levels, using the baculovirus expression system in Sf9 cells or of receptors that have been rendered constitutively active by site-directed mutagenesis, have provided an experimental system in which agonist-independent, spontaneous adenylyl cyclase activity can be measured. This phenomenon has been best characterized for the β_2 -adrenergic receptor but has also been observed for the bradykinin, dopamine D₁, and 5-HT_{2C} receptors, suggesting it is likely to be a general feature of the superfamily. A group of well-characterized β_2 -adrenergic receptor antagonists were found to inhibit, to varying degrees, receptor dependent spontaneous activity (17). Negative antagonists have been shown to promote the dissociation of spontaneously occurring receptor–G-protein interactions, using constitutively active mutant receptors. As an example, the ligand ICI118551 has been shown to inhibit the basal signaling activity of the β_2 -adrenergic receptor, thus acting as a negative antagonist (18).

Inverse agonist drugs may also span a range of activities, and thus current ligand and drug classification schemes recognize full agonists, partial agonists, neutral antagonists, and partial and full inverse agonists. For the ligand-gated ion channels, pharmacological studies of the interaction of the nicotinic acetylcholine receptor with a series of *N*-substituted analogs of decamethonium identified several compounds as antagonists but unique in their capacity to antagonize receptors that had been previously exposed to agonists. Through an extensive series of equilibrium and kinetic binding experiments, it was shown that these antagonists preferentially bound to the desensitized receptor state and had the ability to promote the conversion of the nicotinic acetylcholine receptor to this inactive state. In contrast, classical antagonists, such as tubocurarine, demonstrated the same affinity for both resting and desensitized receptor states. Antagonists that demonstrated enhanced affinity for the desensitized receptor state were termed metaphilic antagonists.

Noncompetitive Antagonists

Noncompetitive antagonists interact with a spatially distinct site (NCI site), which may be allosterically regulated by the agonist binding site(s) and block receptor function through an allosteric mechanism. In the absence of agonist, the allosteric constant, M , is increased by binding of these inhibitors, and in the presence of agonist, the rate of agonist-elicited conversion to the desensitized state is accelerated. For the LGCRs, the allosterically coupled NCI site is fundamentally different from the agonist site in its pharmacology. A study of the interaction of a phenylphenanthridium ligand, ethidium, with the high-affinity NCI site of the nicotinic acetylcholine receptor determined that it bound with extremely high selectivity to the desensitized state relative to a state stabilized by interaction with snake α -toxin at equilibrium, showing a ratio greater than 2800 (13). Conversely, ethidium binding at the NCI site and its occupation converted the receptor to a state of higher agonist affinity. Similarly, association rates for the binding of radiolabeled phencyclidine (PCP) to the NCI site increased by a factor of 1000–10,000 for the transient open-channel state relative to the toxin stabilized receptor state (19). Potential explanations for these findings are that the binding site location within the transmembrane ion channel is sterically inaccessible in a toxin-stabilized receptor conformation (Fig. 3) and that binding occurs only to selected conformational states, such as the open channel state (20). Numerous ligands act through allosteric mechanisms to regulate receptors in the LGCR superfamily. As in the case of the AChR, drugs such as MK-801 and PCP bind within the ion channel domain of NMDA receptors, blocking agonist-dependent activation. GABA_A receptor function is also regulated by the binding of benzodiazepines, barbiturates, and steroids at distinct allosteric sites on the receptor.

Methods to Study Drug–Receptor Interactions in Real Time

Drug–receptor interactions may quantitatively differ among rapidly converting multiple receptor states that are induced by agonist binding. While most studies commonly focus on analysis of ligand–receptor interactions and cellular responses under presumed steady-state conditions, experimental techniques available for the quantitative analysis of drug–receptor interactions with transient receptor conformational states are more limited. Since the actions of many receptors regulate rapid (within seconds) and transient changes in intracellular calcium and

protein kinase activity, the magnitude and the duration of the response to either transient or sustained agonist stimulation will determine the integrated cellular response. Thus, the effects of drugs on the temporal integration of individual receptor responses are critical in determining the response of a cell to extracellular stimuli.

Molecular pharmacological techniques have been applied to study rapid, real-time analysis of ligand–receptor conformational states and cellular signaling. A variety of rapid mixing techniques have provided sensitive and rapid measurements capable of defining receptor conformational transitions induced by agonist or drug binding to membrane-bound receptors. The earliest ligand–receptor kinetic studies measured radioligand binding of agonists and functional responses, using rapid filtration, but were technically limited in kinetic resolution. Direct spectroscopic measurements extending into the millisecond time domain have been made of conformational transitions for the nicotinic acetylcholine receptor (LGCR) and the formyl peptide (fMLP, GPCR) receptor because the affinity of fluorescent agonist ligands is greatly altered by interaction with their respective receptors (21). Real-time analysis of fluorescent formyl peptide ligand binding to intact human neutrophils by both spectrofluorometric and flow cytometric methods has led to a model of signal transduction dynamics and ligand–receptor–G-protein ternary complex interactions (22). Extrinsic fluorescence labeling of receptors by covalent modification of the receptors with reporter groups has also provided evidence for conformational changes in LGCRs and GPCRs. Spectroscopic signals originating from incorporation of an environmentally sensitive fluorescent probe into the purified, human β 2-adrenergic receptor provided kinetic information indicative of both agonist and negative antagonist-mediated receptor conformational changes (23).

Electrophysiological analysis of drug interactions with single receptor ligand-gated ion channels and voltage-dependent channels has provided another invaluable approach to analyze the activated open-channel states and closed states of LGCRs and voltage-activated ion channels with time resolution unmatched by other techniques. Using the methods of patch clamping, these high-resistance seals (gigaohm) attached to membrane patches or intact cells have recorded changes in conductance arising from the opening of individual transmembrane channels. Studies of this nature have been crucial in defining functional characteristics of many subtypes of receptor-ion channels and allowed the investigation of the mechanism of interactions of drugs with open channel states of these receptors, which are otherwise not amenable to study. In a classic study, the channel blocking mechanism of the local

anesthetic QX-222 with the nicotinic acetylcholine receptor was defined (24). Rapid and repeating flickering changes in conductance of one channel were observed as single drug molecules stochastically entered, blocked, and left the receptor ion channel. Kinetic constants derived from such measurements can define state-dependent blockade of receptor channels.

Flow cytometry has been utilized as a powerful approach to collect multiparameter kinetic data to evaluate multiple activation parameters simultaneously in individual cells and to correlate these parameters with the ligand occupancy of each cell. Cellular functional parameters that can be measured employing this technique include intracellular calcium, magnesium, pH, and membrane potential. Choices of functional parameters depend only upon the availability and utilization of a wide range of specific fluorescent probes that can act as reporters of the desired cellular parameter. Now a common procedure is to employ a fluorescent probe for calcium to measure continuous changes in intracellular calcium in populations of cells in suspension or in cellular monolayers of immobilized cells, using conventional spectrofluorometers. Utilized in combination with fluorescence digital imaging microscopy, spatial gradients reflecting changes in ion concentrations within individual living cells in the millisecond time domain have been visualized. New instrumentation has enabled adaption of this technique to a high-density plate format that is suitable for high-throughput screening in drug discovery programs.

Progress in defining the spatial and temporal cellular location of receptors in living cells during signal transduction has advanced through the use of confocal fluorescence microscopy coupled with the development of receptor-specific fluorescent probes. Receptor fluorescent probes, in particular, green fluorescent protein (GFP), have become a popular reporter molecule that has allowed measurements agonist mediated events in real time. Tagging of GPCRs with the GFP has enabled the direct visualization of real-time trafficking of GPCRs in living cells (11). The approach utilized has been to generate expression constructs that contains GFP fused to the carboxyl terminus of the receptor. Synthesis and expression of over 20 different GPCR–GFP chimeric fusion receptors has shown rapid changes in GPCR receptor cellular distribution in the membrane and the subsequent internalization of ligand and receptor. Such analyses have provided crucial insight into the mechanisms involved in controlling GPCR function and enable the studies of the actions of GPCR drug agonists. While basic receptor studies employing fluorescent ligands are still few, this nascent area is expanding as additional

specific fluorescent ligands, recombinant receptors, and instrumentation become widely available. These studies have also established fluorescence-based detection methods as the preferred format for ultrahigh throughput screening in drug discovery programs.

FUNCTIONAL ROLES FOR A MULTIPLICITY OF RECEPTOR SUBTYPES

The approach of molecular cloning has greatly expanded our knowledge of many receptor subtypes and has revealed an impressive heterogeneity of receptors not previously envisioned. The discovery of multiple receptor subtypes that are highly homologous has provoked examination of key questions related to their physiological functional significance and their importance as individual drug discovery targets. Substantial efforts have been made to define specific physiological roles and advantages conferred to the organism by expression of multiple receptor subtypes that have apparently identical endogenous ligand binding and signal transduction properties. Despite examples of receptor subtypes that exhibit indistinguishable binding and functional coupling properties, these closely related subtypes and isoforms may exhibit differences in properties that include the time course of activation, ionic selectivity, desensitization, regulation, and efficacy in coupling to distinct second messenger systems. Furthermore, homologous receptor subtypes within a family often display differentially regulated expression in particular cell types and tissues in response to specific stimuli.

Molecular Mechanisms Generating Diversity

Small differences in receptor structure (one to several amino acids) have been found to endow important pharmacological and functional differences to receptor subtypes and species homologs. Within the GPCR superfamily, a change in a single amino acid residue was found to underlie the major pharmacological differences between the rat 5-HT_{1B} and the 5HT_{1DB} receptors (25). Similarly, a change of one residue (Asn351Asp) in the dopamine D₅ receptor results in a 10-fold decrease in dopamine binding affinity while mutations in the D₂ receptor lead to differences in binding affinities for typical and atypical neuroleptic drugs used in the treatment of psychotic disorders. Essential differences in coupling of splice variants of the dopamine D₂ receptor, mglu1 receptor, somatostatin sst₂ receptor, and prostaglandin EP₃ receptor have been observed. Within the ligand-gated ion

channel superfamily, it has been found that the charge and size of a single amino acid residue in a critical channel site (the Q/R site of the GluR-B subunit) determines the particular ionic selectivity and permeability properties that characterize subtypes of AMPA-kainate-activated channels (26). For the muscle subtype of the nicotinic acetylcholine receptor, the functional properties of the receptor (channel conductance and open time) are altered during development by substitution of one of the subunits in the pentameric complex; from (α 1)₂ β 1 γ δ in embryonic muscle to (α 1)₂ β 1 ϵ δ in the adult muscle. Thus, the highly significant changes in receptor structure and function that can result from subtle variations in the genetic sequence encoding receptors can affect target identification, characterization, and validation in the drug discovery process.

Cellular and tissue expression patterns for related receptor subtypes may overlap but are often highly specific and tightly regulated. For example, there are striking differences in the cellular distribution of CXCR1 (IL-8R1) and CXCR2 (IL-8R2) receptors (77% sequence identity overall). CXCR1 expression is restricted to neutrophils, monocytes, and a few myeloid cell lines, whereas CXCR2 is widely distributed in myeloid cell lines but also in lymphocytes, melanoma cells, melanocytes, and fibroblasts (27). These specific cellular distributions suggest that the main function of the CXCR1 receptor is neutrophil activation and mobilization of phagocytes in host defense while the CXCR2 receptor serves to mediate migration and growth of cells not involved in defense. In situ hybridization studies of mRNAs encoding subunits of the neuronal nicotinic acetylcholine receptor gene family (LGCRs) and subtypes of GPCRs have documented that transcripts corresponding to homologous receptor subunits and subtypes within a receptor family exhibit distinct anatomical distributions in the mammalian brain (28). Among the family of human dopamine receptor subtypes, D₂ receptors are most highly expressed in the striatum, D₃ and D₅ receptors exhibit a preferential localization in the limbic area of the brain and low expression in motor areas such as the basal ganglia, while D₁ and D₄ receptors are both abundant in the cerebral cortex. Thus, the operative paradigm is that functional diversification and heterogeneity in expression patterns for receptor subtypes allows each subtype and isoform to perform a unique physiological role.

Defining Receptor Subtype Function

The process of defining receptor subtypes associated with cloned sequences and their associated physiological and pharmacological properties remains a major challenge,

particularly for the ligand-gated ion channel superfamily. The molecular diversity of subunits that comprise the nicotinic acetylcholine, GABA_A, and glutamate LGCRs families allows for a plethora of subtypes composed of unique heteromeric subunit combinations. Many of the subunits do not directly bind ligands but nevertheless dictate binding properties for the receptor complex. Expression of distinct combinations of subunits is required to understand the individual roles each subunit plays in creating pharmacologically distinct receptor subtypes when expressed in heterologous cells. Such expression studies have shown that recombinant GABA_A receptors composed of different α -subunit isoforms in combination with invariant β and γ subunits display distinct benzodiazepine pharmacology, GABA–benzodiazepine interactions, and steroid modulation. The existence of multiple receptor subtypes with unique pharmacological characteristics and differential subanatomical localization provides the rationale for their use as molecular targets in the development of novel subtype-selective drugs that are targeted to specific tissues and are hopefully devoid of the side effects associated with existing nonselective drugs.

CLONED HUMAN RECEPTORS AS DRUG DISCOVERY TARGETS

An important application of recombinant DNA technology has been the capability to provide human, cloned receptor subtypes as expressed functional proteins for drug discovery efforts. Prior to the availability of the cloned receptor targets, screening programs relied upon utilization of receptors obtained from animal tissue homogenates that had the inherent disadvantages of receptor heterogeneity, nonhuman pharmacology, and low receptor expression levels. The development and utilization of stable cell lines that express high levels of a single human receptor subtype for drug screening allow for a more accurate and selective pharmacological screening method compared with conventional methods and have enabled the discovery of highly selective drugs that discriminate between various receptor subtypes. Furthermore, cell lines have been engineered to express a recombinant receptor linked to a second messenger or other downstream biochemical response so that the functional consequences of receptor occupancy can be detected. Thus, not only can drug binding affinity of novel lead compounds be quantitatively defined, but the use of such recombinant cell lines permits rapid characterization of agonist or antagonist activity and measurements of compound efficacy at the cellular level.

Many human recombinant receptor subtypes expressed in a wide variety of human and other mammalian host cell backgrounds (e.g., CHO, HeLa, L-cells) have been found to exhibit ligand binding properties indistinguishable from their nonrecombinant “native” receptor counterpart. However, the use of recombinant receptor systems for receptor classification and pharmacological characterization can generate misleading data if heterologous host cell-specific properties have been imposed on the receptor. Ligand binding and signal transduction properties of many GPCRs have been found to depend upon the host cell as well as the expression level of the receptor, and thus agonist affinities and coupling mechanisms determined in different recombinant expression systems may not be identical. In the case of the expression of the human 5HT_{1A} in HeLa cell lines, certain compounds acted as agonists in a cell line with a high receptor density but acted as pure antagonists in a cell line with only a sixfold lower density of receptors. Similarly, data for other GPCRs have been obtained showing the dependence of agonist efficacy and pharmacological properties on receptor density.

The question of analyzing receptor function must be viewed within the context of the functional level for analysis as to whether the molecular entity is the single receptor, a signal transduction pathway, or an integrated cellular response. For example, activation of a specific signal transduction pathway is often viewed as a constituting a receptor response. Insight into the complexities of analyzing the responses of recombinant receptors in heterologous cells is provided by studies of the signal transduction properties resulting from the activation of the D₂ dopamine receptor stably expressed in two different cell lines. Activation of D₂ receptors in either GH₄C₁ rat pituitary cells or mouse Ltk[−] fibroblasts produced inhibition of adenylyl cyclase activity (29). However, while D₂ receptor activation caused a rapid stimulation of phosphatidylinositol (PI) hydrolysis and an increase in intracellular calcium in Ltk[−] cells, it failed to effect PI hydrolysis and induced a decrease in intracellular calcium in GH₄C₁ cells. Similarly, activation of the 5HT_{1A} receptor expressed in Ltk[−] cells was coupled to an increase in intracellular calcium but resulted in a decrease in calcium influx when expressed in GH₄C₁ cells. Since the cell-specific differences in the signaling pathways for the D₂ and 5HT_{1A} receptors were the same, these effects may most likely be attributed to the complement of G-proteins present in each host cell type. Thus, the effector systems and responses of a transfected receptor depend not only on the receptor but upon the particular cell type in which it is expressed.

The high level of membrane expression for recombinant GPCRs in insect cells (1–40 pmol/mg in Sf9 or Sf21

cells) using the baculovirus system has demonstrated appropriate receptor pharmacology for antagonist ligands (affinities and rank order potencies), whereas anomalous agonist binding properties and lack of functional coupling have been observed for some receptors. The GPCRs expressed using baculovirus typically show only a single, low-affinity binding state for agonists, whereas the same receptors demonstrate high- (coupled) and low-affinity (uncoupled) agonist binding states in mammalian cells. Similarly, the pharmacological properties of GPCRs expressed in yeast (M_1 muscarinic and D_2 dopamine receptors) were not comparable with mammalian cells since only a single low-affinity agonist binding state was detected, indicating that the recombinant receptors do not couple to the endogenous G-proteins present in these cells (30, 31).

Cellular engineering to coexpress recombinant GPCRs and the appropriate G-protein has been found to provide an approach to overcome such limitations. Toward this end, it was recently found that heterologous coexpression of the G protein α -subunit $G\alpha_{16}$ with a variety of GPCRs in mammalian cells enabled coupling to $PLC\beta$ activity and increases in intracellular calcium. In yeast, functional coupling was only achieved using a mutant strain lacking one yeast G-protein α -subunit and including a cDNA encoding an appropriate G-protein (31).

The integration of molecular biological approaches into receptor pharmacology has been useful in elucidating receptor structure–function relationships. Construction of chimeric receptors (hybrid polypeptides composed of adjacent portions of two related receptor subtypes) has helped to identify functional domains of many individual receptors. Such studies have successfully identified ligand-binding domains and cytoplasmic domains mediating signal transduction of many receptors. Using *in vitro* mutagenesis to change single amino acids to create point-mutated receptors has defined critical residues involved in receptor–ligand interactions, conformational transitions, ion channel selectivity, signal transduction, and sites of receptor phosphorylation. Due to the lack of high-resolution structural data for most membrane receptors, this approach often requires making the assumption that overall receptor conformation and the equilibrium distribution between receptor states (resting, active, and desensitized) have not been altered in the mutated receptor. Nevertheless, this experimental approach has been the basis for a large body of molecular modeling aimed at understanding the specific functional groups on receptors that interact with drug ligands. The combination of such molecular pharmacological and modeling approaches can be expected to reveal the structural determinants of drug binding and novel allosteric sites on

the receptor protein for drug interaction. Ultimately, the ability to produce large amounts of purified receptor protein from recombinant expression systems will enable the use of physical methods (NMR, electron microscopy,

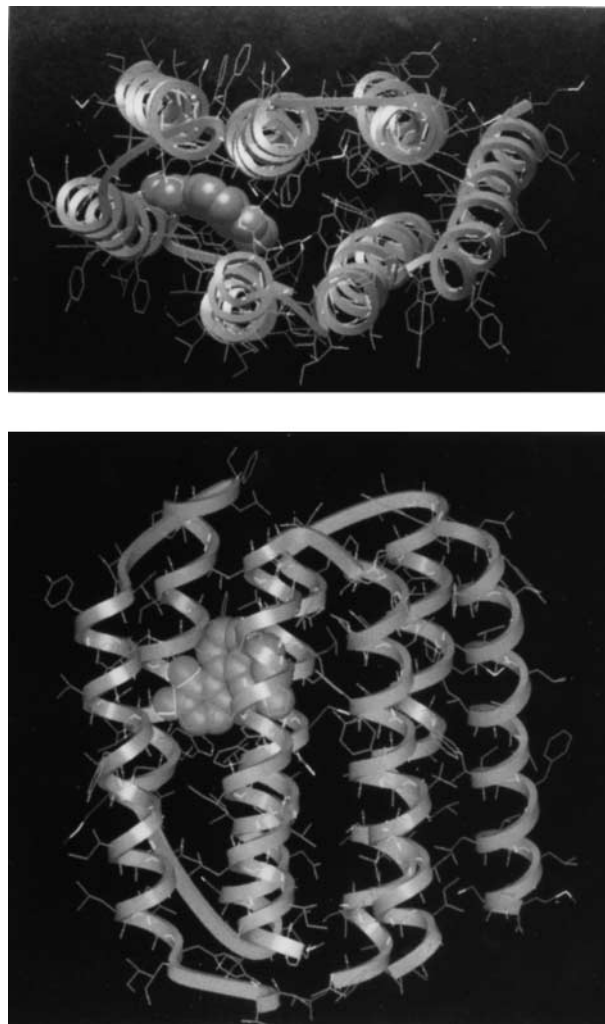


Fig. 4 A molecular model of the dopamine D_2 receptor with a ligand docked in the binding site. The model of the D_2 receptor transmembrane helices was constructed from the coordinates of the bacteriorhodopsin structure derived from two-dimensional electron diffraction experiments and is consistent with the projection structure for rhodopsin. The transmembrane helices are represented by a solid ribbon and the drug, apomorphine, is a space filling representation. The top view looking down the helical axis of the receptor clearly delineates the seven transmembrane helices that are the key structural motif for the GPCR superfamily. Some of the helices are inclined relative to the perpendicular to the membrane plane. The bottom view is in the plane of the membrane with the extracellular space at the top of the figure. (Adapted from Ref. 57.)

X-ray crystallography) to obtain high-resolution receptor structural data that will delineate ligand-binding sites and precisely define drug–receptor interactions. The recent elucidation of the high-resolution structure of rhodopsin is a major step towards the detailed understanding of the 3D structures of the GPCR superfamily, as shown in Fig. 4 for the dopamine D₂ receptor.

Drug Discovery Using Adopted Orphan Receptors

There are now numerous examples of orphan receptors that have been identified as novel types or subtypes of existing receptor families for which functional roles have been characterized as proven therapeutic drug targets. Many of these orphan receptors are of immediate interest as potential drug discovery targets since they represent novel receptor subtypes that extend existing receptor families that have members that are therapeutically important drug receptor targets. The close structural relatedness that is the molecular basis for organization of the receptor superfamilies has allowed a rational-based search for new members of the ligand-gated ion channel, G-protein coupled, receptor tyrosine-kinase, and nuclear receptor superfamilies. In general, discovery of new receptors (or subunits) has been accomplished by screening cDNA libraries, using sequences conserved among the receptor family at low stringency (homology screening). An alternative method has been to employ PCR amplification of human genomic DNA with degenerate oligonucleotides encoding conserved receptor domains.

Initial efforts to identify orphan receptors exploited PCR methodology to clone several novel orphan receptors belonging to the GPCR superfamily (32). These orphan receptors were subsequently shown to include the adenosine A₁ and A_{2a} receptors (33), a 5-HT_{1D} receptor (34, 5), and a central cannabinoid receptor (CB₁). PCR using degenerate primers was subsequently used to identify cDNAs encoding the NK₁, NK₂, dopamine D₁ and D₅ (36), histamine H₃, adenosine A₃(37), olfactory receptors, and numerous other GPCRs. The application of these approaches has been highly successful and has led to the cloning of a series of orphan receptors (or novel subunits) in each superfamily for which functions have not yet been assigned. To date, more than 40 additional members of the nuclear receptor superfamily have been cloned for which ligands have not been identified. In addition, the ligands for the many orphan members of the receptor–protein tyrosine phosphatase superfamily have also not yet been established. Numerous orphan receptor–tyrosine kinases belonging to the EPH family have been identified that likely function by transducing signals

initiated by direct cell–cell interaction (38). Many of the EPH receptors are specifically expressed in the nervous system (39), and these receptors have been implicated in the control of axon guidance, in regulating cell migration, and in defining compartments in the developing embryo.

Strategies for Ligand Identification for Orphan GPCR Receptors

Since GPCRs constitute the largest superfamily of receptors in man and these receptors also represent important drug targets that have been subject to intense biochemical and pharmacological studies, the following discussion will focus on the orphan receptors that belong to the GPCR superfamily. An explosion of the number of orphan GPCRs has been primarily due to advances in both cloning and sequencing technologies. Specifically, directed studies employing a combination of low stringency hybridization, polymerase chain reaction, an increasing database of ESTs (expressed sequence and tagged cDNA) and high throughput sequencing technologies have led to an exponential increase in the number of orphan GPCRs that await ligand identification.

The new paradigm for orphan receptor-based drug discovery is shown diagrammatically (Fig. 5). A first step toward ligand identification for a given orphan GPCR is to generate a full-length clone if the cDNA sequence is incomplete. To accomplish this, a specific tissue is identified that contains the orphan receptor gene and a full-length cDNA clone is typically generated using RT-PCR. Next, the cDNA is fully sequenced, and the sequence is compared with the sequence of known GPCRs through the use of genomic databases. If the identity of an orphan GPCR approaches 45% with a known receptor, it is likely that the orphan receptor and the highly homologous GPCR will share the same ligand (40). Successful application of this approach led to the identification of the ligand for the human 5-HT_{1D}(41), C3a (42), and CGRP (43) receptors. However, there are subfamilies of GPCRs that do not show this degree of similarity. Furthermore, many orphan GPCRs may reveal no significant identity with known GPCRs but will demonstrate high homology to other orphan receptors, leading to the identification of new GPCR families, as typified by the important discovery of the edg family of lysophospholipid receptors (44).

Most orphan receptors share a low degree of overall sequence homology, which may be in the range of 20–30% overall amino acid identity. In these cases, a useful step in ligand identification is to evaluate the both the tissue and cell type distribution of the orphan receptor mRNA through Northern blot analysis or RT-PCR. A very

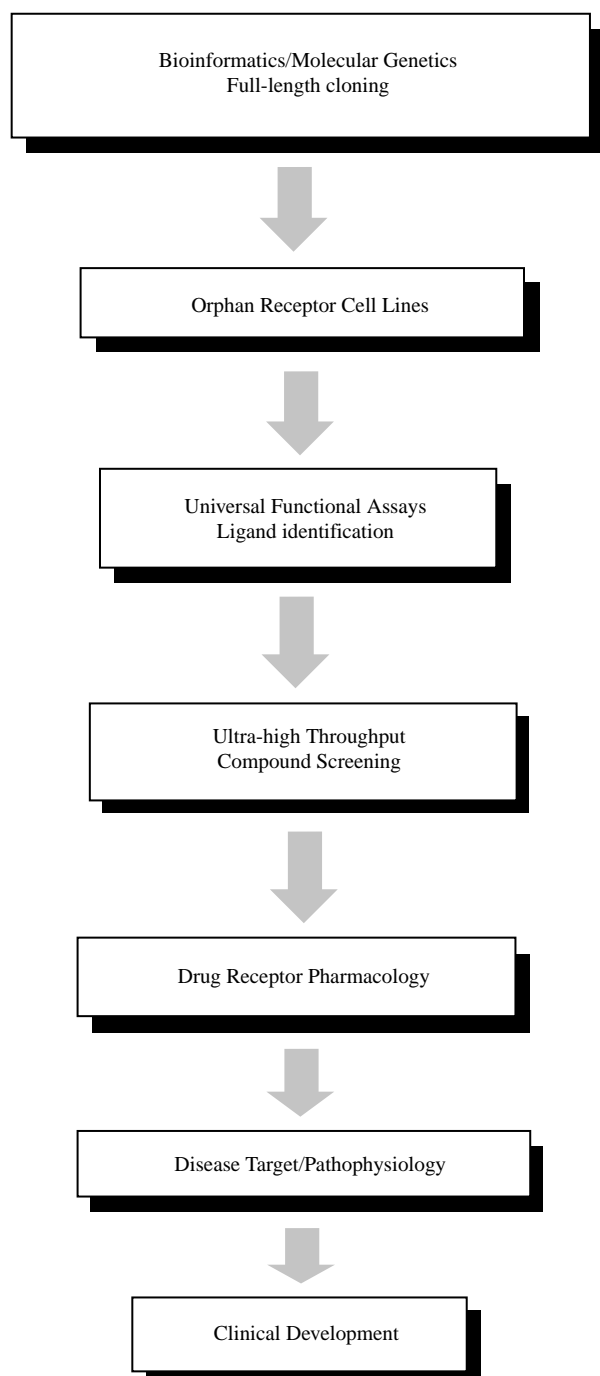


Fig. 5 Strategy for using orphan receptors as targets for drug discovery. The flow diagram illustrates the paradigm shift in receptor-based drug discovery, which is based upon a reverse molecular pharmacological strategy. The starting material for drug discovery is a sequence corresponding to an orphan receptor of unknown function.

restricted distribution pattern for an orphan GPCR may help provide clues concerning the physiological function of the receptor. Differential expression between normal and diseased tissue may indicate a role in the pathophysiology and suggest a potential therapeutic relevance for drugs targeted to the orphan receptor. Knowledge of the distribution of an orphan GPCR can also guide selection of cells or tissues to be used as sources of the endogenous cognate ligand for reverse pharmacology or ligand fishing studies. With the advent of chip technology, potentially thousands of orphan receptor cDNAs can be attached to a chip, thereby enabling the simultaneous evaluation of the expression levels under diverse pathological conditions.

Ligand Fishing for Orphan Receptors

Orphan receptors of interest may exhibit no significant homology with known GPCRs, and distribution studies may not provide any clues to the identity of the endogenous ligand. In this case, reverse pharmacology or ligand fishing studies are pursued (Fig. 5), especially for peptide receptors that often exhibit high affinity for their endogenous ligand. In this approach, the orphan receptor is used as bait to isolate the endogenous ligand from cell or tissue extracts suspected to contain the ligand (40, 44, 45). This approach has led to the discovery of novel endogenous peptide ligands orphanin FQ/nociceptin (46), orexin (47), prolactin-releasing peptide (48), melanin concentrating hormone (49), and APJ (50) as the cognate ligands for their respective orphan receptors. The specific high-affinity interaction of the endogenous ligand contained in extracts with the expressed orphan GPCR is measured in an appropriate functional second messenger assay. Once the active extract is identified, it can be fractionated, purified by HPLC, and the identity of the active component may be determined by analytical methods that may include peptide sequencing, mass spectroscopy, and NMR.

Expression of Orphan GPCRs

A critical step in ligand fishing for an orphan GPCR is to express the receptor in an appropriate cellular expression system. As previously discussed, choices may include mammalian cells, insect cells, bacteria, oocytes, or yeast. Initial considerations for choosing an expression system may focus on the type of functional assay to employ in screening procedures. A basic requirement for any given functional assay is the presence of the appropriate signal transduction machinery in the expression system. While it

is often not possible to predict specific coupling mechanisms and signal transduction pathways activated by an orphan receptor, several mammalian cells have been confirmed as suitable for human orphan receptor expression, including CHO and HEK 293 cells. All of these contain a full repertoire of signal transduction components required for measurement of most GPCR-mediated changes in second messenger systems (44). Another factor to be weighed in selecting an appropriate expression system is the potential interference of endogenous receptors present in the host cell line. This factor can be experimentally addressed through parallel screening of the parental cell line. Additionally, it is important to establish that the orphan receptor is expressed at sufficient densities and present on the membrane surface. Expression levels have been evaluated by epitope tagging of the receptor and measurement of receptor on the membrane surface by fluorescence activated cell sorting (FACS) or, alternatively, by Western blot analysis.

Functional Assays Used for Orphan Receptor Ligand Identification

For the full value of the receptor gene sequence to be realized, the function of the expressed orphan receptor, as well as its regulation and expression need to be elucidated. In many cases, exemplified by the nociceptin receptor, the orphan receptor will be activated by an as yet unknown endogenous ligand (transmitter, peptide, or hormone) or the endogenous ligand may be identified as a previously characterized molecule whose function has not been assigned. Strategies for identification of either the endogenous or surrogate ligands for an orphan receptor are frequently based upon utilization of high and ultrahigh throughput "agonist screening" assays in which ligand binding is coupled to a cellular functional response.

Functional assays based upon measurements of common cellular signal transduction pathways (i.e., mobilization of intracellular calcium, cAMP, transcription response elements) have been employed. Since many types of receptors (LGCRs, GPCRs, and RTKs) mediate cellular responses through elevations in intracellular calcium, real-time measurements of intracellular calcium, using fluorescent probes, have provided a rapid, generic assay for receptor activation. In addition, cell-based transcriptional activation assays represent a functional screening method that requires engineering of stable cell line(s) expressing the recombinant orphan linked to a luminescent or fluorescent-based reporter gene (i.e., luciferase or green fluorescent protein) under the

transcriptional control of a promoter element (i.e., cAMP response element (CRE), NFAT, or STAT binding elements). The advantages inherent in these one-step screening approaches include high detection sensitivity, compatibility with automation, and high throughput.

The first consideration for developing a functional assay for screening extracts or compounds to identify endogenous ligands or surrogate agonists for orphan GPCRs is to determine which G-protein(s) mediate coupling to the receptor. Certainly, this is difficult to determine unless the orphan receptor of interest contains a high level of homology to other known GPCRs activating well-characterized signal transduction pathways. In some cases, evaluation of extracts or compounds in both G_s , G_i and G_q -associated functional assays is necessary. Once the type of assay endpoint is chosen, there is a variety of high throughput and robust functional assays that can be utilized in ligand fishing. Advances in commercially available technology have enabled the development of ultrahigh throughput screening methods and systems that have led to the miniaturization of assays from 96-well plate formats to 1536-well formats that require only microliter volumes of reagents. Functional assays based upon one of several fluorescence detection techniques, including fluorescence resonance energy transfer, polarization, time-resolved fluorescence, and luminescence, now allow receptor-activated signaling events in live cells to be rapidly measured.

Reporter gene assays

Reporter gene assays have been widely used as high-throughput, cell-based functional assays for numerous GPCRs. In these assays, gene transcription is activated by a receptor-associated transduction event, such as increases in cellular cAMP or intracellular calcium. Reporter genes typically consist of a specific responsive element placed upstream of a minimal promoter, which together control the expression of the reporter gene, such as β -galactosidase luciferase. Commonly used response elements include the cAMP response element that is suitable for either G_s - and G_i -coupled receptors or a calcium-sensitive response element suitable for G_q -coupled receptors (51).

Another cell-based functional gene transcription assay has been developed by Tsien and co-workers that permits real-time receptor activation of gene expression to be measured in live cells. For this assay, mammalian cells are transfected with receptor and the gene for β -lactamase that acts as a reporter by hydrolyzing an exogenous substrate that is loaded intracellularly. The highly sensitive fluorescence assay permits clonal selection of single cells by flow cytometry and forms the basis for a high-throughput screen.

A bioluminescent reporter assay for G_q -coupled receptors has been described in which the receptor of interest is cotransfected with the calcium-sensitive protein aequorin. Agonist activation of receptor-mediated increases in intracellular calcium lead to activation of aequorin and subsequent luminescent responses (52). Due to the considerable amplification occurring after agonist activation of the GPCR, this assay can be significantly miniaturized and run in 96-, 384-, and 1536-well formats. Transient expression of apoaequorin in CHO cells and reconstitution with the co-factor coelenterazine resulted in a large, concentration-dependent agonist-mediated luminescent response following cotransfection with the endothelin ET_A , angiotensin AT_2 , thyrotropin-releasing hormone (TRH), and neurokinin NK_1 receptors, all of which interact predominantly with the G_{α_q} -like phosphoinositidase-linked G-proteins. To generate a system amenable for the study of agonist activity at virtually any G-protein-coupled receptor, the α -subunit of the receptor promiscuous G-protein $G_{\alpha_{16}}$ was either transiently or stably expressed in CHO cells together with apoaequorin. In cells expressing $G_{\alpha_{16}}$, but not in its absence, agonists at a series of receptors that normally interact with either G_{α_s} or G_{α_i} were now able to cause a luminescent response from mitochondrially targeted apoaequorin. In the case of the A_1 adenosine receptor, this response was clearly a result of activation of $G_{\alpha_{16}}$ and not a consequence of the release of the G_{α_i} -associated β/γ complex, as the luminescent response was unaffected by pertussis toxin treatment of the cells, whereas agonist-mediated inhibition of adenylyl cyclase activity was attenuated. These studies describe the use of coexpressed apoaequorin as a reporter for G-protein-coupled receptor-mediated calcium signaling. Furthermore, coexpression of $G_{\alpha_{16}}$ and apoaequorin provides a basis for a generic mammalian cell microplate assay for the assessment of agonist action at virtually any GPCR, including orphan receptors for which the physiological signal transduction mechanism may be unknown.

Use of Engineered Cells for Developing Cell-Based Functional Assays

Advances in molecular engineering have facilitated the design of recombinant cells that can be used to design universal assays for orphan receptors. As a tool to analyze the function of newly discovered orphan receptor genes, "promiscuous" G-proteins have been utilized to develop screens for orphan GPCRs that can be used to search for compounds that activate such receptors. For example, coexpression of the "universal adapter" $G_{\alpha_{16}}$ G-protein with an orphan GPCR of interest allows the activated

receptor to couple to increases in intracellular calcium, although the receptor may not couple to G_q in nonengineered cells. Another common approach to designing a universal functional ligand-screening system for any GPCRs is to construct and coexpress chimeric G-proteins with an orphan GPCR (53). Chimeric G-protein α -subunits have been constructed in which the backbone of either G_s , G_i , or G_q , is combined with small peptidic sequences corresponding to the C-terminus of G_s , G_i , and G_q . The C-terminal end of the chimera dictates receptor coupling and the G-protein backbone determines which effector system is activated. For example, short C-terminal portions of G_s or G_i can be combined with the backbone of G_q , resulting in an engineered G-protein chimera that can universally signal through increases in intracellular calcium or activate calcium sensitive reporter assays (53, 54). Conversely, chimeras can be constructed with the backbone of the G_s - α -subunit and the C-terminus of G_q to develop an assay system in which cAMP is measured. The design of universal assay systems for orphan GPCRs can greatly facilitate throughput for analysis of large numbers of candidate receptors.

Recently, a number of GPCRs have been reported to activate second messenger responses, even in the absence of their natural ligands or surrogate agonists. GPCRs that possess this ligand-independent activity are referred to as constitutively active GPCRs. Three different methods have been employed to produce constitutively activated receptors and include: 1) overexpression of the native receptor in heterologous expression systems, 2) overexpression of the appropriate G-protein that couples to the receptor, or 3) mutation of specific aminoacid residues of the receptor, typically in transmembrane regions or in intracellular loops of the receptor. A number of pharmaceutical companies are now utilizing constitutively activated orphan GPCRs for identification of either inverse agonists or agonists from small molecule libraries using functional ligand-independent screens. Thus, drug discovery for orphan GPCRs does not now require previous identification of the endogenous ligand or a surrogate agonist to develop a functional screen for drug discovery. Molecules discovered in these screens can be used not only as potential therapeutics but also as tools to further evaluate the physiological functions of the orphan GPCRs and their potential therapeutic relevance.

Screening for Endogenous and Surrogate Ligands

After a reliable and robust functional assay is developed for an expressed orphan receptor, screening can ensue.

Several approaches can be taken. First, cell and tissue extracts or biological fluids can be evaluated for activation of the orphan receptor. If an active extract is identified, it is further fractionated and then subjected to HPLC followed by further testing of fractions in the functional assay. After substantial purification, the identity of the active component can be elucidated by numerous analytical methods including mass spectroscopy and NMR. Another approach is to screen a collection of compounds with known pharmacological actions and biological compounds for which the mechanism of action is unknown. A third approach is to screen peptide or small molecule libraries. This can easily be accomplished rapidly due to advances in high-throughput screening and laboratory automation. Once the surrogate ligand is identified, clues as to the physiological function of the orphan receptor and possibly the endogenous ligand can be often obtained.

The identification of gene sequences is only the beginning of the process for the development of the useful small molecule therapeutic agent. Analysis of a stretch of only 10–25 amino acids from an orphan receptor sequence may be sufficient to identify whether the orphan of interest is homologous to either a known protein or a recognizable motif that corresponds to a ligand recognition or functional domain characteristic of a receptor superfamily. Based upon the level of observed sequence homology with known receptors, it may be possible to either deduce the class of ligand bound by the orphan receptor and/or postulate a mechanism of signal transduction. The systematic search for natural cognate ligands of orphan receptors is more difficult when the properties of such receptors are not well predicted by sequence comparisons to known receptors and/or the orphan receptors exhibit novel pharmacology.

As an example of the first strategy, the process leading to the characterization of the orphan ORL1 (Opioid Receptor-Like I) receptor, which has become known as the orphanin or nociceptin receptor (46), is of interest since it represents one model that can be readily adapted to other orphan receptors. The ORL1 receptor is a novel G-protein coupled receptor that is most closely related to the opioid receptors. Based upon substantial sequence identity of the ORL1 receptor with opioid receptors (50% overall, 65% within transmembrane domains), it was reasonable to hypothesize that the related orphan receptor would share signal transduction properties in common with the μ -, δ -, and κ -opioid receptor subtypes. Since opioid receptors are all negatively coupled to adenylyl cyclase, a stable recombinant CHO cell line expressing ORL1 was constructed for use in a functional screen, using untransfected CHO cells as a control. A survey of opiate ligands identified etorphine, a nonselective opiate agonist, as mediating inhibition of forskolin-induced accumulation

of cAMP, although its potency was found to be about three orders of magnitude less compared with other opioid receptors. However, additional efforts to characterize the pharmacology of the ORL1 receptor by analysis of agonist effects induced by endogenous opioid peptides (endorphins, enkephalins, and dynorphins) or other synthetic opioid ligands were not successful.

Based upon the structural homology of ORL1 with opioid receptors, in particular, the acidic extracellular loop 2 of the κ -opioid receptor, it was hypothesized that the endogenous ligand might be a peptide that resembled dynorphin. A biochemical fractionation procedure was used to isolate a pituitary peptide whose structure was identified as a 17 amino acid neuropeptide that shares many features in common with other opioid peptides and is now known as nociceptin. Isolation of the endogenous ligand was achieved through a functional assay on the basis of its ability to inhibit adenylyl cyclase in a stable recombinant cell line. The identity of nociceptin was confirmed through synthesis of a radiolabeled derivative that was found to bind in a saturable manner with high affinity and to be a potent and specific activator of the ORL1 receptor. In vivo activity of nociceptin induced hyperalgesia when administered intracerebroventricularly to mice, indicating the agonist of the ORL1 receptor appears to possess pronociceptive properties. The unique pharmacology, physiology, and brain distribution of this novel receptor make it an important target for drug discovery.

An alternative approach may be to employ an immobilized orphan receptor to capture the cognate ligand from biological extracts, fractions, and other combinatorial chemistry libraries. Orphan receptor screening efforts will be able to take advantage of large-scale synthetically produced random peptide, peptidomimetic, and combinatorial chemical libraries, as well as phage display libraries that are available as novel sources of chemical and structural diversity for discovery of potent and selective chemical entities for each of the receptors. Finding new natural product-based lead compounds by screening fermentation broths and extracts from plant and marine organisms may yield ligands with novel structures, as was found in the case of the subtype-selective endothelin receptor peptidic antagonists.

Among the GPCR superfamily, the discovery and characterization of the large number of novel receptor subtypes belonging to the chemokine, dopaminergic, serotonergic, somatostatin, and opioid receptor families provide examples of the tremendous impact of gene cloning. As a consequence of the original cloning of the orphan receptor sequence corresponding to the 5-HT_{1D} receptor in 1989 (55), the human receptor sites involved in the action of acute antimigraine drugs were subsequently

identified. Studies of these receptors have led to significant insights into the pathophysiology of migraine and the development of new antimigraine drugs, such as naratriptan and zolmitriptan, which were selected for development based upon their high affinity and selectivity for the human recombinant 5-HT_{1D} receptors. Clearly, the value inherent in orphan receptor characterization for drug discovery has been proven and will continue to lead to the discovery of substantial numbers of novel, human receptors designated as “orphans” that will provide novel targets for drug development.

Human Genome Sequencing—New Orphan Receptors and Subtypes

The most intensive research project in biomedicine is under way—the immense task of sequencing the human genome and identifying all of the expressed human genes. This effort is driving a paradigm shift in the fundamental approach to drug discovery as genomic data becomes the initiation point for the drug discovery process. Considerable progress has been made by the Human Genome Project as advances in cloning, mapping, and sequencing technologies have jointly contributed to an explosion in the volume of human sequence data. The rapidly approaching completion of the human genome sequence will lead to structure of all of the roughly 25,000–35,000 different human genes. To integrate and analyze the vast database of genomic data, the field of bioinformatics is providing powerful computational algorithms and methods for searching databases and complex data sets to allow molecular classification and extrapolate receptor structure and function from numerous new sequences. The impact of this effort will be an exponential increase in the number of novel orphan receptors requiring further study to define their potential role as targets for drug discovery. At the conclusion of this project, the identification and sequence of all receptor genes in the human genome will be available for study. Based upon estimates that 2–3% of the genome is likely to consist of receptors, this would suggest that the total number of receptors to be identified will be approximately 600–900. Yet another benefit of these genomic efforts will be the identification single disease-causing genes that are responsible for common or inherited diseases. Numerous examples exist for disease states that are associated with either deficient receptor responses, unregulated signaling function, or enhanced responses to neurotransmitter or hormonal signaling, indicating the biomedical importance for identifying the molecular etiology of a disease state and gaining access to disease-relevant receptor targets. The recent cloning of a Kaposi's sarcoma-associated herpesvirus (KSHV, or

human herpesvirus 8) genome fragment revealed an open reading frame encoding a putative GPCR that was homologous to human CXCR1 and CXCR2 receptors (56). Expression studies showed that it was indeed a bona fide signaling receptor that exhibited binding characteristics of a chemokine receptor, with affinity for a range of chemokines in the CXC and CC families. The receptor demonstrated constitutive (agonist-independent) activity in COS cells and stimulated cellular proliferation, making it a candidate viral oncogene.

Technological extensions based upon the human genome sequencing platform that enhance receptor-based drug discovery include development and construction of biochips containing DNA microarrays of gene sequences. These biochips enable high-throughput screening using fluorescent hybridization of total cDNA or mRNA libraries from particular cell types or rare tissues and permit quantitative examination of large numbers of specific receptor sequences expressed in normal and disease samples. The large-scale quantitative examination of large numbers of specific receptor sequences expressed in normal and disease samples is called expression profiling. Such differential gene expression studies can detect different levels and patterns of a receptor gene expression in different tissue states and can be used as a tool to decipher receptor gene function and potentially define a relationship to a particular disease state.

In summary, the application of molecular genetics to large-scale sequencing, analysis of DNA sequence information, and new experimental approaches for rapid functional analysis of receptor molecules from these sequences can be expected to lead to the identification of an abundance of novel receptors that will become molecular targets for the development of highly selective drugs. Indeed, gene-based discoveries are already providing an entirely new approach to the development of a novel drug pharmaceuticals. The parallel utilization of all receptor subtypes that comprise a family of human receptors in high-throughput primary screening programs has demonstrated the value of this approach for the identification of receptor-subtype specific drugs. This development in the field of drug discovery holds great promise since drugs with therapeutic activity that emerge from such screening programs should possess high molecular specificity.

The current challenge in receptor functional genomics will be to develop successful experimental strategies and new technologies for the identification of ligands that can activate orphan receptors or inhibit these receptors, should they exhibit constitutive, basal activity. Although progress has been made, the continued application of molecular biological approaches to receptor pharmacology coupled with advances in biophysical methods should serve as

tools that will enable a precise delineation of receptor binding sites and drug–receptor interactions at the molecular level. A more complete understanding the molecular events that transduce ligand binding into receptor activation will ultimately aid in the discovery of more selective agents for specific conformational states of existing receptor subtypes. Furthermore, the rapid elucidation of receptor–protein interaction partners and mapping intracellular signal transduction pathways associated with receptor signaling will accelerate understanding of novel receptor function. Finally, methods to localize and image the expression of receptor genes in living cells and animals can be expected to provide a foundation for novel pharmacological interventions targeted to specific cell types and tissues.

The application of many new technologies is synergistically accelerating the process of moving from novel receptor gene sequences to validated drug discovery receptor targets and ultimately, new drugs. The ongoing sequencing of the human genome and ensuing characterization of orphan receptors coupled with the precise molecular analysis of receptor structure, function, expression, and role of these receptors in disease processes will continue to reveal new drug receptor targets that will serve as a platform for much of the drug discovery enterprise.

REFERENCES

- Kenakin, T.P.; Bond, R.A.; Bonner, T.I. Definition of Pharmacological Receptors. *Pharmacol. Rev.* **1992**, *44*, 351–362.
- Alexander, S.; Peters, J. TIPS Receptor and Ion Channel Nomenclature Supplement. *Trans. Pharmacol. Sci.* Elsevier.
- Hulme, E.C.; Birdsall, N.J.; Buckley, N.J. Muscarinic Receptor Subtypes. *Annu. Rev. Pharmacol. Toxicol.* **1990**, *30*, 633–673.
- Changeux, J.P.; Devillers-Thiery, A.; Chemouilli, P. Acetylcholine Receptor: An Allosteric Protein. *Science* **1984**, *225*, 1335–1345.
- Bowery, N.G. GABA_B Receptor Pharmacology. *Annu. Rev. Pharmacol. Toxicol.* **1993**, *33*, 109–147.
- Burt, D.R.; Kamatchi, G.L. GABA_A Receptor Subtypes: from Pharmacology to Molecular Biology. *FASEB J.* **1991**, *5*, 2916–2923.
- Seeman, P.; Van Tol, H.H. Dopamine Receptor Pharmacology. *Trends Pharmacol. Sci.* **1994**, *15*, 264–270.
- Hamblin, M.W.; Metcalf, M.A. Primary Structure and Functional Characterization of a Human 5-HT_{1D}-Type Serotonin Receptor. *Mol. Pharmacol.* **1991**, *40*, 143–148.
- Adham, N.; Ellerbrock, B.; Hartig, P.; Weinshank, R.L.; Branchek, T. Receptor Reserve Masks Partial Agonist Activity of Drugs in a Cloned Rat 5-Hydroxytryptamine 1B Receptor Expression System. *Mol. Pharmacol.* **1993**, *43*, 427–433.
- Hill, P.; Hnilo, J.; Karla, M.; Bounds, S.; Herz, J.M. Cloning, Expression and Comparison of the Binding Characteristics of the Known Human Dopamine Receptors. *Adv. Neurolo.* **1996**, *69*, 41–52.
- Kallal, L.; Benovic, J.L. Using Green Fluorescent Proteins to Study G-Protein-Coupled Receptor Localization and Trafficking. *Trends Pharmacol. Sci.* **2000**, *21*, 175–180.
- Sine, S.M.; Taylor, P. The Relationship Between Agonist Occupation and the Permeability Response of the Cholinergic Receptor Revealed by Bound Cobra Alpha-Toxin. *J. Biol. Chem.* **1980**, *225*, 10144–10156.
- Herz, J.M.; Johnson, D.A.; Taylor, P. Interaction of Noncompetitive Inhibitors with the Acetylcholine Receptor. The Site Specificity and Spectroscopic Properties of Ethidium Binding. *J. Biol. Chem.* **1987**, *262*, 7238–7247.
- Sokoloff, P.; Andrieux, M.; Besancon, R. et al. Pharmacology of Human Dopamine D3 Receptor Expressed in a Mammalian Cell Line: Comparison with D2 Receptor. *Eur. J. Pharmacol.* **1992**, *225*, 331–337.
- Chio, C.L.; Drong, R.F.; Riley, D.T.; Gill, G.S.; Slightom, J.L.; Huff, R.M. D4 Dopamine Receptor-Mediated Signaling Events Determined in Transfected Chinese Hamster Ovary Cells. *J. Biol. Chem.* **1994**, *269*, 11813–11819.
- Lahti, R.A.; Evans, D.L.; Stratman, N.C.; Figur, L.M. Dopamine D4 Versus D2 Receptor Selectivity of Dopamine Receptor Antagonists: Possible Therapeutic Implications. *Eur. J. Pharmacol.* **1993**, *236*, 483–486.
- Chidiac, P.; Hebert, T.E.; Valiquette, M.; Dennis, M.; Bouvier, M. Inverse Agonist Activity of Beta-Adrenergic Antagonists. *Mol. Pharmacol.* **1994**, *45*, 490–499.
- Samama, P.; Pei, G.; Costa, T.; Cotecchia, S.; Lefkowitz, R.J. Negative Antagonists Promote an Inactive Conformation of the Beta 2-Adrenergic Receptor. *Mol. Pharmacol.* **1994**, *45*, 390–394.
- Oswald, R.E.; Bamberger, M.J.; McLaughlin, J.T. Mechanism of Phencyclidine Binding to the Acetylcholine Receptor from Torpedo Electropaque. *Mol. Pharmacol.* **1984**, *25*, 360–368.
- Herz, J.M.; Atherton, S.J. Steric Factors Limit Access to the Noncompetitive Inhibitor Site of the Nicotinic Acetylcholine Receptor. *Fluorescence Studies. Biophys. J.* **1992**, *62*, 74–76.
- Neubig, R.R.; Sklar, L.A. Subsecond Modulation of Formyl Peptide-linked Guanine Nucleotide-Binding Proteins by Guanosine 5'-O-(3-Thio)Triphosphate in Permeabilized Neutrophils. *Mol. Pharmacol.* **1993**, *43*, 734–740.
- Posner, R.G.; Fay, S.P.; Domalewski, M.D.; Sklar, L.A. Continuous Spectrofluorometric Analysis of Formyl Peptide Receptor Ternary Complex Interactions. *Mol. Pharmacol.* **1994**, *45*, 65–72.
- Gether, U.; Lin, S.; Kobilka, B.K. Fluorescent Labeling of Purified β 2-Adrenergic Receptor. Evidence for Ligand-specific Conformational Changes. *J. Biol. Chem.* **1995**, *270*, 28268–28275.
- Neher, E.; Steinbach, J.H. Local Anaesthetics Transiently Block Currents Through Single Acetylcholine-Receptor Channels. *J. Physiol.* **1978**, *227*, 153–176.
- Oksenberg, D.; Marsters, S.A.; O'Dowd, B.F.; et al. A Single Amino-acid Difference Confers Major Pharmacological Variation Between Human and Rodent 5-HT_{1B} Receptors. *Nature* **1992**, *360*, 161–163.

26. Burnashev, N.; Villarel, A.; Sakmann, B.; Jingami, H. Dimensions and Ion Selectivity of Recombination AMPA Kainate Receptor Channels and Their Dependence on Q/R Site Residues. *J. Physiol.* **1996**, *496*, 165–173.
27. Moser, B.; Barella, L.; Mattei, S. et al. Expression of Transcripts for Two Interleukin 8 Receptors in Human Phagocytes, Lymphocytes and Melanoma Cells. *Biochem. J.* **1993**, *294*, 285–292.
28. Goldman, D.; Deneris, E.; Luyten, W.; Kochhar, A.; Patrick, J.; Heinemann, S. Members of a Nicotinic Acetylcholine Receptor Gene Family are Expressed in Different Regions of the Mammalian Central Nervous System. *Cell* **1987**, *48*, 965–973.
29. Vallar, L.; Muca, C.; Magni, M.; et al. Differential Coupling of Dopaminergic D2 Receptors Expressed in Different Cell Types. Stimulation of Phosphatidylinositol 4,5-Bisphosphate Hydrolysis in LtK-fibroblasts, Hyperpolarization, and Cytosolic-Free Ca^{2+} Concentration Decrease in GH4C1 Cells. *J. Biol. Chem.* **1990**, *265*, 10320–10326.
30. Payette, P.; Gossard, F.; Whiteway, M. Dennis M. Expression and Pharmacological Characterization of the Human M1 Muscarinic Receptor in *Saccharomyces Cerevisiae*. *FEBS Lett.* **1990**, *266*, 21–25.
31. King, K.; Dohlman, H.G.; Thorner, J.; Caron, M.G.; Lefkowitz, R.J. Control of Yeast Mating Signal Transduction by a Mammalian β 2-Adrenergic Receptor and Gs Alpha Subunit. *Science* **1990**, *250*, 121–123.
32. Libert, F.; Vassart, G.; Parmentier, M. Current Developments in G-Protein-Coupled Receptors. *Curr. Opin. Cell Biol.* **1991**, *3*, 218–223.
33. Libert, F.; Schiffmann, S.N.; Lefort, A. et al. The Orphan Receptor CDNA RDC7 Encodes an A1 Adenosine Receptor. *EMBO J.* **1991**, *10*, 1677–1682.
34. Hamblin, M.W.; Metcalf, M.A. Primary Structure and Functional Characterization of a Human 5-HT1D-Type Serotonin Receptor. *Mol. Pharmacol.* **1991**, *40*, 143–148.
35. Zgombick, J.M.; Weinshank, R.L.; Macchi, M.; Schechter, L.E.; Branchek, T.A.; Hartig, P.R. Expression and Pharmacological Characterization of a Canine 5-Hydroxytryptamine 1D Receptor Subtype. *Mol. Pharmacol.* **1991**, *40*, 1036–1042.
36. Zhou, Q.Y.; Grandy, D.K.; Thambi, L. et al. Cloning and Expression of Human and Rat D1 Dopamine Receptors. *Nature* **1990**, *347*, 76–80.
37. Zhou, Q.Y.; Li, C.; Olah, M.E.; Johnson, R.A.; Stiles, G.L.; Civelli, O. Molecular Cloning and Characterization of an Adenosine Receptor: The A3 Adenosine Receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7432–7436.
38. Lindberg, R.A.; Hunter, T. cDNA Cloning and Characterization of ECK, an Epithelial Cell Receptor Protein-Tyrosine Kinase in the EPH/ELK Family of Protein Kinases. *Mol. Cell. Biol.* **1990**, *10*, 6316–6324.
39. Bartley, T.D.; Hunt, R.W.; Welcher, A.A. et al. B61 is a Ligand for the ECK Receptor Protein-Tyrosine Kinase. *Nature* **1994**, *368*, 558–560.
40. Marchese, A.; Sawzdargo, M.; Nguyen, T. et al. Discovery of Three Novel Orphan G-Protein-Coupled Receptors. *Genomics* **1999**, *56*, 12–21.
41. Maenhaut, C.; Van Sande, J.; Massart, C. et al. The Orphan Receptor CDNA RDC4 Encodes a 5-HT1D Serotonin Receptor. *Biochem. Biophys. Res. Commun.* **1991**, *180*, 1460–1468.
42. Ames, R.S.; Li, Y.; Sarau, H.M. et al. Molecular Cloning and Characterization of the Human Anaphylatoxin C3a Receptor. *J. Biol. Chem.* **1996**, *271*, 20231–20234.
43. Kapas, S.; Clark, A.J. Identification of an Orphan Receptor Gene as a Type 1 Calcitonin Gene-Related Peptide Receptor. *Biochem. Biophys. Res. Commun.* **1995**, *217*, 832–838.
44. Wilson, S.; Bergsma, D.J.; Chambers, J.K. et al. Orphan G-Protein-Coupled Receptors: The Next Generation of Drug Targets? *Br. J. Pharmacol.* **1998**, *125*, 1387–1392.
45. Civelli, O. Functional Genomics: The Search for Novel Neuro-transmitters and Neuropeptides. *FEBS Lett.* **1998**, *430*, 55–58.
46. Meunier, J.C.; Mollereau, C.; Toll, L. et al. Isolation and Structure of the Endogenous Agonist of Opioid Receptor-like ORL1 Receptor. *Nature* **1995**, *377*, 532–535.
47. Sakurai, T.; Amemiya, A.; Ishii, M. et al. Orexins and Orexin Receptors: A Family of Hypothalamic Neuropeptides and G-Protein-Coupled Receptors that Regulate Feeding Behavior. *Cell* **1998**, *92*, 573–585.
48. Hinuma, S.; Habata, Y.; Fujii, R. et al. A Prolactin-Releasing Peptide in the Brain. *Nature* **1998**, *393*, 272–276.
49. Chambers, J.; Ames, R.S.; Bergsma, D. et al. Melanin-Concentrating Hormone is the Cognate Ligand for the Orphan G-Protein-Coupled Receptor SLC-1. *Nature* **1999**, *400*, 261–265.
50. Tatemoto, K.; Hosoya, M.; Habata, Y. et al. Isolation and Characterization of a Novel Endogenous Peptide Ligand for the Human APJ Receptor. *Biochem. Biophys. Res. Commun.* **1998**, *251*, 471–476.
51. Colquhoun, D.; Hawkes, A.G. A Note on Correlations in Single Ion Channel Records. *Proc. R. Soc. London, Ser. B. Biol. Sci.* **1987**, *230*, 15–52.
52. Stables, J.; Green, A.; Marshall, F. et al. A Bioluminescent Assay for Agonist Activity At Potentially Any G-Protein-Coupled Receptor. *Anal. Biochem.* **1997**, *252*, 115–126.
53. Milligan, G.; Rees, S. Chimeric G Alpha Proteins: Their Potential Use in Drug Discovery. *Trends Pharmacol. Sci.* **1999**, *20*, 118–124.
54. Conklin, B.R.; Herzmark, P.; Ishida, S. et al. Carboxyl-Terminal Mutations of Gq Alpha and Gs Alpha that Alter the Fidelity of Receptor Activation. *Mol. Pharmacol.* **1996**, *50*, 885–890.
55. Maenhaut, C.; Van Sande, J.; Massart, C. et al. The Orphan Receptor CDNA RDC4 Encodes a 5-HT1D Serotonin Receptor. *Biochem. Biophys. Res. Commun.* **1991**, *180*, 1460–1468.
56. Arvanitakis, L.; Geras-Raaka, E.; Varma, A.; Gershengorn, M.C.; Cesarman, E. Human Herpesvirus KSHV Encodes a Constitutively Active G-Protein-Coupled Receptor Linked to Cell Proliferation. *Nature* **1997**, *385*, 347–350.
57. Teeter, M.; Fromanitz, M.; Stec, B.; Du Ravid, J. Homology Modelling of the Dopamine D2 Receptor and Its Testing by Docking of Agonist and T Ricyclic Antagonists. *J. Mod. Chem.* **1994**, *37*, 2874–2878.

Rheology of Pharmaceutical Systems

Fridrun Podczeck

University of London, London, U.K.

INTRODUCTION

The American Society of Rheology, which was founded in 1929, has defined “rheology” as the science of deformation and flow of matter. Deformation describes the change of matter in terms of shape or volume, or both. If the changes due to deformation are at least partly nonreversible, then matter flows. The full spectrum of possible types of deformation of liquid or solid matter is illustrated in Fig. 1. The gaseous state is usually not part of rheological investigations.

Typical modern rheological investigations do not consider Newtonian liquids in the sense of the Navier–Stokes model or fully elastic solids, but are concerned with the behavior of materials between these two extremes.^[1] In the pharmaceutical sciences, however, Newtonian behavior and its characterization are of interest, and methodology and equipment are described in various pharmacopeias to provide standards for the quality of pharmaceutical materials.

There is an increasing interest in rheological methods in the medical and biological sciences. A change of the rheological behavior of certain body fluids such as mucus,^[2,3] saliva,^[4] blood,^[5–7] or synovial fluid could be used as an indicator of the severeness of a disease or the effectiveness of drug substances used to treat certain medical conditions such as mucoviscidosis. The rheological properties of raw materials and formulations can cause problems during the manufacture of dosage forms in the pharmaceutical, cosmetic, or food industry. For example, more viscous materials require larger amounts of energy during mixing. The viscosity might be reduced by the application of heat, which could reduce the mixing time and improve product homogeneity. However, gel–sol transformations are not always reversible, and in some instances, the heat development due to shear might already destroy the product. Transport and handling might also result in a change in viscosity of the product.^[8] A decrease of the viscosity of suspensions could hence result in increased sedimentation and, for example, hard shell capsules containing a thixotropic filling might start to leak.

The rheological properties of pharmaceutical products are also important with respect to patient compliance. Stiff creams might complicate their application or could cause pain, and hence the patient will be more reluctant to use his medicine. Suspensions and emulsions should flow out of the bottle to allow easy metering and, for example, toothpaste should be squeezed out of a tube easily, yet should form a string without leakage of fluid. Several injections rely on gel–sol transformation, for example, those containing aluminum stearate. Aluminum stearate is added to produce a stiff gel for prolonged-release delivery of antibiotics via intramuscular injection. To ease the pain involved, the gel should form a sol solely on shaking, yet after injection gelation should occur quickly to maintain the prolonged-release principle. Hence, the knowledge of rheological techniques can play an important role in pharmaceutical development and manufacture.

DEFINITIONS AND TERMINOLOGY

Stress and Strain

Deformation is the result of a force acting on or within a body, yet its extent depends not on the force itself, but on the magnitude of force per unit area, i.e., the stress (unit: Pa or N m^{-2}). Forces acting within a body are the result of gravity or inertia and are effective throughout the whole body. Surface forces act on defined points along the periphery of the material.

As a consequence of the stress applied, the body will change its shape, and as a result, there will be a change in length of the body, i.e., strain. The relationships between stress and strain are illustrated in Fig. 2.

In many of the rheological test methods, the sample is subjected to “simple shear.” The best example for illustration is that of a pack of cards (Fig. 3). If a shear stress τ is applied to the top (“reference”) card of the pack, the displacement of each individual card of the pack is proportional to its distance from the reference card. The larger the distance is, the less is the card displaced. The relative displacement of two cards divided by their separation ($\delta l/l$) is called “simple shear strain.”

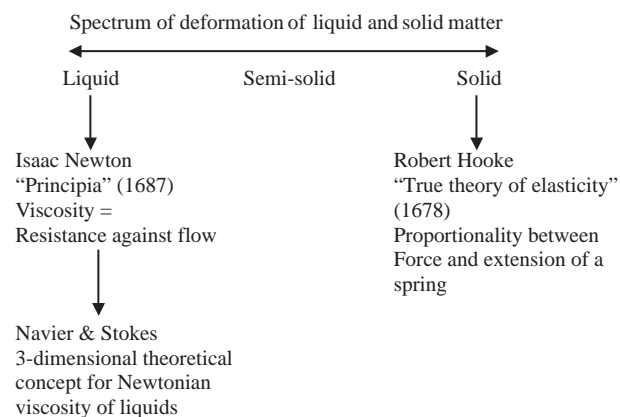


Fig. 1 Spectrum of deformation of liquid (viscous) and solid (elastic) matter first described by Newton and Hooke, and further developed by Navier and Stokes. Semisolid matter behaves viscoelastic, i.e., viscous at low strain rates and elastic at high strain rates.

Rate of Strain and Shear Rate

The rate of strain ("strain rate") considers the magnitude of the changes in length as a function of time (unit: sec^{-1}). For the tensional strain, the strain rate is hence defined as $\dot{\epsilon} = d\epsilon/dt$, whereas the shear strain rate ("shear rate") is defined as $\dot{\gamma} = d\gamma/dt$ (for symbols refer to Fig. 2). In simple shear, the shear strain rate can be explained as the velocity gradient occurring in a sample during flow.

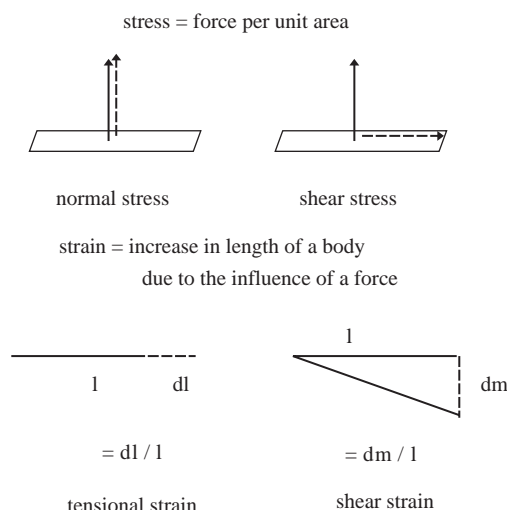


Fig. 2 Definitions and relationships between stress and strain: the vectors of normal force (solid line) and normal stress (dashed line) are in one and the same direction, whereas these vectors are orientated perpendicular in shear.

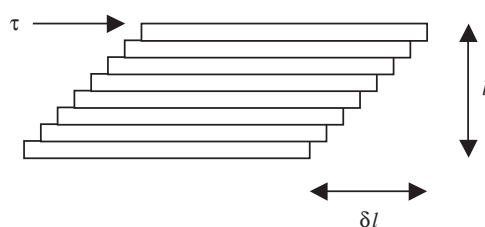


Fig. 3 Simple shear illustrated on a pack of cards: τ = shear stress, l = distance between two cards, δl = displacement of two cards relative to each other.

However, in cases where the flow lines are curved, the velocity gradient is not equal to the shear strain rate. Such complicated flow can happen in concentric cylinder viscometers (e.g., "cup and bob viscometer"), in particular at higher angular velocities and with the inner cylinder rotating ("Searle principle").^[9] Centrifugal forces stabilize the flow when the outer cylinder rotates ("Couette principle"), and turbulent flow is hence less likely.^[10] In cone-plate viscometers, deviations from laminar flow are sufficiently suppressed when using small cone-plate angles^[11] as is done in most practical instruments.

Modulus of Elasticity and Compliance

In the ideal case of a Hookean body, the relationship between stress and strain is fully linear, and the body returns to its original shape and size, after the stress applied has been relieved. The proportionality between stress and strain is quantified by the modulus of elasticity (unit: Pa). The proportionality factor under conditions of normal stress is called modulus of elasticity in tension or "Young's modulus" (E), whereas that in pure shear is called modulus of elasticity in shear or "modulus of rigidity" (G). The relationships between E , G , shear stress, and strain are defined by:^[12]

$$G = \frac{E}{2(1 + \nu)} \quad (1)$$

$$\gamma = \frac{\tau}{G} \quad (2)$$

where ν is the Poisson's ratio. For soft solids and liquids, where the stress needed to change the shape is considerably smaller than the stress required to change the volume, Poisson's ratio is approximately 0.5.^[10] Hence, $E \approx 3G$.

The moduli of elasticity in tension and shear are important for viscoelasticity measurements, where the stress applied is small.

The compliance is defined as the reciprocal of the modulus of elasticity (unit: Pa^{-1}). Hence, shear compliance is defined as $J = 1/G$, while tensile compliance is defined by $D = 1/E$.

Viscosity

The flow of liquids or semisolids is described by viscosity, or, more precisely, by shear viscosity η (unit: Pa sec). The viscosity defines the resistance of the material against flow. Viscosity is not a coefficient, because it is a function of the shear strain rate $\dot{\gamma}$ [$\eta = f(\dot{\gamma})$]. In the classical fluid mechanics, the dynamic viscosity is obtained using a viscometer. (A viscometer is a rheometer, i.e., an instrument for the measurement of rheological properties, limited to measuring the viscosity only.) The ratio between dynamic viscosity and material density is termed “kinematic viscosity” ($\nu = \eta/\rho$; unit: $\text{m}^2 \text{sec}^{-1}$). The kinematic viscosity provides a normalized value and characterizes the flow velocity. In the engineering sciences this value is, for example, used to calculate the throughput of liquids in pipes.

The reciprocal of the viscosity is called “fluidity” (unit: $\text{Pa}^{-1} \text{sec}^{-1}$) and is an indicator of the mobility of the material.

TYPES OF RHEOLOGICAL BEHAVIOR IN SIMPLE SHEAR

Newtonian Flow

The relationships between stress, strain, and viscosity are usually depicted in the so-called rheograms. In the pharmaceutical sciences, typical flow curves are presented, i.e., $\tau = f(\dot{\gamma})$. In the engineering sciences, the viscosity is usually drawn as a function of the shear stress [$\eta = f(\tau)$]. This is sensible as most viscometers control the shear stress applied rather than the shear strain rate. However, the entity of interest is the viscosity as a function of the shear strain rate [$\eta = f(\dot{\gamma})$].

In the following text, descriptions such as “liquid,” “fluid,” or “semisolid” have been avoided where possible to stress similarities and differences in the rheological behavior of matter.

Newtonian behavior can be observed only for ideally viscous bodies. The flow curve shows a direct proportionality between shear stress and shear strain rate (Fig. 4a) with the straight line going through zero. The viscosity remains constant over the complete range of shear stresses applied and is independent of the shear strain rate (Fig. 5a). The stress in the material goes back to zero immediately

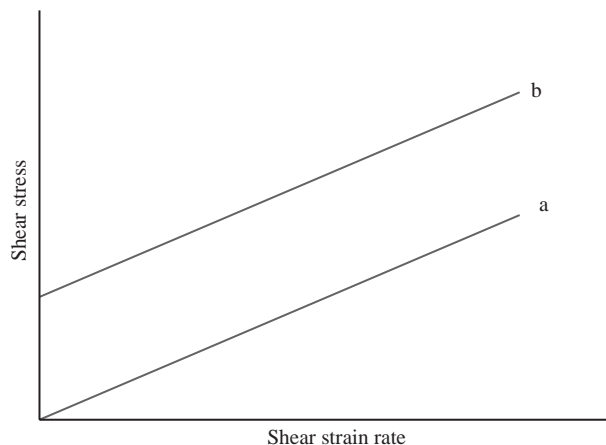


Fig. 4 Typical rheogram of a Newtonian (a) and a Bingham body (b).

when shearing has stopped, and repeated shearing always results in the same value of viscosity. Even after extended shearing times, the viscosity of Newtonian bodies does not change. Hence, for Newtonian bodies, the viscosity is a material constant at a given temperature and pressure. Many liquids such as water, glycerol, fatty oils, or organic solvents behave like Newtonian bodies. Also molten Vaseline behaves Newtonian, whereas at room temperature Vaseline is classified as a non-Newtonian body. While many liquid materials behave ideally viscous, their mixtures often show non-Newtonian behavior.

Non-Newtonian Flow

Shear thinning (“pseudoplastic flow”)

Shear thinning materials show a decrease in viscosity with increasing shear strain rate. A typical flow curve is

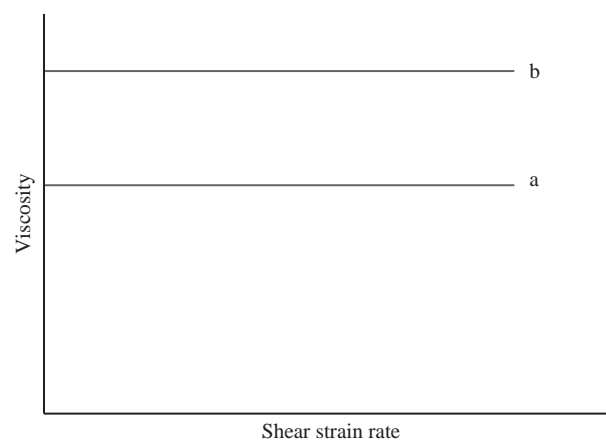


Fig. 5 Viscosity as a function of shear strain rate for a Newtonian (a) and a Bingham body (b).

presented in Fig. 6a, and the viscosity as a function of shear strain rate is depicted in Fig. 7a. The curves indicate that at very low shear stress values/shear strain rates and again at very high shear stress values/shear strain rates, the viscosity reaches a constant value. These two regions are often referred to as lower (i.e., lower shear strain rate) and upper (i.e., higher shear strain rate) Newtonian regions. The viscosity representing the lower Newtonian region is also known as the “zero-shear viscosity.” The upper Newtonian region cannot be obtained experimentally in some instances due to the limitations of some older equipment.

Shear thinning is often observed in solution containing thread-like macromolecules such as celluloses or unbranched polymers. Here the stress applied causes the macromolecules to align in the direction of the shear stress vector. This results in a reduced resistance to flow, and hence in a decrease in viscosity. After the orientation process is complete, the solution shows Newtonian flow.

Mathematical models, which can predict the shape of a flow curve of a shear thinning material including lower and upper Newtonian regions, require at least four parameters. The Cross model^[13] is one such model:

$$\frac{\eta - \eta_\infty}{\eta_0 - \eta_\infty} = \frac{1}{[1 + (K\dot{\gamma})^m]} \quad (3)$$

where η_0 and η_∞ are the asymptotic values of viscosity in the lower and upper Newtonian region, K is a constant with the dimension of time describing the steepness of the curve, and m is a dimensionless constant describing the overall curvature. Examples of a successful application of the Cross model to aqueous solutions of macro-

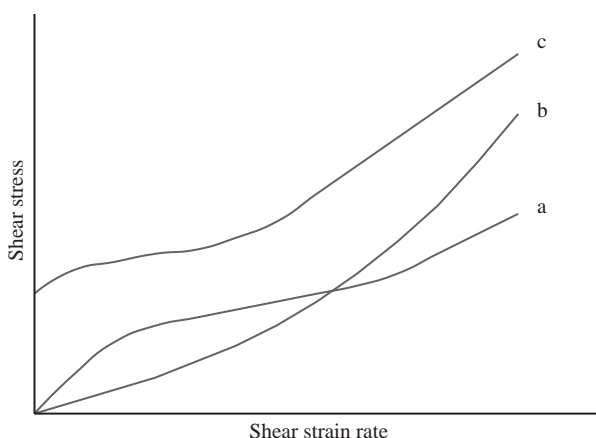


Fig. 6 Typical rheogram for a shear thinning (a), shear thickening (b), and a Casson body (c).

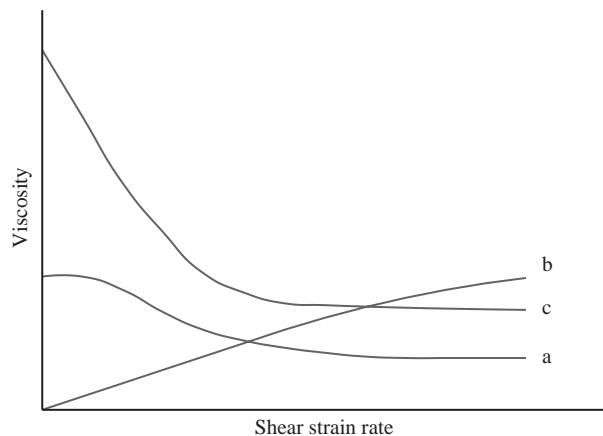


Fig. 7 Viscosity as a function of the shear strain rate for a shear thinning (a), shear thickening (b), and a Casson body (c).

molecules can be found by Boger^[14] and Whitcomb and Macosko.^[15]

Carreau^[16] devised another general equation to model the shear thinning behavior, which has become a popular alternative to the Cross model:

$$\frac{\eta - \eta_\infty}{\eta_0 - \eta_\infty} = \frac{1}{[1 + (k_1\dot{\gamma})^2]^{m_1/2}} \quad (4)$$

The constants k_1 and m_1 are of similar importance for the shape of the flow curves as are K and m in the Cross model.

In acknowledging that the upper Newtonian region is sometimes not observed due to experimental difficulties, the Cross and the Carreau models reduce to Eqs. 5 and 6, respectively:

$$\frac{\eta}{\eta_0} = \frac{1}{1 + (K\dot{\gamma})^m} \quad (5)$$

$$\eta = \frac{\eta_0}{[1 + (k_1\dot{\gamma})^2]^{m_1}} \quad (6)$$

Similarly, the two models could be reduced by neglecting the lower Newtonian region. This has been done by Sisko^[17] for the Cross model in order to account for the rheological properties of grease in bearings:

$$\eta = \eta_\infty + \frac{\eta_0}{(K\dot{\gamma})^m} \quad (7)$$

Other practical applications for the Sisko model have been reported for yogurt^[18] and polymer liquid crystals.^[19]

Considering that in many cases the viscosity measured is significantly smaller than that of the lower Newtonian region ($\eta \ll \eta_0$), and significantly larger than that of the upper Newtonian region ($\eta \gg \eta_\infty$), the Cross model can be further simplified:

$$\eta = \frac{\eta_0}{(K\dot{\gamma})^m} \quad (8)$$

It is standard practice to redefine the constants so as to $m = 1 - n$ and $\eta_0/K_m = K_2$.^[1] Using these definitions in Eq. 7 provides the so-called “powder law” model:

$$\eta = K_2 \dot{\gamma}^{n-1} \quad (9)$$

where n is the power law index, and K_2 is the consistency. The latter has the units of Pa sec^n and is hence difficult to interpret in physical terms. The larger the values of K_2 are, the stiffer the materials appear at rest. The value of n describes the curvature of the flow curve. Values of n below unity are found for shear thinning non-Newtonian bodies, whereas a value of 1 would be found for Newtonian flow. It should also be noted that Eq. 9 only describes the shear thinning region of a flow curve without considering the possible existence of a lower or upper Newtonian region.

Shear thickening (“dilatant flow”)

Shear thickening materials show an increase in viscosity with increasing shear strain rate. An idealized flow curve is presented in Fig. 6b, and the viscosity as a function of shear strain rate is depicted in Fig. 7b. The shear thinning region usually extends only over about one decade of shear rate (power law index $n > 1$) in contrast to shear thinning, which usually covers at least two or three decades. Also, in many cases, shear thickening is preceded by a short phase of shear thinning at low shear strain rates.^[1,20]

Shear thickening is typical for the flow of pastes and suspensions, when these have high concentrations of solid particles. However, the particle size needs to be below $50 \mu\text{m}$. The minimum solid concentration is usually quoted to be more than 50%, and suspensions show shear thickening if they are deflocculated. At rest and low shear the particles are covered with a liquid film coating acting as a lubricant. At low shear strain rates this liquid flows showing shear thinning. However, the liquid layer will be disrupted above a critical stress. Solid particles come into close interparticulate contact, which results in friction and hence in an increase in viscosity, i.e., shear thickening. Hoffman^[21] has studied this effect in more detail using an optical diffraction system.

Complex Flow Behavior

Bingham bodies

Newtonian flow, shear thinning, and thickening represent comparatively simple flow patterns. The description of flow becomes more complex, if these flow patterns occur in combination with a yield stress, i.e., a minimum stress required for flow to be visible in an adequately short time span. The observation of a yield stress is mostly related to the existence of a defined inner structure such as that of Vaseline or cubic liquid crystals.

Doubt has been cast on the existence of a true yield stress a long time ago, and it is usually Professor Reiner, who is quoted to have said “everything flows if you wait long enough,” some 60 years ago. Barnes and Walters^[22] found an apparently well-defined yield stress for PVA latex adhesives employing a shear strain rate-controlled rheometer. However, when they lowered the mean shear strain rate, the yield stress decreased proportionally. Finally, when they measured these materials at extremely low shear strain rates using a stress-controlled rheometer, Newtonian flow and absence of a yield stress was observed. Hence, it might be better to quote an “apparent” yield stress valid for defined experimental conditions only.

The Bingham body model describes materials with an apparent yield strength above which Newtonian flow is observed. This is illustrated in Figs. 4b and 5b, which show a typical flow curve and viscosity as a function of shear strain rate, respectively.

Flocculated suspensions often have high viscosities at low shear strain rates, which give the impression of the existence of a yield stress.^[23] The nearly Newtonian flow observed is due to a dramatic reduction in the effective number of particles as a result of the particle association process.

Mathematically, the Bingham body can be described as:

$$\tau = \tau_Y + \eta_P \dot{\gamma} \quad (10)$$

where τ_Y is the apparent yield stress, and η_P is, for historical reasons, called the “plastic” viscosity.

Casson bodies

The Casson body model^[24] is a semiempirical model used to describe shear thinning behavior after an apparent yield stress has been exceeded. It is, in fact, a combination of the apparent yield stress with the power law model, whereby

the power law index n is arbitrarily set to 0.5:

$$\tau^{0.5} = \tau_Y^{0.5} + k_2 \dot{\gamma}^{0.5} \quad (11)$$

An example of such a flow curve is presented in Fig. 6c, and viscosity as a function of shear strain rate is depicted in Fig. 7c. Many pharmaceutical ointments and creams show a similar shape of the flow curve with an extended upper Newtonian region. In these cases, an extrapolation of the linear portion of the flow curve to zero shear strain rate in order to obtain a “dynamic yield stress” is often utilized.^[25]

A more general mathematical description of shear thinning materials comprising an apparent yield stress is given by the Herschel–Bulkley model:^[26]

$$\tau = \tau_Y + k_2 \dot{\gamma}^n \quad (12)$$

which again is a simple extension of the power law model. For $n = 1$ Eq. 12 reduces to Eq. 10. Hence, power law and Bingham model are two extremes of the Herschel–Bulkley model.

Influence Factors on the Viscosity

Pressure

The viscosity of Newtonian and non-Newtonian materials increases exponentially with an increase in pressure. However, these changes are extremely small, and under atmospheric conditions of ± 1 bar, they are hardly detectable. Hence, the pressure is normally not controlled during rheological measurements. In some circumstances, the pressure exerted on, for example, oils and lubricants, can take up values in excess of 1 GPa (oil rigs, lubricants in gears), and the increase in viscosity is substantial. In such applications, it is thus required to consider the pressure as a factor when studying their rheological properties.

Temperature

An increase in temperature is accompanied by a decrease in viscosity. For Newtonian materials, this relationship can be approximated to the Arrhenius equation. To obtain an accuracy of $\pm 1\%$ in the measurements of the viscosity of water requires a temperature control of $\pm 0.3^\circ\text{C}$.^[1] The temperature dependence increases with an increase in viscosity. Thus, the demands on temperature control are even higher for more viscous materials. The shearing of the samples itself also generates heat. To ensure that the heat is not effective in altering the

temperature of the sample and equipment, the heat has to be removed quickly.

Shear time

Pharmaceutical preparations often provide lyogels with a colloid-like structure. For example, polymers with unbranched linear structures form matrices with parallel arrangement of individual chains. The van der Waals interactions between these chains lead to a higher degree of order. Areas of high order are separated from each other by amorphous areas. Laminar or sphero-colloidal matrices are formed, for example, from Bentonite and colloidal silicon dioxide, respectively. In all these cases, the flow elements will not be able to adapt to an applied stress instantaneously, and shear thinning will hence result. At the end of shear, the original structure may be completely destroyed. At rest, Brownian motion will lead to a rebuilding of the inner structure in many cases, but in some instances, the rheodestruction process is irreversible. The observation of a gradual decrease in viscosity with shear followed by a gradual recovery of the original structure is called “thixotropy” and is usually related to shear thinning materials. For shear thickening samples, a similar hysteresis can be observed in some instances, which are referred to as “negative thixotropy” or “antithixotropy.” An example of negative thixotropy is polyethylene glycol containing semisolid formulations.

Thixotropic behavior can be detected in two different ways. In the “step test” the material is sheared initially at a constant high shear strain rate, which is, after a defined time span has elapsed, reduced to a very small value. The shear stress recorded during the high shear phase decreases in a shear thinning fashion, indicating structural breakdown. The inner structure is rebuilt during the low shear period and results in a gradual increase of the shear stress recorded. Alternatively, the “loop test” can be employed. Typical flow curves obtained from loop tests on thixotropic and negative thixotropic materials are illustrated in Fig. 8. If highly elastic colloidal matrices are subjected to step or loop tests, viscoelastic contributions to the changes in stress have to be taken into account.^[10]

LINEAR VISCOELASTICITY

Definition of Viscoelasticity

The particular response of a sample to applied stress or shear strain rate depends on the time scale of the experiment. If the shear strain rate is kept low, many materials appear to behave viscous, whereas at high shear

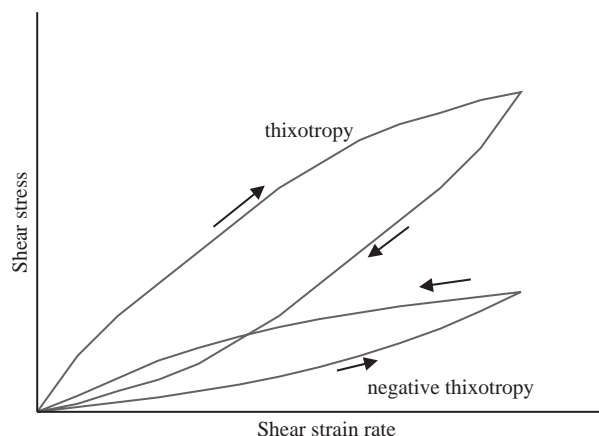


Fig. 8 Typical flow curves for thixotropic and negative thixotropic materials as obtained in a loop test.

strain rates, materials might behave rather elastically. The simultaneous existence of viscous and elastic properties in a material is called viscoelasticity, and one can assume that all real materials are viscoelastic in nature.

One very impressive example of a viscoelastic material is Silly Putty® (“bouncy putty”), which is a silicon-based polymer. If dropped onto a surface, it bounces back higher than a rubber ball, yet under light pressure, the material can be flattened with ease and remains in its new shape. If placed on the edge of, for example, a table, the material will slowly creep over the edge like liquid due to the influence of the force of gravity. Most working materials, however, do not behave in these extremes.

Linear and Nonlinear Viscoelastic Behavior

Materials can show linear and nonlinear viscoelastic behavior. If the response of the sample (e.g., shear strain rate) is proportional to the strength of the defined signal (e.g., shear stress), i.e., if the superposition principle applies, then the measurements were undertaken in the linear viscoelastic range. For example, the increase in shear stress by a factor of two will double the shear strain rate. All differential equations (see, for example, Eq. 13) are linear. The constants in these equations, such as viscosity or modulus of rigidity, will not change when the experimental parameters are varied. As a consequence, the range in which the experimental variables can be modified is usually quite small. It is important that the experimenter checks that the test variables indeed lie in the linear viscoelastic region. If this is achieved, the quality control of materials on the basis of viscoelastic properties is much more reproducible than the use of simple viscosity measurements. Nonlinear viscoelasticity experiments are

more difficult to model and hence rarely used compared to linear viscoelasticity models.

Mechanical Modeling

A general differential equation for linear viscoelasticity is:^[1]

$$\left(1 + \alpha_1 \frac{\partial}{\partial t} + \alpha_2 \frac{\partial^2}{\partial t^2} + \cdots + \alpha_n \frac{\partial^n}{\partial t^n}\right) \tau = \left(\beta_0 + \beta_1 \frac{\partial}{\partial t} + \beta_2 \frac{\partial^2}{\partial t^2} + \cdots + \beta_m \frac{\partial^m}{\partial t^m}\right) \gamma \quad (13)$$

where α_i ($i = 1, n$) and β_j ($j = 0, m$) are the constants, and τ and γ the shear stress and shear strain, respectively.

In order to derive some simple linear viscoelastic models, it is necessary to introduce the mechanical equivalents for a Newtonian and a Hookean body. First, it is assumed that β_0 is the only nonzero constant in Eq. 13:

$$\tau = \beta_0 \gamma \quad (14)$$

This equation describes Hookean elasticity, and $\beta_0 = G$ (G is the modulus of rigidity). In Fig. 9, the classical mechanical spring model representing Eq. 14 is illustrated. If, however, it is assumed that β_1 is the only nonzero constant in Eq. 13, then:

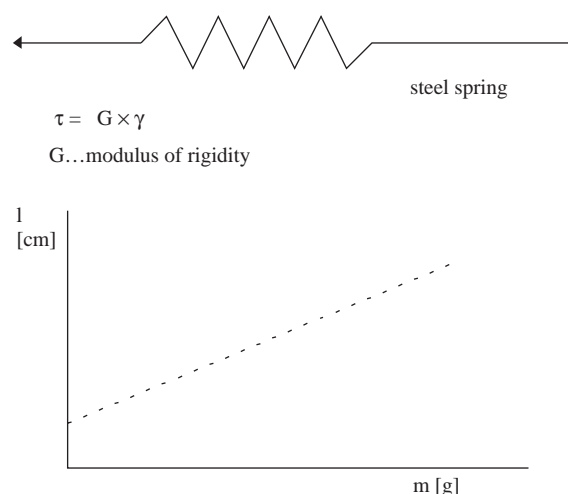


Fig. 9 Steel spring as the mechanical model for an ideal Hookean body: the length of the spring increases proportionally to the force applied, which is here represented by a weight that stretches the spring.

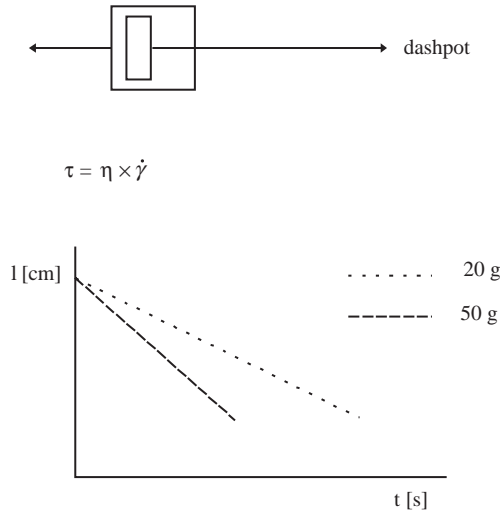


Fig. 10 Dashpot as the mechanical model of an ideal viscous material: the dashpot flows to similar extent but in a different time span depending on the force applied.

$$\tau = \beta_1 \frac{\partial \gamma}{\partial t} = \beta_1 \dot{\gamma} \quad (15)$$

The value of β_1 is equal to the viscosity η , and Eq. 15 thus describes the Newtonian body. The mechanical model of a Newtonian body is the “dashpot,” which is illustrated in Fig. 10. The rate of extension of the dashpot depends on the stress exerted, and if the dashpot becomes stress free at any time, it will remain in its current state of extension.

If it is now assumed that all constants except $\beta_0 (= G)$ and $\beta_1 (= \eta)$ are zero, the so-called Kelvin model results:

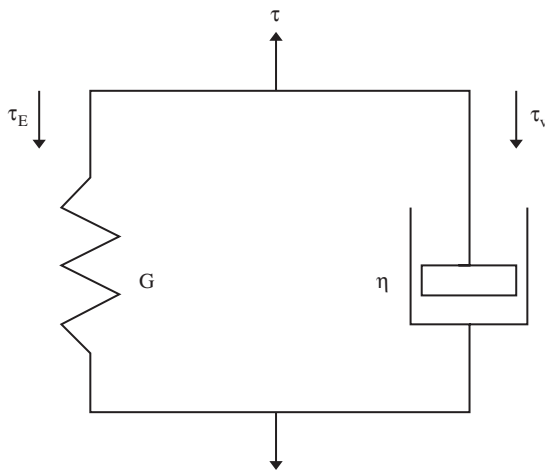


Fig. 11 Kelvin model: spring and dashpot are combined parallel.

$$\tau = G\gamma + \eta\dot{\gamma} \quad (16)$$

In mechanical terms, the Kelvin model is a parallel combination of a dashpot and a spring as illustrated in Fig. 11. The extension of dashpot and spring is always equal, as is the induced strain in both elements. If suddenly a shear stress is applied, the dashpot can only respond slowly but continuously. The response of the spring is hence inhibited, and its maximum elongation is only reached with delay. The viscosity of the dashpot determines the response time, i.e., the higher the viscosity, the slower the response of the system. When the application of the stress is terminated, the spring will try to contract to get back to its original state. Depending on the viscosity of the dashpot, this process is again more or less delayed.

The numerical description of the growth of strain in the Kelvin body is:^[1]

$$\gamma = \frac{\tau}{G} [1 - \exp^{-t/t_k}] \quad (17)$$

where t is the time and t_k the retardation time ($t_k = \eta/G$).

If all constants except α_1 and β_1 in Eq. 13 are set to zero, the following equation results:

$$\tau + \alpha_1 \dot{\tau} = \beta_1 \dot{\gamma} \quad (18)$$

Replacing α_1 with a time constant t_M ($t_M = \eta/G$; “relaxation time”) and β_1 as before with η leads to the so-called Maxwell model, which is depicted in Fig. 12. The Maxwell model comprises a serial connection of dashpot and spring, which respond to the stress with different velocities.

The stress growth for the model is described by:

$$\tau = \eta\dot{\gamma}[1 - \exp^{-t/t_M}] \quad (19)$$

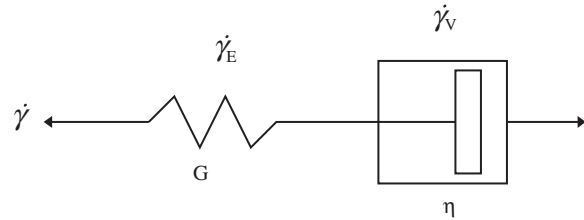


Fig. 12 Maxwell model: spring and dashpot are combined in series.

The stress decline after the removal of the employed shear strain rate is modeled as:

$$\tau = \eta \dot{\gamma} \exp^{-t/t_M} \quad (20)$$

Pharmaceutical materials are rarely described by simple mechanical equivalents such as the Kelvin (solid-like behavior) or the Maxwell (liquid-like behavior) model.

Combinations of Kelvin and Maxwell units with different characteristic time constants (t_k or t_M) are usually required. Although models using Maxwell or Kelvin units have been reported,^[27,28] it is numerically sufficient to combine either Maxwell or Kelvin units into line spectra.^[29,30] Such line spectra are easier to handle, but require care in the choice of the properties of each individual element. If the experimenter cannot relate the elements chosen to structural features in the sample, such modeling is of little value. Examples of appropriate use of line spectra can be found in the literature.^[31,32]

Creep Experiments

In practice, viscoelastic properties can be determined by static and dynamic tests. The typical static test procedure is the “creep test.” Here, a constant shear stress is applied to the sample over a defined length of time and then removed. The shear strain is monitored as a function of time. The level of stress employed should be high enough to cause sample deformation, but should not result in the destruction of any internal structure present. A typical creep curve is illustrated in Fig. 13a, together with the four-element mechanical model that can be used to explain the observations. The creep compliance represents the ratio between shear strain rate and constant stress at any time t .

When the stress is first applied to the sample, the creep compliance rises from zero to a defined value instantaneously (A–B in Fig. 13a). One can assume that this is reflected in an extension of spring 1 (Fig. 13b). Primary structural units are stretched elastically. If the stress were removed at this point, the sample would recover completely. The value of the creep compliance at point B is called the instantaneous compliance:

$$J_0 = \frac{1}{E_0} = \frac{\gamma_0(t)}{\tau} \quad (21)$$

with E_0 being the instantaneous elastic modulus and $\gamma_0(t)$ the instantaneous strain for $\lim_{t \rightarrow 0} t = 0$.

If shear continued, more links between the structural units would break and re-form, but as weaker links do so at smaller time points, there is some retardation of this process. In Fig. 13a, this phase, which is called the retarded elastic region, is presented by the curved

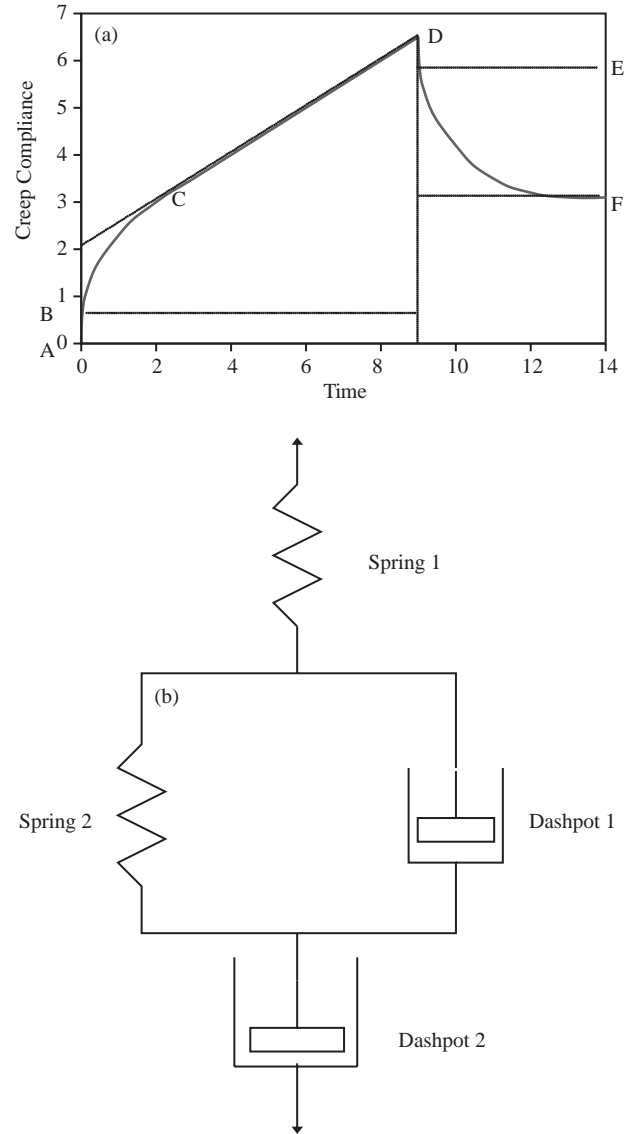


Fig. 13 (a) Typical creep curve: refer to the text for details. (b) Mechanical model to describe a typical creep curve.

compliance–time profile between the points B and C. In the mechanical model (Fig. 13b), this region corresponds to a slow movement of spring 2 and dashpot 1, i.e., the Kelvin unit. The value of the retarded compliance can be obtained from:

$$J_R = J_M [1 - \exp^{-t/t_m}] = \frac{\gamma_R(t)}{\tau} \quad (22)$$

with

$$t_m = J_M \eta_M = \frac{\eta_M}{E_R} \quad (23)$$

and J_M being the mean compliance, $\gamma_R(t)$ the retardation strain, t_m the mean retardation time, η_M the mean viscosity in the retarded phase, and E_R the retarded elasticity.

Finally, the material will flow as if it were a Newtonian body (C–D in Fig. 13a). Here, the ruptured links have no time to reform, and the linearity of this part of the curve indicates fully viscous behavior. In the mechanical model, this region refers to the deformation of dashpot 2 (Fig. 13b). The Newtonian compliance can be calculated from:

$$J_N = \frac{t}{\eta_N} = \frac{\gamma_N(t)}{\tau} \quad (24)$$

where η_N is the Newtonian viscosity, and $\gamma_N(t)$ is the Newtonian strain.

After the stress has been removed (point D in Fig. 13a), the recovery phase follows a pattern mirroring the creep compliance curve to some degree: First, there is some instantaneous elastic recovery (D–E; return of spring 1 into its original shape; see Fig. 13a, b). Second, there is a retarded elastic recovery phase (E–F; slow movement of the Kelvin unit into its original state; see Fig. 13a, b). However, during the Newtonian phase, links between the individual structural elements had been destroyed, and viscous deformation is nonrecoverable. Hence, some deformation of the sample will remain; this is in the mechanical model reflected in dashpot 2, which remains extended (Fig. 13b).

Oscillation

The most popular dynamic test procedure for viscoelastic behavior is the application of an oscillatory stress of small amplitude. This shear stress applied produces a corresponding strain in the material. If the material were an ideal Hookean body, the shear stress and shear strain rate waves would be in phase (Fig. 14a), whereas for an ideal Newtonian sample, there would be a phase shift of 90° (Fig. 14b), because for Newtonian bodies the shear strain is at a maximum, when a maximum of stress is present. The shear strain, when assuming an oscillating sine function, is at a maximum in the middle of the slope, because there is the steepest increase in shear strain due to the change in direction. For a typical viscoelastic material, the phase shift will have a value between $> 0^\circ$ and $< 90^\circ$ (Fig. 14c).

The shear stress applied:

$$\tau = \tau_0 \cos \omega t \quad (25)$$

hence produces a strain of:

$$\gamma = \gamma_0 \cos(\omega t - \delta) \quad (26)$$

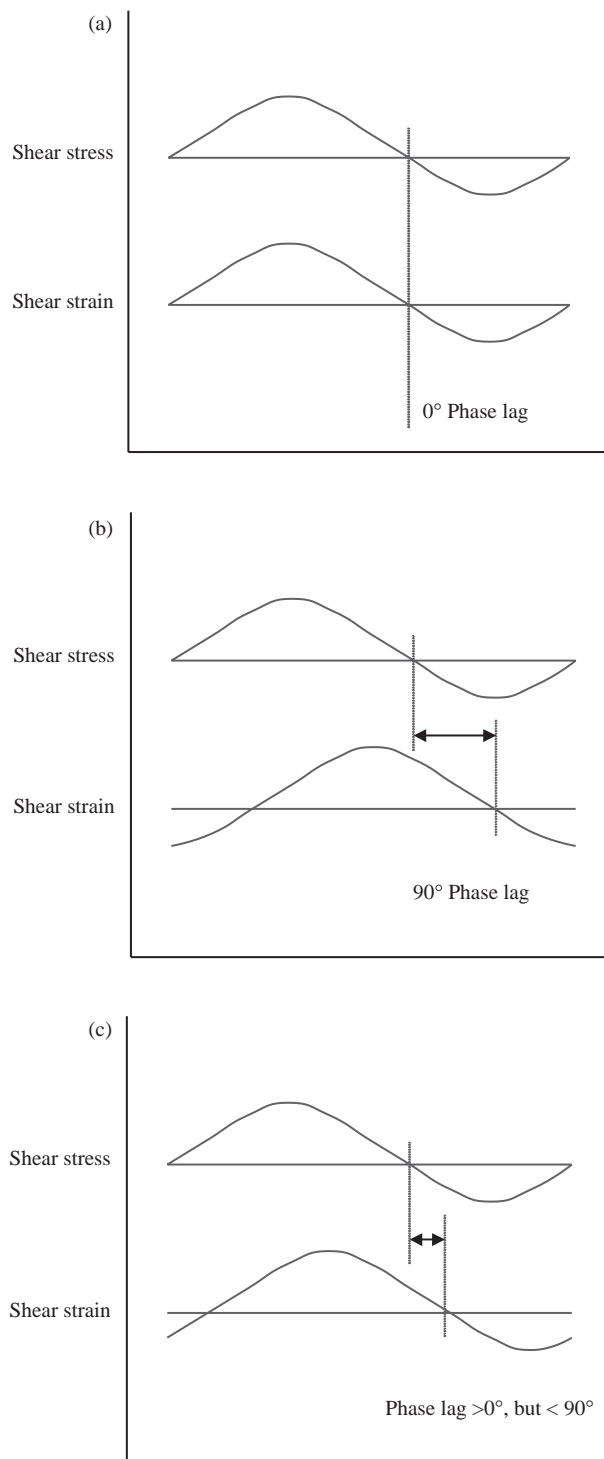


Fig. 14 Oscillating shear stress and resulting shear strain: Hookean body (a); Newtonian body (b); viscoelastic material (c).

where τ_0 and γ_0 are the stress and strain amplitudes, respectively, ω is the oscillation frequency, and δ is the phase lag (“phase angle,” “loss angle”), which quantifies

the advance of the oscillatory stress to the oscillatory strain. The values for δ and the amplitude ratio γ_0/τ_0 depend on the material. In the linear viscoelastic range, these are material properties, although both vary with the frequency employed.

Eq. 26 can be rewritten as:

$$\gamma = \gamma_0 \cos \delta \cos \omega t + \gamma_0 \sin \delta \sin \omega t \quad (27)$$

where the term $\gamma_0 \cos \delta$ accounts for the amplitude of the component of strain in phase with the stress, and the term $\gamma_0 \sin \delta$ quantifies the amplitude of the component of strain 90° out of phase.^[10] From these terms, the storage (J') and the loss compliance (J'') can be defined by normalizing with the stress amplitude:

$$J' = \frac{\gamma_0 \cos \delta}{\tau_0} \quad (28)$$

$$J'' = \frac{\gamma_0 \sin \delta}{\tau_0} \quad (29)$$

This simplifies Eq. 27 to:

$$\gamma = \tau_0(J' \cos \omega t + J'' \sin \omega t) \quad (30)$$

In the case of an ideal Hookean body with $\delta = 0^\circ$, the loss compliance is zero and J' relates to the elastic energy, which has been stored in the material. No energy has been lost. Similarly, for a Newtonian body $\delta = 90^\circ$, while the storage compliance is zero. Here, the value of J'' is related to the rate with which energy was dissipated. In a viscoelastic material the values of J' and J'' are not equal to zero, and the relative proportion is quantified by:

$$\tan \delta = \frac{J''}{J'} \quad (31)$$

When controlling the strain applied rather than the stress, a similar analysis can be undertaken to derive the storage modulus G' and the loss modulus G'' , assuming that $\tau = \tau_0 \cos(\omega t + \delta)$ and $\gamma = \gamma_0 \cos \omega t$, instead of using Eqs. 25 and 26. Such analysis shows that:

$$\tau = \gamma_0(G' \cos \omega t - G'' \sin \omega t) = G' \gamma + \frac{G''}{\omega} \dot{\gamma} \quad (32)$$

with $G' = \tau_0 \cos \delta / \gamma_0$ and $G'' = \tau_0 \sin \delta / \gamma_0$. Again, the storage modulus G' describes the elasticity of the material and has a value of unity, if the body behaves fully elastically or is highly structured such as a gel. If G' is zero, then the body behaves as fully viscous and has no inner structure (e.g., a sol). The loss modulus refers to the

viscous properties and quantifies the degree of energy dissipation due to viscous flow. G'' is zero, if the body behaves Hookean. As before, the proportion between viscous and elastic behavior at a given frequency is described by $\tan \delta = G''/G'$.

The quantities of J' , J'' , G' and G'' are related via:^[10]

$$\tan \delta = \frac{J''}{J'} = \frac{G''}{G'} \quad (33)$$

$$J' = \frac{G'}{G'^2 + G''^2} \quad (34)$$

$$J'' = \frac{G''}{G'^2 + G''^2} \quad (35)$$

$$G' = \frac{J'}{J'^2 + J''^2} \quad (36)$$

$$G'' = \frac{J''}{J'^2 + J''^2} \quad (37)$$

Finally, one can consider the strain rate amplitude ($\dot{\gamma}_0$) instead of stress or strain, which results in:

$$\tau = \dot{\gamma}_0(-\eta' \sin \omega t + \eta'' \cos \omega t) \quad (38)$$

where $\eta' = \tau_0 \sin \delta / \dot{\gamma}_0$ is called the dynamic viscosity, and $\eta'' = \tau_0 \cos \delta / \dot{\gamma}_0$ is an unnamed parameter related to the storage modulus via:^[11]

$$G' = \eta'' \omega \quad (39)$$

The loss modulus is related to the dynamic viscosity accordingly, i.e.,

$$G'' = \eta' \omega \quad (40)$$

Typically, a spectrum of either pair of quantities as a function of frequency is reported, whereby the frequency normally covers values between 1000 Hz and 0.001 Hz.^[1] The changes in the parameters measured are utilized to describe the viscoelasticity of the material.

RHEOLOGICAL MEASURING TECHNIQUES

Classification of Rheometers

There are a vast number of rheometers available from a series of manufacturers, and Whorlow^[10] classified these into “steady rotating instruments” and “tube and slit rheometers.” Walters^[33] wrote a detailed account on the

fundamentals of the design of rheometers, and Ferry^[34] reviewed the instrumental techniques available for the investigation of viscoelastic properties. Special requirements and techniques available for the industry were analyzed by Sherman.^[35]

Physical Principles

The main requirement for exact and reproducible results is that the sample undergoes laminar flow during testing. Laminar flow appears as streamlined with a velocity profile of parabolic shape when monitored in a tube viscometer. For a Newtonian sample, shear stress and shear strain rate are at a maximum in close proximity to the walls of the test tube, and zero in the center. Shear thinning samples show truncated flow patterns, and the tendency to “plug flow” increases with a decrease in the power law index.^[1] In rotational viscometers laminar flow is the movement of the sample in concentric circles around the axis of rotation of the instrument.

Departures from laminar flow, which are attributed to inertia and/or viscoelasticity,^[11] result in turbulences, i.e., an uneven flow pattern with locally clear deviations from the flow direction. In the extreme, the flowing sample can start to circulate locally, which is known as “Taylor vortices” and mainly observed in concentric cylinder instruments, where the inner cylinder rotates,^[10] i.e., in “cup and bob viscometers.”

Reynolds number Re , which describes the ratio between inertial and viscous forces, can be used to determine whether the flow in tubes is laminar or turbulent:

$$Re = \frac{\rho d v}{\eta} \quad (41)$$

where ρ is the density of the sample, d is the diameter of the tube, v is the mean flow rate, and η is the dynamic viscosity of the sample. For rotating viscometers, Reynolds number combines the relevant parameters such as cylinder, cone, or plate radii with the physical properties of the material and the flow velocity. As these equations are very equipment specific, special literature should be consulted.

Other errors, which could influence the results obtained, are, for example, wall effects (“slipping”), the dissipation of heat, and the increase in temperature due to shear. In a tube, the viscosity of a flowing medium is less near the tube walls compared to the center. This is due to the occurrence of shear stress and wall friction and has to be minimized by the correct choice of the tube diameter. In most cases, an increase in tube diameter reduces the influence of wall slip on the flow rate measured, but for Newtonian materials of low viscosity, a large tube diameter

could be the cause of turbulent flow.^[10] When investigating suspensions with tube viscometers, constrictions can lead to inhomogeneous particle distributions and blockage. Due to the influence of temperature on viscosity (see “Influence Factors on the Viscosity”), heat dissipated must be removed instantaneously, and temperature increase due to shear must be prevented under all circumstances. This is mainly a constructional problem of rheometers. Technically, the problem is easier to control in tube rheometers than in rotating instruments, in particular, the concentric cylinder viscometers.^[1]

Steady Rotating Instruments

Concentric cylinder viscometers

The schematic principle of a simple arrangement, i.e., the “cup and bob viscometer,” is shown in Fig. 15. The driveshaft rotates with a given number of rotations per minute. If there were no friction between the two

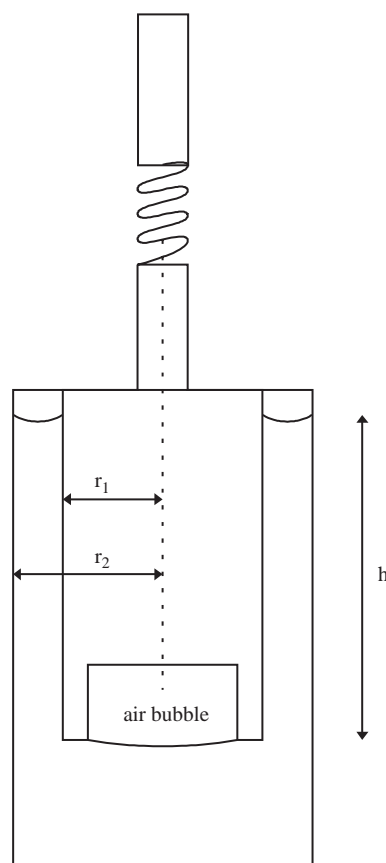


Fig. 15 Cup and bob viscometer: r_1 = radius of the outer cylinder, h = length of the inner cylinder immersed in the liquid.

concentric cylinders, the spring would transfer the rotation completely. However, the material to be investigated, which fills the gap between the two concentric cylinders, is viscous and causes resistance against the rotation. As a result, the spring cannot follow the rotation of the drive shaft and gets twisted. Thus, a torque is produced, which is proportional to the shear stress applied.

For equipment with a narrow gap width between inner and outer cylinder, i.e., a ratio of greater than 0.97 for inner (r_1) to outer (r_0) cylinder, the shear strain rate equals:

$$\dot{\gamma} = \frac{r_0 \omega}{r_0 - r_1} \quad (42)$$

where ω is the angular velocity of the inner (rotating) cylinder. The shear stress in the sample depends on the couple C and the immersed length of the inner cylinder (see Fig. 15):

$$\tau = \frac{C}{2\pi r_0^2 h} \quad (43)$$

Combining Eqs. 42 and 43, the viscosity can be calculated from:

$$\eta = \frac{C(r_0 - r_1)}{2\pi r_0^3 \omega h} \quad (44)$$

A narrow gap between the two cylinders is problematic with respect to the alignment, which must be perfectly parallel, and the use of suspensions with larger particle size.^[1] Hence, commercial equipment often provides an alternative assembly with a wider gap. Corrections are required here to allow for the shear strain rate gradient in the gap, i.e., as the shear stress is larger in close proximity to the wall of the rotating cylinder, the shear strain rate will vary accordingly. For such assemblies, Krieger and Mason provided corrections allowing the calculation of the shear strain rate, shear stress, and viscosity at the inner cylinder:

$$\dot{\gamma} = \frac{2\omega}{n[1 - (r_1/r_0)^{2/n}]} \quad (45)$$

$$\tau = \frac{C}{2\pi r_1^2 h} \quad (46)$$

$$\eta = \frac{Cn[1 - (r_1/r_0)^{2/n}]}{4\pi r_1^2 h \omega} \quad (47)$$

where n is the power law index. The latter can be determined from a double-logarithmic graph of $C = f(\omega)$.

^aSee "Extrusion and Extruders."

End effects can lead to an overestimation of the shear stress applied, and one method of reducing such errors on the measurements involves the entrapment of an air bubble beneath the inner cylinder (see Fig. 15). The method, although popular, cannot completely eliminate frictional end effects,^[37] and hence an additional calibration should be performed. The viscosity of known Newtonian samples is utilized to calculate the theoretical angular velocity for that material from Eq. 44. The difference between the experimental and theoretical angular velocity as a function of viscosity renders a straight line, from which a correction factor can be obtained.

Parallel plate systems

Parallel plate systems consist of a stationary and a rotating plate, and the sample is sheared between them (Fig. 16). The volume of the enclosed sample is equal to:

$$V = \pi R_{pp}^2 H \quad (48)$$

where R_{pp} is the radius of the parallel plates, and H the gap width (see Fig. 16).

The shear strain rate at a given radius r is:

$$\dot{\gamma} = \frac{\omega r}{H} \quad (49)$$

and hence, the shear strain rate increases from the center of the plates toward the periphery, where it reaches its maximum value. The maximum shear stress, which occurs

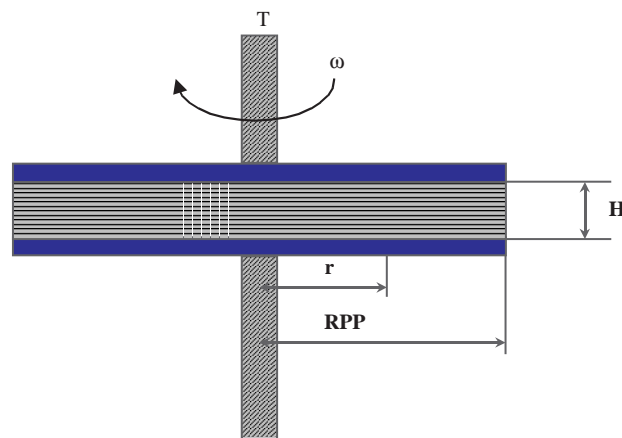


Fig. 16 Parallel plate system: r = distance of a point from the center of the plates, R_{pp} = radius of the plates, H = gap between the plates, T = torque produced, ω = angular velocity of the rotating plate.

at the periphery, equals:

$$\tau_{\max} = \frac{2C}{\pi R_{pp}^2} \left(\frac{3}{4} + 0.25 \frac{\ln C}{\ln \dot{\gamma}_{\max}} \right) \quad (50)$$

The viscosity of the sample at the periphery of the plates is then calculated from:^[33]

$$\eta = \frac{3CH}{2\pi R_{pp}^4 \omega} \left(1 + \frac{1}{3} \frac{d \ln C}{d \ln \omega} \right) \quad (51)$$

However, the use of the parallel plate geometry is not recommended for viscosity measurements, because the shear strain rate variation along the gap between the plates is larger than that experienced in concentric cylinder systems.^[10] However, there might be advantages when using the geometry in oscillatory studies.^[1]

Cone and plate viscometers

A cone and plate geometry is illustrated in Fig. 17. The plate remains stationary, while the cone rotates, or vice versa. The angle between the cone and plate surfaces is usually less than 5°. For larger angles, the analysis of the results obtained from non-Newtonian materials would be complex or even impossible. For the small angles, sample ejection is less pronounced and temperature control can be easily achieved.^[10]

At a small enough gap the shear strain rate at any point in the gap is constant. Sample volumes can be kept small and the results are highly reproducible. However, the tip of the cone should just be in contact with the plate without exertion of a force. If there were a gap between the tip and

the plate, erroneous shear strain rates would be obtained. To avoid damage of the cone tip and plate midpoint, the tip might be truncated, yet the adjustment of the gap between the truncated end of the cone and the plate still has to be very precise. Truncation facilitates a more even sample distribution throughout the gap, and specially angled plate designs can prevent sample ejection and particle movement in suspensions due to centrifugal forces.

The shear strain rate at any point r in the gap between cone and plate is given by:

$$\dot{\gamma} = \frac{\omega r}{d(r)} = \frac{\omega}{\tan \theta} \quad (52)$$

where $d(r)$ is the distance between plate and cone surface at the point r , and $r = R$ for $d(r) = d$ at the perimeter of the geometry.

The shear stress can be obtained from:

$$\tau = \frac{3C}{2\pi R^3} \quad (53)$$

where R is the radius of the cone. Finally, the viscosity of the sample can be calculated from:

$$\eta = \frac{3C \tan \theta}{2\pi R^3 \omega} \quad (54)$$

These equations are still valid with an error sufficiently less than 1%, if the radius of the truncated tip of the cone, R_1 , is less than $0.2R$.^[10]

Tube and Slit Rheometers

Hydrostatic head viscometers ("capillary viscometers")

Hydrostatic head viscometers are capillary viscometers in which a hydrostatic head of fluid provides the driving pressure for the test material to flow through a capillary. Typically, these viscometers are used with Newtonian liquids or liquid-like samples. The simplest hydrostatic head viscometer is the U-tube of Ostwald. It consists of two bulbs of similar radius, linked by a capillary. A more advanced apparatus is the suspended level viscometer of Ubbelohde. Here, the additional bulb directly below the capillary ensures that there is the same pressure over the full length of the capillary. Both the Ostwald and the Ubbelohde viscometers are filled with the sample and suspended in a water bath to equilibrate the equipment and material at a given test temperature. The time for the sample to flow through the capillary, which is proportional to the kinematic viscosity of the material, is then measured.

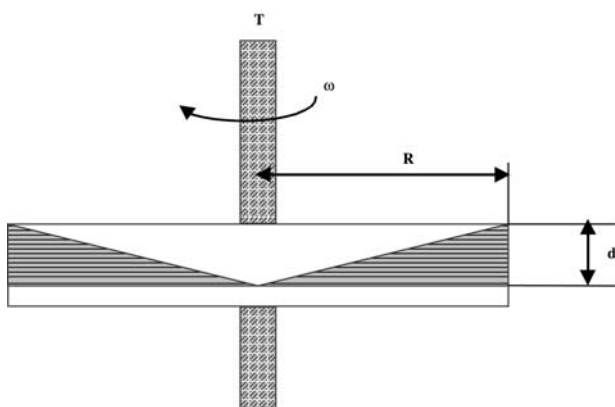


Fig. 17 Cone and plate geometry: R = radius of the cone and plate, d = gap between cone and plate at the perimeter of the geometry, ω = angular velocity of the rotating cone, T = torque produced.

The pressure gradient along the capillary of an Ostwald viscometer is given by the Hagen–Poiseuille law:

$$\frac{dP}{dl} = \frac{8(dV/dt)\eta}{\pi r^4} \quad (55)$$

where dV/dt is the volume flow rate, i.e., the liquid volume passing the capillary in a defined time span, and r is the capillary radius. The shear stress τ in the liquid varies between $(r/2)(dP/dl)$ at the capillary wall and zero in the middle of the capillary, while the shear strain rate varies between $4(dV/dt)/(\pi r^3)$ and zero.^[1]

Special developments of capillary viscometers to be used with non-Newtonian liquids are available.^[1,10] However, these require that the sample does not show any thixotropy or negative thixotropy.

Slit viscometers

Flow between two parallel stationary walls as a result of an applied pressure gradient is known as slit flow. It can be considered as the two-dimensional analog of capillary flow. This technique is used when studying practical industrial processes such as injection molding of plastics, for example, for bottle-pack lines. The shear stress is here calculated from:

$$\tau = \frac{h}{2} \frac{dP}{dl} \quad (56)$$

where h is the slit height, and the shear strain rate is given by:^[1]

$$\dot{\gamma} = \frac{2(dV/dl)}{bh^2} \left(2 + \frac{d \ln(dV/dl)}{d \ln \tau} \right) \quad (57)$$

where b is the slit width.

Falling ball viscometers

The falling ball viscometer consists of a wider tube containing the test sample. The viscosity of Newtonian-behaving materials is obtained by measuring the time for a ball of known density and size to fall through the sample. The resistance of the liquid against the fall of the ball (w) is:

$$w = 6\pi r \eta v \quad (58)$$

and the force exerted on the ball due to gravity (F_B) is:

$$F_B = \frac{4}{3} \pi r^3 (\rho_B - \rho_N) g \quad (59)$$

Under equilibrium conditions, i.e., the ball is suspended by

the liquid and hence does not fall. Eqs. 58 and 59 are equal ($w = F_B$; $v = 0$). In Eqs. 58 and 59, r is the radius of the ball, v is the velocity of the falling ball, ρ_B and ρ_N are the densities of the ball and the Newtonian sample, respectively, and g is the gravitational constant (9.81 m sec^{-2}). Combining the two equations, the dynamic viscosity of the material can be calculated from:

$$\eta = \frac{2r^3(\rho_B - \rho_N)g}{9v} \quad (60)$$

The velocity of the falling ball is obtained by measuring the time the ball needs to fall a defined distance.

The falling ball can cause turbulent flow of the liquid that is replaced, and hence the velocity of fall should be kept low. The above equations also assume that the tube has an infinite diameter. However, as this is practically not the case, the viscosity obtained from Eq. 60 should be corrected for wall effects as follows:^[10]

$$\eta_{\text{corr}} = \eta \left(1 - 2.104 \frac{d}{D} + 2.09 \frac{d^3}{D^3} \right) \quad (61)$$

where d is the diameter of the falling ball, and D is the inner diameter of the tube. The Höppler-falling ball viscometer tries to reduce wall effects and turbulent flow by a nonvertical alignment of the tube.

Viscoelasticity Measurements

Creep tests can be performed with steady rotating viscometers. Conventional forced-oscillation measurements can also be performed with ordinary shear geometries such as plate–plate or plate–cone systems in rotating rheometers.

Wave propagation^[10,34,36] involves the generation of waves at the surface of the sample and the evaluation of the velocity and attenuation of these waves when traveling through the sample. “Steady-flow rheometers”^[37] produce a sample flow so that the velocity at a fixed point in the apparatus remains unchanged during the experiment. However, the individually flowing elements of the sample undergo an oscillatory shear, because the flow velocity varies from point to point in the equipment. In the Maxwell orthogonal rheometer, this is achieved with a plate–plate geometry. One of the plates rotates with a defined constant angular velocity, while the other plate is free to rotate and its velocity is usually close to that of the other plate.^[33]

RHEOLOGY OF PHARMACEUTICAL SYSTEMS

Liquid and Semisolid Dosage Forms

Rheological work on semisolid dosage forms such as ointments and creams has not been the main focus of research in recent years. Pharmacopoeias use viscosity measurements as a quality control parameter for liquids, which behave as Newtonian. The pharmacopoeial tests for non-Newtonian semisolid dosage forms are less concerned with rheological properties, and the use of penetrometer methods was preferred over true rheometry in some pharmacopoeias (e.g., DAB 10). The European Pharmacopoeia contains a monograph also describing the use of steady rotation instruments, but does not provide details for their use to characterize or classify samples. The simple approach of penetrometry and the unspecific employment of rheometers appear unjustified, if one considers the vast amount of literature from the 1960s and 1970s on ointment bases such as soft white paraffin,^[38–44] lanolin,^[45,46] and others,^[43,45,47] where rheological measurements were shown to be able to detect and explain among other things batch to batch and source to source variability of these materials. That the problems, which these materials can pose during manufacture, are still not resolved, is indicated by the occasional reoccurrence of rheological investigations, yet now combined with other modern analytical techniques^[48–50] such as low-frequency dielectric spectroscopy and thermoanalysis. Rheological investigations on gels, which are also important as semisolid dosage forms due to their gel–sol transition, have been reviewed extensively by Gallegos and Franco.^[51]

Suspensions

Suspensions are dispersions of solid particles in a liquid. The amount of solid material suspended occupies a certain fraction of space in the total volume of the suspension. This fraction is called the phase volume ϕ . Rheological observations on suspensions are all based on a volume solid per volume liquid basis, because the rheological behavior depends mainly on hydrodynamic forces, which act on the surface of the particles or particle aggregates, irrespective of the particle density.

For dilute suspensions (10% or less phase volume) of spherical particles in a Newtonian fluid, the Einstein equation^[52,53] can be used to predict the viscosity of the suspension, η , which also behaves as Newtonian:

$$\eta = \eta_L(1 + 2.5\phi) \quad (62)$$

where η_L is the viscosity of the suspending liquid. Eq. 62 does not consider interparticulate forces, particle size, or particle shape.

Flow of any concentrated suspension will become impossible when the solid particles can form a continuous three-dimensional network of contacts throughout the sample. This so-called maximum packing fraction ϕ_m depends mainly on the particle size distribution^[54] and the particle shape.^[55] Broader particle size distributions result in lower values of ϕ_m , because the smaller particles can fill the gaps between the bigger ones, and a deviation from spherical shape results in lower values of ϕ_m due to steric hindrance of packing.^[56] Also flocculation will result in a decrease in the value of ϕ_m , because the individual flocs are only loosely packed.

Concentrated suspensions can also behave as Newtonian, but more commonly these suspensions show shear thinning or shear thickening behavior.

For Newtonian flow, the viscosity of a concentrated suspension can be calculated from the Krieger–Dougherty equation:^[57]

$$\eta = \eta_L \left(1 - \frac{\phi}{\phi_m}\right)^{-\eta\phi_m} \quad (63)$$

Shear thinning of concentrated suspensions is typical for submicron particles dispersed in a low viscosity Newtonian fluid.^[58] At low shear strain rates, Brownian motion leads to a random distribution of the particles in the suspension, and particle collision will result in viscous behavior. At high shear strain rates, however, particles will arrange in layers, which can slide over each other in the direction of flow. This results in a reduced viscosity of the system in agreement with the principles of shear thinning. A pronounced apparent yield stress can be found for shear thinning suspensions, if the Brownian motion is suppressed by electrostatic repulsion forces, which result in three-dimensional crystal-like structures of the particles with low mobility.^[59]

Shear thickening will always follow at higher shear strain rates. This is mainly due to the breakdown of the particle layers mentioned before,^[58] or due to liquid film rupture between the particles and hence direct interparticulate contact.^[1] The latter appears to be more typical for suspensions with $\phi \geq 0.5$.

The rheology of pastes, which are highly concentrated suspensions, is very important for the process of extrusion and spheronization to produce pellets. This topic is covered by J.M. Newton elsewhere in this encyclopedia.^a

^aSee “Extrusion and Extruders.”

Emulsions, Microemulsions, and Liquid Crystals

Emulsions, microemulsions, and liquid crystalline systems are suspensions of deformable particles, and many of the principles stated earlier for suspensions are valid to a similar extent. The effect of the phase volume, however, is less pronounced.^[1]

Emulsions are increasingly non-Newtonian, i.e., shear thinning with increasing concentration of the lipophilic phase.^[1] The relative effects of droplet size and shape and their distributions are less well studied, compared to solid particle suspensions, but in general, an increase in viscosity is observed for smaller and more monosized droplets. Hence, an increase in energy input, for example, by homogenization, will always result in an increase in viscosity.

The viscosity of microemulsions has been studied several times in order to determine hydration and interactions between the dispersed droplets.^[60–64] It was found that an increase in hydration of the surfactant molecules resulted in rheological behavior more similar to that of suspensions containing solid particles in low concentrations. In any case, the microemulsions showed Newtonian flow characteristics.

Liquid crystals can be classified into lyotropic and thermotropic systems. The rheology of thermotropic liquid crystals is less documented, but in general, nematic liquid crystals were found to show Newtonian flow, whereas smectic and cholesteric liquid crystals demonstrated more or less pronounced plug flow.^[65] Plug flow is typical for non-Newtonian, shear thinning liquids.^[1]

The behavior of lyotropic crystals is usually non-Newtonian.^[66] Lamellar liquid crystals show thixotropic or negative thixotropic behavior,^[67] but shear thinning/thixotropy is more common.^[68–71] The shear thinning effect is ascribed to orientation of the liquid crystalline domains and some structural breakdown. Viscoelasticity has also been observed.^[72,73] If the concentration of the surfactant molecules is, however, low and close to the CMC, Newtonian flow will be observed.^[73]

Flow of hexagonal liquid crystalline systems is presumably a function of the alignment of the rod-like aggregates along their long axis in the direction of flow.^[74] The shear thinning flow process can be accompanied by an apparent yield stress.^[75] Viscoelastic behavior has also been reported.^[76]

Cubic liquid crystalline systems have been described as clear, stiff gels.^[77] As such, they show shear thinning after an apparent yield stress has been exceeded. The viscoelastic properties are also typical for the gel character: a broad

linear viscoelastic range and a frequency-independent elastic component, which is considerably higher than the viscous component, are observed.^[78,79]

Film Coatings

Rheological techniques such as creep tests or oscillation can be applied to film coatings in order to assess their viscoelastic properties. However, the tests are undertaken typically in tension, not in shear. A typical creep–time curve for polymer films can be divided into three parts.^[80] After an instantaneous initial strain has developed, polymer films will often undergo a period of transient response to the stress applied. The tensional strain rate decreases with time and reaches a minimum steady-state value. In the third stage, the creep rate increases again leading to final rupture of the film. In typical experiments, however, the recovery phase is initiated while the tensional strain rate value is still in the steady-state region. The decrease in the tensional strain rate during the transient stage is related to changes in the substructure of the film. These changes result in an overall increase in the resistance to dislocation motion. A constant tensional strain rate during the steady-state creep stage, however, is indicative for the substructures in the film to have stabilized, and that there is a dynamic balance between strain hardening and softening processes. Creep curves obtained on polymer films might not possess a transient or steady-state stage. In the former case, the creep curve is linear throughout, whereas in the latter case, the transient part of the creep curve ends in a linear part parallel to the time axis. If the creep–time curve is fully linear, steady flow will result in permanent deformation, whereas the occurrence of only a transient stage should result in complete recovery.

SURFACE RHEOLOGY

The measurement of rheological properties at the surface of a solution or the interface between a solution and, for example, a biological film is called surface or interfacial rheology. In this technique also, experiments are performed either in tension, compression or shear, and phenomena observed in bulk rheology such as flow and viscoelasticity are also observable. An introduction to the techniques available and some key findings are discussed by Warburton.^[81,82]

REFERENCES

1. Barnes, H.A.; Hutton, J.F.; Walters, K. *An Introduction to Rheology*; Elsevier: Amsterdam, 1989.

2. Marriott, C.; Shih, C.K.; Litt, M. Changes in the gel properties of tracheal mucus induced by divalent cations. *Biorheology* **1979**, *16*, 331–337.
3. Martin, G.P.; Loveday, B.E.; Marriott, C. The effect of bromhexin hydrochloride on the viscoelastic properties of mucus from the mini-pig. *Eur. Respir. J.* **1990**, *3*, 392–396.
4. Barnes, D.M.W.; Warburton, B. The effect of pH upon the surface film characteristics of whole saliva. *J. Pharm. Pharmacol.* **1988**, *40*, 35P.
5. Dormandy, J.A. Influence of blood viscosity on blood flow and the effect of low molecular weight dextran. *Br. Med. J.* **1971**, *4*, 716–719.
6. Dormandy, J.A. Hyperviscosity angina. *Lancet* **1975**, *1*, 679–680.
7. Barnes, A.J.; Locke, P.; Scudder, P.R.; Dormandy, T.L.; Dormandy, J.A.; Slack, J. Is hyperviscosity a treatable component of diabetic microcirculatory disease? *Lancet* **1977**, *2*, 789–791.
8. Chawla, B.S. Thixotropy and rheopexy of propylidone suspensions in arachis oil: Effect of median particle size. *J. Pharm. Pharmacol.* **1968**, *20*, 168S–175S.
9. Schlichting, H.T. *Boundary Layer Theory*; 7th Ed., (translated by Kestin, J.) McGraw-Hill: New York, 1979.
10. Whorlow, R.W. *Rheological Techniques*; 2nd Ed. Ellis Horwood: New York, 1992.
11. Giesekeus, H. Sekundärströmung in viskoelastischen Flüssigkeiten bei stationärer und periodischer Bewegung. *Rheol. Acta* **1965**, *4*, 85–101.
12. Timoshenko, S.P.; Goodier, J.N. *Theory of Elasticity*, 3rd Ed. McGraw-Hill: New York, 1987.
13. Cross, M.M. Rheology of non-Newtonian fluids: A new flow equation for pseudo-plastic systems. *J. Colloid Sci.* **1965**, *20*, 417–437.
14. Boger, D.V. Demonstration of upper and lower Newtonian fluid behaviour in a pseudo-plastic fluid. *Nature* **1977**, *265*, 126–128.
15. Whitcomb, P.J.; Macosko, C.W. Rheology of xanthan gum. *J. Rheol.* **1978**, *22*, 493–505.
16. Carreau, P.J. Rheological equations from molecular network theories. *Trans. Soc. Rheol.* **1972**, *16*, 99–127.
17. Sisko, A.W. The flow of lubricating greases. *Ind. Eng. Chem.* **1958**, *50*, 1789–1792.
18. De Kee, D.; Turcotte, G.; Code, R.K. Rheological Characterisation of Time-Dependent Foodstuffs. In *Rheology Volume 3: Applications*; Astarita, G., Marrucci, G., Nicolais, L. Eds.; Plenum Press: New York, 1980; 609–614.
19. Onogi, S.; Asada, T. Rheology and Rheo-Optics of Polymer Liquid Crystals. In *Rheology Volume 1: Principles*; Astarita, G., Marrucci, G., Nicolais, L. Eds.; Plenum Press: New York, 1980; 127–147.
20. Beazley, K.M. Industrial Aqueous Suspensions. In *Rheometry: Industrial Applications*; Walters, K., Ed.; John Wiley & Sons: New York, 1980; 339–413.
21. Hoffman, R.L. Discontinuous behaviour in concentrated suspensions. *Trans. Soc. Rheol.* **1972**, *16*, 155–173.
22. Barnes, H.A.; Walters, K. The yield stress myth? *Rheol. Acta* **1985**, *24*, 323–326.
23. Ippolito, M. Rheology of Disperse Systems—Influence of NaCl on the Viscous Properties of Aqueous Bentonite Suspensions. In *Rheology Volume 2: Fluids*; Astarita, G., Marrucci, G., Nicolais, L., Eds.; Plenum Press: New York, 1980; 615–658.
24. Casson, N. Flocculated Particles. In *Rheology of Disperse Systems*; Mill, C.C., Ed.; Pergamon Press: London, 1959; 84–104.
25. Voigt, R.; Bornschein, M. *Lehrbuch der Pharmazeutischen Technologie*; Verlag Volk und Gesundheit: Berlin, 1987.
26. Von Herschel, W.H.; Bulkley, R. Konsistenzmessungen von Gummi-Benzollösungen. *Kolloid-Z.* **1926**, *39*, 291–300.
27. Fröhlich, H.; Sack, R. Theory of the rheological properties of dispersions. *Proc. Roy. Soc. London* **1946**, *A185*, 415–430.
28. Oldroyd, J.G. The elastic and viscous properties of emulsions and suspensions. *Proc. Roy. Soc. London* **1952**, *A218*, 122–132.
29. Alfrey, T. Methods of representing the properties of viscoelastic materials. *Quart. Appl. Math.* **1945**, *3*, 143–150.
30. Roscoe, R. Mechanical models for the representation of viscoelastic properties. *Br. J. Appl. Phys.* **1950**, *1*, 171–173.
31. Shama, F.; Sherman, P. Rheology of pharmaceutical and cosmetic semisolids. *J. Food Sci.* **1966**, *31*, 699–706.
32. Barry, B.W. Rheology of pharmaceutical and cosmetic semisolids. *Adv. Pharm. Sci.* **1974**, *4*, 1–72.
33. Walters, K. *Rheometry*; Chapman & Hall: London, 1975.
34. Ferry, J.D. *Viscoelastic Properties of Polymers*; John Wiley & Sons: Chichester, 1980.
35. Sherman, P. *Industrial Rheology*; Academic Press: London, 1974.
36. Harrison, G. *The Dynamic Properties of Supercooled Liquids*; Academic Press: London, 1976.
37. Maxwell, B.; Chartoff, R.P. A polymer melt in an orthogonal rheometer. *Trans. Soc. Rheol.* **1965**, *9*, 41–52.
38. Schulte, K.E.; Kassem, M.A. Rheologische Studien an Salbengelen. *Pharm. Acta Helv.* **1963**, *38*, 358–370.
39. Birdwell, B.F.; Jessen, F.W. Crystallisation of petroleum waxes. *Nature* **1966**, *209*, 366–368.
40. Davis, S.S.; Shotton, E.; Warburton, B. Some limitations of continuous shear methods for the study of pharmaceutical semi-solids. *J. Pharm. Pharmacol.* **1968**, *20*, 157S–167S.
41. Longworth, A.R.; French, J.D. Quality control of white soft paraffin. *J. Pharm. Pharmacol.* **1969**, *21*, 1S–5S.
42. Barry, B.W.; Grace, A.J. Grade variation in the rheology of white soft paraffin BP. *J. Pharm. Pharmacol.* **1970**, *22*, 147S–156S.
43. Barry, B.W.; Grace, A.J. Investigation of semisolid lipophilic preparations by small strain and continuous

- shear viscometry and their application to texture profile. *J. Pharm. Sci.* **1971**, *60*, 814–820.
44. Barry, B.W.; Grace, A.J. Rheological and sensory evaluation of work softening and recovery of pharmaceutical white soft paraffins. *J. Pharm. Sci.* **1971**, *60*, 1198–1203.
 45. Boylan, J.C. Rheological study of selected pharmaceutical semisolids. *J. Pharm. Sci.* **1966**, *55*, 710–715.
 46. Boylan, J.C. Rheological estimation of the spreading characteristics of pharmaceutical semisolids. *J. Pharm. Sci.* **1967**, *56*, 1164–1169.
 47. Davis, S.S. Viscoelastic properties of pharmaceutical semisolids. I: Ointment bases. *J. Pharm. Sci.* **1969**, *58*, 412–417.
 48. Fu, C.-C.; Lidgate, D.M. Characterisation of the shear sensitivity property of petrolatum. *J. Pharm. Sci.* **1985**, *74*, 290–294.
 49. Pena, L.E.; Lee, B.L.; Stearns, J.F. Structural rheology of a model ointment. *Pharm. Res.* **1994**, *11*, 875–881.
 50. Goggin, P.L.; He, R.; Craig, D.Q.M.; Gregory, D.P. An investigation into the use of low-frequency dielectric spectroscopy as a means of characterising the structure of creams based on aqueous cream BP. *J. Pharm. Sci.* **1998**, *87*, 559–563.
 51. Gallegos, C.; Franco, J.M. Rheology of food, cosmetics and pharmaceuticals. *Curr. Opin. Colloid Interf. Sci.* **1999**, *4*, 288–293.
 52. Einstein, A. Eine neue Bestimmung der Moleküldimension. *Ann. Phys.* **1906**, *19*, 289–306.
 53. Einstein, A. Berichtigung zu meiner Arbeit: Eine neue Bestimmung der Moleküldimension. *Ann. Phys.* **1911**, *34*, 591–592.
 54. Liu, D.-M. Particle packing and rheological property of highly-concentrated ceramic suspensions: ϕ_m determination and viscosity prediction. *J. Mater. Sci.* **2000**, *35*, 5503–5507.
 55. Turian, R.; Yuan, T.-F. Flow of slurries in pipelines. *AIChE J.* **1977**, *23*, 232–243.
 56. Wakeman, R. Packing densities of particles with log-normal size distribution. *Powder Technol.* **1975**, *11*, 297–299.
 57. Krieger, I.M.; Dougherty, T.J. A mechanism for non-Newtonian flow in suspensions of rigid spheres. *Trans. Soc. Rheol.* **1959**, *3*, 137–152.
 58. Hoffman, R.L. Interrelationships of particle structure and flow in concentrated suspensions. *MRS Bull.* **1991**, *16*, 32–37.
 59. Hiltner, P.A.; Papir, Y.S.; Krieger, I.M. Diffraction of light by nonaqueous ordered suspensions. *J. Phys. Chem.* **1971**, *75*, 1881–1886.
 60. Attwood, D.; Currie, L.R.J.; Elworthy, P.H. Studies of solubilised micellar solutions: 3. The viscosity of solutions formed with nonionic surfactants. *J. Colloid Interf. Sci.* **1974**, *46*, 261–265.
 61. Baker, R.C.; Florence, A.T.; Ottewill, R.H.; Tadros, T.F. Investigations into the formation and characterisation of microemulsions: 2. Light scattering, conductivity and viscosity studies of microemulsions. *J. Colloid Interf. Sci.* **1984**, *100*, 332–349.
 62. Peyrelasse, J.; Moha-Ouchane, M.; Boned, C. Viscosity and the phenomenon of percolation in microemulsions. *Phys. Rev. A* **1988**, *38*, 4155–4161.
 63. Ktistis, G. A viscosity study on oil-in-water microemulsions. *Int. J. Pharm.* **1990**, *61*, 213–218.
 64. Kahlweit, M.; Strey, R.; Hasse, D.; Kunieda, H.; Schmeling, T.; Faulhaber, B.; Borkovec, M.; Eicke, H.F.; Busse, G.; Eggers, F.; Funck, T.H.; Richmann, H.; Magid, L.; Soderman, O.; Stills, P.; Winkler, J.; Dittrich, A.; Jahn, W. How to study microemulsions. *J. Colloid Interf. Sci.* **1987**, *118*, 436–453.
 65. Brown, G.H.; Doane, J.W.; Neff, V.D. *A Review of the Structure and Physical Properties of Liquid Crystals*; Butterworth: London, 1971.
 66. Israelachvili, J. The science and applications of emulsions—An overview. *Colloids Surf. A: Physicochem. Eng. Aspects* **1994**, *91*, 1–8.
 67. Solyom, P.; Ekwall, P. Studien über die rheologischen Eigenschaften der Mesophasen in Natriumkaprylat–Dekanol–Wasser Systemen. *Rheol. Acta* **1969**, *8*, 316–327.
 68. Pohlin, L.; Fontell, K. Flow properties of lamellar liquid crystalline lipid–water systems. *J. Colloid Interf. Sci.* **1978**, *67*, 272–283.
 69. Hirsch, E.; Wittmann, J.C.; Candan, F. Structure and viscometric behaviour of gel phases in quaternary systems containing anionic surfactants. *J. Dispers. Sci. Technol.* **1982**, *3*, 351–372.
 70. Gallegos, C.; Nieto, M.; Nieto, C.; Muñoz, J. Influence of surfactant concentration on the time dependant rheological behaviour of the lamellar liquid crystals. *Proc. Colloid Polym. Sci.* **1991**, *84*, 236–240.
 71. Franco, J.M.; Muñoz, J.; Gallegos, C. Transient and steady flow of a lamellar liquid crystalline surfactant/water system. *Langmuir* **1995**, *11*, 669–673.
 72. Tamamushi, T.; Kodaira, Y.; Matsumura, M. A comparative study of the thermotropic and lyotropic mesophases formed by ammonium dodecanoate. *Colloid Polym. Sci.* **1976**, *254*, 571–576.
 73. Teipel, U.; Heymann, L.; Aksel, N. Rheologische Eigenschaften von wässrigen nicht-ionischen Tensidlösungen in oszillatorischer Scherströmung. *Chem.-Ing.-Tech.* **2000**, *72*, 349–355.
 74. Groves, M.J.; Ahmad, A.B. Some rheological properties of lyotropic liquid crystals formed by phosphated polyoxyethylene surfactants, *n*-hexane and water. *Rheol. Acta* **1976**, *15*, 501–507.
 75. Muñoz, J.; Alfaro, M.C. Rheological and phase behaviour of amphiphilic lipids. *Grasas y Aceites*. [Int. J. Fats Oils] **2000**, *51*, 6–25.
 76. Gallegos, C.; Muñoz, J.; Rodríguez, J.M.; Flores, V. Rheology of Lyotropic Liquid Crystals for Binary Polyoxyethylene Fatty Alcohol/Water Systems. II. The Hexagonal Mesophase. In *Process and Trends in Rheology*

- II; Giesekus, H., Hibberd, M.F., Eds.; Steinkopf Verlag: Berlin, 1988; 282–284.
77. Ekwall, P. Composition, Properties and Structures of Liquid Crystalline Phases in Systems of Amphiphilic Compounds. In *Advances in Liquid Crystals 1*; Brown, G.H. Ed.; Academic Press: New York, 1975; 1–142.
 78. Gradzielski, M.; Hoffmann, H.; Panitz, J.-C.; Wokaun, A. Investigations on L₂ phase and cubic phase in the system AOT/1-octanol/water. *J. Colloid Interf. Sci.* **1995**, *169*, 103–109.
 79. Linemann, R.; Langer, J.; Schmidt, G.; Kratzat, K.; Richtering, W. Linear and nonlinear rheology of micellar solutions in the isotropic, cubic and hexagonal phase probed by rheo-small-angle light scattering. *Rheol. Acta* **1995**, *34*, 440–449.
 80. Hertzberg, R.W. *Deformation and Fracture Mechanics of Engineering Materials*, 4th Ed.; John Wiley & Sons: New York, 1996.
 81. Warburton, B. Interfacial rheology. *Curr. Opin. Colloid Interf. Sci.* **1996**, *1*, 481–486.
 82. Warburton, B. Rheology of Pharmaceutical Systems. In *Encyclopedia of Pharmaceutical Technology*, 1st Ed. Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker: New York, 1996; Vol. 13; 371–407.



Roller Compaction Technology for the Pharmaceutical Industry

Ronald W. Miller

Bristol-Myers Squibb, Global Pharmaceutical Technologies, New Brunswick, New Jersey, U.S.A.

Paul J. Sheskey

The Dow Chemical Company, Midland, Michigan, U.S.A.

INTRODUCTION

The pharmaceutical industry uses granulation methods to enlarge and densify small powder particles into larger ones. This improves powder flow so that the material can be processed effectively and efficiently further into solid dosage forms. There are two pharmaceutical methods of granulation, wet and dry. Wet granulation methods are widely described in the literature; for example, *Encyclopedia of Pharmaceutical Technology (Granulations)* is solely devoted to this technology.^[1]

Dry granulation processing does not use moisture or heat to process powder into densified granules. In the pharmaceutical industry, there are two methods of dry granulation: slugging and roller compaction. Very little has been written about pharmaceutical dry granulation technology. Its contemporary use in the industry is about 50 yr–60 yr, beginning in the late 1940s. However, its popularity has risen in the last 20 years in parallel with the increased research for new efficacious active drug molecules in the pharmaceutical industry. These new active molecules cannot be processed so easily using wet and heat granulation processing steps, because of their chemical fragility and sensitivity. Therefore, this pushes the necessity for the use of dry granulation processing techniques to advance new sensitive active solid dosage forms in the 21st century.

BACKGROUND

Briefly, in dry pharmaceutical granulation processing, the powder particles are aggregated under high pressure, typically 30 bar–70 bar pressure. Particulate matter can be aggregated when compressed at high pressure because of bonding forces developed by the direct contact between the solid surfaces. The high pressure serves to improve the contact area between the surfaces and thus the overall bonding strength. Sometimes, a binding agent is needed to provide additional bonding strength. Later in the chapter, there is more about the kind of excipients used in roller

compaction formulations to aid the manufacturing of pharmaceutical products.

In the pharmaceutical industry, dry granulation processing in the 1950s–1970s favored a process called slugging. This process design consisted of feeding powder into a large compression machine, such as a Stokes D3 type compression machine, where the powder was compressed into large tablets or slugs, typically in the order of 1-in. diameter with a tablet gauge of about 0.25 in. The tablet slugs were subsequently milled by a separate sizing machine to an appropriate particle size distribution, and further processed into pharmaceutical capsules, powder for oral suspensions, sachets, or tablet dosage forms. The slugging process is still used today by only a few manufacturing firms that have old pharmaceutical or over-the-counter formulation processes. Today, modern pharmaceutical formulation processes introduced into the Americas, Western Europe, Australia, and parts of Asia do not use this kind of dry granulation equipment in newly developed formulations. The slugging machine process is a relic of the past in the 21st century of modern pharmaceutical technology; roller compaction is the key technology to dry granulation processing of the future.

Some characteristics are described briefly about the slugging process to complete the technology information for the reader. The slugging process is externally influenced by raw material feed properties such as powder cohesiveness, density, flow characteristics, and powder particle size distribution. The slugging machine's design characteristics such as machine type, feed hopper, feed frame, die diameter and tooling features, compression speed, and slugging pressure also influence the slugging process and the final product properties. In general, the key processing operational aspect of slugging is to maintain a uniform powder fill weight into the dies during the dynamics of the slugging process. This assures the best chance to manufacture uniform powder slugs and ultimately, uniform densified granules. The compression-slugging setup is a key essential to maximizing the slugging throughput and minimizing the hopper feed frame and die powder flow problems associated with the process. Slugging

**Table 1** Disadvantages of slugging

Single batch processing	Excessive air and sound pollution
Frequent maintenance change-overs	Increased use of storage containers
Poor process control	Increased needs of manufacturing space
Poor economies of scale	Increase of logistics
Low manufacturing throughput per hour	More energy and time required to produce 1 kg of slugs than 1 kg of roller compact

(Courtesy of Handbook of Pharmaceutical Granulation Technology.^[2])

compression is normally performed at 4 tn–6 tn hydraulic pressure, at a rate of 10–30 turret revolutions per minute. The specific machine tonnage, turret speed, and roll dwell time required for the process are dependent on the powder blend's physical properties, the tooling configuration, machine parts, and ultimately the slug specifications. Typical slugging machine output ranges from 30 kg hr⁻¹ to 50 kg hr⁻¹. Slugging machines are not instrumented with modern devices to control their performance. There are numerous disadvantages with the slugging technology in the pharmaceutical industry as listed in Table 1.

BENEFITS OF ROLLER COMPACTION

This chapter identifies key aspects of roller compaction technology. Unlike the slugging process technology, roller compaction technology is well suited for dry granulation agglomeration in the era of modern development of active pharmaceutical ingredients and the design of modern pharmaceutical plants. The increasing scale of manufacturing pharmaceutical products worldwide and the need for high processing rates, together with increased levels of good manufacturing practices necessitate controlled dry granulation processes with as few processing steps as possible. This has been accomplished by instrumenting roller compactors to automate and control the mechanical process. Roller compaction technology plays a very important role in providing competitive cost control, safety, and quality products in the pharmaceutical industry. Key roller compaction benefits observed in the pharmaceutical industry are identified in Table 2.

A number of these attributes, best technology practices and features, were rated for their industrial and pharmaceutical importance and reported by Miller and Sheskey in 2001.^[3] Ultimately, today's roller compaction technology offers a continuous process with better process controls, manufacturing efficiencies, and environmental protection than an archaic slugging process technology.

PREFACE: POWDER GRANULATION AND COMPACTION

Powder granulation is a process of powder size enlargement that incorporates small particles into larger ones. The definition of granulation comprises a range of different size enlargement methods that can be classified as either dry or wet. In wet methods, a suitable liquid is used to agglomerate the small powder particles into a mass. The wet mass is subsequently dried and sized for further down-stream processing needs. Wet granulation methods have been the most widely used powder granulation technology in the production of pharmaceutical products, particularly in modern pharmaceutical manufacturing.

The chief reasons to granulate powders for the manufacture of pharmaceutical dosage forms are described by Kristensen and Schaefer.^[1]

- To improve powder flow properties for dosage filling and compression processes.
- To eliminate wet granulation induced degradants and to improve product stability.
- To prevent active product ingredient from segregating.
- To reduce bulk volume thereby minimizing storage and enhancing transport.
- To reduce potential environmental and safety hazards

Kristensen and Schaefer provide ample literature references in their chapter in the Encyclopedia of Pharmaceutical Technology about granulation size enlargement methods.^[1] Capes,^[4,5] Pietsch,^[6,7] Sherrington and Oliver,^[8] Kapur.^[9] Others also referenced about wet granulation technologies: Kristensen and Schaefer,^[10] Lindberg,^[11] Fonner, Anderson, and Banker,^[12] Anderson, Banker, and Peck,^[13] and Ghebre-Sellassie.^[14]

COMPACTION THEORY

The bonding forces in a dry aggregate are important to granulation properties such as granule integrity, flowability, friability, density, compressibility, and size for down-stream manufacturing process steps.^[1] Rumpf and coworkers described the bonding mechanisms occurring during dry granulation as a mixture of van der Waals' forces, mechanical interlockings, and a recombination of bonds established between freshly created surfaces and solid bridges, created because of partial melting and solidification during compression.^[15]

A general theory describes particle bonding related to roller compaction in the *Handbook of Pharmaceutical Granulation Technology*.^[2] The process of dry granulation relies on interparticulate bond formation. Granule bond

Table 2 Advantages of roller compaction

Simplifies processing	Uses less raw materials
Eliminates aqueous and solvent granulating	Eliminates water-induced degradants
Facilitates powder flow	Improves process cycle time
Uses minimal energy to operate	Prevents particle segregation
Requires less man-hours to operate	Facilitates continuous manufacturing
Improves drug dosage weight control	Improves content uniformity
Reproduces consistent particle density	Does not require explosion-proof room or equipment
Produces good tablet and capsule disintegration	Produces a dry product that is process scaleable

(Courtesy of Handbook of Pharmaceutical Granulation Technology.^[2])

formation is characterized in different stages, which usually occur in the following order:

1. Particle rearrangement.
2. Particle deformation.
3. Particle fragmentation.
4. Particle bonding.

Particle rearrangement occurs initially as powder particles begin filling void spaces. Air begins to leave the powder blend's interstitial spaces, and particles begin to move closer together. This action increases the powder blend's density. Particle shape and size are key factors in the rearrangement process. Spherical particles will tend to move less than other-shaped particles because of their close initial packing to one another. Particle deformation occurs as compressional forces are increased. This deformation increases the points of contact between particles where bonding occurs and is described as plastic deformation.^[2]

Particle fragmentation follows as the next bonding stage. This occurs at increased compression force levels. At this stage, particle fracturing creates multiple new surface sites, additional contact points, and potential bonding sites. Particle bonding occurs when plastic deformation and fragmentation happen. It is generally accepted that bonding takes place at the molecular level, and this is due to the effect of van der Waals' forces.^[2]

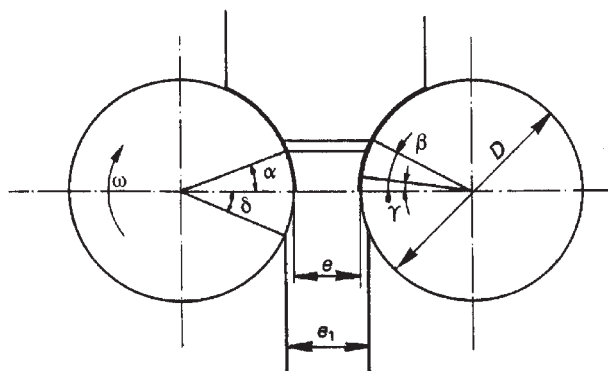
When powder granules undergo an applied force or stress, a stress force is released from the granules. The granules attempt to return to their original shape or form; this is described as *elastic deformation*. A deformation, which does not totally recover after the stress is released, is a *plastic deformation*. Elastic and plastic deformations can occur simultaneously, but one effect usually predominates.

Parrot identified that three theories of compressional bonding exist: mechanical, intermolecular, and liquid-surface film. Mechanical bonding purports that individual particles undergo elastic, plastic, and brittle deformation. Bonding of this nature occurs because particle surfaces intertwine, forming mechanical bonds. Intermolecular

theory identifies that there are some unsatisfied surface ions that have a potential need to bond to one another. Under pressure, intermolecular forces become pushed together close enough so that van der Waals' forces can act to consolidate particles. The liquid-surface film theory identifies that bonding occurs because of the existence of a thin liquid film. The thin liquid film is generated from pressure induced by the energy of compression. This mechanism acts as a bonding agent promoting mechanical strength and an enlarged particle.^[16] Very little information has been written about this last theory.

Dehont et al. provided a simplified approach to roller compaction theory.^[17] They described that powder granules move through stages in the feed area. The material is drawn into the gap by rubbing against the roll surfaces. The densification that occurs in this area is particle rearrangement. At this stage, the speed of the powder is slower than the peripheral speed of the rollers. Fig. 1 represents compactor rolls in the horizontal plane; powder is pushed vertically downward into the compaction area.

Note in Fig. 1, α is the nip angle and β the material in volume space. The material is located in the compaction area between α and the horizontal axis (Fig. 1). At this stage, the material undergoes additional compaction


Fig. 1 Front view of compactor rolls in horizontal plane. (Courtesy of Ellis Horwood.^[17])

forces. The particles undergo plastic deformation and are bonded. Dehont's team noted that nip angle varies according to the material characteristics of particle size and density and the angle is about 12° .^[17] They defined the neutral angle, γ , which corresponds to the point where the pressure applied by the rollers is the greatest on the material. They also defined elastic deformation, δ , and that occurs after the compact begins leaving the compression roll area. Compacted flakes may increase in size due to material elastic deformation and actually may have a larger thickness than the roll gap, e .^[17]

Dehont et al. developed Eq. 1 for the linear variation of flake thickness at a specific roll diameter:^[17]

$$e_1 = D(d_0/d_1 - d_0)(1 - \cos \alpha) \quad (1)$$

where e_1 is the flake thickness, D , the roll diameter, d_0 , the material density at angle α , and d_1 , the flake density.

Dehont et al. assumed that the material in the compaction area remains horizontal and moves at the peripheral speed of the rollers. They also considered that the angle α is independent of the roller diameter size and noted that the flake thickness e_1 depends on the roller speed, the roller surface, and the compaction pressure. All these parameters influence the density of the flake, d_1 . Dehont et al. concluded that if the same flake thickness was obtained with different roller diameters, the flake density would be greater with larger diameter rollers.^[17] This is due to the greater nip angle formed, with the larger rolls allowing more material to be compacted.

Heckel considered the compaction of powders analogous to that of a first-order chemical reaction. The pores were the reactant and the densification of the material the product. The proportionality between the change of density with pressure and the pore fraction was the process kinetics.^[18] Heckel explained mathematical constants that described the compaction behavior of a given powder and developed a mathematical relationship, Eq. 2.^[18]

$$\ln(1/1 - D) = KP + \ln(1/1 - D_0) \quad (2)$$

where D is the relative powder density, D_0 , the relative loose powder density at zero pressure, P , the pressure applied, $1 - D$, the pore fraction, and K , the proportionality constant.

The expression of density–pressure relationship permitted the determination of density values in the range of the pressures investigated. Heckel described mathematically that the curved region when plotting $\ln(1/1 - D)$ vs. P is associated with powder densification.^[18] This occurred by a mechanism of individual particle movement in the absence of interparticle bonding. Heckel concluded that the densification represented by the linear region of the plot, $\ln(1/1 - D)$ vs. P , occurred by

plastic deformation of the compact after an appreciable amount of interparticle bonding had taken place. For quantitative reasons, Heckel redefined the mathematical expression denoting a constant, $A = \ln(1/1 - D_0)$, which is quantitatively valid except at low pressures. He postulated that the constant A , which is somewhat larger than $\ln(1/1 - D_0)$, represents the degree of packing achieved at low pressures as a result of rearrangement processes before appreciable amounts of interparticle bonding take place.^[18] He also postulated that the constant K , the slope of the linear region, gives a measure of the compact to densify by plastic deformation.

Heckel concluded that density–pressure data indicate the rate of the change of density with pressure, at any pressure, is proportional to the pore fraction in the compact at that pressure. Additionally, density–pressure curves may be described by two parameters. He theorized that one was related to low pressure densification by interparticle motion. The second measured the ability of the compact to densify by plastic deformation after appreciable interparticle bonding.

Johanson identified, through very comprehensive mathematical models and relationships, material properties, press dimensions, and operating conditions for roll compactors. For more information, the interested person should read the entire reference.^[19] In part, he described that roller compaction involves a continuous shear deformation of the granules into a solid mass.

To satisfy the theory's assumption, it was postulated that the material be isotropic, frictional, cohesive, and compressible. Fig. 2 depicts material in a press that undergoes shear deformation into a solid mass.^[19]

In Fig. 2, P_0 is the horizontal pressure between rolls, θ , the angular position of roll bite, α , the nip angle, $2d$, the

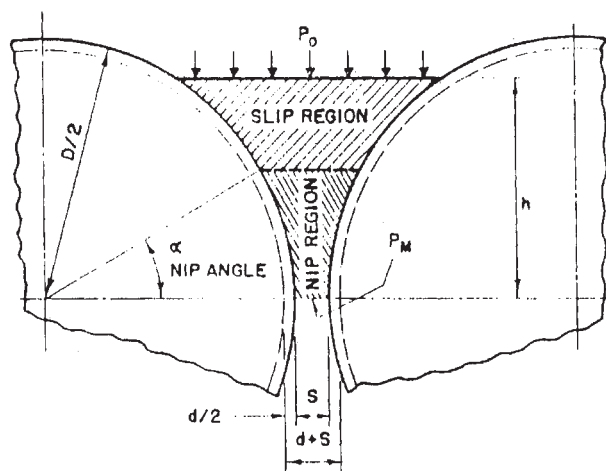


Fig. 2 Front view of compactor rolls in horizontal plane depicting powder regions under different compaction forces. (Courtesy of Transaction ASME.^[19])

roll diameter D , h , the height above the roll center line at which feed pressure P_o is applied, P_m , the horizontal pressure at $\theta = 0$, and S , the roll gap.

Johanson pointed out that no roller compactor theories at that time determined the angle of the nip and the bulk density at $\theta = \alpha$, except by actually rolling the granular solid in a roll press. He also provided a method to calculate the nip angle and the pressure distribution between the rolls. His calculations determined the pressure distribution above the nip area and the pressure in the nip area.^[19]

He provided the technical rationale to calculate the nip pressures in the nip region. He showed that material trapped in a volume V_α between arc-length segments ΔL must be compressed to volume V_θ between the same arc-length segments. The relationship requires that the bulk densities γ_α , γ_θ in volumes V_α , V_θ be related by Eq. 3 (Fig. 3):^[19]

$$\gamma_\alpha / \gamma_\theta = V_\theta / V_\alpha \quad (3)$$

Fig. 3 defines the following: P_o —horizontal pressure between rolls; θ —angular position of roll bite; α —nip angle; $2d$ —roll diameter D ; h —height above the roll center line at which feed pressure P_o is applied; P_m —horizontal pressure at $\theta = 0$; S —roll gap; ΔL —arc-length segments; V_α —material trapped in volume space described by arc-lengths; V_θ —compressed volume space described by arc-lengths; γ_α and γ_θ —respective powder bulk densities in volume spaces V_α and V_θ ; and K —a material property constant for a given moisture content, temperature, and time of compaction.

Johanson stated that the pressure σ_θ at any $\theta < \alpha$ can be determined as a function of the pressure σ_α , at $\theta = \alpha$, by the pressure–density relationship. It was understood that for increasing pressures, log density was a linear function of log pressure.^[20]

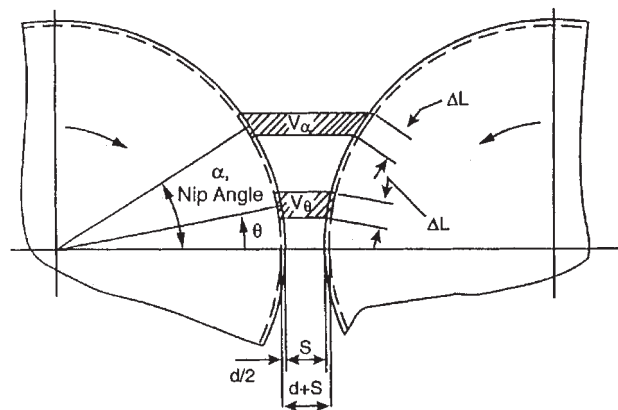


Fig. 3 Front view of compactor rolls in horizontal plane depicting nip angle. (Courtesy of Transaction ASME.^[19])

Johanson found that the nip angle does not depend on the magnitude of the roll force or the roll diameter. He demonstrated that the nip angle was affected very little by the geometry of the press or the cut grooves on the roll surface. It was mostly influenced by the nature of the materials that were compressed. The very compressible materials, with small K -values, had very large nip angles. On the other hand, incompressible materials, with large K -values, had very small nip angles. Ultimately, Johanson's results showed that material properties determine the maximum pressure that a roller press can apply to a material.

Parrott evaluated compaction of several pharmaceutical powders into compact sheets that were sized by an oscillating granulator into granules. He studied flake thickness, bulk and tap densities, angle of repose, and flow rate after sizing. His work demonstrated that one could improve most granule densities but not necessarily improve granule flow properties.^[21]

Nearly all particulate matter can be aggregated when compressed at high pressure. In some cases, pressure alone cannot achieve sufficient bonding strength, therefore, a binding agent is required to be added to the powder blend.^[1] Polymeric binders added to mixtures at certain percentages form highly viscous bridges between particle-to-particle powders. When the polymeric mixtures are compressed, they increase the overall compact strength and enhance particle-to-particle bonding and can improve powder flowability. Later, in this chapter, additional information about the use of binders will be discussed.

Pietsch described a compaction capacity throughput equation for a roller compactor using flat rolls. The equipment capacity equation is theoretical in nature; it assumes that all compacts are 100% useable for downstream processing needs. Compact capacity throughput, C_c , for a roller compactor can be determined by Eq. 4.^[6]

$$C_c = \pi D l h_A n 60 \gamma \quad (4)$$

where C_c is the roller compactor throughput (kg hr^{-1}), D , the roller diameter (cm), l , the roller length, working width (cm), h_A , the gap width between the rollers, sheet thickness (cm), n , the revolutions per minute (min^{-1}), and γ , the apparent sheet density (kg cm^{-3}).

DESIGN FEATURES OF ROLLER COMPACTORS

Certainly, a key enhancement that highlights today's pharmaceutical industry's state-of-the-art roller compactors, is programmable logic controllers (PLCs). They are used to control and monitor mechanical parts that regulate screw feed rate, roll speed, roll pressure, roll gap, vacuum deaeration, and mill speed. This section will not discuss

PLC designs as it relates to compactor design features and advancements.

Briefly, a key machine innovation, vacuum deaeration, was a new important feature design added by some roller compactor vendors in the early-mid-1990s. The design feature has been shown to help premodify raw material density prior to compacting and increase throughput.^[22] Other equipment features such as multiple horizontal or angled feed screws have assisted manufacturing a uniform raw material feed across the rolls.^[22,23] Newly designed roll machine blocks, featuring cantilever roll systems, offer more efficient ways to clean, handle, and facilitate product and equipment changeovers. New storage hoppers and various screw feeder designs have improved delivering poor flow powder to the rolls.

A history of hopper and feed screw designs showed that each design evolved to facilitate and improve powder flow

to the compactor feed screw conveyance system. Feed hopper designs depicted in Fig. 4 show designs incorporated in older compactor models [examples (a)–(c)]. Dehont et al. described the powder compaction feeding systems that were in use up to 1989.^[24]

Hopper (a) was used when the material had good rheological characteristics. Materials that had good but unique flows, and needed to be evenly distributed across the roll surface, such as granular salts, used hoppers that had guiding flaps (b). When the powder material had poor rheological characteristics, a vertical flap distribution box (c) was fitted under the hopper. It was primarily used when compacting ores. Force-feeder designs, examples (d)–(f), are more commonly used in the pharmaceutical industry. They are employed with various configured designs and shapes to supply poorly flowing powder to the nip roll region.

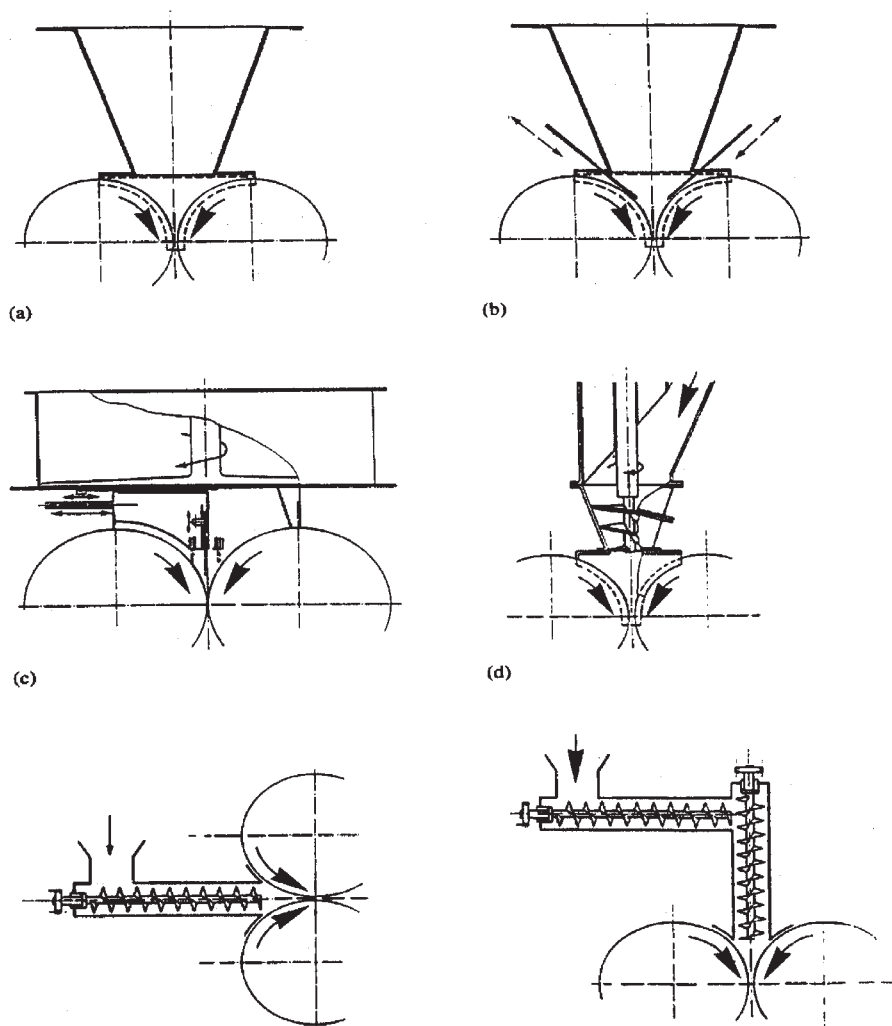


Fig. 4 Feed hopper designs: (a) simple hopper, (b) simple flap hopper, (c) flap distribution box, and (d) force-feeder. (Courtesy of Ellis Horwood Limited.^[17])

Pietsch described, in Fig. 5, modern feed screw system designs; (a) depicts vertical force-feeding screw with slightly tapered end, (b) an inclined feed screw, (c) a vertical tapered and blade angled feed screw, and (d) horizontal (single or dual) straight feeder screw(s).^[25]

Innovatively designed feed screw systems are used commercially in roller compactors worldwide. An example is shown in Fig. 6.

Sizing devices are now trimly fitted to the compactor body and controlled by variable speed drives. Most compactors no longer require a second machine (mill) in tandem to size compacts as required by slugging technology. Roller compactors have clean-in-place (CIP) systems that offer environmental and safety features. These systems minimize human exposure to chemicals and improve cleaning efficiencies, Fig. 7.

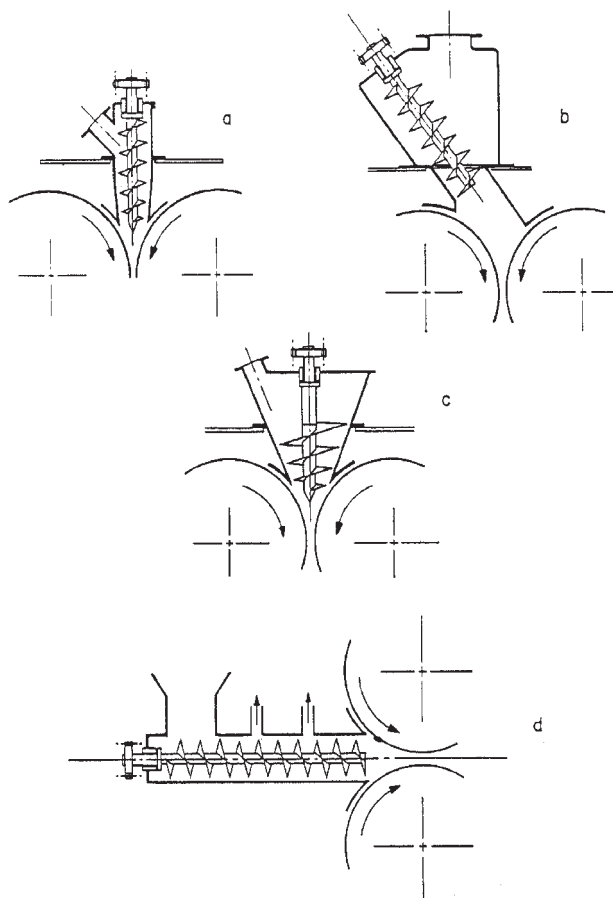


Fig. 5 Force-feeders: (a) vertical straight and lightly tapered screw feeder, (b) straight inclined screw feeder, (c) tapered vertical screw feeder, and (d) horizontal straight screw feeder. (Courtesy of Handbook of Powder Science and Technology.^[25])



Fig. 6 Twin horizontal feed screw system with vacuum deaeration. (Courtesy of Alexanderwerk Inc.)

ROLL CONFIGURATION

Compactor design features have evolved over the years. By the mid-1970s, research revealed a number of roll design improvements that increased compacting efficiency. Three key conditions were identified, at that time, which optimized the roll compact throughput and minimized leakage of noncompacted powder.^[26]

- Adequate powder supply must enter the gripping zone.
- Powder must be conveyed fully into the narrowest part of the roller gap.
- Compaction pressure must be distributed as uniformly as possible across the whole roller-gripped powder mass.

Equipment engineers and researchers worked on improving feeding equipment systems and roll designs to satisfy and maximize the above conditions. Some of the key advances are identified in Ref. 22 and are re-emphasized in this chapter.

Because of powder feed variability at the nip and in the roll gap regions, powder leakage is produced during the compaction process. This situation produces excessive fines and possible undesirable processed material. Usually, this problem is caused by uneven powder flow and compact formed when the powder is fed towards the middle of the roll width. Granules produced under these conditions are typically not optimal for further pharmaceutical processing.

Trials performed by Funakoshi, Asogawa, and Satake using rectangular aperture chutes, fitted at the end of a feed screw, aided in preventing uneven powder flow across the roll width. Funakoshi's team demonstrated the positive effects of having concavo-convex roller pair fitted with inner ring walls. This feature counteracted side seal effects (fractured or incomplete compacted edges). They also designed and tested other rolls (concavo-convex with rim), which allowed the powder to distribute more uniformly

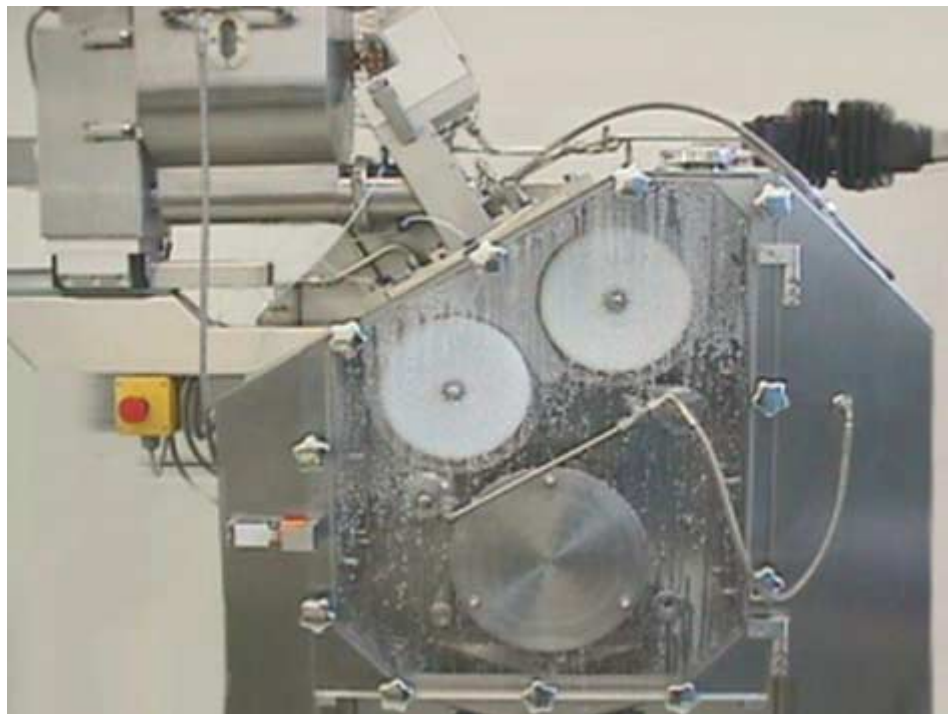


Fig. 7 Roller compactor CIP system. (Courtesy of Gerteis Co.)

across the roll width. This design reportedly minimized powder leakage during compaction.^[26]

Funakoshi's team also showed, when the roll rim inner angle is zero that the powder is not adequately and uniformly delivered to the gripping and compacting zones. This occurred because the stationary side seals acted as resistors to the powder flow.^[26] Their work demonstrated that the formed compact and the compaction pressure (using standard rolls) produced an uneven compact across the roller ends because of side seal friction. When they employed the newly designed concavo-convex rimmed rolls, it protected the compact from the adverse side seal fracturing effect. Funakoshi, Asogawa, and Satake also determined a proper selection of rimmed rollers, which delivered adequate powder to the compaction zone and also conveyed powder fully across the roller gap region. They concluded that the height and slope of the inner walls' rims optimally influenced the side seal effect. A 65° inner wall slope produced 2.5%–3.0% fines' leakage during the compaction of lactose. When the rolls had no inner wall rims, the fines' leakage was 15%. In summary, they optimized the compact and pressure distribution along the rolls by using the concavo-convex rimmed shaped rolls. They acknowledged that the best roll design was still dependent on the powder raw material properties.^[26]

Parrot further substantiated the usefulness of rimmed concavo-convex roller pair to increase the density of

several pharmaceutical powders. His work resulted in optimizing a process with an uncompacted leakage rate of 5%. The optimization depended on the physical properties of the powder and the machine operating conditions such as the roller gap, feed screw speed, and roll speed.^[27] Some of their key design feature findings are summarized:

- Installing concavo-convex roll surface rather than flat roller pair.
- Installing rectangular feed chute and flaps.
- Designing cylindrical, conical/cylindrical, or tapered-shaped variable speed auger feed screws.
- Optimizing roll rim angle to 65°.
- Installing digital and analog variable feed screw controllers.
- Developing horizontal and variable screw feed systems.

Jerome et al. in the 1980s studied the effects of compactor adjustments on powder blend properties. His team's research showed that pressure applied by the movable roller is not a predominant factor. They found that the most important variable was the compactor feed screw in relation to the roll speed. They correlated improved tablet compression hardness to the feed screw speed in relation to the speed of the rolls.^[28]

Typical compactor roll configurations are shown in Fig. 8. Most compactors now have one floating roll and



Fig. 8 Roll configuration designs. (Courtesy of Alexanderwerk Inc.)

one fixed roll each on separate bearing blocks. An illustration of this is shown in Fig. 9.

Currently, one manufacturer of roller compactor equipment for the pharmaceutical industry has a fixed roll-pair system. This type of roll design system relies on powder gravity feed and a specially designed compression feed screw to continuously deliver powder feed stock to the fixed roll pair, Fig. 10.

Pietsch described two pairs of rollers with different diameters D_1 and D_2 , with identical roll gaps h_A . He theorized that if the peripheral speed of both pairs of rollers is the same, roller compaction takes place more gradually in the case of the larger roll pair. At the same time, the larger diameter roll pair pulls a larger powder

volume into the nip region, resulting in a higher density of compacted product. The larger diameter roll pair also minimizes air entrainment more efficiently than the smaller diameter roll pair when both are operating at the same peripheral speed.^[25]

Pietsch noted that the peripheral roll speed and particulate powder speed are not equivalent in the entire compaction zone. Throughput does not increase proportionally with roll speed. There are two effects that hinder throughput: starved conditions in the feed zone, and secondly, too much squeezed air from the particle mass flows upward and against the powder flow, reducing the supply of material to the nip area.^[25]

What is the optimum roll speed for a compaction process? What factors does the formulating scientist or process engineer need to consider to maximize compact quality and compaction throughput? Johanson^[29] attempted to answer these questions by predicting roll-limiting speeds for briquetting presses. He developed mathematical expressions considering even the gas and liquid effects as they can theoretically be squeezed from a solid mass. Solid properties, press dimensions, and operating conditions were evaluated to predict optimum roll speeds. The results necessary for a quality briquette are most critical for low density fine particles. Johanson's work showed the relation between feed pressure and roll speed to essentially be proportional to the material's permeability (porosity). For example, when the initial

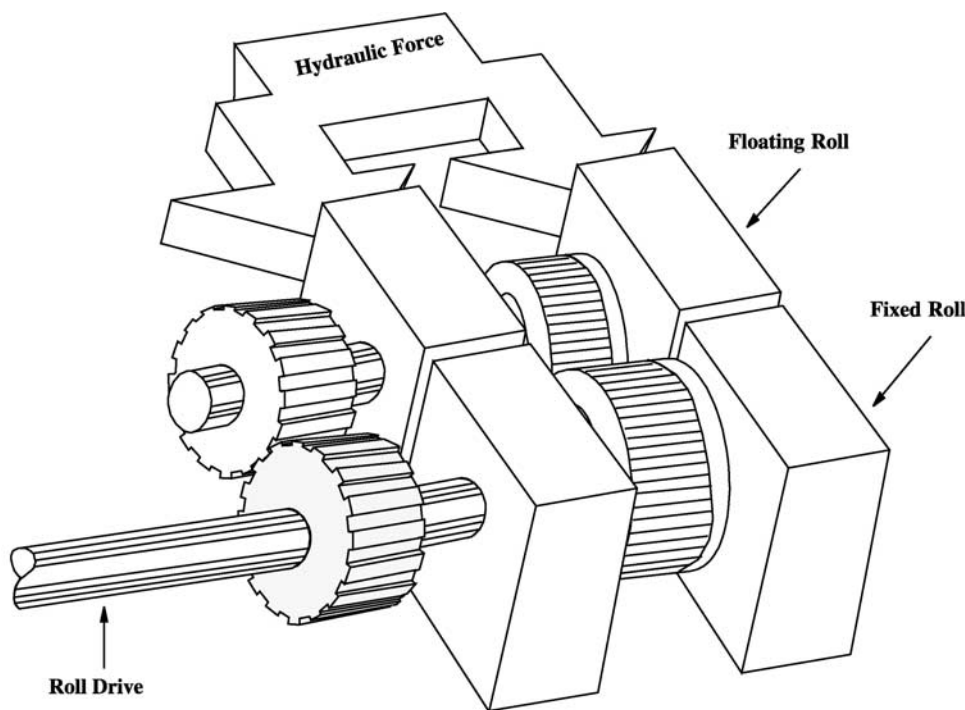


Fig. 9 Compactor roll blocks. (Courtesy of Handbook of Pharmaceutical Granulation Technology.^[21])

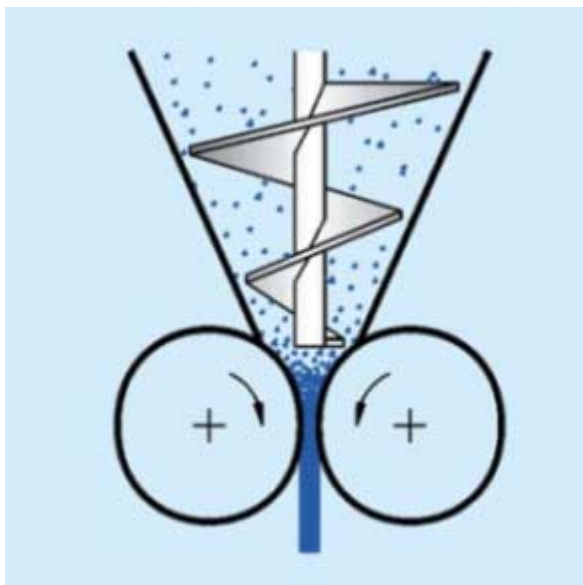


Fig. 10 Compression feed screw and fixed roll system and no vacuum deaeration. (Courtesy of Bepex-Hosokawa Co.)

powder was compacted, sized, and recompact, the bulk powder permeability increased; however, when the powder was compressed, the compression force decreased, and compactor feed pressure force requirement decreased significantly. Johanson^[29] also demonstrated that if the compactor feed pressure is kept constant and the press speed is increased, the maximum pressure applied to the briquette decreases as the roll speed increases. As a result, the briquette density and strength decrease likewise. For additional information, the interested reader should peruse the reference list.

Sheskey and Hendren in 1999 studied the effect of roll surface configuration on the drug release and physical properties of a HPMC matrix controlled-release dosage form.^[30] Smooth and axial-grooved roll surface designs were studied using a Vector Model TF-Mini roller compactor (Vector Corporation, Marion, Iowa). Their hypothesis was that the greater the depth of the concavity on the roll surface, the pressure exerted on the powder being compacted would not be as evenly displaced as with a smooth roll surface. Therefore, as with tablet tooling design, the top of the crown area of the compacted ribbon would theoretically be softer than the rest of the ribbon. However, the results of a particle size distribution test performed on milled ribbons generated using both smooth and axial-grooved roll pairs showed similarity between tested samples. In addition, results showed little difference in tablet crushing strength values between samples manufactured using either type (smooth, axial-grooved) roll surface design. Drug release profiles of tablets prepared from roller compacted granulations using both

roll surface configurations were also similar. It was noted that the authors did not measure product throughput rates to address the possibility of improved efficiency when using an axial-grooved roll pair vs. a smooth roll pair.

FEED SCREW DESIGN

The consistency and evenness of the powder feed into a roll pair determines to a large extent, how complete a compact is made and ultimately the success of a compaction process. Most roller compacting systems suffer the disadvantage of leakage, i.e., 20%–30% powder particles (depending on the formulation) are not compacted. This primarily occurs because of uneven powder feed and powder slippage between individual loose particles and the roll surfaces. Under these conditions, it is usually necessary to recycle the uncompact powder or fines. Recycling a compaction process is a significant draw back because of additional capital expenditure, labor costs, and increased throughput time.

The chief objective of roller compaction is to consistently make an agglomerate of sufficient strength that meets required density, granulometry, and powder flow specifications. Operationally, the key goal of a compactor is to maintain a range of pressure on the feedstock, independent of the fluctuating powder granulometry and flow fed into the rolls, so that a consistent compact is made.

Typically, the compaction process is managed by controlling the input material, the quantity per unit of time, the roll speed, and the roll gap. Allowing the roll gap to float unchecked can influence the production rate and the compact quality. Therefore, it is important to control the compaction process by setting a constant powder feed rate during the compaction operation.

Design innovations on the powder feed input side of the roller compactor are complex. The complexity is double-edged: powder materials, which flow and move well, are more easily handled on the feed side of the compactor. These types of materials generally do not necessarily need significant densification to marginally improve their flow handling characteristics on a large scale. On the other hand, poor flowing and low density powders require special equipment and considerations to feed a compactor. Additionally, it is necessary to maintain a constant powder flow and quantity of material to the rolls during the compaction process. Therefore, the delivery feed system plays a very important role in delivering poor flowing low density bulk powder materials to the roll compactor.

Johanson, in 1995, described an arch-breaking hopper design that effectively delivered poor flowing powder

from the hopper to a horizontal feed screw. The unique hopper design eliminated the “rat-holing” effect of poor flowing powders during hopper voiding and is described in Fig. 11.^[31]

Miller noted that the total compaction power requirement is the sum of the power required for the feed screw(s) and the roll drive.^[2] Feed screws not only convey the powder material from the compactor storage hopper but they also help deaerate the powder in the process. The deaeration of the powder acts as a mini compactor by precompacting the material just prior to roll compaction. Optimum compaction pressures and feed screw designs vary widely for different powder material properties. Changes in bulk powder density and feed screw speeds will affect the roll gap, the compaction pressure, the throughput, and the quality of the compact.^[2]

Weggel indicated that feed screw torque varies directly with the precompaction pressure.^[32] He suggested that by maintaining a constant feed screw pressure, the compactor operator can control the compact quality. Variations in precompaction pressure and in the compacting pressures are directly related to feed screw amperages and the roll drive motor. The compaction pressures are dependent on a continuous flow of powder into the feed screw area. If the powder feed flow is intermittent or interrupted, this will affect the feed screw amperage readings and eventually the roll drive readings. Weggel noted that force-feeding the material deaerates the feedstock and reduces the roll pressure loads. He found generally that it is more efficient

to achieve maximum densification during several stages of the roller compaction process.^[32]

Miller noted in 1996 that the design of vertical cylindrical feed screws tends to feed powder uniformly to the circumference of the feed screw.^[2] He concluded that the powder feeding to the rolls did not coincide with the general rectangular shape of the compactor throat. The observed draw back was that the powder did not get delivered evenly across the rolls; the middle of the rolls received more powder than the rolls' edge. This potentially created a poor quality compact; a strong middle compact that is weak at the ends because of frayed edges. Frayed compact edges give rise to uncompacted material and excess fines.^[2] An example of the phenomena was depicted when using a light yellow-green color feedstock, which after roller compaction using a twin feed screw system, produced a homogeneous colored compact. The same feedstock passed through a compactor designed with a single vertical feed screw, exhibited a sinusoidal colored compact as shown in Fig. 12.

Guigon and Simon, in 2001, showed a similar phenomenon using a horizontal single feed screw without vacuum deaeration when compacting sodium chloride.^[33]

Miller indicated that a multiple horizontal feed screw design system provided a more uniform powder distribution across the rolls in area and volume than a single vertical feed screw system. This concept applied to either the vertical or the horizontal single feed screw systems.^[2]

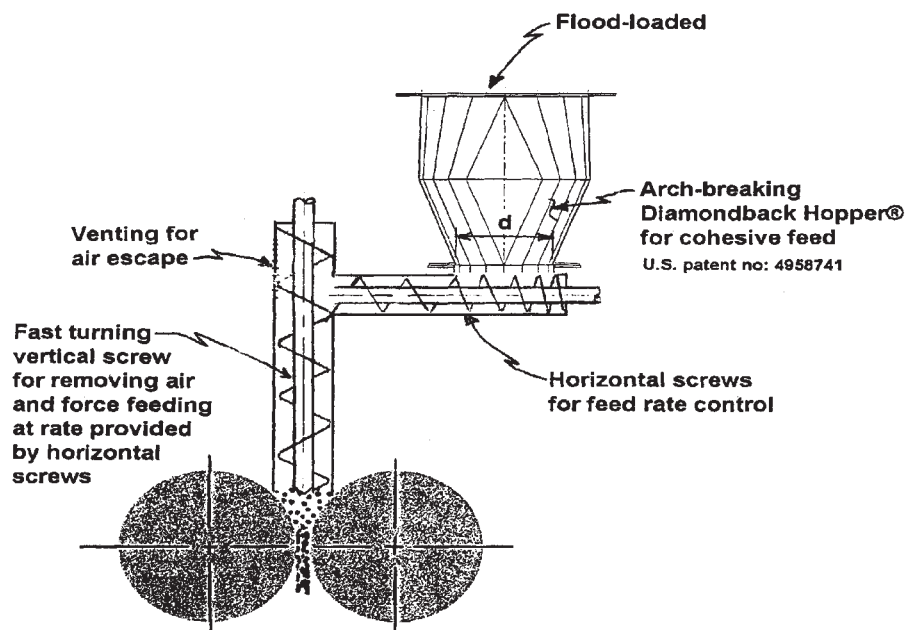


Fig. 11 Arch-breaking hopper design. (Courtesy of the Institute for Briquetting and Agglomeration.^[31])

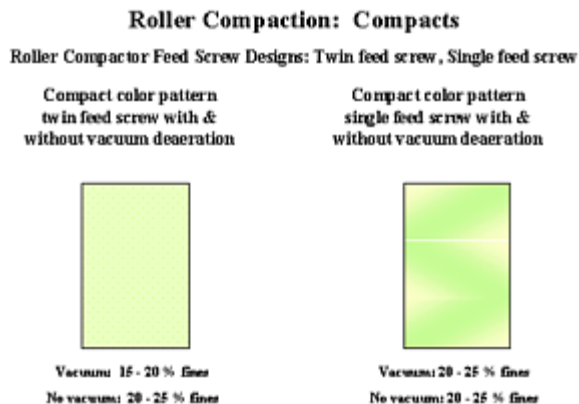


Fig. 12 Roller compaction: compacts. Roller compactor feed screw designs: twin feed screw and single feed screw. (Courtesy of R. W. Miller personal files.)

DEAERATION THEORY

A key factor limiting compaction throughput and quality is air entrapment in powder materials. During compression, air-occupying voids between particles are compressed and squeezed. The gas pushes through the powder causing powder fluidization and a nonuniform level of powder at the roll gap. It is best described in Fig. 13. This situation limits compact throughput and creates a nonuniform compact density. It also creates excess fines prior to sizing because of “spidering” compact edges.

The spidering condition occurs when gases rush across the inside of a compact to thinly and weakly formed flaked edges. The flake edges break apart perpendicular to the compaction direction. The compact edge breakage appears “saw tooth” in structure and varies in length depending on the nature of the powder binding properties, the amount of air entrainment, and the roll dwell time.

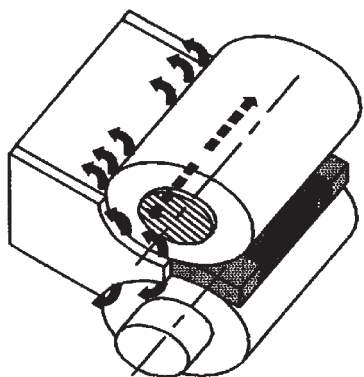


Fig. 13 Pattern of gas escape from roll nip region. (Courtesy of the Institute For Briquetting and Agglomeration.^[35])

Johanson predicted theoretical compactor operating conditions to handle air entrapment effects in materials. In general, he concluded that the critical compacting pressure level is dependent on a number of factors: roll speed, roll diameter, powder permeability, compressibility, and compact strength. Johanson indicated, when applying these principles in commercial application, a compactor operator would have to operate the press at slightly less than maximum pressures to allow for material inconsistencies and variable in-feed flow rates.^[34]

Pietsch, Johanson, and Dec described the forces at work in a roll press when powders are compacted.^[25,34,35] Pietsch, in Fig. 14, illustrated the three typical work phases during compaction.^[25] The first region is the initial solids' contact pressure with the feed screw. Pressure at this point is nil, p_0 . Later, as the solids begin to get pushed into the rolls and gripped by the rolls, pressure continues to build.

Ultimately the solids are moved into the narrowest point of the nip angle where the maximum pressure occurs at p_{max} . The second region describes the solid's density increase through compaction. As is shown in Fig. 14, densification occurs rapidly, in less than a second. Some materials have properties that after discharge from the compactor undergo elastic deformation, which can reduce the compact's density, expressed by the dotted curve. Third region of force at work is air pressure. Models such as the one described take into account that it is dependent on the porosity of the particulate solids compacted. If the porosity of the compacted material remains high enough during the compaction process, air pressure can escape and vent during and after compaction. On the other hand, if the material has low porosity, air pressure builds to high levels because it cannot escape easily. This can be seen in several ways, one of which is through expansion of the compact bursting. Bursting can be associated with a popping sound when operating the compactor. The second way occurs

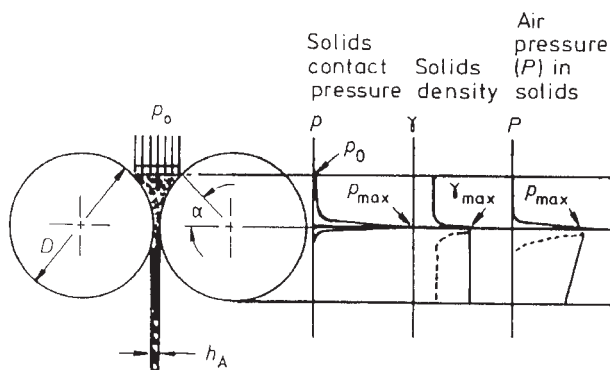


Fig. 14 Compaction work phases. (Courtesy of the Handbook of Powder Science.^[25])

when the compacted sheet breaks into slivers or “spidering” at the compact edges.

Both Johanson and Pietsch reported that expanding gas in a compact is detrimental to the compaction process: by reducing the compaction throughput and increasing the amount of fine particles. The effects of roll speed and powder porosity on air pressure in a compacted sheet are also illustrated in Fig. 15 by Pietsch.^[25] This graphic shows a relative large roll speed operating range when compacting a permeable (porous) powder. Air entrainment does not limit roller speed for coarse granular powders. On the other hand, when compacting very fine powders, the operating roll speed range is significantly reduced because of air entrainment.

Miller indicated that the evenness of the powder feed into the rolls determines, to a large extent, the success of compaction. Roller compactor systems suffer from two disadvantages: as the powder feed bulk density approaches 0.3 g cm^{-3} or less, the compaction throughput efficiency decreases. Secondly and concurrently, the uncompacted powder leakage generally increases around the rolls.^[22] Miller, in 1994, described a new machine design improvement that used vacuum deaeration to remove air entrainment from the powder just prior to the nip angle during roller compaction. The multiple benefits of such action are significant, and remarkable results have been observed when compacting low density raw materials.^[22]

- More uniform powder feed to the rollers.
- Less voltage and amperage variability for the roll pair.
- More uniform and strong compact.
- Less powder leakage.
- Greater yield.
- Less powder adhering to the compact prior to sizing.
- Higher compact throughput.
- Less airborne particles.

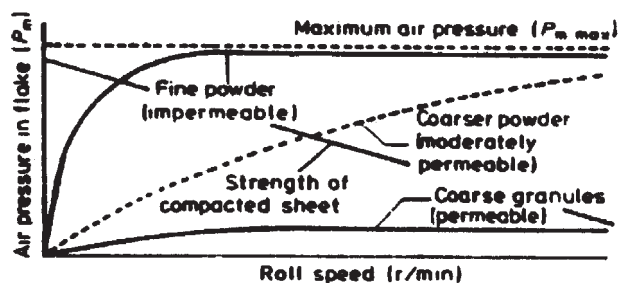


Fig. 15 Effects of roller speed and permeability on air pressure in the compacted sheet. (Courtesy of the Handbook of Powder Science.^[25])

The newly designed equipment involved a compactor fitted with two horizontal feed screws, which featured vacuum deaeration. Specifically, the roll compactor was equipped with a conical storage hopper containing a variable speed agitator. Bulk powder was fed directly from the top of the hopper to the top of twin horizontal auger feed screws, which directly transported the powder to the nip roll area, Fig. 16.^[2]

A novel stainless steel encasing that leads to the compactor rolls encloses the variable speed auger screws. Just before the nip area, a pair of sintered stainless steel segments are assembled within the horizontal auger feed system, which can operate under partial vacuum. A small, self-contained vacuum pump draws negative pressure through a dry filter and a stainless steel line to the sintered assembly plates. The partial vacuum is adjustable from -0.1 bar to -0.8 bar . The compaction rolls operate at different speeds and are supported on heavy-duty bearings in such a way that the lower roll is fixed and the upper roll is slightly movable in the vertical plane. The deaerated material passes through the roll pair, which is under infinitely variable hydraulic pressure. The deaeration, auger feed screws' design and speed, roll speed, and hydraulic roll pressures are the main factors in producing a compact with specified properties.

In several experiments, Miller studied the effects of using horizontal twin auger feed screws under partial vacuum (Alexanderwerk Inc., Horsham, Pennsylvania, model 50/75 compactor). The feed powder was vacuum deaerated just before roller compaction. The experimental design showed that the compactor's deaeration feed system significantly increased compaction output and minimized powder leakage when compacting very low density blends ($< 0.35 \text{ g cm}^{-3}$).^[22]

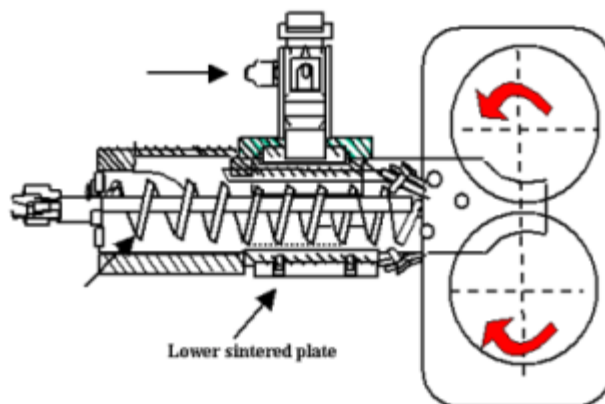


Fig. 16 Side view of feed screw system and vacuum deaeration with sintered plate segments. (Courtesy of Handbook of Pharmaceutical Granulation Technology.^[2])



To evaluate the effectiveness of the compactor's deaeration system, a test was designed to process a low density active drug blend with and without the activated deaeration system. The test showed how much material was compacted with the deaeration system engaged and how much material was compacted when it was not engaged. The test also determined how much material was not compacted (powder roll leakage) in each case. During multiple trials, weighed powder blends (0.25 g cm^{-3} – 0.35 g cm^{-3} density) were processed with the deaeration system engaged. The material was compacted at a specific screw feed rate, roll speed, and roll pressure. The compacted material was not sized. Instead, it was carefully collected on a 10 mesh screen to determine the compact and leakage quantities. In the first experiment, under the influence of vacuum deaeration, the compactor produced throughputs of 100 kg hr^{-1} and the resultant noncompacted material leakage rate was $< 2\%$. The vacuum deaeration was so effective, resulting in only limited leakage of noncompacted material, there was no need to recirculate the noncompacted powder back to the rolls to meet processing specifications.

Under a second set of compacting conditions, vacuum deaeration was not activated and the process typically produced 70 kg hr^{-1} – 80 kg hr^{-1} of densified compact. The powder leakage rate increased to 20%–30%. During this set of conditions, the powder flow to the roll pair was uneven and the ribbon compact was not uniform. Processing this formulation for a long period of time under these conditions would have required recirculation equipment to return the uncompacted powder to the rolls for further processing.^[22,36,37] The compactor parameters are described in Table 3 (Trials 1 and 2).

A second experiment was conducted with a different formulation involving the same active drug substance and the same active quantity per unit dose. This experiment replicated the design of the first experiment. The powder feed was processed and compacted with and without vacuum deaeration. The rates of compaction were

measured in the same manner as the first experiment. The compaction parameters were purposefully optimized at a roll speed of 8 rpm. The compactor yielded 150 kg hr^{-1} and the powder leakage rate was only 0.9% using the deaeration system. The compaction process typically produced 100 kg hr^{-1} – 110 kg hr^{-1} with 20%–30% powder leakage when vacuum deaeration was not in use. Table 4 (Trials 1 and 2) contains parameter settings and comparative results.^[22,36,37]

Both experiments clearly demonstrated the importance of vacuum deaerating the bulk powder just prior to compaction. In addition, the second formulation provided increased compaction throughput (with or without the vacuum deaeration) compared to the first formulation. The research formulators had improved the compaction rate by reducing the powder blend fluidization level (defined as the amount of powder airborne at a specific agitation rate) without changing the bulk density or other bulk parameters. Miller noted that the fluidization level appeared to be a key parameter that should be minimized to improve compaction yield rates.^[22]

In several other trials, Miller also studied the effects of using two different compactors' vacuum deaeration systems. The tests, similar to the ones previously described in this section, were carried out using differently designed compactors (employing and not employing vacuum deaeration). The objective of these trials was to increase the active bulk density of the feedstock from 0.22 g cm^{-3} to 0.61 g cm^{-3} .^[2]

Miller reported that the compactor, machine one, containing the horizontal twin screws and the vacuum deaeration design as described in Fig. 16, produced granules with 0.63 g cm^{-3} – 0.64 g cm^{-3} density at 80 kg hr^{-1} yield. When the vacuum deaeration was not activated the desired granule density specification could not be achieved. See Table 5 (Trials 1 and 2) for operating parameters and results of compactor machine design one.

A second compactor, machine 2, Fig. 17, did not densify the active drug granules during the first

Table 3 Compactor operating conditions and throughput yields for poor flowing low density active drug blend, formulation 1, vacuum deaeration and nonvacuum deaeration trials

Conditions	Trial 1	Trial 2
Powder density (g cm^{-3})	0.25–0.35	0.25–0.35
Screw feed (rpm)	52	52
Roll speed (rpm)	8	8
Vacuum (bar)	– (0.78–0.80)	0
Roll pressure (bar)	60–65	60–65
Compact rate (kg hr^{-1})	100	70–80
Compact leakage rate (kg hr^{-1})	2	15–20

(Courtesy of Handbook of Pharmaceutical Granulation Technology.^[21])

Table 4 Compactor operating conditions and throughput yields for poor flowing low density active drug blend, formulation 2, vacuum deaeration and nonvacuum deaeration trials

Conditions	Trial 1	Trial 2
Powder density (g cm^{-3})	0.25–0.35	0.25–0.35
Screw feed (rpm)	52	52
Roll speed (rpm)	8	8
Vacuum (bar)	– (0.78–0.80)	0
Roll pressure (bar)	60–65	60–65
Compact rate (kg hr^{-1})	150	100–110
Compact leakage rate (kg hr^{-1})	1.3	20–30

(Courtesy of Handbook of Pharmaceutical Granulation Technology.^[21])

Table 5 Operating parameters and results of compactor machine design 1 when compacting poor flowing low density active drug substance

Conditions	Trial 1	Trial 2
Initial density (g cm^{-3})	0.22	0.22
Powder density (g cm^{-3})	0.63–64	0.45
Screw feed (rpm)	52	52
Roll speed (rpm)	8	8
Vacuum (bar)	0.65	0
Roll pressure (bar)	50	50
Compact rate (kg hr^{-1})	80	42

(Courtesy of Handbook of Pharmaceutical Granulation Technology.^[2])

compaction run when employing vacuum deaeration. After a second compaction pass with vacuum deaeration, Trial 2, the powder density was increased slightly to 0.58 g cm^{-3} . It appeared that the level of vacuum being applied to the active drug substance, in machine 2, was insufficient, ineffective, or not optimally positioned, (even at its maximum vacuum deaeration level) to effectively help predensification.^[2] See Table 6 (Trials 1, 2, and 3) for machine two operating parameters and results.

Miller's experiments evaluated the effects of powder density, screw feed speed, roll speed, roll pressure, vacuum deaeration pressure, compaction rate, and the compaction leakage rate. Test results demonstrated that the first compactor's deaeration and feed system designs significantly increased compaction output. The new equipment design and process provided high compact yields and virtually eliminated powder leakage, obviating the need for expensive powder recirculation equipment. Compactor one's vacuum deaeration design (Fig. 16) proved to be

Table 6 Operating parameters and results of compactor machine design 2 when compacting a poor flowing low density active drug substance

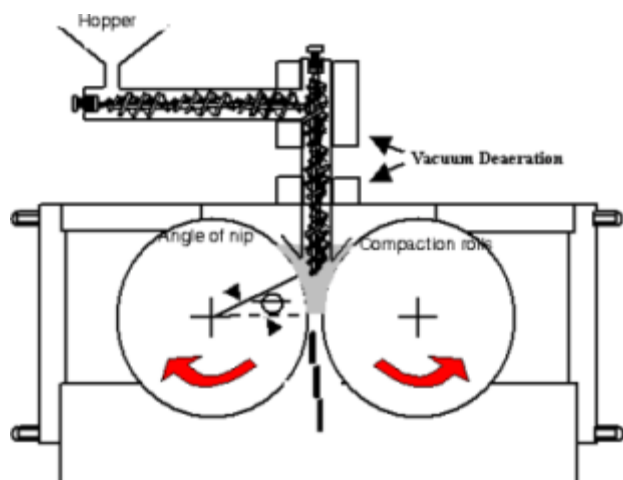
Conditions	Trial 1	Trial 2	Trial 3
Initial density (g cm^{-3})	0.22	0.45	0.22
Powder density (g cm^{-3})	0.45	0.58	0.46
Screw feed (horizontal) (rpm)	30	30	30
Screw feed (vertical) (rpm)	260	260	260
Roll speed (rpm)	6	6	6
Vacuum (bar)	0.34	0.34	0
Roll pressure (bar)	50	50	50
Compact rate (kg hr^{-1})	40	48	36

(Courtesy of Handbook of Pharmaceutical Granulation Technology.^[2])

superior to the second (Fig. 17) when compacting an active bulk drug with a density of approximately 0.2 g cm^{-3} . In summary, a new critical condition, vacuum deaeration, had been identified in optimizing roller compaction effectiveness and efficiency.^[2]

Miller concluded that four key processing conditions must exist to optimize roller compaction throughput and minimize powder leakage around the rolls:^[2]

- Adequate powder supply must enter the gripping zone.
- Powder must be fully conveyed into the narrowest part of the roller gap.
- Compaction pressure must be distributed as uniformly as possible over the whole of the roller-gripped powder mass.

**Fig. 17** Compactor front view of vertical feed screw system with vacuum deaeration. (Courtesy of Handbook of Pharmaceutical Granulation Technology.^[2])**Fig. 18** Computer controlled roller compactor with horizontal feed screws, vacuum deaeration, and cantilevered rolls. (Courtesy of Alexanderwerk Inc.)

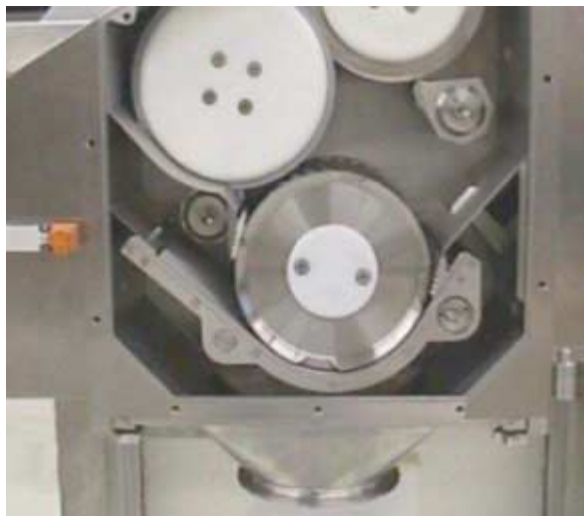


Fig. 19 Compactor attached sizing system. (Courtesy of Gerteis Co.)

- Sufficient vacuum deaeration must be effectively distributed prior to the nip roll region, particularly for low bulk density powder feed stock.

SIZING COMPACTS

There are no published studies describing compactor-milling studies. The theory of comminution or size reduction is well written about by Rekhi and Vuppala.^[38] They described the criteria for mill selection, mill classification, and variables affecting the sizing process and scale-up. The interested reader is also recommended to peruse Parrot, Johnson, and Lantz.^[39–41]

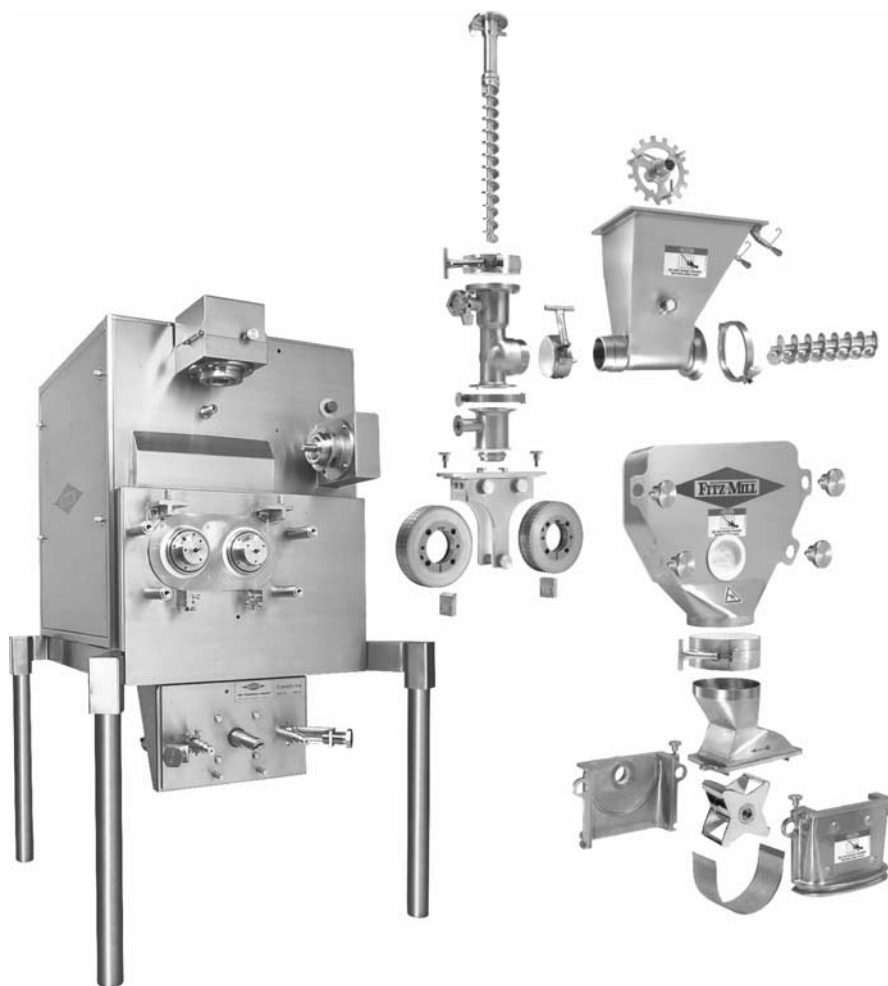


Fig. 20 Compactor and mill parts disassembled. (Courtesy of Fitzpatrick Co.)

EQUIPMENT FEATURES

A closer look at state-of-the-art roller compactors now available for the pharmaceutical industry is described in the following pages. Careful attention to detail was taken by the authors to represent the key industrial compactors' feature designs. Electronic pictures and equipment descriptions were obtained from the equipment manufacturers. This was the best way to fairly represent each company's machine design characteristics and features, facilitate the subject matter, and make the chapter factual and meaningful for the reader. All of the equipment innovations presented offer high levels of process control, throughput efficiency, safety, and a quality product. In this section, sizing equipment will be referenced but not described in detail.

The Alexanderwerk Inc. roller compactor (Fig. 18) is modular designed with a horizontal twin feed screw system that feeds two cantilevered rolls positioned vertically (6 and 12 o'clock). One roll is fixed and the other floats; the rolls can be cooled. The horizontal feed screws are specially enclosed in a stainless steel shroud that contains a unique vacuum deaeration system. Operator interface screens allow for on-line monitoring and controlling feed screw speed, roll gap and pressure, and sizing granulators. Diagnostic feature functions are displayed on-line, such as operation, maintenance, and calibration. All functions are interfaced and adjustable through PLC for process control and report monitoring. The two sizing granulators (coarse and fine units) are in series and integral to the compactor.

The Gerteis Co. roller compactor (Fig. 19) consists of three main units: feeding system, compaction, and granulating units. Feeding and tamping augers transport the powder to the compacting area and the machine is fitted with vacuum deaeration. Monitoring and controlling the roll force, gap distance, as well as the speed and torque distribution between the rolls can be adjusted through PLC for process controls. The rolls are mounted at an angle 30° to the horizontal plane; one roll is fixed, the other floats and they are cantilevered. The sizing unit is attached to the machine and consists of two prebreakers, a rotor with and interchangeable roll, and sieve basket. The system has interface capability to collect data for process documentation. The rolls can be cooled and the machine has CIP capability.

The Fitzpatrick Co. roller compactor (Fig. 20) features two feed screws in series: one horizontal and the second vertical. This system is designed to transfer powder bulk into a set of cantilevered rolls positioned in the horizontal mode; one is fixed, the other floats. Operator interface allows for on-line monitoring and controlling feed screw speeds, roll pressure, and roll gap. Features such as on-line help and diagnostic functions, maintenance, and calibration screens are displayed. All functions are interfaced

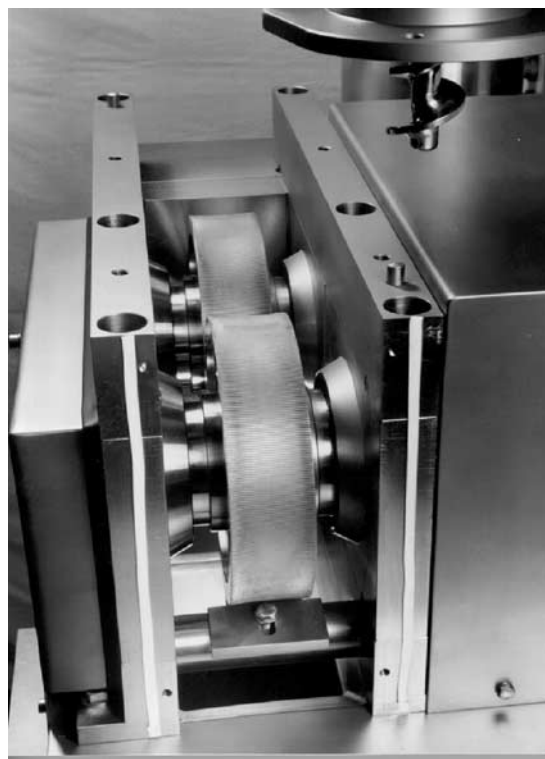


Fig. 21 Fixed rolls on two bearing blocks, rolls are concave. (Courtesy of Bepex-Hosokowa Co.)

and adjustable by PLC for process control and report generating. The machines have vacuum deaeration and roll cooling capabilities. The sizing unit is separate from the compactor and is either a rotor bar or a hammer mill.

Bepex-Hosokowa Co. compactor (Fig. 21) consists of a pair of rolls that are both fixed and not cantilevered. The rolls are mounted on a horizontal plane in fixed bearing blocks and there is a set fixed gap distance. The Bepex roll designs are concave and can be water cooled.

A specially designed compression feed screw and upper arm assembly are shaped to closely conform to the inner wall of the feed hopper. The feed screw (usually noncylindrical) has uniquely pitched edges. The feed screw design features and gravity aid to drive the powder-stock into the compaction area. There is no vacuum deaeration system. Predensification and deaeration occur in the screw feed region during operation.

Monitoring and controlling the roll speed and force and the feed screw speed and torque are adjusted through PLC for process controls. Prebreaker and flake crushers are attached sizing features. CIP capability exists.

Vector Co. compactor (Fig. 22) consists of cantilevered rolls that are mounted on a horizontal plane; one roll is fixed, the other floats. The feed screw designs are specially tapered and consist of noncylindrical and cylindrical



Fig. 22 Cantilevered compactor with one fixed and floating rolls, no vacuum deaeration, sizing device not attached. (Courtesy of Vector Co.)

shapes. Monitoring and controlling the roll force and speed, gap distance, as well as the feed screw speed and torque can be adjusted through PLC for process controls. There is no vacuum deaeration system. Predensification and deaeration occur in the screw feed region during operation. The Vector compactor rolls can be water cooled. The feed screw design features and gravity aid to drive the powder-stock into the compaction area.

ACKNOWLEDGMENTS

The authors gratefully acknowledge and thank Carol G. Miller for her contributions and assistance in editing and data manuscript control. The authors thank Dr. Pierre Guigon, Professor and Director of Chemistry at University of Technology of Compiègne France for his collaborative contributions to the text. The authors thank Mr. William Early of Global Pharmaceutical Technology Bristol-Myers Squibb for his text editing support. The authors acknowledge and thank the following persons and companies who contributed electronic files of photographs and pictures: Mr. Leonard Minervini, Alexanderwerk, Inc., Mr. Thomas Durig, Aqualon-Hercules Inc., Mr. Paul Gerteis, Gerteis

Maschinen AG, Mr. Scott Wennerstrum, Fitzpatrick, Co., Mr. Steven Kray, Vector Co., Mr. Albin Friedrich, Bepex-Hosokawa Inc.

REFERENCES

1. Kristensen, H.G.; Schaefer, T. Granulations. In *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boyland, J., Eds.; Marcel Dekker, Inc.: New York, 1993; Vol. 7, 121–126.
2. Miller, R.W. Roller Compaction Technology. In *Handbook of Pharmaceutical Granulation Technology*; Parikh, D.M., Ed.; Marcel Dekker, Inc.: New York, 1997; Vol. 81, 99–150.
3. Miller, R.W.; Sheskey, P.J. Survey of Current Industrial Practices and Preferences of Roller Compaction Technology and Excipients Year 2000. *Am. Pharm. Rev.* **2001**, 4 (1), 24–35.
4. Capes, C.E. Particle Size Enlargement. In *Handbook of Powder Technology*; Williams, J., Allen, T., Eds.; Elsevier: Amsterdam, 1980; Vol. 1.
5. Capes, C.E. Particle Size Enlargement. In *Kirk-Othmer Encyclopedia of Chemical Technology*, 3rd Ed.; Grayson, M., Eckroth, D., Eds.; Wiley: New York, 1978; Vol. 21, 77–105.
6. Pietsch, W.B. Size Enlargement by Agglomeration. In *Handbook of Powder Science and Technology*; Fayed, M., Otten, L., Eds.; van Nostrand Reinhold Co.: New York, NY, 1984; Chap. 7.
7. Pietsch, W.B. In *Agglomeration Techniques for the Manufacturing of Granular Materials with Specific Product Characteristics*; Roth, D., Ed.; The Institute for Briquetting and Agglomeration Proceedings; Gannon University: Erie, PA, 1997; Vol. 25, 49–164.
8. Sherrington, P.J.; Oliver, R. *Granulation*; Heyden & Son: London, 1981.
9. Kapur, P.C. *Adv. Chem. Eng.* **1978**, 10, 55.
10. Kristensen, H.G.; Schaefer, T. *Drug Dev. Ind. Pharm.* **1987**, 13, 803.
11. Lindberg, N. Industrial Wet Granulation. *Spl. Issue Acta Pharma. Suec.* **1988**, 25, 185–280.
12. Fonner, D.E.; Anderson, N.R.; Banker, G.S. Granulation and Tablet Characteristics. In *Pharmaceutical Dosage Forms: Tablets*; Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1982; Vol. 2, Chap. 5.
13. Anderson, N.R.; Banker, G.S.; Peck, G.E. Principles of Improved Tablet Production System Design. In *Pharmaceutical Dosage Forms: Tablets*; Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1982; Vol. 3, Chap. 1.
14. Ghebre-Sellassie, I. *Pharmaceutical Pelletization Technology*; Marcel Dekker, Inc.: New York, 1989.
15. Pietsch, W.B. *Roll Pressing*; Heyden: London, 1976.
16. Parrot, E.L. *Pharmaceutical Dosage Forms*; Marcel Dekker, Inc.: New York, 1990; Vol. 2, 203–204.



17. Dehont, F.R.; Hervieu, P.M.; Jerome, E.; Delacourte, A.; Guyot, J.C. Briquetting and Granulation by Compaction: A New Granulator-Compactor. In *Pharmaceutical Technology, Tableting Technology*; Wells, J., Rubinstein, M., Eds.; Ellis Horwood: London, 1993; Vol. 2 (Compression), 1–11.
18. Heckel, R.W. Density-Pressure Relationships in Powder Compaction. *Trans. Metallurgical Soc. AIME* **1961**, *221*, 671–675.
19. Johanson, J.R. Rolling Theory for Granular Solids. *Trans. Am. Soc. Mech. Eng.* **1965**, 842–848.
20. Jenike, A.W.; Shield, R.T. Plastic Flow of Coulomb Solids Beyond Original Failure. *J. Appl. Mech.* **1959**, *26*, 599–602.
21. Parrott, E.L. Densification of Powders by Concavo-Convex Roller Compactor. *J. Pharm. Sci.* **1981**, *70* (30), 288–291.
22. Miller, R.W. Advances in Pharmaceutical Roller Compactor Feed System Designs. *Pharm. Technol.* **1994**, *18* (3), 154–162.
23. Shileout, G.; Lammens, R.L.; Kleinebudde, P. Dry Granulation with a Roller Compactor Part 1: The Function Units and Operational Modes. *Pharm. Technol. Eur.* **2000**, 24–35.
24. Dehont, F.R.; Hervieu, P.M. Briquetting and Granulation by Compaction New Granulator Compactor for the Industry. *Drug Dev. Ind. Pharm.* **1989**, *15* (14–16), 2245–2263.
25. Pietsch, W. Size Enlargement by Agglomeration. In *Handbook of Powder Science, and Technology*, 2nd Ed.; Fayed, M., Otten, L., Eds.; Chapman & Hall: New York, NY, 1997; 347–364.
26. Funakoshi, Y.; Asogawa, T.; Satake, E. Use of a Novel Roller Compactor with a Concavo-Convex Roller Pair to Obtain Uniform Compacting Pressure. *Drug Dev. Ind. Pharm.* **1977**, *3* (6), 555–573.
27. Parrott, E.L. Densification of Powders by Concavo-Convex Roller Compactor. *J. Pharm. Sci.* **1981**, *70* (3), 288–289.
28. Jerome, E. Measurement of Resulting Forces on a Roller Compactor. *Drug Dev. Ind. Pharm.* **1991**, *17* (12), 1571–1591.
29. Johanson, J.R. Predicting Limiting Roll Speed for Briquetting Presses. *Proceedings of the 13th Institute for Briquetting and Agglomeration*, 1975; Vol. 13, 89–99.
30. Sheskey, P.J.; Hendren, J. The Effects of Roll Compaction Equipment Variables, Granulation Technique, and HPMC Polymer Level on a Controlled-Release Matrix Model Drug Formulation. *Pharm. Technol.* **1999**, *23* (3), 90–106.
31. Johanson, J.R. Roll Press Feed Systems. *Proceedings of the 24th Institute for Briquetting and Agglomeration*, October 1995; Vol. 24, 149–163.
32. Weggel, R.W. The Basics of Force-Feed Roll Briquetting. *Manuf. Eng.* **1982**, 107–109.
33. Guigon, P.; Simon, O. Correlation Between the Geometry of Feeding System and the Stress Distribution Applied on the Compact. In *Proceedings of the 27th Institute for Briquetting and Agglomeration*, Providence, RI, November 2001; Roth, D., Ed.; Gannon University: Erie, PA, 2001; Vol. 27, 31–41.
34. Johanson, J.R.; Cox, B.D. Fluid Entrainment Effects in Roll Press Compaction. *Proceedings of the 20th Institute for Briquetting and Agglomeration*, 1987; Vol. 20, 251–263.
35. Dec, R.T. Problems with Processing of Fine Powders in Roll Press. *Proceedings of the 24th Institute for Briquetting and Agglomeration*, October 1995; Vol. 24, 199–210.
36. Miller, R.W. Using Vacuum-Deaeration Feed System to Minimize Powder Leakage During Roll Compaction. *Powder Bulk Eng.* **1997**, *11* (2), 71–75.
37. Miller, R.W. Vacuum Deaeration Advances in Pharmaceutical Roller Compaction Technology. *Proceedings of the 24th Institute for Briquetting and Agglomeration*, October 1995; Vol. 24, 165–173.
38. Rekhi, G.S.; Vuppala, M.K. Sizing of Granulation. In *Handbook of Pharmaceutical Granulation Technology*; Parikh, D.M., Ed.; Marcel Dekker, Inc.: New York, 1997; Vol. 81, 389–418.
39. Parrot, E.L. Milling. In *The Theory and Practice of Industrial Pharmacy*; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea and Febiger: Philadelphia, 1986; 21–46.
40. Johnson, C. Comminution Variables and Options. *Powder Bulk Eng.* **1989**, *3*, 40–44.
41. Lantz, R.J., Jr. Size Reduction. *Pharmaceutical Dosage Forms: Tablets*; Marcel Dekker, Inc.: New York, 1990; Vol. 2, 107–157.

SCALE-UP AND POSTAPPROVAL CHANGES (SUPAC)

Jerome P. Skelly

Consultant, Alexandria, Virginia, U.S.A.

OVERVIEW

In the process of developing a new drug product, the batch sizes used in the earliest human studies are small. As one proceeds through Phase 1 testing (i.e., the first introduction of a new chemical entity to humans), Phase 2 (discovering an indication for use), and Phase 3 (determining dose, side-effect profile, etc.), the size of the batches is gradually increased. When a New Drug Application (NDA) is approved by the Food and Drug Administration (FDA), the drug product is scaled up to a significantly larger batch size to meet the demands of the anticipated market. Similarly, in the development of a generic version of an already approved marketed product, a small batch is produced and tested for, among other things, bioequivalence to the FDA reference listed drug product. When the generic product meets FDA approval criteria, the Abbreviated New Drug Application (ANDA) or generic antibiotic application (AADA) is approved for marketing. It, too, is then scaled up to meet the demands of its anticipated market.

Whether a new chemical entity being brought to market for the first time or an approved generic version of previously marketed product, the size of the batch is almost inevitably scaled up to a significantly larger batch. In the process of scaling up, certain changes in the formula (composition) and/or in the manufacturing process and/or in the equipment may be necessary. In addition, the site at which the product will be manufactured may differ from where the smaller (pilot) batches were manufactured. The scale-up process and the changes made after approval in the composition, manufacturing process, manufacturing equipment, and change of site have become known as Scale-Up and Postapproval Changes, or SUPAC. The FDA has issued various guidances for SUPAC changes designated SUPAC-IR (1) (for immediate-release solid oral dosage forms), SUPAC-MR (2) (for modified-release solid oral dosage forms), and SUPAC-SS (3) (for non-sterile semisolid dosage forms including creams, ointments, gels, and lotions).

Although scale-up may occur at any point in the lifetime of a product, it most often occurs after the firm has been notified that the drug product is approvable, i.e., it meets all the conditions required by the FDA for marketing. With

the submittal of Final Printed Labeling, a showing that the marketed product will meet the conditions for marketing as approved by the FDA (and in the case of generics, production of three consecutive scaled-up batches), and satisfactory completion of a preapproval inspection by the local FDA district office, the product is formally approved to be manufactured and sold in the United States. At this point, SUPAC begins to exert its effect.

Although SUPAC is a means of decreasing regulatory burden by empowering industry to make regulatory decisions, it does not affect any compliance or inspection requirement. It also is limited to scale-up and postapproval changes, even though the underlying science applies to preapproval changes as well. The major affect of SUPAC is a significant decrease in the time required to implement changes.

BACKGROUND

For years, the FDA had approved generic drug applications (ANDAs), using bioequivalence as a surrogate for clinical effectiveness. Aware that changes in the scale of the batch or the manufacturing equipment could affect a product's bioavailability, the FDA required that batches used in the ANDA-submitted bioequivalence studies be "production-sized batches made on production equipment." The FDA failed to verify the truthfulness of this statement, which had been submitted in virtually every ANDA and AADA. At one point, however, FDA inspectors noted that one firm, having obtained approval of a generic version of an approved marketed drug, had great difficulty scaling up that product to market-size batches. On investigation, the FDA discovered that the production-size batch made on production equipment consisted only of a few hundred dosage units produced on laboratory equipment. Although the samples from this very small batch when tested were bioequivalent to the innovator product, the firm could not scale up to the million plus units per batch they wished to market.

The FDA immediately promulgated rule 22-90, requiring that the minimum batch size for bioequivalence studies would henceforth be 100,000 units or 10% of

the anticipated market batch size, whichever is greater. Although this rule solved an immediate problem, it was not based on science. To address the question, the FDA used a method previously used to address quite controversial issues involving controlled-release dosage forms (4, 5) and bioanalytical methods (6). With the American Association of Pharmaceutical Scientists, it held three public workshops, bringing together outstanding industrial, academic, and FDA scientists to openly discuss the scientific and regulatory issues. The discussions were followed by the publication of consensus White Papers (7–9).

SUPAC

The premise of the consensus White Papers was that if: 1) the source of the drug substance for the smaller and larger batches was the same; 2) the drug substance particle size (both mean and distribution) was the same; 3) the excipients were the same; 4) the excipient particle size (both mean and distribution) was the same; 5) the order of addition was the same; 6) the equipment was the same; 7) the processing was the same; and, most important, 8) a surrogate test for bioequivalence testing (dissolution) was the same, the two batches were indeed the same. Over the previous 20 years the FDA Biopharmaceutics Program had established that within definable limits, dissolution was predictive of in vivo bioequivalence, for the same formulation, processed under the same conditions, on the same equipment. These criteria became the fundamental principle of the SUPAC initiative. (The percutaneous diffusion test is similarly used as a surrogate bioequivalence test for nonsterile, semisolid formulations.)

To establish the validity of the approach recommended by the three consensus papers, the FDA contracted the College of Pharmacy of the University of Maryland to study several drug products chosen on the basis of their solubility and permeability. The data revealed that the workshop recommendations were conservative and could be safely implemented. In fact, the studies showed that even broad differences in in vitro dissolution that resulted from major compositional changes failed to translate into bioavailability differences. Subsequently, the FDA published its *SUPAC Guidance for Immediate Release Solid Oral Dosage Forms* and followed with guidances for modified-release (controlled-release) and nonsterile semisolid dosage forms. In November 1999 (modified slightly in December 1999), the FDA extended the SUPAC concept to address changes in analytical methodology, packaging, and labeling and sterile semisolid dosage forms (10). This last guidance also updated the previously published guidances

on immediate-release, modified-release, and nonsterile, semisolid dosage forms. In particular, the issue of multiple postapproval changes (which had been addressed differently in the previously published guidances) were now the same. The FDA now allowed multiple postapproval changes for every solid oral dosage form, using the same requirements as its SUPAC Semisolid Guidance.

The SUPAC Guidances published by the FDA define various levels of change and for each level of change specifies the 1) recommended chemistry, manufacturing, and control tests; 2) in vitro dissolution testing and/or in vivo bioequivalence tests; and 3) documentation that the FDA requires to be filed in the NDA, ANDA, or AADA to support the change. These guidances do not affect other compliance or inspection documentation required by the FDA Center for Drug Evaluation and Research Office of Compliance (CDER-OC) or the FDA field investigation units.

DOCUMENTATION

An annual report must be filed for every active application pending at the FDA. This report, which must be filed annually, updates all activity covered by the various FDA regulations affecting drug products under development or approved for marketing in the United States. Any change to an FDA-approved product must be submitted in a Supplemental New Drug Application, which requires review and approval. Until SUPAC, all supplements were Prior Approval Supplements (PAS). However, with SUPAC, the Agency allowed the use of a new Changes Being Effectuated (CBE) supplement, which allows industry to implement certain moderate changes on the very day the supplement is sent to the FDA. This eliminated the previously required and often lengthy FDA prereview and comment period. On the other hand, the Agency, after its review, reserves the right to require a PAS rather than a CBE if it disagrees with the industry decision. When the FDA published its new and updated guidance in December 1999, it provided for a new supplemental application called Changes Being Effective in 30 days (CBE-30). This procedure primarily applies to sterile dosage forms for which the Agency, obviously, requires a more exacting level of review.

LEVELS OF CHANGE

SUPAC lists three levels of change (Fig. 1) including: 1) minor changes, which are unlikely to have any

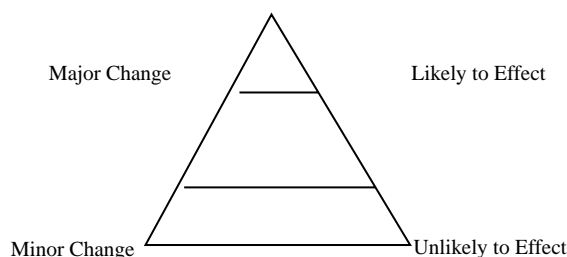


Fig. 1 Levels of change to FDA-approved products.

detectable effect on formulation quality and/or performance; 2) moderate changes, which could have a significant effect on formulation quality and/or performance; and 3) major changes, which are quite likely to have a significant effect on formulation quality and/or performance of the dosage form.

For a minor change, the chemical testing required and submitted to the FDA is routine. For example, accelerated stability testing and a commitment to perform long-term stability testing, plus those tests contained in the compendia and/or in the NDA/ANDA/AADA, would have to be performed. The data must be filed in the next annual report. On the other hand, a major change could involve dissolution-testing profiles to be determined in several media, an *in vivo* bioequivalence study, and a requirement for all data to be submitted in a supplemental filing to the application, which would require FDA preauthorization before implementation.

CHEMICAL, MANUFACTURING, AND CONTROLS TESTS

The FDA chemical testing for immediate-release, solid oral dosage forms (including the surrogate dissolution test) takes into consideration whether the drug has a narrow therapeutic index. The chemical, manufacturing, and control testing (including dissolution) is more onerous for narrow therapeutic index drug products than for other immediate-release, solid oral dosage forms.

In addition, although the testing criteria for changes in the composition of nonrelease-controlling excipients in delayed release (i.e., enteric-coated) dosage forms and controlled-release (i.e., extended) dosage forms are very similar to those for immediate-release, solid oral dosage forms, the criteria for those excipients affecting the release rate are, understandably, more onerous.

On the other hand, the comparability of the drug substance and excipients with those of the innovators product listed by the FDA as the "Reference Listed Drug"

(RLD), assumes (under the FDA's SUPAC guidelines) considerably greater importance for sterile and nonsterile semisolid formulations than, for solid oral dosage forms. With the semisolid dosage forms, the FDA uses various types of diffusion analyses in lieu of dissolution as a surrogate for *in vivo* bioequivalence.

In general, semisolid dosage forms are complex formulations having complex structural elements. Often, they are composed of two phases (oil and water), one of which is a continuous (external) phase and the other of which is a dispersed (internal) phase. The active ingredient is often dissolved in one phase, although occasionally the drug is not fully soluble in the system and is dispersed in one or both phases, creating a three-phase system. The physical properties of the dosage form depend on various factors including the size of the dispersed particles, the interfacial tension between phases, the partition coefficient of the active ingredient between phases, and the product rheology. These factors combine to determine the release characteristics of the drug as well as other characteristics such as viscosity.

Although for a true solution, the order in which solutes are added to the solvent is usually unimportant, this is not true for dispersed formulations. Because dispersed matter can distribute differently depending on the phase to which a particulate substance is added, the order of addition for these formulations is of critical importance. Any change in the order of addition, therefore, is a major change.

Over a period of years, the FDA, monitoring the performance of a variety of physical and chemical tests commonly performed on semisolid products and their components (e.g., solubility, particle size, and crystalline form of the active component; viscosity; and homogeneity of the product), has determined that these tests provide evidence of consistent performance. Although the evidence available at this time is less convincing than *in vitro* dissolution as a surrogate for *in vivo* bioequivalence in the case of solid oral dosage forms, the FDA is using diffusion testing (i.e., *in vitro* release-rate testing) to assess product "sameness" to allow certain SUPAC for semisolid products.

STABILITY

In addition, the FDA requires that every marketed product be stable, meeting its approved specifications throughout its marketed shelf-life. Generally, for minor changes, in addition to the application or compendial chemistry and manufacturing control requirements

(CMC), long-term stability data from one batch must be filed when available in the annual report. For moderate changes, the general requirement is for accelerated (3-month) stability data from one batch filed in either a CBE supplement or a PAS, plus long-term stability data from one batch filed in an appropriate annual report. For major changes, in addition to the long-term stability data for one batch filed in an annual report, 3-month accelerated stability data must be submitted in a PAS for either one or three batches, depending on how long that particular product has been in the marketplace. For new chemical entities, the period is 5 years, whereas for a new dosage form of an approved chemical entity, the period is 3 years. If the product meets this market criterion, a "substantive body of information" is said to exist, and data from one batch are required to be submitted. If a product does not meet this marketing criterion, a "significant body of information" is said not to exist, and data from three batches are required to be submitted. In any case, the data must be filed in a supplement requiring FDA preapproval (i.e., PAS).

SURROGATE MEASURES OF EFFICACY

For 30 years, the FDA has satisfactorily used in vivo bioequivalence as a surrogate of clinical efficacy for the development of new dosage forms and strengths of innovative products and for approval of generic versions of approved marketed drugs. It additionally requires human in vivo bioequivalence studies for most major postapproval changes, as well as for moderate postapproval changes for narrow therapeutic index drugs. In those cases where an in vivo/in vitro correlation has been established for a particular product, it may be used in lieu of in vivo bioequivalence.

However, the FDA uses in vitro dissolution as a surrogate for in vivo bioequivalence testing for most minor and moderate postapproval changes. For implementation of minor changes, the dissolution requirement is usually the same as that required by the FDA for the release of an approved product into the marketplace. On the other hand, for moderate changes, the FDA usually requires additional in vitro dissolution testing. The type of dissolution study required for solid oral dosage forms varies depending on several drug variables, e.g., solubility, permeability, therapeutic index, and dosage form.

Minor changes may be implemented for drugs that are both soluble (at the largest strength marketed) and permeable (bioavailability >90%) if the product meets

the SUPAC single-point dissolution criteria of 85% in 15 minutes. Those that fail this test must meet a more onerous multipoint dissolution test in which samples are taken every 15 min until either 85% is dissolved or an asymptote is reached, thus establishing a dissolution rate profile. Modified-release preparations must additionally meet the profile criterion in water and buffered aqueous solutions. The dissolution profiles comparing product manufactured after the change with that manufactured before the change must be comparable. The FDA accepts comparability if the profiles meet the FDA "F-2" similarity test criterion (11). Failure to establish comparable dissolution profiles would normally necessitate human in vivo bioequivalence studies or use of an in vivo/in vitro correlation. In addition delayed-release solid oral dosage forms must pass the compendial 2-h acid-dissolution test.

In the case of semisolid dosage forms, the FDA uses a standard open-chamber diffusion cell fitted with a synthetic membrane as the in vitro release test to be used as a surrogate for in vivo bioequivalence studies. It uses this test for semisolid dosage form SUPAC changes, just as it used dissolution in the case of solid oral dosage forms. The FDA guidance recommends using the test on all creams, ointments, and gels. In this test, a plot of the amount of drug released per unit area of membrane ($\mu\text{g}/\text{cm}^2$) versus the square root of time should yield a straight line. The regression (slope of the line) represents the drug release rate. Comparisons of the release rate before and after making the change are required. In the case in which the change is not similar, in vivo bioequivalence data would be required to implement the change.

POSTSCRIPT

SUPAC is a revolutionary change in drug regulation. For nearly all of the 20th century, congressional action increased industry's regulatory burden in response to what was perceived as industry's irresponsible actions. This was accompanied by an ever-increasing cost to develop and market new drugs and a concomitant increase in the time required to obtain marketing approval from the FDA. In the last 2 decades, many perceived that the United States was falling behind in the drug development process and that, as a result, Americans were not able to obtain new state-of-the-art treatments for serious diseases in timely manner. With the passage of the Food and Drug Administration Modernization Act (FADAMA), Congress allowed the FDA to employ

additional scientists while at the same time mandated management efficiencies in the drug review process. SUPAC is a reversal of the ever-increasing authority of the U.S. drug regulatory body. It empowers the industry to take responsibility for its actions, while simultaneously allowing adequate overview of the drug development and marketing process by the FDA.

Because of FADAMA, the SUPAC process has been extended by the FDA to include sterile dosage forms, analytical methods, product labeling, and product packaging postapproval changes.

Two important updates of previously published SUPAC guidances include the change in multiple SUPAC changes and the redefinition of "site of manufacturer". For all SUPAC guidances, multiple postapproval changes are now permitted. The requirements that must be fulfilled default to those for the most restrictive change proposed. For example, three level-one changes and one level-two change would have to meet the more restrictive level-two criteria.

In addition, although the definition of "same site/different site" will remain the same for foreign manufacturers until a May 14, 1999, proposed regulation requiring registration of foreign manufacturers who want to market drugs in the United States is published in final form, the definition for domestic manufacturers has been changed. In the latter case, Same Site is defined as the site in which the new and old buildings are included under the same drug establishment registration number, and the operations in both are inspected by the same FDA district office. Different site is defined as the site where the new and old buildings have different drug establishment registration numbers or when the operations are inspected by different FDA district offices.

REFERENCES

1. *Immediate Release Solid Oral Dosage Forms: Scale-Up and Post Approval Changes*, Center for Drug Evaluation and Research (CDER): March 1995; FDA Guidance for Industry.
2. *SUPAC-MR—Modified Release Solid Oral Dosage Forms: Scale-Up and Post Approval Changes*, Center for Drug Evaluation and Research (CDER): Sept 1997; FDA Guidance for Industry.
3. *Non-Sterile Semisolid Dosage Forms: Scale-Up and Post Approval Changes*, Center for Drug Evaluation and Research (CDER): May 1997; FDA Guidance for Industry.
4. Skelly, J.P.; Barr, W.; Benet, L.; Doluisio, J.; Goldberg, A.; Levy, G.; Lowenthal, D.; Robinson, J.; Shah, V.; Temple, R.; Yacobi, A. Report of a Workshop on Controlled Release Dosage Forms: Issues and Controversies. *Pharm. Res.* **1987**, 4 (1), 75–77.
5. Skelly, J.P.; Amidon, G.L.; Barr, W.; Benet, L.; Carter, J.R.; Robinson, J.; Shah, V.; Yacobi, A. In Vitro and In Vivo Testing and Correlation for Oral Controlled/Modified Release Forms. *J. Pharm. Sci.* **1990**, 79, 849–854.
6. Shah, V.; Midha, K.K.; Dighe, S.; Skelly, J.P.; Layloff, T.; McGilveray, I.J.; Viswanathan, C.T.; Yacobi, A. Analytical Method Validation: Bioavailability, Bioequivalence, and Pharmacokinetics Studies. *Pharm. Res.* **1992**, 9, 588–590.
7. Skelly, J.P.; Van Buskirk, G.; Savello, D.R.; Amidon, G.L.; Arbit, H.M.; Dighe, S.; Fawzi, M.B.; Gonzalez, M.; Malick, A.W.; Malinowski, H.; Nedich, R.; Pearce, D.M.; Peck, G.E.; Schwartz, J.B.; Shah, V.P.; Shangraw, R.F.; Truelove, J.P. Scale-Up of Solid Oral Immediate Release Dosage Forms. *J. Pharm. Res.* **1993**, 10, 313–316.
8. Skelly, J.P.; Van Buskirk, G.A.; Savello, D.; Augsburger, L.; Theeuwes, F.; Leeson, L.; Lesko, L.; Barr, W.H.; Amidon, G.; Gonzalez, M.; Malinowski, H.; Berge, S.; Fox, D.; Wheatly, T.; Nixon, P.; Porter, S.; Clevenger, J.; Dighe, S.; Arbit, H.; Pearce, D.M.; Peck, G.E.; Jerussi, R.; Hoiberg, C.; Schwartz, J.; Shah, V.P.; Robinson, J.; Shangraw, R. Scale-Up of Oral Extended Release Dosage Forms. *Pharm. Res.* **1993**, 10, 1800–1805.
9. Van Buskirk, G.A.; Shah, V.; Adair, D.; Arbit, H.M.; Dighe, S.; Fawzi, M.; Feldman, T.; Flynn, G.L.; Gonzalez, M.; Gray, V.; Guy, R.; Herd, A.K.; Hem, S.L.; Hoiberg, C.; Jerussi, R.; Kaplan, A.S.; Lesko, J.; Malinowski, H.; Meltzer, N.; Nedich, R.; Pearce, D.; Peck, G.E.; Rudman, A.; Savello, D.; Schwartz, J.B.; Skelly, J.P.; Vanderlaan, R.K.; Wang, J.C.; Weiner, N.; Winkel, D.; Zatz, J.L. Workshop III Report: Scale-Up of Liquid and Semisolid Disperse Systems. *Pharm. Res.* **1994**, 11 (8), 1216–1220.
10. *Changes to an Approved NDA or ANDA*, US, DHHS, FDA, Center for Drug Evaluation and Research (CDER), Nov 1999; FDA Guidance for Industry.
11. *Dissolution Testing of Immediate Release Solid Oral Dosage Forms* US, DHHS, FDA, Center for Drug Evaluation and Research (CDER), Aug 1997; FDA Guidance for Industry.

SECONDARY ELECTRON MICROSCOPY IN PHARMACEUTICAL TECHNOLOGY

Peter C. Schmidt

University of Tübingen, Tübingen, Germany

BASIC PRINCIPLES OF ELECTRON MICROSCOPY

Introduction

Basic principles of imaging

The principle of an electron microscope is based on the light microscope except that electrons are used instead of light. The resolving power of any microscope is given by Abbe's equation:

$$d_0 = \frac{0.61 \cdot \lambda}{n \sin \alpha} [\text{nm}] \quad (1)$$

where d_0 = minimum resolvable separation distance; λ = wavelength of the light; n = refractive index of the medium between the object and the objective lens; and α = half-angle subtended by the objective at the object.

The lower the minimum resolvable distance the higher the resolution of the instrument. Therefore, the resolution is enhanced by a short wavelength of the light, a high refractive index of the medium between the objective lens and the object, and a small distance between the object and the lens. The product $n \sin \alpha$ is called the numerical aperture of the objective lens. Since α can never exceed 90° , an objective in air can never resolve distances smaller than 0.61λ . The limits of a light microscope are given in Table 1.

De Broglie (2) showed that an electron has a dual character. It can be regarded either as a moving charged single particle or as a radiation with a distinct wavelength. The relation between the two is given by Eq. 2:

$$\lambda = \frac{h}{mv} [\text{nm}] \quad (2)$$

where λ = wavelength; h = Planck's constant; m = mass of the electron; and v = velocity of the electron.

The velocity of the electron depends on the voltage applied and can be used to calculate the wavelength, λ , and the minimum resolvable distance, d_0 , as shown in Table 2.

The advantage of the use of electrons as a light source compared with visible light is obvious.

Types of electrons in electronic imaging

When an electron beam is interacting with a specimen surface, different reactions can occur. The various signals arising from the specimen's surface on incidence of the primary electron beam are shown in Fig. 1.

Secondary electrons (SEs): A collision of electrons from the primary electron beam with the surface of the specimen results in a detachment of the so-called SEs. The number of SEs depends on the surface topography, the accelerating voltage, and the atomic number of the surface elements. They create the SE current, which is collected for imaging.

Backscattered electrons (BSEs): These are primary electrons that have been reflected from the specimen surface in a way that they return back out of the specimen again. Depending on the individual collision, they have energies ranging from the full primary energy of the electron beam down to the level of secondary electrons. Their intensity increases with increasing atomic number. They give information on the topography of the sample and the atomic number of the sample elements.

X-ray photons (X): Photons of X-radiation are emitted from the specimen under electron bombardment. They are characteristic for elements and could be used to determine the element distribution on the surface of the sample under investigation.

Transmitted electrons (TEs): TEs are those penetrating through a thin specimen that are focussed into images on a phosphor luminescence screen or on a photosensitive material.

Absorbed electrons (AEs): Some of the primary electrons are absorbed by the specimen and are grounded as a current. If there is no ground connection, the specimen will charge up.

Cathodoluminescence (CL): This is a light emission of the specimen under electron bombardment. The light could be visible or invisible.

In secondary electron microscopy (SEM), in most cases, SEs are used for imaging. BSE detectors are closely connected with element analysis of the specimen. Other types are more or less seldom used in electron microscopy.

Table 1 Minimum resolvable distance, d_0 , of a light microscope depending on the wavelength λ , the refractive index n , and the angle α

Wavelength λ (nm)	Angle α (°)	Refractive index n	Minimum resolving distance d_0 (nm)
800	15	1.000 ^a	1885
400	15	1.000 ^a	942
800	30	1.000 ^a	976
400	30	1.000 ^a	488
800	60	1.000 ^a	563
400	60	1.000 ^a	281
800	60	1.516 ^b	371
400	60	1.516 ^b	185

^aAir.
^bOil.
(Adapted from Ref. 1.)

Table 2 Variation of electron velocity, wavelength, and minimum resolvable distance with accelerating voltage for an electron microscope

Acceleration voltage (kV)	Velocity (km/h)	λ (nm)	d_0^a (nm)
1	18, 370	0.03876	0.0169
10	58, 460	0.01220	0.0053
100	164, 400	0.00370	0.0016

^aAccording to Eq. 1, assuming that $n \cdot \sin \alpha = 1.4$.
(Adapted from Ref. 3.)

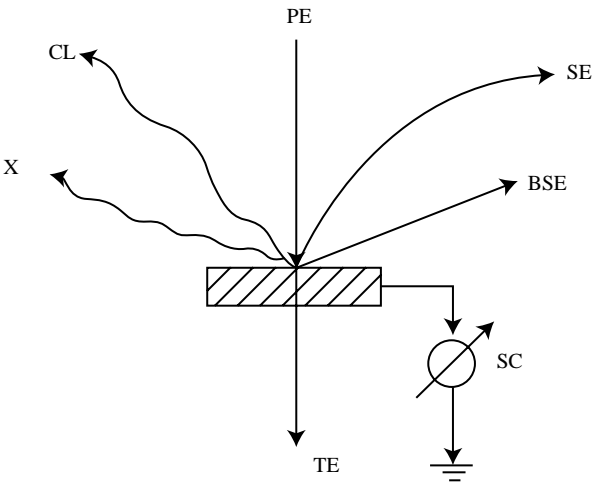


Fig. 1 Signals arising from the specimen’s surface on incidence of the primary electronic beam (4), where PE = primary electronic beam; SE = secondary electrons; BSE = backscattered electrons; X = X-ray photons; TE = transmitted electrons; SC = absorbed electrons; and CL = cathodoluminescence.

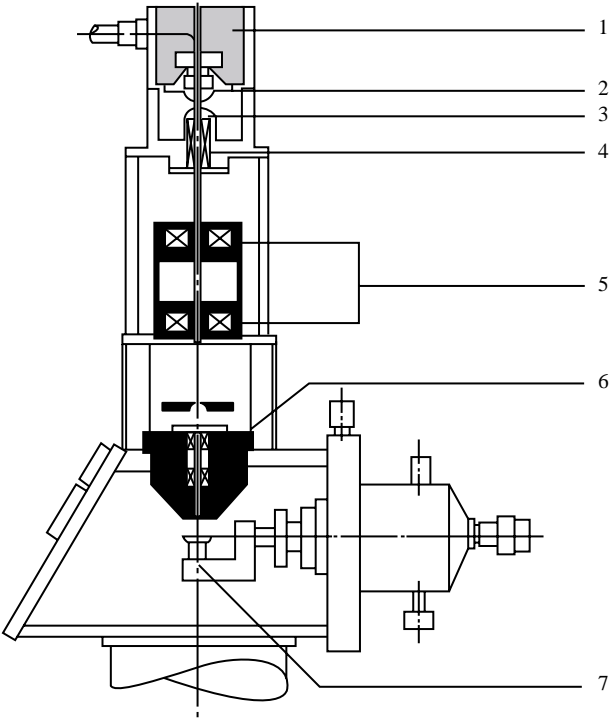


Fig. 2 The design of a secondary electron microscope (4), where 1 = electronic gun; 2 = Wehnelt cylinder; 3 = anode; 4 = beam alignment coils; 5 = condensor lenses; 6 = objective lens; and 7 = specimen holder.

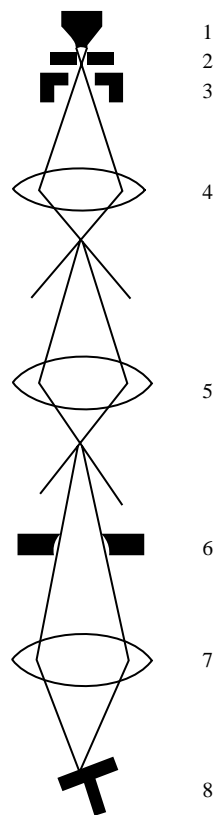


Fig. 3 The electronic beam arrangement modified from (4), where 1=cathode; 2=Wehnelt cylinder; 3=anode; 4 + 5=electromagnetic condenser lenses; 6=aperture; 7=objective lens; and 8=specimen surface.

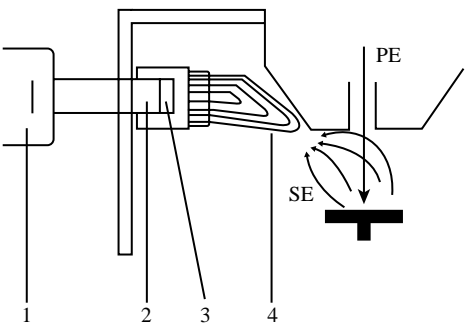


Fig. 5 The secondary electron detector (4), where 1=photomultiplier; 2=light guide; 3=scintillator; 4=collector; PE=primary electron beam; and SE=secondary electron beam.

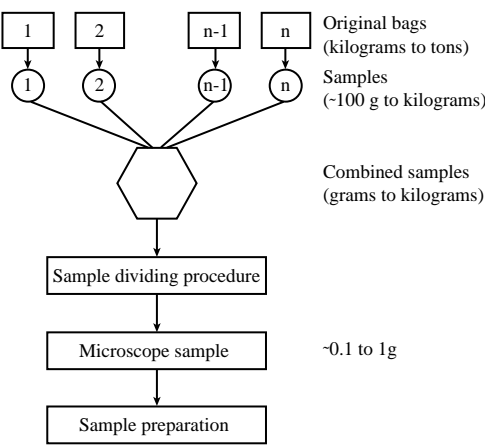


Fig. 6 Sampling procedure to create a small sample ready for microscopic examination.

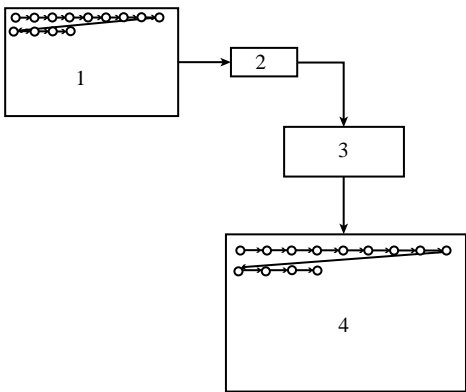


Fig. 4 Principle of the scanning mode of a secondary electron microscope (4), where 1=specimen surface; 2=detector; 3=video processing; and 4=monitor screen.

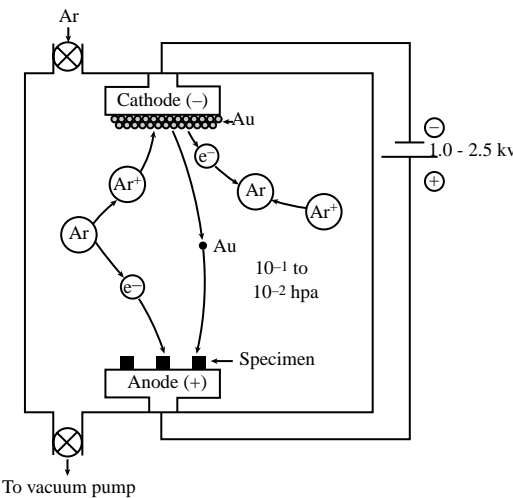


Fig. 7 Principle of the cathodic sputtering process.

In transmission electron microscopy (TEM), transmitted electrons will be processed to give an image.

Secondary Electron Microscope

Among the different types of electron microscopes such as the secondary electron microscope (SEM), the transmission electron microscope (TEM), the scanning transmission electron microscope (STEM), and the field emission scanning transmission electron microscope (FESTEM), the SEM finds greater application in the field of pharmaceutical technology, followed by the TEM. This article will focus only on the principle of the SEM.

The design of a SEM is shown in Fig. 2.

It consists of the electronic gun (1), the Wehnelt cylinder (2), the anode (3), and beam alignment coils (4) on the top of the instrument. The condensor lenses (5), the aperture, and the objective lens (6) focus the beam onto the specimen that is mounted on the specimen holder (7). The latter one could be moved in X-, Y-, and Z-direction within the specimen chamber. In addition, the sample could be moved by rotation. The arrangement to create the electronic beam is shown in Fig. 3.

The electrons are emitted by the cathode (1) from a filament, normally a heated tungsten wire. The emitted electron current is controlled by a Wehnelt cylinder (2) having a negative polarity against the cathode. The electrons are picked up by the anode (3) after being

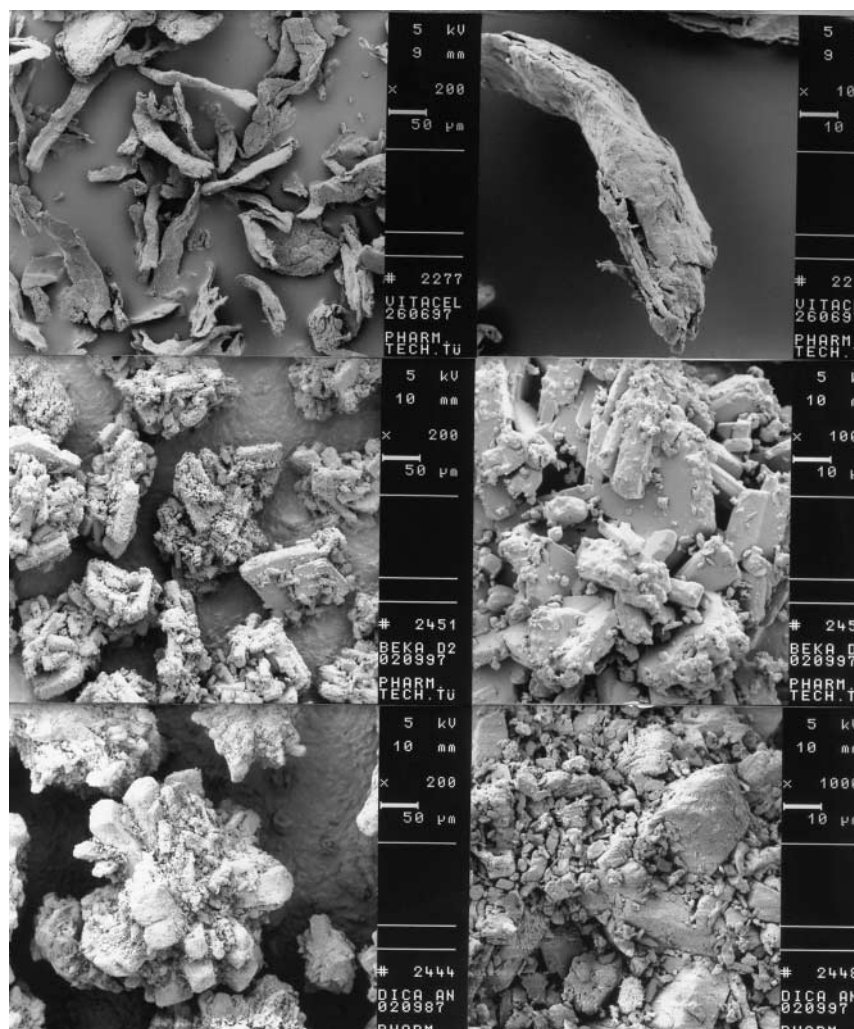


Fig. 8 Microcrystalline cellulose (Vitacel) (top), dicalcium phosphate dihydrate (Bekapress D2) (middle), and dicalcium phosphate anhydrous (Dicaphos AN) (bottom).

accelerated by the accelerating voltage of approximately 1–25 kV. The cathode assembly and the anode are arranged so as to produce a crossover of the electron beam between the components. Through the bore hole of the anode the electron beam enters two electromagnetic condensor lenses (4 and 5), which reduce the crossover. After passing the aperture (6), the electron beam is focussed by the objective lens (7) so that the focal spot is imaged on the specimen surface (8).

The specimen is scanned point by point by the electron beam as shown in Fig. 4. Secondary electrons that are emitted from the specimen's surface (1) are collected by the detector (2) and undergo a video processing (3) leading to the image formed on the video screen (4).

The electron beam is moved on the surface of the specimen by an electromagnetic deflection system that is integrated in the objective lens moving the beam in a raster over the specimen as mentioned above. The deflection system consists of two sets of crossed saddle coils for deflection in X and in Y direction. The saddle coils produce distortion-free images at lowest magnifications and permit large deflection angles.

The secondary electron detector (Fig. 5) (4) is mounted to the side of the microscope chamber. Primary electrons

(PEs) liberate secondary ones (SEs) from the specimen's surface, which are caught by the collector (4) having a positive potential for the detection of SE. The electrons pass through the grid and move toward the scintillator (3). This is biased to +10 kV and accelerates the low-energy SEs to a higher energy level. These electrons strike the scintillator, where they generate photons, which are guided out of the detector chamber through a light guide (2) to the photomultiplier (1). The photomultiplier converts the light current by amplification again into an electron current that presents the video signal at the output of the subsequent preamplifier. If the collector is at a negative potential, only backscattered electrons (BSE) can strike the scintillator and a BSE image is produced in this mode.

Electrons move over long distances only under vacuum. The whole system, therefore, has to be operated under a vacuum of 10^{-5} to 10^{-7} hPa. The vacuum system consists of a rotary pump giving a vacuum up to 10^{-3} hPa combined with a water-cooled turbomolecular pump leading to the desired high vacuum. In some instruments, an oil-diffusion pump is used instead of a turbomolecular pump. An automatic control ensures that the system cannot be run without water cooling and at a low vacuum level.

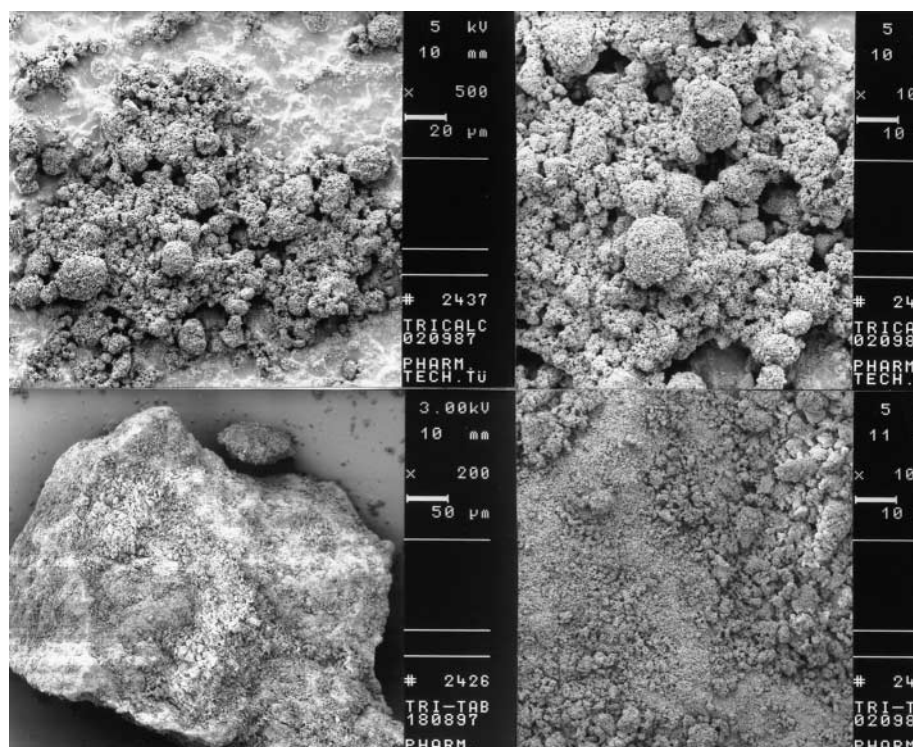


Fig. 9 Tricalcium phosphates (hydroxyapatite). Tricaphos, a spray-dried product (top), and Tritab, prepared by roller compaction (bottom).

SAMPLE PREPARATION IN SCANNING ELECTRON MICROSCOPY

Introduction

Most of the samples being investigated by SEM in the field of pharmaceutical technology are powders prepared by different methods, granules, pellets, tablets, and films from coated tablets. The sample preparation of bulk materials includes the following steps:

1. choosing a representative sample from the bulk material;
2. mounting the sample on a suitable sample holder (specimen); and

3. preparing the specimen for observation by coating, if it is not self-conducting.

In pharmaceutical technology, materials are handled quite often as a bulk. Due to the fact that electron microscopic imaging is based on very small samples, the sampling procedure itself and the preparation of the final sample become very important. An overview for sampling procedures is given by Sommer (5). A scheme for powder sampling is given in Fig. 6.

The initial samples from the original bags are normally taken by hand. Depending on the size of the bag they have to be taken from different places. They are then combined by mixing before undergoing the sampling procedure.

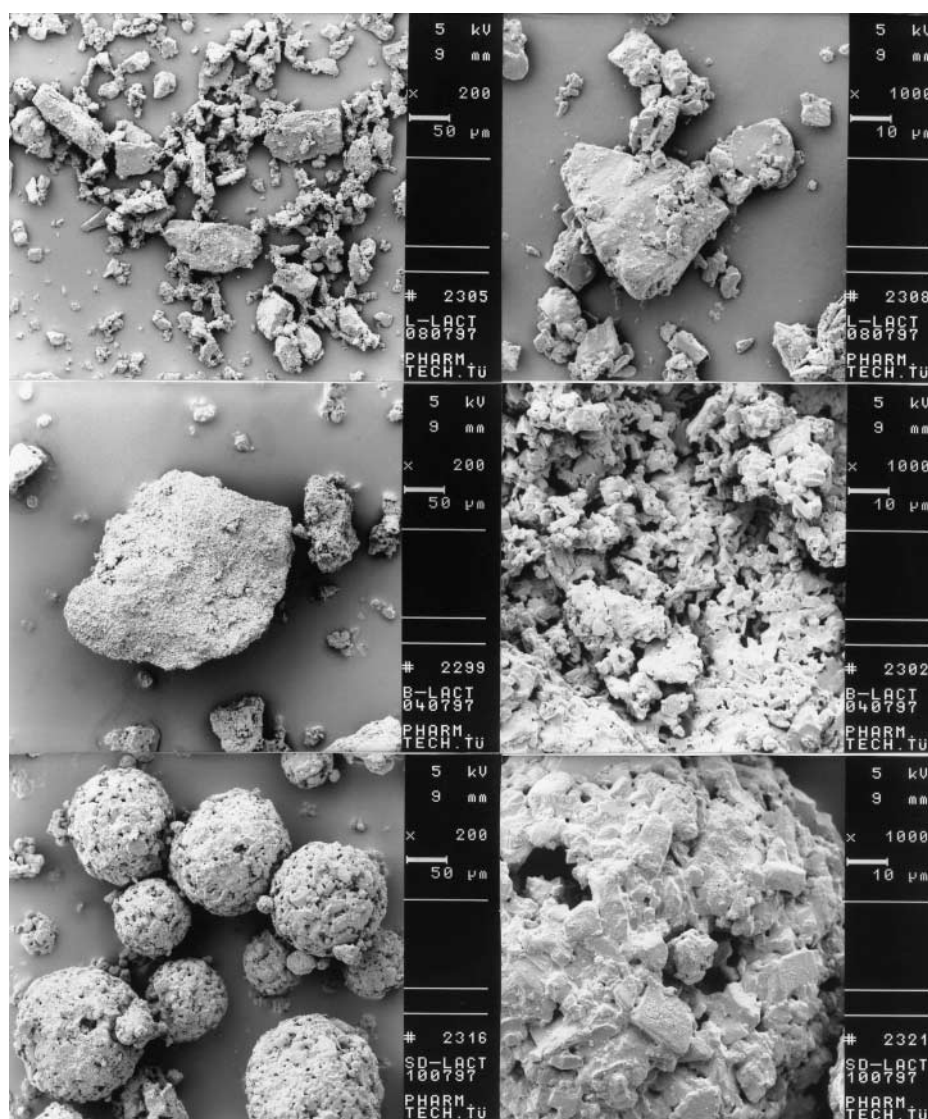


Fig. 10 Lactoses. α -lactose monohydrate (top), β -lactose (middle), and spray-dried lactose (bottom).

Today, these procedures are carried out by sampling machines like a rotary sample divider.

Depending on the size of the combined sample, the procedure has to be done in several steps if necessary. At the end a sample size directly applicable to the microscopic sample preparation must be achieved.

Specimen Preparation for Secondary Electron Microscopy

In pharmaceutical technology, most of the samples that are examined by an SEM are powders. After the sampling procedure, the final sample has to be distributed

uniformly onto a stub and later on—if it is not self-conducting—covered by a layer of a conducting material. The stubs are either pin-type mushroom or cylinders normally made from aluminum. In most of our work, simply metal rivets were used instead of the expensive aluminum stubs offered by the suppliers of SEM accessories. The stubs have to be cleaned carefully by a surfactant solution followed by distilled water and acetone in an ultrasonic bath. Finally, the traces of acetone should be removed by a warm air stream. The surface of the stubs is then covered by a double-adhesive tape. Conductivity of the tape is ensured by a small droplet of conducting silver fluid, which is placed between one end of the tape and the stub. Coarse powders

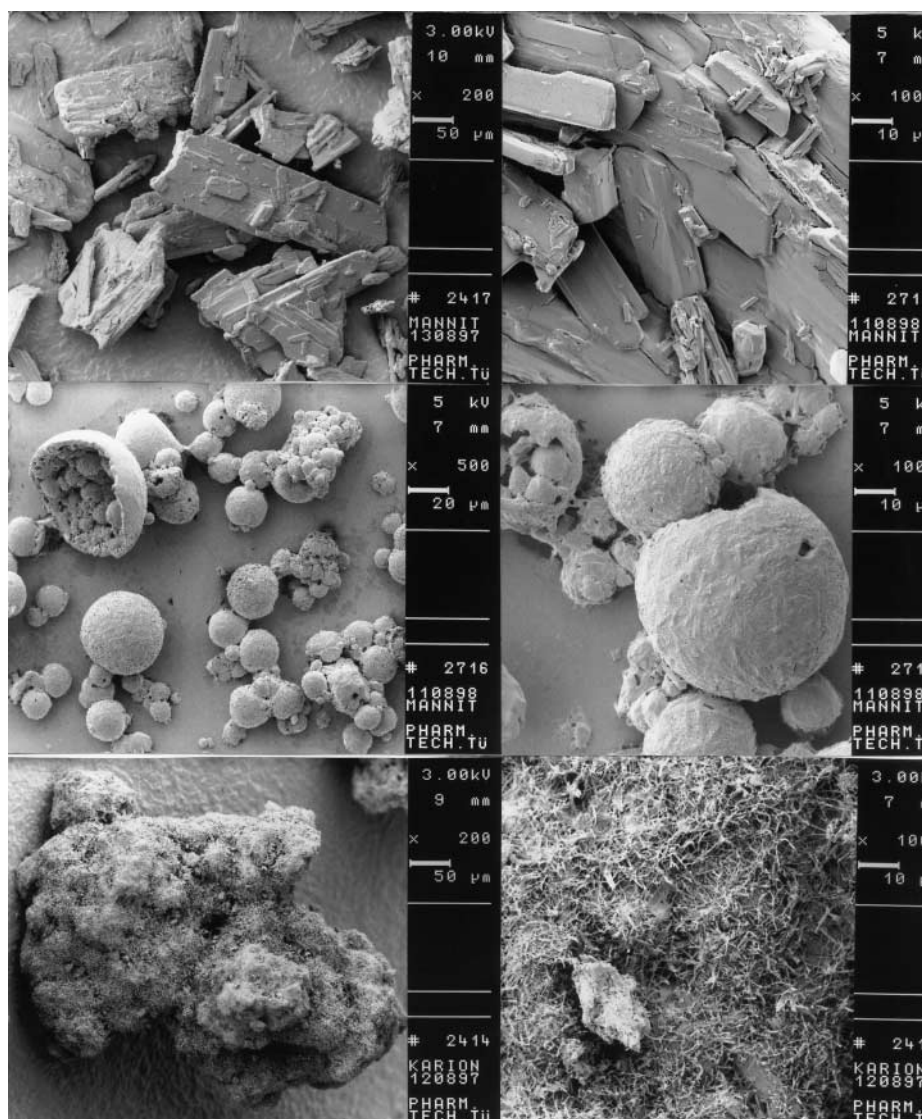


Fig. 11 Sugar alcohols. Mannitol, crystalline (top), mannitol, spray dried (middle), and sorbitol-instant, spray dried (bottom).

are directly sprinkled onto the so-prepared stubs, and fine ones have to be desagglomerated. Depending on the degree of agglomeration this could be done either by blowing the material onto the stub by an air stream or by sucking it through a pipe by vacuum. Also, the “Rhodos” dry powder desaggregation unit of the Sympatec-Helos Particle Sizer (6) could be used to blow the powder onto a stub. Other samples like tablets, coated tablets, pellets, packaging materials, and needles of syringes can be directly mounted to the stub by the use of colloidal silver paste, bonding the sample to the stub mechanically and electrically. For other preparation methods the reader is recommended to refer to a special literature (7).

A part of the primary electron beam, amounting 10^{-12} to 10^{-6} A, is emitted as secondary or backscattered electron current. The difference between the two currents must be allowed to leak away to earth via the stub, otherwise the specimen will charge up. If a sample is not self-conducting it has to be provided with a thin conducting coat. Only at a low accelerating voltage in

the range of 1 kV, a sample does not need to be conductive. The coating is performed under vacuum either by evaporating carbon, gold/palladium, or platinum or by d.c. sputtering. The latter is the most popular and easy technique to apply conductive coatings to a sample. The principle of sputtering is shown in Fig. 7.

The apparatus consists of an evacuated chamber, a cathode, and an anode. The air within the chamber is replaced by an inert gas, preferably argon, and the voltage applied to the system is in the range of 1.0–2.5 kV. Under these conditions a glow discharge is set up between cathode and anode. Neutral gas atoms are ionized to positively charged argon ions and electrons. The ions are attracted by the cathode, where they cause an emission of a gold atom and an electron from the cathode material. The electron neutralizes an argon ion, the Au atom is moving toward the anode. On the way there are a lot of collisions with gas atoms and as a result the Au atoms arrive at the anode from different directions where they settle onto the specimen's surface. Free electrons are accelerated in

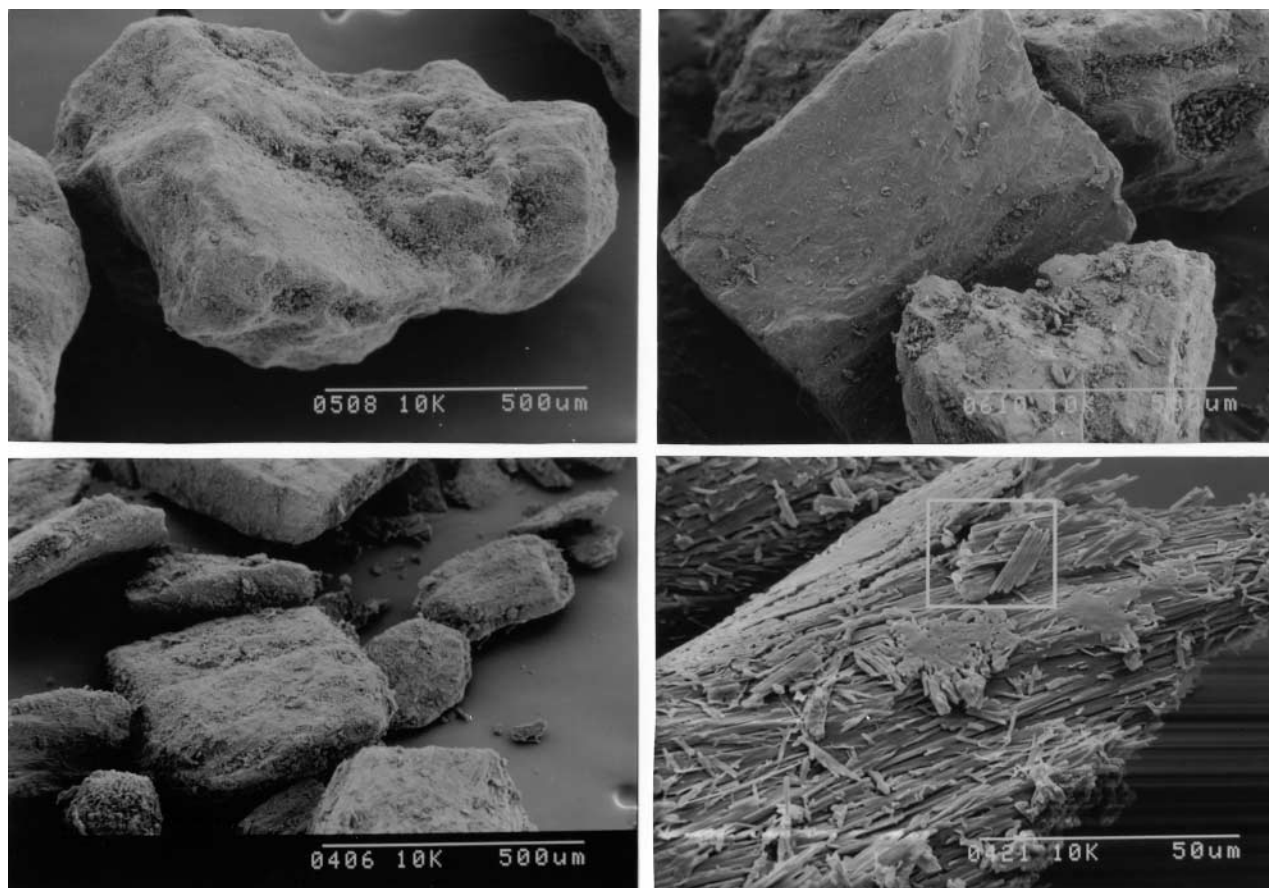


Fig. 12 Different types of sorbitol.

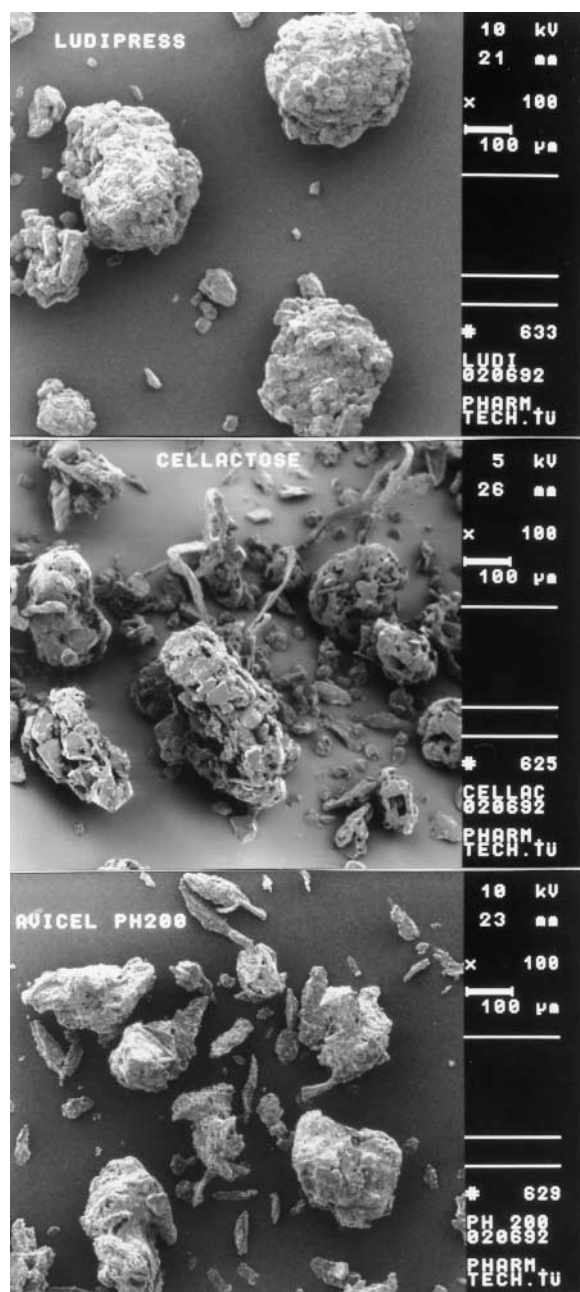


Fig. 13 Ludipress (93% α -lactose monohydrate, 3.5% soluble povidone, 3.5% cross-linked povidone) (top), Cellactose (75% α -lactose monohydrate, 25% powdered cellulose) (middle), and Avicel PH 200 (agglomerated microcrystalline cellulose) (bottom).

the direction of the anode, and when reaching the anode, create heat in the specimen.

The process itself and the deposition of metal atoms onto the specimen's surface depends on the target material, the type and pressure of the inert gas within the

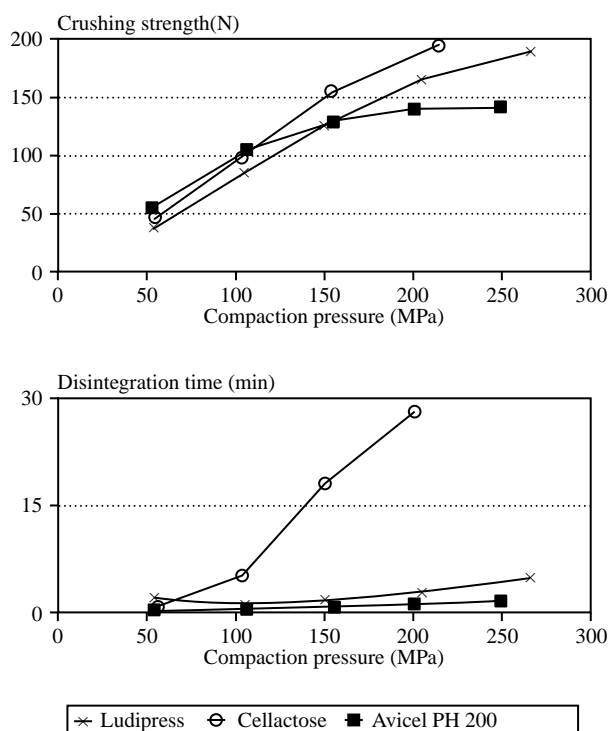


Fig. 14 Compressional pressure/hardness-profile and compressional pressure/disintegration profile of the coprocessed materials, Ludipress, Cellactose, and Avicel PH 200.

chamber, the cathode potential, the current, the time of sputtering, and the distance between the target, and the sample. The preferred gas is argon, having the lowest price of all inert gases. Air could not be used because it causes oxidation of metals, e.g., gold. A black coating indicates the presence of air during the sputtering process. With argon the sputtering rate of gold and silver is high compared with carbon. The coating thickness is determined by the sputtering rate, the distance between the target and the sample, and the time. There is a linear correlation between the thickness of the gold layer and the sputtering time at constant conditions.

The quality of the coat has to be judged by the homogeneity of the thickness, adhesion onto the surface of the sample, and its mechanical stability. Plain surfaces result in a uniform coating thickness. At rough surfaces, a uniform coating can only be achieved when the metal atoms settle diffuse from all directions. This could be predetermined by the vacuum applied. The optimum coating thickness is a compromise between the conductivity of the coated sample and the resolution of fine surface structures. If the conductivity is not high enough, a charge up of the specimen is observed. For

routine imaging a coating thickness of 1 to 2.0 nm is sufficient.

APPLICATION OF SECONDARY ELECTRON MICROSCOPY IN PHARMACEUTICAL TECHNOLOGY

The application of SEM in the field of pharmaceutical technology is widespread. Solid starting materials for any kind of formulation can be visualized as well as powder

mixtures, granulations, pellets, tablets, coatings, spray- and freeze-dried products, microparticles, liposomes, and packaging materials. Although there are articles dealing with SEM investigations of raw materials (8–10) and some other materials, a general review on the applications of SEM in pharmaceutical technology is not available. Thus, this section attempts to fill the gap in the literature. The following examples were taken from our own experience over 13 years. They do not cover the whole field, but they have been selected to demonstrate that a lot of problems can be solved by the view through a microscope.

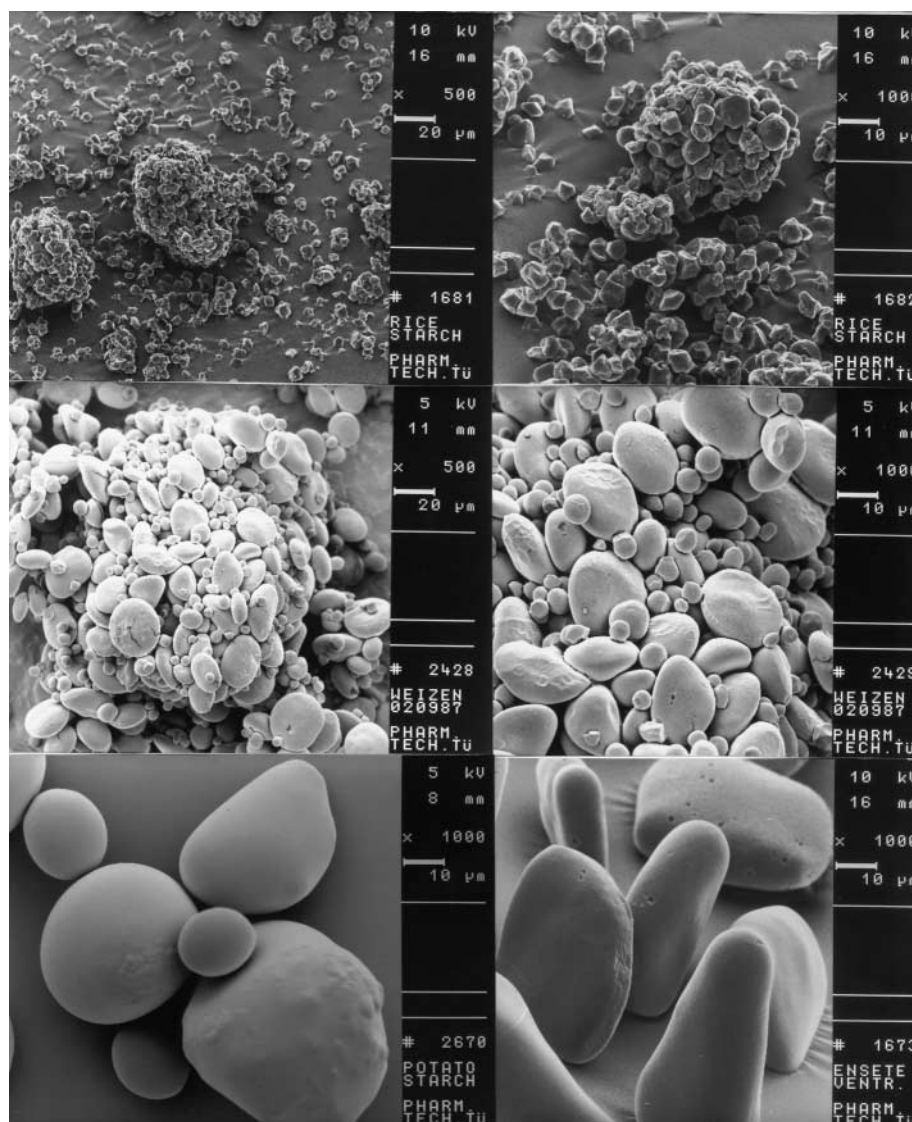


Fig. 15 Starches used in pharmacy. Rice starch (top), wheat starch (middle), potato starch (bottom left), and enset starch (false banana) (bottom right).

Starting Materials

Filler/binders

Filler/binders are used in direct compression of tablets. The requirements they have to fulfill are summarized by Khan and Rhodes (11). An important factor in the evaluation of filler/binders is their so-called “dilution

potential,” i.e., giving an amount of active ingredient that can be taken up by the filler/binder to produce a tablet according to previously fixed specifications.

Filler/binders belong to the following classes of substances: microcrystalline and powdered celluloses, lactoses, phosphates, particularly calcium phosphates, sugar alcohols, and the so-called “co-processed materials,”

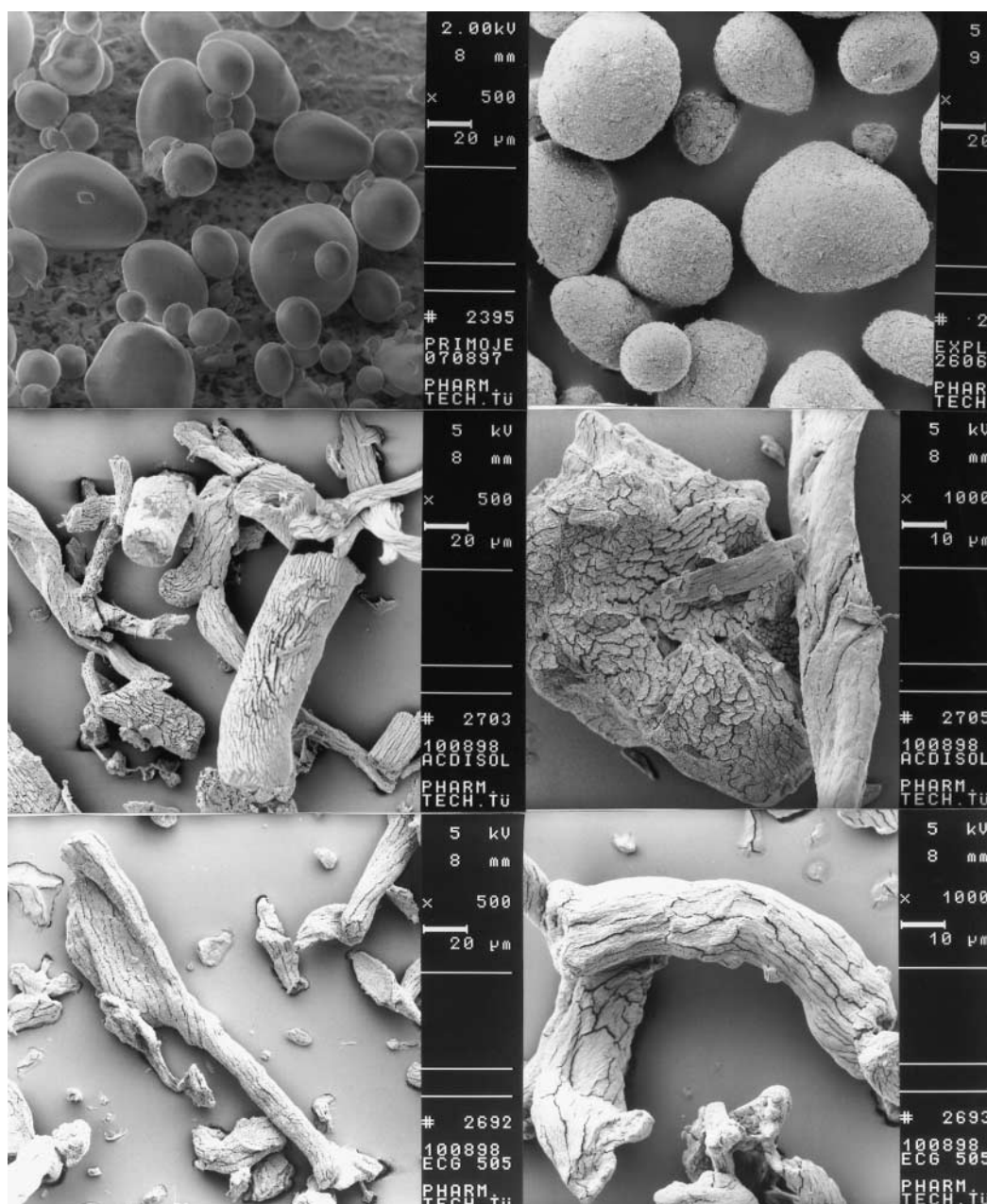


Fig. 16 Disintegrants. Primojel (top left), Explotab, both sodium starch glycolates (top right), cross-linked carboxymethyl cellulose (middle), and calcium carboxymethyl cellulose (bottom).

which are especially designed combinations for direct compression (12). The following figures (Figs. 8–13) show filler/binders at different magnifications. Figure 8 (top) depicts microcrystalline cellulose (Vitacel). The fibrous structure is typical for all products of this type and differs significantly from the calcium phosphates shown below. The dicalcium phosphate dihydrate (Bekapress D2) in the middle is composed of small agglomerates, whereas

the anhydrous type (Dicaphos AN) is agglomerated to a higher extent. In Fig. 9 two types of tricalcium phosphates are presented. Crystallographically both are hydroxyapatites. Tricaphos (top) is a spray-dried product showing typically rounded particles built from very small crystallites, whereas Tritab is a granulation prepared by roller compaction (bottom). For detailed description of calcium phosphates see the relevant literature (13–16).

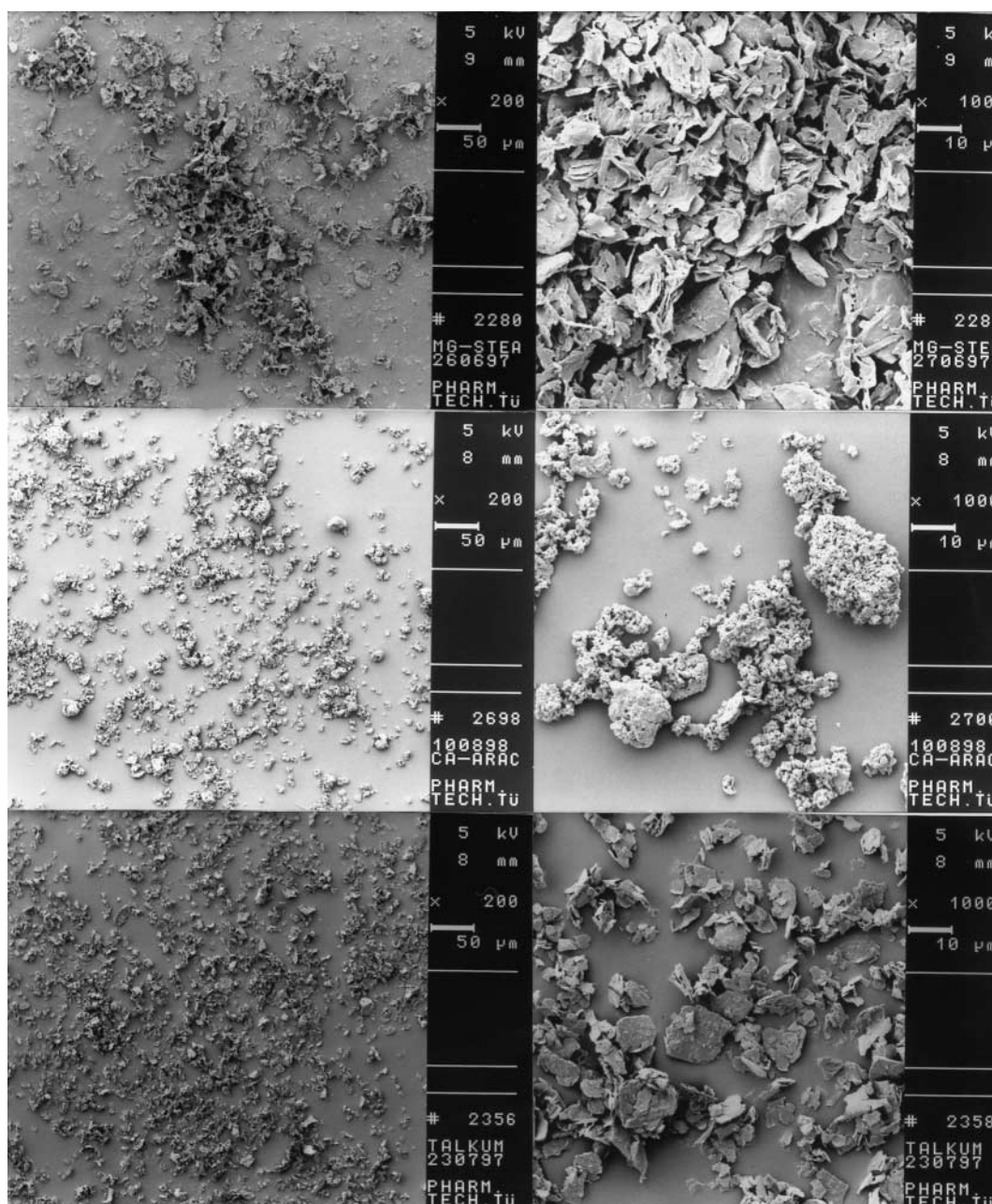


Fig. 17 Lubricants. Magnesium stearate (top), calcium arachinate (middle), and talc (bottom).

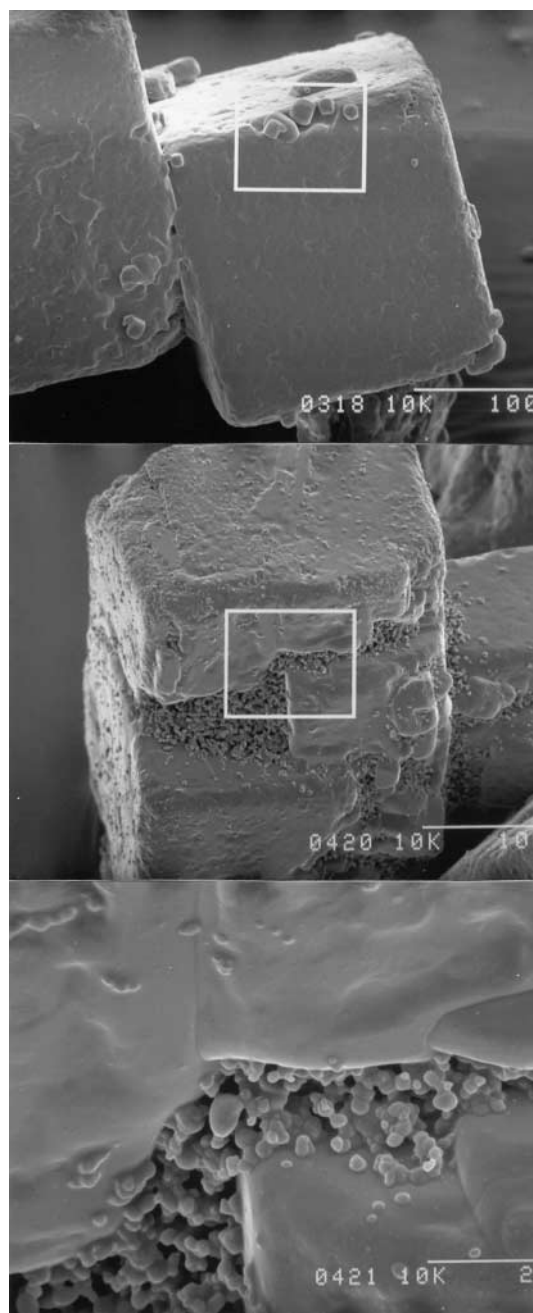


Fig. 18 Partially interactive powder mixture between sodium chloride (top) and micronized etilefrine HCl. The bottom picture is an extension of the middle. Etilefrine HCl settle at crystal irregularities of the cubic sodium chloride.

Figure 10 presents three different types of lactoses. The typically wedge-shaped α -lactose monohydrate (top) differs significantly from the roller dried β -lactose (middle), and the ball-shaped, spray-dried lactose (bottom), which shows the highest compactability due to

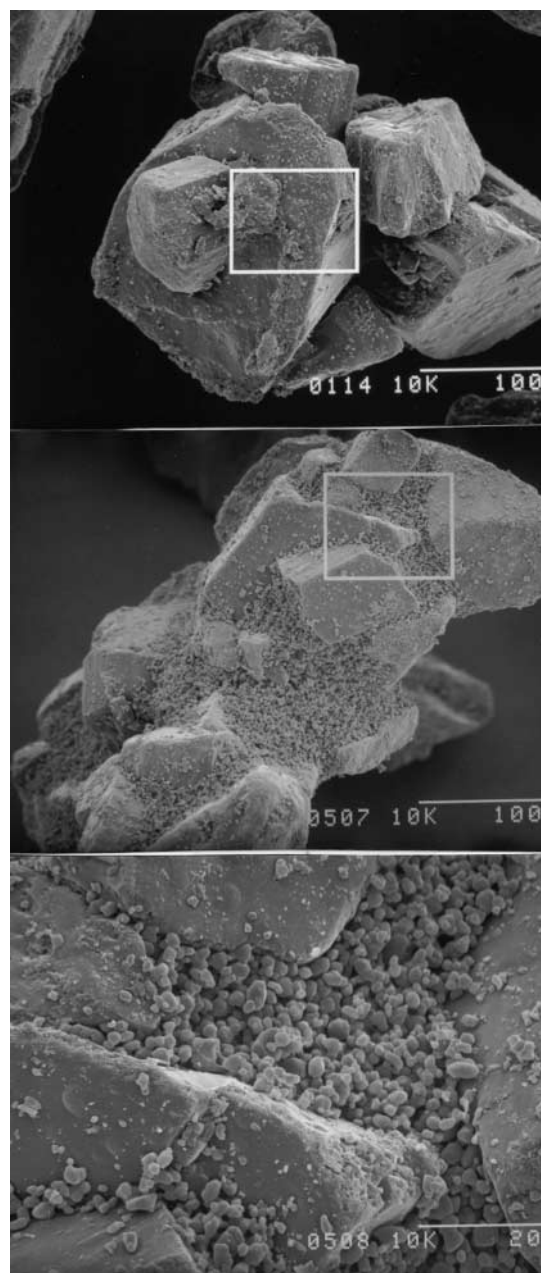


Fig. 19 Interactive mixture between lactose (top) and micronized etilefrine HCl. The bottom picture is an extension from the middle. Due to more crystal irregularities, there is a higher adherence of the etilefrine HCl onto the carrier.

its agglomerated form and its content of 8–15% of amorphous lactose (17).

The sugar alcohols, mannitol and sorbitol are quite often used in the preparation of buccal and sublingual tablets as well as lozenges. Fig. 11 shows the morpho-

logical differences between a crystallized (top) and a spray-dried mannitol (middle). Sorbitol, of which the γ -polymorph is used in direct compression, has to be regarded as the “chameleon” of tablet excipients. The spray-dried instant product (Fig. 11, bottom) is of irregular shape and shows fine needles on higher magnification. A product obtained from a melt (Fig. 12, top left) has a more regular surface as well as products that have been produced by crystallization (top right and bottom left). At higher magnifications the latter product exhibits a completely different structure compared to spray-dried sorbitol (18–20).

Figure 13 depicts the SEM pictures of three different co-processed materials (21). Ludipress, being mostly α -lactose monohydrate, shows fairly round agglomerates,

Cellactose, containing 25% of a powdered cellulose, is composed of lactose agglomerates from which the incorporated cellulose fibers protrude like tentacles. Avicel PH 200, a pure microcrystalline cellulose is an agglomerated irregular-shaped product (22).

The compressional properties of these materials with respect to hardness are similar, whereas the disintegration time shows for Cellactose a tremendous increase at compressional pressures above 100 MPa as shown in Fig. 14 (23).

Disintegrants

Starches are used as disintegrants in tablets, as binders after gelatinization in granulations, as carriers in spray drying, and as fillers and moisture absorbers in external

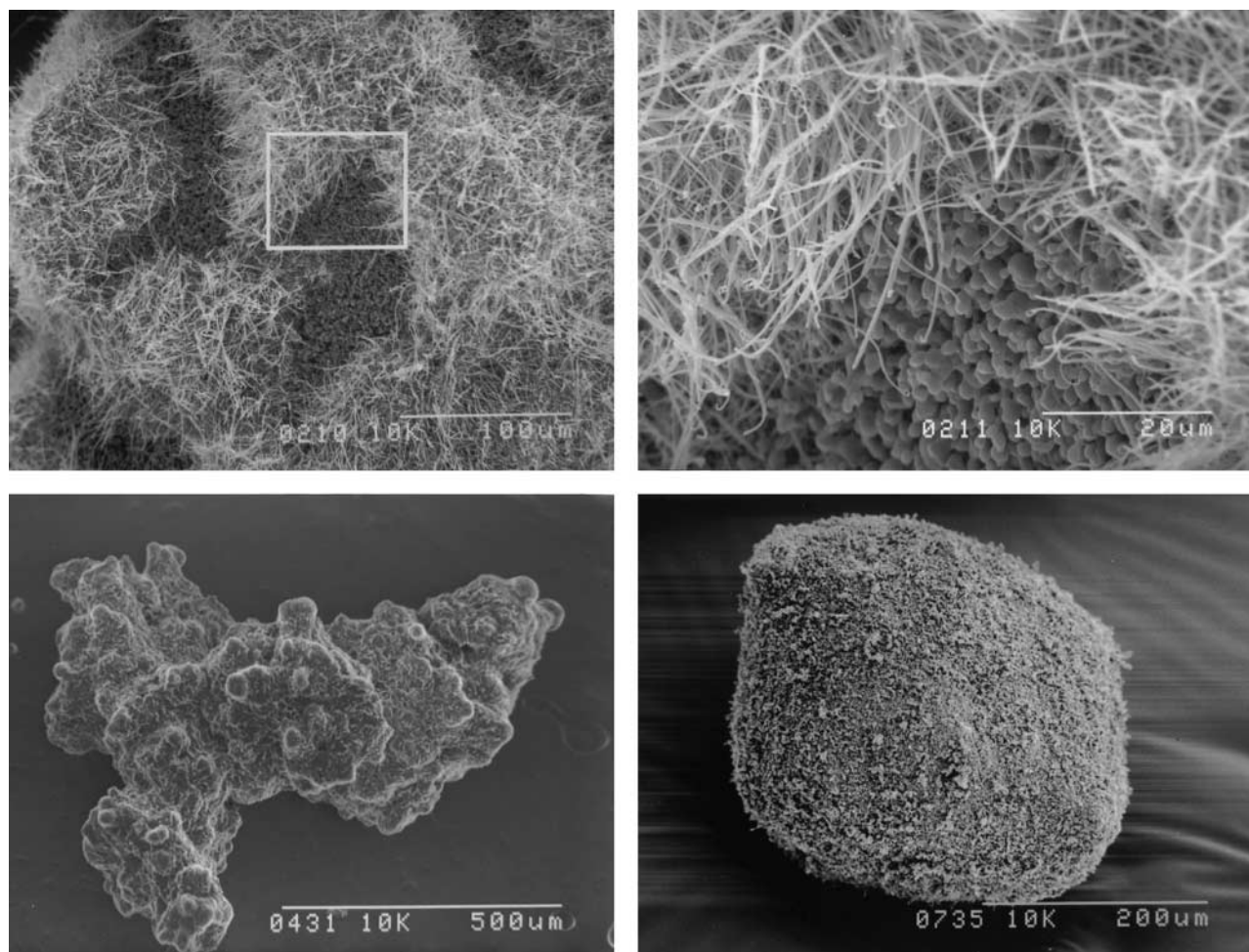


Fig. 20 Interactive mixtures with sorbitol as a carrier. Sorbitol plus 16% micronized etilefrine HCl, part of the granule (top left), sectional view from left showing the fine sorbitol needles and the micronized particles of etilefrine (top right). Pure sorbitol instant (bottom left), sorbitol plus 16% propranolol (bottom right).

powders. As disintegrants they are used in quite high concentrations ranging from 10 to 15% of a tablet formulation. Figure 15 presents rice (top) and wheat (middle) starches in two different magnifications. On the bottom left potato starch (one of the most commonly used starches) is shown together with starch from an Ethiopian plant called Enset false banana (*Ensete ventricosum*) showing particles of a very distinctive shape having a similar size as potato starch (24).

Besides starches, starch derivatives like sodium starch glycolate (Primojel, Explotab) are used as disintegrants as they exhibit enhanced disintegrating power. Although both products (Primojel and Explotab) are chemically identical, their microscopic behavior is different (Fig. 16, top). Primojel has a smooth surface and Explotab has a rough surface, resulting from their preparation methods that attribute to the differences in their disintegrating action (25).

Derivatives of celluloses like Ac-di-Sol and E.C.G. 505 act as the so-called “super disintegrants” (26, 27). Their microscopic pictures are similar to cellulose fibers. The

materials, however, could not be distinguished by microscopic examination.

Lubricants

Among lubricants, magnesium stearate is the most widely used one. It appears in different crystal forms, shows different particle size and shape, and occurs in several hydrate forms (28, 29). Its lubricating effectiveness was described by Delacourte et al. (30). The platelet form (Fig. 17) seems to have some advantages. Magnesium stearate is capable of forming films on other tablet excipients during prolonged mixing, leading to a prolonged drug liberation time (31), a decrease in hardness, and an increase in disintegration time (32, 33). For comparison reasons, calcium arachinate and talc are presented in Fig. 17. Calcium arachinate differs from the others by showing a microparticle structure instead of platelets.

The lubricating ability of talc is less compared with magnesium stearate, although both substances are microscopically similar (34).

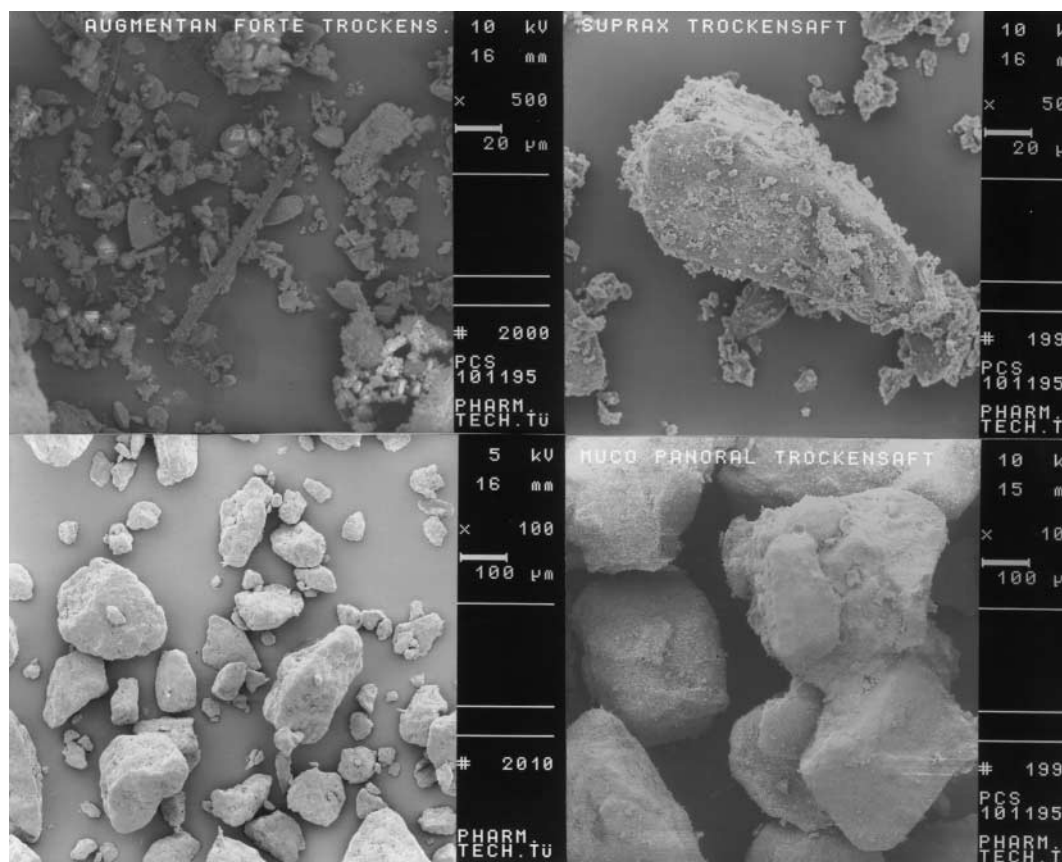


Fig. 21 Dry syrup preparations. Powder mixture (top left), partially granulated product (top right), granulated product (bottom left), and coated crystals, partially agglomerated (bottom right).

Powder Mixtures

Powder mixtures could be random (35–39) or “ordered” mixtures (40). The latter were later on called “interactive” mixtures (41) due to the fact that they are not of a higher degree of order. The mixing homogeneity of an interactive mixture is not better compared with a random one, but it leads to an interaction between a fine component and a coarse carrier and then to a higher stability of the system. Pharmaceutical powder mixtures are quite often between a random and an interactive mixture, and hence they are partially interactive. The degree of interaction depends on the following parameters:

1. particle size of the “adherent” fine powder;
2. particle size and surface structure of the coarse carrier;
3. cohesion forces between the “fines”;
4. adhesion forces between the “fines” and the coarse carrier particles; and
5. relation between the amount of “fines” and coarse particles.

The following examples deal with interactive powder mixtures (42, 43). Figure 18 shows on top a sodium chloride crystal. When mixing the sodium chloride with 16% of a micronized etilefrine HCl, a partially interactive powder mixture results (Fig. 18, middle and bottom). The micronized active ingredient is preferentially adhered onto crystal irregularities. In these regions multilayers of etilefrine HCl are obtained. When choosing lactose as a carrier instead of sodium chloride, the situation changes (Fig. 19). Due to the more irregular crystals, there is a higher degree of visible adhering particles (middle). The extension of this figure (bottom) shows again the concentration of fines at crystal irregularities. Special effects could be achieved when using sorbitol instant as a carrier. Sorbitol instant is produced by spray drying and subsequent instantization. Fig. 20 presents such a particle showing an irregular shape and a rough surface. Mixing this product with 16% of etilefrine HCL (top left) results in an adherence tendency to the “valleys” of the sorbitol particle. The extension (top right) shows the fibrous structure of the spray-dried sorbitol and the micronized particles. When using 16% of micronized propranolol HCl instead of etilefrine as the active ingredient being mixed, a total coverage of the sorbitol particle is obtained. Since propranolol has a higher adhesion and cohesion tendency, after filling up the sorbitol “valleys” it covers the whole surface.

Dry Syrup Preparations

Dry syrup preparations are powdered or granular formulations prepared for reconstitution. They are filled

up to volume by a liquid, normally water, just before use. During production and storage they are solids and during administration they are either solutions or suspensions. The main reason for the development of dry syrups is the instability of the active ingredient in the liquid state. They

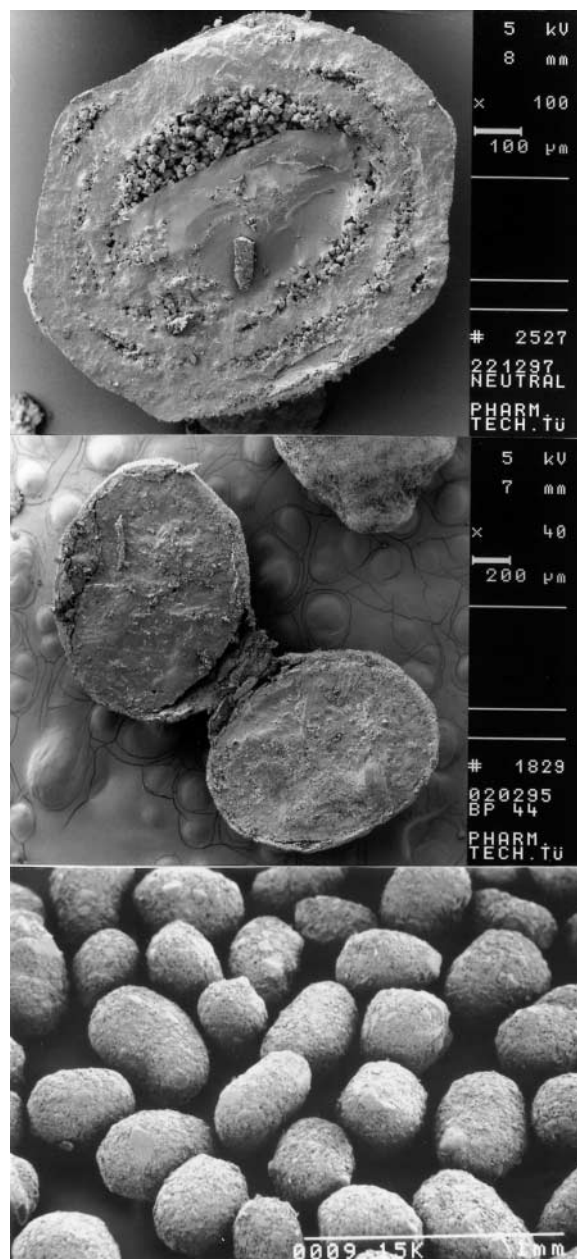


Fig. 22 Pellets prepared by different methods. Cross-sectional view of a pellet prepared by powder layering (top), cross-sectional view of a pellet prepared by extrusion/spheronization (middle), and size and shape of pellets prepared by extrusion/spheronization (bottom).

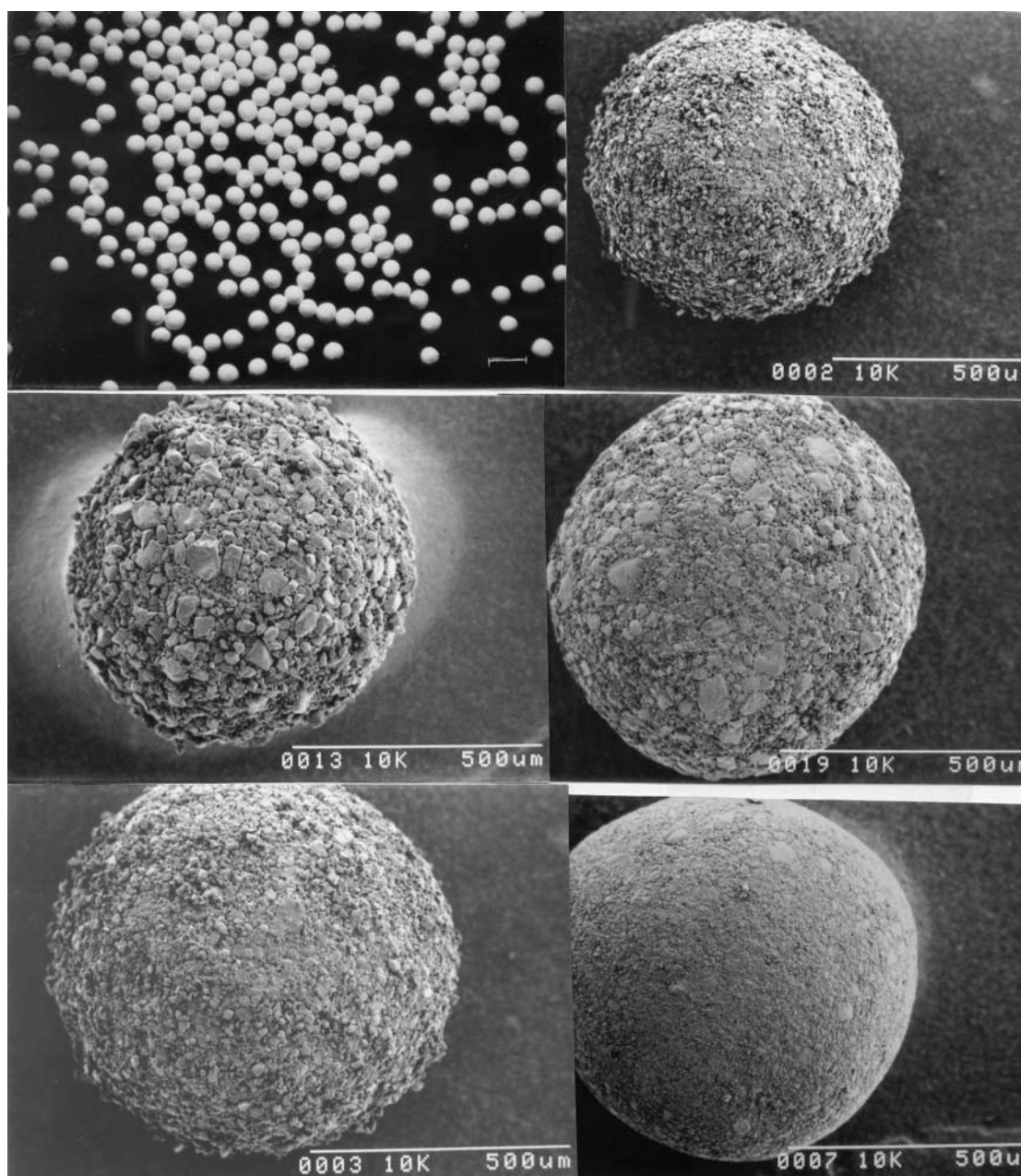


Fig. 23 Pellets prepared by spherical agglomeration. Overview about ascorbic acid pellets prepared from powder of a mean particle diameter of $8.8\ \mu\text{m}$. The bar is 1 mm (top left). SEM picture from left (top right). Ascorbic acid pellet prepared from powder of a mean diameter of $8.8\ \mu\text{m}$ without the addition of povidone (middle left). Pellet from left prepared with the addition of 2% PVP (middle right). Ascorbic acid pellet prepared from powder of a mean particle diameter of $3.0\ \mu\text{m}$ without the addition of povidone (bottom left). Pellet from left with the addition of 2.0% povidone (bottom right).

have to be used within a limited period of time after conversion into a suspension or solution. Most of the medicaments formulated as dry syrups belong to the group of antibiotics. It was shown that the bioavailability of a dry syrup could be superior compared with an oily

suspension (44). In addition to the active ingredient, the formulations (45) contain carriers like saccharose, sorbitol, and xylitol, acting also as sweeteners, microcrystalline cellulose and starch, viscosity enhancers, mainly xanthan gum, artificial sweeteners, preservatives,

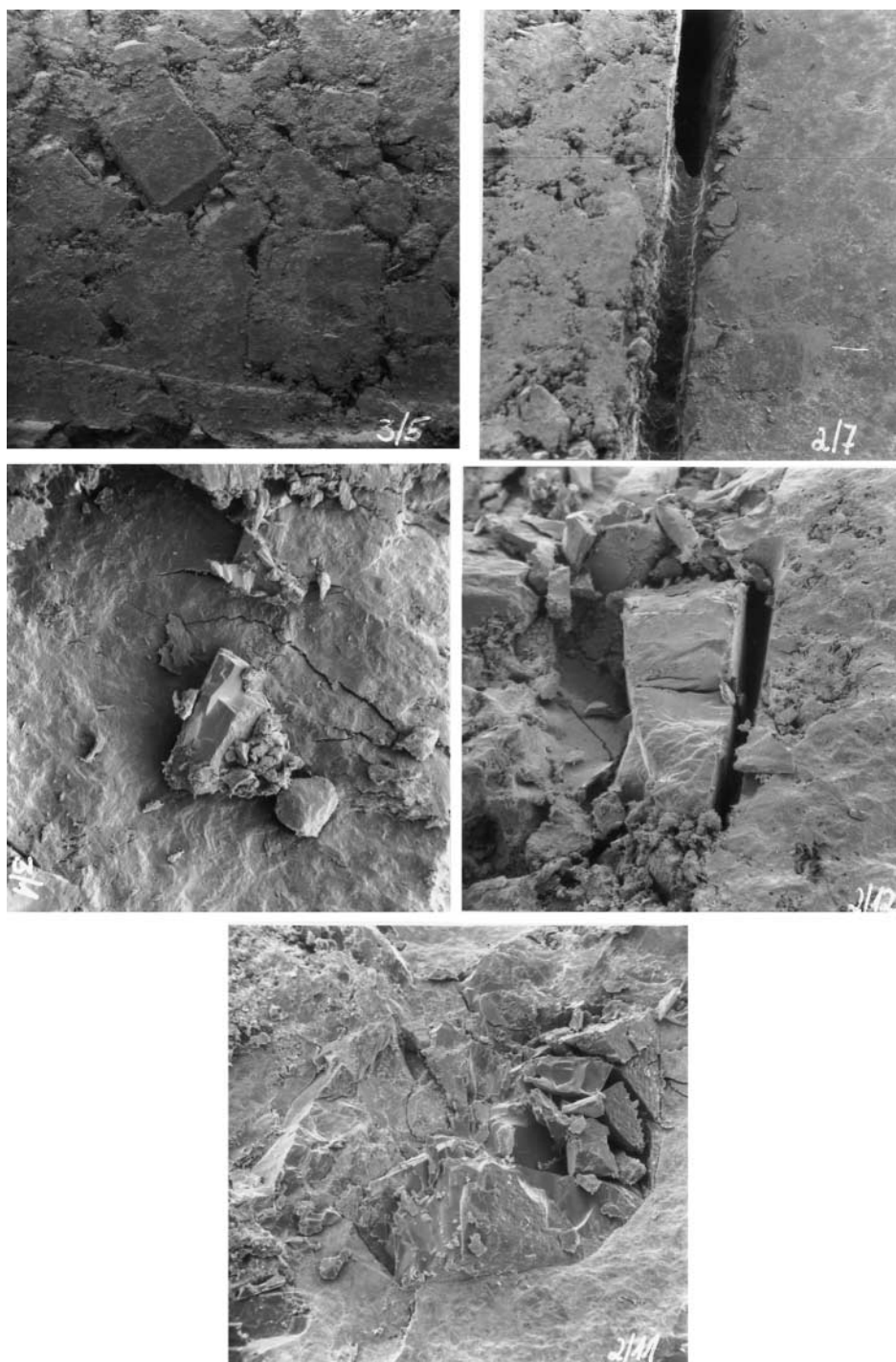


Fig. 24 Differences in the deformation behavior of substances during tableting demonstrated by sorbitol instant as an example for a plastically deforming substance and ascorbic acid as a more brittle one. Tablet compressed at 5 kN (10 mm in diameter) (top left) and 30 kN on top right. Inner part of a broken tablet showing an ascorbic acid crystal being totally embedded into plastically deformed sorbitol (middle left) and a crystal being partially removed from the sorbitol matrix through the break down of the tablet (middle right). A cracked ascorbic acid crystal in a sorbitol matrix (bottom).

wetting and complexing agents, and buffer substances. They could be prepared as powder mixtures, completely or partially granulated products, and as coated coarse carriers. Fig. 21 (top left) shows as an example for a powder mix, amoxicilline trihydrate, and clavunlinic acid containing preparation. A lot of small particles of different size and shape indicate a mixture. A partially

granulated product containing cefixim trihydrate is shown on top right. Besides small size particles granules are visible. A granulated dry syrup prepared from acetylcysteine is presented on bottom left, while coated carriers that are partially agglomerated containing bromhexine hydrochloride and cefaclor monohydrate are shown on bottom right. In three of the four formulations

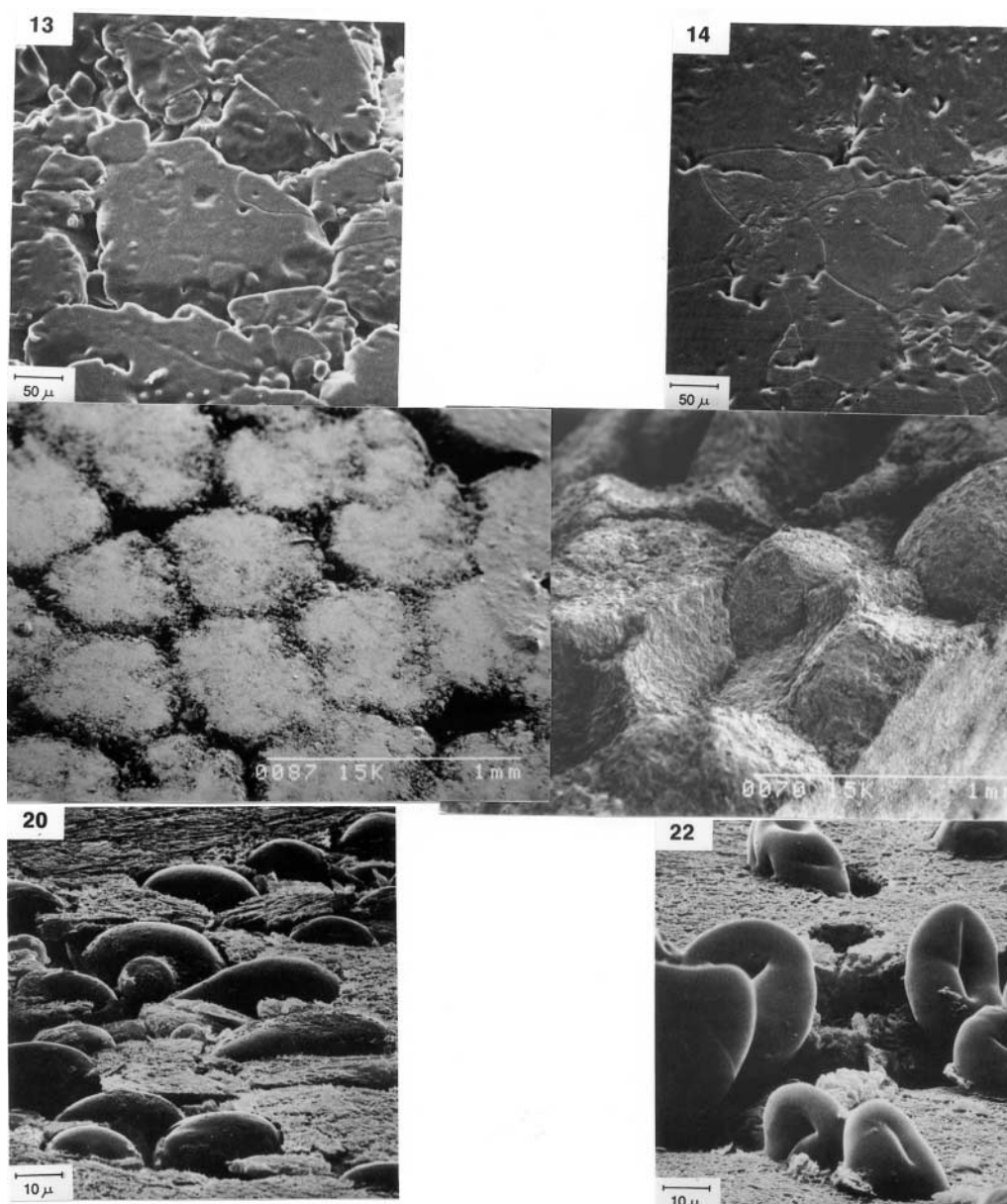


Fig. 25 Differences observed in the deformation of substances during tableting. Potassium chloride at low (top left) and high compressional forces (top right), pellets prepared by spherical agglomeration in a tablet surface (middle left) and in the inner part of a broken tablet (middle right), surface of an ASA tablet containing 10% sodium carboxymethyl starch (bottom left) and the same surface after moistening with a finger (bottom right) (54).

mentioned above xanthan gum is used as viscosity enhancing agent.

Pellets

Pellets are mainly prepared by four different methods: powder layering, rotating fluidized bed, extrusion/spheronization, and the agglomeration method. The oldest one is the so-called, "powder layering method" (46), where seed crystals in a drum are moistened by a binder solution and subsequently dusted by a fine powder. After a drying phase, the next moistening step is carried out followed by a powder application and so on. These types of pellets are characterized by

a core corresponding to the seed crystal, surrounded by fine particles. Figure 22 (top) represents such a pellet.

Pellets prepared by extrusion/spheronization (47) have uniform cross-sectional area (Fig. 22, middle) but differ slightly in shape and size, which is based on the spheronization process following extrusion (Fig. 22, bottom). They are harder compared with pellets produced by powder layering. Their appearance is similar to pellets obtained by fluidized-bed granulation using the so-called "rotating fluidized-bed" (48). Due to the fact that ideally round pellets are desired, several attempts were made to quantify their roundness (49, 50). The spherical agglomeration is a completely different method that is

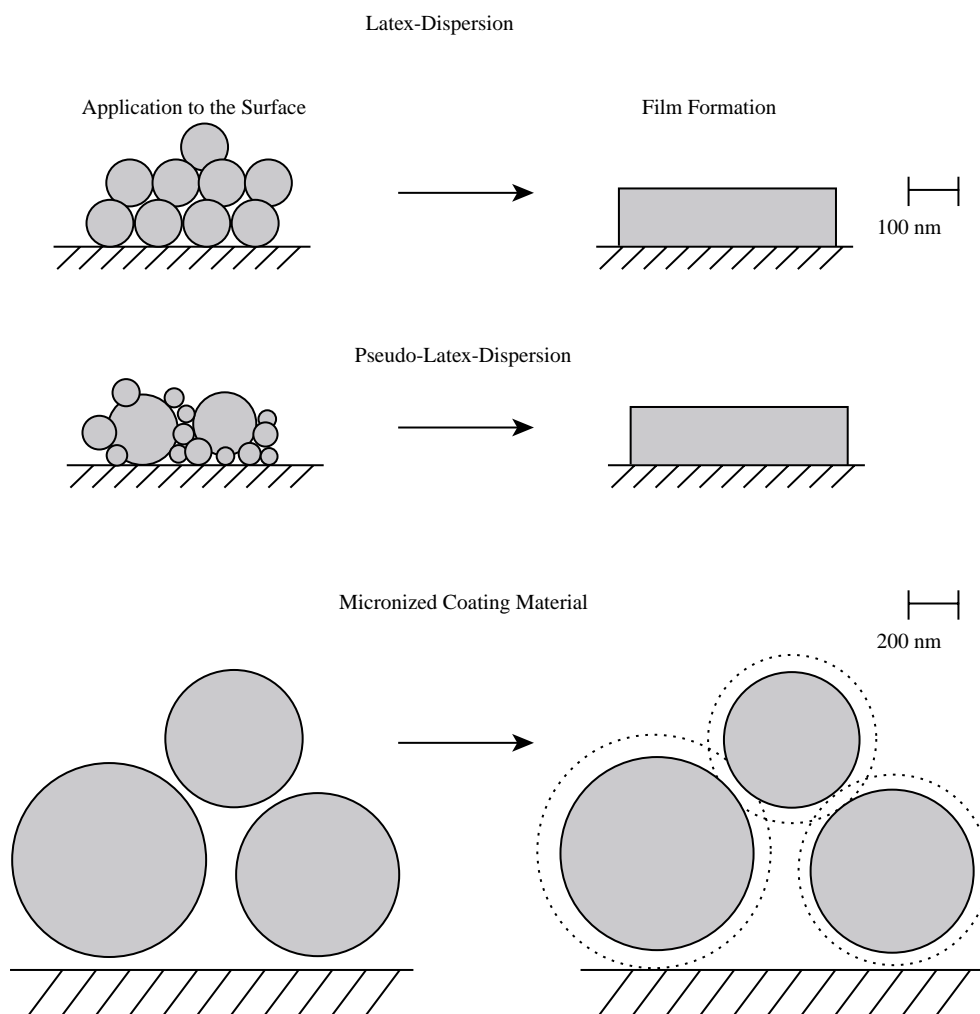


Fig. 26 Differences observed in the mechanism of film formation from latex and pseudolatex dispersions-and from micronized coating materials.

employed to develop pellets that are round and uniform in size. In this, particles are agglomerated in suspension (51, 52).

The resulting pellets are round and uniform in size (Fig. 23, top left); the shape of an ascorbic acid pellet prepared from a fine powder having a mean diameter of $3.0\text{ }\mu\text{m}$ is shown in Fig. 23 top right. Increasing the particle size of the starting material to a mean diameter of $8.8\text{ }\mu\text{m}$, the surface of the pellets becomes more rough (middle left). The addition of 2% of povidone during pellet preparation smoothes the surface (middle right). A finer starting material of $3\text{ }\mu\text{m}$ leads to a smoother surface (bottom left) that could become very smooth by the addition of 2% of povidone (bottom right). This example clearly demonstrates that by spherical agglomeration, it is possible to adapt the surface of the pellets being produced by a proper choice of the particle size of the starting material as well as by the addition of a polymer (53).

Tablets

SEM was widely used to investigate the structure of tablets. An excellent review is presented by Hess (54). Imaging of tablets is a useful tool to demonstrate differences in compression behavior of substances. Figure 24 shows an example where the plastically deforming spray-dried sorbitol instant was compressed together with the more brittle ascorbic acid in one tablet. At low compressional force of 5 kN for a 10 mm tablet, the rectangular ascorbic acid crystals as well as the partially deformed sorbitol particles are visible (top left). On top right the surfaces of two tablets compressed at 5 kN (left) and 30 kN (right) are compared. At higher compressional forces, a uniform, flat, and smooth tablet surface is formed, but within this surface a single unchanged ascorbic acid crystal could be detected. Observation of a broken tablet (Fig. 24, middle), which was prepared at a high compressional force of 30 kN, reveals that the ascorbic

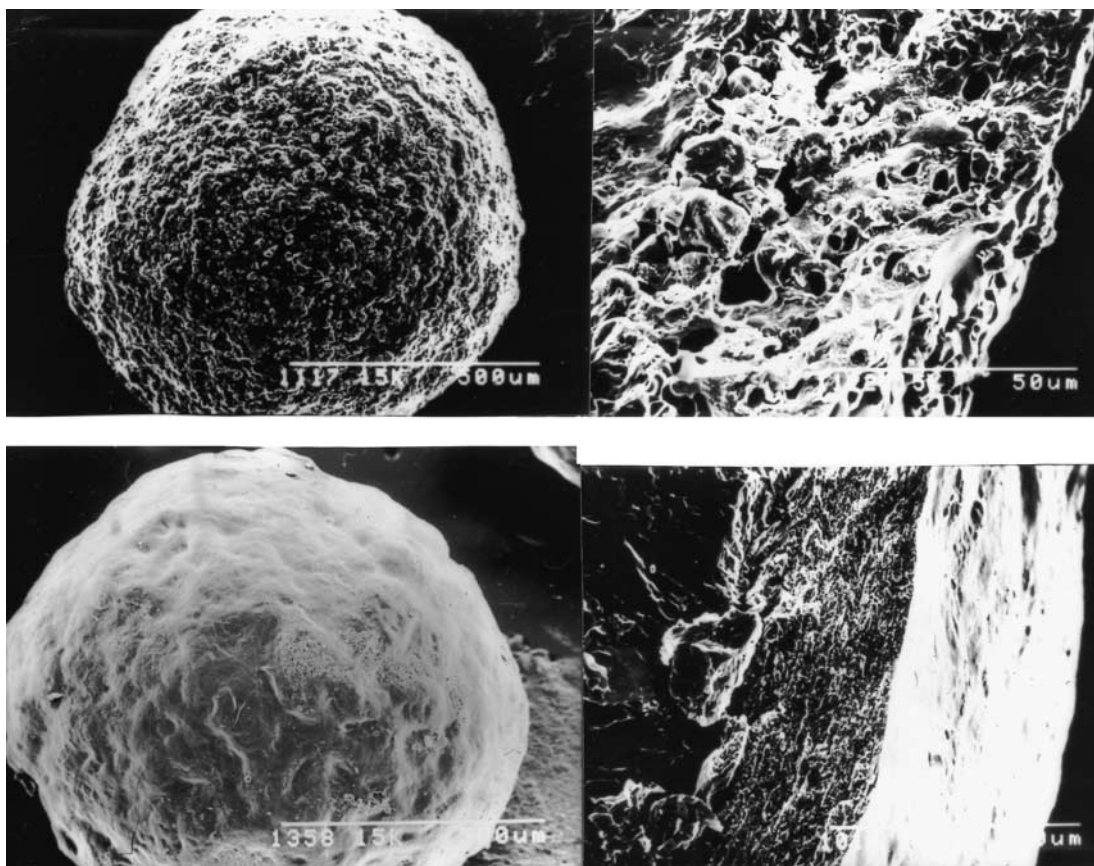


Fig. 27 Influence of the type of polymer application on film properties (top and side view). HP55 (HPMCP)-coated pellet, prepared from a micronized film dispersion (top). HP55 (HPMCP)-coated pellet, prepared from an organic solution (bottom).

acid crystal is totally fixed within a matrix of plastically deformed sorbitol. Even when the crystal was partially removed from the sorbitol matrix, during the break down of the tablet, the sorbitol matrix is still intact (middle right). This is supported by the bottom picture, where an ascorbic acid crystal was cracked during the breaking of the tablet, but it is still surrounded by plastically deformed sorbitol. These findings are similar to investigations of Hess (54), who used the plastically deforming potassium chloride (Fig. 25, top). On the top right, after compressing at low compressional forces, the single potassium chloride particles are separated by dips between them, while after compressing at high forces—although the original particles are still detectable—there are no visible dips.

When compressing pellets into tablets, their deformation depends on their hardness (55). When hard pellets produced by spherical agglomeration are compressed (Fig. 25, middle), the single pellets are still visible at the surface (left) and are only partially deformed in the inner part of the tablet [right, (53)]. This is of high importance, when pellets having an enteric or a sustained-release coat are compressed with the aim not to damage the coating.

The bottom pictures of Fig. 25 show differences in the elastic recovery of substances after compression. The left picture depicts the surface of an ASA-tablet containing 10% of sodium carboxymethyl starch as a disintegrant. The plastically deforming ASA remains plain after compression, whereas the starch derivative shows elastic recovery. After moistening the surface with finger the starch derivative particles start swelling moving out of the tablet surface (54).

Coatings

Coatings are applied to solid dosage forms for several reasons: taste masking, moisture prevention, gastric resistancy, and sustained-release action being the most common ones (56). They could be handled as a solution, a latex or pseudolatex dispersion, or as a suspension containing a micronized polymer powder as the film-forming agent. The film formation is easiest from a polymer solution because single polymer molecules can interact to build the coating. Using latex or pseudo latex dispersions or even micronized polymers the situation is different as demonstrated in Fig. 26. The size of true latex particles is uniform and it is in the range of 100–200 nm. A pseudolatex shows a broader particle size distribution although the mean particle diameter could be in the same range. In contrast, the particle size of a micronized particle of a film forming agent is 10 to 30 times higher compared to a latex, which is in the range of 1 to 3 or even more

micrometers. For a complete film formation, the plasticizer has to penetrate uniformly into the polymer. This is easily achieved when a latex dispersion is used. With a pseudolatex dispersion, however, the larger particles would not be plasticized completely. With a micronized polymer, on the other hand, only the outer part of the particles will be penetrated by the plasticizer. The resulting differences are presented in Fig. 27. On top, the rough surface and cross-sectional area of a pellet coated with the micronized polymer hydroxypropyl methylcellulose (HPMCP, HP55) are shown. The cross-sectional view

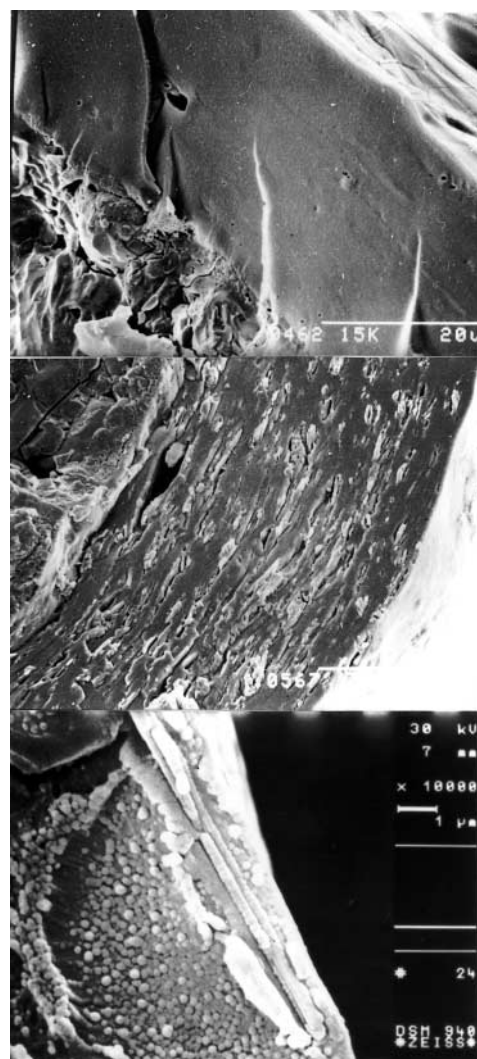


Fig. 28 Cross-sectional areas of different films. Eudragit L 30 D without pigments, film thickness approx. 25 μm (top), Eudragit L 30 D film containing 50% of magnesium stearate, film thickness approx. 25 μm (middle), incomplete film formation at higher magnification, showing individual latex particles from an Eudragit L 30 D dispersion in the film (bottom).

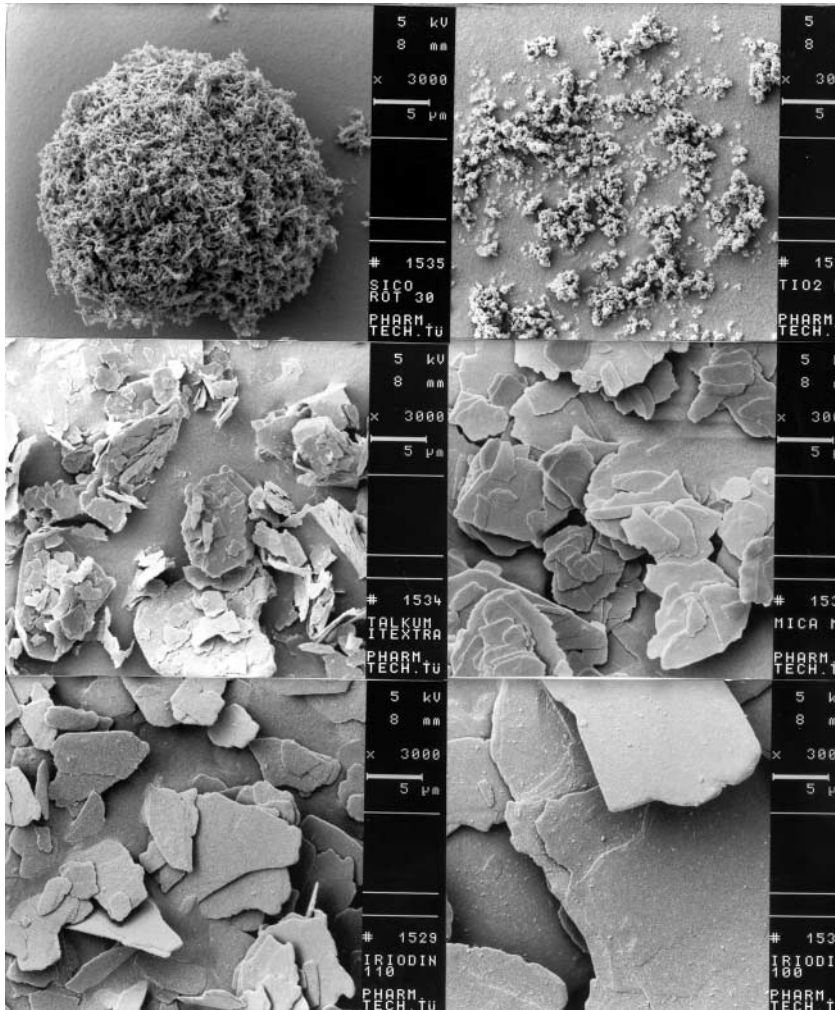


Fig. 29 Comparison of differences in size of pigments at the same magnification. Red iron oxide (top left), titanium dioxide (top right), talc (middle left), mica (middle right), and two pearl luster pigments Iridodin 110 (bottom left) and Iridodin 100 (bottom right).

hardly allows distinction between the coat and the inner part of the particle. The thickness of the film is approximately 25 µm. The pellet shown on the bottom was coated using an organic solution of HP55 showing a smooth surface and a more uniform cross-sectional view of the film. A more detailed structure of films is depicted in Fig. 28, where the influence of composition and conditions of preparation is apparent. The top picture represents a film that is produced without pigments. The addition of 50% of magnesium stearate results in a more structured layer (middle). On the bottom of Fig. 28 the cross-sectional view of an incomplete film at a higher magnification is presented. A pigment platelet is located near the surface. The latex particles around the platelet form a film, whereas those that one in the inner part of the coat still exist as latex

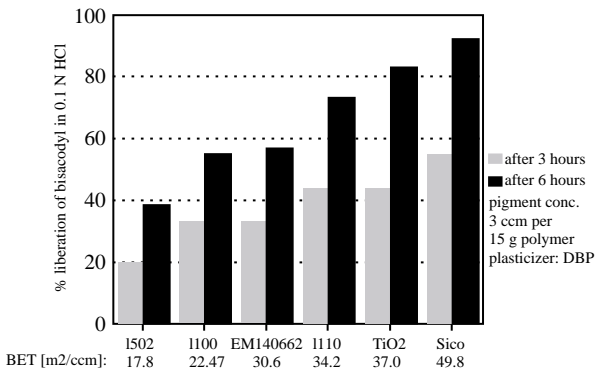


Fig. 30 Liberation of theophylline from aquacoat ECD 30 coated pellets with different pigments in relation to the BET surface of the pigments.

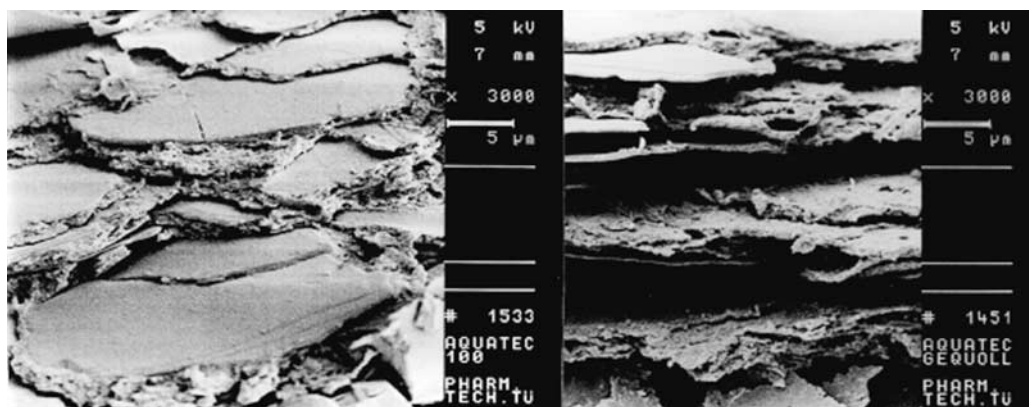


Fig. 31 Cross-section of films from Aquacoat ECD 30 pigmented with 20% (V/V) of the pearl luster pigment Iridin 100.

particles in the range of 100–200 nm. Obviously the minimum film-forming temperature was not reached in this region. The pictures clearly demonstrate the differences between the various application forms of film-forming agents.

Pigments could influence the mechanical properties of a film as well as the dissolution behavior of a drug substance

from coated tablets or pellets. According to Rowe (57) different pigments should be compared on the bases of the pigment volume concentration, which takes density differences into account (Eq. 3),

$$PVC = \frac{V_P}{V_P + V_B} \quad (3)$$

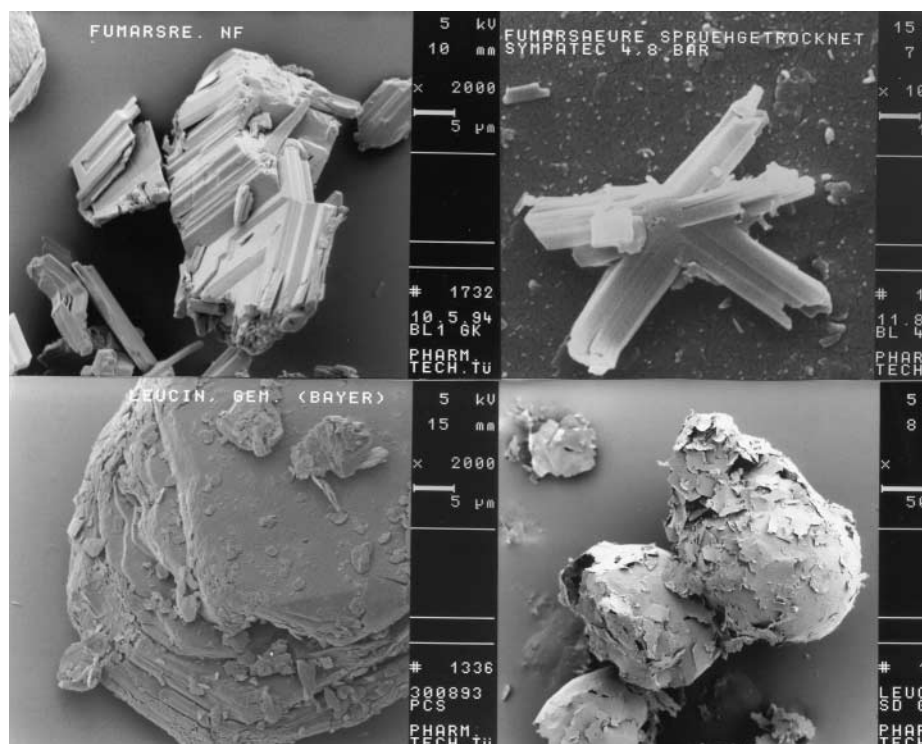


Fig. 32 SEM picture showing starting materials and their corresponding spray-dried products. Fumaric acid milled (top left) and spray dried (top right), L-leucine milled (bottom left) and spray dried (bottom right). Before spray drying both substances show a layered crystal structure but differ significantly after the spray-drying process.

where PVC = pigment volume concentration; V_P = pigment volume; and V_B = polymer volume.

Pigments exhibit different shape and size. Figure 29 clearly demonstrates that for red iron oxide (top left) and titanium dioxide (top right) on one hand, and talc, mica (middle), and two pearl luster pigments (bottom) on the other. The primary particle size of the two metal oxide crystallites is below $1\ \mu\text{m}$. The red iron oxide is more agglomerated compared with titanium dioxide. Talc (middle left) and mica (middle right) show platelet structure with a broader particle size distribution of talc. The two pearl luster pigments, both based on mica, coated with metal oxides also show the typical platelet structure and higher particle sizes. For titanium dioxide (58) and talc (59) the influence of the amount of pigment added to the film-coating formulation on the drug dissolution was shown. Up to the critical pigment volume concentration, the permeability of the films is decreased and therefore the sustained release effect is prolonged. Above the critical concentration the permeability of the films increases due to defects in the uniform film layer caused by the pigments (60).

With pearl luster pigments, depending on their particle size and/or BET-surface, respectively, this effect could be used to control drug liberation of sustained-release preparations. Figure 30 compares the in vitro dissolution of theophylline in water after 3 and 6 h from pellets that were coated with an Aquacoat ECD 30 dispersion pigmented with 20% of different pigments based on the amount of polymer used. Sicopharm (red iron oxide) and titanium dioxide, both fine crystallites, show the highest release rates. The others having a more or less pronounced platelet structure show a decrease in drug liberation that is directly correlated to the BET-surface of the pigments (60). The reason for the reduced film permeability lies in the special structure of films containing platelet-shaped pigments. Figure 31 (left) shows a section of such a film under an angle of approximately 45° . The roof-like structure of the coating is obvious. Therefore, substances penetrating the film have to permeate around the solid platelets, which prolongs the diffusion. Path consequently, the drug liberation is delayed. During the dissolution process, the film swells due to the swelling properties of the polymer. At the end of the process a picture shown in Fig. 31 (right) is obtained. The cross-sectional area of the swollen film in this case was observed under an angle of 90° . The pigment platelets are still in their original overlapping position. The degree of swelling depends on the type of polymer used. Aquacoat ECD 30 shows only a slight effect compared with Eudragit RL 30 D and RS 30 D.

Spray-Dried Products

Spray drying is used to convert solutions, emulsions, or suspensions into powders (61). The applications in pharmaceutical technology are numerous. Raw materials are spray dried, for instance, to enhance the compressional properties of substances such as lactose (62) and tricalcium phosphate (13), to distribute a minor component like digoxin more uniformly in a matrix, and to enhance

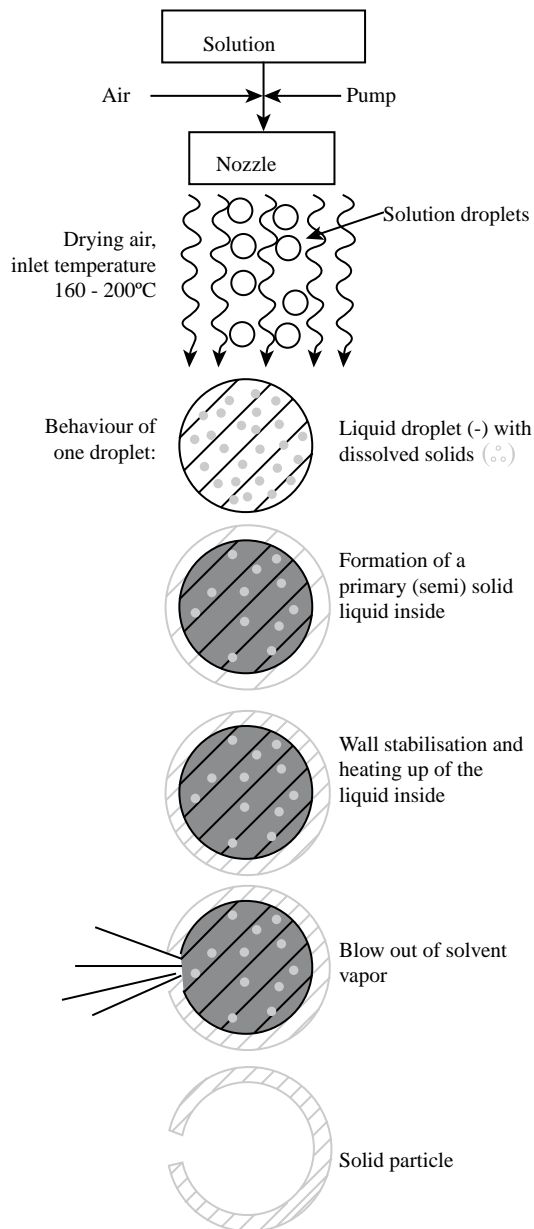


Fig. 33 Spray-drying process of extracts with respect to the behavior of a single droplet.

their dissolution rate as described for digoxin in combination with a hydrophilic polymer (63).

Normally one would expect round- to nearly ball-shaped particles from a spray-drying process when starting from a solution or an emulsion. When using a suspension, the resulting particles could also be more or less irregular agglomerates as shown earlier for spray-dried lactose and some of the so-called “multi-purpose excipients.” But even when starting from a solution round particles may not be obtained. This is demonstrated in Fig. 32. Fumaric acid and L-leucine in their original state are both crystals compasing of thin lamellae. After spray-drying fumaric acid forms small star-like agglomerates (Fig. 32, top right), whereas L-leucine forms hollow spheres built from small and very thin platelets (Fig. 32, bottom right). These spheres have advantages as lubricants in effervescent tablets compared with a milled product because the spheres are broken during compression, forming a thin

film of small L-leucine platelets on the punches and die of a tablet press. Thus, an amount of 3.75% of spray dried L-leucine in an effervescent tablet formulation was more effective compared with 5% of the milled product (64). Spray drying of extracts is quite often used to prepare readily soluble tea preparations (65). The advantages are a quicker solubility, a uniform distribution of active ingredients, and a free-flowing, dust-free powder. The drying process of a single droplet can be schematically described as shown in Fig. 33.

In the first step, liquid droplets are formed by the nozzle of the spray dryer with the aid of a pump and compressed air or by an airless system. Due to the high temperature of the drying air, the solvent evaporates first from the outer surface of the droplet and a solid “skin” is formed around the liquid that acts as a barrier for further evaporation. As a result, the pressure within the droplet increases until the wall breaks and the solvent is blown out of the droplet. At

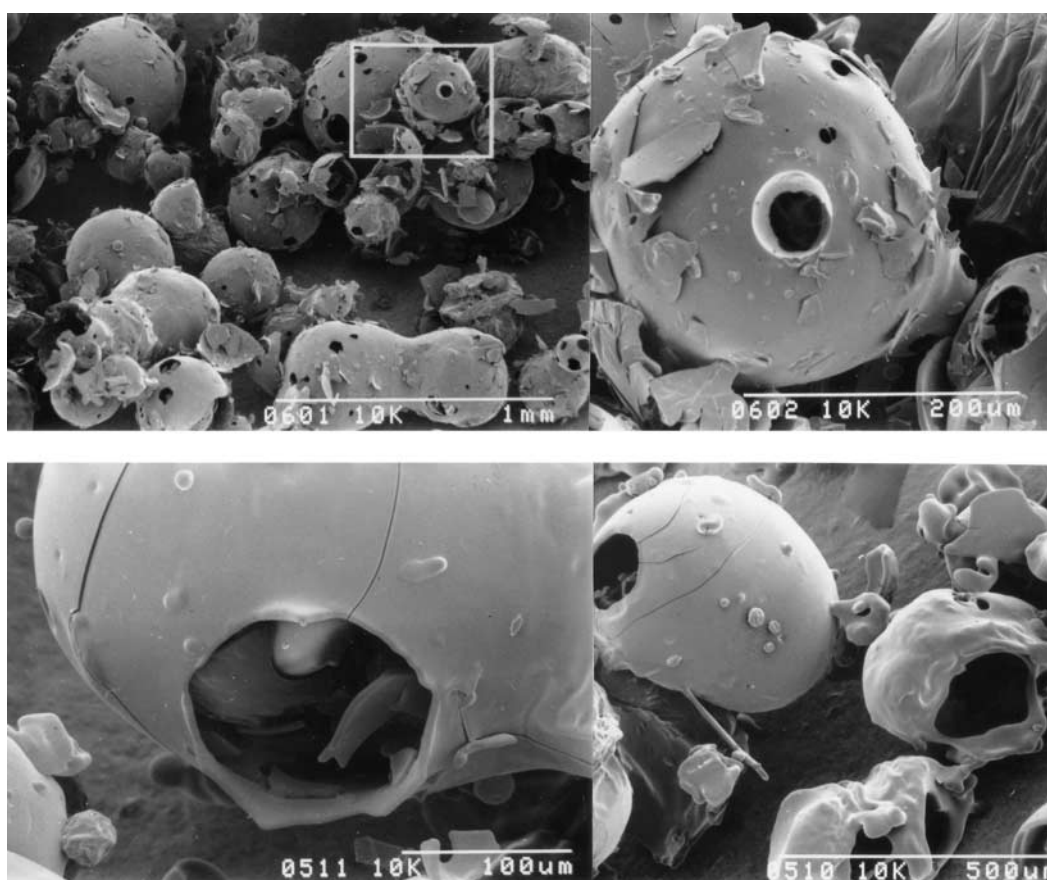


Fig. 34 Spray-dried plant extract solutions used to prepare readily soluble tea preparations. Spray-dried product from an airless high-pressure system leading to a coarse particle size distribution (top left) and a magnification thereof (top right), detailed view of one particle showing fragments inside (bottom left), and the product from above after a wrong sample preparation (bottom right).

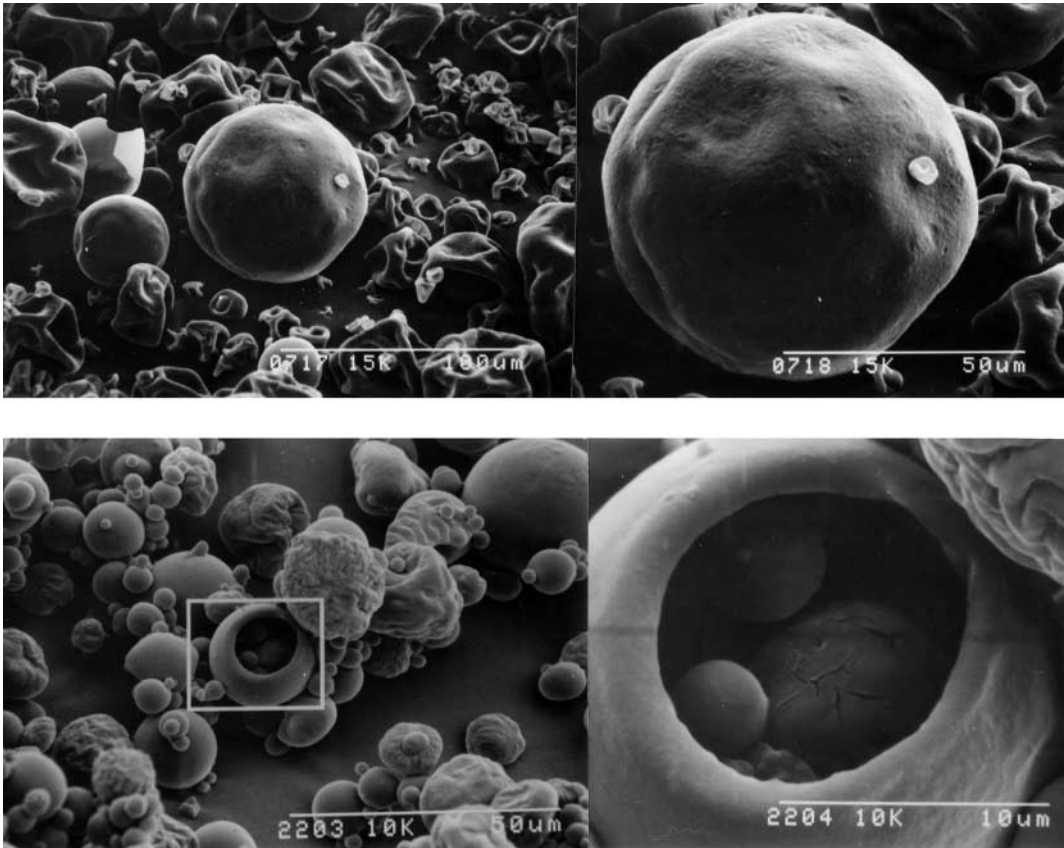


Fig. 35 Spray-dried extracts and pantothenate solution. Spray dried calcium pantothenate: overview (top left) and detail (top right), showing mainly particles that were damaged during preparation of the SEM sample either by the high vacuum of the SEM or by the heat created during the sputter coating process. Spray dried *Tussilago farfara* extract prepared by a system using atomizing air to form droplets and leading to small particle sizes (bottom left) and a magnification thereof (bottom right).

the end, a hollow solid particle results showing small holes, where the liquid was blown out. Depending on the excipients in the formulation particles could more or less

break down during the process. The fragments then adhere to other fragments and to the spheres. This behavior is demonstrated by Fig. 34 (top left). The extract particles on

Table 3 Manufacturing data of porous ceramic pellets of β -Tricalcium Phosphate and hydroxyapatite, respectively

Type (see Fig. 44)	Compaction pressure (MPa)	Firing temperature (°C)	Size (d/h; mm) ^a	Density ^b (%)
a ^c	48	1075	4.4/4.0	64.9
b ^c	48	1125	4.1/3.7	78.4
c ^c	168	1075	4.4/4.0	79.8
d ^d	111	1150	4.4/4.0	63

^ad = Diameter; h = Height.
^bPercent density based on the true density of the starting materials.
^c β -Tricalcium phosphate.
^dHydroxyapatite.

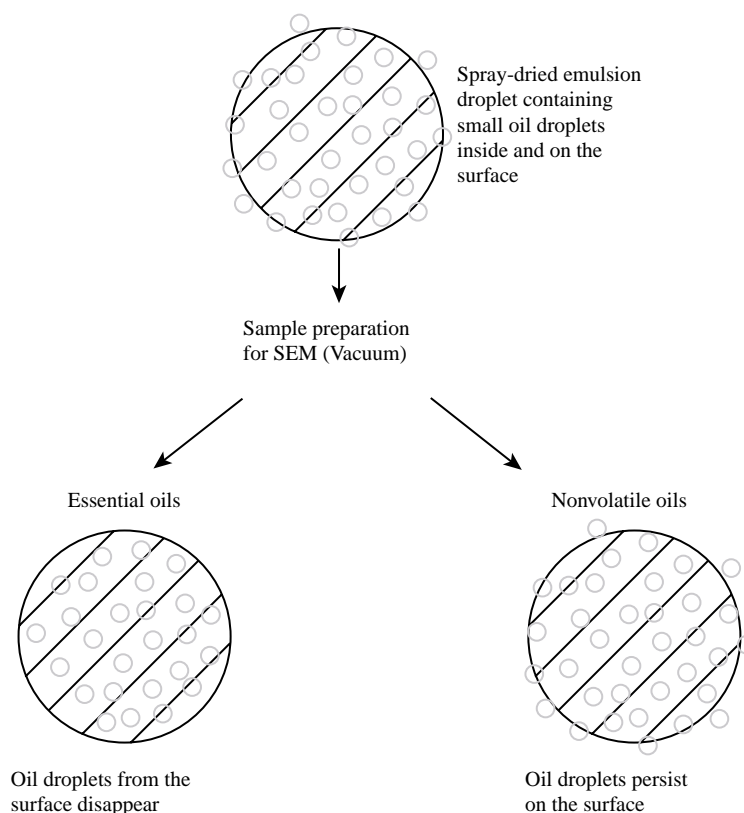


Fig. 36 Differences between spray-dried emulsions containing essential oils and nonvolatile oils during sample preparation. The oil droplets of the volatile essential oils disappear during sample preparation under vacuum, while nonvolatile oil droplets stay permanently.

that figure originating from an airless spraying system result in particle size ranges from approximately 100 to 500 μm . Fragments adhering to other particles are present as well as agglomerates of two or three spheres. The higher magnification of one of the particles (Fig. 34, top right) shows the "blow out hole" and also adhering fragments. A higher magnification of one of the "blow out holes" (bottom left of Fig. 34) demonstrates the presence of fragments and smaller spheres within a bigger particle. In the picture, at bottom right, some of the particles show irregular surfaces. These are artifacts, which are formed during sample preparation. They are built as a result of overheating during the sputter process with gold. A similar case is demonstrated in Fig. 35 (top). Most of the spray-dried calcium pantothenate particles are collapsed due to instability of the previously formed thin walled spheres under the high vacuum of the SEM. From time to time one can observe the collapsing of the particles under the SEM.

The bottom pictures of Fig. 35 demonstrate the difference to particles mentioned before when using an air-driven nozzle instead of an airless one. The particle size range of the *Tussilago farfara* extract shown in the figure is now in the range from 1 to 20 μm . Some of them are damaged by heat during sputtering (bottom left).

At higher magnification (bottom right) one single particle of roughly 20 μm in diameter is shown containing smaller ones inside. The wall thickness is in the range of 1 μm or below. In some cases emulsions containing oil in the inner phase are submitted to spray drying. The spray-dried particles contain the small oil droplets inside as well as on the surface as schematically drawn in Fig. 36. During the sputtering process of the samples under vacuum essential oil droplets are evaporated from the surface of the spray-dried particles, whereas nonvolatile oil droplets stay on the surface and are coated by gold. The different situations are demonstrated in Fig. 37 by SEM pictures. On

top, an azulene containing spray-dried particle to which smaller ones are adhered is shown. Here, no oil droplets on the surface are visible due to the fact that azulene belongs to the group of essential oils. In the middle, the surface of a 50% vitamin E acetate containing spray-dried particle is depicted. The surface is rough and higher magnification of the same (bottom of Fig. 37) clearly demonstrates the presence of small oil droplets on the surface of the spray-dried particle.

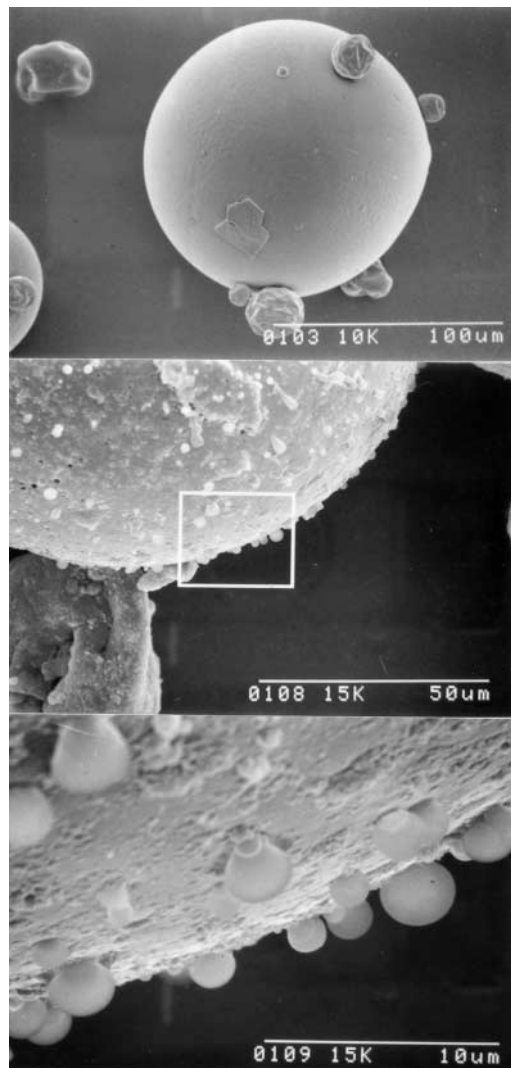


Fig. 37 Spray-dried emulsions. Particle from an emulsion containing a volatile essential oil (azulene) (top). Particle from an emulsion containing a nonvolatile oil (vitamin E acetate) (middle and bottom).

Freeze-Dried Products

Freeze drying was first carried out in 1890 by Altman (66) but became well known through the industrial development of the process. The process is used in the food industry e.g., for the production of instant coffee, tea, and other products. In the field of pharmaceutical technology, it was for a long time restricted to only few formulations for injection containing the active ingredient in the freeze-dried state in an ampoule to be dissolved just before application. With the increasing interest in protein and peptide formulations, freeze-dried products became more important. These are discussed under “microparticles for injection” in the following section.

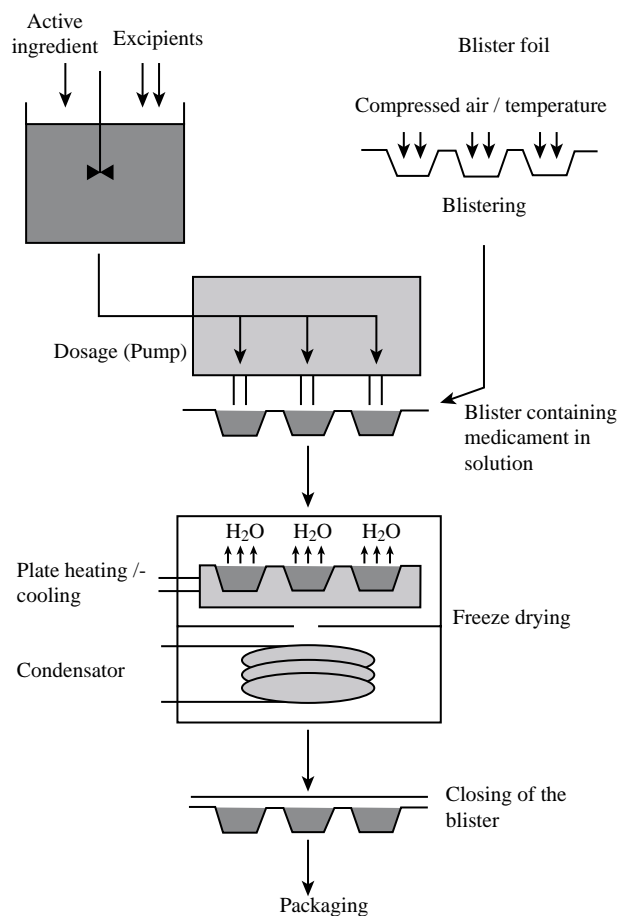


Fig. 38 Preparation scheme of medicaments containing-freeze dried platelets.

Instantaneously soluble freeze-dried platelets (67) became known during the last few years. They were mainly developed for people having problems of swallowing a tablet. Their production scheme is presented in Fig. 38. A solution containing the active ingredient together with excipients like mannitol and if necessary polymers, sweeteners, colors, and flavors is dosed by a pump into preformed blisters of PVC or a

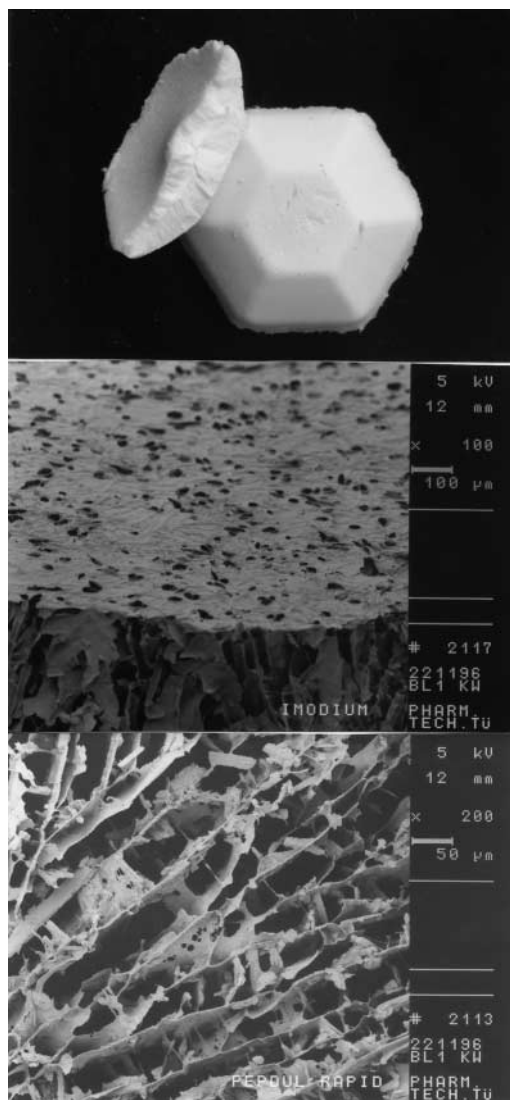


Fig. 39 Instantaneously soluble freeze-dried medicament platelets/tablets. Cross-sectional side view and top view (top), border line between the surface (upper part), and the cross-section (lower part of the picture) (middle), inner part of the platelet (bottom).

combination of PVC and PVDC. The blisters containing the solution are freeze dried and an aluminum foil covering the blister is applied at the end of the process. The platelets have the size of tablets as shown on top of Fig. 39 and they adapt to the form of the blister through the freeze-drying process. The surface shows small holes from channels where water was evaporating during drying. These channels form a percolating system within the tablet (Fig. 39, middle). Their microstructure is shown on bottom of Fig. 39 indicating the loose structure of the platelets having a high internal surface. Through the channels, water is penetrating quickly into the platelet leading to a very fast dissolution.

Microparticles for Injection

During the last decade, peptides and proteins became more and more important as medicaments. Most of them are unstable in an acidic environment and in the

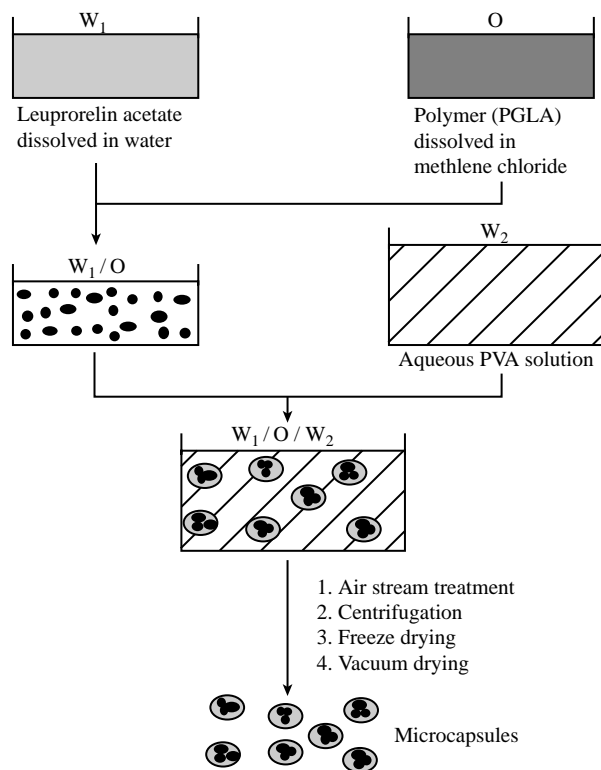


Fig. 40 Preparation of leuporelin acetate containing microcapsules according to the “in-water-drying” process.

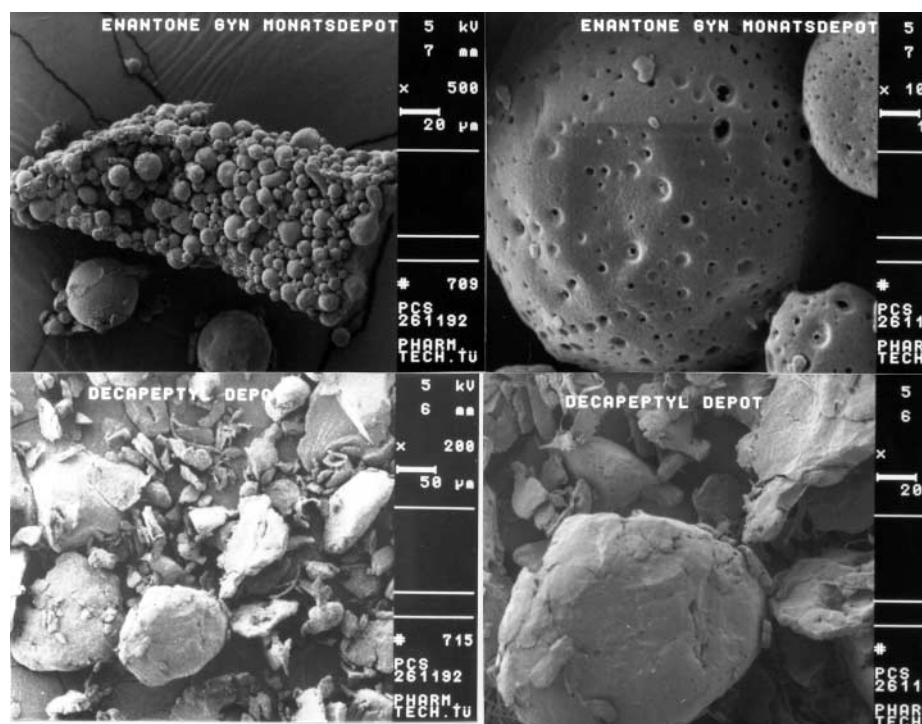


Fig. 41 Microparticles for injection prepared by the “in-water-drying” process (top) and by freeze drying (bottom) for two LH-RH super agonists.

presence of enzymes. Therefore, they could not be administered or applied via the peroral route (68). Since the development and registration of poly-lactide-glycolides as biodegradable polymers for injectable

sustained release preparations (69), methods of preparation of injectable microparticles became very interesting. As preparation methods, spray and freeze drying and the so-called “in-water-drying” process are in use (70). The latter is demonstrated in Fig. 40. Leuporelin acetate dissolved in water and a polylactide-glycolide solution in methylene chloride are mixed forming a W/O-Emulsion, where the methylene chloride solution is the outer phase of the emulsion. By the addition of an aqueous polyvinyl alcohol (PVA) solution, a water/oil/water multiple emulsion is formed with PVA solution in the outer phase. The main part of the methylene chloride is evaporated by an air treatment leading to a precipitation of polylactide-glycolide. The resulting particles are centrifuged, freeze dried, and dried under vacuum to remove traces of methylene chloride (71). Fig. 41 shows (top) the resulting round particles as a bulk before redispersion for injection (left) and single particles at higher magnification (right). The single particle exhibit pores with small holes where the methylene chloride has evaporated. For comparison reasons, a freeze-dried product is shown on the bottom of Fig. 41.

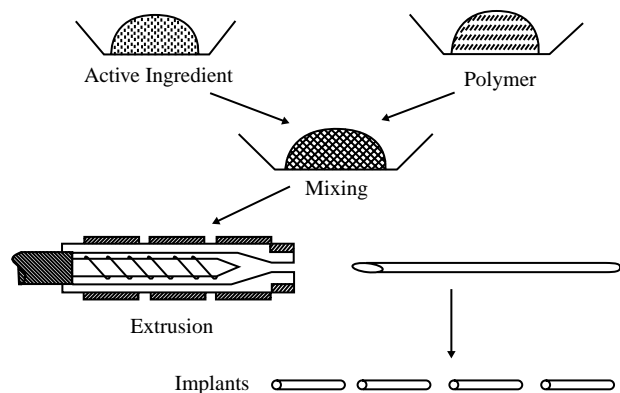


Fig. 42 Preparation of sterile injectable implants by extrusion of a mixture of *D,L*-poly(lactide-glycolide) and an active ingredient.

Implants

Implants can be biodegradable or nonbiodegradable ones. In earlier times tablets were implanted by a small surgery and removed after a certain time period. Modern implants consist of biodegradable organic or inorganic materials. The most commonly used organic material since the introduction of Zoladex® is *d,l*-poly(lactide-glycolide) (71). The implants are cylindrical in shape and have a diameter enabling the application by syringes and injection needles. A scheme of the preparation of this type of implants is presented in Fig. 42.

The active ingredient and the polymer are mixed, extruded to a cylindrical rod, and cut into pieces of approximately 1 cm in length. These pieces are introduced into syringes and applied to the patient as an implant through a needle. Fig. 43 presents one end of the cylindrical implant at two magnifications. The higher magnification shows small irregularities at the end of the rod that could be responsible for the pain sometimes reported during the application of the systems.

A second type of implant that is used for the local antibiotic treatment of bone infections is based on β -tricalcium phosphate ceramic pellets (72). The pellets are prepared by granulation and subsequent thermal treatment between 975 and 1300°C. They are loaded with antibiotic drugs like gentamycin. The drug liberation is affected by the pore volume of the pellets leading to a sustained release of the antibiotic. Fig. 44 presents a SEM picture of some materials, showing the influence of the thermal treatment and compaction pressure of the green pellets on the structure of the material. The manufacturing data are given in Table 3.

An increase in firing temperature from 1075 to 1125°C leads to a smaller pellet and a higher density corresponding to a lower porosity. Similar compacts can be obtained at the lower firing temperature by the application of a higher compaction pressure. Hydroxyapatite shows differences compared with β -tricalcium phosphate leading to low densities at high thermal treatment. A prolonged drug liberation up to 14 days was achieved by this method.

SUMMARY

The application of secondary electron microscopy (SEM) has become widespread in pharmaceutical technology. Besides other techniques like particle size analysis, surface measurements, X-ray diffraction, and differential scanning calorimetry, it is a useful tool to

characterize raw materials. Intermediates like powder mixtures, granules, and other bulk materials could be visualized by SEM. For interactive powder mixtures it is the only method making the interaction visible. Pellets could be investigated with respect to size and shape as well as for surface characteristics and properties of coatings being applied to them. Tablets show their internal structure in SEM. Spray-dried and freeze-dried products are further interesting examples for the application of SEM. In general one could state that more than 50% of all problems could be solved by the view through a microscope.

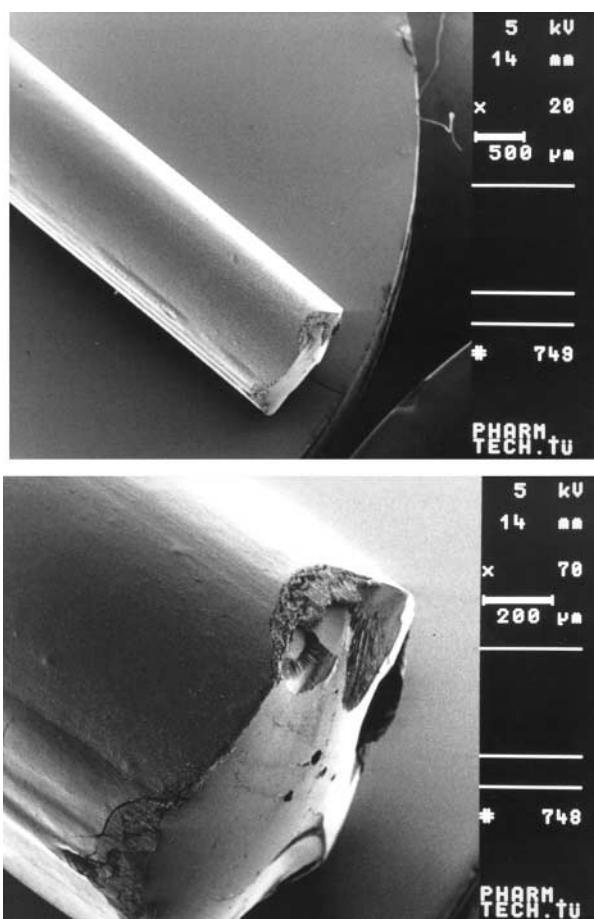


Fig. 43 SEM picture of a Zoladex® cylindrical implant. The stripes at the lower surface result from small irregularities of the extrusion nozzle (top). One end of a Zoladex® implant at a higher magnification. The cross-sectional area shows irregularities resulting from the cut off of single cylinders (bottom).

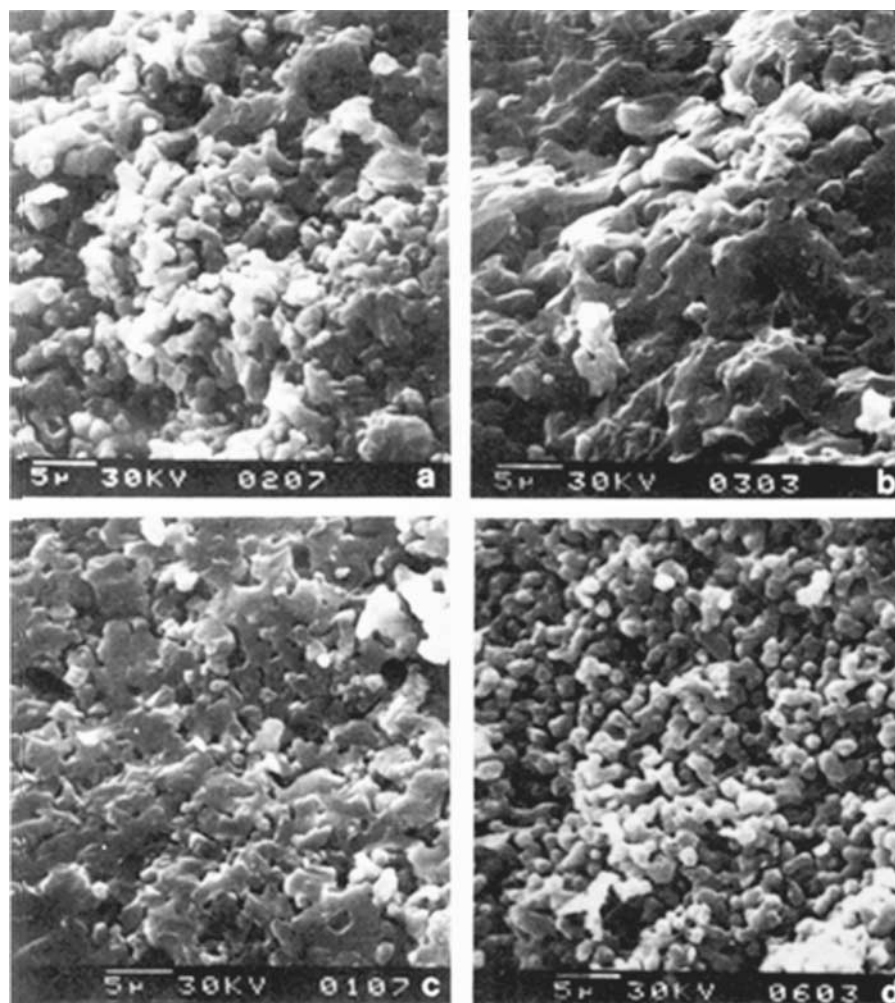


Fig. 44 Scanning electron micrographs of the structure cross-section of porous ceramic pellets of β -tricalcium phosphate (a–c) and hydroxyapatite (d).

ACKNOWLEDGMENTS

This chapter was written on the basis of a 13 years experience with SEM in pharmaceutical technology. The author would like to thank Mr. Klaus Weyhing for his permanent involvement in the practical work with the SEM and for his graphical support as well as Mrs. Renate Beer for typing and editing this work. Thanks should also be given to Professor Dr. Tsige Gebre-Mariam, University of Addis Ababa, for correction of the manuscript.

REFERENCES

1. Watt, I.M. Microscopy with Light and Electrons. *The Principle and Practice of Electron Microscopy*, 2nd Ed.; Cambridge University Press: New York, Melbourne, 1997; 8–10.
2. de Broglie, L.A. Tentative Theory of Light Quanta. *Phil. Mag.* **1924**, *47*, 446.
3. Watt, I.M. Microscopy with Light and Electrons. *The Principle and Practice of Electron Microscopy*, 2nd Ed.; Cambridge University Press: New York, Melbourne, 1997; 14.
4. DSM 940. A-Design of Instrument. *Operation Manual*; Zeiss: Oberkochen, 1990; 4–5.
5. Sommer, K. Sampling Procedures. *Sampling of Powders and Bulk Material*; Springer-Verlag: Berlin, New York, Tokyo, 1986; 80–124.
6. Sympatec Rhodos 12SR Rhodos Operationg Instructions. *Operation Manual*; Sympatec GmbH: Clausthal-Zellerfeld, 1985; 3.1–3.52.
7. Watt, I.M. Specimen Preparation for Electron Microscopy. *The Principle and Practice of Electron Microscopy*, 2nd Ed.; Cambridge University Press: New York, Melbourne, 1997; 136–187.

8. Shangraw, R.F.; Wallace, J.W.; Bowers, F.M. Morphology and Functionality in Tablet Excipients for Direct Compression. Part I. Pharm. Technol. **1981**, *5*, 69–78.
9. Doelker, E.; Massuelle, D.; Veuille, F.; Humbert-Droz, P. Morphological, Packing, Flow, and Tableting Properties of New Avicel Types. Drug Dev. Ind. Pharm. **1995**, *21*, 643–661.
10. Larhrib, H.; Wells, M.H.; Rubinstein, M.H.; Ricart, G. Characterization of PEGs using Matrix-Assisted Laser Desorption/Ionisation Mass Spectrometry and Other Related Techniques. Int. J. Pharm. **1997**, *14*, 187–198.
11. Khan, K.A.; Rhodes, C.T. The Production of Tablets by Direct Compression. Can. J. Pharm. Sci. **1973**, *8*, 1–5.
12. Shangraw, R.F. The Design and Evaluation of Coprocessed Tablet Excipients and Direct Compression Active Ingredients. FMC **1981**, *1*, 1–20.
13. Schmidt, P.C.; Herzog, R. Calcium Phosphates in Pharmaceutical Tableting. Part I: General Properties. Pharm. World Sci. **1993**, *15*, 105–115.
14. Schmidt, P.C.; Herzog, R. Calcium Phosphates in Pharmaceutical Tableting. Part II: Comparison of Tableting Properties. Pharm. World Sci. **1993**, *15*, 116–122.
15. Carstensen, J.T.; Ertel, L. Physical and Chemical Properties of Calcium Phosphates for Solid-State Pharmaceutical Formulations. Congr. Int. Technol. Pharm., 5th, Ass. Pharm. Gal. Ind. **1989**.
16. Hou, X.P.; Carstensen, J.T. Compression Characteristics of Basic Tricalcium Phosphate ($\text{Ca}_3(\text{PO}_4)_2$, $\text{Ca}(\text{OH})_2$). Int. J. Pharm. **1985**, *25*, 207–215.
17. Lerk, C.F. Physikalisch-Pharmazeutische Eigenschaften Von Lactose. Pharm. Unserer Zeit **1987**, *16*, 39–46.
18. Schmidt, P.C.; Vortisch, W. Einfluß Der Herstellungsart von Füll- und Bindemitteln auf Ihre Tablettierfähigkeit: Vergleich von Acht Marktüblichen Sorbit-Typen. Pharm. Ind. **1987**, *49*, 495–503.
19. Schmidt, P.C. Tableting Characteristics of Sorbitol. Pharm. Technol. **1983**, *7*(11), 65–74.
20. DuRoss, J.W. Modification of the Crystalline Structure of Sorbitol and Its Effects on Tableting Characteristics. Pharm. Technol. **1984**, *8*, 42–53.
21. Schmidt, P.C.; Rubensdörfer, C. Evaluation of Ludipress as a “Multi-Purpose Excipient” for Direct Compression. Part I: Powder Characteristics and Tableting, Drug Dev. Ind. Pharm. **1994**, *20*, 2899–2925.
22. Pesonen, T.; Paronen, P.; Puurunen, T. Evaluation of a Novel Cellulose Powder as a Filler-Binder for Direct Compression of Tablets. Pharm. Weekbl. Sci. Ed. **1989**, *11*, 13–19.
23. Schmidt, P.C.; Rubensdörfer, C. Evaluation of Ludipress as a “Multi-Purpose Excipient” for Direct Compression. Part II: Interactive Blending and Tableting with Micronized Glibenclamide Drug, Dev. Ind. Pharm. **1994**, *20*, 2927–2952.
24. Gebre-Mariam, T.; Schmidt, P.C. Isolation and Physico-Chemical Properties of Enset Starch. Starch/Stärke **1996**, *48*, 208–214.
25. Bolhuis, G.K.; van Kamp, H.V.; Lerk, C.F. On the Similarity of Sodium Starch Glycolate from Different Sources. Drug Dev. Ind. Pharm. **1986**, *12*, 621–630.
26. Ferrero, C.; Munoz, N.; Velasco, M.V.; Munoz-Ruiz, A.; Jiménez-Castellanos, R. Disintegrating Efficiency of Croscarmellose Sodium in a Direct Compression Formulation. Int. J. Pharm. **1997**, *147*, 11–21.
27. Bertoni, M.; Ferrari, F.; Bonferoni, M.C.; Rossi, S.; Caramella, C. Functionality Tests for Tablet Disintegrants: The Case of Sodium Carboxymethylcelluloses. Pharm. Technol. Eur. **1995**, *7*, 17–24.
28. Ertel, K.D.; Carstensen, J.T. Chemical, Physical, and Lubricant Properties of Magnesium Stearate. J. Pharm. Sci. **1988**, *77*, 625–629.
29. Steffens K.-J. Koglin J. The Magnesium Stearate Problem III, International Conference on Pharmaceutical Ingredients and Intermediates 1992; 118–125
30. Delacourte, A.; Guyot, J.C.; Colombo, P.; Catellani, P.L. Effectiveness of Lubricants and Lubrication Mechanism in Tablet Technology. Drug Dev. Ind. Pharm. **1995**, *21*, 2187–2199.
31. Frömming, K.-H.; Gröbler, S. Einfluß Von Füllmitteln Und Von Magnesiumstearat Auf Die Wirkstofffreisetzung Aus Hartgelatine kapseln. Pharm. Ztg. **1983**, *128*, 786–793.
32. Roblot, L.; Puisieux, F.; Duchêne, D. A Study on Lubrication by Magnesium Stearate. The Influence of the Proportion of Lubricant and the Mixing Process on the Tablet Characteristics. Labo. Pharma. Probl. Tech. **1983**, *31*, 843–847.
33. Jarosz, P.J.; Parrott, E.L. Effect of Lubricants on Tensile Strengths of Tablets. Drug Dev. Ind. Pharm. **1984**, *10*, 259–273.
34. Williams, R.O., III; McGinity, J.W. A Study of the Influence of Magnesium Stearate or Talc on the Compaction Properties of Direct Compression Excipients using Tableting Indices. Pharm. Tech. Tabl. Tech. **1987**, *1*, 157–165.
35. Stange, K. Beurteilung Von Mischgeräten Mit Hilfe Statistischer Verfahren. Chem. Ing. Tech. **1954**, *26*, 150–155.
36. Stange, K. Die Mischgüte Einer Zufallsmischung Als Grundlage Zur Beurteilung Von Mischversuchen. Chem. Ing. Tech. **1954**, *26*, 331–337.
37. Egermann, H.; Frank, P. Novel Approach to Estimate Quality of Binary Random Powder Mixtures: Samples of Constant Volume. I: Derivation of Equation. J. Pharm. Sci. **1992**, *81*, 551–555.
38. Egermann, H.; Frank, P. Novel Approach to Estimate Quality of Binary Random Powder Mixtures: Samples of Constant Volume. II: Applications of Equation to Optimize Tableting Conditions. J. Pharm. Sci. **1992**, *81*, 667–669.
39. Egermann, H.; Krumphuber, A.; Frank, P. Novel Approach to Estimate Quality of Binary Random Powder Mixtures: Samples of Constant Volume. III: Range of Validity of Equation. J. Pharm. Sci. **1992**, *81*, 773–776.
40. Hersey, J.A. Ordered Mixing: A New Concept in Powder Mixing Practice. Powder Technol. **1975**, *11*, 41–44.
41. Egermann, H.; Orr, N.A. Ordered Mixtures—Interactive Mixtures. Powder Technol. **1983**, *36*, 117–118.
42. Schmidt, P.C.; Ben, E.S. Vereinfachung Der Kapselrezeptur Durch Bildung Inaktiver Mischungen. Pharm. Ztg. **1987**, *132*, 2550–2557.
43. Schmidt, P.C.; Walter, R. Investigation of the Cohesion Behavior of Powders and Their Adhesion to a Carrier by an Electronic Tensiometer. Pharmazie **1994**, *49*(2,3), 183–187.
44. Sievert, M.; Blume, H.; Stenzhorn, G.; Lenhard, G.; Kieferndorf, U. Zur Qualitätsbeurteilung Von Phenoxy-methylpenicillinhaltigen Fertigdarzneimitteln 2. Mitteilung:

- Vergleich Der Bioverfügbarkeiten Eines Trockensaftes Und Einer Üligen Suspension. *Pharm. Ztg. Wiss.* **1990**, *135*, 103–107.
45. Ryder, J. The Formulation of Dry Syrup Preparations. *Int. J. Pharm. Technol. Prod. Manuf.* **1979**, *1*, 14–25.
 46. Rothgang, G. APV-Info. **1974**, *20*, 39.
 47. Vervaet, C.; Baert, L.; Remon, J.P. Extrusion-Spheronisation—A Literature Review. *Int. J. Pharm.* **1995**, *116*, 131–146.
 48. Schmidt, W.G.; Mehnert, W.; Frömming, K.-H. Controlled Release from Spherical Matrices Prepared in a Laboratory Scale Rotor-Granulator—Release Mechanism Interpretation using Individual Pellet Data. *Eur. J. Pharm. Biopharm.* **1996**, *42*, 348–350.
 49. Eriksson, M.; Alderborn, G.; Nyström, C.; Podczek, F.; Newton, J.M. Comparison between and Evaluation of Some Methods for the Assessment of the Sphericity of Pellets. *Int. J. Pharm.* **1997**, *148*, 149–154.
 50. Lindner, H.; Kleinebudde, P. Anwendung Der Automatischen Bildanalyse Zur Charakterisierung Von Pellets. *Pharm. Ind.* **1993**, *55*, 694–701.
 51. Kawashima, Y. An Experimental Study of the Kinetics of Spherical Agglomeration in a Stirred Vessel. *Powder Technol.* **1974**, *10*, 85–92.
 52. Kawashima, Y. Spherical Crystallization as a Novel Particle Design Technique for the Development of Pharmaceutical Preparations. *Kona*. **1987**, *5*, 75.
 53. Wolf, J. Untersuchung Der Hergestellten Agglomerate. *Herstellung, Charakterisierung und Tablettiereigenschaften Von Sphärischen Arzneistoffagglomeraten*; Thesis University of Tübingen: Tübingen, 1990; 59–61.
 54. Hess, H. Tablets Under the Microscope. *Pharm. Technol. Int.* **1978**, *1*, 18–43.
 55. Beckert, T.E.; Lehmann, K.; Schmidt, P.C. Compression of Enteric-Coated Pellets to Disintegrating Tablets. *Int. J. Pharm.* **1996**, *143*, 13–23.
 56. McGinnity, J.W. *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*; Marcel Dekker, Inc.: New York, 1997.
 57. Rowe, R.C. The Effect of Pigment Type and Concentration on the Incidence of Edge Splitting on Film-Coated Tablets. *Pharm. Acta. Helv.* **1982**, *57*, 221–225.
 58. Thomas, N.L. The Barrier Properties of Paint Coatings. *Prog. Org. Coat.* **1991**, *19*, 102–121.
 59. Asbeck, W.K.; Van Loo, M. Critical Pigment Volume Relationships. *Ind. Eng. Chem.* **1949**, *41*, 1470–1475.
 60. Maul, K.A.; Schmidt, P.C. Influence of Different-Shaped Pigments and Plasticizers on Theophylline Release from Eudragit RS30D and Aquacoat ECD30 Coated Pellets. *S. T. P. Pharm. Sci.* **1997**, *7*, 498–506.
 61. Herbener, R. Staubarme Pulver Durch Sprühtrocknen—Erfahrungen Mit Neueren Techniken. *Chem. Ing. Tech.* **1987**, *59*, 112–117.
 62. Vromans, H.; Bolhuis, G.K.; Lerk, C.F.; van de Biggelaar, H.; Bosch, H. Studies on Tableting Properties of Lactose. VII. The Effect of Variations in Primary Particle Size and Percentage of Amorphous Lactose in Spray-Dried Lactose Products. *Int. J. Pharm.* **1987**, *35*, 29–37.
 63. Nürnberg, E.; Dölle, B. Darstellung Von Digoxin-Sprüheinbettungen Aus wäßrigen Systemen: Eigenschaften Dieser Produkte. *Pharm. Ind.* **1980**, *42*, 1019–1026.
 64. Röscheisen, G.; Schmidt, P.C. Preparation and Optimization of L-Leucine as Lubricant for Effervescent Tablet Formulations. *Acta Pharm. Helv.* **1995**, *70*, 133–139.
 65. Rinkel, R. Sprühtrocknung Pharmazeutischer Produkte. *APV-Info.* **1970**, *15*, 170–180.
 66. Oetjen, G.-W. *Gefriertrocknen*; VCH-Verlagsgesellschaft mbH: Weinheim, 1997; 3.
 67. Corveleyn, S.; Remon, J.P. Formulation and Production of Rapidly Disintegrating Tablets by Lyophilisation using Hydrochlorothiazide as a Model Drug. *Int. J. Pharm.* **1997**, *152*, 215–225.
 68. Chen, T. Formulation Concerns of Protein Drugs. *Drug Dev. Ind. Pharm.* **1992**, *18*, 1311–1354.
 69. Brannon-Peppas, L. Recent Advances on the Use of Biodegradable Microparticles and Nanoparticles in Controlled Drug Delivery. *Int. J. Pharm.* **1995**, *116*, 1–9.
 70. Lill, N. Injizierbare Arzneiformen Auf Basis Biodegradabler Polymere. *Pharm. Ztg. Prisma* **1995**, *2*, 269–278.
 71. Schmidt, P.C.; Maul, K.A. Moderne Arzneiformen für Urogenitalerkrankungen. *Dtsch. Pharm. Ztg.* **1993**, *138*, 837–842.
 72. Thoma, K.; Alex, R.; Fensch-Kleemann, E. Biodegradable Controlled Release Implants Based on β -Tricalcium Phosphate Ceramic. I. Preparation and Characterization of Porous β -Tricalcium Phosphate Ceramic Pellets. *Eur. J. Pharm. Biopharm.* **1992**, *38*, 101–106.

FURTHER READING

- Alexander, H. *Physikalische Grundlagen der Elektronenmikroskopie*; Teubner Studienbücher: Stuttgart, 1997.
- Flegler, S.L.; Heckman, J.W.; Klomparens, K.L. *Elektronenmikroskopie: Grundlagen, methoden, Anwendungen*; Spektrum: Heidelberg, 1995.
- Watt, I.M. *The Principles and Practice of Electron Microscopy*, 2nd Ed.; Cambridge University Press: Cambridge, 1997.

SEMISOLID PREPARATIONS

Guru Betageri
Sunil Prabhu

Western University of Health Sciences, Pomona, California, U.S.A.

INTRODUCTION

Pharmaceutical semisolid preparations may be defined as topical products intended for application on the skin or accessible mucous membranes to provide localized and sometimes systemic effects at the site of application. In general, semisolid dosage forms are complex formulations having complex structural elements (1). They are often composed of two phases (oil and water), one of which is a continuous (external) phase and the other a dispersed (internal) phase. The active ingredient is often dissolved in one or both phases, thus creating a three-phase system. The physical properties of the dosage form depend on various factors, including the size of the dispersed particles, the interfacial tension between the phases, the partition coefficient of the active ingredient between the phases, and the product rheology. These factors combine to determine the release characteristics of the drug as well as other characteristics such as viscosity. Although a majority of semisolid preparations contain medicinal agents for therapeutic effect, some nonmedicated semisolid preparations are used for their physical effects as protectants and lubricants. The design of a semisolid preparation is based on its ability to adhere to the surface of application for a reasonable duration before they are washed or worn off. The adhesion is brought about by a plastic rheologic behavior that allows the semisolids to retain shape and cling as a film until acted on by an outside force, whereupon they deform and flow. One can assess this behavior by thrusting a finger through the semisolid mass, which leaves a track that does not fill up when the finger is withdrawn (2, 3). Semisolids are characterized by a three-dimensional structure that is sufficient to impart solid-like character to the undisturbed system but that is easily broken down and realigned under an applied force. In broad terms, semisolid preparations may be classified as ointments, creams, pastes, and gels.

Ointments are composed mostly of fluid hydrocarbons meshed in a matrix of higher melting solid hydrocarbons. Common examples of ointment bases include mineral oil, petrolatum, and polyethylene glycol. Creams are semisolid emulsion systems with an opaque appearance. Their

consistency and rheologic properties are based on whether the emulsion is o/w or w/o and on the nature of the solid in the internal phase. Pastes are basically ointments into which a high percentage of insoluble solids has been added. Powders such as zinc oxide, titanium dioxide, starch, and kaolin are incorporated in high concentrations into a preferably lipophilic greasy vehicle to form a paste-like mass. Gels are semisolid systems in which a liquid phase is constrained within a three-dimensional polymeric matrix in which a high degree of physical cross-linking has been introduced. Most of the semisolid preparations are applied to the skin for topical relief of dermatologic conditions, whereas a lesser portion of these preparations are applied to mucous membranes such as rectal and buccal tissue, vaginal mucosa, urethral membrane, external ear lining, nasal mucosa, and cornea. Whereas normal skin permeation is limited to localized effects, the mucous membranes permit more ready access to systemic circulation. Overall, even though the nature of the underlying structures differ remarkably across all the different semisolid systems, they all share the property that their structures are easily broken down, rearranged, and reformed.

OINTMENTS AND CREAMS

Ointments utilize certain bases that act as vehicles to deliver the drug and to impart emollient and lubricant properties to the preparation. Usually, but not always, they contain medicinal substances (4). Properties of ointments may vary from product to product depending on their specific use, ease, and extent of application. In general, ointment bases may be classified into four general groups: hydrocarbon, absorption, water-removable, and water-soluble bases.

Hydrocarbon Bases

Also known as oleaginous bases, the hydrocarbon bases are essentially water-free, incorporating aqueous preparations

only in small amounts and with considerable difficulty. The primary features of this type of base include its emollient effect, retention on the skin for prolonged periods, prevention of escape of moisture from the skin to the atmosphere, and difficulty in washing off. They act as occlusive dressings, thus increasing skin hydration by reducing the rate of loss of surface water. Also, they do not dry out or change noticeably on aging. Hydrocarbon-based semisolids comprise fluid hydrocarbons C₁₆ to C₃₀ straight chain and branched, entrapped in a fine crystalline matrix of yet higher-molecular-weight solid hydrocarbons. The high-molecular-weight fraction precipitates out substantially above room temperature, forming interlocking crystallites (5). The extent and specific nature of this structure determine the stiffness of the ointment. In general, hydrocarbon-based ointments liquefy on heating because the crystallites melt. Moreover, when cooled very slowly, they assume fluidity much greater than when rapidly cooled because slow cooling leads to fewer and larger crystallites and, therefore, less total structure.

Common examples of these bases include:

1. *Petrolatum, USP*—a mixture of semisolid hydrocarbons obtained from petroleum. It is an unctuous mass, varying in color from yellowish to light amber. It melts at temperatures between 38 and 60°C and may be used alone or in combination with other agents as an ointment base. Petrolatum of varying melting ranges and consistencies are commercially available as given in Table 1.
2. *White petrolatum, USP*—a petrolatum that has been decolorized, either partially or wholly. It is used for the same purpose as petrolatum but is more esthetically acceptable to a patient than petrolatum because of its lighter color. White petrolatum is particularly useful to treat diaper rash (impervious to

urine and protects the baby’s skin) and dry skin (helps the skin retain moisture). Example 1 illustrates an ointment preparation that incorporates 95% white petrolatum resulting in an ointment base with firm consistency.

3. *Yellow ointment, USP*—the purified wax obtained from the honeycomb of the bee. It contains 5% yellow wax and 95% petrolatum in the formulation.
4. *Mineral oil*—a mixture of liquid hydrocarbons obtained from petroleum. These are useful as levigating agents to wet and incorporate solid substances (e.g., salicylic acid, zinc oxide) into the preparation of ointments that consist of oleaginous bases as their vehicle. There are two types of mineral oils listed in the *U.S. Pharmacopeia/National Formulary* (USP/NF). Mineral oil USP is also called heavy mineral oil with a specific gravity between 0.845 and 0.905 and a viscosity of not less 34.5 cSt (cSt = mm²/s) at 40°C. Light mineral oil, NF has a specific gravity between 0.818 and 0.880 and a viscosity of not more than 33.5 cSt. Table 2 lists the commercially available mineral oil fractions.

Blending of increasing quantities of mineral oil with petrolatum can produce ointments of various consistencies as desired. For example, blending 10% w/w mineral oil with 90% w/w white petrolatum base can produce an ointment with less drag or resistance to spreading, making it ideal for application to burns or other painful areas. As illustrated in Example 2, melting the petrolatum and mineral oil together and allowing them to cool forms a soft base or vehicle. By increasing the quantity of the mineral oil in the mixture, a base is

Table 1 Specifications of a range of petrolatum^a

	Melting point °F (°C)	USP or ASTM consistency	Sabolt viscosity at 210°F	Typical congealing point °F (°C)
White petrolatum, USP	122/135 (50–57)	175/205	64/75	125 (51.6)
	118/130 (47–54)	210/240	57/70	120 (48.8)
	130/140 (54–60)	155/190	60/70	130 (54.4)
Yellow petrolatum, USP	118/130 (47–54)	210/240	57/70	118 (47.7)
	122/135 (50–57)	175/205	57/70	123 (50.5)

^aWitco Chemical Corporation, Greenwich, CT.

Table 2 Specifications of a range of mineral oils^a

Type	Viscosity, cSt at 40°C	Specific gravity at 60°C
Mineral oil, USP	65.8/71.0	0.870/0.887
	72.0/79.5	0.864/0.878
	60.0/63.3	0.863/0.883
	38.4/41.5	0.859/0.882
	34.9/37.3	0.858/0.882
Light mineral oil, NF	28.1/30.3	0.856/0.882
	24.2/26.3	0.854/0.873
	17.7/20.2	0.842/0.870
	14.2/17.0	0.845/0.860
	7.6/8.7	0.831/0.842

^aPenzoil Products Company, Karns City, PA.

obtained that has a gel-like consistency indicating a more viscous preparation (Example 3). Hydrocarbon vehicles have several advantages, such as stability and emolliency; however, they suffer from one major disadvantage—greasiness—which may stain clothing and is usually difficult to remove (6).

Ophthalmic Ointments

Ophthalmic ointments differ from conventional ointments in that they must be sterile. Although the USP does not specify any particular base in the ophthalmic drug monographs, it states that the ingredients used must be sterilized under rigid aseptic conditions and again in its final container (7). In choosing an ointment base for an ophthalmic preparation, it must meet several qualities such as having the ability to diffuse the medication throughout the secretions bathing the eye, being nonirritating to the eye, and having the ability to diffuse the medication throughout the secretions bathing the eye. It must also have a softening point close to body temperature, both for patient comfort and for drug release. To avoid the risk of damage to the eye, it is imperative that the ointment be free of large particles and any metal particles. Mixtures of mineral oil and white petrolatum are commonly used as the base in medicated and nonmedicated (lubricating) ophthalmic ointments. Usually, the medicinal agents are added to an ointment base either as a solution or as a finely micronized powder and are subjected to fine milling to render the preparation uniform and smooth. The ophthalmic ointments also must meet the USP sterility tests and the test for metal particles in ophthalmic ointments. Rendering an ophthalmic ointment sterile requires special

Example 1 White ointment, USP

White petrolatum	95 (% w/w)
White wax	5

Procedure: Melt the white wax and add the petrolatum; continue heating until a liquid melt is formed. Congeal with stirring. Heating should be gentle to avoid charring (steam is preferred), and air incorporation by too vigorous stirring is to be avoided.

Example 2 Soft petrolatum base

White petrolatum, USP	90.0 (% w/w)
Mineral oil, NF	10.0

Procedure: Melt the white petrolatum and mineral oil together and allow to cool.

Example 3 Gelled petrolatum base

White petrolatum, USP	75.0 (% w/w)
Mineral oil, NF	25.0 (% w/w)

Procedure: Melt the white petrolatum and mineral oil together and allow to cool.

aseptic techniques and processing. Each drug, along with other components, is rendered sterile separately, aseptically weighed, and incorporated in preparing a final product that meets the sterility requirement (8). This is done because of difficulty in terminal product sterilization, such as lack of penetration of steam into the ointment base and instability of components owing to high dry heating. Antimicrobial preservatives such as methylparaben (0.05%) and propylparaben (0.01%) and its combinations phenylmercuric acetate (0.0008%), chlorobutanol (0.5%), and benzalkonium chloride (0.008%) are used as needed.

Absorption Bases

Absorption bases, as such, are hydrophilic, anhydrous materials (w/o emulsions) or hydrous bases (w/o emulsions that have the ability to absorb additional water). Addition of lanolin, lanolin isolates, cholesterol, lanosterol, or acetylated sterols renders the hydrocarbon base hydrophilic. Such hydrophilic mixtures have been known as absorption bases; however, the word absorption is a misnomer. Although the bases do eventually absorb aqueous solutions to be considered w/o emulsions, they do not absorb water on contact, but after sufficient agitation

only. They are conventional ointments that contain w/o emulsifiers in appreciable quantity. A w/o emulsion is formed when an aqueous medium, perhaps containing the drug in solution, is worked into the base. Thus, these bases can be classified as: 1) those that permit the incorporation of aqueous solutions, resulting in the formation of w/o emulsions (e.g., hydrophilic petrolatum), and 2) those that are already w/o emulsions (emulsion bases) that permit the incorporation of small, additional quantities of aqueous solutions (e.g., cold cream). These bases are useful as emollients, although they do not provide the degree of occlusion as oleaginous bases. They are also used pharmaceutically to incorporate aqueous solutions of drugs (e.g., sodium sulfacetamide solution) into oleaginous bases. A typical example of an anhydrous absorption base is hydrophilic petrolatum, USP (Example 4). Here, cholesterol confers the w/o emulsion property, whereas inclusion of stearyl alcohol and white wax enhances firmness and heat stability. Diverse additives including cholesterol, lanolin (which contains cholesterol, cholesterol esters, and other emulsifiers), semisynthetic lanolin derivatives, and assorted ionic and nonionic surfactants are used to emulsify water into these systems, singularly or in combination. Lanolin is probably the best known substance for the emulsification of water in an anhydrous base. Anhydrous lanolin, USP is capable of absorbing up to 30% of its weight of water to form an emulsion. Its water-absorbing capacity is improved to 50% by the addition of cholesterol. Among the absorption bases, lanolin is the oldest and best known. However, because of its tackiness and viscous nature and reports of allergic reactions, the use of lanolin is very limited. Advents in technology have successfully produced a number of emulsifiers for w/o emulsions. Used in specific formulations, emulsifiers such as polyglyceryl esters form w/o emulsions (Example 5). Absorption bases impart excellent emolliency and a degree of occlusiveness on application. However, they are also greasy when applied and are difficult to remove. The anhydrous types of absorption

Example 4 Hydrophilic petrolatum, USP (anhydrous absorption base)

White petrolatum	86.0 (% w/w)
Cholesterol	3.0 (% w/w)
Stearyl alcohol	3.0 (% w/w)
White wax	8.0 (% w/w)

Procedure: Melt the stearyl alcohol and white wax together on a steam bath, then add the cholesterol and stir until it completely dissolves. Add the white petrolatum and mix. Remove from the bath and stir until mixture congeals.

Example 5 Water-in-oil emulsion formulation

Mineral oil	40.0 (% w/w)
White wax	6.0 (% w/w)
Polyglyceryl 5 trioleate	8.0 (% w/w)
Isopropyl palmitate	3.0 (% w/w)
Sorbitol monostearate	3.5 (% w/w)
Polysorbate 60	2.5 (% w/w)
Propylene glycol	4.0 (% w/w)
Water	33.0 (% w/w)
Preservative, q.s.	

bases can be used when the presence of water causes stability problems with specific drug substances such as antibiotics. Commercially available absorption bases include Aquaphor (Beiersdorf, Hamburg, Germany) and Polysorb (Fougera, Melville, NY). Absorption bases, either hydrous or anhydrous, are seldom used as vehicles for commercial drug products because the w/o emulsion is more difficult to handle than the more conventional o/w system, and also there is reduced patient acceptance because of greasiness.

A new group of w/o emulsifiers based on the linkage of polymethoxysiloxane chains with alkyl side chains and polyol groups has been constituted (9). These emulsifiers, known as organosilicone polymers, have the capability of producing w/o emulsions with a high water content. Although the polymethylsiloxane chains possess both hydrophilic and lipophilic properties, the alkyl side chains supply the necessary lipophilic and polyol groups to provide the hydrophilic characteristics of the emulsifier. Other examples of w/o emulsifiers include cetyl dimethicone copolymer (10), polyethylene glycol-20-com glycerides (11), and a series of caprylic-capryl stearates (12).

**Water-Removable Bases
(Water-Washable Creams)**

These are the most commonly used o/w emulsion bases that are capable of being washed from skin or clothing with water. They may contain water-soluble and -insoluble components. From a therapeutic viewpoint, they have the ability to absorb serous discharges in dermatologic conditions. The water-removable bases form a semipermeable film on the site of application after the evaporation of water. As such, the base consists of three component parts: the oil phase, the emulsifier, and the aqueous phase. The oil phase, also called the internal phase, is typically made up of the petrolatum

and/or liquid petrolatum. Other ingredients such as cetyl and stearyl alcohol may be added to make up the oil phase. A typical water-removable emulsion base is hydrophilic ointment, USP, as shown in Example 6. The stearyl alcohol serves as an adjuvant emulsifier. Petrolatum in the oil phase contributes to the water-holding ability of the overall formulation. The aqueous phase contains preservative materials, emulsifier, and humectant. Humectants are added to minimize water loss in the finished composition and to add to the overall physical product acceptability. Common examples of humectants used include glycerin, propylene glycol, and a polyethylene glycol. The aqueous phase also contains the water-soluble components of the emulsion system, together with any additional stabilizers, antioxidants, buffers, etc., that may be necessary for stability, pH control, and other considerations associated with aqueous systems. Another example of an o/w emulsion base is illustrated in Example 7 (vanishing cream). The vanishing creams are so called because on application and rubbing into the skin, there is little or no visible evidence of their presence. The addition of an emulsifying agent is critical to formulating an emulsion. Emulsifiers must meet the following criteria before incorporation:

1. be a surfactant to reduce surface tension;
2. be able to prevent coalescence by being absorbed quickly around the dispersed droplets;
3. facilitate mutual repulsion between particles by imparting an adequate electrical potential to the droplets;
4. be able to increase viscosity to ensure a semisolid system; and
5. be effective at low concentrations. Emulsifiers used in cream formulations may be classified into three different categories: anionic, cationic, and nonionic emulsifiers.

Anionic emulsifiers

The active portion of this class of emulsifiers is the anion. In general, these emulsifiers are more acid-stable and permit adjustment of the emulsion pH level to the desirable range of 4.5 and 6.5. Common examples include sodium lauryl sulfate and soaps such as triethanolamine stearate. Triethanolamine stearate is one of the most popular emulsifiers for creams and lotions in use today. It is usually prepared in situ during manufacture from stearic acid in the hot oil phase and from triethanolamine in the hot aqueous phase. The amount of triethanolamine controls the pH level of the resulting product.

Example 6 Hydrophilic ointment, USP

Stearyl alcohol	25.0 (% w/w)
White petrolatum	25.0 (% w/w)
Methylparaben	0.025 (% w/w)
Propylparaben	0.015 (% w/w)
Sodium lauryl sulfate	1.0 (% w/w)
Propylene glycol	12.0 (% w/w)
Purified water	37.0 (% w/w)

Procedure: Melt the stearyl alcohol and white petrolatum in a steam bath and warm to 75°C. Add the other ingredients previously dissolved in water and warmed to 75°C, and stir mixture until it congeals.

Example 7 Vanishing cream (cream base o/w)

<i>Oleaginous phase</i>	
Stearic acid	13.0 (% w/w)
Stearyl alcohol	1.0
Cetyl alcohol	1.0
<i>Aqueous phase</i>	
Glycerin	10.0
Methylparaben	0.1
Propylparaben	0.05
Potassium hydroxide	0.9
Purified water, q.s.	100

Procedure: Heat the oil phase and water phase to approximately 65°C. Add the oil phase slowly to the aqueous phase with stirring to form a crude emulsion. Cool to approximately 50°C and homogenize. Cool with agitation until congealed.

Cationic emulsifiers

Cationic compounds are highly surface-active but are used less frequently as emulsifiers. The cation portion of the molecule is usually a quaternary ammonium salt including a fatty acid derivative such as dilauryldimethylammonium chloride. These emulsifiers are irritating to the skin and eyes and have a considerable range of incompatibilities, including anionic materials.

Nonionic emulsifiers

This class of emulsifiers shows excellent pH and electrolyte compatibility in emulsions, owing to the fact that they do not ionize in solution. Although nonionic emulsifiers range from lipophilic to hydrophilic members, a typical emulsifier system may include a mixture of both a lipophilic and a hydrophilic member to produce a hydrophilic-lipophilic balance (HLB). As devised by Griffin (13), an HLB scale can be used to establish a range of optimum efficiency for a suitable emulsifier for a certain activity such as wetting agent,

Table 3 Activity and HLB value of emulsifiers

Activity	Assigned HLB
Antifoaming	1–3
Emulsifiers (w/o)	3–6
Wetting agents	7–9
Emulsifiers (o/w)	8–18
Solubilizers	15–20
Detergents	13–15

antifoaming agent, detergent, etc. Although the numbers have been assigned up to 40, the usual range is between 1 and 20. Materials that are highly polar or hydrophilic are assigned higher numbers than materials that are less polar and more lipophilic. Generally, those emulsifiers with an assigned value between 3 and 6 are greatly lipophilic and produce w/o emulsions, whereas those in the range between 8 and 18 are hydrophilic and produce o/w emulsions. By using this scale, it is possible to relate various emulsifiers to suitable applications. Table 3 illustrates the relationship between HLB numbers and the type of activity expected from emulsifiers. Emulsions containing nonionic emulsifiers are prepared by dissolving or dispersing the lipophilic component in the oil phase and the hydrophilic component in the aqueous phase. The two phases are heated separately and combined. Emulsions containing nonionic emulsifiers are generally low in irritation potential, stable, and have excellent compatibility characteristics. Examples 8 and 9 illustrate the use of these emulsifiers in the formulation of creams.

The Cosmetic, Toiletry and Fragrance Association's *International Cosmetic Ingredient Dictionary* (14)

Example 8 Oil-in-water cream formulation HLB 13

Part A	
Stearic acid	10.0 (% w/w)
Mineral oil	8.0 (% w/w)
White petrolatum	6.0 (% w/w)
Sorbitan monostearate	2.0 (% w/w)
Part B	
POE ^a sorbitan monostearate	1.0 (% w/w)
Water	73.0 (% w/w)
Preservative, q.s.	

Procedure: Part A is heated at 70°C until completely melted. Part B is added at 70°C and the mixture is cooled to room temperature.

^aPolyoxyethylene.

Example 9 Water-in-oil cream formulation HLB 12

Part A	
Mineral oil	50.0 (% w/w)
White wax	15.0 (% w/w)
Sorbitan monostearate	2.0 (% w/w)
Part B	
POE ^a sorbitan monostearate	3.0 (% w/w)
Water, q.s.	30.0 (% w/w)

Procedure: Same as in Example 8.

^aPolyoxyethylene

provides an exhaustive listing of the variety of emulsion-base components, particularly the oil phase components. Over 6000 ingredients, cross-referenced to more than 25,000 trade names and synonyms, are presented. These include emulsifiers, coemulsifiers, surfactants, and stabilizers. Another source is *The McCutcheons Detergents and Emulsifiers* (15), which also lists thousands of emulsifiers and surfactants. These lists are important sources of information and help formulators in the pharmaceutical and cosmetic industry. Commonly used oils and surfactants in formulating semisolid emulsions are shown in Tables 4 and 5.

Microemulsions

Microemulsions are fluid, transparent, thermodynamically stable oil and water systems, stabilized by a surfactant usually in conjunction with a cosurfactant that may be a short-chain alcohol, amine, or other weakly amphiphilic molecule (16, 17). An interesting characteristic of microemulsions is that the diameter of the droplets is in the range of 100–1000 Å, whereas the diameter of droplets in a kinetically stable macroemulsion is 5000 Å. The small droplet size allows the microemulsion to act as carriers for drugs that are poorly soluble in water. The suggested method of preparation of microemulsions is as follows: the surfactant, oil, and water are mixed to form a milky emulsion and titrated with a fourth component, the cosurfactant, until the mixture becomes clear (18). If more oil is added to a w/o system, the system becomes cloudy, but the addition of more cosurfactant again gives a clean transparent emulsion. As illustrated in Example 10, the formulation results in clear, transparent “ringing gels” with good stability. Microemulsions are optically clear because the diameter of the particles in the colloidal system is less than one-quarter of the wavelength of incident light. Because of this effect, the particles do not scatter light and result in a transparent system (19, 20). Microemulsions, once made, must be packaged while hot

Table 4 Oil phase emulsion components

Type	Composition	Physical state
Hydrocarbon	Mineral oils	Liquid
Hydrocarbon	Petrolatum	Semisolid
Hydrocarbon	Polyethylene polymers	Solid
Hydrocarbon	Synthetic waxes	Solid
Esters	Vegetable oils	Liquid
Ester	Lanolin	Liquid
Esters	Synthetics (e.g., butyl stearate)	Liquid or solid
Alcohols	Long chain	Liquid or solid
Fatty acids	Long chain	Liquid or solid
Ethers	Polyoxyethylene	Liquid or solid
	Polyoxypropylene	Liquid or solid
Silicones	Polymers	Liquid or solid
Natural products	Plant and animal waxes	Solid

and in the liquid form. After they are cooled, they cannot be reheated and melted because they will lose transparency because of coalescence of the oil globules to a size where they reflect light.

Water-Soluble Bases

These bases contain only water-soluble components. Water-soluble bases are also referred to as greaseless because of a lack of oleaginous materials. Incorporation of aqueous solutions is difficult because they soften greatly with the addition of water. They are better used for nonaqueous or solid substances. Polyethylene glycols (PEG) make up the majority of components of the water-soluble base. PEGs may exist as liquids or waxy solids, identified by numbers that are an approximate indication

of their molecular weight. The lowest number signifies a liquid state, which transitions to a waxy solid state as the numbers increase. For example, PEG 400 is a liquid, whereas PEG 4000 is a waxy solid. Polyethylene glycol is a polymer of ethylene oxide and water represented by the formula $\text{OHCH}_2(\text{CH}_2\text{OCH}_2)\text{CH}_2\text{OH}$. They are nonvolatile, water-soluble, or water-miscible compounds and chemically inert. PEGs of interest as vehicles include the 1500, 1600, 4000, and 6000 products, ranging from soft, waxy solids to hard waxes. PEG, particularly 1500, can be used as a vehicle by itself, but better results are often obtained using blends of high- and low-molecular-weight glycols as in polyethylene glycol ointment, NF (Example 11). PEGs also serve as excellent bases for suppository insert dosage forms. Various combinations of PEGs may be used to achieve a suppository base of desired

Table 5 Emulsifiers for semisolid emulsions

Anionic	Cationic	Nonionic
Alkyl sulfates	Quaternary ammonium compounds	Polyoxyethylene alkyl-aryl ethers
Soaps	Alkoxyalkylamines	Polyoxypropylene alkyl-aryl ethers
Dodecyl benzenesulfonates		Polyoxyethylene fatty acid esters
Lactylates		Polyoxyethylene sorbitan esters
Sulfosuccinates		Polyoxyethylene-polyoxypropylene block polymers
Monoglyceride sulfates		
Phosphate esters		Sorbitan fatty acid esters
Silicones		Glyceryl fatty acid esters
Sarcosinates		
Taurates		

Example 10 Microemulsion

Mineral oil, NF	16.0 (% w/w)
Sucrose distearate	5.0 (% w/w)
Sucrose stearate	5.0 (% w/w)
Diethylamine oleylether—10 phosphate	2.0 (% w/w)
Diethylamine oleylether—3 phosphate	5.0 (% w/w)
1,3-Ethylhexane diol	2.5 (% w/w)
Propylene glycol	5.0 (% w/w)
Water	59.5 (% w/w)

consistency and characteristics. For example, PEG 1000 blended with PEG 4000 results in a higher melting product, whereas blending with a liquid polyethylene glycol lowers the melting point. Some drugs lower the melting points of PEG and, therefore, each ingredient must be considered in selecting a base. The advantage of PEG suppositories is that they may be designed to prevent melting at body temperature but rather to dissolve slowly in the body fluids. Thus, they can be prepared at temperatures higher than that of the body. They also can be stored without the need for refrigeration and do not tend to leak from the orifice on insertion because of higher melting temperatures.

PASTES

Pastes maybe defined as ointments incorporating a high percentage of insoluble particulate solids, sometimes as much as or more than 50% (2). The use of this high amount of insoluble particulate matter renders a stiffness to the system as a result of direct interactions between the dispersed particulates and by absorption of the liquid hydrocarbons from the vehicles onto the surface of the particles. Because of the stiffness, they remain in place after application and are used effectively to absorb serous secretions. Pastes as such are not suited for application to hairy parts of the body. Examples of insoluble ingredients serving as the dispersed phase include starch, zinc oxide, and calcium carbonate. Pastes

Example 11 Polyethylene glycol ointment, NF

Polyethylene glycol 3350	40 (% w/w)
Polyethylene glycol 400	60 (% w/w)

Procedure: Heat the two ingredients in a water bath to 65°C. Allow to cool and stir until congealed.

Example 12 Zinc oxide paste, USP

Zinc oxide	25.0%
Starch	25.0%
Calamine	5.0%
White petrolatum, q.s.	100%

Procedure: Titrate the calamine with the zinc oxide and starch and incorporate uniformly in the petrolatum by levigation in a mortar or on a glass slab with a spatula. Mineral oil should not be used as a levigating agent, because it would soften the product. A portion of petrolatum can be melted and used as a levigating agent if so desired.

make good protective barriers for the following reasons. In addition to forming an unbroken film, pastes also absorb and neutralize certain harmful chemicals before they reach the skin surface. This last feature is attributed to the presence of insoluble particulate matter within the paste formulations. For example, for the treatment of diaper rash, when spread over the baby's bottom, the pastes absorb irritants formed by bacterial action on urine. Pastes also provide a protective layer over skin lesions and, when covered with suitable dressings, prevent excoriation of the patient's skin by scratching. Pastes afford emollient action as do ointments. In addition, the water-impermeable film formed on application is opaque and thus can often serve as a sunblock. Pastes are less greasy than ointments because of the absorption of the fluid hydrocarbon fraction to the insoluble particles. A clinically distinctive feature, which is generally attributed to pastes, is the ability to absorb exudates by nature of the powder or other absorptive components (21). Among the few pastes in use today is zinc oxide paste (Lassar's Plain Zinc Paste) (Example 12), which is prepared by levigating and then mixing 25% each of zinc oxide and starch with white petrolatum. The product is very firm and is better able to protect the skin and absorb secretions than is zinc oxide ointment.

GELS

Gels are defined as semisolid preparations consisting of dispersions of small or large molecules in an aqueous liquid vehicle rendered jelly-like through the addition of a gelling agent (22). Gels are an intermediate state of matter, containing both solid and liquid components. The solid component comprises a three-dimensional network of interconnected molecules or aggregates that immobilizes the liquid in the continuous phase (23). Gels may be

Example 13 Carbomer 941 gel

Carbomer 941	0.5 (% w/w)
Glycerine	10.0 (% w/w)
Triethanolamine	0.5 (% w/w)
Water	89.0 (% w/w)
Preservative q.s.	

Procedure: Water, glycerine, and preservative are mixed and the carbomer added by sprinkling on the surface while constantly mixing at high speed. Triethanolamine is added with slow agitation until a clear viscous gel forms.

classified into two primary types: hydrogels, which have an aqueous continuous phase, and organogels, which have an organic solvent as the liquid continuous medium. Gels may also be classified based on the nature of the bonds involved in the three-dimensional solid network: chemical gels form when strong covalent bonds hold the network together, and physical gels form when hydrogen bonds and electrostatic and van der Waals interactions maintain the gel network (24). Gelling agents commonly used are synthetic macromolecules (e.g., carbomer 934), cellulose derivatives (e.g., carboxymethylcellulose and hydroxypropylmethylcellulose), and natural gums (e.g., tragacanth). Carbomers in particular are high-molecular-weight water-soluble polymers of acrylic acid cross-linked with allyl ethers of sucrose and/or pentaerythritol. The NF contains monographs for six such polymers: carbomers 910, 934, 934P, 940, 941, and 1342. They are used as gelling agents at concentrations of 0.5–2.0% in water. Carbomer 940 yields the highest viscosity: between 40,000 and 60,000 CP as a 0.5% aqueous dispersion. Depending on their polymeric composition, different viscosities result. Gels may be classified as two-phase or single-phase systems. A two-phase gel system consists of floccules of small distinct particles rather than large molecules, thus called

a two-phase system often referred to as a magma. Milk of magnesia (or magnesia magma), which comprises a gelatinous precipitate of magnesium hydroxide, is an example of such a system. The gel structure in the two-phase systems is not always stable and thus may thicken on standing, forming a thixotrope, and must be shaken before use to liquefy the gel and enable pouring. Single-phase systems are gels in which the macromolecules are uniformly distributed throughout a liquid with no apparent boundaries between the dispersed macromolecules and the liquid. Examples of such gels include tragacanth and carboxymethylcellulose. A typical gel formulation may contain, apart from the gelling agent and water, a drug substance, cosolvents such as alcohol and/or propylene glycol, antimicrobial preservatives such as methylparaben and propylparaben or chlorhexidine gluconate, and stabilizers such as edetate disodium. Medicated gels may be prepared for administration by various routes including topically to the skin or eye, nasally, vaginally, and rectally. Some simple gel formulations are shown in Examples 13 and 14. Example 15 gives the formulation for a firm, rigid gel.

MANUFACTURING METHODS

Ointments

There are two primary methods of manufacturing ointment dosage forms: incorporation and fusion. Each may be used on a small scale (laboratory) or large scale (industry) for the manufacture of semisolid dosage forms.

Incorporation (laboratory scale)

On a small scale, as in a pharmacy, small quantities of ointments may be prepared using a mortar and pestle or

Example 14 Carbomer 934 alcoholic gel

Carbomer 934 resin	3.0 (% w/w)
Glycerine	10.0 (% w/w)
Ethanol	40.0 (% w/w)
2-Ethylhexylamine	2.5 (% w/w)
Water	44.5 (% w/w)

Procedure: The carbomer is dispersed in the glycerine and water, and a solution of the 2-Ethylhexylamine in ethanol is added to the water solution with mixing until a clear transparent gel is formed.

Example 15 Solubilized mineral oil gel

Polyethylene glycol 3 oleyl ether phosphate	6.8 (% w/w)
Polyethylene glycol 3 oleyl ether	4.0 (% w/w)
Polyethylene glycol 5 oleyl ether	2.7 (% w/w)
Mineral oil 220 SUS	13.6 (% w/w)
2-Ethyl-1,3-Hexanediol	13.4 (% w/w)
Propylene glycol	1.4 (% w/w)
Water	68.1 (% w/w)

Procedure: All components except water are heated to 75°C. The water is added at 75°C to the oil with mixing to form a soft gel, which on stirring and cooling sets at approximately 35°C.

an ointment slab (porcelain or glass) and spatula. The finely divided drug material is dispersed into an appropriate vehicle by mixing the components or rubbing the ingredients together to form an ointment. Sometimes, nonabsorbent parchment paper may be used to cover the working surface, which is then subsequently disposed of, thus eliminating the time to clean the ointment slabs.

When incorporating solids by the method of spatulation, the ointment base is placed on one side of the slab, and the finely divided drug powder components are placed on the other side. A small portion of the powder is mixed with a small portion of the base and worked in together to form a uniform mass. The procedure is repeated with the remaining base and powdered materials until all of the components are uniformly combined and blended. A broad blade spatula, made of metal or hard plastic, may be used for this task. Hard plastic spatulas are used when a component (e.g., phenol) of the mixture sometimes reacts with the metal.

Often it is desirable to keep the particle size of powder components in an ointment to a minimum to avoid roughness or grittiness of the preparation. Methods of particle size reduction include use of the mortar and pestle and the process of levigation. In this, the finely divided powdered material is levigated thoroughly with a small quantity of ointment base to form a concentrate. Sometimes, levigating agents such as mineral oil (for oleaginous bases) and glycerin (for aqueous bases) can also be used to form smooth dispersions. The levigating agent should be physically and chemically compatible with the drug and base. The concentrate is then diluted geometrically with the remainder of the base until a uniform mix is achieved. The process of levigation is usually carried out using a mortar and pestle to reduce particle size as well as to disperse the powdered material into the concentrate. After levigation, the dispersion may either be transferred onto an ointment slab or mixed further using the mortar and pestle with the remainder of the ointment base. If the drug substance is soluble in solvents such as alcohol or water, the powdered materials are first dissolved in the solvent of choice and then incorporated with the remainder of the base using a mortar and pestle. Spatulation is inconvenient because the water- or alcohol-dissolved components are free-flowing.

When incorporating liquids or solutions of drugs into ointment bases, care must be taken to select the appropriate ointment base with regard to its capacity to accept the volume of drug in solution. To incorporate liquids into hydrophobic bases such as oleaginous bases,

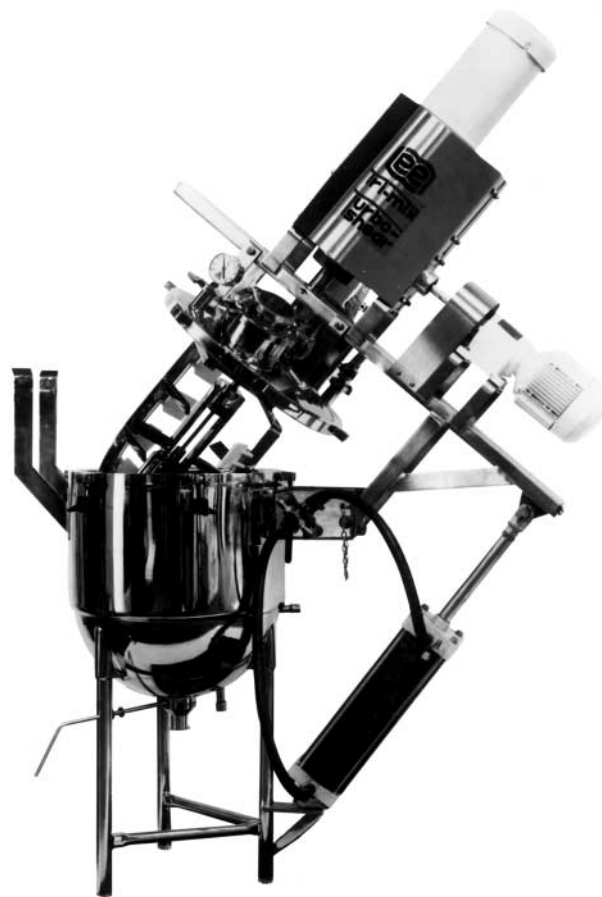


Fig. 1 Stainless steel jacketed kettle with agitator. (Courtesy of Lee Industries, Inc., Philipsburg, Pennsylvania.)

the liquid solution must first be incorporated in a minimum amount of hydrophilic base and subsequently added to the hydrophobic base. Care must be taken not to exceed the liquid-retaining capacity of the bases beyond which they become too soft or semiliquid in state.

Incorporation (industrial scale)

Mechanical mixers are used to prepare ointments in large quantities, especially in the Pharmaceutical industry. Stainless steel kettle mixers may be used to manufacture hydrocarbon and water-soluble base ointments (Fig. 1). The stainless steel kettle is jacketed for heating and cooling and is equipped with a mixing device. The kettle configuration is especially well-suited for the melting and mixing of oils and waxes and for complete bottom-emptying. Depending on the formulation to be processed, the mixer may be either a propeller or an anchor-agitator design. If the formulation is primarily liquid, a propeller

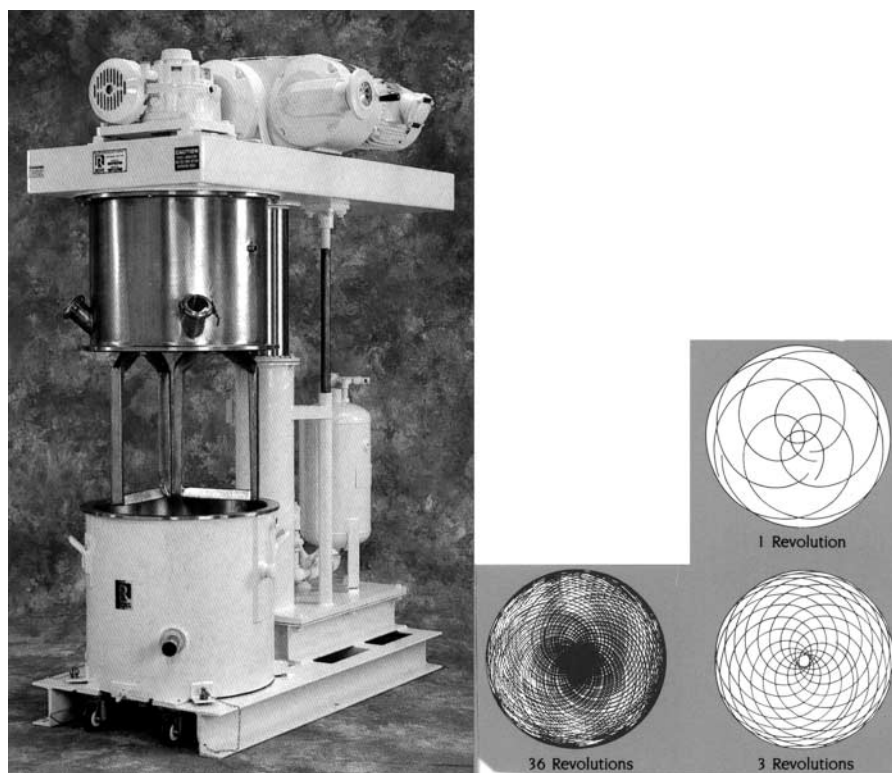


Fig. 2 Double planetary mixer showing mixing patterns at different stirring speeds. (Courtesy of Charles Ross & Son Co., Hauppauge, New York.)

mixer is more suitable. If the formulation consists of solid waxes, lanolin, petrolatum, and similar components, a kettle with a removable agitator, as shown in Fig. 1, is indicated.

Planetary mixers (Fig. 2) may also be used for large-scale preparations of the ointments. Here, the finely divided powdered materials are added slowly or sifted into the vehicle (ointment base) previously placed inside the rotating mixer. On achieving a uniform consistency, the ointment preparation may be processed through the roller mill to ensure complete dispersion and reduce the size of aggregates that may have formed during processing. The roller mills force the coarsely formed ointments through stainless steel rollers to produce ointments that are uniform in composition and smooth in texture.

Fusion Process

The process of fusion uses heat to melt all or some of the components of the ointment; the mixture is then allowed to cool with constant stirring until congealed. Additional components of the ointment preparation that were not

subject to the initial melting process are added to the congealing mixture as it is being cooled and stirred. Heat-labile substances are added to the preparation after careful observation that the temperature of the mixture is sufficiently low so as not to cause decomposition or volatilization of the components.

Substances may also be added by levigating a portion of the base with the milled drug to form a concentrate. The concentrate is then added to or dispersed in the remainder of the vehicle using a mixer such as porcelain dish or glass beaker (on a small scale) or a steam-jacketed vessel (on a large scale) and allowed to cool and congeal as noted earlier. Once congealed, the mixture is rubbed with a spatula or mortar and pestle (on a small scale) or allowed to pass through an ointment roller mill (on a large scale) to ensure a uniform and smooth texture. Some common examples of components that are subjected to the fusion process include beeswax, paraffin, stearyl alcohol, and high-molecular-weight polyethylene glycols (PEG). The order of addition of components in the process of fusion plays a role in the manufacture of ointments. In general, one way to mix all components together is to choose the component with

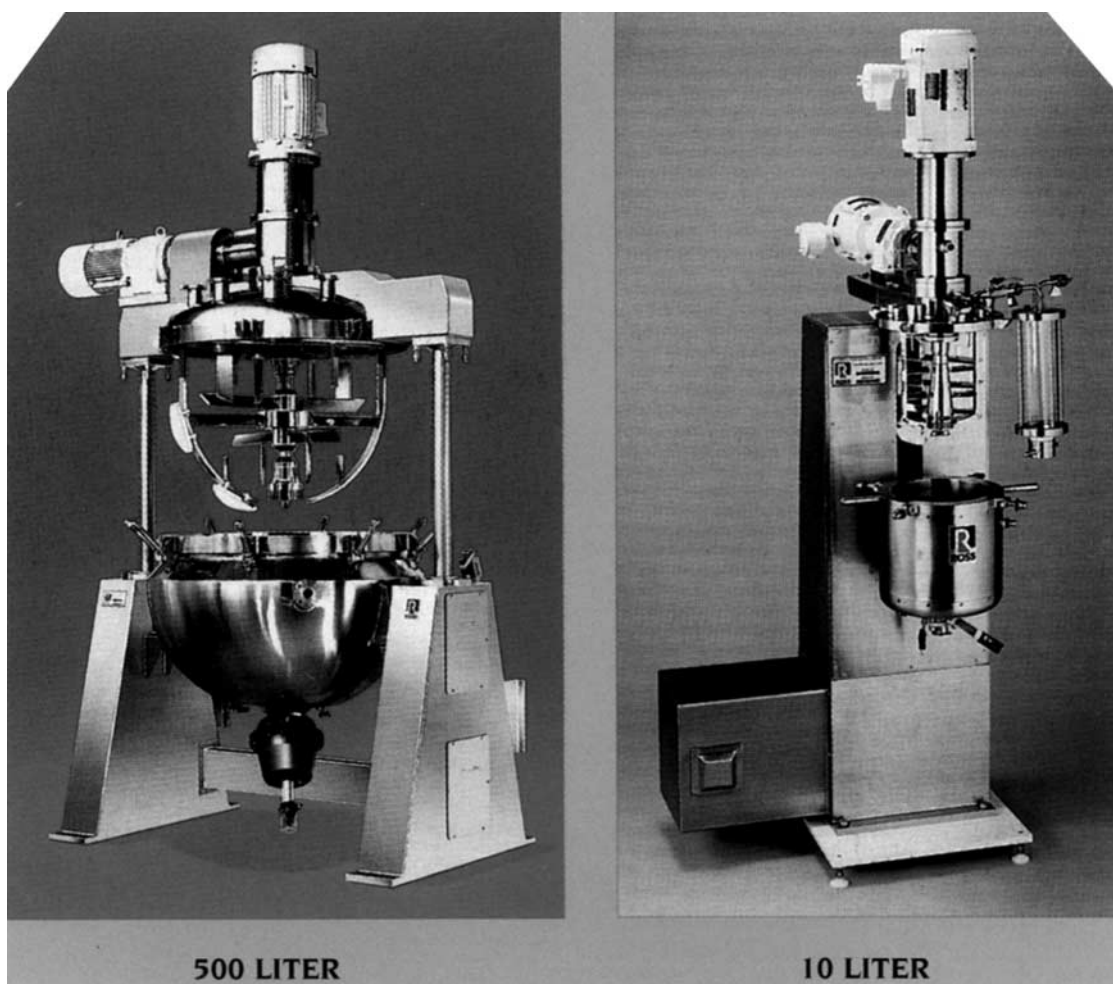


Fig. 3 High-turbulence mixers (turboemulsifiers) of varying capacities. (Courtesy of Charles Ross & Son Co., Hauppauge, New York.)

the highest melting point and use the minimum amount of heat to melt it. Subsequently, all the other components may be added with constant stirring and cooling of the melt until the mixture is congealed. Alternatively, the component with the lowest melting point may be added first. With increasing temperature, the remaining materials may be added to melt the entire mixture and then the process of cooling and congealing followed as before.

Emulsion Products

Emulsions are prepared by means of a two-phase heat system. In the first phase, the oil phase ingredients are combined in a jacketed tank and heated between 70 and 75°C to melt or liquefy all the ingredients to a uniform state. In a separate tank, the second phase, aqueous phase

ingredients are heated together to slightly above 75°C. The aqueous phase is then slowly added to the oil phase through constant agitation. The mixture is cooled slowly with continuing agitation as the emulsion is formed. The medicinal ingredients may be added when the emulsion is in the cooling stage. Usually, these ingredients are added as concentrated slurry that has been previously milled to a finely divided state. At times, when the aqueous phase of the emulsion system is larger, a large kettle may be required with a more complex mixer. As the oil phase is heated in a kettle, the aqueous phase is prepared in a kettle almost twice the size of the oil phase kettle. This size difference accommodates the volume of the final emulsion. Both phases may be heated between 70 and 80°C. For convenience and efficiency, the oil phase kettle should be mounted directly above the aqueous phase so as to introduce the oil phase mixture into the lower emulsifying

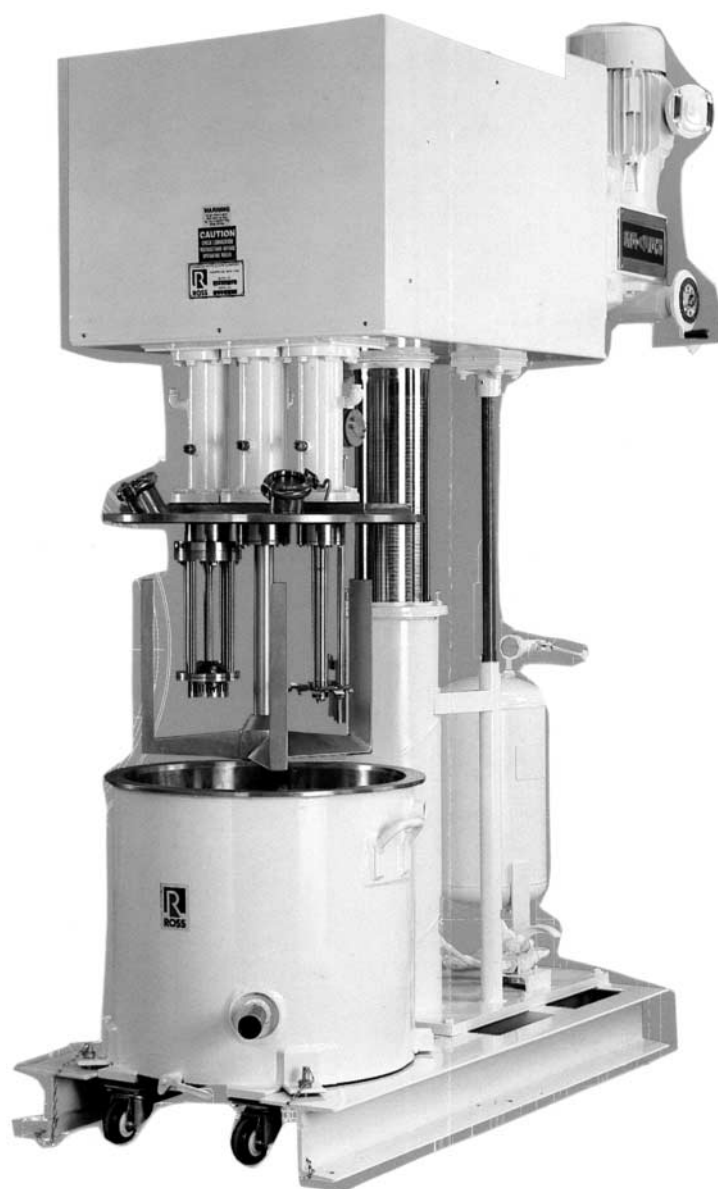


Fig. 4 Large-scale multiagitator mixer. (Courtesy of Charles Ross & Son Co., Hauppauge, New York.)

unit. The oil phase is introduced into the aqueous phase kettle in a relatively small stream to allow rapid dispersion and emulsification by an appropriately sized agitator. For capacities of 200 gallons or less, a single high-turbulence agitator may be used in most cases. Figure 3 shows two models of a high-turbulence emulsifier consisting of three top-entering coaxial agitators including an anchor-agitator and self-adjusting Teflon scrapers for thorough mixing of the preparation. For larger batches of up to 1000 gallons, a large-scale

multiagitator mixer may be used (as shown in Fig. 4) that involves a mixer-emulsifier (for high degree of shear), a high-speed disperser (to disperse solids into viscous liquids), and an anchor-agitator (to provide maximum movement of product under low shear conditions in the mix vessel). Another design for large batches uses a double-motion, counterrotating agitation combined with a homogenizing action that allows for versatility in mixing applications (Fig. 5). After the addition of the phases, the rate of cooling is generally slow to allow for adequate

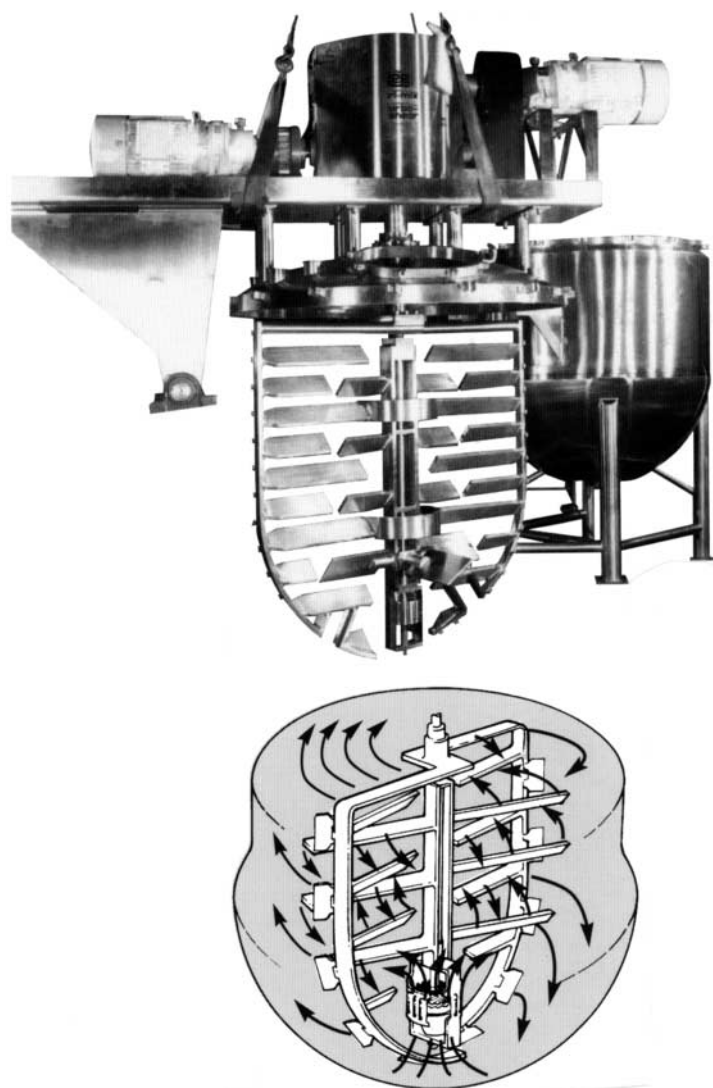


Fig. 5 Large-scale manufacturing unit (Tri-mix Turboshear) with counter-rotating mixing bars. (Courtesy of Lee Industries, Inc., Philipsburg, Pennsylvania.)

mixing while the emulsion is still in liquid form. Cooling should be at a rate consistent with the mixing of the emulsion and scraping of the kettle walls to prevent formation of congealed masses of ointment or cream, especially when the semisolid contains a large amount of high-melting substances. If drugs are introduced during the manufacture of the product, oil-soluble drugs should be dissolved in the oil phase and water-soluble drugs in the aqueous phase. All emulsions are susceptible to bacterial contamination, thus, preservative agents should be added while the emulsion is still hot to effect complete solution. To improve the stability of the oil phase, colloid mills may be used to disperse the oil phase further once the emulsion is formed. As shown in Fig. 6, a colloid mill

can be a portable unit that can be lowered into the kettle and operated as the mixing of phases continues. It operates by a shearing action of a high-speed rotor against a stationary stator with a clearance of a few thousandths of an inch. The colloid mill may also be located outside of the kettle and the emulsion pumped through it to the filling equipment or holding tank (Fig. 7). Sometimes, a homogenizer may be used to reduce the size of the oil globules by exerting a smearing action in which the emulsion is forced through a small orifice under high pressure. It is located between the emulsifying kettle and the filling line or holding tank. The advantage of the homogenizer over the colloid mill is that it does not incorporate air into the emulsion. However, the throughput



Fig. 6 Portable batch mixer on mobile hydraulic floor stand (Courtesy of Silverson Machines, Inc., East Long Meadow, Massachusetts).

of the homogenizer is 10–100 times lower than that of the colloid mill. When the emulsion is pumped from one vessel to another, an in-line colloid mill may be installed. Instead of producing an emulsion in a kettle or other vessel, several phases may be introduced into the pipeline in their proper proportions using metering pumps and emulsified en route by the action of the in-line colloid mill (Fig. 8).

Another in-line method of emulsification is the ultrasonic process, which uses a sonolator and the

principle of the Pohlman whistle (25). This process uses a very high-intensity mixing device that mechanically generates ultrasonic acoustic energy to produce emulsions and dispersions. Liquid to be processed is pumped through a special orifice, forming a flat, high-pressure stream. This jet impinges on the edge of a flat blade enclosed in a tube, causing it to vibrate at ultrasonic frequencies. Cavitation produces violent local pressure changes that act on the liquid, causing instantaneous and intense dispersion of any immiscible liquids or insoluble particles. Water-in-oil

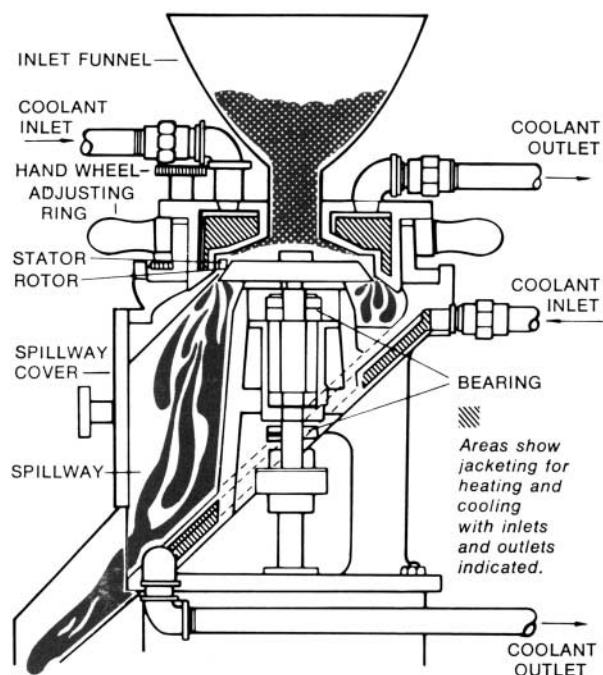
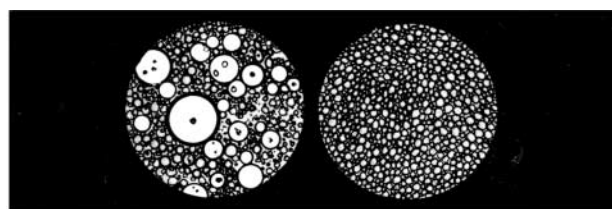


Fig. 7 Schematic representation of colloid mill and microphotographs showing emulsion globule size before and after milling. (Courtesy of Premier Mill Corp., Reading, Pennsylvania.)

emulsions are manufactured in the same equipment as o/w emulsions except that the phases are reversed in the equipment.

Miscellaneous Preparations

Heating, melting, and cooling in a single kettle in the same manner as with ointment preparations may be used to process the water-soluble bases composed of polyethylene glycols, ranging from liquids to waxy solids. They may be allowed to cool and solidify and to be stored until needed. Then they are remelted and packaged. Gels can be processed similarly in a single mixing kettle with slow agitation to avoid air entrapment, which may occur if mixed rapidly. This is especially important because all gels appear to be transparent or translucent, and air bubbles appear as a visible contamination.

PRESERVATIVES

Semisolid dosage forms must meet acceptable standards for microbial content, and preparations that are prone to microbial growth must contain antimicrobial preservatives. Antimicrobial preservative substances are included in ointment formulations to maintain the potency and integrity of product forms and to protect the health and safety of the consumer. The USP addresses this subject in its monograph *Microbiological Attributes of Non-Sterile Pharmaceutical Products*. An ideal preservative system should have attributes such as effectiveness at low concentrations against a host of microorganisms, solubility, nontoxicity, nonsensitizing, compatibility with other components, absence of odor, stability, and inexpensive. In addition, they must have the appropriate oil/water partition coefficient and be effective at the pH level of the products.

Several sources of contamination include water, raw materials, and poor sanitation conditions during manufacture, storage, and use. As such, preparations that contain water tend to support microbial growth to a greater extent than do preparations that are water-free. In the case of emulsion manufacture, the process of heating between 70 and 80°C usually kills most of the microorganisms. However, contamination may still occur as a result of unclean transfer lines and from improperly cleaned, sanitized, and protected filling equipment. It is not unusual to use a combination of preservatives to broaden their spectrum of activity because use of a single antimicrobial compound is generally insufficient for adequate preservation.



Fig. 8 Ultra-hygienic in-line mixer. (Courtesy of Silverson Machines, Inc., East Long Meadow, Massachusetts.)

Two categories of microorganisms are cause for concern in preservation of topical semisolid products. They are those liable to cause pathogenic symptoms and include staphylococci and hemolytic streptococci, *Pseudomonas aeruginosa* and *cepacia*, and *Escherichia coli*, and microorganisms liable to cause spoilage, which include water and airborne molds and yeasts. The following section describes the properties of some commonly used preservatives (26).

Parabens

The two most widely used agents are methylparaben and propylparaben. They are effective against molds and yeasts, but less effective against bacteria. They are more effective against Gram-positive than against Gram-negative bacteria. Parabens are most active at acidic pH levels less so in alkaline media. They are usually combined; an effective combination is 0.20% methylparaben and 0.05% propylparaben.

Dowcill 200 (Dow Chemical Co., Midland, MI) is a water-soluble, broad-spectrum antimicrobial and antifungal compound. It is not inactivated by nonionic, cationic, or anionic formulations, and it is particularly effective against *Pseudomonas*. Its activity is independent of pH; effective concentration is 0.02–0.30%.

Glydant (MDMH) (Lonza, Inc. Fairlawn, NJ): dimethylol-5,5,dimethylhydantoin is a water-soluble, highly active, broad-spectrum preservative. It functions over a wide range of temperatures and pH levels. It is noncorrosive, toxicologically acceptable, and biodegradable; effective concentration is 0.005–0.10%.

Germall 115 (Sutton Laboratories, Chatham, NJ): imidazolidinyl urea is a hygroscopic water-soluble white powder compatible with essentially all cosmetic ingredients including surfactants, proteins, and other special ingredients. Germall 115 acts synergistically with all other preservatives. It is effective against Gram-negative bacteria including *Pseudomonas*. Combined with parabens, it provides a broad spectrum of activity against yeasts and molds. It is recommended in the combination of Germall 115, 0.30%; methylparaben, 0.20%; and propylparaben, 0.01%.

Germall II (Sutton Laboratories, Chatham, NJ): diazolidinyl urea is a water-soluble, hygroscopic powder effective against Gram-positive and Gram-negative species including *Pseudomonas*. It is synergistic with other preservatives including the parabens. It is not inactivated by surfactants, proteins, or emulsifiers and is effective at all usual pH levels. It is recommended in the combination of Germall II, 0.20%; methylparaben, 0.20%; and propylparaben, 0.10%.

Germaben II-E (Sutton Laboratories, Chatham, NJ) is a clear liquid preservative system, readily soluble at a concentration of 1.0% in both w/o and o/w emulsions, but not in water alone. It has the composition of Germall II, 20%; methylparaben, 10%; propylparaben, 10%; and propylene glycol, 60%.

Suttocide A (Sutton Laboratories, Chatham, NJ) is a 50% aqueous solution of sodium hydroxymethylglycinate. It is a broad-spectrum antimicrobial preservative active against Gram-positive and Gram-negative bacteria and against yeast and mold. It remains active at alkaline pH levels.

LiquaPar Oil (Sutton Laboratories, Chatham, NJ) is a 100% clear, active, stable liquid blend of isopropyl, isobutyl, and *n*-Butyl esters of *p*-Hydroxybenzoic acid. This combination of parabens is a very effective preservative even at low concentrations against Gram-positive and Gram-negative bacteria, yeasts, and molds. The higher alkyl esters are more active and more stable and resistant to hydrolysis than are the lower alkyl esters; effective concentration is 0.1–0.4%.

Busan 1504 (CTFA) (Buckman Laboratories, Memphis, TN): dimethylhydroxy-methylpyrazole is a water-soluble, broad-spectrum bactericide and fungicide compatible with anionic, cationic, and nonionic ingredients. It is stable over a broad pH range and compatible with proteins; effective concentration is not given.

Ethylenediaminetetraacetic acid (EDTA) is a chelating agent that binds certain metals, especially iron and copper, which are essential to the nutrition of certain microorganisms. In this manner, it is a strong booster or enhancer of the activity of preservatives (27, 28) especially the parabens. Alone, it has the ability to increase the permeability of the bacterial cell wall and can kill *Pseudomonas aeruginosa* and *E. coli* by this activity; effective concentration is 0.05–0.10%.

Several factors may influence the success or failure of a preservative to protect a formulation against microbial contamination. These factors include the interaction of the preservative with surfactants, active substances, other components of the vehicle, sorption by the polymeric packaging materials, and product storage temperature. Although hundreds of chemicals can function as germicides, only a few substances have made it to the marketplace. The small list is not based as much on a compound's effectiveness as an antimicrobial agent as on the compound's safety and effectiveness in the final product.

The packaging of semisolid products, usually in jars and tubes, represents the best and worst conditions for microbial contamination. When using jars, the risk of contamination is higher every time a jar is opened or each

time fingers touch the product. For these reasons, semisolid products in jars require a highly effective, long-acting preservative system to remain active throughout the life of the product. With tubes, the risk of microbial contamination is significantly reduced because they expose only a very small area each time a cap is removed. Tubes also provide control of dose by an amount expressed with minimum exposure to the environment or human contact. Therefore, because products packaged in tubes are less likely to become contaminated, the use of preservative is also reduced. Thus, products packaged in tubes are less likely to cause skin irritation or sensitization.

PACKAGING

Topical dermatologic products are packaged in jars or tubes, whereas ophthalmic, nasal, vaginal, and rectal semisolid products are almost always packaged in tubes. Regardless of the container a semisolid product is eventually placed in, it must meet with the guidelines of the FDA (29) for drug products, as follows:

Containers, closures and other component parts of drug packages, to be suitable for their intended use, must not be reactive, additive or absorptive to the extent that the identity, strength, quality or purity of the drug will be affected. All drug product containers and closures must be approved by stability testing of the product in the final container in which it is marketed. The stability test includes testing filled containers at room temperature (e.g., 70°F) as well as under accelerated conditions (e.g., 105 and 120°F).

Because it is chemically inert, impermeable, strong, and rigid, glass has FDA clearance and is the ideal container for most drug products. With the proper closure system, it provides an excellent barrier to practically every element except light. Opaque jars are used to protect light-sensitive products and are available as porcelain white or dark green or amber in color. Commercially available empty ointment jars vary in size from 0.5 ounce to 1 pound.

Plastic containers are increasing in popularity over glass and metal (aluminum). This is because of qualities such as light weight, less breakage-prone, lower shipping costs, and the convenience of silk screen and plastic heat-transferred labeling. However, the disadvantage of plastics is the risk of permeation in two directions: from the product through the plastic from the inside out and from

the ambient environment through the plastic from the outside into the product.

Plastic containers used for emulsion systems must be thoroughly evaluated for physical and chemical changes in the emulsion as well as for physical changes in the container. Product-plastic interactions may be divided into five categories: permeation, leaching, sorption, chemical reaction, and alteration in the physical properties of the plastics or products (30, 31). For adequate protection of the product, container evaluation during the development stage of the product is imperative. The protocols and standard for tests to demonstrate resin equivalence are found in the USP/NF and include three categories of tests for chemical and spectral characteristics and moisture barrier (7). Some examples of plastics include high- or low-density polyethylene (HDPE or LDPE) or a blend of each, polypropylene (PP), polyethylene terephthalate (PET), and various plastic/foil/paper laminates, sometimes 10 layers thick (32). Each of these plastics has its own characteristics and advantages that make it conducive to packaging of semisolid products. For example, LDPE is soft and resilient and provides a good moisture barrier. HDPE provides a superior moisture barrier but is less resilient. PP has a high level of heat resistance, and PET offers transparency and a high degree of product chemical compatibility. Thus, plastics and plastic laminates are generally preferred over metal tubes for packaging of pharmaceutical and cosmetic products.

QUALITY CONTROL

The purpose of undertaking quality control measures is to ensure a quality product to the consumer. Quality control encompasses a broad range of responsibilities including component and final product testing as well as in-process testing and controls. Compliance with current Good Manufacturing Practices (cGMP) regulations, process validation, and good management of operations such as design, production, packaging, testing, and distribution are all integral parts in the total quality assurance effort. Quality control provides critical support in the testing and decision making required for process validation, and the activities and responsibilities of the quality control unit cover a broad range of testing and other activities (33). The goal of quality control is to ensure that different lots of a product are essentially the same by confirming that all lots meet specifications. The parameters used to determine the quality assurance of semisolid preparations include raw material specifications, in-process controls,

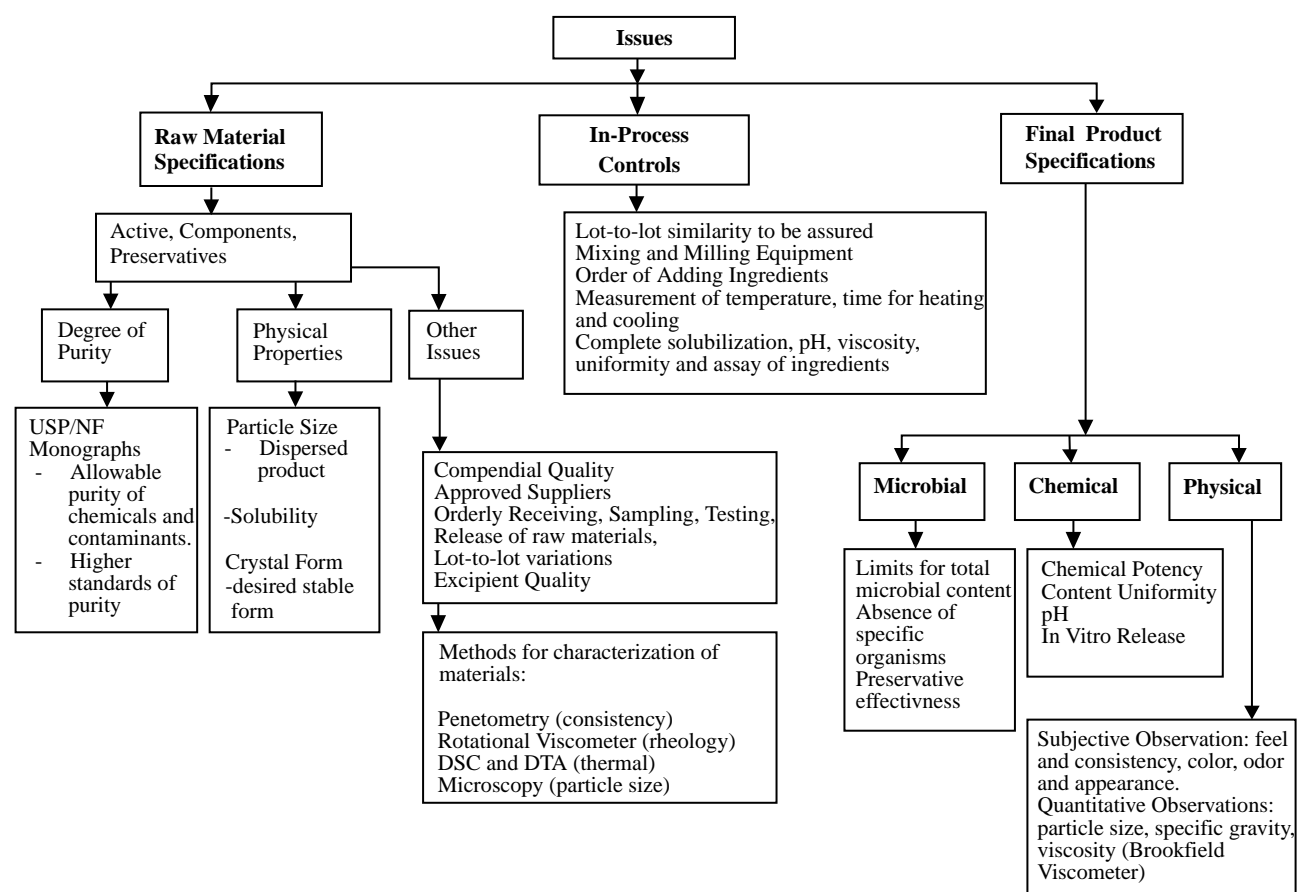


Fig. 9 Quality control of semisolid preparations.

and finished product specifications. The key parameter for any drug product is its efficacy as demonstrated in controlled clinical trials. However, the time and expense associated with such trials make them unsuitable as routine quality control methods. Therefore, *in vitro* surrogate tests are often used to ensure that product quality and performance are maintained over time and in the presence of change. As illustrated in the flow sheet (Fig. 9), a variety of physical and chemical tests commonly performed on products and their components (e.g., solubility, particle size and crystalline form of the active component, viscosity and homogeneity of the product) have historically provided reasonable evidence of consistent performance. More recently, *in vitro* release testing has shown promise as a means to comprehensively ensure consistent delivery of the active components from semisolid preparations (34). An *in vitro* release rate can reflect the combined effect of several physical and chemical parameters, including solubility and particle size of the active ingredient and rheological properties of the dosage form. In most cases, *in vitro* release rate is a useful

test to assess product sameness between prechange and postchange products. Recently, an *in vitro* test developed for a peptide containing gel formulation demonstrated that the release test can be modified to ensure product sameness after scale up and postapproval change (SUPAC) (35). However, with any test, the metrics and statistical approaches to documentation of sameness in quality attributes should be considered. Bioequivalence to previous lots may be considered as an additional parameter for some pharmaceutical products (36).

STABILITY

Stability testing is a routine procedure performed on drug substances and products. It is involved at various stages of product development. In early stages, accelerated stability testing (at relatively high temperatures and/or humidities) can be used to determine the types of degradation products that may be found after long-term

storage. Testing under more gentle conditions (those recommended for long-term shelf storage) and slightly elevated temperatures can be used to determine product shelf life and expiration dates.

According to the FDA draft guidelines to the industry (1), semisolid preparations should be evaluated for appearance, clarity, color, homogeneity, odor, pH, consistency, viscosity, particle size distribution (when feasible), assay, degradation products, preservative and antioxidant content (if present), microbial limits/sterility, and weight loss (when appropriate). Additionally, samples from production lots of approved products are retained for stability testing in case of product failure in the field. Retained samples can be tested along with returned samples to ascertain if the problem was manufacturing- or storage-related. Appropriate stability data should be provided for products supplied in closed-end tubes to support the maximum anticipated use period, during patient use, and after the seal is punctured allowing product content with the cap/cap liner. Ointments, pastes, gels, and creams in large containers including tubes should be assayed by sampling at the surface, top, middle, and bottom of the container. In addition, tubes should be sampled near the crimp.

For a product to be considered stable, it must retain the same properties and characteristics, within specific limits, that it possessed at the time of manufacture (7). A product must remain stable when it is stored for a period of time after the date of manufacture. When a product is subjected to stability testing, the objective is to determine whether the product has adequate shelf life under market and use conditions, when manufactured and packaged on a commercial scale. The FDA requires stability data on at least three lots of the same formulation and packaging—these data should appear on the product made by the commercial or any other equivalent process (in terms of critical parameters) to the proposed commercial process. To establish a stability profile, product stability storage tests must be carefully planned and initiated.

The commonly used predictive method of accelerated and stress testing depends on storage of products under conditions of rigidly controlled environment. Studies under accelerated conditions provide useful data for establishing expiration date and product stability information for future product development (e.g., preliminary assessment of proposed manufacturing changes such as change in formulation and scale up). These studies also assist in validation of analytical methods for the stability program and generate information that may help elucidate the degradation profile of the drug substance or product. Studies under stress conditions may be useful in

determining whether accidental exposures to conditions other than those proposed (e.g., during transportation) are deleterious to the product and also for evaluating which specific test parameters may be the best indicators of product stability.

A stability program typically involves aging the product under accelerated conditions of temperature, humidity, and light. Other types of accelerated or stress testing include centrifugation, shipping tests, and some form of product use test. The purpose of the product use test is primarily to evaluate preservative efficacy and ensure that the physical properties of the product remain stable under the conditions of use and are suitable for the intended application. Another useful test is to subject the product to low-intensity shaking for a period of hours or days. The key parameters of a stability program include the selection of specific accelerated storage conditions, how stability samples are stored at the test conditions (or stability stations), testing intervals, and testing the stability samples. A representative stability program may be illustrated as follows:

Samples: Each sample is one unit placed inverted in the storage chamber. An equal number of units are placed upright as control samples. The sample may also include units placed sideways. Additional samples should be placed at each stability station to make provision for samples for further investigations if this becomes necessary.

Stability Stations and Sampling Periods: Three 24-h freeze-thaw cycles are used: 1, 2, and 3 months at 50°C; 1, 3, 6, and 12 months at 40°C; and 1, 6, 12, 18, 24, and 36 months at 25°C (room temperature). At the end of each period, the sample is evaluated for chemical and physical properties as applicable (as noted earlier). The sample should also be examined to evaluate product-package compatibility. Microscopic examination of the sample can provide a very useful evaluation of changes in particle size and crystal structure. Tests for evaluating the compatibility of the product with the container include weight loss and moisture loss (if applicable). Reduction in the concentration of critical ingredients would indicate loss owing to binding with the container material or permeation through the container walls. The container should also be examined to evaluate interaction of the product with the container surface. The purpose of a stability program is to measure the changes observed in samples subjected to various stress conditions. The most difficult aspect of stability testing is to apply this information to determine the stability profile of the

product and to project an expected shelf life under market conditions. This is particularly difficult with semisolid preparations because the accelerated environmental and other stress tests used for stability testing are likely to cause substantial changes in critical properties of formulation ingredients. Higher storage temperatures can cause changes in the solubility and partitioning profile of drug and preservatives and alteration of the rheological profile of a product system. There may also be changes in the interfacial properties of ingredients that help stabilize emulsion systems. Therefore, results of stability testing can be misleading if the stress conditions used for stability evaluation do not reflect realistic product exposure conditions. It is generally recognized that application of stress testing to develop the stability profile of a product is useful. However, subjecting the product to unrealistic stress conditions is not useful for predicting shelf life under normal market and use conditions (37). Therefore, it is important to recognize the specific properties that are affected by the stress conditions of the stability test and whether the sample will return to normal after stress is removed. For this assessment, the pharmaceutical scientist must depend on the knowledge developed by appropriate studies to characterize the physical and chemical properties of the system.

REFERENCES

- Shah, V.P. FDA Guidance for Industry Document. *Nonsterile Semisolid Dosage Forms*; Center for Drug Evaluation and Research: Rockville, MD, 1997; 1–40.
- Flynn, G.L. Cutaneous and Transdermal Delivery: Processes and Systems of Delivery. *Modern Pharmaceutics*, 3rd Ed.; Banker, G.S., Rhodes, C.T., Eds.; Marcel Dekker, Inc.: New York, 1995; 272–298.
- Idson, B.; Lazarus, L. Semisolids. *Theory and Practice of Industrial Pharmacy*, 3rd Ed.; Williams and Williams: Baltimore, MD, 1986; 534–563.
- Block, L.H. Medicated Applications. *Remington: The Science and Practice of Pharmacy*, 17th Ed.; Gennaro, A.R. Ed.; Mack Publishing Co. Easton, PA, 1995; 1577–1597.
- Erdi, N.Z.; Cruz, M.M.; Battista, O.A. Rheological Characteristics of Polymeric Gels. *J. Colloid Interface Sci.* **1968**, 28, 36.
- Franks, A.J. Hydrocarbon Bases for Ointments and Creams. *Soap, Perfumes and Cosmetics* **1964**, 37, 221–319.
- Ophthalmic Ointments. In *United States Pharmacopoeia*, 23rd Ed.; The Pharmacopoeial Convention, Inc.: Rockville, MD, 1995; 1812.
- Lee, J.Y. Sterilization Control and Validation for Topical Ointments. *Pharm. Tech.* **1992**, 16, 104–110.
- Hameyer, P. Novel Emulsifiers for w/o Emulsion Preparations. *Household Profess. Prod. Indust.* **1993**, 4, 27–31.
- Information About Cosmetic Ingredients. *Dow Corning Bulletin*; Midland, MI, 1988.
- Surface Active Agents. *Croda Formulary Update*; Croda, Inc.: New Jersey, 1999; 11–15.
- In *Creanova Product Literature*; Degussa-Huls America, Inc.: Piscataway, NJ, 2000.
- Griffin, W.C. Hydrophilic-Lipophilic Balance. *J. Soc. Cosmet. Chem.* **1949**, 1, 311.
- International Cosmetic Ingredients Dictionary*; The Cosmetic, Toiletries and Fragrance Association: Washington, DC, 2000; 1250–1261.
- McCutcheons Surfactants and Emulsifiers*; McCutcheon Directories: Princeton, WI, 2000; 150–166.
- Paul, B.K.; Moulik, S.P. Microemulsions. Overview. *J. Disper. Sci. Tech.* **1997**, 18, 301–367.
- Lawrence, M.J. Surfactant Systems: Microemulsions and Vesicles as Vehicles for Delivery. *Eur. J. Drug Metab. Pharm.* **1994**, 3, 257–269.
- Bhargava, H.N.; Narurkab, A.; Lieb, L.M. Using Micro-Emulsions for Drug Delivery. *Pharm. Tech.* **1987**, 11, 46.
- Prince, L.M. Microemulsions. *Theory and Practice*; Academic Press: New York, 1977; 511–531.
- Friberg, S.E.; Venable, R. Microemulsions. *Encyclopedia of Emulsion Technology*; Becher, P., Ed.; Marcel Dekker, Inc.: New York, 1983; 1, 128–185.
- Juch, R.D.; Ruffi, Th.; Surber, C. Pastes: What Do They Contain? How Do They Work? *Dermatology* **1994**, 189, 373–377.
- Ansel, H.C.; Allen, L.V., Jr.; Popovich, N.G. Ointments, Creams and Gels. *Pharmaceutical Dosage Forms and Drug Delivery Systems*, 7th Ed.; Lippincott, Williams and Wilkins: Philadelphia, 1999; 244–262.
- Murdan, S.; Gregoriadis, G.; Florence, A.T. Novel Sorbitan Monostearate Organogels. *J. Pharm. Sci.* **1999**, 6, 608–614.
- Hermans, P.H. *Colloid Science*; Kruyt, H.R., Ed.; Elsevier: Amsterdam, 1969; 11, 483–651.
- Gennaro, A.R. *Remington's Pharmaceutical Sciences* 18th Ed.; Mack Publishing Co., 3rd Ed.; Easton PA, 1990; 1640.
- Bandelin, F.J.; Sheth, B.B. Semisolid Preparations. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1989; 31–61.
- Kabara, J.J. GRAS Antimicrobial Agents for Cosmetic Products. *J. Soc. Cos. Chem.* **1980**, 31, 1–10.
- Kabara, J.J. Structure-Function Relationships of Surfactants as Antimicrobial Agents. *J. Soc. Cos. Chem.* **1978**, 29, 733–741.
- FDA Good Manufacturing Practices Regulations*; Fed. Register, 1974; 9.
- Modern Plastics Encyclopedia*; McGraw-Hill Inc.: New York, 1992; 49.
- Varsona, J.; Gilbert, S.G. *Proceedings of the Plastic Packaging Institute*; Grosset and Dunlap: New York, 1991; 122–131.
- Forcinio, H. Tubes the Ideal Package for Semisolids. *Pharm. Tech.* **1998**, 22 (1), 32–36.
- Hanna, S.A. Quality Assurance. *Pharmaceutical Dosage Forms: Disperse Systems*; Lieberman, H.A., Rieger, M.M.,

- Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1989; 2, 631–680.
34. Shah, V.P.; Elkins, J.S.; Williams, R.L. Evaluation of the Test System Used for In Vitro Release of Drugs for Topical Dermatological Drug Products. *Pharm. Dev. Tech.* **1999**, *4* (3), 377–385.
35. Badkar, A.; Talluri, K.; Tenjarla, S.; Jaynes, J.; Banga, A.K. In Vitro Release Testing of a Peptide. *Gel. Pharm. Tech.* **2000**, *1*, 44–52.
36. Van Buskirk, G.A.; Shah, V.P.; Adair, D.; Arbit, H.M.; Dighe, S.V.; Fawzi, M.; Feldman, T.; Flynn, G.L.; Gonzalez, M.A.; Gray, V.A.; Guy, R.H.; Herd, A.K.; Hem, S.L.; Hoiberg, C.; Jerussi, R.; Kaplan, A.S.; Lesko, L.J.; Malinowski, H.J.; Meltzer, N.M.; Nedich, R.L.; Pearce, D.M.; Peck, G.; Rudman, A.; Savello, D.; Schwartz, J.B.; Schwartz, P.; Skelly, J.P.; Vanderlaan, R.K.; Wang, J.C.T.; Weiner, N.; Winkel, D.R.; Zatz, J.L. Workshop III Report: Scale Up of Liquid and Semisolid Disperse Systems. *Pharm. Res.* **1994**, *11* (8), 1216–1220.
37. Rieger, M.M. Skin Care: New Concepts vs. Established Practices. *Cosm. Toiletr.* **1991**, *106*, 59–69.

SOLUBILIZATION OF DRUGS IN AQUEOUS MEDIA

Paul B. Myrdal
Samuel H. Yalkowsky

The University of Arizona, Tucson, Arizona, U.S.A.

INTRODUCTION

Therapeutic drugs are often given systemically. Once given systemically, a drug will distribute throughout the body. By distributing in the body, the drug is essentially diluted out from its original concentration in the formulation/dosage form. Hence, the formulation is really a drug concentrate. For solid dosage forms, the dose to be delivered is not normally a physical problem. However, dose can become a significant formulation challenge for parenteral preparations, due to limitations in aqueous solubility and volume. Therefore, in order to obtain a solution formulation for drugs with poor solubility, it is necessary to alter the formulation to facilitate solubilization.

The choice of solubilization method will depend upon how efficiently the drug can be solubilized, stability in the system, and upon the biocompatibility of the vehicle for a given delivery route. For solid dosage forms, it may be possible to alter the solid phase to enhance dissolution. For parenterals, the four most commonly used techniques for solubilization are: pH adjustment; cosolvent addition; micelle inclusion through surfactant addition and complexation. The following chapter is designed to summarize the theoretical as well as practical use of each of the above techniques. More extensive discussion on techniques for drug solubilization can be found in books dedicated to the subject (1, 2).

SOLUBILITY

Before attempting random laboratory experimentation, it is prudent to understand what physicochemical properties are making a given drug poorly soluble. There are two key components that govern the solubility of an organic solute in water, namely the crystal structure (melting point and enthalpy of fusion) and the molecular structure (activity coefficient). The aqueous solubility, X_w , of an organic nonelectrolyte is simply described by the addition of these two terms, i.e.,

$$\log X_w = \log X_i^c - \log \gamma^w \quad (1)$$

where X_i^c is the ideal solubility and γ^w is the activity coefficient of the compound in water.

Ideal Solubility

The ideal solubility pertains to the effect of the crystalline structure on solubility. A solute molecule must first dissociate from this crystal lattice before it can go into solution. This dissociation from the crystalline lattice is accompanied by a free energy change. The more energy it takes to free a solute molecule from its crystal (i.e., the higher the melting point), the lower the solubility.

It is possible to quantitate the molar free energy necessary to produce a hypothetical supercooled liquid at a given temperature T , through the use of an enthalpy-temperature thermodynamic cycle (3). Rigorously, it is necessary to sum the corresponding enthalpy and entropy changes that it takes to heat the crystal to the melting point, melt the crystal, then cool the liquid back to the reference temperature. For this discussion it is convenient to omit any change in enthalpy with temperature (heat capacity). This allows for the simplified mathematical expression

$$\log X_i^c = \frac{\Delta S_m(T_m - T)}{2.303RT} \quad (2)$$

where X_i^c is the ideal solubility (mole fraction), R is the universal gas constant (1.98 cal/deg mol), ΔS_m is the entropy of melting and T_m is the melting point of the compound in Kelvin.

If the entropy of melting, ΔS_m , is not known experimentally, an approximation for rigid organic molecules of 13.5 cal/degmol can be made (4). If the temperature of interest is 298 K then Eq. 2 becomes

$$\log X_i^c = -0.01(T_m - 298) \quad (3)$$

or

$$\log X_i^c = -0.01(MP - 25) \quad (4)$$

where MP is the melting point of the compound in Celsius. If the melting point is less than 25°C (i.e., liquid at room temperature) then there is no crystal limitation on solubility and Eq. 4 is equal to zero.

For those compounds that are crystalline, it is important to emphasize that the ideal solubility is strictly a function of the pure crystal, and as a result, is solvent independent. Eq. 4 also assumes that the solvent does not effect the solid phase in any way to change the free energy.

Table 1 demonstrates the effect of melting point on solubility as described by Eq. 4. It can be seen that for every 100 degree difference in melting point, there is a corresponding change in solubility of 10 times. Thus, for a hypothetical compound that melts at 325°C, the crystal structure would be responsible for approximately a 3-fold (1000 times) decrease in solubility, relative to it as a liquid at room temperature.

Activity Coefficient

The aqueous activity coefficient of a compound describes the effect of molecular structure on aqueous solubility. If a compound mixes with water and forms an ideal solution, then the activity coefficient would be taken as unity and the last term in Eq. 1 would become zero. In such a case, then the solid phase, if any, would be the sole physical property inhibiting solubility. Most drugs are relatively nonpolar and do not form ideal solutions with water. Therefore in order to understand the extent by which the inherent molecular structure is limiting solubility, it is helpful to obtain an estimate for the aqueous activity coefficient.

Numerous methods have been developed to estimate the aqueous activity coefficient (3, 5–13). A convenient method that has been successfully applied within the pharmaceutical industry utilizes the octanol/water partition coefficient. Yalkowsky and Valvani (14) have shown that the molar activity coefficient can be directly related to the octanol/water partition coefficient, $K_{o/w}$, by the following molar relationship;

$$\log \gamma^w = -\log K_{o/w} + 0.80 \quad (5)$$

Table 1 Comparison of different hypothetical melting points and the resulting effect on solubility as described by Eq. 4

Hypothetical melting point (°C)	Log X_i^c (Eq. 4)	Decrease in solubility
0	0	No effect (reference)
25	0	No effect (reference)
125	-1	10×
225	-2	100×
325	-3	1000×

There are numerous methods by which to estimate octanol/water partition coefficients. One of the most widely recognized is the group contribution method CLOGP (15).

Equation (5) enables a convenient tool for assessing the effect of structure on aqueous solubility. Table 2 demonstrates the decrease in solubility as partition coefficient increases. As $K_{o/w}$ increases by a factor of 10 ($\log K_{o/w} = 1$), the decrease in solubility is also a factor of 10. Thus, a compound with a partition coefficient of 1000 ($\log K_{o/w} = 3$) will have a solubility that is approximately 3-fold (1000 times) less than a reference compound that has a $K_{o/w}$ of 1 ($\log K_{o/w} = 0$).

Estimating Solubility

With the knowledge or estimate of the octanol/water partition coefficient and melting point, solubility can be estimated, by incorporating Eqs. 4 and 5 into Eq. 1. This gives

$$\log S_{\text{est}} = -0.01(\text{MP} - 25) - \log K_{o/w} + 0.80 \quad (6)$$

where S_{est} is the estimated molar solubility, MP is the melting point (°C) and $K_{o/w}$ is the octanol/water partition coefficient. For drugs that are liquid at room temperature, the melting point term is zero, giving simply

$$\log S_{\text{est}} = -\log K_{o/w} + 0.80 \quad (7)$$

While Eqs. 6 and 7 are simplified equations for estimating aqueous solubility, they can also be used to facilitate the understanding of why a drug is poorly soluble. The equations yield quantitative descriptions of which physicochemical properties are limiting solubility. If a compound has a low, or no melting point at room temperature, however and a high $\log K_{o/w}$, then molecular structure is limiting solubility and the aqueous media must be modified to facilitate solubilization. If the melting point is very high, and the $\log K_{o/w}$ is low, then modification of the aqueous media may not significantly increase solubility and manipulation of the solid phase may be necessary. If a compound has a high melting point and a high $\log K_{o/w}$, it can be appreciated that a formulator will have a significant challenge in drug solubilization.

Table 2 Comparison of different hypothetical $\log K_{o/w}$ s and the resulting effect on solubility as described by Eq. 5

$K_{o/w}$	$\log \gamma^w$ [Eq. 5]	Decrease in solubility
1	-0.8	Reference
10	-1.8	10×
100	-2.8	100×
1000	-3.8	1000×

SOLID PHASE

As discussed above, the solid phase for a solute molecule is a fundamental component of a compound's solubility. Consequently, understanding and characterizing the solid phase of a drug, in a given system, is of vital importance from a formulation standpoint. A comprehensive review on the theoretical and practical aspects of solid state chemistry and the significance to the pharmaceutical industry has been given by Byrn et al. (16).

It is important to always consider, that for any raw drug substance or formulation, only one solid phase is thermodynamically stable for a given set of environmental conditions. The most stable solid will have the lowest free energy and correspondingly the lowest solubility. The unknowing use of a metastable crystal can lead to formulation stability problems with respect to solubility. In light of this, it would seem as though little can be done in regards to increasing solubility through the use of crystal modification. It is clear that the most stable crystal is the most prudent choice for suspensions. The kinetics for solvent mediated transformation are generally too rapid or too uncontrollable for a stable product. However for solid dosage forms, it may be possible to utilize a less stable solid phase due to the decreased molecular mobility in the dry state. The key determinant for success will be if the metastable form has an intrinsically high energy barrier to reversion and is well defined from a kinetic standpoint. To facilitate stability, it may be necessary to add excipients or impose packaging constraints to eliminate or slow down conversion.

Apparent Solubility and Dissolution Rate

The crystal form with the higher free energy may exhibit an apparent solubility that is higher than the true equilibrium solubility for the system. An apparent solubility increase can occur anytime the starting solid material is not the most stable for the given system. However, a metastable crystal will produce only a transient increase in solubility. The most stable crystal will eventually precipitate and the apparent solubility gain will diminish until the thermodynamic equilibrium solubility is reached. The degree of supersaturation and duration will depend on the characteristics of the starting material and on the nucleation rate and growth kinetics of the stable form. Consequently, an inherent difficulty in working with metastable systems is that the kinetics of conversion often cannot be predicted or controlled.

By virtue of having a higher apparent solubility, a metastable crystal will have an increased dissolution rate

over the more stable form. The change in mass, M , as a function of time, t , for a solute is directly proportional to its apparent solubility, S_{app} , and is given by

$$dM/dt = K \times A(S_{app} - C) \quad (8)$$

where A is the solvent accessible surface area, C is the concentration of the solute in solution, and K is a constant that includes the diffusion coefficient of the solute and other hydrodynamic properties of the system (17). Hence, the larger the apparent solubility of a metastable form, the greater the dissolution rate that it leads to. Hamlin et al. (18) have illustrated the dependency of dissolution rate upon solubility for a large number of pharmaceutical compounds. Yoshihashi et al. (19) have correlated initial dissolution rates for terfenadine based on heats of fusion of the starting material. In addition, since the crystal contribution to solubility is independent of the solvent, this general relationship is applicable to virtually any solvent system. Nicklasson and Brodin (20) have successfully correlated dissolution rate with solubility for ethanol cosolvent systems.

Surface area is also directly proportional to the dissolution rate of a solute. Particle size reduction is another common and often efficient means by which to achieve higher levels of drug in solution at earlier time points (21–28). As particle size decreases, the surface area per unit volume of solute increases and consequently more drug is exposed to the solvent. Also, as particle size decreases the surface molecules are of higher free energy which increases dissolution. And finally, the processing of solid material can often lead to crystal defects within a particle or surface area where crystallinity is lost (amorphous), both of which can increase the apparent solubility. Mosharraf et al. have demonstrated the effect of crystal structure disorder on solubility and dissolution rate (29).

The significance of solubility and dissolution is one to be highlighted. Insufficient oral bioavailability can be the result of a compound's low intestinal solubility and/or a slow dissolution rate. It may be practical at times to use a solid material that gives a higher apparent solubility and/or an increase in dissolution rate in order to enhance bioavailability. Aguiar et al. (30, 31) were able to show that polymorph B of chloramphenicol palmitate, which can produce solubility that is roughly two times that of polymorph A, gives greater blood levels in vivo. They were able to show that as the percentage of polymorph B increases in the dosage form, there is a linear increase in blood concentration of chloramphenicol palmitate. Several other investigators (32–34) have also found that with the proper choice of the solid-state form,

biological activity can be enhanced due to increased solubility or dissolution rate.

Apparent Solubility Enhancement from Different Solid Phases

In order to assess the relative increase in solubility of a metastable solid phase with respect to another, a simple solubility ratio can be defined. Here the solubility ratio is

defined as the value for the higher solubility phase divided by the lower. Table 3 contains a selected list of solubility ratios for different solid phases of a selected list of example drugs.

From Table 3, it can be seen that a solubility ratio greater than one can be obtained, however there are cases in which there is no apparent increase in solubility between phases. For the data given, the transient solubility increase for solvates is similar to that of polymorphs, having solubility ratios that often range between 1 and 3.

Table 3 Examples of different drugs and their observed solubility ratios for different solid phases^a

Drug (temp. in °C)	More soluble phase/less soluble phase	Solubility ratio	Reference
Acetohexamide (37)	II/I	1.2	35
Benzoxoprofen (25)	I/II	1.5	36
Glibenclamide (37)	II/I	1.6	37
Mebendazole	C/A	3.6	38
Mebendazole	B/A	7.4	38
Meprobamate (25)	II/I	1.9	39
Oxyclozamide (25)	II/I	2.6	40
Oxyclozamide (25)	III/I	3.9	40
Ampicillin (not cited)	Anhydrate/trihydrate	1.3	41
Ampicillin (20)	Anhydrate/hydrate	2.2	42
Ampicillin (30)	Anhydrate/hydrate	1.5	42
Calcium gluceptate (37)	Anhydrate/3.5 hydrate	18.6	43
Erythromycin (30)	Anhydrate/dihydrate	2.2	44
Lamivudine (25)	Anhydrate/0.2 hydrate	1.2	45
Paroxetine HCl (20)	Anhydrate/hemihydrate	1.7	46
Phenobarbital (20)	Anhydrate/hydrate	1.1	47
Phenobarbital (25)	Anhydrate/hydrate	1.4	47
Phenobarbital (35)	Anhydrate/hydrate	1.0	47
Piroxicam (25)	Anhydrate/monohydrate	1.0	48
Sulfamethoxazole (25)	Anhydrate/monohydrate	1.2	49
Theophylline (25)	Anhydrate/monohydrate	2.0	50
Uric acid (25)	Hydrate/anhydrate	1.6	51
Uric acid (37)	Hydrate/anhydrate	1.9	51
DMHP (25)	Formate solvate/anhydrate	8.2	52
Furosemide (37)	Dioxane solvate/anhydrate	1.3	53
Furosemide (37)	DMF solvate/anhydrate	1.2	53
Glibenclamide (37)	Pentanol solvate/anhydrate	31.8	37
Glibenclamide (37)	Toluene solvate/anhydrate	2.4	37
Caffeine (25)	Amorphous/crystalline	6.5	54
Diacetylmorphine (25)	Amorphous/crystalline	16	54
Theophylline (17)	Amorphous/crystalline	58	54
Theobromine (16)	Amorphous/crystalline	50	54
Morphine (20)	Amorphous/crystalline	268	54
Hydrochlorothiazide (37)	Amorphous/crystalline	1.1	55
Bendrofluazide (37)	Amorphous/crystalline	2.8	55
Cyclothiazide (37)	Amorphous/crystalline	6.2	55
Cyclopenthiazide (37)	Amorphous/crystalline	8.3	55
Polythiazide (37)	Amorphous/crystalline	9.8	55

^aThe different solid phases used as well as the temperature of the experiment study are also given.

While most compounds produce increases that are less than a factor of ten, calcium gluceptate and glibenclamide exhibited significant increases of nearly 20 and 30, respectively (37, 43).

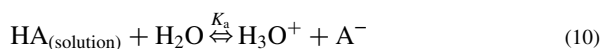
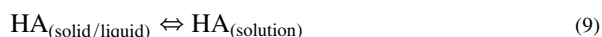
Consideration of Eq. 4, would suggest that the largest gain in solubility should be realized for those solid fractions that are actually not crystalline, i.e., amorphous. For a purely amorphous material, there would be no crystal contribution to the solubility and the compound could be treated as a liquid. Since it is difficult to produce a purely amorphous material, there is often some degree of crystallinity to a "amorphous" material. Nonetheless, amorphous materials clearly have the highest free energy, relative to any crystalline solid fraction. As a result, amorphous materials can potentially give some of the largest solubility ratios. Toffoli et al. (54) and Corrigan et al. (55) have investigated the use of amorphous materials and have found a wide range of apparent solubility ratios. Some of their data is listed in Table 2, and clearly shows that the use of an amorphous material has the greatest potential for solubility enhancement. Other investigators have also observed significant solubility differences with amorphous compounds (56–58). Some investigators have seen practical in vivo benefits by using amorphous materials (59, 60). However, due to the high free energy of these systems their use is often not practical. Giron (61) has noted a few compounds that have relatively stable amorphous forms.

Through alteration of the solid-state form it is possible to increase the dissolution rate, or apparent solubility of a drug. These increases can potentially increase the bioavailability of a poorly water soluble drug. However, the practical use of a higher energy solid form is limited due to physical and chemical stability issues. Significant investigation must be made in order to assure that a dosage form using a metastable crystal will maintain integrity throughout the product life.

pH CONTROL

Solid-state manipulation is often not advantageous or practical, as a result, it is then necessary to alter the solvent. For organic solutes that are ionizable, changing the pH of the system may be the simplest and most effective means of increasing aqueous solubility.

For a weak monoprotic acidic drug, $HA_{(solid/liquid)}$, in water:



where $HA_{(solution)}$ represents the free acid (unionized form) in solution and A^- is the ionized acid in solution. The total concentration of drug in solution (S_T) is equal to $[HA_{(solution)}] + [A^-]$, where $HA_{(solution)}$ is the intrinsic solubility of the drug (S_w). From Eq. 10, the ionized concentration of drug is

$$[A^-] = \frac{K_a[HA]}{[H^+]} \quad (11)$$

giving

$$S_T = S_w + \frac{K_a[HA]}{[H^+]} \quad (12)$$

Since $[HA] = S_w$, $\log K_a = pK_a$ and $\log H^+ = pH$, Eq. 12 can be simplified to,

$$S_T = S_w(1 + 10^{(pH-pK_a)}) \quad (13)$$

which relates the total solubility to the intrinsic solubility and pK_a of the weak monoprotic acid and the pH of the system.

For a weak base in an aqueous solution,



By analogy to the weak acid the total solubility is described by

$$S_T = S_w(1 + 10^{(pK_a-pH)}) \quad (16)$$

where the pK_a refers to HA^+ . Fig. 4 illustrates the general pH-solubility profiles for three solutes having the same intrinsic solubility but different pK_a s.

Fig. 1 illustrates that solubilization by ionization can be very efficient for a mono acidic or basic drug. The linear solubilization slope corresponds to a 10-fold increase in solubility for a one-unit change in pH. Zwitterionic compounds, which have both acidic and basic functional groups, will have a pH solubility plot that will be a combination of those given in the top and bottom portions of Fig. 1. These compounds have a minimum solubility at a pH that is equal to the average of the pK_a s (62–65). The solubilization of divalent acids and bases is similar to their monoprotic counterparts. However, upon the ionization of the second functional group, the solubilization slope is 2 instead of 1, which corresponds to a 100-fold increase in solubility for a one-unit change in pH. A complete mathematical description for the solubilization of a dibasic compound has been given by Garren and Pyter (66).

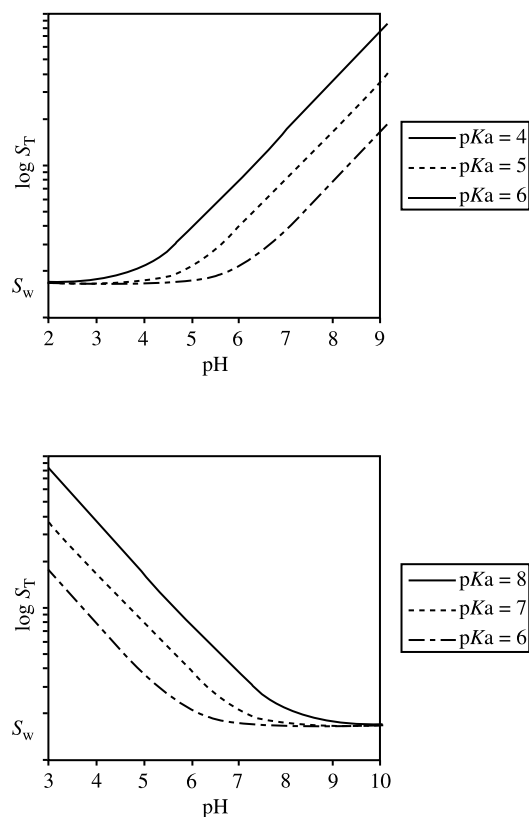
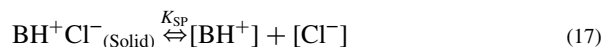


Fig. 1 (Top) Log solubility, $\log S_T$ vs. pH for typical weak acids having pK_a s of 4, 5, and 6 and the same intrinsic solubility, $\log S_w$. (Bottom) Log solubility, $\log S_T$, vs. pH for typical weak bases (HA^+) having pK_a s of 8, 7, and 6 and the same intrinsic solubility.

Salt Formation

Eqs. 13 and 16 assume that the ionized species of a solute has infinite solubility. However, an ionized solute can form salts with appropriate counterions. The formation of a salt is governed by the solubility product, K_{sp} , of the salt complex. For example, the thermodynamic equilibrium for a chloride salt, BH^+Cl^- , is



where $BH^+Cl^-_{(Solid)}$ represents the solid chloride salt, BH^+ is the ionized base and Cl^- is the chloride counterion. As a result, the concentration of the ionized species will be limited by the solubility of the salt. Some organic salts are very soluble in aqueous systems, others are not and can significantly limit the solubility of a solute. Chowhan (67) demonstrated the effect of four different salts on the pH-solubility profile of naproxen. As shown in Fig. 2, all four salts behave identically between the pH values of 1 and 6, and can be described by Eq. 13. However, depending on

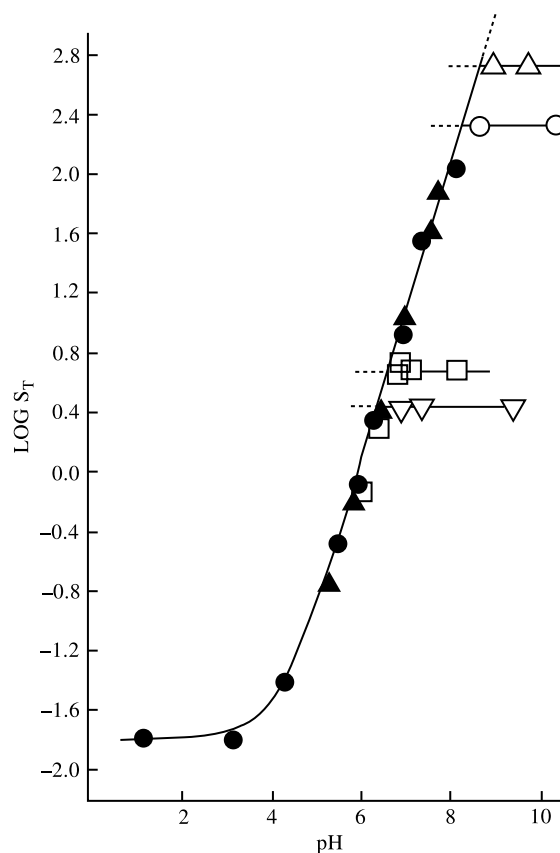


Fig. 2 Solubility-pH profiles, S_T of naproxen and four salts: Δ , potassium salt; \circ , sodium salt; \square , magnesium salt; and ∇ , calcium salt. (From Ref. 67.)

the given salt, the solubility of naproxen is limited at pH values greater than 6.5. The solubilities of the corresponding salts are indicated by plateaus, with the potassium salt having nearly 200 times the solubility of then calcium salt. Hence, at a given plateau the salt is the solid fraction that is in equilibrium with the system, which once again emphasizes the importance of characterizing the solid phase in order to understand what factors are influencing solubility.

Kramer and Flynn (68) have also previously investigated the effect of salt formation on solubility. A graph from their work is given in Fig. 3 and shows the pH-solubility profile for an organic hydrochloride. As can be seen from the data, there is an exponential increase in solubility as would be expected from Eq. 16. The solubility diverges from the expected line and roughly plateaus at pH levels under about 7.2. As pH decreases further, so does solubility, due to the disappearance of the free base from solution after the hydrochloride is formed. Kramer and Flynn illustrated how Eq. 16 can be modified to take into account salt

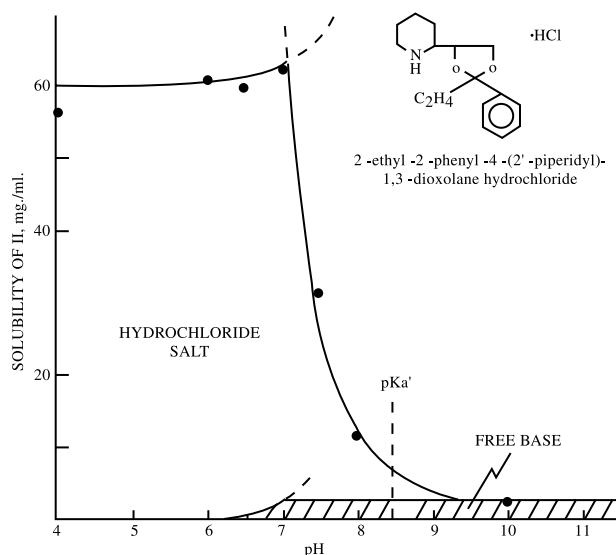


Fig. 3 Solubility–pH profile of a substituted piperidyl-dioxolane (II) in 0.05 *M* succinate buffer at 30°C. Lines represent theoretical curves. (From Ref. 68.)

formation. Additional mathematical expressions have been developed more recently to characterize pH solubility profiles with the appreciation of counterion affects (69–71).

In addition to overall solubility, salts can have a significant impact on the dissolution rate of a solute. Hence, solubility considerations are important factors to be assessed when trying to make a rational choice of an appropriate pharmaceutical salt (72–77).

Buffers

One practical use of a buffer is to simply maintain the pH of the system over time. For pH solubilized drugs, another practical use of a buffer is to reduce or eliminate the potential for precipitation of the drug upon dilution. It has been shown that drug precipitation upon intravenous injection has been linked to phlebitis (78–82). As shown through the previous discussion, the solubility of a drug can be increased exponentially with pH alteration. However, it also then follows that the solubility can decrease exponentially with pH alteration. If a pH solubilized formulation is diluted with a medium by one half, then the drug concentration will decrease by one half. At the same time the pH of the new mixture may change. If it changes by one pH unit, in a direction that decreases ionization of the drug, then the solubility of that drug will decrease 10-fold. The drug can precipitate if the concentration in the solution exceeds the new solubility.

The degree and extent of precipitation will depend on the ability of a formulation to resist pH change when diluted. The pH change on dilution of one solution by another will depend on the initial pHs and buffer capacities of both solutions. Surakitbanharn et al. (83) illustrated the affect of initial pH and phosphate buffer concentration on pH change when diluted with Sorensen's Phosphate Buffer. Myrdal et al. (82) used the computational model of Surakitbanharn et al. as a means of selecting buffer concentration and initial pH to eliminate the precipitation potential of a weakly basic drug, levemopamil–HCl. Myrdal et al. showed that the unbuffered formulation precipitated upon dilution and resulted in phlebitis *in vivo*, whereas the buffered formulations did not precipitate and did not elicit any phlebitis *in vivo*.

Solubilization through the Use of pH Control

Under the proper conditions, the solubility of an ionizable drug can increase exponentially by adjusting the pH of the solution. From a structural point of view, a drug that can be efficiently solubilized by pH control should be either a weak acid with a low pK_a or a weak base with a high pK_a . Note that the effect of pH on solubilization is independent of the value of the solubility of the unionized form of the drug. In other words the solubility of an unionized acid with a pK_a of 4.2 and a solubility of 1.0 mg/ml and one with the same pK_a and a solubility of 0.001 mg/ml will both be increased by a factor of 1001 at pH 7.4.

Although the use of a buffer will aid in reducing the risk of precipitation upon dilution, there are pH, concentration and buffer type limitations. Physiological compatibility will depend on factors such as the route of administration, tonicity, and contact time. Other factors such as chemical stability as a function of pH must also be considered. Table 4 lists some marketed parenteral products and associated buffers, concentrations and pH values.

COSOLVENTS

A common and effective way by which to increase the solubility of a non polar drug is through the use of cosolvents. A cosolvent system is one in which a water miscible or partially miscible organic solvent is mixed with water to form a modified aqueous solution. Cosolvents have some degree of hydrogen bond donating and or hydrogen bond accepting ability as well as small hydrocarbon regions. The resulting solution will have physical properties that are intermediate to that of the pure

Table 4 Selected list of marketed products that contain buffers

Drug (product)	Route of administration	Buffer, concentration (% w/v) and pH
Methohexital sodium	IV/IV infusion reconstituted	Sodium carbonate pH 9–11
Chlorpromazine–HCl (Thorazine)	IM/IV after dilution	Ascorbic acid 0.2% pH 3.0–5.0
Amikacin sulfate	IM/IV infusion after dilution	Sodium citrate 2.8% pH 3.5–5.5
Biperidan lactate (Akineton)	IV/IM	Sodium lactate 1.4%
Thiopental sodium (Pentothal sodium)	IV infusion reconstituted	Sodium carbonate pH 10–11
Epinephrine–HCL (SusPhrine)	SC	Ascorbic acid 1.0%
Etidocaine HCl (Duranest)	Infiltration	Citric acid pH 3–5
Etoposide (VePesid)	IV infusion	Citric acid 0.2% pH 3–4
Methoxamine HCl (Vasoxyl)	IV/IM	Citric acid 0.3% pH 3–5
Methyldopate HCl (Aldomet ester HCl)	IV infusion after dilution	Sodium citrate 0.3% Citric acid 0.5% pH 3.5–4.2
Nalbuphine HCl (Nubain)	IV/IM/SC	“Citrates” 2% pH 3.5
Perphenazine (Trilafon)	IM/IV/SC	Citric acid pH 4–5.5
Chlordiazepoxide–HCL (Librium)	IM/IV	Maleic acid 1.6% pH 3
Topotecan (Hycamtin)	IV infusion	Tartaric acid 2% pH 2.5–3.5
Diazepam (Valium)	IM/IV	Sodium benzoate/benzoic acid 5.0% pH 6–7
Midazolam–HCL	IM/IV	pH 3
Octreotide acetate 84-(Sandostatin)	SC/IM/IV after dilution	Lactic acid 0.34% pH 4.2
Prochlorperazine edislate (Compazine)	IM/IV/IV infusion after dilution	Sodium biphosphate 0.5%
		Sodium tartrate 1.2% pH 4–6
Phenytoin sodium (Dilantin)	IM/IV	Sodium hydroxide pH 10–12.3

(From Refs. 84–86.)

organic solvent and water through the reduction of water–water interactions. This affords a system that is more favorable for nonpolar solutes.

Table 5 gives some physical properties of commonly used pharmaceutical cosolvents in water. Note that *n*-octanol is added in the table as a reference, since compounds that have large octanol/water partition coefficients have poor aqueous solubility. The use of cosolvent systems that have physical properties that are more similar to those of *n*-octanol would be expected to have greater success in solubilizing a nonpolar drug. In the following section it will be demonstrated that aqueous

solubilization via cosolvency is dependent on both the solute and cosolvent physical properties.

Mathematical Description for Cosolvency

Numerous methods have been proposed to predict or describe the effect of a particular cosolvent system (86–102) on drug solubility. A practical cosolvent model was developed by Yalkowsky and coworkers (103–106) by assuming the mixed solvent system is a linear

Table 5 Physical properties of some common cosolvents and their reference to water and *n*-octanol

Solvent	log $K_{o/w}$	Solubility parameter	Surface tension	Dielectric constant
Water	–4.00	23.4	72.0	81.0
Glycerin	–1.96	16.5	64.9	42.5
Propylene glycol	–0.92	12.6	37.1	32.0
PEG-400	–0.88	11.3	46.0	13.6
Dimethyl sulfoxide	–1.09	12.0	38.0	46.7
Dimethyl acetamide	–0.66	10.8	35.7	37.8
Ethanol	–0.31	12.7	22.2	24.3
<i>n</i> -Octanol	2.94	10.3	20.5	10.3

combination of the pure components. They found that this approach yields the log-linear relationship

$$\log S_{\text{mix}} = \log S_w + \sigma f_c \quad (18)$$

where the logarithm of the solubility of a nonpolar solute S_{mix} will increase with the cosolvent fraction (f_c), having a slope of σ , and an intercept of $\log S_w$. On a linear scale, an exponential increase in solubility is observed with an increase in cosolvent composition. Figure 4 illustrates the linear and exponential solubilization profiles.

For solid solutes, the melting point contributes only to the intercept of the solubilization profile. The independence of the intrinsic solubility and the solubilization slope can be demonstrated by the isomers benzo(a)pyrene and perylene. Figure 5 (top) shows the solubilization of benzo(a)pyrene and perylene in ethanol/water systems.

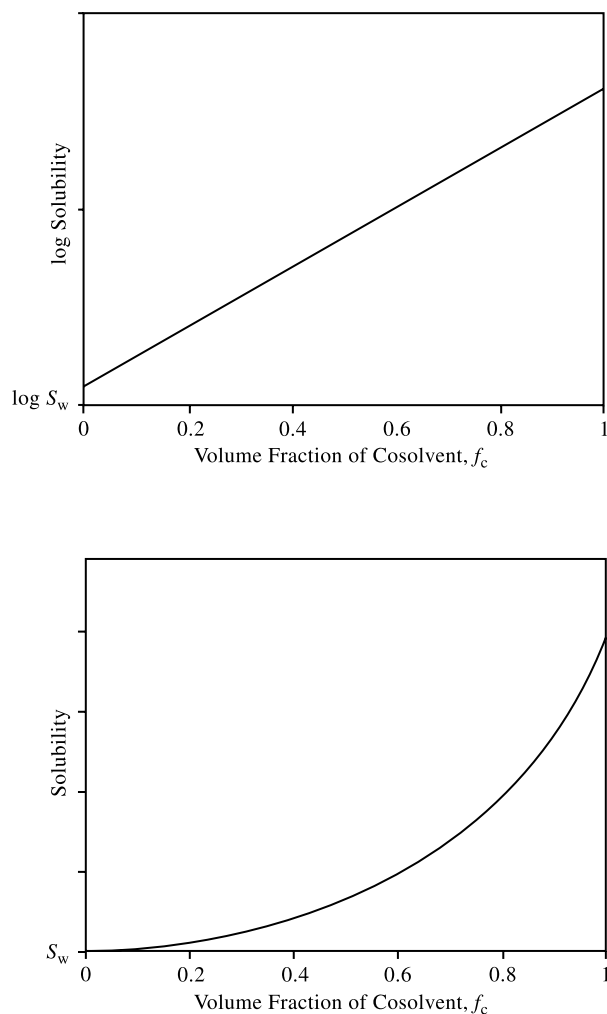


Fig. 4 Typical cosolvent solubilization curves for a nonpolar drug. (Top) Log-linear scale. (Bottom) Linear scale.

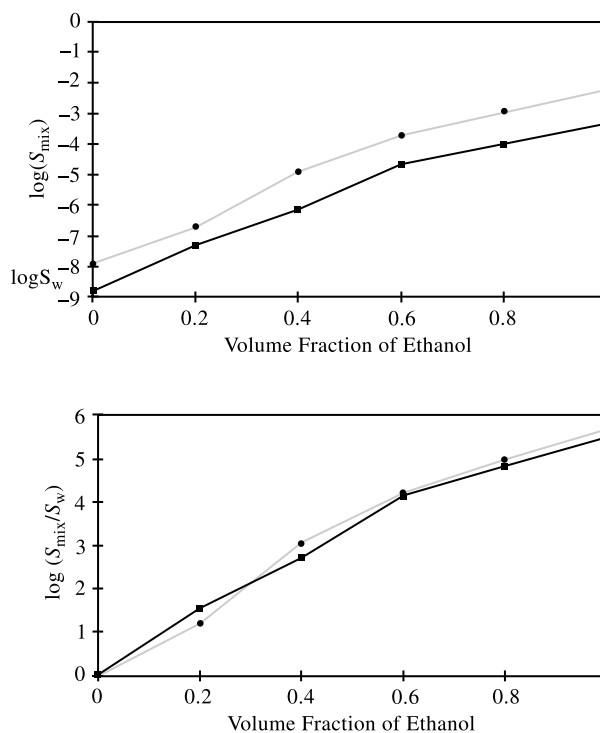


Fig. 5 (Top) Solubilization profile of \blacklozenge , benzo(a)pyrene and \blacksquare , perylene in ethanol–water mixtures. (Bottom) Relative solubilization ($\log(S_{\text{mix}}/S_w)$) of \blacklozenge , benzo(a)pyrene and \blacksquare , perylene in ethanol–water mixtures.

The $\log K_{o/w}$'s (or aqueous activity coefficients) can be considered to be equal for these two compounds, however, their melting points are different. Benzo(a)pyrene has a melting point of 179°C, while perylene has a melting point of 273°C, and as a result, the latter has a lower intrinsic solubility. Yet both compounds have solubilization slopes (σ) which are nearly identical. This is because σ is solely a function of the activity coefficients ($\log K_{o/w}$).

Since the intrinsic solubility does not directly affect the solubilization slope, it is convenient to arrange Eq. 18 to

$$\log(S_{\text{mix}}/S_w) = \sigma f_c \quad (19)$$

By plotting $\log(S_{\text{mix}}/S_w)$ versus f_c , solubilization lines are normalized with respect to the intrinsic solubility and pass through the origin on a semi log plot. This facilitates comparisons among different drugs as well as different cosolvents. Fig. 5 (bottom) shows the solubilization plots of benzo(a)pyrene and perylene using the expression of Eq. 19.

The ability of a cosolvent to solubilize a given solute can be related to the properties of both the solute and cosolvent. The solubilization slope, σ , for a given cosolvent system is dependent upon the polarity of

the drug and the polarity of the cosolvent and can be related by,

$$\sigma = s \log K_{o/w} + t \quad (20)$$

where polarity of the drug is indicated by $\log K_{o/w}$ and the polarity and physical properties of the cosolvent are represented through the parameters s and t .

Dependence of Solubilization on Solute Properties

As illustrated in Fig. 5, isomeric compounds or compounds that have almost identical aqueous activity coefficients (i.e., similar $\log K_{o/w}$'s) will have very similar solubilization slopes for a given cosolvent system. Differences in overall solubility are once again attributed to differences in the crystal contribution to solubility.

For nonpolar solutes, or those solutes that are more nonpolar than the cosolvent of choice, solubilization generally follows the log-linear model of Eq. 18. The degree of solubilization is directly related to the polarity of the solute. From Eq. 20 it has been shown that the solubilization slope, σ , can be related directly to the logarithm of the octanol–water partition coefficient ($\log K_{o/w}$). Fig. 6 shows data for a series of hydrocortisones in propylene glycol. As $\log K_{o/w}$ increases it follows that the solubilization slopes increase. Li and Yalkowsky (107) have also illustrated this relationship for nonpolar solutes in ethanol–water systems. Once again the trend maintains that as $\log K_{o/w}$ increases, solubilization is increased for a given cosolvent system (i.e., s and t are constant). For the

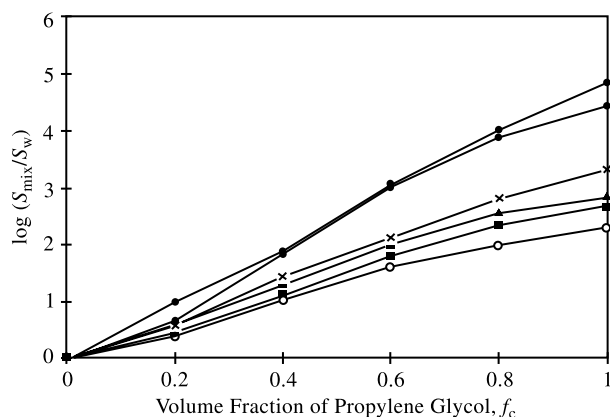


Fig. 6 Solubilization of different hydrocortisones in propylene glycol–water mixtures; ●, hydrocortisone acetate, ◆, hydrocortisone propionate, ×, hydrocortisone butyrate, ▲, hydrocortisone pentanoate, ■, hydrocortisone hexanoate, ○, hydrocortisone heptanoate.

most nonpolar compounds (largest $\log K_{o/w}$'s) solubilization approaches several orders of magnitude.

As $\log K_{o/w}$ decreases so does the ability of an organic cosolvent to solubilize the solute. For solutes that are semipolar and polar (polarities that are between water and a given cosolvent) little gain in solubility, if any, can be expected with the addition of an organic cosolvent. Semipolar solutes will have a maximal solubility at some mixed composition at a polarity that matches the solute. After a maximum solubility the slope becomes negative, and any additional cosolvent will decrease the solubility of the solute. The overall gain in solubility for semipolar solutes is generally less than a factor of five, which is significantly less than that of nonpolar solutes.

The solubility of polar solutes decreases with the addition of an organic cosolvent to water. Although it is obvious that cosolvents would not be used for solubilizing polar solutes, it is important to understand that for formulations that have other components or excipients, the use of a cosolvent for a nonpolar drug may cause solubility problems for polar excipients. If it is a polar drug that is insoluble, a reflection to Eq. 6 will aid in understanding that it is likely the crystal contribution that is limiting the solubility. This is generally the case since polar compounds have significant polar functionalities such as hydrogen bonding moieties that generate a strong crystal structure. If any of the functional groups are acidic or basic, the choice of pH may be the most efficient means by which to solubilize these compounds. Once again the trend maintains that as incremental hydrocarbon groups are added to a base structure, solubilization is increased for a given cosolvent system (i.e., s and t are constant). For the most nonpolar compounds (largest $\log K_{o/w}$) solubilization approaches 5–6 orders of magnitude.

Dependence of Solubilization on Cosolvent Properties

The reduction of intermolecular hydrogen bonding interactions of water when an organic cosolvent is mixed with water creates a solvent system that favors the dissolution of a nonpolar solute. The more nonpolar the cosolvent, the more nonpolar the cosolvent/water system will become and the greater the solubilization of a nonpolar drug. In addition to polarity considerations, it is also useful to take into account hydrogen bond donating and accepting capabilities when evaluating cosolvent systems (108). As can be seen from solubilization of benzocaine with different cosolvent systems in Fig. 7, there is a clear dependence between cosolvent polarity and solubilization. The more nonpolar the cosolvent the greater the solubilization slope.

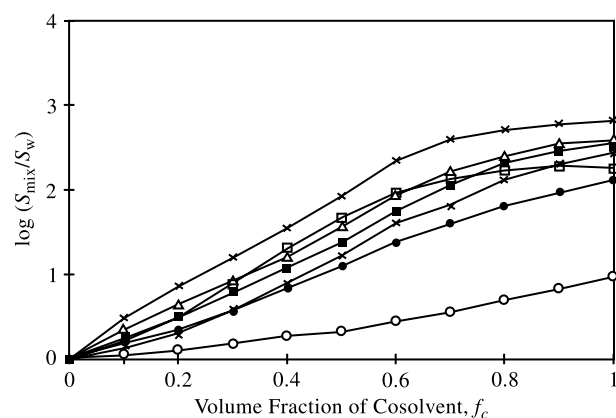


Fig. 7 Solubilization of benzocaine by *, dimethylacetamide, Δ , PEG-400, \blacksquare , PEG-200, \square , ethanol, \times , methanol, \blacklozenge , propylene glycol; and \circ , glycerin.

Relationship between Solubilization Slope, σ , and Solute Polarity

It is important to recognize that cosolvent solubilization does not rigorously follow a log-linear relationship. In fact a slight negative deviation is observed at low cosolvent compositions and a downward curving positive deviation is often observed at higher cosolvent compositions. Several investigators have attempted to explain or predict this type of behavior (89–102), however, the simplest and most useful remains the log-linear model. Another advantage of the log-linear model is that only two data points are needed to estimate the solubility of a solute in any ethanol–water composition.

With appreciation of the curvature observed for the cosolvent systems given above, a good correlation can still be obtained from simply looking at the terminal slope, σ , of a cosolvent system. That is, by using the log-linear approach (Fig. 4), the slope is simply the difference between the solubility of a solute in the pure cosolvent and in water. The general applicability of this has been demonstrated by Morris et al. (109) and Li and Yalkowsky

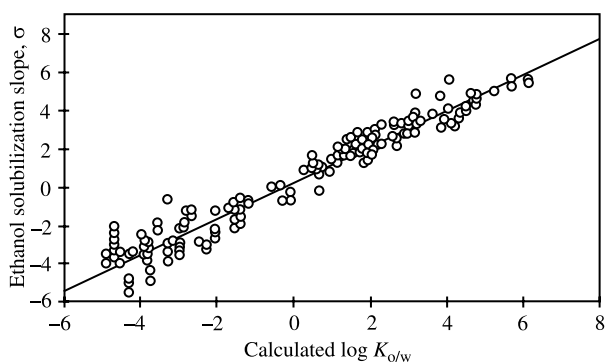


Fig. 8 Relationship between solubilization slope, σ and calculated $\log K_{o/w}$ (via $C \log P$) for a variety of different solutes in ethanol cosolvent systems. (From Ref. 107.)

(107). Figure 8 illustrates the relationship between solubilization slope, σ , and $\log K_{o/w}$. The gain in solubilization potential (increase in σ) is directly related to an increase in the calculated $\log K_{o/w}$.

Millard and Yalkowsky (110) have recently evaluated several cosolvent systems using $\log K_{o/w}$'s calculated by CLOGP (15). They found reasonable correlations between the solubilization slope and $\log K_{o/w}$ for ethanol, propylene glycol, polyethylene glycol 400, and glycerol cosolvent systems. A summary of their work is given in Table 6.

From Table 6 it is clear that cosolvent solubilization effectiveness (σ) is directly related to $\log K_{o/w}$. The least polar cosolvent, EtOH, produces the highest σ value, and the most polar cosolvent, glycerol, produces the lowest σ value.

Multiple Cosolvents

The use of multiple cosolvents can be a valuable method for solubilizing a poorly water soluble drug when a dosage form necessitates limits on the amount and type of cosolvent that can be utilized. The effects of multiple cosolvents on solubility can be reasonably approximated

Table 6 Regression analysis of solubilization slope, σ , and $\log K_{o/w}$ (Eq. 20) for different cosolvent systems

Cosolvent	Cosolvent $\log K_{o/w}$	σ	t	n	r^2	SE
Ethanol	−0.31	0.93	0.38	120	0.96	0.53
Propylene glycol	−0.92	0.76	0.57	93	0.93	0.48
Polyethylene glycol 400	−0.88	0.73	1.19	25	0.77	0.62
Glycerol	−1.96	0.34	0.29	26	0.81	0.34

Given are the slope, σ , intercept, t , number of compounds used in the regression, n , and the regression correlation values, r^2 , and standard errors, SE.

by simply expanding Eq. 18 to include a liner addition of cosolvents, i.e.,

$$\log S_{\text{mix}} = \log S_w + \sigma_1 f_{c1} + \sigma_2 f_{c2} + \sigma_3 f_{c3} + \dots \quad (21)$$

where the subscripts 1, 2, and 3 represent the slope and fraction of cosolvents 1, 2, 3, etc. Chien and Lambert (111) found that Eq. 21 adequately represented the solubility of a steroid in mixed solvent systems. Similarly, Prammar and Das Gupta (112) found that the solubility of spironolactone (Fig. 9) in multiple cosolvent systems followed a log-linear relationship. Data from Prammar and Das Gupta are presented in Fig. 10, showing the solubility of spironolactone for some ternary systems. Figure 10 demonstrates a nearly linear increase in solubility of spironolactone as the amount of propylene glycol or glycerin is added to a given PEG-400 concentration. All of the systems include 10% ethanol, giving a ternary system. Further investigation of the data, reveal that propylene glycol is a more efficient solubilizing agent than glycerol. In fact, the slope for propylene glycol is nearly twice that of glycerol, which is very similar to the findings above for the single cosolvent systems (Table 8).

Eq. 21 assumes that each cosolvent interacts independently. Of course this is not always the case. In fact the data of Prammar and Das Gupta (112) show a synergistic affect for the solubilization of the nonpolar solute, spironolactone. In addition, the use of one cosolvent can be used to increase the solubility of a partially miscible cosolvent in water. The facilitated cosolvency effect of multiple species in water has been shown by Gupta et al. (113) and Riley (114).

Solubilization through the Use of Cosolvents

Cosolvents can be a powerful tool by which to solubilize drugs. For a given cosolvent selected, it is clear that the degree of solubilization attained for any solute is directly proportional to the octanol/water partition coefficient of

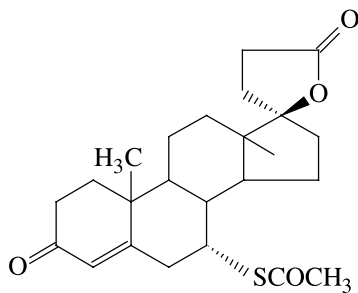


Fig. 9 Structure of spironolactone.

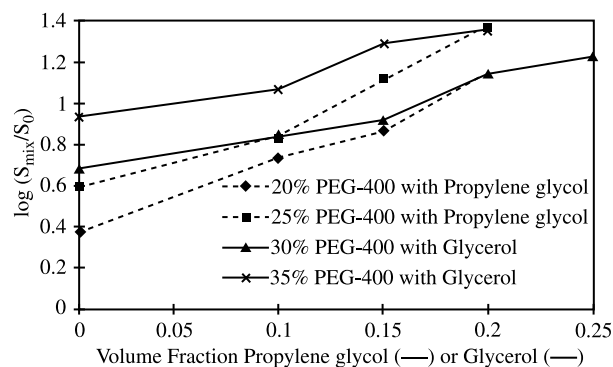


Fig. 10 Ternary cosolvent systems and resulting solubilization of spironolactone. S_0 represents the solubility of spironolactone in a 10% ethanol/water system. Given in the figure are different concentrations of PEG-400 with varied propylene glycol or glycerin volume fractions. (From Ref. 112.)

that solute. However, as with pH solubilization, solubilization through cosolvents is logarithmic, and as a result there is the potential for precipitation upon dilution.

A potential limitation to the use of cosolvents may be the choice and amount of cosolvent needed. For pharmaceuticals there are relatively few organic cosolvents that are generally regarded as safe. Rubino (115) has discussed the biological effects of many of the commonly utilized cosolvents such as ethanol, propylene glycol, glycerol, polyethylene glycols, and dimethylacetamide (DMA). The permissible amount of a given cosolvent will be dependent upon the dosage form. Table 7 gives some examples of products that contain cosolvents. The introduction of new, safe cosolvents would greatly enhance the technique of using cosolvents for solubilization. However, even with the limited number of acceptable cosolvents, cosolvency is clearly a valuable methods by which to solubilize poorly water soluble drugs.

SURFACTANTS

Surfactants are molecules with distinct polar and nonpolar regions. Most surfactants consist of a hydrocarbon segment (usually in the form of a long aliphatic chain segment) connected to a polar group. The polar group can be anionic (such as a carboxylate, sulfate, or sulfonate), cationic such as ammonium, trialkylammonium, or pyridinium), zwitterionic (such as glycine or carnitine) or nonionic such as polyethylene glycol, glycerol, or sugar. An illustration of the general structure of classical surfactants, as well as different pictorial representations, are illustrated in Fig. 11. Selected structures of non-classical surfactants are illustrated in Fig. 12. Although

Drug (product)	Route of administration	Cosolvent composition (% w/v)
Busulfan (Busulfex)	IV infusion after dilution	67% PEG 400 33% + <i>N,N</i> DMA
Chlordiazepoxide (Librium)	IV after dilution	20% Propylene glycol
Diazepam (Valium)	IM/IV	10% Ethanol 40% Propylene glycol
Digoxin (Lanoxin)	IM/IV	10% Ethanol 40% Propylene glycol
Dihydroergotamine mesylate (D.H.E.)	IM/IV	6.1% Ethanol 15% Glycerin
Esmolol–HCl (Brevibloc)	IV after dilution	25% Ethanol 25% Propylene glycol
Etoposide (VePesid)	IV infusion after dilution	30 % Ethanol 60% PEG 300
Fenoldopam (Corlopam)	IV after dilution	50% Propylene glycol
Lorazepam (Ativan)	IM/IV after dilution	18% Polyethylene glycol 400 80% Propylene glycol
Melphalan–HCl (Alkeran)	IV infusion after dilution	5% Ethanol 60% Propylene glycol
Nitroglycerin (Nitro-Bid)	IV infusion after dilution	70% Ethanol 4.5% Propylene glycol
Pentobarbital Sodium (Nembutal sodium)	IV/IM after dilution	10% Ethanol 40% Propylene glycol
Paricalcitol (Zemlar)	IV	20% Ethanol 30% Propylene glycol
Phenytoin Sodium (Dilantin)	IM/IV	10% Ethanol 40% Propylene glycol
Epinephrine (Sus-Phrine)	SC	32.5% Glycerin
Methocarbamil (Robaxin)	IM/IV after dilution	50% Polyethylene glycol 300
Oxytetracycline (Terramycin)	IM	67–75% Propylene glycol
Teniposide (Vumon)	IV infusion after dilution	42% Ethanol + 6% DMA

Fig. 11 Representation of classical surfactant structures: (a) Space-filling model; (b) chemical structure diagram; (c) simple line “stick” figure; (d) hybrid of diagrams a-c. (From Ref. 1.)

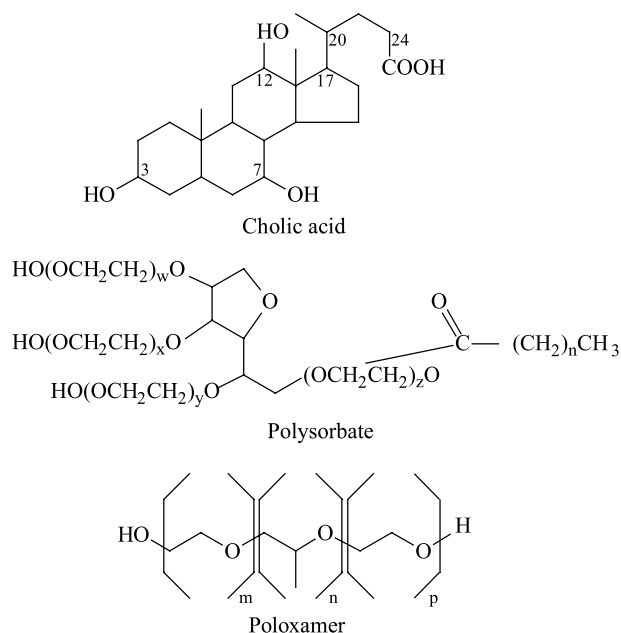


Fig. 12 Structure for the nonclassical surfactants cholic acid and generic structural features for polysorbates and poloxamers.

representation for a solute's total solubility, S_T , in a surfactant system is

$$S_T = S_w + \kappa(C_{\text{surf}} - \text{CMC}) \quad (22)$$

where C_{surf} is the total concentration of the surfactant and κ is the solubilization capacity. The quantity in parenthesis in Eq. 22 represents the micelle concentration. The solubilization capacity reflects the number of surfactant molecules that are required to solubilize a single solute molecule. Deviations from Eq. 22 are usually the result of changes in micelle shape or size that occur as the concentration of surfactant increases.

A micelle is a dynamic aggregation of any number of individual surfactant molecules, or monomers. Although the molecules are intertwined, they are in constant motion like those of a liquid. Thus, the interior of a micelle can be thought of as a separate phase and a micellar solution can be thought of as a microdispersion of that phase in water. If the micelle is considered to be a separate phase, it is then convenient to evaluate the solubilization capacity (κ), in terms of the partition between the micelle and water. The micellar partition coefficient, K_m , is defined as the ratio of the solute concentration in the micelle, C_m , to that of water, C_w :

$$K_m = C_m / C_w \quad (23)$$

The micelle/water partition coefficient for many solutes have been shown to correlate to the octanol/water partition

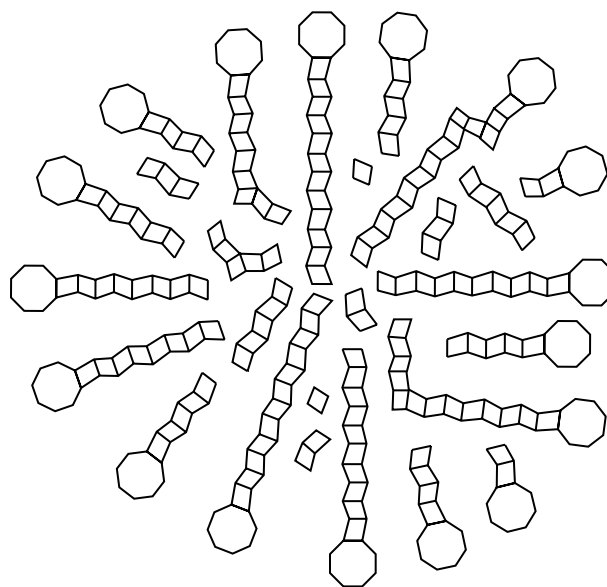


Fig. 13 Illustration of a micellar structure with nonpolar solutes in the core of the micelle.

coefficient (116–121). Data in Table 8, from Azaz and Donbrow (116), show that the micellar partition coefficients of the methylphenols increase with the number of methyl groups. Collett and Tobin (119) showed that the micellar partition coefficients of several benzoic acid derivatives are proportional to their octanol–water partition coefficient for poloxamers (Table 9). Tomida et al. (119) also illustrated that most of the 34 monosubstituted benzoic acids with Brij 35 have micellar partition coefficients that are inversely proportional to their aqueous solubilities and proportional to their octanol–water partition coefficients. The data of Tomida et al. (121) for some steroid hormones (Table 10), further illustrate the parallelism between octanol–water and micelle–water partition coefficients.

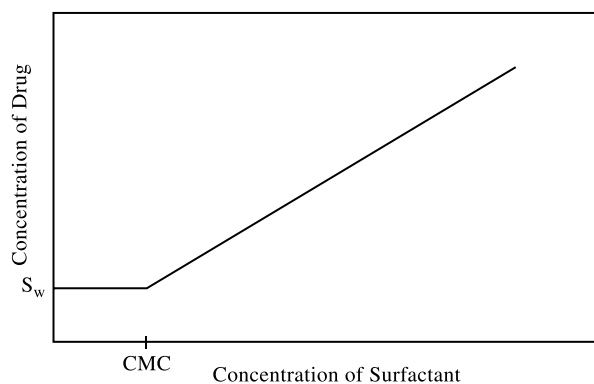


Fig. 14 General curve for solubilization with surfactants.

Table 8 Micellar partition coefficients of methylphenols by cetomacrogol

Solute	log S_w	K_M	log K_M	CLOGP
Phenol	—	42.00	1.62	1.49
2-Methylphenol	−0.62	79.50	1.90	1.95
3-Methylphenol	−0.70	76.40	1.88	1.96
4-Methylphenol	−0.85	85.10	1.93	1.94
2,3-Dimethylphenol	−1.43	169.00	2.23	2.42
2,4-Dimethylphenol	−1.29	125.00	2.10	2.30
2,5-Dimethylphenol	−1.54	197.00	2.29	2.33
2,6-Dimethylphenol	−1.31	114.00	2.06	2.36
3,4-Dimethylphenol	−1.41	151.00	2.18	2.23
3,5-Dimethylphenol	−1.40	132.00	2.12	2.35

(From Ref. 116.)

Semipolar solutes are generally solubilized by the polar regions of nonionic surfactants. The solute is absorbed into the polyoxyethylene mantle rather than the hydrocarbon core. The micellar partition coefficients of the semipolar solutes is also dependent upon their octanol–water partition coefficient. However, since these solutes are solubilized primarily in the mantle, the surfactant property that is most important is the number of polyoxyethylene units, and the alkyl chain length of the surfactant plays only a minor role in determining the micellar partition coefficient of a semipolar solute.

Materials that are solubilized in polyethylene glycol can be solubilized in the polyoxyethylene chains on the surface of a nonionic micelle. Ismail et al. (122) found that the micellar partition coefficients of barbiturates in polysorbates 20, 40, 60, and 80 is a function of the solute substituents and is proportional to the octanol–water partition coefficient of the barbiturate. Similarly, Ikeda

Table 10 Micellar partition coefficients of some steroids in Brij 35

Solute	log S_w	log K_M	CLOGP
Hydrocortisone	−2.97	1.99	1.55
Corticosterone	−3.24	2.27	1.94
Deoxycorticosterone	−3.45	2.60	2.90
Cortisone	−3.27	1.82	1.42
Hydrocortisone acetate	−4.34	2.36	2.19
Cortisone acetate	−4.21	2.31	2.10
Deoxycorticosterone	−4.63	2.86	3.08
11-Hydroxyprogesterone	−3.82	2.43	2.36
Progesterone	−4.42	3.12	3.87
Testosterone	−4.08	2.80	3.29
Prednisolone	−3.18	2.04	1.61
Prednisolone acetate	−4.37	2.45	2.40
Triamcinolone	−3.68	1.98	1.03
Betamethasone	−3.77	2.45	1.94
Dexamethasone acetate	−4.90	2.99	2.91
Betamethasone-17-valerate	−4.71	3.25	3.48

(Based on Ref. 121.)

et al. (123) showed that the solubilization of alkyl barbiturates by polyoxyethylene lauryl ether is not dependent upon the number of carbons in the substituents. Since the different polysorbates contain different aliphatic groups, the rather small dependence of solubilization upon polysorbate number (i.e., upon alkyl chain length) suggests that the barbiturates are not solubilized primarily in the hydrocarbon portion of the micelle. Gouda et al. (124) showed that the solubilization of barbiturates in polyoxyethylene stearates is proportional to the number of polyoxyethylene units in the surfactant.

Jafvert et al. (125) developed the following equation for predicting the molar solubilization capacity of organic

Table 9 Micellar partition coefficients of some benzoic acid derivatives by poloxamer L64 at 37°C

Solute	log S_w	K_M	log K_M	CLOGP
Benzoic acid	−1.52	43.50	1.64	1.81
2-Hydroxybenzoic acid	−1.76	83.00	1.92	2.38
3-Hydroxybenzoic acid	−1.21	28.50	1.45	1.50
4-Hydroxybenzoic acid	−1.23	30.80	1.49	1.58
4-Chlorobenzoic acid	−3.31	277.00	2.44	2.65
4-Bromobenzoic acid	−3.54	445.00	2.65	2.86
4-Iodobenzoic acid	−3.95	678.00	2.83	3.02
4-Nitrobenzoic acid	−2.66	92.70	1.97	1.89
4-Methoxybenzoic acid	−0.63	66.50	1.82	1.96

(From Ref. 119.)

compounds in nonionic surfactants on the basis of surfactant structure as well as solute structure.

$$K_M = K_{o/w}(0.030\sim L_1 - 0.0058\sim L_2 - 0.0056\sim L_3 + 0.0319\sim L_4) \quad (24)$$

where:

L_1 = number of straight chain aliphatic carbons in the hydrophobic tail

L_2 = number of repeating ethoxy groups

L_3 = number of carbons in a sorbitan group

L_4 = number of total carbons in alkylbenzene groups

This equation relates the micellar partition coefficient to the octanol–water partition coefficient. It also accounts for both the polar and nonpolar surfactant moieties.

Alvarez-Nunez and Yalkowsky (126) showed that the molar micellar partition coefficient is related to the octanol–water partition coefficient for a structurally diverse set of pharmaceutically important compounds through the relationship

$$K_M = a \log K_{o/w} + b \quad (25)$$

Which for posysorbate 80 is equal to

$$K_M = 0.92 \log K_{o/w} - 0.07 \quad (26)$$

Eq. 26 is then used to estimate total solubility by the equation

$$S_{\text{tot}} = S_w [1 + C_{\text{polysorbate}} \times 10^{(0.92 \log K_{o/w} - 0.07)}] \quad (27)$$

where $C_{\text{polysorbate 80}}$ is the concentration of polysorbate 80.

While both Eqs. 24 and 26 are quite useful, they must be employed with caution, especially for solutes that have some degree of amphiphilic character. Solutes that have surface-active physicochemical properties themselves (i.e., they have separate polar and nonpolar regions) tend to be more soluble than expected because they can accumulate at the core/mantel interface or the core/water

interface. In essence, weakly amphiphilic solutes can act as cosurfactants and form mixed micelles with nonionic surfactants as well as with ionic surfactants. This can alter both the CMC and the size of the micelle. In most cases this leads to a higher degree of solubilization than would be predicted on the basis of either molecular size or partition coefficient.

In summary, micellar solubilization for both nonpolar and semipolar surfactants can be obtained. Due to the limitation on surfactant quantity that can be practically used, due to biological constraints, micellar solubilization is typically more advantageous for low dose formulations. In addition, although a large number of anionic, cationic and nonionic surfactants are available for use as solubilizing agents (128–131), only polysorbate-80 and cremophor EL have been used to any significant extent in parenteral products. Table 11 lists some parenteral formulations that contain surfactants. It should be noted that adverse clinical events have been observed for patients receiving cremophor EL.

COMPLEXATION

Complexation is the association between two or more molecules to form a nonbonded entity with a well-defined stoichiometry. The two types of complexation that are most useful for increasing the solubility of drugs in aqueous media are stacking and inclusion. Stacking complexes are formed by the overlap of the planar regions of aromatic molecules, while inclusion complexes are formed by the insertion of the nonpolar region of one molecule into the cavity of another molecule (or group of molecules).

The mathematical description for the equilibrium constant of a 1:1 complex, $K_{1:1}$, is defined by

$$K_{1:1} = [SL]/[S][L] \quad (28)$$

Table 11 Examples of some products that contain surfactants

Drug (Product)	Route of administration	Surfactant composition (% w/v)
Taxol (Paclitaxel)	IV infusion after dilution	51% Cremophor EL
Valrubicin (Valstar)	IV infusion after dilution	50% Cremophor EL
Chlordiazepoxide HCl (Librium)	IM/IV (diluent)	4% Polysorbate 80
Cyclosporin (Sandimmune)	IV infusion after dilution	65% Cremophor EL
Etoposide (VePesid)	IV infusion after dilution	8% Polysorbate 80
Amiodarone (Cordarone)	IV infusion after dilution	10% Polysorbate 80

(From Refs. 84–86.)

where S is the concentration of the free solute, L is the concentration of the free ligand, and $[SL]$ is the concentration of the solute/ligand complex. The equilibrium constant is also commonly referred to as the stability constant or the complexation constant. If it takes two ligand molecules to complex with a solute molecule the complexation constant is defined by

$$K_{1:2} = [SL_2]/[S][L]^2 \quad (29)$$

The total solubility of the solute, S_T , for a solute that forms a 1:1 complex is

$$S_T = S_w + [SL] \quad (30)$$

where S_w is the intrinsic solubility of the solute in water. Similarly the total concentration of ligand, $[L^{tot}]$, in the system is

$$[L^{tot}] = [L] + [SL] \quad (31)$$

Combining Eqs. 28, 30, and 31, gives the general equation for solubilization by 1:1 complexation,

$$S_T = S_w + [K_{1:1}S_w/(1 + K_{1:1}S_w)]L^{tot} \quad (32)$$

where the intercept is the aqueous solubility of the solute. Eq. 32 represents a linear increase in the solubility of the solute with increasing ligand concentration, with a slope, $\sigma_{complex}$, of

$$\sigma_{complex} = [K_{1:1}S_w/(1 + K_{1:1}S_w)] \quad (33)$$

and the stability constant is

$$K_{1:1} = \sigma_{complex}/[S_w(1 - \sigma_{complex})]. \quad (34)$$

From the above it can be seen that, as the stability constant of a 1:1 complex increases, the slope will increase until the value converges to unity for a strong complex in which one ligand molecule solubilizes one solute molecule. The initial segment of the curve in Fig. 15 illustrates this. This linear region will continue until the solubility of the complex itself is reached, at which point the total solubility of the solute remains constant, as indicated by the central segment of the curve. A plateau is analogous to the maximum solubility of a salt as described previously. Further addition of the complexing agent can result in a reduction in of the concentration of the free solute and a leveling off of the curve at the solubility of the pure complex as illustrated by the final segment of the curve in Fig. 15.

The solubilization curve for a solute molecule that complexes with two ligand molecules is more complicated than those shown in Fig. 15. If the complexation constant for a second ligand is significantly lower than the first, a 1:1 complex will be formed at lower ligand concentration.

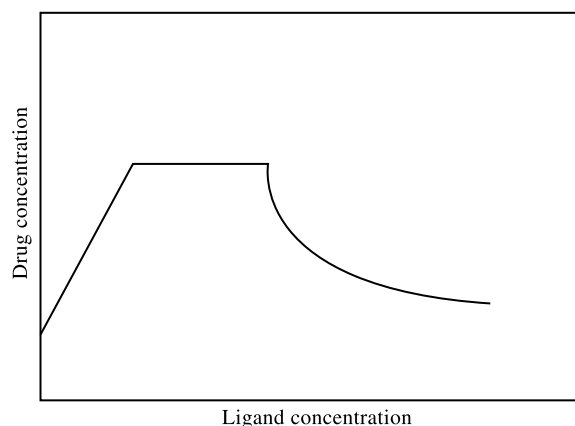


Fig. 15 General solubilization profile for complexation.

It will then combine with a second ligand to produce a 2:1 complex. Assuming that the latter is more soluble than the 1:1 complex, the solubilization curve will have two distinct slopes. If each ligand is equally capable of complexing with the solute, they will complex simultaneously to produce a convex up solubilization curve.

According to Eq. 32, the factors that determine the degree of solubilization of the solute are the complexation constant and the solubilities of the solute, the ligand, and the complex. As a result, the most useful ligands for solubilization in aqueous media are highly water soluble, and produce soluble complexes.

Self-Association and Stacking Complexation

Nonpolar moieties tend to be squeezed out of water by the strong hydrogen bonding interactions of the water. This causes some molecules to minimize the contact with water by aggregation of their hydrocarbon moieties. This aggregation is favored by large planar nonpolar regions in the molecule. Just as micelles can be pure or mixed, stacked complexes can be homogenous or mixed. The former is known as self-association and the latter as complexation. Some examples of substances that interact in an aqueous media by stacking are shown in Fig. 16.

Higuchi and Kristiansen (132) found that complexes formed between compounds containing aromatic acids or amides and compounds with aromatic nitrogens, such as those depicted in the second and third row of Fig. 16, consistently had higher stability constants than complexes made of a single compound. There have been a number of attempts to quantitatively relate or classify the chemical structures of the solute and ligand with their ability to complex (132–137).

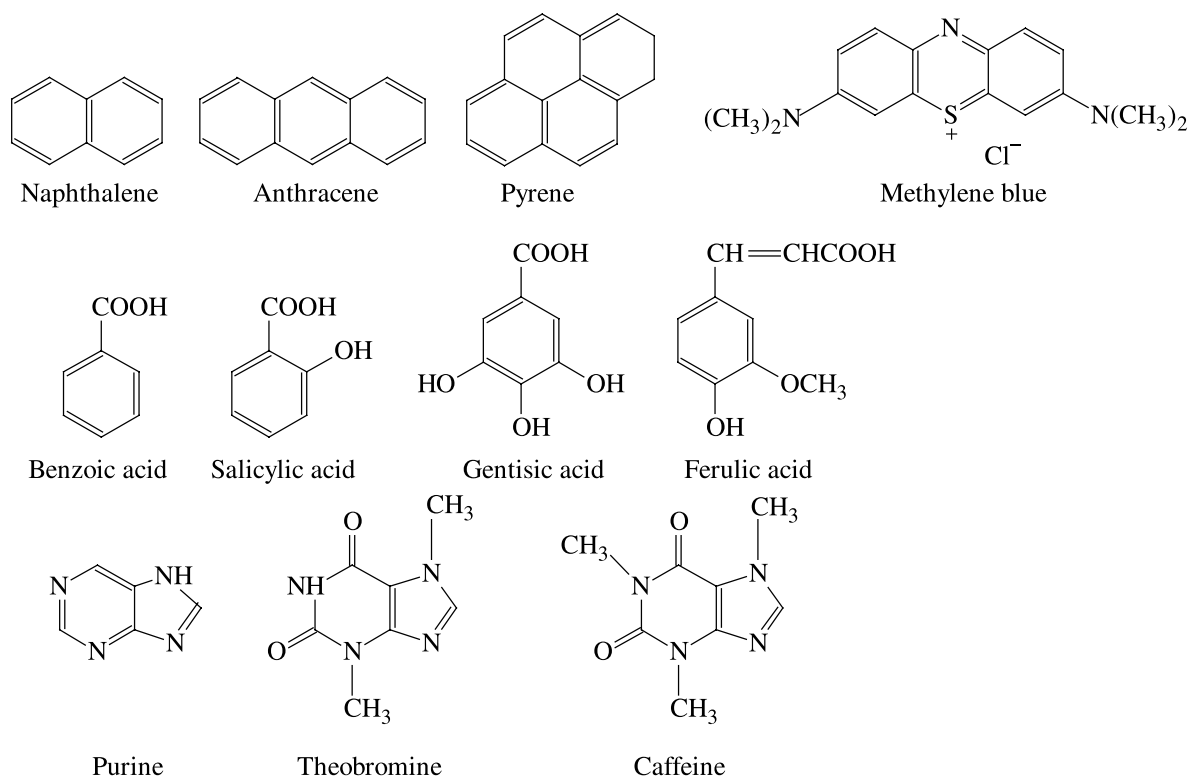


Fig. 16 Some compounds that are known to form stacking complexes.

Inclusion Complexes

An inclusion complex is produced by the inclusion of a nonpolar molecule or the nonpolar region of a molecule (known as the guest) into the nonpolar cavity of another molecule or group of molecules (known as the host). When the guest molecule enters the host molecule the contact between water and the nonpolar regions of both is reduced. Thus, inclusion phenomena are the result of the same driving force that produces micellization, self-association, and stacking; namely the squeezing out from water of nonpolar moieties.

The major structural requirement for inclusion complexation is a snug fit of the guest into the host cavity. The host cavity must be large enough to accommodate the guest and small enough to eliminate water so that the total contact between water and the nonpolar regions of the host and the guest is reduced.

The most commonly used host molecules are the cyclodextrins. These cyclic oligomers of glucose are relatively soluble in water and have cavities large enough to accept nonpolar portions of common drug molecules. The naturally occurring cyclodextrins contain 6, 7, and 8 glucopyranose units and are termed α , β , and γ ,

respectively. These are represented in Fig. 17 where “G” represents a glucopyranose unit. Modified cyclodextrins have one or more of the hydroxy groups of one or more of the glucopyranose units modified. Some of the more common modifications are with alkyl or hydroxyalkyl groups, or with anionic or cationic functionalities. Many of these modified cyclodextrins are more soluble than their naturally occurring precursors (138–140).

The size of the cavity in the cyclodextrin is the major factor in determining which guest solutes will be most acceptable for complexation. In general, alkyl groups will

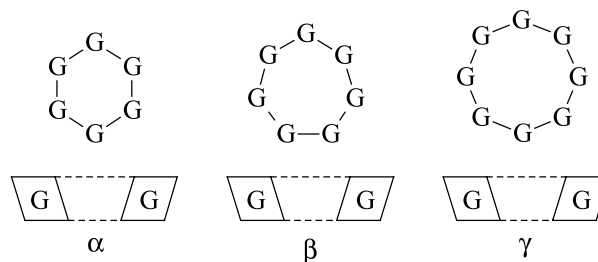


Fig. 17 Schematic representation of two views of the naturally occurring α , β , γ cyclodextrins.

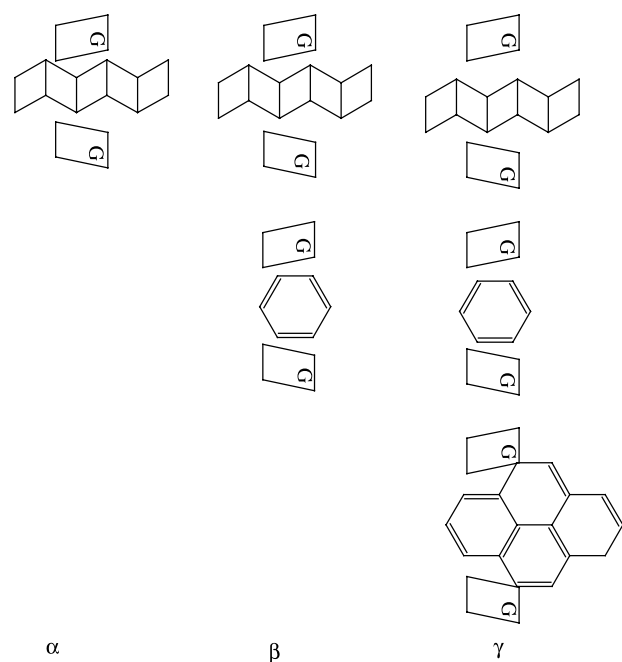


Fig. 18 Representation of the incorporation of pentane, benzene, and pyrene into α , β , and γ cyclodextrins.

fit well into the cavity of the α -cyclodextrins. The β -cyclodextrins are most well suited for accepting single aromatic rings, and the γ -cyclodextrins have large enough cavities to accommodate larger hydrocarbons such as pyrene. Fig. 18 illustrates these as well as other representative inclusion complexes.

The degree to which a solute molecule will be solubilized by a cyclodextrin molecule will depend on several properties. First the solute molecule must have a significant nonpolar portion in order to be squeezed out of the water and into the cyclodextrin cavity. Since the interior dimensions of a given cyclodextrin are fixed, a significant part of the molecule (or whole molecule) must then fit inside the cyclodextrin. Once presented to the interior cavity, it will be the fit, as well as the intermolecular interactions between the two molecules that will determine the strength of the complex. The stability constants for some compounds are given in Table 12. Note that the compounds in the upper third of the table contain aliphatic chains and are preferentially solubilized by α -cyclodextrin, while those in the middle contain aromatic rings and are best solubilized by β -cyclodextrin. The bottom third of the table contains fused ring compounds that are best solubilized by the larger γ -cyclodextrin.

Seo et al. (141) have shown the dependence of stability constant and overall solubility upon cyclodextrin ring size for spironolactone. From the data given in Fig. 19, it can be

Table 12 Apparent stability constants for complexes of α -, β -, and γ -cyclodextrins

Solute	α	β	γ
1-Butanol	89	16	
1-Octanol	6309	1479	
Carmofur	1200	670	180
Prostaglandin E1	1430	1700	530
Nonyl- <i>p</i> -hydroxybenzoate	4558	4327	
Decyl- <i>p</i> -hydroxybenzoate	4236	3306	
Octyl- <i>p</i> -hydroxybenzoate	3747	3920	
Ethyl- <i>p</i> -aminobenzoate	290	500	
Methyl- <i>p</i> -hydroxybenzoate	218	870	
Ethyl- <i>p</i> -hydroxybenzoate	178	1055	
Phenobarbital	30	1400	110
Phenytoin	90	1120	120
Anthracene	75	2000	220
Phenanthrene	16	1500	770
Benzo(<i>a</i>)anthracene	88	3225	605
Pyrene	148	543	1125
Benzo(<i>a</i>)pyrene	173	2219	63245
Triamcinolone	121	2370	9920
Triamcinolone acetonide	256	3230	26100
Triamcinolone diacetate	300	3530	12100
Dexamethasone	169	4660	26600
Dexamethasone acetate	316	9560	37300
Digitoxigenin	1700	130000	640000
Digitoxin	350	37000	78000

(Based upon Ref. 1.)

seen that for the relatively large spironolactone molecule (Fig. 9) the stability constant increases with size of the host cavity that is reflected through an increase in the linear slopes. The lowest maximum solubility was obtained with the β -cyclodextrin, which has the lowest solubility of the three cyclodextrins (142).

As discussed earlier and shown in Fig. 19, a significant limitation to the naturally occurring cyclodextrins is their limited solubility. Based on this and other physical property limitations, numerous derivatives have been prepared from the base cyclodextrins (143–147). Among the most studied of these is that of the hydroxypropyl- β -cyclodextrin, HP- β -CD and the sulfobutylethers of β -cyclodextrin, SBE- β -CD. These compounds are very soluble in water and can be used in high concentrations. Pitha et al. (148) investigated the use of high concentrations of HP- β -CD for a variety of compounds and found a wide range of solubility enhancement. The data from Pitha et al. is given in Table 13. SBE- β -CDs, have also shown good complexation characteristics. Table 14 shows the solubility enhancement for some selected compounds in a solution of Sulfobutylether

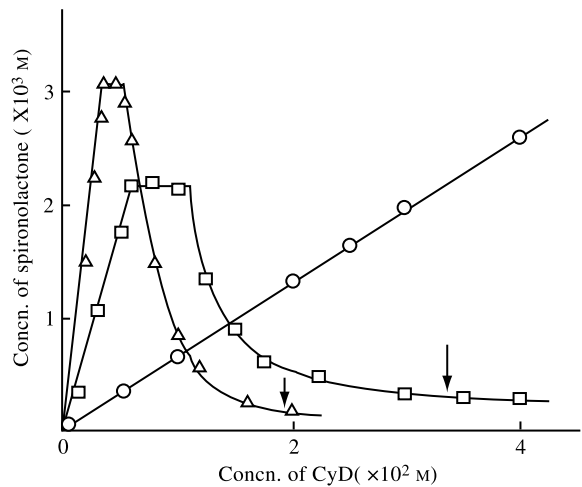


Fig. 19 Solubility of spironolactone as a function of cyclodextrin and concentration. ○, α-cyclodextrin, □, β-cyclodextrin, and, △, γ-cyclodextrin. (From Ref. 14.)

β-cyclodextrin IV. The selected data given in Tables 13 and 14 provide a good illustration of the practical range of solubilization that can be achieved by complexation.

Combination of pH and Complexation

The effect of pH on solubilization by complexation will depend entirely on the given solute and ligand. Tinwalla et al. (150) found that the combined use of ionization and complexation can be a powerful method for solubilization. They were able to significantly increase the solubility for a

Table 13 Solubility enhancement through the use of high concentrations of HP-β-CD

Solute	% HP-β-CD	Solubility enhancement
Estriol	50	13666
Estradiol	40	7000
Progesterone	40	2266
Spironolactone	40	1400
Testosterone	40	1461
Digoxin	50	971
Dexamethasone	50	240
Chlorthalidone	50	87.5
Diphenylhydantoin	50	57
Furosamide	50	24
Nitroglycerin	40	8.3
Acetamidopen	50	6
Apomorphine	50	5.8
Theophylline	50	1.3

(From Ref. 148.)

Table 14 Solubility enhancement through the use of 0.1 M SBE-β-CD (IV)

Solute	Solubility enhancement
Testosterone	2020
Prednisolone acetate	426
Dexamethasone	208
Dapsone	189
Prednisolone	106
Methylprednisolone	88
Hydrocortisone	87
Menadione	69
Benzyl guanine	68
Chloramphenicol	8

(From Ref. 149.)

thiazolobenzimidazole derivative, TBI (shown in Fig. 20), through the use of both pH and complexation via HP-β-CD.

Li et al. (151, 152) derived a mathematical expression to explain the combined effect of pH and complexation on solubilization. If either the ligand or solute ionizes with pH, the stability constant will typically decrease. However, even with a decrease in the stability constant the solubility can sometimes increase. Although the

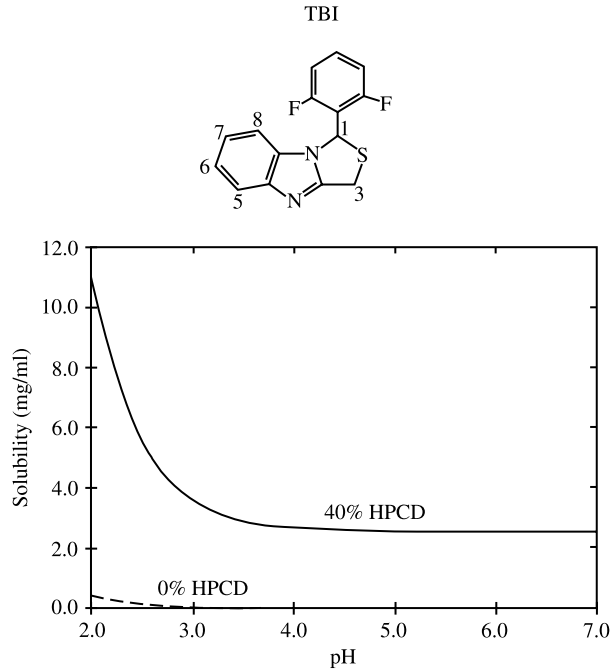


Fig. 20 Solubility of TBI (structure top, $pK_a = 3.55$), as a function of pH for 0 and 40% HP-β-CD solutions. (From Ref. 150.)

Table 15 Examples of some parenteral products that contain complexation agents

Drug (Product)	Route of administration	Surfactant composition (% w/v)
Alprostadil (Edex)	Intracavernosal (reconstituted)	α -Cyclodextrin
Itraconazole (Sporanox)	IV Infusion after dilution	40% Hydroxypropyl- β -cyclodextrin
Ziprasidone mesylate (Phase III)	IM/SC	Sulfobutylether- β -cyclodextrin
Prednisolone phosphate sodium (Hydeltrasol)	IV Infusion after dilution	2.5% Niacinamide

(From Refs. 84–86.)

stability complex for the protonated form of TBI was nearly 13-fold less than that of the neutral complex, they were still able to attain a 3-fold increase in solubility.

Solubilization through Complexation

With the general characterization for stacking complexes being planar, the list of possible candidates is extensive. Of course, not all possible ligands can be considered acceptable. Some of the planar molecules that form stacking complexes (such as those given in Fig. 16) are pharmacologically active. Others like the parabens or benzyl alcohol can be used as antimicrobial agents. Furthermore, the addition of these organic compounds can also facilitate solubilization through cosolvency. Inclusion complexation has been proven to be useful for compounds with isolated hydrocarbon regions. Solutes with large or multiple hydrocarbon regions are most efficiently solubilized by complexation, however a shape requirement must be met in order for inclusion complexation to occur. An abbreviated list of excipients used in some parenteral products is given in Table 15.

REFERENCES

1. Yalkowsky, S.H. *Solubility and Solubilization in Aqueous Media*; Oxford University Press: New York, 1999.
2. Liu, R. *Water-Insoluble Drug Formulations*; Interpharm Press: Englewood, CO, 2000.
3. Prausnitz, J.M.; Lichtenthaler, R.N.; Gomes de Azevedo, E. *Molecular Thermodynamics of Fluid-Phase Equilibria*, 2nd Ed., Prentice-Hall: New Jersey, 1986.
4. Yalkowsky, S.H. *Ind. Eng. Chem. Fundam.* **1979**, *18*, 108–111.
5. Myrdal, P.B.; Ward, G.H.; Dannenfelser, R.-M.; Mishra, D.; Yalkowsky, S.H. *Chemosphere* **1992**, *24*, 1047–1061.
6. Myrdal, P.B.; Ward, G.H.; Simamora, P.; Yalkowsky, S.H. SAR and QSAR in Environ. Res. **1993**, *1*, 53–61.
7. Myrdal, P.B.; Manka, A.M.; Yalkowsky, S.H. *Chemosphere* **1995**, *30*, 1619–1637.
8. Lee, Y.-C.; Myrdal, P.B.; Yalkowsky, S.H. *Chemosphere* **1996**, *33*, 2129–2144.
9. Pinsuwan, S.; Myrdal, P.B.; Lee, Y.-C.; Yalkowsky, S.H. *Chemosphere* **1997**, *35*, 2503–2513.
10. Slusher, J.T. *Fluid Phase Equilib.* **1988**, *153*, 45–61.
11. Ruelle, P. *J. Org. Chem.* **1999**, *12*, 769–786.
12. Abraham, M.H.; Le, J. *J. Pharm. Sci.* **1999**, *88*, 868–880.
13. Lin, S.-T.; Sandler, S.I. *AIChE J.* **1999**, *45*, 2606–2618.
14. Yalkowsky, S.H.; Valvani, S.C. *J. Pharm. Sci.* **1980**, *69*, 912–922.
15. *ClogP[®]*, version 2.2.02b, Bio-Byte Corporation: 1997.
16. Byrn, S.R.; Pfeiffer, R.R.; Stowell, J.G. *Solid State Chemistry of Drugs*, 2nd Ed.; SSCI, Inc.: West Lafayette, IN, 1999.
17. Noyes, A.A.; Whitney, W.R. *J. Am. Chem. Soc.* **1897**, *19*, 930–934.
18. Hamlin, W.E.; Northam, J.I.; Wagner, J.G. *J. Pharm. Sci.* **1965**, *54*, 1651–1653.
19. Yoshihashi, Y.; Kitano, H.; Etsuo, Y.; Terada, K. *Int. J. Pharm. Sci.* **2000**, *204*, 1–6.
20. Nicklasson, M.; Brodin, A. *Int. J. Pharm. Sci.* **1984**, *18*, 149–156.
21. Allen, D.J.; Kwan, K.C. *J. Pharm. Sci.* **1969**, *58*, 1190–1193.
22. Goldberg, A.H.; Gibaldi, M.; Kanig, J.L. *J. Pharm. Sci.* **1966**, *55*, 487–472.
23. Kaneniwa, N.; Watari, N. *Chem. Pharm. Bull.* **1977**, *25*, 867–875.
24. Florence, A.T.; Salole, E.G. *J. Pharm. Pharmacol.* **1976**, *28*, 637–642.
25. Florence, A.T.; Salole, E.G.; Stanlake, J.B. *J. Pharm. Pharmacol.* **1974**, *26*, 479–480.
26. Ikekawa, A.; Hayakawa, S. *Bull. Chem. Soc. Jpn.* **1981**, *54*, 2587–2591.
27. Picolo, T.; Sakr, A. *Pharm. Ind.* **1984**, *46*, 1277–1279.
28. Pereira De Almeida, L.; Simoes, S.; Brito, P.; Portugal, A.; Figueiredo, M. *J. Pharm. Sci.* **1997**, *86*, 726–731.
29. Mosharraf, M.; Sebhatu, T.; Nystrom, C. *Int. J. Pharm.* **1999**, *177*, 29–51.
30. Aguiar, A.J.; Krc, J.; Kinkel, A.W.; Samyn, J.C. *J. Pharm. Sci.* **1967**, *56*, 847–853.
31. Aguiar, A.J.; Zelmer, J.E. *J. Pharm. Sci.* **1969**, *58*, 983–987.
32. Kato, Y.; Kohetsu, M. *Chem. Pharm. Bull.* **1981**, *29*, 268–272.
33. Yokoyama, T.; Umeda, T.; Kuroda, K.; Kuroda, T.; Asada, S. *Chem. Pharm. Bull.* **1981**, *29*, 194–199.

34. Kokubu, H.; Morimoto, K.; Ishida, T.; Inoue, M.; Morisaka, K. *Int. J. Pharm.* **1987**, *35*, 181–183.
35. Kuroda, K.; Yokoyama, T.; Umeda, T.; Takagishi, Y. *Chem. Pharm. Bull.* **1978**, *26*, 2565–2568.
36. Umeda, T.; Matsuzawa, A.; Ohnishi, N.; Kuroda, T. *Chem. Pharm. Bull.* **1984**, *32*, 1637–1640.
37. Suleiman, M.S.; Najib, N.M. *Int. J. Pharm.* **1989**, *50*, 103–109.
38. Himmelreich, M.; Rawson, B.J.; Watson, T.R. *Austral. J. Pharm. Sci.* **1977**, *6*, 123–125.
39. Clements, J.A.; Popli, S.D. *Can. J. Pharm. Sci.* **1973**, *8*, 88–92.
40. Pearson, J.T.; Varney, G. *J. Pharm. Pharmacol.* **1973**, *25*, 62–70.
41. Wells, J.I. *Pharmaceutical Preformulation. The Physico-chemical Properties of Drug Substances*; Ellis Horwood, Ltd.: United Kingdom, 1988; 94–95.
42. Poole, J.W.; Bahal, C.K. *J. Pharm. Sci.* **1968**, *57*, 1945–1948.
43. Suryanarayanan, R.; Mitchell, A.G. *Int. J. Pharm.* **1985**, *24*, 1–17.
44. Fukumori, Y.; Fukuda, T.; Yamamoto, Y.; Shigitani, Y.; Hanyu, Y.; Takeuchi, Y.; Sato, N. *Chem. Pharm. Bull.* **1983**, *31*, 4029–4039.
45. Jozwiakowski, M.J.; Nguyen, N.-A.T.; Sisco, J.M.; Spancake, C.E. *J. Pharm. Sci.* **1996**, *85*, 193–199.
46. Buxton, P.C.; Lynch, I.R.; Roe, J.M. *Int. J. Pharm.* **1988**, *42*, 135–143.
47. Sekiguchi, K.; Kanke, M.; Nakamura, N.; Tsuda, Y. *Chem. Pharm. Bull.* **1973**, *21*, 1592–1600.
48. Kozje, F.; Golic, L.; Zupet, P.; Palka, E.; Vodopivec, P.; Japelj, M. *Acta Pharm. Jugosl.* **1985**, *35*, 275–281.
49. Abdallah, O.; El-Fattah, S.A. *Pharm. Ind.* **1984**, *46*, 970–971.
50. Rodriguez-Hornedo, N.; Lechuga-Ballesteros, D.; Wu, H.-J. *Int. J. Pharm.* **1992**, *85*, 149–162.
51. Mentasti, E.; Rinaudo, C.; Boistelle, R. *J. Chem. Engin. Data* **1983**, *28*, 247–251.
52. Ghosh, S.; Grant, D.J.W. *Int. J. Pharm.* **1995**, *114*, 185–196.
53. Matsuda, Y.; Tatsumi, E. *J. Pharmacobio-Dyn.* **1989**, *12*, 38.
54. Toffoli, F.; Avico, U.; Signoretti, Ciranni E.; Di Francesco, R.; Di Palumbo, V.S. *Ann. Chem.* **1973**, *63*, 1–4.
55. Corrigan, O.I.; Holohan, E.M.; Sabra, K. *Int. J. Pharm.* **1984**, *18*, 195–200.
56. Otsuka, M.; Kaneniwa, N. *Chem. Pharm. Bull.* **1983**, *31*, 230–236.
57. Higuchi, W.I.; Lau, P.K.; Higuchi, T.; Shell, J.W. *J. Pharm. Sci.* **1963**, *52*, 150–153.
58. Doherty, C.; York, P. *Drug Dev. Ind. Pharm.* **1989**, *15*, 1969–1987.
59. Mullins, J.D.; Macek, T.J. *J. Pharm. Sci.* **1960**, *49*, 245–248.
60. Fukuoka, E.; Makita, M.; Yamamura, S. *Chem. Pharm. Bull.* **1987**, *35*, 2943–2948.
61. Giron, D. *Thermochim. Acta.* **1995**, *248*, 1–59.
62. Cohn, E.J.; McMeekin, T.L.; Edsall, J.T.; Weare, J.H. *J. Am. Chem. Soc.* **1934**, *56*, 2270–2282.
63. Tsuji, A.; Nakashima, E.; Hamano, S.; Yamana, T. *J. Pharm. Sci.* **1978**, *67*, 1059–1066.
64. Tsuji, A.; Nakashima, E.; Yamana, T. *J. Pharm. Sci.* **1979**, *68*, 308–315.
65. Tsuji, A.; Nakashima, E.; Nishide, K.; Deguchi, Y.; Hamano, S.; Yamana, T. *Chem. Pharm. Bull.* **1983**, *31*, 4057–4069.
66. Garren, K.W.; Pyter, R.A. *Int. J. Pharm.* **1990**, *63*, 167–172.
67. Chowhan, Z.T. *J. Pharm. Sci.* **1978**, *67*, 1257–1260.
68. Kramer, S.F.; Flynn, G.L. *J. Pharm. Sci.* **1972**, *61*, 1896–1904.
69. Avdeef, A. *Pharm. Pharmacol. Commun.* **1998**, *4*, 165–178.
70. Streng, W.H.; Hsi, S.K.; Helms, P.E.; Tan, H.G.H. *J. Pharm. Sci.* **1984**, *73*, 1679–1684.
71. Streng, W.H. *Int. J. Pharm. Sci.* **1999**, *186*, 137–140.
72. Berge, S.M.; Bighley, L.D.; Monkhouse, D.C. *J. Pharm. Sci.* **1977**, *66*, 1–19.
73. Bighley, L.D.; Berge, S.M.; Monkhouse, D.C. *Salt Forms of Drugs and Absorption. Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1996; 13, 453–499.
74. Agharkar, S.; Lindenbaum, S.; Higuchi, T. *J. Pharm. Sci.* **1976**, *65*, 747–749.
75. Gould, P.L. *Int. J. Pharm.* **1986**, *33*, 201–217.
76. Morris, K.R.; Fakes, M.G.; Thakur, A.B.; Newman, A.W.; Singh, A.K.; Venit, J.J.; Spagnuolo, C.J.; Serajuddin, A.T.M. *Int. J. Pharm.* **1994**, *105*, 209–217.
77. Spurlock, C.H. *J. Parent. Sci., Tech.* **1986**, *40*, 70–72.
78. Jusko, W.J.; Gretch, M.; Gassett, R. *J. Am. Med. Assoc.* **1973**, *225*, 176.
79. Yalkowsky, S.H.; Valvani, S.C.; Johnson, B.W. *J. Pharm. Sci.* **1983**, *72*, 1014–1017.
80. Ward, G.H.; Yalkowsky, S.H. *J. Parent. Sci. Tech.* **1993**, *47*, 161–165.
81. Jamerson, B.D.; Dukes, G.E.; Brouwer, K.L.R.; Donn, K.H.; Massenheimer, J.A.; Powel, J.R. *Pharmacotherapy* **1994**, *14*, 47–52.
82. Myrdal, P.B.; Simamora, P.; Surakitbanharn, Y.; Yalkowsky, S.H. *J. Pharm. Sci.* **1995**, *84*, 849–852.
83. Surakitbanharn, Y.; Simamora, P.; Ward, G.H.; Yalkowsky, S.H. *Int. J. Pharm.* **1994**, *109*, 27–33.
84. Strickley, R.G. *PDA J. Pharm. Sci. Technol.* **1999**, *53*, 324–349.
85. Strickley, R.G. *PDA J. Pharm. Sci. Technol.* **2000**, *54*, 69–96.
86. Strickley, R.G. *PDA J. Pharm. Sci. Technol.* **2000**, *54*, 152–169.
87. Paruta, A.N.; Sciarone, B.J.; Lordi, N.G. *J. Pharm. Sci.* **1965**, *54*, 1325–1333.
88. Sorby, D.L.; Liu, G.; Horowitz, K.N. *J. Pharm. Sci.* **1965**, *54*, 1811–1813.
89. Martin, A.; Newburger, J.; Adjei, A. *J. Pharm. Sci.* **1980**, *69*, 487–490.
90. Martin, A.; Paruta, A.N.; Adjei, A. *J. Pharm. Sci.* **1981**, *70*, 1115–1120.
91. Williams, A.; Amidon, G.L. *J. Pharm. Sci.* **1984**, *73*, 9–13.
92. Williams, A.; Amidon, G.L. *J. Pharm. Sci.* **1984**, *73*, 14–17.
93. Williams, A.; Amidon, G.L. *J. Pharm. Sci.* **1984**, *73*, 18–23.

94. Ochsner, A.B.; Belloto, R.J., Jr.; Sokoloski, T.D. *J. Pharm. Sci.* **1985**, *74*, 277–282.
95. Acree, W.E., Jr.; McCargar, J.W.; Zvaigzne, A.E.; Teng, I.L. *Phys. Chem. Liq.* **1991**, *23*, 27–35.
96. Escalera, J.B.; Bustamante, P.; Martin, A. *J. Pharm. Pharmacol.* **1994**, *46*, 172–176.
97. Barzegar-Jalali, M.; Hanaee, J. *Int. J. Pharm.* **1994**, *109*, 291–295.
98. Li, A.; Andren, A.W. *Env. Sci. Tech.* **1995**, *29*, 3001–3006.
99. Barzegar-Jalali, M.; Jouyban-Gharamaleki, A. *Int. J. Pharm.* **1997**, *152*, 247–250.
100. Jouyban-Gharamaleki, A.; Barzegar-Jalali, M.; Acree, W.E., Jr. *Int. J. Pharm.* **1998**, *166*, 205–209.
101. Jouyban-Gharamaleki, A.; Acree, W.E., Jr. *Int. J. Pharm.* **1998**, *167*, 177–182.
102. Jouyban-Gharamaleki, A.; Valaee, L.; Barzegar-Jalali, M.; Clark, B.J.; Acree, W.E., Jr. *Int. J. Pharm.* **1999**, *177*, 93–101.
103. Yalkowsky, S.H.; Amidon, G.L.; Zograf, G.; Flynn, G.L. *J. Pharm. Sci.* **1972**, *64*, 48–52.
104. Yalkowsky, S.H.; Valvani, S.C.; Amidon, G.L. *J. Pharm. Sci.* **1976**, *65*, 1488–1495.
105. Yalkowsky, S.H.; Roseman, J.T. *Solubilization of Drugs. Techniques of Solubilization of Drugs*; Yalkowsky, S.H., Ed.; Ch. 3 Marcel Dekker, Inc.: New York, 1981.
106. Yalkowsky, S.H.; Rubino, J.T. *J. Pharm. Sci.* **1985**, *74*, 416–421.
107. Li, A.; Yalkowsky, S. *J. Pharm. Sci.* **1994**, *83*, 1735–1740.
108. Rubino, J.T.; Yalkowsky, S.H. *Pharm. Res.* **1987**, *4*, 220–230.
109. Morris, K.R.; Abramowitz, R.; Pinal, R.; Davis, P.; Yalkowsky, S.H. *Chemosphere* **1988**, *17*, 285–298.
110. Millard, J.W.; Alvarez-Nunez, F.A.; Yalkowsky, S.H. submitted. *J. Pharm. Sci.* **2001**.
111. Chein, Y.W.; Lambert, H.J. *Chem. Pharm. Bull.* **1975**, *23*, 1085–190.
112. Pramdar, Y.V.; Das, Gupta V. *Pharmazie* **1994**, *49*, 661–665.
113. Gupta, B.; Mishra, D.S.; Cheng, C-H.; Yalkowsky, S.H. *Tox. Environ. Chem.* **1991**, *33*, 7–21.
114. Riley, C.M. *Solubilization of Some Poorly Soluble Drugs by Cosolvents*; M.S. Thesis University of Arizona: Tucson, AZ, 1990.
115. Rubino, J.T. *Cosolvents and Cosolvency. The Encyclopedia of Pharmaceutical Technology*; Marcel Dekker, Inc. 1990; 3, 375–398.
116. Azaz, E.; Donbrow, M. *J. Colloid and Interface Sci.* **1976**, *57*, 11–15.
117. Fashelelbom, K.M.S.; Timoney, R.S.; Corrigan, O.I. *Pharm. Res.* **1993**, *10*, 631–634.
118. Collet, J.H.; Koo, L. *J. Pharm. Sci.* **1975**, *64*, 1253–1255.
119. Collet, J.H.; Tobin, E.A. *J. Pharm. Pharmacol.* **1979**, *31*, 174–177.
120. Tomida, H.; Yotsuyanagi, T.; Ikeda, K. *Chem. Pharm. Bull.* **1978**, *26*, 2824–2831.
121. Tomida, H.; Yotsuyanagi, T.; Ikeda, K. *Chem. Pharm. Bull.* **1978**, *26*, 2832–2837.
122. Ismail, A.A.; Gouda, M.W.; Motawi, M.M. *J. Pharm. Sci.* **1970**, *59*, 220–224.
123. Ikeda, J.; Kato, K.; Tukamoto, T. *Chem. Pharm. Bull.* **1971**, *19*, 2510–2517.
124. Gouda, M.W.; Ismail, A.A.; Motawi, M.M. *J. Pharm. Sci.* **1970**, *59*, 1402–1405.
125. Jafvert, C.T.; Van Hoof, P.L.; Heath, J.K. *Water Res.* **1994**, *28*, 1009–1017.
126. Alvarez-Nunez, F.A.; Yalkowsky, S.H. *Int. J. Pharm.* **2000**, *200*, 217–222.
127. Nakagawa, T.; Tori, K. *Koll. Z.* **1960**, *168*, 132–139.
128. Elworthy, P.H.; Florence, A.T.; Macfarlane, C.B. *Solubilization by Surface-Active Agents*; Chapman Hall: New York, 1968.
129. Attwood, D.; Florence, A.T. *Surfactant Systems*; Chapman Hall: New York, 1983.
130. Rosen, M. *Surfactants and Interfacial Phenomena*; Second Ed. John Wiley & Sons: New York, 1989.
131. Christian, S.D.; Scamehorn, J.F. *Solubilization in Surfactant Aggregates*; Marcel Dekker, Inc.: New York, 1994.
132. Higuchi, T.; Kristiansen, H. *J. Pharm. Sci.* **1970**, *59*, 1601–1608.
133. Kakemi, K.; Sezaki, T.; Mitsunaga, T.; Nakano, M. *J. Pharm. Sci.* **1970**, *59*, 1579–1601.
134. Kenley, R.A.; Jackson, S.E.; Winterle, J.S.; Shunko, Y.; Visor, G.C. *J. Pharm. Sci.* **1986**, *75*, 648–653.
135. Fawzi, M.B.; Davison, E.; Tute, M.S. *J. Pharm. Sci.* **1980**, *69*, 104–106.
136. Boje, K.M.; Sak, M.; Fung, H-L. *Pharm. Res.* **1988**, *5*, 655–659.
137. Connors, K.A. *Binding Constants*; Wiley-Interscience Publication: New York, 1987; 1–101.
138. Yoshida, A.; Arima, H.; Uekama, K.; Pitha, J. *Int. J. Pharm.* **1988**, *46*, 217–222.
139. Muller, B.W.; Brauns, U. *Int. J. Pharm.* **1985**, *26*, 77–85.
140. Okada, Y.; Kubota, Y.; Koizumi, K.; Hizukuri, S.; Ohfuchi, T.; Ogata, K. *Chem. Pharm. Bull.* **1988**, *36*, 2176–2185.
141. Seo, H.; Tsuruoka, M.; Hashimoto, T.; Fujinaga, T.; Otagiri, M.; Uekama, K. *Chem. Pharm. Bull.* **1983**, *31*, 286–291.
142. Jozwiakowski, M.J.; Connors, K.A. *Carbohydr. Res.* **1985**, *143*, 51–59.
143. Muller, B.W.; Brauns, U. *Int. J. Pharm.* **1985**, *26*, 77–88.
144. Duchene, D.; Wouessidjewe, D. *Pharm. Tech.* **1990**, *14*, 26–34.
145. Szejtli, J. *Pharm. Tech.* **1991**, *15*, 36–44.
146. Szejtli, J. *J. Inclus. Phenom.* **1992**, *14*, 25–36.
147. Irie, T.; Uekama, K. *J. Pharm. Sci.* **1997**, *86*, 147–162.
148. Pitha, J.; Milecki, H.; Fales, H.; Pannell, L.; Uekama, K. *Int. J. Pharm.* **1986**, *29*, 73–82.
149. Data from CyDex, L. C., Overland Park, Kansas, 66212.
150. Tinwalla, A.Y.; Hoesterey, B.L.; Xiang, T.; Lim, K.; Anderson, B.D. *Pharm. Res.* **1993**, *10*, 1136–1143.
151. Li, P.; Tabibi, S.E.; Yalkowsky, S.H. *J. Pharm. Sci.* **1998**, *87*, 1535–1537.
152. Li, P.; Zhao, L.; Yalkowsky, S.H. *J. Pharm. Sci.* **1998**, *88*, 1107–1111.

SPECTROSCOPIC METHODS OF ANALYSIS—ATOMIC ABSORPTION AND EMISSION SPECTROPHOTOMETRY

John P. Oberdier

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

INTRODUCTION

Often it is necessary to measure the metal content of different kinds of samples in the pharmaceutical industry. These samples can be actual products that have metals, such as calcium or magnesium added for their therapeutic value or biological samples, such as blood, urine, and tissue. More recently, work has been done on the role of metals in biological processes. In addition, products can be monitored for trace metal contamination, that can lead to unexpected degradation of product. A number of techniques are available for metal analyses. Some of the more common techniques are colorimetry, titrimetry, atomic absorption spectrophotometry (AA), fluorescence spectrophotometry, and emission spectrophotometry. The most common form of emission spectrophotometry is inductively coupled plasma (ICP). Thus, the analyst must choose the best approach for the sample being analyzed. The purpose of this article is to give a general overview of the strengths and limitations of atomic absorption spectrophotometry and ICP so that the best technique can be selected for the problem to be solved.

PRINCIPLE OF OPERATION FOR ATOMIC ABSORPTION SPECTROPHOTOMETRY (1)

The basis of atomic absorption spectrophotometry is the absorbance of light by the free, ground-state atoms of the element of interest. The ground state of an atom is the electronic state in which all the electrons are in their most stable configuration or orbitals. When light is absorbed by an atom, one or more of the electrons is excited to a higher energy orbital. The word free refers to the lack of any effects that would alter the amount of energy or the wavelength of the energy that is needed to cause the electrons in the atoms to shift from the ground state to an excited state(s).

Because each element absorbs at very discrete wavelengths, the lamp used for analysis of a particular metal emits light only at the desired wavelengths and is

specific for that element. The two kinds of lamps used in atomic absorption spectrophotometry are the hollow cathode lamp (HCL) and the electrodeless discharge lamp (EDL). These lamps contain a deposit of the metal of interest in an inert gas atmosphere and then use either a charged anode/cathode or a radio frequency to excite the metal atoms. As the excited atoms relax, they emit the absorbed energy as light energy at the characteristic wavelength(s) for that particular metal. This light energy is focused through the sample chamber onto the monochromator and detector as shown in Fig. 1.

The concentration of the analyte in the sample chamber is proportional to the amount of light absorbed and, under the proper experimental conditions, follows Beer's law. This relationship is defined as $A = abc$, where A is the absorbance, a is the absorption coefficient of the analyte, b is the cell path length, and c is the concentration of the analyte. The concentration of unknown samples can be readily determined by comparing the absorbance of a sample of known concentration(s) with the absorbance of the test sample.

SAMPLE INTRODUCTION TECHNIQUES AVAILABLE FOR ATOMIC ABSORPTION SPECTROPHOTOMETRY

One limitation of atomic absorption spectrophotometry is that the samples generally have to be in solution, preferably aqueous. Thus, either the sample must be directly soluble in a suitable solvent or some type of pretreatment, such as acid digestion, is necessary. One exception is that some instruments using a graphite furnace can be modified for direct injection of solids. Another limitation is that only one metal can be analyzed at a time. There are four primary methods of accomplishing this:

1. flame (aspiration of sample solution);
2. cold vapor (or chemical vaporization);
3. hydride generation (or chemical vaporization); and
4. Graphite furnace (or electrothermal atomization).

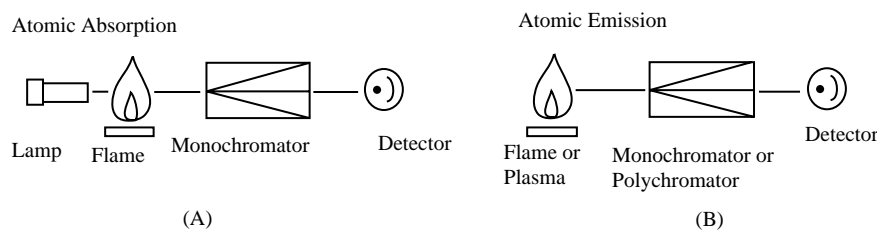


Fig. 1 Comparison of major components of atomic absorption and (A) emission spectrophotometers (B). (Courtesy of Perkin-Elmer Instruments).

All these methods have one common goal; to get the analyte atoms into the light path as free, ground-state atoms. Thus, there must be a balance of supplying enough energy so that the analyte atoms are free from association with other atoms (or molecules) in the sample matrix, versus not having excess energy, thereby allowing the electrons of the analyte atoms to remain in the ground state.

The technique of flame atomic absorption spectrophotometry accomplishes this by aspirating the sample solution into a burner chamber, where it is mixed with a fuel gas and an oxidant gas. The mixture is then burned in a specially designed burner head (Fig. 2). The light beam is directed lengthway down the burner, and the absorption of the analyte atoms in the flame is measured. The most commonly used gas mixtures are air with acetylene and nitrous oxide with acetylene. Experimental conditions are

well-defined in the literature, and “cookbook” conditions are available from most instrument manufacturers. In addition, many instruments are computer-controlled, and typical conditions are available directly on the operating screen. Figure 3 shows the type of information usually available. These conditions provide excellent starting points for new methods, and in many cases the manufacturers provide information on possible interferences and the linear range for each element.

Table 1 shows the detection limits of atomic absorption spectrophotometry for various metals. In general, flame atomic absorption spectrophotometry is quantitative in the lower parts-per-million levels and is readily automated for routine, high-volume samples. The other three techniques are used primarily for trace analysis and are quantitative to the lower parts-per-million levels for many elements.

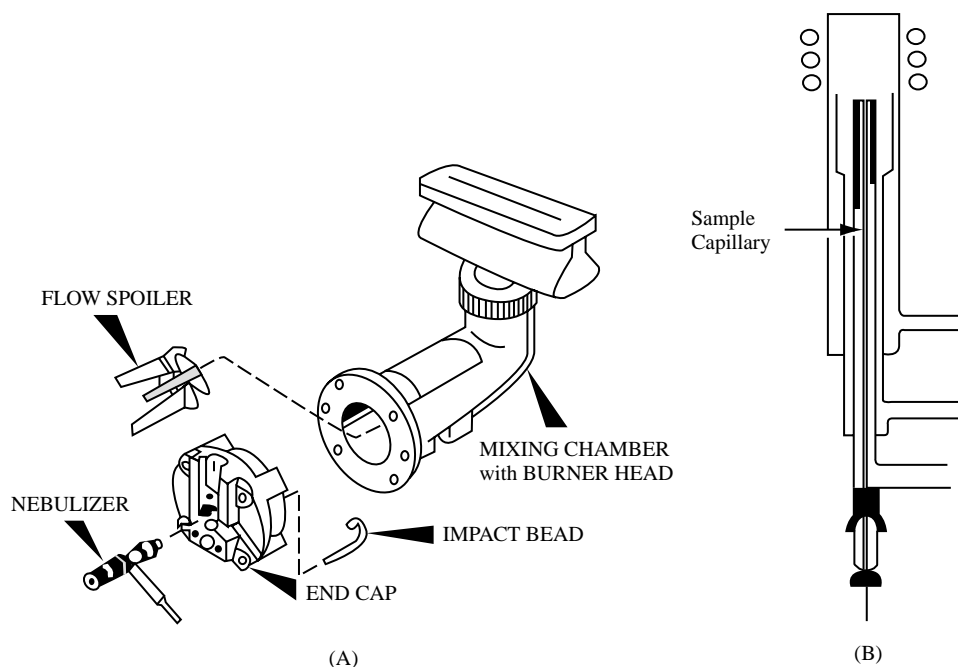


Fig. 2 Comparison of atomic absorption burner (A) and ICP argon plasma torch (B). (Courtesy of Perkin-Elmer Instruments.)

Recommended Conditions

Identification
Element: **Cu**

Atomic Absorption

Wavelength (nm)	Slit Width (nm)	Relative Noise	Char. Conc (mg/L)	Sensitivity Check (mg/L)	Linear to (mg/L)
324.8	0.7	1	0.077	4	5
327.4	0.7	1.1	0.17	8	5
216.5	0.2	7.2	0.47	20	20

Oxidant: Air Oxidant Flow (L/min): 10.0 Fuel Flow (L/min): 2.0

Emission
Wavelength (nm): 327.4 Slit (nm): 0.2
Oxidant: N2O Oxidant Flow (L/min): 6.0 Fuel Flow (L/min): 7.8

Data Remarks

Fig. 3 Typical operation screen for copper using flame atomic absorption spectrophotometry (Courtesy of Perkin-Elmer Instruments.)

The cold vapor technique is used for mercury. This technique involves reducing the mercury to the zero valence state with either sodium borohydride or stannous chloride. The mercury is then swept into a gas cell aligned in the light path of the spectrophotometer, using a stream of nitrogen or air. Fig. 4 shows a diagram of a typical unit.

A broad peak is obtained on the graph. Either the peak height or the integrated peak area is measured and compared with that obtained for a specific volume of a known concentration. Note that this technique measures the total mercury present in the sample aliquot. Thus, both the mercury concentration and the sample volume analyzed are important in determining the sensitivity of the assay. Generally, the sample volume is less than 50 ml; however, volumes up to 100 ml have been used. One potential problem with this technique is that the mercury must be in a form that is easily reduced. If there are complexing agents or organo mercury compounds that do not yield Hg atoms, these compounds must be treated before analysis. Frequently, acid digestion of the sample before analysis brings the sample to the desired state for analysis.

The technique of hydride generation is very similar to the cold vapor technique addressed previously. Seven metals (arsenic, bismuth, germanium, antimony, selenium, tin, tellurium) can be analyzed using this method. First, the volatile hydrides are formed by reacting the sample with sodium borohydride. The metal hydride is then swept into a heated gas cell using a stream of argon. Anything in the sample matrix that prevents the metal hydride from forming easily causes an interference. However, this technique is applicable to many complicated samples that

are very difficult or impossible to analyze using other techniques.

The final technique described for atomic absorption spectrophotometry is the graphite furnace. In this technique, a tubular, high-temperature furnace is placed in the light path of the spectrophotometer so that the light is focused down the center of the sample tube. The sample tube is a hollow cylindrical tube made of compressed graphite, with or without a pyrolytic graphite coating. The tube may also be equipped with a L'vov platform. One of the primary advances of this technique have been in furnace and sample tube design to improve the uniformity of heating. Uneven heating causes reproducibility problems and spattering or loss of sample. An atmosphere of inert gas is maintained in the sample chamber to prevent oxidation of the graphite tube during the subsequent heating steps. Typically 5–50 μ l of the sample are deposited on the bottom side of the tube or on the L'vov platform, and the tube is electrically heated to evaporate the sample solvent (100–200°C). If the sample contains organic material, the sample is usually ashed (600–1000°C). After this step, the inert gas flow is interrupted, and the metal is volatilized (1500–3000°C) and then measured. Matrix modifiers may be added to the sample to prevent loss of the analyte, before the volatilization step. All of these parameters are set up and optimized during initial method development. The preliminary settings are readily available in the literature and from most instrument manufacturers. Once the conditions have been chosen, the entire sequence can be automated, including use of an autosampler. An autosampler is especially recommended for this technique because one of the primary causes of irreproducibility is improper or inconsistent placement of the sample in the sample tube. Many autosamplers will add matrix modifiers and/or known amounts of standards to samples for standard addition-recovery experiments.

One of the inherent problems in the graphite furnace technique is background correction. When the metal atoms are volatilized, there may be absorption or loss of light that is not related to the metal atom concentration. For example, there may be small amounts of partially carbonized material that also volatilizes and causes light scattering. There are at least two methods available to correct for background effects: continuum correction and Zeeman correction. Continuum background correction assumes that absorption or loss of light, caused by effects other than atomic absorption, are constant over the entire spectral range. Thus, the sample signal is corrected by electronically subtracting the background absorption at a wavelength where the metal being measured does not absorb. This technique works well for background effects

Table 1 Typical detection limits (mg/L)

Element	Flame AA	Hg/Hydride	GFAA	ICP emission	ICP-MS	Element	Flame AA	Hg/Hydride	GFAA	ICP emission	ICP-MS
Ag	1.5		0.02	0.9	0.003	Mo	45		0.08	3	0.003
Al	45		0.1	3	0.006	Na	0.3		0.02	3	0.003 ^a
As	150	0.03	0.2	50	0.006	Nb	1500			10	0.0009
Au	9		0.15	8	0.001	Nd	1500			2	0.002
B	1000		20	0.8	0.09	Ni	6		0.3	5	0.005
Ba	15		0.35	0.09	0.002	Os	120			6	
Be	1.5		0.008	0.08	0.03	P	75000		130	30	0.3
Bi	30	0.03	0.25	30	0.0005	Pb	15		0.06	10	0.001
Br					0.2	Pd	30		0.8	3	0.003
C				75	150	Pr	7500			2	<0.0005
Ca	1.5		0.01	0.02	0.05 ^a	Pt	60		2.0	10	0.002
Cd	0.8		0.008	1	0.003	Rb	3		0.03	30	0.003
Ce				5	0.0004	Re	750			5	0.0006
Cl					10	Rh	6			5	0.0008
Co	9		0.15	1	0.0009	Ru	100		1.0	6	0.002
Cr	3		0.03	2	0.02	S				30	70
Cs	15				0.0005	Sb	45	0.15	0.15	10	0.001
Cu	1.5		0.1	0.4	0.003	Sc	30			0.2	0.02
Dy	50			2	0.001	Se	100	0.03	0.3	50	0.06
Er	60			1	0.0008	Si	90		1.0	3	0.7
Eu	30			0.2	0.0007	Sm	3000			2	0.001
F					10000	Sn	150		0.2	60	0.002
Fe	5		0.1	2	0.005 ^a	Sr	3		0.025	0.03	0.0008
Ga	75			4	0.001	Ta	1500			10	0.0006
Gd	1800			0.9	0.002	Tb	900			2	<0.0005
Ge	300			20	0.003	Te	30	0.03	0.4	10	0.01

(Continued)

Table 1 Typical detection limits (mg/L) (Continued)

Element	Flame AA	Hg/Hydride	GFAA	ICP emission	ICP-MS	Element	Flame AA	Hg/Hydride	GFAA	ICP emission	ICP-MS
Hf	300			4	0.0006	Th					<0.0005
Hg	300	0.009	0.6	1	0.004	Ti	75		0.35	0.4	0.006
Ho	60			0.4	<0.0005	Tl	15		0.15	30	0.0005
I					0.008	Tm	15			0.6	<0.0005
In	30				9	<0.0005	U	15000		15	<0.0005
Ir	900		3.0	5	0.0006	V	60		0.1	0.5	0.002
K	3		0.008	20	0.015 ^a	W	1500			8	0.001
La	3000			1	0.0005	Y	75			0.3	0.0009
Li	0.8		0.06	0.3	0.0001 ^a	Yb	8			0.3	0.001
Lu	1000			0.2	<0.0005	Zn	1.5		0.1	1	0.003
Mg	0.15		0.004	0.07	0.007	Zr	450			0.7	0.004
Mn	1.5		0.035	0.4	0.002						

^a Denotes that the ICP-MS detection limit was measured under cold plasma conditions.

All detection limits are given in micrograms per liter and were determined using elemental standards in dilute aqueous solution. All detection limits are based on a 98% confidence level (3 S.D.). All atomic absorption (Model 5100) detection limits were determined using instrumental parameters optimized for the individual element, including the use of system 2 electrodeless discharge lamps where available. ICP emission (Optima 3000) detection limits were obtained under simultaneous multielement conditions with a radial plasma. Detection limits using an axial plasma (Optima 3000 XL) are typically improved by 5–10 times.

Cold vapor mercury detection limits were determined with a FIASTM-100 or FIAS-400 flow-injection system with amalgamation accessory. The detection limit without an amalgamation accessory is 0.2 µg/L with a hollow cathode lamp, 0.05 µg/L with a System 2 electrodeless discharge lamp. (The Hg detection limit with the dedicated FIMSTM-100 or FIMS-400 mercury analyzers is <0.010 µg/L without an amalgamation accessory and <0.001 µg/L with an amalgamation accessory.) Hydride detection limits shown were determined using an MHS-10 Mercury/Hydride system. Graphite furnace AA detection limits were determined using 50-µl sample volumes, a L'vov platform, and full STPF conditions (Model 5100 PC with 5100 ZL Zeeman Furnace Module or Model 4110 ZL). SIMAA 6000 detection limits are similar in its multisource and are typically two to five times better in its single-source mode. Graphite furnace detection limits can be enhanced further by the use of replicate injections.

ICP-MS detection limits were determined using a 3-S integration.

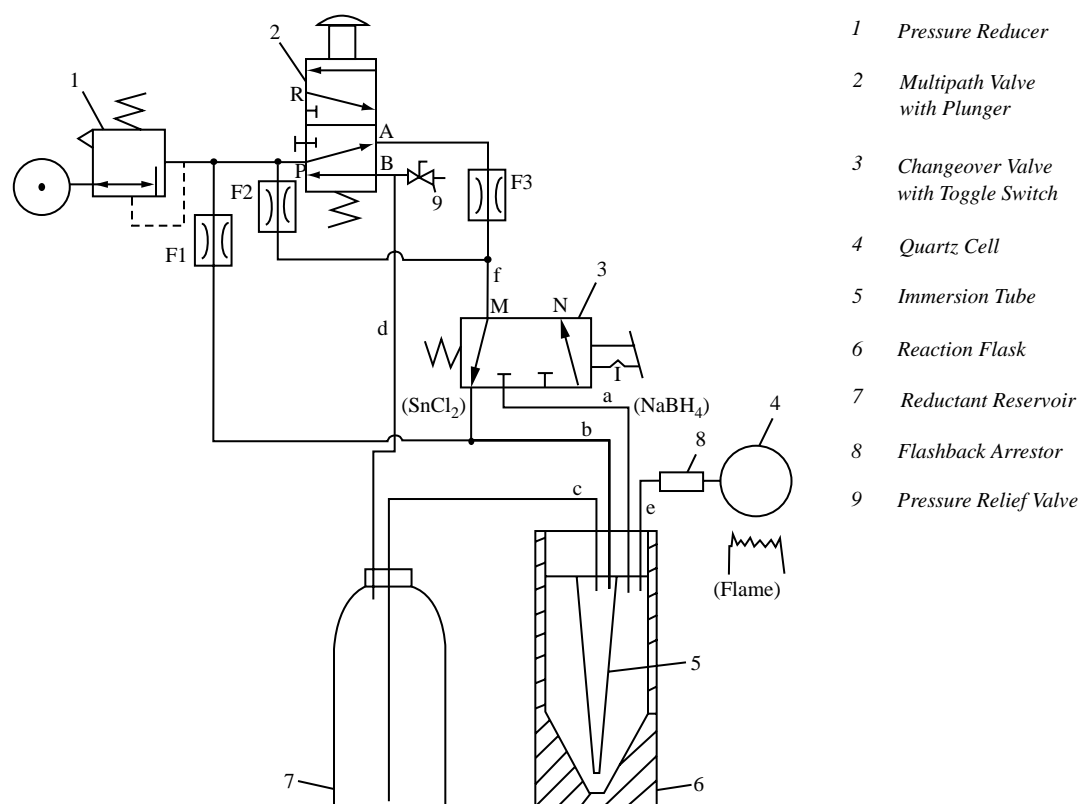


Fig. 4 Diagram of a typical unit for vapor or hydride generation technique. (Courtesy of Perkin-Elmer Instruments.)

such as light scattering that are fairly continuous over the entire spectral range. However, many background effects are not continuous, and the analyte signal is either under- or overcorrected. Zeeman background correction relies on the Zeeman effect. In simplest terms, Professor Zeeman discovered in 1896 that when an atomic line spectrum was subjected to a magnetic field, the lines were split into multiple components. Thus, the corrected sample signal can be obtained by electronically subtracting the sample absorbance in the presence and absence of a magnetic field. Because the background is actually being measured at the analyte wavelength, the measurement is generally less susceptible to error.

COMMON SOURCES OF ANALYTICAL ERROR IN ATOMIC ABSORPTION SPECTROPHOTOMETRY

The two common sources of analytical error are interference during the analytical measurement and contamination during sample preparation. During

the analytical measurement, potential interferences include, matrix effects: chemical interferences, ionization interferences, and spectral interferences.

Matrix effects usually involve physical parameters, which differ between samples and standards. For example, in flame atomic absorption spectrophotometry, if the viscosity of the sample and standard solutions is significantly different, the solutions are aspirated into the burner chamber at different rates and the responses are different. Viscosity effects can often be eliminated by adding 1–2% of a metallic salt to the analyte solutions. Chemical interference can include both complexes and organically bound metals. When these type of interference are encountered, the samples often have to be digested with acid to destroy the organic material. For chemical complexes, adding a large excess of a competing ion is an alternative to acid digestion. Ionization interference is very common for sodium, potassium, and lithium. To overcome this problem, a large excess of one of the other elements is added. For example, 1–2% lithium might be added to a sample being analyzed for potassium. Ionized potassium atoms can collide with the large excess of lithium atoms, resulting in the loss of energy for the potassium atoms. These atoms will return to the ground state

and be measured. Finally, there may be spectroscopic interference such as molecular or atomic absorption at the analyte wavelength. In practice, this interference can normally be overcome, and very complex samples can be analyzed rapidly and accurately, provided the analyst recognizes the potential interference and takes appropriate steps to minimize it.

Contamination during sample preparation is probably the single major source of error in this technique and is especially important when analyzing for metals at the parts-per-million level. Contamination can occur because of improperly cleaned containers or the wrong type of container (such as glass versus plastic). In addition, contamination can come from reagents used to prepare the sample. Special grade acids or matrix modifiers are usually necessary for trace metal analysis. Another source of contamination is the general environment. For example, even a dust particle can cause errors of 10-fold or more in the analysis of aluminum. In summary, serious thought must be given to every material and reagent the sample will contact during the entire sample-handling process. A series of review articles (2–6) are listed in the references for additional reading on atomic absorption spectrophotometry.

PRINCIPLE OF OPERATION AND TECHNIQUES USED IN EMISSION SPECTROPHOTOMETRY (7)

Emission Spectrophotometry is a complementary technique to atomic absorption. In this technique, the electrons in the atoms are excited by electrical or thermal energy to an unstable energy state. When the atom returns to a more stable state or the ground state, energy as light is emitted. This emitted light is detected and quantitated in emission spectrophotometry. The wavelength(s) of the emitted light is specific to the element(s) that are present. Figure 1 compares the major components of atomic absorption and emission spectrophotometry.

This emission of light from elements was reported as early as the mid-1700s and was later the basis of discovering four elements by Kirchoff and Bunsen in the mid-1800s. However, it was not until the 20th century that the technique really became a useful analytical tool. The first analytical instruments used a spark discharge to provide the energy to excite the atoms so that they would emit light. This approach gave valuable qualitative and semiquantitative information but did not have the accuracy and precision needed for many applications. The one exception was the analysis of alkali earth metals, which require very little energy to raise the electrons to an

excited state. A simple propane flame is sufficient. This led to the development of a clinical analyzer for sodium and potassium in blood and urine samples that is still used routinely in clinical laboratories. It was not until the argon plasma flame was developed that a reliable source of high energy was available to excite a wide variety of atoms so that they would emit light. This led to the development of the technique commonly referred to as inductively coupled plasma (ICP) spectrophotometry.

As noted previously, the argon plasma flame provided the reliable energy source needed for ICP to become a reality. The plasma is created by passing argon gas between two concentric quartz tubes while applying an RF field (Fig. 5). This plasma flame reaches temperatures up to 10,000°K, with the sample reaching temperatures of 5500–8000°K. The shape of the flame confines the sample emission to an optically thin source resulting in a wide dynamic range for ICP. The sample is aspirated into the argon plasma flame, where it is believed to go through several steps: desolvation, vaporization, atomization, and finally excitation or ionization. As the different atoms relax to a lower energy state, the emitted light strikes a monochromator or polychromator, and the intensity of the line spectra for each element is measured at the appropriate wavelength(s) for that element.

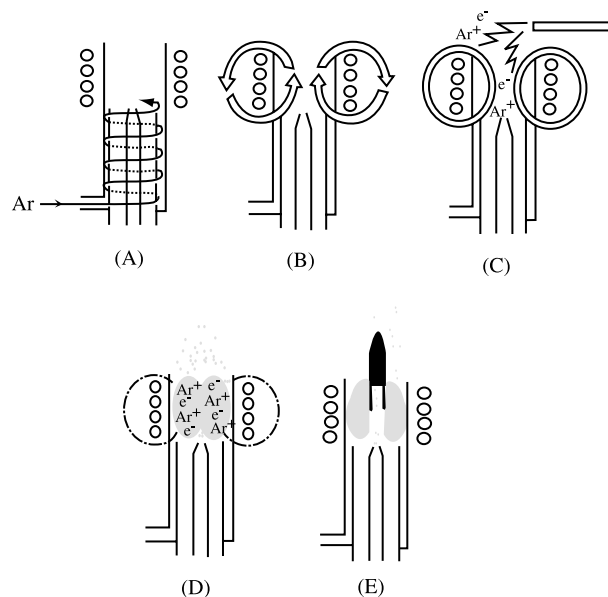


Fig. 5 Cross section of an ICP torch and load coil depicting an ignition sequence. (A) Argon gas is swirled through the torch; (B) RF power is applied to the load coil; (C) a spark produces some free electrons in the argon; (D) the free electrons are accelerated by the RF fields, causing further ionization and forming a plasma; (E) the sample aerosol-carrying nebulizer flow punches a hole in the plasma. (Courtesy of Perkin-Elmer Instruments.)

The most commonly used technique of sample introduction is aspiration of the solution into the argon plasma flame. Because of the high temperatures in the flame, many of the problems associated with atomic absorption are eliminated. However, matrix effects such as significant differences in viscosity between sample and standard solutions can still have an effect. When needed, most of the techniques of sample introduction used in atomic absorption spectrophotometry can also be used for sample introduction in emission spectrophotometry [see the review articles (8–13) listed in the references].

HOW TO CHOOSE AMONG THE TECHNIQUES

The major advantage of ICP is that multiple elements can be analyzed at the same time. The rule of thumb is if one typically assays less than four to six metals in a sample, atomic absorption is the better route. However, for more than six metals, ICP is clearly the choice. However, other factors such as sensitivity and interferences also have to be considered. One other variation of ICP that is becoming important is the hyphenated technique of ICP–MS (mass spectrophotometry). This combination gives excellent sensitivity and can even differentiate ionic species or isotopes if needed. Table 1 summarizes the sensitivities of the various techniques. As noted previously, all these techniques are complimentary, and the various strengths and weaknesses must be understood to choose the best technique.

THE FUTURE

The instrumentation for atomic absorption spectrophotometry is very well-defined and can range from a relatively simple manually operated instrument to a completely automated system that is on line to a central database. In addition, accurate results can be obtained on a wide range of samples. The future lies in using this technique to solve problems rather than to further develop instrumentation. For example, the FDA published new guidelines for the aluminum content of products used in total parenteral nutrition.

The role of ICP in the future will continue to expand, especially with the development of ICP–MS. The role of metals in biological processes and disease states is just beginning to be understood, with numerous studies being published. For multivalent metals, the exact species of metal involved in the biological process can be determined with ICP–MS. Understanding this type of detail will

hopefully lead to important breakthroughs in the future. See the References (14–17) for additional information on future applications.

REFERENCES

1. Beatty, R.D.; Kerber, J.D. *Concepts, Instrumentation, and Techniques in Atomic Absorption Spectrophotometry*; Perkin Elmer: Minneapolis, KS, 1993.
2. Komaromy-Hiller, G. Flame, Flameless, and Plasma Spectroscopy. *Anal. Chem.* **1999**, *71* (12), 338R–342R.
3. Lagalante, A.F. Atomic Absorption Spectroscopy. A Tutorial Review. *Appl. Spectrosc. Rev.* **1999**, *34* (3), 173–189.
4. Hill, S.J.; Dawson, J.B.; Price, W.J.; Shuttler, I.L.; Smith, C.M.M.; Tyson, J.F. Advances in Atomic Absorption and Fluorescence Spectrometry and Related Techniques. *J. Anal. At. Spectrom.* **1999**, *14* (8), 1245–1285.
5. Sturgeon, R.E. Graphite Furnace Atomic Absorption Spectrometry and Environmental Challenges at the Ultratrace Level—A Review *Spectrochim. Acta, Part B* **1997**, *52B* (9, 10), 1451–1457.
6. Harnly, J.M. The Future of Atomic Absorption Spectrometry: A Continuum Source with a Charge Coupled Array Detector. *J. Anal. At. Spectrom.* **1999**, *14* (2), 137–146.
7. Boss, C.B.; Fredeen, K.J. *Concepts, Instrumentation, and Techniques in Inductively Coupled Plasma Optical Emission Spectrometry*, 2nd Ed.; Perkin Elmer: New York, 1999.
8. Suzuki, K.T. Hyphenated Techniques as Tools for Speciating Biological Metals: Metallothionein and Metal-binding Proteins. *Analisis* **1998**, *26* (6), M57–M61.
9. Sutton, K.L.; Caruso, J.A. Liquid Chromatography-Inductively Coupled Plasma Mass Spectrometry. *Chromatogr., A* **1999**, *856*, 243–258.
10. Horlick, G.; Montaser, A. Analytical Characteristics of ICPMS Induct. Coupled Plasma Mass Spectrom. **1998**, 503–613.
11. Taylor, A.; Branch, S.; Halls, D.J.; Owen, L.M.W.; White, M. Atomic Spectrometry Update: Clinical and Biological Materials, Foods and Beverages. *J. Anal. At. Spectrom.* **1999**, *14* (4), 717–781.
12. Pruszkowski, E.; Neubauer, K.; Thomas, R. An Overview of Clinical Applications by Inductively Coupled Plasma Mass Spectrometry. *At. Spectrosc.* **1998**, *19* (4), 111–115.
13. Lobinski, R.; Potin-Gautier, M. Metals and Biomolecules—Bioinorganic Analytical Chemistry. *Analysis* **1998**, *26* (6), M21–M24.
14. Taylor, H.E.; Huff, R.A.; Montaser, A. Novel Applications of ICPMS Induct. Coupled Plasma Mass Spectrom. **1998**, 681–807.
15. Welz, B. Speciation Analysis. The Future of Atomic Absorption Spectrometry. *J. Anal. At. Spectrom.* **1998**, *13* (5), 413–417.
16. Vanhaecke, F.; Moens, L. Recent Trends in Trace Element Determination and Speciation Using Inductively Coupled Plasma Mass Spectrometry. *J. Anal. Chem.* **1999**, *364* (5), 440–451.
17. Durrant, S.F. Laser Ablation Inductively Coupled Plasma Mass Spectrometry: Achievements, Problems, Prospects. *J. Anal. At. Spectrom.* **1999**, *14* (9), 1385–1403.

SPECTROSCOPIC METHODS OF ANALYSIS—DIFFUSE REFLECTANCE SPECTROSCOPY

Herbert Michael Heise

Institute of Spectrochemistry and Applied Spectroscopy, Dortmund, Germany

INTRODUCTION

In the field of molecular spectroscopy, absorption measurements with electromagnetic radiation covering the wavelengths between 200 nm and 20 μ m can be used for analytical applications. By convention, tags such as ultraviolet (UV), visible (VIS), and infrared (IR) were given to special intervals of the spectrum, each yielding different specific information on organic and inorganic substances. The energy for electronic transitions between outer molecular orbitals corresponds to radiation of the first two spectral regions noted, whereas vibrational spectroscopy is connected to the infrared. Here, it can be further distinguished between near- and mid-infrared spectral subregions, which are important for routine analytical work.

In the classic spectroscopy experiment, transmission measurements are performed to analyze the radiation absorption of samples. Today, an additional powerful measurement technique is diffuse reflectance spectroscopy, which can be applied advantageously to scattering samples such as powders and other dispersed systems. It often requires less sample preparation than for a spectrum recorded in transmission. After sample penetration, the radiation is diffusely scattered and partially absorbed before part emerges back at the surface, from where it is detected using various optics and detectors. Beside identification of compounds, quantitative assays for active ingredients and fillers and in general formulation testing are available that often require special chemometric tools. Trace analysis and the study of adsorbed chemicals are additional fields. The combination with separation techniques, e.g., thin-layer chromatography, has been established, where diffuse reflectance measurement techniques are involved.

INSTRUMENTAL ASPECTS OF THE MEASUREMENT TECHNIQUE

Molecular spectroscopy plays an important role in the characterization of pharmaceutical substances (1). For

chemical quality and process control, near-infrared spectroscopy has received much attention in the last years. The pharmaceutical industry was forced to develop fast measurement equipment and techniques for the identification of raw materials on receipt as well as to verify composition of pharmaceutical formulations before the products leave the premises. Some techniques are suited for process analysis because of their speed and lack of sample preparation. Further applications are concerned with the analysis in the laboratory during synthesis of compounds, identification of constituents of competitive products, and determination of additives and coatings. Additional cases can be discussed for which low detection limits are required. For best performance and high precision, instrumentation and sample preparation must be adapted to meet such demanding applications.

Instruments and Accessories

The spectral ranges noted above usually cannot be covered by a single instrument. Spectrometers with a grating monochromator are versatile instruments for recording UV/VIS and near-infrared spectra, which are equipped, depending on the spectral range to be measured, on different detectors [commonly used are silicon detectors: 200–1100 nm, indium gallium arsenide (InGaAs): 800–1700 nm (extended wavelength options are available), and lead sulfide: 1100–2500 nm]. Beside scanning instruments, also diode array spectrometers exist to measure the whole spectrum simultaneously. Other measurement technology uses a set of optical filters; such instruments are widespread in assays of food and agricultural products. A special technology that has been implemented in fast near-infrared spectrometers are acousto-optical tunable filters. Radiation sources such as tungsten halogen lamps, which provide a broad emission spectrum of electromagnetic radiation, may be replaced by light-emitting diodes or tunable lasers, broadening the applicability of future spectrometer systems.

The infrared spectral range is usually served by Fourier transform spectrometers, which contain an interferometer for wavelength dispersion. Today, the spectra are

immediately calculated by using Fourier transformation of the measured interferograms. Such instrumentation can provide better spectral quality in shorter time than monochromator based systems. For the mid-infrared, thermal detectors such as those from deuterated triglycine sulfate (DTGS) or, for fast and sensitive measurements, liquid nitrogen cooled photodetectors such as those from mercury cadmium telluride (MCT) are used, whereas for the near-infrared InGaAs, germanium or indium antimonide (InSb) photodetectors are used. An overview on recent instrumentation is given by Coates (2). It is usual practice to display the infrared spectra with wavenumbers as abscissa (the transformation from wavelength λ , for example in micrometers, to wavenumbers $\tilde{\nu}$, in cm^{-1} , is given by $\tilde{\nu} = 10,000/\lambda$).

Various accessories were designed for recording diffuse reflectance spectra. Apart from special devices developed by different groups and described in the literature, several commercially available types must be noted. A few accessories are shown schematically in Fig. 1, which are representative of the diversity of optics. For a long time, integrating spheres have been in use, in particular for UV/VIS and near-infrared spectroscopy, although a few applications with sensitive MCT detectors also can be found within the mid-infrared (3). Usually, a baffle is placed within the sphere to prevent “first strike” radiation from immediately hitting the detector, which measures the scatter over all angles [hemispherical collection; see Fig. 1(a)].

As an alternative, reflection optics based on ellipsoidal mirrors or segments of such type became widespread. Several commercially available types can be noted, one is typified by Spectra-Tech’s Collector [Fig. 1(b)]. To reduce the specular reflection, a blocker can be applied, which is placed vertically on the sample surface, perpendicular to the optical axis. Another accessory is Harrick’s so-called praying mantis (the name was coined after its appearance), which uses an out-of-plane optical geometry to avoid the Fresnel reflection from the sample surface. The device shown in Fig. 1(c) was designed by us to improve the collection of diffusely back-reflected photons, observing the figures of merit for the case that irradiation and detection optics were to be optimized for optical throughput. Another positive feature of such a design is that small spots on large samples can be studied; for applications, see Heise (4). A critical evaluation of different types of diffuse reflectance accessories, especially for discrimination of specular and diffusely reflected radiation, has been given by Yang et al. (5). For a critical test using a high-absorbing inorganic sample—within an inert low-absorbing matrix of KCl powder, it was found that the specular component within the measured

spectrum was primarily sample rather than accessory-dependent.

Additional accessories are constructed from optical fibers. One affordable material suited for the UV/VIS and near-infrared is quartz, whereby a special quality of low-OH grade quartz provides better transmission, especially at longer wavelengths. Compared with mirror-based systems, the acceptance cone for radiation-gathering is reduced owing to the core/cladding configuration of the fiber light-guides [see Fig. 1(d)]. Bifurcated fiber bundles are often used as accessories, whereby the radiation-delivering fibers are mixed at random with the second fiber bundle, by which the reflected radiation is guided to the detector. The flexibility of such fiber optic probes is enormous and also allows measurements from powdered samples, for example, in opened casks.

The diffuse reflectance of a sample is usually measured against a nearly 100%-reflecting material, analogously to a transmittance measurement [a splendid coverage of standards for reflectance measurements is provided by Springsteen (6)]. Traditionally, freshly prepared MgO or BaSO₄ was used for coating the inner surface of an integrating sphere. For the spectral range of 200–2500 nm SpectralonTM (Labsphere, North Sutton, NH) is now the preferred material because it is inert and easy to handle and process. It is a white thermoplastic resin that exhibits a reflectance larger than 0.99 over the spectral range of 400–1500 nm and larger than 0.95 for wavelengths up to 2500 nm. For the mid-infrared, powdered potassium bromide is usually available or gold-coated sandpaper of different grades, which we preferred. A commercially available standard material, obtainable from the same company as above, is InfraGold, which is a multilayer metallic on a mechanically roughened substrate. It exhibits a reflectance between 0.94 and 0.96 through the spectral ranges, with wavelengths between 1 and 25 μm . For dilution of samples, additional materials others than alkali halides have been proposed, which is presented below.

THEORETICAL CONSIDERATIONS

For spectroscopic studies, different measurement techniques are available. Radiation interacting with the sample may be absorbed, transmitted, or reflected. Usually, Lambert–Beer’s law can be applied for absorbing samples showing no scattering. Reflection at a sample surface can be specular, in particular for a plane geometry as obtained for polished surfaces (also called Fresnel reflection), or diffuse, especially for rough surfaces as found in powders.

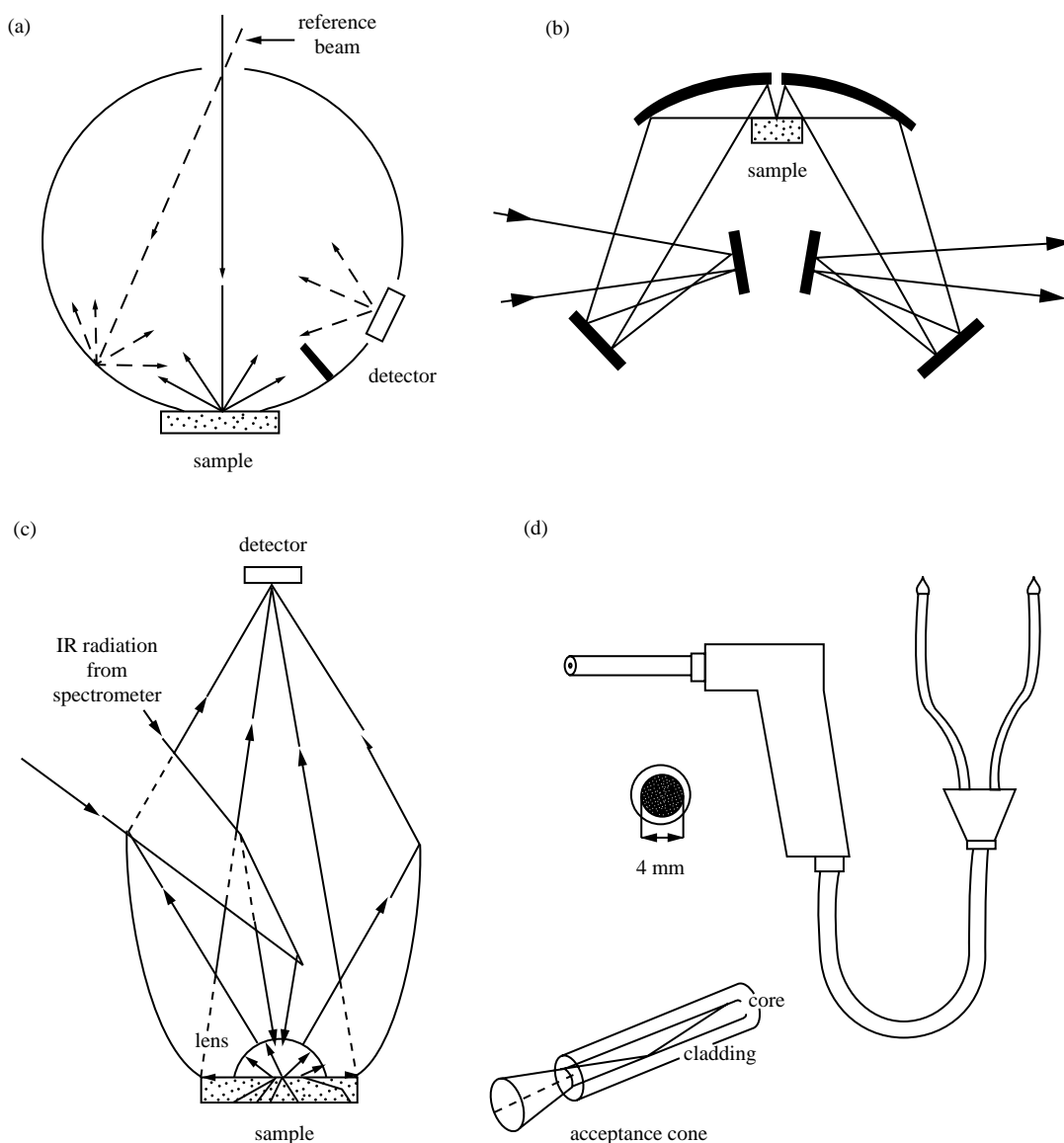


Fig. 1 Accessories for diffuse reflectance spectroscopy: (a) Integrating sphere with hemispherical radiation collection; (b) Accessory based on ellipsoidal mirrors, used within the sample compartment of the spectrometer; (c) Rotational ellipsoidal mirror device with dedicated detector; (d) Bifurcated fiber optic-based accessory (also shown is the random mixture of fibers for illumination and detection; compared with devices based on reflection optics the acceptance cone for radiation delivery and collection is limited and depends on the refractive indices of the core and cladding material).

Many reflections are a combination of both diffuse and specular components. Whereas specular reflectance can be rigorously treated using the Fresnel equations, more complex processes are responsible for the diffuse reflection inside a scattering system. With dispersed systems, further phenomena of interaction between radiation and materials must be addressed. When radiation penetrates into a dispersed sample, the impinging radiation is partly absorbed and partly scattered by the particles so that

radiation is also reflected in a diffuse manner. This diffusely reflected radiation is useful for analytical applications and, with special functions for linearization of the diffuse reflectance signal, quantitative relations between transformed signal and concentration can be found.

As several researchers have shown empirically, the use of $-\log(\text{reflectance})$ can provide, analogous to a transmittance measurement, a linear relationship between

the transformed reflectance and concentration, if the matrix is not strongly absorbing as can be found for many samples studied by near-infrared spectroscopy (7). This issue is presented in detail below. A different approach based on a physical model was considered for UV/VIS measurements and later also applied within the mid-infrared. A theory was derived by Kubelka and Munk (8) for a simple, one-dimensional, two-flux model, although it must be noted that Arthur Schuster (1905) had already come up with a reflectance function for isotropic scattering. A detailed description of theoretical and practical aspects was given by Kortüm (9). The optical absorption and scattering characteristics are described by its absorption and scattering coefficients k and s , respectively. For diffuse reflectance spectra, the transformation of the reflectance by the so-called Kubelka–Munk (K–M) function can provide a linear relationship for absorption band intensities versus concentrations (the units are often referred to as K–M units):

$$F(R) = (1 - R_\infty)^2(2R_\infty) = k/s$$

R_∞ is the reflectance of an infinitely thick sample (in the near-infrared, this means an approximate 5-mm thickness and more). The theory was recently revisited by Loyalka and Riggs (10), who reinvestigated the accuracy of the Kubelka–Munk equations. They found that the coefficient k must be replaced by $k = 2a$ with the absorption coefficient $a = \ln(10) \varepsilon c$, as derivable from Beer's law; for the latter equation: $\ln(10) = 2.303$, ε the molar absorptivity, and c the molar concentration. Such a dependency for k was stated earlier by other researchers when comparing more refined radiation transport theories for biomedical applications, e.g., Ref. (11).

It must be kept in mind that the K–M theory does little more than provide an empirical model useful for quantitative work. To derive at the above equation, several assumptions were necessary such as infinite lateral extension and optical thickness, isotropy of the sample, and diffuse illumination, of which in particular the last is not met in practice when a directional sample illumination is prevailing. For isotropically scattering samples, a so-called three-flux approximation to the near-infrared radiative transfer within pharmaceutical powders and mixtures was presented recently for the derivation of scattering and absorption coefficients (12). Earlier, Hecht had studied different approaches and compared modeling results with measurements primarily in the visible (13). When the scattering inside the medium shows anisotropy as, for example, with forward scattering, more sophisticated modeling of the radiation transport inside the scattering medium is required. One tool applied recently

for spectrum prediction is Monte Carlo simulations, by which statistical averages are obtained for a large population of photons using probability functions for the scattering events. In particular, such an approach is often used for biomedical studies to analyze the photon distribution in tissues after irradiation.

Most spectroscopic software can convert spectra to K–M units on demand. With a constant scattering coefficient, at least within a limited spectral range, the transformed spectra resemble those obtained by conventional transmission spectroscopy. The relative error for K–M unit-based intensities is minimal for reflectance values between 0.2 and 0.7, which is a similar statement as obtainable for absorbance values derived from transmittance measurements and their error estimate. Studies with concern of a varying change in the reflectance of the reference material and its influence on the sample K–M units were carried out by Krivácsy and Hlavay (14).

PRACTICAL ASPECTS

The measurement of diffuse reflectance spectra from powdered samples requires less sample preparation than for the recording of transmittance spectra from solids, for which the embedding of such samples in a transparent medium is usually carried out. For mid-infrared measurements the KBr pellet technique is routinely used in the laboratory, which requires the application of high pressures to get a transparent tablet. The limitations are that the inertness of this embedding material is sometimes not given with the result of unwanted reactions. Scattering within the pellet can still exist, which is evident from evolving baseline shifts. The absorptivities of vibrational combination and overtone bands in the near-infrared are weaker than those of related bands in the mid-infrared, where many intensive bands can be assigned to fundamental vibrations. This leads to the consequence that the amount of substances required for achieving significant absorbances is much larger than for KBr pellet measurements in the mid-infrared.

In Fig. 2, a comparison of two spectra of anhydrous glucose is shown. The upper trace was recorded in transmission. For this, the pellet was placed in front of the spectrometer's detector to reduce the attenuation effects by scattering from inhomogeneities inside the pellet, because placing the pellet into the conventional sample compartment did not allow a satisfactory spectral signal-to-noise ratio. The lower spectrum was obtained using the diffuse reflectance technique on the powder, which was placed in a cup and slightly compressed by a spatulum,

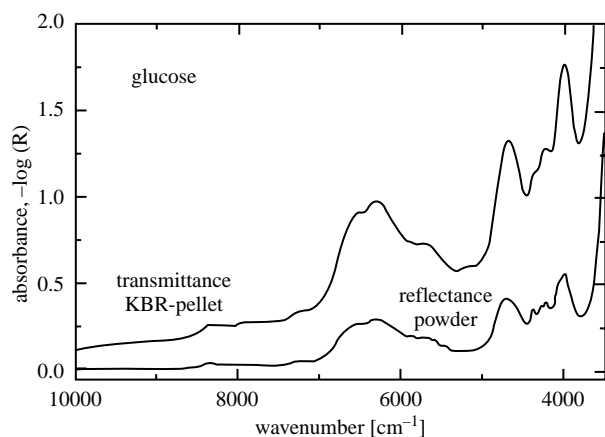


Fig. 2 Comparison of near-infrared spectra of crystalline anhydrous glucose. Spectrum measured in transmittance using the KBr pellet technique [spectral resolution 64 cm^{-1} ; intensity data in absorbance (top trace) and diffuse reflectance spectrum using a spectral resolution of 32 cm^{-1} and intensity data transformed to $-\log(R)$ (bottom trace)].

illustrating the ease of sample preparation. Slight differences in the spectroscopic features arise only from the reduced spectral resolution, as used for the transmittance spectrum shown here.

The high-information density of mid-infrared spectra is well-known. In comparison, the discrimination power of near-infrared spectroscopy is illustrated in Fig. 3(a) for different monosaccharides, for which diffuse reflectance spectra were recorded. Fig. 3(b) shows the near-IR spectra of two pharmaceutical substances. In Fig. 4, the influence of finer particles, obtained by grinding the crystal material from the supplier pack on the spectrum is presented. With coarser particles, the band absorbance intensities are enlarged owing to a decrease in scattering, which means that radiation can penetrate deeper into the sample. In addition, a nearly constant attenuation can be observed, which leads just to a positive offset in the $-\log(R)$ spectrum. However, there are more severe distortions in a K–M transformed spectrum owing to baseline shifts, as pointed out by Griffiths (7).

The spectra of undiluted substances are influenced by effects from Fresnel reflection owing to strong absorption bands found within the mid-infrared spectral range. Additionally, the contrast between band intensities of strong and weak absorption bands is dramatically reduced. An example is provided in Fig. 5, which compares a KBr pellet spectrum of caffeine with that obtained from a powdered undiluted sample by using the diffuse reflectance technique. Two different representations were chosen, one with a transformation into K–M units and a

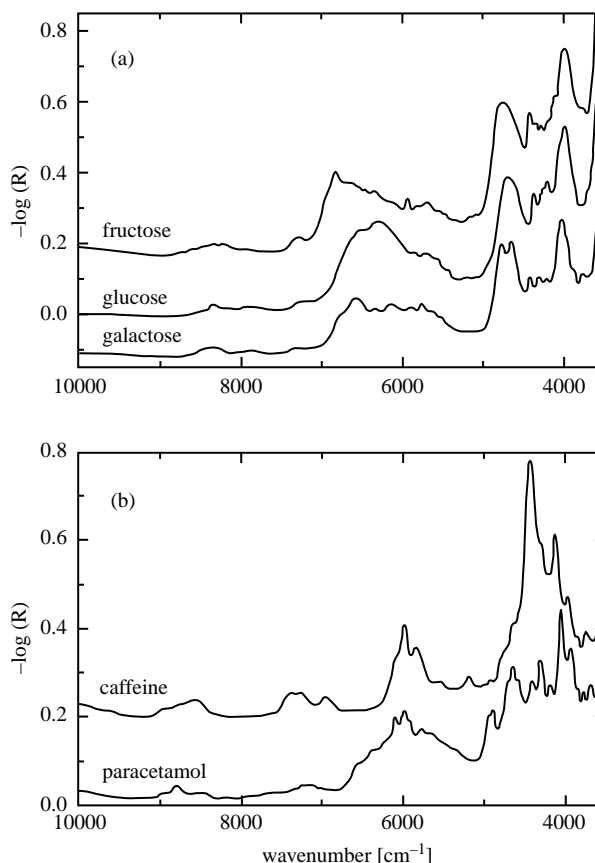


Fig. 3 Diffuse reflectance near-infrared spectra of several pharmaceutical compounds: (a) Comparison of the spectra of three different monosaccharides measured as crystalline powders; (for clarity, the spectra of fructose and galactose are offset; spectral resolution 32 cm^{-1}); (b) Comparison of spectra of pure caffeine and paracetamol (offset, spectral resolution 32 cm^{-1}).

second with $-\log(R)$. The latter shows uniquely that by using the reflectance technique, a tremendous amplification of weak absorption bands can be achieved. The distribution of intensity ratios is closer to that of the transmittance spectrum, when a dilution of the caffeine by KBr powder, for example, by a factor of 100, is used. A close-up of the mid-infrared fingerprint region in Fig. 6 provides us with the clues as to why a dilution by embedding the compound into the KBr powder must be recommended to obtain a spectrum by diffuse reflectance, which is comparable with the KBr pellet absorbance spectrum.

For strong absorption bands, a large Fresnel component shows up within the diffuse reflectance spectrum, leading to severe band distortions, which are also unmodeled by the K–M theory; the undiluted substance spectrum, in this

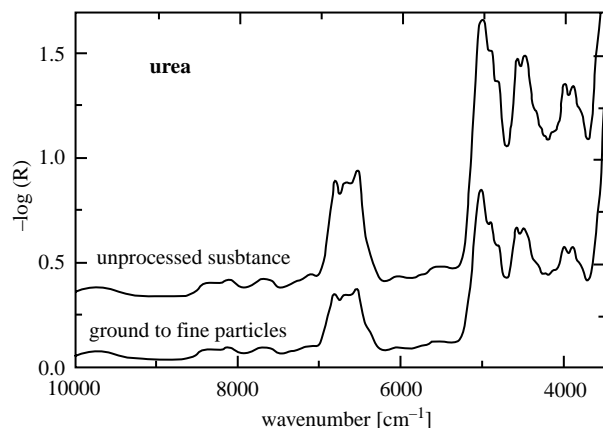


Fig. 4 Diffuse reflectance spectra of crystalline urea illustrating the effect of particle size (for the top trace, the crystalline material was unprocessed from the supplier's container), whereas for the bottom spectrum, the substance was ground to finer particles.

example of pure caffeine, shows significant shifts and band shape changes that cause certainly difficulties when comparing such a spectrum with those from spectral libraries, built from absorptivity data that were obtained from standard transmission or similar attenuated total reflection measurements. With a high-absorption index κ , which is related to the absorption coefficient by $a = \ln(10) (4\pi/\lambda)\kappa$, with λ the wavelength of the impinging radiation, the Fresnel reflection can reach unity. In this case, the band is also called the Reststrahlen band. There are applications to use the effect, e.g., for narrow optical reflection band pass filters.

Similar cases with high absorptivities can be found in the UV/VIS region, in which fundamental electronic absorptions exist. When particle size increases, the proportion caused by specular reflectance may become stronger and readily evident in the diffuse reflectance spectrum. A discussion of previous UV/VIS work is given by Wendlandt and Hecht (15).

In the past, extensive investigations were made to obtain better insight into the limitations of the diffuse reflectance measurement technique. Studies demonstrated that sample properties such as particle size and packing affect, in addition to the optical constants, the diffuse reflectance spectrum (9, 15). The characteristics of the diffuse and specular components were studied for different particle sizes and dilution within a nonabsorbing inert matrix. It was found that specularly and diffusely reflected radiation coexist in the measured diffuse reflectance spectrum, even in KCl-diluted samples. In addition, the specular component, which is certainly sample-dependent, is not necessarily the same as from the front-surface

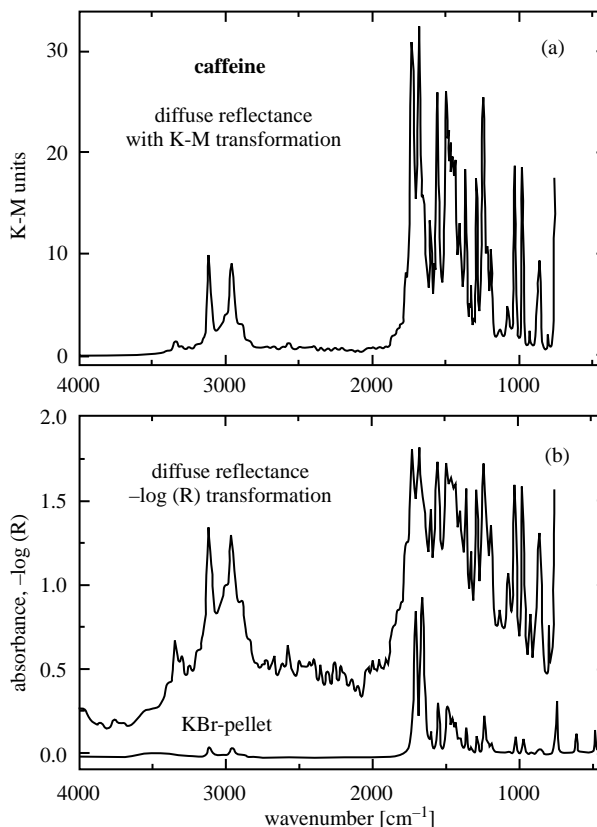


Fig. 5 Comparison of mid-infrared spectra of caffeine obtained by diffuse reflectance and transmission spectroscopy. (a) Diffuse reflectance spectrum of the pure powdered substance with transformed intensity data in K-M units. (b) Same diffuse reflectance spectrum, but using $-\log(R)$ transformation (top trace), the lower spectral range was limited by the cut-off of the MCT detector used; the bottom trace shows a transmission spectrum using the conventional KBr pellet technique transformed into absorbance, i.e., $-\log(\text{transmittance})$.

reflection (16). To prove this, a top layer of pure KCl powder was repeatedly put on top the sample to prevent front-surface reflectance. Another study to be noted is by Brimmer et al. (17) who also discussed the effect of Fresnel reflection with accessories of different optical geometries on the diffuse reflectance spectra.

The penetration depth of the probing IR-radiation is determined by both the scattering and the absorbing characteristics of the sample. Theoretical calculations were carried out by Fraser and Griffiths (18), showing that for strong and medium absorptivities, the criterion of an infinitely thick sample layer is valid for 1 mm, whereas for weaker bands, approximately about 5 mm of sample thickness was claimed. For pure KCl, under the conditions of a defined particle size distribution for the powder and the application of a certain pressure to obtain a

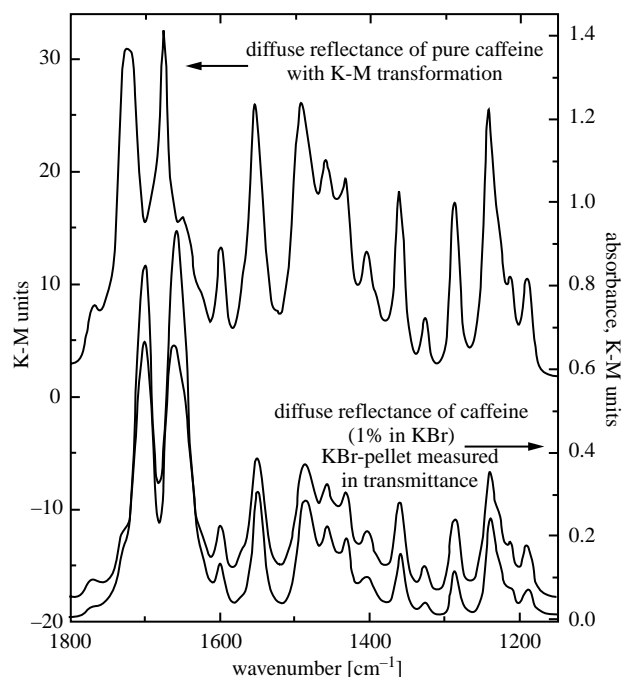


Fig. 6 Spectral differences in mid-infrared spectra of caffeine measured under different conditions. Diffuse reflectance spectrum of the pure crystalline powder (top trace), diffuse reflectance spectrum of 1% of caffeine in KBr-powder (middle trace, intensity data for both spectra in K–M units), and absorbance spectrum as recorded using the KBr pellet technique (bottom trace).

reproducible compression of the disperse sample, a depth of 2 mm was found to reach the K–M requirements of an “infinite” sample thickness, i.e., a further increase in depth causes no change in the measured reflectance (14). The packaging density of the sample, as just noted above, is another parameter that affects the spectrum. The effect of pressure applied to powders for reproducible measurements was studied also by several other spectroscopists [see, for example, Ref. (19)].

Many other materials were studied as matrix materials for diffuse reflectance spectroscopy. Beside the previously noted alkali halides, other examples are silver halides, diamond, germanium, silicon, and chalcogenide glass, preferably of small particle diameters. For an extensive survey on different matrices, see Ref. (20).

A study of the precision and accuracy limiting factors of quantitative assays using diffuse reflectance was carried out by Krivácsy and Hlavay (21), comparing these against transmission measurements. Beside a reproducible sample preparation, the effect of using an underestimated reflectance value of the reference standard ($R_{\infty,r} < 1$) in the K–M transformation approach for the quantitative

evaluation of the spectra leads to a negative intercept for the calibration curves and a slightly reduced linear regression coefficient. Another prerequisite is that scattering stays constant, which is achieved by embedding the sample in a non-absorbing matrix, e.g., of KBr powder with a defined particle size distribution.

The linearity of the K–M function has been investigated for several different conditions. For caffeine, as diluted in KCl powder and reaching up to 100% of weight, maxima of weak and strong absorption bands were tested for this important analytical parameter. It was found that under the conditions given, the use of an accessory with off-axis optical geometry and with application of crossed polarizers before and after the sample, reduced contributions from specular Fresnel reflection from particles on the upper sample surface. As a result, the linear region of the absorbance–concentration dependencies was extended (22). For near-infrared spectra, an alternative linearization function for the measured diffuse reflectance based on $-\log(R)$ is noted above. An intensive investigation on the effect of an absorbing matrix was carried out by Olinger and Griffiths (23), who stated that the concentration range over which linearity holds depends on particle size and matrix absorption strength. The effects of nonlinearity are even more pronounced for mid-infrared spectra (24). Such consequences have to be taken into account when analyzing the *in situ* spectra of adsorbates on substrates, such as from thin-layer-chromatography plates. The combination of spectroscopy with separation techniques is presented below in greater detail.

An interesting study for the interpretation of diffuse reflectance spectra of powders and bulk materials was carried out by Chaffin and Griffiths (25) for the extended near-infrared spectral range, reaching from 10,000 to 2500 cm^{-1} as defined by the authors. The effects of absorption saturation and Fresnel reflection for different amounts of sample scattering are presented. Reducing the scattering inside the studied material, e.g., by application of a transparent liquid for matching the matrix refractive index, a tremendous enhancement of the absorption could be demonstrated, as predicted by the K–M theory.

Because a diffuse reflectance spectrum is scatter-dependent, information on mean particle sizes is also obtainable, which is a parameter of great importance in powder technology. An approach using near-infrared spectrometry combined with multivariate calibration has been presented by Ilary et al. (26). Included was a spectrum standardization by multiplicative scatter correction (see later).

A nonconventional approach for the qualitative analysis of difficult solid samples is to use silicon carbide abrasive paper. The sample is rubbed against the rough surface, and

a small amount sticks to the substrate. A diffuse reflectance accessory is used for recording the spectrum, obtained from a "sample loaded" substrate measurement that is ratioed against that of a clean spot. Recently, metal-coated substrates were found to be even better suited, because the signal-to-noise ratio could be improved owing to much higher reflectivity, compared with the results from using uncoated abrasive paper (recommended grit size, 400). For extremely hard materials, a diamond-in-metal sampling device was used successfully (27).

For qualitative and quantitative mid-infrared studies, a dilution of a sample by a scattering, but transparent matrix is highly recommended. The diffuse reflection technique can be used for low concentrations because of enhancement effects for weak absorptions, which do not exist for transmission measurements. On the other hand, the dilution by KBr powder can add spectral impurities, and the grinding process itself carried out in a humid atmosphere could lead to an increase in water absorption band intensities owing to the hygroscopic nature of the KBr powder. Therefore, great care has to be taken when working with such a material, which should be kept dry in a desiccator or dried again in an oven. In Fig. 7 an example of a diffuse reflectance spectrum of a KBr powder is shown, for which the reference measurement was carried

out using a gold-coated diffusely reflecting substrate. As evident from the baseline offset in the spectrum, the reflectance of the KBr powder is lower than that of the gold-coated standard. The intensification of absorption bands from impurities, either inherent to the KBr used or from a polluted atmosphere, is again to be stressed when using the diffuse-reflectance technique. In this context, suprapure KBr batches and a clean bench are prerequisites for improving assay limits (28).

Diffuse reflectance has also been used extensively for trace analysis. In such a case, the sample to be studied is dissolved using a volatile solvent and transferred to KBr powder, preferably in microcups. The solvent is evaporated, and the sample has coated the KBr particles. With microsampling techniques, sample loadings lower than 100 ng can give satisfactory spectra. If strict rules for avoiding impurities and contamination are observed, detection limits even at the low nanogram level can be reached. Examples from the literature are the identification of different pharmaceuticals (28) and of chlorodibenzo-dioxins isomers (29).

When coupling separation techniques such as thin-layer chromatography (TLC) and liquid chromatography (LC) with infrared spectroscopy for component identification and, less successfully, for quantification, the diffuse reflectance technique has often been applied. Lowest detection limits were usually obtained when the analyte was isolated from its matrix, i.e., the TLC substrate or the eluent in the case of LC-IR coupling. The dissolved compound is usually transferred to a transparent powder, from which it is detected after solvent elimination [see, for example, Ref. (28)]. In situ measurements on TLC plates were also carried out by this and other groups (30). However, because of the strong and broad background absorptions from the stationary phase, such as cellulose, Al_2O_3 , silica gel, and others, analyte detection limits deteriorate significantly compared with the case of a transparent matrix. In addition, substrate interactions with the separated component can often be noticed, which may cause drastic spectrum changes and certainly lead to difficulties with library searching if only normal solid-phase spectra are implemented for reference. Liquid chromatography with polar eluents, as used in reversed-phase LC, complicate the solvent elimination step, although interfaces have been constructed to enable an on-line analysis of effluents from reverse-phase high-performance liquid chromatography (RP-HPLC) (31). A review of analyte deposition-based concepts for such separation techniques is given by Somsen et al. (32). The incorporation of HPLC/FT-IR into pharmaceutical research programs and its automation are discussed by Pivonka and Kirkland (33).

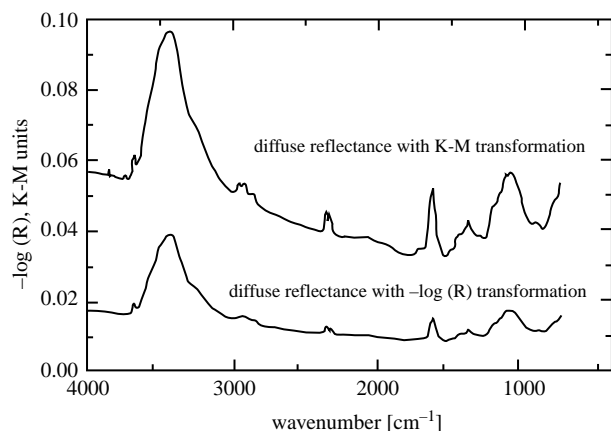


Fig. 7 Diffuse reflectance spectrum of KBr powder measured against a gold-coated diffuse reflective substrate transformed to either $-\log(R)$ or K-M units. The spectral reflectance of KBr powder is lower than that of the gold-coated standard. Furthermore, various absorption bands are noticeable: the broad band approximately at 3500 cm^{-1} and that at 1640 cm^{-1} is from adsorbed water. The doublet at 2350 cm^{-1} is from uncompensated atmospheric CO_2 within the spectrometer and accessory; bands between 1500 and 1000 cm^{-1} arise from inorganic impurities, and bands below 3000 cm^{-1} are from compounds containing aliphatic groups.

CHEMOMETRICS

Quality control, pharmaceutical product identity checks, and quantification are important fields in the broad application of the different spectroscopic methods. There are many spectroscopic aspects, e.g., concerning sample preparation, influences from different accessories, and possibly spectrometer effects, which certainly influences quantitative measurements. The latter problems could be solved using calibration transfer between different spectrometer types, for example, a scanning and an FT-near-IR spectrometer (34).

For qualitative spectrum interpretation, the conventional method for routine identification of chemical species is a library-search, based on spectral mapping algorithms. Before library-searching spectral preprocessing, i.e., elimination of baseline effects and noise, standardization, etc., is performed on the sample spectrum. Comparison of such a processed spectrum with a candidate library spectrum can be based on different principles such as correlation of spectra, similarity and distance measures, and logical operations. For an overview see Ref. (35). For the mid-infrared spectral range, many libraries are available, although only a limited number of spectra might have been recorded with the diffuse reflectance technique. On the other hand, industrial activities are on the way for near-IR libraries, in particular, recorded using the diffuse reflectance technique. For the UV/VIS range, only limited digitized information is available, whereas most spectra are from solution work. A more difficult situation is found for mixtures aiming at component identification. For a discussion and applied strategies, see Ref. (36). Procedures for classifying and interpreting spectra can be based on chemometric tools such as factor analysis, pattern recognition, cluster analysis, and artificial intelligence algorithms (37). Spectral deviations can sometimes be traced back to differences in chemical purity (e.g., water content), particle size, and size distribution or polymorphism (38). Recently, pattern recognition methods were applied for near-infrared detection of polymorphic forms and their conversion. The methods were robust enough when applied to different spectrometers to avoid the need for multivariate instrument standardization (39).

Quantification is another important field in which fast results are needed for process analysis and quality control. Multivariate techniques play an important role in quantitative assays, for which different strategies and algorithms are in use for calibration. Individual wavelength and full-spectrum methods are in use: multiple linear regression (MLR) is an example of the first category, whereas partial least-squares (PLS) and principal

component regression (PCR) usually make use of broad spectral interval information. For more details, see Refs. (40, 41). Apart from so-called full-spectrum methods, spectral variable selection is a playground of chemometricians for achieving improved and more robust calibration models. A recent study also provides a review on the activities in this field (42).

It must be mentioned again that quantification based on diffuse reflectance spectra requires a reproducible sample preparation, because the reflectance depends on particle size and shape, packing density, and the texture of the sample surface. Therefore, sample preparation must be controlled (using same grinding time when using a ball and mill grinder and pressurizing the sample reproducibly). A recent example was provided by Blanco et al., who tested multivariate calibrations for active compounds in pharmaceutical mixtures based on partial least-squares regression in combination with different spectral-preprocessing methods (43). Often, baseline effects are eliminated by taking first or second derivatives of the spectral data, and, in particular, multiple scatter correction is an important preprocessing method used for the standardization of diffuse reflectance spectra (44).

PHARMACEUTICAL APPLICATIONS

The number of applications of diffuse reflectance spectroscopy for pharmaceutical analysis has increased tremendously in recent years. An overview to the activities up to 1986 is found in the first edition of the *Encyclopedia of Pharmaceutical Technology* (45). The on-line identification of pharmaceutical raw materials has been noted above because of the requirements of Good Manufacturing Practice (GMP), which is most often performed now using near-infrared spectroscopy. On the other hand, mid-infrared spectroscopy can provide more information owing to the fact that more vibrational bands from all possible molecular substructures show up within this spectral range, whereas the near-infrared comprises primarily bands corresponding to overtones and combination bands, just involving vibrations with C—H, O—H, and N—H molecular bonds. The superiority of mid-infrared spectroscopy compared with measurements in the near-infrared was shown when a special formulation type and the quantification of the active ingredients were required (46). Other applications are concerned with the development of rapid, reliable, and noninvasive testing assays for whole tablets or intact capsules. Lodder et al. described analyzing several kinds of inorganic compounds through a gelatin wall; their aim was the detection of

capsule-tampering using a special clustering algorithm, which was based on near-infrared reflectance analysis (47). Another direct control of the active ingredient included in tablets without extraction or pulverization by near-infrared spectrometry was possible, as shown by other scientists (48).

Near-infrared spectroscopy was proposed as an alternative to several compendial test methods. As an example, ampicillin trihydrate was used to demonstrate that in total, eight quality criteria could be checked and its "conformity index" calculated (38). Additional chemometrics for quality control were developed by the same group (49). Their strategy was used to distinguish differences in mean particle size of various lactose samples and to study the effectiveness of a blending process for homogenous mixture preparation.

Another task often required is the determination of the water content. Quantification of moisture in hard gelatin capsules was described by Berntsson et al. who discussed its importance in monitoring for at-line process control (50). A sparse MLR calibration model with three optimized near-infrared wavelengths yielded an absolute prediction error of 0.1% over a range of 5.6–18% water content. For monitoring the production process, Han and Faulkner (51) analyzed the spectral range of 1100–2500 nm. Second derivative spectra were used for PLS calibration of moisture. Single-wavelength analysis was applied for coating thickness. Furthermore, identification of tablets inside blister packaging was also possible.

A fiber optic probe was used by Blanco et al. (52) for analysis of spasmocetyl samples with the active compound otilonium bromide and cellulose, maize starch, sodium starch glycolate, and glyceryl palmitostearate as excipients. Another study from this group covered the identification, qualification of the substance, and the quantification of the active component. A library search with a comparison to the near-infrared spectra of 163 pharmaceuticals was involved (53). An on-line monitoring for the determination of the endpoint of polymorph conversions in pharmaceutical processes was recently described (54); further investigations into this field were published and are noted previously (39).

An interesting application is the nondestructive identification of pharmaceutical tablets directly through a blister packing using a fiber optic probe (55). Three different pattern recognition methods based on full near-infrared spectra and subset wavelength ranges were tested to discriminate between film-coated and nonfilm-coated tablets as well as between active and placebo tablets. Dreassi et al. (56) used near-infrared FT spectroscopy for the quality control of solid pharmaceutical formulations. Quantitative assays were developed

for solid drug preparations, for pills containing ibuprofen and tablets with paracetamol, and for powders containing benzydamine hydrochloride and tricetol. Additional work for qualitative and quantitative analysis aimed at all stages of the production, including determination of moisture, was also described by Dreassi et al. (57). Another study was concerned with the characterization of several pharmaceuticals on the basis of their physical properties, i.e., crystalline states and densities (58).

In conclusion, it can be stated that spectroscopic techniques will further dominate the analytical tools of the future with respect to qualitative and quantitative assays. This is because of their speed and the enormous information content of the spectra, especially in the infrared, and the fact that reagent-free multicomponent methodologies are available. The widespread diffuse reflection technique certainly has to compete with others in the laboratory and at the production site. However, for the study of bulk and dispersed systems, it will often be the method of choice. There are additional developments concerned with dedicated instruments and user-friendly interfaces, in which chemometrics play an important role. It is hoped that the sophisticated algorithms presented in the literature will be available soon within expert systems, allowing adaptations to special applications and broadening the general acceptance of these spectroscopic methods presented above.

ACKNOWLEDGMENTS

I am indebted to Mrs. M. Hillig for support in preparing the diagrams. Financial support from the Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen, and the Bundesministerium für Bildung und Forschung is gratefully acknowledged.

REFERENCES

1. Bugay, D.E.; Williams, A.C.; Brittain, H.G., Eds.; *Physical Characterization of Pharmaceutical Solids*; Marcel Dekker, Inc.: New York, 1995.
2. Coates, J. Vibrational Spectroscopy: Instrumentation for Infrared and Raman Spectroscopy. *Appl. Spectrosc. Rev.* **1998**, *33*, 267–425.
3. Workman, J. Jr.; Springsteen, A. *Applied Spectroscopy—A Compact Reference for Practitioners*; Academic Press: San Diego, 1998.
4. Heise, H.M. Clinical Applications of Mid- and Near-Infrared Spectroscopy. *Infrared and Raman Spectroscopy of Biological Materials*; Gremlich, H.-U., Yan, B., Eds.; Marcel Dekker, Inc.: New York, 2001; 259–322.

5. Yang, P.W.; Mantsch, H.H.; Baudais, F. A Critical Evaluation of Three Types of Diffuse Reflectance Infrared Accessories. *Appl. Spectrosc.* **1986**, *40*, 974–978.
6. Springsteen, A. Standards for Reflectance Measurements. *Applied Spectroscopy—A Compact Reference for Practitioners*; Workman, J., Jr., Springsteen, A., Eds.; Academic Press: San Diego, 1998; 247–267.
7. Griffiths, P.R. Letter: Practical Consequences of Math Pre-Treatment of Near Infrared Reflectance Data: $\log(1/R)$ vs $F(R)$. *J. Near Infrared Spectrosc.* **1995**, *3*, 60–62.
8. Kubelka, P.; Munk, F. Ein Beitrag Zur Optik Der Farbanstriche. *Z. Techn. Phys.* **1931**, *12*, 593–601.
9. Kortüm, G. *Reflectance Spectroscopy*; Springer-Verlag: New York, 1969.
10. Loyolka, S.K.; Riggs, C.A. Inverse Problem in Diffuse Reflectance Spectroscopy: Accuracy of the Kubelka–Munk Equations. *Appl. Spectrosc.* **1995**, *49*, 1107–1110.
11. van Gemert, M.J.C.; Star, W.M. Relations Between the Kubelka–Munk and the Transport Equation Models for Anisotropic Scattering. *Lasers Life Sci.* **1987**, *1*, 287–298.
12. Burger, T.; Fricke, J.; Kuhn, J. NIR Radiative Transfer Investigations to Characterise Pharmaceutical Powders and Their Mixtures. *J. Near Infrared Spectrosc.* **1998**, *6*, 33–40.
13. Hecht, H.G. A Comparison of the Kubelka–Munk, Rozenberg, and Pitts–Giovannelli Methods of Analysis of Diffuse Reflectance for Several Model Systems. *Appl. Spectrosc.* **1983**, *37*, 348–354.
14. Krivácsy, Z.; Hlavay, J. Effect of Sample Packing on the Scattering Properties of the Reference Material in Diffuse Reflectance Infrared Spectrometry. *Talanta* **1994**, *41*, 1143–1149.
15. Wendlandt, W.; Hecht, H.G. *Reflectance Spectroscopy*; Interscience Publishers: New York, 1966.
16. Yang, P.W.; Mantsch, H.H. Diffuse Reflectance Infrared Spectrometry: Characteristics of the Diffuse and Specular Components. *Appl. Optics* **1987**, *26*, 326–330.
17. Brimmer, P.J.; Griffiths, P.R.; Harrick, N.J. Angular Dependence of Diffuse Reflectance Infrared Spectra. I. FT-IR Spectrogoniophotometer. *Appl. Spectrosc.* **1986**, *40*, 258–265.
18. Fraser, D.J.J.; Griffiths, P.R. Effect of Scattering Coefficient on Diffuse Reflectance Infrared Spectra. *Appl. Spectrosc.* **1990**, *44*, 193–199.
19. Yeboah, S.A.; Wang, S.H.; Griffiths, P.R. Effect of Pressure in DRIFT Spectra of Compressed Powders. *Appl. Spectrosc.* **1984**, *38*, 259–264.
20. TeVrucht, M.L.E.; Griffiths, P.R. Quantitative Investigation of Matrices for Diffuse Reflectance Infrared Fourier Transform Spectrometry. *Talanta* **1991**, *38*, 839–849.
21. Krivácsy, Z.; Hlavay, J. Method for the Reliable Quantitative Analysis by Diffuse Reflectance Infrared Spectroscopy. *J. Mol. Struct.* **1995**, *349*, 289–292.
22. Brimmer, P.J.; Griffiths, P.R. Angular Dependence of Diffuse Reflectance Infrared Spectra. III. Linearity of Kubelka–Munk Plots. *Appl. Spectrosc.* **1988**, *42*, 242–247.
23. Olinger, J.M.; Griffiths, P.R. Quantitative Effects of an Absorbing Matrix on Near-Infrared Diffuse Reflectance Spectra. *Anal. Chem.* **1988**, *60*, 2427–2435.
24. Brimmer, P.J.; Griffiths, P.R. Effect of Absorbing Matrices on Diffuse Reflectance Infrared Spectra. *Anal. Chem.* **1986**, *58*, 2179–2184.
25. Chaffin, N.C.; Griffiths, P.R. Role of Scattering Coefficients in Extended Near-Infrared Diffuse Reflection Spectrometry. *Appl. Spectrosc.* **1998**, *52*, 218–221.
26. Ilari, J.L.; Martens, H.; Isaksson, T. Determination of Particle Size in Powders by Scatter Correction in Diffuse Near-Infrared Reflectance. *Appl. Spectrosc.* **1988**, *42*, 722–728.
27. Hoult, R.A.; Spragg, R.A. Abrasive Sampling for Diffuse Reflectance Using Reflective Substrates, Itoh, K., Tasumi, M., Eds.; Fourier Transform Spectroscopy, 12th International Conference, Waseda University Press: Tokyo, 1999; 506–507.
28. Otto, A.; Bode, U.; Heise, H.M. Experiences with Infrared Microsampling for Thin-Layer and High-Performance Liquid Chromatography. *Fresenius Z. Anal. Chem.* **1988**, *331*, 376–382.
29. Gurka, D.F.; Billels, S.; Brasch, J.W.; Riggall, C.J. Tetrachlorodibenzodioxin Isomer Differentiation by Micro Diffuse Reflectance Fourier Transform Infrared Spectrometry At the Low Nanogram Level. *Anal. Chem.* **1985**, *57*, 1975–1979.
30. Glauninger, G.; Kovar, K.-A.; Hoffmann, V. Possibilities and Limits of an On-Line Coupling of Thin-Layer Chromatography and FTIR-Spectroscopy. *Fresenius. J. Anal. Chem.* **1990**, *338*, 710–716.
31. Castles, M.A.; Azarraga, L.V.; Carreira, L.A. Continuous On-Line Interface for Reversed-Phase Microbore High Performance Liquid Chromatography/Diffuse Reflectance Infrared Fourier Transform Analysis. *Appl. Spectrosc.* **1986**, *40*, 673–680.
32. Somsen, G.W.; Gooijer, C.; Brinkman, U.A. Th. Analyte-Deposition-Based Detection in Column Liquid Chromatography: Concept and Examples. *Trends Anal. Chem.* **1998**, *17*, 129–140.
33. Pivonka, D.E.; Kirkland, K.M. Research Strategy for the HPLC/FT-IR Analysis of Drug Metabolites. *Appl. Spectrosc.* **1997**, *51*, 866–873.
34. Lin, J.; Lo, S.-C.; Brown, C.W. Calibration Transfer from a Scanning Near-IR Spectrophotometer to a FT-near-IR Spectrophotometer. *Anal. Chim. Acta* **1997**, *349*, 263–269.
35. Luinge, H.J. Automated Interpretation of Vibrational Spectra. *Vibr. Spectrosc.* **1990**, *1*, 3–18.
36. Chen, C.-S.; Li, Y.; Brown, C.W. Searching a Mid-Infrared Spectral Library of Solids and Liquids with Spectra of Mixtures. *Vibr. Spectrosc.* **1997**, *8*, 9–17.
37. Luinge, H.J. Multivariate Methods for Automated Spectrum Interpretation. *Computing Applications in Molecular Spectroscopy*; George, W.O., Steele, D., Eds.; Royal Society of Chemistry: Cambridge, 1995; 87–103.
38. Plugge, C.; van der Vlies, C. The Use of Near Infrared Spectroscopy in the Quality Control Laboratory of the Pharmaceutical Industry. *J. Pharm. Biom. Anal.* **1992**, *10*, 797–803.
39. Aldridge, P.K.; Evans, C.L.; Ward, H.W., II; Colgan, S.T.; Boyer, T.; Gemperline, P.J. Near-Infrared Detection of Polymorphism and Process-Related Substances. *Anal. Chem.* **1996**, *68*, 997–1002.
40. Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley: Chichester, 1989.
41. Marbach, R.; Heise, H.M. On the Efficiency of Algorithms for Multivariate Linear Calibration Used in Analytical Spectroscopy. *Trends Anal. Chem.* **1992**, *11*, 270–275.

42. Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J.P.; Munck, L.; Engelsen, S.B. Interval Partial Least-Squares Regression (IPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 413–419.
43. Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; de la Pezuela, C. Effect of Data Preprocessing Methods in Near-Infrared Diffuse Reflectance Spectroscopy for the Determination of the Active Compounds in a Pharmaceutical Preparation. *Appl. Spectrosc.* **1997**, *51*, 240–246.
44. Isaksson, T.; Næs, T. The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. *Appl. Spectrosc.* **1988**, *42*, 1273–1284.
45. Bornstein, M. Diffuse Reflectance Spectroscopy. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Vol. 4, Marcel Dekker, Inc.: New York, 1991; 23–35.
46. Ryan, J.A.; Compton, S.V.; Brooks, M.A.; Compton, D.A.C. Rapid Verification of Identity and Content of Drug Formulations Using Mid-Infrared Spectroscopy. *J. Pharm. Biomed. Anal.* **1991**, *9*, 303–310.
47. Lodder, R.A.; Selby, M.; Hieftje, G.M. Detection of Capsule Tampering by Near-Infrared Reflectance Analysis. *Anal. Chem.* **1987**, *59*, 1921–1930.
48. Jensen, R.; Peuchant, E.; Castagne, I.; Boirac, A.M.; Roux, G. One-Step Quantification of Active Ingredient in Pharmaceutical Tablets Using Near-Infrared Spectroscopy. *Spectrosc. Int. J.* **1988**, *6*, 63–72.
49. Plugge, W.; van der Vlies, C. Near Infrared Spectroscopy as a Tool to Improve Quality. *J. Pharm. Biomed. Anal.* **1996**, *14*, 891–898.
50. Berntsson, O.; Zackrisson, G.; Oestling, G. Determination of Moisture in Hard Gelatin Capsules Using Near-Infrared Spectroscopy: Applications to At-Line Process Control of Pharmaceuticals. *J. Pharm. Biomed. Anal.* **1997**, *15*, 895–900.
51. Han, S.M.; Faulkner, P.G. Determination of SB 216469-S during Tablet Production Using Near-Infrared Reflectance Spectroscopy. *J. Pharm. Biomed. Anal.* **1996**, *14*, 1681–1689.
52. Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; de la Pezuela, C. Quantitation of the Active Compound and Major Excipients in a Pharmaceutical Formulation by Near Infrared Diffuse Reflectance Spectroscopy with Fiber Optical Probe. *Anal. Chim. Acta* **1996**, *333*, 147–156.
53. Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; de la Pezuela, C.; Russo, E. Control Analysis of a Pharmaceutical Preparation by Near Infrared Reflectance Spectroscopy. A Comparative Study of a Spinning Module and a Fiber Optic Probe. *Anal. Chim. Acta* **1994**, *298*, 83–191.
54. Norris, T.; Aldridge, P.K.; Sekulic, S.S. Determination of End-Points for Polymorph Conversions of Crystalline Organic Compounds Using On-Line Near Infrared Spectroscopy. *Analyst* **1997**, *122*, 549–552.
55. Dempster, M.A.; MacDonald, B.F.; Gemperline, P.J.; Boyer, N.R. A Near Infrared Reflectance Analysis Method for the Non-Invasive Identification of Film-Coated and Nonfilm-Coated, Blister-Packed Tablets. *Anal. Chim. Acta* **1995**, *310*, 43–51.
56. Dreassi, E.; Ceramelli, G.; Corti, P.; Massacesi, M.; Perruccio, P.L. Quantitative Fourier Transform Near-Infrared Spectroscopy in the Quality Control of Solid Formulations. *Analyst* **1995**, *120*, 2361–2365.
57. Dreassi, E.; Ceramelli, G.; Corti, P.; Perruccio, P.L.; Lonardi, S. Application of Near-Infrared Reflectance Spectrometry to the Analytical Control of Pharmaceuticals: Ranitidine Hydrochloride Tablet Production. *Analyst* **1996**, *121*, 219–222.
58. Dreassi, E.; Ceramelli, G.; Corti, P.; Lonardi, S.; Perruccio, P.L. Near-Infrared Reflectance Spectrometry in the Determination of the Physical State of Primary Materials in Pharmaceutical Production. *Analyst* **1995**, *120*, 1005–1008.

FURTHER READING

- Adams, M.J. *Chemometrics in Analytical Spectroscopy*; Royal Soc. Chem.: Cambridge, 1995.
- Knowles, A.; Burgess, C., Eds.; *Practical Absorption Spectrometry. Techniques in Visible and Ultraviolet Spectrometry*; Chapman & Hall: London, 1984; 3.
- Korte, E.-H.; Röseler, A. Foundations and Features of Infrared Reflection Techniques. *Infrared and Raman Spectroscopy*; Schrader, B. Ed.; Wiley-VCH Weinheim 1995; 572–602.
- Kramer, R. *Chemometric Techniques for Quantitative Analysis*; Marcel Dekker, Inc.: New York, 1998.
- Mirabella, F.M. *Modern Techniques in Applied Molecular Spectroscopy*; John Wiley & Sons: New York, 1998.
- Workman, J.J., Jr. Review of Process and Non-Invasive Near-Infrared and Infrared Spectroscopy: 1993–1999. *Appl. Spectrosc. Rev.* **1999**, *34*, 1–89.

SPECTROSCOPIC METHODS OF ANALYSIS— FLUORESCENCE SPECTROSCOPY

Stephen G. Schulman

Jeffrey A. Hughes

University of Florida, Gainesville, Florida, U.S.A.

INTRODUCTION

Luminescence processes in molecules can be classified according to the source from which the excitation energy is derived. In photoluminescence, which encompasses fluorescence and phosphorescence, excitation is achieved by the absorption of light by the potentially luminescent molecule. In chemiluminescence, excitation occurs as the result of an energetic chemical reaction. Fluorescent analytical methods make up the great majority of the ultrasensitive assays carried out by light emission spectroscopy in the pharmaceutical sciences. Although they are more complicated, there has been some interest in phosphorescence and more in chemiluminescence. Accordingly, these phenomena are also briefly treated here.

MOLECULAR ENERGY LEVELS AND ELECTRONIC PROCESSES

Light absorption by an organic molecule results in the promotion of a single electron to a higher energy molecular orbital. The excited state of the molecule thus formed then has two electrons in singly occupied orbitals. The promoted electron and the one it left behind in the orbital it originally occupied can have spins that are opposite (as in the ground state), or they may have the same spin. If the electrons have opposite spins, the molecule is said to be in an excited singlet state. If, on the other hand, they acquire the same spin, the excited molecule is said to be in a triplet state. For each excited singlet state, there is a triplet state that, owing to lower electrostatic repulsion in the latter, has a somewhat lower energy than the corresponding singlet state. Luminescence originating from the electronic transition of a molecule in passing from an excited singlet state back to the ground state is called fluorescence. That originating from the transition from an excited triplet state to the ground state is called phosphorescence.

Transition may occur from the ground singlet state to a number of excited singlet states and to any of a number of vibrational levels within each of these excited singlet states as a result of scanning the sample of interest with the range of wavelengths (energies) of light that lie in the visible and near-ultraviolet regions of the electromagnetic spectrum. These electronic transitions produce the electronic absorption spectrum of the compound of interest. The “vibrational breadth” of each upper electronic state results in the broad-banded appearance of the absorption spectra of molecules, which is unlike the line-like appearance of the absorption spectra of atoms.

Subsequent to excitation, excess energy is rapidly lost by the excited molecules in collisions with the solvent. Each collisional loss of energy causes vibrational and sometimes even electronic changes in the excited molecule, which are radiationless (i.e., no light is emitted). These changes are called vibrational relaxation in the case of vibrational change and internal conversion in the case of radiationless electronic transition. In unconjugated molecules where there is great vibrational freedom, these processes may carry the excited molecule back to the ground electronic state. However, in highly conjugated molecules, especially aromatic molecules, vibrational motion may be restricted, and vibrational relaxation and internal conversion may carry the excited molecule only back to the lowest vibrational level of the lowest excited singlet state. From this state, which is about 10^6 times longer-lived than the upper vibrational and electronic states, the molecule may return to the ground electronic state by light emission (fluorescence), or it may undergo photochemical reaction or a change in spin of one of its electrons. When the latter process occurs, the molecule is said to undergo intersystem crossing, a process that results in the creation of the lowest triplet state. If light emission occurs in the demotion of the molecule from the triplet state back to the ground singlet state, it is called phosphorescence. Phosphorescence entails a change in angular momentum of the emitting molecule and is classically forbidden. This is manifested in the long decay-time of phosphorescence (10^{-4} –10 s). Fluorescence decay-times

(reciprocal probability of decay) are typically 10^{-11} – 10^{-7} s. Both forms of light emission are analytically useful, but fluorescence occurs under a much wider variety of circumstances and has been used for analytical purposes far more than has phosphorescence.

Because fluorescence and phosphorescence originate exclusively from the lowest vibrational level of the lowest excited singlet and triplet states, respectively, and terminate in any of a number of vibrational levels of the ground electronic state, there will be only one possible fluorescence band and one possible phosphorescence band observable from a single chemical species capable of fluorescing and/or phosphorescing. More than one of either is indicative of extraneous luminescing species, either impurities or products of photochemical reaction.

CHARACTERISTICS OF LUMINESCENCE SPECTRA

Fluorescence

Fluorescence is characterized by two spectra corresponding to excitation and emission. In solutions, both spectra usually appear as broad bands that may or may not show vibrational fine structure. The excitation and the absorption spectra of a compound should be the same, but with ordinary fluorescence spectrophotometers, a distorted version of the true absorption spectrum is obtained. The degree of distortion depends on the light output of the spectral excitation source and the responses of the monochromators and detector as functions of the wavelengths of exciting light. The excitation spectrum can be useful in the qualitative and quantitative analysis of mixtures. For example, the excitation spectrum of one component of a mixture can be isolated by setting the emission monochromator at the proper wavelength, provided the absorbance spectra of compounds present in the mixture differ sufficiently.

The difference in energy between the longest wavelength absorption band and the fluorescence band of an analyte is called a Stokes' shift and is usually no more than 100 nm. If the fluorescence band shows an unusually large Stokes' shift, then either the solute exhibits some form of photochemical behavior such as isomerism or prototropic dissociation in the excited state, or more than one fluorescent species is present.

The shape of the fluorescence spectrum is independent of the wavelength of light used to excite it because the fluorescence always takes place from the same level, no matter to what level the molecule was originally excited.

Quantum yield and decay-time of fluorescence

Because of the other processes that compete with fluorescence for deactivation of molecules in the lowest excited singlet state, only a fraction of the excited molecules will return to the ground state by fluorescence. This fraction is called the quantum yield of fluorescence (Φ_f) or fluorescence efficiency. Under given conditions of temperature and environment, Φ_f is a physical constant of the excited molecular species. Φ_f usually decreases with increasing temperature. The actual mean time the molecule spends in the excited state before fluorescing is referred to as the lifetime of the lowest excited singlet state (or fluorescence decay-time) and is represented by the symbol τ_R^0 . The time it would spend in the lowest excited singlet state if there were no other processes competing with fluorescence is called the natural life time, τ_R . The quantum yield of fluorescence, then, may be expressed as being equal to τ^0/τ_R^0 . Because it is a fraction (never greater than unity), τ_R^0 is always greater than or equal to τ^0 . The intensity of fluorescent light emitted depends on the concentration and molar absorptivity of the absorbing (ground state) species and the quantum yield fluorescence of the fluorescing (excited state) species.

Phosphorescence

Phosphorescence is also characterized by an excitation and an emission spectrum. Because the lowest triplet state invariably lies lower in energy than the lowest excited singlet state of the same molecule, phosphorescence will occur at wavelengths longer than those of fluorescence and, therefore, at wavelengths much longer than those of the excitation spectrum. As in the case of fluorescence, the phosphorescence spectrum and the phosphorescence excitation spectrum are distorted by the instrumental components and therefore do not represent "true" spectra.

Because of the long radiative lifetime of the lowest triplet state, most phosphorescence in fluid solutions is obviated by collisional quenching, especially by dissolved molecular oxygen. Phosphorescence, when it occurs, is usually observed at low temperatures (e.g., that of liquid nitrogen) in rigid matrices where it may demonstrate high quantum yields. In the past three decades, much interest has been focused on phosphorescence at room temperature (RTP), which sometimes can be observed in samples adsorbed on solid substrates such as filter paper. Unfortunately, the quantum yields observed in room temperature phosphorescence are low, leading to poor analytical sensitivity, and the method has not enjoyed wide popularity. Phosphorescent measurements at low temperatures require special handling techniques and are difficult to

reproduce. Consequently, although low-temperature phosphorimetry is inherently a very sensitive analytical method, it too has yet to attract a substantial group of users. However, there are a substantial number of molecules of pharmaceutical interest that phosphoresce but do not fluoresce (e.g., metronidazole) or fluoresce weakly and phosphoresce intensely (e.g., warfarin), and it is probably worthwhile to pursue phosphorimetry as a valuable aspect of molecular luminescence spectroscopy.

Chemical Structural Aspects of Fluorescence Spectra

Fluorescence spectra of analytical and pharmaceutical interest arise from functionally substituted aromatic molecules, particularly those derived from benzene, naphthalene, and anthracene or their heteroaromatic analogs pyridine, quinoline, isoquinoline, and acridine.

The intensity of fluorescence observable from a given molecular species depends primarily on the quantum yield of fluorescence, which may affect the intensity of fluorescence over about four orders of magnitude and may determine whether fluorescence is observable at all. The quantum yield of fluorescence is dependent on the rates of processes competing with fluorescence for the deactivation of the lowest excited singlet state. These, in turn, depend on molecular structure.

Aromatic molecules, containing lengthy aliphatic side chains, generally tend to fluoresce less intensely than those without the side chains. This is brought about by the introduction of a large number of vibrational degrees of freedom by the aliphatic moieties, providing an efficient pathway for internal conversion. In the unsubstituted aromatic molecules, the rigidity of the aromatic ring results in wide separation of the ground and lowest excited singlet states. In general, molecular rigidity and high quantum yield of fluorescence are closely related.

The fluorescence efficiencies of aromatic molecules are reduced by heavy atom substituents such as bromine and iodine and by certain other groups such as aldehyde and keto as well as nitro groups. However, in many cases the substituents that decrease the intensity of fluorescence enhance the intensity of phosphorescence. Consequently, aromatic nitro compounds, bromo- and iodo-derivatives, aldehydes, ketones, and some *N*-heterocyclics tend to fluoresce very weakly or not at all. However, most of them phosphoresce quite intensely. On the other hand, many substituents that are electron donors such as amino, hydroxy, and methoxy often tend to increase the quantum yields of fluorescence of molecules to which they are attached.

The energies of the ground and lowest excited singlet states of fluorescing molecules are affected by such features of molecular structure as the presence of substituents and the molecular geometry. This is reflected in the position (energy) of the fluorescence band maximum in the spectrum. The greater the separation between the ground and lowest excited singlet states, the greater will be the frequency and the shorter will be the wavelength of fluorescence. This separation depends on the energy difference between the highest occupied and lowest unoccupied molecular orbitals and the repulsion energy between the electronic configurations corresponding to the ground and lowest excited singlet states. For aromatic molecules, the greater the extension of the conjugated system, the smaller the energy separation between the highest occupied (π) and lowest unoccupied (π^*) orbitals. Thus, the wavelengths of fluorescence (λ_f) increase with increasing extension of conjugation in the series benzene ($\lambda_f = 262$ nm), naphthalene ($\lambda_f = 314$ nm), and anthracene ($\lambda_f = 379$ nm).

The fluorescence maxima of aromatic compounds with substituents differ from those of the parent hydrocarbons. Groups such as $-\text{NH}_2$, $-\text{OH}$, and $-\text{SH}$ have unshared electron pairs of energy higher than that of the π -electrons of aromatic molecules. These can be transferred into vacant π -orbitals belonging to the aromatic ring. Thus, the energy gap between the highest occupied and lowest unoccupied orbitals of the substituted molecule is considerably smaller than that between the highest occupied and lowest unoccupied orbitals of the unsubstituted (parent) molecule, and the wavelengths of fluorescence in the former are somewhat longer. For example, 1-naphthylamine fluoresces at 372 nm in hexane whereas naphthalene fluoresces at 314 nm in the same solvent.

Groups such as carboxyl and nitrile, which are electron-withdrawing, have localized, vacant, low-energy π -orbitals in the ground state of the substituted molecule. This type of substituent introduces a vacant orbital between the highest occupied and lowest unoccupied π -orbitals of the unsubstituted molecule. The energy gap between the highest occupied and lowest unoccupied orbitals of the substituted molecule is therefore smaller than the gap between the highest occupied and lowest unoccupied orbitals of the unsubstituted molecule. Fluorescence thus appears at longer wavelengths in the substituted molecule than in the unsubstituted molecule. For example, the fluorescence of 2-naphthoic acid lies at 344 nm in hexane, 30 nm longer in wavelength than the fluorescence of naphthalene.

When an electron-withdrawing group and an electron-donating group are attached to the same aromatic ring,

fluorescence may be viewed as involving transition between the lone-pair orbital of the donor group and the vacant π -orbital of the acceptor group. In this case, the energy of the transition is lower and, thus, the fluorescence wavelength is longer than when either group alone is attached to the ring. When two donor groups or two acceptor groups are attached to the aromatic ring, the position of the fluorescence band is usually determined by the donor group with the highest energy lone pair or the acceptor group with the lowest energy vacant orbital.

Solvent Effects

The solvents in which fluorescence spectra are observed play a major role in determining the spectral positions and intensities with which fluorescence bands occur. In some cases, the solvent may determine whether fluorescence is to be observed at all.

Solvent interactions with solute molecules are predominately electrostatic in nature and may be classified as dipolar or hydrogen-bonding. The position of the fluorescence band maximum in one solvent, relative to that in another, depends on the relative separations between ground and excited state in either solvent and, therefore, the relative strengths of ground- and excited-state solvent stabilization.

If the excited state of a polar molecule has a higher dipole moment than its ground state (most molecules are in this class), the excited state will be more stabilized by interaction with a polar solvent than will the ground state. As a result, upon going from a less polar to a more polar solvent, the fluorescence spectrum will shift to longer wavelengths. In a few cases, the ground state of a solute is more polar than the excited state. In this case, going to a more polar solvent stabilizes the ground state more than the excited state, causing a shift to shorter wavelengths with increasing solvent polarity.

Hydrogen-bonding solvents having positively polarized hydrogen atoms are said to be protic solvents. They interact with the nonbonded and lone electron pairs of solute molecules. Hydrogen-bonding solvents having atoms with lone or nonbonded electron pairs are called hydrogen-bond acceptor solvents. They interact with positively polarized hydrogen atoms on electronegative atoms belonging to the solute molecules (e.g., in $-\text{COOH}$, $-\text{NH}_2$, $-\text{OH}$, $-\text{SH}$). Because most hydrogen-bonding solvents are also polar, hydrogen-bonding and nonspecific dipolar interaction are usually both present as modes of solvation of functional molecules. Accordingly, the spectral shifts actually observed on going from one solvent to another are a composite of dipolar and hydrogen bonding effects that may be constructively or destructively additive.

Protic solvents interacting with functional groups that are electron-withdrawing in the excited state (e.g., carbonyl) enhance charge transfer by introducing a partial positive charge into the electron acceptor group. This stabilizes the excited state relative to the ground state so that the fluorescence spectra shift to longer wavelengths, and often the quantum yield of fluorescence increases with increasing hydrogen-bond donor capacity of the solvent. Increasing protic nature of the solvent produces shifts to shorter wavelengths and usually lower fluorescence efficiencies when interacting with lone pairs on functional groups, which are electron donors in the excited state (e.g., $-\text{OH}$, $-\text{NH}_2$). Hydrogen-bond acceptor solvents produce shifts to longer wavelengths and higher fluorescence efficiencies by solvating hydrogen atoms on functional groups that are electron donors in the excited state (e.g., $-\text{OH}$, $-\text{NH}_2$). This is a result of partial withdrawal of the positively charged proton from the functional group that facilitates transfer of electronic charge away from the functional group. Finally, solvation of hydrogen atoms on functional groups that are charge transfer acceptors in the excited state (e.g., carboxyl) inhibits charge transfer by leaving a residual negative charge on the functional group. This results in shifting of the fluorescence spectrum to higher frequency and gives rise to lower quantum yields of fluorescence.

The same considerations are applicable to the influence of the solvent on the phosphorescence wavelength, but the shifts are much smaller than in the case of fluorescence.

pH Effects

The effects of solution acidity and basicity on luminescence spectra result from the dissociation of acidic functional groups or protonation of basic functional groups associated with the aromatic portions of fluorescing and phosphorescing molecules. Protonation or dissociation can alter the natures and rates of nonradiative processes competing with luminescence and, thereby, affect the quantum yields of emission. For example, the antimalarial mefloquine fluoresces very weakly but phosphoresces well in neutral aqueous media. However, at $\text{pH} < 1$, its protonated form fluoresces intensely, and its phosphorescence is very weak.

Protonation and dissociation alter the relative separations of the ground and excited states of the reacting molecules and thereby cause shifting of the luminescence spectra. The shifts tend to be greater in fluorescence spectra than in phosphorescence spectra and are attributable to the electrostatic stabilization or destabilization of the excited state, relative to the ground state, produced by protonation and dissociation. The protonation

of electron-withdrawing groups such as carboxyl, carbonyl, and pyridinic nitrogen causes shifts of the luminescence spectra to longer wavelengths, whereas the protonation of electron-donating groups such as the amino group produces spectral shifts to shorter wavelengths. The protolytic dissociation of electron-donating groups such as hydroxyl or sulfhydryl produces spectral shifts to longer wavelengths, whereas the dissociation of electron-withdrawing groups such as carboxyl produces shifting of the photoluminescence spectra to shorter wavelengths.

One of the interesting aspects of acid-base reactions of fluorescent molecules in fluid solutions is derived from the occurrence of protonation and dissociation during the lifetime of the excited state. This phenomenon affects the dependence of fluorescence on pH and must be considered in the development of a fluorometric analysis in aqueous solutions.

The lifetimes of molecules in the lowest excited singlet state are typically of the order of 10^{-11} – 10^{-7} s. Typical rates of prototropic reactions are 10^{11} dm³ mol⁻¹ s⁻¹ or lower. Consequently, excited-state proton-transfer reactions may be competitive with radiative deactivation of the excited molecules.

In the event that excited-state proton transfer and radiative deactivation are temporally competitive, the quantum yield of fluorescence will demonstrate a complex dependence on the pH level of the solution. Because the electronic distribution of a molecule is usually much different in an excited state than in the ground electronic state, the variation in fluorescence intensity caused by the pH dependence of the quantum yield of fluorescence will occur in a pH region different from the pH region in which the fluorescence intensity depends on the absorbance of the analyte. This means that in an arylammonium ion, for example, which becomes more acidic in the lowest excited singlet state, one might observe fluorescence from the conjugate base at a pH level of zero, even though the pK_a level might be as high as 5.

Quenching

The emission of light by photoexcited luminescent molecules may be decreased or even eliminated by interactions with other chemical species. This phenomenon is called quenching of luminescence.

Two kinds of quenching are distinguished. In static quenching, complexation between the potentially luminescent molecule and the quencher takes place in the ground state. The complex, when excited, fails to luminesce. The efficiency of quenching is governed by the formation constant of the complex as well as by the

concentration of the quencher. The quenching of the fluorescence of doxorubicin by Fe(III) is an example.

In dynamic quenching, the quenching species and the potentially luminescent molecule react subsequent to photoexcitation of the latter and during the lifetime of its excited state. Dynamic quenching is also called diffusional quenching, and its efficiency depends on the viscosity of the solution, the lifetime of the excited state (τ°) of the luminescent species, and the concentration of the quencher (Q). This is summarized in the Stern–Volmer equation (Eq. 1):

$$\Phi/\Phi_f^\circ = \frac{1}{1 + K_Q\tau^\circ[Q]} \quad (1)$$

where k_Q is the rate constant for encounters between quencher and potentially luminescing species, and Φ_f° and Φ_f are the quantum yields of luminescence in the absence and presence of concentration of the quencher (Q), respectively.

Dynamic quenching is also characterized by the dependence of the actual lifetime of the excited state on (Q). In static quenching, the lifetime of the excited state is invariant with respect to (Q).

Concentration Effects

As the concentration of a potentially luminescent solute is increased, the frequency of encounters between solute molecules is increased. This often results in the formation of solute complexes at the expense of monomeric solute molecules. Obviously, such interactions will affect the fluorescence expected from a given solution based strictly on the formal concentration of monomers and can seriously affect the results of a fluorimetric analysis.

In solutions of moderate concentration, there are two types of excited state solute–solute interaction, which are common. An excited polymer, or excimer, may form through the aggregation of excited solute molecules with ground-state molecules of the same type. Fluorescence quenching and/or a spectral red shift may result. Also, a heteropolymeric excited-state complex may form between two different solute molecules. Such a complex is referred to as an exciplex.

The transfer of electronic excitation energy from one molecule to another is another phenomenon that is related to the concentration of potentially luminescent molecules. Energy transfer occurs quite frequently in nature, either by direct collision or even over distances as great as 50 Å or more by a radiationless mechanism by which the excitation energy is transmitted from the molecules that are originally excited (donors) to the recipient molecules

(acceptors). The efficiency with which an excited donor will transfer its excitation energy to an acceptor molecule (rather than fluoresce) is a function of the lifetime of the excited state of the donor, the orientations of the donor and acceptor molecules with respect to one another, and the inverse sixth power of the distance between the donor and acceptor. The energy-transfer process quenches the fluorescence of the donor and often sensitizes the fluorescence of the acceptor.

Temperature Effects

At low temperature, the efficiency of luminescence is usually increased, and the spectra often become sharper. Both effects result from the increase in solution viscosity and thus a decrease in collisional deactivation. In addition to changes in intensity, wavelength shifts also occur when the temperature of a medium is changed. At very low temperature, fluorescence originates from an excited state that is locked into the ground-state equilibrium geometry and solvent cage. At room temperature, in fluid solutions, fluorescence occurs only after the molecular geometry and solvent cage of the fluoresce have adapted to the new electronic distribution characteristic of the excited state. The emitting excited state under fluid conditions is lower in energy than under rigid conditions and, therefore, the fluorescence under the latter circumstance will be at higher energy (shorter wavelength) than in the former.

QUANTITATIVE ANALYSIS

Luminescence spectroscopy is used more often in quantitative analysis than in any other application. The quantitative relationship between fluorescence intensity F and analyte concentration C is derived from the Beer–Lambert law:

$$I = I_0 10^{-\epsilon Cl} \quad (2)$$

where I and I_0 are the intensities of exciting light transmitted through and incident on the sample, l is the length of the light path through the sample, and ϵ is the molar absorptivity of the molecular species of interest at the nominal wavelength of excitation. The intensity of light absorbed I_a is then:

$$I_a = I_0 - I = I_0 (1 - 10^{-\epsilon Cl}) \quad (3)$$

If all molecules absorbing light fluoresced, then I_a would be the intensity of fluorescence. However, only the

fraction Φ_f fluoresces. The remaining fraction $1 - \Phi_f$ returns to the ground state by nonradiative means. Thus:

$$F = \phi_f I_a = \phi_f I_0 (1 - 10^{-\epsilon Cl}) \quad (4)$$

If the absorbance (ϵCl) of the sample is less than 0.02, then to within an approximately 2% error, Eq. 4 becomes:

$$F = 2.3 \phi_f I_0 \epsilon Cl \quad (5)$$

in which F is linear with respect to C . It then remains only to prepare a standard solution of concentration C_s and having fluorescence intensity F_s to be compared with the unknown solution of concentration C_u and fluorescence intensity F_u according to:

$$C_u = \frac{F_u}{F_s} C_s \quad (6)$$

It should be noted that the conditions for Eq. 5 to obtain can be met by either working with very dilute solutions or by exciting in a region of the absorption spectrum where ϵ is small.

Of course, it is always prudent to prepare a calibration curve for several standard solutions to check the linearity of the F versus C plot before attempting the simple relative fluorimetry expressed in Eq. 6 for a large number of samples.

In the case of phosphorescence, the intensity P is related to the concentration of analyte, following similar lines of reasoning applied to fluorescence, according to:

$$P = \phi_{ST} \phi_P I_0 (1 - 10^{-\epsilon Cl}) \quad (7)$$

where Φ_{ST} and Φ_P are the quantum yields of intersystem crossing (the fraction of excited molecules converted from the lowest excited singlet state to the lowest triplet state) and phosphorescence (the fraction of molecules arriving in the lowest triplet state that are deactivated by phosphorescence), respectively. As is the case for fluorescence, Eq. 7 can be expanded and for weakly absorbing solutions can be made analogous to Eq. 6, according to:

$$C_u = \frac{P_u C_s}{P_s} \quad (8)$$

where P_u and P_s are the phosphorescence intensities of the unknown and standard samples, respectively. Eq. 8 is, therefore, the basis of simple quantitative phosphorimetry.

INSTRUMENTATION

The basic components of luminescence instrumentation are generally arranged as shown in Fig. 1. The sample is placed in a sample cell and excited by either ultraviolet or

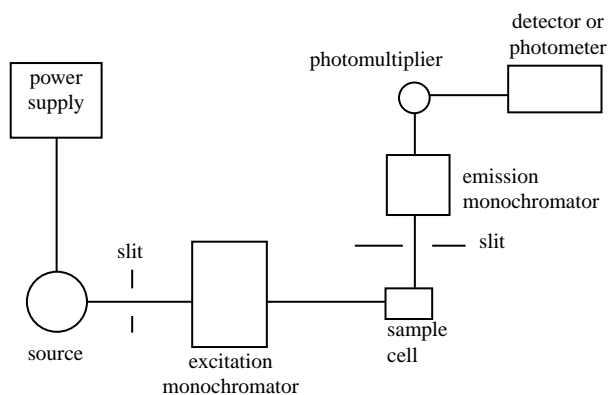


Fig. 1 General layout of luminescence instrumentation.

visible light from a source. A filter or monochromator may select a particular excitation wavelength region. The sample emits luminescence radiation in all directions, but only a portion is observed by the detection system that is usually set at 90° from the source of excitation to minimize the possibility of detecting stray light from the source. Another filter or monochromator may select the emission wavelength region, and a detector would convert the luminescence radiant flux into an electrical signal. An amplifier-readout system amplifies and processes the signal and displays it in the required form.

Sources

Sources of radiant energy are generally classified according to the spectral distribution of the radiation emitted. Continuum sources have a broad spectral distribution covering a large wavelength range. Line sources have spectra consisting of a number of sharp lines or bands. Lasers have allowed the use of extremely narrow bandwidths.

Continuum sources

Incandescent lamps are the simplest type of light source. They usually consist of a heated rod or filament of tungsten in an inert gas such as nitrogen or argon at a pressure of approximately 1 atm.

Radiation emitted by the lamp is caused by the temperature achieved by its filament, which in turn depends on the power input on the lamp. The radiation spectral distribution is a function of the temperature of the source and the wavelength, and this relationship is defined by the Planck black body equation:

$$B = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/kT} - 1} \quad (9)$$

where B = spectral radiance in $\text{J/s m}^2 \text{ nm sr}$; h = Planck's constant, $6.6262 \times 10^{-34} \text{ J/s}$; c = velocity of light in a vacuum, $2.99792 \times 10^8 \text{ m/s}$; λ = wavelength in nanometers; k = Boltzmann's constant, $1.38066 \times 10^{-23} \text{ J/K}$; and T = absolute temperature in Kelvin.

A typical 500 W tungsten lamp operates at 3000 K with an emissivity of 0.4. If the spectral radiance is plotted against wavelength, the spectral radiance will have a maximum at approximately 1000 nm and swiftly drops at approximately 300 and 5000 nm, so that the tungsten lamp is most useful in the visible and near-infrared range.

Incandescent lamps are not sufficiently intense to be useful in luminescence spectrometry, although their stability cannot be matched by gas discharge and arc lamps.

Discharge lamps involve the application of voltage across two electrodes in a gas or metal vapor consisting of neutral atoms and/or molecules. If free electrons are introduced (through a high-voltage spark), then electrons are accelerated and excited, producing light emission. Gas discharge lamps use hydrogen, nitrogen, and the inert gases. Metal vapor discharge lamps use the more volatile metals such as mercury, cadmium, zinc, gallium, indium, and thallium.

With an increased voltage across the gap, higher states of excitation occur, and radiation at shorter wavelengths is emitted. When the voltage exceeds the ionization potential, the gas will be ionized by collisions resulting in increased current such that protons are accelerated to the cathode and electrons to the anode. The reformation of protonated gas atoms or molecules is very improbable at low pressures and currents because secondary electrons generated as ions may strike the cathode, neutralizing the protonated gas atom or molecule. When the secondary electrons are sufficient to maintain the discharge current, a glow discharge occurs. When the current is further increased, the cross-sectional area of the discharge increases proportionally, and the voltage becomes constant. To increase the current further, the voltage must be increased, resulting in increased brightness with a change to an arc discharge where a voltage drop occurs. A contraction in discharge into a smaller area on the cathode may also follow.

Depending on the pressure in the lamp, both line and continuum emission may be observed. Low-pressure lamps operated at low current densities and temperatures produce sharp atomic lines with little or no continuous background. Increasing pressure and temperature cause the lines to broaden and increase the intensity of the background continuum.

High-pressure xenon, mercury, or mercury–xenon arc lamps are the most common excitation sources because of

their relatively high intensity, wide spectral range, and low cost. Commercially available xenon lamps are available in a wide range of power ratings and are usually operated on DC for greatest stability and longest life. Below 300 nm, the light output of the xenon lamp falls off sharply, such that the long wavelength peaks of the excitation spectrum are exaggerated and appear different from those in the absorption spectrum. Mercury–xenon lamps produce very intense light corresponding to the line emission spectrum of mercury (principal lines are at 254, 313, and 365 nm), and generally the light produced is more intense than that produced at any given wavelength by a xenon lamp of comparable power rating. However, these lamps emit very little light at other wavelengths, limiting the choice of excitation wavelength that may be used. However, because many substances have absorption spectra that overlap the mercury emission lines, they may be excited using the mercury–xenon lamp.

Pulsed discharge lamp (flash lamps) are used when high-intensity, short-lived pulses of exciting light are needed. A high-voltage capacitor is discharged across the gap, causing a short, high-current pulse to the lamp, and producing a very intense pulse of light.

The energy input to the lamp per flash is characteristic of lamp manufacture and is related to the capacitance and the voltage across the lamp:

$$Q = CV^2 \quad (10)$$

where Q = energy input in joules; C = capacitance in farads; and V = operating potential in volts.

The duration of the light flash is approximately equal to the duration of the current pulse through the lamp:

$$t_f = 1/2 R_a C \quad (11)$$

where t_f = flash duration in seconds, and R_a = effective arc resistance in ohms.

The effective arc resistance is proportional to the arc gap length and is inversely related to the 2/3 power of the operating voltage. This implies that the shortest, most intense pulses are obtained from lamps with very short gap tubes that are operated under high voltage and low capacitance. The value of R_a for a particular lamp is given by the manufacturer and is specific for a set of conditions.

The peak input power per flash in watts is given by:

$$P_i = Q/t_f \quad (12)$$

and the peak light output per flash in watts is:

$$P_0 = \varepsilon P_i \quad (13)$$

where ε = efficiency of conversion (usually 0.25–0.5 for xenon lamps).

The average input electrical power in watts is:

$$\bar{P}_i = Qf \quad (14)$$

where f = repetition rate in hertz.

The manufacturer generally recommends the maximum P_i value; higher-powered types may need to be forced air-cooled.

The average current drawn from the power supply is given by:

$$\bar{i} = VCf \quad (15)$$

The value of f should not exceed approximately $0.1 t_f^{-1}$ to prevent continuous ionization of the lamp.

The charging resistance R_c is selected so that its power rating is greater than the average input electrical power.

$$R_c = \frac{1}{5RC} \quad (16)$$

Xenon, nitrogen, and hydrogen flash lamps are available. The nature and pressure of the fill-gas affect the spectral distribution of the flash lamp.

The arc may appear narrow and filamentary when a xenon flash tube is operated below its maximum input power per flash because not all of the gas is ionized. The position of the arc tends to shift from flash to flash, causing erratic light output. It is generally not advisable to operate a flash tube very far below its maximum input.

Line sources

Glow discharge lamps (hollow cathode lamps) producing line spectra are quite similar to those propagating

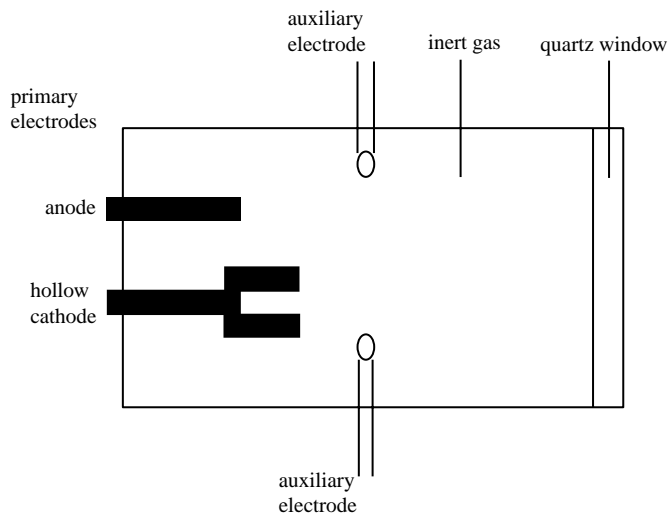


Fig. 2 Schematic diagram of a hollow cathode lamp.

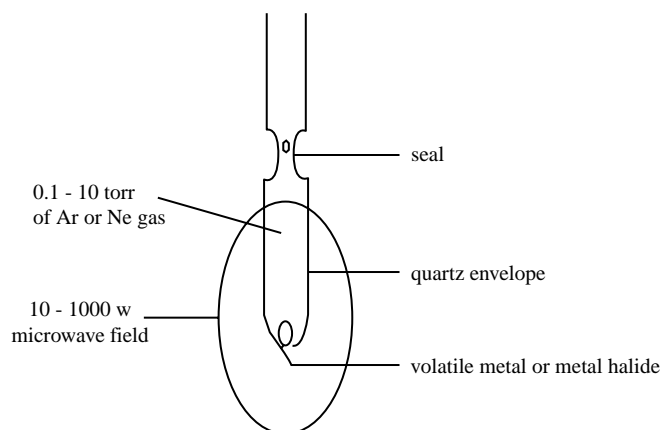


Fig. 3 Schematic diagram of an electrodeless discharge lamp.

continuous spectra presented above. Accelerated electrons generate positive ions on collision with atoms of a carrier gas (Fig. 2). The ions gain energy in an electric field and collide with the cathode. Atoms of the cathode are released and are excited by inelastic collisions with the electrons. The excited metal atoms then emit light.

The lamps are usually sealed and may have auxiliary electrodes to excite atomic vapor. They are operated at DC but may be modulated and pulsed. If the primary current i_p introduced is too high, the lines emitted are self-reversed. However, this may be avoided by keeping i_p low and i_a (auxiliary current) high.

Electrodeless discharge lamps, or EDLs (Fig. 3), are easy to make, but the results obtained are not always reproducible owing to chemical effects with the walls of the lamp, poor outgassing, poor seal-off, etc. A tesla coil starter produces a few free electrons that are accelerated by a high-frequency electric field (through the use of radio- or microwaves). The electrons acquire enough energy to excite and ionize atoms produced by the thermal heating. The actual discharge is confined to a "skin effect," wherein the outer surface of the envelope minimizes self-reversal. Low pressures (0.1–1 torr) of argon or neon gas are used; helium is usually not used because it diffuses out of the quartz envelope and is not very stable in intensity.

The lamps may be operated as continuous wave sources or be modulated or pulsed. EDLs have a poor stability or poor shelf life in some cases, but they have high intensities, produce narrow spectral lines, and are relatively inexpensive.

Lasers

Modern tunable dye lasers have any of several advantages over conventional sources, including 1) extremely narrow linewidths, 2) high intensity, 3) collimation and excellent

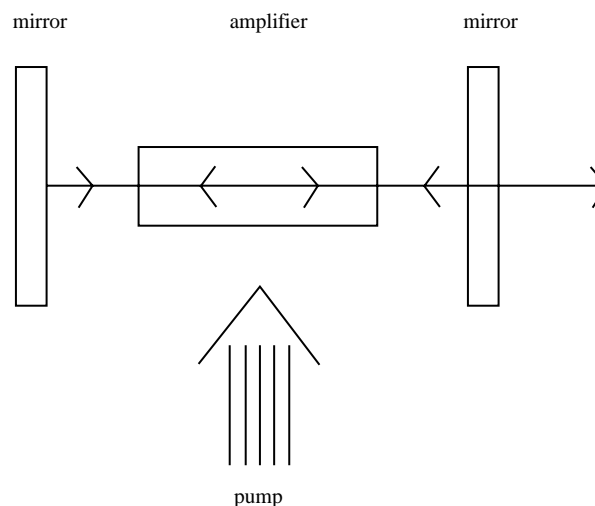


Fig. 4 Laser with conventional two-mirror cavity.

focusing abilities, 4) continuous tunability over certain wavelength regions, and 5) extremely short pulse duration. The primary disadvantages are their expense, relative complexity, and the difficulty to obtain wavelength-tunable laser output at wavelengths below 320 nm (nonlinear frequency mixing techniques may be used).

Lasers have three primary components (Fig. 4): 1) an active medium that amplifies incident electromotive waves, 2) an energy pump that selectively pumps energy into the active medium to populate selected levels and to achieve population inversion, and 3) an optical resonator, or cavity, composed of two opposite mirrors a set distance apart that store part of induced emission concentrated in a few resonator modes. A population inversion must be produced in the laser medium, deviating from the Boltzman distribution; thus, the induced emission rate exceeds the absorption rate, and an electromotive wave passing through the active medium is amplified rather than attenuated. The optical resonator causes selective feedback of radiation emitted from the excited species in the active medium. Above a pump threshold, feedback converts the laser amplifier to an oscillator, resulting in emission in several modes.

The wavelength range over which the laser would operate depends on the spectral range over which the active medium and cavity have a net gain. Wavelength selection devices such as gratings, prisms, and etalons are placed in the cavity when restricting the lasing wavelength is desired; these reduce the gain of the cavity below lasing threshold for all the desired wavelengths.

The active medium in tunable dye lasers is a dye solution. The dyes are organic compounds with conjugated double bonds and have delocalized π electrons that can

produce large absorption cross-sections. Because of the large dipole moments that can be generated, the spontaneous lifetime is short (<10 ns), and the quantum yield of fluorescence is large. Dye lasers are pumped by flashlamps, N_2 lasers, Cu vapor lasers, Nd-YAG, excimer, or Argon ion lasers. They oscillate within a broad band. Therefore, short pulses can be obtained.

Semiconductor lasers (laser diodes) have recently appeared as luminescence excitation sources. They have several outstanding advantages over gas and dye lasers and arc lamps as excitation sources. They are more powerful and more coherent light sources than arc lamps. They also have the advantage of producing a polarized beam that can be useful for certain applications. They are smaller, more compact, and much less expensive than other kinds of lasers, and the fact that most semiconductor lasers emit in the far-red or infrared means that less energy can be deposited in the sample so that considerably less thermal decomposition can occur than with conventional ion, excimer, and dye lasers. That until recently, emission from semiconductor lasers has been generally confined to the far-red and infrared has been a mixed blessing. On the one hand, the fact that so few substances demonstrate electronic absorption in that region of the electronic spectrum means that excitation with semiconductor laser light will be very selective. For example, the interfering luminescences of tryptophan and bilirubin from serum samples will not be problematic with semiconductor laser excitation because these substances are not excited by red light. On the other hand, because so few substances absorb red light, a new generation of fluorescent probes and labels is required for the labeling of analytes to be detected subsequent to chromatographic or immunochemical analysis.

Several of the polymethine dyes absorb in the red, far-red, and near-infrared and fluoresce efficiently as well in this spectroscopic region. Functional derivatives of these may provide excellent fluorescent labels for semiconductor laser excitation. Several research groups are currently actively involved in the synthesis and development of large polyunsaturated dye molecules that are excited and show luminescence in the red and near-infrared. This promises to be one of the most exciting areas of luminescence spectroscopy for the foreseeable future.

Wavelength Selection

The choice of limited spectral wavelength bands for excitation and emission is necessary to prevent the possibility of excitation and detection of emission from extraneous species and to spurious stray light. Mono-

chromators and filters are used to select the wavelength areas of interest.

Filters

Filters are absorption type, made of tinted glass, gelatin-containing organic dyes lacquered or sandwiched between glass, or a solution of absorbing substances, or are interference filters.

Absorption filters selectively absorb portions of incident polychromatic light and allow the transmission of light of preferred wavelengths. There are three general classifications: 1) neutral tint, 2) cut-off filters, and 3) bandpass filters. Neutral tint or neutral-density filters have a nearly constant transmission over a range of wavelengths; they are used with strongly fluorescing compounds and decrease light intensity uniformly. Cut-off filters have a sharp cut-off in their transmittance characteristics; beyond a certain wavelength there is little or no transmittance. These are especially useful in preventing stray or unwanted light from falling on the detector. Bandpass filters either transmit or refuse a set wavelength band and are usually made from a series of cut-off filters.

An interference filter consists of two highly reflective but partially transmitting films of silver separated by a spacer film of completely transparent material (e.g., MgF_2). Light of wavelength corresponding to the optical separation of the silver films (± 5 – 9 nm), and integral multiples of this principal wavelength, are transmitted, whereas other wavelengths are eliminated by destructive interference. These filters are suited for intense sources because they absorb very little energy and rarely need to be cooled.

Monochromators

Monochromators are usually diffraction gratings with slit arrangements. Prisms were once used as monochromators but are now almost obsolete because gratings are much less expensive and less cumbersome. Polychromatic light is dispersed by the grating into its component wavelengths through constructive and destructive interference. Two types of gratings are available: transmission and reflection gratings.

A large number of parallel transparent and opaque lines are arranged alternately on a transmission grating; when a source is incident on one side of the grating, each transparent line acts as an independent line source of the original radiation. Interference will occur among the monochromatic light waves transmitted by the closely spaced transparent lines, causing an augmentation in light intensity at certain points and elimination at others.

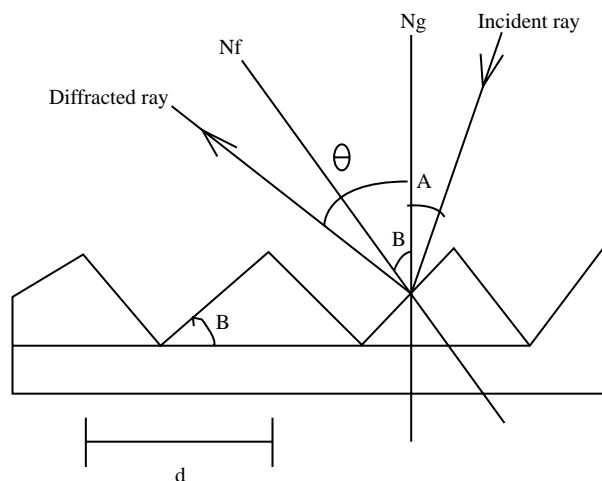


Fig. 5 Diffraction grating. Ng = normal to the grating; Nf = normal to the face of the grating; A = angle of incidence; Θ = angle of diffraction; d = grating spacing; and B = blaze angle.

A reflection grating is similar to a transmission grating except that it uses a large number of grooves on a reflective surface to cause the interference.

Fig. 5 shows a parallel light beam incident onto two adjacent grooves of a diffraction grating. At a wavelength and an angle of incidence A normal to the grating surface, constructive interference for an angle B is obtained:

$$m\lambda = d(\sin A \pm \sin B) \quad (17)$$

where m = order of interference.

The plus sign is taken if A and B are on the same side of the grating normal and the minus sign if they are not. If the light is reflected back into the direction of the incident light, the equation reduces to:

$$m\lambda = 2d \sin B \quad (18)$$

When $m = 0$, the angle of the incident radiation is equal to the angle of diffraction, and this corresponds to specular reflection of the incident radiation. When $m = 1$, the diffraction is said to be first-order and is the primary image of the spectrum. When $m = 2$, the diffraction is second-order and is the second image of the spectrum with a dispersion twice that of the first-order, and so forth for higher orders. This may cause interference. Therefore, a filter is used to eliminate higher-order spectra.

Two important factors in selecting a grating are its blaze angle and resolving power. The blaze angle ϵ is dependent on the blaze wavelength. The output of a grating at a particular wavelength is compared with that of an aluminum mirror, and the blaze corresponds to the

wavelength of maximum efficiency:

$$\Theta = \frac{A - B}{2} \quad (19)$$

A is fixed by the construction of the instrument, whereas B varies with A according to Eq. (17). The resolving power R depends on the number of grooves in the grating N and the order of interference:

$$R = mN \quad (20)$$

R is relatively independent of wavelength. A grating with a high resolution would produce better separation of the polychromatic light to its component wavelengths.

The excitation grating is on a turntable that when rotated allows light of different wavelengths to be focused through a slit and onto the sample. The sample would fluoresce, and the light would be channeled by the emission grating, which is also on a turntable, and is focused through another slit onto the detector.

Slits

Slits help focus light onto the monochromators and the detector; they regulate the wavelength range that excites and is emitted by the sample. Smaller slit widths are more selective, producing a narrower range of spectral bandpass (bandwidth at half the peak transmittance), but a decrease in transmitted light is noted and, therefore, there is a decrease in sensitivity. There are two types of slits: fixed and variable.

Fixed slits are cut in an opaque material. A series of these are found in instruments using fixed slits; the desired slit width is obtained by using any one in the group, and the results so obtained are reproducible.

Variable slits are used more commonly than fixed slits. There are two types available: unilateral or bilateral. Both involve the use of beveled blades attached to micrometers that allow the blades to shift a specified distance, although unilateral slits have only one movable blade; the other is fixed. Bilateral slits are preferred over unilateral slits because a constant center line is maintained with the former whereas that using the latter continuously changes as the slit width changes. Variable slits are more expensive but are more convenient for routine analysis. With constant use, the micrometer mechanism and blade edges become worn, resulting in a decrease in sensitivity and reproducibility. This is not considered a major problem in analytical work unless a precise knowledge of the slit widths is imperative.

Cell Compartments and Cells

Cell compartments are usually painted flat black to minimize stray light and are covered when the fluorimeter is in use to eliminate the entrance of external light. They are usually set such that the fluorescence emitted by the sample would be at a 90° angle to the line of entry of the exciting light. This arrangement decreases the interference of stray exciting light. If the sample has strong absorption at the excitation wavelength or is a solid, concentration quenching may occur, especially if the sample has a low quantum yield. Using a front-surface configuration, with the fluorescence emitted at 30° to the line of entry of the exciting light, may solve this problem. Solid sample holders with glass, quartz, or silica windows are also available.

Fluorescence cells are made of quartz, silica, or special optical glass. Glass cells are cheaper and are suitable for use with excitation wavelengths above 330 nm. For shorter wavelengths, quartz or silica cells have to be used. If analysis is to be performed at low temperatures, cryorefrigerators or Dewar flasks (Fig. 6) are used.

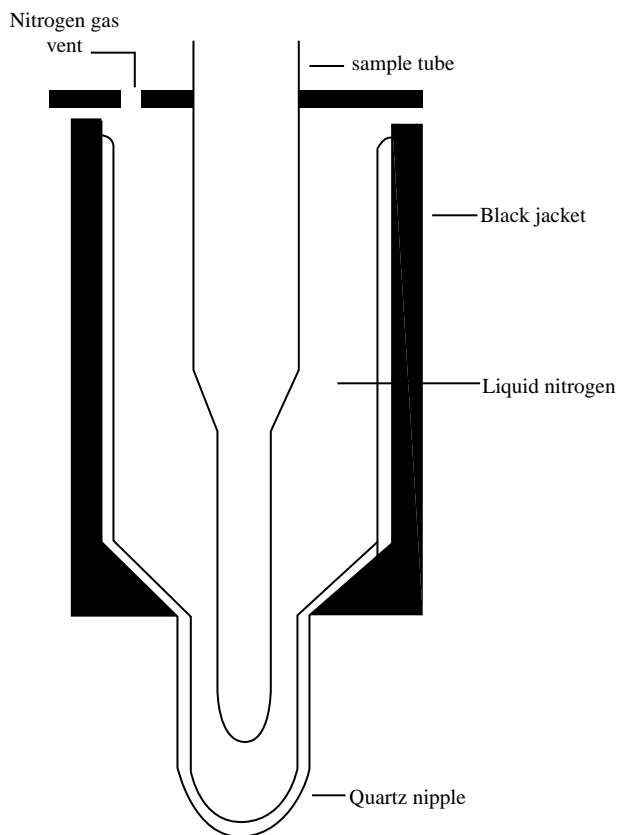


Fig. 6 Dewar flask.

At low temperatures, both fluorescence and phosphorescence may be detected when the sample in the Dewar flask is illuminated. If only phosphorescence is emitted by the sample, this setup may be used to measure phosphorescence. To isolate phosphorescence when both types of luminescence are present, a mechanical chopper with openings 180° apart is mounted in the sample compartment and rotated at a speed of several thousand revolutions per minute. This effectively cuts off the fluorescence because its lifetime is 10^{-11} – 10^{-7} s, and the emitted light would have vanished completely in one quarter of a revolution. Phosphorescence has a longer lifetime (10^{-4} to several seconds) and will be registered by the detector.

Detectors

Photomultipliers

Photomultipliers are the most widely used photodetectors in fluorescence spectroscopy. They detect very low levels of light using secondary emission from a dynode chain (Fig. 7). A photomultiplier is made up of an evacuated glass or quartz tube containing a photocathode, a series of dynodes made of electron emissive material, and an anode. Fluorescence emitted by the sample strikes the photocathode, which in turn emits photoelectrons. These electrons are accelerated toward the first dynode, and more electrons are produced, hitting the second dynode and producing even more electrons, and so forth, until the last dynode, when electrons coming from the last dynode strike the anode. The anode current flows through the load resistor, generating the signal voltage.

The amplification factor of the photomultiplier G depends on the electron gain per dynode g and the number of dynodes x :

$$G = g^x \quad (21)$$

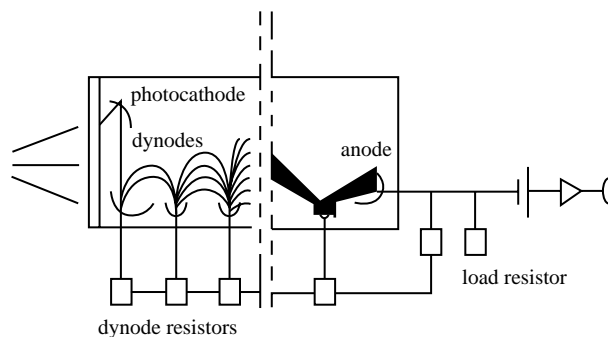


Fig. 7 Photomultiplier.

The amplification and gain per dynode of an amplifier rely on the operating voltage. Thus, photomultipliers need a very stable high-voltage power supply. The spectral response depends on the cathode type and envelope material and is not affected by the electron multiplication process.

At room temperature, the noise in photomultipliers is predominantly shot noise, which is because of the fundamentally quantized nature of light energy and electrical current. Photons arrive at the cathode randomly, even though the overall intensity of light is constant. Photoelectrons emitted from the cathode arriving at the anode would thus do so randomly, even though the long-term rate of photoelectron pulses is constant and parallel to the light intensity:

$$i_s = \sqrt{2eBG\Delta f i_a} \quad (22)$$

where $B = 1 + g^{-1} + g^{-2} + \dots + g^{-Z} = \sum_{y=0}^Z g^{-y}$ and f = electrical noise bandwidth in hertz.

The equation for shot noise current i_s owing to the total anodic current (photocurrent plus thermionic current) i_a takes into consideration the gain of the tube G . Shot noise may often be reduced by increasing the measurement time because this would increase the signal-to-noise ratio. The thermionic component of the noise may be reduced by cooling the tube. There is a minimum temperature to which the tube may be cooled. Exposure of the tube to high light levels may cause changes in the sensitivity of the tube, and photomultipliers must never be exposed to room light or irreversible damage may result. A fatigued tube may often be restored by operating it in the dark for some time at the normal operating voltage. The cathode or anode currents must not exceed the ratings specified by the manufacturer. Photomultipliers usually have a response time of less than 10^{-8} s, which is rapid enough to detect most fluorescence decay-times.

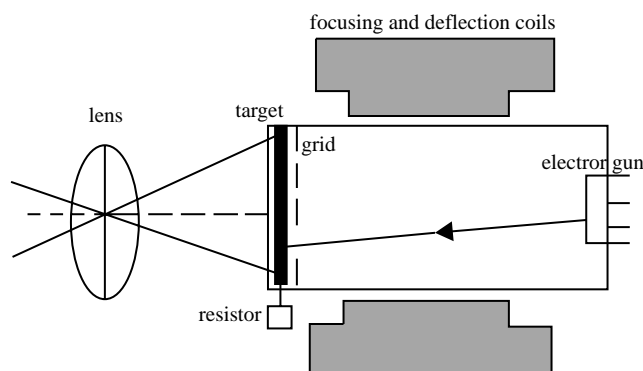


Fig. 8 Photoconductive target vidicon.

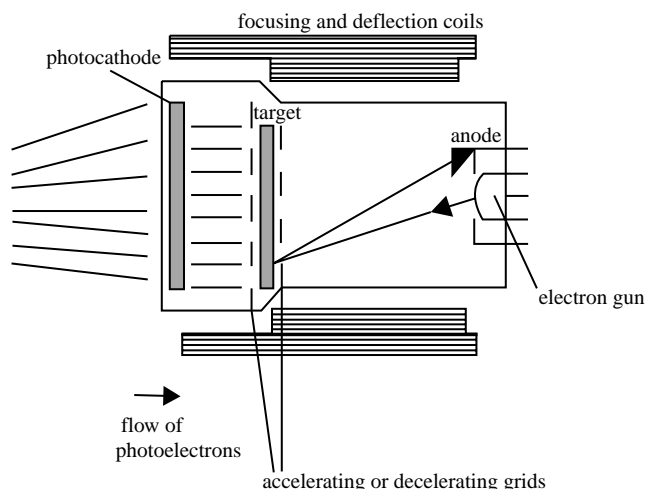


Fig. 9 Return beam vidicon.

Image detectors

Vidicons are of two types: the photoconductive target vidicon (1) and the image orthicon, or return beam vidicon (2). In the former (Fig. 8), light incident on a target previously biased by an electron beam would cause a charge separation in the photosensitive material. The carriers of the charge discharge the bias proportionate to the amount of light striking each local area, or pixel. When the raster-scanned beam returns to this region, it replaces the bias, and a signal is detected as the voltage across the load resistor is detected. The illuminated pixels cause an increase in the signal.

The return beam vidicon (Fig. 9) uses a photocathode to detect the light image. The photoelectrons are formed on a target composed of a secondary electron emitter similar to the dynodes in a photomultiplier, and secondary electrons produced are collected by a mesh, leaving the target with an image stored in the form of positively charged areas corresponding to the lit parts of the photocathode image. A low-velocity electron beam scans the target surface, and electrons are absorbed from the beam where the target is charged; the beam is scattered only in areas of little or no charge. An anode finally collects the scattered electrons. Unlike the photoconductive target vidicon, illuminated areas in the image orthicon cause a decrease in signal current.

Intensified vidicons (Fig. 10) detect photoelectrons instead of photons. There are two groups: in SEC vidicons, the target is a secondary electron conductor such as MgO, KCl, MgF, and Ag, whereas silicon-intensified target (SIT) vidicons have silicon diode arrays as their targets. SEC vidicons are approximately 10 times more sensitive

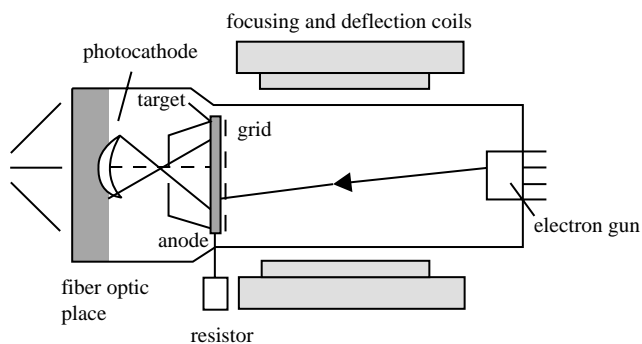


Fig. 10 Intensified target vidicon.

than ordinary vidicons, and SIT vidicons are up to 500 times more sensitive: they are capable of single-photon detection (3). The use of the latter is limited by its high noise level; however, cooling the tube and shutting the readout beam off can lower the noise by increasing the integration time (4).

Charge-coupled devices (CCD)

CCD sensors provide a higher sensitivity to light than do common detectors and photographic film, allowing for the acquisition of weaker signals. This technology has revolutionized the use of fluorescent microscopes in the biomedical sciences and has begun to have an impact on fluorescent-based analytical systems.

A CCD chip is an array of light-sensitive elements that are in fact small electronic capacitors. These capacitors are charged by the electrons generated by the emitted light. In fact, each photon that reaches the CCD array's atoms displaces some electrons that are providing a current source. These current sources are localized in small, delimited areas (the capacitors) called pixels.

Common CCD chips are composed of several pixels, for example, 192×165 , 512×512 , 1024×1024 , or more. It is easy to understand that it is physically impossible to individually access each pixel. In fact, a CCD sensor provides only one serial output, through which each capacitor can be discharged (each pixel can be read). The capacitors are disposed in lines, and there are some control gates that allow the transfer of one pixel line into the next. The last line of the array can be transferred into a horizontal shift register. This shift register allows the transfer of one pixel to the next, and the last pixel of this horizontal register is connected to the output gate.

The CCD chip has one clock entry for the vertical transfer, one for the transfer gate, and one for the shift register. Thus, read out the signal, one must first transfer

the last line into the shift register by providing a clock pulse on the transfer gate control pin. Then one has to provide one clock pulse on the shift register control pin for the reading of each point of this line. Finally, one must provide one clock pulse on the vertical transfer. The high number of pixels allows increases sensitivity and spatial resolution.

If a CCD array is used in a warm environment (20°C), the photosensitive area atoms produce a current, which produces noise. To reduce this noise, the CCD array atoms should be cooled so that their thermal excitation is reduced. The CCD is cooled by two primary processes: 1) electrical and 2) chemical.

Cooling a CCD by use of liquified gas is most efficient but is still reserved for the highest sensitivity purposes. This is primarily owing to the equipment needed to store and use such liquids. Thus, the most commonly used process is the use of a thermoelectric cooler, also known as a Peltier cell. It is a device working as a heat pump. The heat contained in one face of the cells is transferred to the other. These cells should be powered by direct current. Efficiency improvement of such a cell can be performed by cooling its hot face. To provide substantial cooling of the CCD, it is possible to use two Peltier cells in addition to a water- or glycol-cooling system. The first cell cools down the CCD, the second cools the first, and the liquid circuit evacuates the heat provided by the second Peltier cell.

Signal Processing

These systems amplify and process the electrical signal from the detector and show it in a read-out form. The electrical noise bandwidth is usually determined by the system, which in turn influences the amount of noise of the signal and determines the response time of the system.

DC amplifiers

DC amplifiers are the simplest and least expensive of the electronic measurement systems. They are most commonly found in commercially available fluorimeters. Ideally, the amplifier stage of a circuit contributes little or no noise to the system; the photomultiplier should produce only the noise associated with perfect performance of the photocathode and the electron multiplication process described above. However, the anode dark current of the photomultiplier adds to the noise in the signal, and the amplifier makes its own contribution to the total noise. It is therefore imperative to select the proper photomultiplier tube with low dark current so as to have a higher signal-to-noise ratio.

Photon counting

The current pulse from a single photon event would be noted at the anode as an approximate million electrons and a pulse width in nanoseconds. The amplifier changes the anode current pulse into a pair of voltage signals with sufficient gain and output impedance to couple it to the discriminator, which then shapes the pulse and rejects those below a set minimum threshold. The maximum count rate of less than 10^7 pulses/s determined by the pulse pair resolution of the amplifier-discriminator limits pulse counting, but methods to increase the upper limit by multilevel discrimination (5) and DC switching at high light levels (6) have been described.

Correction of Spectra

Excitation and emission spectra obtained from most instruments are dependent on the wavelength-sensitive components of the particular system. To compensate for these and the different amounts of light loss per instrument, corrections for instrumental characteristics are necessary to obtain the true spectra. The results allow intrainstrument spectral comparisons to be made. Although self-correcting instruments are commercially available, these are generally too complicated and expensive for routine analytical work.

Excitation spectra may be corrected using a constant response thermopile, a chemical actinometer, a standardized phototube, or a reference compound. The latter two methods also may be used for correcting emission spectra. By reporting fluorescence intensities in relation to reference compounds, one guards against the source and photomultiplier aging characteristics of a given instrument. Any stable fluorescent compound with broad excitation and emission bands is suitable for this purpose.

SPECIAL TOPICS

The following spectroscopic topics are viewed as current areas of analytical interest. Many review articles give a more thorough discussion of these subjects, and these are cited accordingly.

Chemiluminescence

Fluorescence initiated by chemical reactions is called chemiluminescence. It is a common phenomenon in flames in which free radicals are oxidized by molecular

oxygen but occurs rarely in solution because many of the reactive species necessary for the generation of chemiluminescence have very short lifetimes in solution. Phosphorescence as a form of chemiluminescence does not exist in solution because of quenching of the triplet state in the liquid phase.

A number of chemiluminescent reactions are known. The oxidation of 3-aminophthalhydrazide (luminol) has been studied extensively (7–14). The emission resulting from the reaction of certain oxalic acid derivatives is another outstanding example (15). When chemiluminescence occurs in living organisms, it is called bioluminescence. The best known example is observed in the firefly *Photinus pyralis*. Many review articles have been written about different aspects of bioluminescence; for overviews one should consult those by Gundermann and McCapra (16) and by Campbell (17).

Because chemiluminescence does not require photoexcitation, a lamp and excitation monochromator are unnecessary. Conventional fluorometric spectrophotometers may be used to measure chemiluminescence except that the excitation optics are not used.

Fiber Optic Sensors

The fiber optic fluorometer is a recent development that shows great promise for in situ fluorometry. In this device, the sample compound and cell are replaced by a fiber optic cable. Light from a source travels along an optically conducting fiber to its end where absorption, reflection, or scattering of light or fluorescence occurs. Fluorescence sensors comprise the largest group of fiber optic sensors; fluorescence sensors based on principles of direct fluorescence measurement, both single- and dual-wavelength (18, 19), fluorescence quenching (20, 21), or competitive binding (22) are available. Owing to the paucity of directly fluorescent analytes, most fluorescent optical sensors (FLOPS) are based on quenching by the analyte of the sensor fluorophore.

Instrumentation essentially consists of a light source, optical filters (if necessary), the fiber optic, a sensing zone (to which a fluorescent probe is usually covalently bound), and a detector. Lasers; xenon lamps; hydrogen-, deuterium-, mercury-, and halogen lamps; and light-emitting diodes (LEDs) have been used as excitation sources. Lasers provide high-intensity excitation with a narrow bandwidth and are especially suited for remote sensing where light losses are large. However, they are quite expensive and require a heavy power supply. The most promising light sources are the LEDs; these run at a low voltage and current, are small, have long life spans,

and do not generate heat. However, they do have the disadvantage of working only from the IR region down to the blue. Fiber optic fluorescence spectrometers may be operated in a continuous or pulsed excitation mode. The latter allows for the application of time-resolved fluorimetry. The material of the optical fiber would determine the wavelength used. Fused silica, glass, and plastic fibers have all been used. Silica can be used from the ultraviolet range down to 220 nm, but the fibers are expensive. Glass is suitable for use in the visible region and is reasonably priced. Plastic fibers are the lowest in cost but are limited to use above 450 nm. Measurements may be processed directly using analog and digital circuits or a microprocessor. Photomultipliers, photodiodes, photoconductor cells, photovoltaic light detectors, and CCDs have all been used as detectors.

Fiber optic sensors are less expensive, more rugged, and smaller than electrodes, which currently dominate in sensor-based analysis. In the future, we may see the former replacing the latter in various areas of analytical and clinical chemistry. Fields of application include groundwater and pollution monitoring; process control; remote spectroscopy in high-risk areas with radioactive, explosive, biological, or other hazards; titrimetry; and biosensing. Different aspects of optical sensing have been the subject of several studies (23–25). The field of FLOPS is currently one of the most active in analytical spectroscopy.

Luminescence Detection in Separation Methods

A limitation of the application of luminescence spectroscopy to the analysis of real samples is its lack of specificity owing to similarities in spectral bandshapes and spectral positions of the luminescence spectra of many compounds. An obvious solution to this problem is the separation of the analytical sample's interfering constituents from each other before quantitation by fluorescence. High-performance liquid chromatography (HPLC) and related separation methods can be coupled to fluorescence spectroscopy to take advantage of the sensitivity of the spectroscopic method and the specificity of the separation method.

Luminescence affords a very sensitive means of detection in flowing systems such as HPLC, electrophoresis, flow injection, and flow cytometry. HPLC fluorescence detectors are similar in operation to conventional fluorimeters. Most fluorescence detectors use filters for crude monochromation. Filters pass light in a wider band than do monochromators. This favors spectral sensitivity because more light excites the sample and is collected by the detector. Grating monochromators, on the other hand,

favor selectivity. The fluorimetric detector is susceptible to the usual interferences that hinder fluorescence measurements, namely, background fluorescence and quenching.

In the operation of a fluorimetric HPLC detector, the light from a UV or visible source is monochromated to some degree and focused on the cell that is on line with and receiving the effluent of the separatory column. Fluorescence is emitted by the sample in all directions so that the emitted light can be measured with the detector at right angles to the path of the exciting light and the direction of fluid flow. Stray or scattered excitation light is then blocked from the detector by a filter, and the emitted energy is then measured using a suitable detector. It is important to recognize that the solvent has a strong effect on the intensity of fluorescence. The solvent also scatters light owing to the Raman effect. This light is of slightly lower energy than the exciting light. For this reason, the emission wavelength is selected to be longer than that of the scattered Raman light. Flow cells have been designed that can be used for both fluorescence and absorption measurements. Quantitation can be considerably improved by simultaneously monitoring the absorbance and fluorescence signals that extend the linear dynamic range for fluorescent samples. At high sample concentrations in which the absorbance at the excitation wavelength is greater than 0.02, fluorescence intensity becomes non-linear with concentration. At these higher concentrations, however, the light absorbance of the sample is often measurable and linear with concentration. Fluorescence, however, is often detectable at concentrations from 10^3 to 10^9 times lower than those in which absorbance is detectable.

The laser as an excitation source for luminescence detection in chromatography has become popular because of its high excitation power and its highly focused beam. The high degree of collimation of the laser beam has made it an excellent excitation source with the microbore cells, and tubing used is capillary zone electrophoresis. A typical laser beam can be focused into a spot a few microns in diameter. With laser excitation and fluorescence detected, capillary zone electrophoresis analyte concentrations below 1 attomole (at nanomole concentrations) have been measured. The advantages offered by the power and high degree of collimation of the laser as an excitation source can be exploited to measure the properties of a variety of particles including cells, algae, bacteria, larvae, etc., that either show native fluorescence or can form fluorescent adducts. A well-defined flow of isolated particles is achieved by using hydrodynamic focusing. The particle flux is interrogated by a laser beam, which excited one particle at a time. Single-cell passage is guaranteed by adjusting the two flows; thus, collision between two cells

(particles) is avoided. The method allows high-speed analysis of the cells and is called flow cytometry or fluorescent cell-sorting.

Time and Phase-Resolved Fluorimetry

The spectroscopy described thus far is based on the measurement of the intensity of fluorescence produced under steady-state conditions of excitation. Steady-state fluorimetry is derived from the excitation of the sample with a continuous beam of exciting radiation. The lamps and the power supplies used in conventional fluorimeters are sources of continuous radiation. After a short period of initial excitation of the sample, a steady state is established in which the rate of excitation of the analyte is equal to the sum of the rates of all processed, deactivating the lowest excited singlet state including fluorescence. When the steady state is established, the observed fluorescence intensity becomes time-invariant and produces the temporally constant signal that is measured by the photodetector. With the development of modern electro-optical components, it has become possible to excite a potentially fluorescent sample with a pulsed flash lamp that emits its radiation in bursts of 2–10 ns in duration with approximately 0.2 ms between pulses. Pulsed lasers generate even shorter duration pulses (typically down to 1 ps). A fluorescent sample excited by a single pulse will decay exponentially until the next pulse again excites the sample. The fluorescence from the sample excited by the pulsed source can be represented, after detection, as a function of time on a fast sampling oscilloscope used in conjunction with a multichannel pulse analyzer. The former approach is called pulse fluorimetry and the latter time-correlated single-photon counting. In either case, fluorescences with decay-times much longer than the lamp-pulse characteristics can be analyzed from a semilogarithmic plot of fluorescence intensity against time, which will yield a straight line (or a series of overlapping lines if several fluorophores have comparable but not identical decay-times) whose slope is proportional to the decay-time and whose vertical axis intercept can be compared with that of a standard solution of the fluorophore for quantitative analysis. If, however, the lamp pulse-time and the decay-time of the fluorophore are comparable, the lamp characteristics must be subtracted from the observed signal to obtain the fluorophore's decay characteristics. This is usually accomplished by using a computer to deconvolute the composite temporal characteristics of the lamp and the fluorophore output.

The pulsed source (time-resolved) method then distinguishes between the emissions of several fluorescing

species by using their decay-times rather than their fluorescence intensities. This means that several strongly overlapping fluorescences such as those of catecholamines can be quantitated simultaneously without chemical or mechanical separation.

Time-resolved fluorimetry is also useful for the elimination of interferences from stray light caused by Rayleigh and Raman scatter. The latter phenomena occur on a time scale of 10^{-14} – 10^{-13} because they have a much shorter duration than lamp or laser pulses. This light associated with them can be eliminated from the signal that ultimately reaches the detector. Data taken on one time-resolved fluorometer can be directly compared with that from another. This is not possible in steady-state fluorimetry unless correcting wavelength variable instrumental response has been affected.

Phase fluorimetry is another useful fluorimetric technique for the determination of substances with overlapping fluorescence spectra. In phase fluorometry, the phase angle between the lamp pulse and the emission of fluorescent light allows for discrimination between fluorescences of different origin. For additional reading on this subject, see the excellent review by Demas (26).

With the introduction of powerful computers and advanced programs, traditional fluorescent devices are pushing the fringes of detection. It is now possible to observe events at a single molecule level. Information from these studies is refining our understanding of biochemistry and cell trafficking of molecules (28).

REFERENCES

1. Lubszynski, H.G. Review of TV Camera Tubes and Electron Optics. *Adv. Electron. Electron. Phys.* **1979**, 52, 11.
2. Redington, R.W. *Photoelectronic Imaging Devices*; Biberman, L.M., Nudelman, S., Eds.; Plenum Press: New York, 1971; 2, 193–202.
3. Milch, J. Slow Scan Sit Detector for X-Ray-Diffraction Studies using Synchrotron Radiation. *IEEE Trans. Nucl. Sci.* **1979**, NS-26, 338.
4. Gruner, S.M.; Milch, J.R.; Reynolds, G.T. Evaluation of Area Photon Detectors by a Method Based on Detective Quantum Efficiency (DQE). *IEEE Trans. Nucl. Sci.* **1978**, NS-25, 562–565.
5. Gustafson, T.L.; Lytle, F.E.; Tobias, R.S. Sampled Photon Counting with Multilevel Discrimination. *Rev. Sci. Instrum.* **1978**, 49, 1549–1550.
6. Marino, D.F.; Ingle, J.D., Jr. Microprocessor-Based Data Acquisition System for Chemiluminescence Measurements. *Anal. Chem.* **1981**, 53, 1175–1179.
7. Gundermann, K.D. Chemiluminescence in Organic Compounds. *Agnew. Chem. Int. Ed. Engl.* **1965**, 4, 566–572.

8. McCapra, F.Q. Chemiluminescence of Organic Compounds. *Rev. Chem. Soc.* **1996**, *20*, 485–572.
9. Haas, J.W., Jr. Chemiluminescent Reactions in Solutions. *J. Chem. Educ.* **1967**, *44*, 396–402.
10. White, E.H.; Roswell, D.F. The Chemiluminescence of Organic Hydrazides. *Acc. Chem. Res.* **1970**, *3*, 54–62.
11. Cormier, M.J.; Hercules, D.M.; Lee, J., Eds.; *Chemiluminescence and Bioluminescence*; Plenum Press: New York, 1973.
12. Gundermann, K.D. Recent Advances in Research on the Chemiluminescence of Organic Compounds. *Top. Curr. Chem.* **1974**, *46*, 61–65.
13. Roswell, D.F.; White, E.H. *Methods of Enzymology*; DeLuca, M.A., Ed.; Academic Press: New York, 1978; 57, 409.
14. Rauhut, M.M. *Encyclopedia of Chemical Technology*, 3rd Ed.; John Wiley & Sons: New York, 1979; 5, 416.
15. De Jong, G.J.; Kwakman, P.J.M. Chemiluminescence Detection for High Performance Liquid-Chromatography of Biomedical Samples. *J. Chromatogr.-Biomed.* **1989**, *492*, 319–343.
16. Decker, K.A.; Hinkkanen, A. Luminometric Determination of Flavin Adenindinucleotide. *Methods Enzymol.* **1986**, *122*, 185–192.
17. Kricka, L.J. Chemiluminescence and Bioluminescence. *Anal. Chem.* **1999**, *71*, 305R–308R.
18. Saari, L.A.; Seitz, W.R. Immobilized Morin as Fluorescence Sensor for Determination of Aluminum (III). *Anal. Chem.* **1983**, *55*, 667–670.
19. Zhujun, Z.; Seitz, W.R. Optical Sensor for Oxygen Based on Immobilized Hemoglobin. *Anal. Chem.* **1986**, *58*, 220–222.
20. Lubbers, D.W. Opitz, N. Proceedings of the International Meeting on Chemical Sensors Fukuoka Japan. Elsevier: Amsterdam, 1983; 609–619.
21. Soper, S.A.; Warner, I.M.; McGown, L.B. Molecular Fluorescence, Phosphorescence, and Chemiluminescence Spectrometry. *Anal. Chem.* **1998**, *70*, 477R–494R.
22. Schultz, J.S. Affinity Sensor-A New Technique for Developing Implantable Sensors for Glucose and Other Metabolites. *Diabetes Care* **1982**, *5*, 245–253.
23. Seitz, W.R. Chemical Sensors Based on Fiber Optics. *Anal. Chem.* **1984**, *56*, 16A–34A.
24. Peterson, J.I.; Vurek, G.G. Fiber-Optic Sensors for Biomedical Applications. *Science* **1984**, *224*, 123–127.
25. Seitz, W.R. Optical Sensors for Clinical Applications. *J. Clin. Lab. Anal.* **1987**, *1*, 313–316.
26. Demas, J.N.; Schulman, S.G., Eds.; *Molecular Luminescence Spectroscopy: Methods and Applications*; Part 2 Wiley-Interscience: New York, 1988; Chap. 2.
27. Weiss, S. Fluorescence Spectroscopy of Single Biomolecules. *Science* **1999**, *283*, 1676–1683.
28. Pederson, T. Movement and Localization of RNA in the cell nucleus. *FASEB J.* **1999**, *13*, S238–S242.
29. Hercules, D.M. *Fluorescence and Phosphorescence Analysis*; Wiley: New York, 1966.
30. Guilbault, G.G. *Fluorescence*; Marcel Dekker, Inc.: New York, 1968.
31. Parker, C.A. *Photoluminescence of Solutions*; Elsevier: Amsterdam, 1968.
32. Udenfriend, S. *Fluorescence Assay in Biology and Medicine*; Academic Press: New York, 1962, 1968; 1, 2.
33. Becker, R. *Theory and Interpretation of Fluorescence and Phosphorescence*; Wiley: New York, 1970.
34. Pesce, A.; Rosen, G.; Pasby, T. *Fluorescence Spectroscopy*; Marcel Dekker, Inc.: New York, 1971.
35. White, C.; Argauer, R. *Fluorescence Spectrometry: A Practical Approach*; Marcel Dekker, Inc.: New York, 1971.
36. Winefordner, J.D.; Schulman, S.G.; O'Haver, T.C. *Luminescence Spectrometry in Analytical Chemistry*; Wiley: New York, 1972.
37. Zander, M. *Phosphorimetry*; Academic Press: New York, 1968.
38. Guilbault, G. *Practical Fluorescence*; Marcel Dekker, Inc.: New York, 1973.
39. Schenk, G. *Absorption of Light and Ultraviolet Radiation: Fluorescence and Phosphorescence Emission*; Allyn and Bacon: Boston, 1973.
40. Schulman, S.G. *Fluorescence and Phosphorescence Spectroscopy: Physicochemical Principles and Practice*; Pergamon: Oxford, 1977.
41. Wehry, E.L., Ed. *Modern Fluorescence Spectroscopy*; Plenum: New York, 1976, 1981; 1–2, 3–4.
42. Lakowicz, J.R. *Principles of Fluorescence Spectroscopy*, 2nd Ed.; Plenum: New York, 1983, 1999.
43. Schulman, S.G., Ed. *Molecular Luminescence Spectroscopy: Methods and Applications*; Part 1, Part 2, Part 3 Wiley-Interscience: New York, 1985, 1988, 1993.
44. Sharma, A.; Schulman, S.G. *Introduction to Fluorescence Spectroscopy*; Wiley-Interscience: New York, 1999.
45. Ichinose, N.; Schwedt, G.; Schnepel, F.M.; Adachi, K. *Fluorescence Analysis in Biomedical Sciences*; Wiley Interscience: New York, 1991.

SPECTROSCOPIC METHODS OF ANALYSIS—INFRARED SPECTROSCOPY

Marilyn D. Duerst

University of Wisconsin-River Falls, River Falls, Wisconsin, U.S.A.

INTRODUCTION

Infrared (IR) spectroscopy refers broadly to the study of the interaction between matter and infrared radiation. Infrared radiation falls in the region between the visible and microwave parts of the electromagnetic spectrum, with wavelengths from 0.7 to 500 μm . Infrared spectroscopists often express this region of the spectrum in units of wavenumbers (symbol cm^{-1}) which refers to the number of waves per centimeter. This region (14,000 to 20 cm^{-1}) is usually subdivided into three regions: from 14,000 to 4000 cm^{-1} is called “near-infrared,” from 4000 to 400 cm^{-1} the “mid-infrared,” and from 400 to 20 cm^{-1} the “far-infrared.” The mid-infrared region is widely used in the analysis of drugs and pesticides, and will be the focus of discussion here.

Chemical compounds absorb infrared radiation when there is a dipole moment change (in direction and/or magnitude) during a molecular vibration, molecular rotation, or molecular rotation-vibration. Absorptions are also observed with combinations, differences or overtones of molecular vibrations. A specific type of molecule is limited in the number of vibrations and rotations it is allowed to undergo. Therefore, each chemical compound has its own specific set of absorption frequencies and thus exhibits its own characteristic IR spectrum. This unique property of a compound allows the organic chemist to identify and quantify an unknown sample. (A special infrared technique called vibrational circular dichroism (VCD) is required to distinguish optical isomers).

Certain functional groups in a molecule (e.g., hydroxyl, carbonyl, and amine) absorb IR radiation and exhibit absorption bands at characteristic frequencies regardless of the structure of the rest of the molecule. These bands are termed “group frequencies.” They are predictable and allow the analyst to deduce important structural information about an unknown molecule. An IR spectrum can be rapidly recorded for any phase, i.e. solid, liquid, or vapor. By coupling IR spectroscopy with other analytical techniques such as nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry, the organic

chemist can determine the structure of an unknown compound.

The introduction of the Fourier transform infrared (FTIR) spectrometer has revolutionized the entire field of IR spectroscopy. The typical FTIR spectrometer has a number of advantages over the conventional dispersive IR spectrometer in terms of the resolution of bands in the spectrum and the speed of acquiring the spectrum. Because of continuous improvements in sampling techniques, analysis software, and instrumentation hardware, the FTIR spectrometer has become increasingly more valuable in many areas from fundamental research to quality control to on-line process control. Recently, much interest has been focused on coupling the FTIR spectroscopic technique with chromatographic and other analytical techniques such as gas chromatography (GC-FTIR), GC-mass spectrometry (GC-MS-FTIR), high-performance liquid chromatography (HPLC-FTIR), supercritical fluid chromatography (SFC-FTIR), thin-layer chromatography (TLC-FTIR), and thermogravimetric analysis (TGA-FTIR). These methods, often termed “hyphenated techniques,” employ IR spectroscopy as a means of detection to identify an unknown substance in a chromatographic eluent. Used in conjunction with a proper interfacing and sampling techniques, the limit of detection in some studies can also be in the picogram range, especially with the FTIR microscope.

This general approach will discuss the fundamental principles, instrumentation, and current applications of IR spectroscopy. Numerous reference materials, covering all aspects of infrared spectroscopy, have been published over the years (1–3). For a more in-depth understanding of the theory and instrumentation, readers can refer to the bibliography and references listed at the end of this article.

THEORY

The electromagnetic spectrum includes radiation from cosmic rays to radio waves, with wavelengths ranging from 10^{-9} nm to longer than 1000 km. In progression from

short to long wavelengths, the types of radiation included in the spectrum are gamma rays, X-rays, far-, middle-, and near-ultraviolet rays, visible light, infrared rays, and microwaves.

Electromagnetic radiation may be conceptualized as a wave traveling at the speed of light (c) in a vacuum, where c is approximately 3.0×10^8 m/s. The frequency (ν) is related to the wavelength by Eq. 1.

$$c = \nu\lambda \quad (1)$$

The wavelength, in units of cm or μm , is defined as the distance between peaks or troughs of the wave. The frequency of the wave is the number of peaks passing a fixed point per unit of time. The unit of frequency is the Hertz, i.e., the cycles or waves per second.

In infrared spectroscopy, the wavenumber, $\bar{\nu}$, is often used instead of frequency, and is related to wavelength according to Eq. 2.

$$\bar{\nu} = 1/\lambda \quad (2)$$

The wavenumber, in units of cm^{-1} , is the number of waves per centimeter. It is directly proportional to the frequency, but is not the same as frequency, as shown in Eq. 3.

$$c = \nu\lambda, \text{ thus } \bar{\nu} = \nu/c \quad (3)$$

The relationship between wavenumber (in cm^{-1}) and wavelength (in μm) is given by Eq. 4.

$$\bar{\nu} = 10^4/\lambda \quad (4)$$

INFRARED SPECTROSCOPY

Electromagnetic radiation may also be considered as photons traveling at the speed of light. The energy of a photon (E) is related to the wavelength (λ), frequency (ν), and velocity of light (c) by Eq. 5, where h is Planck's constant ($h = 6.626 \times 10^{-34}$ J/s).

$$E = h\nu = hc/\lambda \quad (5)$$

In all spectroscopic techniques, the interaction of radiation with the molecule results in the transfer of energy to the molecule. The energy absorbed by the molecule is quantized, that is, the absorption is specific to the frequency of the radiation. Depending on the energy of the photons, the effect of an increase in energy in the molecule could cause reorientation of nuclear or electron spin states (if the molecule were in a magnetic field), changes in the energy of the valence electrons, changes in the vibrational-rotational energy, ejection of inner electrons from the molecule, or changes in the nucleus.

The absorption of IR radiation changes the rotational and/or vibrational energy states of the molecules. The IR absorption spectrum of a molecule may appear as a set of broad bands rather than discrete lines because each vibrational energy change is accompanied by a number of rotational energy changes. Three main types of absorption involving vibrations include the fundamental vibration, overtones, and combinations. The fundamental vibrational absorption bands usually appear in the mid-infrared region. It is important to realize that a molecule absorbs incident infrared radiation only when the molecular vibrational frequency is the same as the frequency of the incident radiation, and if the vibration of the atoms results in a change in the dipole moment. A symmetrical vibration does not result in a change of dipole moment; therefore, no infrared absorption will occur.

Two main types of fundamental vibrations are stretching and bending. Stretching vibrations usually appear at higher frequencies than bending vibrations. A stretching vibration consists of the movement of the atom along the bond axis, which results in changes in bond length. A bending vibration consists of a change in bond angles between bonds with a common atom. Fig. 1 shows the typical stretching and bending vibrational modes of a $-\text{CH}_2-$ group within a molecule.

The intensity of vibrational absorption is proportional to the square of the magnitude of the dipole moment change. For a small dipole-moment change, the absorption band is weak, whereas a large, permanent dipole moment, such as that in a carbonyl group, exhibits strong absorption in the infrared when the carbon-oxygen distance is changed.

The stretching frequency is affected by the masses of the atoms and the strength of the bonds (i.e., the "stiffness" of the chemical bond) connecting them. Their relationship can be approximated by Eq. 6.

$$\bar{\nu} = (1/2\pi c)(f/u)^{1/2} \quad (6)$$

where $\bar{\nu}$ is the vibrational frequency (cm^{-1}), c the velocity of light (cm/s), f the force constant of the bond (mN/m or dyn/cm), u the reduced mass of the two atoms, m_1 and m_2 , or $u = (m_1 m_2)/(m_1 + m_2)$.

The spatial orientation of atoms with respect to each other also affects their vibrational frequency. For example, the coupling or interaction of two fundamental vibration groups of similar frequencies in close proximity within a molecule and the inter- or intramolecular hydrogen bonding affect the vibrational frequencies of the molecule.

In general, the vibrational frequency is higher for atoms with a smaller reduced mass and a larger force constant. Vibrations of groups where one atom is hydrogen have a higher frequency than those with other, heavier atoms.

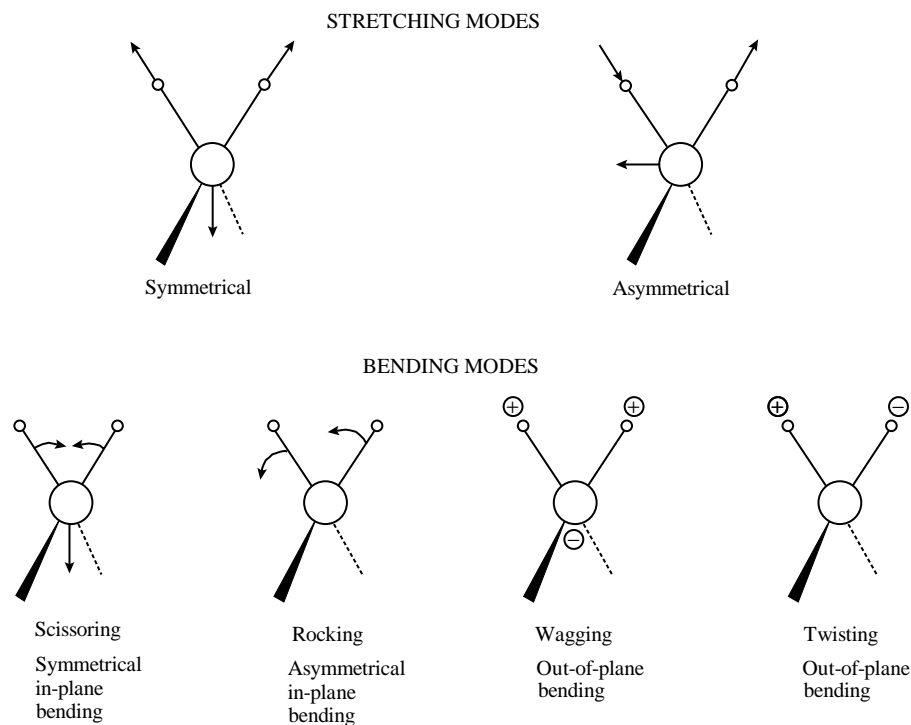


Fig. 1 Stretching and bending vibrations of a $\text{—CH}_2\text{—}$ group.

For example, if the hydrogen atom in the C—H group is replaced by a deuterium atom, the vibrational frequency of C—D is lower than the vibrational frequency of C—H by a factor of about 1.414. The force constant is proportional to the bond strength (and thus to the bond order) between atoms. Therefore, for groups with atoms of the same mass, the vibrational frequency for triple-bond groups is about three times higher than that for double-bond groups. The vibrational frequency for the double-bond group is twice as high as that for the single-bond group.

Absorption bands that are attributed to overtone and combination vibrations are also observed in the IR spectrum of polyatomic molecules. Overtone vibrations occur at frequencies of approximately integral multiples of the fundamental frequencies. Combination vibrations appear at frequencies that are the sum or the difference of the frequencies of two or more fundamental vibrations. Overtone and combination bands are much less intense than fundamental bands.

Fermi resonance, which is the interaction between fundamental vibrations and overtone or combination vibrations, is also common in IR spectra. The result of Fermi resonance is the formation of two new vibration modes with one frequency higher and the other one lower than that observed when the interaction is absent. The unique quality of an IR absorption spectrum for a molecule

is partly attributed to overtone bands, combination bands, and Fermi resonance bands.

INSTRUMENTATION

Two main classes of infrared instrumentation are dispersive and nondispersive. A dispersive instrument uses a prism or grating to separate light into its frequencies. An IR spectrum generated by a dispersive instrument is normally a plot of wavenumber (cm^{-1}) or wavelength (μm) versus percent transmittance. In a nondispersive instrument, the IR radiation does not pass through a prism or grating, but through an interference filter or an interferometer, as in a Fourier transform infrared spectrometer, and is then collectively sent through the sample. An interference pattern generated by absorption within the sample is called an interferogram. It can be converted into a conventional IR spectrum by application of a mathematical operation (the Fourier transform) to the data. Using modern computer technology, this mathematical manipulation can be completed in a matter of seconds.

A typical IR spectrometer consists of the following components: radiation source, sampling area, monochromator (in a dispersive instrument), an interference filter

or interferometer (in a nondispersive instrument), a detector, and a recorder or data-handling system. The instrumentation requirements for the mid-infrared, the far-infrared, and the near-infrared regions are different. Most commercial dispersive infrared spectrometers are designed to operate in the mid-infrared region ($4000\text{--}400\text{ cm}^{-1}$). An FTIR spectrometer with proper radiation sources and detectors can cover the entire IR region. In this section, the types of radiation sources, optical systems, and detectors used in the IR spectrometer are discussed.

RADIATION SOURCES

Infrared radiation is generated by electrically heating a source to a certain temperature. Several radiation sources (sometimes called “black body” sources) are commonly used for mid-infrared spectrometry. Table 1 summarizes the properties of each source, including the sources used for the near-IR and far-IR regions. Most sources have a maximum energy output at a certain wavelength range when heated at their optimum temperature. The energy of the radiation decreases gradually at longer wavelengths. In a dispersive instrument, however, widening the slit to allow more radiation to reach the detector can sometimes compensate for this effect.

Monochromator

A monochromator, mainly used in a dispersive instrument, is an optical device capable of separating infrared radiation into its constituent wavelengths. The monochromator

normally consists of several components, including mirrors for collimating, focusing, or changing the direction of the radiation beam, and filters for attenuating the radiation and reducing stray radiation or eliminating unwanted grating orders. The most important component is the prism or the diffraction grating. By rotating the prism or grating, different wavelengths pass across a fixed slit, thereby allowing the entire wavelength range to be scanned. In most cases, the grating is preferable to the prism because the prism has several disadvantages, including lower dispersion (resulting in lower resolving power) and the introduction of unwanted reflection off its front and back surfaces.

Interferometer

The interferometer used in a nondispersive instrument is a device that divides the beam of radiation into two paths and recombines the two beams after a path difference has or has not been introduced. The basic concept of the interferometer was introduced by Michelson almost a century ago (Fig. 2). It consists of a stationary mirror, a moving mirror, and a beam splitter. The radiation from the infrared source is divided at the beam splitter; half the beam is passed to a fixed mirror and the other half is reflected to the moving mirror. The two beams are later recombined at the beam splitter and passed through the sample to the detector. For any particular wavelength, the two beams interfere constructively or destructively, depending on the difference between the optical paths of the beams in the two arms of the interferometer. The sample placed in the beam path absorbs radiation of a certain wavelength. The recorded interferogram is the sum

Table 1 Radiation sources for infrared spectrometers

Source	Composition	Operating temperature (°C)	Infrared range	Remarks
Nichrome coil	Nichrome	1100	Mid-IR	Air cooled, reliable, inexpensive, low temperature, less intense than other sources
Globar	Silicon carbide	1300	Mid-IR or Far-IR to 2000 cm^{-1}	Cooled by water, emission can be down to 80 cm^{-1}
Nernst glower	Oxides of zirconium, thorium, and yttrium	1500	Mid-IR $10,000\text{--}500\text{ cm}^{-1}$	Poor emission at high wavenumbers, brittle, preheating required
High-pressure mercury arc	Mercury		Far-IR	Best for range between 200 and 10 cm^{-1}
Tungsten lamp	Tungsten		Near-IR	Best for between $33,000$ and 400 cm^{-1}

of all waves except those absorbed by the sample. By a Fourier transformation of the interferogram, a conventional IR spectrum is obtained.

Detector

The detector converts infrared radiation into an electrical signal. The two main classes of detectors are thermal and quantum detectors. The heating caused by impinging infrared radiation changes some physical properties of the thermal detector itself. In quantum detectors, the quantum nature of infrared radiation changes the detector's electrical properties.

The responsivity (E) or specific detectivity (D^*) and the noise equivalent power NEP (W_n), are often used to measure the sensitivity of a detector. The responsivity depends on the wavelength of the radiation and the temperature of the detector. The NEP, also called minimum detectable power, is the quotient of detector noise (N) divided by voltage responsivity (E). The D^* is the reciprocal of NEP, thus $W_n = N/E$ and $D^* = 1/W_n$. A more sensitive detector has a smaller NEP and larger D^* , which results in less noise and a faster response time.

Thermal Detectors

Thermal detectors absorb infrared radiation and convert it to heat, resulting in a temperature change; a temperature-dependent detector property is then measured. The properties of the detector affected by the infrared radiation, depending upon the type of detector, include

the expansion of a liquid, solid, or gas, electrical resistance, voltage, and electric polarization. Thermal detectors respond to radiation over a wide range of wavelengths. Since the response depends upon a change in temperature, such detectors are usually slow, typically responding in about 0.01–0.1 s. Table 2 summarizes the characteristics of various types of thermal detectors.

Quantum Detectors

Quantum detectors are usually made of semiconductor materials or mixtures. Some commonly used quantum detectors are made of lead sulfide (PbS), lead selenide (PbSe), indium antimony (InSb), or mercury cadmium telluride (MCT, HgTe–CdTe). The absorption of infrared radiation in quantum detectors excites electrons from a non-conducting state into a conducting state, resulting in a change in the current or voltage. Since the excitation of the electrons to a higher energy state is quantized, the detector exhibits a sharp cutoff frequency toward the far-infrared. Cooling is generally required to avoid thermal agitation, which produces internal electrical noise. The temperature to which the detector must be cooled depends on its sensitivity; for example, an MCT detector is usually cooled to liquid nitrogen temperature (77 K), whereas PbS and PbSe can be operated at a temperature just below ambient. Quantum detectors have much higher sensitivity and faster response times than thermal detectors; they thus are most often used in high-performance FTIR spectrometers. The response time for quantum detectors is measured in microseconds rather than in milliseconds. The sensitivity of quantum detectors is strongly dependent upon the composition of the semiconductors.

DISPERSIVE INFRARED SPECTROMETERS

Dispersive IR spectrometers can be either single-beam or double-beam; usually they are double-beam instruments. The use of a double-beam instrument significantly reduces the problem of possible interference from carbon dioxide and water vapor in the atmosphere during analysis. The absorption due to a solvent may also be nearly cancelled by placing an equivalent pathlength solvent cell in the reference beam.

A typical double-beam spectrometer is shown in Fig. 3. Radiation from the source is split into the reference beam and the sample beam. Each beam passes through a comb-shaped attenuator that regulates the beam intensity. The two beams are alternately sent through the slit into the

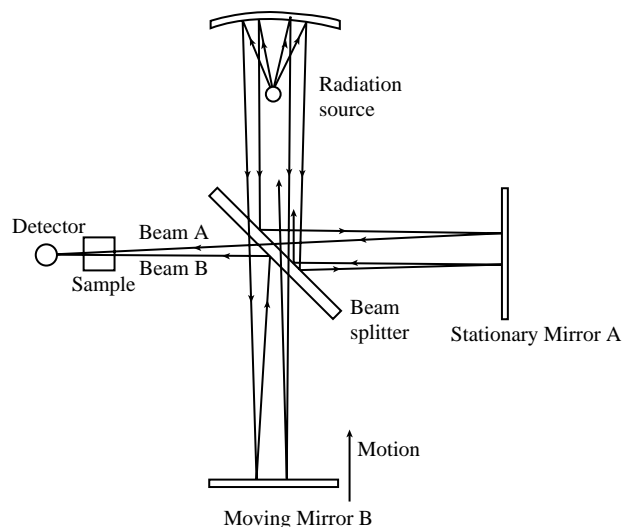


Fig. 2 Diagram of an interferometer. (Adapted from Ref. 23.)

Table 2 Thermal detectors for infrared spectrometers

Detector	Composition	Physical property affected	Remarks
Thermocouple and thermopile	Bismuth and antimony	Voltage at the junction of the two metals	Simplest and most direct measurement; slow response (30 ms)
Thermistor or thermistor bolometer	Oxides of Mn, Co, and Ni	Resistance changes	Slow response
Golay pneumatic detector	Xenon gas	Gas expansion, combined with the moving of a diaphragm, changes the illumination of a photocell	Extends to far-IR region with proper window; 15 ms response time; sensitive to intense light; microphonic disturbances; diaphragm wears out easily
Pyroelectric detector	Polarized pyroelectric materials ^a at temperatures below its Curie point	Electric polarization produces voltage changes	Low noise and fast response; used for fast-scan FTIR

^aSuch as triglycine sulfate (TGS), deuterated triglycine sulfate (DTGS), LiTaO₃, LiNbO₃.

monochromator by the chopper motor, which typically rotates at a frequency of 11 or 13 Hz. After the radiation has been dispersed into its constituent wavelengths by the monochromator, the beam is passed through the other slit and then onto the detector. The detection in a double-beam spectrometer is based on the “optical null” principle. In the absence of a sample, the two beams reach the detector with equal intensity and the recording pen is at 100% transmittance. When the sample is present, the beams are not of equal intensity. The difference in the intensity is amplified by an amplifier, which controls the pen motor. The attenuator comb in the reference beam is driven by the pen motor to adjust the intensity of the reference beam and make it equal to the sample beam. A pen moves up and down a chart to record the percent transmittance of the sample beam, and thus monitors the extent to which the attenuator must be driven. The scan motor rotates the prism or grating in the monochromator to scan the spectrum, simultaneously moving the chart under the pen. The chart records the percent transmittance versus the wavenumber (or wavelength). An alternative to the comb is to plot the difference in voltage between the reference and sample beams.

NONDISPERSIVE INFRARED SPECTROMETERS

One type of nondispersive spectrometer employs filters to isolate the wavelength desired; the other type uses an

interferometer, i.e., the Fourier transform infrared (FTIR) spectrometer. Infrared analyzers utilizing simple filters to select the desired wavelength range for analysis are very useful in industrial on-line process control for monitoring a gas or liquid stream. This type of IR analyzer has high sensitivity and selectivity. A more sophisticated model (with microprocessor) can perform multicomponent analyses, calibration, and data reduction; it can also generate reports. The instrumental design of infrared analyzers is beyond the scope of this article.

Fourier transform infrared spectrometers can be single- or double-beam. Most commercial FTIR spectrometers are of single-beam design (Fig. 4). Double-beam FTIR spectrometers (Fig. 5) are less common and only a few are commercially available, and are mostly used for research purposes rather than routine tasks. A double-beam instrument is designed to compensate for atmospheric interference. In most modern IR spectrometers the optical components are manufactured in a sealed and desiccated compartment with the goal of reducing water and carbon dioxide interferences. A laser beam generates a reference signal to monitor the sampling rate and the velocity of the moving mirror, resulting in very accurate wavelengths in the spectra.

The FTIR technique has several advantages over the conventional dispersive technique. It provides a higher accuracy in wavenumber values, a better resolution, a better signal-to-noise ratio in a shorter time, and faster response time, making it useful for in-process or on-line monitoring. Dispersion or filtering of the radiation is not required.

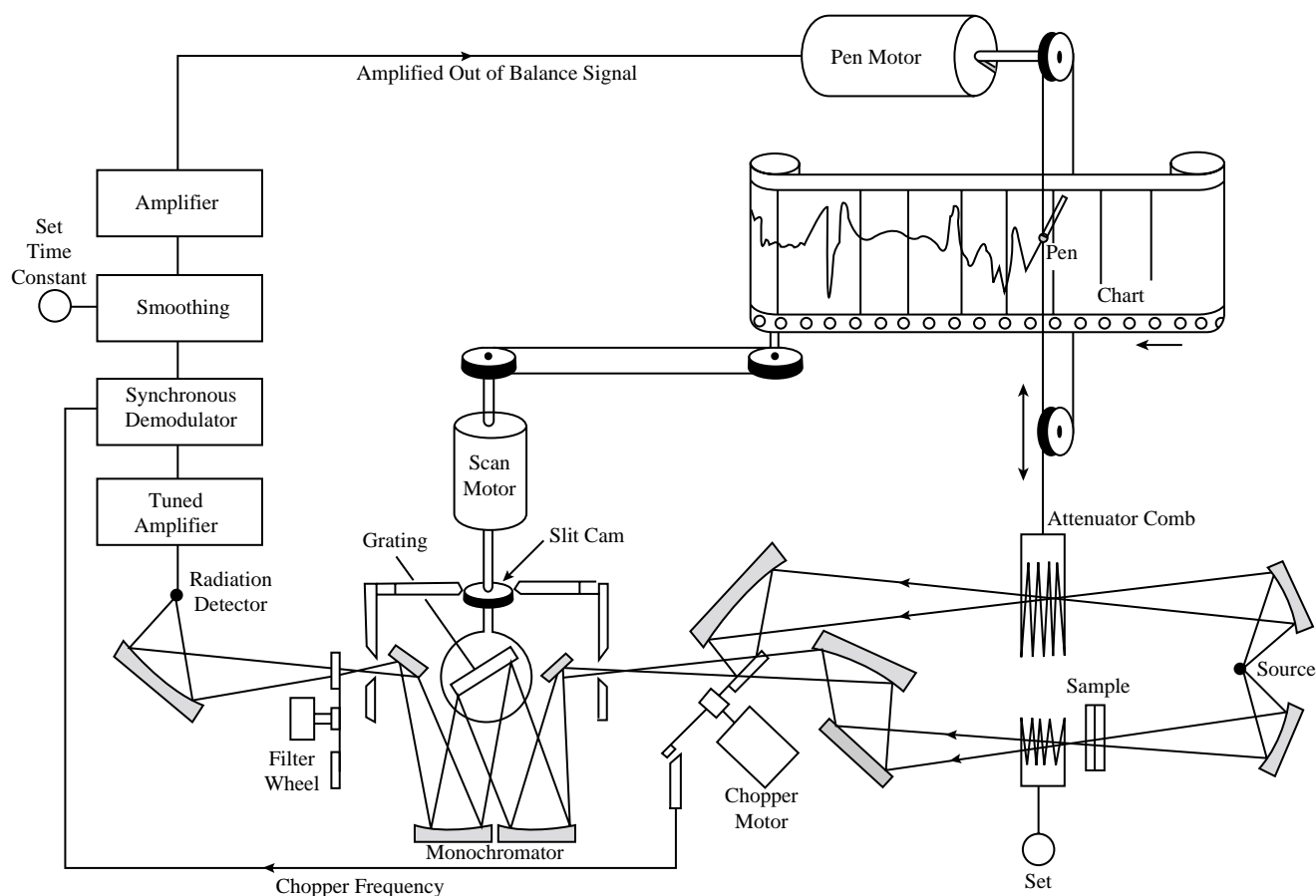


Fig. 3 Schematic of a typical double-beam spectrometer.

Data-Handling System

Most modern infrared spectrometers are equipped with a computer or microprocessor. The computer can record and store spectra, plot either absorbance or transmittance, overlay spectra for comparison, subtract one spectrum from another (to determine the difference or to remove the solvent or impurity spectrum from the sample spectrum), smooth and/or correct base lines, perform multiple scans to increase the signal-to-noise ratio and improve sensitivity, plus many other functions. A computer facilitates the Fourier transform process as well as the solving of complex mathematical matrices for multiple-component analyses.

SAMPLE HANDLING

Infrared spectra may be obtained for gases, liquids, or solids. For transmittance infrared spectroscopy, the sampling techniques may involve a solution, a film, a mull,

or a pellet, depending on the type of sample. Reflectance spectroscopy differs from transmittance spectroscopy in that infrared radiation reflected from the surface of a material is studied. With a proper sampling accessory (obtainable from commercial sources), the materials analyzed by reflectance techniques normally require little or no sample preparation. The method is nondestructive, noninvasive, and very useful for analyzing materials that are too thick or have too much absorbance to be analyzed by transmittance spectroscopy.

Transmittance Spectroscopy

Obtaining the spectrum by transmittance spectroscopy normally requires sample cells with “window” materials that must be:

- Transparent in the wavelength area where the spectrum will be measured
- Chemically stable (nonreactive with the sample or the environment)

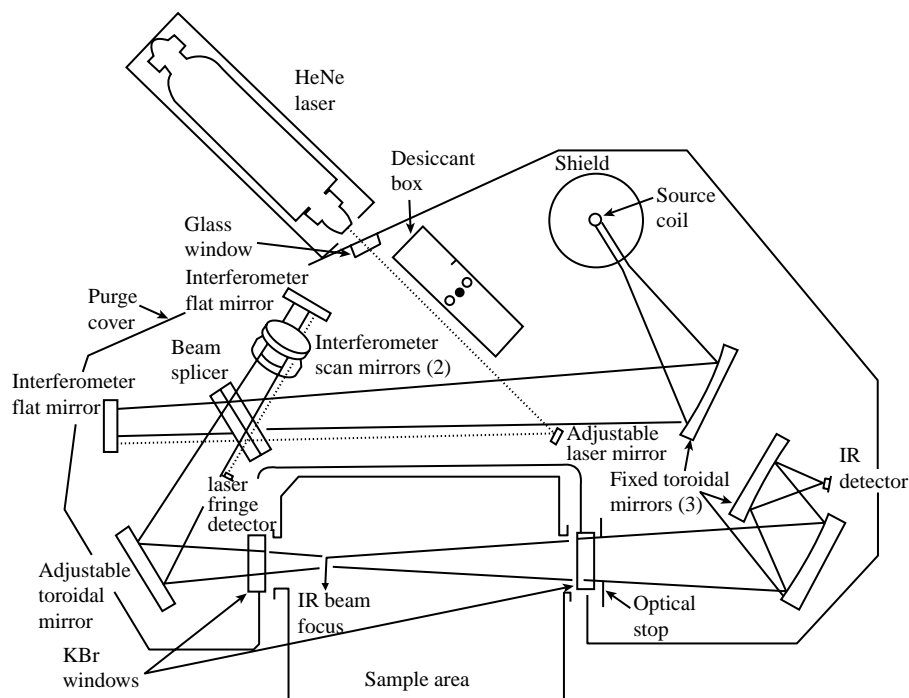


Fig. 4 Optical diagram of a single-beam FTIR spectrometer. (From Perkin-Elmer Corporation.)

- Easy to shape, grind, and polish to optical quality
- Stiff enough to retain their shape.

The most widely used window materials (Table 3) are alkali halides, particularly sodium chloride and potassium bromide.

Gases and Low-Boiling Liquids

For routine infrared analysis of gaseous samples, the type of cell normally used has a fixed path length of 10 cm. For trace level analysis, such as air monitoring, a cell with a longer path length is necessary to increase the sensitivity. A variable path cell (Fig. 6) can provide path lengths (in steps of 1.5 m) for 20-, 40-, and 120 m cells with the same volume of sample. A normal sample compartment in an infrared spectrometer does not hold a cell 120 m long. The long-path gas cell therefore utilizes a folded-path design in which the entering radiation is reflected back and forth several times before leaving the cell. Changing the angular adjustment of an internal pair of mirrors with a dial located outside the cell can vary the number of passes through the cell. Increasing the sample pressure to 10 atm in a gas cell is another way of improving the sensitivity.

Liquids and Solutions

For the analysis of liquid samples, the concentration and path lengths are selected in such a way that the transmittance lies between 15 and 75%. For neat liquids, a very thin layer (0.001–0.05 mm) is sufficient. For 0.05–10% (w/w) solutions, a cell length between 0.1 and 1 mm is normally used. For solution analysis, a compensating cell containing pure solvent is placed in the reference beam of dispersive instruments (a disadvantage for single beam FT instruments). In general, the solvent selected should not interact strongly with the solute. Several solvents are usually required to cover the entire spectrum. For example, carbon tetrachloride, transparent for the region between 4000 and 1333 cm^{-1} , and carbon disulfide, transparent for the region between 1333 and 650 cm^{-1} , may be used together. Some commonly used solvents are deuterated chloroform, methylene chloride, acetonitrile, and acetone. Information regarding the transparent regions for these solvents can be found in (1). A dilute solution in a nonpolar solvent usually gives the best spectrum.

Three types of sample cells are used in liquid analysis: sealed, demountable, and variable thickness cells (fixed and rotating windows). Fig. 7 shows a circular demountable cell and a circular sealed cell, both designed by Perkin-Elmer. A sample cell with rectangular shaped

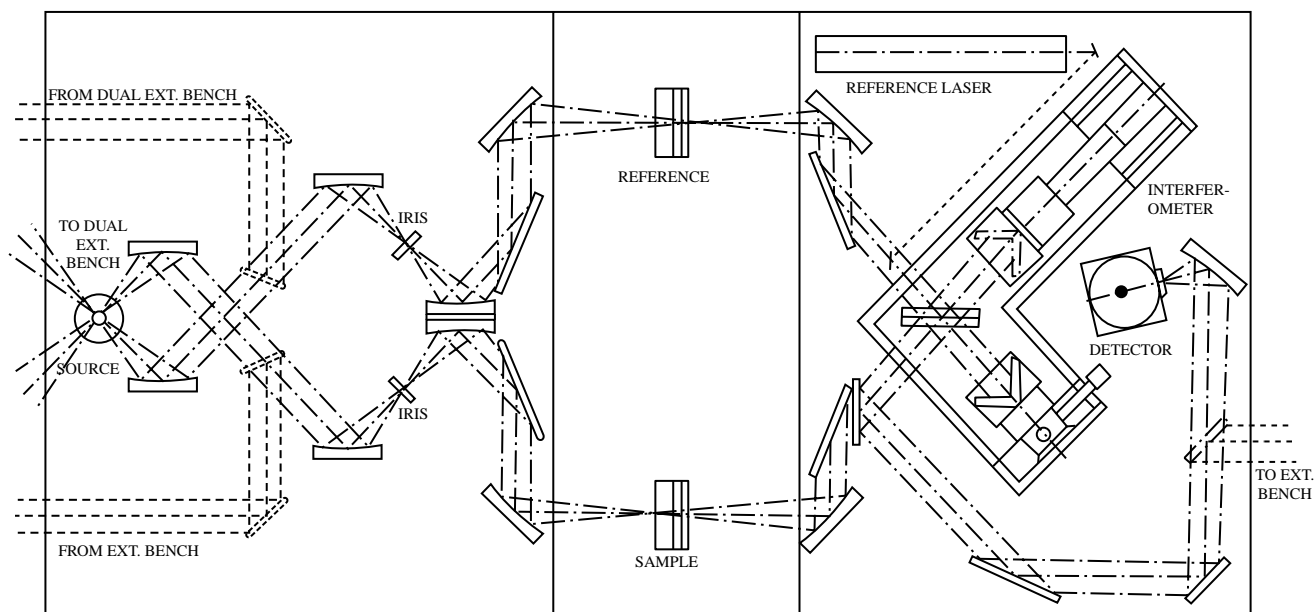


Fig. 5 Schematic of a double-beam FTIR spectrometer.

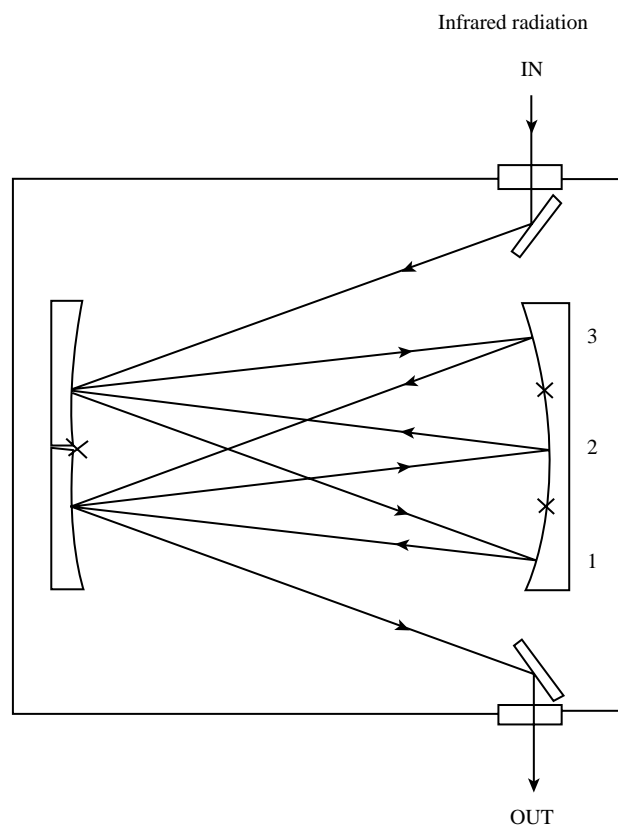


Fig. 6 A variable-path gas cell. (Adapted from Ref. 30.)

windows is also commercially available. The sealed cells are designed for quantitative analysis. The demountable cells are easily cleaned and suitable for routine sample analysis. The path length of a demountable cell can be changed with a suitable spacer. The path length of a variable-path cell in some cases can be varied continuously from 0.025 to 6 mm. This type is useful for determining extinction coefficients in liquid studies and for compensating for solvent absorption in dispersive double-beam spectrometers. Flow-through cells are available for continuous analysis of liquids such as the column effluent when an infrared spectrometer is coupled to a chromatographic system.

Solids

Solids may be examined as a deposited film, a mull, or a pressed disk, or by infrared microscopy.

Films

The deposited-film technique is particularly useful for polymers, resins, and amorphous solids. The samples are dissolved in a reasonably volatile solvent, the solution poured onto a suitable window, and the solvent evaporated by gentle heating or vacuum treatment. One disadvantage of films is that they may cause excessive light-scattering in transmittance spectrometry. This shortcoming can be

Table 3 Properties of infrared-transmitting materials

Material	Transmission range wavenumber (cm ⁻¹)	Refractive index ^{a,b} , at 1000 cm ⁻¹	Solubility g/100 g water at 20°C
Sodium chloride, NaCl	40,000–625	1.49	36.0
Potassium bromide, KBr	40,000–385	1.52	65.2
Potassium chloride, KCl	40,000–500	1.46	34.7
Cesium bromide, CsBr	10,000–270	1.67 (5000 cm ⁻¹)	124.3
Cesium iodide, CsI	33,000–200	1.74	160.0 (61°C)
Fused silica, SiO ₂	50,000–2,500	1.42 (3333 cm ⁻¹)	Insoluble
Calcium fluoride, CaF ₂	50,000–1,100	1.39 (2000 cm ⁻¹)	1.51 × 10 ⁻³
Barium fluoride, BaF ₂	50,000–770	1.42	0.12 (25°C)
Thallium bromide-iodide, KRS-5	16,600–250	2.37	<4.76 × 10 ⁻²
Silver bromide, AgBr	20,000–285	2.20	12 × 10 ⁻⁶
Silver chloride, AgCl	25,000–435	2.00 (5000 cm ⁻¹)	Insoluble
Zinc sulfide, ZnS (Irran-2) ^a	10,000–715	2.20	Insoluble
Zinc selenide, ZnSe (Irran-4) ^a	10,000–515	2.41	Insoluble
Polyethylene (high density)	625–33	1.54 (5000 cm ⁻¹)	Insoluble
Germanium, Ge ^a	20,000–600	4.0 (5000 cm ⁻¹)	Insoluble
Silicone, Si	8,300–1,500 and 360–70	3.4	Insoluble
Sapphire ^a , Al ₂ O ₃	50,000–1,780	1.74	Insoluble
Magnesium oxide, MgO (Irran-5)	25,600–1,060	1.71 (5000 cm ⁻¹)	Insoluble

^aFor attenuated total-reflectance spectrometry.^bUnless otherwise indicated.

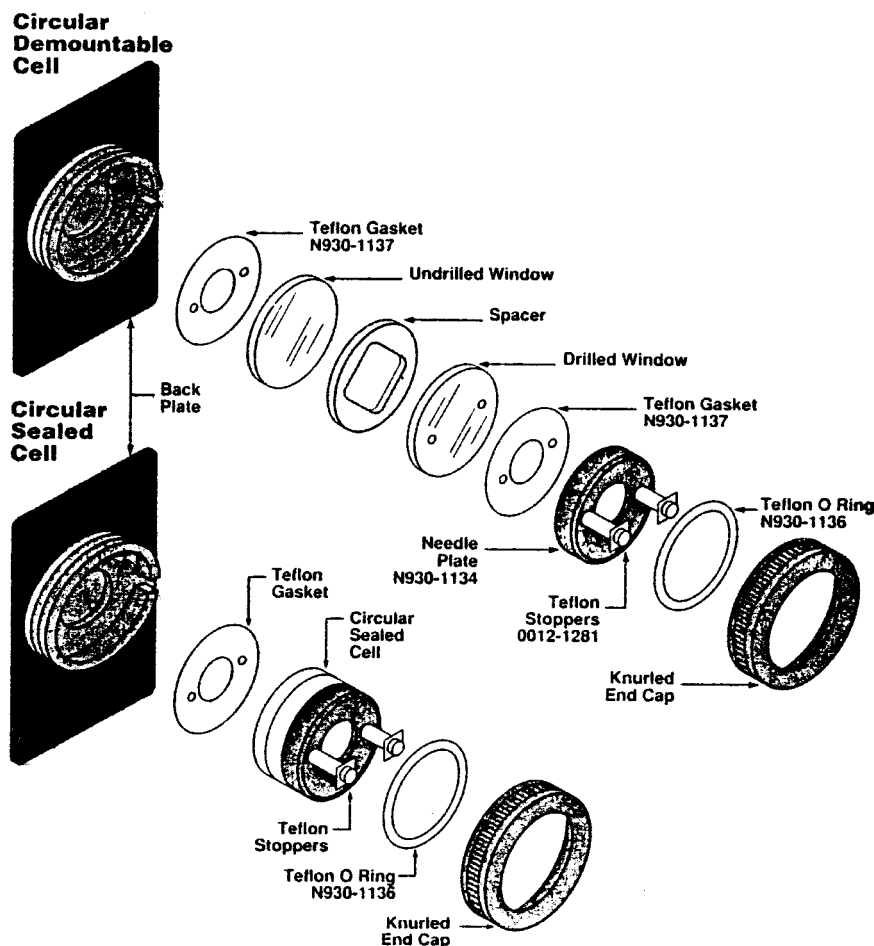


Fig. 7 Diagram of a circular demountable cell and a circular sealed cell for liquid samples. (From Perkin-Elmer Corporation.)

overcome by using attenuated total-reflection or internal-reflection sampling techniques.

Mulls

Grinding 2–5 mg of sample in a smooth agate mortar, adding a drop or two of mulling oil to the sample, and continuing the grinding forms a mull of a solid sample. The sample should be ground until the particle size is below $2\ \mu\text{m}$ to avoid excessive radiation scattering. The mull is placed on the mull plates and the film thickness is adjusted so that the transmittance is between 10 and 90%*T*. Commonly used mulling agents are Nujol (a high-boiling mineral oil), hexachlorobutadiene, perfluorokerosene, and chlorofluorocarbon greases (fluorolubes).

Pressed disks

A pressed disk is prepared by grinding a mixture of 0.5–1.0 mg of sample and 100 mg of dry, powdered potassium bromide in a mortar or Wig-l-bug to reduce the

particle size to less than $2\ \mu\text{m}$. The resulting mixture is pressed into a disk under a pressure of 68.9–103.4 MPa (10,000–15,000 psi). A simple-to-operate minipress is also available. Moisture bands near 3448 and $1639\ \text{cm}^{-1}$ appear frequently in spectra obtained by this technique.

REFLECTANCE SPECTROMETRY

A problem that may be encountered when analyzing a solid sample by transmittance spectroscopy is radiation scattering. Employing reflectance spectroscopy can sometimes reduce this problem. With this technique, the infrared spectra of most solid materials are easily obtained with little or no sample preparation. Spectra of a wide range of solid samples can be characterized with this technique, such as coatings on beverage containers and silicon wafers, polymer films, or other intractable samples.

The reflectance technique, however, is less sensitive than the transmittance technique since about 80% of the infrared radiation is lost after being reflected off the sample surface.

Different types of reflectance spectroscopy depend upon the reflecting behavior of the radiation on the solid. Fig. 8 illustrates various categories used to distinguish techniques for reflecting radiation off solids. Specular reflection spectroscopy is used to measure the reflectance spectrum of a smooth, glossy surface. In reflection-absorption spectroscopy, the radiation passes through a thin surface film on a reflective (typically metallic) surface twice. The thickness of the surface film is normally between 0.2 and 20 μm . This technique is often used in studying the coatings on beverage containers and on silicon wafers of semiconductor devices. Diffuse-reflectance spectroscopy is applied to samples with a coarse, grainy texture such as powders, fibers, and rough surfaces.

Attenuated total-reflectance (ATR) spectroscopy is a widely used sampling technique, in which a sample is placed in contact with a reflecting medium (a plate or prism shaped material called an internal reflectance element). A beam of radiation entering the prism is reflected internally if the angle of incidence at the interface between sample and prism is greater than the critical angle (a function of the refractive index of the sample and the prism).

The sample absorbs the radiation during this process and an absorption spectrum characteristic of the sample is thus obtained. The spectrum of a sample acquired by the ATR technique is generally comparable to the transmission spectrum, but band shifts have been observed. The ATR technique has been used for materials that are too thick or absorb too strongly to be analyzed by transmission spectroscopy. Aqueous solutions can be analyzed by the liquid ATR technique, which was previously very difficult, using transmission spectroscopy.

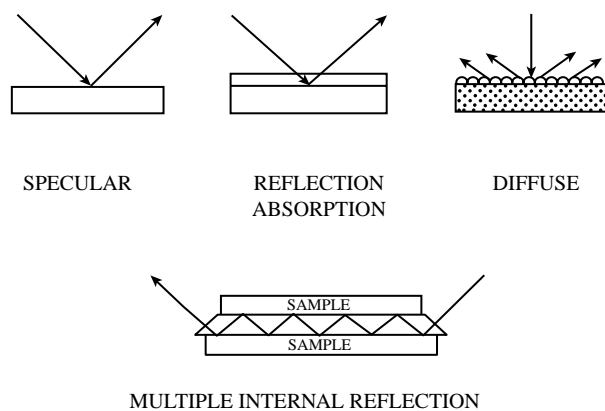


Fig. 8 Main forms of reflective behavior.

Most reflectance spectroscopy is carried out utilizing an accessory that can be easily inserted into and removed from the sampling compartment of a conventional spectrometer. These accessories are designed for each application and usually consist of mirrors or prisms for reflecting or focusing the radiation. Most of the sampling accessories for the FTIR spectrometers and the FTIR microscopes are available commercially.

QUALITATIVE ANALYSIS

Applications of IR spectroscopy to qualitative analysis are mainly for the identification of unknown compounds. For a pure substance, an exact match of the IR spectrum of the compound with that of the reference standard is positive identification. It is important that both spectra are measured under the same conditions. A solid substance might have various crystalline forms. Therefore, the IR spectra of standard and sample in a solid state might not be identical. In such cases, equal portions of the test sample and the standard should be dissolved in equal volumes of a suitable solvent. The solutions are evaporated to dryness in similar containers under identical conditions and the test is repeated on the residues. The compound can also be recrystallized from the same solvent and the determination repeated.

INTERPRETATION OF INFRARED SPECTRA

For an unknown compound without a reference standard, important structural information can be obtained from the IR spectrum. Fig. 9 is a simplified illustration of the correlation between the absorption frequency in cm^{-1} and the functional groups [A more comprehensive description of this type of correlation chart is given in (4).] By observing the presence or absence of certain "group frequencies", related to common functional groups such as $-\text{OH}$, $-\text{NH}_2$, $-\text{CH}_3$, $-\text{C}=\text{O}$, $-\text{CN}$, $-\text{C}-\text{O}-\text{C}$, $-\text{COOH}$, etc., the gross structural features of an unknown compound can be quickly determined.

The IR spectrum in the mid-IR region is divided into the functional-group region, $4000\text{--}1300\text{ cm}^{-1}$ and the fingerprint region, $1300\text{--}400\text{ cm}^{-1}$. To interpret an IR spectrum, the hydrogen-stretching vibrations, which appear between 4000 and 2500 cm^{-1} , are investigated first to determine whether the compound is aromatic or aliphatic. A sharp band at 3300 cm^{-1} suggests the presence of $-\text{C}=\text{C}-\text{H}$, a terminal acetylene moiety, and/or an $-\text{NH}$ group. Aromatic and unsaturated

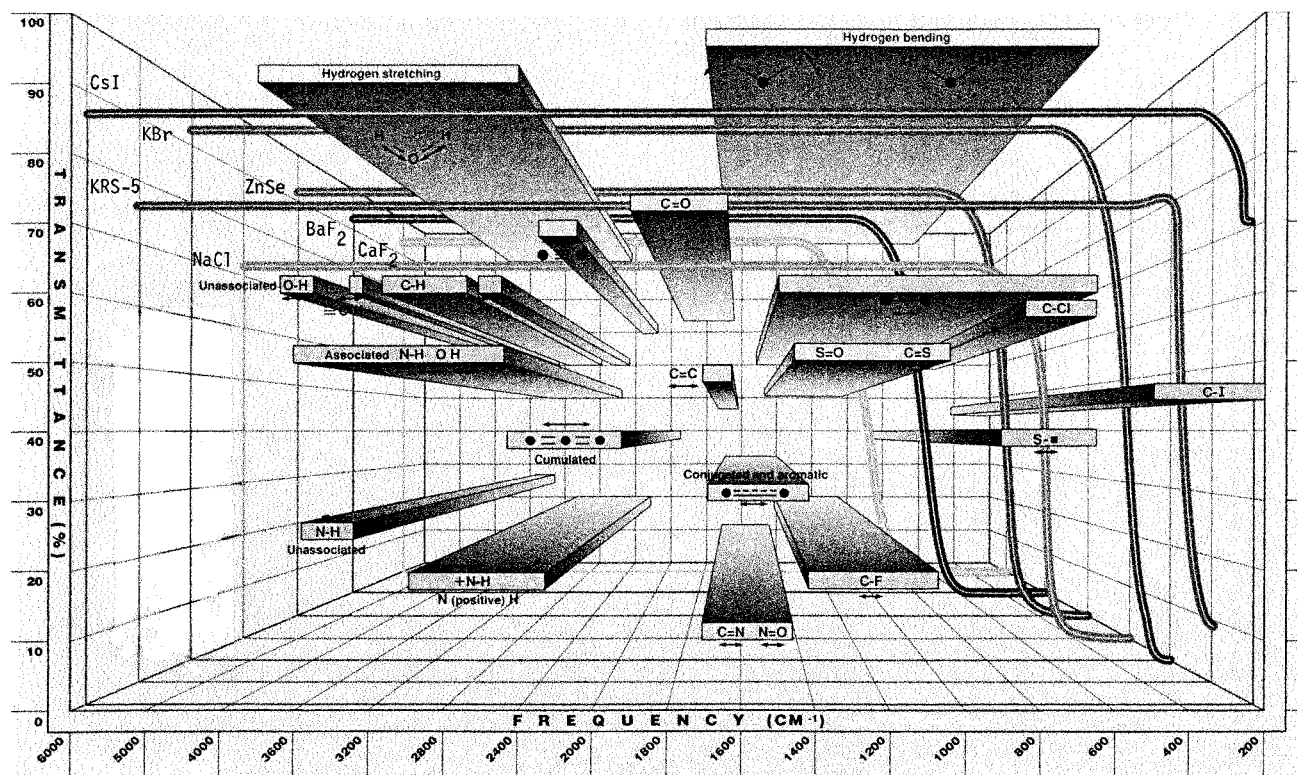


Fig. 9 Structure and infrared absorption bands correlation chart. Transmission ranges for infrared window materials CsI, KBr, ZnSe, KRS-5, BaF₂, CaF₂, and NaCl are also indicated on the chart. (From Perkin-Elmer Corporation.)

compounds appear at 3100–3000 cm^{-1} and aliphatic compounds appear at 3000–2800 cm^{-1} . Unsaturation may also be suggested by C=C and C≡C stretches between 2500–1540 cm^{-1} . Next, the group frequency is examined to establish the presence or absence of certain functional groups such as —OH, —NH, and C=O. The stretching vibration of carbonyl groups occurs in the 1850–1540 cm^{-1} range. The fingerprint region, 1300–909 cm^{-1} , is characteristic of each molecule and, when examined in reference to the other regions, provides positive identification of certain functional groups. The lack of strong absorption bands in the 909–650 cm^{-1} region usually indicates a nonaromatic structure. The group frequency might vary when the molecular environment changes; thus, the shift in frequency provides further structural information for an unknown substance. Some typical examples are the effects of inter- or intramolecular hydrogen bonding, the dimerization of a carboxylic acid, and ring strain within the molecule.

After the IR spectrum has been interpreted, the structural characteristics of the unknown can be narrowed

down to a few compounds in a specific category. For positive identification of an unknown, information on the physical state, melting or boiling point, solubility, and the history of the compound should be obtained. The flame test should be done, and the NMR and mass spectra should be determined. Once the structure of the unknown has been deduced, the IR spectrum of the unknown can be compared with available IR reference spectra (5, 6). With FTIR spectrometers, a large collection of IR spectra can be stored in the computer's library. When the IR spectrum of the unknown is obtained, the instrument can do a library search and find the spectrum that offers the best match to the unknown.

If the unknown substance is not pure or is mixed with other components, separation of the compound from the matrix may be required. However, in many cases mixtures may be characterized. The compound can be purified by recrystallization, distillation, solvent-partitioning, pH manipulation, or chromatographic separation. Once the appropriate separation procedures have been completed, the purified unknown can undergo the structural characterization steps described above.

QUANTITATIVE ANALYSIS

The amount of infrared radiation absorbed by the sample is proportional to the concentration of the sample and the cell path length. The Beer–Lambert law, shown in Eq. 7, illustrates the mathematical relationship:

$$A = abc = \epsilon b(\text{cmol. wt.})A = \log_{10}(I_0/I) = \log(1/T) \quad (7)$$

The absorbance A is the logarithm (to the base 10) of the reciprocal of the transmittance (T), which in turn is the ratio of the intensity of the transmitted radiation I to the intensity of the incident radiation I_0 . The absorptivity, a , is the quotient of the absorbance (A) divided by the product of concentration c in grams per liter of the substance and the absorption path length b in cm. The more commonly used molar absorptivity ϵ is used when the concentration is in mol/L rather than g/L.

At a constant cell path length, Beer's law shows that the absorbance of radiation through a medium is proportional to the concentration of the solute. Beer's law is strictly valid only for monochromatic radiation. Stray light (i.e., scattered radiation), which reaches the detector without having passed through the desired beam path, molecular interactions such as hydrogen bonding, which varies with the sample concentration, and other instrumental factors such as slit width, all affect molar absorptivity and result in some deviations from Beer's law. For an accurate analysis of the concentration of an unknown sample, it is usually necessary to first create a calibration curve from standard solutions and then determine the absorbance of the sample under identical conditions. The concentration of the unknown is determined from the calibration curve. Diluted solutions (less than 2%) are recommended for this analysis.

If a calibration curve has been established previously and the instrumental conditions have not been changed, the unknown concentration can sometimes be determined by using a single calibration preparation. The concentration of the unknown is calculated by ratioing absorbances and concentrations, according to Eq. 8.

$$\begin{aligned} A_{\text{std}} &= c_{\text{std}} \\ A_{\text{unk}} &= c_{\text{unk}} \end{aligned} \quad (8)$$

In such cases, the calibration curve of the sample should be determined periodically to make sure that the linearity of Beer's law is still valid.

For quantitative analysis of a single component, a strong absorption band that is relatively free of overlapping bands or interference is selected from the IR

spectrum. The intensity of the band is measured either in units of percent transmittance or absorbance.

The intensity of the incident radiation (I_0) and the intensity of the transmitted radiation (I) can be measured by the base-line method illustrated in Fig. 10. In the FTIR spectrometer, the computer is capable of converting the transmittance into absorbance, subtracting the background or the solvent interference from the sample spectrum, generating a calibration curve, and calculating the unknown concentration.

For quantitative analysis of multiple components in a mixture where overlapping of the bands generally occurs, the number of absorption bands chosen should be equal to the number of components to be determined (n components). A series of standard mixtures (at least as many as n) containing a known amount of the substance is prepared. The resulting equations, involving the science of chemometrics, are solved by matrix algebra methods with the help of a computer. The accuracy of this type of quantitative analysis technique decreases with an increasing number of components in the mixture.

APPLICATIONS AND RECENT DEVELOPMENTS

With the continuous improvement of sampling techniques, software, and instrumental designs, more and more sophisticated FTIR spectrometers are now available. *Analytical Chemistry* biannually reviews the most recent developments in many fields of analytical techniques, including infrared spectroscopy, many times on a yearly basis.

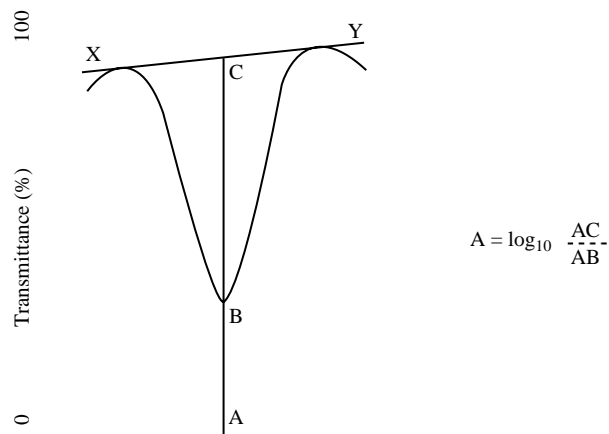


Fig. 10 Illustration of a base-line measurement method. (Adapted from Ref. 29.)

Hyphenated Techniques

Hyphenated techniques refer to the combination of one or more analytical techniques for problem solving and fundamental research. Numerous types have been reported in the scientific journals. Frequently the FTIR spectrometer is coupled with chromatographic instruments for the structural characterization of a column eluent. These systems are designed to monitor the eluent or to obtain its spectrum. Much of the development has been focused on the sampling technique and the design of the interface between the chromatographic system and the FTIR spectrometer to improve the performance of the system.

For GC–FTIR, headspace GC–FTIR, or GC–FTIR–MS, a light-pipe gas cell, direct deposition, or a matrix isolation apparatus is usually incorporated. The mobile phase is helium or nitrogen, which does not interfere. In general, the matrix-isolation technique gives a better quality of infrared spectrum and a lower detection limit. Organic material present as low as a picogram can be detected by this technique (7, 8).

For HPLC–FTIR, GPC–FTIR, or SFC–FTIR, the design of the interface is more challenging since the mobile phases used for these chromatographic systems normally have strong infrared absorbencies; thus, it is important to remove the mobile phase prior to measuring the spectrum. For the interface between the two systems flow-cells or mobile-phase elimination techniques may be used. Some recent developments point toward the elimination of mobile-phase techniques (9–11). A microbore column can help to reduce the mobile-phase volume in the system (12, 13).

FTIR Microscopy

Fourier transform infrared microscopy is the primary infrared technique for structural identification of materials at microquantities. The method is nondestructive and noninvasive. When using a proper transmittance sampling technique and a proper detector, the limit of detection can be as low as the picogram level. In the pharmaceutical industry, FTIR microscopy is used to analyze bulk drugs, excipients, and particulate contaminants (14). Recent studies have shown that by coupling FTIR microscopy with GC, HPLC, SFC, or GPC systems, the detection limit of the method is substantially improved (15, 16).

Infrared Analyzers

The combination of the fast-response FTIR spectrometer and proper sampling techniques has made the infrared

analyzer the instrument of choice for the on-line monitoring of a process stream. The IR analyzer is generally compact, rugged, and easy to operate in the field or industrial production. The infrared analyzer equipped with a microprocessor or a computer is capable of identification and quantitative analysis of single or multiple components in the industrial process stream in a short time. FTIR analyzers are widely used for in-process on-line monitoring of manufacturing, waste streams, and environmental air monitoring (17–21).

SUMMARY

Infrared spectroscopy continues to be a rapidly growing analytical technique. Improvements and innovations have been made in sampling techniques and analysis software. A large variety of samples in various forms can be analyzed by IR spectroscopy. This flexibility has made infrared spectroscopy one of the most important tools in today's analytical laboratories, especially for analysis of trace contaminants.

REFERENCES

1. Duerst, R.W.; Duerst, M.D.; Stebbings, W.L. *Transmission Infrared Spectroscopy. Modern Methods of Applied Molecular Spectroscopy*; John Wiley & Sons: New York, 1998.
2. Duerst, R.W.; Stebbings, W.L.; Lillquist, Gerald, J.; Westberg, James, W.; Breneman, William, E.; Spicer, Colleen, K.; Dittmar, Rebecca, M.; Duerst, Marilyn, D. *Depth Profiling and Defect Analysis. Practical Guide to Infrared Microspectroscopy*; Marcel Dekker, Inc.: New York, 1995.
3. Willard, H.H.; Merritt, L.L., Jr.; Dean, J.A.; Settle, F.A., Jr. *Instrumental Methods of Analysis*, 7th Ed.; Wadsworth Publishing Co.: Belmont, CA, 1988.
4. Silverstein, R.M.; Bassler, G.C.; Morrill, T.C. *Spectroscopic Identification of Organic Compounds*, 5th Ed.; John Wiley & Sons: New York, 1991.
5. Pouchert, C.J. Ed. *The Aldrich Library of Infrared Spectra*, 2nd Ed.; Aldrich Chemical Co.: Milwaukee, 1975.
6. Sadtler Research Laboratories. *Catalog of Infrared Spectrograms*; Philadelphia (A Continuously Updated Subscription Service).
7. Holloway, T.T.; Fairless, B.J.; Freidfine, C.E.; Kimball, H.E.; Kleopfer, R.D.; Wurrey, C.J.; Jonoby, L.A.; Palmer, H.G. *Appl. Spectrosc.* **1988**, *42*, 359–369.
8. Bourne, S.; Haefner, A.M.; Norton, K.L.; Griffiths, P.R. *Anal. Chem.* **1990**, *62*, 2448–2452.
9. Lange, A.J.; Griffiths, P.R.; Fraser, D.J. *Anal. Chem.* **1991**, *63*, 782–787.

10. Griffiths, P.R.; Haefner, A.M.; Norton, K.L.; Fraser, D.J.J.; Pyo, D.; Makishima, H.J. *High Resol. Chrom.* **1989**, *12*, 119–122.
11. Norton, K.L.; Lange, A.J.; Griffiths, P.R.J. *High Resol. Chrom.* **1991**, *14*, 225–229.
12. Jinno, K.; Fujimoto, C.J. *Chrom.* **1990**, *506*, 443–460.
13. Fujimoto, C.; Jinno, K. *Trends Anal. Chem.* **1989**, *8* (3), 90–96.
14. Clark, D.A.; Nichols, G. *Anal. Proc.* **1990**, *27*, 19–21.
15. Fraser, D.J.J.; Norton, K.L.; Griffiths, P.R. *Pract. Spectrosc.* **1988**, *6*, 197–210.
16. Bergin, F.J. A. *Appl. Spectrosc.* **1989**, *43* (3), 511–515.
17. Coates, J.P.; Rein, A.J.; Morris, K.S. *Am. Lab.* **1988**, *20* (2), 117–124.
18. Morris, K.S.; Rein, A.J. *Am. Lab.* **1988**, *20* (11), 46–52.
19. Fuller, M.P.; Garry, C.; Stanek, Z. *Am. Lab.* **1990**, *22* (15), 58–69.
20. Wilks, P.A. *Am. Lab.* **1991**, *26C*–26E.
21. McIntosh, B.C.; Vidrine, D.W.; Doyle, W.M. *Am. Lab.* **1991**, *23* (18), 19–22.
22. Beckett, A.H.; Stenlake, J.B. *Practical Pharmaceutical Chemistry*, 3rd Ed.; Athlone Press, University of London: London, England, 1976; *2*, 331–360.
23. Crooks, J.E. *The Spectrum in Chemistry*; Academic Press: New York, 1978.
24. Duerst, R.W.; Duerst, Marilyn D. *Transmission Infrared Spectroscopy. Modern Techniques in Applied Molecular Spectroscopy*; Mirabella, F., Ed.; John Wiley & Sons, Inc.: New York, 1998.
25. Duerst, R.W.; et al. *Depth Profiling and Defect Analysis of Films and Laminates: An Industrial Approach. A Practical Guide to Infrared Microscopy*; Humecki, H., Ed.; Marcel Dekker, Inc.: New York, 1995.
26. Griffiths, P.R.; de Haseth, J.A. *Fourier Transform Infrared Spectrometry. Chemical Analysis*; John Wiley & Sons, Inc.: New York, 1986; 83.
27. Hannah, R.W.; Swinehart, J.S. *Experiments in Techniques of Infrared Spectroscopy*; Perkin-Elmer: Norwalk, CT, 1974.
28. Krishnan, K.; Hill, S.L. *FT-IR Microsampling Techniques*; Bio-Rad Digilab Division: Cambridge, MA, 1990.
29. Miller, R.G.J.; Stace, B.C. *Laboratory Methods in Infrared Spectroscopy*, 2nd Ed.; Heyden Press: (Wiley), New York, 1979.
30. Stewart, J.E. *Infrared Spectroscopy: Experimental Methods and Techniques*; Marcel Dekker, Inc.: New York, 1970.

SPECTROSCOPIC METHODS OF ANALYSIS—MASS SPECTROMETRY

Mike S. Lee

Milestone Development Services, Newtown, Pennsylvania, U.S.A.

INTRODUCTION

The dramatically increased expenditures for both in-house and outsourced pharmaceutical research and development (R&D) have led to a greater dependence on technology. New technologies are constantly introduced into drug development to address throughput issues and improve development cycles. The incorporation of new technologies has resulted in fundamental change in the drug development paradigm. Recently, sample generating-based technologies such as high throughput biomolecular screening and automated parallel synthesis have shifted the bottleneck to sample analysis-based technologies.

The current focus on analytical techniques in the pharmaceutical industry emphasizes four primary figures of merit: sensitivity; selectivity; speed; and high throughput. Mass spectrometry (MS) provides each of these key attributes, and therefore, has been benchmarked an effective solution for pharmaceutical analysis in each stage of drug development (1). Perhaps more enabling than the MS-based technology itself is the diverse applications of MS in conjunction with sample preparation, chromatographic separation, and informatics. It is within this context that MS has played an increasingly vital role in the pharmaceutical industry and has become the preferred analytical method for trace-mixture analysis (Fig. 1).

A variety of MS formats are widely accepted and applied in the pharmaceutical industry. The specific MS application is often defined by the sample introduction technique. The pharmaceutical applications highlighted in this article feature two types of sample introduction techniques: dynamic and static. Dynamic sample introduction involves the use of high-performance liquid chromatography (HPLC) on-line with MS. The resulting liquid chromatography/mass spectrometry (LC/MS) format provides unique and enabling capabilities for pharmaceutical analysis. The electrospray ionization (ESI) (2) and atmospheric pressure chemical ionization (APCI) (3) modes are the most widely used. Static sample introduction techniques primarily use matrix-assisted laser desorption/ionization (MALDI) (4).

The advances in MS instrumentation (5) and role of MS within the pharmaceutical industry (1) have been recently

reviewed. This article will focus on MS technologies with regard to specific applications in drug development. The intent of this article is to provide an overview of MS applications and describe the significant integration of this technology into drug development. A detailed and in-depth overview of current MS technologies and applications can be obtained from the recent proceedings of the American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics (www.asms.org) and the Association of Biomolecular Resource Facilities (www.abrf.org).

GENOMICS

Though the contributions of MS have been somewhat limited in the field of genomics, there has been increased participation and interest (6). Certainly, the worldwide recognition received from the Human Genome Project created a sense of urgency toward determining genetic variation.

Genomics refers to the study of genetic data to draw correlation between individual genetic inheritance and medically or biologically important parameters. For example, these parameters may involve a patient's response to a specific drug. Knowledge of the genetic basis of individual drug response may provide understanding of the observed variability in drug response arising as a result of genetically determined differences in drug absorption, disposition, metabolism, or excretion (7). Furthermore, knowledge of genetic variation may be useful during the target selection process when multiple targets are available within a specific disease state. Thus, the pharmaceutical industry has great interest in determining the genetic variation in patient populations.

Due to the existence of sequence variations, or polymorphisms, no two human genomes are identical. Single nucleotide polymorphisms (SNPs) are the most abundant genetic variation with an estimated frequency of 1 SNP per 500 basepairs. Since SNPs are so prevalent in the genome, they can act as markers that are linked with a phenotype to provide a comprehensive measure of interaction with a specific drug. The validation of

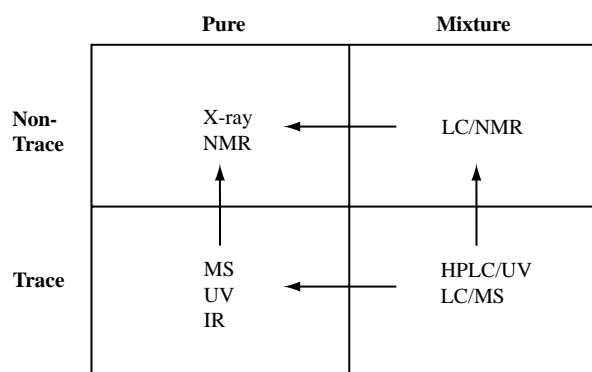


Fig. 1 Structure analysis matrix that illustrates pharmaceutical analysis preferences for four specific sample types: nontrace/pure; nontrace/mixture; trace/pure; and trace/mixture. (Courtesy of Milestone Development Services, Newtown, PA.)

a particular SNP represents an important stage for establishing SNPs as a routine clinical diagnostic marker. The validation of SNPs via MALDI-TOF MS is emerging as a valuable genotyping tool (8, 9). A schematic of a MALDI-TOF MS instrument is shown in Fig. 2.

A recent study performed by Stroh et al. (10) compared the performance of MALDI-TOF MS with restriction fragment length polymorphism (RFLP) and fluorescence polarization (FP). The study involved the analysis of known mutations of the IL-1 β gene. The procedure involved amplification of patient DNA samples using standard PCR techniques followed by a primer extension step where a separate post-PCR primer is hybridized

directly adjacent to the SNP site. The resulting MALDI-TOF MS data provided a direct confirmation of molecular weight for fast analysis of polymorphisms.

Knowledge of the DNA sequence flanking the SNP site allows for the optimum choice of post-PCR primer size, primer location, and dideoxy-nucleotide(s). Thus, the assay can be designed to extract complete information about the SNP regardless of its state. This emerging MS-based approach for SNP genotyping has potential to be highly automated without the requirement for fluorescent tags. Furthermore, "multiplexing" can be attained by selecting post-PCR primers of varying lengths to dedicate predetermined regions of the mass spectrum to specific SNPs.

PROTEOMICS

The study of protein structure, function, quantity, and interactions during maturation and progression of disease is referred to as proteomics. Analytical approaches that use a combination of two-dimensional (2-D) gel electrophoresis for protein separation and MS analysis for protein identification followed by database searches is a widely practiced proteomics strategy (11). The tryptic peptides extracted from gels are analyzed by MALDI-TOF MS and microcolumn or capillary LC tandem mass spectrometry (MS/MS) techniques. Typically, the MALDI-TOF MS techniques are used to quickly identify peptide fragments and confirm the presence of known proteins. Nano-scale

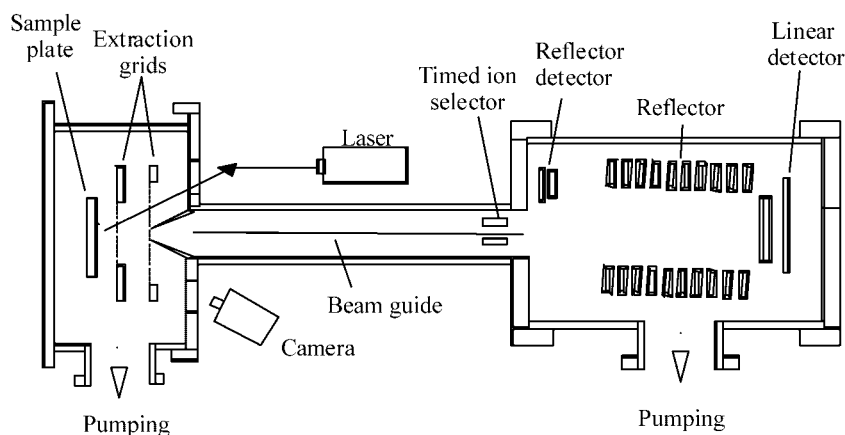


Fig. 2 Schematic of a MALDI-TOF MS instrument. MALDI-TOF samples are prepared with a matrix that contains a small organic molecule capable of absorbing ultraviolet light. A laser is used to desorb ions from the sample plate and the resulting ions are forced into the flight tube by application of the acceleration voltage from extraction grids. All ions leave the source with the same kinetic energy and travel down the flight tube toward an ion reflector. Separation is based on mass with lighter ions traveling faster than heavier ions. The ion reflector is used to correct for small kinetic energy differences between ions of the same mass resulting in improved resolution and mass accuracy. (Courtesy of Applied Biosystems, Framingham, MA.)

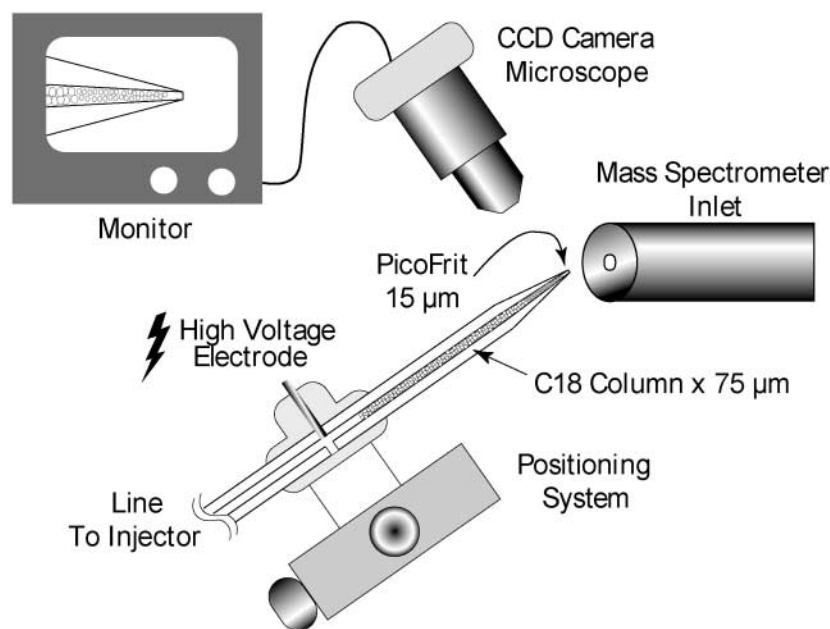


Fig. 3 Schematic of a nano-scale capillary ESI interface. This specialized LC/MS interface, operating at flow rates from 20 to 500 nL/min and using 50 to 100 μm ID columns, typically provides low femtomole sensitivity. Fully automated sample handling and preparation procedures (i.e., desalting and preconcentration) combined with specialized devices for high separation and variable nL gradient flow rates provide unique capabilities for high-throughput analysis of proteins. (Courtesy of New Objective, Cambridge, MA.)

capillary LC/MS/MS techniques (using 50–100 μm diameter columns, operating at flow rates of 20–500 nL/min) are used to further interrogate the complex protein mixture at the low femtomole level. These techniques require the use of a specialized ESI source shown in Fig. 3.

The combination of MALDI-TOF MS and capillary LC/MS/MS was recently described for the identification of disease state markers in human urine (12). In this study, urine proteins obtained from emphysema patients were separated on 2-D gels and selected spots were digested with trypsin and analyzed by MALDI-TOF. A database search using Protein Prospector identified a potential biomarker for emphysema as human α -1-antitrypsin (A1AT). The corresponding MALDI spectrum contained nine out of 18 peptides with masses that match the expected tryptic digest fragments for A1AT.

The same tryptic digest protein sample was analyzed by capillary LC/MS/MS using an ion trap mass spectrometer followed by a database search with SEQUEST. Figure 4 illustrates the components of an ion trap mass spectrometer. The highly automated data-dependent MS/MS analysis provided excellent sequence coverage for 11 tryptic peptides related to A1AT in a single LC/MS run. A tryptic peptide that corresponds to the A1AT sequence SVLGQLGITK observed at retention time (r_t) 30.5 min. was observed in the spectrum. The LC/MS data also

provided sequence information on unmatched MALDI peaks.

The need to detect lower concentration of protein and peptide mixtures has resulted in the increased use of hybrid quadrupole/orthogonal TOF (QTOF) MS/MS instruments (13) in conjunction with microcolumn LC. Fig. 5 shows a schematic of a QTOF instrument. This LC/MS approach provides a resolution of ca. 0.1 mass units allowing for the analysis of complex product ion spectra (14, 15). A recent publication by Chalmers and Gaskell highlights the current challenges in proteome analysis with regard to MS instrumentation (16).

NATURAL PRODUCTS DEREPLICATION

Historically, an excellent source of novel lead drug compounds is natural products. Natural product screening activities typically occur during drug discovery and involve the testing of crude extracts obtained from microbial fermentation broths, plants, or marine organisms. When activity above a certain level is detected, active components are isolated and purified for identification. This process is often time-consuming, where the physicochemical characteristics of the active components

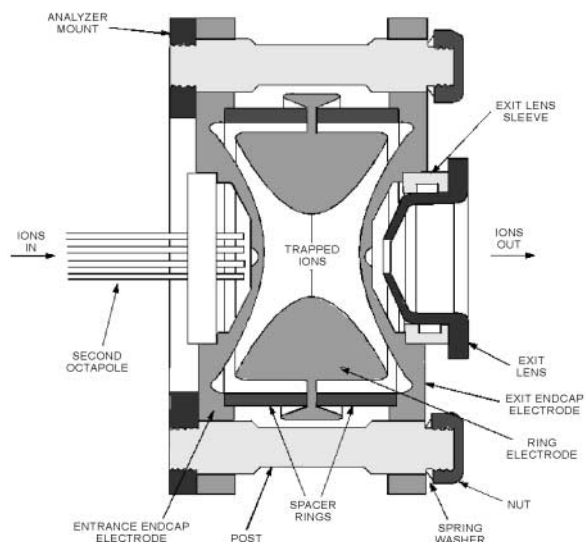


Fig. 4 Schematic of an ion trap MS instrument. This device consists of two endcap electrodes (entrance and exit) and a ring electrode. An ion trap MS separates ions based on mass-to-charge ratio (m/z). Once ions are introduced into the ion trap MS, the radiofrequency (rf) amplitude is increased so that ions are sequentially ejected (by increasing mass) and detected. This type of MS provides a routine (i.e., benchtop) and sensitive detector using either GC and LC interfaces. Furthermore, this instrument provides a unique format for multiple stages of MS analysis (MS^n). (Courtesy of Thermo-Finnigan, San Jose, CA.)

are determined, known compounds are identified (dereplication), and the novel compounds are scaled-up for more detailed investigation.

Analysis strategies that use on-line ESI-LC/MS approaches provide an integrated format for natural product dereplication by combining traditional fraction collection, sample preparation, and multi-component analysis into a single step. In this way, crude extracts are screened without extensive purification and chemical analysis. Furthermore, less material is required due to the sensitivity of the technique and chromatographic resolution is retained.

The key to natural products analysis using this approach is dependable molecular weight determination. This information is used with existing natural product databases that contain information on the bioactive compounds, the physical descriptions of the microorganisms from which they come, their spectrum of activity, the method of extraction and isolation, and physical data (i.e., molecular weight, UV absorption maxima). Molecular weight is the most critical information for initial searches because of its link to structural specificity. This information is used to make pivotal decisions on whether or not to proceed to more time-consuming isolation steps based on novelty of the compound. In recently reported studies (17, 18), LC/MS is used to increase sensitivity and accelerate analysis. These features serve to significantly reduce labor. A recent review described LC/MS-based approaches for

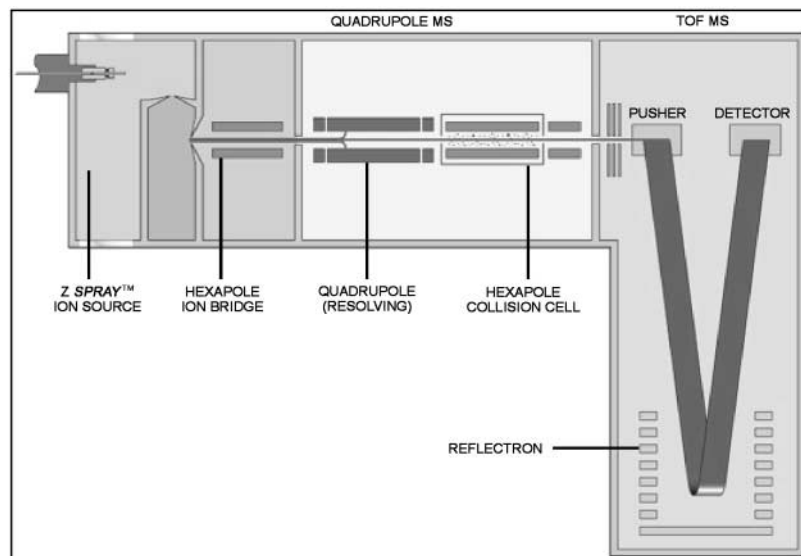


Fig. 5 Schematic of a QTOF MS instrument. Ions formed in the source region are introduced into a quadrupole mass filter (see Fig. 7) that separates ions based on mass-to-charge ratio (m/z). Selected ions are then transferred into the TOF MS for detailed analysis (i.e., high resolution capabilities). (Courtesy of Micromass, Manchester, UK.)

the characterization of natural product mixtures in conjunction with high-throughput screening (19).

The instrumental configuration of the LC/MS system developed by Ackermann et al. (17) features an HPLC, UV detector, fraction collector, ESI-tandem quadrupole (MS/MS), and MALDI-TOF MS. Filtered fermentation broths were extracted with butanol or ethyl acetate, and eluted on a gradient C18 reversed-phase HPLC separation. The eluent was split 1:10 between a tandem quadrupole MS/MS instrument (scanning 250–2000 amu/3 sec in the full scan mode) and a single wavelength UV detector (254 or 230 nm). One-min fractions were collected after the UV detector. Of these fractions, 20–50 μ L is used for MALDI-TOF analysis, and the remainder is concentrated for microbiological testing. The LC/UV chromatogram was compared to the bioactivity assay histogram to highlight the peaks that contain activity. The molecular weights of the active peaks were obtained for novelty assessment of the compounds.

Similar approaches that use on-line LC/MS and LC/MS/MS techniques have been recently described for natural products dereplication (20, 21). Approximately 1,000 natural product extracts can be screened annually for *in vivo* and *in vitro* activity using LC/MS-based systems. A standard approach for dereplication involves a comparison of retention time, full scan mass spectra (i.e., molecular weight information), and MS/MS spectra with those from known biologically active standards. Thus, previously identified components are rapidly eliminated and do not require time-consuming structure elucidation studies. The savings of effort allow researchers to focus efforts on novel chemistries. Samples of novel compounds can then be infused into an ion trap mass spectrometer, and a multiple stage mass analysis (MS^n) fragmentation map is generated.

COMBINATORIAL CHEMISTRY

Recent reviews (22, 25) describe MS-based methods ranging from the analysis of complex molecular libraries (26) to open-access formats for drug discovery and development (27). High throughput criteria were central to each application.

An important development in the quest for high throughput combinatorial library analysis was the multiple ESI interface described by Wang et al. (28). This novel ESI interface enabled effluent flow streams from an array of four HPLC columns to be sampled independently and sequentially using a quadrupole MS instrument. The interface featured a stepping motor and rotating plate

assembly. The effluent flow from the HPLC columns was connected to a parallel arrangement of electrospray needles co-axial to the mass spectrometer entrance aperture. The individual spray tips were positioned 90° relative to one another in a circular array. Each spray position was sampled multiple times per second by precise control of the stepping motor assembly.

The parallel sample analysis format using a multiplexed LC/MS interface with an orthogonal time-of-flight (TOF) MS was described by de Biasi et al. (29). This approach illustrated the high-throughput capabilities of a multiplexed ESI interface in combination with an MS format that accommodates fast chromatography methodologies. The system featured a four-way multiplexed electrospray interface attached directly with the existing source of the TOF-MS instrument. A rotating aperture driven by a variable speed stepper motor permitted the sampling of the spray from each electrospray probe tip. The data files were synchronized with the corresponding spray.

As the preparation of large libraries for lead discovery became routine, the burden placed on analysis techniques focused mainly on throughput and quality (30). However, biological assay requirements typically required pure compounds. Thus, the focus shifted toward the use of automated high throughput purification methods applied to libraries of discrete compounds (31). Reverse-phase analytical and preparative HPLC formats in conjunction with MS techniques have been critical for the high throughput purification approaches for parallel synthesis libraries. A variety of approaches that featured the use of gradient methods, short columns, and high flow rates have been described (32). Highly automated LC/MS approaches for purification at the multimilligram level were described for a quadrupole system by Zeng et al. (33). These methods featured the use of short columns that were operated at ultra high flow rates. Preparative columns were operated at flow rates in excess of 70 ml/min to match the linear velocity of the short analytical columns (4.0 ml/min). Analytical LC/MS analyses of compound libraries were achieved in 5 min for chromatographically, well-behaved compounds. Slightly longer preparative LC/MS analysis times (8–10 min/sample) were required for compounds that exhibited poor chromatographic peak shapes and/or for compound mixtures the required higher resolution separations. The fraction collection process is initiated in real-time once the reconstructed ion current is observed for a specific m/z value that corresponds to the compound of interest. This design permitted the collection of one sample per fraction. Thus, the need for very large fraction collector beds and postpurification screening and pooling was eliminated. Unattended and automated

operation of this system led to the purification of overcompounds (mg quantities) per day.

BIOAFFINITY SCREENING

With the integration of highly automated parallel synthesis techniques into drug discovery programs, hundreds of thousands of compounds are now screened against a particular biological target. Once activity is determined for a mixture, the identification of the active component(s) is necessary. A recent study described the use of bioaffinity selection LC/MS methods for the identification of active mixture component(s) (34). This approach features an integrated bioaffinity-based LC/MS screening method to separate and identify compounds from mixtures.

A mixture of compounds is incubated with the target protein and the components bound to the protein are selected by using a size exclusion chromatography (SEC) "spin column." In this experiment, the unbound compounds are retained on the column. The bound components are eluted and identified with LC/MS. Increased specificity is obtained by dissociating the bound compounds and performing a second equilibration incubation with the protein. This procedure preferentially selects for the compounds with higher affinity, and results in an enhancement of the quantitative LC/MS response. Iterative stages of incubation, size-exclusion, and LC/MS allow the tighter binding components to be enriched relative to weaker binding components.

In this study, the peroxisome proliferator-activated receptor (PPAR γ), which is a target for anti-diabetic drugs (construct molecular weight of 32,537 Da), is incubated with 10 ligands that range in molecular weight from 283 to 587 units. A spin column of 6000 Da cutoff is used for SEC purposes. The retained mixture of components is analyzed by fast perfusive chromatography (35, 36), using a standard full-scan LC/MS strategy. This analysis procedure allows for the identification and quantitation of the protein and the ligands, compared to their responses prior to incubation. The ligand-protein complex that dissociated under the reversed-phase chromatographic conditions is selectively detected.

This analysis scheme provided a quick measurement of binding affinity, and serves as a screening tool during drug candidate selection. Spreadsheets were constructed and used to calculate the binding affinity of the components. In the example described above, two incubation cycles followed by the SEC separation provided an enhancement of strong binders to weak binders. This LC/MS-based method provides a unique approach to obtain information

in situations when lower concentrations of tighter binding ligands are present in the same mixture with higher concentrations of weaker binding ligands. Furthermore, this method is more efficient than synthetic deconvolution procedures and does not require the use of radioligands.

Combinatorial chemistry initiatives have created a tremendous challenge for activities that deal with the screening of these mixtures for activity against a specified target (37). MS-based approaches that use affinity selection (38), encoding methodologies (39–41), pulsed ultrafiltration (42), and anti-aggregatory approaches (43) have been described.

The use of MS formats that provide accurate mass capabilities have been recently illustrated for screening combinatorial libraries (44–47). The unambiguous confirmation/identification of combinatorial library components from small quantities of material have been illustrated using a hybrid quadrupole/orthogonal TOF (QTOF) (44, 45) and Fourier transform ion cyclotron resonance (FTICR) (47) mass spectrometry. A schematic of a FTICR-MS system is shown in Fig. 6. Accurate isotope patterns or "isotopic signature" and unique mass differences between isobaric compounds can be obtained using these two MS formats.

OPEN-ACCESS SYSTEMS

Chemists now routinely use open-access MS systems in the same way that they previously used thin-layer chromatography (TLC) to monitor reaction mixtures for the desired product and to optimize reaction conditions. In practice, medicinal chemists require only molecular weight data, and are comfortable with a variety of MS ionization methods to obtain this information. However, confidence in the actual method and procedure is a requisite. Today, molecular mass measurement has quickly become a preferred means of structure confirmation over NMR and IR during the early stages of synthetic chemistry activities (i.e., drug discovery), where sample quantities are limited.

In the open-access LC/MS procedure described by Pullen et al. (48), the samples are directly introduced from solution for ease of automation and sample preparation. Chemists prepare samples in solvent to a suggested concentration range, then log the samples into the system. The sample log-in is done at any time during the continuous automated queue. Autosampler vials are used to hold the samples, and autosamplers are used to directly deliver samples in solution to the mass spectrometer. The system uses a standard method to analyze the samples in

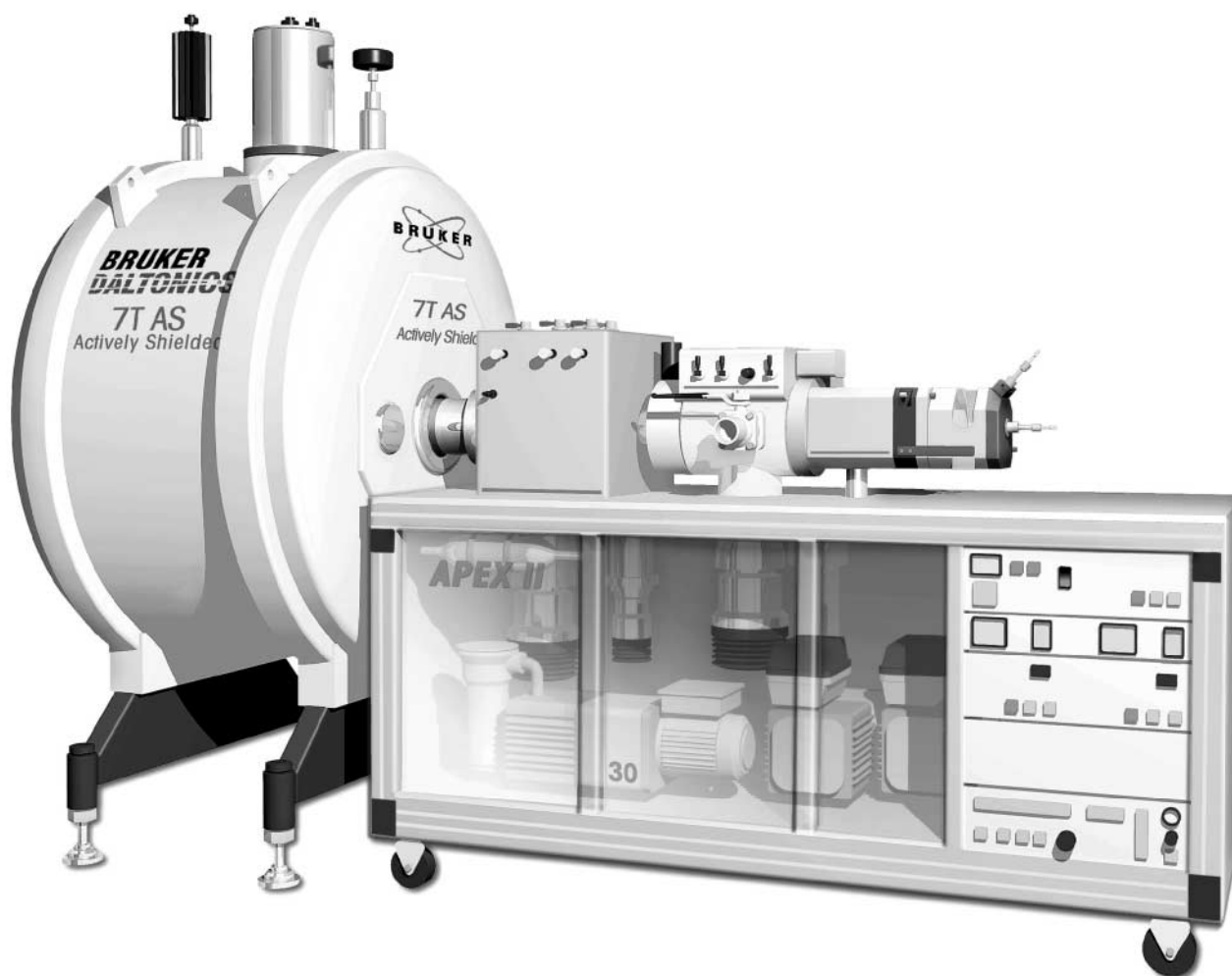


Fig. 6 Schematic of a FTICR MS instrument. This type of MS consists of an ion cyclotron resonance (ICR) analyzer cell that is situated in the homogeneous region of a large magnet. The ions introduced into the ICR analyzer are constrained (trapped) by the magnetic field to move in circular orbits with a specific frequency that corresponds to a specific mass-to-charge ratio (m/z). Mass analysis occurs when radiofrequency (rf) potential is applied (pulsed) to the ICR analyzer so that all ions are accelerated to a larger orbit radius. After the pulse is turned off, the transient image current is acquired and a Fourier transform separates the individual cyclotron frequencies. Repeating this pulsing process to accumulate several transients is used to improve the signal-to-noise ratio. (Courtesy of Bruker Daltonics, Billerica, MA.)

queue, average spectra according to a preset scheme, and print out a spectrum for the chemist. Fail-safe procedures for untrained users and instrument self-maintenance at start-up and shutdown were also developed.

Taylor et al. (49) further demonstrated the value of open-access LC/MS systems for generating a widened scope of pharmaceutical analysis applications, including: 1) characterization of synthetic intermediates and target compounds; 2) reaction monitoring; 3) reaction optimization; 4) analysis of preparative HPLC fractions; and 5) analysis of TLC plate spots. The availability of these methods led to the increased use of LC/MS for structural analysis. The short analysis time and reliable structure

confirmation resulted in the use of LC/MS as a first choice for structure characterization for synthetic chemistry applications.

Open-access LC/MS formats have spawned new dimensions in access and data management. The use of a direct exposure probe (DEP) for automated sample introduction has been developed for quick (ca. 3 min) molecular weight determination of new lead compounds and quantitative analysis (50). Figure 7 illustrates an automated direct probe system for molecular weight determination. Versatile software packages for data manipulation and processing has been a popular approach for integrating analysis and information (51–53). These software programs are

efficiently implemented with stand-alone computers and servers that are networked with open-access mass spectrometer data systems. In this configuration, the data are generated, visualized, processed, and automatically reported for the chemist. The program compares a template of predicted molecular ions with the actual ions generated by ESI and APCI for the quick analysis of synthetic products, intermediates, reactants, reagents, and contaminants. A list of observed ions along with known artifact ions is generated and used to provide a measure of the quality-of-fit to the predicted product(s).

Open-access MS systems provide an effective means for maintaining the high-throughput characterization of synthetic compounds. These systems offer an efficient laboratory- to bench-scale integration of sample generation and analysis activities. Advances in analytical instrumentation and electronic communication have also

played a major role in the emergence and acceptance of MS as a front-line tool for structure characterization.

IN VIVO DRUG SCREENING

The application of APCI-LC/MS techniques for the rapid determination of protein binding and pharmacokinetics in drug discovery were recently described by Allen et al. using a single quadrupole instrument (54). A “cocktail” approach consisted of four experimental compounds and a control compound dosed orally at 1 mg/kg with plasma samples obtained at 0.5, 1, 2, 4, and 8 h post dose. To insure reproducibility, the control compound was tested with each cocktail. This approach generated timely systemic exposure (AUC and Cmax)

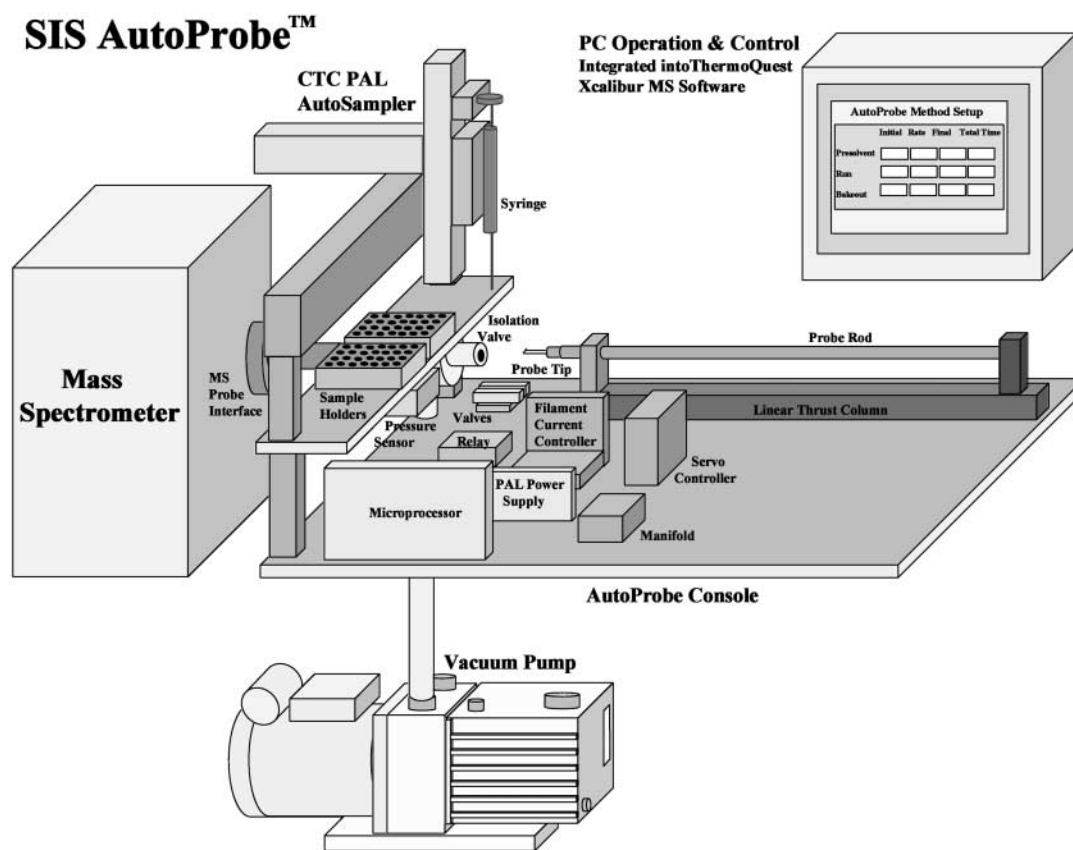


Fig. 7 Schematic of an automated direct probe system for molecular weight determination that features an ion trap MS. Samples are dissolved in a suitable solvent and injected via an automated syringe system onto the DEP wire. The probe is injected into the MS via an automated isolation valve system and the temperature is ramped to the programmed temperature. After sample analysis the probe is removed from the source and heated to a high temperature to clean the DEP wire in preparation for the next sample. This type of integrated, open-access MS-based application provides routine, unattended support for medicinal chemistry needs such as reaction monitoring and the optimization of reaction conditions. (Courtesy of Scientific Instrument Services, Ringoes, NJ.)

data on 44 test compounds in three work days, using two laboratory scientists.

The use of LC tandem quadrupole MS/MS-based screening approaches for quantitative bioanalytical measurements allow a large, chemically diverse, range of potential drug candidates to be analyzed quickly and confidently. A schematic of a tandem quadrupole MS/MS instrument is shown in Fig. 8. The development of unique LC/MS-based systems for *in vivo* pharmacokinetic screening reduces the analysis to a manageable number of samples, and results in a cost-effective approach to evaluate new lead compounds. Approaches to this type of methodology will likely vary, according to the behavior of the molecules of interest, standard operating procedures (SOPs), performance capabilities of the mass spectrometer, and integration of automated sample preparation, and data analysis procedures. Success will likely be dependent on the above parameters, as well as on the degree of tolerance to which the specific screen is set.

The simultaneous pharmacokinetic assessment of multiple drug candidates in one animal has been termed “n-in-one” or “cassette dosing.” As discussed for the previous example, this parallel approach results in an increased productivity for bioanalysis during drug discovery. Beaudry et al. (55) recently investigated the extension of this methodology to study larger numbers of compounds in each mixture, and to integrate sample preparation with the LC/MS/MS system for increased efficiency.

The number of analytes studied in parallel was extended to 63 plus an internal standard. The increased

number of analytes was made possible due to improvements to the collision region of the MS/MS system that provide increased sensitivity and reduced “memory effects.” In addition, robotic systems for sample handling and on-line (solid phase extraction) SPE of plasma samples were integrated with the LC/MS/MS system. An isocratic reversed-phase HPLC method provided a cycle time of 4.5 min per sample. The on-line sample preparation and short analysis resulted in an increased sample throughput that required less time from the scientist. The method produced good performance, in terms of extraction efficiency, linearity, and limit of detection (LOD), and has the capability of analyzing 320–960 samples per day. The strategic emphasis of this approach is on providing high throughput LC/MS methods for evaluating large numbers of drug candidates during drug discovery to eliminate poor pharmacokinetic performers.

METABOLIC STABILITY SCREENING

The use of fast gradient elution LC/MS techniques on a single quadrupole instrument was described for high throughput metabolic stability screening (56). The method uses as HPLC column-switching apparatus to desalt and analyze lead candidates incubated with human liver microsomes. Substrates were selected whose *in vivo* clearance is controlled predominantly by phase I oxidative metabolism as opposed to phase II metabolism or renal

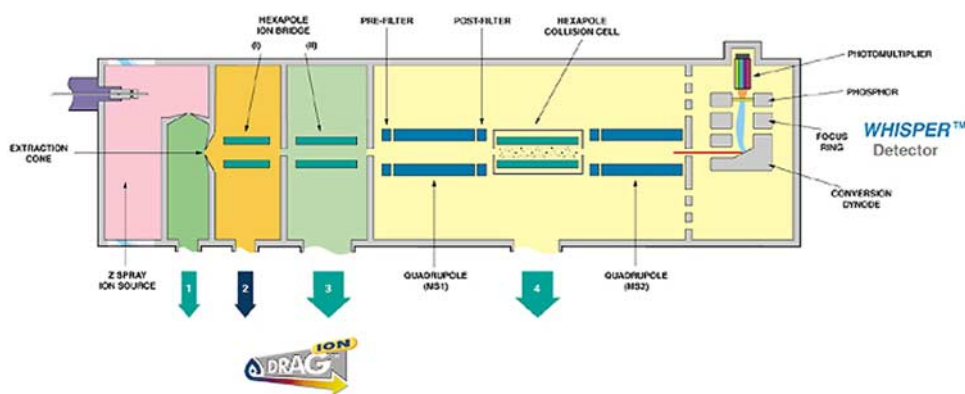


Fig. 8 Schematic of a tandem quadrupole MS/MS instrument. A tandem quadrupole MS/MS instrument consists of two quadrupole MS filters, MS1 and MS2, separated by a collision cell. Each quadrupole MS filter consists of four cylindrical or hyperbolic shaped rods. A unique combination of direct current (dc) potential and radiofrequency (rf) potential is applied to each pair of rods (one pair 180° out of phase with the other). A mass spectrum results by varying the voltages at a constant rf/dc ratio. A variety of scan modes (e.g., full scan, product ion, precursor ion, neutral loss) provide unique capabilities for quantitative and qualitative structure analysis. (Courtesy of Micromass, Manchester, UK.)

clearance. In this way, the resulting data could be resolved into four categories of metabolic stability: high ($\geq 60\%$); moderate ($\geq 30\text{--}59\%$); low ($\geq 10\text{--}29\%$); and very low ($< 10\%$).

The rapid structure identification of metabolites is a powerful complement to previously described quantitative approaches. The utility of an automated metabolite identification approach, using LC/MS/MS with an ion trap mass spectrometer has been demonstrated (57). In this study, MS^n analysis is automated to provide maximum structural information in combination with predictive strategies for biotransformation. Automated data-dependent scan functions are used to generate full scan, MS/MS, and MS^n mass spectra of metabolites within a single chromatographic analysis. This feature is unique and avoids the multiple (2–4) injections that are necessary with other MS/MS configurations (e.g., tandem quadrupole). Along with the significant savings in time, detailed structure information is generated, which enables a comprehensive analysis of substructure relationship to be constructed for each metabolite. These automated studies provide unique advantages during drug discovery, and provide an early perspective on the metabolically labile sites, or “soft spots” of a drug candidate. This knowledge is useful during lead optimization activities, and can lead to the initiation of proactive research efforts that deal with metabolism-guided structural modification and toxicity.

METABOLITE PROFILING AND IDENTIFICATION

The application of LC/MS-based techniques for the structure identification of drug metabolites has played a significant role in drug development. The early identification of drug metabolites provides valuable insights into the pathways of metabolism and biotransformation. Once metabolites are identified/confirmed, metabolism-guided structural modification during the drug discovery stage is initiated to facilitate the selection of drug candidates for subsequent development.

The identification of metabolite structures with LC/MS and LC/MS/MS techniques using quadrupole-based MS instruments are an effective approach due to their ability to analyze trace mixtures from complex samples of urine, bile, and plasma. The key to structure identification approaches is based on the fact that metabolites generally retain most of the core structure of the parent drug (58, 59). Therefore, the parent drug and its corresponding metabolites would be expected to undergo similar fragmentation and to produce mass spectra that indicate major substructures.

Kerns et al. demonstrated the application of LC/MS and LC/MS/MS standard method approaches in preclinical development for the metabolite identification of buspirone, a widely used anxiolytic drug (60). The success of this method relies on the performance of the LC/MS interface and the ability to generate abundant ions that correspond to the molecular weight of the drug and drug metabolites. The production of abundant molecular ions is an ideal situation for molecular weight confirmation because virtually all the ion current is consolidated into an adduct of the molecular ion (i.e., $[M + H]^+$, $[M + NH_3]^+$).

Full-scan mass spectra generally contain an abundant $[M + H]^+$ ion signal with little detectable fragmentation. Product-ion spectra are obtained to reveal product ions and neutral losses that are associated with diagnostic substructures of the buspirone molecule. To assist with the MS/MS structure identification, the gross substructure of buspirone is categorized into *profile groups* (61). Profile groups directly correlate specific product ions and neutral losses with the presence, absence, substitution, and molecular connectivity (62) of specific buspirone substructures and their modifications. The profile groups of buspirone are identified with abbreviations that correspond to the three specific substructures: azaspirone decane dione (A), butyl piperazine (B), and pyrimidine (P). Substituted substructures are designated with a subscript (s), and a dash (–) denotes substructure connectivity. Thus, the buspirone molecule is represented by A–B–P. The A_s –B–P designation refers to metabolite structures that contain the azaspirone decane dione, butyl piperazine, and pyrimidine substructures with substitution on the azaspirone decane dione substructure. The profile group categorization within a corresponding database allows the rapid visual recognition of primary substructures affected by metabolism.

Metabolite structure databases can be easily constructed and contain information on the structure, molecular weight, UV characteristics, RRT, and product ions of metabolites obtained from rat bile, urine, and liver S9 samples. Using this format, Kerns et al. reported the predominant buspirone metabolite profile groups as A_s –B–P, A–B–P_s, and A_s –B–P_s. These profile groups indicate azaspirone decane dione and pyrimidine as metabolically active sites of attack and the presence of multiple substitution sites on each of these substructures.

IMPURITY PROFILING AND IDENTIFICATION

Synthetic impurities are of particular concern during process research and safety evaluation activities. Often,

impurities are the result of synthetic by-products or starting materials of the scale-up process. Impurities provide a comprehensive indicator of the chemical process and are diagnostic of overall quality. Process chemists use this information to guide process optimization. Knowledge of the identity and relative amount of impurities is used to diagnose process reactions so that changes in reagents and reaction conditions leads to better yields and higher quality material.

With an increasing number of novel lead candidates that enter into preclinical development, considerable resources are needed to identify impurities. LC/MS-based approaches provide integrated sample clean-up and structure analysis procedures for the rapid analysis of impurities. This advantage was demonstrated during the preclinical development of TAXOL[®] (18). LC/MS played an important role for the identification of impurities contained in extracts and process intermediates from *Taxus brevifolia* and *T. baccata* biomass. Because drugs derived from natural sources often have a very diverse set of structural analogs, it is important to determine which analogs are carried through the purification process and ultimately appear as impurities. This task presents a unique challenge during the early stages of drug development due to the highly complex nature of the samples.

Kerns and coworkers described a structure identification strategy that incorporates LC/MS and LC/MS/MS techniques using quadrupole-based instruments for rapid, sensitive, and high-throughput impurity analysis (18). This approach integrates traditional steps of sample preparation, separation, analysis, and data management into a single instrumental method. The resulting multidimensional data include retention time, molecular weight, UV, and substructure information. A structure database is developed for each candidate and is used to rapidly identify the same impurities in new samples. Structures are proposed based on using the drug candidate as a structural template and, with the use of a standard method approach, consistency for comparison of results throughout the preclinical development process is ensured.

Nearly all of the impurities contained the characteristic paclitaxel core substructure as indicated by the characteristic product ion at m/z 509 with variations due to modifications. Many of these taxanes contained a side-chain similar to paclitaxel, with variations occurring on the terminal amide of the side chain. The product ions that differed from the characteristic side-chain ions of paclitaxel (m/z 286) by values indicative of specific substructures were used to identify these terminal amide variations. A comparison with the paclitaxel substructural template indicated structural differences beyond the position of the amide group in the side-chain substructure.

When a new impurity was encountered during chemical process research, retention time and molecular weight information were compared to the database for rapid identification. This approach is similar to the procedure described for natural product dereplication. If the compound is not contained in the structure database, then the corresponding LC/MS/MS analysis is performed to obtain substructural detail and the proposal of a new structure.

A standard reversed-phase HPLC method was used for all the samples that are associated with a drug candidate to reduce time-consuming method development/method refinement procedures. Standard reversed-phase methods typically involve a 20–30 min cycle time and provide information for a wide range of compounds. The incorporation of a standard method strategy allows the use of autosampling procedures and standard system software for data analysis.

During the development of TAXOL[®], 90 taxane impurities were rapidly identified and added to the structure database. This MS/MS information was routinely obtained for impurities down to the 100 ng level (injected), and required approximately 2–3 h for the analysis of each sample. The compounds are structurally categorized with profile group terminology. The LC/MS-based methods were significantly faster than the previously used analytical methods based on scale-up, isolation, fractionation, and individual structural analysis. Software tools capable of sample tracking, interpretation, and data storage facilitate the structure profiling of impurities, degradants, and metabolites (63). Key pharmaceutical analysis elements that deal with sample preparation, real-time analysis decisions, databasing, distribution/visualization of results (Fig. 9) and prediction of fragmentation are now highly integrated.

DEGRADANT PROFILING AND IDENTIFICATION

During the course of drug development, the bulk drug and drug formulation are studied under a variety of stress conditions such as temperature, humidity, acidity, basicity, oxidization, and light. Qin et al. described the utilization of stressing conditions that may cause degradation (64). The resulting samples may be used to validate analytical monitoring methods and to serve as predictive tools for future formulation and packaging studies.

A traditional approach to study degradant formation involves similar time-consuming scale-up and preparation steps as described for metabolite and impurity analysis. Similarly, this area of pharmaceutical analysis

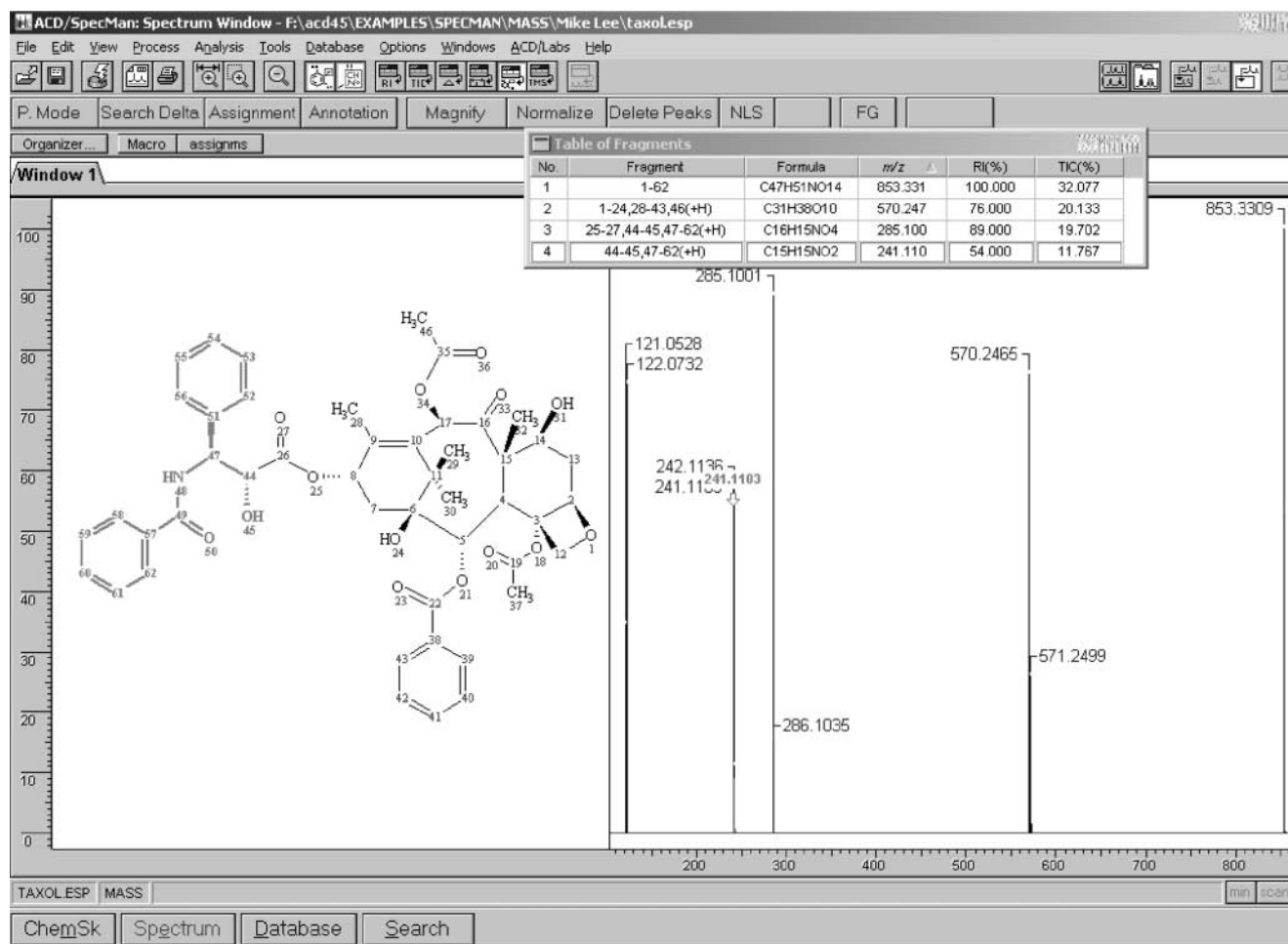


Fig. 9 Visualization of molecular fragments using a “lasso tool” application. The lasso tool is used to identify a particular fragment and, if a signal corresponding to its mass is present in the spectrum, the fragment is highlighted and the corresponding assignment is added to an assignment table. (Courtesy of Advanced Chemistry Development, Toronto, Ontario, Canada.)

has experienced the issues associated with faster drug development cycles. Rourick and coworkers recently described proactive approaches to obtain degradant information with quadrupole LC/MS methods during the preclinical development stage (65). The corresponding structural information provides insight for decisions on which leads to further develop for clinical testing. The early structural information on degradants of a drug candidate offers a unique capability for synthetic modification to minimize degradation. Structural information can also facilitate planning of preclinical drug development in process research, formulation development, and safety assessment.

The strategy for impurity and degradant identification described by Rourick et al. subjects lead candidates to various development conditions followed by LC/MS and

LC/MS/MS analysis protocols. A structure database is constructed from the corresponding results and is used to reveal unstable regions within the drug structure as well as to ascertain which candidate or homologous series of drug candidates may be the most favorable for further development. High capacity and throughput speed are necessary so that many lead candidates may be evaluated. Applicability of the method to a wide range of compound classes is desirable. Once the drug candidate enters clinical development and manufacturing, the structure database is useful for the rapid identification of impurities and degradants in samples generated during these stages of development.

The method exposes drug candidates to forced degradation conditions, (e.g., acid, base, heat, and moisture) as a *predictive profile*. The coordinated use of LC/MS and

LC/MS/MS provide structure identification for speed, sensitivity, and high throughput. Standard methods, useful for 80% of the compounds, are applied. Various types of structural data are obtained for elucidation purposes (e.g., retention time, molecular weight, MS/MS), and unknown compounds are elucidated with the candidate drug as a structural template. The LC/MS analysis provides retention time and molecular weight data, whereas LC/MS/MS provides substructural detail for structure identification. Drug candidates are incubated under drug processing, storage, and physiological conditions that were expected to occur throughout drug lifetime.

Using this approach, 10 degradants of cefadroxil, an orally effective semisynthetic cephalosporin antibiotic, were elucidated in a 2-day study. The use of standard LC/MS methods provided consistency from sample to sample throughout the development process, and allowed for the construction and use of a structural database for the rapid identification of impurities and degradants during development. The reversed-phase HPLC conditions provided a general measure of the polarity of each compound, useful for interpretation of substructural differences between related compounds. Due to the mass-resolving capability of the mass spectrometer, chromatographic resolution of co-eluting or unresolved components was not required. Abundant protonated molecule ions, $[M + H]^+$, provided reliable molecular weight information, and product-ion spectra generated valuable substructure information for each degradant. The product-ion spectrum of cefadroxil was used as a template for interpretation; specific product ions and neutral losses were compared to the spectra obtained from the unknown degradants. Product ions common to each spectrum provided evidence of substructures unchanged by the degradation conditions and differences were indicative of structural variations.

QUANTITATIVE BIOANALYSIS—SELECTED ION MONITORING

The quantitative analysis of targeted components in physiological fluids is a major requirement in clinical development. In 1991, Fouda et al. (66) pioneered the use of APCI-LC/MS on a single quadrupole instrument for the quantitative determination of the renin inhibitor, CP-80,794, in human serum. Because the pharmacological action is below 200 pg/ml, a quantitative assay in the low pg/ml range was required to monitor the drug's pharmacokinetic and pharmacodynamic properties. Also,

the structure of the CP-80,794 molecule lacked a significant chromophore for UV detection with conventional HPLC methods. Furthermore, the low volatility and thermal instability precluded analyses with GC/MS methods.

Quantitative LC/MS assays in clinical development generally involve four intensive steps: sample preparation; assay calibration; sample analysis; and data management. In the method developed by Fouda and coworkers, human serum samples were prepared with a liquid-liquid extraction procedure. Assay calibration involved the use of human serum samples fortified with CP-80,794 at 11 concentrations (6 replicates per concentration) ranging from 0.05 to 10 ng/ml. The LC/MS analysis involved the use of the SIM mode to monitor the molecular ions $[M-H]^-$ that correspond to the drug (m/z 619) and internal standard (m/z 633). In this particular LC/MS application, the negative ion mode was highly sensitive for this class of compound. Samples were loaded onto an HPLC auto-sampler and 80 μ L aliquots are injected onto the column at 4-min intervals. The elution times of the drug and internal standard were 3.1 and 3.4 min, respectively.

At the time, this application provided a powerful benchmark for the use of quadrupole LC/MS-based methods in the pharmaceutical industry and paved the way for the tremendous growth of MS-based applications in support of clinical development. This particular assay successfully supported several clinical studies with sensitive and reliable results. This performance was benchmarked on more than 4000 clinical samples, and led to a widened scope of MS application for quantitative bioanalysis (67–71).

QUANTITATIVE BIOANALYSIS—SELECTED REACTION MONITORING

The use of selected reaction monitoring (SRM) methods for quantitative bioanalysis represents increased dimensions of mass spectrometry analysis. A SRM method that features a tandem quadrupole MS/MS instrument for the quantitative analysis of an antipsychotic agent, clozapine, in human plasma was recently described by Dear et al. (72). Preclinical development studies of clozapine in rats and dogs used HPLC with fluorescence detection (FLD). With this method, a better limit of quantitation (LOQ) of 1 ng/ml was obtained. As the compound moved into the clinical stages of development, a more sensitive method of analysis was required to obtain rapid metabolic information in support of drug safety evaluation studies. As a result, a standard LC/MS/MS method was developed for

the quantitative analysis of clozapine (I) and four metabolites (II-V) in human plasma.

The LC/MS/MS strategy deployed is similar to previously described approaches for protein, natural products, metabolite, and impurity identification. An ionization technique that generates abundant molecular ion species with very little fragmentation is desirable. The product-ion spectrum is obtained to generate the substructural template of the molecule. Abundant and structurally unique transitions (molecular ion \rightarrow product ion) are identified from the spectrum, and are used in the corresponding SRM experiment for quantitation. The SRM experiment provides a high degree of selectivity and better LOD than full-scan or SIM experiments for the analysis of complex mixtures (73, 74). The selectivity of MS/MS reduces the requirements for complete chromatographic resolution of each component. Therefore, LC/MS/MS experiments for quantitation typically emphasize short analytical run times to provide high sample throughput.

The inter- and intra-assay precision (% C.V.) of this method were reported to be less than 8% across the range of the limits of quantification (0.05–10 ng/ml). The accuracy (% bias) for all spiked control concentrations did not exceed $\pm 4\%$. Same-day turnaround of results for over 100 samples was possible and was used to support an acute dose tolerance and pharmacokinetic study that involved the analysis of 1600 samples.

PEPTIDE MAPPING IN QUALITY CONTROL

Quality control involves a carefully designed series of analysis and protocols. The purpose of this activity during the manufacturing stage of drug development is to ensure the production of safe, high quality drug products. These measures are helpful for the producers of the product as well as the regulators (i.e., FDA). In this way, adherence to protocols and procedure are carefully monitored on a routine basis. When any uncertainty in the manufacturing process occurs, procedures are referenced and data are analyzed to determine the specific stage of manufacturing to begin examining. Thus, the responsibility of drug manufacturers and regulating agencies are to determine *when* and *how* a process went awry. The ability to do so in an efficient, straightforward manner is helpful to both parties, and ultimately, the consumer. Thus, the ability to provide this information is highly dependent on the manufacturer's ability to control the process.

The use of LC/MS in a manufacturing quality control environment for biologicals was reported by Chang et al. (75). Their approach involved three steps.

First, the expected cleavage sites (in this case, trypsin) within the amino acid sequence of the protein somidobove are indicated. This "tryptic map" serves as a template for the expected peptide fragments. Second, an analytical method using chromatography columns and conditions that provide the best resolution and reproducibility is developed. An opportunity exists to optimize the analysis based on chromatography, digestion, and LC/MS performance. Finally, the resulting LC/MS data are profiled, according to amino acid sequence, peak number, and $[M + H]^+$. Other properties such as relative retention time can also be added in this format.

IMPURITY IDENTIFICATION IN QUALITY CONTROL

Regulatory authorities strictly scrutinize the leachables (e.g., plasticizers, impurities) that may come from medical devices and drugs. It is the responsibility of the drug or medical device company to identify the leachables and to provide adequate testing of their toxicity. Monitoring methods must be developed and validated to effectively control toxic leachables during the manufacture of high quality pharmaceuticals.

As a result, materials for medical devices and drug products must be tested for leachable components. Once a known toxic compound is discovered, it must be identified for the assessment of toxicity, followed by the monitoring of levels using validated methods as required by the FDA. This identification procedure could be a time-consuming process with traditional methods that are based on fractionation and individual component analysis.

Tiller et al. demonstrated an analytical strategy with on-line LC/UV/MS and LC/MS/MS to rapidly obtain structural information for leachables from a drug-delivery device (76). Similar to proteomics-based applications, the analysis strategy makes use of "data-dependent" analysis, wherein the mass spectrometer first obtains molecular ions using full-scan techniques, and makes real-time decisions about MS/MS product-ion spectra that must be obtained. In this way, molecular weight and substructural information are both obtained for many components during a single HPLC run.

Many components were readily observed from the ESI-LC/MS chromatogram, and several polyester leachables were identified. The ESI LC/MS chromatogram revealed 15 components, compared to the three components that were observed in the 220 nm UV chromatogram. This difference illustrates the capability of ESI-LC/MS to provide a more universal detection when the analytes

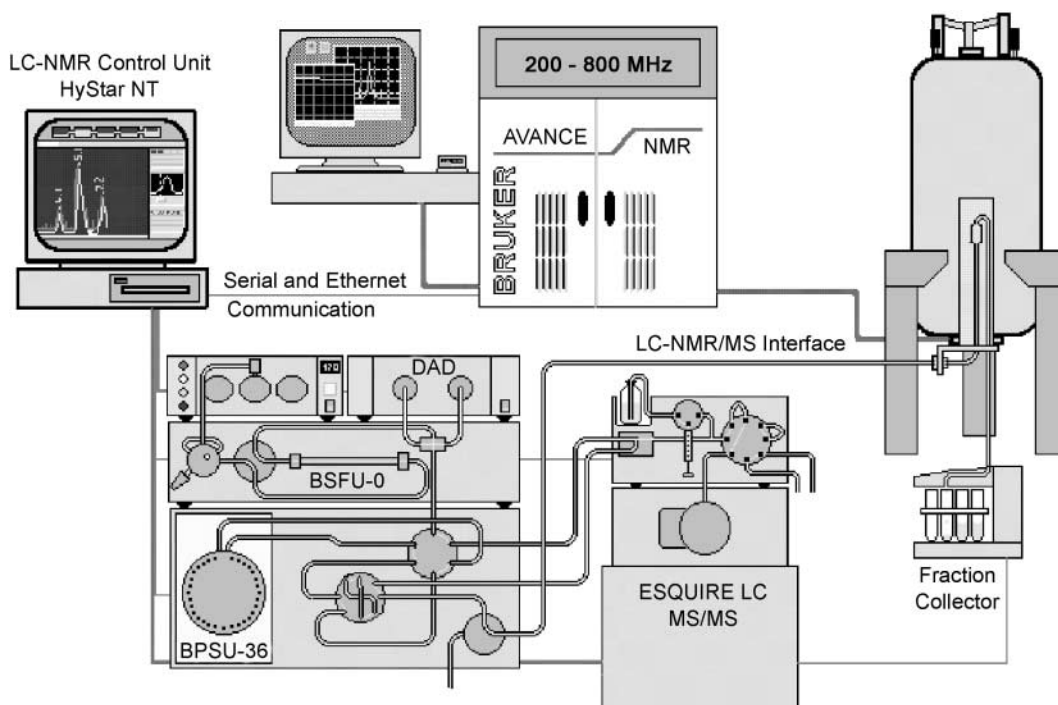


Fig. 10 Schematic of an LC/NMR/MS system. In this configuration, NMR is used to provide structural, stereochemical, or ligand binding information while MS is used to provide molecular weight information. (Courtesy of Bruker Daltonics, Billerica, MA.)

do not contain strongly UV-absorbing substructures (e.g., aromatic). The MS-based method proved to be highly efficient because molecular weight and substructural information via the full-scan and product-ion experiments, respectively, could both be obtained for the sample components.

CONCLUSIONS AND FUTURE PROSPECTS

The understanding and application of MS in the pharmaceutical industry has experienced tremendous growth due to unprecedented sample requirements (i.e., trace mixture analysis) and commercial pressure (i.e., faster development time-lines). Today, MS is an essential component of the modern pharmaceutical laboratory as well as an important complement to traditional methods of analysis.

An expanded role for highly sensitive assays for both quantitative and qualitative analysis will likely continue. Novel pharmaceutically relevant formats and systems that expand current approaches as well as create low-flow, miniaturized formats (77–79) are envisioned. Continued advances in integrated approaches that feature MS will generate significant interest with regard to pharmaceutical

analysis. For example, the use of integrated systems that feature NMR/MS have been described for applications that involve LC/NMR/MS (80,81) and MS/NMR (82). A schematic of an LC/NMR/MS system is shown in Fig. 10. In these configurations, NMR is used to provide structural and stereochemical information (80, 81) or to verify binding/interaction (82) while MS is used to provide molecular weight information. New challenges that deal specifically with probing mechanism of action will likely generate the need for a broader application of MS as well as spur the development of novel technologies for sample preparation, chromatography, and information management.

REFERENCES

1. Lee, M.S.; Kerns, E.H. *Mass Spectrom. Rev.* **1999**, *18*, 187–279.
2. Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, S.F.; Whitehouse, C.M. *Science* **1989**, *246*, 64–71.
3. Bruins, A.P.; Covey, T.R.; Henion, J.D. *Anal. Chem.* **1987**, *59*, 2642–2646.
4. Hillenkamp, F.; Karas, M.; Beavis, R.C.; Chait, B.T. *Anal. Chem.* **1991**, *63*, 1193A–1203A.
5. Burlingame, A.L.; Boyd, R.K.; Gaskell, S.J. *Anal. Chem.* **1998**, *70*, 647R–716R.

6. Kelleher, N.L. *Chem. Biol.* **2000**, *7*, R37–R45.
7. Wolf, C.R.; Smith, G. *Pharmacogenetics. Br. Medical Bull.* **1999**, *55*, 366–386.
8. Ross, P.; Hall, L.; Smirnov, I.; Haff, L. *Nat. Biotechnol.* **1998**, *18*, 1347–1351.
9. Haff, L.A.; Smirnov, I.P. *Genome Res.* **1997**, *7*, 378–388.
10. Pezzullo, L.II.; Hall, S.K.; Affourtit, J.P.; Seymour, A.B.; Stroh, J.G. *Proceedings of the 48th ASMS Conference on Mass Spectrometry and Allied Topics*, Long Beach California, 2000; 1552.
11. Yates, J.R. *J. Mass Spectrom.* **1998**, *33*, 1–19.
12. Gale, B.L.; Bleibaum, J.L.; MacKenzie, R.T.; Martin, R.L.; Straub, K.M. *Proceedings 48th ASMS Conference Mass Spectrometry and Allied Topics*, Long Beach, California, 2000; 62.
13. Morris, H.R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R.S.; Hoyes, J.; Bateman, R.H. *J. Rapid Commun. Mass Spectrom.* **1996**, *10*, 889–896.
14. Stevenson, T.I.; Loo, J.A. *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, Texas, 1999; 215.
15. Buko, A.; Tang, Q.; Trevillyan, J.; Chiou, G. *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, Texas, 1999; 248.
16. Chalmers, M.J.; Gaskell, S.J. *Curr. Opin. Biotechnol.* **2000**, *11*, 384–390.
17. Ackermann, B.L.; Regg, B.T.; Colombo, L.; Stella, S.; Coutant, J.E. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1227–1237.
18. Kerns, E.H.; Volk, K.J.; Hill, S.E.; Lee, M.S. *J. Nat. Prod.* **1994**, *57*, 1391–1403.
19. Strege, M.A. *J. Chromatogr. B* **1999**, *725*, 67–78.
20. Gilbert, J.R.; Lewer, P. *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, Florida, 1998; 21.
21. Janota, K.; Carter, G.T. *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, Florida, 1998; 557.
22. Burdick, D.J.; Stults, J.T. *Methods Enzymol.* **1997**, *289*, 499–519.
23. Fitch, W.L. *Mol. Divers.* **1998-99**, *4*, 39–45.
24. Süßmuth, R.D.; Jung, G. *J. Chromatogr. B Biomed. Sci. Appl.* **1999**, *725*, 49–65.
25. Swali, V.; Langley, G.J.; Bradley, M. *Curr. Opin. Chem. Biol.* **1999**, *3*, 337–341.
26. Dunayevskiy, Y.; Vouros, P.; Carell, T.; Wintner, E.A.; Rebek, J., Jr. *Anal. Chem.* **1995**, *67*, 2906–2915.
27. Cepa, S.; Searle, P. *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, Florida, 1998; 283.
28. Wang, T.; Cohen, J.; Kassel, D.B.; Zeng, L. *Comb. Chem. High Throughput Screen* **1999**, *2*, 327–334.
29. de Biasi, V.; Haskins, N.; Organ, A.; Bateman, R.; Giles, K.; Jarvis, S. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1165–1168.
30. Van Hijfte, L.; Marciniak, G.; Froloff, N. *J. Chromatogr. B Biomed. Sci. Appl.* **1999**, *725*, 3–15.
31. Weller, H.N. *Mol. Divers.* **1998-99**, *4*, 47–52.
32. Weller, H.N.; Young, M.G.; Michalczyk, S.J.; Reitnauer, G.H.; Cooley, R.S.; Rahn, P.C.; Loyd, D.J.; Fiore, D.; Fischman, S.J. *Mol. Divers.* **1997**, *3*, 61–70.
33. Zeng, L.; Wang, X.; Wang, T.; Kassel, D.B. *Comb. Chem. High Throughput Screen* **1998**, *1*, 101–111.
34. Davis, R.G.; Anderegg, R.J.; Blanchard, S.G. *Tetrahedron* **1999**, *55*, 1653–1667.
35. Regnier, F.E. *Nature* **1991**, *350*, 634–635.
36. Afeyan, N.B.; Fulton, S.P.; Regnier, F.E. *J. Chromatogr.* **1991**, *544*, 267–279.
37. Schriemer, D.C.; Hindsgaul, O. *Comb. Chem. High Throughput Screen* **1998**, *1*, 155–170.
38. Kaur, S.; McGuire, L.; Tang, D.; Dollinger, G.; Huebner, V. *J. Prot. Chem.* **1997**, *16*, 505–511.
39. Geysen, H.M.; Wagner, C.D.; Bodnar, W.M.; Markworth, C.J.; Parke, G.J.; Schoenen, F.J.; Wagner, D.S.; Kinder, D.S. *Chem. Biol.* **1996**, *3*, 679–688.
40. Hughes, I. *J. Med. Chem.* **1998**, *41*, 3804–3811.
41. Wagner, D.S.; Markworth, C.J.; Wagner, C.D.; Schoenen, F.J.; Rewerts, C.E.; Kay, B.K.; Geysen, H.M. *Comb. Chem. High Throughput Screen* **1998**, *1*, 143–153.
42. van Breeman, R.B.; Huang, C.-R.; Nikolic, D.; Woodbury, C.P.; Zhao, Y.-Z.; Venton, D.L. *Anal. Chem.* **1997**, *69*, 2159–2164.
43. Park, S.; Wana, L.; Johnson, M.E.; Venton, D.L. *J. Comb. Chem.* **2000**, *2*, 314–317.
44. Blom, K.F.; Combs, A.P.; Rockwell, A.L.; Oldenburg, K.R.; Zhang, J.H.; Chen, T. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1192–1198.
45. Lane, S.J.; Pipe, A. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 798–814.
46. Fang, A.S.; Vouros, P.; Stacey, C.C.; Kruppa, G.H.; Laukien, F.H.; Wintner, E.A.; Carell, T.; Rebek, J., Jr. *Comb. Chem. High Throughput Screen* **1998**, *1*, 23–33.
47. Speir, J.P.; Perkins, G.; Berg, C.; Pullen, F. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1937–1942.
48. Pullen, F.S.; Kerkins, G.L.; Burton, K.I.; Ware, R.S.; Teague, M.S.; Kiplinger, J.P. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 394–399.
49. Taylor, L.C.E.; Johnson, R.L.; Raso, R. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 387–393.
50. Scientific Instrument Services. *AutoProbeTM Application Note*; Ringoes: New Jersey, 2000.
51. Whitney, J.L.; Kerns, E.H.; Rourick, R.A.; Hail, M.E.; Volk, K.J.; Fink, S.W.; Lee, M.S. *Pharm. Tech.* **1998**, May, 76–82.
52. Tong, H.; Bell, D.; Keiko, T.; Siegel, M.M. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1174–1187.
53. Richmond, R.; Grolach, E.; Seifert, J.M. *J. Chromatogr. A* **1999**, *835*, 29–39.
54. Allen, M.C.; Shah, T.S.; Day, W.W. *Pharm. Res.* **1998**, *15*, 93–97.
55. Beaudry, F.; Yves Le Blanc, J.C.; Coutu, M.; Brown, N.K. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1216–1222.
56. Ackermann B.L.; Ruterbories K.J.; Hanssen B.R.; Lindstrom, T.D. *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, Florida, 1998; 16.
57. Lopez, L.L.; Yu, X.; Cui, D.; Davis, M.R. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1756–1760.
58. Perchalski, R.J.; Yost, R.A.; Wilder, B.J. *Anal. Chem.* **1982**, *54*, 1466–1471.
59. Lee, M.S.; Yost, R.A.; Perchalski, R.J. *Annu. Rep. Med. Chem.* **1986**, *21*, 313–321.

60. Kerns, E.H.; Rourick, R.A.; Volk, K.J.; Lee, M.S. *J. Chromatogr. B* **1997**, *698*, 133–145.
61. Kerns, E.H.; Volk, K.J.; Hill, S.E.; Lee, M.S. *J. Rapid Commun. Mass Spectrom.* **1995**, *9*, 1539–1545.
62. Lee, M.S.; Klohr, S.E.; Kerns, E.H.; Volk, K.J.; Leet, J.E.; Schroeder, D.R.; Rosenberg, I.E. *J. Mass Spectrom.* **1996**, *31*, 1253–1260.
63. Williams, A.; Lee, M.S.; Lashin, V. *Spectroscopy* **2001**, *16*, 38–49.
64. Qin, X.Z.; Ip, D.P.; Chang, K.H.; Dradransky, P.M.; Brooks, M.A.; Sakuma, T. *J. Pharm. Biomed. Anal.* **1994**, *12*, 221–233.
65. Rourick, R.A.; Volk, K.J.; Kerns, E.K.; Lee, M.S. *J. Pharm. Biomed. Anal.* **1996**, *14*, 1743–1752.
66. Fouda, H.; Nocerini, M.; Schneider, R.; Gedutis, C. *J. Am. Soc. Mass Spectrom.* **1991**, *2*, 164–167.
67. Wang-Iverson, D.; Arnold, M.E.; Jemal, M.; Cohen, A.I. *Biol. Mass Spectrom.* **1992**, *21*, 189–194.
68. Ayrton, J.; Dear, G.J.; Leavens, W.J.; Mallett, D.N.; Plumb, R.S. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1953–1958.
69. Kaye, B.; Herron, W.J.; Macrae, P.V.; Robinson, S.; Stopher, D.A.; Venn, R.F.; Wild, W. *Anal. Chem.* **1996**, *68*, 1658–1660.
70. Olah, T.V.; McLoughlin, D.A.; Gilbert, J.D. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 17–23.
71. Ayrton, J.; Plumb, R.; Leavens, W.J.; Mallett, D.; Dickens, M.; Dear, G.J. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 217–224.
72. Dear, G.J.; Fraser, T.J.; Patel, D.K.; Long, J.; Pleasance, S. *J. Chromatogr. A* **1998**, *794*, 27–36.
73. Johnson, J.V.; Yost, R.A. *Anal. Chem.* **1985**, *57*, 758A–768A.
74. Kusmierz, J.J.; Sumrada, R.; Desiderio, D.M. *Anal. Chem.* **1990**, *62*, 2395–2400.
75. Chang, J.P.; Kiehl, D.E.; Kennington, A. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1266–1270.
76. Tiller, P.R.; El Fallah, Z.; Wilson, V.; Juysman, J.; Patel, D. *Rapid Commun. Mass Spectrom.* **1997a**, *11*, 1570–1574.
77. Schultz, G.A.; Corso, T.N.; Prosser, S.J.; Zhang, S. *Anal. Chem.* **2000**, *72*, 4058–4063.
78. Li, J.; Wang, C.; Kelly, J.F.; Harrison, D.J.; Thibault, P. *Electrophoresis* **2000**, *21*, 198–210.
79. Li, J.; Kelly, J.F.; Chernushevich, I.; Harrison, D.J.; Thibault, P. *Anal. Chem.* **2000**, *72*, 599–609.
80. Holt, R.M.; Newman, M.J.; Pullen, F.S.; Richards, D.S.; Swanson, A.G. *J. Mass Spectrom.* **1997**, *32*, 64–70.
81. Shockcor, J.P.; Unger, S.E.; Savina, P.; Nicholson, J.K.; Lindon, J.C. *J. Chromatogr. B Biomed. Sci. Appl.* **2000**, *748*, 269–279.
82. Moy, F.J.; Haraki, K.; Mobilio, D.; Walker, G.; Powers, R.; Tabei, K.; Tong, H.; Siegel, M.M. *Anal. Chem.* in press.

SPECTROSCOPIC METHODS OF ANALYSIS—NEAR-INFRARED SPECTROMETRY

Emil W. Ciurczak

Purdue Pharma LP, Ardsley, New York, U.S.A.

INTRODUCTION

In recent years, there has been an increased interest in near-infrared (NIR) spectrometry. New instrumentation and algorithms have made the technique more powerful yet simple to use. The physics of NIR allow the samples to be analyzed as-is with, essentially, no preparation. NIR absorbance bands originate in the mid-infrared (MIR) region, but absorbencies are far lower (than those in the MIR), detectors are quieter, and light sources are stronger. The depth of penetration of the NIR light is such that pure materials may be analyzed, allowing whole tablets and bulk materials to be assayed quickly and easily.

Near-infrared radiation is the region of the electromagnetic spectrum between the visible and the (traditional) infrared. The nominal wavelengths for NIR are between 650 and 2500 nm (15,400 to 4000 cm^{-1}). The MIR or traditional infrared derives from the fundamental stretching, bending, and rotating motions of the molecule in question. Its wavelength range is between 2500 and 25,000 nm (4000 to 400 cm^{-1}).

Several aspects of a typical NIR spectrum derive directly from the physics of the region:

1. As the absorptivities in the NIR are 10–1000 times lower than that in the mid-range IR, the peaks are concomitantly smaller. This allows concentrated or pure samples to be scanned.
2. All the higher (1st through 6th) overtones of the O–H, N–H, C–H, and S–H bands from the IR are seen in the (much smaller) NIR region. This, in addition to combination bands (e.g., C=O stretch + N–H bend in protein), gives rise to a crowded spectrum with severely overlapping bands.
3. With lower absorptivities, the radiation penetrates deeper into a sample of either a pure material or mixture. Without dissolving, grinding, or diluting with an inert salt, the integrity of the compound is maintained. Thus, crystallinity, polymorphic form, particle size, and ratio or surface to lattice moisture is maintained.

With these differences from the mid-range IR, standard IR and UV/Vis/NIR instruments could not be counted on to produce satisfactory analyses of nonconventional pharmaceutical samples.

HARDWARE

Initially, work in the NIR was performed on instruments primarily designed for ultraviolet or visible analyses. The NIR portion of the spectrum was an afterthought in many instruments commercially available at that time: Beckman, Cary, Coleman, Perkin-Elmer, Shimadzu, Unicam, and Zeiss. The earliest instruments dedicated to NIR were based upon interference filters. The work performed by Karl Norris of the U.S. Department of Agriculture (Beltsville, MD) was initially done on the UV/Vis/NIR instruments mentioned previously. He established a number of important wavelengths associated with grains and other agricultural products. This group of “useful” wavelengths was then used to guide the wavelengths of the filters first used in the NIR instruments. Later, filters were added for tobacco and textile industries.

The first commercial instrument dedicated to NIR was developed by Dickey-John and presented in 1971 at the Illinois State Fair. The technology was licensed by Technicon and introduced soon thereafter. Technicon then developed its own scanning monochromator-based instrument. This was followed by instruments from Pacific Scientific (now FOSS-NIR Systems) and LT Industries, to name a few companies.

In the late 1980s and early 1990s, interferometer-based instruments were introduced by companies already producing IR equipment: Nicolet, Bomem, Perkin-Elmer, and others. Often, these FT-NIR instruments were merely adapted FT-IRs that were already in existence and were not suited for pharmaceutical samples: pure raw materials, blends and granulations, tablets and capsules, and larger plastic containers.

A number of instrument suppliers as well as engineering specialty houses are able to install in-line NIR hardware in

Table 1 Typical applications of NIR in pharmaceutical production

Quantitative	Qualitative
Moisture in pure substances	Identification of pure substances
Moisture in freeze-dried materials	Identification of mixes
Mean particle size of pure substances	Qualification of pure substances
Mean particle size of mixtures/granulations	Qualification of mixes
Assay/uniformity of blends	Identification of clinical dosage forms
Content uniformity of tablets/capsules	Identity of packaging materials
Level of coating	
Prediction of dissolution times	
Tablet hardness	
Degradation products (tablets)	
Polymorphism	
Crystallinity	

production settings. For up-to-date listings and manufacturers, a newcomer (or even experienced user) should scan the Internet, attend meetings such as PittCon, IFPAC, and more specialized conferences such as the International Diffuse Reflectance Conference (Chambersburg, PA on even-numbered years) or the International NIR Conferences (Europe and Asia on various years).

SOFTWARE

Software, per se, was not available for NIR until the 1980s. Technicon (Tarrytown, NY) and Pacific Scientific (Silver Spring, MD) offered software for their scanning monochromator-based (holographic grating) instruments. The software ran their instruments, collected the spectra, and performed simple calculations. At that time, the only algorithms for NIR used multiple linear regression as the core math treatment (“step-up search” and “best-combination”). Spectral treatments such as smoothing and derivatives were available in the late 1980s, but little in the way of “higher math” treatments.

Throughout the 1980s, researchers were developing private software packages for principle components analyses (PCA) and partial least squares (PLS). These matrix-based algorithms have been commercially available for over a decade from the instrument vendors or from third-party vendors. The PCA and PLS algorithms supplied by instrument vendors are more pragmatic and tend to change more slowly than those provided by vendors whose primary function is to provide software. Recently, artificial neural networks (ANNs or NNs) have attracting attention. These are best discussed in a text devoted to the subject and have not yet been used in industrial settings.

Third-party software vendors perform an important part in supplying software for equation development. These vendors exist solely for providing arithmetic tools for the analyst. Instrument vendors are good at providing software to run their instruments and collect quality spectra. The strong point of software providers is their ability to write software solely for analysis. These programs work, in many cases, better than the software instrument companies because they are single-purposed and not multifunctional (i.e., complex and prone to longer development times and, therefore, bugs.) A listing of these vendors would be at best, ephemeral. Large conferences would be the place to start looking for current suppliers.

The first commercial qualitative analysis software was introduced in 1984 (1, 2) by Technicon. Based on Mahalanobis distance, the program used the absorbances of materials at wavelengths chosen by the software to classify them. This was first applied to incoming raw materials. Since then, virtually every manufacturer of hardware also supplies software capable of performing either qualitative or quantitative analyses. As will be touched on in applications, identification and quality assessment of incoming raw materials is commonplace in the industry.

APPLICATIONS

Owing to the late development of fast, small, and inexpensive computers, little work was performed on pharmaceutical samples until the 1980s. In 1966 (3), many pharmacologically active amine salts were investigated both in solid state and in solution. Two drugs were quantified in 1967 (4): allylisopropylacetureide and

Table 2 Some types of NIR instrumentation and suppliers

Instrument type	Representative manufacturers
Interference filter	Bran & Leubbe Dickey-John, Churchill Factory of L.I. (Hungary) Foss Electric General Analysis Corp. Infrared Engineering Moisture Systems Corp. Oxford Analytical, Inc. Percon GmbH Perten Trebor Industries
Scanning grating	Bran & Leubbe EG&G Princeton Applied Research Foss-NIRSystems Guided Wave, Inc. Hitachi Instruments, Inc. L.T. Industries Shimadzu Trebor Industries Tecator AB Varian
Interferometer-based	Bran & Leubbe Bomem, Inc. Buchi Mattson Instruments Midac Corp. Nicolet Instrument Corp. Perkin-Elmer Corp.
Acousto-optic tunable filter	Bran & Leubbe Brimrose Corp. EG&G Princeton Applied Research Infrared Fiber Systems L.T. Industries Rosemont
Diode-array	EG&G Princeton Applied Research KES Analysis Perten

phenacetin were dissolved in chloroform and simultaneously quantified at 1983 and 2019 nm, respectively.

Water was determined in several matrices (5) in 1968. Solid samples were analyzed for hydrous and anhydrous forms of strychnine sulfate, sodium tartrate, and ammonium oxalate mixed with KCl and compressed into disks containing 100 mg KCl and 25 mg drug. The water band at ~1940 nm was used for the hydrate analysis.

Since the middle 1980s, there has been an ever-expanding list of applications, both qualitative and

quantitative, of NIR in pharmaceuticals. These applications are as follows:

1. Particle Size. Since the first reports of particle size measurement by NIR in 1985 (6) and 1986 (7), numerous researchers have reported using NIR for particle size assessment, both in pure samples and mixes (e.g., 8, 9).
2. Hardness. As NIR spectra are affected by physical and chemical changes within a sample, physical parameters such as hardness may be measured. In 1997, a paper by Morrisseau and Rhodes (10) was published wherein hardness from 2 to 12 kg was estimated by NIR. Several other workers have reported success in measuring hardness: Ebube et al. in 1999 (11) and Kirsch and Drennen (12) have done some quality work on the topic.
3. Polymorphism crystallinity. One of the earliest reported uses of NIR for polymorphism was for the polymorphs of caffeine in 1985 (13). In 1987, Gimet and Luong (14) used NIR to ascertain changes in crystallinity during processing.

Differentiation between polymorphs was performed by pattern recognition in 1995 by Aldrich et al. (15). The actual control of a process was reported in 1998 by DeBraekeleer et al. (16), where they described using PCA, SIMPLISMA, and orthogonal projections to correct for temperature variation during the monitoring of polymorph conversion. This is a real-time, in-line procedure.

Several recent papers have also highlighted NIR for polymorphism (17–20). The advantage of NIR for reading polymorphic changes is that the C–C, C–O, and other non-H bonds are not visible. The N–H and O–H bands dominate the spectrum, making crystal changes obvious for qualitative analysis.

4. Enantiomers/structural isomers. The particular optical isomer of a drug being used in a formulation is quite important. For example, quinine is used to treat malaria; quinidine, its optical isomer, is used for heart arrhythmia. In 1985, Ciurczak (21) observed that although pure D- and L-amino acids gave identical spectra, the racemic mixtures (DL-) produced an entirely different spectra. Some work was presented by Ciurczak (22) in 1986, which was later expanded and published (23) by Buchanan et al. in 1988, where the enantiomer ratio was determined via NIR.

Mustillo and Ciurczak (24) presented a paper discussing the spectral effect of optically active solvents on enantiomer mixes. This information was used as a technique to screen for polar modifiers in normal-phase chromatography of racemic mixtures (25).

In 2000, the enantiomeric composition of ibuprofen in solid-state mixtures was performed by Agatonovic-Kurstin et al. (26).

Structural or geometric isomers as well as nearly identical molecules may be discriminated by NIR. Some early work was performed by Kradjel and Ciurczak (27) and presented in 1985. Work has been performed where the type of polyvinylpyrrolidone is determined by NIR (28). The only differences between types are polymerization conditions.

5. Identification Quality Assessment. In a first of its kind paper, Rose in 1982 showed that a number of structurally similar penicillin-type drugs could be identified and determined by NIR. (29). At Sandoz, Ciurczak in 1984 reported using Mahalanobis distance-based algorithms for the identification of raw materials (30). Ciurczak also reported the use of spectral matching (SM) and principle component analysis (PCA) for raw materials (31) and suggested a method for introducing variations into samples for more robust equation development in 1986 (32). NIR has been in use for raw material ID since then in companies worldwide.
6. Blend Uniformity. An early presentation (33) of the potential for following a dry blend through its stages of mixing by Ciurczak was expanded and published in 1990 (34). A fiber probe was used on a mixture, probing at several locations to several depths. The spectra were then compared with a fully blended sample, using both a SM and a PCA program.

Another similar paper was published by Blanco et al. in 1999 (35). They also used a fiber optic probe on stopped blenders and showed for yet another blend that NIR is a useful tool. Further work was reported on blend uniformity in 1996 by Drennen et al. (36), where samples were sieved and assayed for hydrochlorothiazide by reflection NIR.

Workers at Pfizer patented an automated system specifically designed for a "V-Blender." Its development is described in 1996 in a paper by Sekulic et al. (37). This design calls for a single-fiber probe to be permanently inserted in the shaft of the blender. This device was used by DeMaesschalck et al. in 1998, and a paper was published in this work. The significance of this paper was that workers outside Pfizer also used this device successfully.

A study that used both ports along the body of a blender and a two-dimensional imaging technique to measure homogeneity is described in (38). This approach uses multiple measurement points and the speed of NIR. The paper also describes a two-dimensional imaging of the surface to assess uniformity.

Papers by DeMaesschalck et al. (39) and Bemtsson et al. (40) are significant in that they discuss the mathematical treatments involved in following blend uniformity and in-line analysis of active.

7. Moisture. In NIR, as in the MIR, water has the strongest absorption of light. It stands to reason that moisture in solid products would be an important assay. For instance, Warren et al. (41) used NIR to determine the amount of water in glycerides. Transmission spectra of standard propylene glycol and glycerin were used to calibrate and measure for the water content.

Correlation of total, bound, and surface water in raw materials was the topic of a paper by Torlini and Ciurczak in 1987 (42). The NIR was calibrated by Karl Fischer titration, differential scanning calorimetry (DSC), and thermal gravimetric analysis (TGA) for the various water types. NIR could distinguish surface from bound water, whereas standard loss on drying (LOD) could not.

In a recent paper, Derksen et al. (43) showed how NIR could be used to determine the moisture content through the glass vials for freeze-dried samples with varying amounts of active.

In-line moisture of a granulation was discussed by Rantanen et al. (44) in 2000. A binder was added in a solution and the granulation evaluated for residual moisture. Derivatives were used to minimize the particle size effects. Although the measurements were performed (initially) with a full scanning instrument, the authors were able to use fixed wavelength instruments for routine usage. This is an important introduction to simple process control.

Ciurczak discusses similar simple mono-wavelength devices in a recent article (45), where he shows "spark plug" type of mini-infrared source-detector combination. This device should soon be available in the NIR range for similar applications.

8. Clinical Supplies. A rapidly growing use for NIR is the discrimination between clinical dosage forms and placebos that are made to look identical to them. Normally, unmarked tablets or capsules are arranged in a specific manner in blister packs. The identity of which is the active is kept from both the patient and the attending doctor (thus the term "double-blind").

An early application of NIR to clinical samples was published by Aldrich et al. (46) in 1994. Using a custom fiber optic assembly, samples were scanned directly through the polymeric material of the blister packs. This demonstrated the power of discrimination of NIR.

In two papers, Ritchie et al. (47, 48) described an approach for performing NIR qualitatively with an eye

to cGMPs (current good manufacturing practices). As clinical lots are often ad hoc formulations, it is difficult to gather enough lots to generate a discriminant equation prior to an actual clinical trial. They developed a procedure whereby equations are quickly generated for a particular study, based on that placebo and active, used for the study, then discarded. Thus, a full validation procedure is shortened to half-day per study.

9. Content Uniformity/assay. The question most often asked is when NIR will be able to be used as a release test. In the earliest NIR assays, tablets and capsules were not analyzed intact. Prior to scanning, the active was extracted and scanned in a clear liquid. The first use of NIR for tablets was reported in 1968 (49). Sherken assayed the meprobamate content in tablet mixtures and commercial preparations. Allen (50) used NIR to analyze a three-component mix: carisoprodol, phenacetin, and caffeine. The powder was extracted with chloroform and scanned in the NIR. Several other publications took advantage of the "dissolve and scan" approach (51, 52).

The first published analysis for powders was written in 1981 (53) by Beconsall et al. It described the analysis of propranolol and magnesium carbonate mixes. In 1987, Ciurczak and Torlini (54) published on the analysis of solid and liquid dosage forms using reflectance and transreflectance, respectively.

In 1987, Chasseur assayed cimetidine granules (55). This was the first reported use of a synthetically produced extended range for calibration; calibration samples were produced with a range of content from 70 to 130% of label. In 1987, Osborne (54) used NIR to assay for nicotinamide in vitamin premixes.

In 1988, Jensen et al. used NIR to analyze amiodarone tablets (57). The approach was somewhat primitive: the tablets were scraped of their coating and glued to an aluminum plate. Reflectance was used and six wavelengths were needed to analyze the tablets.

Two chapters of collections were published in the early 1990s wherein solid dosage form analysis was discussed. Based on a presentation at the Fourth International Conference on NIRS (58), Stark used a newly developed diode-array spectrometer scanning between 520 and 1800 nm to analyze specimens. Samples were placed on glass slides and the light collected at a fixed angle. Intact acetaminophen, ibuprofen, and antacid tablets were collected.

In a second chapter (59), Monfre and DeThomas wrote about the calibration of a NIR instrument for the analysis of a prescription vasodilator. Individual tablets were used, but still needed to be crushed for analysis. The important

part of the calibration was that no synthetic tablets were produced to extend the range. Using only production tablets ranging from 96% to 102% of label claim, the NIR results were within 0.5 mg of the HPLC values.

Borer et al. used NIR to evaluate the key sources of variation in tablet analysis in 1998 (60). The effects of active variation as well as instrumental variations were evaluated statistically. Factors such as the opening (light beam width), derivative segment size, days between measurement and calibration of the instrument, and tablet orientation were found to be significant. This study tends to affirm the value of transmission measurements where the tablet orientation and optical configurations are fixed.

NIR was used to determine the effects of changes in magnesium stearate concentration and variations in compression pressure on tablet analysis (61). Various types of Avicel (microcrystalline cellulose), varying mostly in particle size, were compressed into tablets with or without magnesium stearate as lubricant by using various compression pressures. Various mathematical treatments were used to either measure the differences or obviate their effects on the analysis.

CONCLUSIONS

The above list of applications is not intended to be a comprehensive listing of all workers or findings in the field of NIR. Rather, it is an overview and can be used as a starting point for informational purposes. By performing a citation search on the authors cited here, the people who have used their work as a basis of current research can be known and will give the analyst an extended list of authors to follow-up for additional discoveries.

REFERENCES

1. Mark, H.L.; Tunnell, D. *Anal. Chem.* **1985**, *57*, 1449.
2. Ciurczak, E.W. *Annual Symposium on NIRA*. 7th Technicon: Tarrytown NY, 1984.
3. Sinsheimer, J.E.; Keuhnelian, A.M. *J. Pharm. Sci.* **1966**, *55*, 1240–1244.
4. Oi, N.; Inaba, E. *Yakugaku Zasshi* **1967**, *87*, 213–215.
5. Sinsheimer, J.E.; Poswalk, N.M. *J. Pharm. Sci.* **1968**, *57*, 2006–2010.
6. Ciurczak, E.W. *Proceedings of the 1st Pan American Conference*, October, 1985, San Juan, PR.
7. Ciurczak, E.W.; Torlini, R.P.; Demkowitz, M.P. *Spectroscopy* **1986**, *1* (7), 36.
8. Ilari, J.L.; Martens, H.; Isaksson, T. *Appl. Spectrosc.* **1986**, *45* (5), 722.

9. O'Neil, A.J.; Jee, R.D.; Moffat, A.C. *The Analyst* **1999**, *124*, 33–36.
10. Morisseau, K.M.; Rhodes, C.T. *Pharm. Res.* **1997**, *14* (1), 108.
11. Ebube, N.K.; Thosar, S.S.; Roberts, R.A.; Kemper, M.S.; Rubinovitz, R.; Martin, D.L.; Reier, G.E.; Wheatley, T.A.; Shukla, A.J. *Pharm. Develop. Tech.* **1999**, *4* (1), 19–26.
12. Kirsch, J.D.; Drennen, J.K. *J. Pharm. Biomed. Anal.* **1999**, *19*, 351–362.
13. Ciurczak, E.W. *Proceedings FACSS* October 1985, Philadelphia, PA.
14. Gimet; Luong, T.J. *Pharm. Biomed. Anal.* **1987**, *5*, 205–211.
15. Aldrich, P.K.; Evans, C.L.; Ward, H.W.; Colgan, S.T.; Gemperline, P.J. *Anal. Chem.* **1996**, *68* (6), 997.
16. DeBraekeleer, K.; Cuesta Sanchez, F.; Hailey, P.A.; Sharp, D.C.A.; Pettman, A.J.; Massert, D.L. *J. Pharm. Biomed. Anal.* **1998**, *17*, 141.
17. Caira, M.R. *Topics in Current Chemistry*; Springer Verlag: Berlin, 1998.
18. Blanco, M.; Coello, J.; Iturriaga, H.; MasPOCH, S.; Perez-Maseda, C. *Anal. Chim. Acta* **2000**, *407*, 247–254.
19. Patel, A.D.; Luner, P.E.; Kemper, M.S. *Internat. J. Pharmaceut.* **2000**, *206*, 63–74.
20. Bauer, H.; Herkert, T.; Bartels, M.; Kovar, K.-A.; Schwarz, E.; Schmidt, P.C. *Pharm. Ind.* **2000**, *62* (3), 231–235.
21. Ciurczak, E.W. *Proceedings FACSS* October 1985, Philadelphia, PA.
22. Ciurczak, E.W. *Proceedings of the 26th Annual Conference Pharmaceutical Analysis* August, 1986, Merrimac, WI.
23. Buchanan, B.R.; Ciurczak, E.W.; Grunke, A.; Honigs, D.E. *Spectroscopy* **1988**, *3* (9), 54.
24. Mustillo, D.M.; Ciurczak, E.W. *Proceedings of the Eastern Analytical Symposium* November Somerset NJ
25. Ciurczak, E.W.; Dickinson, T.A. *Spectroscopy* **1991**, *6* (7), 36.
26. Agatonovic-Kustrin, S.; Beresford, R.; Razzak, M. *Anal. Chim. Acta* **2000**, *417*, 31–39.
27. Kradjel, C.; Ciurczak, E.W. *Proceedings of the 1st Pan American Chemical Conference*, October 1985, San Juan PR.
28. Kreft, K.; Kozamernik, B.; Urleb, U. *Int. J. Pharmaceut.* **1999**, *177*, 1–6.
29. Rose, J.J.; Prusik, T.; Mardekian, J.J. *Parenter. Sci. Technol.* **1982**, *36* (2), 71.
30. Ciurczak, E.W. *Proceedings of the 7th Annual Symposium on NIRA* Technicon, Tarrytown, NY, 1984.
31. Ciurczak, E.W. *Proceedings of the AAPS National Meeting* November 1990, Las Vegas, NV.
32. Ciurczak, E.W. *Proceedings of the 9th Annual Symposium on NIRA* Technicon, Tarrytown, NY, 1986.
33. Ciurczak, E.W. *Proceedings of the AAPS National Meeting* November 1990, Las Vegas, NV.
34. Ciurczak, E.W. *Pharm. Tech.* **1991**, *15* (9), 141.
35. Blanco, M.; Coello, J.; Eustaquio, A.; Iturriaga, H.; MasPOCH, S. *The Analyst* **1999**, *124*, 1089–1092.
36. Wargo, D.J.; Drennen, J.K.J. *Pharm. Biomed. Anal.* **1996**, *14* (11), 1414.
37. Sekulic, S.S.; Ward, H.W., II; Brannegan, D.R.; Stanley, E.D.; Evans, C.L.; Sciavolino, S.T.; Hailey, P.A.; Aldrich, P.K. *Anal. Chem.* **1996**, *68*, 509–513.
38. El-Hagrasy, A.S.; Morris, H.R.; D'Amico, F.; Lodder, R.A.; Drennen, J.K. *J. Pharm. Biomed. Anal.* in press.
39. De Maesschalck, R.; Cuesta-Sanchez, F.; Massart, D.L.; Doherty, P.; Hailey, P. *Appl. Spectrosc.* **1998**, *52* (5), 725–731.
40. Berntsson, O.; Danielsson, L.-G.; Johansson, M.O.; Folestad, S. *Analyt. Chim. Acta* **2000**, *419*, 45–54.
41. Warren, R.J.; Zaembo, J.E.; Chong, C.W.; Robinson, M.J. *J. Pharm. Sci.* **1970**, *59* (1), 29.
42. Torlini, R.P.; Ciurczak, E.W. *Proceedings of PittCon March*, 1987, Atlantic City.
43. Derksen, M.W.J.; van de Oetelaar, P.J.M.; Maris, F.A. *J. Pharm. Biomed. Anal.* **1998**, *17*, 473–480.
44. Rantanen, J.; Rasanen, E.; Tenhunen, J.; Kansakoski, M.; Mannermaa, J.-P.; Yliruusi, J. *Eur. J. Pharm. Biopharm.* **2000**, *50*, 271–276.
45. Ciurczak, E.W. *Spectroscopy* **2001**, *16* (1), 17–18.
46. Aldrich, P.K.; Mushinsky, R.F.; Andino, M.M.; Lewis, C.L. *Appl. Spectrosc.* **1994**, *48*, 1272–1276.
47. Ritchie, G.E.; Mark, H.L.; Ciurczak, E.W. *9th Diffuse Reflectance Conference*, August 1998; Chambersburg, PA.
48. Ritchie, G.E.; Tso, C.; Ciurczak, E.W. *Proceedings of PittCon March 2000*, New Orleans, LA.
49. Sherken, S.J. *Assoc. Offic. Anal. Chem.* **1968**, *51*, 616–618.
50. Allen, L. *J. Pharm. Sci.* **1974**, *63*, 912–916.
51. Zappala, A.F.; Post, A. *J. Pharm. Sci.* **1977**, *66*, 292–293.
52. Corti, P.; Dreassi, E.; Corbini, G.; Lonardi, S.; Gravina, S. *Analisis* **1990**, *18*, 112–116.
53. Beconsall, J.K.; Percy, J.; Reid, R.F. *Anal. Chem.* **1981**, *53*, 2037–2040.
54. Ciurczak, E.W.; Torlini, R.P. *Spectroscopy* **1987**, *2* (3), 41–43.
55. Chasseur, J.C. *Chim. Oggi.* **1987**, *6*, 21–27.
56. Osborne, B.G. *Analyst* **1987**, *112*, 313–315.
57. Jensen, R.; Peuchant, E.; Castagne, I.; Roux, G. *Spectrosc. Int. J.* **1988**, *6*, 63–72.
58. Stark, E. *Making Light Work: Advances in Near-Infrared Spectroscopy*; Murray, I., Cowe, I., Eds.; VCH: Weinheim Germany, 1991; 27–34.
59. Monfre, S.L.; DeThomas, F.A. *Near-Infrared Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications*; Hildrum, K.I., Ed.; Ellis-Hornwood: Chichester UK, 1992; 435–440.
60. Borer, M.; Zhou, X.; Hays, D.M.; Hofer, J.D.; White, K.C.J. *Pharm. Biomed. Anal.* **1998**, *17*, 641–650.
61. Ebube, N.K.; Thosar, S.S.; Roberts, R.A.; Kemper, M.S.; Rubinovitz, R.; Martin, D.L.; Reier, G.E.; Wheatley, T.A.; Skukla, A.J.

SPECTROSCOPIC METHODS OF ANALYSIS—ULTRAVIOLET AND VISIBLE SPECTROPHOTOMETRY

R. Raghavan
Jose C. Joseph

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

INTRODUCTION

Many compounds of pharmaceutical interest are colorless. In the early part of the 20th century, analytical methods were developed using color-forming reactions and measurement of color for quantitation. With the development of instrumentation and based on an understanding of optics, ultraviolet and visible spectrophotometric techniques advanced further. However, with rapid development of atomic absorption spectroscopy, plasma atomic spectroscopy, and GC and HPLC techniques, the interest in spectrophotometry waned in the late 1980s and in 1990s. HPLC methods and HPLC instruments advanced further. As a result, rapid scan techniques and microprocessor control of such instruments led to the development of sophisticated UV–vis spectrophotometric detectors. With renewed interest in the application of UV–vis spectrophotometric techniques, a whole new generation of spectrophotometers with microprocessor instrument control, data acquisition, data handling, and data smoothing are currently available.

BASIC CONCEPTS AND DEFINITIONS

Electromagnetic Radiation

Spectrophotometry involves the measurement of the chemical species using light energy. Light is a form of radiation called electromagnetic radiation. Electromagnetic radiation spans a very large region involving very low energy waves (<0.0012 kJ/mol) to extremely energetic cosmic radiation ($>1.2 \times 10^8$ kJ/mol). This energy can be related by the equation $E = h\nu$, where h is a constant known as Planck's constant (equal to 6.62×10^{-24} J/s), ν is the frequency, which is related to the velocity of propagation of radiation as waves; and $\nu = c/\lambda$, where λ is the wavelength of radiation, which is the distance between either of two crests or two troughs of such a wave. Electromagnetic radiation is continuous; however, it is separated into different regions with differing energies.

The instruments and techniques needed to measure radiation are different in different regions of radiation. It should be noted that this artificial separation resulted from historical development of understanding radiation phenomenon. Table 1 lists the different regions of electromagnetic radiation.

The human eye responds to radiant energy between 150 and 310 kJ/mol (350–780 nm, wavelength). For each color observed, there is a complementary color; this complementary color corresponds to the wavelength of light that is absorbed by the material. Table 2 shows the approximate relationship between color characteristics and wavelength. The wavelength region of interest for pharmaceutical analysis is between 200 and 800 nm (UV and visible region).

Absorption of Light

When light of a particular wavelength passes through a dilute solution in an appropriate container (quartz, glass, or plastic cells), the light may interact with matter in a number of ways. A portion of the light may be absorbed and the rest passed through unaffected but redirected. The redirection phenomenon is known as scattering. The absorbed energy may be emitted with energy of lower wavelength. This process is known as emission. In UV–vis spectrophotometry, the difference between the incident and emerging beam of light is measured. This property, known as absorption, is used in quantitative and qualitative characterization of molecules of pharmaceutical interest. If the intensity of an incident beam of radiation is monitored and compared with that of an emerging beam, there is attenuation in the emergent beam. This attenuation is caused by 1) reflection at the air/cell surface interface; 2) absorption by the cell wall; 3) reflection at the cell/liquid interface; 4) absorption by solute; 5) absorption by solvent; 6) scatter by solution; and 7) refraction or dispersion by cell.

Therefore, to observe only the effect caused by the substance of interest, all other factors have to be eliminated. This is normally achieved by the use of two matched cells. By comparison of the energy of the incident

Table 1 Electromagnetic spectrum

Radiation (rays)	≈Frequency (Hz)	Wavelength (m)	Energy (kJ/mol)
Cosmic rays	$>10^{20}$	$<10^{-12}$	$>1.2 \times 10^8$
Gamma rays	$10^{19.5}-10^{20}$	$10^{-12}-10^{-11}$	$1.2-12 \times 10^7$
X-Rays	$10^{16.5}-10^{19.5}$	$10^{-11}-10^{-8}$	$12,000-1.2 \times 10^7$
UV	$10^{15}-10^{16.5}$	$10^{-8}-0.38 \times 10^{-7}$	310–12,000
Visible	$10^{14.6}-10^{15}$	$0.38 \times 10^{-7}-0.8 \times 10^{-7}$	150–310
Infrared	$10^{11}-10^{14.6}$	$0.8 \times 10^{-7}-10^{-3}$	0.12–150
Microwave	10^7-10^{11}	$10^{-3}-10^{-1}$	0.0012–0.12
Radio	$<10^7$	$<10^{-1}$	<0.0012

and emergent beams from these matched cells, the extent of light absorbed by the sample of interest (or analyte) can be expressed in terms of a quantity called “absorbance.”

Beer Lambert’s Law (or Beer’s Law)

Instruments used for UV–vis spectra measure the intensity of incident (I_0) and emerging light (I). It is related to absorbance as follows: $A = -\log(I/I_0) = -\log(T)$. T is called transmittance. The instruments can display data as either percent of transmittance or, more commonly, as absorbance A . According to Beer’s law, absorbance is proportional to the concentration of the absorbing species and the length of the path through which light travels: $A = abc$, where a is a constant called the absorption coefficient or absorptivity; a is dependent on the wavelength and the unit of concentration, c , and b is the path length in centimeters. When c is expressed as mol/L (M) $A = \epsilon bc$, where ϵ is called the molar absorptivity. Beer’s law is applicable under the following conditions:

- The incident radiant beam is monochromatic (monochromatic light is light of a single wavelength or better, light of a wavelength within a very narrow range, in terms of what is experimentally feasible).
- The incident radiation travels through equidistant parallel paths through the sample.
- There is no chemical reaction occurring followed by absorption of radiant energy.
- The sample solution is homogeneous and, thus, there is no loss of radiant energy through scattering or by other reflection processes.
- Each molecule absorbs independently and is not affected by other molecules in solution. (Note: This may not be true for high solute concentration).

The common terms, symbols, and definitions of terms, per International Union of Pure and Applied Chemistry (IUPAC) in absorption measurement are given in Table 3.

Applicability of Beer’s Law

The direct proportionality between absorbance and concentration for a fixed path length must be established experimentally for a given instrument under specified set of conditions. If this relationship is linear within the specified range of concentrations, the system is said to obey Beer’s law. Departure from this direct proportionality arises from both instrumental and chemical reasons.

In Beer’s law, c is the total analytical concentration of the absorbing species in a particular molecular state. If there is a disturbance in the concentration of this particular species as a function of solvent concentration, these deviations can be eliminated by appropriate choice of experimental conditions.

Sometimes, in organic acids and bases, the extent of hydrolysis or ionization can be affected as a function of pH. Then, the pH of the analyte should be chosen such that it is

Table 2 Visible spectrum characteristics

Wavelength (nm)	Color absorbed	Color observed
380–420	Violet	Green-yellow
420–440	Violet-blue	Yellow
440–470	Blue	Orange
470–500	Blue-green	Red
500–520	Green	Purple
520–550	Green-yellow	Violet
550–580	Yellow	Violet-blue
580–620	Orange	Blue
620–680	Red	Blue-green
680–780	Purple	Green

Table 3 Spectrophotometric nomenclature (IUPAC)

Name	Symbol	Definition
Transmittance	T	P/P_0
Absorbance	A	$\log_{10} P/P_0$
Path length	b	—
Absorptivity	a	A/bc^*
Molar absorptivity	ϵ	$A/bc^\#$
Wavelength unit	nm	10^{-9} m
Absorption max	λ_{\max}	Wavelength of maximum absorption

b is in cms; c^* units other than moles per liter; $\#$ moles per liter. Other terms used in the older literature are given in Raghavan and Joseph (1997) under Bibliography.

three units more or three units less than the pK_a of the monoprotic acid. Otherwise, the absorbance measurement can be made at a wavelength at which molar absorptivities of both species are identical. The wavelength at which the absorbencies are identical is known as isosbestic point. At this wavelength, the molar absorptivity of the two species is the same and, thus, measurements at this wavelength do not deviate from Beer's law. At millimolar concentrations at which absorbance measurements are made, the refractive index of the sample and that of the reference do not differ considerably. However, if the concentration is increased, as in derivative spectrophotometry, refractive index may be altered and, as a result, deviation from Beer's law may become significant. The second type of deviation from Beer's law occurs under derivative and dual wavelength spectrophotometric conditions. Deviation from Beer's law is negligible if:

- instrumental deviations are minimal;
- the solvent absorption is negligible in comparison with the analyte absorbance;
- the analyte concentration is within a specified range;
- there is no chemical interaction between solute molecules or between solute and solvent molecules; and
- the temperature is maintained constant, especially for flow injection analysis and reaction rate measurements.

Deviations from Beer's law caused by instrumental parameters are discussed later.

COMPONENTS OF THE SPECTROPHOTOMETER

List of Instrument Components and Functions

A spectrophotometer measures the change in light intensity when light passes through a liquid sample or

other medium. Therefore, appropriate light source(s) is needed. The light source gives out radiation consisting of a whole range of wavelengths and different intensities. Therefore, these different wavelengths should be separated. The light thus separated has to be passed through the analyte sample. To monitor the emerging beam, knowledge of incident beam is gathered from a reference solution containing all the matrix elements except that of the analyte. The emerging light needs detection, and an appropriate display system is needed for proper presentation of data. These different components and their functions are presented briefly in Table 4.

Light source

Two light sources, one for each region, are commonly used. For the visible region, a tungsten filament lamp sealed inside a quartz envelope containing a small amount of iodine is used. The source is operated at 3500 K. This type of lamp has excellent stability and a useful life of more than 10,000 h. For the entire UV region, a deuterium arc discharge lamp is used. This lamp consists of a quartz envelope filled with deuterium gas at a pressure of <0.01 atm (2–3 mm) and two electrodes separated by a small distance. When electrically discharged by passing a current through the electrodes, light is emitted as a result of ionization of gas and produces a continuous emission spectrum below 375 nm. This lamp has a useful life of more than 1000 h. A typical spectrophotometer contains both types of lamps, with a mechanical source selector that automatically switches between the two lamps as necessary. (It should be noted that mercury-vapor lamps are normally used as light sources for UV detectors of high-performance liquid chromatographs. Because a majority of organic compounds of pharmaceutical interest absorbs at 254 nm, and this lamp has an intense line spectrum at 253.7 nm, this lamp finds wide use in HPLC detectors).

Monochromators and filters

In forward optic spectrophotometers, electromagnetic radiation of a specified wavelength is passed through sample solutions. Because radiation from source is polychromatic (consists of light of all wavelengths), the various wavelengths of radiation must be separated within a narrow range to form a monochromatic beam before it is passed through the cell and directed on to the detector. Two types of wavelength selectors are used. These are 1) filters, which provide a limited wavelength selection with a high bandwidth, and 2) monochromators, which provide continuous variation of wavelengths with narrow bandwidths. Monochromators serve two basic functions, namely, they isolate the narrow bandwidth of wavelengths very close to the specific

Table 4 Components of a spectrophotometer and their functions

Component	Functions of the component
Power supply	Regulates and provides constant power needed for the light sources
Light source	Provides electromagnetic radiation over a range of wavelengths; the intensity may vary as a function of wavelength
Monochromator	Selects a narrow band of wavelength of choice
Entrance and exit slits	These slits are part of a monochromator to select a narrow beam of light
Sample port	This port accommodates a cell containing sample analyte and any other reference cells
Detector	The detector consists of a photosensitive device, which converts radiation energy into electrical energy yielding an output signal; this may consist of electronic circuitry and microprocessors and other components to amplify the initial signal received from the radiant energy
Digital read-out/computer	The digitized signals from the detector are stored in a computer for data-handling and for graphical display or print-out later

(selected) wavelength and automatically adjust exit slits such that the intensity of the light at the chosen wavelength is maximum, and light of other wavelengths is minimal or preferably, negligible. In addition, the monochromator lends itself to rapid scans of the entire spectral region of interest. The monochromator is a system consisting of combinations of lenses, a dispersion device (a prism or more commonly a grating), slits, and filters.

The entrance and exit slits of a monochromator isolate narrow portions of the light beam and direct them on to the grating. The slit width may be varied automatically by instruments to increase or decrease radiant power reaching the sample and detector. The grating is automatically driven by what is called a sine-bar mechanism, in which the bar that holds the grating is attached to a stepper motor-driven micrometer screw. Thus, the instrument covers a range of approximately 200–900 nm, with wavelength variation of ± 0.2 nm.

The grating is used to split polychromatic light into a series of monochromatic radiation of selected wavelengths. The diffraction grating used in UV–vis spectrophotometer is a very simple optical device that consists of parallel grooves. Most gratings in spectrophotometers are prepared holographically by coating a flat glass plate with a photosensitive material. An interference pattern is then generated by the use of laser beams. When the photosensitive layer is developed, glass plates are left with perfectly grooved patterns. To make the grating reflective, a thin film of aluminum is vacuum-deposited on its surface. A protective layer of silicon dioxide is deposited on top of this aluminum layer to keep it from tarnishing. Given the use of lasers, and holographic techniques, a precision level of <0.001 nm can be attained for these gratings.

When polychromatic radiation is incident upon the grating, it is rotated, resulting in a specific wavelength of light being allowed to pass through the slit and then onto the sample or detector depending on the nature of the spectrophotometer. Such rotation of the grating results in scanning a series of wavelengths each time with a specified wavelength. Thus, polychromatic light is split into monochromatic light. Gratings, therefore, provide stable, parallel, and highly monochromatic light, and stray radiation is greatly reduced in holographically generated grating monochromators. For UV–vis spectrophotometers, the gratings used normally have 1000–3000 lines per millimeter. Thus, a single grating, as a dispersion device, is adequate to cover the entire range between 200 and 900 nm.

Cells

Sample cells are available in a variety of shapes and sizes that can be used for many different applications. The most common is a rectangular-shaped cell, made of quartz, with dimensions of $12 \times 12 \times 45$ mm, with an inside length of 1 cm, and holds approximately 4 ml of solution. Cylindrical cells are normally 20 mm in diameter with a path length ranging from 5 to 200 mm. Because glass and plastic materials absorb in the UV region, quartz cells are required for measuring spectra below 340 nm, but cheaper glass cells or disposable plastic cells are adequate for the visible region. Cells with built-in water jackets are available for maintaining constant temperature for kinetic studies. Flow-through cells for flow injection and HPLC applications typically have cell volumes between 0.5 and 2 μ l, with a path length of approximately 1 cm. For limited samples, typically with a cross-sectional area of 2×2.5 mm, holding at least 60 μ l of sample is available. With special ultramicrocells, samples volume as low as 5 μ l can be used.

The cell windows should be maintained clean with no scratches and adhering liquids. The placement of the cuvettes is very important. Otherwise, cell walls can block the radiation beam, resulting in irreproducible erroneous results. Cells should always be stored in matched pairs, dry, and preferably inside a protective container. Catalogs provide figures and a list of the different types of cells available commercially.

Detectors

A detector is a device in which light intensity is transformed into an electrical signal. An ideal detector should have the following characteristics:

- high sensitivity irrespective of wavelength;
- very low drift in sensitivity, if any, over the entire wavelength range;
- high signal-to-noise ratio;
- no dark signal (or, if any, it should be very low);
- rapid response time;
- linear response to radiant power; and
- ease in calibration and standardization.

Even though no detector meets the entire criteria, many commercially available instruments, come very close to ideal with respect to performance characteristics.

Typically, light is allowed to fall on a metal coated with a light-sensitive substance. When light falls on this substance, electrons are released. By maintaining a potential difference of approximately 100 V, the released electrons from the cathode travel to the anode, producing a photocurrent. The resulting photocurrent thus generated is proportional to the intensity of light and is amplified, digitized, and measured. The measured current is plotted as a function of wavelength. However, in a majority of spectrophotometers, photomultiplier tubes are used to produce secondary amplification. The secondary amplification is produced such that for every primary electron generated, secondary electrons are attracted to a secondary electrode called a dynode. By using 8 to 10 dynodes, that are suitably oriented, connected in series, and having increasing positive potential, a million-fold amplification in photocurrent is attained. Very efficient silicon diodes are also used in many newer instruments.

Semiconductor silicon diodes

Semiconductor silicon diodes, called photodiodes, with special conductance properties are generally used in diode array spectrophotometers. Photodiodes are very useful because scan times of the order of a few milliseconds and thus many scans may be obtained in a very short time. In addition, the data obtained, being digital, can be further

processed to generate spectra with a very high degree of resolution and accuracy.

Charge coupled device detectors

These solid state detectors are multichannel devices. They have large number of photon-detecting registers. The registers are semiconductor capacitors that have been formed on a silicon chip. These detectors, which can detect very low levels of light, are primarily used in probe-type, near-infrared, and visible spectrophotometers and in fluorometers.

Read-out devices

The electrical signal from the detector is fed into an instrument with read-out capabilities. These read-out systems may be direct-reading or null point systems. In direct-reading systems, the output of the detector is allowed to drive a sensitive meter. In the null point systems, the detector output is balanced by a reference circuit. Because of the imbalance in the current, a servomotor is activated, and this stops when the two circuits are balanced. The extent of movement of the motor directly gives a digital readout. Most popular digital read-out device provide a visual numerical display using light-emitting diodes or tubes. These are directly controlled by the signal from the detector. These digital read-out devices have very rapid response. Current spectrophotometers and photometers are microprocessor-controlled instruments with digital read-out capabilities.

INSTRUMENTATION

The components of a spectrophotometer can be arranged in one of two ways as shown in Fig. 1.

The two arrangements differ only in the relative arrangement of the sample and the monochromator. However, the complexity, the quality, and the speed at which data are acquired and presented are very different. A brief account of these is given below.

Forward Optic Spectrophotometers

Forward optic spectrophotometers are either single-beam or double-beam spectrophotometers. The single-beam instruments can be either very simple or expensive depending on the sophistication desired or needed. Simple single-beam instruments have poor stability and excessive drift. These advantages are eliminated in systems equipped with a high-resolution monochromator with adjustable slits, controlled by microprocessors for rapid data acquisition and evaluation of data.

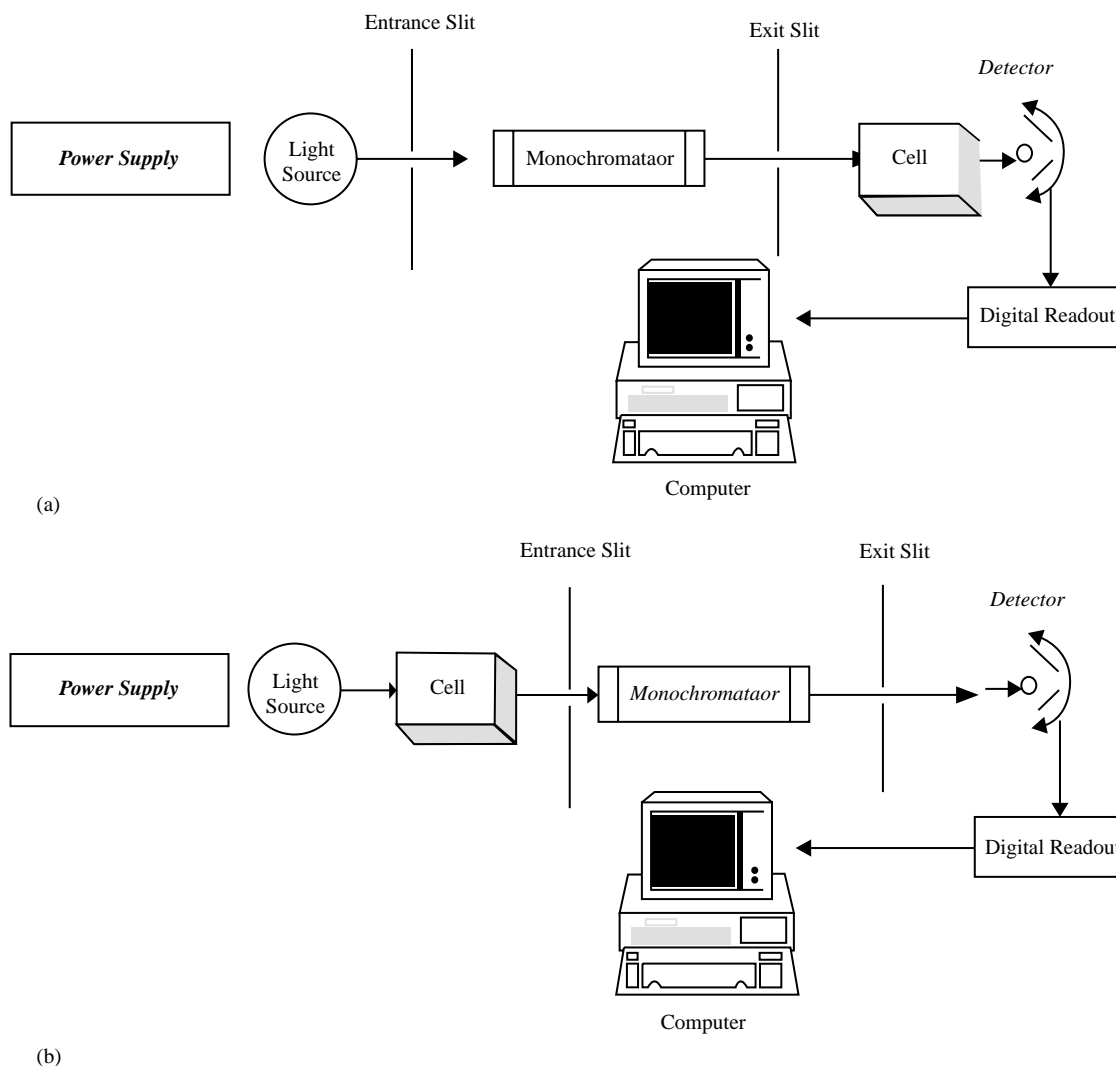


Fig. 1 (a) Components of a typical UV-vis spectrophotometer (forward optics). (b) Components of a reverse optics spectrophotometer.

There are two different types of double-beam systems. One such double-beam system, referred to as a double-beam-in-time, is shown in Fig. 2a and the other, known as double-beam-in-space, is shown in Fig. 2b.

In the double-beam-in-space system the components of a single beam are exactly duplicated except for the light source. In this system, the primary light beam is split into two beams with a mirror. One of the beams passes through a reference cell containing the solvent alone, whereas the second beam passes through the sample cell. The signal from the detector of the reference cell is subtracted from that of the sample cell, and the absorbance is plotted against the wavelength.

The dual-beam, single-deflector system is referred to as a double-beam-in-time spectrophotometer. In this

system, the primary beam is split into two beams by a mechanical chopper; the chopper is a rotating wheel with alternate silvered and cut sections. The rotating chopper, in addition to splitting the beam, introduces a dark beam into each beam so that the detector records a series of alternating dark-reference dark-sample signals. Using an appropriate chopping frequency and a synchronizing signal, the three signals are measured within 10–50 ms. As a result, this configuration, more easily being amenable to microprocessor control, is used in many rapid scan spectrophotometers.

The double-beam instruments compensate for

- fluctuations in the radiation output signal;
- drift in the detector and amplifier; and

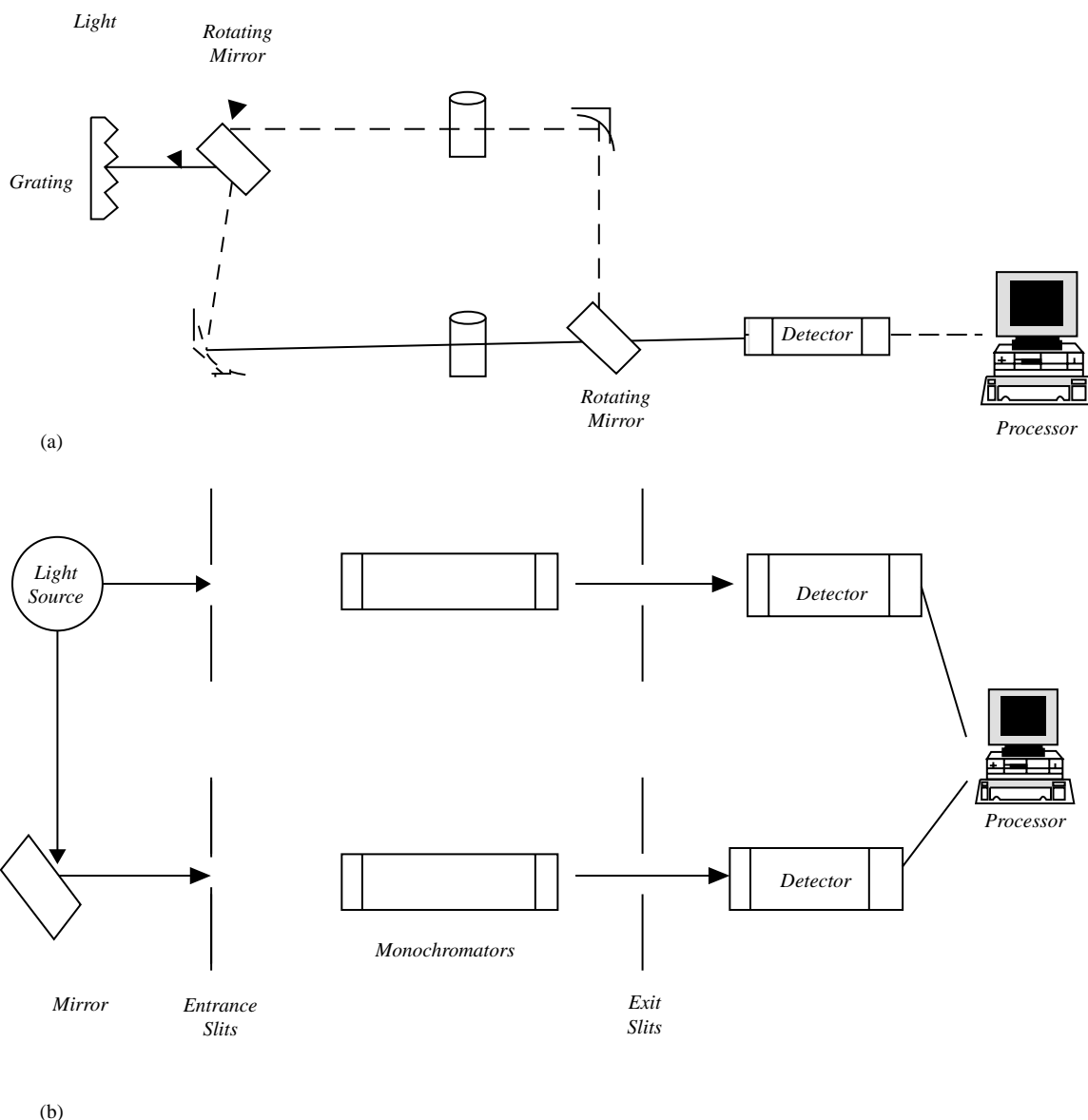


Fig. 2 (a) Block diagram of a double-beam-in-time spectrophotometer (single-detector system with a beam-splitter); a chopper motor can be used in place of the mirror. (b) Block diagram of a double-beam-in-space spectrophotometer.

- wide variations in the intensity of the light source as a function of wavelength.

Reverse Optic Spectrophotometers—Photodiode Array Detectors

The photodiode array detectors are based on reverse optics as shown in Fig. 1b. In this mode, a pair of vibrating mirrors passes the polychromatic light through either the reference cell or the sample cell. The emerging light

dispersed by a dispersion grating (in the monochromator) reaches the diode array detector consisting of a linear array of miniature diodes. The number of such photodiodes may vary between 256 and 4096 per array, with 2048 being the most common. Many such diodes can be formed along the length of a silicon chip. Typically each chip is 1- to 6-cm long, and the widths of the individual diode may vary from 15 to 50 μM . The monochromatic slit width is adjusted to correspond to the width of one of the silicon diodes. Therefore, the output signal of each diode corresponds to

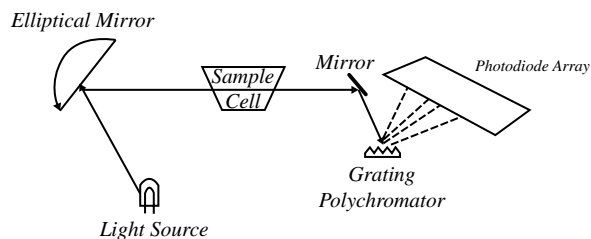


Fig. 3 Schematic of a photodiode-array detector.

the radiation of a different wavelength. By scanning the output of each diode sequentially, a spectrum can be obtained. Because the electronic processing is very rapid, the entire spectrum in a specified range can be acquired in less than 5 ms. The analog signals from the photodiodes are digitized and transferred to a computer. Many different types of multichannel detectors and other diode array detectors are commercially available. These rapid scanning detectors are used for

- analysis of multiple components in a mixture;
- following reactions pathway of short-lived intermediates in reaction kinetics;
- process control by continuous monitoring;
- measuring the dissolution rate of tablets; and
- obtaining spectra of a mobile solution (detection in HPLC).

These new instruments containing several microprocessors with appropriate hardware and software control the following:

- autosampling, which involves loading cell holders and measuring number of samples sequentially;
- automatic mixing, adding reagents, if necessary, diluting, and other sample preparations;
- setting up operating conditions such as selecting a lamp, filter, wavelength drive, and data acquisition mode;
- controlling autocalibration, automatic baseline noise storage, and baseline subtraction;
- multiple data handling characteristics such as spectrum smoothing, data graphical display, and data and peak expansion;
- addition and subtraction of spectra, creation of overlapped spectra for comparison and the like;
- control-associated hardware peripherals such as printers, plotters, etc.; and
- self-test procedures and record keeping of lamp use, maintenance records, etc. to comply with GLP regulations.

PERFORMANCE PARAMETERS OF A SPECTROPHOTOMETER

Specification and List of Parameters

The performance of a spectrophotometer is dependent on a number of parameters that affect the quality of the data, namely, wavelength accuracy, photometric (absorbance) accuracy, spectral bandwidth variations, extent of stray light, and linear response of the instrument. The performance parameters can be evaluated from the specifications provided by the manufacturer. A standard list of specifications of a typical double-beam spectrophotometer is provided in Table 5.

A brief outline of the specification terms, the values of which determine the performance parameters of the instrument, are presented below.

Wavelength Accuracy and Wavelength Calibration

The wavelength accuracy corresponds to the accuracy with which wavelengths measured in different instruments can be compared. In many spectrophotometric procedures, the analyte is quantitated by comparing the absorbance of the analyte with that of a reference standard. The standard is usually prepared at virtually the same concentration of the analyte and absorbance measured at the λ_{\max} of the chromogenic species. This approach is satisfactory for most filter photometers because comparisons are made at a fixed wavelength and with the same spectral bandwidth. However, when continuous choice of wavelength is available, as in many grating spectrophotometers, the accuracy and reproducibility of wavelength as well as the intensity of analyte absorbance have to be verified. This is accomplished by wavelength calibration. For wavelength calibration, holmium oxide glass, usually supplied by the instrument manufacturer, is used. Holmium oxide glass, when scanned between 280 and 650 nm, give very short principal peaks at 241.5, 279.3, 287.6, 333.8, 360.8, 418.5, 536.4, and 637.5 nm. Wavelength accuracy can also be checked using the emission lines from a low-pressure mercury source or deuterium lamp. The spectral bright lines used for this check are 486.0 and 656.1 nm.

For broader bandpass filters, neodymium glass filters are available for additional wavelength calibration in the visible region. Special NIST reference materials referred to as SRM (standard reference materials) numbers are available for calibration and verification from the National Institute of Standards and Technology. Information about these standards can be accessed

Table 5 Standard list of specifications for a typical spectrophotometer

Specification	Typical value for a double-beam instrument
Wavelength range (nm)	190–900 nm
Wavelength accuracy (nm)	± 0.3 nm at slit width of 0.2 nm
Wavelength repeatability (nm) ^a	± 0.1 nm
Wavelength scanning speed	Fast, medium, slow, and super slow
Bandwidth (nm)	0.1–7.5 nm (variable and selectable)
Resolution (nm)	± 0.1 nm
Stray light (%)	$<0.015\%$ at 220 and 340 nm $<0.0003\%$ at 220 and 340 nm (drift instrument)
Photometric accuracy (AU) ^b	± 0.002 in 0–0.5 AU ± 0.004 in 0.5–1.0 AU
Photometric repeatability ^c	± 0.001 in 0–0.5 AU (with NIST 930 D filter)
U (AU)	± 0.002 in 0.5–1.0 AU $\pm 0.3\%T$ (0–100% <i>T</i>)
Photometric mode	Absorbance (Abs), transmittance (%), reflectance (%), and energy (E)
Base-line correction	Selectable with storage firm ware
Drift (AU/h)	<0.0004 after warm up
Response time or time constant	≈ 0.1 –5 s
Photometric range (AU)	–4.0–5.0
Baseline flatness ^d (AU)	Within ± 0.001 Abs
Scan speed	50 nm/min to ≈ 2100 nm/min

^aWavelength repeatability is a measure of the precision of the wavelength measured.

^bAU = Absorbance units.

^cPhotometric repeatability refers to the precision of measured absorbance especially using a standard.

^dBaseline flatness refers to the average deviation from 0.0 AU for a blank versus blank spectrum in the wavelength range in which the spectrum is taken. *Note:* This may be dependent on scan speed.

directly from the Internet (<http://www.i-nist.gov/itl/div898/index.html>).

Wavelength Repeatability and Bandwidth

Spectral Bandwidth (SBW)

Wavelength repeatability is a measure of the precision of wavelength measured. The bandwidth refers to the width of an emission band (from the monochromator) at half peak height. This value, normally provided by the manufacturer is accepted. Using a mercury vapor lamp one can also check the spectral width. A number of well defined emission lines at 243.7, 364.9, 404.5, 435.8, 546.1, 576.9, and 579 nm can be used to check spectral bandwidth. However, the accuracy of the absorbance measured is dependent on the ratio of spectral bandwidth to the normal bandwidth (NBW) of the absorbing species. Most active pharmaceutical compounds have a normal bandwidth of approximately 20 nm or greater. With a SBW/NBW ratio of 0.1 or

less, absorbance can be measured with an accuracy of 99.5% or greater.

Resolution

Spectral resolution is a measure of the separation between two adjacent wavelengths. The two adjacent peaks are said to be resolved when the minimum absorbance between the two peaks is 80% or lower of the maximum absorbance. Most double-beam spectrophotometers have resolution below 0.5, typically 0.1. However, many diode array and rapid scan spectrophotometric detectors have resolution between 1 and 2 nm.

Photometric Accuracy

Photometric accuracy is determined by comparing measured absorbance to a certified absorbance of accepted standards. Solutions of potassium dichromate are recommended for evaluation of photometric

accuracy. [See Raghavan and Joseph (1997) under Bibliography] details about these standards. However, instrumental noise, drift, and stray light influence both photometric accuracy and precision. A discussion of these follows.

Noise

Noise in UV–vis spectrophotometry refers to uncertainties caused in the measurement of the absorbance signal. Essentially, there are two sources of noise. One is dependent on the source intensity (Schott noise) and the other independent of it. The effect of noise in spectrophotometric measurement can be significantly reduced if the concentration of the analyte is adjusted such that measured absorbance is between 0.3 and 1.2 absorbance units. A significant source of noise in double-beam instruments arises when sample and reference cells are not positioned properly. Minor imperfections in the cells cause reflections and scattering losses, which vary as a result of variations in the exposure of cell windows to the beam of light.

Drift

Photometric accuracy is also affected by drift. Drift is caused by changes in lamp intensity between measurements. This effect is nullified by subtracting the background absorbance of the analyte at a reference wavelength in a nonabsorbing region of the spectra. This process of “internal referencing” improves accuracy considerably. If the drift varies as a function of wavelength, two reference wavelengths, one before and one after the wavelength of interest, but in nonabsorbing regions, are chosen. This drift is assumed to be linear, and the drift calculated by linear fit at the wavelength of interest is subtracted.

Stray light and slitwidth

The spectral bandwidth is directly proportional to the slit width. When the slit width is narrowed, the bandwidth is less and thus the spectral purity increases. However, as the slit width is reduced, the signal-to-noise ratio also decreases proportionately, causing poorer precision. A slit width of 2 nm is adequate for most bands. A slit width of 0.5–1 nm is preferable for sharp bands. In HPLC detectors, to maximize the sensitivity, the slit width is automatically kept at 5 nm unless a rapid scan is carried out. During rapid scan, the slit width is normally changed to 1 nm. Stray light is any radiation identified by the detector outside of the narrow band of the wavelength selected by the monochromator. Scattering and diffraction inside the monochromator introduce stray light. Stray light is usually expressed as a ratio (or percent) of stray

light into the sample compartment or into the detector housing. In the specification, stray light is given to be 0.015% at 220 and 340 nm. Many currently available double-beam instruments have very low stray light values. Stray light values below 0.1% gives absorbance measurement errors below 0.5%, when the absorbancies measured are approximately 0.5 AU. Stray light becomes significant for quantitative measurement if the measurement is below 220 nm or above 330 nm because the source intensity is low in these regions. Manufacturers typically give stray light data at 220 and 340 nm. Stray light contributions increase with the age of the instrument. Therefore, the extent of stray light contribution needs to be assessed when necessary.

Factors that Influence Spectra

The position and intensity of the different bands in a spectrum depends on a number of factors. The effects of some of these may be considerable.

Solvent effects

The position and intensity of the different bands in a spectrum depend on the nature and conditions of the solvent used. In addition to electronic excitation, which causes absorption, vibrational and rotational energy level changes occur when light is absorbed. As a result, when spectrum is generated in the vapor phase vibration and rotation, fine structure is observable. However, in a solvent medium, because of thermal agitation and different types of interaction of the solute with the solvent, the position of λ_{max} may differ from solvent to solvent. In general, in polar solvents, interaction degrades the spectrum into broad bands because of molecular, electronic, and hydrogen-bond interactions. In aqueous solution, pH, concentration of analyte, ionic strength, and temperature can alter the position and intensity of maxima observed.

Deviation from Beer's law: Instrumental factor

The radiation entering a sample cell is a narrow band of polychromatic radiation. Thus, the effective molar absorptivity is an average of multiple wavelengths. As a result, the positive deviations from Beer's law may occur at low concentrations, whereas negative deviations may occur at higher wavelengths. Nonlinear behavior may also occur as a result of varying path length. When radiant beam is reflected at the solution/wall interface back into the solution, the beam traverses multiple paths. Thus, the net effective path length is increased and, therefore, the absorbance increases. This becomes pronounced (~0.3%) at low concentrations.

APPLICATION OF SPECTROPHOTOMETRIC MEASUREMENTS

In pharmaceutical analysis, spectrophotometric measurements are normally used for quantitative determination of known constituents. They are also used for qualitative analysis of compounds. These are addressed briefly. [See Gilpin and Pachla (1997, 1999) for extensive review of these methods.]

Qualitative Analysis

UV–vis spectrum of an analyte in solution is usually broad with very few narrow bands and shoulder. However, based on the nature of specific functional groups present in a molecule, some general spectral behavior was observed. Based on these studies and with the aid of quantum mechanical calculations, empirical relationships between structure and functional groups have been established. The terms commonly used in defining spectra are shown in Table 6.

Organic Functional Groups and Structure Elucidation

In a molecule, the absorption of ultraviolet or visible light represents the excitation of bonding electrons. Therefore, from the position of λ_{\max} and also from the intensity, the structure of functional groups present can be inferred. UV–vis spectroscopy can be used as a powerful tool to supplement identification of functional groups by other techniques. The electrons present in the ground state (lower energy state) of a molecule are excited to higher energy states by the absorption of light energy. This energy that is absorbed is measured. Some typical chromaphoric groups and their spectral characteristics are presented in Table 7.

The presence of an absorption band at a particular wavelength is a strong indication that a particular chromophore is present in the molecule. When additional functional groups are added, the characteristics change. For example, C–C double bond in butadiene absorb approximately at 214 nm. When a third double bond is conjugated, the absorption maximum is shifted to a longer wavelength by 30 nm. Also, alkyl substitution at a carbon atom leading to branching increases by another 5 nm. Empirical relationships, known as Woodward–Feiser rules, help predict the position of λ_{\max} in organic molecules. These empirical rules are applied to assign structures of organic molecules or distinguish between some isomeric compounds. Organic chemists use spectral data as an additional powerful tool to assign structures of organic compounds. [For a review of structure/spectra relationships, see Thomas (1997) in Bibliography.]

Spectra of Inorganic Ions

In the case of inorganic anions, generally excitation from a nonbonding lower energy state to an excited state (n to π^* transitions) is observed. For example, NO_3^- shows absorption at 313 nm. Chromate and dichromate ions have well-defined spectral absorption, which is used in calibration. However, from such data, conformation of identity is not possible. Inorganic cationic species, which are colored, usually exhibit very characteristic spectrum. These compounds with very well-defined absorption spectra are transition and inner transition metal cation coordination complexes. For example, a solution of copper is pale blue in aqueous slightly acidic medium, whereas in ammoniacal medium it is strong blue. These are attributed to splitting of “d” orbitals in the presence of ligands that complexes with the metal ion: one set of three d orbitals is called t_{2g} and the other set is known as e_g orbitals.

In aqueous solutions containing copper, the metal ion is octahedrally coordinated, i.e., the metal ion is at the center,

Table 6 Spectral characteristics of some organic functional groups

Species/functional groups	λ (nm)	ϵ ($\text{L mol}^{-1} \text{ cm}^{-1}$)
Satd. hydrocarbons	<200	—
Alkenes, alkynes, aromatic compounds, aromatic heterocyclic, highly conjugated aliphatics such as carotenes	200–500	$\sim 10^4$ – 10^5
Ethers, amines sulfides, alkyl halides	150–260	100–1000
Ketones, aldehydes, amides, esters, acids (—COOH), nitrates, <i>N</i> -heterocyclics such as pyridine	250–600	10–100

Table 7 Definition of spectral terms

Spectral term	Definitions and comments
Chromophore	A functional group, when present in a molecule, shows absorption in the region between 200 and 800 nm
Auxochrome	An additional functional group in the molecule, which causes shift to longer wavelengths
Hypsochromic (blue) shift	Shift of λ_{\max} to shorter wavelengths
Bathochromic (red shift)	Shift of λ_{\max} to longer wavelengths
Hyperchromic	Increase in intensity of absorption
Hypochromic	Decrease in intensity of absorption

and it is surrounded by six ligands, one at each apex of an octahedron. As a result of these coordinating ligands, the d orbital electrons, which are at the axis of the approaching ligand, experience forces of repulsion and attain higher energy state compared with those orbitals that are far removed from the axis of approach. The five equi-energetic d orbitals are split into two sets of equi-energy orbitals. The magnitude of the separation between the two energy states depends on the ligand field established by the ligands. In copper, of the nine d electrons, a set of six electrons occupy the lower energy state t_{2g} orbitals, and the other three occupy the excited state e_g orbitals. The e_g orbitals can accommodate four electrons. By absorbing visible light, an electron from the ground state can be promoted to the vacant excited state e_g orbital. This results in the characteristic absorption of many metals ions. EDTA, which is used in many pharmaceutical preparations, forms a hexa-coordinated complex, thus stabilizing and solubilizing metals. The example cited is for octahedral splitting. Similarly, in a tetrahedral field, the ligands can occupy the apex of a tetrahedron, and in a square planar complex, the corners of square plane. The tetrahedral and square planar complexes give rise to similar but different splitting of the d orbitals. Depending on which of these different ligand fields are formed different spectral patterns are observable for the same metal ion.

Charge Transfer Complexes

A third type of electronic transition that causes light absorption in the visible region is attributed to the formation of charge-transfer complexes. One example is the formation of the color blue by I_3^- when starch is used as an indicator in iodimetric titrations. A second example is the reaction between Fe^{3+} and thiocyanate ion SCN^- . In this case, an electron is transferred from a donor (SCN^-) to the acceptor (Fe^{3+}). The net effect is an internal redox reaction-forming Fe^{2+} and SCN radical. This type of internal redox reaction is generally reversible. However,

occasionally such charge transfer complexes may undergo irreversible redox reactions resulting in photochemical oxidation and reduction reactions. The formation of similar charge-transfer complexes in organic molecules results in the characteristic absorption in the UV region that can be used for qualitative analysis.

Qualitative Identification

The direct measurement of UV–vis spectra is rarely used for confirmation of identity of a component. However, the comparison of analyte spectrum obtained on a chromatographically separated component to that of a reference compound under identical chromatographic conditions is extensively used for identification and chromatographic purity of the eluting peak. Diode array/rapid scan spectrophotometers are very commonly used for this purpose. A perfect point-to-point comparison of spectra of two or more compounds is possible with the use of the proper computer algorithm. Under chromatographic conditions, the approach is to first establish that a particular drug or chemical substantially contains a single chromatographic component. Spectral data generated are then used for identity and for establishing the chromatographic purity of the analyte of interest. A second approach is to deduce identity by comparing the ratio of absorbance at peak maximum versus absorbance at peak minimum. *USP* 24 specifies such an identification procedure for cisplatin and methyldopate. For cisplatin, an absorbance ratio of 4.5 ($\lambda_{301}/\lambda_{246}$) is specified. However, in methyldopate hydrochloride, absorbance difference between a reference standard and the bulk drug at the same concentration should not differ by more than 3.0%.

Quantitative Analysis

Absorption spectroscopy finds extensive use in every field in which the quantitation of chemical component is

required. The characteristics of the spectrophotometric methods include the following:

- extensive applicability to a large number of systems;
- very high sensitivity (ppm to ppb levels can be determined);
- by the choice of wavelength, solvent, pH conditions, and added reagents, selectivity, which is normally poor, can be assured;
- excellent accuracy. Relative uncertainties of 0.5–3.0% can be easily obtained (under certain conditions, with appropriate precautions, errors can be minimized to 0.1%);
- very easy and convenient; and
- on-site data acquisition with certain type of probe spectrophotometers using hand held (palm) computers.

Application to Naturally Absorbing Species

Virtually all organic and many inorganic compounds absorb in the region between 210 and 700 nm. The conventional methods generally involve the following steps.

Selection of solvents

Any solvent exhibiting solubility of a few 10ths of a milligram to milligrams per milliliter is adequate for the purpose. Choice of the solvent is important such that it is transparent, i.e., background absorbance is extremely small. Organic solvents of spectrophotometric quality that is guaranteed to very small background absorbance are commercially available. Only these solvents are recommended for use. For aqueous solutions, the common variables such as pH, temperature, ionic strength of the medium, and the presence of interferants should be adequately controlled.

Selection of wavelength

Usually, from a spectral scan, the wavelength corresponding to the absorption maximum is selected. At λ_{\max} , sensitivity is maximum and, over a narrow region close to the maximum, the absorbance is constant. Therefore, minor wavelength differences do not alter the response, and also the absorbance at λ_{\max} is less sensitive to instrumental uncertainties. However, if adequate sensitivity exists, measurements can be made at λ_{\min} , if interfering absorbance is negligible at λ_{\min} .

Both reference and sample cells should be matched, properly cleaned, and rinsed before use. To eliminate lint and films from the outer surface of the cells, swipe cleaning with a lens paper soaked in methanol or

acetonitrile is recommended. When all these conditions are met and precautions are observed, quantitation is facilitated by comparison measurement. The reference standard concentration (C_s) is adjusted very close to the expected analyte concentration (C_u). Then standard (A_s) and sample (A_u) absorbencies are measured: $(A_u/A_s) \times C_s = C_u$.

This direct measurement can be used routinely for in-process assay of pharmaceutical formulations. The formulations during in-process conditions are not likely to contain potential degradants and thus, normally, the degradant interference is absent. Sometimes interference can be eliminated by solvent extraction of the impurities or of the analyte itself. Use of suitable nonabsorbing additives, which selectively reacts or masks the interference, can eliminate interference.

Method of standard additions

The method of standard additions may take one of several forms. If the matrix interference is constant and small, the sample is diluted to a known volume and the absorbance is measured. Then multiple volumes of a reference standard of the analyte at a known concentration are added and diluted to the same volume. The absorbance of each of the sample thus prepared is measured. From a plot of volume added versus absorbance and by extrapolation to zero volume of addition, the concentration of the analyte can be determined. Fig. 4 shows a plot of calculated absolute weight of a drug added to a fixed volume of solution as a function of measured absorbance. The x coordinate value of 4.08 corresponding to zero absorbance indicates the weight of drug in the sample.

Analysis of mixtures

The total absorbance of a solution is the sum of absorbance of the individual components in a mixture. For example, for a 1-cm cell, the total absorbances

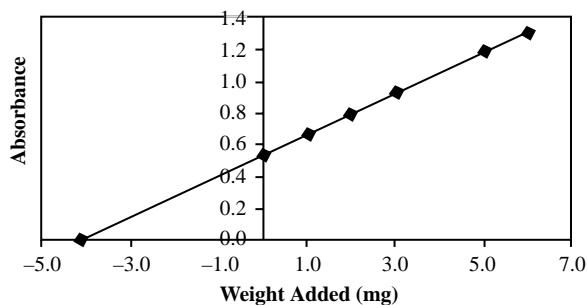


Fig. 4 Plot of weight added versus absorbance.

measured at each of two wavelengths are related to concentration as follows:

$$A_{\lambda 1} = A_1 + A_2 = \varepsilon_1 c_1 + \varepsilon_2 c_2$$

$$A_{\lambda 2} = A'_1 + A'_2 = \varepsilon'_1 c_1 + \varepsilon'_2 c_2$$

Then a known concentration of each analyte is prepared. The absorbance of these solutions are measured at each of the two wavelengths. Solving the two simultaneous equations, two equations for c_1 and c_2 are obtained. Then the known values of individual concentrations and calculated values of the four molar extinction coefficients are substituted in the derived equations to arrive at the unknown concentrations of the two components in the mixture. Multicomponent analysis is normally used in the dissolution testing of tablets. Standard hardware and software components for multicomponent dissolution testing based on compendial method are available from instrument manufacturers.

Derivative Spectroscopy

The UV–vis spectrum is a display of absorbance as a function of wavelength $A = f(\lambda)$. At the absorbance maximum, the derivative $dA/d\lambda = 0$; the second-order derivative $d^2A/d\lambda^2$ is a negative maximum. For all even-order derivatives, the spectra show well-defined peaks, and these can be used for analyte determination. The advantages of derivative spectroscopy are as follows.

- Positions of local maximum are precisely defined in the derivative plot compared with the diffuse absorption maximum. As a result, distinguishable derivative spectra are obtained for different compounds, even though normal spectra are likely to be very similar.
- Overlapping spectra, which could not be resolved by conventional techniques, are easily resolved.
- In quantitative analysis, selectivity and sensitivity are increased.
- Combined with algebraic calculations, derivative spectroscopy can be easily adopted for multicomponent analysis.

Difference Spectrophotometry

This technique uses information obtained as the difference of two absorbance measurements when two components having very close λ_{\max} values are present in a sample or when there is matrix interference in the sample, the sample at the same concentration is introduced into the reference and sample cells. Then, composition of one of the cells is

modified by changing pH, or by adding a suitable reactant or by changing the temperature of the solutions. The difference in the absorbance is measured. For example, at pH 3.0, a particular species absorbs, and its absorbance is measured. By changing the pH to 7.0, the absorbance changes at the same wavelength. However, the contribution to absorbance of the interference remains the same regardless of pH. Then the difference between the absorbances measured cancels the absorbance owing to the matrix. This difference ΔA is proportional to the concentration.

In turbid or cloudy solutions, in which absorbance is affected by scattering of the sample, two absorbance measurements are made: one at λ_{\max} and the other at a different wavelength where the analyte does not absorb. Then, the difference being proportional to concentration of the analyte, the analyte concentration can be calculated. Difference spectrophotometry is used in the determination of constituents in tablets, complex pharmaceutical preparations, plant extracts, syrups, biological matrices such as blood and serum, injectable oil preparations, and the like. For this technique, in addition to other general requirements such as well-matched cells that are positioned accurately, the solution should be made homogeneous, and the instrumental stray radiation at the wavelength of interest should be extremely small.

Application to Nonabsorbing Species

The limitations of naturally absorbing species for the determination of analyte concentration more often suffer from poor selectivity, sensitivity, and interference in mixtures of components. In such cases, the target analyte is reacted with a suitable reagent that results in a product, that usually absorb at longer wavelengths either in the UV or, more often, in the visible region. An appropriate reagent is selected to meet one or more of the following needs:

- to quantitate spectrophotometrically inactive compounds. These compounds usually have active functional groups such as hydroxy group, amino groups, and the like, which react to yield spectrally active products.
- to increase selectivity. When multiple components, each of which has absorbance in the wavelength maximum of the analyte are present, suitable chemical reaction selective to the analyte, is selected so the effect of the absorbing components is eliminated.
- to increase sensitivity. In many cases, the analyte may have natural absorption. However, that may not be adequate to determine at a required level (ppb or ppm

levels). In pharmaceutical manufacturing, it may be necessary to establish that the active analyte has been cleaned of the manufacturing systems below a certain level. In these cases, sensitivity enhancements are possible using an appropriate chemical reaction to yield a product with high molar absorptivity. For example, in cisplatin, although it has good absorbance sensitivity at approximately 210 nm, the sensitivity is good only for few tenths of milligrams per milliliter of cisplatin and not adequate for trace-level determination. However, by reacting cisplatin with diethyldithiocarbamate (DDTC), platinum DDTC complex with high molar extinction coefficient is obtained. Increasing the concentration using a column-switching technique and chromatographic separation, a detection limit of 0.2 ppb was attained. [See Raghavan and Mulligan (2000) and Raghavan et al. (2000).] It is not always necessary to add chemical reagents. An oxidation product may be generated on electrochemical oxidation of an analyte or by using pulsed photochemical activation, species thus generated may be more spectrally amenable for quantitation. However, if reagents are chosen for reaction, the following need to be considered for successful sensitivity and/or selectivity optimization:

- appropriate reagent concentration, solvent, and composition of the mixture; ionic strength, pH, and temperature of the reaction mixture;
- adequate reaction time for completion or for reproducible response that is proportional to the concentration of the analyte; and
- similar treatment of reagent blanks to allow for matrix changes that might affect the reaction.

Excellent reviews of application of spectrophotometric methods based on chemical reactions are available in the literature (see Bibliography).

Measurement of Color

Color comparison tubes

In pharmaceutical formulations, pale yellow color may be imparted by the excipients or as a result of formation of colored degradants, or feebly colored products may be generated intentionally as in limit tests for heavy metals. In such cases for visual testing glass color comparison tubes, Nessler tubes, of uniform cross-section and flat bottoms are used. These tubes should be matched as closely as possible in internal diameter and length. The tubes are viewed downward through the solutions against a white background. The color intensity of the unknown is

compared with two successive standards, one above and one below the level of the unknown. The color is reported to be less than the most intense color standard with which it is compared. To meet compendial color requirements, special instruments are commercially available.

Probe-Type Spectrophotometers

Many dip-type UV–vis spectrophotometers, use fiber optic sampling technology coupled with highly sensitive charge coupled device (CCD) array detectors. The probe is immersed directly in the sample being measured. Then radiation from a light source is transmitted through an optic fiber to the tip. The light travels through the 0.5 cm through the sample and reflects off the back surface of curved mirror over another 0.5 cm through the sample and then back to an input filter. The sample absorbs part of the light and reflects back a small portion of the light that carries the spectral information of the sample. This reflected light travels back through suitably oriented fibers back to CCD array detectors. The spectrum is displayed through the use of the appropriate device. Direct measurement of concentrated solutions is made possible by amplifying the weak reflected light. As a result, this lends itself to on-line monitoring. Many remote filter optic sampling options for measurements not only on liquids but also on powder and solid objects are available. These spectrophotometers are miniaturized so that they can be hand-held and transportable for on-site measurements. Probe-type spectrophotometers are also useful for spectrophotometric titrations.

Spectrophotometric Titrations

The absorbance of a selected species is monitored as a function of volume of titrant added. The resulting absorbance data are plotted against volume of titrant. These titration curves have different shapes, depending on the absorbance characteristic of the analyte, added reagent, or the product formed. However, in all cases, the slope changes gradually near the equivalence point. Extrapolation of the two linear sections before and after the titration equivalence point yields an intersection point. The titrant volume corresponding to this intersection point is the equivalence point of the titration. Of the analyte, reagent added, or product in the solution, any one or two of these can be spectrally active. Therefore, a total of six possible titrant shapes are possible. Because the titration represents a series of analytical measurements, analytical accuracy is very high.

A variation of these titrations is used to determine stoichiometry and formation constant(s) of complexes. In these methods, typically the metal concentration and total volume of solutions are kept constant, but the ligand-to-metal ion ratio is continuously varied. From a plot of absorbance versus mole fraction, the stoichiometry between the metal and ligand can be obtained. Numerous variations of this method are adopted to obtain stoichiometry and formation constants or binding constants in many biochemical determinations.

VALIDATION OF SPECTROPHOTOMETRIC METHODS

The UV-vis spectrophotometric methods require validation of the method for the analysis of pharmaceutical compounds. Once the method is developed, data regarding precision, accuracy, linear response behavior, limit of quantitation, limit of detection, selectivity, and ruggedness are generated. These terms are very well-defined in many compendial monographs and in ICH, USP, and FDA guidelines. These are not addressed further. However, it should be emphasized that without appropriate validation data and reasonable understanding of how the method results will be affected by minor day-to-day variation of experimental parameters, routine generation of acceptable data may be difficult.

BIBLIOGRAPHY

1. Banwell, C.N.; McCash, E.M.; *Fundamentals of Molecular Spectroscopy*, 4th Ed.; McGraw-Hill: New York, 1994.
2. Burtis, C.A.; Ashwood, R.E., Eds. *Tietz Textbook of Clinical Chemistry*, 3rd Ed.; W. B. Saunders Co.: Philadelphia, 1999.
3. Clark, B.J.; Frost, T.; Russell, M.A. *UV Spectroscopy: Techniques, Instrumentation, Data Handling*; Chapman & Hall: London UK, 1993.
4. George, W.O.; Willis, H.A. *Computer Methods in UV, Visible and IR Spectroscopy*; Royal Society of Chemistry: Cambridge, UK, 1990.
5. Gilpin, R.K.; Pachla, L.A. Pharmaceutical and Related Drugs. *Anal. Chem.* **1997**, *69* (12), 145R–163R.
6. Gilpin, R.K.; Pachla, L.A. Pharmaceutical and Related Drugs. *Anal. Chem.* **1999**, *71* (12), 217R–233R.
7. Görög, S. *Ultraviolet and Visible Spectrophotometry in Pharmaceutical Analysis*; CRC Press, Inc.: New York, 1995.
8. Görög, S.; Bihar, M.; Csizér, É.; Dravetz, F.; Gazdag, M.; Herényi, B. Estimation of Impurity Profiles of Drugs and Related Materials. Part 14: The Role of HPLC/Diode-Array UV Spectroscopy in the Identification of Minor Components (Impurities, Degradation Products, Metabolites) in Various Matrices. *J. Pharm. Biomed. Anal.* **1995**, *14*, 85–92.
9. Owen, T. *Fundamentals of Modern UV-Visible Spectroscopy, A Primer*, Hewlett-Packard Publication Number, 12-5965-5123E; Hewlett-Packard Company: Germany, 1996.
10. Perkampus, H.H., Ed. *UV-VIS Atlas of Organic Compounds*, 2nd Ed.; VCH Publishers: 1992.
11. Raghavan, R.; Joseph, J.C. *Ultraviolet and Visible Spectrophotometry in Pharmaceutical Analysis*; Swarbrick, J., Boylan, J., Eds.; Marcel Dekker, Inc.: New York, 1997; 15, 293–339.
12. Raghavan, R.; Mulligan, J.A. Low-Level (PPB) Determination of Cisplatin in Cleaning Validation (Rinse Water) Samples. I an Atomic Absorption Spectrophotometric Method. *Drug Dev. Ind. Pharm.* **2000**, *26* (4), 423–428.
13. Raghavan, R.; Burchett, M.; Loffredo, D.; Mulligan, J.A. Low-Level (PPB) Determination of Cisplatin in Cleaning Validation (Rinse Water) Samples. II. A High-Performance Liquid Chromatographic Method. *Drug Dev. Ind. Pharm.* **2000**, *26* (4), 429–440.
14. Silverstein, R.M.; Weber, F.X.; Silverstein, R.M. *Spectrometric Identification of Organic Compounds*; John Wiley & Sons: New York, 1997.
15. Skoog, D.A.; Holler, F.J.; Niemann, T.A. *Principles of Instrumental Analysis*, 5th Ed.; Saunders College Publishing: Chicago, 1998; 299–351.
16. Soares, O.D.D.; Costa, J.L.C. Spectrophotometers Intercomparison for Spectrocolorimetric Scale Harmonization. *Rev. Scientific Instruments* **1999**, *70* (12), 4471–4481.
17. Sykes, P.A.; Shiue, H-C.; Walker, J.R.; Bateman, R.C., Jr. Determination of Myoglobin Stability by Visible Spectroscopy. *J. Chem. Ed.* **1999**, *76* (9), 1283–1284.
18. Talsky, G., Ed. *Derivative Spectrophotometry*; VCH Publishers: New York, 1994.
19. Thomas, M. *Ultraviolet and Visible Spectroscopy. Analytical Chemistry by Openlearning*; Ando, D.J. Ed.; John Wiley & Sons: New York, 1997.
20. Thomas, M. *Ultraviolet and Visible Spectroscopy. Analytical Chemistry by Openlearning*; Ando, D.J. Ed.; John Wiley & Sons: New York, 1997.
21. *United States Pharmacopeia, 24th Ed., National Formulary*, 19th Rev. The United States Pharmacopeial Convention Inc.: Rockville MD, 1999, General Chapter 851.
22. Williams, R. Application of Fourier Transform Spectrometry in the Ultraviolet, Visible, and Near-IR. *Appl. Spectrosc. Rev.* **1989**, *25*, 63–79.

STARCHES AND STARCH DERIVATIVES

Ann W. Newman

SSCI, Inc., West Lafayette, Indiana, U.S.A.

Ronald L. Mueller

GlaxoSmithKline, King of Prussia, Pennsylvania, U.S.A.

Imre M. Vitez

Chris C. Kiesnowski

Bristol-Myers Squibb Pharmaceutical Research Institute, New Brunswick, New Jersey, U.S.A.

INTRODUCTION

Starch is one of the most commonly used excipients in the pharmaceutical industry due to its disintegration and binding properties. A number of sources of starch are commercially available, with corn starch being the most common. An overview of starches and starch derivatives and their use in the pharmaceutical industry is presented.

DESCRIPTION

Molecular Structure

Starch is a polymeric material with a molecular formula of $(C_6H_{10}O_5)_n$, where n ranges from 300 to 1000. Common starches contain two types of D-glucopyranose polymers called *amylose* and *amylopectin*. Amylose is a linear polymer of α -D-glucopyranosyl units linked $(1 \rightarrow 4)$ as shown in Fig. 1a. These molecules can be comprised of 100 to over 1000 glucose units. Amylopectin is a branched polymer of α -D-glucopyranosyl units containing $(1 \rightarrow 4)$ linear linkages and $(1 \rightarrow 6)$ linkages at the branch points, as shown in Fig. 1b. This polymer is three or more times larger than amylose. Most naturally occurring starches contain approximately 30% amylose, however, specific starches and their properties are determined by the size and amount of each type of polymer molecule present in the material. Attractive forces between the polymeric molecules form the starch granules. The linear portions tend to associate into micelles, which bind the molecules together to form a somewhat ordered structure. Models of this structure have been proposed (1), and it is known that the structure is rigid and insoluble in water.

Types of Starch

Starch can be derived from a number of natural sources, including those listed in Table 1 (2). It is found in various parts of the plants and several extraction methods are used to isolate the material. The most common type of starch used in the pharmaceutical industry is corn, although studies with other types of starch have been performed (3–6).

The most common preparation of starch is wet milling, although dry milling is also performed. A series of milling, separation, concentration, and washing steps result in a suspension of starch granules (7). After processing, cornstarch is a white powder with a pale yellow tint. Bleaching is required to achieve absolute whiteness.

A number of starch modifications are used in pharmaceutical applications. Pregelatinized or compressible starch has been chemically or mechanically processed to rupture all or part of the granules in water. It is then dried to yield an excipient material suitable for direct-compression formulations. Sterilizable maize starch contains magnesium oxide (not greater than 2.2%) and has been chemically or physically treated to prevent gelatinization on exposure to moisture or steam sterilization. Soluble starch results when potato or maize starch has been chemically treated to destroy the gelatinizing ability of starch.

General Properties

Starch is a fine white powder that is odorless and tasteless. It is composed of very small spherical or elliptical granules. The botanical origin of the starch material determines the granule shape and size, and these characteristics are summarized in Table 1. It is insoluble in alcohol, most solvents, and cold water. Alkaline solutions, however, will degrade starch and its polysaccharide components. Starch

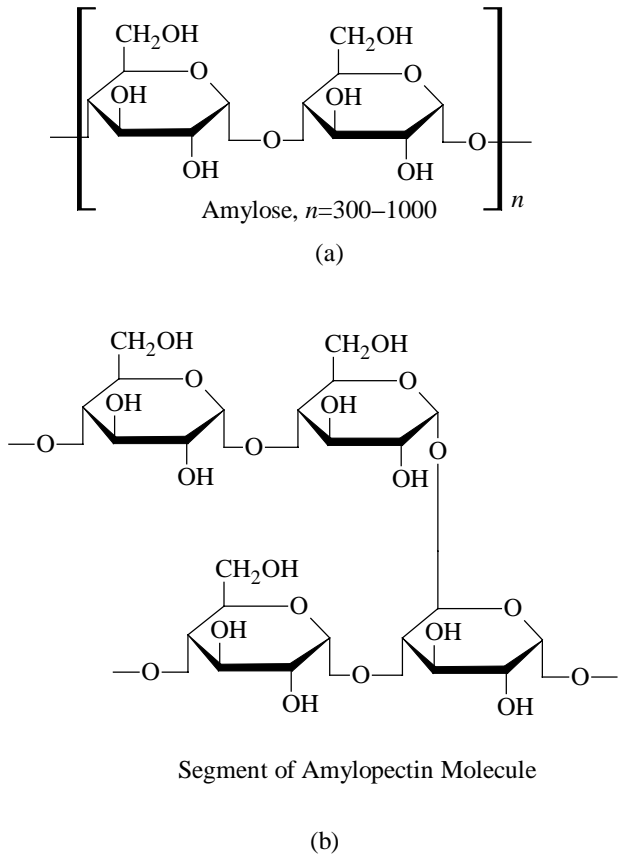


Fig. 1 (a) Linear amylose starch molecule. (b) Branched amylopectin starch molecule.

is relatively resistant to enzymes other than α-amylase and amylglucosidase.

When starch is suspended in water and heated to a critical point called the gelatinization temperature, water will penetrate the granules and swell them to produce a viscous mass. With the rising temperature, the hydrogen bonds that hold the micellar structural units and the water molecules in an aggregated state tend to dissociate. The dissociated water molecules are then able to penetrate the weakened starch structure and gradually hydrate the many hydroxyl groups along the length of the starch molecule. Gelatinization temperatures vary from starch to starch, but range from 60 to 75°C (2). Starch granules lose their characteristic shape as gelatinization proceeds.

The reaction of starch with iodine is a common identity test for starch. A dilute solution of iodine stains starches a blue to bluish red color. It is believed that the amylose portion complexes with iodine by forming a helix around it (7). This blue color has been used both as a qualitative and quantitative test for starch in various systems.

The National Formulary (8) contains numerous assays for starch including botanical characteristics, identification, microbial limits, pH, loss on drying (LOD), residue on ignition (ROI), iron, oxidizing substances, sulfur dioxide, and organic volatile impurities.

PHYSICAL PROPERTIES

Structural Information

Starch is a semicrystalline polymer. The linear amylose molecules are amorphous in nature, but the branched

Table 1 Sources and characteristics of various starches

Type of Starch	Extracted from	Granule shape	Granule size (μm)
Corn (Maize)	Seed	Round or polygonal	5–25
Tapioca	Root	Round or oval	2–25
Potato	Root	Egg-shaped	15–100
Wheat	Seed	Round or elliptical	2–10 or 20–35
Sago	Stem	Oval or egg-shaped	20–60
Arrowroot	Root	Oval	15–70
Rice	Seed	Polygonal	3–8
Barley	Seed	Round or elliptical	2–6 or 20–35
Waxy sorghum	Seed	Round or polygonal	6–30
Sweet potato	Root	Polygonal	10–25
Waxy maize	Seed	Round or polygonal	5–25

(From Ref. 2.)

Table 2 Crystallographic data for A and B type starch

Property	A type ^a	B type ^b
Lattice	Monoclinic	Hexagonal
<i>a</i> (nm)	2.124	1.85
<i>b</i> (nm)	1.172	1.85
<i>c</i> (nm)	1.069	1.04
α (°C)	90	90
β (°C)	90	90
γ (°C)	123.5	120
Space group	B2	—
Density	1.48	—
Volume	2.218 nm ³	—

^a(From Ref. 11.)^b(From Ref. 12.)

amylopectin portion has been reported as partially crystalline. It is believed that the crystalline regions in the starch granule are interspersed in a continuous amorphous phase (1, 9, 10).

X-ray diffraction studies have shown that starch exists in three crystal forms designated as A, B, and C (10). These forms are dependent on the botanical source of the starch. Pattern A is observed for cereal grain starches, whereas pattern B is characteristic of tuber, fruit, and stem starches. Pattern C is intermediate between the A and B patterns and has been attributed to mixtures of A and B type crystallites. The A type pattern is commonly observed for cornstarch.

Single crystal X-ray diffraction data for the crystalline portion of A type starch have been reported and the crystallographic data are summarized in Table 2 (11). The unit cell contains 12 glucose residues located in two left-handed, parallel-stranded double helices packed in a parallel fashion. Four water molecules are located between these helices. It has been reported that the B type starch also contains chains arranged in double helices (12). The currently accepted hexagonal unit cell dimensions are in Table 2. The A and B structures differ in crystal packing of the chains and in moisture content.

Diffuse reflectance (DR) infrared (IR), Raman, and solid-state ¹³C cross polarization/magic angle spinning (CP/MAS) nuclear magnetic resonance (NMR) spectra were acquired for unmodified and pregelatinized cornstarch (13, 14). Differences were not evident in the spectra for the various starch samples. Solution-phase, ¹H-NMR studies have also been performed on amylose and model compounds (15). Spectral assignments and intramolecular hydrogen bonding suggest that the same conformation is perpetuated along the amylose chain.

Thermal Properties

Thermal analysis has also been used to characterize the structure of starch. A melting endotherm due to the crystalline portions of starch has been investigated (16), but it is not clearly resolved in all samples due to the small amount of crystalline material present in the samples. This transition is also dependent on the sample preparation and moisture content of the material. Melting points of 168–210° C have been reported (16, 17).

The glass transition temperature (*T_g*) in starch, measured using differential scanning calorimetry (DSC), was found to be dependent on the moisture content of the preparation and covered a temperature range of 22–130°C (18). Gelatinization studies of starch using DSC showed that concentrated starch/water suspensions produced a well-defined endotherm under select conditions. The endotherm was integrated to obtain the heat of gelatinization for various starches (19).

Microscopy

A number of excellent reviews on the microscopy of starches have been published (2, 20–23). As summarized in Table 1, the granule shape and size is characteristic of the botanical origin and can be used to identify the materials. It has been reported that the floury granules, as found for potato and tapioca starches, tend to be larger and more regular in shape (round, elliptical, or oval). Horny starches, such as corn and rice, are usually

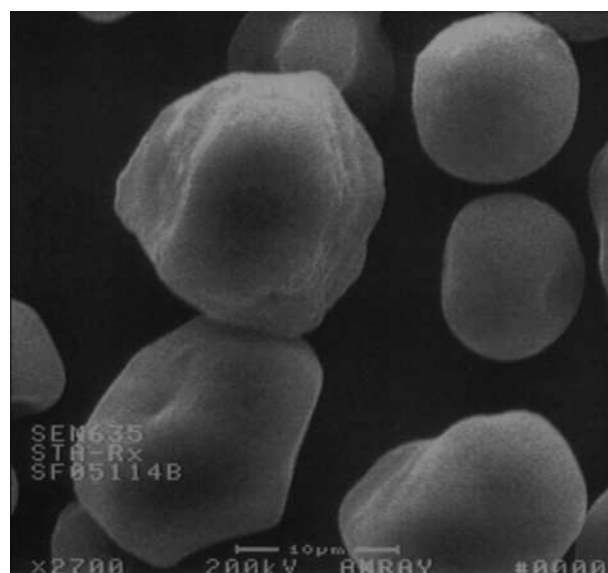


Fig. 2 Scanning electron micrograph of unmodified cornstarch granules (magnification 2700×).

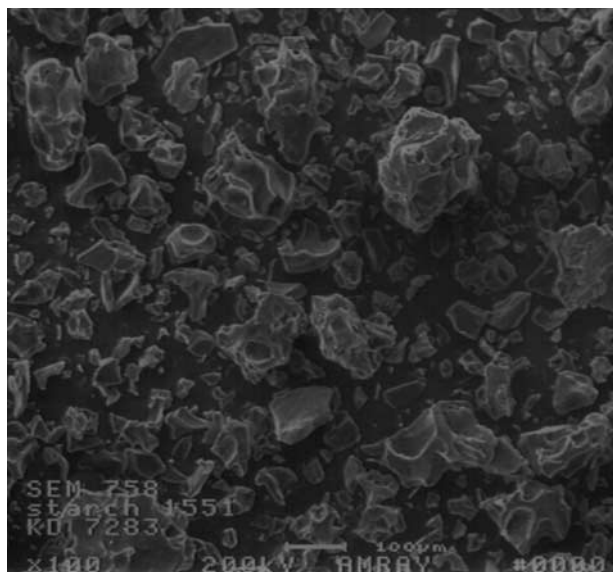


Fig. 3 Scanning electron micrograph of pregelatinized starch (magnification 100 \times).

described as polygonal because of the angular sides of the granules caused by the close packing of the granules in the kernel. An example is given in Fig. 2 for unmodified cornstarch. Starches are found as individual granules, but aggregated materials are also observed and are attributed to the drying conditions. Extensive heat and moisture during drying will produce a slight gelatinization of the surface of the granule and cause the granules to adhere together to form the aggregates. Pregelatinized starch exhibits an entirely different morphology, as shown in Fig. 3. The particles are irregular and the starch granules

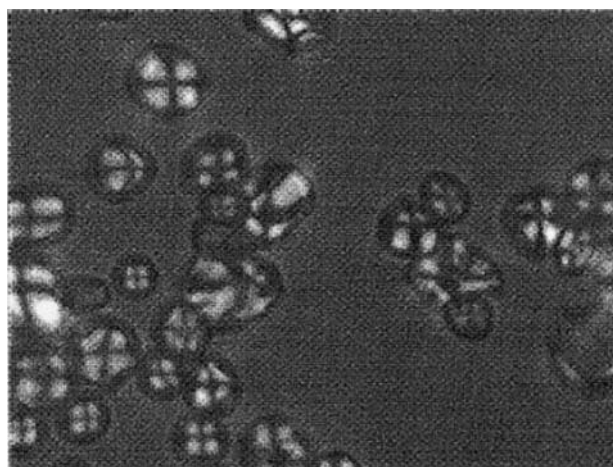


Fig. 4 Optical micrograph of unmodified cornstarch under crossed polarizers (magnification 400 \times).

are not evident as expected from the pregelatinization process.

When unmodified starch granules are observed using polarized light, two dark lines will form a cross or a V-shape (Fig. 4). The shape of the cross can be used to help identify the type of starch. One explanation for this feature suggests that the density and distribution of moisture throughout the granule are not uniform (2). As the granules dry, stresses are formed within the granule resulting in the bright regions observed under the polarized light. When the starch swells or is gelatinized, the cross is no longer visible with the polarizing microscope.

Micromeritic Properties

Micromeritic properties, such as particle size, surface area, density, and flow properties can affect the disintegration, handling, and tableting properties of starch materials. A number of methods for determining the particle size of various starches have been used. For bulk powder analysis, sieving is employed for large amounts of material. Another common method for particle size determination is optical microscopy because it gives a direct measurement of the individual particles (14, 24, 25). Laser light scattering analysis has also been utilized to measure the size of dry particles and suspensions (26). This analysis was found to be dependent on the model used to fit the data, and better reproducibility was obtained with samples suspended in liquid. Surface area measurements of starches have been obtained by air permeametry (27) or nitrogen adsorption (14). Relatively low surface areas ranging from approximately 0.1–0.5 m²/g have been reported.

Bulk powder characteristics are important in understanding the handling properties of an excipient or a granulated product. A classification system to evaluate the flow properties of powders has been introduced by Carr (28, 29). A flowable powder is defined as free-flowing and tends to flow steadily and consistently, whereas a floodable powder exhibits an unstable, discontinuous, and gushing type of flow. A number of studies have investigated the bulk powder properties of starch (14, 30) and granulations made with starch (31, 32). The starch materials were found to exhibit poor to borderline flow properties.

Hygroscopicity

Starch has been classified as a moderately hygroscopic material (33). Water sorption studies have been conducted using static methods (saturated salt solutions in closed chambers) (33–35) modified inverse frontal gas chromatography (36), and automated moisture balance systems

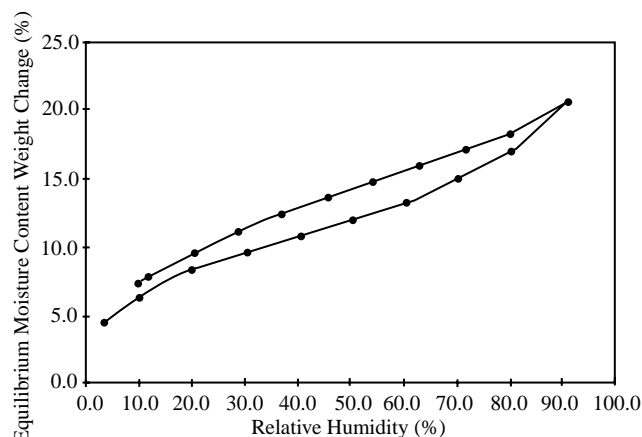


Fig. 5 Moisture isotherm for unmodified cornstarch.

(14). The isotherms are typically type II curves exhibiting hysteresis, as shown in Fig. 5. Hysteresis in starch samples has been attributed to intra- and intermolecular hydrogen bonding of water with the hydroxyl groups of the starch molecule (33, 37). The extent of hydration and swelling depends on the accessibility of the hydroxyl groups in the starch to the water, and it has been suggested that the amorphous regions are responsible for the reversible swelling of starch upon the adsorption of water (9, 37).

USES AND APPLICATIONS

Starch is widely used in the pharmaceutical industry because, among its other properties, it is readily available, inexpensive, white, and inert. Excipient compatibility studies of starch and various active drugs have been performed using thermal methods of analysis. As an example, starch has been found to be compatible with cephalexin (38) and acetylcysteine (39) using this method of excipient screening.

Starch is described as a tablet/capsule diluent, tablet disintegrant, and glidant. The function of starch can depend on how it is incorporated into the formulation. Starch will function as a disintegrant when it is added in the dry state prior to adding a lubricant. It may exhibit both binding and disintegrant properties when it is incorporated either as a paste or dry before granulation with other agents. It has been reported that starches deform mostly by plastic flow during compression, but this was found to be dependent on the particle size, size distribution, and particle shape (40).

Starch is used in many formulations for its disintegration properties. The most common explanation for the disintegration properties is the swelling of the starch

granules when exposed to water, and it has been proposed that amylose is the component responsible for the disintegration properties of starch due to swelling. A second mechanism suggests that the disintegrating action of starch in tablets is due to capillary action rather than swelling. A third proposed mechanism is based on the particle-particle repulsion forces between the tablet constituents when in contact with water and the hydrophilic nature of starch (41).

A number of new applications have been reported for the use of starch in formulations. A new linear, short-chain, high surface area starch product was prepared by gelatinization followed by enzymatic degradation for use in directly compressible controlled-release systems (6). The combination of ibuprofen-starch granulates and hydroxypropyl cellulose was used to produce controlled-release formulations of ibuprofen (42). Gelatinization and freeze-drying has been used to produce cold water-swelling starch as well as a matrix forming excipient in sustained release tablets (43). Grafted starch microcapsules are being investigated for the oral administration of vaccines to prevent degradation (44).

STARCH DERIVATIVES

Unmodified starch has proven to be an effective and inexpensive disintegrating agent, however, relatively high concentrations have been required and the flow properties have been poor. The major shortcoming of starch is that while it is effective at disintegrating tablets, it is less effective at disintegrating the granules from which the tablets are made. When other properties, such as rapid disintegration, were necessary, new materials called starch derivatives were investigated.

Starches undergo many reactions characteristic of alcohols because of the many hydroxyl groups present in the structure. Modification of the D-glucopyranosol units can occur by oxidation, esterification, etherification, or hydrolysis. The resulting starch derivatives are defined by a number of factors such as plant source, prior treatment (acid-catalyzed hydrolysis or dextrinization), amylose/amylopectin ratio or content, molecular weight distribution or degree of polymerization, type of derivative (ester, ether, oxidized), nature of the substituent group, and physical form (granular, pregelatinized) (9, 45).

The degree of substitution (DS) is a common method of characterizing starch derivatives and is a measure of the average number of hydroxyl groups on each D-glucopyranosyl unit. It is expressed as the moles of substituent per D-glucopyranosyl units, and the maximum DS is 3 since

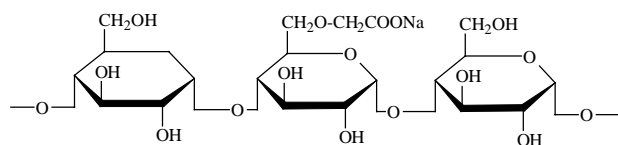


Fig. 6 Sodium starch glycolate molecule.

three hydroxyl groups are available in the unit for substitution. Most commercially produced derivatives have a DS less than 0.2. Molar substitution is used when the substituent group reacts further with the reagent to form a polymeric substituent. It is defined as the level of substitution in terms of mole of monomeric units (in the polymeric substituent) per mole of D-glucopyranosyl unit and can be larger than 3.

Sodium Starch Glycolate

Sodium starch glycolate is the sodium salt of a poly- α -glucopyranose in which some of the hydroxyl groups are in the form of the carboxymethyl ether (Fig. 6). Starch is carboxymethylated by reacting it with sodium chloroacetate in an alkaline medium followed by neutralization with acid. Cross-linking is achieved by physical or chemical methods. The sodium starch glycolate commonly prepared from potato starch is marketed as *Explotab* and *Primojel*. It is used as a tablet and capsule disintegrant and exhibits superior disintegration properties when compared to various starches (27). Sodium starch glycolates prepared from numerous starch sources (potato, corn, wheat, rice, sago, tapioca, and enset) have exhibited variations in disintegration efficiency, which have been related to the DS, cross-linking, and purity of the materials (46, 47).

Commercial sodium starch glycolate is a white to off-white, odorless, tasteless powder. It is an amorphous material with a T_g ranging from 128 to 156°C depending on the moisture content of the sample (33). It is considered to be a hygroscopic material based on the moisture uptake observed above 50% RH. It should be stored in a closed container to prevent exposure to wide variations in humidity and temperature that may cause caking. The material is composed of oval or spherical granules in the range of 30–100 μm . Flow properties measured for this material show that it is a free flowing powder that produced formulations exhibiting excellent mechanical properties (30).

Maltodextrins

Maltodextrins are carbohydrate materials prepared by controlled acid or enzyme hydrolysis (depolymerization)

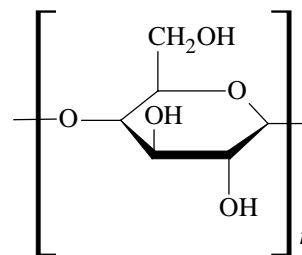


Fig. 7 Maltodextrin molecule.

of corn starch. The D-glucose units are linked primarily by α -(1-4) bonds but have branched segments linked by α -(1-6) bonds, as shown in Fig. 7. The physical properties of these materials are determined by the degree of starch hydrolysis, which is expressed as the dextrose equivalent (D.E.) value. The D.E. is defined as the amount of reducing sugars present, which is reported as grams of D-glucose per 100 g of dry substance. A high D.E. value represents a low degree of polymerization of the maltodextrins.

The maltodextrin solids are amorphous, white, or off-white powders or granules, which are nonsweet and odorless. The materials are processed by spray drying, fluidized bed agglomeration, and roller compaction to improve their handling properties. Maltodextrins are hygroscopic above 50% RH and need to be stored below this RH in tightly closed containers. The powders will begin to gel when the materials are stored above 75% RH. The hygroscopicity of maltodextrins will increase with an increase in the D.E. value.

Maltodextrins are used as a coating agent, tablet and capsule diluent, tablet binder, and viscosity-increasing agent. They have been evaluated as a direct compression tablet excipient and were found to exhibit similar compression properties to those of other direct compression excipients (48). Studies relating the degree of polymerization and compression properties show that tablet properties such as density and tensile strength could be correlated to the degree of polymerization and the moisture content of the maltodextrins (49). Matrix pellet formulations using microcrystalline wax, maltodextrin, and various binders were used for sustained release and immediate release formulations (50).

SUMMARY

Starches and starch derivatives are important in the formulation of pharmaceutical drug substances. Various starch sources, starch modifications, and starch derivatives provide a wide range of solids, which can be used in

pharmaceutical applications. The various properties exhibited by the materials can be exploited for specific uses during the formulation process.

REFERENCES

- Lineback, D.R. Current Concepts of Starch Structure and Its Impact on Properties. *J. Jpn. Soc. Starch Sci.* **1986**, *33* (1), 80–88.
- Kerr, R.W., Ed. *Chemistry and Industry of Starch*; Academic Press: New York, 1950.
- Gadalla, M.A.F.; Abd El-Hameed, M.H.; Ismail, A.A. A Comparative Evaluation of Some Starches as Disintegrants for Double Compressed Tablets. *Drug. Dev. Ind. Pharm.* **1989**, *15* (3), 427–446.
- Mitrevej, A.; Sinchaipanid, N.; Faroongsarng, D. Spray-Dried Rice Starch: Comparative Evaluation of Direct Compression Fillers. *Drug Dev. Ind. Pharm.* **1996**, *22*(7), 587–594.
- Gebre-Mariam, T.; Schmidt, P.C. Characterization of Enset Starch and Its Use as a Binder and Disintegrant for Tablets. *Pharmazie* **1996**, *51*(5), 303–311.
- Te Wierik, G.H.P.; Eissens, A.C.; Bergsma, J.; Arends-Scholte, A.W.; Bolhuis, G.K. A New Generation Starch Product as Excipient in Pharmaceutical Tablets. III. Parameters Affecting Controlled Drug Release from Tablets Based on High Surface Area Retrograded Pregelatinized Potato Starch. *Int. J. Pharm.* **1997**, *157*, 181–187.
- Whistler, R.L.; Daniel, J.R. *Starch. Kirk-Othmer Encyclopedia of Chemical Technology*; Grayson, M., Ed.; 3rd Ed.; John Wiley & Sons: New York, 1983; 21, 492–507.
- The National Formulary*. NF 19 United States Pharmacopoeial Convention, Inc. Rockville MD, 2000; 2524–2525.
- French, A.D.; *Starch: Chemistry and Technology*; Whistler, R.L., BeMiller, J.N., Paschall, E.F., Eds.; 2nd Ed.; Academic Press: New York, 1984; 183–247.
- Banks, W.; Greenwood, C.T. *Starch and Its Components*; John Wiley & Sons: New York, 1975.
- Imberty, A.; Chanzy, H.; Perez, S. The Double-Helical Nature of the Crystalline Part of A-Starch. *J. Mol. Biol.* **1988**, *201*, 365–378.
- Imberty, A.; Buleon, A.; Tran, V.; Perez, S. Recent Advances in Knowledge of Starch Structure. *Stærke* **1991**, *43* (10), 375–384.
- Bugay, D.E.; Findlay, W.P. *Pharmaceutical Excipients*; Marcel Dekker, Inc.: New York, 1999.
- Newman, A.W.; Mueller, R.L.; Vitez, I.M.; Kiesnowski, C.C.; Bugay, D.E.; Findlay, W.P.; Rodriguez, C. *Starch. Analytical Profiles of Drug Substances and Excipients*; Brittain, H.G., Ed.; Academic Press: New York, 1996; 24, 523–577.
- St-Jacques, M.; Sundararajan, P.R.; Taylor, K.J.; Marchessault, R.H. Nuclear Magnetic Resonance and Conformational Studies on Amylose and Model Compounds in Dimethyl Sulfoxide Solution. *J. Am. Chem. Soc.* **1976**, *98* (15), 4386–4391.
- Donovan, J.W. Phase Transitions of the Starch–Water System. *Biopolymers* **1979**, *18*, 263–265.
- Lelievre, J. Starch Gelatinization. *J. Appl. Polym. Sci.* **1973**, *18*, 293–296.
- Zeleznek, K.J.; Hoseny, R.C. The Glass Transition in Starch. *Cereal Chem.* **1987**, *64* (2), 121–124.
- Stevens, D.J.; Elton, G.A.H. Thermal Properties of the Starch/Water System Part 1. Measurement of the Heat of Gelatinisation by Differential Scanning Calorimetry. *Stærke* **1971**, *23* (1), 8–11.
- Sjostrom, O.A. Microscopy of Starches and Their Modifications. *Ind. Eng. Chem.* **1936**, *28* (1), 63–74.
- Moss, G.E. The Microscopy of Starch. *Examination and Analysis of Starch and Starch Products*; Radley, J.A., Ed.; Applied Science Publishers, Ltd.: London, 1976; 1–32.
- Schoch, T.J.; Maywald, E.C. Industrial Microscopy of Starches. *Starch: Chemistry and Technology*; Whistler, R.L., BeMiller, J.N., Paschall, E.F., Eds.; 2nd Ed.; Academic Press: New York, 1984; 637–647.
- Gallant, D.J. Electron Microscopy of Starch and Starch Products. *Examination and Analysis of Starch and Starch Products*; Radley, J.A., Ed.; Applied Science Publishers, Ltd.: London, 1976; 33–59.
- Paronen, P.; Juslin, M. Compressional Characteristics of Four Starches. *J. Pharm. Pharmacol.* **1983**, *35*, 627–635.
- Juslin, M.; Kahela, P.; Paronen, P.; Turakka, L. Comparative Evaluation of Starches as Tablet Adjuvants. *Acta. Pharm. Fenn.* **1981**, *90*, 83–93.
- Merkku, P.; Yliruusi, J.; Kristofferson, E. Particle Size Determination of Some Pharmaceutical Fillers by Laser Light Diffraction Part II. *Acta Pharm. Nord.* **1992**, *4* (4), 265–270.
- Smallenbroek, A.J.; Bolhuis, G.K.; Lerk, C.F. The Effect of Particle Size of Disintegrants on the Disintegration of Tablets. *Pharm. Weekblad.* **1981**, *116*, 1048–1051.
- Carr, R.L. Evaluating Flow Properties of Solids. *Chem. Eng.* **1965**, *72*, 163–168.
- Carr, R.L. Classifying Flow Properties of Solids. *Chem. Eng.* **1965**, *73*, 69–72.
- Vennat, B.; Gross, S.; Pourrat, A.; Pourrat, H. Tablets of Hamamelis Dry Extract by Direct Compression: Comparative Study of Natural Starches and Starch Derivatives. *Drug Dev. Ind. Pharm.* **1993**, *19* (11), 1357–1368.
- Kottke, M.K.; Chueh, H.-R.; Rhodes, C.T. Comparison of Disintegrant and Binder Activity of Three Corn Starch Products. *Drug Dev. Ind. Pharm.* **1992**, *18* (20), 2207–2223.
- Nasipuri, R.N.; Kuforiji, F.O. Effect of Granule Size of Starch as a Direct Compression Carrier on the Physical Properties of Chlorpheniramine Tablets. *Pharm. Ind.* **1981**, *43* (10), 1037–1041.
- Faroongsarng, D.; Peck, G.E. The Swelling and Water Uptake of Tablets III: Moisture Sorption Behavior of Tablet Disintegrants. *Drug Dev. Ind. Pharm.* **1994**, *20* (5), 779–798.
- Sair, L.; Fetzer, W.R. Water Sorption by Starches. *Ind. Eng. Chem.* **1944**, *36*, 205–208.
- Malamataris, S.; Goidas, P.; Dimiriou, A. Moisture Sorption and Tensile Strength of Some Tableted Direct Compression Excipients. *Int. J. Pharm.* **1991**, *68*, 51–60.
- Paik, S.W.; Gilbert, S.G. Water Sorption Isotherms on Sucrose and Starch by Modified Inverse Frontal Gas Chromatography. *J. Chromatogr.* **1986**, *351*, 417–423.

37. Das, B.; Sethi, R.K.; Chopra, S.L. Sorption and Desorption of Water Vapour on Starch. *Isr. J. Chem.* **1972**, *10*, 963–965.
38. El-Shattawy, H.H.; Kildsig, S.O.; Peck, G.E. Cephalexin-Direct Compression Excipients: Formulation Stability Screening Using Differential Scanning Calorimetry. *Drug. Dev. Ind. Pharm.* **1982**, *8* (6), 897–909.
39. Kerc, J.; Srcic, S.; Urleb, U.; Kanalec, A.; Kofler, B.; Smid-Korbar, J. Compatibility Study Between Acetylcysteine and Some Commonly Used Tablet Excipients. *J. Pharm. Pharmacol.* **1992**, *44*, 515–518.
40. McKenna, A.; McCafferty, D.F. Effect of Particle Size on the Compaction Mechanism and Tensile Strength of Tablets. *J. Pharm. Pharmacol.* **1982**, *34*, 347–351.
41. Guyot-Hermann, A.M. Disintegration Mechanisms of Tablets Containing Starches. Hypothesis About the Particle–Particle Repulsive Force. *Drug Dev. Ind. Pharm.* **1981**, *7* (2), 155–177.
42. Palmieri, G.F.; Lovato, D.; Martelli, S. New Controlled-Release Ibuprofen Tablets. *Drug Dev. Ind. Pharm.* **1999**, *25* (5), 671–677.
43. Sanchez, L.; Torrado, S.; Lastres, J.L. Gelatinized/Freeze-Dried Starch as Excipient in Sustained Release Tablets. *Int. J. Pharm.* **1995**, *115*, 201–208.
44. Singh, M.; O'Hagan, D. The Preparation and Characterization of Polymeric Antigen Delivery Systems for Oral Administration. *Adv. Drug. Delivery Rev.* **1998**, *34*, 285–304.
45. Wurzburg, O.B. Starch in the Food Industry. *Handbook of Food Additives*; Furia, T.E., Ed.; The Chemical Rubber Company: Boca Raton FL, **1972**, *1*, 361–395.
46. Bolhuis, G.K.; Arends-Scholte, A.W.; Stuut, G.J.; de Vries, J.A. Disintegration Efficiency of Sodium Starch Glycolates, Prepared from Different Native Starches. *Eur. J. Pharm. Biopharm.* **1994**, *40* (5), 317–320.
47. Gebre-Mariam, T.; Winnemoller, M.; Schmidt, P.C. An Evaluation of the Disintegration Efficiency of a Sodium Starch Glycolate Prepared from Enset Starch in Compressed Tablets. *Eur. J. Pharm. Biopharm.* **1996**, *43* (2), 124–132.
48. Parrott, E.L. Comparative Evaluation of a New Direct Compression Excipient, SoludexTM 15. *Drug Dev. Ind. Pharm.* **1989**, *15*, 561–583.
49. Li, L.C.; Peck, G.E. The Effect of Moisture Content on the Compression Properties of Maltodextrins. *J. Pharm. Pharmacol.* **1990**, *42*, 272–275.
50. Zhou, F.; Vervaet, C.; Schelkens, M.; Lefebvre, R.; Remon, J.P. Bioavailability of Ibuprofen from Matrix Pellets Based on the Combination of Waxes and Starch Derivatives. *Int. J. Pharm.* **1998**, *168*, 79–84.

STATISTICAL METHODS

Charles Bon

AAI International, Wilmington, North Carolina, U.S.A.

INTRODUCTION

Death and taxes, as the old adage goes, are the only certainties in life. Although this is an overstatement, it does emphasize the uncertain world in which the pharmaceutical scientist lives and works. Faced with "estimates" of product characteristics such as potency, content uniformity, impurity levels, and dissolution performance, the scientist must make important go/no-go decisions. These estimated values, as is true for most measured quantities, inherently vary from the true values on which correct decisions depend. Statistical methods provide tools that enable the pharmaceutical scientist to act decisively in an uncertain world. An understanding of basic statistical methods is important for all who work in the pharmaceutical field.

SOME BASIC CONCEPTS

In the developmental process for many drugs, the drug product is administered under controlled conditions to healthy, normal individuals or to the targeted patient population. This is done to characterize the rate and extent of absorption, the bioavailability, of the active drug contained in the product. The bioavailability is estimated from the measured concentrations of the drug that appear in serial blood specimens collected over a period of time after product administration. Basic statistical principles govern the behavior of the typical bioanalytical procedure used to measure these concentrations in the collected blood specimens.

The first step of a typical procedure might involve the transfer of 0.5 ml of the specimen into a screw-cap culture tube using a disposable serological pipette. Next, 100 μ l of internal standard, a chemical structurally similar to the drug of interest, is added. A small volume, 1.0 ml, of a buffer solution at an appropriate pH level is added to decrease the aqueous solubility of the drug and internal standard. After thorough mixing, 7 ml of diethyl ether is added to the tube, and it is capped and shaken to extract the drug and internal standard into the organic ether phase. The tube is centrifuged to obtain a clean separation of

the aqueous and the organic phases. A 5-ml portion of the organic phase is transferred to a clean culture tube and is evaporated to dryness under nitrogen. The residual of the extract is reconstituted in 100 μ l of mobile phase, an acetonitrile/methanol/buffer solution, and 10 μ l is injected onto a high-pressure liquid chromatograph (HPLC) equipped with a 3- μ , C8 column. The column effluent is monitored for ultraviolet absorbance at 280 nm, and the drug and internal standard peak area, or height, responses are determined. The drug-to-internal peak response ratio is calculated, and an estimate of the concentration of drug in each specimen is obtained through interpolation of the ratio on a calibration curve. The calibration curve is constructed from the peak response ratios of extracted calibrator specimens containing known amounts of the drug in interference-free blood.

There are several obvious sources where error can contribute to the uncertainty in the concentration estimate. If, for example, 1.0 ml of the specimen is mistakenly transferred instead of 0.5 ml, but the correct volume was used for preparing the calibrators, systematic error would have occurred. Systematic errors lead to a constant, predictable uncertainty in the estimate. To deal with a systematic error, we must recognize that it has occurred and then correct for the mistake that led to its occurrence. Statistical methods do not address systematic errors. If, however, the correct 0.5 ml-volume of specimen and 100- μ l volume of internal standard were used, we still would have errors affecting the concentration estimate. These errors would be random errors. Random errors are positive and negative deviations that inherently occur in any attempt to exactly measure a quantity, in our case, specific volumes of the specimen and internal standard. We might transfer 0.498 ml of a specimen owing to an unrecognized air bubble interfering with our reading of the meniscus in the transfer pipette. Or, perhaps, 0.502 ml is transferred because of some of the specimen adhering to the outside of the pipette. Statistical methods are our tools for dealing with uncertainty resulting from random error (chance).

After the transfer of the specimen and the addition of internal standard, the drug-to-internal ratio becomes a fixed quantity. Any additional random or systematic volume errors should not affect the concentration estimate. Upon injection of the reconstituted extract onto the HPLC,

however, random errors will occur that do affect the estimate. These are the result of chance deviations in the partitioning of the drug and internal standard between the mobile phase and the column. Random fluctuations in the UV detection system will also affect the concentration estimate. The influences of random errors are statistically additive. In our example, the random errors are independent of each other in that the occurrence, sign (positive or negative), and magnitude of each are unrelated to the occurrence, sign, and magnitude of the others. This is in contrast to correlated errors, which are related to each other in some predictable way.

Two important characteristics of any assay method are its accuracy and its reproducibility. Accuracy is how close an estimate is expected to be in relation to the true value for the specimen. Reproducibility relates to how repeated estimates of the same specimen vary about their average value. Both accuracy and reproducibility are usually defined for a given concentration and may differ between low and high specimen concentrations. In a good assay method, these differences should be inconsequential across the working range of the method.

The expected value of an estimate is the average of an infinite number of determinations of the estimate. These infinite determinations, taken as an aggregate, make up the population of estimates. A population does not have to be infinite in size. Some examples of finite populations are the potencies for all tablets in a given lot of a drug product or the sitting blood pressures of all patients who use a certain antihypertensive medication in the coming year. The mean (μ) of a population is a parameter of the population. The estimate of μ , obtained from a single concentration estimate for a specimen, varies from one determination to the next and is aptly referred to as a variable. If a variable conceptually takes on a continuum of values, as is the case for a concentration estimate, it is called a continuous variable. Variables that take on only certain discrete values, such as the number of tablets produced from 50 kg of active drug material, are referred to as discrete variables.

It is impossible to conduct an infinite number of extractions of a specimen to determine the accuracy of a method. As a result, we estimate the accuracy of an assay by performing a finite number of extractions (n) on the specimen. We report the accuracy as the mean ($\bar{x} = \sum x_i/n$, $i = 1, 2, \dots, n$) of the multiple determinations, expressed as a percent of the known concentration. The finite group of determinations is a sample from the population, and its mean is referred to as the sample mean. The sample mean is a statistic that estimates the population parameter μ . If we could obtain the means from an infinite number of same-size samples, regardless of their size, then the mean of these infinite sample means would equal μ . In statistical

terminology, we say that the sample mean is an unbiased estimator of the population mean. Unbiasedness is a desirable property for any statistic. Some statistics only become unbiased when sample sizes are large. These statistics are said to be asymptotically unbiased. Such statistics have their greatest utility when used with larger samples, such as those consisting of 20–30 determinations rather than smaller numbers such as 2–10.

Reproducibility, or precision, of a method relates to how individual estimates fluctuate around the average value. The magnitude of the fluctuation in the population is expressed by the parameter variance (σ^2). Variance is the average of the squared deviations about μ for all values x_i in the population: $\sum (x_i - \mu)^2/N$. An unbiased estimate of σ^2 is obtained from the deviation of each value (x_i) around the mean (\bar{x}) for a sample taken from the population: $s^2 = \sum (x_i - \bar{x})^2/(n - 1) = (\sum x_i^2 - (\sum x_i)^2/n)/(n - 1)$.

The form of the equation not involving \bar{x} is a convenient calculating formula that avoids rounding problems that can occur when individual values are very close to the mean. It is common practice to report the standard deviation s , which is the square root of the sample variance. The standard deviation is often normalized as the percent coefficient of variation CV% by dividing it by the sample mean and expressing the result as a percent. In the analyses of pharmaceutical dosage units and in FDA regulations governing these analyses, the term relative standard deviation (RSD) is used for this calculated quantity instead of the term CV%.

Given two estimates of a statistic, one from a sample of size n and the other from a sample of size $2n$, one might expect that the estimate from the larger sample would be more reliable than that from the smaller sample. This is, in fact, supported by statistical theory. If the variance in the population is σ^2 , then the variance of the sample mean for samples of size n is σ^2/n . The square root of this is the standard error of the mean. Consistent with the variance of the sample mean being $1/n$ times that of a single determination (σ^2), the standard deviation and the CV% of the sample mean are reduced by the square root of n . As a direct consequence, an assay method that relies on the mean of two independent concentration determinations has a CV $1/\sqrt{2}$ that of the same method based on a single determination. This provides an easy way to increase the precision (reduce variability) of a method. An example of this is found in radioimmunoassay in which it is common for a concentration estimate to be calculated from the mean response of two determinations of a specimen.

As noted previously, fluctuations in concentration estimates about the true value arise from multiple, independent, random errors. Each of the independent errors (σ_i^2) is statistically additive, such that the total assay error

$\sigma^2 = \Sigma \sigma_i^2$. Similarly, the CV% of the assay will be the square root of the sum of the squared CV% values for all independent sources of error in the method. It is interesting to examine the impact of this on the determination of the important pharmacokinetic measure C_{\max} , the maximum concentration of a drug after administration of a drug product. For many drugs, the biological variability, the degree to which the true C_{\max} value varies during replicate administrations of the product to the same individual, can have a CV of 25% or greater. If an assay method has a 10% CV and the biological variability for C_{\max} is 25%, then the C_{\max} estimate would have a CV of 26.9%. This is simply the square root of the sum of the two squared independent error CVs: $25^2 + 10^2$. If the precision of the assay method was improved to 5% CV, C_{\max} would be estimated with a 25.5% CV. The effort required to reduce a 10% CV assay to half that level, to obtain a mere 1.4% increase in overall precision would seldom prove to be cost-effective.

A USEFUL STATISTICAL DISTRIBUTION

The normal distribution appears ubiquitously throughout science and nature. References to applications of normal theory are found throughout the pharmaceutical literature. The distribution is one of the earliest introduced, having been published in 1733 by De Moivre. Most scientists have a basic familiarity with the distribution and its characteristic bell-shaped curve (Fig. 1). Those familiar with column chromatography might recognize this shape as that of the perfect chromatographic peak. In fact, the

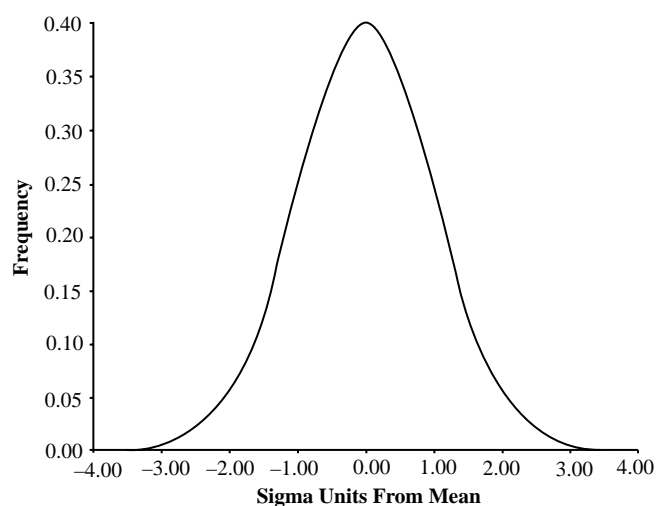


Fig. 1 The normal distribution.

principles of chromatographic peak symmetry and peak-to-peak resolution can be derived from normal theory. The normal distribution is defined by two parameters. The first, μ , defines the central location of the distribution, and the second, σ , defines the spread of the distribution about its center. The distribution has some unique properties. Its mean is the same as its median, the value at which 50% of the population are below and the remaining 50% are above. The distribution is symmetrical, its shape below its center is the mirror image of its shape above its center. Its mode, the value that occurs with the greatest frequency, also coincides with the mean and median. Approximately 68% of the distribution lies within 1σ of the mean, 95% lies within 2σ of the mean, and 99.8% lies within 3σ of the mean. The behavior of many observations in nature and many measurements in science can be approximated using the normal distribution. An important property that leads to the nearly universal application of the distribution is found in the central limit theorem. This theorem states that regardless of whether a given population is normally distributed, the distribution of the mean of randomly selected samples from the population will tend toward normality. This tendency increases as the size of the sample increases. If the population is, in reality, normally distributed, then a sample size of 1 is all that is needed. The more deviant the population distribution is from normality, the larger the sample size needs to be for its mean to be normally distributed.

It is reasonable to question whether the distribution of the estimates of a drug concentration in a blood specimen might be approximated by the normal distribution. Table 1 presents the results of repeated analyses of a specimen of interference-free plasma spiked to contain a known amount of drug. These data are taken from a comparative bioavailability study in which single doses of an unmarketed generic product and the marketed brand product of a drug were administered on separate occasions to healthy males. The values presented are the first of duplicate determinations of a quality control (QC) specimen that was included with each batch of subject specimens. This was done to verify that the in-process accuracy and precision of the assay method were consistent with the values observed during the assay validation.

A frequency histogram of the results is shown at the top of Fig. 2. The bottom of the figure shows the plot of the normal distribution with mean 207.6 ng/ml and σ 14.1 ng/ml. The shape of the histogram plot is similar to the plot of the normal distribution. The greatest deviation between the two is in the region of the center histogram. The sample distribution is higher peaked in its center, containing 42% of the values, than is the normal distribution, which has a 38% frequency at its center. The mean,

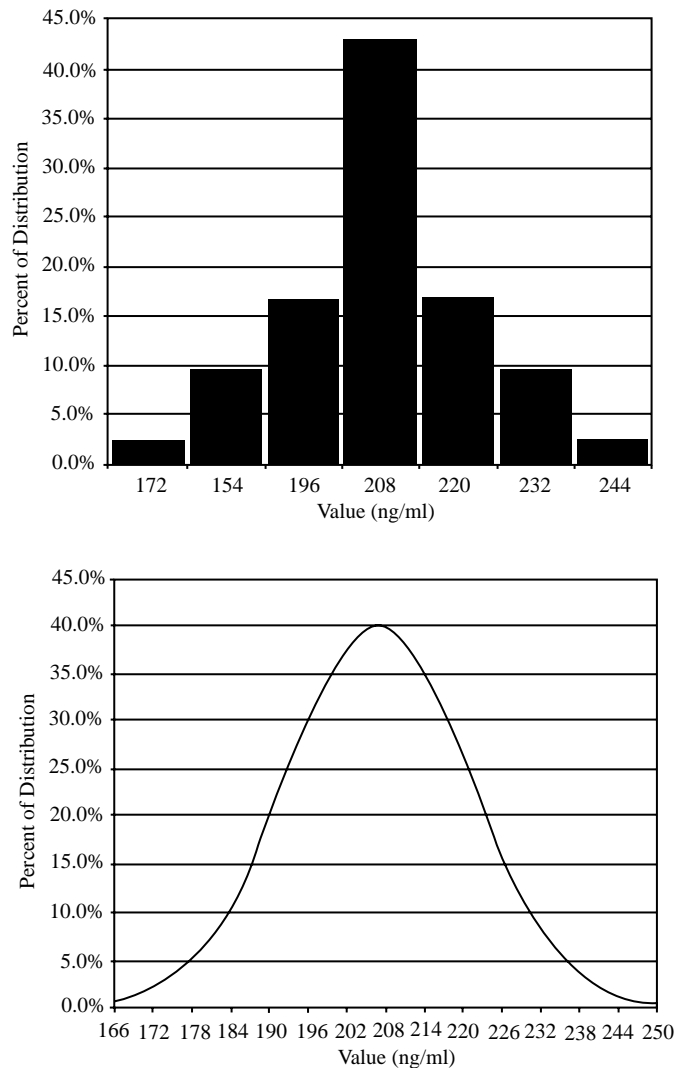
Table 1 Results of quality control specimen analyses

Batch	Conc. (ng/ml)	Batch	Conc. (ng/ml)
1	188	24	201
2	216	25	197
3	201	26	199
4	166	27	180
5	214	28	237
6	209	29	212
7	226	30	239
8	183	31	207
9	210	32	216
10	213	33	213
11	209	34	226
12	222	35	213
13	214	36	204
14	213	37	194
15	205	38	218
16	226	39	207
17	203	40	196
18	188	41	208
19	215	42	210
20	211	Mean	207.6
21	201	Median	209
22	205	Mode	213
23	206	Standard deviation	14.1

median, and mode of the QC values are close to being equal, as would be expected if the values had come from a normal distribution.

A particularly useful form of the normal distribution is obtained by transforming each value x_i to its standard normal value. This transformation converts the distribution to one that is independent of μ and σ . The conversion is $Z_i = (x_i - \mu)/\sigma$, where Z_i is known as the standard normal deviate and is normally distributed with $\mu = 0$ and $\sigma = 1$. Tables of Z -values can be found in any elementary statistics textbook. An example is presented in Table 2. The standard normal deviate table typically provides the proportion (area under the curve) of the distribution that lies between $-\infty$ and various Z -values (the lower tail) or between various Z -values and $+\infty$ (the upper tail). The proportion of the distribution lying within a given Z range around μ is calculated by taking the difference between the tabled proportions for $+Z$ and $-Z$.

Table 3 provides the expected percentages of the standard normal distribution that lie within some selected Z ranges about μ and compares these with the percentage of the QC sample distribution that falls within these ranges. The percentages for the sample distribution are

**Fig. 2** QC frequency histogram plot and corresponding normal distribution.

calculated by taking the number of QC values within each range, dividing it by 42, the total number of values in the sample, and expressing this as a percent. Because μ and σ are unknown, the sample mean and standard deviation are used in the calculations. The lower limit and upper limit values, in ng/ml, for each range is calculated as: $[x - Z \cdot s]$, $[x + Z \cdot s]$.

As seen in the frequency histogram, the observed distribution of the QC values in the vicinity of the mean, between $Z = -1$ and $Z = 1$, is higher than predicted by normal theory. However, the distribution outside this region closely resembles what would be expected for a normally distributed variable. A goodness-of-fit test to determine whether a variable follows a certain statistical

Table 2 Cumulative areas under the standard normal curve ($-\infty$ to Z)^{a,b}

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

^aEnter Table by Z-value to obtain cumulative area entry. As an example, area for $Z = 1.96$ (entry at row 1.9, column 0.06) is 0.9750, indicating that 97.5% of the standard normal distribution is below this Z-value, and 2.5% is above. Areas for negative Z-values are calculated by subtracting the area for the positive Z-value from 1. For example, the area for $Z = -1.96$ is calculated as $1 - 0.9750$ or 0.0250.

^bTable values generated using the SAS System.

distribution can be constructed using a chi-square statistic (Table 4). The range of the sample values is divided into intervals, and the expected number of values (E) that should fall in each interval is calculated. It is important to keep the intervals large enough so that at least five observations are expected in each. The number of observed values (O) in the sample that falls within each interval is then determined. The chi-square statistic for this test is $\chi^2 = \sum[(O - E)^2/E]$. If the calculated statistic value

exceeds that of the critical upper tail, chi-square value at $\alpha = 0.05$ ($p = 0.05$ testing level), we reject the hypothesis that the sample distribution is consistent with the assumed statistical distribution. If our calculated value is less than the critical value, we accept the hypothesis.

The chi-square table (Table 5) is entered according to the significance level (e.g., $p = 0.05$) and the degrees of freedom (df) for the calculated statistic. The degrees of freedom for the goodness-of-fit test are the total number of

Table 3 Distribution of results of quality control specimen analyses

$\pm Z$ units about mean	Expected percent	Observed percent
3.0	99.7	100
2.5	99	98
2.0	95	93
1.5	87	88
1.0	68	74
0.7	52	64
0.5	38	55
0.2	18	24
0.1	8	12

intervals less one. If, as in our case, population parameters for the assumed statistical distribution are not known and must be estimated from the sample, then the degrees of freedom are further reduced by the number of parameters estimated. In our example using eight intervals and estimating the mean and variance of the normal distribution from the sample, there are $8 - 1 - 2 = 5$ degrees of freedom. The critical chi-square value at $\alpha = 0.05$ with 5 df is 9.24. As Table 4 shows, our calculated statistic value, 6.33, is less than the critical value. Accordingly, we accept the assumption that the QC values appear to come from the assumed normal distribution with $\mu = 207.6$ and $\sigma = 14.1$.

Normal distribution theory can be used to test whether a particular sample value is consistent with other values or with our past experience. If the mean μ and the variance σ^2 are known, then we can determine how deviant an observed value x_i appears to be by calculating the statistic $Z = (x_i - \mu)/\sigma$ and comparing this with the table of standard normal deviates. Suppose that one of the values for our QC specimen was 170 ng/ml. Past experience has led us to believe that the results for this QC specimen are normally distributed with $\mu = 207.6$ ng/ml and $\sigma = 14.1$. Is the value of 170 ng/ml consistent with the assumption (hypothesis) that the assay is functioning properly with only random errors operative. To test this, we calculate the Z-statistic, which is equal to $(170 - 207.6)/14.1 = -37.6/14.1 = -2.77$. From the table of standard normal deviates, we see that the proportion of the normal distribution that lies within the range $-\infty$ to -2.77 (or similarly, between 2.77 and $+\infty$) is 0.0028. Therefore, the probability of encountering a value this far removed from the mean (± 37.6 ng/ml) is $p = 2 \times 0.0028 = 0.0056$. Typically, when there is less than a 5% probability ($p < 0.05$) of observing a value, we would question

whether our hypothesis was correct. We conclude, therefore, that either we have just observed a relatively rare event or that the assay was not working as expected. The hypothesis that the observed value deviates from its expected value owing only to random fluctuation is called the null hypothesis. If the null hypothesis is rejected, then we accept some stated alternative hypothesis. In this example, the alternative hypothesis is that the assay method was not functioning properly.

The Z-test can also be used to test if the mean \bar{x} from a sample of size n is consistent with the known mean μ of the population. The Z-statistic for this test is equal to $(\bar{x} - \mu)/(\sigma/\sqrt{n})$, where σ/\sqrt{n} is the standard error of the mean. The statistic is evaluated against the table of standard normal deviates just as we did in determining whether the QC value of 170 ng/ml was acceptable.

It should be noted that although it is common to use the 5% level of significance ($p < 0.05$) for testing, this level is not always appropriate. If the rejection of the 170 ng/ml QC value causes us to evaluate if the conduct and performance of the assay for a particular batch of specimens were correct, then the 5% level is probably appropriate. At this level, we would expect to have to investigate the performance of 1 in 20 of the batch runs even when everything was functioning properly. This would result in what is statistically known as a Type I error. A 5% level of extra circumspection would generally not be a problem. However, if the rejection of the 170 ng/ml value led to our dropping the value from estimates of in-process accuracy and precision, then it is not generally acceptable to erroneously exclude 5% of the values, thereby calculating estimates on the best 95% of the results. In this case, a 1% level of significance ($p < 0.01$), or even lower, would be more appropriate. At this lower level of significance, we would reject fewer values that deviated from the expected value simply owing to random error. Although this appears to be desirable, testing at a lowered level of significance also reduces our ability to reject values that differ from expected owing to true assay performance problems. This nondetection of truly aberrant values is referred to as a Type II error. Generally, the only way to decrease both Type I and Type II errors is to increase the size of the sample used in the statistical test.

THE DISTRIBUTION OF THE SAMPLE MEAN

We seldom know the population variance and, therefore, must estimate it from a sample. If the population is normally distributed or if the sample is a large one from a

Table 4 Chi-Square Goodness-of-Fit Test

Z_{low}	Z_{high}	Theoretical proportion	$x \cdot - Z_{\text{low}} * s$	$x \cdot + Z_{\text{high}} * s$	Expected (E)	Observed (O)	$(O - E)^2/E$
-4.00	-1.18	0.1190	151.2	191.0	5	5	0.000
-1.18	-0.72	0.1168	191.0	197.5	5	3	0.800
-0.72	-0.37	0.1199	197.5	202.4	5	4	0.200
-0.37	0.00	0.1443	202.4	207.6	6	7	0.167
0.00	0.37	0.1443	207.6	212.9	6	7	0.167
0.37	0.72	0.1199	212.9	217.8	5	9	3.200
0.72	1.18	0.1168	217.8	224.3	5	2	1.800
1.18	4.00	0.1190	224.3	264.1	5	5	0.000

Sum = 6.33; critical $\chi^2 = 9.24$.

population that is not normal, then we can construct a test to determine whether the sample mean is consistent with the known, or assumed, mean. To do so, we rely on a statistic based on Student's t -distribution. The t -distribution,

Table 5 Critical values of the Chi-square and student's t Distributions^a

DF	Chi-square values for one-sided test			t Values for one-sided test		
	0.05	0.025	0.005	0.05	0.025	0.005
1	2.71	3.84	5.02	6.31	12.71	63.69
2	4.61	5.99	7.38	2.92	4.30	9.92
3	6.25	7.81	9.35	2.35	3.18	5.84
4	7.78	9.49	11.14	2.13	2.78	4.60
5	9.24	11.07	12.83	2.02	2.57	4.03
6	10.64	12.59	14.45	1.94	2.45	3.71
7	12.02	14.07	16.02	1.89	2.36	3.50
8	13.36	15.51	17.53	1.86	2.31	3.36
9	14.68	16.92	19.02	1.83	2.26	3.25
10	15.99	18.31	20.48	1.81	2.23	3.17
11	17.28	19.68	21.92	1.80	2.20	3.11
12	18.55	21.03	23.34	1.78	2.18	3.05
14	21.06	23.68	26.12	1.76	2.14	2.98
16	23.54	26.30	28.85	1.75	2.12	2.92
18	25.99	28.87	31.53	1.73	2.10	2.88
20	28.41	31.41	34.17	1.72	2.09	2.85
25	34.38	37.65	40.65	1.71	2.06	2.79
30	40.26	43.77	46.98	1.70	2.04	2.75
35	46.06	49.80	53.20	1.69	2.03	2.72
40	51.81	55.76	59.34	1.68	2.02	2.70
50	63.17	67.50	71.42	1.68	2.01	2.68
60	74.40	79.08	83.30	1.67	2.00	2.66
∞	—	—	—	1.645	1.960	2.576

^aTable values generated using the SAS System.

attributed to W.S. Gossett, who wrote under the pseudonym "Student," describes the behavior of the means for samples taken from a normal distribution. The t -distribution is defined entirely by the sample size n , or more typically by its degrees of freedom $n - 1$. The distribution has a shape similar to that of the standard normal distribution, bell-shaped, but for small samples is lower peaked and broader than the standard normal distribution. As the sample size increases, the distribution approaches that of the normal distribution, coinciding with it when the sample size is infinite. The t -statistic is used like the Z -statistic, for testing the consistency of a sample mean with the population mean μ . The difference between the two statistics is that for the t -statistic, σ is replaced by its sample estimate, the standard deviation s . The statistic is $t = (x \cdot - \mu)/(s/\sqrt{n})$. Table 5 provides a listing of critical t -values (think of them as t deviates) for different degrees of freedom ($n - 1$). Note that at infinite degrees of freedom, the critical t -value is simply the standard normal deviate Z .

Assume that the QC specimen was supposed to be prepared to contain 200 ng/ml of drug and that we do not know the true population variance. Is a sample mean equal to 207.6 ng/ml based on 42 determinations consistent with the true value being 200 ng/ml? This typically would be stated in the form of a null (H_0) and alternative hypothesis (H_a):

$$H_0 : \mu = 200 \text{ ng/ml}$$

$$H_a : \mu \neq 200 \text{ ng/ml}$$

Because σ for the population is unknown, we must estimate it from the sample standard deviation s . We calculate $t = (207.6 - 200)/(14.1/\sqrt{42}) = 3.49$. Referring to Table 5 and entering it with 40 degrees of freedom, the closest-value to the $42 - 1 = 41$ degrees of freedom for our calculated statistic, we find a critical, one-sided, t -value at

the 0.005 level equal to 2.70. If the calculated t -statistic value exceeds a tabled critical value ($t > t_{\text{crit}}$) or if it is less than the negation of the critical value ($t < -t_{\text{crit}}$), then the null hypothesis is rejected. In the example, the calculated value 3.49 exceeds the critical value and we reject the hypothesis that the sample comes from a population with a mean of 200 ng/ml at the 0.01 level of significance (2×0.005). We, instead, accept the alternative hypothesis that the QC specimen appears to have been prepared to contain a drug concentration different than 200 ng/ml. This is an example of a one-sample t -test.

Another application of the t -test is the two-sample t -test. This test is used to determine whether two samples come from populations with the same mean ($\mu_1 - \mu_2 = 0$) or whether the population means differ by some hypothesized amount ($\mu_1 - \mu_2 = c$). In both cases, the statistic is calculated as $t = [(x_1 - x_2) - (\mu_1 - \mu_2)]/se$, where x_1 and x_2 are the means of first sample and the second sample. The denominator of the statistic is the standard error for the difference between the two means, calculated as: $se = s_p(1/n_1 + 1/n_2)^{1/2}$, where, $s_p = [((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2)/(n_1 + n_2 - 2)]^{1/2}$, s_1 and n_1 are the standard deviation and size of the first sample, and s_2 and n_2 are the standard deviation and size of the second sample. We enter the t -table with $n_1 + n_2 - 2$ degrees of freedom. If the tabled value is exceeded by our calculated value or if the calculated value is less than the negation of the tabled value, then we reject the hypothesis that the two samples are from populations with the same mean or differ by the hypothesized amount c .

Assume that we have a second set of 20 QC values from a second QC specimen, presumably prepared identically to the specimen from our previous set of values, and found a mean of 199.6 ng/ml and standard deviation 13.5 ng/ml. We can evaluate whether these results are consistent with those from 42 results on the first QC specimen. The null hypothesis for the test is that the two QC specimens have been prepared identically, $\mu_1 - \mu_2 = 0$. The alternative hypothesis would be that they are not identical, $\mu_1 - \mu_2 \neq 0$. The standard error is $se = [(41 \times 14.1^2 + 19 \times 13.5^2)/(41 + 19)]^{1/2}(1/42 + 1/20)^{1/2} = (193.566)^{1/2}(1/42 + 1/20)^{1/2} = 3.78$.

The t -statistic is $(207.6 - 199.6)/3.78 = 2.12$. Entering Table 5 with 60 degrees of freedom, we obtain a critical t -value for a one-sided test at the 0.025 level of significance equal to 2.00. As the alternative hypothesis is that the two QC specimens differ, without regard to which has the higher and which has the lower value, the test is a two-sided test. The critical table value at 0.025, one-sided, is the critical value to use for a two-sided test at 0.05. The calculated statistic exceeds this critical value, and we reject the hypothesis that the two sets of QC samples were

prepared identically. If we have the stock solutions used to prepare the two QC specimens, we would probably analyze them to see whether they have identical concentrations.

The t -distribution can also provide a tool to evaluate whether two samples on which paired determinations had been obtained appear to come from populations with the same mean (or from the same population). A paired t -test can be applied when we want to determine whether a newly trained analyst performs an assay method with the same proficiency as an experienced analyst. Suppose that each analyst processes seven different QC specimens and we obtain the following assay results, in the order (new analyst, experienced analyst): (12.6, 11.3), (3.46, 2.34), (25.4, 22.5), (10.3, 8.80), (5.89, 4.68), (16.4, 14.2), and (9.95, 8.20). The paired t -test deals with the differences for each of the paired results (new-experienced): 1.3, 1.12, 2.9, 1.5, 1.21, 2.2, and 1.75. The mean difference is 1.71, with a standard deviation of 0.641. The t -statistic is calculated by dividing this mean by the standard error (standard deviation/ \sqrt{n}) and comparing the result with the critical t -value with $n - 1$ degrees of freedom. With the $n = 7$ pairs in the example, the calculated t -statistic is $1.71/(0.641/\sqrt{7}) = 7.06$. This exceeds the critical t -value 2.45 for the two-sided, 0.05 level of significance (the one-sided 0.025 value in Table 5) with six degrees of freedom. We conclude that a difference as large as the one observed between the two analysts would not likely be attributable to random error alone. The "new" analyst may need some additional training.

ANOTHER USEFUL DISTRIBUTION

A commonly encountered statistical distribution is the binomial distribution. This distribution deals with the behavior of binary outcomes such as the flip of a coin (heads/tails), the gender of a child (boy/girl), or the determination if a tablet has acceptable potency (pass/fail). When dealing with a sequence of independent binary outcomes, such as multiple flips of a coin or determining whether the potencies of 20 tablets are individually acceptable, the binomial distribution can be used. The probability of observing x successes in n outcomes is $C(x,n) p^x q^{n-x}$. Binomial expansion for $x = 1$ to n is $C(0,n)p^0 q^n + C(1,n)p^1 q^{n-1} + C(2,n)p^2 q^{n-2} + \dots + C(n,n)p^n q^0$. This sum equals 1, as the probability of observing at least one of the possible outcomes is 1 (a certainty). The notation in the expansion $C(a,n)$ is the number of ways of obtaining groups of size a from n distinct items: $n!/a!(n-a)!$. As an example, the number of groups of size three obtainable when there are four

different items is $C(3,4) = 4!/[3!(4-3)!] = (4 \times 3 \times 2 \times 1)/[(3 \times 2 \times 1)(1)] = 4$. Note that $0!$ is defined as 1. The p in the binomial expansion is the success probability for the single binary event such as observing a head with one flip of a coin. The q stands for the single-event failure probability (e.g., observing a tail) and is equal to $1 - p$.

PUTTING IT ALL TOGETHER

In the analyses of blood specimens from subjects participating in bioavailability studies, the FDA instructs laboratories to include quality control specimens (QC) at each of three known concentrations (low, mid, and high). The QC specimens are processed in duplicate with each batch of subject specimens. The acceptance criteria for the batch, based on the results of these QC specimens, is that at least four of the six values must fall within a specified range about their nominal concentrations. In addition, no more than one value at each of the three QC concentration levels can be outside its acceptance range. Combining binomial and normal distribution theory, we can estimate the number of batch runs we expect to reject because of random error.

Assume that the acceptance limit for each QC value is that it must fall within 15% of its nominal concentration. Any value meeting this criterion passes and any not meeting this criterion fails (a binary outcome). The probability of accepting a batch run based on multiple binary outcomes (QC pass/fail determinations) will be governed by the binomial distribution. The probability of a single QC value failing is equal to the probability of obtaining a concentration outside 85–115% of its nominal concentration. The concentration estimates are assumed to be normally distributed with a mean equal to 100% of the nominal concentration. The sigma value for the estimates is equal to the CV of the assay. If the assay CV is 12%, the probability of a QC value passing acceptance criteria is the probability of obtaining a Z-value between $(85 - 100)/12$ and $(115 - 100)/12$ or between -1.25 and $+1.25$. The proportion of the standard normal distribution between -1.25 and $+1.25$ is 0.7888, which is p , the probability of a single QC value passing. The value of q is $1 - 0.7888$ or 0.2112. There are three mutually exclusive ways that at least four of the six QC values can pass the acceptance criterion. All six values could pass, five of the six could pass, or four of the six could pass. These outcomes are mutually exclusive; the occurrence of any one of them excludes the possibility that either of the other two occurs. The probability of at least four of six QC values passing, then, is the sum of the probabilities of each of the three mutually exclusive ways in which this event can occur.

It should be noted that in probability calculations, when an event A can occur through any one of the mutually exclusive events B, C, D, etc., then the probability of A is the sum of the individual probabilities of B, C, D, etc. However, if A occurs only when events B, C, D, etc. all occur, then the probability of A is the product of the individual probabilities of B, C, D, etc.

Applying the binomial expansion, the probability of at least four of the six QC values passing is $C(6,6)p^6q^0 + C(5,6)p^5q^1 + C(4,6)p^4q^2 = 1(0.7888)^6 + 6(0.7888)^5(0.2112) + 15(0.7888)^4(0.2112)^2$, which equals 0.8869. For a batch to pass, however, we have the additional restriction that when only four of six values pass, no more than one of the two failures can occur at the same QC level (low, mid, or high). This restriction reduces the 15 possible ways that four of six QC values can pass to 12. This reduces the probability of the batch being accepted to $0.8869 - 3(0.7888)^4(0.2112)^2$ or 0.8351. There is an 83.5% probability of accepting a batch run, and there is nearly a 16.5% chance of rejecting it because of random error. We might consider improving the precision of the assay rather than proceeding with a method that is anticipated to erroneously reject nearly 17% of our analyses. If the CV% of the method were improved to 8%, the probability of batch acceptance increases to 0.9875, and our expected failure rate is only 1.2%. With 42 batch runs, we expect to have to repeat 1–2 batches, as contrasted with the 12% CV assay, where we expect to repeat seven batches. Here is a case in which a modest improvement in assay precision reaps big rewards.

The chi-square distribution used in the goodness-of-fit test is useful in another important statistical test. Assume that previous experience leads us to believe that the σ of an assay method is no greater than 10 ng/ml. In processing 42 batches, the observed standard deviation was 14.1 ng/ml for the QC values. Is this result consistent with the prior knowledge of the assay ($H_0: \sigma \leq 10$), or does it appear that the precision of the assay has deteriorated ($H_a: \sigma > 10$)? Application of a chi-square test can help answer this question. The appropriate statistic in this case is $\chi^2 = (n - 1)s^2/\sigma^2$, which follows a chi-square distribution with $n - 1$ degrees of freedom. In our example, we replace s in the statistic with 14.1, our sample standard deviation, and σ with our assumed upper limit of 10 for σ . The calculated statistic for this one-sided test is $\chi^2 = (42 - 1)14.1^2/10^2 = 81.51$. This calculated value exceeds the critical chi-square value 59.34 at the 0.005 level of significance and 40 (approximately 41) degrees of freedom. We conclude that the assay precision had deteriorated and would probably launch an investigation as to why.

In manufacturing a drug product, it is common to collect and analyze specimens from the mix of active and inactive

ingredients, the blend. This is done to verify adequate uniformity of the blend before proceeding with the manufacturing process. The specimens are strategically collected from the blend container, for example, from the left, center, and right regions of the top, middle, and bottom of the container. Blend uniformity criteria usually require that the mean assay value for the specimens falls within an acceptable range (e.g., 95–105%) about the label claim (100% potency) for the product. In addition, the relative standard deviation CV% for the analyses of the blend specimens must not be greater than some specified limit (e.g., 5%). Using the characteristics of the normal and the chi-square distributions, we can estimate the chances of passing blend uniformity criteria.

Suppose that prior knowledge of the manufacturing process leads us to believe that we routinely produce blends with potencies between 97 and 103% of label claim. We estimate that the variability for a “good” blend is no more than 3% CV, a composite of true blend inhomogeneity, sampling error, and analytical method variability. What is the probability of accepting a good blend? The standard error of the mean for nine specimens from the blend will be 1%, our worst-case estimate of CV for a single determination (3%) divided by the square root of the number of specimens collected (9). Assuming that the true potency of the blend is 97% (a worst-case estimate), then the probability of observing a mean from nine blend specimens that is within our acceptance range of 95–105% is the same as the probability of observing a Z -value between $(95 - 97)/1$ and $(105 - 97)/1$ or between -2 and $+8$. The proportion of the standard normal distribution contained between -2 and $+8$ is 0.977, the probability of passing the first criterion for blend acceptance. The second criterion is that the CV of the nine blend specimens must be less than 5%. With a true CV of 3%, this is the probability of observing a chi-square value less than $(n - 1)5^2/3^2 = (8)(25)/9 = 22.2$. With eight degrees of freedom, the probability of a chi-square value less than 22.2 is approximately 0.996, which is the probability of passing the second criterion. The probability of accepting the blend material is the product of the individual probabilities of passing criterion 1 and criterion 2, or 0.977×0.996 , which equals 0.973. Therefore, 97% of our blend batches are expected to pass if only random error is operative.

Some have proposed widening the acceptance range (e.g., 90–110%) but requiring that all individual blend specimens fall within the widened range. Is this an easier or harder criterion to meet? The probability of a single-blend specimen being acceptable, is the probability of the standard normal deviate Z being between $(90 - 97)/3$ and $(110 - 97)/3$ or between -2.33 and $+4.33$. This probability is 0.99. For the blend to pass the first criterion,

the first specimen, the second, and the third, up through the ninth, must all, independently, be acceptable. The probability of passing this first criterion becomes 0.99^9 , or 0.914. This, when combined with the probability of meeting the second criterion, results in a probability of blend acceptance being 0.914×0.996 or 0.91. Only 91% of the blends are expected to pass. The chance failure rate goes from approximately 1 in 30, based on using the mean to 1 in 11 based on using individual values, despite the widened acceptance limits.

ONE FINAL DISTRIBUTION

There is one final statistical distribution, the F -distribution, that is an important addition to the basic statistical tool chest. This distribution is used in the evaluation of two variance estimates to determine whether they are consistent with each other. The QC sample based on 42 estimates (41 degrees of freedom) had a standard deviation of 14.1 ng/ml. If we had another set of 31 QC values (30 degrees of freedom), perhaps from a second bioavailability study, with a standard deviation of 19.5 ng/ml, we might want to know whether the assay precision values for the two studies were consistent. The variance ratio statistic is s_1^2/s_2^2 , where s_1 is the higher of the two standard deviations, and s_2 is the lower of the two. The calculated value of the statistic is compared with tables of critical F -values with $n_1 - 1$ numerator and $n_2 - 1$ denominator degrees of freedom. In our case, the critical $F_{30,41}$ is approximately equal to 1.74 at the 5% level of significance (see Table 6 for $F_{30,40}$, the closest value). The calculated ratio is $19.5^2/14.1^2 = 1.91$, which exceeds the critical value. The interpretation is that the assay precision for the second of the two studies differs significantly ($p < 0.05$) from that of the first (it has less precision, greater CV%).

The F -distribution has great utility in a statistical test referred to as analysis of variance (ANOVA). ANOVA is a powerful tool for testing the equivalence of means from samples obtained from normally distributed, or approximately normally distributed, populations. As an example, suppose that the following are the content uniformity values on 20 tablets from each of four different lots: lot A mean = 99.5%, standard deviation = 2.6%; lot B mean = 100.2%, standard deviation = 2.8%; lot C mean = 90.5%, standard deviation = 2.1%; and lot D mean = 100.3%, standard deviation = 2.7%.

Are any of the lots different from the other lots? To answer this question, we need to conduct a statistical test with a null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$

(all means are equal to some unknown μ), and an alternative hypothesis H_a : $\mu_1 \neq \mu$ or $\mu_2 \neq \mu$ or $\mu_3 \neq \mu$ or $\mu_4 \neq \mu$ (at least one mean is not equal to at least one of the other means). It is assumed that all lots have the same, unknown, variance, σ^2 . The F -statistic involves the calculation of two variance estimates $s_1^2 = 1/(k - 1) \sum n_j(x_{.j} - x_{..})^2$ and $s_2^2 = 1/(N - k) \sum \sum (x_{ij} - x_{.j})^2$. In the statistic, $x_{..}$ is the grand mean across all $k = 4$ lots, n_j is the number of values from which the j th lot mean $x_{.j}$ was determined (20 for each lot), and $N = \sum n_j = \sum \sum n_{ij}$ is the total number of x_{ij} values ($20 + 20 + 20 + 20 = 80$). The s_1^2 is a pooled variance estimate of how the category means vary about the grand mean, and s_2^2 is a pooled variance estimate of how the individual values within each category vary about their category mean. If the identification (grouping) of the values into categories (lots) does not affect the variance estimate, then the variance ratio s_1^2/s_2^2 will differ from 1.0 by only random error. The F -distribution describes how the ratio varies about 1.0 owing to random error when a set of values are arbitrarily grouped into categories. If the calculated statistic value exceeds the critical $F_{k-1, N-k}$ tabled value, then the null hypothesis of equal means is rejected.

In our example, $x_{..} = \sum n_j x_{.j} / N = (20 \times 99.5 + 20 \times 100.2 + 20 \times 90.5 + 20 \times 100.3) / 80 = 97.625$. $s_1^2 = (1/3)[(99.5 - 97.625)^2 + (100.2 - 97.625)^2 + (90.5 - 97.625)^2 + (100.3 - 97.625)^2] = 22.689$. Because the square of the standard deviation in each lot j is $s_j^2 = \sum (x_{ij} - x_{.j})^2 / (n_j - 1)$, we can calculate s_2^2 as $(1/(N - k)) \sum (n_j - 1) s_j^2 = 1/(80 - 4) [19 \times 2.6^2 + 19 \times 2.8^2 + 19 \times 2.1^2 + 19 \times 2.7^2] = 69.0$. The calculated variance ratio is $53.8/6.575 = 8.18$ and the critical 5% level $F_{3,76}$ is < 2.76 . The critical value 2.76 is that of $F_{3,60}$ in Table 6. This value is the closest tabled value with the desired numerator degrees of freedom (3) and denominator degrees of freedom 60, which do not exceed $N - k = 76$. As the calculated statistic value exceeds the critical F -value, we reject the null hypothesis that the means of all lots are the same. We accept the alternative hypothesis that at least one of the four means differs statistically from at least one of the other means. Looking at the four means, it appears that lot C differs from the others.

Instead of performing ANOVA, we might have considered pairing each lot mean with each of the other lot means and performing multiple two-sample t -tests to determine which ones differ significantly from the others. This approach would not be acceptable. First, it does not use all of the available information for the pooled denominator variance estimate. Second, it introduces what is known as the multiple comparison problem. The problem arises if there are multiple, separate, statistical evaluations conducted on the same set of data. When we

conduct a single test of a hypothesis at the 5% level of significance ($\alpha = 0.05$), such as $H_0: A = B$, we expect to falsely reject the hypothesis 5% of the time when A actually equals B (Type I error). We have, therefore, a 95% probability of being correct in the assessment when the null hypothesis $A = B$ is true. This is also true for the F -test of $A = B = C = D$ in ANOVA. If we were to independently perform multiple t -tests on our data with hypotheses such as $A = B$, $B = C$, $C = D$, $A = C$, etc., then for each of these 5% level tests, we have a 95% probability of obtaining a correct assessment when the null hypothesis that all means are equal is true. To correctly accept the null hypothesis of all means being equal, we must simultaneously conclude that $A = B$ and $B = C$ and $C = D$ and $A = C$, etc., from the multiple t -tests. The probability of doing so is $(0.95)(0.95)(0.95) \dots (0.95) = (0.95)^n$, where n is the number of pair-wise comparisons conducted by t -tests. The probability of being incorrect in this multiple pair-wise approach is $1 - (0.95)^n$, which exceeds the desired 5% level for any $n > 1$. In our case, there are $C(4, 2)$ or six possible pair-wise comparisons. The multiple pair-wise approach has a probability of an incorrect assessment, Type I error when the null hypothesis is true, equal to $1 - (0.95)^6 = 0.265$. This approach would essentially be testing at a 26.5% level of significance, rather than at the desired 5% level. For this reason, we only consider pair-wise examination of the data when the global assessment of equality of means is rejected by ANOVA. This maintains the desired 5% significance level.

For the post-ANOVA, pair-wise evaluations, there are procedures to deal with the multiple comparison problem. One such procedure is based on the F -distribution with one and $N - k$ degrees of freedom. This test also relies on the value of s_2^2 from ANOVA. The test statistic is $F = (1/s_2^2) [(x_{.1} - x_{.2})^2 / (1/n_1 + 1/n_2)]$, where $x_{.1}$, $x_{.2}$ are the means of the n_1 and n_2 values for the two lots in the pair-wise comparison. Comparing lot A and lot C: $F = (1/6.575) [(99.5 - 90.5)^2 / (1/20 + 1/20)] = 123.2$. This far exceeds the critical $F_{1,76}$ value at even a 1% level, which is < 7.08 , based on $F_{1,60}$, and we therefore reject the hypothesis that lot A and lot C means are equal. Because, the means for lot B and lot D differ from that of lot C by an even greater amount, they also are found to be statistically different from the lot C mean. By contrast, the comparison of lots A and D, with means of 99.5% and 100.3%, respectively, have an F -test value of 0.97, far less than the critical 5% value, which is < 4.00 .

The test of the four lots is an example of a one-way ANOVA. The one-way comes from the fact that there is only one category (lot) into which the data is classified. Often, we have more than one category (class variable) in which we need to classify data. Although our interest may

Table 6 Critical values of the *F*-distribution for 5% level of significance^a

Numerator degrees of freedom:																				
Df ^b	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	60	100	
1	161	200	216	225	230	234	237	239	241	242	243	244	246	248	249	250	251	252	253	
2	18.5	19.0	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.70	8.66	8.63	8.62	8.59	8.57	8.55	
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.52	4.50	4.46	4.43	4.41	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.94	3.87	3.83	3.81	3.77	3.74	3.71	
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.51	3.44	3.40	3.38	3.34	3.30	3.27	
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.22	3.15	3.11	3.08	3.04	3.01	2.97	
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.01	2.94	2.89	2.86	2.83	2.79	2.76	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.85	2.77	2.73	2.70	2.66	2.62	2.59	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.72	2.65	2.60	2.57	2.53	2.49	2.46	
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.62	2.54	2.50	2.47	2.43	2.38	2.35	
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.53	2.46	2.41	2.38	2.34	2.30	2.26	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.46	2.39	2.34	2.31	2.27	2.22	2.19	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.40	2.33	2.28	2.25	2.20	2.16	2.12	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.35	2.28	2.23	2.19	2.15	2.11	2.07	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.31	2.23	2.18	2.15	2.10	2.06	2.02	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.27	2.19	2.14	2.11	2.06	2.02	1.98	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.20	2.12	2.07	2.04	1.99	1.95	1.91	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.01	1.93	1.88	1.84	1.79	1.74	1.70	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.92	1.84	1.78	1.74	1.69	1.64	1.59	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.87	1.78	1.73	1.69	1.63	1.58	1.52	
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.84	1.75	1.69	1.65	1.59	1.53	1.48	
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.77	1.68	1.62	1.57	1.52	1.45	1.39	
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.67	1.57	1.51	1.46	1.39	1.32	1.24	

^aDf are the denominator degrees of freedom.

^bTable values generated using the SAS System.

be to determine only whether a particular class variable has meaning, it is important to include other class variables that may influence the variability of the data. ANOVA involves a null hypothesis for each classification variable that proposes that the means at each different level of the class (category) are all equal. If we reject the null hypothesis we conclude in favor of the alternative hypothesis, that at least one mean in the class differs from at least one other mean in the class. This is also a conclusion that categorizing the data by that class variable has meaning.

Assume that we have two analysts to determine the drug concentrations in plasma specimens from a bioavailability study, using our previously described analytical method. Before starting the analyses, we want to determine whether they are equally proficient with the method. We might set up the test by obtaining plasma specimens at three different concentrations of drug. We then have each analyst process in duplicate each specimen on each of two different days. The resulting drug concentration values can be categorized by three different class variables: analyst, day, and specimen. The variable analyst has two possible levels, analysts one and two. Day has two levels (days 1 and 2), and specimen has three (A, B, and C). ANOVA on the results is a three-way one, named for the three classification variables included. Our interest is in the classification by analyst, but the other two variables are necessary to properly define how the experiment was conducted. Table 7 shows the results of the experiment.

ANOVA calculations are straightforward in this example and are easily expanded to situations in which there are higher numbers of categories. The first ANOVA quantity we calculate is C , the correction factor. C is simply the square of the sum of all the individual values, divided by N , the total number of values: $(\sum x_{ijkl})^2/N$, where $i = 1-2$ (analysts), $j = 1-2$ (days), $k = 1-3$ (specimens), and $l = 1-2$ values for each specimen for each analyst on each day. Table 8 demonstrates the calculation of this and the other calculated ANOVA quantities. The quantity A is calculated as the sum of the squared individual values: $\sum x_{ijkl}^2$. The ANOVA quantity for each classification variable (category) is the sum of the squared totals for each level of

Table 7 Evaluation of the proficiency of two analysts to process plasma samples

Specimen	Day	Analyst 1	Analyst 2
A	1	52.8	57.9
A	1	52.3	57.5
A	2	53.3	58.0
A	2	52.1	57.0
B	1	42.2	45.6
B	1	42.4	44.2
B	2	41.7	43.8
B	2	40.4	44.1
C	1	56.8	64.0
C	1	57.0	64.9
C	2	59.6	62.6
C	2	57.2	61.8

the classification, divided by the number of values in each level of the classification. As seen with the variable specimen, this is the sum of the squared totals for the values for each of the specimens A, B, and C, all divided by 8, the number of values for each specimen (2 values for each of 2 days for each of 2 analysts). The principle of ANOVA is that the total sum of squares SS_{Total} is divisible into its component sums of squares for each classification variable plus random error: $SS_{\text{Total}} = SS_{\text{Analyst}} + SS_{\text{Day}} + SS_{\text{Specimen}} + SS_{\text{Error}}$. The total sums of squares are calculated as the difference between two ANOVA quantities, $A - C$. The sums of squares for each classification variable is its ANOVA quantity minus C . The sums of squares for random error is determined by difference: $SS_{\text{Error}} = SS_{\text{Total}} - SS_{\text{Analyst}} - SS_{\text{Day}} - SS_{\text{Specimen}}$. The degrees of freedom are also additive with $df_{\text{Total}} = df_{\text{Analyst}} + df_{\text{Day}} + df_{\text{Specimen}} + df_{\text{Error}}$. The total degrees of freedom are simply the total number of values less one. The degrees of freedom for each class variable are the number of levels within the class less one. The error degrees of freedom are determined by difference. Dividing each sum of squares by its degrees of freedom provides the mean square MS , which is an estimate of the population variance σ^2 .

Table 8 Calculation of ANOVA quantities

	Totals	No. of observations	Σ (Totals squared)/No.
All 24 values	1269.2	24	$C = 67119.53$
Days 1 and 2	637.6, 631.6	12	$T = 67121.03$
Analysts 1 and 2	607.8, 661.4	12	$K = 67239.23$
Specimens A, B, and C	440.9, 344.4, 483.9	8	$R = 68395.42$
		Square of each value	$A = 68541.24$

Table 9 Calculation of ANOVA table

ANOVA				
Source of variation	DF	SS	MS = SS/DF	F-ratio
Day	1	1.5000 ($T - C$)	1.5000	1.16
Analyst	1	119.7067 ($K - C$)	119.7067	92.42 ^a
Specimen	2	1275.8958 ($R - C$)	637.9479	492.51 ^a
Error	19	24.6108 ($A - T - K - R + 2C$)	1.2953	
Total	23	1421.7133 ($A - C$)		

^aIndicates statistical significance ($p < 0.05$).

Table 9 shows the construction of the ANOVA table. If the variance estimate of a class variable MS_{variable} deviates significantly from that obtained by that for random error MS_{Error} , then the null hypothesis that the means at the different levels for that variable are equal is rejected. In other words, the classification of data by that variable is explanatory of the variation observed in the data. We conduct the test by using the variance ratio test $F = MS_{\text{variable}}/MS_{\text{Error}}$, with numerator degrees of freedom equal to those of the variable and denominator degrees of freedom equal to that of the error term. As shown in Table 9, we reject the hypothesis that the two analysts process the specimens equivalently. We also see, as we expected, that the specimen levels *A*, *B*, and *C* are not equal. There are no detectable differences in the values obtained over the 2 days of processing.

In the previous example, we tested the hypothesis of “equality.” In reality, our interests are usually not in equality but in “comparability.” We generally do not require, or even logically expect, that two lots of the same pharmaceutical product will have the exact same potency. If we had several different lots of a drug product and analyzed enough units from each (e.g., 100–200 tablets), we could detect as statistically significant even a 1% difference in the potency between the lots. Although statistically significant, such a difference would have no practical significance. We require, however, that the potencies, as with the analysts’ performance on an assay, be comparable, allowing for a reasonable margin of error. Although hypothesis testing deals with tests of equality, the closely related confidence interval approach deals with comparability. The confidence interval calculation can be applied to comparisons of means from samples drawn from normal populations or from any population if the samples are large (thanks to the central limit theorem). First, we calculate the difference between two means from samples from the two population $\bar{x}_1 - \bar{x}_2$, which estimates the difference between the population means $\mu_1 - \mu_2$. Next, we need to calculate the standard error for this difference.

The standard error is calculated as $se = (s_p^2/n_1 + s_p^2/n_2)^{0.5}$, where n_1 and n_2 are the number of values in the first and second samples, and s_p^2 is a variance estimate (s_p^2 from t -test and s_2^2 from one-way ANOVA or MS_{Error}). We have $(1 - \alpha) \times 100\%$ confidence that the true difference between the population means falls within the interval $(\bar{x}_1 - \bar{x}_2) \pm se \times t_{\alpha/2}$, where $t_{\alpha/2}$ is the critical one-sided t -value at the $\alpha/2$ level of significance and degrees of freedom equal to those for our variance estimate s_p^2 .

In ANOVA for the results from the two analysts, $MS_{\text{Error}} = 1.2953$ with 19 degrees of freedom. The means for analyst 1 was 607.8/12, or 50.65 and for analyst 2 was 661.4/12, or 55.12. The difference between the means is -4.47 . The standard error se for the difference would be $(s_p^2/n_1 + s_p^2/n_2)^{0.5} = (2s_p^2/n)^{0.5} = (2MS_{\text{Error}}/12)^{0.5} = (2 \times 1.2953/12)^{0.5} = 0.465$. The critical t -value for a 95% confidence interval is the one-sided, 0.025 level value at 19 degrees of freedom, 2.095. The 95% confidence interval is $(-4.47) \pm (0.465)(2.095)$ or from -5.44 to -3.50 . We have 95% confidence that the results of analyst 1 are between 6.34% ($100 \times 3.50/55.12$) to 9.88% ($100 \times 5.44/55.12$) lower than those of analyst 2 using the assay method. Because the value 0 is not included in the 95% confidence interval, we can conclude at the 5% level that there is a statistically significant difference between the two analysts, the same conclusion we had with ANOVA. However, a difference of 9.88%, the maximum confidence interval limit, might not be large enough for us to reject that the two analysts are comparable in their performance of the assay method. This is a decision that was not possible based solely on ANOVA results.

LINEAR REGRESSION

Those familiar with analytical methods probably have familiarity with “fitting a straight line” through data points. The statistical method generally used is known as

linear regression or ordinary least-squares. Even the simplest scientific calculators and spreadsheet programs contain the methods for determining the slope (m) and intercept (b) of the line relating the dependent variable y (measured with random error) to the independent variable x (without random error), $y = mx + b$. The statistical form of this is $y_i = mx_i + b + e_i$, where e_i is the random error for the observation y_i at a specific value of x_i . The calculations for fitting the line are easy, with $m = [\Sigma(x_i - \bar{x})(y_i - \bar{y})]/[\Sigma(x_i - \bar{x})^2]$ and $b = \bar{y} - m\bar{x}$, where \bar{y} and \bar{x} are the mean values for all the y_i and x_i values, respectively. It is common to have more than a single y value for each x value. A basic assumption is that the variance (error) in the determination of each y_i value is independent of its corresponding x_i value.

As previously indicated, the determination of a drug concentrations in plasma specimens requires the construction of a calibration response curve. This curve is often constructed as a straight line from the measured peak response ratios (y_i) plotted against their respective calibrator concentrations (x_i). The drug concentration in a specimen or the apparent (back-calculated) calibrator concentration is obtained from a rearrangement of the equation for the calibration line (without error) $x_i = (y_i - b)/m$. An example of a calibration curve with

back-calculated concentrations is given the first four columns of Table 10.

It is obvious that the lower end of the calibration curve does not provide an accurate representation of the calibrator concentrations, a problem that does not exist at the higher end of the curve. This problem illustrates what can happen when there is violation of the assumption of equal (homogeneous) variance for the y_i values. In bioanalytical methods, the largest component of the random error can often be attributed to volume errors. One often finds that the standard deviation of the response (y_i) is proportional to the concentration (x_i), that is, $e_i = kx_i$, where k is a constant. This violates the assumption of homogeneity of variance for y_i . To correct for this, we can use a variance-stabilizing transformation. If we divide the equation of the line by x_i , we obtain $y_i/x_i = m + b(1/x_i) + e_i/x_i$. However, $e_i/x_i = K$ is a constant, resulting in equal variance for $y'_i = y_i/x_i$. When y'_i is regressed against $1/x_i$, we obtain a line with slope equal to our desired intercept b and an intercept equal to the desired slope m . The last four columns of Table 10 demonstrate with calibrator data how this transformation provides accurate estimates for the back-calculated concentrations. The variance-stabilized line provides a slope estimate that differs only slightly from that of the original line ($0.938 \rightarrow 0.941$) and an

Table 10 Linear regression analyses of calibrator curve data

Calibrator (ng/ml) (x)	Response ratio (y)	Back ^a calculated (ng/ml)	Accuracy (%)	1/x (x')	y/x (y')	Back ^b calculated (ng/ml)	Accuracy (%)
0.250	0.321	0.134	54	4.000	1.285	0.253	101
0.250	0.311	0.123	49	4.000	1.244	0.242	97
0.500	0.545	0.373	75	2.000	1.090	0.490	98
0.500	0.59	0.419	84	2.000	1.178	0.537	107
1.00	1.042	0.902	90	1.000	1.042	1.02	102
1.00	0.994	0.851	85	1.000	0.994	0.967	97
5.00	4.981	5.10	102	0.200	0.996	5.20	104
5.00	4.466	4.55	91	0.200	0.893	4.66	93
10.0	9.911	10.4	104	0.100	0.991	10.4	104
10.0	9.247	9.65	96	0.100	0.925	9.73	97
25.0	24.132	25.5	102	0.040	0.965	25.5	102
25.0	22.448	23.7	95	0.040	0.898	23.8	95
50.0	44.061	46.7	93	0.020	0.881	46.7	93
50.0	51.171	54.3	109	0.020	1.023	54.3	109
100	101.132	108	108	0.010	1.011	107	107
100	88.314	93.9	94	0.010	0.883	93.7	94
250	227.438	242	97	0.004	0.910	242	97
250	241.472	257	103	0.004	0.966	256	103

^aSlope = 0.938; intercept = 0.195.

^bSlope = 0.941; intercept = 0.0834.

intercept estimate that differs more substantially (0.195 \rightarrow 0.0834). The approach of weighted least-squares, available in many advanced regression programs, gives the same results, using a $1/x^2$ weighting, and provides a greater choice for variance stabilization.

A FINAL WORD

The methods presented will hopefully provide a basic foundation for the application of statistics to pharmaceutical problems. Because of space limitations, the discussion has been limited to situations in which known statistical distributions could be assumed (parametric analyses). This is not always the case with real-life data. Fortunately, there are a number of nonparametric, distribution-free methods available to the pharmaceutical scientist to deal with analyses of such data. A general knowledge of statistical methods is a necessity for the pharmaceutical scientist. It is, however, also important for the scientist to realize when a problem or experimental design requires consultation with a statistician. The pharmaceutical statistician has a toolbox of methods considerably more advanced than the few basic methods presented here.

BIBLIOGRAPHY

- Bolton, S. *Pharmaceutical Statistics*, 3rd Ed.; Marcel Dekker, Inc.: New York, 1997.
- Brown, B.W., Jr.; Hollander, M. *Statistics: A Biomedical Introduction*; John Wiley & Sons: New York, 1977.
- Chatterjee, S.; Price, B. *Regression Analysis by Example*; John Wiley & Sons: New York, 1977.
- Dixon, W.J.; Massey, F.J., Jr. *Introduction to Statistical Analysis*, 3rd Ed.; McGraw-Hill Co.: New York, 1969.
- Draper, N.R.; Smith, H. *Applied Regression Analysis*, 2nd Ed.; John Wiley & Sons: New York, 1981.
- Kanji, G.K. *100 Statistical Tests*; Sage Publications, Inc.: Newbury Park, CA, 1993.
- Li, C.C. *Introduction to Experimental Statistics*; McGraw-Hill Co.: New York, 1964.
- Matthews, D.E.; Farewell, V.T. *Using and Understanding Medical Statistics*, 3rd revised Ed.; Karger AG: Basel; Switzerland, 1996, 5.
- Meyer, S.L. *Data Analysis for Scientists and Engineers*; John Wiley & Sons: New York, 1975.
- Snedecor, G.W.; Cochran, W.G. *Statistical Methods*, 8th Ed.; Iowa State University Press: Ames, 1989.
- SAS[®] System. *Base SAS and SAS/STAT[®] Software*; Release 6.12 TS060 for Windows SAS Institute, Inc.: Cary, NC, 1989–1996.
- Zar, J.H., Ed. *Biostatistical Analysis*; Prentice-Hall: Englewood Cliffs, NJ, 1974.

Statistical Process Control and Process Capability

Thomas D. Murphy

Consultant, Morristown, New Jersey, U.S.A.

Shailesh K. Singh

Merlin L. Utter

Wyeth, Pearl River, New York, U.S.A.

INTRODUCTION

Scope

Statistical process control (SPC) provides a statistical approach for evaluating processes and for improving the quality of these processes through elimination of special causes. When SPC is effectively implemented within a company, benefits can be derived through a reduced cost of manufacture, improved quality, fewer troubleshooting crises, and improved relationships with customers. Process capability is a companion tool—one that can be used once a state of statistical control is achieved—to assess the performance of a process relative to its product specifications. Process capability can be used to determine whether processes are capable of continually operating within their stated specification limits.

These tools have been used successfully in such industries as the automotive and semiconductor industries and have been credited for helping to provide great improvements in quality. They have been used to a lesser extent in the pharmaceutical industry but do provide the same potential for providing improvements in quality. This chapter provides a clear explanation of these concepts and how they can be used in practice. The relevance of these concepts to the pharmaceutical industry and some of the cautions associated with their use will be discussed. Some examples will be provided to aid in the understanding of their use. In addition, some possible reasons why these concepts have been slow in gaining acceptance within the pharmaceutical industry will be discussed.

Basic Principles

A *process* is defined as a set of interrelated or interacting activities that transform inputs into outputs. *Statistical process control* is a set of techniques for improving the quality of its output. Although manufacturing processes first come to mind, SPC can apply to any process, such as services, office procedures, and quality assurance func-

tions. *Process capability* analysis assesses the ability of the process to consistently meet specifications.

In any process, regardless of how well designed or maintained, a certain amount of natural or inherent variability will exist. This natural variability has been called a “stable system of chance causes.” A process operating with only chance causes is said to be in *statistical control*, and the variation is said to be because of *common causes*. Statistical process control evaluates whether or not a process is in statistical control with respect to one or more process or product characteristics.

If an upset occurs in a process (indicated by a spike, sustained shift, or trend in the process mean or by a change in the process variability), then an investigation is initiated to determine the *special cause* of the upset. The special cause, when determined, is either eliminated or controlled by process adjustment. By eliminating or controlling the special causes, process improvement occurs. *Root cause analysis* is a structured investigation that aims to identify a special cause and the actions necessary to eliminate it, and involves the use of a collection of tools including flow charts, brainstorming, cause-and-effect diagrams, and Pareto analysis, as described in other references.^[1,2]

The *control chart* is the basic analytical tool of SPC and is used for first *assessing* a process, then for *monitoring* a process output with respect to on-target control and process variability. A control chart is basically a time plot of a statistic calculated from a variable associated with a process. This variable may either be a process variable, such as temperature or flow rate, or a product variable, such as fill weight or potency. Examples of statistics are an individual measurement, an average of two or more measurements, a percentage of defective output items, a count of defect occurrences in time or space, or a measure of variation such as a range or standard deviation of two or more measurements.

Superimposed on this time plot are the upper and lower *control limits*, traditionally set at a distance of ± 3 times the standard error (SE) of the statistic from the *center line* or process mean. This controls the risk of a false alarm at a low level (a chance of 3 of 1000 if the



distribution is normal). The process is said to be in a *state of statistical control* if the plotted points appear to occur in a random pattern and are contained within the control limits. The centerline and control limits are calculated from retrospective data from the process.

STATISTICAL PROCESS CONTROL

The Shewhart Chart

The *Shewhart Chart* is applicable when the data are numeric and multiple data values are taken at each time point. For discrete parts processes, multiple items are sampled from the process at each time point and measured for one or more characteristics. These multiple measurements constitute a *subgroup*. The number n of measured items in the subgroup is the *subgroup size*. The values of the measured characteristic in a subgroup are denoted as X_1, X_2, \dots, X_n . The subgroup *average* $\bar{X} = \sum_{i=1}^n X_i/n$ is calculated and plotted on the Average Chart (also called the \bar{X} -Bar Chart) as a measure of the process average at that time point. The subgroup *range* $R = \text{Max}(X_i) - \text{Min}(X_i)$ is also calculated and plotted on the Range Chart (also called the R -Chart) as a measure of the process variation at that time point. The two charts are used as a unit and constitute the Shewhart Chart.

On each of the component charts, the average value is represented as a solid line, called the central line (CL). For the \bar{X} chart, the CL is the average of the averages $\bar{\bar{X}} = \sum_{j=1}^k \bar{X}_j/k$, and for the R -Chart, the CL is the average range $\bar{R} = \sum_{j=1}^k R_j/k$, where k is the number of subgroups.

Around the CL are the control limits, set at ± 3 SE of the statistic being plotted. If the statistic value falls outside the control limits, this is a signal that the process is not in a state of statistical control. Because the standard errors are functions of the process standard deviation σ , an estimate of this quantity is necessary. This can be supplied by the average range. The lower control limit (LCL) and upper control limit (UCL) are calculated as follows:

Chart	CL	LCL	UCL
Average	$\bar{\bar{X}}$	$\bar{\bar{X}} - A_2\bar{R}$	$\bar{\bar{X}} + A_2\bar{R}$
Range	\bar{R}	$D_3\bar{R}$	$D_4\bar{R}$

The values of A_2 , D_3 , and D_4 are given in Table 1.

The d_2 factor is used to provide an estimate of the process standard deviation $\hat{\sigma} = \bar{R}/d_2$, where the caret (^) denotes an estimated value. More extensive tables of control chart factors are available.^[3]

Table 1 Control chart factors

n	A_2	D_3	D_4	d_2
2	1.880	0	3.267	1.128
3	1.023	0	2.575	1.693
4	0.729	0	2.282	2.059
5	0.577	0	2.114	2.326
6	0.483	0	2.004	2.534
7	0.419	0.076	1.924	2.704
8	0.373	0.136	1.864	2.847
9	0.337	0.184	1.816	2.970
10	0.308	0.223	1.777	3.078

Shewhart Chart Example

An example of a Shewhart Chart is shown below for a hypothetical powder fill process. Five vials of product were sampled every hour and the net content of each vial was determined. The Shewhart Chart is shown in Fig. 1. The Average Chart indicates lack of statistical control at subgroups 3, 4, 9, 14, and 17. Further study of the Average Chart indicates a possible shift in the process mean at subgroup 12, and the Range Chart shows an increase at subgroup 6. Subsequent special cause investigation determined that the shift in process mean was caused by a change to a new feed hopper containing powder of a higher bulk density before subgroup 12 (the filling was controlled by volume). A line stoppage before subgroup 6 because of a shift change caused increased variability in the fill weights of the product on startup.

The purpose of SPC is to detect special causes so that they can be eliminated. In this example, two special causes were uncovered: the change in input powder bulk density and the unnecessary line stoppages.

Calculations for the Shewhart Chart

This section illustrates the calculations required to generate the control chart provided in the preceding example. In this case, the control limits were calculated from the available line data. In subsequent phases of SPC, these control limits would be applied to future data as they are measured from the process. The averages and ranges calculated for each subgroup for the current data are in Table 2.

The overall average of the $k=20$ subgroup averages and the overall average range of the subgroup ranges were calculated using the formulas $\bar{\bar{X}} = \sum_{i=1}^k \bar{X}_i/k = 6017.6/20 = 300.9$ and $\bar{R} = \sum_{i=1}^k R_i/k = 51.6/20 = 2.6$. These values were plotted as the centerlines of the Average and Range Charts, respectively.

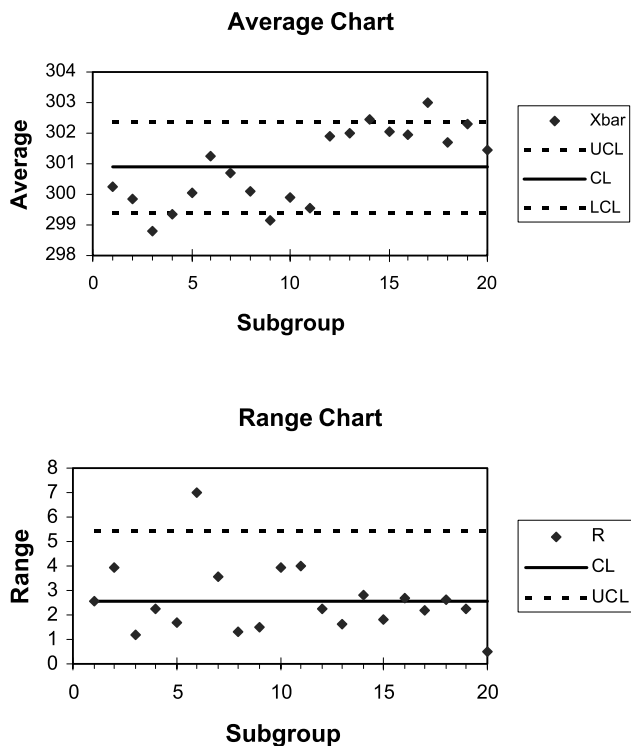


Fig. 1 Shewhart control chart for a hypothetical powder fill process. (Go to www.dekker.com to view this figure in color.)

For a subgroup size of $n=5$, the control chart factors taken from Table 1 were:

n	A_2	D_3	D_4	d_2
5	0.577	0	2.114	2.326

The control chart limits for the Average Chart were:

$$\text{LCL} = \bar{\bar{X}} - A_2\bar{R} = 300.9 - (0.577)(2.6) = 299.4$$

$$\text{UCL} = \bar{\bar{X}} + A_2\bar{R} = 300.9 + (0.577)(2.6) = 302.4$$

The control chart limits for the Range Chart were:

$$\text{LCL} = D_3\bar{R} = (0)(2.6) = 0$$

$$\text{UCL} = D_4\bar{R} = (2.114)(2.6) = 5.5$$

An estimate of the process standard deviation was $\hat{\sigma} = \bar{R}/d_2 = 2.6/2.326 = 1.1$.

Control Chart Design

A basic concept in the Shewhart Chart methodology is the division of observations into *rational subgroups* to

Table 2 Averages and ranges calculated for each subgroup for the current data

Subgroup	Data					Average	Range
1	299.7	298.7	300.2	301.3	301.2	300.2	2.6
2	301.7	297.8	299.8	301.1	298.9	299.9	3.9
3	299.3	298.3	298.2	299.0	299.2	298.8	1.2
4	297.9	299.4	299.6	300.1	299.6	299.3	2.3
5	299.7	299.6	301.3	299.9	299.8	300.1	1.7
6	299.5	302.0	300.9	302.4	299.3	301.2	7.0
7	301.7	298.4	300.5	300.9	301.9	300.7	3.5
8	299.9	299.5	300.7	299.6	300.8	300.1	1.3
9	298.6	299.2	298.5	299.6	300.0	299.2	1.5
10	300.0	299.7	302.2	298.3	299.3	299.9	3.9
11	297.4	301.4	298.7	299.3	300.8	299.5	4.0
12	300.5	300.9	300.6	298.6	298.9	301.9	2.2
13	300.7	300.3	299.1	299.8	300.1	302.0	1.6
14	300.6	300.1	299.1	301.9	300.5	302.4	2.8
15	300.1	300.8	300.9	299.4	299.1	302.0	1.8
16	301.1	298.8	298.4	300.7	300.6	301.9	2.7
17	302.2	301.4	301.3	300.1	300.0	303.0	2.2
18	300.5	300.0	298.9	298.2	300.8	301.7	2.6
19	300.4	300.6	300.2	299.0	301.2	302.3	2.3
20	299.7	299.2	299.2	299.6	299.5	301.4	0.5
Sums over subgroups						6017.6	51.6
Averages over subgroups						300.9	2.6

maximize variability between subgroups and to minimize variability within subgroups. As an example, if we are sampling four bottles from a filling line each hour, it would be better to sample them all at one time, rather than to sample one bottle every 15 min. Sampling over a longer period of time can capture sources of variability other than inherent variability.

Frequency of sampling depends on the dynamics of the process; that is, it depends on how quickly the process reacts to changes in process inputs and on the consequences (costs) of failing to react quickly to changes.^[4] Another consideration is the testing time turnaround, because a high sampling frequency is not useful if the test results are not available before the time that the next sample is taken. Typically, physical measurements may be made in real time whereas chemical measurements may require laboratory assay.

The subgroup size is based on a number of considerations: cost and ease of sampling, cost and ease of measurement, and the need to quickly detect shifts from target of a given magnitude. A subgroup size of four or five is the most common, especially for physical measurements, because it assures that the average value is normally distributed if the distribution of the individuals is reasonably symmetric. If testing is lengthy or expensive, the subgroup size can be reduced to two or three. Large subgroup sizes, such as 10 or 20, are unusual and recommended only in those situations where the control chart must be sensitive to small shifts in the process mean and the testing is rapid and inexpensive.

The chart in Table 3 indicates how quickly the X-Bar Chart would react to a shift of h standard deviations ($h\sigma$), in terms of the average number of subgroups or average run length (ARL), as a function of subgroup size n .^[1] The larger the subgroup size is, the more sensitive the chart is to a given shift.

Table 3 Average run length to detect a shift of $h\sigma$ vs. subgroup size

h	Subgroup size (n)					
	1	2	4	6	10	16
0.5	167.3	92.4	44.0	26.4	12.8	6.3
0.6	124.4	63.9	27.8	15.9	7.4	3.6
0.8	72.3	32.4	12.4	6.7	3.1	1.7
1.0	44.0	17.7	6.3	3.4	1.8	1.2
1.2	27.8	10.4	3.6	2.1	1.3	1.0
1.4	18.3	6.5	2.4	1.5	1.1	1.0
1.6	12.4	4.3	1.7	1.2	1.0	1.0
1.8	8.7	3.1	1.4	1.1	1.0	1.0
2.0	6.3	2.3	1.2	1.0	1.0	1.0
2.5	3.2	1.4	1.0	1.0	1.0	1.0
3.0	2.0	1.1	1.0	1.0	1.0	1.0

Supplemental Statistical Process Control Rules for Detecting Lack of Statistical Control

The process is in a state of statistical control if the values are within the control limits and the pattern is random. To help detect special causes, a set of supplemental rules is available from the "Statistical Quality Control Handbook,"^[5] and a subset of these rules is given below. The process is deemed "out of control" when:

- 2 of 3 points fall outside the warning limits (shift)
- 8 points in a row fall above or below the CL (shift)
- 6 points in a row are steadily increasing or decreasing (trend)
- 14 points in a row alternate up and down (two feed sources or overadjustment).

To implement these rules, it is helpful to also show on the chart the lines for ± 2 SE (warning limits) in addition to the control limits. These rules should be used with discretion; studies have shown that the use of the supplemental rules causes an increase of "false alarm signals" vs. using the ± 3 SE control limits only.^[6]

Control Charts for Multiple Processes

Many processes consist of multiple streams, such as multiple-head fillers for powder fills or multicavity molds for tablet compression. Each stream must be considered a separate process.^[7] There are two general situations:

- (1) The stream number can be identified at time of product inspection.
- (2) The stream number cannot be identified.

For case (1), a group control chart could be maintained on all streams but it suffices to plot only the two streams corresponding to the highest and lowest values. The chart limits are set up for a single stream and run rules should not be used. If the output of the heads is highly correlated, then a single chart on one stream may be used as a surrogate for all streams.

Case (2) often applies to high-speed fillers with a large number of heads. For this situation, a random sample of a subset of the heads can be taken at each time period. Once per shift (or, more frequently, if necessary), a complete set from all heads (a "head check") should be taken and these data can be examined for necessary individual head adjustments.

There are two types of special causes in multiple processes: those that affect a single head (e.g., plugging), and those that affect all heads (e.g., viscosity). The head check addresses the first kind, and the control chart of

random samples addresses the second kind. Head-to-head variation can be considered a special cause, but one that cannot be eliminated entirely.

Other Control Charts

A standard deviation chart can be used instead of the Range or *R*-Chart, where the sample standard deviation s is calculated as $s = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1)}$. The statistic s is considered to be more “efficient” than the range R because s makes use of all the data instead of the two extreme data points in the subgroup. However, the *R*-Chart is easier to use on the shop floor and is recommended for subgroups sizes up to seven, where R is at least 90% as efficient as s . Control chart factors for the s chart can be found in Ref. [3].

The *Individual—Moving Range Chart* is used when the subgroup size is one. If the process deals with bulk product in a batch or a continuous process, only single measurements may make sense because multiple measurements would reflect only the assay and sampling variability, not the process variability. Because there is no measure of variability in a single measurement, the *moving range*, or the absolute difference between two successive individual values, is calculated. The average moving range performs the function that the average range does in the Shewhart Chart. The control limits for the Individual Chart are computed using an A_2 value of 2.66. The UCL for the Moving Range Chart uses a value of D_4 equal to 3.267.

Two alternatives to the Shewhart control chart, which are more complicated to calculate but generally more effective to detect small shifts, are the Cumulative Sum (or Cusum) control chart and the Exponentially Weighted Moving Average (EWMA) control chart. These control charts will not be discussed here, but are described in standard references.^[1,2]

The P-Chart is used when the statistic plotted is the fraction defective p of n product units. The C-Chart is used when the statistic plotted is the count c of defects or occurrences of an event in time or space. These charts are termed “attribute charts” and are also described in standard references.^[1,2]

Developing and Maintaining a Statistical Process Control Program

An SPC program can be valuable when used on processes where production is steady. The best chance for successful application occurs when a team is formed, comprising the plant operators, first-line supervisors, technical support people, and plant management. The team must decide on

the most critical variables to track, and these variables should be kept to a minimum.

Ryan^[2] discusses three phases of control chart usage:

Phase I—retrospective analysis of historical data for process assessment

Phase II—process monitoring in real time to find and control special causes

Phase III—process monitoring to confirm that the process remains in statistical control.

In Phase I, if data are not already available, the first step is to set up a plan for gathering and charting the data. These charts should be kept in the operating area and are usually manually plotted. Operators are encouraged to record all process adjustments as well as to record anything that might have happened to the process. These charts may be part of the operating records. A minimum of 100 data points is recommended before control limits are calculated and placed on the charts.^[2] Because most of the subgroups are of size four to five, this amounts to approximately 20–25 subgroups. The initial plots may simply be Run Charts, where the averages are plotted. The team should be discussing potential special causes that might affect the process variable being plotted.

In Phase II, the team should immediately react to an out-of-control signal and investigate for a special cause. During this phase, the team should record any process information that might be relevant in special cause identification. This phase is complete when the process is in control for an extended period of time with infrequent signals (say, one in every 100–300 subgroups). The control limits should be recalculated if a process change has occurred. They may also be recalculated at some specified frequency, such as every 50 subgroups or annually, depending on how often the observations are taken.

During Phase III, the process is monitored for instances of special causes. When special causes cannot be entirely eliminated, an Out-of-Control Action Plan (OCAP) can be developed for routine use by operating personnel.^[2] The OCAP comprises three features: activators, checkpoints, and terminators. The *activators* define the conditions that signal when the OCAP must be activated, and the control chart usually performs this function. The *checkpoints* instruct the operator to investigate specific items as possible special causes for the out-of-control condition. Once a checkpoint has been identified as a special cause, a *terminator* calls for a specific action to be applied to resolve the problem.

The purpose of SPC is to reduce variability and to improve the quality of process output. Such programs derive benefits through reduced costs of manufacture, better quality, fewer troubleshooting crises, and improved



relationships with customers. In a strict sense, SPC is a tool for getting a process into statistical control, not for making process adjustments. For making process adjustments, while data continue to be charted, the centerline is usually set at the target value and process adjustment schemes are needed. Without these schemes, operators will tend to over adjust. One reference that discusses such manual process adjustment schemes in detail is Ref. [4]. Two summary articles on the same topic can be found in Refs. [8] and [9].

PROCESS CAPABILITY

Defining Process Capability Indices

As stated earlier, a process capability analysis assesses the process performance relative to product specifications. Specification limits are denoted as LSL for the lower specification limit, and as USL for the upper specification limit. In some cases, there is only a single specification limit, either a lower limit or an upper limit.

Two *process capability indices* are in common use, and they are as follows:

- (1) The process-centered index $C_P = (USL - LSL) / 6\sigma_C$
- (2) The process-shifted index $C_{PK} = \text{MIN}[(USL - \mu) / 3\sigma_C, (\mu - LSL) / 3\sigma_C]$

where μ is the process average and σ_C is the standard deviation measuring the inherent process variability. The C_P index can be used only when there are two-sided specifications. These indices are defined so that higher is better.

Figs. 2–4 illustrate the meaning of the C_P index for the following situations: $C_P < 1$, $C_P = 1$, and $C_P > 1$. The three situations can be described as follows:

$C_P < 1$: The $\mu \pm 3\sigma_C$ range of this process distribution (the area under the curve) is wider than the specification limits, so that $C_P < 1$ (see Fig. 2). A large percentage of

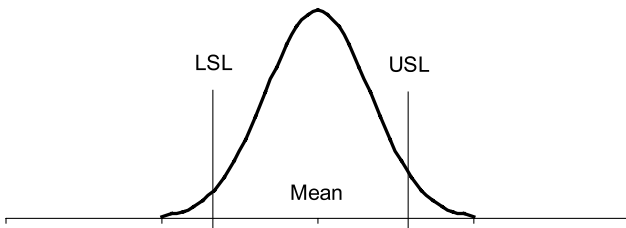


Fig. 2 Comparison of process output vs. specifications for $C_P < 1$.

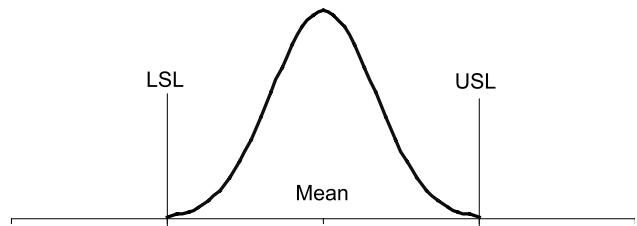


Fig. 3 Comparison of process output vs. specifications for $C_P = 1$.

rejects is produced even when the process is centered and in control. The defects are depicted as the area under the curve outside the specification limits, and the process is clearly not capable of producing a product that consistently meets specifications.

$C_P = 1$: For this process, the $\mu \pm 3\sigma_C$ range coincides with the specification limits; therefore $C_P = 1$ (see Fig. 3). Assuming that the process is centered and in control, the defect rate is 0.27% if the process output is normally distributed.

$C_P > 1$: For this process, the $\mu \pm 3\sigma_C$ range is much tighter than the specification limits, so that $C_P > 1$ (see Fig. 4). The process has room for slight shift and drift without producing defects.

If a process is not centered, than a more meaningful process capability index is C_{PK} . This is shown in Fig. 5.

$C_{PK} < 1$: For this process, $C_P > 1$, but the process is not centered; therefore defects are produced. For this reason, $C_{PK} < 1$ better reflects this situation (see Fig. 5). Another definition of C_{PK} that might be easier to understand is $C_{PK} = |\mu - \text{nearest spec limit}| / 3\sigma_C$.

The relationship between C_P and percent defective (assuming a normal distribution) is given in Table 4.

For these indices to be meaningful as descriptors of process capability, the process must be in a near state of statistical control. A perfect state of statistical control is rarely attainable. A criterion for near-statistical control is that out-of-control signals appear about once in every

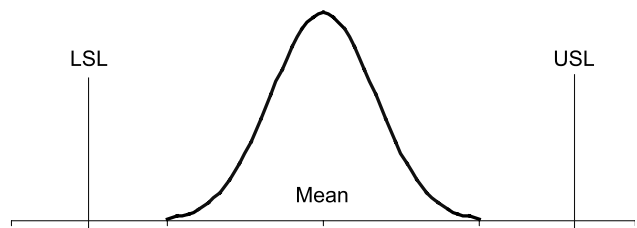


Fig. 4 Comparison of process output vs. specifications for $C_P > 1$.

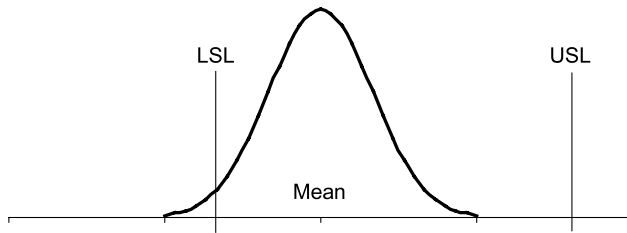


Fig. 5 Comparison of process output vs. specifications for $C_{PK} < 1$.

100–300 subgroups. This is usually the case in Phase III of the SPC program. Until that point is reached, the process capability indices should reflect only the process potential should the special cause variation be eliminated. Some of the many cautions associated with the use of capability indices are described throughout this entry and detailed in Refs. [10] and [11].

Defining Process Performance Indices

Sometimes an analysis of process capability is desired when the process is not in a state of strict statistical control. When this occurs, process performance indices should be used. Process performance indices can be defined as: $P_P = (USL - LSL)/6\sigma_P$, and $P_{PK} = \min[(USL - \mu)/3\sigma_P, (\mu - LSL)/3\sigma_P]$, where σ_P is now the overall standard deviation of the process output characteristic.

For a process that is not in a state of strict statistical control, σ_P is greater than σ_C , the inherent variation, because there is additional variation because of process shift and drift. This additional variation component is σ_S , where the “S” in σ_S denotes special cause variation. Because variances, not standard deviations, are additive, these components can be represented as $\sigma_P^2 = \sigma_C^2 + \sigma_S^2$. Another way to state the objective of SPC is to reduce σ_S as close to zero as possible. One criterion used to assess whether a process is near statistical control is the closeness of the ratio σ_P/σ_C to 1. Table 5 shows the relationship between this ratio and the percentage of total variation represented by special cause variation.

Table 4 Relationship between C_P and percent defective

C_P	% Def	C_P	% Def
0.6	7.19	1.0	0.27
0.7	3.57	1.1	0.097
0.8	1.64	1.2	0.032
0.9	0.69	1.3	0.001

Table 5 Relationship between σ_P/σ_C and the percentage of total variation represented by special cause variation

σ_P/σ_C	σ_S^2/σ_P^2 (%)
1.30	41
1.25	36
1.20	31
1.15	24
1.10	17
1.05	9
1.00	0

For example, when $\sigma_P/\sigma_C = 1.2$, the percentage of special cause variation is 31%, which may represent a practical goal for attaining “a near state of statistical control.”

Estimating Process Capability Indices

To estimate process capability indices from data, the estimates replace the true process values in the following formulas as follows:

The process-centered index

$$\hat{C}_P = (USL - LSL)/6\hat{\sigma}_C$$

The process-shifted index

$$\hat{C}_{PK} = \min[(USL - \bar{X})/3\hat{\sigma}_C, (\bar{X} - LSL)/3\hat{\sigma}_C]$$

where \bar{X} is the estimated process average and $\hat{\sigma}_C$ is the estimated standard deviation measuring the inherent process variability. The estimates of \bar{X} and $\hat{\sigma}_C$ can come from a prior control chart analysis, where $\hat{\sigma}_C = \bar{R}/d_2$.

Use of the \hat{C}_P estimate assumes that the process is centered at or near the exact middle of the specification range of a two-sided specification. This represents the best capability achievable. The \hat{C}_{PK} estimate assesses the process with no assumption that the process is centered, and this estimate is always equal to or less than the corresponding \hat{C}_P estimate. If the distribution of the output is close to a normal distribution and the capability index is based on a large and representative sample size, then acceptance values can be established. Some industries define a process with $\hat{C}_{PK} = 1$ as barely capable and one with $\hat{C}_{PK} = 1.33$ or higher as capable. (Note that a $\hat{C}_{PK} = 1$ represents a situation where the process is centered and the specification range is equal to $6\hat{\sigma}_C$.)



Estimating Process Performance Indices

The estimates of the performance indices are defined analogously to the capability indices as $\hat{P}_P = (USL - LSL)/6\hat{\sigma}_P$ and $\hat{P}_{PK} = \text{MIN}[(USL - \bar{X})/3\hat{\sigma}_P, (\bar{X} - LSL)/3\hat{\sigma}_P]$, using the estimate $\hat{\sigma}_P$, which is the estimated overall standard deviation of the process output characteristic. Thus $\hat{\sigma}_P$, the estimate of σ_P , includes not only the estimated inherent deviation $\hat{\sigma}_C$, but also the estimated variation due to process shift and drift from special causes. Therefore $\hat{\sigma}_P > \hat{\sigma}_C$ for a process not in statistical control.

Process performance indices strictly apply only to the data set being examined because the process is not predictable into the future when it is not in statistical control. However, in practice, the indices are often interpreted to provide an assessment of the process as it currently exists.

If the process is not in a state of strict statistical control (i.e., stable and predictable), every attempt should be made, when practical, to improve control and to put the process into as close to a state of statistical control as possible. If the process cannot be put into statistical control, enough time points should be collected to be able to assess the full degree of shifts and drifts of the process and to determine whether the data are indicative of the long-term process with all of its shifts and drifts. As Pruett and Schneider^[12] indicate, P_{PK} is a measure of process capability over a period of time. If extensive data can be obtained, the ability of the process to meet its specifications can be evaluated, with care, by the estimate \hat{P}_{PK} . If extensive data cannot be obtained, \hat{P}_{PK} can provide only the capability of that particular set of data to meet the specifications.

Before the special causes are fully identified and controlled, \hat{P}_{PK} estimates the process performance that the customer currently experiences, and \hat{C}_{PK} estimates the potential process capability attainable when the process is brought more closely into a state of statistical control. As the process approaches a state of statistical control, \hat{P}_{PK} approaches \hat{C}_{PK} . Similarly, the estimates \hat{P}_P and \hat{C}_P are used when the process is centered within two-sided

Table 7 Sample capability indices that provide at least a 95% confidence that the true process capability is at least 1.00 or 1.33 based on various sample sizes

<i>n</i>	$C_P=1.00$	$C_P=1.33$
10	1.65	2.19
30	1.28	1.70
100	1.13	1.51
200	1.09	1.45

specification limits, or to reflect what the capability would be if the process were centered.

Precision of Estimates of Capability and Performance Indices

It is important to remember that the capability and performance indices calculated from the data are only estimates and thus have uncertainty associated with them. Confidence intervals can be calculated for these indices to reflect this imprecision.^[11] For example, a $100(1-\alpha)\%$ two-sided confidence interval about the estimated process capability index \hat{C}_P is:

$$\hat{C}_P \sqrt{\frac{\chi_{1-\alpha/2, n-1}^2}{n-1}} \leq C_P \leq \hat{C}_P \sqrt{\frac{\chi_{\alpha/2, n-1}^2}{n-1}}$$

where $\chi_{1-\alpha/2, n-1}^2$ and $\chi_{\alpha/2, n-1}^2$ are the lower and upper $\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom.

This interval is constructed to contain the true unknown index C_P with a specified degree of confidence $100(1-\alpha)\%$. For example, 95% two-sided confidence intervals for two different sample values of \hat{C}_P for different sample sizes are as shown in Table 6.

As can be seen from the table, if a sample \hat{C}_P of 1.00 is obtained from a sample of 30 values, then with 95% confidence, it can be stated that the true process capability C_P is between 0.74 and 1.26. If, instead, one is only interested in one direction (e.g., how low the true C_P index might be based on a sample of size n), a one-sided lower confidence interval could be generated.

If, on the other hand, one is interested in determining the minimum acceptable sample process capability index \hat{C}_P required to have some stated confidence $100(1-\alpha)\%$ that the true C_P is above some value (e.g., $C_P=1.0$), one can use the following equation for a particular C_P , sample size n , and risk level α : the minimum \hat{C}_P is $C_P / \sqrt{\chi_{1-\alpha, n-1}^2 / n - 1}$. This can be useful for determining sample sizes during protocol generation where a particular

Table 6 The 95% two-sided confidence intervals for two different sample values of \hat{C}_P for different sample sizes

<i>n</i>	$\hat{C}_P=1.00$		$\hat{C}_P=1.33$	
	Lower	Upper	Lower	Upper
10	0.55	1.45	0.73	1.93
30	0.74	1.26	0.99	1.67
100	0.86	1.14	1.14	1.51
200	0.90	1.10	1.20	1.46

minimum C_P is required for protocol acceptance. For example, the sample capability indices that would be needed to have at least a 95% confidence that the true process capability is at least 1.00 (or 1.33) based on various sample sizes are shown in Table 7.

Confidence intervals for any of the other indices of interest (P_P , C_{PK} , or P_{PK}) have sometimes been estimated in practice by using the above equations but by substituting the index of interest. However, these intervals should only be considered as approximate. More statistically based approximations for these intervals are discussed in Refs. [13] and [14]. One often used approximate confidence interval for C_{PK} assumes that the distribution of \hat{C}_{PK} is normal and a Taylor series approximation is used to construct a $100(1-\alpha)\%$ two-sided confidence interval for C_{PK} as follows:

$$\hat{C}_{PK} \pm Z_{\alpha/2} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{PK}^2}{2n-2}}$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution. For example, to generate a 95% two-sided confidence interval, use $Z_{\alpha/2}=1.96$.

As can be seen from the above examples, reported estimates of capability and performance indices based on small sample sizes are potentially misleading and should not be used unless relatively large samples are obtained, or they are accompanied by a confidence interval when reporting the results.

Nonnormal Distributions

The use of the 6σ factor in defining the process capability indices assumes that the data follow a normal distribution. It is important to remember that large values of these capability indices (indicating a very capable process) may be derived from incapable processes when the data are nonnormal. Similarly, data from a quite capable process may result in low capability index values just because the data are nonnormal. Normality of the data can be checked in a variety of ways. Routine plotting of histograms is a good first step in assessing whether the data are normal. Some of the more detailed techniques for determining whether a set of data is normal are: probability plotting, the chi-square goodness of fit test, and the Anderson–Darling statistic.^[15] If it is determined that the data are not normal, four options are available.^[16]

- Apply a transformation to the data to make the transformed data normal. If the distribution is skewed to the right, one might try a log, inverse, square root,

or cube root transformation of the data to make the data normal. If the data are skewed to the left, an exponential, squared, or cubed transformation might be applied. A histogram can be applied before and after the transformation to assess the ability of the transformation to make the data normal. It is important to remember that the transformation must be applied to the USL and LSL, in addition to the data, before computing the capability index of interest.

- Assume a known distribution and use the multipliers that correspond to the 2.5% and 97.5% points for the known nonnormal distribution (to obtain 95% confidence). One common distribution that is effective in fitting a broad range of distributional forms is the gamma distribution.
- Use the index “as is” as a relative measure and note in report to this effect. Even if the data are nonnormal but symmetrical, the calculations are generally reasonable. However, if the data distribution is heavily skewed, one should be hesitant to use this particular option.
- Omit the computation of capability and performance indices because the values so obtained would be misleading.

Process Capability Example

This example studied the process capability of capsule weight for a hypothetical prescription drug. The specification limits were 290–350 mg. Four capsules were sampled from the process every 20 min and weighed. There were 25 subgroups of four data points each for a total of 100 data points. The results are summarized in Fig. 6.

The control chart analysis indicated that the process was in a state of statistical control because none of the averages and ranges was outside its respective control limits. The data distribution (histogram) was deemed reasonably close to a normal distribution (the bell-shaped curve superimposed on the histogram) because the cumulative data distribution was close to a straight line on the normal probability plot.

In this study, the variability estimates were $\hat{\sigma}_P=6.89$ mg and $\hat{\sigma}_C=6.60$ mg, both in close agreement because the process was in statistical control. The \hat{C}_{PK} estimate was 1.32, which was lower than the \hat{C}_P estimate of 1.52, because the process was not exactly centered between the specification limits. The process performance estimates did not apply because the process was in statistical control, but were calculated by the package and were close to the capability estimates. (Note: In the Minitab output, the estimates do not have the caret (^) over the symbol, but they are in fact estimates.)

If the process were not in a state of statistical control (stable and predictable), then the \hat{P}_P and \hat{P}_{PK} estimates



Capsule Weights

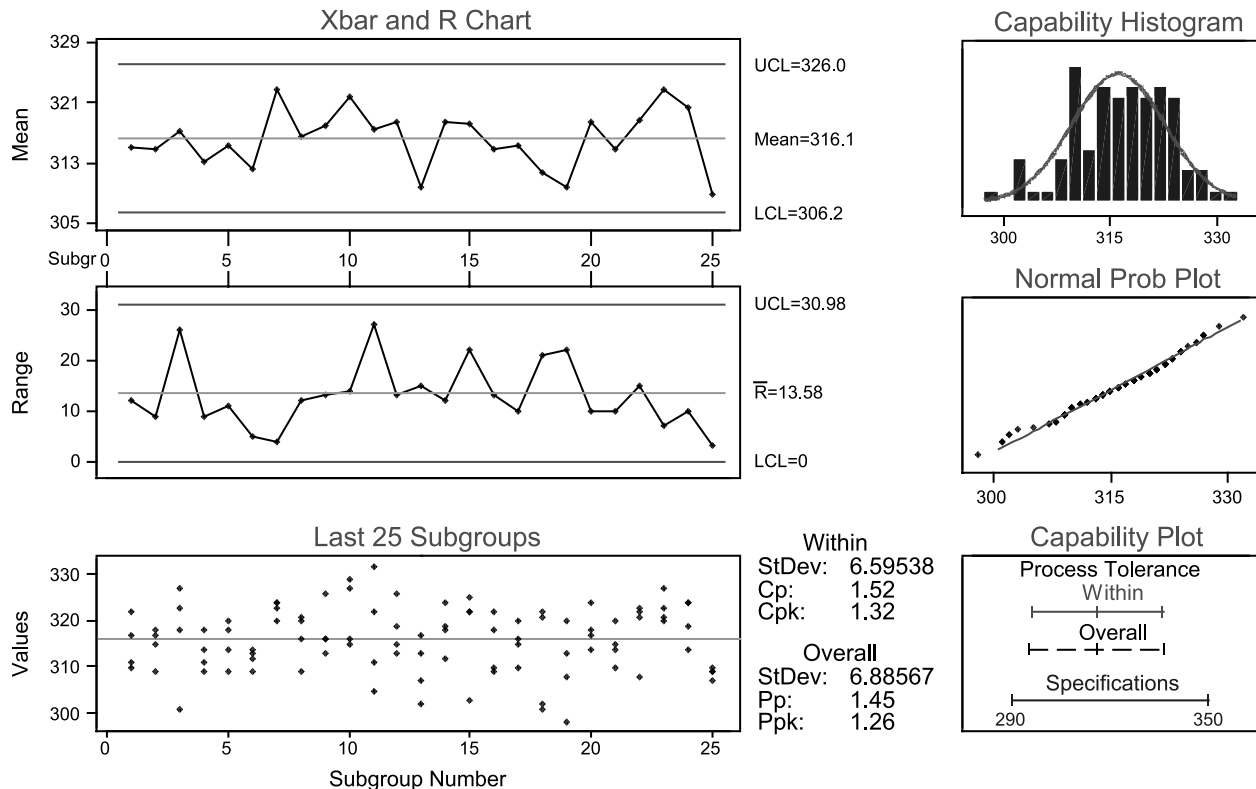


Fig. 6 Summary of capsule weight data using the statistical package, Minitab®. (From Minitab Inc., State College, PA. For details, refer to www.minitab.com.) (Go to www.dekker.com to view this figure in color.)

could be used to judge the process performance at the time the data were collected. The \hat{C}_P and \hat{C}_{PK} indices could be used to assess the process potential if the process were brought into a state of statistical control.

Relationship Between Statistical Control and Process Capability

To summarize, it is important to remember that a process may be in statistical control but still not capable. In a strict sense, it is not appropriate to talk about whether a process is capable if it is not in a state of strict statistical control (because it is not stable and predictive of the future); in practice, the indices are often used to provide an assessment of the process as it currently exists. Caution has been given to the reader to make sure that before an assessment is made of a process that is not in statistical control, enough data have been collected to cover the full range of shifts and drifts in the process. With this information in hand, it is possible to claim, with some

caution, that a process is capable of staying within specifications although it is not in statistical control. However, even though the process may be capable, there are still many benefits in working toward the elimination of all special cause variation.

APPLICATION TO PHARMACEUTICALS

Typical Applications

The concepts discussed in this paper are appropriate for many applications in pharmaceuticals. Some of these areas are as follows:

- Drug potency
- Other finished product attributes (e.g., content uniformity, dissolution)
- Tablet/capsule in-process characteristics (weight, thickness, hardness, disintegration, etc.)

- Powder characteristics (mean particle size, bulk density, etc.)
- Microbial counts
- Drug content application (nasal spray, etc.)
- Fill weight and fill volume
- Air/oxygen-purged systems
- Leachables, residuals
- Liquid characteristics (viscosity, refractive index, etc.)
- Consumer complaints, industrial safety measurements, etc.

Bringing processes into or near a state of statistical control will improve processes by making them less variable centered closer to target, and allow the manufacturer to make a product that will more consistently meet product specifications. This benefits both the manufacturer and the consumers who use their products. The use of SPC methods to evaluate and to improve processes not only can be applied to product characteristics such as tablet weight and tablet hardness, but also to product performance measures such as consumer complaints, line down time, and industrial safety measurements. An SPC approach to process improvement can also lead to reductions in fill overages, reductions in waste, as well as reductions in batch failures. By eliminating special cause variability, it becomes easier to monitor a process to ensure that new special causes do not find their way into the process.

As mentioned earlier, as part of a process capability study, an assessment of variability is conducted. Variability may include both process and measurement variability. This may be especially important when measuring such attributes as potency, content uniformity, and dissolution. It is important to be able to separate process from measurement variability during any investigation of how to improve processes. The samples may need to be tested in a designed fashion to be able to correctly separate these sources of variability. Measurement variability can be independently addressed through assay validation studies.

Analysis of In-Process Data

Control charts are an excellent analysis tool to both monitor and improve in-process performance during process development and later during production, where it is desired to follow process characteristics over time within batches or runs. The most common examples of tablet process characteristics that are measured in-process are weight, thickness, and hardness. The parameters measured need to be controllable so that adjustments can be made. During the initial runs, it is desirable to limit process adjustments to a minimum to observe the process

in its natural state. Any adjustments made should be recorded and explained. Out-of-limit results should never be removed prior to performing a process capability analysis. If special cause variation is detected, then process improvements should be made to eliminate the special cause variation.

For new products, capability should be evaluated and improvements made to the process during development and scaleup, before validation is performed. It should be conducted on the actual full-scale equipment that will be used during routine production. Our experience has been that when evaluating in-process tablet and capsule parameters during production, a good setup is important to ensure that the process starts out at the target value. Additionally, the practice of 1) shutting down the line for breaks and lunch and 2) allowing hopper runout during processing can cause added variability because of segregation of the powder.

Product flow characteristics and interdependencies between parameters, such as tablet weight, thickness, and hardness, may prevent the parameters from each being in strict statistical control. Special studies testing weight, thickness, and hardness on the same tablets will aid in better understanding this interrelationship between these product characteristics. For such characteristics, the in-process specifications may be on either the individual values or the averages (e.g., the average of 10 hardness values), or both. When the average is being evaluated, process capability indices may not be appropriate for evaluating the ability to meet specifications based on individuals.

For example, some process monitoring schemes use an average of 10 tablets to track tablet weight. The 10 tablets are sampled and measured in a single weighing operation at each sampling time, and the average weight is computed and plotted. In this situation, the state of statistical control of tablet weight can be assessed using Individual Moving Range Charts on the tablet averages, but such data may not be applicable for process capability studies when the specification is based on individual tablets.

It is important to remember that a process may be in statistical control but still not capable. Range studies for the critical parameters are an effective way to determine whether the specification limits are tighter than they need to be, or, if no specifications exist, to aid in their development. A range study is intended to determine the high and low ranges for which the parameter is able to demonstrate an ability to meet end-product requirements; for example, high and low hardness ranges that meet both dissolution as well as physicals (e.g., no chipping) help to define appropriate in-process specification limits. If the process is not capable, this may be



because the specifications had originally been set at $\pm 3\sigma$ limits and not what defines acceptable product, as is the purpose of a range study. If this is what was done, capability indices of around 1.0 must be expected.

Analysis of End-Product Data (e.g., for Validation, Annual Product Review, etc.)

Capability and performance indices are appealing as a management process assessment tool because they tie together the process performance (as measured by the sample mean and sample standard deviation) with the product specifications into a single measure. Process performance indices are sometimes used as acceptance criteria during validation, as well as Annual Product Review; however, they need to be used very carefully. Sample sizes may be too small to provide reliable estimates, especially if these indices are used to determine the pass/fail of a validation study or an Annual Product Review. Emphasizing the use of capability indices without the efforts required to initially put the processes into statistical control misses the point. Although not recommended by the authors as strict acceptance criteria for validation or Annual Product Review, capability indices, if used carefully, do provide a useful yardstick for comparison of performance over time periods, or for comparison between products for helping in determining priorities for process improvement efforts. These indices also provide a yardstick to measure the effect of any improvements made to the process in a 'language' that can be easily understood. There are also times when there are just better ways to define process capability for solid dosage forms, possibly by tying it to the end-product test requirements [e.g., the ability to meet the multiple-stage United States Pharmacopeia (USP) tests for content uniformity and dissolution].

The use of common pharmaceutical ingredients and the tendency to "gang test" assays in the laboratory may prevent many of our end-product attributes from being in strict statistical control when measured over time. Therefore it is not recommended that a validation study or Annual Product Review require that a process be in statistical control to be considered acceptable, not even for in-process parameters such as tablet weight, thickness, and hardness. Processes that are not in a state of strict statistical control can be capable of consistently meeting specifications and can be validated. However, if processes are not in statistical control, efforts should always be made to eliminate special causes and get them into as near a state of statistical control as possible. The validation and Annual Product Review data can even be helpful in determining how processes can be improved.

Although these comments have been applied to end-product data, many of these same comments also apply to in-process data as well.

Limitations Within Pharmaceuticals

As discussed earlier, processes in the pharmaceutical industry may be difficult to always get into a state of strict statistical control. Interdependencies may exist between attributes such as the weight, thickness, and hardness of in-process tablets because of product flow characteristics within the batch. Data from sequentially generated batches may be dependent because of the use of common raw materials, or being tested on the same day. But the benefits derived from using the techniques discussed in this entry to improve existing processes and to keep them under control are many. The effort is worth the hard work required to do it.

Nash^[17] points out that what might be possible in other industries may not always be possible for pharmaceutical processes in that a C_{PK} of 1.33 may not always be achievable. Specifications should be customer-directed limits that ensure safety and efficacy, not limits to control the consistency of process. Internal control limits should be used for this purpose. Many of our specifications are either provided to us by the USP or influenced by the Food and Drug Administration (FDA), where industry is encouraged to set specifications as tight as the process will allow. Companies are sometimes asked to agree as to what they view as tight specifications before the full experience that comes with additional production has occurred. In addition, once stability changes over time, batch processing differences, slight biases in the analytical method, and typical measurement and process variability are accounted for, defective levels may not always be in the parts-per-million range as is required for a C_{PK} of 1.33. In fact, if measurement error is large, arbitrary C_{PK} goals may actually be impossible to meet no matter how much the process is improved.^[11]

Because, in the pharmaceutical industry, being the first to market a new drug is of utmost importance, taking the time to put processes into statistical control is too often perceived by companies as slowing down the process of product introduction and not as added value. This attitude needs to change. Companies may also feel that their reward for improving process control and reducing variability is that the FDA will tighten specifications further, thereby risking the production of out-of-specification product. A company's interest in improving processes is often discouraged by a need to revalidate or refile with the FDA whenever significant changes are made to the filed NDA. The authors feel that there needs to be discussions between industry and the FDA on these

issues so that industry is rewarded for the improvement of their processes and everyone, most importantly the consumer, wins. This need for cooperation between the FDA and industry to move the pharmaceutical industry toward real improvements was voiced by Woodcock of the FDA in Ref. [18].

CONCLUSION

The use of SPC to help improve processes and the use of process capability to evaluate the ability of processes to routinely meet specifications are described in detail. They are companion tools that should be used in concert. These tools are described in sufficient detail to allow interested practitioners to correctly apply them within their workplace. Examples of their use are provided for reinforcement as are the assumptions being made when using them. Some of the limitations associated with their use are also discussed, especially in the context of application within the pharmaceutical industry.

With the writing of this entry, the authors hope to encourage the proper use of these important process improvement and process evaluation tools, especially within the pharmaceutical industry. We also hope that some of the barriers to their implementation within the pharmaceutical industry that were discussed within this entry can be overcome through cooperation with the FDA and a mutual desire to improve pharmaceutical processes.

REFERENCES

1. Montgomery, D.C. *Introduction to Statistical Quality Control*, 4th Ed.; Wiley, 2001.
2. Ryan, T.P. *Statistical Methods for Quality Improvement*, 2th Ed.; Wiley, 2000.
3. *Manual on Presentation of Data and Control Chart Analysis*, 7th Ed.; ASTM International, 2002.
4. Box, G.E.P.; Luceño, A. *Statistical Control in Monitoring and Feedback Adjustment*; Wiley, 1997.
5. *Statistical Quality Control Handbook*; Western Electric Company, 1956.
6. Champ, C.W.; Woodall, W.H. Exact results for Shewhart control charts with supplementary runs rules. *Technometrics* **1987**, 29 (4), 393–399.
7. Nelson, L.S. Control chart for multiple stream processes. *J. Qual. Technol.* **1986**, 18 (4), 255–256.
8. Box, G.E. George's column. *Qual. Eng.* **1991–1992**, 4 (1), 143–151.
9. Box, G.E. George's column. *Qual. Eng.* **1991–1992**, 4 (2), 331–338.
10. Kane, K. Process capability indices. *J. Qual. Technol.* **1986**, 18 (1), 41–52.
11. Gunter, B. The use and abuse of C_{pk} , Part 4. *Qual. Prog.* **1989**, 22 (7), 86–87.
12. Pruett, J.M.; Schneider, H. *Essentials of SPC in the Process Industries*, 2nd Ed.; Instrument Society of America, 1996.
13. Rodriguez, R. Recent developments in process capability analysis. *J. Qual. Technol.* **1992**, 24 (4), 176–187.
14. Kushler, R.H.; Hurley, P. Confidence bounds for capability indices. *J. Qual. Technol.* **1992**, 24 (4), 188–195.
15. Shapiro, S. *How to Test Normality and Other Distributional Assumptions*; How To Series, ASQ 1990; Vol. 3.
16. *Quality Assurance for the Chemical and Process Industries, a Manual of Good Practices*, 2nd Ed.; ASQ Quality Press, 1999; 42.
17. Nash, R.A. Understanding the process capability index concept. *J. Valid. Technol.* **1998**, 4 (3), 202–204.
18. Regulatory Leeway sought for process analytical technology. *Gold Sheet* February **2002**, 36 (2), 1–19.



Sterilization by Dry Heat

Andrea Chieppo

Thomas Kupiec

Analytical Research Laboratories, Oklahoma City, Oklahoma, U.S.A.

INTRODUCTION

Sterilization means the destruction of all life. The aim of sterilization is to destroy the ability of microorganisms to survive and multiply with the oldest and most recognized agent of destruction being heat.^[1] Dry heat is the method of choice for sterilizing heat-stable items that might not be adequately penetrated by steam or are damaged by moisture.^[2] Dry-heat sterilization is frequently used in the pharmaceutical industry; items often sterilized by dry heat include powders, oils,^[3] petrolatum, glassware, and stainless-steel equipment.^[2]

Dry-heat sterilization is generally a less complicated process than steam sterilization; it is, however, relatively slow and requires higher temperatures and/or longer exposure times. This is because of the fact that microbial lethality is lower with dry heat than that for steam at the same temperature.^[2] There are various temperatures and periods of treatment for dry heat depending on the pharmacopoeia. The U.S. Pharmacopoeia (USP) states that the dry-heat sterilization process for containers for sterile pharmaceutical products should be at a temperature of 160–170°C for a period of 2 to 4 hr. The British Pharmacopoeia states that items sterilized by dry heat should be kept at a temperature not less than 160°C for at least 1 hr. For the Pharmacopoeia Nordica, the recommendation is 30 min at 180°C. Different materials and sterilization equipment used account for the discrepancies between these pharmacopoeias, but there is also a lack of sufficient information concerning dry-heat sterilization.^[4]

Dry-heat processes have two main targets: microorganisms and their by-products, pyrogens or endotoxins. Depyrogenation is the process that destroys the chemical activity of these by-products. Destruction of microorganisms and endotoxins by dry heat is considered an oxidative process, which is almost a combustion.^[5] Depyrogenation requires a higher temperature than sterilization and can be summarized as follows:

If an effective dry heat depyrogenation is performed, sterilization generally is achieved as well. Effective dry heat sterilization can be performed even without achieving depyrogenation. If moist heat sterilization is performed, in normal operating conditions depyrogenation is not achieved.^[5]

DRY-HEAT STERILIZATION KINETICS

Sterilization is a probability function. In the pharmaceutical industry, an item is deemed sterile if there is less than 1 chance in 1,000,000 that viable microorganisms are present in the sterilized article or dosage form. Therefore there is a 10^{-6} probability of nonsterility.^[6] In all sterilization processes, the inactivation of microorganisms develops as a first-order chemical reaction, i.e., at a rate which is approximately logarithmic.^[7]

The following equation typically describes the order of microbial death:

$$K = 1/t(\log N_0 - \log N)$$

where K =a constant, assuming logarithms to the base 10 and depending on the organism, temperature, and substrate, t =time of exposure in min, N_0 =initial number of viable organisms, and N =number of viable organisms at the end of the time interval.^[7]

Using natural logarithms, the exponential form of the equation is:

$$N = N_0 e^{-Kt}$$

It was found that a 90% reduction in the microbial population resulted in the following equation for K :

$$K = 1/t(\log N_0 - \log 0.1N_0)$$

$$K = 1/t(1) = 1/t$$

$$t = 1/K$$

Time t is defined as the decimal reduction time or D value. Therefore

$$D = 1/K \text{ or } D = 1/10$$

Fig. 1 shows a simplified way to determine the D value where the amount of time can be determined to reduce the microbial population by 90%.^[7]

The kinetics of dry-heat treatments is comparable to that of moist heat sterilization. The organisms that are considered to be representatives for dry-heat sterilization processes are spores of *Bacillus subtilis* var. *niger*.^[4] A D_{170} value, which is the time needed to give a 90% reduction of the spore population at 170°C, of at least 1.5 min may be assumed for these spores.^[6]

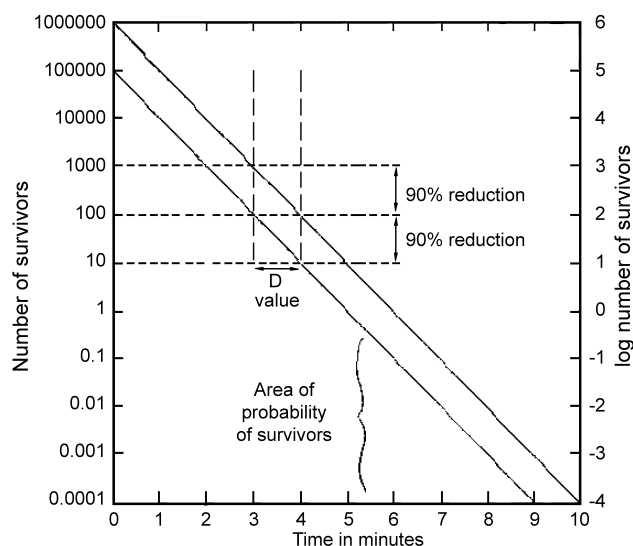


Fig. 1 Microbial death rate curves that illustrate concept of decimal reduction (D values) and probability of survivors. (From Ref. [6].)

F_H Concept

For dry-heat temperatures other than 170°C , F_H values are used. The F_H concept is comparable to the F_O concept for moist heat sterilization and references lethality to equivalent times at 170°C . F_H values are shown in units of minutes or seconds, and the calculations of F_H use the same equations as the calculations of F_O . The only difference is that a z value of 20°C is substituted for 10°C .^[6]

DRY-HEAT STERILIZATION PROCESSES

Great strides have been made in developing and improving other methods of sterilization, but little has been accomplished where dry-heat sterilization is concerned. Sterilizers used today are still basically the same as they were a decade ago.^[1]

The two types of processes currently used in dry-heat sterilization include: 1) dry-heat batch sterilization/oven sterilization and 2) dry-heat tunnel sterilization. Process 1 is the type of dry heat unit widely used in the pharmaceutical industry; it uses the principle of convective heat transfer to heat the load. Process 2 is only found in large-scale processes, and the main application of this process is in the sterilization and depyrogenation of glass.^[6]

Convection Heating Processes

In the pharmaceutical industry, large-scale industrial ovens are used. These large-scale processes usually operate to sterilize only because higher temperatures are

required for depyrogenation and this would be prohibitively long.^[6] Convection heating occurs when heat is transferred through a medium by motion of its parts. The two types of convection heating include natural and forced convection heating. Natural convection heating happens as a result of the buoyancy forces generated by differences in density. In forced convection heating, high-efficiency particulate air (HEPA)-filtered air is heated by passage over the electric heating elements of the oven and then this heat is transferred from the air to the product by forced convection.^[2] The basic equation for convective heat transfer is:

$$q/c = \bar{h}_c A \Delta T$$

where q/c = the rate of heat transfer by convection (BTU/hr), \bar{h}_c = the average unit thermal convective conductance (BTU/hr ft^2 $^\circ\text{F}$), A = area (ft^2), and ΔT = the difference

The main features shown in the sketch are

1. Air-circulation fan
2. Water-cooled battery (for the cooling phase)
3. Circulation HEPA filters
4. Launch/recovery bulkheads
5. Trolley and load
6. Discharge duct
7. HEPA filter on the discharge duct to prevent back-flow contamination
8. Variable-speed fan for chamber pressurization (proportionally controlled)
9. Prefilter and HEPA filter on the chamber pressurization loop
10. Electric heater (proportionally controlled)
11. Four flexible Pt100 4-wire RTDs
12. Main control Pt100 4-wire RTD
13. Pressure transducer

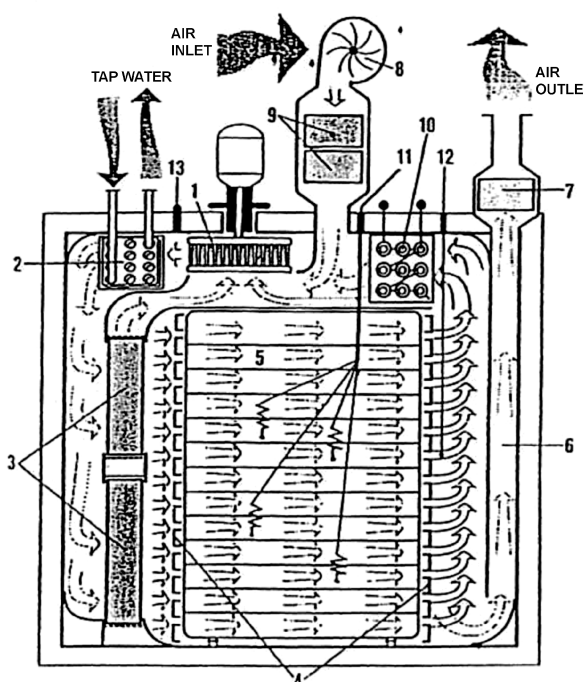


Fig. 2 Dry-heat batch sterilizer. (From Ref. [4].)

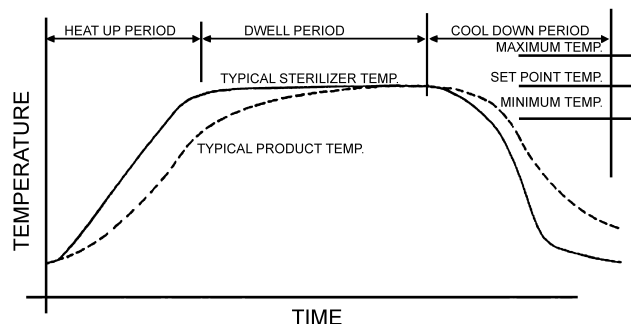


Fig. 3 Example of a batch process. (From Ref. [2].)

between the surface temperature and the temperature of the fluid at some specified location ($^{\circ}\text{F}$).^[2]

Fig. 2 shows a schematic diagram of a modern forced-convection batch sterilizer. The unit is a two-door sterilizer where the unloading door leads to the sterile area. The pressure inside the chamber must continuously be controlled so that the inside is always slightly higher than the pressure in the nonsterile loading area and slightly lower than the pressure in the sterile unloading area.^[5]

There are four stages involved in oven sterilization. They are 1) drying, 2) heat-up, 3) exposure, and 4) cool-down. During the drying stage, moisture is driven off the product to the atmosphere until the air temperature in the sterilizer is around 80°C . At this point, a baffle closes which allows the temperature within the oven to reach the operating level; this is the heat-up phase. Exposure starts the moment the sterilization thermal sensor reaches the set temperature. After the timed exposure period, the heating elements switch off and cooldown begins. Fig. 3 shows a normal batch process.^[6]

Dry-Heat Tunnel Sterilization

Dry-heat tunnel sterilization is a continuous conduction process in contrast to the batch processes in ovens. A continuous conduction process is one in which a predetermined quantity of items is subject to a continuous conduction cycle at a predetermined rate to effect sterilization and/or depyrogenation.^[2]

Conduction heating occurs via two different mechanisms. The first mechanism is accomplished by molecular interaction where molecules at a higher energy level impart energy to molecules at a lower energy level.^[2] The second mechanism takes place through “free” electrons. The ability of different solids to conduct heat varies directly with the free electron concentration. The solids with the highest amount of free electrons are pure metallic solids, and those solids with the lowest concentration are the nonmetals; therefore pure metals are the best conductors of heat.^[2]

Dry-heat tunnels differ from ovens because energy is not lost to heating and reheating between cycles as occurs with ovens. For a majority of the time spent in the tunnel, the product is subjected to rising temperatures, unlike ovens where the product is held at a constantly maintained exposure temperature.^[6] Tunnels basically consist of a thermally insulated “tunnel” which directly connects an upstream cleaning machine to the downstream sterile area. Inside the tunnel is a horizontally rotating transport belt made of a stainless-steel mesh. When a product is sterilized in a tunnel, it is dried, heat-treated by radiant heat or hot air, and then cooled. A typical tunnel sterilization exposure time is 3 or 4 min at 300°C or more. Whether radiant heat or hot air is used, the internal part of the tunnel must remain pressurized by ventilation at an intermediate pressure level that is between the sterile downstream system and the loading room.^[5] Fig. 4 shows a comparison of the two main types of sterilizing tunnels.

Radiant-heat sterilization tunnels

In the pharmaceutical industry, radiant-heat tunnels have had extensive usage. During this process, energy flows from a high-temperature body to a lower-temperature body and the heat is emitted in the form of finite batches of energy without the aid of an intervening medium.^[2] Infrared tunnels transmit heat by exposing the surface of items to radiation by direct rays. This is accomplished by infrared heaters, located in the roof of the tunnel; when products pass through the tunnel, their surfaces are heated along with the internal surface of the tunnel itself. This heat is then diffused throughout the product by radiation, conduction, and turbulent airflow.^[6]

The basic transfer equation for radiant heating processes is as follows:

$$qr = \sigma AT_1^4$$

where qr = rate of heat flow in BTU/hr, A = surface area of the emitting object in ft^2 , T_1 = surface temperature of the emitting object in degrees Rankine, and σ = Stefan-Boltzmann constant 0.174×10^{-8} BTU/hr ft^2 $^{\circ}\text{R}$.^[2]

There are disadvantages of IR tunnels; they are large, their belt speeds are slow, product heat-up may be less uniform, and they may create particulate problems. These particles are generated from the deterioration of the heating element and the moving belt. This problem can be helped by installing HEPA filters at the entrance of the tunnel and at the cooling zone junction.^[6]

Hot-air laminar flow sterilization tunnels

The most effective means for the sterilization of air is incineration. In laminar flow (LF) tunnels, heating occurs



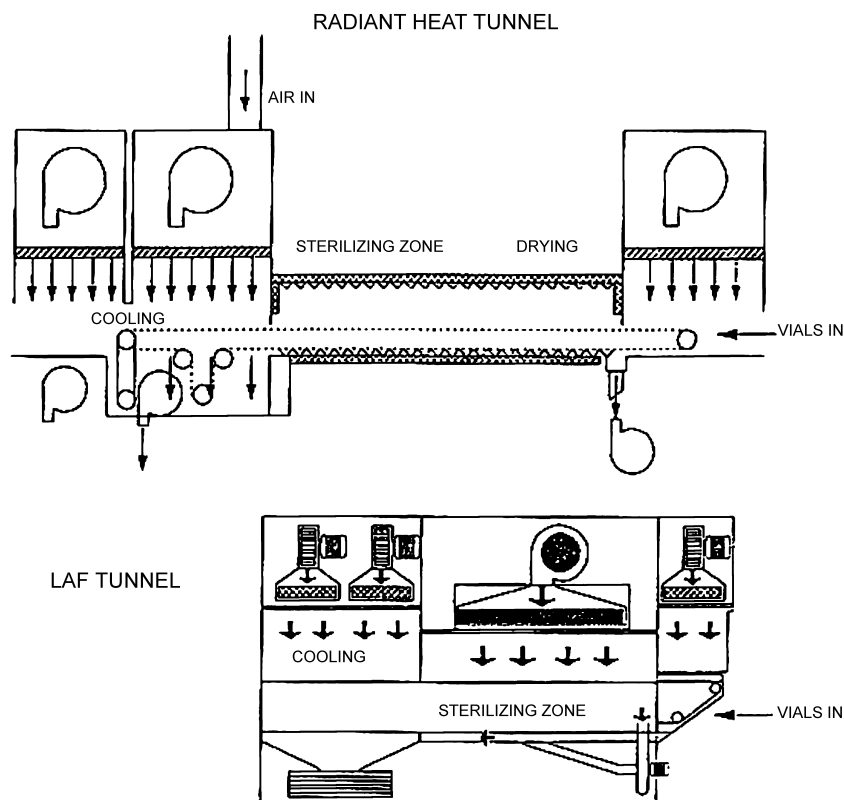


Fig. 4 Sterilizing tunnels. (From Ref. [5].)

by the circulation of hot filtered air, which is forced onto the product. The air is then withdrawn by a circulation fan and leaves the product through heating elements below the transport belt and then is brought back into the tunnel through HEPA filters.^[5] This sweeping action of the laminar airflow helps contribute to the removal of particulate contamination. The air speed within the LF tunnel remains around 0.5 m/sec. Laminar flow tunnels also have the advantage of quicker heating resulting in shorter process time. This fast heat transfer also contributes to a lower risk of product contamination from particles.^[6]

From the point of view of handling the product, continuous tunnels are more favorable than batch ovens. The main reason for this is that there is no batch work required after the unpacking of the components and loading of them into the tunnel until the packaged product is removed from the line after filling. This is important in large-scale productions, but for easier isolation of the sterile area, batch ovens are better. This is because of the fact that there must be a continuous flow of air through the open connection between the sterile area and the tunnel. The pressure of the sterile area must always remain higher than that of the tunnel. A large difference between

the two would cause an excessive escape of air from the sterile area to the tunnel resulting in the disturbance of the laminar airflow and temperature inside the tunnel.^[5]

EFFECTS OF DRY HEAT ON MICROORGANISMS AND ENDOTOXINS

Inactivation of Microbial Populations

As mentioned earlier, inactivation of microorganisms occurs by oxidation, but other possibilities must also be considered. One possibility is the effect on DNA. In *B. subtilis* spores, sublethal temperatures have been shown to induce mutants as a result of depurination. There are also claims that genetically determined differences in water content could result in the dry-heat sensitivity of spores.^[8] A great amount of interest has been generated relating the water activities of spores with their thermal resistance. The water content of dry microbial cells is important in regards to their destruction rate, and water content is determined by the relative humidity of the atmosphere surrounding the cells; therefore the destruction rate varies



Sterilization by Dry Heat

with the relative humidity of the system. In dry-heat sterilization, water is not present in the liquid state. If the relative humidity is 100%, then some water is present in the liquid state. Therefore the relative humidity of a dry-heat sterilization system can be any value between 0% and 100%. To establish the relative sterilization effect in dry-heat sterilization, the relative humidity conditions must be known in addition to the temperature.^[1] At equilibrium, the water activity (A_w) inside the cell is equal to the relative humidity of the atmosphere surrounding the microbial cell.^[8] Water activity is used to describe the relative water availability inside a microbial cell or spore and is related to the concentration and is expressed as a fraction of the pure component in the standard state where:

$$A_w = 1$$

When reporting on dry-heat test data, relative humidity is measured and controlled instead of water activity. This is because of the fact that biological materials more closely parallel vapor pressure than water content, and vapor pressure of the gas atmosphere surrounding dry microbial cells is more easily measured and controlled than the water content inside the cell.^[1]

There are three primary and three secondary variables involved with the mode of action of dry heat. Temperature, water content, and time are the primary variables, and the secondary variables are open and closed systems, physical and chemical properties of the microorganism and adjacent support, and the gas atmosphere.^[8]

Effect of temperature

The most important variable in dry-heat sterilization is temperature. Temperature, which is the measure of the heat energy level, is a function of time. One measure of the change in destruction rate with temperature is the z value. The z values for dry-heat inactivation of microbial spores have been reported from about 15°C to 30°C.^[8]

Effect of microbial water content

As discussed earlier, water has a direct influence on the resistance of microorganisms to dry-heat destruction. The destruction rate of spores is a function of the quantity of water in the cell at the time of heating. This water content is only constant under certain conditions and in most conditions, the moisture content of the cell can change so that the secondary variables cause confusion in analysis of results. The water vapor pressure in the atmosphere surrounding the cell determines the movement of water to or from microorganisms on surfaces. Research found that when the humidity in air passing over spores was in-

creased from 0 to 0.2, the D value also increased by a factor of 100. Spores of intermediate moisture content with an RH between 0.1 and 0.6 were found to be more resistant to the effects of heat at temperatures between 100°C and 135°C than spores of either greater or lesser moisture content.^[8]

Opened and closed systems

Opened and closed systems were first suggested to indicate the relative control of the heating environment on water loss or gain of spores.

In an open system, the atmosphere surrounding the spore determines the spore water content. In this system, cells are heated, and during this heating time, water may be lost or gained by the cell. The cells can gain or lose water either rapidly or slowly, depending on the physical system and the nature of the spore.^[8]

In a closed system, the spore water content is not influenced by the heating environment, but is more a function of condition inside the enclosure. Closed systems are different in that water movement and water availability to the cell are restricted. The quantity of water that is initially present in the enclosure limits the quantity of water that is available or that can move to or from the cell. Initial water content and enclosure volume of the closed system are two important parameters. The relative humidity of the atmosphere in the enclosure at the time of sealing determines the water concentration in the cell during heating as well as the total volume of the enclosure. Changing either of these parameters will alter the relative humidity, which then alters the water content in the spore during heating.^[8]

Dry-Heat Destruction of Microorganisms Associated with Soil

There are two groups of microorganisms that accumulate on an object when it comes in contact with people and air. The first group includes naked or unprotected microbial cells from human or animal origin. The second group is made up of spores found in particles of dust or dirt that fall on the object being sterilized. These spores come either from the air or from the activity of persons working near the object. The naked microbial cells of the first group are generally easily killed by dry-heat sterilization, but the spores found in soil particles can be difficult to kill. Recent findings on the dry-heat destruction characteristics of microflora associated with soil particles include the following:

1. Microorganisms found in soil are widely variable, having different survival characteristics.



2. Different organisms found in soil are highly resistant to dry heat, including *Bacillus xerothermodurans*, which was found to have a $D_{125^{\circ}\text{C}}$ value of 139 hr.
3. Some organisms can become more resistant to heat when encapsulated in particles, such as *B. subtilis*.
4. It is estimated that 1 spore in 10^3 and 10^5 is extremely resistant to dry heat when found in soil.
5. The bound water in soil particles may overshadow the effect of the humidity level on the destruction rate of the spore.^[8]

The possible variation in numbers and species of microorganisms found in soil and the physical conditions of their location are unlimited. Spores found in soil can be located at any point on or within the particle. Spores can become completely encapsulated in soil particles through the various soil wetting and drying cycles, which can increase resistance by a factor of 10 .^[8]

Destruction of Endotoxins

Bacterial endotoxins, or pyrogens, are substances that will cause a variety of symptoms such as a rise in body temperature when injected in large amounts into human or animal bodies. In pharmaceutical products and medical devices intended for parenteral injection, the absence of endotoxins has equal or greater importance than sterility. None of the other large-scale sterilization methods, such as saturated steam, gamma radiation, and ethylene oxide, are capable of destroying endotoxins.^[6]

All microorganisms seem to be capable of producing pyrogens, but the most potent forms of pyrogens are associated with gram-negative bacteria. Located externally to the peptidoglycan cell wall of gram-negative bacteria is a loosely structured envelope. The outer layer of this envelope is made up of lipopolysaccharides linked to phospholipids and proteins that act as a permeability barrier that stops the diffusion of exoenzymes into the greater environment. It is the lipopolysaccharide fractions of the cell envelope that have been shown to stimulate the pyrogenic response and are termed bacterial endotoxins. Purified endotoxin, which consists only of lipopolysaccharide, is pyrogenic in lower doses than naturally occurring endotoxins. There are three distinct chemical layers of lipopolysaccharide; they include an inner core called lipid A, an intermediate polysaccharide layer, and an outer polysaccharide side chain. Lipid A is the region that is responsible for pyrogenicity and consists of a highly substituted disaccharide of glucosamine.^[6]

The dry-heat destruction of endotoxins is complex and poorly understood. Most evidence has shown that destruction follows second-order kinetics with an initial high

rate of destruction followed by a slower terminal rate.^[6] The kinetics of dry-heat destruction for lipopolysaccharide are expressed in terms of D^1 and D^2 , for the initial and secondary first-order reaction rates, respectively. Previous research found that there was greater reduction in the dry-heat resistance of lipopolysaccharide in whole cells than that in the semipurified state. The D^1 value for the whole cells was found to be 1.6, and the D^2 value was 12. In the semipurified state, D^1 was 3.7 and D^2 was 29.4. Thus the rate of the destruction was two times faster when whole cells were used.^[9]

There is basically no endotoxin destruction at temperatures below 80°C , and the D values can be as high as 20 min for dry-heat temperatures of around 170°C . The second-order models give a better estimate of endotoxin destruction kinetics at temperatures above 250°C than in the 170 – 250°C range.^[6] Generally, the following conditions are required for endotoxin destruction:

230°C —60 to 90 min.
 250°C —30 to 60 min.^[5]

The USP and FDA both state that a claim to depyrogenation should be supported by evidence that any endotoxin present on an item prior to treatment has been inactivated to no more than $1/1000$ of the original amount or a three \log_{10} reduction. At 170°C , the minimum time required to obtain three \log_{10} reductions is between 3 and over 66 hr. At 250°C , six \log_{10} reductions have been demonstrated. For this reason, depyrogenation cycles need to be developed and validated empirically. However, the USP recommends a temperature of at least 250°C .^[6]

DISADVANTAGES OF DRY-HEAT STERILIZATION

Dry heat should be used only for materials that cannot be sterilized by steam either because the moisture would damage the materials or they would be impermeable to it. There are many complicating factors associated with dry-heat sterilization. The process of steam sterilization is accomplished by saturated moisture; in dry-heat sterilization, the moisture can vary considerably. Because of the loss of moisture, the death rate of spores might change with the continued application of heat.^[1]

Although air is the least expensive means of sterilization, it is not the best heat transfer medium. In oven sterilization, there are several drawbacks. One drawback is the lack of uniformity of temperature within the oven. Hot air has a tendency to stratify and poorly penetrate masses of cooler material, but moving the air stream can accelerate heat transfer. Slow heating and cooling down



Sterilization by Dry Heat

because of low heat transfer rates from air to product is another problem with ovens that leads to longer sterilization cycles. Other disadvantages of dry-heat sterilization can occur because of the high temperatures, which can cause heat damage or even charring of materials. Materials can also be damaged due to oxidation because the medium that facilitates this destructive action also augments its deleterious effects.^[1]

CONCLUSION

Although much work has been performed to improve and develop other methods of sterilization, very little has been accomplished in the field of dry-heat sterilization. The process is time consuming and difficult to control because of the temperature stratification and slow heating rate.^[10] Dry heat is still the agent of choice for sterilizing items that might not be adequately penetrated by steam and that will tolerate high temperatures, such as oils, petrolatum, and closed containers.^[2]

The development of the infrared radiation tunnel in recent years has opened new possibilities for using dry heat in high-temperature short-time sterilization processes.^[4] The main advantage of dry-heat sterilization is its penetrating power. It is not as corrosive as steam for metals and sharp instruments, and it does not erode ground glass surfaces, which allows glass to be sterilized at much higher temperatures for shorter periods of time.^[1]

REFERENCES

1. Joslyn, L. Sterilization by Heat. In *Disinfection, Sterilization, and Preservation*, 3rd Ed.; Block, S., Ed.; Lea and Febiger: Philadelphia, 1983; p. 3, 27–30, 766–767.
2. *Validation of Dry Heat Processes Used for Sterilization and Depyrogenation*, Technical Report No. 3; Parenteral Drug Association: Philadelphia, 1981.
3. Kupiec, T.; Mathews, P.; Ahmad, R. Dry-heat sterilization of parenteral oil vehicles. *Int. J. Pharm. Compd.* **2000**, *4* (3), 223–224.
4. Molin, G.; Molin, O. Dry-heat sterilization of pharmaceutical glassware using hot air or infra-red radiation. *Acta Pharm. Suec.* **1976**, *13*, 476, 481–483.
5. Remington, J. Dry Heat Treatments. In *The Science and Pharmacy*; Mack Publishing Company: Easton, PA, 1995; 1472–1474.
6. Halls, N. Dry Heat Sterilization and Depyrogenation. In *Achieving Sterility in Medical and Pharmaceutical Products*; Marcel Dekker, Inc.: New York, 1994; 109–120.
7. Bruch, C.W. Levels of sterility: Probabilities of survivors vs. biological indicators. *Bull. Parenter. Drug Assoc.* **1974**, *28* (3), 105–113.
8. Block, S.S. *Disinfection, Sterilization, and Depyrogenation*, 4th Ed.; Lea and Febiger: Malvern, PA, 1991; 35, 97–101, 514–515.
9. Tsuji, K.; Harrison, S. Dry-heat destruction of lipopolysaccharide: Dry-heat destruction kinetics. *Appl. Environ. Microbiol.* **1978**, *36*, 710–714.
10. Molin, G.; Östlund, K. Dry-heat inactivation of *Bacillus subtilis* var. *niger* spores with special reference to spore density. *Can. J. Microbiol.* **1976**, *22*, 359.



Request Permission or Order Reprints Instantly!

Interested in copying and sharing this article? In most cases, U.S. Copyright Law requires that you get permission from the article's rightsholder before using copyrighted content.

All information and materials found in this article, including but not limited to text, trademarks, patents, logos, graphics and images (the "Materials"), are the copyrighted works and other forms of intellectual property of Marcel Dekker, Inc., or its licensors. All rights not expressly granted are reserved.

Get permission to lawfully reproduce and distribute the Materials or order reprints quickly and painlessly. Simply click on the "Request Permission/Order Reprints" link below and follow the instructions. Visit the [U.S. Copyright Office](#) for information on Fair Use limitations of U.S. copyright law. Please refer to The Association of American Publishers' (AAP) website for guidelines on [Fair Use in the Classroom](#).

The Materials are for your personal use only and cannot be reformatted, reposted, resold or distributed by electronic means or otherwise without permission from Marcel Dekker, Inc. Marcel Dekker, Inc. grants you the limited right to display the Materials only on your personal computer or personal wireless device, and to copy and download single copies of such Materials provided that any copyright, trademark or other notice appearing on such Materials is also retained by, displayed, copied or downloaded as part of the Materials and is not removed or obscured, and provided you do not edit, modify, alter or enhance the Materials. Please refer to our [Website User Agreement](#) for more details.

Request Permission/Order Reprints

Reprints of this article can also be ordered at

<http://www.dekker.com/servlet/product/DOI/101081EEPT120012009>



Sterilization by Ethylene Oxide

Terezinha de Jesus Andreoli Pinto

Universidade de São Paulo, São Paulo, Brazil

INTRODUCTION

The biological activity of ethylene oxide (EtO) was initially observed by Cotton and Roark,^[1] who detected it as an insecticide, in concentrations ranging from 3.2 mg L^{-1} to 32.0 mg L^{-1} . In 1937, Gross and Dixon^[2] requested the patent of EtO as a sterilization method due to the considerable experimental evidences it presented.

The first EtO commercial application was related to spices sterilization and/or fumigation.^[3–7] Phillips and Kaye^[8] conducted a series of studies using EtO, focusing on the sterilization of material with thermal resistance problems.

Although its posterior industrial applications included pharmaceutical and cosmetic products, packaging materials and raw materials of biological origin,^[9] EtO is predominantly used in the sterilization process of medical devices, such as intravenous sets, cardiopulmonary and anesthesiology devices. In the hospital environment, the EtO largely contributed as one of the main agents responsible for the controversial increase in the reutilization of single-use medical devices.

ETHYLENE OXIDE CHARACTERISTICS

EtO (oxirane, epoxyethane, and dimethylene oxide) is a colorless gas condensing at low temperatures to a mobile liquid. It is miscible in all proportions with water, alcohol, ether, and most organic solvents. Its vapors are flammable and explosive. The physical properties of EtO are summarized in Table 1.

EtO is highly reactive; industrially, it is a primary chemical intermediate for a wide variety of compounds. The three-membered ring is opened in most of the reactions, which give a series of polyethylene glycol derivatives of increasing chain length and water solubility. Ethylene's strained configuration has been a subject of bonding and molecular structure studies. It is generally noncorrosive to metals and leaves no odor or taste residues on the materials submitted to its contact. The vapors of EtO (not the liquid form) can be flammable and explosive at concentrations above 3% in air. For this reason, it is often combined in sterilant mixtures with an inert gas to eliminate flammability.

EtO is used outside the health care industry as well. Actually, only a small fraction of EtO is used as a sterilant. It is registered as a pesticide by the Environmental Protection Agency (EPA) and is widely used as a fumigant in the food industry. It is also used as a chemical in the manufacture of detergents, plastics, fibers, film, antifreeze, and other products.^[10]

MECHANISM OF ETHYLENE OXIDE STERILIZATION

The activity of EtO, similar to many other disinfectants, preservative, and sterilizing agents, such as formaldehyde, beta-propiolactone, methylbromide, and ethylenimine, depends on an alkylation reaction.^[9] This reaction occurs with some groups within the complex enzymes, proteins, and nucleic acids in the bacterial cell.^[11] These compounds can then no longer be effective or necessary to the vital processes of the microorganism cell. Furthermore, one would expect these effects to vary according to the extensions of reactions: static, mutagenic, or toxic.

The effects of EtO concentration and process temperature were widely studied.^[8,12,13] Ernst and Shull^[13] initially verified that the lethality kinetics of *Bacillus subtilis* var. niger spores is of zero order at high levels of EtO. With the EtO concentration reduction, the reaction becomes first order. Its sterilization curves at different temperatures show the required EtO concentration, and the duplication of the reaction rate of every 10°C .

It has been well established that microorganisms that have been equilibrated with atmospheric humidity—or even dried to very low humidity, however, retaining some free water—are killed easier than those desiccated. It is accepted that organic chemical reactions occur through an activated complex formation; thus, we can infer that water influences this activation. Water must also be present as a reaction medium or solvent if biological entities are to be ionized, so that they can enter a transition state with EtO.^[14] Furthermore, no sterilizing activity can be observed when a nonpolar solvent such as dioxane and chloroform replaces water.

**Table 1** Typical physical and chemical properties of ethylene oxide

Molecular weight	44.05
Specific gravity ($H_2O = 1$) at 68°/39.2°F (20°/4°C)	0.8700
Specific gravity (air = 1)	1.49
Vapor pressure at 68°F (20°C)	22 psia (151.7 kPa abs)
Solubility in water	Complete
Percent volatile by volume	100.0
Evaporation rate (butyl acetate = 1)	High
Boiling point at 1 atm	50.9°F (10.5°C)
Freezing point at 1 atm	− 170.7°F (− 112.6°C)
Appearance, odor, and state	Colorless liquid and gas at normal temperature and pressure. Nonresidual, ether-like odor about 500 ppm. Odor not detectable until well above the PEL. Shipped and stored under nitrogen pressure as a liquid

Water has another function in this sterilization process—the sterilizing agent concentration on the surface of the microorganism. One can visualize microorganisms in an EtO environment in which all the sterilant is in the vapor phase, and in which the reaction rate is dependent on both the number of collisions of EtO molecules with active sites of the organism, and the colliding molecules absorption by the organism, followed by migration of the EtO molecules to the active sites. The collision reaction is quite slow; however, in an atmosphere containing water vapor and EtO, water will equilibrate the organism in the liquid phase, dissolve the EtO, and concentrate the EtO on the organism as a solution, thereby increasing the rate of microorganism killing.^[14]

TECHNOLOGY OVERVIEWS

Occupational Engineering Considerations

Hazard assessment of EtO has involved considerations about its chemical composition, characteristics, and reactivity. In spite of EtO's readiness to be transformed in the body to ethylene glycol and ethylene chlorohydrine, and its biological half-life of about 10 min, it is distributed throughout most body organs. Studies with cells, animals, and also epidemiological ones, showed that EtO is able to produce neuropharmacologic, neurotoxic, reproductive, teratogenic, and mutagenic effects.^[15–17]

The Occupational Safety and Health Administration (OSHA) is responsible for regulating the in-plant occupational exposure to EtO in the United States. Under the current OSHA standard, accepted by several other countries, the permissible exposure level (PEL) of EtO is 1 ppm (1.8 mg m^{-3}) per 8 hr (time-weighted

average). However, an action level of 0.5 ppm is advisable, since it is a guarantee of employers from low employee exposure levels.

The OSHA has also established an exclusion limit for EtO of 5 ppm (9.0 mg m^{-3}) average over a maximum sampling period of 15 min. In addition to these regulatory pressures, high levels of concern about workers protection and the potential liability implications have significantly increased industry's interest in control measures for EtO.

The main EtO sources that potentially contribute to workers' exposure in sterilization plants generally can be divided into four major categories.^[18]

- Exposures associated with sterilization equipment operation: sterilizer (door gasket leaks, piping leaks, leaks from valves and fittings, vent line leaks, and plume recirculation); vacuum pump (seal leaks, degassing from sealant discharged to drain); sterilant gas storage and delivery system, such as tanks, cylinders, or drums (leaks from valves and fittings, gas line leaks, and releases during change over).
- Exposures associated with sterilized product degassing: unloading, transfer, storage, and spore sample removal.
- Exposures associated with maintenance operations: routine maintenance and repair, preventive maintenance.
- Exposures associated with emergency situations: major leaks, major spills, and equipment malfunctions.

Engineering solutions must be studied, developed, and introduced in order to eliminate or, at least, minimize problems in all the steps.^[19]

Sterilization by Ethylene Oxide

3

Ventilation

General area ventilation systems are designed on the basis of how much EtO can be degassed into the specific area mentioned. It is essential to consider the total quantity of EtO remaining in the product after sterilization. This will include packaging EtO in the cardboard vapor spaces and the primary packaging, as well as the absorbed/adsorbed EtO in cardboard, the packaging, the product, and the pallets.

Vacuum pump modifications

Vacuum pump modifications usually involve conversion of a once-through liquid ring pump to a full sealant recovery pump. This will eliminate any continuous liquid discharge to the drain. The sealant can either be water or oil. The use of water places limitations on the maximum vacuum depth, which is reasonably attenuated because of the vapor pressure effects. The use of oil eliminates this maximum vacuum depth limitation.

Hot degassing

The major source of worker exposure to EtO in the plant is the degassing operation after removal from the sterilizer. Different types of products absorb or entrap EtO at different levels, depending upon the nature of the product, its specific materials of construction, how it is packaged, and the specific sterilization cycle to which it was exposed. Unfortunately, not all the absorbed or entrapped EtO is

removed in the sterilizer post-evacuation cycle. Hence, the remaining EtO will degas after the products are removed from the sterilizer. This degassing process typically takes place over an extended period of time (days) at typical plant ambient conditions and usually it is necessary that the products be quarantined for this period.

The use of heat in hot degassing chambers or rooms to accelerate the EtO degassing has been known for a long time. A hot degassing tunnel built directly behind the sterilizer has also been used. In this case, the sterilizer must have double doors and one must provide automatically taking out the pallets from the sterilizer and transporting them into the tunnel.

Automated systems

The most efficient way to protect the worker is keep him away from all operations, like loading, unloading, and starting procedures during the cycle. So the sterilization process should be automated and computerized.^[20] Besides the worker protection, the computerized system allows the cycles “self-certification,” which adds reproducibility advantages to efficiency of the sterilization cycle.

Fig. 1 (derivada do 6, mas modificada) shows a plant design that exempts the worker presence, except near the control panel. The residue monitoring should be included in the system. It must embrace the environment and obviously all processed products and discharges. Among several options, gas chromatography, especially the one with head space, has been the most adjusted.

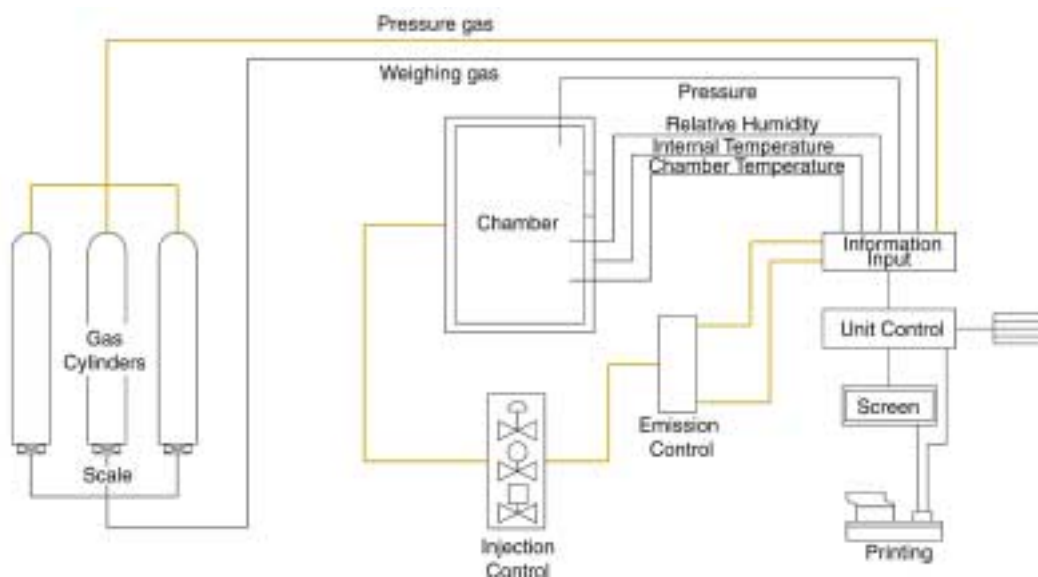


Fig. 1 Example of a sterilization plant provided with automatic system.

Ethylene Oxide Emission Control

During the handling of the ambient controls, the different forms of EtO emission must be controlled, in order to minimize indirect contamination to the population. For example, according to the amendment of the TA-Luft (German air pollution regulations) in 1986, the emission limit of EtO has been fixed to a concentration of $5 \text{ mg N}^{-1} \text{ m}^{-3}$ at a mass flow $> 25 \text{ g hr}^{-1}$. Mayer et al.^[21] described an appropriate plant of compact design that, using installations for ethylene glycol reactions, ensures these limiting values for the sterilization with EtO.

There are current technologies that can be used to control EtO emissions to the environment, with specific advantages and disadvantages.^[22]

Water scrubbing

This technology involves the absorption of EtO into water in a packed column. The gas stream discharged from the sterilizer enters the column near the bottom and flows upward through the packing material. There are systems in which absorption of gaseous EtO into the aqueous solution occurs, with its posterior reaction with water to form ethylene glycol, promoting cooling and recycling of the solution for further EtO removal. The use of an acid reaction (0.1 N is a good concentration, see Fig. 2) with water is a good option; anyway, it will be necessary to neutralize water, before its discharge to the environment (usually once or twice a year).



Thermal incineration

Thermal incineration uses a flame to oxidize EtO to carbon dioxide and water at high temperature (typically

760–982°C or 1400–1800°F). A residence time of 0.3 sec–0.5 sec is enough to achieve a high EtO destruction efficiency.

Difficulties are related to the high cost of the equipment and to limitations in the case of mixtures, when hydrogen chloride (HCl) and hydrogen fluoride (HF) are involved.

Catalytic oxidation

In a catalytic oxidation system, the gas stream exhausted from the sterilization is first diluted with relatively large quantities of air and heated up to a considerably lower temperature (204–426°C or 400–800°F). The mixture is then passed through a catalyst bed where EtO is heated with oxygen and converted to carbon dioxide and water.

Even with the advantage of high EtO oxidation efficiency, there are limitations related to potential decomposition of mixture compounds in toxic by-products (such as phosgene). In addition, the equipment cost is relatively high.

TECHNOLOGY NEW TRENDS

Related to explosive decomposition properties of EtO, pure liquid EtO can inflame in the presence of an ignition source.^[23,24] The precise threshold limits for liquid decomposition are influenced to some extent by the type of ignition source, as well as by the geometry of the vessel used. Pure EtO vapor can explode by decomposition in the presence of common igniters. Pure EtO vapor at normal storage conditions is more difficult to ignite than mixtures of EtO and air or mixtures of hydrocarbons and air. The potential for decomposition can be eliminated by diluting EtO vapor with a specified proportion of inert gas.

Though EtO itself may be used to carry out the sterilization, since the beginning of its use as a sterilant,

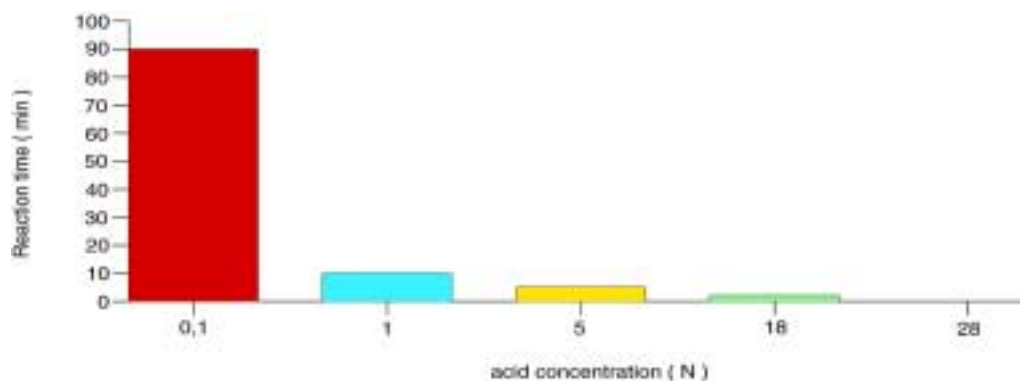


Fig. 2 Reaction time to form ethylene glycol considering time.

Sterilization by Ethylene Oxide

5

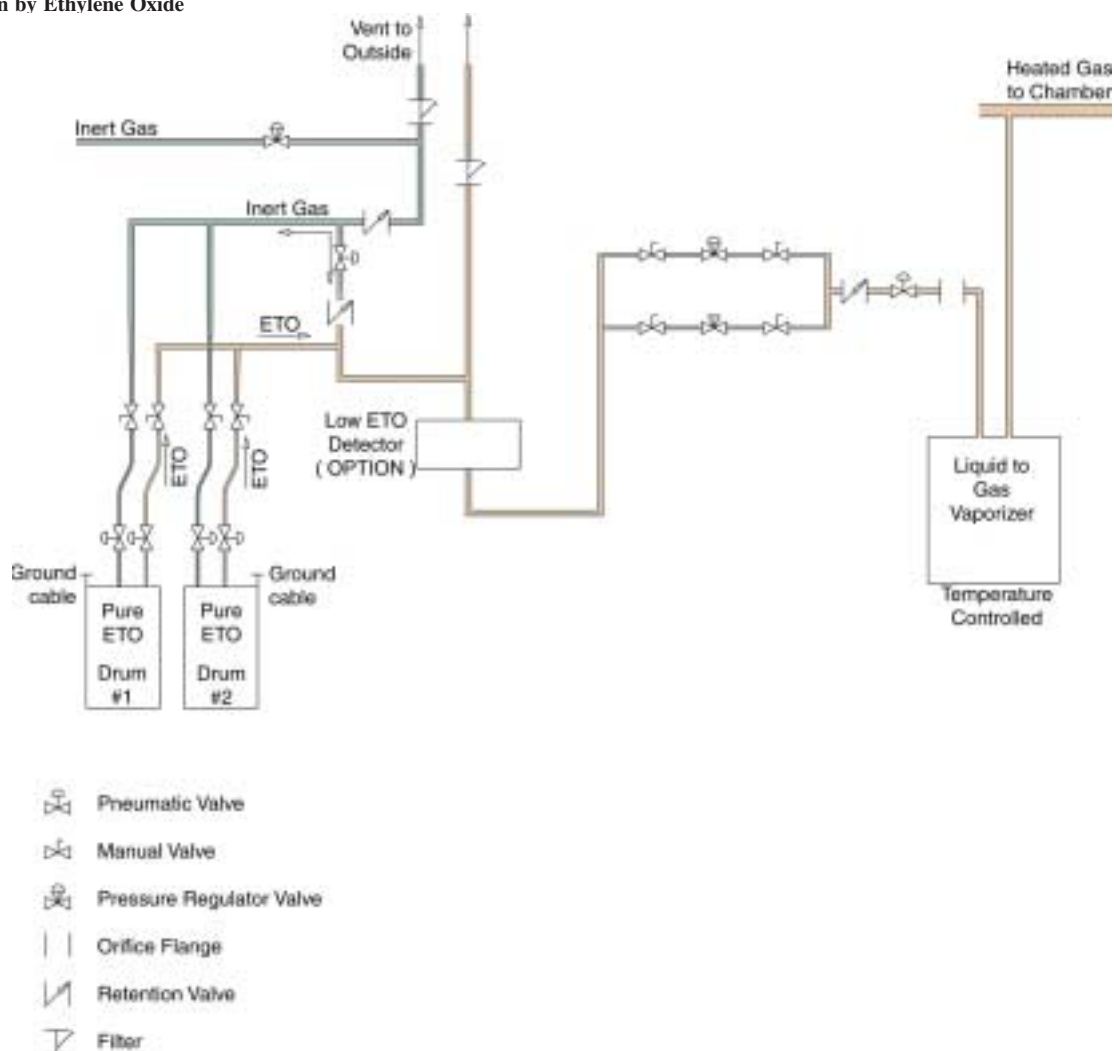


Fig. 3 Example of a single installation for a 100% EtO sterilization system.

this has not been done. Instead, EtO sterilant has been most used in a mixture with a flame retardant. Over the 1970s and 1980s, dichlorodifluoromethane, known in the industry as CFC 12, was the flame retardant of choice for use with EtO in a sterilant mixture. The most commonly used mixture consisted of 27.3 mol% (12 wt%) EtO and 72.7 mol% (88 wt%) CFC 12; this mixture is commonly referred to 12-88 in the industry. Also, a 10:90 mixture of EtO: carbon dioxide (CO₂) was used, but requiring an increasing of 15 psig in the sterilization pressure (this mixture contains only 10% EtO in volume, in contrast to 27% in volume of the 12:88 mixture).^[10]

Due to problems that have arisen with the use of CFC12 (one of the compounds which causes significant damage to the ozone layer in the upper atmosphere), worldwide reduction and elimination of CFC12-use have been started up. This restriction, although related to inert dilutor and not to EtO, reached great part of the users being thus added to the occupational limits requirements. The new

technician-scientific concepts reached the economic area^[25] aggravating extensive fetching of alternatives.

Alternatives that had gained new impulse were the mixtures with carbon dioxide. The 30:70 mixture with inflammability features has led to the construction of new plants, considering the requirements.

Another option is a nonflammable EtO—carbon dioxide mixture that contains less than 40% of the EtO per unit volume, as does 12-88. Thus, sterilization must be carried out either at higher pressures as for longer exposition times. Furthermore, the large difference in the vapor pressures of EtO and carbon dioxide causes the mixture to separate upon withdrawal from the storage tank or cylinder, raising the danger of delivering a sterilant mixture rich in carbon dioxide, which will not sterilize, or rich in EtO, which is explosive.

Several companies also re-evaluated their positions about using 100% EtO.^[25] In fact, the high costs to build an explosion-proof equipment with adequate installations

would be paid back in savings by switching to 100% EtO. Also, in the past, the main difference between sterilization with 100% EtO and 12/88 was that a deep vacuum was generally used for 100% EtO sterilization, so that all of the air in the package was sucked out, then replaced with a mixture of EtO and moisture; this deep intense vacuum required breathable packaging material or vents that would allow the air inside to escape, so the package would not blow up like a balloon and break. A vacuum was generally maintained in the 100% EtO sterilizer, so that if any leaks occurred, outside air would be sucked into the sterilizer, rather than EtO would leak out, as can occur with positive sterilizer pressures.

Nowadays, it is known how to use 100% EtO along with an inert gas like nitrogen to get safe sterilization, even for the most sensitive packaging. Pressures may be balanced and a modern computer that controls 100% EtO sterilizer is a very versatile piece of equipment. Fig. 3 shows schematic installations with an inert gas hookup to the drum. There are also two devices to make sure that the EtO supply to the sterilizer is shut off when the drum is emptied. One is the weight scale, which automatically signals the computer when a preset amount of EtO has been withdrawn and shuts off the drum by a compressed air valve. The another device is the computer that stops the flow when it senses a change from EtO to the inert gas.

Without these two devices, it is possible that the drum could be filled with inert gas and, of course, this would not sterilize the materials.

Of course, there are no electrical switches or an ignition source in the closed sterilizer–heat exchanger system. Even the presence of high amounts of moisture both in the product from the preconditioning rooms and from the humidity injection system would reduce the likelihood of an explosion by absorbing the EtO into the moisture and by eliminating the cause of static charges.

If inert gas lines are hooked into the EtO pipes near the EtO source, inert gas can sweep the EtO from the lines so that they will be either filled with the inert gas or be empty. Inverting the filters and flushing the lines with inert gas can keep any EtO from being collected in the filter housing—that also eliminates the exposure involved in changing filters. Filter changing is probably the major source of exposure for all types of EtO sterilization and this system virtually eliminates that problem.^[22]

The use of 100% EtO sterilization techniques offers low residuals due to the more efficient EtO removal from the package by the deep vacuum. The process of “pulsing” the vacuum as reported by Christensen^[26] not only greatly lowers residuals which form glycols and chlorohydrins by quickly reducing the amount of EtO in the package, but also allows degassing outside the sterilizer to be done

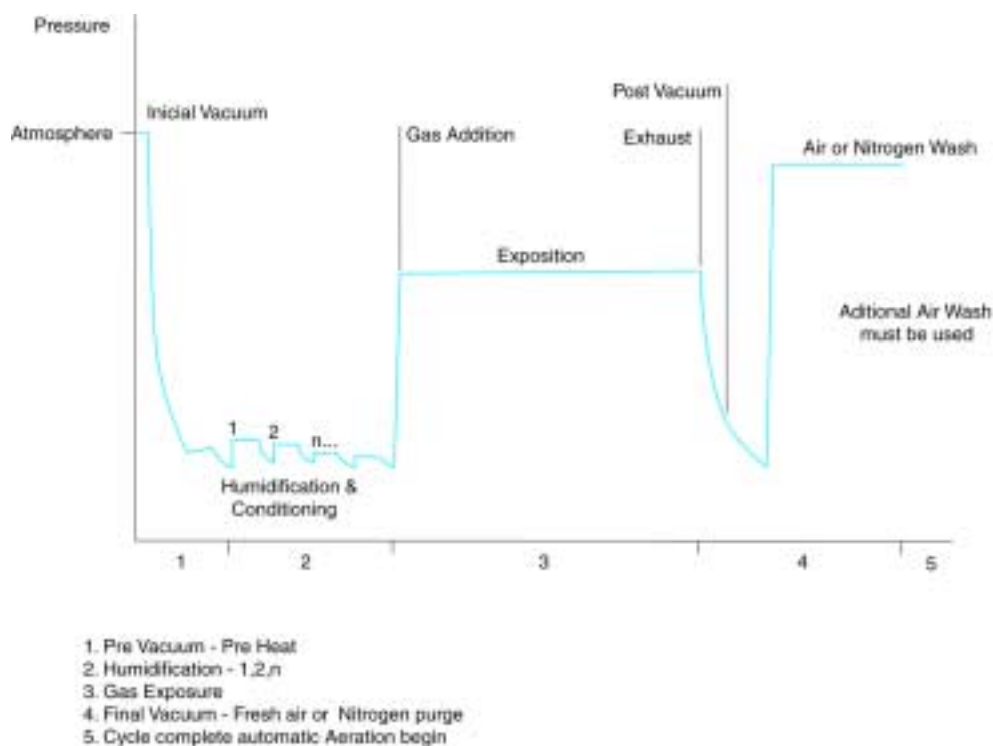


Fig. 4 A typical sterilization cycle observed with single-use 100% EtO cartridges.

Sterilization by Ethylene Oxide

7

within the OSHA limits. In fact, one of the latest trends is to do all the degassing in the sterilizer, so OSHA exposure levels are never even approached. This procedure is ideal for 100% EtO systems, especially those having an EtO disposal unit (Fig. 4).

Recent knowledge about ozone-depleting gases and an international consensus on the need of reducing their effects promoted a search for alternative chemicals to several different CFC applications. From these, one of the most interesting are the hydrochlorofluorocarbons (HCFCs) which, besides being similar to CFC in inhibition of explosiveness and inflammability of EtO, can also be used as transitional compounds while more environmentally suitable compounds are not available. An EtO–HCFC (10:90) mixture was compared to 100% ethylene sterilizing gas to determine its relative ability to kill seven different bacteria. Results demonstrated that the EtO–HCFC mixture was equivalent to the 100% EtO sterilant to kill vegetative organisms, as well as spore suspensions in

the absence of serum and salt, and better with these contaminants on carriers.^[27] Also, a study performed^[28] to compare sterilization effects between Oxyfume[®] 12 (using EtO and CFC12, 12:88) and Oxyfume 2002 (using EtO and HCFC 22 and 124, 10:63:27), under different concentrations (450 mg L^{-1} and 600 mg L^{-1}) and different temperatures (45, 55, and 65°C) revealed similar results, even in different concentrations (similar D value using *Bacillus subtilis*, var. *niger*, ATCC 9372).

Additional advantages of this new mixture are related to the use of the same equipment and installations previously developed for the CFC mixture (Fig. 5), with only a little increase in pressure or time of exposition, during the sterilization cycle. The only concern is that, at present, HCFC is targeted to be phased out by the year 2020 in Canada (2 in 9) and 2030 in the United States (3 in 9).

Different options can require other kinds of equipment, designed with the goal of protecting the workers, environmental preservation and safety of products.

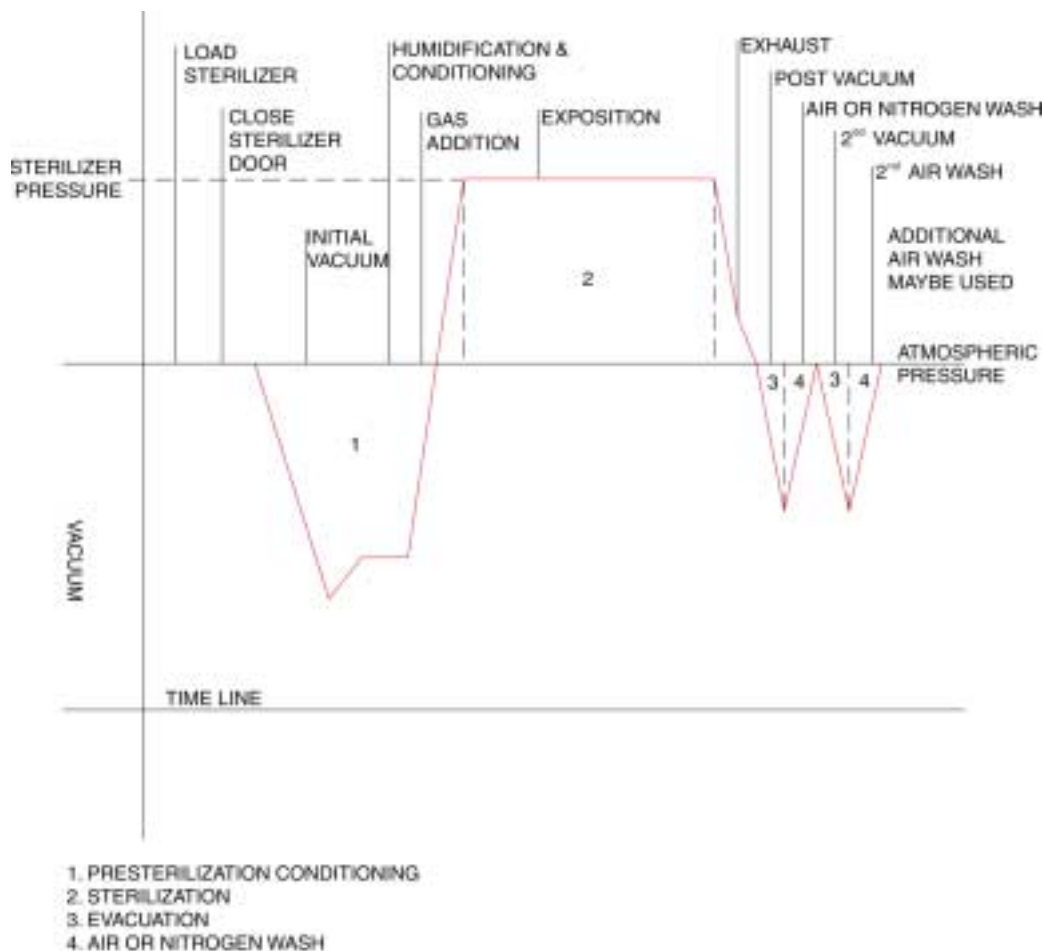


Fig. 5 Typical sterilization cycle with CFC–EtO and HCFC–EtO mixtures.



Validation and Routine Control of Ethylene Oxide Sterilization

Even though the definition of sterility is an absolute condition, the effectiveness of the sterilization process can be determined by measuring the reduction of microbial population. Such measurements reveal the kinetics of microbial inactivation, and it is from the exponential nature of inactivation that the concept of sterility assurance level (SAL) is derived. This value is expressed as a negative power to the base 10. SALs of both 10^{-3} and 10^{-6} have been adopted worldwide. However, there is a tendency to adopt a single standard, 10^{-6} .^[29]

The validation of a sterilization process is a documented procedure, which obtains records and interprets the data required, to show that a process will consistently comply with predetermined specification. Besides respecting the SAL, the biocompatibility and functional properties of a product have to be respected.

The most important current standard is ANSI/AAMI/ISSO 11135—1994 (revision of ANSI/AAMI ST27—1988). Medical devices validation and routine control of EtO sterilization were approved by the Association for the Advancement of Medical Instrumentation (AAMI), the American National Standards Institute, Inc., (ANSI) and the International Organization for Standardization (ISO). This standard considers the following steps for validation: commissioning and performance qualification, which includes physical and microbiological qualification. Medical devices to be sterilized shall be manufactured under conditions that ensure that their bioburden is consistently low. Another important point is about the AAMI specification that the biological indicator evaluator resistometer (BIER) is the appropriate vessel for evaluating resistivity to sterilization of biological indicators.^[30,31]

The adoption of the validation standard document reflects the medical industry ongoing commitment to safety, efficacy, and global harmonization of manufacturing standards. Its purpose is to provide confidence in all markets that the sterilization process is properly designed, tested, and monitored.^[29]

According to the standard, the sterilization process using EtO shall include preconditioning and/or conditioning, sterilization cycle, and aeration. Preconditioning and/or conditioning treatments must be performed under controlled conditions for a defined period of time to achieve specified temperature and relative humidity within the load. The sterilization cycle includes: air removal, conditioning (if used), sterilant injection, maintenance of specified conditions for the exposure time, sterilant removal, flushing (if used), and air admission to atmospheric pressure. Aeration must be performed with product retained under specified conditions, in the sterilizer and/or in a separate chamber or room. One of

the most important differences is that ISO 11135 requires additional temperature and humidity monitoring of the load, both during validation and routine.^[32]

According to the ISO standard, the temperature and humidity of the load must be measured during the performance qualification. These measurements should be used to establish range specifications for the temperature and humidity of the product at the end of both the preconditioning and conditioning steps, before adding the sterilant into the chamber. Only in an Annex of ISO standard (B), included as a guidance, at the end of preconditioning, the measured temperature and humidity ranges within the sterilization load should not exceed $\pm 5^{\circ}\text{C}$ and $\pm 15\%$ humidity. This annex also contains the information that a recorded temperature range within an empty chamber during sterilant exposure of less than or equal to $\pm 3^{\circ}\text{C}$ of the required set point should be obtained.

Throughout the exposure time, the sterilization load should attain the minimum specified temperature and the temperature range across the product load should be less than or equal to 10°C at any given time during sterilant exposure. The actual temperature should be determined during physical performance qualification. Although temperature measurements have long been required, direct humidity measurement of the load has not. Manufacturers have used vapor pressure calculations to ensure that when steam is added to the chamber after initial evacuation, the resulting increase in pressure indicates attainment of a level of humidity appropriate to support microbial inactivation.

By interrupting the sterilization cycle just before the introduction of EtO gas, validation runs can be performed using currently available chamber humidity sensors, even though, care maintenance and calibration of these sensors are very important, and a real challenge is to improve their state-of-art.

A double check for EtO concentration is required during sterilization (weight of sterilant, volume on direct analyses, besides the pressure rise). During aeration, the temperature of the load must be monitored in reference to a specified range established. The microbiological performance qualification should demonstrate the adequacy of the process to the sterilization of the product by the inactivation of indicators for EtO sterilization. The bioburden of the product has to be established, and indicators for EtO sterilization—spores of *B. subtilis* var. niger, which are being considered for another classification as *B. atrophaeus* on the basis of high DNA–DNA reassociation values and confirmatory automated Ribo-Print analysis^[33] must comply with EN 866-2. These indicators shall be placed at representative positions throughout the sterilization load under the cycle conditions selected to deliver less lethality. The indicator should represent, considering the holder where the spores are

inoculated, the load configuration and the package that protects it in the worst sterilizer conditions. This is added to the worst conditions in the configuration of the load.^[10,34]

According to the standard, the microbiological qualification can be performed by the lethality determination of the cycle by three methods. In one of them, the lethality shall be determined by the construction of a survivor curve using direct enumeration of a survivor. At least five points employing graded exposure times to EtO, with all other process parameters remaining constant, except time, shall be included on the survivor curve. The second possibility is the fraction-negative method, in which indicators for EtO sterilization are also exposed to EtO in graded exposure times with all process parameters remaining constant, except time. After exposure, the indicators for EtO sterilization shall be tested by direct immersion in an appropriate culture medium. Both methods permit calculation of D value (Fig. 6).^[35] The method C, or half-cycle method, can be used only for conventional product release. It involves the determination of the minimum time of exposure to EtO at which there are no survivors, with all other process parameters,^[21] except time remaining constant. Two further experiments should be performed to confirm the minimum time. Both should show no growth from biological indicators. The specified exposure time should at least double this minimum time.

Considering the parametric process, it requires more attention in comparison with conventional release. In this case, a secondary measure of chamber temperature is required, and the chamber humidity and gas concentration

must be directly monitored, what constitutes an important challenge, mainly related to safety in 100% EtO cycles (Norma ISO). An inherent characteristic of the sterilization process that reduces the interest in obtaining conditions that allow the parametric liberation is related to the attendance to specified limits of EtO, ethylene glycol and ethylene chlorohydrine previously to the product release, what demands additional time.^[36]

Another possibility, proposed by Rodriguez et al.,^[37] considers the concept of accumulated lethality (Fo), used for thermal sterilization, but applied to optimization of EtO sterilization technology. A mathematical model of the inactivation of biological indicators spores by EtO was developed, along with two formulas: a “response” equation for calculating the number of survivors of a sterilization cycle, and a formula for determining the accumulated lethality of exposure to EtO. Experiments verified that the equations are applicable to processes with relative humidity values between 15% and 90%, enabling users to compare the lethality of dissimilar EtO cycles.

On completion of the validation programme, the test results should be compiled into a test report, and characterize the certification of validation.

OUTLOOK

Despite the serious criticisms against EtO related to toxicity and environmental aspects, sterilization using this gas, 100% or in various compositions, is one of the most

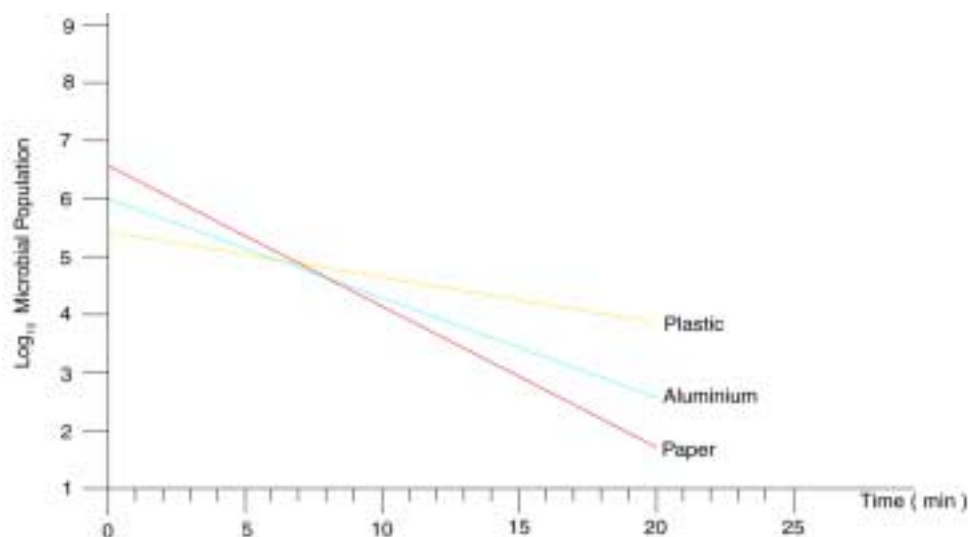


Fig. 6 Microbial death rate (D determination using direct enumeration of survivors) for *B. Subtilis* var. *niger* on paper (D = 4310 min), aluminium (D = 6143 min), and plastic (D = 12,300 min) carrier.

widely used processes. Compared to alternative processes, EtO still remains a suitable choice than irradiation, because it promotes molecular alterations in different polymeric compounds, and it also causes long-term problems with the ramrods of ^{60}Co , when its active life expires. Besides that, the irradiation with difficulties persists with accelerated electrons, in spite of the obtained improvements.

Furthermore, the EtO process presents wide effectiveness and possibility of validation in industrial sterilizers, when compared to the plasma sterilization process. It also shows advantages in comparison to other sterilant gases (formaldehyde and hydrogen peroxide) related to permeability, diffusion, volatilization, polymerization, and compatibility.

So, since there is not a perfect sterilant agent, we can consider the EtO process as ideal, although it demands a certain amount of knowledge for its safe and effective use.

REFERENCES

1. Cotton, R.T.; Roark, R.C. Ethylene Oxide as a Fumigant. *Ind. Eng. Chem. Wash.* **1928**, *20*, 805–809.
2. Gross, P.M.; Dixon, L.F. Method of Sterilizing. U.S. Pat. 2,075,845, 1937. Apud: Ref. 8.
3. Hall, I.A. Sterilized Spices: New Factor in Food Quality Control. *Food Ind.* **1938**, *10*, 424–425. See also pp. 464–467.
4. James, L.H. Reducing the Microbial Content of Spices. *Food. Ind.* **1938**, *10*, 428–429.
5. Jensen, L.B.; Wood, I.H.; Jensen, C.E. Swelling in Canned Chopped Hams. *Ind. Eng. Chem.* **1934**, *26*, 1118–1120.
6. Smith, H.W. Treated Spices Reduce Spoilage. *Food Ind.* **1940**, *12*, 50–72.
7. Yesair, J.; Williams, O.B. Spice Contamination and Its Control. *Food Res.* **1942**, *7*, 118–126.
8. Phillips, C.R.; Kaye, S. The Sterilization Action of Gaseous Ethylene Oxide. I. Review. *Am. J. Hyg.* **1949**, *50*, 270–279.
9. Alguire, D.E.; Yeung, A.C. Making Cosmetics Microbiologically Safe. *Cosmet. Toiletries* **1979**, *94*, 77–80.
10. Pinto, T.J.A. Aspectos Fundamentais na Validação do Monitor Biológico Para a Esterilização por Óxido de Etileno. Tese de Doutorado, Faculdade de Ciências Farmacêuticas—USP, São Paulo, 1991; 203 pp.
11. Lencioni, E.; Panzarasa, L. La Sterilizzazione Con Ossido Di Etilene. *Boll. Chim. Farm.* **1977**, *116*, 378–392.
12. Ernst, R.R.; Doyle, J.E. Sterilization with Gaseous Ethylene Oxide: A Review of Chemical and Physical Factors. *Biotechnol. Bioeng.* **1968**, *10*, 1–31.
13. Ernst, R.R.; Shull, J.J. Ethylene Oxide Gaseous Sterilization. I. Concentration and Temperature Effects. *Appl. Microbiol.* **1962**, *10*, 337–341.
14. Gunther, D.A. The Chemistry and Biology of Eto Sterilization. *MD & DI* **1980**, *6*, 31–35.
15. Golberg, L. *Hazard Assessment of Ethylene Oxide*; CRC Press: Boca Raton, FL, 1986.
16. Turchi, G.; Bonatti, S.; Citti, L.; Gervasi, P.G.; Abbondandolo, A. Alkylating Properties and Genetic Activity of 4-Vinylcyclo-hexene Metabolites and Structurally Related Epoxides. *Mutat. Res.* **1981**, *83*, 419.
17. Vogel, E.; Natarajan, A.T. The Relation Between Reaction Kinetics and Mutagenic Action of Monofunctional Alkylating Agents in Higher Eucaryotic Systems. II. Total and Partial Sex-Chromosome Loss in *Drosophila*. *Mutat. Res.* **1979**, *62*, 101.
18. Desai, P.R.; Buonicore, A.J. Engineering Controls to Minimize Worker Exposure to Ethylene Oxide at Sterilization Facilities. *Plant/Oper. Prog.* **1990**, *9* (2), 103–107.
19. Gunther, D.A. Evolution of Equipment, Techniques for Industrial EtO Sterilization. *Am. Perfume Cosmet.* **1970**, *85*, 35–38.
20. Lencioni, E.; Franchi, G.; Panzarasa, L. L'Autocertificazione nei Processi di Degermazione Con Eto. *Pharm. Acta Helf.* **1987**, *62* (10–11), 306–312.
21. Mayer, V.F.; Agostini, G.; Schaber, K.; Koch, A. Abscheilung von Ethylenoxide bei Sterilisationsprozessen in der Pharma Industrie. *Pharm. Ind.* **1989**, *51* (3), 294–298.
22. Buonicore, A.J.; Desai, P.R. EtO Emission Control Alternatives for Sterilization Facilities. In *A Safe Method in Safe Hands—A Practical Approach to Harmonization*; EUCOMED Conference on EtO Sterilization, Paris, 1989.
23. Britton, L.G. Thermal Stability and Deflagration of Ethylene Oxide. *Plant/Oper. Prog.* **1990**, *9* (2), 75–86.
24. June, R.K.; Dye, R.F. Explosive Decomposition of Ethylene Oxide. *Plant/Oper. Prog.* **1990**, *9* (2), 64–67.
25. Reichert, M. Low Temperature Sterilization Alternatives—What Are the Real Costs? *Surg. Serv. Manag.* **1995**, *1* (2), 38–43.
26. Christensen, D.E. Changing EtO Sterilizer Cycles to Reduce Ethylene Oxide Exposure Levels. *Med. Dev. Diagn. Ind.* **1984**, *6*, 27–35.
27. Alfa, M.J.; Degagne, P.; Olson, N. Bacterial Killing Ability of 10% EtO Plus 90% Hydrochlorofluorocarbon Sterilizing Gas. *Infect. Control Hosp. Epidemiol.* **1997**, *18* (9), 641–645.
28. Oliveira, D.C. Esteriliza ção Por Óxido de Etileno: Estudos De Efetividade Esterilizante de Misturas Não Explosivas e Compatíveis com a Camada de Ozônio. Dissertação, Faculdade de Ciências Farmacêuticas—USP, 2000; 159 pp.
29. Morrissey, R.F.; Bruch, C.W.; Sharbaugh, R.J.; Favero, M.S.; Jarvis, W.R.; Masefield, J. Sterility and Safety Assurance of Medical Devices. *MD & DI* **1992**, *14* (4), 78–81.
30. Manning, C.R. Validation EtO Packaging/Sterilizer Configurations. *MD & DI* **1990**, *1* (3), 52–58.
31. AAMI STANDARD. Biological Indicators for EtO Sterilization Processes in Health Care Facilities. Estados Unidos, 1986.
32. Booth, A.F. ISO 1135: New Standard Presents New Challenges. *MD & DI* **1994**, *16* (2), 64–67.
33. Fritze, D.; Pukall, R. Reclassification of Bioindicator Strains *Bacillus subtilis* DSM 675 and *Bacillus subtilis*



- DSM 2277 as *Bacillus atrophaeus*. Int. J. Syst. Evol. Microbiol. **2001**, 51, 35–37.
34. Pinto, T.J.A.; Saito, T.; Iossif, M. Ethylene Oxide Sterilization: III—Influence of Carrier Nature in a Biological Monitor Performance. PDA J. Pharm. Sci. Technol. **1994**, 48 (3), 155–158.
35. Paulson, D.S. Calculating D-Values for Steam Sterilization Processes. MD & DI **1995**, 17 (5), 198–204.
36. Seille, J.M.; Delattre, L.; Meurice, L.; Jaminet, F. Étude de l'Effet d'une Sterilisation a l'oxide d'ethylene sur les Teneurs Residuelles en Chlorhydrine du Glycol et en Ethyleneglycol Dans des Articles Medico-chirurgicaux a Base de PVC, Prealablement Irradies Au Cobalt 60. J. Pharm. Belg. **1985**, 40 (4), 213–221.
37. Rodriguez, A.C.; Young, B.; Caulk, K.; Zelewski, J.; Kwasnica, S.; Aguirre, S. Calculating Accumulated Lethality and Survivorship in EtO Sterilization Processes—The Availability of Two New Equations Makes It Possible to Compare the Effects of Dissimilar Process Cycles. MD & DI **2001**, 5, 100–107.

STERILIZATION BY MOIST HEAT

Dario Pistolesi

Fedegari Autoclavi S.p.A., Albuzzano, PV, Italy

INTRODUCTION

To sterilize something means to render it “aseptic” (from the Greek word *sepsis*, putrefaction, preceded by a privative *a*). In other words, it means to inactivate the micro-organisms that may produce this putrefactive action (in the broadest sense of the term).

Moist-heat sterilization is achieved when water vapor (or, more generically, moist heat, i.e., a suitable combination of temperature and humidity) at a definite temperature is introduced or generated (even indirectly) at the level of the micro-organisms to be inactivated and is maintained in such conditions for a definite time. As explained in detail hereafter, moist-heat sterilization proceeds as an inverse logarithmic progression. Therefore, only a treatment of infinite duration provides absolute certainty that all micro-organisms have been inactivated.

In pharmaceutical technology, to define an item as sterile, one must be able to demonstrate that on a statistical basis related to the processing conditions, no more than 1 in 10^6 units subjected to sterilization “may be” nonsterile. Therefore, the SAL (Sterility Assurance Level) of the product must be greater than (or equal to) 10^6 . The obvious consequence of this situation is that although the word sterile expresses an absolute concept, the word sterilized, understood as the result of an adequate sterilization process, has a probabilistic meaning.

Current pharmacopeias, standards, and guidelines related to sterilization generally use the following type of wording:

... Sterilization is a special process because its efficacy cannot be verified by simple inspection and testing on the final product.... For this reason, sterilization processes have to be validated before use, the performance of the process monitored routinely and equipment regularly maintained....

Accordingly, installation qualification and operational qualification of the sterilizer and validation of the processes, combined with continuous monitoring of each individual routine process, are now considered fundamental (1). Therefore, sterility tests (i.e., microbiological tests performed on the final product) have lost much of the

significance they had in the past (2), but they are still being performed. So-called parametric release (i.e., release based on the evaluation of tightly controlled physical parameters of each process, without performing sterility tests) is in fact theoretically accepted by many pharmacopoeias, but is (as stated, for example, in the *European Pharmacopoeia*) “... subject to the approval of the competent authorities ...,” which are generally highly reluctant to grant such approval.

All pharmacopeias consider moist-heat sterilization as the method of choice, i.e., the method to be preferred, unless, of course, the product to be sterilized is incompatible with the characteristics of steam. The reason for this preference is the fact that moist-heat sterilization provides the best combination of flexibility, reliability, and low equipment and operating costs.

OVERVIEW

Moist-heat sterilization is achieved when a suitable combination of temperature and humidity can be introduced (or indirectly generated) at the level of the micro-organisms to be inactivated. The classic way to achieve this is by means of pressurized saturated steam at the temperature of 121°C (250°F). However, other sterilizing media (e.g., superheated water or a steam–air mixture) are also frequently used to obviate certain problems that pure steam may pose. Sometimes the load is rotated inside the chamber of the sterilizer to achieve particular results.

MOIST-HEAT STERILIZATION KINETICS (3)

Consider a system contaminated by a single microbiological species (i.e., an ampoule containing an aqueous suspension of a given micro-organism) immersed in pressurized saturated steam at constant temperature. One can demonstrate experimentally that the reaction of thermal inactivation of the micro-organism develops as a first-order chemical reaction (i.e., as a chemical decomposition

reaction) in which the reaction rate is proportional at all times only to the amount of product still to be degraded. The proportionality coefficient is typical of the species and conditions of the given micro-organism. All this seems obvious for dry-heat sterilization, but not for steam sterilization, in which the water vapor molecules would appear to take part in the reaction. Actually, this bimolecular reaction is a first-order reaction, because an excess of steam is always present and its concentration can be considered constant.

The most widely used mathematical equation of the above is:

$$N = N_0 10^{-t/D} \quad (1)$$

where N_0 = initial number of micro-organisms; t = elapsed time (or sterilization time); N = number of surviving micro-organisms after exposure time t ; and D = "decimal decay time," defined as the time interval required, at a specified and constant temperature, to reduce the microbial population to 1/10 of its original quantity.

At 121°C, the D -values generally oscillate between 0.2 and 1.5 min for the various microbial species that can be encountered in pharmaceutical activity. Eq. 1 allows for two important conclusions:

1. the time required to reduce the micro-organism population to any preset value is a function of the initial concentration, and
2. the effect of sterilization in the same conditions (T and t) will be very different according to the D -value of the contaminating micro-organism.

Fig. 1 shows that the same reduction ratio is achieved for different microbial species (at the same constant

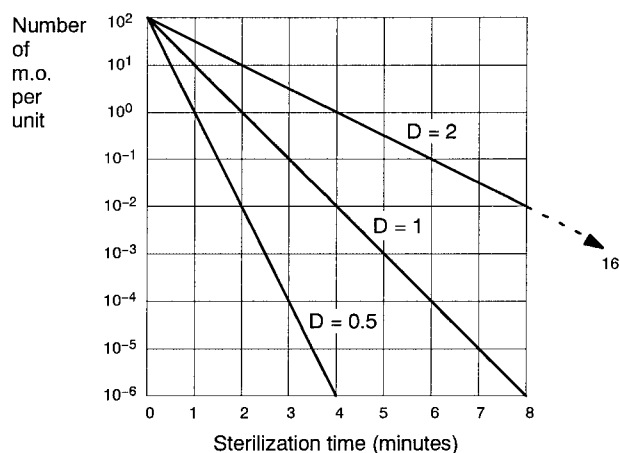


Fig. 1 Death rate curves illustrating decimal reduction concept. (From Ref. 3.)

temperature) with an exposure time that is proportional to the D -value of each species.

Consider a batch of units (e.g., a batch of filled ampoules) with a constant initial population for each unit of 100 micro-organisms or 10^2 . If $D_{121} = 1$, after 1 min at 121°C, the population will be reduced to $10^1 = 10$. After another minute, only $10^0 = 1$ micro-organism will still be surviving. After another minute, the surviving population would be $10^{-1} = 1/10$ micro-organism per unit. In biological terms, such a contamination is obviously meaningless; statistically, it means that there is a probability that $\frac{1}{10}$ of the units of the sterilized batch are still contaminated. Clearly, after another 5 min of sterilization, this probability will be reduced to $1/10^6$ or 10^{-6} . In other words, the SAL (introduced earlier) is 10^{-6} .

A more reassuring SAL, for example 10^{-9} , is very often sought. It is sufficient to extend the sterilization for just 3 additional min. The problem, therefore, is evidently not cost-related; rather, it is simply linked to the risk of subjecting the treated material to thermal degradation.

All the above considerations have been made under the assumption that the temperature is kept constant during the sterilization period. Obviously, the D -value changes when the temperature changes. When the D -values obtained experimentally for a given microbial species are plotted on a semilogarithmic chart as a function of the temperature T , a path such as the one shown in Fig. 2 is obtained.

Clearly, if D is 1.0 at 121°C, it is 0.1 at 131°C and 10 at 111°C. In other words, the value of D decreases or increases by a factor of 10 when the temperature increases

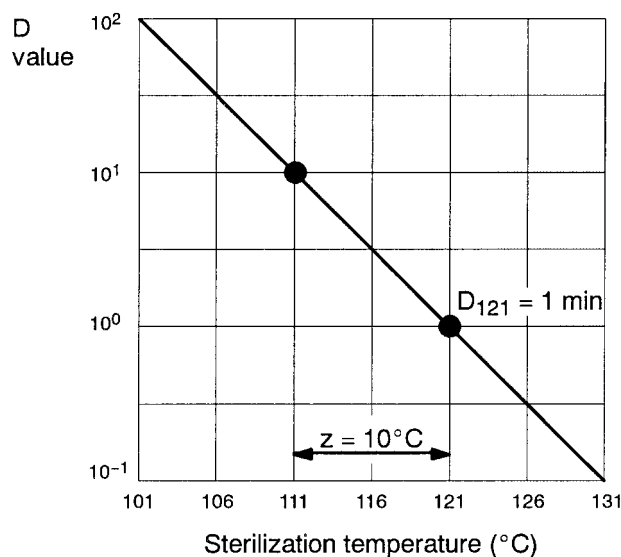


Fig. 2 Logarithm of D decreases linearly as temperature increases. (From Ref. 3.)

or decreases by 10°C. The algorithm z is defined as temperature coefficient of moist-heat sterilization, i.e., the number of degrees of sterilization temperature that causes a 10-fold variation of D or of the sterilization rate. Depending on the micro-organism being considered, the z -value varies between 5 and 15 for the 100–130°C sterilization range. The z -value is frequently assumed to be equal to 10 in the absence of precise experimental data (1).

It is evident that small temperature variations cause dramatic variations in the rate of the sterilization reaction. It is easy to calculate that a variation by only 1°C in the vicinity of 121°C causes a variation of approximately 24% in the value of D , i.e., of the sterilization rate.

F_0 or Equivalent Time

To compare the lethal effect of a sterilization performed for any given time t_x at any temperature T_x (which may vary over the time t_x), it is very useful to be able to express this lethal effect by relating it by calculation to a given reference temperature. When this reference temperature is 121°C (or, more exactly, 121.11°C, which correspond to 250°F) and z is assumed to be 10 (or 18 if the temperature is expressed in °F), the resulting algorithm is known as F_0 and is expressed by:

$$F_0 = \Delta t \sum 10^{\left(\frac{T-121}{z}\right)} \quad (2)$$

where Δt = time interval between two successive temperature measurements; T = actual sterilization temperature in °C at the time t ; and z = temperature coefficient, assumed to be equal to 10.

F_0 is known as equivalent time because its dimension is actually a time expressed in minutes. Clearly, when the values in Eq. (2) are, for example, $\Delta t = 15$ min and $T = 131^\circ\text{C}$, F_0 is 150 min, i.e., the lethal effect is 10 times higher than that of a sterilization lasting 15 min at 121°C. If instead $\Delta t = 15$ min but $T = 111^\circ\text{C}$, F_0 is equal to 1.5 min, i.e., the lethal effect is 10 times smaller than that of a 15-min sterilization at 121°C. An F_0 of 12 (delivered to the coldest point of the load) is generally considered sufficient for adequate sterilization in the pharmaceutical field (4).

MOIST-HEAT STERILIZATION PROCESSES

Current pharmaceutical production practice uses substantially three moist-heat sterilization processes: 1) pressurized saturated steam; 2) superheated water; and 3) steam–air mixture. Process 1 is the traditional multipurpose

process, which obviously uses pure pressurized saturated steam as sterilizing medium. Processes 2 and 3 are so-called counterpressure processes; they were introduced in pharmaceutical production practice approximately 20 years ago and, respectively, use a spray superheated water and a homogeneous mixture of steam and air as sterilizing media. These processes allow the control of the pressure of the sterilizing medium independently of its temperature (which is impossible to accomplish with pure saturated steam). As explained below, these processes are used to treat solutions in containers that cannot tolerate the internal overpressure that is generated inside when sterilized with process 1.

PRESSURIZED SATURATED STEAM METHOD

This is certainly the most widely used and most versatile moist-heat sterilization method. Accordingly, it is widely used not only for sterilization of pharmaceutical products but also for laboratory and hospital sterilization and for the treatment of medical devices. Nonetheless, it has significant limitations, especially in pharmaceutical use, which are described later. The sterilizing medium is obviously pure pressurized saturated steam. The word saturated means that the steam is in thermodynamic equilibrium with its liquid form (water) at the temperature being considered.

Typical operating conditions are 121°C (i.e., 250°F) for 15 min (or even less); this temperature is matched by a saturated steam pressure of 2.05 abs bar (i.e., 205 kPa). However, higher or lower temperatures (and, therefore, pressures) are often used, with obvious appropriate adjustments of the holding time.

The term dry saturated steam is sometimes used. It should be made clear that dryness is a theoretical condition of steam and that in practice, moist saturated steam is used. This also provides assurance that the steam really is saturated and is not superheated. The use of superheated steam may in fact cause problems in process management.

However, the steam must entrain the smallest possible amount of condensate. The term steam dryness fraction defines the amount of condensate entrained by the moist steam. A dryness fraction of 0.95 means that 100 g of moist steam consist of 95 g of dry saturated steam plus 5 g of condensate, which is (or should be) at the same temperature the steam. A dryness fraction of 0.95 is considered the lower limit of adequacy for moist-heat sterilization.

The reliability of sterilization performed by means of saturated steam is based on three essential characteristics of this medium:

1. When steam condenses, it releases heat at a constant temperature and in very large amounts; 1 kg of steam condensing at 121°C (transforming into water at 121°C) releases as much as 2200 kJ (or 525 kcal).
2. The temperatures and pressures of saturated steam have a two-way correlation. Once the steam temperature is determined, so is its pressure, and vice versa. Saturated steam at 121°C inevitably has a pressure of 2.05 abs bar; saturated steam at 3.04 abs bar inevitably has a temperature of 134°C. This entails two very interesting practical possibilities: a pure saturated steam autoclave can be equally temperature- or pressure-controlled, and regardless of the parameter used for control, the second parameter can easily be used to cross-monitor the first.
3. One gram-molecular weight of water (18 g, i.e., 18 ml in the liquid state) as steam at 121°C and 2.05 abs bar occupies a volume of approximately 15 L. This means that when steam condenses at 121°C, it shrinks in volume by almost 1000 times. Consequently, additional steam spontaneously reaches the material to be sterilized. The condensate that forms can be easily removed from the autoclave chamber by means of a condensate trap or by continuous bleeding.

Apart from these three favorable characteristics, other phenomena linked to the use of pure saturated steam must be considered:

1. To perform its microbiological inactivation action, the steam must come into contact with the micro-organisms. This can occur directly or indirectly. It occurs directly when the steam makes contact, for example, with a surgical instrument located in the autoclave chamber. It occurs indirectly when the steam is generated, for example, inside a sealed ampoule that contains an aqueous solution by heat exchange with the steam in the chamber. However, it is evident that it is impossible to steam-sterilize the inside of a closed empty ampoule or its contents if they are, for example, an anhydrous oil-based solution.
2. The air initially present in the chamber and the incondensables (generally CO₂) possibly entrained by the steam have molecular weights, and therefore densities, which are 1.5–2.0 times higher than those of steam. Therefore, at the beginning of the process, the air must be removed from the chamber, and the steam must not contain incondensables. Otherwise, they tend to stratify in the lower portions of the chamber, producing unacceptable temperature gradients.

3. When closed nondeformable containers with aqueous solutions are sterilized, the pressure inside can reach values far higher than the chamber pressure. The reasons for this are explained in detail later but, in any case, the internal overpressure can reach or exceed 1.4 bar and can be intolerable for many types of container. In such cases, it is necessary or convenient to use counterpressure autoclaves (as described later).

SATURATED STEAM AUTOCLAVES

Construction

All sterilizers intended for pharmaceutical use are currently made of class AISI 316 stainless steel, including the valves and piping. Other materials may be acceptable only for service components arranged downstream of the sterilizer (e.g., the vacuum pump or the condensate trap). The service elements arranged upstream of the sterilizer (e.g., heat exchangers or water pumps) also must be made of stainless steel.

Silicone rubber or PTFE and its derivatives are generally used for gaskets (for doors, valves, etc). The chamber of these autoclaves is horizontal, with a rectangular or (rarely) cylindrical cross-section. The dimensions of these chambers can vary considerably, from approximately 100 L to 10 m³ or more.

Doors

Doors are generally rectangular, even though the chamber is cylindrical. In this case, the doors are inscribed in the circumference. There may be one or two doors: two doors are always used when the autoclave leads into a sterile room. Two-door autoclaves are often used when this requirement does not apply, but the need is nonetheless felt to separate the loading area (where nonsterile products are placed) from the unloading area (where only already-sterilized products can be placed). The doors may be of various kinds.

Side-hinged, manually operated doors retained by radial locking bars

At present, these are the most widely used doors. The rim gasket is solid and fixed. The radial bars are moved, during closure/opening, by a central hand-wheel that is operated manually. This locking system requires lubrication, and this can entail microbiological problems, especially if the

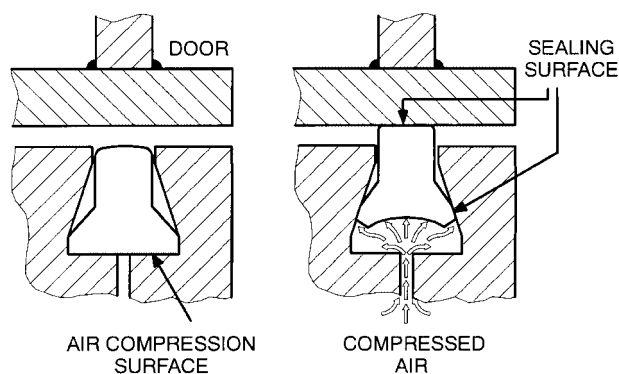


Fig. 3 Dovetail section gasket activated by compressed air. (From Ref. 5.)

door opens onto a sterile room (the closure method that uses perimetric eye-bolts is now obsolete).

Vertically or horizontally sliding, automatically operated doors

These doors have no lubrication problems, but in the horizontally sliding version (which is common for industrial-size autoclaves), they have the drawback that they occupy a considerable amount of floor space. The gasket is generally located on the chamber rabbet and is compressed toward the door by two methods: 1) the gasket is hollow and inflatable (by means of compressed air or steam); and 2) the gasket has a particular dovetail cross-section and is contained in a slot that also has a dovetail cross-section. An adequate pressure of compressed air in the rear part of the slot is sufficient to activate the seal on the door, and the release of the pressure (without requiring vacuum) is sufficient to activate the retraction of the gasket (Fig. 3).

Diagonally moving doors

These doors combine the positive features of the two preceding types. During opening (and, in reverse, during closure), the door is automatically moved slightly upward and laterally, enough to disengage it from the mechanical systems that retain its four sides. Then the operator opens the door manually by turning it about its side hinges. The gasket is generally identical to the second one described previously.

Jacket

Saturated steam autoclaves are generally provided with a jacket, that is constituted by a second wall that more or less fully encloses the inner chamber and thus forms a secondary space around it. For the sake of brevity, the various jacket construction methods are not presented here. The purpose of a jacket is summarized as follows:

- 1) to preheat the autoclave initially and keep it warm during loading/unloading;
- 2) to preheat the load during the initial air removal phase;
- 3) to contribute to the drying of the load in the final vacuum phase;
- and 4) to reduce any condensate entrained by the steam.

Steam can be fed to the jacket-chamber assembly by:

- 1) Single feed in which the steam first enters the jacket, circulates inside it, flows out of it, and then enters the chamber. The reduction of condensate entrained by the steam can be achieved only with this single-feed approach.
- 2) Separate-feed in which, of course, two controls are provided. The jacket may be fed with plant steam, but the chamber is certainly supplied with ultraclean steam. At present, this is the solution generally used for modern autoclaves because it ensures that no microbiological or particulate contamination can reach the chamber from the jacket, which is a closed and convoluted space that is practically impossible to clean and inspect accurately.

Process

Initial air removal from the chamber

The basic reason for air removal has been addressed previously. Moreover, loads are often made of porous or packaged materials, which require reliable and rapid removal of air from the inside of the loads as well.

Today, the so-called gravity removal method is used only for special tasks. In this method, the steam enters the top part of the chamber and is distributed by a suitable sparger (theoretically on a uniform front). The air escapes through a large valve in the lower part of the chamber by way of two actions: gravity and displacement by steam. This method is rather slow and unsuitable for porous and other loads that may trap air inside recessed cavities.

A modern autoclave is generally equipped with a water-ring pump that can produce a vacuum of approximately 70 residual mbar in the chamber. Accordingly, almost 7% of the air is not removed. The following two methods are essentially used for completing air removal.

Pulsed vacuum: Once the maximum initial vacuum has been reached, the pump is stopped and steam enters the chamber until an approximate atmospheric pressure (or a higher pressure) is reached; then vacuum is produced again. Three vacuum/pressure pulses are generally sufficient to achieve suitable air removal.

Dynamic vacuum: Once the maximum initial vacuum has been reached, the vacuum pump is kept running while a 5- to 10-min. steam-injection is performed from the side of the chamber that lies opposite the vacuum drain point. Modern autoclaves can perform both methods depending on the load to be sterilized.

Heating and sterilization phases

Considerable amounts of condensate form in the chamber during the heating and sterilization phases. This condensate must be removed, and there are basically two ways to accomplish this. The first uses a condensate trap at the bottom of the chamber. This is the cheapest and simplest method, but it causes significant drops in pressure (and therefore in temperature) when the trap opens, owing to the inertia of the trap. The second method uses dynamic steam. This is the most reliable method, but it is also slightly more expensive. During the heating and sterilization phases, the vacuum pump runs continuous and extracts the condensate through a small valve. A small amount of steam of course is also extracted continuously, accordingly providing the dynamic condition of the steam.

Poststerilization phases

These phases may be very different and are clearly linked to the sterilized material and to the required results. The most common solutions are the following:

Drying-cooling final vacuum: This is produced by restarting the vacuum pump until a preset vacuum (e.g., 100 mbar) is reached. The pump is then kept running for a preset time. Porous materials (and nonporous materials also) are thus dried and cooled quickly.

Cooling by circulating cold water in the jacket: This method is used with containers that are partially filled with solutions (for example, culture media) and closed with nonhermetic closures. With such containers, drying-cooling final vacuum is not applicable because the solution would boil, and cooling by direct spray (described hereafter) may cause contamination. Steam is removed from the chamber by introducing sterile air at a pressure that is equal to, or greater than the sterilization pressure. Then cold water is circulated in the jacket. Chamber air pressurization has two purposes: to prevent boiling of the solutions and to improve heat exchanges between the load and the jacket.

Cooling by direct spraying of cold water onto the load: This method is generally used for cooling filled and sealed ampoules contained in perforated trays and generally marshaled. It is performed by spraying, or rather by nebulizing, purified water or water for injection onto the load by means of a sparger located in the ceiling of the chamber. Water nebulization produces rapid steam condensation and an equally rapid pressure drop in the chamber while the pressure inside the ampoules remains high (because the temperature of the solution decreases rather slowly). However, good-quality ampoules can withstand this treatment adequately. The water spray is generally stopped when the load temperature reaches

70–80°C. Accordingly, the load still contains enough heat energy to dry spontaneously once removed from the autoclave.

Cooling with cold water sprayed directly onto the load with air counterpressure: Very frequently, the pressure stress that occurs when using method of cooling by direct spraying of cold water, cannot be tolerated by the load. In such cases, it is possible to drain the steam from the chamber by replacing it with sterile compressed air at a pressure equal to or greater than the sterilization pressure. Cooling water is sprayed onto the load only after this replacement has been performed. However, it is obvious that this method only allows for reduction of the pressure stress of the containers in the cooling phase; the pressure stress in the sterilization phase (discussed later) is unavoidable.

Ampoule tightness tests: The purpose of these tests is to allow for rejection of ampoules that have closure defects, fractures, or cracks. These tests fall essentially into two categories: penetration of dyed solutions (usually with methylene blue) in the ampoules and poststerilization pressure stress. Details of these methods are not presented here because of space constraints.

Sterilization of the air that enters the chamber

As noted previously, in many cases it is necessary to introduce air in the sterilization chamber. This air must be sterile, otherwise it recontaminates the sterilized load or the sterile room if a two-door autoclave is connected to it. This air is generally sterilized by filtration with a system built into the autoclave. Therefore, it is necessary to:

1. Provide a filtration cartridge with suitable retention.
2. Allow in situ periodic sterilization of the assembled system by means of an automatic process of the autoclave.
3. Ensure that the filtration system and its piping maintain sterility during successive sterilization programs used for production.
4. Perform the system integrity test before and after each sterilization program of the filtration system.
5. Allow for validation of all procedures noted earlier.

Process controllers

Today, process controllers installed in autoclaves are based on programmable logic controllers (PLCs), personal computers (PCs), customized electronic solutions, or, sometimes, different combinations of the aforementioned systems. However, a very large number of autoclaves managed by old electropneumatic systems are still in operation. Modern process controllers, of course, offer previously inconceivable levels of performance. Today,

temperature and/or pressure control is generally performed with a proportional-integral-derivative algorithm. Sterilization can be time-managed or F_0 -managed (F_0 being accumulated by several flexible temperature probes enabled for this function). Some management systems offer exceptional flexibility in composing programs and setting parameters. Information provided in real time on cathode ray tube (CRT) or liquid crystal display (LCD) or produced/stored on paper/electronic media is highly detailed.

DIFFERENTIAL PRESSURE BETWEEN INSIDE/OUTSIDE OF A RIGID CONTAINER, PARTIALLY FILLED WITH WATER SOLUTION AND SEALED, DURING STEAM STERILIZATION

When a container in the conditions noted earlier is sterilized in a conventional autoclave that operates with pure saturated steam, during sterilization a considerable overpressure with respect to the pressure inside the autoclave

chamber is generated in the container. This is clearly attributable to the fact that the air (or gas) that was present at filling has remained in the container, whereas the air was eliminated from the autoclave chamber at the beginning of the process. Fig. 4 schematically explains the phenomenon in ideal conditions, i.e., considering air a perfect gas.

Experimentally, it turns out that the actual overpressure is higher than the theoretical one. This is attributable to various facts: the thermal expansion of water is greater than the thermal expansion of the glass of the container; the solution contains dissolved gases that come out of the solution as the temperature rises; air is not a perfect gas. Obviously, the overpressure depends on the filling temperature, the sterilization temperature, the ratio between solution volume and head volume, etc., but at 121°C, it is on average approximately 1.4 bar. Clearly, this phenomenon cannot be ignored: suffice it to note that the stopper of glass bottles with a mouth having a cross-section of approximately 4 cm² would be subjected to an expulsion force of approximately 6 kg.

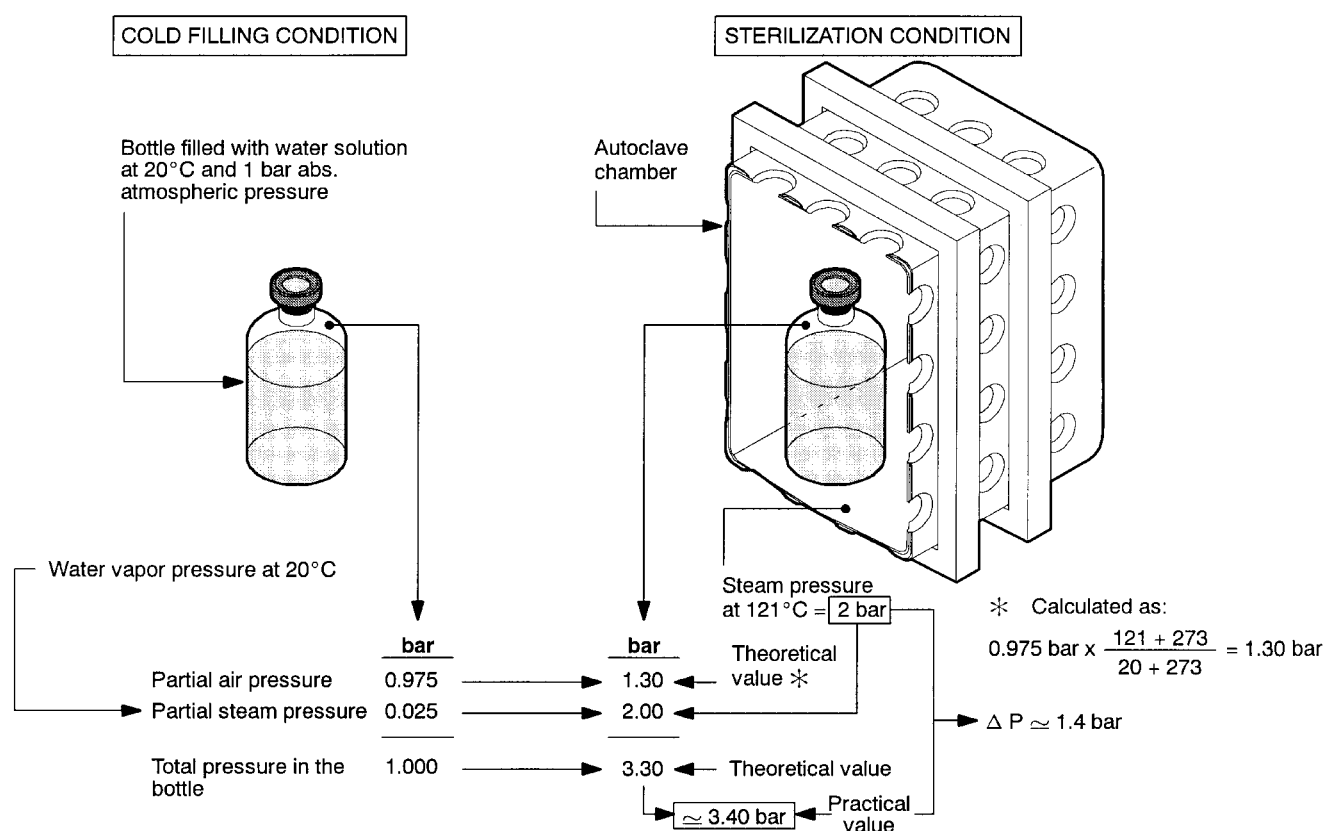


Fig. 4 Schematic description of pressures produced inside a rigid container, partially filled with water solution and sealed, during steam sterilization at 121°C. (Adapted from Ref. 5.)

These conditions therefore prohibit the use of traditional pure saturated steam autoclaves to sterilize solutions contained in a wide variety of containers such as

1. Large-volume parenterals (LVPs) in glass bottles;
2. LVPs and small volume parenterals (SVPs) in plastic containers (flexible, semirigid, rigid);
3. Prefilled glass or plastic syringes;
4. Jars and similar containers with press-on or screw caps;
5. Blisters containing various materials, for example, disposable contact lenses.

To correctly sterilize these products, it is necessary or advisable to use a counterpressure autoclave.

COUNTERPRESSURE MOIST-HEAT STERILIZATION

Moist-heat autoclaves operating with counterpressure are sterilizers capable of controlling the pressure of their sterilizing medium independently of its temperature. They

are used essentially for the terminal sterilization of solutions.

Accordingly, a dual control principle is provided that acts independently on both parameters. Two methods currently in use are superheated water spray and steam–air mixture.

SUPERHEATED WATER SPRAY AUTOCLAVES

Fig. 5 is a typical diagram of these autoclaves. Alternatives are possible, but they do not alter the essential structure. The chamber is horizontal and generally cylindrical, with a single wall and rectangular door(s) inscribed in the circumference.

At the beginning of the process, after loading the product, the lower part of the chamber is filled with water of adequate chemical and bacteriological quality. The air contained in the chamber is not removed. A sanitary-type pump circulates the filling water through a heat exchanger (of the removable-plate or other sanitary

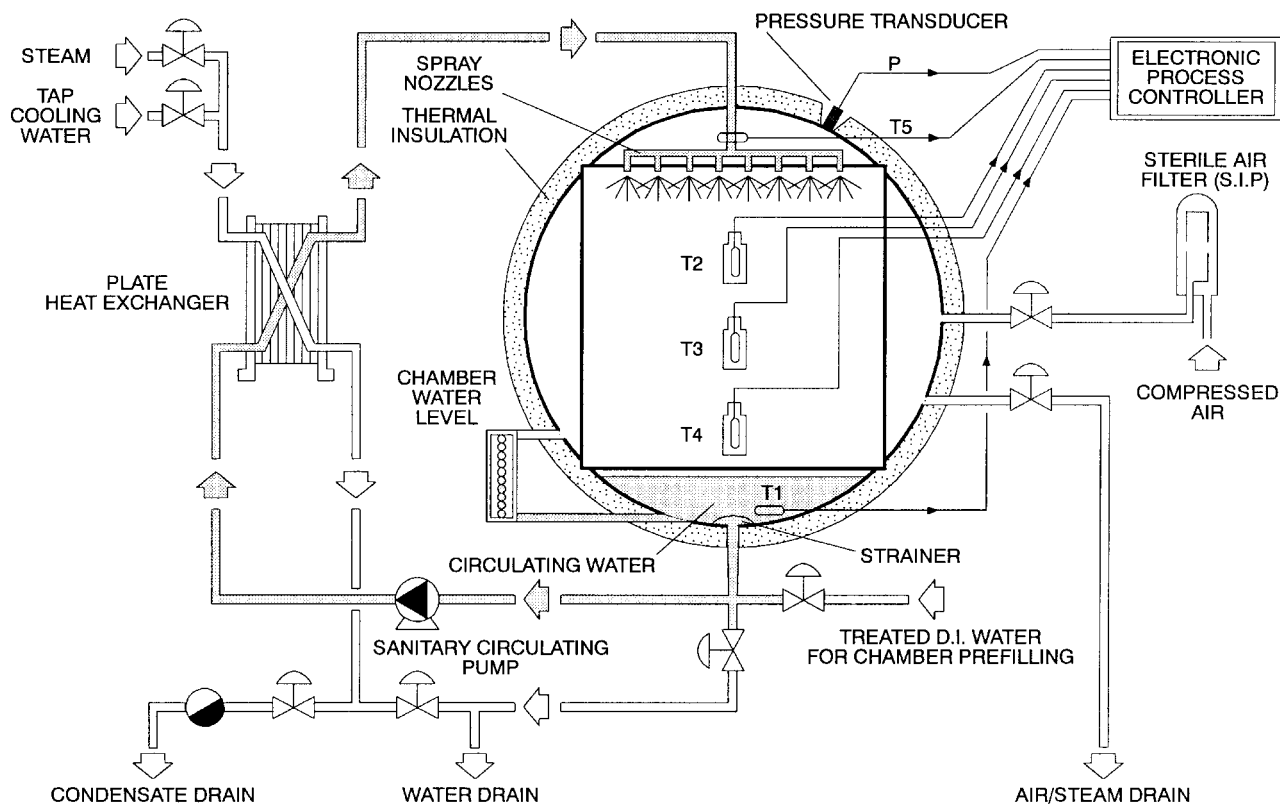


Fig. 5 Superheated water spray autoclave: simplified P.&I.D. (Adapted from Ref. 5.)

type) that is indirectly heated in countercurrent with plant steam. The water is then sprayed onto the load by a sparger located in the upper part of the chamber and equipped with a system of solid-cone spray nozzles. Uniform water redistribution in the lower layers of the load is ensured by suitable perforated racks that support the product. Sometimes additional water spray bars are located on both sides of the chamber.

Heating of the circulating water and, therefore, of the load is very gradual but quite rapid. A temperature of 121°C is typically reached in 25–30 min inside 500-ml containers; the heating rate clearly depends on the characteristics of the solution and its containers. Temperature uniformity in time and space during the sterilization phase is generally very good: much better than $\pm 1.0^\circ\text{C}$. The cooling phase is performed by the same circulating water, which is now sterile and continuously recirculated through the heat exchanger, in which cold water (instead of steam) now flows without contact with the sterile circulating water. The temperature inside 500-ml containers drops to approximately 80°C in 10–12 min.

This temperature is generally suitable to obtain rapid and spontaneous drying of the load once removed from the autoclave.

An appropriate partial pressure of air (sterilized by filtration) is maintained in the chamber during every phase of the process, to compensate for the overpressure inside the containers. Various methods for controlling the total chamber pressure (steam + air) can be used. With computerized process controllers, it is also possible to correlate at any time during each phase the air partial pressure to the average of the solution temperatures of two or more reference containers.

Consequently, the load suffers no thermal or pressure shocks because the differential pressure between the chamber and the containers can be reduced to zero or maintained at all times during the process in a direction that is suitable for the particular type of container during sterilization or, generally, during thermal treatments from 60 to 127°C.

Clearly, these autoclaves have some limitations in their application:

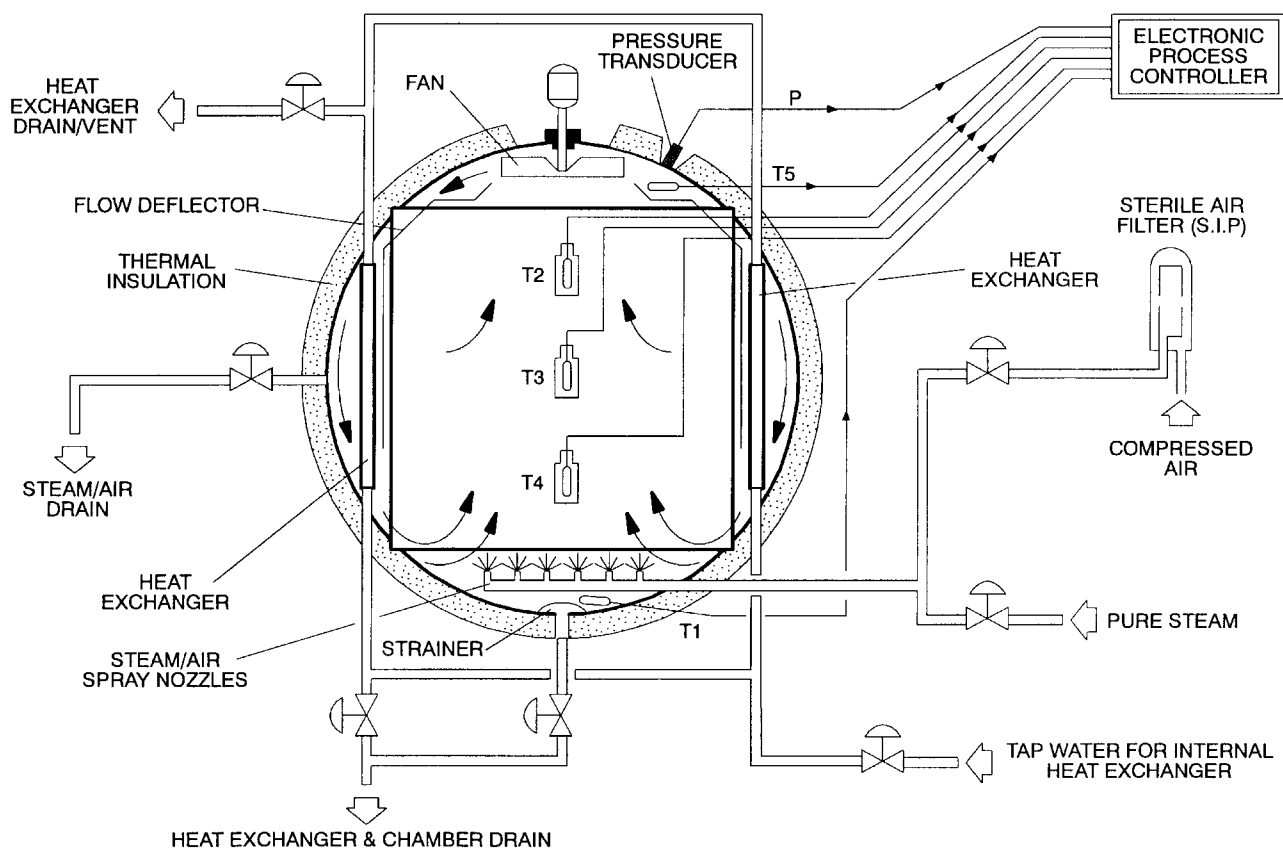


Fig. 6 Steam-air autoclave: simplified P.&I.D. (Adapted from Ref. 5.)

1. It is impossible, or illogical, to dry the load inside the autoclave by pulling vacuum in the chamber or by circulating warm air through the chamber and the load.
2. If materials that have upward-facing concave surfaces are sterilized, these surfaces will be filled with water at the end of the process. The obvious remedy is to load the material upside down.
3. When sterilizing solutions contained in PVC bags, so-called blushing can occur, i.e., the PVC can whiten because of water absorption. The time required for this blushing to disappear can be quite long, depending on the type of PVC and its plasticizer. Blushing does not occur with polypropylene (PP), polyethylene (PE), and poly laminated plastics.

These autoclaves are sterilizers that can vary considerable in size but are generally rather large (1–20 m³ and

more). They are often provided with automated loading/unloading systems.

STEAM–AIR MIXTURE AUTOCLAVES

Figure 6 is a typical diagram of these autoclaves; possible alternatives are addressed later. The chamber is similar to that of superheated water spray autoclaves. At the beginning of the program, steam enters the chamber directly through a suitable sparger located in the lower part of the chamber. The air initially contained in the chamber is not removed. The high-efficiency fan(s) located on the ceiling of the chamber and the flow deflector system have the task of homogenizing and circulating the steam–air mixture that forms inside the chamber. This is an important

	Superheated Water Spray = SWS	Steam–Air Mixture = SAM
Temperature uniformity in time	Good	Good
Temperature uniformity in space	Good	Good
Total pressure control	Good	Good
Counterpressure control	Good	Good
Consumption of high quality water (WFI)	Modest, for initial filling	No
Consumption of tap water for cooling	Acceptable	Approx. 3 times higher than SWS
Consumption of compressed air	Acceptable	Acceptable
Consumption of industrial steam	Acceptable	No
Consumption of ultraclean steam	No	Acceptable
Condensate recovery	Possible and easy	Not possible
Cooling water recovery	Possible, recovered water is initially very hot	Possible, recovered water is initially very hot
Autoclave price	Acceptable	Approx. 10% higher than SWS
Total process duration	Short	Approx. 30% higher than SWS
Autoclave productivity/price	High	Approx. 30% lower than SWS
Operating principle	Quite simple and straightforward	More complex than SWS
Overall machine design	Simple	More complex than SWS
Autoclave qualification/Validation	Normal	Normal
Operating flexibility according to type of load	Suitable for any kind of container with the following remarks: – upward concavities collect water – product is unloaded wet – PVC bags may generate “blushing” effect	Suitable for any kind of container: – upward concavities collect condensate only – some kind of containers may be unloaded slightly damp – limited “blushing” effect of PVC bags
Possibility of combination with saturated steam processes	Not recommended	Feasible and moderately expensive

Fig. 7 Schematic drawing of magnetically driven fan. (Adapted from Ref. 5.)

and demanding task because the air clearly tends to stratify on the bottom. The condensate that forms is removed by continuous and spontaneous bleeding from the chamber.

The cooling phase consists of feeding compressed and sterile air to the chamber to condense and replace all the steam, while maintaining the same total sterilization pressure or possibly increasing it. Cold plant water is then fed to the internal heat exchangers, which are constituted by batteries of hollow plates arranged in the two lateral sectors of the chamber (for simplicity, only one plate is shown in Fig. 6). However, this cooling method uses two solid-gas heat exchanges, which have poor efficiency. An attempt can be made to improve efficiency by increasing the air pressure in the chamber within the limits of the product, thus increasing the density of the air and therefore its exchange efficiency. The fans of course continue to run during the cooling phase as well.

Despite this refinement, the cooling phase is significantly longer than that in superheated spray water autoclaves. A mechanically critical point of steam–air autoclaves is the tightness of the fan shaft. This problem has been completely solved in the more advanced machines by adopting magnetically driven fans (Fig. 7). The air partial pressure during the program is managed as described above for superheated water spray autoclaves, and the dimensions and loading/unloading systems are also similar.

Possible alternatives to the configuration shown in Fig. 6 are 1) horizontal fans (instead of vertical fans) located on one side of the chamber. This solution entails a more severe risk of shaft bending and vibration, which can cause wear of the delicate sealing system. Moreover, it is technically more difficult to manufacture magnetically driven fans with a horizontal shaft; and 2) shell-and-tube heat exchangers (instead of plate-type exchangers).

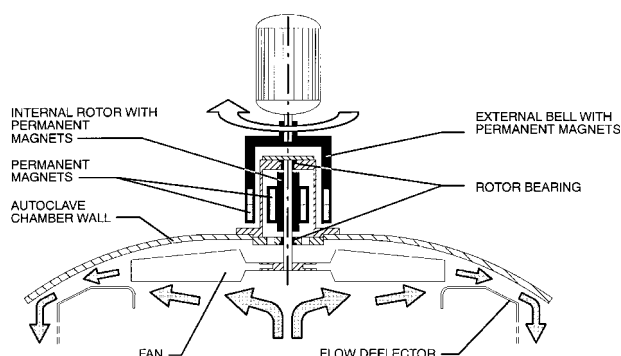


Fig. 8 Critical comparison of superheated water spray (SWS) and steam–air mixture (SAM) autoclaves. (Adapted from Ref. 5.)

With steam–air autoclaves, blushing of PVC bags is generally less intense than with superheated water spray autoclaves and essentially affects only the areas where the bags rest on the supporting racks. Among the positive features of steam–air autoclaves is the relative ease in combining the traditional pure saturated steam cycles, i.e., in manufacturing hybrid pure steam/steam–air autoclaves (in this case, the chamber is equipped with a jacket and a vacuum pump). This combination is instead not recommended for superheated water spray autoclaves, although it is offered by some manufacturers. Figure 8 is a summary comparison of superheated water spray and steam–air autoclaves.

STERILIZING A ROTATING LOAD

Currently, the pharmaceutical market more and more often requires rotating-load sterilization autoclaves. Load rotation can have essentially three goals:

1. To maintain the stability (or homogeneity) of emulsions (or suspensions) that would tend to break out because of the sterilization temperature.
2. To sterilize heat-sensitive products at high temperature, drastically reducing the sterilization time. The logic behind this principle is that, as noted, the temperature coefficient of the moist-heat sterilization reaction is very high (z is on the average equal to 10), whereas the temperature coefficient of a classic thermal degradation reaction is much lower (on the average equal to 2). Obviously, to achieve the goal, the product heating/cooling sterilization rates must be very high and uniform. Because rotation stirs the product, it indeed facilitates the penetration/removal of heat into/from the product, especially if it is dense and viscous.
3. To provide the best possible testing of ampoule tightness with fast poststerilization vacuum. This testing method (presented above) achieves maximum effectiveness when the “open” defects of the ampoules are below the level of the solution. Ampoule rotation is the ideal method for achieving this condition regardless of the location of the defect on the ampoule (tip, shoulder, bottom).

Naturally, the production of this type of autoclave requires highly refined design and construction technology because 1) the load rotation system complicates

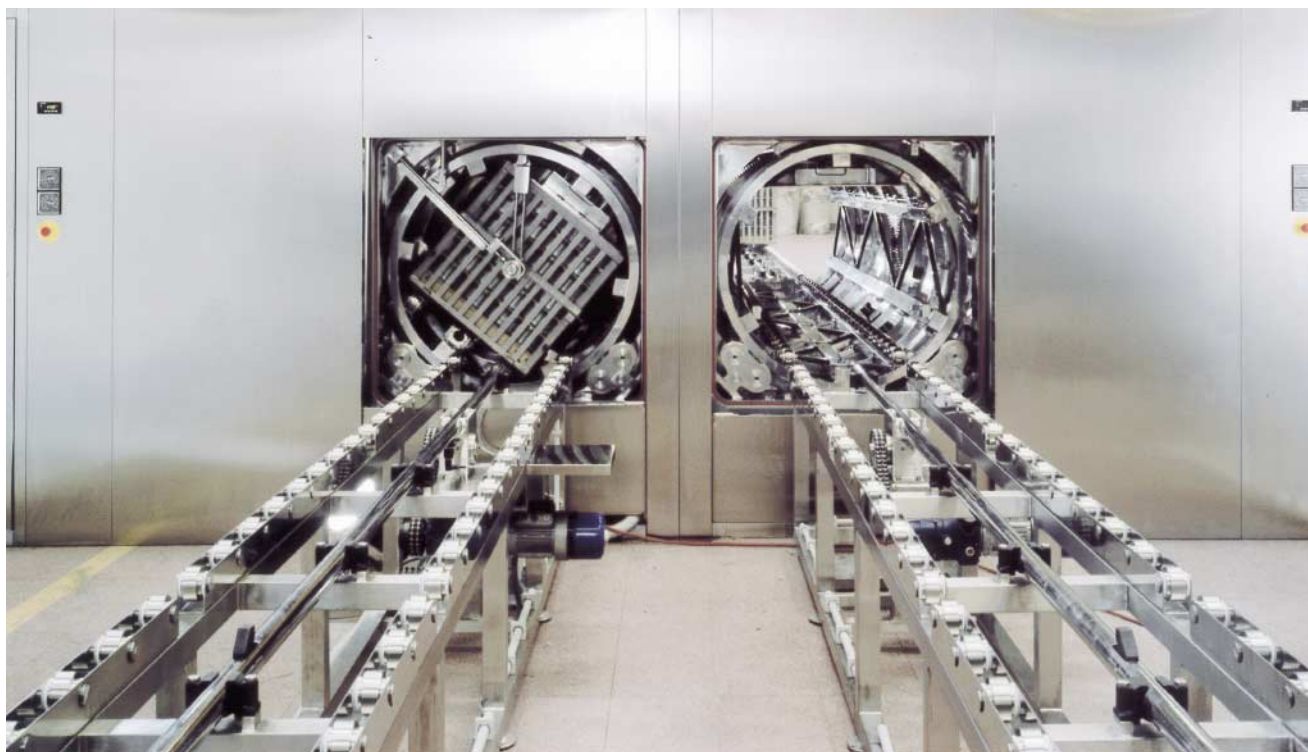


Fig. 9 Two superheated water spray autoclaves with a chamber capacity of approximately 4 m³, with a rotating load and sliding double doors. The man-machine interface of the process controller is not shown. The automated loading/unloading systems, frequently used with large sterilizers to allow faster loading and unloading operations, are shown instead. (Adapted from Ref. 5.)

construction significantly; 2) the loads to be rotated are generally bulky and heavy; 3) it is practically impossible to avoid mass displacements of the load during rotation; and 4) lubrication of the load bearing and rotating system must be avoided for hygiene-related reasons.

Finally, it is evident that the actual loading capacity of the chamber is reduced because of the presence of a cylindrical structure that must rotate inside it and support the load contained in appropriate trays with a lid (the entire system being appropriately perforated). Fig. 9 shows this cylindrical structure both when empty and when filled with the trays. These autoclaves are generally counterpressure sterilizers and the load is rotated throughout the process at an adjustable rate (1–10 rpm) and, if required, intermittently and in alternating directions.

REFERENCES

1. Carleton, F.J.; Agalloco, J.P. *Validation of Aseptic Pharmaceutical Processes*; Marcel Dekker, Inc.: New York, 1986.
2. Lavagna, S.M. *Injectables and Water for Pharmaceutical Use*; Editrice Bias: Milan, 1995.
3. Pflug, I.J. *Syllabus for an Introductory Course in the Microbiology and Engineering of Sterilization Processes*, 4th Ed.; Environmental Sterilization Service: St. Paul, MN, 1980.
4. *Validation of Steam Sterilization Cycles*. Technical Monograph No. 1 Parenteral Drug Association: Philadelphia, 1978.
5. Pistolesi, D.; Mascherpa, V. *Moist and Dry Heat Sterilization Technology*; Fedegari Autoclavi S.p.A.: Albuzzano PV, 1999; CD-ROM.

STERILIZATION BY RADIATION

Stephen G. Schulman

Phillip M. Achey

University of Florida, Gainesville, Florida, U.S.A.

INTRODUCTION

The actions of ionizing radiations on matter and the subsequent interactions of the irradiated molecules are useful for the sterilization of pharmaceutical and surgical supplies. To take maximum advantage of the benefits derived from ionizing radiations, it is desirable to have an understanding of the fundamental processes that result in radiation damage to living and nonliving systems.

This article addresses with the effects of ionizing radiation on biological systems, beginning with the physical and chemical actions on condensed matter. After these sections, the biological responses of living systems, particularly microorganisms and viruses, to ionizing radiation are considered. A section on the current uses and government policies on the applications of ionizing radiation in areas important to the pharmaceutical industry is also included.

Description of Ionizing Radiation

Radiation energy can be either in the form of electromagnetic energy or of particle radiation. Radiation is distinguished as being either nonionizing or ionizing. Examples of nonionizing radiation include the ultraviolet, visible, infrared, and radiofrequency parts of the electromagnetic spectrum. These kinds of radiation are not addressed in this article.

A common feature of all ionizing radiation is that it is of sufficient energy to cause ionizations in the exposed material. These ionizations result in release of orbital electrons from atoms and cause disruptions of covalent bonds. Release of energy acquired by molecules in this manner will result in dramatic changes in the physical and chemical structures of the exposed materials because of the concentration and localized release of the ionizing radiation energy. By comparison, thermal heating is less efficient at causing bond rupture because of the wide distribution and diffuse release of thermal energy in matter. Whole-body exposure of a human to an amount of ionizing radiation energy equivalent to the amount of thermal energy received by drinking a cup of hot coffee will result

in the death of the individual within 30 days. Therefore, because of the potent biological action and the lack of our ability to sense exposure to ionizing radiation (unless the exposure dose rare is extremely high), it is important to practice extreme caution when working with this agent.

Another common characteristic of ionizing radiation is its ability to penetrate material. The two most commonly used forms for sterilization are energetic electron beams and electromagnetic radiation (e.g., γ -rays from cobalt-60 or cesium-137). The penetrating ability of γ -rays and X-rays is much greater than that of electrons. In either case, when developing protocols for using ionizing radiation in sterilization of material, the penetrating ability of the radiation must be considered.

PHYSICAL AND CHEMICAL ACTIONS OF IONIZING RADIATION

This section deals with the fundamental nature of the interactions of high-energy radiations with matter, from the absorption of the radiations to the eventual establishment of chemical equilibrium in the system. The process may be divided into three stages which are illustrated in Fig. 1.

1. The physical stage, consisting of the absorption of the radiant energy by the irradiated system. Its duration is of the order of 10^{-15} s.
2. The physicochemical stage, the processes that lead to the establishment of thermal equilibrium in the system. Its duration is of the order of 10^{-12} s.
3. This chemical stage, which entails diffusion and chemical reaction of the reactive species, ultimately resulting in chemical equilibrium. It lasts upwards of 10^{-8} s, depending on the rate constants and diffusion coefficients of the reactive species.

The Physical Stage

The absorption of the energy associated with ionizing radiation by a medium results, initially, in ionization and electronic excitation. These processes occur regardless of

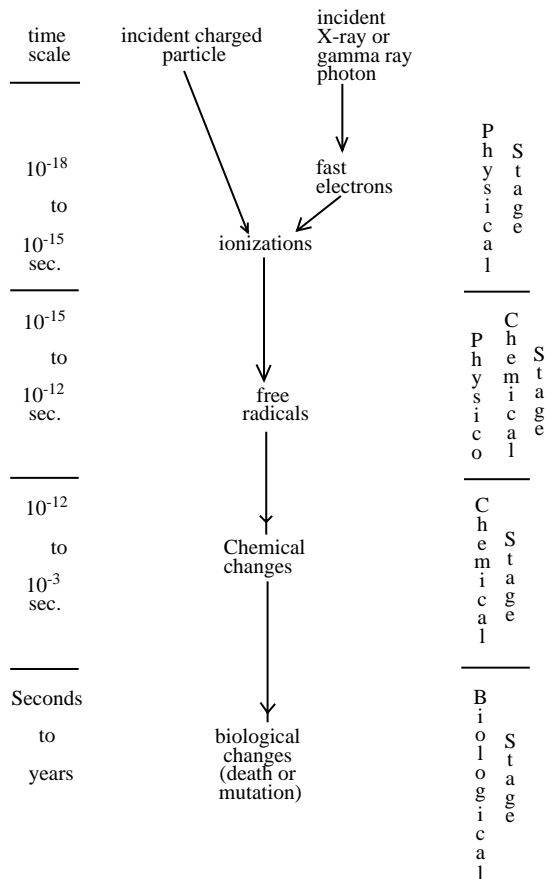


Fig. 1 The stages of radiation action on chemical systems and biological organisms.

the nature of the radiation. The mechanism of excitation and ionization by charged particles is different, however, from that effected by high-energy photons.

Charged particles

The interactions with a medium of charged particulate radiations such as protons, (β -particles, and γ -rays consist predominately of electrostatic coulomb excitation and ionization caused by ejection of atomic and molecular electrons in the medium. According to Bethe's semiclassical treatment, the energy lost to the medium, per unit length of path, by a heavy particle of charge Ze and velocity v is:

$$\frac{-dE}{dx} = \frac{4\pi Z^2 e^4 n}{mv^2} \ln \frac{2mv^2}{I} \quad (1)$$

where n is the electron density (number of electrons per unit volume) of the medium, m is the electronic mass, and I is a mean excitation potential for the medium ($I = 11.5 \text{ Zev}$

for $Z \leq 30$, $I = 8.8Z \text{ ev}$ for $Z > 30$, where Z is the mean atomic number of the medium).

The term $-dE/dx$ is called the linear energy transfer (LET) of the radiation. If electrons (β -rays) are the ionizing particles, the expression for LET is slightly different:

$$\frac{-dE}{dx} = \frac{4\pi Z^2 e^4 n}{mv^2} \ln \left(\frac{mv^2}{2I} \sqrt{\frac{e}{2}} \right) \quad (2)$$

where e is the basic of the natural logarithms.

Several important conclusions can be drawn from Eqs. 1 and 2. First, the rate of energy loss of a charged particle in a given medium is proportional to the electron in the medium. Second, because the factor v^2 outside the logarithmic term is more important than that inside, the rate of energy loss increases as the particle slows down. Third, if two particles of equal energy but different mass are compared, the heavier one will have a smaller velocity and thus a higher LET. Consequently an α -particle will produce many more excitations and ionizations per unit path length and have a shorter path length than a β -particle of the same energy.

High-energy photons

When high-energy photons such as x-rays and γ -rays pass through matter, they lose energy by way of three mechanisms: photoelectric absorption, in which the photon transfers its entire energy to an electron; Compton scattering, in which the photon transfers part of its energy to an electron; and pair production, in which the photon disappears and a high-energy electron and positron are formed. The relative importance of each of these mechanisms depends on the energy of the photon. For photons in the 100-kev to 2-MeV range, the principal mode of absorption by the medium is Compton scattering. Much higher photon energies favor pair production (at least 1.02 MeV are required to produce a pair), whereas lower energies favor photoelectric absorption. The principal effect of the absorption of high-energy photons is the production of energetic electrons that then dissipate their energies by the mechanism described by Eq. 2.

In general, the effect of transfer of energy from an energetic particle to the medium is to produce along its path a variety of electronically excited molecules, ions, and free electrons. Secondary electrons are also formed along with the ions. The electronic transitions resulting in the formation of these species occur in times [10^{-15} s] that are short compared with molecular vibration periods [$10^{-14} - 10^{-12} \text{ s}$]. The amount of energy absorbed by the irradiated system per unit mass is called the dose and is expressed in Rads, where 1 Rad ($= 6.24 \times 10^{13} \text{ eV/gm}$) is

the amount of radiation that will deposit 100 ergs of energy per gram of the irradiated system.

The Physiochemical Stage

This stage lasts about 10^{-14} – 10^{-12} s, which is typical of the period of molecular vibrations. During this period, internal molecular rearrangements can take place.

During the physiochemical stage, the excited molecules and ions dissipate their excess energy by bond rupture, luminescence, internal conversion, and energy transfer to neighboring molecules. Also during this stage, the low-energy, secondary electrons produced during the physical stage interact with molecules in the environment resulting in the formation of free radicals.

The Chemical Stage

During this stage, the reactive intermediates (ions and radicals) produced in the previous stages diffuse away from their sites of production and undergo chemical reactions with each other and with other molecules in the environment.

In condensed systems (liquids and solids), the main reactive species produced in the physiochemical stage that react in the chemical stage are free radicals. Their primary modes of reaction are atomic abstraction, radical recombination, and addition to π -bonds.

Because most systems of interest to pharmaceutical scientists are either liquids or solids, it is useful to present some of the features of radiolysis common to liquid or solid samples.

Liquid

The irradiation of liquids results initially in the ejection of electrons with the consequent formation of ions. The ejected electrons usually lose their excess kinetic energy within the electric field of the parent ion. Most of the ion pairs formed culminate in recapture of the ejected electrons, leaving the molecules in a highly excited, electronic state that may return to the ground state by internal conversion, luminescence, or energy transfer. Alternatively, the highly excited neutral molecules may split into free radicals. Some ion pairs may be sufficiently long-lived to diffuse away from the site of production and react with the surrounding medium. The free radicals are the most important reactive species formed. Once free radicals are formed along the track of an ionizing particle, they may combine with each other or they may diffuse away from the spur and react with molecules in the bulk of

the liquid medium. Those that recombine within the spur react so rapidly that they cannot be detected by physical or chemical methods. They form stable molecular products, which are known as the molecular yield. Those radicals that diffuse away from the spur and react with the medium can be detected by physical methods such as electron spin resonance spectroscopy and by chemical methods such as compound formation with radical scavengers, e.g., iodine and diphenylpicrylhydrazyl. The compounds formed by reaction with radical scavengers are called the radical yield. The mechanisms of chemical radiation effects are frequently determined by comparison of relative molecular and radical yields. The radiation chemistry of liquids has developed along two distinct paths—that of water and aqueous solutions and that of organic liquids. The radiation chemical processes responsible for the destruction of microorganisms are probably most closely related to the former, but damage to membranes and cell walls may actually be better related to the radiation chemistry of organic liquids or even some solids.

Water and Aqueous Solution

The irradiation of pure water is believed to result in two dissociative processes. The first of these is the direct dissociation of water into hydrogen atoms and hydroxyl radicals:



The second reaction is the ionization of water to yield a hydrogen ion, a hydroxyl radical, and a hydrated electron:



The hydrated electron is a powerful reducing agent and will reduce water and the hydrogen ion according to:



and the hydrogen ion according to:



Because the latter two reactions result in the same products as the direct radiolysis and because the products of reduction by the hydrogen atom and the hydrated electron are identical, it is frequently impossible to determine whether the hydrated electron or the hydrogen atom is the principal reducing species in aqueous solutions. In acid solutions, it is reasonable to assume that the hydrated electron will reduce H^+ almost exclusively and that $\text{H}\cdot$ will therefore be the predominant reducing species. However, in neutral and basic solution, the hydrated electron may be

assumed to predominate. The ultimate molecular products of the radiolysis of pure water are hydrogen gas and hydrogen peroxide, formed by the reactions:



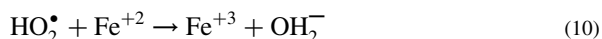
The radiation chemistry of aqueous solutions may be considered from two points of view. The first, called the Target Theory, considers the direct effect of ionizing radiations on the solute molecules. The second approach regards transformations in the solute molecules to be attributed to interactions with the reactive intermediates formed by the radiolysis of water. Because most aqueous systems are relatively dilute, the latter approach seems statistically more reasonable. Kinetic studies of dilute aqueous systems have indeed borne out this supposition. The radiation chemistry of aqueous solutions then becomes the free radical and redox chemistry of $\text{H}\cdot$, $\text{OH}\cdot$, and e_{aq}^- .

The effectiveness of radicals in producing chemical change in aqueous systems depends on the LET of the ionizing radiation that produces these radicals. A high LET particle, such as an α -particle or a proton, will produce a large concentration of radicals along its short track. These radicals are likely to recombine, forming molecular products, before they can diffuse away from the spurs in which they are formed. Low LET particles, on the other hand, produce low radical concentration along their tracks. This minimizes the probability of recombination so that the radicals can diffuse away from the spurs and initiate chemical reactions. Protons and α -particles therefore result in high molecular yields, whereas β - and γ -rays result in high radical yields.

One of the earliest devices for the measurement of radiation dosage, the Fricke dosimeter, is based on the oxidation of the ferrous ion by OH radicals produced in the radiolysis of a dilute aqueous solution of ferrous sulfate:



The presence of dissolved oxygen alters the nature of the redox properties of irradiated water. This is a consequence of the "radical scavenging" property of oxygen. Molecular oxygen has two unpaired electrons. One of these can form a covalent bond with a hydrogen atom, forming the hydroperoxy radical ($\text{HO}_2\cdot$). This species acts principally as an oxidizing agent; e.g.:



An important consequence of this is that, whereas solutions in pure irradiated water have approximately

equal oxidizing and reducing capabilities, the presence of oxygen in these solutions can result, in some cases, in very strong oxidizing properties because of the conversion of the reducing hydrogen atom to the predominantly oxidizing hydroperoxy radical. In general, the presence of oxygen in aqueous solutions will lead to alternations of the mechanisms of radiolyses owing to the "exclusively of oxidation." Radiation damage to microorganisms tends to be far more extensive in the presence of oxygen than in its absence.

Organic Liquids

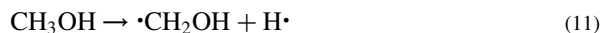
An important difference between the radiation chemistry of water and of organic liquids is that the concept of the spur (a reasonably well-defined volume in which the formation of the reactive species occurs along the track of the ionizing particle) becomes hazy. The radicals formed in water tend to recombine rather than react with the environment immediately after formation. The volume in which recombination is likely defines the spur. The radical products of irradiated organic liquids, however, are more likely to interact with their immediate environment than to undergo recombination. This is evidenced by the low molecular yields of hydrogen from irradiated organic systems.

The radiation chemistry of hydrocarbons and their derivatives has been investigated extensively. An important difference between gas phase and liquid phase radiolysis of hydrocarbons exists in that the breaking of carbon-carbon bonds is an important primary process in the gas phase, whereas in the liquid phase, the rupture of carbon-hydrogen bonds is almost exclusive. Another important difference between analogous reactions in gas and liquid phases occurs in the polymerization process. In gas phase polymerizations, the presence of radical scavengers such as iodine and benzoquinone does not appreciably alter the yields of polymeric products. In the liquid phase, however, the yields of the polymers obtained from the irradiation of materials such as vinyl chloride are seriously curtailed by the addition of radical scavengers. This indicates that polymerization in the liquid state occurs primarily by a free radical mechanism, whereas in the gaseous state, it occurs by an ionic mechanism. The irradiation of polymeric materials results in cross-linking of polymer chains and grafting of dissimilar polymeric materials. This treatment of polymers contributes considerable tensile strength and heat resistance to the irradiated polymers and is already being exploited commercially in the production of stain-resistant textiles and heat-resistant plastic containers.

Irradiation of saturated aliphatic compounds typically results in unsaturation, polymerization, and isomerization. The radiolysis of cyclohexane illustrates all three of these processes. If the radicals are very energetic, cyclohexene can be formed by the abstraction of hydrogen from a cyclohexyl radical either by a hydrogen atom or by another cyclohexyl radical. If the radicals become thermalized, recombination of radicals can occur to give bicyclohexyl. A less frequent process is rearrangement, followed by hydrogen atom capture to yield methylcyclopentane.

The irradiation of alkyl halides results in cleavage of the carbon-halogen bond. The radiolysis of methyl iodide, for example, yields ethane and molecular iodine.

Alcohols, on radiolysis in the liquid state, yield aldehydes and vicinal glycols. For example, consider the radiolysis of methanol:



and



The irradiation of frozen alcohols results in deep coloration of the alcoholic glasses. Methanol turns a brilliant purple, whereas ethanol turns blue. These colored glasses are stable if kept in the dark at low temperature. Exposure to visible or ultraviolet light results in bleaching of the alcoholic glasses as well as in the elimination of the electron spin resonance signal observed in the colored glasses. The colors are believed to be caused by the absorption spectra of trapped free radicals in the glasses. The product yields from the bleached glasses are different from those of irradiated glasses that have not been exposed to light. This suggests that the trapped radicals might be photolyzed by visible and ultraviolet light.

The irradiation of aromatic compounds results in considerably lower yields of radiolysis products than does irradiation of aliphatic compounds of similar molecular weight and functional group composition. This has been attributed to effectiveness of the delocalized π -orbitals in accommodating excitation energy without permitting the molecule to dissociate. Nevertheless, some radiolysis does occur. Benzene is known to yield biphenyl, phenylcyclohexadiene, and a polymeric material of average composition $(\text{C}_6\text{H}_7)_x$, which behaves as if it were an unsaturated hydrocarbon. Dimerization and polymer formation are also characteristic of the radiation chemistry of other aromatic hydrocarbons. The resistance of polystyrene $-(\text{C}_6\text{H}_5)_n$

to cross-linking compared with polyethylene is further evidence of the stability of aromatics to radiation effects.

Aromatic compounds frequently protect other more radiosensitive compounds from radiolysis. For example, liquid cyclohexane is protected from extensive radiolysis by the addition of a small amount of benzene. This is probably due to energy transfer from cyclohexane to benzene, followed by dissipation of the excitation energy by the aromatic π -system.

One of the most important general features of the radiation chemistry of liquids is that so much energy is deposited by the ionizing radiations, excited or reactive molecules are formed in close proximity and are likely to react with one another. This situation is not encountered in photochemistry except when lasers are used for excitation.

SOLIDS

Pharmaceutical and surgical supplies are often in the solid state when irradiated. Certainly, their containers are solid. It is therefore in order to consider some of the radiation chemistry of solids.

Because of the "fixed" positions of atoms in crystalline lattices, the effects of irradiation of solids include atomic displacements as well as electronic excitation and ionization. Although electronic alterations of materials affect their chemical behavior, atomic displacements in solids are found to have a much more pronounced effect on the physical properties of crystals. To dislodge an atom from its normal lattice position, a certain amount of energy must be transferred to the atom by an irradiating particle. Because of the large mass of the atoms, electrons and photons will be relatively ineffective in producing substantial numbers of atomic dislocations. The heavier particles, α -particles, protons, deuterons, and neutrons, will be much more effective at this process. Furthermore, unlike the primary effect of ionizing radiations in producing distal electronic disturbances through electrostatic effects, the predominant process that is required to produce atomic dislocations is direct collision.

There are two types of lattice defects that occur in all real crystals and at very high concentration in irradiated crystals. These are known as point defects and line defects. Point defects occur as the result of displacements of atoms from their normal lattice sites. The displaced atoms usually occupy sites that are not in the lattice framework; they are then known as "interstitials."

The empty lattice site left behind by the interstitial is called a vacancy. A vacancy produced by displacement of an anion or cation, along with its interstitial ion, is called a Frenkel pair, or simply a Frenkel defect. In some cases, the displaced ions are removed so far from their vacancies that they form a new layer at the crystal surface. The vacancies left behind in this case are called Schottky defects. Frenkel and Schottky defects play very important roles in the properties of solids altered by radiation damage.

Line defects (dislocations) are produced by slippage or shear of the crystal lattice. If the slippage is perpendicular to a face of the crystal so that the lattice planes on either side of the dislocation are parallel but displaced with respect to one another, the defect is called an edge dislocation. If the slippage is angular, as if produced by rotation about the shear axis so that lattice planes on either side of the defect are not perpendicular, the defect is called a screw dislocation.

There are four broad classifications of crystal types, according to the nature of the interatomic forces holding the crystal together. In metallic crystals, the atoms are thought to form a quasi-ionic lattice arrangement with the valence electrons, which bind the lattice, delocalized throughout the crystal so that they cannot be identified with any atom. Valence crystals, such as diamond, consist of a lattice in which the atoms are bonded by conventional covalent interaction throughout the lattice. This implies that a valence crystal could be considered a giant molecule. Molecular crystals (e.g., naphthalene and water) are regular arrangements of well-defined molecules that are bound together in the lattice by Van der Waals and hydrogen-bonding forces. Finally, the ultimate in electronic localization occurs in ionic crystals, in which the lattice is composed of alternating positive and negative ions held together by strong electrostatic attractions. Sodium chloride is a typical example of an ionic lattice.

Metallic Crystals

Because of the delocalization of electrons throughout the metallic crystal, no persistence of ionization or chemical decomposition can occur because a positive hole formed by an electron ejection is always refilled by an electron from the conduction band. On the other hand, sufficiently energetic radiations can cause atomic displacements. The production of interstitial atoms swells the lattice, thereby decreasing the density of the crystal.

The most obvious evidence of radiation damage in metallic crystals is decrease in electrical and thermal conductivity. This is attributable to scattering of electrons

and phonons by vacancies and interstitials that destroy the order of the lattice necessary for high conductivity.

The obvious effects of radiation damage in metallic crystals can be reversed by "annealing." Heating the irradiated materials supplies the energy required to push an interstitial back into a vacancy.

Valence Crystals

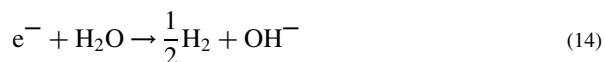
The strong bonding in valence crystals results in the failure of these crystals to demonstrate, on irradiation, quasi-chemical changes such as depolymerization. Unlike metals, however, valence crystals have no conduction electrons. This permits them to retain electronic dislocations as well as atomic displacement. The trapping of dislocated electrons in the crystal by potential wells such as those created by atomic vacancies results in coloration of the normally transparent valence crystals.

Ionic Crystals

Irradiation of ionic crystals results in atomic and electronic dislocations. The trapping of displaced electrons by anion vacancies results in the absorption of visible and near ultraviolet light, which give these crystals their characteristic colors. These pseudoatomic electrons and their vacancies are called color centers.

The exposure of colored ionic crystals to visible or ultraviolet light causes the annealing of trapped electrons and results in bleaching of the colorations induced by irradiation. In some cases in which the crystal remains uncolored upon irradiation, thermoluminescence is observed in the annealing process.

The dissolution of a heavily irradiated crystal of sodium chloride in water will result in the evolution of hydrogen and chloride from the solution. The solution also turns alkaline. This is presumably owing to the reactions of trapped holes and electrons with water:



Trapped electrons also account for the ability of irradiated sodium chloride to initiate polymerization in acrylonitrile.

Irradiation of nitrates, chlorates, perchlorates, and bromates results in the liberation of oxygen. In KClO_4 , irradiation results in explosion of the crystal due to the internal buildup of oxygen.

Molecular Crystals

The irradiation of substances that form crystals containing discrete molecules held together by dispersion forces results in radiolysis in the conventional sense. For example, the radiolysis of aliphatic carboxylic acids in the solid state yields hydrogen, carbon monoxide, and carbon dioxide. The relative yields of these gases depend on the strengths of the bonds involved in radiolysis and their frequency of occurrence. These considerations apply as well to liquids and gases and suggest no special solid-state effects.

Energy transfer in molecular crystals seems to be a well-established phenomenon. Irradiated crystals of anthracene containing only a trace of naphthacene show the characteristic green fluorescence of naphthacene rather than the violet of the primary constituent. If the material is dissolved in benzene, the anthracene fluorescence predominates.

The irradiation of ice results in formation of trapped hydrogen and hydroxyl radicals as well as the hydrated electron.

The irradiation of surface catalysts alters the properties of these catalysts through defect production on their surface. These defects have been observed to enhance and inhibit catalytic activity in specific cases. For example, the irradiation of silica gel enhances the rate of H_2 - D_2 exchange on it.

CHEMICAL PROTECTION FROM IONIZING RADIATIONS

The effects of ionizing radiations on chemical and biological systems may be minimized by the addition of certain chemical compounds to the system to be irradiated. These compounds react either directly with the radiation or, more often, with the reactive species produced by the radiations. In so doing, they are themselves transformed into other substances, but their transformation results in the preservation of the integrity of the original chemical or biological system.

At the molecular level, there are several mechanisms that account for the protection of irradiated systems by chemical agents. These are energy and charge transfer, in which an ionized or excited molecule transfers its charge or excess energy to a protecting molecule either by collision or by resonance transfer; scavenging, in which a protecting radical scavenger reacts with radicals from the initial actions of the radiation before they can attack other molecules in the system; and complex formation, in which a protective molecule can form complexes that are either

more or less susceptible to radiation damage than the original substance.

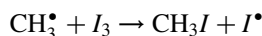
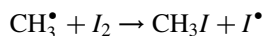
Energy and Charge Transfer

The transfer of charge or excitation energy must be fast enough to compete with dissociation processes if protection is to occur. In some cases, an activated molecule can dissociate within 10^{-14} s, the time for one molecular vibration. However, localization of energy in a particular bond usually requires 10^{-13} – 10^{-9} s. To remove energy or charge from an activated molecule effectively, the protector should have a slightly lower ionization or excitation potential. In fluid systems, the rate of charge transfer can be limited by diffusion, in which case, the donor and acceptor must be in contact. Excitation energy, however, can be exchanged radiationlessly by molecules as much as 70 Å apart by resonance energy transfer, a dipole–dipole interaction. The process is more rapid than fluorescence and competes favorably with dissociation. Transfer processes of the resonance type are extremely efficient in crystalline materials in which the high degree of order permits excitation energy to travel in excitons that transverse the crystal in a time shorter than its vibrational relaxation time. Crystalline structure also facilitates charge transfer by providing conduction bands in which electrons can freely move about.

Energy conversions within a molecule can decrease the probability of decomposition; energy can be distributed so widely that its localization in any one bond is improbable. Aromatic compounds are protective because they can probably dissipate acquired excitation energy throughout their extensively delocalized π -systems.

Scavenging Intermediates

The addition of certain compounds that react readily with free radicals can effectively prevent these products of ionizing radiations, from causing secondary damage in the system. Molecular iodine is a very effective radical scavenger forming iodo compounds with radicals and leaving behind iodine atoms to do further scavenging, e.g.:



Oxygen is a diradical that enhances radiation damage by forming radicals with other radicals. An example is the scavenging of hydrogen atoms by molecular oxygen to form the hydroperoxy radical.

Complexes

Certain compounds may exert protective action by forming molecular complexes with the original molecules of the system. These complexes might be less sensitive to radiolysis or attack by radicals, or they may be better able to transfer charge and excitation energy than the original compound. For example, the radiolytic degradation of polyisobutylene is reduced by copolymerization with styrene. The radiation resistance of the porphyrin ring is enhanced by complexing it with vanadium and other metals.

The most obvious application of chemical protection from ionizing radiations is to biological systems. For a protective agent to be biologically practical, it must be nontoxic at protective concentrations, widely distributed, and remain intact for long periods before irradiation. Many substances have been applied to this problem. To date, the most effective have been compounds such as cysteine because of the scavenging property of the $-SH$ group and the ease of oxidation of the $-NH_2$ group.

Adventitious impurities in pharmaceutical and surgical supplies may act as energy transfer acceptors, scavengers, or complexants and make radiation sterilization more difficult owing to radioprotective action.

MOLECULES OF BIOLOGICAL SIGNIFICANCE

There are two distinct theories of the actions of ionizing radiations on the compounds of living cells that result in chemical transformation that leads to mutation or cell death. The first is the target theory. This approach regards only those events that produce direct ionizations in biologically significant molecules as being important. The evidence for this is that in many cases, the amount of damage to a given organism varies logarithmically with the dose of radiation. This implies that the amount of damage possible in a cell is proportional to the number of radiosensitive molecules remaining undamaged and, therefore, capable of reacting. The other theory is based on an indirect relationship between the incident radiation and the affected, biologically significant molecules. In this approach, the solvent, water, interacts with the radiation, forming ions and radicals. These reactive species, in turn, react with the biologically significant molecules causing radiation damage. Radiation biology, under this approach, is simply a branch of the radiation chemistry of aqueous solutions. There is evidence that both the target and indirect processes occur.

In this section, the effects of ionizing radiations on molecules known to have biological significance and the

relationship of the radiation chemistry of these molecules to radiation effects observed in living organism are considered.

Carbohydrates

The irradiation of aqueous solutions of carbohydrates has the same effect as it does on alcohols. The hydroxyl groups are attacked to yield carbonyl compounds. Under anoxic conditions, dimer products and, ultimately, polymers are also formed. The primary alcohol groups of carbohydrates are especially radiosensitive. Mannitol is readily oxidized to mannose and sorbitol to glucose. Although oxidation of primary alcohol groups is favored by aerobic conditions, high yields from the oxidation of secondary alcohol groups are favored by anoxic conditions.

The irradiation of polysaccharides results predominately in their degradation. This explains why fruits and vegetables become soft on irradiation. This degradation is observed to occur both in solution and in the dry state.

Amino Acids and Peptides

The irradiation of amino acids results in transformation of both the amino and the carboxylic functions. In the dry state, glycine is decarboxylated to methylamine on irradiation; in dilute aqueous solution, however, the amino group is hydrolyzed to give glyoxylic acid, acetic acid, and formaldehyde. In solutions of concentration greater than 2%, methylamine again becomes an important product. The other amino acids are similar to glycine in their radiolytic behavior. Alanine, for example, gives ethylamine and CO_2 in the dry state and pyruvic acid and ammonia in aqueous solution.

The aromatic amino acids, when irradiated in aqueous solution, show effects that are typical of aromatic compounds and amino acids. Phenylalanine is deaminated in aerated solutions with the formation of a ketone. The aromatic ring remains relatively stable to radiolytic decomposition.

The irradiation of peptides results in a chemistry similar to that of the amino acids but also in the breakage of the peptide bond. In aqueous solution, all irradiated peptides give ammonia whether or not free amino groups are present.

The thiol and disulfide containing amino acids degrade to keto acids with the evolution of CO_2 and H_2S , when irradiated in the dry state. In solution, however, the thiol and disulfide groups are excellent radical scavengers, and free radical attack on these groups precludes deamination.

The ultimate result of irradiation of thiol-containing amino acids is their oxidation to disulfides. Thus, irradiation of an aqueous solution of cysteine results in the formation of cystine. The irradiation of the disulfides results in higher oxidation products. For example, cystine gives cystine disulfoxide in aqueous solution. Reduction of disulfides to thiols is not observed.

Proteins and Enzymes

The irradiation of proteins results in the formation of free radicals at the sites of $-S-S-$ bonds. Aromatic amino acids in the proteins are also particularly susceptible to radiolysis; decarboxylation and deamination being common results of irradiation. Rupture of the peptide linkage is characteristic of the radiolysis of proteins. In the case of enzymes, the destruction of peptide linkages is accompanied by a decrease in biological activity. This decrease continues after irradiation is stopped. The mechanisms of radiolysis in the dry state and in solution are different, but the results are usually similar. One of the more important differences between these results is the degradation of proteins by dry-state irradiation compared with increase of molecular weight through cross-linking in solution. In general, the radiation chemistry of proteins and enzymes may be considered a special case of the radiation chemistry of peptides and amino acids.

Respiratory Proteins, Vitamins, and Coenzymes

Respiratory proteins

These substances are iron, porphyrin, protein complexes. Irradiation of these substances may produce effects in the porphyrin ring or in the protein, but oxidation or reduction of the iron is almost always involved. The iron in ferricytochrome-c is reduced to the ferrous state under neutral conditions. Under acid conditions, the ferric form is favored. Hemoglobin and oxhemoglobin are both oxidized from the ferrous to the ferric state, destroying the property of oxygen transport. Large radiation doses result in attack on the porphyrin ring and denaturation of the protein. When irradiated in the dry state in the absence of oxygen, hemoglobin is not oxidized, but it becomes insoluble because of protein denaturation. Myoglobin behaves in a manner similar to that of hemoglobin but is considerably more radiosensitive. Myoglobin is a copper-containing respiratory protein of molecular weight more than 10 times that of hemoglobin. In this case, attack at the protein part of the molecule predominates.

Vitamins and coenzymes

The irradiation of coenzyme I (diphosphopyridine nucleotide) results in reduction of the pyridine-carbox-amido ring. The product of this reduction is probably a dimer that is itself radiosensitive.

The B group vitamins, thiamine and riboflavin, are destroyed on irradiation in dilute aqueous solutions. Riboflavin is reduced in air-free solutions to a semiquinone form. Nicotinic acid is decarboxylated on irradiation in air-saturated aqueous solutions.

Para-aminobenzoic acid is destroyed on irradiation in aqueous solution by deamination and decarboxylation. Sulfanilamide and sulfathiazole are inactivated presumably because of deamination. The cobalt in vitamin B₁₂ is reduced on irradiation from +3 to the +2 state.

The plant hormone auxin has been shown to be radiosensitive. The product of the irradiation of auxin (β -indoleacetic acid) is a polymer similar to that obtained in the radiolysis of indole.

Nucleic Acids

The nucleic acids DNA and RNA are responsible for the transmission of genetic information and protein synthesis. Both of these processes are dependent on the ordering of purine and pyrimidine bases that are bound to the main body of the molecule by phosphoric ester linkages. The main body of these molecules consists of ribose (5-carbon sugar) molecules linked together by phosphoric acid units to form a long strand. The purine and pyrimidine bases branch off from the chain at the ribose sites. The DNA molecule consists of two helically intertwined strands of nucleic acid held together by hydrogen bonding between purine pyrimidine pairs on opposite strands.

The irradiation of nucleic acids ruptures hydrogen bonds that hold DNA strands together results in deamination and dehydroxylation of purine and pyrimidine bases, fission of sugar base linkage, liberation of the purine bases, destruction of the pyrimidine bases, oxidation of the sugar moiety, and breakage of the nucleotide chain with liberation of inorganic phosphates. In the presence of oxygen, irradiation leads to the formation of hydroperoxides of the pyrimidine bases but not of the purine bases.

In general, the purine bases appear to be much more stable to radiolysis than pyrimidine bases. This is probably owing to the greater π -delocalization energy of the purines, which provides a pathway for nondestructive energy dissipation. Furthermore, the pyrimidine bases are known to undergo free radical reactions more readily than the purine bases. The order adenine > guanine > cytosine

> uracil > thymine has been established for the relative stabilities of the bases to radiolysis. In the presence of oxygen in aqueous solution, uracil and thymine form stable hydroperoxides, whereas cytosine forms an unstable hydroperoxide that decomposes to a variety of products.

Irradiation of DNA in the solid state at liquid nitrogen temperature yields radicals in which, electron spin resonance measurements indicate, the unpaired spin is delocalized over the entire chain and does not belong to any one unit of the giant molecule. The addition of small amounts of water to this system does not alter the nature of the DNA radicals produced, but a two-to-one excess of water results in the annihilation of the electron spin resonance signal for DNA with the appearance of a strong signal due to water radicals. It has been postulated that this protective effect is attributable to energy transfer that is made possible in an excess of water by structuring of the water, thus providing a pathway for the formation of a delocalized water radical, or exciton. Electron spin resonance studies of irradiated nucleoprotein solutions indicate that the protein takes most of the radiation damage, protecting the nucleic acid moiety.

The damages caused by ionizing radiations in nucleic acid and their components are obviously detrimental to the passage of genetic information that requires specific order of intact purine and pyrimidine bases in the DNA strands. Alterations in these bases and the DNA molecules in general can lead to mutations and lethal genes. The disruption of RNA molecules interferes with protein synthesis and can result in eventual cell death.

RESPONSE OF BIOLOGICAL SYSTEMS TO IONIZING RADIATION

Identification of the Critical Target

As noted previously, absorption of ionizing radiation energy depends on the atomic weights of the atoms of the material and the density of the material. In contrast to the absorption of ultraviolet and visible radiation, the absorption of ionizing radiation is virtually independent of the nature of the molecular structure. Thus, the release of ionizing radiation energy in biological systems is essentially independent of the molecular bonds contained in the different biological molecules. This complicates the task of identifying the critical actions of this type of radiation. A good understanding of the action of ionizing radiation on living cells requires that the biologically important events should be known.

Determination of the critical sites for killing and mutagenesis by ionizing radiation has required biological

experiments designed to answer this question. Several criteria essential for a target to qualify as a critical site for radiation action include: 1) relatively large size, 2) one or only a few copies in the cell, and 3) that it should serve a critical function for the growth and survival of the cell. The first two criteria result from the statistical distribution of damaging events in the cell. From target theory, it has been shown that the physical distribution of energy released from ionizing radiation follows a Poisson statistical distribution. Therefore, the most likely targets to be damaged are those with the first two criteria. The two cellular components that best satisfy all three criteria are the genomic material (DNA) and the cytoplasmic membrane.

Evidence Supporting DNA as a Critical Target

DNA is considered as one of the plausible critical targets for radiation action because there is only one or a few copies of the genome in each cell, because it is large compared with other molecular components, and because it plays a critical role in the proliferation and survival of the cell.

The most compelling pieces of experimental evidence that DNA is a critical target for the biological action of ionizing radiation are the mutagenic effects of ionizing radiation and the existence of DNA repair-deficient mutants of bacteria and cultured mammalian cells, which display a high sensitivity to killing by ionizing radiation.

Cellular genetic information is determined by the base sequence of the genomic DNA. Any alterations in this base sequence, such as damage to the bases in DNA (see above), will result in changes to this base sequence and, thus, to potentially mutagenic events. Whether the base sequence change will result in a phenotypic change depends on exactly where the base change occurs. If it is in an essential region of the DNA for the gene coding for a protein product or some functional or structural RNA product (tRNA or rRNA), then the base change will give rise to a phenotypic mutant. Otherwise, the mutation will fall into the category of a "silent mutation," which refers to those genotypic changes that do not have any associated phenotypic changes. For either type of mutation, the DNA is the component that must undergo change for the mutagenic event to occur.

An important advance in the understanding of how radiation kills biological systems involved the discovery that certain cells were more sensitive to killing by radiation than others. This discovery was first observed with the killing of bacteria by ultraviolet radiation.

In the case of ultraviolet radiation, it was rather straightforward to establish that the biochemical damage involved in this differential sensitivity of cells was in DNA and, more specifically, that it was the result of the formation of intrastrand (or, much less frequently, interstrand) cyclobutane pyrimidine dimers. Several factors facilitated the identification of this type of damage. First, pyrimidines absorb ultraviolet light strongly at 260 nm, and the efficiency of killing by ultraviolet radiation is maximal at or near this wavelength. Second, there was available a sensitive and quantitative assay for this type of damage. Thus, it was possible to quantitatively measure this biochemical damage and establish its relation to biological killing by the ultraviolet radiation.

Several decades later (in the 1960s), the same differential sensitivity of bacteria to killing by ionizing radiation was observed, and the biochemical damage responsible for this differential sensitivity was determined to be single-strand breaks in the backbone of the DNA. DNA strand breaks can be measured by a sensitive assay, using agarose gel electrophoresis, thus affording the opportunity to quantitatively relate the production of strand breaks with cell killing. Additionally, the isolation of radiation-sensitive mutants that lacked the ability to efficiently repair strand breaks was achieved. Thus, there was good evidence that this type of damage was related to the killing action of ionizing radiation.

Evidence that the Cytoplasmic Membrane Is a Critical Target

The cytoplasmic membrane is critical for the survival of the cell. It is large, and there is only one present for each cell. Therefore, the cytoplasmic membrane has the key characteristics that make it a likely critical target for killing by ionizing radiation. Experimental evidence that the membrane plays a role in the lethal action of ionizing radiation involves the observation that oxygen increases the radiosensitivity of cells to killing by ionizing radiation. This radiosensitization by oxygen during radiation exposure is commonly called the oxygen effect and is active in the killing of bacteria, mammalian cells, and plant cells, but not viruses. Also, the biological inactivation of free DNA (either viral or transforming DNA) by ionizing radiation does not display the oxygen effect.

As a result of the different responses to the oxygen effect, it is common to refer to the killing action of ionizing radiation including a component that is oxygen-dependent and one that is oxygen-independent damage.

Each of these results in cell-killing. The oxygen-independent damage is considered to be damage that involves the DNA, and the oxygen-dependent damage is considered to be that to the cytoplasmic membrane. Typically, the efficiency of killing by ionizing radiation is two to four times greater when exposure occurs in the presence of oxygen compared with when exposure occurs in an anoxic environment. Thus, it is reasonable to predict that between 50 and 75% of the lethal action of ionizing radiation results from membrane damage. For the oxygen effect to occur, the oxygen must be present during radiation exposure.

Relative Sensitivity of Various Biological Systems to Killing by Ionizing Radiation

To design a protocol for sterilization of a solution or material by ionizing radiation, it is necessary to establish the exposure dose required to kill all living organisms present, which is the definition of the sterilization dose. There is a large amount of data that provides measures of the radiosensitivity of diverse biological systems to radiation. For pharmaceuticals, it is important to consider viruses as well as pathogenic bacteria and other pathogenic organisms when determining the exposure dose required for sterilization of a particular product. It is also necessary to consider the physical and chemical environmental factors that cause variations in radiation sensitivity when establishing sterilization doses for the material, for the reasons that have been addressed previously. In general, the most radioresistant biological systems are the viruses. Certain bacterial species (e.g., *Micrococcus radiodurans*) are as resistant as many viruses. The 37% survival dose for these radioresistant bacteria is approximately 20 kGy (2 Mrad).

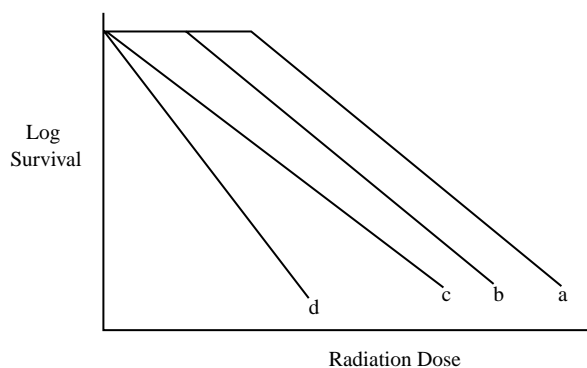


Fig. 2 Biological survival curves.

Survival Curves for Microorganisms

Ultimately, the goal of ionizing radiation treatment of pharmaceuticals is to kill all living organisms, that is, to sterilize the product. To achieve this, it is important to understand and know the survival curve response of microorganisms to killing by radiation. Fig. 2 displays the various kinds of survival curves that may be observed for microorganisms. No numbers have been assigned to the axes, because only the overall kinetic response of killing by the radiation is of interest. Curves *a* and *b* show the killing action for two hypothetical organisms with the same rate of killing but with differing repair capacities. Organisms capable of repairing lethal damage will have a characteristic threshold dose below which there is complete survival from radiation exposure. As the radiation dose increases, the repair system is not able to completely repair all radiation damage, and there is an exponential decrease in the viability of the population with increasing radiation doses. Alternatively, requirement for multiple ionization events in a critical target, or ionization events in multiple targets, can give rise to a threshold dose below which there is no observed radiation killing.

Curves *c* and *d* show the killing of organisms that lack the “shoulder” portion of curves *a* and *b*. The slope of the survival curve for organism *d* is greater than for organism *c* because organism *d* is more sensitive to radiation killing than organism *c*. Two organisms may differ in both the values of their threshold doses and in their kinetics of inactivation by radiation. It is important to keep in mind the shape and rate of inactivation of unwanted organisms when designing protocols for radiation sterilization.

USE OF IONIZING RADIATION IN THE PREPARATION OF STERILE PHARMACEUTICALS FOR HUMAN USE

At this writing, there are no approved procedures for the preparation of sterilized pharmaceuticals by exposure to ionizing radiation, although several packaging materials are approved for sterilization by doses in the 5- to 25-kGy dose region. The U.S. Food and Drug Administration (FDA) is considering rules that would permit the use of ionizing radiation in the terminal sterilization of drugs for human use. The two processes under consideration are aseptic processing and terminal sterilization for the preparation of sterile pharmaceuticals. These two methods differ in the manner by which sterilization of the product is achieved. In the case of aseptic processing, the product and the container are sterilized separately, and then final packaging of the product is carried out under aseptic

conditions. When terminal sterilization is used, the finally packaged product is sterilized. The FDA favors the terminal sterilization approach because fewer failures occur when using this technique compared with the failure rate of the aseptic processing techniques. When using terminal sterilization, the step of transfer of the product into the sterilized container under aseptic conditions is avoided, thus removing the possibility for product contamination during the transfer step. It is recognized that it will not be practical to use terminal sterilization for certain pharmaceuticals, since they might be sensitive to alteration and inactivation by the sterilizing agent, whereas they can be sterilized by filtration, when in solution.

During ongoing deliberations of which techniques should be approved by the FDA for the preparation of sterilized pharmaceuticals, the two physical methods under consideration are heat and ionizing radiation. Concerns similar to those associated with the use of ionizing radiation in food processing have been raised. After extensive discussions and hearings, the FDA has approved the commercial use of ionizing radiation for processing fruits, vegetables, poultry, beef, and spices. Before this technique is approved by the FDA as a physical agent for sterilization of pharmaceuticals, similar hearings will be required. The primary issue to be considered when developing a protocol using ionizing radiation is the establishment of the minimum and maximum allowable exposure doses. A minimum dose limit is required, to ensure that the product has been exposed to a sufficiently high dose that results in sterilization of the product. A maximum dose limit is required to avoid damage to the product resulting from unnecessarily high exposure doses. Exposure dose limits in the FDA regulations for food processing include minimum and maximum acceptable dose limits, which were established with these goals in mind.

Experiments on the action of ionizing radiation on pharmaceuticals and the killing of bacteria by ionizing radiation indicate that this treatment has the potential to be used in terminal sterilization. Reports indicate that there are no products formed in irradiated samples of penicillin G, neomycin, novobiocin, and dihydrostreptomycin, which are different from those that are formed by acidic, basic, hydrolytic, and oxidative treatments of these antibiotics. At the same exposure doses, there was a 1 million-fold reduction in the number of viable bacterial spores, which are the most radiosensitive forms of endospore forming bacteria. Thus, the use of ionizing radiation may provide an alternative to heat or chemical treatment for the sterilization of pharmaceuticals. This could provide a solution to the problem of sterilization of heat-sensitive drugs. Also, avoidance of chemical sterilization removes the possibility of contamination by residues of the chemical that was used

for sterilization. It has been reported that irradiation of the antibiotic cefotaxime formed radiation products from impurities present in the cefotaxime sample. One must be aware of the role of solvents and chemicals other than the drug itself when considering radiolytic changes of a drug during sterilization.

CONCLUSIONS

The purpose of this article is to provide the pharmaceutical industry with an overview of the physical, chemical, and biological actions of ionizing radiation on molecules of interest to the industry, as well as to provide a current perspective on the prospects of the practical use of ionizing radiation as a physical agent in the sterilization of drugs for human use. Advantages of using employing ionizing radiation for pharmaceutical sterilization include:

- effectiveness on all microorganisms;
- no rise in temperature during treatment;
- consistent and reproducible protocols; and
- readily controlled and validated methods.

Although there is currently no approval by the FDA for ionizing radiation sterilization, there are active hearings being held by the agency on this matter. Considering the desire by the FDA to encourage the terminal sterilization method for sterilization of pharmaceuticals and the fact the ionizing radiation has distinct advantages over other treatment methods for use in terminal sterilization as presented above, it is expected that there will be approval of the use of this physical agent in sterilization of drugs for human use after regulatory hearings have provided sufficient information to allow identification of the appropriate conditions under which this agent can be used. Recent developments in the

application of ionizing radiation for sterilization are reported on Web sites <http://www.iaea.org> and <http://www.fda.gov>.

REFERENCES

1. Alper, T. Cellular Radiobiology **1979**.
2. Barbarin, N.; Rollmann, B.; Tilquin, B. Role of Residual Solvents in the Formation of Volatile Compounds after Radiosterilization of Cefotaxime. *Int. J. Pharm.* **1999**, *178*, 203–212.
3. Chemical Protection from Radiation Effects. *Nucleonics* **1960**, *18*, 76–81.
4. Derr D., Radiation Processing What Is It? Where Is It Going? ASTM Standardization News 1993, 25–27
5. Ebert, M., Howards, A. Eds.; *Radiation Effects in Physics, Chemistry and Biology*; Year Book Medical Publishers: Chicago, 1963.
6. Pullman, B.; Pullman, A. *Quantum Biochemistry*; John Wiley & Sons: New York, 1963; 267–283.
7. Radiation Chemistry. *Nucleonics* **1961**, *19*, 37–68.
8. Schwarz, H.A. Radiation Chemistry. *Ann. Rev. Phys. Chem.* **1965**, *16*, 347–374.
9. Stock, D.A.; Achey, P.M. Repair of Single-Strand Breaks in DNA from Cultured *Leptidopteran* Cells Exposed to Gamma Radiation. *Invertebrate Cell System Applications*; Mitsuhashi, J. Ed.; CRC Press: 1989; 1, 45–61.
10. State of the Art Symposium: Radiation Chemistry. *J. Chem. Ed.*; **1981**, *58*, 82.
11. Swallow, A.J. *Radiation Chemistry of Organic Compounds*; Pergamon Press: London, 1960.
12. Symposium on the Development of Radiation Chemistry. *J. Chem. Ed.*; **1995**, *36*.
13. Tilquin, B.; Crucq, A.S. Les Mecanismes Chimiques De La Radiosterilisation De Medicaments Solides. *J. Chim. Phys.* **1999**, *96*, 167–173.
14. Use of Aseptic Processing and Terminal Sterilization, the Preparation of Sterile Pharmaceuticals for Human and Veterinary Use. Federal Register, 1991; 56, No. 198, 51354–51358, October, 11, 1991.

SUPER DISINTEGRANTS: CHARACTERIZATION AND FUNCTION

Larry L. Augsburger

Huijeong A. Hahm

University of Maryland School of Pharmacy, Baltimore, Maryland, U.S.A.

Albert W. Brzezczko

Atlantic Pharmaceutical Services, Owings Mills, Maryland, U.S.A.

Umang Shah

Parke-Davis Division, Warner Lambert, Morris Plains, New Jersey, U.S.A.

INTRODUCTION

Disintegrating agents are substances routinely included in tablet formulations and in some hard shell capsule formulations to promote moisture penetration and dispersion of the matrix of the dosage form in dissolution fluids. An oral solid dosage form should ideally disperse into the primary particles from which it was prepared. Although various compounds have been proposed and evaluated as disintegrants, relatively few are in common usage today. Traditionally, starch has been the disintegrant of choice in tablet formulations, and it is still widely used. However, starch is far from ideal. For instance, starch generally has to be present at levels greater than 5% to adversely affect compactibility, especially in direct compression. Moreover, intragranular starch in wet granulations is not as effective as dry starch. In more recent years, several newer disintegrants have been developed. Often called "super disintegrants," these newer substances can be used at lower levels than starch. Because they can be a smaller part of the overall formulation than starch, any possible adverse effect on fluidity or compactibility would be minimized. These newer disintegrants may be organized into three classes based on their chemical structure (Table 1).

GENERAL CHEMISTRY AND SURFACE MORPHOLOGY

Sodium Starch Glycolate

Sodium starch glycolate is a super disintegrant made from cross-linking sodium carboxymethylstarch (1). Cross-linking involves a chemical reaction with phosphorus oxytrichloride or sodium trimetaphosphate, or a physical

manipulation. Carboxymethylation is performed by reacting starch with sodium chloroacetate in an alkaline medium and neutralizing with a citric or acetic acid, a process known as a Williamson ether synthesis. This synthesis yields carboxymethylation of about 25% of the glucose units. The by-products, which include sodium chloride, sodium glycolate, and sodium citrate or acetate, are partially washed out. The particle sizes of the disintegrants are increased by the substitution and cross-linking processes (3).

Sodium starch glycolates are generally spherical, a characteristic that accounts for their good flowability (4). Figure 1 shows the scanning electron photomicrographs of some of the commercially available sodium starch glycolates.

Croscarmellose Sodium

Croscarmellose sodium is derived from internally cross-linking a cellulose ether, sodium carboxymethylcellulose, which is a water soluble polymer. It is composed of repeating units of cellobiose units, with each unit consisting of two anhydroglucose units linked by 1,4- β -glucoside. Each unit also has three hydroxyl groups. The degree of substitution refers to the average number of hydroxyl groups substituted by carboxymethyl groups. Croscarmellose sodium is made from crude cellulose, which is steeped in sodium hydroxide solution (1). The cellulose is subsequently reacted with sodium monochloroacetate to form carboxymethylcellulose sodium. After completion of the substitution, the excess sodium monochloroacetate slowly hydrolyzes to glycolic acid. The glycolic acid converts a few of the sodium carboxymethyl groups to the free acid, catalyzes the cross-linkage to form croscarmellose sodium, and forms sodium chloride and sodium glycolate as the by-products.

Table 1 Classification of “super disintegrants” (partial listing)

Structural type (NF name)	Description	Trade name (manufacturer)
1. Modified starches (Sodium starch glycolate, NF)	Sodium carboxymethyl starch; the carboxymethyl groups induces hydrophilicity and cross-linking reduces solubility.	Explotab [®] (Edward Mendell Co.) Primojel [®] (Generichem Corp.) Tablo [®] (Blanver, Brazil)
2. Modified cellulose (Croscarmellose, NF)	Sodium carboxymethyl cellulose which has been cross-linked to render the material insoluble.	AcDiSol [®] (FMC Corp.) Nymcel ZSX [®] (Nyma, Netherlands) Primellose [®] (Avebe, Netherlands) Solutab [®] (Blanver, Brazil)
3. Cross-linked poly-vinylpyrrolidone (Crospovidone, NF)	Cross-linked polyvinylpyrrolidone; the high molecular weight and cross-linking render the material insoluble in water.	Crospovidone M [®] (BASF Corp.) Kollidon CL [®] (BASF Corp.) Polyplasdone XL (ISP Corp.)

Most of the by-products can be removed to achieve 99.5% purity by extraction with alcohol. Croscarmellose sodium may be milled to break the polymer fibers into shorter lengths and hence improve its flow properties.

Unlike sodium starch glycolate, crude croscarmellose sodium particles do not flow very well because of their twisted fibrous morphology and varying lengths. Therefore, they are cryogenically milled to improve flowability. The scanning electron photomicrographs show that the croscarmellose sodium particles are fibers with fairly sharp ends, probably because of the milling process (Fig. 2).

Crospovidone

Crospovidone is a cross-linked homopolymer of *N*-vinyl-2-pyrrolidone. The reactants, acetylene and formaldehyde, are used to form butynediol. The hydrogenation and subsequent cyclodehydrogenation of butynediol form butyrolactone. The reaction of butyrolactone with ammonia produces pyrrolidone that is then vinylated with acetylene under pressure. The linear polymerization of the vinylpyrrolidone yields polyvinylpyrrolidone, a soluble binder, while the popcorn (branched) polymerization yields crospovidone, an insoluble super disintegrant. The by-products of popcorn polymerization include vinylpyrrolidone and polyvinylpyrrolidone. Crospovidone contains less than 1.5% of the soluble material, which has been determined to be polyvinylpyrrolidone by infrared spectroscopy.

Crospovidone particles have a very different appearance from those of the other two classes of super disintegrants. Crospovidone particles seem to consist of aggregates of smaller particles that are fused together. This aggregation gives crospovidone a spongy, highly porous

appearance (Fig. 3). Scanning electron photomicrographs show that the reduction of particle size of crospovidone increases the surface area per unit weight, but decreases the intraparticulate porosity and the spongy appearance (4).

DISINTEGRANT ACTION

Although disintegrants are important components in solid dosage forms, their mechanism of action has not been clearly elucidated. The mechanisms proposed in the past include water wicking, swelling, deformation recovery, repulsion, and heat of wetting. It seems likely that no single mechanism can explain the complex behavior of the disintegrants. However, each of these proposed mechanism provides some understanding of different aspects of disintegrant action.

Water Wicking

The ability of a disintegrant to draw water into the porous network of a tablet is essential for effective disintegration. For crospovidone, water wicking has been thought to be the main mechanism of disintegration. Kornblum and Stoopak (5) observed that crospovidone swells very little, yet takes water into its network quite rapidly. Even the extensively swelling sodium starch glycolate showed improved disintegration when the molecular structure was altered to improve water uptake, as observed by Rudnic et al. (6). Unlike swelling, which is mainly a measure of volume expansion with accompanying force generation, water wicking is not necessarily accompanied by a volume increase.

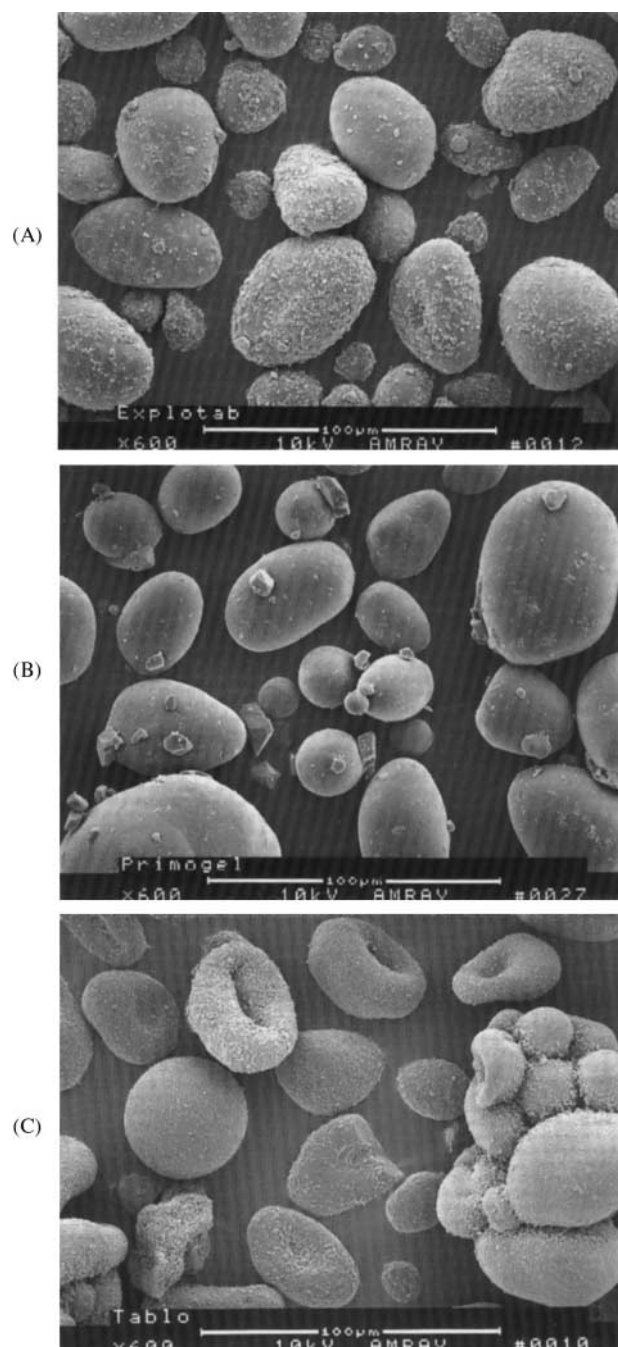


Fig. 1 Scanning electron photomicrograph of sodium starch glycolates: (A) Explotab[®], (B) Primojel[®], and (C) Tablo[®] (600× magnification).

The ability of a system to draw water can be summarized by Washburn's equation (7):

$$L^2 = \left(\frac{\gamma \cos \theta}{2\eta} \right) rt \quad (1)$$

The Washburn equation is too simplistic to apply to a dynamic tablet-disintegration process, but it does show that any change in the surface tension (γ), pore size (r), solid-liquid contact angle (θ), or liquid viscosity (η) could change the water wicking efficiency (L = length of water penetration in the capillary; t = time). For example, when Rudnic et al. (8) evaluated the disintegration efficiency of different particle sizes of crospovidone, those with the largest particle size range (50–300 μm) yielded the shortest disintegration time. Large particle size probably yielded greater pore size and altered the shape of the pore. Indeed, longer fiber length due to greater particle size could improve the efficiency of capillary uptake of water into the dosage form matrix.

Super disintegrants draw water into the matrix system at a faster rate and to a greater extent when compared to traditional starch (9). Van Kamp et al. (10), utilizing a water uptake measurement device, were able to show that tablets that demonstrate greater uptake volume and rate, such as those containing sodium starch glycolate, disintegrated more rapidly. Although the hydrophobic lubricant, magnesium stearate, seemed to negatively affect the wicking process, those containing sodium starch glycolate were less affected by the detrimental effect of mixing with the hydrophobic lubricant. Lerk et al. (11) also observed a decreased rate of wetting when disintegrants were mixed with magnesium stearate for various mixing times. The decrease in the rate of wetting was proportional to the time of mixing. Most likely, this observation reflected a greater delamination of magnesium stearate at longer mixing times.

Swelling

Although water penetration is a necessary first step for disintegration, swelling is probably the most widely accepted mechanism of action for tablet disintegrants. Indeed, most disintegrants do swell to some extent, but the variability of this property between disintegrants reduces its plausibility as a sole mechanism.

The earliest attempt to measure swelling was to measure the sedimentation volume of slurries. Nogami et al. (12) developed a reliable test to measure both swelling and water uptake. Gissinger and Stamm (13) modified this apparatus and found a positive correlation between the rate of swelling and disintegrant action for some disintegrants. List and Muazzam (14) later adapted this apparatus to measure both rate of swelling and swelling force through the application of force and displacement transducers. They found that disintegrants that generate large swelling forces are generally more effective.

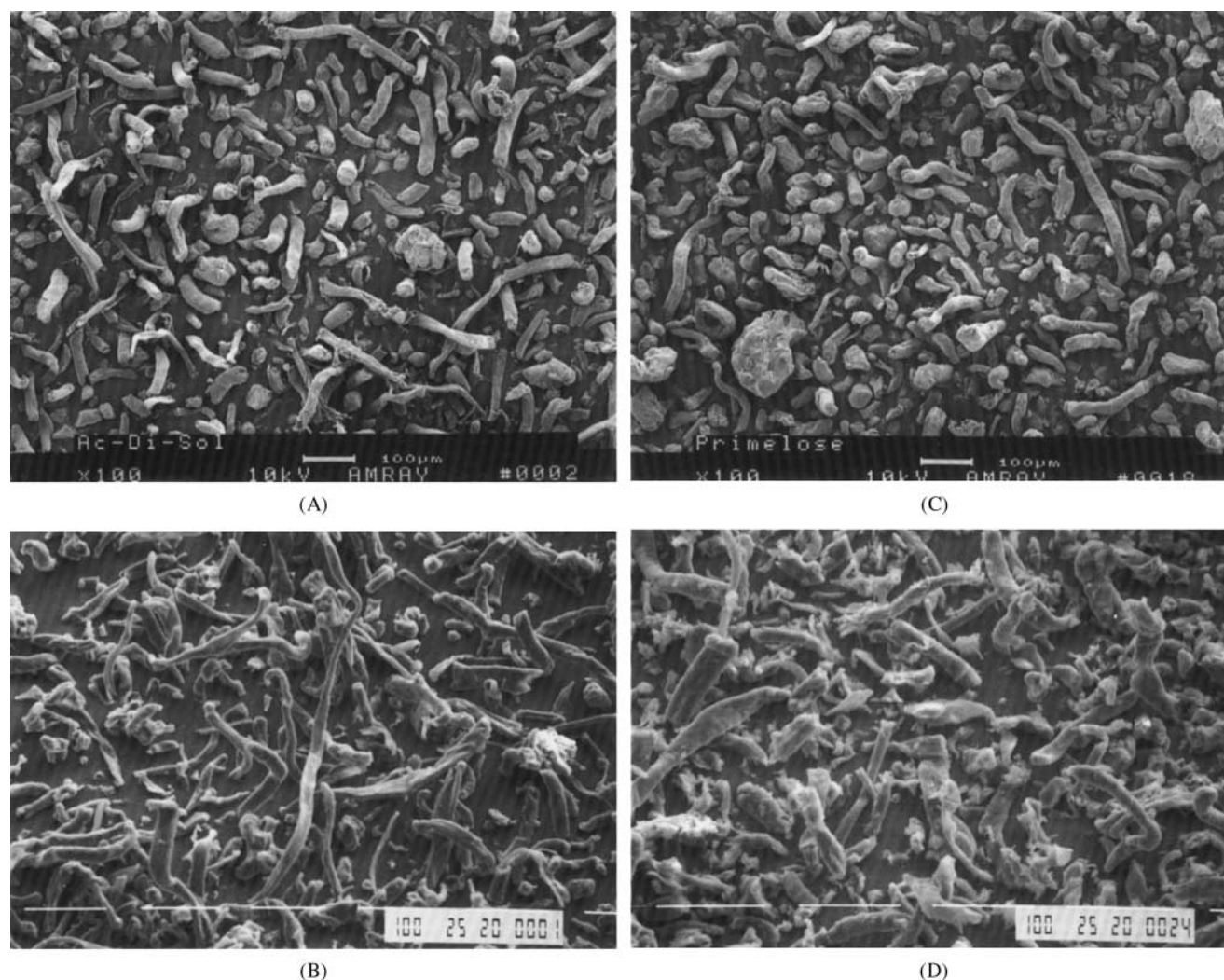


Fig. 2 Scanning electron photomicrograph of croscarmelloses: (A) AcDiSol[®], (B) Nymcel ZSX[®], (C) Primellose[®], and (D) Solutab[®] (100× magnification).

For swelling to be effective as a mechanism of disintegration, there must be a superstructure against which the disintegrant swells. Swelling of the disintegrant against the matrix leads to the development of a swelling force. A large internal porosity in the dosage form in which much of the swelling can be accommodated reduces the effectiveness of the disintegrant. At the same time, a matrix that yields readily through plastic deformation may partly accommodate any disintegrant swelling if swelling does not occur at a sufficient rapidity.

The swelling of some disintegrants is dependent on the pH of the media. Shangraw et al. (3) reported that sedimentation volumes of anionic cross-linked starches and celluloses are altered in acidic media. Polyplasdone XL[®] and Starch 1500[®] were unchanged. In a separate

study, Chen et al. (15) showed that acetaminophen tablets containing Primojel[®] and AcDiSol[®] had longer disintegration and dissolution times in acidic medium compared to neutral medium. Those containing Polyplasdone XL[®] showed no such differences. Mitrevej and Hollenbeck (16) verified the remarkable swelling capacity of some “super disintegrants” by exposing individual particles deposited on slides to high humidities and observing their degree of swelling microscopically.

On the other hand, when Caramella et al. (17–19) evaluated different disintegrants for their ability to swell, no correlation could be observed between the maximum disintegrating force and percent of particle swelling. Because they did observe a correlation between the rate of disintegrating force development and the disintegration

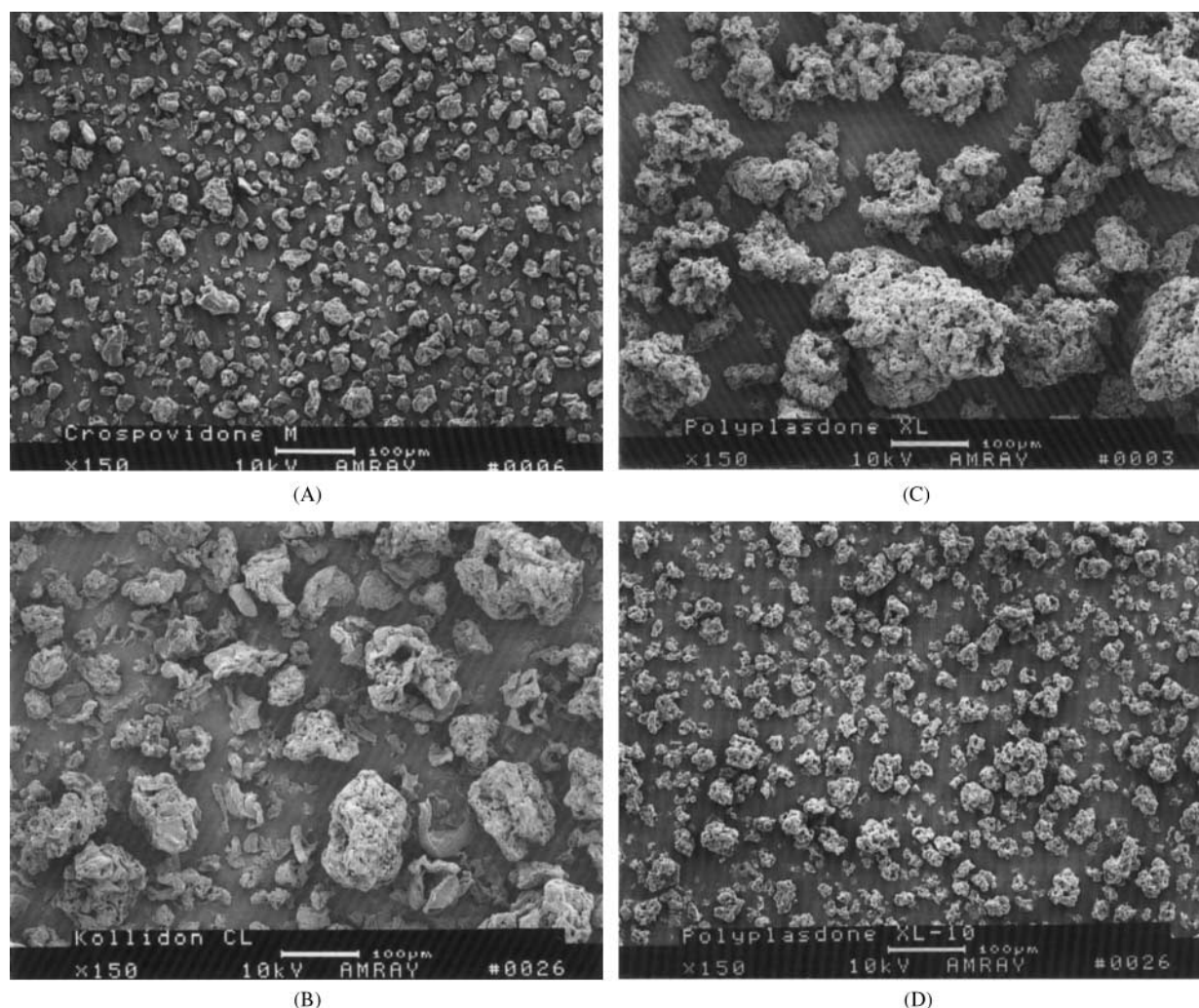


Fig. 3 Scanning electron photomicrograph of crospovidones: (A) Crospovidone M[®], (B) Kollidon CL[®], (C) Polyplasdone XL[®], and (D) Polyplasdone XL-10[®] (150× magnification).

time, therefore, the authors suggested that the rate of development of a disintegrating force is all-important. Swelling capable of rapid force development may be preferred since a slowly developing force could hypothetically allow tablets to relieve the stress generated without bond disruption.

Deformation Recovery

The deformation recovery theory implies that the shapes of the disintegrant particles are distorted during compression, and that the particles return to their precompression shape upon wetting, thereby causing the tablet to break apart. Hess (20), with the aid of photomicrographs, showed that

deformed starch particles returned to their original shape when exposed to moisture.

Fassihi (21) concluded that at higher compression forces, disintegration may become dependent on mechanical activation of the tablet, resulting from the stored energy imparted by the compression process. He examined the disintegration times of tablets made of Emdex[®] powder, magnesium stearate, and 5% disintegrant. Regardless of the disintegrant used (sodium starch glycolate, microcrystalline cellulose, croscarmellose sodium, or starch), the disintegration time increased with increasing compression force, then decreased again when the compression force was above 120 MN/m⁻².

Research on deformation and its recovery in situ as a disintegration mechanism is incomplete. However, such a

mechanism may be an important aspect of the mechanism of action of disintegrants such as croscopovidone and starch that appear to exhibit little or no swelling. The efficacy of such disintegrants likely would be dependent on the relative yield strength of the disintegrant and that of the matrix in which it is compressed, since disintegration efficiency would surely depend on how much deformation is sustained by the disintegrant particles. Also, time-dependent stress relaxation could possibly be a factor in the aging of such tablets in that any deformation induced into the disintegrant, that cannot be sustained by intraparticulate bonding, gradually may recover as the matrix relaxes.

Repulsion Theory

Ringard and Guyot-Hermann (22) have proposed a particle–particle repulsion theory to explain the observation that particles that do not swell extensively, such as starch, could still disintegrate tablets. In this theory, water penetrates into the tablet through hydrophilic pores and a continuous starch network that can convey water from one particle to the next, imparting a significant hydrostatic pressure. The water then penetrates between starch grains because of its affinity for starch surfaces, thereby breaking hydrogen bonds and other forces holding the tablet together. Presently, this theory is not supported by adequate data.

Heat of Wetting

Matsumara (23) noticed that starch particles exhibit slight exothermic properties when wetted, which was thought to cause localized stress resulting from expansion of air retained in the tablet matrix. Unfortunately, this explanation, if valid, would be limited to only a few substances such as aluminum silicate and kaolinite. List and Muazzam (24) found that exothermic wetting reactions were not exhibited with all disintegrants and that even when a significant heat of wetting was generated, disintegration time did not always decrease. Caramella et al. (25) found that an increase in temperature, which should cause air expansion, did not enhance maximum force generation in several formulations. Therefore, they concluded that expansion of air in pores from heat of wetting could not be supported by the data. More recently, Luangtana-anan et al. have examined the heat of wetting of powders and tablets of magnesium carbonate and Emcompress® (26).

Magnesium carbonate tablets with significantly higher heat of wetting disintegrated more readily than the Emcompress® tablets. Indeed, it would be interesting to

develop a model for the mechanism of tablet disintegration using a thermodynamic approach; however, heat of wetting alone probably is inadequate to explain disintegration.

Generation of a Disintegrating Force or Pressure as a Unifying Principle

The rate of generation of a disintegrating force may be a unifying factor in the mechanisms of disintegration (19). Many proposed mechanisms may be visualized as giving rise to a force. Brzezczko (27) developed techniques to simultaneously measure the rate of liquid uptake into a tablet and the rate of generation of both axial and radial swelling forces. As indicated in Figs. 4, 5, and 6, tablet compaction contributes more to the axial pressure than to the radial pressure when super disintegrants representing the three main super disintegrant classes were studied in model tablet formulations. In all three cases, the maximum axial pressure in an anhydrous lactose matrix was well below that observed with a dicalcium phosphate dihydrate matrix when the disintegrants are compared at the same concentration. The differences in disintegrant performance in soluble and insoluble matrices could be rationalized in terms of pressure development and liquid uptake. Fig. 7 compares the maximum axial disintegrating pressure and disintegration times of the tablets containing 2% of the disintegrants in a matrix composed of dicalcium phosphate or lactose (27). As can be seen, a higher disintegration pressure favors rapid disintegration of the dicalcium phosphate-based tablets, but a slower disintegration of lactose-based tablets. Higher initial axial disintegrating pressure rates also yield shorter disintegration times for the dicalcium phosphate-based tablets, but no such correlation is seen with the lactose-based tablets, whose disintegrating pressure rates are much lower than those of the dicalcium phosphate-based tablets (see Fig. 8). Maximum water uptake and water uptake rate seem to be poor predictors of disintegration time, as seen in Figs. 9 and 10. However, lactose-based tablets show a trend toward slower disintegration with faster liquid uptake. It was suggested that faster liquid uptake leads to a faster dissolution of lactose and increased porosity to accommodate swelling and/or structural recovery.

Peppas (28) attribute the difference in disintegration rate between soluble and insoluble matrices to two proposed phenomena — an interface-controlled mechanism and a diffusion-controlled mechanism — as represented in the following equation:

$$F/F_{\infty} = 1 - \exp(-kt^n) \quad (2)$$

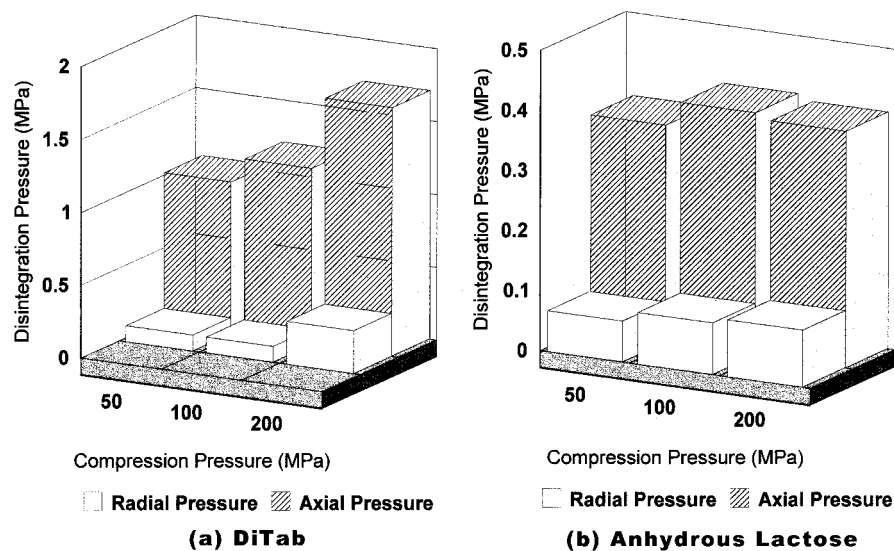


Fig. 4 The effect of compression pressure on the axial and radial disintegrating pressures of compacts made with lactose or DiTab[®], and AcDiSol[®] (5%). (From Ref. 27.)

Here, F is the disintegration force at time t , F_{∞} is the maximum force developed, k is an expansion rate constant, and n signifies which of the two mechanisms controls the disintegration. The interface controlled phenomenon involves tablet particles breaking apart from the interface of the tablet and the diffusion-controlled phenomenon

involves particles diffusing away. Although it is thought that both happen simultaneously, the degree to which disintegration depends on each system can differ. For example, those tablet matrices with a relatively small n of about 0.6 are thought to be dominated by the diffusional mechanism; whereas, those with an n of greater than

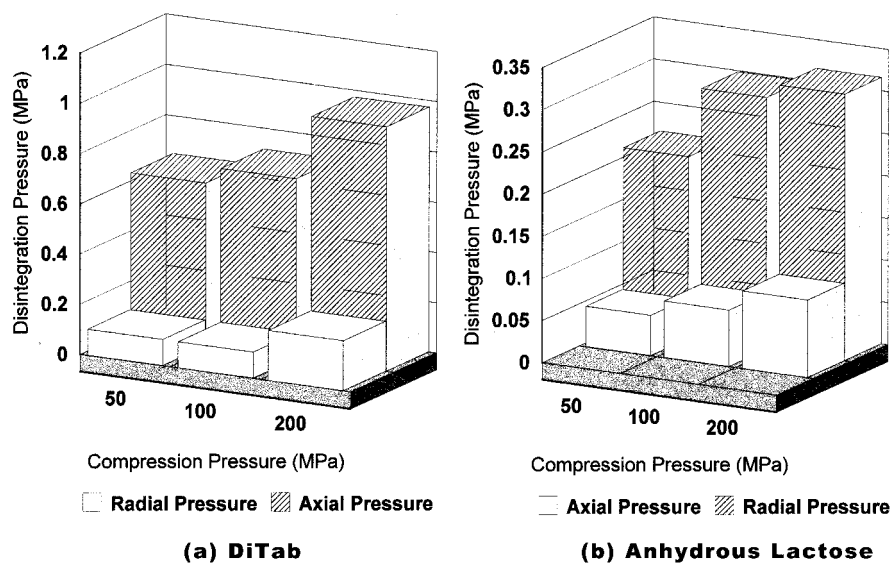


Fig. 5 The effect of compression pressure on the axial and radial disintegrating pressures of compacts made with lactose or DiTab, and Primojel[®] (5%). (From Ref. 27.)

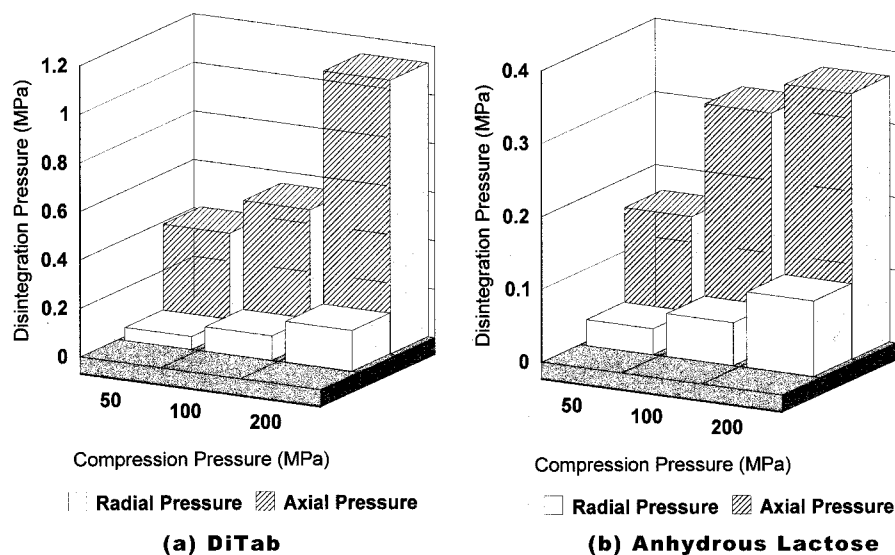


Fig. 6 The effect of compression pressure on the axial and radial disintegrating pressures of compacts made with lactose or DiTab, and Polyplasdone XL[®] (5%). (From Ref. 27.)

0.9 are thought to be interfacial mechanism dominant. The value of n would certainly differ based on the solubility of the matrices.

Since super disintegrants are highly hydrophilic yet insoluble in water, they would be expected to be more effective in breaking the tablet apart interfacially than controlling the diffusion per se. Indeed, Caramella et al. observed that disintegration occurred readily for tablets containing insoluble calcium phosphate; whereas, tablets containing highly soluble β -lactose disintegrated slowly. Such phenomena were explained by a lower value of n for the system containing β -lactose. In other words, the super

disintegrant's interface-controlled mechanism could not overcome the diffusion-controlled mechanism of the β -lactose (29).

FACTORS AFFECTING DISINTEGRANT ACTIVITY

Particle Size

Both the rate and force of disintegrant action may be dependent upon the particle size of the disintegrant. Smallenbroek et al. (30) found that starch grains having

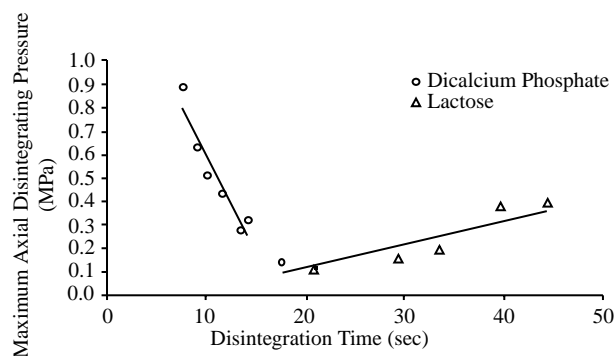


Fig. 7 Maximum axial disintegrating pressure versus disintegration time of dicalcium phosphate and lactose tablets containing 2% super disintegrants. \circ = Dicalcium phosphate: $r^2 = 0.92$, $p < 0.05$, significant correlation. Δ = Lactose: $r^2 = 0.81$, $p < 0.05$, significant correlation. (From Ref. 27.)

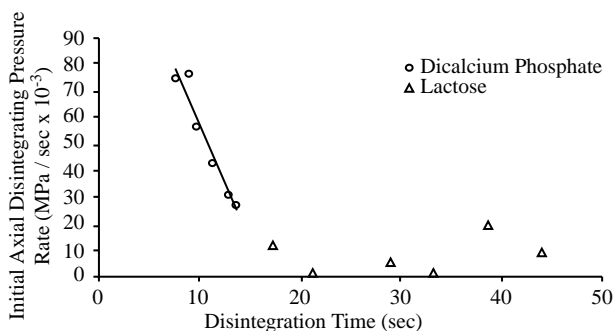


Fig. 8 Initial axial disintegrating pressure rate versus disintegration time of dicalcium phosphate and lactose tablets containing 2% super disintegrants. \circ = Dicalcium phosphate: $r^2 = 0.95$, $p < 0.05$, significant correlation. Δ = Lactose: $r^2 = 0.10$, $p < 0.05$, no significant correlation. (From Ref. 27.)

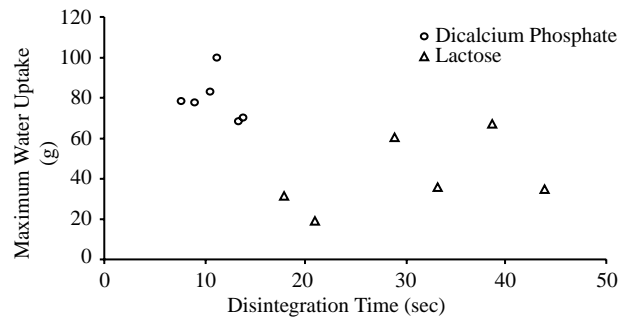


Fig. 9 Maximum water uptake versus disintegration time of dicalcium phosphate and lactose tablets containing 2% super disintegrants. ○ = Dicalcium phosphate: no significant correlation. △ = Lactose: no significant correlation. (From Ref. 27.)

relatively large particle sizes were more efficient than the smaller particle size grades. This is probably because the continuous hydrophilic network of disintegrants is more efficiently accomplished by the bigger particles. Also, Rudnic et al. (8) found that coarser grades of crospovidone (50–100 μm , Grade B; 50–300 μm , Grade C) were more efficient than the finer particles (0–15 μm , Grade A). The differences in disintegration efficiency between Grade B and Grade C were not clear, however. When List and Muazzam (24) evaluated two different grades of crospovidone particles (100–200 μm and >315 μm), the efficiencies between the two grades were very similar (see Table 2). Results for the other disintegrants, Amberlite IRP88[®] and potato starch, support that coarser particle sizes allow more efficient disintegration than finer particles. For disintegrants that swell extensively, such

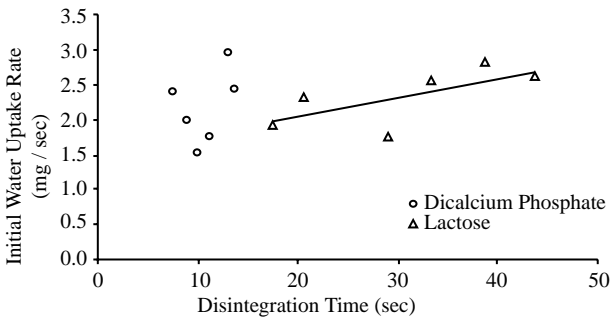


Fig. 10 Initial water uptake rate versus disintegration time of dicalcium phosphate and lactose tablets containing 2% super disintegrants. ○ = Dicalcium phosphate: no significant correlation. △ = Lactose: $r^2 = 0.47$, $p < 0.05$, no significant correlation. (From Ref. 27.)

efficiency can be explained by the observed force development. Indeed, larger particles of sodium starch glycolate swelled more rapidly and to a greater extent than did the smaller particles (6).

Molecular Structure

Disintegrants can vary in molecular structure based on how they are manufactured or processed. Corn starch, for example, contains different ratios of two sugar fractions, amylose and amylopectin. Schwartz and Zelinske (31) concluded that the linear polymer, amylose, was responsible for the disintegrant properties associated with starch, whereas the branched polymer, amylopectin, was responsible for the gummy property.

Table 2 Effect of particle size and compression pressure on swelling pressure and disintegration time

Disintegrant	Compression pressure (bar)	Particle size (μm)	Swelling pressure (bar)	Disintegration time (s)
Amberlite IRP88 [®]	625	<50	0.660	84
	1560	<50	1.121	30
Amberlite IRP88 [®]	625	100–200	1.083	52
	1560	100–200	2.262	22
Potato starch	625	<50	0.165	254
	1560	<50	0.310	164
Potato starch	625	80–100	0.234	160
	1560	80–100	0.445	77
Polyplasdone XL [®]	625	100–200	0.898	31
	1560	100–200	1.772	14
Polyplasdone XL [®]	625	>315	0.760	42
	1560	>315	1.480	17

Tablet composition: 2.5% disintegrant, 1% magnesium stearate, and Emcompress[®]. (From Ref. 24.)

Varying the amylose to amylopectin ratio did not affect the porosities of the resulting tablets. Rudnic et al. (6) evaluated the effect of cross-linking and carboxymethyl substitution in sodium starch glycolate and concluded that the swelling of the disintegrant was largely inversely proportional to the degree of cross-linkage. Swelling also was inversely proportional to the level of substitution, but to a lesser degree. Shah et al. (32) found that carboxymethyl cellulose having high molecular weight and low levels of carboxymethylation was best for tablet disintegration.

Effect of Compression Force

Compression force affects disintegration time in different ways. First, it governs the penetration of dissolution fluids into the matrix by controlling the porosity of the compact. Low compression force can lead to relatively high tablet porosity and can allow rapid penetration of water. However, it has often been observed that tablets containing starch exhibit disintegration times that tend to pass through minimum as compression force increases (21). At low compression forces, any possible swelling or deformation recovery that may take place may be more or less accommodated by the porosity, whereas at intermediate compression forces, a maximal disintegrating effect may develop. At high compression forces, fluid penetration may be impeded by a further reduction of porosity while particle deformation of the disintegrants becomes more important. In general, List and Muazzam (14) found increased swelling pressures at higher compression forces when various amberlite resins, starches, and crospovidones were employed at the 2.5% level in dicalcium phosphate matrix tablets (see Table 2). Similar findings were reported by and Fassihi (21) and Brzezko (27).

In two different studies Khan and Rhodes (33, 34) observed that tablets containing sodium starch glycolate disintegrate relatively slowly at low compression force, quickly at intermediate compression force, and slowly again at high compression force. However, the effects of compression force on the disintegration time of other types of disintegrants, such as cation exchange resin, calcium sodium alginate, and various forms of starches, varied widely. Perhaps the effect of compression force on the disintegration time depends on the nature of the disintegrants, such as their mechanism of disintegration and deformation characteristics.

Munoz et al. (35) found that the effect of compression pressure on disintegration time varied depending on the concentration used of the super disintegrant, Explotab®. Fig. 11 shows that shortest disintegration time could be

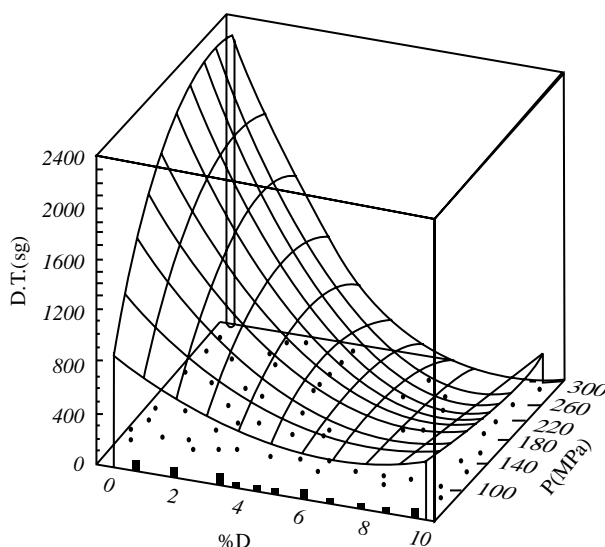


Fig. 11 Surface response of disintegration time as functions of compression pressure and percent disintegrant. (From Ref. 11.)

achieved at around 7% disintegrant concentration. At this concentration, compression force has little effect on disintegration time. The disintegration time was more affected by compression force at low disintegrant concentration, showing fastest disintegration time at intermediate compression force. This type of biphasic effect of compression force on disintegration time also was observed for AcDiSol® (36), and the surface response curve is very similar to that of Explotab®. When disintegration times were studied at 5 and 10% disintegrant, 5% AcDiSol® yielded the lowest porosity, lowest yield pressure in Heckel analysis, and shortest disintegration time. At a 10% disintegrant level, the tablets showed a slight postcompression expansion, which could explain a slightly increased disintegration time compared to the 5% disintegrant level.

The effect of compression force on disintegration efficiency seems, therefore, largely dependent on the mechanism of the disintegrant action. The effectiveness of swelling or structure recovery may well be dependent on attaining a compression force that achieves a critical porosity in the matrix. On the other hand, the capillary uptake of liquid, which is a necessary precursor to these mechanisms could be compromised if the tablet matrix is compressed to too low a porosity.

Matrix Solubility

The disintegrant mechanism seems to depend not only on the disintegrant itself but on the matrix as well. Disintegrants work most effectively in insoluble matrices

(27). Insoluble matrices, such as those containing calcium phosphate do not disintegrate adequately without disintegrants. On the other hand, tablets and capsules that primarily consist of water soluble fillers or drugs tend to dissolve rather than to disintegrate, even when disintegrating agents are present. It has been suggested that during the dissolving process, the water acts as a plasticizer (37), which can potentially reduce the development of disintegrating force. In addition, soluble materials that tend to swell can form viscous plugs, which may impede further penetration of moisture into the matrix. However, the addition of disintegrants almost predictably shortens disintegration time, despite the solubility of the matrix.

Method of Incorporation in Granulation

The method of incorporation of disintegrants in granulation has been controversial. Should the disintegrant be all extragranular, all intragranular, or divided between these two locations? Shotton and Leonard (38) reported that maize starch, sodium calcium alginate, alginic acid, and other disintegrants gave more rapid disintegration when incorporated extragranularly than when incorporated intragranularly in a sulfadiazine granulation. They also reported that the latter method gave a finer dispersion and concluded that the best compromise was to use both intra- and extragranular disintegrant.

Van Kamp et al. (39) evaluated the method of incorporation of Primojel[®], AcDiSol[®], and Polyplasdone XL[®] in prednisone tablets formed from lactose granules. Whether the incorporation of the super disintegrant was intragranular, extragranular, or evenly distributed in both sites, they found little or no difference in disintegration time, crushing strength, or dissolution of prednisone. Interestingly, their results with potato starch showed differences that did not agree with the earlier work of

Shotton and Leonard (38) in that intragranular starch was more effective than extragranular starch (see Table 3). Gordon et al. (40) reported that dissolution of naproxen, a poorly soluble drug at gastric pH levels, was faster when AcDiSol[®] was incorporated intragranularly, compared to when it was incorporated extragranularly or evenly distributed between the intra- and extragranular portions. Even more recently, a study reported by Khattab et al. (41) showed that the combined incorporation of intra- and extragranular disintegrating agents (sodium starch glycolate, croscarmellose sodium, or crospovidone) in a paracetamol granulation resulted in faster disintegration and dissolution than either extragranular or intragranular incorporation alone.

More studies are necessary to elucidate the effect of other factors, such as the type of binder, the type of filler, and the solubility of the matrix, which may significantly affect the effectiveness of disintegrants in different modes of incorporation. For example, Becker et al. (42) found that extragranular crospovidone was more effective in an acetaminophen tablet when the binder was maltodextrin (Licab DSH[®]), pregelatinized maize starch (Lycab PGS[®]), or low substituted hydroxypropylcellulose (L-HPC) than when polyvinylpyrrolidone or hydroxypropylmethylcellulose was the binder. In addition, the difference seen in the effectiveness of starch in different modes of incorporation between the Shotton and Leonard (38) study and the Van Kamp et al. (39) study may be related to the absence or presence of lactose, a soluble filler. Unlike the Shotton and Leonard study, Van Kamp et al. used lactose as a soluble filler, which might have reduced the relative effectiveness extragranular starch, making the intragranular incorporation more favorable.

The observations summarized in Table 4 make any attempt to generalize that one method of incorporation of disintegrant in granulation is better than another difficult.

Table 3 Effect of method of disintegrant addition in granules on the tablet properties

Disintegrant addition method	Crushing strength (kgf)			Disintegration time (s)		
	Intra	Equal	Extra	Intra	Equal	Extra
Control		6.5			664	
4% Primojel [®]	5.3	5.0	5.8	38	41	49
4% Ac-Di-Sol [®]	3.8	4.8	5.7	110	126	148
4% Nymcel zsd 16 [®]	4.0	4.3	6.5	499	540	488
4% Polyplasdone XL [®]	5.8	6.0	6.1	31	40	43
20% Potato Starch	3.3	3.4	2.1	69	80	110

(From Ref. 39.)

Table 4 Comparison of different modes of incorporation of disintegrants in granules on disintegration efficiency of tablets, as reported by different investigators

Investigators	Drug (D), Binder (B), Filler	Disintegrants evaluated	Order of disintegration efficiency	Additional comments
Shotton et al.	Sulfadiazine (D) Povidone (B) No filler	Maize starch	Extra > Intra	Extragranular incorporation yielded fastest disintegration, but intragranular incorporation yielded finer particles. Equal distribution of disintegrants is recommended.
		Na calcium alginate		
		Alginic acid		
		Microcrystalline cellulose Colloidal aluminum silicate		
Van Kamp et al.	Prednisone (D) Gelatin (B) Lactose (F)	Potato starch	Intra > Extra	The difference between the modes of incorporation for super disintegrants is not great.
		Primojel® Polyp lasdone XL® Nymcel® Croscarmellose Na	Intra > Equal > Extra Intra > Equal > Extra Extra > Intra > Equal Equal > Extra > Intra	
Khattab et al.	Paracetamol (D) Polyvinylpyrrolidone (B) No filler	Na starch glycolate Crospovidone	Equal > Intra or Extra Equal > Intra > Extra	Overall, the equal distribution of disintegrants yielded the fastest disintegration and dissolution.

(From Refs. 38, 38, and 41.)

Table 5 Rework efficiency (%RE) of super disintegrants

Disintegrant (2%)	Relative F_s (35% porosity) ^a	Relative F_s (40% porosity) ^a	% RE ^b
Control	0.842	0.848	45
Polyplasdone XL [®]	0.941	0.926	64
Explotab [®]	0.737	0.863	86
AcDiSol [®]	1.045	0.951	45

$$^a\text{Re} \cdot F_s = \frac{\text{Maximum swelling force (1st compression)}}{\text{Maximum swelling force (2nd compression)}}$$

$$^b\%RE = \frac{\text{AUC (1st compression)}}{\text{AUC (2nd compression)}}$$

AUCs are the area under curve from the disintegration time versus compression pressure graphs.

(From Ref. 44.)

However, when all of the data are taken together, it would appear that the combined addition of disintegrants both extragranularly and intragranularly would provide the best opportunity for optimal disintegrant activity.

Effect of Reworking

The effect of recompressing a wet massed microcrystalline cellulose matrix containing super disintegrants on swelling force kinetics also has been considered (43). When the disintegrants were placed extragranularly, only Explotab[®] among those considered retained good efficiency after reworking. When placed intragranularly, all disintegrants had reworking efficiencies equivalent to that of the nondisintegrant control. Adding 2% disintegrant extragranularly prior to the second compression restored disintegrant behavior for Polyplasdone XL[®], but only partial restoration was seen for AcDiSol[®]. In further work (44), reworked tablets containing 2% disintegrant extragranularly were studied. The data in Table 5 illustrate that maximal swelling forces were reduced in all cases, but there was no correlation with tablet disintegration time.

Incorporation in Hard Gelatin Capsules

The utility and performance of super disintegrants in direct-fill powder formulations for hard shell capsules filled on tamping machines are roughly analogous to those of direct compression tablet formulation. In a study where capsules were filled under controlled tamping force conditions using an instrumented Zanasi LZ 64 dosator machine, a dicalcium phosphate based formulation containing hydrochlorothiazide and different super disintegrants were tested for dissolution times (45). The croscarmelloses were found to be more effective than sodium starch glycolate in promoting hydrochlorothiazide dissolution, whereas

crospovidone was the poorest in this regard. In a follow-up multifactorial study, all main factors, including disintegrant type, compression force, level of lubricant, and filler type, were found to have significant effects on dissolution (Figs. 12 and 13) (46). In most cases at lower disintegrant concentration, increasing the tamping force improved the dissolution of hydrochlorothiazide, most likely due to reduced porosity. When the filler was changed from lactose to dicalcium phosphate, the magnitude and the order of effectiveness of the disintegrants changed.

Like the experience with tablets, the effect of disintegrants in already rapidly soluble capsule matrices is lower than that in water insoluble matrices. Perhaps doubling the concentration normally required for tablets is needed to effect efficient disintegration and significantly affect dissolution. This need for higher disintegrant concentration probably reflects the greater pore structure of capsule plugs compared to compressed tablets. At equivalent concentrations in model lactose or dicalcium phosphate-based systems, sodium starch glycolate and croscarmellose sodium were more effective than crospovidone in promoting dissolution of hydrochlorothiazide from capsules manufactured with the same tamping force (47). For either filler, disintegration times and swelling correlated well with dissolution.

NEW DISINTEGRANTS

Gellan gum and Xanthan SM[®] appear to have performance characteristics similar to those of super disintegrants. Gellan gum is an anionic polysaccharide of linear tetrasaccharides, and is derived from *Pseudomonas elodea* (48). When 4% gellan gum was incorporated in an ibuprofen tablets, its disintegration time was 4 min, which

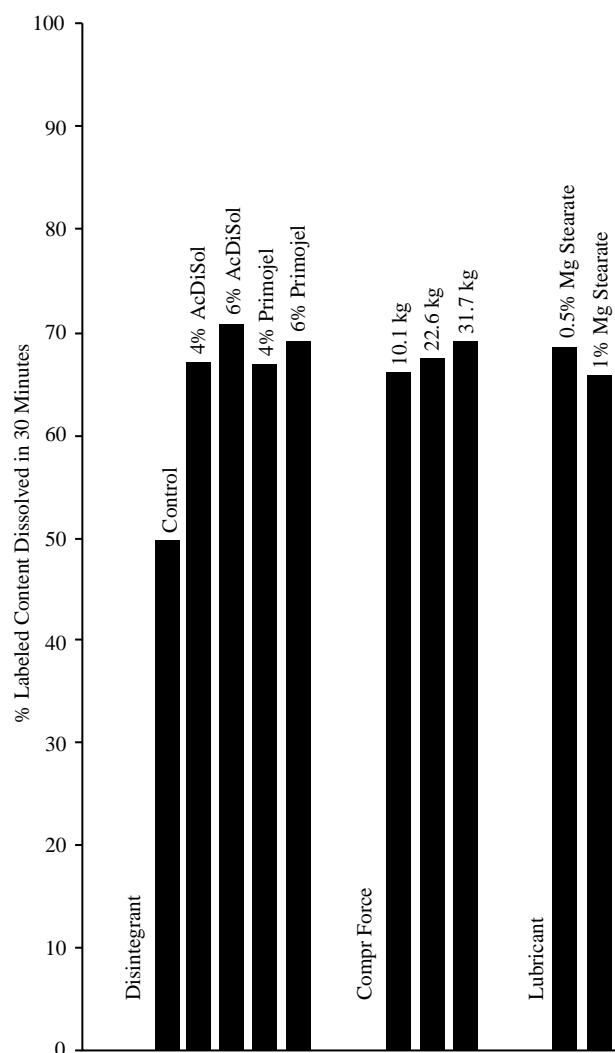


Fig. 12 The averaged effect of disintegrant, compression force, and lubricant on the release of hydrochlorothiazide from anhydrous lactose based capsules. Control = 0% disintegrant. (From Ref. 47.)

was much superior than that obtained using dried starch or Avicel PH 102® (>15 min), and comparable to those of Explotab®, AcDiSol®, and Kollidon CL® (4–7 min).

Xanthan SM® is a new USP xanthan derivative with higher hydrophilicity and lower gelling tendency (49). In aspirin tablets with 3% Xanthan SM® disintegration time was about 10 min. Increasing the concentration of disintegrant beyond 3% did not improve the disintegration time, whereas the most effective concentration of AcDiSol® was 5%, yielding a disintegration time of less than 5 min. Xanthan SM®, like AcDiSol®, has low solubility in water, but is reported to swell more extensively in water.

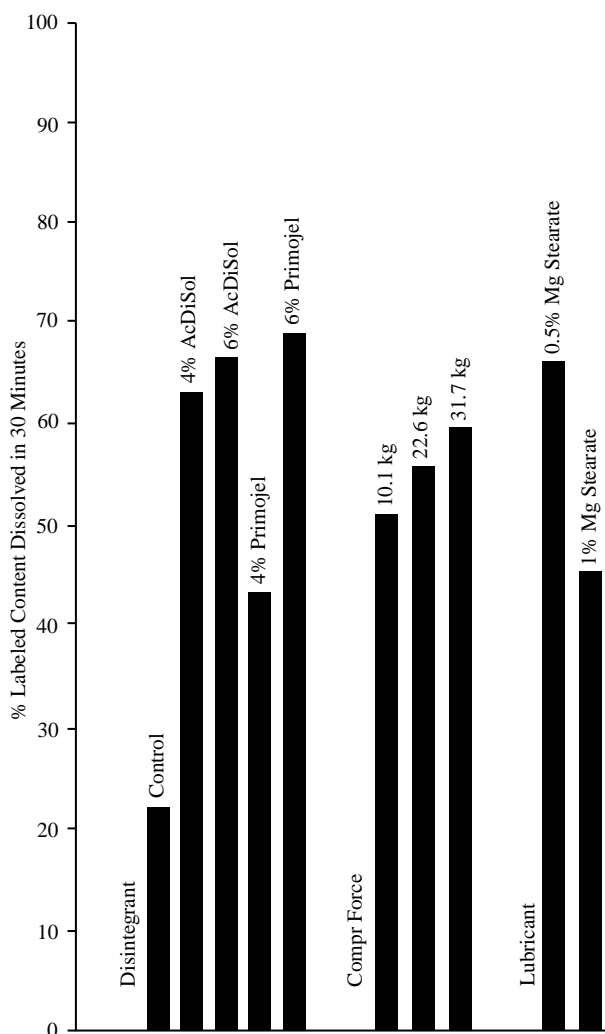


Fig. 13 The averaged effect of disintegrant, compression force, and diluent on the release of hydrochlorothiazide from dicalcium phosphate based capsules. Control = 0% disintegrant. (From Ref. 47.)

SUMMARY

In summary, super disintegrants are excipients used to promote rapid breakdown of oral solid dosage forms to aid dissolution in vivo. Commonly used super disintegrants include sodium starch glycolate, croscarmellose sodium, and crospovidone. Super disintegrants differ from traditional starch in that they are effective at much lower concentrations. This lower concentration provides formulation scientists greater flexibility, particularly in designing direct compression tablets. However, the effectiveness of both starch and super disintegrants depends heavily upon

the composition of the tablet matrix, compression pressure, and in the case of granulation, the method of incorporation.

Because of the complexities involved, the mechanism of action of super disintegrants is not well understood. Some of the proposed mechanisms include water wicking, swelling, deformation recovery, particle repulsion, and heat of wetting. Water uptake is a necessary precursor to all other mechanisms. Not all mechanisms are well supported by research. Disintegrants appear to function by multiple mechanisms, but a predominant mechanism seems to be characteristic of each disintegrant type. Regardless of their validity, all proposed mechanisms at least have the potential to generate a disintegrating force within the matrix and this appears to be a unifying concept.

ACKNOWLEDGMENT

The authors wish to express sincere appreciation and gratitude to Ms. Lynn DiMemmo, FMC Co., Princeton, N.J., for the SEM's of the super disintegrants.

REFERENCES

1. *Handbook of Pharmaceutical Excipients*, 2nd Ed.; American Pharmaceutical Association: Washington, DC, 1994; 141–144, 462–466.
2. Bolhuis, G.K.; Van Kamp, H.V.; Lerk, C.F. Effect of Variation of Degree of Substitution, Crosslinking, and Purity of the Disintegration Efficiency of Sodium Starch Glycolate. *Acta Pharm. Tech.* **1984**, *30*, 24–32.
3. Shangraw, R.F.; Mitrevej, A.M.; Shah, M.N. A New Era of Tablet Disintegrants. *Pharm. Tech.* **1980**, *4*, 49–57.
4. Shah, U.S. *Evaluation of the Functional Equivalence of Different Sources of Super Disintegrants in Pharmaceutical Tablets*; University of Maryland: Baltimore, 1996; 16–26, 92–112.
5. Kornblum, S.S.; Stoopak, S.B. A New Tablet Disintegrating Agent: Cross-Linked Polyvinylpyrrolidone. *J. Pharm. Sci.* **1973**, *62*, 43–49.
6. Rudnic, E.M.; Kanig, J.; Rhodes, C.T. The Effect of Molecular Structure on the Function of Sodium Starch Glycolate in Wet Granulated Systems. *Drug Dev. Ind. Pharm.* **1983**, *9*, 303–320.
7. Washburn, E.W. The Dynamic of Capillary Flow. *Phys. Rev.* **1921**, *17*, 273–283.
8. Rudnic, E.M.; Lausier, J.M.; Chilamkarti, R.N.; Rhodes, C.T. Studies of the Utility of Cross-Linked Polyvinylpyrrolidone as a Tablet Disintegrant. *Drug Dev. Ind. Pharm.* **1980**, *6*, 291–309.
9. Grai, E.; Ghanem, A.H.; Mahmoud, H. Studies of the Direct Compression of Pharmaceuticals. *Pharm. Ind.* **1984**, *46*, 279–284.
10. Van Kamp, H.V.; Bolhuis, G.K.; De Boer, A.H.; Lerk, C.F.; Lie-A-Huen, L. The Role of Water Uptake on Tablet Disintegration. *Pharm. Acta Helv.* **1986**, *61*, 22–29.
11. Lerk, C.F.; Bolhuis, G.K.; Smullenbroek, A.J.; Zuurman, K. Interaction of Tablet Disintegrants and Magnesium Stearate during Mixing. II. Effect on Dissolution Rate. *Pharm. Acta Helv.* **1982**, *57*, 282–286.
12. Nogami, H.; Nagai, T.; Fukuoka, E.; Sonobe, T. Disintegration of the Aspirin Tablets Containing Potato Starch and Microcrystalline Cellulose in Various Concentrations. *Chem. Pharm. Bull.* **1969**, *17*, 1450–1455.
13. Gissinger, D.; Stamm, A. A Comparative Evaluation of the Properties of Some Tablet Disintegrants. *Drug Dev. Ind. Pharm.* **1980**, *6*, 511–536.
14. List, P.H.; Muazzam, U.A. Swelling —The Driving Force Behind Disintegration. *Pharm. Ind.* **1979**, *41*, 1075–1077.
15. Chen, C.R.; Lin, Y.H.; Cho, S.L.; Yen, S.Y.; Wu, H.L. Investigation of the Dissolution Difference Between Acidic and Neutral Media of Acetaminophen Tablets Containing Super Disintegrant and a Soluble Excipient. *Chem. Pharm. Bull.* **1997**, *45*, 509–512.
16. Mitrevej, A.; Hollenbeck, R.G. Photomicrographic Analysis of Water Vapor Sorption and Swelling of Selected Super Disintegrants. *Pharm. Tech.* **1982**, *6*, 48–50.
17. Caramella, C.; Columbo, P.; Conte, U.; Gazzaniga, A.; LaManna, A. The Role of Swelling in the Disintegration Process. *Int. J. Pharm. Tech. Prod. Mfr.* **1984**, *5*, 1–5.
18. Caramella, C.; Columbo, P.; Conte, U.; LaManna, A. Swelling of Disintegrant Particles and Disintegrating Force of Tablets. *Labo-Pharma. Probl. Tech.* **1984**, *32*, 115–119.
19. Caramella, C.; Columbo, P.; Conte, U.; LaManna, A. Tablet Disintegration Update: The Dynamic Approach. *Drug Dev. Ind. Pharm.* **1987**, *13*, 2111–2145.
20. Hess, H. Tablets Under the Microscope. *Pharm. Tech.* **1978**, *2*, 38–57, 100.
21. Fassihi, A.R. Mechanisms of Disintegration and Compactibility of Disintegrants in a Direct Compression System. *Int. J. Pharm.* **1986**, *32*, 93–96.
22. Guyot-Hermann, A.M.; Ringard, J. Disintegration Mechanisms of Tablets Containing Starches. Hypothesis About the Particle-Particle Repulsive Force. *Drug Dev. Ind. Pharm.* **1981**, *7*, 155–177.
23. Matsumara, H. Studies on the Mechanism of Tablet Compression and Disintegration. IV. Evolution of Wetting Heat and Its Reduction by Compressional Force. *Yakugaku Zasshi* **1959**, *79*, 63–68.
24. List, P.H.; Muazzam, U.A. Swelling —A Driving Force in Tablet Disintegration. *Pharm. Ind.* **1979**, *41*, 1075–1077.
25. Caramella, C.; Ferrari, F.; Conte, U.; Gazzaniga, A.; LaManna, A.; Colombo, P. Experimental Evidence of Disintegration Mechanisms. *Acta Pharm. Tech.* **1989**, *35*, 30–33.
26. Luangtana-anan, M.; Catellani, P.L.; Colombo, P.; Dinarvand, R.; Fell, J.T.; Santi, P. The Role of Bond Weakening by Liquids in the Disintegration of Tablets. *Eur. J. Pharm. Biopharm.* **1992**, *38*, 169–171.
27. Brzezczko, A.W. *Ph.D. Dissertation; A Critical Evaluation of Tablet Disintegration Processes*; University of Maryland: Baltimore, 1989; 129–172, 237–255.
28. Peppas, N.A. Energetics of Tablet Disintegration. *Int. J. Pharm.* **1989**, *51*, 77–83.
29. Caramella, C.; Colombo, P.; Conte, U.; Ferrari, F.; Gazzaniga, A.; LaManna, A.; Peppas, N.A. A Physical Analysis of the Phenomenon of Tablet Disintegration. *Int. J. Pharm.* **1988**, *44*, 177–186.

30. Smallenbroek, A.J.; Bolhuis, G.K.; Lerk, C.F. The Effect of Particle Size of Disintegrants on the Disintegration of Tablets. *Pharm. Weekbld.* **1981**, *116*, 172–175.
31. Schwartz, J.B.; Zelinski, J.A. The Binding and Disintegrant Properties of the Corn Starch Fractions: Amylose and Amylopectin. *Drug Dev. Ind. Pharm.* **1978**, *4*, 463–483.
32. Shah, N.H.; Bekersky, I.; Jarowski, C.I. *Carboxymethyl-cellulose: Effect of the Degree of Polymerization and Substitution on Tablet Disintegration and Dissolution. III. Urinary Salicylate from Rapid and Controlled Release Formulas*; Program and Abstracts, APhA. Acad. Pharm. Sci. Mtg., IPT Section, Nov 17, 1983; American Pharmaceutical Assoc. Washington, DC, 1983; 34–35.
33. Khan, K.A.; Rhodes, C.T. Disintegration Properties of Calcium Phosphate Dibasic Dihydrate Tablets. *J. Pharm. Sci.* **1975**, *64*, 166–168.
34. Khan, K.A.; Rhodes, C.T. Effect of Variation in Compaction Force on Properties of Six Direct Compression Tablet Formulations. *J. Pharm. Sci.* **1976**, *65*, 1835–1837.
35. Munoz, N.; Ferrero, C.; Munoz-Ruiz, A.; Velasco, M.V.; Jimenez-Castellanos, M.R. Effect of Explotab® On the Tableability of a Poorly Soluble Drug. *Drug Dev. Ind. Pharm.* **1998**, *24*, 785–791.
36. Ferrero, C.; Munoz, N.; Velasco, M.V.; Munoz-Ruiz, A.; Jimenez-Castellanos, R. Disintegrating Efficiency of Croscarmellose Sodium in a Direct Compression Formulation. *Int. J. Pharm.* **1997**, *147*, 11–21.
37. Murthy, K.S.; Ghebre-Sellassie, I. Current Perspectives on the Dissolution Stability of Solid Oral Dosage Forms. *J. Pharm. Sci.* **1993**, *82*, 113–126.
38. Shotton, E.; Leonard, G.S. Effect of Intragranular and Extragranular Disintegrating Agents on Particle Size of Disintegrated Tablets. *J. Pharm. Sci.* **1976**, *65*, 1170–1174.
39. Van Kamp, H.V.; Bolhuis, G.K.; Lerk, C.F. Improvement by Super Disintegrants of the Properties of Tablets Containing Lactose, Prepared by Wet Granulation. *Pharm. Weekbld. Sci. Ed.* **1983**, *5*, 165–171.
40. Gordon, M.S.; Chatterjee, B.; Chowhan, Z.T. Effect of the Mode of Croscarmellose Sodium Incorporation on Tablet Dissolution and Friability. *J. Pharm. Sci.* **1990**, *79*, 43–47.
41. Khattab, I.; Menon, A.; Sakr, A. Effect of Mode of Incorporation of Disintegrants on the Characteristics of Fluid-Bed Wet-Granulated Tablets. *J. Pharm. Pharmacol.* **1993**, *45*, 687–691.
42. Becker, D.; Rigassi, T.; Bauer-Brandl, A. Effectiveness of Binders in Wet Granulation: A Comparison Using Model Formulations of Different Tableability. *Drug Dev. Ind. Pharm.* **1997**, *23*, 791–808.
43. Gould, P.L.; Tan, S.B. The Effect of Recompression on Disintegrant Efficiency in Tablets Prepared by Wet Granulation. *Drug Dev. Ind. Pharm.* **1985**, *11*, 441–460.
44. Gould, P.L.; Tan, S.B. The Effect of Recompression on the Swelling Kinetics of Wet Massed Tablets, Containing 'Super' Disintegrants. *Drug Dev. Ind. Pharm.* **1985**, *11*, 1819–1836.
45. Botzolakakis, J.E.; Small, L.E.; Augsburger, L.L. Effect of Disintegrants on Drug Dissolution from Capsules Filled on a Dosator-Type Automatic Capsule-Filling Machine. *Int. J. Pharm.* **1982**, *12*, 341–349.
46. Botzolakakis, J.E.; Augsburger, L.L. The Role of Disintegrants in Hard-Gelatin Capsules. *J. Pharm. Pharmacol.* **1984**, *36*, 77–84.
47. Botzolakakis, J.E.; Augsburger, L.L. Disintegrating Agents in Hard-Gelatin Capsules. Part I: Mechanism of Action. *Drug Dev. Ind. Pharm.* **1988**, *14*, 29–41.
48. Antony, P.J.; Sanghavi, N.M. A New Disintegrant for Pharmaceutical Dosage Forms. *Drug Dev. Ind. Pharm.* **1997**, *23*, 413–415.
49. Rizk, S.; Barthelemy, C.; Duru, C.; Guyot-Hermann, A.M. Investigation on a New Modified USP Xanthan with Tablet-Disintegrating Properties. *Drug Dev. Ind. Pharm.* **1997**, *23*, 19–26.

BIBLIOGRAPHY

- Buckton, G. The Role of Compensation Analysis in the Study of Wettability, Solubility, Disintegration, and Dissolution. *Int. J. Pharm.* **1990**, *66*, 175–182.
- Caramella, C.; Colombo, P.; Conte, U.; Ferrari, F.; La Manna, A. Water Uptake and Disintegrating Force Measurements: Towards a General Understanding of Disintegration Mechanisms. *Drug Dev. Ind. Pharm.* **1986**, *12*, 1749–1766.
- Carstensen, J.T.; Chowan, Z.T. Correlation Between Tablet Disintegration and In Vitro Dissolution. *Drug Dev. Ind. Pharm.* **1980**, *6*, 569–571.
- Kanig, J.L.; Rudnic, E.M. The Mechanisms of Disintegrant Action. *Pharm. Tech.* **1984**, *8*, 50–63.
- Mahmoud, H.M.; El-Shaboury, M.H. Effect of Certain Tablet Ingredients on the Penetration Rate of Powder Beds. **1985**, *94*, 125–131.
- Nogami, H.; Fukuzawa, H.; Nakai, Y. Studies on Tablet Disintegration. I. The Effect of Penetrating Rate on Tablet Disintegration. **1963**, *11*, 1389–1398.
- Pesonen, T.; Paronen, P.; Ketolainen Disintegrant Properties of an Agglomerated Cellulose Powder. *Int. J. Pharm.* **1989**, *57*, 139–147.
- Shangraw, R.F.; Mitrevj, A.; Shah, M.N. A New Era of Tablet Disintegrants. *Pharm. Tech.* **1980**, *4*, 49–57.

SURFACTANTS IN PHARMACEUTICAL PRODUCTS AND SYSTEMS

Owen I. Corrigan
Anne Marie Healy

University of Dublin, Trinity College, Dublin, Ireland

INTRODUCTION

Surface-active agents (surfactants) are substances which, at low concentrations, adsorb onto the surfaces or interfaces of a system and alter the surface or interfacial free energy and the surface or interfacial tension. Surface-active agents have a characteristic structure, possessing both polar (hydrophilic) and nonpolar (hydrophobic) regions in the same molecule. Thus surfactants are said to be amphipathic in nature. The wide range of uses for surfactants in pharmaceutical products and systems is the subject of this article.

PHYSICOCHEMICAL BACKGROUND

Surface and Interfacial Tension; Surface and Interfacial Free Energy

Atoms and molecules at surfaces and interfaces possess energies significantly different from those of the same species in the bulk phase. The term “surface” is usually reserved for the region between a condensed phase (liquid or solid) and a gas phase or vacuum, while the term “interface” is normally applied to the region between two condensed phases.

In the case of a liquid–gas interface, molecules of the liquid in the boundary can only develop attractive cohesive forces with molecules situated below and adjacent to them. They can develop attractive adhesive forces with molecules of the gaseous phase. However at the gas–liquid interface, these adhesive forces are quite small. The net effect is that molecules at the surface of the liquid have potential energies greater than those of similar molecules in the interior of the liquid and experience an inward force toward the bulk of liquid. This force pulls the molecules of the interface together and the surface contracts.

Thus, the surface of a liquid behaves as if it were in a state of tension—the surface tension (γ)—due to the contracting force acting in all directions in the plane of the surface.

In order to extend the surface of a liquid it is necessary to bring molecules from the interior to the surface against the inward pull. The work required to increase the surface area by unit area is termed the *surface free energy*.

At the interface between two condensed phases, the dissimilar molecules in the adjacent layers facing each other have potential energies greater than those of similar molecules in the respective bulk phases. This is due to the fact that cohesive forces between like molecules tend to be stronger than adhesive forces between dissimilar molecules. Thus the interfacial tension is the force per unit length existing at the interface between two immiscible or partially miscible condensed phases and the interfacial free energy is the work required to increase the interface by unit area.

Adsorption Phenomena

Adsorption may be defined as the process of enrichment of one or more substances at a surface (1) or as the taking up of one substance at the surface of another (2). It can occur at any type of interface. However, in the context of pharmaceutical systems the interfaces where surfactant adsorption is important are the gas–liquid, liquid–liquid, gas–solid, and liquid–solid interfaces.

Adsorption at liquid–liquid and liquid–gas interfaces

Considering a system of two immiscible phases (e.g., heptane and water), a surface-active molecule that is adsorbed at the interface between the two liquids will tend to orient itself with its hydrophilic end toward the more polar liquid (water), and its hydrophobic end toward the less polar liquid (heptane). Thus the surfactant molecules replace water and/or heptane molecules of the original interface. The interaction across the interface is then between the hydrophilic group of the surfactant and the water molecules on one side of the interface, and between the hydrophobic group of surfactant and heptane on the other side of the interface. These interactions are much stronger than the original interactions between the unlike

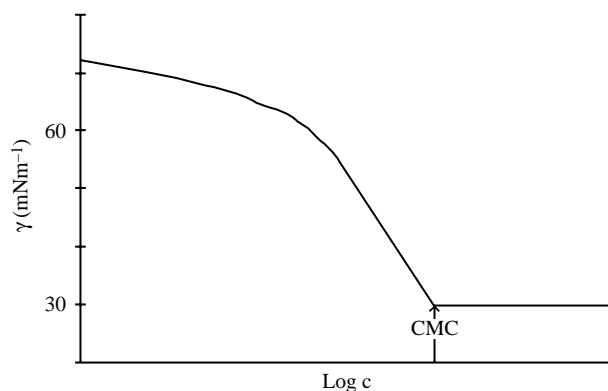


Fig. 1 Schematic plot of surface or interfacial tension (γ) versus logarithm of the surfactant concentration (c).

molecules of heptane and water; therefore the interfacial tension is significantly reduced by the adsorption of surfactant at the interface (i.e., the inward pull for each phase at the interface is reduced).

Air consists of molecules that are mainly nonpolar. Surface tension reduction by surfactants at the air–aqueous interface occurs due to adsorption of surfactants at the interface, with the hydrophilic end of the surfactant oriented toward the liquid. The presence of the surfactant molecules reduces the net inward pull toward the bulk liquid, and therefore reduces the surface tension.

The effect of a surfactant on the lowering of surface tension is shown in Fig. 1. The surface tension is lowered even at low concentrations of surfactant. As the surfactant concentration is increased, the surface layer becomes saturated with surfactant molecules, and micelles form within the bulk liquid as an alternative way of shielding the hydrophobic portions of the surfactants from the aqueous environment; the surface tension tends to a constant value. Micelles are small aggregates of surfactant in which the surfactant molecules are arranged in such a way that the hydrophobic ends are shielded from the surrounding aqueous environment. The concentration at which micelles first appear in solution is termed the critical micelle concentration (CMC).

Adsorption at solid–liquid interfaces

Adsorption of surfactant from an aqueous solution onto a solid surface may involve specific chemical interaction between the surfactant (adsorbate) and the surface (adsorbent).

Common interactions that can occur (3) include:

1. An ion-exchange process
2. An ion-pairing interaction
3. Acid–base interaction via either hydrogen bonding between substrate and adsorbate or Lewis acid–Lewis base reaction
4. Adsorption by polarization of π electrons, where the adsorbate contains electron-rich aromatic nuclei and the adsorbent has strongly positive sites
5. Adsorption by dispersion forces, i.e., London–van der Waals dispersion forces acting between adsorbate and adsorbent
6. Hydrophobic bonding.

Contact Angles and the Wetting of Solids

A drop of liquid when placed on a flat, homogeneous solid surface, comes to equilibrium, assuming a shape which minimizes the total free energy of the system. The angle between the liquid and the solid is called the contact angle (θ), the angle being measured through the liquid (Fig. 2). The contact angle may be calculated if the surface and interfacial tensions are known from Young's equation given in Eq. 1 or 2.

$$\gamma_{SA} = \gamma_{SL} + \gamma_{LA} \cos \theta \quad (1)$$

or

$$\cos \theta = \frac{\gamma_{SA} - \gamma_{SL}}{\gamma_{LA}} \quad (2)$$

where γ_{LA} is the surface tension of the liquid, γ_{SL} is the interfacial tension existing between the solid and liquid phases, and γ_{SA} is the surface tension (or surface free energy) of the solid. If $\theta < 90^\circ$, wetting of the solid is said to take place. If $\theta > 90^\circ$, wetting does not take place.

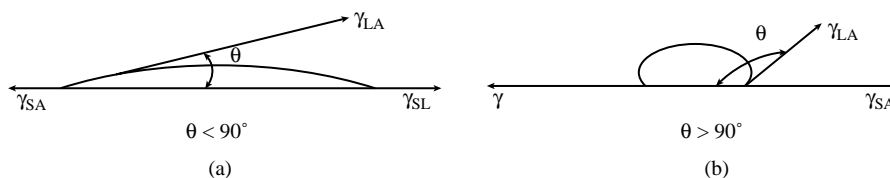


Fig. 2 Contact angles. In (a), $\theta < 90^\circ$, and wetting of the solid occurs; in (b), $\theta > 90^\circ$, and wetting does not take place.

The term “wetting” refers to the displacement from a surface of one fluid by another. It is most commonly applied to the displacement of air from a liquid or solid surface by water or an aqueous solution. The term “wetting agent” is applied to any substance that increases the ability of water or an aqueous solution to displace air from a liquid or solid surface.

For good wetting, $\cos \theta$ should be as close as possible to 1; that is, θ should be as close as possible to 0. From Young's equation, it can be seen that if γ_{LA} or γ_{SL} was minimized, $\cos \theta$ would be maximized, and wetting would be promoted.

Contact angles of water on powders of pharmaceutical importance are usually measured by preparing disks of the powder by compression or melting. However, compaction may change the surface, so making the measured result of little relevance. Contact angles on finely divided solids can also be determined by packing the powder into a tube and measuring the penetration of liquids into the packing.

Three types of wetting phenomena have been described (4): adhesion wetting, spreading wetting, and immersional wetting.

The way in which a particular system behaves depends on the interfacial energies between the solid substrate and any contacting liquid, and between the liquid and the second fluid (air). By manipulating these factors, the wetting process can be controlled. This may be achieved by the use of surfactants.

Modification of the wetting process by the use of surfactants

The effect of surfactants on the wetting process is a result of their adsorption at various interfaces with a resulting alteration of interfacial tensions. As has been noted from Young's equation, the wetting process is promoted if either γ_{LA} or γ_{SL} or both are reduced with γ_{SA} remaining unchanged. Surfactants almost always cause a reduction in γ_{LA} , however, the same cannot be said for γ_{SL} and the effect on the interfacial tension depends on the nature of the adsorption. Thus the addition of a surface-active agent to the system does not always promote wetting, and spreading may in fact be made more difficult (4).

If adsorption of the surfactant molecules at the solid-liquid interface occurs in such a manner that they are oriented with their polar ends toward the substrate and hydrophobic ends toward the liquid, the wettability of an aqueous solution is reduced. This orientation of surfactant molecules at the surface occurs if they are adsorbing to ionic or polar substrates (ion-exchange or ion-pairing mechanism). However, at higher concentrations of surfactant, the surfactant ions adsorb by hydrophobic

interaction with the already adsorbed layer, thus exposing their hydrophilic ends to the solution in such a way that the surface becomes more readily wetted. Thus, the contact angle may first increase and subsequently decrease following the addition of more surfactant to a solution. In contrast, where adsorption occurs onto nonpolar surfaces by, for example, van der Waals attraction, the surfactant molecules are oriented with their hydrophilic groups toward the liquid, the hydrophilicity of the substrate is increased, and it becomes more wettable.

The adsorption of surfactants onto solid surfaces is important with respect to their detergent properties, their use as wetting agents in solid pharmaceutical dosage forms, and as stabilizers for suspension formulations. The mode of action of surfactants in each of these systems is discussed further below.

Micellization

As mentioned previously, surfactant molecules have the ability to form micelles in aqueous solution. These micelles are colloidal-sized clusters of molecules. Micellization is an alternative to interfacial adsorption for removing hydrophobic groups from contact with the aqueous environment, thereby reducing the free energy of the system. In micelles, the hydrophobic groups are directed toward the center of the surfactant aggregate. In cases where there is little distortion of the surrounding solvent by the hydrophobic group, there is little tendency for micellization to occur, such as in water when the hydrophobic group of the surfactant is short or in the case of nonaqueous solvents.

One of the most important applications of micellization in the context of pharmaceuticals is their ability to solubilize drugs of poor aqueous solubility.

Micelles are dynamic species; there is a constant rapid interchange of surfactant molecules between the micelle and the bulk solution. Micelles cannot, therefore, be regarded as rigid structures with a defined shape, although an average micellar shape may be considered.

The main types of micelles recognized (3) are:

1. Small spherical
2. Elongated cylindrical, rodlike micelles with hemispherical ends (prolate ellipsoids)
3. Large, flat lamellar micelles (disklike extended oblate spheroids)
4. Vesicles, more or less spherical structures, consisting of lamellar micelles arranged in one or more concentric spheres.

In nonaqueous solvents, surfactants may form “inverted micelles” where the hydrophilic heads of the surfactant

molecules are present in the center of the micelle with the hydrocarbon chains extending outward into the solvent. Dipole–dipole interactions hold the hydrophilic heads of the surfactant molecules together in the core, and in certain cases hydrogen bonding between head groups can also occur.

Micellar shape can be affected by changes in temperature, concentration and the presence of added electrolyte to the liquid phase. Changes in any of these factors may affect micellar size, shape, and aggregation number (number of surfactant monomers in the micelle).

Phase Behavior of Surfactants

Equilibrium phase structures

As the concentration of a surfactant solution is increased, structures of the types depicted in Fig. 3 may be encountered (5). At concentrations well above the CMC,

a more ordered structuring of the solution occurs. Two main types of liquid crystalline phases may be identified: the middle phase, M, exhibiting a hexagonal array of indefinitely long, mutually parallel rods; and the neat phase, G, with a lamellar structure. The liquid crystalline hexagonal phase, like the micellar phase, can exist either in a normal or reverse orientation. The order of phase structures formed upon increasing surfactant concentration generally follows a well defined sequence (Fig. 4) with a “mirror plane” through the lamellar phase in such a way that normal phase structures can be considered to be “oil-in-water” and the reverse structures to be “water-in-oil” (5).

Modified phase structures

In addition to the equilibrium phase structures mentioned above, nonequilibrium surfactant phase structures exist that are also finding applications in drug delivery. Vesicular forms of surfactants are generally formed by dispersing lamellar phases in an excess of water (or nonaqueous polar

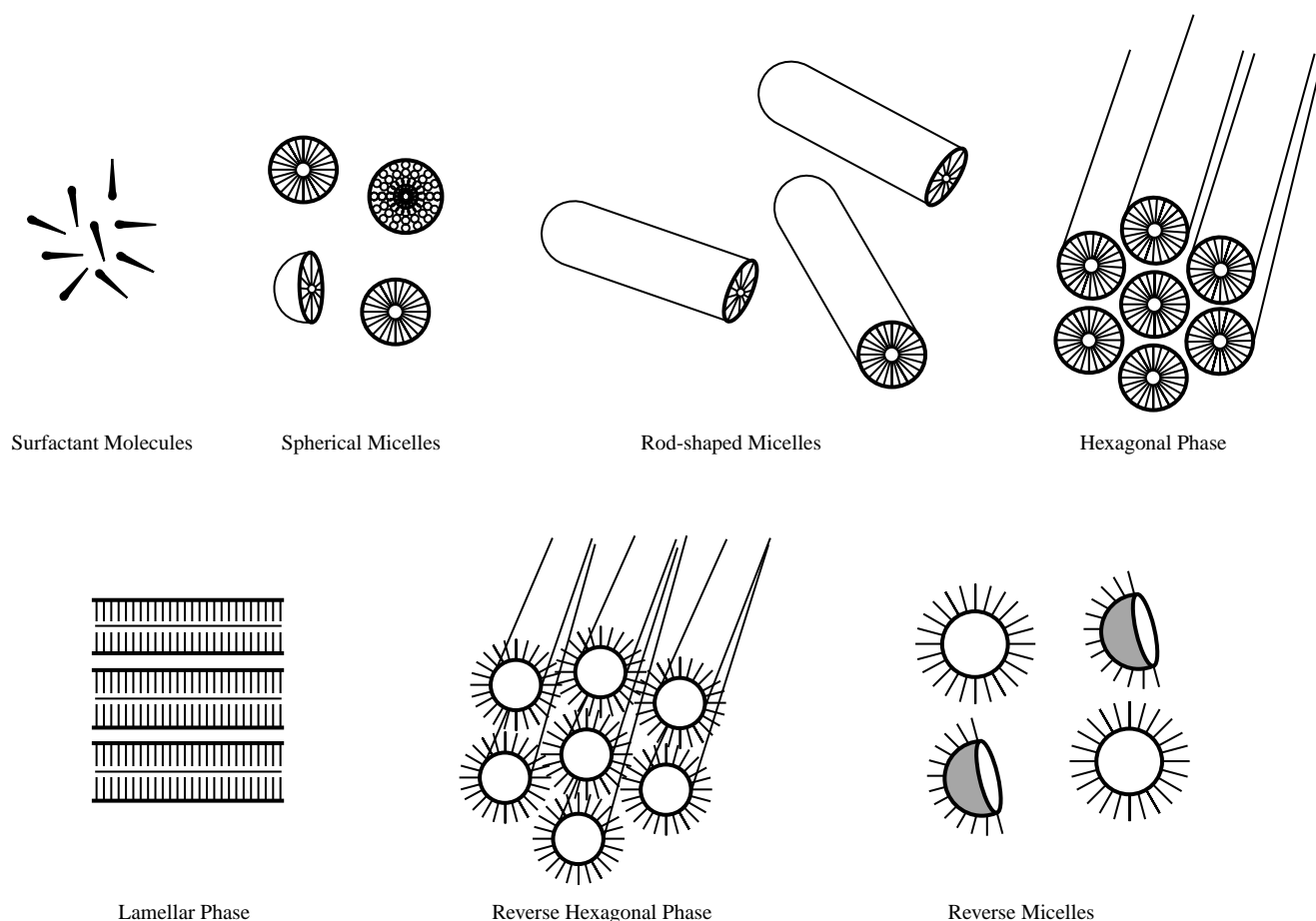


Fig. 3 Equilibrium phase structures of surfactant molecules. (From Lawrence, M.J. *Chem. Soc. Rev.* **1994**, 23 (6), 417–424, reproduced by permission of the Royal Society of Chemistry.)

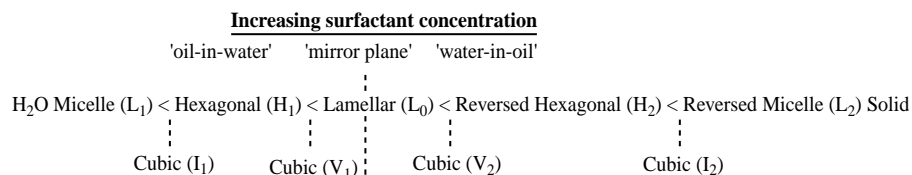


Fig. 4 Idealized phase sequence in surfactant-water systems. (From Lawrence, M.J. *Chem. Soc. Rev.* **1994**, 23 (6), 417–424, reproduced by permission of the Royal Society of Chemistry.)

solvents such as ethylene glycol or dimethylformamide) or, in the case of reversed vesicles, in an excess of oil. With most surfactants, vesicles are nonequilibrium structures that will eventually re-equilibrate back into the lamellar phases from which they originated. Vesicles are structural analogs of liposomes (discussed later); they are approximately spherical structures and have the ability to “solubilize” both lipid soluble and water soluble agents.

Several of the phase structures produced by surfactants have potential as carriers and vehicles for drugs and also as targeting systems, used to direct the drug to a specific site in the body (5).

SURFACTANT CLASSIFICATION

Surfactant molecules may be classified based on the nature of the hydrophilic group within the molecule. The four main groups of surfactants are defined as follows:

1. *Anionic surfactants*, where the hydrophilic group carries a negative charge, such as carboxyl (RCOO^-), sulphonate (RSO_3^-) or sulphate (ROSO_3^-). Examples of pharmaceutical importance include potassium laurate, $\text{CH}_3(\text{CH}_2)_{10}\text{COO}^- \text{K}^+$, and sodium lauryl sulphate, $\text{CH}_3(\text{CH}_2)_{11}\text{SO}_4^- \text{Na}^+$.
2. *Cationic surfactants*, where the hydrophilic group carries a positive charge (e.g., quaternary ammonium halides, $\text{R}_4\text{N}^+\text{Cl}^-$). Examples of pharmaceutical importance include cetrимide, a mixture consisting mainly of tetradecyl (ca. 68%), dodecyl (ca. 22%), and hexadecyltrimethylammonium bromides (ca. 7%), as well as benzalkonium chloride, a mixture of alkylbenzyltrimethylammonium chlorides of the general formula $[\text{C}_6\text{H}_5\text{CH}_2\text{N}^+(\text{CH}_3)_2\text{R}]\text{Cl}^-$, where R represents a mixture of the alkyls from C_8H_{17} to $\text{C}_{18}\text{H}_{37}$.
3. *Ampholytic surfactants (also called zwitterionic surfactants)*, where the molecule contains, or can potentially contain, both a negative and a positive charge, (e.g., the sulfobetaines, $\text{RN}^+(\text{CH}_3)_2\text{CH}_2\text{CH}_2\text{SO}_3^-$). Examples of pharmaceutical importance include

N-Dodecyl-*N,N*-Dimethylbetaine, $\text{C}_{12}\text{H}_{25}\text{N}^+(\text{CH}_3)_2\text{CH}_2\text{COO}^-$.

4. *Nonionic surfactants*, where the hydrophile carries no charge but derives its water solubility from highly polar groups such as hydroxyl or polyoxyethylene ($\text{OCH}_2\text{CH}_2\text{O}$) groups. Examples of pharmaceutical importance include polyoxyethylated glycol monoethers (e.g. cetomacrogol), sorbitan esters (Spans[®]) and polysorbates (Tweens[®]).

Tables 1–4 in the article *Surfactants in Pharmaceutical Products and Systems* in Volume 14 of the first edition of this encyclopedia (6), together with the references cited therein, give listings of some of the surfactants most commonly used in pharmaceuticals, along with the purpose(s) for which they are usually employed.

SURFACTANT USES IN PHARMACEUTICAL PREPARATIONS

Because of their unique functional properties, surfactants find a wide range of uses in pharmaceutical preparations. These include, depending on the type of product, improving the solubility or stability of a drug in a liquid preparation, stabilizing and modifying the texture of a semisolid preparation, or altering the flow properties of a granulate, thus aiding in the processing of the final tablet dosage form. In addition to their use as excipients to improve the physical and chemical characteristics of the formulation, surfactants may be included to improve the efficacy or bioperformance of the product. The properties of surfactants are such that they can alter the thermodynamic activity, solubility, diffusion, disintegration, and dissolution rate of a drug. Each of these parameters influence the rate and extent of drug absorption. Furthermore, surfactants can exert direct effects on biological membranes thus altering drug transport across the membrane. The overall effect of inclusion of a surfactant in a pharmaceutical formulation is complex and may be beyond those initially intended.

Surfactants may reduce the effectiveness of antimicrobials or preservatives included in a formulation (7). They also have the capacity to damage biological membranes.

LIQUID SYSTEMS

Solutions

Surfactants as solubilizing agents

Solubilization can be defined as “the preparation of a thermodynamically stable isotropic solution of a substance normally insoluble or very slightly soluble in a given solvent by the introduction of an additional amphiphilic component or components” (4). The amphiphilic components (surfactants) must be introduced at a concentration at or above their critical micelle concentrations. Simple micellar systems (and reverse micellar) as well as liquid crystalline phases and vesicles referred to above are all capable of solubilization. In liquid crystalline phases and vesicles, a ternary system is formed on incorporation of the solubilize and thus these anisotropic systems are not strictly in accordance with the definition given above (4).

Solubilization by micelles

The location of a solubilized molecule in a micelle is determined primarily by the chemical structure of the solubilize. Solubilization can occur at a number of different sites in a micelle:

1. On the surface, at the micelle–solvent interface,
2. Between the hydrophilic head groups,
3. In the palisades layer, i.e., between the hydrophilic groups and the first few carbon atoms of the hydrophobic groups that comprise the outer regions of the micelle core,
4. More deeply in the palisades layer, and
5. In the micelle inner core.

In aqueous systems, nonpolar additives such as hydrocarbons tend to be intimately associated with the hydrocarbon core of the micelle. Polar and semipolar materials, such as fatty acids and alcohols are usually located in the palisades layer, the depth of penetration depending on the ratio of polar to nonpolar structures in the solubilize molecule.

In reverse micelles (formed in nonpolar solvent systems containing surfactant), polar additives may be solubilized in the core where a polar interaction of head groups occurs.

A preferred location of the solubilize molecule within the micelle is largely dictated by chemical structure. However, solubilized systems are dynamic and the location

of molecules within the micelle changes rapidly with time. Solubilization in surfactant aqueous systems above the critical micelle concentration offers one pathway for the formulation of poorly soluble drugs (7). From a quantitative point of view, the solubilization process above the CMC may be considered to involve a simple partition phenomenon between an aqueous and a micellar phase. Thus the relationship between surfactant concentration C_m and drug solubility C_{tot} is given by Eq. 3.

$$C_{tot} = C_s + PC_s C_m \quad (3)$$

where C_s is the drug solubility in the absence of surface-active agent and P is the distribution coefficient of drug between the micelle and bulk phases. A plot of C_{tot} versus C_m is linear with a slope of PC_s , which is the solubilizing capacity of the micelle (8).

The effect of altering the pH of the vehicle, in the case of a partly ionized drug will be to alter the apparent partition coefficient. Thus the effect of increasing the pH of a vehicle containing an acidic drug is to reduce the proportion of drug in the micellar phase. If the surfactant is a weak electrolyte, it may induce a concentration-dependent change in pH thus altering drug partitioning and solubility (9).

In general the solubilizing capacity for surfactants with the same hydrocarbon chain length increases in the order anionic < cationic < nonionic, the effect being attributed to a corresponding increase in the area per head group, leading to looser micelles with less dense hydrocarbon cores which can accommodate more solute.

The solubilizing capacity for a given surfactant system is a complex function of the physicochemical properties of the two components which, in turn, influence the location or sites where the drug is bound to the micelle. The molar volume of the solubilize together with its lipophilicity are important factors, the former reducing and the latter increasing solubilization (9).

Many pharmaceutical products contain a number of solutes potentially capable of being solubilized within the micellar phase. Thus competition can occur between solutes resulting in an altered solubilizing capacity. Furthermore, the addition of a second highly solubilized component to form a mixed micellar system may greatly alter the structure, size and solubilizing capacity of the system, thereby greatly enhancing drug solubility.

Solubilization has been used for many years in the formulation of phenolic antiseptic and disinfectant solutions. In the case of Cresol and Soap Solution (Lysol) and Chloroxylenol Solution B.P., soap micelles are used to solubilize the phenolic substances. The soap

(anionic surfactant) is formed by reaction of potassium hydroxide with a suitable oil such as linseed oil (in Cresol and Soap Solution) or castor oil (in Chloroxyleneol Solution). The solubilizing potential of surfactant solutions for hydrophobic species has also been exploited in the design of cholelitholytic solvents for gallstone dissolution with some limited success.

Stability of drugs in solubilized systems

Solubilization of a drug by incorporation into micelles may affect its stability (7). In the micelle, the molecular environment of the drug molecules changes their proximity and orientation with respect to each other, which may affect activity. In a micelle, the drug molecules may be protected from attacking species such as hydronium or hydroxide ions and the stability of the drug may be increased. The difference in environment between the micellar and bulk aqueous phases may be such that reaction rates may be radically changed by the transfer of solute to micelles. Micellar systems may be used to deliberately alter the rates and directions of chemical reactions (7).

AB block copolymer micelles

It is well known that block copolymers in a selective solvent (a good solvent for one block but a nonsolvent for the other) form a micellar structure through the association of the insoluble segments (10). In contrast with micelles formed from low molecular weight surfactants, block copolymer micelles dissociate slowly to free polymeric chains. They have a greater capacity for solubilizing aromatic molecules and express lower CMCs. The AB block copolymers are considered useful vehicles for hydrophobic drugs.

Only a few block copolymers form micellar structures in aqueous milieu. One example is a series of polyethylene oxide/polypropylene oxide/polyethylene oxide block copolymers known as Pluronics (tradename) or poloxamers. The poloxamers have been used widely in pharmaceuticals, particularly as emulsifiers for intravenous lipids (7). At low concentrations, poloxamer monomers are thought to form monomolecular micelles by a change in configuration in solution (7). At higher concentrations, aggregation of the monomolecular micelles occurs. The aggregates so formed show the ability to solubilize drugs and increase the stability of solubilized materials. Poloxamers have low toxicity and their solubilization capabilities might prove useful in the delivery of hydrophobic drugs, although multimolecular micelle formation with core-shell structure is uncertain under physiological conditions (11).

Other block copolymers have been prepared and studied as formulation adjuvants for hydrophobic drugs, e.g., poly(ethylene oxide)/poly(aspartic acid) and poly(ethylene oxide)/poly(β -benzyl-L-aspartate) block copolymers have been used with adriamycin (12, 13).

Suspensions

If a suspension is to be produced by a dispersion technique (as opposed to precipitation techniques), surfactants may be used in the formulation to aid dispersion of the solid particles in the liquid. This is particularly important if the powder is not readily wetted by the liquid vehicle. Surfactants can reduce the interfacial tension between the solid particles and the liquid vehicle. The advancing contact angle is reduced, and wetting of the solid particles promoted. Such a system is said to be deflocculated. The inclusion of a surface-active agent to improve powder wettability can often improve the bioavailability of the formulation.

The forces at the surface of a particle affect the degree of flocculation and agglomeration in a suspension. Particles dispersed in a liquid medium may become charged in one of two main ways. Ionic species present in solution may be adsorbed at the surface or, alternatively charges on the surface may arise due to ionization of groups (such as carboxyl groups for example) which may be located at the surface. The surface charge will influence the distribution of ions in the aqueous medium surrounding the solid particles. The result is the formation of what is known as an "electric double layer." If the surface charge is positive, immediately adjacent to the surface will be a region of tightly bound solvent molecules and negative counterions. Thus, the first layer is tightly bound, while the second layer (which still contains an excess of negative ions) is more diffuse (14). As two particles approach each other in aqueous medium, a weak attractive force exists just beyond the range of the double-layer-repulsive forces. This region is responsible for the particle interaction termed "flocculation."

Flocculated particles are weakly bonded, settle rapidly, do not form a cake and so are easily resuspended. For this reason it is frequently desirable to promote flocculation in a suspension.

The inclusion of surfactants in the formulation is one way of achieving what is known as "controlled flocculation." Surfactants can cause dispersed solids to flocculate by a number of different mechanisms (3). The first is where there is an electrostatic attraction of surfactant ions to oppositely charged sites on the particle surface, resulting in a lowering of the electrical energy

barrier to the close approach of two particles to each other. Flocculation may also occur by a bridging mechanism. A long (usually polymeric) surfactant molecule containing functional groups at various sites may adsorb onto sites on the surface of adjacent particles, holding the particles together in a loose arrangement. Alternatively if the surfactant molecules adsorb in such a manner that the molecule extends into the liquid phase, interaction of the extended portions of surfactant molecules adsorbed to different particles result in bridging of those particles.

Another method of employing surfactants to achieve flocculation is to first treat the particles with an ionic surfactant to disperse them. A readily soluble electrolyte is then added which has the effect of compressing the electrical double layer surrounding each particle, allowing flocculation to occur. Subsequent dilution of this type of system will redisperse it (due to a decrease in electrolyte concentration).

Emulsions

Emulsification is one of the most important applications of surface-active agents in pharmaceutical systems. The phenomenon has been extensively studied and many books and chapters of books have been devoted to the subject.

Macroemulsions are either oil in water (o/w) or water in oil (w/o). The type of emulsion formed depends largely on the emulsifying agent used; the process and relative proportions of the oil and water phases are less important. In general, o/w emulsions are produced by emulsifying agents that are more soluble in the water phase than in the oil phase, and w/o emulsions are produced by emulsifying agents that are more soluble in the oil phase. It is also possible to form a multiple emulsion. For example, a small water or aqueous solution droplet may be enclosed in a larger oil droplet which is itself dispersed in an aqueous phase. Such a system is referred to as a "water-in-oil-in-water" (w/o/w) emulsion. It is also possible to form an o/w/o emulsion.

Many medicinal agents which have an unpalatable taste or texture can be made more acceptable for oral administration when formulated as emulsions. Mineral-oil-based laxatives, oil soluble vitamins and high-fat nutritive preparations are frequently administered in the form of o/w emulsions. It has been shown that in some cases the absorption of drugs may be enhanced if formulated as emulsions (15). Emulsions (o/w) have also been used for the intravenous administration of lipid nutrients. Radiopaque emulsions have been used as diagnostic agents in X-ray examinations.

Emulsification is widely used in pharmaceutical products for external application such as lotions and creams, and in aerosol products to form foams. Semisolid emulsified formulations are discussed below.

Based on the size of the dispersed particles or droplets, emulsions may be classified (16) into

1. Macroemulsions, droplets $\sim 0.1-50\ \mu\text{m}$, opaque emulsions
2. Microemulsions, droplets $10-100\ \text{nm}$, transparent dispersions.

Stabilization of the dispersion of one immiscible liquid in another requires the addition of an emulsifying agent which is commonly a surfactant or a mixture of surfactants.

In the formation of an emulsion, one of the two immiscible liquids is broken up into droplets which are dispersed in the other liquid. The dispersion of one liquid in another immiscible liquid leads to a large increase in interfacial free energy because of the increase in the area of the interface. The emulsifying agent stabilises the emulsion by adsorbing at the liquid-liquid interface as an oriented interfacial film. This film reduces the interfacial tension between the liquids and also decreases the rate of coalescence of the dispersed droplets by forming mechanical, steric and/or electrical barriers around them.

A strong mechanical barrier lessens the chance of droplets coalescing on collision. For maximum mechanical stability, the interfacial film of the adsorbed surfactants should be close packed with strong lateral interactions. For this reason, a mixture of two or more surfactants is commonly used as the emulsifying agent, such as a combination of a water-soluble surfactant and an oil-soluble surfactant. In pharmaceutical (o/w) systems a mixture of a sorbitol ester (Span[®]) with a polyoxyethyleneated sorbitol ester (Tween[®]) is often used. The water soluble Tween tends to have a greater interaction with the aqueous phase, its hydrophilic group extending further into the water than that of the nonoxyethyleneated ester. This is believed to facilitate interaction of the hydrophobic groups of the molecules, as they can approach each other more closely in the interfacial film.

The interfacial film can also stabilize the emulsion by producing repulsive electrical forces between approaching droplets. This repulsion is due to surface charge on the droplets. The surface charge effect is believed to be important only in the case of o/w emulsions. The source of surface charge is the hydrophilic head of the surfactant molecules which is oriented toward the aqueous continuous phase. In emulsions containing ionic surfactant molecules, the charge on the disperse phase droplets

is due to the amphipathic ion. In the case of nonionic surfactants, the charge may arise either from adsorption of ions from the aqueous phase or from frictional contact between droplets and the aqueous phase. In the latter case, the phase with the higher dielectric constant is positively charged (3).

Microemulsions

Microemulsions consist of large or “swollen” micelles, containing an internal phase similar to that found in a solubilized solution (16). Unlike macroemulsions, they appear as clear, transparent solutions. They tend to be more thermodynamically stable than macroemulsions and can have essentially infinite lifetimes assuming no change in composition, temperature and pressure. This is in contrast to macroemulsions which, although they may remain stable for long periods of time, will ultimately undergo phase separation to attain a minimum in free energy. Microemulsions can generally be obtained by gentle mixing of the ingredients of the emulsion. In this respect, they differ from macroemulsions which require intense agitation for their formation. Microemulsions are usually prepared with more than one surfactant or using a mixture of surfactant and cosurfactant (e.g., a polar compound of intermediate chain length).

Microemulsions have been studied as drug delivery systems, in particular for topical and transdermal drug delivery (17, 18).

Microspherical particles prepared by emulsification processes

Emulsification–evaporation processes are widely used in the preparation of polymer based microspherical drug-loaded particulates. For example, hydrophobic drug-loaded PLA (polylactic acid) or PGLA (polylactide-co-glycolide) biodegradable microspheres are often prepared from emulsions containing a non aqueous dispersed phase of dichloromethane containing the drug and polymer in an aqueous continuous phase. For the preparation of hydrophilic drug loaded microspheres a double-emulsion process may be necessary. The nature of the surfactants used to stabilize the emulsion phases can greatly influence the size, size distribution, surface morphology, loading, drug release, and bioperformance of the final multiparticulate product.

Aerosols

Surfactants are found in both solution and suspension formulations of metered dose inhalers (MDIs). The most common surfactants found in pressurised aerosol preparations include sorbitan trioleate (Span 85), oleic acid,

and lecithins at concentrations of 0.1–2.0% (w/w). These agents are nonvolatile liquids which dissolve in the propellant blend. Their function in the formulation is to provide lubrication for the metering valves and, in the case of suspension formulations, to maintain the disperse nature of the drug.

The three surfactants commonly used in chlorofluorocarbon (CFC)-based MDI formulations are insoluble in the CFC-replacement propellants, hydrofluoroalkane (HFA) 134a and HFA 227. Possible formulation alternatives involve the use of an adjuvant such as ethanol to aid dissolution of the surfactant or a novel surfactant. Several companies have investigated novel materials among which are fluorosurfactants, polyoxyethylenes and drugs coated with surfactant (19).

Controlled flocculation in metered-dose aerosol suspensions

Controlled flocculation is a widely used technique for stabilizing suspended systems. The aim is to alter particle surface charge or to achieve particle separation via steric hindrance with the help of appropriate stabilizing excipients. However this is particularly difficult to achieve in nonpolar systems such as suspensions in CFC (or HFA) propellants. Controlled flocculation to optimise the stabilisation of MDIs has been recommended by Ranucci et al. (20) but disputed by Hickey et al. (21).

Liposomes

Liposomes are single- or multilayered phospholipids vesicles. They are roughly spherical in shape and consist of lipid bilayers alternating with aqueous regions.

Liposomes have shown potential as drug delivery systems. The exact location of a drug molecule in a liposome depends on its physicochemical composition and the composition of the lipids. Water soluble drugs may be included in the aqueous phase, and oil-soluble drugs may be added to the membrane-forming phospholipid. An extensive account of the pharmaceutical use of liposomes is found in the article “*Liposomes as Pharmaceutical Dosage Forms*,” by Y. Barenholz, and D.J.A., Crommelin, Volume 9 of the first edition of this encyclopedia (22).

SEMISOLID SYSTEMS

Surfactants are major constituents of pharmaceutical, cosmetic, and food semisolid formulations, many of which are emulsions, either oil in water (o/w) or water in oil (w/o). They are included for their stabilizing, wetting,

solubilizing, detergent and penetration-enhancing properties.

Water-in-oil emulsions traditionally contain surfactants of natural origin such as cholesterol, wool fat, wool alcohols, lanolin, divalent salts of fatty acids soaps, calcium oleate and/or synthetic agents of low hydrophilic-lipophilic balance (HLB) (indicating high lipophilicity), such as Spans (fatty acid esters of sorbitan). An example of such a product is Oily Cream B.P. which consists of a 1:1 mixture of wool alcohols and water.

Oil-in-water creams, for topical use, generally contain mixed emulsifiers/surfactants; one of which is a water soluble surfactant with a high HLB, the other being an amphiphile, usually a long chain fatty alcohol (e.g., of chain length C_{14} to C_{18}) or acid (e.g., palmitic or stearic). The water soluble surfactant may be anionic (e.g., sodium lauryl sulphate), cationic (e.g., cetrimide), or nonionic (e.g., cetomacrogol, Tweens).

These mixed-surfactant systems are used not only for their ability to form complex condensed films at the liquid-liquid interface, enhancing the stability of the emulsion, but also because of their ability to impart "body" to the product, resulting in a semisolid product rather than a liquid. Mixed emulsifiers control the consistency of a cream by forming a viscoelastic network throughout the continuous phase of the emulsion. The network results from the interaction of the mixed emulsifier with water, forming a liquid crystalline phase.

Foams

Emulsification is used in aerosol products to produce foams which are generally formulated as o/w emulsions. The liquified propellant forms the disperse phase of the emulsion, and the medication is usually in the aqueous continuous phase. On discharge from the pressurised container, the propellant vaporizes to form bubbles which remain trapped within the aqueous phase giving rise to a foam. These are referred to as "stable foam" products. Nonaqueous stable foams may also be formulated, where the water is replaced by various glycols such as polyethylene glycol. "Quick breaking foams" result when the propellant is in the external phase. The product is emitted as a foam and collapses into a liquid.

Biological Effects on Percutaneous Absorption

Surfactants—traditionally common constituents and stabilizers of topical vehicles, ranging from hydrophobic agents such as oleic acid to hydrophilic sodium lauryl

sulphate—have been tested as penetration enhancers to improve transdermal drug delivery. Ionic surfactants are thought to enhance transdermal absorption by disordering the lipid layer of the stratum corneum and by denaturation of keratin. The use of penetration enhancers in general, and surfactants in particular, in transdermal therapeutic systems has been reviewed by Walters (23).

SOLID DOSAGE FORMS

Surface-active agents have been widely shown to enhance drug dissolution rates. This may be due to wetting effects, resulting in increased surface area, effects on solubility and effective diffusion coefficient or a combination of effects. Consequently surfactants have been included in tablet and capsule formulations to improve wetting and deaggregation of drug particles and thus increase the surface area of particles available for dissolution.

This wetting effect is found to be operative at concentrations below the CMC. The effect of surfactants on the dissolution of solids is complex. In addition to effects on the available surface area, surfactants in concentrations above the CMC can increase drug solubility and hence the effective concentration gradient. However they also reduce the effective rate of drug diffusion as a consequence of drug solubilization within micelles. Models to quantify the effect of surfactant concentration on drug dissolution have been developed (24). For solids whose dissolution is under significant surface control, surfactants may further influence the dissolution process. In this regard the enhancing effect of surfactants on the dissolution rate of cholesterol has been widely studied (25).

Hard Gelatin Capsules and Tablets

Wetting agents

Surfactants are used in capsule (26) and tablet formulations as wetting agents to aid dissolution.

Lubricants, anti-adherents, and glidants

The primary function of tablet lubricants is to reduce the friction arising at the interface of tablet and die walls during compression and ejection. Lubricants also possess antiadherent (prevention of sticking to the punch and, to a lesser extent, to the die wall) and glidant (improvement of flow characteristics of powders or granulates) characteristics and are useful in the processing of hard gelatin capsules.

Magnesium stearate is used extensively as a lubricant in tablet manufacture. It is an example of a “boundary lubricant,” that is, the polar regions of the molecule adhere to the metal surface of the die wall (in tablet manufacture). Adsorption of magnesium stearate to the powder or granule surfaces also prevents agglomeration of the feed material and aids flow.

Lubricants may be classified as water-soluble or water-insoluble. The latter are generally more effective than water-soluble lubricants and can be used at a lower concentration (27). Common water-insoluble lubricants (which are surfactants) include magnesium stearate, calcium stearate, sodium stearate, and stearic acid; water-soluble lubricants include sodium lauryl sulphate and magnesium lauryl sulphate.

Sodium lauryl sulphate is used in the production of hard gelatin capsules where it is added to the gelatin solution during the preparation stage. The stainless steel molds are lubricated prior to dipping into the gelatin solution and sodium lauryl sulphate is added to reduce the surface tension of the mix and cause the mold pins to wet more uniformly (28).

Solid Dispersion Systems

The bioavailability of hydrophobic drugs can be increased by strategies designed to enhance the dissolution rate of the drug. This has been achieved in many cases by forming a solid dispersion of the drug in a suitable carrier, often a hydrophilic polymer such as polyethylene glycol (PEG) or polyvinylpyrrolidone (PVP). The drug is dispersed in the carrier by coprecipitation from a suitable solution containing both drug and carrier, by melting both components together, or by some other process involving a phase change. By using relatively high concentrations of carrier and a rapid precipitating process, the drug may form as an amorphous or molecularly dispersed high energy phase in the carrier. A number of workers have used surfactants as the carrier material to achieve this enhanced dissolution effect. Among the surfactants employed are polyoxyethylene stearate, Renex 650, poloxamer 188, Texafor AIP deoxycholic acid, and Tweens and Spans. Surfactants have also been added to conventional drug-polymer solid dispersions to further improve drug release properties. Sjökvist et al. (29) found that the incorporation of sodium dodecyl sulphate (1–2%) in griseofulvin (3–10%)-PEG solid dispersions eliminated any traces of crystalline drug, griseofulvin being present as a solid solution. Other three-component solid dispersions containing surfactants have also been reported such as Tween 20-Griseofulvin-PEG and Tween 20-Oxodipine-

PEG (30). Problems have been reported however as to the physical stability of surfactant containing systems, dissolution rates decreasing over a 12-month period (31).

Matrix Systems

Drug release from nondisintegrating inert matrices, fabricated from hydrophobic carriers such as polyethylene, is improved by the presence of surfactants in the dissolution medium. Drug release was shown to be a function of the pore size distribution of the matrix and the permeation pressure of the release media defined by its surface tension and contact angle. Inclusion of dioctyl sodium succinate, which reduced the contact angle below 90°, greatly enhanced drug release; increasing the concentration of polysorbate in the range of 0.001–0.1% had the same effect (32). Surfactants have also been included in matrix-type drug delivery systems to aid penetration of the dissolution medium thus increasing the rate and extent of drug release.

Suppositories

Several nonionic surface-active materials have been developed as suppositories vehicles. Many of these bases, known as water-dispersible bases, can be used for the formulation of both water-soluble and oil-soluble drugs (33). The surfactants most commonly used are the polyoxyethylene sorbitan fatty acid esters (Tweens), the polyoxyethylene stearates, and the sorbitan fatty acid esters (Spans). These surfactants may be used alone, blended, or with other suppository base materials to yield a wide range of melting points and consistencies.

Surface-active agents are widely used in combination with other suppository bases. The inclusion of these agents in the formulation may improve the wetting and water-absorption properties of the suppository. In addition, emulsifying surfactants help to keep insoluble substances suspended in a fatty base suppository (33).

The inclusion of a surfactant in the suppository formulation may enhance the rectal absorption of drugs. The effect has been attributed to the formation of mixed micelles. It has been suggested that the presence of the micelle facilitates the incorporation of the lipid component of the mixed micelle into the biological membrane. This lipid then enhances the fluidity and permeability of the membrane to the poorly absorbed drug. It appears that the colorectal mucous membrane is more sensitive to the effects of mixed micelles than the gastrointestinal membrane of the small intestine.

Surfactant Influence on Drug Absorption from the Gastrointestinal Tract

In the context of oral dosage forms containing surfactants, these agents may play a role in reducing the rate of gastric emptying and retarding the movement of drug to the absorption site by increasing the viscosity of the formulation. This is thought to be especially true of polyoxyethylene derivatives. Bile salts, which are physiological surfactants, have been shown to affect the rate of gastric emptying. The presence of bile salts in the stomach has also been shown to affect ionic movement across the gastric mucosa, thus increasing the movement of hydrogen and chloride ions out of the lumen.

Surfactants may also affect the rate and extent of drug absorption by exerting an influence on the permeability of the biomembrane. Competitive binding of the surfactant to the membrane protein is considered to be partially responsible for enhanced drug absorption in many cases. Alternatively, the enhancement may be due to allosteric rearrangement of the membrane protein which is triggered by the binding of one or more permeating species.

Nakanishi et al. (34) studied the effect of a range of surfactants on the rectal absorption of sulphaguanidine and found absorption to be increased. The increase was associated with histological changes in rectal membrane, increasing the rectal permeability. The same authors found that surfactants such as sodium deoxycholate and sodium dodecyl sulphate used together with the chelating agent EDTA could increase the rectal absorption of macromolecules such as inulin, insulin, and albumin.

The membrane effects of surfactants are explained by a combination of membrane-surfactant binding, disruption of membranes through solubilization into lipoproteins, proteins, and mixed micelles, protein-protein interactions, and selective solubilization of some membrane components by the surfactant. The structure of the surfactant may play a role in determining the range and extent of the influence of a particular surfactant on drug absorption. It appears that the greatest effect is achieved by molecules having a C12–C16 hydrocarbon chain, polyoxyethylene chain lengths between 10 and 20, and molecular areas between 1.0 and 1.6 nm² (4). These effects, in the case of drugs of low aqueous solubility, are in addition to the higher absorption rate, arising from an increase in drug solubility (35, 36).

Surfactants, at high concentrations, exhibit some toxicity and have the ability in many cases to disrupt a membrane. Both ionic and nonionic surfactants have been shown to assist the breakdown of the mucous layer covering the epithelium and at high concentrations are thought to interfere with the membrane itself, which may

lead to disruption of membrane metabolism, particularly with regard to enzyme systems associated with the membrane. Adverse reactions to drug formulation agents including surfactants have been reviewed by Weiner and Bernstein (37).

DIRECT ACTIONS OF SURFACTANTS

Detergents

Detergents are surfactants that are used for the removal of foreign matter from a solid surface. The process involves many of the actions specific to surfactant molecules. The surfactant requires good wetting properties to ensure good contact with the solid surface. It must also have the ability to remove dirt into the bulk liquid. This is achieved by a lowering the dirt-liquid and solid-liquid interfacial tensions, thus reducing the work of adhesion between the dirt and the solid and enabling the dirt to be readily detached. Once detached, adsorption of surfactant at the dirt particle surface prevents deposition, allowing the dirt to be washed away. If the dirt is oily it may be emulsified or solubilized by the surfactant.

Antimicrobial Activity

Significant antimicrobial effects have been associated with cationic surfactants, in particular the quaternary compounds. The action mechanism of quaternary surfactants involves disruption of the cell membrane, protein denaturation, and enzyme inhibition. Quaternary compounds are able to lyse cells at relatively low concentration, resulting in leakage of cell contents into the surrounding medium. Quaternary ammonium and some phosphonium surfactants are used as topical disinfectants in commercial dermatological products, in surgical hand scrubs, and in the irrigation of skin wounds. The most commonly used quaternary compounds employed for their antimicrobial effects are cetylpyridinium chloride, benzalkonium chloride, benzethonium chloride and cetyltrimethylammonium bromide (38). Other surfactants, containing more than one quaternary (or positively ionizable group) are among the most active substances known in terms of antimicrobial activity. Included in this group are dequalinium acetate and chlorhexidine gluconate which have been used in throat lozenges and mouthwashes. The lysis of cells can also occur in the presence of anionic surfactants, although these are in general weaker in their antimicrobial activity. A wide range of anionics, in particular sodium lauryl sulphate and its homologs, finds wide application in mouthwashes (38).

Respiratory Distress Syndrome (RDS)

In 1959, surfactant deficiency was identified as the major pathogenic factor in respiratory distress syndrome in infants (39). Pulmonary surfactant is a complex mixture of phospholipids, neutral lipids, and specific proteins which spread as a monolayer at the air-liquid interfaces of the lung and lower surface tension at end-expiration thus preventing alveolar collapse. If the amount or quality of endogenous surfactant is inadequate, inspiratory pressure and the work of breathing must increase in order to re-expand the alveoli with each breath and permit adequate gas exchange. As the infant grows tired, progressive respiratory failure occurs.

Phosphatidylcholine is the major component of endogenous surfactant, constituting about 60% of total phospholipids, and dipalmitoylphosphatidylcholine (DPPC) is the primary surface-tension lowering phospholipid.

The surfactant replacement therapy treatment used may be either "natural" or "artificial." Natural surfactants are derived from bovine or porcine animal lungs or human amniotic fluid. Synthetic or artificial surfactants are composed of DPPC and spreading agents such as unsaturated phosphatidylglycerol or tyloxapol and hexadecanol (40).

NATURALLY OCCURRING SURFACTANTS

Of the naturally occurring surfactants, the bile salts and phospholipids are of particular importance.

Phospholipids

The phospholipids are widely found in biological membranes and can be used as emulsifiers especially for intravenous fat emulsions, and as a key component of liposomes. The elucidation of factors governing the solubilization of drugs in phospholipid dispersions can provide some clues as to the biological role of interactions with lipid systems *in vivo* (4). Phospholipids have been discussed above and in reference (22) in the context of liposomes.

Bile Salts

Bile salts are carboxylic acids (C22–C28) with a cyclopentenophenanthrene nucleus containing a branched chain of 3–9 carbon atoms ending in a carboxyl group. Structurally they form micelles which are different from the conventional spherical micelles

synonymous with amphiphiles having a distinct hydrocarbon chain. The hydrophobic feature of the bile salts is associated with one surface of the steroid nucleus, and consequently intermolecular association is much more restricted. Primary and secondary micelles have been proposed, the former consisting of two to four molecules, the latter being composed of aggregates of the primary micelles. The CMC is less distinct and is highly dependent on the structure of the specific bile salt, in particular the number of hydroxy groups and their orientation.

Many studies have been completed in order to assess the effect of bile salts on the bioavailability of poorly soluble drugs. Bile salts for example, have been shown to enhance the absorption of sulphaguanidine and urogastone. Bile salts may also play a role in enhancing the transport of a compound from the lumen of the intestine to the systemic circulation. Such absorption involves overcoming the resistance of the aqueous boundary layer and the membrane epithelium to the passage of the drug.

Bile salts readily form mixed micelles with lipid-like molecules such as lecithins or fatty acids. These mixed micelles are structurally very different from the simple micelles and generally have a much greater solubilizing capacity for hydrophobic molecules, both biological and synthetic. The solubility of DDT, a nonpolar, water insoluble molecule, for example, in bile salt micellar solution can be increased to a far greater extent by the addition of unsaturated long chain fatty acids, probably because of mixed micelle formation.

Saponins

Saponins are glycosides found in certain plants which are characterized by their property of producing a frothing aqueous solution. The term "saponin" is derived from the Latin "sapo" meaning soap. Plant materials containing saponins have been used for a long time in many parts of the world for their detergent properties, for example, in Europe, the root of *Saponaria officinalis* and in South America the bark of *Quillaja saponaria* (41).

The saponin structure is either of the steroidal (commonly tetracyclic triterpenoids) or pentacyclic triterpenoid type. Triterpenoid saponins are found, for example in Quillaia bark and in liquorice root. Quillaia B.P. is defined as the dried inner part of the bark of *Quillaja saponaria* and other species of *Quillaja* and is used as an emulsifying agent. Liquorice, the root of which also contains triterpenoid saponins, has long been used in pharmacy as a flavoring agent, demulcent, and mild expectorant.

Iscoms

Iscoms (Immune-stimulating complexes) are stable complexes of cholesterol, phospholipid, and Quil A (derived from *Quillaja saponaria*) in size ranges from 40 to 100 nm. They are promising carriers for antigens in subunit vaccines. Iscoms are considered to be multi-micellar structures, shaped and stabilized by hydrophobic interactions, electrostatic repulsion, steric factors and possibly hydrogen bonds (42). Protection has been achieved after immunization with iscom-based vaccines, against viruses like the Epstein–Barr virus (43) and the measles virus (44).

SURFACE ACTIVITY OF DRUGS

A large number of drug molecules exhibit surface activity, that is, they tend to accumulate at interfaces, depress surface tension and associate to form aggregates in solution. Although the hydrophobic groups of most drugs are aromatic, they still behave like typical surfactants (which possess flexible hydrophobic chains), inasmuch as these aromatic groups have a high degree of flexibility. (Drugs that exhibit association characteristics typical of surface active agents and may reduce surface tension are reviewed in Ref. 4.)

Most of the drugs form micelles at concentrations that they do not attain in vivo. It is therefore their surface activity, rather than their self-association tendency which is more important biologically. Surface-active drugs will tend to bind hydrophobically to proteins and other macromolecules and to associate with other amphipathic substances such as bile salts, phospholipids, and receptors. As with other surface-active agents, surface-active drugs may interact directly with biological membranes. The possible biological implications of surface activity is discussed by Attwood and Florence (4) in relation to the phenothiazine tranquillisers and local anesthetics.

REFERENCES

1. Rupperecht, H.; Lee, G. Adsorption at Solid Surfaces. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1988; 1, 73–114.
2. Myers, D. Surfaces and Interfaces: General Concepts. *Surfaces, Interfaces, and Colloids: Principles and Applications*; VCH Publishers Inc.: New York, 1991, 7–24.
3. Rosen, M.J. *Surfactants and Interfacial Phenomena*; Wiley Interscience: New York, 1989.
4. Attwood, D.; Florence, A.T. *Surfactant Systems. Their Chemistry, Pharmacy and Biology*; Chapman and Hall: London, New York, 1983.
5. Lawrence, M.J. Surfactant Systems: Their Use in Drug Delivery. *Chem. Soc. Rev.* **1994**, 23 (6), 417–424.
6. Corrigan, O.I.; Healy, A.M. Surfactants in Pharmaceutical Products and Systems. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1996; 14, 295–331.
7. Florence, A.T. Drug Solubilization in Surfactant Systems, Drugs and the Pharmaceutical Sciences. *Techniques of Solubilization of Drugs*; Yalkowsky, S., Swarbrick, J., Eds.; Marcel Dekker, Inc.: New York, 1982; 12, 15–89.
8. Ong, J.T.H.; Manoukian, E. Micellar Solubilization of Timobesone Acetate in Aqueous and Aqueous Propylene Glycol Solutions of Nonionic Surfactants. *Pharm. Res.* **1988**, 5 (11), 704–708.
9. Faehelebom, K.M.S.; Timoney, R.F.; Corrigan, O.I. Micellar Solubilization of Clofazimine Analogs in Aqueous Solutions of Ionic and Nonionic Surfactants. *Pharm. Res.* **1993**, 10 (4), 631–634.
10. Riess, G.; Hurtrez, G.; Bahadur, P. Block Copolymers. *Encyclopedia of Polymer Science and Engineering*, 2nd Ed.; Mark, H.F., Bikales, N.M., Overberger, C.G., Menges, G., Eds.; Wiley-Interscience: New York, 1985; 2, 324–434.
11. Kataoka, K.; Kwon, G.S.; Yokoyama, M.; Okano, T.; Sakurai, Y. Block-Copolymer Micelles as Vehicles for Drug Delivery. *J. Control. Release* **1993**, 24 (1–3), 119–132.
12. Yokoyama, M.; Miyauchi, M.; Yamada, N.; Okano, T.; Sakurai, Y.; Kataoka, K.; Inoue, S. Polymer Micelles as Novel Drug Carrier—Adriamycin-Conjugated Poly(Ethylene Glycol) Poly(Aspartic Acid) Block Copolymer. *J. Control. Release* **1990**, 11 (1–3), 269–278.
13. Kwon, G.S.; Naito, M.; Yokoyama, M.; Okano, T.; Sakurai, Y.; Kataoka, K. Physical Entrapment of Adriamycin in AB Block Copolymer Micelles. *Pharm. Res.* **1995**, 12 (2), 192–195.
14. Martin, A.N.; Bustamante, P. *Physical Pharmacy: Physical Chemical Principles in the Pharmaceutical Sciences*, 4th Ed.; Lea and Febiger: Philadelphia, 1993, 386–388, 541–542.
15. Carrigan, P.J.; Bates, T.R. Biopharmaceutics of Drugs Administered in Lipid-Containing Dosage Forms I: GI Absorption of Griseofulvin from an Oil-in-Water Emulsion in the Rat. *J. Pharm. Sci.* **1973**, 62, 1476–1479.
16. Eccleston, G. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1992; 5, 137–188.
17. Osborne, D.W.; Ward, A.J.I.; O'Neill, K.J. Microemulsions as Topical Drug Delivery Vehicles I. Characterization of a Model System. *Drug Dev. Ind. Pharm.* **1988**, 14 (9), 1203–1219.
18. Linn, E.E.; Pohland, R.C.; Byrd, T.K. Microemulsions for Intradermal Delivery of Cetyl Alcohol and Octyl Dimethyl PABA. *Drug Dev. Ind. Pharm.* **1990**, 16 (6), 899–920.
19. Bowman, P.A.; Greenleaf, D. Non-CFC Metered Dose Inhalers: The Patent Landscape. *Int. J. Pharm.* **1999**, 86 (1), 91–94.

20. Ranucci, J.A.; Dixit, S.; Bray, R., Jr.; Goldman, D. Application of Controlled Flocculation to Metered Dose Aerosols Suspensions. *Pharm. Res.* **1987**, *4* (2 suppl.), S-25.
21. Hickey, A.J.; Dalby, R.N.; Byron, P.R. Effects of Surfactants on Aerosol Powders in Suspension, Implications for Airborne Particle Size. *Int. J. Pharm.* **1988**, *42*, 267–270.
22. Barenholz, Y.; Crommelin, D.J.A. Liposomes as Pharmaceutical Dosage Forms. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C. Eds.; Marcel Dekker, Inc. New York, 1994; 9, 1–39.
23. Walters, K.A. Penetration Enhancers and Their Use in Transdermal Therapeutic Systems. *Transdermal Drug Delivery*; Drugs and the Pharmaceutical Sciences; Hadgraft, J., Guy, R.H., Eds.; Marcel Dekker, Inc.: New York, 1989; 35, 197–246.
24. Higuchi, W.I. Diffusional Models Useful in Biopharmaceutics. Drug Release Rate Processes. *J. Pharm. Sci.* **1967**, *56*, 315–324.
25. Higuchi, W.I.; Su, C.C.; Park, J.Y.; Gulari, E. Mechanism of Cholesterol Gallstone Dissolution. Analysis of the Kinetics of Cholesterol Monohydrate Dissolution in Taurocholate/Lecithin Solutions by the Mazer, Benedek, and Carey Models. *J. Phys. Chem.* **1981**, *85* (2), 127–129.
26. Newton, J.M.; Rowley, G.; Törnblom, J-F.V. Further Studies on the Effect of Additives on the Release of Drug from Hard Gelatin Capsules. *J. Pharm. Pharmac.* **1971**, *23* (suppl.), 156S–160S.
27. Banker, G.S.; Peck, G.E.; Baley, G. Tablet Formulation and Design. *Pharmaceutical Dosage Forms: Tablets*; Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1980; 1, 61–107.
28. Jones, B.E. Gelatin Additives, Substitutes and Extenders. *Hard Capsules: Development and Technology*; Ridgway, K., Ed.; The Pharmaceutical Press: London, 1987, 54.
29. Sjökvist, E.; Nyström, C.; Aldén, M. Physicochemical Aspects of Drug Release, 13. The Effect of Sodium Dodecyl-Sulfate Additions on the Structure and Dissolution of a Drug in Solid Dispersions. *Int. J. Pharm.* **1991**, *69* (1), 53–62.
30. Veiga, M.D.; Escobar, C.; Bernad, M.J. Dissolution Behavior of Drugs from Binary and Ternary Systems. *Int. J. Pharm.* **1993**, *93* (1–3), 215–220.
31. Sjökvist-Saers, E.; Nyström, C.; Aldén, M. Physicochemical Aspects of Drug Release, 16. The Effect of Storage on Drug Dissolution from Solid Dispersions and the Influence of Cooling Rate and Incorporation of Surfactant. *Int. J. Pharm.* **1993**, *90* (2), 105–118.
32. Singh, P.; Desai, S.J.; Simonelli, A.P.; Higuchi, W.I. Role of Wetting on the Rate of Drug Release from Inert Matrices. *J. Pharm. Sci.* **1968**, *57* (2), 217–226.
33. Coben, L.J.; Lieberman, H.A. Suppositories. *The Theory and Practice of Industrial Pharmacy*, 3rd Ed.; Lachman, L., Lieberman, H.A., Kanig, J.L., Eds.; Lea & Febiger: Philadelphia, 1986; 564–588.
34. Nakanishi, K.; Masada, M.; Nadai, T. Effect of Pharmaceutical Adjuvants on the Rectal Permeability to Drugs. IV. Effect of Pharmaceutical Adjuvants on the Rectal Permeability to Macromolecular Compounds in the Rat. *Chem. Pharm. Bull.* **1984**, *32*, 1628–1632.
35. O'Reilly, J.R.; Corrigan, O.I.; O'Driscoll, C.M. The Effect of Simple Micellar Systems on the Solubility and Intestinal-Absorption of Clofazimine (B663) in the Anesthetized Rat. *Int. J. Pharm.* **1994**, *105* (2), 137–146.
36. O'Reilly, J.R.; Corrigan, O.I.; O'Driscoll, C.M. The Effect of Mixed Micellar Systems, Bile-Salt Fatty-Acids, on the Solubility and Intestinal-Absorption of Clofazimine (B663) in the Anesthetized Rat. *Int. J. Pharm.* **1994**, *109* (2), 147–154.
37. Weiner, M.; Bernstein, I.L. *Adverse Reactions to Drug Formulation Agents*; Marcel Dekker, Inc.: New York, 1989.
38. Rieger, M.M. Surfactants. *Pharmaceutical Dosage Forms: Disperse Systems*, 2nd Ed.; Lieberman, H.A., Rieger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1996; 1, 211–286.
39. Avery, M.E.; Mead, J. Surface Properties in Relation to Atelectasis and Hyaline Membrane Disease. *Am. J. Dis. Child.* **1959**, *97*, 517–526.
40. Fujiwara, T.; Maeta, H.; Chida, S.; Morita, T.; Watabe, Y.; Abe, T. Artificial Surfactant Therapy in Hyaline-Membrane Disease. *Lancet* **1980**, *1*, 55–59.
41. Evans, W.C. Saponins, Cardioactive Drugs and Other Steroids. *Trease and Evans' Pharmacognosy*, 14th Ed.; WB Saunders Company Ltd.: London, 1996; 293–309.
42. Kersten, G.F.A. Aspects of Iscoms. Structural, Pharmaceutical and Adjuvant Properties; Ph.D. Thesis, University of Utrecht, 1990; 93–104.
43. Morgan, A.J.; Finerty, S.; Lövgren, K.; Scullion, F.T.; Morein, B. Prevention of Epstein-Barr (EB)Virus-Induced Lymphoma in Cottontop Tamarins by Vaccination with the EB Virus Envelope Glycoprotein GP 340 Incorporated into Immune-Stimulating Complexes. *J. Gen. Virol.* **1988**, *69*, 2093–2096.
44. de Vries, P.; van Binnendijk, R.; van der Marel, P.; van Wezel, A.L.; Voorma, H.O.; Sundquist, B.; UytdeHaag, F.G.C.M.; Osterhaus, A.D.M.E. Measles Virus Fusion Protein Presented in an Immune-Stimulating Complex (iscom) Induces Haemolysis-Inhibiting and Fusion-Inhibiting Antibodies, Virus Specific T Cells and Protection in Mice. *J. Gen. Virol.* **1988**, *69*, 549–559.

SUSPENSIONS

Robert A. Nash

Consultant, Mahwah, New Jersey, U.S.A.

INTRODUCTION

A suspension is a particular class or type of dispersion or dispersed system in which the internal or suspended phase is dispersed uniformly by mechanical agitation throughout the external phase (called the suspending medium or vehicle). The internal phase, consisting of a homogeneous or heterogeneous distribution of solid particles having a specific size range, is maintained uniformly throughout the suspending vehicle with the aid of a single or a particular combination of suspending agent(s). In addition, unlike in a solution, the suspended particles exhibit a minimum degree of solubility in the external phase. In a colloidal suspension, the solids are less than about 1 μm in size. In a coarse suspension, they are larger than about 1 μm . The practical upper limit for individual suspendable solid particles in coarse suspensions is approximately 50–75 μm . When one or more types of solid particles that constitute the internal phase are pharmaceutically useful and/or physiologically active, the system is known as a pharmaceutical suspension.

In an emulsion, the particles of the internal phase are spherical or liquid droplets that are dispersed throughout a liquid external phase. Even though the particles may be liquid only at elevated temperatures (50–80°C) and semisolid or rigid at room temperature, as long as they appear spherical on careful microscopic examination, they are generally considered to be emulsified rather than suspended. Thus, a clue to the presence of a suspended particle is its lack of sphericity or its definitive lattice structure. Exceptions to this general rule are spherical microspheres and related spherical solid microparticles.

CLASSIFICATION

Martin and Bustamante (1) list three general classes of pharmaceutical suspensions: orally administered (sometimes referred to as mixtures), externally applied (topical lotions), and injectable (parenteral).

Oral Suspensions

The solids content of an oral suspension may vary considerably. For example, antibiotic preparations may contain 125–500 mg of active solid material in a 5-ml (teaspoonful) dose, whereas a drop concentrate may provide the same amount of insoluble drug in a 1- or 2-mL dose. Antacids and radiopaque suspensions contain relatively high amounts of suspended material for oral administration. The vehicle may be a syrup, a sorbitol solution, or a gum-thickened, water-containing artificial sweetener because in addition to ingredients, safety, taste, and mouthfeel are important formulation considerations. In the case of limited shelf life (low chemical stability of the insoluble drug), the dosage form may be prepared as a dry granulation or powder mixture that is reconstituted with water prior to use.

Topical Suspensions

Historically, the externally applied “shake lotion” is the oldest example of a pharmaceutical suspension. Calamine Lotion USP, as well as other dermatological preparations, is closely associated with the technical development of the pharmaceutical suspension (2). Because safety and toxicity considerations are most readily dealt with in terms of dermatological acceptability, many useful suspending agents were first introduced in topical formulations (3). In addition, the protective action and cosmetic properties of topical lotions usually require the use of high concentrations of the dispersed phase, often in excess of 20%. Therefore, topical lotions represent the best example of suspensions that exhibit low settling rates (4). Various pharmaceutical vehicles have been used in the preparation of topical lotions, including diluted oil-in-water or water-in-oil emulsion bases, dermatological pastes, magmas, and clay suspensions.

Parenteral Suspensions

The solids content of parenteral suspensions is usually between 0.5 and 5.0%, except for insoluble forms of penicillin, in which concentrations of the antibiotic may exceed 30%. These sterile preparations are designed for

intramuscular, intradermal, intra-lesional, intra-articular, or subcutaneous administration. The viscosity of a parenteral suspension should be low enough to facilitate injection. Common vehicles for parenteral suspensions include preserved 0.9% saline solution or a parenterally acceptable vegetable oil. The primary factor governing the selection of injectable ingredients is safety. Ophthalmic suspensions that are instilled into the eye must be prepared in a sterile manner. The vehicle is employed is essentially isotonic and aqueous in composition.

UTILITY OF SUSPENSIONS

A suspension is often chosen as pharmaceutical dosage form for drugs insoluble in water and aqueous fluids at the dosage required for administration and when attempts to solubilize the drug would compromise stability and safety. For oral administration, the taste of a bitter or unpleasant drug can often be masked by choosing an insoluble form of the active drug.

An aqueous suspension is a useful oral dosage form for administering insoluble or poorly soluble drugs. The large surface area of the dispersed drug particles often facilitate absorption. Unlike drug particles contained in tablets or capsules, the dissolution of drug particles in suspension and subsequent absorption commence upon dilution in gastrointestinal fluids. Finely divided particles dissolve faster and have higher relative solubilities than do similar macroparticles.

The parenteral suspension is an ideal dosage form for prolonged or “depot” release. In the administration of a drug as an aqueous or oleaginous suspension into subcutaneous or muscular tissue, the drug is deposited at the injection site. The depot acts as a reservoir, slowly releasing drug at a rate related to both the intrinsic aqueous solubility of the drug form and the type of suspending vehicle, either aqueous or oily for the purpose of maintaining prolonged systemic absorption of the drug from the injection site.

HYDROPHILIC/HYDROPHOBIC SOLIDS

Insoluble solids, regardless of particle size, that have a relatively low interfacial tension and are readily wetted by water are called *hydrophilic* solids. These solids include clays (bentonite, kaolin, talc, magnesium aluminum silicate); bismuth salts, barium sulfate, carbonates, hydroxides, or oxides of calcium, magnesium, zinc, and aluminum; and titanium dioxide. The hydrophilicity of a powder surface can be investigated with the help of

moisture absorption studies in which the solid particles are exposed to varying relative humidities. Insoluble powders that absorb moisture below relative humidities of 70–80% at room temperature are said to be hydrophilic solids.

Fine insoluble solids that are not easily wetted by water and have a relatively high interfacial tension are referred to as *hydrophobic* solids. These include a large number of low-density organic materials and pharmaceutical substances, such as charcoal and sulfur. The hydrophobic nature of the latter group is accentuated by entrapped air adsorbed on the surface of these particles. Hydrophobic materials may be wetted by oils and semipolar liquids and are called lipophilic solids; conversely, hydrophilic materials behave like lipophobic solids in oils.

Hydrophilic solids can be suspended easily in water without the aid of a water-dispersible surfactant or wetting agent, and conversely hydrophobic solids can be suspended in oils and nonpolar vehicles without the use of lipid-soluble surfactants. The crystal density of hydrophilic solids usually ranges from 1.5 to 6.9 g/cm³, whereas the crystal density of hydrophobic solids usually ranges from 0.9 to 2.2 g/cm³.

PARTICLE SIZE CONSIDERATIONS

The mean particle diameter and the particle size distributions of suspended insoluble drugs are important considerations in formulating stable pharmaceutical suspensions. Hiestand (5) defines the lower limit of coarse suspensions as those containing particles larger than 0.1 μm . Except for a number of clays, oxides, charcoal, and pigments, the average particle size of most drugs and pharmaceutical rarely falls below 1 μm . Although most submicron inorganic excipients appear to behave like hydrophilic solids, most insoluble drugs and pharmaceutical excipients are usually soft, organic, essentially crystalline hydrophobic solids ranging in particle size from several microns to several hundred or more.

Drug particle size is an important factor influencing product appearance, settling rates, drug solubility, in vivo absorption, resuspendability, and overall stability of pharmaceutical suspensions. Insoluble drug particles are seldom uniform spheres or cubes, even after size reduction and classification. Wide distributions in particle size often lead to high-density suspensions. Systems with widely differing particle shapes (plates, needles, filaments, and prisms) frequently produce low-density slurries. The growth over time of unprotected, slightly soluble drug solids and changes in their particle size distribution in suspension are

a serious problem. Crystal growth of particles is usually attributed to one or more of the following mechanisms:

- “Oswalt ripening” or the growth of large particles at the expense of smaller ones, because of a difference in solubility rates of different size particles. For example, the increase in the solubility rate of a 0.2- μm particle, is 13%. For a 2- μm particle, it is 1%, and for particles above 20 μm , it is negligible.
- Crystal growth due to temperature fluctuations on storage is of minor importance, unless suspensions are subjected to temperature variations of 20°C or more.
- A polymorphic form may change to another more stable crystalline form; changes in crystal habit may be related to the degree of solvation or hydration.
- Crystal growth may also arise when the more energetic amorphous or glassy forms of a drug exhibit significantly higher initial solubility in water than the corresponding crystalline forms.
- Size reduction by crushing and grinding can produce particles whose different surfaces exhibit high and low solubility rates. This effect can be related to differences in the free surface energy introduced during comminution (grinding).

Crystal growth and changes in particle size distribution can generally be controlled by employing one or several of the following procedures and techniques.

1. Selection of particles with a narrow range of particle sizes, such as microcrystals between 1 and 10 μm .
2. Selection of a stable crystalline drug form that usually exhibits lower solubility in water. The crystalline form that is physically most stable usually has the highest melting point.
3. High-energy milling should not be used during particle size reduction. Microcrystals are best formed by controlled precipitation techniques or shock cooling.
4. A water-dispersible surfactant wetting agent dissipates the free surface energy of particles by reducing the interfacial tension between the solid and the suspending vehicle.
5. A protective colloid, such as gelatin, gum, or a cellulosic derivative, is used to form a film barrier around the particles, inhibiting dissolution and subsequent crystal growth.
6. The viscosity of the suspending vehicle is increased to retard particle dissolution and subsequent crystal growth.
7. Temperature extremes during product storage (freeze–thaw conditioning) must not occur.

8. Supersaturation favors the formation of needlelike crystals and should be avoided.
9. Rapid or shock cooling and high agitation favor the formation of thin, small crystals and should be avoided. Slow crystallization by evaporation yields compact crystals.
10. Experimentation with different crystallizing solvents is recommended to change crystal size and shape.
11. Impurities and foreign substances during crystallization affect the reproducibility and aggregation potential of many drug particle systems.
12. Constant crystallizing conditions are essential. Batch-to-batch variation in crystal size and shape is often associated with poor control of processing and crystallization procedures.

Variations in assay results can be avoided by the preparation of homogeneous, well-mixed, or nonsettling fine particle suspensions (size 1–10 μm). Particle size reduction results in slow, more uniform settling rates. The bioavailability of drugs is improved by reducing the size of suspension particles. Furthermore, drug particles smaller than 20 μm produce less pain and tissue irritation when injected parenterally. However, fine particles may have a deleterious effect on chemical stability because of their high dissolution rate.

Particle Size Reduction

Drug solids are easy to grind. Reduction to a particle size of about 50–75 μm usually produces a free-flowing powder. Solids containing particles smaller than 50 μm tend to aggregate or agglomerate in the dry state. Furthermore, below 10–50 μm , the increased free surface energy, as evidenced by the cohesion of small particles, becomes a factor interfering with further size reduction. The powder may become damp, especially if there is a tendency to attract moisture. Material tends to “ball up,” which indicates that the agglomerated masses are larger than the individual particles contained within.

As the pores between powder particles become smaller with decreasing particle size, increases in surface area facilitate liquid penetration. Aggregates behave like hydrophobic solids, entrapping air and becoming difficult to wet.

The most efficient method of producing fine particles is by dry milling prior to manufacture of the suspension. Dispersion equipment, such as colloid mills or homogenizers are normally used to wet-mill finished suspensions to break up poorly wetted fine particle aggregates or agglomerates. Among the several methods of producing small, reasonably uniform drug particles are micropulverization, fluid energy grinding, controlled precipitation, and

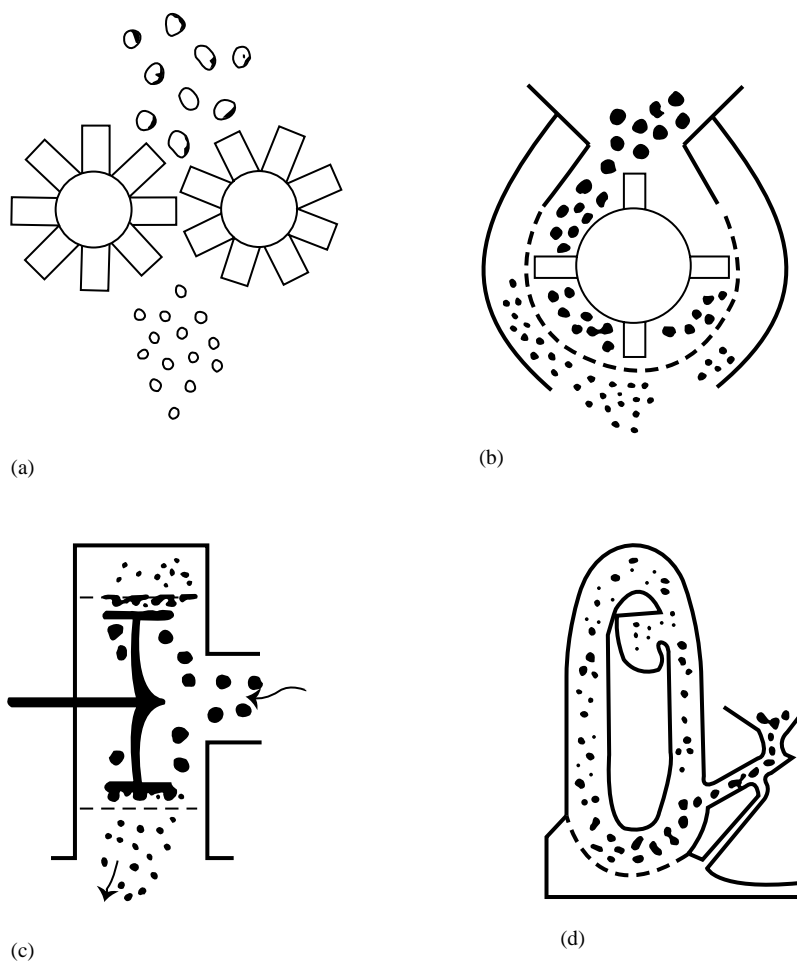


Fig. 1 The four basic types of size reduction equipment used to produce fine solid particles: (a) Crushers and shredders. (b) Hammermills. (c) Colloid mills. (d) Fluid energy mills.

spray drying. Figure 1 illustrates the four basic types of size reduction equipment used in the pharmaceutical industry to produce fine powder particles.

Micropulverization

Micropulverization is one of the most rapid, convenient, and inexpensive methods of producing fine drug powders. The milling equipment includes hammermills, micropulverizers, universal mills, end-runner mills, and ball mills. Micropulverizers are highspeed attrition or impact mills specially adapted for fine grinding. Some mills are fitted with classifiers to facilitate particle separation by centrifugation. Because ultrafine particles smaller than $10\text{ }\mu\text{m}$ are rarely produced, buildup of electrostatic charge on the surface of milled powder is encountered only occasionally. The main disadvantage of micropulverization is the large distribution of particle sizes

produced, normally in the range of $10\text{--}50\text{ }\mu\text{m}$ or higher. Nevertheless, these powders are satisfactory for the preparation of most oral and topical suspensions.

Fluid energy grinding

The process of fluid energy grinding, also referred to as jet milling or micronizing, is the most effective method for reducing particles below $10\text{ }\mu\text{m}$. The ultrafine particles are produced by the shearing action of high-velocity streams of compressed air on particles in a confined space. The main disadvantage of fluid energy grinding is the high electrostatic charge built up on the surfaces of the milled powder, which makes powder classification and collection exceedingly difficult. However, because it is important that a majority of drug particles in parenteral suspensions are below $10\text{ }\mu\text{m}$, fluid energy grinding is the most convenient method for their production.

Controlled crystallization

A solvent that dissolves a solid readily at room temperature may serve as a crystallizing medium when mixed with another solvent in which the compound is only sparingly soluble. A solution that is nearly saturated at a temperature about 10°C below the boiling point of the solvent combination is prepared in a temperature ranging between 60 and 150°C. Separation of microcrystals from such hot concentrated solutions is commonly induced by cooling and stirring. However, when supersaturation is obtained by agitation and shock cooling of the hot solution and through the rapid introduction of another cold miscible solvent in which the drug is only sparingly soluble, formation of minute crystalline particles (nucleates) proceeds without appreciable crystal growth, and uniform microcrystals of the drug are thus obtained. In addition, ultrasonic methods have been used during shock cooling to promote microcrystal formation.

Spray drying

Particles of microcrystalline size can also be obtained by spray-drying procedures, resulting in a porous, free-flowing, easily wetted, essentially monodispersed powder. With proper control of process variables, spherical particles are obtained that may be coated with agents to aid suspension and promote stability. However, the process is not normally considered for the preparation of ultrafine powders.

PHYSICAL ASPECTS

Stability of Suspensions

The chemical stability of a drug in suspension is controlled by the fact that the rate of degradation is related to the concentration of the drug in aqueous solution rather than to the total concentration of the drug in the product. Generally, a suspended drug decomposes only in solution as the solid phase gradually dissolves; that is, a solution concentration equal to the solubility of the drug is maintained. Drug degradation in a suspension usually follows zero-order kinetics, with the rate constant solely dependent on the saturation solubility of the drug in solution. Reducing the solubility of the suspended drug decreases the rate of degradation. The constancy of potency may be improved by selecting a pH value or range where the drug is least soluble or by replacing the drug with a less soluble derivative or salt. Decomposition in suspensions may also be described as a diffusion-controlled process or by catalysis initiated by environmental factors such as oxygen, light, and trace metals.

As a rule, the problem of suspension stability is complicated by the fact that pharmaceutical suspensions are affected at least as much by physical as by chemical factors.

Because a suspension exists in more than one state (liquid and solid), there are different ways in which the system can undergo chemical or physical change. Some of the more obvious difficulties involved in stability predictions are based on the fact that simple hydrostatic relationships (Stoke's law, etc.) used to define settling rates assume a spherical, deflocculated, free-falling particle that is not affected by particle–particle or particle–vehicle interactions. Suspensions that exhibit non-Newtonian flow are difficult to define in terms of the basic stability relationships. In addition, suspensions that are described in terms of a single representative particle do not reflect the influence of the entire particle size distribution.

Chemical stability predictions are sometimes complicated by the difficulty of determining the pH value of suspensions, which often changes because of surface coating of electrodes and differences between bulk-suspension and supernatant-vehicle readings. Accelerated elevated temperature stability testing often has a pronounced adverse effect on viscosity, particle solubility, and size distribution.

Deflocculated Suspensions

The empirical method of producing pharmaceutical suspensions is based on an attempt to prepare a stable deflocculated dispersion of a drug in a suitable suspension vehicle. In the past, a series of suspensions was often prepared using different concentrations of a favored suspending agent to identify the formulation that would produce the most homogeneous looking ("smooth") and stable suspension. The finished preparation was usually passed through a homogenizer or colloid mill to improve the final dispersion. Smooth looking viscous suspensions were produced; however, after some time, the drug particles settled slowly, forming a tightly packed sediment that was almost impossible to resuspend even with vigorous shaking. Primary particles or small aggregates, reaching the bottom of the container during sedimentation (settling), slipped past each other and produced compact layers of solids. The interparticle interaction in such compact sediments is relatively high because the interparticle distances are small, and the weak van der Waals forces of attraction, which decreases exponentially with distance, are appreciable. Such conditions frequently lead to the undesirable phenomenon of "caking or claying" and require extensive agitation for resuspension.

The physical instability of these early deflocculated suspensions led to other methods of producing physically stable pharmaceutical suspensions. For example, the density of the vehicle was made to equal or approach the crystal density of the suspended drug particles. If the drug particles are small enough and the vehicle sufficiently viscous, the particles remain suspended indefinitely in accordance with Stokes' law. Because the crystal density of most organic drug particles lies somewhere between 1.1 and 1.5 g/cm³, the only liquid vehicles for oral use with densities (at 25°C) high enough to be considered are Sorbitol Solution USP (1.29 g/cm³), Syrup USP (1.31 g/cm³), and high-fructose corn syrup (1.41 g/cm³). In practice, however, it is extremely difficult to prepare oral suspensions by the matched-density technique alone because dilution with water and other liquids reduces the vehicle density. Nevertheless, the use of high-density liquids as suspending vehicles often has a beneficial effect on physical stability.

The approximate settling velocities for nonflocculated particles of various average size were determined in the range of 0.2–200 μ m at density differences between solid particles and the suspending liquid of 0.2 and 2.0 g/cm³, respectively, with an absolute viscosity of the suspending liquid between 1 and 1000 centipoise (cP or mPa s). Terminal settling rates were calculated by a method described by Carpenter (6) for two concentrations of suspended solids, namely less than 2 and 20% vol. According to the analysis, permanent-type suspensions, which exhibit a settling viscosity of less than 0.14 cm per year at 25°C, can be obtained with a suspending liquid of a viscosity of ca 1000 cP, a density difference of 0.2 g/cm³ or less and an average particle size of the suspended solid of 0.2 μ m. This is a difficult set of criteria for most pharmaceutical suspensions to meet.

Because the distribution of particle sizes in most pharmaceutical suspension is generally above 1 μ m, deflocculated or peptized systems settle very slowly in stages, with the larger particles settling more rapidly than smaller ones. Ultimately, they form a tight, dense sediment that is difficult to resuspend. When viewed under a microscope, the dispersed suspension consists of individual particles, showing no apparent association.

Deflocculated suspensions are produced by three methods.

Mutual repulsion to large ζ -potential

This is best achieved by the adsorption of an electrolyte (KCl) or polyelectrolyte dispersant (sodium hexametaphosphate) on the surface of suspended particles to create a strong mutual repulsion between the microsize suspended particles. For example, moderate stability is

achieved when the ζ -potential is between ± 30 and ± 60 mV, and good to excellent physical stability is achieved when the ζ -potential is between ± 60 and ± 100 mV. As the size and density of the suspended particles increase beyond 1 μ m and 1.0 g/cm³, respectively, the effect of ζ -potential becomes less important.

Adsorption of a smaller hydrophilic or lyophilic colloid on larger suspended particles

When a strongly hydrated hydrophilic protective colloid, such as gelatin, is adsorbed on the surface of the suspended particles, the affinity for water exceeds the mutual attraction of adjacent particles for each other. The protective colloid and hydrogen-bonded water molecules form a protective hydration layer around each suspended particle.

Steric hindrance due to adsorption of an oriented nonionic surfactant or polyelectrolyte

Adsorption of a nonionic polymer (gum or cellulosic) or surfactant (polysorbate 80) of sufficient chain length creates steric hindrance and prevents adjacent suspended particles from coming close enough to join each other. Steric stabilization has the advantage over electrostatic stabilization in that it is relatively insensitive to the presence of electrolyte in the aqueous vehicle.

Because many pharmaceutical suspensions are not capable of reaching a state of complete electrostatic repulsion and producing and maintaining deflocculated particles, the technique has been regarded as unworkable by many investigators (1, 5).

Flocculated Suspensions

Matthews and Rhodes (8), Haines and Martin (7), Hiestand (5), and Ecanow and co-workers (9, 10) are credited with establishing the "structured particle" concept or flocculated pharmaceutical suspension. The following definitions should prove useful in avoiding confusion among three closely related terms: flocculation, agglomeration, and coagulation. The term aggregation can apply to all three.

Flocculation refers to the formation of a loose aggregation of discrete particles held together in a networklike structure by physical adsorption of macromolecules, bridging during chemical interaction (precipitation), or when the longer-range van der Waals forces of attraction exceed the shorter-range forces of repulsion. The floccule referred to as a "stable floc" usually contains varying amounts of entrapped liquid medium or vehicle within the networklike structure (Table 1).

Table 1 Suspension characteristics

Type	ζ -Potential	Relative settling rate	Sediment volume in vials at equilibrium	Drainage from vial	Resuspendability
Agglomerated or coagulated	0	Nonuniform	Low (lumpy)	Poor	Poor (may cake)
Overflocculated	0	Very high	Very high	Poor	Good
Slightly over flocculated	0	High	Moderate to high	Fair	Good
Slightly underflocculated	0	Moderate	Moderate	Good	Good
Underflocculated	0 to +10 mV	Slow	Moderate to low	Good	Fair
Deflocculated or peptized	+10 to +30 mV	Nonuniform	Low (stratified)	Good	Poor (cakes difficult to resuspend)

In agglomeration, a large number or mass of particles are closely bound together as aggregates in a dry (air) or liquid state. Coagulation or severe overfloculation refers to the massing of particles in a liquid state alone and sometimes in the form of a fluid gel structure. Aggregated particles of overflocculated systems, including adsorbed surface films, are in surface contact with each other, and each mass or coagula acts as a unit. The particles of such coagulated systems are held together by strong film-to-film bonds. Coagulated suspensions, like deflocculated suspensions, tend to “cake” on standing.

Soon after milling and suspension, unless steps are taken to prevent it, microsize particles tend to grow with time. Because the solubility rate of unprotected particles (Ostwald ripening) is higher than that of large crystals, crystal growth is favored until a more thermodynamically stable distribution of particles sizes is achieved. This phenomenon tends to promote “caking and cementing together” of particles.

The creation of a protective coat or boundary layer (with a hydrophilic colloid) about such particles offers the best protection to crystal growth. Because protective barriers may or may not flocculate, the substrate particles, the sign (positive or negative), and the charge potential on the particle surface (including a double layer) govern the choice between flocculation or dispersion (deflocculation).

The processes involved in the formation of suspensions are shown in Fig. 2. The flocculated state (C) may be reached either directly by wetting and dispersing hydrophobic particles (A) with a suitable flocculating surfactant, or indirectly by first wetting and dispersing to produce a dispersed or peptized particle (B) with a suitable surfactant and then flocculating with a suitable agent such as a hydrophilic colloid or polyelectrolyte.

In contrast to peptized or deflocculated particles, flocculated suspensions (C), which are considered pharmaceutically stable (although colloiddally unstable), can always be resuspended with gentle agitation. Severe overfloculation, on the other hand, caused by the addition of too much flocculating agent or by prolonged exposure to extreme thermal conditions (freeze–thaw cycles) tend to produce agglomerated or coagulated irreversible systems (E). The term plaque (platelike) is used to describe essentially flat agglomerates, whereas the term coagula (clumplike) is reserved for thicker, three-dimensional particle masses. In the absence of a protective colloid, the process of crystal growth is indicated by the arrow connecting (A) to (D). Some authors (11) refer to the “stable floc” as the partially flocculated state. Simply stated, the greater the number of particle-to-particle contact points in the cluster, the higher the degree of flocculation.

The main advantages of the stable floc are as follows.

The aggregates tend to break up easily under the application of moderate shear stress, such as gentle agitation of a bottle or vial, or by the flow through a small orifice (hypodermic needle and/or syringe) and reform an extended network of particles after the force is removed. Flocculation, therefore, imparts a structure to the suspension with virtually no increase in viscosity (Fig.3).

In contrast to peptized or deflocculated systems, the stable floc settles rapidly, usually to a high sediment volume, and may be easily resuspended even after standing for prolonged periods of storage time. The settling rate is not rapid and maintain content uniformity with reasonable periodic agitation. The stable floc can be produced by employing aseptic techniques, using vehicle components that are safe for intramuscular injection.

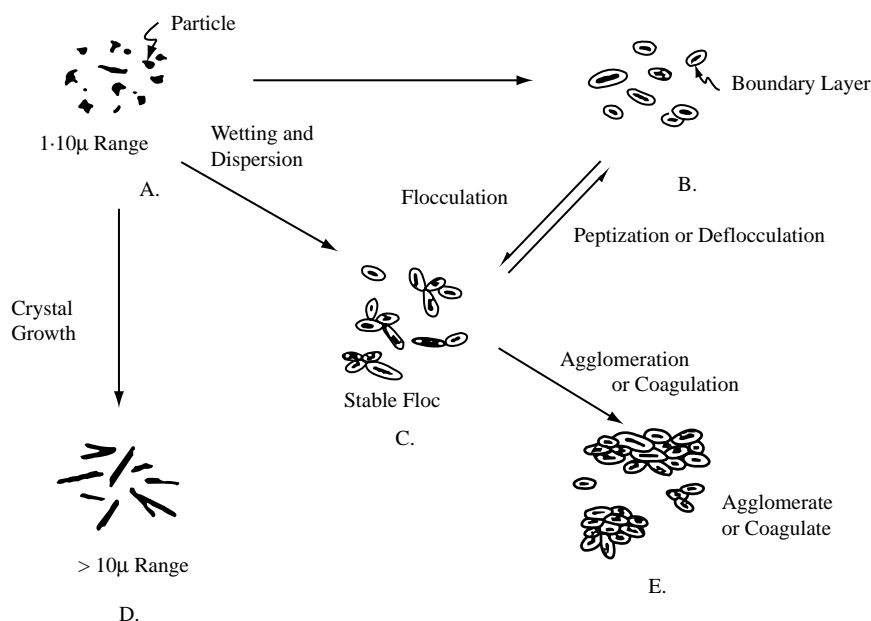


Fig. 2 Processes involved in the formation of suspensions.

Flocculated pharmaceutical suspensions are prepared using several methods. The choice depends on the properties of the drug and the class of suspension desired. The following examples illustrate how suspensions may be prepared by controlled flocculation procedures:

1. The wetting agent, polysorbate 80 (not more than 0.1–0.2% w/v of the final concentration), is dissolved in approximately half the final volume of aqueous vehicle. An anionic surfactant, such as Docusate Sodium USP, may be used as a wetting agent. This agent is, however, sensitive to pH and electrolyte concentration.

2. Ultrafine particles of the drug at the desired final concentration are uniformly and carefully spread over the surface of the vehicle, and the drug is permitted to be wetted undisturbed for as long as 16 h (overnight).
3. The wetted slurry is passed through a fine wire mesh screen (100 mesh size for ca 120 μm or larger) to remove poorly wetted powder. A single pass through a colloid mill can achieve the same result.
4. The slurry concentrate of the drug is agitated gently with an impeller-type mixer.
5. Small amounts of a 10% w/v solution of aluminum chloride hexahydrate are added dropwise, to the drug slurry from a buret or dropping pipette until the flocculation endpoint is reached (zero ζ-potential). To determine the endpoint, small samples are withdrawn and transferred to a graduated cylinder, an equal amount of vehicle is added to each, and the cylinders are gently shaken and permitted to stand undisturbed. The sample with the highest ratio of sediment to total suspension volume, exhibiting a clear supernate and good drainage characteristics is considered to be at the appropriate endpoint. Usually, no more than about 0.1–0.2% aluminum chloride hexahydrate is required to achieve the flocculation endpoint. A 10% solution of calcium chloride dihydrate may also be used as the flocculating agent. In this case, 1–2% of the calcium salt may be required for stable floc formation. If the drug fails to flocculate in the

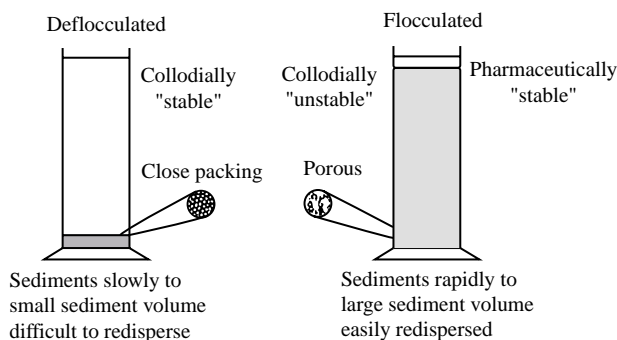


Fig. 3 Characteristics of flocculated and deflocculated suspensions. (From Ref. 5.)

presence of polyvalent aluminum or calcium ions, the water-insoluble drug particles are considered to be positively charged, and the procedure is repeated, this time using a polyvalent anionic flocculating agent such as 10% w/v sodium hexametaphosphate or 10% trisodium citrate.

6. After the flocculation endpoint has been established and verified, the other components (preservative, colorant, flavor, buffer, etc.) are added, dissolved in the liquid vehicle, and the slurry is brought to final volume with liquid vehicle.

Another popular method of preparing an oral suspension consists of suspending the drug in a solution of a hydrophilic colloid (gelatin or gum) or a diluted magma of bentonite, attapulgit, or colloidal magnesium aluminum silicate. The concentration of flocculating agent suspended in water, sorbitol, or syrup solution is between 0.1 and 1%. Overflocculation may be reversed by the careful addition of small amounts of a suitable surfactant or polyvalent deflocculating agent. Because clays cannot be used in injectable products, two other methods are described here.

One method, specially useful for preparing physically stable “noncaking” suspensions, consists of titrating concentrated aqueous solutions of soluble salt forms of acidic or basic drugs with a corresponding solution of a strong acid or a strong base. The water-insoluble free acid or base is precipitated at the pH of minimum solubility of the drug. The concentrations of the reacting solutions and the order of addition may be varied to produce an acceptable stable floc. If necessary, the electrolyte thus formed during precipitation may be reduced through slurry decantation or filtration to adjust tonicity and/or to maintain physical and chemical stability. This procedure can also be carried out under aseptic conditions.

A stable floc may also be produced by dispersing insoluble particles in a turbid or hazy vehicle consisting of finely dispersed or emulsified semipolar, liquid droplets, which cause the droplets to be adsorbed on the surface of the insoluble drug particles, resulting in a stable floc. Turbid aqueous vehicles have been prepared by the interaction of nonionic surfactants and preservatives. The concentration of surfactant and preservative required for haze formation may be reduced by the addition of small amounts of sorbitol to the vehicle.

Structured Vehicle

Another technique for the preparation of a stable suspension is based on the concept of the “structured

vehicle,” in which the viscosity of the preparation, under static conditions of very low shear, on storage approaches infinity. The vehicle is said to behave like a “false body” that is able to maintain the suspended particles in a state of more or less permanent suspension.

Structured vehicles are usually not considered for the preparation of parenteral suspensions because, owing to their high viscosity, such systems lack sufficient syringe ability for ease of use.

Bingham-Type Plastic Flow

Vehicles that exhibit the unusual property of Bingham-type plastic rheological flow are characterized by the need to overcome a finite yield stress before flow is initiated. Permanent suspension of most pharmaceutical systems requires yield-stress values of at least 2–5 Pa (20–50 dyn/cm²). Bingham plastic flow is rarely produced by pharmaceutical gums and hydrophilic colloids. *National Formulary* (NF) carbomers exhibit a sufficiently high yield value at low solution concentration and low viscosity to produce permanent suspensions. The carbomers, however, require a pH value between 6 and 8 for maximum suspension performance. The polymer is essentially incompatible with cationic resins, certain polyvalent ions, and high concentrations of electrolytes.

Thixotropic Flow

Thixotropy is a rheological property that results in yield stress on standing. Thixotropic flow is defined as a reversible, time-dependent, isothermal gel–sol transition. Thixotropic systems exhibit easy flow at relatively high shear rates. However, when the shear stress is removed, the system is slowly reformed into a structured vehicle. The usual property of thixotropy results from the breakdown and buildup of floccules under stress. A small amount of particle settling takes place until the system develops a sufficiently high yield value. The primary advantage of thixotropic flow is that it confers pourability under shear stress and viscosity and sufficiently high yield stress when the shear stress is removed at rest.

The concentrations of Bingham-type and thixotropic suspending agents required to achieve a yield stress of 10 Pa (100 dyn/cm²) is shown in Table 2.

Suspending agents such as a pseudoplastic (sodium carboxymethylcellulose) in combination with a clay (hydrated colloidal magnesium aluminum silicate)

or blends and coprecipitates of sodium carboxymethylcellulose and microcrystalline cellulose exhibit some thixotropic flow characteristics. Other pseudoplastics such as hydroxyethylcellulose or hydroxypropyl methyl cellulose may be required to overcome possible incompatibilities with sodium carboxymethylcellulose.

Emulsion Base

An emulsion base or a waxy-type self-emulsifier to develop structure or “false body” in suspension systems is widely used in both pharmaceutical and cosmetic systems. A dilute emulsion system is not often considered for suspension purposes because of the potential complexities involved in mixing emulsion and suspension systems. The drug particles are dispersed in the primary emulsion component prior to dilution with other vehicle components. Emulsifiers that exhibit Bingham plastic or thixotropic flow characteristics should have acceptable formulating properties (taste, stability, etc.).

FORMULATION OF SUSPENSIONS

During the preparation of physically stable pharmaceutical suspensions, a number of formulation components are used to keep the solid particles in a state of suspension (suspending agents), whereas other components are part of the liquid vehicle itself and have other functions in the dosage form.

- 1. *Components of the suspending system*
 - Wetting agents
 - Dispersants or deflocculating agents

Table 2 Concentration of suspending agents in water at 25°C required to achieve a yield value of 10 Pa

Agent	Concentration (%)
Carbomer 941	0.1
Carbomer934	0.2
Carrageenan	0.5
Carboxymethylcellulose	2.0
Xanthan gum	2.0
Algin	3.5
Magnesium aluminum silicate	5.0
Hydroxyethylcellulose	5.0
Guar gum	5.0
Tragacanth gum	5.0

(From B.F. Goodrich Specialty Chemicals, Cleveland, Ohio.)

Flocculating agents
Thickeners

- 2. *Components of the suspending vehicle or external phase*
 - pH control agents and buffers
 - Osmotic agents
 - Coloring agents, flavors, and fragrances
 - Preservatives to control microbial growth

Wetting Agents

Wetting agents are surfactants that lower the interfacial tension and contact angle between solid particles and the liquid vehicle. When the insoluble powder is added to a liquid vehicle containing a wetting agent, penetration of the liquid phase into the powder will be sufficiently rapid to permit air to escape from the particles. The resulting wetted particles either sink en masse or separate with low-shear agitation. The best range for wetting and spreading by nonionic surfactants is between a hydrophile–lipophile balance (HLB) value of 7 and 10, although surfactants with values higher than 10 are often used for this purpose. Common wetting agents and surfactants include: 1) anionic type (docusate sodium and sodium lauryl sulfate); and 2) nonionic type (polyoxyalkyl ethers, polyoxylakyl phenyl ethers, polyoxy hydrogenated castor oil, polyoxy sorbitan esters, and sorbitan esters).

Deflocculants and Dispersing Agents

Unlike surfactants, these agents do not appreciably lower surface and interfacial tension; thus, they have little tendency to create foam or wet particles. Most deflocculants, however, are not generally considered safe for internal use, and as a result the only acceptable dispersant for internal products is lecithin or a lecithin derivative (naturally occurring mixture of phosphatides and phospholipids). Because lecithins vary in water solubility and dispersibility characteristics, proper control of product specifications must be maintained to obtain reproducibility.

Flocculating Agents

Primary flocculating agents are simple neutral electrolytes in solution that are capable of reducing the ζ -potential of suspended charged particles to zero. Small concentrations (0.01–1%) of neutral electrolytes, such as sodium or potassium chloride, are often sufficient

to induce flocculation of weakly charged, water-insoluble, organic nonelectrolytes. In the case of highly charged, insoluble polymers and polyelectrolyte species, similar concentrations (0.01–1%) of water-soluble divalent or trivalent ions, such as calcium salts, alums, sulfates, citrates, and phosphates, may be required for floc formation, depending on particle charge (positive or negative). These salts are often used together as pH buffers and flocculating agents.

Thickeners, Protective Colloids, and Suspending Agents

Protective or hydrophilic colloids, such as gelatin, natural gums, and cellulosic derivatives, that are adsorbed on insoluble particles, increase the strength of the hydration layer formed around suspended particles through hydrogen bonding and molecular interaction. Because these agents do not reduce surface and interfacial tension, they function best in the presence of a wetting agent. Many of these agents are protective colloids in low concentration (<0.1%) and viscosity builders in higher concentrations (>0.1%). Suspending agents commonly used in pharmaceutical suspensions include:

- *Cellulosics*: sodium carboxymethylcellulose, microcrystalline cellulose (including coprecipitates and blends of the two), hydroxyethylcellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, methylcellulose, starch, sodium starch glycolate, and powdered cellulose
- *Clays*: attapulgite, bentonite, magnesium aluminum silicate, kaolin, silicon dioxides
- *Gums*: acacia, agar, algin, carrageenan, guar, pectin, tragacanth, xanthan
- *Polymers*: carbomers, polyvinyl alcohol, povidone, polyethylene oxide
- *Sugars*: dextrin, maltitol, sucrose
- *Others*: aluminum monostearate, emulsifying waxes, gelatin

The other agents used in pharmaceutical suspensions (pH control agents, buffers, osmotic agents, stabilizers, vehicles, colorants, flavors, fragrances, and preservatives to control microbial growth) are not discussed in this article.

STERILE SUSPENSIONS

Sterile suspensions (injectable and ophthalmic) have characteristics that are not commonly shared by other

suspension systems, such as ease of resuspension, drainage, absence of foreign particulate matter and pyrogens, and syringeability in the case of injectable suspensions.

The preparation of a sterile suspension is a difficult manufacturing procedure. It requires strict attention to detail during the final recrystallization of the active drug substance, size reduction, and sterilization of the active drug substance and the suspending vehicle, aseptic wetting of the sterile drug powder with a portion of the sterile vehicle, aseptic dispersion and milling of the bulk sterile suspension, and aseptic filling of the finished suspension into sterile containers.

Various procedures for the manufacture of sterile suspensions have been reported by Akers et al. (12), Grimes (13), and Portnoff (14). At present, there is no pharmaceutically acceptable chemical agent that can be added to the finished suspension to make it both sterile and safe. Therefore, an elaborate program of sterility checks at critical phases of the operation is required.

An important property of a good parenteral suspension is syringeability; the ability of a parenteral solution or suspension to pass easily through a hypodermic needle, especially during the transfer of a product from vial to hypodermic syringe prior to injection. Increases in vehicle viscosity, vehicle density, and size and concentration of suspended particles make the transfer more difficult.

The most important of these features is probably viscosity. Fortunately, in the case of parenteral suspensions, viscosity is the easiest parameter to control. The preparation of a stable floc contributes little to the overall viscosity of the system (the so-called Einstein relationship) and does not adversely influence syringeability. Although the individual particles are held loosely together in large multiple aggregates, they are easily broken up and reformed during the passage from vial to syringe and syringe to injection site.

Drainage, or the ability of the suspension to break cleanly away from the inner walls of the primary container-closure system, is another important characteristics of a well-formulated parenteral suspension. Deflocculated to flocculated systems show this property, whereas overflocculated systems show some degree of poor drainage, also called "buttermilk appearance," a term that aptly describes this unsightly condition. Silicone coated containers, vials and plugs with dimethicone improve drainage and help reverse the tendency toward poor drainage by slightly overflocculated systems.

Resuspendability, or the ability to distribute settled particles with a minimum of shaking, is an important

characteristic of parenteral suspensions. Stable, flocculated parenteral suspensions that have been undisturbed for long periods of storage time are easily resuspended.

COSMETIC SUSPENSIONS

Cosmetic suspensions are available in two types. The first comprises pigmented products that are suspended in essentially aqueous vehicles (liquid makeup, eye-liners, mascara, and blusher). These products have a high solids content, high density, impalpable powders, and pigments permanently suspended in a primary oil-in-water emulsion-type base or a complex system of hydrophilic cellulose derivatives, clays, and/or polymeric film formers, in which the gelling and suspending properties of the vehicle often are reinforced by a small amount of a Bingham-type plastic such as carbomer.

The second type comprises pigment-containing nail enamels. The coloring tints, pigments, pearls, and lakes are suspended with the aid of an organophilic, thixotropic gellant, such as stearalkonium hectorite in a nonaqueous vehicle, containing butyl acetate, ethyl acetate, and isopropyl alcohol solvents in which the primary plasticized nitrocellulose and toluene sulfonamide—formaldehyde resin film formers are dissolved. Nail enamel is an excellent example of a permanent suspension in a nonaqueous vehicle.

TEST METHODS FOR PHARMACEUTICAL SUSPENSIONS

Tingstad (15) reviewed test methods for determining the physical stability of pharmaceutical suspensions. The procedures outlined are designed to determine the state of flocculation of a formulation. Because there is more than one method of preparing stable suspensions, the following test methods and performance criteria were found useful for determining the stability of both flocculated and dispersed systems.

Appearance

The appearance in a graduated glass cylinder or transparent glass container is noted. The following questions were addressed: At equilibrium, is the color and appearance of the sediment uniform? Are there breaks or air pockets in the sediment? Is the residual drainage

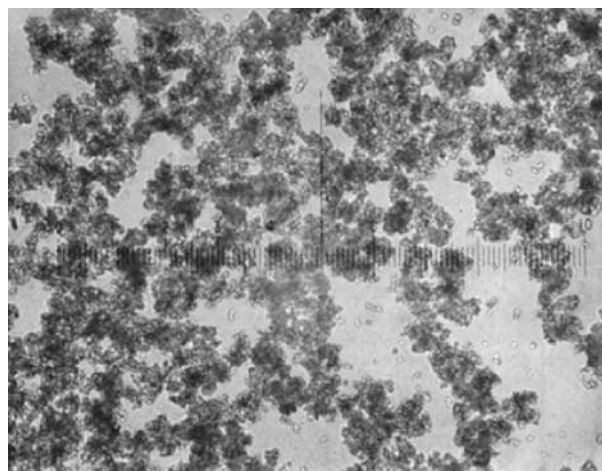


Fig. 4 Photomicrograph of a flocculated steroid suspension.

above the sediment uniform and minimal, or is there coagulated material adhering to the inside walls of the container?

Photomicroscopic Examination

The microscope can be used to estimate and detect changes in particle size distribution and crystal shape. Its usefulness can be enhanced by attaching a Polaroid-type camera to the microscope to permit rapid processing of photomicrographs (Fig. 4). These can be used, for example, to distinguish between flocculated and nonflocculated particles and to determine changes in the physical properties and stability. Sufficient fields and samples should be examined to make these determinations. Dilutions for microscopic examination should be made with supernatant external phase rather than with purified water. Individual particle size distributions can be accurately determined, using suitable electron instrumentation, for example, a Coulter Multisizer II (an electrical sensing zone instrument from Coulter Scientific Instruments, Hialeah, FL) or the Elzone 280 PC systems (Particle Data, Elmhurst, IL). General methods for particle size analysis are given in Fig. 5.

Color, Odor, and Taste

These characteristics are especially important in orally administered suspensions. Variation in color often indicates poor distribution and/or differences in particle size. Variations in taste, especially of active constituents, can often be attributed to changes in particle size, crystal

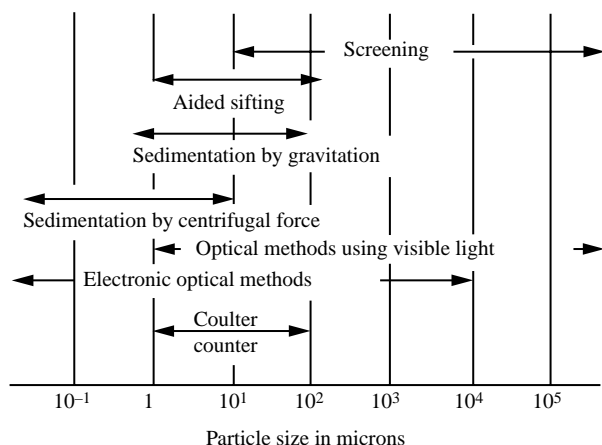


Fig. 5 General methods for particle size analysis.

habit, and subsequent particle dissolution. Changes in color, odor, and taste can also indicate chemical instability.

Sedimentation Rate, Sediment Volume, and Resuspendability

Simple, inexpensive, graduated cylinders (100–1000 mL) are useful for determining the physical stability of suspensions. They can be used to determine the settling rates of flocculated and nonflocculated suspensions and the sediment height at equilibrium. The falling height of the liquid–sediment interface of the suspension is determined as a function of time, and the sedimentation rate test is repeated periodically during storage. The sediment volume at equilibrium should be sufficiently large to support uniform resuspension with gentle agitation. The equilibrium sediment volume should be similar and reproducible batch after batch.

Volumetric graduated cylinders are used to determine the “*F*” or flocculation ratio, a value that represents the ratio of the sediment volume to the original suspension volume at a given time. It is used to measure the relative degree of flocculation and physical stability of suspensions. A sufficiently wide graduated cylinder should be used to overcome the “wall effect” which often influences the settling rate and equilibrium sediment volume of flocculated suspensions. Small graduated cylinders have a tendency to support suspensions because of the adhesive forces acting between the inner surface of the container and the suspended particles.

Viscosity

A Brookfield viscometer with a helipath attachment (Stoughton, MA) is a useful rheological instrument for measuring the settling behavior and structure of pharmaceutical suspensions and for characterizing the properties and stability of flocculated suspensions. The viscometer should be properly calibrated to measure the apparent viscosity of the suspension at equilibrium at a given temperature to establish suspension reproducibility. Apparent viscosity, like pH, is an exponential term, and therefore the log-apparent viscosity is an appropriate way of reporting the results.

Density

Specific gravity or density of the suspension is an important parameter. A decrease in density often indicates the presence of entrapped air within the structure of the suspension. Density measurements at a given temperature should be made using well-mixed, uniform suspensions; precision hydrometers facilitate such measurements.

pH Value

The pH value of aqueous suspensions should be taken at a given temperature and only after settling equilibrium has been reached, to minimize “pH drift” and electrode surface coating with suspended particles. Electrolyte should not be added to the external phase of the suspension to stabilize the pH, because neutral electrolytes disturb the physical stability of the suspension.

Freeze–Thaw Cycling

This is a useful guide to the physical stability of suspensions. If freeze–thaw cycling or elevated temperature exposures are chosen for physical stability testing, companion samples of a closely related marketed suspension should be included in the testing protocol for comparison purpose because pharmaceutical suspensions are not normally designed to withstand temperature extremes during storage (15–30°C optimum).

Drug Content Uniformity

This important testing procedure is best performed using either “unit of use” volume (e.g., 5 mL of oral liquid or a spray actuation of an oral inhalation product) or sampling from a well-mixed dispensing container from the top, middle, and bottom of the suspension.

Dissolution Testing

This technique (16) is still evolving. The favored approach at present is to submerge a small, known amount of suspension inside a secured Durapore (polyvinylidene fluoride) membrane pouch (Millipore Products, Bedford, MA) of suitable porosity in "teabag" fashion in a suitable dissolution medium using the USP Method 1 Paddle Apparatus (17). Optimization of experimental conditions (rate of agitation, volume and type of medium, temperature, etc.) must be established to achieve reproducible results.

Particle Size Measurement

Recently (18–20), with respect to the importance of particle size distribution in terms of particle characterization and product physical stability testing, there has been interest in newer light-scattering methods for particle detection called photon correlation spectroscopy (PCS). PCS methods can be applied to both micro- and nanosuspensions.

The information thus obtained from the use of such equipment includes mean particle size, particle size distribution, particle concentration, molecular weight estimation, polydispersity, particle shape, hydrodynamic interactions, and aggregation mechanisms.

In addition (21), there are several experimental options available for particle size measurement alone. They include single particle optical sensing (SPOS), laser diffraction (LD), and ultrasound attenuation (UA).

Other Procedures

Assays for potency, preservative effectiveness, compatibility with primary container-closure system, off-torque, simulated use testing, etc., should be handled in a manner similar to that used for conventional liquid solutions, with the provision that the container is well-mixed prior to sampling.

NANOSUSPENSIONS

Pharmaceutical nanosuspensions (22) are usually very finely dispersed solid drug particles in an aqueous vehicle for either oral and topical use or for parenteral and pulmonary administration. The key difference from conventional suspensions is that the particle size distribution of the solid particles in nanosuspensions is

usually less than 1 μm , with an average particle size range between 200 and 600 nm.

The techniques used for preparing nanoparticles are similar to those used to prepare more conventional drug particles and include controlled precipitation, ball milling using glass or zirconium oxide pearls, and high-pressure homogenization.

Nanosuspensions can be sterilized for parenteral use by using conventional steam sterilization in an autoclave, γ -irradiation, or membrane microfiltration in certain situations.

The key to long-term physical stability of aqueous nanosuspensions is the selection of a suitable water-soluble surfactant or polymer as an external particle stabilizer to prevent particle growth. Several potential stabilizers are lecithin, phospholipids, poloxamers, and polysorbates.

The physical stability of nanosuspensions may be monitored with the use of electron microscopic analysis.

The major advantage of pharmaceutical nanosuspensions is their ability to increase the in vivo absorption of highly-water-insoluble drugs by dramatically reduced particle size.

REFERENCES

1. Martin, A.; Bustamante, P. *Physical Pharmacy*, 4th Ed.; Lea & Febiger: Philadelphia, 1993; 477–484.
2. Marcus, A.D.; DeKay, H.G. New Developments in Calamine Lotion. *J. Am. Pharm. Assoc., Pract. Ed.* **1950**, *11*, 227–229.
3. Boylan, J.C. The Development of Semi-Solid Dosage Forms: An Overview. *Drug Dev. Commun.* **1976**, *2*, 320.
4. Robinson, M.J. Third Annual National Industrial Pharmaceutical Research Conference, Land O'Lakes: WI, 1961.
5. Hiestand, E.N. Theory of Coarse Suspension Formulation. *J. Pharm. Sci.* **1964**, *53*, 1–18.
6. Carpenter, C.R. Calculate Settling Velocities for Unrestricted Particles on Hindered Settling. *Chem. Engr.* **Nov.** **1983**, *4*.
7. Haines, B.A.; Martin, A.N. Interfacial Properties of Powdered Material. *J. Pharm. Sci.* **1961**, *50*, 228–232, 753–759.
8. Matthews, B.A.; Rhodes, C.T. Use of DLVO Theory to Interpret Pharmaceutical Suspension Stability. *J. Pharm. Sci.* **1970**, *59*, 521–525.
9. Ecanow, B.; Gold, B.; Ecanow, C. Newer Aspects of Suspension Theory. *Am. Perfum. Cosmet.* **Nov.** **1969**, *84*, 27–31.
10. Ecanow, B.; Webster, J.; Blake, M.I. Conductivity Studies of Suspension Systems in Different States of Aggregation. *J. Pharm. Sci.* **1982**, *71*, 456–457.
11. Michaels, A.S.; Bolger, J.C. The Plastic Flow Behavior of Flocculated Kaolin Suspension. *Ind. Eng. Chem. Fundam.* **1962**, *1*, 153–162.

12. Akers, M.J.; Fites, A.L.; Rabinson, R.L. Formulation Design and Development of Parenteral Suspensions. *J. Parenter. Sci. Technol.* **1987**, *41*, 88–96.
13. Grimes, T.L. Scaleup and Manufacture of Sterile Suspensions, APhA 133rd Annual Meeting, San Francisco, March 19, 1986.
14. Portnoff, J.B. The Development of Sterile Suspensions—Case Study, APhA 133rd Annual Meeting, San Francisco, March 19, 1986.
15. Tingstad, J.E. Physical Stability Testing of Pharmaceuticals. *J. Pharm. Sci.* **1964**, *53*, 955–962.
16. Stout, P.J.; Howard, S.A.; Mauger, J.W. Dissolution of Pharmaceutical Suspensions. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1991; 4, 169–192.
17. United States Pharmacopeia 24, The Pharmacopeial Convention, Rockville, MD 2000, 4.
18. Castanho, M. *Light Scattering and Photon Correlation Spectroscopy*, NATO ASI Series, High Technology, 1997; 40, 31–36.
19. Constantinides, P.P.; Yiv, S.H. *Int. J. Pharm.* **1995**, *115*, 225.
20. Ruth, H. *Saint Int. J. Pharm.* **1994**, *116*, 253.
21. Nicoli, D. *Am. Lab.* **Jan.** **2000**.
22. Miiller, R.H.; Jacobs, C.; Kayser, O. Nanosuspensions. *Pharmaceutical Emulsions and Suspensions*; Nielloud, F., Marti-Mestres, G., Eds.; Marcel Dekker, Inc.: New York, 2000.
23. Banker, G.S.; Rhodes, C.T. *Modern Pharmaceutics*, 2nd Ed.; Rev. Exp., Marcel Dekker, Inc. New York, 1989, 339–353.
24. *Handbook of Pharmaceutical Excipients*, 3rd Ed., APhA and Pharmaceutical Press: London, 2000.
25. Martin, A.; Bustamante, P. *Physical Pharmacy*, 4th Ed.; Lea & Febiger: Philadelphia, 1993; 477–484.
26. Nash, R.A. Pharmaceutical Suspensions. *Pharmaceutical Dosage Forms, Dispersed Systems*; Lieberman, H.L., Rieger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1988; 1, 151–198.
27. Zografi, G.; Swarbrick, J.; Schott, H. Disperse Systems. *Remington's Pharmaceutical Sciences*; Gennaro, A.R., Ed.; Mack Publishing Co.: Easton PA, 1990; 257–309.

TABLET COMPRESSION: MACHINE THEORY, DESIGN, AND PROCESS TROUBLESHOOTING

Michael J. Bogda

Barr Laboratories, Inc., Pomona, New York, U.S.A.

INTRODUCTION

The most common method of drug delivery is the oral solid dosage form, of which tablets and capsules are predominant. The tablet is more widely accepted and used compared to capsules for a number of reasons, such as cost, tamper resistance, ease of handling and packaging, ease of identification, and manufacturing efficiency. Over the past several years, the issue of tamper resistance has resulted in the conversion of most over-the-counter drugs from capsules to predominantly all tablets.

Pharmaceutical products have been manufactured into compressed tablets for many years. During the 1950s, much research was devoted to the physics of compression (1, 2). Since that time, the pharmaceutical industry has attained a much greater understanding of the compression process, which resulted in the development of more robust pharmaceutical formulations (3–5). This has been achieved by the use of instrumented tablet presses and sophisticated data collection systems combined with the development of mathematical models.

During this time, a significant portion of the development work has been conducted on older equipment, which has been retrofitted to measure compression and ejection-force signals. Recent advances in the design of tablet compression equipment has resulted in higher-efficiency machines designed to optimize compression efficiency, minimize tablet weight variation, and provide greater flexibility, allowing the production of a greater range of products. However, the modern sophisticated machines still employ the same general concepts of operation: die fill, tablet compression, tablet ejection, and tablet scrape-off. Therefore, studies conducted on older equipment designed to evaluate the compression characteristics of materials, can offer significant insight into material behavior. However, modern machines provide greater accuracy and efficiency as follows:

- Improved material feed systems.
- Improved cam design and material of construction.
- Multistage compression.
- Isolated design for quick cleaning and changeover.

- Improved force-measurement techniques.
- Introduction of electronics to provide force control.
- Integration of on-line weight, thickness, and hardness test units providing weight feedback control to the force control unit, and
- High-speed single-tablet sorting to reject out-of-specification tablets.

Therefore, optimal product development can typically be performed on these machines that offer improved compression designs and material feed systems.

This article provides the basic information necessary to understand the general process of tablet formation. General machine design characteristics and tablet press nomenclature are presented. Tablet press control systems and process automation are discussed, followed by process and product troubleshooting on tablet compression equipment.

THE PROCESS OF TABLET FORMATION

The quality of a compressed tablet is determined by material fill characteristics and compression behavior. During compression, the rate at which tablets can be produced can be limited due to nonuniform material fill characteristics. Pending successful and reproducible material fill (die fill), the powder mass must form a coherent compact that stays intact upon ejection out of the die. Therefore, tablet press performance can be limited due to poor fill characteristics and/or poor compression behavior.

Die fill characteristics depend upon material flow properties that are primarily affected by particle size and shape. Additionally, high interparticle friction can have a detrimental effect on die fill characteristics due to powder bridging and nonuniform flow characteristics. A nonuniform particle size distribution may also lead to material segregation resulting in uniformity problems. Tablet presses employ volumetric filling of the material into the die cavity. Most high-speed tablet presses are equipped with force feeders, which use rotating paddles to promote

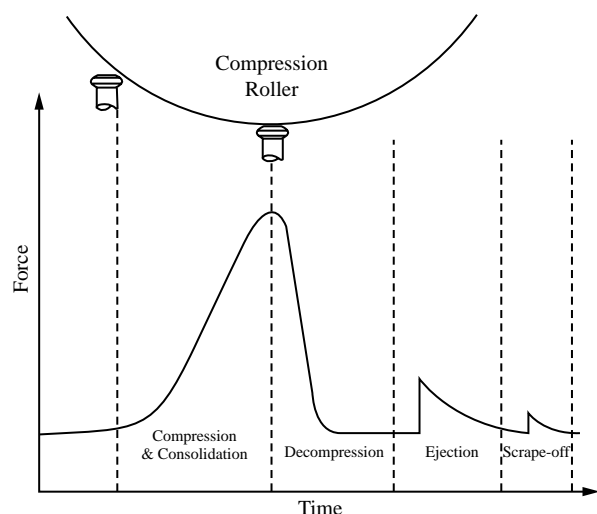


Fig. 1 The compression process.

uniform die fill characteristics. For certain materials, attention must be paid to ensure that the action of the power feeder does not cause overmixing or segregation of the material.

Upon proper die fill, one must consider whether the material will form a tablet. The tableting characteristics of powders depend on the viscoelastic properties of the material. The process of compaction has been defined as “the compression and consolidation of a two-phase system due to an applied load.” (6). The quality of the compact depends on the compression and consolidation of the powder mass, decompression of the compact, ejection from the die, and subsequent scrape-off from the lower punch. A schematic representation of the compression process is shown in Fig. 1. Because these viscoelastic properties are time dependent, both the magnitude and the rate of application (and release) of the compression force affect tablet quality.

Compression and Consolidation

During compression, the bulk volume of the material is reduced, resulting in the displacement of the gaseous phase (air) (6). Further increasing the force leads to particle deformation and rearrangement. At this point, the three principal modes of deformation are as follows:

1. *Elastic deformation*: A spontaneously reversible deformation of the compact in which, upon removal of the load, the powder mass reverts back to its original form. Most materials undergo elastic deformation to some extent. Compression of rubber would be by elastic deformation.

2. *Plastic deformation*: After exceeding the elastic limit of the material (yield point), the deformation may become plastic, that is, the particles undergo viscous flow. This is the predominant mechanism when the shear strength between the particles is less than the tensile or breaking strength. Plastic deformation is a time-dependent process.
3. *Brittle fracture*: Upon exceeding the elastic limit of the material (yield point), the particles undergo brittle fracture if the shear strength between the particles is greater than the tensile or breaking strength. Under these conditions, the larger particles are sheared and broken into smaller particles.

The compression process includes these three mechanisms. The individual characteristics of the material under investigation determine the extent to which each is active. Since some of these deformation characteristics are time-dependent, machine characteristics can have a major effect on tableting performance. These characteristics determine the rate of force application, dwell time (i.e., the time of maximum compression force, which depends on the punch head flat diameter and the tangential velocity), and the rate of decompression (see Fig. 1).

Typically for materials that undergo plastic deformation, as machine speed is increased there is less time for stress relaxation. Under these conditions, the tablets may cap and laminate. However, capping and lamination can be eliminated or minimized by slowing down the compression process (reducing the machine speed), lowering the rate of force application (larger compression roller diameter), or increasing the time of compression (multi-stage compression).

The final tablet properties are also affected by the consolidation (i.e., bonding) mechanisms of the powder which is influenced by its chemical nature, the surface area of the contact points, contamination (including film coatings such as magnesium stearate), and interparticle distance. The predominant consolidation mechanisms are listed below (7–10):

- *Mechanical theory*: As the particles undergo deformation, the particle boundaries intertwine to form mechanical bonds.
- *Intermolecular forces theory*: van der Waals forces bond the molecules together at the newly sheared surfaces of the particle boundaries. Microcrystalline cellulose is believed to undergo significant hydrogen bonding during tablet compression (11).
- *Liquid-surface film theory*: Thin liquid films form which bond the particles together at the particle surface. The energy of compression produces melting or solution at the particle interface followed by subsequent

solidification or crystallization thus resulting in the formation of bonded surfaces. Due to the applied pressure, the particles may melt or dissolve. As the pressure is released, solidification and crystallization occur.

The intermolecular forces theory and the liquid-surface film theory are believed to be the major bonding mechanisms in tablet compression. Many pharmaceutical formulations require a certain level of residual moisture to produce high quality tablets. The role of moisture in the tableting process is supported by the liquid-surface film theory.

During tablet formation, as load is applied to the powder mass, plastic deformation and brittle fracture create clean surfaces that are brought in intimate contact by the applied load. These surfaces bond together by a number of mechanisms such as those listed earlier. Plastic deformation is believed to create the greatest number of clean surfaces. Because it is time-dependent, higher rates of force application should lead to the formation of less new clean surfaces, resulting in weaker tablets. Additionally, because tablet formation is dependent upon the formation of clean new surfaces, high concentrations or overmixing of materials that form weak bonds result in weak tablets. Magnesium stearate, for example, forms weak bonds and easily wets surfaces. Therefore, over-lubrication or overmixing of magnesium stearate may lead to weak tablets.

Fragmentation (the creation of new clean surfaces) continues at the same time at which bonding and densification occur. A high quality tablet can be formed only when the process of bonding and densification exceeds fragmentation. The rates at which these functions occur are dependent upon the rate at which the forces are applied.

During compression, the powder compact typically undergoes a temperature increase (12, 13), which depends on frictional effects that are dependent, in turn, on the specific material characteristics, the lubrication efficiency, the magnitude and rate of application of the compression forces, and the machine speed. Typical temperature increases are between 4 and 30°C (14). As the tablet temperature rises, stress relaxation and plasticity increase, while the elasticity decreases and strong tablets are formed (15). Therefore, compression of material, at elevated temperature with increased ductility should result in stronger tablets.

It is believed that under certain conditions precompression is beneficial because it helps to remove entrapped air (16) and extends dwell time, thereby increasing the degree of stress relaxation and plastic deformation (17) as well as

the number of bonds, thus increasing tablet strength. Additionally, by extending the time of compression, precompression may provide a gradual loading and unloading of force. However, recent studies (18) suggest that a high level of precompression (higher than that of the main compression) may improve the tableting characteristics more than conventional tableting where main compression exceeds precompression. It is theorized that this effect is due to the high initial compression force that raises the tablet temperature, thereby increasing ductility and allowing greater plastic deformation. Application of the lower second compression force increases the formation of particle-particle bonds while minimizing particle-bond rupture, resulting in stronger tablets.

Decompression

After the compression and consolidation of the powder in the die, the formed compact must be capable of withstanding the stresses encountered during decompression and tablet ejection (19). The rate at which the force is removed (dependent on the compression roller diameter and the machine speed) can have a significant effect on tablet quality. The same deformation characteristics that come into play during compression play a role during decompression.

After application of the maximum compression force, the tablet undergoes elastic recovery. While the tablet is constrained in the die, elastic recovery occurs only in the axial direction. If the rate and degree of elastic recovery are high, the tablet may cap or laminate in the die due to rapid expansion in the radial direction only. If the tablet undergoes brittle fracture during decompression, the compact may form failure plains due to the fracturing of the surfaces. Tablets that do not cap or laminate are able to relieve the developed stresses by plastic deformation. Since plastic deformation is time-dependent, stress relaxation is also time-dependent. Therefore, tablet fracture is affected by rates of decompression (machine speed) since the compact may not have sufficient time to relieve the internal stresses created during decompression. Formulations which contain significant concentrations of microcrystalline cellulose typically form good compacts due to its plastic deformation properties. However, if the machine speed and the rate of tablet compression are significantly increased, these formulations exhibit capping and lamination tendencies.

The rate of decompression can also have an effect on the ability of the compacts to consolidate (form bonds). Based on the liquid-surface film theory, the rate of crystallization or solidification should have an effect on the strength of the

bonded surfaces. The rate of crystallization is affected by the pressure (and the rate at which the pressure is removed). High decompression rates should result in high rates of crystallization. Typically, slower crystallization rates result in stronger crystals. Therefore, if bonding occurs by these mechanisms, lower machine speeds (lower rates of decompression and crystallization) should result in stronger tablets.

Ejection

After decompression, the tablet remains in the die until it is ejected. During this time, a residual die wall force is exerted by the tablet on the die wall. Tablet ejection is defined by three stages (6):

1. The initial ejection peak force required to break the tablet adhesion to the die wall. This force is the highest force encountered during ejection and occurs over a very short period of time.
2. The force required to push the tablet up the die wall, which is typically lower than the ejection peak force. However, inadequate lubrication or damaged dies may result in slip-stick behavior where the tablet continues to adhere and break adhesions to the die wall surface. These conditions typically result in tablet failure.
3. Declining forces as the tablet emerges from the die.

Inadequate lubrication typically results in high ejection forces and possibly tablet failure. Ejection forces below 200 N are optimal, although forces up to 400 N are common. Ejection forces ranging between 400 and 800 N are considered high. Forces exceeding 800 N result in excessive heat build-up and could lead to machine damage and product failure. Inadequate lubrication can also result in striations along the tablet side wall, and picking and sticking.

Tablets that undergo significant elastic recovery upon decompression may exhibit capping upon ejection out of the die. Under these circumstances, the tablet builds up significant stress while in the die, which can only be relieved in the axial direction. As the tablet emerges from the die, its top portion can expand in both the radial and axial directions while the portion remaining in the die is confined. Shear stresses develop along the edge of the die and result in tablet failure.

Scrape-Off

Tablet scrape-off occurs immediately after ejection. Figure 2 illustrates a tablet stripper on a rotary tablet press. Typical forces during tablet scrape-off are 2 N or less. Scrape-off forces of 6 N or higher result in tablets

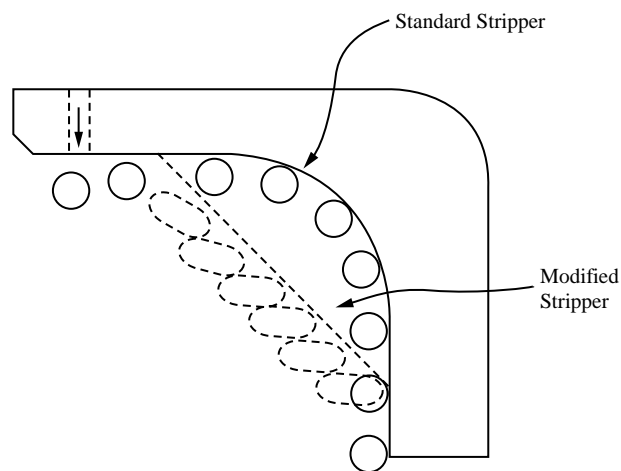


Fig. 2 Tablet stripper.

sticking to the lower punches and subsequently picking or, under extreme circumstances, shearing the bottom of the tablet. Frequently, shearing of the lower portions of the tablet due to scrape-off problems is mistaken for capping. However, this can be easily distinguished by examining the lower punches and rotating the press manually.

At high machine speeds, scrape-off problems may be encountered due to tablets backing up at the point of discharge. Under many circumstances, specially modified tablet strippers have proven to be beneficial for shaped tablets by discharging the tablets off the die table as quickly as possible (see Fig. 2).

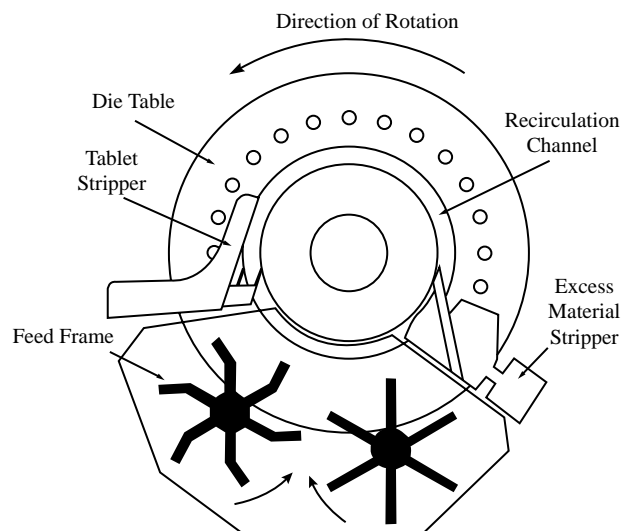


Fig. 3 Rotary tablet press (top view).

ROTARY TABLET PRESS DESIGN

Pharmaceutical tablets are generally produced on rotary tablet presses (Fig. 3), where upper and lower punches reside in the upper and lower turret, respectively. The dies are inserted in the die table and secured by die lock screws. The upper and lower turret and the die table are precisely aligned. The movement of the punches is controlled by cam tracks and compression rollers. As the entire assembly rotates, the upper and lower punches move along the cam tracks to accomplish die fill, tablet compression, ejection, and scrape-off.

Tablet Compression

Tablet compression can be separated into the two distinct yet equally important phases of die fill-weight adjustment and tablet formation as shown in Fig. 4. As die fill begins, the lower punch face is initially flush with the die table surface as the lower punch enters the overfill cam at the entry of the feeder. The lower punch travels under the feeder and is pulled down by the overfill cam. At this point the lower punch has passed through approximately 50% of the feeder and the die cavity contains more material than required.

After overfilling the die cavity, the lower punch is adjusted to a constant height as it passes into the weight-regulation unit. The constant height, known as the fill depth, is measured as the distance between the lower punch face and the die table surface. Since die fill is volumetric, the constant height of the lower punch in the weight-regulation unit provides a constant volume of material. Therefore, the fill depth is affected by the density of the granulation. Variation of granulation density between batches results in different fill depths, whereas variable granulation density within a batch results in fluctuating fill depth requirements.

As the lower punch passes from the fill cam to the weight-regulation cam, the excess material is pushed back into the feeder and scraped off at the top of the die table by the excess material stripper and directed into a recirculation channel. On many modern presses, the lower punch is lowered by approximately 2–4 mm relative to the top of the die table after the excess material stripper. This lowers the material away from the top of the die table, minimizing uncontrolled loss as the upper punch enters into the die cavity after scrape-off. Under these circumstances the upper punch does not contact the top of the material until it enters into the die, minimizing material loss and weight variation. Additionally, lowering the slug of material away

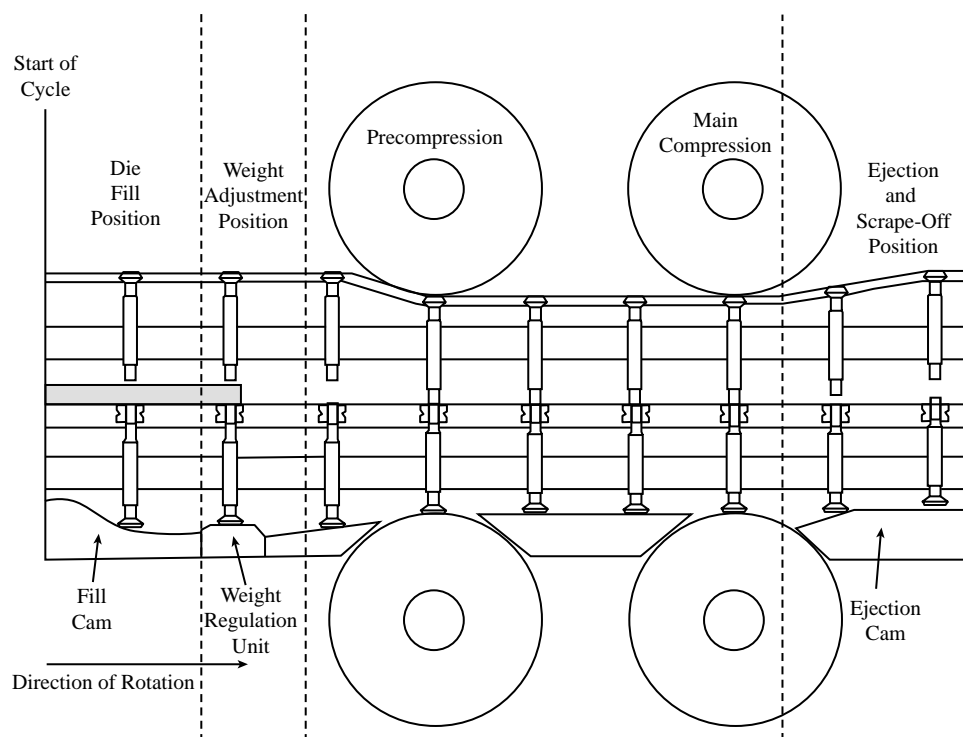


Fig. 4 Rotary press (sequences of compression).

from the die table surface reduces material loss due to the centrifugal force of the rotating die table.

Overfilling of the die cavity is necessary to achieve uniform tablet weights and to optimize machine running conditions. However, at times excessive overfilling can lead to other problems such as excessive wear for abrasive raw materials, particle size reduction for friable granulations, material segregation, and overmixing of lubricant. Therefore, the effect of different machine running conditions must be evaluated for each different product separately.

The material that is directed into the recirculation channel is subsequently introduced back into the feeder at the tablet stripper or at the inside edge of the feeder. It is worth noting that the paddle in the feeder at the point of material entry rotates in the opposite direction as the turret to aid in die fill and induce flow back into the feeder.

Frequently the maximum machine speed may depend on die fill characteristics due to excess tablet weight variation at high machine speeds. However, because the compression characteristics of most pharmaceutical products exhibit viscoelastic properties, the press speed may also have a major effect on the compressibility of the material. For this reason, the ability to compress a tablet adequately is often the overriding factor to consider in tablet compression.

The process of tablet formation begins as the upper punch is lowered directly into the die cavity after the excess material stripper. As mentioned previously, it is advantageous if the slug of material is lower than the die table surface as the upper punch enters to minimize uncontrolled material loss and weight variation. After the upper punch enters into the die, the upper and lower punches begin to move toward each other as the punches ride along cam tracks toward the precompression rollers. At the precompression stage, the initial (and typically the lower) compression force is applied. Traditionally, tablet presses were equipped to apply a 20 kN maximum precompression force using relatively small compression rollers (approximately 100-mm (4-in.) diameter rollers). However, to improve flexibility, many modern rotary tablet presses are equipped with identical precompression and main compression force capabilities, allowing the application of 80–100 kN forces using 250–300-mm-diameter compression rollers.

After application of the precompression force, the punches move toward the main compression rollers where the final (main) force is applied. As the punches impact the rollers, the compression force increases until the punch head flat is tangent to the compression roller and maximum force is applied (see Fig. 1). The applied

compression force is a measured value and depends on the distance between the punches and the quantity of material in the die. After main compression, the upper punch is pulled out of the die cavity while the lower punch impacts the ejection cam to begin the ejection process. As the die table continues to rotate, the lower punch raises the tablet out of the die cavity to eject the tablet to the point of scrape-off.

Press Design and Layout

Modern rotary tablet presses are typically designed in separate machine sections (press zones). Typical sections to provide separation and isolation of the compression area from the other components are as follows:

- Upper cam section
- Compression section
- Lower cam section
- Lower mechanical section
- Electrical section

With the proper separation of these areas, only the compression zone is exposed to material, thus reducing cleaning and change-over time of the tablet press. In addition to the machine sections, an understanding of other machine subsystems is necessary, such as the lubrication system and the diagnostic systems (safety systems) to achieve optimal machine performance.

Modern rotary tablet presses are either single-sided or double-sided. A single-sided machine has one feeding station, one set of precompression and main compression rollers, and one discharge station. These machines produce one tablet per punch station per die table revolution. A double-sided machine has two feeding stations, two sets of precompression and main compression rollers, and two discharge stations, and produces two tablets per punch station per die table revolution. The double-sided machine operates identically to the single-sided machine with the exception that the excess material from the first feeding station passes into the second feeding station. A double-sided machine has a higher output than a single-sided machine. Its pitch circle diameter is also greater, which could result in weight uniformity and compressibility issues.

Bilayer tablet presses employ the same general design concepts as single-sided machines. Typically, a double-sided machine can be converted to a bilayer machine by replacing various cams. The material for each layer is introduced separately into each feeder and is removed from the die table to prevent contamination.

Upper cam section

The upper cam section is typically shrouded and sealed to prevent exposure of material. It consists of the upper cam track, all upper compression rollers, and all adjustments to the position of the upper compression rollers. The primary components of the upper cam section are as follows:

1. *Upper punch removal/dwell cams:* The upper punches are loaded and removed from the machine at this location. These cams typically reside directly above the material feeder. In many press designs, the upper punch dwell cam is designed to measure the tightness of the upper punches in the turret. A spring loaded cam designed to raise the upper punch slightly (1–4 mm) is connected to a proximity sensor. If the punches are too tight then the spring-loaded cam falls instead of raising the upper punches, thus tripping the proximity sensor and shutting down the machine. In alternative press designs, the upper punch tightness is measured in the upper-punch pull-up cam, typically by a strain gauge measurement of the lifting force.
2. *Upper punch lowering cam:* The upper punches are lowered into the die cavity by the upper-punch lowering cam. This cam is typically CAD optimized to minimize the acceleration and velocity of the upper punch as it enters into the die cavity. In this way, the upper punch travels in a smooth and controlled manner as it enters the die cavity, thus improving weight uniformity.
3. *Upper precompression and main compression rollers insertion depth adjustments:* Insertion depth for both precompression and main compression is adjusted in the upper cam section. The insertion depth determines the location of tablet formation in the die cavity relative to the top of the die table as shown in Fig. 5. It is measured as the distance at which the upper punch enters into the die at the tangent between the upper punch head and the compression roller. Insertion depth can be varied between 2 and 6 mm on most machines and is typically maintained between 3 and 4 mm. For precompression and main compression, the insertion depth should be maintained at approximately the same position. On most modern rotary tablet presses, the adjustments for precompression and main compression insertion depth are independent. However, on many older designs, the precompression roller is attached to the main compression roller assembly and its position is measured relative to the main compression roller position. In this way, the ratio of precompression to main compression remains constant as machine adjustments are made.
4. *Upper punch pull-up cam:* After compression, the upper punch enters into the upper-punch pull-up cam, which removes the upper punch from the die cavity. This cam

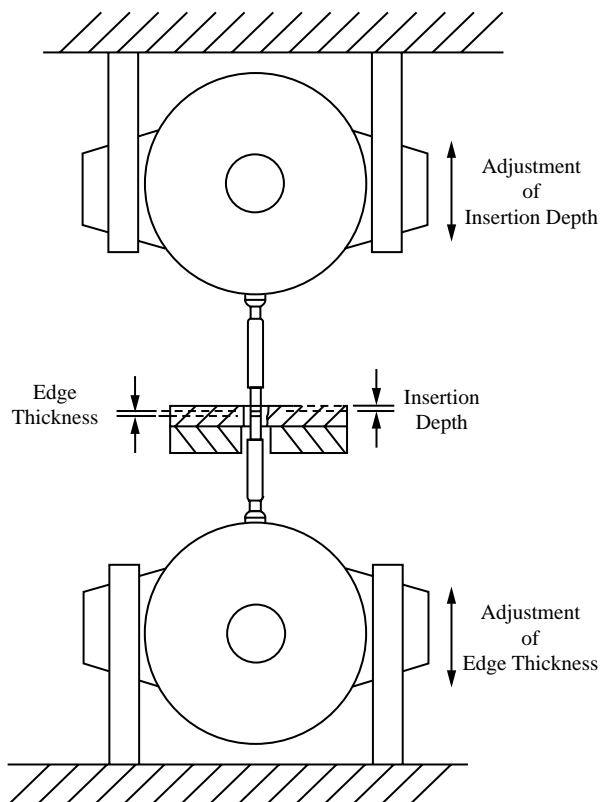


Fig. 5 Rotary tablet press layout.

provides an excellent location to measure the upper punch pull-up force that determines the tightness of the upper punches. Compared to the upper punch dwell cam, this location has the advantage of determining the punch tightness not only in the turret but also in the die cavity. Detection of tight punches at this location prevents almost all possibility of machine damage.

5. *Cam material of construction:* Both the upper and lower cam sections use cams to guide the punches while the turret rotates. These cams are typically made of various materials such as steel, bronze, or alloy. Most of the cam tracks in the turret are designed to smoothly guide the punches. Many modern rotary tablet presses use polymer composite cams for nonimpact points. These cams have excellent qualities in that they provide superior abrasion resistance and have self-lubricating properties minimizing cam and tool wear, heat generation, and noise, and ultimately resulting in increased machine speeds. However, cams that undergo impact (e.g., ejection cam) and stress (e.g., weight-regulation cam) require metal construction with good impact resistance. For this purpose, an aluminum-bronze alloy provides superior abrasion resistance and excellent impact strength.

Compression section

The compression section contains all components that are exposed to the material, such as the material hopper, the feeder, the excess material stripper, the upper and lower turrets, the die table, and the tablet stripper. Additionally, the dust-collection shrouds are located in the compression section. Proper shrouding of this area ensures that none of the upper and lower punch heads, compression rollers, and cam tracks are exposed to material. Proper maintenance and setup of the compression section is critical for optimal press performance. The primary components of the compression section are explained in the following section:

Material hopper: The material hopper is an integral part of the feeding system as shown in Fig. 6. Typically, it is capable of holding approximately 5–10 kg of material. Low level sensors are mounted in the hopper to signal an alarm, shut off the machine or activate a feeding mechanism to deliver more material when the product falls below this level. The material hopper should be symmetrical with steep discharge angles to promote mass flow and prevent funnel flow (rat holing) in the granulation. The discharge outlet of the hopper should be as large as possible reaching into the feeder to prevent material bridging.

On many machines the base of the hopper is equipped with a valve to shut off material flow to the feeder if necessary. Depending on the nature of the granulation, the hopper valve can contribute to material bridging. For materials with very poor flow characteristics, a slide valve may be preferable to a butterfly valve.

Older generation machines typically employ large hoppers, but more recent designs offer options to install

only a straight chute in place of a large overhead hopper. This design minimizes poor flow behavior and segregation due to funnel flow.

Gravity feed frame: Older machines typically employ gravity feed frames which rely on gravitational and turret rotational forces to achieve die fill. These feed frames provide good performance for materials with good flow properties but are typically limited to slow machine speeds. On the other hand, gravity feeders do not agitate the product and impart no energy. Therefore, they offer advantages for products where material segregation and overmixing are of concern. For example, products that are sensitive to overblending of magnesium stearate (i.e., exhibit capping when overblended) may exhibit improved compressibility by using a gravity feeder as opposed to a force feeder.

Force feeder: Force feeders are typically multi-chamber and multipaddle feeders as shown in Fig. 6. These feeders are critical to allow optimal press performance at high machine speeds with minimal weight variation. For products with good flow properties, the feeder should move the material from the overhead hopper to the dies with minimal mixing. Most force feeders contain two or three chambers and paddles. The three chamber/paddle system typically performs better than the two chamber/paddle designs. The top paddle and feed chamber are connected directly to the hopper and move the material from the overhead hopper to the filling chambers located directly above the die cavities. The top chamber eliminates the effect of the head pressure on material flow, thus providing uniform die fill regardless of the quantity of material in the hopper. Alternate systems offer level sensors that are designed to provide a constant quantity of material to the feeding chambers, thus also eliminating the effect of head pressure on material flow.

The force feeder chambers contain material baffles that function to prevent the material from randomly packing in the chambers, which results in nonuniform fill. Optimal systems provide minimal energy input and minimal particle mixing while providing uniform fill. The speed of the paddles can be synchronized with the die table speed minimizing tablet weight variation. The appropriate paddle speed can be determined by using a force-control system that displays the standard deviation of the compression force. The optimal feeder speed is determined by adjusting the feeder speed to achieve the lowest standard deviation in the compression force, which corresponds to the least weight variation.

A rectangular paddle design is typically used to minimize powder mixing in the feeder. However, for materials with poor flow characteristics (bridging in the hopper) due

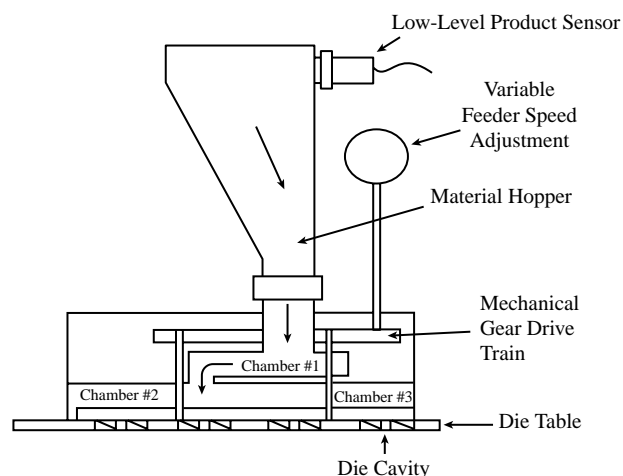


Fig. 6 Material feeding system.

to interparticle friction, a round (or wedge) paddle design can improve flow by forcing interparticle slippage. Under these circumstances, round paddles provide a mixing effect with possible impact on uniformity, compressibility, and dissolution.

The feeder height above the die table surface is very important to minimize product loss and prevent scaling of low melting materials. The feeder height is usually maintained between 0.05 and 0.10 mm (0.002–0.004 in.). Very fine particles may require a feeder height of 0.025 mm (0.001 in.).

Excess-material stripper: The excess-material stripper is located immediately after the feeding system and scrapes off the excess material on the die table after weight adjustment. It is often overlooked during setup although it is one of the most critical components of the tablet press. A brass stripper is employed, which sits flush on the die table under spring tension. The material is scraped off just before the lowering cam. The brass stripper directs the excess material into the recirculation channel. A tail-over-die covers the die cavity after scrape-off to the point of upper punch entry. This design minimizes uncontrolled material loss due to flinging of material out of the die cavity at high rotational speeds.

Precompression and main compression rollers: After die fill and scrape-off, the punches rotate to the precompression station where an initial force is applied to the compact. The tablet is frequently partially formed during the precompression stage. Subsequently, the upper and lower punches move together under the main compression rollers where the final tablet is formed. The main compression roller is usually larger than the precompression roller. However, latest advances suggest that similar sizes for precompression and main compression rollers with the ability to apply similar loads may result in optimal tablet formation. The compression rollers are made of premium tool steels and are surface hardened.

Because the compression characteristics of powders are time-dependent (the exact extent of this dependency depends on the primary modes of deformation), the final tablet properties depend not only on maximum compression forces but also on the rate at which these forces (rate of deformation) are applied and removed. On a rotary tablet press, the rate of deformation is determined by the tangential velocity of the punch and the compression roller diameters. The tangential velocity of the punch is a product of the press speed and the die table circumference (i.e., die table rpm \times 3.14 \times pitch circle diameter). As the tangential velocity increases, the rates of compression and decompression increase while the overall compression time decreases. The roller diameter affects both the rate of compression and decompression. As the diameter

increases, the rates of compression and decompression decrease.

Optimal compression efficiency is achieved on machines that offer multistage compression with high precompression and main compression force flexibility (typically 100 kN maximum). The roller diameters should be as large as possible to provide the lowest possible rates of compression and decompression. If compression problems exist, the longest time for compression should be allowed by running at low press speeds and running on machines with a small pitch-circle diameter.

As mentioned previously, for certain types of products, precompression at a force level higher than that of main compression may increase tablet hardness. The author has found that for materials that primarily undergo brittle fracture, application of a precompression force higher than the main compression force can result in a higher tablet hardness. However, this is typically not the case for materials with elastic properties (e.g., products prone to capping and lamination) because these products require gradual application of force to minimize elastic recovery and allow stress relaxation.

Most heavy tonnage machines (80–100 kN capability for precompression and main compression) have no mechanical linkage between the upper and lower compression rollers. Therefore, for these machines, movement of the upper punch insertion depth does not result in an equal movement of the lower compression roller position. The compression rollers are typically mounted to a block assembly that is adjusted by an eccentric or a vertical slide adjustment (see Fig. 5). An eccentric adjustment typically results in a slight off-center alignment as the roller moves through all of its possible positions, whereas a vertical slide adjustment always maintains the roller along the center line. On older rotary tablet presses the rollers are attached to a rocker arm with one side fixed to the machine roof or base. On the upper compression roller, the other side of the rocker arm is mechanically linked to the lower roller rocker arm. In this way, adjustment to insertion depth results in simultaneous adjustment of both the upper and lower roller positions. As with the eccentric roller adjustment, rocker arm position adjustment results in a slight off-center alignment as the roller moves through all of its possible positions.

Although most rotary tablet presses operate by maintaining fixed roller positions during compression, some designs incorporate a compression compensator system in which the counterforce for compression is air pressure. This system compresses to a constant force and allows roller movement when the preset force is achieved. Under these conditions, potential exists to increase the time that the force is maintained near its peak value (approx-

mately 90% of maximum). Compression to a constant force should theoretically provide a more uniform tablet hardness and more uniform dissolution profiles while allowing a greater variation in tablet thickness.

Tablet stripper: The tablet stripper scrapes off the tablets from the lower punch and directs them down the discharge chute. On high-speed machines, special attention must be paid to the tablet takeoff to prevent tablet backup; modifications are necessary for shaped tablets, as shown in Fig. 2. On high-speed machines it is critical to move the tablets off the die table as quickly as possible. Under some circumstances, repositioning of the Plexiglas cover on the tablet stripper to provide minimal clearance between the tablet and the cover may prevent shingling of tablets. The height of the lower punch at the point of scrape-off should always be checked to verify that it is not below the die table surface. Typically the lower punch should protrude approximately 1–2 mm from the die table surface at the point of scrape-off. For deep concave tablets, a protrusion height above 2 mm may be necessary.

Tablet presses equipped with single-tablet rejection capabilities reject tablets at the point of scrape-off. Based on the compression force of the punch station, single tablets are sorted by using a compressed-air blow off or a mechanical fast gate. On single-sided machines (36 stations) both systems work well for both large and small tablets. However, double-sided, high-speed machines may present difficulties for large tablets at high press speeds.

Material recirculation: Material is recirculated from the center of the turret into the feed frame. Some press designs include recessed recirculation channels to minimize particle attrition and prevent excess material loss to the vacuum system. It is critical not to recirculate too much material because this can result in low product yields and can have a detrimental effect on the powder's physical properties, which could result in poor compressibility, uniformity, and final properties (e.g. reduced dissolution rate).

The point of re-entry of the granulation into the feeder corresponds to the location where the lower punch enters into the fill cam when it is flush with the die table surface. Therefore, the material from the recirculation channel is typically the material in contact with the lower punch face and is the first material to be filled for each die cavity. The effect of the material in the recirculation channel should always be evaluated when compression problems occur, which can be associated with the lower punch. An example would be a case where a granulation exhibits picking only on the lower punch and not routinely on all tablet presses. Additionally, this granulation is slightly underlubricated, very friable, and undergoes minimal brittle fracture.

Previous experiments showed that, as particle size decreases, ejection force and picking tendencies increase. It was determined that, for this product, excessive material recirculation resulted in a reduction in particle size and an increase in picking tendencies. However, the picking problem was seen only on the lower punches because the reduced particle size material came into direct contact only with the lower punches. The problem was minimized by reducing the amount of material in the recirculation channel and redirecting the material from the recirculation channel to incorporate it into the bulk of the material in the feeder, thus preventing it from contacting the lower punch directly.

Dust extraction: Adequate dust extraction is necessary to maintain high-speed operation for extended periods of time. The entire compression area should be shrouded to minimize dust infiltration into other press areas. Effective dust extraction minimizes dust and oil contamination on the surface of the tablets, which could produce black specs. Insufficient dust extraction results in excessive material build-up on the lower and upper punches leading to tight punches. However, the proper balance of dust extraction without high levels of material loss must be determined. If the dust extraction level is too high material could be extracted from the die cavities and the recirculation channel. Furthermore, the dust extraction systems preferentially removes the fine particles. Therefore, if the granulation is a direct-compression blend where the active constituent is of fine particle size, minimum dust extraction levels combined with minimal recirculation may be necessary to prevent a loss of active constituents (resulting in possible low assay).

Lower cam section

The lower cam section is completely sealed from the compression section. It houses the lower compression rollers, the entire lower cam track that guides the lower punches as the turret rotates, and all adjustments for the lower precompression and main compression roller positions. Additionally, any motors necessary for automatic machine adjustment are contained in this section.

Fill cam: The fill cam is designed to lower the punch to overfill the die cavity. Lower-punch fill cams are typically available in a variety of sizes that are changed depending on the final fill depth as determined by the weight regulation cam. Press manufacturers recommend a fill cam in which the weight regulation cam operates in the approximate center of the fill cam. Typical fill cams have a range of approximately 10 mm with an increment range of 4 mm (e.g., 0–10, 4–14, 8–18, and 12–22 mm). Special-order or very shallow fill cams are also available (e.g., 0–6 mm). Alternatively, some manufacturers offer a greater selection

of cams with a narrower range and lower increment (e.g., 5.5 mm range with an increment of 2 mm). These cams should offer greater precision in fill in certain cases. However, with the narrower range, minor changes in the granulation density could result in the necessity to change fill cams throughout the course of the run.

Flexibility in fill cam options offers advantages for specific problem areas. For example, if a material is very abrasive, a shallow fill cam should be chosen to minimize the amount of material that is removed from the feeder and subsequently returned to the feeder and the recirculation channel. A deep fill cam can cause overpacking of the feeder, which could result in jamming, temperature increases, and overmixing of granulation or lubricant.

Weight regulation cam: The lower punch travels from the fill cam to the weight regulation cam, which determines the final volume of material that remains in the die cavity after scrape-off. Proper design and operation of this unit is essential to ensure uniform tablet weights. In general, the unit should operate in a manner to ensure smooth punch travel minimizing punch chatter as the lower punch is raised to a precise and constant height.

The weight-regulation unit consists of several critical components that determine its efficiency of operation as shown in Fig. 7. The lower punch rides on the dosing rail that maintains the lower punch at a constant height at the final fill depth. In order to minimize vertical movement (and subsequently punch chatter) the head of the lower punch is held tight against the lower dosing rail by the holding-down cam which is spring loaded. To ensure that the holding-down cam is tight against the head of the lower punch, a 1–5 mm (0.04–0.20 in.) gap should be maintained between this cam and the lower dosing cam

when the lower punch rests on the dosing cam. During press setup, the fit can be easily checked by placing a lower punch in the weight regulation unit and verifying the absence of vertical movement of the punch. This function is critical to minimize tablet weight variation at high speeds. On many press designs, the weight regulation unit contains a safety cam on its inside.

The lower dosing rail and holding-down cam should always be made of a tough, abrasion-resistant material; aluminum–bronze alloy is highly recommended. The condition of the dosing rail and the holding-down cam should be checked every 6 months to ensure optimal performance. If significant wear is observed, the cams should be replaced or, in the case of the holding-down cam, reworked to provide a tight fit.

As the lower punch leaves the dosing unit, it is pulled down slightly (approximately 2 mm) by the lowering cam. Periodically the condition of the lowering cam should be checked to ensure proper lowering of the punch in order to minimize weight variation.

Many modern presses are equipped to measure the tightness of the lower punches. This function is critical to minimize machine damage as well as cam wear. As mentioned above, excessive cam wear increases tablet weight variation. Lower-punch tight sensors are typically mounted in either of the two locations. The transition cam from the fill cam to the weight regulation unit is an ideal location to measure lower punch tight forces because it is designed to raise the level of the lower punch to the final dosing height. In addition, the lowering cam can measure the counterforce to pulling down the lower punch.

Lower punch brakes: Most rotary tablet presses are equipped with lower-punch brakes that are Teflon tipped and spring loaded to apply constant pressure to the lower punches. Alternatively, some manufacturers apply pressure to a friction belt that provides resistance on the lower punches. The lower-punch brakes act as a “retention” system for holding the lower punches in place during press setup. More importantly, these systems help to minimize lower-punch chatter at high press speeds thus minimizing tablet weight variation. Some press manufacturers use the lower punch seals to retain the lower punches.

Precompression and main compression rails: The precompression rail provides the transition support for the lower punch from the weight regulation unit to the precompression roller, while the main compression rail provides the transition support from the precompression roller to the main compression roller. The optimal press designs provide positive support with these cams by ensuring that the lower punch head flat is always in contact with the rail surface. In this way, there is no abrupt vertical movement of the lower punch as it passes to the

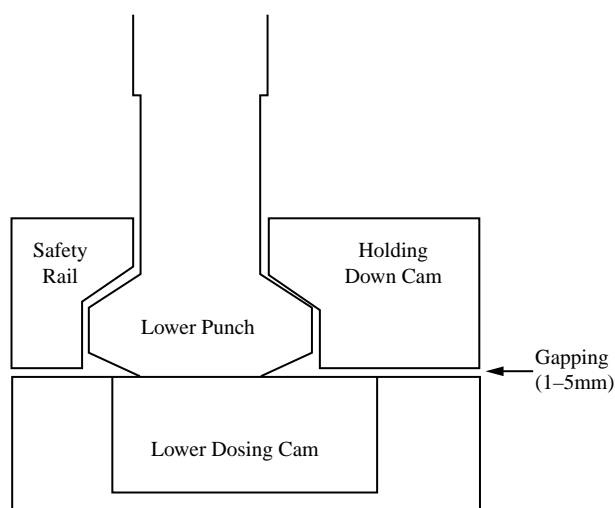


Fig. 7 Weight regulation unit.

compression rollers. Vertical movement of the lower punch before precompression and between the compression stations can cause the introduction of air into the bottom of the compact, resulting in capping at the lower-punch face. Many presses rely on the lower punch brakes or seals to prevent this type of movement. Under these conditions, vertical movement of the lower punches will gradually occur as these components wear causing periodic compression problems.

Adjustment of lower precompression and main compression roller thickness: The position of the lower compression rollers is adjusted from the lower cam section. As shown in Fig. 8, the position of the lower rollers relative to the upper rollers (i.e., insertion depth) determines the tablet edge thickness. Typically, the machine-control panel shows edge thickness on the indicator. However, adjustment of the edge thickness actually results in adjustment of the lower roller position only. For machines that have no mechanical link between the upper and lower compression rollers, the tablet edge thickness indication on the control panel is only valid at the specific insertion depth that was set during edge thickness calibration. However, for some of the electronic, fully automated machines, the machine automatically moves the lower compression roller during the insertion depth adjustment to maintain the same tablet edge thickness.

Ejection rail: After compression the lower punch impacts the ejection rail (or on some machines an ejection roller). Upon impact the tablet is broken free from the die

side wall and begins to move up the die as the machine rotates. The ejection rail should be made of a tough, abrasion resistant material such as aluminum bronze alloy.

Scrape-off rail: After riding up the ejection rail, the lower punch rides on the scrape-off rail to provide a constant height for tablet scrape-off. The height of the lower punch scrape-off can be adjusted to optimize the single tablet rejection height or the tablet scrape-off height.

Force overload system: Most tablet presses are equipped with force overload systems designed to prevent machine and punch damage. As stated previously, the compression force is not a set value, rather a measured value obtained from the fixed punch distance and the quantity of material in the die. On most tablet presses, a maximum allowable compression force can be set. This force setting is actually the counterforce to the measured compression force. If the compression force exceeds this counterforce, the compression assembly will back-off thus reducing the force. Most tablet presses use a hydraulic, air, or spring-loaded system on the lower compression assemblies (both precompression and main compression). In these systems the hydraulic or spring systems are calibrated to the measured force in the die and move instantaneously during an overload condition. Some of the more recent rotary tablet presses use strain gauges to measure the actual compression force for force overload (as opposed to force control). In these systems, the machine mechanically moves the compression assembly when the measured force is exceeded. Therefore, since these systems do not react as quickly as hydraulic, air, or spring systems they typically include a safety margin to initiate an overload condition prior to reaching the actual set value (e.g., 95% of maximum entered value).

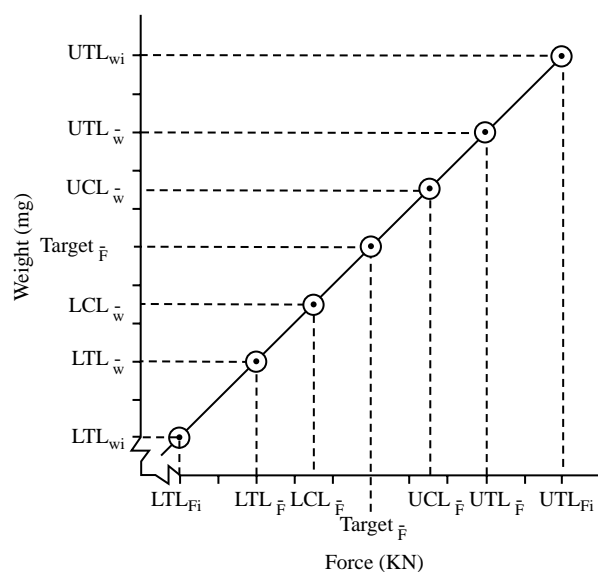


Fig. 8 Force versus weight relationship.

Lower mechanical section

The lower mechanical section houses the main drive motor, the gearbox, the hydraulic pump, the lubrication pump, and the signal wire distribution. Proper venting and cooling of the lower mechanical section is essential to prevent machine damage and minimize heat generation. This section should be equipped with a cooling system for products that are sensitive to heat generation (e.g., contain low melting point components that are prone to picking and sticking).

Electrical section

The electrical section contains all electrical controls and components (e.g., programmable controllers, relays, contacts, and fuses), the signal wire distribution, and the

integrated or remote force control systems. On many machines the electrical section is connected to the front of the press, thus minimizing space requirements. However, modern fully automated machines use computers that may be sensitive to machine vibration, dust, and heat. Therefore, these electronics should be located remotely from the machine.

Lubrication system

Most high speed rotary tablet presses employ automatic lubrication systems during operation. Effective punch lubrication is essential for the movement of the punches in the turret, in the cam tracks, and under the compression rollers. Punch-lubrication systems allow high speed operation over extended periods of time while minimizing cam and tool wear and reducing heat generation.

The lubrication pump is typically maintained in the lower mechanical section and allows variation of both the lubrication interval and time duration. The quantity of oil delivered is normally determined by the oil distribution nozzles connected to the distribution manifolds. The upper punches are usually lubricated on the punch head flat via a felt pad located in the upper punch dwell cam or the upper punch lowering cam. On many modern designs the upper punch barrel is also lubricated. The punch barrel is lubricated via radially drilled holes in the turret that transports oil to the punches by rotational forces, or by overlubricating the punch neck and allowing gravity to transfer the oil down the barrel. Because the upper punch shaft has the greatest area of contact with the upper turret, lubrication of the punch shaft is critical to allow high-speed operation. Poor lubrication of this area can result in heat generation and metal expansion, ultimately causing machine seizure and severe damage.

The lower punches are lubricated at the punch neck. Oil distribution lines are frequently provided to lubricate both sides of the punch neck. Subsequently, gravitational forces distribute the oil over the head of the punch. The lower punch barrel is typically lubricated by radially drilled holes in the turret.

All presses equipped with punch-lubrication systems require oil and dust seals to prevent oil contamination in the product and dust contamination in the turret punch sockets. These seals are normally double lipped, designed to strip oil on one side and powder on the other. As mentioned previously, some press designs use the lower punch seals to retain the lower punches.

Inadequate punch lubrication can lead to excessive heat generation, which could affect tablet properties. An example is a granulation that primarily undergoes elastic and plastic deformation. This product was normally run on

a tablet press without an automatic lubrication system. Production requirements resulted in batch campaigning. At the beginning of the production campaign, the punches were lubricated and placed into the machine. Over the course of the first batch, the product was easily maintained within tablet hardness and thickness specifications. However, as the campaign transitioned into batches two and three, tablet hardness tended to increase while the thickness remained constant. Machine adjustments were made to maintain the hardness and thickness within specification, but in the end the hardness could not be maintained below the maximum limit without exceeding the thickness specification. At this point the machine was inspected and the lower punches were observed to be warm. They were removed, cleaned, lubricated, and placed back in the machine. Upon resuming production, the tablets returned to their original characteristics. In this case, the lack of a punch lubrication system combined with batch campaigning (extended running conditions) resulted in a temperature increase in the machine. For this elastic material, the increased temperature resulted in greater plasticity and more stress relaxation, improving the compressibility of the product. Although under many circumstances this effect would be beneficial, in the present case the improved compressibility resulted in out-of-specification product. The problem was resolved by removing the punches after each batch for cleaning and lubrication.

Instrumentation

Modern rotary production tablet presses are typically equipped to measure precompression and main compression forces. Additionally, measurement and monitoring of tablet ejection force can prove to be beneficial for specific problem products and for production troubleshooting. However, for most pharmaceutical products proper product development and optimization work eliminate the need to instrument a production machine for ejection force. Rotary tablet presses can also be equipped to measure both upper punch and lower punch pull-down forces. These measurements are primarily made to detect tight-running punches and are necessary on production machines only if the machine monitoring system uses direct force measurement for these functions. Tablet scrape-off force can also be measured, but this is only recommended on development machines. Scrape-off forces are typically below 6 N. Therefore, instrumentation of a tablet stripper requires highly sensitive instrumentation that is easily damaged.

Precompression and main compression forces are normally measured for the upper punches (20, 21). These forces are typically measured using strain gauges arranged in a full wheatstone bridge (22). Strain gauges are basically resistors applied to the metal surface in a specific orientation. Under load, the member deflects and the strain changes the resistance of the gauge. The change in resistance is proportional to the applied force. However, due to design differences, some machines measure the lower compression forces as opposed to upper compression forces. The compression force should be measured as close to the compression event as possible. For the most accurate and reproducible readings, the strain gauge should be in line with the compression event as opposed to off center at a remote location. This is easily accomplished on most rotary tablet presses by using a shear pin to support the compression rollers. However, the pin must not be rotated when the position of the compression rollers is changed. Therefore, this system is ideally suited for machines that change roller position by a vertical slide mechanism as opposed to an eccentric mechanism. Shear pins are typically custom manufactured, where quality depends not only on the pin design but also on the strain gauge receptivity and arrangement.

Many modern rotary tablet presses use off-the-shelf load cells for force measurement. These load cells are highly accurate, durable, and easily replaced and calibrated. However, the final accuracy and repeatability of force measurement in the machine not only depend on the quality of the load cell, but also on the design of the compression assembly and the placement of the load cell within the assembly.

Machines that utilize rocker arms with a mechanical linkage between the upper and lower compression assemblies are normally instrumented by applying a strain gauge to the upper rocker arm or to the mechanical linkage connecting the assemblies. The point of strain gauge application is "necked-down" to increase the sensitivity of the member. These instrumented members should be calibrated in the machine to account for the effect of other machine members on the measured force.

The force measurement system (strain gauges or load cells) should be calibrated on a yearly basis or after the compression assembly has been disassembled for any reason. The calibration should be made in the machine to assure accuracy. It is typically performed using a modified punch assembly and a calibrated load cell that are rotated under the compression roller to produce a load. The output of the machine force measurement system is compared to the output of the calibrated load cell. Many machine manufacturers use this single-point calibration to modify the strain gauge factor so that the two outputs are the same

at this load level. Subsequently, different load levels are tested and the error between the machine force measurement system and the calibrated load cell is determined. Alternatively, loads can be applied to the system ranging from the minimum to maximum compression forces. The outputs from the calibrated load cell and the machine force measurement system are recorded and a linear regression is performed on the data to calculate a new strain gauge factor across the entire force measurement range. Subsequently, different load levels are tested and the error between the readings is calculated.

PRESS CONTROL AND AUTOMATION

Conventional rotary tablet presses are controlled by periodically taking tablet samples from the discharge chute and checking their tablet weight, thickness, and hardness. If the tablet weight is outside of the established control limit, the operator increases or decreases the weight by adjusting the fill depth (increasing fill depth increases tablet weight and decreasing fill depth reduces tablet weight). If either the tablet hardness or thickness requires adjustment, the operator typically adjusts the tablet edge thickness on either main compression or precompression. Since there is normally an inverse relationship between the tablet hardness and thickness, the operator usually reduces edge thickness to increase tablet hardness (or decrease thickness) or increases edge thickness to decrease tablet hardness (or increase thickness).

Force and weight control systems use the same basic concepts as conventional machines. Force control systems monitor the tablet compression force and adjust the fill depth to maintain a constant force. Force control systems alone compensate for flow and density variations in the granulation, providing a constant fill quantity. Weight control systems, on the other hand, work in conjunction with the force control system as a secondary control loop and replace the manual function of the machine operator, periodically removing tablet samples to test tablet weight, thickness, and hardness. If the weight control system indicates that the tablet weight must be adjusted, the force control setpoint or the tablet edge thickness is altered resulting in a change in the fill depth from the force control system.

Force Control

During tablet compression, the distance between the rollers remains constant unless a machine adjustment is made to change tablet hardness or thickness. Additionally, all

Table 1 Tablet weight-control points and tolerance limits

Item	Description	Specification (mg)
UTL_{w_i}	Upper tolerance limit of individual tablet weight	105.0
$UTL_{\bar{w}}$	Upper tolerance limit of average tablet weight	103.0
$UCL_{\bar{w}}$	Upper control limit of average tablet weight	101.5
$Target_{\bar{w}}$	Target of average tablet weight	100.0
$LCL_{\bar{w}}$	Lower control limit of average tablet weight	98.5
$LTL_{\bar{w}}$	Lower tolerance limit of average tablet weight	97.0
LTL_{w_i}	Lower tolerance limit of individual tablet weight	95.0

tooling dimensions (tooling length and die cavity size) are constant within established standards. Under these conditions, for a specific material of uniform density, if the same volume of material is delivered to each die, the maximum measured compression force for each punch station is the same. If, on the other hand, different volumes of material are delivered to each die, the maximum measured compression force for each punch station is different. On this basis, adjustment of fill depth (fill volume) to maintain a constant compression force should result in a constant tablet weight. This concept is the general basis of all rotary tablet press force control systems.

The force control systems assume a linear relationship between tablet weight and compression force for a particular granulation, tooling set, and machine tablet edge thickness (i.e., distance between compression rollers). By establishing the relationship between compression force and tablet weight at a specific machine tablet edge thickness (as shown in Fig. 8), a force control system is able to maintain a constant tablet weight by maintaining a constant compression force. Additionally, a force control system is capable of monitoring every compression force and rejecting tablets when the forces exceed specific established limits (for both average forces and individual

forces), thereby essentially monitoring every tablet weight.

The first step in using a force control system is to establish the force versus weight relationship to allow calculation of the appropriate force control set points which correspond to the desired weight control points. Table 1 gives example of weight control points and tolerance limits for a theoretical product.

The tablet press is initially run at target conditions to make the product within specifications. For example, the tablets are made at target conditions of 100 mg at an average compression force of 10 kN. After establishing this point, the fill depth is adjusted to either increase or decrease the tablet weight and the corresponding compression force is measured. In the present example, the tablet weight was increased to 103 mg and the average measured compression force was approximately 13 kN. This procedure should be repeated for several different tablet weights. The data are used for regression analysis to calculate the required force set points that correspond to the weight control points as shown in Fig. 8 and Table 2.

As an alternative to performing regression analysis, fill depth can be adjusted to achieve each average weight

Table 2 Force setpoints

Item	Description	Specification (kN)
UTL_{F_i}	Upper tolerance limit for individual compression force	15.0
$UTL_{\bar{F}}$	Upper tolerance limit for average compression force	13.0
$UCL_{\bar{F}}$	Upper control limit for average compression force	11.5
$Target_{\bar{F}}$	Target for average compression force	10.0
$LCL_{\bar{F}}$	Lower control limit for average compression force	8.5
$LTL_{\bar{F}}$	Lower tolerance limit for average compression force	7.0
LTL_{F_i}	Lower tolerance limit for individual compression force	5.0

requirement and the resultant compression force can be recorded and set.

During normal production with the force control system in operation as specified above, the tablet press will operate to maintain the constant compression force of 10 kN by adjusting the fill depth (see Fig. 9). Most force control systems do not require the user to input the upper and lower force control limits for average compression force that are typically set by the manufacturer within tighter tolerances than those demanded by the weight requirements. The control system typically calculates the average compression force every revolution and compares it to the force set point. If the average measured compression force varies from the set point of 10 kN (outside of the tolerance set by the manufacturer) then the force control system adjusts the fill depth to return the force to 10 kN. If the compression force is outside the average tolerance limits (below 7 kN or above 13 kN), the machine will shut down and reject the tablets. This condition should correspond to average tablet weights outside of the 97.0–103.0 mg limits. Typically, the force control system is capable of adjusting the fill depth to maintain the compression force well within the tolerance limits. However, if the material has poor flow characteristics and exhibits bridging followed by surging, the system may be unable to compensate quickly enough to prevent these out of control conditions.

During the entire operation, if any of the individual measured compression forces goes outside of the individual tolerance limits of 5 kN or 15 kN (corresponding to individual tablet weights below 95 mg or above 105 mg), these individual tablets are rejected at the point of tablet scrape-off. Some machine designs are not effective at sorting individual tablets reliably and reject multiple tablets if this condition occurs. For most machine designs, the user can specify a maximum number of individual tablets that can be sorted per punch location or

per batch before the machine shuts down. Exceeding these maximum limits may indicate a problem punch or excessive weight variation for the batch, possibly related to a setup problem.

As described, a force control system maintaining a constant compression force will adequately compensate for variations in granulation density, thus providing more uniform tablet weights. However, operation of this system still requires an operator to periodically check tablet weight, thickness, and hardness. If during a weight check the operator determines that the average tablet weight has gone beyond the control limit, while the average compression force is still being maintained at its set point, the operator must take one of the following actions:

- The system is placed in manual mode (shut-off force control system) and the tablet weight adjusted back to target by increasing or decreasing the fill depth as necessary. Once the weight is within requirements, the force control set point is changed (and all other limits by the same amount) to the current value that is displayed and the machine is returned to the automatic mode.
- While in the automatic mode, the force control set point is increased or decreased to increase or decrease the tablet weight, respectively. For example, by increasing the force set point, the control system increases fill depth to achieve the higher compression force requirement thus increasing tablet weight. The machine is allowed to stabilize and tablet weight is checked. Adjustments are made until tablet weight is within specifications.
- While in the automatic mode, the machine tablet edge thickness is increased or decreased to increase or decrease the tablet weight, respectively. For example, by increasing the machine tablet edge thickness, the measured compression force is decreased. The force control system will then increase fill depth to return the compression force to its previous value thus increasing tablet weight.

Minor changes in the force to weight relationship over the course of a compression run are common. These may be due to changes in the compressibility of the granulation or to changes in the machine over the course of the run (such as temperature changes).

Tablet presses that use a compression compensator system, which in turn compress to constant compression forces and allow roller movement, operate by the same control theories as those presented for force control. However, these systems measure roller displacement as opposed to compression force and relate it to tablet weight.

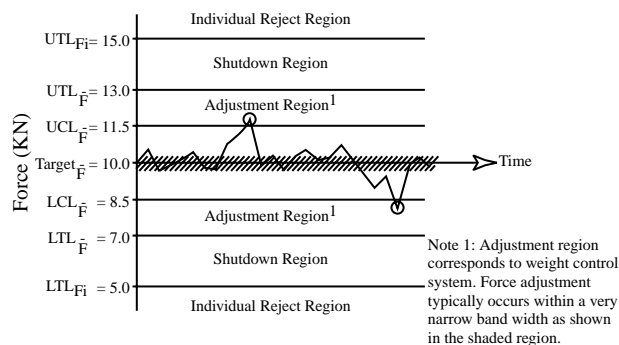


Fig. 9 Force control system.

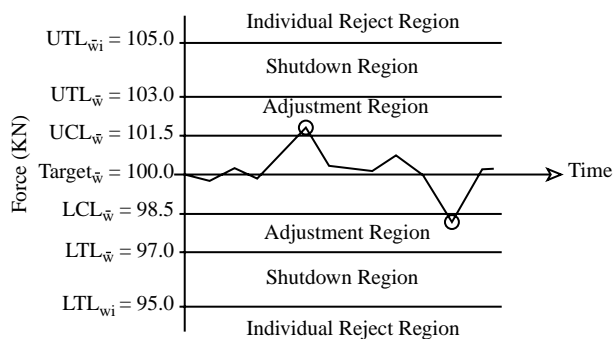


Fig. 10 Weight control system.

Weight Control

Weight control, as a secondary control loop to force control, allows automation of the tableting operation. The weight control system (Fig. 10) maintains the same limits as those presented for force control (Fig. 9). The machine assumes the place of the operator and periodically samples the tablet press to determine the tablet weight. The sampling requirements should be set to the same interval and number of tablets as those required for manual operation. Most weight feedback systems measure individual tablet weights and calculate the average weight for feed back purposes.

In the example given above, it is assumed that the weight feedback system will measure 10 individual tablet weights every 15 min. The force control system is in operation and is maintaining the average compression force at 10 kN. If during each 15 min. check, the average weight of 10 tablets is within the average control limits (see Fig. 10), no machine adjustments from the weight control system will be made. Alternatively, if the average tablet weight goes beyond the average control limits (above 101.5 mg or below 98.5 mg) the weight control system will initiate a change in one of the following ways depending on the machine:

Increase or decrease the force control setpoint to increase or decrease the tablet weight, respectively. The new force control set point is calculated by means of the machine manufacturer's control algorithm. Typically after adjustment to the new force control set point, the weight control system should resample and verify that the average tablet weight is now within the average control limits.

Increase or decrease the machine tablet edge thickness to increase or decrease the tablet weight, respectively. The new machine tablet edge thickness setting is calculated by the manufacturer's algorithm. After the new setting has

been made and the fill depth adjusted to maintain constant compression force, the weight control system should resample and verify that the average tablet weight is now within the average control limits.

The weight control systems that change the force control setpoint result in a change of the compression force throughout the run with a relatively constant tablet thickness. On the other hand, by changing the machine tablet edge thickness, the compression force remains relatively constant throughout the run while the tablet thickness varies. Since these adjustments are usually relatively small, both methods of machine adjustment typically produce similar results.

If during any check the average tablet weight exceeds the average tolerance limits (below 97 mg and above 103 mg) or an individual tablet exceeds the individual tolerance limits (below 95 mg and above 105 mg), the machine will shut off and reject the tablets. It is important to note that individual tablet weights exceeding the tolerance limits indicate that the tablet rejection system is not functioning properly.

Control and Monitoring of Weight/Thickness/Hardness

Tablet presses equipped to measure tablet weight, thickness, and hardness use the same concepts for force and weight control as presented here. These systems offer the additional flexibility of testing and controlling both tablet thickness and hardness. However, most machine manufacturers and tablet manufacturers have found that additional control of either tablet thickness or hardness is both difficult and costly. It is cost effective to only monitor tablet thickness or hardness and shut-off the machine if tolerance limits are exceeded. In this way, fully automated operation is possible without operator testing.

TROUBLESHOOTING

A proper understanding of a material's compression characteristics combined with a knowledge of tablet compression equipment allows efficient troubleshooting of production problems. Although there is no substitute for a robust granulation, a product can be optimized by examining all of the different machine factors that can affect performance. By applying a variety of the concepts discussed here, a large variety of processing problems can be eliminated.

Tablet Weight Variation

Excessive tablet weight variation can be caused by a variety of factors. For many granulations, the inherent poor flow characteristics of the material may be the rate limiting step, and simply slowing down the machine may reduce weight variation. Additionally, optimization of the feeder paddle speeds to minimize the standard deviation in the compression force should help to minimize weight variation. If weight variation is excessive, the following machine components should be examined:

- The tightness of the *hold-down cam* should be examined to verify that it is not excessively worn and is holding the lower punch tight against the dosing cam. If the previous product resulted in tight lower punches, premature wear may have occurred on this cam, causing increased weight variation.
- The condition of the *lower punch pull-down cam* should be examined to verify that it is not overly worn and that it drops the lower punch to pull the material in the die cavity below the die table surface.
- Both the condition and position of the *excess material stripper* should be examined to assure that it sits snug and level on the die table surface for complete scrape-off.
- Different types of *feeder paddles* can be used to promote flow (e.g., round feeder paddles are used for materials that exhibit bridging). The feeder speed should be optimized to minimize force and weight variation.
- The best *fill cam size* is that where the fill depth is centered in its range.
- A minimum amount of *material recirculation* is necessary to provide steady flow and fill. Too much recirculation can result in material back-up and reduction in the granulation particle size. The recirculation channel must be free of obstructions (i.e., broken tablets).
- If large tablets are being produced requiring deep fill depths, then *insertion depth* should be increased. Otherwise, as the punches pass from a deep fill to a shallow insertion depth, the uncompressed granulation will be pushed out of the die cavity resulting in material loss.

Product Yield

Product yield can be affected by a number of factors. Unfortunately, yield problems are not noticed until after the loss occurs. However, by paying attention to the following areas during setup, these problems can be minimized:

- Excessive material loss can be avoided by ensuring that the *excess material stripper* is flush against the die

table. Otherwise the material will be sucked into the dust extraction system.

- The *feed frame height* should be maintained between 0.05 and 0.10 mm (0.003–0.004 in.). For very fine particle size granulation, the clearance should be reduced to 0.025–0.05 mm (0.001–0.002 in.).
- The *fill cam size* should be reduced to minimize overfilling and material recirculation.
- Material in the *material recirculation* channel should be maintained at a minimal level. As more material is recirculated, the likelihood for loss is greater. The piece guiding the material from the recirculation channel to the feeder must be properly positioned.
- If the *insertion depth* is too shallow relative to the fill depth, material will be pushed out of the die and lost to the dust extraction system.
- Excessive *feeder speeds* cause excess material recirculation and increase material loss.
- It may be necessary to reduce *press speed*. As the press speed is increased, the turret rotational forces will sling more material to the outside of the turret from both the recirculation channel and the die cavity.

Low Hardness

Tablet hardness is affected by many factors. For troubleshooting purposes, it should be determined if the low hardness is due to capping or noncompressibility. For formulations that exhibit low hardness without capping, the following guidelines are helpful:

- For *multi-stage compression*, a machine should be equipped for both precompression and main compression with large diameter rollers.
- The *ratio of precompression to main compression* should be adjusted. For tablets that exhibit no capping problems, both precompression and main compression forces should be maximized.
- *Press speed* is reduced in order to increase total compression time.
- Formulations sensitive to lubricant levels may exhibit low hardness due to overmixing in the feeder. If the formulation contains a significant quantity of magnesium stearate or there is a shift in particle size thus extending the lubricant differently, overmixing can reduce tablet hardness. In this case, the *feeder speed* should be reduced to a minimum.

Capping and Lamination

Tablet capping and lamination typically create the most difficult problems, due to a variety of causes. Identification of the cause often leads to the solution. The basic concepts to alleviate these problems center on minimizing elastic

behavior while promoting plastic deformation. Depending on the exact nature of the problem, this can be achieved from a formulation perspective by modifying the formula to incorporate a plastically deforming matrix, by adding components to enhance bonding, or by increasing the moisture level. Alternatively, from a machine perspective the following guidelines should be followed:

- The *rate of force application* should be reduced by applying the compression force as gradually as possible. This can be accomplished by lowering the press speed or using a machine with a small pitch circle diameter.
- The *ratio of precompression to main compression* should be modified with gradual application of the precompression force followed by main compression. A precompression force that is too high can be harmful.
- The effect of reducing the *compression force* should be evaluated. In many circumstances, overcompression of a granulation will result in failure.
- A machine with large *compression-roller diameters* should be used to minimize the rate of force application.
- *Die cavity wear* must be investigated and the condition of the die cavities examined. If the dies are very worn, slip-stick behavior may occur during tablet ejection resulting in tablet failure.
- *Curled/damaged punches* promote tablet capping. Under these conditions the tools should be reworked or replaced.

Picking and Sticking

Tablet picking and sticking problems are typically related to formulation issues. However, in small scale manufacturing these problems do not occur frequently. Regardless, once a product is approved, it is difficult to make significant formulation changes. To minimize these problems the following areas should be considered:

- *Heat of compression*: Excessive heat generation during compression will increase the picking tendency of a low melting material. Use of cooling systems for the compression section or lower mechanical section may be helpful.
- *Press speed*: Lowering press speed and compression force reduces heat generation. Lower press speeds extend the contact time between the material and the punch face.
- *Precompression force*: Elimination of precompression may prevent picking. For example, some materials pick if the compression force is too low. Therefore, application of precompression at low forces may result in tablet picking.

- *Tool condition*: The condition of press tooling should always be evaluated when picking occurs. Polishing the punches and application of various coatings to the tools may help to eliminate picking for certain materials.
- *Start-up conditions*: Start-up should always be close to optimum conditions to prevent fouling the punch faces. If maintaining the force above a minimum is necessary to prevent picking, starting up near aim conditions of compression force prevents initial picking.
- *Tablet stripper*: The point of impact of the lower punches should be repositioned relative to the tablet stripper. Under many circumstances the stress of impact on the stripper can cause failure and sticking in logos.

Tablet Jams and Chipping

Tablet jams and chipping at the stripper reduce tablet quality and may contaminate the feeder resulting in weight variation. High speed machines typically have greater problems than slower machines. These problems are solved by focusing on the tablet stripper and lower punch ejection height.

- *Scrape-off height*: For many tablet types (e.g. deep concave tablets) the height of the lower punch at the point of scrape-off must be increased to ensure that the tablet is removed completely from the die when it impacts the tablet stripper.
- *Modified tablet stripper*: For shaped tablets, a modified tablet stripper that removes the tablets from the die table quickly prevents tablet back-up and breakage.
- *Height of plastic cover*: If the tablets exhibit layering or shingling, the height of the stripper cover should be lowered to just accommodate one tablet between the cover and the die table.
- *Tablet stripper*: The impact point of the stripper can cause chipping. Both the height and position of the tablet stripper must be checked. Damaged punches must be replaced or repaired.
- *Air assist*: If tablet jams cannot be eliminated by modifying the tablet stripper, an air assist blow-off at the tablet stripper may solve the problem.
- *Static eliminator*: In the production of lightweight tablets in a low humidity environment, static electricity may cause tablet back-up. In this case, installation of a static eliminator will improve tablet discharge.

SUMMARY

Optimal manufacturing of tablets requires good equipment and materials. The equipment is only a tool to achieve

the final goal of high quality tablets. However, if the materials have marginal flow and compressibility problems, machine flexibility allows the production of the highest quality product possible.

REFERENCES

1. Higuchi, T.; Arnold, R.D.; Tucker, S.J.; Busse, L.W. J. Am. Pharm. Assoc. **1952**, *41*, 93–96, Sci. Ed.
2. Train, D. J. Pharm. Pharmacol. **1956**, *8*, 745–761.
3. Nystrom, C. Drug Dev. Ind. Pharm. **1993**, *19*, 2143–2196.
4. Wray, P.E. Drug Dev. Ind. Pharm. **1992**, *18*, 627–658.
5. Hoblitzell, J.R.; Rhodes, C.T. Drug Dev. Ind. Pharm. **1990**, *16*, 469–507.
6. Marshall, K. Compression and Consolidation of Powdered Solids. *The Theory and Practice of Industrial Pharmacy*, 3rd Ed.; Lea & Febiger: Philadelphia, 1986.
7. Parrot, E.L. Compression. *Pharm. Dosage Forms: Tablets*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1990; 2.
8. Goetzel, C.G. *Treatise of Powder Metallurgy*; Interscience Publishers, Inc.: New York, 1949; 1, 259–312.
9. Heistand, E.N. Int.: J. Pharm. **1991**, *67*, 217–229.
10. Hiestand, E.N.; Smith, D.P. Int. J. Pharm. **1991**, *67*, 231–246.
11. Reier, G.E.; Shangraw, R.E. J. Pharm. Sci. **1966**, *55*, 510–514.
12. Hanus, E.J.; King, L.D. J. Pharm. Sci. **1968**, *57*, 677–684.
13. Rankell, A.S.; Higuchi, T. J. Pharm. Sci. **1968**, *57* (4), 574–577.
14. *Fette Technical Bulletin 8004*; Fette America Inc.: Rockaway, NJ.
15. Esezobo, S.; Pilpel, N. J. Pharm. Pharmacol. **1986**, *38*, 409–413.
16. Gunsell, W.C.; Kanig, J.L. Tablets. *The Theory and Practice of Industrial Pharmacy*, 2nd Ed.; Lea & Febiger: Philadelphia, 1976.
17. Hiestand, E.N.; Peot, C.B.; Ochs, J.E. J. Pharm. Sci. **1977**, *66*, 510–519.
18. Bateman, S.D.; Rubinstein, M.H.; Thacker, H.S. Pharm. Tech. Int. **June 1990**, 30–36.
19. Rippie, E.G.; Morehead, W.T. J. Pharm. Sci. **1994**, *83*, 708–715.
20. Watt, P.R. *Tablet Machine Instrumentation in Pharmaceutics: Principles and Practice*; Ellis Horwood: Chichester, UK, 1988.
21. Oates, R.J.; Mitchell, A.G. J. Pharm. Pharmacol. **1994**, *46* (4), 270–275.
22. Schmidt, P.C.; Vogel, P.J. Drug Dev. Ind. Pharm. **1994**, *20*, 921–934.

TABLET EVALUATION USING NEAR-INFRARED SPECTROSCOPY

Christopher T. Rhodes

Karen Morisseau

University of Rhode Island, Kingston, Rhode Island, U.S.A.

INTRODUCTION

Near-infrared spectroscopy (NIRS) continues to grow in importance as a useful analytical technique. It offers unique potential as a rapid, nondestructive method of quantitative and qualitative evaluation. NIRS has been used extensively in the food and agricultural industries for many years to determine moisture, protein, and starch content in grains (1). The pharmaceutical industry has been cautiously slow to accept NIRS as a commonly used technique, probably because of the absence of primary absorption bands. In recent years, an increasing amount of academic research is being carried out on the theory behind NIR. The use of NIRS for pharmaceutical applications has grown owing, in part, to technological advances in instrumentation and software.

Several reviews (2–7) have been published recently, attesting to its increasing popularity. Textbooks (8) relating to pharmaceutical uses of NIRS are becoming more common. Literature references pertaining to pharmaceutical uses date back to the early 1980s (9, 10). However, earlier NIR articles exist and were not taken seriously by the pharmaceutical industry at the time. In 1966, Sinsheimer and Keuhnelian (11) reported quantitative NIRS work with pharmaceutically active compounds pressed into pellets.

NIRS involves the multidisciplinary approaches of the analytical chemist, statistician, and computer programmer. The word chemometrics refers to the application of mathematical or statistical methods to measurements made on chemical systems of varying complexity. Chemometrics is defined (12) as the chemical discipline that uses mathematical, statistical, and other methods that apply formal logic to design or select optimal measurement procedures and experiments, and to provide maximum relevant chemical information by analyzing chemical data.

Chemometrics has found widespread use in the interpretation of analytical data and is relied on for the development of NIRS methods.

In pharmaceutical applications, NIRS is more often used as a secondary analytical technique than as a primary

tool. As a secondary method, a reference method is required to determine the reference component values that are to be used in the NIR calibration. The mathematical expression relating the component property (or properties) to absorbance is called a calibration model (also referred to as an algorithm). Using sophisticated spectral software, the analyst can correlate sample spectra to laboratory data, develop a calibration model, and apply that model to similar, new samples to predict constituent properties.

THEORY

The NIR region of the electromagnetic spectrum is from 800 to 2500 nm. The segment from 1100 to 2500, known as the Herschel region (13), is the range most often used in the analysis of pharmaceutical products. In the NIR region, the radiation can penetrate compacted materials such as tablets, providing a vast amount of spectral information about the sample.

The NIR region of the spectrum contains overtones and combination bands that are primarily attributed to hydrogen vibrations (OH, CH, NH). These overtones and combination bands are much weaker than the fundamental vibrations, thus, the molar absorptivities are much smaller than those of the corresponding infrared bands. Smaller molar absorptivities allow the use of undiluted samples and penetration of solid samples with good results.

There are several notable differences among the near-infrared region and other infrared regions of the electromagnetic spectrum. Conventional infrared instruments usually operate in the near-, mid-, or far-infrared regions, depending on the energy source and the detectors used. The wavelength range used for the NIR is just beyond the visible end of the spectrum and is referred to in terms of nanometers. Other regions of the spectrum are referred to in terms of wave numbers. Thus, the near-infrared region is $14,300\text{--}4000\text{ cm}^{-1}$, the midinfrared range is $4000\text{--}200\text{ cm}^{-1}$, and the far infrared is from $200\text{--}10\text{ cm}^{-1}$.

INSTRUMENTATION

NIRS instruments are typically designed as either transmittance or reflectance, with some allowing the user to switch from one to the other. The difference between the two instrument configurations lies in the positioning of the sample and the detector(s). In transmittance mode, the sample is placed between the monochromator and the detector so that the entire pathlength of the sample is integrated into the measurement. Transmittance measurements require higher frequency energy (800–1400 nm) because of the greater depth of penetration into the sample (14). In reflectance mode, the monochromatic light is illuminated directly onto the sample, and the reflected light is collected by detectors positioned at 45° angles to the sample.

Advances in instrumentation have resulted in a wide array of choices for analysts. These include instruments for specific applications with custom sample holders as well as general multiuse types. In the past 10 years, various patents have been issued for sample holders (15), sample supports (16), and a fiber optic system for dissolution (17).

An extensive product review of recent NIR technology was published by Noble (18). Enormous progress has been achieved in chemometrics and computing power, making many new applications possible. There are dozens of manufacturers of NIR spectrophotometers in the United States. There are many vendors of sampling components and software packages for data analysis. Research data of the most recent instrumentation and software are available directly on the World Wide Web, as most manufacturers maintain a Web site. There are numerous Internet (19) sites that provide links to professional spectroscopy societies, publishers of spectroscopy journals, and patents (20) related to pharmaceutical uses of NIR.

The US FDA's Division of Drug Analysis sponsored a cooperative study among manufacturers of NIRS instruments and pharmaceutical manufacturers to analyze a series of tablets, hard and soft gelatin capsules, and powders. One goal of the study was to demonstrate that a near-infrared scan was not unique among manufacturers. Another reason for the study was to establish FDA guidelines for the Instrument Qualification and Performance Verification of NIR instruments used in pharmaceutical analysis. Nine manufacturers participated in the study and represented makers of numerous types of monochromators (acousto-optical tunable filter or AOTF, dispersive and Fourier transform-NIR or FT-NIR). Ciurczak (21) reported a detailed account of the study to compare instrument performance in similar products, thus providing some comparative information to potential

buyers of NIR equipment. Ciurczak concluded that the quality of the spectra depends on the monochromator, and there is a range of noise variability in both FT-NIR and dispersive-type instruments. Examination of the spectra appearing in the report indicated differences in photometric noise and in spectral resolution (in certain regions) among instruments. One of the manufacturers, Analytical Spectral Devices, Inc., Boulder, CO, was identified in the study as the source of the donated spectra for the report and also provided their data over the Internet.

DATA ANALYSIS AND CALIBRATION

Obtaining Sample Spectra

The process of scanning a sample is quite simple and very rapid. The sample holder and surface must first be gently cleaned of debris and a reference (usually an internal ceramic Coor's standard) scan taken. The sample may then be placed in the sample holder, which may hold one or more of a specific type of sample. The sample is positioned, the lid closed, and the scan taken. Scan times are usually approximately 40 s. If multiple scans of the same sample are needed, the sample may be removed and rescanned. Instrument software facilitates the process and allows the spectra to be named and stored in data files.

Calibration

Calibration models are developed to determine the relationship between calibration set spectra and the constituent value of interest for those samples. Calibration involves taking spectra from many samples varying over the measurement range and also measuring the desired parameters. A rugged chemometric model for a complex sample may require hundreds to thousands of samples taken from all possible situations, in and out of specification, that it may encounter. Samples selected for calibration must contain all of the variables affecting the chemical and physical properties of the samples to be analyzed. To characterize each source of variation, it is recommended that 15 to 20 samples be run for each variable. Application of a math treatment, such as second derivative, prepares the raw spectral data for use in a regression and subsequent development of a calibration equation. This type of treatment results in a data file that will yield more information more easily than a raw data file.

Because NIR bands are mixtures of overtones and combinations, the intensity of the absorbance at particular wavelengths do not necessarily respond linearly to a change

in concentration. In the case of a mixture, band mixing may further disrupt any linear relationship between the intensity and the concentration. These are the reasons why the simple application of Beer's law $A = ebc$, where A = absorbance, e = absorptivity, b = path length, and c = concentration to NIR bands may not generate equations suitable for quantitation. Multiple regression techniques help avoid this problem. Linear calibration methods such as multiple linear regression (MLR), principal component analysis (PCA), and partial least-squares regression (PLS) are routinely used in NIR analyses. The choice of regression technique is subjective, depending on many factors. For further description of these techniques, see the introduction to multivariate calibration methods by Thomas (22). Other chemometric techniques useful in NIRS methods include BEAST bootstrap error-adjusted single-sample technique (BEAST), a nonparametric clustering algorithm used by Lodder and co-workers (23) to detect capsule tampering by NIRS. Qualitative NIR methods use pattern-recognition techniques to "train" the computer to identify an unknown material. Residual variance methods and discriminant analyses (Mahalanobis distance) were described by Mark and Tunnell (24), and were compared by Gemperline and Boyer (25). Gerhausser and Kovar (6) reported strategies for optimization of spectral libraries and compared two pattern-recognition methods to identify 117 drug substances.

Validation

Regardless of the mathematics used to interpret NIR spectra, the method must still undergo a validation process. The principal elements of ensuring linearity, accuracy, selectivity, and reproducibility of a quantitative method are required.

The validation process determines the amount of error owing to variation among the values in the population. It is used to check for the existence of a relationship between the calibration set and the validation set. Manufacturers of NIRS instrumentation include software packages that allow the operator to predict analytical results on data files that have been stored, thus allowing for validation of the calibration equation and testing for errors in the developed calibration. This enables calibration equation performance testing in terms of precision. The validity of these models depends on the ability of the calibration set to accurately represent the samples in the prediction set.

One source of prediction error is the inherent accuracy and precision of the reference analytical method used. If the reference method produces erroneous values that are consistently high or low, this bias will be reflected in

the prediction results. Other sources of prediction error relate to the reproducibility, stability, and repeatability of the NIR instrument. Reproducibility (precision) is validated by making repeated measurements of the same sample and removing it between runs. Small changes in conditions may occur owing to multiple insertions of a sample onto the instrument. Stability refers to similar changes that may occur over a longer period (hours or days). Repeatability refers to the instrument's ability to generate consistent measurements under the same conditions (without removing the sample from the instrument) over a relatively short period (seconds or minutes). All of these factors must be addressed to ensure the validity of the NIR calibration model.

Regulatory Issues

The American Society for Testing and Materials (ASTM) recently published an official document (27) providing a guide to spectroscopists for the multivariate calibration of infrared spectrometers. The scope of the publication, entitled *Standard Practices for Infrared Multivariate Quantitative Analysis* includes a description of multivariate calibration methods for the determination of physical or chemical characteristics of materials. This document is the first official standard for the application of chemometric multivariate analysis to near-infrared and infrared instruments.

Regulatory bodies around the world have given approval for NIR methods for a variety of purposes. In June 1995, the Medicines Control Agency (MCA) granted approval to Glaxo Wellcome in the United Kingdom for a NIR method for the identification and assay of Zovirax® (acyclovir) tablets. This is believed to be the first official approval for NIR granted by the MCA as an assay method for tablets.

In Norway, Wieders Farmasøytiske A/S received approval for the use of NIR as an alternative method for identification, assay, and determination of moisture content of paracetamol (acetaminophen) tablets. The Norwegian Medicines Control Authority approved the method in December 1996.

Other regulatory approvals are described in later sections of this chapter.

CURRENT METHODS OF TABLET EVALUATION

Official standards for the evaluation of tablets are given by the *U.S. Pharmacopeia* (USP) and other compendia and include uniformity of dosage units (weight variation,

content uniformity) and disintegration testing. Unofficial tests include those for mechanical strength (hardness, crushing strength) and resistance to abrasion (friability). A major disadvantage of current compendial methods of tablet evaluation is that they are time-consuming and destructive in nature. Once a test is performed on a sample, the integrity of that sample is usually lost (with the exception of weight testing) and no additional testing may be done on it. For example, a traditional quantitative analysis such as high-performance liquid chromatography (HPLC) typically calls for the tablet to be ground, followed by dissolution and dilution in a suitable medium. A significant amount of time and labor is required to run each sample in duplicate or triplicate. Other traditional methods of tablet evaluation also involve time-consuming sample preparation, such as KBr dilution for mid-infrared analysis.

The mechanical strength of a tablet plays an important role in the development and control procedures. Crushing strength is the most widely used test of mechanical strength. It is defined as the compression force that, when applied diametrically to a tablet, just fractures it (28). Tablet hardness depends on the weight of material and the space between the upper and lower punches at the moment of compression. Inconsistent hardness values are likely to result from variation in these parameters. The fundamentals of powder compression are given in a report by Leuenberger and Rohrer (29).

The Erweka hardness tester measures horizontal crushing strength by applying a load at 90° to the longest axis. This type of hardness tester is subject to two sources of inherent error: 1) the possibility of an incorrect zero, and 2) a scale that does not accurately indicate the true load applied. Other commercially used instruments include the Strong-Cobb, Monsanto, and Pfizer hardness testers. Variations in crushing strength values obtained from different types of hardness testers may be attributed to inaccuracies in instrument scale values, incorrect zero adjustment, and varying methods of applying the load. This necessitates calibration when comparing results from different types of testers. The physical dimensions and shape of the tablet may also contribute to the property of crushing strength.

The conventional methods of hardness testing for tablets also involve a subjective operator error. The scale on the Erweka hardness tester is divided into segments of 0.25-kg units. Very often, the sample under evaluation may produce a reading that falls between two divisions, and it is up to the operator to decide on the result.

Tablet-coating processes are commonly used in the pharmaceutical industry. Aqueous film coatings composed of cellulose derivatives and other polymers are useful for

the control of dissolution of drug from the tablet. Gravimetric analysis and HPLC are often used to determine the endpoint of the coating process.

ADVANTAGES OF NIRS TABLET EVALUATION

NIRS is a nondestructive method, thus, 100% inspection of batches is theoretically possible, allowing better control of product uniformity. NIR is also noninvasive, enabling subsequent evaluation of the same tablets by another method. Samples may be retained for further analysis by NIR or other methods, allowing a direct correlation between tests. Economic benefits are obvious for manufacturers, who may increase profits per batch because of the need for fewer retained samples. NIRS is particularly useful in the early stages of product development when the supply of the new drug is limited.

NIR analysis is rapid, requiring less than one minute to analyze a single sample. Also, NIR analyses do not require the use and ultimate disposal of organic solvents, thereby reducing environmental waste. Advances in instrumentation, fiber optics, and software offer many options. Portable NIRS units are not uncommon.

Accelerated degradation samples may be analyzed and returned to the associated storage condition at each of the appropriate time intervals, thus drastically reducing the number of samples taken from the batch.

Measurement of powders can take place directly through the unopened glass jar or vial because glass does not absorb in the NIR region. Direct measurement through the container further reduces sampling errors, which may be introduced when a sample is withdrawn.

Multiple components of a sample may be analyzed simultaneously. A single spectrum can be obtained and compared with several different calibration sets at the same time, allowing the measurement of several constituents at one time. This saves considerable time and labor.

LIMITATIONS OF NIRS TABLET EVALUATION

The initial calibration process for a substance or a product can be quite detailed. A calibration equation is needed for each constituent in the sample. NIR calibrations must be formulation-specific. The accuracy of the NIR method cannot be better than the reference method from which it was built. Ruggedness of the models improves when all expected types of variation are included in the model. Careful selection of representative

samples is imperative to the successful performance of the calibration. The choice of mathematical models depends highly on the character of the sample.

Another issue is that of transferability of the calibration model among instruments. This has been a significant obstacle to more widespread use of NIR methods. Transferability is especially important to multisite facilities, because it is needed to avoid time-consuming recalibration procedures. Calibration errors may occur among instruments because of slight differences in instrument response, especially if full-spectrum multivariate models are used. Shenk and Westerhaus (30) addressed the problem and proposed a standardization algorithm, which was modified by others (31, 32).

Physical attributes of the tablets can affect the calibration process. For example, scored tablets and those of differing geometries may produce more variable NIR spectra than flat, unscored tablets. Work in our laboratory (33) demonstrated a difference in NIR hardness testing of scored Avicel[®]-based chlorpheniramine tablets. Mixed calibration models composed of flat and concave tablets gave variable hardness prediction results, supporting the assertion that calibration models should contain samples of homogeneous composition.

Borer and coworkers (34) published a useful evaluation of sources of variability encountered in the NIR analysis of drug products. Analysts involved in NIRS method development should consider this report. Parameters included in the study included.

1. instrument settings (number of scans averaged per spectrum, iris opened or closed, frequency of reference spectrum collection);
2. data treatment (segment value used for second-derivative calculation); and
3. library design (total number of samples, number of days required to complete the library).

BULK DRUG PROPERTIES

Raw Material Identification

NIRS is useful for the analysis of both raw materials and finished dosage forms. Qualitative determination of pharmaceutical raw materials using NIRS was reported as early as 1982 (35). The largest variations in commercially produced excipients and actives appear to be in moisture content and particle size (36, 37). These parameters may be monitored by NIRS with relative ease. Incoming substances may be tested for immediate identity confirmation on the receiving dock by inserting a fiber

optic probe directly into the barrel. Spectral data for each lot of material purchased may be saved and added to the growing spectral reference library.

The use of NIRS as an identification method involves the establishment of a comprehensive library containing the spectra of many compounds. The sample is scanned and identified by finding the most similar item in the library. Comparison of the sample spectrum to the library reference spectrum is achieved by calculating (via computer) the cosine of the angle between the vectors of both spectra. This value is called the spectral match value (SMV) and may range from -1 to $+1$, with $+1$ being a perfect match. Plugge and Van Der Vlies (38) utilized the SMV for a NIR method of identification of ampicillin trihydrate. In their work, an SMV of 0.9980 was established as a minimum value for positive identity of ampicillin. The conformity index or CI is also calculated to establish the degree of conformity of a batch to specifications. Conformity testing detects deviations from normal processing by comparing the average batch spectra with the library of a representative collection of approved batches of the same material.

Particle Size

The measurement of particle size is a key issue in the formulation of many pharmaceutical products. Particle size distribution is known to directly influence physical properties of powders, such as dissolution rate, powder flow, bulk density, and compressibility. Conventional methods of particle size measurement include sieve analysis and laser diffractometry (39).

Ciurczak and associates (40) reported a NIR method of determination of particle size of pure, granular substances. The method is based on theories of reflected light, in which reflectance increases as the particle size decreases. The reference method was a low-angle laser light-scattering (LALS) particle sizer (Malvern). The researchers found linear results for particles above 85 μm , but less accurate results for smaller particles.

O'Neil et al. (41) described a NIRS method for measuring the median particle size of numerous compounds and excipients. Malvern data and NIR reflectance spectra of sieved fractions and bulk samples were used to construct multiple linear regression (MLR) and quadratic least-squares fit calibration models. The same group (42) reported success in measuring the cumulative percentage frequency particle size distribution of microcrystalline cellulose (Avicel) using NIRS. This study compared the use of three-wavelength MLR and principal component regression (PCR), each model using Malvern and NIR reflectance data. The PCR model

produced smaller SEPs, suggesting a more robust model than the MLR.

Frake and co-workers (43) extensively evaluated numerous chemometric techniques for the NIRS prediction of mass median particle size determination of lactose monohydrate. Models evaluated in zero order (untreated) and second derivative were MLR, PLS (partial least squares), and ANN (artificial neural network). The researchers concluded that there is more than one way to treat data and achieve a good calibration model. The group also confirms previous observations that derivitization of data does not remove "particle size effects" (previously thought to contribute to baseline shift).

Other researchers (44) have also reported NIR particle size studies.

Polymorphism and Racemization

Many drugs have the ability to form several distinct crystalline forms or polymorphs. Although the polymorphs are chemically identical, they are different arrangements of molecules and may exhibit different properties. Each polymorph may possess different energy levels and ultimately affect the dissolution rate of the compound. The infrared spectra of polymorphs may be expected to vary owing to different arrangement of functional groups (hydrogen bonding and polarization may be affected). Traditional methods for the identification of polymorphism are differential scanning calorimetry (DSC) and X-ray crystallography. However, these technique are destructive, and thus multiple runs of the same sample are not possible. Ciurczak (45) reported a NIR method of identification of the polymorphs of caffeine. Aldridge and co-workers (46) reported a NIR method of detection of polymorphism in which pattern-recognition methods were used to discriminate between the desired polymorphic form of a drug substance and another undesired polymorph. A significant outcome of this work was the successful transfer of the calibration between six other NIR instruments without the use of multivariate calibration transfer algorithms.

Gimet and Luong (47) used NIRS to determine whether the processing of a granulation resulted in racemization. Others (48) have also reported the differential identification of optically active forms.

Moisture Content

The classic methods of water determination are Karl Fischer (KF) titration, gas chromatography (GC), and loss on drying (LOD). Using near-infrared methods,

the presence of water is indicated by a major NIR absorption band at approximately 1920–1950 nm. The absorption maximum and peak shape depend on the degree of hydrogen bonding occurring within the environment where the water is located. The stronger the hydrogen bond, the longer the wavelength of the NIR absorption maxima. A second band attributed to water may appear at approximately 1450 nm and is attributed to bound water.

The NIRS determination of moisture content has been documented by numerous researchers (49–51). Plugge and Van Der Vlies (52) described a NIR method for ampicillin, which measures eight quality control criteria:

- identity,
- water content,
- crystallinity,
- ampicillin content.
- fraction of anhydrous ampicillin,
- residual reagent,
- residual organic solvents, and
- residual starting material.

DeThomas and VonBargen (53) patented a NIR fiber optic system for the measurement of moisture or "a constituent" in a powder. Berntsson, and associates (54) described a NIRS method to determine at-line moisture content in bulk hard gelatin capsules. Capsules were equilibrated at various humidities, and reference moisture content was determined by loss on drying.

Zhou and co-workers (55) at Eli Lilly and Company developed a NIRS method to determine moisture content in a freeze-dried drug product using gas chromatography (GC) as the reference method. A standard error of prediction (SEP) as low as 0.07% w/w was obtained using the NIR method in the range of 0.1–5.7% water. At very low water levels, the workers found the GC method to be more precise than the NIR method.

Corti and co-workers (56) developed a NIRS method to determine the water content of crushed ranitidine chlorhydrate tablets. KF titration was used as the reference method. Higher NIRS errors were found with the samples having a water content of less than 1%.

Dziki and coworkers (57) used NIRS to monitor the mobility of water within the sarafloxacin crystal lattice. The study involved the veterinary product sarafloxacin in an aqueous granulation process. A failed lot of the product was indistinguishable from an acceptable lot using X-ray powder diffraction, midinfrared spectroscopy, differential scanning calorimetry (DSC), and TGA. NIRS detected intermediate stages of water absorption in the granulation and enabled the process to be controlled.

An interesting report on NIR measurement of water in skin was published recently by Martin (58) at Helene

Curtis, Inc. NIRS measurements indicated four types of water contained in the skin. The workers presented tentative assignments for the absorption bands:

- lipid bilayers (1875 nm)
- secondary and primary water on protein groups (1909 and 1923 nm), and
- bulk water beneath the stratum corneum (1890 nm).

Studies were performed in vivo on two volunteers as well as in vitro using porcine skin. Although not directly related to tablet evaluation, the work is notable for its contribution to the transdermal delivery of drugs, potential NIR prediction of drug penetration through the skin, and conclusions regarding moisturizer use on the skin.

PRODUCT EVALUATION

Identification and Potency

Reports describing NIR methods of identification of solid and liquid dosage forms have increased in the literature during the 1990s (59–62). Earlier work has been reviewed elsewhere (see Introduction). Virtually all new NIR methods include product identification in the assay. Gottfries and co-workers (63) reported a NIR method for the measurement of metoprolol in controlled-release tablets. In their work, a comparison was made between the diffuse reflectance and transmission modes. The workers found better prediction of tablet strength using transmission mode and reasoned that reflectance spectra are more sensitive to the inhomogeneity of the material.

Ebube and coworkers (64) reported a NIR method that can distinguish between three Avicel® products owing to the varying particle sizes. The method was also designed to predict tablet hardness and lubricant concentration. Regardless of the small sample sizes used, good results were obtained.

North, Young and Leng et al. (65) recently presented an interesting comparison of NIR diffuse reflectance and transmittance for the analysis of tablets. Tablets used were of different drug content and size as well as coated and uncoated. Both methods produced excellent resolution of two tablet strengths. Uncoated and coated tablets produced different spectra by reflectance, but only a minimal difference was seen by transmission. Of two drug peaks seen in the reflectance spectra, only one band was observed in transmission because no transmission through the tablet occurred in the region of the second peak. Both methods produced difference in spectra as a result of difference in tablet size and surface curvature.

Merckle and Kovar (66) reported an assay of effervescent tablets (intact and powdered) by NIRS in transmittance and reflectance modes. Results of quantitative determination of acetylsalicylic acid (ASA) and ASA in combination with ascorbic acid and/or paracetamol were comparable in both transmittance and reflectance modes. Corti and associates (67) described a NIR transmittance analysis of coated tablets, using both whole and milled tablets.

Brashear and coworkers (68) developed a diffuse reflectance NIRS method to quantify lomefloxacin and polyethyleneglycol (PEG) within a polymeric implant. The properties of pore-forming excipients such as PEG are known to affect the rate of drug release from a matrix.

Sondermann and Kovar (69) described a study using NIRS for the identification of “ecstasy” street samples. Ecstasy tablets may contain either *N*-Methyl-3,4-Methylenedioxyamphetamine (MDMA) or *N*-Ethyl-3,4-Methylenedioxyamphetamine (MDE) as well as other amphetamine derivatives. In addition, various excipients were present, and nonstandardized production procedures contributed to inhomogeneous tablets. The authors researchers included a broad range of excipients in their calibration work and succeeded in constructing three PLS models for identification.

Identity testing of blister packaged tablets/supplies for clinical trials

A noninvasive NIR method was reported by Dempster and et al. (70) for the identification of tablets within blister packages. The method identified and discriminated various potencies of an experimental drug, placebo tablets, and clinical comparator tablets.

Aldridge and coworkers (71) described a NIRS method for nondestructive identification of blister-packaged tablets. The NIR method drastically reduced the assay time required by the previous TLC method. The TLC method required a full day to analyze 40 tablets, compared with the NIR method, which analyzed 10 tablets in 7 min.

Detection of adulteration

Lodder, Selby and Hieftje et al. (72) reported a NIR method of detection of capsule tampering. Several nonprescription products were selected for the study and tested with and without the addition of various adulterants. The NIR method provided a rapid, noninvasive way to screen products for known foreign substances. One limitation of the NIR method is the inability to predict which substances might be present in a product. Although the NIR spectra of many adulterants may be present in the spectral library, there is the possibility that a new,

unknown substance could be added to a product and not detected as an error in the NIR product spectrum.

Product development

In the drug-discovery field, chemists can use NIR to monitor the progress of reactions. Hearn and coworkers (73) reported a NIR method to monitor the preparation of compounds for screening as antituberculosis drugs. The reaction of isonicotinic acid hydrazide (INH) with carbonyl compounds in the preparation of Schiff bases was followed. Forbes and coworkers (74) described a NIR conformance test method to assay and identify two chemical intermediates used in the manufacture of Loracarbef, a carbacephalosporin.

Bauer, Dziki and Quick et al. (75) used NIRS to investigate the problem of dissolution failure in an erythromycin tablet formulation. The technique enabled the group to identify the presence of a dehydrated dihydrate produced during formulation. The dihydrate was found to gradually bind with magnesium hydroxide in the tablet formulation, thus delaying the process of dissolution. The use of NIR facilitated the development of a humidifying process that reversed the binding and increased the dissolution rate.

Quality Control Parameters

Tablet hardness

It has been shown (76, 77) that the NIR signal varies with a change in compression force. In other words, changes in tablet hardness result in an alteration in the NIR spectra of the sample. Presumably, increasing the compression force during the tableting process causes the tablet to be smoother as well as harder, thus causing less light scattering, leading to a greater absorbance and higher baseline. NIR hardness testing of tablets can be performed at the same time as other parameters such as identity, moisture content, and coating thickness.

Further work by Kirsch and Drennen (78) on NIR hardness testing has explored the use of various mathematical models for calibration. Other researchers have also reported NIR methods for the measurement of tablet hardness (64, 79).

Tablet coating

NIRS has been used to determine tablet-coating and core thickness. Kirsch and Drennen (80) evaluated the use of NIR at-line to monitor film coating in a Wurster column. The method was successful at predicting coating thickness of two coating formulations at various intervals during the process and was less time-consuming than wet chemical methods. Earlier work by Kirsch and Drennen (81)

described a NIR method to determine film-coating parameters of theophylline tablets. Increasing coating thickness, corresponded to increased NIR absorbance in certain regions of the spectrum. Calibration models were developed for tablet hardness, coating thickness and the prediction of time to 50% dissolution.

Buchanan and coworkers (82) at Merck reported a NIR method for evaluating a new coating-thickness manufacturing process. A precision film-coating process was tested whereby an immediate-release drug-active coating surrounded an extended-release active-drug core. The NIR method enabled the evaluation to proceed more quickly and less expensively than did the reference HPLC method. The implementation of the NIR method allowed rapid evaluation of tablets and assisted in identifying "dead zones" in the Wurster column, thus allowing immediate correction and revision of the process.

Determination of degradation products

Drennen and Lodder (83) reported a nondestructive NIRS method to monitor the decomposition of aspirin. In contrast to the multi step HPLC assay for salicylic acid and the USP identity tests for aspirin, the NIR method involved a 90-s scan of individual intact aspirin tablets. The workers correlated changes in spectra to the mass of water absorbed, the mass of salicylic acid formed, and the time the tablets spent in a hydrator.

Shimoyama and coworkers (84) reported a NIR analysis of photodegradation of poly(methyl methacrylate) using an in situ fiber optic device. This type of technology from a related discipline is notable as a potential application for pharmaceutical systems.

Characterization of powder blends and blend homogeneity

Powder blending is a fundamental step in the process of manufacturing pharmaceutical products. Only a homogeneous mixture can be properly subdivided to provide uniform doses of the active ingredients. The current procedures for monitoring blend uniformity require that the blending process be stopped at defined intervals to obtain samples of the blend. The samples are collected from different locations in the blending vessel using a sample thief. The samples are then sent to a laboratory and analyzed using traditional methods such as HPLC or UV until the active components are within specification for that formulation. This approach requires a significant amount of time and labor and may be subject to errors induced by sampling methods.

NIRS has been shown by several researchers to be useful for evaluating the powder-mixing process. Wargo

and Drennen (85) demonstrated the use of a NIR method to determine homogeneity of powder blends.

In 1995, the European Patent Office granted a patent to Dr. Paul K. Aldridge, Pfizer Central Research, Groton, CT, for an apparatus (86) for mixing and detecting on-line homogeneity. The apparatus involves the use of a diffuse-reflectance fiber optic probe interfaced on-line with a V-blender. Sekulic and co-workers (87) at Pfizer described the use of this apparatus for on-line monitoring of powder blend homogeneity. An 8-quart twin-shell V-blender was interfaced with a fiber optic probe at the axis of rotation. Spectra were collected at prescribed intervals, and data analysis was performed using a series of commercial software. Variability in the NIR spectra as a function of time was measured, and it was shown that this variability reached a minimum level sooner than what traditional blending times suggest. DeMaesschalck et al. (88) used the NIR on-line method described in article by the Sekulic and associates to design an approach for deciding when the blend is homogeneous. They calculated the average standard deviation between spectra taken at each time and used the dissimilarity between each new measurement and the ideal mixture spectrum to monitor changes in the mixture during the blending process.

Scientists at Merck were issued a U.S. patent in 1996 for a method (89) of measuring the homogeneity of tablets using NIR. It can be used to monitor the pharmaceutical material during the tableting process (powder mix, granular mix, and compressed tablets).

CONCLUSIONS

NIRS has proven to be a fast, reliable, and cost-saving method for numerous applications in the pharmaceutical industry. It is no longer the esoteric method it was once believed to be. The pharmaceutical industry has learned a great deal about NIRS from the agricultural and food industries. Concepts and techniques have been borrowed and fitted to the needs of pharmaceutical scientists. Users in all disciplines face common issue, such as calibration transfer, moisture contamination, particle size, and the rigors of calibrating multiple constituents.

NIRS possesses a great and, as yet, incompletely exploited potential in the area of identity testing of drug substances. It has already begun to replace traditional compendial methods of quality control. It has gained recognition from the FDA and other regulatory agencies, a signal to skeptics that NIRS is a solid alternative to traditional methods of analysis. Aggressive workers in the field are moving to develop and receive approval for NIR

methods that bypass the traditional reference methods. A greater understanding of the mathematics involved with NIR analyses has contributed to the wider use of NIR methods. Small companies that wish to use NIR analyses may find it wise to contract out their work to groups with more expertise because initial startup can be expensive and initial calibration work may be time-consuming.

We confidently predict that NIRS will rapidly become an established and standard method for many types of pharmaceutical analyses.

REFERENCES

1. Osborne, B.G.; Fearn, T. *Near Infrared Spectroscopy in Food Analysis*; Longman Scientific & Technical: United Kingdom, 1986; 2.
2. Morisseau, K.M.; Rhodes, C.T. Proven and Potential Uses of Near-Infrared Spectroscopy for the Evaluation of Tablets; Tableting and Granulation Year Book. *Pharm. Tech. Suppl.* 6–11.
3. Morisseau, K.M.; Rhodes, C.T. Pharmaceutical Uses of Near-Infrared Spectroscopy. *Drug Dev. Ind. Pharm.* **1995**, *21*, 1071–1090.
4. Corti, P.; Dreassi, E. Near Infrared Reflectance Analysis: Features and Applications in Pharmaceutical and Biomedical Analysis. *Farmaco.* **1993**, *48* (1), 3–20.
5. MacDonald, B.F.; Prebble, K.A. Some Applications of Near-Infrared Reflectance Analysis in the Pharmaceutical Industry. *J. Pharm. Biomed. Anal.* **1993**, *11*, 1077–1085.
6. Axon, T.G.; Brown, R.; Hammond, S.V.; Maris, S.J. Focusing Near Infrared Spectroscopy on the Business Objectives of Modern Pharmaceutical Production. *J. Near Infrared Spectrosc.* **1998**, *6*, A13–A19.
7. Workman, J.J., Jr.; Review of Process and Non-Invasive Near-Infrared and Infrared Spectroscopy. *Appl. Spectrosc. Reviews*, May 1 **1999**, *34*, 1–2, 1–89.
8. Ciurczak, E.W.; Drennen, J.K. *Near Infrared Spectroscopy in Pharmaceutical and Medical Applications, Practical Spectrosc. Series*; Gadamasetti, K.G., Ed.; Marcel Dekker, Inc.: New York, in press.
9. Rose, J.J.; Prisick, T.; Mindakia, J. J. Parent. Sci. Tech. **1982**, *26*, 71–78.
10. Ciurczak, E.W.; Torlini, R.P. Analysis of Solid and Liquid Dosage Forms Using Near-Infrared Reflectance Spectroscopy. *Spectrosc.* **1987**, *2* (3), 41–43.
11. Sinsheimer, J.E.; Keuhnelian. *Am J. Pharm. Sci.* **1996**, *55*, 1240.
12. Massart, D.L.; Vandeginste, B.G.M.; Deming, S.N.; Michotte, Y.; Kaufman, L. Data Handling in Science and Technology. *Chemometrics: A Textbook*; Vandeginste, B.G.M., Rutan, S.C., Eds.; Elsevier Publishing: New York, 1988; 2, 5.
13. Willard, H.H.; et al. Infrared Spectroscopy. *Instrumental Methods of Analysis*, 7th Ed.; Wadsworth Publishing Company: Belmont, 1988; 287.

14. Workman, J.J., Jr.; Burns, D.A. Commercial NIR Instrumentation. *Handbook of Near-Infrared Analysis*; Burns, D.A., Ciurczak, E.W., Eds.; Marcel Dekker, Inc.: New York, 1992; 37–51.
15. Lodder, R.A. Sample Holders or Reflectors for Intact Capsules and Tablets and for Liquid Microcells for Use in Near-Infrared Reflectance Spectrophotometers, US Patent 165,751, November 21, 1989.
16. Drennen, J.K. Near-Infrared Reflectance Spectrometer System and Related Sample Cell and Sample Support, US Patent 898,454, May 25, 1993.
17. Soloman, S. Non-Destructive Identification of Tablet and Tablet Dissolution by Means of Infrared Spectroscopy, US Patent 338,909, October 21, 1997.
18. Noble, D. Illuminating Near-IR. *Anal. Chem.* **1995**, *67* (23), 735A–740A.
19. — http://kerouac.pharm.uky.edu/asrg/cnirs/ir_spec.htm (accessed May 2000).
20. — <http://leden.tref.nl/~mderksen/patents.html> (accessed May 2000).
21. Ciurczak, E.W. Pharmaceutical Mixing Studies Using Near-Infrared Spectroscopy, Tableting and Granulation Yearbook; Pharm. Tech. Supp. **1997**, 18–28.
22. Thomas, E.V. A Primer on Multivariate Calibration. *Anal. Chem.* **1994**, *66* (15), 795A–804A.
23. Lodder, R.A.; Selby, M.; Hieftje, G.M. Detection of Capsule Tampering by Near-Infrared Reflectance Analysis. *Anal. Chem.* **1987**, *59*, 1921–1930.
24. Mark, H.L.; Tunnell, D. Qualitative Near-Infrared Reflectance Analysis Using Mahalanobis Distances. *Anal. Chem.* **1985**, *57*, 1449–1456.
25. Gemperline, P.J.; Boyer, N.R. Classification of Near-Infrared Spectra Using Wavelength Distances: Comparison to the Mahalanobis Distance and Residual Variance Methods. *Anal. Chem.* **1995**, *67*, 160–16.
26. Gerhauser, C.I.; Kovar, K.A. Strategies for Constructing Near-Infrared Spectral Libraries for the Identification of Drug Substances. *Appl. Spectrosc.* **1998**, *51*, 1504–1510.
27. American Society for Testing and Materials (A.S.T.M). *Standard Practices for Infrared Multivariate Quantitative Analysis*; Practice E1655-00, A.S.T.M: West Conshocken, Pennsylvania, 1995; 1–25.
28. Marshall, K.; Rudnic, E.M. Tablet Dosage Forms. *Modern Pharmaceuticals*, 2 edn; Banker, G.S.; Rhodes, C.T., Eds.; Marcel Dekker, Inc.: New York, 1990; 355–425.
29. Leuenberger, H.; Rohera, B.D. Fundamentals of Powder Compression. *Pharm. Res.* **1986**, *3* (1), 12–22.
30. Shenk, J.S.; Westerhaus, M.O. *Crop Sci.* **1991**, *31*, 1694–1696.
31. Bouveresse, E.; Massart, D.L. Modified Algorithm for Standardization of Near-Infrared Spectrophotometric Instruments. *Anal. Chem.* **1995**, *67*, 1381–1389.
32. Wang, Y.; Kowalski, B.R. *Appl. Spectrosc.* **1992**, *46*, 764–771.
33. Morisseau, K.M. *The Effect of Compression Force on the Near-Infrared Spectra of Tablet Dosage Forms*; University of Rhode Island, 1996; 130–242, Doctoral dissertation.
34. Borer, M.W.; Zhou, X.; Hays, D.M.; Hofer, J.D.; White, K.C. Evaluation of Key Sources of Variability in the Measurement of Pharmaceutical Drug Products by Near Infrared Reflectance Spectroscopy. *J. Pharm. Biomed. Anal.* **1998**, *17*, 641–650.
35. Rose, J.R. Quantitative and Qualitative Analysis with NIRA. Proceeding of the 2nd Annual Symposium on NIRA, Technicon, Tarrytown, New York, July, 1982.
36. Ciurczak, E.W. NIR Analysis of Pharmaceuticals. *Handbook of Near-Infrared Analysis*; Burns, D.A., Ciurczak, E.W., Eds.; Marcel Dekker, Inc.: New York, 1992; 549–563.
37. Burger, T.; Fricke, J. NIR Radiative Transfer Investigations to Characterise Pharmaceutical Powders and Their Mixtures. *J. Near Infrared Spectrosc.* **1998**, *6*, 33–40.
38. Plugge, W.; Van Der Vlies, C. Near-Infrared Spectroscopy as an Alternative to Assess Compliance of Ampicillin Trihydrate with Compendial Specifications. *J. Pharm. Biomed. Anal.* **1993**, *11* (6), 435–442.
39. Aulton, M.E. *Pharmaceutics: The Science of Dosage Form Design*. Churchill Livingstone: Edinburgh, 1988.
40. Ciurczak, E.W.; Torlini, R.P.; Demkowicz, M.P. Determination of Particle Size of Pharmaceutical Raw Materials Using Near-Infrared Reflectance Analysis. *Spectroscopy* **1986**, *1*, 36–39.
41. O'Neil, A.J.; Jee, R.D.; Moffat, A.C. The Application of Multiple Linear Regression to the Measurement of the Median Particle Size of Drugs and Pharmaceutical Excipients by Near-Infrared Spectroscopy. *Analyst* **1998**, *123*, 2297–2302.
42. O'Neil, A.J.; Jee, R.D.; Moffat, A.C. Measurement of the Cumulative Particle Size Distribution of Microcrystalline Cellulose Using Near Infrared Reflectance Spectroscopy. *Analyst* **1999**, *124*, 33–36.
43. Frake, P.; Gill, I.; Luscombe, C.N.; Rudd, D.R.; Waterhouse, J.; Jayasooriya, U.A. Near-Infrared Mass Median Particle Size Determination of Lactose Monohydrate, Evaluating Several Chemometric Approaches. *Analyst* **1998**, *123*, 2043–2046.
44. Dredan, J.; Zelkò, R.; Antal, I.; Bihari, E.; Ràcz, I. Effect of Particle Size and Coating Level on the Diffuse Reflectance of Wax Matrices. *J. Pharm. Pharmacol.* **1998**, *50*, 139–142.
45. Ciurczak, E.W. In Proceedings of FACSS, Philadelphia, 1985.
46. Aldridge, P.K.; Evans, C.L.; Ward, H.W., II; Colgan, S.T. Near-IR Detection of Polymorphism and Process-Related Substances. *Anal. Chem.* **1996**, *68*, 997–1002.
47. Gimet, R.; Luong, A.T. *J. Pharm. Biomed. Anal.* **1987**, *5*, 205.
48. Buchanan, B.R.; Ciurczak, E.W.; Grunke, A.; Honigs, D.E. *Spectroscopy* **1988**, *3*, 54.
49. Kamat, M.; Lodder, R.A.; DeLuca, P.P. Near-Infrared Spectroscopic Determination of Residual Moisture in Lyophilized Sucrose Through Intact Glass Vials. *Pharm. Res.* **1989**, *6*, 961–965.
50. Sinsheimer, J.E.; Poswalk, N.M. Pharmaceutical Applications of the Near-Infrared Determination of Water. *J. Pharm. Sci.* **1968**, *57*, 2007–2010.

51. Zhou, X.; Hines, P.; Borer, M.W. Moisture Determination in Hygroscopic Drug Substance by Near Infrared Spectroscopy. *J. Pharm. Biomed. Anal.* **1998**, *17* (2), 219–225.
52. Plugge, W.; Van Der Vlies, C. The Use of Near Infrared Spectroscopy in the Quality Control Laboratory of the Pharmaceutical Industry. *C. J. Pharm. Biomed. Anal.* **1992**, *10*, 797–803.
53. DeThomas, F.A.; VonBargen, K.P. System for Measuring the Moisture Content of Powder and Fiber Optic Probe Therefore, US Patent 931,783, January 11, 1994.
54. Berntsson, O.; Zackrisson, G.; Ostling, G. Determination of Moisture in Hard Gelatin Capsules Using Near-Infrared Spectroscopy: Applications to At-Line Process Control of Pharmaceuticals. *J. Pharm. Biomed. Anal.* **1997**, *15*, 895–900.
55. Zhou, X.; Hines, P.A.; White, K.C.; Borer, M.W. Gas Chromatography as a Reference Method for Moisture Determination by Near Infrared Spectroscopy. *Anal. Chem.* **1998**, *70*, 390–394.
56. Corti, P.; Dreassi, E.; Corbini, G.; Lonardi, S.; Viviani, E.; Mosconi, L.; Bernuzzi, M. Application of Near Infrared Reflectance to the Analytical Control of Pharmaceuticals: Assay of Ranitidine Chlorhydrate and Water Content in Tablets. *Pharm. Acta Helv.* **1990**, *65*, 28–32.
57. Dziki, W.; Bauer, J.F.; Szpylman, J.J.; Quick, J.E.; Nichols, B.C. The Use of Near-Infrared Spectroscopy to Monitor the Mobility of Water Within the Sarafloxacin Crystal Lattice. *J. Pharm. Biomed. Anal.* **2000**, *22*, 829–848.
58. Martin, K. In Vivo Measurements of Water in Skin by Near Infrared Reflectance. *Appl. Spectrosc.* **1998**, *52*, 1001–1007.
59. Bradfield, K.B.; Forbes, R.A. Development and Validation of an Analytical Method for Identification of Granulated Nicarbazin by Near Infrared Reflectance Spectroscopy. *J. Near Infrared Spectrosc.* **1997**, *5*, 41–65.
60. Jedvert, I.; Josefson, M.; Langkilde, F. Quantification of an Active Substance in a Tablet by NIR and Raman Spectroscopy. *J. Near Infrared Spectrosc.* **1998**, *6*, 279–289.
61. Han, S.M.; Faulkner, P.G. Determination of SB-216469-S During Tablet Production Using Near-Infrared Reflectance Spectroscopy. *J. Pharm. Biomed. Anal.* **1996**, *14*, 1681–1689.
62. Eustaquio, A.; Graham, P.; Jee, R.D.; Moffatt, A.C.; Trafford, A.D. Quantification of Paracetamol in Intact Tablets Using Near-Infrared Transmittance Spectroscopy. *Analyst* **1998**, *123* (11), 2303–2306.
63. Gottfries, J.; Depui, H.; Fransson, M.; Jongeneelen, M.; Josefson, M.; Langkilde, F.W.; Witte, D.T. Vibrational Spectrometry for the Assessment of Active Substance in Metoprolol Tablets: A Comparison Between Transmission and Diffuse Reflectance Near-Infrared Spectrometry. *J. Pharm. Biomed. Anal.* **1996**, *14*, 1495–1503.
64. Ebube, N.K.; Thosar, S.S.; Roberts, R.A.; Kemper, M.S.; Rubinovitz, R.; Martin, D.L.; Reier, G.E.; Wheatley, T.A.; Shukla, A.J. Application of Near-Infrared Spectroscopy for Nondestructive Analysis of Avicel Powders and Tablets. *Pharm. Dev. Tech.* **1999**, *4*, 19–26.
65. North, N.; Young, K.; Leng, M. A Comparison Between NIR Diffuse Reflectance and Transmittance for the Analysis of Pharmaceutical Tablets, 8th FOSS/NIRSystems European Pharmaceutical User Group Meeting, SmithKline Beecham, Harlow, UK, October 1–8, 1997.
66. Merckle, P.; Kovar, K.A. Assay of Effervescent Tablets by Near-Infrared Spectroscopy in Transmittance and Reflectance Mode. *J. Pharm. Biomed. Anal.* **1998**, *17* (3), 365–374.
67. Corti, P.; Ceramelli, G.; Dreassi, E.; Mattii, S. Near Infrared Transmittance Analysis for the Assay of Solid Pharmaceutical Dosage Forms. *Analyst* **1999**, *124*, 755–758.
68. Brashear, R.L.; Flanagan, D.; Luner, P.E.; Seyer, J.J.; Kemper, M.S. Diffuse Reflectance Near-Infrared Spectroscopy as a Nondestructive Analytical Technique for Polymer Implants. *J. Pharm. Sci.* **1999**, *88* (12), 1348–1353.
69. Sondermann, N.; Kovar, K.A. Screening Experiments of Ecstasy Street Samples Using Near Infrared Spectroscopy. *For. Sci. Int.* **1999**, *106*, 147–156.
70. Dempster, M.A.; Jones, J.A.; Last, I.R.; MacDonald, B.F. Near-Infrared Methods for the Identification of Tablets in Clinical Trial Supplies. *J. Pharm. Biomed. Anal.* **1993**, *11*, 1087–1092.
71. Aldridge, P.K.; Mushinsky, R.F.; Andino, M.M.; Evans, C.L. Identification of Tablet Formulations Inside Blister Packages by Near-Infrared Spectroscopy. *Appl. Spectrosc.* **1994**, *48*, 1272–76.
72. Lodder, R.A.; Selby, M.; Hieftje, G.M. Detection of Capsule Tampering by Near-Infrared Reflectance Analysis. *Anal. Chem.* **1987**, *59*, 1921–1930.
73. Hearn, M.J.; Celi, P.; Chanyaputhipong, P.Y.; Chi, W.; Kang, J.O.; Katz, A.; Shah, R.; Thai, M.; Ung, P. Using Near Infrared Spectroscopy to Monitor the Preparation of Compounds for Screening as Antituberculosis Drugs. *J. Near Infrared Spectr.* **1995**, *3*, 19–23.
74. Forbes, R.A.; Persinger, M.L.; Smith, D.R. Development and Validation of Analytical Methodology for Near-Infrared Conformance Testing of Pharmaceutical Intermediates. *J. Pharm. Biomed. Anal.* **1996**, *15*, 315–327.
75. Bauer, J.F.; Dziki, W.; Quick, J.E. Role of an Isomorphic Desolvate in Dissolution Failures of an Erythromycin Tablet Formulation. *J. Pharm. Sci.* **1999**, *88* (11), 1222–1227.
76. Kirsch, J.D.; Drennen, J.K. Determination of Film-Coated Tablet Parameters by Near-Infrared Spectroscopy. *J. Pharm. Biomed. Anal.* **1995**, *13*, 1273–1281.
77. Morisseau, K.M.; Rhodes, C.T. Near-Infrared Spectroscopy as a Nondestructive Alternative to Conventional Tablet Hardness Testing. *Pharm. Res.* **1997**, *14* (1), 108–111.
78. Kirsch, J.D.; Drennen, J.K. Nondestructive Tablet Hardness Testing by Near-Infrared Spectroscopy: A New and Robust Spectral Best-Fit Algorithm. *J. Pharm. Biomed. Anal.* **1999**, *19*, 351–362.
79. Guo, J.H.; Skinner, G.W.; Harcum, W.W.; Malone, J.P.; Weyer, L.G. Application of Near-Infrared Spectroscopy in the Pharmaceutical Solid Dosage Form. *Drug Dev. Indust. Pharm.* **1999**, *25* (12), 1267–1270.

80. Kirsch, J.D.; Drennen, J.K. Near-Infrared Spectroscopic Monitoring of the Film Coating Process. *Pharm. Res.* **1996**, *13*, 234–237.
81. Kirsch, J.D.; Drennen, J.K. Determination of Film-Coated Tablet Parameters by NIR Spectroscopy. *J. Pharm. Biomed. Anal.* **1995**, *13*, 1273–1281.
82. Buchanan, B.R.; Baxter, M.A.; Chen, T.S.; Qin, X.Z.; Robinson, P.A. Use of Near-Infrared Spectroscopy to Evaluate an Active in a Film Coated Tablet. *Pharm. Res.* **1996**, *13*, 616–621, 1996.
83. Drennen, J.K.; Lodder, R.A. Nondestructive Near-Infrared Analysis of Intact Tablets for Determination of Degradation Products. *J. Pharm. Sci.* **1990**, *79*, 622–627.
84. Shimoyama, M.; Matsukawa, K.; Inoue, H.; Ninomiyaa, T.; Ozaki, Y. Non-Destructive Analysis of Photo-Degradation of Poly(Methyl Methacrylate) by Near Infrared Light-Fibre Spectroscopy and Chemometrics. *J. Near Infrared Spectrosc.* **1999**, *7*, 27–32.
85. Wargo, D.J.; Drennen, J.K. Near-Infrared Spectroscopic Characterization of Pharmaceutical Powder Blends. *J. Pharm. Biomed. Anal.* **1996**, *14* (8), 1415–1423.
86. Aldridge, P.K. Pfizer, Inc. Apparatus for Mixing and Detecting On-Line Homogeneity. European Patent Publication **1993**, *A1*, 631–810.
87. Sekulic, S.S.; Ward, H.W.; Brannegan, D.R.; Stanley, E.D.; Evans, C.L.; Sciavolino, S.T.; Hailey, P.A.; Aldridge, P.K. On-Line Monitoring of Powder Blend Homogeneity by Near-Infrared Spectroscopy. *Anal. Chem.* **1996**, *68* (3), 509–513.
88. DeMaesschalck, R.; Sanchez, F.C.; Massart, D.L.; Doherty, P.; Hailey, P.A. On-Line Monitoring of Powder Blending with Near-Infrared Spectroscopy. *Appl. Spectrosc.* **1998**, *52* (5), 725–731.
89. Richmond, E.W.; Buchanan, B.R.; Baxter, M.A.; Duff, A.; Tully, O.M.; Thornton, S.A. Method and System for Determining the Homogeneity of Tablets, US Patent 296,833; April 2, 1996.
90. Williams, P., Norris, K., Eds. *Near-Infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists: St. Paul, MN, 1987.
91. Burns, D.A., Ciurczak, E.W., Eds. *Handbook of Near-Infrared Analysis*, Marcel Dekker, Inc.: New York, 1992.

TABLET FORMULATION

Larry L. Augsburger

University of Maryland, Baltimore, Maryland, U.S.A.

Mark J. Zellhofer

University Pharmaceuticals of Maryland, Inc., Baltimore, Maryland, U.S.A.

INTRODUCTION OBJECTIVES OF TABLET FORMULATION

The best new therapeutic entity in the world is of little value without an appropriate delivery system. Tableted drug delivery systems can range from relatively simple immediate-release formulations to complex extended- or modified-release dosage forms. The most important role of a drug delivery system is to get the drug “delivered” to the site of action in sufficient amount and at the appropriate rate; however, it must also meet a number of other essential criteria. These include physical and chemical stability, ability to be economically mass produced in a manner that assures the proper amount of drug in each and every dosage unit and in each batch produced, and, as far as possible, patient acceptability (for example, reasonable size and shape, taste, color, etc. to encourage patients to take the drug and thus comply with the prescribed dosing regimen).

The discovery of new therapeutic entities always initiates excitement, but the contributions of the formulation specialist are either not well understood or are often taken for granted and thus remain “unsung.” However, the drug and its delivery system cannot be separated. The general design criteria for tablets are given as follows

1. Optimal drug dissolution and, hence, availability from the dosage form for absorption consistent with intended use (i.e., immediate or extended release).
2. Accuracy and uniformity of drug content.
3. Stability, including the stability of the drug substance, the overall tablet formulation, disintegration, and the rate and extent of drug dissolution from the tablet for an extended period.
4. Patient acceptability. As much as possible, the finished product should have an attractive appearance, including color, size, taste, etc., as applicable, in order to maximize patient acceptability and encourage compliance with the prescribed dosing regimen.

5. Manufacturability. The formulation design should allow for the efficient, cost-effective, practical production of the required batches.

That tablets can be formulated to uniquely meet these criteria accounts for their emergence as the most prevalent oral solid dosage form. Although several different types of tablets may be distinguished, they are mostly made by compression, intended to be swallowed whole and designed for immediate release. This paper presents a systematic approach to the design and formulation of immediate-release compressed tablets.

MODERN TABLET FORMULATION DESIGN AND MANUFACTURE

Tablet dosage forms have to satisfy a unique design compromise. The desired properties of rapid or controlled disintegration and dissolution of the primary constituent particles must be balanced with the manufacturability and esthetics of a solid compact resistant to mechanical attrition.

Excipients are critical to the design of the delivery system and play a major role in determining its quality and performance (1). They may be selected to enhance stability (antioxidants, UV absorbers), optimize or modify drug release (disintegrants, hydrophilic polymers, wetting agents, biodegradable polymers), provide essential manufacturing technology functions (binders, glidants, lubricants), enhance patient acceptance (flavors), or aid in product identification (colorants). Thus a tablet formulation is not a random combination of ingredients, but rather a carefully thought out, rational formulation designed to satisfy the above criteria.

A long list of possible excipients is available to the formulation scientist, but certain external factors such as cost, functional reliability, availability, and international acceptance govern their selection. For example, although the official compendia provide standards for identity and purity of excipients, monographs may not provide tests to

assure their functionality. For instance, the NF monograph for Compressible Sugar provides no test for compressibility. The monograph for Lactose USP does not address the many particle size and tableting grades meeting monograph standards. The NF monograph for Pregelatinized Starch refers to grades that are “compressible and flowable in character,” but provides no specifications or tests for these properties. Nor do the monograph tests for disintegrants and lubricants necessarily relate to their functionality. The need to provide functionality tests or tests for properties clearly related to functionality may be as important as controlling identity and purity (2). This point has been made even more apparent in recent years with the emergence of multiple sources of such modern excipients as direct-compression filler-binders and the various classes of “super” disintegrants.

A major problem currently being faced by multinational firms and others who market in the international arena, is the lack of universal acceptability of excipients in different countries. The selection of excipients for international markets is often a compromise between functional efficacy, local restrictions, and cost and availability in the countries where the product is to be made. In recent years, the globalization of the pharmaceutical industry has brought about an intense interest in developing harmonized pharmacopeial excipient standards, Good Manufacturing Practices (GMP) for excipient manufacture, and safety evaluation guidelines for new excipients to eliminate or avoid trade barriers between different countries (3). The International Pharmaceuticals Excipients Council (IPEC), which consists of producers, users, and pharmaceutical scientists, was launched in 1991 to assist regulatory authorities in the United States, Japan, and Europe with harmonization. The separate organizations later formed in the United States (IPEC-Americas), Europe (IPEC-Europe), and Japan (JPEC) are now known as TriPEC and include, as of 1993, more than 100 excipient and pharmaceutical firms (3).

PREFORMULATION

The objective of preformulation studies is to develop a portfolio of information about the drug substance to serve as a set of parameters against which detailed formulation design can be carried out. Preformulation investigations are designed to identify those physicochemical properties of drug substances and excipients that may influence the formulation design, method of manufacture, and pharmacokinetic-biopharmaceutical properties of the resulting product.

Following is a generalized preformulation protocol appropriate for tablet dosage forms. For certain tests, it is assumed that the drug substance is multisourced (a previously new chemical entity whose patent has expired and which is available to the generic market) for which a USP monograph exists.

Identity and Purity

The study of any drug substance must start with the determination of identity and purity. Such tests are necessary to identify degradants and contaminants and may include organoleptic tests for color, odor, and taste. Purity tests can be found in the USP for almost all marketed compounds. Alternative methods can be employed only if they are validated against the USP procedure. Tests other than potency, which can help to identify or determine the purity of compounds, are melting point, specific rotation, pH, heavy metals, residue on ignition, etc. Impurities can occasionally affect stability, and metal contamination can catalyze chemical reactions. Impurities can also alter the color of drug substances. Techniques can be utilized to give a quantitative estimate of impurities such as the impurity index (II) and the homogeneity index (HI). An ordinary impurity test can be found in the USP that estimates impurities by thin-layer chromatography (TLC).

Crystal Properties and Polymorphism

Many drug substances appear in more than one polymorphic form. The form is determined by certain conditions during the crystallization step. Occasionally drug substances are precipitated in such a way that molecules do not organize themselves in any set pattern, resulting in an amorphous powder. It is also possible for solids to entrap solvents stoichiometrically to form solvates.

Even though they are chemically identical, the different polymorphic forms of a compound are associated with different free energies, and, therefore, have different physical properties that can impact significantly on product performance (4). These include differences in solubility and dissolution rate (affecting bioavailability), solid-state stability (affecting potency), deformation characteristics (affecting compactibility), and particle size and shape (affecting powder density and flow properties). The form with the lowest energy is more stable than the others. Although the other polymorphs are thus energetically unfavored, if kept dry, they may persist indefinitely and are called “metastable.” A metastable

form may be preferred, particularly for its ability to dissolve more rapidly.

Polymorphic transformation can take place during pharmaceutical processing, such as particle size reduction, wet granulation, drying, and even during the compaction process (5). Tests employed to determine crystal properties include differential thermal analysis (DTA), differential scanning calorimetry (DSC), and X-ray diffraction (4). See also the article Thermal Analysis of Drugs and Drug Products by D. Giron in this encyclopedia.

Particle Size, Shape, and Surface Area

Probably no characteristics of a drug substance are more important than particle characteristics in determining its performance in a formulation. This is particularly true in those cases where the drug is a poorly soluble nonelectrolyte or a free acid form with poor solubility at low pH values. Such drugs are likely to exhibit dissolution-rate-limited absorption, and if dissolution does not take place rapidly enough, a therapeutic concentration in the body fluids may never be achieved, the peak plasma concentration may be significantly delayed, or much of the drug may bypass that region of the gastrointestinal (GI) tract where absorption is best. Particle size reduction (e.g., micronization) is often utilized to enhance dissolution rate. Small particles present a larger surface area per unit weight to the dissolution media and hence dissolve more rapidly than large particles. Particle size and surface area are two of the most important properties determining the solubility rate of a drug and thus potentially its bioavailability. There are numerous examples of bioavailability problems and bioinequivalence due to the inappropriate particle size of the drug substance.

Particle size and shape also play an extremely important role in the homogeneity of powder blends and the unblending of powders in a mixer. Segregation in handling or during the compaction process has a significant effect on the content uniformity of the finished products. Particle size can also affect the stability of a drug substance in that it governs the surface area available for oxidation and hydrolysis. Surface area is critical for interaction with excipients in tablet dosage forms and can greatly affect stability. Methods to determine particle size and shape include light microscopy, scanning electron microscopy, sieve analysis, and various electronic sensing-zone particle counters. Methods available for surface area measurement include air permeability and various gas adsorption techniques.

Bulk Powder Properties

Bulk powder properties are extremely important in pharmaceutical processing (6). Knowledge of the true and bulk densities of the drug substance as well as of the excipients is extremely useful in

- Providing perspective as to the size of the final tablet and the size and type of processing equipment needed,
- Anticipating problems in the physical mixing of powders and the homogeneity of intermediate and final products because significant differences in true densities can result in segregation,
- Anticipating problems in flow properties, since that property is affected by density, and
- Identifying differences in different lots and raw materials from different suppliers because different polymorphic forms can be expected to exhibit different true densities.

A comparison of true particle density, apparent particle density, and bulk density can provide information on total porosity, interparticle porosity, and intraparticle porosity. Methods include true particle density measurements via helium pycnometry, mercury intrusion porosimetry, and poured and tapped bulk density.

The influence of sorbed moisture on chemical stability and the flow and compaction of powders and granulations is well established. The moisture content and hygroscopicity of excipients is particularly important as total product processing as well as finished product stability can be affected. Hygroscopicity, moisture-sorption isotherms, and equilibrium moisture content can be determined by thermogravimetric analysis and Karl Fisher titration methods.

The compactibility of relatively large-dose drug substances and formulations is another important property. Compactibility is of less concern for smaller-dose drugs for which direct compression fillers may be able to compensate for a lack of ability to form mechanically strong compacts. An instrumented tablet press (7) or compaction simulator (8) may be used to assess the relationship between the mechanical strength of the compact and the force (or pressure) employed to form the tablet. This relationship is the easiest of all compaction measurements to establish and provides important information on the ability of the material to form practical compacts. Measures of compact mechanical strength include hardness (or crushing force), tensile strength, and friability. Other more complex studies, more easily and perhaps best done using a compaction simulator, include measurement of the work or energy of compaction, pressure–density (Athy–Heckel) analysis, strain-rate sensitivity, and elastic recovery (9).

The Athy–Heckel analysis can provide information on deformation mechanism and give an estimate of the mean yield pressure of the material (10). A comparison of yield pressures determined at different punch speeds can give information on the strain-rate sensitivity of the material (11). If the major components of the formulation (including the drug) are strain-rate sensitive, the tablets produced on a high-speed production press may exhibit lamination or capping. Excessive elastic recovery may also indicate such tablet failure. The Hiestand indices (bonding and brittle fracture) may be used to assess the compactibility of materials under laboratory conditions (12).

For the evaluation of flow properties the following test methods may be used:

- Angle of repose
- Minimum orifice diameter,
- Carr index,
- Flow rate, and
- Direct observation of weight variation during tableting runs.

The ultimate goal of flow analysis is to identify the powder or powder blend that provides the least weight variation in the finished tablet. The more fluid the powder is, the more efficiently and reproducibly it should fill the die cavities of a tablet press. This more efficient and reproducible die fill should be reflected in increased tablet weights and reduced intertablet weight variation (13).

Solubility and Permeability

In many cases, the rate of dissolution in gastrointestinal fluids is the rate-limiting step in absorption. The bioequivalence requirements established by the FDA define low solubility as "... <5 mg/mL in water, and slow dissolution rate to be <50% in 30 minutes" (14). However, the solubility of a drug should be considered together with its dose; that is, even a very poorly soluble drug having a sufficiently small therapeutic dose may completely dissolve under physiological conditions. Thus, Amidon et al. (15) have defined a "high solubility" drug as one which at the highest human dose is soluble in 250 ml (or less) water throughout the physiological pH range (1–8) at 37°C. A "low solubility" drug is thus one which requires more than 250 ml of water to dissolve the largest human dose at any pH within the physiological range. The likelihood of having bioavailability problems requires both a consideration of the dose and a solubility volume of the drug and its permeability. Amidon et al. (15) created a Biopharmaceutics Drug Classification System (BCS) based on estimates of these two parameters:

1. Class I: High solubility and high permeability
2. Class II: Low solubility and high permeability
3. Class III: High solubility and low permeability
4. Class IV: Low solubility and low permeability

A jejunal permeability of at least $2-4 \times 10^{-4}$ cm/s, measured in humans by an intubation technique, is considered "high permeability." For many substances, this permeability corresponds to a fraction absorbed of 90% or better. The classification system provides a logical basis for estimating the risk of bioavailability problems. For example, Class I drugs (e.g., propranolol HCl, metoprolol tartrate) are expected to exhibit few bioavailability problems. On the other hand, Class II drugs (e.g., piroxicam) are more likely to exhibit dissolution-rate-limited absorption problems. Class III drugs (e.g., atenolol) are more likely to be prone to absorption (permeability) rate-limited absorption. Class IV drugs (low solubility–low permeability) present formidable obstacles to bioavailability. An in vitro–in vivo correlation (IVIVC) is expected only in the case of Class II drugs. An IVIVC could be expected for Class I drugs if the dissolution rate is slower than the gastric emptying rate. With a sufficiently rapidly dissolving Class I drug, little or no IVIVC is expected because gastric emptying (not dissolution) would be the rate limiting step. Little or no IVIVC is expected for Class III or Class IV drugs.

The FDA has adopted the BCS in developing a guidance that provides relaxed policies on scale-up and postapproval changes of immediate-release oral solid dosage forms (SUPAC-IR). For certain changes, requirements depend on the drug class, with the most liberal policies for Class I drugs, less liberal policies for Classes II and III drugs, and the least liberal policies for Class IV drugs. First issued as a draft on Nov. 29, 1994 for comment (16, 17), a revised version was published in the *Federal Register* on Nov. 30, 1995.

The intrinsic dissolution rate (IDR) of drugs is frequently measured in preformulation tests by the rotating disk method or Wood's apparatus (18). An automated IDR system, based on a modification of a standard dissolution apparatus, allows for attachment to the stirrer of a die in which the pure drug has been compressed with the tablet face flush with the bottom surface of the die (19). The IDR may be used to detect different polymorphs as well as to judge the risk of a drug exhibiting dissolution-rate-limited absorption. Kaplan (20) suggested that an IDR of higher than $1 \text{ mg cm}^{-2} \text{ min}^{-1}$ indicated that dissolution-related absorption problems were unlikely, whereas an IDR lower than $0.1 \text{ mg cm}^{-2} \text{ min}^{-1}$ indicated dissolution-rate-limited absorption.

Drug-Excipient Compatibility Studies

A knowledge of the interaction of drugs and excipients is essential in the initial formulation of a product. It may also be necessary later on during processing scale-up, when problems arise, to determine if incompatibilities exist which affect manufacturing or stability. Drug-excipient interactions are often directly related to the moisture present in one or another of the components or to the humidity to which the formulation is exposed during processing or storage. These studies are always carried out at accelerated temperature and humidity conditions, even though it must be recognized that some interactions are physical (melting and volatilization) and not chemical and that accelerated aging may not be predictive. Tests for excipient-drug interactions are usually conducted on blends of the pure drug and excipient in ratios similar to those in the final dosage form. For example, excipient-to-drug ratios are higher for filler-binders than for lubricants and disintegrating agents. These studies are often performed with the help of a factorial or fractional-factorial experimental design (21). Powders are physically mixed and may be granulated or compacted to accelerate any possible interaction. Samples can be exposed in open pans or sealed in bottles or vials to mimic product packaging. Evaluation of samples includes

1. Visual inspection for changes in color or texture.
2. Both HPLC and TLC are commonly employed with unstressed samples being used as controls. In general, only qualitative results are important initially.
3. Differential thermal analysis is applied and the appearance or disappearance of one or more peaks is noted. Isothermal microcalorimetry can also be employed as well as a thermal activity monitor (TAM) technique.

Compatibility studies are essential in characterizing both raw materials and finished formulations. It has been argued that binary drug-excipient screening studies are inefficient, unrealistic, and ignore processing variables. A better approach may be to carefully select potential excipients based on known chemistry and published compatibility data, and perform miniformulation stability studies (22).

Formulation Design

Based on the preformulation information, decisions can be made regarding formulation design and process strategy. Initial guidance may be provided by the proposed dose. Relatively low-dose drugs can often be tableted by direct compression, a term that is applied to the process by which tablets are compressed directly from blends of the active

ingredient and suitable excipients. No wet or dry granulation is required, although the drug may occasionally be sprayed out of solution onto one of the excipients to ensure uniform dispersion of drug in very low dosage. Larger doses of poorly compactible drugs may be granulated prior to tableting. The process steps required and the choice of excipients are often governed by other properties of the drug.

Analysis of Critical Variables and Formulation Development

Based on the analysis of the preformulation data, likely excipients are selected and small batches may be produced. The number and size of the batches depend on the availability of the drug substance. The batches are intended to assess the feasibility of the formulation, including the types and levels of excipients, as well as the process and its operational variables, such as order of addition, mixing times, compression force, granulation time, etc. The goal is to develop a formulation and process that meets the criteria set forth earlier under Objectives.

MANUFACTURE

Traditionally, tablets have been made by granulation, a process that imparts two primary requisites to formulations: compactibility and fluidity. Both wet granulation and dry granulation (slugging or roll compaction) are used (Table 1). Regardless of whether tablets are made by direct compression or granulation, the first steps, milling and mixing, are the same; the subsequent steps differ.

The wet massing of powders is typically carried out in high-shear mixers prior to wet screening. The wet granules are often dried in fluidized-bed equipment, enhancing the efficiency of the process. Alternatively, wet granulation may be carried out in fluid-bed drier-granulators in which the liquid phase is sprayed onto fluidized powders while the hot air flow dries the granules. This process reduces the number of handling steps and the time and space needed for granulation; it can be automated. The advantages and disadvantages of wet granulation are given in Table 2. See also Granulations by H.G. Kristensen and T. Schaeffer, Vol. 7 (1st Ed.), pp. 121–160, of this encyclopedia.

Regardless of the granulation method, the comparative simplicity of the direct compression process offers obvious advantages, such as

1. Economy
2. Elimination of heat and moisture

Table 1 Typical unit operations involved in wet granulation, dry granulation, and direct compression

Wet granulation	Dry granulation	Direct compression
Milling and mixing of drugs and excipients	Milling and mixing of drugs and excipients	Milling and mixing of drugs and excipients
Preparation of binder solution	Compression into slugs or roll compaction	Compression of tablets
Wet massing by addition of binder solution or granulating solvent	Milling and screening of slugs and compacted powder	
Screening of wet mass	Mixing with lubricant and disintegrant	
Drying of the wet granules	Compression of tablets	
Screening of the dry granules		
Blending with lubricants and disintegrant to produce "running powder"		
Compression of tablets		

3. Optimization of tablet disintegration

4. Stability

The most obvious advantage of direct compression is its greater economy, owing to reduced processing time, less equipment and space required, less process validation, and lower energy utilization. Generally, only blending and compression are required, although prior micronization of the drug may be needed. Unlike wet granulation, processing does not require heat or moisture, which can be detrimental to drug stability. Moreover, direct compression avoids the high pressures associated with slugging or roll compaction. In addition, disintegration is optimized because directly compressed tablets produce primary particles upon disintegration, rather than granules, which must deaggregate to liberate primary particles. Finally, direct compression tablets often exhibit fewer long-term problems of chemical stability or changes in dissolution.

Although there are many significant advantages of direct compression over granulation, there also are important limitations:

1. Uniform blending and prevention of unblending of low-dose drugs
2. Fillers often are costlier than fillers used in granulation
3. Physical properties and functional specifications are more critical; properties of raw materials must be defined and carefully controlled
4. Limitations in producing colored tablets
5. Dust problems
6. Limitations in the dilution capacity of fillet-binders
7. More sensitive to lubricant softening and overmixing than granulations

Limitations in the dilution capacity of excipients can make the direct compression of large-dose, poorly compactible drugs impractical. Lubrication is often

Table 2 Advantages and disadvantages of wet granulation

Advantages	Disadvantages
Enhances fluidity and compactibility. suitable for high-dose drugs with poor flow and/or compactibility	Each unit process brings its own set of complications
Reduces air entrapment	The large number of unit processes increases the chances of problems
Reduces dustiness	Difficult to control and validate
Provides for the addition of a liquid phase (wet granulation) suited to dispersion of low-dose drugs in solution to ensure content uniformity	Potential adverse effects of temperature, time, and rate of drying on drug stability and distribution during drying
Enhances wettability of powders through hydrophilization (wet granulation)	Overall more costly than direct compression in terms of space, time, and equipment requirements
Permits handling of powders without loss of blend quality	

a compromise between the amount and type needed for adequate lubrication and their adverse effect on compactibility. Content uniformity is of greater concern in direct compression tableting, particularly with low-dose drugs. Since the drug is not “locked” into granules, direct compression blends are subject to unmixing in subsequent processing steps. In addition, drugs are often micronized prior to blending to enhance their dissolution rate, and the resulting high surface-to-mass ratios may lead to difficulty in flowing and mixing due to surface interactions. Another important limitation is that unlike granulation, which tends to compensate for variability in excipients, direct compression is heavily dependent upon reproducible properties of the excipients (and the drug). Raw material standards must be carefully defined and address functionality. Lot-to-lot variations in both the drug and the excipients must be avoided. The cost of raw materials and their testing is higher in direct compression.

Thus, direct-compression tableting requires careful attention to the choice of excipients, appropriate flow properties, and blend homogeneity, and to the interplay of formulation and process variables that can affect both compactibility and drug dissolution. See also Direct Compression Tableting by R.F. Shangraw, Vol.4, (1st Ed.) pp. 85–106, of this encyclopedia.

Excipients

The design of the formulation and selection of excipients is especially critical in tablet dosage forms. Products can vary from a relatively simple aspirin tablet containing aspirin and starch to more complex systems that might contain fillers, binders, disintegrating agents, glidants, lubricants, and coating agents. Modified release introduces even more complexity. The appropriate selection of excipients and their concentration are clearly critical to both the ability to manufacture tablets as well as to their performance as a drug delivery system. Since others have illuminated the various excipient classes in great detail, references are provided in Table 3.

Manufacturability

Excipients function to provide compactibility, lubrication, flow properties, disintegration efficiency, wetting, etc. Poor choice of excipients may give rise to poor characteristics (hardness, appearance), which can be important in packaging, storage, and patient acceptance. Problems with excipients may arise from variations in source or lot, particularly in formulations made by direct compression. Examples of excipient problems include variation in performance between *Hoc* cellulose and microcrystalline cellulose relative to particle size, flow, and compactibility,

or different polymorphic forms of sorbitol (α , β and γ) resulting in tablet hardening. Differences in lactose particle size and modification (spray-dried or anhydrous) can provide differences in surface areas over which a lubricant is distributed. Anhydrous lactose hydrates at high humidities with increase in size.

Biopharmaceutics

The formulation of a tablet can affect its bioavailability. Particular care should be given to low-dose poorly soluble drugs, especially those that are micronized. Drugs with low water solubility should never be formulated solely with insoluble fillers, including calcium salts (calcium sulfate, calcium phosphate) which are only soluble at very low pH. Differences in solution rate between hydrated and nonhydrated forms of calcium salts as well as between dibasic and tribasic forms may also be important. In some cases, excipients complex with drug substances such as calcium salts with tetracycline. Varying the ratio of soluble to insoluble fillers in tablets can significantly alter the dissolution pattern of poorly soluble drugs (weak acids and nonelectrolytes), but has little effect on weak bases, which are soluble in gastric fluids.

The type and amount of disintegrating agent can also be important. Differences in source (corn, potato, rice) as well as variations in amylopectin–amylose ratios result in variable disintegration times. The fact that starch included within granules is not as effective as starch added between granules has led to tablets with good disintegration but poor dissolution. Incorporation of super-disintegrants (croscopovidone, croscarmellose, sodium starch glycolate) has improved dissolution from both direct-compression and wet-granulation formulations. Both croscarmellose and sodium starch glycolate can complex with small amounts of cationic drugs in water, but not in physiologic fluids (23).

The type and amount of binder used in granulations affects dissolution rates. Many binders are hydrophilic polymers whose solubility and solubility rate depend upon molecular weight. Quality control tests such as viscosity may be necessary as part of raw material testing of polymeric substances. However, an advantage of the granulation process is that it results in a wetting of drug surfaces, which enhances drug dissolution once the granules have disintegrated. Wetting agents such as sodium lauryl sulfate can significantly improve the dissolution of drugs of poor water solubility formulated into direct compression tablets.

Perhaps the greatest source of concern in excipients is with lubricants, particularly magnesium stearate, which is not only hydrophobic but also has a laminar crystal structure. When blended with other ingredients, it tends to

Table 3 Excipients literature

Excipient class	References (1 st Ed.)
General	Rudnic, E.; Kottke, M. Tablet Dosage Forms. In <i>Modern Pharmaceutics</i> , 3rd Ed.; Banker, G., Rhodes, C., (Eds.); Marcel Dekker, Inc.: New York, 1996; 333.
Diluents and fillers	Czeisler, J.; Perlman, K. <i>Encyclopedia of Pharmaceutical Technology</i> ; Swarbrick, J., Boylan, J., (Eds.); Marcel Dekker, Inc.: New York, 1991; Vol. 4, 37.
Binders	Kristensen, H. <i>Encyclopedia of Pharmaceutical Technology</i> ; Swarbrick, J., Boylan, J., (Eds.); Marcel Dekker, Inc.: New York, 1988; Vol. 1, 415.
Disintegrants	Shangraw, R.; Mitrevej, A.; Shah, M. Pharm. Tech. 1980; 4 (10), 49. Augsburger, L.L.; Brzezczko, A.W.; Shah, U.; Hahm, H. <i>Encyclo of Pharm Technology</i> ; Swarbrick, J., Boylan, J., (Eds.); Marcel Dekker, Inc.: New York, 2001; Vol. 20, Suppl. 3.
Lubricants and glidants	Zanowiak, P. <i>Encyclopedia of Pharmaceutical Technology</i> ; Swarbrick, J., Boylan, J., (Eds.); Marcel Dekker, Inc.: New York, 1994; Vol. 9, 87.
Film coating agents	Radebaugh, G. <i>Encyclopedia of Pharmaceutical Technology</i> ; Swarbrick, J., Boylan, J., (Eds.); Marcel Dekker, Inc.: New York, 1993; Vol. 6, 1.
Controlled-release agents	Chien, Y. <i>Encyclopedia of Pharmaceutical Technology</i> ; Swarbrick, J., Boylan, J., (Eds.); Marcel Dekker, Inc.: New York, 1990; Vol. 3, 281.
Coloring agents	Woznicki, E.; Schoneker, D. <i>Encyclopedia of Pharmaceutical Technology</i> ; Swarbrick, J., Boylan, J., (Eds.); Marcel Dekker, Inc.: New York, 1990; Vol. 3, 65.
Flavor modifiers	Adjei, A., et al. <i>Encyclopedia of Pharmaceutical Technology</i> ; Swarbrick, J., Boylan, J., (Eds.); Marcel Dekker, Inc.: New York, 1993, Vol. 6, 101.

make them hydrophobic by delaminating to coat their surfaces. The problems with magnesium stearate are thus highly process dependent. For example, blending time differences of as little as 2 min can significantly alter the dissolution pattern of finished tablets. Because direct-compression formulations have a higher specific surface area, the same amount of magnesium stearate blended for the same length of time makes direct-compression tablet matrices more hydrophobic than matrices made from granulations. The degree of shear imparted by different mixers during processing significantly affects the distribution of magnesium stearate (24). The characteristics of magnesium stearate vary from supplier to supplier and sometime within the same supplier (25). It is essential to draft raw material specifications and strictly adhere to Standard Operating Procedures (SOPs) during the manufacture of products containing this lubricant. Establishing purchasing specifications beyond those listed in the *National Formulary*, such as bulk and tap density, powder fluidity, particle size, surface area, degree of hydration, and morphology is desirable.

THE EFFECTS OF MANUFACTURING PROCESSES ON FORMULATIONS

Numerous unit processes are involved in making tablets, including particle size reduction and sizing, blending,

granulating, drying, compaction, and (frequently) coating. Various factors associated with these processes can seriously affect content uniformity, bioavailability, or stability. Some of these are given in the following list:

1. Particle Size Reduction
 - Nonuniform particle size can lead to segregation problems
 - Development of electrostatic forces inhibits complete blending
 - Changing the crystalline state can affect solubility
2. Blending
 - Nonhomogeneous distribution of drug substance is the result of poor blending or unblending
 - Overblending of lubricant lowers dissolution rates and affects compactibility
3. Granulation
 - Nonhomogeneous distribution of binder and drug substance gives drug-rich or drug-poor fines
 - Decomposition of drug substance due to residual moisture
 - Uneven granule size (too many or too few fines) leads to compaction or uniformity problems
4. Tableting
 - Uneven compaction pressures affect dissolution
 - Loss of mix quality in hopper and feed frame gives poor content uniformity
 - Additional shearing of lubricant in feed frame lowers dissolution rates

5. Coating

- Nonuniform or incomplete coverage of tablets and beads results in different dissolution patterns

When validating new equipment or procedures, the sampling techniques must reflect the quality of the material being tested, the blend of powders, moisture in granulation, or coating integrity.

Another potential source of problems in manufacturing is reprocessing or reworking. Reworking may be required when finished products fail to meet hardness, content uniformity, disintegration, dissolution, or appearance specifications. Reworking procedures must be in writing and are often part of New Drug Applications (NDAs) and Abbreviated New Drug Applications (ANDAs). Although practiced less today than in past years, given the high cost of some drug substances, reworking continues to be justified, but may involve the following problems: overdistribution of lubricant leading to poor compactibility and dissolution; distribution of particles of coating in the reprocessed tablet may lower the dissolution rate; and loss in compactibility due to work hardening of direct-compression fillers, particularly at higher initial compaction pressures. Reworking is less likely to cause dissolution problems with water-soluble drugs.

SYSTEMATIC FORMULATION DEVELOPMENT

Previously, pharmaceutical experimenters were rarely afforded the luxury of using expanded experimental designs. Time and manpower constraints, imposed by an ever-quicken pace to market, have prevented the thorough examination and full understanding of most commercial formulations. Although the pace of formulating pharmaceutical systems has not lessened recently, a rapid turnaround means formulators must quickly gather information and base final decisions on that information. The window for error is narrower, and the formulator can no longer afford to use empirically gathered data to obtain the critical dosage form information needed for a marketable formulation. Systematic development approaches are also desirable to provide data in anticipation of the use of SUPAC (scale up and postapproval changes) regulatory policies by providing for the establishment of a research database that can help justify such changes to regulatory agencies.

Enormous progress has been made in the direction of systematic formulation development through the use of

such statistical tools as multivariate analysis and response surface methodology, and artificial intelligence.

Experimental Design

In the long term, an efficient experimental design saves time and avoids costly mistakes, and some pharmaceutical firms have departments devoted to the preparation of experimental designs and their analysis. If a statistician is not available, a myriad of commercial experimental-design software packages are on the market (SAS, JMP, STATGRAPHICS, DESIGN EXPERT) with which a formulator can design experiments on a personal computer. These same packages can also aid in data analysis and presentation.

The early phase of such a statistical experimental design approach may include screening designs such as two-level full or fractional factorial designs or Plackett-Burman designs. The number of experimental runs is reduced by intentionally confounding some experimental effects. Ultimately, the screening design narrows the number of potential variables for further study.

The latter phase of such as a central experimental design approach might include response surface methodologies to optimize a formulation such as a central composite or Box-Behnken design. Such methods allow experimenters to assess the effects of several variables at one time without having to study every possible unique combination of variables. This way, a systematic identification of critical variables and an optimization of the formulation and process can be obtained. Response surfaces are generated, which give formulators a graphic demonstration of the effect of variables on various responses, such as drug dissolution. The designs and models available to experimenters are numerous and the proper choice depends on the number of variables and the possible responses. General references for experimental design can be consulted as well as design software manuals, if a statistician is not available for advice.

Artificial Intelligence Approaches

Among the artificial intelligence approaches that have been used to provide support for the formulation process are expert systems and artificial neural networks. An expert system (ES) is an intelligent computer program that attempts to capture the expertise of specialists who have knowledge and experience in a well-defined domain. They are designed to simulate an expert's problem solving process. Developed more than 20 years ago for other

applications, expert systems are a relatively new idea in pharmaceutical technology. In rule-based systems, the knowledge is highly structured and often represented as a set of rules that express the relationship between several pieces of information in the form of conditional statements. These statements specify actions to be taken or advice to be followed. Such systems can shorten development time, simplify formulations, provide the rationale for decisions taken in arriving at a formulation, serve as excellent teaching tools for novices, and accumulate and preserve the knowledge and experience of experts. ESs suffer from the limitation that they are not creative. They can only deal with situations that have been anticipated. They must be designed to handle every contingency. Examples include expert systems developed for formulating tablets (26), for process troubleshooting (27), and for the selection of a mixer (28).

Used in other disciplines for about 40 years, artificial neural networks (ANNs), like ESs, have only recently been applied to pharmaceutical development. ANNs are computer-based programs that attempt to simulate certain functions of the biological brain, such as learning, generalizing, or abstracting from experience. They have the ability to discern relationships or patterns in response to exposure to facts ("learning."). ANN models may be viewed simply as multiple nonlinear regression models. Through ANNs, the data and information generated during experimental work may be transformed relatively easily into knowledge that would enable the formulator to at least construct a few domain specific rules, even though confined, for future cases. However, ANN effectiveness is limited by the training data selected. One limitation is that, in most cases, ANNs lack explanation capability and there is difficulty in obtaining justification for results. Examples of applications to product development include predicting model granulation and tablet characteristics from knowledge of material and process variables (29), and predicting drug release from immediate release formulations (30). The development of a hybrid system, i.e. an Expert Network, that integrates ANNs and ESs has the potential of taking advantage of the strengths of both ANNs and ESs and avoiding the weaknesses of either (31, 32).

Formulation Development

Whether working in a large established laboratory or a small research organization, formulators are well advised to take an overall look at the development process to assess the most rational approach to their particular needs and resources. Among many approaches to rational tablet formulation, the strategy used by researchers at the

University of Maryland in collaboration with FDA scientists (33) examines the drug product development process in light of the SUPAC guidance.

This generalized plan focuses on the assessment the possible influence of formulation and processing variables on bioavailability and manufacture. The research protocol is best visualized by the flow chart in Fig. 1. The model begins with the preformulation stage wherein information is obtained on the physicochemical and biopharmaceutic properties of the drug. One outcome of the preformulation study should be the identification of the biopharmaceutic class (BCS) in which the drug falls since this will provide important guidance in making formulation decisions. For multi-sourced drugs, the different sources are considered. The preformulation study results in a portfolio of information, which provides guidance in formulation design and in the development of appropriate hypotheses to be tested and the critical variables studies that follow.

In the critical variables analysis phase, a statistical experimental design is created (e.g., factorial, Box–Behnken) intended to assess critical formulation and process variables in relatively small-scale manufacture. In these studies, the ranges of composition variables are chosen to at least encompass those noted in the recommendations of the AAPS–FDA Workshop on Scale-up of

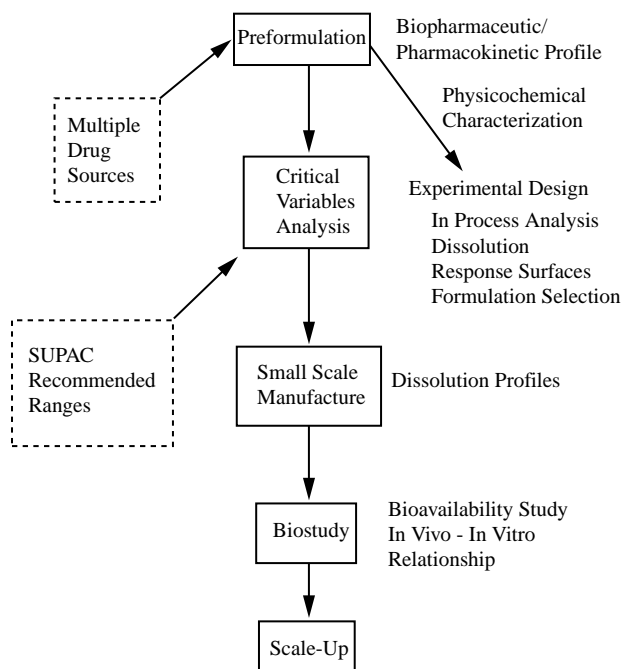


Fig. 1 A research model used in a University of Maryland and FDA collaborative research program. (From Ref. 33.)

Immediate Release Oral Solid Dosage Forms or SUPAC. This phase is usually preceded by a development phase during which variables and levels to be studied are determined and the exact method of manufacture is established. Experimental formulations are assessed at least in terms of dissolution performance, content uniformity, and weight variation. On the basis of these studies, the specific formulations to be manufactured for biostudy are selected.

In the small-scale clinical manufacturing phase, formulations are manufactured under GMP conditions for possible clinical testing. These will be manufactured on a larger scale than those in the previous phases. An experimental design is chosen; however, if some of the variables can be eliminated based on the earlier experiments, the number of formulations produced may be reduced.

The intent of the biostudy phase is to establish an in vitro–in vivo relationship. If an appropriate correlation can be established, dissolution may serve as a surrogate for biostudies in the interpretation of what is significant and what is not among the variables studied.

In the scale-up phase, larger runs of formulations are manufactured. The formulations are selected to determine if the larger scale will enhance the significance of certain variables.

The statistical analysis of the data provides the opportunity to predict changes in dissolution performance resulting from incremental changes in one or more formulation or process variable at a time (e.g., the level of an excipient or the time of mixing).

The development model presented here may differ from some formulation research programs in that biostudies may not be performed on small-scale batches. The major advantage of early biostudies is the potential for early IVIVC and subsequent surrogate use of dissolution testing in further work. Scale-up and pilot-plant roles in formulation changes, while not covered thoroughly here, are reviewed by Racz (34).

REFERENCES

1. Chowhan, Z.T. Excipients and Their Functionality in Drug Product Development. *Pharm. Tech.* **1993**, 17 (9), 72.
2. Czeisler, J.L.; Penman, K.P. Diluents, *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1991; 4, 40–43.
3. Blecher, L.; Ohmae, T.; deJong, H.J. Tri-PEC: Reports from IPEC-Americas, JPEC, and IPEC-Europe. *Pharm. Tech.* **1994**, 18 (8), 53.
4. Raleblian, J.; McCrone, W.J. Pharmaceutical Applications of Polymorphism. *Pharm. Sci.* **1969**, 58, 911.
5. Matsumoto, T.; Nobuyoshi, K.; Iliguchi, S.; Otsuka, M.J. Effect of Temperature and Pressure During Compression on Polymorphic Transformation and Crushing Strength of Chlorpropamide Tablets. *Pharm. Pharmacol.* **1991**, 43, 74.
6. Brittain, H.G. Raw Materials. *Drug Devel. Ind. Pharm.* **1989**, 15, 2083.
7. Schwatz, J.B. The Instrumented Tablet Press: Uses in Research and Production. *Pharm. Tech.* **1981**, 5 (9), 102.
8. Hunter, B.M.; Fisher, D.G.; Pratt, R.W.; Rowe, R.C. A High Speed Compression Simulator. *J. Pharm. Pharmacol.* **1976**, 28, 65.
9. Maganti, L.; Celik, M. Compaction Studies on Pellets I. Uncoated Pellets. *Int. J. Pharm.* **1993**, 95, 29.
10. Muller, F.X.; Augsburger, L.L. The Role of the Displacement-Time Waveform in the Determination of Heckel Behavior Under Dynamic Conditions in a Compaction Simulator and a Fully Instrumented Rotary Tablet Machine. *J. Pharm. Pharmacol.* **1994**, 46, 468.
11. Roberts, R.J.; Rowe, R.C. The Effect of Punch-Velocity on the Compaction of a Variety of Materials. *J. Pharm. Pharmacol.* **1985**, 37, 377.
12. Hiestand, E.; Smith, D.P. Indices of Tableting Performance. *Powder Tech.* **1984**, 38, 145.
13. Augsburger, L.L.; Shangraw, R.F. Effect of Glidants in Tableting. *J. Pharm. Sci.* **1966**, 55, 418.
14. Fed. Reg. *21CFR, Ch. I*; (4/1/87 Ed.) Part 32; 320.
15. Amidon, G.L.; Lennernas, H.; Shah, V.P.; Crison, J.R. A Theoretical Basis for a Biopharmaceutical Drug Classification: The Correlation of In Vitro Drug Product Dissolution and In Vivo Bioavailability. *Pharm. Res.* **1995**, 12, 413.
16. INTGUIDE.1R9. *FDA Center for Drug Evaluation and Research*: Rockville, MD, 1994.
17. Lucisano, L.J.; Franz, R.M. FDA Proposed Guidance for Chemistry, Manufacturing, and Controls Changes for Immediate-Release Solid Dosage Forms: A Review and Industrial Perspective. *Pharm. Tech.* **1995**, 19 (5).
18. Wood, J.H.; Syarto, J.E.; Letterman, H.J. Improved Holder for Intrinsic Dissolution Rate Studies. *Pharm. Sci.* **1965**, 54, 1068.
19. Koparkar, A.D.; Augsburger, L.L.; Shangraw, R.F. Intrinsic Dissolution Rates of Tablet Filler-Binders and their Influence on the Dissolution of Drugs from Tablet Formulations. *Pharm. Res.* **1990**, 7, 80.
20. Kaplan, S.A. Biopharmaceutical Considerations in Drug Formulation Design and Evaluation. *Drug Metab. Rev.* **1972**, 1, 15.
21. Jacobs, A.L. Determining Optimum Drug/Excipient Compatibility Through Preformulation Testing. *Pharm. Manuf.* **1985**, 2 (6), 43.
22. Monkhouse, D.C.; Maderich, A. Whither Compatibility Testing. *Drug Devel. Ind. Pharm.* **1989**, 15, 2115.
23. Hollenbeck, R.G.; Mitrejev, K.; Fan, A. Estimation of the Extent of Drug-Excipient Interactions Involving Croscarmellose Sodium. *Pharm. Sci.* **1983**, 72, 325.
24. Bolhuis, G.K.; deJong, S.W.; Lerk, C.F. The Effect of Magnesium Stearate Admixing in Different Types of Laboratory and Industrial Mixers on Tablet Crushing Strength. *Drug Devel. Ind. Pharm.* **1987**, 13, 1547.
25. Danserean, R.; Peck, G.E. The Effect of the Variability in the Physical and Chemical Properties of Magnesium

- Stearate on the Properties of Compressed Tablets. *Drug Devel. Ind. Pharm.* **1987**, *13*, 975.
26. Rowe, R.C. An Expert System for the Formulation of Tablets. *DTI Manuf. Intel. Newsletr.* **1993**, *14*, 13–15.
 27. Murray, F.J. The Application of Expert Systems to Pharmaceutical Processing. *Pharm. Tech.* **1989**, *13* (3), 100–110.
 28. Lai, F.K.Y. A Prototype Expert System for Selecting Pharmaceutical Powder Mixers. *Pharm Tech.* **1988**, *12* (8), 22–31.
 29. Kesavan, J.G.; Peck, G. Pharmaceutical Granulation and Tablet Formulation Using Neural Networks. *Pharm. Dev. Tech.* **1996**, *1* (4), 391–404.
 30. Ebube, N.K.; McCall, T.; Chen, Y.; Meyer, M.C. Relating Formulation Variables to In Vitro Dissolution Using an Artificial Neural Network. *Pharm. Dev. Tech.* **1997**, *2* (3), 225–232.
 31. Turban, E. *Expert Systems and Applied Artificial Intelligence*; Macmillan Publishing Co.: New York, 1992.
 32. Caudill, M. Expert Networks. *Neural Network PC Tools*; Academic Press: New York, 1990; 189–214.
 33. Augsburger, L.L.; Shangraw, R.; Lesko, L.; Williams, R. An Approach Toward Establishing a Scientific Foundation for Interpreting Regulations and Workshop Reports on Scale-Up and Post Approval Changes. *Pharm. Res.* **1994**, *11* (10), S-161.

34. Racz, I. *Drug Formulation*; John Wiley & Sons: New York, 1989; 25–27.

BIBLIOGRAPHY

- American Pharmaceutical Association and the Royal Pharmaceutical Society of Great Britain. *Handbook of Pharmaceutical Excipients*, 3rd Ed.; American Pharmaceutical Association and Pharmaceutical Press: Washington, 2000.
- Armstrong, A.M.; James, K.C. *Understanding Experimental Design and Interpretation in Pharmaceutics*; Ellis Horwood Ltd: Chichester, UK, 1990.
- Franz, R.M.; Browne, J.; Lewis, A. Experimental Design, Modeling, and Optimization Strategies for Product and Process Development, *Pharmaceutical Dosage Forms: Disperse Systems*, 2nd Ed.; Lieberman, H.L., Rieger, M., Banker, G., Eds.; Marcel Dekker, Inc.: New York, 1996; 1, 427–514.
- Peck, G.E.; McCurdy, V.E.; Banker, G.S. Tablet Formulation and Design, *Pharmaceutical Dosage Forms: Tablets*, 2nd Ed.; Lieberman, H.L., Lachman, L., Schwartz, J., Eds.; Marcel Dekker, Inc.: New York, 1989; 1, 75–130.
- Polderman, J. *Formulation and Preparation of Dosage Forms*; Elsevier/North-Holland: New York, 1977; 3–28.

TABLET MANUFACTURE

Norman Anthony Armstrong

Cardiff University, Cardiff, United Kingdom

INTRODUCTION

The compressed tablet is by far the most widely used dosage form, having advantages for both producer and user. However, the manufacture of tablets can be a complex process, since only a few raw materials inherently possess those properties which are necessary for the production of tablets of satisfactory quality. Hence some preliminary treatment and/or incorporation of excipients in the formulation is usually needed. Tablet manufacture is a paradox. Considerable ingenuity and formulation expertise are required to transform a mass of particles into a low porosity mass. Yet, after the tablet has been ingested, the requirement then is usually for the tablet to release its active ingredient as rapidly as possible, and further ingenuity is needed to bring this about.

Tablets are solid preparations each of which contains a single dose of one or more active ingredients. They are obtained by compressing uniform volumes of particles, and are almost always intended for oral administration.

The earliest reference to a dosage form resembling the tablet is to be found in tenth century Arabic medical literature. Drug particles were compressed between the ends of engraved ebony rods, force being applied by means of a hammer. Details of the tableting process, as it is now known, were first published in 1843 when William Brockedon was granted British Patent 9977 for "manufacturing pills and medicinal lozenges by causing materials when in a state of granulation, dust or powder, to be made into form and solidified by pressure in dies." In this case, too, force was applied by a hammer. Potassium bicarbonate was the first pharmaceutical substance to be so treated.

The use of compressed pills, as they were then known, increased rapidly. It is likely that the term "tablet" for this dosage form was first used in the United States in the 1870s. Power-driven presses replaced Brockedon's hammer, and by 1874 there existed both rotary and excentric presses, which in their mode of operation were fundamentally similar to those in use at the present time. The tablet lent itself to mass manufacture by mechanical means, in contrast to the slower labour-intensive production of older solid dosage forms such as the pill. It is impractical for individual pharmacists to produce

small quantities of tablets on a commercial scale, and this led to the concentration of pharmaceutical manufacture in relatively few industrial sites.

A monograph for Glyceryl Trinitrate Tablets was included in the British Pharmacopoeia of 1885, but no other tablet monograph appeared there until 1945. This was not due to lack of popularity of the dosage form itself, but rather the absence of suitable methods of quality control that were applicable to tablets.

The tablet did not meet with universal approval. In 1895 an editorial in the *Pharmaceutical Journal* in the United Kingdom described the tablet as "one of the evils suffered by legitimate pharmacy," and predicted that tablets "have had their day" (1). Notwithstanding such a prediction, the usage of tablets has continued to increase. The 2000 edition of the British Pharmacopoeia contains 320 monographs for tablets, far in excess of any other dosage form.

The tablet is the most popular dosage form because it provides advantages for all concerned in the production and consumption of medicinal products. Though the initial capital outlay for the manufacturer of tablets is considerable, they can be produced at a much higher rate than any other dosage form, tablet presses capable of producing about one million tablets per hour being available. Furthermore, the fact that the tablet is a dry dosage form promotes stability, and in general, tablets have shelf lives measured in years. They are also convenient to transport in bulk, since they contain relatively small proportions of excipients unlike, for example, oral liquids.

From the viewpoint of the pharmacist, tablets are easy to dispense, while the patient receives a concentrated and readily transportable and consumed dosage form. Furthermore, if properly prepared, tablets provide a uniformity of dosage greater than that of most other medicines, and appropriate coating can mask unpleasant tastes and improve patient acceptance.

The tablet also provides a versatile drug delivery system. Though most tablets are intended to be swallowed intact, the same basic manufacturing process, associated with appropriate formulation, provides medicines for sublingual, buccal, rectal, and vaginal administration, together with lozenges, soluble, dispersible, and effervescent tablets. In addition, techniques that can delay or

otherwise modify the release of the active ingredient from the tablet are available.

Naturally tablets only possess these advantages if they are properly formulated and manufactured. A well-prepared tablet should possess the following qualities:

1. It should, within permitted limits, contain the stated dose of drug.
2. It should be sufficiently strong to withstand the stresses of manufacture, transport, and handling so as to reach the patient intact.
3. It should deliver its dose of drug at the site and at the speed required.
4. Its size, taste, and appearance should not detract from its acceptability by the patient.

TABLET COMPRESSION

All tablets are made by a process of compression. Solid, in the form of relatively small particles, is contained in a die and a compressing force of several tonnes is applied to it by means of punches. The shape of the die governs the cross-sectional shape of the tablet, and the distance between the punch tips at the point of maximum compression governs its thickness. The conformation of the tablet faces, usually flat or convex, is a reflection of those of the punches.

The tip of the lower punch moves up and down within the die, but never actually leaves it. The upper punch descends to penetrate the die and apply the compressive force. It is then withdrawn to permit ejection of the tablet, brought about by an upward movement of the lower punch.

There are two types of tablet press. The excentric press has one die and one pair of punches. The rotary press has a larger number of dies which are fitted, with their corresponding punches, into a rotating turret.

Irrespective of the type of press that is used, the process of tablet compression can be divided into three stages, as shown in Fig. 1.

Stage 1, Filling

The lower punch falls within the die, leaving a cavity into which particulate matter flows under the influence of gravity from a hopper. Though tablets are usually described in terms of weight, the die is filled by a volumetric process. The volume is determined by the depth to which the lower punch descends in the die. Unless this volume is filled reproducibly on each occasion, then

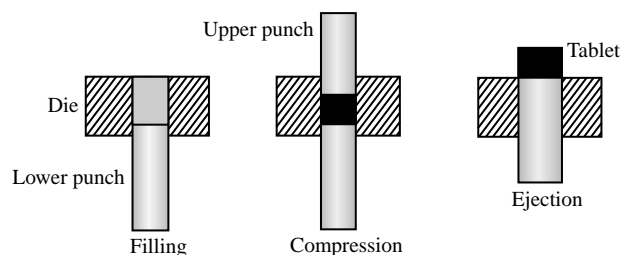


Fig. 1 Cycle of operations of an excentric tablet press.

the mass of the tablet will vary, and with it the drug content of each tablet. Therefore, uniform filling is essential. However, it must be borne in mind that the die cavity has a cross-section of only a few millimetres, and only a fraction of a second is available for filling each die. It therefore follows that the particles must flow easily and reproducibly.

Stage 2, Compression

The upper punch descends, and its tip enters the die, confining the particles. The distance separating the punch faces decreases, either by movement of the upper punch alone (as in excentric presses) or by movement of both punches (as happens in rotary presses). The porosity of the contents of the die is progressively reduced, and the particles are forced into ever-closer proximity to each other. This process is facilitated by the particles fragmenting and/or deforming. Once the particles are close enough together, interparticulate forces then cause the individual particles to aggregate, forming a tablet. The magnitude of the force is governed by the minimum distance separating the punch faces. Therefore, a second essential property of the particles is that they cohere under the influence of a compressive force. It is also essential that this coherence be maintained when the compressing force is removed.

Stage 3, Ejection

The upper punch is withdrawn from the die, and so the force being applied to the tablet is removed. The effect of this might be to cause the deformed particles to return to their former shape, which would result in a decrease in interparticulate contact and hence tablet strength. It is essential that this does not occur. As the upper punch leaves the die, the lower punch moves upwards, pushing the tablet before it. During the compression stage, the particles are forced into intimate contact with the interior die wall. It follows that attempts to remove the tablet will be opposed

by frictional forces and so successful ejection demands lack of adhesion between the tablet and the die wall.

Therefore in summary, for a particulate solid to be successfully transformed into tablets, three key properties need to be present:

1. Good particle flow.
2. The ability of the particles to cohere under the influence of a compressing force. This coherence must be retained after the compressing force has been removed.
3. The ability of the tablet to be ejected from the die after the compressing force has been removed.

Few powders possess all these essentials and some possess none of them. Thus, before successful tableting can take place, some preliminary treatment with the addition of one or more excipients is almost invariably needed.

METHODS OF TABLET PRODUCTION

The pretreatment that is usually necessary takes the form of granulation. The process of granulation is essentially one of size enlargement, and it serves several purposes in the tablet manufacturing process:

1. It improves flow by increasing particle size, since large particles flow more readily than small ones.
2. It improves compression characteristics, adding to the cohesive strength of the tablet.
3. Once a homogeneous mixture has been achieved, segregation is prevented, since particles that are stuck together cannot separate.
4. It reduces dust.

Both wet and dry granulation techniques are available.

Tablet Manufacture by Wet Granulation

This is the traditional method of pretreatment of solids prior to tableting. Despite its complexity and inherent disadvantages, even now about half the tablets produced worldwide are manufactured by this process. Its essence is that particles of active ingredient, with a diluent if necessary, are stuck together using an adhesive, the latter usually being water-based. The result is a granular product which flows more readily and has an improved ability to cohere during compression.

A flow diagram of the wet granulation process together with appropriate excipients is shown in Fig. 2.

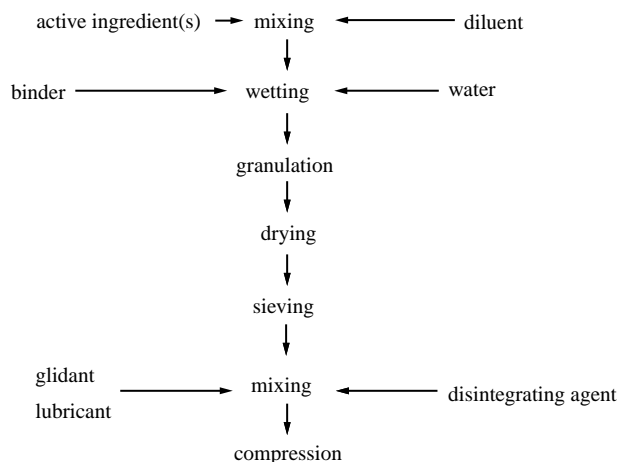


Fig. 2 The wet granulation process of tablet manufacture.

The diluent

The first stage in the wet granulation process is often a dry mixing stage in which the active component is mixed with a diluent. Many drugs need to be administered in doses of only a few milligrams or even less, yet a tablet that weighs less than about 50 mg is difficult for the patient to handle conveniently. It is therefore necessary to increase the bulk of such a tablet with a diluent. Some commonly used diluents are listed in Table 1.

The ideal diluent would be both chemically and physiologically inert, and would not interfere with the bioavailability of the active ingredient. It should also be inexpensive and be easily tableted since, if the proportion of active ingredient is small, the overall tableting properties of the mixture are largely governed by those of the diluent.

Lactose is by far the most frequently used diluent for solid dosage forms. An inexpensive disaccharide obtained as a by-product of the cheese industry; it is available in a number of forms, though α -Lactose monohydrate is the variety that is normally used as the diluent in tablets made by wet granulation. It is freely albeit slowly soluble in water and as such it is a suitable diluent for active ingredients of low water solubility. Lactose is a nonreducing sugar, and is reasonably inert. It can take part in the Maillard reaction when mixed with substances containing primary amine groups, giving highly colored products, and thus its use is contraindicated in such formulations (2).

Probably the second most commonly used diluent in the wet granulation process is dibasic calcium phosphate. This substance is virtually insoluble in water and hence is always used in conjunction with a disintegrating agent. Its properties have been reviewed by Carstensen and Ertell (3).

Table 1 Tablet diluents

Diluent	Comments
Calcium carbonate	Insoluble in water (Cal-Carb ^R , Millicarb ^R , Pharma-Carb ^R , Sturcal ^R)
Calcium phosphate, dibasic	Insoluble in water, good flow properties, available in dihydrate and anhydrous forms (Cyfos ^R , Calstar ^R , Calipharm ^R , Emcompress ^R)
Calcium phosphate, tribasic	Insoluble in water (Tricafos ^R , Tri-Cal ^R , Tri-Tab ^R)
Calcium sulfate	Insoluble in water (Cal-Tab ^R , Compactrol ^R)
Cellulose, microcrystalline	Good compression properties, may not need lubricant, can act as disintegrant (Avicel ^R , Emcocel ^R , Vivacel ^R)
Cellulose, microcrystalline silicified	Combination of microcrystalline cellulose and silica (Prosolv ^R)
Cellulose, powdered	(Elcema ^R , Solka-Floc ^R)
Dextrates	(Emdex ^R)
Dextrose	Hygroscopic, reducing sugar (Tabfine ^R)
Fructose	(Fructofin ^R)
Lactitol	(Finlac ^R)
Lactose monohydrate	The most commonly used diluent. Inexpensive, takes part in Maillard reaction (Fast-Flo ^R , Lactochem ^R , Microtose ^R , Pharmatose ^R , Tablettose ^R , Zeparox ^R)
Magnesium carbonate	
Maltitol	(Maltisorb ^R , Maltit ^R)
Maltodextrin	(Glycidex ^R , Lycatab ^R , Maltrin ^R)
Maltose	(Advantose ^R)
Mannitol	Freely soluble in water, negative heat of solution and therefore cool taste, popular for chewable tablets, noncariogenic (Pearlitol ^R)
Sodium chloride	Freely soluble in water, used in solution tablets
Sorbitol	
Starch	Also acts as disintegrating agent, may give soft tablets
Starch, pregelatinized	Also acts as disintegrating agent. (Lycatab ^R , Pharma-Gel ^R , Pre-Jel ^R , Sepistab ^R , Starch 1500 ^R , Starx 1500 ^R)
Sucrose	Freely soluble in water, sweet taste, hygroscopic, used in lozenges in conjunction with lactose
Sugar, compressible	(Dipac ^R , Nutab ^R)
Sugar, confectioner's	
Sugar spheres	(Nu-Core ^R , Nu-Pareil ^R)
Talc	
Xylitol	Negative heat of solution, cool taste (Xylifin ^R , Xylitab ^R)

Proprietary names are given in parentheses.

Mixing

The purpose of the mixing stage is to ensure that the powder blend and hence the resulting tablets are homogeneous in content. A random mixture is defined as one where the probability of sampling a given type of particle is proportional to the number of such particles in the total mixture. Thus, the aim is to produce a mixture such that when a sample is removed, the relative proportions of the components of that sample are the same as in the mixture as a whole.

Unlike molecules in a fluid, which in time will mix spontaneously by a diffusion mechanism, powder particles do not mix spontaneously but remain in their relative

positions. Therefore before mixing can occur, energy must be put into the system. This causes the powder bed to dilate or expand, the particles separate from one another and this leads to relative motion among them.

It might be intuitively expected that the randomness of a mixture will progressively increase with time, but this is not always the case. Under certain conditions, an optimum mixing time occurs, beyond which the mixture shows a tendency to separate back into its components. This process is known as segregation. Segregation is particularly likely to occur in mixtures where the components differ markedly in size, with differences in shape and density as secondary factors. It is especially likely to occur if regular patterns of movement are set up in the mixing

device, and for this reason, mixers are designed so that an irregular mixing motion occurs (4).

Although in general a size difference between components can lead to segregation, a situation where there is a large difference in sizes between components may be beneficial. In such circumstances, small particles of one component can become trapped in irregularities in the surface of the larger component. These are not random mixtures, as the particles of the two components cannot behave independently. This concept is called "ordered mixing" and it has found applicability in the manufacture of solid dosage forms containing small quantities of highly potent active ingredients (5) (see the article on Blenders and Blending in this encyclopedia).

Granulation

The underlying process of size enlargement in wet granulation is achieved by either one or both of two different mechanisms. Firstly, adjacent solid particles may be stuck together using an adhesive. Such substances are known as binders or granulating agents. Secondly, dissolution of the solid in the granulating liquid can occur, followed by evaporation of the liquid phase of the latter. This will result in the deposition of dissolved material on particle surfaces, forming so-called crystal bridges. The occurrence of this mechanism will depend on the solubility of the solids in the liquid phase. Thus, sucrose will form crystal bridges with an aqueous granulating fluid, whereas calcium phosphate will not.

The process and underlying mechanisms of granulation have been fully described by Sherrington and Oliver (6). Details of commonly used binders are given in Table 2. They are often natural or synthetic polymers and are usually added as aqueous solutions or dispersions. Alternatively, they can be mixed with the other solids in the formulation in the dry state, water then being added.

If the active ingredient is unstable in the presence of water, then a granulation process using nonaqueous liquids can be used. The usual granulating system in such cases is povidone dissolved in isopropanol. The extra costs and environmental problems posed by the use of a volatile and flammable liquid are disincentives to the use of nonaqueous granulation.

The traditional piece of granulating apparatus is the shear granulator. Its function is to homogeneously incorporate an adhesive and viscous liquid such as starch paste into a mass of dry powder to form agglomerates. It follows that a considerable shearing force needs to be exerted. The mixed solids are loaded into the bowl of the mixer, and the liquid added with agitation. The damp solid is then forced through a relatively coarse screen (about 1–2 mm), often by means of oscillating bars, to give

discrete granules. The progression of the granulation process can be monitored by measuring the electrical power consumption by the granulator, and hence optimum granulation times can be established. Ertell et al. have shown that the size of the granulator and the mixing time can be major influences on the physical properties and dissolution rate of the resulting tablets (7).

As described above, the wet granulation process is a long and hence expensive procedure, which has been improved by the introduction of high-speed mixer granulators. These have agitator and chopping blades, which enable mixing, wet massing, and granulation to take place in the same piece of apparatus. In such devices, the granulation process takes place extremely rapidly, and hence the establishment of optimum granulation times is even more important.

A further technique is fluid-bed granulation. Air is passed into the powder bed from below. This causes the particle, of powder to form a suspension in the air and gives effective mixing. The granulating fluid is then sprayed over the particles, which adhere on collision and they are then dried in the heated air stream.

The wet granulation process, apparatus, and pharmaceutical applications have been comprehensively reviewed by Kristensen and Schaefer (8) (see the article on Tablet Granulation in this encyclopedia).

Drying

After the process of granulation, the product exists as a wet mass from which the liquid must be removed, since the presence of water leads to the impairment of flow properties, and perhaps to chemical instability. Water is usually removed by evaporation for which energy is needed. This is normally provided as heat, though microwave energy is being increasingly used for drying in tablet manufacture.

The essential constituents of an effective piece of drying equipment are a heat supply to increase the temperature and thereby reduce relative humidity, a device for removal of evaporated water and a means of minimizing the distance that water molecules must diffuse before they can be evaporated.

The fluidized bed drier is the most commonly used device for drying tablet granules. The solid is fluidized from below by a jet of hot air, and so each granule becomes separated from its neighbors. The air provides an effective means of heat transfer, as well as of removing water vapor. The speed of the drying process is governed by the distance that water molecules must diffuse before they arrive at the evaporative surface. Since the wet granules are present as individual units, the maximum distance over which diffusion occurs is equal to the radius of a granule. Hence, fluidized bed drying is a rapid process.

Table 2 Binders used in the wet granulation process

Binder	Concentration in the granulating fluid (% w/v)	Comments
Acacia mucilage	Up to 20	Yields very hard granules
Alginic acid	1–5	
Carbomer	5–10	(Carbopol [®])
Carboxymethylcellulose calcium	5–15	(Nymcel [®])
Carboxymethylcellulose sodium	5–15	(Nymcel [®])
Cellulose, microcrystalline		(Avicel [®] , Emcocel [®] , Vivacel [®])
Powdered cellulose		(Elcema [®] , Solka Floc [®])
Ethyl cellulose	1–3	(Aquacoat [®])
Gelatin	5–20	Forms gel in cold water, therefore warm solution used, strong adhesive
Glucose, liquid	Up to 50	Strong adhesive, hygroscopic
Guar gum	1–10	
Hydroxyethyl cellulose	2–6	(Cellosize [®])
Hydroxypropyl cellulose	2–6	(Klucel [®] , Methocel [®])
Hydroxypropyl cellulose—low-substituted	5–25	
Hydroxypropylmethyl cellulose	2–5	(Methocel [®] , Pharmacoat [®])
Magnesium aluminum silicate	2–10	(Pharmasorb [®] , Veegum [®])
Maltodextrin	2–10	(Glucidex [®] , Lycatab [®] , Maltrin [®])
Methylcellulose	1–5	(Celacol [®] , Methocel [®])
Polydextrose		(Litesse [®])
Polyethylene oxide	5	(Polyox [®])
Povidone	0.5–5	Also known as PVP or polyvinylpyrrolidone. Soluble in water and some organic solvents, can be used for nonaqueous granulation, very commonly used, synthetic material (Kollidon [®] , Plasdone [®])
Sodium alginate	1–3	(Manucol [®])
Starch paste	5–25	Very commonly used
Starch, pregelatinized	5–10	(Lycatab [®] , Pharma-Gel [®] , Pre-Jel [®] , Sepistab [®] , Starch 1500 [®] , Starx 1500 [®])
Sucrose (syrup)	Up to 70	Hygroscopic, tablets may harden on storage
Water		Suitable for solids that are freely soluble in water

Proprietary names are given in parentheses.

The temperature of the bed can be precisely controlled, and a free-flowing product results. The resemblance to fluid-bed granulation will be apparent, and apparatus based on the fluidized bed principle is available in which mixing, granulation, and drying take place in the same chamber.

Although the apparent turbulence of the air stream may give rise to interparticulate collisions and hence attrition, this is not usually a severe problem. However, the rapid movement of particles in a hot, dry atmosphere can lead to the development of static electrical discharges. Suitable precautions must therefore be taken, especially if flammable liquids have been used in the granulation process.

A more traditional means of drying is the tray drier. Hot air flows over a series of shelves on which the wet material is spread. Compared to the fluidized bed drier, the solid–air interface is smaller, and water molecules may have to diffuse through the whole thickness of the solid layer before the evaporative surface is reached. Thinner layers give quicker evaporation, but this would reduce the overall capacity of the drier. Thus, the drying process is slower in a tray drier than in a fluidized bed drier.

As water diffuses through the bed of solid, it will carry with it any components of the formulation that are soluble in it. This will lead to a nonuniform distribution of these components in the solid. This is not usually a problem with

Table 3 Tablet glidants

Glidant	Concentration in tablet (%)	Comments
Calcium silicate	0.5–2	
Cellulose, powdered	1–2	(Elcema ^R , Solka Floc ^R)
Magnesium carbonate	1–3	
Magnesium oxide	1–3	
Magnesium silicate	0.5–2	
Silicon dioxide, colloidal	0.05–0.5	Excellent glidant (Aerosil ^R , Cab-o-Sil ^R)
Starch	2–10	
Talc	1–10	Insoluble in water but not hydrophobic

Proprietary names are given in parentheses.

fluidized bed drying, but with tray drying, significant differences in composition can occur between the upper and lower surfaces of the solid bed. This can give rise to nonuniform drug content and, if the migrating species is colored, variation in the appearance of the product (9, 10).

Microwaves are being increasingly employed in the pharmaceutical industry for drying purposes. The incident microwave radiation (frequencies of 2450 and 960 MHz are used) causes electrons in substances such as water to resonate, which in turn generates heat and causes the water to evaporate. The water vapor is removed under vacuum, and hence the product dries rapidly at a relatively low temperature. As the bed of solid is stationary, particle attrition does not occur, and dust formation is minimized (see the article on Drying and Driers in this encyclopedia).

Second mixing stage

When the drying process is complete, it is likely that the product will have cohered into relatively large masses, especially if tray drying has been used. The dried material is therefore passed through a sieve (usually 250–700 μm) to break up aggregates and to give a relatively uniformly sized granule. A second mixing stage now follows in which several important ingredients of the formulation are added.

The glidant

The formation of granules from the original powder particles may have improved flow sufficiently for uniform die filling to be achieved. However, if flow is still not good enough, a glidant (also known as an anticaking agent) can be added to improve flow still further.

The most frequently used glidant is colloidal silicon dioxide, which has a mean size of about 20 nm. It is thought to act by lodging in the surface irregularities of the granule, forming a more rounded structure and hence reducing interparticulate friction. Colloidal silica has the added advantage of acting as a moisture scavenger.

Residual water in the formulation is bound to the silica, thereby providing a drier environment for the other ingredients.

Methods of assessing glidant action have been reviewed by Augsburg and Shangraw (11). Lerk et al. showed that a concentration of 0.2% colloidal silica in a tablet formulation had no effect on tablet crushing strength. However, higher concentrations reduced crushing strength especially when associated with prolonged mixing times (12).

Some commonly used glidants are shown in Table 3.

The lubricant

When the tablet formulation is compressed, the sides of the tablet are brought into intimate contact with the die wall. The tablet must then be ejected from the die, involving the movement of the side of the tablet relative to the die wall. Therefore, friction between the tablet and the die wall must be overcome. With materials such as lactose, friction resistance can be considerable, and it may be impossible to remove the tablet from the die without damage to the tablet or to the tablet press. Therefore, a lubricant is almost invariably included in a tablet formulation. A lubricant is a substance that deforms easily when sheared between two surfaces, and hence when interposed between the tablet and the die wall, provides a readily deformable film (13).

Details of some tablet lubricants are shown in Table 4.

Inadequate lubrication can often be recognized by vertical scratches on the sides of the tablet. It may also lead to a build-up of solid on the punch faces, which in turn gives a matt, dimpled appearance to the face of the tablet, a phenomenon known as picking.

In practice, magnesium stearate is by far the most frequently used tablet lubricant, and is extremely effective. Its activity, as with other metallic salts of fatty acids, is believed to derive from adhesion of the polar metallic portion of the molecule to the powder particle surface.

Table 4 Tablet lubricants

Lubricant	Concentration in tablet (wt%)	Comments
Calcium stearate	0.5–2	Water insoluble
Fumaric acid	5	Water soluble
Glyceryl behenate	0.5–4	Water insoluble
Glyceryl palmitostearate	0.5–5.0	Water insoluble (Precirol ^R)
Hydrogenated vegetable oil	1–6	Water insoluble, may be used in conjunction with talc (Lubritab ^R , Sterotex ^R)
Magnesium lauryl sulfate	1–2	Soluble in warm water
Magnesium stearate	0.25–5	Water insoluble, excellent lubricant, reduces tablet strength, prolongs disintegration and dissolution times
Polyethylene glycol 4000 or 6000	2–5	Soluble in water, moderately effective, also known as macrogols (Carbowax ^R)
Sodium lauryl sulfate	1–2	Water soluble, moderate lubricant, but good wetting properties, often employed in conjunction with stearates (Empicol ^R , Stearowet C ^R)
Sodium stearyl fumarate	0.5–2.0	Sparsingly soluble in cold water, soluble in hot water (Pruv ^R)
Starch	2–10	Moderate lubricant
Stearic acid	1–3	Water insoluble
Talc	1–10	Insoluble in water but not hydrophobic. A moderate lubricant
Zinc stearate	0.5–2	Water insoluble

Proprietary names are given in parentheses.

As a consequence, the hydrocarbon portion of the molecule becomes oriented away from the surface (14). Thus, a nonpolar layer is presented to adjacent powder particles and structures such as the press tooling. It is from the formation of this nonpolar layer that the advantages and disadvantages of the use of magnesium stearate in a tablet arise.

To act as an effective lubricant in a tablet, the lubricant must be dispersed over the surface of the powder particles or granules. The more complete this layer, the more effective the lubricant action will be. However, this has two deleterious consequences. The first is that each powder particle presents a hydrophobic and hence water repellent exterior. It is well known that the presence of a lubricant based on fatty acids slows disintegration and dissolution, and has been shown to cause bioavailability problems.

The second consequence is that direct contact between powder particles is, at least in part, replaced by contact between adjacent hydrocarbon layers. Since these by definition have low shear strength, it is not surprising that interparticulate bonding is reduced and hence the tablet structure is weakened. Reduction in tablet strength is particularly marked with substances such as microcrystalline cellulose that undergo deformation on compression, since although the particles may change

shape, the hydrocarbon layer remains intact. Substances which fragment on compression suffer a smaller reduction in strength, since new surface, uncontaminated by lubricant, is created as the particles break up. This new surface can then take part in interparticulate bonding (15).

All these factors, both positive and negative, are consequences of the attrition of particles of lubricant and their spreading around the exterior surface of the other components of the tablet. Therefore, any processing factor that can affect lubricant attrition or the completeness of the film might be expected to influence tablet disintegration, dissolution, bioavailability, and physical strength. The mixing process is extremely important here, and mixing time, mixer type, and batch size (16) have all been shown to influence tablet properties. Thus, there is a need to establish a minimum lubricant concentration and an optimum mixing time within which adequate lubrication is achieved without the development of undesirable tablet characteristics. To ensure batch-to-batch uniformity, the parameters of the mixing process such as type of mixer, batch size, and mixing time must be kept as constant as possible. A mixing time of 2–5 min usually suffices to give adequate lubrication (17).

The water repellent properties of hydrocarbon based lubricants can be countered to a certain extent by the inclusion of a wetting agent such as sodium lauryl sulfate into the formulation. Such materials themselves can have a limited lubricant action. Mixtures of stearates and lauryl sulfates are commercially available.

Sodium stearyl fumarate has been used as an alternative for magnesium stearate. It has about the same lubricating effect, and causes similar tablet strength reduction and prolongation of disintegration time (18).

Lubricants based on fatty acids, because of their low water solubility, are unsuitable for tablets which must be dissolved in water before use. Polyethylene glycol 6000 (macrogol 6000) is soluble in water, but its lubricant activity is limited. Magnesium lauryl sulfate has been suggested as a water-soluble substitute for magnesium stearate. In addition to its lubricant action, this substance, like sodium lauryl sulfate, is an effective wetting agent (19).

It must be stressed that the functions of a glidant and lubricant in a tablet formulation are totally different. A few materials, e.g. talc, can act as both glidant and lubricant, but usually two different excipients are needed. Thus, although colloidal silicon dioxide is an excellent glidant, it has no lubricant activity. Conversely, magnesium stearate, despite its popularity as a lubricant, can hinder rather than promote flow.

The disintegrating agent

Strongly coherent particles are essential for the production of robust tablets, which will have high physical strength and low porosity. However, before it can be absorbed in the gastrointestinal tract, the active ingredient must dissolve, and a physically strong tablet is an impediment to dissolution. Therefore, tablet formulations often include a disintegrating agent, which when it comes into contact with water, disrupts the tablet structure and leads to fragmentation. A larger surface area is thus exposed to the dissolving fluid and dissolution is facilitated. Tablets which contain a large proportion of solids that are freely soluble in water have less need of a disintegrating agent, since such tablets tend to erode from their exterior surfaces rather than disintegrate.

Details of some tablet disintegrating agents are given in Table 5.

For many years, starch was the disintegrating agent of choice. Recently, however, so-called "super disintegrants" have been introduced, which markedly reduce tablet disintegration time. Such substances include croscarmellose, crospovidone, polacrillin potassium, and sodium starch glycolate (20).

The disintegrating agent may be mixed with other powders prior to wetting with the granulating fluid

(intragranular) or at the second mixing stage (extragranular), or both. Shotton and Leonard found that while extragranular disintegrating agents caused the tablet to disintegrate quicker, intragranular disintegrants not only broke down the tablet but also the constituent granules, giving a finer product (21).

The mechanism of action of disintegrating agents has been the subject of some debate (22). Some substances such as starch swell when they come into contact with water, and disruption of the tablet structure has been attributed to this. However, other effective disintegrants do not swell in this way, and are believed to act by providing a network of hydrophilic pathways inside the tablet through which water can diffuse. Irrespective of the precise mechanism of disintegration, it is clear that water uptake into the tablet must be the first step in the disintegration process (23).

Addition of wetting agents such as sodium lauryl sulfate or sodium docusate can assist this water penetration by lowering the surface tension, and they are often used in conjunction with hydrophobic lubricants such as magnesium stearate (see the article on Tablet Disintegrants and Disintegration in this encyclopedia).

Tablet Manufacture by Dry Granulation

Although widely used, the wet granulation method of tablet manufacture suffers from several disadvantages. Water is the usual granulating fluid, and this exposes tablet ingredients to the danger of hydrolysis. Furthermore, the granulating fluid has to be removed, usually by heating. In addition to the energy costs that are incurred, the elevated temperature will accelerate any hydrolytic reaction that might be taking place.

Dry granulation is an alternative method that can be used, and this process is shown in Fig. 3. The components of the formulation are compressed in the dry state. If sufficient bonding strength cannot be achieved by compression alone, a binder is added, also in the dry state.

The initial compression stage can take place by one of two methods. The first uses a conventional tablet press, a process often referred to as "slugging." Because the components of the formulation will not have the necessary attributes for producing good tablets, the tablets produced at this stage (the slugs) will not be of acceptable quality, especially as regards to appearance and weight uniformity. The slugs are then broken down to form a granular product, which after sieving can then be compressed again to give satisfactory tablets. Malkowska and Khan showed that the ease of compressibility of the formulation at the second compression was inversely proportional to

Table 5 Tablet disintegrating agents

Disintegrating agent	Concentration in tablet (wt %)	Comments
Alginic acid	2–10	
Carbon dioxide		Created in situ in effervescent tablets
Carboxymethylcellulose calcium	1–15	(Nymcel [®])
Carboxymethylcellulose sodium	1–5	(Nymcel [®])
Cellulose, microcrystalline	Up to 10	Directly compressible, some lubricant properties (Avicel [®] , Emcocel [®] , Vivacel [®])
Cellulose, powdered	5–15	Solka Floc [®]
Croscarmellose sodium	0.5–5	(Ac-di-Sol [®] , Solutab [®])
Crospovidone	2–5	(Kollidon CL [®] , Polyplasdone XL [®])
Docusate sodium	0.5–1	Acts primarily as a wetting agent
Guar gum	2–8	
Hydroxypropyl cellulose—low-substituted	5–25	
Magnesium aluminum silicate	2–10	(Veegum [®])
Methylcellulose	2–10	
Polacrillin potassium	2–10	Cation exchange resin (Amberlite IRP88 [®])
Poloxamer	5–10	
Povidone	0.5–5	(Kollidon [®] , Plasdone [®])
Sodium alginat	2.5–10	(Manucol [®])
Sodium glycine carbonate		Source of carbon dioxide for effervescent tablets
Sodium lauryl sulfate	0.5–2	Primarily a wetting agent but aids disintegration (Empicol [®])
Sodium starch glycolate	2–8	(Explotab [®] , Primojel [®])
Starch	2–10	Potato and maize starches are most frequently used
Starch, pregelatinized	5–10	(Lycatab [®] , Pharma-Gel [®] , Pre-Jel [®] , Sepistab [®] , Starch 1500 [®] , Starx 1500 [®])

Proprietary names are given in parentheses.

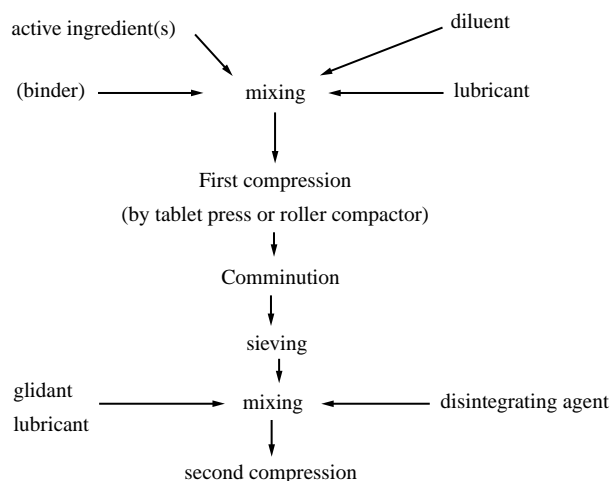


Fig. 3 The dry granulation process of tablet manufacture.

the pressure used at the slugging stage, implying that slugging at high pressure should be avoided (24).

A second method of compression is to use a roller compactor. The powder mixture is passed between two contra-rotating cylindrical rollers to form a cake, which as before is broken down to a product of granular size and then recompressed. Both methods require the addition of a lubricant prior to the first compression stage, though more lubricant will probably be needed before the second compression.

Tablet Manufacture by Direct Compression

Both wet and dry granulation methods of tablet manufacture are complex multistage processes, but are necessary to convert the components of the formulation into a state that can be readily compressed into acceptable tablets. If, however, a major component of the formulation already possesses the necessary degree of fluidity and compressibility, granulation would be unnecessary. This is the basis of the direct compression method of tablet manufacture (25).

The key component here is the diluent. This must not only possess those properties which are necessary for satisfactory tablet formulation, but also retain those properties when mixed with the other constituents of the formulation such as the active ingredient.

The process of direct compression is shown in Fig. 4. The ingredients are mixed together and then compressed. Almost invariably a lubricant must be added, and a glidant and a disintegrating agent included when necessary. The process does not involve the use of a liquid, and hence a drying stage with its attendant energy costs is avoided.

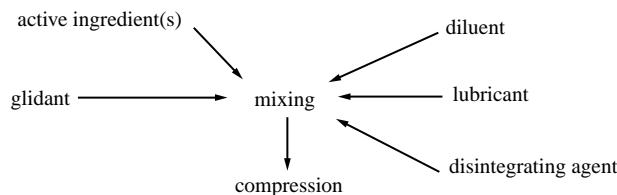


Fig. 4 The direct compression process of tablet manufacture.

Details of some direct compression diluents are given in Table 6. The majority of these are available from only one supplier, though the two most frequently used—spray-dried lactose and microcrystalline cellulose—are available from several sources.

In view of the apparent simplicity of this method of tablet manufacture and the number of suitable diluents that are commercially available, it is perhaps surprising that techniques of tablet manufacture involving granulation are still so widely used. Direct compression can, of course, only be used when a diluent is required by the formulation, i.e., the active ingredient must be relatively potent. Direct compression can offer significant savings in energy, equipment, and material handling costs. Against this must be set higher ingredient costs, since direct compression diluents are more expensive than other diluents.

There are, however, other factors which must be considered. In wet granulation, the properties of the individual drug and diluent particles are, at least to a certain extent, hidden by the binder, whereas in direct compression, the original particles are still present. Therefore, in the latter technique, a premium is placed on batch-to-batch consistency of particulate properties such as size and shape for both drug and diluent. In a direct compression formulation, the components can behave as individual particles, and therefore there is a danger that these can segregate after mixing and prior to compression. In a granulation process, the particles are bound together and so segregation is less likely to happen. Furthermore, the reduction in dust formation brought about by granulation cannot occur in direct compression.

The true direct compression process as described earlier almost invariably applies to formulations containing potent active ingredients and where the direct compression properties derive from the diluent. A few substances do possess adequate flow and cohesive properties without the need for pretreatment. These are usually crystalline inorganic salts such as sodium chloride and potassium chloride. Direct compression forms of less potent active ingredients are available e.g., paracetamol and ascorbic acid. These can be directly compressed into tablets, perhaps after the addition of a lubricant. However, such

Table 6 Direct compression tablet diluents

Diluent	Proprietary name	Comments
Calcium phosphate, dibasic	Emcompress ^R , Di-Tab ^R	Good flow properties, high density, insoluble in water
Calcium phosphate, tribasic	Tri-Tab ^R	Insoluble in water
Calcium sulfate	Compactrol ^R	Insoluble in water
Cellulose, microcrystalline	Avicel ^R , Emcocel ^R , Vivacel ^R	Highly compressible, low bulk density, acts as disintegrant
Cellulose, powdered	Elcema ^R	
Dextrates	Emdex ^R	
Lactitol	Finlac ^R DC	
Lactose		
Anhydrous alpha	Pharmatose DCL30 ^R	Good flow properties
Anhydrous beta	Pharmatose DCL21 ^R	
Spray-dried	Fast-Flo ^R , Zeparox ^R , Pharmatose DCL11 ^R	
Lactose-cellulose coprocessed mixture	Cellactose ^R	
Maltodextrin	Lycatab ^R , Maltrin ^R	Fairly soluble in water, slight lubricant effect
Mannitol	Pearlitol ^R	Freely soluble in water, negative heat of solution
Sorbitol	Neosorb ^R	
Starch, pregelatinized starch	Starch 1500 ^R , Starx 1500 ^R	Disintegrant
Sucrose–maltodextrin coprecipitate	Des-Tab ^R , Dipac ^R , Nu-Tab ^R	Good flow properties, moisture sensitive
Xylitol	Xylitab ^R	Freely soluble in water, negative heat of solution

substances are more accurately described as “pre-granulated,” in that the granulation process—either wet granulation or precompression—has been carried out by the excipient manufacturer (see the article on Direct Compression Tableting in this encyclopedia).

THE BEHAVIOR OF PARTICLES UNDER A COMPRESSIVE LOAD

All tablet manufacture can be regarded as the application of pressure to a population of particles enclosed in a confined space. An understanding of particle behavior under such conditions is therefore the key to understanding the formation and properties of tablets.

Application of a Force to Particles in a Die

Attractive forces exist between any two solid bodies. These forces may be nonspecific, e.g., van der Waal's forces, or may be more specific in nature, e.g., brought about by molecules exhibiting intermolecular hydrogen bonds. However, irrespective of their nature, it is these

forces acting among a large population of particles that enable a coherent tablet to be formed (26). Their magnitude depends directly on the particle mass and inversely on the square of the distance separating the particles. It follows, therefore, that with small particles of small mass, a tablet will only be formed when adjacent particles are forced into intimate contact with each other. This contact is brought about by the application of force.

A representation of what may happen to an individual particle when a force is applied to it can be obtained by considering what happens to a spring when subjected to a load that is applied and then removed. This is shown in Fig. 5. Although the analogy of a powder under compression to a spring undergoing elongation is not exact, it does provide useful comparisons. The load is termed the stress and the change in length the strain.

Initially there is a rectilinear relationship between stress and strain, and if the stress is removed, the spring returns to its original length. This is elastic behavior and is completely reversible. The spring is said to obey Hooke's law and the reciprocal of the slope of this portion of the curve is Young's Modulus for the spring.

If the stress is further increased, eventually a point is reached when the straight-line relationship is lost. This is

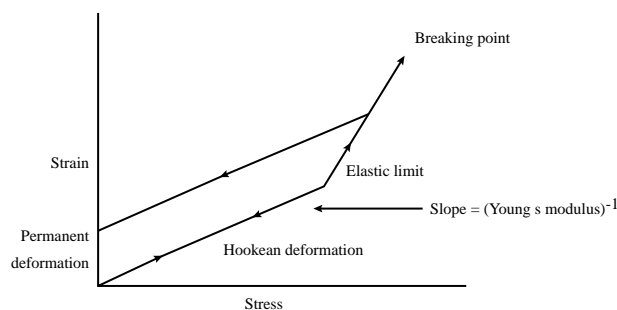


Fig. 5 Stress-strain relationships.

termed the elastic limit. If stresses in excess of the elastic limit are applied and then removed, the spring will not return to its original length. Thus, a fraction of the change in length is permanent or irreversible, and this is termed plastic behavior. Further increase in load will result in more and more plastic deformation until eventually the load is so great that the breaking point of the spring is reached and it snaps. Now, consider now a number of particles constrained in the die of a tablet press and to which a progressively increasing force is applied. A series of events can then occur, perhaps sequentially but there is a greater likelihood that some overlap will occur.

The particles will undergo rearrangement to form a less porous structure. This will take place at very low forces, the particles sliding past each other. This stage will usually be associated with some fragmentation, as the rough surfaces move relatively to one another and asperities are abraded away.

The particles have now reached the stage where relative movement becomes impossible, although the porosity of the powder bed may still be considerable. A further increase in applied force can then induce elastic deformation, plastic deformation, or fragmentation. Which of these alternatives predominates will depend on the properties of the material involved, but the net result will be a further decrease in porosity, and an increase in interparticulate contact.

If only elastic deformation has occurred, then when the compressing force is removed, the particles will return to their former shape. The additional interparticulate contact caused by compression will be lost and a coherent tablet will not be formed Fig. 6.

If, however, the elastic limit has been passed, then as the force is removed, not all the increased interparticulate contact will be lost, cohesion will be retained and a tablet will be formed. Thus, from the point of view of forming a robust tablet, substances with low elastic limits, which undergo plastic deformation at low compressive forces, are preferable to more elastic bodies.

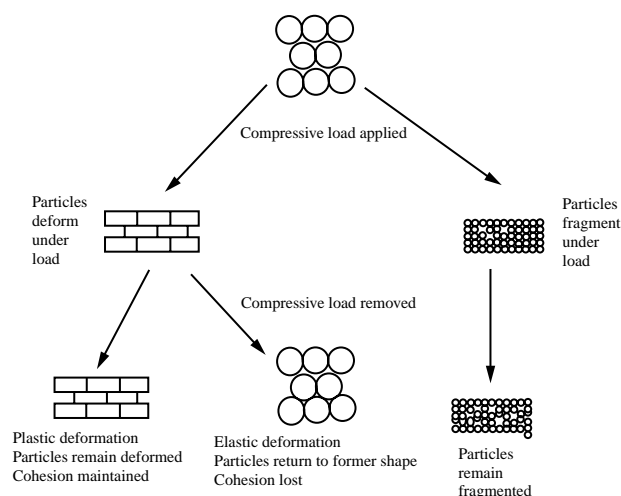


Fig. 6 Elasticity and plasticity in a particulate mass.

If consolidation of the powder mass is brought about by fragmentation, then a large number of points of interparticulate contact are created, from which the strength of the tablet derives. In this case, removal of the compressing force should have no effect on tablet strength, since there is no way the fragments can recombine into the original particles. However, purely fragmentary consolidation is unlikely to have occurred, and so the effect of removal of the force on deformed particles must still be considered.

Force Transmission through a Powder Bed

Consider as before a group of particles in the die of an excentric tablet press. Force is applied by means of the descending upper punch and because the lower punch is passive, the force will be transmitted to it through the powder bed. The distribution of force within the powder bed was investigated by Train, who embedded force transducers (q.v.) in a relatively large mass of powder (27). He found that the diminution of force did not proceed uniformly on descent through the bed, but formed a much more complex pattern. This was caused by the forces being transmitted to and reflected from the die wall. Significant features are zones of high force at the periphery near the moving punch, and much lower in the powder mass on its vertical axis. On the other hand, low force zones occur on the same axis but much nearer to the moving punch Fig. 7. Train's findings were later confirmed by Charlton and Newton using gamma-ray attenuation (28).

The consequences of such a force distribution on tablet strength can be profound. Particle deformation, whether elastic or plastic, will be proportional to the force applied,

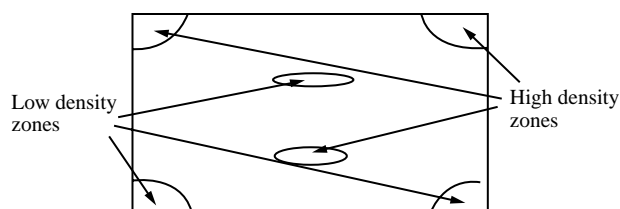


Fig. 7 Density distribution in a compact prepared on an eccentric tablet press.

and as has been discussed, this deformation is an essential preliminary to the formation of the interparticulate bonds on which tablet integrity depends. Thus, the porosity of the tablet, and hence its strength, will vary within the tablet. The weakest points in the tablet structure will be those that receive the lowest force i.e., on the face of the tablet adjacent to the stationary punch and on the central axis near to the moving punch. Thus, because of its nonuniform density, some parts of a tablet are stronger than others.

It should be noted that this discussion assumes that only one punch is actively applying the force to the powder mass while the other is stationary and passive. This is true in the case of an eccentric press, but with a rotary tablet press, both punches move and hence both exert forces on the powder bed. The force distribution so obtained is thus different from that shown in Fig. 7, and results in two low density zones near the faces of the tablet and a high density zone in approximately the centre of the powder mass.

The effect of the removal of the compressing force must now be considered. Elastic recovery will occur to a greater or lesser extent, which will result in a reduction in the strength of interparticulate bonds and an overall weakening of the tablet. It therefore follows that if a tablet is to be disrupted by elastic recovery, this is most likely to occur at its weakest point. This is just below the top surface, and is the phenomenon often encountered in tablet manufacture known as lamination or capping. With this explanation in mind, some effects associated with capping, and some causes and pragmatic solutions to the problem can now be explained.

Capping was for many years considered to be due to the entrapment of air in the tablet, and even the production of tablets in vacuo which still capped did little to dispel this theory. Neither did this suggestion explain why air should cause the fracture just below the face of the tablet. However, by considering the nonuniform density distribution in the tablet, it can be seen that the weakness is not caused by the presence of air per se, but rather the relative absence of solid in those parts of the tablet that have high porosity (29). As compression proceeds, it follows that the pores in the tablet structure are filled with air at

a progressively elevated pressure, and this will obviously assist disruption of the tablet when the compressing force is removed. Thus, any factor which obstructs the expulsion of air from the powder mass during compression will exacerbate capping, though it is not the fundamental cause. Such factors include the clearance between punch and die, the speed at which the force is applied, and the presence of small particles, which makes passage of air through the tablet more tortuous (30).

Any applied stress that exceeds the breaking strength of the tablet will also cause the tablet to break at its weakest point. A number of stresses occur when the tablet is removed from the die after compression. The die may become worn at the point in the die where the tablet is compressed, i.e., the die is fractionally wider at this point than elsewhere. Thus, when the tablet is ejected, it is forced through an aperture, the diameter of which is slightly less than that of the tablet itself. This will obviously stress the tablet, and the interparticulate bonds may be overcome at their weakest point. Also as the tablet is extruded from the die, elastic expansion will occur not just in an axial but in a radial direction. The latter occurs progressively, i.e., one segment of tablet is free to expand while the one below is still constrained by the die. Bond disruption will be an inevitable consequence.

CHARACTERIZATION OF THE COMPACTION PROCESS

A range of parameters has been devised which attempts to describe the process of powder compaction, both to elucidate principles and also to enable predictions to be made regarding compaction properties. The majority of these depend on the availability of methods of accurately measuring applied force and punch position.

Measurement of Applied Force and Punch Movement

The aim of any tableting process is to produce tablets that are of satisfactory quality. Virtually all tablet properties e.g., porosity, physical strength, disintegration time, dissolution time are dependent in some way on the force that is applied by the punches to the particles in the die.

Considerable research on tablet properties was performed for many years, but until a method of accurately measuring compression force was available, meaningful studies could not be carried out.

The key to progress in this field was the introduction of the so-called instrumented tablet press by Higuchi and

others in the mid 1950s (31), in which force transducers were fitted to the press to measure the applied load. This revolutionized research into the tableting process and in addition has led to the development of presses with automatic tablet weight control, since the mass of particles in the die governs the force detected by the transducer (32) (see the article on Automation of Tablet Presses in this encyclopedia).

A force transducer is also known as a strain gauge. In its simplest form, the strain gauge is a network of wires through which an electric current is passed. The wires are bonded very securely to a component of the press, e.g., the upper punch. If a force is applied to the punch, it deforms. The magnitude of the deformation (the strain) is governed by a combination of the magnitude of the force (the stress) and the value of Young's modulus for the material from which the punch is made. The wire of the strain gauge is also deformed, and hence its electrical resistance changes. If the gauge is incorporated into a Wheatstone bridge circuit, then a small change in voltage results. The size of the signal from the strain gauge is proportional to the amount of deformation, which in turn is a function of the applied force. Hence, after amplification and appropriate calibration, the voltage changes can be expressed in terms of the applied force. Signals from the transducers can be fed into an oscilloscope or chart recorder, or stored electronically and subsequently manipulated by computer.

A further advance was the fitting of displacement transducers to tablet presses. These too give out an electrical signal, the magnitude of which is governed by the position of a sensing device in relation to a fixed reference point, and in this way punch movement can be measured. As before, these signals can be amplified, recorded, and stored. If the outputs of the force and displacement transducers are combined, the applied force at a given point in time, and punch position at the same instant can be established.

Instrumentation of rotary presses is more difficult than that for an excentric press if transducers are to be fitted directly to the punches. Since the turret of the press rotates, a fixed connection between transducers and the recording device is impracticable. Alternatives to fixed connections are slip rings and radiotelemetry. A further option is to fit transducers to points of the press remote from the punches and which do not rotate. Techniques for fitting transducers to both excentric and rotary tablet presses, their calibration, and applications of the information derived from instrumented presses have been reviewed by Watt (33).

Fig. 8 shows the output from force and displacement transducers fitted to both punches of an excentric tablet press. The upper punch describes an approximately sinusoidal path as it descends to penetrate the die (point

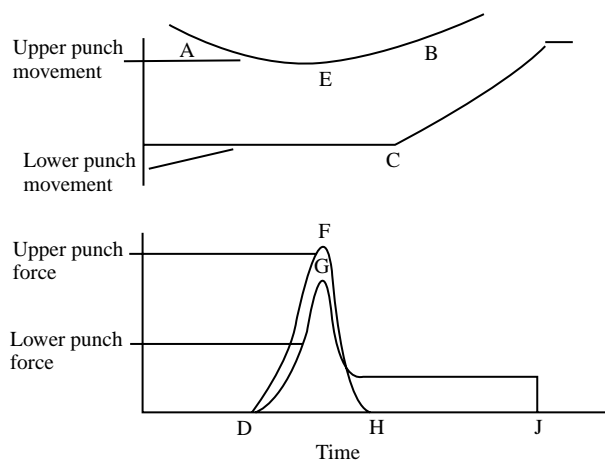


Fig. 8 Force and displacement data from an instrumented excentric tablet press.

A) and then rises after the compression event has taken place, leaving the die at point B. The lower punch remains motionless during the compression event and then rises to eject the tablet from the die (point C).

As the upper punch enters the die and comes into contact with the particles, the height of the bed of particles is reduced and hence porosity decreases. Initially porosity reduction is brought about by particle rearrangement. This requires a very low force which is probably not detectable by the force transducers, the output of which remains zero. The upper punch then encounters a resistance to its motion (point D) as further consolidation by rearrangement becomes impossible. Hence, the output of the upper punch force transducer rises, slowly at first and then more rapidly. Particles are deformed and/or fragmented during this stage to form a coherent tablet. Force is transmitted to the lower punch, and a similar rise is detected by transducers there. As maximum upper punch penetration is achieved (point E), force maxima are detected on both punches (points F and G), that on the lower punch being less than that on the upper.

Once the maxima have occurred, the upper punch begins to rise, and the force detected on both punches falls. That on the upper punch returns to zero as contact is lost between the ascending punch and the top surface of the tablet (point H). That on the lower punch does not fall to zero until ejection is complete (point J). The greater the ejection force, the greater the need for a lubricant in the formulation.

The reason why the lower punch maximum force is less than that of the upper punch is because a fraction of the force applied by the upper punch is transmitted to the die wall, where it results in die wall friction. This is reduced

by the presence of a lubricant. Hence, the ratio between lower punch maximum force and upper punch maximum force, often called the R value, is used as the basis of comparison between lubricants (34). R has a maximum value of unity and lubricants based on stearates usually exhibit R values greater than 0.95.

The time that elapses from point D to point F is known as the consolidation time. This is the time interval when a force is detectable at the upper punch. The contact time is the period when the upper punch is in contact with the original particles or the tablet (point D to point H). The residence time is the period when a force is detected at the lower punch (point D to point J), which ends on tablet ejection.

Transducer output from a rotary tablet press differs in two aspects. Firstly, the lower punch plays an active role in the compression event and moves upwards as the upper punch moves downwards. The second difference is small but important. Because of the sinusoidal movement of the upper punch in an excentric press, the punch speed is only zero at the instant when it reverses direction (point E). Punches on a rotary press have a flat area on the punch head. As the punches pass under or over the pressure rolls, the flat area dictates that there is no punch movement. The period during which this occurs is called the dwell time, and though it only lasts a fraction of a second, it can have a major effect on the consolidation process (35, 36).

Tablet Strength Profiles

The physical strength of a tablet is dependent on the extent and strength of interparticulate bonds and these in turn are related to the compressive force which is applied. Therefore, the relationship between the applied force and some parameter related to tablet strength is a good indication of the ease with which a given substance will form satisfactory tablets, and may also give an insight into the compaction mechanism of the solid and its mechanical properties.

The strength is usually assessed as the force required to fracture a tablet in a defined direction e.g., its diameter. This has traditionally been referred to as the "hardness" of the tablet, which is incorrect terminology in this context. In material science, hardness is a surface property related to resistance to indentation. Terms such as physical strength or mechanical strength are more appropriate.

The compression force is measured using an instrumented tablet press as described earlier. The physical strength of the tablet is measured with crushing apparatus such as that described in most current pharmacopoeias. The results are conventionally presented graphically with compression force as the abscissa and strength as the ordinate.

This technique is satisfactory when comparing tablets of the same size and shape, such as in-process control. However, if virtually identical tablets are not being tested, problems may arise. The force is sometimes expressed as the compression pressure, obtained by dividing the force by the cross-sectional area of the punch. This is valid if the punch has a flat surface, the area of which can be easily calculated; but in practice this is often not the case, and conversion between compression force and compression pressure involves assumptions regarding the area of the punch face, which may not be valid.

The physical strength of a tablet is also dependent on its dimensions. In the construction of a force-strength profile, all tablets will have the same cross-sectional area as the same tooling will have been used. However, as the compressive force is changed, so will the tablet height. Hence, comparisons made on the basis of breaking strength will not be truly valid.

This problem has been circumvented in part by the calculation of the tensile strength of the tablet. The most commonly used formula is shown in Eq. 1, introduced by Rudnick et al. (37) and Fell and Newton (38).

$$T_s = 2P/\pi dt \quad (1)$$

where T_s is the tensile strength of the tablet (MPa), P the crushing strength (N), d the tablet diameter (m), and t the tablet thickness (m).

This equation applies to cylindrical tablets which have a diameter and whose height is constant over the whole tablet surface. Newton et al. have attempted to extend the concept of tensile strength to tablets that are not cylindrical (39).

Confusion can often arise in the units that are used to express compression force or pressure and the strength of the tablet. Table 7 gives examples of units that have been used recently for the axes of tablet strength profiles. Comparison between different tablet formulations would be greatly facilitated if authors used and journal editors insisted on the use of the SI system of units. The SI unit of length is the meter, that of force the Newton and that of pressure the Pascal. The unit for physical strength is the Newton and that of tensile strength calculated from Eq. 1 is the Pascal.

Relationships between Tablet Porosity and Compression Force or Pressure

Fig. 8 shows the movement of both the upper and lower punches of the tablet press in relation to fixed reference points and from these data, the distance separating the two

Table 7 Some units used for the construction of tablet strength profiles

Abscissa parameter	Unit	Ordinate parameter	Unit
Force	kg	Crushing strength (hardness)	kg
	lb		Strong-Cobb units
	kN		N
	N		kp
Pressure	Kg cm ⁻²	Tensile strength	Kg cm ⁻²
	Pa		Pa
	MPa		MPa
	lbin ⁻²		

punch faces can be calculated. If both punch faces are in contact with the tablet, it follows that the distance of their separation is equal to the height of the tablet (h). Consequently, if the cross-sectional area of the die (a), the tablet weight (w) and the true density of the solid from which the tablet is made (ρ) are all known, the porosity of the tablet (ϵ) can be calculated from Eq. 2.

$$\epsilon = 1 - (w/ah\rho) \quad (2)$$

For cylindrical tablets, Eq. 2 becomes Eq. 3

$$\epsilon = 1 - (4w/\pi d^2 h \rho) \quad (3)$$

As the applied force (or pressure) being applied at the same moment in time can also be determined from Fig. 8, it follows that a relationship between force and porosity can be constructed.

A typical relationship is shown in Fig. 9. As force is increased from zero, the porosity of the tablet falls rapidly, but then further increase in force has a progressively smaller effect. The porosity value at which the curve becomes virtually horizontal is dependent on the solid being compressed. Substances that deform plastically typically give tablets of lower porosity than those which fragment.

Many equations have been derived in attempts to provide a mathematical expression of Fig. 9. These have been reviewed by Kawakita and Ludde (40). It must be stressed that such equations are simply mathematical descriptions of the data, and they have no underlying physical significance.

The most widely used of these equations is the Heckel relationship (41):

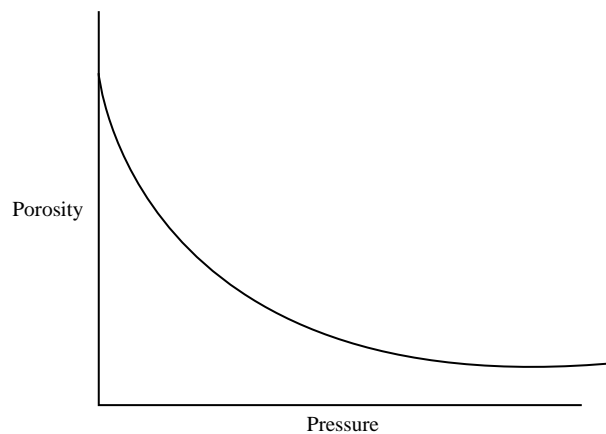
$$\ln 1/(1 - D) = kP + A$$

where D is the relative density of the tablet and hence $(1 - D)$ is the porosity, P is the applied pressure, and k and A are constants.

This equation predicts that a graph of $\ln 1/(1 - D)$ against P should yield a straight line of slope k and intercept A . Heckel surmised that the greater the slope, the greater degree of plasticity in the solid being compressed. Hersey and Rees (42) defined the reciprocal of k to be the mean yield pressure of the solid.

Fig. 10 shows a typical Heckel diagram. Deviations from linearity often occur at low and high pressures. These are to be expected. At low pressures, reduction in porosity is largely by particle rearrangement, and thus the true consolidation mechanism, i.e., fragmentation or deformation, will be a minor component of the total consolidation process. At high pressures, porosity can become very low and hence its reciprocal becomes a very large number.

However, the real problem with Heckel plots is identifying a truly rectilinear section. When pressure–porosity diagrams were first devised, instrumented tablet presses were not widely available. The height of the tablet was measured after ejection from the die and thus the net

**Fig. 9** A force–porosity diagram.

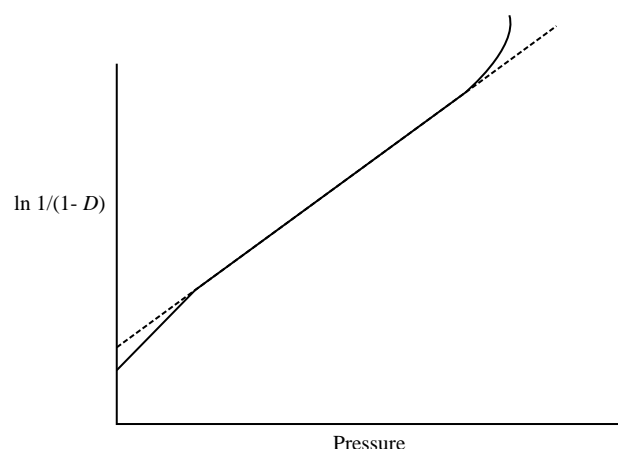


Fig. 10 Pressure and porosity data plotted according to the Heckel equation.

change in height was a combination of height reduction by consolidation and a height increase by elastic expansion. Furthermore, the typical plot comprised only a small number of points, and an apparently rectilinear zone could be identified. With improvements in instrumentation and in particular linkage to a computer, Heckel plots of several hundred points became feasible with tablet height measurements being made before ejection from the die. In the author's experience, the more points there are on a pressure–porosity diagram, the more difficult it is to identify a truly rectilinear section.

Paronen and Ilkka have surveyed the use of the Heckel and other compression equations and have drawn attention to difficulties in their interpretation (43).

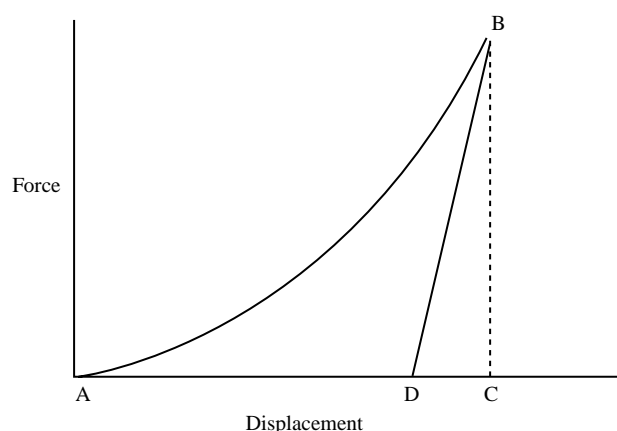


Fig. 11 A force-displacement curve.

Force-Displacement Curves

Fig. 8 shows how force and punch displacement change with time. It is therefore possible to determine both the force and displacement values at any given point in time, and from this, a force–displacement curve can be constructed. Usually the force is plotted as the ordinate and displacement as the abscissa Fig. 11.

The area enclosed by the curve has the units of force (Newtons) multiplied by distance (metres), which is dimensionally equivalent to work or energy (in Joules). Therefore, the force–displacement curve has been used to calculate the work consumed in the compression of a solid (44).

Further refinement of this technique has enabled estimates of elasticity and plasticity of the solid to be made. If the particles were totally nonelastic, then as soon as the upper punch reached its maximum penetration and began to rise again, contact would be lost with the tablet. Therefore, the force recorded by transducers fitted to the upper punch would instantly fall to zero. This clearly does not happen, as demonstrated by the upper punch force values after maximum punch penetration in Fig. 8. This can only mean that as the upper punch ascends, the tablet expands elastically, keeping in contact with the punch face. Thus, the area ABC in Fig. 11 represents the work expended on compressing the solid, and area DBC is the work delivered back to the upper punch by the expanding tablet. The difference between these two areas (i.e., ABD) has been termed the “Net work” of compression.

Use of this technique has shown that the presence of a granulating agent causes a marked increase in the plasticity of the powder mass, with a consequent increase in cohesion and tablet strength. The film of granulating agent between the particles can be regarded as a highly viscous liquid with a large yield value. Application of a force in excess of the yield value causes granules to deform, and this becomes permanent when the force is reduced to below the yield value (45). A somewhat similar mechanism is believed to account for the properties of some direct-compression tablet diluents, e.g., spray-dried lactose, which consists of small crystalline masses embedded in an amorphous and more easily deformed matrix.

The force–displacement curve has been extensively used to characterize the compression event. However, it depends on the availability of sensitive and properly calibrated transducers. Even small inaccuracies in measuring displacement at high force values can have a profound effect on the resultant value of the area under the curve. Ragnarsson has provided a very thorough review of the uses and potential pitfalls of the force–displacement curve (46).

Time-Dependent Effects

It has long been known that for some formulations, changing the speed of operation of the tablet press or changing the type of press can have a profound effect on tablet quality. Such circumstances may arise, for example, when changing from a slow excentric press in formulation development to a high speed rotary in production.

As described earlier, diagrams of transducer output against time enable parameters such as contact time and dwell time to be defined Fig. 8. Also the slope of the displacement–time diagram equals the speed of the punch and the slope of the force–time diagram is the rate of change of the force. If the operating speed of the press is altered, there is a proportional change in all of these.

Such considerations are important because the consolidation of many substances is time-dependent. Fragmentation can be regarded as a virtually instantaneous process. However, substances that undergo deformation behave in a viscoelastic manner, and the time over which the compression force is applied can be crucial. As the time over which they are compressed is reduced, such materials show less consolidation and this in turn can affect the physical strength of the tablet. For example, Armstrong and Palfrey (47) studied the effects of punch speed change on the tablet strength of four direct compression diluents. Though tablet strength was reduced in all cases, solids such as modified starch suffered the greatest reduction. Roberts and Rowe investigated the relative sensitivity of the yield pressures of a range of solids to changes in compression speed, and again found that those solids which deformed on compression showed the greatest change (48).

The rate of consolidation in an excentric press is governed by the speed at which the upper punch moves into the die. This in turn is governed by the lengths of the excentric arms of the press and its speed of operation. In a rotary press, punch speed is governed by the diameter of the die table, the diameter of the pressure roll, the geometry of the punch head, and the speed of operation. Formulae for the calculation of punch speeds at any point of the compression cycle for both types of press have been derived (49, 50).

It is important to distinguish the punch speed from the output of the press. Though rotary presses have a much higher output than excentric presses, this largely derives from their multiplicity of punch stations. It does not necessarily follow that their punch speeds during compaction are significantly greater (51).

If the work of compression in Joules is divided by the time in seconds over which compression occurs, then the power of compression (in Watts) is obtained.

An alternative method of calculating power of compression is to multiply the force by the punch speed when that force was applied. This too can be obtained from the data in Fig. 8. This permits the compression events on different presses to be compared (49).

REFERENCES

1. Editorial: The Passing of the Tablet Fad, *Pharm. J.* **1895**, February 12.
2. Castello, R.A.; Mattocks, A.M. Discoloration of Tablets Containing Amines and Lactose, *J. Pharm. Sci.* **1962**, *51* (1), 106–108.
3. Carstensen, J.T.; Ertell, C. Physical and Chemical Properties of Calcium Phosphate for Solid State Pharmaceutical Formulations, *Drug Dev. Ind. Pharm.* **1990**, *16* (7), 1121–1133.
4. Campbell, H.; Bauer, W.C. Cause and Cure of Demixing in Solid–Solid Mixers, *Chem. Eng.* **Sept. 1966**, 179–185.
5. Staniforth, J.N.; Rees, J.E.; Kayes, J.B. Relation between Mixing Time and Segregation of Ordered Mixes, *J. Pharm. Pharmacol.* **1981**, *33*, 175–176.
6. Sherrington, P.J.; Oliver, R. *Granulation*; Heyden and Sons: London, 1981.
7. Ertel, K.D.; Zoglio, M.A.; Ritschel, W.A.; Carstensen, J.T. Physical Aspects of Wet Granulation. IV. Effect of Kneading Time on Dissolution Rate and Tablet Properties, *Drug Dev. Ind. Pharm.* **1990**, *16* (6), 963–981.
8. Kristensen, H.G.; Schaefer, T. *Granulation*; *Drug Dev. Ind. Pharm.* **1987**, *13* (4/5), 803–872.
9. Rubinstein, M.H.; Ridgway, K. Solute Migration during Granule Drying, *J. Pharm. Pharmacol.* **1974**, *26*, 24P–29P.
10. Armstrong, N.A.; March, G.A. Quantitative Assessment of Surface Mottling of Colored Tablets, *J. Pharm. Sci.* **1974**, *63* (1), 126–129.
11. Augsburger, L.L.; Shangraw, R.F. Effect of Glidants in Tableting, *J. Pharm. Sci.* **1966**, *55* (4), 418–423.
12. Lerk, C.F.; Bolhuis, G.K.; Smedama, S.S. Interactions of Lubricants and Colloidal Silica During Mixing with Excipients, *Pharm. Acta Helv.* **1977**, *52* (3), 33–39.
13. Bowden, F.P.; Tabor, D. *The Friction and Lubrication of Solids*, 2nd Ed.; Clarendon Press: Oxford, 1964.
14. Miller, T.A.; York, P. Pharmaceutical Tablet Lubrication, *Int. J. Pharm.* **1988**, *41*, 1–19.
15. De Boer, A.H.; Bolhuis, G.K.; Lerk, C.F. Bonding Characteristics by Scanning Electron Microscopy of Powders Mixed with Magnesium Stearate, *Powder Tech.* **1978**, *20*, 75–82.
16. Bolhuis, G.K.; Holzer, A.W. Lubricant Sensitivity. *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nystrom, C., Eds.; Marcel Dekker, Inc.: New York, 1996; 517–560.
17. Ragnarsson, G.; Holzer, A.W.; Sjogren, J. Influence of Mixing Time and Colloidal Silica on the Lubricating Properties of Magnesium Stearate, *Int. J. Pharm.* **1979**, *3*, 127–131.

18. Holzer, A.W.; Sjogren, J. Evaluation of Sodium Stearyl Fumarate as a Tablet Lubricant, *Int. J. Pharm.* **1979**, *2*, 145–153.
19. Salpekar, A.M.; Augsburger, L.L. Magnesium Lauryl Sulfate in Tableting: Effect on Ejection Force and Compressibility, *J. Pharm. Sci.* **1974**, *63* (2), 289–293.
20. Shangraw, R.F.; Mitrevej, A.; Shah, M. A New Era of Tablet Disintegrants, *Pharmaceutical Technology* **1980**, *Oct.*, 49–57.
21. Shotton, E.; Leonard, G.S. Effect of Intragranular and Extragranular Disintegrants on the Particle Size of Disintegrated Tablets, *J. Pharm. Sci.* **1976**, *65* (8), 1170–1174.
22. Lowenthal, W. Disintegration of Tablets, *J. Pharm. Sci.* **1972**, *61* (11), 1695–1711.
23. van Kamp, H.V.; Bolhuis, G.K.; de Boer, A.H.; Lerk, C.F.; Lie-A-Huen, L. The Role of Water Uptake on Tablet Disintegration, *Pharm. Acta Helv.* **1986**, *61* (1), 22–29.
24. Malkowska, S.; Khan, K.A. Effect of Recompression on the Properties of Tablets Prepared by Dry Granulation, *Drug Dev. Ind. Pharm.* **1983**, *9* (3), 331–347.
25. Bolhuis, G.K.; Chowhan, Z.T. Materials for Direct Compression. *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nystrom, C, Eds.; Marcel Dekker, Inc. New York, 1996; 419–500.
26. Nystrom, C.; Karehill, P.G. Intermolecular Bonding Forces. *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nystrom, C, Eds.; Marcel Dekker, Inc.: New York, 1996; 17–53.
27. Train, D.; Lewis, C.J. Agglomeration of Solids by Compaction, *Trans. Inst. Chem. Engrs.* **1962**, *40*, 235–263.
28. Charlton, B.; Newton, J.M. Application of Gamma-Ray Attenuation to the Determination of Density Distributions Within Compacted Powders, *Powder Technol.* **1985**, *41*, 123–134.
29. Ritter, A.; Sucker, H.B. Studies of Variables that Affect Tablet Capping, *Pharm. Technol.* **1980**, July, 24–34.
30. Mann, S.C.; Roberts, R.J.; Rowe, R.C.; Hunter, B.M.; Rees, J.E. The Effect of High Speed Compression At Sub-Atmospheric Pressures on the Capping Tendency of Pharmaceutical Tablets, *J. Pharm. Pharmacol.* **1983**, *35*, 44P.
31. Higuchi, T.; Nelson, E.; Busse, L.W. The Physics of Tablet Compression. 3. The Design and Construction of an Instrumented Tableting Machine, *J. Am. Pharm. Assn. Sci. Ed.* **1954**, *43*, 344–348.
32. Murray, F.J. Tablet Press Automation: A Modular Approach to Fully Integrated Production, *Drug Dev. Ind. Pharm.* **1996**, *22* (1), 35–43.
33. Watt, P.R. *Tablet Machine Instrumentation in Pharmaceutics—Principles and Practice*; Ellis Horwood: Chichester, 1988.
34. Nelson, E.; Naqvi, S.M.; Busse, L.W.; Higuchi, T. The Physics of Tablet Compression. 4. Relationship of Ejection and Upper and Lower Punch Forces during the Compression Process, *J. Am. Pharm. Assn. Sci. Ed.* **1954**, *43*, 596–602.
35. Akande, O.F.; Ford, J.F.; Rowe, P.H.; Rubinstein, M.H. The Effects of Lag-Time and Dwell-Time on the Compaction Properties of 1:1 Paracetamol/Microcrystalline Cellulose Mixtures Prepared by Precompression and Main Compression, *J. Pharm. Pharmacol.* **1998**, *50*, 19–28.
36. Munoz-Ruiz, A.; Jiminez-Castellanos, M.R.; Cunningham, J.C.; Katdare, A.V. Theoretical Estimation of Dwell and Consolidation Times in Rotary Tablet Machines, *Drug Dev. Ind. Pharm.* **1992**, *18* (19), 2011–2028.
37. Rudnick, A.; Hunter, A.R.; Holden, F.C. An Analysis of the Diametral Compression Test, *Mater. Res. Stand.* **1963**, *3*, 283–289.
38. Fell, J.T.; Newton, J.M. Determination of Tablet Strength by the Diametral Compression Test, *J. Pharm. Sci.* **1970**, *59* (5), 688–691.
39. Pitt, K.G.; Newton, J.M.; Richardson, R.; Stanley, P. Material Tensile Strength of Convex-Faced Aspirin Tablets, *J. Pharm. Pharmacol.* **1989**, *41*, 289–292.
40. Kawakita, K.; Ludde, K.H. Some Consideration on Powder Compression Equations, *Powder Technol.* **1970**, *4*, 61–68.
41. Heckel, R.W. Density-Pressure Relationships in Powder Compaction, *Trans. Metall. Soc. AIME* **1961**, *221*, 671–675.
42. Hersey, J.A.; Rees, J.E. Deformation of Particles during Briqueting, *Nature* **1971**, *230*, 96.
43. Paronen, P.; Illka, J. Porosity-Pressure Functions. *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nystrom, C, Eds.; Marcel Dekker, Inc.: New York, 1996; 55–75.
44. de Blaey, C.J.; Polderman, J. Compression of Pharmaceuticals. I the Quantitative Interpretation of Force-Displacement Curves, *Pharm. Weekblad* **1970**, *105*, 241–250.
45. de Blaey, C.J.; van Oudtshoorn, M.C.B.; Polderman, J. Compression of Pharmaceuticals. III. Study on Sulphadimidine, *Pharm. Weekblad* **1971**, *106*, 589–599.
46. Ragnarsson, G. Force-Displacement and Network Measurements. *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nystrom, C, Eds.; Marcel Dekker, Inc. New York, 1996; 77–97.
47. Armstrong, N.A.; Palfrey, L.P. The Effect of Machine Speed on the Consolidation of Four Directly Compressible Tablet Diluents, *J. Pharm. Pharmacol.* **1989**, *41*, 149–151.
48. Roberts, R.J.; Rowe, R.C. The Effect of Punch Velocity on the Compaction of a Variety of Materials, *J. Pharm. Pharmacol.* **1985**, *37*, 526–528.
49. Armstrong, N.A.; Abourida, N.M.A.H.; Gough, A.M. A Proposed Consolidation Parameter for Powders, *J. Pharm. Pharmacol.* **1983**, *35*, 320–321.
50. Rippie, E.G.; Danielson, D.W. Visco-Elastic Stress-Strain Behavior of Pharmaceutical Tablets: Analysis during Unloading and Postcompression Periods, *J. Pharm. Sci.* **1981**, *70* (5), 476–482.
51. Armstrong, N.A. Time-Dependent Factors Involved in Powder Compression and Tablet Manufacture, *Int. J. Pharm.* **1989**, *49*, 1–13.

Tablet Manufacture by Direct Compression

Norman Anthony Armstrong

Welsh School of Pharmacy, Cardiff University, Cardiff, U.K.

INTRODUCTION

All tablets are made by compressing a particulate solid between two punches in a die of a tablet press. For an active ingredient to be transformed into tablets of satisfactory quality, the formulation must have three essential attributes.

First, the formulation must flow into the die space of the tablet press sufficiently rapidly and in a reproducible manner. Otherwise, unacceptable variation in tablet weight and content of active ingredient will ensue.

Second, the particles in the formulation must cohere when subject to a compressing force, and that coherence should remain after the compressive force has been removed.

Third, after the compression event is complete, it must be possible for the tablet to be removed from the press without damage to either the tablet or the press.

Very few active ingredients possess all three of these essentials and some possess none of them. Hence some preliminary treatment is almost invariably necessary.

METHODS OF TABLET MANUFACTURES

There are three methods of tablet manufacture designed to confer these essential attributes to a tablet formulation. Wet granulation and direct compression are the most important, with dry granulation (also known as precompression or slugging) used in some circumstances. Fig. 1 shows the processes of wet granulation and direct compression broken down into their constituent stages.

The relative simplicity of the direct compression process is immediately apparent.

Ease of removal of the tablet from the press is, in theory at least, readily achieved. Friction occurs between the tablet and the die and punches of the press, which can be overcome by including a lubricant in the formulation. Hence every formulation, irrespective of the method of manufacture, will include a lubricant. This will usually be a metallic salt of a fatty acid such as magnesium stearate.

The other two prerequisites—flow and cohesion—can only be achieved by more elaborate techniques and are in

fact the reasons why the wet and dry granulation processes were devised.

As part of its complexity, wet granulation involves the addition of a liquid (usually water), followed by its removal, normally by evaporation. In addition to the energy requirements of this drying process, the presence of water might bring about hydrolysis of the active ingredient, which will be exacerbated at the elevated temperatures used for drying.

If a major component of the formulation such as the diluent were to possess the necessary degrees of fluidity and compressibility, granulation would be unnecessary. This is the basis of the direct compression method of tablet manufacture.

Prior to the early 1960s, there were very few materials which possessed these properties. Little and Mitchell^[1] in their text *Tablet Making*, cite sodium chloride and bromide, potassium chlorate, bicarbonate and iodide, ammonium chloride, and hexamine as having properties which permit tableting without some form of prior treatment. No reference is made by name to the process of direct compression.

At about the same time, two materials were introduced that were specifically designed to act as tablet diluents and would not require preliminary treatment. These were spray-dried lactose and microcrystalline cellulose, introduced in 1962 and 1964, respectively. These two substances can be said to have initiated the “direct compression revolution.” Since that time, a wide range of direct compression tablet diluents has become available. The properties of some of these materials will be reviewed later in this article.

It is important to distinguish between true direct compression diluents (i.e., excipients) and active ingredients which are available in a direct compression form. These are usually high dose materials such as aspirin, paracetamol, and ascorbic acid. They can be directly compressed into tablets, the only pretreatment being mixing with a lubricant and perhaps a disintegrating agent. However, such substances are more accurately described as “pregranulated” since the granulation process, either wet or dry, will have been carried out by the excipient manufacturer. It is likely that such materials will contain a

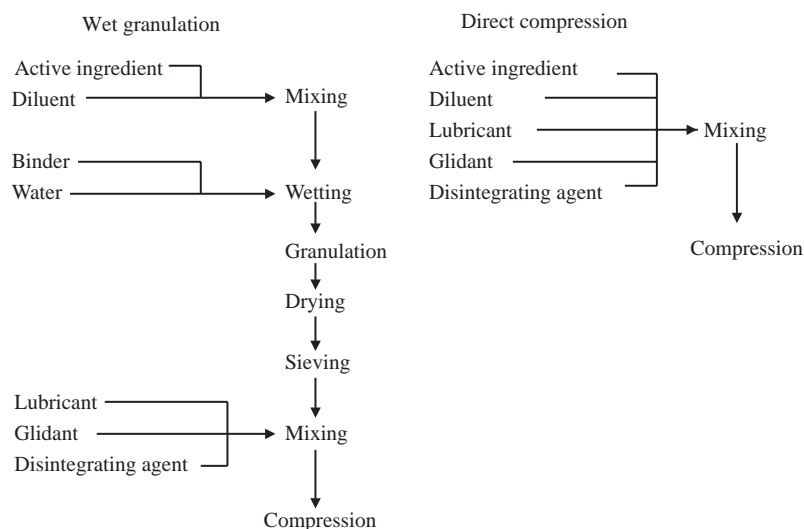


Fig. 1 Comparison of the wet granulation and direct compression processes of tablet manufacture.

binder. For example, ascorbic acid pregranulated with either starch or hydroxypropyl cellulose is commercially available.^[2]

The perceived advantages of the direct compression process of tablet manufacture have given rise to a considerable body of literature. Between 1970 and the end of 2000, there were 598 references to “direct compression” in the index of *International Pharmaceutical Abstracts*. It has been estimated that today, some 40 years after the introduction of diluents specifically designed for direct compression, about 50% of worldwide tablet production is made by this method.^[2] The question must be asked why a process which has so many apparent advantages and for which suitable materials seem to be plentifully available has not made a greater impact. An innate conservatism in the pharmaceutical industry is perhaps a factor, but cannot be the complete answer. It is interesting to consider as an analogy that the process of film coating of tablets, coincidentally introduced at about the same time as direct compression, has practically totally replaced the sugar-coating technique. This shows that a new process can achieve significant and relatively speedy penetration into an industrial environment if it represents a major step forward.

Wet Granulation Process: Advantages and Disadvantages

The wet granulation process is the traditional method of manufacture and is frequently used in the pharmaceutical industry. Expertise in wet granulation is widely available, as is the required equipment. The process improves flow

and cohesion, reduces dust and cross-contamination, and permits the handling of powder blends without loss of homogeneity.

Though it has been practiced for many years and therefore may be perceived as an “old-fashioned” process, it must be borne in mind that the wet granulation process has itself undergone a transformation in recent decades. High-speed mixer–granulators, fluidized bed granulation and drying, and an ever-increasing use of automation have served to make wet granulation a much more efficient and economic process than it once was.^[3]

Nevertheless, the wet granulation process still retains many inherent disadvantages. Problems include choice and method of addition of the binder, and the effect of drying time and temperature on drug stability and its distribution within the solid mass.

Direct Compression Process: Advantages and Disadvantages

The most striking feature of the direct compression process is its simplicity and hence economy. Less equipment is required and the number of stages in the process, each of which will require validation, is greatly reduced. There are also lower labor costs, reduced processing time, and lower power consumption.

An important advantage of the direct compression process is that it is a dry procedure with no need for a drying stage. Thus, exposure to water and the elevated temperatures needed to remove that water are avoided, resulting in a decreased risk of deterioration of the active ingredient.

A further advantage of direct compression is that tablets disintegrate into their primary particles rather than granular aggregates. The resultant increase in surface area available for dissolution should result in faster drug release.

The primary limitation on the use of direct compression is that it depends on the fluidity and compressibility of a tablet diluent. Therefore, it cannot be used for low potency, high dose active ingredients where the inclusion of sufficient diluent in the formulation to permit direct compression would lead to unacceptably large tablets. Thus, active ingredients such as paracetamol and aspirin do not lend themselves to the direct compression process. However, as stated earlier, such ingredients are often available in pregranulated form.

Paradoxically, one of the root causes of difficulties in the direct compression process is its simplicity. It must be regarded as a process in its own right, albeit a simple one, rather than a simplified form of the wet granulation process. The key point to grasp is that wet granulation produces what is in effect a new raw material, i.e., the granule. Minor variations in the properties of the constituents of that granule are covered up by “submerging” them in a mass of binder. This is not so with direct compression. The properties of each particle of every constituent remain essentially unchanged. There is thus a greater need for within-batch and between-batch consistency.

There is also the possibility of segregation of the constituents of the formulation after a homogeneous blend of active ingredient and excipients has been achieved. In a wet granulated product, particles are stuck together by the binder and so there is a much reduced chance of segregation. Because segregation is principally a function of differences in particle size between active ingredients and excipients, it is desirable that the size of the direct compression diluent matches that of the drug. This may not always be feasible.

The simplicity of the direct compression process should obviously bring financial benefits. However, it must be borne in mind that direct compression tablet diluents are considerably more expensive than conventional diluents such as α -lactose monohydrate.

Regulatory considerations also play a part in a decision whether or not to use the direct compression process. Several years may elapse between the finalizing of a tablet formulation and its marketing. During this period, stability testing of the product will have occurred. The formulator must be confident that a chosen direct compression diluent will still be available for a considerable time after product marketing; otherwise reformulation with all its attendant delay and expense will be required. A number of direct

compression diluents have been marketed that were withdrawn after only a few years because of lack of market penetration.

Progression of product development will be accompanied by scale-up of the manufacturing process of the active ingredient to commercial proportions. This may bring about changes in the physical properties of the active ingredient. As stated earlier, the wet granulation process can mask minor changes in physical properties, but this masking cannot occur in direct compression.

For these reasons, direct compression has been most widely adopted by manufacturers of generic (i.e., noninnovative) pharmaceuticals. During the time when the active ingredient is covered by patent protection, its optimum manufacturing process will have been achieved, and so subsequent batch-to-batch variation in its physical properties ought then to be minimal.

FACTORS INFLUENCING THE CHOICE OF A DIRECT COMPRESSION TABLET DILUENT

A wide range of substances is or has been marketed as direct compression tablet diluents. In general, these are commonly occurring materials whose properties have been modified in such a way to give the fluidity and compressibility demanded by the direct compression process. Many direct compression diluents are aggregates of primary particles. For example, an aqueous slurry of α -lactose monohydrate is spray-dried to give an agglomerated product that flows better and is more compressible than the parent substance. A second example is the acid hydrolysis of α -wood cellulose to yield particles containing bundles of microcrystals of cellulose. These can cohere by means of hydrogen bonds to give extremely strong tablets of microcrystalline cellulose.

Most direct compression diluents are available from only one source, but a few can be obtained from more than one manufacturer. If multiple sources are available, they will be offered under individual registered names. For example, microcrystalline cellulose is available under a number of brand names such as Avicel[®] (FMC Corporation), Emcocel[®] (Edward Mendell), and Viva-cel[®] (J. Rettenmaier). Chemical properties of such materials will be similar if not completely identical, especially if there are pharmacopeial standards for identity and purity. However, it cannot be assumed that products from different manufacturers will have the similar physical properties which will govern their performance in the tableting process.

Each brand will probably be accompanied by promotional literature to assist the sale of the product.

It is customary to describe its properties adjectivally, e.g., “excellent flow,” “superior compressibility,” and to present data referring to that material alone. Hence comparison between different excipients can be sometimes extremely difficult.

Several authors have listed the attributes of the “ideal” direct compression diluent.^[2,4] However, it must always be borne in mind that the diluent will invariably form part of a multicomponent mixture. At the very least, the diluent will be mixed with the active ingredient, and almost invariably a lubricant will also be present. The greater the proportion of active ingredient in the formulation, the less influence the diluent will have on the properties of the tablet.

One of the difficulties that beset the product developer is the lack of meaningful tests by which excipients (including direct compression diluents) can be assessed. This has led to the development of the so-called “functionality tests.” Some functionality tests that have been suggested (e.g., particle size, surface area) are in fact physical test methods being carried out under closely defined conditions.^[5] The relation of such a test to the actual function of the excipient needs to be established. After this link has been made, a more suitable title for these tests might be “functionality-related tests”.^[6]

Nevertheless, it is useful to consider some of the desirable properties of a direct compression diluent and how these properties can be appropriately measured.

Properties Required of a Direct Compression Diluent

Fluidity

Good flow is a prerequisite for any tablet formulation to ensure uniformity of tablet weight, which in turn contributes to uniformity of content. Flow can be measured by methods such as angle of repose, flow through an orifice, and by using flow cells, but more meaningful data can be obtained by measuring the uniformity of weight of the tablets themselves. Flow properties can often be improved by the inclusion of a glidant in the formulation.

Ease of mixing and lack of segregation

Achievement of a homogeneous mixture of active ingredient and diluent is essential to obtain tablets with an acceptable uniformity of content of active ingredient. As stated earlier, there is a risk of segregation in a direct compression mixture, because the components are not stuck together as they are in wet granulation. The main

cause of segregation is differences in the particle size of components, with differences in shape and density being secondary factors.^[7] Hence it is desirable that there should not be differences between the particle sizes of active ingredient and diluent. It is unlikely that the size of the active ingredient particles can be changed to match those of the diluent, so the reverse is desirable. Thus, the ideal direct compression diluent should be available in a range of sizes.

An alternative approach is to use the concept of ordered mixing in which fine particles of the active ingredient are dispersed over the surface of much larger diluent particles.^[8] This is only feasible with potent drugs when the diluent will comprise by far the major component of the formulation.

Compression pressure–Tablet strength profile

This is the relationship between the compression pressure applied to the formulation and the physical strength of the resulting tablets. It would seem to be a most important piece of information, yet its derivation is by no means straightforward. Also comparison of data derived by different researchers is often difficult.

If such data are presented graphically, either force or pressure can be used as the abscissa. Unless the cross-sectional dimensions of the tablet are known, interconversion between force and pressure is impossible. Furthermore, a wide variety of units of force and pressure have been used (Table 1), which again makes comparisons difficult if not impossible.

The physical strength of the tablet usually forms the ordinate of a graph. The variety of units used for this is also shown in Table 1. Here too interconversion can be

Table 1 Some units used for the construction of tablet strength profiles

Abcissa		Ordinate	
Parameter	Unit	Parameter	Unit
Force	kg	Crushing strength (hardness)	kg
	lb		Strong-Cobb units
	kN		N
	N		kp
Pressure	kg cm ⁻²	Tensile strength	kg cm ⁻²
	Pa		Pa
	Mpa		MPa
	lb in. ⁻²		

(From Ref. 9.)

extremely difficult. The breaking strength of a tablet is dependent on its physical dimensions. It is therefore logical to use some measure of strength that is independent of tablet size. Therefore the tensile strength of the tablet is often calculated according to Eq. 1.^[10]

$$T_s = 2P/\Pi dt \quad (1)$$

where T_s is the tensile strength of the tablet (MPa), P is the crushing strength (N), d is the tablet diameter (m), and t is the tablet thickness (m).

Use of this equation presupposes that the tablet is circular in cross-section and of uniform thickness, i.e., it is cylindrical. Pitt et al. have attempted to extend the concept of tensile strength to tablets which are not cylindrical.^[11]

Even though comparison is often difficult, it would be rendered considerably easier if the units used conformed to the SI system of weights and measures. Thus, the meter, newton, and pascal should be used as the units of length, force, and pressure, respectively.

Capacity or dilution potential

By definition, direct compression diluents are intended to be mixed with other ingredients. Therefore, not only should the pressure–tablet strength profile of the diluent be determined, but also should those of mixtures of the diluent with an active ingredient. The capacity of a direct compression diluent is the proportion of another ingredient that can be mixed with it while still obtaining tablets of acceptable quality. The definition of “acceptable” will depend on the purpose for which the tablets are required.

The magnitude of the effect that a given active ingredient will have on tablet properties will clearly depend on the tableting properties of that substance. If it is also capable of direct compression, then the effect will not be great. If, however, it is a substance that is difficult to compress into tablets, then it will cause a marked deterioration in tablet quality when mixed with the diluent. Therefore, for a reliable test of capacity, the direct compression diluent should be mixed with a “standard” substance and tableted under standardized conditions. The pressure–strength profiles of the mixtures can then be constructed. Paracetamol^[12,13] and ascorbic acid^[9] have been used as standards.

A method of quantitatively assessing the capacity of a direct compression diluent has been devised by Minchom and Armstrong^[12] who adapted a concept originally introduced by Malkowska and Khan.^[14] Tablets containing only diluent were prepared over a range of compression pressures, their strengths were measured, and a pressure–strength profile was constructed. A mixture of diluent plus

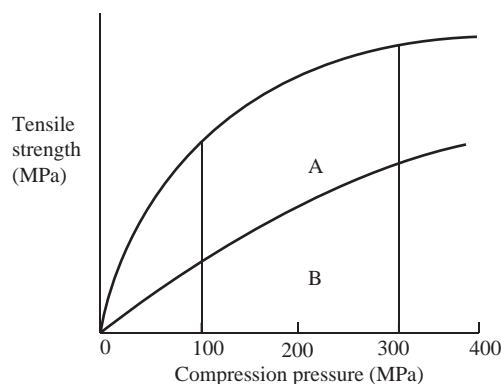


Fig. 2 Calculation of the capacity of a direct compression diluent. (From Ref. 9.)

active ingredient was then compressed under identical conditions, tablet strengths were determined, and the profile was constructed as before (Fig. 2).

The profiles were then fitted by regression analysis to quadratic equations, and integration of these equations between limits (in this case pressures of 100 and 300 MPa) gave the area under the curves. Thus, the change in tablet strength brought about by the addition of the active ingredient is given by the ratio $B/(A + B)$. Details of the application of this method to mixtures of direct compression lactose and ascorbic acid are shown in Table 2.^[9] The technique has been developed further by Habib et al.^[15]

The effect of the active ingredient on flow properties can be determined at the same time by calculating the uniformity of weight of the tablets used to prepare pressure–strength profiles.

Reworking

A faulty batch of tablets can sometimes be recovered by grinding up the tablets and recompressing them, a process

Table 2 Areas under pressure–tablet strength profiles for mixtures of ascorbic acid and direct compression lactitol

Ascorbic acid concentration (%)	Area ratio
0	1.00
20.0	0.75
33.3	0.47
42.9	0.35
50.0	0.32
55.8	0.27
60.0	0.20

(From Ref. 9.)

which is known as reworking and is analogous to the dry granulation method of tablet manufacture. This can sometimes cause problems with a direct compression formulation. Many direct compression diluent particles are in the form of aggregates, e.g., spray-dried lactose is composed of small crystals of lactose embedded in amorphous lactose. If these aggregates are compressed, their structure may be broken down to such an extent that subsequent recompression will result in impaired tablet quality.

The technique of Malkowska and Khan,^[14] used as described before to determine the capacity of a direct compression diluent, was originally developed as a method of expressing the ability of a formulation to be reworked. Referring to Fig. 2, the upper curve represents the strength of tablets prepared without reworking and the lower curve is the strength of reworked tablets. The reworking index is calculated from the ratio of the areas under the curves as described previously.

The mechanism of consolidation

A fundamental property of a solid is the mechanism by which it consolidates under the influence of a compressing force. There are two principal mechanisms—fragmentation and deformation—though most solids will show a mixture of the two with one mechanism predominating. The mechanism can have a major influence on tablet properties.

The effect of compression speed on tablet quality is dependent on the consolidation mechanism. Fragmentation can be regarded as a virtually instantaneous process. Thus, solids which consolidate by fragmentation show little dependence, if any, on the speed at which the consolidation pressure is applied. Deformation on the other hand is time-dependent. It takes a finite time for deformation to occur, and at high rates of punch movement, not enough time may be available for the full effect of the pressure to be exerted.

Table 3 Strain rate sensitivities of powders

Substance	Strain rate sensitivity (%)
Calcium phosphate	0
Calcium carbonate	0
Heavy magnesium carbonate	0
Paracetamol DC	1.8
Paracetamol	10.6
Lactose	16.2
Tabletose	19.2
Anhydrous lactose	20.3
Avicel PH101	38.9
Sodium chloride	39.9
Mannitol	46.4
Maize starch	49.3

(From Ref. 17.)

Changes in punch speed can arise by changing the speed of the tablet press or by changing from a relatively slow-speed excentric press to a high-speed rotary. It must be stressed that the key parameter is punch speed rather than production rate.^[16] Roberts and Rowe^[17] derived a parameter that they called the strain rate sensitivity. This classifies substances according to how tablet strength is affected by changes in punch speed. Details of some of the substances investigated by Roberts and Rowe are shown in Table 3, and many of these are direct compression diluents.

The practical implications of this work are shown in Table 4.^[18] As punch speed is increased, the strength of tablets of dicalcium phosphate is only marginally affected, whereas those made of starch suffer a severe reduction in their strength. As punch speed is increased, speed-sensitive materials give a progressively more porous and hence weaker tablet. It follows that in any report on the tableting properties of a diluent, the type of press and its speed of operation should be stated.

A second area in which consolidation mechanism is important is in the sensitivity of diluents to the effects of

Table 4 Tablet tensile strengths (MPa) produced by a compression pressure of 80 MPa on an excentric press

Press speed (tablets per minute)	Dicalcium phosphate	Lactose	Microcrystalline cellulose	Starch 1500
20	0.63	0.77	5.4	0.28
40	0.62	0.76	5.2	0.25
160	0.60	0.68	3.9	0.15
% reduction	5.0	12	28	47

(From Ref. 18.)

lubricants. In general, addition of a lubricant such as magnesium stearate causes a reduction in tablet breaking strength. As the diluent is mixed with the lubricant, each diluent particle becomes coated with a thin film of lubricant which interferes with interparticulate bonding. However, if fragmentation is the primary method of consolidation, new surface that is uncontaminated by lubricant is continually generated, and so bonding is less affected. Thus, factors which affect the distribution of the lubricant over the diluent surface may have an influence on tablet strength, the magnitude of which depends on the predominant consolidation mechanism. Such factors include mixer design and mixing time and speed.^[19] It follows that the lubricant, its concentration, and method of incorporation should be stated in any publication relating to direct compression diluents.

TABLET DILUENTS USED IN DIRECT COMPRESSION

A wide variety of materials have been used as direct compression diluents, and this has given rise to a considerable literature. Many research reports originate with organizations marketing a specific diluent, or are derived from work sponsored by that organization. Such publications will obviously report data and results relevant to that one diluent.

However, there are some publications which seek to compare a range of diluents, and these are particularly valuable.^[20,21] Perhaps, the most comprehensive review of direct compression diluents has been published by Bolhuis and Chowhan.^[2] Monographs of many direct compression diluents are to be found in the *Handbook of Pharmaceutical Excipients*, which contains details of the physical properties of these substances.^[22]

Direct compression diluents are often commonly occurring substances which have been physically modified to give the necessary degree of fluidity and compressibility. They are most conveniently classified according to their source, viz.:

- Cellulose and cellulose derivatives.
- Inorganic materials.
- Polyols.
- Starch and starch derivatives.
- Sugars.
- Mixtures and coprocessed products.

Information on some direct compression diluents is given in Table 5.

Cellulose and Cellulose Derivatives

Since its introduction as a direct compression diluent in 1964, microcrystalline cellulose has become a pharmaceutical excipient of great importance. It consists of aggregates of microcrystals and is isolated from α -wood cellulose by acid hydrolysis.

When compressed, microcrystalline cellulose particles deform plastically and the surfaces thus brought into contact unite by hydrogen bonding. Because tablets made from microcrystalline cellulose are extremely strong, there is a high dilution potential and they can withstand weakening caused by lubricants.^[19] In fact, microcrystalline cellulose tablets exhibit such a low coefficient of friction that they may need no lubricant. Fluidity of microcrystalline cellulose is low and the addition of a glidant may be necessary. Its bulk density is also low. Unlike many direct compression diluents, it is available in a wide range of particle sizes.

Microcrystalline cellulose is available from several different sources. These can exhibit a range of tableting properties, and so the substitution of one brand of microcrystalline cellulose by another must be approached with caution.^[23] Microcrystalline cellulose is quite hygroscopic, and its tableting properties are dependent on its moisture content, with water causing the interparticulate hydrogen bonds to weaken. Therefore, comparisons between different brands must also take the moisture content into account.^[24]

Powdered cellulose has been used as a direct compression diluent. Though it forms hard tablets, fluidity is poor and dilution potential is low. Like microcrystalline cellulose it has some self-lubricating properties, but addition of a lubricant is usually necessary, causing a marked reduction in tablet strength.^[25]

Silicified microcrystalline cellulose is a coprocessed mixture of microcrystalline cellulose and 2% colloidal silicon dioxide, which has improved flow and binding properties compared to microcrystalline cellulose itself.^[26]

Inorganic Materials

The most widely used inorganic direct compression diluent is calcium phosphate. It is available in several forms, but the unmilled (i.e., coarse) dibasic dihydrate ($\text{CaHPO}_4 \cdot 2\text{H}_2\text{O}$) is the most frequently used. It has good flow and binding properties. Consolidation is principally by fragmentation, so although a lubricant is needed, tablet strength loss is low.^[19] It is nonhygroscopic at humidities up to 80%, and due to its hydrophilic nature, dicalcium phosphate dihydrate tablets are rapidly penetrated by water on immersion. Despite this, the tablets

Table 5 Direct compression diluents, their proprietary names, and their manufacturers

Diluent	Proprietary name	Manufacturer
<i>Cellulose and cellulose derivatives</i>		
Cellulose, microcrystalline	Avicel [®]	FMC
	Emcocel [®]	Penwest
	Vivacel [®]	Rettenmaier
Cellulose, powdered	Solka-Floc [®]	Degussa
	Elcema [®]	Penwest
Cellulose, silicified microcrystalline	Prosolv [®]	Penwest
<i>Inorganic materials</i>		
Calcium phosphate, dibasic anhydrous	Dicafos-A [®]	Budenheim
Calcium phosphate, dibasic dihydrate	Calstar [®]	FMC
	Dicafos [®]	Budenheim
	Emcompress [®]	Penwest
	Di-Tab [®]	Stauffer
	Tricafos [®]	Budenheim
Calcium phosphate, tribasic	Tricompres [®]	Penwest
	Tri-Tab [®]	Rhone-Poulenc
Calcium sulphate	Compactrol [®]	Penwest
<i>Polyols</i>		
Xylitol	Xylitab [®]	Xyrofin
Lactitol	Finlac [®]	Xyrofin
Mannitol	Pearlitol [®]	Roquette
Sorbitol	Neosorb [®]	Roquette
	Sorbitol Instant [®]	Merck
<i>Starch and starch derivatives</i>		
Starch, pregelatinized	Starch 1500 [®]	
	Starx 1500 [®]	Colorcon
Modified rice starch	Primotab [®]	Avebe
Granulated corn starch	Sepistab ST200	Seppic
<i>Sugars</i>		
Compressible sugar	Destab	Desmo
	Dipac	Amstar
	Nutab	Ingredient Technology
Dextrates	Sugartab	Penwest
	Emdex [®]	Penwest
	Celutab [®]	Penick and Ford
<i>Lactose</i>		
Agglomerated	Tablettose	Meggle
	Pharmatose DCL15 [®]	DMV
Anhydrous α	Pharmatose DCL30 [®]	DMV
Anhydrous β	Pharmatose DCL21 [®]	DMV
	Anhydrous DT	Sheffield
Spray-dried	Fast-Flo [®]	Foremost
	Zeparox [®]	Borculo
	Pharmatose DCL11 [®]	DMV
<i>Mixtures and coprocessed products</i>		
Anhydrous lactose–anhydrous lactitol	Pharmatose DCL40 [®]	DMV
Calcium sulphate–microcrystalline cellulose	Cel-O-Cal	FMC
Lactose–cellulose	Cellactose [®]	Meggle
Lactose–starch	Starlac [®]	Roquette
Lactose–povidone	Ludipress [®]	BASF

do not disintegrate because they are almost insoluble in water.^[27] Dicalcium phosphate dihydrate is, however, soluble in acidic media such as gastric juice.

The water of hydration is relatively easily lost from dibasic calcium phosphate dihydrate, and this may have consequences for the stability of products containing it.^[28]

Anhydrous dibasic calcium phosphate and calcium triphosphate can also be used for direct compression. The latter is actually a mixture of calcium phosphates including tricalcium orthophosphate [$\text{Ca}_3(\text{PO}_4)_2$] and hydroxyapatite [$\text{Ca}_5(\text{OH})(\text{PO}_4)_3$]. The preparation and properties of calcium phosphates have been reviewed by Carstensen and Ertell,^[29] and their tableting properties have been studied by Bryan and McAllister.^[30]

A direct compression diluent based on calcium sulphate is also available.^[31]

Polyols

Several solid polyols can be used as direct compression diluents, usually after some physical modification. Most such polyols can be obtained from natural sources, but are usually manufactured by hydrogenation of the parent sugar molecule. Some properties of polyols, together with comparative data for lactose and sucrose, are shown in Table 6.

Sorbitol

This is probably the most commonly used polyol in tablet manufacture. It is freely soluble in water, producing strong tablets. Its negative heat of solution, which produces a cooling effect in the mouth, makes it particularly useful in chewable tablets and “sugar-free” confectionary. Furthermore, most polyols are noncariogenic, which increases their usefulness in this area. It is hygroscopic, and shows an appreciable increase in moisture content with attendant impairment of flow properties when exposed to relative

humidities in excess of 65%. At very high humidities, sorbitol absorbs enough water to bring about dissolution.

Sorbitol can exist in four crystalline forms. Guyot-Hermann et al.^[32] compared 11 commercially available varieties of sorbitol, and found three of these four forms to be present. γ -Sorbitol was found to be the most useful as a tablet diluent. The method of manufacture has also been shown to affect tableting properties, differences being attributed to variations in particle shape and surface properties. Spray-dried varieties of sorbitol are available as direct compression diluents which are claimed to have overcome problems associated with the different crystalline forms.^[33]

Mannitol

Mannitol is an isomer of sorbitol. Like the latter, it has a negative heat of solution which makes it a useful excipient for chewable tablets and lozenges. It is less hygroscopic than sorbitol and has about one-tenth of the solubility in water. Similarly to sorbitol, several polymorphic forms are available which differ in their ability to form tablets.^[34] However, unmodified mannitol cannot be used for direct compression because of poor flow and binding properties. Directly compressible forms are available in a range of particle sizes which are reported to produce excellent tablets.

Lactitol and xylitol

These are both commercially available in forms suitable for direct compression with good flow and binding properties. The former is a water-granulated product of microcrystalline aggregates.^[35] A similar form of xylitol is available, and in the case of xylitol, there are also products pregranulated with either polydextrose or carboxymethyl cellulose.^[36] Both substances are highly soluble in water with negative heats of solution.

Table 6 Properties of some polyols

	Solubility in water (g per 100 cm ³)	Melting point (°C)	Heat of solution (J g ⁻¹)	Water content at			Sweetness (sucrose = 100)
				55% RH	75% RH	100% RH	
Lactitol	170	180	-54	0.0	0.1	37.2	35
Mannitol	18	167	-121	0.5	0.5	83.7	50
Sorbitol	200	110	-113	6.4	10.4	122.2	60
Xylitol	160	94	-153	0.1	0.1	70.6	100
Lactose	22	201	-18	0.8	0.8	7.0	15
Sucrose	200	160	-17	1.0	1.0	38.0	100

Starch and Starch Derivatives

Starch is a very widely used tablet excipient, but in its natural state, it does not possess the fluidity and binding characteristics needed as a tablet diluent. The major consolidation mechanism of starch is by deformation with a high elastic component.^[37] In addition, starch shows a high degree of lubricant sensitivity.^[19]

Pregelatinized starch, often referred to as Starch 1500, contains about 80% unmodified starch, 5% free amylose, and 15% free amylopectin. Though it has been described as a direct compression diluent,^[38] tablets made from pregelatinized starch show low physical strength. The principal application of pregelatinized starch in tablet formulation is as a disintegrating agent. It retains the disintegrating ability of natural starch without the deleterious effects on flow and tablet strength that natural starch would bring about.

Sugars

Lactose

Lactose is a naturally occurring disaccharide containing one galactose unit and one dextrose unit. It is a constituent of all forms of mammalian milk, but is produced commercially from cow's milk, usually as a by-product of the cheese industry. Lactose can exist in two isomeric forms, α -lactose and β -lactose, and can be either crystalline or amorphous. Crystalline α -lactose occurs in both monohydrate and anhydrous forms, but β -lactose only exists in the anhydrous form. The temperature of crystallization is the principal determinant of which form is obtained.^[39]

Though crystalline α -lactose monohydrate is the most common tablet diluent, it is usually used in granulated rather than in direct compression formulations. Neither its flow properties nor its binding properties are good enough to form satisfactory tablets without preliminary treatment. Bonding properties are improved by conversion into aggregates of α -lactose monohydrate crystals by fluid bed granulation. This product is virtually free from amorphous lactose.^[40]

Spray-dried lactose was the first direct compression diluent to be introduced.^[41] It had a major impact on tableting technology and it is still widely used. Spray drying an aqueous suspension of lactose yields a product that largely consists of crystals of α -lactose monohydrate (about 80%) held together in a glass of amorphous material (about 20%). Spray-dried lactose exhibits excellent flow properties due to the spherical shape of the aggregates. However, its ability to form strong tablets is

limited and it has low dilution potential, so it is primarily used in tablets in which it forms the major ingredient. Fragmentation is the major consolidation mechanism, and so tablet strength is not significantly affected by lubricants. Spray-dried lactose can be obtained from several manufacturers whose products differ slightly. A comparative study of spray-dried lactoses from a number of sources has been published by Whiteman and Yarwood.^[21]

Anhydrous lactose is primarily anhydrous β -lactose with up to about 25% anhydrous α -lactose. It consists of agglomerates of fine crystals produced by roller drying a solution of α -lactose monohydrate. Flow properties are good, and tablet strength was found to be superior to other lactose products.^[21] Anhydrous β -lactose is much more soluble than the α -isomer, and extended disintegration times of tablets made from anhydrous lactose have been attributed to the presence of anhydrous α -lactose in the roller dried product.^[42]

Anhydrous α -lactose can be produced by thermal or chemical dehydration of α -lactose monohydrate. During this process, the starting material changes from single crystals to aggregates of anhydrous α -lactose particles. Flowability and binding properties are good, but slow dissolution of tablets made from anhydrous α -lactose has proved a major limitation to its use.^[42]

Sucrose

A nonreducing disaccharide obtained from vegetable sources, sucrose is a widely used pharmaceutical excipient. Because it is more hygroscopic than lactose, it is used less in solid dosage forms. The compactability of pure sucrose is poor, but modified forms of sucrose for direct compression are available. These are collectively termed "compressible sugar," and may contain, depending on source, starch, maltodextrin, or invert sugar, together with a lubricant. Several different types of compressible sugar have been compared.^[43] The minor components obviously played a major role in tablet formation, since significant differences were obtained between varieties. Compressible sugar is often used for lozenges and chewable tablets—because of the high solubility of sucrose, tablets tend to dissolve rather than disintegrate.

Dextrose and dextrates

Dextrose does not lend itself to direct compression. However, a spray-crystallized product, the major constituent of which is dextrose, is used in direct compression. This is known as dextrates, and is produced by the partial hydrolysis of starch. It consists of about 90% dextrose, together with about 5% maltose and higher polysaccharides.^[44]

Though both hydrated and anhydrous forms of dextrates have been described, only the former is commercially available. Dextrose is freely soluble in water and is highly hygroscopic, and hence its use in atmospheres of high humidity may cause problems. Because of its sweet taste and negative heat of solution, it is recommended for use in chewable tablets.

Mixtures and Coprocessed Products

Though numerous direct compression diluents are available, none is ideal. For example, spray-dried lactose flows easily but forms relatively weak tablets, whereas the fluidity of microcrystalline cellulose is poor yet it forms extremely strong tablets. It is understandable, therefore, that the possibilities of combining diluents have been considered, the aim being to combine the advantages of both components without their disadvantages.

The compaction properties of mixtures have been reviewed by Fell,^[45] who concluded that the relationship between the tableting properties of a mixture could only rarely be predicted from knowledge of the same properties of the individual components. Nevertheless, some success has been achieved.^[46]

In recent years, a number of coprocessed excipient combinations have been marketed, which undoubtedly possess advantages over physical mixtures of the same components. Such materials include mixtures of lactose and povidone,^[47] cellulose and lactose,^[48] and anhydrous lactose and anhydrous lactitol. A disadvantage of this approach is that the relative proportions of the components are fixed, and such a combination may not be universally optimal.

REFERENCES

1. Little, A.; Mitchell, K.A. *Tablet Making*; 2nd Ed.; Northern Publishing: Liverpool, U.K., 1963.
2. Bolhuis, G.K.; Chowhan, Z.T. Materials for direct compression. In *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nystrom, C., Eds.; Marcel Dekker, Inc.: New York, 1996; 419–500.
3. Kristensen, H.G.; Schaefer, H.G.; Granulation, T.; Drug Dev. Ind. Pharm. **1987**, *13* (4/5), 803–872.
4. Sheth, B.B.; Bandelin, F.J.; Shangraw, R.F. Compressed Tablets. In *Pharmaceutical Dosage Forms: Tablets*; 1st Ed.; Lieberman, H.A., Lachman, L., Eds.; Marcel Dekker, Inc.: New York, 1980; Vol. 1, 109–185.
5. Brittain, H.G. Functionality testing of excipient materials. Pharm. Technol. **1993**, *9* (9), 72–73.
6. Armstrong, N.A. Functionality related tests for excipients. Int. J. Pharm. **1997**, *155*, 1–5.
7. Campbell, H.; Bauer, W.C. Cause and cure of demixing in solid–solid mixers. Chem. Eng. **1966**, September, 179–185.
8. Staniforth, J.N.; Rees, J.E.; Kayes, J.B. Relation between mixing time and segregation of ordered mixes. J. Pharm. Pharmacol. **1981**, *33*, 175–176.
9. Armstrong, N.A. Selection of excipients for direct compression tablet formulations. Pharm. Technol. Eur. **1997**, September, 24–30.
10. Fell, J.T.; Newton, J.M. Determination of tablet strength by the diametral compression test. J. Pharm. Sci. **1970**, *59* (5), 688–691.
11. Pitt, K.G.; Newton, J.M.; Richardson, R.; Stanley, P. Material tensile strength of convex-faced aspirin tablets. J. Pharm. Pharmacol. **1989**, *41*, 289–292.
12. Minchom, C.M.; Armstrong, N.A. A proposed technique for expressing the capacity of directly compressible tablet diluents. J. Pharm. Pharmacol. **1989**, *39*, 69.
13. Wells, J.I.; Langridge, J.R. Dicalcium phosphate dihydrate–microcrystalline cellulose systems in direct compression tableting. Int. J. Pharm. Tech. and Prod. Manuf. **1981**, *2* (2), 1–8.
14. Malkowska, S.; Khan, K.A. The effect of recompression on the properties of tablets prepared by dry granulation. Drug Dev. Ind. Pharm. **1983**, *9* (3), 331–347.
15. Habib, Y.; Augsburg, L.; Reier, G.; Wheatley, T.; Shangraw, R. Dilution potential: A new perspective. Pharm. Dev. Technol. **1996**, *1* (2), 205–212.
16. Armstrong, N.A. Time-dependent factors involved in powder compression and tablet manufacture. Int. J. Pharm. **1989**, *49*, 1–13.
17. Roberts, R.J.; Rowe, R.C. The effect of punch velocity on the compaction of a variety of materials. J. Pharm. Pharmacol. **1985**, *37*, 526–528.
18. Armstrong, N.A.; Palfrey, L.P. The effect of machine speed on the consolidation of four directly compressible tablet diluents. J. Pharm. Pharmacol. **1989**, *41*, 149–151.
19. Bolhuis, G.K.; Holzer, A.W. Lubricant Sensitivity. In *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nystrom, C., Eds.; Marcel Dekker, Inc.: New York, 1996; 517–560.
20. Bolhuis, G.K.; Reichman, G.; Lerk, C.F.; van Kamp, H.V.; Zuurman, K. Evaluation of anhydrous α -lactose, a new excipient in direct compression. Drug Dev. Ind. Pharm. **1985**, *11* (8), 1657–1681.
21. Whiteman, M.; Yarwood, R.J. The evaluation of six lactose-based materials as direct compression tablet excipients. Drug Dev. Ind. Pharm. **1988**, *14* (8), 1023–1040.
22. Kibbe, A., Ed. *Handbook of Pharmaceutical Excipients*, 3rd Ed.; Pharmaceutical Press: London, 2000.
23. Doelker, E.; Mordier, D.; Iten, H.; Humbert-Droz, P. Comparative tableting properties of sixteen microcrystalline celluloses. Drug Dev. Ind. Pharm. **1987**, *13* (9–11), 1847–1875.
24. Roberts, R.J.; Rowe, R.C. Source and batchwise variability in the compressibility of microcrystalline cellulose. J. Pharm. Pharmacol. **1987**, *39*, 70P.

25. Bos, C.E.; Vromans, H.; Lerk, C.F. Lubricant sensitivity in relation to bulk density for granulations based on starch and cellulose. *Int. J. Pharm.* **1991**, *67*, 39–49.
26. Sherwood, B.E.; Hunter, E.A.; Staniforth, J.N. Silicified microcrystalline cellulose (SMCC™). A new class of high functionality binders for direct compression diluents. *Pharm. Res.* **1996**, *13* (9), S197.
27. Caramella, C.; Colombo, P.; Conte, U.; Ferrari, F.; La Manna, A.; van Kamp, H.V.; Bolhuis, G.K. Water uptake and disintegrating force measurements: Towards a general understanding of disintegration mechanisms. *Drug Dev. Ind. Pharm.* **1986**, *12* (11–13), 1749–1766.
28. De Haan, P.; Kroon, C.; Sam, A.P. Decomposition and stabilisation of the tablet excipient calcium hydrogen phosphate dihydrate. *Drug Dev. Ind. Pharm.* **1990**, *16* (13), 2031–2055.
29. Carstensen, J.T.; Ertell, C. Physical and chemical properties of calcium phosphate for solid state pharmaceutical formulations. *Drug Dev. Ind. Pharm.* **1990**, *16* (6), 1121–1133.
30. Bryan, J.W.; McCallister, J.D. Matrix forming capabilities of three calcium diluents. *Drug Dev. Ind. Pharm.* **1992**, *18* (19), 2029–2047.
31. Elsabbagh, H.M.; Elshaboury, M.H.; Abdel-Aleem, H.M. Physical properties and stability of diazepam and phenobarbitone sodium tablets prepared with compactrol. *Drug Dev. Ind. Pharm.* **1985**, *11* (11), 1947–1955.
32. Guyot-Hermann, A.M.; Leblanc, D.; Dragnet-Brugmans, M. Gamma sorbitol as a diluent in tablets. *Drug Dev. Ind. Pharm.* **1985**, *11* (2/3), 551–564.
33. Basedow, A.M.; Moschl, G.A.; Schmidt, P.C. Sorbitol instant, an excipient with unique tableting properties. *Drug Dev. Ind. Pharm.* **1986**, *12* (11–13), 2061–2089.
34. Debord, B.; Lefebvre, C.; Guyot-Hermann, A.M.; Hubert, J.; Bouche, R.; Guyot, J.C. Study of different crystalline forms of mannitol: Comparative behaviour under compression. *Drug Dev. Ind. Pharm.* **1987**, *13* (9–11), 1533–1546.
35. Armstrong, N.A. Direct compression characteristics of lactitol. *Pharm. Tech. Eur.* **1998**, *10* (2), 42–46.
36. Morris, L.E.; Moore, J.C.; Schwartz, J.B. Characterisation and performance of a new direct compression excipient for chewable tablets: Xylitab. *Drug Dev. Ind. Pharm.* **1996**, *22* (9/10), 925–932.
37. Paronen, P.; Juslin, M. Compression characteristics of four starches. *J. Pharm. Pharmacol.* **1983**, *35*, 627–635.
38. Manudhane, K.S.; Contractor, A.M.; Kim, H.Y.; Shangraw, R.F. Tableting properties of a directly compressible starch. *J. Pharm. Sci.* **1969**, *58* (5), 616–620.
39. Lerk, C.F.; Andreae, A.C.; de Boer, A.H.; de Hoog, P.; Kussendrager, K.D.; van Leverink, J. Transition of lactoses by mechanical and thermal treatment. *J. Pharm. Sci.* **1984**, *73*, 857–859.
40. Shangraw, R.F.; Wallace, J.W.; Bowers, F.M. Morphology and functionality of tablet excipients for direct compression. *Pharm. Tech.* **1981**, *5*, 69–78.
41. Gungel, W.C.; Lachman, L. Comparative evaluation of tablet formulations prepared from conventionally processed and spray-dried lactose. *J. Pharm. Sci.* **1963**, *52* (2), 178–182.
42. Van Kamp, H.V.; Bolhuis, G.K.; Kussendrager, K.D.; Lerk, C.F. Studies on Tableting Properties of Lactose. IV. Dissolution and disintegration properties of different types of crystalline lacose. *Int. J. Pharm.* **1986**, *28*, 229–238.
43. Ondari, C.O.; Kean, C.E.; Rhodes, C.T. Comparative evaluation of several direct compression sugars. *Drug Dev. Ind. Pharm.* **1983**, *9* (8), 1555–1572.
44. Henderson, N.L.; Bruno, A.J. Lactose USP (Beadlets) and Dextrose (PAF 2011), Two new agents for direct compression. *J. Pharm. Sci.* **1970**, *9*, 1336–1340.
45. Fell, J.T. Compaction properties of binary mixtures. In *Pharmaceutical Powder Compaction Technology*; Alderborn, G., Nystrom, C., Eds.; Marcel Dekker, Inc.: New York, 1996; 501–515.
46. Armstrong, N.A.; Lowndes, D.H.L. The use of mixtures of spray-dried lactose and microcrystalline cellulose as direct compression diluents. *Int. J. Pharm. Tech. and Prod. Manuf.* **1984**, *5* (3), 11–14.
47. Schmidt, P.C.; Rubensdorfer, C.J. Evaluation of ludipress as a multipurpose excipient. *Drug Dev. Ind. Pharm.* **1994**, *20* (18), 2927–2952.
48. Armstrong, N.A.; Roscheisen, G.; Al-Aghbar, M.R. Cellac-tose as a tablet diluent. *Manuf. Chem.* **1998**, *67*, 25–26.



Tablet Press Instrumentation

Michael Levin

Metropolitan Computing Corporation, East Hanover, New Jersey, U.S.A.

INTRODUCTION

This article is designed to facilitate the understanding of the general principles of tablet press instrumentation and the benefits thereof by the formulators, process engineers, validation specialists, and quality assurance personnel, as well as production floor supervisors who would like to understand the basic standards and techniques of getting information about their tableting process.

HISTORY OF TABLET PRESS INSTRUMENTATION

In 1952–1954 Higuchi and his group^[1] have instrumented upper and lower compression, ejection, and punch displacement on an eccentric tablet press, and pioneered the modern study of compaction process. This work was followed by Nelson^[2,3] who was the member of the original group.

In 1966, a U.S. patent was granted to Knoechel and co-workers^[4] for force measurement on a tablet press. This patent was followed by two seminal articles in *Journal of Pharmaceutical Sciences* on the practical applications of instrumented rotary tablet machines.^[5,6] A number of other patents related to press instrumentation and control followed from 1973 on ward.^[7–15]

Despite the availability of published designs for instrumenting rotary presses, much of the early work on compaction properties of materials was done on instrumented single punch eccentric presses primarily, due to the relative ease of sensor installation, as well as availability of punch displacement measurement.^[16–18]

In mid-1980s, custom-made press monitoring systems were described,^[19,20] and the first commercial instrumentation packages became available, including both the systems for product development and press control. The first personal computer-based tablet press monitor was sold in 1987, and the first Microsoft Windows-based press instrumentation package in 1995.

A new era of compaction research has begun with the introduction of an instrumented compaction simulator^[21]

in 1976, while a compaction simulator patent was issued as recently as 1996.^[22]

A new generation of “compaction simulators” was born when a mechanical press replicator was patented in the year 2000.^[23]

A good exposé of the early stages of press instrumentation and resulting research is presented in review articles by Schwartz,^[24] Marshall,^[25] and Jones.^[26] A comprehensive review can be found in a relatively recent paper by Çelik and Ruegger.^[27]

For other historical information on the press instrumentation, the reader is encouraged to peruse Ridgeway Watt's^[28] volume. There have been a number of papers published on various disparate instrumentation topics, and in a recent volume by Muños-Ruiz and Vromans,^[29] there are two good articles on the subject—but, unfortunately, they deal with marginal issues of single station press and instrumented punch only.

DATA ACQUISITION PRINCIPLES: FROM TRANSDUCERS TO COMPUTERS

To monitor and control a tablet press, certain sensors must be installed at specific locations on the machine. These sensors are called transducers. In general, a transducer is a device that converts energy from one form to another (e.g., force to voltage). Tablet press transducers typically measure applied force, turret speed, or punch position. Because the signals coming from such transducers are normally in millivolts, they need to be amplified and then converted to digital form in order to be processed by a data acquisition system.

Piezoelectric Gages

Historically, high impedance piezoelectric transducers that employ quartz crystals were used in early stages of press instrumentation. When subject to stress, the crystal accumulates electrostatic charge that is directly proportional to the applied force. Both low and (more modern) high impedance piezoelectric gages have high-frequency response, but may exhibit signal drifting due to charge

leakage (approximately 0.04% decay per second can be seen in modern piezoelectric transducers). Nowadays, the preferred way of instrumenting tablet presses is with strain gages.

Strain Gages

Strain gages are foil, wire, or semiconductor devices that convert pressure or force into electrical voltage. When a stress is applied to a thin wire, it becomes longer and thinner. Both factors contribute to increased electrical resistance. If an electrical current is sent through this wire, it will be affected by the changes in the resistance of the conduit. This principle is used in strain gage-based transducers. Foil gages, known for robust application range, useful nominal resistance, and reliable sensitivity control, are most commonly used for instrumentation of compression, precompression, and ejection forces. Semiconductor-based strain gages are inherently more sensitive but suffer from high electrical noise and temperature sensitivity. Such gages are not commonly used in tablet press instrumentation except for measurement of die-wall pressure and take-off forces.

Wheatstone Bridge

Wheatstone bridge (Fig. 1) is a special arrangement of strain gages that is used to ensure signal balancing. The “full” bridge is composed of two pairs of resistors in a circle, with two parallel branches used for input and two for signal output. By applying the so-called excitation voltage (typically, 10 V DC) to the bridge input and changing the resistance of different “legs” of the bridge by adding special resistors, we can make sure that there is a zero output voltage when no load is applied to the transducer. This is called zero balance. Once the bridge is

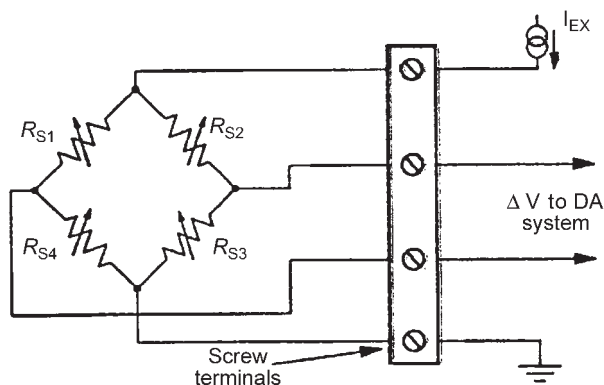


Fig. 1 A typical Wheatstone bridge arrangement of strain gages.

balanced, a small perturbation in the resistance of any “leg” of the bridge results in a nonzero signal output.

The output of the Wheatstone bridge is normally expressed in millivolts per volt of excitation per unit of applied force. For example, sensitivity of 0.2 mV/V/kN means that applying, say, 10 kN force and 10 V excitation will produce 20 mV output. To utilize such output, it usually needs to be amplified several hundred times to reach units of volts.

Another important function of the bridge is balancing of temperature effects. Although individual strain gages are sensitive to temperature fluctuations, Wheatstone bridge arrangement provides for temperature sensitivity compensation, so that the resulting transducer is no longer changing its output to any significant degree when it heats up.

Because the output of these bridges is in the range of millivolts, the cables utilized to carry the signal are normally shielded with a braided or foil-lined sheath around individual wires. The shield, as a rule, is connected to the amplifier, but never touches the actual instrumented equipment (i.e., tablet press). If this rule is violated, a ground loop may generate electrical noise and present a dangerous electrical shock hazard.

Gage Factor

The gage factor is a specific characteristic of a strain gage. It is calculated as the change in resistance relative to unit strain that has caused this change. Strain gages that are commonly used for tablet press instrumentation are made from copper–nickel alloy and have a gage factor of about 2. In a typical bending beam application (such as compression roll transducer), one side of the beam experiences tension while the other side undergoes compression. By mounting two gages on each side, the sensitivity of the transducer can be doubled.

Strain gages have to be bonded to areas of machine parts that are most sensitive to applied force. Such areas can be identified with the use of polarized light technique that “points out” the stress distribution.

Manufacturing a Force Transducer

Usually, instrumentation designs are proprietary and specific detail drawings are held in fiducial capacity. Each force transducer is custom designed for a particular machine part. Overall specifications are taken from the actual party and/or manufacturer approved drawings.

Duplicate steel members, such as pins and cams, are normally made from A2 tool steel in a fully annealed

Tablet Press Instrumentation

(softened) state. Original machine parts are first annealed, if required. The member is then machined, hardened to Rockwell 60–64, tested, and finally, ground to specified dimensions.

It is highly advantageous to determine stress distribution using polarized light beams identifying the areas of maximum strain yet avoiding areas of uneven strain.

A typical procedure of transducer manufacturing in a professional gaging lab might be as follows. Foil type strain gages are bonded to the steel member utilizing a high-performance 100% solid epoxy system adhesive under controlled heating rate conditions. After intrabridge wiring is completed using solid conductor wiring (to prevent electrical noise), multistrand wire cabling with a combination of foil and braided insulation is connected to the bridge. Wire anchoring is achieved utilizing epoxy adhesive and protected with a combination of latex-based adhesives and/or epoxy resins. Lead wire cabling is protected by Teflon outer shield, as well as inner braided wire shield. The Teflon to steel joint is sealed with epoxy but should not be subject to stresses that would cause the cable to kink or yield in such a way as to expose the inner braided wire shield. Protective coatings are then applied and final postcure heating is performed for at least 2 hr at a temperature approximately 50°C above the transducer normal operating temperature (approximately 175 or 105 °C). A silicon-based adhesive, such as RTV, is used to fill large cavities while maintaining a low modulus of elasticity, preventing undue influence upon the actual strain measurement. The coating is resilient to moderate mechanical abrasion, as well as most solvents, oils, cleaners, etc. It is not intended for protection against penetrating sharp objects.

A high-quality connector is then attached to the cable. The next step is to perform offset zeroing with fixed, 1% precision, low-temperature coefficient resistors, followed by NIST (National Institute for Standards and Technology) traceable calibration.

It is worth noting that, in general, the duplicate members are not made of corrosion-resistant steel, because high tensile strength and ability to be easily machined are required. Prior to storage, the surface should be treated like any high-grade tablet press punch steel will be treated; a thin coating of oil should be used after wiping with alcohol.

Load Cells

In some cases, where appropriate, a load cell can be used in place of bonded strain gages.

A load cell is a strain gage bridge in an enclosure forming a complete transducer device.

Like any properly made transducer, it produces an output signal proportional to the applied load. Unlike custom-made transducers that require application of strain gages directly to press parts, load cells are self-contained and can be placed on a press in specially machined cavities to be easily replaced or serviced. As a drawback, load cells generally are less sensitive or less suited to measure the absolute force than custom-made strain gage transducers. Load cells can be used on punch holders in a single station press. Another example of load cell use is a die assembly for calibration of existing traducers on a press.

Linear Variable Differential Transformer

Linear variable differential transformer (LVDT, Fig. 2) is a device that produces voltage proportional to the position of a core rod inside a cylinder body. It measures displacement or a position of an object relative to some predefined zero location. On tablet presses, LVDTs are used to measure punch displacement and in-die thickness. They generally have very high precision and accuracy, but there are numerous practical concerns regarding improper mounting or maintenance of such transducers on tablet presses.

Proximity Switch

Proximity switch (Fig. 3) is a noncontact electromagnetic pickup device that senses the presence of metal. On tablet presses, it is widely used to detect the beginning of a turret



Fig. 2 Linear variable differential transformer (LVDT).

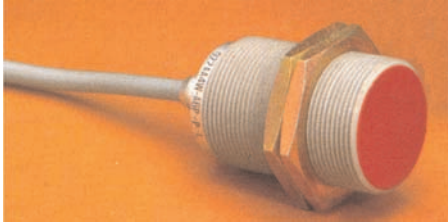


Fig. 3 Proximity switch.

revolution, to identify stations and to facilitate peak detection in tablet force control applications.

Instrumentation Amplifier and Signal Conditioning

Signal conditioning involves primarily an amplifier that provides the excitation voltage, as well as gain (a factor used in converting millivolt output of the strain gage bridge to the volt-based range of the input of the data acquisition device).

Instrumentation amplifier is different from other types of amplifiers in that the signal from each side of the transducer bridge is amplified separately and then the difference between the two amplified signals is, in turn, amplified. As a result, noise from both sides of the sensors is reduced.

Some instrumentation amplifiers also offer filter functionality. A filter circuit combines resistors and capacitors that act to block the undesirable frequencies. It is a fact, however, that a good transducer should provide a clean signal. Use of analog or digital filters may cause the loss of some important portions of the signal that one is attempting to measure. Frequency filters may cause the measured compression force peak, for example, to be skewed toward the trailing edge of the peak and can yield a lower than actual peak force.

Because filters distort the signal, they must be avoided unless absolutely necessary. In many cases, better electronics may make filter use obsolete. For example, the so-called antialiasing filter that is used to condition a high-frequency signal for a slow sampling rate is generally not required for tablet press applications if modern fast speed data acquisition devices are used for signal processing.

In addition to amplification, excitation, and filtering, signal-conditioning devices may provide isolation, voltage division, surge protection, and current-to-voltage conversion.

Instrumentation Terms: Definitions

Several important terms may be now defined with respect to transducers and signal conditioning:

- **Full Scale (FS):** The total interval over which a transducer is intended to operate. Also, it can define the output from transducer when the maximum load is applied.
- **Excitation:** The voltage applied to the input terminals of a strain gage bridge.
- **Accuracy:** The closeness with which a measurement approaches the true value of a variable being measured. It defines the error of reading. Good tablet press transducers have at least 1% accuracy (with this level of accuracy, for example, a compression force transducer with 50 kN FS will produce at most an error of 500 N).
- **Precision:** Reproducibility of a measurement, i.e., how much successive readings of the same fixed value of a variable differ from one another. If a person is shooting darts, for example, the accuracy is determined by how close to the bull's eye the darts have landed, while the precision will be indicated by how close the darts are to each other.
- **Resolution:** The smallest change in measured value that the instrument can detect.
- **Dynamic range of a transducer:** The difference between the maximal FS level and the lowest detectable signal. Measured in decibels (dB), it indicates the ratio of signal maximum to minimum levels:

$$\text{dB} = 10 \log_{10}(S_{\max}/S_{\min})$$

Some press sensors may have a rather narrow dynamic range not necessarily correlated with accuracy. For example, a very accurate compression roll pin designed to measure 50 kN force may not detect 5 kN signal.

- **Calibration:** Comparison of transducer outputs at standard test loads to output of a known standard at the same load levels. A line representing the best fit to data is called a calibration graph.
- **Calibration factor:** A load value in engineering units that a transducer will indicate for each volt of output, after amplifier gain and balance. Calibration factor is usually expressed in relation to FS.
- **Shunt calibration:** A procedure of transducer testing when a resistor with a known value is connected to one leg of the bridge. The output should correspond to the voltage specified in the calibration certificate. If it does not, something is wrong and the transducer needs to be inspected for possible damage or recalibrated.



- **Sensitivity:** The ratio of a change in measurement value to a change in measured variable. For example, if a person ate a 1 lb steak and the bathroom scale shows 2 lb increase in the body weight, then the scale can be called as insensitive (ratio is far from unity).
- **Traceability:** The step-by-step transfer process by which the transducer calibration can be related to primary standards. During any calibration process, transducer is compared to a known standard. National or international institutions usually prescribe the standards. In the United States, such governing body is the NIST.
- **Measurement errors:** Any discrepancies between the measured values and the reported results over the entire FS. Such errors include, but are not limited to:
 - **Nonlinearity:** The maximum deviation of the calibration points from a regression line (best fit to the data), expressed as a percentage of the rated FS output and measured on increasing load only.
 - **Repeatability:** The maximum difference between transducer output readings for repeated applied loads under identical loading and environmental conditions. It indicates the ability of an instrument to give identical results in successive readings.
 - **Hysteresis:** The maximum difference between transducer output readings for the same applied load. One reading is obtained by increasing the load from zero and the other reading is obtained by decreasing the load from the rated FS load. Measurements should be taken as rapidly as possible to minimize creep.
 - **Return to zero:** The difference in two readings: one, at no load, and the second one, after the FS load was applied and removed.

A good transducer is one with the combined (or maximum) error of less than 1% of the FS.

Analog-to-Digital Converters

In order to convey analog output (in volts) from a transducer to a computer, it has to be converted into a sequence of binary digital numbers. Modern analog-to-digital converter (ADC) boards are sophisticated high-speed electronic devices that are classified by the input resolution, as well as the range of input voltages and sampling rates.

Resolution of an ADC board is measured in bits. Bit (abbreviation for binary unit) is a unit of information equal to one binary decision (such as “yes or no,” “on or

off”). A 12-bit system provides a resolution of one part in $2^{12} = 4096$, or approximately 0.025% of FS. Likewise, 16 bits correspond to one part in $2^{16} = 65,536$, or approximately 0.0015% of FS (for tablet press applications, such resolution is usually excessive).

Thus, resolution of ADC board limits not only the dynamic range but also the overall system accuracy. Alternatively, a higher resolution may be required to retain a certain level of accuracy within a given dynamic range. For example, a 0.5% accurate transducer with 80 dB dynamic range requires at least 12-bit ADC resolution.

Amplifying a low-level signal by 10 or 100 times increases the effective resolution by more than 3 and 6 bits, respectively. On the other hand, increasing an ADC resolution cannot benefit the overall system accuracy if other components, such as amplifier or transducer, have a lower resolution.

When an input signal change is smaller than the system's minimum resolution, then that event will go undetected. For example, for an FS of 10 V (corresponding to, say, a compression transducer output of 50 kN), using a 12-bit ADC, any signal that does not exceed 2.44 mV (12.2 N) will not be seen by the system.

The ADC boards also differ by the effective sampling rate range. Sampling rate speed is measured in Hertz (times per second). The signals coming from a tablet press have a frequency of not more than 100 Hz (compression events per second). To avoid aliasing (losing resolution of the incoming signal due to slow sampling rate), the sampling rate should be at least twice the highest frequency of the signal. Most ADC boards used for data collection in tableting applications have a sampling rate of 5–20 times larger than signal frequency. That is why antialiasing filters are not required.

Computers and Data Acquisition Software

Overall accuracy of a data acquisition system is determined by the worst-case error of all its components. One should be aware of the fact that most system errors come neither from transducers (0.5%–1% accuracy) nor from A/D converters (0.025% accuracy) but from the software analyzing the data (round-off errors, improper sampling rate, or algorithms).

The speed and capacity of a data acquisition system depend on the computer's processor and hard drive specifications. The real-time data from transducers is streamlined to both the screen (for monitoring) and the disk (for replay and analysis). Generally speaking, “real-time” processing means reporting any change in the phenomena under study as it happens. Interestingly, but

a high-speed data collection from a tablet press and a bookkeeping home finance program used on a monthly basis to balance the checkbook can both be related to as “real-time” software. The difference is in the time frames. For a tablet press, we are collecting data that need to be sampled and processed in milliseconds (a typical compression event may take 25 msec), while for home bookkeeping once a month will do.

Most vendors supplying transducers, signal conditioning, and computer hardware adhere to practical standards, e.g., there are some accepted norms for strain gage factors, combined errors, sensitivity, ADC resolution, and sampling rate. The difference between vendors is apparent when we compare software because there are no universally established standards of user interface. Yet the hardware is “transparent” (i.e., invisible) to the end user—day in and day out the user is facing the screen, keyboard, and mouse. The ease of software use, bug-free analysis of signals coming from transducers, reliability of statistical computations, quality of graphs and reports, and validatability of the system—all of the above contribute to the quality of the data acquisition software.

Proper validation tests of a data acquisition system should include calculation of an overall system error when the input is known and controlled (e.g., an NIST traceable signal generator providing a sinusoidal signal with a known amplitude and frequency, to simulate compression events on a press). Comparing the output (for example, peak heights as reported by the software) to the known input, the overall system error can be reliably established.

TABLET PRESS INSTRUMENTATION FOR PRODUCT AND PROCESS DEVELOPMENT

R&D Grade Instrumentation

In a production environment, typical tablet weight control mechanism is driven by signals that are coming from a compression force transducer. Strain gages may be installed on a column connecting the upper and lower compression wheel assemblies. This transducer measures what is generally known as “main compression,” i.e., a diluted average of the forces acting on the upper and lower punches. Alternatively, a load cell may be positioned in some convenient location to register a transmitted compression force. It is measured away from the force application axis, and some force is lost in the transmission. The resulting measurement may be adequate for tablet weight control, but, even if properly calibrated, it does not represent the absolute value of the compression force.

The R&D grade instrumentation, in contrast, requires placing the strain gages as close to the punch tips as possible in a vertical alignment with the direction of force application.^[30] In practice, it means placing the gages in a compression roll pin, so that the resulting measurement would reflect the absolute force. Thus, we can differentiate between the force on the upper and lower punch, and moreover, we can compare readings of the compression force from different tablet presses.

Compression Force Measurement

On a typical R&D grade compression transducer, a compression roll pin is machined with incisions made for placing strain gages (Fig. 4). The actual form of these cavities constitutes the very art of the transducer design that is usually proprietary and is based on the know-how of instrumentation vendor.

It has to be noted that there could be an upper or lower instrumented compression roll pin.

The resulting measurement of a compression force is highly correlated with a variety of tablet properties. As compression increases, so does tablet hardness and weight (at constant thickness and true density), along with a force required to eject a tablet. Many variables affect the force of compression: press settings, press speed, punch length variation, punch wear, and damage, formulation and excipient properties.

Precompression Force Measurement

Similar to instrumented compression pin, there is a way to instrument an upper or lower instrumented precompression roll pin. Precompression, if it exists on a press, is used for de-aerating and initial tamping of the powder mass in the die and usually helps to achieve the desired hardness without capping or lamination.^[30,31] The force

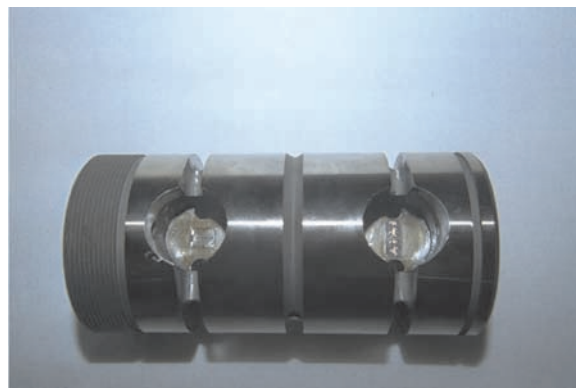


Fig. 4 Compression force transducer.

Tablet Press Instrumentation

of precompression is normally a fraction of the compression force.

Ejection Force Measurement

There are many ways to instrument an ejection cam. A preferred arrangement of strain gages (the so-called “shear force” design) does not require any discontinuities in the cam surface, and, most importantly, it provides for a very good linearity of the resulting signal (Fig. 5).

Larger ejection forces may lead to an increased wear on tooling and eject cam surface. Ejection force may also be used to evaluate the effectiveness of lubrication (of both the press and the product) and punch sticking. Sensitivity and linearity of an ejection transducer are design dependent: shear force designs are always preferable over split cam or cantilevered beam designs.

Take-Off Force Measurement

Take-off force is monitored by mounting a strain gage to a cantilever beam on a press feed frame (in front of the take-off blade, Fig. 6). It is done to measure adhesion of tablets to lower punch face. Such adhesion is indicative of underlubricated granulation, poor tooling face design, die-wall binding, and tablet capping.^[32,33]

Speed Measurement and Station ID Determination

Station identification and press speed are usually obtained by means of a revolution counter (proximity switch). It is installed on the press in order to mark the beginning of a turret revolution and thus enable station identification and speed calculations by system software.

In addition, a linear arrangement of proximity switches can mark the beginning and end of a compression event



Fig. 5 Instrumented ejection cam.



Fig. 6 Instrumented take-off bar.

(information that is vital for tablet weight controllers) and also indicate missing punches.

Other points of instrumentation on a tablet press are as follows:

- A rotary press pull-up cam can be instrumented to measure the upper punch pull-up force (the force required to pull up the upper punch from the die). Likewise, the lower punch pull-down force is measured on a bolt holding the pull-down cam.^[34] It is useful in determining the smoothness of press operation (extent of lubrication, cleanliness of the machine, and long batch fatigue buildup).
- Punch displacement measurements are easily done on a single station press by attaching LVDT to the punch. On a rotary press, such measurements can be done by means of slip ring, telemetry, or instrumented punch. Punch displacement profiles may be used in conjunction with compression force to estimate work of compression and work of expansion (measure of elasticity). Because capping tendency increases with the punch penetration depth, it may be desirable to monitor actual punch movement into the die. The shape of a force–displacement curve is an indication of the relative elasticity or plasticity of the material; whereas plastic deformation is desirable for stronger tablets, excess plasticity usually results in tablets that tend to cap and laminate.^[35–37]
- Radial and axial die-wall force measurements also provide an insight into the compaction mechanism of the material and may indicate a die-wall binding (sticking) that is, in effect, a negative pull on lower punch. The radial die-wall pressure due to friction is material-specific and is more evenly distributed inside the die with an addition of a lubricant.^[38–44] Instrumentation of the die presents a technological challenge because pressure is distributed nonlinearly



with respect to tablet position inside the die and depends on tablet thickness.^[45,46]

- In-die temperature can be monitored for heat-sensitive formulation, such as ibuprofen.^[47,48]

Instrumented Punch

Several vendors offer instrumented punch, i.e., a punch that has strain gages and other instrumentation built-in. The data are then accumulated or transmitted via telemetry to a stationary computer. Such devices are versatile enough to report compression forces and either punch displacement or acceleration, and, at least in theory, they can be easily moved from press to press.^[49,50] However, one should keep in mind that each instrumented punch is limited to one size and shape of the tooling, and is limited to one station, compared to roll pin instrumentation that reports data for all stations and any tooling. In addition, instrumented punches are either rather cumbersome to install, or else they report a useless measurement of punch acceleration instead of punch displacement. Attempts to calculate displacement from acceleration so far could not be validated.

Single Punch Press for R&D

Single punch eccentric presses are often used in early stages of formulation development because they do not require a large amount of powder. Another benefit is that they allow relatively inexpensive measurements of die forces and punch displacement (there is no rotation of die table and therefore no need to use expensive telemetry methods). This is the primary reason why so much basic research and product development were done on eccentric machines.^[51] Negative considerations are that a special tooling is required (usually, F tooling), and also that speed of compaction is too slow compared to rotary presses. As it will be seen later, the speed is a crucial factor in tableting process and therefore the results obtained on single punch presses do not directly correlate with tablets made on rotary machines.

Benefits of Press Instrumentation

Among many benefits of press instrumentation, formulation fingerprinting is perhaps the most obvious. Compressibility and ejection profiles, as well as dissolution and disintegration curves related to compression force, are unique for each formulation and can be used as a batch record. For process optimization, one can include compression or precompression force and speed factors in

experimental design. Compactibility and ejection profiles can be used for excipient and lubricant evaluation. Other useful product development and optimization tools include response surface, Heckel, and force–thickness plots.

Scientifically reliable process scale-up cannot be achieved without instrumentation data providing scale-up parameters such as dwell time, density, and energy of a tablet.

In a pilot plant and on the production floor, proper instrumentation can be used for press troubleshooting, to warn about tooling irregularities, worn-out cams, sticking punches, underlubricated dies, and so on. Finally, instrumentation is widely used for tablet weight control.

Instrumentation for Formulation Development

Much of the current body of knowledge about compaction properties of pharmaceutical materials came from instrumented tablet presses.^[52–56] Many tablet properties, such as tensile strength (hardness) and porosity can be predicted from force profiles.^[57–59]

Work of compaction (a scale-up parameter) can be obtained with proper instrumentation.^[60,61] Information about the plasticity of materials can be derived from force–time curves.^[62–64]

The phenomenon studied with the help of instrumentation is the so-called “lag time” (the time difference between peak of compression and maximum punch penetration). The extent of this lag is indicative of compaction mechanisms of the powder being compressed.^[65,66]

Typical waveform

Let us have a look at the actual waveforms that one may obtain from an instrumented tablet press (Fig. 7). The black proximity switch trace that marks the beginning of a revolution should be noticed. On this press, precompression and compression events coincide in time (they relate to different punches, of course). They are followed by the ejection and take-off.

Compactibility profile

When the average tablet hardness is plotted against the average compression peak force, we get the so-called compactibility profile that allows us to compare different formulations or different processing speeds. Referring to Fig. 8, which formulation is better? Well, formulation No. 2 makes harder tablets for the same compression force, and this would mean less wear and tear on the production press is required to achieve desired hardness. On the other

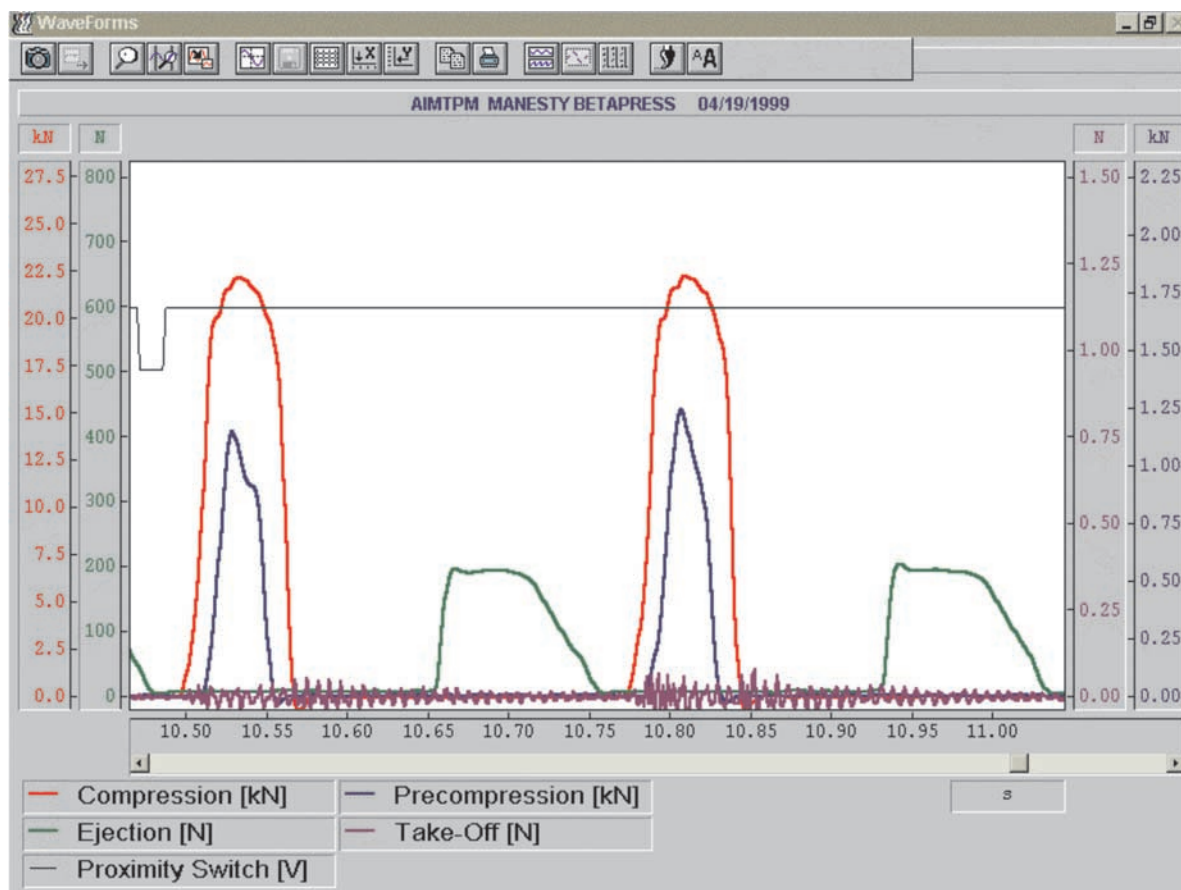


Fig. 7 Typical compression, precompression, and ejection waveforms.

hand, if the hardness tolerance limits are exceptionally narrow, the steeper slope of formulation No. 2 may be a detriment.

Lubricant profile

A certain quantity of lubricant must be present in the granulation to reduce the friction that occurs at the die wall as the tablet is being ejected, as well as to prevent sticking of the tablet to the face of the punches. Without instrumented ejection cam and take-off bar, no objective estimate of an optimal lubricant level is possible, and the “best” formulation is usually the last one prepared.

The plot shown in Fig. 9 will help a formulator to determine the optimal amount of lubricant. Obviously, one would try to minimize the ejection force (again, to reduce wear and tear on the ejection cam of the production press), and yet to avoid the pitfalls of having too much lubricant in the formulation.

Because of the natural association of lubricant properties with lipophilic materials, formulations containing high levels of lubricant can show retarded dissolution of the

active ingredient and the slow dissolution rate could adversely affect the in vivo bioavailability of the drug.

Lubricant study

Another important use of ejection force transducer is the evaluation of lubricants (either different chemically, or similar lubricants coming from different vendors).

As shown in Fig. 10, the preferred lubricant is No. 1 as it results in a smaller ejection force. Early lubricant studies are a must. When lubricant problems occur later on in the scale-up process, corrective measures not only require additional materials and development time, but may also require a supplement of an approved new drug application (NDA).

Response surface plot

Formulation and process optimization can be done statistically with the use of experimental design for estimates of the best processing parameters and excipient

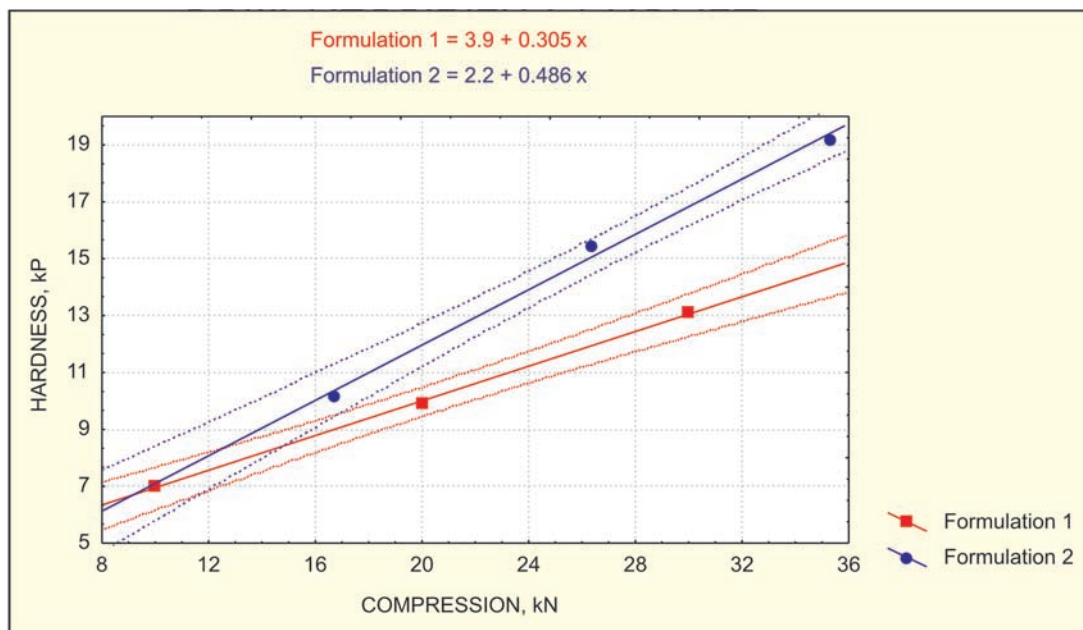


Fig. 8 Compactibility profile.

and lubricant levels. Controllable variables in tableting are mainly the precompression and compression forces and tablet press speed, as well as the formulation component levels. Response variables include the ejection force, tablet hardness and friability, dissolution rate, and drug stability. The purpose of an experimental design is to perform a series of experiments in order to determine some levels of factors that will allow us to achieve an optimal

level of dependent variables. The experiments should be designed so as to minimize effort and maximize statistical reliability of the results. Published work in this area deals mostly with response surface designs that produce a predictor polynomial equation for each response variable under consideration. A multidimensional surface is then searched for the best ranges of factor variables that, when plugged in the equations, result in the optimal value of the

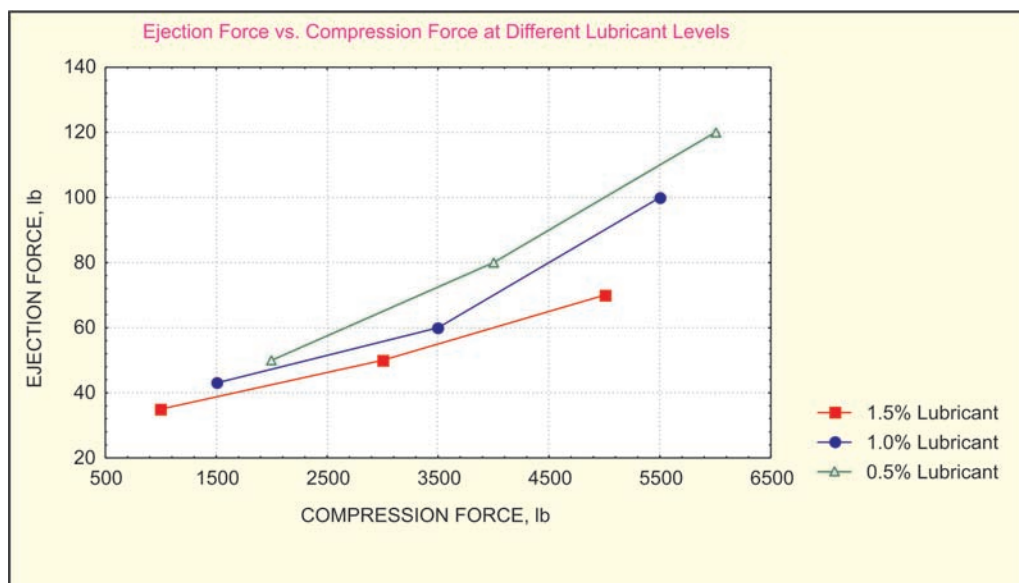


Fig. 9 Lubricant profile.

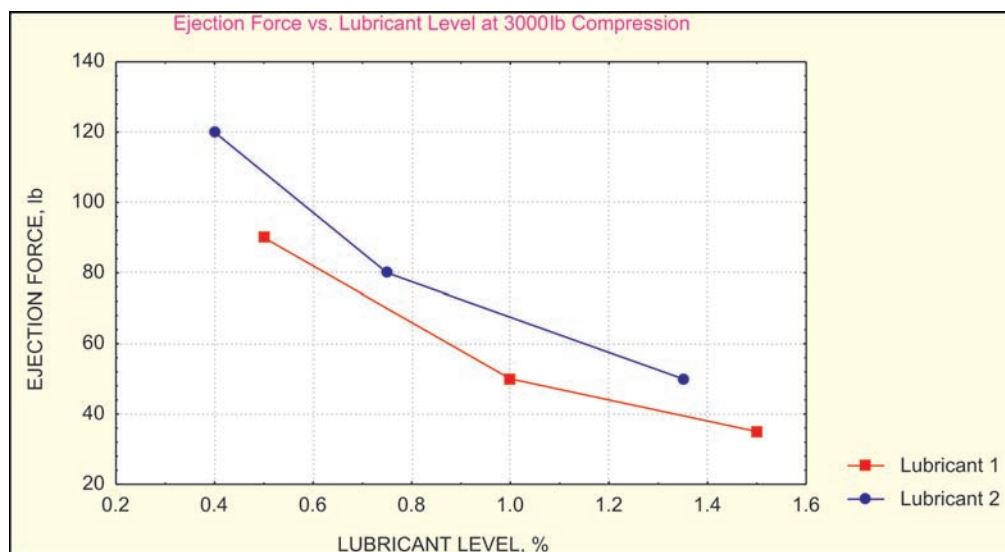


Fig. 10 Lubricant study.

responses. In this plot (Fig. 11), the optimal (highest) hardness is obtained when the compression force is in the range 3000–4000 lb with as much MCC in the formulation as possible.^[67]

Compactibility study—elastic recovery

Many powders, especially with viscoelastic compaction mechanism, such as starch or avicel, exhibit large degree of stress relaxation (with time-dependent deformation).

If you monitor punch displacement and compression force, you can make pretty accurate assessment of the compactibility of your material. If LVDTs are attached to both upper and lower punches, it is possible to actually measure in-die thickness of the tablet at various compression levels. In Fig. 12, one can see how the tablet thickness rapidly decreases in the compression stage and then gains some thickness during the decompression stage of the tableting cycle. The degree of this increase in

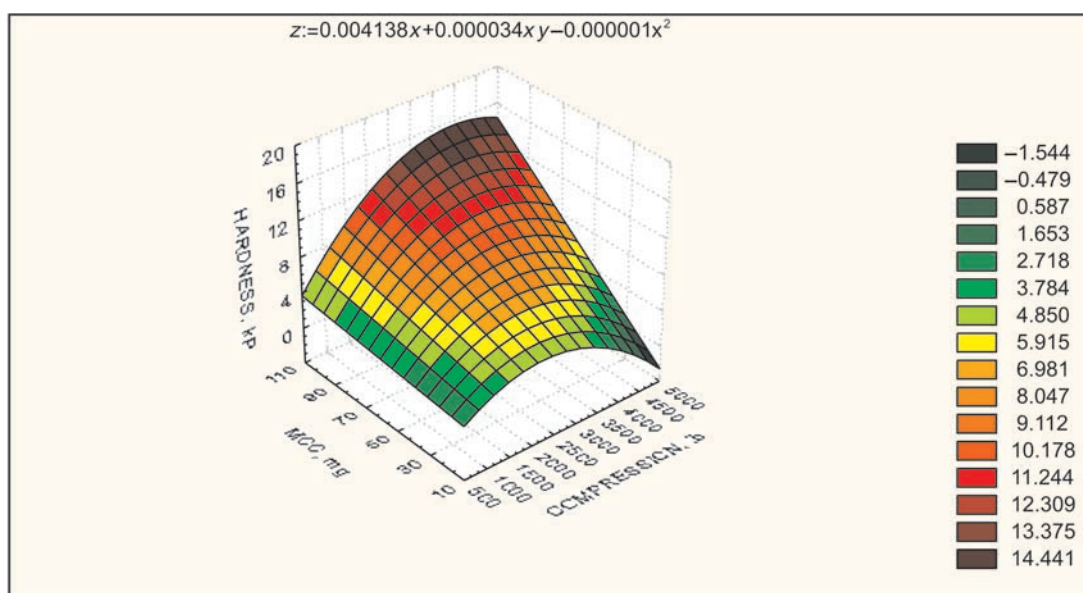


Fig. 11 Response surface plot.

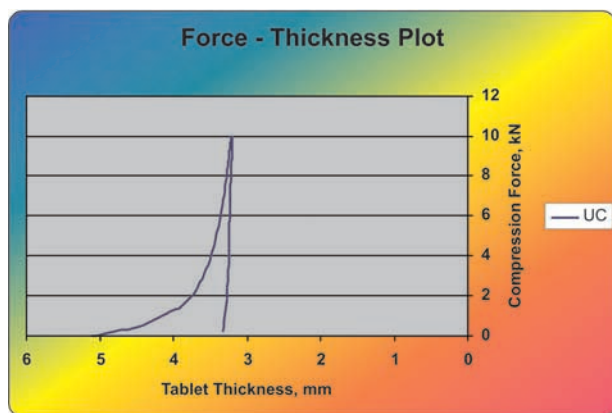


Fig. 12 Compactibility study—elastic recovery.

thickness is indicative of the elastic recovery as the pressure is removed from the tablet.

Elastic recovery and work of compaction were studied extensively using instrumented tablet presses.^[68]

Compactibility study—Heckel plot

In 1961 Heckel^[69,70] postulated a linear relationship between the relative porosity (inverse density) of a powder and the applied pressure. The slope of the linear regression is the Heckel constant, a material-dependent parameter inversely proportional to the mean yield pressure (the minimum pressure required to cause deformation of the material undergoing compression). Large values of the Heckel constant indicate susceptibility to plastic deformation at low pressures, when the tablet strength depends on the particle size of the original powder. The intercept of the line indicates the degree of densification by particle rearrangement (Fig. 13).

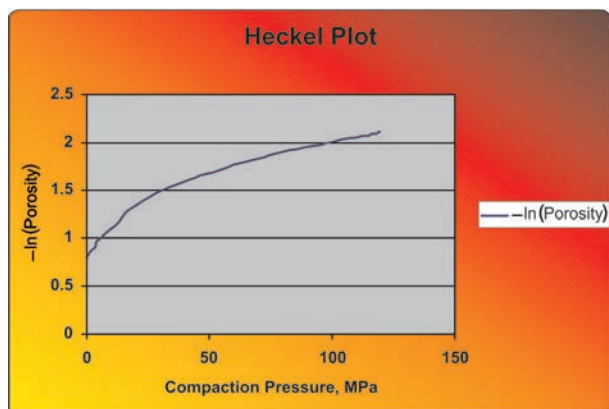


Fig. 13 Heckel plot.

TABLET PRESS INSTRUMENTATION FOR PROCESS TROUBLESHOOTING

Typical problems in tableting include weight variation, capping, lamination, tooling irregularities, die binding, picking, and sticking. Most of such problems can be detected and/or resolved using proper instrumentation.

Upper and Lower Compression

Fig. 14 shows a typical set of upper and lower compression profiles. One can see that the lower trace is smaller than the upper. On a single station press, only the upper punch is usually moving, and the difference is caused, mainly, by the friction of the compressed powder inside the die.

On a rotary press, both punches are moving and therefore the friction of the punches inside the turret and the die causes the difference. The lower punch fits the turret with a larger standard clearance compared to the upper punch. In addition, the lower punch is never leaving the die, while the upper punch leaves and enters the die with each stroke. This results in a comparatively better alignment and lower friction. The close fit of the upper punch does not allow it to penetrate the die smoothly and this can cause the increase in friction. Thus, an excessive difference between the two peaks may indicate an underlubricated die or some punch misalignment problem.

Histogram of Punch Performance

Similar irregularities in punch tolerances become even more visible on a bar histogram where each bar corresponds to peak force produced by each punch (Fig. 15). As the new revolution arrives, the bar lengths

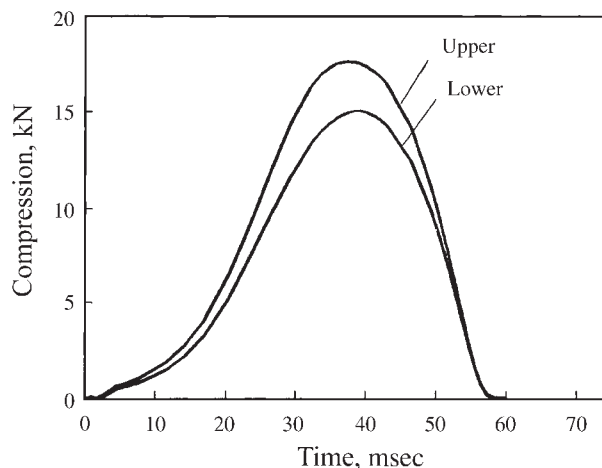


Fig. 14 Upper and lower compression force vs. time.

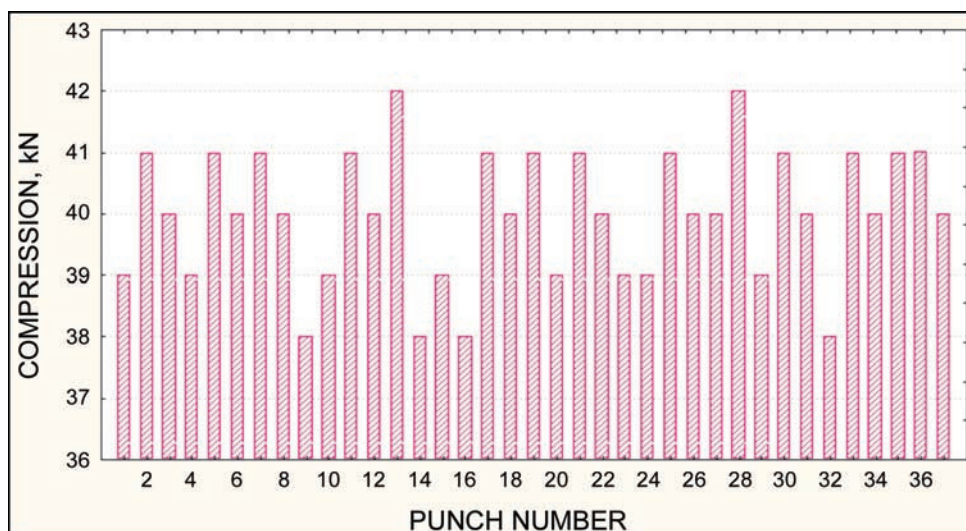


Fig. 15 Histogram of compression force per punch.

are adjusted. If a particular bar persists in being taller or smaller than the rest, it means the presence of a long or short punch.

Press Monitor—Real-Time Screen

On this real-time screen (Fig. 16), one can clearly see that there is an abnormal ejection event in the first station after the proximity switch. This is a clear indication of a sticking punch or a similar problem. In addition, the first compression peak is somewhat taller than others. This may be caused by a long punch.

TABLET PRESS INSTRUMENTATION FOR PROCESS SCALE-UP

Scale-Up Factors

One of the main practical questions facing formulators during development and scale-up is: Will a particular formulation sustain the required high rate of compression force application in a production press without lamination or capping? Usually, such questions are never answered with sufficient credibility, especially when only a small amount of material is available and any trial and error approach may result in costly mistakes along the scale-up path.^[71]

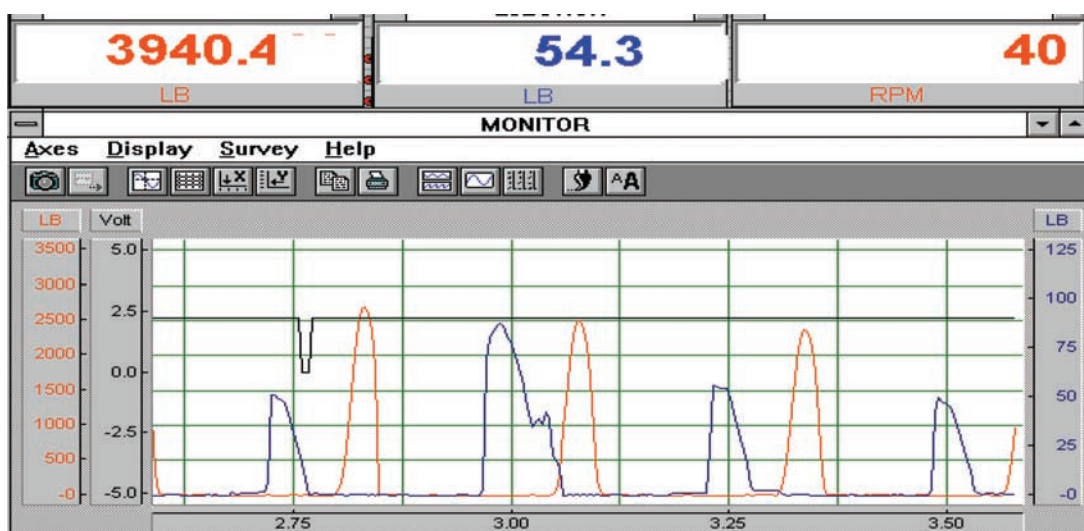


Fig. 16 Real-time screen.



As formulations are moved from small-scale research presses to high-speed machines, potential scale-up problems can be eliminated by simulation of production conditions in the formulation development lab. One way to eliminate potential scale-up problems is to develop formulations that are very robust with respect to processing conditions. A comprehensive database of excipients detailing their material properties may be indispensable for this purpose. However, in practical terms, this cannot be achieved without some means of testing in production environment and, because the initial drug substance is usually available in small quantities, some form of simulation is required on a small scale. Studies of tableting process on a class of equipment, generally known as compaction simulators, are designed to facilitate the development of robust formulations. However, simulators are rarely used to simulate tablet presses for reasons that will be explained later.

In any process transfer from one tablet press to another, one may aim to preserve mechanical properties of a tablet (density, and, by extension, energy used to obtain it) as well as its bioavailability (e.g., dissolution that may be affected by porosity). However, a formulation that was successfully developed on a single station or small rotary press may not stand up to the challenges of scale-up because tablets that were meeting all specifications in the lab or clinical studies may exhibit capping or lamination at higher speeds.^[72,73]

Compression force magnitude and the rate of force application are the most important variables in tableting scale-up.

Force factor

The compression force is the dominant factor of the tableting process. It is directly related to tablet hardness and friability, and is correlated with the phenomena of lamination and capping.

It was also shown to have effect on disintegration times and dissolution profile.

Speed factor

As the punch speed increases, so does the porosity of tablets and their propensity to capping and lamination. The tensile strength of compacts tends to decrease with faster speeds, especially for plastic and viscoelastic materials, such as starch, lactose, avicel, ibuprofen, or paracetamol. Such materials have the tendency to cap or laminate at higher speeds.^[74–88]

The notions of dwell time and contact time, to be discussed in detail later, are the common indicators of press speed and the rate of force application.

Force profile factor

In addition to the level of force and the rate of force application, the shape of the compression force vs. time curve is of a paramount importance because it directly affects tablet properties such as hardness and friability. It is a known fact that the compression part of the compression cycle (during the “rise time” of the force–time profile) is 6–15 times more important than the decompression part as a factor contributing to capping and lamination. On the other hand, reducing the decompression part of the cycle results in the increase of tablet hardness by reducing the extent of elastic recovery.^[89] Alternatively, reducing the compression part of the cycle results in no change of tablet strength for viscoelastic materials and increased hardness for brittle materials.

Other Considerations

Numerous other factors may affect the scale-up process. The quality of the measurements, variation in tooling, powder properties, and tablet weight are some of those factors.

- Instrumentation grade.
- Measurement of speed.
- Measurements of mechanical strength.
- Tooling variation.
- Powder flow variation.
- Excipient/raw material variation.
- Tablet weight variation.

Dwell Time and Contact Time

Matching tablet press speed (rpm) of the research and production presses has, of course, no meaning, because of different number of stations and pitch circle diameter. It is vital, therefore, to translate the rpm into dwell time or contact time.

Many investigators have reported the effect of dwell times and strain rate sensitivity on the compaction of various excipients, especially viscoelastic materials.^[90–94] There are at least two definitions of dwell time in practice today.

Functional definitions

Functionally, the effective dwell time (EDT) at 90% level can be defined as the time it taken by the force–time curve

to traverse the 90% of the peak height. Likewise, the effective contact time (ECT) is the time between points at 10% of the peak height. The shape of the force curve depends, as we know, on the deformation mechanisms of the powder and therefore, all other variables being equal, EDT and ECT will be different for brittle and plastic materials. Although somewhat useful for material characterization, such variables should not be confused with the universally accepted definitions of contact time (time of contact) and dwell time (time of immobility).

Mechanical definitions

Mechanical Definitions of dwell and contact times disregard material properties and concentrate on press and punch geometry (Fig. 17). Contact time can be defined as the time the punch is in contact with the compression wheel. Dwell time is defined as the time the flat portion of punch head is in contact with the compression wheel (time at maximum punch displacement, or time when the punch does not move in vertical direction). In dwell time calculations, the length of the punch head flat and horizontal component of punch speed (as determined by RPM and pitch circle diameter) are used. In case of a round head tooling, the dwell time, as defined here, is zero. But it should be kept in mind that mechanical definition is given here as a convention, a yardstick, or a common measure, to compare press speeds for different presses, and its absolute value is meaningless. A proposed convention to quantify linear speed of a press is to use an Industrial Pharmaceutical Technology (IPT) Type B tooling with a known punch head flat as a standard for press speed comparisons.

In what follows, we will use the mechanical, rather than functional, definitions because they serve as an objective material-independent measure of compaction speed.

Dwell time comparison for different presses

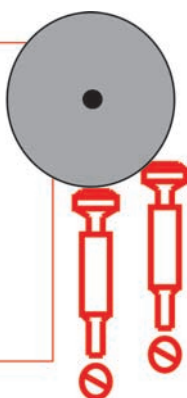
Comparing the dwell time ranges for a number of presses can be a gratifying experience. Here one can see that even for the same brand name, there is a wide distribution of ranges. It has to be noted that proper scale-up will be possible only when the ranges overlap. Dwell time range may also be used to classify or identify tableting equipment. The dwell time ranges vary considerably in various tablet presses. As shown in Fig. 18, Manesty Betapress is positioned well within the range of production speeds of the high-speed presses. That is, probably, why this press is often used for R&D work. On the other hand, small presses such as Korsch PH106 or Piccola do not even come close to benchmark production speeds of 6 msec–15 msec in terms of dwell time. The MCC Presster™ can reach the production dwell times while making one tablet at a time.

Dwell time vs. production rate

For a benchmark production speed of 100,000 tablets per hour (Fig. 19), dwell time distribution follows an inverse power relationship, which is expected because dwell time is a reciprocal function of the press speed. In general, all production presses should be qualified with respect to dwell and contact times per benchmark output. Ideally, product development must be done on a laboratory press that can match (in terms of the dwell time) the target production output.

Dimensional Analysis of Tableting Process

Dimensional analysis is a method for producing dimensionless numbers that completely characterize the process. It is widely used in many areas, including chemical engineering and pharmaceutical industry.^[71] Because all


$$DT \text{ (msec)} = \frac{L \cdot NS \cdot 3,600,000}{\pi \cdot PCD \cdot TPH}$$

L = Length of a flat portion of the punch head (mm)
 NS = Number of Stations
 π = 3.14159265
 PCD = Pitch Circle Diameter of the turret (mm)
 TPH = press speed in terms of Tablets Per Hour

Fig. 17 Dwell time and contact time—mechanical definitions.

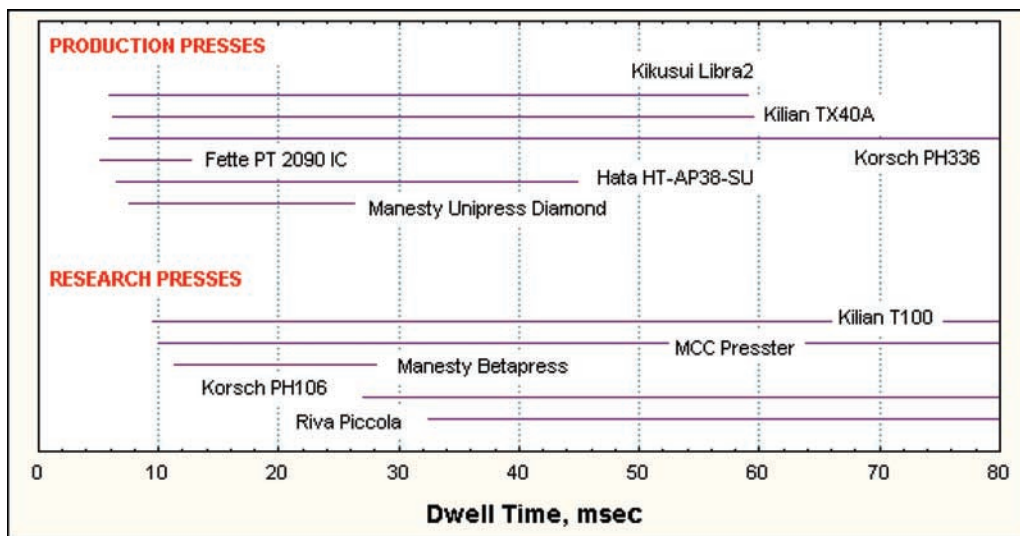


Fig. 18 Dwell time comparison for different presses.

the dimensionless numbers necessary to describe the process in similar systems must have the same numerical value,^[95] matching such values on different scales is a sure way to success in any scale-up operation. This dimensionless space in which the measurements are presented or measured will make the process scale invariant.

In tableting applications, the process scale-up involves different speeds of production in what is essentially the same unit volume (die cavity in which the compaction takes place). Thus, one of the conditions of the theory of models (similar geometric space) is met. However, there are still kinematic and dynamic parameters that need to be investigated and matched for any process transfer.

Scientifically sound approach would be to use the results of the dimensional analysis to model a particular production environment to facilitate the scale-up of tableting process, by matching several major factors, such as compression force and rate of its application (punch velocity and displacement) in their dimensionless equivalent form.^[96]

TABLET PRESS AND COMPACTION SIMULATORS

It can be seen that, as a rule, tablets are formulated at speeds that are very slow compared to production. If

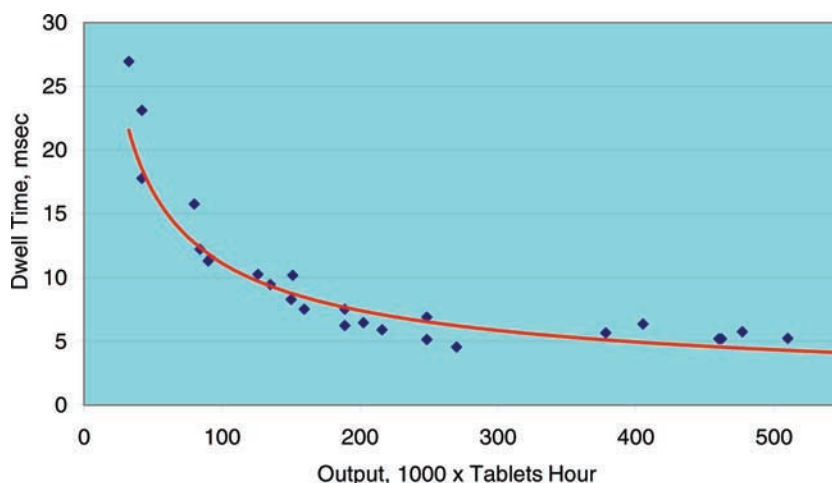


Fig. 19 Dwell time vs. production rate.

simulation of a production press is required to minimize the scale-up effort, some way of speeding up the tableting process in development is required. Once the linear speed of the punch is attained, the rate of force application (i.e., the instantaneous change in compression) should also be matched. This is, of course, an infinitely more difficult task.

Hydraulic Compaction Simulator

A small number of devices known as compaction simulators exist in the world. Invented more than 20 years ago, they become popular in basic compaction research. A wealth of studies have been generated in the last 20 years.^[97–117]

Compaction simulators (Fig. 20) were designed to mimic the compression cycle of any prescribed shape by using hydraulic control mechanisms that are driving a set of two punches (upper and lower) in and out of the die. All hydraulic compaction simulators are similar in design and construction. A compaction simulator consists of several main units: the load frame (column supports and crossheads with punches), the hydraulic unit (pumps and actuators that move the crossheads), and the control unit (electronic console and computer). Usually, a simulator

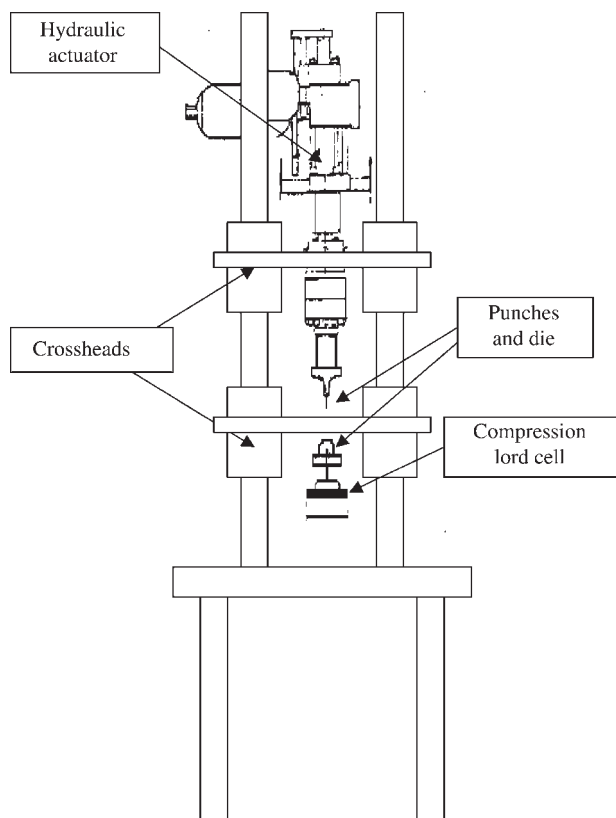


Fig. 20 Schematic depiction of a compaction simulator.

accepts F tooling only, but can be retrofitted to use standard IPT B tooling. Under computer control, the hydraulic actuators maintain load, position, and strain associated with each punch.

The simulation can be achieved by one of the two procedures: matching either the force (load control) or the movement of punches (position control) at any given moment of time. Thus, when running a simulator, one has a choice: to mimic the force/time path (compression profile) or the motion of the punches (punch displacement curves). It is impossible to mimic both at the same time on any hydraulic compaction simulator.

Load control profile

Matching the force–time profile of a production tablet press is the primary goal of any tablet press simulation. However, the rate of force application and the shape of the resulting signal are not usually known in advance. The compression profile (force vs. time) is impossible to calculate theoretically because it has a different shape for different materials and tooling.

Using an instrumented punch to collect force data is cumbersome because it is limited to a particular punch size and shape. Recalculation to pressure values is not always adequate. One can, however, monitor and record force waveform from a properly calibrated R&D grade compression transducer. Once the production press brand, model, speed, and tooling are specified, a waveform can be recorded and then fed into hydraulic simulator.

This approach is impractical for formulation development work because one would need to record force waveform on a rotary press using formulation similar to the one being developed. Usually, there is no such powder available and the actual formulation being developed is available in limited quantities.

Position control profile

Because load control profiles are not practical, users of hydraulic compaction simulators overwhelmingly prefer to utilize punch displacement profiles in hope that, once the punches are forced to move in the same pattern as in the production press, the force–time curve will follow. For example, a recently built laboratory compaction simulator does not have a force control functionality at all.^[118]

There are three possible sources of the punch position trace:

- Prerecorded data.
- Artificial profiles.
- Theoretical profiles.

To obtain a position control profile from any tablet press, one has to record the movement of the punches using LVDT. Besides the technological challenge that this objective may present, the punch movements on a press depend on many factors, including brand name and model of the press, speed of the turret, shape of punches and die, size and shape of the tablet, and most importantly, compaction properties of the powder. The problem with this is that data from production presses are inevitably obtained using material other than the one being developed. It is a vicious circle: the profiles before developing a formulation and the formulation before obtaining punch displacement profiles are needed.

Many compaction studies were done on compaction simulators using artificial punch displacement profiles, for example, the so-called “single-ended” profile, i.e. when the lower punch is stationary (like in a single station press). Other studies were using a “saw tooth” profile, i.e., a constant speed profile where punch displacement speed is constant at any given time interval under load. It is obvious that such profiles have nothing to do with simulation, although they provide a degree of uniformity for basic compaction studies. The very name “compaction simulator” is a misnomer as is acknowledged by a number of researchers in the field. The machine is best described as a compaction research system because it is well suited for the basic compaction studies (densification and bonding properties of materials).

To simulate tablet presses, compaction simulator users most frequently employ the theoretical position control profiles. Theoretical path is calculated from the geometry of the press and punches, using the radius of the compression roll, the radius of the curvature of the punch head rim, the radius of the “pitch circle” (distance between turret and punch axes), and the turret angular velocity.^[119]

The resulting sinusoidal equation is used in order to “simulate” punch movement in a tablet press.

In practice, theoretical and actual punch displacement profiles on a rotary press have very little in common because the theoretical profile does not account for several mechanical factors, such as punch head flat. Moreover, the punch movement equation was derived for a punch moving in and out of an empty die. The effect of material resistance to pressure and elastic recovery is not accounted for in the equation. The discrepancies between the calculated and real punch movements are rather striking.^[120–122]

In addition, it was shown that the lower and upper punches may not move synchronously. Moreover, maximum force does not coincide in time with the minimum punch gap.^[123] These and other considerations (press deformation, contact time, etc.) make the effort of simulating a production press on a hydraulic compaction

simulator rather impractical. That is why, to quote from a paper by Muller and Augsburger,^[124] “Although compaction simulator have been designed to mimic the displacement time behavior of any tablet press, they rarely have been used in that fashion.”

Literature sources reporting the use of compaction simulators in simulating actual tablet presses are rather scarce. Some studies suggest that, for whatever reason, tablets made on a Manesty Betapress were significantly softer than those made on a compaction simulator using the theoretical Betapress punch displacement profiles.^[120,125]

To summarize, one can say that hydraulic compaction simulators are ideally suited for basic compaction research but are not very practical for simulation of production presses.

Tablet Press Replicator: The Presster™

Recently, a new type of machine was introduced to mimic production presses on a small scale. Known under a brand name of “Presster”—kind of an agglomeration of the words “press,” “presto,” and “tester”—this machine can be classified as a mechanical compaction simulator.

The Presster™ is a high-speed single station press that is also a tablet press simulator (Fig. 21). It was designed to simulate production presses without any use of hydraulic controls, and, consequently, there is no need to feed in any artificial, theoretical, or prerecorded punch displacement profiles. Built around a linear carriage that moves a set of punches and a die between two compression rolls, it can mimic press geometry by matching the compression wheels match press speed using a variable speed motor drive match tablet weight and thickness by adjusting depth of fill and the distance between the rolls match tooling by installing standard IPT or any special tooling.

Thus, using mechanical similarity, all of the scale-up fact are matched, namely, the compression force, the speed, and the shape of the force profile. To use Presster, first a production press to be simulate should be selected, and the compression wheels with a matching diameter should be installed. Then, production speed should be selected in terms of tablets per hour, RPM, or dwell time. The Presster™ will mimic the selected production press speed and compression force profile and will allow us to make one tablet at a time.

As a high-speed single station press, mechanical compaction simulator will be able to plot compressibility profiles, Heckel graphs, calculate work of compaction, and virtually any other imaginable variable that is of interest to formulators. Tensile strength of tablets made on a Betapress and The Presster™ was similar.^[126]



Fig. 21 Tablet press replicator: The Presster™.

Precompression and ejection steps of the tableting cycle can be included in simulation. Some current limitations of Presster should be pointed out: It will neither follow any artificial punch movement profile, nor will it address, at least in its present implementation, the issues of feeding and die fill at high speeds, or speed-related temperature fluctuations.

PRESS INSTRUMENTATION AND CONTROL ON THE PRODUCTION FLOOR

Tablet Weight Control and Tablet Force Control

To keep tablet weight within the prescribed tolerance limits, the required instrumentation includes compression transducer and several proximity switches for station identification and pinpointing the compression event. Tablet weight controller can be just one, albeit a major, unit of a larger press automation system. Press automation system may include:

- Tablet weight control.
- Material handling interface.
- Feeding system.

- Collection system.
- Packaging system.
- Supervisory control (SCADA) station.

The latter can be a computer positioned on the supervisor's desk that monitors the performance of each press on the production floor, with timely status report for each batch.

There are many reasons why it is imperative to control tablet weight variation:

- Production costs are lowered because there is less waste.
- Productivity is increased because there is a better equipment utilization.
- Batch-to-batch variability is minimized for obvious reasons.

When the cost of each out-of-spec wasted tablet is significant, the savings produced by weight control quickly add up. Product quality is improved when there is less of a chance to get an out-of-spec tablet into the acceptable batch. Last, but not least, is the improved safety because the automated process requires less human intervention.



Control mechanisms

For constant tablet thickness, in a small target area of tablet weight, compression force is directly proportional to tablet weight. Force control systems maintain compression force within prescribed limits by adjusting the depth of fill cam. The limits are established empirically for each formulation recipe.

Alternatively, a weight control system would require an expressed correlation between force and tablet weight. A few tablets at different force levels can be made to correlate the resulting tablet weights with the force values. Next, one can express tablet weight tolerance limits in terms of the compression force. The weight control system can adjust the dosing to keep the tablet weight within the desired limits. For this control theory to work consistently, one needs to recalibrate force–weight relationship periodically, as the powder properties and tablet mechanical condition can vary in time.

Control functions include alarm and shutdown (when any individual force peak or revolution average exceeds preset limits), force or weight control (usually done on revolution average only), and rejection of out-of-spec tablets. With a mechanical gate, usually several tablets are rejected at a time, with the bad tablet being rejected along with the adjacent good tablets. With a fast pneumatic–air stream rejection mechanism, the control system can pinpoint and reject one bad tablet only.

Control criteria

Control decisions are usually done as follows:

1. Adjust depth of fill cam if revolution average is outside some redefined acceptable limits.
2. Reject each tablet that exceeds individual limits or if the press speed slows down beyond some level.
3. Sound an alarm if the same station repeatedly produces bad tablets or if revolution average exceeds alarm limits.
4. Shut down the press if any tablet is made with a force exceeding the shutdown limits, or if revolution average is outside the shutdown limits.

Control algorithms

One-point control is a description of a control when any deviation of the compression force from a preset target level results in a corrective action of the dosing cam. This action can be proportional (P) to the deviation from the target, or it could be correlated with the integral (I) of

the deviation over some time, or it could be tied in with the rate of change (D, for derivative) of the compression force. This would correspond to proportional, integral, or derivative control types, respectively. There also can be a combination of the control types. Thus, one can have P, I, P + I, P + D, and P + I + D control algorithms.

Alternatively, a two-point control is enacted when the compression force is outside an acceptable band outlined by the upper and lower tolerance limits. Thus, there are separate control limits, rejection limits, alarm limits, and shutdown limits, and no respective action is taken when the signal is inside these limits.

Strictly speaking, one-point control can be viewed as a special case of the two-point control when the bandwidth of the control limits is tightened to approach zero.

Control Systems

Original equipment manufacturer (OEM) control systems

Almost each tablet press manufacturer offers a system that is designed to control the press. Some of these systems are very sophisticated devices that monitor and control a vast array of tableting functions. If, however, there are several brand names on the production floor, any standards in the control system implementation for different manufacturers should not be expected. Likewise, software interfaces exhibit quite a range of user-friendliness.

In one brand name press, tablet weight control is achieved by regulating the dosing cam based on powder bed thickness in the precompression cycle. This clever approach is possible because precompression force is kept relatively constant by means of pneumatic compensating mechanism. Under these conditions, tablet weight is directly proportional to thickness. The subsequent compression cycle can be done to constant thickness, like on any other press.

Generic control systems

An alternative to OEM control systems is generic controllers that may offer plug-in compatibility with the brand name controllers and may provide a degree of standardization.

However, such generic controllers may lack the degree of sophistication and versatility of the controllers that are made by the press manufacturers. One thing to keep in mind is that not all control systems are created equal, but all control systems use the same principles.

A decent tablet weight control system should be based on product recipes, provide instant display of compression force distribution, control charts and batch reports on

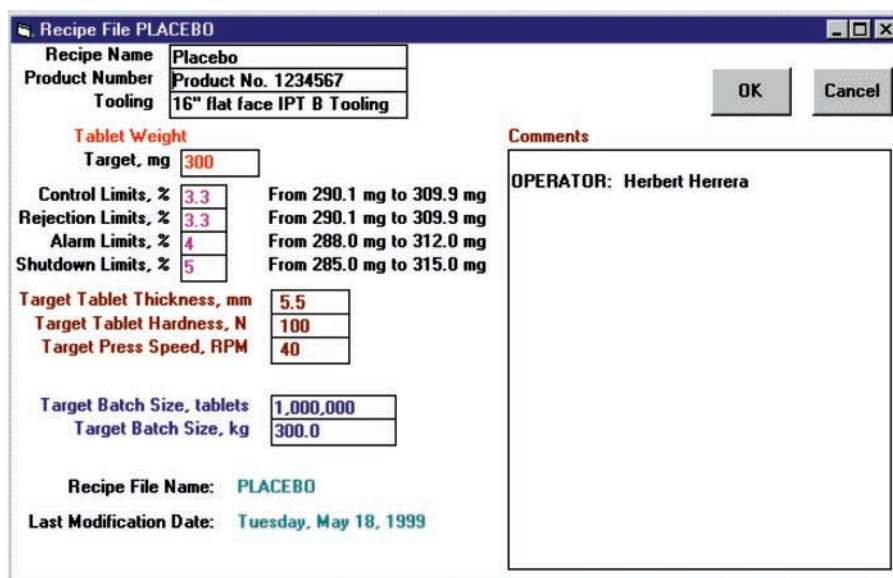


Fig. 22 Recipe example.

demand, archiving data for subsequent analysis or documentation, give some measure of standardization, be fully compliant with current validation requirements, and provide a multilevel security, e.g., password protection for operator and supervisor when they need to change recipes, etc. Some examples of displays, available to operators of instrumented production presses equipped with controllers, follow.

Recipe example

In Fig. 22, an example of a typical control recipe with dosing cam adjustment, rejection, alarm, and shutdown limits expressed as a percent deviation from the target

tablet weight is shown. The software then converts all values into the corresponding compression force levels for control purposes.

Bar histogram

This chart is similar to one used in tablet press monitoring for press troubleshooting (Fig. 15). It is vital for any production press operator.

Control chart

Control chart is a simple graph of peak compression force vs. time. Each point on the chart corresponds to the

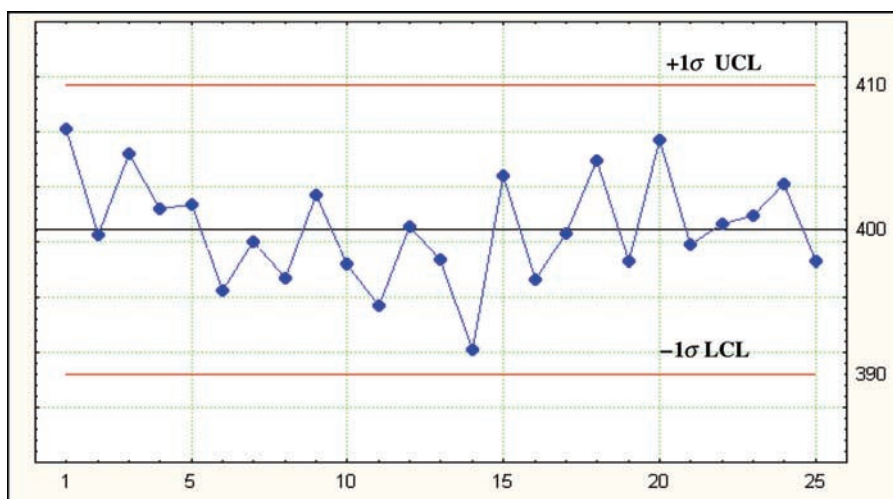


Fig. 23 Statistical process control X-bar chart.



average of N tablets made, with N ranging from 1 to several revolutions. The horizontal lines would indicate the control limits.

Statistical process control chart

Statistical process control (SPC) chart (Fig. 23) of the averages is another “must have” real-time display. Each point on the chart represents a revolution average of the compression forces or corresponding tablet weights. The limit lines are calculated at one standard deviation of the mean, and there are certain rules that are used to determine when and if the process gets out of control. These rules are available in any textbook on the SPC.

ACKNOWLEDGEMENT

Selected excerpts and figures from M. Levin, “Tablet Press Instrumentation for Product Development, Process Scale-Up and Production Quality Control, Education Anytime, CD-ROM Short Course, AAPS 1999” are reprinted with permission.

REFERENCES

1. Higuchi, T.; Nelson, E.; Busse, L.W. The Physics of Tablet Compaction 3: Design and Construction of an Instrumented Tableting Machine. *J. Am. Pharm. Assoc.* **1954**, *43* (6), 344–348.
2. Nelson, E.; Naqvi, S.M.; Busse, L.W.; Higuchi, T. The Physics of Tablet Compression IV: Relationship of Ejection, and Upper and Lower Punch Forces During Compression Process. *J. Am. Pharm. Assoc.* **1954**, *XLIII*(10), 596–602.
3. Nelson, E. The Physics of Tablet Compression VIII: Some Preliminary Measurements of Die Wall Pressure During Tablet Compression. *J. Am. Pharm. Assoc.* **1955**, *44*, 494–497.
4. Knoechel, E.L.; Ross, H.E.; Sperry, C.C. Measurement of Force Within a Tableting Machine. US Patent 3,255,716, 1966.
5. Knoechel, E.L.; Sperry, C.C.; Ross, H.E. Instrumented Rotary Tablet Machines 1: Design, Construction, and Performance as Pharmaceutical Research and Development Tools. *J. Pharm. Sci.* **1967**, *56* (1), 109–115.
6. Knoechel, E.L.; Sperry, C.C.; Lintner, C.J. Instrumented Rotary Tablet Machines 2: Evaluation and Typical Applications in Pharmaceutical Research Development and Production Studies. *J. Pharm. Sci.* **1967**, *56* (1), 116–130.
7. Holm, J.P., Arming Control for Servo-Adjusted Tablet Compression Machines. US Patent 3,734,663, 1973.
8. Hooker, D.B. Potted Strain Gage. US Patent 3,791,205, 1974.
9. Williams, J.J. Applied Instrumentation Providing Tabletting Compression Force. US Patent 4,016,744, 1977.
10. Williams, J.J. Force Measurement and Analysis Particularly Relating to Rotary Tablet Presses. US Patent 4,030,868, 1977.
11. Stiel, D.M.; Browski, C.R.; Neugroschel, E.; Williams, J.J. Tablet Press Related Instrumentation for Use in Development and Control of Formulations of Pharmaceutical Granulations. US Patent 4,100,598, 1978.
12. Williams, J.J. Force Measurement and Analysis Particularly Relating to Rotary Tablet Presses. US Patent 4,099,239, 1978.
13. Stiel, D.M. Method Invoking Tableting Compression Force Control for Optimizing Tableted Formulation Parameters. US Patent 4,121,289, 1978.
14. Breen, D.T., et al. Tablet Press Controller and Method. US Patent 4,570,229, 1986.
15. Lewis, D.A., Pharmaceutical Tablet Press Control Mechanism. US Patent 4,817,006, 1989.
16. Shotton, E.; Deer, J.J.; Ganderton, D. The Instrumentation of a Rotary Tablet Machine. *J. Pharm. Pharmacol.* **1963**, *5*, 106T–114T.
17. Lammens, R.F.; Liem, T.B.; Polderman, J.; de Blayey, C.J. Evaluation of Force–Displacement Measurements During Powder Compaction. Part 1: Precision and Accuracy of Force Measurements. *Int. J. Pharm.* **1980**, *1* (3), 26–35.
18. Walter, J.T.; Augsburg, L. A Computerized Force–Displacement Instrumentation System for a Rotary Press. *Pharm. Tech.* **1986**, *10* (2), 26–33.
19. Jerzewski, R.L.; Rudnic, E.M. The Development of a Low-Cost Intelligent Tablet Compression Monitor Using a Personal Computer. *Pharm. Tech.* **1986**, September, 32–41.
20. Zeiss, J.C.; Beck, B.; Patel, K. Development of a Monitoring System for an Instrumented Rotary Tablet Press. *Pharm. Tech.* **1986**, August, 60–67.
21. Hunter, B.M.; Fisher, D.G.; Pratt, R.M.; Rowe, R.C. A High Speed Compression Simulator. *J. Pharm. Pharmacol.* **1976**, *28S*, 65P.
22. Pento, T. Procedure for Simulating Tablet Compression. US Patent 5,517,871, 1996.
23. Levin, M.; Tsygan, L.; Dukler, S. Press Simulation Apparatus. US Patent 6,106,262, 2000.
24. Schwartz, J.B. The Instrumented Tablet Press: Uses in Research and Production. *Pharm. Tech.* **1981**, *9*, 102–132.
25. Marshall, K. Instrumentation of Tablet and Capsule Filling Machines. *Pharm. Tech.* **1983**, March, 68–82.
26. Jones, T.M.; Ho, A.Y.K.; Barker, J.F. The Use of Instrumentation in Tablet Research, Development and Production. *Pharm. Tech.* **1985**, March, 42–47.
27. Çelik, M.; Ruegger, C.E. An Overview of Tableting Technology, Part I: Tablet Presses and Instrumentation. *Pharm. Tech. Suppl.—TabGran Yearbook* **1996**, 20–39.



28. Ridgeway Watt, P. *Tablet Machine Instrumentation in Pharmaceuticals: Principles and Practice*; Ellis Horwood Ltd.: Chichester, UK, 1988.
29. Muños-Ruiz, A.; Vromans, H. *Data Acquisition and Measurement Techniques*; Interpharm Press: Buffalo Grove, IL, 1998.
30. Vezin, W.R.; Pang, H.M.; Khan, K.A.; Malkowska, S. The Effect of Precompression in a Rotary Machine on Tablet Strength. *Drug Dev. Ind. Pharm.* **1983**, 9 (8), 1465–1474.
31. Vezin, W.R.; Khan, K.A.; Pang, H.M. Adjustment of Precompression Force to Reduce Mixing-Time Dependence of Tablet Tensile Strength. *J. Pharm. Pharmacol.* **1983**, 35, 555–558.
32. Mitrejev, A.; Augsburger, L.L. Adhesion of Tablets in a Rotary Tablet Press. I. Instrumentation and Preliminary Study of Variables Affecting Adhesion. *Drug Dev. Ind. Pharm.* **1980**, 6 (4), 331–337.
33. Mitrejev, A.; Augsburger, L.L. Adhesion of Tablets in a Rotary Tablet Press. II. Effects of Blending Time, Running Time, and Lubricant Concentration. *Drug Dev. Ind. Pharm.* **1982**, 8 (2), 237–282.
34. Goodhart, F.W.; Mayorga, G.; Ninger, F.C. Measurement of Lower Punch Pulldown Force and Its Significance. *J. Pharm. Sci.* **1969**, 58 (2), 248–251.
35. Hoblitzell, J.R.; Rhodes, C.T. Determination of a Relationship Between Force–Displacement and Force–Time Compression Curves. *Drug Dev. Ind. Pharm.* **1990**, 16, 201–229.
36. Ragnarsson, G.; Sjogren, J. Force–Displacement Measurements in Tableting. *J. Pharm. Pharmacol.* **1984**, 37, 145–150.
37. Von Schmidt, P.C.; Tenter, U. Displacement Measurements of Rotary Presses. *Pharm. Ind.* **1985**, 426–430.
38. Walter, J.T.; Augsburger, L. A Computerized Force–Displacement Instrumentation System for a Rotary Press. *Pharm. Tech.* **1986**, 9, 26–43.
39. Nelson, E. The Physics of Tablet Compression VIII: Some Preliminary Measurements of Die Wall Pressure During Tablet Compression. *J. Am. Pharm. Assoc.* **1955**, 44, 494–497.
40. Long, W.M. Radial Pressures in Powder Compaction. *Powder Met.* **1960**, 6, 73–86.
41. Windheuser, J.J.; Misra, J.; Eriksen, S.P.; Higuchi, T. The Physics of Tablet Compression. 13. Development of Die Wall Pressure During Compression of Various materials. *J. Pharm. Sci.* **1963**, 52, 767–772.
42. Carless, J.E.; Leigh, S. Compression Characteristics of Powders: Radial Die Wall Pressure Transmission and Density Changes. *J. Pharm. Pharmacol.* **1974**, 26, 289–297.
43. Carstensen, J.T.; Toure, P. Correlation Between Hysteresis Loop Areas of Lower Punch and of Die Pressures Versus Upper Punch Pressures. *Drug Dev. Ind. Pharm.* **1981**, 7 (5), 645–648.
44. Schrank-Junghani, H.; Bier, H.P.; Sucker, H. The Measurement of Die Wall Forces to Determine the Minimum Concentration of Lubricant Needed for Tablet Formulations. *Acta Pharm. Technol.* **1984**, 30 (3), 224–234.
45. Cocolas, H.G.; Lordi, N.G. Axial to Radial Pressure Transmission of Tablet Excipient Using a Novel Instrumented Die. *Drug Dev. Ind. Pharm.* **1993**, 19, 2473–2497.
46. Hoag, S.; Renuka, N.; Muller, F. Force–Transducer–Design Optimization for the Measurement of Die-Wall Stress in a Compaction Simulator. *Pharm. Pharmacol. Commun.* **2000**, 6, 293–298.
47. Nurnberg, E.; Hopp, A. Temperature Measurement During Tableting. *Pharm. Technol.* **1981**, 9, 81–101.
48. Travers, D.N.; Merriman, M.P.H. Temperature Changes Occurring During the Compression and the Recompression of Solids. *J. Pharm. Pharmacol.* **1970**, 22S, 11S–16S.
49. Ilkka, J. Instrumentation of Rotary Tablet Machine by a Portable Measuring System. In *Data Acquisition and Measurement Techniques*; Munoz-Ruiz, A., Vromans, H., Eds.; Interpharm Press: Buffalo Grove, IL, 1998.
50. Matz, C.; Bauer-Brandl, A.; Rigassi, T.; Schubert, R.; Becker, D. On the Accuracy of a New Displacement Instrumentation for Rotary Tablet Presses. *Drug Dev. Ind. Pharm.* **1999**, 25 (2), 117–130.
51. Barra, J.; Doelker, E. Instrumentation of an Eccentric Tablet Press. In *Data Acquisition and Measurement Techniques*; Muñoz-Ruiz, A., Vromans, H., Eds.; Interpharm Press: Buffalo Grove, IL, 1998.
52. Hoblitzell, J.R.; Rhodes, C.T. Recent Developments in the Use of Instrumented Tablet Presses. *Pharm. Int.* **1985**, 6, 45–50.
53. Hoblitzell, J.R.; Rhodes, C.T. Preliminary Investigations on the Parity of Tablet Compression Data Obtained from Different Instrumented Tablet Presses. *Drug Dev. Ind. Pharm.* **1986**, 12, 507–525.
54. Marshall, K. Monitoring Punch Forces and Punch Movements as an Aid to Developing Robust Tablet Formulations. *Drug Dev. Ind. Pharm.* **1989**, 15 (13), 2153–2176.
55. Martinez-Pacheco, R.; Vila-Jato, J.L.; Souto, C. Prediction of Tablet Properties by Compression Force–Time Profile Parameters. *Drug Dev. Ind. Pharm.* **1990**, 16 (2), 231–241.
56. Hoblitzell, J.R.; Rhodes, C.T. Instrumented Tablet Press Studies on the Effect of Some Formulation and Processing Variables. *Drug Dev. Ind. Pharm.* **1990**, 16 (3), 469–507.
57. Martinez-Pacheco, R.; Vila-Jato, J.L.; Souto, C. The Effect of Maximum Compression Force and Drug Content on Compression Force–Time Profile Parameters. *Drug Dev. Ind. Pharm.* **1990**, 16 (2), 243–254.
58. Obiorah, B.A. Possible Prediction of Compression Characteristics from Pressure Cycle Plots. *Int. J. Pharm.* **1978**, 1, 249–255.
59. Pitt, K.G.; Newton, J.M.; Stanley, P. Effects of Compaction Variables on Porosity and Material Tensile



- Strength of Convex-Faced Aspirin Tablets. *J. Pharm. Pharmacol.* **1990**, *42*, 219–225.
60. Oates, R.J.; Mitchell, A.G. A New Method of Estimating Volume During Powder Compaction and the Work of Compaction on a Rotary Tablet Press from Measurements of Applied Vertical Force. *J. Pharm. Pharmacol.* **1993**, *46*, 270–275.
61. Oates, R.J.; Mitchell, A.G. Calculation of Punch Displacement and Work of Compaction on a Rotary Tablet Press. *J. Pharm. Pharmacol.* **1989**, *41*, 517–523.
62. Ragnarsson, G.; Sjogren, J. Work of Friction and Net Work During Compaction. *J. Pharm. Pharmacol.* **1983**, *35*, 201–204.
63. Vogel, P.J.; Schmidt, P.C. Force–Time Curves of a Modern Rotary Tablet Machine. II. Influence of Compression Force and Tableting Speed on the Deformation Mechanisms of Pharmaceutical Substances. *Drug Dev. Ind. Pharm.* **1993**, *19*, 1917.
64. Juppo, A.M.; Kervinen, L.; Yliruusi, J. Skewness and Kurtosis of Force–Time Profile Obtained from Compression of Lactose, Glucose and Mannitol Granules. *Eur. J. Pharm. Biopharm.* **1995**, *41* (6), 374–381.
65. Ho, A.Y.K.; Jones, T.M. Punch Travel Beyond Peak Force During Tablet Compression. *J. Pharm. Pharmacol.* **1988**, *40*, 75P.
66. Akande, O.F.; Ford, J.L.; Rowe, P.H.; Rubinstein, M.H.; The Effects of Lag-Time and Dwell-Time on the Compaction Properties of 1:1 Paracetamol/ Microcrystalline Cellulose Tablets Prepared by Pre-compression and Main Compression. *J. Pharm. Pharmacol.* **1998**, *50*, 19–28.
67. Down, G.R.B.; Miller, R.A.; Chopra, S.K. Use of a Desktop Computer in a Search-Type Optimization of Tablet Formulations. *Drug Dev. Ind. Pharm.* **1987**, *38*, 23–31.
68. Dwivedi, S.K.; Oates, R.J.; Mitchell, A.G. Estimation of Elastic Recovery, Work of Decompression and Young Modulus Using a Rotary Tablet Press. *J. Pharm. Pharmacol.* **1992**, *44*, 459–466.
69. Heckel, R.W. An Analysis of Powder Compaction Phenomena. *Trans. Metall. Soc.* **1961**, *221*, 1001–1008.
70. Heckel, R.W. Density–Pressure Relationships in Powder Compaction. *Trans. Metall. Soc.* **1961**, *221*, 671–675.
71. Levin, M., Ed. *Pharmaceutical Process Scale-Up*; Marcel Dekker: New York, 2002.
72. Schwartz, J.B. Scale-Up of the Compaction and Tableting Process. In *Pharmaceutical Process Scale-Up*; Levin, M., Ed.; Marcel Dekker: New York, 2002.
73. Strathy, W.; Gomez, A. Practical Aspects of Tableting Scale-Up. In *Pharmaceutical Process Scale-Up*; Levin, M., Ed.; Marcel Dekker: New York, 2002.
74. Bateman, S.D.; Rubinstein, M.H.; Thacker, H.S. Properties of Paracetamol Tablets Produced Using High Precompression Pressures. *J. Pharm. Pharmacol.* **1989**, *41S*, 32P.
75. Bateman, S.D.; Rubinstein, M.H.; Wright, P. The Effect of Compression Speed on the Properties of Ibuprofen Tablets. *J. Pharm. Pharmacol.* **1987**, *39*, 66P.
76. Cook, G.P.; Summers, M.P. Effect of Compression Speed on Tensile Strength of Tablets of Binary Mixtures Containing Aspirin. *J. Pharm. Pharmacol.* **1990**, *42*, 462–467.
77. Danjo, K.; Kimura, H.; Otsuka, A. Influence of Punch Velocity on the Compressibility of Granules. *Drug Dev. Ind. Pharm.* **1996**, *22* (9&10), 933–942.
78. Fell, J.T.; Newton, J.M. Effect of Particle Size and Speed of Compaction on Density Changes in Tablets of Crystalline and Spray-Dried Lactose. *J. Pharm. Sci.* **1971**, *60*, 1866–1869.
79. Garr, J.S.M.; Rubinstein, M.H. An Investigation into the Capping of Paracetamol at Increasing Speeds of Compression. *Int. J. Pharm.* **1991**, *72*, 117–122.
80. Garr, J.S.M.; Rubinstein, M.H. The Effect of Rate of Force Application on the Properties of Microcrystalline Cellulose and Dibasic Calcium Phosphate Mixtures. *Int. J. Pharm.* **1991**, *73*, 75–80.
81. Holman, L.E.; Leuenberger, H. Effect of Compression Speed on the Relationship Between Normalized Solid Fraction and Mechanical Properties of Compacts. *Int. J. Pharm.* **1989**, *57*, R1–R5.
82. Marshall, P.V.; York, P.; MacLaine, J.Q. An Investigation of the Effect of the Punch Velocity on the Compaction Properties of Ibuprofen. *Powder Technol.* **1993**, *74*, 171–177.
83. Marshall, P.V.; York, P.; Richardson, R. The Effect of Duration of Compression on the Axial Recovery Properties of Compacts of a Crystalline Drug Substance. *J. Pharm. Pharmacol.* **1986**, *38*, 47P.
84. Monedero, M.; Jimenez-Castellanos, M.R.; Velasco, M.V.; Muñoz-Ruiz, A. Effect of Compression Speed and Pressure on the Physical Characteristics of Maltodextrin Tablets. *Drug Dev. Ind. Pharm.* **1998**, *24* (7), 613–621.
85. Rees, J.E.; Rue, P.J. Time-Dependent Deformation of Some Direct Compression Excipients. *J. Pharm. Pharmacol.* **1978**, *30*, 601–607.
86. Ritter, A.; Sucker, H.B. Studies of Variables That Affect Tablet Capping. *Pharm. Tech.* **1980**, March, 57–62.
87. Roberts, R.J.; Rowe, R.C. The Effect of Punch Velocity on the Compaction of a Variety of Materials. *J. Pharm. Pharmacol.* **1985**, *37*, 377–384.
88. Roberts, R.J.; Rowe, R.C. The Effect of Relationship Between Punch Velocity and Particle Size on the Compaction Behaviour of Materials with Varying Deformation Mechanisms. *J. Pharm. Pharmacol.* **1986**, *38*, 567–571.
89. Yliruusi, J.K.; Antikainen, O. New Parameters Derived from Tablet Compression Curves. Part I. Force–Time Curve. *Drug Dev. Ind. Pharm.* **1997**, *23* (1), 69–79.
90. Armstrong, N.A. Time-Dependent Factors Involved in Powder Compression and Tablet Manufacture. *Int. J. Pharm.* **1989**, *49*, 1–13.
91. Armstrong, N.A.; Blundell, L.P. The Effect of Machine Speed on the Compaction of Some Directly Compressible Tablet Diluents. *Pharmacology.* **1985**, *37*, 28P.



92. Armstrong, N.A. Considerations of Compression Speed in Tablet Manufacture. *Pharm. Tech.* **1990**, 9, 106–116.
93. Muñoz-Ruiz, A., et al. Theoretical Estimation of Dwell and Consolidation Times in Rotary Tablet Machines. *Drug Dev. Ind. Pharm.* **1992**, 18 (9), 2011–2028.
94. Akande, O.F.; Ford, J.L.; Rubinstein, M.H. Effect of Dwell Times on the Pre- and Main Compression of a 1:1 Paracetamol: Microcrystalline Cellulose. *J. Pharm. Pharmacol.* **1995**, 47, 1102.
95. Zlokarnik, M. *Dimensional Analysis and Scale-Up in Chemical Engineering*, Springer-Verlag: Berlin, 1991.
96. Levin, M.; Zlokarnik, M. Dimensional Analysis of the Tableting Process. In *Pharmaceutical Process Scale-Up*, Levin, M., Ed.; Marcel Dekker: New York, 2002.
97. Mann, S.C.; Roberts, R.J.; Rowe, R.C.; Hunter, B.M. The Influence of Precompression Pressure on Capping. *J. Pharm. Pharmacol.* **1982**, 34, 49P.
98. Mann, S.C.; Roberts, R.J.; Rowe, R.C.; Hunter, B.M.; Rees, J.E. The Effect of High Speed Compression at Sub-atmospheric Pressure on the Capping Tendency of Pharmaceutical Tablets. *J. Pharm. Pharmacol.* **1983**, 35, 44P.
99. Yang, L.; Venkatesh, G.; Fassihi, R. Characterization of Compactibility and Compressibility of Poly(Ethylene Oxide) Polymers for Modified Release Application by Compaction Simulator. *J. Pharm. Sci.* **1985**, 85, 1085–1090.
100. Bateman, S.D. A Comparative Investigation of Compression Simulators. *Int. J. Pharm.* **1989**, 49, 209–212.
101. Bateman, S.D. High Speed Compression Simulators in Tableting Research. *Pharm. J.* **1988**, 240, 632–633.
102. Bateman, S.D.; Rubinstein, M.H.; Rowe, R.C.; Roberts, R.J.; Drew, P.; Ho, A.Y.K. A Comparative Investigation of Compression Simulators. *Int. J. Pharm.* **1989**, 49, 209–212.
103. Çelik, M.; Marshall, K. Use of a Compaction Simulator System in Tableting Research. Part 1: Introduction to and Initial Experiments with the System. *Drug Dev. Ind. Pharm.* **1989**, 15 (5), 759–800.
104. Çelik, M.; Lordi, N.G. The Pharmaceutical Compaction Research Laboratory and Information Center. *Pharm. Tech.* **1991**, 3, 112–116.
105. Rubinstein, M.H. Compaction Properties of a Cellulose–Lactose Direct-Compression Excipient. *Pharm. Tech.* **1991**, 4, 76–80.
106. Rubinstein, M.H. Applications of Compaction Simulators. *Pharm. Manuf. Int.* **1992**, 177–182.
107. Çelik, M.; Driscoll, C.E. An Overview of the Effects of Some Physico-Chemical and Mechanical Characteristics of Particulates on the Compaction and Post-compaction Properties of Compacts. *Drug Dev. Ind. Pharm.* **1993**, 19, 2119–2141.
108. Çelik, M.; Ong, J.T.H.; Chowhan, Z.T.; Samuel, G.J. Compaction Simulator Studies of a New Drug Substance: Effect of Particle Size and Shape, and Its Binary Mixtures with Microcrystalline Cellulose. *Pharm. Dev. Tech.* **1996**, 1, 119–126.
109. CSUG. Compaction Simulator Users Group Meeting—Standards. <http://www.bath.ac.uk/> (accessed March 23, 1996).
110. CSUG. Compaction Simulator Users Group Meeting—Standards. <http://www.bath.ac.uk/> (accessed July 3, 1996).
111. CSUG. Minutes of the Compaction Simulator Users Group. <http://www.bath.ac.uk/> (accessed October 2, 1996).
112. Asgharnejad, M.; Storey, D. Application of a Compaction Simulator to the Design of a High-Dose Tablet Formulation. Part I. *Drug Dev. Ind. Pharm.* **1996**, 22 (9&10), 967–975.
113. Nokhodchi, A.; Rubinstein, M.H. Compaction Simulators in Tableting Research. *Pharm. Tech. Suppl.—TabGran Yearbook* **1996**, 6, 18.
114. Razaghi, A.M.; Venkatesh, G.; Fassihi, R. Consolidation and Compaction Behavior of Sorbitol in Pharmaceutical Technology: A Compaction Simulator Study. AAPS Meeting, November 1997.
115. Yang, L.; Venkatesh, G.; Fassihi, R. Compaction Simulator Study of a Novel Triple-Layer Tablet Matrix for Industrial Tableting. *Int. J. Pharm.* **1997**, 152, 45–52.
116. Barnett, M.E.; Muller, F.X.; Venkatesh, G.M., Characterization of the Tableting Properties of Four Drug Substances Using a Compaction Simulator and Hiestand Indices of Tableting Performance. AAPS Meeting, November 1998.
117. Kim, H.; Venkatesh, G.; Fassihi, R. Compactibility Characterization of Granular Pectin for Tableting Operation Using a Compaction Simulator. *Int. J. Pharm.* **1998**, 161, 149–159.
118. Carlson, G.T.; Christie, H.R.; Curtiss, A.C.; Hausberger, A.G.; Jarvas, R.E.; Jaxheimer, B.A.; Schelhom, J.J.; Sinko, C.M. Design, Fabrication and Start-Up of a Laboratory Scale Compaction Simulator. AAPS Meeting, November 1998.
119. Rippie, E.G.; Danielson, D.W. Viscoelastic Stress/Strain Behavior of Pharmaceutical Tablets: Analysis During Unloading and Postcompression Periods. *J. Pharm. Sci.* **1981**, 70, 476–482.
120. Ruegger, C.D. An Investigation of the Effect of Compaction Profiles on the Tableting Properties of Pharmaceutical Materials. Ph.D. Thesis, 1996.
121. Oates, R.J.; Mitchell, A.G. Comparison of Calculated and Experimentally Determined Punch Displacement on a Rotary Tablet Press Using Both Manesty and IPT Punches. *J. Pharm. Pharmacol.* **1990**, 42, 388–396.
122. Pedersen, M.A.; Rubinstein, M.H.; Bateman, S.D. Measuring Punch–Time Displacement Profiles on a Rotary Tableting Machine Using a High-Speed Video System. *Pharm. Tech.* **1993**, 4, 44–52.
123. Dwivedi, S.K.; Oates, R.J.; Mitchell, A.G. Peak Offset Times as an Indication of Stress Relaxation During



- Tableting on a Rotary Tablet Press. *J. Pharm. Pharmacol.* **1991**, *43*, 673–678.
124. Muller, F.X.; Augsburger, L.L. The Role of the Displacement–Time Waveform in the Determination of Heckel Behavior Under Dynamic Conditions in a Compaction Simulator and a Fully Instrumented Rotary Tablet Machine. *J. Pharm. Pharmacol.* **1994**, *46*, 468–475.
125. Venkatesh, G., Pudipeddi, M., Faulkner, P., Correlations Between Compaction Simulator and Instrumented Beta-press. AAPS Meeting, November 1993.
126. Venkatesh, G.M.; Lamey, K.A.; Levin, M.; Murphy, S., Correlations Between a Hydraulic Compaction Simulator, Instrumented Manesty Betapress, and the PressterTM. AAPS Meeting, November 1999.



Tablet Testing

Saeed A. Qureshi

*Therapeutic Products Directorate, Health Canada, Tunney's Pasture,
Ottawa, Canada*

INTRODUCTION

A tablet is perhaps the oldest and most common pharmaceutical dosage form. Its popularity is due to its convenience in the administration of a drug without the help or supervision of a health care practitioner, thus providing patients' freedom and a very cost-effective means of providing a reproducible medication. A tablet seldom consists of only the active ingredient. In fact, a tablet represents a mixture of one or more active ingredient(s) with a number of "inactive" ingredients or excipients. There are many reasons for formulating a tablet product with excipients, ranging from management of small dosage amounts of active ingredient to esthetic reasons of color and shape of a product. However, the most fundamental and critical objective of a tablet product is to provide/deliver the active ingredient accurately and reproducibly. Therefore, from this perspective, a tablet is now commonly considered as a drug delivery device. There has also been an increased emphasis in developing tablets that provide, unlike the conventional-release tablets that are of fast disintegrating/release characteristics, controlled disintegration/release process of the active ingredient. These tablet products are therefore known by different names, such as slow-, extended-, controlled-, sustained-, or delayed-release tablets to reflect their drug release characteristics. These modified drug release products provide further convenience to patients by reduced frequency of drug administration, e.g., once a day instead of three doses per day, thus increasing chance of compliance as well. However, for establishing the quality of a tablet product, the fundamentals remain the same, i.e., to ascertain that the product delivers the intended active ingredient in an accurate and reproducible manner. Therefore, tablet testing can be broadly divided into three aspects or categories:

1. Confirmation of the nature of the active ingredient and the product (identity, quantity, impurities, integrity, etc.).

2. Establishing pharmaceutical availability of the active moiety both in vitro and in vivo in humans and, if required, also in animals.
3. Establishing stability profiles to achieve shelf life.

This article provides an overview of the current techniques used for this purpose and discussion of the generally accepted standards. Using conventional-release products as a model product and North American standards/specifications, logistics of conducting testing and the general practices and instrumentation used are described. Relevant references are also provided^[1-4] for further and specific details of testing products and possible differences in national and regional guidelines.

TESTING OF NATURE OF THE TABLET PRODUCTS

In this category, one seeks to establish whether the tablets are within specifications, for example, the nature of the active ingredient (identification), expected amount (assay), purity (related compounds), and uniformity of the amount of drug from tablet to tablet (uniformity of dosage units). Commonly these testing procedures are described in pharmacopeias under a specific name, for example, the names given in parentheses are referred to as the USP (U.S. Pharmacopeia) terminology. In addition to these tests, some other tests such as friability, hardness, disintegration, etc. are also conducted and are described below. Although a number of procedures could be described for individual tests, most emphasis will be given to procedures described in the pharmacopeias because these are usually relatively simpler to conduct and are generally recognized around the world.

Identification

The first and foremost important test in tablet testing is to establish that the tablets contain the labeled active ingredient. For this purpose, usually a fixed number of tablets, e.g., 10–20, are ground and extracted with appropriate solvent extraction. The extract, with or

without a concentration step, is usually chromatographed along with an authentic standard solution. The identity of drug is confirmed based on similarity of ultraviolet (UV) spectrum and/or retention times using chromatographic analysis. This test is generally qualitative in nature. More sophisticated techniques such as chromatographic techniques with or without coupling with mass-spectrometry may be used. However, for routine quality control purposes, the simpler techniques such as thin layer chromatography (TLC) or high-performance liquid chromatography (HPLC) with UV detection are mostly employed.

Assay

One may assume that this test as a quantitative version of the identification test. Again, 10–20 tablets are ground and the active ingredient is dissolved or extracted in a suitable solvent using the described procedure. The concentration of the extracted solution is determined using a specific and validated spectroscopic or chromatographic method against a solution of reference standard. These results are reported as percent of expected/labeled value. Although the specifications for assay results differ from product to product, generally the expected range for individual active ingredient is to be within 90%–110% of the labeled amount.

Uniformity of Dosage Units

This test is conducted to establish consistency in the content of active ingredient from tablet to tablet. There are generally two approaches taken in establishing this: weight variation or content uniformity.

If the active ingredient represents not less than 50% weight of the tablet and greater than 50 mg, then one may establish uniformity of dosage units using the weight variation method. A sample of 10 tablets are weighed individually and results of these weighing are recorded.

In the case of the content uniformity approach, a sample of 10 tablets are individually analyzed using the analytical method described under the assay procedure. It is mandatory to use content uniformity for tablets with less than 50 mg of active ingredient and/or representing less than 50% total mass of the tablets. The content uniformity approach is preferred over the weight variation approach as it more precisely reflects the variation of the active ingredient from tablet to tablet.

The required specification for this test is that uniformity of dosage unit should be within a range of 85%–115% with a relative standard deviation of less than or equal to 6%.

Impurities

The impurities present in a tablet product may fall under different categories, such as: foreign substances that are introduced by contamination or adulteration; toxic impurities that have significant undesirable biological activity, even as a minor component, and require individual identification and quantitation by specific tests; concomitant components that are characteristic of many bulk pharmaceutical preparations, but are not considered as impurities, for example, optical and geometrical isomers; ordinary impurities are those that are innocuous by virtue of having no significant biological activity in the amount present. These impurities may arise out of the synthesis, preparation, and/or degradation of the product.

Generally, most of the impurity profile is established from bulk pharmaceutical material (the raw material). However, from the finished product perspective, products are only analyzed to the extent of toxic impurities. In this case, usually it must be established that product is either free from the specific impurities or the impurities are within acceptable specifications.

Friability

This test is intended to determine, under defined conditions, the friability of uncoated tablets, the phenomenon whereby tablet surfaces are damaged and/or show evidence of lamination or breakage when subjected to mechanical shock or attrition. Commercially available apparatuses known as friabilators are used for the test. Basically, it consists of a drum with diameter between 283 mm and 291 mm and having width of 36 mm–40 mm, made of transparent plastic material (Fig. 1). The drum is attached to the horizontal axis of a device that rotates at 25 ± 1 rpm. The tablets are tumbled at each turn of

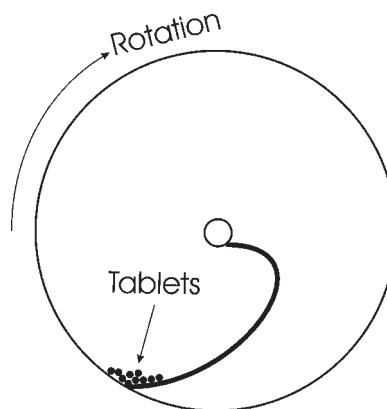


Fig. 1 Schematic of friability testing apparatus.

Tablet Testing

the drum by a curve projection with an inside radius of 75.5 mm–85.5 mm that extends from middle of the drum to outer wall. Thus, at each turn, the tablets roll or slide and fall onto the drum wall or onto each other. Usually, a sample of 10 tablets are tested at a time, unless tablet weight is 0.65 g or less, where 20 tablets are tested. After 100 turns, the tablet samples are evaluated by weighing. If the reduction in the total mass of the tablets is more than 1%, the tablets fail the friability test. Generally, the test is done once. If cracked, cleaved, or broken tablets are obvious, then the sample also fails the test.

Hardness Testing

A tablet requires a certain amount of mechanical strength to withstand the shocks of handling in its manufacturing, packing, shipping, and dispensing. As discussed before, hardness and friability are most common measures used to evaluate tablet strength. The need for testing hardness or crushing strength, in addition to friability, may be explained with an analogy that friability determines how fragile a tablet is. If a tablet is more fragile than expected, then the friability test will detect its substandard quality. However, on the other hand, if the tablets are more robust than desired, a friability test would not detect this deficiency. It is the tablet hardness test that will detect the deficiency.

The most widely used apparatus to measure tablet hardness is the Schleuniger apparatus. This, and other newer electrically operated test equipment, eliminates the operator variability inherent in the measurement using older apparatuses.^[5]

Generally, the force required to break a tablet may be expressed in either kilograms or pounds. It is of critical importance to note that because as results will vary with the specific make and type of the test apparatus used, direct comparison of results obtained on different types of apparatuses cannot be made. Thus, the same instrument must be used consistently throughout a particular study.

Disintegration Test

A disintegration test is a test to establish how fast a tablet disintegrates into aggregates and/or finer particles. The test assumes that if product disintegrates within a short period of time, such as within 5 min, then the drug would be released as expected and one should not anticipate a problem in the quality of a drug product. Although this test is in use for some products in pharmacopeias, its use is generally diminishing in favor of drug dissolution testing, which is described later in this article.

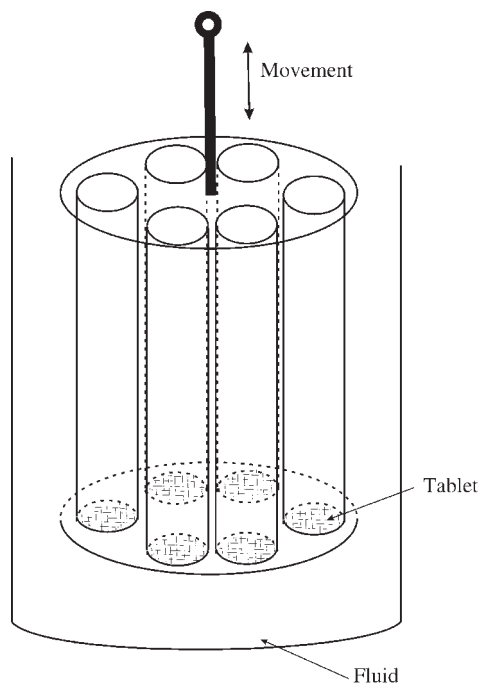


Fig. 2 Schematic of a tablet disintegration apparatus.

When required, the test is conducted using a specially designed instrument known as disintegration apparatus. A schematic representation of the apparatus is shown in Fig. 2. The apparatus employs a basket of six tubes with a base of metal sieve. A tablet is placed in each tube and is held in place by a plastic weight. The six-tube assembly, containing six tablets, is suspended using a hanger with a mechanism of vertical motion at a fixed speed. While hanging the six-tube assembly on the hanger, the assembly is moved in vertical motion in water or a buffer solution. The time for disintegration of each tablet is recorded and should meet the required time specification.

PHARMACEUTICAL AVAILABILITY

Checking the quality of drug products using chemical analyses, as described above, ensures that the product is of acceptable quality in its contents. However, this does not guarantee that the active ingredient in the product will be released in an acceptable manner. There are a number of factors, such as particle size, crystalline form, and compression during manufacturing that, individually or collectively, can severely impact the release of drug from a product thus affecting its efficacy. That is why drug release characteristics must be evaluated and established. The quality assurance around this concept represents one of the major components of product development and later

for quality assurance purposes. The gold standard to establish drug release characteristics of a product is based on an in vivo study, i.e., testing bioavailability (BA) of the drugs in humans.

Furthermore, in reality, changes in product formulation and manufacturing are anticipated and in many cases are unavoidable, for example, if there is a change in ingredient from the supplier, or change in a manufacturing component or facility itself. However, such changes create concern about the impact on the drug release profile and thus the safety and efficacy of the product. The only “gold standard” analytical test is to conduct a BA study to establish the similarity of the new product to the earlier (or the reference) version. However, conducting such BA studies is often one of the most expensive and time-consuming processes in the manufacturing of a product. In addition to the cost and time considerations, ethical concerns also limit the conduct of these studies in humans.

Thus, because of cost, time, and ethical considerations, it is not always possible to conduct drug release studies in humans, and an in vitro drug release evaluation test is the most desirable alternative. For this purpose, a dissolution test for tablet products has been developed and has become a tool for both product development and quality assurance. The test is conducted routinely at every stage of drug product development, during the manufacturing and postmanufacturing stages. A relatively detailed description of the concepts concerning pharmaceutical availability is presented here along with the issues and difficulties of the current approaches so that the reader may be aware of the directions of future development.

In Vivo Drug Release Assessment

To assess release characteristics, one would require to determine BA and/or bioequivalence (BE) characteristics of a product. Both BA and BE areas are subdisciplines of pharmacokinetic studies. Therefore, one would require to have some familiarity with the basic principle of pharmacokinetics. For the purpose of this article, the necessary pharmacokinetic concepts are described below. For further details see Refs. 6 and 7.

Bioavailability Studies

When a drug is absorbed from the gastrointestinal (GI) tract into the blood stream (systemic circulation), it is distributed throughout the body that results in reaching at the site(s) of its action. Concurrent to distribution, drug elimination from the blood due to metabolism or excretion to urine commences. Therefore, at any given time, drug

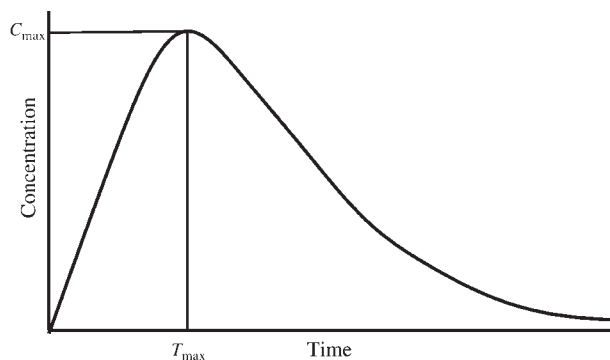


Fig. 3 Plasma drug concentration–time profile.

levels in the blood are the net result of this absorption and elimination processes. In the initial phase, the rate and extent of drug absorption is greater than that for the elimination; therefore, the absorption phase dominates, while at later times when all the drug is absorbed, it is the elimination process that dominates. Although exceptions exist, a typical observed blood drug concentration profile after dosing is shown in Fig. 3. From the analytical chemistry perspective, commonly it is not the blood, but the extracted plasma or serum that is analyzed and is used to reflect drug levels in blood.

The time to reach the maximum plasma drug concentration (C_{\max}) is termed as T_{\max} . The slope reflecting the increase of drug levels in the initial plasma drug profile represents rate of appearance/delivery of drug to the systemic circulation. Depending on the nature of the product, the rate of delivery could be critical. The slope reflecting the decrease of drug levels in the terminal phase of plasma drug concentration represents rate of elimination of drug from the systemic circulation. If one were to draw the curve shown in Fig. 3 on a semilogarithmic scale, i.e., time vs. $\log(\text{plasma drug concentration})$, then the terminal portion of the curve becomes a straight line, as the elimination rate is usually exponential. This rate of elimination is represented by a parameter k_e and is commonly known as elimination rate constant. These parameters, collectively or individually determine the rate of absorption/appearance of drug into systemic circulation.

Another important parameter is the area under the plasma concentration–time curve (AUC). It reflects the extent of drug absorption. From a product quality aspect, knowledge of the described pharmacokinetic parameters is important. Utilizing these parameters, one would establish drug release characteristics of a tablet in vivo.

Calculation of pharmacokinetic parameters

Obviously C_{\max} and T_{\max} are observational and determined directly from experimental results or the curve. Generally,



AUC is determined using the trapezoidal rule, as the arithmetic sum of average plasma concentration at two sampling times divided by the differences in sampling times, i.e., $(C_2 + C_1)/2(t_2 - t_1)$, up to the last measured plasma concentration. Such an area under the curve is termed as AUC_{0-t} . However, as the drug is still in the body and exponentially eliminated after the last determined concentration, the extent of drug remaining is determined by a formula $AUC_{t-\infty} = C_t/k_e$. Thus, the total $AUC_{Total} = AUC_{0-t} + AUC_{t-\infty}$.

To establish drug release characteristics of a product such as a tablet, one has to determine BA of the drug. Bioavailability is defined as a measure, relative to some standard, of the rate and amount of drug that reaches the systemic circulation. As C_{max} , T_{max} , and AUC are reflections of availability of drug in plasma, these parameters are used for determining BA of products. By comparing the derived pharmacokinetics from a BA study, similarity or dissimilarity of drug release in vivo is established.

The concept of BA is very critical from the drug quality perspective. Although, as stated above, chemical and physical tests are critical for testing the quality of tablet products, from the safety and efficacy perspective of drug products, it is the BA studies that are necessary. Even when content and quality of the tablets are within specifications, but release of drug in humans is not as expected, the substandard quality of the products could be dangerous to the health of its users. Therefore, BA studies are conducted at many stages of tablet product development and manufacturing, in particular, product development and receiving permit for marketing. A sample protocol for conducting BA studies for a tablet product testing is described below.^[8]

A comparative BA study is usually conducted in healthy humans, usually involving 12–30 volunteers depending upon the expected variability in the derived pharmacokinetic parameters. To minimize variability in derived parameters, the healthy volunteers should have certain characteristics such as age between 18 and 50 years, within normal weight range, and free from disease or any medication. The study protocol has to be approved by the institutional ethics committee.

Following an overnight fast, a dose is administered to the subjects with a controlled amount of fluid (250 ml of water). At a prescribed time, following the dose, a standardized meal is provided to each volunteer, again to standardize conditions so that results from any of the volunteers would not affect adversely.

Following the dosage administration, blood samples (10–15) are withdrawn from the volunteers. The idea behind selecting the sampling times is that one would be

able to accurately determine the C_{max} . Therefore, generally more frequent sampling times are needed at the lower part of the curve. Similarly, the terminal phase should have a sufficient number of sample times to establish the elimination rate constant. The duration of sampling should be such that most (at least 80%) of the drug, based on the area under the curve, should be accounted for.

The plasma samples are stored for later analysis. The samples are analyzed using validated analytical methods.^[9] The most commonly used methods are chromatographic, i.e., HPLC or gas chromatography. These methods, including the storage conditions, must be validated so that accurate and precise results are assured. Pharmacokinetic parameters from the plasma drug release profiles are determined for individual volunteers. The average values of these parameters reflect the BA of the product.

Bioequivalence Studies

A BE study is not as such an experimental study, but a form of reporting for a comparative BA study. In a BE study, a BA study is conducted to compare the results between reference and test products. A test product may be a product after a manufacturing or an ingredient change or a generic product. In fact, the whole generic industry is based on bringing products to the market based on conducting BE studies between innovators and their products. The study has to be conducted following a statistically valid approach and design, usually a crossover design, e.g., in one period, one-half of the volunteers are given one set of dosage form and in the second period, the order is reversed.

Once the experimental part of a comparative BA study is completed and respective pharmacokinetic parameters are derived and compared, the products are declared bioequivalent when they meet the set and expected specifications for the parameters. The requirements and parameters and their specifications may vary from country to country. However, the most common standard followed is that of the U.S. Food and Drug Administration (FDA). In this case, a 90% confidence interval of the ratios of the log-transformed values of parameters (C_{max} and AUC) should fall within the range 80–125.

Therefore, if one would like to establish impact of any change on pharmaceutical availability from a product or to develop a new formulation or product in comparison with an old product, these BA/BE studies remain the standard and are the requirements for showing the safety and efficacy of the altered/new product. However, from the routine quality control prospective as well establishing quality of a product after some minor changes in formulation or

manufacturing processes, the justification for BA/BE studies on economic and ethical reasons becomes difficult. Therefore, often only in vitro drug release testing is conducted. In the following sections, circumstances for the use of in vitro and in vivo methods are provided. However, a brief description of the current practice of dissolution testing is first given.

In Vitro Drug Release Assessment

Drug dissolution testing is a procedure used to evaluate drug release characteristics of solid oral products such as tablets. As in this article, it is also referred to synonymously as in vitro drug release assessment as opposed to in vivo drug release, which is commonly assessed in humans with studies using blood/plasma concentration profile information from BA or BE studies. The rationale behind conducting dissolution testing is that if a drug is to be absorbed from the GI tract, it usually has to dissolve. Therefore, for a drug to be absorbed, it has to be released from the product and dissolved in the GI fluid. Thus, a dissolution test is an established analytical test to assess the qualities of a drug product, based on its rate and extent of dissolution, i.e., release characteristics.

Commonly, dissolution tests are conducted with apparatuses comprising two major determinants that can be varied: 1) dissolution medium (nature and volume); and 2) a stirring and mixing mechanism.

Choice of dissolution media

As a dissolution test is conducted to simulate drug release in the human GI tract, the generally recommended media are based on aqueous buffers in the pH range of 1–8. Commonly used media are HCl (0.01–0.1 *N*) to simulate gastric fluid and phosphate or acetate buffer in the range of pH 4.0–6.8 to represent intestinal fluids. Recently, suggestions^[10,11] have also been made for other relevant dissolution media reflecting composition, volume, flow rates, and mixing patterns of the fluids in the GI tract. Volumes are usually in the range of 250 ml–1000 ml, however, the use of smaller and larger volumes are becoming common as well. A critical consideration in the choice of a dissolution medium volume is that drug should dissolve freely, i.e., without any concern of reaching saturation in the medium. In the case of low solubility drugs, solubilizing agents, such as sodium lauryl sulfate, are added to enhance the solubility. A low concentration of alcohol may be used to facilitate dissolution for low solubility drugs. However, larger concentrations of alcohol are to be avoided, as this might result in misleading or

unrepresentative results. The dissolution tests are conducted in media kept at a temperature of 37°C.

Stirring and mixing mechanism

The second major determinant for drug dissolution testing concerns the device for stirring or mixing of the product with the dissolution medium. The instrumentation for mixing, in effect the dissolution apparatuses, provides this aspect of the variant. Four different types of apparatuses based on different types of mixing approaches are available commercially and have compendial recognition. These apparatuses are known as: 1) paddle; 2) basket; 3) flow-through; and 4) reciprocating cylinder. For a detailed description and specifications of these apparatuses, readers should refer to any of the pharmacopeias such as the USP.^[12] A brief description of the apparatuses and stirring mechanism is given in the following sections.

Paddle Apparatus. In this case, the agitation is achieved by an inverted T-shape stirrer. The product is placed in the medium that is contained in a glass beaker with round-shaped bottom (Fig. 4). The stirring is achieved by rotating the spindle commonly between 50 rpm and 100 rpm. At specific times, samples are withdrawn and the percent of the drug dissolved is determined using any of the conventional analytical methods such as UV or liquid chromatography.

Basket Apparatus. This method is very similar to the one described above except that the inverted T-shape stirrer is replaced with a cylindrical metallic-wire basket (Fig. 4). The tablet product is placed in the basket that is attached to the spindle, which provides rotation to the basket. The dissolved drug comes out of the basket and mixes with the bulk medium. In this apparatus, it is possible that when the product is disintegrated and

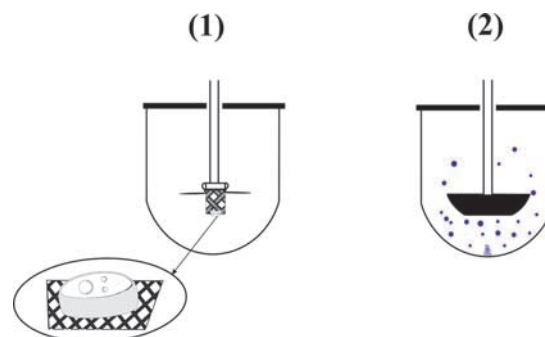


Fig. 4 Schematic of USP drug dissolution apparatuses 1 (basket) and 2 (paddle). (From Qureshi, S.A. Calibration: The USP Dissolution Apparatus Suitability Test. *Drug Inf. J.* **1996**, 30, 1060, Fig. 2.)

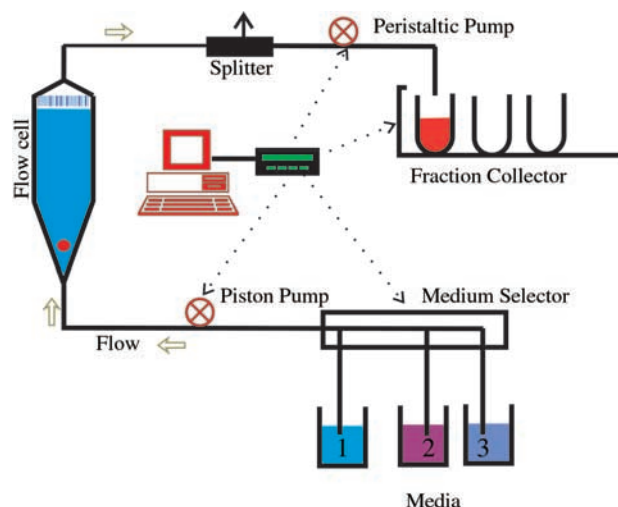
Tablet Testing

Fig. 5 Schematic of a flow-through drug dissolution testing apparatus, also referred to as apparatus 4 in the USP.

particles drop and settle at the bottom of the vessel without dissolving thus providing limited dissolution. Therefore, one may anticipate erratic dissolution results using such an apparatus, at least for some products.

Flow-Through Apparatus. In this type of apparatus (Fig. 5), no stirrer is present. The product is placed in the path of a flowing stream of dissolution medium. The dissolved drug is continuously removed and the exiting solution is measured for the dissolved drug. If required, the exiting liquid may be returned to the medium reservoir for recirculation, thus providing a closed-circuit operation with a limited or finite volume. However, its best use is achieved because of its open-ended operation with a large dissolution medium volume. The system is gaining popularity for products of low solubility drugs. Its main disadvantage compared to the paddle and basket apparatuses is that it is somewhat more challenging to operate.

Reciprocating Cylinder Apparatus. This system may be envisioned as a collection of tubes/vessels in series, as a set of six tubes. Each set of vessels contains a medium of similar or different characteristics such as pH. A hollow cylinder (Fig. 6) with a wire gauze base contains the product and moves in a vertical motion in each vessel. The dissolved drug is mixed with the external medium. At the end of preset time intervals, the vertical moving cylinder containing the remaining product is lifted and moved to the next set of vessels with fresh medium and the process is repeated a number of times. This instrument provides the strongest agitation. Aside from the advantage of the system that drug product can be tested at different pH values, the system appears to have significant

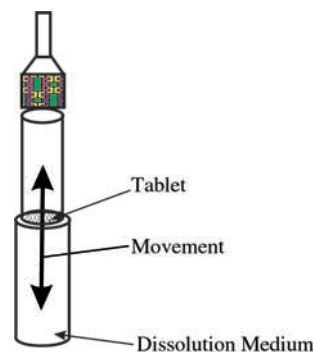


Fig. 6 Schematic of a USP 3 dissolution-testing apparatus, also referred to as reciprocating cylinder apparatus.

disadvantages. It provides limited volume; thus, solubility of drug can become an issue. There could be fluid carryover when the product in the vertical moving tube is transferred from one medium containing tube to the next. In addition, automation of such a system for unattended monitoring is not currently available.

Among the four apparatuses described above, most of the experimental work reported in literature is based on basket and paddle apparatuses.

Reporting Results and Tolerances

Generally, dissolution results are reported as cumulated percent drug release vs. time. Presently, most of the tolerances are based on a single time point, such as not less than 85% dissolved or released in 30 min. However, more reports are appearing with percent drug release values at multiple time points, resulting in a drug release pattern, commonly known as a “profile.”

Commonly, dissolution tests are conducted in a set of six units (tablets). Most commercially available apparatuses can run a test in a set of six units, thus resulting in saving of time and resources and leading to better reproducibility. Although tests are conducted in a set of six, generally results are reported for individual tablets. The most common form of tolerances is of the pharmacopeial ones such as the USP, which has been adopted by other pharmacopeias as well. These compendial standards are described in levels or stages, where each stage has its own set of tolerances that may be based on individual and mean results. For profiles, a similar approach of tolerances is followed while considering every sampling time point as a discrete set of results. It is important to note that there are no requirements and standards of variance around the mean such as standard deviation or coefficient of variation. A lack of such



standards for variation appears to be one challenge (weakness) in current dissolution testing practice.^[13,14]

Recently, a relatively new approach, known as pooled sampling, has also been adopted for reporting the results.^[15] This approach is based on reporting a single result obtained by pooling multiple samples, usually six. The advantages of this approach appear rather arguable. However, one significant disadvantage is that it would hide variability in results from tablet to tablet from a single lot or multiple lots, thus decreasing the quality of information obtained.

With regard to profiles, another parameter that has been introduced to compare drug release is (test vs. reference product) known as similarity factor or f_2 factor.^[16] In essence, it reflects a negative logarithmic sum of differences of percent drug release of two products over multiple sampling times. A value of this parameter in the range of 50–100 is considered to reflect similarity of profiles and thus similarity of products for drug release characteristics. However, there is limited information currently available in support of the usefulness of this approach.

Choice Between In Vivo and In Vitro Studies

Both of these types of studies are conducted to assess the drug release characteristics of tablet products. The gold standard remains in vivo studies. However, as stated earlier, it is not necessary, due to ethical and economical reasons, that one has to do the in vivo studies all the time. There are situations wherein the safety and efficacy aspect can be addressed by conducting only the in vitro studies.^[17,18] The confidence in the in vitro results as a reflection of in vivo results is obtained if an established in vivo–in vitro correlation (IVIVC) is available, i.e., in vivo drug release obtained from plasma drug concentration–time profiles reflects the observed drug dissolution characteristics. There are numerous ways of observing and reporting IVIVC.^[19–23] However, for the purpose of reliability, one generally compares the observed in vitro cumulative drug release results with the cumulative drug release in vivo derived from the pharmacokinetic data. Although there is significant enthusiasm in developing and reporting IVIVC in the scientific community, examples of successful IVIVC are rare for predicting in vivo characteristics based on in vitro results and vice versa. On the other hand, examples of lack of IVIVC may be found more frequently. For example, in a study^[24] to establish IVIVC, drug release characteristics of various primidone tablet products were assessed both in vitro (dissolution tests) and in vivo (BA/BE study). The results of the evaluation

indicate that even though there were large differences in the in vitro dissolution of the primidone tablets, the in vivo results were shown to be very similar.

In addition, one should also not expect an IVIVC for all drug products. The drug absorption into systemic circulation depends upon two aspects: 1) drug release in the GI tract and 2) permeability characteristics of the drug through GI tissue. If the absorption of a drug is permeability dependent, then one should not anticipate IVIVC because the IVIVC aspect only relates in vitro release to in vivo absorption or drug appearance in the body.

Based on the drug solubility and permeability characteristics, drugs may be divided into four groups: 1) high solubility–high permeability drugs; 2) low solubility–high permeability drugs; 3) high solubility–low permeability drugs; and 4) low solubility–low permeability drugs. This classification of drug into four groups is known as Biopharmaceutic Classifications System.^[25] Based on this classification system, one would anticipate a successful IVIVC with drugs in low solubility–high permeability group. In this type of drugs, as permeability is high thus as soon as drug is dissolved, one would anticipate corresponding absorption through the GI tissue. For the two categories that have drugs with low permeability values, because the drug absorption is permeability dependent, the IVIVC is unlikely to be observed. Furthermore, it is also unlikely that one would see IVIVC for drugs in the category 1, i.e., high solubility and high permeability category. The reason being that if a drug is fast dissolving, e.g., in less than 15 min, then one would anticipate saturation or overloading of the GI tract absorption system. Thus, changes in vivo dissolution would unlikely be reflected in plasma drug concentrations. However, if a high solubility–high permeability drug is formulated as a slow-release product, then the drug would appear as low solubility/dissolution drug and would behave as low-solubility–high permeability drug where IVIVC should be possible. Therefore, IVIVC is not only a drug-dependent characteristic; it is also a product-dependent characteristic.

Another reason for the lack of success of IVIVC may be due to the lack of relevant testing environment in vitro. The currently used methodologies not only offer very poor reproducibility of results, but also very poor mixing and stirring characteristics in vitro.^[14] Therefore, it is highly unlikely that general success of IVIVC could be achieved soon. There is a clear need for some improvement in this area.

Notwithstanding the above discussion, recently, there has been recognition of accepting in vitro results in lieu of in vivo results for the assessment of safety and efficacy of pharmaceutical products. Particularly, there has been

Tablet Testing

a guidance document^[26] from U.S. FDA that allows in vitro studies in lieu of in vivo studies. In this particular example, if a drug has high permeability and high solubility characteristics and no history of problem regarding the drug absorption, inference to safety and efficacy of the product after changes in product manufacturing can be granted provided if the products show equivalent in vitro drug release characteristics. The acceptance of in vitro results for in vivo is not based on IVIVC for this particular group of products, i.e., products with high solubility and permeability drugs and very fast in vitro release characteristics (more than 85% dissolved in 15 min), but because experience with such drugs is that these have no BA problems and are unlikely to cause a potential health risk.

In short, it can be said that BA/BE testing remains the gold standard for establishing the pharmaceutical ability of products in vivo. Due to ethical and economical reasons, BA/BE availability studies are mostly conducted at product development and major formulation and manufacturing process changes. However, an in vitro test, in conjunction with other analytical tests, is performed whether its relevance to in vivo is established or not that appears to provide adequate assurance of quality of products in particular anticipated release characteristics in vivo. Thus, drug dissolution testing has become an important and essential quality control test for establishing lot-to-lot consistency in the evaluation of products. Furthermore, in some cases such as for highly soluble and permeable drugs in fast release product characteristics, a dissolution test can also be used for a substitute of a BA study.

PRODUCT STABILITY EVALUATION

Product stability testing is an essential component of establishing the quality of the products. The stability parameters of a drug dosage form can be influenced by environmental conditions of storage (e.g., temperature, light, air, and humidity) as well as by the packaging styles.

Knowledge of the physical stability of a formulation is very important for two reasons. First, a pharmaceutical product must look fresh and elegant as long as it remains on the shelf. Second, the active ingredient must be available to the patient throughout the expected shelf life of the product. The causes of deterioration in quality of active ingredient could be due to incompatibility with excipients, oxidation, reduction, hydrolysis, and racemization to name a few.

With the exception of BA studies that are not conducted for the stability studies, all the above-mentioned tests are

Table 1 An example of commonly used protocol for stability testing of tablet products

Type	Conditions	Duration (mo)
Long-term testing	$25 \pm 2^\circ\text{C}/60\% \text{ RH} \pm 5\%$	12
Accelerated testing	$40 \pm 2^\circ\text{C}/75\% \text{ RH} \pm 6\%$	6
Alternate testing	$30 \pm 2^\circ\text{C}/60\% \text{ RH} \pm 5\%$	

Alternate testing is required if significant changes occur during 6 mo storage under conditions of accelerated testing.

performed after storing the product in the package material to declare that products of acceptable quality after the manufacturing. In vitro dissolution tests are relied on for the drug release characteristic in vivo. The analytical methods employed for stability testing must be validated to show that the accuracy and precision have not been effected by the interferences of any potential degradation product(s).

The stability studies on the active substances and packaged dosage forms are conducted by “real time.” Long-term tests at specific temperatures and humidity representing storage conditions experienced in the distribution chain of the climatic zone(s) of the country or region of the world concerned.

There are generally two types of stability studies conducted: 1) regular and 2) accelerated. In the case of accelerated stability study, the product is stored under elevated temperature and relative humidity (RH) conditions to force or expedite its degradation pathway and assess the changes. The stability testing conditions depend on the anticipated market environment. An example of a recommended protocol^[27] is provided in Table 1.

REFERENCES

1. *USP XXIV*; The United States Pharmacopoeial Convention, Inc.: Rockville, MD, 2000; 1773–1774.
2. *British Pharmacopoeia (BP) 2001*; British Pharmacopoeia Commission: London, UK, 2001; 1810.
3. *European Pharmacopoeia*, 4th Ed.; Council of Europe: Strasbourg, France, 2002; 561.
4. Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances. International Conference on Harmonization (ICH), Geneva, Switzerland, October 1999.
5. Gennaro, A.R. (Editor-in-Chief), *Remington: The Science and Practice of Pharmacy*, Lippincott Williams & Wilkins: PA/MD, 2000; 883.
6. Gibaldi, M. *Biopharmaceutic and Clinical Pharmacokinetics*, 4th Ed.; Lea/Febriger: Philadelphia/London, 1991.



7. Abdou, H.M. *Dissolution, Bioavailability and Bioequivalence*, Mack Publishing Company: Easton, PA, 1989.
8. United States Food and Drug Administration Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Consideration, October 2000.
9. United States Food and Drug Administration Guidance for Industry: Bioanalytical Method Validation, May 2001.
10. Dressman, J.B.; Reppas, C. In Vitro–In Vivo Correlations for Lipophilic, Poorly Water-Soluble Drugs. *Eur. J. Pharm. Sci.* **2000**, *11*, (S-73)–(S-80).
11. Nichlaides, E.; Galia, E.; Efthymiopoulos, C.; Dressman, J.B.; Reppas, C. Forecasting the In Vivo Performance of Four Low Solubility Drugs from Their In Vitro Dissolution Data. *Pharm. Res.* **1999**, *16*, 1876–1882.
12. *USP XXIV*; The United States Pharmacopoeial Convention, Inc.: Rockville, MD, 2000; 1941–1951.
13. Qureshi, S.A.; McGilveray, I.J. Typical Variability in Drug Dissolution Testing: Study with USP and FDA Calibrator Tablets and a Marketed Drug (Glibenclamide) Product. *Eur. J. Pharm. Sci.* **1999**, *7*, 249–258.
14. Qureshi, S.A.; Shabnam, J. Cause of High Variability in Drug Dissolution Testing and Its Impact on Setting Tolerance. *Eur. J. Pharm. Sci.* **2001**, *12*, 271–276.
15. *USP XXIV*; The United States Pharmacopoeial Convention, Inc.: Rockville, MD, 2000; 1943.
16. Shah, V.P.; Tsong, Y.; Sathe, P.; Liu, J.-p. In Vitro Dissolution Profile Comparison, Statistics and Analysis of the Similarity Factor, f_2 . *Pharm. Res.* **1998**, *15*, 889–896.
17. Augsburger, L.; Shangraw, R.; Lesko, L.J.; Williams, R. An Approach Towards Establishing a Scientific Foundation for Interpreting Regulations and Workshop Reports of Scale-Up and Postapproval Changes. *Pharm. Res.* **1994**, *11*, S-143.
18. Eddington, N.D.; Ashraf, M.; Augsburger, L.L.; Lesli, J.L.; Fossler, M.J.; Lesko, L.J.; Shah, V.P.; Rekhi, G.S. Identification of Formulation and Manufacturing Variables That Influence In Vitro Dissolution and In Vivo Bioavailability of Propranolol Hydrochloride Tablets. *Pharm. Dev. Tech.* **1998**, *3*, 535–547.
19. Gillespie, W.R. Convolution-Based Approaches for In Vivo–In Vitro Correlation Modeling. *Adv. Exp. Med. Biol.* **1997**, *423*, 53–65.
20. Polli, J.E. Dependence of In Vitro–In Vivo Correlation Analysis Acceptability on Model Selections. *Pharm. Dev. Tech.* **1999**, *4*, 89–96.
21. Dowell, J.A.; Hussain, A.; Devane, J.; Young, D. Artificial Neural Network Applied to the In Vitro–In Vivo Correlation of an Extended-Release Formulation: Initial Trials and Experience. *J. Pharm. Sci.* **1999**, *88*, 154–160.
22. Eddington, N.D.; Marroum, P.; Uppoor, R.; Hussain, A.; Augsburger, L. Development and Internal Validation of an In Vitro–In Vivo Correlation for a Hydrophilic Metoprolol Tartrate Extended Release Tablet Formulation. *Pharm. Res.* **1998**, *5*, 466–473.
23. Langenbucher, F.; IVIVC: Indices for Comparing Release and Response Profiles. *Drug. Dev. Ind. Pharm.* **1999**, *25*, 1223–1225.
24. Meyer, M.C.; Straughn, A.B.; Mhatre, R.M.; Shah, V.P.; Williams, R.L.; Lesko, L.J. Lack of In Vitro/In Vivo Correlations of 50mg and 250mg Primidone Tablets. *Pharm. Res.* **1998**, *15*, 1085–1089.
25. Amidon, G.L.; Lennernas, H.; Shah, V.P.; Crison, J.R. A Theoretical Basis for the Biopharmaceutic Drug Classification: The Correlation of In Vitro Drug Product Dissolution and In Vivo Bioavailability. *Pharm. Res.* **1995**, *12*, 413–420.
26. FDA Guidance for Industry: Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System, August 2000.
27. Harmonized Tripartite Guidelines: Stability Testing of New Drug Substances and Products. International Conference on Harmonization (ICH), Geneva, Switzerland, November 2000.

TARGETED DRUG DELIVERY: MONOCLONAL ANTIBODIES AS DRUG TARGETING AGENTS

Ban-An Khaw

Northeastern University, Boston, Massachusetts, U.S.A.

INTRODUCTION

The concept of targeted drug delivery may be considered to have originated with the proposed “magic bullet” of Paul Ehrlich for specific eradication of the spirochete of syphilis at the beginning of the 20th century (1). Yet this vision of targeted drug delivery did not begin until Pressman and Keighley in 1948 (2) demonstrated that radiolabeled antibodies could be specifically targeted to rat kidneys. Thus, antibodies became specific pharmaceuticals for targeting radioisotopes to various pathological disorders for diagnosis and therapy including various tumor. The next major thrust in the advancement of targeted drug delivery, using antibodies was provided by the revolutionary publication of Kohler and Milstein on monoclonal antibody technology (3). Monoclonal antibodies have since dominated the “targeting” in the targeted drug delivery systems.

The term targeted drug delivery conjures up a vision of a) a targeting moiety, and b) the drug to be targeted. The targeting moiety can be receptors, ligands, oligonucleotides, hormones, or antibodies. The drug to be targeted could be radioisotopes, pharmaceuticals, immunochemicals, chemicals, toxins, or biologicals. Thus, the concept of targeted drug delivery becomes almost all encompassing. Yet in practice, targeted drug delivery is somewhat restricted. Literature search of the term “targeted drug delivery” invariably resulted in the reference of articles relating to antibody mediated liposomes delivery system. Yet it should also include antibody, bispecific antibody, or avidin–biotin antibody delivery system. In oncology, antibodies may provide the primary targeting device for targeted drug delivery. Similarly antibodies form the primary targeting reagents in the cardiovascular applications. This chapter will restrict targeted drug delivery to targeting with antibodies and stress their applications in the cardiovascular arena, with cursory introductory review of the most pertinent targeting aspects in oncology.

ANTIBODIES IN CANCER IMAGING

Pressman and Keighley were the first to demonstrate that radioiodinated antibodies to normal rat kidneys targeted specifically to the kidneys in vivo (2). The potential for targeted delivery of radioisotopes for tumor therapy and diagnosis was immediately recognized by many investigators. However, progress for the next decade or so was disappointing due to lack of specific targets associated with tumors, pure antibodies for in vivo trials, and appropriate radiolabels for tagging the available antibodies. Never the less, Mach et al. (4), and Goldenberg et al. (5) succeeded in using polyclonal antibodies to carcinoembryonic antigen (CEA) to image carcinomas. Subsequently these same investigators used monoclonal antibodies to show that carcinomas can be imaged with radiolabeled anti-CEA antibodies (6, 7).

Since those early days of tumor imaging with immunoglobulin fractions, or affinity purified antibody fractions radiolabeled with I-131, monoclonal antibodies, and new and improved radiolabeling methods have enabled investigators to image breast (8), colon (9), lung (10), ovarian (11), and prostate cancers (12). Melanomas (13), T-cell lymphomas (14), pancreatic cancer (15), and indeed every form of cancer have been targeted with polyclonal or monoclonal antibodies in either experimental or clinical trials. Improvements in labeling with radioisotopes, such as Tc-99m, In-111, and I-123 to antibodies, have also led to better and more efficient gamma imaging. Further improvements in radioimmuno-cintigraphy include use of bispecific monoclonal antibodies (16) and multistep avidin–biotin conjugated antibodies to reduce nontarget organ activities and enhanced target to background ratios (17).

In the therapeutic arena, antibodies have also been used for immunotherapy and radioimmunotherapy. Immunotherapy itself is essentially targeted drug therapy where specific antibody may activate complement, attract inflammatory cells, and/or induce production of cytokines, or as in the case of Herceptin inhibit cell growth by

binding to ErbB2 that are overexpressed in metastatic breast cancer (18). Similarly, Rituxan is an antibody that binds to overexpressed CD20 in non-Hodgkins lymphoma (19) inducing apoptosis in these cancer cells (20). On the other hand, radioimmunotherapy may under certain oncologic conditions provide greater benefit. Use of murine–human chimeric monoclonal antibodies for targeting of ovarian cancer with β emitters for therapy has also been encouraging (21).

Immunotoxins consisting of tumor associated antigen specific antibody linked to the toxic chains of ricin, abrin, or other toxins have also been tested. Although phenomenal results in cancer therapy were anticipated, the progress has been disappointing. The only unequivocal success has been the *in vitro* treatment of bone marrow cells to rid them of malignant cells for subsequent injection into patients with hematological malignant cancer after whole body irradiation (22). Targeting of cancer cells with immunoconjugates of target specific antibody covalently linked with drugs, such as doxorubicin, have also been undertaken (23). Other methods for targeted drug delivery include use of immunoliposomes for delivery to specific tumors (24).

Despite these novel methods for targeted drug delivery directed with monoclonal antibodies, monoclonal antibodies have not totally met the expectations of the “magic bullet” of Paul Ehrlich in oncology. Therefore, apart from diagnostic imaging with radiolabeled monoclonal antibodies and direct immunotherapy, other innovative targeted drug delivery systems have had only limited clinical success.

TARGETED DRUG DELIVERY IN THE CARDIOVASCULAR SYSTEM

As indicated above, antibodies can provide target specificity as well as function as carriers of those drugs for targeted delivery. Whether the drug to be delivered is a radioisotope for imaging or therapy, or an antimetabolite, the prototypical model exemplified will be the radiolabeled antibodies for targeted drug delivery.

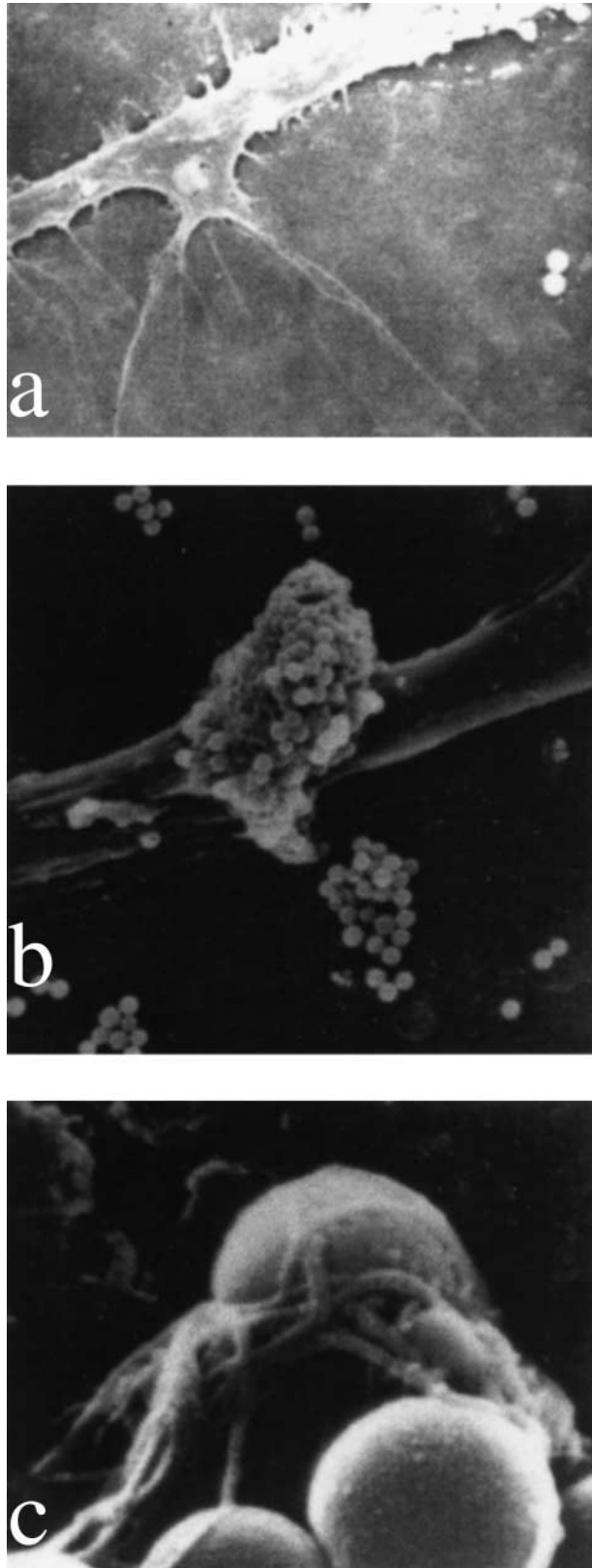
The application of antibodies in cardiovascular targeting *in vivo* originated with the experimental demonstration of the feasibility of using radiolabeled antimyosin antibody for diagnosis of acute myocardial infarction in 1976 (25). Since then, the use of antibodies in the cardiovascular system has encompassed imaging of myocarditis (26), heart transplant rejection (27), dilated cardiomyopathy (28), alcohol induced cardiomyopathy (29), adriamycin cardiotoxicity (30), various other cardiomyopathies (31), vascular clots (32), atherosclerotic lesions (33), and even

certain cancers such as soft tissue sarcomas (34). Yet the best characterized and studied antibody for cardiovascular diagnostic targeting is monoclonal antimyosin Fab for its exquisite specificity in the detection of myocardial cell death associated with various forms of cardiomyopathies.

Targeted Delivery of Radiolabeled Antimyosin Antibody in Acute Myocardial Infarction

To enable targeted delivery of any drug, there must be target specific structures not present on normal nontarget organs or cells that can be recognized by the targeting reagents, or neostructures must be exposed during the development of the pathological conditions. Since the cell membrane of normal cardiac cells are composed of the same lipid bilayer as that found in necrotic cells, one cannot envision a target specific cell membrane structure that is present only on the membranes of necrotic cells. However, it is reasonable to assume that structures such as the cardiac myofilaments are not exposed to the extracellular environment in viable cardiocyte, but following myocardial necrosis they would become exposed to the extracellular milieu (35). Therefore, such structures should provide new targets for delineation of the necrotic from nonnecrotic myocardium. Anticardiac myosin antibody was chosen as the model targeting moiety for the delivery of the model drug for diagnostic imaging of acute myocardial infarction (25). The above hypothesis was validated using hypoxic neonatal murine myocytes in primary cultures treated with antimyosin antibody attached covalently to 1 μ m diameter polystyrene beads, where the beads represented potential drugs for targeting (35). Normal myocytes with intact sarcolemma prevented accumulation of antimyosin beads (Fig. 1a) whereas hypoxic myocytes with sarcolemmal lesions were targeted by antimyosin beads at the lesions sites (Fig. 1b). Upon higher magnification ($\times 100,000$), targeting of the beads can be discerned as the myofilaments are seen wrapped around the antimyosin-beads (Fig. 1c). Such specific targeting of the necrotic myocardium can also be demonstrated *in vivo* in experimental and clinical situations by gamma imaging technology.

To establish without equivocation that antimyosin Fab can target the necrotic myocardium with highest specificity, a mixture antimyosin antibody labeled with I-125 and normal IgG labeled with I-131 was administered by intracoronary delivery into dogs with experimental acute myocardial infarction (36). Targeting of both radiolabeled immunoglobulin species was assessed by gamma scintillation counting (36). I-125 labeled antimyosin antibody localized in the necrotic myocardium with a target to nontarget ratio of $\approx 32:1$ at the infarct center, whereas



nonspecific localization was only about 6.5:1 in the same tissue samples (36). Similarly, in vivo demonstration of the specificity of antimitosin antibody for the necrotic myocardium was performed with antimitosin Fab (Fig. 2A) and nonspecific Fab radiolabeled with In-111 (Fig. 2B). Dogs with experimental acute myocardial infarction injected with In-111 labeled antimitosin Fab showed in vivo localization of the radiolabel conjugated to antimitosin Fab in the infarct. The nonspecific Fab control also radiolabeled with In-111 did not localize in the infarcted region of the myocardium in control infarcted dogs (37).

To further demonstrate the exquisite specificity of antimitosin for targeting acute myocardial infarction, two monoclonal antimitosin antibodies, one with high and the other with low affinity, were employed to visualize canine acute myocardial infarction. The low affinity antimitosin Fab (3H31E6) despite being specific for myosin, with an apparent affinity of $\approx 6.5 \times 10^6$ L/mole, showed no in vivo targeting of the radiolabeled antibody even at 5 h after antibody administration (Fig. 2D). On-the-other-hand, with the higher apparent affinity antimitosin Fab (R11D10) (1×10^9 L/mole), the infarct may be visualized as early as 1 hour but unequivocally some time later after intravenous administration (Fig. 2C) (37). The ratios of target (*T*) to blood pool (*B*) activities of the in vivo gamma images were determined by computer planimetry to be 1.7 ± 0.38 (mean \pm SD) for R11D10 Fab, which was significantly greater than that of 3H31E6 Fab (0.85 ± 0.12). The latter was the same as the *T/B* ratios obtained with nonspecific Fab (0.75–0.771) (37). Another monoclonal antimitosin Fab (2G42D7) with an apparent affinity similar to R11D10 also provided *T/B* ratio of 1.5 ± 0.27 ($p = 0.13$). These studies confirmed that not only is antimitosin specific for targeting acute myocardial necrosis, the requirement of sufficiently high affinity for successful in vivo visualization makes this targeting phenomenon exquisitely target specific.

Similarly, gamma images of patients with persistently occluded left anterior descending coronary artery (LAD) and another with reperfused LAD injected with In-111 labeled monoclonal antimitosin R11D10 Fab (38) showed

Fig. 1 Scanning electron micrographs of murine neonatal primary myocytes in culture treated with antimitosin linked 1 μ diameter fluorescent polystyrene beads. a) A normal myocyte showing the intact cell membrane and a lack of antimitosin-bead binding. b) A necrotic myocyte with a region of sarcolemmal disruption showing antimitosin-bead binding in that region. c) 100,000 times magnification of the region of cell membrane disruption showing antimitosin-beads wrapped around by the myofilaments that contain myosin.

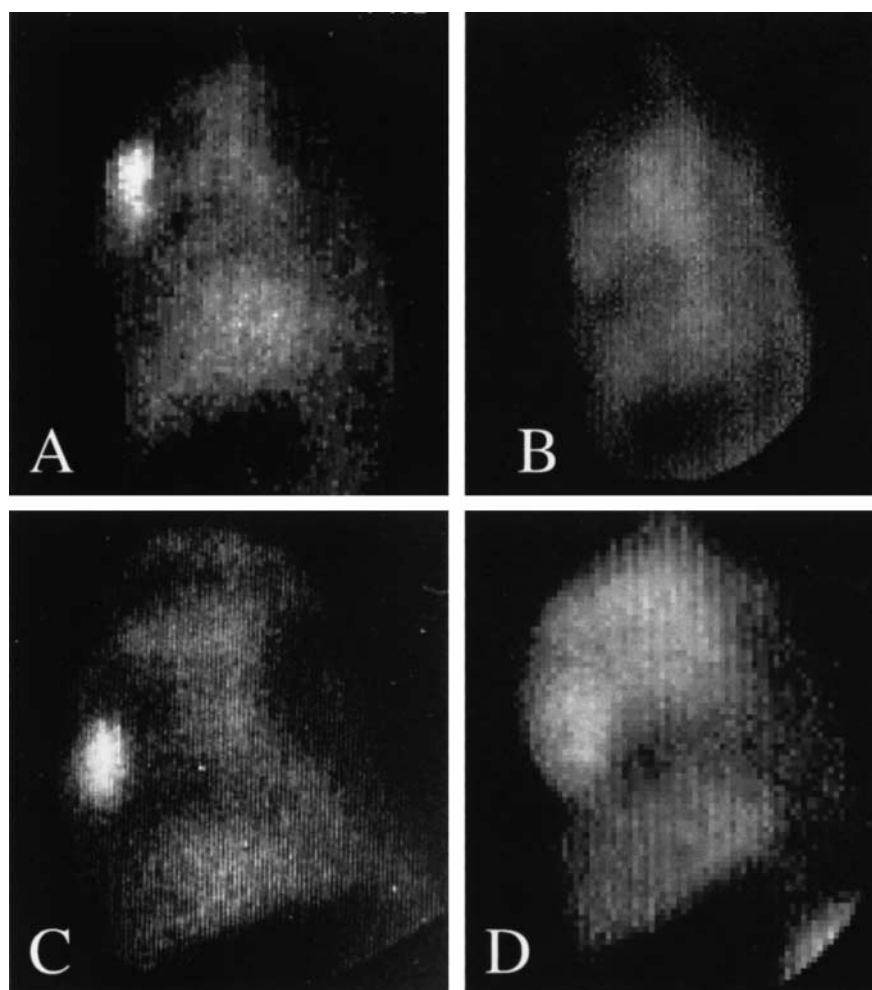


Fig. 2 Gamma images of four dogs with acute experimental myocardial infarction, injected with high affinity ^{111}In -labeled antimyosin Fab ($K_a = 0.5\text{--}1 \times 10^9 \text{ L/M}$) (A and C), ^{111}In -labeled nonspecific monoclonal Fab (B) and the low affinity ^{111}In labeled 3H3 antimyosin Fab ($K_a = 5 \times 10^6 \text{ L/M}$) (D). At 5 h post intravenous administration of these antibodies, only the images with high affinity antimyosin showed unequivocal infarct uptake, whereas the images with nonantimyosin specific monoclonal Fab and the low affinity antimyosin Fab showed only blood pool activity. (From Ref. 37.)

unequivocal targeting of the radiopharmaceutical in the region of the hearts corresponding to the areas subtended by the occluded coronary vessels (Fig. 3). In patients with no infarction, no targeting of radioactivity in the myocardial region was obtained.

Despite the ability of antimyosin Fab to specifically delineate acute myocardial infarction, its utility is hampered by the slow blood clearance of the Fab fragments (38). This causes delayed development of high enough target to background ratios for in vivo visualization by gamma scintigraphy, especially in clinical use to $>12 \text{ h}$ after intravenous administration. Usually, images were acquired 18–24 h after IV administration, however, small MI may require 48 h of blood clearance to enable

unequivocal diagnosis. If on the other hand, qualitative diagnostic end-point is the desired outcome, irrespective of visual confirmation of the infarct size, then infarcts may be detected rather early after intravenous antibody administration over and above the blood pool activity (39). These rate-limiting steps may be overcome by using smaller antibody fragments, such as sFv (40), CDR (41), and mimetics (42). However, by increasing the clearance rate, there is a concomitant decrease in the absolute available antibody for target localization. Therefore, we developed a new approach to improve target to background ratios in experimental myocardial infarct visualization by decreasing the background activity without affecting the target activity at any time point.

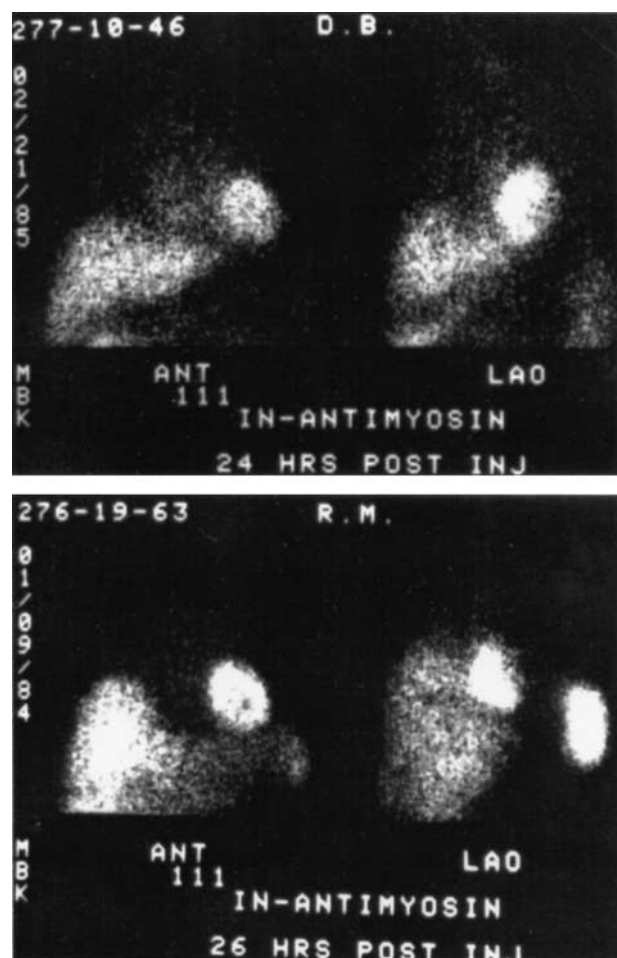


Fig. 3 Anterior and 45° LAO images of two patient with acute MI at 24 and 26 h after intravenous administration of In-111 labeled antimyosin Fab. Images of a patient with persistently occluded LAD (no reperfusion) (top) and those of a patient with successful reperfusion (bottom). (From Ref. 38.)

ENHANCING IN VIVO TARGETED DRUG DELIVERY WITH NEGATIVE CHARGE-MODIFIED MONOCLONAL ANTIBODIES

Since antibodies are basic glycoproteins, under physiological conditions they are positively charged (43). Cells and extracellular matrices, on the other hand are negatively charged due to the presence of cell membrane bound acidic residues such as sialic acids (44,45) and heparan sulfate proteoglycans (46), respectively. Therefore, the potential for nonspecific ionic interaction between the positively charged molecules and the negatively charged cell surfaces or extracellular matrices exists. Such nonspecific interaction has been utilized to

deliver methotrexate–polylysine conjugates to malignant cells (47), protracted release of basic fibroblast growth factor for the salvage of the infarcted myocardium (46), as well as for the delivery of genetic constructs by lipofection or cationic liposomes (48). Alternatively, we proposed that if the basic (positively charged) antibodies were to be modified to carry a highly anionic polymer so that the isoelectric point of the modified antibody becomes low (e.g., $PI < 5$), then this modified antibody should have reduced nonspecific ionic interaction with nontarget cells and extracellular matrices (49). However, the affinity of the antibody at $\approx 1 \times 10^9$ L/mole can easily overcome the repulsive ionic forces when the charge-modified antibody approaches its homologous antigen. This would permit the same targeting of the charge-modified antibody as with the noncharge modified antibody due to the same antigenic specificity and affinity, but should result in lower nontarget background activity and thereby allow earlier development of high enough target to background ratios for visualization of the target (49). Furthermore, since negatively charged diethylene triamine pentaacetic acid chelate modified polymers are used, very high specific radioactivity radiolabeled antibody preparations can be prepared.

Antimyosin Fab was covalently modified with multi-DTPA conjugated polylysine that has been made totally anionic by succinylation (50). The immunoreactivity of the negatively-charge-modified AM-Fab carrying a polymer of approximate 3.3 or 17 kD was identical to that of the unmodified AM-Fab (50). When these negative charge-modified AM-Fab preparations were labeled with In-111, a specific radioactivity of 50–100 mCi/mg AM-Fab was obtained. The specific radioactivity of the conventional AM-Fab was only 2–10 mCi/mg AM-Fab. Thus, less negative charge-modified AM-Fab was needed to deliver the same radioactivity relative to the conventional dose. When the negative charge-modified AM-Fab was administered into dogs with reperfused experimental MI, myocardial infarcts were visualized within 30 min of intravenous administration of the antibody preparation (Fig. 4), whereas the conventional In-111 labeled AM-Fab required 1–2 h of antibody circulation and clearance before infarcts were visualized.

This earlier target visualization could be, as stated previously, due to quicker development of high target to background ratios for visualization by gamma imaging, which requires a ratio of about 10:1. It could also be due to a new concept developed in our laboratory where visualization could be achieved not based on target to background ratios, but on the absolute difference in the radioactivity between target and background (51). Assuming that the same amount (500 μ g) of charged

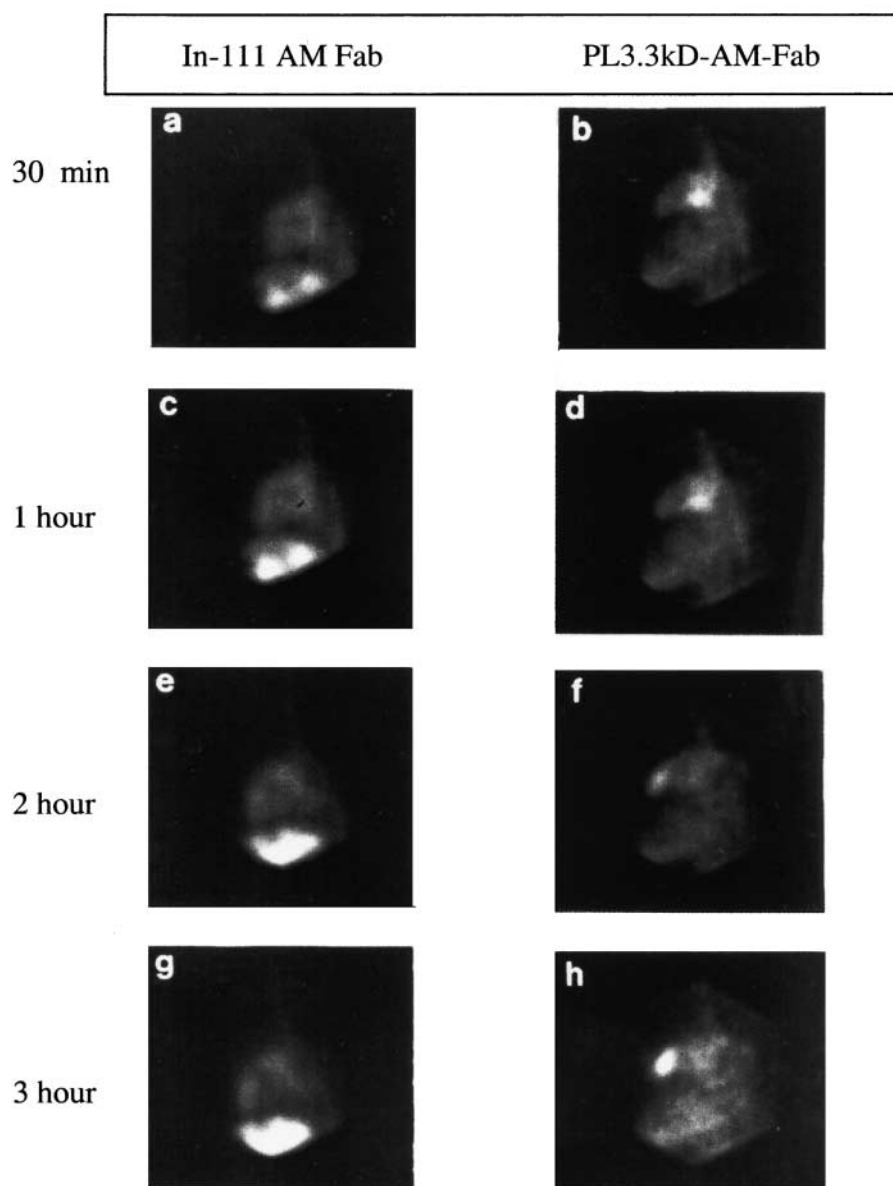


Fig. 4 Serial left lateral gamma images of two dogs with acute experimental myocardial infarction injected with negative charge-modified In-111 labeled antimyosin Fab (right panels) and conventionally In-111 labeled antimyosin Fab (left panels); a and b = 30 min images, c and d = 1h, e and f = 2h, and g and h = 3h post intravenous administration. (From Ref. 50.)

modified and unmodified antimyosin Fab radiolabeled with In-111, were injected into dogs with experimental MI, and at 5 h post injection there were $0.1718 \pm 0.0201\%ID/g$ (52) of negative-charge modified antimyosin Fab in the infarct, since the specific radioactivity of negative-charge modified antimyosin Fab was 100 mCi/mg, the absolute radioactivity in 1 g of infarct would be 85.9 μCi . On the other hand, 500 μg of unmodified antimyosin Fab with a specific radioactivity of

5 mCi/mg with $0.2041 \pm 0.0204\%ID/g$ would only have 5 $\mu Ci/g$ of the infarcted myocardium. Since normal myocardial activities were $0.0056 \pm 0.0004\%ID/g$ with charge-modified antimyosin Fab and $0.0263 \pm 0.0037\%ID/g$ with unmodified antimyosin Fab, the absolute background radioactivities were 2.8 and 0.658 μCi , respectively. Therefore, the difference between target and background for charge-modified antimyosin Fab would be 83.1 μCi and the ratio between

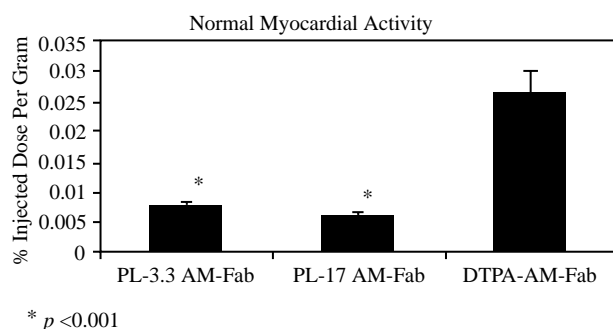


Fig. 5 Normal myocardial activity of polymer modified antimyosin Fab and conventionally In-111 labeled antimyosin Fab. (From Ref. 50.)

target to background would be 30.7:1. The difference in the absolute activity between target and background on the other hand for unmodified antimyosin Fab would be only 4.34 μCi and the ratio 7.6:1. Therefore, with a differential of 83 $\mu\text{Ci/g}$ between target to background, one should be able to clearly visualize the target quite early, whereas a differential of 4.34 μCi may not permit easy early visualization.

To determine whether earlier visualization was due to increased infarct activity or decreased nontarget organ activities, biodistribution data were also compared. Figure 5 shows that even in the normal myocardium (nontarget tissue), the radioactivity was significantly lower for 3.3 and 17 kD negative charge-modified polymer-AM-Fab than that of the conventionally labeled AM-Fab ($p < 0.001$) (50). The applicability of negative charge modification was further demonstrated in a rabbit model where an infarct as small as 216 mg was visualized by in vivo imaging within 3 h after intravenous administration (Fig. 6) (52). The percent injected dose localization in this animal was as high as 1.73% ID/g providing $\approx 86.3 \mu\text{Ci/g}$ of the target. The mean maximal target to nontarget (minimal) ratio from the seven rabbits was 53.9 ± 18.4 (52).

Therefore, negative charge modification of AM-Fab not only imparted lower nontarget organ activity without affecting the target activity, it also provided antibody preparations with very high specific radioactivity enabling the use of less proteinaceous compounds for in vivo administration. Furthermore, it also showed that when high target activity is achieved, visualization was possible earlier based on the difference between target and background activities, rather than target to background ratios. Whether this process of negative charge modification will ever find clinical application may not be determined by scientific feasibility but more by

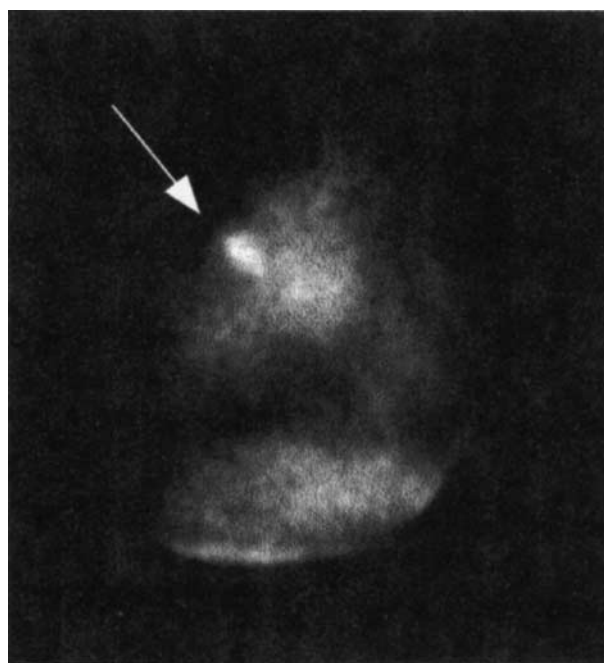


Fig. 6 Left lateral gamma image of a rabbit with a 216 mg experimental myocardial infarct visualized with In-111 labeled negative charge-modified antimyosin Fab. Arrow points to the small infarct visualized in vivo. (From Ref. 52.)

commercial considerations. Nevertheless, a murine-human chimeric antibody F(ab')_2 specific for proliferating smooth muscle cells associated with atherosclerotic lesions has been negatively charge modified, labeled with In-111 and shown to localize in the carotid artery atherosclerotic lesions in all nine patients studied to date (53).

Antimyosin is highly specific and sensitive for diagnosis of myocardial necrosis associated with acute myocardial infarction. Despite its requirement of approximately 24 h lag time for unequivocal diagnostic imaging, it is highly useful for diagnosis of equivocal myocardial infarction, myocardial necrosis associated with unstable angina, right ventricular infarction (54), and perioperative myocardial infarction in by-pass surgery (55).

ANTIMYOSIN IMAGING IN DIAGNOSIS OF VARIOUS CARDIOMYOPATHIES

Myocarditis

Myocarditis is a cardiomyopathy of highly variable clinical manifestations that can in more severe cases, lead to dilated cardiomyopathy and heart failure (56).

It is believed to have a viral origin but the chronic component of the etiology is believed to be autoimmune in nature. To unify the diagnosis of the disease, the Dallas Criteria were formulated (57). The criteria mandated the presence of mononuclear cell infiltration and myonecrosis demonstrated in endomyocardial biopsies for unequivocal diagnosis of myocarditis. However, the criteria focused only at a limited phase of myocarditogenesis. Irrespective of the inflammatory obligatory component of the criteria, the presence of myonecrosis led us to propose that antimyosin immunoscintigraphy should be able to target the myonecrotic component of the disease and provide a very sensitive diagnostic indicator for noninvasive diagnosis of myocarditis. In the initial study of 28 patients with histories and clinical findings suggestive of myocarditis, In-111 antimyosin immunoscintigraphy was positive in 17 patients (61%) (Fig. 7, left panel) and negative in 11 (39%) (Fig. 7, right panel) (26). All antimyosin negative patients were also negative by endomyocardial biopsy criteria, whereas all biopsy positive patients were also positive by biopsy criteria (26). A potential complication in the use of In-111 antimyosin Fab for diagnosis of acute myocarditis is the misinterpretation of residual blood pool activity for myocardial uptake. Since myonecrosis associated with myocarditis is generally diffused and not as intense as seen in acute myocardial infarction, uptake of In-111 labeled antimyosin could also be diffused and of low contrast. Therefore, it is recommended that single photon emission tomographic imaging be performed in case of equivocation. The tomographic reconstruction images in the transverse, sagittal, and coronal views should show myocardial activities rather than blood pool activities.

Dilated Cardiomyopathy

It has been clinically suspected that active myocarditis is capable of resulting in heart failure and acute dilated cardiomyopathy. Although the exact number of cases of active myocarditis or ongoing inflammation that ultimately resulted in idiopathic cardiomyopathy is not known, Dec et al. (28) studied 74 patients (50 men and 24 women) with dilated cardiomyopathy with global ejection fraction less than 0.45 with In-111 antimyosin Fab. Thirty-nine patients (53%) were positive by imaging criteria. Of these 39, 11 patients had histologically verified myocarditis, whereas 28 showed no evidence of myocarditis on biopsy. Out of the remaining 35 patients with normal antimyosin images, 33 were also negative by endomyocardial biopsy criteria. However, two were false negative since they had biopsy verified myocarditis. Thus, the sensitivity of antimyosin imaging was 85% and the predictive value of a normal scan was 94%. However, the specificity was only 54%, using the Dallas Criteria as the gold standard. This low specificity is most probably due to the need to use a Gold Standard that is highly insensitive and because myocarditis is a disease of either right or left or both ventricles, and endomyocardial biopsies are primarily obtained from the right ventricles. Irrespective of this low specificity of antimyosin imaging for diagnosis of myocarditis, patients with abnormal antimyosin scan and biopsy results showed significant improvement in the mean ejection fraction from 0.27 ± 0.02 to 0.43 ± 0.04 within 6 months follow-up examination. Whereas, patients with normal scans and nondiagnostic biopsy results had only slight improvement in the ejection fraction from 0.19 ± 0.02 to 0.24 ± 0.03 . Furthermore,

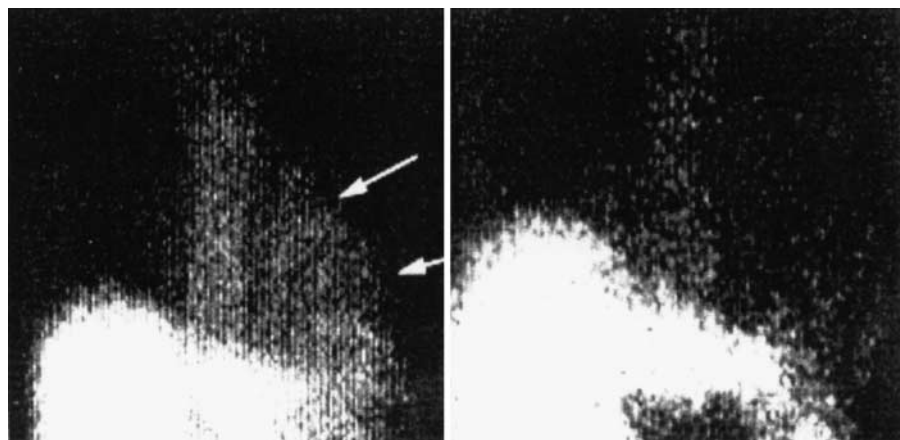


Fig. 7 In-111 Antimyosin gamma images of two patients suspected of having myocarditis: a) positive image (left panel), and b) a negative image (right panel). The arrows point the myocardial activity.

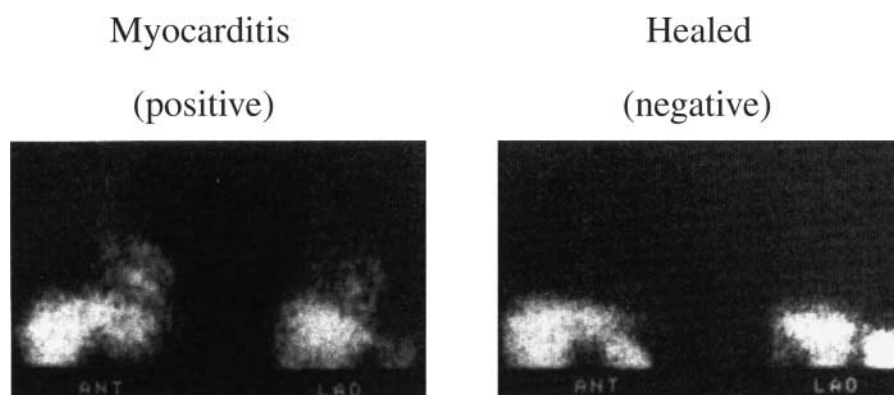


Fig. 8 Planar anterior and LAO In-111 antimyosin gamma images of a patient with biopsy positivity for myocarditis obtained initially when the left ventricular ejection fraction was only 34% (left panels) and after 6 months of steroid therapy (right panels) when the LVEF had normalized to 55% are shown.

those patients with positive antimyosin scans and negative biopsy results showed significant improvement in ejection fraction at follow-up indicating that since spontaneous improvement in cardiac function is a recognized feature of active myocarditis, the subset of patients with antimyosin scan positivity and spontaneous improvement in the cardiac function had myocarditis, which biopsy failed to detect. Negative antimyosin imaging can also be used to follow efficacy of therapy. Fig. 8 shows that although the initial antimyosin images were positive, after 6 months of steroid therapy, the images were negative for myonecrosis component of the Dallas Criteria.

Heart Transplant Rejection

Since commencement of immunosuppressive treatments for acute heart transplant rejection is predicated on the presence of myonecrosis on endomyocardial biopsies, antimyosin imaging should also be applicable for directing therapy. Frist et al. (58) showed that antimyosin could be used to detect myonecrosis associated with acute rejection. However, they also noted that immediately after transplantation, there appeared to be a basal uptake of In-111 labeled antimyosin Fab probably due to myocardial injury associated with transplantation related procedures. Subsequently Ballester et al. showed systematically that this initial antimyosin positivity could return to baseline as quickly as 3 months after transplantation (59). Furthermore, they showed that if antimyosin uptake assessed as Heart to Lung activity ratios remained elevated for greater than 1 year after transplantation, prognosis for the patient was poor and that patient was

a prime candidate for re-transplantation. Various degrees of intensity of In-111 antimyosin uptake were shown in Fig. 9, ranging from normal antimyosin scan (A), to mild uptake (B), to moderate uptake (C), to significant uptake (D) (60).

Antimyosin imaging appears to be highly sensitive for the detection of acute heart transplant rejection. Relative to the gold standard of endomyocardial biopsy, antimyosin has a sensitivity of 95% (61), however, specificity was only 33%. This discrepancy may be due to a lack of specificity of antimyosin or to a low sensitivity of biopsy in the detection of rejection. However, sampling error of endomyocardial biopsy is a distinct possibility that may account for the discrepancy since transplant rejection is histologically a patchy process. Therefore, if antimyosin were taken as the gold standard, then the sensitivity and specificity of endomyocardial biopsy would be 31 and 95%, respectively (61). This is consistent with the low diagnostic yield of endomyocardial biopsy reported for the diagnosis of active myocarditis (62, 63)

Other Cardiomyopathies

Due to the mechanism of targeting of antimyosin Fab, it appears that antimyosin could also be used to delineate various cardiomyopathies as long as there is an association of the disease process with irreversible myocardial injury where the integrity of the cell membrane has been compromised. Therefore, antimyosin Fab has been used successfully to demonstrate myocardial injury due to Adriamycin cardiotoxicity (64), in Rheumatic carditis (65), in Lyme carditis (66), Churg–Strauss disease (67) and Cardiac contusion (68).

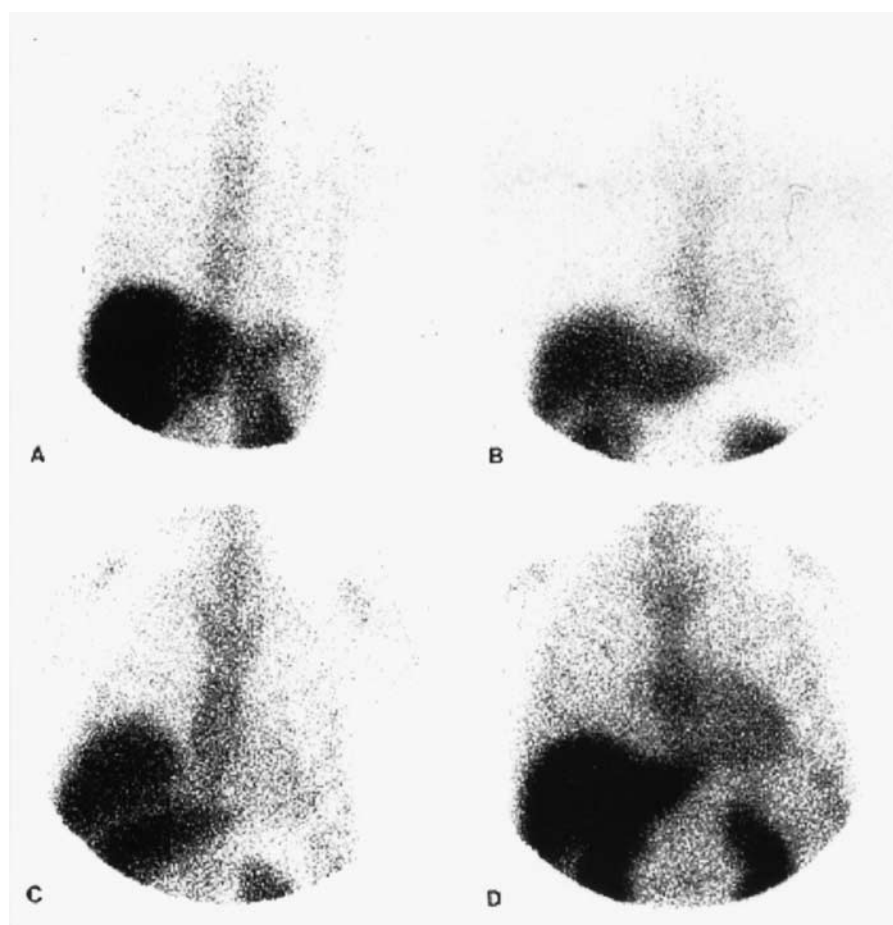


Fig. 9 Anterior gamma images of patients with various degrees of rejection detected by In-111 antimyosin imaging. A = normal scan, B = mild uptake, C = moderate uptake and D = significant uptake. (From Ref. 60.)

TARGETING VASCULAR DISORDERS WITH RADIOLABELED ANTIBODIES

Imaging Blood Clots

Attempts to image blood clots preceded the era of monoclonal antibodies. Spar et al. used polyclonal antifibrinogen antibodies to detect thrombi in vivo in the mid 1960s (69). Since the antibody reacted with fibrinogen as well as fibrin, the specificity for the detection of preformed clots was poor. Hui et al. in 1983 (70) developed monoclonal antibodies (59D8 and 64C5) specific for the β chain of the fibrin molecule. They were made to a 7 amino acid *N*-terminal sequence of the β chain of fibrin. Since the *N*-terminus of the β chain of fibrin constitutes neoantigen generated when fibrinogen is cleaved by thrombin, these monoclonal antibodies did not cross react with fibrinogen. Kudryk et al. (71) also generated a similar monoclonal antifibrin T2G1S that

also reacted with the same *N*-terminal of the β chain of fibrin. Although there are other monoclonal antibodies with varying specificities for fibrin, only 59D8 and T2G1S have seen clinical studies. Fig. 10 (left panel) shows a 24-h-image of the lower extremities of a normal subject, whereas Fig. 10 (right panel) shows a set of spot images in a patient with venographically documented DVT.

It appears that imaging DVT was quite easily feasible in experimental and clinical trials. Whether imaging of pulmonary emboli in a clinical situation with monoclonal antifibrin antibody would be successful is not known. Experimentally, Kanke et al. (72) showed that 64C5 monoclonal antifibrin was able to image PE. However, small PE of <50 mg were not visualized in vivo by gamma imaging despite a correlation between clot size and total antifibrin uptake. Further improvements are needed to make this method of detection of DVT or PE a clinical reality.

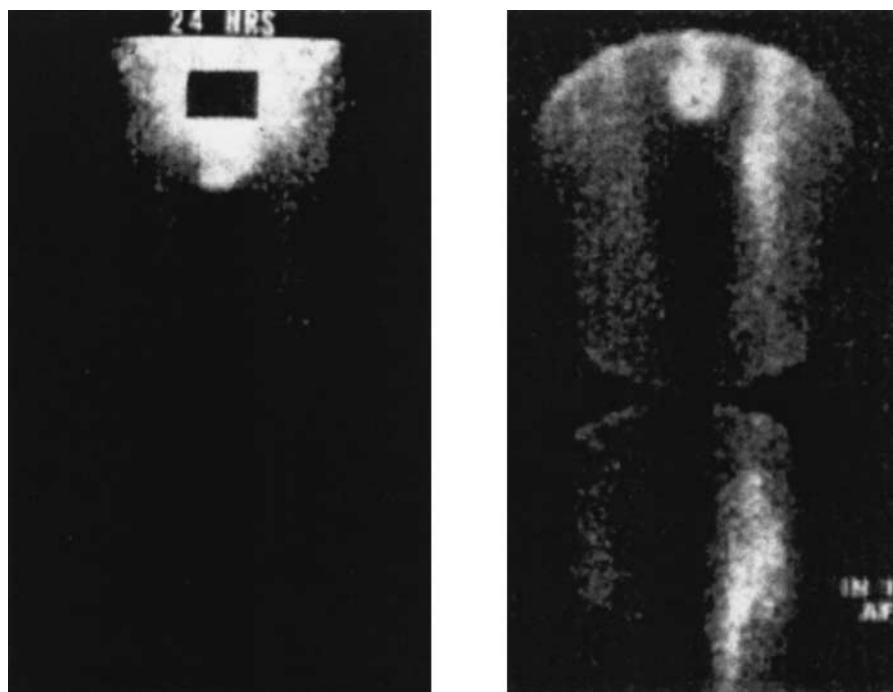


Fig. 10 Anterior gamma images of normal control (left) and patient with DVT (right) with Tc-99m labeled antifibrin Fab. Radiolabeled antifibrin uptake can be seen in the left thigh and calf regions. (From Ref. 32.)

Targeting Atherosclerotic Lesions with Radiolabeled Antibodies

Atherosclerosis is another intravascular pathological disorder that appears to be amenable to targeting with monoclonal antibodies. It was initially thought that atherosclerotic lesions possessed no unique compounds that could serve as specific targets. These lesions are composed of macrophages that had become foam cells by the ingestion of oxidized LDL, and smooth muscle cells of the synthetic phenotype. No differentiated antigens were thought to exist in the lesions. Therefore, conventional modes of diagnosis, such as arteriography or ultrasonography, were believed to suffice for providing information on the anatomical narrowing of the affected vessels. Although these methods are effective, they cannot provide pathophysiologic information that may shed light on the stability as well as the pathogenesis of the lesions.

To elucidate the potential metabolic component of the pathogenesis of atherosclerotic lesions, several approaches have been taken experimentally. Radiolabeled oxidized LDL (73) as well as antibodies to the activated macrophages (74) have been tried with some

results. A monoclonal antibody that is specific for a complex antigen produced only by the proliferating smooth muscle cells of atherosclerotic lesions was developed by Scotgen Biopharmaceuticals, Inc. (75). This IgM isotype antibody designated Z2D3 (76) was shown to be able to target experimental atherosclerotic lesions in a rabbit model (77). The IgM subclass of Z2D3 was class switched to IgG as well as genetically engineered to produce a murine-human chimeric IgG Z2D3 antibody. Fig. 11 (top panel) shows a set of gamma images of a rabbit with experimental atherosclerotic lesions targeted with In-111 labeled chimeric Z2D3 F(ab')₂. The lesions could be visualized in the region of the descending aorta that was the site of experimentally induced lesions. In this model, the lesions were induced by de-endothelialization of the descending aorta from the region of the diaphragm to the bifurcation of the femoral arteries (75). The rabbits were kept on 6% peanut oil, 2% cholesterol enriched chow for at least 3 months. This protocol produced lesions that are more akin to human fibrous lesions, unlike the fatty streak lesions observed in Watanabi hyperlipidemic rabbits. As controls, rabbits with similar lesions were injected with In-111 labeled charge-modified normal human IgG F(ab')₂ (Fig. 11, lower panels). No specific targeting was seen in vivo in

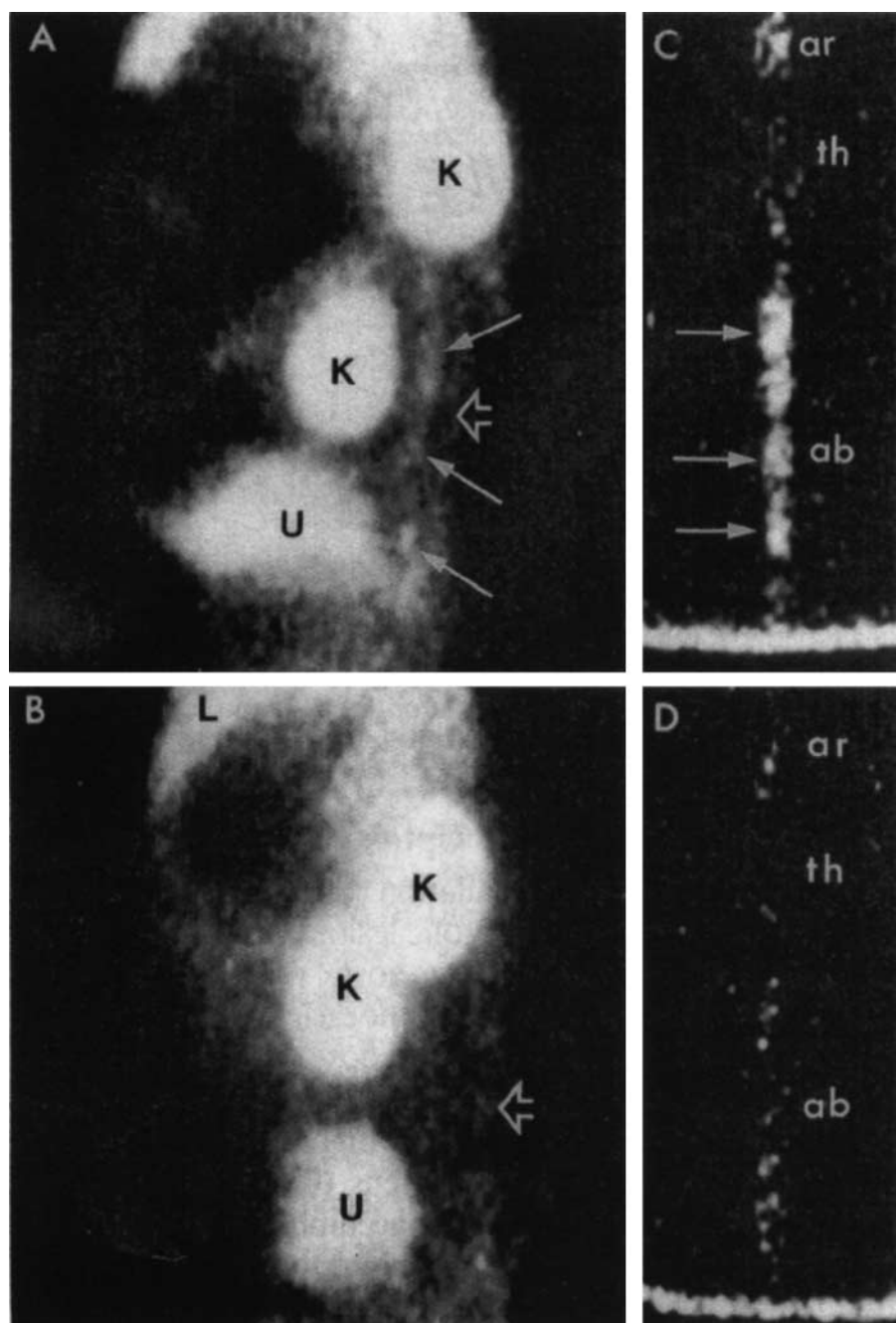


Fig. 11 Left lateral oblique images of rabbits with experimental atherosclerotic lesions induced in the descending aorta imaged with In-111 labeled murine-human chimeric Z2D3 F(ab')₂ (Top panels) and In-111 labeled human IgG F(ab')₂ (Bottom panels). The in vivo gamma images are shown in the left panels (k = kidney, U = urinary bladder activity, solid thin arrows = atherosclerotic lesions, and open larger arrow = spinal activity), and the ex vivo images of the excised aortas from the aortic arch to the femoral bifurcation are shown in the right panels.

the region of the experimental lesions. Immunohistologically, this antibody stains the region of smooth muscle cell proliferation and not the quiescent smooth

muscles of the contractile phenotype of the media (Fig. 12). Ultimately, the negative charge-modified chimeric Z2D3 F(ab')₂ was used to determine whether

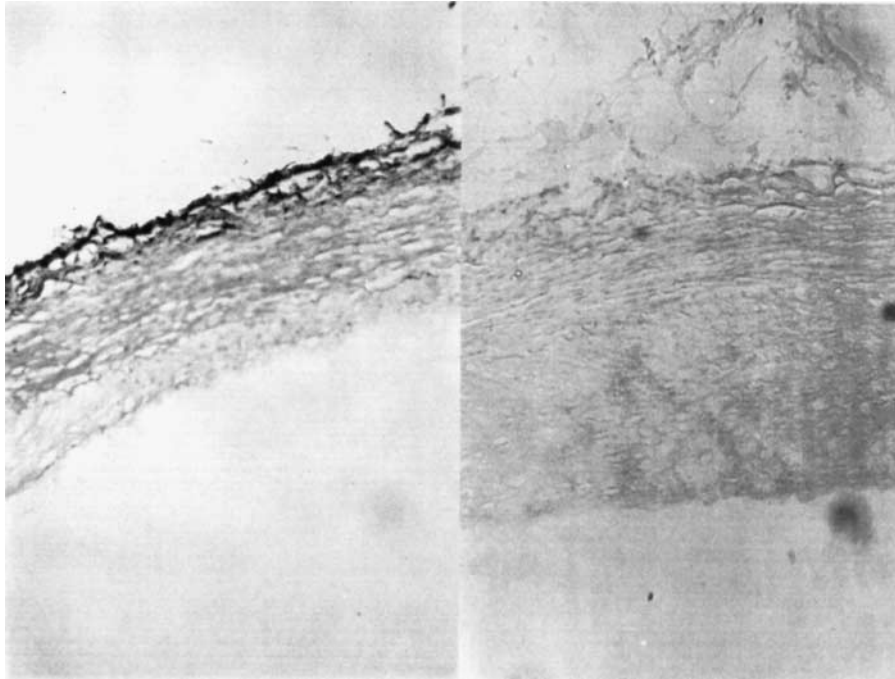
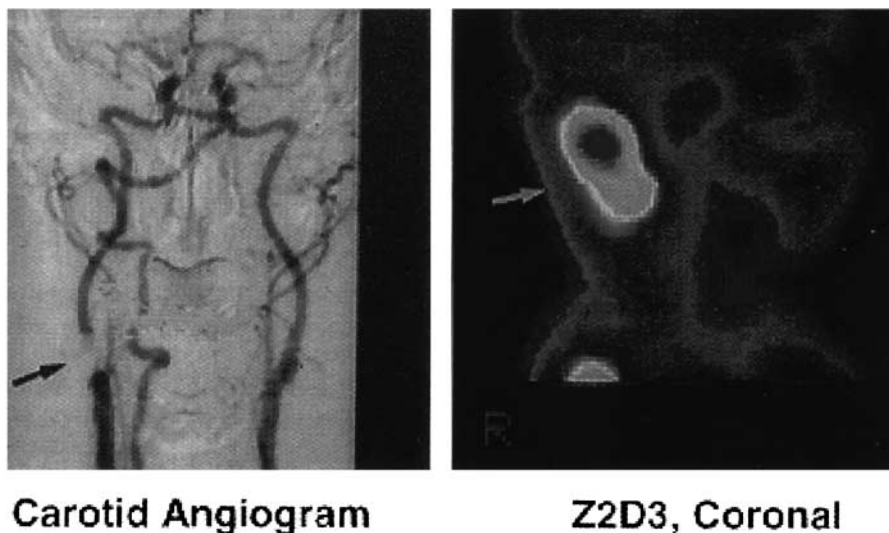


Fig. 12 Immunoperoxidase staining of frozen sections of atherosclerotic rabbit descending aorta (left panel) and normal rabbit aorta (right panel).

it could delineate atherosclerotic lesions in carotid lesions in patients (78). Using planar and SPECT imaging, it was observed that the lesions were better

detected with SPECT imaging (Fig. 13) (79). Whether this antibody Z2D3 will have wide clinical application must await additional trials.



Carotid Angiogram

Z2D3, Coronal

Fig. 13 Coronal tomographic image of a patient with carotid atherosclerotic lesions obtained 4 h after IV administration of In-111 labeled negative-charge modified murine-human chimeric Z2D3 F(ab')₂ (right panel). The arrow denotes the carotid lesion. The acrotid angiogram demonstrating a severe right internal carotid artery lesion (arrow) is shown in the left panel).

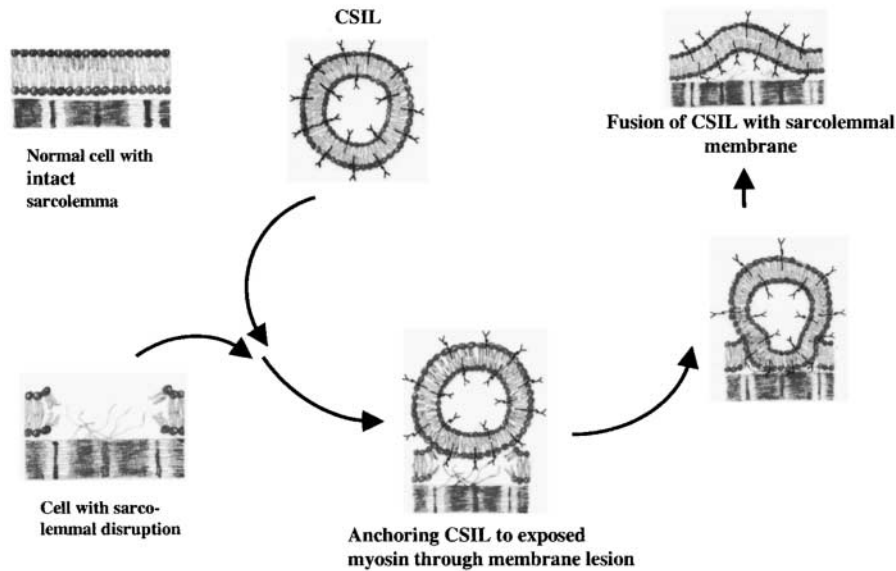


Fig. 14 Diagrammatic representation of the hypothesized mechanism of cell membrane sealing and salvage by antimyosin liposomes. (Adapted From Ref. 84.)

TARGETING CELL MEMBRANE LESION WITH CYTOSKELETAL-ANTIGEN SPECIFIC IMMUNOLIPOSOMES

Two primary kinds of cell deaths are now known to exist in the animal kingdom. Oncotic (formerly known as necrotic) and apoptotic cell deaths differ in that the former is due to external nonphysiological injury to the cell (80) and the latter "apoptosis is due to internally mandated process of suicide cell attrition or programmed cell death known" (81). Under pathological conditions of an ischemic insult to organs, such as the heart and the brain, the primary mode of cell death in the center of infarction is oncotic (82). Apoptosis plays a role in the periphery and in reperfused regions (83). Acute myocardial infarction is believed to be the result of oncosis, where development of cell membrane lesions constitute the irreversible phase. The presence of these lesions, which is initially represented by sub-microscopic holes in the sarcolemma in the acute phase of myocardial injury, permits washout of intracellular macromolecules into the circulation. At the same time, certain intracellular proteins including the components of the cytoskeleton, such as myosin and vimentin, become exposed through these holes to the surroundings milieu. Appropriately labeled antibodies against intracellular cytoskeletal antigens have been used to demarcate these cell membrane lesions (25).

Therefore, we hypothesized that if these cell membrane lesions could be sealed at the time or prior to reperfusion, then viability of myocardial cells should be preserved (84). To achieve cell membrane lesion sealing, we proposed that liposomes would be the ideal agent. However, the liposomes must be provided with target recognition sites that are specific for the lesions in the cell membrane. To achieve this, a monoclonal antibody that specifically recognizes myosin of the myofilaments that constitutes the cytoskeleton of the muscle cells was used. Thus, antimyosin immunoliposomes should be able to target the cytoskeletal myosin exposed through the small lesions of the cell membrane during the process of ischemic injury before the lesions become too large for repair (Fig. 14).

Preservation of Cell Viability by Cytoskeleton-specific Immunoliposomes

Proof of concept studies were performed in vitro in cell cultures of H9C2 rat embryonic cardiocytes. Two million H9C2 cardiocytes were incubated in aliquots of 10% fetal calf serum in DMEM at 37°C, 5% CO₂. After overnight incubation, the cells were washed in phosphate-buffered saline and recultured in fresh medium (with or without different liposome preparations) (84). Nitrogen was bubbled through the medium vigorously for 4 min dislodging all cells (>95%) from the bottom of the flasks. The flasks were then closed tightly to maintain hypoxia

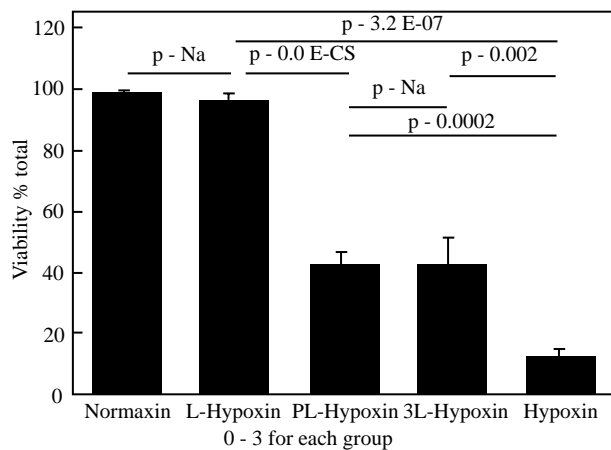


Fig. 15 Comparison of cell viability of H9C2 cardiocytes in culture by Trypan Blue dye uptake criteria. Viability of hypoxic (24 h) cells treated the IL were compared to PL or IgG-liposome. Positive control was the normoxic cultures and the negative control was hypoxic cells with no treatment. (From Ref. 84.)

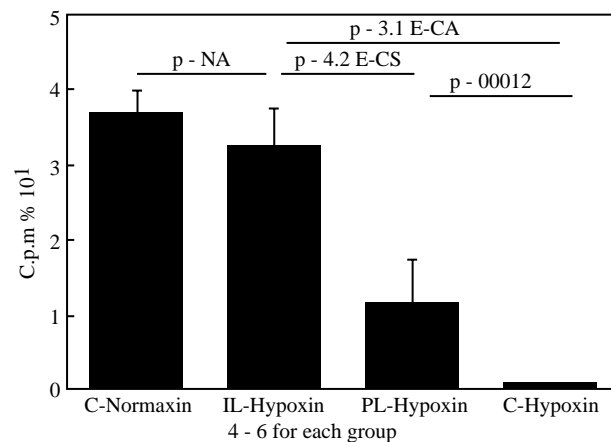


Fig. 16 Comparison of cell viability of H9C2 cardiocytes by ³H-Thymidine uptake criteria. Viability of hypoxic cells (24 h) treated with IL was compared to Normoxic cells ($p = \text{NS}$), PL treated hypoxic cells ($p < 0.01$) and untreated hypoxic cells ($p < 0.01$). (From Ref. 84.)

through overnight incubation at 37°C. H9C2 cardiocytes in hypoxic culture conditions were incubated with anti-myosin-immunoliposomes (IL), plain liposomes (PL) and control nonspecific IgG liposomes (IgL). Hypoxic and normoxic cardiocytes without liposome treatment were included as additional controls. Liposomes were prepared from a mixture of egg phosphatidylcholine and cholesterol at a molar ratio of 7:3. For incorporation of antimyosin antibody 2G42D7 into liposomes, the antibody was modified with N-glutaryl phosphatidyl ethanolamine (NGPE) (85,86) and added to the detergent-solubilized lipids. Liposome preparations containing fluorescent rhodamin-labeled lipids were also used for fluorescent microscopy and confocal microscopy (84).

At the end of the hypoxic period, viability of the cells were assessed by the Trypan Blue dye exclusion criteria. Fig. 15 demonstrated that the viability of hypoxic cells treated with IL were not significantly different from the viability of the normoxically cultured H9C2 cardiocytes. The viability of hypoxic cardiocytes treated with plain liposomes or IgG-liposomes was significantly lower than either those of immunoliposome treated hypoxic cells or normoxic cells. Since there was no difference in viability in cardiocytes treated with PL or IgG-liposomes, subsequent studies utilized only PL treatment as control. The phenomenon of preservation of cell viability with IL treatment was reaffirmed by ³H-thymidine uptake studies (84). DNA replication as reflected by ³H-thymidine uptake, was similar in the normoxic cells and the hypoxic cells treated with IL. However, cells treated with PL had only about 10% of ³H-thymidine

uptake of either the normoxic or IL treated hypoxic cells. Hypoxic cells with no treatment at all showed almost no

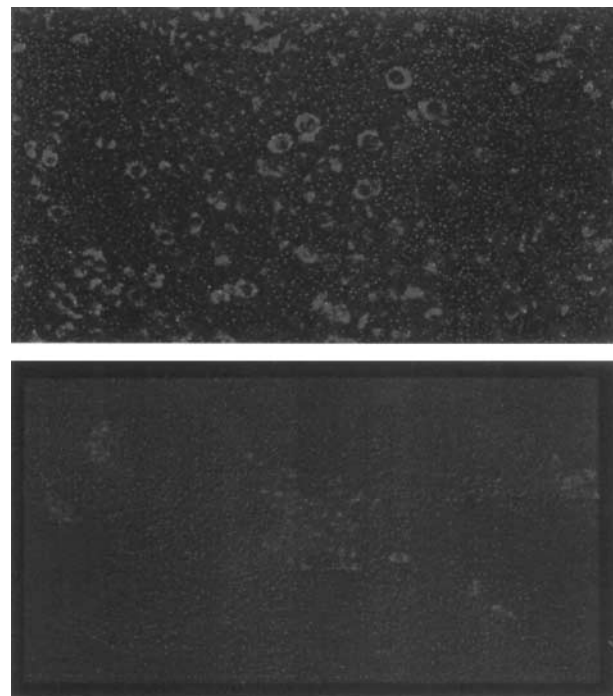


Fig. 17 Epifluorescent micrographs of 24 h hypoxic H9C2 cardiocytes treated with rhodamine labeled IL (right) or rhodamine labeled PL (left). Cells treated with IL were all viable; however, they all showed attached rhodamine fluorescent indicating attachment of IL to the cells. (From Ref. 84.)

^3H -thymidine uptake, indicative of a lack of viable cells (Fig. 16).

The sites of targeting of IL on hypoxic cardiocytes was visualized by fluorescent microscopy, using rhodamine labeled antimyosin immunoliposomes. Fig. 17 shows that only cells treated with antimyosin–rhodamine labeled IL were still confluent in the culture and that almost all cells were labeled with fluorescent liposomes. Those cells treated with rhodamine labeled PL showed extremely sparse number of cells still attached to the culture plates at 24 h of incubation, with essentially no or minimal fluorescence. Confocal microscopic examination of the cultures treated with rhodamine labeled IL showed that the cells still retained their morphology and shape with scattered fluorescent liposomes attached to the cell membranes (Fig. 18, left panel). Those cells treated with rhodamine labeled PL were shrunken and only a few random cells showed some nonspecific attachment of fluorescent PL (Fig. 18, right panel). In this study, untreated hypoxic cells were all dead and since there was no fluorescent compounds added in them, no micrographs were obtained.

It is also very important, how efficacious this protective effect can last. Therefore, we undertook a study to determine whether antimyosin IL can protect severely injured cardiocytes cultured under hypoxic conditions for 1–5 days (87). The experiment was designed as already described, however, cells were kept under hypoxic conditions during different times. Cell viability was assessed by ^3H -Thymidine (^3H -T) uptake. Untreated hypoxic cardiocytes (HC), normoxic cardiocytes (NC), and cardiocytes treated with PL were used as controls. Survival of NC increased from 100% (the mean ^3H -T uptake in control NC was assigned 100%) to ca. 250%

after 24 h indicative of cell replication, whereas virtually no cardiocytes survived after 24 h of hypoxia. PL added to HC provided some protection. Cell viability after 24 h was ca. 78%, which dropped to ca. 4% after 2 days of hypoxia, and to less than 1% after 3 or more days. However, when hypoxic cells were treated with IL, not only did IL conferred protection, IL also permitted cell replication. This is evidenced by the similarity in the increase of ^3H T uptake after 24 h in NC (250%) and in IL-treated HC (ca. 225%, $p = \text{NS}$). These data show that the protective effect imparted by IL on hypoxic cells leads to a long-term preservation of cardiocyte viability that might be especially apropos for preservation of cell viability during organ transportation for transplantation.

Preliminary In Vivo Experiments on Decreasing Infarct Size with CSIL

In our preliminary study utilizing a model of myocardial infarction in rabbits, when antimyosin IL, IgG-L, PL, or saline placebo were delivered concomitant with circumflex coronary artery occlusion for 45 min followed by reperfusion, the infarct size determined at 6 h of reperfusion, of IL treated rabbit hearts were approximately 5–10% of the control PL, IgG-L, or saline placebo treated rabbit hearts (88). Fig. 19 shows nitroblue tetrazolium stained heart sections of IL treated rabbit (top left panel) relative to PL (top right panel), IgG-L (bottom left panel), or saline placebo (bottom right panel) treated rabbit hearts. In nitroblue tetrazolium staining, normal myocardium stained as purple to dark brown, where as the infarcted myocardium is seen as light colored regions. Whether intravenously delivered IL would have any efficacy

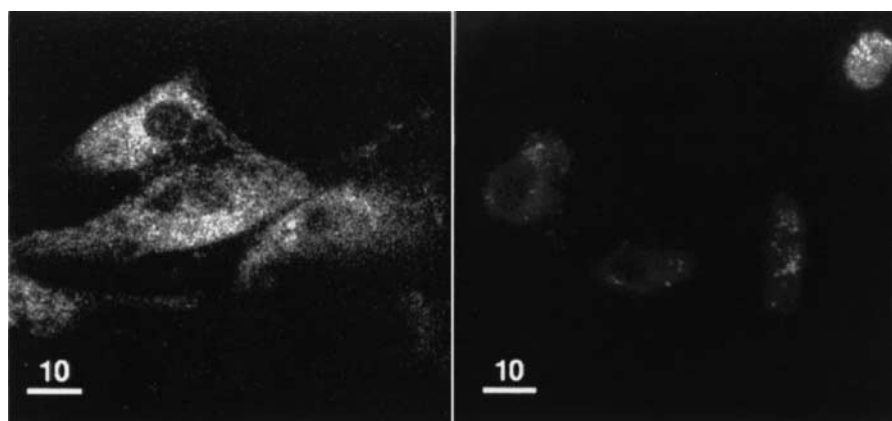


Fig. 18 Confocal micrographs of 24 h hypoxic H9C2 cardiocytes treated with rhodamine labeled IL (right) or rhodamine labeled PL (left). The micrographs are shown in pseudocolors. Cells treated with IL showed retention of membrane integrity and cell morphology (left panel). Liposomes represented as yellow colored regions are also discernable on the cells. Cells treated with PL showed presence of only dead cells with only a few cells having nonspecifically attached PL (right panel). (From Ref. 84.)

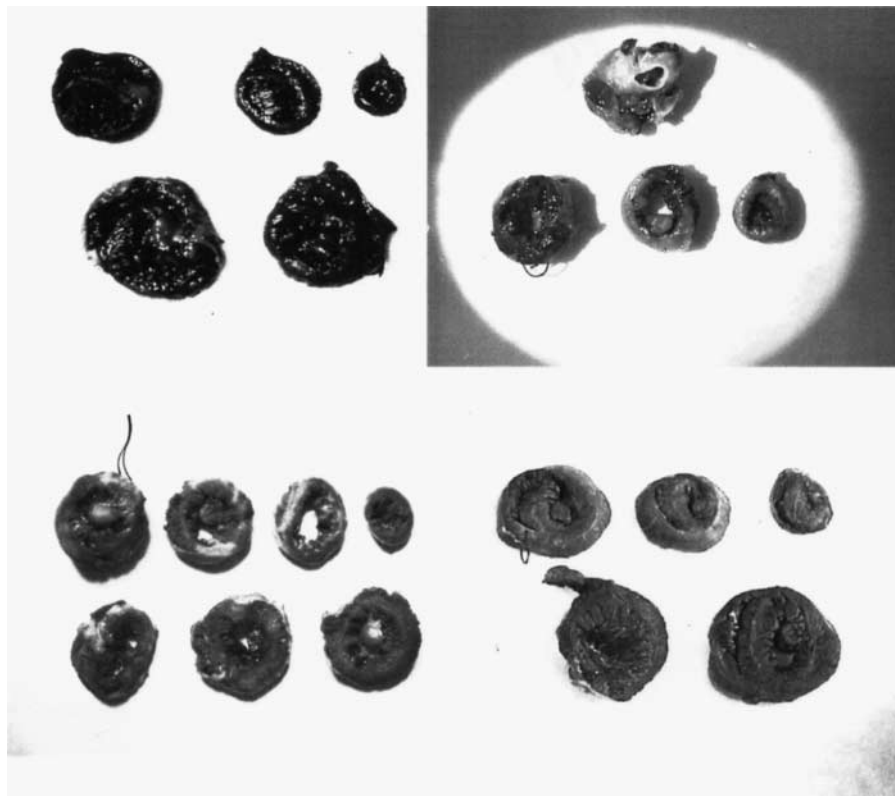


Fig. 19 Nitroblue tetrazolium stained heart sections of rabbits with acute myocardial infarction treated with PL (left panel), saline placebo (middle panel) and IL (right panel). The sites of left circumflex coronary artery occlusion can be seen in the middle and right panels to be at the 4th slices by the presence of the silk sutures. Normal myocardium is stained purple or brown, whereas the infarcted regions remain as light colored regions.

relative to direct intraatrial delivery of the liposomes is not known at this time. None the less, in acute myocardial infarction, angioplasty is an alternative method to thrombolytic reperfusion therapy, therefore, in combination with angioplasty, it may be possible to deliver the IL directly into the infarct zone at the time of angioplasty.

Fusion of Cytoskeleton-specific Immunoliposomes with Hypoxic Cells

To demonstrate that preservation of cell viability is due to fusion of IL with the cell membranes, IL were prepared with intraliposomal silver grains. The rationale was that the only way silver grains can enter the treated hypoxic cardiocytes is by fusion of the IL with the cell membrane that resulted in preservation of cardiocyte viability. If endocytosis of the IL is the process of internalization without fusion, then plain liposomes should also show internalization of the silver grains. To prepare such liposomes (84), silver nitrate was added in a buffer solution during liposome preparation by sonication.

Liposomes, purified from the nonentrapped silver nitrate by dialysis were dialyzed in 0.12 M NaCl solution and exposed to light for 1 h that led to the formation of fine electron-dense precipitate of silver oxide inside liposomes.

Figure 20 shows that only cells treated with silver grains impregnated antimyosin IL retained the cell morphology and showed internalization of the silver grains (Fig. 20, top panel). Cardiocytes treated with silver grains impregnated PL showed no viable cells. However, after extensive search one viable cell was identified. The silver grains were observed outside the cell (Fig. 20, bottom panel) (89). The size of each clump of silver grains seen in Fig. 20 top panel is consistent with the size of the liposomes that were prepared and was approximately 200–300 μm in diameter. From this study, it was not possible to determine whether silver grains would also migrate to the nucleus. Silver grains, once internalized may not be as mobile as soluble pharmaceuticals or genetic constructs. Furthermore, since no internalization of silver grains was observed with PL, the mechanism of internalization seen with IL does not appear to be due to endocytosis or pinocytosis. Therefore,

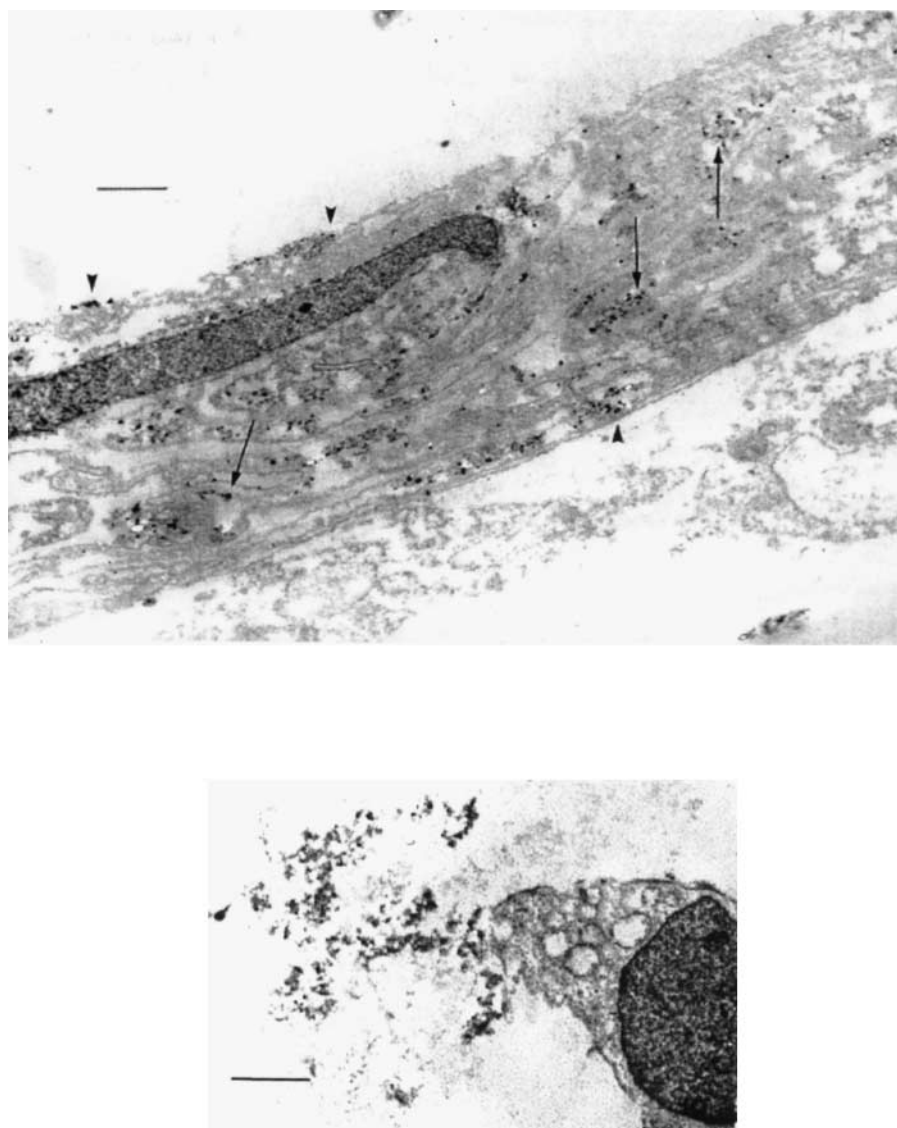


Fig. 20 Transmission electronmicrographs of hypoxic H9C2 cardiocytes treated with silver grains impregnated IL or PL. Silver grains in seen intracellularly in IL treated cells; arrows, arrow heads represents 1 μ m. (From Ref. 84.)

internalization of the silver grains delivered by IL is consistent with the proposal that cytoskeleton-specific IL is able to fuse with the membrane of hypoxic cells releasing their contents into the cytoplasm.

Intracytoplasmic DNA Delivery

Since antimyosin immunoliposomes can be used to seal cell membrane lesions (84) and intraliposomally entrapped silver grains can be delivered directly into the cytoplasm (88), we also hypothesize that if we substitute genetic constructs for the silver grains, then such constructs should

also be deliverable directly into the cytosol (90). Delivery of genetic constructs directly into the cytosol, bypassing the endocytic route, might result in higher efficiency of gene expression. To demonstrate this hypothesis, we initially used antimyosin sFc vector (90), then pGL2 luciferase vector (91) and finally bacterial beta galactosidase vector (92). The rationale for using three different vectors is to show that irrespective of the origin of the vectors, one murine, one insect, and one bacteria origin, they all can transfect the cardiocytes and result in very efficient gene expression. All these in vitro studies showed that vectors delivered directly into the cytoplasm were



Fig. 21 Evidence of successful gene transfection with bacterial β -galactosidase vectors with IL in hypoxic cells, or with cationic liposomes or IgG-liposomes. The β -galactosidase activity was developed with X-Gal at pH 7.0. Cells treated with IL- β -galactosidase vector showed many cells with blue color. The culture was also confluent. Cells treated with cationic liposome- β -galactosidase vector showed only two cells with gene expression. The cells were also not confluent indicating cell attrition. IgG-liposome treated cells also showed no β -galactosidase expression.

expressed with very high efficiency. Comparative studies with standard cationic liposomes showed that although transfection efficiency of cationic liposomes was very efficient, due to the over-loading of the cell membrane with approximately 40,000 liposomes per cell (93), the viability of transfected cells declined dramatically. Furthermore, since internalization of the vectors is via endosomes, the factor of vector loss due to lysosomal activity was considerable as evidence by a lower number of cells expressing the gene products.

When IL was used for transfection of bacterial β -gal vector into hypoxic cells, the majority of the cells showed bacterial β -galactosidase activity when stained with X-gal color reagent. Although individual cells transfected with IL displayed less β -galactosidase activity than those few cells expressing the enzyme activity after cationic liposome transfection, the total number of cells demonstrating gene expression were at least 40 times more (94). This result is consistent with our calculation that only 3–4 copies of vectors were delivered per cell by IL transfection, whereas when cationic liposomes were used, more than 3000 copies of the vectors were delivered per cell (94). Fig. 21 shows the comparison of the photomicrographs of cells with successful expression of the β -galactosidase vector transfected with IL (A) or cationic liposomes (B) and IgG liposomes (C). Those cells treated with PL similar to IgG-L also showed no transfection.

Although the present novel method of gene transfection requires institution of an insult to the cell membrane, our studies with cell viability already showed that almost no cells were lost due to cell death in these treated cultures. Therefore, for efficient *in vitro* cell transfection and gene expression, this immunoliposome method would be highly desirable. Whether this method would be valuable *in vivo*, additional studies are necessary.

CONCLUSIONS

Monoclonal antibodies are the targeting moieties of the “targeted drug delivery systems.” Whether the drugs to be targeted are radioisotopes for diagnosis or therapy, or chemotherapeutic agents such as doxorubicin that are directly attached to the antibodies as immunoconjugates or entrapped in liposomes and targeted as immunoliposomes, the antibodies provide the most easily available and versatile targeting reagent. Although this article has concentrated on the demonstration of antibodies in targeted drug delivery in the cardiovascular system, their potential is by no means restricted to this system. The imaging and therapeutic applications described for the cardiovascular system should be amenable to adaptation to other organ systems. A cursory discussion

of monoclonal antibodies in oncological uses was provided. Although the potential targeted drug delivery applications in neurological and other systems were not discussed in this article, their utilities are by no means negligible. However, it would be the subject of another article. Therefore, only the cardiovascular application of monoclonal antibodies in targeted drug delivery has been attempted in this review.

REFERENCES

- Ehrlich, P.; Herter, C.A. Ueber Einige Verwendungen der Naphtochinonsulfosaure. *Z. Physiol. Chem.* **1904**, *61*; 379–392.
- Pressman, D.; Keighley, G. Zone of Activity of Antibodies as Determined by the Use of Radioactive Tracers: Zone of Activity of Nephrotoxic Antikidney Serum. *J. Immunol.* **1948**, *59*; 141–146.
- Kohler, G.; Milstein, C. Continuous Cultures of Fused Cells Secreting Antibodies of Predefined Specificity. *Nature: (London)* **1975**, *256*; 495–497.
- Mach, J.-P.; Carrel, S.; Forni, M.; Ritschard, J.; Donath, A.; Alberto, P. Tumor Localization of Radiolabeled Antibodies Against Carcinoembryonic Antigen in Patients with Carcinoma. *New Eng. J. Med.* **1980**, *303*; 5–10.
- Goldenberg, D.M.; DeLand, F.H.; Kim, E.E.; Primus, F.J.; van Nagell, J.R., Jr.; Estes, N.; DeSimone, P.; Rayburn, P. Use of Radiolabeled Antibodies to Carcinoembryonic Antigen for the Detection and Localization of Diverse Cancers by External Photoscanning. *New Eng. J. Med.* **1978**, *298*; 1384–1388.
- Mach, J.P.; Buchegger, F.; Forni, M. Use of Radiolabeled Monoclonal Anti-CEA Antibodies for the Detection of Human Carcinomas by External Photoscanning and Tomoscintigraphy. *Immunol. Today* **1981**, *2*; 239–247.
- Bischof-Delaloye, A.; Delaloye, B.; Buchegger, F. Clinical Value of Immunoscintigraphy in Colorectal Carcinoma Patients: A Prospective Study. *J. Nucl. Med.* **1989**, *30*; 1646–1656.
- Rainsbury, R.M.; Westwood, J.H.; Coombes, R.C. Localization of Metastatic Breast Carcinoma by Monoclonal Antibody Chelate Labeled with In-111. *Lancet* **1983**, *i*; 934–938.
- Chatal, J.F.; Saccavini, J.C.; Fumoleau, P. Immunoscintigraphy of Colon Carcinoma. *J. Nucl. Med.* **1984**, *25*; 307–314.
- Kramer, E.L.; Noz, M.E.; Liebes, L.; Murthy, S.; Tiu, S.; Goldenberg, D.M. Radioimmunodetection of Non-Small Cell Lung Cancer Using Technetium-99m-Anticarcinoembryonic Antigen IMMU-4 Fab' Fragment. Preliminary Results. *Cancer* **1994**, *73* (3 Suppl); 890–895.
- Epenetos, A.A.; Sheperd, J.; Britton, K.E. I-123 Radioiodinated Antibody Imaging of Occult Ovarian Cancer. *Cancer* **1985**, *55*; 984–987.
- Babaian, R.J.; Lamki, L.M. Radioimmunoscintigraphy of Prostate Carcinoma. *Semin. Nucl. Med.* **1989**, *19*; 309–321.
- Larson, S.M.; Carrasquillo, J.A.; Krohn, K.A. Diagnostic Imaging of Malignant Melanoma with Radiolabeled Antitumor Antibodies. *JAMA* **1983**, *249*; 811–812.
- Carasquillo, J.A.; Mulshine, J.L.; Bunn, B.A. Tumor Imaging of In-111 T101 Monoclonal Antibody is Superior to Iodine-131 T101 in Cutaneous T-Cell Lymphoma. *J. Nucl. Med.* **1987**, *28*; 281–287.
- Mariani, G.; Molea, N.; Bacciardi, D.; Boggi, U.; Fornaciari, G.; Campani, D.; Salvadori, P.A.; Giulianotti, P.C.; Mosca, F.; Gold, D.V. Initial Tumor Targeting, Biodistribution, and Pharmacokinetic Evaluation of the Monoclonal Antibody PAM4 in Patients with Pancreatic Cancer. *Cancer. Res.* **1995**, *55* (23 Suppl); 5911s–5915s.
- Peltier, P.; Curtet, C.; Chatal, J.-F. Radioimmunodetection of Medullary Thyroid Cancer Using a Bispecific Anti-CEA/Anti-insulin-DTPA Antibody and an Indium-111 Labeled DTPA-Dimer. *J. Nucl. Med.* **1993**, *34*; 1267–1273.
- Pimm, M.V.; Fells, H.F.; Perkins, A.C.; Baldwin, R.W. Iodine-131 and Indium-111 Labelled Avidin and Streptavidin for Pre-Targeted Immunoscintigraphy with Biotinylated Anti-Tumour Monoclonal Antibody. *Nucl. Med. Commun.* **1988**, *9* (11); 931–941.
- Brenner, T.L.; Adams, V.R. First Mab Approved for Treatment of Metastatic Breast Cancer. *J. Am. Pharm. Assoc.* **1999**, *39*; 236–238.
- McLaughlin, P. Clinical Status and Optimal Use of Rituximab for B-Cell Lymphomas. *Oncology* **1998**, *12*; 1763–1777.
- Nielsen, U.B.; Marks, J.S. Internalizing Antibodies and Targeted Cancer Therapy: Direct Selection from Phage Display Libraries. *Pharm. Sci. Technol. Today* **2000**, *3* (8); 282–291.
- Mahe, M.A.; Fumoleau, P.; Fabbro, M. A Phase II Study of Intraperitoneal Radioimmunotherapy with Iodine-131-labeled Monoclonal Antibody OC-125 in Patients with Residual Ovarian Carcinoma. *Clin. Cancer. Res.* **1999**, *5* (S10); 3249s–3253s.
- Winkler, U.; Barth, S.; Schnell, R.; Diehl, V.; Engert, A. The Emerging Role of Immunotoxins in Leukemia and Lymphoma. *Ann. Oncol.* **1997**, *8* (Suppl 1); 139–146.
- Trail, P.A. Cure of Xenografted Human Carcinoma by BR96-doxorubicin Immunoconjugates. *Science* **1993**, *261*; 212–215.
- Bernstein, N. Antibody-directed Targeting of Liposomes to Human Cell Lines: Role of Binding and Internalization on Growth Inhibition. *Cancer Res.* **1987**, *47*; 5954–5959.
- Khaw, B.A.; Beller, G.A.; Haber, E.; Smith, T.W. Localization of Cardiac Myosin-Specific Antibody in Myocardial Infarction. *J. Clin. Invest.* **1976**, *58*; 439–446.
- Yasuda, T.; Palacios, I.F.; Dec, G.W.; Fallon, J.T.; Gold, H.K.; Leinbach, R.C.; Strauss, H.W.; Khaw, B.A.; Haber, E. Indium-111 Monoclonal Antimyosin Antibody Imaging in the Diagnosis of Acute Myocarditis. *Circulation* **1987**, *76*; 306–311.
- Ballester, M.; Carrio, I.; Abada, M.L. Patterns of Evolution of Myocyte Damage After Human Heart Transplant Detected by 111 In Monoclonal Antimyosin. *Am. J. Cardiol.* **1988**, *62*; 623–627.
- Dec, G.W.; Palacios, I.; Yasuda, T.; Fallon, J.T.; Khaw, B.A.; Strauss, H.W.; Haber, E. Antimyosin Antibody Cardiac Imaging: Its Role in the Diagnosis of Myocarditis. *J. Am. Col. Cardiol.* **1990**, *16* (1); 97–104.

29. Ballester, M.; Martí, V.; Carrió, I. Spectrum of Alcohol-Induced Myocardial Damage Detected by Indium-111-Labeled Monoclonal Antimyosin Antibodies. *J. Am. Coll. Cardiol.* **1997**, *29* (1); 160–167.
30. Estorch, M.; Carrio, I.; Berna, L. 111 In-Antimyosin Scintigraphy After Doxorubicin Therapy in Patients with Advanced Breast Cancer. *J. Nucl. Med.* **1990**, *31*; 1965–1969.
31. Narula, J.; Khaw, B.A.; Southern, J. Monoclonal Antibodies in Cardiovascular Diseases. Khaw, B.A., Narula, J., Strauss, H.W., Eds.; Lea & Febiger: Philadelphia, 1994; 118–126.
32. Knight, L.C. Antifibrin Antibody for Detection of Deep Vein Thrombosis. *Monoclonal Antibodies in Cardiovascular Diseases*; Khaw, B.A., Narula, J., Strauss, H.W., Eds.; Lea & Febiger: Philadelphia, 1994; 171–186.
33. Narula, J.; Ditlow, C.; Chen, F.; Khaw, B.A. Monoclonal Antibodies for Detection of Atherosclerotic Lesions. *Monoclonal Antibodies in Cardiovascular Diseases*; Khaw, B.A., Narula, J., Strauss, H.W., Eds.; Lea & Febiger: Philadelphia, 1994; 206–215.
34. Kairemo, K.J.; Wiklund, T.A.; Liewendahl, K. Imaging of Soft-tissue Sarcomas with In-111-Labeled Monoclonal Antimyosin Fab Fragments. *J. Nucl. Med.* **1990**, *31*; 23–31.
35. Khaw, B.A.; Scott, J.; Fallon, J.T.; Haber, E.; Homcy, C. Myocardial Injury: Quantitation by Cell Sorting Initiated with Anti-Myosin Fluorescent Spheres. *Science* **1982**, *217*; 1050–1053.
36. Khaw, B.A.; Gold, H.K.; Leinbach, R.C.; Fallon, J.T.; Strauss, H.W.; Pohost, G.M.; Haber, E. Early Imaging of Experimental Myocardial Infarction by Intracoronary Administration of 131 I-Labeled Anticardiac Myosin (Fab')₂ Fragments. *Circulation* **1978**, *58*; 1137–1142.
37. Khaw, B.A.; Petrov, A.; Narula, J. Complementary Roles of Antibody Affinity and Specificity in In Vivo Diagnostic Cardiovascular Targeting: How Specific is Antimyosin for Irreversible Myocardial Damage. *J. Nucl. Cardiol.* **1999**, *6*; 316–323.
38. Khaw, B.A.; Yasuda, T.; Gold, H.K.; Leinbach, R.C.; Johns, J.A.; Kanke, M.; Barlai-Kovach, M.; Strauss, H.W.; Haber, E. Acute Myocardial Infarct Imaging with Indium-111-Labeled Monoclonal Antimyosin Fab. *J. Nucl. Med.* **1987**, *28*; 1671–1678.
39. Khaw, B.A. The Current Role of Infarct Avid Imaging in Cardiovascular Nuclear Medicine-I. *Seminars in Nuclear Medicine*. **1999**, *XXIX* (3); 259–270.
40. Nedelman, M.A.; Shealy, D.J.; Boulton, R.; Bruntm, E.; Seasholtz, J.I.; Allen, I.E.; McCartney, J.E.; Warren, F.D.; Oppermann, H.; Pang, R.H.L.; Berger, H.J.; Weisman, H.F. Rapid Infarct Imaging with a Technetium-99m-Labeled Antimyosin Recombinant Single-Chain Fv: Evaluation in a Canine Model of Acute Myocardial Infarction. *J. Nucl. Med.* **1993**, *34*; 234–241.
41. Calcutt, M.J.; Komissarov, A.A.; Marchbank, M.T.; Deutscher, S.L. Analysis of a Nucleic-Acid-Binding Antibody Fragment: Construction and Characterization of Heavy-Chain Complementarity-Determining Region Switch Variants. *Gene* **1996**, *168* (1); 9–14.
42. Saragovi, H.U.; Fritzpatrick, D.; Raktabutr, A.; Nakanishi, H.; Kahn, M.; Greene, M.I. Design and Synthesis of a Mimetic from an Antibody Complementarity Determining Region. *Science* **1991**, *253*; 792–795.
43. Painter, R.H.; Freedman, M.H. Isolation and Characterization of Electrophoretically Homogenous Rabbit Antihapten Antibody Populations. II. Survey of the Isoelectric Properties of Antihapten Antibodies Directed Against Charged and Uncharged Proteins. *J. Biol. Chem.* **1971**, *246*; 1742–1751.
44. Silva Filho, F.C.; Santos, A.B.S.; de Carvalho, T.M.U.; de Souza, W. Surface Charge of Resident, Elicited, and Activated Mouse Peritoneal Macrophages. *J. Leukocyte Bio.* **1987**, *41*; 143–149.
45. Gallagher, J.E.; George, G. Brody AR. Sialic Acid Mediates the Initial Binding of Positively Charged Inorganic Particles to Alveolar Macrophage Membranes. *Am. Rev. Respir. Dis.* **1987**, *135*; 1345–1352.
46. Yanagisawa-Miwa, A.; Uchida, Y.; Nakamura, F.; Tomaru, T.; Kido, H.; Kamijo, T.; Sugimoto, T.; Kaji, K.; Utsuyama, M.; Kurashima, C.; Ito, H. Salvage of Infarcted Myocardium by Angiogenic Action of Basic Fibroblast Growth Factor. *Science* **1992**, *257*; 1401–1403.
47. Shen, W.-C.; Ryser, H.J.-P. Conjugation of Poly-L-Lysine to Albumin and Horseradish Peroxidase: A Novel Method of Enhancing the Cellular Uptake of Proteins. *Proc. Natl. Acad. Sci.* **1978**, *75*; 1872–1876.
48. Gao, X.; Huang, L. Cationic Liposomes and Polymers for Gene Transfer. *J. Liposome. Res.* **1993**, *3*; 17–30.
49. Torchilin, V.P.; Klibanov, A.L.; Nossiff, N.D.; Slinkin, M.A.; Strauss, H.W.; Haber, E.; Smirnov, V.N.; Khaw, B.A. Monoclonal Antibody Modification with Chelate-linked High-Molecular-Weight Polymers: Major Increases in Polyvalent Cation Binding without Loss of Antigen Binding. *Hybridoma* **1987**, *6*; 229–240.
50. Khaw, B.A.; Klibanov, A.; O'Donnell, S.M.; Saito, T.; Nossiff, N.; Slinkin, M.A.; Newell, J.B.; Strauss, H.W.; Torchilin, V.P. Gamma Imaging with Negatively-Charge-Modified Monoclonal Antibody: Modification with Synthetic Polymers. *J. Nucl. Med.* **1991**, *32*; 1742–1751.
51. Narula, J.; Petrov, A.; Ditlow, C.; Pak, K.Y.; Chen, F.W.; Khaw, B.A. Maximizing Radiotracer Delivery for Scintigraphic Localization of Experimental Atherosclerotic Lesions with High-Dose Negative-Charge-Modified Z2D3 Antibody. *J. Nucl. Cardiol.* **1997**, *4*; 226–233.
52. Narula, J.; Torchilin, V.P.; Petrov, A.; Khaw, S.; Trubetskoy, V.S.; O'Donnell, S.M.; Nossiff, N.D.; Khaw, B.A. In-Vivo Targeting of Acute Myocardial Infarction with Negative Charge, Polymer-Modified Antimyosin Antibody: Use of Different Cross-linkers. *J. Nucl. Cardiol.* **1995**, *2*; 26–34.
53. Carrio, I.; Pieri, P.; Prat, L.; Tison, V.; Pedrini, L.; Moscatelli, G.; Sardi, G.; Estorch, M.; Berna, L.I.; Rimbau, V.; GuJiu, M.; Pak, Ch; Ditlow, Ch; Chen, F.; Khaw, B.A. In-111 Chimeric Negative-Charged-Z2D3 PL-F(ab')₂ in the Detection of Atherosclerotic Plaques. *J. Nucl. Med.* **1995**, *36* (5); 133.
54. Johnson, L.L.; Seldin, D.W.; Tresgallo, M.E.; Bhatia, K.; Rodney, R.A.; Gibbons, J.F.; Esser, P.D. Right Ventricular Infarction and Function from Dual Isotope Indium-111-Antimyosin/Thallium-201 SPECT and Gated Blood Pool Scintigraphy. *J. Nucl. Med.* **1991**, *32* (Abstract); 1018.
55. van Vlies, B.; van Royen, E.D.; Visser, C.A.; Meyne, N.G.; van Buul, M.M.; Peter, R.T.; Dunning, A.J. Frequency of Myocardial Indium-111 Antimyosin Uptake After

- Uncomplicated Coronary Artery Bypass Surgery. *Am. J. Cardiol.* **1990**, *66*; 1191–1195.
56. Narula, L.; Khaw, B.A.; Yasuda, T. Antimyosin Imaging for Acute Myocarditis. *Monoclonal Antibodies in Cardiovascular Diseases*; Khaw, B.A., Narula, J., Strauss, H.W. Eds.; Lee & Febiger: Philadelphia, 1994; 67.
 57. Aretz, H.T.; Billingham, M.E.; Edwards, W.D.; Factor, S.M.; Fallon, J.T.; Fenoglio, J.J.; Olsen, E.G.J.; Schoen, E.J. Myocarditis: A Histopathologic Definition and Classification. *J. Cardiovasc. Pathol.* **1986**, *1*; 3–14.
 58. Frist, W.; Yasuda, T.; Segall, G.; Khaw, B.A.; Strauss, H.W.; Gold, H.K.; Stinson, E.; Oyer, P.; Baldwin, J.; Billingham, M.; McDougall, R.; Haber, E. Noninvasive Detection of Human Cardiac Transplant Rejection with in Antimyosin (Fab) Imaging. *Circulation* **1987**, *76*; V81–85.
 59. Ballester, M.; Obrador, D.; Carrio, I.; Caralps-Riera, J.M. Indium-111 Monoclonal Antimyosin Antibody Studies in the Diagnosis of Rejection and Management of Patients After Heart Transplantation. *Monoclonal Antibodies in Cardiovascular Disease*; Khaw, B.A., Narula, J., Strauss, H.W., Eds.; Lee & Febiger: Philadelphia, 1994; 79–98.
 60. Ballester, M.; Obrador, D.; Carrio, I.; Caralps-Riera, J.M. Indium-111 Monoclonal Antimyosin Antibody Studies in the Diagnosis of Rejection and Management of Patients After Heart Transplantation. *Monoclonal Antibodies in Cardiovascular Diseases*; Khaw, B.A., Narula, J., Strauss, H.W., Eds.; Lee & Febiger: Philadelphia, 1994; 81.
 61. Ballester, M.; Obrador, D.; Carrio, I.; Moya, C.; Auge, J.M.; Bordes, R.; Marti, V.; Bosch, I.; Berna, L.; Estorch, M.; Pons, G.; Camara, M.L.; Padro, J.M.; Aris, A.; Caralps, J.M. Early Postoperative Reduction of Monoclonal Antimyosin Antibody Uptake is Associated with Absent Rejection Related Complications After Heart Transplantation. *Circulation* **1992**, *85*; 61–68.
 62. Chow, L.H.; Radio, S.J.; Sears, T.D.; McManus, B.M. Insensitivity of Right Ventricular Biopsy in the Diagnosis of Myocarditis. *J. Am. Coll. Cardiol.* **1989**, *14*; 915–920.
 63. Becker, A.E.; Heijmans, C.D. Essed CE. Chronic Non-ischemic Heart Disease and Endomyocardial Biopsies. Worth the Extra. *Eur. Heart. J.* **1991**, *12*; 218–223.
 64. Carrio, I. Diagnosis of Doxorubicin Cardiotoxicity with Indium-111 Antimyosin Fab. *Monoclonal Antibodies in Cardiovascular Diseases*; Khaw, B.A., Narula, J., Strauss, H.W., Eds.; Lee & Febiger: Philadelphia, 1994; 99–108.
 65. Narula, J.; Malhotra, A.; Yasuda, T.; Talwar, K.K.; Reddy, K.S.; Chopra, P.; Southern, J.F.; Vasan, R.S.; Tandon, R.; Bhatia, M.L.; Khaw, B.A.; Strauss, H.W. Usefulness of Antimyosin Antibody Imaging for the Detection of Active Rheumatic Myocarditis. *Am. J. Cardiol.* **1999**, *84* (8); 946–950.
 66. Casans, I.; Villar, A.; Almenar, V.; Blanes, A. Lyme Myocarditis Diagnosis by Indium-111 Antimyosin Antibody Scintigraphy. *Eur. J. Nucl. Med.* **1989**, *15*; 330–331.
 67. Krause, T.H.; Schumichen Check, A.; Lang, B.; Hohnloser, S.; Moser, E. Scintigraphy Using ¹¹¹In-Labeled Antimyosin in Churg–Strauss Vasculitis with Myocardial Involvement. *Nuklearmedizin* **1990**, *29*; 177–179.
 68. Hendel, R.C.; Cohen, S.; Aurigemma, G.; Whitfield, S.; Dalberg, S.; Pape, L.; Leppo, J. Focal Myocardial Injury Following Blunt Chest Trauma: A Comparison of Indium-111 Antimyosin Scintigraphy with Other Noninvasive Methods. *Am. Heart J.* **1992**, *123*; 1208–1215.
 69. Spar, I.L.; Goodland, R.L.; Schwartz, S.I. Detection of Preformed Venous Thrombi in Dogs by Means of I-131-Labeled Antibodies to Dog Fibrinogen. *Circ. Res.* **1965**, *17*; 322–329.
 70. Hui, K.Y.; Haber, E.; Matsueda, G.F. Monoclonal Antibodies to a Synthetic Fibrin-Like Peptide Bind to Human Fibrin but not Fibrinogen. *Science* **1983**, *222*; 1129–1132.
 71. Kudryk, B.; Rohoza, A.; Ahadi, M.; Chin, J.; Wiebe, M. Specificity of a Monoclonal Antibody for the NH₂-terminal Region of Fibrin. *Mol. Immunol.* **1984**, *21*; 89–94.
 72. Kanke, M.; Matsueda, G.R.; Strauss, H.W.; Yasuda, T.; Liao, C.S.; Khaw, B.A. Localization and Visualization of Pulmonary Emboli with Radiolabeled Fibrin Specific Monoclonal Antibody. *J. Nucl. Med.* **1991**, *32*; 1254–1260.
 73. Roberts, A.B.; Lees, A.M.; Lees, R.S.; Strauss, H.W.; Fallon, J.T.; Taveras, J.; Kopiwoda, S. Selective Accumulation of Low Density Lipoproteins in Damaged Arterial Wall. *J. Lip. Res.* **1983**, *24*; 1160–1167.
 74. Virgolini, I.; Muller, C.; Fitscha, P.; Chiba, P.; Sinzinger, H. Radiolabeling Autologous Monocytes with in Oxine for Reinjection in Patients with Atherosclerosis. *Prog. Clin. Biol. Res.* **1989**, *355*; 271–280.
 75. Harrison, D.C.; Calenoff, E.; Chen, F.; Parmley, W.; Khaw, B.A.; Ross, R. In Plaque-Associated Immune Reactivity as a Tool for the Diagnosis and Treatment of Atherosclerosis. *Trans. Am. Clin. Climatol. Assoc.* **1992**, *103*; 210–217.
 76. Khaw, B.A.; Calenoff, E.; Chen, F.; O'Donnell, S.M.; Nossiff, N.D.; Strauss, H.W. Localization of Experimental Atherosclerotic Lesion with Monoclonal Antibody Z2D3. *J. Nucl. Med.* **1991**, *32* (5); 1005.
 77. Narula, J.; Petrov, A.; Bianchi, C.; Ditlow, C.C.; Dilley, J.; Pieslak, I.; Chen, F.W.; Torchilin, V.P.; Khaw, B.A. Noninvasive Localization of Experimental Atherosclerotic Lesions with Mouse/Human Chimeric Z2D3 Antibody Specific for the Proliferating Smooth Muscle Cells of Human Atheroma: Imaging with Conventional Antibody and Image Enhancement with Negative Charge-Modified Antibody. *Circulation* **1995**, *92*; 474–484.
 78. Carrio, I.; Pieri, P.L.; Narula, J.; Prat, L.; Pedrini, L.; Riva, P.; Sarti, G.; Pak, Ch; Ditlow, Ch; Chen, F.; Khaw, B.A. In-111 Chimeric Negative-Charged Z2D3 PL-F(ab')₂ Imaging of Proliferating Smooth Muscle Cell in Atherosclerotic Lesions. *J. Nucl. Med.* **1996**, *37* (5); 49.
 79. Khaw, B.A.; Carrio, I.; Pieri, P.L.; Narula, J. Radionuclide Imaging of the Synthetic Smooth Muscle Cell Phenotype in Experimental Atherosclerotic Lesions. *Trends Cardiovas. Med.* **1996**, *6*; 226–232.
 80. Majno, G.; Joris, I. Apoptosis, Oncosis, and Necrosis: An Overview of Cell Death. *Am. J. Pathol.* **1995**, *146* (1); 3–15.
 81. Kerr, J.F.R.; Wyllie, A.H.; Currie, A.R. Apoptosis: A Basic Biological Phenomenon with Wide Ranging Implications In Tissue Kinetics. *Br. J. Cancer.* **1972**, *26*; 239.
 82. Jennings, R.B.; Schaper, J.; Hill, M.L.; Steenberg, C., Jr.; Reimer, K.A. Effect of Reperfusion Late in the Phase of Reversible Ischemic Injury: Changes in Cell Volume, Electrolytes, Metabolites, and Ultrastructure. *Circ. Res.* **1985**, *56*; 262–278.
 83. Bailik, S.; Geenen, D.L.; Sasson, I.E.; Cheng, R.; Horner, J.W.; Evans, S.M.; Lord, E.M.; Koch, C.J.; Kitis, R.N. Myocyte Apoptosis During Acute Myocardial Infarction in

- the Mouse Localizes to Hypoxic Regions but Occurs Independently of p53. *J. Clin. Invest* **1997**, *100*; 1363–1372.
84. Khaw, B.A.; Torchilin, V.P.; Vural, I.; Narula, J. Plug and Seal: Prevention of Hypoxic Cardiocyte Death by Sealing Membrane Lesions with Antimyosin-Liposomes. *Nature Med.* **1995**, *1*; 1195–1198.
85. Weissig, V.; Lasch, J.; Klibanov, A.L.; Torchilin, V.P. A New Hydrophobic Anchor for the Attachment of Proteins to Liposomal Membranes. *FEBS Lett* **1986**, *202*; 86–90.
86. Torchilin, D.L.; Klibanov, A.L.; Huang, L.; O'Donnell, S.; Nossiff, N.D.; Khaw, B.A. Targeted Accumulation of Polyethylene Glycol-Coated Immunoliposomes in Infarcted Rabbit Myocardium. *FASEB J.* **1992**, *6*; 2716–2719.
87. Khaw, B.A.; Vural, I.; Narula, J.; Haider, N.; Torchilin, V.P. Cardiocyte Viability by Immunoliposome-Cell Membrane Sealing at 1, 2, 3, 4 and 5 Days of Hypoxia Proceedings of 23rd International Symposium on Controlled Release of Bioactive Materials. Kyoto, Japan 1996, 617, 618.
88. Khaw, A.B.; Vural, I.; DaSilva, J.; Torchilin, V.P. In Vivo Myocardial Preservation with Cytoskeleton Specific Immunoliposomes in Acute Myocardial Infarction Proceedings of the 27th International Symposium on Controlled Release of Bioactive Materials, in press.
89. Khaw, B.A.; Vural, I.; DaSilva, J.; Torchilin, T.P. Use of Cytoskeleton-Specific Immunoliposomes for Preservation of Cell Viability and Gene Delivery. *STP Pharma Sci.* **2000**, *10* (4); 279–283.
90. Khaw, B.A.; Vural, I.; Torchilin, V.P.; Haider, N.; Narula, J. Expression of Antimyosin sFv Gene in Cardiocytes: Use of Cytoskeleton-Specific Immunoliposomes for Transfection Proceedings of 23rd International Symposium on Controlled Release of Bioactive Materials Kyoto Japan 1996, 135, 136.
91. Khaw, B.A.; Vural, I.; Torchilin, V.P.; Narula, J. Intracytoplasmic Delivery of Plasmid pGL2 Vector by Cytoskeleton Specific Immunoliposome Transfection Proceedings of the 25th International Symposium on Controlled Release of Bioactive Materials 1998; 25, 178, 179.
92. Khaw, B.A.; Vural, I.; Narula, J.; Torchilin, V.P. Comparative Efficiencies of Transfection by Cationic Liposomes and Cytoskeleton Specific Immunoliposomes Proceedings of the 26th International Symposium on Controlled Release of Bioactive Materials 1999; 26, 218, 219.
93. Felgner, P.L.; Ringold, G.M. Cationic Liposome-Mediated Transfection. *Nature* **1989**, *337*; 387–388.

TECHNOLOGY TRANSFER CONSIDERATIONS FOR PHARMACEUTICALS^a

Ira R. Berry

Duramed Pharmaceuticals, Inc., Somerset, New Jersey, U.S.A.

INTRODUCTION

Technology transfer for pharmaceutical products is a program that has been followed for some time (1–11). Issues to be considered when organizing the transfer of technology from the research arena to the production and quality assurance environments are reviewed in this article. The discussion focuses on the coordination and implementation of a transfer program for a product, with emphasis given to those factors special to the pharmaceutical industry. The success of any program is highly dependent on the effectiveness of the communication preceding its implementation; therefore, the preparation and distribution of a complete document summarizing raw material and equipment requirements, manufacturing and packaging processes, process validation parameters, quality control procedures, and safe handling procedures—as well as a detailed plan of action outlining expected results and time frames—must be distributed before the scale-up experience. Input from the marketing and manufacturing centers must be integrated into the plan to ensure that the right product is developed at the right price within the desired time frame. An outline encompassing these critical aspects of a transfer program is presented.

Whether a tablet, a transdermal patch, a topical ointment, or an injectable, the transformation of a pharmaceutical prototype into a successful product requires the cooperation of many individuals. To complete the task efficiently, the transfer of a product from the research and development area to production must be organized. Planning for process commercialization is one area where tangible rewards can be realized. The successful transfer of a project to a production site from the research arena does not happen on its own. Organizing the transfer of a technology or new product from research to production may be one of the most perplexing problems that development scientists, engineers, and marketers may encounter during their careers. This article provides some

insight into the issues that should be considered during the transfer program and offers a sequence of events toward completing the task.

A major decision focuses on that point where the idea or process is advanced from a research-oriented program to one targeted toward commercialization. Generally, the cost of product development rises dramatically during the pilot scale-up and initial production batch efforts. In other words, the critical path for success is dependent on the completion of the technology transfer to the production site at an affordable cost.

The three primary considerations to be addressed during an effective technology transfer are the plan, the persons involved, and the process. A plan must be devised to organize the personnel and the process steps. Once prepared, the plan must be communicated to the involved parties in research, at the corporate level, and at the production site. The success of any program is highly dependent on the effectiveness of the communication preceding its implementation. Therefore, identifying the parties involved in the development process is one of the most important tasks to be confronted and must be completed early in the transfer process.

PERSONNEL

The proper personnel must be informed of their involvement, desired contributions, and responsibilities. This helps identify potential problem areas that may hinder the accomplishment of the challenge at hand. It is desirable to appoint a project leader or liaison from the research and development group as the focus of the communication pattern. This individual has the responsibility to coordinate the assembly of the necessary information to support the product's advancement for process development. The trap to avoid here is to assign someone to this role and not give him or her the authority commensurate with the responsibilities with which he or she has been charged. The practice of pairing a seasoned project manager as a mentor with a less-experienced future

^a(Revised from Ref. 1.)

project coordinator is recommended. In this manner, team players can be nurtured. The communication and networking skills of the successful manager can be shared across the organization. On the other hand, it is expected that individuals who do not possess the necessary talents to be effective are identified before they encounter the stressful world of project management alone and, more importantly, before they have a chance to contribute to the failure of a potentially successful product.

Through mentoring, the one-to-one contact offers the unique opportunity for the sharing of ideas, skills, and observations. If properly organized with two-way evaluations, both participants should benefit. Not all individuals, however, are suited for the role of mentor. Care should be exercised in the selection and pairing of mentors for the less experienced. Perhaps a human resource manager could assist in identifying those people with the interpersonal skills required for this teaching position.

Information from the product development area would be gathered from the formulator, analytical and microbiology testing groups, and the packaging development unit. Issues to be considered from these groups will be discussed below. A safety evaluation from the toxicology group and industrial hygienist should be completed before a scale-up effort. In today's regulatory climate, a concise and understandable summary must be provided regarding the risks associated with and procedures for proper handling of all chemical substances, whether drug actives or excipients. Failure to provide sufficient information to those not skilled in the art to make the decision to initiate a handling, weighing, or processing operation may lead to an unfortunate employee injury. Although an employee has a "right to know," employers have an obligation to provide the necessary warnings and training to minimize placing an employee at risk. The legal ramifications of improper training or notification are outside the scope of this article.

An opinion should be solicited from the patent department or a patent attorney. With the implementation of the General Agreement on Tariffs and Trade (GATT) accord, patents take on a new meaning. The impact of this legislation is far reaching, not only in the United States, where major revisions in the patent laws have been required and implemented, but also in many foreign countries where patents may be essentially worthless or not enforceable. In general, patents have an effective life of 20 years from the date of initial filing. Additional periods of exclusivity may be allowed, on application and when certain criteria are met, that extend the effective term of a patent. In summary, care must be exercised to ensure that proper legal protection of the novel concept has been secured in those areas where desired.

The drug regulatory affairs unit must also be involved with the product transfer to specify "how much" of "what" information is needed either to submit for a drug approval or to introduce a product to the marketplace. The goal here is to collect the proper amount and kind of data necessary to support the prerequisite filings, either internal or external to the company.

The corporate office commonly involves personnel from the production planning unit, engineering group, and new product coordination section, as well as from marketing, because each of these divisions has a vested interest in the success of the venture. It is important to identify the needs of those people at the corporate level to minimize delays caused by the "we did not know that we had to do this now" club. Activities such as product label preparation, copy for advertisements and promotional pieces, and package graphics may need to be initiated during the transfer program as not to lose valuable lead times. The acceptance or approval process in some companies is very labor-intensive and therefore time-consuming. This is one area where a detailed time and event plan has been shown to yield a significant impact on timely completion of a project.

Keeping the corporate participants informed becomes a pivotal task for the project manager in obtaining the final acceptance of the product by the corporation. Therefore, any time taken with those people involved at the corporate level to explain the steps involved in the transfer program is time well spent. Like any educational process, consistency and repetition enforce learning. It is incumbent on management to foster cooperation among individuals at the functional research and corporate centers, because the loss of time is the worst enemy an organization can face.

There are a number of individuals at the manufacturing site who must work as a team to ensure the timely and efficient completion of this transfer effort. The plant manager, technical director, production planning group, manufacturing area supervisor, and quality control and quality assurance units—as well as the plant engineering, packaging and transportation supervisors—must be informed as to their responsibilities. Personnel training must also be considered if the technology is new to the site. Last but not least are the contributions that the line mechanics and chemical operators can make to the program. They perform the necessary production functions daily. Their insight and practical experience are an invaluable resource that should not be overlooked.

The success of the transfer is dependent on the ability of the project leader to motivate employees to work toward a mutually beneficial goal, namely, introducing a new product that will improve consumer health and create jobs and increased profit for the company.

MOTIVATING PARTICIPANTS

Implementation of a positive return on involvement is one mechanism by which changes in responsibilities and tasks can be implemented. Changes here include the manufacture of new products or using new procedures to produce existing products. By involving personnel in the planning of changes, in discussing the resources necessary to complete the tasks, and in creating an environment in which innovation can thrive, the project manager should strive to realize the completion of the program quickly and efficiently with reliable quality at an affordable cost. At the same time, those who have worked to bring the program to completion should share a feeling of teamwork and a sense of accomplishment. Return on involvement encompasses the philosophy that employees are a key element in the successful introduction of any technology or product approval. Project managers must keep this in mind because they frequently depend on the cooperation of individuals outside their direct control to accomplish their goals.

PRETRANSFER CONSIDERATIONS

Several assumptions must be satisfied before the advancement of any product to plant scale-up trials. First, the marketing division should have examined the proposed product prototypes and agreed that the product meets their needs. Second, the intended commercial package configurations should have been selected. Although it is not uncommon to package portions of the first scale-up batch in a variety of formats, it is incumbent on the project leader to eliminate unnecessary packages to minimize the dilution of effort. Third, any constraints, such as cost or time, must be identified so that they may be given due consideration.

There is an important question to be asked, i.e., does the product meet the needs of the consumer? The development staff believes they have captured marketing's vision with their product offering; however, the marketing and sales units must concur. The decision is generally in the sales unit, because those people are responsible for making the product available to the consumer and, more important, they are responsible to see that the product meets the consumer demands, real or perceived. It is expected that any pertinent focus groups or market research studies would be completed before the scale-up effort. These studies help confirm the product concept and its viability in the marketplace.

The cost of a development program increases dramatically as the number of package configurations to

be advanced to commercialization increases. Selecting the proper package sizes, closures, colors, and neck finishes is compromised by the package composition and availability, and, finally, the intended use and cost of the unit. For example, a smaller unit lasting 1 month, such as a calendar pack, makes marketing sense for oral contraceptives for several reasons. Dispensing and unit sales are generally cyclic and more predictable, allowing for better profit projections. The calendar pack and monthly cost to the consumer not only position the product toward enhanced consumer acceptance but also toward better patient compliance, especially with expensive medications. In contrast, inhalation aerosol units, because of their pattern of chronic use, frequently contain sufficient doses to last several months. Suffice it to note, it behooves the marketing unit to work with financial analysts to determine the optimal product configuration and cost profile that maximizes consumer acceptance and convenience, unit turns, profits, and resources.

Constraints always exist but occasionally are not communicated accurately or promptly. Competition in the marketplace frequently causes introduction deadlines and endproduct cost constraints that must be considered during the initial phases of a development program. On the other hand, lead times for materials and personnel resources must be appreciated when commitments are made to timelines. Without planning in advance, the task may not be possible at all. Factors that influence decisions may originate externally as well as internally. Care must be exercised to address those issues that affect the timely completion of a project.

Physicochemical properties of raw materials and the finished dosage form should be characterized before any scale-up effort. Having methodology available and validated to compare batches is essential. For example, drug-release profiles and viscosities have the potential of being altered in scale-up by manufacturing procedures and equipment. Care must be exercised to maintain the desired profiles and other product specifications. The effect of batch size and process scale-up should be monitored closely.

Formulation and/or development of advanced drug-delivery systems such as microencapsulated molecules, transdermal patches, or liposomes are frequently accomplished in the laboratory. However, large-scale production of these dosage forms may be problematic because the same conditions of manufacture may not be attainable or desirable in the plant setting. Consultation with process development personnel during the finalization of the prototype development phase is one way of minimizing scale-up difficulties.

An area occasionally overlooked by the development staff is the necessity of securing confidentiality

agreements from vendors supplying technologies or services to a firm. All contractors should be required to execute a confidentiality agreement that specifically encompasses the technology and product being developed. These documents should be prepared, reviewed, and

executed by the appropriate legal and executive officers of both organizations. Especially when the science or product is not well-defined and patent protection has not been secured, such as with the development of novel, specialized drug delivery systems or new chemical

Activity	Completion Dates	
	Target	Actual
1. Formulation selected	_____	_____
2. Site for plant trial established	_____	_____
3. Planning meeting scheduled	_____	_____
a. Review of development report	_____	_____
b. Manufacturing formulation/procedures	_____	_____
b. Handling and safety issues	_____	_____
c. Raw material specifications and suppliers	_____	_____
d. Packaging procedures	_____	_____
e. Packaging component specifications and suppliers	_____	_____
f. Testing procedures and validation	_____	_____
g. Process validation protocols	_____	_____
4. Date of plant trial established	_____	_____
a. Manufacturing	_____	_____
c. Packaging	_____	_____
d. Quality assurance testing	_____	_____
c. On-site review of experience	_____	_____
d. Shipment of samples to R&D for testing	_____	_____
5. Date of shipment delivery	_____	_____
a. Confirmation of results	_____	_____
b. Stability initiation	_____	_____
c. Product evaluation	_____	_____
Safety	_____	_____
Efficacy	_____	_____
Drug release	_____	_____
6. Postproduction review meeting	_____	_____
7. Assembly of final pilot plant documents	_____	_____
a. Manufacturing	_____	_____
b. Packaging	_____	_____
c. Handling and safety procedures	_____	_____
d. Quality control specifications and analytical methods	_____	_____
e. Stability data	_____	_____
f. Material safety data sheet	_____	_____
g. Shelf-life projection	_____	_____
h. Process validation summary	_____	_____
i. Bibliography	_____	_____
8. Review of validation report	_____	_____
a. Research and Development	_____	_____
b. Manufacturing	_____	_____
c. Engineering	_____	_____
d. Quality Assurance/Quality Control	_____	_____
e. Regulatory Affairs	_____	_____
f. Corporate	_____	_____
g. Marketing	_____	_____
9. Issuance of final product specifications	_____	_____
10. First commercial batch	_____	_____

Fig. 1 Project technology transfer checklist. (Modified from Ref. 1.)

entities, this task must be completed expeditiously, and no work should be initiated until the agreements have been properly executed.

TRANSFER PROGRAM

Any development and technology transfer program should be reduced to a written document such as that shown in Fig. 1. An outline or checklist must be compiled to ensure that appropriate consideration has been given to relevant issues. This helps ensure also that all parties are approaching the task from the same perspective and priority. A manufacturing site must be designated and the appropriate personnel notified as to their involvement. The necessary information must be collected and disseminated to the involved parties. At a minimum, copies of the proposed formula, manufacturing and testing procedures, and safe handling considerations should be distributed to allow sufficient time for review and comment. A planning review session should be convened with representatives from the research, corporate, marketing, manufacturing, quality assurance, regulatory affairs, and engineering departments in attendance. Selection of the time and location of this meeting should be made to encourage maximum participation.

The meeting should be chaired by the project leader. It is his or her responsibility to determine the relevant issues to be discussed, to follow up that an agenda for the meeting is distributed, and to establish that minutes of the meeting are taken. Concerns that arise during this meeting should be noted and addressed because the purpose of the meeting is to draw from the experience of the participants to identify potential problem areas in the program. The tone of this session should be one of consensus and not one of autocratic rule. Motivation, communication, and cooperation must be stressed in the voice and actions of the project leader. This is the first step toward accomplishing the primary program objective, namely, the timely and informed transfer of a new product from the research arena to the production site. The following subjects should be discussed at the planning meeting:

- Validated analytical methods
 - Regulatory considerations
 - Rework procedures
 - Transportation
- Formula—handling and safety considerations
 - Raw materials
 - Manufacturing equipment
 - Manufacturing precautions
 - Manufacturing procedures
 - Packaging
 - Process validation
 - Specifications for raw materials, packaging components, and in-process and finished product

Each aspect should be reviewed to ensure that critical issues have been addressed. If an aspect is not relevant, it should be stated that it is not applicable to the program. An issue perceived as unimportant in one department may be a monumental task in another department. With new drugs and drug-delivery systems, this effort is critical to the success of a program.

FORMULA

Understanding the formula, its derivation, and its constraints is one of the first prerequisites to any development program. The feasibility of the formula may be established by reviewing the ingredients of the composition and explaining their function in the formula. It would be appropriate to review the claims and physical characteristics of the product while evaluating a sample so that the participants of a planning meeting appreciate the appearance and attributes of the product.

The safety evaluation completed by the toxicology group and/or industrial hygienist should be reviewed with the participants because they, as managers, will most likely assume the responsibility to protect the safety of employees who will work on this project. The employees' right to know must be protected further by keeping them informed of the potential risks to which they may be exposed.

Prerequisite safety information can be transmitted by draft material safety data sheets, especially for new drug actives and new drug products. Preparation of the draft document may lead to the question: Are the process handling procedures for this product necessary and appropriate? Those personnel involved in the manufacturing of drug products are cognizant of the concepts of inherent risk due to an agent's toxicity, and potential risk, due to exposure. Whether the new drug and/or product should be handled in an open environment, contained area, or in isolation must be determined by those people who are responsible for safety. In all cases, the procedures must be reviewed and approved by the individual responsible for the involved manufacturing and testing sites.

Comments about the prestability and finished product stability profiles should be presented as an overview of the new product's chemical stability. This will support the anticipated shelf life of the formula and its ability to withstand the "process shocks" normally encountered

during production scale-ups. The toxicity of the finished product should be discussed. This information ensures that a decision regarding the handling of the formula has been made based on generated data and experience. A draft or tentative material safety data sheet may be one route to disseminate this information.

Constraints on specific ingredients or sources of ingredients, cost of goods, or manufacturing equipment should be reviewed. Constraints must be considered when formula optimization is undertaken; however, optimizing formulas may be best addressed in the production environment because batch size and manufacturing equipment sometimes have been shown to render viable laboratory and small-scale formulas virtually inoperative in the plant. The experience gained during the manufacture of laboratory and scale-up batches is invaluable and should be shared with the participants, especially the production and quality assurance staffs, in written reports and follow-up meetings.

Many firms use a laboratory development report to record these activities. These data can be incorporated into a full project-development history. For new drug products, development histories are needed to fulfill regulatory directives. The detail necessary in any summary report depends greatly on the magnitude of the problems encountered during development and the corporate structure in which one lives. For some multinational companies, the product-development report serves as the basis of spreading interest in a new product across global borders.

RAW MATERIALS

Sources of raw materials, especially those critical to a new product's functionality, should be identified. Availability and costs should be ascertained to aid in the planning process. Care should be taken to ensure that new drug actives and, when possible, excipients are secured from vendors with a current acceptable FDA compliance profile and a drug master file.

Testing monographs including methods to ascertain a lot's chemical and, if necessary, microbiological integrity should be provided to the selected manufacturing site in advance so that the methods may be applied to the incoming supplies. Handling of materials including storage, disposal, and employee precautions should be documented, especially for new or potentially hazardous materials. Again, material safety data sheets (MSDS) for all materials used should be available and disseminated to the plant personnel before any exposure.

MANUFACTURING

Equipment

When new drugs and drug-delivery systems are developed in the laboratory, the correlation of the necessary production equipment may be very difficult indeed. For example, the shear needed to create the desired particle size of an emulsion with the help of laboratory equipment may pose serious problems in the selection of plant equipment necessary to reproduce the attributes of the product. Recording the speed of a laboratory mixer is not sufficient by itself for this task; definition of the operating principle and equipment design is necessary to accomplish the task.

Any equipment used in drug manufacturing, including packaging, should have undergone an equipment evaluation, including installation, operational, and performance qualification, following a written protocol. This is an important step in a process-validation program. Through this effort, the operating parameters and capabilities of a given piece of equipment are documented. Furthermore, should a major component of the equipment fail, installation of a replacement part of known specifications reduces and in many cases eliminates the need for revalidation of every product processed with that equipment. The equipment itself, however, must be shown to meet its previous operating capabilities before being placed back into service.

The availability, size, and surfaces or composition of the required equipment should be specifically identified so that the scale-up effort may be representative of a production run. A preliminary compatibility screen of contact surfaces should be completed before the selection of scale-up equipment. The location of the equipment in reference to other requirements, such as services or the packaging area, may be a factor in the selection of equipment. A cleaning-validation study should be conducted to ensure that no residues of active ingredient or cleaning agent remain after cleaning and that the equipment is suitable for production use again. Alternative equipment may be considered and used; however, experience will dictate its suitability.

Precautions

Any concerns regarding the handling of equipment or product by employees should be addressed in the planning stage. This is especially important with regard to environmental (particulate contamination or sterility) or atmospheric (oxygen, moisture, or light sensitivity) problems.

Procedures

Procedures should be clear and concise. Specific descriptions should be used when possible. For example, “pass the emulsion through a suitable colloid mill (Eppenbach mill) at a setting of 0.005 inch (0.12 mm)” is preferred to a description referencing a more general description of a piece of equipment. Process-validation testing is necessary using specific equipment, as described by design and operating principle.

Procedures should be realistic, and any instructions must be scale-oriented. Specific parameters may be necessary for manufacturing areas. For example, cooling or heating times are typically equipment-dependent. Cooling 1 kg in the laboratory in 15 min may take 4 hours for a 40,000-kg batch in the plant. Similarly, filtration of small batches in the laboratory may not provide the necessary information to predict filter life or flow rates needed for large-scale manufacturing.

Based on the experience gained during the pilot scale-up effort, a process flow chart should be constructed. It helps identify steps and issues in need of process-validation review. In addition, the timing of activities toward the scheduling of manpower needs, such as for in-process testing, is generally more apparent when viewed in the context of the total process.

Process-optimization parameters as identified during the pilot scale-up effort should be monitored during the production scale-up batch. In this manner, appropriate recommendations based on experience may be integrated into the production of future batches. Many optimization experiments may be efficiently incorporated into the process-validation program.

Packaging

The description, specifications, and test methods for any packaging configuration should be available to the plant before the production scale-up. Unit functionality and fit should be included as a practical use test in any specification. The plant equipment to be used in packaging should be evaluated for feasibility, speed, and contact surface compatibility. Preliminary evaluations of surface compatibility, discussed previously, should suffice as an early indication of packaging equipment suitability.

The availability of or lead time to secure the necessary packaging components that are representative of the commercially marketed package frequently places stress on the project timeline. Package costs and possible alternative packaging can be evaluated with bulk produced from this batch. Therefore, a course of action to minimize project failure resulting from an unsatisfactory package

may be appropriate. The resultant dilution of resources and increased project expense must be weighed in accepting this course of action.

The number of various package sizes to be filled from a batch may be critical. For example, for generic drugs, entire batches must be filled for the batch to be accepted by the FDA. Although the batch may be filled into numerous package formats, care must be exercised to fill a sufficient number of each format to ensure proper equipment set-up and that the filling speeds used are representative of a full-scale production effort.

Procedures for packaging the batch including fill tolerances and precautions such as aseptic handling or nitrogen gassing should be reviewed to state the requirements for acceptance in advance. As a part of this production scale-up effort, it may be desirable to evaluate the product's bulk stability in the storage tank or storage drums to establish limits on the length of time a batch may be held before final packaging. This is especially important if the bulk product is to be manufactured at one site and transported to another site for packaging. Storage container compatibility deserves appropriate attention also.

Finally, the personnel involved in packaging should be instructed on any safety and handling issues that might affect them or compromise the product's integrity. Whether it is for the development of new package formats, such as for intranasal drug administration or transdermal patches, or for more traditional delivery systems, such as cycle packs, solutions, or aerosols, the need to educate employees involved in the processing is essential to the transfer program's success. Identifying and controlling process variables are necessary while experience is gained, and the process is optimized and validated.

PROCESS VALIDATION

Each class of product has specific issues to be addressed for process validation. In general, variable process steps such as mixing times and temperatures should be validated. Many articles have been written regarding the validation of processes affecting pharmaceutical products (12–14). Protocols to evaluate those parameters that may affect a product's integrity should be agreed on by the R&D, regulatory, production, QA, and engineering staffs. During the preparation and packaging of the pilot production batch, generation of data toward improving the efficiency of these processes, as well as minimizing batch-to-batch variations, is very important because this information will serve as documentation to support the new product's commercial feasibility. Also, the

establishment of cleaning procedures and documentation of cleaning validation can be accomplished at this time.

There is no magic number of batches required to prove that a process is validated. Generally, the number of batches accepted is three. The technical complexities of and product sensitivities to variable parameters dictate how extensive a validation program is needed. Suffice it to note that the process must be controllable and reproducible and yield a product that meets the desired specifications.

QUALITY CONTROL AND QUALITY ASSURANCE

One of the purposes of the pilot production batch is to introduce the new product in its entirety to the functional areas of the site (manufacturing, packaging, and testing), that is, the release of raw materials and packaging components as well as in-process, bulk, and finished product testing should be completed at the site as if the pilot scale-up were a commercial batch.

In-process testing, bulk release before packaging, and finished product specifications are proposed. Specification limits are proposed based on the experience gained with smaller laboratory and scale-up batches. Any comments or concerns regarding the test methods and specifications should be addressed at this time. Reagents and equipment to complete the required testing must be available at the plant. A contact at the R&D analytical laboratory should be established to explain aberrant values or to answer questions.

Communication of requirements of time and manpower to the quality control department is a critical issue that must not be overlooked. Prompt attention to analysis needs does help keep the pilot production batch process moving forward. If microbiological release of bulk product is needed before packaging, the project timeline should reflect the time period (3–5 days) generally needed for this activity.

Sampling must be scheduled for release and stability testing using a statistically valid sampling program. This is especially important for stability studies. In this manner, the chosen samples are documented to be representative of the entire batch.

Batch documentation is an important factor. Preparation of master batch records in accordance with plant standard operating procedures (SOPs) should be followed by an approval of the document by the sponsoring division, usually the formulator or process development staff of the R&D unit. On completion of a batch, review of the batch records by the quality assurance group ensures compliance to GMP and that all necessary deviations from

and modifications to the manufacturing records are properly explained and documented.

THE MANUFACTURING TRIAL SCHEDULING DATE

At the conclusion of the planning meeting, any actions that must be undertaken before the scale-up should be documented and a responsible party identified. At the discussion of the trial date, time constraints must be considered, along with the availability of raw materials, packaging components, plant scheduling time, and personnel. Coordination of personnel and supplies is the responsibility of the project leader. The ability to lead and negotiate another individual's priorities helps bring the trials to completion on schedule.

Finally, arrangements for transporting raw materials and packaged finished products must be made to ensure that the scale-up effort is completed on schedule and that stability studies are initiated expeditiously. Participants should leave the meeting under the impression that one person is in charge of the project, that the program has been well thought out and documented, and that commitments will be honored. Early issuance of meeting minutes will reinforce the importance of individual responsibilities and serve as notice to the participants and their superiors that their cooperation has been solicited, is needed, and is expected.

COMPLETION

In review, the activities to be completed at the manufacturing site are:

- Release of raw materials and packaging components
- Manufacture and packaging of the trial batch
- Generation of data from in-process, bulk, and finished product samples
- Process validation, including equipment qualification and reviews of batch records, processing, cleaning validation, and on-site experiences
- Shipment of finished product to the research facility for testing

An exit interview with involved plant personnel offers an opportunity for their comments to be heard. Their efforts should be acknowledged and their input seriously considered and incorporated into the manufacturing document. Any differences that cannot be resolved at this time should be noted and studied further. The art of

listening and diplomacy must be used inasmuch as this forum must be one of cooperation and not one of confrontation.

At this meeting, a discussion about possible rework procedures may be appropriate. The ability to recover materials, especially expensive drug actives, is desirable. Early identification of steps where rework may be possible allows for procedures to be tested, verified, and put in place, should they be needed. For pharmaceuticals, rework procedures may be used only if they are appropriately documented, validated, and approved by the responsible corporate and government bodies. Rework procedures are not an automatic means for handling out-of-specification product lots but rather for identifying where an effort may be implemented successfully. Logistics and economics, as always, dictate whether a rework should be considered.

POSTPRODUCTION ACTIVITIES AND EVALUATION

Once the plant experience has been concluded, confirmatory analyses on duplicate samples for in-process, bulk release, and finished product previously tested at the plant site should be completed at R&D. In this manner, the proposed methods are challenged to yield similar results from different analysts in different locations. Discrepancies in values generated at this point must be investigated and resolved. Samples should be placed into the stability testing system according to the organization's procedures.

Finally, a report must be prepared and issued expeditiously summarizing the experience, reviewing each area's involvement, and proposing, if necessary, changes in the process or control methods. Timely and factual communication of project progress to the other corporate areas not directly involved in the scale-up, such as the marketing, finance, and purchasing units, draws their attention and commitment of resources. By fostering informed decision-making through directed written communication, the time required to plan or complete these activities to bring the product to market is minimized, and resource usage is optimized.

A formal postproduction trial review meeting should be held with representatives from the R&D, corporate, marketing, quality assurance, regulatory affairs, and production centers. Plans to commercialize the product or to submit documentation for government approval if this is the next step in the development scheme are outlined, contingent on the successful completion of a defined stability program. Agreement as to the suitability of all the factors involved in the preparation of the product

should be the result of this meeting, with a substantiating document in the form of meeting minutes or a signed "statement of concurrence" generated and distributed.

A monograph of all pertinent sections to support the product's introduction should be assembled, reviewed, and disseminated to the appropriate parties. This document should include:

- Manufacturing formula
- Draft label copy
- Raw material tests and specifications
- Manufacturing procedure
- In-process tests and specifications
- Finished product test methods and specifications
- Packaging component specifications and drawings
- Packaging component test methods
- Stability data on bulk product
- Stability data on packaged product
- Shelf-life projection of expiration dating
- Material safety data sheet for the product
- Bibliography

A projection of shelf life should be included to document the recommendation for the expiration date. Although accelerated stability data are frequently used to support expiration dating for up to 24 months, extension of dating beyond 24 months is based on real-time test results.

A bibliography of all project reports and memos should be assembled, including the development and validation reports. Ideally, the project leader should maintain a file in chronological order of all communications and reports. Compiling the file into a bibliography is a tedious task but one that, should the need to answer a question arise, will be well worth the effort. Product development histories and testing "quirks" regularly appear in these documents and in no other place. Skilled project leaders maintain an ongoing listing as a means of tracking a project. A bibliography for a new pharmaceutical product may be voluminous, and a reference to its location may be useful. The transfer of the project is considered complete when the first commercial batch is produced under the supervision of the manufacturing site without problems.

SUMMARY

This article has reviewed issues to be considered when organizing the transfer of pharmaceutical technology from the research arena to the production environment. Critical areas affecting the manufacture, packaging, safety, and quality of pharmaceutical products are discussed in relationship to their impact by the technology transfer

process. The necessity of a plan with input from the various organizational centers is emphasized. The success of the program is highly dependent on the communication and cooperation shared throughout the process.

REFERENCES

1. Popp, K.F. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1996; 14, 419–432.
2. Amico, L.A.; Pilot Plant Operation. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C. Eds.; Marcel Dekker, Inc.: New York, 1995; 12, 187–207.
3. Franz, R.M.; Copeland, R.D.; Lewis, L.D.; Stagner, W.C. Pilot Plant Design. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1995; 12, 171–186.
4. Groen, J.J. Project Management. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1995; 13, 121–150.
5. Owen, V.M. Technology Transfer in the Diagnostic Industry. *Med. Dev. Technol. (Lond)* **1994**, 5, 23–26.
6. Popp, K.F. *Drug Dev. Ind. Pharm.* **1987**, 13 (13), 2339–2362.
7. Popp, K.F. Organizing the Transfer of Pharmaceuticals from Research to Production. *Specialized Drug Delivery Systems—Manufacturing and Production Technology*; Tyle, P., Ed.; Marcel Dekker, Inc.: New York, 1990; 37–50.
8. Rao, A.V.; Rajan, J.V. Transfer of Drug Technology from Laboratory to Industry. *Eastern Pharmacist (India)* **1985**, 28, 63–65.
9. Testa, E.G.; Lepiti, S. Product Technology Transfer. *Pharmaz. Ind. (Germany)* **1981**, 43 (12), 1231–1234.
10. Upupa, N. Research, Product Development, and Pilot Plant Scale Up. *Pharmacy Times (India)* **1990**, 22, 63–65.
11. Evanoff, B.J.; Hofmann, K.L., Jr. *Validation of Active Pharmaceutical Ingredients*, 2nd Ed.; Berry, I.R., Harpaz, D., Eds.; Interpharm Press: Englewood, CO, 2001.
12. Berry, I.R. Process Validation: Practical Applications to Pharmaceutical Products. *Drug Dev. Ind. Pharm.* **1988**, 14 (2 & 3), 377–389.
13. Berry, I.R.; Nash, R.A. *Pharmaceutical Process Validation*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1993.
14. Berry, I.R.; Harpaz, D. Eds.; *Validation of Active Pharmaceutical Ingredients*, 2nd Ed.; Interpharm Press: Englewood, CO, 2001.

THERMAL ANALYSIS OF DRUGS AND DRUG PRODUCTS

Daniëlle Giron

Novartis Pharma AG, Basel, Switzerland

INTRODUCTION

Thermal analysis techniques, in which a physical property is monitored as a function of temperature or time while the analyte is heated or cooled under controlled conditions, are fundamental techniques for the characterization of drugs and drug products, not only while processing or aging conditions may be simulated but while the methods gives access to thermodynamic data. Due to the different informations delivered, thermal analysis methods are concurrent or complementary to other analytical techniques, such as spectroscopy, chromatography, melting, loss on drying, assay, for identification, purity, and quantitation. They are basic methods in the field of polymer analysis and in physical and chemical characterization of pure substances as well as for mixtures. They find good applications for preformulation, processing, and control of the drug product. The introduction of automation considerably increases the advantages of these methods. New horizons are open with the availability of combined techniques and microthermal analysis.

PRINCIPLES AND EXPERIMENTAL FACTORS

Considering the number of physical parameters of a substance that may be measured, the number of techniques derived is very large. Details on most techniques are well described by Wendlandt (1). For pharmaceutical applications, the methods generally used are differential scanning calorimetry (DSC), thermogravimetry (TG) (or thermogravimetric analysis: TGA), and, to a lesser extent thermomechanical analysis (TMA). All techniques are automated and have data acquisition. Hyphenated techniques and modulated DSC are growing techniques, “state of the art” for the 21st century. Excellent books or review articles dealing with the principle, instrumentation, and applications of thermal analysis methods for pharmaceuticals are given in (1–15). As emphasized by Cheng et al. (14), tendency in the next two decades will be more precise and meaningful measurements in these techniques and new developments in obtaining

the temperature dependence of a material’s structure and dynamics.

Differential Scanning Calorimetry (DSC)

When a material is heated or cooled, there is a change in its structure or composition. These transformations are connected with a heat exchange. Differential scanning calorimetry (DSC) is used for measuring the heat flow into and out of the sample, as well as for determining the temperature of the thermal phenomenon during a controlled change of temperature. The first method developed by Le Chatelier in 1887 was differential thermal analysis (DTA), where only the temperature induced in the sample was measured.

The principle of DSC is as follows: two ovens are linearly heated; one oven contains the sample in a pan, the other contains an empty pan as a reference pan. If no change occurs in the sample during heating, the sample pan and the reference pan are at the same temperature. If a change such as melting occurs in the sample, energy is used by the sample and the temperature remains constant in the sample pan while the temperature of the reference pan continues to increase. Therefore a difference of temperature occurs between the sample pan and reference pan.

Manufacturers use two methods of measurements. In the first method called “heat flux DSC,” the instrument measures this temperature difference (DTA). Through calibration, this temperature difference is transformed into a heat flow, dq/dt . Therefore, there is a thermal factor that may vary with temperature. In the second method, called “power compensation DSC,” two individual heaters are used in order to monitor the individual heating rates of the two individual ovens. A system controls the temperature difference between sample and reference. If any temperature difference is detected, the individual heatings are corrected in such a way that the temperature is kept the same in both pans. That is, when an endothermic or exothermic process occurs, the instrument delivers the compensation energy in order to maintain equal temperature in both pans.

In the first case temperature is primarily measured, and in the second case, energy is primarily measured.

The differentiation of measuring principles is with modern instrumentation not very significant under normal applications. Due to calibration and integrated data handling, the instruments produce similar qualities of reported results.

Each instrument can deliver the same information, that is, heat flow as a function of temperature (or time). The peak shape, the resolution, and the sensitivity depend on the principle of measurement and the specification of the instrument.

For first-order transitions such as melting, crystallization, sublimation, boiling, etc., the integration of the curve gives the energy involved in the transition. For second-order transitions, the signal gives the change in the specific heat, for example, glass transitions.

Fig. 1 shows typical transitions. Melting and crystallization are first-order transitions. The extrapolated onset temperature (T_e) is the melting or boiling point. The peak temperature (T_m) is dependent on instrument and measurement parameters. The glass point is determined as inflexion point. Manufacturers represent the heat flow in different ways: the endotherms in the positive side for power compensation DSC and in the negative side for heat flux DSC. Melting, boiling, and sublimation are endothermic, which means they need energy. Crystallization is exothermic, which means that it supplies energy. Desolvations without melting are generally endothermic. Solid–solid phase transition and decomposition may be endothermic or exothermic.

Modern instruments provide heating, cooling and isotherms between subambient temperatures (with a cooling device) and higher temperatures in the range of 1200–1500°C. In order to avoid reactions with the atmosphere the measurements are carried out under nitrogen. The major components of the systems include the DSC sensors, the furnace, the programmer, and the data handling. The temperature plotted on the abscissa is the programmed temperature, not the temperature of the sample.

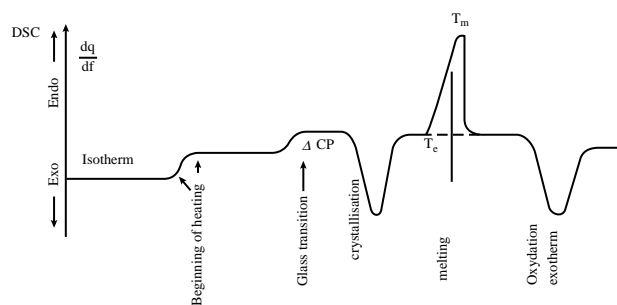


Fig. 1 Theoretical DSC scans.

The difference between the programed and the actual temperature of the sample is called “thermal lag.” It depends on the thermal resistance of the instrument and the heating rate. In modern instruments dedicated to accurate analytical measurements for pharmaceuticals, the sensors are in direct contact with the bottom of the pans and the sample size is in the milligram range or less. Therefore this correction is not very high, but it has to be taken into consideration. Generally pure indium (>99.9999%) is used for the correction of the thermal lag.

Great efforts have been made in recent years in order to validate the different instruments not only in comparing principle and results but also in determining the critical parameters such as heating and cooling rates, particle size, weight, resolution, atmosphere, and type of pans (crimped pan, sealed pan, open pan, etc.).

The instruments include automation and data acquisition. The calibration of the instrument should be done at a yearly basis. This includes the measurement of temperature and enthalpy. Most certified standards are highly purified metals. Indium is the preferred reference standard, but it covers only one temperature. It is recommended for pharmaceuticals to include several organic substances for which the melting point or the melting enthalpy has been accurately determined. Sarge et al. (16,17) proposed several organic substances and metals. The heat determination of quartz was also recommended. Sabbah (18) published recently a broad review of data of organic substances. For pharmaceuticals, it is suitable to have certified materials covering a broad range corresponding to the thermal events of interest (19).

Tables 1 and 2 are examples of calibration of the temperature and of the calorimetric response of PE-DSC-7 instruments by using different materials (20). Very important for pharmaceutical industry is the confidence of the laboratory that delivers the reference. Since the heating rate may have an influence on the data, it is recommended to compare the melting point and the melting enthalpy of organic standards, additionally to indium, at different heating rates covering the measurement range. For very accurate determinations, it is recommended to use standards with a melting point in the range of the considered temperature in a series of measurements.

Pressure DSC

In Pressure DSC (PDSC), the sample can be submitted to different pressures, which allows to characterize substances at the pressures of processes or to distinguish overlapping peaks observed, for example, by desolvation (21).

Table 1 Example of calibration of Perkin-Elmer DSC-7 instruments with melting standards at 10 K min⁻¹ under nitrogen

Certified substances	Onset T (°C) certificate	Instrument 1 Onset T (°C)	ΔT (°C)	Instrument 2 with intracooler Onset T (°C)	ΔT (°C)
Iodobenzene	-31.3			-32.2	0.9
H ₂ O	0.0			0.1	0.1
4-Nitrotoluene	51.5	50.4	1.1	51.2	0.3
Biphenyl	69.3	68.2	1.1	68.6	0.7
Naphthalene	80.2	79.4	0.8	80.1	0.1
Benzil	94.7	94.21	0.6	94.5	0.2
Acetanilide	114.0	113.9	0.1	113.6	0.4
Benzoic acid	122.1	122.0	0.1	121.8	0.4
Diphenylacetic acid	146.5	146.9	0.4	146.9	0.4
Indium	156.6	156.8	0.2	156.5	0.1
Anisic acid	183.1	183.6	0.5	183.2	0.1
2-Chloro-anthraquinone	210.0	210.1	0.1	210.1	0.1
Tin	231.9	232.7	0.8	232.7	0.8
Anthraquinone	284.5	285.2	0.7	284.8	0.7
Lead	327.5	328.6	1.1	—	—
Zinc	418.9	420.3	1.4	—	—

Modulated DSC

This new technique introduced in 1993 (22) has been thoroughly examined and discussed (23). Main advantages are the separation of overlapping events in the DSC scans. In conventional DSC, a constant linear heating or cooling rate is applied. In modulated DSC (MDSC), the normally linear heating ramp is overlaid with a sinusoidal function (MDSC) defined by a frequency and an amplitude to produce a sinusoidal-shaped temperature versus time function. Using Fourier mathematics, the DSC signal is split into two components: one reflecting non-reversible events (kinetic) and the other reversible events.

$$T = T_0 + bt + B \sin(\omega t)$$

$$dq/dt = C(b + B\omega \cos(\omega t)) + f'(t, T) + K \sin(\omega t)$$

where T is temperature, C the specific heat, t the time, ω the frequency, $f'(t, T)$ is the average underlying kinetic function once the effect of the sine-wave modulation has been subtracted. K is the amplitude of the kinetic response to the sine-wave modulation and $(b + B\omega \cos(\omega t))$ is the measured quantity dT/dt or “reversing” curve.

The total DSC curve, the reversing curve giving reversible transitions and the nonreversing curve giving irreversible transitions (e.g. the glass transitions), is

Table 2 Examples of calorimetric measurements of standards with two different DSC-7 instruments and measurement cell at different times (A, B, C) at 10 K min⁻¹ under nitrogen

Standard substance	ΔH (J/g) (Theory)	A		B		C	
		ΔH (J/g)	% deviation	ΔH (J/g)	% deviation	ΔH (J/g)	% deviation
Naphthalene (80.2°C)	148.6	147.1	1.0	148.6	0.0	—	—
Benzil (94.7°C)	112.0	110.1	1.7	112.8	0.7	—	—
Benzoic acid (80.2°C)	147.2	—	—	—	—	146.6	0.4
Biphenyl (69.3°C)	120.4	120.0	0.6	—	—	120.5	0.1
Diphenyl-acetic acid (146.5°C)	146.9	—	—	146.8	0.1	—	—
Indium (156.6°C)	28.7	28.63	0.2	28.8	0.35	28.7	0.1
Tin (231.9°C)	60.2	60.0	0.3	—	—	60.8	1.0

obtained. MDSC is a valuable extension of conventional DSC. Its applicability (24) is recognized for precise determination of the temperature of glass transitions and for the study of the energy of relaxation, and it depends on a number of important parameters to be studied. It has been recently applied for the determination of glass transitions of hydroxypropylmethylcellulose films (25) and for the study of amorphous lactose (26), as well as for the study of some glassy drugs (27).

Microwave thermal analysis (MWTa)

In this new technique (28), microwaves are used both to heat a material and to detect thermal transitions.

Micro-DSC

The instruments of conventional DSC allows to measure very small amounts of material. The author was able to characterize the melting peak of indium with 0.032 mg by using a DSC-7 of Perkin-Elmer. New instrument generation will permit to increase sensitivity and amount of material to be studied decrease to nanorange (29).

Microcalorimetry

Microcalorimetry is a growing technique (30,31) complementary to DSC for the characterization of pharmaceuticals. Larger sample volume and high sensitivity means that phenomena of very low energy (unmesurable by DSC) may be studied. The output of the instrument is measured by the rate of heat change (dq/dt) as a function of time with a high sensitivity better than $0.1 \mu\text{W}$. Microcalorimetry can be applied to isolated systems in specific atmospheres; or for batch mode where reactants are mixed in the calorimeter. Solution calorimetry can be used in adiabatic or isoperibol modes in microcalorimeters at constant temperature. (See the corresponding article about calorimetry of this edition.)

Thermogravimetric Analysis

In thermogravimetry (TG or TGA) the change in sample mass is determined as a function of temperature and/or time. The instrument is a thermobalance that permits the continuous weighing of a sample as a function of time. The sample holder and a reference holder are bounded to each side of a microbalance. The sample holder is in a furnace, without direct contact with the sample, the temperature of which is controlled by a temperature programmer. The balance part is maintained at a constant temperature. The instrument is able to record the mass loss or gain of the

sample as a function of temperature and time [$m = f(T)$]. Most instruments also record the DTG curve, which is the rate of the mass change $dm/dt = f(T)$.

The DTG curves, allow a better distinction of overlapping steps, as demonstrated in Fig. 2, for $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$. The area under the DTG curve is proportional to the mass change, and the height of the DTG peak at any temperature gives the rate of mass change. The real advantage of DTG is to permit accurate location of the end of a desolvation process if decomposition follows desolvation by use of the minima in the DTG curve.

The instrument used in thermogravimetry is a thermobalance (balance controller, sample chamber, furnace, furnace controller) with data processing. In order to check the stability of the system a baseline at the highest sensitivity has to be done for all heating rates in the temperature range of analysis: The highest deviation will be observed at the highest heating rate. The thermobalance may have a vertical or horizontal construction. The sensitivity of new thermobalances attains $0.1 \mu\text{g}$. Some manufacturers offer combined DSC/TG instruments.

The mass accuracy is generally not a problem of modern TG. Calibration of the mass with certified mass can be used as for all other balances. Electrostatics, temperature fluctuation, sensitivity of the sensor, and thermal lags have to be known, what is best done with regular calibration. For automatic TG, the pans have to be tightly closed and pierced just before the measurement; therefore, the TG curves of desolvation may be different as for open pans. The use of a protective gas and its flow, as well as the sample mass and the heating rate play a role in the comparison of the temperature of thermal events. The influence of heating rate is exemplified in Fig. 3 with $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$. The limit of detection can be calculated by determining the maximum

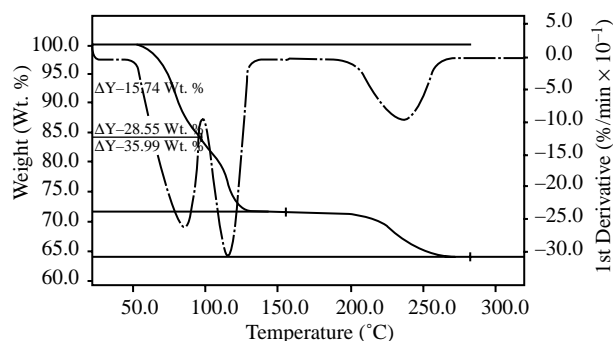


Fig. 2 TG of copper(II) sulfate pentahydrate with the heating rate 20 K min^{-1} . Use of DTG for the different steps.

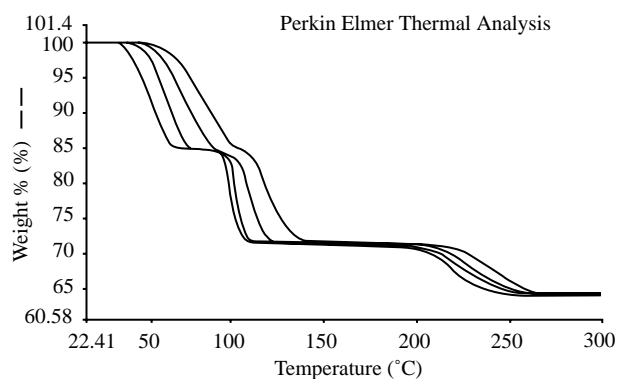


Fig. 3 TG of copper(II) sulfate pentahydrate. Influence of the heating rate. TG curves from the top: 20 K min⁻¹, 10 K min⁻¹, 5 K min⁻¹, 2.5 K min⁻¹.

of deviation of the base line in the temperature range of interest.

Table 3 shows an example of calibration performed with hydrates, which cover the starting of dehydration temperature from 50 to 120°C and the end of dehydration from approx. 150°C until to 270°C with different heating rates.

Since there is no contact between pan and furnace, the thermal lag is higher than in DSC. The standards recommended by ICTA and distributed by NBS are ferromagnetic standards exhibiting loss of ferromagnetism at their curie point temperature within a magnetic field: Nickel (354°C), Permanorm 3 (266°C), Numetal (386°C), Permanorm 5 (459°C), Trafoperm (754°C). The method does not permit the temperature measurement with high precision. These standards have been studied by several authors (32). The ICTA temperatures are within 5–10°C. McGhie et al. (33) proposed a calibration technique in which a small inert platinum weight is suspended by a fusible link composed of a calibration standard that releases the platinum weight at the temperature of melting. The Mettler instrument TGA 850 is constructed so that the melting curve of standards can be measured and used as calibration, as demonstrated in Table 4.

TG can be used with different atmospheres and under vacuum. TG has a huge number of pharmaceutical applications. Automated TG is extremely efficient to replace the loss on drying assay in drug substances, being able to separate loss of solvent from decomposition by using very small amounts of substance. Solvent entrapped or bounded as solvate is easily determined (9, 34, 35). A comprehensive article on TG has been recently written by Dunn and Sharp (36). Ozawa proposes the use of modulated TG for kinetic analysis (37).

Water sorption–desorption isotherms can be carried out by using thermobalances. Now specific instruments allow to measure water sorption–desorption isotherms at different constant temperatures (e.g., dynamic vapor sorption instrument (DVS), Surface Measurement Systems Ltd., Monarch Beach, US).

Thermomechanical Analysis

In thermomechanical analysis (TMA) the deformation of the sample under stress is monitored against time or temperature while the temperature increases or decreases proportionally to time. Changes are detected by mechanical, optical, or electrical transducers. The stress may be a compression, penetration, tension, flexure, or torsion. Generally the instruments are also able to measure the sample dimensions, a technique called thermodilatometry. The stress (F/A) expressed in N/m² or Pa may be a normal tensile stress σ , a tangential shearing stress τ , or a pressure change Δp ; the force applied is F and A is the area.

The deformation is measured by the strain, which is the deformation per unit dimension.

$$\text{Elongation} \quad \varepsilon = \Delta L/L_0$$

$$\text{Volume strain} \quad \theta = \Delta V/V_0$$

$$\text{Shear strain} \quad \gamma = \Delta x/y$$

For an elastic material, the Young's modulus is defined by

$$E = (F/A)/(\Delta L/L_0) \quad \text{or} \quad E = \sigma/\varepsilon$$

Creep is the gradual irreversible elongation of the sample.

Table 3 Example of calibration of loss of mass with three hydrate standards

Substance	Theoretical amount water	Result		
		5 K min ⁻¹	10 K min ⁻¹	20 K min ⁻¹
Sodium-tartrate dihydrate	15.7%	15.73%	15.60%	15.73%
Calcium-oxalate monohydrate	12.3%	12.55%	12.51%	12.48%
Copper-sulfate pentahydrate	36.1%	36.08%	36.03%	36.04%

Table 4 Example of calibration of the temperature of TGA 850 with melting standards

Substance	Theory	Result		
		5 K min ⁻¹	10 K min ⁻¹	20 K min ⁻¹
Nitrotoluene	51.5°C	51.49°C	51.64°C	53.78°C
Indium	156.6°C	157.62°C	157.38°C	157.74°C
Tin	231.9°C	233.44°C	233.42°C	233.68°C

These parameters depend on the temperature. The coefficient of thermal expansion is

$$\alpha = \left[\frac{dL}{dT} \right] \left[\frac{1}{L_0} \right]$$

The instruments have a furnace, and mostly a linear variable differential transformer (LVDT) to produce an electrical signal from a linear movement. An additional unit controls the force applied. Special attachments allow the same instrument to work in different modes such as elongation, compression, penetration, or tension.

The slope of the TMA trace may also be obtained by DTMA.

$$\frac{dL}{dt} = L_0 \alpha \frac{dT}{dt}$$

where dT/dt is the heating rate.

Thermomechanical methods are very useful for the determination of phase transformations such as polymorphic solid–solid transitions or glass transitions. Fig. 4 shows some theoretical curves for glass transition and polymorphic transition in extension or in penetration mode. Recently TMA has been proposed for the measurement of the internal stress of tablets of ethylcellulose of different molecular weight (38) and for measurement of swelling of polysaccharide hydrogels (39) and of polymeric films (40).

Calibration of the instrument for its response to length may be carried out with a standard length piece of metal or ceramic. The temperature can be calibrated in the same

way as for DSC. Metal standards such as indium, tin, or lead are mostly used. Recent publications (41) deal with calibration and errors of TMA.

Dynamic Mechanical Analysis

In this technique, the mechanical response of a sample is measured as it is deformed under oscillating load against temperature or time. Dynamic mechanical analysis (DMA) is a further development of TMA, but the instruments are different.

DMA is mostly applied to the study of polymers. Relevant parameters are the storage modulus and the loss modulus. Generally the loss $\tan \delta$, which is the ratio of these modulus, is plotted against temperature. A recent overview of the pharmaceutical applications of DMA has been published by Craig and Johnson (42).

Torsional braid analysis (TBA) is a particular case where the sample supported by a fiberglass braid is subjected to a torsional strain.

Hyphenated Techniques (Combined Techniques)

A comprehensive characterization of the physical properties of materials often requires a multidisciplinary approach since no single technique is capable of characterizing pharmaceuticals completely.

Thermomicroscopy or hot stage microscopy is a well-established method (43, 43–45) for the microscopic observation with polarization of the sample while heating

Table 5 Study of two polymorphs with unique melting curve

Property	Crystalline form A	Crystalline form B
Melting point °C	304	311
Melting enthalpy in kcal/mol	12	11
Water uptake after 1 day 92% RH	0%	3.2% (Hydrate)
Transformation in alcohols	A	B → A

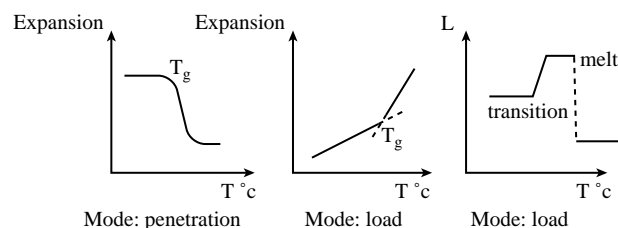


Fig. 4 Theoretical TMA curves for glass transition T_g and polymorphic transition.

or cooling, allowing to see desolvation, melting, crystallization, eutectic formation, and even transformations in suspensions in solvents. The combination of hot stage microscopy to new technology such as high-resolution color camera, image manipulation software makes the technique very attractive for inducing metastable states, for observation of crystal habit, and for better interpretation of other methods. Thermophotometry is the measurement of the light intensity and thermoluminescence of the light emitted by the sample. FT-IR microscopy (15, 45, 46), Raman microscopy (47–49) are excellent additional tools to thermomicroscopy. Calibration of microscope, of heating unit, and of spectroscopic methods should be done. TEM (transmission microscopy) and SEM (scanning electron microscopy) with EDX have been combined to DSC for the study of solid-state reactions.

Newly born, the scanning thermal microscopy derived from atomic force microscopy brings a revolution in the instrumentation for measuring thermophysical and thermomechanical properties of the matter, and the TA instrument was awarded at Pittsburgh 1998. The instrument has been applied for the characterization of Ibuprofen compacts as model substance (50).

Temperature-resolved X-ray diffraction with a heating cell is widely used (12, 51–54). Crystalline changes are clearly assigned; the X-ray diffraction patterns obtained *in situ* allow to predict quantitative methods if, for kinetic reasons, forms that are present at high temperatures occur at ambient conditions. Low-temperature X-ray diffraction cell has been developed for the study of frozen aqueous solutions (55). The introduction of XRD-DTA cell (56) and recently of the DSC-XRD instrument of Rigaku presented at the Denver X-ray Conference in 1999 (57) demonstrates the advantage of this direct combined technique. The observation of polymorphic transformation by using variable temperature synchrotron X-ray diffraction method is a promising technique with the new computerized ability for obtaining structural data (58).

Thermogravimetry can be coupled with DSC. Most companies offer the TG-IR (59) or the TG-MS coupling

(60). Synergic chemical analysis by coupling TG-FT-IR, TG-MS or TG-GC-MS has been recently discussed (61). Fig. 5 demonstrates the ability of TG-MS for the study of dehydration and decomposition of calcium oxalate dihydrate. The steps correspond to the dehydration into anhydrous calcium oxalate, followed by the transformation into calcium carbonate then by the formation of calcium oxide (10, 11). The sample studied in Fig. 5 contains additionally to crystal water some free water. The TG-MS shows that some amount of the CO evolved during the decomposition also transforms into CO₂.

These new emerging combined techniques enables the observation of extremely small samples with a high degree of information. They find a good place for proper screening (55) (according to the polymorphic studies required by ICH) (62) in order to choose the proper form and to justify its choice. They also permit to analyze more easily the complex matrixes of drug products.

APPLICATIONS OF THERMAL ANALYSIS TECHNIQUES FOR DRUG SUBSTANCES

The transitions observed by thermal analysis techniques are based upon the Gibbs phase rule and phase diagrams, p , T , concentration.

All transitions or reactions involving energy changes may be measured by DSC. Transitions involving mass changes are detected by TG. For a single product, specific heat, glass transition, melting, boiling, sublimation, decomposition, and phase transitions induced by polymorphism during heating are important for the choice of the salt form and for safety studies where the DSC exothermic peaks are relevant. The use of DSC for the measurement of the melting point of raw materials has been proposed (20). Hydrates or solvates, or volatile compounds in the formulations can be investigated by DSC combined with TG and TG-MS. DSC curves of mixtures of solid compounds depend upon the phase diagrams in solid state. If there is no interaction in the solid state and if there is a miscibility in the melt, an eutectic behavior is observed. This enables the purity determination of raw materials, the analysis of enantiomers, and the study of “physical interactions” in preformulation. If the compounds are not miscible in the liquid state, the DSC curve of the mixture is the addition of the DSC curves of each compounds. Interaction is observed in the solid state in case of formation of solid solution or complex formation between components or in the case of chemical reaction.

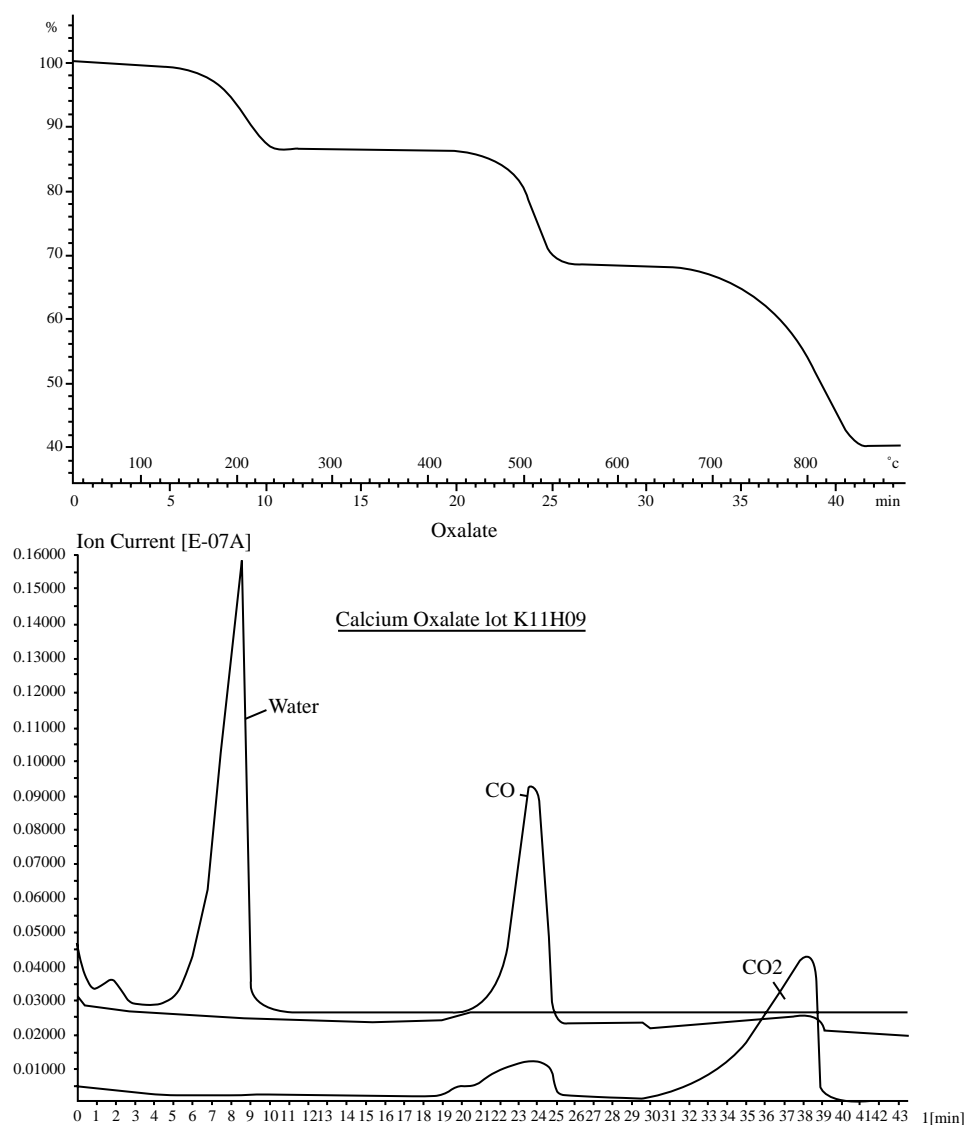


Fig. 5 TG and TG-MS of calcium oxalate monohydrate. The TG steps correspond to the dehydration followed by the formation of calcium carbonate, then calcium oxide is obtained. The evolved gas is detected by MS.

Polymorphism/Pseudopolymorphism and Amorphous State

Polymorphism is the tendency of any substance to crystallize into different crystalline states. The solid forms of the same compound are called polymorphs or crystalline modifications. On melting, they produce the same liquid. Polymorphs show the same properties in the liquid or gaseous state but they behave as different substances in the solid state. The best known example of polymorphism is carbon, which can exist in the form of graphite or as a diamond.

The amorphous state characterizes crystallization in a nonordered, random system, related to the liquid state. The name “glassy state” is given to amorphous products that liquify by undergoing a glass transition.

The expression pseudopolymorphism applies to hydrates and solvates.

Different solid phases that may occur during crystallization or galenical processes are polymorphs, amorphous phases, or solvates as the result of compound formation with the solvent (63–66).

A recent detailed review about thermal analysis of polymorphs and pseudopolymorphs (67) listed more than 300 drug substances presenting this behavior in the literature. Polymorphism of excipients and their thermal analysis has been reviewed in (68).

DSC gives not only temperature of events, but also melting energies. The Burger's rule and the energy diagrams (69–74) help considerably to approach the thermodynamic equilibrium of a polymorphic system.

For each polymorph (single compound), there is a solid–liquid equilibrium curve and a solid–gas equilibrium curve. The solid–gas curves meet at a point. If the liquid–gas equilibrium curve meets the two solid–gas curves *after* this point of intersection, there will be a solid 1, solid 2 equilibrium curve and a reversible transition point $1 \leftrightarrow 2$ at a specific pressure. This is known as *enantiotropy*. At the transition point, the free energy of the two forms is the same.

The term *monotropy* applies in the case of an irreversible transition from one form to another. Monotropy is bound to the existence of metastable thermodynamic forms. The liquid–gas curve crosses the solid–gas curves for the two forms *before* their point of intersection.

Knowing the relationship between the thermodynamic quantities H (enthalpy), G (free energy), S (entropy), and T (temperature), it is often simple to represent equilibrium states by plotting the free energy G as a function of the temperature for each form. If the two curves intersect before the melting point, there is reversibility, i.e.

enantiotropy, and if the reverse is true, there is *monotropy*.

The relationship between melting enthalpies of two solid phases A and B and the heat of transition is:

$$\Delta H_t = \Delta H_A^f - \Delta H_B^f$$

Figs. 6A and 6B illustrate the plots of the functions G and H versus temperature (energy/temperature diagrams) for each polymorph and for the liquid. The thermodynamic reversibility of the solid transition between two crystalline forms is characteristic of enantiotropic systems. Each form has its thermodynamic stability range. The lower melting form is stable in the temperature range below the transition point; the higher melting form is stable in the temperature range above the transition point. In case of monotropy only one form is stable whatever the temperature range. The Burger's rule is as follows in the case of *enantiotropy*, the lower melting form has the higher melting enthalpy and the transition into the high melting form by heating is endothermic; in the case of *monotropy*, the thermodynamic stable form is the higher melting form with the higher melting enthalpy. The transformation of the unstable form into the stable form is exothermic.

Because of kinetic factors, metastable forms are encountered in temperature ranges outside the thermodynamic range. Crystallization processes generally imply the cooling of concentrated solutions or precipitation by addition of cosolvent. Depending on the relative positions of

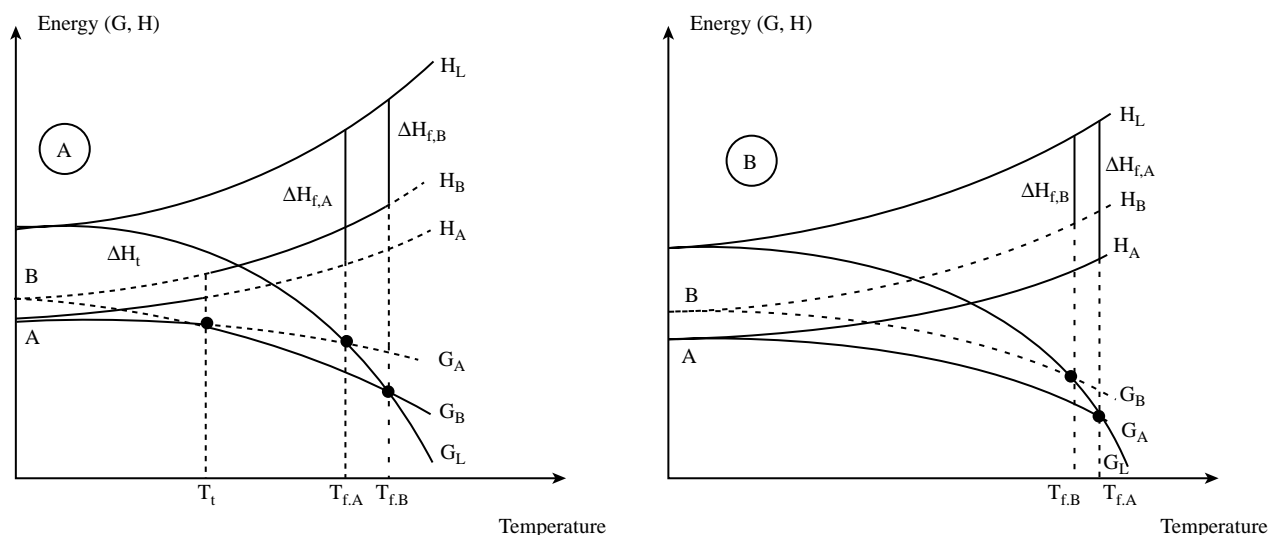


Fig. 6 Energy diagrams showing plots of enthalpy H and Gibbs free energy G , against temperature T , for the solid and liquid phases of a single compound, showing (A) enantiotropy and (B) monotropy.

the solubility curves of the metastable polymorphic forms and the metastable curve of supersaturation, the first nucleus can be a metastable form. Transformation to the stable crystalline form may or may not occur, depending on kinetic factor. Furthermore solvates exist at lower temperatures and their presence should be considered and finally due to the humidity of the air or from water activity of the solvents, hydrates may be formed. Polymorphism of solvates and hydrates is not uncommon. This phenomenon of concomitant polymorphs has been recently reviewed (75).

Fig. 7 illustrates the behavior of polymorphs A and B in case of enantiotropy (Fig. 7a) and monotropy (Fig. 7b) during heating. For all analysis where a temperature change is involved, kinetic factors have to be considered for proper interpretation of the results. The DSC scans will differ if the sample being analyzed is stable or metastable at ambient temperature. A is the stable form at ambient temperature in both cases.

In the case of *enantiotropy* (Fig. 7a), only the form A should be encountered below the transition point and the behavior upon heating is illustrated by the DSC scan 1, the endothermic transition $A \rightleftharpoons B$ should be observed followed by the melting peak of the form B. For kinetic reasons, e.g., in case of too quick heating rate, the transition $A \rightleftharpoons B$ does not occur and form A melts. Then two possibilities may be found: no other signal occurs (DSC scan 2) or an exothermic crystallization of B from the melt occurs and later on, the melting peak of B is observed (DSC scan 3). If the metastable form B occurs below the transition, upon heating it can be transformed into A with an exothermic transition (DSC scan 4);

thereafter, the form A transforms into the form B according to scan 1 or scan 2). Finally (scan 5), the metastable form B can melt without any transformation.

In the case of monotropy (Fig. 7b) only the form A should exist and the DSC scan should show the melting peak of the stable form A (scan 1). If the metastable form B is heated, then scan 2 or 3 may be observed: the form B transforms exothermically into the stable form A (scan 2) or the form B melts and the stable form A grows from the melt and its melting peak is observed.

In the case of racemate, the situation is more complex since one racemate polymorph can be a true racemate and the other one a conglomerate (73, 74) and one has to consider additional solid phase transitions such as peritectoids and eutectoids, which are not polymorphic transitions (76).

Very often some substances have two melting points separated by an exotherm. Such a DSC scan can correspond to a monotropy or to an enantiotropy. The sample may be a pure form or a mixture. Using different heating rates and tempering in DSC, one may be able to measure melting points and melting enthalpies and to use the Burger' rule.

The TG curve is extremely valuable for preventing misinterpretations. McCauley (77) describes the DSC curve of phthalylsulfathiazole, which presents such a DSC curve with two endothermic peaks separated by an exothermic peak. The TG curve demonstrates a strong decomposition during the first melting. The resulting degradation product then recrystallizes and melts at higher temperatures. We observed such an effect for aspartam

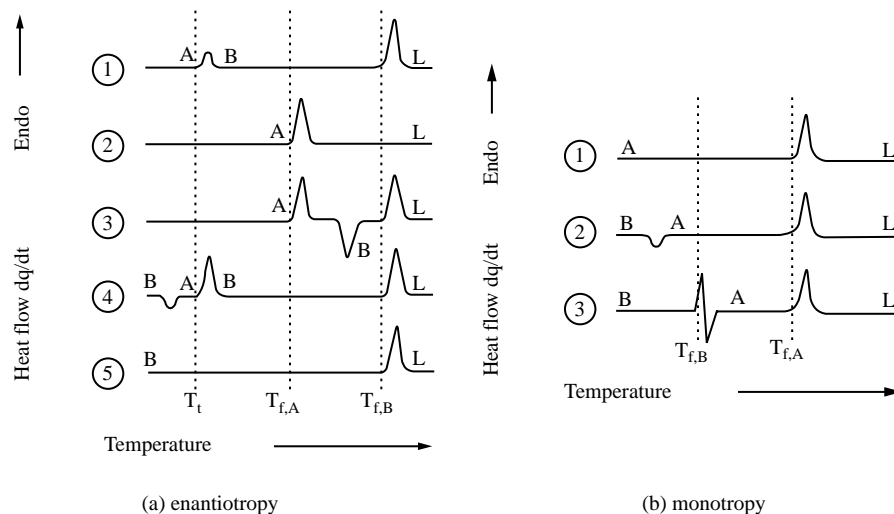


Fig. 7 Possible DSC curves for two polymorphs: (a) Enantiotropy $A \rightleftharpoons B$, B is the highest melting form, A is the stable form below the transition point. (b) Monotropy $B \rightarrow A$, A is the highest melting form. (For explanations of the scans, see text.)

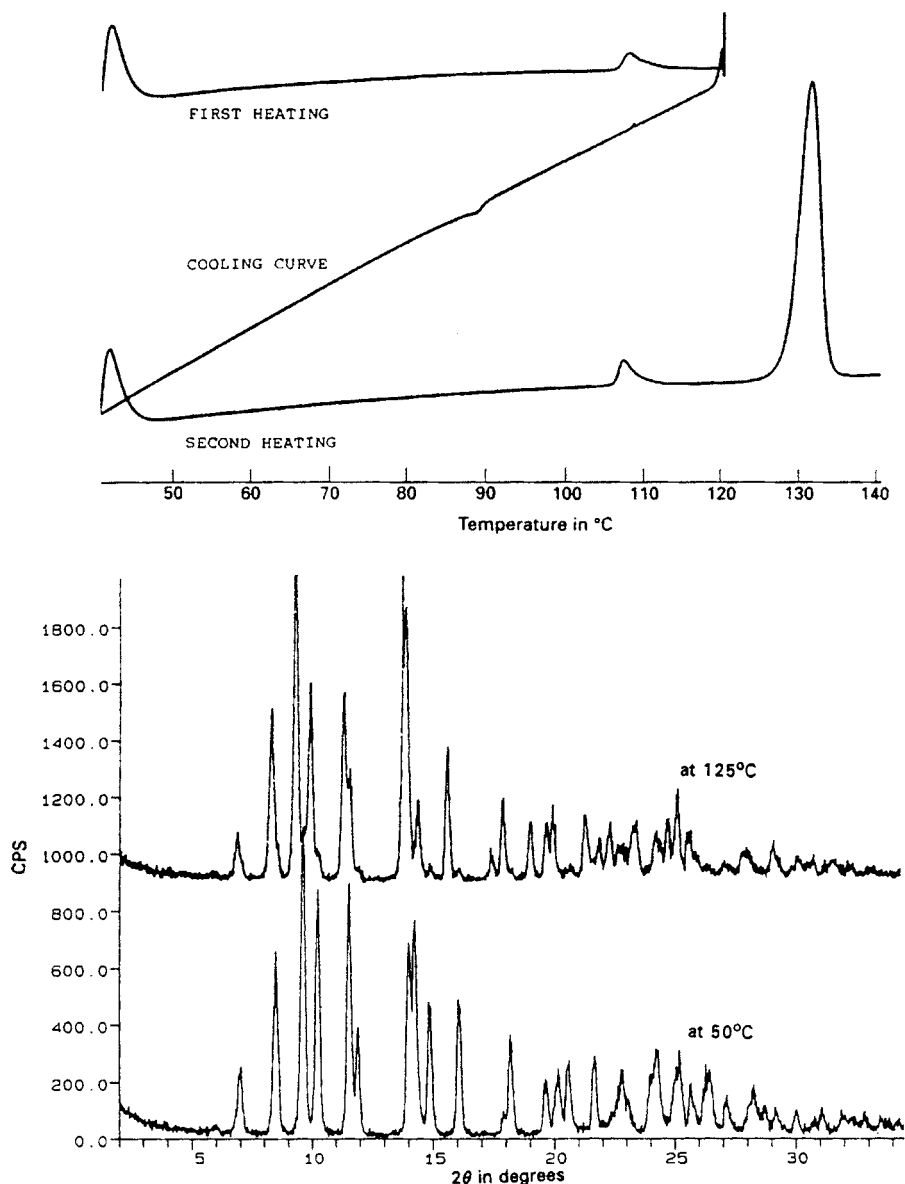


Fig. 8 Reversible transition in a purine derivative studied by DSC and temperature-resolved X-ray powder diffraction (From Ref. 67.)

(13) and for a malonate salt, which decomposed in the corresponding base (12). If an isomerization takes place, then only the analytical data of the product obtained allow us to have an accurate interpretation.

These curves illustrate the complex transitions that may occur when heating or cooling polymorphs. Therefore, combined techniques are very useful for complete interpretation of observations given by DSC as emphasized in (54). Fig. 8 illustrates a reversible enantiotropic transition followed by DSC and by temperature-resolved X-ray diffraction of a purine derivative. For this substance

six crystalline forms were found. The Burger's rule as well as the study of the supersaturated solutions and the use of combined techniques allowed us to find out that the form of Fig. 8 was the stable form below the transition point. All other forms were monotrops to this form (78).

Table 5 deals with the example of a benzoquinoline hydrochloride for which both forms presented a melting that was followed by decomposition. No change was observed by slow heating rate. Since the melting enthalpies differed only by 10%, the proper interpretation needed the verification of the hypothesis:

enantiotropic transition. The analysis of the insoluble solid in the equilibration of both forms in alcohols (solvent mediated transition) showed that form A is always obtained, what confirmed the observation of the Burger's rule.

Amorphous State

If a physical property of a crystalline substance is plotted against temperature, a sharp discontinuity occurs at the melting point. For amorphous substances, there is no melting point, and a change of slope occurs at the so-called glass transition temperature T_g . The glass transition is characterized by a change of heat capacity. Below this temperature, the amorphous phase has certain properties of a crystalline solid (e.g., plastic deformation) and is termed "glassy." Above this temperature, the substance retains some of the properties of a liquid, e.g., molecular mobility, and is termed "rubbery." Above this temperature, the increase in molecular mobility facilitates spontaneous crystallization into the crystalline form with an exothermic enthalpy change after the glass transition. The use of amorphous forms is attractive, particularly for sparingly soluble compounds because of the enhanced solubility and dissolution rate over the crystalline state leading to increased bioavailability. However, the amorphous state is thermodynamically unstable. The glass transition temperature T_g is lowered by water or other additives, facilitating conversion to the rubbery state and hence facilitating crystallization (79). Since it may be interesting to maintain the amorphous state, the temperature of the glass transition and its behavior should be characterized. DSC and modulated DSC are commonly used. Hancock and Zografi studied intensively the amorphous state of drug substances and used the relaxation energy at the glass transition as well as the dependency of the heating rate for the study of the "fragility" of the amorphous state (80–82). Depending on the temperature, the isothermal crystallization in one or the other polymorphic form may be favored as demonstrated for amorphous indomethacin (83).

The study of the amorphous state (for drug substances and also for excipients) is based upon the changes observed in the glass transition. Fig. 9 shows the influence of the heating rate for the determination of the glass transition by DSC for the polymeric excipient Carbopol 974. It is classical to perform a first run for the elimination of water and relaxation energy and to determine the glass transition temperature T_g of the pure compound, if no degradation occurs during the first run. However, this procedure is not valid if it is desirable to study the role of the matrix.

When the amorphous material does not transform into the crystalline material, the measurement of the melting

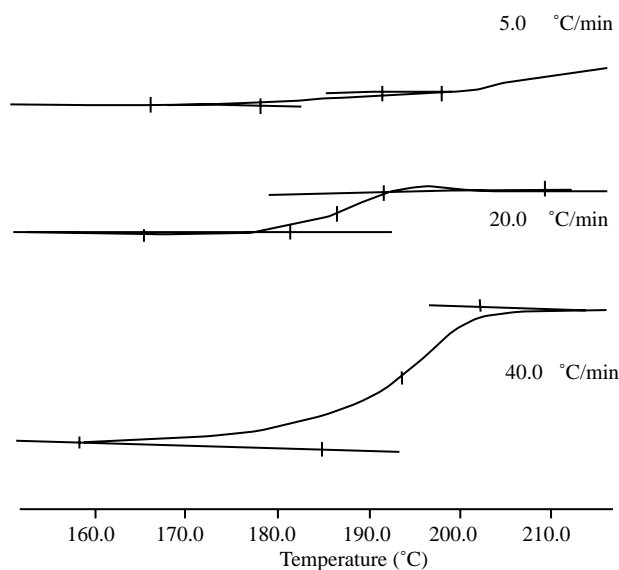


Fig. 9 Quality control of excipients by using the glass transition. Example of Carbopol 974 P. Second DSC run in order to eliminate water and relaxation: Influence of the heating rate.

peak allows to determine the degree of crystallinity of mixtures by comparing its value with the melting enthalpy of a pure crystalline material. If the amorphous sample crystallizes upon heating, then the crystallization peak may be used for the determination of the amorphous content. Such an example is given in Fig. 10. In this case it was possible to attain a limit of detection of 1%.

Pseudo-polymorphism

In the case of solvates, binary phase diagrams of temperature versus concentration of the solvent (or water) at a given pressure are useful for the understanding of the phase transitions. The characterization of solvates and hydrates need the use of both DSC and TG. Desolvation can be complex: melting of the solvate followed by exothermic recrystallization into the anhydrous form or solid-state transformation with evaporation and possibly further endothermic or exothermic events corresponding to a cascade of phase transitions. In such complex situations, combined techniques TG-IR or TG-MS and temperature-resolved X-ray diffraction are extremely helpful since the identity of the volatile component or of possible volatile decomposition product can be identified on line (54, 67). In case of hydrates, water sorption–desorption isotherms as well as the X-ray diffraction in humid chambers are needed (54, 84).

The DSC, TG curves of solvates and hydrates are related to the phase diagrams between substance and solvent (or

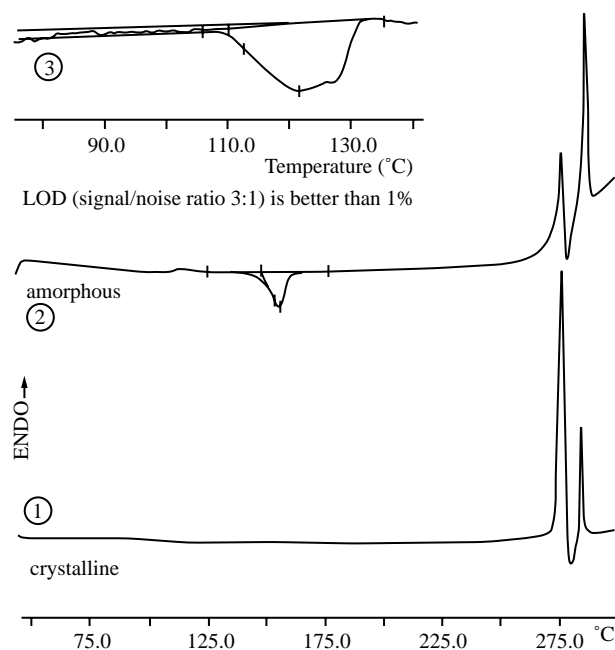


Fig. 10 Determination of amorphous content by DSC at 20 K min⁻¹. 1) Crystalline sample, 2) amorphous sample, 3) exotherm of a sample containing 4% amorphous as calculated by DSC. Calculation of the limit of detection by peak/noise ratio.

water). Eutectic are observed (85). Fusion or decomposition of the solvate may occur during heating. Therefore, one may observe the melting of the solvate followed by recrystallization into the anhydrous form or the endothermic desolvation in the solid state. In certain cases both phenomena may overlap. Details about experimental factors and examples can be found in (67). If the anhydrous form is metastable, further phase transitions follow the desolvation. If several solvates or hydrates exist, the transitions observed depend on the pressure, as demonstrated by Soustelle (86) in the case of copper sulfate pentahydrate. Depending on the pressure, the direct dehydration into the anhydrous or the dehydration via the monohydrate, or the three dehydration steps trihydrate, monohydrate and anhydrous forms may be obtained. Hydrates have been the subject of several reviews (87, 88). Polymorphism of hydrates (89) is also frequent. Figure 11 shows the case for a drug substance with polymorphic behavior of the trihydrate. The anhydrous form shows a dual melting (enantiotropy). This form was very hygroscopic and transformed into the trihydrate H_I at room temperature. The DSC scan of this trihydrate shows a dehydration peak in solid state, followed by the same scan as on the original anhydrous form. After storage for several months in a tropical climate (30°C/75% RH), a second trihydrate H_{II} is obtained. The same trihydrate H_{II} is also

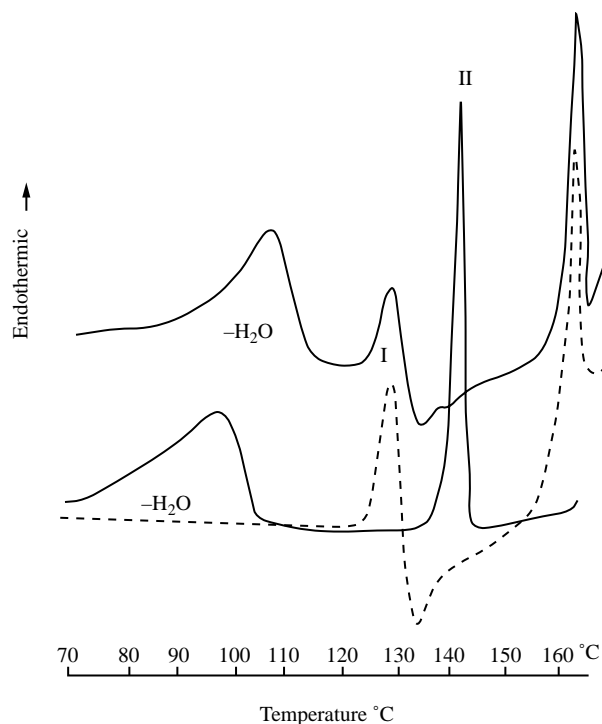


Fig. 11 Polymorphism of a trihydrate. (From Ref. 10.)

obtained by crystallization in water or after equilibration of the original anhydrous form with saturated aqueous solution. The DSC scan of this trihydrate differs from the first one because the dehydration gives rise to a new anhydrous form. This interpretation was confirmed by X-ray Guinier de Wolff diagrams, purity analysis, Karl Fischer, and TG (10, 67).

Often the solvates (hydrates) are not detected since, according the corresponding phase diagram, at ambient temperature, they can be partly or completely dissociated. Suspensions of hydrates in water should shift the equilibrium toward the formation of the stable hydrated form. The ability of DSC measurements at subambient temperatures allow to determine phase transitions. Giron et al. (90) proposed to use the melting peak of freezable water for the analysis of suspensions of drug substances in water in combination with TG for the determination of the number of molecules of water bounded as hydrates.

Study of Transition: Kinetics

Since transitions may occur during milling, processing, and aging, thermal methods are widely used for the kinetic study of all transformations. The purpose of any kinetic study is to obtain information concerning the reaction mechanism through comparison of a series of measured

fractions converted versus time. Most mechanisms in solid state are a nucleation period, a growth zone, and an unreacted core. For phase transitions of polymorphs and pseudopolymorphs, only heterogeneous kinetic applies (at least two modifications or two phases and a gas). In heterogeneous kinetic a great number of factors should be considered as temperature gradient in the sample, particle size, activation, nucleus, or diffusion.

A summary of current kinetic methods used with thermal analysis techniques can be found in (67).

Microcalorimetric Techniques

Solution calorimetry allows us to investigate processes that involve enthalpy changes. Adiabatic microcalorimeters and isoperibol calorimeters used in batch modes or flow modes allow for the precise determination of the heat of solution. Mixing the reactants is accomplished by breaking a bulk allowing reactants to mix or by special chambers where the reactants are mixed together.

If a compound exists in two or more different crystalline or amorphous configurations with different lattice energies, the heating solution in any given solvent will differ. The difference in the heats of solution will be equal to the difference in lattice energy of the solids, provided that the solid compounds are identical chemically. For example, we measured the energy of solution in water of the two modifications of a drug substance. The difference of 9.7 kJ mol^{-1} was found very close to the difference of the melting energies of 9.1 kJ mol^{-1} measured by DSC (54). However, the DSC information is superior since the temperatures of melting of both forms are measurable. Since the lower melting form had a lower melting energy, it was the stable form and both forms were enantiotropically related. For review see Ref. 67.

Byström (91) developed a technique in order to determine the crystallinity of drug substances by isothermal microcalorimetry. Considering that micronization introduces amorphous regions not measurable by X-ray diffraction, the method should be an analytical tool

for analyzing batch-to-batch quality. The principle of the measurement lies in the transformation of the amorphous state in the crystalline one at high humidity levels. The amorphous substance adsorbs water and the glass transition is lowered, permitting the acceleration of the crystallization. The energy evolved is measured in function of time by isothermal microcalorimetry. Results were found comparable with X-ray diffraction but a quite lower limit of detection is possible (better than 1%) (54, 92, 93).

DSC Purity Analysis

The basis of any calorimetric purity method is the relationship between the melting depression of a substance and the level of impurities according to van't Hoff's law. The purity is readily calculated from the DSC curve of a single melting event of a few milligrams of the substance, without the need for reference standard of the drug substance and its impurities.

The DSC impurity analysis is described in USP. With modern equipment including robotic systems and data acquisition, the DSC purity analysis is a state-of-the-art technique for pharmaceutical development.

The determination of purity by means of DSC is based on the assumption that impurities depress the melting point of a pure material according to the eutectic phase diagram behavior.

Figure 12 shows the phase diagram for the two component mixture with the so-called eutectic point. At the eutectic point E (e.g., 40% A, 60% B), the crystals A and B melt together at the temperature T_E , below the melting temperature of the pure compounds. If a mixture of A and B (containing, e.g., 90% A) is heated, the melting of eutectic mixture (which is 40% in A) is observed initially, until all of B is melted. During the melting of the eutectic (40% A, 60% B) a part of A is melted with B, with the corresponding amount $2/3 \times 10\%$ of A, i.e., 6.66% of A.

Then as the temperature increases, pure A melts between T_E and T_m . T_m is the temperature at the end of the melting. For the corresponding DSC curve, an endotherm

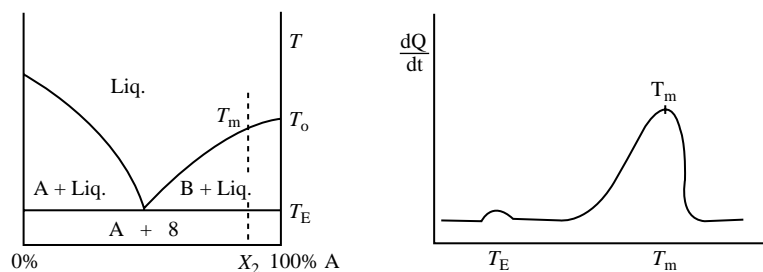


Fig. 12 Binary phase diagram with eutectic, and DSC curve of the composition x_2 .

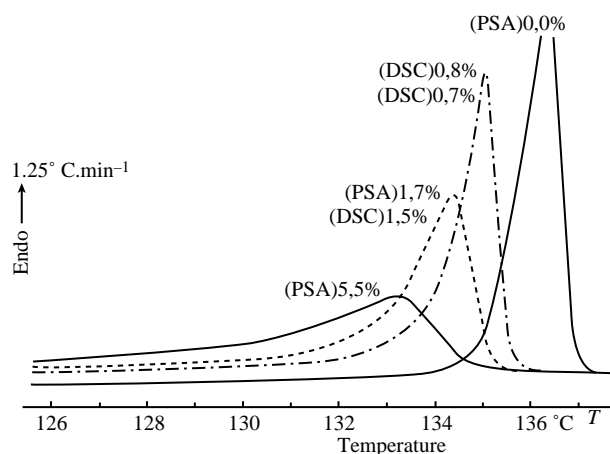


Fig. 13 Broadening effect of the melting curve of β -hydroxypropyltheophylline due to impurities. PSA = phase solubility analysis. All batches have the same TLC purity results. (From Ref. 6.)

at the eutectic temperature is observed, then the melting of crystals A occurs. The effect of impurity on the DSC curve is a melting depression and a broadening of the melting curve (Fig. 13).

The amount of impurities is calculated from the melting-point depression $\Delta T = T_0 - T_m$.

The van't Hoff's law for *diluted solutions* is

$$x = \frac{(-\Delta T \Delta H_f)}{RT_0^2}$$

where x is the mole fraction of impurities, ΔT the melting point depression, ΔH_f the melting point of pure material, T_m the melting of the analyte, T_0 the melting point of the pure compound, and R the gas constant.

The DSC procedure does not directly measure ΔT , but can be used to calculate it from the melting curve. At the eutectic point, all of B is in the liquid phase. During the melting of A after the eutectic point the concentration of B varies in the liquid phase. This causes the broadening of the DSC curve. With no solid solution formation, the concentration of impurity in the liquid phase at any temperature during the melting is inversely proportional to the fraction melted at that temperature, and the melting-point depression is directly proportional to the mole fraction of impurity. A plot of the observed analyte temperature T_i versus the reciprocal of the fraction melted $1/F_i$ at temperature T_i should yield a straight line with the

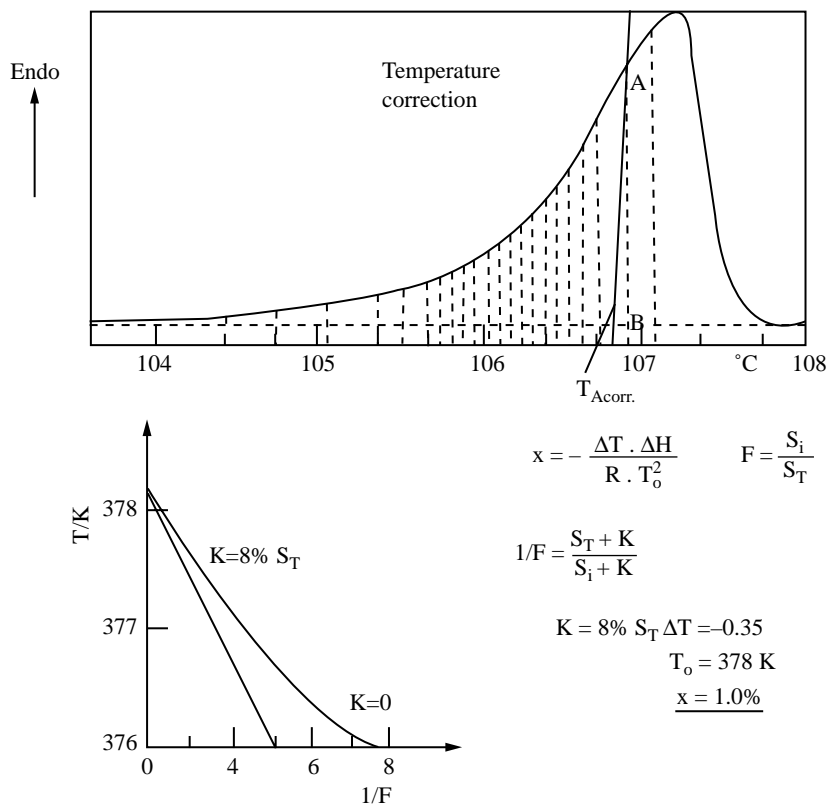


Fig. 14 Purity calculations by DSC.

slope equal to the melting-point depression ($T_0 - T_m$). The theoretical melting point of the pure compound T_0 is obtained by extrapolation to $1/F_i = 0$:

$$T_i = T_0 - \frac{RT_0^2(1/F_i)}{\Delta H_f}x$$

This relation may be expressed as

$$T_i = T_0 - \Delta T(1/F_i)$$

Substituting the experimentally obtained values for ΔT , ΔH_f , and T_0 in the first equation yields the mole fraction of the total eutectic impurities, which, when multiplied by 100, gives the mole percentage of total eutectic impurities.

The temperature of each point T_i is the sample temperature, not the programmed temperature. Due to the thermal lag, a correction depending on the instrument has to be done for each point.

The melting curve is divided into small portions and each area S_i is calculated. The melted fraction F_i ,

$$F_i = \frac{S_i}{S_{\text{total}}}$$

is calculated for each point and the curve T_i is plotted as a function of $1/F_i$, where T_i is the temperature at fraction F_i (Fig. 14). The slope ΔT and the ordinate T_0 can be calculated.

Partly because of the lack of the eutectic-point detection, the curve is not a straight line, and a correction factor K must be added to each fraction of the curve. Formation of solid solution or artefacts during melting may also be responsible.

$$F_i = \frac{S_i + K}{S_{\text{total}} + K}$$

Software from manufacturers mostly use iterative linearization, which gives the best value of K . Characteristics of this determination are as follows:

- Impurities are measured, which have an eutectic behavior (i.e., solubles in the liquid phase and insolubles in the solid phase).
- The sum of impurities should be $\leq 2\%$.
- The result is expressed in mol % without knowledge of impurities.
- Pure material is not needed.
- Small amounts (1–2 mg or less) of material are used.
- If decomposition occurs during melting, it can give erroneous results.
- The purity results are obtained after less than 1 h.

The influence of products parameters and instrument parameters have been discussed in detail by Giron et al.

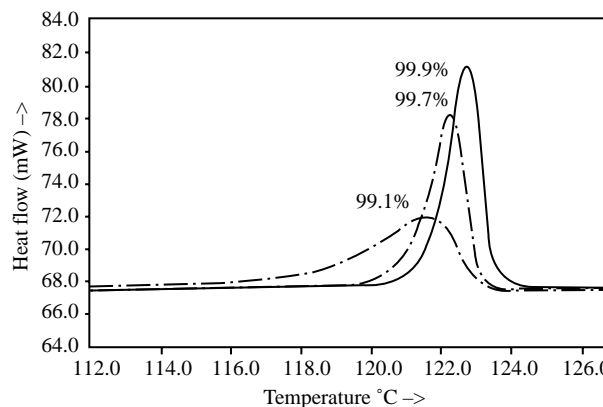


Fig. 15 Stability screening. Example of DSC purity results obtained for a drug substance sensitive to moisture. From the top to the bottom: Initial sample (purity 99.9%), sample stored under humid conditions at 50°C (purity 99.7%), sample stored under humid conditions at 80°C (purity 99.1%).

(94), who proposed a validation scheme for DSC purity method. The authors showed the advantage of the method as support to chromatographic techniques, for the monitoring of purification for the study of stability behavior of raw materials under stress conditions, for establishing purity profiles. Fig. 15 shows the DSC scans corresponding to a stability screening of a drug substance sensitive to moisture.

Enantiomers

Enantiomers are stereoisomers, which are mirror images of each other. An equimolecular mixture of two enantiomers is called a racemate. Crystalline racemates occur in three different types. The first is termed a conglomerate, that is, a mechanical mixture of crystals of pure enantiomers that is formed from two solid phases. The most common type is the racemic compound, which consists only of one crystalline phase in which the two enantiomers are present in equal quantities. The third type is the pseudoracemate in which a solid solution of the two enantiomers is present.

Conglomerates that are equimolecular mixtures of two crystalline enantiomers are easily separated by crystallization. There are two phases in the solid state and only one phase in the liquid state (miscibility). The equation of Schröder–Van Laar in its simplified form correlates the composition of mixtures to the end of fusion T^f :

$$\ln x = \frac{\Delta H_A^f}{R} \left[\frac{1}{T_A^f} - \frac{1}{T^f} \right]$$

where x is the mole fraction of the more abundant enantiomer ($0.5 \leq x \leq 1$) of a mixture that melts at T^f (in

K). ΔH_A^f and T_A^f are respectively the enthalpy of fusion and the melting point of the pure enantiomers, and R is the gas constant.

Racemic compounds, or true racemates, exhibit two eutectic points each between the pure enantiomer and the racemic compound. The shape of the DSC diagrams can vary, depending on the relative positions of temperatures of eutectics and racemic compound and on the composition of the eutectics (Figs. 16b and 16c). In the case of Fig. 16c it is difficult to distinguish by DSC a racemic compound from a conglomerate. Other methods as IR or X-ray are suitable for proper interpretation. For example, propanol hydrochloride has been described as conglomerate (95) or as racemate compound (96).

The equation of Schröder–Van Laar, which permits the calculation of the liquidus curve, may be applied to the point of the liquidus between the pure enantiomers and the corresponding eutectics (T_{AE}^f and T_{AS}^f). For the part E_RRE_S , the equation of Prigogine and Defay applies:

$$\ln 4x(1-x) = \frac{2\Delta H_R^f}{R} \left[\frac{1}{T_R^f} - \frac{1}{T^f} \right]$$

where x represents the mole fraction of the enantiomer in the mixture that melts at T^f and ΔH_R^f and T_R^f are respectively the enthalpy of fusion and the melting point of the racemic compound. Polymorphism can occur for the racemic compound and the enantiomers. Phase diagrams with monotropic transformation or enantiotropic transformation have been discussed (75, 76). Quite interesting is the transformation of the racemic compound into a conglomerate since this phenomenon can be used for purification via crystallization, as described for nimodipine (97). DSC is applied for the establishment of phase diagrams, for the determination of thermodynamic data (98), for the purity determination, or for the monitoring of industrial resolutions. For the establishment of phase diagrams it is suitable to add spectroscopic or crystallographic methods (99).

Two methods can be used for the purity determination: the direct method or the indirect method (75). The direct

method is applied for mixtures when the phase diagram is established using Schröder–Van Laar or Prigogine–Defay equation. For enantiomers of high purity (>95%) the general DSC purity method for eutectic impurities is applicable. The same limitations remain (polymorphism, degradation, during melting). The method gives the sum of impurities without differentiation of the type of impurity. For purity > 95% the same results have been obtained using the indirect and the direct methods (100, 101).

Conglomerates are easy to purify by crystallization or by the entrainment technique described by Jacques et al. (75), which involves introducing seed crystals of the desired enantiomer into a cooling saturated solution of the racemic mixture. According to these authors, conglomerate formation is observed three times, more frequently with salts. Therefore it is advantageous to compare the behavior of salts forms. On the other hand, for a synthesis in several steps, it is useful to study the behavior of each step in order to choose the step exhibiting a conglomerate behavior resulting in an efficient enantiomeric resolution. In the case of racemic compound, the entrainment technique can be used in those cases where the eutectics are situated very close to the racemate. This underscores the value of systematic searches for derivatives that form conglomerates or at least racemic compounds whose eutectics are close to that of the racemate.

Another further way of purification is the formation of diastereomeric salts. However, partial solid solution are often observed and it is difficult to achieve high purity. Resolution via diastereomeric salt formation based on a part of the DSC curve has been discussed (102).

The change from racemates to enantiomers has implication in the galenical formulations. Propanolol-base enantiomer has a better skin permeation than racemate, what was explained by a lowering of the melting point (103). The stereospecific preformulation of ibuprofen has also been discussed. The theoretical phase diagram has been calculated from the DSC curves of the (+) enantiomer and of the racemate by using the equations of Schröder–Van Laar and Prigogine–Defay equations. Experimental data confirm the calculated data (104, 105).

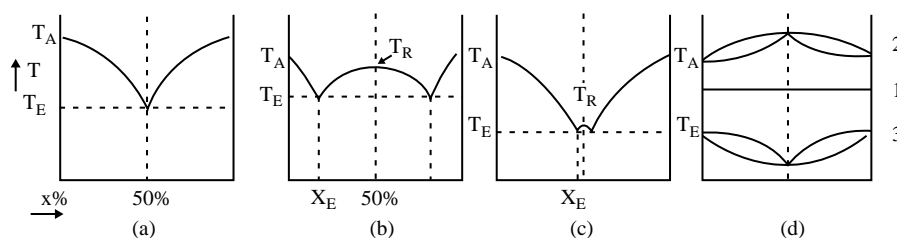


Fig. 16 Phase diagrams for enantiomers: (a) conglomerate, (b) and (c) racemic compound, (d) solid solution.

The low melting point of the enantiomer may have implications of the pharmaceutical process. Relationships between physical properties and crystal structures of chiral drugs have been discussed (106).

APPLICATION OF THERMAL TECHNIQUES FOR THE DRUG PRODUCT: PREFORMULATION, PROCESSING, AND AGING

The preformulation includes the choice of the salt form and of the polymorph of the drug substance. Melting points, solubilities, dissolution, hygroscopicity, stability, feasibility, processability, and polymorphic behavior have to be considered. The second step is to study the behavior of the drug substance with excipients (107).

Excipients

The use of thermal analysis techniques for pharmaceuticals implies the knowledge of the thermal behavior of the excipients. A great number of publications deal with polymorphic behavior of excipients and especially with the amorphous forms as prerequisite knowledge for freeze drying and milling processes.

Lactose exists in two isomeric forms α and β . It is possible to obtain the α -monohydrate, anhydrous crystalline α and β forms, as well as an amorphous form. The pharmaceutical properties of these various types are different. The hardness of tablets obtained using amorphous lactose produced by lyophilization is 10 times that obtained using crystalline forms. During milling, it has been observed that the monohydrate loses part of its water of crystallization and of its crystallinity. Heat treatment of different lactoses has permitted the discovery of an anhydrous, unstable α form and a crystal containing α and β in the ratio 1:1. Under the influence of high degrees of humidity, amorphous lactose crystallizes and anhydrous forms tend to reconvert to the monohydrate (68, 108, 109).

The thermal analysis studies of different forms of sorbitol, mannitol, glucose, magnesium stearate are reviewed in (68). Mono-, di-, and trihydrate of magnesium stearate as well as the amorphous form may be found. According to Wada et al. (110), thermal analysis is the most appropriate method for characterization of magnesium stearate.

Thermal analysis are widely used for polymers and copolymers analysis (4). Glass transitions, melting, and decomposition processes are analyzed. Since the glass transition temperature T_g is marked by changes in

the thermal capacity, expansion coefficient, and rigidity, TMA technique as well as DSC may be used. T_g increases with molecular mass up to certain values. Plasticizers and water depress this temperature. Thermal stability and influence of antioxidants and fillers may be analyzed by TG or DSC, under oxygen.

The compatibility of polymers in blends is tested by comparing DSC curves of the components. Immiscible crystalline blends such as polyethylene–polypropylene show the DSC peaks of polyethylene and polypropylene. Immiscible amorphous blends exhibit two glass transitions; in miscible blends a new glass point is observed. Partially miscible blends have two glass points situated between the glass points of each polymer. Different equations such as Gordon–Taylor express the relation between the new glass transition T_g and the glass transition points T_{g1} and T_{g2} of components. W_1 and W_2 are the weight fractions and K is the ratio $\Delta C_{p2}/\Delta C_{p1}$.

$$T_g = W_1 T_{g1} + K W_2 T_{g2} / W_1 + K W_2$$

The polymers mostly used in pharmaceutical packaging are polyethylene, polypropylene, PVC, polyamide, polystyrol, nylon, cellulose acetate, polyethylene terephthalate, and blends thereof. Copolymers and rubbers are also used. The DSC melting curve of polyethylene used for packaging purposes is characteristic. Low- and high-density polyethylene are differentiated by their melting points (111). Melting point and density of polyethylene are linearly correlated (112). Crystallinity may be determined as described above for amorphous state.

Polyethylene glycols (PEG) have been intensively characterized by thermal analytic methods. The melting points of PEG increase with the molecular weight and decreases with the content of water as a result of eutectic formation (10). Corrigan (113) studied the different DSC peaks of PEG: Once folded chain crystals and extended chain crystals are present in PEG 6000, which results in two DSC peaks. In PEG of higher molecular mass, only folded chain crystals are present. Lower molecular mass PEG contain only extended chain crystals. Craig reviewed the thermal studies of PEG, including properties in aqueous solutions (114, 115). Phase diagram of PEG 4000 was recently discussed (116). Higher molecular mass PEG are generally used in solid dispersion systems and their melting behavior is relevant for the temperature of galenical preparation (5, 117).

Glass transitions of polyvinylpyrrolidones of different molecular weight may be used as identity (10, 118). Water depresses the glass transitions and effects the physical properties of polyvinylpyrrolidone, as demonstrated by Tan and Challa (119).

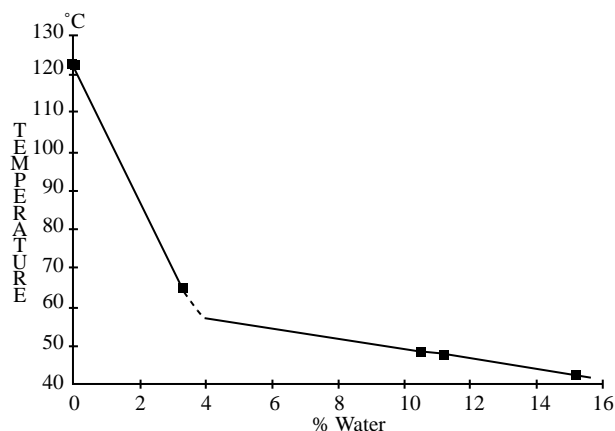


Fig. 17 Influence of the water content on the glass transition temperature of HPMC 4000 measured by DSC in sealed pans.

A large number of cellulose derivatives have been studied for food as well as for pharmaceutical applications (120–123). The measurement of the glass transition is often difficult due to the broad endotherm of dehydration of water. Figure 17 shows the effect of water on the depression of the glass transition of hydroxypropyl methylcellulose HPMC4000.

Film coating processes need the knowledge of the glass transitions for a proper film. Ethylcellulose, cellulose phthalate, polyvinylalcohols, polymethyl methacrylates have been studied and the critical parameters for film formation discussed (124–126).

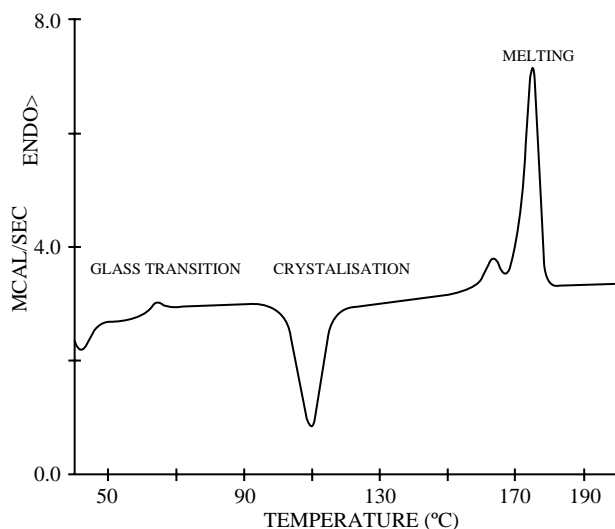


Fig. 18 DSC curve of L-poly(lactic acid) after quenching from the melt. (From Ref. 10.)

Polysaccharides and water interaction are especially studied for their use in spray-dried products in food and in biotechnology (127–129).

Biodegradable polymers can be crystalline or amorphous. Poly-L-lactic and D,L-poly(lactic acids) have been studied by Pitt and Gu (130). Figure 18 deals with the second run of poly-L-lactic acid after quenching. The glass transition is followed by crystallization, the melting of the crystalline form. Aging and crystallinity of biodegradable polymers have been studied by Akhtar et al. (131). TG is useful for the determination of the entrapped solvent (often methylene chloride) and of moisture (5, 10). Microspheres containing biodegradable polymers are intensely analyzed by thermal analysis techniques allowing both the glass transition point of the polymer and the physico-characterization of the drug substance to be analyzed; thermal techniques generally with electron microscopy are used for the optimization of the drug loading and of the process (132). The physical aging of the polymer can be assessed in DSC by the amplitude associated with the glass transition of the matrix. This relaxation energy increases with aging. Rosilio studied the progesterone poly(D,L-lactide-co-glycolide) microspheres. Aging acts on the solid state of the drug substance loaded. Different polymorphs are obtained, depending on the copolymer composition (133).

Phase Diagrams

The thermodynamic phase diagrams are the basis for understanding DSC curves of formulation: eutectics, solid solutions, eutectics with partial solid solutions, and compound formation with congruent or incongruent melting. Figure 19 exemplifies the building of the phase diagram of propyphenazone and butesamide, using the DSC curves. In order to save the number of DSC scans, theoretical curves can be added (134). Such phase diagram was performed between a drug substance and stearic acid: no formation of the salt was observed (13). Recently thermal analysis has been used for the complex phase diagram of propanol/oleic acid for which the salt and a mesomorphic phase have been found (135).

DSC was proposed for compatibility studies comparing DSC curves of components and mixtures. Unfortunately some misinterpretation may occur. DSC curves can only reflect physical behavior. The formation of eutectic is not an incompatibility. Furthermore, water generally is not present in the mixtures at the temperatures of the melting peaks. Giron et al. (9) performed the DSC curves of a drug substance and several excipients, the initial DSC curves of the mixtures and the DSC curves of the mixtures after 1 month at 50°C/<30% RH and 50°C/75% RH. Chromatography was used for the chemical analysis

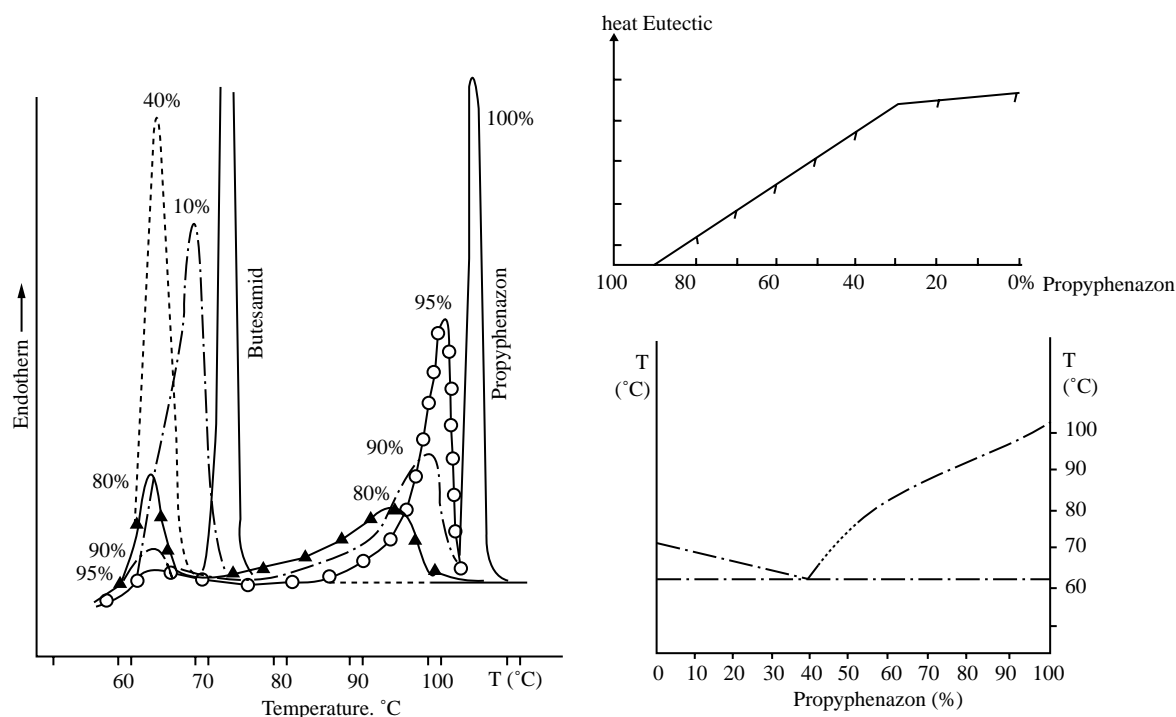


Fig. 19 Example of building a phase diagram by DSC of propyphenazone and butesamide. (From Ref. 6.)

of the drug. Figure 20 shows the comparison of the DSC curves of the mixtures. The effect of the humidity on the degradation is very striking: The drug substance is completely decomposed with calcium sulfate, avicel, and stearic acid. For talc and dicalcium phosphate, there is no degradation. The observation of the initial DSC curves for the five excipients would have wrongly given the conclusion to the same compatibility.

For such compatibility studies, isothermal microcalorimetry has been suggested (e.g., Ref. 136).

The DSC study of mixtures of drug and excipients is very useful, for the information gained (e.g., if the eutectic melts at ambient temperature). In other cases, one may target interaction with excipients as it is the case of solid dispersions, solid solutions, or complex formation. The choice of the carrier defines the characteristics of dissolution of the dispersed drug. Poorly water-soluble active ingredients are combined with water-soluble carriers in order to increase the dissolution of the active ingredient. A good water-soluble drug combined with a slightly soluble carrier leads to a retardation of drug release from the matrix. For reviews on solid dispersions, see Refs. 5 and 137. For the development of solid dispersions, the method of preparation, the type of systems, the dissolution characteristics, and the aging problems should be considered. DSC is appropriate for the study of solid dispersions by comparing DSC curves of pure compounds, of physical mixtures or

melted mixtures to the DSC curve of the solid dispersion. TG, X-Ray diffraction, and hyphenated techniques are good complements. The transitions observed are those transitions expected from the phase diagrams, or the amorphous state with glass transition or very often new metastable forms of drug substance as well as metastable forms of excipients.

The phase diagram given in Fig. 21 for darodipine–polyethylene glycol 6000 results from DSC experiments carried out by two techniques (10). The DSC of physical mixtures obtained by grinding were scanned just after the end of melting. After cooling a second scan at 5 K min^{-1} was performed. In the second technique the mixtures were dissolved in methanol and the solvent evaporated. With both techniques, the same results were obtained.

The knowledge of the phase diagram allows the choice of the temperature of the process. It can be suitable to choose higher loading than the eutectic composition. Once the solid dispersion manufactured and milled, DSC is very advantageous for checking the batch reproducibility or the stability behavior by the measurement of the melting enthalpy of the eutectic. For a solid dispersion with 40% darodipine, a standard deviation of 1.4% for the melting point of the eutectic located at 58°C and 2.4% for the heat of eutectic of 27 cal g^{-1} was found.

PEG used for solid dispersion vary from molecular weight 1000 to molecular weight 20,000. PEG of lower

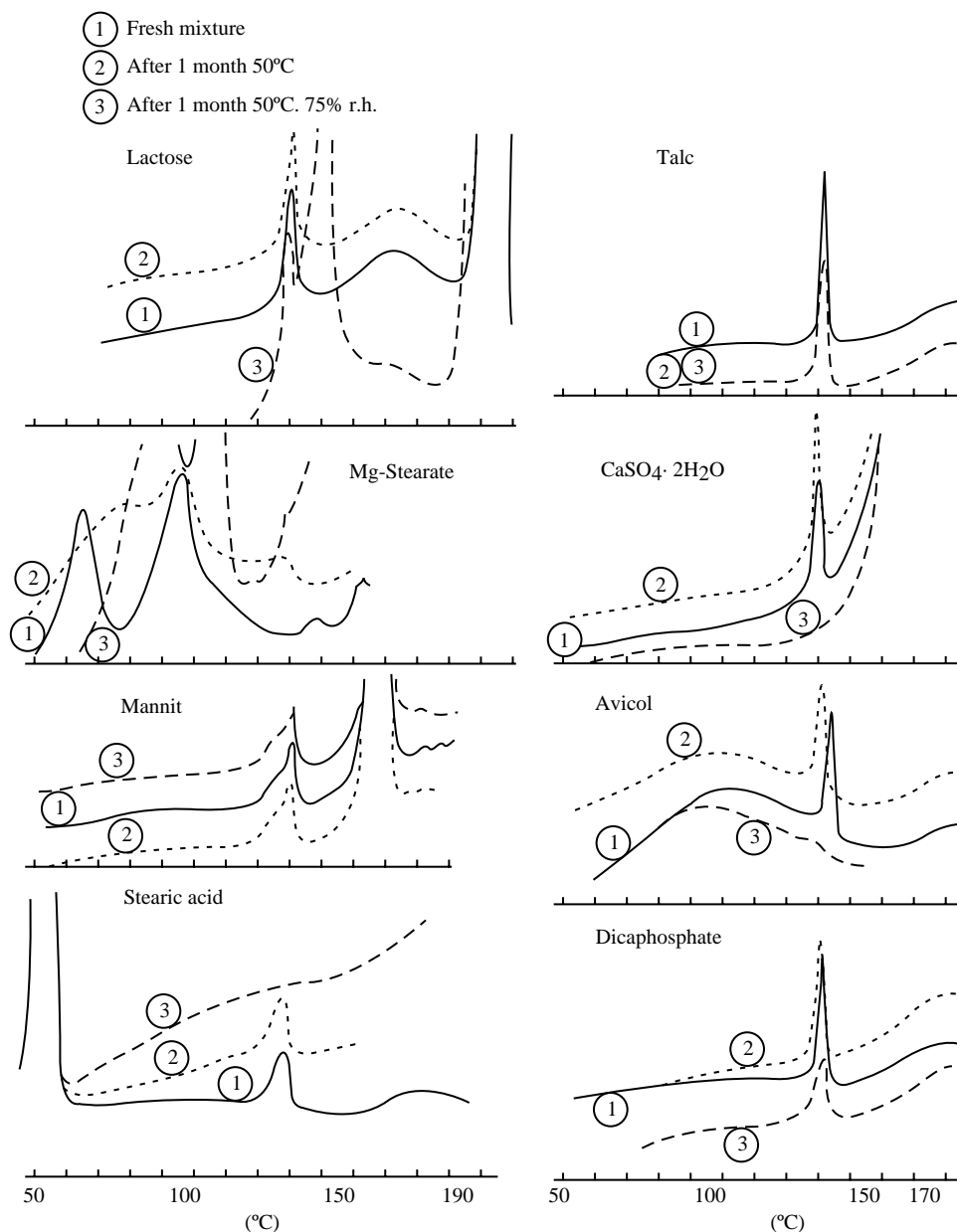


Fig. 20 DSC curves of 10% drug substance with excipients: 1: Original curves; 2: curves after 1 month at 50°C/<30% RH; and 3: 50°C/75% RH. The initial DSC curves suggest an incompatibility only with mannitol and magnesium stearate. The DSC curves (curves 3) after storage at 50°C/75% RH allow the differentiation of the excipients: degradation with stearic acid, calcium sulfate dihydrate, and avicel.

molecular weight are liquid and therefore not suitable. The quality mostly used are PEG 4000 and PEG 6000. The composition of the eutectic of most drug substances lies in small concentration of drug substance. Some monotectics (0% drug substance) have been described. The highest amount of drug is obtained with glutethimide with an eutectic composition of 32% (5).

Solid solutions: The drug and the carrier are miscible in the solid state. Polyvinylpyrrolidones (PVP) with different molecular weights dissolve drug substances such as diazepam (138) through hydrophobic interactions. The disparition of the DSC peak of the drug substance demonstrates the formation of the solid solution.

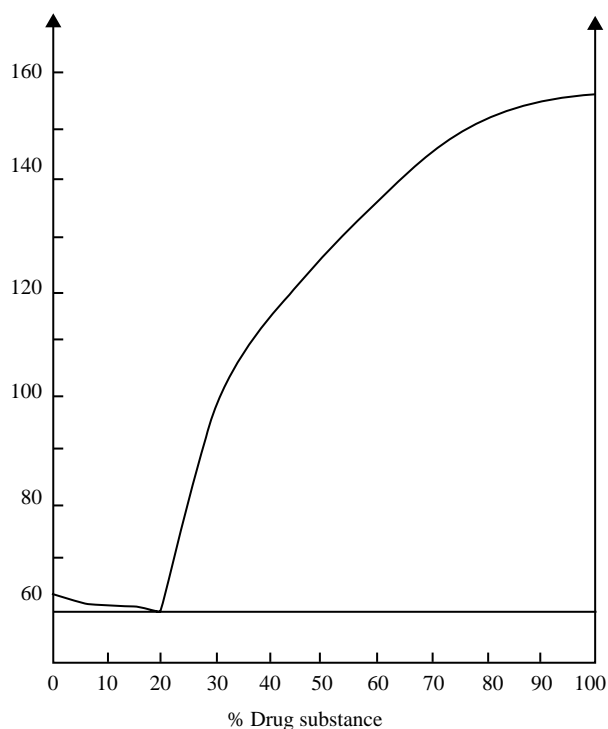


Fig. 21 Phase diagram of a solid dispersion of darodipine-PEG 6000.

Drug substance and carrier can be both in the glassy state. Total miscibility, partial miscibility, or total immiscibility has been observed. This type of system is very sensitive to temperature and moisture and crystallizations are often observed after long storage.

New derivatives of starch, mainly cyclodextrins (139) which forms inclusion compounds, are best carriers for the monitoring of the dissolution characteristics of the drug. These cyclic oligosaccharides contain six (α -CyD), seven (β -CyD), and eight (γ -CyD) α -(1,4)-linked glucose units. A great number of chemically modified cyclodextrins have been manufactured in the last decade. Complexes are formed through inclusion in the cavity or through interactions with chemical groups. In order to obtain complexation, the compounds have to bind a complex first in solution. From the usual methods of preparation, kneading, coprecipitation, freeze drying, or spray drying, often the spray-drying technique gives the best results.

Giordano et al. (140) suggested a method of calculation of the ratio guest/host. They performed DSC analysis of dispersions of different compositions with an excess of the guest molecule. The remaining energy in the melting peak of the guest molecule allows the calculation of the amount of free drug. They plot this amount for different compositions versus the total guest fraction and

compared the plot obtained with the theoretical plots for ratios 1:1, 1:2, 1:3.

In most cases the DSC peak of the drug substance disappears and no new peak is observed. With hydrocortisone butyrate (141), a new peak corresponding to the complex has been observed. For this drug, the complexation increases in the order α -CyD, β -CyD, γ -CyD, and dimethyl- β -CyD. Thermogravimetry analysis has also been used (142). The temperatures of vaporization, sublimation, or degradation of the drugs are displaced to higher temperatures, due to the complexation.

Analysis of the Drug Product

If the components are not miscible in the solid and the liquid states, their DSC peaks remain unchanged. This allows to identify components of the drug product, to follow aging problems with polymorphism and in favourable cases to quantify the drug substance and the excipients (11, 20, 143, 144).

Fatty Acids and Glycerides Derivatives

Most fats and similar compounds show polymorphism behavior, including fatty acids, fatty alcohols,

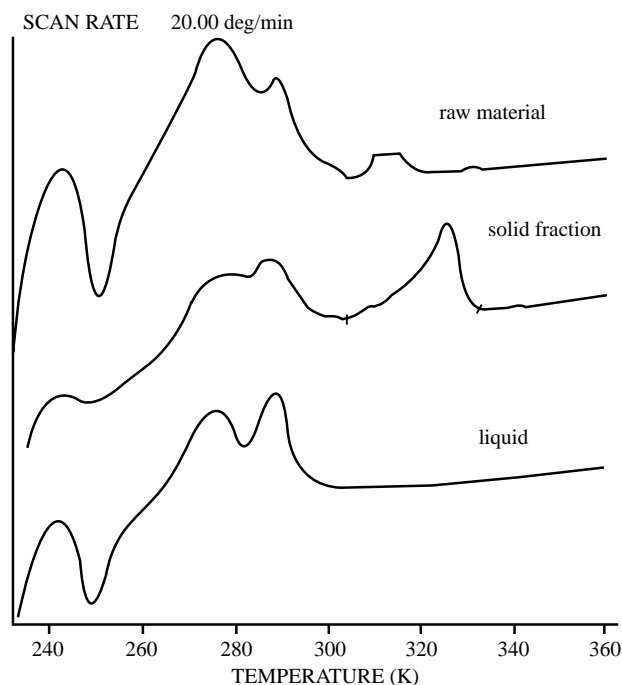


Fig. 22 DSC curves for the monitoring of fractionation of a liquid excipient. Scan rate 20 K min⁻¹.

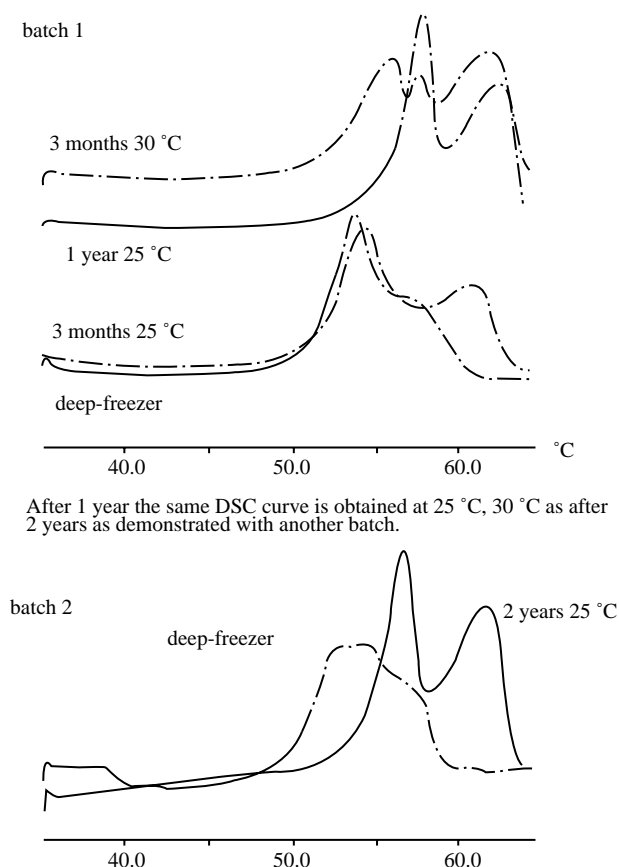


Fig. 23 Aging of precirrol.

cetostearyl alcohols, glycerides, oils, hydrogenated or transesterified oils, and suppositories (145). This polymorphic behavior is characterized by change of melting by aging, giving rise to hardening effects of suppositories and precipitations of liquid excipients that are complex mixtures of glycerides. Examples of aging problems are given in (10). Fig. 22 is an example of the control of the fractionation of a liquid excipient, using DSC. Aging of the precirrol excipient is demonstrated in Fig. 23. Sub ambient DSC was also used for the preformulations of microemulsions (11–13). The solid fat index, calculated as a percentage of the solid as function of the temperature, is very valuable for the evaluation of suppositories (146). Analysis of the drug is also possible. Chemical reaction was observed for aminophylline (147). Since most fatty acid derivatives are unsaturated, they are very sensitive to oxidation. The oxidability and the influence of antioxidants can be measured by DSC and TG, comparing the starting of the degradation (148).

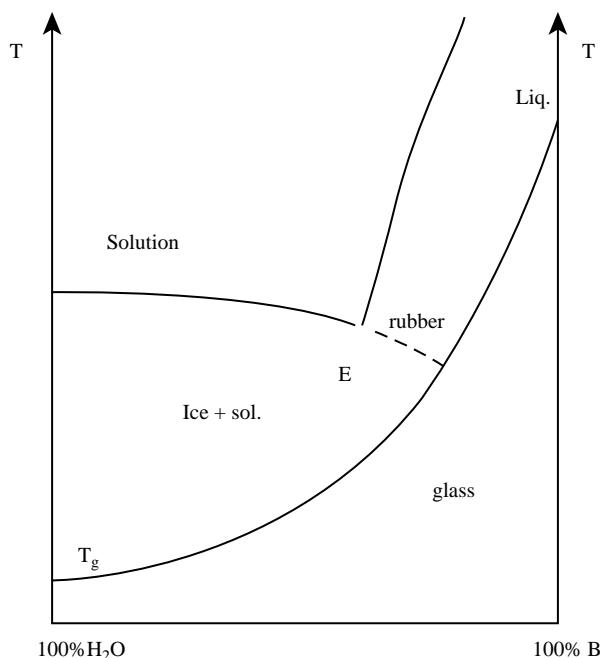


Fig. 24 Phase diagram of water and amorphous substance during freeze drying.

Interaction with Water

The phase diagrams of drug substance and excipients with water as well as the study of the temperature of glass transition (see Fig. 24) are the basis of the choice of the conditions of freeze-dried or spray-dried formulations (149–152). The lyophilizates are well characterized by DSC and TG. The polymorphic behavior of all components can be studied by thermal techniques. For proteins, it is suitable to have excipients in the formulation that remains amorphous. Trehalose was found to be a very efficient lyoprotectant (153).

Freezable water is determined by the measurement of the melting peak of ice. This technique is currently applied for creams, and gels (154, 155). Examples of the melting peak of freezable water are given in Fig. 25. Swelling properties of modified release matrices and their interaction during dissolution have been studied (156, 157).

Monoglycerides–water systems have been characterized (158). Three classes of phases (lamellar, cubic, and hexagonal) have been determined. Water/oil systems and their aging on the phase transitions have been studied using DSC (159, 160). Thermogravimetry can be successfully used additionally (161).

Liposomes are multilayered vesicles consisting of concentric bilayers of phospholipids interdispersed with

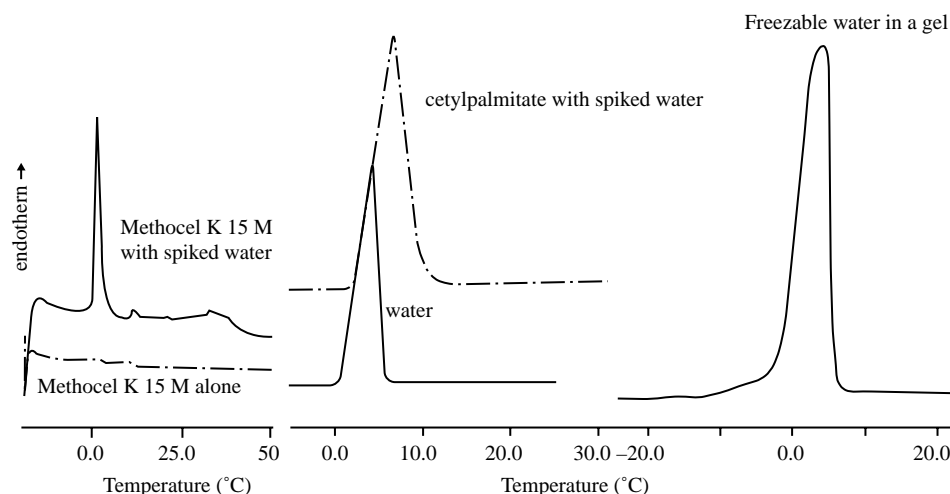


Fig. 25 Determination of freezable water, using the melting peak of ice. (a) Methocel K15 M, (b) cetyl palmitate, and (c) pharmaceutical gel.

aqueous phases. In aqueous media, the phospholipids undergo gel, liquid crystalline transitions easy to detect by DSC. The study of the change of these transitions, temperature, peak width, and energy allows to characterize the hydrated phospholipid bilayers and to study the liposome formation with drugs (162–164). Drug targeting studies with phospholipids membranes have been proposed using DSC technique (165).

CONCLUSION

Thermal analysis methods are widely used in all fields of pharmaceuticals. They are unique for the characterization of single compounds. The information correlated with the thermodynamic phase diagrams is extremely helpful for rational preformulation and development of new delivery systems. Very rapid and requiring only very small samples of material, these methods are applicable in development and also in production for quality control. The combination with spectroscopic and crystallographic data will allow better insight in complex phase changes behavior.

REFERENCES

1. Wendlandt, W.W. *Thermal Analysis. Chemical Analysis*; 3rd Ed., Elving, P.J., Winefordner, J.D., Kolthoff, I.M., Eds.; John Wiley: New York, 1986; 19.
2. Haines, P.J. *Thermal Methods of Analysis, Principles, Applications and Problems*; Blackie, Academic Professional: London, 1995.
3. Wunderlich, B. *Thermal Analysis*; Academic Press: New York, 1990.
4. Turi, E.A. *Thermal Characterization on Polymeric Materials*; 2nd Ed.; Academic Press: New York, 1997.
5. Ford, J.L.; Timmins, P. *Pharmaceutical Thermal Analysis Techniques and Applications. Series in Pharmaceutical Technology, Ellis Horwood Books in Biological Sciences*; Rubinstein, M.H. Ed.; John Wiley & Sons: New York, 1989.
6. Giron-Forest, D. *Thermoanalytische Verfahren. Pharmazeutischer Qualitätskontrolle*; Feltkamp, H., Fuchs, P., Sucker, H., Eds.; Stuttgart, Leipzig Germany, 1983; 298–310, Georg Thieme Verlag.
7. Barnes, P.A. Applications of New Methods and Instrumentation in Thermal Analysis. *Thermochim. Acta* **1987**, *114*, 1–8.
8. Brennan, W.P. Some Applications of Thermal Analysis as a Supplement to or Replacement for ASTM Testing Standards. *Thermochim. Acta* **1977**, *18*, 10–13, and 101–111.
9. Giron, D. Applications of Thermal Analysis in the Pharmaceutical Industry. *J. Pharm. Biochem. Anal.* **1986**, *40*, 755–770.
10. Giron, D. Thermal Analysis in Pharmaceutical Routine Analysis. *Acta Pharm. Jugosl.* **1990**, *40*, 95–157.
11. Giron, D. Thermal Analysis of Drugs and Drug Products. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1995; 15, 1–79.
12. Giron, D. Thermal Analysis, Microcalorimetry and Combined Techniques for the Study of Pharmaceuticals. *J. Therm. Anal. Calorim.* **1999**, *56*, 1285–1304.
13. Giron, D. Contribution of Thermal Methods and Related Techniques for Rational Development of Pharmaceuticals. *P.S.S.T.* **1998**, *1*, 191–262.
14. Cheng, S.Z.D.; Li, C.Y.; Calhoun, B.H.; Zhu, L.; Zhou, W.W. Thermal Analysis: The Next Two Decades. *Thermochim. Acta* **2000**, *355*, 59–68.

15. Thompson, K.C. Pharmaceutical Applications of Calorimetric Measurements in the New Millenium. *Thermochim. Acta* **2000**, 366, 83–87.
16. Sarge, S.M.; Gmelin, E. Temperature, Heat and Heat Flow Rate Calibration of Differential Scanning Calorimeters. *Thermochim. Acta* **2000**, 347, 9–13.
17. Sarge, S.M.; Hohne, G.W.H.; Cammenga, H.K.; Eysel, W.; Gmelin, E. Temperature, Heat and Heat Flow Rate Calibration of Scanning Calorimeters in the Cooling Mode. *Thermochim. Acta* **2000**, 361, 1–20.
18. Sabbah, R.; Xu-wu, A.; Chickos, J.S.; Leita, M.L.; Roux, M.V.; Torres, L.A. Reference Materials for Calorimetry and Differential Thermal Analysis. *Thermochim. Acta* **1999**, 331, 93–204.
19. Giron, D.; Goldbronn, C.; Piechon, P. Thermal Analysis Methods for Pharmacopea Materials. *J. Pharm. Biochem. Anal.* **1989**, 7, 1421–1430.
20. Giron, D. Characterization of Pharmaceuticals by Thermal Analysis. *Am. Pharm. Rev.* **2000**, 3(2), 53–61, 3 (3), 43–53.
21. Han, J.; Gupte, S.; Suryanarayanan, R. Applications of Pressure Differential Scanning Calorimetry in the Study of Pharmaceutical Hydrates. II. Ampicillin Trihydrate. *Int. J. Pharm.* **1998**, 170, 63–72.
22. Gill, P.S.; Sauerbrunn, S.R.; Reading, M. Modulated Differential Scanning Calorimetry. *J. Thermal. Anal.* **1993**, 40, 931–939.
23. Hohne, G.W.H. Modulated DSC. *Thermochimica Acta* **1999**, 330, 45–51.
24. Coleman, N.J.; Craig, D.Q.M. Modulated DSC: A Novel Approach to Pharmaceutical Thermal Analysis. *Int. J. Pharm.* **1996**, 135, 13–29.
25. McPhillips, H.; Craig, D.Q.M.; Royall, P.G.; Hill, V.L. Characterization of the Glass Transition of HPC Using Modulated DSC. *Int. J. Pharm.* **1999**, 180, 83–90.
26. Craig, D.Q.M.; Barsnes, M.; Royall, P.G.; Kett, V.L. An Evaluation of the Use of Modulated Temperature DSC as a Means of Assessing the Relaxation Behavior of Amorphous Lactose. *Pharm. Res.* **2000**, 17(6), 696–700.
27. Bottom, R. The Role of MTDSC in the Characterisation of a Drug Molecule Exhibiting Polymorphic and Glass Forming Tendencies. *Int. J. Pharm.* **1999**, 192(1), 47–53.
28. Parkes, G.M.B.; Barnes, P.A.; Bond, G.; Charsley, E.L. Qualitative and Quantitative Aspects of Microwave Thermal Analysis. *Thermochim. Acta* **2000**, 356, 85–96.
29. Olson, E.A.; Efremov, M.Y.; Kwan, A.T.; Lai, S.; Petrova, V. Scanning Calorimeter for Nanoliter-Scale Liquid Samples. *Appl. Phys. Lett.* **2000**, 77(17), 2671–2673.
30. Buckton, G. Application of Isothermal Microcalorimetry in the Pharmaceutical Sciences. *Thermochim. Acta* **1995**, 248, 117–129.
31. Wadsö, I. Isothermal Calorimetry. *Thermochim. Acta* **1999**, 294, 1–11.
32. Charsley, E.L.; Warne, S.St.; Warrington, S.B. Thermogravimetric Apparatus Temperature Calibration Using Melting Point Standards. *Thermochim. Acta* **1987**, 114, 53–61.
33. McGhie, A.R.; Chiu, J.; Fair, P.G.; Blaine, R.L. Studies on ICTA Reference Materials Using Simultaneous TG-DTA. *Thermochim. Acta* **1983**, 67, 241–250.
34. Komatsu, H.; Yoshii, K.; Okada, S.; *Chem. Pharm. Bull.* **1994**, 42, 1631–1635.
35. Giron, D.; Goldbronn, C.; Pfeffer, S. Automation in Thermogravimetry: Application in Pharmaceutical Industry Poster Presented at the 4th Symposium on Pharmacy and Thermal Analysis. Karlsruhe, 1999.
36. Dunn, J.G.; Sharp, J.H. *Treatise on Analytical Chemistry*, Kolthoff, I.M. Ed.; 2nd Ed.; Part 1 John Wiley & Sons: New York, 1993; 13, 127–267.
37. Shlieout, G.; Zessin, G. Investigation of the Internal Stress of Ethylcellulose Tablets by TMA. *Pharmazie* **1997**, 52, 713–717.
38. Ozawa, T. Kinetic Analysis by Repeated Temperature Scanning. Part 1. Theory and Methods. *Thermocim. Acta* **2000**, 356, 173–180.
39. Nakamura, K.; Kinoshita, E.; Hatakeyama, T.; Hatakeyama, H. TMA Measurement of Swelling Behavior of Polysaccharide Hydrogels. *Thermochim. Acta* **2000**, 352, 171–176.
40. Knop, K.; Mattheé, K. Quellungsmessungen Von dünnen Polymerfilmen Mittels TMA. User Com (Mettler Toledo Application) **1998**, 2, 9–10.
41. Syler, R.; ASTM Spec. Techn. Publ. **1991**, 1136, 22–31, Matsumari N., ASTM Spec. Techn. Publ., 1991, 1136, 32–48.
42. Craig, D.M.Q.; Johnson, F.A. DMA for Pharmaceuticals. *Thermochim. Acta* **1995**, 248, 97–115.
43. Kühnert-Brandstätter, M. *Thermomicroscopy in the Analysis of Pharmaceuticals Intern. Ser. Monog. Anal. Chem.*, Belcher, R., Freiser, M. Eds.; Pergamon Press: Oxford, 1971; 45.
- 43a. Kühnert-Brandstätter, M. *Thermomicroscopy of Organic Compounds in Comprehensive Analytical Chemistry*; Svehla, G., Ed.; Elsevier Publisher: Amsterdam, 1982; XVI, 329–428.
44. Cooke, P.M. Chemical Microscopy. *Anal. Chem.* **1996**, 68, 333R–378R.
45. Vitez, I.M.; Newman, A.W.; Davidovich, M.; Kiesnowski, C. The Evolution of Hot Stage Microscopy to Aid Solid-state Characterizations of Pharmaceutical Solids. *Thermochim. Acta* **1998**, 324, 187–196.
46. Kellner, R.; Kühnert-Brandstätter, M.H.; Malissa, H. FTIR-Microscopic Investigations of Microphases and Microphase-Transitions in Organic Substances. *Mikrochim. Acta* **1998**, 2, 133–137.
47. Moss, R. La Spectroscopie Raman à Transformée De Fourier Dans Un Laboratoire d'analyse Industrielle. *Analisis Magazine* **1994**, 22, M17–M22.
48. Szelagiewicz, M.; Marcolli, C.; Cianferani, S.; Hard, A.P.; Vit, A.; Burkhard, A.; von Raumer, M.Ch.; Hofmeier, Ch.A.; Zilian, A.E.; Francotte, E.; Schenker, R. In Situ Characterization of Polymorphic Forms. The Potential of Raman Techniques. *J. Therm. Anal.* **1999**, 57, 23–43.
49. Griesser, U.J.; Auer, M.E.; Burger, A. Micro-Thermal Analysis, FTIR-and Raman-Microscopy of (R,S)-Proxiphylline Crystal Forms. *Microchem. J.* **2000**, 65(3), 283–292.
50. Craig, D.Q.M.; Royall, P.G.; Reading, M.; Price, D.M.; Lever, T.J.; Furry, J. Micro-Thermal Analysis for the Characterization of Pharmaceutical Materials. *Proc. Conf. North Am. Thermal Anal. Soc.*; 26th Omnipress: Madison Wisconsin, 1998; 610–615.
51. Epple, E.; Cammenga, H.K. Temperature Resolved X-ray Diffractometry as a Thermoanalytical Method: A

- Powerful Tool for Determining Solid State Reaction Kinetics. *J. Therm. Anal.* **1992**, *38*, 619–626.
52. Conflant, P.; Guyot-Hermann, A.M. Contribution of X-ray Powder Diffraction Versus Temperature to the Solid State Study of Pharmaceutical Raw Materials. *Eur. J. Pharm. Biopharm.* **1994**, *40*, 388–392.
53. Giron, D.; Draghi, M.; Goldbronn, C.; Pfeffer, S.; Piechon, P. Study of the Polymorphic Behaviour of Some Local Anesthetic Drugs. *J. Therm. Anal.* **1997**, *49*, 913–927.
54. Giron, D. Investigation of Polymorphism and Pseudo-Polymorphism in Pharmaceuticals by Combined Thermo-analytical Techniques. *J. Therm. Anal. Calorim.* **2000**, ICTAC, in press.
55. Cavatur, R.K.; Suryanarayanan, R. Characterization of Phase Transformations during Freeze-drying by In Situ X-Ray Powder Diffractometry. *Pharm. Dev. Technol.* **1998**, *3*, 579–586.
56. Ashizawa, K.; Netsu Sokutei **1998**, *25*, 97–104.
57. Arai, T.; Kishi, A.; Kobayashi, Y. Coupled DSC and X-Ray Diffraction Instrument. *Thermochim. Acta* **1999**, *325*, 151–156, Rigaku Application Note Presented at the 48th Annual Denver X-ray Conference, 1999.
58. Cair, M.R. Crystalline Polymorphism of Organic Compounds. *Top. Curr. Chem.* **1998**, *198*, 163–208.
59. Materazzi, F. Application of Coupled TG-IR. To the Polymorphism of Pharmaceuticals. *Appl. Spectrosc. Rev.* **1997**, *32*, 385.
60. Ozawa, T. Application of TG-MS. *Mass Spectrom. Soc.* **1998**, *11*, 52.
61. Breen, C.; Last, P.M.; Taylor, S.; Komadel, P. Synergic Chemical Analysis—The Coupling of TG with FTIR, MS and GC-MS. *Thermochimica Acta* **2000**, *363*, 93–104.
62. Decision Tree: Investigating the Need to Set Acceptance Criteria for Polymorphism in Drug Substances and Drug Products International Conference on Harmonization (ICH) Guideline Specification Q6A, 1999.
63. Halebian, J.; McCrone, W.J. Pharmaceutical Applications of Polymorphism. *Pharm. Sci.* **1969**, *58*, 911–929.
64. Halebian, J.K. Characterization of Habits and Crystalline Modifications of Solids and Their Pharmaceutical Applications. *Pharm. Sci.* **1975**, *64*, 1269–1288.
65. Giron, D. Le Polymorphisme. *Labo-Pharma-Probl. Techn.* **1981**, *307*, 151–160.
66. Masse, J.; Bauer, M.; Billot, P.; Broquaire, M.; Chauvet, A.; Doveze, J.; Garinot, O.; Giron, D. Etude Du Polymorphisme Et Du Pseudopolymorphisme à L'aide Des Méthodes Thermo-Analytiques, Rapport d'une Commission SFSTP. *S.T.P. Pharma Prat.* **1997**, *7*, 235–246.
67. Giron, D. Thermal Analysis and Calorimetric Methods in the Characterization of Polymorphs and Solvates. *Thermochim. Acta* **1995**, *248*, 1–59.
68. Giron, D. Le Polymorphisme Des Excipients. *STP Pharma (Hors serie)* **1990**, *6*, 87–98.
69. Burger, A.; Ramberger, R. On the Polymorphism of Pharmaceuticals and Other Molecular Crystals. I. Theory of Thermodynamic Rules. *Mikrochim. Acta* **1979**, *2*, 259–271.
70. Burger, A.; Ramberger, R. On the Polymorphism of Pharmaceuticals and Other Molecular Crystals. II. Applicability of Thermodynamic Rules. *Mikrochim. Acta* **1979**, *2*, 273–316.
71. Burger, A. Thermodynamic and Other Aspects of the Polymorphism of Drugs. *Pharm. Int.* **1982**, *3*, 158–163.
72. Burger, A.; Henck, J.O. The Presentation of Polymorphic Systems by Energy/Temperature Diagrams. *Biopharm. Pharm. Technol.* **1995**, *1*, 10–11.
73. Toscani, S. An Up-To-Date Approach to Drug Polymorphism. *Thermochim. Acta* **1998**, *321*, 73–79.
74. Jacques, J.; Collet, A.; Wilen, S.H. *Enantiomers, Racemates and Resolutions*; John Wiley, 1981.
75. Bernstein, J.; Davey, R.J.; Henck, J.O. Concomitant Polymorphs. *Angew. Chem. Int. Ed.* **1999**, *38*, 3340–3461.
76. Coquerel, G. Review on the Heterogeneous Equilibria Between Condensed Phases in Binary Systems of Enantiomers. *Enantiomer* **2000**, *5*, 481–498.
77. McCauley, J.A. Detection and Characterization of Polymorphism in the Pharmaceutical Industry. *AIChE Symp. Ser.* **1991**, *87*, 58–63.
78. Giron, D.; Piechon, P.; Goldbronn, C.; Pfeffer, S. Thermal Analysis, Microcalorimetry and Combined Techniques for the Study of the Polymorphic Behavior of a Purine Derivative. *J. Therm. Anal. Calorim.* **1999**, *57*, 61–72.
79. Hancock, B.C.; Zografi, G. The Relationship Between the Glass Transition Temperature and the Water Content of Amorphous Solids. *Pharm. Res.* **1994**, *11*, 471–417.
80. Hancock, B.C.; Zografi, G. Characteristics and Significance of the Amorphous State in Pharmaceutical Systems. *J. Pharm. Sci.* **1997**, *86*, 1–12.
81. Shalae, E.Y.; Zografi, G. How Does Residual Water Affect the Solid State Degradation of Drugs in the Amorphous State. *J. Pharm. Sci.* **1996**, *85*, 1137–1141.
82. Hancock, B.; Dalton, C.R.; Pikal, M.J.; Shamblin, S.L. A Pragmatic Test of a Simple Calorimetric Method for Determining the Fragility of Some Amorphous Pharmaceutical Materials. *Pharm. Res.* **1998**, *15*, 762–767.
83. Andronis, V.; Zografi, G. Crystal Nucleation and Growth of Indomethacin Polymorphs from the Amorphous State. *J. Non-Cryst. Solids* **2000**, *271*, 236–248.
84. Kontny, M.J.; Zografi, G. Sorption of Water by Solids. *Drugs Pharm. Sci.* **1995**, *70*, 387–418.
85. Di Martino, P.; Piva, F.; Conflant, P.; Guyot-Hermann, A.M. Thermal Analysis and Powder X-Ray Diffraction Study of Terpin. Evidence of a Eutectic (Hydrate/Anhydrous Form). *J. Thermal Anal. Calorim.* **1999**, *57*, 95–111.
86. Soustelle, M. Physico-Chemical Transformation of Powders. *Powder Technology and Pharmaceutical Processes*; Chulia, D., Deleuil, M., Pourcelot, Y. Eds.; Elsevier Science: Amsterdam, 1994; 27–58.
87. Morris, K.R. Structural Aspects of Hydrates and Solvates. *Drugs Pharm. Sci.* **1999**, *95*, 125–181, Polymorphism in Pharmaceutical Solids; Brittain, H.G., Ed.; Marcel Dekker, Inc.: New York.
88. Kankhari, R.K.; Grant, D.J.W. Pharmaceutical Hydrates. *Thermochim. Acta* **1995**, *248*, 61–79.
89. McCauley, J.A.; Varsolona, R.J.; Levorse, D.A. The Effect of Polymorphism and Metastability on the Characterization and Isolation of Two Pharmaceutical Compounds. *J. Phys. D, Appl. Phys.* **1993**, *26*, B85–B89.

90. Giron, D.; Goldbronn, C. Use of Sub-Ambient DSC to Complement Conventional DSC And TG. *J. of Thermal Anal.* **1997**, *49*, 907–912, 1998; *51*, 727.
91. Byström, K. Thermometric Company. Application Note 22004 **1990**, .
92. Humera, A.; Buckton, G.; Rawlins, D.A. The Use of Isothermal Microcalorimetry in the Study of Small Degrees of Amorphous Content of Hydrophobic Powder. *Int. J. Pharm.* **1996**, *130*, 195–201.
93. Giron, D.; Remy, P.; Thomas, S.; Vilette, E. Quantitation of Amorphicity by Microcalorimetry. *J. Thermal Anal.* **1997**, *48*, 465–472.
94. Giron, D.; Goldbronn, C. Place of DSC Purity Analysis in Pharmaceutical Development. *J. Thermal. Anal.* **1995**, *44*, 217–251.
95. Neah, S.H.; Shinwari, M.K.; Hellmuth, E.W. Melting Point Phase Diagrams of Free Base and Hydrochloride Salts of Bevantolol, Pindolol and Propanolol. *Int. J. Pharm.* **1993**, *99*, 303–310.
96. Elsabee, M.; Frankerd, R.J. Solid State Properties of Drugs. DSC Of Chiral Drugs Mixtures Existing as Racemic Solid Solutions, Mixtures or Racemic Compounds. *Int. J. Pharm.* **1992**, *86*, 221–230.
97. Grunenberg, A.; Kell, B.; Henck, J.O. Polymorphism in Binary Mixtures, as Exemplified by Nimodipine. *Int. J. Pharm.* **1995**, *118*, 11–21.
98. Li, Z.; Zell, M.T.; Munson, E.J.; Grant, D.J.W. Characterization of Racemic Species of Chiral Drugs Using Thermal Analysis, Thermodynamic Calculation and Structural Studies. *J. Pharm. Sci.* **1999**, *88*, 337–346.
99. Rustichelli, C.; Gamberini, M.C.; Ferioli, V.; Gamberini, G. Properties of the Racemic Species of Verapamil Hydrochloride and Gallopamil Hydrochloride. *Int. J. Pharm.* **1999**, *178*, 111–120.
100. Pitré, D.; Nebuloni, M.; Ferri, V. Calorimetric Determination of the Enantiomeric Purity of (1R,2R)-2 Amino-1 (4-nitrophenol) 1,3 Propanediol. *Arch. Phar. (Weinheim)* **1991**, *324*, 325–328.
101. Pitré, D.; Nebuloni, M. Further Characterization of the Solid Form of Iopanoic Acid and Its Enantiomers. *Arch. Phar. (Weinheim)* **1992**, *325*, 385–388.
102. Madarasz, J.; Kozma, D.; Pokol, G.; Acs, M.; Fogassay, E. Merit of Estimation from DSC Measurements for the Efficiency of Optical Resolutions. *J. Thermal Anal.* **1994**, *42*, 877–894.
103. Tonitou, E.; Chow, D.D.; Lawter, J.R. Chiral Beta-blockers for Transdermal Delivery. *Int. J. Pharm.* **1994**, *104*, 19–28.
104. Romero, A.J.; Rodhes, C.T. Approaches to Stereospecific Preformulation of Ibuprofen. *Drug Dev. Ind. Pharm.* **1992**, *17*, 777–792.
105. Dwivedi, S.K.; Sattari, S.; Jamaki, F.; Mitchell, A.G. Ibuprofen Racemate and Enantiomers. Phase Diagrams, Solubility and Thermodynamic Studies. *Int. J. Pharm.* **1992**, *87*, 95–104.
106. Li, Z.; Grant, D.J.W. Relationship between Physical Properties and Crystal Structure of Chiral Drugs. *J. Pharm. Sci.* **1997**, *86*, 1073–1078.
107. Doelker, E. Physicochemical Behavior of Active Substances. Consequences for the Feasibility and Stability of Pharmaceutical Forms. *S.T.P. Pharma Prat.* **1999**, *9*, 399–409.
108. Buckton, G.; Darcy, P. Water Mobility in Amorphous Lactose Below and Close to the Glass Transition Temperature. *Int. J. Pharm.* **1996**, *136*, 141–146.
109. Figura, L.O. The Physical Modification of Lactose and Its Thermoanalytical Identification. *Thermochim. Acta* **1993**, *222*, 187–194.
110. Wada, Y.; Matsubayara, H. Magnesium Stearate. *Thermochim. Acta* **1992**, *196*, 63–84.
111. Containers <661> Physico-Chemical Tests Plastics. *USP* **24**; 2000, 1933–1935.
112. Schneider, H.A.; Di Marzio, E.A. The Glass Transition of Polymer Blends. *Polymer* **1992**, *33*, 3453–3460.
113. Corrigan, O.I. Retardation of Polymeric Carrier Dissolution by Dispersed Drugs. *Drug. Dev. Ind. Pharm.* **1986**, *12*, 1777–1791.
114. Craig, D.Q.M.; Newton, J.M. Characterization of Polyethylene Glycols using Differential Scanning Calorimetry. *Int. J. Pharm.* **1991**, *74*, 33–41.
115. Craig, D.Q.M. A Review of Thermal Methods Used for the Analysis of the Crystal Form, Solution Thermodynamics and Glass Transition Behavior of Polyethylene Glycols. *Thermochim. Acta* **1995**, *248*, 189–203.
116. Sunol, J.J.; Farjas, J.; Berlanga, R.; Saurina, J. Thermal Analysis of a Polyethylene Glycol (PEG 4000): T-CR-T Diagram Construction. *J. Thermal Anal. Calorim.* **2000**, *61*, 711–718.
117. Ford, J.L. The Use of Thermal Analysis in the Study of Solid Dispersions. *Drug Dev. Ind. Pharm.* **1987**, *13*, 1741–1777.
118. Barabas, E.S.; Adeyeye, C.M. Povidone. *Analytical Profiles of Drug Substances and Excipients*; Brittain, H.G., Ed.; Academic Press: New York, 1993; *22*, 555–685.
119. Tan, Y.; Challa, G.; *Polymer* **1976**, *17*, 739–750.
120. Sakellariou, P.; Rowe, R.C.; White, E.F.T. The Thermo-mechanical Properties and Glass Transition Temperatures of Some Cellulose Derivatives. *Int. J. Pharm.* **1985**, *27*, 267–277.
121. Ford, J.L. Thermal Analysis of Hydroxypropylmethylcellulose and Methylcellulose: Powders, Gels and Matrix Tablets. *Int. J. Pharm.* **1999**, *179*, 209–220.
122. Kaloustian, J.; Pauli, A.M.; Pastor, J. Thermal Analysis of Cellulose and Some Etherified and Esterified Derivatives. *J. Thermal. Anal.* **1997**, *48*, 791–804.
123. Sanghavi, N.M.; Sheikl, F.; Fruitwala, M.; *Drug Dev. Ind. Pharm.* **1994**, *20*, 1923–1931.
124. Sakellariou, P.; Rowe, R.C.; White, E.T.F. The Solubility Parameters of Some Cellulose Derivatives and Polyethylene Glycols Used in Tablet Film Coating. *Int. J. Pharm.* **1986**, *31*, 175–177.
125. Wu, C.; McGinity, J.W. Non Traditional Plasticization of Polymeric Films. *Int. J. Pharm.* **1999**, *177*, 15–27.
126. Amighi, K.; Moes, A.J. Determination of Thermal Properties and Film Formation Characteristics of Different Acrylic Aqueous Dispersions Used for Film Coating. *Pharm., Biopharm. Pharm. Technol.* **1995**, *1*, 52–53.
127. Appelqvist, I.A.M.; Cooke, D. Thermal Properties of Polysaccharides at Low Moisture: An Endothermic Melting Process and Water–Carbohydrate Interactions. *Carbohydr. Polym.* **1993**, *20*.

128. Cesaro, A. Thermodynamics of Phase Equilibria in Biotechnological Polysaccharides. *Indian J. Technol.* **1992**, *30*, 565–577.
129. Corrigan, O.I. Thermal Analysis of Spray Dried Products. *Thermochim. Acta* **1995**, *248*, 245–258.
130. Pitt, G.G.; Gu, Z.W. Modification of the Rates of Chain Cleavage of Poly(ϵ -caprolactone) and Related Polyesters in the Solid State. *J. Controlled Release* **1987**, *283*(4), 283–291.
131. Aktar, S.; Ponton, C.W.; Notariami, L.J.; Gould, P.L.; Pharm. Res. **1989**, *5*, 556–560.
132. Dubernet, C. Thermal Analysis of Microspheres. *Thermochimica Acta* **1995**, *248*, 259–269.
133. Rosilio, V. Study of Progesterone Loaded Poly (D,L - Lactide-co-Glycolide). *Pharm Res.* **1998**, *13*, 794–800.
134. Giordano, F.; Bettinetti, G.P. An Experimental and Theoretical Approach to the Analysis of Binary Systems. *J. Pharm. Biomed. Anal.* **1988**, *6*, 951–955.
135. Crowley, K.J.; Forbes, R.T.; York, P.; Nyqvist, H.; Camber, O. Oleate Salt Formation and Mesomorphic Behavior in the Propanol/Oleic Acid Binary System. *J. Pharm. Sci.* **1999**, *88*, 586–591.
136. Koenigbauer, M.J. Pharmaceutical Applications of Microcalorimetry. *Pharm. Res.* **1994**, *11*, 777–783.
137. Serajuddin, A.T.M. Solid Dispersion of Poorly Water-Soluble Drugs: Early Promises, Subsequent Problems and Recent Breakthroughs. *J. Pharm. Sci.* **1999**, *88*, 1058–1066.
138. Keipert, S.; Voigt, R. Interactions between Macromolecules Excipients and Drugs. Improvement of Dissolution of Benzodiazepines Derivatives by PVP. *Pharmazie* **1986**, *41*, 400–404.
139. Duchêne, D. *New Trends in Cyclodextrins and Their Derivatives*; Santé Paris, 1991.
140. Giodarno, F.; Bruni, G.; Bettinetti, G.P. Solid State Microcalorimetry on Drug Cyclodextrin Binary Systems. *J. Thermal. Anal.* **1992**, *38*, 2683–2690.
141. Chun, I.K.; Yun, D.S.; Int. J. Pharm. **1993**, *96*, 916103.
142. Funk, O.; Schwab, L.; Frömming, K.H.; J. Inclusion Phenom. Mol. Recognit. Chem. **1993**, *16*, 299–314.
143. Giron, D.; Goldbronn, C. Use of DSC and TG For Identification and Quantification of the Dosage Form. *J. Thermal. Anal.* **1997**, *48*, 473–483.
144. Bucci, R.; Magri, A.D.; Magri, A.L. DSC in the Chemical Analysis of Drugs. Determination of Diclofenac in Pharmaceutical Formulations. *J. Therm; Anal. Calorim.* **2000**, *61*(2), 369–376.
145. Garti, N. *Crystallization and Polymorphism of Fats and Fatty Acids*; Marcel Dekker, Inc.: New York, 1988.
146. Giron, D.; Riva, A.; Steiger, M. DSC as Support for the Development of Suppositories. *Thermochim. Acta* **1985**, *85*, 509–512.
147. Prya-Jones, R.H. Aminophylline Suppository Decomposition: An Investigation Using DSC. *Int. J. Pharm.* **1992**, *86*, 231–237.
148. Riga, A.; Patterson, G.H. *Oxidative Behavior of Materials by Thermal Analytical Techniques*; ASTM Special Technical Publication 1326, 1997, ASTM.
149. Franks, F. Freeze-Drying: From Empiricism to Predictability. The Significance of Glass Transition. *Dev. Biol. Stand.* **1992**, *74*, 9–19.
150. Franks, F. The Glassy State in Foods. *Int. J. Food Sci. Technol.* Blanshard, J.M.V., Lillford, P.J., Eds.; 1995; *30*, 89–100.
151. Liapiz, A.I.; Pikal, M.J.; Bruttini, R. Research and Development Needs and Opportunities in Freeze-Drying. *Drying Technol.* **1996**, *14*, 1265–1300.
152. Ross, H.; Franks, F. Stabilization of Pharmaceutical Drug Substances by Freeze-Drying: A Case Study. *Drug Stab.* **1996**, *1*, 73–85.
153. Mehl, P.M. Solubility and Phase Transitions in the System Trehalose/Water. *J. Thermal Anal.* **1997**, *49*, 817–821.
154. Junginger, H.E. Polyhydroxy Methacrylate-co-methacrylic Acid: Effect of Water, PEG 400 and PEG6000 on the Glass Transition Temperature. *Polymer* **1989**, *30*, 1946–1953.
155. Kodama, M.; Kato, H.; Aoki, H. Water Molecules in Subgel Phase of Dimyristoylphosphatidylethanolamine/Water. *Thermochim. Acta* **2000**, *352*, 213–221.
156. Katzhendler, I.; Mader, K.; Friedman, M. Structure and Hydration Properties of Hydroxypropyl Methylcellulose Matrices Containing Naproxen and Naproxen Sodium. *Int. J. Pharm.* **2000**, *200*, 161–179.
157. Aoki, S.; Ando, H.; Ishii, M.; Watanabe, S.; Ozawa, H. Water Behavior During Release from a Matrix Observed Using Differential Scanning Calorimetry. *J. Controlled Release* **1995**, *33*, 365–374.
158. Morley, W.G.; Tiddey, G.J. Phase Behaviour of Momoglyceride-Water Systems. *J. Chem. Soc. Faraday Trans.* **1993**, *89*, 2823–2831.
159. Eccleston, G.M.; Beattie, L. Microstructural Changes During Storage of Systems Containing Cetostearyl Alcohol/Polyoxyethylene Alkyl Ether Surfactants. *Drug Dev. Ind. Pharm.* **1988**, *14*, 2499–2518.
160. Eccleston, G.M. Phase Transitions in Ternary Systems and Oil-Un-Water Emulsions Containing Cetrimide and Fatty Alcohols. *Int. J. Pharm.* **1985**, *27*, 311–32.
161. Regola, A.; Canesi, M.; Dolfini, A.; Gazzaniga, A. Use of Thermogravimetry in the Physicochemical Characterization of Hydrophilic Creams. *Boll. Chim. Farm.* **1996**, *135*, 67–71.
162. Blume, A. Biological Calorimetry: Membranes. *Thermochimica Acta* **1991**, *193*, 292–347.
163. Taylor, K.M.G.; Morris, R.M. Liposomes. *Thermochimica Acta* **1995**, *248*, 289–302.
164. Van Winden, E.C.A.; Zhang, W.; Crommelin, D.J.A. Effect of Freezing Rate on the Stability of Liposomes During Freeze-drying and Rehydration. *Pharm. Res.* **1997**, *14*, 1151–1156.
165. Puglisi, G.; Fresta, M.; Ventura, C.; Mazzone, G.; Vandelli, M.A. Methotrexate Interaction with a Lipid Membrane Model of DPPC. *J. Thermal Anal.* **1995**, *44*, 1287–1299.

TITRIMETRY

Vesa Virtanen

Orion Pharma, Kuopio, Finland

DEFINITION OF TERMS

Titrimetry or titrimetric analysis is any method of quantitative chemical analysis in which the amount of a substance is determined by measuring the volume that it occupies or the volume of a second substance that is needed to react completely with the substance being determined. Titration is a process of chemical analysis in which the quantity of some constituents of a sample is determined by adding to the measured sample an exactly known quantity of another substance with which the desired constituent reacts in a definite, known proportion. The reagent of exactly known composition used in a titration is called a standard solution.

The goal of titration is the equivalence point, where the addition of standard solution in an amount that is chemically equivalent to the substance with which it reacts. In fact, its position can be estimated only by observing physical changes associated with it in the solution. These changes manifest themselves at the endpoint of the titration.

An indicator is a supplementary chemical compound that exhibits a change in color as a result of concentration changes occurring near the equivalence point.

PRECIPITATION TITRATIONS

In these titrations, the determination of the substance is effected by precipitating it in the form of an "insoluble" compound of known composition. The equivalence point is reached when sufficient reagent to complete precipitation has been added. In practice, the insoluble reaction product formed will be very slightly soluble and to an extent that depends on the amount of solvent present as well as on the nature and amounts of other ions and compounds present.

In the simplest general case of a slightly soluble salt (AB) formed by the reaction of the oppositely charged univalent anion (A^-) and cation (B^+) of two soluble salts:



where K_{ppn} is the precipitation constant. Assuming that interferences are small enough to be neglected, the concentration of the insoluble salt AB may be regarded as being constant:

$$K_{ppn}[AB] = [A^-][B^+] = \text{constant} = S_{AB}$$

where the constant S is the solubility product. If the solution is pure, equivalent concentrations of A^- and B^+ will be present, and therefore:

$$[A^-] = [B^+] \text{ or } [A^-]^2 = [B^+]^2 = S_{AB}.$$

In this case, if the product $[A^-][B^+]$ exceeds $(S_{AB})^{0.5}$, the solution is saturated with respect to AB, and this substance separates as a precipitate (1).

Titration Curves

Titration curves are based on the negative logarithm (to the base 10) of the molar concentration of the species; p -values or p -functions are useful for deducing the properties required of an indicator as well as the titration error that its use is likely to cause.

The general shape of a titration curve for the precipitation titration of a solution containing one anion, Cl^- , is shown in Fig. 1. Curves plotted from calculated pX values are always symmetrical. A curve plotted by using experimentally obtained values for $[Cl^-]$ is not symmetrical. The lack of symmetry is due to those ions that are being adsorbed in excess by the precipitate in different amounts. This fact is not taken into account in theoretical calculations. Accurate precipitation titrimetry requires a long, steep AB section because the length of AB depends on the initial concentrations of the reactants and the value of the solubility product of the compound precipitated. An increase in analyte and reagent concentration enhances the change in pX in the equivalence point region.

Precipitation titrations can be extended to mixtures that form precipitates of different solubilities. The titration curve in Fig. 2 for a chloride/iodide mixture is a composite of the individual curves for the two anionic species. Because silver iodide has a much lower solubility than does silver chloride, the initial additions of the reagent

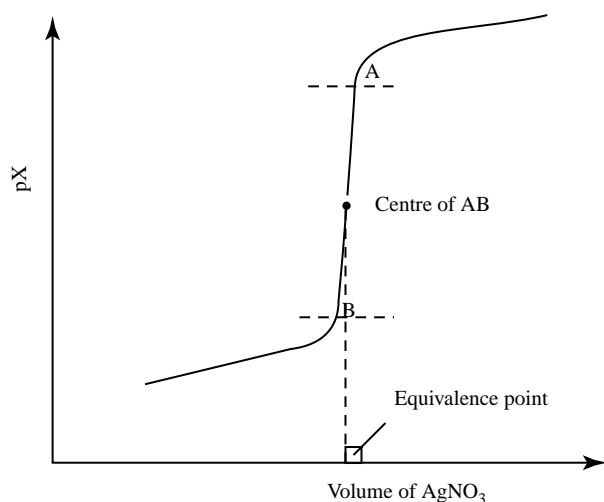


Fig. 1 A theoretical precipitation titration curve.

result exclusively in formation of iodide. Thus, two equivalence points are evident.

Endpoint Detection

The stoichiometric equivalence point should be immediately detectable. This usually requires a large change in some physical or chemical property of the solution. This point in the reaction is often located by means of a

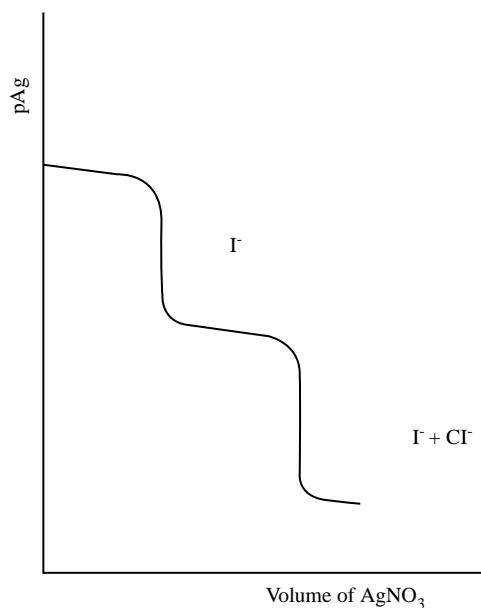
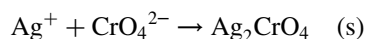
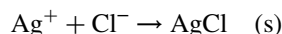


Fig. 2 Titration curve of solution with iodide and chloride.

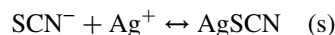
secondary system, which provides an observable endpoint. This secondary system must be reproducible, clearly identifiable, and ideally coincident with the stoichiometric equivalence point. Because coincidence is not always achieved, the difference between the endpoint and the equivalence point should be easily measurable. Often, so-called blank solution is used for this correction. A chemical indicator produces a visually detectable change in the solution, usually color or turbidity, or may form another precipitate that has a distinctive color. The indicator functions by reacting competitively with one of the reactants or products of the titration and, therefore, its concentration must be kept low, favoring an intense color at the endpoint. For some determinations, an electrochemical sensor provides an accurate way of locating the equivalence point.

The formation of a second, highly colored precipitate is the basis of the Mohr method of endpoint detection. Chloride and bromide ions are titrated with standard silver nitrate using chromate ion as indicator, the endpoint being indicated by the appearance of brick-red silver chromate (2).

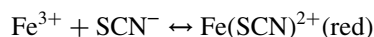


Because silver chromate is more soluble, the K_{sp} value (soluble product constant) of silver chromate is not exceeded until the precipitation of Cl^- is complete. The endpoint can be corrected by using the Mohr method to standardize the silver nitrate solution against pure sodium chloride.

The Volhard method of endpoint detection involves using Fe^{3+} ions as the indicator (3). This procedure requires a suitably acidic solution to prevent precipitation of iron(III) as the hydrated oxide. It has the disadvantage that it is useful only for the reaction:



After the first excess of thiocyanate ions is added, the indicator reaction is:



This means that to determine halide ion (except F^-) back titration is required. In this case, a measured aliquot of standard silver nitrate solution is added to the sample, and the excess silver ion is determined by back-titration with a standard thiocyanate solution.

An adsorption indicator is typically an organic dye, such as fluorescein and its derivatives. Most adsorption indicators are weak acids. Their use is thus confined to basic, neutral, or slightly acidic solutions in which

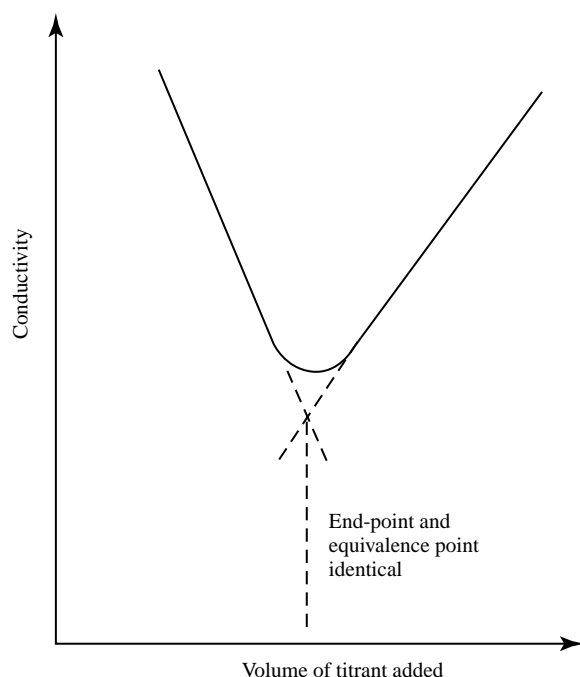


Fig. 3 A conductivity curve for precipitation titration.

the indicator exists predominantly as the anion. Some cationic adsorption indicators are suitable for titrations in strongly acidic solutions. In this case, adsorption of the dye and coloration of the precipitate occur if the precipitate particles possess a negative charge.

Because the total amount of ions in solution decreases as the analyte is precipitated, the conductance of the solution must decrease. For an accurate result, the plot of conductivity against volume of titrant added has the characteristics shown in Fig. 3. This technique is suitable for both dilute and concentrated solutions as well as for colored solutions.

The electromotoric force (emf) of a cell depends on the ionic concentration of the solutions. To locate the equivalence point, the variation in emf is monitored as the concentration of the analyte changes. When the measured emf is plotted against the total volume of titrant added, the curve produced is similar to that of a titration curve Fig. 4. This technique has all the advantages of the conductometric method and gives an experimental curve from which the endpoint can be detected accurately.

Applications

Most applications of precipitation titrations are based on the use of a standard silver nitrate solution and are therefore sometimes called argentometric methods (Table 1).

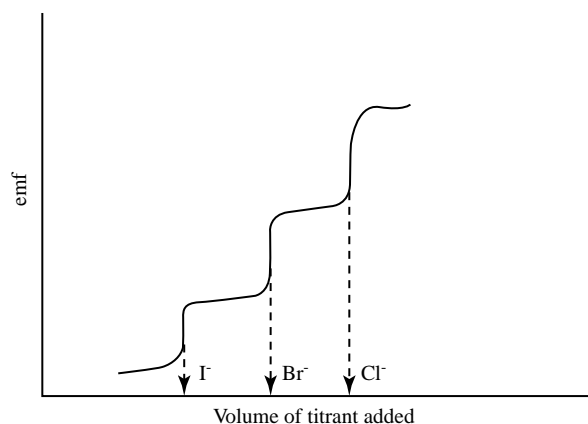


Fig. 4 Titration curves for a mixture of halides.

NEUTRALIZATION TITRATIONS

Acid–Base Equilibria

Aqueous solutions always contain hydronium ions as well as hydroxide ions as a consequence of the dissociation of water:



Certain solutes cause changes in the concentrations of the two species, often with profound effect on the chemical characteristics of the solution.

Application of the mass law to the dissociation of water leads to $K_w = [\text{H}_3\text{O}^+][\text{OH}^-]$, where K_w is called the ion-product constant for water. At 25°C the ion-product constant has the numerical value of $1.0 \times 10^{-14} \text{ mole}^2/\text{L}^2$. In pure water, the concentrations of hydronium and hydroxide ions are identical, and pure water is neutral.

A useful relationship is obtained from the negative logarithm of both sides of the ion-product constant expression. Thus:

$$\begin{aligned} -\log K_w &= -\log[\text{H}_3\text{O}^+][\text{OH}^-] \\ &= -\log[\text{H}_3\text{O}^+] - \log[\text{OH}^-] \end{aligned}$$

Table 1 Typical applications of precipitation titrations

Analyte	Titrant	Endpoint Detection
Cl^- , Br^- , I^-	Ag^+	Potentiometric
Cl^- , Br^- , I^-	Ag^+	Precipitate
Ag^+	SCN	Fe(III), potentiometric
Cl^- , Br^-	$\text{Hg}(\text{NO}_3)_2$	Precipitate
SO_4^{2-} , MoO_4^{2-}	$\text{Pb}(\text{NO}_3)_2$	Precipitate

from which it follows that:

$$pK_w = -\log K_w = pH + pOH$$

where pK_w represents the negative logarithm of the ion-product constant of water. At 25°C, pK_w is 14; that is, $pH + pOH = 14.00$.

Buffer Solutions

A buffer solution resists changes in pH as a result of dilution or small additions of acids or bases. The most effective buffer solutions contain high and approximately equal concentrations of a conjugate acid–base pair. The resistance of buffer mixtures to pH is changed by adding acids or bases. The ability of a buffer to prevent a significant change in pH is directly related to the total concentration of the buffering species and their concentration ratios. The buffer capacity of a solution is defined as the number of equivalents of strong acid or base needed to cause 1.00 L of the buffer to undergo a 1.00-unit change in pH (4).

Acid–Base Indicators

Endpoint detection in a neutralization titration is ordinarily based on the abrupt change in pH that occurs near the equivalence point. A noninstrumental method of pH measurement much used in simple titrations uses indicators. These are generally organic dyes, or weak acids or bases, that on dissociation or association undergo internal structural changes at or near the equivalence point of the neutralization, resulting in a color change.

A list of some compounds possessing acid-base indicator properties is shown in Table 2. Ethanol is the common solvent for indicator solutions.

Titration curves of simple systems

When both reagent and analyte are strong, the net neutralization reaction can be expressed as follows:



The hydronium ions in an aqueous solution of a strong acid derive from two sources: the reaction of the solute with water and the dissociation of water itself.

Titration of a strong acid with a strong alkali starts with pure acid in the start that is gradually diluted, changing increasingly to a neutral salt until only neutral salt remains at the equivalence point. Immediately beyond the equivalence point, the amount of strong alkali increases. The change in pH near the equivalence point will be sharp and large. This may or may not be at pH 7, depending on

Table 2 Some acid-base indicators

Indicator	pH range	Color change
Cresol red	0.2–1.8	red–yellow
Thymol blue	1.2–2.8	red–yellow
	8.0–9.6	yellow–blue
Bromophenol blue	3.0–4.6	yellow–blue
Methyl yellow	2.9–4.0	red–yellow
Methyl orange	3.1–4.4	red–orange
Congo red	3.0–5.0	blue–red
Bromocresol green	3.8–5.4	yellow–blue
Methyl red	4.2–6.3	red–blue
Bromocresol purple	5.2–6.8	yellow–purple
Bromothymol blue	6.0–7.6	yellow–blue
Phenol red	6.8–8.4	yellow–red
Cresol purple	7.6–9.2	yellow–purple
Thymol blue	8.0–9.6	yellow–blue
Phenolphthalein	8.3–10.0	colorless–red
Thymolphthalein	9.3–10.5	colorless–blue
Alizarin yellow GG	10–12	colorless–yellow

the degree of ionization of the acid and base, that is, their strengths as acid or base.

Titration of a weak acid such as acetic acid against strong base gives a titration curve as that shown in Fig. 5. At first, partly ionized acid is present owing to a pH too high for a strong acid. As neutralization continues, more acid ionizes and mainly ionized acid and salt are present. Therefore, the pH changes gradually; this is called the buffer effect. If a strong acid is titrated with a weak base such as ammonia, the converse occurs.

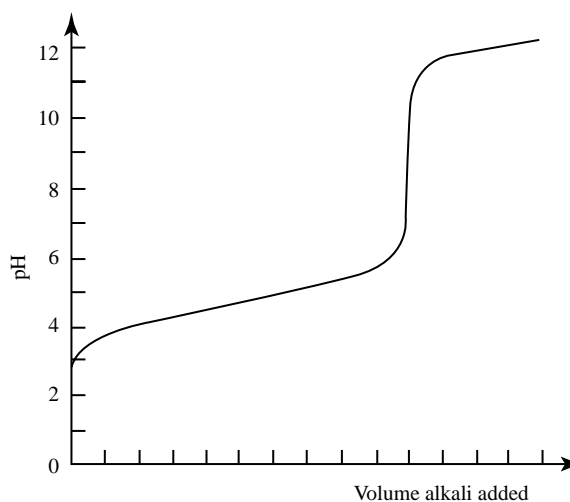


Fig. 5 A weak acid-strong base titration curve.

When a weak acid is titrated with a weak base, the titration curve shows a continuously and gradually changing pH with the addition of base. No region with a sharp shift in pH is obtained for small additions of titrant. If a sharp shift occurs, it is still less than 2 pH units and not detectable by indicators. On both sides of the equivalence point, buffers are present, and at the equivalence point, the pH depends on the relative strengths of acid and base.

Titration curves of complex systems

Complex systems include solutions containing two acids or bases, which contain or consume two or more protons, and amphiprotic substances that act as both acids and bases. A characteristic of all such systems is that two or more equilibria must be considered in describing their behavior. As a consequence, the techniques for pH data derivation are often more complex than for simple systems.

A titration curve for the titration of weak and strong acids with strong base is shown in Fig. 6. The stronger acid is neutralized first, along with some of the weaker acid. As a result, a curve more sloping is obtained. The sharp pH change of the strong acid is still obtained, although the weaker acid interferes. After all the strong acid has gone (point A), the weaker acid is titrated (point B).

Compounds with two or more acidic or basic functional groups will yield multiple endpoints in a titration, provided the acidic or basic groups differ sufficiently in strength. Computational techniques permit the derivation of reasonably accurate theoretical titration curves for polyprotic acids or bases, provided the ratio K_1/K_2 is above 103. K_1 and K_2 are dissociation constants. The titration curve of a dibasic weak acid with NaOH resembles that shown in Fig. 6.

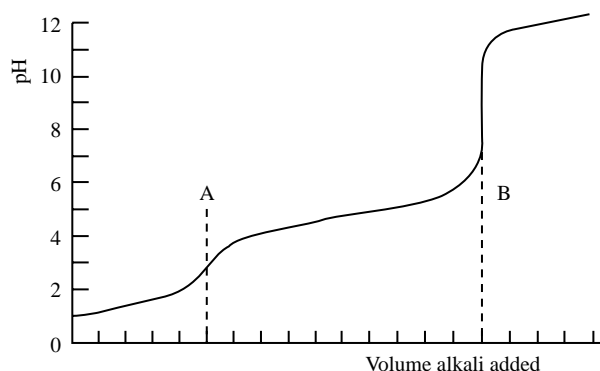


Fig. 6 Titration of a mixture of a strong and a weak acid with alkali.

The derivation of a titration curve for a polyfunctional base is similar to the previous one for acid.

Amphiprotic substance, when dissolved in a suitable solvent, behaves both as a weak acid and as a weak base. If either its acidic or its basic character predominates sufficiently, titration of the species with a strong base or a strong acid may be feasible.

Applications in Aqueous Media

Several important elements that occur in organic and biological systems are conveniently determined by methods that involve an acid-base titration as the final step. Generally, these elements are nonmetallic such as carbon, nitrogen, chlorine, bromine, sulfur, phosphorus, and fluorine. However, in addition, similar methods exist for several less commonly encountered species. In each case, the element is converted to an inorganic acid or base that can be titrated.

Numerous inorganic species can be determined by titration with strong acids or bases. For example, ammonium salts are determined by conversion to ammonia with strong base and distillation in the Kjeldahl apparatus. Also, inorganic nitrates and nitrites can be determined using Kjeldahl method (5) by reducing these species to ammonium ion.

Carboxylic and sulfonic acid groups impart acidity to organic compounds. Most carboxylic acids and sulfuric acids are readily dissolved in water, and their titration with a base is straightforward. If solubility in water is not sufficient, the acid can be dissolved in ethanol and titrated with aqueous base. Aliphatic amines and many saturated cyclic amines can be titrated directly with a solution of a strong acid. Esters are determined by saponification with a measured quantity of standard base. The excess base is titrated with standard acid.

Many aldehydes and ketones can be assayed with the aid of a solution of hydroxylamine hydrochloride. The liberated hydrochloric acid is titrated with base. The total salt content of a solution is determined by converting the salt to an equivalent amount of an acid or a base by passage through a column packed with an ion-exchange resin and then titrated with either base or acid, respectively.

Applications in Nonaqueous Media

Many useful titrations can be performed in glacial acetic acid, which is used widely for the titration of aromatic amines, amides, ureas, and other weak nitrogen bases. Direct titration of most amino acids with a standard acid can be performed in glacial acetic acid.

Basic solvents such as ethylenediamine, dimethylformamide, pyridine, and dimethylsulfoxide can be used for acids that are too weak to be titrated in water. Amine salts, inorganic salts, carboxylic acids, phenols, and imides are soluble in ethylenediamine, where they exhibit enhanced acidic characteristics.

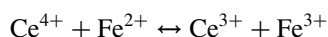
Many sulfa drugs such as sulfanilamide, sulfathiazole, and sulfathalidine can be titrated in ethyleneamine with tetrabutylammonium hydroxide. The ability of dimethylformamide to dissolve salts, polymers, and many organic compounds accounts for its wide use as a titration solvent. Inert solvents such as acetone can be used for titration of acids, acetonitrile for both acids and bases, and ethyl acetate for amines. A suitable nonaqueous medium such as methyl isobutyl ketone makes it possible to discriminate among various mineral acids that are not leveled by the solvent as they are in water.

OXIDATION-REDUCTION TITRATIONS

Equilibrium

Oxidation of a substance (element or a compound) involves the loss of electrons. Conversely, reduction of a substance involves the gain of electrons. Thus, in oxidation-reduction system, electron transfer occurs.

In titration of iron(II) with cerium(IV):



equilibrium is attained after each addition of titrant. After the first addition of cerium(IV), all four species will be present in the solution in amounts dictated by the equilibrium constant of the reaction. At chemical equilibrium of an oxidation-reduction system, the electrode potentials E for the two half-reactions are identical. Thus, at any point in the titration:

$$E_{\text{Ce}^{4+}} = E_{\text{Fe}^{2+}} = E_{\text{system}}$$

If the solution contains a reversible oxidation-reduction indicator as well, its potential must be the same as that for the system:

$$E_{\text{In}} = E_{\text{Ce}^{4+}} = E_{\text{Fe}^{2+}} = E_{\text{system}}$$

Titration Curves

The potential of a system can be measured experimentally by determining the emf of a suitable cell. Thus, for the titration of iron(II) with cerium(IV), the cell is:

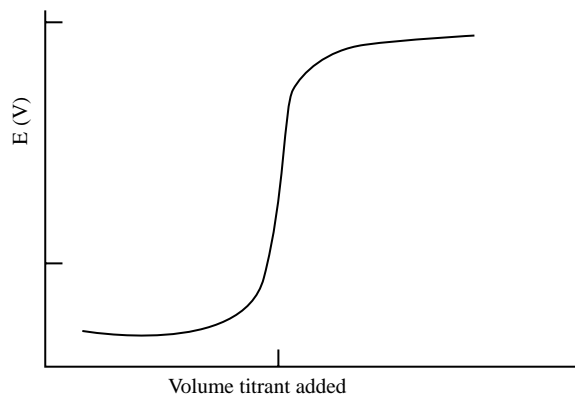
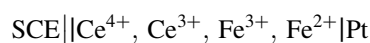
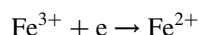


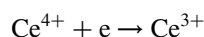
Fig. 7 Titration curve for iron(II) with cerium(IV).



The potential of the platinum electrode versus a saturated calomel electrode (SCE) is determined by the affinity of Fe^{3+} for electrons:



and that of Ce^{4+} :



In deriving Esystem data for a titration curve, either $E_{\text{Ce}^{4+}}$ or $E_{\text{Fe}^{3+}}$ can be used. Short of the equivalence point, the concentration of cerium(IV) is vanishingly small. The concentrations of iron(II), iron(III), and cerium(III) can be calculated from the amount of titrant added.

Applying the Nernst equation for the iron(III)/iron(II) couple gives a value for the potential of the system directly. Using the couple cerium(IV)/cerium(III) would give the same answer, but it would first be necessary to calculate a value for the concentration of cerium(IV), which in turn would require evaluation of the equilibrium constant for the reaction.

As shown in Fig. 7 there is a rapid change in the value of E as the titration is proceeded through the endpoint. In fact, the titration curve has the same general form as that of an acid-base titration. An exact value for the endpoint can be calculated using the Nernst equations for the half-reactions.

The shape of the curve for an oxidation-reduction titration depends on the nature of the system under consideration. The titration curve in Fig. 7 is symmetric about the equivalence point because the molar ratio of oxidant to reductant is equal to unity. An asymmetrical curve results if the ratio differs from this value. Solutions containing two oxidizing or reducing agents yield titration curves containing two inflection points if the standard

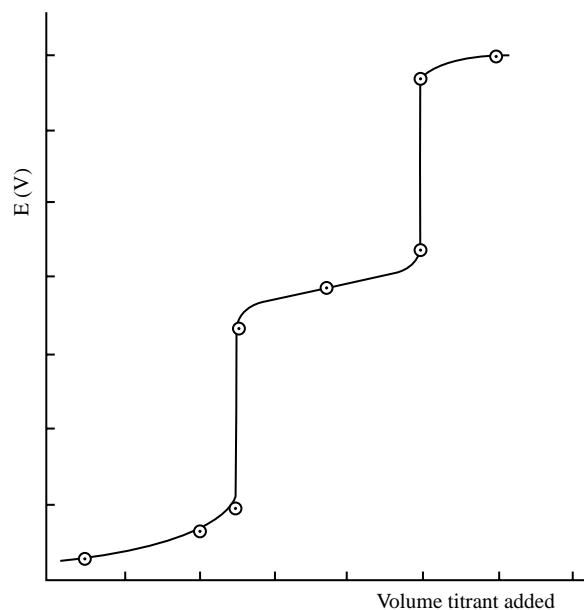


Fig. 8 Titration of iron(II) and titanium(III) with cerium(IV).

potentials for the two species are different by more than approximately 0.2 V. Fig. 8 shows the titration curve for a mixture of iron(II) and titanium(III) with cerium(IV). The first additions of cerium are used by more readily oxidized titanium(III) ion, thus, the first step in the titration curve corresponds to titanium and the second to iron.

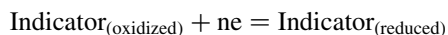
Oxidation–Reduction Indicators

Specific indicators owe their behavior to a reaction with one of the participants in the titration. The best-known specific indicator is starch, which forms a dark blue complex with triiodide ion. Also, potassium thiocyanate is

used as a specific indicator, for example, in titration of iron(III) with solutions of titanium(III) sulfate.

Oxidation–reduction indicators

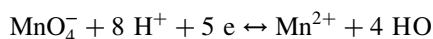
A redox indicator is essentially a compound that has oxidized and reduced forms with different, and preferably intense, colors (Table 3). The half-reaction responsible for color change in a typical oxidation–reduction indicator can be written as:



Applications

Standard oxidants

The powerful oxidant potassium permanganate, KMnO_4 , is a widely used oxidizing agent. It is most commonly used with solutions that are 0.1 *N* or greater in mineral acid. Under those conditions, the product is manganese(II):



However, the tendency of permanganate to oxidize chloride ion is a disadvantage because hydrochloric acid is such a useful solvent and, furthermore, KMnO_4 solutions have limited stability.

For example, determinations of Sn, H_2O_2 , Fe, V, Mo, W, U, Ti, Mg, Ca, Zn, Co, Pb, Ag, K, Na, and HNO_2 potassium permanganate are used as oxidizing agents.

A sulfuric acid solution of cerium(IV) is nearly as potent an oxidizing reagent as is permanganate and can be substituted for the latter in most of the applications described with potassium permanganate. Cerium(IV) does not oxidize chloride ion at a detectable rate as does permanganate, and the reagent is indefinitely stable. On the other hand, the color of cerium(IV) is not sufficiently intense to serve as an indicator. In

Table 3 Some redox indicators

Indicator	Color change oxidized-reduced	Transition potential, V
Indigomonosulphonate	Blue—colorless	+ 0.26
Phenosafranin	Red—colorless	+ 0.28
Indigotetrasulphonate	Blue—colorless	+ 0.36
Methylene blue	Blue/green—colorless	+ 0.53
Diphenylamine	Violet—colorless	+ 0.76
<i>p</i> -Ethoxychrysoidine	Yellow—red	+ 0.76
Diphenylamine sulfonic acid	Red/purple—colorless	+ 0.85
Eriocaulin A	Red—yellow/green	+ 0.98
1,10-phenanthroline iron(II)	Pale blue—red	+ 1.11
2,3-diphenylamine dicarboxylic acid	Blue/violet—colorless	+ 1.12

addition, the reagent cannot be used in neutral or basic solutions.

Use of potassium dichromate is limited because of its lesser oxidizing strength and slowness of some of its reactions. This technique is used for determination of iron(II), nitrate, chlorate, permanganate, and organic peroxides, among others.

Iodine, a relatively weak oxidant, is used for the selective determination of strong reducing agents. Iodine can be used for determination of As, Sb, Sn, H_2S_2 , SO_2 , $\text{S}_2\text{O}_3^{2-}$, and N_2H_4 , among others. Its great advantage is its ready availability as a sensitive and reversible indicator. Its disadvantages include the low stability of iodine solutions and the incompleteness of reactions between iodine and many reducing agents.

Potassium bromate is a convenient source of bromine in organic analysis. Few organic compounds react sufficiently rapidly for a direct titration, and, thus, a measured excess of the bromate is added to the sample, and after the reaction is complete, the excess bromine is back-titrated with an arsene(III) solution. Some organic compound analyzed using bromine substitution are, for example, phenol, *p*-chlorophenol, salicylic acid, acetylsalicylic acid, *m*-cresol, aniline, sulfanilic acid, and *b*-naphthol. Bromine addition is used most often in the estimation of olefinic unsaturation in fats, oils, and petroleum products.

Reductants

Because of the readiness of reducing agents to react with atmospheric oxygen, the titrations must be carried out in and the reagents must be stored under an inert atmosphere. Alternatively, a stable standard oxidizing agent can be used for titration of an aliquot of the reductant to determine the current concentration of the reducing agent.

A variety of oxidizing agents such as Cr(VI), Ce(IV), Mo(VI), NO_3^- , NH_2OH , and organic peroxides can be determined by reaction with a measured excess of standard iron(II) solution. Standard potassium dichromate is frequently used for the back-titration.

Iodide ion is a moderately effective reducing agent. In its applications, a standard solution of sodium thiosulfate is used to titrate the iodine liberated by reaction of the analyte with an unmeasured excess of potassium iodide. Some substances determined by using iodometric method are IO_4^- , IO_3^- , BrO_3^- , ClO_3^- , Br_2 , Cl_2O_2 , O_3 , Cu^{2+} , NO_2^- , and organic peroxide.

The water content in solids and many organic acids, alcohols, esters, ethers, anhydrides, and halides can be

determined using Karl–Fischer reagent (6), composed of iodine, sulfur dioxide, pyridine, and methanol.

COMPLEX-FORMATION TITRATIONS

Metal ions can act as electron-pair acceptors, reacting with electron donors to form coordination compounds or complex ions. The donor species, or ligand, must have at least one pair of unshared electrons to form the bond. Remarkable growth in the analytical applications of complex-formation reactions is attributable to a particular class of coordination compounds called chelates. These compounds are made by the reaction of a metal ion and a ligand that contains two or more donor groups. The properties of chelates can differ markedly from the parent cation.

Titration Curves

The data necessary to plot theoretical *p*-functions versus reagent volumes require the use of formation (or stability) constants (Fig. 9). Equilibrium between a metal ion M,

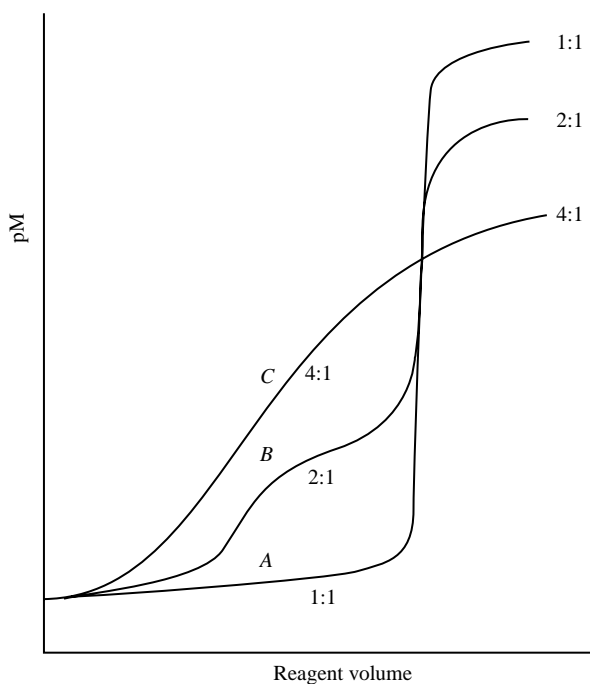


Fig. 9 Curves for complex formation titrations. Titration of the tetradentate ligand D (curve A), bidentate ligand (curve B), and unidentate ligand (curve C).

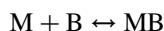
possessing a coordination number of 4, and the quadridentate ligand D can be represented by:



In similar manner, the equilibrium between M and bidentate B can be represented by:



Overall formation of MB_2 occurs in two steps and involves the intermediate formation of MB:



The formation constants for these individual processes are:

$$K_1 = [MB]/[M][B]$$

and

$$K_2 = [MB_2]/[MB][B]$$

The equilibrium constant for the overall reaction is given by the product of the individual steps:

$$K_{\text{overall}} = [MB_2]/[M][B]^2 = K_1 K_2$$

The complex between M and a simple, unidentate ligand A results in the overall equilibrium:



This process occurs also in a stepwise manner, and the equilibrium constant for the overall reaction is therefore numerically equal to the product of the constants for the four constituent processes.

Endpoint Detection

In some complex-formation titrations, the endpoint is noted by the formation or disappearance of a solid phase. For example, in the titration of cyanide with silver ion, the solution remains clear, but the first excess of silver causes formation of a white solid that marks the endpoint. The electron-donor groups of most common ligands tend to combine not only with metallic ions but also with protons; thus, the equivalence point in a complex-formation titration is often accompanied by a marked change in pH, which can be detected with an acid-base indicator.

Formation or disappearance of a soluble colored complex can also indicate an endpoint. Many reagents that form colored complexes with certain metals have been developed only for use as indicators in these titrations. If

the cation being titrated produces a color with the indicator, the endpoint will be characterized by the disappearance of this color. When the cation does not give a colored complex, a second cation that does is introduced, and the first excess of titrant then decolorizes this complex.

Some Applications

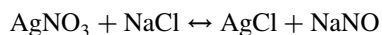
Inorganic complexing reagents such as $Hg(NO_3)_2$, $AgNO_3$, $NiSO_4$ and KCN can be used for complex-formation titrations. Mercury(II) ion forms neutral complexes with most of the anions that precipitate with silver nitrate such as Br^- , Cl^- , SCN^- , CN^- and thiourea.

$AgNO_3$ reacts with CN^- forming $Ag(CN)_2^-$. Iodine used as indicator and endpoint is detected by formation of solid AgI. $NiSO_4$ can also be used for determination of CN^- with AgI as indicator. The endpoint is also detected here by formation of AgI. KCN reacts with Cu^{2+} , Hg^{2+} , and Ni^{2+} forming $Cu(CN)_4$, $Hg(CN)_2$ and $Ni(CN)_4$. Various indicators for endpoint detection can be used (7).

Numerous tertiary amines that also contain carboxylic acid groups form remarkably stable chelates with many metal ions. Ethylenediamine tetra-acetic acid (EDTA) can be used for determination of 40 elements by direct titration using metal-ion indicators for endpoint detection (8). Direct titration procedures are limited to metal ions that react rapidly with EDTA. Back titration procedures are useful for the analysis of cations that form very stable EDTA complexes and for which a satisfactory indicator is not available. EDTA is also used for determining water hardness; the total concentration of calcium and magnesium expressed in terms of the calcium carbonate equivalent.

POTENTIOMETRIC TITRATIONS

In potentiometric titrations, the activity of one species is continuously monitored as it changes during the course of the titration, for example, in the titration:



where silver nitrate is added to aqueous sodium chloride. The change in activity of the Cl^- can be monitored. The potential between the reference electrode and the indicator electrode is usually measured at the start of and after the addition of small amounts of titrant. A reproducible equilibrium is of little concern. Requirements for reference electrodes are greatly relaxed. In potentiometric titrations, accuracy is higher than that in direct potentiometric measurements because measured potentials are used to

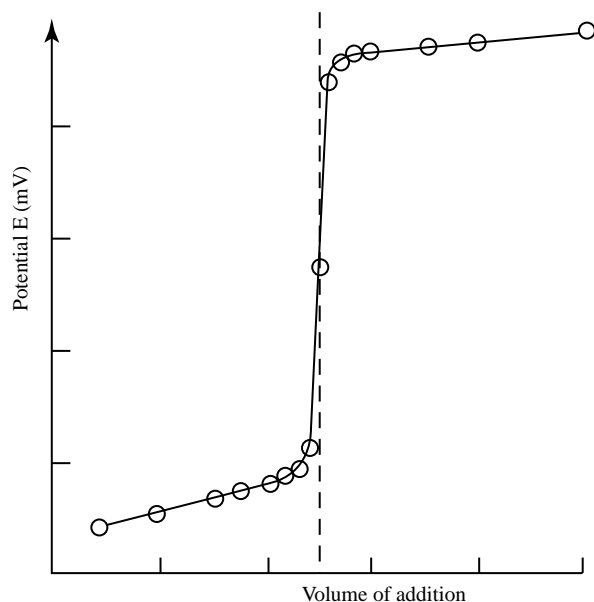


Fig. 10 Potentiometric titration curves.

detect rapid changes in activity that occur at the equivalence point of the titration. The change in emf versus titration volume is more interesting than the absolute value of the emf. In potentiometric titrations, the influence of liquid-junction potentials and activity coefficients is minimized. The acid and alkali errors of the glass electrode occur at extremes of pH and have little effect near the interesting equivalence point.

Titration Curves and Endpoint Detection

In potentiometric titrations, the titration curve can be followed point by point, plotting as the ordinate successive values of the cell emf versus the corresponding volume of titrant added as the abscissa (Fig. 10). The titrant should be added in the smallest accurately measurable increments that provide an adequate density of points, particularly in the vicinity of the equivalence point. The greatest change in emf occurs around the equivalence point. The most straightforward method takes the midpoint in the steeply rising portion of the curve as the endpoint.

A second approach is to calculate the change in potential-per-unit change in volume in reagent ($\Delta E/\Delta V$). By inspection, the endpoint can be located from the inflection point of the titration curve. This is the point that corresponds to the maximum rate of change of cell emf per unit volume of titrant added (usually 0.05 or 0.1 mL). The first-derivative method is based on the sigmoid shaped curve.

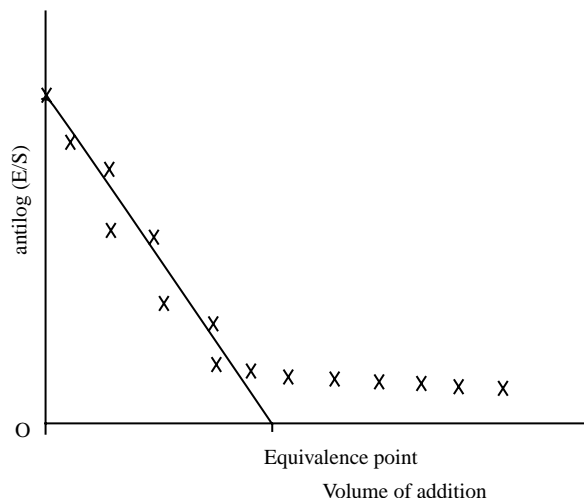


Fig. 11 Gran's plot for a potentiometric titration.

The second-derivative method is an extension of the first-derivative method. The second-derivative of the data changes sign at the point of inflection in the titration curve. This change is often used as the analytical signal in automatic titrators.

Grans method (9,10) involves the use of the Nernst equation:

$$E = E^* + S \log C_i$$

where

S is electrode slope, E^* is $E' \pm (0.0591/n) \log i$, and

and E' is the constant incorporating the potential of the reference electrode and the standard potential of the half-cell containing the solution under investigation and the ion selective electrode. The rearrangement of this equation gives:

$$\log C_i = (E - E^*)/S$$

or

$$\begin{aligned} C_i &= \text{antilog}[(E - E^*)/S] \\ &= [\text{antilog}(-E^*/S)][\text{antilog}(E/S)] \end{aligned}$$

The first antilog term is a constant and therefore:

$$C_i = \text{const} \cdot [\text{antilog}(E/S)]$$

If the species i is disappearing during the course of the titration, then it will do it linearly. The antilog term reduces linearly, and a plot of antilog (E/S) against volume of reagent added should give a straight line whose intercept with the volume axis will be the equivalence point, as shown in Fig. 11.

Applications

Precipitation and complex-formation titrations

The indicator electrode for a precipitation titration is often the metal from which the reacting cation is derived. Membrane electrodes that are sensitive to one of the ions involved in the titration process may be used. For example, fluoride-sensitive membrane electrode is used in the determination of the fluoride content of toothpastes. Lanthanum(III) solution is used as a precipitant.

Silver nitrate is the most commonly used reagent for precipitation titration, including potentiometric determination of endpoint. Argentometric methods exist for the determination of halides, halogenoids, mercaptans, sulfides, phosphates, oxalates, and arsenates. A silver electrode serves as indicator for the potentiometric determination of these ions. If the concentration of reagent and analyte is 0.1 *M* or higher, then a calomel electrode can be used. In dilute solutions, the slight leakage of chloride ions from the salt bridge could be a source of significant error, but not in more concentrated solutions.

Mercury electrode can be used for the potentiometric determination of 29 divalent, trivalent, and tetravalent cations with EDTA (11, 12).

Neutralization titrations

Potentiometric acid–base titrations are particularly useful for the analysis of mixtures of acids or polyprotic acids (or bases) because often, discrimination between the endpoints can be made. An approximate numerical value for the dissociation constant of a weak acid or base can be estimated from potentiometric titration curves. In theory, this quantity can be obtained from any point along the curve, but it is most easily derived from the pH at the point of half-neutralization.

In the case of weak acid HA, it can be assumed that at the midpoint, $[HA] = [A^-]$, and thus:

$$K_a = [H^+][A^-]/[HA] = [H^+]$$

which gives

$$pK_a = pH$$

It must be noted that this constant is not correct because activities should be used in calculations:

$$K_a = [A^-]f_A/[HA]f_{HA}$$

and because

$$[HA] \text{ and } [A^-] \text{ are equal } K_a = f_A/f_{HA}$$

Oxidation–reduction titrations

In oxidation–reduction titrations, an electrode potential related to the concentration ratio between the oxidized and reduced forms of either of the reactants is determined as a function of the titrant volume. The indicator electrode must be responsive to at least one of the couples involved in the reaction. Indicator electrodes for oxidation–reduction titrations are generally constructed from platinum, gold, mercury, or palladium. The metal chosen must be unreactive with respect to the components of the reaction. The indicator metal is merely a medium for electron transfer.

Automatic Titrators

An entire titration can be performed automatically by titrators equipped with microcomputers and analog-to-digital converters and using dedicated software (13). The most widely used apparatus for automatic reagent addition consists of a calibrated syringe that is activated by a motor-driven micrometer screw. The volume is determined from the number of turns the screw makes during the titration.

Another method uses a preset equivalence point potential applied across the electrodes by means of a calibrated potentiometer. A difference between this potential and that of the electrodes causes an “error” signal, which is amplified. This causes the electronic switch to close, permitting a flow of electricity through the solenoid-operated valve of the burette. As the signal approaches zero, the flow of titrant ceases as the current to the solenoid is switched off.

Second-derivative titrators have the advantage that no preknowledge of the equivalence point potential is required. The signal processor calculates the second derivative of the electrode potential of the indicator electrode. Change in the sign of the second derivative causes a switching device to turn off the flow of the titrant.

A fully automated unit accepts a series of samples placed on a turntable. After each titration, the turntable rotates, indexes the next sample beneath the electrode assembly, and starts the titration.

CONDUCTOMETRIC TITRATIONS

Theory

Conduction of an electric current through an electrolyte solution involves migration of positively charged species toward the cathode and negatively charged species toward the anode. The conductance of a solution is

a measure of the current flow that results with application of a given electrical force. It is directly dependent on the number of charged species the solution contains (14). The conductance L of a solution is also the reciprocal of the electrical resistance and has the units of ohm:

$$L = 1/R$$

where R is the resistance in ohms.

The conductance is directly proportional to the length (l) of a uniform conductor. Thus:

$$L = kA/l$$

where k is a proportionality constant called the specific conductance, and A is the surface area.

The equivalent conductance, Λ , is defined as the conductance of a 1-g equivalent of solute contained between electrodes spaced 1 cm apart. It is equal to L when 1-g equivalent of solute is contained between electrodes spaced 1 cm apart. The volume V of solution (cm^3) that will contain 1 gram equivalent solute is given by:

$$V = 1000/C$$

where C is the concentration in equivalents per liter.

Volume can also be expressed in terms of the dimensions of the cell:

$$V = lA$$

when l is 1 cm

$$V = A = 1000/C$$

when $L = kA/l$,

$$\text{when } L = kA/l$$

$$\text{and } \Lambda = L, \text{ then } \Lambda = 1000k/C$$

A conductometric titration involves measurement of the conductance of the sample after successive additions of reagent. The endpoint is determined from a plot of either the conductance or the specific conductance as a function of the volume of added titrant. Throughout a titration, the volume of the solution is always increasing. Unless the conductance is corrected for this effect, nonlinear titration curves result. The titrant should be at least 10 times as concentrated as the solution being titrated to keep the volume change small. Some temperature control is ordinarily required during a conductometric titration because the temperature coefficient for conductance measurements is approximately 2% per °C.

Titration Curves

Titration curves for conductometric titrations take a variety of shapes, depending on the chemical system under investigation. In general, they are characterized by straight line portions with dissimilar slopes on either side of the equivalence point, as shown previously in Fig. 3. To establish a conductometric endpoint, after correcting for volume changes, the conductance data are plotted as a function of titrant volume. The two linear portions are then extrapolated, and the point of intersection is taken as the equivalence point. Frequently, reactions fail to proceed to absolute completion, and the conductometric titration curves invariably show departures from strict linearity in the region of the equivalence point.

An advantage of the conductometric titration is that it can be used for titrations based on relatively unfavorable equilibria.

Applications

Acid–base titrations

Neutralization titrations are particularly well-adapted to the conductometric titration because of the very high conductance of the hydronium and hydroxide ions compared with the conductance of the reaction products. In neutralization of strong acids, hydronium ions are being replaced by an equivalent number of less mobile sodium ions, and the conductance decreases as a result of this substitution. At the equivalence point, the concentration of hydronium and hydroxide ions are at a minimum, and the solution exhibits its lowest conductance. After the endpoint, a reversal of slope occurs as the sodium ion and the hydroxide ion concentration from the excess base increase. There is an excellent linearity between conductance and the volume base added, except at very near equivalence point region. Very dilute solutions can be analyzed accurately.

In conductometric titrations, a change in conductance is caused after the endpoint owing to the increase in concentration of the mobile hydroxide ion. Very weak acids (such as boric acid) and moderately weak acids (such as acetic acid) can be titrated using conductometric titration.

If the titrant is a weak electrolyte (such as ammonia), the curve is essentially horizontal past the equivalence point, which causes less uncertainty to the extrapolation of a curve. In titration of a weak base, such as acetate ion, with a strong acid, a salt and undissociated acetic acid are formed. After the endpoint is passed, a sharp rise in conductance attends the addition of excess hydronium

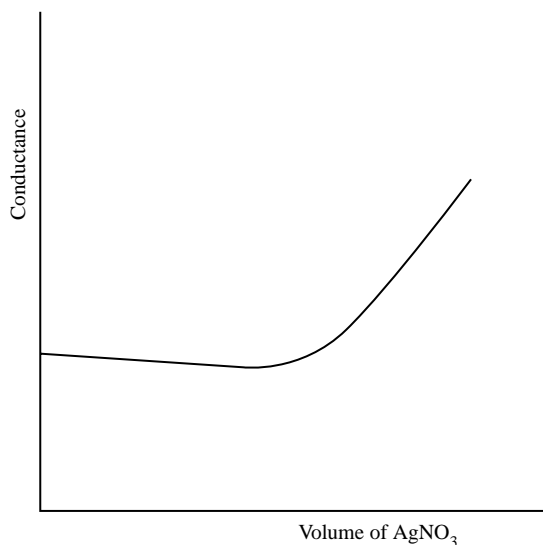


Fig. 12 Titration of chloride ion with silver nitrate.

ions. Salts whose acidic or basic character is too weak to give satisfactory endpoints with indicator are conveniently titrated with the conductometric method. The conductometric titration of a mixture of two acids that differ in degree of dissociation is frequently more accurate than a potentiometric titration.

Precipitation and complex-formation titrations

Figure 12 illustrates the titration of sodium chloride with silver nitrate. After all chloride is precipitated, the addition of excess silver nitrate causes a rapid increase in conductivity. The slope of the initial portion of the curve may be either downward or upward, depending on the relative conductance of the ion being determined and the ion of like charge in the reagent that replaces it. Slow reactions and coprecipitation are sources of difficulty with precipitation and complex-formation titrations.

COULOMETRIC TITRATIONS

Theory

A coulometric titration uses an electrolytically generated titrant for reaction with the analyte. In some analyses, the active electrode process involves only generation of the reagent. In other titrations, the analyte may also be directly involved at the generator electrode (15). The current in a coulometric titration is carefully maintained at a constant and accurately known level. The product of this current and

the time required to reach the equivalence point for the reaction yield the number of coulombs and thus the number of equivalents involved in the electrolysis. The coulomb (C) is the quantity of electricity that is transported in 1 by a constant current of 1 ampere. The Faraday constant (F) is the quantity of electricity that produces one equivalent of chemical change at an electrode (16).

The so-called primary titration technique is attempted only with electrodes of silver metal, silver–silver halide, or mercury amalgams, which are the source of the electrogenerated species. The substance to be determined reacts directly at the electrode or with a reactant electrogenerated from the working electrode. This class of titrations is limited generally to nondiffusible reactants such as mercury amalgams, silver ions generated by anodization of silver metal, and halides liberated by reduction of the appropriate silver–silver halide electrode.

In secondary coulometric titrations, an oxidation–reduction buffer serves as the titrant precursor. An active intermediate from the titrant precursor must first be generated with 100% efficiency by the electrode process. The intermediate must react rapidly and completely with the substance being analyzed.

The endpoint may be detected by addition of colored indicators, provided the indicator itself is not electroactive. Potentiometric and spectrophotometric indication is used in acid–base and oxidation–reduction titrations. Amperometric procedures are applicable to oxidation–reduction and ion-combination reactions especially for dilute solutions.

The most obvious advantage of coulometric titration is the elimination of problems associated with the preparation, standardization, and storage of standard solutions. This is particularly important for labile reagent substance such as chlorine or bromine. Another advantage is the requirement of a small quantity of reagent. Instrumentation is simple; a single constant-current source can be used to generate precipitation, oxidation–reduction, or neutralization reagents. Furthermore, coulometric titration is easily automated.

Typical sources of error in coulometric titrations are variation in the current during electrolysis and errors in measurement of current and time. Departure of the process from 100% current efficiency is the primary error source. In some cases, the equivalence point is not the endpoint of the titration.

Applications

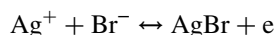
Neutralization titrations

Both weak and strong acids can be titrated with a high degree of accuracy using electrogenerated hydroxide ions.

Table 4 Coulorimetric titrations of oxidation–reduction reactions

Reagent	Substance determined
Br ₂	As ³⁺ , Sb ³⁺ , U ⁵⁺ , NH ₃ , N ₂ H ₄ , NH ₂ OH, phenol, aniline
Cl ₂	As ³⁺ , I [−]
I ₂	As ³⁺ , Sb ³⁺ , S ₂ O ₃ ^{2−} , H ₂ S
Ce ⁴⁺	Fe ²⁺ , Ti ³⁺ , U ⁴⁺ , As ³⁺ , I [−] , Fe(CN) ₆ ^{3−}
Fe ²⁺	Cr ⁶⁺ , Mn ⁷⁺ , V ³⁺ , Ce ⁴⁺
Ti ³⁺	Fe ³⁺ , V ⁵⁺ , Ce ⁴⁺ , U ⁶⁺
C ₂ Cl ₃ [−]	V ⁵⁺ , Cr ⁶⁺ , IO ₃ [−]

In this application, the platinum anode must be isolated by some sort of diaphragm to eliminate potential interference from the hydrogen ions produced. Alternatively, chloride or bromide ions can be added to the solution, and a silver wire used as an anode. The reaction at this electrode is:



Thus, this anode product does not interfere with the neutralization reaction.

Strong and weak bases can be titrated with hydrogen ions generated at a platinum anode.

Here also the cathode must be isolated from the solution to prevent interference from the hydroxide ions produced at that electrode.

Precipitation and complex-formation titrations

Anodically generated silver ions can be used for precipitation titrations of various substances. A cell constructed from a piece of heavy silver wire can be used. Substances precipitated include Cl[−], Br[−], I[−], and mercaptans. Similar applications using mercury(I) ion formed at a mercury anode have been used for the determination of Cl[−], Br[−], and I[−].

The complexing ability of ethylenediaminetetra-acetic acid (EDTA, H₄Y) has been exploited in the coulometric titration of metal ions. The method depends on the reduction of the mercury(II) or cadmium chelate of EDTA and on the titration of the metal ion (for example, magnesium) to be determined by the anion of EDTA that is released.

Oxidation–reduction titrations

Applications of coulometric titrations involving oxidation–reduction reactions are shown in Table 4. Electrogenerated oxidizing agents such as bromine have proved to be useful, especially in organic analysis.

AMPEROMETRIC TITRATIONS

Theory

Polarographic methods can be used to estimate the equivalence point of a reaction, provided that at least one of the participants or products of the titration is oxidized or reduced at the microelectrode. When the potential applied across the two electrodes is maintained at some constant value, the current may be measured and plotted against the volume of the titrant, thus, the term amperometric titration. In the case of working electrode-reference electrode pair, the potential of the indicator electrode is maintained at a constant value with respect to a reference electrode, measuring a limiting current, which is proportional to the concentration of one or more of the reactants or products of the titration.

Amperometric titration is easily automated. The titrator can be programmed to shut off when a specified current level is reached. Titrant is run into a blank solution until a specified current is reached, the sample is added, and the titrant is again added until the specified current level is reached.

A correction for dilution is necessary to attain a linear relationship between current and volume of titrant. By working with a reagent that is 10-fold more concentrated than the solution being titrated, this correction becomes negligible. Amperometric and conductometric titrations are similar in the respect that the data for each are collected well away from the equivalence point. Therefore, reactions that are relatively incomplete can be used.

Titration Curves

Typical forms of amperometric titration curves are shown in Fig. 13. A titration in which the substance being analyzed reacts (is reduced) at the electrode while the reagent does not is shown in Fig. 13a. A sufficiently high potential is applied to give a diffusion current for

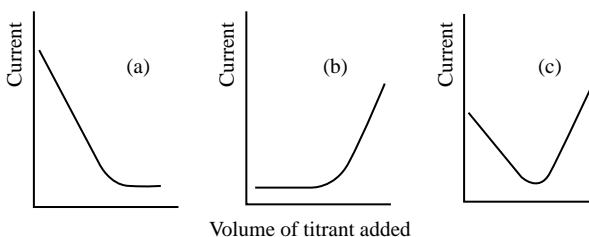


Fig. 13 Amperometric titration curves: (a) analyte is reduced, reagent is not; (b) reagent is reduced, analyte is not; (c) both reagent and analyte are reduced.

the substance, and a linear decrease in current is observed as substance ions are removed from the solution by precipitation. The curvature near the equivalence point reflects the incompleteness of the analytical reaction in this region. The endpoint is obtained by extrapolation of the linear portions. Nearly a mirror image curve is obtained for a titration in which the reagent reacts (is reduced) at the microelectrode, and the substance being analyzed does not (Fig. 13b). The third common curvature for amperometric titrations is obtained when both reagent and substance analyzed react (are reduced) at the electrode (Fig. 13c).

Amperometric Titrations with Two Polarized Microelectrodes

A modification of amperometric method involves the use of two stationary microelectrodes immersed in a well-stirred solution of a sample (17). A small potential is applied between these electrodes, and such current that flows is followed as a function of the volume of reagent added.

Oxidation–reduction titrations

Twin polarized platinum microelectrodes are conveniently used for endpoint detection for oxidation–reduction titrations. Consider a titration curve for oxidation–reduction titration where both reactants behave reversibly at the electrodes. An example of this kind of titration is titration of iron(II) with cerium(IV) (Fig. 14a). At the starting point of the titration, no current is observed because no suitable cathode reactant is available. With addition of cerium(IV), a mixture of iron(II) and iron(III) is produced, which permits the passage of current. Beyond the midpoint in the titration, iron(III) becomes in excess, and the current is then regulated by decreasing iron(II) concentration. At the equivalence point, the current approaches zero because iron(III) are present, and the applied potential is not great enough to cause these to react at the electrode. Beyond the equivalence point, the current

rises again because both cerium(III) and cerium(IV) are present to react at the electrodes.

In cases where only reagent behaves reversibly, a different form for titration curve is obtained. Although the reagent added can serve as an anode reactant, no cathode reactant is available because of the slow rate at which the substance analyzed is reduced at a platinum surface. Therefore, no current is observed. Beyond the equivalence point, depolarization of the cell can occur, and the current is dependent on the concentration of the reagent (Fig. 14b). In cases where only the species titrated behave reversibly, before equivalence point, a current is observed that depends on the concentration of the species present in lesser amount. At equivalence point, a zero current is reached, and beyond the equivalence point, no current is observed because the reagent does not behave reversibly at the electrodes (Fig. 14c).

Precipitation titrations

Twin silver microelectrodes permit observation of the endpoint for various titrations using silver nitrate as precipitation reagent (Table 5). Short of the equivalence point, essentially no current exists when a small potential is applied between two such electrodes during a titration of substance analyzed with silver ions because no easily reduced species is present in the solution. After equivalence, the cathode becomes depolarized owing to the presence of a significant amount of silver ions that can react to give silver. Current is permitted as a result of this half-reaction and the corresponding oxidation of silver at the anode. The magnitude of the current is directly proportional to the concentration of the excess reagent.

PHOTOMETRIC TITRATIONS

The change in absorbance of a solution may be used to follow the change in concentration of a radiation-absorbing

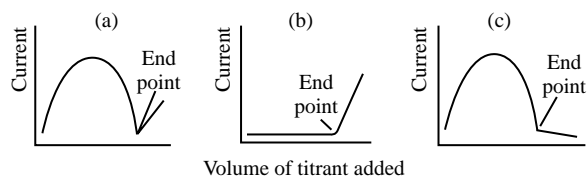


Fig. 14 Amperometric titration curves with twin polarized electrodes: (a) both reactants behave reversibly at the electrode; (b) only reagent behaves reversibly; (c) only analyte titrated behaves reversibly.

Table 5 Applications of amperometric titrations with precipitation products

Reagent	Substance determined
AgNO ₃	Cl [−] , Br [−] , I [−] , CN [−] , RSH
Cupferron	Cu ²⁺ , Fe ³⁺
Dimethylglyoxime	Ni ²⁺
Pb(NO ₃) ₂	SO ₄ ²⁺ , MoO ₄ ²⁺ , F [−] , Cl [−]
8-Hydroxyquinoline	Mg ²⁺ , Cu ²⁺ , Zn ²⁺ , Cd ²⁺ , Al ³⁺ , Bi ³⁺ , Fe ³⁺
K ₂ CrO ₄	Pb ²⁺ , Ba ²⁺

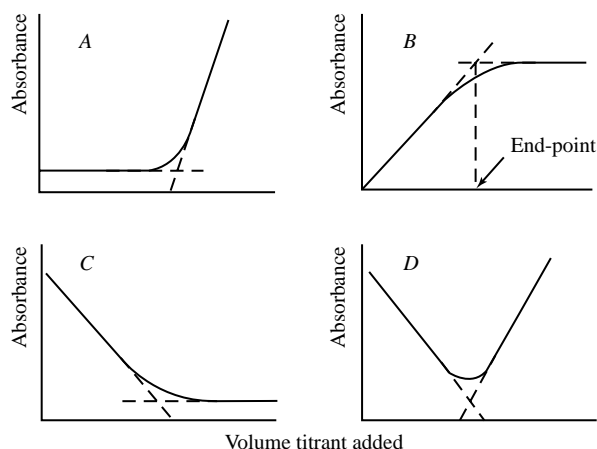


Fig. 15 Different shapes of photometric titration curves.

constituent during a titration (18). The absorbance is directly proportional to the concentration of the absorbing constituents. A plot of absorbance versus titrant consists, if the reaction is complete, of two straight lines that intersect at the endpoint. If the reaction is appreciably incomplete, extrapolation of the two linear segments of the titration curve establishes the intersection and endpoint volume.

Photometric titrations have several advantages over a direct photometric determination. The presence of other absorbing species at the analytical wavelength does not necessarily cause interference because only the change in absorbance is significant. However, if the absorbance of nontitratable components is intense, the absorbance readings are shifted to the undesirable upper end of the absorbance scale. Ideally, only a single absorber is present among the reactant, titrant, and products.

The analytical wavelength is selected to avoid the interference caused by other absorbing substances. Also, there is a need for a molar absorptivity that causes the change in absorbance during the titration to fall within a convenient range. Often, the chosen wavelength lies well apart from an absorption maximum.

Titration Curves

Volume change, caused by addition of titrant, is seldom negligible, and straight lines for titration curve are obtained only if there is correction for dilution.

If correction is not made, the lines are curved downward toward the volume axis, and erroneous intersections are obtained. Use of a microsyringe and a relatively concentrated titrant is desirable.

Possible shapes of photometric titration curves are shown in Fig. 15. Curve A is typical of the titration where

the titrant alone absorbs and where the absorbance readings are taken at the wavelength of the titrant. Curve B is characteristic of systems where the product of the reaction absorbs, and curve C is typical for reactions where the analyte is converted to a nonabsorbing product. When a colored analyte is converted to a colorless product by a colored titrant, curves similar to D are obtained.

Applications

Areas of particular applicability are for solutions so dilute that the indicator blank is excessive or when the color change is not sharp perhaps because titration reactions that are incomplete in the vicinity of the equivalence point or when extraneous colored materials are present in the sample. Titrations in solutions of high or low ionic strength and even in nonaqueous solution are easily performed. Sensitivity of the measurement can be changed easily by simply changing the wavelength or the length of the cell path, a feature that also makes photometric titration attractive. Indicators can be used in titrations where self-indicating systems are lacking.

REFERENCES

1. Grant, J. *Sutton's Volumetric Analysis*, 13th Ed.; Butterworth's Scientific Publications: London, 1955; 28–29.
2. Skoog, D.A.; West, D.M. *Fundamentals of Analytical Chemistry*, 2nd Ed.; Holt, Rinehart, Wiston: London, 1972; 225.
3. Skoog, D.A.; West, D.M. *Fundamentals of Analytical Chemistry*, 2nd Ed.; Holt, Rinehart, Winston: London, 1972; 226–227.
4. Laitinen, H.A. *Chemical Analysis*; McGraw-Hill: New York, 1960; 37.
5. Grant, J. *Sutton's Volumetric Analysis*, 13th Ed.; Butterworth's Scientific Publications: London, 1955; 140.
6. Mitchell, J. *Anal. Chem.* **1951**, 23, 1069.
7. Meites, L. *Handbook of Analytical Chemistry*; McGraw-Hill: New York, 1963; 3–226.
8. Reilley, C.N.; Barnard, A.J., Jr. *Handbook of Analytical Chemistry*; McGraw-Hill: New York, 1963; 3–166 to 3–200.
9. Gran, G. *Acta Chem. Scand.* **1950**, 4, 559.
10. Gran, G. *Analyst* **1952**, 77, 661.
11. Reilley, C.N.; Schmid, R.W. *Anal. Chem.* **1958**, 30, 947.
12. Reilley, C.N.; Schmid, R.W.; Lamson, D.W. *Anal. Chem.* **1958**, 30, 953.
13. Foreman, J.K.; Stockwell, P. *Automatic Chemical Analysis*; Wiley: New York, 1975; 42–62.

14. Shadlovsky, T.; Weissberger, A. *Physical Methods of Organic Chemistry*, 3rd Ed.; Interscience: New York, 1960; 1, 3011–3048.
15. Clem, R.G. *Ind. Research* **1973**, (September), 50.
16. Skoog, D.A.; West, D.M. *Fundamentals of Analytical Chemistry*, 4th Ed.; Holt, Rinehart, Winston: Hong Kong, 1986; 443–444.
17. Lingane, J.J. *Electroanalytical Chemistry*; Interscience: New York, 1958; 280–294.
18. Goddu, R.F.; Hume, D.N. *Anal. Chem.* **1954**, 26, 1679.
19. Headridge, J.B. *Photometric Titrations*; Pergamon Press: New York, 1961.
20. Laitinen, H.A.; Harris, W.E. *Chemical Analysis*, 2nd Ed.; McGraw-Hill: New York, 1975.
21. Lingane, J.J. *Electroanalytical Chemistry*; Interscience: New York, 1958.
22. Ma, T.S.; Hassan, S.S.M. *Organic Analysis Using Ion-Selective Electrodes*; Academic Press: London, 1982; 1, 2.
23. Rechnitz, G.A. Bioanalysis with Potentiometric Membrane Electrodes. *Anal. Chem.* **1982**, 54, 1194A.
24. Serjeant, E.P. *Potentiometry and Potentiometric Titrations*; John Wiley: New York, 1984.
25. Svehla, G. *Automatic Potentiometric Titrators*; Pergamon Press: New York, 1978.

TONICITY

Jaymin C. Shah

Pfizer Inc., Groton, Connecticut, U.S.A.

INTRODUCTION

Parenteral formulations, both large and small volume, have been discussed in depth in Volume 11 of the *Encyclopedia of Pharmaceutical Technology* (V 11, pp. 201–217, 217–237). However, no discussion of parenteral formulations is complete without an adequate description of tonicity. Tonicity is an important factor in the formulation of products intended for application to sensitive mucous membranes of organs such as eye, ear, and nose. In this article, an attempt is made to first introduce tonicity with respect to its physiological significance, followed by a discussion of the physicochemical basis for tonicity and colligative properties. Then, a brief review of methods of measuring and/or calculating tonicity is given, followed by the established methods of adjusting tonicity and the examples illustrating each of the methods. Tables listing the established values necessary to do the calculations are provided at the end of the article. Excellent comprehensive reviews dealing with various aspects of tonicity are available in the literature in the form of chapters in various textbooks for pharmacy students (Remington's *Pharmaceutical Sciences*, *Physical Pharmacy* by Martin, etc.). A list of such reading material is provided at the end of the article in the Bibliography.

Dosage forms are drug-delivery systems designed to deliver drug to the systemic circulation or to a localized region of the human body. These dosage forms should ideally be free of any undesired adverse effects from the drug and from the formulation components. Reasonable risks associated with the drug substance are sometimes tolerated with an objective of realizing significant therapeutic advantages, as in the case of cancer chemotherapeutic agents. However, any untoward side effect, even as minor as irritation, resulting from an excipient or the finished dosage form cannot be accepted and should not be tolerated. This concern is particularly important to parenteral formulations that breach the normal defensive barriers of the human body to deliver the drug. Therefore, any formulation that comes in contact with sensitive mucous membranes of organs such as the eye should not result in tissue irritation and pain attributable to the formulation itself. One of the

physicochemical means by which a formulation may result in pain and tissue irritation is caused by the nonphysiological concentration of dissolved solutes coming in contact with sensitive tissues. Tonicity is a formulation property that has a direct influence on the ability of the formulation to result in tissue irritation, as described by the following example.

If a small quantity of blood defibrinated to prevent clotting is mixed with a solution containing 0.9% w/v of NaCl, the red blood cells remain intact and retain their normal size and shape. The NaCl solution is considered to be isotonic and has essentially the same salt concentration as does the red blood cell. In contrast, if the blood is mixed with 1.8% w/v NaCl solution, erythrocytes shrink and become wrinkled or crenated as if the cell content has been sucked out. The salt solution that causes this is considered hypertonic with respect to the red blood cell contents. It is because the red blood cell contains a lower salt concentration than the surrounding 1.8% w/v salt solution and as if the water from the erythrocytes passes through the cell membrane to dilute the surrounding salt solution to equalize the two salt concentrations across the membrane. The opposite phenomenon occurs if blood is mixed with 0.45% w/v NaCl solution. Water from the surrounding salt solution enters the erythrocytes, causing them to swell and finally burst, with the liberation of hemoglobin. The 0.45% w/v salt solution is considered hypotonic, and the phenomenon is known as hemolysis. The physiological significance of hemolysis was reconfirmed recently by the report of 10 episodes of hemolysis among patients who received hypotonic 25% human albumin because of dilution with sterile water instead of isotonic sodium chloride (1). Two of these 10 recipients exhibited significant hemolysis and adverse pathological conditions, with one resulting in death (1). Also, it has been observed that hypertonic and hypotonic salt solutions tend to irritate sensitive mucous membranes of the eye, the nose, and the muscle when applied. However, an isotonic solution causes no tissue irritation when it comes in contact with the tissue. The crenation and the hemolysis of red blood cells in hypertonic and hypotonic salt solution, respectively, can be explained by the movement of water across the cell membrane. A membrane is defined as semipermeable if it allows only the movement of solvent molecules

across it. The process of diffusion of a solvent through a semipermeable membrane from a less-concentrated solution to a more-concentrated solution is known as osmosis. The pressure that must be applied to the concentrated solution side of the membrane to prevent the flow of pure solvent across the membrane from the diluted solution is known as the osmotic pressure. In crenation, water diffuses from the inside of the erythrocyte across the membrane into the exterior hypertonic salt solution. Hemolysis occurs when water diffuses from the exterior hypotonic solution into the erythrocyte, causing it to swell and burst. An isotonic solution is an aqueous solution that generates the same "tone," or osmotic pressure, as the body fluids across biological membranes and thus prevents water flow in either direction and hence is nonirritating when injected, instilled, perfused, or brought into contact with sensitive mucous membranes. When a solution is hypertonic or hypotonic, osmotic water flow occurs and tone of the membrane is affected. Thus, formulators need to adjust the tone, or tonicity, of the solution to be isotonic with physiological fluids. To be able to adjust the tonicity of a formulation to the isotonic state, one has to understand the principles behind the generation of the osmotic pressure resulting from the dissolved solutes and how it can be altered to that of the physiological fluids.

Osmotic pressure is a colligative property unlike the additive and constitutive properties of solution. Simply stated, the colligative properties of a solution are dependent solely on the number of nonsolvent (solute) particles (molecule/ions) dissolved or in a true solution form in a given solvent and are independent of the specific physicochemical characteristics of the nonsolvent dissolved substance(s). For example, two nonelectrolyte solutes A and B when prepared as 0.1M solutions will exhibit the same osmotic pressure irrespective of the chemical nature of A and B. Experimentally, it has been found over the years that the colligative properties are indeed independent of the solute nature and dependent solely on the number of independent particles in dilute solutions for a wide variety of solutes, provided the number of particles is properly assessed. An implicit assumption in the statement above is that the solute is nonvolatile relative to the solvent and the reasoning will be clear from the discussion that follows. The colligative properties of solution are vapor pressure lowering, boiling-point elevation, freezing-point depression, and osmotic pressure. These four properties are effects of solute on the solvent, in that it reduces the escaping tendency of the solvent, and all of them can be related to vapor pressure lowering of the solution. Osmotic pressure is of primary importance from the formulation standpoint; however, it is

cumbersome to measure, and therefore other colligative properties are determined because they are all interrelated. The theory of the colligative properties has been well-established and reviewed in depth in the various textbooks and in the pharmaceutical literature (see the Bibliography). And because the focus of this *Encyclopedia* is on pharmaceutical technology rather than on the theoretical foundation, we briefly overview the theory of colligative properties as needed to understand the concepts used in methods to adjust tonicity of parenteral and ophthalmic formulations.

RAOULT'S AND HENRY'S LAWS AS BASIS FOR COLLIGATIVE PROPERTIES

Raoult's law states that in an ideal solution, the partial vapor pressure of each volatile constituent is equal to the vapor pressure of the pure component multiplied by its mole fraction in the solution. Thus, for two constituents A and B in solution:

$$P_B = P_B^o X_B \quad (1)$$

$$P_A = P_A^o X_A \quad (2)$$

where P_A and P_B are the partial vapor pressures of constituents A and B over their solution when the fractional molar concentrations are X_A and X_B , and the vapor pressure of the pure constituents are P_A^o and P_B^o . Therefore, it can be inferred that the vapor pressure of B above the solution by dilution with A is reduced relative to its vapor pressure in pure state and vice versa for A. This diminishes the escaping tendencies of each component, leading to a reduction in the rate of escape of the molecules of A and B from the surface of the solution. This law is valid only in ideal solutions in which there are no intermolecular interactions between components A and B (adhesive interactions) or in which interactions between the two components A and B are identical to the interactions of the pure components A and pure B (cohesive interactions between A and A and between B and B). In essence, the molecule of each component sees an environment identical to its molecular environment in the pure state. This refers to an infinitely dilute solution in which a component's thermodynamic activity is equal to its concentration. However, in the real solutions, the assumption noted above may not apply, and negative deviation from Raoult's law may occur when adhesive attractions between A and B are greater than cohesive attraction within pure A or pure B molecules; i.e., the vapor pressure of the solution or the partial vapor pressure

of each component is lower than that expected based on Raoult's law applied to ideal solution. Similarly, positive deviations from Raoult's law can occur when interactions between A and B are less than the cohesive interactions of pure A or pure B, resulting in vapor pressures higher than that expected based on Raoult's law applied to ideal solution. In general, Raoult's law states that when a component A is diluted with another component B, the partial vapor pressure of A is reduced; in essence, a dilution effect.

Raoult's law does not apply over the entire concentration range in a nonideal, real solution. However, when one component is in a large enough excess to be considered a solvent, Raoult's law may be expressed as:

$$P_{\text{solvent}} = P_{\text{solvent}}^{\circ} X_{\text{solvent}} \quad (3)$$

in such a dilute solution and is valid only for the solvent component of a nonideal solution that is sufficiently dilute for the other component, i.e., the solute in a dilute nonideal solution. In such a dilute solution, the solute molecule is completely surrounded by solvent molecules such that the solute molecule can interact only with the solvent molecules because there are very few solute molecules. Further dilution beyond this point does not alter a solute molecule's environment and, even if the solute molecule interacts with the solvent molecule, the solute's partial vapor pressure or the thermodynamic activity becomes proportional to its fractional molar composition as:

$$P_{\text{solute}} = K_{\text{solute}} X_{\text{solute}} \quad (4)$$

Eq. 4 is known as Henry's law, and K_{solute} is the Henry's law constant, which is less than $P_{\text{solute}}^{\circ}$. Therefore, Henry's law applies to the solute in dilute solutions, and Raoult's law applies to solvent in dilute nonideal solutions. Note the similarities between Eqs. 1 and 2 and between Eqs. 3 and 4 for the nonideal dilute solution case. When the solution is ideal, Henry's law becomes identical to Raoult's law, and K_{solute} becomes identical to $P_{\text{solute}}^{\circ}$. When the partial pressures of the solute and the solvent are directly proportional to their mole fractions over the entire range, the solution is ideal. In a nonideal solution, Raoult's law will apply to the solvent over the entire concentration range, whereas Henry's law will apply to the solute in a limited concentration range in which it is in a sufficiently diluted form.

When a nonvolatile solute is dissolved in a solvent, the partial vapor pressure of the solvent above the solution is equal to the vapor pressure of the solution. And because

the mole fraction of the solvent is $X_{\text{solvent}} = 1 - X_{\text{solute}}$, Eq. 3 can be rewritten as:

$$P_{\text{solution}} - P_{\text{solvent}} = P_{\text{solvent}}^{\circ} (1 - X_{\text{solute}}) \quad (5)$$

for the dilute solution of a nonvolatile solute of mole fraction X_{solute} . Therefore, the important conclusion from Raoult's and Henry's laws is that the thermodynamic activity of a solvent as measured by its vapor pressure is proportional solely on the mole fractional composition of the solute, irrespective of the physical and chemical nature of the dissolved species. The vapor pressure of the solvent above the solution thus depends solely on the number of particles (molecules/ions) of the dissolved solute and not on the weight concentration of the solute in solution. Therefore, the vapor pressure of a solvent above a dilute solution that obeys Henry's law is a colligative property of the solution. Henry's law has been found to be applicable to nonelectrolyte-type solute mole fractional concentrations of 0.1; however, the range is much smaller for electrolyte type solutes owing to the long-range nature of interionic interactions as noted above. Its impact is evaluated later below.

COLLIGATIVE PROPERTIES

Raoult's law forms the basis for the colligative properties, and Henry's law sets the limits of the applicability of Raoult's law to colligative properties of a solution as increasing amounts of solute are added to solution. As noted above, colligative properties are a consequence of the number of dissolved particles in solution and are all related to the escaping tendency of the solvent from solution.

The four colligative properties that are of importance are: 1) the vapor pressure lowering, 2) the elevation of boiling point, 3) the freezing-point depression, and 4) the osmotic pressure. An attempt is made below to describe qualitatively and quantitatively each colligative property of solutions, with an emphasis on their interrelationship and their application later in measurement and adjustment of the tonicity of solutions, with particular reference to parenteral formulations. Although theoretical derivations based on thermodynamics can be used to show how each of the colligative properties of solution arises and relate to each other, textbooks on physical chemistry for theoretical derivations are recommended.

Lowering of Vapor Pressure

The addition of a nonvolatile solute to a solvent leads to a reduction in the vapor pressure of the solvent because of a

reduction in thermodynamic activity of the solvent. Also, because the solute is nonvolatile, the vapor pressure of the solvent is the vapor pressure of the solution, as seen from Eq. 5. Qualitatively, one can imagine that fewer numbers of solvent molecules are escaping per unit surface area of the solution than from the pure solvent because fewer solvent molecules are present per unit surface area of the solution owing to displacement by solute molecules. However, these solute molecules will not affect the condensation of solvent molecules with insufficient kinetic energy present in the vapor phase. The result is a net reduction in escaping tendency of solvent molecules on the surface, causing a lowering of vapor pressure and, consequently, the rate of vaporization. The resulting vapor pressure lowering ($P_{\text{solvent}}^{\circ} - P_{\text{solution}}$) and the relative vapor pressure lowering as a function of mole fractional concentration of solute can be obtained by rearranging Eq. 5 to Eq. 6:

$$\frac{P_{\text{solvent}}^{\circ} - P_{\text{solution}}}{P_{\text{solvent}}^{\circ}} = \frac{\Delta P}{P_{\text{solvent}}^{\circ}} = X_{\text{solute}} \quad (6)$$

The left term is the relative vapor pressure lowering, which is solely dependent on the mole fraction concentration of a single solute or the sum of mole fraction of each solute dissolved in the solution. Thus, the relative vapor pressure lowering is a direct measure of the total number of dissolved solute particles, irrespective of their physicochemical nature. The mole fractions can be converted into molality (m moles of solute per 1000 g of solvent) to result in the following equation for water as the solvent:

$$\frac{\Delta P}{P_{\text{solvent}}^{\circ}} = X_{\text{solute}} \cong \frac{m M_{\text{solvent}}}{1000} \quad (7)$$

$\cong 0.018 m$ for Aq. solutions

where m is the concentration of solute expressed in molality, and M_{solvent} is the molecular weight of the solvent in grams. For water, $M = 18$, and because the density of water is close to 1, for dilute aqueous solutions, the molality and molarity (moles/liter) can be used interchangeably, and Eq. 7 can be used to calculate the relative vapor pressure lowering from the molar concentration of the nonelectrolyte solute.

Elevation of the Boiling Point

The boiling point of a liquid is the temperature at which the vapor pressure of the liquid becomes equal to the external pressure acting on the liquid, which is 760 mm Hg at one atmospheric pressure. Therefore, the boiling point of a solution of nonvolatile solute will be higher than that of the pure solvent owing to the solute reducing the vapor

pressure of the solvent above the solution according to Raoult's law. The solution has to be heated to a higher temperature to achieve the same vapor pressure to result in boiling of the solvent. The elevation of the boiling point ($T_{\text{solution}} - T_{\text{b}}^{\circ}$) is directly proportional to the relative vapor pressure lowering based on the Clausius–Clapeyron equation, resulting in an equation relating it to molality as follows:

$$(T_{\text{solution}} - T_{\text{b}}^{\circ}) = \delta T_{\text{b}} = \Delta T_{\text{b}} = K \Delta P = K_{\text{b}} m \quad (8)$$

where in K_{b} is called the molal elevation constant or the ebullioscopic constant, a characteristic constant for each solvent, and is the boiling point elevation of an ideal 1-molal solution of a nonvolatile solute. K_{b} can be obtained by measuring the $\delta T_{\text{b}}/m$ of several molal concentrations of solute in solutions and extrapolating the $\delta T_{\text{b}}/m$ versus molality curve to zero solute concentration. The value of K_{b} for water is $0.515^{\circ} \text{ kg/mole}$. By measuring the boiling point elevation of a solvent in a solution and knowing the K_{b} for that solvent, one can calculate the molal concentration of a solute in the solution. From Eq. 8, it is evident that the elevation of boiling point is a colligative property like vapor pressure lowering because it is strictly dependent on the molal concentration of the solute: the number of particles in solute and, thus, independent of the physicochemical nature of the solute.

Freezing-Point Depression

The freezing point of a liquid or the melting point of a solid phase of a pure compound is the temperature at which the solid and liquid phases are in equilibrium at a pressure of 1 atm. The freezing point of a pure compound is described by a unique point in the phase diagram of the compound, and, at that point, the solid and liquid phases are in equilibrium, and the vapor pressure of the liquid phase coincides with the vapor pressure of the pure solid phase. Because in a solution, the vapor pressure of the solvent is lowered relative to the pure solvent, no freezing (or crystallization) takes place at the equilibrium temperature of the liquid and solid phases of the pure solvent; i.e., the freezing point. The phase with the lower vapor pressure is the more stable phase thermodynamically. Therefore, cooling of the solution below the freezing point of the pure solvent results in a greater reduction in vapor pressure of the pure solid phase than the solution phase, and when the vapor pressure of the two phases eventually coincides, freezing (crystallization) of the pure liquid solvent occurs. The dissolved solute reduces the escaping tendency of the solvent molecules to crystallize, and thus the temperature must be reduced to reestablish

equilibrium between the solid and liquid phases and hence the depression of freezing point. It must be noted that the dissolved solute only affects the freezing of the liquid phase and does not alter the tendency of the molecules to leave the solid phase, although both processes are occurring in equilibrium at the freezing point. Also, the solvent should form a pure solid; if the solute cocrystallizes with the solvent, the phenomenon is complex and cannot be described as a colligative property. The more concentrated the solution, the greater the freezing-point depression, and using thermodynamic principles, Raoult's law, and Clausius–Clapeyron equation, the freezing-point depression can be related to solute concentration expressed in molality as follows:

$$(T_{\text{solution}} - T_f^o) = \delta T_f = \Delta T_f = K \Delta P = K_f m \quad (9)$$

where in K_f is called the molal depression constant or the cryoscopic constant, a characteristic constant for each solvent dependent on the physicochemical nature of the solvent, and is the freezing-point depression of an ideal 1-molal solution of a nonvolatile solute. K_f can be obtained experimentally by measuring the $\delta T_f/m$ of several molal concentrations of solute in solutions and extrapolating the $\delta T_f/m$ versus molality curve to zero solute concentration. The value of K_f for water is $1.86^\circ \text{ kg/mole}$. By measuring the freezing-point depression of a solvent in a solution and knowing the K_f for that solvent, one can calculate the molal concentration of a solute in the solution. Freezing-point depression is a colligative property as seen from Eq. 9, because it is proportional to molal concentration of solute; the number of particles in solution, and not on the physicochemical characteristics of the solute.

Osmotic Pressure

As described in the Introduction, the process of diffusion of a solvent through a semipermeable membrane from a less-concentrated solution into a more-concentrated solution is osmosis. This results in the development of a hydrostatic pressure head on the more-concentrated solution side of the membrane. Alternatively, pressure may be applied to the more-concentrated solution side of the semipermeable membrane to prevent the diffusion of solvent. This applied pressure on the concentrated solution is identical to the hydrostatic pressure head that may develop owing to osmosis. It is known as the osmotic pressure and is directly proportional to the solute concentration in an ideal solution. A semipermeable membrane is one that allows the movement of only solvent molecules, and if the membrane is not semipermeable,

osmosis may not be observed because the solute will diffuse quickly through the membrane to equalize the concentration on two sides of the membrane.

Osmosis tends to equalize the escaping tendencies of the solvent on both sides of the semipermeable membrane. Escaping tendency can be measured in terms of partial vapor pressure of solvent above the solution. Alternatively, one can see that at the beginning, the solution and pure solvents have different thermodynamic activities for the solvent because they have different vapor pressures. For the solution and pure solvent on two sides of the semipermeable membrane to be in equilibrium, they should have identical escaping tendency or identical vapor pressure and, thus identical thermodynamic activity. The equilibrium is therefore established by the generation of the osmotic pressure that compensates for the difference in solvent concentration on the two sides of the membrane, which is responsible for the different vapor pressures, escaping tendencies, and thermodynamic activity. Therefore, osmosis is a process to reach equilibrium state whereby solvent spontaneously flows from the high-free-energy (low-vapor-pressure) side of the membrane to the low-free-energy (high-vapor-pressure) side of the membrane, until the solvent's free energies on both sides of the membrane are equal and identical. Obviously, the solute will not be able to attain equilibrium because it cannot diffuse through the semipermeable membrane. Because osmotic pressure is attributable to the difference in vapor pressure of the solvent above solution, it is also a colligative property as explained below.

Using thermodynamics and considering free energy of the solvent as a function of vapor pressure, the osmotic pressure (π) that develops when a solution is separated from pure solvent by semipermeable membrane can be related to vapor pressures as shown below:

$$\Pi = \frac{RT}{V_{\text{solvent}}^M} \ln \frac{P_{\text{solvent}}^o}{P_{\text{solution}}} \quad (10)$$

where in R is the gas constant, T is temperature, V_{solvent}^M is the partial molar volume, (the volume occupied by 1 mole of solvent), and π is the developed osmotic pressure when a solution with vapor pressure P_{solution} is separated from a solvent with vapor pressure P_{solvent}^o by a semipermeable membrane. Applying Raoult's law and substituting mole fractions for the vapor pressures from Eq. 5 into Eq. 10 results in the following:

$$\Pi = \frac{-RT}{V_{\text{solvent}}^M} \ln(1 - X_{\text{solute}}) \cong \frac{RT}{V_{\text{solvent}}^M} X_{\text{solute}} \quad (11)$$

since $\ln(1 - X_{\text{solute}}) \cong -X_{\text{solute}}$

In a dilute solution, X_{solute} is approximately equal to the molar ratio $n_{\text{solute}}/n_{\text{solvent}}$, and Eq. 11 becomes:

$$\Pi = \frac{n_{\text{solute}}}{n_{\text{solvent}} V_{\text{solvent}}^M} RT = \frac{n_{\text{solute}}}{V_{\text{solution}}} RT = mRT \quad (12)$$

in which the number of moles of solvent multiplied by the partial molal volume is equal to the volume of the solvent in solution. In a dilute solution, the volume of solvent can be approximated to the volume of solution, which results in the above equation relating osmotic pressure to molar or molal concentration of a solute in solution. Eq. 12 is known as Morse's expression and demonstrates how osmotic pressure is a colligative property directly proportional only on the number of particles dissolved in the solvent irrespective of the nature of the solute. Van't Hoff had recognized early on that there is a direct proportionality among osmotic pressure and concentration of solute and temperature, and suggested a relationship that was similar to the equation for an ideal gas as follows:

$$\Pi V_{\text{solution}} = n_{\text{solute}} RT \quad (13)$$

Eq. 13 is analogous to the ideal gas equation, and van't Hoff concluded that osmotic pressure of a dilute solution was a pressure that the solute would exert if it behaved like a gas occupying that volume. Eq. 13 can also be expressed as:

$$\Pi = \frac{n_{\text{solute}}}{V_{\text{solution}}} RT = cRT \quad (14)$$

which shows that osmotic pressure is directly proportional to the concentration of solute expressed in molarity. This equation is similar to Morse's expression, Eq. 12; however, it has been shown theoretically and experimentally that more accurate results can be obtained when solute concentration is expressed in molality rather than in molarity. Although the resemblance of Eq. 13 to the ideal gas equation is striking, osmotic pressure is a result of differences in the escaping tendencies of the solvent on two sides of the membrane rather than of the behavior of a solute such as a gas.

From Eqs. 12–14, one finds that 1-molar solution of any solute will generate an incredibly high osmotic pressure of approximately 24 atm at room temperature, which has been verified experimentally. Although this estimate is based on the assumption that the solution is dilute and behaving ideally, at high concentrations of solute, the theory overestimates the experimental findings. The discussion above deals primarily with the thermodynamic basis for the generation of osmotic pressure; however, it does not address the issue of how fast the equilibrium is attained or how fast the osmotic pressure will be generated. The rate of generation of osmotic pressure is a kinetic process and depends to a great extent

on the characteristics of the semipermeable membrane. Red blood cell membrane or the mucous membrane in the eye are very thin and moist, and water can diffuse very rapidly through the membrane to generate the enormous osmotic pressure. However, the osmotic pressure may develop very slowly across synthetic and semisynthetic polymeric membranes, across which the diffusion is very slow.

COLLIGATIVE PROPERTIES OF ELECTROLYTES AS COMPARED WITH NONELECTROLYTE SOLUTES

The colligative properties, by definition, should be independent of the nature of the solute. Therefore, 0.1-molal solutions of sucrose and NaCl should exhibit similar colligative properties. It was observed by van't Hoff that colligative properties of dilute solutions of nonelectrolytes such as sucrose were expressed satisfactorily by the equations above. However, solutions of strong electrolytes such as salts gave osmotic pressure twice or three times as large as would be expected based on Eq. 14, depending on the electrolyte investigated. To account for this anomaly, van't Hoff proposed the following modification of Eq. 14 as shown below:

$$\Pi = i c RT \quad (15)$$

in which i can be considered to be a factor to account for the deviation of concentrated solutions of electrolytes and also nonelectrolytes from Raoult's law as applied to ideal solutions. After Arrhenius developed the theory of ionization or dissociation of salts into ions, van't Hoff and others recognized that the value of i approached or equaled the number of ions into which the electrolyte or the molecule dissociated as the solution was made more dilute. For example, a dilute 0.1- M solution of NaCl would be twice as active osmotically as a 0.1- M solution of sucrose, and i for NaCl was two. Similarly, 0.1- M solutions of CaCl_2 and MgCl_2 would generate three times the osmotic pressure of 0.1- M sucrose solution, and i for both is equal to three. Therefore, it was realized that i reflected the number of ions the electrolytes dissociated into and, thus, electrolytes at equimolar concentrations were more effective in generating osmotic pressure based on the number of ions they produced on dissociation. However, it was also observed that at moderate concentrations of electrolytes, osmotic pressures were less than that expected based on complete dissociation. In fact, this led the scientific community to suggest partial dissociation for even strong electrolytes and to use

colligative properties as a measure of degree of dissociation. However, we know now that strong electrolytes do dissociate completely even in concentrated solutions from other measurements such as conductivity techniques. The lower-than-expected values of osmotic pressure in moderate concentrations of electrolytes is attributable to the influence of long-range ionic interactions that come into play as the solution gets increasingly concentrated. The basic assumption in Raoult law was that there was no interaction between solute particles and, even if there was any, it should equal that between the solute and the solvent. However, the strong attractive forces between ions of opposite charges do predominate in increasingly concentrated solutions of electrolytes, and their thermodynamic activity is reduced relative to that in an infinitely dilute solution. Also, the effect of ionic strength of a solution has been shown to influence the activity of electrolytes, and thus, it is the interionic forces rather than the partial dissociation that seems to influence the colligative properties of electrolytes being lower than expected based on complete dissociation.

All the colligative properties of all solutes with the modification by the van't Hoff factor i can be expressed as:

$$\begin{aligned}\Delta P &= i P_{\text{solvent}}^{\circ} m; & \Pi &= i RT m; \\ \Delta T_f &= i K_f m; & \Delta T_b &= i K_b m\end{aligned}\quad (16)$$

where in i for nonelectrolyte should be 1 and for strong electrolytes equal the number of ions formed on dissociation. For example, i should be 2 for NaCl, 3 for CaCl_2 , and 4 for FeCl_3 . However, in reality, i is less than that calculated, based on the number of ions produced in concentrated solutions, but will approach the theoretical number in infinitely dilute ideal solutions. When the i value is calculated, for a number of solutions with increasing concentration of the solute and then extrapolated to zero concentration of the solute, one can obtain the theoretical i value. The van't Hoff factor has also been considered the ratio of any colligative property of a real solution to that of an ideal solution of a nonelectrolyte.

PHYSIOLOGICAL AND CLINICAL SIGNIFICANCE OF TONICITY

Osmotic pressure becomes important from a physiological standpoint because a majority of biological membranes are semipermeable, and body fluids such as blood and tears exhibit significant osmotic pressure owing to a

number of solutes dissolved in them. As noted above in the Introduction, if a small quantity of blood is mixed with a solution containing 0.9% w/v NaCl, the red blood cells remain intact and retain their normal size and shape. The NaCl solution is considered to be isotonic because it maintained the tone of the membrane of the red blood cell. In contrast, if the blood is mixed with the hypertonic 1.8% w/v NaCl solution, cells shrink and become wrinkled or crenated owing to its content being sucked out. It is because the red blood cell content exerts a lower osmotic pressure than does the surrounding hypertonic 1.8% w/v salt solution, and the water inside the cells diffuses through the cell membrane to dilute the surrounding salt solution to equalize the osmotic pressure across the membrane. The exact opposite phenomenon occurs if blood is mixed with hypotonic 0.45% w/v NaCl solution. Water from the surrounding salt solution enters the cells, causing them to swell and finally burst with the liberation of hemoglobin, and the phenomenon is known as hemolysis. The crenation and hemolysis of red blood cells in hypertonic and hypotonic salt solution, respectively, are explained by the movement of water across the cell membrane owing to the osmotic pressure differential. However, it is well-known that the different physiological membranes are different with respect to their permeability characteristics. The red blood cell membrane has been found to be permeable to small polar and semipolar solutes such as alcohol, boric acid, and urea, etc. Thus, the erythrocyte membrane is not truly semipermeable, and although 2% boric acid solution is iso-osmotic with erythrocyte cell contents, it causes hemolysis because boric acid moves freely across the membrane and its solution being hypotonic acts like pure water in its effect on erythrocytes. However, the same 2% boric acid solution is both iso-osmotic and isotonic with eye secretions and causes no irritation when instilled in the eye because the mucous membrane of the eye is a true semipermeable membrane. To resolve the confusion created by different permeability characteristics of biological membranes, the word isotonicity was created. The word isotonic refers to solutions that are iso-osmotic with the cell contents, across a specific membrane, and in addition, maintains the tone of the membrane, i.e., no solvent movement across the membrane. Thus, 2% boric acid solution is iso-osmotic with blood but it behaves like a hypotonic solution with erythrocytes while it is both iso-osmotic and isotonic with respect to the mucous membrane of the eye.

It has been also observed that hypertonic and hypotonic salt solutions tend to irritate sensitive tissue and cause pain when applied to mucous membranes of the eye, ear, and nose, etc., whereas isotonic solution causes no tissue

irritation when it comes in contact with the tissue. Obviously, the tonicity of formulations that come in to direct contact with blood, muscle, eye, nose, and delicate tissues is critical. Therefore, the issue of tonicity is important in small- and large-volume injectables, ophthalmic products, and products intended for tissue irrigation. The degree of tissue irritation or hemolysis or crenation observed depends on the degree of deviation from isotonicity, the volume injected, the speed of injection, the concentration of the solutes in the injection, and the nature of the membrane. The parenteral and ophthalmic formulations are therefore adjusted to isotonicity if possible. Hypotonic solutions can be easily adjusted to isotonicity by adding solutes such as dextrose or sodium chloride, commonly used for this purpose. However, at times, the formulation may be hypertonic and may have to be diluted with water to maintain isotonicity. This dilution of the hypertonic solution may be precluded owing to other limitations such as poor aqueous solubility of the drug. In such a case, the hypertonic solution can be administered slowly in small volumes into a large vein such as the subclavian vein in which the formulation will be diluted and distributed rapidly, minimizing chances of crenation of erythrocytes, pain, and tissue irritation on injection. It has also been observed that minor deviations such as 10% from isotonicity may result in no effect or only temporary effects at the site of injection. However, the effects of deviation from isotonicity of large-volume parenterals can be fairly severe and, thus, parenteral nutrient solutions and infusions of large volume need to be adjusted to isotonicity. Large-volume infusion of hypotonic solutions has been observed to cause effects ranging from hemolysis to water-retention problems such as convulsions and pulmonary edema. This was exemplified by the recent report of 10 episodes of hemolysis, with two patients exhibiting significant hemolysis and renal insufficiency resulting in one death. These severe episodes of hemolysis occurred because of the large volume infusion of 25% human albumin diluted with sterile water instead of with isotonic sodium chloride for therapeutic plasma exchange (1). In contrast, large-volume infusion of hypertonic solutions can result in severe conditions such as intracellular dehydration, osmotic diuresis, hyperglycemia, glycosuria, dehydration from loss of water, and coma. Also, hypertonic solution infusion should be terminated gradually to avoid sudden changes in osmotic pressure. In summary, any formulation that comes in to contact with sensitive tissues of the human body needs to be adjusted to isotonicity to minimize any adverse effects. To be able to adjust the formulation to isotonicity, a method to measure the tonicity and/or the osmotic pressure of the formulations has to be used. The methods to measure tonicity or

osmotic pressure of solutions are reviewed briefly as below.

MEASUREMENT OF TONICITY OF SOLUTIONS

The most direct method for measurement of tonicity obviously would be to observe changes in erythrocytes on mixing solution with blood. If hemolysis or crenation or a marked change in the appearance of erythrocytes occurs, the solution is not isotonic. If the cells retain their normal size and shape, the solution is isotonic. Grosicki and Husa used this method early on; however, one has to be mindful of the fact that solutions may be iso-osmotic with erythrocyte contents, yet may cause hemolysis because solutes such as boric acid are permeable through erythrocyte membrane and, thus, solution is not isotonic (2). Therefore, Grosicki and Husa recommended that the word isotonic should be used with reference to solutions having equal osmotic pressures with respect to a particular membrane (2). Because hemolysis due to hypotonic solution results in release of oxyhemoglobin directly proportional to the number of cells hemolyzed, a quantitative method has been developed to calculate osmotic pressure and the van't Hoff *i* factor noted above. A limitation of observing changes in erythrocytes as a measure of tonicity is the fact that the specific chemical interaction of the solute with the cell, pH of the solution, presence of solvents, lipid solubility of the solute, and denaturant activity of solute may have influences on the cell membrane and, thus, osmotic pressure differences alone are not responsible for hemolysis. Furthermore, it was shown recently that hemolysis is related to the contact time in addition to hypotonicity of the formulation (3). To overcome this limitation, some investigators have used measurements of erythrocyte cell volumes as a function of tonicity of solution, which influence solvent (water) uptake or loss from erythrocytes. This method is more sensitive, objective, and reliable than observation of hemolysis. Recently, a method using fluorescence anisotropy for fluidity of erythrocyte membranes demonstrated differences between hypotonic and isotonic solutions (4). However, the method is involving, and more data need to be obtained to correlate tonicity with fluidity of the membrane to be reliable.

An alternative approach is based on the theoretical foundation described earlier for the colligative properties. If the solution is isotonic with blood, its osmotic pressure, vapor pressure, boiling-point elevation, and freezing-point depression should also be identical to those of blood. Thus, to measure isotonicity, one has to measure the osmotic

pressure of the solution and compare it with the known value for blood. However, the accurate measurement of osmotic pressure is difficult and cumbersome. If a solution is separated from blood by a true semipermeable membrane, the resulting pressure due to solvent flow (the head) is accurately measurable, but the solvent flow dilutes the solution, thus not allowing one to know the concentration of the dissolved solute. An alternative is to apply pressure to the solution side of the membrane to prevent osmotic solvent flow. In 1877, Pfeffer used this method to measure osmotic pressure of sugar solutions. With the advances in the technology, sensitive pressure transducers, and synthetic polymer membranes, this method can be improved. However, results of the search for a true semipermeable membrane are still elusive, and this method is still cumbersome and inconvenient. The measurement of osmotic pressure using this method has been applied successfully to colloidal solution of proteins to measure their molecular weight because they are of relatively large molecular weight and are impermeable across number of membranes. Numerous instruments, known as osmometers, are commercially available to measure osmotic pressure (5,6). Only the Knauer membrane osmometer and colloid osmometer are true osmometers using a semipermeable membrane (5). Vapor pressure osmometers such as the Wescor osmometer, using the principle of vapor pressure lowering should not be called osmometers. These types of instruments measure the vapor pressure using the isopiestic method, or the thermoelectric method, or the measurement of the dew point of unknown solution in comparison with a standard, and then calculate osmotic pressure and osmolality using the theory of colligative properties. These instruments require a few microliters, and the method is fairly precise, simple, and totally automated. The presence of a volatile solvent such as ethanol will create problem with this method because the inherent assumption is that only water is present in the vapor phase. This can be a serious limitation because many parenteral and, ophthalmic formulations contain organic solvents for the purposes of drug solubility and, sometimes of stability. Commercial osmometers of this type have been found to measure osmolalities in the range of 100–3000 mmol/kg reliably.

Boiling-point elevation can also be used to measure osmotic pressure and tonicity of a solution using just a reflux condenser and a thermometer. The commercially available instrument is the Cottrell boiling-point apparatus. However, this method is affected by the ambient barometric pressure and the presence of volatile solvents in the solution.

Osmometers based on the freezing-point depression are the most commonly and widely used instruments for measurements of tonicity because of the simplicity,

reliability, and ease of use. Freezing-point depression of solutions of a number of drugs at various concentrations has already been determined, and thus, an extensive database is available for adjustment of tonicity of solutions of these drugs, as addressed below. The freezing-point depression of a solution can be simply measured using a salt–ice bath, Dewars flask, and Beckmann's thermometer. Numerous commercial instruments requiring small quantities of solution such as Osmette from Precision Systems that use the principle of Beckmann's freezing-point method are now available. One of the problems with this method is the disengagement of ice and the need for determination of the actual equilibrium freezing point. The latter limitation can be overcome by use of the equilibrium method, in which solid solvent (ice) is placed in contact with solution (aq.) and the freezing point measured and compared with that of the pure solvent (water) in contact with the solid solvent (ice). Also, one has to consider the presence of other solvents in influencing the freezing-point depression. The freezing-point depression method is precise to the extent that the differences in freezing points of two systems within $\pm 0.0002^{\circ}\text{C}$ can be measured. Freezing-point osmometry can be used for all samples with osmolalities less than 550 mmol/kg, including those that contain volatile solutes (6). Once the freezing point of the solution is known, inert solute such as sodium chloride is added to match the freezing point of solution with that of blood and lacrimal fluids. After considerable debate and experimentation, it is now well-established that -0.52°C is the freezing point of blood and lacrimal fluids, following the work of Lund et al. (7). This is also the freezing point of 0.9% sodium chloride solution, which is therefore considered to be isotonic with both blood and lacrimal secretions. Therefore, to determine the tonicity of a solution, one has to measure its freezing-point depression and compare it with that of blood (-0.52°C). The freezing-point depressions of a number of drugs are listed in Table 1. A more comprehensive list of the freezing-point depressions of various concentrations of drugs can be found in The Merck Index, Remington's Pharmaceutical Sciences, and other literature sources listed in the Bibliography. An important consideration when using this method is that although all the solutes present in solution contribute to its freezing-point depression, those that permeate the biological membrane will not maintain the tone, for example, boric acid. In addition, association of solute molecules by processes such as complexation and micellar association, which are temperature-dependent, may have to be considered. The viscosity and presence of suspended particles can also affect the freezing point by altering the crystallization of the solvent. Nevertheless, freezing-point

Table 1 Freezing point depressions ($T_f^{1\%}$), L_{iso} values and sodium chloride equivalents (E) for drugs and excipients for adjusting their solutions to isotonicity^a

Substance	MW^b	$T_f^{1\%c}$	L_{iso}^d	E^e	V^f
Alcohol, dehydrated	46.07	0.41	1.9	0.70	23.3
Aminophylline	456.46	0.10	4.6	0.17	5.7
Amphetamine sulfate	368.49	0.13	4.8	0.22	7.3
Antipyrine	188.22	0.10	1.9	0.17	5.7
Antazoline (Antistine) hydrochloride	301.81	0.11	3.2	0.18	6.0
Apomorphine hydrochloride	312.79	0.08	2.6	0.14	4.7
Ascorbic acid	176.12	0.11	1.9	0.18	6.0
Atropine sulfate	694.82	0.07	5.3	0.13	4.3
Aureomycin hydrochloride	544	0.06	3.5	0.11	3.7
Barbital sodium	206.18	0.29	3.5	0.29	10.0
Benadryl hydrochloride (diphenhydramine hydrochloride)	291.81	0.34	3.4	0.20	6.6
Boric Acid	61.84	0.29	1.8	0.50	16.7
Butacaine sulfate	710.95	0.12	8.4	0.20	6.7
Caffeine	194.19	0.05	0.9	0.08	2.7
Calcium gluconate	448.39	0.09	4.2	0.16	5.3
Calcium lactate	308.30	0.14	4.2	0.23	7.7
Camphor	152.23	0.12	1.8	0.20	6.7
Chloramphenicol (chloromycetin)	323.14	0.06	1.9	0.10	3.3
Chlorobutanol (chloreton)	177.47	0.14	2.5	0.24	8.0
Cocaine hydrochloride	339.81	0.09	3.2	0.16	5.3
Dextrose-H ₂ O	198.17	0.09	1.9	0.16	5.3
Dibucaine hydrochloride	379.92	0.08	2.9	0.13	4.3
Ephedrine hydrochloride	201.69	0.18	3.6	0.30	10.0
Ephedrine sulfate	428.54	0.14	5.8	0.23	7.7
Epinephrine bitartrate	333.29	0.11	3.5	0.18	6.0
Epinephrine hydrochloride	219.66	0.17	3.7	0.29	9.7
Fluorescein sodium	376	0.18	6.9	0.31	10.3
Glycerin	92.09	0.20	1.8	0.34	11.3
Homatropine hydrobromide	356.26	0.10	3.6	0.17	5.7
Lactose	360.31	0.04	1.7	0.07	2.3
Magnesium sulfate · 7H ₂ O	246.50	0.10	2.5	0.17	5.7
Menthol	156.26	0.12	1.8	0.20	6.7
Meperidine hydrochloride	283.79	0.12	3.7	0.22	7.3
Methamphetamine hydrochloride	185.69	0.22	4.0	0.37	12.3
Morphine hydrochloride	375.84	0.09	3.3	0.15	5.0
Morphine sulfate	758.82	0.08	6.2	0.14	4.8
Naphazoline hydrochloride	246.73	0.16	3.3	0.27	7.7
Neomycin sulfate	—	0.06	—	0.11	3.7
Neostigmine bromide	303.20	0.11	3.2	0.22	6.0
Nicotinamide	122.13	0.15	1.9	0.26	8.7
Penicillin G potassium	372.47	0.11	3.9	0.18	6.0
Penicillin G Procaine	588.71	0.06	3.5	0.10	3.3
Penicillin G sodium	356.38	0.11	3.8	0.18	6.0
Phenacaine hydrochloride	352.85	0.11	3.3	0.20	5.3
Phenobarbital sodium	254.22	0.14	3.6	0.24	8.0
Phenol	94.11	0.20	1.9	0.35	11.7
Phenylephrine hydrochloride	203.67	0.18	3.5	0.32	9.7
Physostigmine salicylate	413.46	0.09	3.9	0.16	5.3
Physostigmine sulfate	648.45	0.08	5.0	0.13	4.3
Pilocarpine nitrate	271.27	0.14	3.7	0.23	7.7

(Continued)

Table 1 Freezing point depressions ($T_f^{1\%}$), L_{iso} values and sodium chloride equivalents (E) for drugs and excipients for adjusting their solutions to isotonicity^a (Continued)

Substance	MW^b	$T_f^{1\%c}$	L_{iso}^d	E^e	V^f
Potassium acid phosphate (KH_2PO_4)	136.13	0.25	3.4	0.43	14.2
Potassium chloride	74.55	0.45	3.3	0.76	25.3
Potassium iodide	166.02	0.20	3.3	0.34	11.3
Procaine hydrochloride	272.77	0.12	3.4	0.21	7.0
Quinine hydrochloride	396.91	0.08	3.3	0.14	4.7
Scopolamine hydrobromide	438.32	0.07	3.1	0.12	4.0
Silver nitrate	169.89	0.19	3.3	0.33	11.0
Sodium acid phosphate ($NaH_2PO_4 \cdot H_2O$)	138.00	0.24	3.2	0.40	13.3
Sodium benzoate	144.11	0.24	3.4	0.40	13.3
Sodium bicarbonate	84.00	0.38	3.2	0.65	21.7
Sodium bisulfite	104.07	0.36	3.7	0.61	20.3
Sodium borate $\cdot 10H_2O$	381.43	0.25	9.4	0.42	14.0
Sodium chloride	58.45	0.58	3.4	1.00	33.3
Sodium iodide	149.92	0.23	3.4	0.39	13.0
Sodium nitrate	85.01	0.39	3.4	0.68	22.7
Sodium phosphate, anhydrous	141.98	0.31	4.4	0.53	17.7
Sodium phosphate $\cdot 2H_2O$	178.05	0.25	4.4	0.42	14.0
Sodium phosphate $\cdot 7H_2O$	268.08	0.17	4.6	0.29	9.7
Sodium phosphate $\cdot 12H_2O$	358.21	0.13	4.6	0.22	7.3
Sodium propionate	96.07	0.36	3.4	0.61	20.3
Sodium sulfite, exsiccated	126.06	0.38	4.8	0.65	21.7
Streptomycin sulfate	1457.44	0.04	6.0	0.07	2.3
Strong silver protein	—	0.0	—	0.08	2.7
Sucrose	342.30	0.05	1.6	0.08	2.7
Sulfacetamide sodium	254.25	0.14	3.4	0.23	7.7
Sulfadiazine sodium	272.27	0.14	3.8	0.24	8.0
Sulfamerazine sodium	286.29	0.14	3.9	0.23	7.7
Sulfanilamide	172.21	0.13	2.2	0.22	7.3
Sulfathiazole sodium	304.33	0.13	3.9	0.22	7.3
Tetracaine hydrochloride	300.82	0.11	3.2	0.18	6.0
Tetracycline hydrochloride	480.92	0.08	4.0	0.14	4.7
Tripelennamine hydrochloride	291.83	0.17	3.8	0.30	7.3
Urea	60.06	0.35	2.1	0.59	19.7
Zinc chloride	139.29	0.37	5.1	0.62	20.3
Zinc Sulfate $\cdot 7H_2O$	287.56	0.09	2.5	0.15	5.0

^aValues vary somewhat with concentration, and those in the table are for 1 to 3% solutions of the drugs in most instances.

A more comprehensive table of E and T_f values is found in Ref. 8.

^b MW is the molecular weight of the drug.

^c $T_f^{1\%}$ is the freezing point depression of a 1% solution of the drug.

^d L_{iso} is the molar freezing point depression of the drug at a concentration approximately isotonic with blood and lacrimal fluid.

^e E is the sodium chloride equivalent of the drug.

^f V is the volume in milliliters of isotonic solution that can be prepared by adding water to 0.3 g of the drug (the weight of drug in one fluid ounce of a 1% solution).

(Adapted with modifications from Ref. 9.)

depression has become popular because of its simplicity, reliability, and availability of commercial instruments. The methods of osmometry, the technology, and the limitations inherent in each method have been reviewed recently (5, 6) and should be consulted for more details.

Based on the theory of colligative properties and the principles of osmometry, it is understood that osmometer will read osmolalities and not osmolarities because colligative properties are directly proportional to the total solute concentration expressed in molality (see Eqs. 1–16).

The relationship between osmolality and osmolarity and its significance can be found in the Remington's Pharmaceutical Sciences and in a review article by Deardorff (10). However, it is more convenient to use osmolarity because it is based on weight/volume rather than on weight/weight as in osmolality. The U.S. Pharmacopeia also recommends that the labeling of parenteral and ophthalmic formulation should list osmolarity while the experimentally determined quantity is osmolality. Methods to convert osmolality to osmolarity using determinations of solution density and solute content (11, 12) or using partial molal volume of solute and solvent have been described (13, 14).

THEORETICAL METHOD TO CALCULATE TONICITY USING L_{iso} VALUE

In the discussion above of colligative properties of electrolytes, the equations were modified by introduction of the van't Hoff i factor as shown in Eq. 16. Because the freezing-point depressions of strong and weak electrolytes are always greater than those calculated from Eq. 9, because of different degrees of ionization and interionic interaction, a new factor, $L = iK_f$, is introduced. The L value obtained from freezing-point depression of a solution of a particular type of electrolyte at a molar concentration (c) that is isotonic with blood is defined as L_{iso} ($L_{\text{iso}} = 0.52^\circ/c$). For example, the L_{iso} value for an isotonic sodium chloride solution (0.9%w/v) is 3.4, and its freezing-point depression is 0.52°C . Because the colligative properties are independent of the chemical nature of the electrolyte and the interionic interactions in dilute solutions are similar, all electrolytes of the same type will have identical L_{iso} values. Therefore, all nonelectrolytes have $L_{\text{iso}} = 1.9$, whereas uni-univalent electrolyte's $L_{\text{iso}} = 3.4$ and triunivalent electrolyte's $L_{\text{iso}} = 6.0$. Thus, if the ionic nature of the solute and its molecular weight are known, using the appropriate L_{iso} value, freezing-point depression can be calculated for a solution of a given concentration. The average L_{iso} values for all types of solutes are available in the literature (9). This method is simple and does not require experimentation, but it is only approximate, with potential for some error. Also, one has to know the ionic nature of the solute, which can be difficult to determine for a new drug compound with a complex structure.

METHODS OF ADJUSTING TONICITY

From the theoretical background presented above, one can easily devise his or her own methods to adjust tonicity of

solutions using the principles of colligative properties. However, in the practice of pharmacy, a number of simple methods to adjust tonicity of formulation in a prescription order on an extemporaneous basis were developed to help the pharmacist. The methods of adjusting tonicity could be classified into two types. In class I methods, some inert substance such as sodium chloride or dextrose is added to the solution to lower its freezing point to match that of blood (-0.52°C) i.e., made isotonic by the addition of inert excipient. In class II methods, a calculated quantity of water is added to the total solute content (drug) of the prescription to make it isotonic, which is then diluted with sufficient isotonic diluting solution to bring it to the final volume. These methods are explained below, followed by a simple example illustrating the method. However, the assumptions inherent in all the methods need to be considered carefully. The first assumption is that colligative properties are additive for mixture of solutes and that they are related linearly to their concentration expressed in molarity, molality, or in percentages. This assumption is true in dilute solutions of nonelectrolytes and electrolytes. However, when dealing with concentrated solutions, this assumption may not be valid. In cases of chemical interaction, association, complexation, or micellar interaction among solutes in solution, the colligative properties of solutes may not be additive. The second assumption is that they consider all solutes present in solution to be contributing to its tonicity. However, it is known from the discussion above that all biological membranes are not truly semipermeable, and thus, some solutes will not contribute to the tonicity of the solution across that membrane, for example, boric acid across erythrocytes membrane. Nevertheless, the errors introduced are small, and slight deviations from isotonicity on either side do not result in significant adverse effects. In the literature, one may also find methods known as the L value method or the L_{iso} method, which are identical to the theoretical method described above for solutes, for which one can calculate the freezing-point depression based on molecular weight and ionic nature. By knowing the freezing-point depression, any of the class I or class II methods can then be used to adjust their solution to isotonicity.

Class I Methods

Freezing-point depression method (cryoscopic method)

The freezing-point depression of a number of drugs and excipients, either experimentally determined by the method described above or calculated theoretically using the L_{iso} method, is available in the literature (8, 9, 15). Table 1 (adapted from Ref. 9.) lists the freezing point

depression of 1% solution of various drugs. Basically, from the percentage of drug present in solution, the freezing point of the solution is calculated. This number is then subtracted from the freezing point of blood (-0.52°C) to obtain the freezing-point depression to be achieved by the addition of sodium chloride. Knowing that 0.9% sodium chloride is isotonic and freezes at -0.52°C , the amount of sodium chloride to be added is calculated as shown in the example below.

Example 1: Calculate the amount of sodium chloride needed to prepare 100 ml of 2% isotonic physostigmine salicylate solution.

Freezing-point depression of 2% physostigmine salicylate = $2 \times 0.09^{\circ}\text{C} = 0.18^{\circ}\text{C}$ (Table 1). Therefore, the freezing-point depression to be achieved by adding sodium chloride = $0.52^{\circ} - 0.18^{\circ}\text{C} = 0.34^{\circ}\text{C}$. Sodium chloride (0.9%) produces a freezing-point depression of -0.52°C ; therefore, the percentage of sodium chloride needed = $(0.34^{\circ}\text{C}/0.52^{\circ}\text{C}) \times 0.9\% = 0.59\% = 0.59 \text{ g}/100 \text{ ml}$.

Sodium chloride equivalent (E) method

The sodium chloride equivalent (E) is the amount of sodium chloride equivalent to 1 g of the drug in exerting the same osmotic effect. The E value for a new drug can be calculated from its L_{iso} value or from the freezing-point depression as shown below.

The freezing-point depression of a 1 g/L-solution of a new drug can be expressed as:

$$\Delta T_f = L_{\text{iso}} \frac{1 \text{ g}}{MW} \quad (17)$$

By definition, E gram of sodium chloride ($MW = 58.45$ and $L_{\text{iso}} = 3.4$) in 1 L will have similar freezing-point depression as shown below:

$$\Delta T_f = 3.4 \frac{E \text{ g}}{58.45} \quad (18)$$

Therefore, equating Eqs. 17 and 18, results in the following equation for E :

$$E = 17 \frac{L_{\text{iso}}}{MW} \quad (19)$$

Wells developed a nomogram based on the above equation to readily calculate E values from the MW and L_{iso} value of the drug (16). Thus, the E value for physostigmine salicylate ($MW = 413.46$) calculated using $L_{\text{iso}} = 3.4$ for a uni-univalent electrolyte is equal to 0.14, which is close to 0.16 (E value) in Table 1. This small deviation is attributable to the difference between the experimentally determined L_{iso} (3.9) of physostigmine salicylate and the theoretical value of 3.4 for a uni-univalent

electrolyte. By knowing the E value, the solution can be adjusted to isotonicity as shown below.

Example 2: Calculate the amount of sodium chloride needed to prepare 100 ml of 2% isotonic physostigmine salicylate solution.

Physostigmine salicylate (2 g/100 ml) is equivalent to $2 \times 0.16 (E) = 0.32 \text{ g}/100 \text{ ml}$ of sodium chloride. Therefore, 0.58 g (0.9–0.32 g) of sodium chloride has to be added to 100 ml of this solution to make it isotonic.

Note that the answers given by the two methods are not identical, but very close.

Class II Methods

The class II methods involve the calculation of a quantity of water needed to make an isotonic solution for a given amount of drug, followed by dilution with an isotonic solution to make up the volume. These methods were developed to enable pharmacists to prepare parenteral and ophthalmic formulations with simplicity and ease.

The White-Vincent method

In this method, the weight of the drug (w) is first multiplied by its sodium chloride (E) to obtain the quantity of sodium chloride osmotically equivalent to weight per gram of drug (17). Because 0.9 g of sodium chloride dissolved in 100 ml results in an isotonic solution, the volume of isotonic solution that can be prepared from weight per gram of drug is given by the following equation:

$$V = wE \frac{100}{0.9} = 111.1 wE \quad (20)$$

Thus, dissolving weight per gram of drug in V ml of water will result in an isotonic solution that can be further diluted with isotonic solutions such as 0.9% sodium chloride or isotonic dextrose solution to make up the volume. The method can be illustrated by the following example.

Example 3: Prepare 100 ml of 2% physostigmine salicylate solution isotonic with blood.

Using Eq. 20 and E of physostigmine salicylate = 0.16 from Table 1, the volume of water needed to prepare isotonic solution, $V = 2 \text{ g} \times 0.16 \times 111.1 \text{ ml/g} = 35.55 \text{ ml}$. This solution can be diluted with 64.45 ml of any isotonic diluting solution to obtain 100 ml of 2% isotonic physostigmine salicylate solution. To verify the results, if we assume that we dilute the above solution with 64.45 ml of isotonic sodium chloride solution, the equivalent amount of sodium chloride added is 0.58 g, which matches with results obtained using the class I methods.

The Sprowls method

In the early days of pharmacy practice, many prescriptions were written to prepare one fluid ounce of a 1% drug solution,

thus, the amount of drug ($w = 0.3$ g) and the final volume were fixed (one fluid ounce or 30 ml). Sprowls, recognizing this fact, suggested a modification of the White–Vincent method to further simplify the calculations for the practicing pharmacist (18). In this method, the amount of drug is fixed at 0.3 g (30 ml of 1% solution), and the volume of water required to prepare the isotonic solution is calculated using Eq. 20 and listed in a table such as column 4 in Table 1 for all drugs that are commonly used in parenteral and ophthalmic formulations and for which sodium chloride equivalents are known. The pharmacist then makes up the volume of the preparation to 30 ml with an isotonic diluting solution to fill the prescription. For example, if one fluid ounce of 1% physostigmine salicylate solution is to be prepared, from Table 1, column 4, we recognize that 5.3 ml of water is required for 0.3 g of physostigmine salicylate to prepare an isotonic solution. After the preparation of this 5.3-ml solution, it can be diluted with any isotonic diluting solution to make up the volume to one fluid ounce. If one needed to prepare 100 ml of a 1% solution, the volume of water (V) should be multiplied by 3.33 to obtain the amount of water necessary to make it isotonic.

FUTURE DIRECTIONS

The theory of colligative properties is well-understood and successfully applied to parenteral formulations for making them isotonic and, thus, safe and acceptable. The techniques of osmometry have been refined, and now instruments that can estimate freezing-point depression, vapor pressure, or osmotic pressure from microliter quantities of samples in a few minutes are commercially available. At the same time, very few pharmacists are required to compound prescriptions requiring the knowledge of the various methods of adjustments of tonicity. Because of ever-increasing complexities in the structure of new drug entities, there is an increasing problem of their inadequate aqueous solubility, exemplified by drugs such as Cyclosporine and Taxol. A number of organic solvents and new classes of surfactants are being developed and used to aid in solubilization and, thus, in the formulation of these drugs for parenteral administration. The issue of tonicity needs to be addressed from this perspective because organic solvents and the surfactants behave differently in solution than do the traditional solutes whose characteristics in solution are well-understood. Also, dispersed systems such as nanocapsules, liposomes, and microemulsions are being developed as parenteral formulations. The colligative properties of these systems need to be investigated too. There is an increasing concern regarding tissue irritation and muscle injury at the

site of injection resulting from formulations. With the advent of biotechnology, more peptide and protein drugs are in clinical trials than before, and, also, gene therapy is being considered for few diseases. The parenteral formulations of these newer drugs are more complex to maintain the integrity of their higher-order structure. Therefore, the issue of tonicity needs to be revisited with a newer approach and from a different perspective.

REFERENCES

1. Morbidity & Mortality Weekly Report **1999**, 48 (8), 157–159.
2. Grosicki, T.S.; Husa, W.J. J. Am. Pharm. Assoc., Sci. Ed. **1954**, 43, 632–636.
3. Krzyzaniak, J.E.; Raymond, D.M.; Yalkowsky, S.H. PDA J. Pharm. Sci. Tech. **1996**, 50 (4), 223–226.
4. Au, K.S. Biochem. Mol. Bio. Int. **1994**, 32, 49–53.
5. Bevan, D.R. Anesthesia **1978**, 33, 794–800.
6. Sweeney, T.E.; Beuchat, C.A. Am. J. Physio. **1993**, 264, R469–480.
7. Lund, C.G.; Nielsen, P.; Pedersen-Bjergaard, K. *The Preparation of Solutions Iso-Osmotic with Blood, Tears, and Tissue*; Danish Pharmacopeia Commission: Einar Munksgaard Copenhagen, 1947; 2.
8. Budavari, S.; O'Neil, M.J.; Smith, A.; Heckelman, P.E.; Kinneary, J.F., Eds. *The Merck Index: An Encyclopedia of Chemicals, Drugs and Biologicals*, 12th Ed.; Merck & Co. Inc.: Whitehouse Station NJ, 1996; 47–57, MISC.
9. Martin, A.; Bustamante, P.; Chun, A.H.C. Physical Pharmacy: Physical Chemical Principles. *The Pharmaceutical Sciences*; Lea & Febiger: Philadelphia, 1993; 101, 142, 169–189.
10. Deardorff, D.L. Am. J. Hosp. Pharm. **1980**, 37, 504–509.
11. Murty, S.R.; Kapoor, J.N.; DeLuca, P.P. Am. J. Hosp. Pharm. **1976**, 33, 546–551.
12. Gatlin, L.; Kulkarni, P.; Hussain, A.; DeLuca, P.P. Am. J. Hosp. Pharm. **1979**, 36, 1357–1361.
13. Streng, W.H.; Huber, H.E.; Carstensen, J.T. J. Pharm. Sci. **1978**, 67, 384–386.
14. Huber, H.E.; Streng, W.H.; Tan, H.G.H. J. Pharm. Sci. **1979**, 68, 1028–1032.
15. Siegel, F.P. Tonicity, Osmoticity, Osmolality and Osmolarity. *Remington's Pharmaceutical Sciences*; Gennaro, A.R., Chase, G.D., Marderosian, A.D., Harvey, S.C., Hussar, D.A., Medwick, T., Rippie, E.G., Schwartz, J.B., Zink, G.L., Eds.; Mack Printing Co.: Easton Pennsylvania, 1990; 1481–1498.
16. Wells, J.M. J. Am. Pharm. Assoc. Prac. Ed. **1944**, 5, 99–106.
17. White, A.I.; Vincent, H.C. J. Am. Pharm. Assoc. Prac. Ed. **1947**, 8, 406–411.
18. Sprowls, J.B. J. Am. Pharm. Assoc., Prac. Ed. **1949**, 10, 348–352.

BIBLIOGRAPHY

- Flynn, G.L. J. Parenteral Drug Assoc. **1979**, 33, 292–314.
 Hadzija, B.W. Am. J. Pharm. Ed. **1995**, 59, 191–195.



Tooling for Tableting

Annette Bauer-Brandl

The University of Tromsø, Tromsø, Norway

INTRODUCTION

Introducing high-speed tableting in 24-hr shifts has tremendously increased productivity, which in turn places heavy demands not only on the tablet presses and the compressibility of the tableting materials but also on precision with respect to dimensions and the wear resistance of the toolings (punch and die sets). Furthermore, in many cases of high tableting-speeds, adhesiveness of tablets to the toolings becomes a problem in the course of long-term production. Tableting materials need to be optimized in line with new demands. At the same time, tooling requirements with respect to adhesion-reducing finish have become more important than ever.

Special attention must be paid to handling and maintaining the tool sets during their entire lifetime. Not only does timely mechanical work-over prolong their total lifetime, but also trace patterns of wear on the tooling itself can be indicative of numerous different events during production, inter alia pressures that are generally too high, inadequate lubrication resulting in excessive friction, blockage of pressure rollers, tilted punches, or dies. Furthermore, worn-out punches not only affect the quality of the tablets (increased variation of mass, reduced surface shininess etc.), but can also cause serious problems in the process (powder loss with dust sticking to lubrication oil as well as excessive wear of the tablet press itself). Scratched surfaces can lead to grave problems with tablets sticking to the punches at the discharge point, making tablet production impossible. A deviation in the length of a lower punch translates directly into mass variability in the tablet batch; variation in the length of the entire set also leads to varying compression forces and consequently affects the dissolution properties of the tablets.

On the other hand, tablet tooling design opens up for many possibilities for overcoming common tableting problems that cannot be solved by modifying the materials or the compression process, such as lubrication problems which in general are also a function of tooling size and shape.

BACKGROUND

In the following text, only examples and principles will be discussed. Mentioning or omitting a specific make or

model does not constitute any value judgment. Comprehensive recommended publications are “The Tablet Specification Manual”^[1] with respect to technical and practical problems of tooling, and for all aspects of tableting including development of tableting materials, optimization of compression cycles, and handling of tooling “Die Tablette.”^[2]

GENERAL TERMINOLOGY OF TOOLING

A set of tooling consists of upper punch, lower punch, and die. The upper punch has a shorter stem; the lower punch stem is longer because it travels longer distances up and down in the die for filling, compression, and ejection, thereby sealing the die hole from below during the entire process of compression.

In single-punch tablet presses (reciprocating tablet presses), the punches are fixed to the punch holders which travel up and down, and in many cases fixation screws are used. Each make has its own holder design for punches and dies, and consequently for the tooling sets as well. Therefore little can be said in general terms, but it would appear that on the whole the remarks on the rotary press tooling apply correspondingly.

In rotary tablet presses, only the dies are fixed in the die table, whereas the punches are vertically mobile and are not fixed to a holder. They slide up and down in the turret bores, driven by cam tracks guiding the punch heads. The dwell time, i.e., the period during which the tablet is under full compression, is determined by the diameter of the flat part of the punch head.

The amount of clearance (space) between interacting parts like punch/die and barrel/turret guide, depends on the tolerance range (deviation from theoretical dimensions due to practical manufacturing reasons) of tooling dimensions. The clearance between the die wall and the punch tip affects the tableting process as well as the mechanical properties of the finished product: while the powder is compressed, air needs to be released, and when the compressed tablet is moved upwards within the die bore, friction on the die walls is affected by the clearance.

A deviation in the overall length or the working length of a lower punch in a set of punches for a rotary press translates directly into mass variability in the tablet batch;

a deviation of not more than ± 0.1 mm can lead to a mass variability of 2% in the case of small tablets. Moreover, variability of the lengths of both lower and upper punches affects the variability of compression forces with consequences for tablet dimensions, their tensile strength, and dissolution behavior. In the case of a short lower and upper punch, overfilling would not necessarily lead to an increased force signal.^[3]

TOOLING FOR SINGLE-PUNCH PRESSES AND COMPACTION SIMULATORS

Nowadays, single-punch presses are, in general, only used in R&D, but hardly in pharmaceutical production. Useful research tools are instrumented special hydraulic presses, “compaction simulators,” which allow for online measurements of forces and displacements during the tableting process. Compaction simulators are in many cases designed to accommodate standard single-punch tooling, which is specific for each machine type, e.g., Korsch EK0 (<http://www.korsch.de>), Kilian SP300 (<http://ourworld.compuserve.com/homepages/Kilian>). Dimensions of the tooling and tolerances thereof are up to the press and tooling manufacturer.

Two-piece configuration—i.e., a punch with interchangeable punch tips—is common for single-punch presses as, in general, single-punch tooling is less exposed to abrasion and—compared to tooling for rotary presses—can relatively be aligned easily. In the case of two-piece tools, the lower punch tips are not rigidly fixed but allow for some tilting in order to avoid excess friction in the die bore. Multiple tools comprise several punch tips on a single holder in order to increase productivity. They have been common for single-punch presses used in small-scale production, particularly in the form of multipiece tools, where any cracked tip can be exchanged separately.

TOOLING FOR ROTARY TABLET PRESSES

Fig. 1 shows the general terminology of punches and dies for rotary tablet presses. The punch consists of head, neck, barrel, and stem. The barrel diameter and the overall length are determined by the machine turret dimensions. The working length (i.e., overall length minus any cup depth at the punch tip) determines thickness and mass of the tablets and needs to be controlled regularly. The variation in working length should be within $\pm 10 \mu\text{m}$ to ensure low variability of tablet properties. The die depth should also be accurately cut in order to guarantee a flush fitting with the die table.

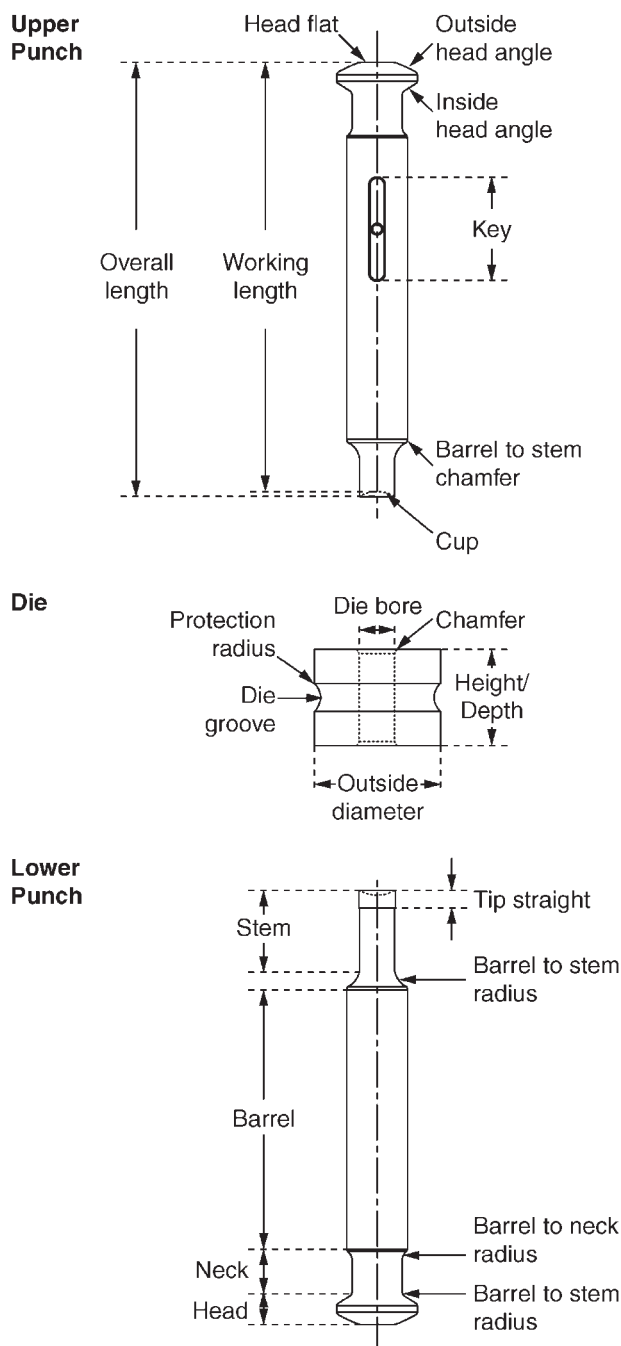


Fig. 1 General terminology of tooling (sets of upper and lower punches and dies) for rotary tablet presses.

International Standards

Standardization of tooling dimensions has been agreed to facilitate interchangeability between presses and to reduce the number of spare parts. In many cases, a certain press is available with different turrets in order to standardize tooling within a tablet manufacturing company irrespective of

Tooling for Tableting

3

the make of tablet presses used in R&D and production, respectively. A special turret for R&D on which all standard tooling can be used on one and the same turret is available from Pharmachine India (<http://www.serenader.com/pharmachine/develop.asp>). It allows to interchange punches and dies of any sizes for trial compression.

There are two main standards: 1) B-tools with subcategories EU19 (Europe)/TSM 19 (USA)/Japan Norm; 2) D-Tooling, also called EU1, TSM 1, which are classified according to the dimensions of the barrel diameter, overall length, and the overall diameter of the punch head. Fig. 2 compares dimensions of B- and D-tooling. D-tooling is thicker than B-tooling; European toolings (both B- and D-types) are longer compared to TSM types. Fig. 3 explains terminology of the punch head in order to differentiate subcategories of B- and D-toolings. The head angle or head radius (in the case of “domed heads”), respectively, determines the lead under the compression roller. The head flat determines the dwell time of the compression event (i.e., the period during which the tablet is under maximum pressure), which should preferably be long so as to yield mechanically resistant tablets. The head flat must in all cases be smaller in diameter than the neck in order to guarantee pressure transmission onto the whole punch. Neck shape and neck

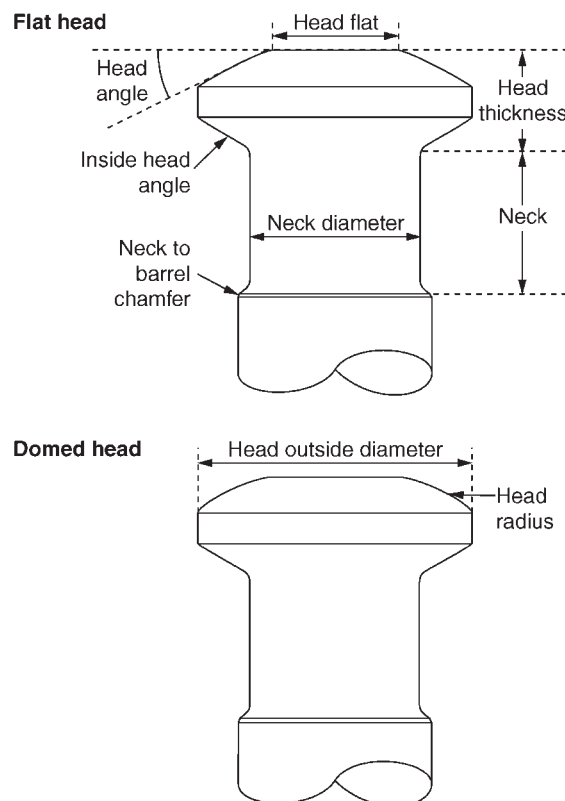


Fig. 3 Punch head terminology.

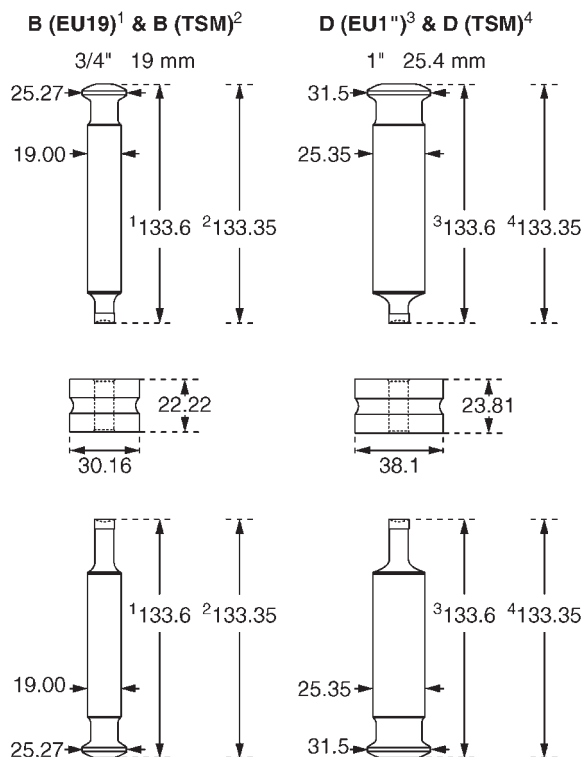


Fig. 2 Standard sizes of B- and D-tooling. Comparison of the respective subcategories according to the European standard and the U.S. standard (TSM).

diameter are determined by the dimensions of the cam track. The head thickness needs to be precisely cut in order to ensure a smooth gliding of the punches through the cam track and, in the case of lower punches, to guarantee a reproducible dosing of granule in the die as well. Three subcategories of B-tooling according to the European and the US (TSM) standard as well as the Japanese norm are compared in Fig. 4(a) and (b) in general terms, and their respective head dimensions are accurately depicted and closely compared in Fig. 5(a). The respective head dimensions for the three subcategories of D-tooling are shown in Fig. 5(b).

There is another head shape to the D-tooling called Fette EU1 441 (<http://www.fette.de>), which is characterized by the absence of a neck and a larger head diameter (38.00 mm) allowing for increased dwell times (see “Special Tooling”).

A further subcategory of B-tooling, which is referred to as B2-type (Stokes, <http://www.stokesdti.com>), has the same barrel diameter as the standard B-type, but the overall length of the lower punch is smaller (Fig. 6). B2 is used on some older presses, but not manufactured any more. There is yet another variation on standard B-type tooling, where the overall diameter of the die is smaller (Fig. 7). The die diameter determines the maximum tablet

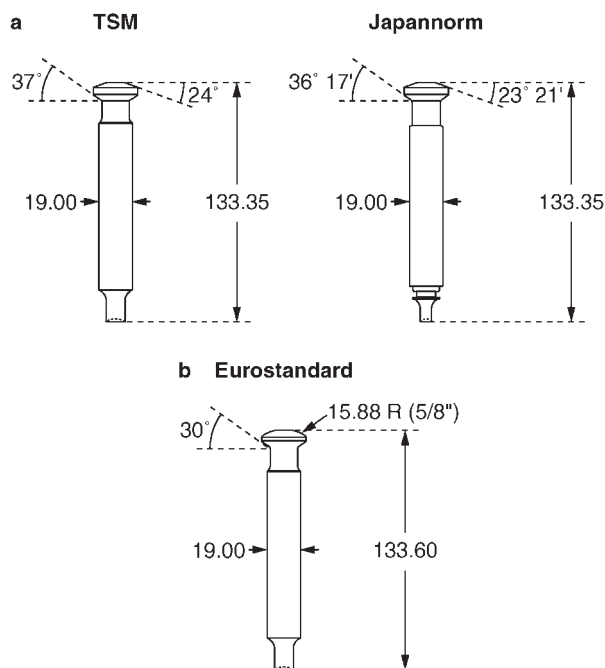


Fig. 4 Subcategories of B-tooling, defined by overall length and head shape: (a) U.S. (TSM) standard, and Japan norm; (b) European norm (domed head).

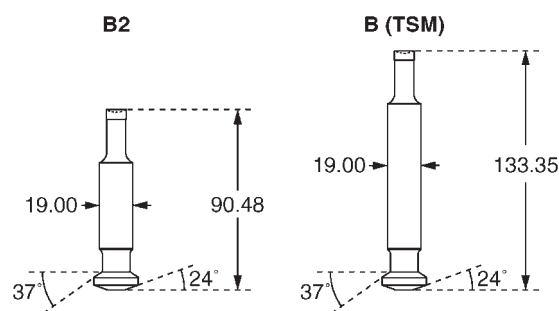
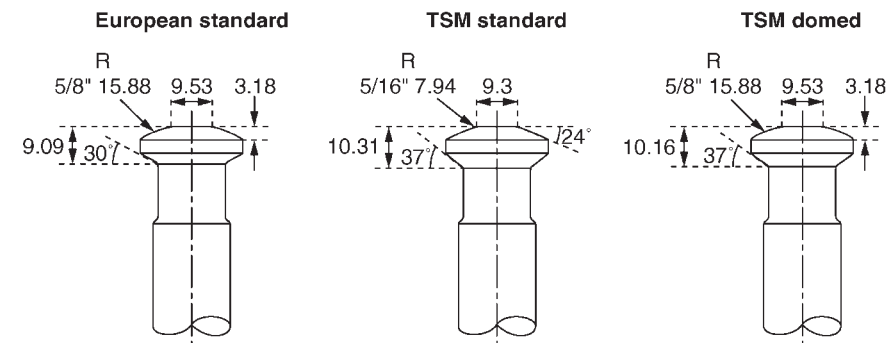


Fig. 6 Subcategories of B-tooling: B2 tooling compared to standard B.

diameter as well as the maximum number of die bores in the case of multiple tools. A larger die is more resistant although pressure is not directly applied to it but only by lateral extension. The terminology of standard dies is explained in Fig. 8(a). The height (depth) of the die ensures a flush fitting with the die table. The overall die diameter—nominal size and manufacturing tolerances—enables a tight fit in the turret avoiding lateral movement, which is also supported by the die groove accommodating the locking screws. A protection radius or protection

B-Head profiles



D-Head profiles

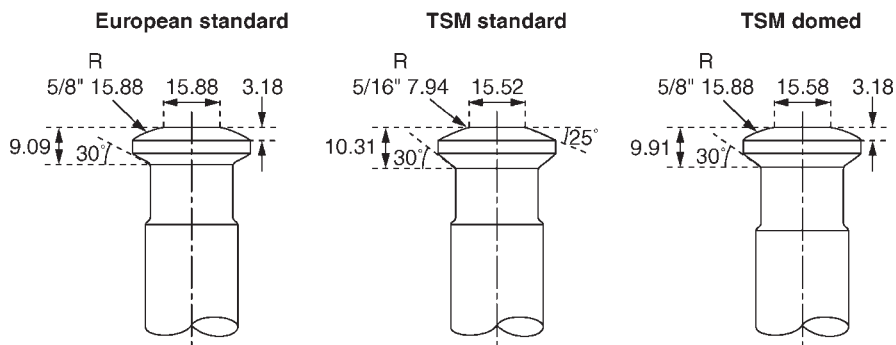


Fig. 5 Head profiles of standard tooling. (a) Heads of B-tooling with subcategories European, TSM, and Japan norm. (b) Heads of D-tooling with subcategories European, TSM, and Japan norm.

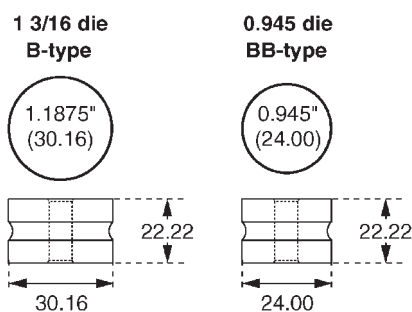


Fig. 7 Subcategories of B-tooling: BB tooling die diameter compared to standard B.

shoulder prevents damage to the die locking screw or scoring of the die pockets in the turret caused by burrs or sharp edges on the die. The die bore should be made as close to its nominal size as possible because it is this, which determines the tablet size, not the punches. A bore chamfer provides a lead for the upper punch to enter the die.

Any clearance between the die wall and the punch affects the tableting behavior, as entrapped air will escape through the gap when pressure is applied. Fig. 9 presents the terminology. Table 1 summarizes manufacturing tolerances of punches and dies as well as the clearance between them.

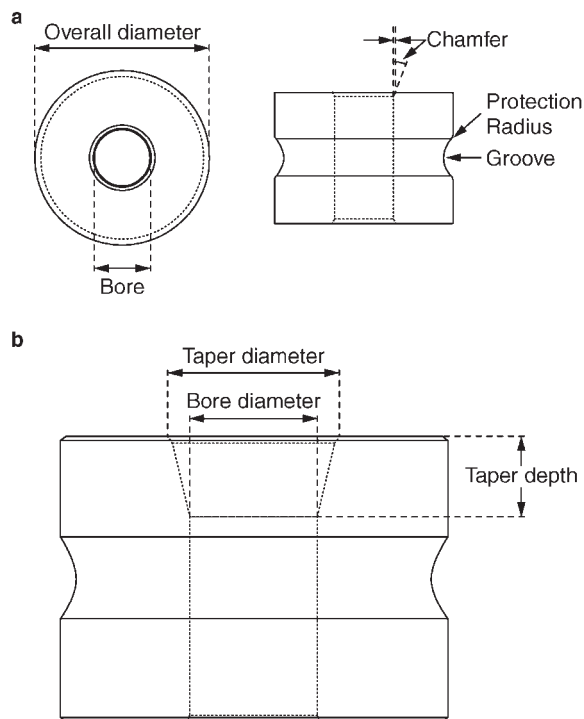


Fig. 8 Die terminology: (a) Standard die. (b) Tapered die.

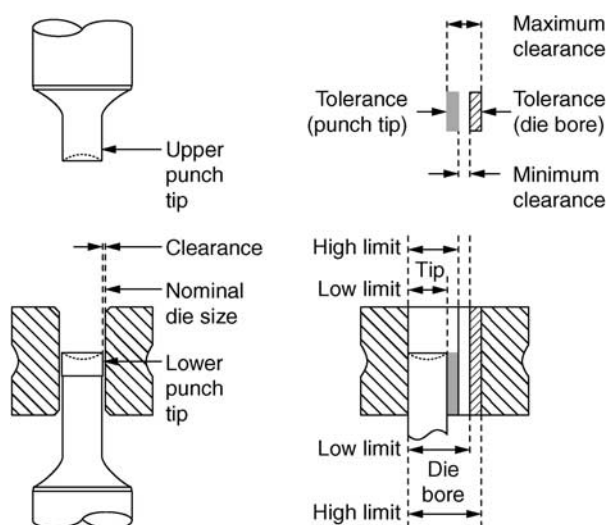


Fig. 9 Terminology of tolerances and clearances between punches and dies.

Die bore tapers with a conical opening as is depicted in Fig. 8(b) have the same effect as the clearance. Tapered dies are more expensive than standard dies. Tapers can be applied to both ends of the bore to allow for an inversion

Table 1 Tolerances and clearances in standardized tooling

<i>a) Tolerances for B- and D-type round tooling:</i>		
Upper punch	Overall length	$\pm 50 \mu\text{m}$
	Working length	$\pm 10 \mu\text{m}$
	Head diameter	$+ 0 \mu\text{m} - 100 \mu\text{m}$
Lower punch	Overall length	$\pm 50 \mu\text{m}$
	Working length	$\pm 10 \mu\text{m}$
	Head diameter	$+ 0 \mu\text{m} - 100 \mu\text{m}$
Die	Height	$\pm 10 \mu\text{m}$
	Diameter	$+ 0 \mu\text{m} - 10 \mu\text{m}$
	Concentricity	$25 \mu\text{m}$
<i>b) Tolerances and clearances for round and shaped tooling</i>		
	Round tooling (μm)	Shaped tooling (μm)
Nominal die bore size	$+ 10 - 0$	$+ 20 - 0$
Upper punch tip tolerance	$- 40$	$- 60$
Lower punch tip tolerance	$- 20$	$- 40$
Upper punch/die minimum clearance	40	40
Upper punch/die maximum clearance	60	80
Lower punch/die minimum clearance	20	20
Lower punch/die maximum clearance	40	60

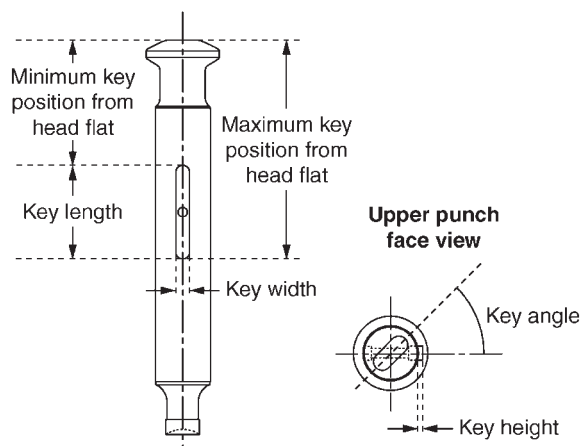


Fig. 10 Keys on punches: terminology.

of the dies. Tapered dies not only let air escape, but also assist ejection as the stress is gradually released by ejecting to an increasing diameter thereby preventing capping and generally increasing the tensile strength of the tablets. Furthermore, their use also allows higher machine speeds.

Shaped punches—all other shapes than round ones—and multtip punches need to have keys on their barrel in order to prevent any rotation. The terminology of keys is explained in Fig. 10. Keys are available in two configurations: the Woodroff key and the flat key (European key) can be distinguished by the shape of the slot milled into the barrel where the key is in many cases screwed into position. The vertical position of the key on the punch barrel and the length of the key itself are determined by the turret dimensions in such a way as to allow for clearance in the punch-guide keyways. The optimum angle of the key with respect to tablet shape is determined by the rotation direction of the tableting machine and the design of the take-off fence at the discharge point with respect to the punch tip, and therefore not standardized. Tablets hitting the fence with a corner should be avoided; they should rather hit it with a flat face in order to ensure proper discharge without fragmentation. Common configurations for punch keys and their lengths can be found in Tablet Specification Manual.^[1]

Barrel flutes, usually on round upper and lower punches, are vertical sharp-edged grooves along the length of the punch barrel. During production, they would rotate and thereby scrape off possible depositions of material in the punch guideways.

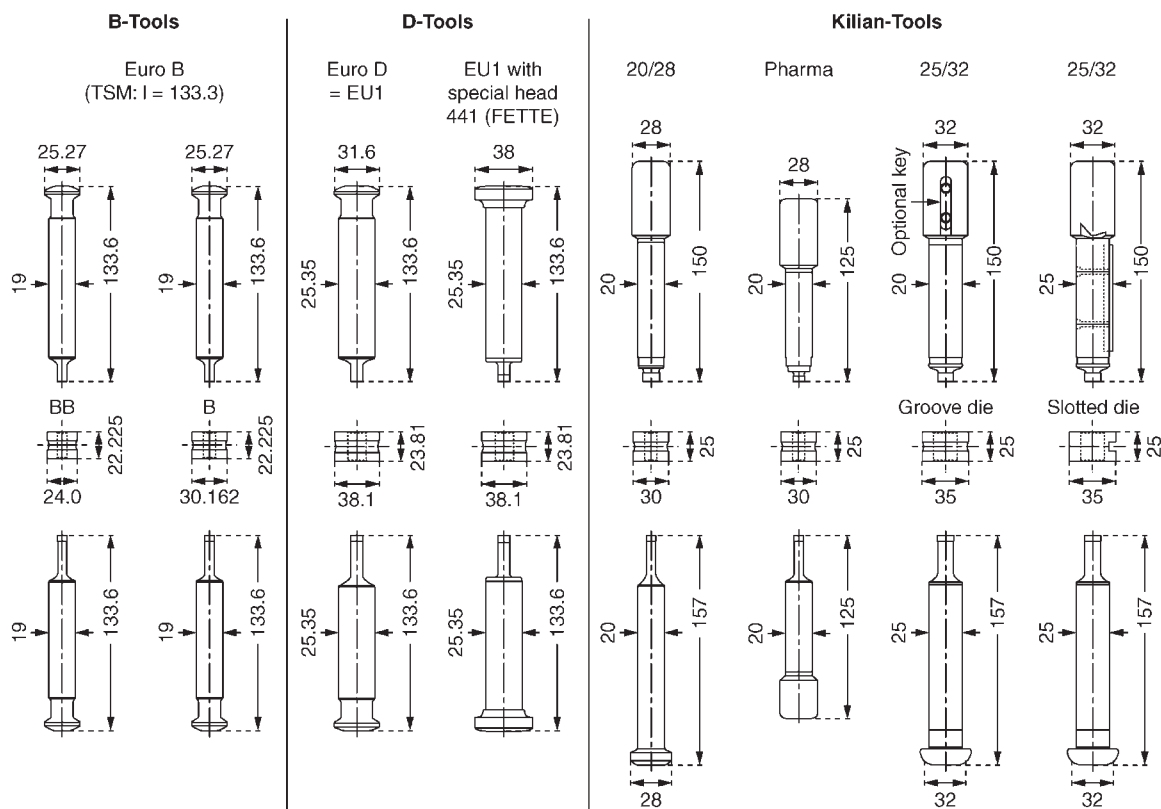


Fig. 11 Comparison of shapes and dimensions of well-known toolings.

Several types of dust caps and sealings fitted to the barrel-to-stem chamfer (e.g., Ref. 4) are available to prevent contamination of the tablets by machine oil dripping from the upper punch. Alternatively, punch bellows in silicone rubber fitted into special seal grooves for Euro-standard are available (e.g., from Holland, <http://www.iholland.co.uk>).

SPECIAL TOOLING

Some makes of tablet presses have their own tooling specifications allowing for longer dwell times and higher compression forces, such as the Fette EU1 441 mentioned earlier, and Kilian 25/32, with subtypes in die shape, i.e., grooved and slotted, which are depicted together with standard types in a tooling identification chart (Fig. 11). Most tablet presses from the manufacturers cited are alternatively available with turrets for standard size tooling.

Two-piece configurations with interchangeable punch tips can also be used in punches for rotary presses. Special care and experience are necessary as these punches can be difficult to assemble and align; they are mostly used with multiple tools in order to increase productivity, thus offsetting additional costs.

Korsch (www.korsch.de) has a special machine for solving sticking problems, where the entire punches are rotated before and after the compression phase in order to prevent adhesion. These punches have a cogwheel around their neck.

From Manesty (<http://www.manesty.com>) a special rotating-head tooling is available which enables the punch head of shaped tooling to rotate in order to increase the lifetime of punches and cams.

By IMA (<http://www.ima.it>), a special high-speed machine design has been developed. Feeding happens at the center, taking advantage of increasing centrifugal forces. This design requires special tooling, which is made in two pieces. This means that only the punch tips need to be changed, which saves time in this special machine.^[5]

Instrumented punches and dies, which enable online measurement of forces on the punch and relative displacement or acceleration, electrical resistance, and acoustic release as well as measurement of die wall forces, are special tools used exclusively in research; they are discussed in a separate chapter.

PUNCH TIPS

First of all, one has to make a general distinction between round and shaped tablets. The nomenclature of tablet shapes of round flat and convex tablets as well as the respective punch tip forms are shown in Fig. 12. The tip edge design of the punch and the forces it is exposed to during compression, and the resulting deflection forces are shown in Fig. 13. It is impossible to maintain an infinite sharp edge as this would wear off immediately; the figure also shows how a land is introduced by polishing the punch tip in order to increase its pressure tolerance. The land/cup connection can be designed as a sharp corner, a standard curvature (radius), or a blend radius.

Round tablet terminology

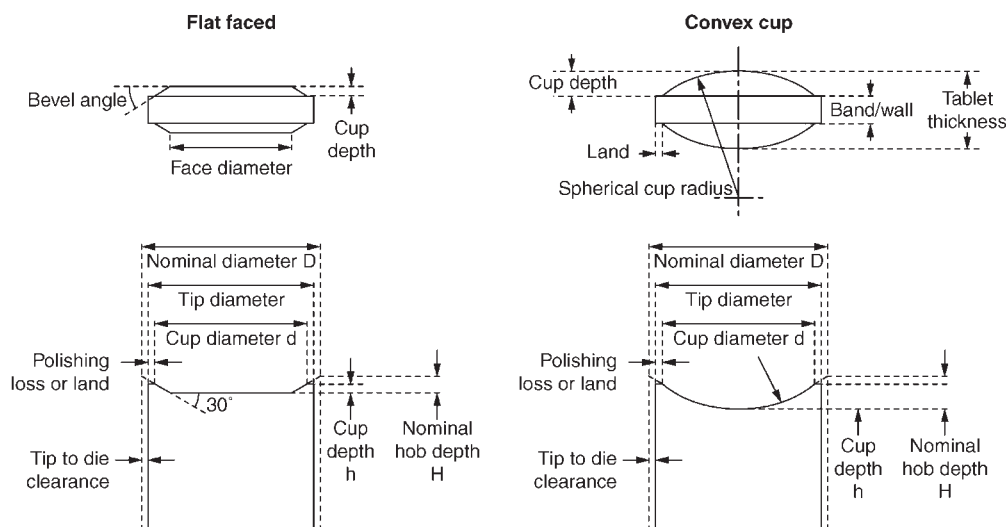
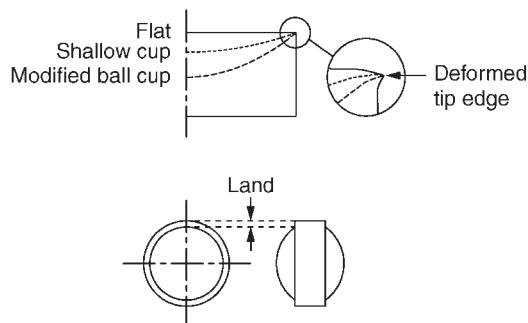


Fig. 12 Terminology of round tablets for flat-faced and convex tablets.

Sharp tip edge deformation



Tip deflection under load

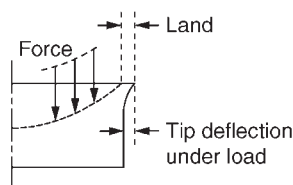


Fig. 13 Punch tip edges: wear and deformation under load, effect on tablet shape and tip deflection.

Round tablet profile terminology can be found in Fig. 14.

Similarly, for shaped tablets, terminology is explained in Fig. 15, and well-known examples of shapes and profiles are depicted in Fig. 16.

In order to avoid sticking, it is recommended to avoid corners. Compared to the profile of the land/cup connection, an alteration of the common connection between the bevel and the flat by rounding the beveled edge of flat-faced punches, thus reducing the possibility of sticking of material to the sharp edge, has been proposed to reduce sticking problems.^[6]

Break lines are most often bisects or quadrisects, whereas engravings often comprise characters and graphic objects facilitating identification of the tablets. Engravings on tablets can be embossed or debossed (Fig. 17), the punch has to be the respective opposite, and the characters and figures are mirrored. In many cases, break lines and engravings trigger sticking problems. In order to quantify the pull-off force of a tablet sticking to a punch surface, a special instrumentation was developed^[7] and used for the optimization of the angle between the engravings and the surface.^[8] As sticking in most cases starts at the lateral faces of the engravings due to shear forces occurring there

Round tablet profiles

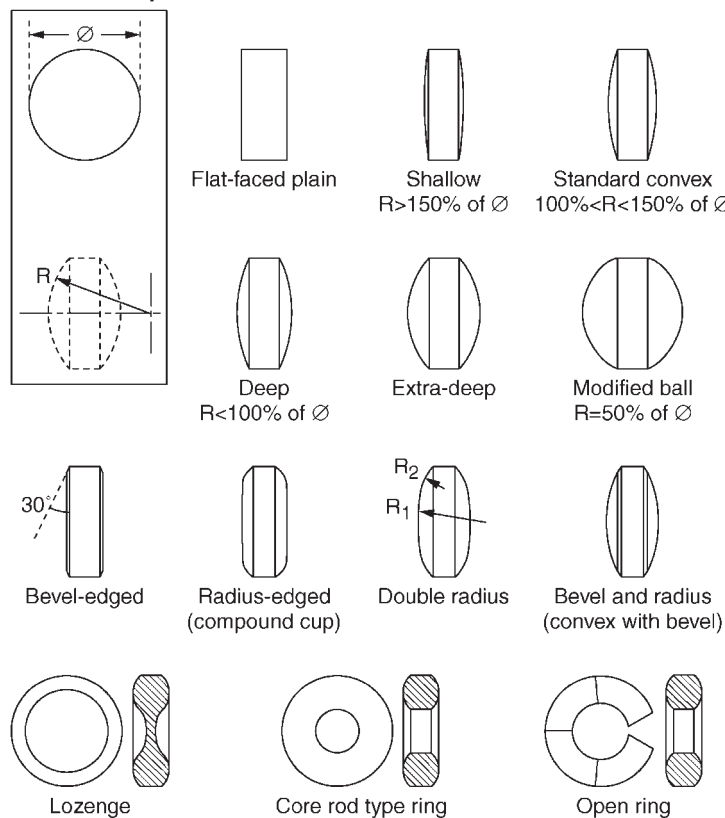


Fig. 14 Round tablet profiles. (Open ring tablet: From Notter, A. R. German Patent DE 4342146 C1, December 10, 1993.)

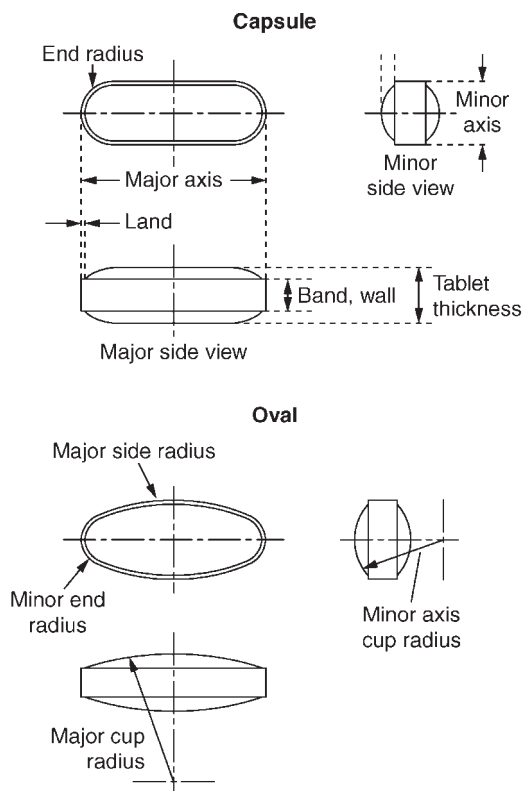
Shaped tablet terminology

Fig. 15 Terminology of shaped tablets: profiles of capsule-shaped (oblong) and oval tablets.

during compression, the authors quoted above generally recommend a small angle of the stroke to the surface (between 45 and 75° —as to the question of how to define the opening angle, see Fig. 18). This is in contrast to findings of Sabir, Evans, and Jain,^[9] who recommend an angle of 90° , but their study lacks conclusiveness as the steep angle tooling, unlike the reference tools, was chromium-plated. Waimer et al. used a cone for their measurements,^[7,8] and it can be assumed that engravings comprising a greater number of lateral faces per area would need even more attention. Empirically, it is recommended to avoid sharp corners as in characters like “A” and to use radii instead; tooling manufacturers have their special rounded character type sets (Fig. 18). For film coating, the profile of the engravings must be particularly flat-angled in order to ensure wetting with the coating material. The maximum area on the surface of the tablet that can be used for engravings depends on the cup depth.

Requirements for break lines have been increased during the past years. All break lines must ensure that the tablet breaks into equal parts and lines for decoration purposes are not legal any more. Guaranteeing

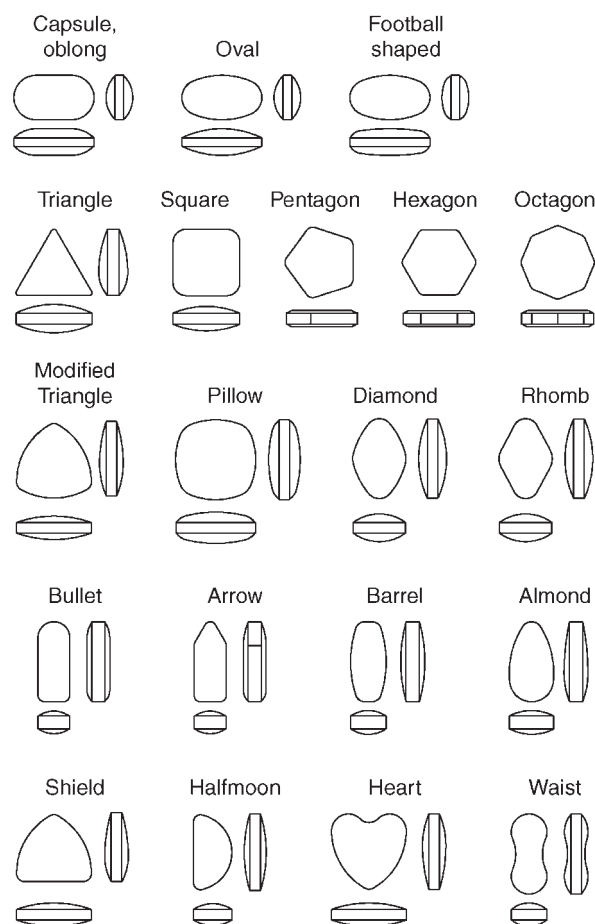


Fig. 16 Examples of well-known tablet shapes. (Courtesy of Ritter Pharmatechnik, D-Hamburg.)

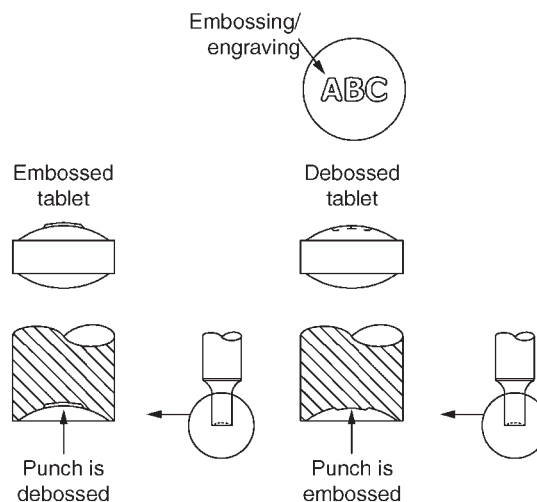


Fig. 17 Embossing/debossing of engravings on tablets and punches.

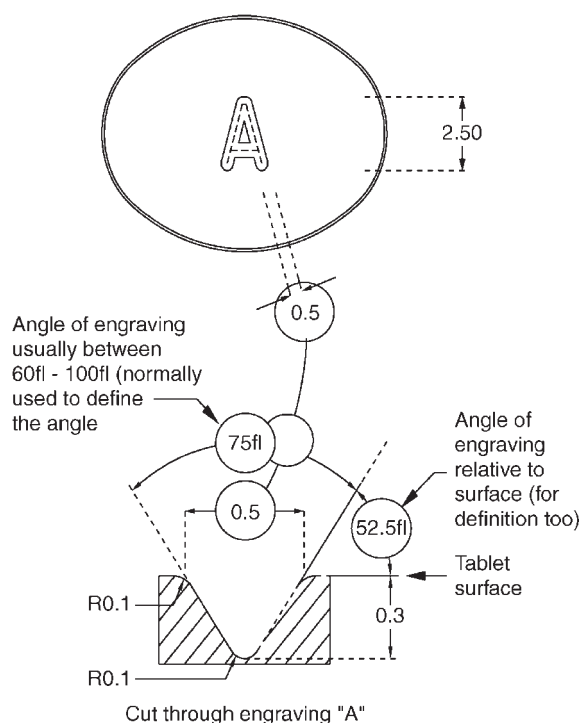


Fig. 18 Terminology of engravings, exemplified by the character "A." (Courtesy of Ritter Pharmatechnik, D-Hamburg.)

a reproducible crack is not trivial as the powder located right under the break line is exposed to the highest compression stress and most compaction and is therefore in many cases mechanically strongest. Therefore, break lines should not be too deep, and notching of the outer tablet wall serves to ensure a defined crack line. Notching requires coves in the die wall and, consequently, punch rotation has to be prevented. Some typical standard bisect line types are shown in Fig. 19. The shape, angle, and depth of bisects with regard to the shaped tablet surface has been optimized in order to facilitate the breaking of tablets by pressing them against a hard pad with one finger (Fig. 19, lower line, and Figs. 20 and 21).

In bigger tablets other than those commonly used pharmaceutically, the requirements for reproducibility are lower, which, e.g., makes it possible to press a trough into the tablet surface so that a second mixture of tensides for dish-washing can be applied after tableting. The punch can be designed as an interchangeable cone made from synthetic material.^[10]

Multiple tools have several punch tips on one and the same barrel; the corresponding die has as many bores in order to increase productivity. One disadvantage results from the more delicate handling of the tools, but multiple tools are definitely useful for small tablets in order to increase productivity. Minitablets of 2 mm diameter are

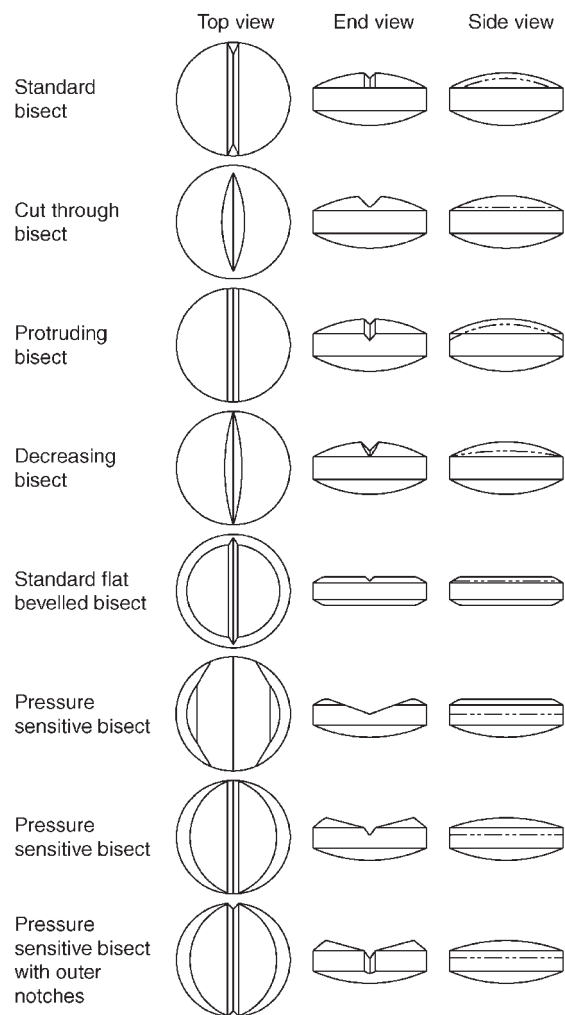


Fig. 19 Forms of commonly used bisects. (Courtesy of Ritter Pharmatechnik, D-Hamburg.)

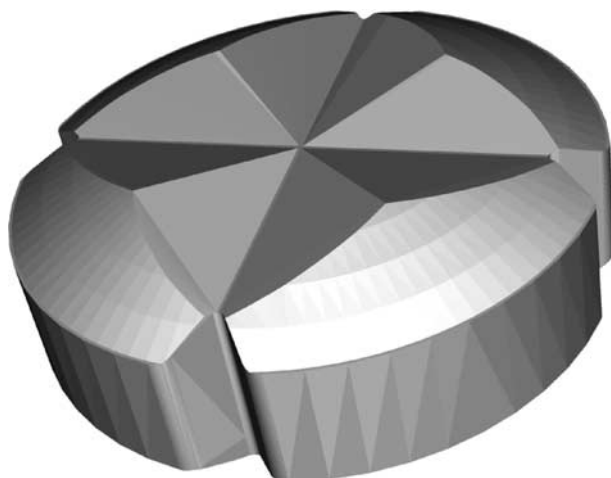


Fig. 20 Example of a pressure-sensitive quadrisect tablet. (Courtesy of Ritter Pharmatechnik, D-Hamburg.)

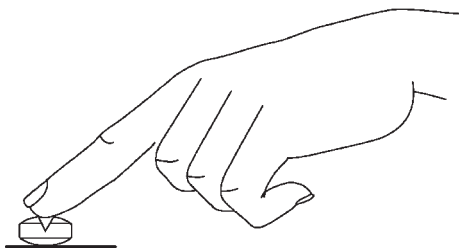


Fig. 21 Method for dividing pressure-sensitive tablets into equal parts.

produced as a substitute for pellets and offer the advantage of better reproducibility of filling into capsules as well as better reproducibility of their specific surface area, which is important for the coating process and the subsequent reproducible release of the active drug. Multiple punch tips are frequently manufactured as multiple tools in a several-piece punch design. A 19-fold punch set is available and working in standard production of minitabets. Up to 55-fold tooling has proved to function in experimental studies (<http://www.ritter-pharmatechnik.de>). The setup of the tiny tooling is depicted in Fig. 22.

MATERIALS, MANUFACTURE, FINISH, AND PRESSURE TOLERANCE

Materials and Manufacture

Steels are basically classified as carbon and alloy steels which are defined by national and international standards, e.g., ASTM and DIN. Carbon steels contain less than 1.65% manganese, 0.6% silicon, 0.6% copper, plus boron and deoxidizers as well as carbon as the principal additive increasing hardness. Manganese also increases hardness, particularly the ability to be hardened by tempering, and is present in all steels. Silicon acts as a deoxidizer. Alloy steels exceed these limits or contain additional elements, e.g., chromium, vanadium, tungsten, molybdenum, nickel, and cobalt. A high-nickel content increases resistance to corrosion caused by HCl released from drug salt.^[11] The composition of special steel types for tooling, their provenance, and their thermal treatment are part of the trade secret of tooling manufacturers.

A high abrasion resistance usually corresponds to high hardness of a steel type, whereas its ductility and toughness is low and vice versa: very hard tooling may

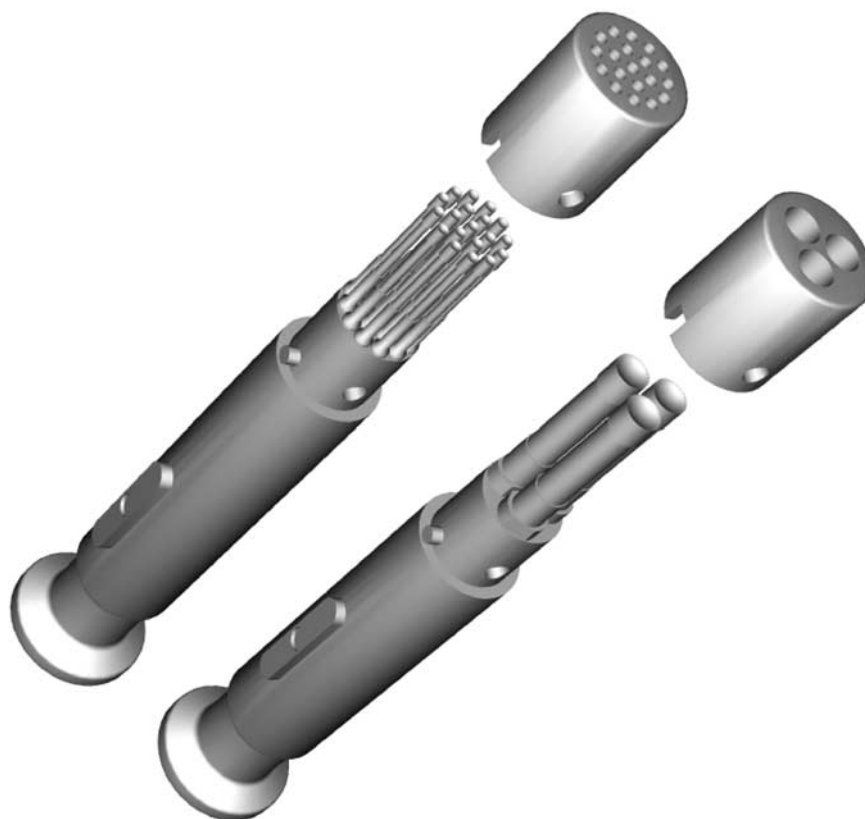


Fig. 22 Setup of multitools, left: minitabets, 2 mm diameter; right: triple-tool for 6-mm tablets. (Courtesy of Ritter Pharmatechnik, D-Hamburg.)

Table 2 Hardnesses of steel qualities used for tooling (Rockwell hardness)

Steel	S1(BS1)	S7	BD2 (D2)	BD3(D3)	440C stainless
Rc-hardness	54–56	54–58	58–60	58–60	56–58

fracture, ductile material may wear off quickly. Hardness and ductility do not only depend on the chemical composition of the steel but also on its heat-treatment (tempering). A second tempering procedure for punch tips, which is computer-controlled and carried out at high temperatures in a vacuum furnace, softens the material and prevents fracture.

Punches usually have a longer lifetime than dies if they are made from the same material; that is why manufacturers often choose different steel qualities. The surface of a die bore can become smoother or rougher during its use. The ejection force does not—apart from the composition of the tableting material and the compaction force—necessarily depend on roughness of the die wall, but to a high degree depends on the metal type of the die.^[12]

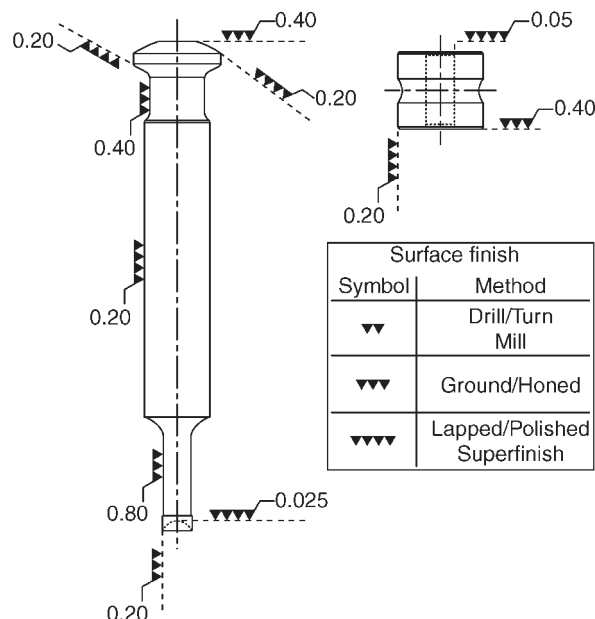
For specific tableting tasks, the tooling manufacturer will be able to provide optimum tool properties. It may be helpful in this respect to supply him with a sample of the tableting material in question.

Table 2 gives hardness characteristics of some steel qualities that are frequently used. 440C has the highest chromium content (17%), which increases wear resistance. D2 and D3 have a high-chromium content (12%) and are additionally carbon-rich (over 1.5%). Chromium and carbon together form stable carbide, which increases hardness. S1 contains tungsten (2%), which makes heat-treatment easier. More about steel qualities and international standardization thereof is available from steel manufacturers or the relevant authorities via the World Wide Web (e.g., <http://www.sz-metal.si>).

Finish

Adhesion, sticking, picking, and filming on punches and dies can in many cases be avoided by a suitable surface finish of the tooling. In some cases, the surface is smoothened during continuous use of punches, whereas in other cases, depending on the material compressed, e.g., in the case of crystalline lactose, small scratches may occur on the punch tips after a few thousand compression events.^[13] On those small scratches and defects in the surface, corrosion will start.

Tolerances for the surface finish of tooling are also specified by manufacturers and in international standards, see Fig. 23. Any roughness of the finishing is below 1 μm

**Fig. 23** Surface finish of punches; tolerances for top-to-valley differences (in μm).

peak-to-valley height, and particularly punch-tip faces and die bore walls are polished to achieve lower tolerances.

Coatings and Platings on Punch Tips

Reduced ejection forces can help to avoid both lamination and capping and give tablets a shiny surface as well as high mechanical strength. Internal lubrication (lubricants added to the tableting material) is counter-productive not only because its admixture is another critical step in production, but also because the most effective lubricants are hydrophobous and make the tablets softer and increase disintegration time as well as dissolution rate. Attempts to coat the tooling with fluid or solid lubricants just before each compression event are relatively complicated with respect to machinery, reproducibility, dust formation, and GMP demands, e.g., batch definition. In many cases, an easier alternative may be to apply a finish to the punches that reduces friction by special coating or plating of the punch tips so that ejection forces are reduced. Another reason for coating the punches may be to decrease the adhesiveness of tablets to the punches as this does not necessarily decrease with reduced ejection forces. It may also be necessary to increase corrosion resistance in cases where aggressive chemicals are compressed. A critical demand on all coatings and platings is their wear resistance. The thickness of the coat must be taken into account as it affects clearances and tolerances. As this is also a part of the trade secret of manufacturers and specialized companies, little is published on this subject.



Hard chromium plating is applied galvanically with a thickness of approximately $30\text{ }\mu\text{m}$. It should not be confused with blank chromium plating, which consists of several layers of other metals with a very thin chromium layer on top and is not suitable for tooling. Hard chromium platings, although appearing shiny and smooth, are more porous than the blank coatings, brittle, and in many cases not sufficiently wear-resistant. Moreover, any transition of traces of chromium into the product would be totally unacceptable for pharmaceutical use. Not in all cases do they have the desired effect with respect to sticking.^[14]

Nodular thin dense chromium is an FDA-approved coating which is applied to the entire punch and all its faces.

Chromium can be implanted into the surface in the form of chromium nitride at a depth of up to $1\text{ }\mu\text{m}$ by nitrogen ion bombardment of the surface. This treatment proved to have a favorable effect on sticking problems^[14] and makes the surface relatively wear-resistant. Titanium can be applied in the same way in the form of TiN. This treatment leaves the tools in a “bronzed” appearance because the surface is relatively rough as seen in the electron microscope. Increased hardness and greater wear and corrosion resistance is achieved, but the effect on friction is smaller than in the case of CrN.

Nickel–chromium–boron alloying coat applied on both punches and dies has been described^[15,16] and compared to untreated tools. Surprisingly, depending on the material under compression, the coating had the expected effect, or there was a total lack of effect, or they even increased ejection forces. The authors concluded that anyway, the tablets were superior in appearance.

Titanium carbide is by far the hardest material available for the treatment of surfaces.

Teflon coating has not proven to be wear-resistant enough to tolerate tableting.

Carbon in layer thickness of $0.1\text{ }\mu\text{m}$ – $10\text{ }\mu\text{m}$, doped with hydrogen in 5 at.%–30 at.% has proven to be extremely wear-resistant^[17] and to prevent sticking of tablets. The amorphous carbon layer is applied by CVP (chemical vapor deposition), which can also be used to apply coatings of metal, alloy, carbide, and nitride. It generally yields lower adhesion and superior wear resistance of the coatings, but the high temperatures necessary for the process (1000°C) would soften the steel and distort it. Ion plating is an alternative ensuring low adhesion and minimum distortion of the punches.

Alternative Materials

Inserted (lined) dies are fitted with an insert (liner) in the outer shell of the metal and consist of a much harder material such as tungsten carbide or a ceramic. Liners are used for compression of abrasive or corrosive materials.

The outer shell of steel protects the harder, more brittle liner from possible failure, e.g., by the die locking screw.

Maximum Pressure Tolerance

The pressure tolerance depends on the smallest diameter of the punch, e.g., the backing-off lathe at the punch tip and the cup depth. It is calculated from the load capacity of the tool steel. A guide for maximum pressure tolerance is given in Table 3.

Lifetime

Fatigue failure is cumulative because any over-stress inflicts microscopic damage. It correlates with the maximum stress in each loading cycle: a 25% reduction in stress may prolong the lifetime tenfold. It is advisable not to exceed the stress level over the maximum allowable limit for infinite lifetime; it is much better to change the design of the punch-tip cup in order to eliminate fatigue failure.

Average lifetime is most affected by granule composition and lubrication properties, lying between 200,000 and several millions of compression events.^[18]

Principal reasons for termination of punch life are damage to heads and tips, rolled-in or burred tip edges, pitted faces, distorted and/or flattened bisects or engravings, and undersized tip lands. For dies, it is the wear in the die bore that terminates lifetime, whereas for carbide lined dies it is often distortion leading to die screw groove. Increased clearances accelerate wear, e.g., by extrusion of

Table 3 Maximum pressure tolerated by round tools (kN); estimated values as a guideline

Diameter		Flat-faced	Flat with bevel	Concave (depending on cup depth)	Deep concave and modified ball
mm	in.				
4	5/32	10	6	4–7	2
5	3/16	15	10	6–11	3
6	7/32	22	14	8–15	4
7	9/32	30	19	11–20	6
8	5/16	39	25	15–23	8
9	11/32	50	31	18–34	10
10	13/32	61	38	23–41	13
11	7/16	81	50	30–55	16
12	15/32	88	55	38–65	18
13	1/2	104	65	50–82	20
15	19/32	115	77	60–110	27
20	25/32	185	144	90–180	51
25	1	240	210	180–220	90

10 kN = 1 t.

(From Ref. [2].)



powder between punch and die. During lifetime the deflection of the upper punch in the turret can double, which leads to rolled-in punch tips. Cam design and cam track tolerances determine wear on punch heads with smaller tolerances used on modern machines prolonging punch life.

MAINTENANCE OF PUNCH SETS

Sets of toolings, i.e., upper punch, lower punch, and die, should not be interchanged. Therefore, they are usually marked by engraved numbers, punches at the neck, dies on the outer wall. Special care must be taken when toolings are fitted to the tablet press. Fitting instructions usually follow with the manual for the respective press. In general, careful aligning is crucial, for hands-on hints see Ref. 19 and ask the tooling manufacturer. Special maintenance kits useful for changing punches and containing a set of aids are available from press manufacturers and tooling manufacturers.

Immediately after the end of the production campaign, and in the case of long-lasting production cycles also during runs (e.g., daily), tooling should be carefully cleaned. Unless a special washing machine is used, warm soap water and a soft brush are suitable for thorough cleaning; if necessary, a few minutes in an ultra-sound bath can be helpful. Be careful not to scratch the surfaces! Upon rinsing with warm water, ethanol or isopropanol is used to remove any remaining oil. Removing punches and dies for washing takes a lot of time. Some tablet presses (e.g., from Fette, Korsch) are constructed in a way that enables interchangeability of the whole turret including tooling, and the cleaning of the turret takes place in a fully automatic washing machine at the same time as the rest of the tablet press is cleaned in place, which very much reduces turn-over-time, especially if two separate turrets are available.

Tooling should not be touched with bare fingers after washing, and usually all the tooling is thoroughly coated with an acid-free oil in order to prevent corrosion. Instead of expensive commercial products, paraffin and vaseline in pharmacopoeial quality are also suitable. Some users want to avoid all oiling as oil may contribute to sticking problems; in this case punches are washed with alcohol, possibly dusted with magnesium stearate, and stored in a humidity-controlled atmosphere.

Surface roughness and alteration in tooling dimensions are results of use and wear. Tooling surfaces should be frequently checked (magnifying glass). Timely polishing will prolong the lifetime of the tooling. Polishing can be done conventionally by hand on a lathe using ultra-fine polishing paste, polishing wool, and felt or a polishing brass brush. Alternatively, one can use cost-effective “drag

finishing” machines (e.g., from Otec, <http://www.otec.de>), which drag and rotate punches and dies through a granulate consisting of slightly abrasive particles coated with a special paraffin base. This process takes about 1 hr, depending on the abrasiveness of the granulate. The process does not round off the tooling but only decreases surface roughness and can be used whenever tooling appears dull or discolored. Furthermore, the punch barrels are polished too, which may help to prevent rough motion of the punches.

The wear pattern on punch heads, punch tips, and in die bores may indicate specific problems during tableting. Measuring dimensions of punches and dies is a task which has to be performed frequently and can be carried out computer-aided. Dimensions to be checked are overall length, cup depth, working length, diameters, tip diameter, and die bore dimensions. Use a micrometer.

For estimating the clearance of the upper punch in the turret and wear of the turret guide, it is easiest to measure the deflection of the upper punch tip, which should be approximately 100 μm . If it is more than 0.3 mm, the punches and/or guides are worn out.

Wear is unavoidable and it alters the dimensions of the tooling sets with time. It is, therefore, recommendable to use spare sets of tooling in regular turn-over in order to equalize wear of all tooling sets.

Storage cabinets for tooling should be clearly laid out for individual punch sets and protect them from clashing, dust, and humidity. Cabinets where punches lie or hang are available in many models made from different materials and with different degree of air-tightness.

A special tooling utility case system, consisting of individual punch and die trays enabling the combination of as many trays as needed, has an open design for easy inspection and cleaning (Ref. 20, available from Holland, <http://www.iholland.co.uk>).

Demands from GMP and the necessity of connecting tool sets to tablet products led to the introduction of “tooling management” with the help of special software (e.g., TM II from Natoli, <http://natoli.com/>, and Vali-Scan from Holland), where information on each individual encoded punch and die set is collected to provide a tooling life record. The software may be installed on a mobile trolley and the code on each individual tool will be checked while it is set onto the machine.

CONCLUSION

Many common tableting problems (e.g., sticking, capping, and corrosion) which cannot be overcome by optimizing the properties of the tableting material and/or the compression cycle, can possibly be solved by changing the tooling design. Options are the form,



the material and/or the finish of the respective toolings. Although there are software programs available for designing any possible punch tip shape and for theoretically calculating tablet volumes, surface areas, and other useful features (e.g., TabletCAD, from Natoli), this is no guarantee for being able to compress the respective tablets in practice. Moreover, it may be difficult to imagine on the basis of a punch design how the tablet itself would look like. For this purpose, simulated prototypes made quickly and inexpensively from a special plastic material can be bought under the trade name Elizatab (Elizabeth Carbide, <http://www.eliz.com>). As tooling manufacturers have been working with effective tablet designs for many years, it is generally recommendable to ask their advice whenever questions arise as to the design of tooling or tablets.

ACKNOWLEDGMENTS

The author wants to thank Mrs. S. Ritter, Ritter Pharma-Technik GmbH, Hamburg, Germany, for sharing her experience in tooling manufacture for production and research, Mr. Rod Wolstenholm, University of Tromsø, for doing the drawings, and Mrs. Gudrun Hansen, Tromsø, for help with the English.

REFERENCES

1. *Tablet Specification Manual (TSM)*; American Pharmaceutical Association: Washington, 1995.
2. Ritschel, W.; Bauer-Brandl, A. *Die Tablette*, 2nd Ed.; Editio Cantor Verlag: Aulendorf, Germany, 2002.
3. Ling, W.C. Tooling as a Factor in Tablet Weight Variation. *J. Pharm. Sci.* **1973**, *62* (12), 2007–2011.
4. Notter, M.; Notter, K. Tablettierwerkzeug, Insbesondere Tablettierstempel. European Patent 0 841 152 A2, October 24, 1997.
5. Fabbri, G.B. Tool-Holder for Tablet Making Machine. US Patent 5,882,696, March 16, 1999, I.M.A. Industria Macchine Automatiche.
6. Cargille, J.J. Tooling Face Configuration—Particularly Adapted for Forming Tablets (Cargille Curve). US Patent 5,164,206, November 17, 1992.
7. Waimer, F.; Krumme, M.; Danz, P.; Tenter, U.; Schmidt, P.C. A Novel Method for the Detection of Sticking of Tablets. *Pharm. Dev. Technol.* **1999**, *4* (3), 359–367.
8. Waimer, F.; Krumme, M.; Danz, P.; Tenter, U.; Schmidt, P.C. The Influence of Engravings on the Sticking of Tablets. Investigations with an Instrumented Upper Punch. *Pharm. Dev. Technol.* **1999**, *4* (3), 369–375.
9. Sabir, A.; Evans, B.; Jain, S. Formulation and Process Optimization to Eliminate Picking from Market Image Tablets. *Int. J. Pharm.* **2001**, *215* (1–2), 123–135.
10. Holderbaum, T. Tablettierstempel. German Patent DE 199 08 027 C1, February 2, 1999, Henkel KGaA.
11. Narurkar, A.N.; Purkaystha, A.R.; Sheen, P.C. Effects of Various Factors on the Corrosion and Rusting of Tooling Material Used for Tablet Manufacturing. *Drug Dev. Ind. Pharm.* **1985**, *11* (8), 1487–1495.
12. Otz, M.; Thoma, H. Effects of the Metal Type and the Roughness of the Die Wall on the Expanded Work for Tablet Ejection. *Pharm. Dev. Technol.* **2000**, *5* (1), 19–26.
13. Müller, B.W. Über Veränderungen der Pressflächen von Tablettierwerkzeugen. *Die Pharmazeutische Industrie* **1972**, *34* (12), 972–976.
14. Schumann, S. The Effects of Chromium Nitride Ion Bombardment Treatment of Tablet Tooling on Tablet Adherence. *Drug Dev. Ind. Pharm.* **1992**, *18* (19), 1037–1061.
15. Tsiftoglou, T.B.; Mendes, R.W. Effect of Boron Alloy Coating of Tableting Tools. *Pharm. Technol.* **1982**, *6* (April), 30–32, 36–38.
16. Shah, K.B.; Augsburg, L.L.; Shangraw, R.F. Effect of Boron Alloy Coating of Tableting Tools. *Pharm. Technol.* **1982**, *6* (April), 31–34, 41–42, 44, 49–50, 52, 54.
17. Hieke, A.; Grischke, M.; Brand, J. Vorrichtung zum Verpressen von Fließförmigen Feststoffen. German Patent DE 198 10 969 A1, March 13, 1998.
18. Swartz, C.J.; Anschel, J. Evaluation of Tableting Tool Life Records. *J. Pharm. Sci.* **1968**, *57* (10), 1779–1782.
19. Ritschel, W.; Bauer-Brandl, A. Tablettierwerkzeuge. *Die Tablette*, 2nd Ed.; Editio Cantor Verlag: Aulendorf, Germany, 2002, 552–567.
20. Bal, G.W. Tooling Utility System. US Patent 6,112,896, September 5, 2000.

TRANSDERMAL DELIVERY: SONOPHORESIS

Samir S. Mitragotri

University of California, Santa Barbara, California, U.S.A.

Hua Tang

Daniel Blankschtein

Robert Langer

Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.

TRANSDERMAL DRUG DELIVERY

Systemic as well as topical delivery of drugs via the transdermal route is limited by the low skin permeability which is attributed to the stratum corneum (SC), the outermost layer of the skin (1). The SC consists of disk-like dead cells (keratinocytes) containing keratin fibers and water, surrounded by densely-packed lipid bilayers. The highly-ordered structure of the lipid bilayers confers a highly impermeable character to the SC. A variety of approaches have been suggested to enhance transdermal drug transport. These include: 1) use of chemicals to either modify the skin structure or to increase the drug concentration in the transdermal patch (2,3); 2) applications of electric fields to create transient transport pathways [electroporation] (4,5) or to increase the mobility of charged drugs through the skin [iontophoresis] (6); and 3) application of ultrasound [sonophoresis] (7–56).

Sonophoresis was shown to enhance transdermal drug transport about half a century ago by Fellingner et al. (16) who showed that application of ultrasound increases transport of hydrocortisone across the skin. Following this study, attempts were made to enhance transdermal transport of more than 15 drugs including steroidal anti-inflammatory drugs such as hydrocortisone, dexamethasone; non-steroidal anti-inflammatory drugs such as salicylates and ibuprofen; anesthetic agents such as lidocaine; and proteins such as insulin. This chapter provides a review of these studies with emphasis on associated techniques, mechanistic studies, and safety.

GENERATION AND APPLICATION OF ULTRASOUND FOR SONOPHORESIS

Generation of Ultrasound

Ultrasound is a sound wave possessing frequencies above 20 kHz (57,58). These waves are characterized by two

main parameters: frequency and amplitude. Amplitude of ultrasound waves can be represented in terms of peak wave pressure (in Pascals) or in terms of intensity (in the units of W/cm^2). Ultrasound can be applied either continuously or in a pulsed manner. In the latter case, an additional parameter, duty cycle, is required to characterize ultrasound application. Duty cycle is the fraction of time for which ultrasound is ON.

Ultrasound is generated using a device referred to as a sonicator. It consists of an electrical signal generator which generates an electrical AC signal at the desired frequency and amplitude. This signal is applied across a piezo-electric crystal (transducer) to generate ultrasound. The thickness of the piezo-electric crystal is selected so that it resonates at the operating frequency. Sonicators operating at various frequencies in the range of 20 kHz to 3 MHz are available commercially and can be used for sonophoresis.

If a sonicator operating at the desired frequency is not available commercially, it is possible to assemble one using commercially available signal generators, amplifiers, and transducers. Such sonicators operating at frequencies of 10 MHz and 16 MHz have been assembled by Bommannan et al. (11). (For a discussion of the relevant methods for making a custom sonicator, see Ref. 11.)

For sonophoretic delivery, the desired drug is dissolved in a solvent and applied on the skin. Ultrasound is applied by contacting the transducer with the skin (see Fig. 1A–C) through a coupling medium to ensure a proper contact between the transducer and the skin. This medium can be the same as the solvent used to dissolve the drug or it can be a commercially available ultrasound coupling gel (for example, Aquasonic, Polar, NJ).

Transmission of Ultrasound from the Transducer to the Skin

Transmission through the medium

Ultrasound requires a coupling medium for transmission from the transducer to the desired tissue. The coupling

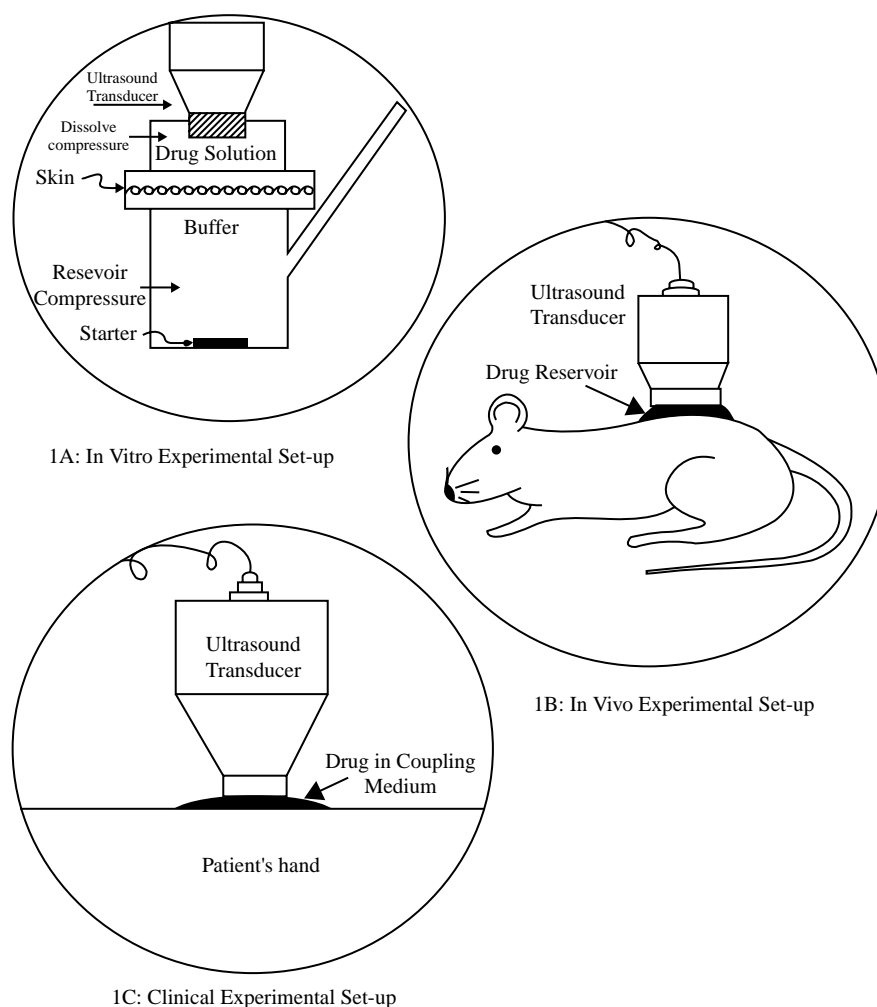


Fig. 1 Experimental set-up for sonophoresis delivery.

medium should result in proper transmission of ultrasound from the transducer to the skin. The transmissive properties of a medium are indicated by its acoustic impedance (Z). A coupling medium is appropriate for sonophoresis if its acoustic impedance (Z) is comparable to that of skin ($1.6 \times 10^6 \text{ kg/m}^2/\text{s}$). Z -values for various materials can be found in (57, 58). For example, water has a Z -value of $1 \times 10^6 \text{ kg/m}^2/\text{s}$ and is a reasonable coupling agent. Z -values for several media are listed in Table 1.

Absorption of ultrasound

Every medium absorbs ultrasound to a certain extent. The ability of a medium to absorb ultrasound is indicated by the absorption coefficient (α). The extent of absorption is given by the following equation.

$$f(\tau) = 1 - \exp(-\alpha\tau)$$

where $f(\tau)$ is the fraction of ultrasound intensity absorbed as the ultrasound beam propagates in a medium with absorption coefficient α and thickness τ . An estimation of α -values for various materials at various ultrasound frequencies may be found in Refs. (57, 58). In the case of water, the α -value is 0.0006 at an ultrasound frequency of 1 MHz, suggesting that a 1 cm thick column of water absorbs less than 0.1% of ultrasound (1 MHz) intensity, i.e., water is a reasonable coupling medium. Absorption coefficients for several media are listed in Table 1.

Ultrasound reflection

Ultrasound is reflected at the boundary of two media possessing different acoustic impedances. 99.99% of ultrasound is reflected at the air-water boundary when an ultrasound beam is incident upon it from either side. Hence occurrence of air bubbles should be minimized in

Table 1 Acoustic impedances and absorption coefficients of materials

Material	Acoustic impedance, Z ($\text{kg/m}^2/\text{s}$)	Absorption coefficient (α) at 1 MHz (cm^{-1})
Water	1.5×10^6	0.0006
Blood	1.6×10^6	0.028
Bone	6.3×10^6	3.22
Skin	1.6×10^6	0.62
Fatty tissue	1.54×10^6	0.14
Muscle	1.6×10^6	0.76
Air	0.0004×10^6	2.76

(From Hoogland, R., *Ultrasound Therapy*; Ernaf Nonius: Delft, Holland, 1986.)

the coupling medium in order to avoid ultrasound reflection. The reflection coefficient for various interfaces may be estimated from the acoustic impedances of two media forming the interface using equations described in (57, 58).

Selection of Ultrasound Parameters

Proper selection of ultrasound parameters is required to ensure safe and efficacious sonophoresis. Ultrasound parameters such as frequency, intensity, duty cycle, and distance of transducer from the skin influence the efficiency of sonophoresis. Below we present a general discussion of the role played by various ultrasound parameters in sonophoresis. Note that the objective of this discussion is not to point out the exact values of ultrasound parameters to be selected, but rather to present information regarding the dependence of sonophoretic enhancement on each parameter.

Ultrasound frequency

Ultrasound at various frequencies in the range of 20 kHz to 16 MHz has been used for sonophoresis. These studies of sonophoresis can be classified into three categories based on the ultrasound frequency used, i.e., therapeutic, high-frequency, and low-frequency ultrasound.

Therapeutic frequency ultrasound (1–3 MHz): This is the most commonly used ultrasound frequency range for sonophoresis. Specifically, over 90% of the previous studies of sonophoresis have been conducted using therapeutic ultrasound. A summary of these studies is provided in Table 2. Interestingly, in the therapeutic frequency range of 1–3 MHz, frequencies closer to 1 MHz

have been preferably used for sonophoresis. No reason has been given by investigators for the use of this particular frequency. Mitragotri et al. (36) reported that the sonophoretic enhancement in the therapeutic frequency range varies inversely with ultrasound frequency. They found that while 1 MHz ultrasound enhances transdermal transport of estradiol across human cadaver skin in vitro by 13-fold, 3 MHz ultrasound at the same intensity induces an enhancement of only 1.5-fold. They further hypothesized that the observed inverse dependence of sonophoretic enhancement on ultrasound frequency occurs since cavitation effects, which are primarily responsible for sonophoresis, vary inversely with ultrasound frequency (37, 59).

High-frequency ultrasound (above 3 MHz): Bommanan et al. (11, 12) performed sonophoresis of salicylic acid and lanthanum tracers across hairless rat skin in vivo using high-frequency ultrasound ($f = 2, 10$, and 16 MHz) (Table 3). They investigated the dependence of sonophoresis on ultrasound frequency in the high-frequency region and found that 10 MHz ultrasound is more effective in enhancing transdermal transport of salicylic acid than that at 16 MHz, which in turn, is more effective than that at 2 MHz. They proposed that the sonophoretic enhancement in the high-frequency region should vary directly with ultrasound frequency, though the anomalously high efficiency of sonophoresis at 10 MHz was due to higher efficiency of the transducer operating at that frequency.

Low-frequency ultrasound (below 1 MHz): Tachibana et al. (52, 53) have reported use of low-frequency ultrasound (48 kHz) to enhance transdermal transport of lidocaine and insulin across hairless mice skin. Very low-frequency ultrasound has also been used by Mitragotri et al. (37, 38) to enhance transport of various low-molecular weight drugs including salicylic acid, corticosterone as well as high-molecular weight proteins including insulin, γ -interferon, and erythropoietin across human cadaver skin in vitro. They investigated the dependence of sonophoretic enhancement in low-frequency region using two ultrasound frequencies, 20 and 40 kHz, and found that the sonophoretic enhancement of transdermal salicylic acid flux induced by 20 kHz ultrasound is up to 7-fold higher than that induced by 40 kHz ultrasound at the same intensity. The inverse dependence of sonophoretic enhancement on ultrasound frequency was hypothesized to occur due to inverse dependence of cavitation effects on ultrasound frequency (59).

Ultrasound intensity

Various ultrasound intensities in the range of 0.1–2 W/cm² have been used for sonophoresis. In most

Table 2 Literature reports of therapeutic sonophoresis

Drug	Molecular weight (Da)	Experimental system	Ultrasound conditions	Experimental conclusions ^a	Ref.
Caffeine	194	Human skin in vitro	1 MHz, 2 W/cm ²	0.2 ± 0.4	(37)
		Hairless rat in vitro	1 MHz, 2 W/cm ²	1	(33)
Corticosterone	346	Human skin in vitro	1 MHz, 2 W/cm ²	3 ± 0.6	(37)
Dexamethasone	392	Swine	1 MHz, 1.5 W/cm ²	Significant enhancement	
Estradiol	272	Human skin in vitro	1 MHz, 2 W/cm ²	12 ± 1.5	(37)
Fluocinolone acetonide	452	Human skin in vivo	1 MHz, 2 W/cm ²	Significant enhancement	(35)
		Dogs	1 MHz, 0.3–1 W/cm ²	Significant enhancement	(42)
		Human skin in vivo	1 MHz, up to 2 W/cm ²	Significant enhancement	(42)
Hydrocortisone	362	Human skin in vivo	1 MHz, up to 3 W/cm ²	Significant enhancement	(42)
		Human skin in vivo	1 MHz, 1.5 W/cm ²	Significant enhancement	(42)
		Swine	1 MHz, 1.5 W/cm ²	Significant enhancement	(42)
		Pigs	1 MHz, up to 3 W/cm ²	Significant enhancement	(42)
Indomethacin	357	Rats	1 MHz, 0.75 W/cm ²	Significant enhancement	(40)
		Human skin in vitro	1 MHz, 2 W/cm ²	0.1 ± 0.6	^b
Lidocaine	234	Human skin in vivo	1 MHz, 0.25 W/cm ²	No enhancement	(8)
Phenylbutazone	308	Human skin in vivo	1–3 MHz, 1.5 W/cm ²	No enhancement	(53)
		Human skin in vivo	1 MHz, 2 W/cm ²	Significant enhancement	(13)
Physostigmine	275	Hairless rats in vivo	1 MHz, 3 W/cm ²	Significant enhancement	(32)
Progesterone	274	Human skin in vitro	1 MHz, 2 W/cm ²	0.1 ± 0.5	(37)
Salicylate	138	Human skin in vivo	1 MHz, 1.5 W/cm ²	No significant enhancement	(15)
		Human skin in vivo	1 MHz, 1.5 W/cm ²	No significant enhancement	(14)
Testosterone	288	Human skin in vitro	1 MHz, 2 W/cm ²	4 ± 1.1	(37)

^a The experimental conclusions are reported either as statistically significant or insignificant enhancement or in terms of a quantitative ratio of sonophoretic and passive skin permeability.

^b Unpublished data by J. Kost and R. Langer.

cases, use of higher ultrasound intensities is limited by thermal effects. Several investigations have been performed to assess the dependence of sonophoretic enhancement on ultrasound intensity. Miyazaki et al. (41) found a relationship between the plasma concentrations of indomethacin transported across the hairless rat skin by sonophoresis (therapeutic conditions) and the ultrasound intensity used for this purpose. Specifically, the plasma indomethacin concentration at the end of three hours after sonophoresis (0.25 W/cm^2) was about 3-fold higher than controls at the same time. However, increasing intensity by 3-fold (to 0.75 W/cm^2) further increased sonophoretic enhancement only by 33%. Mortimer et al. (42) found that application of ultrasound at 1 W/cm^2 increased transdermal oxygen transport by 40% while that at 1.5 W/cm^2 and 2 W/cm^2 induced an enhancement by 50% and 55%, respectively.

In the very low-frequency ultrasound region (20 kHz), Mitragotri et al. (37) have reported that permeability of human skin in vitro to insulin increased by more than 100-fold as the ultrasound intensity increased from 12.5 to 125 mW/cm^2 . This variation of sonophoretic skin permeability with ultrasound intensity is quite different from that observed in therapeutic frequency region described above.

Pulse length

Ultrasound can be applied either in a continuous or a pulsed mode. A pulsed mode of ultrasound application is used many times because it reduces the severity of adverse side effects of ultrasound, such as thermal effects. However, pulsed application of ultrasound may have a significant effect on the efficacy of sonophoresis. As will be discussed later, cavitation effects, which play a crucial role in sonophoresis, vary significantly with the pulse length. For example, the cavitation threshold in an aqueous solution at 1 MHz changes from approximately 0.3 W/cm^2 (60) to 33 W/cm^2 (61) as the mode of ultrasound application changes from continuous to pulsed, with a pulse length of 1 ms applied every 10 ms. This is because under pulsed ultrasound, during the intervals between pulses, gas nuclei formed during the previous pulse have time to dissolve back into solution, and therefore, making it more difficult to cavitate the solution (62). Mitragotri et al. (36) reported that while a continuous application of therapeutic ultrasound (1 MHz, 2 W/cm^2) increased human skin permeability to estradiol by 13-fold, a pulsed application (2 ms pulses applied every 10 ms) did not significantly enhance transdermal estradiol flux. In very low-frequency ultrasound region, Kost et al. (28) reported that urea permeability of cuprophane membranes

increased from 6 to 56% as the ultrasound (20 kHz) pulse length increased from 100 to 400 ms (applied every second).

Distance of the transducer from the skin

The ultrasound pressure (or intensity) field around a transducer is quite complex. The intensity of the ultrasound passes through a series of maxima and minima in a region near the transducer and beyond a certain distance, decreases monotonically with distance. The region in which the ultrasound intensity passes through the series of minima and maxima is referred to as the near field, and the region beyond the near field is referred to as the far field. The length of the near field of a transducer having an area of 1 cm^2 operating at 1 MHz is 1.66 cm (57). In most of the experiments reported in the literature, for which ultrasound frequencies of 1 MHz or above have been used, the distance of the skin from the transducer was probably less than 1.66 cm. As a result, in most reported experiments, the skin was in the near-field region. In very low-frequency ultrasound region, Julian et al. (23) studied the effect of transducer distance on the sonophoretic permeability of benzoic acid through a polydimethylsiloxane membrane under low-frequency conditions. They observed that the effect of 20 kHz ultrasound on the permeability of the membrane is insensitive to the distance of the transducer from the membrane. This probably occurs because of the successive reflections of ultrasound waves in the diffusion cell which prevents any systematic pressure pattern from forming in the diffusion cell.

Ultrasound energy dose

In a recent systematic study of the dependence of 20 kHz sonophoresis on ultrasound parameters, Mitragotri et al. showed that the enhancement of skin permeability varies linearly with ultrasound intensity and ultrasound on-time (for pulsed ultrasound, ultrasound on-time equals the product of total ultrasound application time and duty cycle), while is independent of the ultrasound duty cycle. Based on those findings, the authors reported that there is a threshold energy dose for ultrasound induced transdermal drug transport. Once the threshold value is crossed, the enhancement of skin permeability varies linearly with the ultrasound energy dose (J/cm^2), which is calculated as the product of ultrasound intensity and ultrasound on-time. This result indicates that ultrasound energy dose can be used as a predictor of the effect of 20 kHz sonophoresis. The authors also indicated that it is important to determine the threshold energy dose for each individual sonophoresis system, for example, the

Table 3 Literature reports of high frequency sonophoresis

Drug	Molecular weight (Da)	Experimental system	Ultrasound conditions	Experimental conclusions ^a	Ref.
Salicylic acid	138	Hairless rat in vivo	2, 10, and 16 MHz 200 mW/cm ²	2–4 fold enhancement	(11)
Lanthanum tracers	—	Hairless rat in vivo	10 and 16 MHz 200 mW/cm ²	Significant enhancement	(12)

^aThe experimental conclusions are reported as statistically significant or insignificant enhancement.

real in vivo situation, because it may vary from system to system. Specifically, it may vary between different skin models, as well as with the ultrasound frequency and the distance of the transducer from the skin surface, etc.

PREVIOUS STUDIES OF SONOPHORESIS

Numerous attempts of sonophoresis have been performed over the last 40 years. As described earlier, these attempts can be classified into three categories: therapeutic frequency, high-frequency and low-frequency ultrasound.

Therapeutic Frequency Sonophoresis

The therapeutic ultrasound conditions correspond to a frequency in the range of 1–3 MHz and an intensity in the range of 0–2 W/cm². Therapeutic ultrasound has been attempted to enhance transdermal transport of more than 15 drugs (7–10, 14, 15, 18–22, 25, 22, 27, 28, 30–33, 36, 33, 42–45, 47–50, 55, 56, 63), a summary of which is provided in Table 2.

Historically, the transdermal route of drug administration has been considered for topical rather than systemic delivery of drugs. Accordingly, most of the sonophoresis experiments reported in Table 2 were intended for topical delivery of various drugs. Among all the drugs that have been used for sonophoresis, much attention has been focused on anti-inflammatory drugs. These include steroidal drugs such as hydrocortisone and dexamethasone and non-steroidal drugs such as indomethacin and salicylate. Sonophoresis of anti-inflammatory drugs offers an advantage over their passive topical delivery in that ultrasound may deliver drugs deeper into tissues. This is especially advantageous in the case of delivery of anti-inflammatory drugs to muscles which lie deeper into the body. Griffin et al. (18) reported that application of

ultrasound (1 MHz, 2 W/cm²) delivered hydrocortisone about 5 cm deep into pig tissues. This characteristic property of sonophoresis has been used effectively by these investigators to deliver hydrocortisone to joints for the treatment of rheumatoid arthritis.

The most commonly used technique of sonophoresis in these studies was to apply hydrocortisone in the form of an ointment on the skin and then apply ultrasound by keeping the transducer in contact with the ointment. In some cases, the transducer was moved in circular patterns to avoid a continuous exposure of a certain part of the skin to ultrasound. Although these studies were performed using different animal models, application techniques, hydrocortisone concentrations in the ointment, and exposure time, a measurable enhancement of hydrocortisone transport was reported in almost all cases. In contrast, most of the attempts to enhance transdermal transport of lidocaine and salicylates have been less successful. In the case of lidocaine, the sonophoretic enhancement was measured in terms of reduction of onset time for anesthesia or prolonging duration of anesthesia. In most cases, no significant effect of ultrasound application on either induction time or duration of anesthesia has been reported (55). Similarly sonophoresis of salicylates from ointments has not been found to induce any significant increase in plasma salicylate levels (14).

Literature data reported in Table 2 indicate that except in the case of steroids including hydrocortisone, dexamethasone, testosterone, estradiol, and corticosterone, application of therapeutic ultrasound results in either minor or no enhancement of transdermal drug transport. Mitragotri et al. (36) presented a hypothesis for this variation of sonophoretic enhancement from drug to drug based on their mechanistic conclusion that ultrasound induces disorganization of the SC lipid bilayers, thus increasing drug diffusivity and hence permeability of the SC. This mechanism suggests that drugs such as steroids which possess low passive diffusion coefficients through the SC bilayers (36) compared to those through the disordered SC bilayers should be significantly

enhanced by ultrasound application. On the other hand, drugs such as lidocaine and salicylic acid possessing a passive diffusion coefficient comparable to that through the disorganized bilayer phase (36) may not be significantly enhanced by ultrasound application. A mathematical equation was developed (39) to predict a priori whether application of therapeutic ultrasound under a typical condition, that is, 1 MHz, 2 W/cm² will enhance transdermal transport of a given drug:

$$E \approx \frac{K_{o/w}^{0.75}}{4 \times 10^4 P}$$

where E is sonophoretic enhancement of skin permeability (dimensionless), $K_{o/w}$ is the octanol–water partition coefficient, and P is the passive skin permeability (cm/h). A list of $K_{o/w}$ and P values for various drugs may be found in (64) and (65) respectively.

Over the last 20 years, transdermal route of delivery has been considered as a means for systemic drug administration. Over this period, sonophoresis has been attempted to enhance systemic transdermal delivery. Levy et al. (30) showed that 3–5 min. of ultrasound exposure (1 MHz, 1.5 W/cm²) increased transdermal permeation of mannitol and physostigmine across hairless rat skin in vivo by up to 15-fold. They also reported that the lag time typically associated with transdermal drug delivery was nearly-completely eliminated after exposure to ultrasound (30).

Although several attempts have been made to enhance transdermal drug transport using therapeutic ultrasound, a typical enhancement induced by therapeutic ultrasound is about 10-fold or smaller. This enhancement, may be sufficient for local delivery of certain drugs such as hydrocortisone, but is not sufficient for the systemic delivery of many drugs. Accordingly, despite of significant attention dedicated to sonophoresis, there is no commercially available sonophoresis system for the systemic drug delivery.

Low-Frequency Sonophoresis

Less attention has been given to sonophoresis in this ultrasound region (Table 4). Tachibana and Tachibana (51–53) reported that application of low-frequency ultrasound (48 kHz) enhances transdermal transport of lidocaine and insulin across hairless rat skin in vivo. They found that the blood glucose level of a hairless rat immersed in a beaker filled with insulin solution (20 U/ml) and placed in an ultrasound bath (48 kHz, 5000 Pa or 37 mm Hg) decreased by 50% in 240 min. (51). They also showed that application of ultrasound under similar conditions prolongs the anesthetic effect of transdermally administrated lidocaine in hairless rats (53) and enhances

transdermal transport of insulin in rabbits (52). Mitragotri et al. (37, 38) have shown that application of ultrasound at even lower frequencies (20 kHz) enhances transdermal transport of various low-molecular weight drugs including corticosterone, and salicylic acid as well as high-molecular weight proteins such as insulin, γ -interferon, and erythropoietin across the human skin in vitro (37) Table 4 provides a summary of the literature reports of low-frequency sonophoresis.

Transdermal transport enhancement induced by low-frequency ultrasound has been found to be much more significant than that induced by therapeutic ultrasound. For example, although application of therapeutic ultrasound has no effect on transdermal permeation of lidocaine, low-frequency ultrasound has been shown to significantly enhance lidocaine transport across hairless rat skin in vivo. Quantitatively, Mitragotri et al. (38) compared the enhancement ratios (ratio of the sonophoretic and passive permeabilities measured in vitro across human cadaver skin) induced by *therapeutic ultrasound* (1 MHz, 2 W/cm², continuous) (36) and *low-frequency ultrasound* (20 kHz, 125 mW/cm², 100 msec pulses applied every second) in the case of four permeants, butanol, corticosterone, salicylic acid, and sucrose. They found that the enhancement induced by low-frequency ultrasound is up to 1000-fold higher than that induced by therapeutic ultrasound (38).

Low-frequency ultrasound has been found to enhance transdermal transport of drugs which do not permeate skin passively, for example, large-molecular weight proteins. Application of low-frequency ultrasound (20 kHz, 125 mW/cm², 100 ms pulses applied every second) has been shown to enhance transdermal transport of proteins including insulin, γ -interferon, and erythropoietin across human cadaver skin in vitro (37). Figure 2 shows the variation of in vitro human skin permeability to insulin with ultrasound (20 kHz, 100 ms pulses applied every second) intensity. Skin insulin permeability increases by more than 100-fold as the ultrasound intensity increases from 12.5 mW/cm² to 125 mW/cm². It has also been shown that application of ultrasound under conditions the same as those mentioned above delivers therapeutic doses of insulin across hairless rat skin in vivo from a chamber glued on the rats back and filled with an insulin solution (100 U/ml) (37). Figure 3 shows blood glucose levels of diabetic rats during ultrasound-insulin treatment. Insulin-ultrasound treatment (20 kHz, 225 mW/cm², 100 ms pulses applied every second) reduces the blood glucose level of diabetic hairless rats from about 400–200 mg/dl (the blood glucose level of normal rats) in 30 min. A corresponding change in the plasma insulin levels was observed during sonophoresis. Normal hairless

Table 4 Literature reports of low frequency sonophoresis

Drug	Molecular weight (Da)	Experimental system	Ultrasound conditions	Experimental conclusions ^a	Ref.
Corticosterone	346	Human skin in vitro	20 kHz, 125 mW/cm ²	80	(38)
Estradiol	272	Human skin in vitro	20 kHz, 125 mW/cm ²	3	(38)
Salicylic Acid	138	Human skin in vitro	20 kHz, 125 mW/cm ²	400	(38)
Aldosterone	360	Hairless rat in vitro	20 kHz, 125 mW/cm ²	1400	(38)
Water	18	Human skin in vitro	20 kHz, 125 mW/cm ²	113	(38)
Lidocaine	234	Rat skin in vivo	48 kHz	Significant enhancement	(53)
Insulin	6000	Human skin in vitro Rat skin in vivo	20 kHz, up to 225 mW/cm ²	Significant enhancement	(52)
Insulin	6000	Rat skin	48 kHz,	Significant enhancement	(37)
γ -Interferon	17000	Human skin in vitro Rat skin in vivo	20 kHz, up to 225 mW/cm ²	Significant enhancement	(37)
Erythropoietin	48000	Human skin in vitro Rat skin in vivo	20 kHz, up to 225 mW/cm ²	Significant enhancement	(37)

^aThe experimental conclusions are reported either as statistically significant or insignificant enhancement or in terms of a quantitative ratio of sonophoretic and passive skin permeability.

rats possessed a plasma insulin level of 101 ± 31 picomolar while diabetic hairless rats possessed a value below our detection limit (34 picomolar). During sonophoresis, the levels of transdermally delivered human insulin in rat plasma reached a value of 77 (± 28) picomolar after 30 min. and a value of 178 (± 84) picomolar after 1 h (37). No significant change in the plasma concentration of indigenous rat insulin was observed during sonophoresis. These results indicate that sonophoresis at 20 kHz delivers therapeutic doses of insulin across hairless rat skin in vivo.

High-Frequency Sonophoresis

This region of ultrasound corresponds to a frequency higher than 3 MHz. Bommanan et al. (11) hypothesized that since the absorption coefficient of the skin varies directly with the ultrasound frequency, high frequency

ultrasound energy would concentrate more in the epidermis, thus leading to higher enhancements. In order to assess this hypothesis, they studied the effect of high-frequency ultrasound (2–15 MHz) on the permeability of salicylic acid (dissolved in a gel) through hairless guinea pig skin in vivo. They found that a 20 minute application of ultrasound (0.2 W/cm²) at a frequency of 2 MHz did not significantly enhance amount of salicylic acid penetrating the skin. However, 10 MHz ultrasound under otherwise same conditions resulted in about a 4-fold increase and 16 MHz ultrasound resulted in about a 2.5-fold increase in transdermal salicylic acid transport (11). They also investigated the effect of shorter (5 min.) ultrasound exposures under similar conditions on transdermal salicylic acid transport and found that while 10 MHz ultrasound enhances transdermal transport by 1.6-fold, that at 16 MHz enhances it by about 1.8-fold. Application of high-frequency ultrasound was also found to reduce the long lag time associated with transdermal

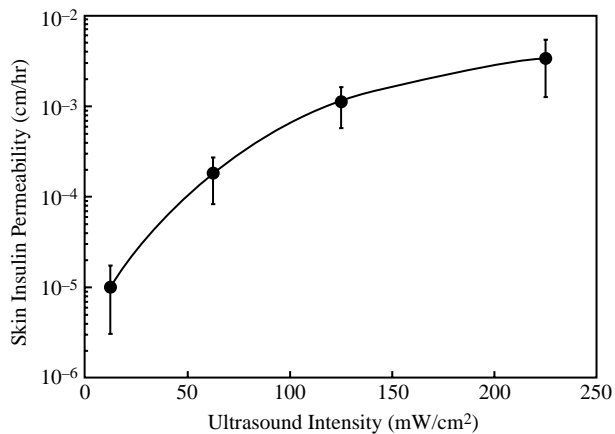


Fig. 2 Human skin permeability to insulin (37).

transport. Bommanan et al. (11) found that transdermally delivered salicylic acid appears much sooner in the urine if driven by sonophoresis than by passive permeation. These researchers also found that an electron dense tracer, such as lanthanum, was driven deep into the dermis by a 5 min. application of high-frequency ultrasound in hairless mouse *in vivo*.

MECHANISMS OF SONOPHORESIS

In order to understand the mechanisms of sonophoresis, it is important to identify various effects of ultrasound exposure on the human tissue since one or more of these effects may contribute to the mechanism of sonophoresis.

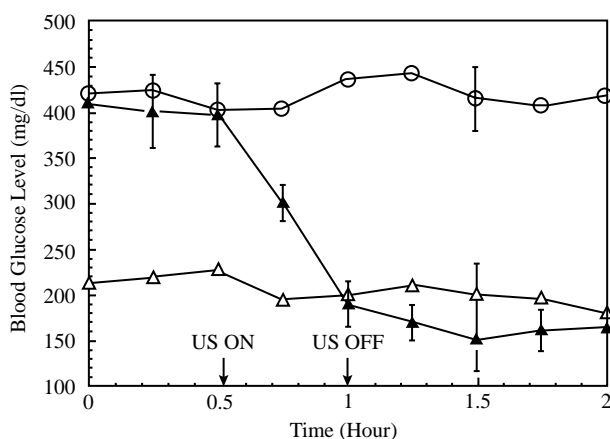


Fig. 3 Blood glucose levels of rats; ○ – diabetic hairless rats; △ – normal rats, ▲ – diabetic rats treated by sonophoresis and insulin for 30 min (37)

A brief description of the various biological effects of ultrasound is provided below.

Exposure Effects

Thermal effects

Absorption of ultrasound results in a temperature increase of the medium. Materials which possess higher ultrasound absorption coefficients, such as bones, experience severe thermal effects as compared to muscle tissues which have a lower absorption coefficient (α). α -values for several biological tissues can be found in (60,66) (Table 1). The absorption coefficient of a medium increases proportionally with the ultrasound frequency indicating that, the thermal effects of ultrasound are proportional to the ultrasound frequency. The increase in the temperature of a medium upon ultrasound exposure at a given frequency varies proportionally with the ultrasound intensity and exposure time. The thermal effects can be substantially decreased by pulsed application. For a detailed discussion of the thermal effects of ultrasound (60).

Acoustic streaming

Acoustic streaming, by definition, is the development of time independent large fluid velocities in a medium under the influence of an ultrasound wave. The primary causes of acoustic streaming are the reflections and other distortions of the wave propagation. Oscillations of cavitation bubbles may also contribute to acoustic streaming. The shear stresses developed by streaming velocities may affect the neighboring structures (67).

Cavitation effects

Cavitation is the formation of gaseous cavities in a medium upon ultrasound exposure. The primary cause of cavitation is the ultrasound-induced pressure variations in the medium. Cavitation involves either rapid growth and collapse of a bubble (transient cavitation) or slow oscillatory motion of a bubble in ultrasound field (stable cavitation). Cavitation affects tissues in several ways. Specifically, collapse of cavitation bubbles releases a shock wave which can cause structural alterations in its surroundings. Biological tissues contain numerous air pockets trapped in the fibrous structures which act as nuclei for cavitation upon ultrasound exposure. Accordingly, a significant cavitation activity is known to occur in biological tissues upon ultrasound exposure (60). The cavitation effects vary inversely with ultrasound frequency and directly with ultrasound intensity. Significant attention has been devoted to explore which of the

above mentioned phenomena plays an important role in sonophoresis.

Mechanisms of Therapeutic Sonophoresis

Mortimer et al. (42) performed sonophoresis of oxygen across frog skin *in vitro*. They found that the sonophoretic enhancement of transdermal oxygen transport depends on ultrasound intensity, rather than pressure amplitude. Based on this observation, they hypothesized that cavitation cannot be responsible for sonophoresis. They hypothesized that the observed enhancement occurs due to acoustic streaming in the solution around the skin (42). Levy et al. (30) performed an *in vitro* investigation of the roles played by thermal effects, cavitation, and mixing in sonophoretic enhancement of urea transport across polymer membranes. They found that the observed enhancement can not be explained by the thermal effects or mixing. In an attempt to elucidate the role played by cavitation, they performed sonophoresis experiments using degassed solutions. Since degassing a solution decreases the cavitation activity in the solution, they hypothesized that if a decrease in the sonophoretic enhancement is observed upon degassing, it would indicate the importance of cavitation. Indeed, they found that degassing procedure reduced the sonophoretic enhancement of urea permeation by 2-fold suggesting that cavitation may play a role in sonophoresis.

Cavitation occurs in a variety of mammalian tissues, including muscle, abdominal tissues, brain, cardiovascular tissues, and liver upon exposure to ultrasound at a variety of conditions (60). As explained earlier, the occurrence of cavitation in biological tissues is attributed to the existence of a large number of gas nuclei. These nuclei are gas pockets trapped in either intracellular or intercellular structures. Simonin et al. (68) hypothesized that cavitation occurs in the follicles of the skin upon ultrasound exposure and enhances transdermal permeation by convective velocities through follicles. However, no evidence was presented to support this hypothesis. Mitragotri et al. (36) presented results of the experiments indicating that cavitation inside the skin plays an important role in sonophoresis performed using therapeutic ultrasound.

In the first set of experiments, the known effect of static pressure on cavitation was utilized. It is known that cavitation in fluids and porous media (21) can be suppressed at high pressures. This effect is believed to occur due to the dissolution or collapse of the gaseous nuclei under the influence of pressure. Sonophoresis experiments were performed using skin compressed at

30 atm (between two smooth glass plates soaked in water placed in a compression press for two hours prior to sonophoresis experiments). They found that while application of ultrasound (1 MHz, 2 W/cm², continuous) enhances estradiol permeability of the normal human epidermis by 13-fold, the corresponding enhancement for compressed skin is only about 1.75-fold.

In the second set of experiments, the heat-stripped human cadaver skin was degassed (under a pressure of 0.05 mm Hg) prior to the permeability experiments. The authors hypothesized that when a skin piece soaked in buffer is subjected to high vacuum, the resulting low pressures should reduce the dissolved gas concentration in the buffer thereby forcing small gaseous nuclei in the skin to dissolve. When the degassed skin was exposed to ultrasound, once again, the effect of ultrasound on the estradiol permeability was minimal (1.5-fold), compared to 13-fold across the normal skin. Based on these two results, the authors concluded that cavitation inside the skin plays a major role in enhancing transdermal transport upon therapeutic ultrasound exposure. They provided the following hypothesis for the mechanism of sonophoresis performed using therapeutic ultrasound.

Ultrasound exposure in the therapeutic range causes cavitation in the keratinocytes of the stratum corneum. Oscillations of the ultrasound-induced cavitation bubbles near the keratinocyte-lipid bilayer interfaces may, in turn, cause oscillations in the lipid bilayers, thereby causing structural disorder of the SC lipids (Fig. 4). Shock waves generated by the collapse of cavitation bubbles at the interfaces may also contribute to the structure-disordering effect.

Since diffusion of permeants through a disordered bilayer phase can be significantly higher than that through a normal bilayer, transdermal transport in the presence of ultrasound is expected to be higher than passive transport.

Mechanisms of Low-Frequency Sonophoresis

Since cavitation effects in fluids vary inversely with ultrasound frequency (59), it is likely that cavitation effects should play an even more important role in low-frequency sonophoresis. Tachibana et al. (53) hypothesized that application of low-frequency ultrasound results into acoustic streaming in the hair follicles and sweat ducts of the skin, thus leading to enhanced transdermal transport. Mitragotri et al. (38) hypothesized that transdermal transport during low-frequency sonophoresis occurs across the keratinocytes rather than hair follicles. They provided

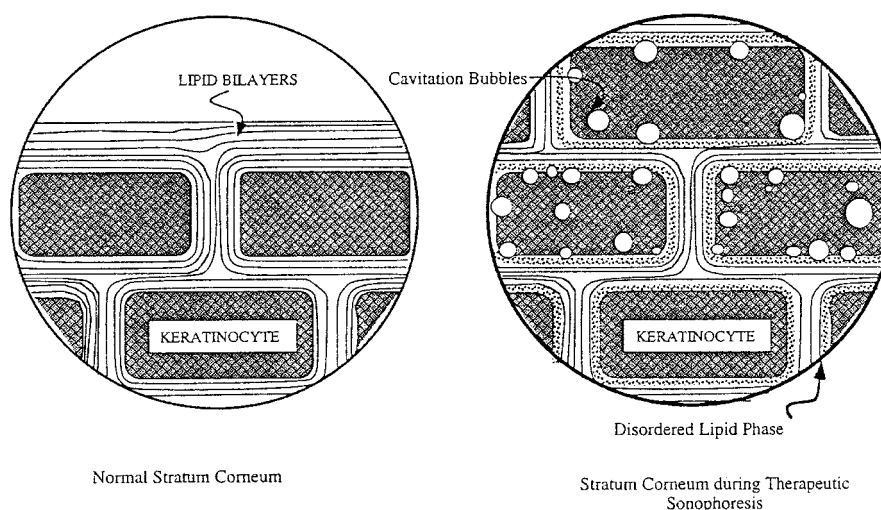


Fig. 4 Ultrasound-induced cavitation bubbles in stratum corneum.

the following hypothesis for the higher efficacy of low-frequency sonophoresis.

Cavitation induced by low-frequency ultrasound may cause disordering of the SC lipids. In addition, oscillations of cavitation bubbles may result in significant water penetration into the disordered lipid regions. This may cause the formation of aqueous channels through the intercellular lipids of the SC through which permeants may transport (Fig. 5). The occurrence of transdermal transport through aqueous channels across the disordered lipid regions may enhance transdermal transport as compared to passive transport because i) the diffusion coefficients of permeants through water, which is likely to primarily occupy the channels generated by ultrasound, are up to 1000-fold higher than those through the ordered lipid bilayers (38), and ii) the transport path length of these aqueous channels may be much shorter (by a factor up to 25 (69)) than that through the tortuous intercellular lipids in the case of passive transport.

This hypothesis also explains why low-frequency ultrasound can induce transdermal transport of drugs which exhibit very low passive transport. Drugs possessing low passive permeabilities are either i) hydrophilic, which makes their partitioning into the SC bilayers difficult, or ii) large in molecular size (for example, proteins), which reduces their diffusion coefficients in the SC. Low-frequency ultrasound may overcome both of these limitations by providing aqueous transport channels across the skin. Since these channels are filled with saline, hydrophilic drugs can easily partition into the SC. In addition, diffusion of drugs through water is much faster

than that through ordered lipid bilayer regions, thus allowing drugs to transport across the skin at a faster rate. Therefore, molecules such as hydrophilic drugs or proteins, may permeate skin with relative ease in the presence of low-frequency ultrasound.

Mechanisms of High-Frequency Sonophoresis

Bommanan et al. (12) performed sonophoresis of lanthanum tracers across hairless mice skin at an ultrasound frequency of 16 MHz in order to understand the transport pathways during high-frequency sonophoresis. They observed the skin under the electron microscope after sonophoresis and found that 5 min. of sonophoresis results in penetration of lanthanum tracers to dermal levels of the skin. They further reported that the tracer was patchily distributed within the intercellular lipid bilayers of the SC. They provided the following hypothesis for the mechanism of high-frequency sonophoresis. The micronuclei (air-pockets) present in the SC oscillate in response to oscillating pressure fields of ultrasound and eventually collapse. The oscillations of these bubbles result in enhanced skin permeation. They also hypothesized that the patchy distribution of the lanthanum tracer revealed in the micrographs corresponds to the location of oscillating air pockets in the SC. In a later report, Menon et al. (35) presented additional microscopic studies of the hairless mice skin after undergoing sonophoresis of lanthanum tracer. They reported the presence of long confluent channels in the intercellular lipids filled with lanthanum tracers in the hairless rat skin exposed to ultrasound. They

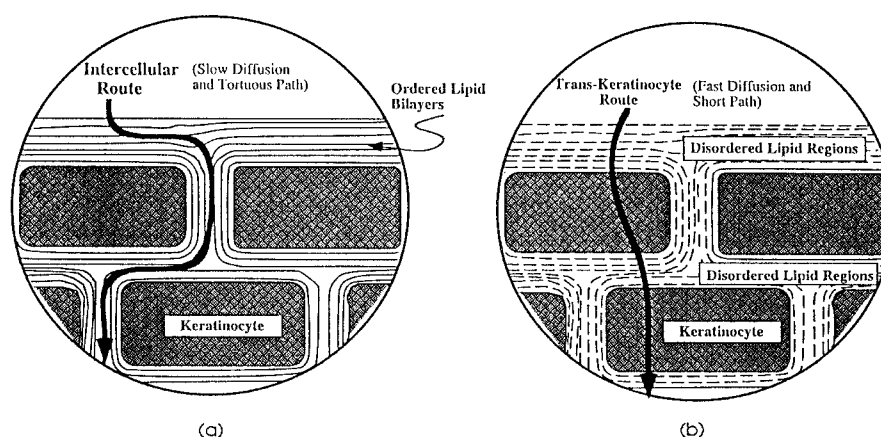


Fig. 5 Transdermal transport through the stratum corneum. (a) Passive and (b) during low-frequency sonophoresis.

presented the following hypothesis for the mechanism of sonophoresis: application of ultrasound opens and expands gas-filled cavities in the SC much like pumping air through a collapsed rubber tubing. Enhanced transport of drugs may then occur through these confluent channels across the SC.

SAFETY

The safety aspects of sonophoresis involve the reversibility of the skin barrier properties after turning ultrasound off, and the effect of ultrasound on the living parts of the skin and underlying tissues. Many reports exist in literature describing preliminary assessments of sonophoresis with respect to these two issues.

Recovery of the Skin Barrier Properties after Sonophoresis

Numerous reports exist to suggest that application of therapeutic ultrasound (1–3 MHz, 0–2 W/cm²) does not induce any irreversible change in the skin permeability to drugs *in vivo*. Quantitative measurements of estradiol transport across human skin (*in vitro*) have also shown that application of therapeutic ultrasound (1 MHz, 2 W/cm²) does not induce any statistically significant irreversible change in skin barrier properties (36). Similar studies have also been performed using very low-frequency ultrasound (20 kHz, 125 mW/cm², 100 ms pulses applied every second) to assess whether application of low-frequency ultrasound results in any permanent loss of the barrier properties of skin measured in terms of water permeability (38). It has been found that in the case of a 1 h long

ultrasound exposure, the skin permeability to water measured within 2 h post-exposure was comparable to the passive skin permeability to water. In the case of a 5-h long ultrasound exposure, the skin permeability 2 h post-exposure was about 6 times higher than the passive permeability to water. However, this value continued to decrease, and was within a factor of 2 of the passive skin water permeability 12 h post-exposure. Studies have also been performed (35) to assess whether application of high-frequency ultrasound induces any irreversible damage to the barrier properties of the skin measured in terms of trans-epidermal water loss (TEWL) across hairless mice skin exposed to high-frequency ultrasound (16 MHz). No significant difference in TEWL values of the skin exposed to ultrasound and that not exposed to ultrasound was found (35).

Biological Effects of Low-Frequency Ultrasound

Ultrasound over a wide frequency range has been used in medicine over last century. For example, therapeutic ultrasound (1–3 MHz) has been used for massage, low-frequency ultrasound has been used in dentistry (23–40 kHz) (70, 71), and high-frequency ultrasound (3–10 MHz) has been used for diagnostic purposes (66). In view of this, significant attention has been dedicated to investigate the effects of ultrasound on biological tissues. However, no conclusions have been reached regarding the limiting ultrasound conditions required to ensure safe exposure.

As described earlier, ultrasound affects biological tissues via three main effects, thermal effects, cavitation effects, and acoustic streaming. Conditions under which these effects become critical are given below (60).

Thermal effects may be important when

1. The tissue has a high protein content.
2. A high intensity continuous wave ultrasound is used.
3. Bone is included in the heated volume.
4. Vascularization is poor.

Cavitation may be important when

1. Low-frequency ultrasound is used.
2. Gassy fluids are exposed.
3. Small gas-filled spaces are exposed.
4. The tissue temperature is higher than normal.

Streaming may be important when

1. The medium has an acoustic impedance different from its surroundings.
2. The fluid in the biological medium is free to move.
3. Continuous wave application is used.

Numerous investigators have performed histological studies of animal and human skin exposed to ultrasound under various conditions in order to assess the effect of ultrasound on living skin cells. Levy et al. (30) exposed in vivo hairless rat skin to therapeutic ultrasound (1 MHz, 1.5 W/cm² continuous wave or 3 W/cm² pulsed wave, 3–5 min) and reported no damage to the skin. Nevertheless, Machet et al. (72) reported both epidermal and dermal alteration of human and hairless mouse skin when exposed to therapeutic ultrasound in vitro (3.3 MHz, 3 W/cm², 10 min). The authors proposed that a combined cavitation and thermal effect could explain these cellular lesions. Due to the higher likelihood of occurrence of cavitation under low-frequency ultrasound, the effect of low-frequency ultrasound on skin histology has drawn the most attention in the recent years. Tachibana et al. (52) exposed rabbit skin to 105 kHz ultrasound (5000 Pa, 5-s pulses applied at 5-s intervals, for 90 min) and reported no inflammation or destruction of skin. Mitragotri et al. (37) performed histological studies of hairless rat skin exposed to 20 kHz ultrasound (125 mW/cm², 100-ms pulses applied every second, for 1 h) and found no damage to the epidermis and underlying living tissues. In a more recent studies on skin microscopy, Yamashita et al. (54) exposed both human skin in vitro and hairless mouse skin in vivo to 48 kHz ultrasound at 0.5 W/cm² for 5 min, and reported that there was slight and complete removal of keratinocytes from human skin and hairless mouse skin, respectively. From the above reports, we can deduce that the effect of ultrasound, especially in the low-frequency range, on living skin cells strongly depends on the ultrasound parameters, i.e., frequency, intensity, exposure time and duty cycle. Further research focusing on safety

issues is required to evaluate limiting ultrasound parameters for safe exposure.

CONCLUSIONS

Application of ultrasound enhances transdermal drug transport, a phenomenon referred to as sonophoresis. Proper choice of ultrasound parameters including ultrasound energy dose, frequency, intensity, pulse length, and distance of transducer from the skin is critical for efficient sonophoresis. The numerous attempts made over the last 50 years can be classified into three categories: therapeutic frequency, high-frequency and low-frequency ultrasound; the first represents the most commonly used ultrasound condition for sonophoresis, although recently, attention has been more focused on low- and high-frequency conditions. Mechanistic experiments performed by several investigators suggest that cavitation plays a major role. It has been suggested that cavitation disorganizes the lipid bilayers of the skin through which enhanced transport of drugs may occur. Various studies have indicated that application of ultrasound under conditions used for sonophoresis does not cause any permanent damage to the skin or underlying tissues, although more work is required before arriving at definite conclusions regarding the safety of ultrasound exposure.

REFERENCES

1. Jarrett, A. *The Physiology and Pathology of the Skin*; Ed.; Academic Press: London, 1978.
2. Walters, K.A. Penetration Enhancers and Their Use in Transdermal Therapeutic Systems. *Penetration Enhancers and their Use in Transdermal Therapeutic Systems*; Hadgraft, J., Guy, R.H., Eds.; Marcel Dekker, Inc.: New York, 1989; 197–233.
3. Junginger, H.E.; Bodde, H.E.; de Haan, F.H.N. Penetration. In *Visualization of Drug Transport across Human Skin and the Influence of Penetration Enhancers*; Hsieh, D.S., Ed.; Marcel Dekker, Inc.: New York, 1994; 59–90.
4. Bommanon, D.; Tamada, J.; Leung, L.; Potts, R. Effects of Electroporation on Transdermal Iontophoretic Delivery of Leutinizing Hormone Releasing Hormone. *Pharm. Res.* **1994**, *11*, 1809–1814.
5. Prausnitz, M.R.; Bose, V.; Langer, R.; Weaver, J.C. Electroporation of Mammalian Skin: Mechanism to Enhance Transdermal Drug Delivery. *Proc. Natl. Acad. Sci.* **1993**, *90*, 10504–10508.
6. Burnette, R.R. Iontophoresis. In *Transdermal Drug Delivery: Development Issues and Research Initiatives*; Hadgraft, J., Guy, R.H., Eds.; Marcel Dekker, Inc.: New York, 1989; 247–291.

7. Antich, T.J. Phonophoresis: The Principles of the Ultrasonic Driving Force and Efficacy in Treatment of Common Orthopedic Diagnoses. *J. Orth. Sports Phys. Ther.* **1982**, *4*, 99–102.
8. Benson, H.A.E.; McElnay, J.C.; Harland, R. Phonophoresis of Lingocaine and Prilocaine from Emla Cream. *Int. J. Pharm.* **1988**, *44*, 65–69.
9. Benson, H.A.E.; McElnay, J.C.R.H. Use of Ultrasound to Enhance Percutaneous Absorption of Benzydamine. *Phys. Ther.* **1989**, *69*, 113–118.
10. Benson, H.A.E.; McElnay, J.C.J.H. Influence of Ultrasound on the Percutaneous Absorption of Nicotinate Esters. *Pharm. Res.* **1991**, *9*, 1279–1283.
11. Bommannan, D.; Okuyama, H.; Stauffer, P.; Guy, R.H. Sonophoresis. I. The Use of High-Frequency Ultrasound to Enhance Transdermal Drug Delivery. *Pharm. Res.* **1992**, *9*, 559–564.
12. Bommannan, D.; Menon, G.K.; Okuyama, H.; Elias, P.M.; Guy, R.H. Sonophoresis. II. Examination of the Mechanism(s) of Ultrasound-Enhanced Transdermal Drug Delivery. *Pharm. Res.* **1992**, *9*, 1043–1047.
13. Brondolo, W. *Arch. Orthop.* **1960**, *73*, 532–540.
14. Ciccone, C.D.; Leggin, B.Q.; Callamaro, J.J. Effects of Ultrasound and Trolamine Salicylate Phonophoresis on Delayed-Onset Muscle Soreness. *Phys. Ther.* **1991**, *71*, 666–678.
15. Davick, J.P.; Martin, R.K.; Albright, J.P. Distribution and Deposition of Tritiated Cortisol Using Phonophoresis. *Phys. Ther.* **1988**, *68*, 1672–1675.
16. Fellingner, K.; Schmidt, J. *Klinik und Therapie des Chronischen Gelenkreumatismus*, Maudrich, Vienna, Austria **1954**, 549–554.
17. Fogler, S.; Lund, K. Acoustically Augmented Diffusional Transport. *J. Acous. Soc. Am.* **1973**, *53*, 59–64.
18. Griffin, J.E.; Touchstone, J. Ultrasonic Movement of Cortisol into Pig Tissue. *Am. J. Phys. Med.* **1965**, *44*, 20–25.
19. Griffin, J.E. Physiological Effects of Ultrasonic Energy as it is Used Clinically. *J. Am. Phys. Ther. Assoc.* **1966**, *46*, 18–26.
20. Griffin, J.E.; Echternach, J.L.; Proce, R.E.; Touchstone, J.C. Patients Treated with Ultrasonic Driven Hydrocortisone and with Ultrasound Alone. *Phys. Ther.* **1967**, *47*, 600–601.
21. Griffin, J.E.; Touchstone, J.C. Low-Intensity Phonophoresis of Cortisol in Swine. *Phys. Ther.* **1968**, *48*, 1136–1144.
22. Griffin, J.E.; Touchstone, J.C. Effects of Ultrasonic Frequency on Phonophoresis of Cortisol into Swine Tissues. *Am. J. Phys. Med.* **1972**, *51*, 62–78.
23. Julian, T.N.; Zentner, G. Ultrasonically Mediated Solute Permeation through Polymer Barriers. *J. Pharm. Pharmacol.* **1986**, *38*, 871–877.
24. Julian, T.N.; Zentener, G.M. Mechanism for Ultrasonically Enhanced Transmembrane Solute Permeation. *J. Control. Rel.* **1990**, *12*, 77–85.
25. Kleinkort, J.A.; Wood, F. Phonophoresis with 1 Percent Versus 10 Percent Hydrocortisone. *Phys. Ther.* **1975**, *55*, 1320–1324.
26. Kost, J.; Levy, D.; Langer, R. Ultrasound as a Transdermal Enhancer. *Percutaneous Absorption Mechanisms-Methodology-Drug Delivery*; Bronaugh, R., Maibach, H.I. Eds.; Marcel Dekker, Inc.: New York, 1989; 595–601.
27. Kost, J.; Langer, R. Ultrasound-Mediated Transdermal Drug Delivery. *Topical Drug Bioavailability, Bioequivalence, and Penetration*; Shah, V.P., Maibach, H.I., Eds.; Plenum: New York, 1993; 91–103.
28. Kost, J. Ultrasound for Controlled Delivery of Therapeutics. *Clin. Mater.* **1993**, *13*, 155–161.
29. Lenart, I.; Auslander, D. The Effects of Ultrasound on Diffusion through Membranes. *Ultrasonics* **September 1980**, 216–217.
30. Levy, D.; Kost, J.; Meshulam, Y.; Langer, R. Effect of Ultrasound on Transdermal Drug Delivery to Rats and Guinea Pigs. *J. Clin. Invest.* **1989**, *83*, 2974–2078.
31. Machluf, M.; Kost, J. Ultrasonically Enhanced Transdermal Drug Delivery. Experimental Approaches to Elucidate the Mechanism. *J. Biomat. Sci.* **1993**, *5*, 147–156.
32. McElnay, J.C.; Matthews, M.P.; Harland, R.; McCafferty, D.F. The Effect of Ultrasound on the Percutaneous Absorption of Lingocaine. *Br. J. Clin. Pharmacol.* **1985**, *20*, 421–424.
33. McElnay, J.C.; Kennedy, T.A.R.H. The Influence of Ultrasound on the Percutaneous Absorption of Fluocinolone Acetonide. *Int. J. Pharm.* **1987**, *40*, 105–110.
34. McElnay, J.C.; Benson, H.A.E.; Harland, R.; Hadgraft, J. Phonophoresis of Methyl Nicotinate: A Preliminary Study to Elucidate the Mechanism. *Pharm. Res.* **1993**, *4*, 1726–1731.
35. Menon, G.; Bommanon, D.; Elias, P. High-Frequency Sonophoresis: Permeation Pathways and Structural Basis for Enhanced Permeability. *Skin Pharmacol.* **1994**, *7*, 130–139.
36. Mitragotri, S.; Edwards, D.; Blankschtein, D.; Langer, R. A Mechanistic Study of Ultrasonically Enhanced Transdermal Drug Delivery. *J. Pharm. Sci.* **1995**, *84*, 697–706.
37. Mitragotri, S.; Blankschtein, D.; Langer, R. Ultrasound-Mediated Transdermal Protein Delivery. *Science* **1995**, *269*, 850–853.
38. Mitragotri, S.; Blankschtein, D.; Langer, R. Transdermal Drug Delivery Using Low-Frequency Sonophoresis. *Pharm. Res.* **1996**, *13*, 411–420.
39. Mitragotri, S.; Blankschtein, D.; Langer, R. An Explanation for the Variation of the Sonophoretic Transdermal Transport Enhancement from Drug to Drug. *J. Pharm. Sci.* **1997**, *86*, 1190–1192.
40. Mitragotri, S.; Farrell, J.; Tang, H.; Terahara, T.; Kost, J.; Langer, R. Determination of Threshold Energy Dose for Ultrasound-Induced Transdermal Drug Transport. *J. Control. Rel.* **2000**, *63*, 41–52.
41. Miyzaki, S.; Mizuoka, O.; Takada, M. External Control of Drug Release and Penetration: Enhancement of the Transdermal Absorption of Indomethacin by Ultrasound Irradiation. *J. Pharm. Pharmacol.* **1990**, *43*, 115–116.
42. Mortimer, A.J.; Trollope, B.J.; Roy, O.Z. Ultrasound-Enhanced Diffusion through Isolated Frog Skin. *Ultrasonics* **1988**, *26*, 348–351.
43. Newman, J.T.; Nellermo, M.D.; Crnett, J.L. Hydrocortisone Phonophoresis: A Literature Review. *J. Am. Pod. Med. Assoc.* **1992**, *82*, 432–435.
44. Novak, E.J. Experimental Transmission of Lidocaine through Intact Skin by Ultrasound. *Arch. Phys. Med. Rehab.* **May 1964**, 231–232.
45. Oziomek, R.S.; Perrin, D.H.; Herold, D.A.; Denegar, C.R. Effect of Phonophoresis on Serum Salicylate Levels. *Med. Sci. in Sports and Exer.* **1990**, *23*, 397–401.

46. Policoff, L.D. Effective Use of Physical Modalities. *Orthopedic Clinics of North America* **1982**, *13*, 579–586.
47. Pottenger, J.F.; Karalfa, L.B. Utilization of Hydrocortisone Phonophoresis in United States Army Physical Therapy Clinics. *Milit. Med.* **1989**, *154*, 355–358.
48. Pratzel, H.; Ditrich, P.; Kukovetz, W. Spontaneous and Forced Cutaneous Absorption of Indomethacin in Pigs and Humans. *J. Rheumat.* **1986**, *13*, 1122–1125.
49. Quillen, W.S. Phonophoresis: A Review of the Literature and Technique. *Athlet. Train.* **1980**, *15*, 109–110.
50. Skauen, D.M.; Zentner, G.M. Phonophoresis. *Int. J. Pharm.* **1984**, *20*, 235–245.
51. Tachibana, K.; Tachibana, S. Transdermal Delivery of Insulin by Ultrasonic Vibration. *J. Pharm. Pharmacol.* **1991**, *43*, 270–271.
52. Tachibana, K. Transdermal Delivery of Insulin to Alloxan-Diabetic Rabbits by Ultrasound Exposure. *Pharm. Res.* **1992**, *9*, 952–954.
53. Tachibana, K.; Tachibana, S. Use of Ultrasound to Enhance the Local Anesthetic Effect of Topically Applied Aqueous Lidocaine. *Anesthesiology* **1993**, *78*, 1091–1096.
54. Yamashita, N.; Tachibana, K.; Ogawa, K. Scanning Electron Microscopy Evaluation of the Skin Surface after Ultrasound Exposure. *Anat. Rec.* **1997**, *247*, 455–461.
55. Williams, A.R. Phonophoresis: An In Vivo Evaluation Using Three Topical Anaesthetic Preparations. *Ultrasonics* **1990**, *28*, 137–141.
56. Wing, M. Phonophoresis with Hydrocortisone in the Treatment of Temporomandibular Joint Dysfunction. *Phys. Ther.* **1981**, *62*, 32–33.
57. Kinsler, L.W.; Frey, A.R.; Coppens, A.B.; Sandes, J.V. *Fundamentals of Acoustics*; John Wiley & Sons: New York, 1982.
58. Hueter, T.F.; Bolt, R.H. *Sonics: Techniques for the Use of Sound and Ultrasound in Engineering and Science*; John Wiley & Sons: New York, 1962.
59. Gaertner, W. Frequency Dependence of Acoustic Cavitation. *J. Acoust. Soc. Am.* **1954**, *26*, 977–80.
60. Suslick, K.S. *Ultrasound: Its Chemical, Physical and Biological Effects*; VCH Publishers: New York, 1989.
61. Crum, L.A.; Folwkes, J.B. Acoustic Cavitation Generated by Microsecond Pulses of Ultrasound. *Nature* **1986**, *321*, 319.
62. Ciaravino, V.; Flynn, H.G.; Miller, M.W. Pulsed Enhancement of Acoustic Cavitation: A Postulated Model. *Ultrasound Med. Biol.* **1981**, *7*, 159–166.
63. Tyle, P.; Agrawala, P. Drug Delivery by Phonophoresis. *Pharm. Res.* **1989**, *6*, 355–360.
64. Hanch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biological Sciences*; Wiley: New York, 1979.
65. Flynn, G.L. Physiochemical Determinants of Skin Absorption. *Physiochemical Determinants of Skin Absorption*; Gerrity, T.R., Henry, C.J., Eds.; Elsevier: New York, 1990; 93–127.
66. Wells, P.N.T. *Biomedical Applications of Ultrasound*; Plenum Press: New York, 1977.
67. Nyborg, W.L. Acoustic Streaming. *Acoustic Streaming*; Mason, W.P., Ed.; Academic Press: New York, 1965; II B, 265–331.
68. Simonin, J.P. On the Mechanisms of In Vitro and In Vivo Phonophoresis. *J. Control. Rel.* **1995**, *33*, 125–141.
69. Edwards, D.; Langer, R. A Linear Theory of Transdermal Transport Phenomena. *J. Pharm. Sci.* **1994**, *83*, 1315–1334.
70. Walmsley, A.D. Applications of Ultrasound in Dentistry. *Ultrasound Med. Biol.* **1988**, *14*, 7–14.
71. Walmsley, A.D. Potential Hazards of the Dental Ultrasonic Descaler. *Ultrasound Med. Biol.* **1988**, *14*, 15–20.
72. Machet, L.; Pinton, J.; Patat, F.; Arbeille, B.; Pourcelot, L.; Vaillant, L. In Vitro Phonophoresis of Digoxin across Hairless Mice and Human Skin: Thermal Effect of Ultrasound. *Int. J. Pharm.* **1996**, *133*, 39–45.

ULTRASONIC NEBULIZERS

Kevin M.G. Taylor

University of London, London, United Kingdom

Orla McCallion

Vandsons Research, Islington, London, United Kingdom

INTRODUCTION

Ultrasonic nebulizers use ultrasonic energy to convert liquid, usually an aqueous solution, into an aerosol for inhalation. They are used to deliver β_2 -agonists, corticosteroids, antiallergics, anticholinergics, and antiviral and mucolytic agents to the respiratory tract (1). Recent innovations have increased the popularity of nebulizers, both ultrasonic and air-jet, and new devices with improved portability, compared with traditional models, are capable of generating aerosols with high respirable fractions (deep lung delivery) with a high drug output. Nebulized drugs may be inhaled during normal tidal breathing through a mouthpiece or face mask, permitting their use for patients, such as the hospitalized, the elderly, children, and patients with arthritis, who experience difficulties with other devices. Nebulizers represent ideal delivery systems for drugs that cannot be conveniently formulated into pressurized metered-dose inhalers (pMDIs) or dry powder inhalers (DPIs) or when the therapeutic dose is too large for delivery with these systems.

Ultrasonic nebulizers were first developed in the 1960s and were initially used for air humidification in respiratory care units (2). Generally, ultrasonic nebulizers have higher mass outputs than do air-jet nebulizers, in which compressed gas is used as a means of generating the aerosol. However, this is achieved at the expense of a large aerosol droplet size (3–7). Consequently, the majority of ultrasonically generated aerosols have a mean droplet size considered unsuitable for efficient targeting of a drug to the alveolar region of the lung. However, the size distribution of droplets within aerosols produced by ultrasonic nebulization is generally less than that from air-jet nebulizers. Nevertheless, the generated aerosols are polydisperse in nature, and as with air-jet nebulizers, they require baffles to remove the larger droplets from the emitted aerosol. Because ultrasonic nebulizers have a high mass output, with increased droplet concentration, which is independent of the airflow, the duration of nebulization

is shorter than with air-jet nebulizers. This, together with their quieter mode of operation, may make them particularly attractive to patients. Ultrasonic nebulizers are reported to produce bronchodilator responses comparable with those of air-jet nebulizers and pMDIs (8).

MECHANISM OF AEROSOL GENERATION

The energy required to atomize a liquid is produced by a piezoelectric crystal transducer, usually a synthetic ceramic material, vibrating at a high frequency (1–3 MHz). When an alternating electric current is applied, the crystal shrinks and expands and the resultant vibrations are transmitted to the nebulized fluid, either directly or via a coupling liquid, usually water. A fountain of liquid is produced at the liquid surface, with large droplets being emitted from its apex and a “fog” of small droplets being produced from the lower part.

Fig. 1 illustrates the two mechanisms proposed for the processes of liquid disintegration and aerosol generation within ultrasonic nebulizers (9). The capillary-wave theory relates to the production of capillary waves in the bulk liquid. These waves constructively interfere to form peaks and a central geyser. When the amplitude of the applied energy is sufficiently high, the crests of the capillary waves break off, and droplets are formed. The rate of generation of capillary waves is dependent on both the physicochemical properties of the nebulized fluid and the intensity of the ultrasonic vibration. Mercer (3) used Eq. 1 to calculate the threshold amplitude for the generation of capillary waves:

$$A = 4 \frac{\nu}{f\lambda} \quad (1)$$

where A is the threshold amplitude, ν is the kinematic viscosity of the liquid, f is the acoustic frequency, and λ the capillary wavelength. When the amplitude exceeds the threshold value by a factor of approximately four, droplets are formed. Lang (10) noted that the mean droplet size generated from thin liquid layers was proportional to

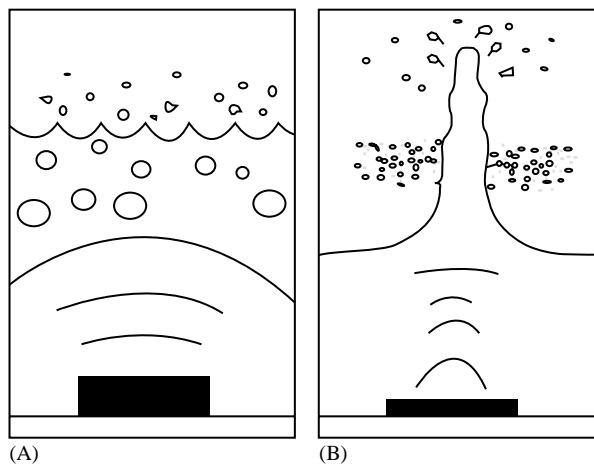


Fig. 1 Proposed ultrasonic nebulizer aerosolization mechanisms. (A) Cavitation bubble formation at low frequency; (B) capillary-wave formation at high frequency. (From Ref. 9.)

the capillary wavelength on the liquid surface. Using an experimentally determined factor of 0.34, the droplet diameter D is given by Eq. 2;

$$D = 0.34\lambda \quad (2)$$

where D is the number median diameter and λ is the capillary wavelength. Lobdell (11) concurred with these findings and calculated a theoretical value of 0.36 for the proportionality constant. The capillary wavelength can be calculated from Kelvin's equation, as in Eq. 3:

$$\lambda = (8\pi\gamma/\rho f^2)^{1/3} \quad (3)$$

where γ is the surface tension, ρ is the density, and f the acoustic frequency. When γ is in mN/m (dyne/cm), ρ is in g/cm³, and f is in megacycles/s, then D is given in micrometers. Good correlation exists between calculated and experimentally derived values (12).

The second mechanism proposed for aerosol generation is based on the piezoelectric crystal operating at low frequency and imparting vibrations to the bulk liquid. This results in the formation of cavitation bubbles, which move to the air-liquid interface (13). The internal pressure within the bubbles equilibrates with that of the atmosphere, causing their implosion. When this occurs at the liquid surface, portions of the liquid break free from the turbulent bulk liquid, resulting in droplet formation. The dependence of atomization on cavitation phenomena has been demonstrated for frequencies between 0.5 and 2.0 MHz (14, 15). Boguslavskii and Eknadiosyants (16) combined these theories with their proposal that droplet formation resulted from capillary waves initiated and driven by cavitation bubbles.

NEBULIZER DESIGN

Ultrasonic nebulizers exist in a number of basic designs that differ in the configuration of the piezoelectric crystal transducer, nebulizer chamber, baffles, and auxiliary airflow systems (Fig. 2) (9). Once the aerosol cloud is generated from the nebulizer fluid, it is transferred from the chamber and made available to the patient. Ultrasonic nebulizers produce a large number of droplets per unit volume, which tend to aggregate and settle in the absence of air circulating through the device. Larger droplets impact on the baffles or internal surfaces to return to the reservoir surface for recirculation under the influence of gravity. Smaller droplets leave the device aided by an internal fan (e.g., Medix Electronic[®], Easimist[®]) or by entrainment into the inspiratory flow of the patient (e.g., DeVilbiss Pulmosonic[®]). Air velocity over the reservoir surface may be modified by fan speed (and flow constrictors), thereby influencing both droplet size and aerosol output rate. For instance, changing the fan speed

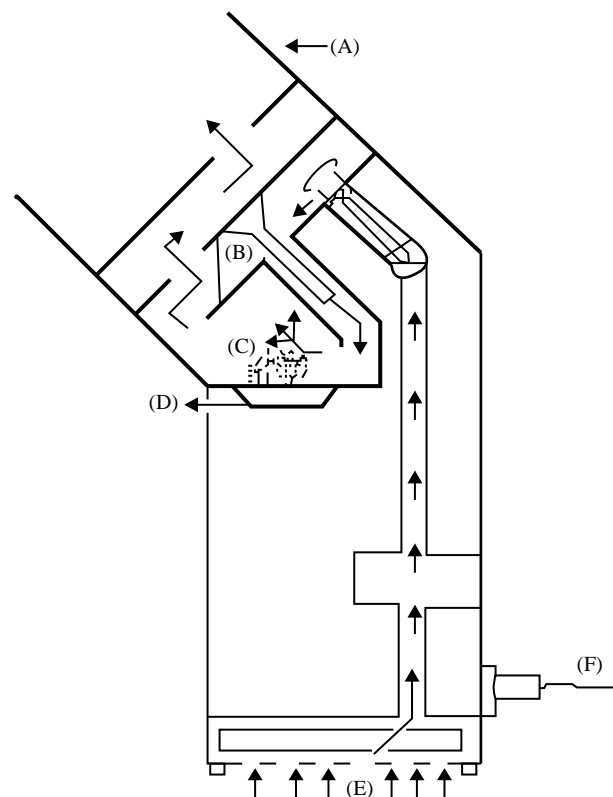


Fig. 2 Schematic diagram of a typical ultrasonic nebulizer. (A) Face mask or mouthpiece; (B) baffles; (C) geyser of respiratory solution or suspension; (D) piezoelectric crystal; (E) internal fan; (F) battery or electrical source. (From Ref. 9.)

Table 1 Design features of ultrasonic nebulizers determining particle size distribution and mass output

Design features	Nebulizer characteristics
Piezoelectric crystal	Frequency of vibration Amplitude of vibration Surface morphology (flat or curved) Coupling between crystal and fluid
Fluid reservoir	Size Shape Baffles
Auxiliary air flow	Velocity

(Modified from Ref. 7.)

setting of the Sonix 2000[®] (Medix) ultrasonic nebulizer can vary flow output between 1 ml per min and 1 ml per 6 min. The important design features influencing mass output and the particle size of the generated aerosols are summarized in Table 1.

Whereas air-jet nebulizers are usually disposable or sterilizable, ultrasonic nebulizers are too expensive to be produced as disposable units and are thus used repeatedly (17), running the risk of bacterial contamination (18). Cleaning nebulizers and connecting tubing is difficult, and the transfer of Gram-negative bacteria between nebulization equipment and patients has been reported (19).

Although most ultrasonic devices share a basic design, some novel devices have been developed. The Respimat[®] (Boehringer Ingelheim) offered direct delivery of a metered dose from a valve similar to that used in pMDIs to the surface of a vibrating crystal (2.5 MHz). Although the mass median aerodynamic diameter (MMAD) of the aerosol generated was closed to 10 μm , generally considered too large for effective inhalation therapy, equivalence to an equivalent pMDI formulation was claimed (20).

The Bepak Piezo Electric Actuator[®] is a novel aerosol delivery system based on a piezoelectric crystal combined with an electroformed mesh (Fig. 3). It produces droplets of “adjustable” size from a single metered drop or fluid reservoir (21). The mesh hole dimension (as small as 3 μm) determines the size of the droplets produced, whereas the size and density of the holes control the rate of fluid delivery. These can be varied according to the formulation. Although solutions are more readily nebulized, suspensions can be aerosolized if the particle size of the

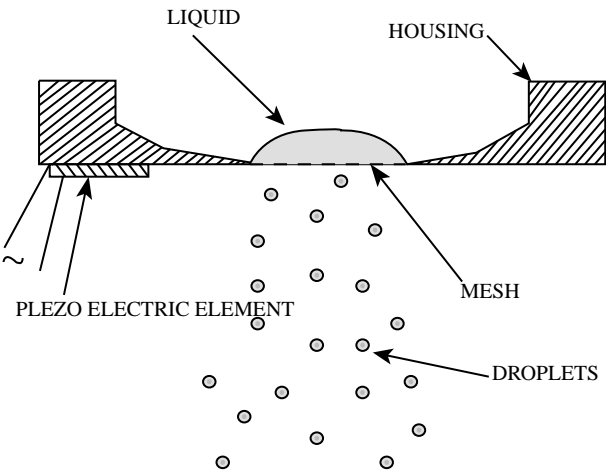


Fig. 3 Schematic diagram of the sprayhead of the Bepak Piezo Electric Actuator. (Reproduced courtesy of Bepak plc, UK.)

suspended particles is two to three times smaller than the mesh size.

More recently, The Technology Partnership has created several new forms of a perforated mesh atomizer known as TouchSpray[®] (22). TouchSpray devices deliver suspensions without requiring these to be formulated with small suspended particles. Furthermore, they can produce small droplets from large mesh holes. For example, TouchSpray devices can deliver 5 μm droplets of suspensions containing normally micronized drug suspensions, which typically include particles of 3 μm or larger. These droplets can be produced by devices having mesh hole sizes of 15 μm . TouchSpray allows small droplets to be produced from solutions. A review of the development of perforated mesh devices is given in Humberstone et al. (23). An inhalation droplet spray plume produced by the TouchSpray is shown in Fig. 4.

Future devices are likely to be hybrids of ultrasonic nebulizer technology and pMDI or DPI technology. Vibrating transducer or grids, incorporated into pMDIs, could break down large propellant droplets that would otherwise be wasted into a more respirable size range. Alternatively, the dependence of DPIs on patient-generated turbulent airstreams could be minimized by deaggregating powder by vibrational elements within the inhalation device or by generating aerosols electrostatically (Spiros[®], Dura).

AEROSOL DROPLET SIZE

The efficacy of a therapeutic inhalation aerosol depends on its ability to penetrate the respiratory tract. This is

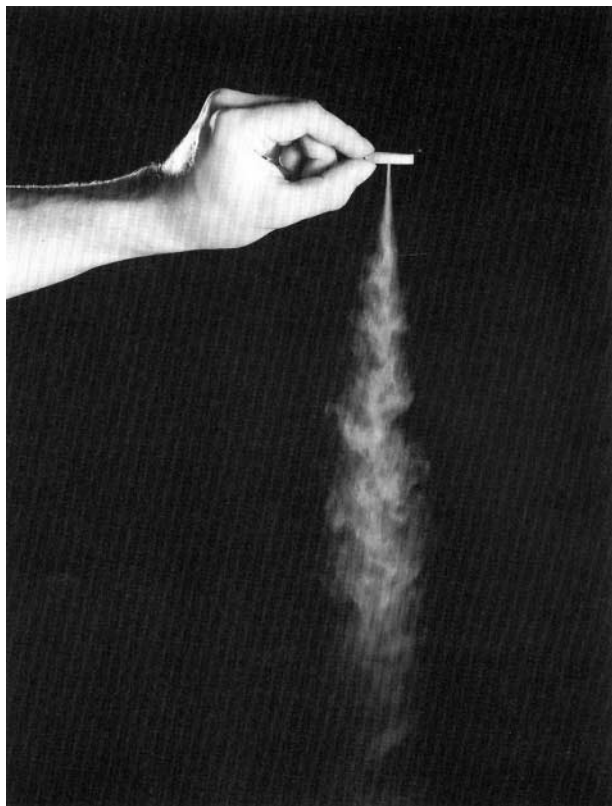


Fig. 4 Inhalation droplet spray plume produced by TouchSpray device. (Reproduced courtesy of The Technology Partnership plc, UK.)

primarily dependent on the particle size of the particles or droplets. To penetrate to the peripheral (respiratory) airways, aerosols generally require a size less than approximately $5\text{--}6\text{ }\mu\text{m}$, with a size less than $2\text{ }\mu\text{m}$ being optimal for alveolar deposition (24, 25).

Clinical performance is often predicted from in vitro measurements of the droplet size of an aerosol. Nebulized aerosols are usually assessed by a multistage liquid impinger (MSLI), cascade impactor, or laser diffraction methods. The MSLI and cascade impactors comprise a series of progressively finer jets and collection plates, permitting fractionation of aerosols according to their MMADs (26). Problems associated with the cascade impactors for the assessment of nebulized aerosols are that the high flow rates involved (typically $28\text{--}90\text{ L/min}$) give rise to rapid solvent evaporation and droplet entrainment. Additionally, these inertial impaction techniques are invasive, laborious, and time-consuming. Alternatively, nebulized aerosols may be sized by passing the spray through a beam of a laser diffraction analyzer. The volume or mass median diameter is then calculated from

the generated diffraction pattern. Provided the liquid density is known, the MMAD can be calculated.

Numerous investigators have compared the aerosol droplet size of nebulized aerosols from ultrasonic and air-jet devices (3, 4, 6, 27–31). Because droplet size is inversely proportional to the acoustic frequency, smaller droplets are generated from ultrasonic devices with higher frequencies. Ultrasonic nebulizers with high operating frequencies ($2\text{--}3\text{ MHz}$) are capable of generating droplets of sizes comparable to air-jet devices. However, devices with low frequencies ($<0.5\text{ MHz}$) tend to generate droplets outside the respirable range (4, 32). Most commercially available ultrasonic nebulizers have an operating frequency between 1 and 2 MHz. Although these produce aerosol droplets, which are significantly larger than those produced by air-jet nebulizers (3, 4), the size distributions are generally less polydisperse (3, 33).

NEBULIZATION TIME, DRUG OUTPUT, AND RESIDUAL VOLUME

Patient compliance with prescribed nebulization regimes is primarily determined by the duration of the therapy. Nebulizer fluids may be atomized for a set period, or more usually, a measured volume of therapeutic liquid is nebulized to “dryness.” The time taken to achieve this is directly related to the volume to be delivered. However, not all the fluid in the nebulizer can be atomized, and some fluid remains associated with the baffles, internal structures, and walls of the nebulizer as the “dead” or “residual” volume (34). The proportion of fluid remaining as the residual volume and thus unavailable to patients is higher for smaller fill volumes.

Ultrasonic devices typically produce a much larger fluid output per unit of time than do jet nebulizers. Although Sterk et al. (4) suggested that high airflow through these devices resulted in lower droplet concentrations per unit volume than for air-jet nebulizers, more recent findings indicate that the reduced volume of dilution air in modern ultrasonic devices gives a more concentrated aerosol cloud. Ultrasonic nebulizers may retain a higher residual volume than do comparable air-jet nebulizers (35), but they show less tendency to increase the solute concentration within the nebulizer chamber during operation (27, 29).

In comparing DeVilbiss ultrasonic and jet devices, Newman et al. (29) reported shorter nebulization times for the ultrasonic devices with higher fluid outputs. Furthermore, the air-jet devices tended to increase the concentration of drug in the nebulizer chamber more

than did the ultrasonic counterparts. The output from nebulizers comprises droplets of drug solutions and suspensions and solvent vapor, which saturates the outgoing air, causing solute concentration to increase during nebulization. Because ultrasonic nebulizers generally have higher fluid outputs and larger droplet sizes than do air-jet nebulizers, there is more solvent available within the dense aerosol cloud to saturate the outgoing air. Thus, the changes in solute concentration within the dead volume are much smaller than for air-jet devices.

THERMAL EFFECTS

Excess energy within the nebulizer is converted to heat, causing the temperature of the liquid in the reservoir to increase until the input energy balances the energy removed by evaporation and by conduction to the surroundings and circulating air (12). The temperature of a liquid in an ultrasonic nebulizer may thus increase by as much as 20°C above ambient temperature during use (6, 36). This results in changes in the properties of the fluid, including surface tension and viscosity. Although such changes may affect the aerosol particle size characteristics and overall drug output, little variation was seen during prolonged operation of Medix Electronic or Easimist nebulizers (37) or when the temperature in a DeVilbiss Aerosonic nebulizer was increased using a heating coil (38).

Warming may have a beneficial effect. For instance, the temperature of fluids atomized in air-jet nebulizers decreases by approximately 10–15°C during use (34, 39), resulting in bronchoconstriction in some asthma sufferers (40). Bronchoconstriction, which is most marked at 5°C, disappears at 37°C and thus may be minimized by using an ultrasonic device. Furthermore, when solutions of drugs with low solubility are to be nebulized, ultrasonic nebulizers, which warm the solutions, may be preferable (36) to air-jet devices, which cool them and may cause precipitation (39). However, the heat generated may harm heat-labile materials such as diethylenetriaminepentaacetic acid (^{99}mTc -DTPA) (41), proteins (31), and some antibiotic solutions (42). Thus, ultrasonic nebulizers are specifically prohibited for aerosolization of recombinant human deoxyribonuclease (rhDNase) (1).

FORMULATION OF NEBULIZER FLUIDS

Respiratory solutions are the mainstay of nebulized inhalation therapy. These typically contain drug dissolved

in aqueous, isotonic solvent systems, with the inclusion of cosolvents (e.g., ethanol) if necessary. Suspensions are less common for nebulization, although corticosteroid preparations such as budesonide are available. Antioxidant and antimicrobial preservatives may be included in the formulations; however, some (e.g., sodium metabisulfate, benzalkonium chloride, EDTA) may have paradoxical effects and cause coughing and bronchoconstriction (43). To prevent this, "preservative-free" unit dose products are marketed. Isotonicity is generally achieved using sodium chloride. Although iso-osmotic solutions of pH 3–8.5 are usually employed, the osmolality and pH may change during use, resulting in bronchospasm (44).

The size and output characteristics of aerosols generated from such liquids depend on their physicochemical properties (density, surface tension, viscosity) in conjunction with the nebulizer design and operating conditions. Empirical and semiempirical formulas predict that the droplet size of the aerosols is proportional to surface tension and inversely related to viscosity, whereas the effect of density, over the concentration range normally encountered, is negligible. The filtering effects of the baffles and solvent evaporation may modify the secondary aerosol produced, but studies (32, 37, 45, 46) have found viscosity to be a major determinant of aerosol size and output characteristics. High-viscosity fluids offer greater resistance to the integral fountain-disintegration process, thereby producing not only lower output but also larger droplets. Gershenzon and Eknadosyants (14) noted higher droplet outputs with lower viscosity fluids, whereas Boucher and Kreuter (32) reported that it was difficult to ultrasonically aerosolize fluids with viscosities exceeding 10 cP. Gershenzon and Eknadosyants (14) stated that the atomization rate for a wide range of fluids (except water) is given by the proportionality of A^2 to $\pi p/\eta\sigma$, where A is the atomization rate, p is the liquid vapor pressure, η is the viscosity, and σ is the surface tension.

Il'in and Eknadosyants (15) suggested that the dynamic viscosity coefficient was the most important property determining the nebulization rate. The rate of nebulization for solutions of tyloxapol or *N*-acetyl-L-cysteine decreased progressively as the solution viscosity increased. When certain oily and viscous liquids (e.g., Lipiodal®; a radio-opaque diagnostic agent) were nebulized, a fountain of liquid was generated, although this did not disintegrate to produce an aerosol. Similar results were reported by McCallion et al. (37), who compared a range of different model fluids, finding that ultrasonic nebulizers could not efficiently atomize the more viscous liquids and tended to produce poor total fluid outputs. Furthermore, when the fluid viscosity increased, larger droplets were generated.

Surface tension is also important because it represents the force resisting the formation of new surfaces. These forces tend to impair atomization by opposing any distortion or irregularity on the liquid surface, thereby delaying the onset of fountain formation. In suspension formulations, surfactants may be present as suspending agents. Reduction in liquid surface tension, through the addition of these agents, may decrease nebulization rate. This may be attributed to reduction in capillary wavelength, causing an increase in threshold amplitude (45) or through their influence on the diffusion of gas into cavitation bubbles (47). When a range of pharmaceutically relevant surfactant systems was nebulized, an inverse relationship was found between droplet size and surface tension over the entire concentration range investigated or to a peak value (33). There was no relationship between this peak value and a specific surface tension or the critical micelle concentration. In most cases, the total fluid output was unchanged.

Most nebulizer formulations are solutions, but a few corticosteroid suspension formulations have been marketed. In general, ultrasonic nebulizers are less efficient and more variable in delivering suspensions than are air-jet nebulizers. Although soluble radiopharmaceuticals may be more appropriate for delivery from ultrasonic nebulizers (48), Lin et al. (49) successfully used an ultrasonic nebulizer to deliver radiolabeled sulfur and tin colloids for lung imaging. McCallion et al. (50) nebulized a range of latex sphere suspensions in air-jet and ultrasonic nebulizers. No correlation was found between the size of the suspended spheres and the size distribution of the nebulized droplets. Higher outputs of smaller spheres were reported, with a concentrating effect occurring in the residual volume. The ultrasonic nebulizer studied was less efficient than the jet nebulizers, degrading some of the larger spheres and being unable to atomize suspended spheres of a specific size range.

Nebulizers, particularly air-jet devices, have been studied extensively for the delivery of liposomes to the lung (e.g., 51, 52). Because in ultrasonic nebulizers, the temperature of fluid in the reservoir is raised during use, they have generally been avoided for delivering liposomes, which exhibit temperature-dependent drug release. Barber and Shek (53) reported that egg phosphatidylcholine (egg PC) liposomes with a mean size of 281 nm or smaller were stable to nebulization in a DeVilbiss Ultra-Neb 99[®] ultrasonic nebulizer. However, dipalmitoylphosphatidylcholine liposomes of 499 nm increased in size within the nebulizer reservoir, suggesting fusion of vesicles, which could result in loss of entrapped hydrophilic materials. A later study (54) showed that the size of large multilamellar egg PC

liposomes remaining in a Medix Electronic nebulizer decreased markedly during nebulization, suggesting vesicle disruption, which was reduced by including cholesterol in the formulations.

Ultrasonic nebulizers are less suitable than jet nebulizers for delivery of proteins to the airways because of their thermal sensitivity. In a comparison of eight air-jet and two ultrasonic nebulizers, all air-jet nebulizers maintained the enzymatic activity of rhDNase in both the collected aerosol and the residual volume (31). With the ultrasonic nebulizers, some thermal denaturation of the enzyme was evident toward the end of the nebulization period when the liquid volume was minimal and its temperature highest. The maximum temperature of the rhDNase solution was 58°C, which was near the thermal transition temperature (approximately 65°C) of the enzyme (31).

Ip et al. (55) investigated DeVilbiss Aerosonic[®], Mountain Medical Microstat[®], and Medix Electronic[®] ultrasonic nebulizers for the delivery of recombinant consensus α -interferon. The extent of protein aggregation was governed by the type of nebulizer used with, the Easimist causing the least and Microstat the most aggregation. This was related to the increase in temperature and could be minimized by cooling the nebulizer solution during use. The Aerosonic[®] nebulizer completely inactivated lactate dehydrogenase after 20 min (38). The profile of inactivation differed from that of an air-jet nebulizer and was associated both with the temperature increase of the nebulizer fluid and the aerosol production. The activity of the enzyme was almost completely retained if Tween 80 or PEG 8000 was included in the nebulizer fluid.

GUIDELINES AND STANDARDIZATION FOR THE USE OF NEBULIZERS

The British Thoracic Society Nebulizer Project Group published a set of guidelines for nebulizer treatment (18). Their findings relating to ultrasonic devices are summarized as follows.

1. The aim of treatment is to deliver a therapeutic dose of the drug as an aerosol in the form of respirable particles within 5–10 min.
2. Nebulizers are useful when large doses of drug are to be inhaled by patients too ill to use alternative devices and when drugs are not available for delivery from pMDIs and DPIs.
3. Nebulizers are generally used to treat acute exacerbations of asthma or chronic obstructive pulmonary

disease. Other indications include long-term bronchodilator treatment of chronic airflow obstruction; prophylactic treatment for asthma; antimicrobial drugs for cystic fibrosis, bronchiectasis, and HIV/AIDS; and symptomatic relief in palliative care.

4. When drugs other than bronchodilators are being nebulized, equipment known to provide a suitable output should be used, and specific instructions should be given to patients. Such treatment should be supervised by hospital specialists.
5. Nebulization times for bronchodilators should be less than 10 min. Nebulization times and correct operation of devices should be familiar to the patients.

Convenience and patient preference determine whether a mouthpiece or face mask is used.

Currently, there are no *U.S. Pharmacopeia* or *European Pharmacopoeia* standard tests for characterizing the output of respiratory solutions or suspensions generated by air-jet or ultrasonic nebulizers. If nebulizers are to truly compete with pMDIs and DPIs, standardization of methods to measure dead volume, quantity of aerosol emitted from the device, nebulization times, and the size distribution of nebulized clouds is a logical and fundamental requirement.

REFERENCES

1. British National Formulary 38, British Medical Association and Royal Pharmaceutical Society of Great Britain: London, 2000.
2. Tovell, R.M.; D'Ambruso, D.C. Humidity in Inhalation Therapy. *Anesthesiology* **1962**, *23*, 452–459.
3. Mercer, T.T. Output Characteristics of Three Ultrasonic Nebulizers. *Chest* **1981**, *80*, 813–817.
4. Sterk, P.J.; Plomp, A.; Van der Vate, J.F.; Quanjer, P.H. Physical Properties of Aerosols Produced by Several Jet and Ultrasonic Nebulizers. *Bull. Eur. Physiopathol. Respir.* **1984**, *20*, 65–72.
5. Matthys, H.; Kohler, D. Pulmonary Deposition of Aerosols by Different Mechanical Devices. *Respiration* **1985**, *48*, 269–276.
6. Phipps, P.R.; Gonda, I. Droplets Produced by Medical Nebulizers: Some Factors Affecting Their Size and Solute Concentration. *Chest* **1990**, *97*, 1327–1332.
7. Hardy, J.G.; Newman, S.P.; Knoch, M. Lung Deposition from Four Nebulizers. *Respir. Med.* **1993**, *87*, 461–465.
8. Ballard, R.D.; Bogin, R.M.; Pak, J. Assessment of Bronchodilator Response to a β -Adrenergic Delivered from an Ultrasonic Nebulizer. *Chest* **1991**, *100*, 410–415.
9. Dalby, R.N.; Hickey, A.J.; Tiano, S.L. Medical Devices for the Delivery of Therapeutic Aerosols to the Lungs. *Inhalation Aerosols: Physical and Biological Basis for Therapy*; Hickey, A.J., Ed.; Marcel Dekker, Inc.: New York, 1996; 441–473.
10. Lang, R.J. Ultrasonic Atomization of Liquids. *J. Acoust. Soc. Am.* **1962**, *36*, 6–8.
11. Lobdell, D.D. Particle Size-Amplitude Reactions for the Ultrasonic Atomizer. *J. Acoust. Soc. Am.* **1968**, *43*, 229–231.
12. Mercer, T.T.; Goddard, R.F.; Flores, R.L. Output Characteristics of Three Ultrasonic Nebulizers. *Ann. Allergy* **1968**, *26*, 18–27.
13. Sollner, K. The Mechanism of the Formation of Fogs By Ultrasonic Waves. *Trans. Faraday Soc.* **1936**, *32*, 1532–1536.
14. Gershenzon, E.L.; Eknadiosyants, O.K. The Nature of Liquid Atomization in an Ultrasonic Fountain. *Sov. Phys. Acoust.* **1964**, *10*, 156–162.
15. Il'in, B.I.; Eknadiosyants, O.K. Nature of Liquid Atomization in an Ultrasonic Fountain. *Sov. Phys. Acoust.* **1966**, *12*, 310–318.
16. Boguslavskii, Y.Y.; Eknadiosyants, O.K. Physical Mechanism of the Acoustic Atomization of a Liquid. *Sov. Phys. Acoust.* **1969**, *15*, 14–21.
17. Greenspan, B.J. Ultrasonic and Electrodynamical Methods of Aerosol Generation. *Inhalation Aerosols: Physical and Biological Basis for Therapy*; Hickey, A.J., Ed.; Marcel Dekker, Inc.: New York, 1996; 313–335.
18. The Nebuliser Project Group of the British Society Standards of Care Committee. Current Best Practice for Nebuliser Treatment. *Thorax* **1997**, *52* (Suppl. 2), S1–S106.
19. Rhoades, E.R.; Ringrose, R.; Mohr, J.A.; Brooks, L.; McKown, B.A.; Felton, F. Contamination of Ultrasonic Nebulization Equipment with Gram Negative Bacteria. *Arch. Intern. Med.* **1971**, *127*, 228–232.
20. Zierenberg, B.J. The Respimat, a New Inhalation System Based on the Piezoelectric Effect. *Biopharmaceut. Sci.* **1992**, *3*, 85–90.
21. Baker, P.G.; Stimpson, P.G. Electronically Controlled Drug Delivery Systems Based on the Piezo Electric Crystal. *Respir. Drug Del.* **1994**, *4*, 273–285.
22. Humberstone, V.C.; Newcombe, G.C.F.; Sant, A.J.; Palmer, M.R. Fluid Droplet Production Apparatus and Method. EP-615-470: 1994.
23. Humberstone, V.C.; Sant, A.J.; van Rensburg, R.J. Piezoelectric Aerosol Inhalers: Technology to Product. Conference Proceedings of New Horizons in Pulmonary Drug Delivery, Management Forum: London, October 1–2, 1996.
24. Stahlhofen, W.; Gebhart, J.; Heyder, J. Experimental Determination of the Regional Deposition of Aerosol Particles in the Human Respiratory Tract. *Am. Ind. Hyg. Assoc. J.* **1980**, *41*, 385–398.
25. Newman, S.P.; Clarke, S.W. Therapeutic Aerosols. I. Physical and Practical Considerations. *Thorax* **1983**, *38*, 881–886.
26. Hallworth, G.W.; Andrews, H.G. Size Analysis of Suspension Inhalation Aerosols by Inertial Separation Methods. *J. Pharm. Pharmacol.* **1976**, *28*, 898–907.
27. Ferron, G.A.; Kerrebijn, K.F.; Weber, J. Properties of Aerosols Produced with Three Nebulizers. *Am. Rev. Respir. Dis.* **1976**, *114*, 899–908.
28. Ryan, G.; Dolovich, M.B.; Obminski, G.; Cockcroft, D.W.; Juniper, E.; Hargreave, F.E.; Newhouse, M.T. Standardization of Inhalation Provocation Tests: Influence of

- Nebulizer Output, Particle Size and Method of Inhalation. *J. Allergy Clin. Immunol.* **1981**, *67*, 156–161.
29. Newman, S.P.; Pellow, P.G.D.; Clarke, S.W. In Vitro Comparison of DeVilbiss Jet and Ultrasonic Nebulizers. *Chest* **1987**, *92*, 991–994.
30. Smalldone, G.C.; Perry, R.J.; Deutsch, D.G. Characteristics of Nebulizers Used in the Treatment of AIDS-Related *Pneumocystis Carinii* Pneumonia. *J. Aerosol Med.* **1988**, *1*, 113–126.
31. Cipolla, D.C.; Clarke, A.R.; Chan, H.-K.; Gonda, I.; Shire, S.J. Assessment of Aerosol Delivery Systems for Recombinant Human Deoxyribonuclease. *STP Pharm. Sci.* **1994**, *4*, 50–62.
32. Boucher, R.M.G.; Kreuter, J. The Fundamentals of the Ultrasonic Atomization of Medicated Solutions. *Ann. Allergy* **1968**, *26*, 591–600.
33. McCallion, O.N.M.; Taylor, K.M.G.; Thomas, M.; Taylor, A.J. The Influence of Surface Tension on Aerosols Produced by Medical Nebulisers. *Int. J. Pharm* **1996**, *129*, 123–136.
34. Clay, M.M.; Pavia, D.; Newman, S.P.; Lennard-Jones, T.; Clarke, S.W. Assessment of Jet Nebulisers for Lung Aerosol Therapy. *Lancet* **1983**, *2*, 592–594.
35. Flament, M.-P.; Leterme, P.; Gayot, A.T. Factors Influencing Nebulizing Efficiency. *Drug Dev. Ind. Pharm.* **1995**, *21*, 2263–2285.
36. Taylor, K.M.G.; Hoare, C. Ultrasonic Nebulisation of Pentamidine Isethionate. *Int. J. Pharm.* **1993**, *98*, 45–49.
37. McCallion, O.N.M.; Taylor, K.M.G.; Thomas, M.; Taylor, A.J. Nebulization of Fluids of Different Physico-chemical Properties with Air-Jet and Ultrasonic Nebulizers. *Pharm. Res.* **1995**, *12*, 1682–1688.
38. Niven, R.W.; Ip, A.Y.; Mittelman, S.; Prestrelski, S.J.; Arakawa, T. Some Factors Associated with the Ultrasonic Nebulization of Proteins. *Pharm. Res.* **1995**, *12*, 53–59.
39. Taylor, K.M.G.; Venthoeye, G.; Chawla, A. Pentamidine Isethionate Delivery from Jet Nebulisers. *Int. J. Pharm.* **1992**, *85*, 203–208.
40. Lewis, R.A. Nebulisers for Lung Aerosol Therapy. *Lancet* **1983**, *2*, 849.
41. Waldman, D.L.; Weber, D.A.; Oberdörster, G.; Drago, S.R.; Utell, M.J.; Hyde, R.W.; Morrow, P.E. Chemical Break-down of Technetium-99m DTPA During Nebulization. *J. Nucl. Med.* **1987**, *28*, 378–382.
42. Dennis, J.H.; Hendrick, D.J. Design Characteristics for Drug Nebulizers. *J. Med. Eng. Technol.* **1992**, *16*, 63–68.
43. Beasley, R.; Rafferty, P.; Holgate, S.T. Adverse Reactions to the Non-Drug Constituents of Nebuliser Solutions. *Br. J. Clin. Pharmacol.* **1988**, *25*, 283–287.
44. Schöni, M.H.; Kraemer, R. Osmolarity Changes in Nebulizer Solutions. *Eur. Respir. J.* **1989**, *2*, 887–892.
45. Davis, S.S. Physico-Chemical Studies on Aerosol Solutions for Drug Delivery I. Water-Propylene Glycol Systems. *Int. J. Pharm.* **1978**, *1*, 71–83.
46. Newman, S.P.; Pellow, P.G.D.; Clarke, S.W. Dropsizes from Medical Atomisers (Nebulisers) for Drug Solutions with Different Viscosities and Surface Tensions. *Atom. Spray Tech.* **1987**, *3*, 1–11.
47. Kapustina, O.A. Effect of Surface Active Substances on Bubble Growth Kinetics in a Sound Field. *Sov. Phys. Acoust.* **1969**, *15*, 110–111.
48. Istiman, A.T.; Manoli, R.; Schmidt, G.H.; Holmes, R.A. An Assessment of Alveolar Deposition and Pulmonary Clearance of Radiopharmaceuticals After Nebulization. *Am. J. Roentgenol.* **1974**, *120*, 776–781.
49. Lin, M.S.; Hayes, T.M.; Goodwin, D.A.; Kruse, S.L. Distal Penetration in Radioaerosol Inhalation with an Ultrasonic Nebulizer. *Radiology* **1974**, *112*, 443–447.
50. McCallion, O.N.M.; Taylor, K.M.G.; Thomas, M.; Taylor, A.J. Nebulisation of Monodisperse Latex Sphere Suspensions in Air-Jet and Ultrasonic Nebulisers. *Int. J. Pharm.* **1996**, *133*, 203–214.
51. Farr, S.J.; Kellaway, I.W.; Parry-Jones, D.R.; Woolfrey, S.G. 99m-Technetium as a Marker of Liposomal Deposition and Clearance in the Human Lung. *Int. J. Pharm.* **1985**, *26*, 303–316.
52. Taylor, K.M.G.; Taylor, G.; Kellaway, I.W.; Stevens, J. The Influence of Liposomal Encapsulation on Sodium Cromoglycate Pharmacokinetics in Man. *Pharm. Res.* **1989**, *6*, 633–636.
53. Barber, R.F.; Shek, P.N. Liposome Stability during Ultrasonic Nebulization. Proceedings of the 33rd Harden Conference, The Biochemical Society: Ashford, UK, 1989; 53.
54. Leung, K.K.M.; Bridges, P.A.; Taylor, K.M.G. The Stability of Liposomes to Ultrasonic Nebulisation. *Int. J. Pharm.* **1996**, *145*, 95–102.
55. Ip, A.Y.; Arakawa, T.; Silvers, H.; Ransone, C.M.; Niven, R.W. Stability of Recombinant Consensus Interferon to Air-Jet and Ultrasonic Nebulization. *J. Pharm. Sci.* **1995**, *84*, 1210–1214.

UNIT PROCESSES IN PHARMACY— THE OPERATIONS

Anthony J. Hickey

University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

David Ganderton

Vectura Ltd., Bath, United Kingdom

INTRODUCTION

The fundamentals of unit processes have been described elsewhere (1, 2). The following is a summary of the major operations. Although it is broad in scope, the abbreviated form of this article has resulted in selected omissions. References are given, wherever possible, to make up for this deficiency.

EVAPORATION AND DISTILLATION

Evaporation

The term evaporation, in the pharmaceutical industry, is primarily associated with the removal of water and other solvents by boiling in batch processes.

Heat transfer to boiling liquids in an evaporator

The heat required to boil a liquid in an evaporator is usually transferred from a heating fluid, such as steam or hot water, across the wall of a jacket or tube in or around which the liquid boils. A qualitative discussion of the methods used to secure high rates of heat flow can be used on Equation (1),

$$Q = UA \Delta T \quad (1)$$

where Q is the rate of heat flow, U is the overall heat transfer coefficient, A is the area over which heat is transferred, and ΔT is the difference in temperature between the fluids.

Other factors described by Eq. 1 are the area of the heat transfer surfaces, which should be as large as possible, and the temperature difference between the heating surface and the boiling liquid. As long as the critical heat flux is not exceeded, the latter also should be large.

The physical properties of solution and liquids

A number of physical factors, which are inter-related, are relevant to the study of evaporation. For a given heating

fluid, the temperature difference across the wall of an evaporator is determined by the boiling temperature, a variable controlled by the external pressure and the concentration of the solute in the solution. The boiling temperature and the solute concentration both influence the viscosity of the solution, a factor which greatly affects the heat transfer coefficient. The boiling temperature also determines the solubility of dissolved constituents and the degree of concentration which can be carried out without separation of solids.

The effect of heat on the active constituents of a solution: The thermal stability of components of a solution may determine the type of evaporator to be used and the conditions of its operation. If a simple solution contains a hydrolyzable material and the rate of its degradation during evaporation depends on its concentration at any time, an exponential relation between the remaining fraction, F , and the time, t , characteristic of a first-order reaction, is obtained, as shown in Eq. 2.

$$F = e^{-kt} \quad (2)$$

The dependence of the reaction velocity constant, k , on the absolute temperature, T , is expressed by Eq. 3,

$$k = Ae^{-B/T} \quad (3)$$

where A and B are constants characteristic of the reaction. Thus, at temperatures T_1 , T_2 , and T_3 , where $T_1 > T_2 > T_3$, the relation between the remaining fraction and the time of heating becomes clear, as shown in Fig. 1. This indicates the importance of the temperature and time of heating. If the latter can be shortened, the temperature of evaporation can be increased greatly without increasing the fraction which is degraded. If, therefore, the effect of temperature on the rate of evaporation is known, it is possible to define the conditions of time and temperature at which decomposition is minimum.

In practice, the kinetics of degradation and the relation of evaporation rate and temperature are usually not known. This is particularly true when the criteria by which the product is judged are color, taste, or smell. In addition, this

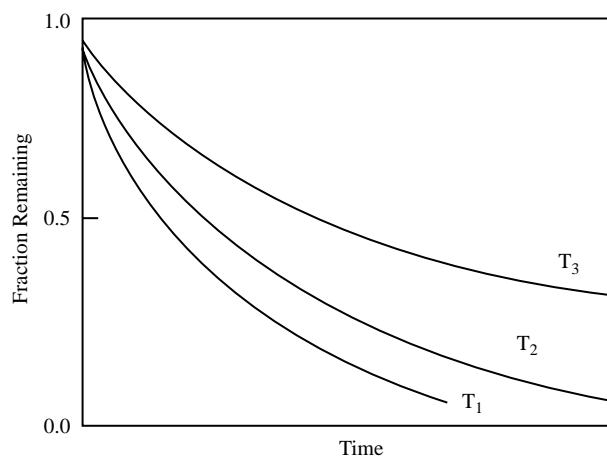


Fig. 1 The effect of time and temperature of degradation.

analysis neglects the temperature variation in the evaporating liquid and the degradation in boundary films where temperatures are higher. Therefore, experiments are often necessary to determine the suitability of an evaporation process.

In batch processes, the time of exposure to heat is well defined. This is also true of continuous processes in which the liquid to be evaporated is passed only once through the heater. In continuous processes, where the liquid is recirculated through the heater, the average residence time, a , is given by the ratio.

The volumetric discharge is only an indication of the damage that may be caused by prolonged heating. If perfect mixing occurs in the evaporator, the fraction, f , which is in the unit for time t or less, is given by Eq. 4.

$$f = 1 - e^{-t/a} \quad (4)$$

This relation shows that an evaporator, for example, with an average residence time of 1 h holds 13.5% of active principles for 2 h and about 2% for 4 h.

Evaporators

It is convenient to classify evaporators into natural circulation evaporators, forced-circulation evaporators, and film evaporators.

Natural circulation evaporators

Small-scale evaporators consist of a simple pan heated by a jacket or a coil, or both. Admission of the heating fluid to the jacket induces the liquid in the vessel to boil. Very small evaporators may be open, the vapor escaping to atmosphere or into a vented hood. Larger pan evaporators are closed, the vapor escaping through a pipe. Small

jacketed pans are efficient and easy to clean and may be fitted for the vacuum evaporation of thermolabile materials. Their size, however, is limited because the ratio of heating area to volume decreases as the capacity increases. Larger vessels must employ a heating coil which increases their evaporating capacity but it also makes cleaning more difficult.

Forced-circulation evaporators

On the smallest scale, forced-circulation evaporators are similar to the pan evaporators described previously, modified only by the inclusion of an agitator. Vigorous agitation increases the boiling film coefficient, the degree of which depends on the type and speed of the agitator. An agitator should be used for the evaporation of viscous materials to prevent degradation of material at the heated surfaces. Some large-scale continuous units are similar to the natural circulation evaporators already described.

Film evaporators

In the short tubes of the calandria, an intimate mixture of vapor and liquid is discharged at the top as shown in Fig. 2a. If the length of the tube is greatly increased, progressive phase separation occurs until a high velocity core of vapor is formed which propels an annular film of liquid along the tube. This phenomenon, which is one stage of flow when a liquid and a gas pass in the same direction along a tube, is employed in film evaporators. The turbulence of the film gives very high heat transfer coefficients, and the bubbles and vapor evolved are rapidly swept into the vapor stream. Although recirculation may be adopted, it is possible with the high evaporation is found in long tubes to sufficiently concentrate the liquid in a single pass. Because a very short residence time is obtained, very thermolabile materials may be concentrated at relatively high temperatures. Film evaporators are also

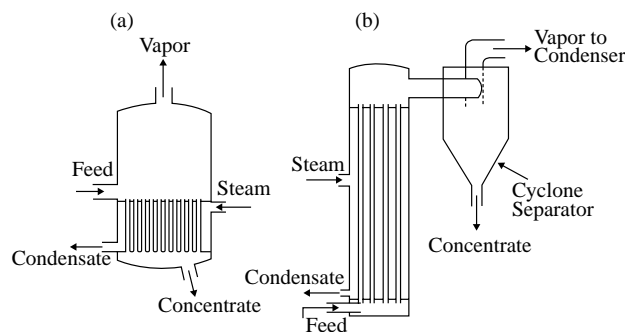


Fig. 2 (a) Evaporator with calandria; (b) climbing film evaporator.

suitable for materials which foam strongly. Various types have been developed, but all are essentially continuous in operation and their capacity starts from a few liters per hour upward.

The climbing film evaporator, which is the most common film evaporator, consists of tubes 15 to 30 feet (4.5–13.6 m) long and 1 to 2 in. (2.5–5 cm) in diameter mounted in a steam chest (Fig. 2b). The feed liquid enters the bottom of the tubes and flows upward for a short distance before boiling begins. The length of this section, which is characterized by low heat transfer coefficients, may be minimized by preheating the feed to its boiling point. The pattern of boiling and phase separation follows and a mixture of liquid and vapor emerges from the top of the tube to be separated by baffles or by a cyclone separator. Climbing film evaporators are not suitable for the evaporation of viscous liquids.

In the falling film evaporator, the liquid is fed to the top of a number of long heated tubes. Because the gravity assists the flow down the tube, this arrangement is better suited to the evaporation of moderately viscous liquids. The vapor evolved is usually carried downward and the mixture of liquid and vapor emerges from the bottom for separation. Even distribution of liquid must be secured during feeding. The tendency to channel in some tubes leads to drying in others.

The rising-falling film evaporator concentrates a liquid in a climbing film section and leads the emerging liquid and vapor into a second tube section which forms a falling film evaporator. Good distribution of liquid is claimed in the falling film section and this type is particularly suitable for liquids which greatly increase in viscosity during evaporation.

In mechanically aided film evaporators, a thin film of material is maintained on the heat transfer surface irrespective of the viscosity. This is usually achieved by means of a rotor, concentric with the tube, which carries blades that either scrape the tube or ride with low clearance in the film. Mechanical agitation permits the evaporation of highly viscous materials or those that have a low thermal conductivity. Because the temperature variations in the film are reduced and residence times are shortened, the vacuum evaporation of viscous thermolabile materials becomes possible.

Evaporation without boiling

During heating, some evaporation takes place at the surface of a batch of liquid before boiling begins. Similarly, liquids that are very viscous or excessively frothy may be concentrated without boiling. The diffusion of vapor from the surface is described by Eq. 5,

$$N_A = k_g RT(P_{Ai} - P_{Ag}) \quad (5)$$

where N_A is the number of moles evaporating from unit area in unit time, k_g is the mass transfer coefficient across the boundary layer, R is the gas constant, T is the absolute temperature, P_{Ai} is vapor pressure of the liquid, and P_{Ag} is the partial pressure of the vapor in the gas stream; k_g is proportional to the gas velocity.

Distillation

Distillation is a process in which a liquid mixture is separated into its component parts by vaporization. The vapor evolved from a boiling liquid mixture is normally richer in the more volatile components than the liquid with which it is in equilibrium. Distillation rests upon this fact. Although multicomponent mixtures are most common in distillation processes, an understanding of the operation can be based on the vapor pressure characteristics of two-component or binary mixtures.

Binary mixtures of immiscible liquids: steam distillation

If the components of a binary mixture are immiscible, the vapor pressure of the mixture is the sum of the vapor pressures of the two components, each exerted independently and not as a function of their relative concentrations in the liquid. This property is employed in steam distillation, a process particularly applicable to the separation of high boiling substances from nonvolatile impurities. The steam forms a cheap and inert carrier. The principles of the process also apply to other immiscible systems.

The composition of the distillate is calculated in the following way. For two components, A and B, the total vapor pressure, P , is the sum of the vapor pressures of the components, P_A and P_B . Since the partial pressure of a component in a gaseous mixture is proportional to its molar concentration, the composition of the vapor is given by Eq. 6,

$$\frac{n_A}{n_B} = \frac{P_A}{P_B} \quad (6)$$

where n_A and n_B are the number of moles of A and B in the vapor, respectively. If W_A and W_B are the weights of A and B in the vapor, Eq. 7 holds,

$$\frac{W_A}{M_A} = \frac{M_B}{W_B} = \frac{P_A}{P_B} \quad (7)$$

where M_A and M_B are the respective molecular weights. The distillate obtained from the vapor is $W_A + W_B$, and the percentage of A in the distillate is expressed by Eq. 8.

$$\frac{W_A}{M_A + W_B} \times 100 = \frac{P_A M_A}{P_A M_A + P_B W_B} \times 100 \quad (8)$$

The ratio of immiscible organic liquid to water in the distillate is increased if the former has a high molecular weight or a high vapor pressure. Steam distillation under vacuum may be employed when the thermal stability of the material prohibits temperatures of approximately 100°C.

The relation of vapor pressure and mixture composition: In a binary mixture of two completely miscible components, the vapor pressure is a function of the mixture composition as well as of the vapor pressures of the two pure components. If the liquids are ideal, the relation of vapor pressure and composition is given by Raoult's law. At a constant temperature, the partial vapor pressure of a constituent of an ideal mixture is proportional to its mole fraction in the liquid. Thus, for a mixture of A and B, the partial vapor pressure of A is given by Eq. 9,

$$P_A = P_A^0 \cdot x_A \quad (9)$$

where P_A^0 is the vapor pressure of pure A and x_A is its mole fraction. Similarly, the partial vapor pressure of B is expressed by Eq. 10.

$$P_B = P_B^0 \cdot x_B \quad (10)$$

The total pressure of the system, P , is simply $P_A + P_B$.

Very few liquid mixtures rigidly obey Raoult's law. Consequently, the vapor pressure data must be determined experimentally. Mixtures that deviate positively from this law give a total vapor pressure curve which lies above the theoretical straight line. Negative deviations fall below the line. In extreme cases, deviations are so large that a range of mixtures exhibits a higher or lower vapor pressure than either of the pure components.

Returning to ideal systems, the partial pressure of a component in the vapor is proportional to its mole fraction, as shown in Eq. 11 for component A,

$$P_A = y_A P \quad (11)$$

where P_A is the partial pressure of A in the vapor and y_A is its mole fraction. Because, $P_A = P_A^0 x_A$, Eqs. 12 and 13 can be written as:

$$y_A = \frac{x_A P_A^0}{P} \quad (12)$$

similarly,

$$y_B = \frac{x_B P_B^0}{P} \quad (13)$$

If A is the more volatile component, P_A^0 is greater than P_B^0 . Therefore y_A is greater than x_A , that is, the vapor is richer

in the more volatile component than the liquid with which it is in equilibrium.

Systems that form minimum boiling mixtures are common. Ethyl alcohol and water provide an example, the azeotrope containing 4.5% by weight of water. The boiling point at atmospheric pressure is 78.15°C that is 0.25°C lower than the boiling point of pure alcohol. Maximum boiling mixtures are less common. The most familiar example is hydrochloric acid which forms an azeotrope boiling at 108.6°C containing 20.2% by weight of hydrochloric acid. Mixtures that form azeotropes cannot be separated into the pure components by normal distillation methods. However, separation into the azeotrope and one pure component is possible.

Simple or differential distillation

In simple or differential distillation, the vapor evolved from the boiling mixture is immediately removed and condensed. Unless the boiling points of the two pure components differ widely, a reasonable degree of separation is not possible. This method may be used to remove low boiling solvents from aqueous solutions.

Rectification or fractionation

In simple distillation, vapor enrichment is small. In fractionation, a term synonymous with rectification, the vapor leaving the boiling liquid is led up a column to meet a liquid stream or reflux which originates higher in the column as part of the condensate. In a series of partial condensations and vaporizations, the rising vapor becomes richer in the more volatile component at the expense of the falling liquid and high degrees of separation become possible. The columns, called fractionating columns, are of two basic types: packed columns and plate columns.

Packed columns: These are used for laboratory and small-scale industrial distillation and are usually operated as a batch process. The column consists of a vertical, hollow, cylindrical shell containing a packing designed to offer a large interfacial contact area between liquid and vapor. The form of the packing varies but Raschig rings, which consist of small metallic or ceramic cylinders, are the most commonly used. In general, packed columns operate under widely varying conditions without serious loss of efficiency.

Plate columns: A plate column consists of a series of plates or trays on which the liquid is retained for some period during its movement down the column. The rising vapor is bubbled through this liquid, providing intimate contact between the phases. Liquid in reflux moves downward between plates and is usually carried by a downcomer. Contact between the vapor and liquid takes place in stages.

Plate columns operate efficiently over a limited range of conditions. They are mainly used in large-scale continuous installations in which the conditions of distillation can be closely maintained.

Molecular distillation

Molecular distillation is carried out without boiling at very low pressures of the order 0.001 mm Hg (0.133 Pa). At these pressures, collision of molecules in the evolving vapor and reflection back to the liquid surface is greatly decreased and the mean free path of the molecules is of the same order as the distance between the evaporating surface and a condenser placed a short distance away. It then becomes possible to distil liquids of very high boiling point although the degree of separation cannot exceed one theoretical plate. The process therefore is used primarily to concentrate nonvolatile components in a high boiling medium. The vitamins in cod liver oil can be concentrated this way. For the separation of liquids of comparable volatility, several separate distillation stages are necessary.

Because agitation due to boiling is absent, an alternative method of maintaining the more volatile component at the evaporating surface must be adopted. In the industrial molecular still shown in Fig. 3, the feed is introduced at the bottom of a heated conical rotor and flows upward as a thin liquid layer under the action of centrifugal force. The residue is caught in a gutter at the top. The vapor is condensed on a concentric, water-cooled condenser a short distance away and discharged.

AIR CONDITIONING AND HUMIDIFICATION

General principles of the supply of air in pharmaceutical processes are similar to conventional air conditioning. The control of its quality, however, may be more stringent. In areas where sterile materials are made and handled, for example, the cleaning process must remove bacteria. In other situations, it may be necessary to remove water vapor. The flow of powders is a sensitive function of moisture content, and the equilibrium moisture content of a material is determined by the humidity. Some tableting processes break down if the humidity is too high. In such processes, the scale of the air conditioning varies. It may be necessary to supply a whole room with air of a certain quality. Alternatively, conditioning may be restricted to a small area surrounding a particular piece of equipment.

Vapor and Gas Mixtures

The study of the properties of the air–water vapor mixture is called psychrometry, and data are presented in various forms of psychrometric charts presenting various data. In Fig. 4, humidity is plotted as ordinate and temperatures as abscissa. Percentage of relative humidity is plotted as curves running across the chart. The use of this simplified chart is demonstrated later.

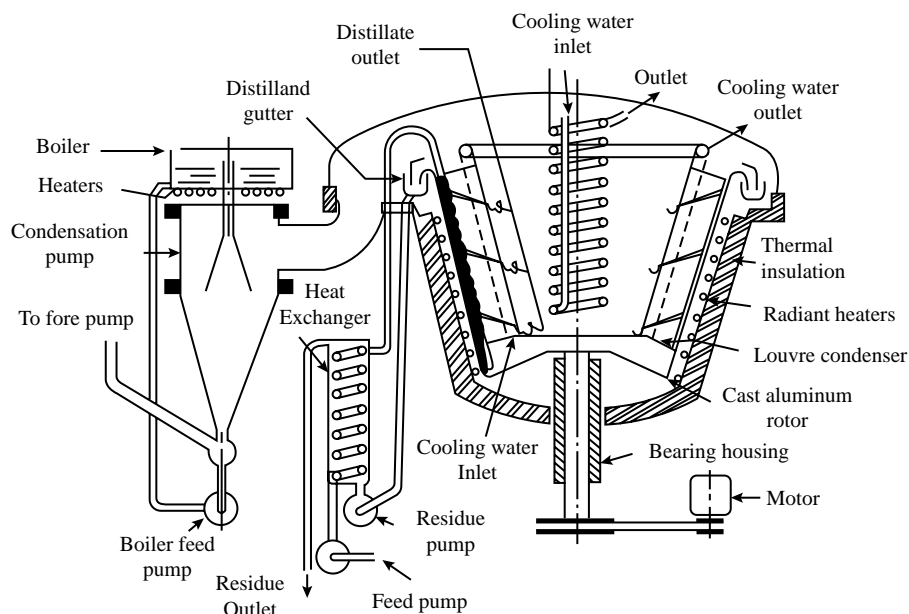


Fig. 3 Large-scale molecular still. (From Ref. 3)

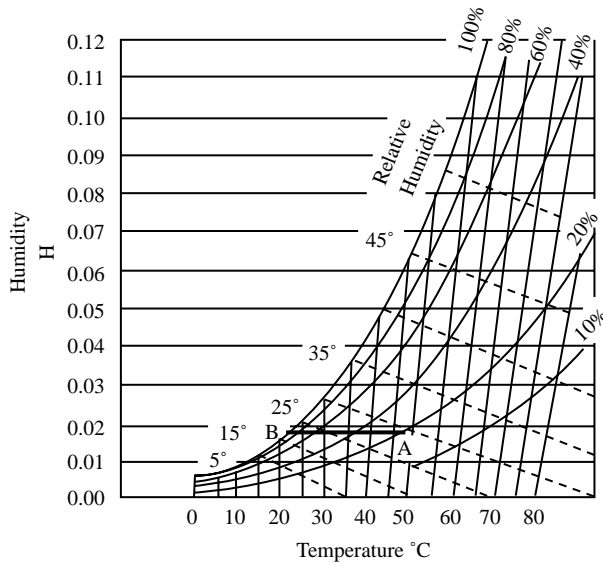


Fig. 4 A psychrometric chart.

Hygrometry, the Measurement of Humidity

The accurate determination of the humidity of air is carried out gravimetrically. The water vapor present in a known volume of air is chemically absorbed with a suitable reagent and weighed. In less laborious methods, the humidity is derived from the dew point or the wet-bulb depression of a water–vapor mixture.

The dew point is the temperature at which a vapor condenses when cooled at constant pressure. If air of the condition denoted by point A in Fig. 4 is cooled, the relative humidity increases until the mixture is fully saturated. This condition is given by point B; the temperature coordinator is the dew point, which can be measured rapidly by evaporating ether in a silvered bulb. The temperature at which dew deposits from the surrounding air is noted and the humidity is read directly from a psychrometric chart.

The derivation of the humidity from the wet-bulb depression requires a preliminary study of the transfer of mass and heat at a boundary between air and water. The difference between the air temperature and the wet-bulb temperature is the wet-bulb depression. If these temperatures are denoted by T_a and T_{wb} , the rate of heat transfer, Q , is given by Eq. 14,

$$Q = hA(T_a - T_{wb}) \quad (14)$$

$$Q = \rho k_g A(H_i - H_a) \quad (15)$$

Equating Eqs. 14 and 15 gives Eq. 16.

$$H_i - H_a = \frac{h}{\rho k_g \lambda} (T_a - T_{wb}) \quad (16)$$

Both the heat and mass transfer coefficients are functions of air velocity. However, at air speeds greater than about 15 ft/s (4.5 m/s), the ratio $h:k_g$ is approximately constant. The wet-bulb depression is directly proportional to the difference between the humidity at the surface and the humidity in the bulk of the air. In the wet- and dry-bulb hygrometer, the wet-bulb depression is measured by two thermometers, one of which is fitted with a fabric sleeve wetted with water. These thermometers are mounted side by side and shielded from radiation, an effect neglected in the derivation above. Air is drawn over the thermometers by means of a small fan. The derivation of the humidity from the wet-bulb depression and a psychrometric chart are discussed later.

Many wet- and dry-bulb hygrometers operate without any form of induced air velocity at the wet bulb. This may be explained by examining another air–water system. If a limited quantity of air and water is allowed to equilibrate under conditions in which heat is neither gained nor lost by the system, the air becomes saturated and the latent heat required for evaporation is drawn from both fluids which cool to the same temperature. This temperature is the adiabatic saturation temperature, T_∞ . It is a peculiarity of the air–water system that the adiabatic saturation temperature and the wet-bulb temperature are the same. If water at this temperature is recycled in a system through which air is passing, the incoming air is cooled till it reaches the adiabatic saturation temperature at which point it is saturated. The temperature of the water, on the other hand, remains constant and all the latent heat required for evaporation is drawn from the sensible heat of the air. Equilibrium is expressed by Eq. 17,

$$(T_a - T_\infty)S = (H_\infty - H_a)\lambda \quad (17)$$

where T_a is the temperature of the incoming air and S is its specific heat, H_a and H_∞ are the humidities of the incoming air and the saturated air, and λ is the latent heat of evaporation for water.

When both wet- and dry-bulb temperatures have been found, the humidity is read from the psychrometric chart in the following way. The point on the saturation curve corresponding to the wet-bulb temperature is found first. An adiabatic cooling line is then interpolated and followed until the coordinate corresponding to the dry-bulb temperature is reached. The humidity is read from the other axis.

Humidification and Dehumidification

Most commonly, air is humidified by passage through a spray of water. Small quantities of air are easily dehumidified by adsorbing the water vapor with alumina or silica gel arranged in columns. These are mounted in pairs so that one can be regenerated while the other is in use. Alternatively, the air can be cooled below the dew point. Excess water vapor condenses and the cold saturated air is reheated.

CRYSTALLIZATION

The term crystallization describes the production of a solid, single-component, crystalline phase from a multi-component fluid phase. The importance of crystallization lies primarily in the purification achieved during the process and in the physical properties of the product. A crystalline powder is easily handled, stable, and often possesses good flow properties and an attractive appearance.

Crystallization from a vapor, which occurs naturally, e.g., in the formation of hoar frost, is employed in sublimation processes and for the condensation of water vapor during freeze-drying.

Crystallization in Melts

A melt may be defined as the liquid form of a single material or the homogeneous liquid form of two or more materials which solidify on cooling. Crystallization in such a system passes through the following stages: supercooling, nuclei formation, and crystal growth.

If a single-component liquid is cooled, some degree, often high, of supercooling must be established before crystal nuclei form and growth begins. A metastable liquid region exists below the melting point which only can be entered by cooling. In this metastable, supercooled region, the absence of nucleation precludes the formation and growth of crystals. If, however, a crystal seed is added, growth occurs. The deliberate seeding of a metastable system is commonly employed in industrial crystallization. With further cooling, spontaneous nucleation usually takes place and the released heat of crystallization raises the temperature of the melt to its true melting point.

Nucleation

In certain single-component systems, such as piperine, nucleation and crystal growth are independent and can be

separately studied. The rate of nucleation as a function of supercooling is studied by maintaining the melt for a certain time at the given temperature and then quickly raising the temperature to the metastable region where further nucleation is negligible but the already formed nuclei can grow.

Spontaneous nucleation occurs when sufficient molecules of low kinetic energy come together in such a way that the attraction between them is sufficient to overcome their momentum.

Crystal Growth

If nucleation and crystal growth are independent, the latter can be studied by seeding a melt with small crystals under conditions of little or no natural nucleation. The rate of growth can then be measured. The form of the crystal growth curve is again explained by the kinetics of the molecules. At temperatures just below the melting point, molecules have too much energy to remain in the crystal lattice. As the temperature falls, more molecules are retained and the growth rate increases. Ultimately, however, diffusion to and orientation at the crystal surface is depressed.

For crystal growth in a single component melt, the molecules at the crystal surface must reach the correct position at the lattice and become suitably orientated, losing kinetic energy. These energy changes appear as heat of crystallization, which must be transferred from the surface to the bulk of the melt. The rate of crystal growth is influenced by the rate of heat transfer and the changes taking place at the surface. Agitation of the system increases heat transfer by reducing the thermal resistance of the liquid layers adjacent to the crystal until the changes at the crystal face become the controlling effect.

In multicomponent melts and solutions, deposition of material at the crystal face depletes the adjacent liquid layers and a concentration gradient is set up with saturation at the face and supersaturation in the liquid. Diffusion of molecules to the crystal face is discussed in the next section.

Crystallization from Solutions

During crystal growth, a high degree of supersaturation promotes a high growth rate. A reaction at the surface, in which, solute molecules become correctly orientated in the crystal lattice, provides a second resistance to the growth of the crystal. Simultaneously, the heat of crystallization must be conducted away.

For given conditions of temperature and saturation, agitation modifies the rate of crystal growth. Initially,

agitation quickly increases the growth rate by decreasing the thickness of the boundary layer and the diffusional resistance. However, as agitation is intensified, a limiting value is reached which is determined by the kinetics of the surface reaction.

As with melts, soluble impurities may increase or reduce nucleation rate. Insoluble materials may act as nuclei and promote crystallization. Impurities may also affect crystal form and, in some cases, are deliberately added to secure a product with good appearance, absence of caking, or suitable flow properties.

Crystallizers

Although other methods may be adopted, crystallizers can be classified conveniently in the same way, a solution is supersaturated. This leads to the self-explanatory terms, cooling crystallizer and evaporate crystallizer. In vacuum crystallizers, evaporation and cooling both take place.

FILTRATION

Filtration may be defined as the removal of solids suspended in a liquid or gas by passage through a pervious medium on which the solids are retained. The pervious medium or septum is normally supported on a base and these, together with a suitable housing providing free access of fluid to and from the septum, comprise the filter.

Methods

Clarification

Clarification of parenteral solutions eliminates unwanted solids normally present in very small concentrations. This may be carried out with the help of thick media, which allow the penetration and arrest of particles by entrapment, impingement, and electrostatic effects. This procedure leads to the concept of depth filtration in which particles, perhaps a hundred times smaller than the dimensions of the passages through the medium, are removed. Such filters are not absolute and must be designed with sufficient depth so that the probability of the smallest particle under consideration passing right through the filter is extremely small.

Depth filtration fundamentally differs from the use of media in which pore size determines the size of particle retained. Such filters may be said to be "absolute" at a particle diameter closely related to the size of the pore, so that there is a relatively sharp division between particles which pass the filter and those that are retained. An analogy

with sieving may be drawn for this mechanism. The life of such filters depends on the number of pores available for the passage of fluid. Once a particle is trapped at the entrance to the pore, the contribution of the latter to the overall flow of liquid is very much reduced. Coarse straining with a wire mesh and membrane filter employ this mechanism. Sterilization of liquids by filtration could be regarded as an extreme application of clarification in which the complete removal of particles as small as 0.3 μm must be ensured.

Cake filtration

The most common industrial application is the filtration of slurries containing a relatively large amount of suspended solids, usually 3 to 20%. The septum acts only as a support in this operation, the actual filtration being carried out by the solids deposited as a cake. In such cases, solids may completely penetrate the septum until the deposition of an effective cake occurs. Until this time, cloudy filtrate may be recycled. The physical properties of the cake largely determine the method employed. Washing and partial drying or dewatering are often integral parts of the process. Effective discharge of the cake completes the process. The solids, the filtrate, or both may be wanted.

The Theories of Filtration

Filtration theory has two important aspects. The first describes the flow of fluids through porous media and is applicable to both clarification and cake filtration. The second, which is of primary importance only in clarification, is the retention of particles on a depth filter.

Flow of fluids through porous media

The concept of a channel with a hydraulic diameter equivalent to the complex interstitial network which exists in a powder bed leads to Eq. 18,

$$Q = \frac{KA \Delta P}{\eta L} \quad (18)$$

where Q is the volumetric flow rate, A is the area of the bed and L its thickness, ΔP is the pressure difference, and η is the viscosity of the fluid. The permeability coefficient, K , is given by

$$\frac{\varepsilon^3}{5(1 - \varepsilon)^2 S_0^2}$$

where ε is the porosity of the bed and S_0 its specific surface area (cm^2/cm^3).

In clarification, high permeability and filtration rate oppose good particle retention. In the formation of clarify-

ing media from sintered or loose articles, accurate control of particle size, specific surface and porosity is possible, and, a medium can be designed which offers the best compromise between permeability and particle retention. The analysis of permeability given above can be accurately applied to these systems. Because of extremes of shape, this is not so with the fibrous media used for clarification. Here it is possible to develop a material of high permeability and high retentive capacity. Such a material is, however, intrinsically weak and must be adequately supported.

A mathematical account of the theories of clarification with depth filters is found in the work of Ives (4, 5) and Maroudas and Eisenklam (6).

Filters

The method by which the filtrate is driven through the filter medium and cake, if present may be used to classify filters into:

- Gravity filters
- Vacuum filters
- Pressure filters

Each group may be further subdivided into filters employed in continuous or batch processes although, due to technical difficulties, continuous pressure filters are uncommon and expensive. Centrifugation is another means of removing filtrate. Extensive surveys can be found in the literature (7, 8).

Many small-scale filters simply consist of a fixed, rigid medium, robust enough to withstand limited pressures, mounted in a suitable housing. These filters, which are also vacuum operated, are used to clarify by depth filtration. Media are composed of sintered metals, ceramics, plastics, or glass. Filters prepared from closely graded and sintered chemical powders are suitable for the sterilization of solutions by filtration on a manufacturing scale.

Filter Media

In cake filtration, the medium must oppose excessive penetration and promote the formation of a junction with the cake, to high permeability. The medium should also give free discharge of cake after washing and dewatering.

Rigid media

Rigid media may be loose or fixed. The former is exemplified by the deposition of a filter aid on a suitable support. Filtration characteristics are governed mainly by particle size, size distribution, and shape in a manner described earlier. These factors may be varied for different filtering requirements.

Fixed media vary from perforated metals used for coarse straining for the removal of very fine particles with a sintered aggregate of metal, ceramic, plastic, or glass powder. The size, size distribution, and shape of the powder particles together with the sintering conditions control the size and distribution of the pores in the final product. The permeability may be expressed in terms of the coefficient given in Eq. 18. Alternatively, the medium may be characterized by air permeability. The maximum pore size, which is important in the selection of filters for sterilization, may be determined by measuring the pressure difference required to blow a bubble of air through the medium while it supports a column of liquid with a known surface tension.

Flexible media

Flexible media may be woven or unwoven. Filter media, woven from cotton, wool, synthetic and regenerated fibers, and glass and metal fibers, are used as septa in cake filtration. Cotton is the most widely used natural fiber, nylon is predominant among synthetic fibers. Terylene is a useful medium for acid filtration. Penetration and cake discharge are influenced by twisting and plying of fibers and by the adoption of various weaves such as duck and twill. The choice of a particular cloth often depends on the chemical nature of the slurry.

Nonwoven media in the form of felts and compressed cellulose pulps, are used for clarification by depth filtration. Unless carefully prepared, they have the disadvantage of losing fibrous material from the downstream side of the filter. The application of sheet media has been discussed earlier. High wet strength is conferred on paper sheets by resin impregnation. An alternative technique employs asbestos fibers supported in a cellulose framework.

Mechanism of air filtration

A theoretical foundation for the filtration of air by passage through fibrous media was laid in the early 1930s by studies of the flow of suspended particles around various obstacles. In studies of the filtration of smokes (9, 10) it has been shown that the following factors operate simultaneously in the arrest of a particle during its passage through a filter, although their relative importance varies with the type of filter and the conditions under which it is operated.

- Diffusion effects due to Brownian movement
- Electrostatic attraction between particles and fibers
- Direct interception of a particle by a fiber
- Interception as a result of inertial effects acting on a particle and causing it to collide with a fiber
- Settling and gravitational effects

Air filters operate under conditions of streamline flow as indicated by the streamlines drawn around a cylindrical fiber. It was assumed that capture of a particle takes place if any contact is made during its movement around the fiber. Once captured, the particle is not re-entrained in the air stream and deposited deeper in the bed. Support for this assumption has been found by using an atomized suspension of *Staphylococcus albus* and spores of *Bacillus subtilis* (11). Nevertheless, some fiber filters are treated with viscous oils, presumably to make capture more positive and to reduce re-entrainment.

Deviation of particles from streamlines can occur in a number of ways (10, 12). The chance of capture increases if Brownian movement causes appreciable migration across streamlines, an effect only important for small particles (less than 0.5 μm) and low air speeds, when the time span spent in the vicinity of a fiber is relatively long. These conditions also apply to capture which is the result of electrostatic attraction.

Sampling efficiency has been demonstrated for bacterial aerosols (13) in a study of the efficiency with which a glass fiber mat collected *B. subtilis* spores atomized as particles just over a micrometer in radius. A theoretical approach to the removal of industrial dusts has been developed (14–16).

Design, operation, and testing of air filters

Granular beds, fibrous media, and “absolute filters” prepared from cellulose and asbestos are used for high-efficiency air filtration. With fibrous and granular filters, the fractional reduction in particle content is assumed to be the same through successive incremental thicknesses of the filter, expressed by Eq. 19,

$$\frac{dC}{dx} = -kC \quad (19)$$

where C represents the number of particles entering a section of thickness dx . The constant, k , is a measure of the filter's ability to retain a particle. It is a complex function of fiber diameter, interfiber distance, and the operational air velocity. Integration between inlet and outlet conditions gives Eq. 20.

$$\log \frac{C_{\text{out}}}{C_{\text{in}}} = -kC \quad (20)$$

The use of this log penetration effect in filter design has been described elsewhere (17). If a certain filter thickness is capable of retaining 90% of the entering particles and 10^6 particles enter, 10^5 penetrate. If six thicknesses are used, Eq. 21 predicts that only one particle penetrates. The

log-penetration effect has been confirmed for fibrous filters (13) and granular beds (18).

Centrifugal Operations

An object moving in a circular path is subjected to an outward centrifugal force which balances the centripetal force moving the object toward the center of rotation. This principle is used in the mechanical separations called centrifugal filtration and centrifugal sedimentation. In the former, a material is placed in a rotating perforated basket which is lined by a filter cloth used to separate a solid, which is retained at the cloth, from a liquid. It is essentially a filtration process in which the driving force is of centrifugal origin. This does not depend upon a difference in the density of the two phases.

In centrifugal sedimentation, the separation is due to the difference in the density of two or more phases. This is the more important process, where both solid–liquid mixtures and liquid–liquid mixtures can be completely separated. If, however, the separation is incomplete, there is a gradient in the size of the dispersed phase within the centrifuge due to the faster radial velocity of the larger particles. Operated in this way, the centrifuge becomes a classifier.

Centrifugal sedimentation

The motion of a particle in a liquid is described by Stokes' equation. If its diameter is d , the rate u at which it settles by gravity in a liquid of viscosity η and density ρ is given by Eq. 21

$$u = \frac{1}{18} d^2 \frac{\rho_s - \rho}{\eta} g \quad (21)$$

where g is the acceleration due to gravity, and ρ_s is the density of the particle. In the centrifuge, the gravitational force causing separation is replaced by a centrifugal force. If the particle has a mass m and moves at an angular velocity ω in a circle of radius r , the centrifugal force is $\omega^2 r(m - m_1)$, where m_1 is the mass of the displaced liquid. The expression

$$\frac{\omega^2 r}{g}$$

is, therefore, the ratio of the centrifugal and gravitational forces in the example described previously. Its value can exceed 10,000. The separation is quicker, more complete, and effective in systems containing very fine particles which do not settle by gravity because of Brownian movement.

Expressing the mass of the particle in terms of its volume and effective density, the centrifugal force can be written as in Eq. 22.

$$\frac{\pi}{6} d^2 (\rho_s - \rho) \omega^2 r \quad (22)$$

In streamline conditions, the opposing viscous force is $3\pi \eta u$, where u is the terminal velocity of the particle. Equating these expressions gives Eq. 23.

$$u = \frac{1}{18} d^2 \frac{(\rho_s - \rho)}{\eta} \omega^2 r \quad (23)$$

The rate of sedimentation is proportional to the radius of the basket and the square of the speed at which it rotates. Centrifugal sedimentors can be divided into a number of types. For operations at very high speeds, the centrifuge bowl is tubular with a length/diameter ratio from 4 to 8. The solids are periodically discharged by scraping the walls of the centrifuge tube. Uses include the cleaning of fats and waxes, the fractionation of blood, and the recovery of viruses.

DRYING

Drying may be defined as the vaporization and removal of water or other liquid from a solution, suspension, or other solid-liquid mixture to form a dry solid. The change of phase from liquid to vapor distinguishes drying from the mechanical methods of separating solids from liquids such as filtration. The latter often precede drying because they offer a cheaper method for removing a large part of the liquid, where applicable.

Adjustment and control of moisture levels by drying, is important in the manufacture and development of pharmaceutical products. Apart from the obvious requirement of dry solids for many operations, drying may be carried out in order to:

- Improve handling characteristics, as in bulk powder filling and other operations involving powder flow, and
- Stabilize moisture-sensitive materials, such as aspirin and ascorbic acid.

Theory

The following terms are employed in discussing drying: humidity and humidity of saturated air, relative humidity, wet-bulb temperature, and adiabatic cooling line. Other terms may be defined as:

- **Moisture Content.** It is usually expressed as weight per unit weight of dry solids.
- **Equilibrium Moisture Content.** If a material is exposed to air at a given temperature and humidity, it will gain or lose moisture until equilibrium is reached. The moisture present at this point is defined as the equilibrium moisture content for the given exposure conditions. At a given temperature, it will vary with the partial pressure of the water vapor in the surrounding atmosphere.

Equilibrium moisture content curves vary greatly with the type of material examined. Insoluble, nonporous materials, such as talc or zinc oxide, have equilibrium moisture contents of almost zero over a wide humidity range. A moisture content between 10 and 15% may be expected for cotton fabrics under normal atmospheric conditions. Drying below the equilibrium moisture content for room conditions may be deliberately undertaken, particularly if the material is unstable in the presence of moisture; subsequent storage becomes important.

The effects of storage after drying also may be assessed from the equilibrium moisture content curves. Storage conditions are not critical for the lactose granulation (19, 20). If the antacid formulation is stored at a relative humidity of only 65% it would, given sufficient time, absorb moisture until the content was 91%. This could be associated with poor flow characteristics and its attendant difficulties during compression.

Evaporation of Water into an Air Stream

The evaporation of moisture into a warm air stream, with the latter providing the latent heat of evaporation, is a common drying mechanism although it is not easily adapted to the recovery of the liquid. In the evaporation from a liquid surface which, with the passage of air, falls to the wet bulb temperature corresponding to the temperature and humidity of the air, the rate at which water vapor is transferred from the saturated layer at the surface to the drying stream is described by Eq. 15,

$$N = \frac{k_g}{RT} (P_{wi} - P_{wa})$$

where P_{wi} is the partial pressure of the water vapor at the surface, P_{wa} is the partial pressure of water vapor in the air, k_g is a mass transfer coefficient, and N the number of moles vapor transferred from unit area in unit time. Rewriting this in terms of the total mass, W , transferred in unit time from the entire drying surface A gives Eq. 6,

$$W = \frac{M_w A}{RT} k_g (P_{wi} - P_{wa})$$

where M_w is the molecular weight of water vapor, R is the gas constant, and T the absolute temperature. The mass transfer coefficient, k_g , is a function of the temperature, the air velocity, and the angle of air incidence. A high velocity or angle of incidence diminishes the thickness of the stationary air layer in contact with the liquid surface and therefore lowers the diffusional resistance.

The rate of evaporation may also be expressed in terms of the heat transferred across the laminar film from the drying gases to the surface, as shown in Eq. 9,

$$Q = hA(T_a - T_s)$$

where Q is the rate of heat transfer, A is the area of the surface, T_a and T_s , are the temperatures of the drying air and the surface, respectively, and h is the heat transfer coefficient. The last is also a function of air velocity and the angle of impingement. If the latent heat of evaporation is λ , this affords a mass transfer rate, W , which is given by Eq. 24.

$$W = \frac{hA}{\lambda}(T_a - T_s) \quad (24)$$

When these conditions pertain to drying, the surface temperature, T_s , which is the wet bulb temperature, is normally much lower than the temperature of the drying gases. This is of great importance in the drying of thermolabile materials. If solids are present in the surface, the rate of evaporation is modified, the overall effect depending on the structure of the solids and the moisture content.

Static Beds of Nonporous Solids

The drying of wet granular beds containing nonporous particles, which are insoluble in the wetting liquid, has been extensively studied. The operation is presented as the relation of moisture content and time of drying in Fig. 5a.

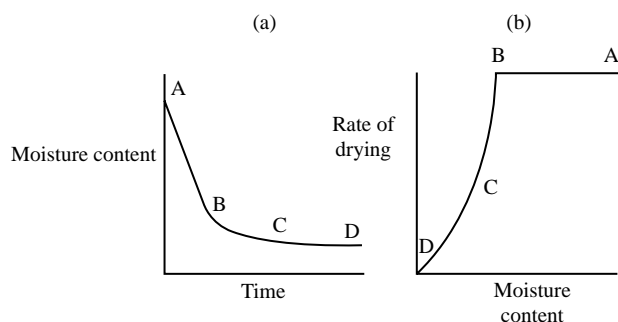


Fig. 5 (a) Relation of moisture content and time of drying; (b) rate of drying and moisture content.

It should be noted that the equilibrium moisture content is approached slowly. A protracted period may be required for the removal of water just above the equilibrium value. This is not justified if a small amount of water can be tolerated in further procession establishing realistic drying requirements.

The data have been converted to a curve relating the rate of drying to moisture content in Fig. 5b. The initial heating period during which equilibrium is established is short and has been omitted from both figures. Assuming that sufficient moisture is initially present, the drying-rate curve exhibits three sections limited by the points A, B, C, and D. In section A-B, called the constant-rate period, moisture is evaporating from a saturated surface at a rate governed by the stationary air film in contact with it. An analogy with evaporation from a plain water surface can therefore be drawn and Equations (11) and (13) apply. The rate of drying during this period depends upon the air temperature, humidity, and speed, which in turn determine the temperature of the saturated surface. Assuming that these are constant, all variables in the drying equations given earlier are fixed, and a constant rate of drying is established which is largely independent of the material being dried. The drying rate is somewhat lower than for a free-water surface and to some extent depends on the particle size of the solids. During the constant-rate period, liquid must be transported to the surface at a rate sufficiently high to maintain saturation. The mechanism of transport is discussed later.

At the end of the constant-rate period B, a break in the drying curve occurs. This point is called the critical moisture content, and a linear fall in the drying rate occurs with further drying. Because section, B-C, is called the first falling-rate period. At and below the critical moisture content, the movement of moisture from the interior is no longer sufficient to saturate the surface. As drying proceeds, moisture reaches the surface at a decreasing rate and the mechanism which controls its transfer influences the rate of drying. Because the surface is no longer saturated, it tends to rise above the wet bulb temperature.

For any material, the critical moisture content decreases as the particle size decreases. Eventually, moisture ceases to reach the surface which becomes dry. The plane of evaporation recedes into the solid, the vapor reaching the surface by diffusion through the pores of the bed. This section is called the second falling-rate period and is controlled by vapor diffusion, a factor which is largely independent of the conditions outside the bed but markedly affected by the particle size due to its influence on the dimensions of pores and channels. During this period, the surface temperature approaches the temperature of the drying air.

Considerable migration of liquid occurs during the constant-rate and first falling-rate periods. Associated with the liquid is any soluble constituent which forms a concentrating solution in the surface layers as drying proceeds. Deposition of these materials takes place when the surface dries. Considerable segregation of soluble elements in the cake can occur, therefore, during drying. These effects have been fully investigated (21).

The Internal Mechanism of Drying

Capillary forces offer a coherent explanation for the drying periods of many materials. If a tapered capillary is filled with water and exposed to a current of air, the meniscus at the smaller end remains stationary while the tube empties from the wider end. A similar situation exists in a wet particulate bed and the phenomenon is explained by the concept of suction potential. A negative pressure exists below the meniscus of a curved liquid surface which is proportional to the surface tension, λ , and inversely proportional to the radius of curvature, r . (The meniscus is assumed to be a part of a hemisphere.) This negative pressure or suction potential may be expressed as the height of liquid, expressed by Eq. 25,

$$h = \frac{2\lambda}{\rho g r} \quad (25)$$

where ρ is the density of the liquid.

The suction potential, h_x , acting at a depth x below the meniscus is given by Eq. 26.

$$h_x = h - x \quad (26)$$

The particles of the bed enclose spaces called pores connected by passages, the narrowest part of which is called the waist. The dimensions of the latter are determined by the size of the surrounding particles and the manner in which they are packed. In a randomly packed bed, pores and waists of varying sizes are found. Thus, the radius of a capillary running through the bed varies continuously. The depletion of water in this network is controlled by the waists because the radii of curvature are smaller and the suction potentials are greater than for the pores. The application of this mechanism has been described fully elsewhere (22, 23).

Through Air-Circulation Drying

If the particles are in a suitable granular form, it is often possible to pass the air stream downward through the bed of solids. Drying follows the pattern described previously, except that each particle or agglomerate behaves as a

drying bed. The surface area exposed to the drying gases is greatly increased, and the drying rates are 10 to 20 times higher than those encountered when air is passed over a free surface.

Methods Involving Movement of the Solid

As an extension of drying by passing the air stream through a static bed of solids, it is possible to project air upward through the bed at a velocity high enough to fluidize the particles. Alternatively, the material may be mechanically subdivided and introduced into the drying stream. Both methods give high drying rates due to high interfacial contact between the drying surfaces and the air stream. Fluidized bed driers and spray driers, respectively, are based on these principles.

Solids Moving Over a Hot Surface

Conditions in which the solids move over a heated surface are employed in tumbling and agitated driers. Drying rates are higher than those obtained in static beds because fresh solids are continually exposed to the hot surface. The heat treatment received by the solid is more uniform.

Batch Driers

Hot air ovens

Ovens operating by passing hot air over the surface of a wet solid which is spread over trays arranged in racks, provide the simplest and cheapest drier. In small installations, the air is passed over electrically heated elements and once through the oven. Larger units may employ steam-heated, finned tubes, and thermal efficiency is improved by recirculating the air. This is controlled by manually set dampers, and a common operating position gives 90% recirculation and 10% bleed-off. The heater bank is placed in such a position that the solids do not receive radiant heat and incoming air may be filtered.

The chief advantage of the hot air oven, apart from its low initial cost, is its versatility. With the exception of dusty solids, materials of almost any physical form may be dried. Thermostatically controlled air temperatures between 40 and 120°C permit heat-sensitive materials to be dried. For small batches this may be the equipment of choice. However, the following characteristics have led to development of other small driers:

- A large floor space is required for the oven and tray loading facilities.

- Labor costs for loading and unloading the oven are high.
- Long drying times, usually of the order of 24 h, are necessary.
- Solvents can be recovered from the air only with difficulty.
- Unless carefully designed, nonuniform distribution of air over the trays results in variations in temperature and drying times within the oven. Variations of $\pm 7^{\circ}\text{C}$ in temperature have been found from location to location during the drying of tablet granules (19). Poor air circulation may permit local saturation and the cessation of drying.

An extensive analysis of tray drying and the effect of operational variables has been given by Shepherd et al. (24).

If the material is of suitable granular form, drying times may be reduced to 1 h or less by passing the air downward through the material laid on mesh trays. The oven in this form is called a batch through-circulation drier.

Vacuum tray driers

Vacuum tray driers offer an alternative method for drying small quantities of material. When scaled up, construction becomes massive to withstand the applied vacuum and cost is further increased by the associated vacuum equipment. Vacuum tray driers are, therefore, only used when a definite advantage over the hot air oven is secured, such as low temperature drying of thermolabile materials or the recovery of solvents from the bed. The exclusion of oxygen may also be advantageous or necessary in some operations.

Heat is usually supplied by passing steam or hot water through hollow shelves. Radiation from the shelf above may cause a significant increase in temperature at the surface of the material if high drying temperatures are used. Drying times are long, usually of the order of 12 to 48 h.

Tumbling driers

The limitations of ovens, particularly with respect to the long drying times, has, where possible, promoted the design and application of other batch driers. The simplest of these is the tumbler drier. Its most common shape is the double cone (19). Operating under vacuum, it provides controlled low temperature drying, the possibility of solvent recovery, and increased drying rates. Heat is supplied to the tumbling charge by contact with the heated shell and heat transfer through the vapor.

A normal charge would be about 60% of the total volume and, for driers 2–7 ft (0.6–2 m) in diameter, drying times of 2–12 h may be expected. In studying the

application of tumbler driers to drying tablet granules, periods of 2–3.5 h were sufficient instead of 18 h required by hot air ovens (25). The mixing and granulating capacity of the tumbling action has suggested that these operations could precede drying in the same apparatus.

Fluid-bed driers

The term “fluidization” is applied to processes in which a loose, porous bed of solids is converted to a fluid system, having the properties of surface leveling, flow, and pressure-depth relationships, by passing the fluid up through the bed.

Fluidized-bed techniques, employing air as the fluidizing medium, have been successfully applied to the drying of solids of the suitable physical form. The high interfacial contact between drying air and solids gives drying rates 10 to 20 times higher than those obtained during tray drying.

Fluidized-bed driers are particularly suitable for granulated materials and are being increasingly used for tablet granulations, providing that product changeover is not too frequent. Machines vary in size, handling up to 250 kg. Drying times, maximum, minimum, and optimum air velocities, air temperature, and the tendency to cake and channel are established experimentally as those cannot be predicted accurately at present.

Agitated batch driers

Agitated batch driers consist of a jacketed cylindrical vessel with agitator blades designed to scrape the bottom and walls. They may operate at atmospheric pressure or under vacuum. Pasty materials that could not be handled in tumbling or fluidized-bed driers, may be successfully dried at rates higher than can be achieved in an oven.

Freeze drying

Freeze drying is an extreme form of vacuum drying in which the solid is frozen and drying takes place by subliming the solid phase (26–30) at low temperatures and pressures. Establishing and maintaining these conditions, together with the low drying rates obtained, constitutes the most expensive method of drying which is only used on a large scale when other methods are inadequate.

Freeze drying is extensively used when rapid decomposition occurs during normal drying. Another application concerns substances that can be dried at high temperatures but are thereby changed in some way.

Freeze drying is theoretically a simple technique. Pure ice exhibits an equilibrium vapor pressure of 4.6 mm Hg (611 Pa) at 0°C and 0.1 mm Hg (13.3 Pa) at -40°C . The vapor pressure of ice containing dissolved substances is, of course, lower. If, however, the pressure above the

frozen solution is less than its equilibrium vapor pressure, the ice sublimes, eventually leaving the solute as a sponge-like residue equal in apparent volume to the original solid.

Continuous Driers

Although many types of continuous driers are available, the scale of the operation for which they are designed is rarely appropriate to pharmaceutical manufacture. As with most continuous operations, the cost is disproportionately high for small units. Spray and drum driers provide an exception, because residence times in the driers are short and thermal degradation is minimized. Under some conditions, freeze drying may be the only practicable alternative.

Spray driers

The solution or suspension to be dried is sprayed into a hot air stream and circulated through a chamber. The dried product may be carried out to cyclone or bag separators or may fall to the bottom of the drying chamber and be expelled through a valve. The chambers are normally cylindrical with a conical bottom although proportions vary widely. The process can be divided into four sections: atomization of the fluid, mixing of the droplets, drying, and finally removal and collection of the dry particles.

In vertical spray driers, the flow of the drying gas may be concurrent or counter-current with respect of the movement of droplets. The movement of the gas is, however, complex and highly turbulent. Good mixing of droplets and gas occurs, and the heat and mass transfer rates are high. In conjunction with the large interfacial area conferred by atomization, these factors give very high evaporation rates. The residence time of a droplet in the drier is only a few seconds (5–30 s). Since the material is at wet bulb temperature for much of this time, high gas temperatures of 150–200°C may be used even with thermolabile materials. Although the temperature of the material rises above the wet-bulb temperature at the end of the process, the drying gas is cooler and the material is almost dry, a condition in which many materials are thermally less sensitive.

Drying is considered to take place by simple evaporation rather than by boiling and it has been observed that a droplet reaches a terminal velocity within about one foot of the atomizer. Beyond this, there is no relative velocity between the droplet and the drying gas unless the former is very large. The droplets may dry to form a solid, spherical particle. If, however, the emerging solids form a skin, internal pressure may inflate the particle and the final dry form will consist of hollow spheres which

may or may not have a blow hole. These xenospheres may also fragment, resulting in a final product of agglomerates of finely divided solids.

The capital and running costs of spray driers are high, but if the scale is sufficiently large, it may provide the cheapest method. When thermolabile materials are dried on a small scale, costs will be 10 to 20 times higher than for oven drying. Air used to dry fine chemicals or food products is heated indirectly, thus reducing thermal efficiency and increasing costs. In some other installations, hot gases from combustion may be used directly.

Drum driers

The drum drier consists of one or two slowly rotating, steam-heated cylinders. These are coated with solution or slurry by means of a dip feed in which the lower portion of the drum is immersed in an agitated trough of feed material or, in the case of some double-drum driers, by feeding the liquor into the gap between the cylinders. Spray and splash feeds are also used. In dip feeding, the hot drum must not boil the liquid in the trough. Drying takes place by simple evaporation rather than by boiling. The dried material is scraped from the drum by a knife at a suitable point.

Drying capacity is influenced by the speed of the drum and the temperature of the feed, which may be preheated. With the double-drum drier, the gap between the cylinders determines the thickness of the film.

Drum driers, like spray driers, are relatively expensive in small sizes and their use in the pharmaceutical industry is largely confined to drying thermolabile materials where the short contact time is advantageous. Drums are normally fabricated from stainless or chrome-plated steel to reduce contamination. The heat treatment to which the solid is subjected is more intense than in spray drying and the physical form of the produce is often less attractive during drying, the liquid approaches its boiling point and the dry solids attain the temperature of the drum surface.

SIZE REDUCTION AND CLASSIFICATION

The theoretical strength of crystalline materials can be calculated from interatomic attractive and repulsive forces. The strength of real materials, however, is found to be many times smaller than the theoretical value. The discrepancy is explained in terms of flaws of various kinds, such as minute fissures or irregularities of lattice structure known as dislocations. These have the capacity to concentrate the stress in the vicinity of the flaw. Failure

may occur at a much lower overall stress than is predicted from the theoretical considerations. Failure occurs with the development of a crack tip which propagates rapidly through the material, penetrating other flaws which may, in turn, produce secondary cracks. The strength of the material depends therefore on the random distribution of flaws and is a statistical quantity varying within fairly wide limits. This concept explains why a material becomes progressively more difficult to grind. Since the probability of containing an effective flaw decreases as the particle size decreases, the strength increases until, with the achievement of faultless domains, the strength of the material equals the theoretical strength. This position is not realized in practice due to complicating factors such as aggregation.

The strength of most materials is greater in compression than in tension. It is therefore unfortunate that technical difficulties prevent the direct application of tensile stresses. The compressive stresses commonly used in comminution equipment do not cause failure directly but generate by distortion sufficient tensile or shear stress to form a crack tip in a region away from the point of primary stress application. This is an inefficient but unavoidable mechanism. Impact and attrition are the other basic modes of stress application. The distinction between impact and compression is referred to later. Attrition, which is commonly employed, is difficult to classify but is probably primarily a shear mechanism.

The deformation and subsequent failure of a brittle material is not only a function of stress but also of the rate at which the stress is applied. Different results may be obtained from slow compressive breaking and impact breaking at the same energy level. Particle shape, size, and size distribution may be affected. In impact breaking, the rate of stress application is so high that the limiting strain energy may be exceeded several times by the suddenness of the operation. The reason is that fracture is time dependent, a lag occurring between the application of maximum stress and failure.

Stress application is further complicated by "free crushing" and "packed crushing" mechanisms. In free crushing, the stress is applied to an unconstrained particle and released when failure occurs. In packed crushing, the application of stress continues on the crushed bed of particles. Although further size reduction occurs, the process is less efficient due to vitiation of energy by the effects of interparticulate friction and stress transmission via particles which do not themselves fracture. This is easily demonstrated when a crystalline material is ground in a pestle and mortar. The fine powder initially produced protects coarser particles. If the material is sieved and

oversize particles are returned, the operation may be completed with far less effort.

Various hypotheses relate the net grinding energy applied to a process and the size reduction achieved. The first, proposed by Karl von Rittinger in 1867, states in Eq. 27 that the energy necessary for size reduction is directly proportional to the increase in surface area,

$$E = k(S_p - S_f) \quad (27)$$

where E is the energy consumed, and S_p and S_f are the surface area of the product and feed materials, respectively. The constant, k , depends on the grinding unit employed and represents the energy consumed in enlarging the surface area by one unit. The relation between surface area and particle size has already been derived, and Eq. 28 may therefore be written,

$$E = k \left(\frac{1}{d_p} - \frac{1}{d_f} \right) \quad (28)$$

where d_f and d_p are the particle sizes of feed and product particles, respectively.

The hypothesis indicates that energy consumption per unit area of new surface produced increases faster than the linear ratio of feed and product dimensions, a phenomenon already noted and explained. The proportionality of net energy input and new surface produced has been confirmed in some grinding operations.

Conversion of grinding energy to surface energy is neglected in Kick's law, promulgated in 1885. It is based on the deformation and brittle failure of elastic bodies and states that the energy required to produce analogous changes of configuration of geometrically similar bodies is proportional to the weight or volume of those bodies. The energy requirements are independent of the initial particle size and depend only on the size reduction ratio. Kick's law predicts lower energies than the relation proposed by Rittinger. The theory, however, demands that the resistance to crushing does not change with particle size. The role of flaws present in real materials is not considered, with the result that the energy required for fine grinding, when the apparent strength may have greatly risen, is underestimated.

A third theory of comminution gives results intermediate between the predictions of the laws of Kick and Rittinger (31). It rests upon three principles: the first states that any divided material must have a positive energy register. This can only be zero when the particle size becomes infinite. The input energy, E , for any size reduction process then equals the product energy register minus the feed energy register. The energy associated with a powder increases as the particle size decreases, and it

may be assumed that the energy register is inversely proportional to the particle size to an exponent, n . Hence, Eq. 29 is valid.

$$E = E_p - E_f = \frac{K}{d_p^n} - \frac{K}{d_f^n} \quad (29)$$

The second principle of this theory (31) assigns to n a value of i , stating that “the total work useful in breaking, which has been applied to a stated weight of an homogeneous material, is inversely proportional to the square root of the diameter of the product particles.”

The third principle states that breakage of the material is determined by the flaw structure. This aspect of size reduction has already been discussed.

A modification to Kick’s law, sometimes known as the fourth law of comminution, has also been proposed (32). For its discussion, the reader is referred to the original paper.

An empirical, but realistic approach to mill efficiency is gained through experiments in which the energy consumed and size reduction achieved are compared with values obtained in a laboratory test operating under free crushing conditions. All energy supplied in the latter is available for crushing and the test is assumed to be 100% efficient. Both slow crushing and impact tests are used. A large number of single particles may be simultaneously crushed and the work done is measured (33). The latter is related to the size reduction. Similar measurements can be made during practical milling, expressing the efficiency of the process as a percentage of the free crushing value. On this basis, the approximate efficiency of the roll crusher is 80%, of the swing hammer mill 40%, of the ball mill 10%, and of the fluid energy mill only 1%.

The Operation of Mills

Heywood (34) has stated that any type of crushing or grinding machine exhibits optimal comminution conditions for which the ratio of the energy to new surface is minimal. If finer grinding is attempted in such a machine, the ratio is increased. Mills may thus become grossly inefficient if called upon to grind at a size for which they were not designed. A limited size reduction ratio is imposed upon a single operation, larger ratios being obtained by the adoption of several stages, each employing a suitable mill. The fluid energy mill, which presents a size reduction ratio of up to 400, is exceptional.

A low retention time is inherent in free crushing machines. Little overgrinding takes place and the production of excessive undersize material or fines is avoided. Protracted milling times are found with many

low-speed mills, with the result that considerable overgrinding takes place. Accumulation of product particles within the mill reduces the effectiveness of breaking stresses and the efficiency of milling progressively decreases.

Dry and wet grinding

Between the approximate limits of 5 and 50% moisture, materials cake and do not flow. Both factors oppose effective grinding. Dry grinding is carried out at low moisture contents, the upper limit depending on the nature of the material. Although 5% or more moisture may be permissible for vegetable drugs, it would prove excessive during the milling of a coarse, impervious solid.

Wet grinding is a common procedure when a fluid suspension is required and drying, which would provide a significant drawback, is unnecessary. An excellent dispersion can be produced simultaneously, which in some operations provides the primary objective, size reduction being of secondary importance. Wet grinding also may be adopted when the size reduction achieved during dry grinding is prematurely linked by aggregation.

Certain general advantages are secured during wet grinding, including increased mill capacity, a lower energy consumption, the elimination of hazards from dust, and easier handling of materials. The principal disadvantage, apart from the possible inclusion of a drying stage, is the increased wear of the grinding medium.

Temperature sensitivity

Care must be exercised during the milling of temperature-sensitive materials, especially for a very fine product; caking results if the softening point is exceeded. Materials may be chilled before grinding or facilities provided for cooling the mill during grinding. Waxy solids can be successfully ground with dry ice, the low temperatures conferring brittle characteristics on the material. Chemical degradation may occur at high grinding temperatures. Oxidative changes can be prevented by grinding in an inert atmosphere such as nitrogen.

Structural changes

Several examples of change of physical structure during very fine grinding have been reported, for example, changes in the crystal form of calcium carbonate after ball milling (35), distortion of the kaolinite lattice (36), and formation of various barbiturate polymorphs (37). Changes such as these could affect solubility and other physical characteristics which, in turn, might influence formulation and therapeutic value.

Dust hazards

Hazards from dust may become acute during dry grinding. Extremely potent materials require dust proofing of machines and dust-proof clothing and masks for operators. Danger may also arise from the explosive nature of many dusts.

Grinding Equipment

The following equipment are in regular use for dry-grinding pharmaceutical materials: edge- and end-runner mills, hammer mills, pin mills, ball mills, vibratory mills, fluid energy mills, colloid mills, and roller mills.

Classification or Size Separation

Although a number of particle properties can be used to classify a powder, only two are important. The first is based on the ability of a particle to pass through an aperture. This is sieving or screening. The second employs the drag forces on a particle moving through a fluid. The term “classification” is sometimes restricted to this method of separation but here the terms “elutriation” and “sedimentation” are used. In general, screening is applied to the separation of coarse particles and sedimentation to the separation of fine particles.

Sieving and screening

Sieves and screens are widely used for the classification of relatively coarse materials. For very large particles (>0.5 in.) a robust plate perforated with holes is used. However, the pharmaceutical applications of screening are for much smaller particles and screens are in the form of woven meshes. Unless special methods are used to prevent clogging and powder aggregation, the lower useful limit is in a cloth woven with 200 mesh/in. (70–80 μm). Fine screens of this type are extremely fragile and must be used with great care.

As the scale of the operation increases it becomes, in general, less precise. For continuous screening, the feed material is made to move across the screen to a point of discharge. The residence time on the screen is usually short and many undersize particles traverse it without falling through. With an increase of sieving area, the meshes become more fragile and the finest meshes must be supported with a coarser wire. An example of a large-scale separator utilizes a circular screen, up to 5 ft (1.5 m) in diameter, vibrated in a horizontal plane, the gyratory movement being imparted by an out-of-balance fly wheel connected to the assembly. In other machines, the mesh is rectangular and inclined at a shallow angle (5–30°).

A gyratory movement is developed and the material to be classified is fed to the top. These machines may bear more than one deck, thus allowing the separation of the powder into several fractions at one time.

Elutriation and sedimentation

The simplest classifier is a rising current of fluid in which the particles are suspended. In this case, the force opposing the upward drag is gravitational. If the opposition develops a terminal velocity higher than the current speed, the particle falls. This is the principle of elutriation; the particle size d at which the separation is made follows from a rearrangement of Eq. 29 for conditions in which Stokes' law is valid; it is given by Eq. 30,

$$d = \sqrt{\frac{18\eta\mu}{(\rho_s - \rho)g}} \quad (30)$$

where $\rho_s - \rho$ is the density difference between solid and fluid, η is the viscosity of the fluid, and μ is the speed of the upward current.

In practice, fluctuations in flow conditions due to natural convection and a violation of the conditions for which Stokes' law is valid, blur the point of separation.

The centrifuge is normally operated to completely separate two phases. If, however, the rate at which the feed passes through does not allow all particles to settle, the action of a classifier is developed. This is illustrated by a solid-bowl centrifuge which consists of a steel shell in the form of a frustum mounted horizontally. It contains a conveying screw at the wall which rotates at a slightly higher speed than the shell. Particles that settle at the wall are conveyed to the narrow end of the shell and discharged. Fine particles are entrained with the overflow to the other end. Further details of this and other centrifugal classifiers have been given by Treasure (38).

MIXING

Mixing has been defined (39) as an operation in which two or more ingredients in separate or roughly mixed condition are treated so that each particle of any one ingredient is adjacent to a particle of each of the other ingredients, as nearly as possible. The term “blending” is synonymous and “segregation” or “demixing” is the opposite.

Mixing has been classified (40) as follows:

- *Positive mixing* which applies to systems that, given time, would spontaneously and completely mix. Examples are provided by two gases or two miscible

liquids; mixing apparatus is used on such systems to accelerate mixing.

- *Negative mixing* is demonstrated by suspensions of solids in liquids. Any two-phase system in which the phases differ in density separates unless continuously agitated.
- *Neutral mixing* occurs when neither mixing nor demixing takes place unless the system is acted upon by a system of forces. Examples are found in the mixing of solids and of solids with liquids when the concentration of the former is high.

Theoretical knowledge is, however, insufficient to predict the performance of mixers. More commonly, choice is based upon broad empirical principles which are supported by practical tests.

Mixing of Solids

The mixing of all systems of matter involves a relative displacement of the particles, whether they are molecules, globules, or small crystals, until a state of maximum disorder is created and a completely random arrangement is achieved.

In 1953, Lacey (41) showed that the variation in the composition of samples drawn from a random mixture of two materials could be expressed by Eq. 31,

$$s = \sqrt{\frac{p(1-p)}{n}} \quad (31)$$

where s is the standard deviation of the samples, p is the proportion of one component, and n is the number of particles in the sample. The relation requires that the two components are alike in particle size, shape, and density and only can be distinguished by some neutral property, such as color. If very many samples are withdrawn from a mixture of equal parts of two materials, each containing a given number of particles, the results of analysis can be presented in the form of a frequency curve in which the samples are normally distributed around the mean content of the mixture, and 99.7% of the samples will fall within the limits $p = 0.5 \pm 3\sigma$. The standard deviation of the samples is inversely proportional to the square root of the number of particles in a sample. If the particle size is reduced to the extent that the same weight of sample contains four times as many particles, the standard deviation is halved.

In a critical examination of pharmaceutical mixing, Train (42) showed that samples of a random mixture of equal parts A and B must contain at least 800 particles if 997 out of every 1000 samples were to lie between $\pm 10\%$ of the stated composition, that is, the proportion, p , of

$A = 0.5 \pm 0.05$, where $\sigma = 0.05/3$. If limits of $\pm 1\%$ were substituted, 90,000 particles must be present in each sample. The true standard deviation is given by σ . The standard deviation estimated by the withdrawal of a number of samples is denoted as s .

If, instead of equal parts A and B, the proportion of an active ingredient, A, in the mixture was 0.1 (10%), imposition of limits of $\pm 10\%$ (in 997 cases out of 1000) requires that each sample shall contain over 8000 particles. If the proportion of active constituent is 0.01, or 1%, a figure of 90,000 particles per sample is obtained, and if the limits are reduced to $\pm 1\%$, the active constituent is 0.01, or 1% a figure of 90,000 particles per sample is obtained, and if the limits are reduced to $\pm 1\%$, the figure is 9×10^6 .

The theoretical derivation of these results is based on component particles which vary in size, shape, and density. This condition is not encountered in the practical mixing of solids and, as described later, any of these factors may prevent the formation of a random mixture. The value of the number of particles per sample derived in any example must therefore be raised if the limits given are to be maintained.

As already shown, a series of samples drawn from a random mix exhibits a standard deviation of s_r . An index of mixing, M , suggested by Lacey (43) is given by Eq. 32,

$$M = \frac{s_r}{s} \quad (32)$$

where s is the standard deviation of samples drawn from the mixture under examination. This approaches unity as mixing is completed. Eq. 33 has been suggested,

$$M = \frac{s_0 - s}{s_0^2 - s_r^2} \quad (33)$$

where s_0 is the standard deviation of samples drawn from the unmixed materials. It is equal to $p(1-p)$, where p is the proportion of the component in the mix. It has been modified (43) to Eq. 34, using the variance of the samples,

$$M = \frac{s^2 - s_r^2}{s_0^2 - s_r^2} \quad (34)$$

This is a fundamental equation for expressing the state of the mixture, the index M varying from zero to one.

The binomial and Poisson distributions have also been used to examine the state of a mixture. If the proportion of black particles in a random mixture of black and white particles is p , the probability, $P(x)$, of obtaining x black particles in a sample of n particles is given by Eq. 35.

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (35)$$

If p is small (<0.15) and n is large, the Poisson distribution can be used, applying Eq. 36,

$$P(x) = e^{-m} \frac{m^x}{x!} \quad (36)$$

where $m = np$, the mean number of black particles in the samples of n particles. This relation may be used in an assessment of dry mixing equipment (44). If m is greater than 20 and more than 10 samples are taken, then:

- About 10 of the samples have the number of black particle's outside the limits $m \pm 1.7\sqrt{m}$,
- About 5% of the samples have the number of black particles outside the limits $m \pm 2.0\sqrt{m}$, and,
- About 1% of the samples has the number of black particles outside the limits $m \pm 2.6\sqrt{m}$.

Mechanism of mixing and demixing

The randomization of particles by relative movement, one to another, is achieved by the following mechanisms:

- *Convective mixing*, where groups of adjacent particles are transferred from one location in the mass to another.
- *Diffusive mixing*, where the particles are distributed over a freshly developing surface, and
- *Shear mixing*, where slip planes are set up within the mass.

Convective mixing predominates in machines utilizing a mixing element moving in a stationary container, for example, the horizontal ribbon mixer. Groups of adjacent particles are moved from one position to another, steadily decreasing the scale of segregation.

Diffusive mixing predominates in tumbler mixers. The material is tumbled as it is lifted past its angle of repose. Mixing occurs when a particle changes its path of circulation through a collision or by being trapped in voids presented by another layer of particles.

Shear mixing occurs when forces acting on the particles induce the formation of a slip place, resulting in relative displacement of two regions. Shear mixing occurs, for example, in the rearrangement of shapes as the main charge falls from end to end in a double cone mixer. Train (42) has stressed the importance of expansion or dilation of the material so that shear forces may be effective. A practical corollary is that efficiency will be reduced if the machine is overfilled.

As long as one type of particle is not preferentially caught, random mixing eventually occurs in the radial plane. If, however, one component is smaller, denser, or has certain shape characteristics, it is preferentially trapped and moves into the lower layers of the mixing zone until it finally concentrates as a central core running

the length of the mixer. Similar effects occur in axial mixing, and the final shape of the segregated zone formed under the influence of axial and radial movement depends on the flow properties of the material. Similar effects have been reported with a double-cone blender (44). Segregation also occurs with materials dumped from the mixer.

Mixing rate

Because mixing is a process of achieving uniform randomness, the rate of mixing is proportional to the amount of mixing still to be done. If, at the start a particle changes its path of circulation, it is most likely to find itself in a different environment. The mixing rate is therefore high. At the end of the process, the particle is less likely to find a different environment, and such a change gives no useful mixing. Fewer mixing events take place, and the mixing rate finally reaches zero. It can be represented for any mixing mechanism by Eq. 37,

$$\frac{dM}{dt} = k(1 - M) \quad (37)$$

where M , the index of mixing, has already been defined. Integration of Equation (37) gives Eq. 38.

$$M = 1 - e^{-kt} \quad (38)$$

The rate constant, k , depends on the physical nature of the materials being mixed and on the geometry and operation of the mixer.

Mixing Machines

Trough and ribbon mixers

A simple trough mixer consists of a semicircular trough in which an impeller, such as a number of paddles mounted at diverse angles on a shaft running the length of the trough, rotates, lifting and distributing the material in an irregular manner. Convective and shear mixing occurs, as well as some fine-scale diffusive mixing when the impeller lifts material clear of the main charge.

The ribbon mixer employs a ribbon-like conveying scroll. The helix, which may be continuous or interrupted, is rotated in a semicircular trough and mixing again occurs through convection and shear, giving rapid coarse-scale dispersion. Two ribbons set to convey material in opposite directions are frequently fitted to the shaft. Although little axial mixing in the vicinity of the shaft occurs, mixtures with high homogeneity can be produced by prolonged mixing, even when components differ in particle size, shape, or density or tend to aggregate.

Tumbler mixers

Tumbler mixers operate primarily by a diffusive mechanism; their use is confined to freeflowing and granular materials. The mild forces are employed, which preclude the mixing of materials that aggregate strongly, allow friable materials to be handled satisfactorily. The more elaborate geometrical forms are most commonly used because movement of material in all planes, which is necessary for rapid overall mixing, is induced. Internal baffles and lifter blades may also be incorporated. For example, axial movement of material along the length of a simple drum mixer is slow and can be enhanced by these methods.

Mixing of Liquids

Miscible liquids are most commonly mixed by impellers rotating in tanks, including paddles, propellers, and turbines. All the material should pass through the impeller zone at frequent intervals of time, the design of the mixer preventing the formation of “dead” zones. The turbulent, high velocity flow of liquid from the impeller causes mixing by projecting eddies into, and entraining liquid from, the neighboring zones. The thin ribbons of one component in another rapidly become diffuse and finally disappear through molecular diffusion.

The flow pattern may be analyzed in terms of its three components of motion:

- Radial flow, in a direction perpendicular to the impeller shaft.
- Longitudinal or axial flow, in a direction parallel to the shaft, and
- Tangential flow, in which the liquid follows a circular path around the shaft.

A satisfactory flow pattern depends on the correct balance of these components. In a cylindrical tank, radial flow gives rise to axial flow by reaction at the wall of the tank. Tangential flow receives no such modification. Its predominance as laminar flow circulation supports stratification at various levels. Furthermore, a vortex is created at the surface of the liquid which may penetrate to the impeller, causing air to be dispersed in the liquid. In general, tangential flow should be minimized by moving the impeller to an off-center position, thus destroying the symmetry of the mixer, or by modification of the flow pattern by means of baffles. Tanks with vertical agitators may be baffled by one, two, or more strips mounted vertically on or just away from the vessel wall. These reduce but do not eliminate tangential flow, whereas little modification of radial and axial flow occurs. Baffles produce additional turbulence.

Paddle mixers

For a simple paddle, with upper and lower blades, suitable for mixing miscible liquids of low viscosity a tangential flow pattern predominates with zones of turbulence to the rear of the blades (10–100 rpm). The gate paddle is suitable for mixing liquids of higher viscosity and the anchor paddle with low clearance between pan and blade is useful for working across a heat transfer surface. Stationary paddles intermeshing with the moving element suppress swirling in the mixer. In other examples, baffles are also necessary. Unless paddle blades are pitched, poor axial turnover of the liquid occurs. Paddles are therefore not suitable for mixtures that separate.

Propeller mixers

Propellers are commonly used for mixing miscible and immiscible liquids of low viscosity. The marine propeller is typical of the group. High speed rotation (400–1500 rpm) of the relatively small element provides high shear rates in the vicinity of the impeller and a flow pattern with mainly axial and tangential components. They may be used in unbaffled tanks when mounted in an off-center position or inclined from the vertical. In large-scale operations, horizontal mounting in the side of the vessel is frequently used.

Turbines

Turbine designs are intermediate between paddles and propellers. Turbines are effective mixers over a wide viscosity range and provide a very versatile mixing tool. The ratio of radial to tangential flow, the predominating parameters with this impeller, increases as the operating speed increases. Pitched-blade turbines are sometimes used to increase axial flow. Baffles must be used to limit swirling unless the turbine is shrouded. This impeller produces a discharge with no tangential component.

STERILIZATION

Sterilization processes do not result in a product that can be described as absolutely sterile or nonsterile inasmuch as the process is a statistical phenomenon. A variety of techniques are available (45), including heat, radiation, ethylene oxide sterilization, and sterile filtration.

Thermal Sterilization

The amount of heat required to sterilize depends upon the magnitude (T), duration (t), and amount of moisture

present, $t \propto 1/T$. For example, heat coagulates protein in the living cell. The temperature required for this phenomenon to occur is inversely proportional to the moisture present.

Dry heat

Relatively stable substances that resist degradation at high temperatures ($>140^\circ\text{C}$) are suitable candidates for dry heat sterilization. A 2-h exposure at 180°C or 45 min at 260°C kills spores as well as vegetative forms of micro-organisms. These exposure periods do not include the lag time from loading of the oven until sterilization temperature is reached. The lag time depends on the geometry and operating features of the oven and the characteristics of the load.

Both natural and forced-convection oven types can be employed; they have been described in the section on drying. The forced-convection oven offers the advantages of uniformity of heat distribution and reduction in lag time in comparison with the natural-convection system. The dry-heat method is reserved almost exclusively for glass or metal as other materials char (cellulose), oxidize (rubber), or melt (plastic) at these temperatures.

Moist heat

Moist heat offers the advantage of greater effectiveness at low temperatures. The thermal capacity of steam is much greater than that of hot air. Spores and vegetative forms of bacteria may be effectively destroyed in an autoclave employing steam (121°C) under pressure (15 psig) for 20 min or (27 psig at 132°C), for 3 min. The lag time to complete exposure of the material to be sterilized is important.

Radiation

Ultraviolet light is frequently employed to reduce airborne microbial contamination. Surface sterilization is usually achieved by employing a mercury vapor lamp with an emitted light of 253.7 nm.

Radiation sterilization includes the use of the ionizing radiation of x-rays and gamma-rays. The former are derived from bombardment of a heavy metal target with electrons. Gamma-rays are obtained from atomic nucleus decay from excited to ground state.

The energy evolved from radiation can be equated to photon behavior where $E = h\nu$ and $\nu = C/\lambda$, (E and ν are the energy and frequency of a photon, respectively), h is Planck's constant, and C and λ are the speed and wavelength of light, respectively. The energy absorbed from the radiation sources equates to the dose.

$$1 \text{ rad} = 100 \text{ erg/g of material absorbing}$$

$$= 6.24 \times 10^{13} \text{ eV/g}$$

$$= 2.4 \times 10^{-6} \text{ cal/g} (10 \times 10^{-6} \text{ J/g})$$

There are a variety of radiation sources. ^{60}Co decays to ^{59}Co in the core of a nuclear reactor to emit two photons (1.17 and 1.33 MeV) and an electron (0.31 MeV). The half-time for decay is 5.3 years. ^{137}Cs decays emitting one photon (0.661 MeV). Cesium has a 33-year half-life. An electron beam can be accelerated to an energy equivalent of 5–10 MeV. At energies below 5 MeV, penetration is insufficient for sterilization. Depth of penetration can be correlated with energy levels; for example, materials with density equivalent to water ($\rho = 1 \text{ g/cm}^3$) are penetrated 0.5 cm/MeV. ^{60}Co gives rise to radiation that penetrates 30 cm through water. Accelerating electrons have a high dose rate and exposure is only required for seconds. ^{60}Co has a lower dose rate, and an exposure for hours is required.

Ionizing radiation arises from the photoelectric effect, the Compton effect, or ion pair production. Gamma radiation causes local and intense damage and may break chemical bonds. The primary target is the deoxyribonucleic acid (DNA) of the micro-organism. In addition, free radicals may be formed, such as peroxides that result in intracellular and extracellular peroxides by a chain reaction that causes damage.

Resistance to damage

Damage depends on the amount of energy absorbed relative to the number and resistance of the micro-organisms being irradiated. Unicellular organisms have greater resistance than multicellular ones. Gram-positive bacteria have greater resistance than gram-negative bacteria. Finally, bacterial spores have greater resistance than vegetative forms. Viruses are more resistant than bacteria. The energy required to reduce the population of viruses by 90% (D value), is 0.5 Mrad (5 mGy). Fungi are equivalent to bacterial spores in their resistance.

In order to evaluate the dose, a number of parameters must be known. What magnitude of source (e.g., ^{60}Co) is available? A typical source ranges from 500,000 to 2×10^6 Curies (Ci) where 1 Ci is 3.7×10^{10} disintegrations per second. The product geometry and the speed of the conveyor carrying it to the source must be known. The dose can be evaluated by a variety of dosimetric techniques. In bulk or ampoules containing liquids, ferric ammonium sulfate and ceric sulfate can be used and the absorbance change evaluated by UV spectrophotometry; however, this is only accurate for ^{60}Co and ^{137}Cs .

Radiochromic solids can be utilized and evaluated by visible spectrophotometry. Amber and red polymethyl methacrylate are used to evaluate 0.1–1.0 Mrad and 0.5–5.0 Mrad, respectively. Nylon film is examined for opacity following exposure and may be used to evaluate exposures of 0.1–5.0 Mrad.

Validation requires the determination of the bioburden and the *D* value. These represent the dose required to achieve sterilization and the estimated dose. If low *D* values are obtained, the dose may be regarded as overkill. *Bacillus pumulis* exhibits inherently high resistance to gamma-ionization radiation (*D* values 0.15–0.22 Mrad). The FDA prefers a 12-log reduction in microorganisms. The dose required is approximately 2.6 Mrad.

Ethylene Oxide

Ethylene oxide (bp, 10.8°C) is a gaseous alkylating agent. It alkylates proteins and ribonucleic and deoxyribonucleic acid in micro-organisms. It replaces labile hydrogen with hydroxyethyl groups. Ethylene oxide is utilized as a surface sterilant. Bulk crystalline materials can occlude vegetative bacterial cells or spores with crystals. Consequently, ethylene oxide does not reach them. The final step prior to sterilization is an aseptic recrystallization step.

Ethylene oxide is a colorless gas with an aromatic odor. The threshold limit for the odor is 700 ppm. The OSHA specification for worker exposure is 10 ppm. The toxicity of ethylene oxide is similar to that of ammonia. It causes conjunctival and respiratory irritation, dizziness, headaches, and vomiting. It is known to be mutagenic and may be carcinogenic. By-products include ethylene glycol (bp, 198.9°C) and ethylene chlorhydrin (bp, 128.4°C). Pure ethylene oxide is flammable and explosive. It is generally mixed with propellant (88:12) or carbon dioxide (90:10). Ethylene oxide polymerizes in the liquid state in 90–120 days. In this form it may plug lines or deposit polymerized sludge.

Ethylene oxide inactivates all micro-organisms. The sterilizing rate depends upon its concentration, the temperature, the duration of exposure, and the water content of the micro-organism. Inactivation follows classical first-order kinetics and is irreversible. Relative humidity is synergistic, at 30–60% the micro-organism hydrates. The water acts as a vehicle to transport the gas through polyethylene and polypropylene. Polystyrene traps ethylene oxide and dissipates it over years and thus is not appropriate for ethylene oxide sterilization. Temperatures of 40–60°C are suitable for heat-sensitive articles. Cycle times are longer at low temperatures, relative humidities, or ethylene oxide concentrations.

Generally, concentrations of 350–700 mg/ml are employed; cycle times vary from 4 to 12 h.

Following sterilization the load is degassed by a dynamic process wherein filtered air is passed over the product for 12–72 h. Degassing usually takes place in the treatment chamber but may be moved to a sterile facility. The process is monitored using *Bacillus subtilis* var. *niger* as a biological indicator, commercially available as spore strips (10^6 spores per strip). In addition, the load is probed with thermocouples during validation. The gaseous mixture is sampled at different points in the sterilizer for gas chromatographic analysis.

Sterile Filtration

Several filter geometries are available for sterile filtration. They consist of flat membranes in a stainless steel press (<293 mm), pleated membranes housed in stainless steel cartridges, and stacked plates in the form of flat segments of membrane filters.

Matrix filters consist of fibers with pores having a depth up to 120 μm . Cellulose nitrate may be dissolved in a highly volatile solvent, such as amyl acetate, ether, or dioxane. A gel-forming solvent, acetone, ethanol, or propanol, may be added. The mixture is poured on a flat plate and placed in a controlled-temperature environment to dry. Pore size is dependent on the concentration of the gel-forming solvent. A number of other substances may be used as filter material, including cellulose, acetate and butyrate, polyamides (nylon), polysulfones, fluorocarbons (Durapore membranes), polyvinylidene difluoride (hydrophobic), or surfaces modified with organic amides (hydrophilic), acrylic polymers, or polyvinyl chloride. To make some membranes hydrophilic, surfactants may be added including Tween 80, Triton X-100, hydroxypropyl cellulose, or glycerol. Sieve filters are made of polycarbonate (Nucleopore, 10 μm thick). Collimated uranium fission products form nucleation tracks in film. Exposure to chemical etching determines the pore size.

Adsorption and screening

Most membrane filters, when wetted, have a negative charge. Bacteria have a similar negative charge and do not necessarily remain on the filter. Filters with other characteristics can be selected under these circumstances. Positively charged (AMF Zeta Plus Membrane) or protein- and peptide-adsorbing (Pall Posidyne Nylon 66) filters can be selected.

Ionic strength, pH, pressure, and flow rate affect particle adsorption. The flow rate through a filter is expressed by Eq. 39,

$$Q = \frac{C_i A P}{V} \quad (39)$$

where C_i is the inherent resistance of the filter to flow (a function of void volumes), A is the surface area, P is the pressure, and V is the viscosity. Filters are rated according to nominal pore size and absolute pore size (the largest pore in the filter); this recognizes that a pore size distribution exists.

Filter integrity

The filter integrity can be evaluated by a number of techniques. The destructive test involves filtering a suspension of bacterial cells (*Pseudomonas diminuta*, $0.3 \times 1 \mu\text{m}$, through a $0.2 \mu\text{m}$ -filter. If 6 L of suspension containing 1×10^7 organisms per mL are passed through a $1\text{-}\mu\text{m}$ filter, there should be no microorganism and an 8-log reduction would have occurred. The bubble-point test assumes that pores can be characterized as capillaries. When totally wetted, all the capillaries should be full of water or solution. The pore length is generally much greater than the diameter. Pressure is applied to the wetted filter. The bubble-point pressure, P , may be described by Eq. 40,

$$P = \frac{4\gamma \cos \Theta}{D} \quad (40)$$

where γ is the surface tension (72 dynes/cm² or 7.2 Pa), θ is the contact angle, and D is the diameter of capillary. The bubble-point test is performed before and after sterile filtration.

A specified area of filter must be soaked in a specified volume of product for a designated time. The accelerated stability of active ingredients at 40–60°C for 60 days must be established prior to the selection of a filter for a particular purpose. The extent of damage, and the nature and quantity of extractables and their potency have to be evaluated.

EXTRACTION AND LEACHING

Leaching or solid–liquid extraction are terms that describe the extraction of soluble constituents from a solid or semisolid by means of suitable solvents. The process, which is used whenever tea or coffee is made, is an important stage in the production of many fine chemicals found naturally in animal and vegetable tissue. Examples are found in the extraction of fixed oils from seeds, in the preparation of alkaloids, such as strychnine from *Nux vomica* beans or quinine from Cinchona bark; and in the isolation of enzymes, such as rennin, and hormones, such as

insulin, from animal sources. In the past, a wider importance attended the process because the products of simple extraction procedures, known as galenicals, formed the major part of the ingredients used to fulfill a doctor's prescription.

Whatever the scale of the extraction, leaching is performed in one of two ways. In the first, the raw material is placed in a vessel, forming a permeable bed through which the solvent or menstrum percolates. The wanted constituents are dissolved, and the solution issues from the bottom of the bed. This liquid is sometimes called the miscella and the exhausted solids, the marc. The process is called leaching by percolation. The second process employs immersion and consists of immersing the solid in the solvent and stirring. After a suitable period of time, solid and liquid are separated.

Percolation

Coarsely ground material is placed in the body of the extractor which may be jacketed for control of the extraction temperature. The packing must be even or the solvent flows preferentially through a limited volume of the bed and leaching is inefficient. In large extractors, channeling is prevented or reduced by horizontal, perforated plates placed at intervals in the bed; these redistribute the percolating liquid.

Solvent inhibition swells dried materials and reduces the permeability of the bed. This is most marked with aqueous solvents. If swelling occurs, it is necessary to moisten the material with water or with the solvent before it is packed into the extractor.

Immersion

In pharmaceutical processes, leaching by immersion is carried out in simple tanks which may be agitated by a turbine or paddle. If the solids are adequately suspended, intimate contact between the phases promotes efficient extraction. Incomplete extraction due to channeling is avoided and difficulties due to swelling do not arise. Problems arise, however, in the subsequent separation of the phases. The materials to which leaching by immersion is applied are normally either finely divided or coarse and compressible. When agitation ceases, the solids settle and the leach liquid can be siphoned or pumped off by lines suitably placed in the tank. The sediment, however, contains a large volume of the leach liquid which must be recovered by resuspending the solids in fresh solvent, allowing the solids to sediment and decanting the supernatant liquid. Cake filtration provides an alternative

method of separation. The leach liquid remaining in the cake is displaced by passing a wash liquid. In some cases, a filter press may be used for both extraction and separation.

Solvent

The ideal solvent is cheap, nontoxic, and nonflammable. It is highly selective, dissolving only the wanted constituents of the solid. It should have a low viscosity, allowing easy movement through a bed of solids, and, if the resulting solution is to be concentrated by evaporation, have a high vapor pressure. Water and alcohol, and mixtures of the two, are widely used. Both, however, are nonselective, leaching varying proportions of gums, mucilages, and other unwanted components. Most of the tinctures and liquid extracts used in pharmacy are simple, impure extracts made with, water or mixtures of water and alcohol. Acidified or alkaline mixtures of water and alcohol are used to extract insulin from comminuted pancreas. A more selective extraction is given by petroleum solvents, benzene, and related solvents. In the preparation of many pure alkaloids, the powdered material is moistened with an alkaline solution, packed into a bed, and leached with petroleum. Subsequent purification by fractional crystallization is facilitated by the absence of gums. Acetone and chlorinated hydrocarbons also find applications in leaching. In some cases, specific properties of the wanted constituents may suggest a particular solvent. Eugenol, for example, can be readily extracted from cloves with a solution of potassium hydroxide.

Leaching Rate

Whatever method is adopted, leaching consists of a number of consecutive diffusional or mass transfer processes. The solvent first penetrates the raw material and dissolves the soluble elements. These diffuse in the opposite direction to the surface of the solid matrix and through the liquid layers at its surface to reach the bulk solution. These processes are under the influence of an overall concentration gradient, the concentration being lowest in the bulk solution. Any of these processes may be responsible for limiting the rate at which leaching proceeds. In pharmaceutical leaching, however, the solid matrix is usually cellular, a structure which normally offers the highest diffusional resistance. The complexity of such structures does not permit a strict analysis of the processes of mass transfer. Nevertheless, the simple diffusional concepts expressed in Fick's law suggest that the following factors influence the leaching rate:

- The size distribution of the leached particles,
- The temperature of leaching,
- The physical properties of solvent, and
- The relative movement imposed upon the solids and the liquid.

Size and Size Distribution of the Solid Particles

The particle size of the solids determines the distance which solvent and solute must diffuse within the solid matrix. Since this distance offers the major diffusional resistance, its reduction by comminution raises the rate of leaching, the concentration gradient being effectively increased. In addition, the inverse relationship between particle size and surface area requires an increase in the area of contact between the matrix and the surrounding liquid. Transfer of solute at this boundary is therefore facilitated. In leaching by immersion, a further advantage conferred by size reduction is the ease with which finer particles are suspended. Finally, extensive cell rupture occurs during grinding, allowing more direct contact between solvent and solute and more rapid dissolution and diffusion.

Other factors, however, operate against size reduction. Leaching by percolation demands the formation of a permeable bed. Low permeability gives low flow rates and low extraction rates. Permeability is a complex function of both particle size and porosity, the former determining how a given void space is to be disposed within the bed. The disposition of the void space consists of few channels of relatively large diameter, that is, a bed of high permeability, if the particle size is large. In leaching by immersion, the difficulties of separating solid and liquid increase as the particle size decreases.

The opposition of the factors suggests an optimum particle size for any particular extraction. This is determined to some extent by the physical nature of the solids. A dense, woody structure would be extracted as a fine powder. An example is given by the root of *Ipecacuanha*. A leafy structure, on the other hand, would be more satisfactorily leached as a coarse powder.

Both porosity and permeability are influenced by the particle size distribution. A high porosity is secured if the distribution is narrow. Small particles may otherwise fill the interstices created by the contact of larger particles. After grinding it is often necessary, therefore, to classify the product and remove undersize material. This material would then be bulked with the fines from other batches and extracted separately. A further advantage arising from a narrow size distribution is even packing and the creation of a regular system of pores and waists. This promotes even movement of solvent and solution through the bed.

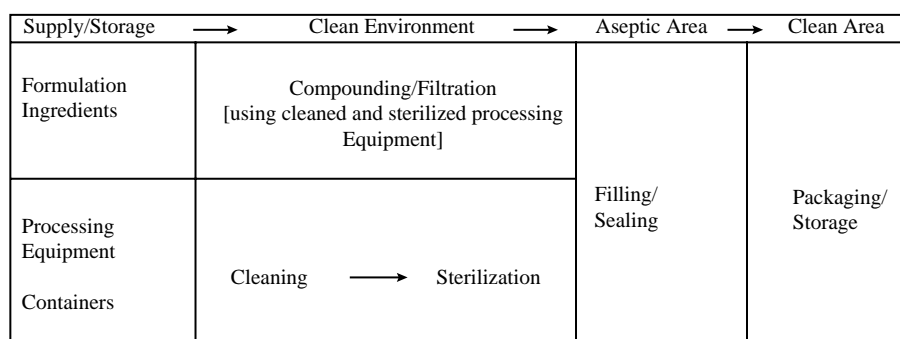


Fig. 6 Flow diagram illustrating the various processes in sterile parenterals production.

Temperature

Within the limits imposed by the thermal stability of the wanted constituents, a high extraction temperature appears desirable. The solubility of most materials increases with increasing temperature, allowing higher solute concentrations and higher concentration gradients. Both this and the increased diffusivity result in higher extraction rates. In many cases, however, materials are susceptible to heat degradation and cold extraction must be used. In addition, the selectivity of a solvent may be impaired at high temperatures. An example of the use of moderately high temperatures is the extraction of Rauwolfia alkaloids with boiling methanol.

The Relative Movement Imposed Upon the Solids and the Liquid

The major and controlling resistance to the diffusion of the solute to the bulk solution is normally found in the cell matrix. Increase in the rate of movement of the solution past the surface does not, therefore, greatly affect the rate of extraction. This is in marked contrast to the processes of dissolution and crystallization. Nevertheless, movement is imposed upon the solvent in both general methods described above.

In the percolation of a liquid through a bed of solids, mass transfer of the solute from the surfaces of the solid to the liquid in the interstices of the bed takes place by molecular diffusion and by natural convection arising from the density changes created by dissolution. Although these processes are slow, they are much faster than mass transfer in the matrix under the same concentration differences. Concentration gradients in the liquid outside the particles are, therefore, very low. At any point in the bed, the introduction of dilute solution from above and the loss of concentrated solution to below decrease the interstitial concentration by dilution or displacement. This effect can be considered simply to reduce the solute concentration at

the junction of solid and solution, thus imposing a favorable concentration gradient within the matrix.

APPLICATIONS

The processes described here are integrated in order to facilitate the production of pharmaceutical dosage forms. The following examples are intended to illustrate the application of the processes in the dominant pharmaceutical settings. They have been selected to demonstrate the broad application of unit operations in pharmaceutical manufacturing. Unique processes are associated with each dosage form. This is no less the case for dermatologics, intranasal and inhalation products, and the range of alternative pharmaceuticals than for the examples given. Nevertheless, the majority of processes and their underlying principles (11) are similar from one dosage form to the next.

Parenteral Products

Parenteral products are intended for injection into a variety of subdermal and submucosal locations (46). Their manufacture can be defined as a sequence of operations intended to be performed in certain environments or under specific conditions. Fig. 6 illustrates the sequence in which these processes may be combined. This flow diagram shows the relationship of the unit operations to the underlying physical-organic chemistry of compounding and the subsequent processes pertaining to packaging.

Solid Dosage Forms

The majority of solid dosage forms are intended for oral ingestion. The drug released from the dosage form is available at the site of absorption or action within the gastrointestinal tract.

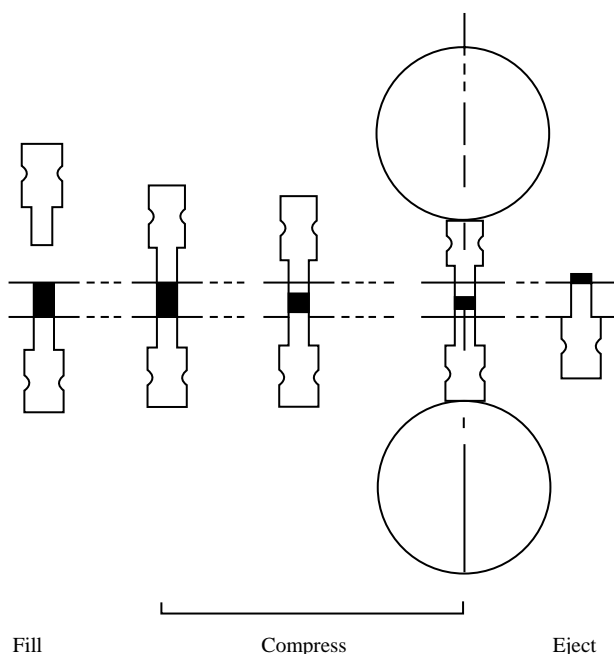


Fig. 7 Sequence of events in tablet press operation.

Additional processes are required for the production of tablets beyond those described previously. As these processes are not ubiquitous in pharmaceutical manufacturing they are dealt with only briefly here.

Granulation

Following particle size reduction and blending the formulation may be granulated (47), which provides

homogeneity of drug distribution in the blend. In addition, it may help flow properties and compression characteristics of the powder. Large granules can be prepared from primary particles by drying from a slurry (with techniques described above) or spraying with granulating solution. Fig. 7 shows a top-spray granulator. An alternative method (Fig. 7) employs an auger to force the blend between rollers, thereby forming a compressed solid which disintegrates into large aggregates (48).

The steps involved in granulation begin with transferring powders to a mixer and blending the product. The granulation solution can be added and coarse milling or wet granulation begun. Finally, the product is dried and milled to an appropriate size.

Compression

Compressed solids, tablets, or caplets, are prepared by placing the blend of component additives in a cylinder or die, above a moveable piston or punch. An upper punch is brought into the top of the piston, and pressure applied to the distal ends of the punches forces the powder into a compact (Fig. 7). The quality of the product depends upon the cohesive forces acting on the powder upon compression. These cohesive forces are influenced by the selection of additives in the dosage formulation. One method of evaluating tablet manufacture considers the effect of the applied pressure on porosity of a compressed powder (49). Data may be plotted as the negative natural logarithm of porosity against applied pressure in the form of a Heckel plot (50). The slope of this plot is proportional to the yield value (ϕ , elastic limit) with a value of $1/3\phi$.

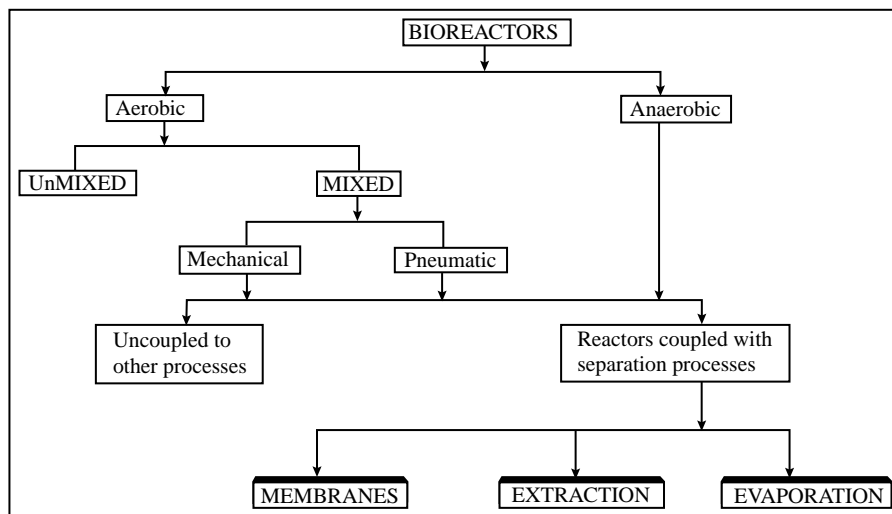


Fig. 8 Schematic of types of bioreactors.

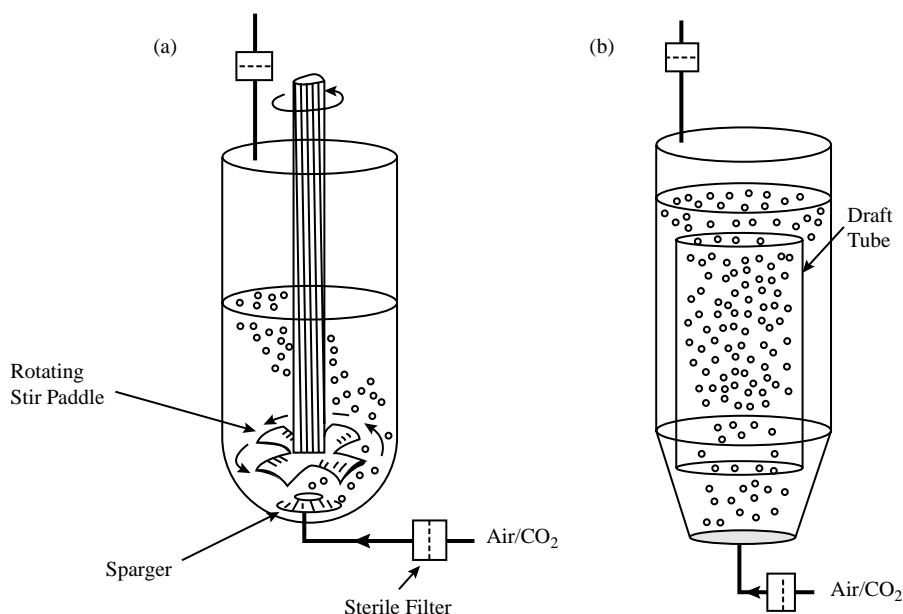


Fig. 9 Bioreactors: (a) stirred tank reactor; (b) airlift fermentor.

Bioprocessing

The use of biotechnology in the manufacture of pharmaceuticals is of increasing interest. Consequently these techniques require attention in the planning of unit processes. Bioprocessing can be considered in terms of small-scale bioreactors, or fermenters, and the translation of such processes into large-scale economically viable production operations (51, 52).

Bioprocessing is by no means a new field. The topicality of this subject is due to the increased interest in the use of isolated cells and microorganisms as manufacturing tools. It might well be argued that this technology was developed millennia ago for the purposes of wine and beer production. More recently, the use of attenuated microorganisms or isolated antigenic materials for vaccination resulted in further developments. In the last decade, the interest in genetic engineering and manipulation of the genetic code of certain microorganisms has produced a revolution in the manufacturing of pharmaceuticals.

Bioreactors

The major difference between a biotechnological process and other pharmaceutical manufacturing operations is the need for a bioreactor (Fig. 8). These bioreactors may be required to produce expressed proteins utilizing bacteria, yeast, insect, or mammalian cells. It would be difficult to describe the various bioreactor elements and their

permutations. Some of the simplest examples of bioreactors are shown in Fig. 9.

CONCLUSION

Pharmaceutical manufacturing entails the combination of a number of unit processes. The major processes have been described in this article. Brief outlines of the applications of these processes to parenteral, solid dosage form, and biological materials production are given.

The efficiency, quality, and economy of manufacturing depends upon an understanding of the individual operations involved in processing. In many cases, unlike in other industrial processing, safety and efficacy of a therapeutic agent may be affected. A guide or introduction to the practical aspects of unit processes in pharmacy is provided here.

REFERENCES

1. Ganderton, D.; Hickey, A.J. Unit Process in Pharmacy: Fundamentals. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; 1st Ed.; Marcel Dekker, Inc.: New York, 1996; 15, 341–398.
2. McCabe, W.L.; Smith, J.C.; Harriott, P. *Unit Operations in Chemical Engineering*; 5th Ed.; Ind. McGraw-Hill, Inc.: New York, 1993.

3. Hickman, K.C.D. Commercial Molecular Distillation. *Eng. Chem.* **1947**, 39, 686.
4. Ives, K.J. *Proc. Inst. Civ. Engrs.* **1963**, 25, 345.
5. Ives, K.J. Symposium: Interaction Between Fluids and Particles. *Inst. Chem. Engr.* **1962**, 260.
6. Maroudas, A.; Eisenklam, P. Clarification of Suspensions: A Study of Particle Deposition in Granular Media. *Chem. Eng. Sci.* **1965**, 20, 867.
7. Salter, K.C.; Hosking, A.P. *Chem. Eng. Pract.* **1958**, 6, 487.
8. Dickey, G.D. *Filtration*; Reinhold: New York, 1961.
9. Suits, C.G. *The Collected Works of Irving Langmuir*; Pergamon: New York, 1961; 10, 394.
10. Hinds, W.C. *Aerosols Technology, Properties, Behavior and Measurement of Airborne Particles*; John Wiley and Sons: New York, 1982; 164–186.
11. Terjsten, S.G.; Cherry, C.B. *Trans. Inst. Chem. Engrs.* **1947**, 25, 89.
12. Reist, P.C. *Aerosol Science and Technology*; 2nd Ed.; McGraw-Hill, Inc.: New York, 1992.
13. Gaden, E.L.; Humphrey, A.E. Fibrous Fitters for Air Sterilization. Design Procedure. *Ind. Eng. Chem.* **1956**, 48, 2172.
14. Stairmand, C.J. *Trans. Inst. Chem. Engrs.* **1950**, 28, 130.
15. Hesketh, H.E.; EI-Shobokshy, M.S. *Predicting and Measuring Fugitive Dust*; Technomic Publishing Co., Inc.: Lancaster, PA, 1985, 1–33.
16. Fuchs, N.A. *The Mechanics of Aerosols*; Dover Publications, Inc.: New York, 1964.
17. Humphrey, A.E.; Gaden, E.L. Air Sterilization by Fibrous Media. *Industr. Eng. Chem.* **1955**, 47, 924.
18. Cherry, G.B.; McCann, E.P.; Parker, A. The Removal of Bacteria from Air by Filtration: Application to Industrial-Scale Fermentations. *J. Appl. Chem.* **1951**, 1S103.
19. Scott, M.W.; Lieberman, H.A.; Chow, F.S.; Rankell, A.S.; Johnston, G.W. Drying as a Unit Operation in the Pharmaceutical Industry. I. *J. Pharm. Sci.* **1963**, 52, 284.
20. Scott, M.W.; Lieberman, H.A.; Chow, F.S. Pharmaceutical Applications of the Concept of Equilibrium Moisture Contents. *J. Pharm. Sci.* **1963**, 52, 994.
21. Newitt, D.M.; Na Nagara, P.; Papadopoulos, A.L. *Trans Inst. Chem. Engrs.* **1960**, 38, 273.
22. Ceaglske, N.H.; Hougen, O.A. Drying Granular Solids. *Industr. Eng. Chem.* **1937**, 29, 805.
23. Pearse, J.F.; Oliver, T.R.; Newitt, D.M. *Trans. Inst. Chem. Engrs.* **1949**, 27, 1.
24. Shepherd, C.B.; Hadlock, C.; Brewer, R.C. Drying Materials in Trays. Evaporation of Surface Moisture. *Industr. Eng. Chem.* **1938**, 30, 388.
25. Cooper, J.; Swartz, C.J.; Suydam, W. Drying of Tablet Granulations. *J. Pharm. Sci.* **1961**, 50, 67.
26. Pikal, M.J.; Roy, M.L.; Shah, S. Mass and Heat Transfer in Vial Freeze-Drying of Pharmaceuticals: Role of the Vial. *J. Pharm. Sci.* **1984**, 73, 1224–1237.
27. Nail, S.N. The Effect of Chamber Measure on Heat Transfer in the Freeze Drying of Parenteral Solutions. *J. Parent. Drug Assoc.* **1980**, 34, 358–368.
28. Jennings, T.A. Discussion of Primary Drying During Lyophilization. *J. Parent. Sci. Techn.* **1988**, 42, 118–121.
29. Dushman, S.; Lafferty, J.M. *Scientific Foundations of Vacuum Technique*; 2nd Ed.; John Wiley and Sons: New York, 1962; 48.
30. Ho, N.F.H.; Roseman, T.J. Lyophilization of Pharmaceutical Injections: Theoretical Physical Model. *J. Pharm. Sci.* **1979**, 68, 1170–1174.
31. Bond, F.C. The Third Theory of Comminution. *Trans. Amer. Inst. Min. Metall. Engrs.* **1952**, 193, 484.
32. Holmes, J.A. *Trans. Inst. Chem. Engrs.* **1957**, 3S, 125.
33. Carey, W.F.; Stairmand, C.J. Recent Advances in Mineral Dressing. *Inst. Min. Metall.* **1953**, 117.
34. Heywood, H. *Chem. Eng. Pract.* **1957**, 3, 8.
35. Gammage, R.B.; Glasson, D.R. Crystal Changes in Vateritic Calcium Carbonate During Ball Milling. *Chem. Ind.* **1963**, 1466.
36. Gregg, S.J. *Trans. Br. Ceram. Soc.* **1955**, 54, 257.
37. Cleverley, B.; Williams, P.P. Polymorphism and Changes of Infrared Spectra of Barbiturates During Sample Preparation. *Chem. Ind.* **1959**, 49.
38. Treasure, C.R. G. *Trans. Inst. Chem. Engrs.* **1965**, 43, T199.
39. Perry, R.H.; Chilton, C.H. *Chemical Engineer's Handbook*; 5th Ed.; McGraw-Hill: New York, 1973.
40. Dankwerts, P.V. The Theory of Mixtures and Mixing. *Research* **1953**, 6, 355.
41. Lacey, P.M.C. *Trans. Inst. Chem. Engrs.* **1953**, 21, 53.
42. Train, D. Pharmaceutical Aspects of Mixing Solids. *Pharm. J.* **1960**, 185, 129.
43. Lacey, P.M.C. Developments in the Theory of Particle Mixing. *J. Appl. Chem.* **1954**, 4, 257.
44. Adams, J.F.E.; Baker, A.G. *Trans. Inst. Chem. Engrs.* **1956**, 34, 91.
45. Avis, K.E.; Akers, M.J. Sterilization. *The Theory and Practice of Industrial Pharmacy*, 3rd Ed.; Lachman, L., Lieberman, H.A., Kanig, J.L. Eds.; Lea & Febiger: Philadelphia, 1986; 619–638.
46. Boylan, J.C.; Fites, A.L. *Modern Pharmaceutics*; 2nd Ed. Marcel Dekker, Inc.: New York, 1990, 491–538.
47. Carstensen, J.T. Heterogenous Systems. *Theory of Pharmaceutical Systems*; Academic Press: New York, 1973; 2, 223.
48. Doelker, E. Assessment If Powder Compaction. *Powder Technology and Pharmaceutical Processes*; Chulia, D., Deleuil, M., Pourcelot, Y., Eds.; Elsevier: Amsterdam, 1994; 403–471.
49. Carstensen, J.T. *Pharmaceutical Principles of Solid Dosage Form*; Technomic Publishing Company: Lancaster PA, 1993; 73.
50. Heckel, R.W. Density Pressure Relationships in Powder Compaction. *Trans. Metal. Soc. AIME* **1961**, 221, 671.
51. Propkop, A.; Bajpai, R.K. *Recombinant DNA Technology and Applications*; Prokop, A., Bajpai, R.K., Ho, C., Eds.; McGraw-Hill, Co.: New York, 1991; 415–459.
52. Hofmann, F.K. Scaleup of Production and Purification of Cell Expressed Proteins. *Pharmaceutical Biotechnology, Fundamentals and Essentials*; Klegerman, M.E., Groves, M.J., Eds.; Interpharm Press: Buffalo Grove IL, 1992; 138–164.

BIBLIOGRAPHY

Allen, T. *Particle Size Measurement*, 4th Ed.; Chapman and Hall: New York, 1990.

- Andrews, G.A.; Kniseley, R.M.; Wagner, H.N. *Radioactive Pharmaceuticals*; US Atomic Energy Commission: Washington, 1966.
- Ansel, H.C.; Popovich, N.G.; Allen, L. *Pharmaceutical Dosage Forms and Drug Delivery Systems*, 6th Ed.; Williams and Wilkins: Malvern, PA, 1995.
- Avis, K.E. *Process Engineering Applications* Interpharm Press, Inc.: Buffalo Grove, IL, 1995.
- Banker, G.S.; Rhodes, C.T. *Modern Pharmaceutics*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1990.
- Carey, V.P. *Liquid-Vapor Phase-Change Phenomena* Hemisphere Publishing Corporation: New York, 1992.
- Carstensen, J.Y. *Pharmaceutical Principles of Solid Dosage Forms*; Technomic Publishing Co.: Lancaster, PA, 1993.
- Cheremisinoff, P.N. *Air/Particulate Instrumentation and Analysis*; Ann Arbor Science: Ann Arbor, 1981.
- Chulia, D.; Deleuil, M.; Pourcelot, Y. *Powder Technology and Pharmaceutical Processes*; Elsevier Science B.V.: Amsterdam, 1994.
- Groves, M.J. *Parenteral Technology Manual*, 2nd Ed.; Interpharm Press: Buffalo Grove, IL, 1988.
- Groves, M.J.; Olson, W-P.; Anisfeld, M.H. *Sterile Pharmaceutical Manufacturing*; Interpharm Press: Buffalo Grove IL, 1991.
- Hesketh, H.E.; El-Shobokshy, M.S. *Predicting and Measuring Fugitive Dust*; Technomic Publishing Company: Lancaster, 1985.
- Hyman, D. Mixing and Agitation. *Advances in Chemical Engineering*; Academic Press: New York, 1962; 3.
- Klegerman, M.E.; Groves, M.J. *Pharmaceutical Biotechnology Fundamentals and Essentials*; Interpharm Press: Buffalo Grove, IL, 1992.
- Lachman, L.; Lieberman, H.A.; Kanig, J.L. *The Theory and Practice of Industrial Pharmacy*; 3rd Ed.; Lea & Febiger: Philadelphia, 1986.
- Lefebvre, A.H. *Atomization and Sprays*; Hemisphere Publishing Corporation: New York, 1989.
- Little, A.; Mitchell, K.A. *Tablet Making*, 2nd Ed.; The Northern Publishing Co., Ltd. Liverpool, 1963.
- Masters, K. *Spray Drying Handbook*, 5th Ed.; Longman Scientific and Technical and John Wiley and Sons, Inc.: New York, 1991.
- Mullin, J.W. *Crystallization*, 3rd Ed.; Butterworth-Heinemann: London, 1993.
- Wert, C.A.; Thomson, R.M. *Physics of Solids* 2nd Ed. McGraw-Hill, Inc.: New York, 1970.
- Mullin, J.W. *Crystallization*; Butterworth: London, 1961.
- Pietsch, W. *Size Enlargement by Agglomeration*; John Wiley and Sons: New York, 1991.
- Prokop, A.; Bajpai, R.K.; Ho, C. *Recombinant DNA Technology and Applications*; McGraw-Hill, Inc.: New York, 1991.
- Tien, C. *Granulator Filtration of Aerosols and Hydrosols* Butterworths: Boston, 1989.
- Van-Hook, A. *Crystallization: Theory and Practice*; A.C.S. Monograph No. 152 Chapman and Hall: London, 1961.
- Weidenbaum, S.S. Mixing of Solids. *Advances in Chemical Engineering*; Academic Press: New York, 1958; 2.

UNIT PROCESSES IN PHARMACY: FUNDAMENTALS

Anthony J. Hickey

The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

David Ganderton

Vectora Ltd., Bath, United Kingdom

INTRODUCTION

Pharmaceutical manufacturing can be divided into a number of unit process on the basis of a few fundamental principles. The following monograph describes briefly fluid flow and heat and mass transfer.

FLUID FLOW

Fluids (liquids and gases) are a form of matter that cannot achieve equilibrium under an applied shear stress but deform continuously, or flow, as long as shear stress is applied.

Streamlines are hypothetical lines without width drawn parallel to all points to the motion of the fluid. As velocity increases, pressure decreases. Pressure field around an object is the reverse of velocity field. This may appear to contradict common experience. However, it follows from the principle of conservation of energy and finds expression in Bernoulli's theorem.

Bernoulli's Theorem

At any point in system through which a fluid is flowing, the total mechanical energy can be expressed in terms of potential energy, pressure energy, and kinetic energy.

Potential energy of a body is its capacity to do work by reason of its position relative to some center of attraction. For unit mass of fluid at a height z above some reference level, the potential energy is zg .

Pressure energy or flow energy is an energy form peculiar to the flow of fluids. The work done and the energy acquired in transferring the fluid is the product of the pressure P and the volume.

Volume of unit mass of the fluid is the reciprocal of the density ρ . For an incompressible fluid, the density is not dependent on the pressure, so for unit mass of fluid the pressure energy is P/ρ .

Kinetic energy is a form of energy possessed by a body by reason of its movement. If the mass of the body is m and

its velocity u , the kinetic energy is $1/2 mu^2$, and for unit mass of fluid the kinetic energy is $u^2/2$.

The total mechanical energy of unit mass of fluid is therefore

$$\frac{u^2}{2} + \frac{P}{\rho} + zg$$

and proved no energy is lost or gained by the system, the mechanical energy at two points A and B is the same as that expressed by Eq. 1.

$$\frac{u_A^2}{2} + \frac{P_A}{\rho} + z_A g = \frac{u_B^2}{2} + \frac{P_B}{\rho} + z_B g \quad (1)$$

This relationship neglects the frictional degradation of mechanical energy that occurs in real systems. A fraction of the total energy is dissipated in overcoming shear stresses induced by velocity gradients in the fluid. Energy E lost during flow between A and B must be considered. The dimensions of each component are $L^2 T^{-2}$, Length²Time⁻². In practice, each term is divided by g (LT^{-2}) to give the dimension of length. The terms are then referred to as velocity head, pressure head, potential head, and friction head, the sum giving the total head of the fluid as shown in Eq. 2.

$$\frac{u_A^2}{2g} + \frac{P_A}{\rho g + z_A} = \frac{u_B^2}{2g} + \frac{P_B}{\rho g} + z_B + \frac{E}{g} \quad (2)$$

The evaluation of the kinetic energy term requires consideration of the variation in velocity found in the direction normal to flow. The kinetic energy, given by the term $u_{\text{mean}}^2/2$, differs from the true kinetic energy found by summation across the flow direction. The former can be retained, however, if a correction factor a is introduced. Then

$$\text{velocity head} = \frac{u_{\text{mean}}^2}{2ga},$$

where a has a value of 0.5 in laminar flow and approaches unity in fully turbulent flow. Mechanical energy is added to the system at some point by means of a pump. The work W done, in absolute units, on a unit mass of fluid at A is given by Eq. 3.

$$\frac{W}{g} + \frac{u_A^2}{2g} + \frac{P_A}{\rho g} + z_A = \frac{u_B^2}{2g} + \frac{P_B}{\rho g + z_B} + \frac{E}{g} \quad (3)$$

The power required through a system at a certain rate may be calculated using Eq. 3 to drive a liquid. The sum of heads, ΔH , being the total head against which the pump must work is therefore

$$\frac{W}{g} + H$$

If the work performed and energy acquired by unit mass of fluid is ΔHg , the power required to transfer mass m in time t is given by

$$\text{Power} = \frac{Hgm}{t}$$

As the volume flowing in unit time Q is $m/\rho t$, the power is given by Eq. 4.

$$\text{Power} = QHg\rho \quad (4)$$

Flow Measurement

The Bernoulli theorem can also be applied to the measurement of flow rate. The passage of an incompressible fluid through a constriction results in an increase in velocity from u_1 to u_2 , which is associated with a decrease in pressure from P_1 to P_2 , which can be measured directly. The volumetric flow rate $Q = u_1 a_1 = u_2 a_2$ by algebraic rearrangement (which is not shown). The final linear velocity u_2 can be described by Eq. 5.

$$\frac{u_2^2}{2} - \frac{u_1^2}{2} = \frac{P_1 - P_2}{\rho} \quad (5)$$

The volumetric flow rate can be described by Eq. 6.

$$Q = a_2 \sqrt{\frac{2(P_1 - P_2)}{\rho(1 - a_2^2/a_1^2)}} \quad (6)$$

This derivation neglects the correction of kinetic energy loss due to nonuniformity of flow in both cross sections and the frictional degradation of energy during passage through the constriction. This is corrected by the introduction of a numerical coefficient, C_D , known as the coefficient of discharge, as shown in Eq. 6.

$$Q = C_D a_2 \sqrt{\frac{2(P_1 - P_2)}{\rho(1 - a_2^2/a_1^2)}}$$

The value of C_D depends on conditions of flow and shape of the constriction. For a well-shaped constriction (notably circular cross section), it would vary between 0.95 and 0.99 for turbulent flow. The value is much lower in laminar flow because the kinetic energy correction is larger. The return of the fluid to the original velocity by means of a diverging section forms a flow-measuring device known as a Venturi meter.

The Venturi meter is shown in Fig. 1(a). The converging cone leads to the narrowest cross section known as the throat. The change in pressure is measured across this part of the meter and the volumetric flow rate found by substitution into Eq. 6. Values of the coefficient of discharge are given previously. The diverging section or diffuser is designed to induce a gradual return to the original velocity. This minimizes eddy formation in the diffuser and permits the recovery of a large proportion of the increased kinetic energy as pressure energy. The permanent loss of head due to friction in both converging and diverging sections is small. The meter is therefore efficient.

When the minimization of energy degradation is less important, the gradual economic return to the original velocity may be abandoned, compensation for loss of efficiency being found in a device that is simpler, cheaper, and more adaptable than the Venturi meter. The orifice meter, to which this statement applies, consists simply of a plate with an orifice. A representation of flow through the meter is depicted in Fig. 1(b), indicating convergence of the fluid stream after passage through the orifice to give across section of minimum area called the *vena contracta*. The downstream pressure tapping is made at this cross section. The volumetric flow rate is given by Eq. 6 in which a_2 is the jet area at the *vena contracta*. The measurement of this dimension is inconvenient. It is therefore related to the area of the orifice, a_0 , which can be accurately measured by the coefficient of contraction C_c by the relation $C_c = a_2/a_0$.

The coefficient of contraction, frictional losses between the tapping points, and kinetic energy corrections are absorbed in the coefficient of discharge. The volumetric flow rate is then given by Eq. 7.

$$Q = C_D a_0 \sqrt{\frac{2(P_1 - P_2)}{\rho(1 - a_0^2/a_1^2)}} \quad (7)$$

If the orifice is small compared with the pipe cross section, the term $(1 - (a_0^2/a_1^2))$ approaches unity. As $P_1 - P_2 = \Delta h \rho g$, Δh being the difference in head developed by the orifice, Eq. 7 reduces to Eq. 8.

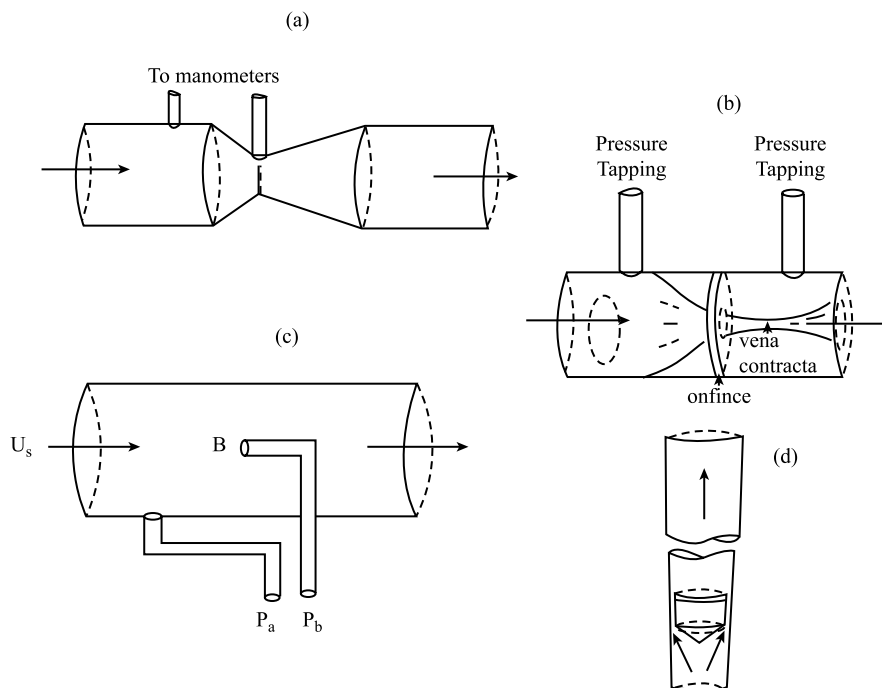


Fig. 1 (a) Venturi meter; (b) orifice meter; (c) Pitot tube; (d) rotameter.

$$Q = C_D a_0 (h \rho g)^{1/2} \quad (8)$$

The value of C_D for the orifice meter is about 0.6, varying with construction, the ratio a_0/a_1 , and flow conditions within the meter. Due to its complexity, it cannot be calculated.

After passage through the orifice, flow disturbance during retardation causes the dissipation of most of the excess kinetic energy as heat. The permanent loss of head is therefore high, increasing as the ratio of a_0/a_1 falls, ultimately reaching the differential head produced within the meter.

The Bernoulli theorem may be used to determine the change in pressure caused by retardation of fluid at the upstream side of a body immersed in a fluid stream. This principle is applied in the Pitot tube, shown in Fig. 1(c). The fluid velocity is reduced from u_a , the velocity of the fluid filament in alignment with the tube, to zero at B, an position known as the stagnation point. The pressure, P_b , is measured at this point by the method shown in Fig. 1(c). The undisturbed pressure, P_a , is measured in this example with a tapping point in the wall connected to a manometer. As the velocity at B is zero, Eq. 10 reduces to

$$\frac{u_a^2}{2} = \frac{p_b - P_a}{\rho}$$

and u_2 can be calculated. As only a local velocity is measured, variation of velocity in a section can be studied by altering the position of the tube. This procedure must be used if the flow rate in a pipe is to be measured. The mean velocity is derived from velocities measured at different distances from the wall. This derivation and the low pressure differential developed render the Pitot tube less accurate than either the venturi tube or the orifice meter for flow measurement. Hence, the tube is small in comparison with the pipe diameter and, therefore, produces no appreciable loss of head.

The rotameter (a variable area meter), shown in Fig. 1(d), is commonly used, giving a direct flow rate reading by the position of a small float in a vertical, calibrated glass tube through which the fluid is flowing. The tube is internally tapered toward the lower end so that the annulus between float and wall varies with the position of the float. Acceleration of the fluid through the annulus produces a pressure differential across the position of the float and an upward force upon it. At the equilibrium position, which may be stabilized by a slow rotation of the float, this upward force is balanced by the weight force acting on the float. If the equilibrium is disturbed by increasing the flow rate, the balance of weight force and the pressure differential are produced by movement of the float upward to a position at which the area of the annulus is bigger. For accurate measurement, the rotameter is

calibrated with the fluid to be metered. Its use is, however, restricted to that fluid.

Laminar and Turbulent Flow

The nature of flow may be examined by introducing a dye into the axis of the tube. At low speeds, the dye forms a coherent thread, which grows very little in thickness with distance down the tube. However, with progressive increase in speed, the line of dye begin to waver and then break up. Secondary motions crossing and recrossing the general flow direction occur. Finally, at very high speeds, no filament of dye is detected and mixing is instantaneous. In this experiment, flow changes from laminar to turbulent, the change occurring at a critical speed. Generalizing, in laminar flow, the instantaneous velocity at a point is always the same as the mean velocity in both magnitude and direction. In turbulent flow, order is lost and irregular motions are imposed upon the main steady motion of the fluid. At any instant of time, the fluid velocity at a point varies in both magnitude and direction, having components perpendicular as well as parallel to the direction of net flow. Over a period of time, these fluctuations even out to give the net velocity in the direction of flow.

In turbulent flow, rapidly fluctuating velocities produce high velocity gradients within the fluid. Proportionately large shear stresses are developed, and to overcome them, mechanical energy is degraded and dissipated in the form of heat. The degradation of energy in laminar flow is much smaller.

The turbulent mechanism that carries motion, heat, or matter from one part of the fluid to another is absent in laminar flow. The agency of momentum transfer is the shear stress arising from the variations in velocity, that is, the viscosity. Similarly, heat and matter can only be transferred across streamlines on a molecular scale, heat by conduction and matter by diffusion. These mechanisms that are present but less important in turbulent flow are comparatively slow. Velocity, temperature, and concentration gradients are, therefore, much higher than in turbulent flow.

The Flow of Liquids in Pipes

The many pharmaceutical processes that involve the transfer of a liquid confer great importance on the study of flow in pipes. This study permits the evaluation of pressure loss due to friction in a simple pipe and assesses the additional effects of pipe roughness, changes in diameter, bends, exits, and entrances. When the total pressure drop due to friction is known for the system, the equivalent head

can be derived and the power requirement for driving a liquid through the system can be calculated from Eq. 9.

Streamline flow in a tube

The mathematical analysis of streamline flow in a simple tube results in the expression known as Poiseuille's law, one form of which is Eq. 9.

$$Q = \frac{P\pi d^4}{128\eta l} \quad (9)$$

where Q is the volumetric flow rate or discharge, ΔP is the pressure drop across the tube, d and l are the diameter and length of the tube, respectively, and η is the viscosity of the fluid.

Where flow in the tube is streamline or turbulent, an infinitesimally thin stationary layer is found at the wall. The velocity increases from zero at this point to a maximum at the axis of the tube. The velocity profile of streamline flow is shown in Fig. 9(a). The velocity gradient du/dr varies from a maximum at the wall to zero at the axis. In flow through a tube, the rate of shear is equal to the velocity gradient, and Eq. 1 dictates the same variation of shear stress.

To derive Poiseuille's law, the form of the velocity profile must first be established. For a fluid contained within a radius r flowing in a tube of radius R , the pressure drop across length l is ΔP ; therefore, the pressure force driving this section is $\Delta P\pi r^2$. If the flow is steady, this force can only be balanced by opposing viscous forces acting on the "wall" of the section. This force is

$$\tau = \frac{Pr}{2l}$$

Substituting for Eq. 1 gives

$$-\frac{du}{dr} = \frac{Pr}{2\eta l}$$

The velocity gradient is negative because u decreases as r increases. If $r = R$, then $u = 0$. Integration gives

$$\int_0^u du = \frac{P}{2\eta l} \int_R^r r dr$$

Therefore, Eq. 10 can be written as

$$u = \frac{P}{2\eta l} \left(\frac{R^2 - r^2}{2} \right) \quad (10)$$

This relation shows that the velocity distribution across the tube is parabolic. For such a distribution, the maximum velocity is twice the mean velocity.

The volumetric flow rate across an annular section between r and $(r + dr)$ is given by

$$Q = 2\pi r dr u.$$

Substituting for u for Eq. 10 gives

$$Q = \frac{P\pi}{2\eta l} (R^2 r - r^3) dr$$

The total volumetric flow rate is the integral between the limits $r = R$ and $r = 0$:

$$Q = \frac{P\pi}{2\eta l} \int_0^R (R^2 r - r^3) dr = \frac{P\pi}{2\eta l} R^2 \left(\frac{r^2}{2} - \frac{r^4}{4} \right)_0^R$$

Therefore Eq. 9 is valid:

$$Q = \frac{P\pi R^4}{8\eta l} = \frac{P\pi d^4}{128\eta l}$$

where d is the diameter of the tube.

As $Q = u_{\text{mean}} \pi (d^2/4)$, substitution and rearrangement gives Eq. 11.

$$P = \frac{32u_{\text{mean}}\eta l}{d^2} \quad (11)$$

The Significance of Reynolds Number Re

The Reynolds number describes the ratio of the inertia and viscous or frictional forces. Higher the Reynolds number, greater is the relative contribution of inertial effects. At very low Re , viscous effects predominate and the contribution of inertial forces can be ignored.

Calculation of the Pressure Drop in a Pipe Due to Friction

If the volumetric flow rate of a liquid of density ρ and viscosity η through a pipe of diameter d is Q , the derivation of the mean velocity u from the flow rate and pipe area completes the data required for calculating Re . If the pipe roughness factor is known, the equivalent value of $R/\rho u^2$ can be read from Fig. 2, and the shear stress at the pipe wall calculated. The total frictional force opposing motion is the product of R and the surface area of the pipe, πdl , where l is the length of the pipe. If the unknown pressure drop across the pipe is ΔP , the force driving through the pipe is $\Delta P \cdot \pi d^2/4$. Equating pressure force and frictional force,

$$P = \frac{\pi d^2}{4} = R\pi dl$$

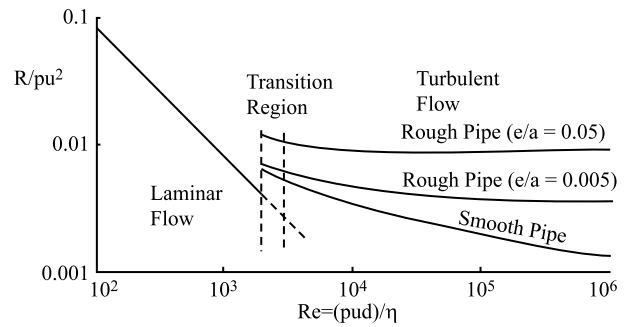


Fig. 2 Pipe friction chart $R/\rho u^2$ versus Re .

Therefore, Eq. 12 follows.

$$P = \frac{4Rl^2}{d} \quad (12)$$

Division by ρg gives the pressure loss as a friction head.

Flow in Tubes of Noncircular Cross Section

The foregoing arguments may be applied to turbulent flow in noncircular ducts by introducing a dimension equivalent to the diameter of a circular pipe. This is known as the mean hydraulic diameter, d_m , which is defined as four times the cross-sectional area divided by the wetted perimeter. The following examples are given:

For a square channel of side b ,

$$d_m = \frac{4b^2}{4b} = b.$$

For an annulus of outer radius r_1 and inner radius r_2 ,

$$\frac{4(\pi r_1^2 - \pi r_2^2)}{2\pi r_1 + 2\pi r_2} = 2(r_1 - r_2)$$

This simple modification does not apply to laminar flow in noncircular ducts.

Frictional Losses at Pipe Fittings

In addition to the friction losses at the wall of a straight pipe, losses occur at the various fittings and valves used in practical systems. In general, these losses are derived from sudden changes in the magnitude or direction of flow induced by changes in geometry. They can be classified as losses due to a sudden contraction or enlargement, losses at entrance or exit, and losses due to pipe curvature. Losses can be conveniently expressed as a length of straight pipe offering the same resistance. This is usually in the form of a number of pipe diameters. For

example, the loss at a right-angled elbow is equivalent to the length of a straight pipe equal to 40 diameters. The sum of the equivalent lengths of all fittings and vales is then added to the actual pipe length and the total frictional loss estimated by Eq. 20.

Motion of Bodies in a Fluid

Viscous and inertial forces operate to determine the flow pattern and drag force on a body moving relative to a fluid. The Reynolds number, which expresses their ratio, is used as a parameter to predict flow behavior. The relation between the drag force and its controlling variables is presented in a manner similar to that employed for flow in a pipe. Considering a sphere moving relative to a fluid, the projected area normal to flow is $\pi d^2/4$, where d is the diameter of the sphere. The drag force acting on the unit projected area R' is determined by the velocity u , the viscosity η , and the density ρ of the fluid and the diameter of the sphere d . Dimensional analysis yields Eq. 13.

$$\frac{R}{\rho u^2} = f(Re) = f\left(\frac{ud\rho}{\eta}\right) \quad (13)$$

This form of Reynolds number, Re' , employs the diameter of the sphere as the linear dimensions.

With the exception of an analysis at very low Re values, the form of this function is established by experiment. Results are presented on logarithmic coordinates in Fig. 3. When Re' is less than about 0.2,

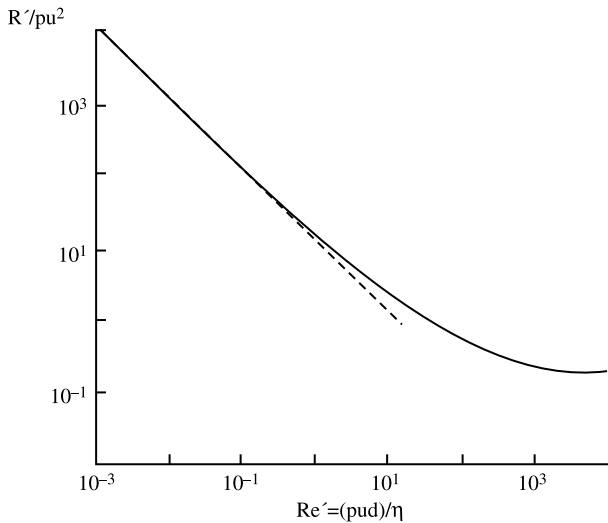


Fig. 3 $R'/\rho u^2$ versus Re' number for a smooth sphere.

viscous forces are solely responsible for drag on the sphere and Eq. 13 becomes

$$\frac{R}{\rho u^2} = \frac{12}{Re}$$

Therefore,

$$\text{Total drag force} = R \frac{\pi d^2}{4} = \rho u^2 \frac{12 \pi d^2}{Re \cdot 4} = 3\pi\eta du \quad (14)$$

This is the normal form of Stokes law.

At larger Re' values, the experimental curve progressively diverges from this relation, ultimately becoming independent of Re' and giving a value of $R/\rho u^2$ equal to about 0.22. As Re' increases, the form drag increases, ultimately becoming solely responsible for the force opposing motion.

For nonspherical particles, the analysis employs the diameter of a sphere of equivalent volume. A correction factor, which depends upon the shape of the body and its orientation in the fluid, must be applied.

An important application of this analysis is the estimation of the speed at which particles settle in a fluid. Under the action of gravity, the particle accelerates until the weight force mg is exactly balanced by the opposing drag. The body then falls at a constant terminal velocity u . Equating weight and drag forces gives Eq. 15.

$$mg = \frac{\pi}{6} d^3 (\rho_s - \rho) g = R \frac{\pi d^2}{4} \quad (15)$$

where ρ is the density of the particle.

For a sphere falling under streamline conditions ($Re' < 0.2$), $R' = \rho u^2 (12/Re')$. Substituting in Eq. 15 gives Eq. 16.

$$u = \frac{d^2 (\rho_s - \rho) g}{18\eta} \quad (16)$$

This expression follows more simply from the equation $mg = 3\pi d\eta u$.

Flow of Fluids Through Packed Beds

The analysis of the flow of fluids through a permeable bed of solids is widely applied in filtration, leaching, and several other processes. A first approach may be made by assuming that the interstices of the bed correspond to a large number of discrete, parallel capillaries. If the flow is streamline, the volumetric flow rate Q is given for a single capillary by Eq. 14:

$$Q = \frac{P\pi d^4}{128\eta l}$$

where l is the length of the capillary and d its diameter, ΔP is the pressure drop across the capillary, and η is the viscosity of the fluid. The length of the capillary exceeds the depth of the bed by a value that depends upon its tortuosity. The depth of bed, L , is however, proportional to the capillary length l so that

$$Q = \frac{Pd^4}{k\eta L},$$

where k is a constant for a particular bed. If the area of the bed is A and it contains n capillaries per unit area, the total flow rate is given by

$$Q = \frac{Pd^4 nA}{k\eta L}$$

Both n and d are not normally known. However, they have certain values for a given bed, expressed by Eq. 17.

$$Q = KA \frac{P}{\eta L} \quad (17)$$

where $K = d^4 n/k$. This constant is a permeability coefficient and its reciprocal, $1/K$, is the specific resistance. Its value characterizes a particular bed.

The postulate of discrete capillaries precludes valid comment on the factors that determine the permeability coefficient. Channels are not discrete but are interconnected in a random manner. Nevertheless, the resistance to the passage of fluid must depend on the number and dimensions of the channels. These can be expressed in terms of the fraction of the bed which is void, that is, the porosity, and the manner in which the void fraction is distributed. Illustrating this reference to a specific example, water would flow more easily through a bed with a porosity of 40% than through a bed of the same material with a porosity of 25%. It would also flow more quickly through a bed of coarse particles than through a bed of fine particles packed to the void fraction or porosity. The latter effect can be expressed in terms of the surface area offered to the fluid by the bed. This property is inversely proportional to the size of the particles forming the bed. Permeability increases as the porosity increases and the total surface of the bed decreases and these factors may be combined to give the hydraulic diameter d' of an equivalent channel. This is defined by:

$$d = \frac{\text{Volume of voids}}{\text{Total surface of material forming bed}}$$

The volume of voids is the porosity, and the volume of solids is $(1 - \varepsilon)$. If the specific surface area, that is, the surface area of unit volume of solids, is S_0 , the total surface presented by unit volume of the bed is $S_0(1 - \varepsilon)$.

Therefore, Eq. 18 applies.

$$d = \frac{\varepsilon}{S_0(1 - \varepsilon)} \quad (18)$$

Under streamline conditions, the rate at which a fluid flows through this equivalent channel is given by Poiseuille's Eq. 14.

The velocity u' in the channel is derived by dividing the volumetric flow rate by the area of the channel, $k'd'^2$. Combining the constants,

$$u = \frac{Q}{kd^2} = \frac{Pd^2}{k\eta L}$$

This velocity, when averaged over the entire area of the bed, solids, and voids, gives the lower value u . These velocities are related by

$$u = u\varepsilon$$

Therefore:

$$\frac{u}{\varepsilon} = \frac{Pd^2}{k\eta L}$$

Substituting for d' by means of Eq. 18 gives

$$\frac{u}{\varepsilon} = \frac{P}{k\eta L} \frac{\varepsilon^2}{(1 - \varepsilon)^2 S_0^2}$$

and Eq. 19.

$$u = \frac{P}{k\eta L} \frac{\varepsilon^3}{(1 - \varepsilon)^2 S_0^2} \quad (19)$$

In this equation, known after its originator as Kozeny's equation, the constant k'' has a value of 5 ± 0.5 . As $Q = uA$, where A is the area of the bed, Eq. 19 can be transformed to Eq. 20.

$$Q = \frac{PA}{\eta L} \cdot \frac{\varepsilon^3}{5(1 - \varepsilon)^2 S_0^2} \quad (20)$$

This analysis shows that permeability is a complex function of porosity and surface area, the latter being determined by the size distribution and shape of the particles. The appearance of specific surface in Eq. 20 offers a method for its measurement and provides the basis of fluid permeation methods of size analysis. This equation also applies in the studies of filtration.

Pumps

Eqs. 8 and 9 examine the power requirement for driving a liquid through a system against an opposing head. This energy is normally generated by a pump. In different

processes, the quantities to be delivered, the opposing head, and the nature of the fluid vary widely, and many pumps are made to meet these differing requirements. Basically, however, pumps can be divided into positive-displacement pumps, which may be reciprocating or rotary, and impeller pumps. Positive-displacement pumps seek to displace a fixed volume of fluid with each stroke or revolution; impeller pumps, on the other hand, impart high kinetic energy to the fluid which is subsequently converted to pressure energy. The volume discharged depends upon the opposing head.

The equipment for pumping gases and liquids is essentially similar. Machines delivering gases are commonly called compressors or blowers; compressors discharge at relatively high pressures and blowers at relatively low pressures. The lower density and viscosity of gases results in higher operating speeds and, to minimize leakage, smaller clearance between moving parts.

HEAT TRANSFER

Heat transfer in process vessels has been described previously in this encyclopedia (1). The following is a more complete review of heat transfer as a unit operation in pharmacy. Heat energy can only be transferred from a region of higher to a region of lower temperature. Understanding heat transfer requires the study of the mechanism and rate of this process.

Heat is transferred by three mechanisms: conduction, convection, and radiation. It is unusual for the transfer to take place by one mechanism only.

Conduction is the most widely understood mechanism of heat transfer and the main method in solids. The flow of heat depends upon the transfer of vibrational energy from one molecule to another and, in the case of metals, the movement of free electrons. Radiation is rare in solids but examples are found among glasses and plastics. Convection by definition, is not possible under these conditions. Conduction in the bulk of fluids is normally overshadowed by convection, but it assumes great importance at fluid boundaries.

Heat is transferred from or to a region by the motion of fluids and the phenomenon of convection. In natural convection, the movement is caused by buoyancy forces induced by variations in the density of the fluid; these variations are caused by differences in temperature. In forced convection, movement is created by an external agency such as a pump.

All bodies with a temperature above absolute zero radiate heat in the form of electromagnetic waves.

The radiation may be transmitted, reflected, or absorbed by matter, the fraction absorbed being transformed into heat. Radiation is of great importance at very high or very low temperatures and under circumstances in which the other modes of heat transmission are suppressed. Although the heat losses can, in some cases, equal the losses by natural convection, the mechanism is, from the standpoint of pharmaceutical processing, least important and needs only brief consideration.

Heat transfer in many systems occurs as a steady-state process, and the temperature at any point in the system does not vary with time. In other important processes, temperatures in the system do vary with time. This situation, which is common among the small-scale, batch-operated processes of the pharmaceutical and fine chemical industry, is known as unsteady heat transfer and, as warming or cooling occurs, the thermal capacity (i.e., the size and specific heat), of the system becomes important. Analysis of unsteady heat transfer is complex.

Heat Transfer Through a Wall

Heat transfer by conduction through walls follows the basic relation given by Fourier's equation (Eq. 21), which states that the rate of heat flow, Q , is proportional to the temperature gradient, dT/dx , and to the area normal to the heat flow, A .

$$Q = -kA \frac{dT}{dx} \quad (21)$$

As the distance x increases, the temperature T decreases. Hence, measuring in the x direction, the temperature gradient, dT/dx , is algebraically negative. The proportionality constant k is the thermal conductivity. Its numerical value depends on the material of which the body is made and on its temperature.

Metals exhibit high conductivity, although values vary widely. Nonmetallic solids normally have lower conductivities than do metals. For the porous materials of this group, the overall conductivity lies between that of the homogeneous solid and the air that permeates the structure. Low resultant values lead to wide use as heat insulators. Carbon is an exception among nonmetals. Its relatively high conductivity and chemical inertness permit its wide use in heat exchangers.

Steady nondirectional heat transfer through a plane wall of thickness x and area A is represented in Fig. 4(a). Assuming that thermal conductivity does not change with temperature, the temperature gradient is linear and equal to $(T_1 - T_2)/x$, where T_1 is the temperature of the hot face and

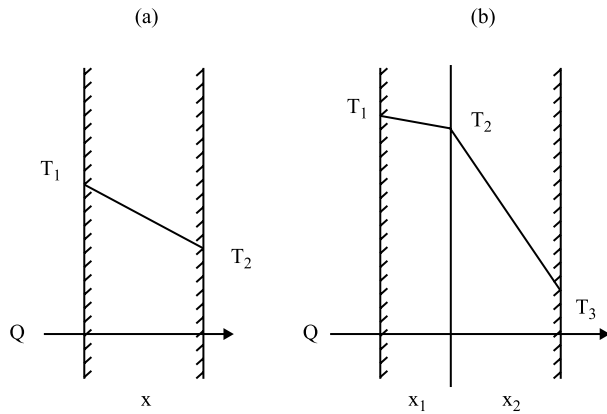


Fig. 4 Conduction of heat through (a) a plane wall and (b) a composite wall.

T_2 the temperature of the cool face. Eq. 21 then becomes Eq. 22

$$Q = kA \frac{T_1 - T_2}{x} \quad (22)$$

This may be rearranged to Eq. 23

$$Q = A \frac{T_1 - T_2}{x/k} \quad (23)$$

where x/k is the thermal resistance. Thus, for a given heat flow, a large temperature drop must be created if the wall or layer has a high thermal resistance.

Conversely, the heat flow promoted by a given temperature difference is reduced if the thermal resistance is increased. This is the principle of insulation by lagging, and it is illustrated by a composite wall, as shown in Fig. 4(b). If steady-state heat transfer exists, the rate of heat transfer is the same for both materials. Therefore,

$$Q = \frac{k_1 A (T_1 - T_2)}{x_1} = \frac{k_2 A (T_2 - T_3)}{x_2}$$

The major temperature drop occurs across the distance x_2 , indicating that this material provides the major thermal resistance. (In the case of heavily lagged, thin metal walls, the temperature drop and thermal resistance of the metal are so small that they can be ignored). Rearrangement of this equation and the elimination of the junction temperature gives Eq. 24.

$$Q = A \frac{(T_1 - T_3)}{(x_1/k_1 + x_2/k_2)} \quad (24)$$

Equations of this form can be applied to any number of layers.

Heat Transfer in Pipes and Tubes

Pipes and tubes are common as barriers over which heat exchange takes place. Conduction is complicated in this case by the changes in the area over which heat is transferred. If Eq. 21 is to be retained, some value of A must be derived from the length of the pipe, l , and the internal and external radii, r_1 and r_2 , respectively. If the pipe is thin-walled and the ratio r_2/r_1 is less than approximately 1.5, the heat transfer area can be based on an arithmetic mean of the two radii, and Eq. 21 becomes Eq. 25.

$$Q = k2\pi \frac{r_2 + r_1}{2} l \frac{T_1 - T_2}{r_2 - r_1} \quad (25)$$

This equation is inaccurate for thick-walled pipes, where the heat transfer area must be based on a logarithmic mean radius r_m . Heat transfer is then expressed by Eq. 26

$$Q = k2\pi r_m l \frac{T_1 - T_2}{r_2 - r_1} \quad (26)$$

where

$$r_m = \frac{r_2 - r_1}{\log_{e} \frac{r_2}{r_1}}$$

Heat Exchange Between a Fluid and a Solid Boundary

Conduction and convection contribute to the transfer of heat from a fluid to a boundary. The distribution of temperatures at a plane barrier separating two fluids may be considered. If the fluids are turbulent motion, temperature gradients are confined to a relatively narrow region adjacent to the wall. Outside this region, turbulent mixing, the mechanism of which has been explained earlier, is very effective in the transfer of heat. Temperature gradients are quickly destroyed and equalization at values T_1 and T_2 occurs. Within the region, there exists a laminar sublayer across which heat is transferred by conduction only. The temperature gradients produced by a given heat flow are correspondingly high. Outside the laminar layer, eddies contribute to the transfer of heat by moving fluid from the turbulent bulk to the edge of the sublayer, where heat can be lost or gained, and by corresponding movements in the opposite direction. The temperature gradients in this region, where both convection and conduction contribute to heat transfer, are smaller than in the sublayer.

The major resistance to the flow of heat resides in the laminar sublayer. Its thickness, therefore, is of critical importance in determining the rate of heat transfer from

the fluid to the boundary. It depends on the physical properties of the fluid, the flow conditions, and the nature of the surface. Increase in flow velocity, for example, decreases the thickness of the layer and, therefore, its resistance to heat flow.

To evaluate the rate of heat transfer at a boundary, a film, transmitting heat only by conduction is postulated. This fictitious film presents the same resistance to heat transfer as the complex turbulent and laminar regions near the wall. If on the hot side of the wall the fictitious layer had a thickness x_1 , the equation of heat transfer to the wall would be

$$Q = kA \frac{T_1 - T_{1\text{ wall}}}{x_1}$$

where k is the thermal conductivity of the fluid. A similar equation applies to heat transfer at the cold side of the wall. The thickness of the layer is determined by the same factors that control the extent of the laminar sublayer. In general the layer thickness is not known and the above equation is rewritten as Eq. 27.

$$Q = h_1 A (T_1 - T_{1\text{ wall}}) \quad (27)$$

where h_1 is the heat-transfer coefficient for the film under discussion. It corresponds to the ratio k/x_1 and is the $\text{Btu/ft}^2 \cdot \text{h} \cdot ^\circ\text{F}$. It is a convenient numerical expression of heat flow by conduction and convection at a boundary. The approximate evaluation of these coefficients is discussed later.

The film coefficient h_2 may be used to characterize heat transfer from the barrier to the colder fluid.

$$T_1 - T_{1\text{ wall}} = \frac{Q}{h_1 A} \quad T_{1\text{ wall}} - T_{2\text{ wall}} = \frac{Q x_w}{h_w A}$$

where k_w is the thermal conductivity of the wall, and

$$T_{2\text{ wall}} - T_2 = \frac{Q}{h_2 A}$$

Addition and rearrangement of these equations gives Eq. 28.

$$Q = \frac{A}{(1/h_1 + x_w/k_w + 1/h_2)} (T_1 - T_2) \quad (28)$$

The quantity

$$\frac{1}{(1/h_1 + x_w/k_w + 1/h_2)}$$

is called the overall heat-transfer coefficient U . A general expression of the rate of heat transfer then becomes Eq. 29.

$$Q = UAT \quad (29)$$

or

$$\frac{ql}{Tk} = \text{constant} \left(\frac{l^3 Tag \rho^2}{\eta^2} \right) \left(\frac{C_p \eta}{k} \right)^q \quad (30)$$

Heat transfer by free convection can thus be presented as a relation between three dimensionless groups: $C_p \eta / k$ is known as the Prandtl number, the combination $l^3 \Delta T a g \rho^2 / \eta^2$ as the Grashof number, and $ql / \Delta T$, the Nusselt number may also be written hl/k .

The specific relation in which these groups stand is established for a particular system experimentally.

The fluid properties, C_p , k , η , and ρ are themselves temperature dependent. In establishing a correlation, the temperature at which these properties are to be measured must be chosen. This is usually the temperature of the main body of the fluid or the mean of this temperature and the temperature of the surface.

The exponents r and q are usually found experimentally to be equal to 0.25 in streamline flow and 0.33 in turbulent flow. The constant varies with the configuration. As an example, the heat transfer to gases and liquids from a large horizontal pipe by free convection is described by Eq. 31.

$$\frac{qd}{kT} = 0.47 \left(\frac{d^3 Tag \rho^2}{\eta^2} \right) \left(\frac{C_p \eta}{k} \right)^{0.25} \quad (31)$$

The linear dimension in this correlation is the pipe diameter d . The fluid properties are to be measured at the mean of the wall and bulk fluid temperatures.

In forced convection, the fluid is moved over the surface by a pump or blower neglecting natural convection are usually neglected. The study of forced convection is of great practical importance and vast amount of data have been amassed for streamline and turbulent flow in pipes, across and parallel to tubes, across plane surfaces, and in other important configurations such as jackets and coils.

In forced convection, the heat transferred per unit area per unit time q , is determined by a linear dimension which characterizes the surface l , the temperature difference between the surface and the fluid, ΔT , the viscosity η , the density ρ , and the velocity u , of the fluid, its conductivity k , and its specific heat C_p . Dimensional analysis yields Eq. 32

$$\frac{qd}{kT} = \text{constant} \left(\frac{C_p \eta}{k} \right)^x \left(\frac{ul\rho}{\eta} \right)^y \quad (32)$$

where ql/kT is the Nusselt number, Nu , $C_p \eta / k$ is the Prandtl number, Pr , and $ul\rho / \eta$ is the Reynolds number, Re , a parameter discussed previously. The values of the indices, x and y and of the constant are established for a particular system experimentally. In the case of turbulent flow in

pipes, the correlation for fluids of low viscosity is given by Eq. 33.

$$Nu = 0.023 Pr^x Re^{0.8} \quad (33)$$

where x has the value 0.4 for heating and 0.3 for cooling. The linear dimension used to calculate Re or Nu is the pipe diameter, and the physical properties of the fluid are to be measured at the bulk fluid temperature. This relation shows that in a given system, the film coefficient varies as the fluid velocity^{0.8}. If the flow velocity is doubled, the film coefficient increases by a factor of 1.7.

Heat Transfer to Boiling Liquids

Heat transfer to boiling liquids occurs in a number of operations, for example, distillation and evaporation. Heat is transferred by both conduction and convection in a process further complicated by the phase change that occurs at the heating boundary. When boiling is induced by a heater in contact with a pool of liquid, the process is known as pool boiling. Liquid movement is derived only from heating effects. In other systems, the boiling liquid may be driven through or over heaters, a process referred to as boiling with forced circulation.

Pool boiling

If a horizontal heating surface is in contact with a boiling liquid, a sequence of events occurs as the temperature difference between the surface and the liquid increases.

When ΔT is small, the degree of superheating of the liquid layers adjacent to the surface is low and bubble formation (growth and disengagement), if present, is slow. Liquid disturbance is small and heat transfer can be estimated from expressions for natural convection given, for example, in Eq. 29.

Vapor formation becomes more vigorous and bubble chains rise from points that progressively increase in number and finally merge. This movement increases liquid circulation. This phase is called nucleate boiling and is the practically important regime. A peak flux occurs and a maximum heat-transfer coefficient is obtained. At this point, ΔT is known as the critical temperature drop. For water, the value lies between 40 and 50°F (4–10°C). The critical temperature drop for organic liquids is somewhat higher. Beyond the critical temperature, vapor formation is so rapid that escape is inadequate and a progressively larger fraction of the heating surface becomes covered with a vapor film. This represents a transition from nucleate boiling to film boiling. When this transition is complete, the vapor entirely covers the surface, film boiling is fully established, and the heat flux rises again.

The low heat-transfer coefficient renders film boiling undesirable and equipment is designed for and operated at temperature differences which are less than the critical temperature drop. If a constant temperature heat source such as steam or hot liquid is employed, exceeding the critical temperature drop results simply in a drop in heat flux and process efficiency. If, however, a constant heat-input source is used, as in electrical heating, decreasing heat flux as the transition region is entered causes a sudden and possibly damaging increase in the temperature of the heating element; such damage is known as boiling burn-out.

Boiling heat-transfer coefficients depend upon both the physical character of the liquid and the nature of the heating surface. Through the agencies of wetting, roughness, and contamination, the latter greatly influences the formation, growth, and disengagement of bubbles in the nucleate boiling regime. There is, at present, no reliable method of estimating the boiling coefficients of heat transfer from the physical properties of the system.

Boiling inside a vertical tube

Heat transfer to liquids boiling in vertical tubes is common in evaporators. If a long tube of suitable diameter in which liquid lies at a low level is heated, the pattern of boiling is established Fig. 5. At low levels, boiling may be suppressed by the imposed head [Fig. 5(a)]. Higher in the tube, bubbles are produced, which rise and coalesce [Fig. 5(b)]. Slug formation due to bubble coagulation occurs [Fig. 5(c) and 5(d)]. The slugs finally break down [Fig. 5(e)]. Escape is hindered and both liquid and vapor move upward at an increasing speed. Draining leads to separation of the phases, giving an annular film of liquid dragged upward by a core of high velocity vapor [Fig. 5(f)]. In long tubes, the main heat transfer takes place in this region by either forced convection or nucleate boiling. At low temperature differences between wall and film, heat transfer occurs quietly as in forced convection. This is the normal regime in a climbing film evaporator and heat flux can be calculated from correlations of the type given in Eq. 40. At high temperatures differences, nucleate boiling takes place in the film and the vigorous movement leads to an increase in heat transfer coefficient.

Boiling with forced circulation

In many systems, movements other than those caused by boiling are imposed. For example, boiling in agitated vessels is common in many batch processes. The boiling heat-transfer coefficients depend upon the properties of the liquid, the nature of the surface, and the agitation. The coefficients obtained are slightly higher than those of pool boiling. Inside the tubes, the pattern of forced circulation

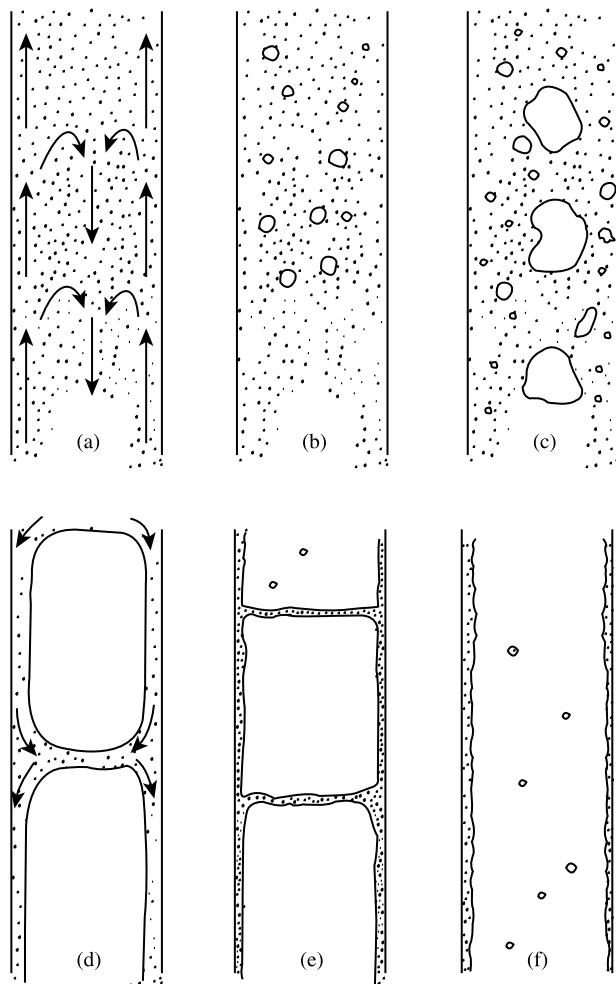


Fig. 5 Boiling in a narrow vertical tube. (a) Boiling suppressed by head, natural convection is shown; (b) bubble formation; (c) slug formation due to bubble coagulation; (d) fully developed slug flow; (e) breakdown of slugs at high vapor rates; (f) annular-flow-climbing film.

boiling is similar to that described in the previous section. Coefficients, however, are higher because the velocities attained are higher.

Heat Transfer from Condensing Vapors

When a saturated vapor is brought into contact with a cool surface, heat is transferred to the surface and a liquid condenses. The vapor may consist of a single substance or a mixture, some components of which may be non-condensable.

The process is described by the following sequence. The vapor diffuses to the boundary where actual condensation takes place. In most cases, the condensate

forms a continuous layer over the cooling surface, draining under the influence of gravity. This is known as film condensation. The latent heat liberated is transferred through the film to the surface by conduction. Although this film offers considerable resistance to heat flow, film coefficients are usually high.

Dropwise condensation

Under some surface conditions, the condensate does not form a continuous film. Droplets are formed which grow, coalesce, and then run from the surface. As a fraction of the surface is always directly exposed to the vapor, film resistance is absent, and heat-transfer coefficients, which may be ten times those of film condensation, are obtained. This process is known as dropwise condensation. Although highly desirable, its occurrence, which depends upon the wettability of the surface, is not predictable and cannot be used as a basis for design.

Condensation of a pure vapor

For film condensation, a theoretical analysis of the laminar flow of a liquid film down an inclined surface and the progressive increase in thickness due to condensation yields Eq. 34 for the mean-transfer coefficient h_m

$$h_m = \text{constant} \left(\frac{\rho^2 k^3 \lambda g}{T \eta x} \right)^{0.25} \quad (34)$$

where λ is the latent heat of vaporization; ρ , k , and η are the density, thermal conductivity, and viscosity of the liquid, respectively; ΔT is the difference in temperature between the surface and vapor. Experimentally determined coefficients confirm the validity of Eq. 34. In practice, however, the coefficients are somewhat higher due to disturbance of the film arising from a number of actors. As the condensation rate rises, the thickness of the condensate layer increases and the film coefficient falls. However, a point may be reached in long vertical tubes at which flow in the layer becomes turbulent. Under these conditions, the coefficient rises again and Eq. 34 is not valid. Coefficients may also be increased if high vapor velocities induce ripples in the film.

Condensation of mixed vapors

If a mixture of condensable and noncondensable gases is cooled below its dew point at a surface, the former condenses, leaving the adjacent layers richer in the latter, thus creating an added thermal resistance. The condensable fraction must diffuse through this layer to reach the film of condensate and heat-transfer coefficients are normally very much lower than the corresponding value for the pure vapor. For example, the presence of 0.5% of

air has been found to reduce the heat transfer by condensation of steam by as much as 50%.

Heat Transfer by Radiation

Of the radiation that falls on a body, a fraction a is absorbed, a fraction r is reflected, and a fraction t is transmitted. These fractions are called absorptivity, reflectivity, and transitivity, respectively. Most industrial solids are opaque such that the transmissivity is zero and Eq. 35 holds.

$$a + r = 1 \quad (35)$$

Reflectivity, and therefore, absorptivity depend greatly on the nature of the surface. The limiting case, that of a body which absorbs all and reflects none of the incident radiation, is called a black body.

Exchange of radiation

The exchange of radiation is based upon two laws. The first, known as Kirchoff's law, states that the ratio of the emissive power to the absorptivity is same for all bodies in thermal equilibrium. The emissive power of a body, E , is the radiant energy emitted from unit area in unit time. A body of area A_1 and emissivity E_1 , therefore, emits energy at a rate $E_1 A_1$. If the radiation falling on unit area of the body is E_b , the rate of energy absorption is $E_b a_1 A_1$, where a_1 is the absorptivity. At thermal equilibrium, $E_b a_1 A_1 = E_1 A_1$. For another body in the same environment, $E_b a_2 A_2 = E_2 A_2$, leading to Eq. 36

$$E_b = \frac{E_1}{a_1} = \frac{E_2}{a_2} \quad (36)$$

For a black body, $a = 1$. The emissive power is therefore E_b . The black body is a perfect radiator and is used as the comparative standard for other surfaces. The emissivity e of a surface is defined as the ratio of the emissive power E of the surface to the emissive power of a black body at the same temperature E_b , as shown by Eq. 37.

$$e = \frac{E}{E_b} \quad (37)$$

Emissivity is numerically equal to absorptivity. As emissive power varies with wavelength, the ratio should be quoted at a particular wavelength for many materials. However, the emissive power is a constant fraction of the black body radiation, that is, the emissivity is constant. These materials are known as gray bodies.

The second fundamental law of radiation, known as the Stefan-Boltzmann law, states that the rate of energy

emission from a black body is proportional to the fourth power of the absolute temperature T , as shown by Eq. 38

$$E = \sigma T^4 \quad (38)$$

where E is the total emissive power and σ is the Stefan-Boltzmann constant. It is sufficiently accurate to say that the heat emitted in unit time Q from a black body of area A is given by

$$Q = \sigma A T^4$$

and for a body that is not perfectly black by Eq. 39

$$Q = \sigma e A T^4 \quad (39)$$

where e is the emissivity.

The net energy gained or lost by a body can be estimated by these laws. The simplest case is that of a gray body in black surroundings. These conditions, in which none of the energy emitted by the body is reflected back, are approximately those of a body radiating to atmosphere. If the absolute temperature of the body is T_1 , the rate of heat loss is $\sigma e A T_1^4$ (Eq. 40), where A is the area of the body and e its emissivity. Surroundings at a temperature T_2 emit radiation proportional to σT_2^4 , and a fraction, determined by area and absorptivity a , is absorbed by the body; this heat is $\sigma a A T_2^4$, and as absorptivity and emissivity are equal, Eq. 40 is valid.

$$\text{Net heat-transfer rate} = \sigma e A (T_1^4 - T_2^4) \quad (40)$$

If part of the energy emitted by a surface is reflected back by another surface, the calculation of radiation exchange is more complex. Equations for various surface configurations take the following general form:

$$Q = F_1 F_2 \sigma A (T_A^4 - T_B^4)$$

where F_1 and F_2 are factors determined by the configuration and emissivity of surfaces at temperatures T_A and T_B .

MASS TRANSFER

Mass transfer in unit operations has been described in detail previously in this encyclopedia (2). the following is a brief review to complete the overview of unit processes in pharmacy.

In mass transfer operations, two immiscible phases are normally present, one or both of which are fluid. In general, these phases are in relative movement and the rate at which a component is transferred from one phase to the other is greatly influenced by the bulk movement imposed

upon the fluids. In most drying processes, for example, water vapor diffuses from a saturated layer in contact with the drying surface into a turbulent air stream. The boundary layer, as described earlier, consists of a sublayer in which flow is laminar and an outer region in which flow is turbulent. The mechanism of diffusion differs in these regimes. In the laminar layer, movement of water vapor molecules across streamlines can occur only by molecular diffusion. In the turbulent region, the movement of relatively large units of gas, called eddies, from one region to another causes mixing of the components of the gas. This is called eddy diffusion. Eddy diffusion is a more rapid process and although molecular diffusion is still present, its contribution to the overall movement of material is small. In still air, eddy diffusion is virtually absent and evaporation occurs by molecular diffusion.

Molecular Diffusion of Gases

Transport of material in stagnant fluids or across streamlines of a fluid in laminar flow occurs by molecular diffusion. Two adjacent compartments, separated by a partition containing pure gases A or B may be envisaged. Random movement of all molecules occurs so that after a period of time, molecules are found quite remote from their original positions. If the partition is removed, some molecules of A move toward the region occupied by B, their number depending on the number of molecules at the point considered. Concurrently, molecules of B diffuse toward regions formerly occupied by pure A.

Ultimately, complete mixing occurs. Before this point in time, a gradual variation in the concentration of A occurs along an axis, designated x , which joins the original compartments. This variation, expressed mathematically, is $-dC_A/dx$, where C_A is the concentration of A. The negative sign arises because the concentration of A decreases as the distance x increases. Similarly, the variation in the concentration of gas B is $-dC_B/dx$. The rate of diffusion of A, N_A , depends on the concentration gradient and the average velocity with which the molecules of A move in the x direction. This relationship is expressed by Fick's law, given by Eq. 41

$$N_A = -D_{AB} \frac{dC_A}{dx} \quad (41)$$

where D is the diffusivity of A in B, a property proportional to the average molecular velocity and, therefore, dependent on the temperature and pressure of the gases. The rate of diffusion N_A is usually expressed as the number of moles diffusing across unit area in unit time. As

with the basic equations of heat transfer, Eq. 41 indicates that the rate of a process is directly proportional to a driving force, which, in this context, is a concentration gradient.

This basic equation can be applied to a number of situations. Restricting discussion exclusively to steady-state conditions, in which neither dC_A/dx or dC_B/dx change with time, equimolecular counterdiffusion is considered first.

Equimolecular counterdiffusion

If no bulk flow occurs in an element of length dx , the rates of diffusion of the two gases A and B must be equal and opposite, that is, $N_A = -N_B$.

The partial pressure of A changes by dP_A over the distance dx . Similarly, the partial pressure of B changes dP_B . As there is no difference in total pressure across the element (no bulk flow), dP_A/dx must equal $-dP_B/dx$. For an ideal gas, the partial pressure is related to the molar concentration by the relation

$$P_A V = n_A RT$$

where n_A is the number of moles of gas A in a volume V . As the molar concentration C_A is equal to n_A/V , therefore

$$P_A = C_A RT$$

Consequently, for gas A, Eq. 40 can be written as in Eq. 42,

$$N_A = -\frac{D_{AB}}{RT} \frac{dP_A}{dx} \quad (42)$$

where D_{AB} is the diffusivity of A in B. Similarly,

$$N_B = -\frac{D_{BA}}{RT} \frac{dP_B}{dx} = \frac{D_{AB}}{RT} \frac{dP_A}{dx}$$

It, therefore, allows that $D_{AB} = D_{BA} = D$. If the partial pressure of A at x_1 is P_{A1} and at x_2 is P_{A2} , integration of Eq. 42 gives Eq. 43.

$$N_A = -\frac{D}{RT} \left(\frac{P_{A2} - P_{A1}}{x_2 - x_1} \right) \quad (43)$$

A similar equation may be derived for the counterdiffusion of gas B.

Diffusion through a stationary, nondiffusing gas

An important practical case arises when a gas A diffuses through a gas B, there being no overall transport of gas B. It arises, for example, when a vapor formed at a drying surface diffuses into a surrounding gas. At the liquid surface, the partial pressure of A is dictated by the temperature. For water, it would be 12.8 mm Hg (1.7 kPa) at 15°C. Some distance away the partial pressure is lower,

and the concentration gradient causes diffusion of A away from the surface. Similarly, a concentration gradient for B must exist, the concentration being lowest at the surface. Diffusion of this component takes place toward the surface. There is, however, no overall transport B so that diffusional movement must be balanced by bulk flow away from the surface. The total flow of A is, therefore, the diffusional flow of A plus the transfer of A associated with this bulk movement.

Molecular Diffusion in Liquids

Equations describing molecular diffusion in liquids are similar to those applied to gases. The rate of diffusion of material A in a liquid is given by Eq. 40.

$$N_A = D \frac{dC_A}{dx}$$

Fick's law for steady-state equimolar counterdiffusion is then expression by Eq. 44

$$N_A = -D \frac{C_{A2} - C_{A1}}{x_2 - x_1} \quad (44)$$

where C_{A2} and C_{A1} are the molar concentrations at points x_2 and x_1 respectively.

Equations for diffusion through a layer of stagnant liquid can also be developed. The applicability of these equations is, however, limited because diffusivity in a liquid varies with concentration. In addition, unless the solutions are very dilute, the total molar concentration varies from point to point. These complications do not arise with diffusion in gases.

Diffusivities in liquids are very much lower than diffusivities in gases, commonly by a factor of 10^4 .

Mass Transfer in Turbulent and Laminar Flow

As already explained, movement of molecules across the streamlines of a fluid in laminar flow can occur only by molecular diffusion. If the concentration of a component A varies in a direction normal to the streamlines, the molar rate of diffusion is given by Eq. 44.

When a fluid flows over a surface, the surface retards the adjacent fluid region, forming a boundary layer. If the flow throughout the fluid is laminar, the equation for molecular diffusion may be used to evaluate the mass transferred across the boundary layer. In most important cases, however, the flow in the bulk of the fluid is turbulent. The boundary layer is then considered to consist of three distinct flow regimes. In the region of the boundary layer most distant from the surface, flow is

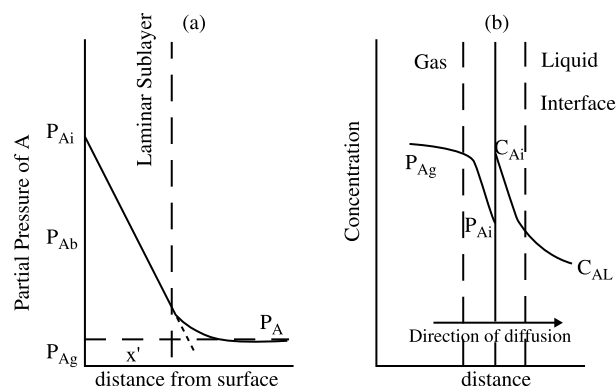


Fig. 6 Mass transfer at (a) a boundary and (b) an interface.

turbulent, and mass transfer is the result of the interchange of large portions of the fluid. Mass interchange is rapid and concentration gradients are low. As the surface is approached, a transition from turbulent to laminar flow occurs in the transition or buffer region. In this region, mass transfer by eddy diffusion and molecular diffusion are of comparable magnitude. In a fluid layer at the surface, a fraction of a millimeter thick, laminar flow conditions persist. This laminar sublayer, in which transfer occurs by molecular diffusion only, offers the main resistance to mass transfer as shown in Fig. 6(a). As flow becomes more turbulent, the thickness of the laminar sublayer and its resistance to mass transfer decrease.

An approach to the evaluation of the rate of mass transfer under these conditions lies in the postulation of a film, the thickness of which offers the same resistance to mass transfer as the combined laminar, buffer, and turbulent regions. The analogy with heat transfer by conduction and convection is exact and quantitative relations between heat and mass transfer can be developed for some situations. This, however, is not attempted here. The postulate of an effective film is explained by Fig. 6(b). As a gas flows over a surface, equimolecular counterdiffusion of components A and B occurs, A away from the surface and B toward the surface. The variation in partial pressure of A with distance from the surface is shown in Fig. 6a. At the surface, the value of P_{Ai} , a linear fall to P_{Ab} occurs over the laminar sublayer. Beyond this, the partial pressure falls less steeply to the value P_A at the edge of the boundary layer. A value slightly higher than this is P_{Ag} , the average partial pressure of A in the entire system. In general, the gas content of the laminar layer is so small that P_A and P_{Ag} are virtually equal. If the molecular diffusion were solely responsible for diffusion, the partial pressure P_{Ag} would be reached at some

fictional distance x' from the surface, over which the concentration gradient $(P_{Ai} - P_{Ag})/x'$ exists. The molar rate of mass transfer would then be

$$N_A = \frac{D}{RT} \cdot \frac{P_{Ai} - P_{Ag}}{x}$$

However, x' is not known, and Eq. 45 may be written as

$$N_A = -\frac{k_g}{RT}(P_{Ai} - P_{Ag}) \quad (45)$$

where k_g , is a mass transfer coefficient, the units of which are cm sec^{-1} . As $C_A = P_A/RT$, it can also be written as

$$N_A = k_g(C_{Ai} - C_{Ag})$$

where C_{Ai} and C_{Ag} are the gas concentrations at either side of the film. Similar equations describe the diffusion of B in the opposite direction.

Diffusion across a liquid film is described by Eq. 46

$$N_A = k_l(C_{Ai} - C_{Al}) \quad (46)$$

where C_{Ai} is the concentration of component A at the interface and C_{Al} is its concentration in the bulk of the phase.

In all cases, the mass-transfer coefficient depends upon the diffusivity of the transferred material and the thickness of the effective film. The latter is largely determined by the Reynolds number of the moving fluid, that is, its average velocity, density, and viscosity, and some linear dimension of the system. Dimensional analysis gives the following relation:

$$\frac{kd}{D} = \text{const}(Re)^q \left(\frac{\eta}{\rho D} \right)^r$$

where Re is the Reynolds number, k the mass transfer coefficient, D the diffusivity, and d a dimension characterizing the geometry of the system.

This relation is analogous to the expression for the heat transfer by forced convection given earlier. The dimensionless group kd/D corresponds to the Nusselt group in heat transfer. The parameter $\eta/\rho D$ is known as the Schmidt number and is the mass-transfer counterpart of the Prandtl number. For example, the evaporation of a thin liquid film at the wall of a pipe into a turbulent gas is described by the equation

$$\frac{kd}{D} = 0.023 Re^{0.8} Sc^{0.33}$$

where Sc is the Schmidt number. Although this equation expresses experimental data, comparison with Eq. 41

again demonstrates the fundamental relation of heat and mass transfer.

Similar relations have been developed empirically for other situations. The flow of gases normal to and parallel to liquid surfaces can be applied to drying processes, and the agitation of solids in liquids can provide information for crystallization or dissolution.

Interfacial Mass Transfer

On a macroscopic scale, the interface can be regarded as a discrete boundary. On the molecular scale, however, the change from one place to another takes place over several molecular diameters. Due to movement of molecules, this region is in a state of violent change, the whole surface layer changing many times a second. Transfer of molecules at the actual interface is, therefore, virtually instantaneous and the two phases are, at this point in equilibrium.

As the interface offers no resistance, mass transfer between phases can be regarded as the transfer of a component from one bulk phase to another through two films in contact, each characterized by a mass-transfer coefficient. This is the two-film theory and the simplest of the theories of interfacial mass transfer. For the transfer of a component from a gas to a liquid, the theory is described in Fig. 6b. Across the gas film, the concentration, expressed as partial pressure, falls from a bulk concentration P_{Ag} to an interfacial concentration P_{Ai} . In the liquid, the concentration falls from an interfacial value C_{Ai} to bulk value C_{Al} .

At the interface, equilibrium conditions exist. The break in the curve is due to the different affinity of component A for the two phases and the different units expressing concentration. The bulk phases are not, of course, at equilibrium and it is the degree of displacement from equilibrium conditions that provides the driving force for mass transfer. If these conditions are known, an overall mass-transfer coefficient can be calculated and used to estimate the mass-transfer rate.

Transfer of a component from one mixed phase to another, as described above, occurs in several processes. Liquid-liquid extraction, leaching, gas absorption, and distillation are examples. In other processes such as drying, crystallization, and dissolution, one phase may consist of only one component. Concentration gradients are set up in one phase only, with the concentration at the interface given by the relevant equilibrium conditions. In drying, for example, a layer of air in equilibrium (i.e., saturated) with the liquid is postulated at the liquid surface and mass transfer to a turbulent air stream is described by Eq. 44, the

interfacial concentration being the saturation concentration. The rate of solution is determined by the difference between the interfacial concentration and the concentration in the bulk solution, and the mass-transfer coefficient.

REFERENCES

1. Markovitz, R.E. Heat Transfer in Process Vessels. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 2001.
2. Graham Nairn, J. Mass Transfer in Unit Operations. *Encyclopedia of Pharmaceutical Technology*; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 2001.

BIBLIOGRAPHY

Beard, J. *Dynamics of Fluids in Porous Media*; Dover Publications, Inc.: Mineola, NY, 1972.

Bird, R.B.; Stewart, W.E.; Lightfoot, E.N. *Transport Phenomena*; John Wiley & Sons: New York, 1960.

Chhabra, R.P. *Bubbles, Drops and Particles in Non-Newtonian Fluids*; CRC Press: Boca Raton, FL, 1993.

Cheremisinoff, N.P. *Practical Fluid Mechanics for Engineers and Scientists*; Technomic Publishing Co., Inc.: Lancaster, PA, 1990.

Cho, Y.I.; Hartnett, J.P. Non-Newtonian Fluids in Circular Pipe Flow. *Adv. Heat Transfer* **1982**, *15*, 59–141.

Coulson, J.M.; Richardson, J.R. *Chemical Engineering*; Pergamon Press: New York, 1977; 1.

The Mathematics of Diffusion; Crank, J. Ed.; reprinted 1992 Oxford University Clarendon Press: Oxford, 1975.

Cussler, E.L. *Diffusion Mass Transfer in Fluid Systems*; Cambridge University Press: New York, 1984.

Fayed, M.E.; Otten, L. *Handbook of Powder Science and Technology*; Van Nostrand Reinhold, Co. New York, 1984.

Hartnett, J.P.; Kostic, M. Heat Transfer to Newtonian and Non-Newtonian Fluids in Rectangular Ducts. *Adv. Heat Transfer* **1989**, *19*, 24–356.

McCabe, W.L.; Smith, J.C.; Harriott, P. *Unit Operations of Chemical Engineering*; McGraw Hill, Inc.: New York, 1993.

Meyer, R.E. *Introduction to Mathematical Fluids Dynamics*; Dover Publications, Inc.: Mineola, NY, 1982.

Neumann, B.S. The Flow Properties. In *Advances in Pharmaceutical Sciences*; Bean, H.S., Beckett, A.H., Carless, J.E. Eds.; Academic Press: New York, 1967; 2, 181–221.

Orr, C. *Particulate Technology*; Macmillan: New York, 1966.

Ozisik, M.N. *Boundary Value Problems of Heat Conduction*; Dover Publications, Inc.: Mineola, NY, 1968.

Prandtl, L.; Tietjens, O.G. *Fundamentals of Hydro- and Aeromechanics*; Dover Publications, Inc.: Mineola, NY, 1957.

Perry, R.H.; Chilton, C.H. *Chemical Engineer's Handbook*; McGraw Hill: New York, 1973.

Taylor, R.; Krishna, R. *Multicomponent Mass Transfer*; John Wiley & Sons: New York, 1993.

VACCINES AND OTHER IMMUNOLOGICAL PRODUCTS

Suresh K. Mittal
Harm HogenEsch
Kinam Park

Purdue University, West Lafayette, Indiana, U.S.A.

VACCINES

Introduction

The concept of vaccination was introduced in the late 18th century by Edward Jenner when he used cowpox virus as a vaccine to protect humans against smallpox virus infections. This led to the development of vaccines over the next 2 centuries to provide protection against various bacterial and viral pathogens. Undoubtedly, the effective vaccination against infectious diseases is the best method of reducing suffering of human and animals caused by viral, bacterial, and parasitic infections. Over the last 200 years, the technology of vaccine development and production has not changed significantly. This usually involves the use of either a killed pathogen combined with an adjuvant or a live pathogen with reduced virulence. Apart from the tremendous success of killed and attenuated virus vaccines over the years, many of such vaccines do not provide satisfactory protection, and there are a number of other disadvantages associated with these vaccines. Additionally, there are important pathogens against which attempts to develop effective vaccines using traditional approaches were unsuccessful. Various protective viral antigens (envelope and/or capsid proteins or glycoproteins and other viral proteins) and bacterial antigens (surface, internal, or fimbria proteins; bacterial polysaccharides; bacterial toxins; and other proteins involved in bacterial metabolism) have been identified as potential vaccine candidates. These protective antigens are used by various means to develop effective vaccines. The field of vaccine technology is not limited to infectious diseases but has shown potential in other areas, such as cancer treatment, reproduction, and modulation of animal productivity. An overview of vaccine strategies is depicted in Fig. 1.

Conventional Vaccines

Inactivated vaccines

Inactivated (killed) pathogenic organisms can be used in vaccines. This is the simplest way to produce vaccines,

provided the organisms can be cultured easily. Therefore, this method is often first tested to develop a potential-vaccine. As with any other technique of vaccine production, this procedure is only good for some organisms. There are a number of methods of inactivating pathogenic organisms; the most common are treatment with chemicals (formalin, formaldehyde, or propiolactate), heat, or γ -irradiation. In some instances, the procedure of inactivation may enhance antigenicity of some antigens important in protection. Inactivated vaccines usually result in good humoral immune response after multiple inoculations. Because inactivated vaccines in general fail to elicit effective mucosal and cell-mediated immune responses, they may provide limited protection against mucosal and intracellular pathogens. Failure to inactivate the pathogenic organisms completely could result in disease instead of protection. During the 1950s, some lots of poliovirus vaccine were not inactivated completely (1, 2). Now, the methods used to detect residual infectivity are more stringent, therefore, inactivated vaccines are considered safe with extremely low or no chance of infection.

There have been instances in which inactivated vaccines led to atypical disease or enhanced disease severity. For example, in the 1960s, formalin-inactivated respiratory syncytial virus (RSV) vaccine actually enhanced the disease symptoms when vaccinated children were naturally exposed to RSV (3, 4). It was later discovered that a change in the antigenicity of RSV F and G glycoproteins (5) resulted not only in alteration in humoral immune response but also in the Th1 and Th2 components of the CD4⁺ T-cell response to RSV (6).

Live attenuated vaccines

Mostly attenuated organisms are being used as live virus vaccines; however, in some instances, even virulent organisms could be used, provided they are not administered via the natural route of infection. For example, human adenovirus types 4 and 7 may cause acute respiratory infections in humans when administered via the oronasal route but provide protection when given orally in enteric-coated capsules (7).

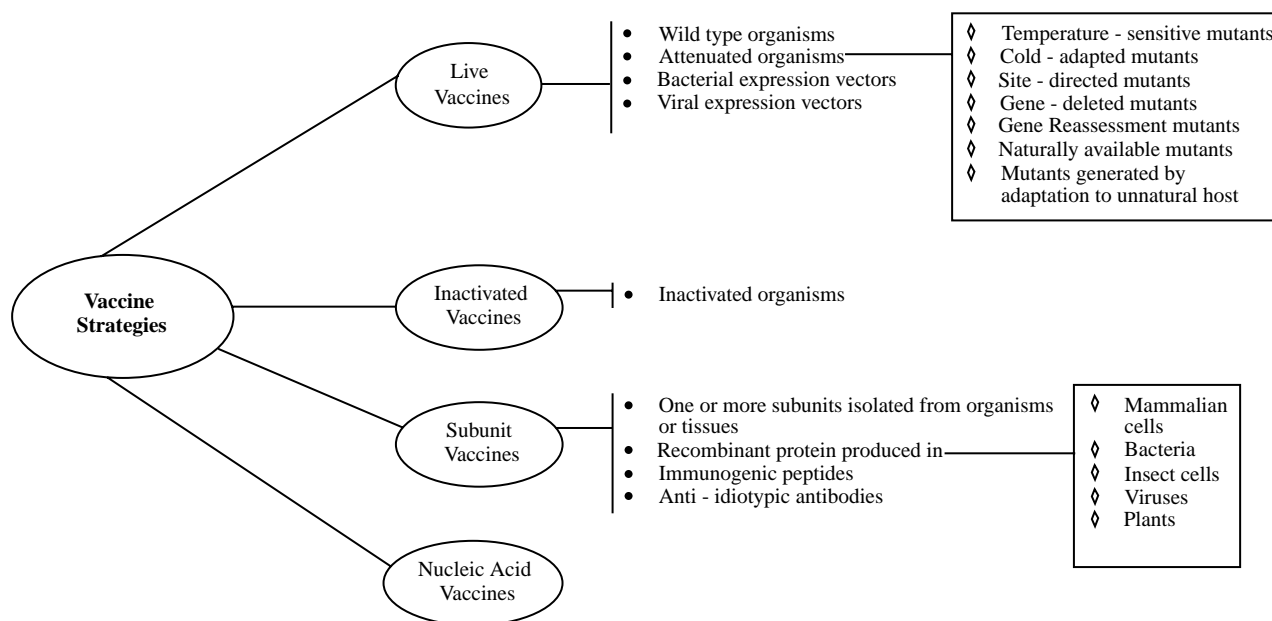


Fig. 1 Overview of vaccine strategies.

There are different ways to attenuate pathogens for vaccine production. Attenuation of organisms can be achieved by growing them under abnormal conditions, which include cultivation in unnatural hosts or cell lines. Some organisms are attenuated when they replicate at different pH levels and/or temperatures. In cells infected with multiple viruses with a segmented genome (e.g., influenza virus, reovirus), genome segments are randomly recombined in the progeny. This process of recombination is known as reassortment and is also useful in generating attenuated viruses. A natural pathogen of one host may be attenuated for another host, e.g., vaccinia virus worked as an attenuated vaccine for smallpox virus eradication program during the 1960s and 1970s, and turkey herpesvirus works as an attenuated vaccine for Marek's disease virus (a chicken herpesvirus). In an inoculated host, the attenuated organism replicates without causing disease symptoms, thereby leading to induction of immune response somewhat similar to the natural infection with the disease-causing organism. The Bacille Camet-Guerin (BCG) strain of *Mycobacterium tuberculosis* was attenuated after more than 200 passages on media containing increasing amounts of bile. The Sabin poliovirus vaccine was attenuated by a number of passages in monkeys and in monkey kidney epithelial cells (8). Measles virus was initially adapted to monkey kidney cells and subsequently attenuated in duck embryo and human tissue culture cell lines (9–11).

Temperature-sensitive (*ts*) mutants have proven to be the most useful type of mutants for a number of viruses and bacteria because of their conditional-lethal phenotype. The (*ts*) mutants are produced by alteration in the nucleotide sequence of a gene so that the resulting protein product of the gene is unable to assume or maintain its functional configuration at the nonpermissive (37–39°C) temperature. The protein, however, is able to assume a functional configuration at the permissive temperature (32–34°C), e.g., herpesviruses, adenoviruses, and influenza viruses. Thus, these mutants can replicate in mucosal sites with a lower temperature, e.g., the nasal cavity, but are unable to cause systemic infections and disease.

A number of advantages associated with live vaccines are that: 1) they are cheap to produce because the inoculum dose is relatively less, 2) they require fewer inoculations, 3) they do not require adjuvants, 4) they elicit both humoral and cell-mediated immune responses, and 5) they can be inoculated by the natural route of infection. Some of the disadvantages associated with live vaccines are that 1) they are usually less stable than inactivated vaccines and may require refrigeration for storage, 2) some of these vaccines under certain situations may revert to virulent form in the host and thereby lead to clinical disease, 3) they may not be recommended for immunosuppressed, immature, older, or pregnant hosts, 4) they may have a low level of residual virulence, and 5) they may be contaminated with other adventitious organisms.

Recombinant Vaccines

Recombinant vector vaccines

Viral vectors: For the development of an effective vaccine strategy for protection against mucosal pathogens such as respiratory and enteric viruses, a vaccine-delivery system that can induce a protective mucosal immunity in the form of secretory IgA antibody, in addition to a systemic immune response, is extremely important. The route of vaccine delivery also plays an important role in determining the type of resultant immunity induced. A number of viruses, such as adenoviruses, poxviruses, herpesviruses, picornaviruses, togaviruses, orthomyxoviruses, paramyxoviruses, and others, have demonstrated considerable potential as vectors for antigen delivery at mucosal surfaces (12). Immunogenic foreign epitopes can be expressed on the virus surface by modifying the viral capsid or envelope protein (13). A wide variety of foreign viral antigens has been expressed in viral vectors, and vaccination-challenge studies in experimental animals have demonstrated moderate to complete protection. Immunization with such vectors leads to the foreign viral antigen expression similar to that of natural infection without causing disease. Antigenic peptides are expressed along with major histocompatibility (MHC) class I and class II antigens and, thus, result in both humoral and cytotoxic T-cell responses.

Both adenovirus-and poxvirus-based vectors have a number of common advantages including that 1) vector construction is easy, ii) relatively high levels of foreign protein expression are easily attained, 3) relative thermostability, 4) they have a large capacity for foreign DNA insertion, 5) vector derivatives are nonpathogenic, and 6) they have a wide host range. More than one foreign antigen can be expressed in the same vector to provide protection against a number of diseases by inoculation with a single vector.

Vaccinia virus expressing rabies glycoprotein has been licensed for use to control rabies in the wildlife population, especially raccoons, foxes, skunks, and coyotes (14). Baits containing a live vaccinia-rabies glycoprotein recombinant virus vaccine are distributed in the rabies endemic area with the intention that rabies-susceptible wild animals that eat these baits will become immunized against rabies virus (15), and this approach has demonstrated satisfactory results. Vaccinia virus expressing the F and H gene of rinderpest virus has shown potential for its use to control rinderpest in developing countries (16, 17).

To increase the safety of viral vectors for immunocompromised hosts and to control their indiscriminate spread, replication defective viral vectors have been developed. These vectors can be grown to high titers in

vitro, but they are defective for in vivo replication. Replication-defective vectors undergo an abortive infection in an inoculated host that leads to foreign antigen expression similar to replication-competent vectors. Replication-defective adenovirus vectors are generated by deleting the early region 1 (E1) genes (18,19). E1-deleted vectors can be grown in an E1-complementing cell line, and animals immunized with such vectors elicit a protective immune response (20). Avian poxviruses grow normally in avian cells but would result in an abortive infection in mammalian hosts. Dogs and cats immunized with an avipox-rabies glycoprotein recombinant are protected against rabies virus infection (21).

Bacterial vectors: Similar to viral expression vectors, attenuated bacteria can be developed as vectors for foreign gene expression and delivery for the purpose of multivalent vaccines. Immunogenic foreign epitopes can be expressed on bacteria surfaces by modifying cell surface proteins, fimbria, or flagella. It has been demonstrated that *M. bovis* BCG strain induces both strong humoral and cell-mediated immunity, therefore, it has been developed as a delivery vector with the assumption that foreign proteins expressed by *M. bovis* in inoculated individuals will also raise a strong protective immune response (22). Because *Salmonella* and *Vibrio* colonize in the intestinal tract, attenuated strains of these bacteria were developed as vectors for mucosal delivery (23–26).

Various bacterial vectors have been used to express a number of bacterial (*B. pertussis*, *S. pneumoniae*, *Y. pestis*, and *L. monocytogenes*), viral (herpesvirus, influenza virus, human immunodeficiency virus, simian immunodeficiency virus, and hepatitis B virus), and parasitic (*S. mansoni*, and *L. major*) antigens (26). Significant improvements in attenuation of bacteria, and the stability, localization, and expression levels of heterologous antigens are required to market the bacterial vector-based vaccines for use in humans or animals.

To enhance foreign gene expression, “balanced lethal,” plasmid-based expression vehicles have been developed (27). A foreign antigen may form inclusion bodies or localize in intracellular compartment of the vector thereby affecting the type, levels, and duration of immune response elicited against the antigen. The *Escherichia coli* α -hemolysin secretion system (HSS) that includes HlyB, HlyD, and TolC is involved in exporting the HlyA-fused foreign antigens to extracellular compartment (28). Using the HSS system for attenuated *Shigella dysenteriae*, the expression and secretion of Shiga toxin-B subunit were obtained (29).

Gene-deleted vaccines

Many attenuated vaccines are derived after introduction of random mutations in the genomes of various pathogens. In

situations in which these random mutations may be point mutations, attenuated organisms may regain virulence owing to back mutations. Because of our increased understanding of virulence of various pathogens at the molecular level, one or more genes responsible for virulence has been identified in many pathogens. The genes associated with virulence may be genes involved with nucleic acid replication and other nonstructural and structural components of the organism. This has made it possible to delete one or more of these genes involved in virulence—another strategy to produce safer attenuated vaccines.

Pseudorabies virus has been attenuated by deleting genes associated with viral virulence. These genes include the thymidine kinase gene (nonstructural protein) involved in viral DNA replication and the gC, gG, and gE genes (nonessential glycoproteins) involved in virus assembly (30, 31). A gene-deleted vaccine of pseudorabies virus has proved highly effective in controlling this viral infection under field conditions. It has been demonstrated that *Salmonella typhimurium* aroA, aroB, and aroC deletion mutants fail to grow in its host because of the absence of aromatic amino acid production. These genes have been targeted to reduce the virulence of the bacterium. *S. typhimurium* gene-deleted mutants are capable of replication at least for a short period in its host, thus raising a protective immune response (32). Vaccination with gene-deleted vaccines also allows eradication of wild-type pathogens from the population. Because antibodies against the deleted gene product will only be developed in infected animals, it is feasible to differentiate between vaccinated and naturally infected animals (33, 34). The process of gene deletion not only attenuates the pathogen but also offers a unique opportunity to insert foreign genes for developing viral or bacterial-vectored vaccines.

Subunit vaccines

A subunit vaccine consists of one or more immunogenic epitopes, proteins, or other components of a pathogenic organism. Immunogenic epitopes can be chemically synthesized and are known as peptide vaccines, e.g., peptide vaccine candidates for foot-and-mouth disease virus (35, 36). The pathogen could be disrupted, and one or more immunogenic proteins such as bacterial cell wall proteins; flagella or pili; and viral envelope, capsid, or nucleoproteins can be purified. The isolation of such components in purified form is sometimes cumbersome and expensive. However, bacterial exotoxins can be easily purified, inactivated, and used as toxoid vaccines.

A number of expression systems including bacteria, yeasts, mammalian cells, insect cells, and plants are now

available for foreign protein expression. High amounts of a foreign protein can be produced in a bacterial-expression system at a low cost. Because scale-up and downstream processing have been well worked out for bacterial-expression systems, they are usually first tested for subunit vaccine production. Many of the immunogenic proteins, especially of viral origin, require secondary modifications that are important for their antigenicity. A bacterial-expression system may produce proteins of altered immunogenicity because the bacterial system lacks many posttranslational processes. However, some viral glycoproteins expressed in bacteria induce protective immunity, e.g., the gp 70 gene of feline leukemia virus (37). A yeast-expressed hepatitis B virus surface antigen (HbsAg)-based subunit vaccine is currently in use for humans and has demonstrated excellent protection against hepatitis B virus infection (38). This vaccine is an excellent example of the potential of recombinant subunit vaccines for providing protection against many viral and bacterial infections.

Because mammalian cells are known to process viral glycoproteins to their functional form by secondary modifications, they are considered one of the means to produce viral antigens for subunit vaccine production. However, the expression of such proteins in mammalian cells is usually too low. It was demonstrated that the stable expression of the transmembrane anchor-deleted form of many viral glycoproteins in mammalian cells results in the secretion of truncated products in the medium in large quantities that could be used as a subunit vaccine without further purification. However, the removal of transmembrane anchor may potentially alter antigenicity of the secreted protein. A number of viral glycoproteins that were expressed either in mammalian or in insect cells and secreted in form of proteins were suitable for providing protective immune response include F and G genes of respiratory syncytial virus (39), the HN and F genes of parainfluenza virus (40), and the gD gene of bovine herpesvirus type 1 (41).

Immunogenic antigen production in plants: In the past decade, significant progress has been made in the stable integration and expression of a wide variety of genes in plant cells, resulting in the creation of novel plants for agricultural and industrial use. The inserted genes confer resistance to insect pathogen and herbicides; enhanced tolerance to drought, salt, and frost; and improved agricultural production. Undoubtedly, improvements in plant attributes by genetic engineering will have a great impact on agriculture production. However, it has been estimated that the major economic (over 90%) gain of plant biotechnology will result from the use of plants as bioreactors to produce high-valued products such as vaccines, industrial enzymes, and other pharmaceuticals.

Production of subunit vaccines in mammalian cells is usually expensive because of the low level of foreign gene expression and high processing cost. High levels of foreign gene expression can be obtained in bacteria and yeast, but many animal viral or mammalian proteins expressed in these systems fail to undergo proper secondary modifications such as glycosylation, phosphorylation, sulfation, etc. Therefore, these recombinant proteins may have altered antigenicity. Because most mechanisms regulating secondary modifications of proteins are present in plants, transgenic plants offer an attractive alternative to produce functional viral, bacterial, or parasitic proteins in large quantities at a very low cost for subunit vaccine production (42). Similarly, the production of functional multimeric antibody molecules in plants has made it possible to manufacture antibodies in bulk amounts for passive immunization (43).

Two major strategies have been devised to produce foreign proteins in plants. These are: 1) the stable integration of chimeric gene into the plant genome under a suitable constitutive or inducible plant promoters (44, 45), and 2) manipulation of plant pathogenic viruses (46). Foreign protein expression in plants usually range from 0.01 to 1% of the total plant protein.

The hepatitis B virus (HBV) surface antigen HBsAg produced in transgenic tobacco elicits an immune response when injected in mice (47). Mice fed transgenic potato tuber expressing B subunit of heat-labile enterotoxin (LT-B) of enterotoxigenic *E. coli* developed antibodies to LT-B, particularly IgA antibodies (44). Dalsgaard et al. (46) demonstrated that immunization of mink with the VP2 capsid protein of mink enteritis virus, expressed in cowpea after infection with modified cowpea mosaic virus, elicited a protective immune response. Protection against challenge with virulent foot-and-mouth disease virus (FMDV) in mice inoculated with the structural protein VP1 of FMDV produced in transgenic *Arabidopsis* has been shown (45). It has been hypothesized that transgenic plants could serve as “edible vaccine,” thereby providing a very inexpensive mean of oral immunization (48).

Anti-idiotypic vaccines

Another approach to provide protective immune response is the use of anti-idiotypic antibodies as vaccines. Antibodies have unique sequences in the variable (V) region in their binding site known as “idiotypic determinants”. Some of the idiotypic determinants make up the antigen-binding site (paratope) of the antibody. The part of the antibody that binds to the antigen is called a paratope. Antibodies to a specific paratope of an idiotypic mimic the epitope of immunizing antigen and are known as anti-idiotypic antibodies. Thus, anti-idiotypic antibodies are mirror images of

antigens and can be used instead of immunogens to elicit a protective immune response. Monoclonal anti-idiotypic antibodies could serve as a source of antigen. Anti-idiotypic vaccines are useful in cases in which actual antigen is poorly immunogenic or similar to host antigens. Some of the pathogens against which anti-idiotypic vaccines have been tested include *Listeria monocytogenes*, *Streptococcus pneumoniae*, hepatitis B virus, Semliki forest virus, and Sendai virus (49, 50). This type of vaccine is still in the developmental stage.

DNA Vaccines

Immunization of mammalian hosts with a plasmid DNA containing a gene under control of a heterologous promoter has introduced a new approach in the area of recombinant vaccine design. The introduced DNA is taken up by cells, and the gene of interest is expressed. The cells expressing the foreign antigen are recognized by the host immune system, leading to humoral and cell-mediated immune responses. DNA vaccines can also be called polynucleotide vaccines or nucleic acid (NA) vaccines. Such vaccines appear to have the primary advantages of both attenuated and inactivated vaccines but without their known limitations. NA vaccines elicit an immune response similar to that obtained with live attenuated vaccines. They also provide safety similar to that of inactivated vaccines, however, without the obvious side effects of adjuvants or animal-derived proteins.

The concept of NA vaccine evolved from initial studies in experimental animals in which the inoculation with naked plasmid DNA resulted in a protective immune response (51). After inoculation into a muscle, the efficiency of cellular uptake of the naked DNA is poor, and a large portion of the DNA is degraded before it reaches the nucleus for transcription. To increase the efficiency of DNA uptake by host cells and to reduce DNA degradation within the cell, a number of delivery systems,

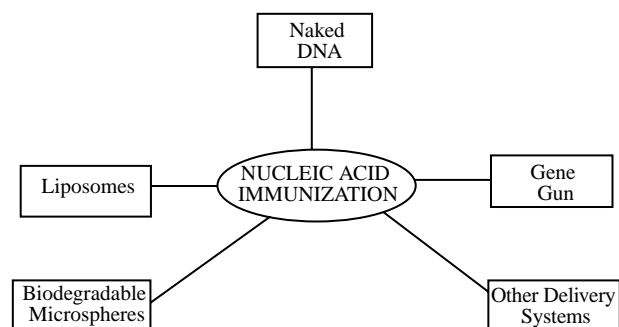


Fig. 2 Methods of nucleic acid delivery.

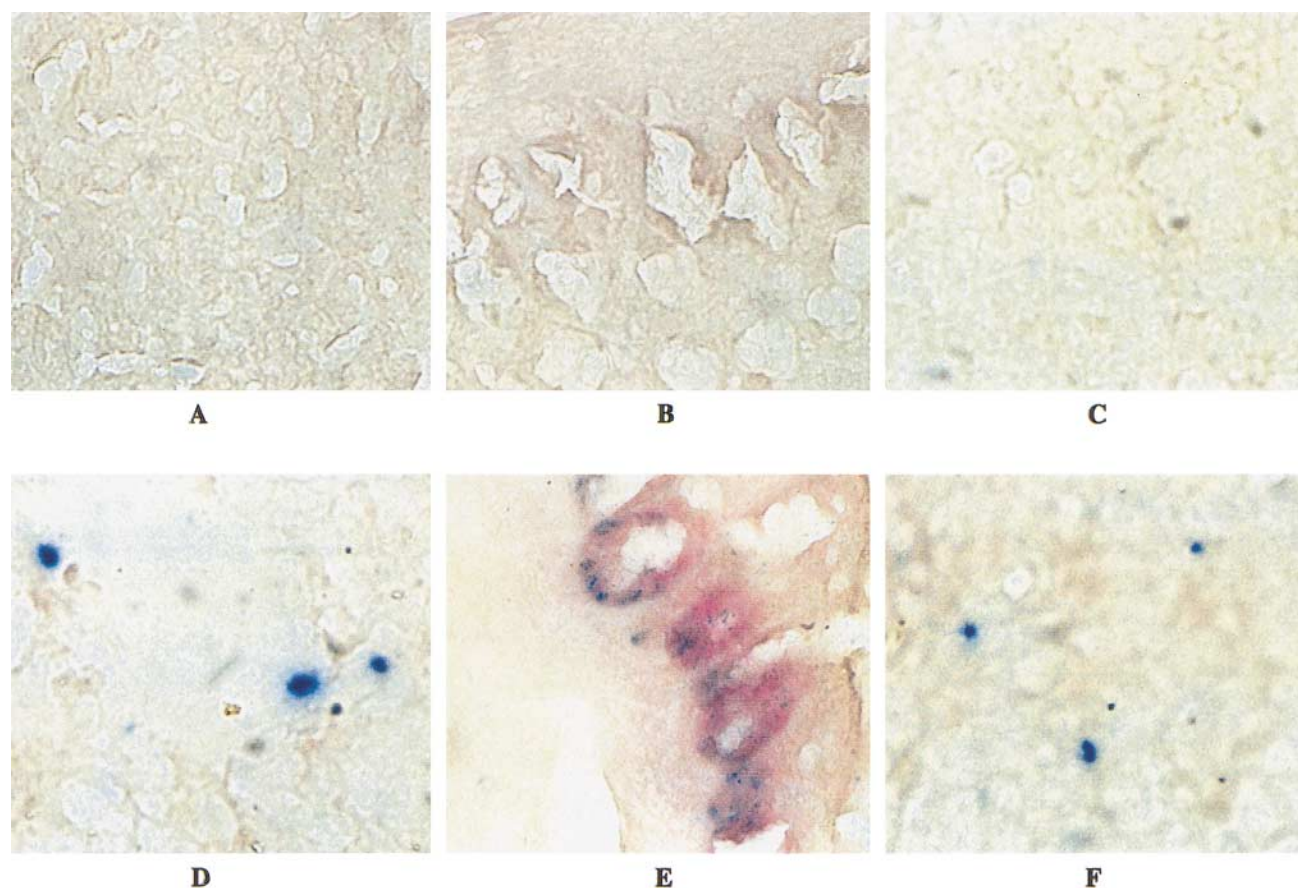


Fig. 3 β -Galactosidase expression in tissues of mice inoculated orally with alginate microspheres containing plasmid DNA: (A) liver, (B) intestine, and (C) spleen sections from the animal inoculated with microspheres containing bovine adenovirus type 3 (BAd3). (D) Liver, (E) intestine, and (F) spleen sections from the animal inoculated with microspheres containing LacZ plasmid +, BAd3. (From Ref. 57.)

such as bombardment with gold microparticles coated with NA (52, 53), incorporation of NA into liposomes and other polycationic lipids (54, 55), biological erodable polymers (56), and others, have been developed (Fig. 2). Recently, it has been demonstrated that alginate microspheres can be used for the encapsulation, delivery, and expression of plasmid DNA (57) (Fig. 3). Inoculation of mice with microspheres containing both plasmid DNA and bovine adenovirus type 3 (BAd3) resulted in a significant increase in transgene expression compared with those inoculated with microspheres containing only the plasmid DNA. As with other delivery systems, alginate microspheres led to a stronger mucosal or systemic immune response, depending on route of inoculation (58). Because alginate microspheres are most likely taken up by macrophages and dendritic cells, it may have a positive effect on the type of immune response elicited.

A number of factors that have an impact on the level and type of immune response produced by an NA vaccine

include the type of immunogen, the dosage and number of inoculations, the heterologous regulatory sequences, the delivery system, the route of inoculation, and the presence or absence of immunomodulatory molecules. A variety of immunogenic antigens including HIV-1, SIV, HTLV-1, influenza virus, hepatitis B virus, hepatitis C virus, herpesvirus, *M. tuberculosis*, *Leishmania*, malaria, and many more have been expressed by NA vaccines and have demonstrated encouraging results (59–63).

Adjuvants

Adjuvants are compounds that, when administered in combination with antigens, enhance the immune response to those antigens. This enhanced immunogenicity can be measured as an increase of antigen-specific antibody levels in serum and/or mucosal secretions, a response against an increased number of epitopes, an increase of cell-mediated immune responses, or a combination thereof. Adjuvants

are particularly important for the induction of protective immune responses against weak immunogens such as subunit vaccines. The mechanisms by which adjuvants enhance the immunogenicity of antigens are not completely understood, but they include immunostimulation, altered processing of antigens, and sustained release of antigens (depot effect). A different type of immune response is obtained by administration of antigens via the oral route, and this has different delivery requirements.

Many compounds can act as adjuvants. Their classification is made difficult by the variety in chemical composition and the overlapping, often poorly understood, mechanisms of action. Only aluminum adjuvants are approved by the FDA for use in human vaccines. Quil A is a saponin that is commonly used as an adjuvant in veterinary vaccines and is also a component of immune-stimulating complexes (ISCOM). These adjuvants are addressed in some detail below. A detailed discussion of other types of adjuvants can be found in recent books (64–66) and reviews (67) on this subject.

Immunostimulation

The immune system can be divided into the adaptive immune system, comprising of B and T lymphocytes, and the innate immune system, which includes neutrophils, macrophages, dendritic cells, and soluble factors such as the complement system. The innate immune system plays a critical role in the activation of the adaptive immune system. Dendritic cells are antigen-presenting cells that integrate the signals from the innate immune system and activate T-cells and possibly B-cells. T-cells have antigen-specific receptors that recognize peptides displayed by MHC I molecules ($CD8^+$ cytotoxic T-cells) and MHC II molecules ($CD4^+$ T helper cells). Engagement of the antigen-specific T-cell receptor is not sufficient, and T-cells also need to receive costimulatory signals delivered via CD28 and CD40-ligand. Dendritic cells express both MHC I and MHC II and, on activation, increase the expression of the costimulatory molecules CD80 and CD86 (ligands for CD28) and CD40. The signals that activate dendritic cells include microbial molecules. The innate immune system is equipped with receptors (called pattern-recognition receptors) that can recognize molecules that are expressed by pathogens, but not by mammalian cells, and alert the innate immune system on infection. These molecules, pathogen-associated molecular patterns, include lipopolysaccharides (LPS), mannose, and bacterial DNA with unmethylated CpG motifs. In addition, dendritic cells are stimulated by host cell components that are expressed and/or released by cells when they undergo stress and pathologic cell death (necrosis). The identity of these components, called danger signals, is uncertain but

may include heat shock proteins. The microbial molecules and danger signals can directly activate dendritic cells, or they can activate other components of the innate immune system resulting in the secretion of cytokines and other mediators that activate dendritic cells. The activated dendritic cells, in turn, activate T- and B-cells.

Immune responses can be divided into type 1 and type 2, based on the pattern of cytokine secretion and functional outcome of the immune response. Type 1 immune responses are characterized by secretion of IFN- γ , production of IgG2a in mice, and activation of macrophages, NK cells, and cytotoxic T-cells. Type 2 responses are characterized by secretion of IL-4, IL-5, and IL-13 and by IgG1 and IgE production. The responses are reciprocally regulated. How the polarization of the immune response toward type 1 or type 2 is determined is not exactly understood. IL-12 is an important factor that drives the type 1 response, and IL-4 is implicated in the type 2 response. Microbial products such as LPS and bacterial DNA stimulate the secretion of IL-12 by dendritic cells and preferentially induce type 1 immune responses.

It is likely that the primary mechanism by which adjuvants stimulate the immune response is by direct or indirect signaling through pattern-recognition and danger signal receptors. Very strong adjuvants are often composed of or include microbial components such as LPS and mycobacteria or derivatives thereof. These type of adjuvants bind to pattern-recognition receptors to stimulate IL-12 production and a type 1 immune response. Coadministration of cytokines can directly activate and influence dendritic cells and the outcome of the immune response. This was clearly demonstrated with an experimental *Leishmania* vaccine using IL-12 as an adjuvant. Immunization of genetically susceptible BALB/c mice with a *Leishmania* antigen did not result in protection, but when IL-12 was injected with the antigen, the mice became markedly resistant to infection. The effect of IL-12 correlated with increased IFN- γ and decreased IL-4 secretion by antigen-specific T-cells in vitro.

Altered processing of antigens

Most T-cells that carry the α - β -T-cell receptor do not recognize and react with intact proteins. Instead, the T-cells recognize small peptides that are derived from proteins and that are linked to MHC I and MHC II molecules. The MHC I-linked peptides are generated in the cytoplasm (endogenous pathway) and recognized by $CD8^+$ T-cells. Proteins in the cytoplasm are degraded by a complex of proteolytic enzymes, the proteasome, and the peptides are transported into the rough endoplasmic reticulum where they associate with MHC I molecules. Peptide binding

stabilizes the MHC I molecules, and the complexes are transported to the cell surface. In contrast, proteins that enter cells by endocytosis are partially degraded into peptides in endosomal vesicles. The peptides bind MHC II molecules that have been transported from the endoplasmic reticulum to the endosomes. The MHC II-peptide complexes are then displayed on the cell surface and are available for recognition by CD4⁺ T-cells.

Vaccines that contain single proteins or inactivated pathogens can readily activate CD4⁺ T-cells because the antigens are endocytosed and processed by MHC II-positive antigen-presenting cells. Activation of the CD4⁺ T-cells can result in a type 1 or a type 2 immune response, depending on the type of adjuvant included. However, such vaccines usually do not activate CD8⁺ cytotoxic T-cells because activation of CD8⁺ T-cells requires processing of antigen via the endogenous pathway. Certain adjuvant formulations such as liposomes, the saponin QS-21, and poly-(lactic-co-glycolic acid) (PLGA) are able to induce cytotoxic T-cell responses to protein antigens (68). These adjuvants appear to target some of the injected antigens into the cytosol of antigen-presenting cells for processing via the endogenous pathway. The mechanism by which this occurs is not known.

Sustained release of antigens

The slow and continued release of antigens has been postulated to induce a strong immune response through continued activation of the immune system. This may contribute to the adjuvant effect of aluminum-based adjuvants and mineral oils. Newer technologies may allow for the design of vaccines that release antigens from a depot at certain time intervals after a single injection. One example is the use of poly PLGA microspheres for encapsulation of antigens. By varying the polymer composition and size of the microspheres, the release of antigen can be varied. Pulsatile release of antigen can be attained by combining multiple variations of PLGA microspheres in a single dose of the vaccine (69). Relatively little is known about the desired pattern of antigen release to obtain a maximal response. It was recently suggested that continued release of antigen is not desirable for the induction of strong memory cell responses. Mathematical models may help design appropriate strategies for the release of antigens from depots after a single injection (70).

Aluminum

Aluminum adjuvants in human vaccines are either aluminum hydroxyphosphate (commonly referred to as aluminum phosphate) or aluminum oxyhydroxide (aluminum hydroxide) (71). Aluminum-based vaccines are prepared

by adsorption of antigen to commercial aluminum hydroxide or aluminum phosphate gels or by mixing antigen with alum (potassium aluminum sulfate), resulting in precipitation. The alum-precipitated adjuvants resemble aluminum phosphate in their chemical and physical properties (71). The surface charge and morphology of the aluminum adjuvants affect their adsorptive capacity. The rate and degree of adsorption are further dependent on the pH, ionic strength of the antigen solution, and isoelectric point of the antigen.

Aluminum adjuvants are universally used in diphtheria-tetanus-pertussis (DTP) vaccines and in most hepatitis B vaccines and have an excellent safety record. They are not ideal adjuvants, however, because the enhancement of the immune response is relatively weak, they are not effective with all antigens, and, most important, they only enhance the humoral (type 2) immune response and have little effect on the cell-mediated (type 1) immune response.

The mechanism by which aluminum enhances the immune response is not clear. Early studies suggested that aluminum adjuvants slowly release the adsorbed antigen over time (depot effect). However, recent experiments demonstrated that antigens are rapidly desorbed after injection in animals. Moreover, aluminum phosphate enhanced the immune response to DNA-encoded antigen after DNA immunization, clearly indicating that adsorption may not be critical to the adjuvant effect of aluminum compounds. These data indicate that aluminum enhances the immune response via other mechanisms. A satisfactory explanation of the adjuvant effect of aluminum also needs to take into account its selective mode of enhancing the immune response, i.e., a predominant type 2 immune response. Aluminum adjuvants induced differentiation toward type 2 immune responses, even in the absence of IL-4 or IL-13. Aluminum stimulated a type 1 and type 2 immune response in genetically engineered mice with a defective IL-4 and IL-13 response, suggesting that aluminum-induced IL-4 and/or IL-13 secretion suppresses the type 1 response but are dispensable for a type 2 response in intact animals (72). The lack of a type 1 immune response is a drawback for the use of aluminum in vaccines for intracellular pathogens and tumors. A recent study demonstrated that aluminum adjuvant with adsorbed IL-12 induces a strong type 1 response, indicating that it is possible to overcome the aluminum-induced suppression of type 1 responses (73).

Saponins

The saponins of the bark of the *Quillaja saponaria* Molina tree have long been known to have immunostimulatory activity. A partially purified fraction, Quil A, has reduced

toxicity and more potent adjuvant activity and is used in veterinary vaccines. Quil A can be further fractionated into fractions that have different degrees of toxicity. QS-21 is a less toxic component with strong adjuvant activity. Saponins probably act by direct stimulation of the immune system (74). They stimulate both the humoral (primarily IgG2a antibodies in the mouse) and cell-mediated immune responses. QS-21 causes protein antigens to be processed and presented via the MHC I pathway, resulting in cytotoxic T-cell responses. Cytokine analysis indicates that QS-21 stimulates type 1 cytokine production.

Immune-stimulating complexes (ISCOMs) are 30–40 nm particles consisting of Quil A, cholesterol, antigen, and phospholipids (74). They are used in a commercial vaccine for equine influenza. ISCOM-adjuvanted vaccines stimulate a strong humoral and cell-mediated immune response caused by the immunostimulatory actions of Quil A and targeting of the particles to macrophages. As with Quil A, ISCOMs target antigens for processing via the MHC I pathway, resulting in induction of cytotoxic T-cell responses.

Delivery of Vaccines

Parenteral versus mucosal route

The success of vaccination depends primarily on the method of presenting the antigen to the host immune system. Antigens have usually been delivered by parenteral (such as intravenous, intramuscular, intraperitoneal, intradermal, and subcutaneous) administration, but recent studies have shown that other routes of delivery such as intranasal, oral, and transdermal delivery have also been effective. In some cases, vaccination through mucosal routes resulted in better responses in IgA production. Because nonparenteral vaccine delivery presents many obvious advantages, numerous attempts have been made on the development of nonparenteral delivery of vaccines.

Parenteral route: Parenteral vaccination remains the immunization method of choice for most antigens because it provides more effective immune response than do any other routes of vaccination in most cases. Every year millions of people receive inactivated influenza vaccine by parenteral administration. Subcutaneous vaccination with inactivated influenza vaccine is known to induce simultaneous immune responses in the blood and upper respiratory tract of subjects. The immune response, i.e., the increase in the number of influenza virus-specific antibody-secreting cells in peripheral blood and tonsils, increased rapidly to reach a peak within 1 week after vaccination (75). Parenteral vaccination of a DNA vaccine encoding glycoprotein D of herpes simplex virus type 2

resulted in systemic cellular and humoral responses. The mucosal humoral responses generated by intramuscular and intradermal vaccination were comparable with those obtained by mucosal vaccination. The DNA vaccine was able to stimulate a response in the Peyer's patches, a major inductive site for mucosal responses (76). For many other antigens, however, the usefulness of parenteral vaccination is limited by the insufficient induction of mucosal immune responses.

Parenteral vaccination is difficult for those living in the developing countries where medical care is not well-established. Vaccination of a large number of subjects using hypodermic needles, which is a highly labor-intensive procedure requiring healthcare personnel, is not practical. The problem becomes even more significant for vaccination of millions of animals. For example, vaccination for routine control of Newcastle disease in chickens by intramuscular injection requiring individual handling of the birds is not practical (77). Recent advances in needleless injectable systems have made the parenteral vaccination easier, but it still requires individual handling. Examples of needleless injection systems are Powder-Ject[®], Medi-Jector[®], Biojector[®], Vitajet[®], Bio-Set[®], and Intraject[®]. They all use high pressure released in a very short period to deliver drugs through the skin. A jet-immunization technique was used for intraoral administration of DNA in the cheek, resulting in high IgA mucosal responses (78). The intraoral jet-injection technique for DNA vaccine delivery has the advantages of being a simple and rapid way to administer the DNA in solution and to provoke specific mucosal IgA after administration in the mucosal-associated lymphoid tissue.

The results of parenteral vaccination depend on the route of administration. For plasmid DNA vaccines, the highest levels of antibodies were induced by intramuscular and intravenous injections, although significant titers were also obtained with sublingual and intradermal delivery. Delivery to the skin by the gene gun induced exclusively IgG1 antibodies (Th2-like) at 4 weeks and only very low IgG2a levels at later times. Other routes, such as intraperitoneal, intraperineal, subcutaneous, intranasal inhalation, intranasal instillation, intrarectal, intravaginal, ocular, and oral, did not result in significant immune responses (79).

Dual-chamber syringe. For delivery of two established vaccines (e.g., polyribosyl ribitol phosphate conjugated to tetanus toxoid and diphtheria–tetanus–whole cell pertussis and inactivated poliovirus vaccine) at the same time, a dual-chamber syringe delivery system can be used. The proximal chamber may contain a vaccine in the freeze-dried solid state, and the distal chamber contains a vaccine in the liquid formulation that allows reconstitution of the

vaccine in the proximal chamber. The immune response by the dual-chamber delivery of vaccination was equivalent to that by the separate-injection method of vaccination. The dual-chamber syringe can be used for safe and effective delivery of two different vaccines that are not yet available as a single formulation for pediatric applications (80). The primary advantage of the dual-chamber syringe is that it reduces the cost of vaccine delivery and, at the same time, increases the vaccine acceptability and coverage rate of vaccines (81).

Mucosal route: Vaccination through mucosal routes provides new avenues of vaccination with a unique advantage of mucosal immunity, that may not be obtained, through parenteral vaccination. Mucosal immunization presents a realistic alternative to parenteral administration for inducing protective immune responses. Vaccination by mucosal route provides a number of advantages over parenteral vaccination. First, mucosal vaccination does not involve hypodermic needles, which are not user-friendly. Second, the total surface area of the mucosal surfaces in the gastrointestinal, respiratory, and urogenital tracts where many infectious pathogens come into contact with the host is huge. Thus, preventing infections at the mucosal surface

provides an immunological first line of defense against diseases (82). This makes priming of the mucosal-associated lymphoid tissue (MALT) by vaccination most desirable. Parenteral vaccination alone is quite often insufficient in inducing mucosal immune responses, because stimulation of the MALT usually requires direct contact between the immunogen and the mucosal surface (83). The mucosal tissues are protected by interconnected local immune system, which is essentially separated from systemic immunity (84). In a common mucosal-defense system, an antigen interacting with localized lymphoid tissue can stimulate IgA precursor cells that may then migrate to other mucosal surfaces to elicit immune reaction in other mucosal tissues. It is known that the mucosal immune system produces 70% of the body’s antibodies (85). Fig. 4 shows a schematic description of the common mucosal-immunization system. Mucosal delivery of numerous antigens by a variety of routes (oral, nasal, tracheal, and rectal) has been shown to elicit immunity at mucosal surfaces mediated by secretory IgA. The presence of MALT indicates that mucosal vaccination at a certain site in the body can be achieved by mucosal immunization at the distal site of the body. Although the mucosal and

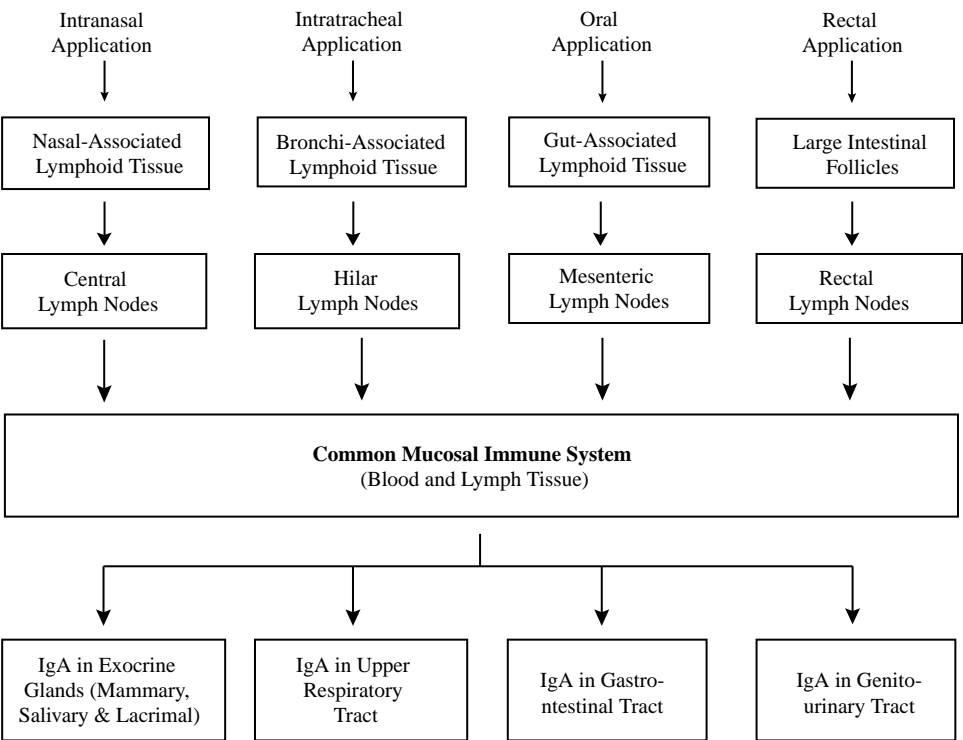


Fig. 4 Mucosal immunization and production of IgA antibodies in various mucosal surfaces via the common mucosal-immunization system. Nasal and rectal vaccinations usually result in IgA production in upper respiratory tract and genitourinary tract, respectively, whereas effector sites by oral vaccination are expected to include many mucosal surfaces.

systemic humoral immune systems function essentially independent of each other, an antigen administered by one route can modify responsiveness to subsequent immunization by an alternate route (86).

Oral vaccination of the various mucosal routes, oral vaccination is the most preferable mode of vaccination because of its ease of use and low cost of manufacturing (87). Furthermore, the gastrointestinal (GI) tract provides the largest component of the mucosal immune system that has been well-characterized. Oral administration of vaccines has high acceptability, by avoidance of injection, to individuals of all ages. Fig. 5 shows the current understanding of oral vaccination. After oral vaccination, an antigen, which is typically loaded in microspheres, is taken up by M-cells in the Peyer's patch of the gut-associated lymphoid tissue. The antigen is then passed to the macrophages and B-cells (B). These cells in turn present the antigen to T helper lymphocytes. These cells migrate into the blood via the mesenteric lymph nodes (MLN) and the thoracic duct (TD). These cells subsequently localize in the effector sites, i.e., mucosal membranes of the GI tract, upper respiratory tract, genitourinary tract, and glandular tissue. At the effector sites, the migrating B-cells develop into plasma cells that produce IgA antibodies. Polymeric IgA is then released as secretory IgA (sIgA) through epithelial cells.

The maximal intestinal immunization can be achieved by intra-Peyer's patch immunization, and thus this method can be used to screen oral vaccine candidate antigens without the added complication of simultaneously testing

oral-delivery systems (88). Immunization of subjects against *Helicobacter pylori* by intra-Peyer's patch resulted in an 84–91% reduction in *H. pylori* infection compared with unimmunized controls. The therapeutic efficacy of the recombinant *H. pylori* urease vaccine in mice was shown to be comparable with that achieved with the combined antibiotic/antacid treatment in humans. The oral vaccination is preferred to conventional treatment of ulcers because it is a very simple and quick procedure compared with long-term conventional treatment. In addition, vaccines use the defense mechanisms of the body to establish long-lasting immunity (89).

One of the limitations of oral vaccination is that it does not always induce sufficient immunity. There are a few good reasons for this. First, the GI tract is designed to digest proteins by acidic and enzymatic degradation for absorption. Because most antigens are proteins in nature, they may be degraded by enzymes in the GI tract as well as by acids in the stomach. This is why soluble antigens administered orally are not effective. Thus, prevention of the antigen degradation is the first step toward successful oral vaccination. Adding protease inhibitors before oral vaccination may induce complete immunity, but this approach is not practical. There are many different enzymes that may not be inhibited by a particular protease inhibitor, and, more important the action of protease inhibitors may not occur at the same time that the antigens are present in the GI tract. Second, the systemic uptake of antigens from the GI tract is very poor. Even after oral intake of gram quantities of antigen, only a nanogram range of antigenic material was found to pass the intestinal barrier (90). It is also possible that for certain antigens, oral vaccination may simply be less effective than parenteral vaccination in induction of systemic immunity (91). The protection resulting from oral vaccination is known to last for a relatively short period, ranging from a few months to 1 year. To obtain the desirable immunity equivalent to systemic immunization, oral vaccination requires much higher and more frequent oral doses. The use of highly effective adjuvants in oral vaccine formulation may result in strong and long-lasting immunity in mucosal tissues.

The issues of degradation of antigens in the GI tract and the poor systemic uptake of antigens from the GI tract have led to encapsulation of antigens in microparticles (also called microcapsules or microspheres). Antigens that are encapsulated in microparticles are protected from degradation, and the microparticulate nature allows better uptake by the M-cells in the Peyer's patches. A large number of studies have shown that antigens orally delivered in microparticles resulted in good mucosal immunity. It is noted here that virus itself can be regarded as a particulate vaccine-delivery system. Many viruses are

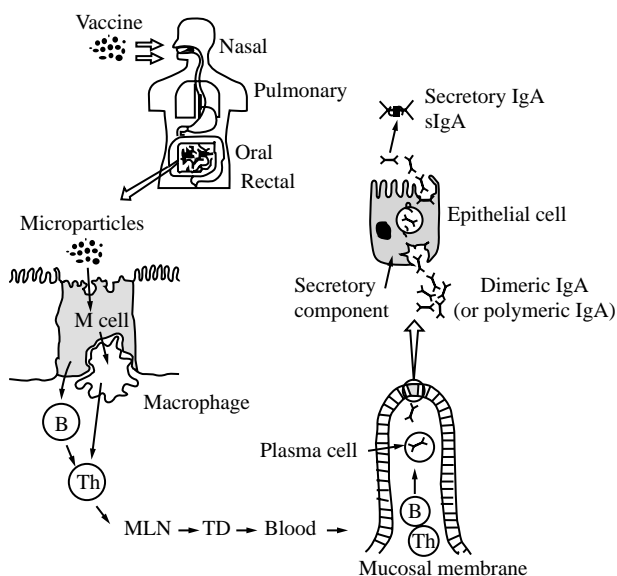


Fig. 5 Mucosal immunization by oral vaccination.

highly effective in inducing immunization after oral vaccination. Norwalk virus, which is a major cause of epidemic gastroenteritis, was immunogenic in healthy human adults even when administered without adjuvants (92). Influenza virus can also elicit immune response after oral administration. Successful oral vaccination relies on targeting of microparticles to the Peyer's patches. It is known that the surface chemistry of microparticles affects the targeting to and uptake by M-cells in the Peyer's patches (93). The exact relationships between the surface chemistry and the uptake by Peyer's patches, however, have not been fully understood. Development of better oral vaccines requires understanding of such relationships.

Intranasal vaccination route has received growing interest for noninvasive immunization. Intranasal immunization has been quite effective for various vaccine-delivery systems. Both solution and microsphere formulations tend to show good immune responses after intranasal administration. Immunization of mice with tetanus toxoid, in solution and microsphere-encapsulated formulations, resulted in high levels of specific IgG and IgA antibodies (94). Nasal vaccine delivery is known to be superior to oral delivery in inducing specific IgA and IgG antibody responses in the upper respiratory tract (95). Nasal immunization is also known to be preferable to the oral route for distant mucosal vaccination that might be used to prevent adhesion of pathogens to the urogenital tract (95). It is interesting to note that the volume of the nasally instilled vaccine is important (94). The larger-volume (e.g., 50 μ L) of microsphere suspension resulted in the higher percentage of particles entering the lungs than did the lower, volume (e.g., 10 μ L) instillation.

It is generally believed that microspheres that adhere to the nasal mucus elicit better immune response, and for this reason, many microspheres made of mucoadhesive polymers, such as chitosan, have been used extensively in the preparation of nasal vaccine formulations.

Transdermal vaccination or transcutaneous immunization, is attractive, because it does not require specially trained personnel necessary for needle injections. Topical application of antigens to intact skin has shown promising results for the administration of DNA-based vaccines. Noninvasive gene delivery by pipetting adenovirus- or liposome-complexed plasmid DNA onto the outer layer of skin was able to achieve localized transgene expression within a restricted subset of skin in mice. It also elicited an immune response against the protein encoded by the DNA (96).

For improved results, transdermal electroporation was also tried to explore the feasibility of nonadjuvant, needle-free skin immunization (97). The transdermal electroporation route elicited higher responses to a myristylated

peptide than did intradermal immunization. For diphtheria toxoid, however, the result was the opposite. It appears that transdermal electroporation is a promising technique for nonadjuvant skin immunization, especially with low-molecular-weight, weakly immunogenic antigens. Topical application of antigen and cholera toxin or bacterial exotoxin to the skin surface resulted in detectable antigen-specific IgG in plasma and mucosal secretions (98, 99). It appears that transcutaneous immunization can induce potent, protective immune responses to both systemic and mucosal challenge (100).

Pulmonary vaccination is especially useful in mass vaccination campaigns. A conventional method of pulmonary delivery of drugs using metered-dose, propellant-driven, small-particle aerosols was used to deliver killed whole bacterium vaccines. The results showed good stimulation of mucosal immunity against respiratory infections in animals (101). Recent advances in powder inhaler devices have made it possible to deliver vaccines via the pulmonary route using dry powder inhalation technologies (102). Dry powder vaccine in the size range from 1 to 5 μ m in diameter is used for the maximum alveolar (deep lung) deposition (101).

Direct gene transfer into the respiratory system can be carried out for either therapeutic or immunization purposes. Cells in the lung can take up and express plasmid DNA whether it is administered in naked form or formulated with cationic liposomes. For a given dose of DNA, the results can be improved when the DNA is mixed with the minimum amount of lipid that can complex it completely (103). Such a complex formation can be considered a formation of microparticles that can enhance cellular uptake and subsequent immune responses.

Parenteral and mucosal combination vaccination:

The combination of mucosal and systemic immunization routes (e.g., parenteral immunization followed by oral immunization or vice versa) generally induces mucosal immune responses that are superior to immunization by either route alone (91). Pigs showed some protection after intramuscular inoculation with formalin-inactivated *M. hyopneumoniae* vaccine in incomplete Freund's adjuvant and a booster inoculation with the same vaccine in microspheres onto the mucosal surface of Peyer's patches by a surgical operation (104).

Antigen delivery systems

The primary goal of antigen-delivery systems is to maintain a stable dosage form during storage and, when administered to present antigens to elicit a vigorous immune response in vivo. It is necessary to develop vaccine formulations that would preserve the antigen and deliver it

to a specific target organ over a desired period. Continuous release or multiple pulsatile release during the desired period would eliminate the inconvenience of multiple vaccine administration for obtaining satisfactory immune responses. The antigen-delivery system plays one of the most crucial roles in the outcome of the immunization. The way that antigens are delivered affects the immune response significantly. Currently, antigen-delivery systems are classified into two systems: live attenuated microorganisms and nonliving microparticulate systems.

Live attenuated organisms: Live attenuated bacteria and viruses have been used not only as vaccines but also as a delivery system that elicits humoral, mucosal, and cellular immune responses against exogenous antigens. Since the success with live attenuated oral vaccines against tuberculosis and polio more than 3 decades ago, a number of live attenuated microorganisms have been used as antigen-delivery systems. Live vaccines are relatively easy and cheap to manufacture, because they do not require purification of antigens or formulation with adjuvants (82). Attenuated strains of microorganisms can be formed spontaneously or induced by heat, chemical, or UV mutagenesis. Another advantage of the attenuated live vaccines is that they can be administered by the natural route of infection. Recently, pathogenic microorganisms have been attenuated by genetic engineering, i.e., mutating specific genes or removing some toxic genes. Because much of the infection occurs through the mucosal surfaces, live attenuated vaccines are best suited for protection against pathogens that access the body through the mucosal surfaces. Live attenuated oral vaccines are expected to provide the most convenient and effective means of vaccinating against enteric disease (105). Orally administered attenuated *Salmonella* are known to interact with the MALT (82). Other examples of live attenuated microorganism vaccines are BCG (bacilli Calmette–Guérin), adenovirus, and poliovirus.

Some viruses and bacteria are inherently quite stable. For example, polio virus can be formulated as a frozen liquid. A live poliovirus vector expressing a foreign antigen generates both antibody and cytotoxic T-lymphocyte responses in mice (106). Most live bacteria and viruses, however, are usually stored as powders after freeze-drying or lyophilization. Preserving the live state through freeze-drying often requires the presence of a stabilizer, which is selected primarily through trial and error. The most widely used nonspecific stabilizers are sugars, amino acids, polyols, and neutral salts which are known to act as bound water substitutes for maintaining the conformational integrity of proteins. An example of lyophilized vaccine products is *S. typhi* bacteria lyophilized to a powder that is encapsulated into gelatin for oral administration (107).

One of the drawbacks of using live microorganisms is that attenuated pathogens may invoke the very disease they are designed to prevent if they are insufficiently attenuated. Even if they are sufficiently attenuated, they still may cause severe infections in immunocompromised individuals. In addition, they always have a potential to revert to full virulence if lesions causing attenuation are not fully characterized (82). If pathogens are over-attenuated, they fail to trigger an appropriate immune response. Thus, it is highly important to attain the right balance between minimal virulence and maximal immunogenicity. This balance can be achieved in a normal population but may not be the same in a population with even minor defects in immune competence (108). Another aspect to notice in using live vaccines is that the distribution of live vaccines requires a cold chain that may not be readily accessible in many developing countries, and this may offset advantages of using live-vectored vaccines (109).

Nonliving microparticulate delivery systems: Nonliving immunogens generally result in immune responses of lesser magnitude and of shorter duration than do those by living immunogens (82). Nonliving immunogens are usually made of microparticulate forms to protect antigens and to improve cellular uptake. Nonliving microparticulates that can be used as antigen-delivery systems include polymeric microparticles, liposomes, virus-like particles, neosomes, and cross-linked protein crystals. The definition of microparticles should be broad enough to include all other forms, such as protein aggregates. The size of microparticles used in the vaccine area is usually less than 50 μm (110). It is common, however, to call any particles less than a few hundred micrometers microparticles. For this reason, it is important to specify the average size of microparticles for particular applications, because the size of microparticles often affect the outcome.

Polymeric microparticles and liposomes have been used extensively as controlled-release dosage forms for many drugs including antigens. They have been quite useful in oral delivery of antigens because encapsulation in microparticles can protect antigens from acidic and enzymatic degradation in the GI tract, and thus serve as a stable vaccine vehicle with extended shelf life. Delivery of antigens by microparticulate-delivery systems has the potential benefits of reducing the number of inoculations, enhancing the immune response via both parenteral and oral vaccination routes, and reducing the total antigen dose required to achieve immune protection (111). Microparticulate vaccine-delivery systems show improved immune responses because of the protection of the loaded antigens from degradation and the slow release of the

antigens. For this reason, microparticulate-delivery systems are often considered adjuvants (66).

Polymer microparticles, a large number of polymers, such as poly(methyl methacrylate), poly(butyl cyanoacrylate), poly(lactide-co-glycolide), polyaracyl starch, dextran, albumin, and alginic acid, have been used for making microparticles for vaccine delivery. All the polymers that have been used for controlled drug delivery can be used for vaccine delivery (112). Preparation of microparticles from water-insoluble polymers [e.g., poly(methyl methacrylate), poly(butyl cyanoacrylate), and poly(lactide-co-glycolide)] requires use of organic solvents or high temperature, both of which may not be good for maintaining tertiary structures of antigens. Preparation from water-soluble polymers frequently requires cross-linking reaction to make the polymers remain insoluble. It is possible that cross-linking agents cross-link not only polymer chains but also antigen molecules. Absorption of water into hydrophilic polymers results in swelling of the network, i.e., formation of hydrogels, or aquagels. Preparation of microparticles from hydrophilic polymers is preferred because it does not require organic solvents or high temperature. Polymers that have been used in the immunization vary depending on the route of administration.

For parenteral vaccination, biodegradable polymeric microparticles made of poly(lactide-co-glycolide) are commonly used as vaccine carriers. Poly(lactide-co-glycolide) has been well-characterized and known to be highly biocompatible. The size of microparticles can be easily controlled, and microparticles of less than 100 μm in diameter can be easily administered by injection through standard-sized needles (22 gauge or smaller). Because of the slow degradation of the polymer, antigens are slowly released from the microparticles for long term in much the same way as do alum adjuvants, and this results in enhanced immune responses. Other polymers, such as chitosan, have been used for preparation of vaccine formulations. Because one of the important roles that microparticles play in immunization is the slow release of antigens, a number of approaches have been tried to achieve antigen release at desired rates. The surface of microparticles can be modified to alter the adsorption and desorption kinetics of antigens. Alternatively, the pore size can be varied to control the release of antigens from microparticles.

The size of microparticles is known to play a critical role in oral immunization. In addition to protecting antigens from acidic and enzymatic degradation in the GI tract, microparticulates are known to enhance uptake by M-cells in the Peyer's patches, and the effectiveness of the uptake depends on the size of microparticles. It is generally thought that microparticles smaller than 10 μm

are preferentially absorbed by M-cells, and the smaller the size, the better the absorption. One study using microparticles of different sizes showed that the efficiency of uptake of 100-nm particles by the intestinal tissue was 15- to 250-fold higher than that of larger size microparticles (113). In addition to the small size, microparticles with more hydrophobic surface property are absorbed better than those with more hydrophilic surface property. There are, however, no definite studies confirming or supporting these assumptions. Once microparticles are placed in the GI tract, adsorption of numerous proteins and polysaccharides present in the GI tract would alter the surface chemistry drastically, and it is difficult to correlate a particular surface chemistry of the native microparticles with the absorption ability.

Virus-like particles (VLPs) consist of one or more viral-coat proteins. They are very immunogenic molecules that allow for covalent coupling of the epitopes of interest (114). Recently, parvovirus-like particles have been engineered to express foreign polypeptides in certain positions, resulting in the production of large quantities of highly immunogenic peptides, and to induce strong antibody, helper T-cell, and cytotoxic T-lymphocyte responses (114). Parenteral administration of recombinant VLPs of papillomavirus induced VLP-specific humoral and cellular immune responses (115). Immunization of VLPs without adjuvant via mucosal route is also known to elicit specific antibody at mucosal surfaces and also systemic VLP epitope-specific T-cell responses (115).

Liposomes are vesicles composed of naturally occurring or synthetic phospholipids. The bilayer structure can be single- or multicompartiment. The size can also vary from smaller than 1 μm to larger than 10 μm . When negatively charged lipid molecules, which form liposomes, interact with divalent cations, a solid, multilayered, crystalline structure called cochleate is formed. Because liposomes and cochleates can protect antigens from the GI tract and deliver them to the Peyer's patches, they have been exploited as an effective delivery system for oral vaccination.

Liposomes, like other vaccine-delivery systems, can exert immunoadjuvant effects. The surface charge of liposomes is known to affect the immune responses. Positively charged liposomes containing soluble antigens were reported to function as a more potent inducer of antigen-specific, cytotoxic T-lymphocyte responses and delayed-type hypersensitivity responses than negatively charged and neutral liposomes containing the same concentrations of antigens (116). Studies showed that the positively charged liposomes delivered proteinaceous antigens efficiently into the cytoplasm of the macrophages/antigen-presenting cells where the antigens are

processed to be presented by class I MHC molecules to induce the cell-mediated immune response (116).

Liposomes containing highly immunogenic glycoproteins of the Sendai virus on their surface, which are called fusogenic liposomes, showed enhanced antigen-specific humoral immunity in mice. The levels of antiovalbumin antibody were markedly increased in serum from mice immunized with OVA encapsulated in fusogenic liposomes. It appears that the fusogenic liposomes function as an immunoadjuvant in inducing antigen-specific antibody production (117).

Virosomes are liposomes containing viral fusion proteins that allow efficient entering into cells fusion with endosome membranes. Viral fusion proteins become activated in the low pH environment in the endosome to release its contents into the cytosol (118). Hepatitis A and influenza vaccines constructed on virosomes elicited fewer local adverse reactions than did their classic counterparts and displayed enhanced immunogenicity. Virosome-formulated influenza vaccine has also been shown to be safe and immunogenic when administered by the intranasal route (119). Other studies have suggested that immunopotentiating reconstituted influenza virosomes can be a suitable delivery system for synthetic peptide vaccines. The virosomes have a great potential for the design of combined vaccines targeted against multiple antigens and multiple pathogens (120).

Micelles are aggregates of detergent molecules in aqueous solution. Detergents are water-soluble, surface-active agents composed of a hydrophilic head group and a hydrophobic or lipophilic tail group. They can also align at aqueous/nonaqueous interfaces, reducing surface tension, increasing miscibility, and stabilizing emulsions. Polymeric micelles made of block copolymers, such as poly(ethylene oxide)-poly(propylene oxide)-poly(ethylene oxide), have been used as a delivery system for hydrophobic drugs. They can also encapsulate antigens for vaccination.

Niosomes are nonionic surfactant vesicles. They have been used to develop a vaccine-delivery system by peroral and oral routes. Ovalbumin was encapsulated in various lyophilized niosome preparations consisting of sucrose esters, cholesterol, and dicetyl phosphate. Encapsulation of ovalbumin into niosomes consisting of 70% stearate sucrose ester and 30% palmitate sucrose ester (40% mono-, 60% di/triester) resulted in a significant increase in antibody titers in serum, saliva, and intestinal washings (121).

Cross-linked protein crystals have been used as antigens. The immunogenicity of cross-linked protein crystals of human serum albumin was 6- to 30-fold higher in antibody titer than that of the soluble protein over an almost 6-month study (122). It is likely that the

cross-linked protein crystals release antigen in a slow-release manner, and in this sense, the cross-linked protein crystals function as a depot. The cross-linked protein crystals present high stability, purity, biodegradability, and ease of manufacturing, all of which are highly attractive features for vaccine formulation (122). Because the cross-linked protein crystals are microparticulates, they can also be used for vaccination through various routes.

IMMUNOMODULATION

Immunomodulation refers to treatments that alter immune responsiveness in a nonantigen-specific manner. Enhancement of the immune response is desired in the treatment of chronic infectious diseases and neoplastic diseases, whereas suppression is needed in cases of inappropriate or exaggerated immune response, including allergies and autoimmune diseases. There are numerous treatments that affect the activity of the immune system. The effect of currently available immunosuppressive drugs is very broad, giving these drugs undesirable side effects. The aim of the research in this area is to design treatments that selectively enhance or suppress immune responses. Some of the newer treatment options are those that target costimulatory molecules, and the use of CpG DNA, and cytokines.

Costimulation

Activation of T-cells requires two signals. The first signal is provided by recognition of MHC/peptide complex by the T-cell receptor. This does not result in proliferation and differentiation of the T-cell unless the T-cell receives a second, costimulatory signal. Several costimulatory signals have been identified, but the major costimulatory signal appears to result from the binding of CD28 on T-cells to B7 molecules on antigen-presenting cells. There are at least two B7 molecules, B7-1 (CD80) and B7-2 (CD86). Activation of antigen-presenting cells results in increased expression of B7-2, followed by B7-1. A second T-cell ligand of the B7 molecules is cytotoxic T lymphocyte antigen-4 (CTLA-4 or CD152) that, other than its name implies, is rapidly expressed on both CD4⁺ and CD8⁺ T-cells after binding of the T-cell receptor to the MHC/peptide complex on antigen-presenting cells. However, in contrast to the positive signal provided by CD28, CTLA-4 downregulates T-cell responses (123). CTLA-4 has a higher affinity for the B7-molecules than does CD28 and may prevent the activation of T-cells when B7 expression by dendritic cells is low and terminate

the immune response when its expression is strongly increased. A soluble chimeric protein, CTLA4Ig, blocks the binding of both CD28 and CTLA-4 to the B7 molecules and, thus, may prevent T-cell activation. Administration of this protein to patients with psoriasis vulgaris, an immune-mediated skin disease, in a phase I clinical trial resulted in significant improvement in approximately 50% of the patients (124). Selective inhibition of CTLA-4 with specific antibodies may boost the immune system. The combination of surgery and anti-CTLA-4 antibody therapy was highly effective in the prevention of metastatic recurrence in a mouse prostatic carcinoma model (125).

Other CD28 and B7 homologs continue to be identified and appear to play a role in costimulation (126). These molecules may provide additional targets for immunomodulation and suggest that it may be possible to fine-tune the immune response through pharmacologic intervention.

CpG DNA

Bacterial DNA has a higher content of the CpG dinucleotide than does vertebrate DNA, and, in contrast to vertebrate DNA, the CpG is not preferentially methylated. The unmethylated CpG DNA sequences provides a strong stimulus for the immune system (127). CpG DNA stimulates the secretion of IL-12 by macrophages and dendritic cells and thus provides a potent stimulus for type 1 immune responses. It also directly stimulates B cells to proliferate and differentiate into immunoglobulin secreting cells. A cellular receptor for CpG DNA has not been identified. The DNA appears to enter the cell via endocytosis, and some of the DNA escapes the endosomes into the cytoplasm of the cell where it activates various signaling pathways.

Applications for oligonucleotides containing unmethylated CpG sequences (CpG-ODN) are being explored in various areas of immunotherapy. Administration of CpG-ODN to mice protected against subsequent challenge with the intracellular bacteria *Listeria monocytogenes* and the intracellular protozoa *Leishmania major*. In addition, the CpG-ODN cured established *L. major* infections. The strong type 1 immunostimulatory property of CpG-ODN makes this compound a good candidate for vaccine adjuvants. Indeed, coadministration of CpG-ODN with antigen markedly boosts the humoral and cell-mediated immune responses. Allergic diseases such as asthma and atopic dermatitis are caused by type 2 immune responses directed against otherwise innocuous antigens. Treatment with CpG-ODN cleared established disease in a mouse model of airway hyper-reactivity, suggesting a CpG-induced reversal to type 1 immune

responses. CpG DNA may also have a place in immunotherapy of cancer because of its ability to activate NK cells through the induction of IL-12. Administration of CpG-ODN in combination with monoclonal antibodies directed against tumor antigens greatly enhanced the survival of mice that had been inoculated with tumor cells.

Cytokines

Cytokines play a critical role in the regulation of the immune and inflammatory response, and they are potential targets for therapy. Important limitations, however, are the pleiotropy and redundancy in the cytokine system and the short half-life and short action range of most cytokines. In spite of these limitations, considerable effort is spent on developing reagents that either block or enhance the activity of a specific cytokine.

Two remarkable successes of cytokine therapy are the treatment of multiple sclerosis with interferon- β and the treatment of rheumatoid arthritis and inflammatory bowel disease with tumor necrosis factor- α inhibitors.

Interferon- β

Clinical trials have demonstrated that subcutaneous injections of recombinant or natural interferon- β reduces the rate of exacerbation of relapsing-remitting multiple sclerosis (128, 129). The mode of action of interferon- β has not been determined. Interferon- β reduces the production of tumor necrosis factor- α and increases the secretion of IL-10 in vitro. TNF- α is a proinflammatory cytokine that may contribute to demyelination in multiple sclerosis. IL-10 suppresses macrophage function and the production of TNF- α . In addition, interferon- β may reduce the entry of leukocytes into the central nervous system, a critical component in the inflammation that causes the lesions in multiple sclerosis.

Tumor necrosis factor- α inhibitors

Tumor necrosis factor- α (TNF- α) is a cytokine with multiple biological effects. It is produced as a transmembrane precursor molecule by various cells in the body. It is cleaved by the TNF- α -converting enzyme and forms trimeric aggregates that bind to either the TNF-receptor (TNFR) I or the TNFR II that are expressed on many different types of cells. The extracellular domains of the TNFR can be cleaved by enzymes and can inhibit TNF- α activity by preventing binding of TNF- α to cell-bound receptors. Recent studies have demonstrated that inhibition of TNF- α activity resulted in significant improvement of the clinical condition of many patients with rheumatoid arthritis and inflammatory bowel disease (130, 131). These studies clearly demonstrate an important role of TNF- α in

rheumatoid arthritis and inflammatory bowel disease, although the precise mechanisms remain to be determined. The inhibition of TNF- α activity is achieved by treatment with anti-TNF- α monoclonal antibodies or with soluble TNFR-fusion protein. To reduce the induction of antibodies against the mouse monoclonal antibodies, the monoclonal antibodies are chimeric (i.e., the constant portion is derived from human immunoglobulins and the TNF- α -specific variable portion is derived from mice) or humanized (all of the immunoglobulin is human except for the complementarity determining regions that fold into the TNF- α -binding region). The TNFR-fusion protein is constructed from the extracellular domain of TNFRII and the Fc portion of human immunoglobulins. This construct has a much longer half-life than does the naturally occurring soluble TNFR.

CHALLENGES IN FUTURE VACCINE FORMULATIONS

Recent advancements in microbial pathogenesis, immunology, genetic engineering, plant genetics, and expression vector technology have formed the foundation for a new generation of vaccines and other pharmaceutical products. New developments in the delivery system have provided us with novel ways to enhance the immunogenicity of subunit antigens or nucleic acids by their controlled release and reduced degradation.

For more convenient and more effective immunization, current vaccine-delivery technologies need to be improved. Currently, vaccination of many inactivated or subunit antigens requires booster doses because of the lack of inherent immunogenicity found in the natural organism. Thus, reducing the number of doses is one of the primary goals in vaccination. Theoretically, various controlled-release technologies can be used to release antigens over time in a sustained or pulsatile manner and to direct antigens to specific antigen-presenting cells for increased vaccine efficacy. In addition to controlled-release technology, the single-shot vaccination requires development of better adjuvants. The mechanism of action of such adjuvants should be known so that reproducible results can be obtained in a mass vaccination program. The requirements and problems of immunizing immunocompromised, immature, older, or pregnant hosts need to be addressed effectively. Further improvement in our understanding of how to modulate Th1 and Th2 responses effectively would certainly help us design better vaccines. Another means of improvement is to combine a number of vaccines into multivalent vaccines. This will improve the

immunization compliance in people living in developed or developing countries. Because the majority of pathogens enter their hosts via mucosal routes, the new-generation vaccines should have the advantage of providing effective protection at the mucosal sites. An ideal vaccine would be one that provides life-long protection with a single inoculation. The new-generation vaccine formulations should also have high stability, thus avoiding the problems commonly observed during storage.

REFERENCES

1. Peterson, L.J.; Benson, W.W.; Graeber, F.O. Vaccination Induced Poliomyelitis in Idaho. *J. A. Med. Assoc.* **1955**, *159*, 241–244.
2. Nathanson, N.; Langmuir, A.D. The Cutter Incident. Poliomyelitis Following Formaldehyde-Inactivated Poliovirus Vaccination in the United States During the Spring of 1955. I. Background. *Am. J. Hygiene* **1963**, *78*, 16–28.
3. Kapikian, A.Z.; Mitchell, R.H.; Chanock, R.M.; Shvedoff, R.A.; Stewart, C.E. An Epidemiologic Study of Altered Clinical Reactivity to Respiratory Syncytial (RS) Virus Infection in Children Previously Vaccinated with an Inactivated RS Virus Vaccine. *Am. J. Epidemiol.* **1969**, *89*, 405–421.
4. Kim, H.W.; Canchola, J.G.; Brandt, C.D.; Pyles, G.; Chanock, R.M.; Jensen, K.; Parrott, R.H. Respiratory Syncytial Virus Disease in Infants Despite Prior Administration of Antigenic Inactivated Vaccine. *Am. J. Epidemiol.* **1969**, *89*, 422–434.
5. Murphy, B.R.; Prince, G.A.; Walsh, E.E.; Kim, H.W.; Parrott, R.H.; Hemming, V.G.; Rodriguez, W.J.; Chanock, R.M. Dissociation Between Serum Neutralizing and Glycoprotein Antibody Responses of Infants and Children Who Received Inactivated Respiratory Syncytial Virus Vaccine. *J. Clin. Microbiol.* **1986**, *24*, 197–202.
6. Connors, M.; Kulkarni, A.B.; Firestone, C.Y.; Holmes, K.L.; Morse, H.C., III; Sotnikov, A.V.; Murphy, B.R. Pulmonary Histopathology Induced by Respiratory Syncytial Virus (RSV) Challenge of Formalin-Inactivated RSV-Immunized BALB/c Mice is Abrogated by Depletion of CD4⁺ T Cells. *J. Virol.* **1992**, *66*, 7444–7451.
7. Top, F.H., Jr.; Buescher, E.L.; Bancroft, W.H.; Russell, P.K. Immunization with Live Type 7 and Type 4 Adenovirus Vaccines. II Antibody Response and Protective Effect Against Acute Respiratory Disease Due to Adenovirus Type 7. *J. Infect. Dis.* **1971**, *124*, 155–160.
8. Sabin, A.B.; Boulger, L. History of Sabin Attenuated Poliovirus Oral Live Vaccine Strains. *J. Biol. Standardization* **1973**, *1*, 115–118.
9. Parkman, P.D.; Meyer, H.M., Jr.; Kirschstein, R.L.; Hopps, H.E. Attenuated Rubella Virus. I. Development and Laboratory Characterization. *New Engl. J. Med.* **1966**, *275*, 569–574.
10. Plotkin, S.A.; Farquhar, J.; Katz, M.; Ingalls, T.H. A New Attenuated Rubella Virus Grown in Human Fibroblasts: Evidence for Reduced Nasopharyngeal Excretion. *Am. J. Epidemiol.* **1967**, *86*, 468–477.

11. Bunyak, E.B.; Hilleman, M.R.; Weiber, R.E.; Stokes, J., Jr. Live Attenuated Rubella Virus Vaccines Prepared in Duck Embryo Cell Culture. I. Development and Clinical Testing. *J. Am. Med. Assoc.* **1968**, *204*, 195–200.
12. Morrow, C.D.; Novak, M.J.; Ansardi, D.C.; Porter, D.C.; Moldoveanu, Z. Recombinant Viruses as Vectors for Mucosal Immunity, Defence of Mucosal Surfaces: Pathogenesis, Immunity and Vaccines. *Current Topics in Microbiology and Immunology*; Kraehenbuhl, J.P., Neutra, M.R., Eds.; Springer-Verlag: Berlin, 1999; 236, 255–273.
13. Muster, T.; Ferko, B.; Klima, A.; Purtscher, M.; Trkola, A.; Schulz, P.; Grassauer, A.; Engelhardt, O.G.; Garcia-Sastre, A.; Palese, P. et al. Mucosal Model of Immunization Against Human Immunodeficiency Virus Type 1 with a Chimeric Influenza Virus. *J. Virol.* **1995**, *69*, 6678–6686.
14. Pastoret, P.P.; Brochier, B.; Languet, B.; Thomas, I.; Paquot, A.; Bauduin, B.; Kieny, M.P.; Lecocq, J.P.; De Bruyn, J.; Costy, F. et al. First Field Trial of Fox Vaccination Against Rabies Using a Vaccinia-Rabies Recombinant Virus. *Vet. Rec.* **1988**, *123*, 481–483.
15. Robbins, A.H.; Borden, M.D.; Windmiller, B.S.; Niezgoda, M.; Marcus, L.C.; O'Brien, S.M.; Kreindel, S.M.; McGuill, M.W.; DeMaria, A., Jr.; Rupprecht, C.E.; Rowell, S. Prevention of the Spread of Rabies to Wildlife by Oral Vaccination of Raccoons in Massachusetts. *J. Am. Vet. Med. Assoc.* **1998**, *213*, 1407–1412.
16. Yilma, T.; Hsu, D.; Jones, L.; Owens, S.; Grubman, M.; Mebus, C.; Yamanaka, M.; Dale, B. Protection of Cattle Against Rinderpest with Vaccinia Virus Recombinants Expressing the HA or F Gene. *Science* **1988**, *242*, 1058–1061.
17. Yilma, T. Vaccinia Virus Recombinant Vaccines for Rinderpest. *Dev. Biol. Standardization* **1995**, *84*, 201–208.
18. Graham, F.L.; Prevec, L. Adenovirus Expression Vectors and Recombinant Vaccines. In *Vaccines; New Approaches to Immunological Problems*; Ellis, R.W., Ed.; Butterworth-Heinemann: Boston, 1992; 363–390.
19. Imler, J.-L. Adenovirus Vectors as Recombinant Viral Vaccines. *Vaccine* **1995**, *13*, 1143–1151.
20. Mittal, S.K.; Papp, Z.; Tikoo, S.K.; Baca-Estrada, M.; Yoo, D.; Benko, M.; Babiuk, L.A. Induction of Systemic and Mucosal Immune Responses in Cotton Rats Immunized with Human Adenovirus Type 5 Recombinants Expressing the Full and Secretory Forms of Bovine Herpesvirus Type 1 Glycoprotein gD. *Virology* **1996**, *222*, 299–309.
21. Taylor, J.; Paoletti, E. Pox Viruses as Eukaryotic Cloning and Expression Vectors: Future Medical and Veterinary Vaccines. *Prog. Vet. Microbiol. Immunol.* **1988**, *4*, 197–217.
22. Jacobs, W.R.; Snapper, S.B.; Lugosi, L.; Bloom, L.; Bloom, B.R. Development of BCG as a Recombinant Vaccine Vehicle. *T-Cell Paradigms in Parasitic and Bacterial Infections. Current Topics in Microbiology and Immunology*; Kaufmann, S.H.E., Ed.; Springer-Verlag: Berlin, 1990; 155, 153–160.
23. Ryan, E.T.; Crean, T.I.; John, M.; Butters, J.R.; Clements, J.D.; Calderwood, S.B. Vivo Expression and Immunoadjuvancy of a Mutant of Heat-Labile Enterotoxin of *Escherichia coli* in Vaccine and Vector Strains Of *Vibrio cholerae*. *Infect. Immun.* **1999**, *67*, 1694–1701.
24. Cardenas, L.; Dasgupta, U.; Clements, J.D. Influence of Strain Viability and Antigen Dose on the Use of Attenuated Mutants of Salmonella as Vaccine Carriers. *Vaccine* **1994**, *12*, 833–840.
25. Cardenas, L.; Clements, J.D. Oral Immunization Using Live Attenuated Salmonella Spp. As Carriers of Foreign Antigens. *Clin. Microbiol. Rev.* **1992**, *5*, 328–342.
26. Killeen, K.; Spriggs, D.; Mekalanos, J. Bacterial Mucosal Vaccines: *Vibrio Cholerae* as a Live Attenuated Vaccine/Vector Paradigm, Defence of Mucosal Surfaces: Pathogenesis, Immunity and Vaccines. *Current Topics in Microbiology and Immunology*; Kraehenbuhl, J.P., Neutra, M.R., Eds.; Springer-Verlag: Berlin, 1999; 236, 237–254.
27. Nakayama, K.; Kelly, S.M.; Curtiss, R., III Construction of an ASD+ Expression-Cloning Vector: Stable Maintenance and High Level Expression of Cloned Genes in a Salmonella Vaccine Strain. *Biotechnology* **1998**, *6*, 693–697.
28. Gentschev, I.; Hess, J.; Goebel, W. Change in the Cellular Localization of Alkaline Phosphatase by Alteration of Its Carboxy-Terminal Sequence. *Mol. Gen. Genetics.* **1990**, *222*, 211–216.
29. Tzschaschel, B.D.; Klee, S.R.; de Lorenzo, V.; Timmis, K.N.; Guzman, C.A. Towards a Vaccine Candidate Against Shigella Dysenteriae 1: Expression of the Shiga Toxin B-subunit in an Attenuated Shigella Flexneri AroD Carrier Strain. *Microbial Pathogenesis* **1996**, *21*, 277–288.
30. Kit, S.; Kit, M. Genetically Engineered Herpesvirus Vaccines. Accomplishments in Pigs and Prospects in Humans. *Prog. Med. Virol.* **1991**, *38*, 128–66.
31. Kit, S. Genetically Engineered Vaccines for Control of Aujeszky's Disease (Pseudorabies). *Vaccine* **1990**, *8*, 420–424.
32. Chatfield, S.N.; Roberts, M.; Dougan, G.; Hormaeche, C.; Khan, C.M.A. The Development of Oral Vaccines Against Parasitic Diseases Utilizing Live Attenuated Salmonella. *Parasitology.* **1995**, *110*, S17–S24.
33. Mettenleiter, T.C. New Developments in the Construction of Safer and More Versatile Pseudorabies Virus Vaccines. *Dev. Biol. Standardization* **1995**, *84*, 83–87.
34. Kimman, T.G.; Gielkens, A.L.J.; Glazenburg, K.; Jacobs, L.; de Jong, M.C.M.; Mulder, W.A.M.; Peeters, B.P.H. Characterization of Live Pseudorabies Virus Vaccines. *Dev. Biol. Standardization* **1995**, *84*, 89–96.
35. Briand, J.P.; Benkirane, N.; Guichard, G.; Newman, J.F.; Van Regenmortel, M.H.; Brown, F.; Muller, S. A Retro-Inverso Peptide Corresponding to the GH Loop of Foot-and-Mouth Disease Virus Elicits High Levels of Long-Lasting Protective Neutralizing Antibodies. *Proceedings of the National Academy Science USA* **1997**, *94*, 12545–12550.
36. Brown, F. Foot-and-Mouth Disease and Beyond: Vaccine Design, Past, Present and Future. *Arch. Virol. (Suppl)* **1999**, *15*, 179–188.
37. Marciani, D.J.; Kensil, C.R.; Beltz, G.A.; Hung, C.H.; Cronier, J.; Aubert, A. Genetically-Engineered Subunit Vaccine Against Feline Leukaemia Virus: Protective Immune Response in Cats. *Vaccine* **1991**, *9*, 89–96.

38. Valenzuela, P.; Medina, A.; Rutter, W.J.; Ammerer, G.; Hall, B.D. Synthesis and Assembly of Hepatitis B Virus Surface Antigen Particles in Yeast. *Nature* **1982**, *298*, 347–350.
39. Brideau, R.J.; Walters, R.R.; Stier, M.A.; Wathen, M.W. Protection of Cotton Rats Against Human Respiratory Syncytial Virus by Vaccination with a Novel Chimeric FG Glycoprotein. *J. Gen. Virol.* **1989**, *70*, 2637–2644.
40. Brideau, R.J.; Oien, N.L.; Lehman, D.J.; Homa, F.L.; Wathen, M.W. Protection of Cotton Rats Against Human Parainfluenza Virus Type 3 by Vaccination with a Chimeric FHN Subunit Glycoprotein. *J. Gen. Virol.* **1993**, *74*, 471–477.
41. Kowalski, J.; Gilbert, S.A.; van Drunen-Littel-van den Hurk, S.; van den Hurk, J.; Babiuk, L.A.; Zamb, T.J. Heat-Shock Promoter-Driven Synthesis of Secreted Bovine Herpesvirus Glycoproteins in Transfected Cells. *Vaccine* **1993**, *11*, 1100–1107.
42. Ma, J.K.-C.; Vine, N.D. Plant Expression Systems for the Production of Vaccine, Defence of Mucosal Surfaces: Pathogenesis, Immunity and Vaccines. *Current Topics in Microbiology and Immunology*; Kraehenbuhl, J.P., Neutra, M.R., Eds.; Springer-Verlag: Berlin, 1999; 236, 275–292.
43. Ma, J.K.-C.; Hiatt, A.; Hein, M.; Vine, N.D.; Wang, F.; Stabila, P.; van Dolleweerd, C.; Mostov, K.; Lehner, T. Generation and Assembly of Secretory Antibodies in Plants. *Science* **1995**, *268*, 716–719.
44. Haq, T.A.; Mason, H.S.; Clements, J.D.; Arntzen, C.J. Oral Immunization with a Recombinant Bacterial Antigen Produced in Transgenic Plants. *Science* **1995**, *268*, 714–716.
45. Carrillo, C.; Wigdorovitz, A.; Oliveros, J.C.; Zamorano, P.I.; Sadir, A.M.; Gomez, N.; Salinas, J.; Escribano, J.M.; Borca, M.V. Protective Immune Responses to Foot-Mouth-Disease Virus with VP1 Expressed in Transgenic Plants. *J. Virol.* **1998**, *72*, 1688–1690.
46. Dalsgaard, K.; Utenthal, A.; Jones, T.D.; Xu, F.; Merryweather, A.; Hamilton, W.D.O.; Langeveld, J.P.M.; Boshuizen, R.S.; Kamstrup, S.; Lomonosoff, G.P.; Porta, C.; Vela, C.; Casal, J.I.; Meloen, R.H.; Rodgers, P.B. Plant-Derived Vaccine Protects Target Animals Against a Viral Disease. *Nature Biotechnol.* **1997**, *15*, 248–252.
47. Thanavala, Y.; Yang, Y.-F.; Lyons, P.; Mason, H.S.; Arntzen, C. Immunogenicity of Transgenic Plant-Derived Hepatitis B Surface Antigen. *Proceedings of the National Academy Science USA* **1995**, *92*, 2258–3361.
48. Arntzen, C.J.; Mason, H.S.; Shi, J.; Haq, T.A.; Estes, M.K.; Clements, J.D. Production of Candidate Oral Vaccines in Edible Tissues of Transgenic Plants. *Vaccines '94: Modern Approaches to New Vaccines Including Prevention of AIDS*; Brown, F., Chanock, R.M., Ginsberg, H.S., Lerner, R.A., Eds.; Cold Spring Harbor Laboratory Press Cold Spring Harbor: New York, 1994; 339–344.
49. Zanetti, M.; Secarz, E.; Salk, J. The Immunology of New Generation Vaccines. *Immunol. Today* **1987**, *8*, 18–22.
50. Oosterlaken, T.A.; Harmsen, M.; Jhagjhoor-Singh, S.S.; Ekstijn, G.L.; Kraaijeveld, C.A.; Snippe, H.A. Protective Monoclonal Anti-Idiotypic Vaccine to Lethal Semliki Forest Virus Infection in BALB/c Mice. *J. Virol.* **1991**, *65*, 98–102.
51. Wolff, J.A.; Malone, R.W.; Williams, P.; Chong, W.; Acsadi, G.; Jani, A.; Felgner, P.L. Direct Gene Transfer into Mouse Muscle In Vivo. *Science* **1990**, *24*, 1465–1468.
52. Yang, N.S.; Burkholder, J.; Roberts, B.; Martinell, B.; McCabe, D. In Vivo and In Vitro Gene Transfer to Mammalian Somatic Cells by Particle Bombardment. *Proceedings of the National Academy Science USA* **1990**, *87*, 9568–9572.
53. Williams, R.S.; Johnston, S.A.; Riedy, M.; DeVit, M.J.; McElligott, S.G.; Sanford, J.C. Introduction of Foreign Genes into Tissues of Living Mice by DNA-Coated Microprojectiles. *Proceedings of the National Academy Science USA* **1991**, *88*, 2726–2730.
54. Felgner, P.L.; Ringold, G.M. Cationic Liposome-Mediated Transfection. *Nature* **1989**, *337*, 387–388.
55. Barthel, F.; Remy, J.S.; Loeffler, J.P.; Behr, J.P. Gene Transfer Optimization with Lipospermine-Coated DNA. *DNA Cell. Biol.* **1993**, *12*, 553–560.
56. Mathiowitz, E.; Jacob, J.S.; Jong, Y.S.; Carino, G.P.; Chickering, D.E.; Chaturvedi, P.; Santos, C.A.; Vijayaraghavan, K.; Montgomery, S.; Bassett, M.; Morrell, C. Biologically Erodable Microspheres as Potential Oral Drug Delivery Systems. *Nature* **1997**, *386*, 410–414.
57. Aggarwal, N.; HogenEsch, H.; Guo, P.; North, A.; Suckow, M.; Mittal, S.K. Biodegradable Alginate Microspheres as a Delivery System for Naked DNA. *Can. J. Vet. Res.* **1999**, *63*, 148–152.
58. Mittal, S.K.; Aggarwal, N.; Sailaja, G.; van Olphen, A.; HogenEsch, H.; North, A.; Hays, J.; Moffatt, S. Immunization with DNA, Adenovirus or Both in Biodegradable Alginate Microspheres: Effect of Route of Inoculation on Immune Response. *Vaccine* **in press**.
59. Ulmer, J.B.; Donnelly, J.J.; Parker, S.E.; Rhodes, G.H.; Felgner, P.L.; Dwarki, V.J.; Gromkowski, S.H.; Deck, R.R.; DeWitt, C.M.; Friedman, A. Heterologous Protection Against Influenza by Injection of DNA Encoding a Viral Protein. *Science* **1993**, *259*, 1745–1749.
60. Donnelly, J.J.; Ulmer, J.B.; Liu, MA. DNA Vaccines. *Life Sci.* **1993**, *60*, 163–172.
61. Lowrie, D.B.; Silva, C.L.; Colston, M.J.; Ragno, S.; Tascon, R.E. Protection Against Tuberculosis by a Plasmid DNA Vaccine. *Vaccine* **1997**, *15*, 834–838.
62. Lowrie, D.B.; Tascon, R.E.; Bonato, V.L.; Lima, V.M.; Faccioli, L.H.; Stavropoulos, E.; Colston, M.J.; Hewinson, R.G.; Moelling, K.; Silva, C.L. Therapy of Tuberculosis in Mice by DNA Vaccination. *Nature* **1999**, *400*, 269–271.
63. Koprowski, H.; Weiner, D.B. *DNA Vaccination/Genetic Vaccination. Current Topics in Microbiology and Immunology*; Springer-Verlag: Berlin, 1998; 226.
64. Powell, M.F.; Newman, M.J. *Vaccine Design: The Subunit and Adjuvant Approach*; Plenum Press: San Diego, 1995.
65. Stewart-Tull, D.E.S. *The Theory and Practical Application of Adjuvants*; John Wiley & Sons: Chichester, UK, 1995.
66. O'Hagan, D.T. *Vaccine Adjuvants. Preparation Methods and Research Protocols*; Humana Press: Totowa, NJ, 2000; 342.
67. Singh, M.; O'Hagan, D. Advances in Vaccine Adjuvants. *Nature Biotechnol.* **1999**, *17*, 1075–1081.

68. Raychaudhuri, S.; Rock, K.L. Fully Mobilizing Host Defenses: Building Better Vaccines. *Nature Biotechnol.* **1998**, *16*, 1025–1031.
69. Cleland, J.L. Design and Production of Single-Immunization Vaccines Using Polylactide Polyglycolide Microsphere Systems. *Vaccine Design: The Subunit and Adjuvant Approach*; Powell, M.F., Newman, M.J., Eds.; Plenum Press: New York, 1995; 439–462.
70. Rundell, A.; DeCarlo, R.; HogenEsch, H.; Doerschuk, P. The Humoral Immune Response to Haemophilus Influenzae Type B: A Mathematical Model Based on T-Zone and Germinal Center B-Cell Dynamics. *J. Theor. Biol.* **1998**, *194*, 341–381.
71. Hem, S.L.; White, J.L. Structure and Properties of Aluminum-Containing Adjuvants. *Vaccine Design. The Subunit and Adjuvant Approach*; Powell, M.F., Newman, M.J., Eds.; Plenum Press: New York, 1995; 249–276.
72. Brewer, J.M.; Conacher, M.; Hunter, C.A.; Mohrs, M.; Brombacher, F.; Alexander, J. Aluminium Hydroxide Adjuvant Initiates Strong Antigen-Specific Th2 Responses in the Absence of IL-4- or IL-13-Mediated Signaling. *J. Immunol.* **1999**, *163*, 6448–6454.
73. Jankovic, D.; Caspar, P.; Zweig, M.; Garcia-Moll, M.; Showalter, S.D.; Vogel, F.R.; Sher, A. Adsorption to Aluminum Hydroxide Promotes the Activity of IL-12 as an Adjuvant for Antibody as well as Type 1 Cytokine Responses to HIV-1 GP120. *J. Immunol.* **1997**, *159*, 2409–2417.
74. Barr, I.G.; Sjolander, A.; Cox, J.C. ISCOMs and Other Saponin Based Adjuvants. *Adv. Drug Del. Rev.* **1998**, *32*, 247–271.
75. Brokstad, K.A.; Cox, R.J.; Olofsson, J.; Jonsson, R.; Haaheim, L.R. Parenteral Influenza Vaccination Induces a Rapid Systemic and Local Immune Response. *J. Infect. Dis.* **1995**, *171*, 198–203.
76. Shroff, K.E.; Marcucci-Borges, L.A.; de Bruin, S.J.; Winter, L.A.; Tiberio, L.; Pachuk, C.; Snyder, L.A.; Satishchandran, C.; Ciccarelli, R.B.; Higgins, T.J. Induction of HSV-gD2 Specific CD4(+) Cells in Peyer's Patches and Mucosal Antibody Responses in Mice Following DNA Immunization by Both Parenteral and Mucosal Administration. *Vaccine* **1999**, *18*, 222–230.
77. Fontanilla, B.C.; Silvano, F.; Cumming, R. Oral Vaccination Against Newcastle Disease of Village Chickens in the Philippines. *Prev. Vet. Med.* **1994**, *19*, 39–44.
78. Lundholm, P.; Asakura, Y.; Hinkula, J.; Lucht, E.; Wahren, B. Induction of Mucosal IgA By a Novel Jet Delivery Technique for HIV-1 DNA. *Vaccine* **1999**, *17*, 2036–2042.
79. McCluskie, M.J.; Brazolot Millan, C.L.; Gramzinski, R.A.; Robinson, H.L.; Santoro, J.C.; Fuller, J.T.; Widera, G.; Haynes, J.R.; Purcell, R.H.; Davis, H.L. Route and Method of Delivery of DNA Vaccine Influence Immune Responses in Mice and Non-Human Primates. *Mol. Med.* **1999**, *5*, 287–300.
80. Langue, J.; Ethevenaux, C.; Champsaur, A.; Fritzell, B.; Begue, P.; Saliou, P. Safety and Immunogenicity of Haemophilus Influenzae Type B-Tetanus Toxoid Conjugate, Presented in a Dual-Chamber Syringe with Diphtheria-Tetanus-Pertussis and Inactivated Poliomyelitis Combination Vaccine. *Eur. J. Pediatr.* **1999**, *158*, 717–722.
81. Kanra, G.; Yurdakok, K.; Ceyhan, M.; Ozmert, E.; Turkay, F.; Pehlivan, T. Immunogenicity and Safety of Haemophilus Influenzae Type B Capsular Polysaccharide Tetanus Conjugate Vaccine (PRP-T) Presented in a Dual-Chamber Syringe with DTP. *Acta Paediatr. Jpn.* **1997**, *39*, 676–680.
82. Roberts, M.; Chatfield, S.N.; Dougan, G. *Salmonella as Carriers of Heterologous Antigens, In Novel Delivery Systems for Oral Vaccines*; O'Hagan, D.T., Ed.; CRC Press Inc.: Ann Arbor MI, 1994; 27–58.
83. McGhee, J.R.; Mestecky, J.; Dertzbaugh, M.T.; Eldridge, J.H.; Hirasawa, M.; Kiyono, H. The Mucosal Immune System: From Fundamental Concepts to Vaccine Development. *Vaccine* **1992**, *10*, 75–88.
84. Hornquist, E.; Lycke, N.; Czerkinsky, C.; Holmgren, J. Cholera Toxin and Cholera B Subunit as Oral-Mucosal Adjuvant and Antigen Carrier Systems. *Novel Delivery Systems for Oral Vaccines*; O'Hagan, D.T. Ed.; CRC Press, Inc.: Ann Arbor MI, 1994; 157–173.
85. Service, R.F. Triggering the First Line of Defense. *Science* **1994**, *265*, 1552–1554.
86. Heritage, P.L.; Underdown, B.J.; Brook, M.A.; McDermott, M.R. Oral Administration of Polymer-Grafted Starch Microparticles Activates Gut-Associated Lymphocytes and Primes Mice for a Subsequent Systemic Antigen Challenge. *Vaccine* **1998**, *16*, 2010–2017.
87. Dale, J.W.; Dellagostin, O.A.; Norman, E.; Barrett, A.D.T.; McFadden, J. Multivalent BCG Vaccines. *Novel Delivery Systems for Oral Vaccines*; O'Hagan, D.T. Ed.; CRC Press, Inc.: Ann Arbor MI, 1994; 87–109.
88. Dunkley, M.L.; Harris, S.J.; McCoy, R.J.; Musicka, M.J.; Eysers, F.M.; Beagley, L.G.; Lumley, P.J.; Beagley, K.W.; Clancy, R.L. Protection Against *Helicobacter pylori* Infection by Intestinal Immunisation with a 50/52-kDa Subunit Protein. *FEMS Immunol. Med. Microbiol.* **1999**, *24*, 221–225.
89. Cortes-Thoulaz, I.; Porta, N.; Glauser, M.; Saraga, E.; Vaney, A.C.; Haas, R.; Kraehenbuhl, J.P.; Blum, A.L.; Michetti, P. Oral Immunization with *Helicobacter pylori* Urease B Subunit as a Treatment Against *Helicobacter* Infection in Mice. *Gastroenterology* **1995**, *109*, 115–121.
90. Husby, S.; Jensenius, J.C.; Svehag, S.-E. Passage of Undegraded Dietary Antigen into the Blood of Healthy Adults. Further Characterization of the Kinetics of the Uptake and the Size Distribution of the Antigen. *Scand. J. Immunol.* **1986**, *24*, 447–455.
91. Eaton, K.A.; Krakowka, S. Chronic Active Gastritis Due to *Helicobacter pylori* in Immunized Gnotobiotic Piglets. *Gastroenterology* **1992**, *103*, 1580–1586.
92. Ball, J.M.; Graham, D.Y.; Opekun, A.R.; Gilger, M.A.; Guerrero, R.A.; Estes, M.K. Recombinant Norwalk Virus-Like Particles Given Orally to Volunteers: Phase I Study (see Comments). *Gastroenterology* **1999**, *117*, 40–48.
93. Foster, N.; Clark, M.A.; Jepson, M.A.; Hirst, B.H. Ulex Europaeus 1 Lectin Targets Microspheres to Mouse Peyer's Patch M-Cells In Vivo. *Vaccine* **1998**, *16*, 536–541.
94. Eyles, J.E.; Williamson, E.D.; Alpar, H.O. Immunological Responses to Nasal Delivery of Free and Encapsulated Tetanus Toxoid: Studies on the Effect of Vehicle Volume. *Int. J. Pharm.* **1999**, *189*, 75–79.

95. Rudin, A.; Riise, G.C.; Holmgren, J. Antibody Responses in the Lower Respiratory Tract and Male Urogenital Tract in Humans After Nasal and Oral Vaccination with Cholera Toxin B Subunit. *Infect. Immun.* **1999**, *67*, 2884–2890.
96. Shi, Z.; Curiel, D.T.; Tang, D.C. DNA-based Noninvasive Vaccination Onto the Skin. *Vaccine* **1999**, *17*, 2136–2141.
97. Misra, A.; Ganga, S.; Upadhyay, P. Needle-Free, Nonadjuvanted Skin Immunization by Electroporation-Enhanced Transdermal Delivery of Diphtheria Toxoid and a Candidate Peptide Vaccine Against Hepatitis B Virus. *Vaccine* **1999**, *18*, 517–523.
98. Schariton-Kersten, T.M.; Glenn, G.; Vassell, R.; Yu, J.; Walwender, D.; Alving, C.R. Principles of Transcutaneous Immunization Using Cholera Toxin as an Adjuvant. *Vaccine, Suppl 17* **1999**, *2*, s37–43.
99. Glenn, G.M.; Schariton-Kersten, T.; Vassell, R.; Matyas, G.R.; Alving, C.R. Transcutaneous Immunization with Bacterial ADP-Ribosylating Exotoxins as Antigens and Adjuvants. *Infect. Immun.* **1999**, *67*, 1100–1106.
100. Glenn, G.M.; Schariton-Kersten, T.; Vassell, R.; Mallett, C.P.; Hale, T.L.; Alving, C.R. Transcutaneous Immunization with Cholera Toxin Protects Mice Against Lethal Mucosal Toxin Challenge. *J. Immunol.* **1998**, *161*, 3211–3214.
101. Brown, A.R.; George, D.W.; Matteson, D.K. Vaccinator Device for Delivering Propellant-Driven Aerosols of *Streptococcus suis* Bacterin into the Respiratory Tracts of Swine. *Vaccine* **1997**, *15*, 1165–1173.
102. LiCalsi, C.; Christensen, T.; Bennett, J.V.; Phillips, E.; Witham, C. Dry Powder Inhalation as a Potential Delivery Method for Vaccines. *Vaccine* **1999**, *17*, 1796–1803.
103. McCluskie, M.J.; Chu, Y.; Xia, J.L.; Jessee, J.; Gebyehu, G.; Davis, H.L. Direct Gene Transfer to the Respiratory Tract of Mice with Pure Plasmid and Lipid-Formulated DNA. *Antisense Nucleic Acid Drug Dev.* **1998**, *8*, 401–414.
104. Weng, C.N.; Tzan, Y.L.; Liu, S.D.; Lin, S.Y.; Lee, C.J. Protective Effects of an Oral Microencapsulated *Mycoplasma hyopneumoniae* Vaccine Against Experimental Infection in Pigs. *Res. Vet. Sci.* **1992**, *53*, 42–46.
105. Forrest, B.D. Clinical Evaluation of Attenuated *Salmonella typhi* Vaccines in Human Subjects. *Novel Delivery Systems for Oral Vaccines*; O'Hagan, D.T., Ed.; CRC Press, Inc.: Ann Arbor, MI, 1994; 59–85.
106. Crotty, S.; Lohman, B.L.; Lu, F.X.; Tang, S.; Miller, C.J.; Andino, R. Mucosal Immunization of Cynomolgus Macaques with Two Serotypes of Live Poliovirus Vectors Expressing Simian Immunodeficiency Virus Antigens: Stimulation of Humoral, Mucosal, and Cellular Immunity. *J. Virol.* **1999**, *73*, 9485–9495.
107. Aunins, J.G.; Lee, A.L.; Volkins, D.B. Vaccine Production. *The Biomedical Engineering Handbook*; Bronzino, J.D., Ed.; CRC Press: Boca Raton, FL, 1995; 1502–1517.
108. Bussiere, J.L.; McCormick, G.C.; Green, J.D. Preclinical Safety Assessment Considerations in Vaccine Development. *Vaccine Design: The Subunit and Adjuvant Approach*; Powell, M.F., Newman, M.J., Burdman, J.R., Eds.; Plenum Press: New York, 1995; 61–79.
109. Lawrence, D.N.; Goldenthal, K.L.; Boslego, J.W.; Chandler, D.K.F.; La Montagne, J.R. Public Health Implications of Emerging Vaccine Technologies. *Vaccine Design: The Subunit and Adjuvant Approach*; Powell, M.F., Newman, M.J., Burdman, J.R., Eds.; Plenum Press: New York, 1995; 43–60.
110. O'Hagan, D.T. Microparticles as Oral Vaccines. In *Novel Delivery Systems for Oral Vaccines*; CRC Press, Inc.: Ann Arbor, MI, 1994; 175–205.
111. Morris, W.; Steinhoff, M.C.; Russell, P.K. Potential of Polymer Microencapsulation Technology for Vaccine Innovation. *Vaccine* **1994**, *12*, 5–11.
112. Kuntz, R.M.; Saltzman, W.M. Polymeric Controlled Delivery for Immunization. *Trends Biotechnol.* **1997**, *15*, 364–369.
113. Desai, M.P.; Labhasetwar, V.; Amidon, G.L.; Levy, R.J. Gastrointestinal Uptake of Biodegradable Microparticles: Effect of Particle Size. *Pharm. Res.* **1996**, *13*, 1838–1845.
114. Casal, J.I.; Rueda, P.; Hurtado, A. Parvovirus-Like Particles as Vaccine Vectors. *Methods* **1999**, *19*, 174–186.
115. Liu, X.S.; Abdul-Jabbar, I.; Qi, Y.M.; Frazer, I.H.; Zhou, J. Mucosal Immunisation with Papillomavirus Virus-Like Particles Elicits Systemic and Mucosal Immunity in Mice. *Virology* **1998**, *252*, 39–45.
116. Nakanishi, T.; Kunisawa, J.; Hayashi, A.; Tsutsumi, Y.; Kubo, K.; Nakagawa, S.; Nakanishi, M.; Tanaka, K.; Mayumi, T. Positively Charged Liposome Functions as an Efficient Immunoadjuvant in Inducing Cell-Mediated Immune Response to Soluble Proteins. *J. Controlled Rel.* **1999**, *61*, 233–240.
117. Hayashi, A.; Nakanishi, T.; Kunisawa, J.; Kondoh, M.; Imazu, S.; Tsutsumi, Y.; Tanaka, K.; Fujiwara, H.; Hamaoka, T.; Mayumi, T. A Novel Vaccine Delivery System Using Immunopotentiating Fusogenic Liposomes. *Biochem. Biophys. Res. Commun.* **1999**, *261*, 824–828.
118. Lasic, D.D. *Liposomes in Gene Delivery*; CRC Press: Boca Raton, FL, 1997; 67–112.
119. Cryz, S.J. BERNAL: A Century of Immunobiological Innovation. *Vaccine* **1999**, *17*, S1–S5.
120. Poltl-Frank, F.; Zurbriggen, R.; Helg, A.; Stuart, F.; Robinson, J.; Gluck, R.; Pluschke, G. Use of Reconstituted Influenza Virus Virosomes as an Immunopotentiating Delivery System for a Peptide-Based Vaccine. *Clin. Exp. Immunol.* **1999**, *117*, 496–503.
121. Rentel, C.O.; Bouwstra, J.A.; Naisbett, B.; Junginger, H.E. Niosomes as a Novel Peroral Vaccine Delivery System. *Int. J. Pharm.* **1999**, *186*, 161–167.
122. St Clair, N.; Shenoy, B.; Jacob, L.D.; Margolin, A.L. Cross-Linked Protein Crystals for Vaccine Delivery. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9469–9474.
123. Thompson, C.B.; Allison, J.P. The Emerging Role of CTLA-4 as an Immune Attenuator. *Immunity* **1997**, *7*, 445–450.
124. Abrams, J.R.; Lebowitz, M.G.; Guzzo, C.A.; Jegasothy, B.V.; Goldfarb, M.T.; Goffe, B.S.; Menter, A.; Lowe, N.J.; Krueger, G.; Brown, M.J.; Weiner, R.S.; Birkhofer, M.J.; Warner, G.L.; Berry, K.K.; Linsley, P.S.; Krueger, J.G.; Ochs, H.D.; Kelley, S.L.; Kang, S.W. CTLA4Ig-Mediated Blockade of T-Cell Costimulation in Patients with Psoriasis Vulgaris. *J. Clin. Invest.* **1999**, *103*, 1243–1252.
125. Kwon, E.D.; Foster, B.A.; Hurwitz, A.A.; Madias, C.; Allison, J.P.; Greenberg, N.M.; Burg, M.B. Elimination of Residual Metastatic Prostate Cancer After Surgery and

- Adjunctive Cytotoxic T Lymphocyte-Associated Antigen 4 (CTLA-4) Blockade Immunotherapy. *Proceedings of the National Academy Science USA* **1999**, *96*, 15074–15079.
126. Abbas, A.K.; Sharpe, A.H. T-cell Stimulation: An Abundance of B7s. *Nature Med.* **1999**, *5*, 1345–1346.
127. Krieg, A.M.; Yi, A.-K.; Hartmann, G. Mechanisms and Therapeutic Applications of Immune Stimulatory CpG DNA. *Pharmacol. Therapeutics* **1999**, *84*, 113–120.
128. Hall, G.L.; Compston, A.; Scolding, N.J. Beta-Interferon and Multiple Sclerosis. *Trends Neurosci.* **1997**, *20*, 63–67.
129. Yong, V.W.; Chabot, S.; Williams, G. Interferon Beta in the Treatment of Multiple Sclerosis. *Mechanisms of Action. Neurology* **1998**, *51*, 682–689.
130. Moreland, L.W. Inhibitors of Tumor Necrosis Factor for Rheumatoid Arthritis. *J. Rheumatol.* **1999**, *57*, 7–15, Suppl. 26.
131. Sandborn, W.J.; Hanauer, S.B. Antitumor Necrosis Factor Therapy for Inflammatory Bowel Disease: A Review of Agents, Pharmacology, Clinical Results, and Safety. *Inflamm. Bowel Dis.* **1999**, *5*, 119–133.

Validation of Analytical Procedures

Joachim Ermer

Aventis Pharma, Frankfurt am Main, Germany

John S. Landy

Aventis Pharma, Bridgewater, New Jersey, U.S.A.

INTRODUCTION

To the pharmaceutical world, the meaning of analytical methods validation is the process to confirm that a method does what it purports to do, that is, to document through laboratory studies that the measurement procedure can reliably assess the identity, strength, and/or quality of a bulk drug substance, excipient, or finished pharmaceutical product. To provide consistent, worldwide regulatory expectations, previously unavailable, the International Conference on Harmonization [ICH,^{[1]a}] has defined the methods validation process for the release and stability testing of all new products. This chapter interprets these ICH regulatory definitions and requirements,^[2,3] as well as provides direction toward rational and efficient validation.

Regulatory methods can be compendial or noncompendial. Wherever possible, methodologies are to be employed which are documented, generally recognized as official pharmacopoeia or compendia. Compendial methods are considered valid; however, suitability must be verified under actual conditions of use. Noncompendial methods require validation, and must be selected if a compendial method does not exist. A noncompendial method can be chosen over an existing compendial method, if it can be demonstrated to be superior to the compendial test.

Before a product dossier has been submitted to an agency for regulatory market approval, analytical laboratories have utilized validated methods to support toxicological, clinical, stability, development, scale-up, optimization, process, and cleaning validation studies. Unreliable data for any of these studies have the potential to completely undermine the speed and success of approval. A method's "life cycle"^[9] parallels the drug development process. Starting with early (preclinical) development projects, the related methods for drug substance and finished drug product require only some rudimentary validation to provide sufficient confidence in

the results, eventually leading to a complete methods validation package for the final stages of product development and commercialization (see "Validation During Drug Development and Manufacturing"). The methods life cycle concludes with methods transfers, monitoring of routine quality control (QC) usage, and revalidation. We define revalidation as repeating those parts of validation that are affected by a modification, for example, specificity, if the column has changed. Repeating the whole validation periodically is superfluous; instead, continuous monitoring of the performance of the analytical procedure should be performed (see "Maintenance of the Validation Status"). Many tests may be specified in the early development of a product or process that will not be ultimately selected for routine release testing. Clearly not all products reach the approved and marketing stage, due to toxicology, efficacy, or even business conditions. Multiple other changes can occur along the way toward approval such as active pharmaceutical ingredient (API, drug substance), synthetic route changes, drug product formulation, and process changes, as well as newly identified degradation pathways. All of these affect the method applied. Therefore, methods (development) validation requires efficient planning of resources to match the accuracy and precision requirements needed to assess product quality.

Methods Selection and Applicability for Routine Use

The target laboratory [development (R&D) or QC] where the methods will be utilized and the stage of development are just as important as the analyte/sample to be measured when considering the selection of a method as these will affect the accuracy and precision requirements, as well as laboratory economic, environmental, and ergonomic factors. Methods should be selected that are adequate for testing the attribute to be measured. They must be sufficiently selective, accurate, precise, and robust to demonstrate conformance to proposed specifications.

^a<http://www.ifpma.org/ich5q.html> (accessed March 2001).

It should be noted that a precise method is an extremely important attribute in QC. A precise method permits a fluctuation in the manufacturing process to be detected before it can cause an out-of-specification (OOS) event. With a sufficiently precise method, subsequent tightening of the specification should not necessitate modification or revalidation of the test method.

Additionally, these procedures should be devised with the explicit objective of transferring the methods to a qualified laboratory. The development laboratory should specify test techniques, parameters, and other details in the final dossier submission that have the highest likelihood of being transferred successfully to a QC site. Methods shall be selected that experience shows will likely be reproducible. Reproducibility is assessed late in the validation process. Methods should be designed to minimize direct analysis and turnaround time, to maximize efficiency, and to shake the convenience, safety, health and environmental impacts of the test as consistent with the other primary objectives of method development. Development considerations should include equipment and expertise limitations of the receiving laboratory, expected sample volumes, and achieving a balance between acceptable traditional validation parameter values (accuracy, precision, etc.) and analysis time and robustness. The receiving laboratory should participate in the early review of these methods to provide and prepare for the transfer (e.g., purchase equipment, obtain reference standard supply) and should also assess safety, meaningful system suitability parameters (derived from robustness studies), and overall clarity of the written method. Automated tests may be desirable. Where a method has been developed and performed on automated equipment, an early assessment should be made as to whether automation is critical to the accuracy, precision, reproducibility, and robustness of the method compared to a manual method.

Validations and Specifications (Limits)

Method validations and drug substance or finished pharmaceutical product specifications are intimately linked. To ensure transferability of the method and to ensure the method will operate successfully in a QC site, the method variation (from the intermediate precision) should be known and monitored. The method variation is an estimate of the variation that will be experienced in routine use of the method. More method variation will create unacceptable random failure rates, and provide no room for reasonable product variation or even minor stability changes. The method variation should be less than one-third of the interval from the mean or target (typically

the midpoint of the upper and lower specification limit) value to the nearest specification limit, or one-sixth of the in-specification operating range of the method, whichever is smaller. Otherwise, OOS results may occur, even for product that was produced at target and with zero process variation.

If a reasonable opportunity exists (based on common belief and experience) to improve the method variation to one-fourth or preferably less of the specification interval, this effort should be made. In the example of a drug product where the specification is 95%–105% with a target of 100%, then the maximum method variation must be less than 1.67%; however, if the method variation is above 1.25%, efforts should be made to reduce its variation. Ideally, the method variation should not be greater than 2.0% even if the specification range is wider than 10%. The 2.0% maximum method variation requirement does not apply to tests where the variation of the method is substantially confounded with variation of the product being measured.

Validation as a Good Business Practice

While compliance to legal requirements is paramount to the pharmaceutical industry and its associated regulatory bodies, it is by far not the only reason to judiciously develop and validate analytical procedures. Well-developed and validated methods represent good business practices, as haphazardly chosen and/or poorly validated methods can haunt a company financially for the short and long term.

Extensive sample preparation with long cycle times, and excessive hazardous solvents usage and disposal are easily calculable financial and safety losses. Imprecise stability data can easily lead to erroneous or shortened expiration dating assignments. Poorly validated methods increase the chances of OOS results and investigations absorb costly laboratory resources. Incapable methods (inadequate precision for the associated specification) cannot readily discriminate passing from failing products: therefore, one can release unacceptable product, leading to product recalls. Inaccurate methods can lead to sub- or superpotent products. Nonreproducible methods do not transfer quickly or efficiently and can become limiting factors of pre-approval inspections with potential to affect product launches.

REGULATORY REQUIREMENTS

Due to the importance of demonstrating the suitability of new analytical procedures described in submission dossiers, in the 1980s many regulatory agencies published

requirements for analytical validation, in varying details. The U.S. Food and Drug Administration (FDA) issued two guidelines, one for the applicant,^[10] the other for inspectors and reviewers.^[11] The first one is also intended to ensure that the analytical procedure can be applied in an FDA laboratory and requires, therefore, a detailed description of the procedure, reference materials, discussion of potential impurities, etc. With respect to validation, data should be provided to demonstrate an appropriate accuracy, precision, linearity, selectivity, and quantitation limit (QL) for impurities and degradation products. For drug product, recovery, lack of interference from placebo, and variability with respect to time, laboratory, operator, and column should be demonstrated.

The second guideline is focused on reversed-phase chromatography and gives details regarding critical methodological issues, as well as acceptable results for parameters. A revised draft for the first guideline, published in 2000,^{[12]b} focused on providing raw data in detail. However, this has nothing to do with validation and should be an inspection issue, as there is a real danger to burden the validation documentation with huge amounts of data, which make it difficult to concentrate on a scientifically justified demonstration of suitability. Additionally, there are some inconsistencies with the ICH documents, which are not helpful for the harmonization process.

The same validation characteristics were described in the *U.S. Pharmacopoeia (USP)*.^[13] Three categories of analytical procedures were distinguished: quantitation procedures for main components in drug products, procedures for determination of impurities, and procedures for pharmaceutical–technical characteristics (e.g., dissolution). The guidance on validation of the European Community^[14] was rather general and incorporated in the respective sections of the submission documentation. In the Canadian guideline,^[15] a very detailed discussion is provided for requirements and especially acceptance criteria. Although this gives some orientation, the given acceptance criteria were sometimes a bit ambiguous, for example, the intermediate precision/reproducibility of below 1% for drug substances.

International Conference on the Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH)

This process was initiated in 1990 in order to harmonize the submission requirements for new pharmaceuticals in

the three main regions of Europe, the United States, and Japan and to avoid duplication, inefficiencies, and delays. A forum was created for a constructive dialogue between regulatory authorities and industry. The six cosponsors of ICH were the European Commission, the European Federation of Pharmaceutical Industry Association (EFPIA), the Japanese Ministry of Health (MHW), the Japanese Pharmaceutical Manufacturers Association (JPMA), the Food and Drug Association (FDA), and the Pharmaceutical Research and Manufacturers of America (PhRMA). Several organizations such as the Canadian Health Protection Branch (HPB), the USP, and the European Pharmacopoeia (EP) participated as observers. Within the Quality section, seven topics were taken into account:^[1] stability, validation,^[2,3] impurities,^[4–6] pharmacopoeial harmonization, biotechnological products, specifications,^[7,8] and good manufacturing practice (GMP).

The ICH was very valuable in harmonizing terms and definitions^[2] as well as basic requirements^[3] for analytical validation. Of course, due to the nature of the harmonization process, there are some compromises and inconsistencies, but the importance of a proper validation is currently widely known and accepted. In Table 1, the required validation characteristics for the various types of analytical procedures are shown. In the following, the main ICH requirements for the validation characteristics are summarized. However, the ICH guidelines must not be regarded as a checklist. “It is the responsibility of the applicant to choose the validation procedure and protocol most suitable for their product”.^[2]

Specificity

“Specificity is the ability to assess unequivocally the analyte in the presence of components which may be expected to be present. Typically, these might include impurities, degradants, matrix, etc.”.^[2] For identification, discrimination between closely related compounds likely to be present should be demonstrated by positive and negative samples. For assay and impurity tests, available impurities/degradants can be spiked to the corresponding matrix or degraded samples can be used. Specificity can also be demonstrated by verification of the result with an independent analytical procedure. The overall specificity can (and often will) be obtained by a combination of several analytical procedures, for example, in case of a partly specific titration with a chromatographic impurity determination. In case of chromatographic separations, resolution factors should be obtained for critical separations. Tests for peak homogeneity, for example, by diode array detection (DAD) or mass spectrometry (MS)

^b<http://www.fda.gov/cder/guidance/2396dft.htm>.

Table 1 Validation characteristics normally evaluated for the different types of test procedures and the minimum number of determinations required

Validation characteristics	Minimum number	Analytical procedure			
		Identity	Impurities		Assay ^a
			Quantitative	Limit	
Specificity ^b	Not applicable	+	+	+	+
Linearity	5	—	+	—	+
Range	Not applicable	—	+	—	+
Accuracy	9 (e.g., 3 × 3)	—	+	—	+
Precision					
Repeatability	6 or 9 (e.g., 3 × 3)	—	+	—	+
Intermediate precision/ reproducibility ^c	2 series	—	+	—	+
DL	Approach dependent	—	— ^d	+	—
QL	Approach dependent	—	+	—	—

(+) normally evaluated; (—) normally not evaluated.

^a Including dissolution, content/potency.

^b Lack of specificity of one analytical procedure could be compensated by other supporting analytical procedure(s).

^c Reproducibility not needed for submission.

^d May be needed in some cases.

(From Refs. 2, 3).

are recommended. All tested substances should be documented including the rationale of their selection. All relevant results, e.g., as tables or chromatograms, should be provided, discussed, and evaluated.

Linearity

“The linearity of an analytical procedure is its ability (within a given range) to obtain test results which are directly proportional to the concentration (amount) of analyte in the sample”.^[2] It can be distinguished between the linearity of the detector/instrument, obtained from a dilution of the analyte and the linearity of the analytical procedure, obtained from independent preparations (spiking, weighing) including (as far as possible) the complete sample pretreatment. At least five concentrations over the whole working range should be analyzed. Besides a visual evaluation of the analyte signal as a function of the concentration, appropriate statistical calculations are recommended, such as a linear regression. The parameters slope and intercept, the sum of squares, and the coefficient of correlation should be reported.

Range

“The range of an analytical procedure is the interval between the upper and lower concentration (amounts) of analyte in the sample (including these concentrations) for which it

has been demonstrated that the analytical procedure has a suitable level of precision, accuracy and linearity”.^[2]

The following minimum ranges are required:

- Assay: 80–120% of the test concentration.
- Content uniformity: 70–130% of the test concentration.
- Dissolution: 20% below to 20% above the specified range.
- Impurities: reporting level to 120% of the specification.
- 100% method: reporting level of the impurity to 120% assay specification.

Accuracy

“The accuracy of an analytical procedure expresses the closeness of agreement between the value which is accepted either as a conventional true value or an accepted reference value and the value found”.^[2]

This validation characteristic is an indication of systematic errors (bias). Accuracy can be demonstrated by the following procedures:

- Inferred from precision, linearity, and specificity.
- Comparison of the results with those of a well characterized procedure.
- Application to a reference material (drug substance).
- Recovery of drug substance added to placebo (drug product).

- Recovery of the impurity added to drug substance or drug product (impurities).

For the quantitative approaches, at least nine determinations across the specified range should be obtained, for example, three determinations at three concentration levels. The percent recovery or the difference between the mean and the accepted true value together with the confidence intervals (CIs) are recommended.

Precision

“The precision of an analytical procedure expresses the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogeneous sample under the prescribed conditions. Precision may be considered at three levels: repeatability, intermediate precision, and reproducibility”.^[2] Precision measures random errors and should be obtained preferably with authentic samples. As parameters, the standard deviation, the relative standard deviation (RSD) (coefficient of variation), and the CIs should be calculated for each level of precision. Repeatability reflects the analytical variability under the same operating conditions over a short interval of time (within-assay, intra-assay). At least nine determinations across the specified range or six determinations at 100% test concentration should be performed. Intermediate precision includes the influence of additional random effects according to the intended use of the procedure, i.e., within laboratories variations, for example, different days, analysts, equipment, etc. Reproducibility, i.e., the precision between laboratories, is not required for submission, but can be taken into account for standardization of analytical procedures. The variations should be selected according to the intended use of the analytical procedure.

Detection and quantitation limit

“The detection limit (DL) of an individual analytical procedure is the lowest amount of analyte in a sample which can be detected but not necessarily quantified as an exact value. “The QL of an individual analytical procedure is the lowest concentration of analyte in a sample which can be quantitatively determined with suitable precision and accuracy”.^[2]

Various approaches can be applied:

- Visual definition.
- Calculation from the signal-to-noise ratio (DL and QL correspond to the 3- or 2- and 10-fold of the noise level, respectively).

- Calculation from the standard deviation of the blank (Eq. 1).
- Calculation from the calibration line at low concentrations (Eq. 1).

$$DL; QL = \frac{F \times SD}{b} \quad (1)$$

where F is the factor of 3.3 (DL) or 10 (QL); SD is the standard deviation of the blank, standard deviation of the ordinate intercept, or residual standard of the linear regression; and b is the slope of the regression line.

The estimated limits should be verified by analyzing a suitable number of samples containing the analyte at the corresponding concentrations. DL or QL and the procedure used for determination as well as relevant chromatograms should be reported.

Robustness

“The robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage”.^[2] During the development phase of the analytical procedure, susceptible parameters should be identified, for example, stability of analytical solutions, extraction time, pH and composition of mobile phase, column lots and suppliers, temperature, flow rate, etc. A factorial design is encouraged.

System suitability tests

“System suitability testing is an integral part of many analytical procedures ...”.^[3] Suitable parameters depend on the type of analytical procedure; reference is given to pharmacopeias.

RATIONAL AND EFFICIENT VALIDATION

As analytical procedures are used throughout drug development and the manufacturing and release of drug substances and drug products, the reliability of their results is essential. Important decisions such as the establishment of the shelf-life from stability studies, the need for additional toxicological trials if new impurities appear or if known impurities exceed the qualified levels, and the reworking of batches and batch release or rejection are based on analytical results. Therefore, an appropriate validation to demonstrate the performance and suitability

of the analytical procedures is much more than a formal requirement.

Suitability of the Analytical Procedure

What does “suitability for its intended purpose” mean? Basically, suitability is determined by the specification limits (or the aim of the analytical investigation) and the design of the given test item. For some applications, the requirements are explicitly defined in the ICH guidelines. For example, the reporting level for unknown impurities in drug substances is set to 0.05% or 0.03%;^[4] thus the corresponding test procedure must be able to quantify impurities at this concentration with an appropriate level of precision and accuracy. A maximum permitted analytical variability can be calculated from assay specification limits^[16] (see Fig. 1 and Eq. 2), based on the concept of CIs to describe analytical and manufacturing variability.^[17] At least, the compatibility between specification limits and analytical variability should be verified. Preferably, the limits are established taking the analytical and manufacturing variability into account. In this case, an analytical variability normally expected for the given kind of analytical procedure (analytical state of the art) can be defined as acceptance criterion. Of course, the thus-obtained limits must primarily meet necessary quality and safety requirements. However, if this is satisfied, limits can reflect analytical variability.^[7,8] With respect to

assay determinations, the variability of the analytical procedure is often larger than the variability of the manufacturing.

$$s_{\max} = \frac{|(BL - SL)|\sqrt{n}}{t_{n-1,0.95}} \quad (2)$$

where BL is the basic limit, obtained from the theoretical content and the manufacturing variability, with respect to the critical “half” of the specification range; SL is the (overall) specification limit with respect to the critical “half” of the specification range; n is the number of repetitions in the assay; t is the Student’s t -factor for $(n - 1)$ degrees of freedom in the validation and 95% statistical confidence (one-sided).

It must also be demonstrated that the design of the analytical procedure, for example, the intended calibration mode, the number of repetitions, etc. is suitable. Therefore, suitability is strongly connected with the given, individual analytical procedure. The performance parameters are of varying importance and have different acceptance criteria. In consequence, a “checklist” approach to validation must be avoided. The analyst has to identify critical parameters, which are of importance for the required performance of the individual analytical procedure, to design the experimental studies accordingly and to define acceptance criteria for the results.

Statistical tests should only be carefully (directly) applied as acceptance criteria due to the small number of data normally obtained. Sometimes, because of abnormally small variabilities in the analytical series, differences are identified as significant which are of no practical relevance.^[18] In addition, when comparing independent methods for the proof of accuracy, different specificities can be expected which add a systematic bias, thus increasing the risk of the aforementioned danger. The analyst must decide whether detected statistical differences are of practical relevance. On the other hand, a large variability can also obscure differences which are not acceptable. If validation software is used, it must be flexible enough to meet these precautions.^{[19]c}

The analytical state of the art should be taken into account, although it is not the ultimate goal to optimize an individual analytical procedure as well as possible. It is also very important to recognize that the release of a given batch is based on a whole set of test procedures which complement and supplement each other. Their selection in the specification design^[7,8] has, therefore, considerable influence on the required performance of the individual control test and, hence, on its validation.

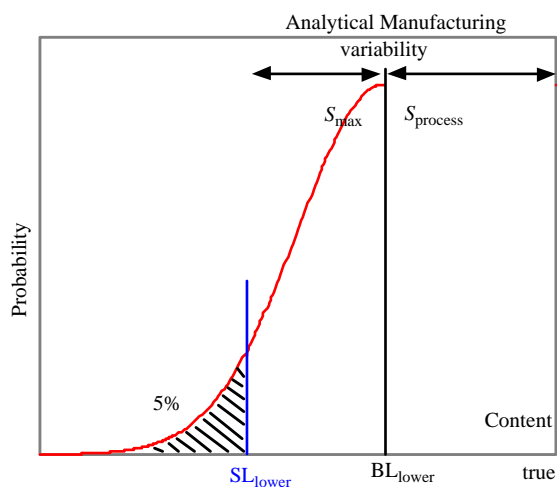


Fig. 1 Construction of specification limits from 95% CIs of the probability distribution of experimental results. The procedure is shown for the lower limits. BL = basic limits, imposed by the variability of the manufacturing process. SL = overall specification limit, combination of BL and the (lower) limit of the one-sided 95% confidence interval of the analytical variability. (From Ref. 16.)

^c<http://www.novia.de>.

Validation Characteristics

Specificity

There has been some controversy regarding the technical term for this validation characteristic, i.e., specificity vs. selectivity.^[20] This may be partly attributed to the fact that in contrast to an isolated test procedure in analytical chemistry, in pharmaceutical analysis, the sum of various control tests and hence their combined specificity is used for the overall evaluation.

In spite of all discussions, there is a broad agreement that this validation characteristic is of crucial importance and builds the critical basis for each analytical procedure. As no absolute and quantitative measure exists (at least for the overall specificity), the requirements depend on the individual analytical procedure as well as on its combination with others.

Chromatographic Separation. With respect to chromatographic techniques, specificity can be demonstrated by a sufficient separation of the substances present. For the assay, appropriate separation means an adequate resolution between the peak of interest and other peaks (e.g., impurities, placebo or matrix components), which need not to be separated from each other. In contrast, universal procedures for the determination of impurities require a sufficient separation of all relevant impurity peaks. The required resolution is strongly dependent on the difference in the size of the corresponding peaks as well as on their elution order.^[21] Therefore, if separation factors are determined, the typical concentration levels or the specification limits (as worst case) of the impurities should be maintained. Resolution factors can be calculated according to EP (Eq. 3) and USP (Eq. 4) at half height and at the baseline, respectively. However, this is only sensible for baseline-separated peaks. The USP approach is less sensitive toward tailing, but more complex to determine.

$$R_s = \frac{1.18(t_{Rb} - t_{Ra})}{w_{0.5a} + w_{0.5b}} \quad (3)$$

$$R_s = \frac{2(t_{Rb} - t_{Ra})}{w_a + w_b} \quad (4)$$

where $t_{Ra,b}$ is the retention time of peaks a and b with $t_{Rb} > t_{Ra}$, $w_{0.5a,b}$ is the peak width a and b at half height, and $w_{a,b}$ is the peak width a and b at baseline.

In case of incomplete separations, especially for peaks of different magnitude, calculations according to Eqs. 3 and 4 are not possible or are biased due to the additivity of the peak curves. Here, other separation parameters such as

the peak-to-valley ratio (p/v) should be used.^[23] This approach, which measures the height above the extrapolated baseline at the lowest point of the curve separating the peaks with respect to the height of the minor (impurity) peak (Fig. 2), is directly related to the peak integration and independent on tailing or “smearing” effects in the elution range behind the main peak.^[22,23]

Besides a “physical” improvement of the separation, it can also be optimized “visually” if the spectra of the two concerning peaks are different. Then, a suitable wavelength can be selected to suppress interferences. However, such an approach has to be balanced with respect to the QL and the robustness of the quantitation if the detection wavelength does not correspond to a stable region of the spectra, such as (relative) maxima or minima or shoulders.

Peak Purity Investigations. In order to be able to detect the coelution of unknown substances, peak homogeneity (also termed peak purity) investigations should be performed.

In case of not too large concentrations differences of the coeluting peaks, simple methods of one-dimensional detection can be applied. For normal eluting pure substances, the peak width at half height is proportional to the retention time. Therefore, performing isocratic chromatography with different concentrations of organic modifier, a plot of the peak width vs. the retention time will give a linear relationship for homogeneous peaks. For a single chromatographic separation, the symmetry factor is independent on the height of pure substance peaks.^[24]

In case of small amounts of coeluting impurities, rechromatography of suspected peaks represents a simple, universally available and sensitive approach. The more the two applied chromatographic methods differ, the

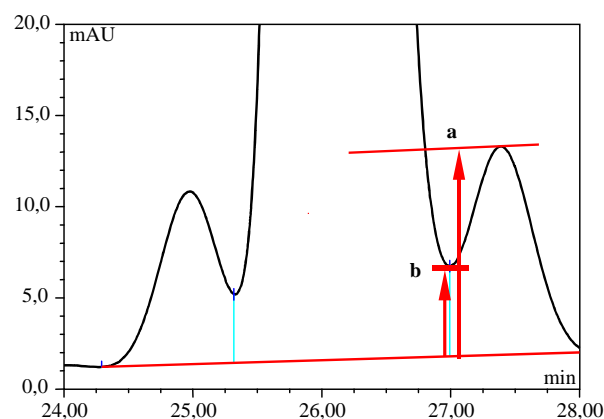


Fig. 2 Peak separation indices: peak-to-valley ratio $p/v = a/b$. If the calculation is modified as $p/v = 1 - b/a$, the parameter is normalized and has the same sense as the resolution factor, i.e., increasing values indicate better separation, with a maximum of 1.

greater is the power of the investigation. Various combinations can be taken into consideration such as coupling of reversed phase (RP) chromatography with another RP method (different eluent, pH, column), size exclusion chromatography, ion chromatography, thin layer chromatography (TLC), capillary electrophoresis (CE),^[25] gas chromatography (GC), etc. The rechromatography can be performed off-line with isolated peak fractions, or as a direct orthogonal coupling of the two methods. Working with isolated peak fractions, care should be taken to avoid artifacts due to degradation.

With diode array detectors or scanning detectors, the spectral peak homogeneity can be investigated.^[26] However, such an approach requires a difference in both the spectra and in the retention time of the coeluting substances. If this is fulfilled, detection of inhomogeneities with commercially available software is easy, if the concentration difference is not too large (Fig. 3b). However, impurities below 1% are difficult to recognize (Fig. 3c).

The most discriminating technique for investigation of the peak purity is mass spectrometric detection.^[27,28] Using on-line LC-MS coupling, mass spectra are taken over the whole elution range of the suspected peak (Fig. 4). In a first step, the obtained signals (mass-to-charge ratio, m/z) must be assigned to the main substance. In the given example, m/z 275.8 and 295.6 belong to the drug substance (Fig. 4, inset B) and represent the doubly charged molecular ion and a cluster formed by the doubly charged ion and acetonitrile (from the eluent). If during the spectra "scrolling" additional masses are detected such as m/z 324.9 (Fig. 4, inset A), the corresponding mass chromatogram is extracted. Differences in the retention time or elution behavior with respect to the UV peak is proof of a coeluting impurity. Even peaks with identical retention can be attributed to impurities if a

change in the mass-to-charge ratio cannot be explained by the drug substance itself. In the given example, the impurity amounts to 0.5%. Of course, the DL depends on the individual MS response of the concerned substances and diastereomers cannot be detected. If LC procedures with nonvolatile buffers are validated, the corresponding peak fractions can be isolated and rechromatographed under MS-compatible conditions. Alternatively, the coupling can be performed on-line.^[29] Although not often applied in routine (pharmaceutical) analysis, MS detection offers tremendous gains in efficiency and reliability of the procedures, such as highly specific detection (largely) without interferences, for monitoring of impurity profiles and identification.^[28]

If samples from stress testing are used to demonstrate appropriate separation power, care should be taken to avoid overdegradation as this would result in secondary (or even higher order) degradants which are of no practical relevance. Therefore, degradation should be restricted to about 10%. Alternatively, samples from regular stability studies (accelerated storage conditions) may be used.

Linearity and range

Although a linear relationship between the analyte concentration and the measured signal will exist for most methods used in pharmaceutical analysis, there are some exceptions such as TLC and fluorescence detection. Therefore, the term "analytical response" would be more appropriate.

Prerequisites for the Calibration Types. It depends on the design of the analytical procedure as to which regression parameters are meaningful and which results are acceptable. In other words, the model to be used for

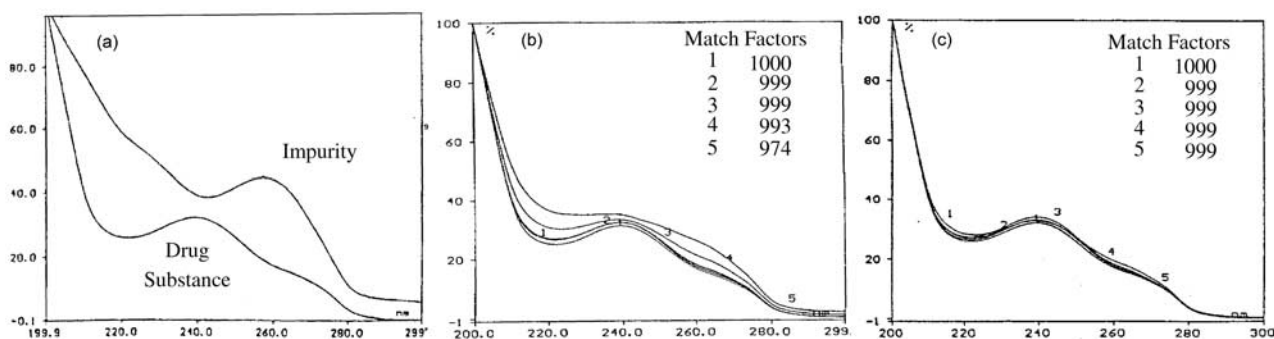


Fig. 3 Investigation of chromatographic peak purity by diode array detection: (a) spectra of drug substance and impurity; (b) coelution of a mixture containing about 10% impurity; (c) coelution of a mixture containing about 0.5% impurity. The spectra were obtained in the peak maximum, at about 5% and 50% of each side of the peak. The normalization was performed with respect to the first spectrum at the peak front (a matchfactor of 1000 means identical spectra) using commercial software.

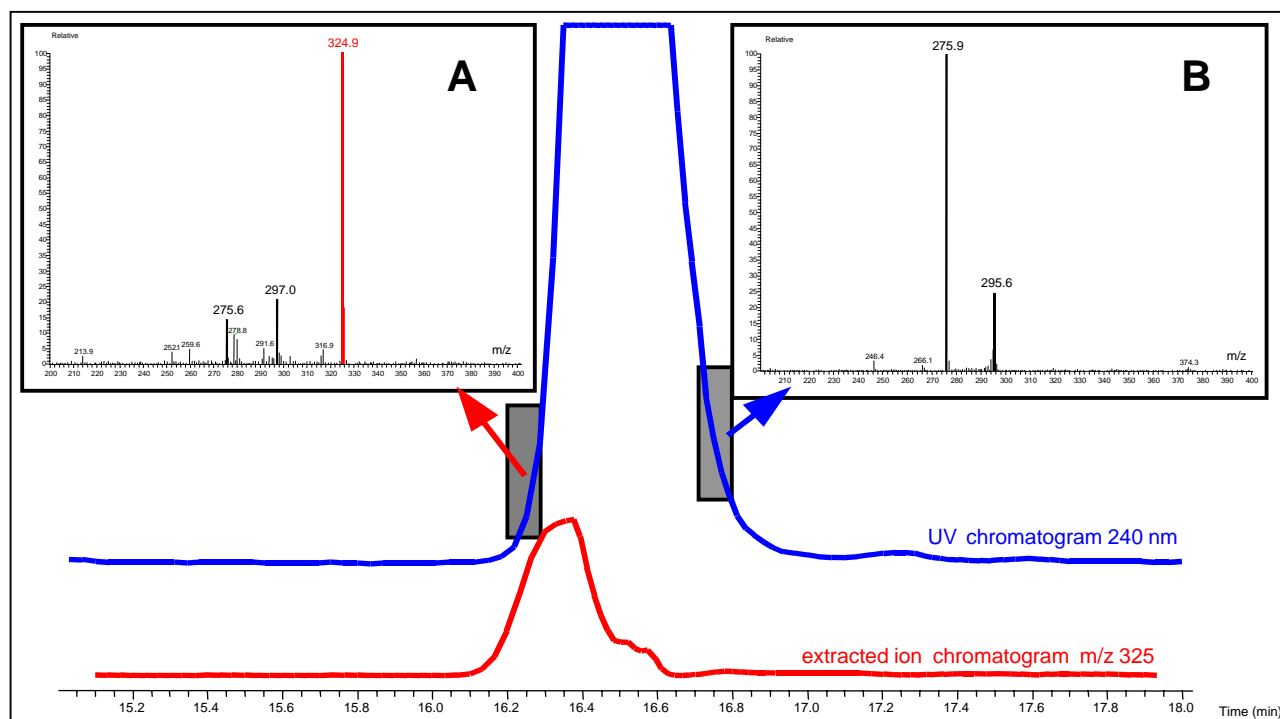


Fig. 4 Investigation of chromatographic peak purity by means of LC-MS. The top and bottom trace show the UV chromatogram at 240 nm and the extracted ion chromatogram for a mass-to-charge ratio (m/z) 325. Insets A and B display (two of the) mass spectra taken at different retention time ranges of the UV-peak. Optimizing the chromatographic conditions resulted in a separation of the impurity which amounts to 0.5% area at 240 nm (data not shown).

quantitation must be justified. For a single-point calibration (external standardization), a linear function, zero intercept, and the homogeneity of variances are required. The prerequisites for a linear multiple-point calibration are a linear function and in case of an unweighted calibration also the homogeneity of variances. A nonlinear calibration requires only a continuous function. With respect to the 100% method (area normalization for impurities), both for the main peak and the impurities, a linear function and a zero intercept are required, within their working ranges.

The required linearity range must be obtained from the working range of the analytical procedure. It may be useful to extend the minimum concentration range when applying a single-point calibration in order to avoid an extensive extrapolation to zero. This might cause problems in the evaluation of the significance of the intercept. If the calibration for an analyte determination in a complex matrix (drug product, impurities, and degradants by external standards) or with a complex sample preparation is intended to be performed with a simple solution of the reference standard, this linearity should be compared with the linearity of the whole procedure using reconstituted (spiked) samples.

The homogeneity of variances over the whole range (homoscedasticity) is a prerequisite for an unweighted linear regression in order to ensure the same influence of all concentrations (Fig. 5). This can be verified by performing a suitable number of repeated measurements ($n = 6-10$) at the minimum and the maximum of the required linear range, and comparing the variances with an F -test. However, in pharmaceutical analyses, as the concentration range of interest is usually not very large, it is not required to perform a separate test for the homogeneity of variances. In general, the homogeneity of variances is maintained over about two orders of magnitude when using UV absorbance (within the linear range of the detector/instrument). If a calibration is extended over several orders of magnitude (which is more common for bioanalysis, but may also be considered for impurity determinations), variances are not likely to be homogeneous. As an unweighted linear regression minimizes the absolute residual sum of squares, higher concentrations with larger absolute scattering dominate (Fig. 5a) and are better fitted to the regression line. Consequently, large deviations will occur if such a regression line is used for quantitation of small concentrations (Fig. 5b). For a better

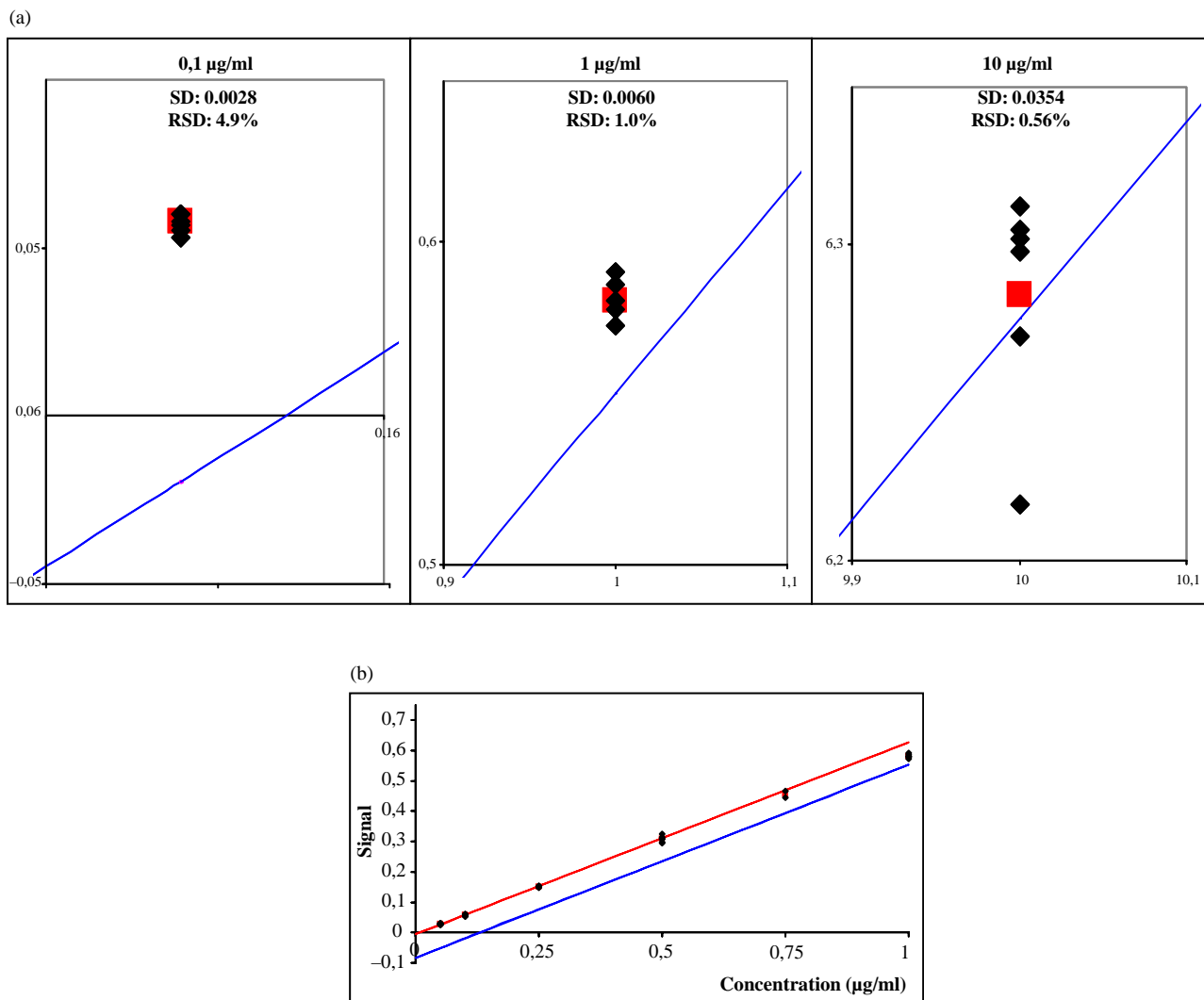


Fig. 5 Dependence of the variability on the analyte concentration (a) and its influence on the regression line (b). (a) Three concentration levels with dispersion parameters, calculated from six determinations each. The same scaling with respect to the signal axis is used. Whereas the absolute standard deviation decreases about 14-fold over three orders of concentration magnitude, the RSD increases considerably. The line obtained by unweighted linear regression of experimental data ranging from 0.05 µg/ml to 100 µg/ml is shown in blue. (b) The lines obtained from weighted (red line, weighting factor $1/x$ and unweighted (blue line) linear regression from 0.05 µg/ml to 100 µg/ml are shown in the lower concentration range.

representation of the lower concentrations, additional weight must be given to their signals performing a weighted linear regression. As weighting factors, the reciprocals of the individual standard deviations or variances can be used or, as an approximation the reciprocals of the concentrations themselves or of their squares. The equations can be found in statistical textbooks or in corresponding software products.

If deviations from linearity are detected or known, nonlinear response functions must be applied, for example, quadratic regressions. However, these models should be

used carefully if a linear relationship is normally assumed for the analytical procedure. Otherwise, it would also fit erroneous experimental values due to the larger flexibility of the regression curve. According to the principle of Ockham's blade, the model should be kept as simple as possible. Even in the case of intrinsic nonlinear response functions, it can be investigated if deviations resulting from linear regression are acceptable in the required working range. In the example given in Fig. 6, the difference between the results obtained by the quadratic regression and a linear regression from 400 ng to 700 ng

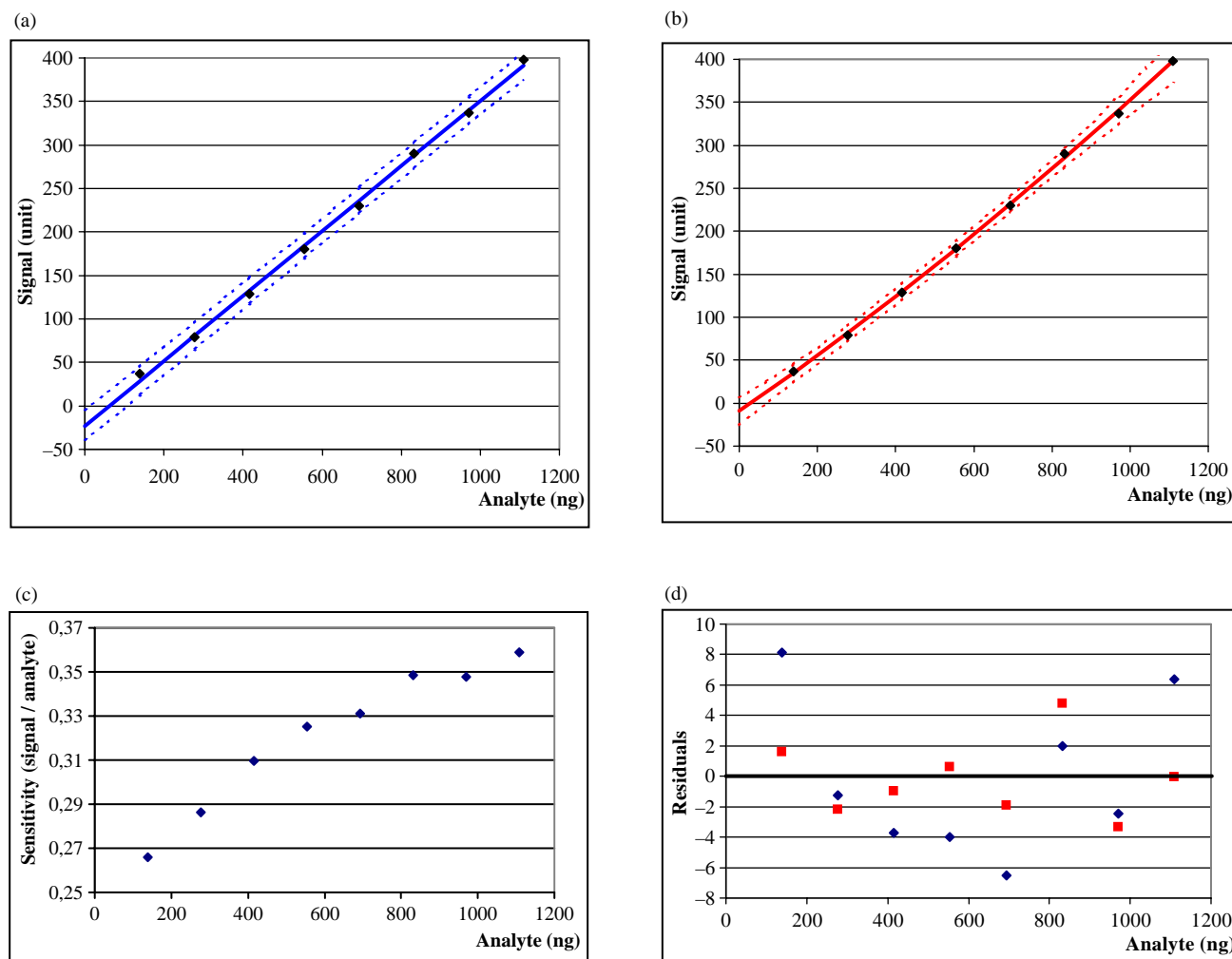


Fig. 6 Deviation from a linear response function. The experimental data were obtained by an LC procedure with nitrogen detection. (a) Unweighted linear regression ($y = a + bx$). The regression line and the limits of the 95% prediction interval are shown. The following parameters were calculated: slope with 95% CIs: 0.374 ± 0.015 ; intercept with 95% CIs: -23.09 ± 10.72 ; residual sum of squares: 189.68; residual SD: 5.62; relation standard error of slope: 2.41%; RSD of sensitivities: 10.12%; coefficient of correlation: 0.99916. (b) Quadratic regression ($y = a + bx + cx^2$). The regression line and the limits of the 95% prediction interval are shown. The following parameters were calculated: coefficient a : -9.217 ; coefficient b : 0.314; coefficient c with 95% CIs: $0.000048\% \pm 65.04\%$; residual sum of squares: 45.965; residual SD: 3.032. (c) Sensitivity plot (ratio of signal to analyte vs. analyte). (d) Residuen plot (difference of experimental to calculated signals vs. analyte) for the unweighted linear regression (blue) and the quadratic regression (red).

analyte (to represent a multiple-point linear calibration) is calculated to 0.4% and 0.55% at the lower and upper limit of the range, respectively. This must be evaluated with respect to a precision of about 2% for this type of procedure.

Evaluation of Linearity. In the ICH guideline, there are only scarce hints how to evaluate linearity. Primarily, a qualitative statement is sufficient for the evaluation of linearity (linear function): Does a linear relationship exist

in the required working range? When aiming at a single-point calibration, a positive answer is sufficient (in addition to a zero intercept) because the regression parameters obtained during validation are not used further in routine testing. In pharmaceutical analysis, this is also the case using a multiple-point calibration. However, relevant parameters may be selected to define acceptance limits for a system suitability test in routine calibration.

The coefficient of correlation is generally expected (and also mentioned in the ICH guideline), but it is not a

quantitative measure for the degree of linearity. It just gives an indication of whether a relationship exists between two sets of data.^[30] This is illustrated in Fig. 6. The coefficient of correlation for a linear regression shown in Fig. 6a was calculated as 0.99916. However, the more narrow 95% prediction interval of the quadratic regression (Fig. 6b) as a measure for the expected deviation of (future) experimental data from the regression line indicates a deviation from linearity. This is also easily recognized by investigation of the sensitivities, i.e., the ratio of analytical signal and the corresponding concentration (Fig. 6c). Their graphical presentation as a function of the concentration results in a horizontal line for the linear range of the procedure (Fig. 7). The ASTM^[31] recommends an interval of $\pm 5\%$ around the sensitivity mean. This interval should be adjusted to the concentration range in question; an acceptable precision can be used for orientation purposes. For concentrations around the QL, a wider interval can be accepted than for an assay procedure (e.g., $\pm 2\%$). Plotting the differences between the experimental values and the values calculated using the regression function vs. the concentration (residual plot, scatterplot) is another possibility of graphical linearity evaluation. With the proper response function, the residuals display a random (and narrower) scattering around the zero line; otherwise a systematic pattern is observed (Fig. 6d). However, in case of only few data (e.g., five according to ICH), a systematic pattern might be difficult to recognize. Here, the sensitivity plot is preferable as trends are better detectable as an upward or downward sloping of the data points.

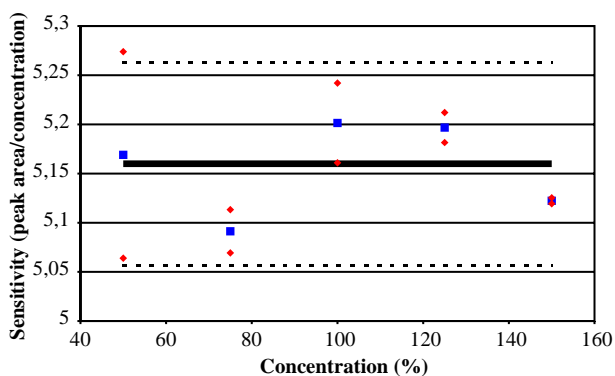


Fig. 7 Sensitivity plot (ratio of signal to analyte vs. analyte) for an LC-assay procedure with UV detection. The means (blue) and the individual (red) sensitivity values, the average sensitivity (solid line) and the $\pm 2\%$ limits (dotted lines) are displayed. The larger scattering for the lower concentrations is caused by their larger influence (weight) on the ratio, compared with larger concentrations.

The standard error of slope is a suitable numerical parameter for the evaluation of linearity. It measures the deviation of the experimental values from the regression line and thus represents a performance parameter with respect to the precision of the regression. Expressed in percent (relative standard error of slope), it is comparable to the RSD obtained in precision studies in the given concentration range (e.g., 10%–20% RSD at the QL, 1%–2% for assay). Therefore, this parameter is better suited for evaluation purposes than the residual sum of squares and the residual standard deviation, which are also measures of the deviation between the experimental data and the regression, but they depend on the absolute magnitude of the signal values and are difficult to compare with results from other equipment or with other procedures.

As statistically based linearity test, the significance of the quadratic coefficient can be checked (see Fig. 6b). If the CI of the quadratic coefficient c ($y = a + bx + cx^2$) includes zero, the quadratic term becomes zero and the equation is reduced to a linear function. However, if a statistical difference is obtained, their practical relevance should be evaluated. For example, if the data display a very small scattering, combined with a slight trend, a quadratic regression could result in a (statistically) better fit. But this does not justify the use of the more complex model, if the difference between the results of the two regression models is negligible in the working range compared with the expected precision (e.g., 0.1% maximum deviation and 1.4% RSD at the working concentration).

Evaluation of the Intercept. The occurrence of (non-negligible) ordinate intercept is an indication of systematic errors, as it can be seen in Fig. 6a. If there are no matrix effects, adsorption of the analyte, etc. the regression line must pass through the origin (within the random error of the experimental values). This is a prerequisite for the establishment of a single-point calibration and of the 100% method for the determination of impurities. The so-called single-point calibration represents, in fact, a two-point calibration line where one point equals zero and the other the standard concentration. The zero intercept can be demonstrated statistically when the CI (usually at 95% level of significance) of the intercept includes zero. Again, such a statistical test should be interpreted carefully and a statistical significance should be evaluated with respect to its practical relevance. On the other hand, a large variability can obscure a substantial deviation of the intercept from zero. Therefore, as an absolute parameter, the intercept should be expressed as a percentage of the analytical signal of a 100% working concentration. For the acceptance limit, a basis for orientation may be sought in an acceptable value for the precision (e.g., 1%–2% for

assay). In fact, this approach can be regarded as an extrapolation of the variability at the working concentration to the origin. In order to avoid weighting effects, very large extrapolation may be disadvantageous. Even if the required range for an assay determination is 80%–120%, the linearity for a single-point calibration should be validated starting with approximately 10%–50%.

Accuracy

Accuracy by Comparison. For drug substance, the only possibility of a quantitative assessment of accuracy is the comparison to the results of another analytical procedure or to a reference (if established with other procedures and/or additional characterization). This can be performed statistically with a *t*-test (see statistical textbooks or corresponding software). However, the shortcomings of these statistical tests (or better the justification of their use) are especially important here. It must be taken into consideration that two independent analytical procedures most probably differ in their specificity. This may lead to a systematic influence on the results (Table 2). If the effect can be quantified, the means should be corrected before performing the statistical comparison. If a correction is not possible, the presumptions of the statistical test are violated and the *t*-test should

Table 2 Investigation of accuracy by comparison of an LC assay procedure with a nitrogen determination according to Dumas. The content obtained from the latter is corrected for the amount of impurities determined with LC. The critical value for the mean *t*-test is larger than the tabulated one, therefore, a (statistically) significant difference of the means is observed. However, the absolute difference between the two procedures, which have a completely different methodology, is only 0.89%

No.	Content LC (%)	Content N (%)
1	92.93	94.18
2	91.32	93.22
3	92.66	92.84
4	93.50	92.77
5	92.41	92.58
6	92.58	92.77
7	91.27	93.03
8	90.66	92.96
9	91.56	92.96
10	93.17	
Mean	92.21	93.03
95% CI	91.53–92.88	92.68–93.39
RSD	1.02%	0.50%
<i>t</i> -test (<i>df</i> = 17, 95%)		
Critical value	2.11	
Test value	2.38	

consequently not be performed. Instead, it should be evaluated if the absolute magnitude of the difference is below an acceptable value (e.g., 2%).

Accuracy by Recovery. Interferences between the matrix (placebo) and the analyte, adsorption within the equipment, incomplete extraction of the analyte during the sample treatment, degradation, etc. can be verified by spiking known amounts of the analyte to the matrix (placebo). For example, drug substance is spiked to placebo (reconstituted drug product) or impurities/degradants are added to drug substance or drug product and are subsequently analyzed. This should be performed as near as possible to the authentic conditions. Ideally, the drug product is prepared with different contents of active ingredient. The least authentic approach would be the addition of standard stock solutions to a placebo solution.

The recovery can be calculated either at each level separately as a percent recovery, or as a linear regression of the found analyte vs. the added one (recovery function) (Table 3).

In the former case, it can be tested, whether the recovery mean differs significantly from the theoretical value of 100% (e.g., by the inclusion of 100 within the 95% CI). However, again it should also be taken into consideration if the absolute magnitude of the difference is acceptable, especially with respect to impurities in low concentration ranges.

With respect to the recovery function, the slope and intercept can be tested vs. the theoretical values of 1 and 0 (by their 95% CIs) or vs. acceptable limits for deviation. Due to the different weighting effects, the two approaches might lead to different results. The percent recovery calculation gives easily interpretable results and should therefore be preferred, at least for narrow working ranges.

It is absolutely essential that the accuracy be validated with the same quantitation method that is used in the control test procedure. Recovery deviations from the theoretical values while performing a calibration with a drug substance alone may indicate interferences between the analyte and placebo components. In such a case, the calibration should be done with a synthetic mixture of placebo and drug substance standard. Such interferences may also be detected by the separate determination of linearity for dilutions of the drug substance and for a spiked placebo.

Precision

Precision should be measured using homogenous, authentic samples. According to the ICH recommendation of nine determinations over the whole range, it may also be

Table 3 Investigation of accuracy by recovery. Drug substance was added in the solid state to a granulation of placebo tablets in a range from 80% to 120% of the nominal content of 50 mg. The further sample preparation is performed according to the control test

	Analyte added (mg)	Analyte found (mg)	Recovery (%)
	39.880	39.640	99.40
	40.030	39.950	99.80
	40.590	40.140	98.89
	49.930	49.780	99.70
	49.880	50.180	100.60
	50.200	50.450	100.50
	60.470	60.050	99.31
	60.190	60.490	100.50
	60.130	59.830	99.50
Percent recovery (%)			
Average	99.80		
Difference to theoretical value	0.20		
95% CI	99.33–100.27		
	no significant difference to 100%		
RSD	0.61%		
Recovery function			
Slope	1.005		
95% CI	0.975–1.036		
	no significant difference to 1		
Intercept	–0.353		
	–1.909–1.202		
	no significant difference to 0		
Relative standard error of slope	0.63%		
Coefficient of correlation	0.99942		

measured using artificially prepared samples or sample solutions, combining the validation characteristics linearity, accuracy, and precision in one experimental series. However, it should be noted that a larger variability can result due to the additional preparation steps. On the other hand, problems of the sample homogeneity and of the sampling itself cannot be detected.

It is essential to be absolutely aware of the different levels of precision, especially if acceptance limits are defined or if the resulting variabilities are used for further calculations. Often, the real uncertainty of results is underestimated, especially with respect to long-term applications.^[32] The large variability of the experimentally determined standard deviation has also to be taken into account (Fig. 8).

In addition to the ICH levels repeatability and intermediate precision, the system precision, i.e., repeated injections/determinations of a single sample solution (also referred to as injection repeatability or injection precision), provides valuable information. Evaluation of these data will help us to show that the chosen equipment is suitable for its intended use. Injection precision will also

become part of the system suitability requirements of the method and an acceptance criterion should be appropriately set. In the EP, a procedure is described which links the maximum permitted injection precision with the specification limits, thus allowing a specific evaluation of the suitability.^[23] However, this approach requires that the specification limits are established appropriately, i.e., taking the analytical and manufacturing variability into account.^[33]

The various levels of precision may be calculated by means of an analysis of variances.^[34] The overall variation is divided into the contributions within and between the series, allowing us to assess the most sensitive part of the analytical procedure as well as the robustness (see Table 4). Acceptance limits for assay determinations can be derived from specification limits established on the basis of experience and the analytical state of the art. With the former approach, the suitability of either the specification limits or the precision of the analytical procedure is tested. Typical RSDs for system precision of LC assay procedures should range below 1%, for repeatabilities up to 1–2%, and for intermediate precision/reproducibility twice the

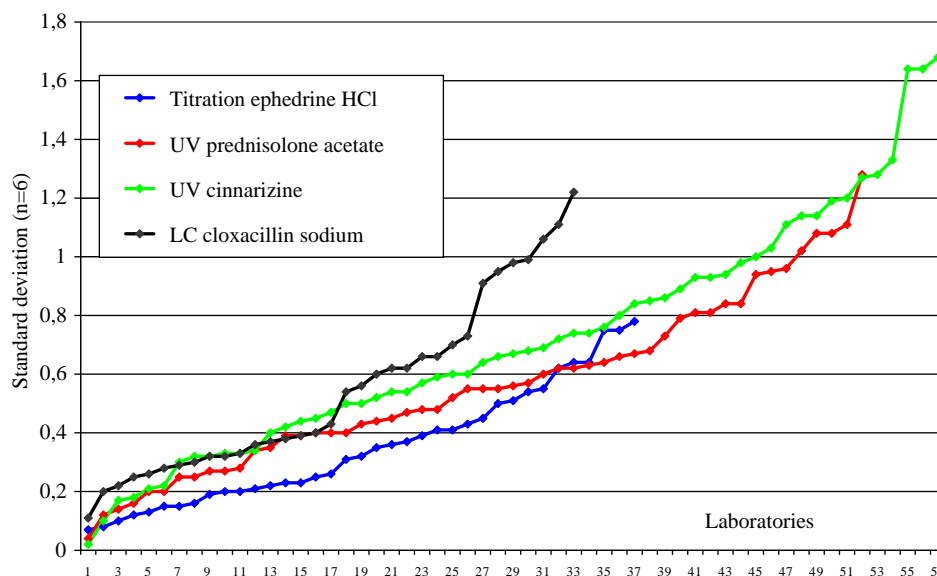


Fig. 8 Repeatability standard deviations for several types of drug substance assay procedures obtained from collaborative studies. The standard deviations from six determinations are ranked in order of increasing values. Cochran's test was applied to remove outlying SDs (not shown). The reproducibility was calculated from the remaining means after passing Grubb's outlier test. For procedures a–c, the true content is 100%, for d the overall mean is calculated to 96.77%. (Blue line) Potentiometric titration of racemic ephedrine hydrochloride. Thirty-nine laboratories participated, two were removed. The mean repeatability and the reproducibility are 0.41 and 0.64, respectively. (Red line) UV spectrometry of prednisolone acetate (specific absorbance). Sixty-four laboratories participated, 12 were removed. The mean repeatability and the reproducibility are 0.63 and 1.38, respectively. (Green line) UV spectrometry of cinnarizine (specific absorbance). Sixty-four laboratories participated, seven were removed. The mean repeatability and the reproducibility are 0.82 and 1.91, respectively. (Black line) LC of cloxacillin sodium. Thirty-four laboratories participated, one was removed. The mean repeatability and the reproducibility are 0.65 and 0.60, respectively. (Data from Ref. 36.)

value for the (average) repeatability can be expected (depending on the amount of variations, time period, etc.).^[35,36] For impurity determinations, the variability is strongly dependent on the concentration level. The reproducibility of the sum of impurities can sum up between 10% and 30%.^[37]

Detection and quantitation limit

One should be aware that the determined QL (or DL) is strongly related to the equipment used at the time of determination. They may represent more system parameters than characteristics of the analytical procedure.^[16,37] They are also dependent on the calculation procedure applied.

In cases where a general QL is required, as in pharmaceutical analysis, it is essential to define a realistic QL (or DL) for the analytical procedure, independently from the equipment used, because this limit has important consequences (e.g., for the consistent reporting of

impurities or for method transfer). They may be derived by taking QL (or DL) from various instruments into account ("intermediate QL," during the development process) or can be defined taking the requirements of the control test (specification limits imposed by toxicology or by a qualified impurity profile) into consideration. For example, a QL which amounts to 50% of the specification limit would allow an appropriate quantitation. For unknown impurities, the ICH reporting thresholds of 0.05% or 0.03%^[4,5] can be defined as QL. During validation, it is just verified, that the actual QL is below the defined limit (regardless of how much below).^[38] For this purpose, each of the approaches described in the ICH guideline can be used.

Calculation from noise. For chromatographic methods, DL and QL can be calculated from the noise. Here the (random) fluctuations of the baseline performing a blank injection is regarded as "noise." It is recommended to obtain the difference between the highest and the lowest signal (amplitude) in a range corresponding to at least 20 peak widths of the (expected) analyte peak.^[39] This is a very straightforward procedure, but is dependent on the

^dhttp://dg3.eudra.org/pharmacos/gmp_doc.htm.

Table 4 Reproducibility study for the content of active drug in a tablet formulation (mg/g). Material from ground tablets was weighed six times in two laboratories by two analysts each and analyzed according to the control test

	Lab 1		Lab 2	
	Analyst A	Analyst B	Analyst C	Analyst D
	698.9	701.5	687.9	695.9
	699.5	695.2	695.2	695.9
	696.5	699.0	687.8	694.6
	701.7	704.1	696.0	699.4
	709.1	703.8	700.1	693.9
	714.0	704.9	698.3	692.0
Mean	703.283	701.417	694.217	695.283
95% CI	± 7.131	± 3.914	± 5.485	± 2.608
RSD	0.97%	0.53%	0.75%	0.36%
BARTLETT-test for homogeneity of variances (99%)				
Critical value	11.300			
Test value	4.692			
Analysis of variances (ANOVA)				
Intra-group variance	23.381			
Inter group variance	16.127			
Overall variance	39.508			
Overall mean	698.550			
95% CI	± 2.654			
Repeatability	0.69%			
Reproducibility	0.90%			

operator.^[40] Care should also be taken to avoid the interpretation of baseline trends as “noise.” Only noise signals with a “peak width” similar to the analyte to be determined can influence its detection and determination. It should also be taken into account that the results represent signal (peak) heights.

Calculation from Linearity. For approaches based on a regression line, it must be taken into account that the residual standard deviation is strongly influenced by the absolute magnitude of the signal values (weighting effect, homogeneity of variances) and therefore by the concentration range used for the linear regression. This range should not exceed the 10-fold of the DL. In this range, the variances can be expected to be homogeneous.

Another approach is directly based on the dispersion of the experimental data in the low concentration range,^[41] represented by the prediction interval of the regression line. This interval can be interpreted as the probability distribution of (future) determinations, which can be experimentally expected. The upper 95% limit of the analyte concentration, whose probability distribution has a 50% overlapping with the distribution of the blank (Fig. 9, B) (and therefore a 50% error rate), is defined as

the DL (Fig. 9, DL). With respect to the QL the overlapping is reduced to 5% (Fig. 9, QL). Hence, a reliable quantification is possible in the latter case.

Estimation from Precision. The QL can also be obtained from precision studies. For this approach, decreasing analyte concentrations are analyzed repeatedly. The RSD is plotted against the corresponding concentration. If a predefined limit is exceeded (e.g., 10% or 20%), the corresponding concentration is established as the QL.^[15] However, a sufficiently large number of analyte concentrations must be analyzed due to the large scattering of the standard deviations in the low concentration range (Fig. 10). This approach can be simplified if a defined QL is to be verified. Then, a sample containing the impurity at the defined QL is repeatedly analyzed. If the obtained RSD is below the predefined limit, the analytical procedure has a suitable sensitivity.

If varying concentrations are used to obtain DL and QL, available impurities and degradants should be spiked to the drug substance or drug product. Different contents can also be obtained by mixing impurity-free and impurity-containing batches. The concentration of the active or matrix components (placebo, cleaning solutions) should be maintained

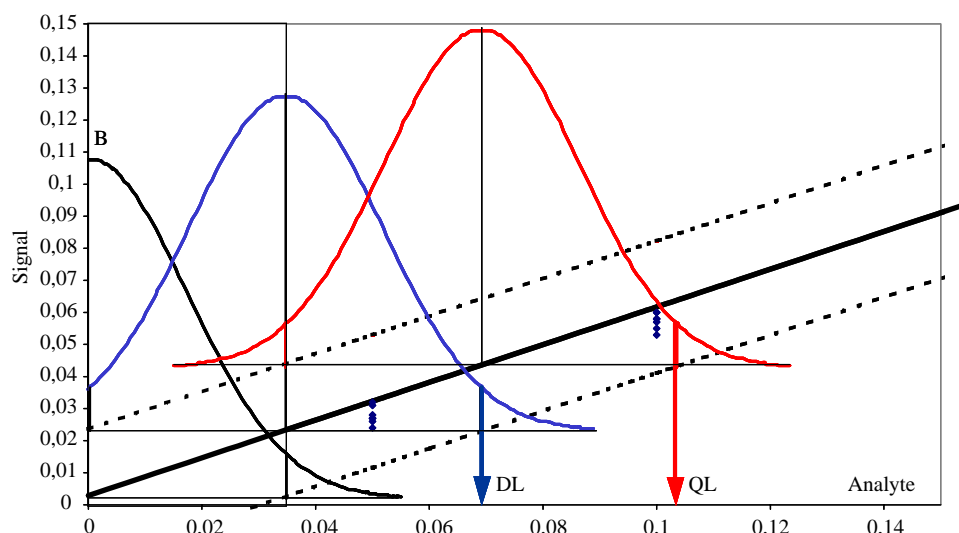


Fig. 9 Calculation of the detection and quantitation limits from the 95% prediction interval of the regression line. LC peak areas were obtained for six analyte concentrations between 0.05 $\mu\text{g/ml}$ and 1 $\mu\text{g/ml}$ with six repeated injections each. The regression line is shown with the 95% prediction intervals (dotted lines) in the lower concentration range. The probability distributions of the blank (B, black curve), of the analyte concentration corresponding to the detection limit (DL, blue curve) and the quantitation limit (QL, red curve) are indicated. The limits of the 95% prediction interval represent the 95% limits of the probability distribution of the corresponding concentration, i.e., the interval around the concentration where (future) experimental results can be expected in 19 out of 20 cases. All experimental results within the probability distribution of the blank (gray shaded area) cannot be distinguished from the blank. This amounts to 50% of all values in the case of the detection limit and 5% for the quantitation limit.

at the nominal level of the test. The QL of unknown substances can be obtained using representative peaks or inferred from the QL of known impurities/degradants.

The validity of the determined DL/QL should be routinely confirmed within the system suitability test (visually or from the signal-to-noise ratio).^[37]

Robustness

Deliberate variations in the parameters of a procedure will provide an estimate of its reliability in routine use. The extent of robustness studies will depend upon the type of method but this evaluation should be considered during the development phase.

Typical variations might include extraction time, flow rate through a testing device, stability of test solutions, etc. For chromatographic methods, typical variations for robustness studies would include different columns, temperature, flow rate, mobile phase pH, and organic content. Experimental design may be used during robustness experiments to develop a matrix approach to these studies,^[42,43] as well as HPLC optimization software. However, care should be taken to incorporate relevant variables into a quantitative statistical design. Otherwise, unnecessary workload is created, for example, if the investigated effects will influence the quantitation of

sample and standard identically such as flow rate, organic modifier, column, etc. These effects should only be investigated with respect to resolution.

A quantitative measure of the robustness of an analytical procedure is the different levels of the intermediate precision and their comparison (analysis of variances) (Table 4). Of course, this approach only addresses random effects, which depend on the extent of variations designed in the inter-

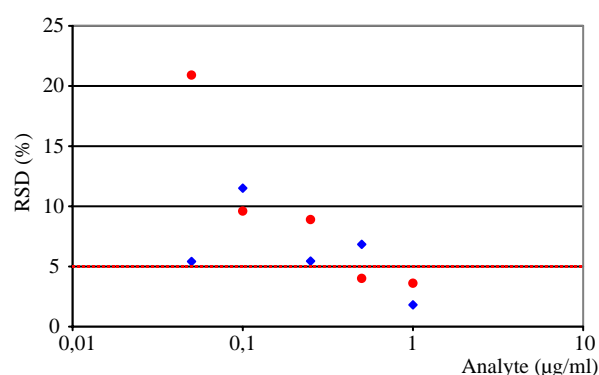


Fig. 10 Relative standard deviations obtained from six repeated injections of analyte concentrations from 0.05 $\mu\text{g/ml}$ to 1 $\mu\text{g/ml}$. Two series on the same LC system are shown, approximately 8 weeks apart. For illustration purposes, an arbitrary acceptance limit of 5% RSD for the injection precision is displayed.

mediate precision studies. In case of insufficient robustness, it will be often not possible to identify its source. In these cases, a systematic investigation is required.

If the robustness studies indicate that the experimental data are susceptible to variations in the method parameters, the appropriate system suitability requirements should be established to ensure the validity of the analytical procedure during routine use.

System suitability tests

The validation of an analytical procedure produces performance parameters of a “well-behaved” and “well-conditioned” system/instrument (including defined variations) which are more or less “snapshots” of the combination procedure/system. In order to routinely confirm the suitability of the integrated measurement instrumentation used with a given procedure, system suitability test parameters should be defined on the basis of the validation results and robustness studies. Larger variations under routine conditions or multiple laboratories should also be taken into consideration.

The extent of the system suitability testing is dependent on, for example, the stage of drug development, the objective of the analytical procedure, the availability of reference substances, etc. The following parameters may be considered:^[23,37,44,45]

- Comparison with a test chromatogram of reference standard (cf. retention times).
- System precision (injection repeatability); acceptance limit for LC: 2% RSD,^[44] for drug substance dependent on the (upper) specification limit [e.g., 102.0% and $n = 6$: 0.85%].^[23]
- Resolution factor—acceptance criteria: not less than 2.0 (recommended with respect to the main peak).
- Number of theoretical plates.
- Tailing factor—acceptance criteria: between 0.5 and 2.0 (recommended).
- Pressure drop.
- Baseline drift.
- (Detector) linearity or sensitivity at different concentrations.
- DL/QL (e.g., as signal-to-noise ratio—acceptance criteria: not less than 3 and 10, respectively).

Validation During Drug Development and Manufacturing

The ICH requirements are intended for the submission of new pharmaceuticals. Good manufacturing practice regulations, which include analytical procedures for

in-process controls, QC of excipients, starting materials, intermediates or during drug development prescribe no details of validation requirements. In these applications, the ICH guidelines should be used for orientation with an increasing expenditure of effort and requirements as development or manufacturing progresses. For validation purposes, three distinct levels can be described for which validation requirements can be defined.

Level I

Level I is intended for analytical procedures during the early development (preclinical phase), for starting materials and in-process controls. At this stage of development, relatively small quantities of drug substance are available, analytical reference standards have not been established, and little information is available regarding synthesis or degradation impurities. Also, frequent changes/improvements in drug substance and formulations are expected. As such, the validation that is conducted is at a minimum. Validation data may be extracted from routine testing. The specificity can be demonstrated using available batches and samples from stress testing in combination with simple variations in the chromatographic conditions (gradient, elution time, pH). For demonstration of linearity, a dilution of the analyte is performed (detector linearity). Accuracy is inferred from specificity, linearity, and precision. A system precision is performed with six injections. QL is defined visually. The results are summarized and evaluated in order to make a statement on the suitability.

Level II

Level II is applied for drug substances and products during the early clinical phases and for intermediates. At this stage of development, drug substance synthetic processes are becoming finalized and higher quality material is available in larger quantities. A drug substance reference standard is established and information is increasingly available on synthetic and degradation impurities. Reference standards are established for impurities as they become available. Formulation development activities are giving rise to finished products to be used in clinical studies. It is at this stage that formal control tests must be established. The validation effort is significantly extended from Level I. Specificity is investigated using various batches, available impurities, excipients, and samples from stability testing. Investigations on peak purity may be performed. The linearity of the detector or of the procedure, depending on the complexity of the latter, is validated. Accuracy is inferred (drug substance)

or obtained from three determinations of the recovery at the working concentration. A repeatability study ($n = 6$), or depending on the application of the procedure, an intermediate precision is performed. Quantitation limit is obtained visually or verified with one of the quantitative approaches. The documentation consists of a short summary of the performed investigations and an evaluation of the results.

Level III

Level III is applied during clinical phase III and for the submission, as well as for important intermediates, and corresponds to the ICH requirements. At this stage of development, the commercial synthetic process for drug substance is established, and important process and degradation impurities have been identified and synthesized. The final formulations and dosages have also been established.

A full validation report is required with description and justification of the design of the experimental studies, and a detailed discussion and evaluation of the results.

In all levels, validation software can be used to standardize and to improve the efficiency of documentation of the data and calculations. However, it should be flexible enough to allow the necessary case-by-case (or type-by-type) design of the validation.^[19]

Maintenance of the Validation Status

In several guidelines, the term revalidation is mentioned,^[2,10,12] sometimes in more detail.^[15] However, one should be aware that validation is always more or less a “snapshot” of the actual state of the overall system equipment—analytical procedure. Therefore, it should be regarded as a basis, i.e., the proof of the general suitability. If adjustments of the analytical procedure are restricted to the validated and recommended ranges,^[23,45] there is no rationale, for a repeated validation. Only if a procedure is changed, a new validation is required.

Of course, it must be guaranteed that the analytical procedure is consistently in a validated state. This can be achieved by monitoring appropriate SST results. Another possibility is to perform (regularly) supplementary investigations such as further intermediate precision studies, which can be included in the routine analysis. With this approach, it is possible to supplement and extend existing information and to take the very important time factor into proper consideration. Information can be obtained on the long-term behavior of an analytical procedure and therefore on the reliability in routine use,

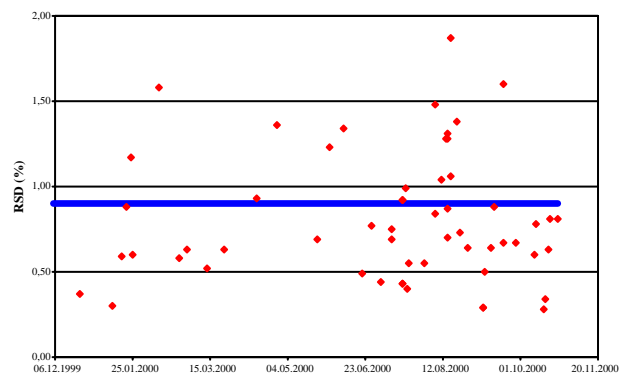


Fig. 11 Relative standard deviations of reference standard preparations for an LC assay procedure over a period of about 10 months. The RSDs were calculated from four injections of two independent preparations each. Fifty-eight results were obtained, two of them exceeded the defined limit of 2.0% (3.85% and 3.94%, not shown). The average RSD (0.88%) is indicated by a blue line. (Data courtesy of Dr. C. Korn, Aventis, Germany.)

for example, better estimates on the true (long term) analytical variability.

In Fig. 11, SST precision results for more than 10 months during drug development are shown. The RSDs were calculated from the standard preparations of an assay procedure, i.e., from four injections of two separate preparations (single-point calibration). Therefore, the variability represents a combination of system precision and repeatability. It should be noted that with an average of 0.88% (representative for the true value with 58 runs over 10 months), the single results scatter regularly up to 1.6%. Two-thirds of all RSDs range between 0.5% and 1.3%. In two cases (about 3%) the RSDs exceed the SST limit of 2.0%. This was attributed to gross (weighing) errors.

Quality control charts are a very effective approach to investigate not only the analytical variability, but also the accuracy. Due to the accumulation of results, additional forms of statistical analysis (statistical process control, SPC) become possible. The continuous monitoring and the proof that the analytical procedure is in statistical control allows a reliable evaluation of the underlying process (e.g., batch manufacturing) and an early detection of trends. Fig. 12 represents a control chart for an LC content determination of an injection solution. A control batch (“dummy”) was analyzed with a single determination each time the analytical procedure was applied (over about 6 months). The target mean and the standard deviation were calculated from the first 20 determinations and normalized to 100%. All but one result are within the control limits (\pm three standard deviations around the

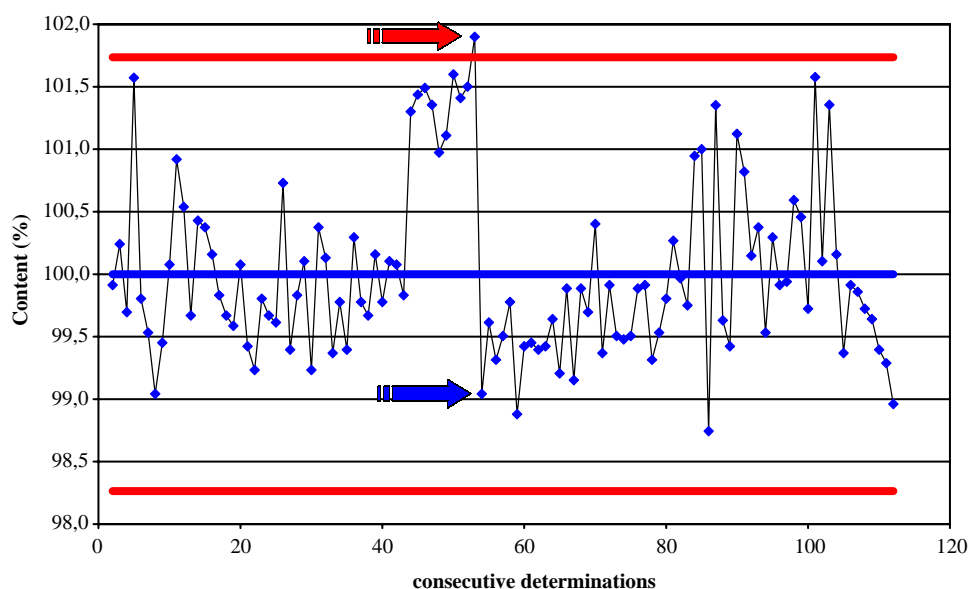


Fig. 12 Control chart of a single LC determination of an injection solution. The mean and the SD were calculated from the first 20 determinations. The mean is normalized to 100% (blue line). The control limits (red lines) correspond to threefold of the RSD. The experimental results are displayed consecutively and connected by a black line to facilitate the recognition of systematic pattern. The blue arrow indicates the use of a new standard solution after an out-of-control event (red arrow). Data by courtesy of Dr H.-J. Ploss, Aventis Germany.)

mean). The group of values preceding the outlying one (red arrow in Fig. 12) was obtained with the same standard solution and is shifted toward the upper control limit. For the next run (blue arrow in Fig. 12), a new standard solution was prepared. The “normal” result of 99.0% indicates that the shift was caused by a deviating standard preparation. The deviation from the mean is with less than 1.5% in the range of the expected analytical variability (for a single determination) and therefore difficult to detect in routine batch analyses. In contrast, such shifts and also trends can easily be detected in a control chart representation. Besides limit violations, the control chart can also inspected for “abnormal” (systematic) patterns, such as six points steadily increasing or decreasing, 14 points in a row alternating up and down, etc. (for details, see [46]). This provides the chance not only to obtain reliable (long term) information on the analytical system, but also to adjust it proactively.

Such a continuous process of analytical validation can be regarded in analogy to the qualification process.^[47d,48] With respect to the four Qs (design, installation, operational, and performance qualification), three Vs can be defined: design, operational, and performance validation. “Design validation” involves the method

development, where the basic performance requirements are established. During the “operational validation,” it is verified that these requirements are generally met, i.e., this part corresponds to the “conventional” validation. Performance validation consists of a continuous monitoring of the validation status of an analytical procedure.

CONCLUSIONS

Instead of an isolated and formal “checklist” activity, validation of analytical procedures should be regarded as part of an integrated quality assurance concept to guarantee the accuracy and reliability of the analytical results and therefore quality, safety, and efficacy of pharmaceuticals.

For such an integrated system, all parts (such as equipment qualification, performance of the analytical procedure, analytical variability, specification limits, etc.) must be compatible with each other to prevent (analytically caused) out-of-control or OOS results. On the basis of the validation characteristics and requirements of the ICH guidelines, each analytical procedure must be validated with respect to parameters, which are relevant to its performance. It is the responsibility of the analyst to identify these parameters and design the

^dhttp://dg3.eudra.org/pharmacos/gmp_doc.htm.

validation study accordingly. Acceptance criteria should be defined in the validation protocol. They can be established from previous experiences (analytical state of the art) or calculated from specification limits. For the intended use of the test procedure acceptable absolute acceptance limits are preferred. Statistical tests should be used carefully and preferably for orientation purposes. The evaluation of the validation results is the responsibility of the analyst and must not be left or reduced to the outcome of a statistical test! However, statistical analysis and consideration are very helpful in verifying the compatibility of specification limits and analytical variability, calculating acceptance limits, and performing simulations in order to predict future risks.

By monitoring the validation status of the applied analytical procedure, for example, with control charts, reliable information on the long-term behavior of the procedure can be obtained and trends can be detected very early. Transferring in this way data to information to knowledge, the analytical system (or the production process) can be adjusted before problems such as OOS results occur (action instead of reaction). In analogy to equipment qualification, a continuous system is proposed: design, operational, and performance validation (three Vs).

REFERENCES

1. International Conference on the Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH).
2. ICH Q2A: Validation of Analytical Methods (Definitions and Terminology), October 1994, Implementation (Step 5): EU: Adopted by CPMP, November 94, issued as CPMP/ICH/381/95; MHW: Adopted July 95, PAB/PCD Notification No. 755; FDA: Published in the Federal Register **1995**, 60, 11260.
3. ICH Q2B Analytical Validation—Methodology, November 1996, Implementation (Step 5): EU: Adopted by CPMP, December 96, issued as CPMP/ICH281/95; MHW: Adopted October 97, PMSB/ELD Notification No. 338; FDA: Published in the Federal Register **1997**, 62 (96), 27463–27467.
4. ICH Q3A, Impurities in New Drug Substances, March 1995, Revised Guideline: Consultation (Step 3): EU: Released for Consultation, November 1999, issued as CPMP/ICH/2737/99; MHW: Released for Consultation, PAB/PCD Notification No. 1829, 17 December 1999, deadline for comments on 31 March 2000; FDA: Published in the Federal Register **2000**, 65 (140), 45085–45090.
5. ICH Q3B, Impurities in New Drug Products, November 1996, Revised Guideline Consultation (Step 3): EU: Released for consultation, November 1999, issued as CPMP/ICH/2738/99; MHW: Released for consultation, PAB/PCD Notification No. 1829, 17 December 1999, deadline for comments on 31 March 2000; FDA: Published in the Federal Register **2000**, 65 (139), 44791–44797.
6. ICH Q3C, Residual Solvents, July 1997, Implementation (Step 5): EU: Adopted by CPMP, September 97, issued as CPMP/ICH/283/95; MHW: Adopted March 1998, PMSB/ELD Notification No. 307; FDA: Published in the Federal Register **1997**, 62 (247), 67377.
7. ICH Q6A, Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products, Chemical Substances, October 1999, Implementation (Step 5): EU: Adopted by CPMP, November 1999, issued as CPMP/ICH/367/96; MHW: To be notified; FDA: Published in the Federal Register **2000** 65 (251), Notices, 83041–83063.
8. ICH: Q6B, Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products, March 1999, Implementation (Step 5):EU: Adopted by CPMP, March 99, issued as CPMP/ICH/365/96; MHW: To be notified; FDA: Published in the Federal Register **1999**, 64FR, 44928.
9. Hokanson, G.C. A life cycle approach to the validation of analytical methods during pharmaceutical product development, Part I. Pharm. Technol. **1994**, 18 (9), 118–130, Part II: **1994**, 18 (10), 92–100.
10. Guidelines for Submitting Samples and Analytical Data for Methods Validation, U.S. Food and Drug Administration, Center for Drugs and Biologics, Department of Health and Human Services, 1987.
11. CDER Guideline on Validation of Chromatographic Methods, Reviewer Guidance of Chromatographic Methods, U.S. Food and Drug Administration, Center for Drugs and Biologics, Department of Health and Human Services, 1994.
12. Draft Guidance Analytical Procedures and Methods Validation. U.S. Food and Drug Administration, Center for Drugs and Biologics, Department of Health and Human Services, 2000.
13. United States Pharmacopeia 24, National Formulary 19, Section 1225 Validation of Compendial Methods United States Pharmacopeial Convention, Rockville, 2000.
14. The Rules Governing Medicinal Products in the European Community, Vol. 3, Addendum, 1990.
15. Acceptable Methods. Drug Directorate Guidelines, National Health and Welfare, Health Protection Branch, Health and Welfare Canada, 1992.
16. Ermer, J. Validation in pharmaceutical analysis. Part I: An integrated approach. J. Pharm. Biomed. Anal. **2001**, 24 (5–6), 755–767.
17. Van de Vaart, F.J. Content limits—Setting and using. Pharmeuropa **1997**, 9.1, 139–143.
18. Whitfield, R.G.; Hughes, D.W.; Layloff, T.P.; Cox, R.R.; Gressett, G.E.; Jimenez, P.J.; Andersen, J.; Reck, R.R.; Schniepp, S. Interpretation and treatment of analytical data. Pharmacop. Forum **1998**, 24, 7051–7056.

19. MVA—Method Validation in Analytics (PC-software), NOVIA GmbH, Frankfurt, Germany.
20. Persson, B.-A.; Vessman, J.; McDowall, R.D. Is your method specific or just selective? LC–GC Int. March **1998**, 160–164.
21. Meyer, V.R. Quantitation of chromatographic peaks in the 0.1 to 1.0% range. *Chromatographia* **1995**, *40*, 15–22.
22. Riley, C.M. Statistical parameters and analytical figures of merit. In *Development and Validation of Analytical Methods*; Riley, C.M., Rosanske, T.W., Eds.; Elsevier: Oxford, 1996; 15.
23. European Pharmacopoeia, 3rd Ed.; Supplement 2001, General Monograph 2.2.46, Chromatographic Separation Techniques, 28–32.
24. Ebel, S. Validation of analysis methods. *Fresenius J. Anal. Chem.* **1992**, *342*, 769–778.
25. Altria, K.D.; Dave, Y.K. Peak homogeneity determination and micro-preparative fraction collection by capillary electrophoresis for pharmaceutical analysis. *J. Chromatogr.* **1993**, *633*, 221–225.
26. Castledine, J.B.; Fell, A.F. Strategies for peak-purity assessment in liquid chromatography. *J. Pharm. Biomed. Anal.* **1993**, *11* (1), 1–13.
27. Bryant, D.K.; Kingswood, M.D.; Belenguer, A. Determination of liquid chromatographic peak purity by electrospray ionization mass spectrometry. *J. Chromatogr. A* **1996**, *721*, 41–51.
28. Ermer, J.; Vogel, M. Applications of hyphenated LC–MS techniques in pharmaceutical analysis. *Biomed. Chromatogr.* **2000**, *14*, 373–383.
29. Ermer, J.; Kibat, P.-G. A quality concept for impurities during drug development — Use of the hyphenated LC–MS technique. *Pharm. Sci. Technol. Today* **1998**, *1*, 76–82.
30. Analytical Methods Committee; Uses (proper and improper) of correlation coefficients. *Analyst* **1988**, *113*, 1469–1471.
31. American Society for Testing and Materials, *ASTM Designation E 1303-89*; ASTM: Philadelphia, 1989.
32. Renger, B. System performance and variability of chromatographic techniques used in pharmaceutical quality control. In *The Future Face of the European Pharmacopoeia*, International Symposium, Cannes, EDQM, February, 8–9, 2001.
33. Daas, A.G.J.; Miller, J.H.McB. Relationship between content limits, system suitability for precision and acceptance/rejection criteria for assays using chromatographic methods. *Pharmeuropa* **1999**, *11* (4), 571–577.
34. DIN ISO 5725-2: Accuracy (trueness and precision) of measurement methods and results; A basic method for the determination of repeatability and reproducibility of a standard measurement method, 1990.
35. Küppers, S.; Renger, B.; Meyer, V.R. Autosamplers — A major uncertainty factor in HPLC analysis precision. *LC/GC Eur.* February **2000**, *13* (12), 114–118.
36. Daas, A.F.J.; Miller, J.H.McB. Content limits in the European Pharmacopoeia. *Pharmeuropa* **1998**, *10* (1), 137–146.
37. Miller, J.H.McB. System suitability criteria—A case study: The determination of impurities in dicloxacillin sodium. *Pharmeuropa* **2000**, *12* (1), 8–17.
38. Validation of compendial methods. *Pharmacop. Forum* **1998**, *24*, 6582–6589.
39. Carr, G.P.R.; Wahlich, J.C. Analytical validation. In *International Pharmaceutical Product Registration*; Cartwright, A.C., Matthews, B.R., Eds.; Horwood Ellis: New York, 1994; 246–286.
40. Vial, J.; Jardy, A. Experimental comparison of the different approaches to estimate LOD and LOQ of an HPLC method. *Anal. Chem.* **1999**, *71*, 2672–2677.
41. Funk, W.; Dammann, V.; Vonderheid, C.; Oehlmann, G., Eds. *Statistische Methoden in der Wasseranalytik*; Verlag Chemie: Weinheim, 1985.
42. Bosque-Sendra, J.M.; Nechar, M.; Cuadros Redriguez, L. Decision protocol for checking robustness with previous outlier detection in the validation of analytical methods. *Fresenius J. Anal. Chem.* **1999**, *365*, 480–488.
43. Vander Heyden, Y.; Luypaert, K.; Hartmann, C.; Massert, D.L.; Hoogmartens, J.; De Beer, J. Ruggedness tests on the high-performance liquid chromatography assay of the United States pharmacopoeia XXII for tetracycline hydrochloride. A comparison of experimental designs and statistical interpretations. *Anal. Chim. Acta* **1995**, *312*, 245–262.
44. United States Pharmacopoeia 24, National Formulary 19, Section 621, Chromatography, United States Pharmacopoeial Convention, Rockville, 2000.
45. Furman, W.B.; Dorsey, J.G.; Snyder, L.R. System suitability tests in regulatory liquid and gas chromatographic methods: Adjustments versus modifications. *Pharm. Technol. Eur.* June **1998**, 30–33.
46. Shainin, D.; Shainin, P.D. Statistical process control. In *Juran's Quality Control Handbook*; 4th Ed. Juran, L.M., Gryna, F.M., Eds.; Section 24, McGraw-Hill: New York, 1988.
47. EU Guide to Good Manufacturing Practice, Draft revised version of Annex15: Validation master plan: Design qualification, installation and operational qualification, non-sterile process validation, cleaning validation.
48. Bedson, P.; Rudd, D. The development and application of guidance on EQ of analytical instruments: HPLC. *Accred. Qual. Assur.* **1999**, *4*, 50–62.

VALIDATION OF PHARMACEUTICAL PROCESSES

Robert A. Nash

Consultant, Mahwah, New Jersey, U.S.A.

INTRODUCTION

Process validation is a requirement of current Good Manufacturing Practices (GMPs) for finished pharmaceuticals (21 CFR 211) and of the GMP regulations for medical devices (21 CFR 820) and therefore applies to the manufacture of both drug products and medical devices.

According to the FDA Guidelines on General Principles of Process Validation (1), process validation is defined, "... as establishing documented evidence, which provides a high degree of assurance, that a specific process will consistently produce a product meeting its predetermined specifications and quality characteristics." The process for making a drug product consists of a series (flow diagram in logically defined steps) of unit operations (modules) that result in the manufacture of the finished pharmaceutical.

There is much confusion regarding the definition of process validation and what constitutes process validation documentation. The term validation is used here generically to cover the entire spectrum of current GMP concerns, essential most of which are; facility, equipment, component, method, and process qualification. Based upon the FDA process validation guidelines (1), the specific term should be reserved for the final stage(s) of the product and process development sequence. The essential or key steps or stages of a successfully completed development program are shown in Table 1.

The end of the development sequence, which should be assigned to formal (three-batch) process validation, derives from the fact that the specific exercise of process validation should never be designed to fail. Failure in carrying out the formal process validation assignment is often the result of incomplete or faulty understanding of the process capability, in other words, what the process can and cannot accomplish under a given set of operational requirements.

In a well-designed validation program, most of the effort should be spent on facilities, equipment, components, methods, and process qualification. In such a program, the formalized, final three-batch validation sequence provides only the necessary process validation documentation required by the FDA to show product reproducibility and a manufacturing process in a state of

control. Such a strategy is consistent with the FDA preapproval inspection program directive (2).

PROCESS VALIDATION OPTIONS

The guidelines on general principles of process validation (1) mention three options: prospective process validation (also called premarket validation), retrospective process validation, and revalidation. Actually there are four, if concurrent process validation is included.

Prospective validation is carried out prior to the distribution of a new product or an existing product made under a revised manufacturing process where such revisions may affect product specifications or quality characteristics. The prospective approach features critical step analysis in which the unit operations are challenged during the process qualification stage to determine those critical process variables that may affect overall process performance, using either worst-case analysis or a fractional-factorial design. During formal, three-batch, prospective validation, critical process variables should be set within their operating ranges and should not exceed their upper and lower control limits during process operation. Output responses should be well within finished-product specifications.

Retrospective validation is recognized in both current GMPs [21 CFR 211.110(b)] and the FDA process validation guidelines (1). It involves accumulated in-process production and final product testing and control (numerical) data to establish that the product and its manufacturing process are in a state of control. Valid in-process results should be consistent with the final specifications of the drug product and shall be derived from previous acceptable process average and process variability estimates where possible and determined by the application of suitable statistical procedures (quality control charting) where appropriate.

The retrospective validation option is chosen for established products whose manufacturing processes are considered to be stable and when, on the basis of economic considerations and resource limitations, prospective qualification and validation experimentation cannot be justified.

Table 1 Development sequence with respect to process validation

Developmental stages	Batch size
Product design	1 X
Product characterization	
Product selection	
Process design	
Product optimization	10 X
Process characterization	
Process optimization	
Process qualification	
Process qualification	100 X
Process validation	
Process certification	
Process revalidation	100 X to 1000 X

Prior to undertaking either prospective or retrospective validation, the facilities, equipment, and subsystems used in connection with the manufacturing process must be qualified in conformance with cGMP requirements.

Concurrent validation studies are carried out under a protocol during the course of normal production. The first three production-scale batches must be monitored as comprehensively as possible. The evaluation of the results is used in establishing the acceptance criteria and specifications of subsequent in-process control and final product testing. Some form of concurrent validation, using statistical process control techniques (quality control charting) may be employed throughout the product manufacturing life cycle.

Revalidation is required to ensure that changes in the process and/or in the process environment, whether introduced intentionally or unintentionally, do not adversely affect product specifications and quality characteristics (2). There should be a quality assurance program (change control) in place which requires revalidation whenever there are significant changes in formulation, equipment, process, and packaging that may impact on product and manufacturing process performance (3). Furthermore, when a change is made in a raw material supplier, the drug manufacturer should be made aware of subtle, potentially adverse differences in raw material characteristics that may adversely affect product and manufacturing process performance.

It is recommended that every requested change be reviewed by the validation or CMC committee. Such a committee should judge if a change is significant for revalidation and decide on a course of action to be taken. The following conditions require revalidation study and documentation:

1. Change in a critical component (usually refers to active pharmaceutical ingredient, key excipients, or primary packaging);
2. Change or replacement in a critical piece of modular (capital) equipment;
3. Significant change in processing conditions that may affect subsequent unit operations and product quality;
4. Change in a facility and/or plant (usually location, site, or support systems);
5. Significant increase or decrease in batch size that affects the operation of modular equipment; and
6. Sequential batches that fail to meet product and process in-process specifications.

In some situations process performance requalification studies may be required prior to undertaking specific revalidation assignments. With the exception of sterile products manufacture, periodic revalidation is not required at the present time. The performance and state of control of the product and its manufacturing process can be adequately covered during the annual product and process review. The FDA has issued an interim guidance document that addresses what constitutes major and minor formulation and manufacturing changes for immediate-release solid dosage forms (4). Such documentation and others to follow should simplify manufacturing decisions about the need to revalidate.

VALIDATION PRIORITIES

There is a basic concept with respect to which pharmaceutical processes should be given a higher priority over others. All pharmaceutical manufacturing processes require process validation documentation, but there is an accepted logical approach to priority selection, in the following order:

Sterile products and their processes

- large volume parenterals (LVPs) infusions greater than 100 ml
- small volume parenterals (SVPs) single and multiple dose injections
- ophthalmics and sterile devices

Nonsterile products and their processes

- low dose high potency tablets and capsules
- drugs with inherent stability problems
- transdermal delivery (TDD) and inhalation products
- the rest of the oral solid dosage forms
- oral liquids and topical products

The best approach to assessing problems with respect to a terminal sterilization method (i.e., moist heat, dry heat, radiation, and chemical methods) is to first establish the qualification, validation and stability of the pharmaceutical process prior to conducting a given sterilization procedure.

THE VALIDATION COMMITTEE

In most companies, the validation or Chemistry, Manufacturing and Control (CMC) committee is charged with the responsibility of establishing and operating the complete validation program for the specific manufacturing site. In some companies the program is led by a validation manager whereas in others, quality assurance personnel have taken on expanded responsibilities in this regard.

Specific process validation assignments are carried out by those with the necessary training and experience. The specifics of how the committee is organized to conduct process validation assignments is beyond the scope of this article. The responsibilities that must be carried out and the traditional organizational structures best equipped to handle each of these assignments are outlined in Table 2. Other members may include Quality Control, Regularity affairs, etc.

VALIDATION MASTER PLAN

The creation of a master plan permits the development of a logical overview of the validation effort. It lays out in a logical sequence the activities or key elements or both to be performed in accordance with the approximate time schedule in a Gantt or PERT chart format. The master plan establishes the critical path through the chart against which progress can be monitored.

The validation program starts with the design and development of raw materials and components, followed by the IQ/OQ of facilities, equipment, and systems through performance and process qualification stages, and terminates in the protocol-driven, three-batch, formal process validation

program. Most of these activities move forward in series. However, by combining activities and elements and moving in parallel, where possible, on independent tracks with respect to active pharmaceutical ingredients (APIs) analytical methods development, facilities, equipment, support systems, and the drug product design and manufacturing process development, a great deal of time can be saved before the individual elements or grouping of activities are combined prior to the formal process validation program. A Process Validation progress chart is shown in Fig. 1.

Installation Qualification (IQ)

This includes procedures and documentation to show that all important aspects of the installation of the facility, support system, or piece of modular equipment, having been properly calibrated, meet its design specifications and that the vendor’s recommendations had been suitably considered.

Operational Qualification (OQ)

Following IQ, procedures and documentation show that the facility, support system, or piece of modular equipment perform as intended throughout all anticipated operating ranges under a suitable load.

Performance Qualification (PQ)

Following IQ and OQ, actual demonstrations during the course of the validation program show that the facility, support system, or piece of modular equipment perform according to a predefined protocol and achieve process reproducibility and product acceptability.

VALIDATION PROTOCOL AND REPORT

The following validation protocol and format for the completed validation report have been suggested in the WHO Guidelines on Validation of Manufacturing Processes (TRS 823) (5).

Table 2 Composition of the process validation committee

Representative of	Function
Engineering	Qualifies for plant, facilities, equipment, and support systems
Development	Qualifies for products and their specific manufacturing processes
Manufacturing	Operates plant, facilities, equipment, support systems and the various manufacturing processes
Quality assurance	Audits plant, facilities, equipment, support systems, the various manufacturing processes, and their products

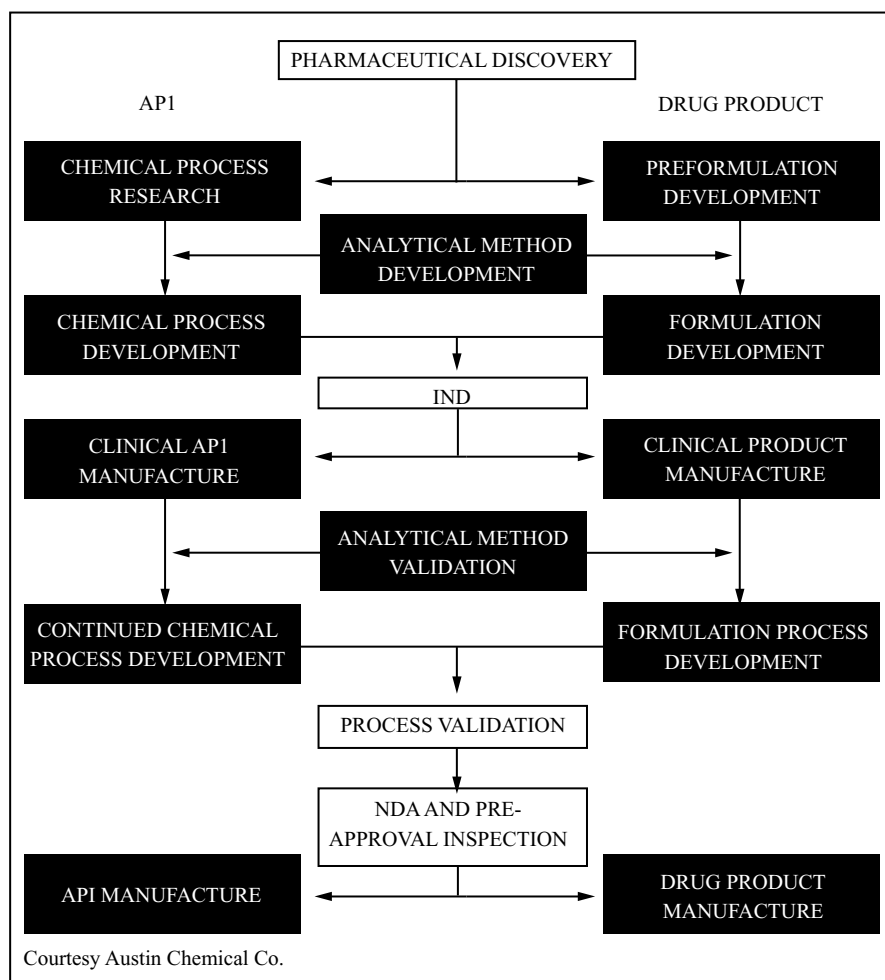


Fig. 1 Process validation progress chart.

1. Purpose (for the whole validation) and prerequisites
2. Presentation of the whole process and subprocesses including flow diagram and critical step analysis
3. Validation protocol approvals
4. Installation and operational qualifications, including blueprints or drawings
5. Qualification report(s)

- Subprocess 1
- Purpose
- Methods and procedures
- Sampling and testing procedures, release criteria
- Reporting function
- Calibration of test equipment
- Test data
- Summary of results
- Approval and requalification procedure
- Subprocess 2 (repeat)

6. Product qualification, test data from prevalidation batches
7. Product validation, test data from three formal validation batches
8. Evaluation and recommendations (include revalidation and requalification requirements)
9. Certification (approval)
10. Summary report with conclusions

The validation protocol and report may also include copies of the product stability report or its summary as well as validation documentation on cleaning and analytical methods.

PREAPPROVAL INSPECTION

The FDA Preapproval Inspection Program (2) is designed to provide a basis for determining the adequacy and

accuracy of reported and factual information in New Drug Application (NDA) and Abbreviated New Drug Application (ANDA) submissions with respect to the suitability of cGMP product development, analytical laboratory, and manufacturing facilities.

A preapproval inspection checklist should include the following documentation which may be required prior to the formal inspection:

1. API development and validation report(s) including impurity profile and polymorphic forms;
2. Pharmaceutical (dosage form) development report;
3. Stability and clinical batch records and history, including phase-III program;
4. Data for API and key excipients used in the manufacture of clinical and biobatches;
5. Bioequivalency report;
6. Technical transfer report (development to manufacturing/QA/QC);
7. Copy of the CMC section of the NDA including information on suppliers and vendors;
8. Copy of proposed production monograph and master batch record;
9. Equipment validation report establishing IQ and OQ;
10. Cleaning validation report;
11. Analytical methods validation and computer systems validation reports;
12. Stability report establishing expiry dating; and
13. Process validation protocol for formal three-batch validation of production-size batches.

During preapproval inspection, the FDA accepts a process-validation protocol based on the company's commitment to complete successfully three production-size validation batches prior to product launch. In some situations a prevalidation (process demonstration qualification) production-size batch is completed before the entire formal three-batch program is carried out.

PILOT SCALE-UP AND TECHNOLOGY TRANSFER

The pilot-production program may be carried out as a shared responsibility between the development laboratories

and their appropriate manufacturing counterpart or as a process demonstration by a separate, designated pilot-plant or process development department. Supporting technology transfer documentation applies to both the specific process and system being qualified and validated and the related testing standards and testing methods. The formal technology transfer is normally made from the development laboratories or the process development pilot-plant to pharmaceutical production function.

In actuality, a number of technology transfer points and documents are generated as prospective validation proceeds through the various stages of product development. These stages of technology transfer in terms of scale-up are illustrated in Fig. 2.

Solid pharmaceutical dosage forms (tablets and capsules) are used to illustrate the various stages of product and process development. These principles and practices also apply in a general way to the development of liquid and semisolid pharmaceutical dosage forms (not discussed here).

STAGES OF VALIDATION

Elements of the validation concept should be incorporated during each of the various stages of the product and process development continuum. These stages can be summarized as follows.

Stage Preformulation Studies: APIs plus key excipients

- Stage I Product design and development
- Stage II Preparation of clinical and biobatches
- Stage III Process scale-up and evaluation
- Stage IV Formal process validation

Preformulation Studies: API

Preformulation testing of the specific API of interest and key excipients to be used in the product design stage, alone and in combinations with the API, should be included as a preliminary first step in the product and process development sequence. A simple check list of items

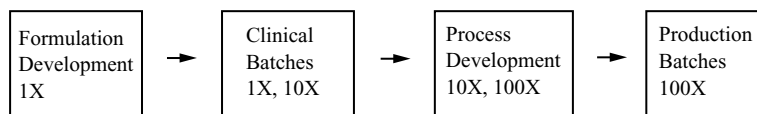


Fig. 2 Technology transfer stages.

worth consideration in preformulation studies with APIs and important or critical excipients is provided as follows:

API

- Key excipients
 - Fillers and diluents
 - Binders
 - Disintegrants
 - Glidants and lubricants
- Establish chemical and physical compatibility
- Minimize lot-to-lot variability in properties
- Worldwide availability from comparable suppliers
- Properties for possible evaluation
 - Color, odor, taste, solubility;
 - Particle morphology (DSC, TGA, x-ray diffraction);
 - Particle size distribution and surface area;
 - Crystal and bulk density, compaction index;
 - Angle of repose and flowability index;
 - Spectrophotometry (UV, FTIR, NMR, OR);
 - Water content, LOD, moisture uptake;
 - Microbial limits and heavy metals;
 - HPLC assay and impurity profile.

Before preformulation studies are undertaken, two-way technical communication between the manufacturers of the API (laboratory and plant) and the pharmaceutical product development laboratories must be established (see Fig. 1). It should start early and be maintained through-out the product and process development program.

In addition to potency, purity, and stability considerations of the API, the product development department is especially interested in the chemical and physical form (free acid or base, salts, esters, amides, polymorphs, solvates, particle size and shape) of the API. Time spent early in the cycle in establishing these particular factors often aids and/or simplifies the subsequent product and process development program.

Not every subject shown above must be tested or addressed. However, aspect, particle morphology and size, compaction and flowability, water content, spectrophotometric and chromatographic data should be studied and monitored throughout the product and process development program (6, 7).

Because key excipients are well established in most new product and process development programs, the same degree of preformulation scrutiny is often not required. Compatibility studies with the API, however, should be performed to study possible untoward interactions between the active ingredients and the excipients. It should be kept in mind that small or minor changes in

physical and possibly chemical properties upon intimate contact in binary studies with key excipients should not automatically exclude a favored excipient without further critical testing.

Stage I: Product Design and Development

Following successful preformulation studies, the API is transferred to the formulations laboratory for preliminary product design and development studies. In most cases, the drug is mixed with an appropriate diluent or filler and glidant combination and filled into two-piece opaque hard-shell capsules for preliminary stability and subsequent phase I clinical studies versus matching placebo capsules (8). At or about the same time, initial studies of a prototype tablet formulation should be started. The key steps in the product design and development sequence are given below.

Stage I: Product Design: 1× Laboratory Scale (1–10 kg)

- Hard-shell capsule (phase I clinical trials) followed by prototype tablet dosage form
 - Direct compression versus wet granulation
 - Maximize chemical and physical stability
 - Minimize product and process costs
 - Product characterization
 - Product selection
 - Process design
- Excipients are selected among the following categories:
 - Binder, diluents, and disintegrants including alginates, calcium phosphate, cellulose, dextrates, gelatin, povidones, starch and derivatives, sorbitol, sucrose, and derivatives. Glidants and lubricants including colloidal silicon dioxide, hydrogenated vegetable oil, mineral oil, PEG, silica gel, sodium lauryl sulfate, stearates, talc.

Although the work is conducted in the research or formulations laboratory using small-scale processing equipment, it is important to gain early experience with colorant systems that have been selected for the finished tablet product; color aids in blend-uniformity evaluation.

In addition to excipient screening and selection, it is important to gauge processing parameters that are more fully explored during the scale-up phases. These processing factors include flowability, compaction and compressibility of powders and granules, content uniformity of powder and granule blends and finished tablets, moisture uptake, in vitro dissolution release profiles, and subsequent full-scale stability testing.

Products used in human clinical trials must, of course, conform to good laboratory, good clinical, and good manufacturing practice requirements (1, 9).

Stage II: Process Development: Pilot Laboratory (Clinical)

After the (1X) “go” laboratory batch has been determined to be both physically and chemical stable, based upon accelerated, elevated temperature testing (1 month at 45°C or 3 month at 40°C and 80% relative humidity) the next step (stage II) is to scale the product and its process to (10X) pilot-laboratory size batch(es). This batch represents the first replicated scale-up of the designated formula. Its size usually ranges between 10 and 100 kg, 10 and 100 L, or 10,000 to 100,000 units. Often these pilot-laboratory batches are used in clinical trials and bioequivalency studies. According to the FDA, the minimum requirement for a biobatch is 100,000 units (10).

Pilot-laboratory batches are usually prepared in small pilot equipment within a designated current GMP approved facility. The number and size of these pilot-laboratory batches may vary, depending on one or more of the following factors:

- Equipment availability
- API availability
- Cost of raw materials
- Inventory requirements for both clinical and nonclinical studies

Process development (process qualification) or process capability studies are normally started in this important stage II of the scale-up sequence. The scope of stage-II process development consists essentially of product optimization and process characterization studies.

Product Optimization:

- Establish formula rationale and boundary conditions for API and excipients

Process Characterization:

- Define unit operations, process variables, and response parameters.
 - Define critical process variables and response parameters using simple experimental designs.
 - Establish provisional control limits for critical process variables and their response parameters based on process replication.
- Maintain product stability.

Unit operations for solid dosage-form development include:

- Granulation
- Drying
- Sizing
- Blending and mixing
- Encapsulation and/or tablet compression
- Coating
- Filling and packaging

Unit operations are selected for the development of a tablet (coated or noncoated) or capsule (hard shell or soft-gel) process (11). Unit operations that are considered to be critical are determined through analysis of the process variables and their respective measured response for each unit operation (Table 3) (12–14).

In order to determine critical control parameters and their unit operations, constraint analysis techniques (15) followed by fractional factorial designs (Table 4) are used to challenge the tentative control limits (so-called worst-case analysis) established for the process at this intermediate stage. Time and effort spent to qualify the process at the 10X stage often simplifies the work that follows during stages III and IV.

Von Doehren et al. (13) and Chowhan (16) have described the various stages of solid dosage form process development as it relates to technology transfer and process validation. Their respective approaches to the topic have been integrated in this article.

Fahrner (17) raises the following issues regarding the new role for pilot plants in product development.

- Too much time is devoted to preliminary or applied research and not enough to the proper development of the process.
- Often a suitable manufacturing strategy is lacking during the early phases of the program, which results in poorly planned technology transfer and an inappropriate division of responsibility with respect to the overall program.
- Most laboratory processes are rarely scalable, since piloting is a scaled-down version of manufacturing not a scaled-up version of the laboratory batch.

Fahrner makes the case for a separate pilot facility (process development function) to bridge the communication gap between R & D and production.

Stage-III: Pilot Production

The technology transfer of the product and process from the traditional product development function to a separate process development (pilot plant) function or production

Table 3 Control parameters for solid-dosage-form development

Unit Operation	Process variables (X) ^a	Measured responses (Y) ^b
Granulation (Power type)	Load Speed (main chopper) Liquid addition rate Granulation time	Power consumption
Drying	Load Inlet temperature Air-flow rate Drying time	Moisture content Bulk density
Sizing (Screening)	Load Screen size Speed Feed rate	Particle size distr. Bulk density
Blending (Mixing)	Load Speed Mixing time	Blend uniformity
Encapsulation	Fill volume Tamper setting Speed Glidant (type, amount)	Capsule weight Moisture content Dissolution Content uniformity Potency
Tablet compression	Press speed Feed rate Precompression force Compression force	Tablet weight Moisture content Hardness/friability Thickness Dissolution/disintegration Content uniformity Potency
Coating (Film type)	Load Pan speed Spray rate Air flow	Weight gain

^a7–23 possible variables.^b11–16 possible responses.

itself is normally carried out at the (100×) pilot-production batch stage (100–1000 kg):

- Full-scale production batch
- For possible future commercial or clinical use
- Evaluate critical process parameters; product and process are scaled to another order of magnitude (100×)
- Process optimization
 - Mixing and blending times
 - Drying times
 - Milling operations
 - Press speed, compression force
 - Encapsulation speed, tamping settings
 - Speed, air flow, spray settings, temperature

- Process qualification (prevalidation batches); determine process capability, challenge in-process control limits
- Maintain product stability

The creation of a separate pilot plant or process development unit has been favored in recent years because it is ideally suited to carry out key process qualification and/or process validation studies in a timely manner (18, 19).

The objective of the pilot-production batch is to scale the product and its process by another order of magnitude (100×). For most solid dosage forms it represents a full production scale batch, in standard equipment. The technology transfer documents should include the technical information normally required for preapproval inspection:

Table 4 Fractional factorial design for process development^a

Trials	Key variables ^b							Sums
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	
1	—	—	—	—	—	—	—	0/7
2	—	—	—	+	—	—	—	1/6
3	—	—	+	—	—	+	—	2/5
4	+	+	—	—	+	—	—	3/4
5	+	+	—	—	—	+	+	4/3
6	+	—	+	+	+	—	+	5/2
7	—	+	+	+	+	+	+	6/1
8	+	+	+	+	+	+	+	7/0
Sums	4/4	4/4	4/4	4/4	4/4	4/4	4/4	28/28

^aAdapted from Hendrix, C. D., What every technologist should know about experimental design, *CHEMTECH* (March 1979).

^bKey variables are randomly assigned an “X” value.

1. Preformulation information
2. Product development report
3. Product stability report
4. Analytical methods report
5. Proposed manufacturing formula, manufacturing instruction, in-process and final product specifications at the 100X-batch size

The objectives of prevalidation trials at stage III (100X pilot production) is to qualify and optimize the process in full-scale production equipment and their facilities.

Rushing through the first (100X) pilot-production batch in order to proceed with formal validation should be discouraged. Small problems that often arise during (100X) scale-up should be addressed immediately and not ignored. Such problems are often best addressed by returning to the laboratories (10X) for supplemental process characterization and qualification studies.

Many companies, however, proceed directly to three-batch formal validation without stage III prevalidation work and often complete formal trials prior to preapproval inspection. The downside of this alternative strategy is that finished production batches often remain in the warehouse beyond their approved expiry dating period.

When faced with a choice of strategies, there is no one ideal way of completing the pilot scale-up and validation sequence other than depending on prior experience with related products and their processes.

Stage-IV: Formal Process Validation

In the normal course of events and following a successfully completed preapproval inspection, formal, three-batch

process validation is carried out in accordance with the protocol approved during the preapproval inspection. The primary objective of the formal process validation exercise is to establish process reproducibility and consistency. The program is not designed to challenge upper and lower control limits (so-called worst-case analysis) of critical process variables. Such upper and lower control limit challenging is normally conducted during the stage II (10X size) process characterization, optimization, and qualification program, using suitable and reasonable experimental designs (Table 4).

The documentation to be established before, during, and after formal process validation is shown below. The protocols and the subsequent formal validation studies are designed to establish uniformity among the three batches with respect to granulation, blending, finished tablet, and finished capsule stages (1, 2, 10).

100X production batches

- Complete product development program and report
- Prepare protocol for prospective process validation
- Complete preapproval inspection requirements; conduct three-batch formal process validation, establish reproducibility for mixing, blending, and compression or encapsulation operations
- Establish process documentation
 - Preformulation report
 - Analytical methods validation report
 - IQ/OQ and cleaning validation reports
 - Formula development report
 - Process feasibility report
 - Manufacturing bioequivalency report
 - Product development report
 - Process validation protocol
 - Process validation report
 - Product stability report

In that respect, the following test data and results are used to show process reproducibility and consistency among validation batches: particle or granule size distribution, bulk density, moisture content, hardness, thickness, friability, weight uniformity, potency uniformity, disintegration–dissolution profile, and product stability. Not every one of these categories have to be addressed nor followed both during in-process and final product testing. Nevertheless, testing must be sufficient to establish process reproducibility and demonstrate, with a high degree of certainty, that the product and process are in a state of control.

Whenever possible, formal validation studies should continue through packaging and labeling operations (whole or in-part), so that machinability and stability of

the finished product can be established and documented in the primary container–closure system.

Recently the FDA, with the cooperation of the Pharmaceutical Industry has developed a series of guidance procedures to speed approval of post approval changes with or without process scale-up changes. At the present time the SUPAC (scale-up post approval change) program covers the following product categories:

- Immediate Release (IR) solid dosage forms
- Extended Release (ER) solid dosage forms
- Delayed Release (MR) solid dosage forms
- Semisolids (SS) dosage forms
- API changes presently called BACPAC
- Packaging changes called PACPAC
- Analytical Methods changes called AMPAC
- Sterile Aqueous Solutions called SUPAC-SAS

If the program is successful, other dosage form categories will be added later.

The following changes are covered:

- Components and Composition Changes
- Manufacturing Equipment and Process Changes
- Batch Size (scale-up) Changes
- Manufacturing Site Changes

The program consists of three levels:

- Level 1 or minor changes that are made without FDA approval and reported in the Annual Report (AR).
- Level 2 or intermediate changes that may be instituted by first filing a Change Being Effected (CBE) Supplement with the FDA and waiting 30 days for a reply before instituting the change.
- Level 3 or major change in which a prior approval supplement (PAS) is filed with the FDA and approval must be obtained from the FDA before proceeding with the change.

This new program is off to a good start and is intermittently involved with the need for adequate process validation studies and documentation to support the changes requested.

CHANGE CONTROL

Procedures with respect to establishing change control should be in place before, during, and after the completion of the formal validation program. A change control system maintains a sense of functionality as the process evolves and provides the necessary documentation trail that ensures that the process continues in a validated,

operational state, even when small noncritical adjustments and changes have been made. Such minor, noncritical changes in materials, methods, and machines should be reviewed by the validation committee (development, engineering, production, and QA/QC) to ensure that process integrity and comparability have been maintained and documented before the specific change that has been requested can be approved by the head of the quality control unit.

The change control system, based upon an approved standard operating procedure(s) (SOPs), takes on added importance as the vehicle or instrument through which innovation and process improvements can be made more easily and more flexibly without prior formal review on the part of the NDA and ANDA reviewing function of the FDA. If more of the supplemental procedures with respect to the chemistry and manufacturing control sections of NDAs and ANDAs could be covered through annual review documentation procedures, with appropriate safeguards, process validation will become more innovative (20–22).

OUT-OF-SPECIFICATIONS

Probably the single most important technical issue facing the pharmaceutical industry at the present time is the question: what constitutes process or batch failure in terms of an out-of-specification (OOS) assay value? The concept of product and/or process failure appears twice in the cGMPs (3).

According to 21 CFR Sect. 211.165(f), “Drug products failing to meet established standards or specifications and other relevant quality control criteria shall be rejected.” In CFR Sect. 211.192 it is stated:

Any unexplained discrepancy (including a percentage of theoretical yield exceeding the maximum or minimum percentages established in master production and control records) or the failure of a batch or any of its components to meet any of its specifications shall be thoroughly investigated [regardless] whether the batch has already been distributed. The investigation shall extend to other batches of the same drug product and other drug products that may have been associated with the specific failure or discrepancy. A written record of the investigation shall be made and shall include the conclusions and follow-up.

The key to establishing product and/or process failure is to verify the accuracy, relevance, and reproducibility of

deviant assay value(s), test result(s), and recorded number(s) that are reported (23). All companies should have SOPs in place that cover first the verification of deviant numbers in the quality control laboratories and, following that investigation and a report showing the test result to be deviant, a second set of SOPs covering follow-up actions taken as described in the following steps:

1. A written procedure for full investigation when there is not a verified laboratory error.
2. Scientific criteria for retesting and resampling during the formal investigation.
3. Description and results of the formal investigation into possible causes of the OOS result(s).
4. Results of all testing involved during the investigation.
5. A scientific basis and justification for discarding any OOS test result and accepting the batch in question.
6. Final determination of conformity to appropriate specifications and justification of the actions taken, and
7. Signature of individual(s) responsible for final decision(s) and the action(s) taken.

Even though the responsibility for batch acceptance or rejection lies with the head of the quality control unit, the help of the validation committee should prove useful in reviewing the process of OOS investigation and arriving at a recommendation for action taken.

CLEANING VALIDATION

According to 21 CFR Sect. 211.67, Equipment Cleaning and Maintenance of cGMP regulations (3), equipment and utensils should be cleaned, maintained, and sanitized at appropriate intervals to prevent malfunction or contamination that would alter the safety, identity, strength, quality, or purity of the drug product. Written procedures shall be established and followed for cleaning and maintenance of equipment. These procedures shall include, but are not limited to the assignment of responsibility for cleaning and maintaining equipment; maintenance, cleaning, and sanitizing schedules where appropriate; description in sufficient detail of methods, equipment, and materials used in cleaning and maintenance operations, and the methods of disassembling and reassembling equipment as necessary to assure proper cleaning and maintenance; removal or obliteration of previous batch identification, protection of clean equipment from contamination prior to use; and inspection of equipment for cleanliness immediately before use.

Records shall be kept of maintenance, cleaning, sanitizing, and inspection.

The objective of cleaning validation of equipment and utensils is to reduce the residues of one product below established limits so that the residue of the previous product does not affect the quality and safety of the subsequent product manufactured in the same equipment.

According to 21 CFR Sect. 211.63, Equipment Design, Size, and Location, of cGMP regulations (3), equipment used in the manufacture, processing, packing, or holding of a drug product shall be of appropriate design, adequate size, and suitably located to facilitate operations for its intended use and for its cleaning and maintenance. Some of the equipment design considerations include type of surface to be cleaned (stainless steel, glass, plastic), use of disposables or dedicated equipment and utensils (bags, filters, etc.), of stationary equipment (tanks, mixers, centrifuges, presses, etc.), of special features (clean-in-place systems, steam-in-place systems), and identifying the difficult-to-clean locations on the equipment (so-called hot spots or critical sites).

The specific cleaning procedure should define the amounts and the specific type of cleaning agents and/or solvents used. The cleaning procedure should give full details as to what is to be cleaned and how it is to be cleaned. The cleaning method should focus on worst-case conditions, such as highest-strength, least-soluble, most difficult to clean formulations. Cleaning procedures should identify the time between processing and cleaning, cleaning sequence, equipment dismantling procedure, need for visual inspection, and provisions for documentation.

The choice of a particular analytical method (HPLC, TLC, spectrophotometric, total organic carbon (TOC), pH, conductivity, gravimetric, etc.) and sampling technique chosen (direct surface by swabs and gauze, or by rinsing) depends on the residue limit to be established, based upon the sampling site, type of residue sought, and equipment configuration (critical sites vs. large surface area) considerations. The analytical and sampling methods should be challenged in terms of specificity, sensitivity, and recovery.

The established residue limits must be practical, achievable, verifiable, and assure safety. The potency of selected drug and presence of degradation products, cleaning agents, particulates and microorganisms should be taken into consideration.

The following residue limits have been suggested: not more than (NMT) 10 ppm or NMT 0.001% of the dose of any product appears in the maximum daily dose of another product and no residue visible on the equipment after cleaning procedures have been performed.

PHARMACEUTICAL EXCIPIENTS

Pharmaceutical Excipients are components of finished drug products and site active components are recognized in the USP/NF. A list of key excipients for solid dosage forms is given:

Diluents, Fillers, and Binders:

- Calcium phosphate, dibasic
- Dextrates
- Dextrin
- Lactose (anhydrous, fast flow)
- Mannitol
- Starch (corn)
- Sugar, compressible

Tablet Disintegrants:

- Cellulose, microcrystalline
- Croscarmellose sodium
- Crospovidone
- Sodium starch glycolate
- Starch, pregelatinized

Tablet and Capsule Lubricants:

- Magnesium stearate
- Mineral oil, light
- Polyethylene glycol
- Sodium stearyl fumarate
- Stearic acid, purified
- Talc
- Vegetable oil, hydrogenated

Other Excipients:

- Carboxymethylcellulose sodium
- Cellulose acetate phthalate
- Ethylcellulose
- Hydroxypropyl cellulose
- Hydroxypropyl methylcellulose
- Hydroxypropyl methylcellulose pathalate
- Methacrylic acid copolymer
- Polysorbates
- Polyvinyl acetate phthalate
- Povidone
- Sodium lauryl sulfate

ACTIVE PHARMACEUTICAL INGREDIENTS (API)

A chemical is considered to be an API if it is intended for medicinal purposes. Regulatory agencies however, place greater emphasis and priority on the manufacture and validation of APIs than UPS/NF excipients.

According to Sect. 501 (a)(2)(b) of the Food, Drug, and Cosmetic (FD&C) Act, all drugs must be manufactured, processed, packed, and held in accordance with cGMPs. No distinction is made between APIs and finished drug products.

Elements common to both APIs and finished drug products include facilities and equipment qualification (IQ/OQ, and PQ), cleaning validation, validation of water supplies, microbial limits for nonsterile material, manufacture of sterile and pyrogen-free material, in-process blending and mixing, analytical methods validation, laboratory controls and in-process testing, change control procedures and revalidation, reprocessing, packaging and labeling, and stability testing.

Process

There are four primary processes used in the manufacture of APIs. They are chemical synthesis, fermentation, extraction, and purification. A flow diagram (Fig. 3) and a description of the chemistry involved are helpful in defining the process. The process description should include appropriate parameters, such as charging quantities or volumes of reactants or solvents, reaction times, temperatures, pressures, etc. Critical processing steps and critical operating parameters should be maintained to ensure batch-to-batch consistency, product yield, and quality.

Where in the chain of unit operations (chemical process) does API validation start? As long as key intermediates are made in the plant, they and their reaction and processing steps should be subjected to an appropriate cGMP and process qualification-validation program. A key intermediate is defined as an intermediate in which an essential molecular characteristic, usually related to stereochemical configuration, is introduced into the final API structure (moeity).

Physical Characteristics

Besides purity (chemical potency), the physical characteristics and properties of the API are extremely important to the end user (drug product manufacturer). Characteristics such as crystal morphology, particle size and shape, bulk density, melting point, optical rotation, etc., have a profound effect upon the final drug product and its performance and stability. In addition to the reaction or extraction step, crystallization, milling, and blending unit operations must be subject to qualification and validation.

Impurity Profile

The USP permits up to 2% of ordinary nontoxic impurities. However, impurities above 0.1% should be

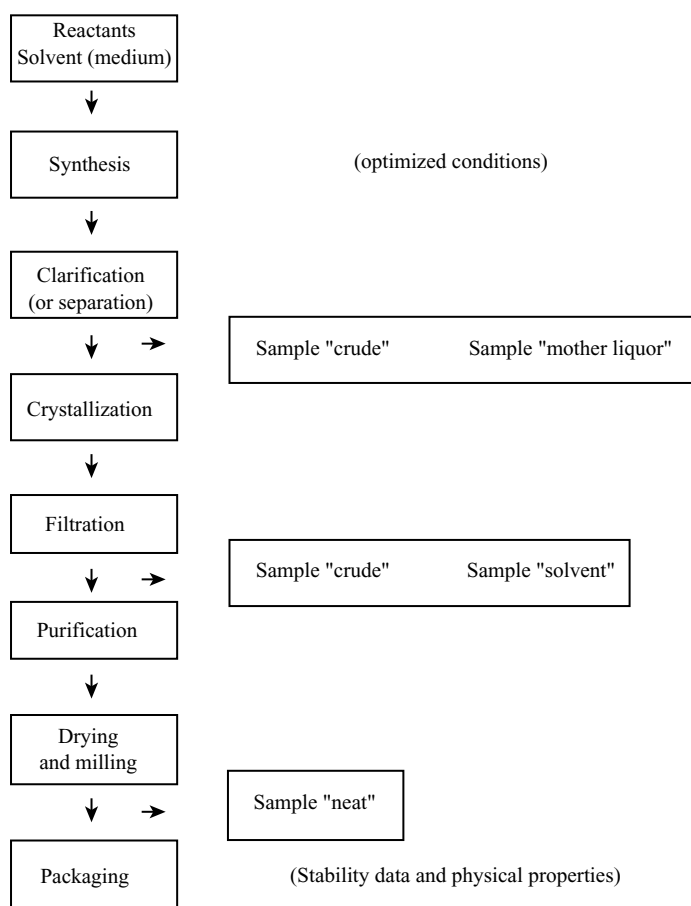


Fig. 3 Manufacture of active pharmaceutical ingredients (APIs).

fully characterized and quantified. Impurities may include starting materials, by-products, intermediates, degradation products, reagents, catalysts, heavy metals, electrolytes, filter aids, and residual solvents. Known toxic impurities must be held to a tighter standard, i.e., below 0.1%.

The ISO quality standards series

The ISO 9000 Series was developed in 1987 by the International Organization for Standardization (ISO) in Geneva, Switzerland. It is a comprehensive set of management standards governing the operation of quality assurance to help develop and document a quality system that is useful for individual companies.

The ISO 9000 Quality Management and Quality Assurance Standards, Guidelines for Selection and Use, provide basic definitions and concepts and explain how to use the rest of the series (9001, 9002, 9003, and 9004).

The ISO 9001 Quality System—Model for Quality Assurance in Design and Development provides for quality assurance in the areas of design, installation, servicing,

development, and production. It is useful primarily for companies that design and develop their own products.

The ISO 9002 Quality System—Model for Quality Assurance in Production, Installation and Service applies to manufacturers, distributors, and service vendors whose products have been designed and serviced by a subcontractor. Such companies are exempt from design control requirements. Both ISO 9001 and ISO 9002 are directly applicable to cGMPs. The connection between the two independent systems is shown in Table 5. Except for language, shades of meaning, and stresses the documents are similar.

The ISO 9003 Quality Systems—Model for Quality Assurance in Inspection and Testing, designed for testing laboratories and equipment distributors only requires conformance to final inspection and testing procedures.

The ISO 9004 Quality Management and Quality Systems Elements—Guidelines provide standards and guidelines for quality management planning and implementation.

Table 5 Meeting quality standards

ISO 9000 series		cGMPs
1.	Management responsibility	Organization and personnel
2.	Quality system	Organization and personnel, laboratory, controls, records, reports
3.	Contract review	Holding and distribution
4.	Design control	Production and process controls
5.	Document control	Records and reports
6.	Purchase products	Control of components, drug product containers, closures
7.	Supplied product	Holding and distribution
8.	Product identification, traceability	Packaging and labeling control
9.	Process control	Production and process control
10.	Procedures for inspection, testing	Laboratory controls
11.	Inspection, measuring, testing equipment	Laboratory controls production, and process controls
12.	Inspection, testing status	Laboratory controls
13.	Control of nonconforming products	Returned and salvaged drug products
14.	Corrective action	Records and reports
15.	Handling, storage, packaging, delivery	Packaging and labeling control holding, distribution
16.	Quality records	Records and reports
17.	Internal quality audits	Laboratory controls
18.	Training needs	Organization and personnel
19.	Servicing procedures	Equipment
20.	Statistical techniques	Building, facilities production and process controls

REFERENCES

1. Berry, I.R.; Harpaz, D. *Validation of Active Pharmaceutical Ingredients*; Interpharm Press, Inc.: Buffalo Grove, IL, 2001.
2. Validation of Pharmaceutical Processes, Sterile Products. Carleton, F.J., Agalloco, J.P., Eds.; Marcel Dekker, Inc.: New York, 1998.
3. Chow, S.-C., Liu, J.-P., Eds. *Validation, Process Controls and Stability, Statistical Design and Analysis in Pharmaceutical Science*; Marcel Dekker, Inc.: New York, 1995.
4. Gibson, W.; Powell-Evans, K. *Validation Fundamentals*; Interpharm Press, Inc.: Buffalo Grove, IL, 1998.
5. Sucker, H., Ed. *Validation in Practice*; Wissenschaftliche Verlagsgesellschaft GmbH: Stuttgart, 1983.
6. Validation of Manufacturing Processes Fourth European Seminar on Quality Control, Geneva, Sept. 25 1980.
7. FDA. Guideline on General Principles of Process Validation, FDA, Rockville, MD, May 11, 1987.
8. FDA. Pre-Approval Inspection/Investigations Guidance Manual, FDA, Rockville, MD, Oct. 1, 1990; 7346.832.
9. Fed. Reg. 43(190) Human and Veterinary Drugs, Current Good Manufacturing Practice in Manufacturing, Processing, Packing or Holding, Sept. 29 1978; 45,014–45,089.
10. Lucisano, L.J.; Franz, R.M. FDA Proposed Guidance for CMC Changes: A Review and Industrial Perspective. *Pharm. Technol.* **1995**, *19* (5), 30–40.
11. WHO. Good Manufacturing Practices for Pharmaceutical Products—Guidelines on the Validation of Manufacturing Processes, WHO, Geneva, 1993.
12. Skelly, J.P. Scale-Up of Immediate-Release Oral Solid Dosage Forms. *Pharm. Res.* **1993**, *10*, 313–316.
13. Skelly, J.P. Scale-Up of Oral Extended-Release Dosage Forms. *Pharm. Res.* **1993**, *10*, 1800–1804.
14. Bolton, S. Process Validation for Hard Gelatin Capsules. *Drug Cosm. Ind.* **1984**, *134* (3), 42–48–85–87.
15. Avallone, H.L. Development and Scale-Up of Pharmaceuticals. *Pharm. Eng.* **1990**, *10* (4), 38–41.
16. FDA, Guide to Inspections of Oral Solid Dosage Forms Pre/Post Approval Issues for Development and Validation, FDA, Rockville, MD, Jan 1994.
17. Berry, I.R. Process Validation for Soft Gelatin Capsules. *Drug Cosm. Ind.* **1984**, *134* (4), 26–35.
18. Nash, R.A. Process Validation for Solid Dosage Forms. *Pharm. Technol.* **1979**, *3* (6), 105–107.
19. von Doehren, P.J.; St. John, Forbe F.; Shively, C.D. An Approach to the Characterization and Technology Transfer of Solid Dosage Form Processes. *Pharm. Technol.* **1982**, *6* (9), 139–156.
20. Sucker, H., Ed. *Validation in Practice*; Wissenschaftliche Verlagsgesellschaft GmbH: Stuttgart, 1983.
21. Berry, I.R., Nash, R.A., Eds. *Pharmaceutical Process Validation*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1993; Revised and Expanded.
22. Chowhan, Z.T. Development of a New Drug Substance into a Compact Tablet. *Pharm. Technol.* **1992**, *16* (9), 58–67.
23. Fahrner, R. New Role for Pilot Plants in Product Development. *Biopharm* **1993**, *6* (3), 34–37.
24. Avallone, H.L.; D'Eramo, P. Scale-Up and Validation of ANDA/NDA Products. *Pharm. Eng.* **1992**, *12* (6), 36–39.
25. Bala, G. An Integrated Approach to Process Validation. *Pharm. Eng.* **1994**, *14* (3), 57–64.
26. Chapman, K.G. A History of Validation in the United States, Part I. *Pharm. Technol.* **1991**, *15* (10), 82–96.
27. Akers, J. Simplifying and Improving Process Validation. *J. Parent. Sci. Technol.* **1993**, *47*, 281–284.
28. Tomamichel, K. Pharmaceutical Quality Assurance: Basics of Validation. *Swiss Pharma* **1994**, *16* (3), 13–23.
29. Bolton, S. When is it Appropriate to Average and its Relationship to the Barr Decision. *Clin. Res. Reg. Affairs* **1994**, *11*, 171–179.

VETERINARY DOSAGE FORMS

J. Desmond Baggot

Monash University, Parkville, Victoria, Australia

INTRODUCTION

Veterinary dosage forms are drug preparations designed for use in, or topical application to, one or more species of domestic animal and/or other species of veterinary interest. Although the majority of veterinary dosage forms contain the same drugs as human dosage forms, some veterinary preparations contain drugs that are not widely used in humans. Examples include benzimidazole anthelmintics, macrolide endectocides, salicylanilide flukicides, synthetic pyrethroids, chloramphenicol derivatives, α_2 -adrenoceptor agonists (sedative–analgesics), and antagonists. The converse applies to some classes of pharmacological agent (e.g., benzodiazepine derivatives, tricyclic antidepressants) because of limited clinical indications for use in animals. Veterinary pharmacology differs from human pharmacology both in the diversity of species of interest and in the emphasis placed on the various classes of drug. Some types of dosage forms are suitable for use in humans and certain animal species. They include parenteral solutions (although the concentration of the drug, the nature of the vehicle, and other constituents of the preparation must be considered); conventional tablets and capsules (amount of the drug relative to dosage requirement must be considered); oral solutions and suspensions (pediatric preparations are generally appropriate for administration to small animals, especially cats); and conventional ophthalmic preparations. Sustained-release tablets and controlled-release transdermal drug-delivery systems designed for use in humans could be used in medium to large breeds of dog. It is because of the unique physiological characteristics of each species, or closely related group of species, and the variation among species in the dose–effect relationship of pharmacological agents that veterinary dosage forms of drugs are required. Veterinary dosage forms designed specifically for use in certain animals species include long-acting parenteral dosage forms of antimicrobial agents for intramuscular injection, oral pastes containing antimicrobial agents or anthelmintics for horses, granules containing nonsteroidal anti-inflammatory drugs or anthelmintics for addition to the feed for horses or pigs, respectively, modified-release ruminal boluses containing anthelmintics for cattle or sheep,

intramammary antimicrobial preparations for cows, preparations containing ectoparasiticides for topical application to various species, and darts containing sedative drugs for capture and restraint of exotic animal species. The avoidance of drug residues in tissues and animal products (milk and eggs) is a regulatory requirement of veterinary preparations administered to food-producing species.

VETERINARY DRUGS

Although the majority of drugs available as veterinary dosage forms were initially developed for use in humans, based on experimental findings in laboratory animals, some drugs have been developed specifically for veterinary use. Anthelmintics and ectoparasiticides/insecticides are primarily veterinary drugs. Some anthelmintics (e.g., ivermectin, mebendazole, albendazole, pyrantel, piperazine, levamisole, praziquantel, bithionol) have been adopted for the treatment of parasitic infections in humans. Most ectoparasiticides are exclusively veterinary drugs. Some antimicrobial agents within certain classes, which include sulfonamides, fluoroquinolones, macrolides, chloramphenicol derivatives, and carboxylic ionophores, are available only for use in animals. Relatively few pharmacological agents were initially developed as veterinary drugs, although some drugs have indications for use in animals that do not apply to humans (Table 1). Other pharmacological agents have largely become veterinary drugs by virtue of their clinical efficacy in animals and replacement by alternative drugs for use in humans. Examples include phenylbutazone, quinidine, phenobarbital, and thiopental.

Steroid (sex) hormones, gonadotrophins, gonadotrophin-releasing hormones (synthetic forms and GnRH analog), and synthetic prostaglandins ($F_{2\alpha}$ type or analog) are used in female animals (cows, ewes, sows, and mares) to regulate various stages of the estrous cycle, whereas synthetic prostaglandins may be used to induce parturition. Melatonin, a modified-release dosage form implanted subcutaneously behind the ear of ewes, is used to stimulate the onset of cyclical ovarian activity. Progestogens are used to synchronize estrus in groups of animals or to enable

Table 1 Some pharmacological agents that are used exclusively in animals

Drug	Classification	Clinical indications
Xylazine	α_2 , (α_1)-adrenoceptor agonist	Sedation; analgesia preanesthetic medication
Yohimbine	α_2 , (α_1)-adrenoceptor antagonist agonist	Xylazine reversal
Detomidine	α_2 -adrenoceptor agonist	Preanesthetic medication
Atipamezole	α_2 , (α_1)-adrenoceptor antagonist	Detomidine reversal
Acepromazine	Phenothiazine-derivative tranquilizer	Preanesthetic medication; sedation
Etorphine–Acepromazine	Potent opioid agonist-PTZ-derivative tranquilizer	Neuroleptanalgesia
Diprenorphine	Opioid antagonist	Etorphine reversal
Droperidol–Fentanyl	Butyrophenon μ -opioid agonist	Neuroleptanalgesia
Azeperone	Butyrophenone	Preanesthetic medication; Behavior modification (pigs)
Alfadolone–Alfaxalone	Steroid anaesthetic(contraindicated in dogs)	Anesthesia
Flunixin	Cyclo-oxygenase inhibitor	Anti-inflammatory; analgesia; antipyresis
Metamizole (Dipyrone)	Cyclo-oxygenase inhibitor	(“Similar” to other NSAIDs)

prediction of the onset of estrus. On removal of the progestogen source (intravaginal device for cows and ewes; added to feed for sows and mares), the negative feedback effect on the pituitary and hypothalamus is terminated, and estrus is initiated. Progestogens (medroxy-progesterone, megestrol, proligestone), generally administered by subcutaneous injection, are used to suppress ovarian activity (estrus) in dogs and cats, whereas altrenogest (added to feed) is used to suppress estrus in cycling mares or to synchronize estrus in gilts and sows.

Monensin, a carboxylic ionophore antibiotic, is available as a premix (for addition to feed) used for the prevention of coccidiosis caused by *Eimeria* spp. in broiler chickens. Turkeys over 16 weeks of age, guinea fowl, and other game birds should not be given access to monensin-containing feed. Monensin is used as a production enhancer (improves feed conversion efficiency and growth rate) in beef cattle and dairy heifers up to the time of first service. It is available for use in cattle as a premix (sodium salt) for addition to the feed and as a modified-release ruminal bolus. The oral LD50 (mg/kg) of monensin differs among species: horses, 2–3; sheep, 12; pigs, 16; cattle, 22; and chickens, 200. Extreme care should be taken not to feed cattle, pig, or poultry rations or supplements to horses and to avoid accidental contamination in feedmills. Steroid hormone-growth promoters (bovine somatotropin, porcine somatotropin, bovine growth hormone-releasing factor) are available for use as production enhancers in some countries, whereas in others (European Union member states), their use is banned. The properties of somatotropins and the dosage forms that have been studied were comprehensively reviewed by Foster (1). The use of β_2 -adrenoceptor agonists such as clenbuterol for production enhancement in cattle is illegal.

Some antidotal substances used in the treatment of plant or heavy metal toxicity in animals could be considered veterinary drugs. They include methylthionium chloride (methylene blue), sodium nitrite, sodium thiosulfate, ammonium molybdate, and sodium calciumedetate. Phytomenadione (vitamin K₁) is indicated for the treatment of warfarin and coumarin poisoning in animals. The substances are present in sweet vernal grass *Anthoxanthum odoratum* and in spoiled sweet clover (*Melilotus officinalis* and *M. alba*) hay, and silage. Acetylcysteine and ascorbic acid are used concurrently in the treatment of acetaminophen (paracetamol) toxicity in cats. The acetylcysteine, which is administered by intravenous injection, serves as a precursor for glutathione replenishment, whereas ascorbic acid (administered intravenously or orally) reduces methemoglobin to hemoglobin. Feline hemoglobin is particularly susceptible to oxidative damage (methemoglobinemia). Atropine sulfate is widely used in animals for preanesthetic medication (although glycopyrronium is preferred in horses) and at a much higher dose (25 to 40 times the preanesthetic dose, 44 μ g/kg), often in conjunction with a cholinesterase-reactivating agent (pralidoxime mesilate), in the treatment of organophosphate toxicity.

This overview of veterinary drugs shows applications of these drugs in animals and how some aspects of veterinary and human pharmacology differ in their orientation.

CATEGORIZATION OF SPECIES

Even though mammalian species differ in physical characteristics (notably body weight) and behavior, they

possess the same body systems that perform generally similar physiological functions. There are, however, species-related adaptations that account for the uniqueness of each species. The character of the adaptations underlies the feasibility of extrapolating scientific information obtained in certain species (such as laboratory animals) to other species (domestic animals) and human beings. Because the pattern of quantifiable adaptations can be generally described mathematically, it is possible to make some predictions regarding interspecies extrapolation. The reliability of predictions on drug bioavailability and disposition depends on knowledge of both the anatomical and physiological similarities of and differences among the species of interest.

The anatomical arrangement of the gastrointestinal tract serves as a basis for broadly categorizing domestic animals as ruminant species (cattle, sheep, and goats) or monogastric species (horses, pigs, dogs, and cats). Consideration of dietary habit somewhat refines the categorization and enables the pattern of drug absorption to be explained as well as the potential interaction among commensal microbial flora in the digestive system and drugs, especially antimicrobial agents. Ruminant species are herbivores with a voluminous forestomach compartment in which microbial fermentation takes place continuously. Horses are monogastric herbivores with a small-capacity stomach and large-capacity colon where microbial digestion takes place. Dogs and cats are monogastric carnivores, whereas pigs are monogastric omnivores (similar to humans) fed a vegetable diet. Because the urinary pH reaction is determined primarily by the composition of the diet, the usual pH range differs among herbivorous species (pH 7.2–8.4) and carnivorous species (pH 5.5–7.0), whereas in omnivorous species, urinary pH can vary over a wide range (pH 4.5–8.2), but would be expected to be alkaline in pigs, whereas it is usually acidic in humans. Urinary pH influences the extent of reabsorption from the distal renal tubules and the half-life of weak organic acids and bases when a significant fraction (arbitrarily >20%) of the systemically available dose is eliminated by renal excretion.

The character of the female reproductive (estrous) cycle varies with the animal species in several respects, that include duration of cycle, length of estrus (sexual receptivity), and time of ovulation (Table 2) (2). In seasonal breeding species (mare, ewe, doe, and queen), the time of year during which estrous cycles occur is strongly influenced by the photoperiod. The mare and queen become anestrus in late autumn owing to decreasing daylight hours, and cycles are re-established with increasing daylight in early spring. The converse situation applies to ewes and does. The plane of nutrition can affect the onset of

estrous cycles in seasonal breeding species. The queen is unique among domestic animal species in that ovulation is induced by coitus. Pharmacological intervention at any stage of the reproductive cycle, whether it be to induce ovulation in mares, ewes, or cows or to suppress estrus or prevent ovum implantation in bitches or queens, is based on changing the plasma concentrations of the reproductive hormones that influence the particular process. To be successful, knowledge of the temporal pattern of the various hormone concentrations in plasma is essential.

Because avian and mammalian species differ in many respects, these distinct classes of animal should be considered separately. Avian species that are “farmed” include chickens, turkeys, ducks, geese, ostriches, guinea fowl, quail, and pheasants. The term poultry refers to farmed domestic fowl that, in common usage, includes chickens, turkeys, ducks and geese, whereas other farmed avian species are considered game birds. Application of the collective term poultry overlooks species differences in dosage requirements and drug residues in tissues as well as differences in susceptibility to toxicity of some drugs.

Birds (and reptiles) have a well-developed renal portal system that drains blood from the caudal portion of the body. Consequently, drugs administered parenterally in the lower extremities (hind limbs) of those species pass through the kidneys before entering the systemic circulation. This feature of blood flow to the kidneys provides the opportunity for first-pass excretion of water-soluble ionized drugs (e.g., β -lactam and aminoglycoside antibiotics) to occur.

Fish, reptiles (which include crocodiles and alligators), and amphibians are poikilothermic, i.e., cold-blooded animals. In contrast to homeothermic animals, the disposition of drugs in poikilothermic animals is influenced by environmental temperature. When applied to fish, the rate of drug elimination varies with the temperature of the water to which the fish are acclimatized. Studies of drug disposition in fish should generally be carried out at more than one water temperature, and whether fresh water or sea water is contained in the tank depends on the species of fish. A complication arises in the case of salmon for example, because adult salmon live in sea water but spawn and grow as fingerlings in fresh water.

DOSAGE FORMS AND ROUTES OF ADMINISTRATION

The type of dosage form, the route of administration, and site of injection of parenteral preparations depend on the animal species or group of related species (such as

Table 2 Average length of various stages of reproductive cycles of domestic animals

Species	Duration of estrus cycle	Length of estrus	Time of ovulation	Time fertilized ova enter uterus (after conception)	Time of implantation (after conception)	Type of placenta	Length of pregnancy
Mare ^a	21 days	5–6 days	Last day of estrus	3–4 days	30–35 days	Epitheliochorial	345 days
Cow	21 days	18 h	12 h after end of estrus	3–4 days	30–35 days	Epitheliochorial	280 days
Ewe ^a	17 days	36 h	30 h after beginning of estrus	3–5 days	15–18 days	Syndesmochorial	147 days
Doe ^a (goat)	20 days	40 h	30–36 h after beginning of estrus	4 days	20–25 days	Syndesmochorial	147 days
Sow	21 days	45 h	36–40 h after beginning of estrus	3–4 days	14–20 days	Epitheliochorial	113 days
Bitch	In estrus at 7–8 month intervals depending on breed	Proestrus, 9 days; estrus, 7–9 days	First or second day of estrus	5–6 days	15 days	Endotheliochorial	64 days
Queen ^a	16 days (nonbred) (pseudopregnancy lasts 36 days)	5–6 days	Induced 24–32 h after coitus	4 days	13 days	Endotheliochorial	65 days

^aSeasonally polyestrous. (Modified from Ref. 2.)

ruminant animals or poultry). The greatest differences relate to oral dosage forms and topical preparations. Whether a veterinary dosage form is intended for individual animal treatment or for administration/application to a large group of animals (herd or flock medication) should be decided before the development of a dosage form. Convenience of administration and cost of the drug preparation are foremost considerations in determining the use of a veterinary dosage form by animal owners.

GASTROINTESTINAL ABSORPTION

The anatomical arrangement of the gastrointestinal tract and associated digestive physiology govern the pattern of drug absorption. In pigs, dogs, cats, and humans, the plasma concentration profile after the administration of an oral solution or conventional (immediate-release) dosage form is generally similar in that it shows a reasonably well-defined single peak. Because the rate of gastric emptying differs among monogastric species and the anterior (upper) portion of the small intestine is the principal site of absorption, the time at which the peak plasma concentration occurs may vary. An effective pH value of 5.3 in the microenvironment of the intestinal mucosal surface, rather than the reaction of intestinal contents (average pH 6.6), is consistent with observations on the absorption of drugs that are weak organic acids or bases. Under normal conditions, weak acids with pK_a values above 3.0 and bases with pK_a values below 7.8 are well-absorbed from the small intestine (3). An alteration in the pH of the stomach or small intestine contents can markedly change the degree of ionization of drugs that are weak organic electrolytes (acids or bases). At pH values below the pK_a , weak acids exist primarily in the nonionized form, which is the moiety that can readily be absorbed; the converse applies to weak bases. Lipid-soluble neutral molecules (digoxin, chloramphenicol) and fluoroquinolones (amphoteric compounds) are well-absorbed in dogs, although the systemic availability of norfloxacin is much lower than that of enrofloxacin/ciprofloxacin or marbofloxacin. Because of their polar nature, aminoglycoside antibiotics are poorly absorbed from the gastrointestinal tract and must be administered parenterally in the treatment of systemic bacterial infections.

After the administration of a sustained-release oral dosage form, the duration of drug availability for absorption is limited by the sum of the residence times of the dosage form in the stomach and small intestine. This has been estimated to be 9–12 h in dogs. In the development of a sustained-release dosage form for use

in dogs, the aim is to provide an effective plasma concentration of the drug throughout the dosage interval (12 h) with an acceptable degree of fluctuation in steady-state concentrations. The latter is determined both by the half-life of the drug and the dosage interval (4). Suitable candidate drugs should have reasonably high oral bioavailability, which implies reliable absorption from the gastrointestinal tract and no more than partial inactivation by the first-pass effect unless active metabolites are formed, a half-life in the range of 4 to 6 h, and a relatively high potency but reasonably wide range of therapeutic plasma concentrations. Because of variation between dogs and humans in the oral bioavailability and the rate of elimination of most lipid-soluble drugs, those that would be suitable for formulating as sustained-release dosage forms often differ between the two species. Theophylline meets the criteria that make it suitable for formulating as a sustained-release oral preparation for administration to dogs. Of the sustained-release oral dosage forms that are commercially available, anhydrous theophylline in tablet form (200 and 300 mg) is preferred for use in dogs. This product has an oral bioavailability (theophylline) of 76%, and the dosage regimen (20 mg/kg administered at 12-h intervals) has been predicted to maintain plasma concentrations within the therapeutic range (6–16 $\mu\text{g/ml}$) with less fluctuation in peak-to-trough theophylline concentrations than other sustained-release dosage forms (5). The dosage regimen for the conventional dosage form (aminophylline tablets) is 10 mg/kg administered at 8-h intervals to dogs.

Other drugs of veterinary interest that have been formulated as sustained-release oral dosage forms include morphine, propranolol, quinidine, procainamide, and verapamil or diltiazem. Of these, sustained-release morphine sulfate (tablets, 15 mg) has the greatest potential for use in dogs over 10 kg body weight. The proposed dosage regimen, 1–2 mg/kg administered at 8- or 12-h dosage intervals, may be effective in the management of chronic pain (6). The sustained-release dosage form overcomes the shortcoming of conventional oral preparations of morphine, which is their short duration of action in dogs. Phenytoin is a likely candidate drug for formulating as a sustained-release dosage form because it would enable phenytoin to be used in dogs for the prevention/treatment of generalized tonic-clonic seizures (grand mal epilepsy). The average oral bioavailability of phenytoin administered as the conventional dosage form is 36%, the half-life in dogs is 3.5–4.5 h (dose-dependent), and the therapeutic range of plasma concentrations is 10–20 $\mu\text{g/ml}$. Oral bioavailability of phenytoin in humans is 90%, and the apparent half-life, which is dose-dependent, is 15–24 h. The slower elimination of phenytoin in humans obviates the

need for the development of a sustained-release dosage form.

Even though the horse is a monogastric species, the stomach has a small capacity (8.5% of the gastrointestinal tract) compared with that of the pig (29%) and the dog (62%). Expressed on the basis of volume capacity, the stomach of the horse, pig, and dog can hold 7–14, 5.5–7, and 3–8 L, respectively. The average pH value of gastric contents in horses is less acidic (pH 5.5; range, 4.5–6.0) than in other monogastric species (pH 3–4), and a substantial portion of the stomach lining is composed of stratified squamous epithelium. Under natural conditions of management, horses feed continuously. The fibrous component of the feed is digested primarily in the large intestine, although horses digest fiber less efficiently than do ruminant species. Microbial digestion of fibrous carbohydrates to volatile fatty acids and break down of undigested dietary protein to peptides and amino acids takes place in the large intestine. The combined capacity of the caecum and colon occupies approximately 55% of the gastrointestinal tract, and the pH of large intestinal contents is 6.6–6.8. Two unrelated features of the digestive system are that horses do not possess a gall bladder (similarly camelids and giraffes) and are unable to vomit. The temporal relationship between feeding and oral dosing can greatly influence the pattern of drug absorption. Because the systemic availability of most antimicrobial agents administered orally (paste formulations) or by nasogastric tube (aqueous suspensions) to horses is significantly decreased by feeding shortly before dosing, food should be withheld for up to 2 h after administration of an antimicrobial agent. Feeding horses close to the time of administering phenylbutazone in various oral dosage forms changed the pattern without altering the extent of absorption of the nonsteroidal anti-inflammatory drug (7). The availability of a drug for absorption from the small intestine of the horse, particularly when administered in

conjunction with or shortly after feeding, may be limited by adsorption to the ingested feed, especially hay. Under these circumstances, absorption may occur in two phases, initially (1–2 h after drug administration) from the small intestine and several (8–10) hours later from the large intestine (principal site) after microbial digestion of the fibrous material in feed (7, 8). The inadvertent contamination of horse feed with a feed additive premix approved for use as a production enhancer/growth promotant in cattle or pigs, may cause toxicity, even death, in horses.

The anatomical arrangement of the gastrointestinal tract distinguishes ruminant (cattle, sheep, and goats) from monogastric(horses, pigs, dogs, and cats) species (Table 3). The difference in digestive physiology between the two groups of species determines the types of oral dosage forms appropriate to administer and influences oral bioavailability of drugs. Susceptibility to ingested plant toxicity differs between ruminant and monogastric species. The volume capacity of the mature reticulorumen is 100–225 L in cattle and 10–25 L in sheep and goats and accounts for approximately 60% of the total capacity of the gastrointestinal tract. The forestomach contents vary from liquid to semisolid consistency, and the pH reaction is normally maintained within the range 5.5 to 6.5 owing to copious secretion of alkaline saliva (pH 8.2–8.4) particularly rich in bicarbonate and phosphate buffers but devoid of amylase. In addition to its buffering action, saliva has an antifoaming action that serves to prevent dietary bloat. Salivary secretion, at a daily rate of 100–190 L in cattle and 6–16 L in sheep, is essential for microbial digestion, which takes place continuously in the reticulorumen. Based on average values of saliva flow and volume of the rumen liquid pool, the turnover rate for reticulorumenal fluid was estimated to be 2.0/day for cattle and 1.1–2.2/day for sheep (9). Despite the stratified squamous nature of its epithelial lining, the rumen has considerable absorptive

Table 3 Relative capacity of components of digestive tract of domestic animal species

Component	Relative capacity (%)			Component	Relative capacity (%) (Sheep and Goat)
	Horse	Pig	Dog		
Stomach	8.5	29.2	62.3	Rumen	52.9
Small intestine	30.2	33.5	23.3	Reticulum	4.5
Cecum	15.9	5.6	1.3	Omasum	2.0
Large colon	38.4			Abomasum	7.5
		31.7	13.1		
Small colon and rectum	7.0			Small intestine	20.4
				Cecum	2.3
				Colon and rectum	10.4

capacity (10, 11). Because absorption takes place by passive diffusion, lipid-soluble drugs, whether neutral molecules or the nonionized form of weak organic acids or bases, may be absorbed from the rumen. The theoretical equilibrium distribution, expressed as concentration ratio, of weak organic acids and bases of different pK_a values between saliva (pH 8.2) or ruminal contents (pH range 5.5–6.5) and plasma (pH 7.4) is presented graphically (Fig. 1) (12).

Ruminal micro-organisms are capable of at least partially metabolizing some drugs (e.g., trimethoprim, chloramphenicol, nitroxylin, digitalis glycosides) either by hydrolysis or reduction, which would decrease the amount of drug available for absorption. Intraruminal biotransformation makes ruminant species susceptible to toxicity caused by ingestion of plants containing cyanogenic glycosides (cyanide poisoning) or accumulated nitrate (nitrite poisoning). Overuse of nitrogenous fertilizers contributes to the incidence of nitrite poisoning. Antidotal substances that are used (injected intravenously) to treat these toxicities are sodium nitrite, followed by sodium thiosulfate for cyanide poisoning and methylthioninium chloride (methylene blue) for nitrite poisoning. In ruminant species, in which chemical compounds can be

altered (activated or detoxified) by microbial action in the forestomach and in which the whole physiological tempo of the body is so dependent on ruminal activity, no pharmacological or toxicological investigation can be interpreted without full consideration of the basic diet and feeding regimen (13).

After comminution by rechewing and microbial digestion, the liquid component with suspended particles of reticuloruminal contents is “pumped” by the omasum (third compartment of the forestomach) into the abomasum (true stomach). During the two-stage transfer process, water and electrolytes are absorbed, and the size of particulate matter in the digesta is reduced. The abomasum, which accounts for approximately 4–5% of the capacity of the gastrointestinal tract in adult cattle and 7.5% of gastrointestinal capacity in sheep and goats, is the only compartment of the ruminant stomach that secretes digestive juices. Secretions from the fundic area of the abomasum contain hydrochloric acid, pepsin and, in suckling preruminant animals, rennin (a milk-coagulating enzyme). Mucus is produced by the columnar epithelial cells that line the abomasum. The pH reaction of abomasal contents does not vary much and is usually close to pH 3.0 (14).

Because of the large volume of reticuloruminal contents, a drug can attain only a low concentration in the reticulorumen, whether administered in solution, suspension, or solid dosage form. In dissolution of solid dosage forms, dilution in the large volume of fluid and binding to particulate matter would decrease the rate, but not necessarily the extent, of drug absorption. Lipid-soluble neutral molecules and the nonionized form of weak organic electrolytes, particularly organic acids, should normally be well-absorbed from the reticulorumen. When aspirin (pK_a 3.5) in a solid dosage form (60 g, which is equivalent to 3.9 g of oral bolus) was administered to cows, salicylate was slowly absorbed, and systemic availability was 50–70% (15). The 12-h dosage interval for aspirin in adult cattle is based on the rate of absorption rather than on the half-life of salicylate, which is 0.8 h.

Benzimidazole anthelmintics (albendazole, fenbendazole, oxfendazole, which is fenbendazole sulfoxide); probenzimidazoles (netobimin and febantel, which are metabolically converted to albendazole and fenbendazole, respectively); and salicylanilide flukicides (closantel, rafoxanide, oxclozanide) are administered as oral suspensions to ruminant animals. For anthelmintic drugs oral suspensions have an advantage over oral solutions in that the delayed availability of drug for absorption extends the duration of action. Reducing the level of feed intake (for 36 h before and 36 h after dosing) delays the onward passage of ruminal fluid with suspended matter from the reticulorumen to the abomasum and small intestine and

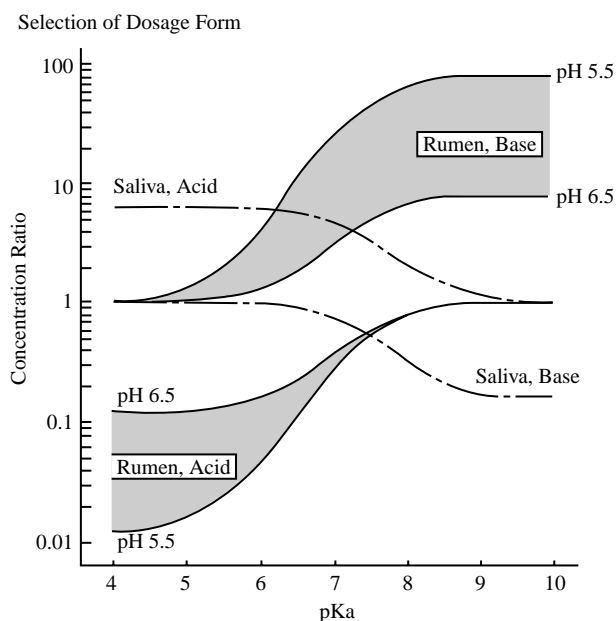


Fig. 1 Expected equilibrium distribution between saliva or rumen contents and plasma of acids and bases of differing pK_a . Concentration ratio is the ratio of the salivary or ruminal concentration to concentration free in the plasma, calculated separately for acids and bases, for saliva of pH 8.2 and rumen contents over a range of pH 5.5–6.5, assuming plasma is pH 7.4. (From Ref. 12.)

would allow more time for drug dissolution and absorption to occur. When the rate of onward passage of ruminal fluid was decreased, by temporarily reducing feed intake, in sheep, the systemic availability of oxfendazole was increased (16). The systemic availability of rafoxanide and triclabendazole, administered as oral suspensions, was higher in housed lambs fed hay and concentrates than in lambs grazing on pasture (17). The effect of dietary regimen/composition on systemic availability of the flukicides could be attributed to the time allowed for dissolution and absorption of the drugs.

Closure of the reticular groove would allow an oral solution to flow directly from the cardiac orifice of the rumen into the abomasum, bypassing the rumen. Spontaneous closure of the groove occurs reflexly in some animals although at least partial closure can be chemically induced. Should rapid absorption of a drug (such as a nonsteroidal anti-inflammatory drug) from the gastrointestinal tract be desired, an oral solution of the drug could be administered immediately after inducing closure of the reticular groove. This can be achieved generally in cattle by administering orally a solution containing sodium bicarbonate and in sheep by administering orally a copper sulfate solution or injecting intravenously a dose (0.3 U/kg) of lysine-vasopressin. Closure of the reticular groove would be unwanted in the case of an oral liquid, such as dimeticone emulsion or poloxaline, used to treat frothy bloat in cattle. In young ruminant animals, reflex closure of the reticular groove, induced by suckling, allows milk to flow directly to the abomasum.

Advantage can be taken of the anatomical arrangement of the forestomach of ruminant animals by administering a modified-release ruminal bolus that will remain lodged in the reticulorumen for a prolonged period. Slow-release ruminal boluses containing trace elements (cobalt oxide, copper oxide, cobalt and copper in sodium phosphate glass matrix, and selenium as sodium selenate) are commercially available for administration to cattle or sheep, and the compound ruminal bolus can be administered to farmed adult deer. Controlled-release ruminal boluses containing certain anthelmintic drugs (ivermectin, fenbendazole, oxfendazole, morantel) or the production enhancer monensin are available for oral administration, using a specialized delivery device, to beef cattle within a specified range of body weight (100–400 kg). These systems are designed either to continuously release the drug into ruminal fluid for a prolonged period (generally 120–140 days) or to intermittently release pulse doses (oxfendazole bolus) at a predetermined interval (approximately 3 weeks). This interval generally coincides with the prepatent period of the major gastrointestinal trstrongylids of cattle (18). Controlled-release ruminal boluses that continuously

release ivermectin, fenbendazole, or morantel into ruminal fluid are available for use in beef cattle, whereas controlled-release boluses that deliver albendazole or ivermectin over a period of 100 days are available for use in sheep (35–65 kg in body weight). The delivery systems are retained in the reticulorumen at least throughout the entire period of drug release. The retention of oral controlled-release drug delivery systems designed for use in humans is limited by gastric emptying time, which can be up to 12 h (19).

FIRST-PASS EFFECT

Having traversed the gastrointestinal mucosal barrier, drug molecules are conveyed in hepatic portal blood to the liver, where they are subjected to the first-pass effect before entering the systemic circulation. The first-pass effect applies to all animal species and, owing to the generally higher capacity of the liver of herbivorous species (ruminant animals and horses) to metabolize lipid-soluble drugs by microsomal oxidative reactions, is likely to decrease the systemic availability of these drugs to a greater extent in herbivorous than in nonherbivorous species (dogs, cats, pigs, humans). In ruminant species (cattle, sheep, goats) and *Equidae* (horses, ponies, donkeys), triclabendazole, administered as an oral suspension, is converted by hepatic first-pass metabolism to triclabendazole sulfoxide (active metabolite), which is subsequently converted to the sulfone (inactive) metabolite. In many species, the first-pass effect can substantially reduce the systemic availability of orally administered lipid-soluble drugs that undergo extensive biotransformation in the liver (e.g., diazepam, propranolol, verapamil). Although presystemic metabolism occurs primarily in the liver, it can also take place in the small intestinal mucosa during the absorption process. Some drugs that show incomplete systemic availability after oral administration to dogs are listed in Table 4. Presystemic metabolism in the liver is responsible for incomplete systemic availability of the majority of these drugs. The oral bioavailability of digoxin was enhanced in humans when the drug was administered as an aqueous alcoholic solution in gelatin capsules rather than in tablet form, even though the tablets had a satisfactory dissolution rate (20–22). In a similar manner, increased relative bioavailability, based on comparison of areas under the plasma concentration–time curves of flufenamic acid, was achieved in dogs when the drug was administered in soft rather than hard gelatin capsules; the average increase was 34% (23). In both instances, the higher oral bioavailability was attributed to physicochemical factors brought into play by adjuvants in the soft gelatin capsules. The influence that

Table 4 Systemic availability of some orally administered drugs in dogs

Drugs (dosage form)	Dose (mg/kg)	Systemic availability (%)	Site of metabolism/ other factor
Phenobarbitone (tablet)	10	86–96	Liver
Valproic acid(tablet)	40	78	Liver
Phenytoin (tablet)	15	36 ^a	Liver
Salicylate (aspirin tablet)	250 mg total	45	Liver
Ibuprofen (gelatin capsule)	5	60–86	Liver
Naproxen ^b (gelatin capsule)	5	68–100	—
Theophylline (conventional aminophylline tablet)	10	91	Liver
Diazepam (tablet)	2	1–3 86 ^c	Liver + (intestinal mucosa)
Lidocaine (solution)	10	15	Liver
Procainamide (tablet)	25	85 ^{a,d}	
Propranolol (conventional tablet)	80 mg total	2–17	Liver
Verapamil	0.5	15	Liver
Digoxin (Lanoxin tablet)	1 mg total	80	Dissolution
Cephalexin monohydrate(capsule)	20	57	Dissolution
Norfloracin (tablet)	5	35 ^a	Dissolution + liver
Enrofloxacin (tablet)	5	100 ^e	—
Allopurinol (tablet)	15	70 ^f	Intestine + liver (xanthine oxidase)

^aAverage oral bioavailability; wide individual dog variation.

^bNaproxen has an unusually long half-life in dogs.

^cTotal active benzodiazepine.

^dDogs do not form *N*-acetylprocainamide (active metabolite).

^eTotal antimicrobial active fluoroquinolone.

^fAverage oral bioavailability is 14% in the horse.

formulation, or rather dosage form, can have on the systemic availability of an orally administered drug was shown in horses given racemic ketoprofen (2.2 mg/kg). The systemic availability of the S(+)- and R(–)-enantiomers was 54.2 and 50.5%, respectively, after administration of micronized racemic ketoprofen powder in hard gelatin capsules to horses with restricted access to feed. When ketoprofen powder from the same batch was administered as an oil-based paste, systemic availability of the S(+)- and R(–)-enantiomers was 5.75 and 2.7%, respectively, regardless of the feeding schedule (24). To avoid hepatic first-pass metabolism, glyceryl trinitrate (nitroglycerin) is formulated in a variety of dosage forms, which include parenteral solution, spray for sublingual application, sustained-release tablet, and transdermal therapeutic systems, for use in humans, and as an ointment for topical application to dogs (cardiogenic pulmonary edema) or horses (acute laminitis). Sublingual administration avoids the first-pass effect, but it is not feasible to administer solid dosage forms by this route to animals.

Although the first-pass effect is a major source of species variation in systemic availability of orally administered drugs that undergo extensive hepatic metabolism, another

important source of variation is metabolism by ruminal microorganisms. Some drugs (nitroxylin, chloramphenicol, digitalis glycosides) are metabolized in the rumen to such an extent that parenteral administration is required for clinical efficacy.

An advantage of rectal over oral administration of lipid-soluble drugs is partial avoidance of the first-pass effect. The extent to which the first-pass effect is avoided appears to be less in dogs than in humans. Because venous drainage of the rectum of the horse, unlike in the humans and dog, appears to be substantially into the hepatic portal vein, minimal avoidance of hepatic first-pass metabolism of drugs absorbed from the rectum could be anticipated in horses.

ORAL DOSAGE FORMS

The oral route of drug administration is safer than parenteral routes and avoids tissue irritation at injection sites. However, wide variation, both inter- and intraspecies, in systemic availability is a feature of orally

administered drugs. The convenience of oral drug administration depends on the species of animal and the dosage form of the drug. To facilitate drug administration and take into account the anatomy and physiology of the digestive system and the average body weight of the various animal species, the requirements of oral dosage forms differ among species. Consideration must be given to size of the total dose (amount of drug) to be administered and oral bioavailability in the animal species.

Oral dosage forms available for administration to animals include oral solutions, liquids, suspensions, gels, pastes, capsules, tablets, ruminal boluses, powders and granules for addition to feed, soluble powders for addition to drinking water or fish medicating baths, and premixes for addition to feed for livestock or poultry. The type of dosage form is determined by the solubility and physicochemical properties of the drug, the species of animal for which the dosage form is intended, and whether a reasonably rapid onset or a prolonged duration of effect is required. Liquid dosage forms (oral solutions and liquids) provide readily available drug for absorption, particularly in monogastric species. Oral liquids, such as dimeticone emulsion and poloxaline (anionic surfactant), are used for the treatment of frothy bloat in cattle. An oral liquid containing propylene glycol (glucose precursor) for addition to drinking water or for preparation of an oral solution is indicated for adjunctive treatment of ketosis in cattle and sheep. Aqueous suspensions are administered by nasogastric tube to horses, by mouth (as a drench) to ruminant animals and by mouth (sometimes with the aid of an oral syringe) to dogs and cats. Oral pastes and gels are semisolid dosage forms supplied in preloaded calibrated syringes designed for convenience of drug administration to horses by their owners. The formulation of a paste must be such that it is syringeable over a wide range of ambient temperatures, is moderately tenacious so that it will adhere to the tongue, and is tasteless or suitably flavored (mint or apple flavor appears to be preferred by horses). Classes of drugs that could be formulated as oral pastes or gels for administration to horses include anthelmintics, nonsteroidal anti-inflammatory drugs, and some antimicrobial agents. Application of an anthelmintic paste, at the appropriate dose of the drug, to the distal forelimbs of cats is a convenient alternative to oral administration of a tablet or capsule. Solid dosage forms must undergo disintegration and dissolution in the stomach before absorption of the drug can occur. Drug release from solid dosage forms delays the rate of absorption compared with that from an oral solution. With regard to the quality of solid dosage forms, capsules offer an advantage over tablets in that the particle size and distribution of the active ingredient are rarely altered by the capsule-filling process, whereas

physical stresses (compression and heat) are imposed in the manufacturing process of tablets. In addition, capsules provide more rapid dissolution than do conventional tablets of the same active ingredient. The pattern of release of coated pellets or granules is more predictable when packed in capsules than compressed into tablets. Capsules, however, are generally more expensive than tablets containing the same drug. Because of the higher manufacturing cost, soft gelatin capsules should probably be used in preference to hard gelatin capsules only when the fill is liquid or bioavailability of the drug is significantly superior. An encapsulated active ingredient, until released, would be protected against metabolism by ruminal micro-organisms and/or chemical degradation in the ruminal environment. Encapsulation of insect development inhibitors (e.g., diflubenzuron) or insect growth regulators (e.g., methoprene) has potential application in cattle. These substances interfere with the metamorphosis and reproductive capacity of arthropod parasites, many of which breed in cattle dung. Dichlorvos, an organophosphorus compound with activity against gastrointestinal nematodes including whipworms (*Trichuris* spp.), is incorporated into polyvinyl chloride resin pellets (plasticization) for addition to the feed for pigs or canned food for dogs. This dosage form protects dichlorvos from hydrolytic degradation and provides slow release into the gastrointestinal tract, which greatly increases the margin of safety of the drug.

A sustained-release dosage form provides an initial amount of drug sufficient to provide a desired therapeutic concentration and continuously releases the drug at a constant (zero-order) rate for an extended period. The margin of safety of the drug is an important consideration, because sustained-release dosage forms contain a large amount of drug and the potential exists for "dose-dumping" with resultant toxicity. For drugs with half-lives in the range of 4 to 6 h, sustained-release dosage forms could provide a dosage interval of 12 h in dogs or 24 h in horses. Very few sustained-release preparations have been developed for use in either species. This type of dosage form has to be swallowed intact.

Controlled-release ruminal boluses are designed either to continuously release drug (an anthelmintic or production enhancer) at a constant rate for a prolonged specified period or to intermittently deliver pulse (usually five) doses at predetermined intervals. Each product is designed for use in either cattle or sheep within a specified range of body weight. Retention of orally administered controlled-release systems in the reticulorumen is dependent on either density or geometry of the system. The ivermectin (Ivomec) ruminal bolus, designed for use in cattle between 100 and 400 kg body weight,

contains 1.72 g of ivermectin that is continuously released into ruminal fluid at a constant rate (12.5 mg/day) over a period of 135 days. The bolus is a cylindrical device that consists of an outer semipermeable membrane enclosing a metal density element at one end, an osmotic energy source at the other end, and a formulation containing the drug in the center. The osmotic energy source is composed of a tablet containing a polymeric salt mixture. Absorption of water through the semipermeable membrane causes expansion of the tablet, which drives the ivermectin wax formulation through the exit port (channel) in the density element (25) (Fig. 2a). The

fenbendazole ruminal bolus (Panacur), designed for use in cattle between 100 and 300 kg body weight, contains 12 g of the anthelmintic which, is continuously released into ruminal fluid over a period of up to 140 days. Release of fenbendazole is controlled by the erosion of two magnesium alloy tubes; each tube contains five tight-fitting cylindrical tablets (6 g of fenbendazole). The magnesium alloy tubes are protected from the ruminal contents by a series of close-fitting rigid plastic rings; the central ring connects both tubes. The bolus is completely biodegradable and leaves no residual device in the forestomach (26). The morantel tartrate (Paratect Flex) ruminal bolus, designed for use in cattle over 100 kg body weight, contains 11.8 g of morantel base. The bolus consists of a drug-impregnated polymer packaged in a cylindrical trilaminar cartridge. Release of the drug (approximately 150 mg/day) takes place by diffusion through the semipermeable membrane into ruminal fluid over a period of at least 90 days. There is no withdrawal period associated with the use of the morantel tartrate trilaminar bolus because the combination of poor absorption and first-pass hepatic metabolism makes systemic availability of morantel exceedingly low. The Captec, also known as the Laby (27), device is a controlled-release ruminal capsule designed for use in sheep within a specified range of body weight that continuously releases the anthelmintic contained within, either albendazole (32.5 mg/day) or ivermectin (1.6 mg/day), over a period of 100 days. The device consists of a polypropylene barrel with a spring-loaded plunger at the closed (top) end and an orifice (series of perforations) at the base (Fig. 2b). A pair of polypropylene wings attached to the top of the barrel is secured in a folded position (compressed configuration) by water-soluble tape to enable oral administration. The barrel is loaded with tablets containing the anthelmintic, either albendazole (3.85 g) or ivermectin (160 mg). After administration of the ruminal capsule, the tape securing the pair of wings dissolves, and the wings open to a predetermined angle (expanded configuration) that prevents both regurgitation and reticulo-omasal passage. The rate of drug release is determined by a combination of factors that include the formulation of the tablets, the diameter of the orifice, and the pressure exerted on the tablets by the plunger. The ruminal fluid bathing the orifice of the device brings about dissolution of the tablets.

The modified-release ruminal bolus containing oxfendazole, designed for use in cattle in the body weight range of 200 to 400 kg, delivers pulse doses (1.25 g) at 3-week intervals over a period of 105 days (18). The bolus containing five tablets (Autoworm 5) releases the first dose at 21 days after administration whereas the bolus

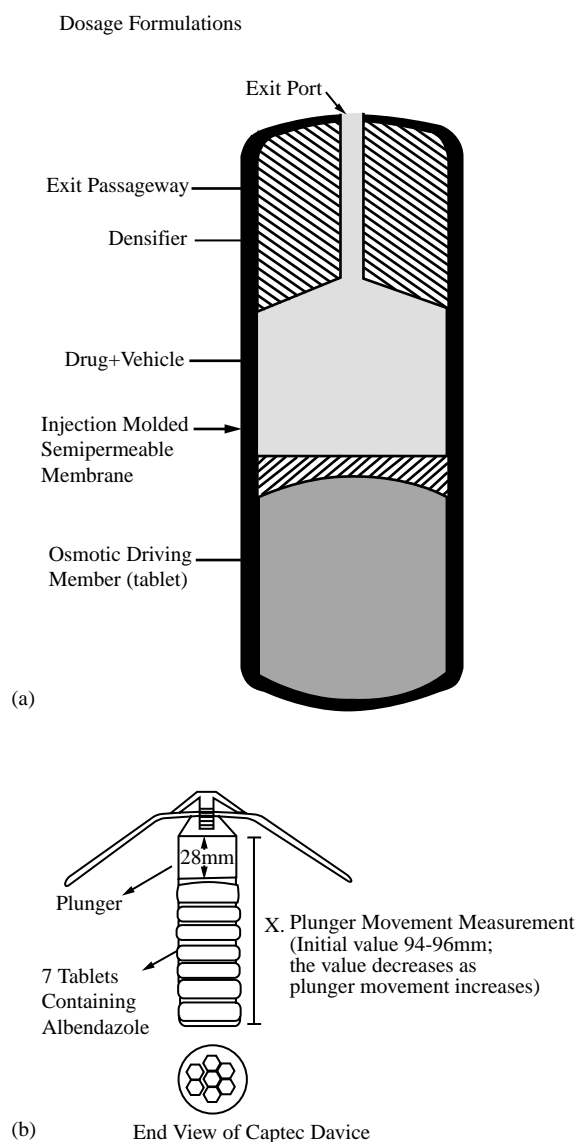


Fig. 2 (a) IVOMEC schematic; (b) captec schematic.

containing six tablets (Autoworm 6) releases the first dose on the day of administration.

Drugs in powder form or prepared granules can be added to the feed for pigs. Shortcomings associated with this method of administration include uncertainty as to the amount of drug (dose) ingested and interanimal variation in oral bioavailability. The drug must have a wide margin of safety and be palatable in the feed and, importantly, the animal must be feeding. Inappetence or indifference to feeding is a usual feature of illness in animals. Addition to feed, often referred to as top dressing of feed, is generally an unreliable method of drug administration to horses. However, the packaging of powders or granules in unit dose sachets provides convenience of drug administration for owners when the number of animals to be dosed is small. Unit dose sachets must be prepared on an individual species basis because the dose level (mg/kg) may differ among species, and the average body weight range of domestic animal species differs by approximately 200-fold. Soluble salts of drugs in prepared solutions or powder form can be administered to poultry and other farmed avian species by addition to drinking water or to fish by addition to the water in a medicating bath of known volume capacity. Sachets containing various trace elements (cobalt, copper, iodine, selenium) for addition to drinking water delivered by dispenser are available for the correction of mineral deficiencies in cattle. The mineral content of each sachet suffices for 25 cattle for a period of 7 days. Several factors to consider in the design of premix formulations and to ensure their satisfactory mixing in bulk feed are addressed in detail by Klink et al. (28). Premixes must always be diluted to the approved use level, usually parts per million (ppm) [g/tonne (feed) or mg/L(water)], for the animal species.

INTRAMUSCULAR INJECTION

Intramuscular injection is a commonly used route for drug administration in animals. This route offers convenience in that it is easy to apply and, because a sizeable proportion of veterinary parenteral preparations are formulated as long-acting dosage forms, either a single dose may suffice or a long dosage interval can be used. There are, however, a number of factors to consider before choosing this route of administration. They include the purpose for administering the drug, the acceptable time interval between drug administration and onset of the desired effect, the suitability of available parenteral preparations for use in the particular animal species, the withdrawal period(s) in food-producing animals, and the cost of the drug preparation

for the anticipated course of treatment. The suitability of a parenteral preparation for use in a particular species primarily depends on the formulation of the preparation. Significant formulation variables include the concentration of drug in a preparation, which determines the volume to be injected, the nature of the vehicle and other ingredients (solubilizers, preservatives), and the likelihood of causing irritation at the injection site. The horse is the least tolerant of domestic animal species to injection-site irritation, and drugs in oily vehicles should never be administered by injection to horses.

Whereas the formulation of a preparation determines the pattern of absorption and influences systemic availability of the drug, the volume of the preparation deposited in muscle and the vascularity of the injection site determines the rate of absorption. The barrier to absorption is the capillary endothelium, which most drugs readily penetrate by passive diffusion, although small water-soluble molecules may enter capillaries by bulk flow through intercellular "pores" in the endothelial membrane. Aqueous channels (pores) in the capillary endothelium are approximately 10-fold larger in diameter (40–80 Å) than those in the intestinal epithelium. Polar drugs such as aminoglycoside antibiotics are rapidly and completely absorbed from intramuscular injection sites, whereas they are very poorly absorbed from the gastrointestinal tract when the epithelial membrane is intact. Drug absorption from intramuscular injection sites is generally assumed to be a first-order process. However, this assumption may not be entirely valid, particularly during the initial stage when absorption may obey zero-order (nonlinear) kinetics. Local and systemic factors that influence the rate of absorption seldom remain constant throughout the absorption process. Increased blood flow to skeletal muscle at the site of injection promotes absorption, whereas the administration of a drug that causes local vasoconstriction or the presence of a disease state that decreases skeletal muscle perfusion delays absorption. Deposition of the injected preparation between muscle masses (intermuscular) or in adipose tissue and the injection of a preparation that causes tissue irritation or precipitation of the drug at the injection site produce an erratic pattern of absorption that is reflected in the plasma concentration profile of the drug.

The site of intramuscular injection can affect the plasma concentration profile and bioavailability of a drug, particularly when a long-acting preparation is administered. The variation in the pattern of absorption can be attributed to regional differences in blood flow to skeletal muscles and in absorptive surface area. In cattle and goats, intramuscular injection in the lateral neck provides superior absorption to injection in the buttock

(*M. semitendineus*) and thigh muscle (*M. quadriceps femoris*), respectively, or subcutaneous injection in the lateral neck. In pigs, the lateral neck should always be used as the site for intramuscular injection of parenteral preparations. The reasons for selecting this site include better absorption than in other sites, less residual drug at the injection site, and avoidance of damage to the carcass. In dogs and cats, the quadriceps muscle mass is the preferred site for intramuscular injection, which should be performed slowly, of parenteral preparations. After the injection of procaine penicillin G (20,000 IU/kg) at various intramuscular sites and subcutaneously in the cranial part of the pectoral area in horses, the peak plasma concentration and systemic availability of penicillin G were highest when the long-acting preparation was injected i.m. in the neck region (*M. serratus ventralis cervicis*). This was followed, in descending order of injection site, by *M. biceps* > *M. pectoralis* > *M. gluteus*, or subcutaneously (Fig. 3) (29). The intramuscular injection technique must ensure that inadvertent intravenous administration does not occur.

Potential disadvantages of intramuscular injection are the deposition of the preparation in adipose tissue or intermuscular fascial planes and the production of tissue damage with persistence of drug residue at the site of injection (30–32). Tissue damage is more likely to be caused by constituents of the formulation than the drug substance *per se*. Useful antemortem methods of assessing the extent of tissue irritation and the rate of resolution at an intramuscular injection site include the echographical examination of the muscle tissue in the immediate vicinity of the injection site (33) and the monitoring of plasma creatine kinase (CK) activity (34, 35). These methods have

the distinct advantage of being applicable to the live animal. Should moderate to severe tissue damage be evident, the extent of the damage and precise nature of the lesion can be described on postmortem examination. The use of some tissue-damaging parenteral preparations in food-producing animals is unavoidable, and the specified withdrawal period(s) must be applied. The withdrawal period for a drug varies with the formulation of the dosage form (preparation), which should be administered only by the recommended route and may differ between animal species. Fish are ectothermic and their basal metabolic rate, which markedly influences the rate of elimination of drugs, varies with the temperature of the water to which they are acclimatized. It follows that withdrawal periods, stated in degree days, for drugs used in “farmed” fish (primarily by addition to feed or bath water) will vary with ambient water temperature.

In cats, puppies, and piglets, particular attention should be given to the drug concentration in parenteral preparations and, when giving an intramuscular injection in the thigh (particularly of cats), to avoid causing damage to the sciatic nerve. Because avian and reptilian species have a well-developed renal portal system, first-pass renal excretion may decrease the systemic availability of drugs that are primarily eliminated by the kidneys (e.g., β -lactam and aminoglycoside antibiotics) when injected intramuscularly in the thigh of birds or caudal half of the body of reptiles.

SUBCUTANEOUS INJECTION

Subcutaneous injection is an alternative route to intramuscular injection for administration of parenteral preparations to domestic animals. This route of administration is used most often in dogs and cats and only occasionally in horses. Most of the factors that influence drug absorption from skeletal muscle also apply to subcutaneous sites. They include the concentration of drug in the parenteral preparation, the nature of the vehicle and other components of the formulation, the total volume of the preparation administered, blood supply to subcutaneous tissue and area of the absorptive surface, tissue irritation, precipitation of drug at the site of injection, and persistence of drug residues. Absorption from subcutaneous sites is often slower and more erratic than from intramuscular sites because of the limited and more variable blood flow to subcutaneous tissue. However, the absorptive area may be larger and can be expanded by massaging the skin covering the region of the injection site. The degree of tissue irritation caused by some parenteral preparations appears to be greater after subcutaneous than intramuscular

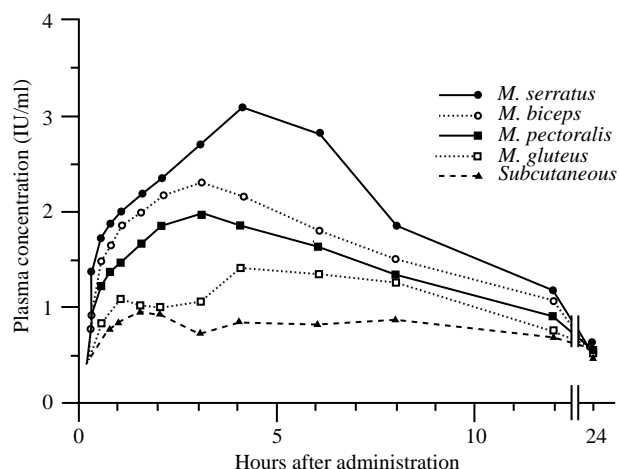


Fig. 3 Mean plasma penicillin concentration–time curves after 20,000 IU of procaine penicillin g/kg was administered to 5 animals (4 horses, 1 pony) at 5 different sites. (From Ref. 29.)

injection(36, 37). Because the damaged tissue surrounding a subcutaneous injection site can be trimmed from the carcass of a slaughtered animal, the likelihood of violative drug residues in the meat is reduced. Subcutaneous injection is the preferred route of administration for some poorly soluble suspensions. When formulating sustained-release parenteral preparations, a compromise may have to be made between the rate of drug release and the degree of tissue irritation caused by the preparation. Various modified-release devices for subcutaneous implantation have been developed as an alternative to injection of sustained-release oily suspensions. Devices developed include an implantable osmotic pump that delivers growth hormone-releasing factor over a period of 7 days in steers and wethers (38). Another subdermal implant is a diffusion polymeric matrix system that releases estradiol at an approximately zero-order rate. A modified-release melatonin implant designed to advance the time of onset of cyclical ovarian activity in ewes is commercially available. Although great efforts have been made toward technological development of modified-release devices for subcutaneous implantation in cattle and sheep, the use of implants in food-producing animals does not appear to have gained wide acceptance over the past decade.

INTRAVENOUS INJECTION

When a parenteral drug solution is administered intravenously, it is assumed that the total dose of the drug is completely available systemically. This assumption overlooks the fact that an intravenously administered drug passes through the lungs, which constitute an organ of biotransformation, before being distributed throughout the body. Intravenous injection produces a prompt pharmacological response and overcomes the variability in absorption associated with other routes of drug administration. Parenterals, excluding long-acting preparations, that are too irritant to be injected by other routes may be cautiously administered intravenously. It must be ensured that perivascular leakage at the site of injection does not occur. Intravenous injection of irritant solutions can cause thrombophlebitis. Once a drug has entered the systemic circulation, its removal from the body is entirely dependent on the elimination(biotransformation and excretion) processes. Pharmacokinetic parameters that describe the disposition(distribution and elimination) of drugs are based on the plasma concentration profiles after intravenous injection of single doses. Assumptions made in the estimation of absolute bioavailability, which is based on comparison of total areas under the plasma concentration–

time curves after extravascular and intravenous administration of the drug, are that the intravenously administered dose is completely available systemically and that the clearance of the drug is not changed by the route of administration.

The vein commonly used for injecting drugs intravenously varies with the species of animal. In horses, cattle, sheep, and goats, the jugular vein is used, whereas in dogs and cats, either the cephalic (usually) or the jugular vein is used, and in pigs, an ear vein is often used. In anesthetized dogs and cats, the jugular vein is convenient to use for administering supplemental doses of intravenous anesthetics, whereas in dogs, the sublingual vein may be used in emergency situations for injecting drug solutions of small volume. In pharmacokinetic studies of drug bioavailability and disposition, blood samples are collected from the jugular vein of all species apart from pigs, in which the anterior vena cava is generally used. It is very unlikely that the vein used for collection of blood samples, with the exception of the vein used for injection of the drug, would influence the results obtained in a pharmacokinetic study involving the measurement of plasma drug concentrations over an extended period.

Because of practical considerations, intravenous infusion is used far less widely for drug administration to animals than to humans. Intravenous administration, especially constant-rate infusion, has two important advantages over other routes. It provides complete systemic availability of the drug and allows greater control over the intensity (magnitude) of the effects produced. The relatively short dosage interval associated with multiple dosing is the principal disadvantage of the use of intravenous injection in animals. Long-acting preparations administered by intramuscular injection constitute the parenteral dosage form most widely used in farm animals.

Intraosseous administration is a feasible alternative to intravenous injection of some antimicrobial agents (sodium ampicillin or amoxicillin, cefotaxime, gentamicin or amikacin sulfate) in neonatal foals (less than 7 days of age). This particularly applies in the treatment of septicemia in neonatal foals that are in a state of septic shock or dehydration or both. The plasma concentration profiles for amikacin administered intraosseously and intravenously to neonatal foals are similar (Fig. 4) (39).

PARENTERAL DOSAGE FORMS

Parenteral dosage forms for use in animals include aqueous, aqueous organic and oily solutions, emulsions,

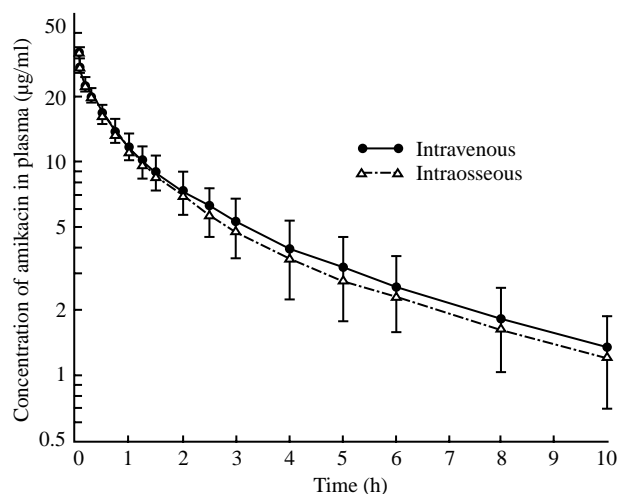


Fig. 4 Concentrations (mean \pm S.D.) of amikacin in plasma after i.v. and i.o. administration of amikacin sulphate (7 mg/kg body weight) to 6 foals at 3 to 5 days of age.

aqueous and oily suspensions for injection, and modified-release devices for subcutaneous implantation. Parenteral preparations must be sterile and pyrogen-free; liquid formulations should, if possible, be buffered close to physiological pH and preferably be isotonic with body fluids. The advantages and disadvantages associated with the various types of parenteral dosage forms have been addressed elsewhere (40).

The formulation of an extravascularly injected parenteral preparation primarily determines the plasma concentration profile of the drug. A drug is immediately available for absorption only when administered as an aqueous solution and when neither precipitation nor tissue irritation occurs at the injection site. Gentamicin sulfate (50 mg/ml aqueous solution) is rapidly and completely absorbed in that the peak plasma concentration is attained within 30–60 min, and systemic availability exceeds 90% from intramuscular injection sites in horses, dogs, and cats. After intravenous, intramuscular, and subcutaneous injection of gentamicin (3 mg/kg) in dogs, the peak plasma concentration (10.7 μ g/ml i.m.; 10.2 μ g/ml s.c.) and absolute bioavailability (96% i.m.; 94.3% s.c.) of the antibiotic provided by the intramuscular and subcutaneous routes were similar (41). Administration of amikacin sulfate (50 mg/ml aqueous solution) by subcutaneous injection at three dose levels (5, 10, and 20 mg/kg) to dogs showed that area under the curve, which reflects the extent of absorption, was proportional to the dose administered (42). After intramuscular injection of ceftiofur sodium (as reconstituted aqueous solution) at dose levels of 1.1 and 2.2 mg of ceftiofur free acid

equivalents per kilogram of body weight to sheep, the peak plasma concentrations (C_{max}) of ceftiofur and metabolites (measured as desfuoylceftiofur acetamide by HPLC) were 4.33 and 7.13 μ g/ml, the apparent half-lives were 6.5 and 7.65 h, respectively, and the area under the curve (from time zero to the limit of quantification of the assay) was proportional to the dose administered (43). The drug was rapidly (t_{max} , 0.5–1 h) and completely absorbed from the intramuscular injection site. Ketamine hydrochloride (100 mg/ml aqueous solution) is rapidly absorbed from the intramuscular injection site in cats, but pain may be evident before onset of the anesthetic effect. In horses, dogs and cats, ketamine administration is preceded by sedative premedication with an α_2 -adrenoceptor agonist (detomidine, medetomidine, or xylazine). Parenteral solutions of an irritant nature (thiopental sodium, ticarcillin sodium-clavulanate potassium combination, cefuroxime, cefotaxime, ceftriaxone, flunixin meglumine) and digoxin injection should be administered only by the intravenous route. Formulations containing sparingly soluble drugs in a water-miscible solvent, such as propylene glycol, may cause precipitation of drug at the intramuscular injection site. This makes these preparations unsuitable for intramuscular administration (e.g., diazepam, phenytoin). Diazepam injection (5 mg/ml solution or emulsion) may be administered to dogs and cats by intravenous injection and, after appropriate dilution, by intravenous infusion.

In parenteral preparations containing a poorly soluble salt of a drug, availability of the drug for absorption, which is dependent on dissolution rather than on the absorption process per se, generally controls the rate of absorption, the peak plasma concentration, and the length of time over which effective plasma concentrations are maintained. Long-acting parenteral preparations are formulated in a nonaqueous vehicle (such as an oil), or a poorly soluble salt of the drug is used (usually an aqueous suspension). These preparations provide slow absorption of the drug over an extended period owing to its gradual or staged availability for absorption (44). Avermectins and milbemycins (macrocyclic lactones) are highly lipophilic substances that determine the extent of their distribution and deposition in body fat, whereas the formulation of parenteral dosage forms (for subcutaneous injection in cattle or sheep, but not horses) influences the plasma concentration profile after the administration of a single dose (200 μ g/kg). The commercially available preparations differ in formulation; ivermectin is a nonaqueous preparation (60% propylene glycol/40% glycerol formal), doramectin is an oil-based preparation containing sesame oil/ethyl oleate (90/10), and moxidectin is an aqueous-based solution. The pharmacokinetic parameters describing the rates of absorption and elimination of these

Table 5 Pharmacokinetic parameters describing the rate of absorption and elimination of ivermectin (IVM), doramectin (DRM), and moxidectin (MXD) after subcutaneous injection (shoulder region) of single doses (200 µg/kg) of the commercially available preparations to 10-month-old hereford calves (180–210 kg body weight)

Kinetic parameter	IVM	DRM	MXD
C _{max} (ng/ml)	42.8 ± 3.8	37.5 ± 3.9	39.4 ± 3.4
t _{max} (days)	4.00 ± 3.94 ^a	6.00 ± 1.35	0.32 ± 0.0 ^c
t _{1/2} (days)	17.2 ± 4.26 ^b	6.25 ± 0.16	14.5 ± 1.20 ^c
Cl _B /F(ml/day·kg)	457 ± 52.5 ^a	322 ± 164	938 ± 62.5 ^c

^aMean kinetic parameters for IVM are significantly different from those obtained for MXD at $P < 0.05$.

^bMean kinetic parameters for IVM are significantly different from those obtained for DRM at $P < 0.05$.

^cMean kinetic parameters for MXD are significantly different from those obtained for DRM (c) at $P < 0.05$.

Results are expressed as mean ± sem ($n = 4$).

endectocides after administration by subcutaneous injection (shoulder region) in 10-month-old Hereford calves are compared in Table 5 (45). Because the systemic availability (F) of the drugs was not determined in this study, the term clearance/systemic availability (Cl_B/F) is used. Comparison of three commercially available parenteral preparations (one conventional and two long-acting) of oxytetracycline (20 mg/kg) injected intramuscularly in the lateral neck of pigs showed statistically significant differences between the preparations in peak plasma concentration (C_{max}), time of peak concentration (t_{max}), and mean residence time (MRT), whereas area under the curve (AUC) did not significantly differ among the preparations (Table 6) (46). The results indicate that a 24-h dosage interval should be used for the conventional preparation, whereas a 48-h interval would be appropriate for either of the long-acting preparations. A single intramuscular dose (20 mg/kg) of a long-acting preparation of oxytetracycline provides plasma

concentrations above 0.5 µg/ml for 48 h in pigs, ruminant calves, cattle, goats, red deer (*Cervus elaphus*), fallow deer (*Dama dama*), and camels (*Camelus dromedarius*). The commercially available long-acting preparations of oxytetracycline should not be administered to *Equidae* (horses, ponies, and donkeys). When comparing parenteral preparations of an antimicrobial agent on the basis of potential clinical efficacy, it is generally useful to compare the areas under the inhibitory plasma concentration–time curves (AUIC = AUC/MIC₉₀) for the duration of the recommended dosage interval because this term indicates the degree of exposure of susceptible micro-organisms to the drug.

The plasma concentration profiles can vary widely among different parenteral preparations of the same drug administered by intramuscular injection, even when the same injection site is used. Comparison of the plasma concentration profiles obtained after intramuscular injection

Table 6 Pharmacokinetic parameters describing the absorption and disposition of three oxytetracycline formulations administered intramuscularly (lateral neck) to pigs ($n = 8$) at a dose of 20 mg/kg body weight

Pharmacokinetic term	Product A	Product B	Product C
C _{max} (µg/ml)	6.27 ± 1.47	5.77 ± 1.0	4.68 ± 0.61
t _{max} (h)	3.0 (2.0–4.0)	0.5 (0.083–2.0)	0.5 (0.083–2.0)
AUC (µg h/ml)	79.22 ± 25.02	91.53 ± 20.84	86.64 ± 14.21
MRT (h)	11.48 ± 2.01	25.27 ± 9.22	37.66 ± 15.62
C _{p(24 h)} (µg/ml)	0.81 ± 0.34	1.01 ± 0.26	0.97 ± 0.29
C _{p(48 h)} (µg/ml)	< LOQ	0.40 ± 0.17	0.50 ± 0.09

LOQ = limit of quantification (0.1 µg/ml).

Product A, Engemycine 10% in polyvinylpyrrolidone.

Product B, Oxyter LA 20% in dimethylacetamide.

Product C, Terramycin LA 20% in pyrrolidone-2 and polyvinylpyrrolidone.

Results are expressed as mean ± S.D.

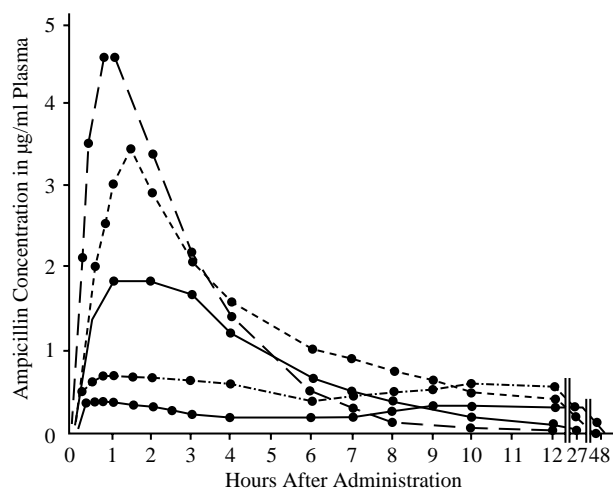


Fig. 5 Mean plasma ampicillin concentrations after i.m. injection of 5 different parenteral ampicillin formulations at similar dose levels (7.7 ± 1.0 mg/kg) to 5 calves. (From Ref. 47.)

tion in the lateral neck of ruminant calves of five different parenteral preparations of ampicillin at a similar dose level (7.7 ± 1.0 mg/kg) serves to illustrate the variation that can exist among preparations (Fig. 5) (47). Estimation of the bioavailability of each preparation relative to a reference preparation is useful. A crossover design incorporating an appropriate washout period between the phases of the bioavailability study should be used whenever feasible.

The drug concentration in a parenteral suspension can influence the plasma concentration profile. When different concentrations of amoxicillin (100 and 200 mg/ml) in aqueous suspensions of amoxicillin trihydrate were administered intramuscularly at the same site and same dose level (10 mg/kg) to horses, the preparation of lower concentration (10%) provided relatively better absorption and a more consistent plasma concentration profile. Intramuscular injection of amoxicillin trihydrate (15% in a mixed oil base) in the neck (10 cm behind the ear) of pigs produced two peaks, 1.7 and 0.8 µg/ml at 1.3 and 6.6 h, respectively, rather than a single peak in the plasma concentration profile and an eight-fold longer mean residence time of the antibiotic than a preparation of the same concentration in oil (48). The two peaks in the plasma concentration profile could be ascribed to the mixture of oil vehicles releasing amoxicillin at different rates.

Age and body weight (45, 55, and 62, kg) of calves was shown to influence the systemic availability of amoxicillin administered intramuscularly as amoxicillin trihydrate (10% aqueous suspension). When the same preparation was administered intramuscularly to different animal species, the trend was for smaller animals (piglets, dogs, cats) to show an early high peak concentration followed by

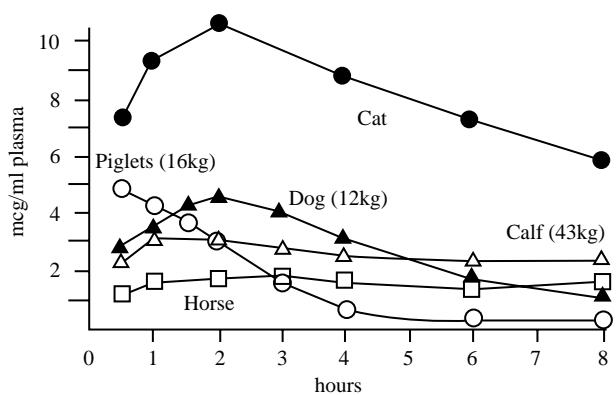


Fig. 6 Effect of species and weight on the bioavailability of amoxicillin after i.m. injection of amoxicillin trihydrate aqueous suspension (100 mg/ml) at the same dose (7 mg/kg) in the various species except cats (10–12 mg/kg).

a rapid decline, whereas larger animals (calves, horses) show a lower and relatively constant plasma concentration of amoxicillin over at least an 8-h period (Fig. 6) (49).

BOVINE MAMMARY GLAND

Antimicrobial agents, as with other drugs, cross the blood–milk barrier, which is a somewhat restrictive functional rather than anatomical barrier, primarily by passive diffusion. Both nonpolar lipid-soluble compounds and polar substances with sufficient lipid solubility passively diffuse through the predominantly lipoidal blood–milk barrier (50). At a moderate level of milk production, the ratio of the volume of blood circulating through the mammary gland to the volume of milk produced has been estimated to be 670:1; this provides ample opportunity for drugs to diffuse passively from the systemic circulation into milk. The composition (g/dL) of milk varies among breeds of cow in protein, from 3.1 (Holstein) to 3.9. (Jersey, Zebu, and buffalo) and fat, from 3.5 (Holstein) to 5.5 (Jersey, Zebu, and buffalo) content.

The milk-to-plasma equilibrium concentration ratio of total (nonionized plus ionized) drug is determined by the degree of ionization of the drug, which is pK_a/pH -dependent, in blood and milk, the charge on the ionized moiety, and the extent of binding to plasma proteins and milk macromolecules. It has been shown that only the lipid-soluble, nonionized moiety of a weak organic acid or base that is free (not bound to proteins) in the plasma can diffuse through cellular barriers and enter the milk (51). In normal lactating cows (milk pH range 6.5–6.8), weak organic acids attain milk ultrafiltrate-to-plasma

ultrafiltrate equilibrium concentration ratios less than 1; oxytetracycline and rifampin, amphoteric drugs with moderate and high lipid solubility, attain equilibrium concentration ratios of 0.75 and approximately 1, respectively; weak organic bases, apart from aminoglycosides, spectinomycin, and polymyxin B (drugs with low solubility in lipid) attain milk ultrafiltrate-to-plasma ultrafiltrate concentration ratios greater than 1 (Table 7) (52). The high concentration ratios attained by lipophilic organic bases (macrolides, lincosamides, trimethoprim) are attributed to the ion-trapping effect in acidic milk. Enrofloxacin and its active metabolite ciprofloxacin, formed by *N*-deethylation (a microsomal-mediated oxidative reaction in the liver), would be expected to attain concentrations in milk that would be effective against susceptible Gram-negative aerobic bacteria, in particular *Escherichia coli*.

In the presence of mastitis, the pH of milk increases to within the range of 6.9 to 7.2. As a consequence, the ion-

trapping effect on lipophilic organic bases is reduced, whereas the concentrations attained by weak organic acids are somewhat increased. The higher pH of milk does not change the concentrations attained by amphoteric drugs (fluoroquinolones, tetracyclines, rifampin); however, antimicrobial activity of these drugs is often lower in milk than in extracellular fluid or in vitro determination would predict.

Antimicrobial therapy is generally applied in the treatment of clinical mastitis during lactation and in treating subclinical mastitis at the end of lactation, whereas the implementation of preventive measures is essential for decreasing the incidence of mastitic infection in the dairy herd. The common causative pathogenic microorganisms of clinical mastitis are *Streptococcus uberis*, coliforms (*E. coli*, *Klebsiella* spp.), *S. aureus*, *S. dysgalactiae*, and *S. agalactiae*. It is usual to treat clinical mastitis both systemically using a parenteral antimicrobial preparation and locally with a quick-release intramammary prep-

Table 7 Comparison of calculated and experimentally obtained milk: plasma concentration ratios for antimicrobial agents under equilibrium conditions

Drug	Lipid solubility	pK _a	Milk pH	Concentration ratio (Milk ultrafiltrate: plasma ultrafiltrate)	
				Theoretical	Experimental
Acids					
Penicillin G	Low	2.7	6.8	0.25	0.13–0.26
Cloxacillin	Low	2.7	6.8	0.25	0.25–0.30
Ampicillin	Low	2.7, 7.2	6.8		0.24–0.30
Cephaloridine	Low	3.4	6.8	0.25	0.24–0.28
Cephaloglycin	Low	4.9	6.8	0.25	0.33
Sulfadimethoxine	Moderate	6.0	6.6	0.20	0.23
Sulfadiazine	Moderate	6.4	6.6	0.23	0.21
Sulfamethazine	Moderate	7.4	6.6	0.58	0.59
Rifampin ^a	Moderate/high	7.9	6.8	0.82	0.90–1.28
Bases					
Tylosin	High	7.1	6.8	2.00	3.5
Lincomycin	High	7.6	6.8	2.83	3.1
Spiramycin	High	8.2	6.8	3.57	4.6
Erythromycin	Very high	8.8	6.8	3.87	8.7
Trimethoprim	High	7.3	6.8	2.32	2.9
Aminoglycosides	Low	7.8 ^b	6.8	3.13	0.5
Spectinomycin	Low	8.8	6.8	3.87	0.6
Polymyxin B	Very low	10.0	6.8	3.97	0.3
Amphoteric					
Oxytetracycline	Moderate	—	6.5–6.8	—	0.75
Doxycycline	Moderate/high	—	6.5–6.8	—	1.53

^aRifampin is an amphoteric drug substance; theoretical concentration ratio was based on its behavior as an organic acid.

^bpK_a value given for aminoglycosides is unconfirmed.

aration. The infusion via the teat canal of an intramammary preparation alone would be inadequate for the treatment of moderate to severe infection because of the decreased ability of an infused drug to ascend partially occluded milk ducts and the requirement, particularly in coliform mastitis, for frequent milkout (stripping) of the infected quarter of the mammary gland. In mild cases of mastitis diagnosed at an early stage of infection, the infusion of a quick-release intramammary preparation may suffice without concurrent systemic treatment. Slow-release intramammary preparations are used at the end of lactation (after the last milking) to treat subclinical mastitis and to prevent the establishment of new infections, including summer mastitis commonly caused by *Actinomyces pyogenes* during the nonlactating (dry) period. Because *S. aureus* is the principal causative microorganism of subclinical mastitis, the slow-release intramammary preparation selected for treatment should contain an antimicrobial effective against all strains of the bacterium.

PARENTERAL PREPARATIONS

An “ideal” antimicrobial agent for systemic therapy of bovine mastitis should possess the following properties (53):

1. low minimum inhibitory concentration (MIC) for the majority of mastitis-causing pathogenic microorganisms;
2. high systemic availability after intramuscular injection;
3. lipid-soluble and predominantly nonionized in the blood and have a low degree of binding to plasma proteins;
4. long apparent half-life to ensure that concentrations above (preferably severalfold) the MIC are maintained at the site of infection in the mammary gland throughout the recommended dosage interval (12 h is desirable);
5. minimal adverse effects in cows treated at effective dosage; and
6. short withdrawal periods (milk and slaughter).

Parenteral preparations with antimicrobial activity, which depends on the causative pathogenic microorganism, and pharmacokinetic properties that meet most of these criteria include procaine penicillin G (aqueous suspension), amoxicillin trihydrate–clavulanate potassium combination (aqueous suspension), and enrofloxacin (solution). Enrofloxacin is not approved for use in cows producing milk for human consumption. To attain effective concentrations in the mammary gland, oxytetracycline

hydrochloride (conventional preparation) has to be administered by slow intravenous injection. Even though macrolide antibiotics attain high concentrations in milk, they also diffuse passively into ruminal fluid (pH 5.5–6.5) where the ion-trapping effect applies. This feature of their distribution may be undesirable. Moreover, slow intravenous injection is the preferred manner of administration because the available parenteral preparations cause tissue irritation at intramuscular injection sites. Because spiramycin binds avidly to tissue components, long withdrawal periods would be associated with the use of this antimicrobial agent.

INTRAMAMMARY PREPARATIONS

Intramammary preparations are formulated to provide either quick release of the antimicrobial agent or slow release of the antimicrobial over an extended period. A requirement of all intramammary preparations is that they be reasonably nonirritating to the parenchyma (epithelial tissue) of the udder. Quick-release preparations are used primarily in lactating cows for the treatment, often in conjunction with systemic therapy, of clinical mastitis. They should have short withdrawal periods. The vehicle used and viscosity of the formulation should allow rapid release of the antimicrobial while ensuring that effective concentrations will be maintained throughout the recommended dosage interval. Access of the released antimicrobial to the site of infection is determined by its uptake and distribution in mammary tissue, which are governed by the chemical nature and physiochemical properties (in particular, lipid solubility) of the drug. Binding to milk proteins or components of mammary tissue limits distribution and extends the withdrawal periods. The transfer of an antimicrobial agent from treated to untreated quarters of the udder takes place via the bloodstream and involves passive diffusion in both directions across the blood–milk barrier. Examples of intramammary preparations formulated as suspensions and having a recommended dosage interval of 12 h include cloxacillin sodium, ampicillin sodium–cloxacillin sodium combination, trimethoprim–sulfadiazine combination, and oxytetracycline hydrochloride (oily suspension), whereas erythromycin is formulated as an intramammary solution. Cefuroxime sodium and cefoperazone sodium (third-generation cephalosporins) are formulated as an oily paste and oily suspension, respectively. Quick-release intramammary preparations have short withdrawal periods, typically slaughter, 7 days and milk-withholding, 3.5 days, or either may be shorter depending on the preparation.

Slow-release intramammary preparations may be infused at the end of lactation (after the last milking) and into the teat canal of nonlactating cows to treat subclinical mastitis and to prevent the establishment of new infections during the nonlactating period. Either a poorly soluble salt of an antimicrobial agent may be used or the formulation of the preparation be such that the rate of antimicrobial release is relatively constant, approaching zero-order. The antimicrobial must remain active (be stable) throughout the extended duration in the udder, and the preparation should not cause tissue irritation. Antimicrobial binding to mammary tissue components is not of particular concern because slow-release preparations are not used in lactating cows. However, the ability to penetrate cell membranes is important because, in chronic staphylococcal mastitis, the pathogenic bacteria often reside within epithelial cells, neutrophils, and macrophages (54). Examples of slow-release intramammary preparations include cloxacillin (benzathine salt) with aluminium monostearate (suspension), ampicillin (trihydrate) and cloxacillin (benzathine) with aluminium monostearate (suspension) or formulated without aluminium monostearate as an oily suspension, procaine penicillin G (oily paste), dihydrostreptomycin sulfate and procaine penicillin G (oily paste). Penicillins and especially aminoglycosides have limited ability to penetrate cell membranes, whereas fluroquinolones, rifampin, and macrolides have this capacity.

Compound preparations containing one or more antimicrobial agents and a corticosteroid (hydrocortisone or prednisolone) are available for intramammary infusion in lactating cows. The reduction in inflammation of the mammary gland is desirable; however, the immunosuppressant effect and decrease in phagocyte function produced by glucocorticoids are undesirable. Amoxicillin trihydrate–clavulanate potassium combination with prednisolone is an example of a compound preparation (oily suspension) that can be administered by intramammary

infusion at 12-h intervals and has short withdrawal periods (slaughter, 7 days; milk, 2 days). Unlike glucocorticoids, the nonsteroidal anti-inflammatory drugs do not cause immunosuppression. Flunixin meglumine (2.2 mg/kg) administered by intravenous injection at 24-h dosage intervals may have a place in the treatment of acute *E. coli* (endotoxin) mastitis. The significance of the antipyretic, anti-inflammatory, and analgesic effects produced by the drug is primarily dependent on the stage of the inflammatory process at which treatment is begun. Early diagnosis of coliform mastitis and prompt initiation of treatment with flunixin greatly increase the beneficial effect of the drug. The half-life of flunixin in cows is 8.1 h and the withdrawal periods are short (slaughter, 7 days; milk, 12 h). Flunixin does not interfere with the activity of concurrently administered antimicrobial agents, should antimicrobial therapy be applied. There is no evidence to support the contention that systemic antimicrobial therapy causes a massive release of endotoxin in cows with coliform mastitis. Frequent stripping of the infected quarter(s) to remove bacteria and cellular debris is important, perhaps essential, in coliform mastitis (55). The slow intravenous injection of oxytocin, 5–10 U of diluted solution (10 U/ml), facilitates the completeness of stripping (milkout).

THERAPEUTIC CONCENTRATIONS

The principal pharmacological effects produced are generally associated with the same range of therapeutic plasma concentrations in domestic animals as in humans (Table 8). The mechanisms of action of drugs appear to be the same in mammalian species. The calculation of a dosage regimen (dose and dosage interval) for a drug preparation is based on a knowledge of the therapeutic concentration range and the pharmacokinetic parameters

Table 8 Principal pharmacological effect and range of therapeutic plasma concentrations of some drugs

Drug	Pharmacological effect	Therapeutic concentrations
Quinidine	Antiarrhythmic	0–6 µg/ml
Procainamide	Antiarrhythmic	6–14 µg/ml
Lignocaine	Antiarrhythmic	1.5–5.0 µg/ml
Propranolol	Antihypertensive	20–80 ng/ml
Digoxin	Positive inotropic	0.6–2.4 ng/ml
Phenobarbitone	Anticonvulsant	10–25 µg/ml
Pethidine	Analgesic	0.4–0.7 µg/ml
Theophylline	Bronchodilator	6–16 µg/ml

that describe bioavailability and disposition of the drug. Complete systemic availability can be assumed only when a drug is administered intravenously. Species differences in the dosage regimen for a drug preparation administered by a particular route can generally, but not always, be attributed to variation among species in pharmacokinetic behavior of the drug. Species differences in susceptibility (dosage requirement) to certain drugs (morphine, xylazine) appear to be attributable to pharmacodynamic variation (affinity and/or efficacy).

Dosage of anthelmintics, which usually refers to a single dose, is based on semiquantitative assessment of clinical efficacy, although, when appropriate, bioavailability and the plasma concentration profile of the active moiety (parent drug and/or active metabolite) are considered in the development of anthelmintic dosage forms and selection of their routes of administration for various animal species (56). Because anthelmintics are administered in different dosage forms to ruminant species (cattle, sheep, and goats), pigs, horses, and small animals (dogs and cats), a wide variety of preparations are commercially available for use in animals.

DRUG DISPOSITION

Species variations in the disposition of a drug may be attributed to differences in the extent of distribution and/or the rate of elimination of the drug. Variations are usual between ruminant and monogastric species and between herbivorous and nonherbivorous species, depending on the classification system used for mammalian species. The rate of elimination of drugs differs widely between homeothermic and poikilothermic species. Little is known regarding drug distribution in poikilothermic species (fish and reptiles).

EXTENT OF DISTRIBUTION

Significant differences in the extent of distribution of drugs, particularly lipid-soluble organic bases, are usual between ruminant and monogastric species. After parenteral administration, lipophilic bases diffuse passively from the systemic circulation into ruminal fluid (pH 5.5–6.5), where they become “trapped” by ionization. These drugs are slowly reabsorbed or, if they possess functional groups suitable for metabolism by hydrolysis or reduction, they may be partially inactivated by ruminal micro-organisms.

The volume of distribution at steady state ($V_{d(ss)}$) is useful for determining the significance of changes in the extent of distribution of a drug in the presence of disease states or by plasma protein binding displacement (drug interaction) and for comparing the extent of drug distribution in neonatal and adult animals of the same species. The volume of distribution at steady state, unlike $V_{d(area)}$, is at least theoretically independent of changes in the rate of drug elimination (57).

PLASMA PROTEIN BINDING

Although the extent of plasma protein binding of a drug varies among domestic animal species, the range of binding is reasonably narrow in the collective species, and the difference among individual species is often not of clinical significance. Species variation in the binding of acidic drugs could be attributed to differences in the conformation of plasma albumin; the lower concentration of plasma albumin in birds than in mammals may generally account for the lower binding of acidic drugs in birds.

ELIMINATION PROCESSES

In domestic animal species, the liver constitutes 1.25–2.5% of live body weight, and hepatic arterial blood flow represents 26–29% of cardiac output. The hepatic portal vein contributes additionally to liver blood flow. The kidneys constitute 0.25–0.6% of live body weight and receive 22–24% of cardiac output. In avian species and reptiles, the renal portal system contributes to blood flow to the kidneys. In proportion to organ weight (mass), the kidneys are more richly perfused with blood than is the liver.

The liver metabolizes drugs, other foreign chemical compounds (xenobiotics), and certain endogenous substances (e.g., steroid hormones, bilirubin) by a variety of pathways. They include hepatic microsomal-mediated oxidative reactions, reductive and hydrolytic reactions (phase I), and conjugation (synthetic) reactions with various endogenous substances (phase II) (58, 59). Phase I metabolic reactions occur ubiquitously and are qualitatively similar in mammals, birds, and fish, but differ widely and unpredictably among species in the rates at which they take place. Although molecular structure is the primary determinant of the phase I reaction that will most likely occur, at least a moderate degree of lipid solubility is required for microsomal oxidation. The situation is

Table 9 Domestic animal species with defects in certain conjugation reactions

Species	Conjugation reaction	Major target groups	State of synthetic reaction
Cat	Glucuronide synthesis	–OH, –COOH, –NH ₂ , =NH, –SH	Present, slow rate
Dog	Acetylation	Ar–NH ₂	Absent
Pig	Sulfate conjugation	Ar–OH, Ar–NH ₂	Present, low extent

(From Ref. 50.)

different with regard to phase II metabolic reactions because some of these are either defective or absent in certain species, which makes the final pathways of metabolism somewhat predictable (Table 9) (50). Requirements of phase II reactions include the presence in a drug molecule (either parent drug or phase I metabolite) of a functional group that is suitable for undergoing conjugation, an endogenous reacting substance (conjugating agent), and a transferring enzyme. The major transferring enzymes are UDP-glucuronyltransferase, sulfotransferase, *N*-acetyltransferase, GSH-*S*-transferase, and methyltransferase. Conjugates of drugs are polar, less lipid-soluble, and consequently less widely distributed extravascularly than the parent drug or phase I metabolite, and the vast majority are pharmacologically inactive. Apart from a few exceptions (e.g., the *N*⁴-acetyl derivative of most sulfonamides), drug conjugates are more water-soluble than the parent drug. Glucuronide conjugates are especially suitable for carrier-mediated active excretion in urine or bile or both. Species variations in conjugate formation could be attributed to the availability of the conjugating agent, the ability to form the “activated” nucleolide, or the activity of the transferring enzyme (60). Fish may have a low capacity to form the activated nucleolide uridine-diphosphate–glucuronic acid, which would limit glucuronide synthesis. Prodrugs belong to a different category of drug, in that prodrugs per se are devoid of activity but are metabolically converted by phase I, usually hydrolytic reactions to the active parent drug. Examples of orally administered prodrugs include enalapril (enalaprilat), pivampicillin (ampicillin), netobimin (albendazole), and febantel (fenbendazole). Metabolic conversion (activation) of these prodrugs occurs presystemically in that it is brought about by ruminal micro-organisms or takes place during passage through the intestinal mucosa or the liver (first-pass metabolism).

Ruminal microorganisms are capable of at least partially inactivating orally administered drugs (chloramphenicol, trimethoprim, nitroxylin, digoxin) by hydrolytic and reductive reactions. The susceptibility of ruminant animals to toxicity caused by ingestion of plants containing cyanogenetic glycosides (e.g., *Prunus* spp., *Acacia* spp.,

Eucalyptus cladocalyx) is attributed to ruminal hydrolysis of the glycosides with liberation of hydrogen cyanide. Intestinal micro-organisms containing β -glucuronidase can reactivate (by hydrolysis) glucuronide conjugates of drugs excreted in bile. Renal excretion is the principal process for elimination of drugs that are predominantly ionized in the plasma, polar drugs, and drug metabolites. Extensive (>80%) binding to plasma proteins limits the availability of drugs for glomerular filtration but does not hinder carrier-mediated active tubular secretion. Although a drug may enter tubular fluid both by glomerular filtration and by proximal tubular secretion, its renal clearance may nonetheless be low when substantial reabsorption occurs in the distal nephron. Because reabsorption takes place by passive diffusion, it is influenced by the concentration of drug and its degree of ionization in distal tubular fluid and by the rate of passage of glomerular filtrate through the distal nephron. The degree of ionization is determined by the pK_a value of the drug and the urinary pH reaction. The usual urinary reaction in carnivorous species is acidic (pH 5.5–7.0), whereas in herbivorous species it is alkaline (pH 7.2–8.4). The excretion of weak organic acids is enhanced under alkaline and decreased under acidic urinary conditions; the converse applies to weak organic bases. Species variations in the rate of elimination of renally excreted drugs (penicillins, most cephalosporins, aminoglycosides, most diuretics, nondepolarizing neuromuscular blocking drugs) are primarily attributable to differences in the glomerular filtration rate (GFR). Based on inulin clearance, mean values of GFR (ml/min \times kg) are 3.96 in dogs; 2.94 in cats; 2.80 in pigs; 2.26, 2.25, and 2.20 in goats, cattle, and sheep, respectively; and 1.65 in horses; the estimated value of GFR in humans is 1.84 ml/min \times kg. Urinary pH influences the rate of elimination of drugs that are moderately lipid-soluble and undergo elimination both by hepatic metabolism and by renal excretion (amphetamine, trimethoprim, phenobarbital, sulfonamides).

The rate of hepatocyte secretion of bile in domestic animal species is 12–24 ml/kg per day; the lower end of the range applies to the dog and cat, and the upper end applies to the horse (a species that does not have a gall bladder). Compounds excreted in bile have molecular weights

exceeding 300 and a degree of polarity that enables them to be transported by a carrier-mediated process from hepatic parenchymal cells into bile. Drugs and drug metabolites (primarily conjugates) excreted in bile enter the duodenum, from which some (depending on their lipid solubility) may be reabsorbed by passive diffusion.

RATE OF ELIMINATION

Half-life is the pharmacokinetic parameter used to measure the overall rate of drug elimination. The half-life of most drugs that are primarily eliminated by hepatic biotransformation varies widely among species (Table 10). The usual trend is that half-life is shorter in cattle and horses (herbivorous species) than in dogs and cats (carnivorous species), whereas the half-life of several drugs is longer in humans than in domestic animals. There are, however, notable exceptions to this trend, such as the methylxanthines (caffeine and theophylline) in horses and phenylbutazone in cattle. The half-life of phenylbutazone in cattle varies from 42 to 66 h and, unlike in horses, dogs, and humans, does not appear to be dose-dependent. Because of differences in the rate of hepatic elimination (presumably by biotransformation), the half-life of various drugs (sulfamethazine, trimethoprim, ceftiofur/desfuroylceftiofur, closantel, clorsulon) is shorter in goats (especially dwarf goats) than in sheep, whereas the half-life of some drugs (phenylbutazone, norfloxacin) is shorter in donkeys than in horses and ponies.

Because phenytoin, valproate, carbamazepine and clonazepam have a much shorter half-life in dogs than in humans (Table 11), the dosage intervals that would be required for anticonvulsant effectiveness in dogs make conventional dosage forms of these drugs impractical for therapeutic use (61). The half-life of phenobarbital and naproxen is 2-fold longer in mongrel (mixed-breed) dogs than in Beagles. The long half-life of naproxen in dogs (74 h in mongrels, 35 h in Beagles) is unusual for a nonsteroidal anti-inflammatory drug. Naproxen half-life is 13.9 h in humans, 8.3 h in horses, 4.8 h in minipigs, and 1.9 h in Rhesus monkeys.

The half-life of salicylate, which is primarily eliminated by glucuronide conjugation, is 25–35 h and dose-dependent in cats, compared with 10–15 h in humans, 8.6 h in dogs, 1 h in horses, and 0.8 h in cattle. A relative deficiency of hepatic microsomal glucuronyl-transferase activity appears to be characteristic of *Felidae*, because it applies not only to the domestic cat (*Felis catus*) but also to the lion (*Panthera leo*), African civet (*Viverra civetta*), and forest genet (*Genetta pardina*) (62). Even though dogs and foxes (*Canidae*) are unable to acetylate aromatic (ArNH_2) and hyrazine (RNHNH_2 or ArNHNH_2) amino groups, the absence of this metabolic pathway does not appear to delay the elimination of drugs with these functional groups (sulfonamides, procainamide). Either a larger fraction of the systemically available dose is excreted as parent drug in the urine or an alternative metabolic pathway compensates for the absence of acetylation of these amino groups. Amino acid conju-

Table 10 Species variations in the half-life of some drugs that are primarily eliminated by hepatic metabolism

Drug	Cattle	Horse	Dog	Human
Pentobarbitone	0.8	1.5	4.5	22.3
Thiopentone	3.3	2.5	8.3	11.5
Salicylate	0.8	1.0	8.6	12.0
Phenylbutazone	42–66	4.1–4.7 ^a	2.5–6.0 ^a	72.0 ^a
Flunixin	6.9	1.9	3.7	—
Morphine	—	1.0	0.95	1.9
Ketamine	0.9	0.7	1.0	2.5
Caffeine	3.8	18.2	4.25	4.9
Theophylline	6.9	14.8	5.7	9.0
Norfloxacin	2.4	6.4	3.6	5.0
Enrofloxacin	1.7	5.0	3.4	—
Chloramphenicol	3.6	0.9	4.2	4.6
Metronidazole	2.8	3.9	4.5	8.5
Trimethoprim ^b	1.25	3.2	4.6	10.6
Sulphadiazine ^b	2.5	3.6	5.6	9.9
Sulphadimethoxine ^b	12.5	11.3	13.2	40

^aHalf-life is dose-dependent.

^bHalf-life may be influenced by urinary pH reaction.

Table 11 Comparison of the average half-life (h) after intravenous administration (apart from carbamazepine) of a single dose of anticonvulsant drugs in dogs and humans

Drug	Dog	Human	Therapeutic range of plasma concentrations (µg/ml)
Phenobarbitone	64	96	10–25
Phenytoin	3.5–4.5 ^a	15–24 ^a	10–20
Sodium valproate	2	14	40–100
Carbamazepine (PO)	1.5	15	4–10
Clonazepam	1.5–2.5 ^a	24–36 ^a	0.01–0.08
Diazepam	7.6 ^b	32.9 ^b	>0.15

^aHalf-life is dose-dependent.^bParent drug and active metabolites.

gation in domestic animal species often involves glycine as the endogenous reacting substance. Glycine conjugation takes place in both the liver and the kidneys of domestic animals other than the dog, in which it occurs only in the kidneys. The transferring enzyme, which is located in mitochondria, is acyl-CoA glycinetransferase. Glycine is replaced by ornithine as the conjugating agent in birds classified as anseriformes (ducks, geese) and galliformes (chickens, turkeys) but not in columbiformes (pigeons, doves) (59).

Even though exceptions exist, which is inevitable in view of the variety of metabolic pathways, the half-life of the majority of drugs that undergo extensive hepatic biotransformation is shorter in laboratory animal species (mice, rats, guinea pigs, rabbits) and Rhesus monkeys than in domestic animal species and humans. Consistent with this trend is that the average half-life of antipyrine, a marker substance used to indicate hepatic microsomal oxidative activity, is shorter in laboratory animals (0.2–1.4 h) and Rhesus monkeys (1.2 h) than in domestic animals (1.75–3.25 h) and especially humans (10.3–12.7 h). It has been hypothesized (63) that the lesser capacity of humans to metabolize drugs by hepatic microsomal oxidation may be correlated with their enhanced longevity, expressed as maximum lifespan potential. Based on an equation using brain weight and body weight, maximum lifespan potential is estimated to be 93 y in humans, 39 y in horses, 20 y in dogs, 14 y in cats and, for comparative purposes, 77 y in African elephants.

As with the situation in mammals, wide variations exist among avian species in the half-life of drugs that are primarily eliminated by hepatic biotransformation (64). The half-life of antimicrobial agents is prolonged in poikilothermic species (fish and reptiles), which is consistent with their much lower metabolic turnover rate (65), and is influenced by ambient (in the case of fish, water) temperature (Table 12). The rate of drug elimination

increases (i.e., half-life decreases) with increase in ambient temperature and varies among fish species.

Species variations in the half-life of drugs that are eliminated by renal excretion is less pronounced than for lipid-soluble drugs that undergo extensive hepatic biotransformation. The half-life of gentamicin, which is eliminated solely by glomerular filtration, is 0.5–1 h in laboratory animals, 1.25–2.5 h in domestic animals, 2.75 h in humans, 1.25–3.4 h in various avian species, and 12 h in channel catfish (*Ictalurus punctatus*) acclimatized at 22°C. In mammalian species, the half-life of gentamicin reflects the relative (not actual) rate of glomerular filtration and is unrelated to urinary pH reaction. The volume of distribution of gentamicin is similar (250–300 ml/kg) in the various species.

The half-life of drugs that undergo a high degree of enterohepatic circulation may vary widely among species and may be relatively more prolonged in fish. The half-life of oxytetracycline is 80.3 h in African catfish (*Clarias gariepinus*) acclimatized at 25°C and 89.5 h in rainbow trout (*Salmo gairdneri*) at 12°C (66) compared with half-life of 9 h in humans, 3.4–9.6 h in domestic animals, and 1.3 h in rabbits. A combination of variables, which include the extent of distribution and the degree of enterohepatic circulation, contributes to species variation in the half-life of digoxin, which ranges from 7.8 h in cattle to 35 h in cats.

CLEARANCE

The systemic clearance of propofol, based on measurement of blood concentrations, in humans (31 ml/min × kg) and dogs (59 ml/min × kg) exceeds liver blood flow, which is 24 and 42 ml/min per kilogram of body weight in humans and dogs, respectively. It can be concluded that another organ (the lungs) or extrahepatic tissue contrib-

Table 12 Half-life of some antimicrobial agents in various species of fish

Antimicrobial agent	Fish species	Acclimatization temperature (°C)	Half-life (h)
Trimethoprim	Carp	10	40.7
	(<i>Cyprinus carpio</i> L.)	24	20.0
Sulphadiazine	Carp	10	47.0
	(<i>Cyprinus carpio</i> L.)	24	33.0
Sulphadimidine	Carp	10	50.3
	(<i>Cyprinus carpio</i> L.)	20	25.6
	Rainbow trout	10	20.6
Ciprofloxacin	(<i>Salmo gairdneri</i>)	20	14.7
	Rainbow trout	12	11.2
	(<i>Salmo gairdneri</i>)		
	Carp	20	14.5
	(<i>Cyprinus carpio</i>)		
Florfenicol	African catfish	25	14.2
	(<i>Clarias gariepinus</i>)		
	Atlantic salmon	10.8 ± 1.5	12.2
Oxytetracycline	(<i>Salmo salar</i>)	(Sea water)	
	Rainbow trout	12	89.5
	(<i>Salmo gairdneri</i>)		
	African catfish	25	80.3
Gentamicin	(<i>Clarias gariepinus</i>)		
	Channel catfish	22	12.0
	(<i>Ictalurus punctatus</i>)		

utes to the elimination (metabolism) of propofol. The blood propofol concentration at which dogs return to the sternal position and human beings regain consciousness appears to be the same (1 µg/ml).

When applied to conventional (immediate-release) aminophylline tablets, the dosing rates that would provide an average steady-state plasma theophylline concentration of 10 µg/ml and produce a sustained bronchodilator effect are 10 mg/kg administered at 8-h dosage intervals to dogs and 5 mg/kg administered at 12-h intervals to horses or cats. The systemic clearance of theophylline is 2.5 times higher in dogs (100 ml/h × kg) than in horses and cats (40 ml/h × kg), whereas systemic availability of the drug from the oral dosage form is similar (90–100%).

INTERSPECIES SCALING

The basic assumption in interspecies scaling is that physiological variables and biochemical processes are related to the body weight of mammalian species. The use of clearance, a physiologically based parameter, may be more appropriate than half-life (a hybrid parameter) for

interspecies allometric scaling of drug elimination. A double logarithmic plot of the pharmacokinetic parameter of interest versus body weight of the animal species is used to verify the linearity of the allometric relationship.

The systemic clearance of structurally unrelated drugs that undergo elimination by various processes shows a high degree of correlation with the body weight of several animal species (Table 13). Inulin (marker substance) and gentamicin are eliminated by glomerular filtration and ampicillin by glomerular filtration and proximal tubular secretion, whereas oxytetracycline undergoes enterohepatic circulation and consequently is slowly excreted by glomerular filtration. Enrofloxacin, theophylline, and antipyrine are poorly extracted by the liver, which implies that their clearance may be influenced both by the unbound fraction in blood and by the metabolic capacity of the liver and are eliminated by phase I hepatic metabolism. Enrofloxacin and theophylline are eliminated primarily by microsomal-mediated oxidative reactions, whereas antipyrine (marker substance) is eliminated entirely by microsomal oxidation. Interspecies scaling of antipyrine clearance identifies the human as a nonconforming species, which is attributed to the substantially lower microsomal oxidative capacity of humans relative to the 10 other species studied. Because of species variations in

Table 13 Allometric relationship between the clearance (ml/min) of some drugs and body weight (kg) of various mammalian species

Drug	Elimination process	No. of species	Allometric		Correlation coefficient
			Coefficient	Exponent	
Inulin	E (r)	8 ^a	4.13	0.86	0.989
Gentamicin	E (r)	8 ^a	2.60	0.86	0.970
Ampicillin	E (r)	8 ^b	5.47	0.94	0.959
Oxytetracycline	E (r)	8 ^c	7.96	0.73	0.978
Enrofloxacin	M (h)	8 ^d	12.53	0.93	0.984
Theophylline	M (h)	9 ^c	1.98	0.83	0.978
Antipyrine	M (h)	10 ^f	8.16	0.85	0.989

^a Cat, dog, goat, sheep, pig, human, cattle, horse.

^b Rabbit, dog, sheep, pig, human, donkey, cattle, horse.

^c Rabbit, dog, goat, sheep, pig, donkey, cattle, horse.

^d Rabbit, cat, dog, sheep, pig, ilama, horse, cow.

^e Rat, guinea pig, rabbit, cat, dog, pig, human, cattle, horse.

^f Mouse, rat, guinea pig, rabbit, monkey, dog, goat, sheep, pig, cattle (human is a nonconforming species).

For each drug, the level of significance is $P < 0.001$.

binding to plasma proteins, the use of hepatic intrinsic clearance of unbound drug, rather than systemic clearance, would represent a refinement of the allometric scaling technique. This refinement has been applied to interspecies scaling of antipyrine (67) and theophylline (68). Additional correction could be made for maximum lifespan potential (63). Because the clearance of highly extracted drugs that undergo extensive phase I hepatic metabolism is primarily determined by a single physiological variable (liver blood flow), interspecies allometric scaling should be feasible for drugs in this category. The index of drug elimination (half-life, systemic clearance, organ clearance of unbound drug) to use depends on the level of refinement required.

With regard to organ (liver, kidney, heart) weights, physiological variables (liver blood flow, renal function, cardiac output, basal oxygen consumption), and the pharmacokinetic parameters systemic clearance and volume of distribution, the numerical value of the allometric exponent is generally in the range of 0.67 to 1. For physiological periods (heartbeat time, duration of respiratory cycle), turnover times (serum albumin, total body water, blood circulation), and drug half-life, the allometric exponent is often close to 0.25, which represents that for energy expenditure in mammalian species (69) and the turnover time of endogenous processes (63).

The reliability of predictions based on interspecies allometric scaling of drug disposition depends on the use of a sensitive and precise analytical method for quantification of the drug (and active metabolite) in blood or plasma, a knowledge of the principal elimination

process for the drug, the use of at least four animal species representing a wide range of body weight (based on log body weight ratio), and the identification, for possible exclusion, of any nonconforming species. Because of the uniqueness of each animal species, application of the technique for predictive purposes should be limited to the preclinical stage of drug development.

IMPLICATIONS OF STEREOISOMERISM

Stereoisomerism has implications in the formulating of dosage forms and because a chiral environment exists within the body, in determining both the degree of activity and the disposition of racemates. The enantioselective behavior of drugs used in domestic animals was comprehensively reviewed by Landoni et al. (70).

Because stereoselective processes are species-related, the enantiomeric ratios of plasma concentrations at various times and areas under the plasma concentration-time curves may differ among animal species after the administration of a drug racemate. Chiral inversion, which occurs to a variable extent in different species, can be equivocally established only by administering individual enantiomers to the animal species of interest and measuring, using a sensitive stereospecific analytical method, the enantiomer administered and the optical antipode in biological fluids and tissues. The pharmacokinetic parameters based on plasma concentration-time data for each of the enantiomers can be statistically compared.

The 2-arylpropionic acid ("profen") nonsteroidal anti-inflammatory drugs, each of which contains a single chiral center, are formulated as racemic (50:50) mixtures of the S(+)- and R(-)- enantiomers, with the exception of naproxen, which is formulated as the S(+)-enantiomer. Based on inhibition of cyclo-oxygenase activity, the S(+)-enantiomer is the eutomer (more potent enantiomer). These drugs differ markedly in both pharmacodynamic activity and pharmacokinetic behavior and, in addition, enantiomer pharmacokinetics of each drug varies among animal species. After intravenous administration of racemic ketoprofen to horses, sheep, and 20-week-old calves and measurement of individual enantiomers in plasma, significant differences between the enantiomers were found in systemic clearance in horses and in both systemic clearance and volume of distribution in sheep (Table 14), whereas values of the pharmacokinetic parameters in calves did not differ between the enantiomers (71–73). The S(+)- to R(-) ratio of area under the curve was 1.35:1 in horses, 0.54:1 in sheep, and 1.05:1 in calves. The predominant enantiomer in plasma was S(+) in horses, R(-) in sheep, and both enantiomers were present in equal concentrations in calves. After the administration of each enantiomer separately to these species, the extent of chiral inversion from the R(-)- to S(+)-enantiomer was estimated to be 49% in horses, 5.9% in sheep, and 31% in calves (72, 74, 75). Unidirectional chiral inversion was estimated to be 49% in *Cynomolgus* monkeys (*Macaca*) (76), 9% in humans (77), and varied from 27 to 66% in laboratory animal species (78). The oral bioavailability of the S(+)-enantiomer of ketoprofen in Beagle dogs is not affected by the proportion of the R(-)-enantiomer in the oral dosage form, even though considerable (73%) metabolic inversion from the R(-) to the S(+)-enantiomer occurs in dogs (79).

It is usual in humans for the S(+)-enantiomer of 2-arylpropionic acids to predominate in plasma and for the S(+)- to R(-)-enantiomeric ratio of plasma concentrations to increase with time after administration of the racemate, which is often attributed to metabolic inversion of the chiral center of the R(-)-enantiomers to their S(+)-antipodes (80). In humans, the S(+)-enantiomer is generally eliminated more slowly than is the R(-)-enantiomer. The extent of chiral inversion of fenoprofen, which has been attributed to the differential rate of formation of the CoA-thioester by hepatic microsomes (81, 82), varies widely among species. It has been estimated to be 90% in dogs (83), 80% in sheep (81), 73% in rabbits (84), 60% in humans (85), 42% in rats (86), and 38% in horses (83).

Carprofen, a weak inhibitor of cyclo-oxygenase but which produces a significant antioedematous effect in dogs (87) and horses (88) and has potent antiplatelet-aggregating properties, does not appear to undergo chiral inversion in either direction in horses, calves, dogs, cats, and humans. After intravenous or oral administration of racemic carprofen, which contains a 50:50 mixture of the enantiomers, to horses (i.v.), 8- to 10-week-old calves (i.v.), cats (i.v.), and dogs (p.o.), the R(-)-enantiomer predominated in the plasma and the R(-)- to S(+)-enantiomeric ratio of plasma concentrations increased with time after administration of the racemate. The increasing R(-):S(+) enantiomeric ratio of plasma concentrations with time can be attributed to stereoselective hepatic metabolism, although stereoselective binding to plasma albumin could contribute. A contrasting situation to that which occurs in domestic animals was found in rats and humans, in that the S(+)-enantiomer predominated in plasma and the R(-):S(+) enantiomeric ratio of plasma concentrations decreased (although only slightly in humans) with time after administration of the racemate (89, 90). The R(-)- to S(+) ratio of area under

Table 14 Pharmacokinetic parameters describing disposition of S(+)- and R(-)-enantiomers after intravenous administration of racemic ketoprofen (KTP) to horses ($n = 6$) and sheep ($n = 6$)

Pharmacokinetic parameter	Horses		Sheep	
	S(+)-KTP	R(-)-KTP	S(+)-KTP	R(-)-KTP
$t_{1/2(\alpha)}$ (h)	0.13 ± 0.03	0.10 ± 0.02	0.14 ± 0.01	0.13 ± 0.03
$t_{1/2(\beta)}$ (h)	1.51 ± 0.45	1.09 ± 0.19	0.86 ± 0.08	0.87 ± 0.10
V_d (ml/kg)	491 ± 206	472 ± 146	256 ± 21	168 ± 15^a
CL_B (ml/h \times kg)	202 ± 22	277 ± 35^a	351 ± 50	196 ± 32^a
MRT (h)	2.23 ± 0.15	2.63 ± 0.33	0.79 ± 0.11	0.95 ± 0.13
AUC ($\mu\text{g} \times \text{h/ml}$)	5.67 ± 0.47	4.19 ± 0.37^a	4.74 ± 0.71	8.73 ± 1.22^a

^a $P < 0.05$.

Values are expressed as mean \pm S.E.M.
(From Refs. 71 and 72)

the plasma concentration–time curve (AUC) was 4.5:1 in horses (91), 2.0:1 in cats (92), 1.8:1 in dogs (87), and 1.4:1 in calves (93), after administration of the racemate. Area under the curve enantiomeric ratios in rats and humans were not determined.

Ketamine, which is present in the commercially available preparation as a 50:50 mixture of the S(+)- and R(–)-enantiomers, is metabolized by hepatic microsomal *N*-demethylation to the corresponding norketamine (metabolite I) enantiomers. Based on the reported eudismic ratio of S(+):R(–) for ketamine enantiomers of 2.9:1 and the observed duration of unconsciousness in dogs and the plasma concentrations in humans at time of emergence from anesthesia which are 0.5 [S(+)] and 1.7 [R(–)] $\mu\text{g/ml}$, it can be concluded that the S(+)-enantiomer is three times more active than is the R(–)-enantiomer (94, 95). After intravenous injection of racemic ketamine, the S(+)-enantiomer of norketamine predominated in the plasma of horses (96) and dogs (97). This could be attributed to enantioselective *N*-demethylation. Systemic clearances of racemic ketamine are 15, 28, and 29 ml/min \times kg in humans, horses, and dogs, respectively. Because the disposition of the individual enantiomers administered separately has not been studied, comment cannot be made regarding the extent of chiral inversion.

Whether a racemate or an enantiomer of a chiral drug should be used in formulating dosage forms depends on the relative pharmacodynamic activity and the potential toxicity (or side effects) of the individual enantiomers, their pharmacokinetic profiles, and, importantly, the proportions formed over time in the target animal species. Binding to plasma and tissue proteins, hepatic microsomal oxidative reactions, and probably glucuronide conjugation and carrier-mediated excretion processes are stereoselective and vary among animal species. First-pass metabolism may influence oral bioavailability of the enantiomers of chiral drugs that are highly extracted by the liver and administered as racemic mixtures. When both enantiomers of a drug with a single chiral center show distinct and desirable effects (e.g., most opioids, dobutamine, bupivacaine), even though they differ in pharmacodynamic activity or when their action and the effects produced are not stereoselective, the formulating of racemic mixtures may be entirely justifiable (98). Nonetheless, because of species variation, pharmacokinetic profiles of the individual enantiomers should be determined using stereospecific analytical methods (99–102) to calculate optimum dosage for the various animal species. Use of the more active enantiomer (eutomer) in formulating dosage forms should be considered when the enantiomers differ widely in pharmacodynamic activity [e.g., S(–)-propranolol, the S(+)-enantiomer of the 2-arylpropionic acid nonsteroidal

anti-inflammatory drugs, *d*-propoxyphene] or toxic potential (levamisole is the *l*-isomer of racemic tetramisole, which can produce many side effects). To selectively produce a certain effect, a distomer (less potent enantiomer) could be formulated as a particular dosage form, e.g., R(+)-timolol as eyedrops to reduce intraocular pressure (glaucoma), R(+)-verapamil as a parenteral or oral dosage form for the treatment of angina (103). The use of an enantiomer, a single chemical entity, would increase selectivity of action, reduce total exposure to the racemate, and simplify dose–response relationships. When an enantiomer is used in formulating dosage forms, it must be optically pure. Bioequivalence assessment of a generic drug preparation requires that the generic preparation contain the racemic drug or the enantiomer corresponding to whichever is present in the innovator (reference) dosage form.

PERCUTANEOUS ABSORPTION

The skin accounts for approximately 10% of live body weight in cattle, goats, and dogs; 7.5% in horses; and 3.7% in humans. Although skin receives approximately 6% of cardiac output, cutaneous blood flow rate to various regions differs among species (104). In humans and pigs, the cutaneous circulation supplies blood (in musculocutaneous arteries) to both the skin and the underlying musculature, whereas in dogs and cats (loose-skin species), blood is supplied directly to the skin.

ABSORPTION PROCESS

The absorption process for a topically applied drug essentially involves the following stages: dissolution of the drug in and release from the vehicle, drug penetration (by diffusion) through the *stratum*, and permeation through the “living” layers of the epidermis to the underlying dermis where absorption into the systemic circulation takes place. Although the initial stage is formulation-dependent in that it relates to the form of the drug and the nature of the vehicle, the translocation stages are primarily governed by the molecular structure and physicochemical properties of the drug. Penetration of the stratum corneum is generally the rate-limiting step in the absorption process (105). Only lipid-soluble drugs can diffuse through the dead, compacted, keratinized cells (corneocytes) of the stratum corneum. However, passive diffusion through the epidermis, including the stratum corneum, can take place by one, or more, of the following routes:

transcellular through the corneocytes, intercellular through the lipid matrix (a tortuous path), or along the sweat gland ducts and hair follicles (appendageal path). Even though lipophilic substances may penetrate the stratum corneum by transcellular diffusion, some degree of water solubility is required for passage through the "living" layers of the epidermis. Polar drugs have a low capacity to penetrate the stratum corneum but may gain access to the "living" epidermal layers by shunt diffusion along the appendageal path. Additional factors that influence percutaneous absorption include the nature of the vehicle, the state of hydration of the stratum corneum, drug persistence in the stratum corneum or other strata of the epidermis, biotransformation in the epidermis, and species differences in histological structure of skin. In aquatic mammals, the stratum corneum is very thick, and the corneocytes are solidly apposed, whereas the epidermis is devoid of a stratum granulosum regardless of whether the skin is glabrous (as in whales) or hairy (as in seals) (106).

The average thickness of the stratum corneum in laboratory and domestic animals and humans is in the range of 10 to 35 μm . The thickness of this epidermal layer might not influence the penetration of chemical substances, whereas the density of appendages per unit surface area, which differs among animal species, does influence passage through the epidermis. Human skin contains an average of 40–70 hair follicles and 200–250 sweat glands per square centimeter, whereas cattle skin contains approximately 2000 hair follicles, with associated sweat and sebaceous glands, per square centimeter (107). The mean follicle density in 10 British breeds of sheep varies from 1000 to 2000 per square centimeter of skin, with a secondary-to-primary follicle ratio of 2.4:1–5.9:1 (108). The ratio of secondary to primary follicles appears to be higher in Merino sheep skin in wool-growing regions. The sebaceous glands of cattle and sheep exude large quantities of lipid material (lanolin in the case of sheep) that serve to protect their skin. The emulsifying properties of exocrine secretions may enhance dissolution and thus facilitate percutaneous absorption of topically applied compounds in cattle and sheep. Seasonal changes in the composition of secretions may cause variations in absorption of moderately lipid-soluble drugs (e.g., levamisole applied by "pour on") at different times of the year (109). Because of the larger number of skin appendages per unit of surface area, the appendageal path will likely contribute more to percutaneous absorption of hydrophilic substances in cattle and sheep than in other species. Horses and humans have highly effective sweat glands, whereas cattle, sheep, pigs, dogs, and cats are unable to sweat profusely. Changes in ambient temperature appear to affect animal skin temperature to a higher degree than human skin

temperature, which suggests that skin plays a greater role in thermoregulation in humans than in animals.

The state of hydration of the stratum corneum, which is normally maintained at 10–15%, affects the rate of penetration of chemical substances. By increasing the state of hydration to 50%, the rate of permeation of some chemical substances through the epidermis can be increased up to 10-fold (110). Occlusion has been shown to enhance the pharmacological effect of topically applied hydrocortisone and fluocinolone acetonide (111); however, percutaneous absorption of drugs is not necessarily increased. The degree of occlusion-induced absorption enhancement appears to increase with increasing lipophilicity of drug substances.

Formulations containing an absorption-promoting substance, such as propylene glycol or sodium lauryl sulfate, may increase the permeability of the stratum corneum to water-soluble drugs. Propylene glycol is a commonly used vehicle in topical corticosteroid preparations for veterinary use. Various aprotic solvents, which include dimethylacetamide, dimethylformamide, dimethylsulfoxide, tetrahydrofurfuryl alcohol, and 2-pyrrolidone, serve as penetration enhancers of polar drugs (112). Dimethylsulfoxide (DMSO) is used in formulating some topical veterinary preparations. The penetration-enhancing property of DMSO is markedly concentration-dependent. At concentrations below 50% DMSO is water, the penetration rate of many drugs differs little than from aqueous solutions. The penetration rate of levamisole through skin of cattle and sheep was somewhat slower from a formulation containing DMSO (concentration not specified) than from an aqueous solution of the drug (107). Plasma and gastrointestinal fluid concentrations of levamisole were lower after pour-on application to cattle than after oral administration or subcutaneous injection of the drug (109). Parathion penetrated the skin of pigs more rapidly when formulated in DMSO than in other vehicles (glycerol–formal/isopropanol mixture, octanol, macrogol 400) (Table 15) (113). The formation of a stable 2:1 water hydrate at a concentration of 67% v/v DMSO may explain its dehydrating and penetration-enhancing effects when present at high concentration (114). These effects are accompanied by, or perhaps attributable to, epidermal tissue damage. Mineral oil may be used in formulating long-acting, water-based topical preparations of synthetic pyrethroids (permethrin, cypermethrin) for application to ruminant animals. This type of preparation would not be washed off by rain and could provide protection against flies for an extended period. Water-insoluble substances can be formulated as emulsifiable concentrates. The emulsifiable concentrate contains one or more surfactants and produces an emulsion or micellar solution with the

Table 15 Pharmacokinetic parameters for parathion (50 mg/kg) applied topically in various vehicles to pigs

Pharmacokinetic parameter	Vehicles			
	GFI ^a	DMSO	Octanol	Macroglol 400
AUC (μg × h/L)	1460–1795	1630–3050	2010–3310	595–600
MRT (h)	57–106	9.7–14.5	22–31	54–60
MAT (h)	55–104	7.5–12.5	20–29	52–58
Bioavailability (%)	16–20	19–28	15–29	3.9–54

^aGFI = glycerol–formal/isopropanol mixture.
Bioavailability (absolute) was based on AUC_{topical}/AUC_{IV}, with correction for dose.
(From Ref. 113.)

water-insoluble drug in the nonaqueous phase when mixed with water (107).

SPECIES VARIATIONS

The barrier properties of skin vary with the species of animal and within a species may differ between regions of the body. Based on limited data obtained from in vitro studies of skin permeability, it could be speculated that species can, in general, be ranked in the following order: rabbits > rats > guinea pigs > cats > dogs > pigs and Rhesus monkeys > humans (least permeable skin). Because of a lack of data, horses, cattle, sheep, and goats are not included in the comparison of skin permeability. The emulsifying property and occlusive effect of sebum and the high density of appendages per unit of surface area would be expected to facilitate percutaneous absorption of substances in ruminant species.

The maximal rate of penetration of an organophosphorus compound through skin sections excised from the dorsal thorax of various species generally supports the skin permeability ranking of species vide supra (Table 16) (115). The compound rapidly penetrated the skin of rabbits and rats, whereas penetration through pig skin occurred more slowly than through skin of the other species. Even though pig skin and human skin are similar in many respects (116), percutaneous absorption of a variety of compounds in the pig was found to range from zero to four times that in humans in vivo (117). Advantage can be taken of the often similar permeability characteristics of pig and human skins with avoidance of the systemic/fat distribution difference in vivo by using the isolated perfused porcine skin flap in vitro model (118, 119). The diffusion of chemical substances through skin and

metabolism within the skin can be determined by assay of the perfusate. The sum of the amount of compound that diffused into the perfusate and the residual amount in the skin preparation at the end of the exposure period to the drug provide an estimation of percutaneous absorption.

Regional variations in percutaneous absorption contribute to differences in the systemic availability of a drug depending on the site of topical application. The Rhesus monkey (*Macaca mulatta*) could probably serve as an animal model for human skin regional variation (120).

The absolute bioavailability of a topically applied drug can be determined only by measurement of plasma concentrations and comparing total areas under the curves or, less reliably, the amounts excreted in urine over a period of at least six half-lives after topical application and intravenous injection of the drug. An appropriate washout period must be allowed to elapse between the phases of a crossover study, which is the experimental design that should be used whenever feasible. Because of species

Table 16 Maximal penetration of radiolabeled organophosphorus compound through excised skin from dorsal thorax of various species

Species	Rate (μg/cm ² /min)
Pig	0.3
Dog	2.7
Monkey	4.2
Goat	4.4
Cat	4.4
Guinea pig	6.0
Rabbit	9.3
Rat	9.3

(From Ref. 115.)

variations in ultrastructure of skin, cutaneous blood supply, density of appendages per unit of surface area, and activity of biotransformation pathways, the percutaneous absorption (rate and extent) of a drug is best determined by performing the study in the species of interest.

CUTANEOUS BIOTRANSFORMATION

Epidermal cytochrome P450 and hydrolytic enzymes in the lipid matrix of the stratum corneum as well as in the stratum granulosum may be involved in conversion reactions of topically applied steroids (121). In pigs, epidermal cytochrome P450 has been shown to convert parathion to paraoxon by oxidative desulfuration (122). The oxygen analog formed is rapidly hydrolyzed to inactive metabolites. The significant difference between mammals and insects in the rate at which the hydrolytic reaction takes place accounts for the selective toxicity of thiophosphate insecticides. The extent to which the ultrastructural difference in the epidermis of aquatic and terrestrial mammals affects the activity of drug-metabolizing enzymes in skin is not known. Fish appear to be unable to detoxify thiophosphate insecticides.

Benzoyl peroxide, the active ingredient in some shampoos for dogs, is almost completely metabolized to benzoic acid in the epidermis. Benzoic acid undergoes conjugation with glycine, primarily in the liver, and is excreted in urine as hippuric acid. Methylation of norepinephrine to epinephrine, an *N*-transferase-mediated conjugation reaction, in human and animal skin preparations has been reported (123).

Biotransformation of propranolol, which involves microsomal-mediated oxidative reactions and glucuronide conjugation, is stereoselective in both the liver and the skin. Based on in vitro studies, using intact human skin and microsomal preparations, of percutaneous absorption and metabolism of racemic propranolol, it was concluded that the *S*(-)-enantiomer (eutomer) is metabolized more efficiently by skin than is the *R*(+)-enantiomer (124). The converse applies to hepatocytes, which metabolize the *R*(+)-enantiomer more efficiently (125). The cytochrome P450 mono-oxygenase enzymes in skin, as in the liver, can be induced (by dexamethasone, for example) or inhibited (chloramphenicol, imidazole antifungal agents). The clinical significance of altered microsomal enzyme activity in the epidermis has not been established. Metronidazole, which is available as a topical gel, inhibits acetaldehyde dehydrogenase (a nonmicrosomal enzyme); however, whether the enzyme is present in skin does not appear to be known.

TOPICAL PREPARATIONS

There is a wide variety of veterinary drug preparations available for topical application to the skin. Although most of these preparations are intended to produce local effects, some are formulated to distribute via the systemic circulation to skin covering all regions of the body or to produce systemic effects. Because of the prevalence of ectoparasites, several ectoparasiticide preparations are available for topical application by various methods depending on the animal species (Table 17). Liquid concentrates, which must be appropriately diluted before use, and prepared solutions are convenient to apply and, depending on susceptibility of the ectoparasite and persistence of the drug in skin, may provide effective treatment and protection against reinfestation for an extended period. The usual method of application of prepared solutions is by pour on to cattle, sheep, pigs, and horses, whereas "spot on" is preferred for dogs and cats. The active ingredient in preparations for pour-on application must be sufficiently lipid-soluble for percutaneous absorption to occur, and the preparation should provide residual activity in the stratum corneum and stratum germinativum and not be removed from the skin by environmental conditions (such as rain) or by rubbing. Diluted liquid concentrates are applied by spray to cattle, horses, pigs, and poultry and by either dip or spray to sheep. Because of the large quantity of drug required for dipping sheep and the difficulty of safe disposal of dip, application of diluted liquid (dip) concentrate as a spray, but at a higher

Table 17 Dosage forms and methods of application of topical ectoparasiticide preparations to individual species

Animal species	Dosage form	Method of application
Cattle	Solution	Pour on
	Liquid concentrate ^a	Spray
	Ear tag	Attach to ears
Sheep	Liquid concentrate ^a	Dip, spray
	Solution	Spot on, pour on
Pigs	Solution	Pour on, spot on
	Liquid concentrate ^a	Spray
Horses	Solution	Pour on
	Liquid concentrate ^a	Spray, shampoo
	Lotion	Dab on
Dogs and cats	Solution	Spot on, spray (dogs)
	Collar	Surrounding neck
	Dusting power	Apply to coat
	Liquid concentrate ^a	Sponge on (dogs)
	Shampoo	Wash (dogs)

^aLiquid concentrates must be appropriately diluted before use on animals.

Table 18 Dosage forms of preparations containing synthetic pyrethroids and the methods of topical application to various species

Dosage form	Animal species	Method of application
Solution	Cattle, sheep, horses, pigs, dogs, cats	Pour on Spot on, spray (dogs)
Liquid concentrate	Cattle, sheep, horses, poultry	spray, dip (sheep)
Dusting powder	Dogs, cats	Apply to coat
Shampoo	Dogs	Wash
Collar	Dogs, cats	Surrounding neck
Ear tag	Cattle	Attach to ears

Synthetic pyrethroids: cypermethrin, deltamethrin, fenvalerate, flumethrin, permethrin.

concentration than that used in the dip bath, is becoming a popular method of ectoparasiticide application by sheep farmers. The relative effectiveness of spraying versus dipping sheep for the treatment and prevention of ectoparasite infestation is open to question. When choosing between an organophosphorus compound (dimpylate, propetamphos) and a synthetic pyrethroid (flumethrin, cypermethrin) for dipping sheep, a consideration (in addition to residual protection) that could be of practical relevance is that there is no withdrawal period associated with the use of synthetic pyrethroids. The low toxicity of pyrethroids in mammals is primarily attributed to the rapid biotransformation by ester hydrolysis and/or hydroxylation (phase I metabolic reactions). Unlike mammals, fish are extremely sensitive to pyrethroid toxicity.

Synthetic pyrethroids (cypermethrin, deltamethrin, fenvalerate, flumethrin, permethrin) are available in a variety of dosage forms for topical application to domestic animals (Table 18). The appropriate dosage form of a

pyrethroid is primarily determined by the animal species, whereas due consideration is given to the ectoparasites and insects that affect the species. Aerosol sprays are more suitable for application to dogs than to cats, because cats resent (fearful response) being sprayed and lick the applied preparation from their coat. The combination preparation containing fenvalerate (0.09%), a synthetic pyrethroid, and diethyltoluamide (9.5%), a cutaneous penetration-enhancing substance, has been reported to cause acute toxicity in cats and occasionally in dogs (126, 127).

Macrolide endectocides (ivermectins and milbemycins) have activity at low concentrations against both internal (nematode) and external (arthropod) parasites. Veterinary preparations of ivermectins include parenteral solutions (for subcutaneous injection), various oral dosage forms, and topical solutions (Table 19). The different preparations are specifically designed for use in certain animal species. The topical solution of ivermectin or doramectin is applied to cattle by pouron, whereas the

Table 19 Dosage forms and routes of administration of avermectins to various animal species

Dosage form	Avermectin	Animal species	Route of administration
Parenteral			
Solution	Ivermectin, doramectin	Cattle, sheep	S.C. injection
	Ivermectin	pigs, (cats)	S.C. injection
Oral			
Solution	Ivermectin	Sheep, goats	Drench
Controlled-release ruminal bolus	Ivermectin	Cattle	P.O.
Controlled-release ruminal capsule	Ivermectin	Sheep	P.O.
Premix	Ivermectin	Pigs	Addition to feed
Paste	Ivermectin	Horses	P.O.
Tablet	Ivermectin	Dogs	P.O.
Topical			
Solution	Ivermectin, doramectin	Cattle	Pour on
Solution	Selamectin	Dogs, cats	Spot on

topical solution of selamectin is applied to dogs and cats by spot-on. Preparations for spot-on application contain concentrated solutions of ectoparasiticides and should be applied directly to the skin at one (cats) or two (dogs) locations where the animal is unable to ingest the drug by licking. The back of the neck is one such site. The total dose of drug to apply will vary with the animal species (lower dose for cats) and with the range of body weight (dogs). Selamectin solution, for example, is commercially available in six dose sizes (from 15 to 240 mg).

The classes of ectoparasiticide and the dosage forms used in cattle, an arbitrarily selected species, are shown in Table 20. Even though different classes of drug may have activity against a similar range of ectoparasites, the drugs that belong to the various classes differ in clinical effectiveness. Quantitative differences in clinical effectiveness could be attributed to parasite susceptibility, which would be decreased by the development of resistance, and to the combined effect of drug access to the site where the parasite is located and persistence of the drug in skin. Both access and persistence are influenced by the physicochemical properties of the drug, the formulation of the preparation, and the method of application to the animal.

Dermatological preparations containing antibacterial agents include aerosol sprays (oxytetracycline hydrochloride, neomycin sulfate in propylene glycol), gels

(metronidazole, fusidic acid), creams (gentamicin sulfate, neomycin sulfate and zinc bacitracin, silver sulfadiazine), ointments (chlortetracycline hydrochloride, sodium fusidate, framycetin sulfate and gramicidin, neomycin sulfate, zinc bacitracin, polymyxin B sulfate), and dusting powders (chlortetracycline hydrochloride and benzocaine). The type of preparation influences the stability and release of the active ingredients. Gels and creams are generally easier to apply to animals but are less occlusive than ointments. Ointments provide a longer duration of drug action, and the occlusive effect may enhance preparation of the active ingredient(s) to the site of infection.

The selection of an antibacterial agent should be based on the clinical diagnosis and, whenever feasible, in vitro culture and susceptibility testing (when considered necessary) of the micro-organisms(s) isolated. In the treatment of superficial skin infections, a topically applied antibacterial preparation may suffice. However in severe and deep-seated skin infections, both local and systemic therapy should be applied. The avoidance of antagonism requires that due consideration be given to the mechanisms of action of drugs selected for concurrent use.

Cutaneous mycotic infections caused by filamentous or dermatophytic fungi can be treated either topically or systemically, depending on the location and severity of the skin lesions. Topical antifungal preparations should be

Table 20 Representative ectoparasiticides for application to cattle

Ectoparasiticide class (representative drugs)	Spectrum of activity	Dosage forms (methods of application)
<i>Organophosphate:</i> (Phosmet)	Mites, lice Warble-fly larvae	Solution (pour on)
<i>Pyrethroid:</i> (Cypermethrin) (Deltamethrin) (Fenvalerate) (Permethrin)	Lice, flies	Solution (pour on) Solution (spot on) Liquid concentrate (spray) ^a Ear tag
<i>Amidine:</i> (Amitraz)	Lice, mites, ticks	Liquid concentrate (spray)
<i>Avermectin:</i> (Ivermectin) (Doramectin) (Abamectin) (Eprinomectin)	Mites, lice, Horn fly, Warble-fly larvae	Solution (pour on) Parenteral solution (S.C. injection); Controlled-release ruminal bolus (ivermectin)
<i>Milbemycin:</i> (Moxidectin)	Similar to avermectins	Solution (pour on); parenteral solution (S.C. injection)

^aLiquid concentrates must be appropriately diluted before use.

formulated to promote penetration and persistence of the drug at the site of infection. Drugs that may be applied topically include various imidazole derivatives (clotrimazole, enilconazole, ketoconazole, miconazole), natamycin, nystatin, and tolnaftate, whereas orally administered drugs include ketoconazole, fluconazole, itraconazole, nystatin, and griseofulvin. Superficial infections caused by *Candida* spp. may be treated locally by applying an imidazole derivative or nystatin.

In domestic animals and humans, the common fungal organisms that cause skin infections are *Trichophyton mentagrophytes* (the usual cause of ringworm in calves between 2 and 7 months of age), *Microsporum gypseum*, and *T. verrucosum* (in species other than dogs and cats). Superficial mycotic skin infections are also caused by *T. equinum* (horses, cattle, humans), *M. nanum* (pigs, cattle, humans), and *M. canis* (dogs and cats). Filamentous fungal keratitis in horses is usually caused by *Fusarium* spp. Yeasts, primarily *Candida* spp., can cause mastitis in cows and metritis in mares and occasionally inhabit the skin, primarily at mucocutaneous junctions, in dogs.

For topical application, individual imidazoles and nystatin are formulated as creams, whereas tolnaftate and nystatin are available as ointments. Enilconazole is formulated as a liquid concentrate, that must be diluted before topical application by wash to horses and dogs or by spray to cattle. Topically applied enilconazole may be used in conjunction with systemic (oral) treatment with griseofulvin. Ketoconazole alone and miconazole nitrate combined with chlorhexidine gluconate are commercially available as shampoos for dogs. After topical application, imidazoles attain effective concentrations and persist in the outer layers of the epidermis, and percutaneous absorption appears to be minimal. Natamycin is commercially available as a suspension, to be diluted before application by spray to horses or cattle. In addition to spraying horses affected with ringworm, all grooming utensils and tackle must be thoroughly cleansed and immersed in the natamycin suspension, which should be diluted in plastic or galvanized containers. Natamycin (5% aqueous suspension) is effective in the treatment of filamentous fungal keratitis, particularly when caused by *Fusarium* spp. in horses.

The selectivity of action of antifungal azoles is attributed to their greater affinity for fungal than for mammalian cytochrome P450 enzymes. Because imidazole derivatives (clotrimazole, enilconazole, ketoconazole, miconazole) are less selective in their action than are triazoles (itraconazole, fluconazole), the former subgroup would be expected to cause greater interference with the biosynthesis of endogenous steroid hormones and to show a higher incidence of pharmacokinetic interactions with

other drugs that undergo microsomal-mediated biotransformation.

The therapeutic effectiveness of topically applied corticosteroids is attributed primarily to their anti-inflammatory activity. The relative efficacy of topical corticosteroids appears to be in the following order: hydrocortisone, prednisolone, betamethasone < hydrocortisone valerate or butyrate, betamethasone valerate, triamcinolone acetonide, flucinolone acetonide < betamethasone dipropionate, fluocinonide. In addition to the nature of the corticosteroid, its solubility, and, to a lesser extent, the concentration used, clinical efficacy is influenced by the formulation of the preparation. Glucocorticoids appear to have greater efficacy when formulated in ointment bases than in cream or lotion vehicles. This could be attributed to the occlusive effect provided by ointments. The application of an occlusive dressing further enhances penetration and persistence of the steroid (reservoir effect) in the stratum corneum (112, 128).

TRANSDERMAL THERAPEUTIC SYSTEMS

A transdermal therapeutic system is a rate-controlled drug-delivery system that, applied to the surface of the skin, continuously releases the drug at a rate that will provide a desired steady-state plasma concentration for a specified duration.

The transdermal therapeutic system containing fentanyl [μ (primarily)-opioid agonist], which was designed to release the drug at a constant rate for 72 h, may have application in dogs for the control of postoperative (surgical) pain. Secure placement of the transdermal system that releases 50 μg of fentanyl/h on the dorsal aspect of the thorax of Beagle dogs (11.4–16.5 kg body weight) provided an average steady-state plasma fentanyl concentration of 1.6 ng/ml, which is within the range of plasma concentrations (1–2 ng/ml) considered to provide analgesia without producing other significant effects (129, 130). A steady-state concentration of fentanyl in plasma was reached at approximately 24 h after placement of the transdermal system, as would be expected because the half-life of fentanyl in Beagles is 6 h and was maintained until removal of the system at 72 h. Because the dosing rate ($3.7 \mu\text{g/h} \times \text{kg}$) exceeded the systemic clearance ($2.7 \mu\text{g/h} \times \text{kg}$) and the transdermal bioavailability of fentanyl is 64%, a fraction of the released drug either persists in the stratum corneum or is metabolized in the epidermis before absorption into the systemic circulation, or both may contribute to incomplete systemic availability of the drug. Because of the 24 h delay in achieving plasma

concentrations within the analgesia-producing range, either an intravenous dose (approximately 30 µg/kg) could be administered at the time of placement of the transdermal system or the system could be securely placed on the animal 12 h before performing the surgery.

Whenever a drug-delivery device such as an insecticidal collar or a transdermal system is securely placed on an animal, the increased potential for drug interaction must be kept in mind at the time of selecting another drug for administration by any route and throughout the course of therapy.

REFERENCES

1. Foster, T.P. Protein/Peptide Veterinary Formulations. *Development and Formulation of Veterinary Dosage Forms*; Hardee, G.E., Baggot, J.D., Eds.; Marcel Dekker, Inc.: New York, 1998; 231–282.
2. Hansel, W.; McEntee, K. Female Reproductive Processes. *Duke's Physiology of Domestic Animals*; Swenson, M.J., Ed.; Comstock Publishing Associates (Cornell University Press): Ithaca, NY, 1977; 772–800.
3. Hogben, C.A.M.; Tocco, D.J.; Brodie, B.B.; Schanker, L.S. On the Mechanism of Intestinal Absorption of Drugs. *J. Pharmacol. Exp. Ther.* **1959**, *125*, 275–282.
4. Theeuwes, F.; Bayne, W. Dosage Form Index: An Objective Criterion for Evaluation of Controlled-Release Drug Delivery Systems. *J. Pharm. Sci.* **1977**, *66*, 1388–1392.
5. Koritz, G.D.; McKiernan, B.C.; Neff-Davis, C.A.; Munsiff, I.J. Bioavailability of Four Slow-Release Theophylline Formulations in the Beagle Dog. *J. Vet. Pharmacol. Ther.* **1986**, *9*, 293–302.
6. Dohoo, S. Steady-State Pharmacokinetics of Oral Sustained-Release Morphine Sulphate in Dogs. *J. Vet. Pharmacol. Ther.* **1997**, *20*, 129–133.
7. Maitho, T.E.; Lees, P.; Taylor, J.B. Absorption and Pharmacokinetics of Phenylbutazone in Welsh Mountain Ponies. *J. Vet. Pharmacol. Ther.* **1986**, *9*, 26–39.
8. van Duijken, E.; Vulto, A.G.; Sloet van Oldruitenborgh-Oosterbann, M.M.; Kessels, B.G.F.; van Miert, A.S.J.P.A.M.; Breukink, H.J. Pharmacokinetics of Trimethoprim-Sulphaclorpyridazine in Horses after Oral, Nasogastric and Intravenous Administration. *J. Vet. Pharmacol. Ther.* **1995**, *18*, 47–53.
9. Hungate, R.E. *The Rumen and its Microbes*; Academic Press: New York, 1966; 218.
10. Phillipson, A.T.; McAnally, R.A. Studies on the Fate of Carbohydrates in the Rumen of the Sheep. *J. Exp. Biol.* **1942**, *19*, 199–214.
11. Masson, M.J.; Phillipson, A.T. The Absorption of Acetate, Propionate and Butyrate from the Rumen of Sheep. *J. Physiol. (Lond)* **1951**, *113*, 189–206.
12. Dobson, A. Physiological Peculiarities of the Ruminant Relevant to Drug Distribution. *Federation Proceedings* **1967**, *26*, 994–1000.
13. Clark, R.; Wessels, J.J. The Influences of the Nature of the Diet and of Starvation on the Concentration Curve of Sulphanilamide in the Blood of Sheep After Oral Dosing. *Onderstepoort J. Vet. Res.* **1952**, *25*, 75–83.
14. Masson, M.J.; Phillipson, A.T. The Composition of the Digesta Leaving the Abomasum of Sheep. *J. Physiol. (Lond)* **1952**, *116*, 98–111.
15. Gingerich, D.A.; Baggot, J.D.; Yearly, R.A. Pharmacokinetics and Dosage of Aspirin in Cattle. *J. Am. Vet. Med. Assoc.* **1975**, *167*, 945–948.
16. Ali, D.N.; Hennessy, D.R. The Effect of Temporarily Reduced Feed Intake on the Efficacy of Oxfendazole in Sheep. *Int. J. Parasitol.* **1995**, *25*, 71–74.
17. Taylor, S.M.; Malton, T.R.; Blanchflower, J.; Kennedy, D.G.; Hewitt, S.A. Effects of Dietary Variations on Plasma Concentrations of Oral Flukicides in Sheep. *J. Vet. Pharmacol. Ther.* **1993**, *16*, 48–54.
18. Rowlands, D.ap T.; Shepherd, M.T.; Collins, K.R. The Oxfendazole Pulse Release Bolus. *J. Vet. Pharmacol. Ther.* **1988**, *11*, 405–408.
19. Singh, B.N.; Kim, K.H. Floating Drug Delivery Systems: An Approach to Oral Controlled Drug Delivery Via Gastric Retention. *J. Controlled. Release* **2000**, *63*, 235–259.
20. Mallis, G.I.; Schmidt, D.H.; Lindenbaum, J. Superior Bioavailability of Digoxin Solution in Capsules. *Clin. Pharmacol. Ther.* **1975**, *18*, 761–768.
21. Johnson, B.F.; Bye, C.; Jones, G.; Sabey, G.A. A Completely Absorbed Oral Preparation of Digoxin. *Clin. Pharmacol. Ther.* **1976**, *19*, 746–751.
22. Lindenbaum, J. Greater Bioavailability of Digoxin Solution in Capsules: Studies in the Postprandial State. *Clin. Pharmacol. Ther.* **1977**, *21*, 278–282.
23. Angelucci, L.; Petrangeli, B.; Celletti, P.; Favilli, S. Bioavailability of Flufenamic Acid in Hard and Soft Gelatin Capsules. *J. Pharm. Sci.* **1976**, *65*, 455–456.
24. Landoni, M.F.; Lees, P. Influence of Formulation on the Pharmacokinetics and Bioavailability of Racemic Ketoprofen in Horses. *J. Vet. Pharmacol. Ther.* **1995c**, *18*, 446–450.
25. Baggott, D.G.; Ross, D.B.; Preston, J.M.; Gross, S.J. Nematode Burdens and Productivity of Grazing Cattle Treated with a Prototype Sustained-Release Bolus Containing Ivermectin. *Vet. Rec.* **1994**, *135*, 503–506.
26. Berghen, P.; Hilderson, H.; Vercruysse, J.; Claerebout, E.; Dorny, P. Field Evaluation of the Efficacy of the Fenbendazole Slow-Release Bolus in the Control of Gastrointestinal Nematodes of First-Season Grazing Cattle. *Vet. Q.* **1994**, *16*, 161–164.
27. Laby, R.H. US Patent 3, 844, 284, 1974
28. Klink, P.R.; Ferguson, T.H.; Magruder, J.A. *Formulation of Veterinary Dosage Forms. Development and Formulation of Veterinary Dosage Forms*; Hardee, G.E., Baggot, J.D., Eds.; Marcel Dekker, Inc.: New York, 1998, 145–229.
29. Firth, E.C.; Nouws, J.F.M.; Driessens, F.; Schmaetz, P.; Peperkamp, K.; Klein, W.R. Effect of the Injection Site on the Pharmacokinetics of Procaine Penicillin G in Horses. *Am. J. Vet. Res.* **1986**, *47*, 2380–2384.
30. Rasmussen, F. Tissue Damage at the Injection Site after Intramuscular Injection of Drugs in Food-Producing Animals. *Trends in Veterinary Pharmacology and Toxicology*; van Miert, A.S.J.P.A.M., Frens, J., van der Kreek, F.W., Eds.; Elsevier: Amsterdam, 1980; 27–33.

31. Nouws, J.F.M. Irritation, Bioavailability and Residue Aspects of Ten Oxytetracycline Formulations Administered Intramuscularly to Pigs. *Vet Q* **1984**, 6, 80–84.
32. Nouws, J.F.M. Injection Sites and Withdrawal Times. *Annal. de. Rech. Vet.* **1990**, 21 (Suppl 1), 145S–150S.
33. Banting, A.deL; Tranquart, F. Echography as a Tool in Clinical Pharmacology. *Acta. Vet. Scand. Suppl.* **1991**, 87, 215–216.
34. Aktas, M.; Lefebvre, H.P.; Toutain, P.-L.; Braun, J.P. Disposition of Creatine Kinase Activity in Dog Plasma Following Intravenous and Intramuscular Injection of Skeletal Muscle Homogenates. *J. Vet. Pharmacol. Ther.* **1995**, 18, 1–6.
35. Toutain, P.-L.; Lassourd, V.; Costes, G.; Alvinerie, M.; Bret, L.; Lefebvre, H.P.; Braun, J.P. A Non-Invasive and Quantitative Method for the Study of Tissue Injury Caused by Intramuscular Injection of Drugs in Horses. *J. Vet. Pharmacol. Ther.* **1995**, 18, 226–235.
36. Nouws, J.F.M.; Vree, T.B. Effect of Injection Site on the Bioavailability of an Oxytetracycline Formulation in Ruminant Calves. *Vet. Q.* **1983**, 5, 165–170.
37. Korsrud, G.O.; Boison, J.O.; Papich, M.G.; Yates, W.D.G.; MacNeil, J.D.; Janzen, E.D.; Cohen, R.D.H.; Landry, D.A.; Lambert, G.; Yong, M.S.; Messier, J.R. Depletion of Intramuscularly and Subcutaneously Injected Procaine Penicillin G from Tissues and Plasma of Yearling Beef Steers. *Can. J. Vet. Res.* **1993**, 57, 223–230.
38. Wheaton, J.E.; Al-Raheem, S.N.; Godfredson, J.A.; Dorn, J.M.; Wong, E.A.; Vale, W.W.; Rivier, J.; Mowles, T.F.; Heimer, E.P.; Felix, A.M. Use of Osmotic Pumps for Subcutaneous Infusion of Growth Hormone-Releasing Factors in Steers and Wethers. *J. Anim. Sci.* **1988**, 66, 2876–2885.
39. Golenz, M.R.; Wilson, W.D.; Carlson, G.P.; Craychee, T.J.; Mihalyi, J.E.; Vinox, L. Effect of Route of Administration and Age on the Pharmacokinetics of Amikacin Administered by the Intravenous and Intraosseous Routes to 3- and 5-Day-Old Foals. *Equine. Vet. J.* **1994**, 26, 367–373.
40. Baggot, J.D.; Brown, S.A. Basis for Selection of the Dosage Form. *Development and Formulation of Veterinary Dosage Forms*; Hardee, G.E., Baggot, J.D., Eds.; Marcel Dekker, Inc.: New York, 1998; 7–143.
41. Wilson, R.C.; Duran, S.H.; Horton, C.R., Jr.; Wright, L.C. Bioavailability of Gentamicin in Dogs after Intramuscular or Subcutaneous Injections. *Am. J. Vet. Res.* **1989**, 50, 1748–1750.
42. Baggot, J.D.; Ling, G.V.; Chatfield, R.C. Clinical Pharmacokinetics of Amikacin in Dogs. *Am. J. Vet. Res.* **1985**, 46, 1793–1796.
43. Craigmill, A.L.; Brown, S.A.; Wetzlich, S.E.; Gustafson, C.R.; Arndt, T.S. Pharmacokinetics of Ceftiofur and Metabolites After Single Intravenous and Intramuscular Administration and Multiple Intramuscular Administration of Ceftiofur Sodium to Sheep. *J. Vet. Pharmacol. Ther.* **1997**, 20, 139–144.
44. Rasmussen, F.; Svendsen, O. Tissue Damage and Concentration at the Injection Site After Intramuscular Injection of Chemotherapeutics and Vehicles in Pigs. *Res. Vet. Sci.* **1976**, 20, 55–60.
45. Lanusse, C.; Lifschitz, A.; Virkel, G.; Alvarez, L.; Sanchez, S.; Sutra, J.F.; Galtier, P.; Alvinerie, M. Comparative Plasma Disposition Kinetics of Ivermectin, Moxidectin and Doramectin in Cattle. *J. Vet. Pharmacol. Ther.* **1997**, 20, 91–99.
46. Banting, A.deL; Baggot, J.D. Comparison of the Pharmacokinetics and Local Tolerance of Three Injectable Oxytetracycline Formulations in Pigs. *J. Vet. Pharmacol. Ther.* **1996**, 19, 50–55.
47. Nouws, J.F.M.; van Ginneken, C.A.M.; Hekman, P.; Ziv, G. Comparative Plasma Ampicillin Levels and Bioavailability of Five Parenteral Ampicillin Formulations in Ruminant Calves. *Vet. Q.* **1982**, 4, 62–71.
48. Agerso, H.; Friis, C. Bioavailability of Amoxycillin in Pigs. *J. Vet. Pharmacol. Ther.* **1998**, 21, 41–46.
49. Marshall, A.B.; Palmer, G.H. Injection Sites and Drug Bioavailability. In *Trends in Veterinary Pharmacology and Toxicology*; van Miert, A.S.J.P.A.M., Frens, J., van der Kreek, F.W. Eds.; Elsevier: Amsterdam, 1980, 54–60.
50. Baggot, J.D. *Principles of Drug Disposition in Domestic Animals: The Basis of Veterinary Clinical Pharmacology*; WB Saunders: Philadelphia, 1977, 10–13, 73–112.
51. Rasmussen, F. *Studies on the Mammary Excretion and Absorption of Drugs*; Carl Fr. Mortensen: Copenhagen, 1966.
52. Prescott, J.F.; Baggot, J.D. Principles of Antimicrobial Drug Disposition; Antimicrobial Drug Use in Bovine Mastitis. *Antimicrobial Therapy in Veterinary Medicine*, 2nd Ed.; Iowa State University Press: Ames Iowa, 1993; 37–60, 553–561.
53. Ziv, G. Drug Selection and Use in Mastitis: Systemic Vs Local Drug Therapy. *J. Am. Vet. Med. Assoc.* **1980**, 176, 1109–1115.
54. Pyorala, S. Staphylococcal and Streptococcal Mastitis. *The Bovine Udder and Mastitis*; Sandholm, M., Honkanen-Buzalski, T., Kaartinen, L., Pyorala, S., Eds.; University of Helsinki, Faculty of Veterinary Medicine: Helsinki, 1995; 143–148.
55. Sandholm, M.; Pyorala, S. Coliform Mastitis. *The Bovine Udder and Mastitis*; Sandholm, M., Honkanen-Buzalski, T., Kaartinen, L., Pyorala, S., Eds.; University of Helsinki, Faculty of Veterinary Medicine: Helsinki, 1995; 149–160.
56. Baggot, J.D.; McKellar, Q.A. The Absorption, Distribution and Elimination of Anthelmintic Drugs: The Role of Pharmacokinetics. *J. Vet. Pharmacol. Ther.* **1994**, 17, 409–419.
57. Benet, L.Z. Pharmacokinetic Parameters: Which are Necessary to Define a Drug Substance? *Eur. J. Resp. Dis.* **1984**, 65 (Suppl. 134), 45–61.
58. Williams, R.T. *Detoxication Mechanisms*; Chapman & Hall: London, 1959, 1–22, 732–740.
59. Williams, R.T. Comparative Patterns of Drug Metabolism. *Fed. Proc.* **1967**, 26, 1029–1039.
60. Williams, R.T. Species Variations in Drug Biotransformations. *Fundamentals of Drug Metabolism and Drug Disposition*; LaDu, B.N., Mandel, H.G., Way, E.L., Eds.; Williams & Wilkins: Baltimore, 1971; 187–205.
61. Frey, H.-H.; Loscher, W. Pharmacokinetics of Anti-Epileptic Drugs in the Dog: A Review. *J. Vet. Pharmacol. Ther.* **1985**, 8, 219–233.

62. French, M.R.; Bababunmi, E.A.; Golding, R.R.; Bassir, O.; Caldwell, J.; Smith, R.L.; Williams, R.T. The Conjugation of Phenol, Benzoic Acid, I-Naphthylacetic Acid and Sulphadimethoxine in the Lion, Civet and Genet. *FEBS Letters* **1974**, *46*, 134–137.
63. Boxenbaum, H. Interspecies Scaling, Allometry, Physiological Time, and the Ground Plan of Pharmacokinetics. *J. Pharmacokinet. Biopharm.* **1982**, *10*, 201–227.
64. Dorrestein, G.M.; van Miert, A.S.J.P.A.M. Pharmacotherapeutic Aspects of Medication of Birds. *J. Vet. Pharmacol. Ther.* **1988**, *11*, 33–44.
65. Calder, W.A., III *Size, Function and Life History*; Harvard University Press: Cambridge, MA, 1984.
66. Grondel, J.L.; Nouws, J.F.M.; Schutte, A.R.; Driessens, F. Comparative Pharmacokinetics of Oxytetracycline in Rainbow Trout (*Salmo gairdneri*) and African Catfish (*Clarias Gariepinus*). *J. Vet. Pharmacol. Ther.* **1989**, *12*, 157–162.
67. Boxenbaum, H. Interspecies Variation in Liver Weight, Hepatic Blood Flow, and Antipyrine Intrinsic Clearance: Extrapolation of Data to Benzodiazepines and Phenytoin. *J. Pharmacokinet. Biopharm.* **1980**, *8*, 165–176.
68. Gaspari, F.; Bonati, M. Interspecies Metabolism and Pharmacokinetic Scaling of Theophylline Disposition. *Drug. Metab. Rev.* **1990**, *22*, 179–207.
69. Kleiber, M. Metabolic Turnover Rate: A Physiological Meaning of the Metabolic Rate Per Unit Body Weight. *J. Theoret. Biol.* **1975**, *53*, 199–204.
70. Landoni, M.F.; Soraci, A.L.; Delatour, P.; Lees, P. Enantioselective Behaviour of Drugs Used in Domestic Animals: A Review. *J. Vet. Pharmacol. Ther.* **1997**, *20*, 1–16.
71. Landoni, M.F.; Lees, P. Comparison of the Anti-Inflammatory Actions of Flunixin and Ketoprofen in Horses Applying PK/PD Modelling. *Equine. Vet. J.* **1995**, *27*, 247–256.
72. Landoni, M.F.; Comas, W.; Mucci, N.; Anglarilla, G.; Bidall, D.; Lees, P. Enantiospecific Pharmacokinetics and Pharmacodynamics of Ketoprofen in Sheep. *J. Vet. Pharmacol. Ther.* **1999**, *22*, 349–359.
73. Landoni, M.F.; Cunningham, F.M.; Lees, P. Pharmacokinetics and Pharmacodynamics of Ketoprofen in Calves Applying PK/PD Modelling. *J. Vet. Pharmacol. Ther.* **1995**, *18*, 315–324.
74. Landoni, M.F.; Lees, P. Pharmacokinetics and Pharmacodynamics of Ketoprofen Enantiomers in the Horse. *J. Vet. Pharmacol. Ther.* **1996**, *19*, 466–474.
75. Landoni, M.F.; Lees, P. Pharmacokinetics and Pharmacodynamics of Ketoprofen Enantiomers in Calves. *Chirality* **1995**, *7*, 586–597.
76. Mauleon, D.; Mis, R.; Ginesta, J.; Ortega, E.; Vilageliu, J.; Basi, N.; Carganico, G. Pharmacokinetics of Ketoprofen Enantiomers in Monkeys Following Single and Multiple Oral Administration. *Chirality* **1994**, *6*, 537–542.
77. Rudy, A.C.; Liu, Y.; Brater, D.C.; Hall, S.D. Steiospecific Pharmacokinetics and Inversion of (R)-ketoprofen in Healthy Volunteers. *J. Clin. Pharmacol.* **1998**, *38*, 3S–10S.
78. Aberg, G.; Ciofalo, V.B.; Pendleton, R.G.; Ray, G.; Weddle, D. Inversion of (R)- to (S)-Ketoprofen in Eight Animal Species. *Chirality* **1995**, *7*, 383–387.
79. Garcia, M.L.; Tost, D.; Vitageliu, J.; Lopez, S.; Carganico, G.; Mauleon, D. Bioavailability of S(+)-Ketoprofen After Oral Administration of Different Mixtures of Ketoprofen Enantiomers to Dogs. *J. Clin. Pharmacol.* **1998**, *38*, 22S–26S.
80. Hutt, A.J.; Caldwell, J. The Importance of Stereochemistry in the Clinical Pharmacokinetics of the 2-Arylpropionic Acid Non-Steroidal Anti-Inflammatory Drugs. *Clin. Pharmacokinet.* **1984**, *9*, 371–373.
81. Soraci, A.L.; Benoit, E.; Olivier, L.; Delatour, P. Comparative Metabolism of R(–)-Fenoprofen in Rats and Sheep. *J. Vet. Pharmacol. Ther.* **1995**, *18*, 167–171.
82. Soraci, A.L.; Benoit, E. In Vitro Fenoprophenyl-Coenzyme A Thioester Formation: Interspecies Variations. *Chirality* **1996**, *7*, 534–540.
83. Benoit, E.; Soraci, A.; Delatour, P. *Chiral Inversion as a Parameter for Interspecies and Intercompound Discrepancies in Enantiospecific Pharmacokinetics*, 6th International Congress European Association for Veterinary Pharmacology and Toxicology, Blackwell Scientific Publications: Edinburgh, 1994; S5/7, 07.
84. Hayball, P.J.; Meffin, P.J. Enantioselective Disposition of 2-arylpropionic Acid Nonsteroidal Anti-Inflammatory Drugs. III. Fenoprofen Disposition. *J. Pharmacol. Exp. Ther.* **1987**, *240*, 631–636.
85. Rubin, A.; Knadler, M.P.; Ho, P.P.K.; Bechtol, L.D.; Wolen, R.L. Stereoselective Inversion of (R)-Fenoprofen to (S)-Fenoprofen in Humans. *J. Pharm. Sci.* **1985**, *74*, 82–84.
86. Berry, B.W.; Jamali, F. Presystemic and Systemic Chiral Inversion of R(–)-Fenoprofen in the Rat. *J. Pharmacol. Exp. Ther.* **1991**, *258*, 695–701.
87. McKellar, Q.A.; Delatour, P.; Lees, P. Stereospecific Pharmacodynamics and Pharmacokinetics of Carprofen in the Dog. *J. Vet. Pharmacol. Ther.* **1994**, *17*, 447–454.
88. Lees, P.; McKellar, Q.A.; May, S.A.; Ludwig, B.M. Pharmacodynamics and Pharmacokinetics of Carprofen in the Horse. *Equine Vet. J.* **1994**, *26*, 203–208.
89. Kemmerer, J.M.; Rubio, F.A.; McClain, R.M.; Koechlin, B.A. Stereospecific Assay and Stereospecific Disposition of Racemic Carprofen in Rats. *J. Pharm. Sci.* **1979**, *68*, 1274–1280.
90. Stoltenberg, J.K.; Puglisi, C.V.; Rubio, F.; Vane, F.M. High-Performance Liquid Chromatographic Determination of Stereoselective Disposition of Carprofen in Humans. *J. Pharm. Sci.* **1981**, *70*, 1207–1212.
91. Lees, P.; Delatour, P.; Benoit, E.; Foster, A.P. Pharmacokinetics of Carprofen Enantiomers in the Horse. *Acta. Vet. Scand. Suppl.* **1991**, *87*, 249–251.
92. Taylor, P.M.; Delatour, P.; Landoni, M.F.; Deal, C.; Pickett, C.; Shojaei, Aliabadi F.; Foot, R.; Lees, P. Pharmacodynamics and Enantioselective Pharmacokinetics of Carprofen in the Cat. *Res. Vet. Sci.* **1996**, *60*, 144–151.
93. Delatour, P.; Foot, R.; Foster, A.P.; Baggot, D.; Lees, P. Pharmacodynamics and Chiral Pharmacokinetics of Carprofen in Calves. *Br. Vet. J.* **1996**, *152*, 183–198.
94. Muir, W.W.; Hubbell, J.A. Cardiopulmonary and Anesthetic Effects of Ketamine and Its Enantiomers in Dogs. *Am. J. Vet. Res.* **1988**, *49*, 530–534.

95. White, P.F.; Ham, J.; Way, W.L.; Trevor, A.J. Pharmacology of Ketamine Isomers in Surgical Patients. *Anesthesiology* **1980**, *52*, 231–239.
96. Delatour, P.; Jaussaud, P.; Courtot, D.; Fau, D. Enantioselective N-demethylation of Ketamine in the Horse. *J. Vet. Pharmacol. Ther.* **1991**, *14*, 209–212.
97. Deleforge, J.; Davot, J.L.; Boisrame, B.; Delatour, P. Enantioselectivity in the Anaesthetic Effect of Ketamine in Dogs. *J. Vet. Pharmacol. Ther.* **1991**, *14*, 418–420.
98. Caldwell, J. The Importance of Stereochemistry in Drug Action and Disposition. *J. Clin. Pharmacol.* **1992**, *32*, 925–929.
99. Foster, R.T.; Jamali, F. High-Performance Liquid Chromatographic Assay of Ketoprofen Enantiomers in Human Plasma and Urine. *J. Chromatogr.* **1987**, *416*, 388–393.
100. Delatour, P.; Garnier, F.; Benoit, E.; Claude, I. Chiral Behaviour of the Metabolite Albendazole Sulphoxide in Sheep, Goats and Cattle. *Res. Vet. Sci.* **1991**, *50*, 134–138.
101. Pasutto, F.M. Mirror Images: The Analysis of Pharmaceutical Enantiomers. *J. Clin. Pharmacol.* **1992**, *32*, 917–924.
102. Carr, R.A.; Foster, R.T.; Lewanczuk, R.Z.; Hamilton, P.G. Pharmacokinetics of Sotalol Enantiomers in Humans. *J. Clin. Pharmacol.* **1992**, *32*, 1105–1109.
103. Drayer, D.E. Pharmacodynamic and Pharmacokinetic Differences between Drug Enantiomers in Humans: An Overview. *Clin. Pharmacol. Ther.* **1986**, *40*, 125–133.
104. Monteiro-Riviere, N.A.; Bristol, D.G.; Manning, T.O.; Rogers, R.A.; Riviere, J.E. Interspecies and Interregional Analysis of the Comparative Histologic Thickness and Laser Doppler Blood Flow Measurements at Five Cutaneous Sites in Nine Species. *J. Invest. Dermatol.* **1990**, *95*, 582–586.
105. Riegelman, S. Pharmacokinetics: Pharmacokinetic Factors Affecting Epidermal Penetration and Percutaneous Absorption. *Clin. Pharmacol. Ther.* **1974**, *16*, 873–883.
106. Montagna, W. Comparative Anatomy and Physiology of the Skin. *Arch. Dermatol.* **1967**, *96*, 357–363.
107. Pitman, I.H.; Rostas, S.J. Topical Drug Delivery to Cattle and Sheep. *J. Pharm. Sci.* **1981**, *70*, 1181–1193.
108. Ryder, M.L. A Survey of the Follicle Populations in a Range of British Breeds of Sheep. *J. Agric. Sci.* **1957**, *49*, 275–284.
109. Forsyth, B.A.; Gibbon, A.J.; Pryor, D.E. Seasonal Variations in Anthelmintic Response by Cattle to Dermally Applied Levamisole. *Aust. Vet. J.* **1983**, *60*, 141–146.
110. Idson, B. Vehicle Effects on Percutaneous Absorption. *Drug Metab. Rev.* **1983**, *14*, 207–222.
111. McKenzie, A.W. Percutaneous Absorption of Steroids. *Arch. Dermatol.* **1962**, *86*, 91–94.
112. Barry, B.W. Properties that Influence Percutaneous Absorption. *Dermatological Formulations: Percutaneous Absorption*; Marcel Dekker, Inc. New York, 1983, 127–233.
113. Gyrð-Hansen, N.; Brimer, L.; Rasmussen, R. Percutaneous Absorption and Recovery of Parathion in Pigs. *Acta Vet. Scand. Suppl.* **1991**, *87*, 410–412.
114. Scheuplein, R.J. Site Variations in Diffusion and Permeability. In *The Physiology and Pathophysiology of the Skin*; Jarrett, A., Ed.; Academic Press: New York, 1978; *5*, 1731–1752.
115. McCreesh, A.H. Percutaneous Toxicity. *Toxicol. Appl. Pharmacol.* **1965**, *7* (Suppl. 2), 20–26.
116. Monteiro-Riviere, N.A.; Stromberg, M.W. Ultrastructure of the Integument of the Domestic Pig (*Sus scrofa*) from one through Fourteen Weeks of Age. *Zbl Vet Med [Reihe C]* **1985**, *14*, 97–115.
117. Bartek, M.J.; LaBudde, J.A.; Maibach, H.I. Skin Permeability *In Vivo*: Comparison in Rat, Rabbit, Pig and Man. *J. Invest. Dermatol.* **1972**, *58*, 114–123.
118. Riviere, J.E.; Monteiro-Riviere, N.A. The Isolated Perfused Porcine Skin Flap as an *In Vitro* Model for Percutaneous Absorption and Cutaneous Toxicology. *Crit. Rev. Toxicol.* **1991**, *21*, 329–344.
119. Riviere, J.E.; Monteiro-Riviere, N.A.; Williams, P.L. The Isolated Perfused Porcine Skin Flap as an *In Vitro* Model for Predicting Transdermal Pharmacokinetics. *Eur. J. Pharmac. Biopharmac.* **1995**, *41*, 152–162.
120. Wester, R.C.; Maibach, H.I. Regional Variation in Percutaneous Absorption. *Percutaneous Absorption*; Bronaugh, R.L., Maibach, H.I., Eds.; Marcel Dekker, Inc.: New York, 1999; 107–116.
121. Behrendt, H.; Korting, H.Ch.; Braun-Falco, O. Zum Metabolismus von Pharmaka in der Haut. *Hautarzt* **1989**, *40*, 8–13.
122. Riviere, J.E.; Chang, S.K. Transdermal Penetration and Metabolism of Organophosphate Insecticides. In *Organophosphates: Chemistry, Fate and Effects*; Academic Press: New York, 1992.
123. Kao, J.; Carver, M.P. Cutaneous Metabolism of Xenobiotics. *Drug Metab. Rev.* **1990**, *22*, 363–410.
124. Ademola, J.I.; Chow, C.A.; Wester, R.C.; Maibach, H.I. Metabolism of Propranolol During Percutaneous Absorption in Human Skin. *J. Pharm. Sci.* **1991**, *82*, 767–770.
125. Ward, S.; Walle, T.; Walle, K.; Wilkinson, G.R.; Branch, R.A. Propranolol's Metabolism is Determined by Both Mephenytoin and Debrisoquin Hydroxylase. *Clin. Pharmacol. Ther.* **1989**, *45*, 72–78.
126. Dorman, D.C.; Buck, W.B.; Trammel, H.L.; Jones, R.D.; Beasley, V.R. Fenvalerate/ *N*, *N*-Diethyl-*m*-Toluamide (DEET) Toxicosis in Two Cats. *J. Am. Vet. Med. Assoc.* **1990**, *196*, 100–102.
127. Mount, M.E.; Moller, G.; Cook, J.; Holstege, D.M.; Richardson, E.R.; Ardans, A. Clinical Illness Associated with a Commercial Tick and Flea Product in Dogs and Cats. *Vet. Human. Toxicol.* **1991**, *33*, 19–27.
128. Munro, D.D.; Stoughton, R.B. Dimethylacetamide (DMAC) and Dimethylformamide (DMFA): Effect on Percutaneous Absorption. *Arch. Dermatol.* **1965**, *92*, 585–586.
129. Kyles, A.E.; Papich, M.; Hardie, E.M. Disposition of Transdermally Administered Fentanyl in Dogs. *Am. J. Vet. Res.* **1996**, *57*, 715–719.
130. Holley, F.O.; van Steennis, C. Postoperative Analgesia with Fentanyl: Pharmacokinetics and Pharmacodynamics of Constant-Rate IV and Transdermal Delivery. *Br. J. Anaesthesia.* **1988**, *60*, 608–613.



Viral Inactivation Issues in Aseptically Processed Parenterals

J. Hotta

A. Klos

S. Petteway, Jr.

D. Pifat

Bayer Corporation, Research Triangle Park, North Carolina, U.S.A.

INTRODUCTION

Plasma-derived therapeutic proteins are parenteral biologics that are purified on an industrial scale. All biologics derived from human sources, such as plasma, carry the risk of viral contamination. Thus, in order to market a medicinal product derived from human plasma, manufacturers must assure the absence of specific viral contamination. Virus validation studies are performed to evaluate the capacity of a manufacturing process to remove viral contaminants. Virus clearance across three different terminal inactivation steps, low pH incubation of immunoglobulins (IgG), pasteurization of albumin, and freeze dry/dry heat treatment of plasma-derived products (Factor VIII and Protein G), is discussed in this article. The data show that, like all other upstream virus reduction steps, the methods used for terminal inactivation are process and product dependent, and that the reduction factors for an individual step may be overestimated or underestimated due to inherent limitations or inadequate designs of viral validation studies.

Parenteral drugs are medicinal products that are not ingested but are injected through the skin, directly into a blood vessel, organ, tissue, or lesion. Since the normal body defenses against infection are bypassed during parenteral administration, preparation of these drugs requires the highest level of contamination control. Contamination can be controlled by terminal sterilization or by aseptic processing. Terminal sterilization involves filling final containers with product and sealing them under a high quality, but not necessarily sterile environment. The product in its final container is then subjected to a sterilization process. During aseptic processing, the product, container, and closure undergo separate sterilization processes and are then brought together. Since microorganisms may be inadvertently introduced during filling, these operations must be performed under stringent conditions to minimize the risk of microbial contamination (e.g., operating in a clean room environment, implementing environmental monitoring programs, and training personnel in aseptic techniques).

Plasma-derived therapeutic proteins are parenteral biologics that are purified on an industrial scale. Most

biologics are highly complex, heat-sensitive, protein mixtures that cannot be terminally sterilized and must be aseptically processed. In addition, all biologics derived from human sources, such as plasma, carry the risk of viral contamination, as the use of plasma-derived products has resulted in the transmission of hepatitis and AIDS. Thus, in order to market a medicinal product derived from human plasma, manufacturers must assure the absence of microbial, as well as specific viral contamination.

Viral contamination of complex biologics can be minimized by strict control of the source material (plasma) and the production process. Controlling the source material involves donor selection, rigorous testing of plasma for virus markers, and inventory hold (Fig. 1). Screening of plasma donors begins with a detailed health questionnaire to identify high-risk factors (e.g., family history of Creutzfeldt–Jakob disease) and behaviors (e.g., tattoos), followed by a physical exam and a background check against the National Donor Deferral Registry (NDDR). Source plasma collected by plasmapheresis (i.e., clear liquid remaining after removal and reinfusion of red cells, leukocytes, and platelets back to the donor) from qualified donors, rather than recovered plasma (i.e., plasma removed from entire blood donations) is used. Qualified donors are those applicants who have successfully passed two donor screenings and testing for viral markers within 6 mo. If a donor does not return to donate within 6 mo, the previous plasma donation is destroyed. Plasma units from applicants who are disqualified on the basis of test results are also destroyed and the names of the rejected donors are entered into the NDDR, a national database that lists all donors who are excluded from further donations.

BACKGROUND

Plasma inventory hold refers to the withholding of all plasma units from qualified donors for 60 day. If a donor, on a repeat visit, shows evidence of viral infection (e.g., seroconversion) or indicates involvement in potential high-risk activity, the donor's previous units held in inventory can be retrieved and destroyed.

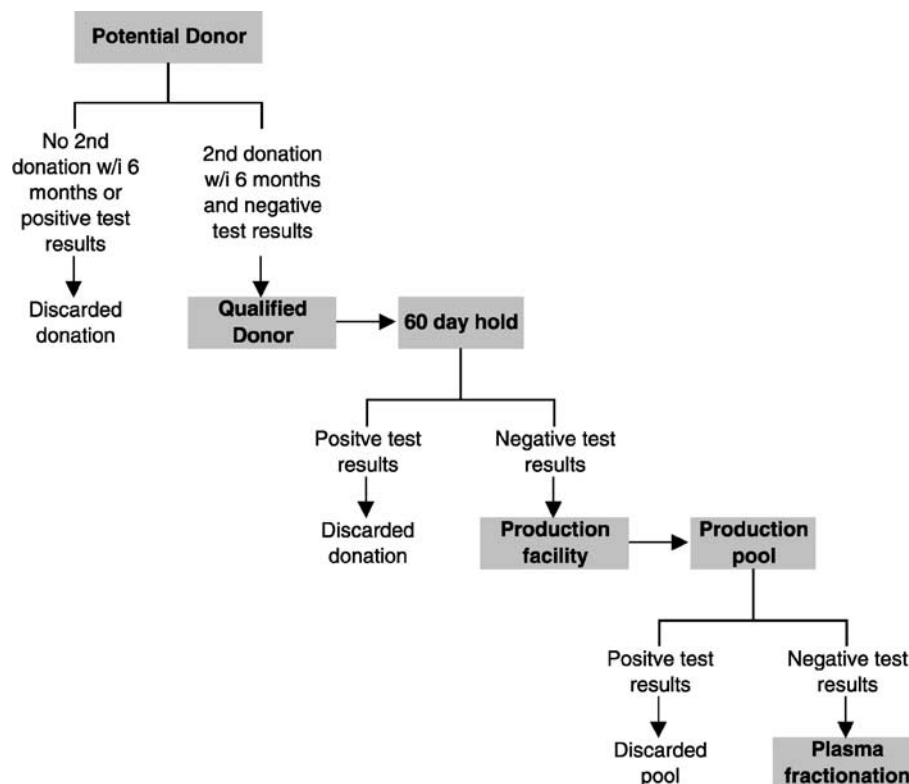


Fig. 1 Controlling source material by the screening and testing of plasma.

All donations must be tested and found nonreactive for syphilis, hepatitis B surface antigen (HBsAg) and negative for antibodies (Ab) to human immunodeficiency virus (HIV) types 1 and 2, and hepatitis C virus (HCV). Many manufacturers are also testing all plasma donations that are negative by standard serological testing, by nucleic acid amplification techniques (NAT) for the presence of viral genomes such as HCV, HIV-1, hepatitis B virus (HBV) and parvovirus B19. Plasma units that are NAT reactive for HCV, HIV-1 and HBV or contain high levels of parvovirus B19 DNA are discarded.

The plasma units that meet viral testing standards by NAT are pooled at the production facility and the pools are retested for the presence of selected viral markers (e.g., HBsAg, Ab to HIV-1/2 and HCV, and HCV RNA). Thus, all source material is screened to reduce or to confirm the absence of many clinically relevant viruses and to minimize any potential viral challenge to the production process.

During the production process, plasma is fractionated or separated into fractions with different biological functions. Because of the complex mixture of proteins, the inability to apply terminal sterilization and the need for tight control of source material, the entire manufacturing

process must follow current good manufacturing practices (cGMP). Biological products are defined more by their manufacturing processes, rather than by analytical testing of final product, to assure the final product is pure, potent, stable, free of microbial contamination, and carries a low risk of viral contamination.

Virus validation studies are performed to evaluate the virus reduction capacity of a manufacturing process and a number of regulatory guidance documents have been issued that describe how to perform and interpret virus clearance experiments.^[1,2] During these studies, production product intermediates are spiked with significant amounts of virus, the material is processed and virus reduction is measured by comparing the level of virus infectivity remaining in the final product-containing fraction with that in the starting material. Process validation is usually performed using the same type and size of equipment that will be used at commercial scale. However, the deliberate contamination of the designated manufacturing equipment and facility with infectious virus would be unacceptable. As a result, virus validation studies are conducted using small-scale replicas and processing conditions representative of production scale.



Reduction, across individual steps of a process, is expressed on a logarithmic scale because virus validation studies can only assure reduction of the virus load by a certain factor. Total reduction of the virus load to zero cannot be guaranteed. The removal/inactivation of the input virus load to below detection or the reduction of $4 \log_{10}$, or more, virus is considered significant. Virus reduction factors for process steps that clear the entire input virus load to below detection, are distinguished from virus reduction factors for steps whose final fractions still contain infectious virus, by the symbol " \geq ." For example, a virus reduction factor of $\geq 3.7 \log_{10}$ would indicate that no infectious virus was detected in the final fraction and that virus reduction could have been greater than $3.7 \log_{10}$ if the input virus spike had been greater. A virus reduction factor of $4.3 \log_{10}$ means a process step cleared significant levels of virus but residual virus could be detected in the final product-containing fraction.

The overall virus reduction capacity of a manufacturing process is the sum of the individual reduction factors, and should be greater than the potential virus load in the starting material. At least one step in the process should clear significant levels of infectious virus so that the overall clearance is not made up of individual small, and possibly negligible, reductions.

Virus reduction steps may be placed upstream or at the very end of the manufacturing process. Reduction in viral infectivity occurs by inactivation of virus or by removal of virus particles. During removal steps, virus is not inactivated but is separated from the protein of interest using methods such as precipitation, chromatography, or filtration. For example, during an ethanol precipitation step, ethanol is added to a suspension to precipitate unwanted contaminating proteins and viruses. The ethanol-containing suspension is then centrifuged so that the contaminants in the precipitated paste fraction can be separated from the product in the effluent fraction.

During inactivation steps, viral infectivity is reduced by treatment with chemicals and/or physical methods. Remnants of virus particles (e.g., viral nucleic acids) may remain in the product-containing fraction but are not infectious. Chemical methods of virus inactivation, such as treatment with solvent-detergent or acetone, must be placed upstream, since subsequent steps are needed to remove or reduce the levels of the toxic chemicals. Terminal inactivation is often achieved using physical methods, such as heat and low pH, because these methods leave no chemical residues. After treatment, the final products are delivered to patients, so aseptic processing conditions must be maintained throughout terminal inactivation steps and the parameters for virus inactivation must be balanced with the conditions to preserve product quality and yield.

The focus of this article is to discuss virus reduction issues that arose during studies of three different terminal inactivation steps: low pH incubation of immunoglobulins (IgG), pasteurization of albumin, and freeze dry/dry heat treatment of plasma-derived products (Factor VIII and Protein G). The data show that, like all other upstream virus reduction steps, the methods used for terminal inactivation are process and product dependent, and the reduction factors for an individual step may be overestimated or underestimated due to inherent limitations or inadequate designs of viral validation studies.

SELECTION OF VIRUSES FOR VIRUS VALIDATION STUDIES

A major issue in performing virus validation studies is determining which viruses should be used. The Committee for Proprietary Medicinal Products (CPMP) has issued guidelines on the selection of viruses to evaluate in validation studies.^[1] Processes must be validated for their capacity to inactivate/remove relevant viruses, or viruses that are known to contaminate plasma or other materials in the production process. If relevant viruses cannot be easily propagated in cell culture or assayed, then validation studies should include specific model viruses with characteristics similar to relevant viruses. If relevant viruses do not represent viruses with a wide range of physico-chemical properties, then validation studies should also include nonspecific model viruses with many different properties. Inactivation/removal of these nonspecific model viruses provides assurance that the manufacturing process is capable of clearing a diverse range of known and unknown viruses.

The virus reduction studies of the three process steps discussed here were performed with HIV-1, Bovine viral diarrhea virus (BVDV), Pseudorabies virus (PRV), Reovirus type 3 (Reo), Hepatitis A virus (HAV), and Porcine parvovirus (PPV). HIV-1 was included as a relevant enveloped virus, while BVDV and PRV were tested as specific model viruses for HCV and HBV, respectively (Table 1). Reo was chosen as a nonspecific model nonenveloped virus, HAV was included as a relevant virus and PPV was used as a surrogate for human parvovirus B19. All viruses were propagated using standard cell culture conditions.^[3,4] The appropriate cell lines were infected, at a low multiplicity of infection, and incubated until 4+ cytopathic effects were observed. The infected cells were frozen and thawed three times to release virus, centrifuged at low speed to remove cell debris and the clarified supernatants were removed for use as virus spikes.

**Table 1** Test viruses

Characteristic	Enveloped viruses		
	Human immuno-deficiency virus (HIV)	Bovine viral diarrhea virus (BVDV)	Pseudorabies virus (PRV)
Relevant or model for Strain	HIV-1, HIV-2 (III B)	HCV Kentucky-22 (BRFF)	CMV, EBV, HSV, HBV PRV dl tk (ATCC VR-2074), Becker (Dupont/Merck)
Nucleic acid	RNA	RNA	DNA
Enveloped	Yes	Yes	Yes
Size (nm)	80–130	80–100	120–220
Resistance to physico-chemical agents	Low	Medium	Medium
	Nonenveloped viruses		
	Reovirus type 3 (Reo)	Hepatitis A virus (HAV)	Porcine parvo virus (PPV)
Relevant or model for Strain	Nonenveloped virus Abney (ATCC VR 232)	HAV HM175 18f (ATCC)	Human parvovirus B19 NADL-2 (ATCC VR 742)
Nucleic acid	RNA	RNA	RNA
Enveloped	No	No	No
Size (nm)	60–80	27–32	18–26
Resistance to physico-chemical agents	High	Very high	Very high

Terminal Inactivation Step: Low pH Final Container Incubation of IgG

Although the partitioning of viruses during ethanol fractionation and the presence of neutralizing antibodies are thought to be the main factors involved in establishing the safety of immunoglobulins,^[5] transmission of HCV by intravenous immunoglobulin (IVIg) has been reported.^[6] Many IgG preparations are formulated at low pH and held for extended periods. The low pH hold step was originally designed to stabilize IgG, reduce anticomplement activity, and allow for intravenous administration.^[7,8] Virus validation studies showed incubation at low pH also inactivated significant levels of enveloped viruses.

The three human IgG products discussed here were purified from Fraction II + III paste by slightly different processing methods but all three shared the same final formulation in 0.2 M glycine, pH 4.25. All three were 10% protein solutions, of which 98% was IgG, and their monomer contents were greater than 90%. At production scale, IgG final bulks (pH 4.25) are typically sterile filtered, aseptically filled into sterile final containers and incubated at 20–27°C for 21 days–28 days.^[9] For the virus studies, however, IgG final bulk material was spiked with virus and adjusted to pH 4.3 or 4.5 before incubating at 5, 20, or 23°C, for up to 28 days. Hanks' Balance Salt Solution (HBSS) or IgG, adjusted to pH 7, were also spiked with virus and included as positive controls.

Aliquots for virus titration were removed on day 0 and at various times during incubation.

The extent and kinetics of reduction were dependent on the test viruses and the exact processing conditions (e.g., pH, time, and temperature). Virus inactivation was greater with lower pH, higher temperatures and longer incubation times and the kinetics of HIV-1 and PRV inactivation were more rapid than BVDV. HIV-1 and PRV were always inactivated to below detection after incubating for 14 days at low pH, 20–23°C but the kinetics of inactivation were slightly different in different IgG solutions (Fig. 2 and Table 2). The entire input load of HIV-1 or PRV was inactivated sooner in IgG solution 1 than in IgG solutions 2 or 3. HIV-1 was below detection after 21 days in IgG solution 1, regardless of pH and temperature, but significant levels of infectious virus were still present after 21 days in IgG solution 2, pH 7 or 5°C. These discrepancies could be due to minor differences in pH (pH 4.3 vs. pH 4.5), temperature (23°C vs. 20°C), or virus preparation.

To control these experimental differences, BVDV inactivation was monitored, in parallel, during low pH incubation, in IgG solutions 1 and 3, using the same stock of virus (Fig. 3 and Table 2). The data show minor changes in pH and temperature resulted in major differences in virus reduction. For example, BVDV reduction in IgG solution 1, 20°C, was 4.2 log₁₀ at pH 4.3 but at pH 4.5, BVDV reduction was only 2.6 log₁₀.

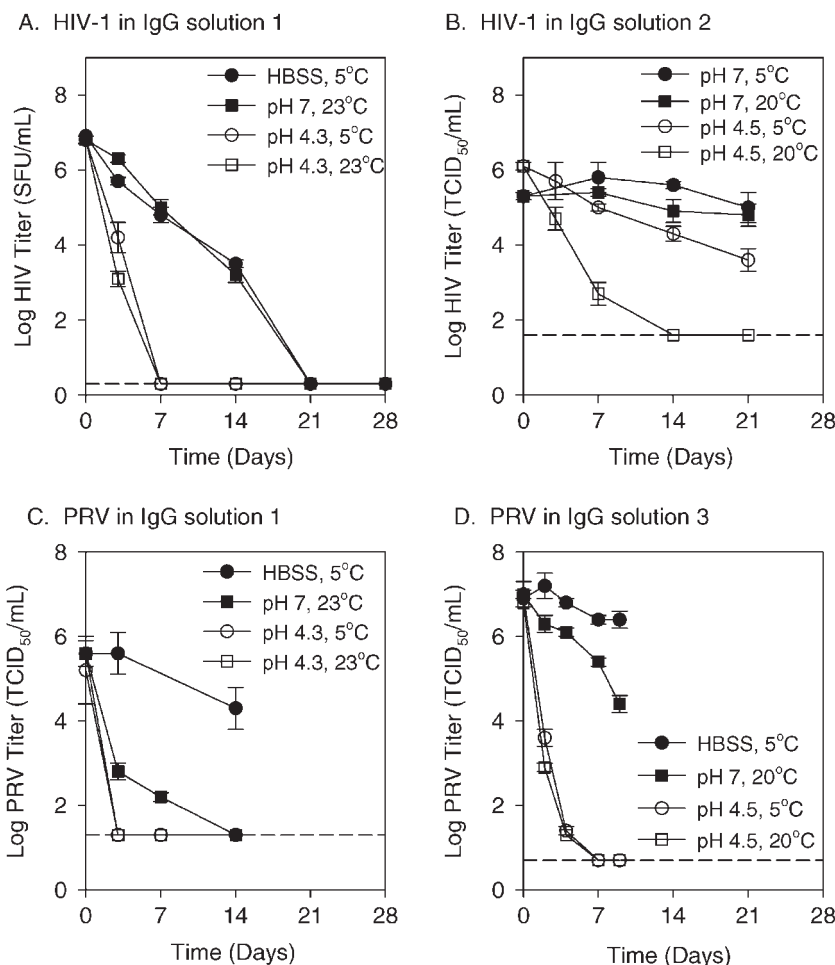


Fig. 2 Kinetics of virus inactivation during low pH incubation in IgG solutions. Virus was spiked into IgG solutions, before adjusting to pH 7, 4.5, or 4.3, and the solutions were incubated at 5, 20, or 23°C. HBSS was also spiked as a positive control. Aliquots for virus titration were removed immediately after spiking ($t = 0$ day) and at various times during incubation. IgG solution 1: HIV-1 (A) or PRV (C), IgG solution 2 or 3: HIV-1 (B) or PRV (D), dashed line, no symbol = virus detection limit.

BVDV reduction was also a function of the IgG test solution. After 21 days in IgG solution 3, pH 4.3, 5°C, no virus could be detected but approximately 4 log₁₀ BVDV was still present in IgG solution 1 under the same conditions. For IgG solution 1, significant BVDV reduction (4.2 log₁₀) was observed at pH 4.3 at 20°C and 23°C but at pH 4.5, less than 4 log₁₀ inactivation was achieved. In contrast, significant BVDV reduction (greater than 4 log₁₀) was observed in IgG solution 3 at both pH 4.3 and 4.5, regardless of temperature.

The results are consistent with other studies in demonstrating the effectiveness of low pH incubation to inactivate enveloped viruses. After incubating IVIg (6% protein, 8% sucrose) at pH 4.4 for one week at 35°C, or for four weeks at 23°C, ≥ 6 log₁₀ Vesicular Stomatitis Virus (VSV) was inactivated.^[10] IVIg (8% protein, 16% maltose) incubated at pH 4 (in the presence of pepsin) for 22 hr at

35°C inactivated mumps, Semliki Forest Virus (SFV), Herpes simplex virus (HSV), and vaccinia virus but had no effect on the nonenveloped poliovirus type 2.^[11] IVIg (7–8% protein, glucose-containing buffer) at pH 4.25 (in the presence of pepsin), 37°C, for 30 hr inactivated HIV, PRV, and BVDV to below detection.^[12]

Immunoglobulins are biologics that differ in many respects from classically synthesized drugs as there are no chemically defined formulas for the many different classes of IgG molecules, each with their own unique antigen recognition capability. Differences in upstream processing may yield product with slightly different protein profiles and these changes could impact virus reduction. For example, in the studies presented here, IgG solutions 1 and 3 were 10% protein solutions with the same formulation but the kinetics and extent of virus inactivation were different in the two products. After incubating at pH 4.5,

Table 2 Summary of virus reduction during incubation at low pH

Test solution	pH	Log ₁₀ virus reduction						
		HIV-1 ^a		PRV ^b		BVDV ^c		
		5°C	20/23°C	5°C	20/23°C	5°C	20°C	23°C
HBSS	7	≥ 6.6	■	1.2	1.0	1.6	■	■
IgG solution 1	7	■	≥ 6.5	■	≥ 4.3	■	■	< 1.0
	4.5	■	■	■	■	■	2.6	3.6
	4.3	≥ 6.5	≥ 6.5	≥ 3.9	≥ 4.3	1.8	4.2	4.2
HBSS	7	na	■	< 1.0	■	na	■	■
IgG solution 2 or 3	7	< 1.0	< 1.0	< 1.0	1.6	■	■	1.2
	4.5	2.5	≥ 4.5	≥ 6.3	≥ 6.1	■	4.1	≥ 4.1
	4.3	■	■	■	■	≥ 4.5	≥ 4.5	≥ 4.4

na = not applicable and gray boxes = not done.

^a HIV-1 studies with IgG solutions 1 and 2 were conducted at different times with different virus stocks and slightly different temperatures. Solution 1 study was 28 days, the incubation temperatures were 5 and 23°C, and titers were calculated as SFU/ml. Solution 2 study was 21 days, the incubation temperatures were 5 and 20°C and titers were calculated as TCID₅₀/ml.

^b PRV studies with IgG solutions 1 and 3 were conducted at different times, using different virus stocks and slightly different temperatures. Solution 1 study was 14 days and the incubation temperatures were 5 and 23°C, while solution 3 study was 9 days and the incubation temperatures were 5 and 20°C.

^c BVDV studies with IgG solutions 1 and 3 were conducted at the same time, using the same virus stock and lasted 28 days.

20°C in IgG solution 3, BVDV inactivation was 4.1 log₁₀ but only 2.6 log₁₀ in IgG solution 1. Significant reduction of BVDV in IgG solution 1 will most likely require incubation at higher temperatures (e.g., 25°C) or lower pH (e.g., pH 4.0). Thus, data which demonstrate the virus safety of a product should not be transferred to other products, even though the two products may appear to be biochemically similar, and each process should be carefully validated.

Terminal Inactivation Step: Pasteurization of Albumin

Pasteurization is the heating of aqueous, stabilized protein solutions. Pasteurization has been used for bulk preparations of Factor VIII, antithrombin III, and alpha₁-antitrypsin but wet heat treatment of these proteins requires the addition of stabilizers that must be removed by a downstream process step.^[13] Removal of stabilizers and denatured proteins from pasteurized materials extends processing times and may involve significant losses in yield.^[14] The development of antibodies to FVIII is a complication of treatment and pasteurization of FVIII concentrates raised concerns regarding the possible induction of FVIII neoantibodies.^[15]

Pasteurizing at 60°C, for 10 hr, in final container, is the pharmacopoeial method to inactivate viruses in albumin preparations. Sodium caprylate and acetyl-tryptophan are often added as stabilizers, which bind to albumin to prevent denaturation or aggregation during heating. So far, only albumin preparations can withstand pasteurization in the final container because the quantity

of caprylate and acetyl tryptophan added are compatible for intravenous administration.^[16] Pasteurization is an effective inactivation step for both enveloped and nonenveloped viruses, as other investigators have shown pasteurization of 20% albumin resulted in ≥ 5 log₁₀ HIV-2, ≥ 7.5 log₁₀ MuLV, ≥ 3.5 log₁₀ PRV, 4.2 log₁₀ Sindbis, ≥ 5.5 log₁₀ Polio, and 6.6 log₁₀ SV40 reduction.^[17] The strongest evidence for safety is in the clinical setting where historical data show that albumin (and plasma protein fraction) solutions have not transmitted viral disease since pasteurization in final container was introduced.^[18]

Evidence for the increased assurance of safety from heating in final container rather than heating in bulk comes from the 1973 outbreak of HBV that was associated with plasma protein fraction (PPF). Investigations revealed that a small amount of PPF had sequestered in a "sampling neck" of a bulk processing container and had not been adequately heated. In contrast, final containers of albumin, made by the same manufacturer, were completely submerged in a waterbath before heating and did not transmit HBV.^[18,19]

The data discussed here are for Plasbumin[®]-25, which is purified from Fraction V paste and consists of 23.5%–26.5% protein, of which no less than 96.5% is albumin. The preparation is stabilized with 0.02 M sodium caprylate and 0.02 M acetyltryptophan and contains 145 mEq/L sodium.^[20] For the virus clearance experiments, virus was spiked into albumin, pH 6.4–7.4, and the solution was heated at 60°C for 10 hr. Aliquots for virus titration were removed as soon as virus was added (preheat), when the temperature of the solution reached 60°C (0 hr) and at

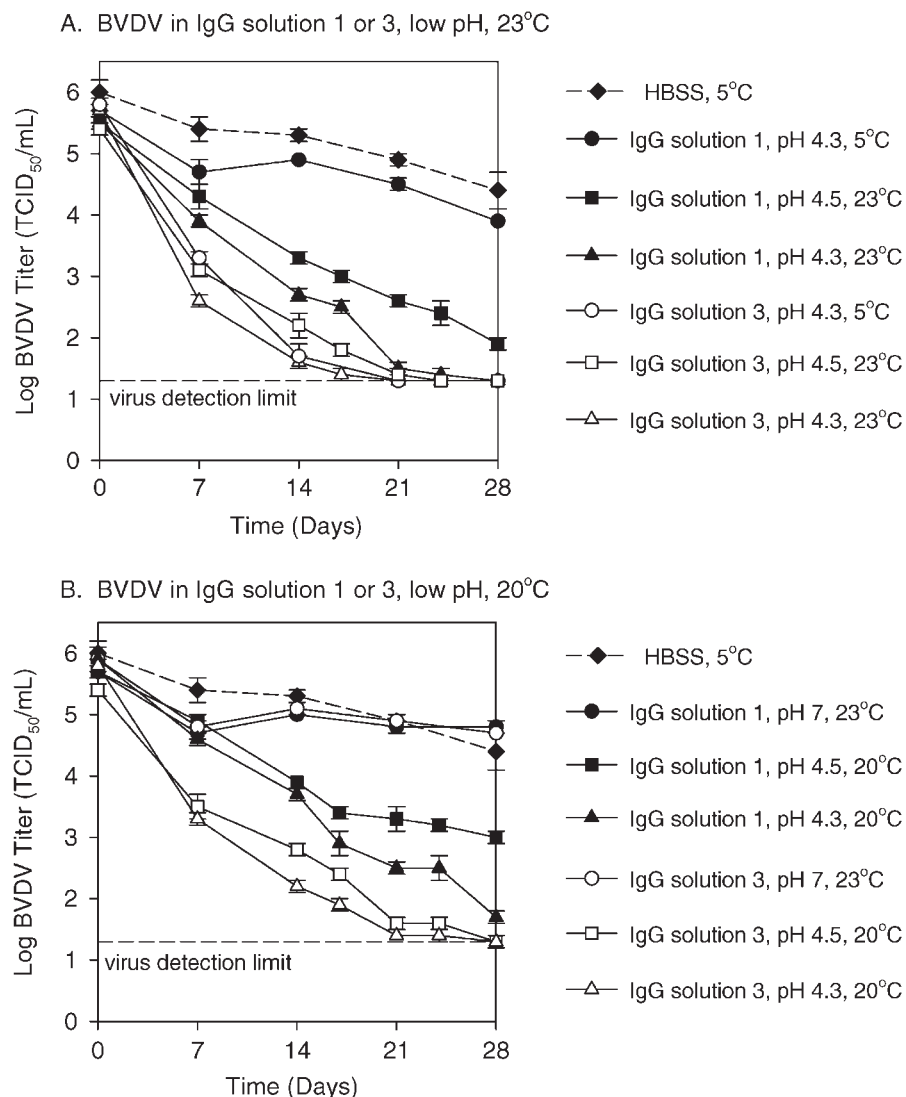


Fig. 3 Kinetics of BVDV inactivation during low pH incubation in IgG solution 1 or 3 at 23°C (A) or 20°C (B). The methods were as described in Fig. 2.

various times during the pasteurization cycle. Unheated albumin and HBSS were also spiked to the same dilution with virus and tested as positive controls.

Enveloped virus inactivation was biphasic, as HIV-1, PRV, and BVDV were quickly inactivated within the first 2 hr of pasteurization. Although inactivation was much slower after 2 hr, all enveloped virus infectivity was below detection after 5 hr (Fig. 4). Although not completely inactivated, the levels of nonenveloped viruses, Reo and HAV (strain HM175/18f), decreased significantly, resulting in 5.6 and 4.4 log₁₀ reduction, respectively (Fig. 4 and Table 3). PPV was the most resistant to pasteurization, as less than 2 log₁₀ virus reduction was observed (Fig. 4 and Table 3).

Additional experiments were performed to examine the robustness of pasteurization to inactivate the most

physico-chemically resistant viruses, HAV and PPV. For these experiments, one parameter (e.g., pH, temperature, or protein concentration) was deliberately set at the extreme limit for processing or set outside normal production processing ranges. All other parameters were maintained at the production setpoints (Fig. 5 and Table 3). The data showed that PPV reduction did not change significantly under any of the conditions tested, while HAV reduction was affected by changes in temperature and protein concentration. When heated below or at the lower limit for temperature, HAV reduction was less than 4 log₁₀. In 25% and 26.5% protein, greater than 4 log₁₀ HAV reduction was achieved but at 5% and 7.5% albumin, HAV reduction was approximately 3 log₁₀. Lower reduction at lower protein concentrations was unexpected since higher concentrations of protein are

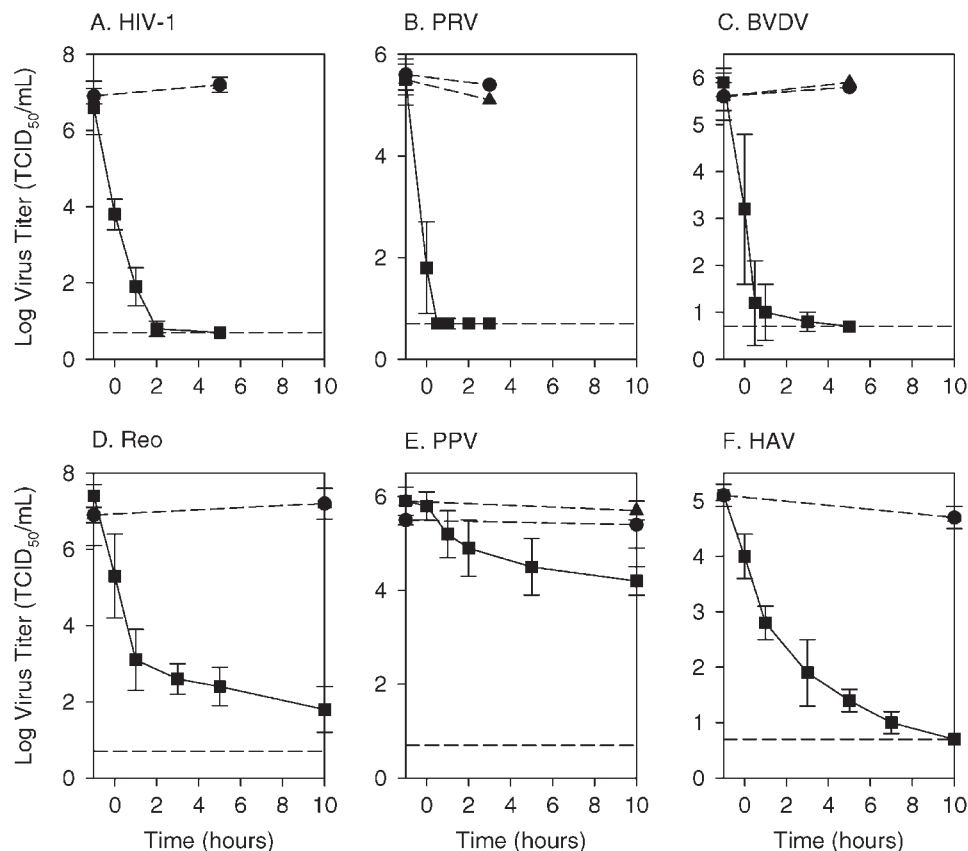


Fig. 4 Kinetics of virus inactivation during pasteurization of 25% albumin. (A) HIV-1, (B) PRV, (C) BVDV, (D) Reo, (E) PPV, and (F) HAV. Virus was spiked into albumin and an aliquot was removed for immediate titration (t = Preheat). Timing of the pasteurization cycle started when the temperature reached 60°C (t = 0 hr). Unheated albumin and Hanks' Balanced Salt Solution (HBSS) were also spiked and incubated at 5°C, as positive controls. Aliquots for virus titration were removed at various times during pasteurization (closed circles = HBSS, 5°C, closed triangles = 25% albumin, 5°C, closed squares = 25% albumin, 60°C, dashed line, no symbol = virus detection limit).

Table 3 Summary of virus reduction during pasteurization of 25% albumin

Test parameter		Log ₁₀ virus reduction					
		HIV-1	PRV	BVDV	Reo	PPV	HAV
pH	6.4	■	■	■	■	2.3	≥ 4.6
	6.9/7.0	≥ 5.9	≥ 4.8	≥ 5.2	5.6	1.6	4.4
	7.4/7.5	■	■	■	■	1.7	4.5
Temperature (°C)	58	■	■	■	■	1.6	3.2
	59	■	■	■	■	1.9	3.3
	59.5	■	■	■	■	■	3.9
	60	≥ 5.9	≥ 4.8	≥ 5.2	5.6	1.6	4.4
	60.5	■	■	■	■	■	≥ 4.4
Protein concentration (%)	5.0	■	■	■	■	1.7	3.3
	7.5	■	■	■	■	1.9	3.5
	25.0	≥ 5.9	≥ 4.8	≥ 5.2	5.6	1.6	4.4
	26.5	■	■	■	■	1.5	4.6

Gray boxes = not tested.

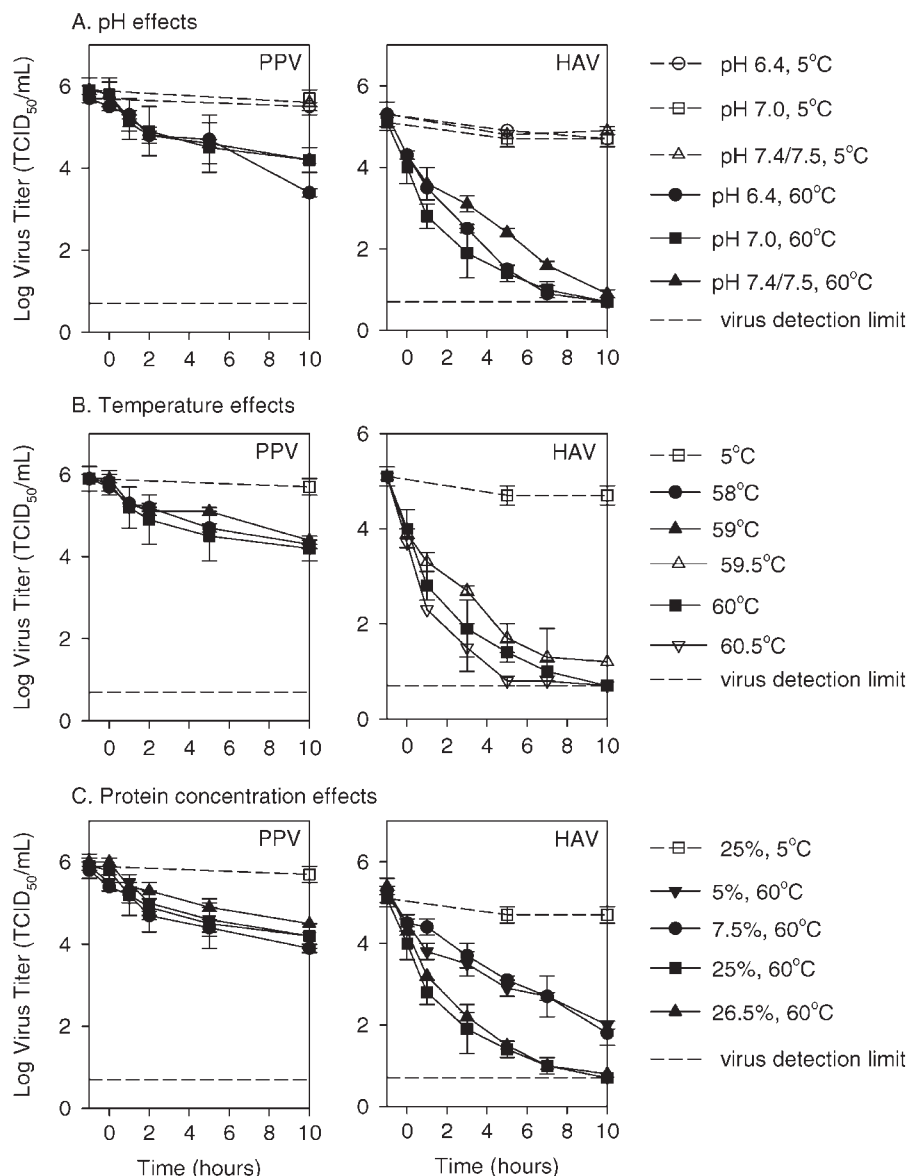


Fig. 5 Effect of (A) pH, (B) temperature, and (C) protein concentration on PPV and HAV inactivation during pasteurization of 25% albumin. Methods were as described in Fig. 4, except different pH, temperature, or protein concentrations were tested.

expected to provide virus with better protection from physico-chemical treatments than lower concentrations. The ratio of stabilizer to protein is the same for both 5% and 25% albumin. However, since 25% albumin contains five times more protein than 5% albumin, the concentrations of stabilizers (sodium caprylate and acetyl tryptophan) are also five times higher. Thus, the discrepancy in HAV reduction in 25% and 5% albumin may be related to the concentrations of stabilizers (or virus destabilizers) in the solutions.

All viruses used in clearance studies are actually model viruses since they have been adapted to grow in cell

culture and laboratory adapted virus strains may be slightly different from clinical isolates. Antigenic and genetic variations have been reported for laboratory adapted strains of HAV.^[21] Two different clones of HAV strain HM175, 18f and 24a, were pasteurized in 25% and 5% albumin and their kinetics of inactivation were compared. The results show HAV HM175/18f was much more resistant to pasteurization than HAV HM175/24a (Fig. 6, Table 4). The titers of HAV HM175/18f were slightly higher than that in the previous experiment but the extent and kinetics of inactivation were similar. HAV HM175/18f inactivation in 25% albumin was 4.5 log₁₀ but

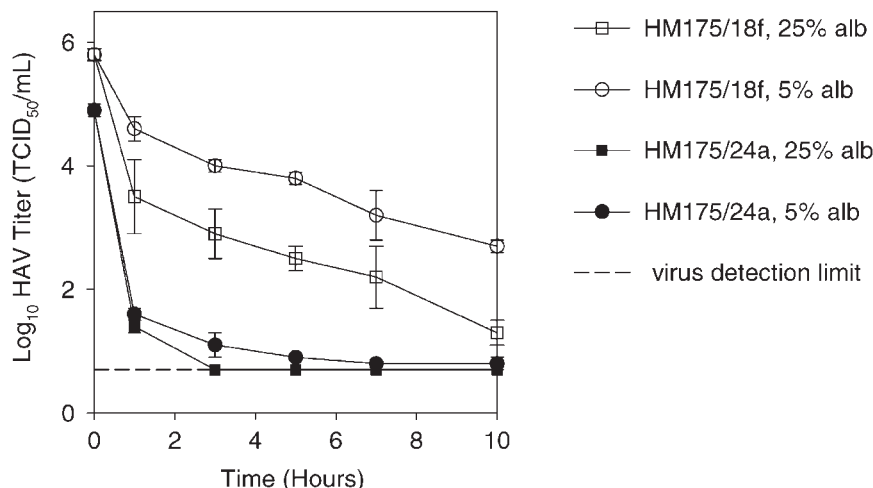


Fig. 6 Kinetics of HAV strain HM175/18f or HM175/24a inactivation during pasteurization in 25% or 5% albumin. The methods were as described in Fig. 4.

was 3.1 \log_{10} in 5% albumin. Inactivation of HAV HM175/24a was not significantly affected by protein concentration, as 4.2 \log_{10} reduction was observed during pasteurization in both 25% and 5% albumin.

Laboratory adapted strains of HAV are often associated with cellular membranes. Propagation of HAV involves freezing and thawing of infected cells to release virus. During its release from disrupted cells, virus may become entangled or associated with cellular membranes, which may shield virus from neutralization by antibody.^[22] To determine if cell associated lipids protect virus from heat denaturation, HAV HM175/18f was extracted with chloroform and inactivation of the extracted virus and untreated virus, during albumin pasteurization, was compared. The results show inactivation of chloroform-extracted virus was significantly faster than untreated virus (Fig. 7).

PPV was also extracted with chloroform, but chloroform extraction had little effect on PPV inactivation during albumin pasteurization (Fig. 7). These results are comparable to previous reports of only 1–2 \log_{10} Minute Virus of Mice, another parvovirus, inactivation after 6 hr of pasteurization in 1% albumin.^[23] Thus, heat stability during pasteurization may be a parvovirus-specific trait,

which is not due to virus association with cellular membranes.

In summary, pasteurization is effective for inactivation of many enveloped and nonenveloped viruses but their kinetics of inactivation are different. In the experiments with 25% albumin, enveloped viruses were below detection after heating for 5 hr, leaving a wide margin of safety. In contrast, nonenveloped virus reduction was slower and required the entire 10 hr of heating. Temperature, time, protein concentration, and possibly stabilizers were critical parameters for virus inactivation.

The data also demonstrated that the subtle differences in viruses should not be underestimated. The different HAV strains/clones and the method of virus preparation had a significant impact on virus clearance. Virus stocks used in virus validation studies are produced in cell culture and the behavior of tissue culture derived viruses may be different from that of native viruses. Laboratory adapted strains of viruses may also have unpredicted properties, such as association with lipids, which may affect their properties. Regardless of the panel of test viruses used in clearance studies, all virus stocks used should be strictly controlled and produced from qualified and traceable cell lines and virus seed banks. Repeated passage of the original stock virus is not recommended because mutations and deviations from the original reference material can result.

Table 4 HAV reduction during pasteurization in albumin

Test solution	Log ₁₀ HAV reduction	
	HM 175/18f	HM175/24a
25% albumin	4.5	4.2
5% albumin	3.1	4.2

Terminal Inactivation Step: Dry Heat Treatment of Freeze Dried Product

Proteins undergo chemical (e.g., deamidation or oxidation) and/or physical degradation (e.g., aggregation and precipitation) in aqueous solutions. Many of these

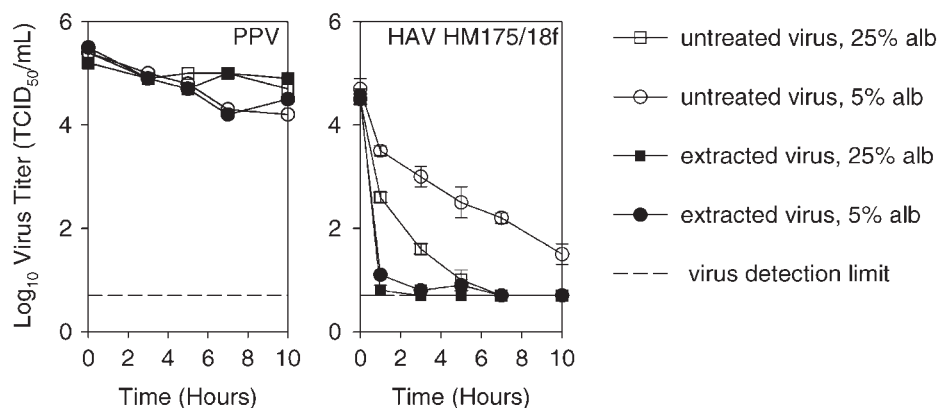


Fig. 7 Effect of chloroform extraction on virus inactivation during pasteurization in albumin. For these experiments, the virus spike was divided into half. One-half was extracted with an equal volume of chloroform while the other-half was prepared as usual. The untreated and chloroform extracted virus were then used as virus spike in pasteurization experiments as described in Fig. 4.

destabilizing reactions are minimized when water is removed and proteins remain in a dried solid state. This process of removing water from the product is called freeze drying or lyophilization, and is an effective way to improve the stability and extend the shelf-life of labile proteins. During the freeze drying step, final containers are aseptically filled, the product is frozen and then the frozen solid is dried by the sublimation of ice under vacuum.

Many freeze dried coagulation concentrates are heated to reduce the risk of virus transmission. During terminal dry-heat treatment, freeze dried FVIII concentrate in final container is heated for periods ranging from several minutes to several days at temperatures ranging from 60 to 100°C.^[4,24–26] This method was originally developed to inactivate hepatitis viruses^[27] but was later recommended by the National Hemophilia Foundation^[28] and the Centers for Disease Control^[29] to reduce the transmission of HIV.

Process validation associated with a freeze dried/dry heated product is more challenging than for products with liquid formulations for some of the following reasons:

1. Assuring specified fill volumes is difficult because many formulations consist of small amounts of the active agent and large amounts of stabilizers. Unlike a powder or liquid fill, a low fill is not readily apparent after freeze drying.
2. The number of variables to ensure aseptic processing is high. Before aseptic fill, different parts of the final product must be sterilized by different methods, such as dry heat to sterilize and depyrogenate glass containers, autoclaving and drying to sterilize and remove moisture from rubber stoppers, and sterile filtration for the liquid formulated product. After aseptic fill, the stoppers are placed on top of the vials

and the partially stoppered vials are transported, loaded into a freeze dryer, and then freeze dried. During freeze drying, the sterilized material in the vials in the freeze dryer will be exposed to the chamber's nitrogen or air, so bacterial retentive filters must be used in all vacuum/air lines to the chamber. Thus, the risk of contamination exists throughout the process until the vials are actually stoppered at the end of the freeze drying cycle.

3. Equipment reuse also carries a risk of contamination, so cleaning procedures for freeze dryers should be validated for their effectiveness to inactivate virus.
4. The technology for freeze drying/dry heat treatment is complex and there are many process parameters to control (e.g., temperatures of the shelf, product and condenser; freezing and drying rates; pressures of the chamber and condenser; and temperature and distribution of heat in the oven).

A process step with many variables to control is equally difficult to scale down for bench scale studies. However, results from animal studies and human clinical trials have been consistent with the “in vitro” model virus clearance studies of the terminal freeze dry/dry heat treatment.

1. Virus clearance studies of Factor VIII (Hemofil T) indicated the freeze dry/dry heat treatment (60°C, 72 hr) step was effective in inactivating HIV (≥ 3 log₁₀ HIV-1 reduction), but was less effective for HBV-like viruses (> 1 log₁₀ PRV reduction, 2.4 log₁₀ VSV reduction) and for HCV-like viruses (2.9 log₁₀ Sindbis reduction).^[30] Studies in animals and humans confirmed the model virus results. Chimpanzees inoculated with Hemofil T, that had been spiked

with HBV (30,000 chimpanzee infectious doses) and freeze dried/dry heated, later developed hepatitis B.^[27] During clinical trials, none of the patients treated with heated Hemofil T became HIV-1 positive^[31] but 11 out of 13 patients developed non-A, non-B hepatitis.^[32]

2. Virus clearance studies of a solvent/detergent treated, freeze dried/dry heated (100°C, 30 min) Factor VIII concentrate (Emoclot), demonstrated significant inactivation of HAV and poliovirus 1 by the terminal freeze dry/dry heat step.^[24] No model parvovirus was included as a test virus in these in vitro studies. During clinical trials, 13 previously untreated patients were given Emoclot and none of them became anti-HAV positive. In contrast, B19 seroconversion and viremia was observed in 5 of 11 susceptible patients.^[33]

Koate®-DVI is a freeze dried concentrate of human plasma-derived Factor VIII (FVIII) that is stabilized

with albumin (10 mg/ml) and formulated in a buffer containing 0.05 M glycine, 3 mM calcium, 0.06 M histidine.^[34] After freeze drying, it is terminally heated in final container for 72 hr at 80°C. Water molecules normally help to maintain proteins in their native state but, in freeze dried material, water molecules may contribute to degradative reactions. The following studies were performed to demonstrate that controlling residual moisture levels in freeze dried/ dry heated cakes may be a method for achieving viral inactivation in the Koate-DVI process.^[4]

Vials of virus-spiked FVIII were freeze dried, sealed with stoppers containing different amounts of moisture, and then heated at 80°C for various times. Moisture in the virus-spiked samples was monitored by Near InfraRed (NIR) spectrometry, whose values had been correlated to loss on drying (LOD) moisture values.^[37] After converting the NIR values to LOD moisture levels, the freeze dried/dry heated materials were reconstituted and titrated for residual virus.

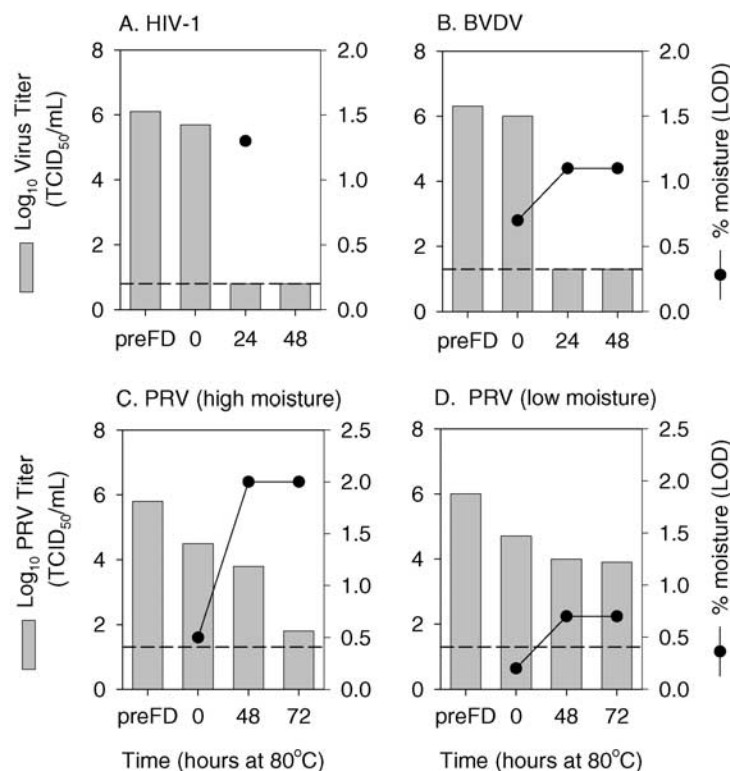


Fig. 8 Kinetics of enveloped virus inactivation during 80°C heat treatment of freeze dried Factor VIII (FVIII) concentrate. (A) HIV-1, (B) BVDV, (C) PRV (> 0.8% moisture) and (D) PRV (< 0.8% moisture). Virus was spiked into FVIII concentrate and a sample was immediately removed for virus titration (pre-FD). Vials were aseptically filled with virus-spiked FVIII concentrate and then freeze dried. After freeze drying, the vials were closed, using stoppers containing different levels of moisture, and sealed before heating at 80°C for up to 72 hr. Near infrared (NIR) readings were taken on vials after freeze drying (0 hr) and at various times after heating at 80°C (24 hr, 48 hr, or 72 hr). After converting the NIR readings to loss on drying (LOD) moisture values, the product in each vial was reconstituted and assayed for residual virus (gray boxes = log₁₀ virus titer, closed circles = % moisture, and dashed line = virus detection limit).

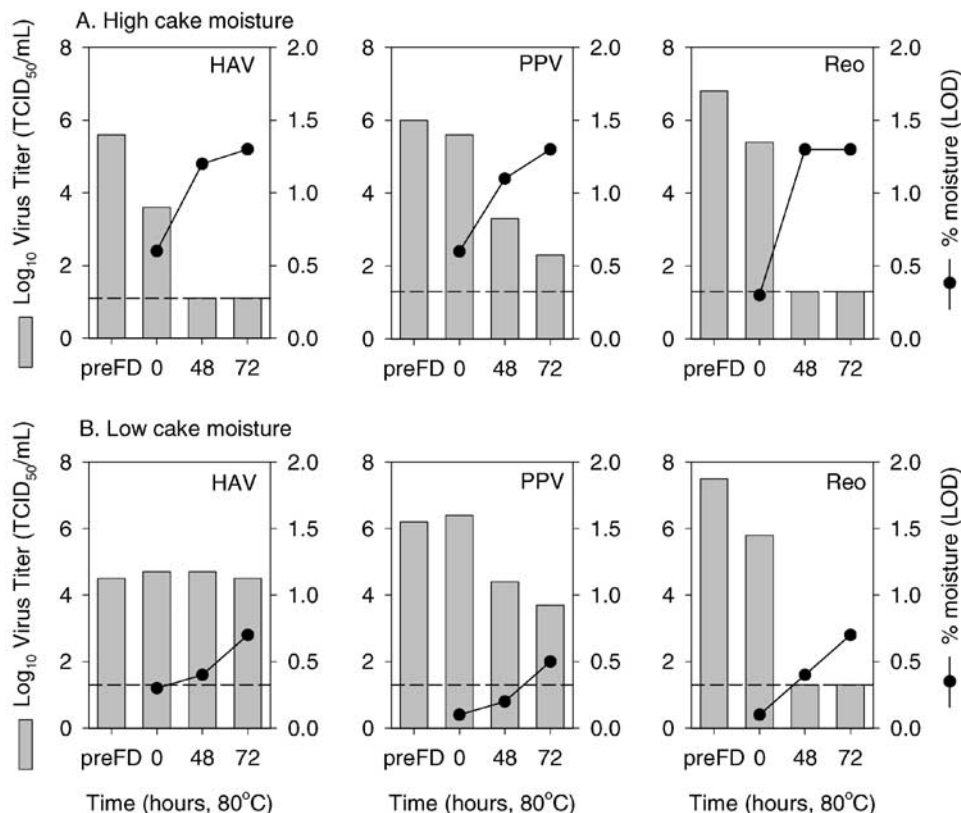


Fig. 9 Kinetics of nonenveloped virus inactivation during 80°C heat treatment of freeze dried Factor VIII (FVIII) concentrate (A) in the presence of high cake moisture (> 0.8% moisture) or (B) in the presence of low cake moisture (< 0.8% moisture). Methods were as described in Fig. 8 (gray boxes = log₁₀ virus titer, closed circles = % moisture, and dashed line = virus detection limit).

As shown in Figs. 8 and 9, the moisture content in freeze dried material was not constant after freeze drying. Autoclaved rubber stoppers hold measurable amounts of water that are transferred to the freeze dried

product^[35,36] and heating at 80°C accelerates this transfer.

Virus reduction data are presented in Figs. 8 and 9 and Table 5. Little or no HIV, BVDV and PPV reduction

Table 5 Summary of virus reduction during 80°C heat treatment of freeze dried AHF

	HIV-1		BVDV		PRV	
	Log ₁₀ virus reduction ^a	Moisture ^b	Log ₁₀ virus reduction	Moisture	Log ₁₀ virus reduction	Moisture
Enveloped viruses						
Low moisture	≥ 5.3	1.3 (24 hr)	≥ 5.7	1.1 (24 hr)	2.1	0.7 (72 hr)
High moisture	■	■	■	■	4.0	2.0 (72 hr)
	HAV		PPV		Reo	
	Log ₁₀ virus reduction	Moisture	Log ₁₀ Virus Reduction	Moisture	Log ₁₀ Virus Reduction	Moisture
Nonenveloped viruses						
Low moisture	0.1	0.7 (72 hr)	2.5	0.5 (72 hr)	≥ 6.2	0.4 (72 hr)
High moisture	≥ 4.5	1.3 (72 hr)	3.7	1.3 (72 hr)	≥ 5.5	1.3 (72 hr)

Gray boxes = not tested and methods were as described in Fig. 8.

^aLog₁₀ virus reduction = Log₁₀ total virus_{pre-FD} - Log₁₀ total virus_{post-80°C heat}.

^bFreeze dried product was measured by NIR spectrometry. The NIR values were then converted to LOD moisture values using a NIR/LOD calibration curve.

occurred during the freeze drying process, but PRV, Reo, and HAV titers, before (Pre-FD) and after freeze drying (0 hr), decreased 1–2 \log_{10} . Regardless of moisture content, HIV, and BVDV were inactivated to below detection after heating 80°C, 24 hr, and Reo was completely inactivated after 48 hr. Inactivation of HAV, PRV, and PPV was dependent on the moisture levels in the cake. When the moisture in the system was 0.8% or greater, virus inactivation was approximately 4 \log_{10} . However, when less than 0.8% moisture was present, virus reduction after heating was $\leq 2.5 \log_{10}$.

These data show the freeze dry/dry heat treatment of the Koate®-DVI process can effectively inactivate significant levels of enveloped and nonenveloped viruses if a minimum threshold level of moisture (0.8%) is attained. Increased moisture levels during dry heat treatment also destroys FVIII activity and at 1.6% moisture, the upper limit for moisture at this step, specific activity decreases an average of 11%.^[38] To compensate for the loss in potency, the bulk is made more concentrated to assure FVIII activity at 9 IU – 22 IU activity per mg protein.

Other investigators have also shown parvovirus inactivation by the terminal freeze dry/dry heat (80°C, 72 hr) step. Canine parvovirus, spiked into FVIII concentrate, Replenate®, was inactivated by 3.0 \log_{10} ,

while 4.7 \log_{10} bovine parvovirus spiked into FVIII concentrate, Liberate®,^[39] was inactivated after heating lyophilized product for 72 hr, 80°C. Unfortunately, in these studies, product loss after treatment was as high as 30%.

Although viruses from different animal species (e.g., porcine, canine, and bovine) were used to test the terminal freeze dry/dry heat treatment of the Koate®-DVI, Replenate®, and Liberate® processes, the differences in parvovirus reduction and product recovery were probably due to the different formulations and freeze drying cycles of all three FVIII products. For freeze dried products, formulation and the freeze drying process are interrelated. Without the appropriate buffers and stabilizers or optimal temperatures and cycle times, many protein preparations would be denatured by the physical stresses associated with the freeze dry/dry heat treatment.

The impact of formulation on virus reduction and product recovery (e.g., potency) was evaluated during terminal freeze dry/dry heat treatment studies with an unlicensed Fraction I derived product, that will be designated “Protein G.” As shown in Fig. 10, the optimum formulation, that achieved $\geq 4 \log_{10}$ PPV inactivation and 80% product recovery, was one that contained 2% albumin, no NaCl, and $\leq 0.3\%$ moisture by the Karl Fischer coulometric method. In contrast, freeze

Freeze dried Protein G formulated in buffer							
No Albumin				Albumin			
No NaCl		NaCl		No NaCl		NaCl	
low moist ¹	high moist ²	low moist	high moist	low moist	high moist	low moist	high moist
Log PPV reduction ³							
≥ 4.0		4.4		3.3		4.3	
				≥ 4.0		ND	
% potency recovered ⁴							
51%		5%		35%		< 2%	
				80%		ND	
				70%		30%	

¹ Low moisture $\leq 0.3\%$ (Karl Fischer coulometric method) after 80°C heat, dry heat, 72 hours

² High moisture $\geq 1.5\%$ moisture after 80°C dry heat, 72 hours

³ Log PPV reduction after 80°C dry heat, 72 hours = Total virus_{pre-FD} – Total virus_{post 80°C heat}

⁴ % potency recovered after 80°C dry heat, 72 hours =

$$[(\text{Potency}_{\text{pre-FD}} - \text{Potency}_{\text{post 80°C heat}}) \div \text{Potency}_{\text{pre-FD}}] \times 100$$

ND = Not Done

Fig. 10 Effect of formulation on virus reduction and product recovery. Protein G was formulated in buffer containing albumin (0% or 2%) and NaCl (0 mM or 150 mM). Virus was added and a sample was immediately removed for virus titration (Pre-FD). The virus-spiked material was aseptically filled into vials, freeze dried, and then heated at 80°C, 72 hr. Mock-spiked Protein G was processed like the virus-spiked samples but was used to measure product recovery (potency).



dry/dry heat treatment of product formulated with no albumin, 150mM NaCl and low moisture resulted in approximately 3 log₁₀ PPV reduction and 35% product recovery. Thus, minor changes in formulations such as the addition of 2% albumin may impact virus reduction and product recovery during a freeze dry/dry heat treatment.

In summary, the large number and complexity of the variables to control when implementing a terminal freeze dry/dry heat treatment makes validation very difficult and increases the probability for error. Since the reliability of the results from virus clearance studies is dependent on the appropriateness of the models used in the studies, it is essential that the freeze dry/dry heated material and the experimental conditions used at small scale be representative of production scale. During the studies, temperature and pressure must be controlled during all stages of the freeze drying cycle and uniform temperatures must be distributed throughout the oven used for the dry heat step. Process parameters such as the amount of moisture retained by rubber stoppers should be examined, as they influence the rate and extent of viral inactivation and the recovery of biological activity.

Perhaps more difficult to duplicate or control during small scale virus clearance studies, are the precise composition and the molecular structure of freeze dried materials. The formulation studies with Protein G indicated virus reduction and product recovery were influenced by the presence of albumin at a prefreeze drying volumetric concentration of 2%. All proteins have unique physico-chemical properties and stabilization requirements, so formulation and freeze drying cycle parameters must be customized for each new protein drug. For terminal viral inactivation steps, it is equally important to develop a formulation and freeze dry cycle that will not stabilize virus.

CONCLUSION

The direct testing for potential viral contamination in a finished product is not considered sensitive or accurate enough to assure the absence of infectious virus because many of the viruses known to contaminate plasma (e.g., HCV, HBV, and parvovirus B19) cannot be easily grown or assayed in cell culture. Testing for the presence of viral nucleic acids in final product is sensitive but cannot distinguish infectious particles from noninfectious particles. Thus, assuring the safety of plasma-derived products with respect to virus transmission is an indirect combinatorial approach: screening and selection of the source material, validation of the production processes for inactivation/removal of virus infectivity, and process control by strict adherence to cGMP.

Screening and selection of the source plasma will only avoid contamination by known pathogens. The protein

purification steps and specific virus reduction methods used in production processes, however, will inactivate and/or remove both known and unknown viruses. Terminal virus inactivation treatments are applied to product in final container and must balance virus inactivation with any modifications to protein immunogenicity, activity, and yield. While many upstream virus inactivation steps rely on chemical methods that involve the addition and subsequent removal of toxic agents (e.g., solvent/detergent), physical methods for virus inactivation, such as pH and heat, are used for terminal steps.

Virus validation studies assess the virus clearance capacity of a process and the evaluation and interpretation of virus clearance data from terminal inactivation and upstream process steps are similar. Limitations in the design and execution of virus validation studies may lead to an incorrect estimate of the ability of a process to inactivate/remove virus infectivity. The three terminal inactivation treatments discussed here illustrate the importance of rigorously controlling virus validation studies.

1. During the low pH incubation step of IgG, minor changes in pH and temperature, as well as in upstream processing steps, significantly impacted the kinetics and extent of virus inactivation. Thus, the process variables evaluated during virus clearance studies must be carefully selected and defined since they may alter the efficacy of the step for virus reduction.
2. The results from the albumin studies demonstrated that virus selection (e.g., HM175/18f and HM175/24a) and preparation (e.g., chloroform extracted and untreated virus) were just as important as the process conditions (e.g., protein concentration and temperature) in determining the outcome of virus reduction during pasteurization. Since laboratory adapted viruses and naturally occurring viruses may differ in their sensitivity to physico-chemical treatments, the results observed with model viruses must be carefully interpreted before they can be extrapolated to relevant viruses of concern.
3. Certain process steps may be easier to model than others and duplicating the freeze dry/dry heat step at small scale is very difficult. The conclusions drawn from virus clearance studies are reliable only when the appropriateness of the small-scale models can be demonstrated. During the freeze dry/dry heat step of Koate[®]-DVI, virus reduction was dependent on moisture levels so even the formulation of stoppers, which could impact the absorption of water during autoclaving and its release to freeze dried material, must be considered and tested.



4. Formulation studies with Protein G suggested that the addition of virus could affect the composition and freeze dry/dry heat treatment of spiked product intermediates. As specified in guidance documents for virus clearance studies,^[1,2] the amount of virus spiked into starting test materials should be as high as possible so that the maximum virus reduction capacity of a production step can be determined. In many clearance studies, the volume of the virus added may be as high as 10%. Since virus reduction and product recovery, during freeze dry/dry heat treatment of Protein G were influenced by the presence 2% albumin, high amounts of virus may alter the composition and structure of freeze dried materials. However, until the analytical tools to study the basic behavior of all proteins during freeze drying and dry heat treatment become available, the stability of viruses and their interactions with plasma proteins will not be fully understood.

The terminal inactivation studies discussed here also demonstrate there is no “magic bullet” to inactivate all viruses. Many inactivation steps are effective against some but not all virus families. Thus, production processes should contain other upstream orthogonal steps to assure broad and effective virus reduction. Regardless of the virus reduction capacity of a manufacturing process, however, absence of virus in plasma-derived products can never be guaranteed because the amount of contaminating virus in source material is generally not known and the presence and/or appearance of new unknown viruses cannot be excluded.

Manufacturing processes have evolved dramatically over the last few years. In the late 1980s, 76% of hemophiliacs were HCV positive,^[40] and between 1979 and 1985, approximately 50% of hemophiliacs had acquired HIV from plasma-derived FVIII.^[41] Since then, however, most U.S.-licensed plasma derivatives have not transmitted HBV, HCV, or HIV as a result of improvements in donor screening and test methods, and the inclusion of effective upstream virus-reduction and terminal virus-inactivation steps in manufacturing processes.^[18] Residual risks of virus transmission from plasma-derived products are now largely associated with nonenveloped viruses.^[42] Thus, the need for additional terminal or upstream virus inactivation/removal steps still exists, but the current challenge is to develop cost effective methods against physico-chemically resistant nonenveloped viruses, such as human parvovirus B19.

ACKNOWLEDGMENTS

The authors thank M. Fournel, J. Wang, and S. F. Chao for their comments on the manuscript. They are also grateful

to K. Dawson, L. Franks, E. Landreth, S. Nicosia, and H. Renfrow for expert technical assistance.

REFERENCES

1. Committee for Proprietary Medicinal Products. Ad Hoc Working Party on Biotechnology/Pharmacy and Working Party on Safety Medicines, EEC Regulatory Document, Note for Guidance: Validation of Virus Removal and Inactivation Procedures. *Biologicals* **1991**, *19*, 247–251.
2. Federal Health Agency and Paul Ehrlich Institute Federal Agency for Sera and Vaccines. Announcement on the Marketing Authorization of Medicinal Products: Requirements of Validation Studies for Demonstrating the Virus Safety of Medicinal Products Derived from Human Blood or Plasma. Federal Gazette No. 84, 4 May 1994.
3. Alonso, W.R.; Trukawinski, S.; Savage, M.; Tenold, R.; Hammond, D. Viral Inactivation of Intramuscular Immune Serum Globulins. *Biologicals* **2000**, *28*, 5–15.
4. Savage, M.; Torres, J.; Franks, L.; Masecar, B.; Hotta, J. Determination of Adequate Moisture Content for Efficient Dry-Heat Viral Inactivation in Lyophilized Factor VIII by Loss on Drying and by Near Infrared Spectroscopy. *Biologicals* **1998**, *26*, 119–124.
5. Finlayson, J.S. Therapeutic Plasma Fractions and Plasma Fractionation (Prologue). *Semin. Thromb. Hemost.* **1979**, *6*, 1–11.
6. Centers for Disease Control. Outbreak of Hepatitis C Associated with Intravenous Immunoglobulin Administration—United States, October 1993–June 1994. *MMWR* **1994**, *43*, 505–508.
7. Barandun, S.; Kistler, P.; Jeunet, F.; Isliker, H. Intravenous Administration of Human γ -Globulin. *Vox Sang.* **1962**, *7*, 157–174.
8. Friedli, H.R. Methodology and Safety Considerations in the Production of an Intravenous Immunoglobulin Preparation. *Pharmacotherapy* **1987**, *7* (Suppl), 36S–40S.
9. Immune Globulin Intravenous (Human), 10%, Gamimune® N, 10%. Bayer Corporation Package Insert 14-7648-002, May 1999.
10. Hamalainen, E.; Suomela, H.; Ukkonen, P. Virus Inactivation During Intravenous Immunoglobulin Production. *Vox Sang.* **1992**, *63*, 6–11.
11. Reid, K.G.; Cuthbertson, B.; Jones, A.D.L.; McIntosh, R.V. Potential Contribution of Mild Pepsin Treatment at pH 4 to the Viral Safety of Human Immunoglobulin Products. *Vox Sang.* **1988**, *55*, 75–80.
12. Bos, O.J.M.; Sunye, D.G.J.; Nieuweboer, C.E.F.; van Engelenburg, F.A.C.; Schuitemaker, H.; Over, J. Virus Validation of pH 4-Treated Human Immunoglobulin Products Produced by the Cohn Fractionation Process. *Biologicals* **1998**, *26*, 267–276.
13. Burnout, T.; Radosevich, M. Reducing the Risk of Infection from Plasma Products: Specific Preventative Strategies. *Blood Rev.* **2000**, *14*, 94–110.



14. Schimpf, K.; Mannucci, P.M.; Kreuz, W.; Brackmann, H.H.; Auerswald, G.; Ciavarella, N.; Moessler, J.; De Rosa, V.; Kraus, M.D.; Brueckman, C.H.; Mancuso, G.; Mittler, V.; Haschke, F.; Morfini, M. Absence of Hepatitis After Treatment with a Pasteurized Factor VIII Concentrate in Patients with Hemophilia and No Previous Transfusions. *N. Engl. J. Med.* **1987**, *316*, 918–922.
15. Barrowcliffe, T.W. Viral Inactivation Vs Biological Activity. *Dev. Biol. Stand.* **1993**, *81*, 125–135.
16. Gellis, S.S.; Neefe, J.R.; Stoks, J.; Strong, L.E.; Janeway, C.A.; Scatchard, G. Chemical, Clinical and Immunological Studies on Products of Human Plasma Fractionation: Inactivation of Virus of Homologous Serum Albumin by Means of Heat. *J. Clin. Invest.* **1948**, *27*, 239–244.
17. Grandgeorge, M.; Veron, J.L. Viral Validation of the Manufacturing Process of High Purity Albumin from Placentas. *Dev. Biol. Stand.* **1993**, *81*, 237–244.
18. Tabor, E. The Epidemiology of Virus Transmission by Plasma Derivatives: Clinical Studies Verifying the Lack of Transmission of Hepatitis B and C Viruses and HIV Type 1. *Transfusion* **1999**, *39*, 1160–1168.
19. Pattison, C.P.; Klein, C.A.; Leger, R.T.; Maynard, J.E.; Berquist, K.R.; Hoofnagle, J.H.; Barker, L.F.; Bryan, J.A.; Smith, D.E. An Outbreak of Type B Hepatitis Associated with Transfusion of Plasma Protein Fraction. *Am. J. Epidemiol.* **1976**, *103*, 398–407.
20. Albumin (Human) 25%, USP, Plasbumin®-25. Bayer Corporation Package Insert 14-7648-001, October 1998.
21. Lemon, S.M.; Murphy, P.C.; Shields, P.A.; Ping, L.-H.; Feinstone, S.M.; Cromeans, T.; Jansen, R.W. Antigenic and Genetic Variation in Cytopathogenic Variants Arising During Persistent Infection: Evidence of Genetic Recombination. *J. Virol.* **1991**, *65*, 2056–2065.
22. Lemon, S.M.; Binn, L.N. Incomplete Neutralization of Hepatitis A Virus In Vitro Due to Lipid-Associated Virions. *J. Gen. Virol.* **1985**, *66*, 2501–2505.
23. Schlegel, A.; Immelmann, A.; Kempf, C. Virus Inactivation of Plasma-Derived Proteins by Pasteurization in the Presence of Guanidine Hydrochloride. *Transfusion* **2001**, *41*, 382–389.
24. Arrighi, S.; Rossi, R.; Borri, M.G.; Lesnikov, V.; Lesnikova, M.; Franco, E.; Divizia, M.; De Santis, M.E.; Bucci, E. "In Vitro" and in Animal Studies on a Double Virus-Inactivated Factor VIII Concentrate. *Thromb. Hemost.* **1995**, *74*, 868–873.
25. Hart, H.F.; Hart, W.G.; Crossley, J.; Perrie, A.M.; Wood, D.J.; John, A.; McOmish, F. Effect of Terminal (dry) Heat Treatment on Non-enveloped Viruses in Coagulation Factor Concentrates. *Vox Sang.* **1994**, *67*, 345–350.
26. Horowitz, B.; Wiece, M.E.; Lippin, A.; Vandersande, J.; Stryker, M.H. Inactivation of Viruses in Labile Blood Derivatives II Physical Methods. *Transfusion* **1985**, *25*, 523–527.
27. Hollinger, F.B.; Dolana, G.; Thomas, W.; Gyorkey, F. Reduction in Risk of Hepatitis Transmission by Heat Treatment of a Human Factor VIII Concentrate. *J. Infect. Dis.* **1984**, *150*, 250–262.
28. National Hemophilia Foundation Medical and Scientific Advisory Council. *Recommendations Concerning AIDS and Therapy of Hemophilia (Revised October 13, 1984)*; National Hemophilia Foundation: New York, 1984.
29. Centers for Disease Control. Update: Acquired Immuno-deficiency Syndrome (AIDS) in Persons with Hemophilia. *MMWR* **1985**, *33*, 589–591.
30. Piskiewicz, D.; Thomas, W.; Lieu, M.Y.; Tse, D.; Sarno, L. Virus Inactivation by Heat Treatment of Lyophilized Coagulation Factor Concentrates. In *Virus Inactivation in Plasma Products*; Morgenthaler, J.J., Ed.; Curr. Stud. Hematol. Blood Transfus. Basel, Karger **1989**, *56*, 44–54.
31. Rouzioux, C.; Chamaret, S.; Montagnier, L.; Carnelli, V.; Rolland, G.; Mannucci, P.M. Absence of Antibodies to AIDS Virus in Hemophiliacs Treated with Heat-Treated Factor VIII Concentrate. *Lancet* **1985**, *1*, 271–272.
32. Colombo, M.; Mannucci, P.M.; Carnelli, V.; Savidge, G.F.; Gazengel, C.; Schimpf, K. Transmission of Non-A, Non-B Hepatitis by Heat-Treated Factor VIII Concentrate. *Lancet* **1985**, *2*, 1–4.
33. Santagostino, E.; Mannucci, P.M.; Gringeri, A.; Azzi, A.; Morfini, M.; Musso, R.; Santoro, R.; Schiavoni, M. Transmission of Parovirus B19 by Coagulation Factor Concentrates Exposed to 100°C Heat After Lyophilization. *Transfusion* **1997**, *37*, 517–522.
34. Antihemophilic Factor (Human), Koate®-DVI. Bayer Corporation Package Insert 14-7665-001, October 2000.
35. De Grazio, F.; Flynn, K. Lyophilization Closures for Protein Based Drugs. *J. Parenter. Sci. Technol.* **1992**, *46*, 54–61.
36. Pikal, M.J.; Shah, S. Moisture Transfer from Stopper to Product and Resulting Stability Implications. *Dev. Biol. Stand.* **1992**, *74*, 165–177.
37. Jones, J.A.; Last, I.R.; MacDonald, B.F.; Prebble, K.A. Development and Transferability of Near-Infrared Methods for Determination of Moisture in a Freeze-Dried Injection Product. *J. Pharm. Biomed. Anal.* **1993**, *11*, 1227–1231.
38. Koate Characterization Report, Bayer Corporation Process Characterization Report, ETR-0799, 1999.
39. Roberts, P.L.; Hart, H. Comparison of the Inactivation of Canine and Bovine Parvovirus by Freeze-Drying and Dry-Heat Treatment in Two High Purity Factor VIII Concentrates. *Biologicals* **2000**, *28*, 185–188.
40. Brettler, D.B.; Alter, H.J.; Dienstag, J.L.; Forsberg, A.D.; Levine, P.H. Prevalence of Hepatitis C Virus Antibody in a Cohort of Hemophilia Patients. *Blood* **1990**, *76*, 254–256.
41. Evatt, B. AIDS and Hemophilia—Current Issues. *Thromb. Hemost.* **1995**, *74*, 36–39.
42. Yee, T.T.; Lee, C.A.; Pasi, K.J. Life-threatening Human Parvovirus B19 Infection in Immunocompetent Hemophilia. *Lancet* **1995**, *345*, 794–795.

Water for Pharmaceuticals

Leonid Shnayder

Aker Kvaerner Pharmaceuticals, Bridgewater, New Jersey, U.S.A.

INTRODUCTION

Water is one of the key utilities in most pharmaceutical facilities. It is used as a solvent, product ingredient, cleaning agent, and for many other applications. Some of those applications require water of higher purity than typically found in municipal water supplies. Therefore companies install private water purification systems that become a critical part of their utility infrastructure.

Why does the potable water from municipal supply require further treatment for pharmaceutical manufacturing? To understand this, let us look into the two main water sources for municipal plants: groundwater aquifers (well water) and surface supplies (rivers and lakes). As rainwater falls onto the Earth's surface, it picks up and dissolves atmospheric gases (oxygen and carbon dioxide), industrial emissions such as nitric and sulfuric oxides, as well as bacteria and dust particles. As it passes through the ground into the aquifers, the water is filtered and becomes quite clean, free of most particles and organic contaminants. However, it now dissolves minerals contained in the limestone and other geological materials. That adds calcium, magnesium, sodium, iron, chlorides, sulfates, and other ions to the ground water. The surface waters found in the lakes and rivers collect organic materials and particles from vegetation and ground runoff.

Municipal water plants provide limited treatment, mostly intended to make the water safe to drink. A lot of contaminants, such as salts, dissolved gases, and organic materials contained in natural sources, remain in the municipal water supply. In addition, chlorine or other disinfectants are often added as part of the treatment process to control microbial contamination. For many critical applications required in pharmaceutical plants, such water quality is not sufficient, and further treatment is necessary.

GRADES OF WATER USED IN THE PHARMACEUTICAL FACILITIES

The water purity required varies widely with the application. For example, water used to feed the boilers producing plant steam only needs to be softened (to avoid scale buildup) and deaerated (to avoid corrosion in the

steam system). On the other hand, water used as an ingredient for an injectable drug has to meet much more stringent quality requirements. The US Pharmacopeia (USP) contains specifications for several grades of water used in the preparation of medicinal products. The two grades most often used in the pharmaceutical plants are USP Purified Water and Water For Injection (WFI). As the name implies, WFI is used for the preparation of injectable drugs, whereas USP Purified Water can be used in the manufacturing of tablets, capsules, creams, lotions, etc. These types of water are called "compendial" because their quality is specified in an official nationally recognized standard such as USP. In addition, many companies use various noncompendial water systems designed for specific needs. While designing the facility, it is important to decide on the proper grades of water to use in various applications. Compendial waters are typically very expensive, not only because of required treatment steps, but also because of extensive validation and testing requirements. Therefore it is prudent to limit their use to processes where such water becomes an ingredient of the pharmaceutical product, is in direct contact with such product, or is used for the final rinse of critical process equipment. For most other applications, various noncompendial grades of water (potable, softened, deionized, etc.) can be successfully used without any regulatory conflicts. It is not uncommon to install a water system that produces water that could easily meet the USP specification yet is not designated as USP Purified Water to avoid unnecessary expenses for regulatory compliance.

The current edition of USP^[1] establishes the following requirements for USP Purified Water:

- Purified Water is prepared from water complying with the federal regulations for drinking water.
- Purified Water contains no added substance.
- Purified Water is obtained by a suitable process.
- Conductivity does not exceed set level (1.3 $\mu\text{S}/\text{cm}$ at 25°C).
- Total organic carbon (TOC) does not exceed set level (500 ppb).

The requirements for WFI are as follows:

- Meets all the requirements for "Purified Water"



- Is obtained by distillation or reverse osmosis (RO)
- Contains not more than 0.25 USP endotoxin units per milliliter
- Is prepared using suitable means to minimize microbial growth.

The USP does not specify allowable concentrations of microorganisms in the official monograph, but recommends action limits of 100 colony-forming units per ml (CFU/ml) for Purified Water and 0.1 CFU/ml for WFI.

Putting aside the items related to the source water and treatment procedures, we can summarize the requirements for Purified Water itself as follows:

1. Low conductivity (high resistivity). This test is intended to show that water contains a minimal amount of ions such as calcium, magnesium, sodium, iron, chloride, sulfate, etc. The inherent presence of hydrogen and hydroxide ions determines the theoretical limit of Purified Water conductivity: approximately 0.05 $\mu\text{S}/\text{cm}$ (resistivity 18 $\text{M}\Omega\text{ cm}$) at a pH of 7.0. The practical limits specified in the USP are in the range of 1–5 $\mu\text{S}/\text{cm}$, depending on pH and temperature.
2. Low TOC (less than 500 ppb). Water shall contain a minimal amount of organic compounds. Such compounds are undesirable for two main reasons: they may be toxic, and/or they may serve as sources of nutrition for microorganisms.
3. Low microbial count. Water shall contain a minimal amount of viable microorganisms, including spores.
4. Low endotoxin level (required for WFI only). The term “endotoxin” applies to organic compounds that cause harmful effects when injected in the bloodstream of laboratory animals. Such compounds are lipopolysaccharides produced as a result of microbial growth or microbial destruction.

WATER PURIFICATION SYSTEMS

As naturally occurring water has a variety of contaminants, there are many treatment processes developed to remove those contaminants. A typical pharmaceutical water purification system contains several unit operations designed to remove various components.

Particle Removal

The municipal water treatment plants use a process called “clarification” based on gravity settling to remove a

majority of the large particles suspended in water. However, some of the suspended materials remain in the potable water and need to be removed as part of the pharmaceutical water purification.

The most common process used for this is depth filtration through a bed of sand or similar material charged in a vertical vessel. The incoming water flows from top to bottom. To improve the efficiency of such filters, two or more layers of media with various particle sizes are used. Coarse and less dense material such as anthracite is located at the top of the bed, whereas finer and denser particles of sand are placed at the bottom. Such multimedia filters can remove most particles larger than 10–20 μm . Periodically the filter bed is backwashed by reversing the flow direction (from bottom to top) and by increasing the flow rate. During backwash, the captured particles are removed and sent to drain, whereas heavier particles of the filter bed remain in the vessel and settle back at the end of the cycle.

In smaller water purification systems, we can often find a cartridge-type depth filter used for the particle removal. The cartridge is made of cellulose, cotton, or polymer fibers, which trap dirt particles throughout its volume. Once a cartridge reaches its holding capacity, it is disposed of and replaced by a new one.

Adsorption of Dissolved Organics and Chlorine

Activated carbon filters often found in pharmaceutical water purification systems are used to remove chlorine and many dissolved organic materials found in the incoming potable water. Physically, a carbon filter looks exactly like the multimedia filter: a vertical tank with a valve manifold next to it. The main difference is that the filter bed consists of activated carbon particles with a very high surface area. These can absorb and retain amazing amounts of organic contaminants and chlorine. The major disadvantage of carbon filters is that they may become a breeding ground for bacteria thriving in an environment without chlorine (which is typically adsorbed in the upper part of the bed) and with plenty of nutrients. Therefore carbon filters are usually periodically sanitized with hot water to contain microbial growth. The bed is also backwashed frequently to remove captured particles. However, activated carbon cannot be economically regenerated in the field; once it has reached its adsorption capacity (typically after 1–5 years), the bed is replaced with fresh carbon. It is recommended to use acid-washed activated carbon for water purification because the minerals resident in the charcoal may lead to leaching of the metal oxides, increasing the water hardness.^[2]

Ion Exchange

An ion exchange process is based on the ability of certain synthetic resins to selectively adsorb either cations or anions, and to release (exchange) other ions based on their relative activity. The water passes through a bed of resin beads in a vertical vessel (again, looking very similar to a multimedia filter), where such exchange takes place. Once all the active sites on the resin bed are filled with the new ions, the resin needs to be regenerated by passing an appropriate chemical solution through the bed.

The most common type of ion exchange system is a water softener. Its main function is to remove calcium and magnesium ions from hard water. These are the components that cause the formation of the hard-to-remove scale in tea kettles; they have a similar effect on boilers, distillation stills, and membrane filtration equipment. The polystyrene resin has a greater affinity for multivalent ions such as calcium and magnesium than for monovalent sodium. When the water passes through the bed, calcium and magnesium ions are captured and exchanged for the sodium adsorbed on the fresh resin. Once the resin is exhausted, it is regenerated by a 10% solution of sodium chloride. The high concentration of sodium ions in such a solution shifts the equilibrium of the adsorption-desorption process, resulting in the displacement of captured multivalent ions with sodium ions. The regeneration cycle can be repeated many times, making softener resin last for years. Softening is a very simple and inexpensive process requiring only salt for resin regeneration, and it is widely used not only in high-purity water systems but also in many utility water systems.

A similar process also takes place in ion exchange deionizers, designed to remove practically all foreign ions from water. Deionizers use two types of resin: one exchanges positively charged ions (cations) such as sodium, calcium, and magnesium with hydrogen ions (H^+); another exchanges negative ions (anions) such as chloride, sulfate, and bicarbonate with hydroxide (OH^-). Although the ion exchange process can produce very-

high-quality water, its major drawback is the type of chemicals required for resin regeneration: cation resins are regenerated by strong acid, and anion resins are regenerated by strong base solutions. Another drawback important for the pharmaceutical water systems is the fact that ion exchange resins can become a breeding ground for microorganisms. For smaller systems, it is common to use service-type ion exchange units, where the vessel with exhausted resin is taken offsite for regeneration, whereas larger systems have to be regenerated on site, resulting in the need to handle concentrated acids and bases and to neutralize waste solutions. A relatively new process gaining popularity in the field is the continuous deionization system, which represents a combination of ion exchange and membrane separation technologies. It uses electrical current to continuously regenerate the ion exchange resin simultaneously with the water treatment process, eliminating the need to handle strong chemicals.

Membrane Filtration

Unlike depth filters where particles are captured throughout the filter volume, membrane filters are surface-type filters: they stop particles larger than the pore size right at the upstream surface of the polymeric membrane. Naturally, such filters have very limited holding capacity: once most of the surface holes are plugged with a single particle each, the filtration process stops. To overcome this inherent limitation, engineers invented a process called “crossflow” or “tangential flow” filtration. In such a system (Fig. 1), fluid moves at high velocity parallel to the membrane, with only a small fraction of the feed flow penetrating through it because of a pressure differential. The particles or molecules that could not penetrate through the membrane are swept by the parallel flow and exit the system in a stream called “concentrate” because it contains a higher concentration of particles than the feed stream. The stream penetrating through the membrane is called “permeate.” Because of the “sweeping” action of the crossflow stream, membrane holes do

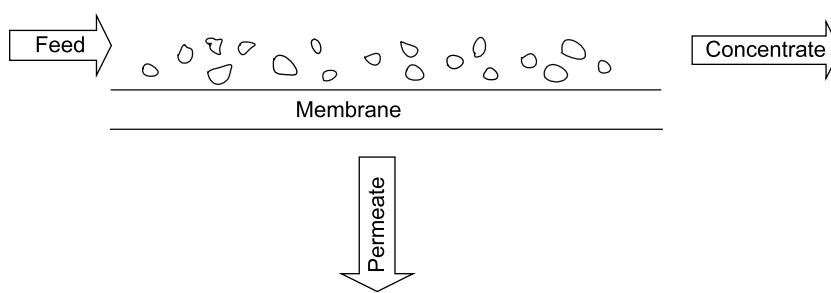


Fig. 1 Tangential flow filtration.



not plug up and such systems can function for a long time between cleanings. Depending on the pore size, membrane filtration processes are called microfiltration, ultrafiltration, and reverse osmosis (RO).

Microfilters use membranes with pores in the 0.1–1 μm range. They can filter out particles of dust, activated carbon, and ion exchange resin fines, and most microorganisms. Microfilters require low differential pressures (5–20 psi) and are available both as normal flow (“dead end”) and crossflow configurations. In pharmaceutical water purification systems, they are often used as disposable cartridge filters after activated carbon filters, softeners, and ion exchange beds.

Ultrafilters use membranes that reject not only solid particles but also dissolved matter with high molecular weight. The “molecular weight cutoff” point of such membranes varies in the range of 10,000–100,000 Da. They can remove bacteria, endotoxins, colloidal contaminants, and large organic molecules. Ultrafilters typically require differential pressure between 10 and 50 psi, and they are also available in both normal and crossflow configurations. In water purification systems, ultrafilters are sometimes used as a polishing step after ion exchange units.

Reverse osmosis is a filtration process using a membrane with the pore size in the range of 1–10 Å. Such membranes reject not only most organic compounds, bacteria, and viruses, but also 90–99% of all ions. Because of the extremely small pore size, RO systems operate at high differential pressures—typically from 100 to 500 psi. Practically all commercial RO systems are

designed in the crossflow configuration. The most common physical membrane configuration for RO systems uses spiral-wound elements (Fig. 2) placed into horizontal pressure vessels (essentially pipes). In the production of the pharmaceutical Purified Water, it is very common to use double-pass RO systems with two filtration stages connected in series. Such systems can easily produce water meeting the requirements for USP Purified Water and even WFI.^[3] However, European regulations do not allow RO to be used as a final treatment step for the production of WFI. This also affects U.S. facilities as the common practice in the United States is to build facilities compliant with both U.S. and European regulations. In the last 10–15 years, RO has become the most common way to produce pharmaceutical Purified Water—either as a final treatment step, or as pretreatment for the distillation stills.

Distillation

Distillation is a process that involves the evaporation of water and then the condensation of the resulting steam. Most dissolved contaminants do not evaporate at the temperature of boiling water, and therefore do not pass into the steam and condensate streams. To improve energy efficiency, distillation is usually conducted in multiple-effect stills designed to recover most of the energy spent on evaporating water. Distillation, while expensive, allows the removal of almost all of the organic and inorganic impurities and achieves very-high-quality water. It is also considered the safest way to avoid microbial

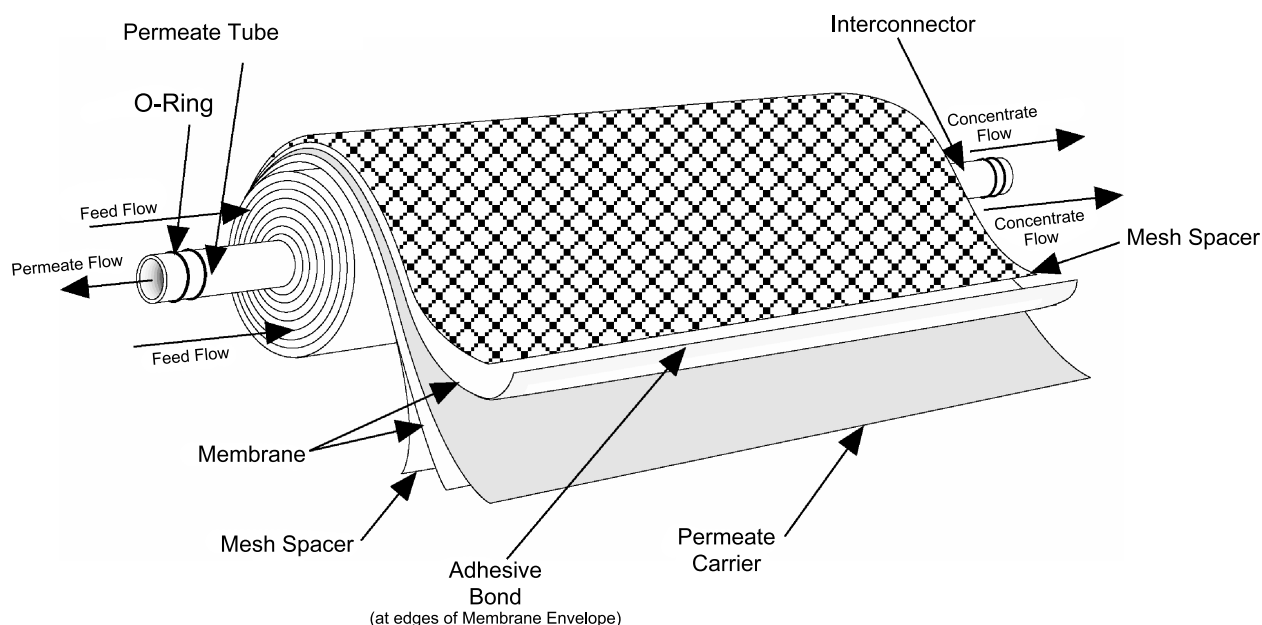


Fig. 2 Spiral-wound membrane element. (Courtesy of Osmonics Inc.)

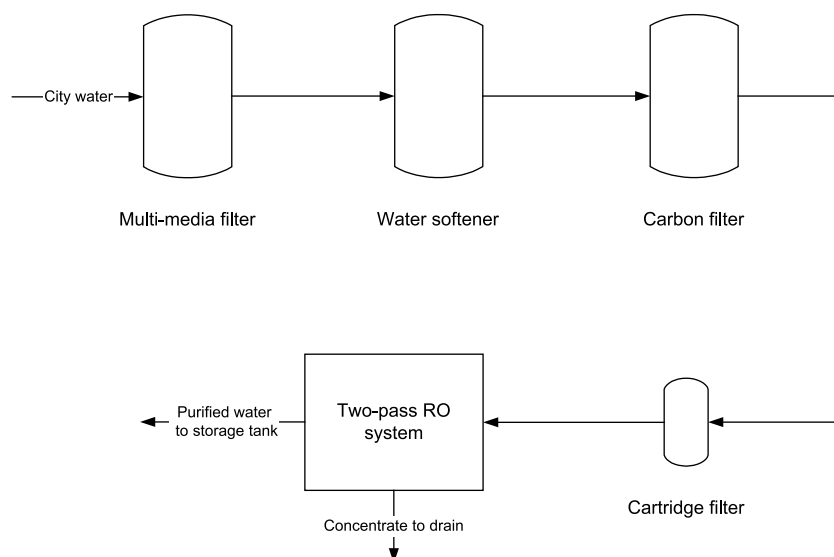


Fig. 3 Pharmaceutical water purification system.

and endotoxin contamination. That is why distillation remains the dominant method of producing WFI. Feed water to the distillation stills is often pretreated with RO and/or other methods to reduce scaling and to increase distillate quality.

One of the possible configurations for the pharmaceutical water purification system is shown in Fig. 3. The first three stages in that system (multimedia filtration, softening, and carbon filtration) are considered pretreatment steps for the RO unit that conducts final purification. The cartridge filter shown before the RO is provided to protect the membranes from the fine carbon particles. There are many variations of the treatment systems: sometimes carbon filter is placed before a softener, sometimes it is eliminated completely, and chlorine removal is accomplished by adding reducing chemicals such as sodium sulfite; often ultraviolet (UV) lights are added to improve microbial control and/or to reduce TOC levels, etc. The systems designed to produce WFI most often include distillation still as a final treatment step.

STORAGE AND DISTRIBUTION SYSTEMS FOR PHARMACEUTICAL WATER

Further discussion assumes that water leaving the purification system and entering the storage tank meets all of the above requirements either for USP Purified Water or WFI. The goal when designing and operating the storage and distribution system is to keep the water at these purity levels, preventing any of the four parameters listed above

from exceeding allowable limits during storage. In particular, the engineers and operators have to ensure that:

- Water is protected against ionic and organic contamination (that would lead to an increase in the conductivity and TOC respectively).
- The system is protected against physical entry of foreign particles and microorganisms.
- Microbial growth is prevented or minimized.

Before moving on to the description of the typical storage and distribution systems and their components, it is beneficial to say a few words about the issue of the microbial growth that presents probably the biggest challenge and causes most concerns among engineers and operating personnel.

Microorganisms are ubiquitous; they can be found in a very wide variety of environments. High-purity water systems are no exception. Although such water contains very few nutrients necessary to sustain life, some types of bacteria developed skills that allow them not only to survive but also to multiply in such waters. They do it for the most part by attaching themselves to the pipe and other system's surfaces and by creating a thin film containing microorganisms, organic and inorganic nutrients (obtained partially from the water and partially from the dead bacteria), and products of bacterial metabolism and decay.^[4-6] Such film, appropriately called biofilm, constitutes a niche environment that is very different from the bulk of the water system's volume, and may be reasonably comfortable for such bacteria. The majority of microorganisms capable of forming biofilms in the



purified water systems are classified as Gram-negative organisms in reference to their quality of not retaining a dye after staining (Cristian Gram was the inventor of the staining method). Such organisms have a tendency to shed lipopolysaccharidic materials from the cell membranes that constitute endotoxins, both during their normal growth and after the cell's destruction. That is why elimination of the Gram-negative microorganisms from the water systems is one of the primary goals. That also explains why test samples taken from the water system soon after sanitization sometimes show higher levels of endotoxins than those observed before: successful sanitization kills all or most of the bacteria in the biofilm, and pieces of their membranes are released in the bulk water stream where they can be detected as endotoxins. Another consequence of bacterial growth in biofilms that must be kept in mind is that results of microbial sampling of the water systems are only indirectly correlated with the actual situation. First, the sample is usually taken from the bulk liquid stream whereas most bacteria reside on the pipe and equipment surfaces. Second, the testing techniques, including the type of growth media used to cultivate the sample before counting colonies, can have a great impact on the test results. Therefore it is no accident that USP does not specify an allowable concentration of microorganisms in the water samples as a pass-fail criteria similar to conductivity or TOC, but only as a helpful indirect indicator that shall be used as an alert or action limit in the efforts to maintain microbial control.

It is important to point out that a small number of microorganisms surviving in the purified water system do not necessarily affect the water quality. Because of the lack of nutrients and other unfavorable conditions, the microbial growth in such waters is very slow. That provides operators with an option to maintain such systems in the state of microbial control by sanitizing them periodically rather than continuously, as long as the frequency is high enough to keep the bacterial counts at acceptable low levels.^[7,8]

Let us consider some of the engineering practices used to achieve the above mentioned goals. For a more detailed description, refer to the ISPE Baseline Pharmaceutical Engineering Guides^[9] recently developed to help industry professionals in designing and operating purified water systems.

Overall System Schematics

One of the typical flow diagrams for the pharmaceutical water storage and distribution system is shown in Fig. 4. The system includes a vertical storage tank, circulation pump, sanitary heat exchanger, and piping designed in the form of complete loop with point-of-use valves located at the minimum distance to the loop. The pump continuously circulates the water from the tank via distribution piping and back to the tank. The water temperature in the system is maintained at 65–85°C to maintain microbial control; any organisms that found their way into the system would be killed or at least prevented from multiplying at such

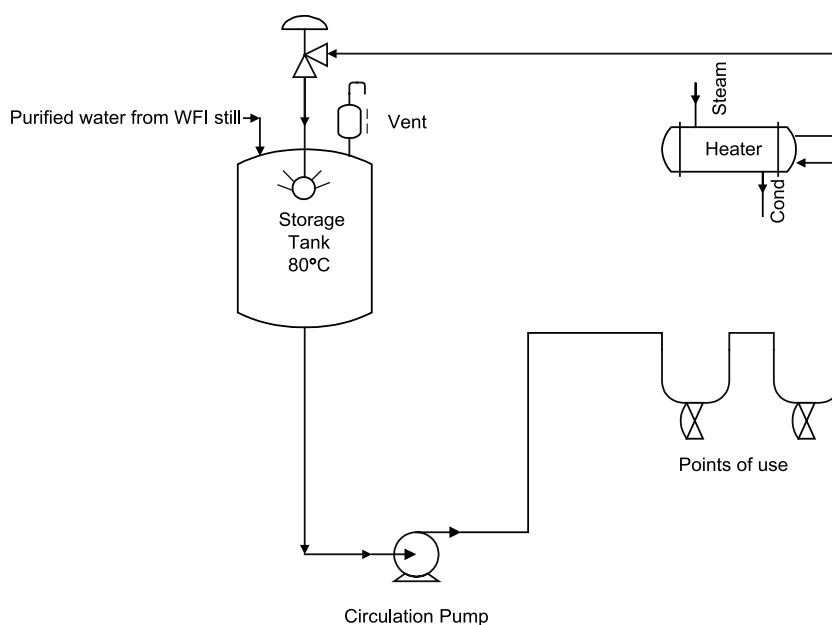


Fig. 4 Purified water storage and distribution system.

temperatures. The heat exchanger is used to compensate for the heat losses occurring in the piping system. The water is returned to the tank via one or more spray balls, thus providing continuous flushing of the tank's top head and upper walls, and keeping them at the high temperature. The entire system (tank, piping, and other components) is insulated for heat conservation. To minimize the "dead legs" at the points of use, it is common to design the loop with U-shaped sections located at each such point, in essence bringing the main header from the ceiling space to a convenient working level where a point-of-use valve would be easily accessible. That adds a lot of elbows and associated pressure drop to the system. In cases where the point-of-use valve is automated, this can sometimes be avoided by locating the valve at the header, away from the actual point of use. In such case, the valve only needs to be accessible for maintenance, but sanitization of the branch line may have to be addressed.

The storage tank is usually equipped with a sterilizing-grade (0.2 μm) hydrophobic vent filter, and sometimes also with a nitrogen blanketing system, to prevent the introduction of particles and organisms during normal

inbreathing. The additional benefits of the nitrogen blanket are the creation of oxygen-free atmosphere in the storage tank's upper space (better for microbial control and corrosion protection), and the avoidance of the slight vacuum conditions during the inbreathing via vent filter. The vent filter on the storage tank is often heat-traced to stay hot, both for microbial control and for prevention of moisture condensation.

The water circulation pump must be of sanitary design and may be equipped either with a single or a double mechanical seal. In the latter case, the purified water from the pump's discharge is used as a seal flush fluid (Fig. 5). If a spare circulation pump is installed for reliability, it is important to ensure that it does not constitute a long "dead leg." This can be avoided either by valving it off and draining it free of water when not in use, or by allowing a small flow of hot water to constantly circulate through the spare pump.

The circulation flow rate is usually selected based not only on the anticipated water usage rates, but also on the requirement to maintain certain minimum velocity (typically 5 ft/sec) during periods of no usage. The minimum

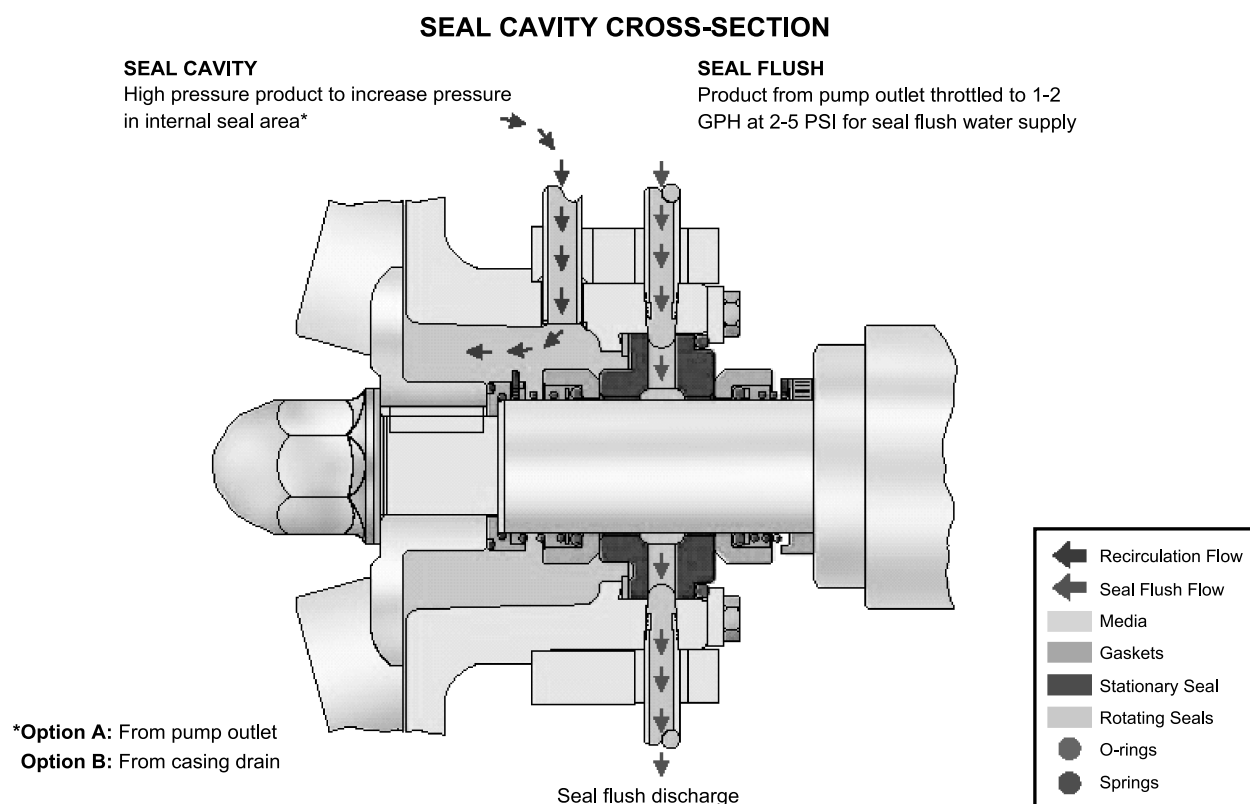


Fig. 5 Double mechanical seal with purified water flush arrangement. (Courtesy of Fristam Pumps Inc.) (Go to www.dekker.com to view this figure in color.)



velocity requirement comes from a widely held belief that 5 ft/sec is necessary to create a turbulent flow and that such flow prevents the accumulation of the biofilm by mechanical shear forces. However, this notion is not supported by experimental results or analysis of the fluid dynamics.^[10,11] First, the flow becomes turbulent at velocities well below 5 ft/sec. Second, even with the turbulent flow, there is always a thin film of liquid adjacent to the piping wall that exhibits laminar flow. The thickness of this film, although dependent on the velocity, is certainly much higher than the size of typical bacteria. However, it would be unfair to say that the circulation velocity does not matter. It shall be selected not only based on the pipeline hydraulics, but also considering its effect on maintaining required water temperature at the various points in the system, including any “dead legs.” The longer the “dead leg” is, the higher is degree of turbulence needed to keep it hot.

Although hot storage and distribution systems similar to the one shown in Fig. 4 above are excellent from the standpoint of microbial control, their main limitation is the fact that water is delivered to all users at the high (65–85°C) temperature. Very often, this is not acceptable for some or all users. Other disadvantages of the hot systems are the higher potential for corrosion of the construction materials, and personnel safety concerns related to the

high water temperature. Various design concepts have been developed to provide the users with purified water at required (usually ambient) temperature.^[12] Two of them are shown in Figs. 6 and 7. In the first case (Fig. 6), water is cooled directly at the point-of-use by a local cooler; in the second case (Fig. 7), a separate ambient temperature loop is created, where one central cooler serves the needs of multiple use points. During the periods of no water usage, the cooler in the system in Fig. 7 does almost nothing (just compensating for the heat gains from the pump). However, once a point-of-use valve is opened and ambient water starts flowing out of the loop, it is replaced by hot water from the storage tank. At this point, the cooler needs to be able to remove a lot of heat to keep the set loop temperature. That usually requires a very large sanitary heat exchanger and a lot of cooling water. From the practical prospective, it is often beneficial to install two exchangers in series: the first one uses cooling tower water to remove most of the heat, whereas the second (“trim cooler”) uses more expensive chilled water to bring the loop to required temperature (typically 20–25°C). It is important to point out that such system cannot be expected to maintain exact loop temperature at all times. Because of the control system delays inherent in its design, the purified water temperature will fluctuate by a few degrees every time a major use point valve opens or closes. This is

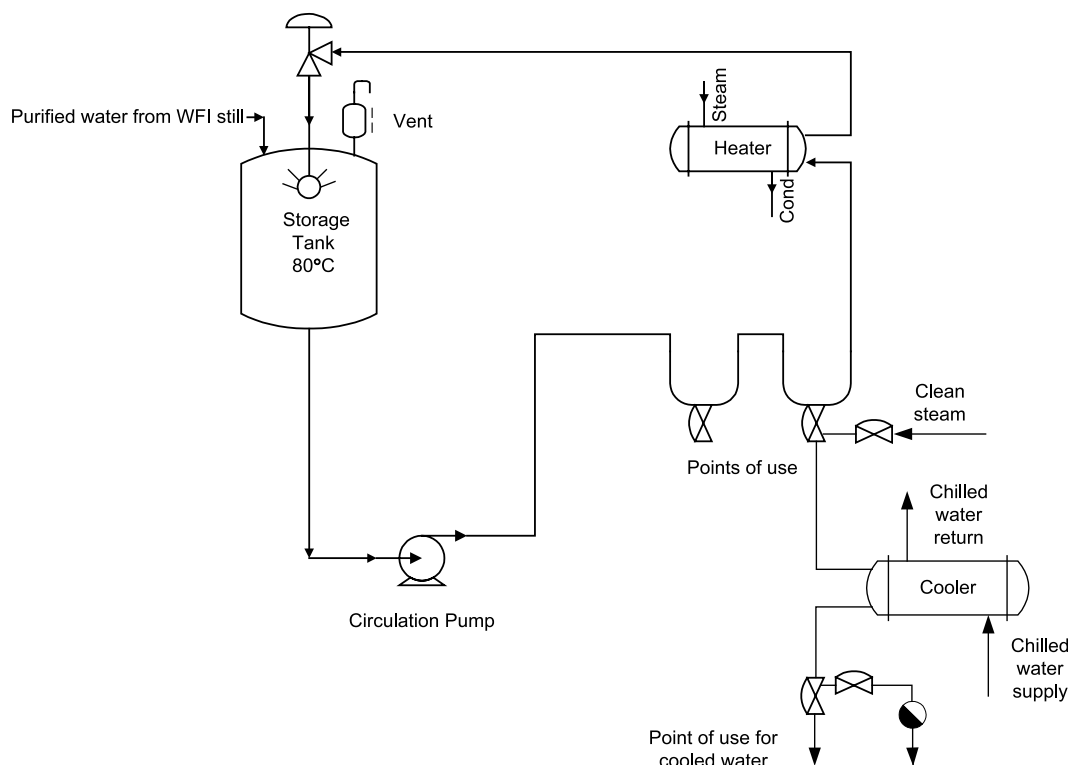


Fig. 6 Purified water storage and distribution system with point-of-use cooler.

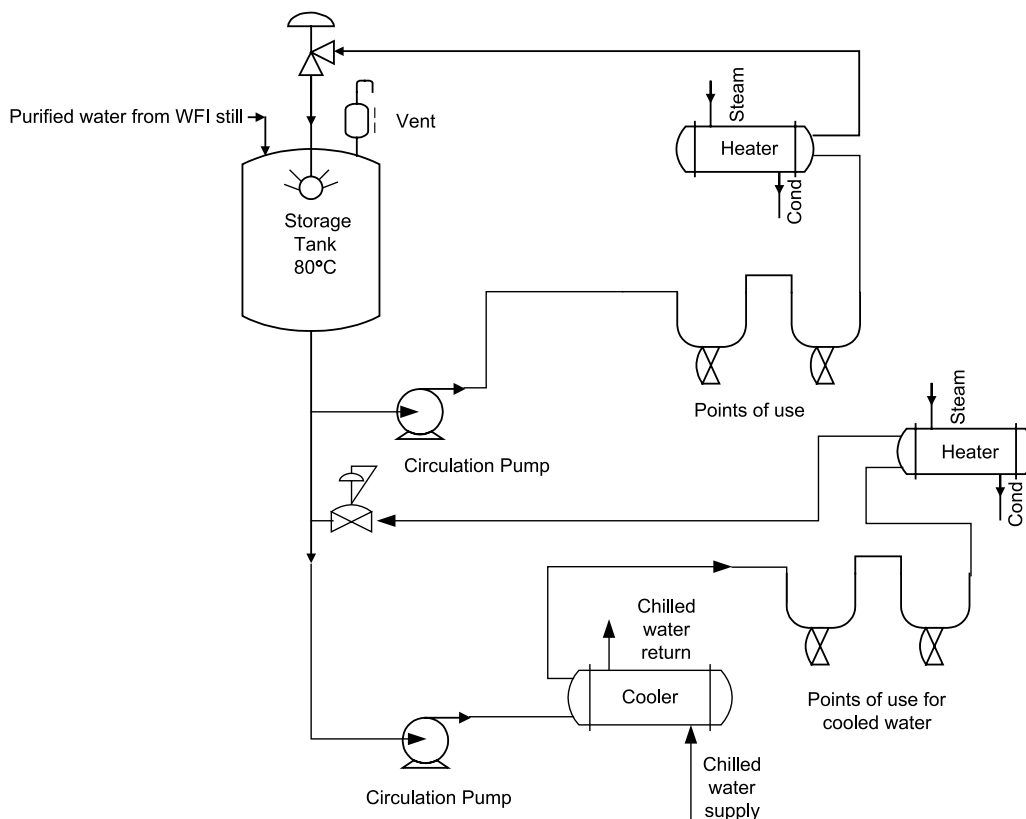


Fig. 7 Purified water storage and distribution system with ambient loop.

normally not a problem, especially in the applications involving filling the water into a process vessel. The heater in the ambient loop (Fig. 7) is provided for periodical loop sanitization; the steam supply is shut off at all other times.

A third approach is to use the system similar to the one shown in Fig. 4, but to keep the water in the storage tank and in the circulation loop at ambient temperature most of the time (this requires an extra heat exchanger with cooling water on the shell side). To achieve microbial control, such system (or any system that includes ambient loop) needs to be periodically sanitized by heating up its contents to about 80°C and holding the temperature for at least 1 hr. The sanitization frequency can vary from once a day to once a week or less: the lower the frequency is, the more attention to the studies proving the effectiveness of the microbial control shall be expected.

An alternative approach to microbial control in the ambient systems is to use ozone as a sanitizing agent. In such case, ozone gas (generated onsite from compressed air or oxygen) is continuously injected into the purified water storage tank. An UV light fixture is installed at the outlet of the circulation pump to decompose ozone in the water directed to the users during normal operation. Periodically, this UV light is shut off for 30 min to 1 hr,

allowing ozonated water to circulate through the distribution piping, thereby achieving its sanitization.

In larger systems, it often becomes impractical to run one large-diameter piping loop to all use points; in such cases, various multiloop arrangements may be employed.^[13]

The selection of a particular system design depends on many factors: number of use points requiring ambient and hot water, flow rates and frequency of usage, availability of capital, etc.^[14]

Materials of Construction

Storage tank, piping, pumps, and other components of the system in contact with purified water should be made out of materials that are chemically resistant to such water and which will not introduce metal ions or other contaminants. The most common material used in the pharmaceutical industry is 316L stainless steel. The suffix “L” in the grade designation stands for “low carbon,” which is important for the components that need to be welded because carbon molecules tend to diffuse to the surface during welding, making weld areas susceptible to corrosion. The piping systems for purified water are usually



constructed of the sanitary tubing. This product has tighter dimensional tolerances than piping, and is available with the wide range of surface finishes, up to the mirrorlike surfaces achieved by the combination of mechanical polishing and electropolishing. It is common to use the tubing polished to 25–30 μm . roughness average (RA) or better.^[15]

The tubing and other components (valves, elbows, tees, etc.) are welded for the most part by the automatic orbital welding machines. Such machines work by creating an electrical arc between the weld electrode and the tubing pieces, melting the ends of the tubes being joined. Unlike conventional welding, this technique creates a very smooth weld containing only the materials from the joined pieces and no metal from the welding electrode. To avoid excessive oxidation of the weld area, it has to be purged with inert gas such as argon during the procedure. In addition to reducing the manual labor, the automatic welding machines help in maintaining a consistent quality of all welds, as well as in documenting parameters (electrical current, electrode speed, purge flow, etc.) affecting such quality. Where automatic welding is not feasible (such as, for example, in the fabrication of zero-static valves), manual welding is allowed; however, it is recommended that it be conducted in the controlled shop environment rather than at the construction site. Welds in sanitary piping are often inspected by borescoping.

To improve the corrosion resistance of the stainless steel components, particularly in the weld-affected areas, it is a common practice to passivate the product contact surfaces of installed piping and equipment by circulating a solution of strong oxidant such as nitric or citric acid, or a chelating agent. Passivation removes free iron from the metal surface and creates a thin protective layer of chromium oxide. It is not uncommon to repeat the passivation treatment periodically (every few years) to restore the passive layer.

In the later years, some of the systems are designed with the plastic materials of construction instead of the stainless steel. The most common is polyvinylidene fluoride (PVDF); it is a partially fluorinated polymer that exhibits many desirable properties (mechanical strength and chemical resistance) approaching those of the fully fluorinated Teflon, but is less expensive and better suited for the installation in piping systems. Properly designed and supported PVDF piping can withstand hot water or even low-pressure steam sanitization, as well as periodical exposure to ozone. Although the PVDF tubing itself costs about as much as the stainless steel and requires continuous support in most cases, it offers advantages such as light weight, easier welding, less expensive valves and fittings, and no need for passivation. These benefits may result in overall installed cost savings compared to the stainless steel systems. Because of the concern about the

leaching of organic impurities from plastics, researchers and suppliers of the sanitary PVDF tubing conducted extensive studies proving that the amount of TOC leached from such tubing into water is very low and does not degrade the water quality.^[16,17] It is fair to say that whereas none of the construction materials can be considered ideal for the purified water systems, plastics such as PVDF emerge as strong competitors to the stainless steel.^[18]

Piping Design and Installation

While designing and installing the piping system for purified water, care shall be taken to avoid any areas

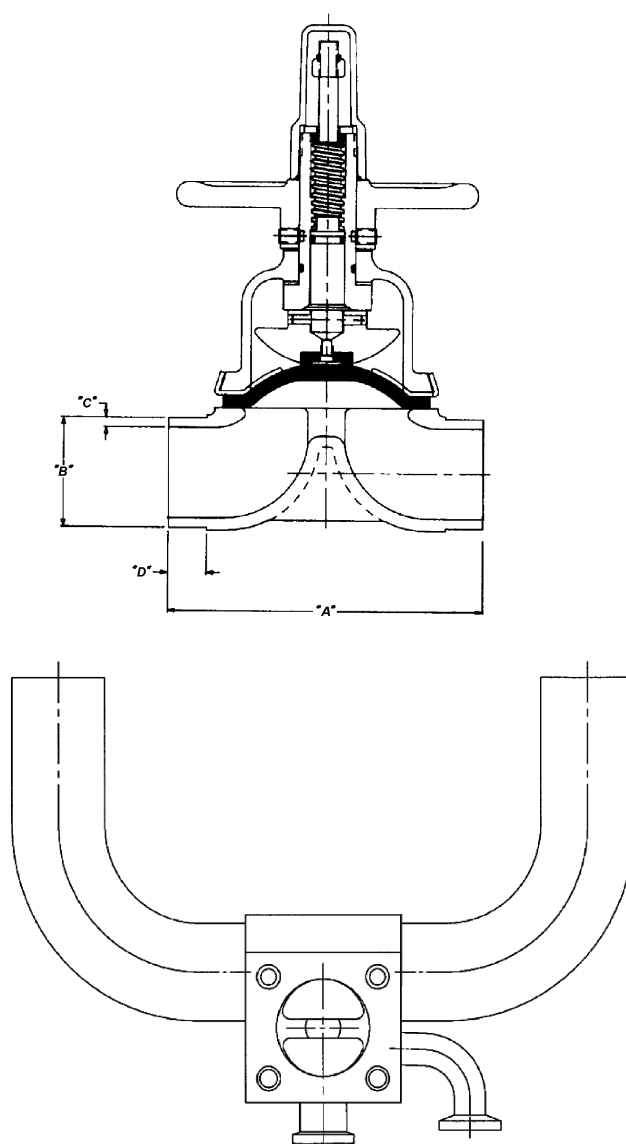


Fig. 8 Weir-style diaphragm valves: shutoff and point-of-use configurations. (Courtesy of ITT Industries.)

where water would remain stagnant (“dead legs”). Where branch connections are necessary, such as at the points of use, instrument ports, etc., the length of the “dead leg” is usually limited to a maximum of three pipe diameters of the branch line, as measured from the main line wall. The importance of this “rule of thumb” in this particular application is in the fact that while the piping system is being sanitized (either periodically or continuously), all parts of it, including the most remote areas, have to reach the required temperature, or, in case of chemical sanitization, be accessible for that chemical. In cases where long “dead legs” cannot be avoided (such as in some existing systems or in smaller new systems), microbial control can be achieved by flushing each of such legs with hot water or chemical solution as part of the sanitization procedure. The piping is usually sloped a minimum of 1/8 in/ft (or about 1:100) toward the nearest drain point to ensure complete drainage. Keep in mind that the thin-walled tubing used for the sanitary systems may sag

between the support points, which would create undrainable pools of liquid in the horizontal runs. Technically speaking, this sloping requirement is only important for the lines sanitized by steam (the condensate needs to drain freely to achieve even heating of all pipe surfaces), but it has become such a common practice that the author is not aware of any purified water systems in the pharmaceutical industry installed without slope. Most of the piping and other system elements are welded, and those components that may require disassembly (pumps, heat exchangers, instruments, etc.) are connected by sanitary joints such as Tri-Clamp, Swagelock TS, etc. The common feature of all such joints is that gasket fits flush to the tube wall, making the joint crevice-free. Diaphragm valves have become a de facto standard for purified water as well as for most other sanitary systems because of their cleanability, absence of crevices, and complete isolation of the wetted parts from the actuator mechanism. The two most widely used styles are weir and radial diaphragm valves, respectively. In addition to simple (one inlet, one outlet) valve configuration, several companies offer a wide variety of point-of-use and other specialty valves designed for the pharmaceutical and related industries in general, and for the purified water systems in particular (Figs. 8 and 9).

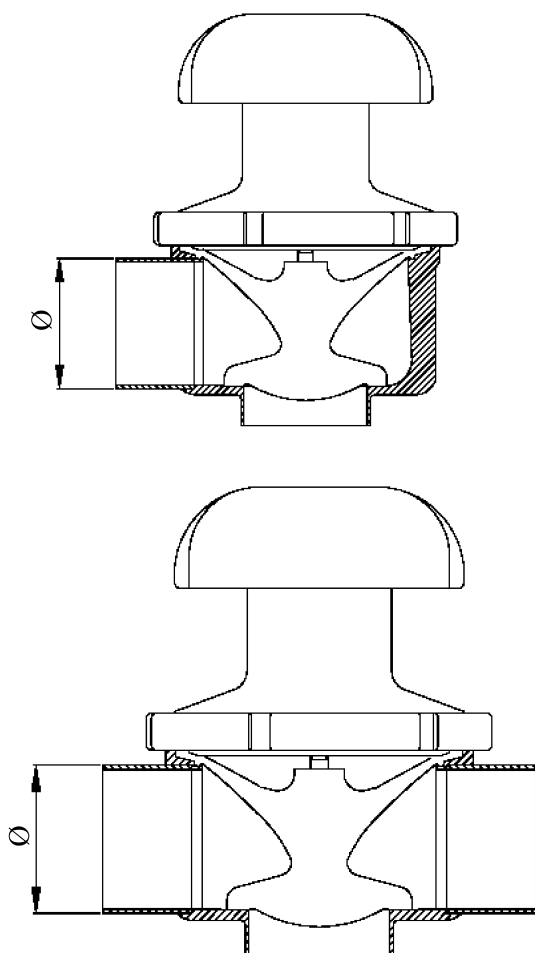


Fig. 9 Radial-style diaphragm valves: shutoff and flow-through configurations. (Courtesy of NovAseptic.)

CONCLUSION

There are many types of purified water systems used in pharmaceutical facilities. Although most of them share common features, each system is custom-designed for a specific application. Developing a proper design requires a good understanding of system operation and careful attention to details. Simply following common “rules of thumb” does not necessarily guarantee a reliable system—no matter how much money is spent. On the other hand, with a good understanding, it is often possible to design, install, and validate a functional and reliable purified water system with less capital investment and lower operating costs.

REFERENCES

1. *United States Pharmacopeia 26 National Formulary 21*; US Pharmacopeia Convention Inc.: Rockville, MD, 2003.
2. Meltzer, T. *Pharmaceutical Water Systems*; Tall Oaks Publishing: Littleton, CO, 1997.
3. Weitnauer, A.; Comb, L. Reverse Osmosis for WFI and PW. Presented at Pharmaceutical Waters'95, Atlantic City, NJ, May 1995.
4. Schaule, G.; Flemming, H. Pathogenic microorganisms in

- water system biofilms. *Ultrapure Water* **1997**, *14* (4), 21–25.
5. Gillis, R.; Gillis, J. A comparative study of bacterial attachment to high-purity water system surfaces. *Ultrapure Water* **1996**, *13* (6), 27–32.
 6. Riedewald, F. Biofilms in pharmaceutical waters. *Pharm. Eng.* **1997**, *17* (6), 8–19.
 7. Collentro, W. Microbial control in purified water systems—Case histories. *Ultrapure Water* **1995**, *12* (3), 30–36.
 8. Byrne, W. Biological Control in High-Purity Water. Presented at Ultrapure Water Expo, Philadelphia, PA, 2001.
 9. *ISPE Baseline Pharmaceutical Engineering Guides: Vol. 4. Water and Steam Systems*; ISPE, 2001.
 10. Gray, G.C. Recirculation velocities in Water For Injection distribution systems. *Pharm. Eng.* **1997**, *17* (6), 28–33.
 11. Klauer, J. An examination of pipe self-cleaning in high-purity water systems. *Ultrapure Water* **2001**, *18* (3), 56–60.
 12. Shnayder, L. Pharmaceutical purified water storage and distribution systems—An engineering prospective. *Pharm. Eng.* **2001**, *21* (6), 66–72.
 13. Dreyer, C.; Du, S.; Sommer, K.; Williams, R. Criteria, tools, and practices for high-purity water distribution systems. *Ultrapure Water* **2000**, *17* (5), 17–28.
 14. McWilliam, A. The design of high purity water distribution systems. *Pharm. Eng.* **1995**, *15* (5), 54–70.
 15. Peterman, L. Design Guidelines for WFI and Other High-Purity Water Systems. Presented at ISPE Conference on Pharmaceutical Water Systems, Amsterdam, The Netherlands, 1991.
 16. Motomura, Y.; Yabe, K. *Piping Materials and Distribution Systems for Advanced Ultrapure Water*; Proceedings of Tenth Annual Semiconductor Pure Water Conference, Santa Clara, CA, 1991; 1–22.
 17. Govaert, R.; Lueghamer, A. Thermoplastic piping systems for pharmaceutical water applications. *Ultrapure Water* **2001**, *18* (10), 32–39.
 18. Sixsmith, T.; Jackson, J. How piping materials for the pharmaceutical industry compare to each other. *Ultrapure Water* **1999**, *16* (4), 53–59.

WATER SORPTION OF DRUGS AND DOSAGE FORMS^a

Mark J. Kontny

Pharmacia, Kalamazoo, Michigan, U.S.A.

James J. Conners

Dura Pharmaceuticals, San Diego, California, U.S.A.

INTRODUCTION

The physical, chemical, and mechanical properties of pharmaceutical drugs and dosage forms are critically dependent on the presence of moisture. Pharmaceutical scientists can cite numerous examples of desirable and undesirable properties that result from varied levels of moisture associated with a particular solid or formulation consisting of mixtures of solids. Flow, compaction, caking, disintegration, dissolution, hardness, and chemical stability are just some of the properties influenced by moisture. Because water is present in bulk liquid form or as a vapor at some relative humidity in virtually all stages of solid manufacture (active ingredient and excipients), storage, processing into formulations, and final product packaging, a fundamental understanding of the role of water in affecting solid properties (and vice versa) is necessary.

Although the properties of individual solids and the performance of solid dosage forms are dependent on moisture, characterization of the underlying water–solid interaction is often nebulous. For example, many solids are described as “hygroscopic” without further reference to whether and how this relates to the rate and amount of moisture uptake as a function of relative humidity and temperature (1). To illustrate this ambiguity, consider that water-soluble, nonhydrating crystalline substances such as sodium chloride sorb very low levels of moisture (e.g., less than 0.1%) below their critical relative humidities, where the solid actually dissolves in the sorbed moisture. On the other hand, some typical excipient materials used in solid dosage forms, such as starches, celluloses, and gelatin capsules, sorb significant quantities of moisture (e.g., 25–50%), and even though they do not dissolve, they do undergo significant morphological change at high relative humidities (i.e., swelling). Moisture uptake rate

for a material depends on both the relative humidity of the environment and the time-dependent moisture content of the solid. For situations in which the environmental relative humidity is significantly different from the relative humidity at which the excipient (gelatin, for example) was previously equilibrated, the initial moisture uptake/loss rate will be significant, but it will approach zero over time. On the other hand, a water-soluble, nonhydrating crystalline substance such as sodium chloride will have a very low moisture uptake/loss rate that will decrease to zero if the environmental relative humidity is kept below its critical relative humidity. However, the uptake rate will be relatively large and continuous, until all the solid has dissolved, if the relative humidity is above the critical value. Obviously, very different mechanisms of water sorption/desorption occur for the different samples. In this light, describing sodium chloride and/or starch as “hygroscopic” offers very little toward understanding the water–solid interactions that might affect their physico-chemical properties. These examples illustrate the need to understand the underlying mechanism(s) of uptake for a particular solid. In this regard, therefore, addressing the following questions provides a basis for studying the various mechanisms of water–solid interaction:

1. How much water is present and what is the corresponding water activity (approximated by relative pressure or percent relative humidity/100)?
2. What are the kinetics of moisture uptake or loss, and is the rate constant or changing over time?
3. Where is the water located (i.e., adsorbed to the external surface of crystals, absorbed into crystals as specific or nonspecific water of hydration, absorbed into amorphous regions, condensed into pores, etc.)?
4. What is the state of the moisture associated with the solid (i.e., bulk water, water of hydration, physisorbed water, etc.)?
5. What form of the solid is present (i.e., particle size and morphology, polymorphic species, degree of crystallinity, state of hydration) and is this form

^aThis work was taken in part from a chapter entitled “Sorption of Water by Solids,” in *Physical Characterization of Pharmaceutical Solids*, H. Brittain, Ed., Marcel Dekker, Inc.: New York, 1995.

thermodynamically stable over the temperature and relative humidity range that the solid is expected to encounter?

It is the objective of this chapter to highlight the various mechanisms whereby water can interact with solid substances, present methodologies that can be used to obtain the necessary data, and then discuss moisture uptake for nonhydrating and hydrating crystalline solids below and above their critical relative humidities for amorphous solids and for pharmaceutically processed substances. Finally transfer of moisture from one substance to another will be discussed.

THE WATER SORPTION ISOTHERM

The most fundamental manner of demonstrating the relationship between sorbed water vapor and a solid is the water sorption–desorption isotherm. The water sorption–desorption isotherm describes the relationship between the equilibrium amount of water vapor sorbed to a solid (usually expressed as amount per unit mass or per unit surface area of solid) and the thermodynamic quantity, water activity (a_w), at constant temperature and pressure. At equilibrium the chemical potential of water sorbed to the solid must equal the chemical potential of water in the vapor phase. Water activity in the vapor phase is related to chemical potential by

$$\mu = \mu^0 + RT \ln a_w \quad (1)$$

where μ is the chemical potential of water in the system at equilibrium, μ^0 is the standard chemical potential of water at a specific reference temperature and pressure, R is the gas constant, and T is the absolute temperature. Lewis et al. (2) defined the relative activity of any pure substance or component (such as water) as a ratio of fugacities:

$$a_w = \frac{f_w}{f_w^0} \quad (2)$$

where f_w is the fugacity of water in the system at equilibrium and f_w^0 is the fugacity of pure water at a standard temperature and pressure. For all practical purposes, the fugacity (or “escaping tendency”) of water vapor can be approximated by the water vapor pressure in the system. This assumption is valid as long as the water vapor behaves as an ideal gas. For the water vapor pressure range of relevance for pharmaceutical systems at temperatures less than 50°C, this approximation is excellent (<0.2% relative error) (3). Thus, the relative

pressure of water vapor, P/P^0 , is usually employed as an estimate of the relative water activity in the system:

$$a_w = \frac{P}{P^0} \quad (3)$$

where P is the water vapor pressure in the system and P^0 is the vapor pressure above pure water at the temperature of interest. Relative humidity (RH) is defined as the relative pressure expressed on a percentage basis:

$$\text{RH} = 100 \frac{P}{P^0} \quad (4)$$

The sorption branch of the isotherm is obtained experimentally by measuring the equilibrium amount of water sorbed to a solid at known relative pressure, beginning with a known mass of absolutely dry solid and then progressively increasing the relative pressure in the system. Drying the solid sample under heat, possibly using vacuum to facilitate the removal of desorbed water vapor, is usually necessary to eliminate residual moisture. One must be aware, however, of the effects of such conditions on the chemical and physical stability of the solid. The desorption portion of the isotherm is obtained by progressively decreasing the relative pressure in the system from a relative pressure of approximately unity, again monitoring the equilibrium amount of moisture sorbed at each relative pressure. Remember that the moisture sorption isotherm is an equilibrium measurement of the interaction of water with a solid. In theory, information regarding the kinetics of moisture uptake is not explicitly derived from this experiment. This distinction is an important one that will be explored in more depth later.

Generation of water sorption–desorption isotherms for a particular solid can lend considerable insight into the nature of the water–solid interaction, as well as the surface characteristics of the solid. This information is readily obtained from the amount of moisture sorbed at lower relative humidities in comparison with the specific surface area of the sample, from the general shape of the isotherm, from whether or not water uptake is a completely reversible process (i.e., whether hysteresis is observed between sorption and desorption), and from the shape of the hysteresis loop if it is present. With knowledge of the aforementioned, one can usually obtain an indication of the mechanism of moisture sorption for the material of interest. For example, a material that exhibits sorption at lower relative humidities in much greater amounts than one might expect based on the specific surface area of the sample, and that exhibits hysteresis over the complete

range of relative humidities, is most likely absorbing water into its internal structure. On the other hand, a material exhibiting a closed hysteresis loop over the higher relative humidity range while sorbing moisture over the lower relative humidity range similar to what might be expected based on its specific surface area, is probably quite porous in nature and is most likely sorbing water via capillary condensation over the higher relative humidity range.

MODELS DESCRIBING VAPOR ADSORPTION

Brunauer, Emmett, and Teller Equation

The model most commonly referred to in the literature describing vapor adsorption onto solid surfaces was put forth in 1938 by Brunauer, Emmett, and Teller (4). The so-called BET model was originally derived using kinetic arguments in a manner very similar to those used by Langmuir (5). The BET model has since also been derived using statistical mechanics (6–8). The BET model assumes that vapor molecules, behaving as an ideal gas, exist in a state of equilibrium with a solid that consists of identical, homogeneous adsorption sites. The first vapor molecule adsorbed to an adsorption site on the solid is proposed to be bound, whereas molecules adsorbing beyond the first layer are assumed to have the properties of the bulk liquid. Furthermore, adsorption is proposed to occur such that the adsorbed molecules do not interact laterally. The linear form of the BET equation is

$$\frac{1}{W \left[\frac{P^0}{P-1} \right]} = \frac{(C_b - 1) * \frac{P}{P^0}}{W_m C_b} + \frac{1}{W_m C_b} \quad (5)$$

where W is the mass of vapor adsorbed per gram of solid at a particular relative pressure, P/P^0 ; W_m is the theoretical quantity of vapor adsorbed when each adsorption site has one vapor molecule adsorbed to it; and

$$C_b = k \exp \left(\frac{H_1 - H_L}{RT} \right) \quad (6)$$

where H_1 is the heat of adsorption of the first vapor molecule adsorbed to a site, H_L is the heat of condensation of the bulk adsorbate, R is the universal gas constant, T is the absolute temperature, and k is a constant, usually assumed to be close to unity. The two BET constants, W_m and C_b , can easily be obtained from the linear plotting form of the BET equation given in Eq. 5. Plotting the quantity $1/[W(P^0/P - 1)]$ versus P/P^0 gives a slope equal

to $(C_b - 1)/W_m C_b$ and an intercept equal to $1/W_m C_b$. Algebraic manipulation gives

$$W_m = \frac{1}{\text{slope} + \text{intercept}} \quad (7)$$

and

$$C_b = 1 + \frac{\text{slope}}{\text{intercept}} \quad (8)$$

In general, the BET equation fits adsorption data quite well over the relative pressure range 0.05–0.35, but predicts considerably more adsorption at higher relative pressures than is experimentally observed. This is consistent with an assumption built into the BET derivation that an infinite number of layers are adsorbed at a relative pressure of unity. Application of the BET equation to nonpolar gas adsorption results is carried out quite frequently to obtain estimates of the specific surface area of solid samples. By assuming a cross-sectional area for the adsorbate molecule, one can use W_m to calculate specific surface area by the following relationship:

$$S = \frac{W_m X N_{av}}{M \Sigma} \quad (9)$$

where S is the specific surface area in m^2/g ; W_m is the mass of adsorbate adsorbed at monolayer coverage; X is the cross-sectional area of an adsorbed adsorbate molecule (assumed to be 19.5 \AA^2 for krypton, 16.2 \AA^2 for nitrogen, 12.5 \AA^2 for water (9, 10)); N_{av} is Avogadro's number; M is the molecular weight of adsorbate; and Σ is the mass of the sample. Obviously, calculating surface areas from moisture uptake data that does not lead to monolayer coverage at W_m [either incomplete coverage (see the section on "Water Sorption onto Nonhydrates" below) or absorption into the solid (see the section on "The Meaning of Specific Surface Areas Calculated from Water Absorption Studies")] will result in incorrect values that have no physical meaning. Therefore, comparison of the surface area measured by nonpolar gas adsorption to that calculated from a moisture sorption isotherm can lend insight into the fundamental interactions between the water and the solid. This will be explored in some depth later.

Guggenheim and deBoer Equation

Many attempts to modify the BET adsorption theory have been made since its original derivation. Its simplicity and ability to fit adsorption data extremely well at lower relative pressures, however, have made it the model of choice for estimating surface areas from nonpolar gas

adsorption. Most modifications of the BET model, developed to analyze data over the entire range of relative pressures, usually add at least one fitting parameter to the equation. This makes computer fitting a necessity, because only two measurable parameters, W and P/P^0 , are available. From a modeling perspective, additional fitting parameters of unknown or undefined physical meaning that arise from such approaches are often a deterrent to the use of multiparameter models because of the consequent difficulty in interpreting results. In this regard, therefore, only a single modification of the BET model, which has been shown to extend the relative pressure range over which vapor adsorption data can be fit, will be considered here. This extension of the BET model, independently derived by Guggenheim (11) and deBoer (12), accounts for the adsorption of an intermediate state of vapor between the tightly bound first molecule adsorbing to an adsorption site and the condensed molecules adsorbed at very high relative pressures. Molecules adsorbed in the intermediate range can be considered to interact with the solid, but the interaction is assumed to be considerably less than that of the first molecule sorbed at an adsorption site. Obviously, this addition of a "third state" of interaction is an approximation. In all likelihood, there is a continuum of interaction states. However, from a computational point of view, the existence of three or more states is indistinguishable. The equation for the three-state interaction model is given as

$$W = \frac{W_m C_G K \frac{P}{P^0}}{\left[1 - K \frac{P}{P^0}\right] \left[1 - K \frac{P}{P^0} + C_G K \frac{P}{P^0}\right]} \quad (10)$$

where P , P^0 , H_L , W , and W_m are identical to the parameters used in the BET equation, and

$$K = B \exp\left(\frac{H_L - H_m}{RT}\right) \quad (11)$$

where B is a constant and H_m is the heat of adsorption of vapor adsorbed in the intermediate layer. The constant C_G is defined as

$$C_G = D \exp(H_1 - H_m RT) \quad (12)$$

where D is a constant, H_1 is the heat of adsorption of the first molecule adsorbed at a site, and H_m is the heat of adsorption of the intermediately bound molecule.

Water Vapor Absorption by Amorphous Solids

The process of water vapor interaction with amorphous solids has been likened to the production of a solid solution in which the water is dissolved in the solid matrix. As more

water is absorbed, the fundamental properties of this solid matrix (e.g., viscosity) can undergo significant change that can then result in visually apparent changes in physical properties (e.g., collapse, recrystallization) (13–17). This will be discussed in further detail later.

Although water vapor is absorbed into amorphous solids and not simply adsorbed on the surface, it still has been found that such sorption isotherms can be fit to the BET equation up to a P/P^0 of about 0.40 as with vapor adsorption, and over the entire range of P/P^0 using its extension, Eq. 10. Because this was first reported by Anderson (18) to be the case for water absorption, Eq. 10, when applied to water vapor sorption, is often called the GAB equation for Guggenheim, Anderson, and deBoer (19). Because the theoretical process for the derivation of the original equation does not translate directly to the absorption process, which involves dissolution of water in the amorphous solid, the significance of fit to the GAB equation is somewhat limited. It is, however, a very useful equation because it does allow one to describe the entire isotherm and to draw out some useful parameters (to be discussed in what follows).

Since water vapor dissolves in the solid during absorption, several models based on solution theory, proposing that the sorbate is taken up into the solid as a solid solution, have been derived and used to describe water sorption on polymers [e.g., Flory-Huggins (20), Hallwood-Horrobin (21)]. More recently, Vrentas et al. (22, 23) developed a solution-based model that accounts for the plasticizing effect of water on a polymer that has been shown to describe the entire moisture uptake isotherm for the polymer (24). While these sorption theories and the many modifications of the BET adsorption model are based on meaningful physicochemical principles, further work is still required to elucidate the molecular mechanisms underlying moisture uptake into polymeric systems. From this perspective, other models based on entirely different theoretical concepts will not be considered further in this chapter. For further reference, the reader is directed to several excellent literature reviews of the many sorption theories that have been proposed (25, 26).

Capillary Condensation

Vapor sorption onto porous solids differs from vapor uptake onto the surfaces of flat materials in that a vapor (in this case, water) will condense to a liquid in a pore structure at a vapor pressure, P_r , below the vapor pressure, P^0 , where condensation occurs on flat surfaces. This is generally attributed to the increased attractive forces

between adsorbate molecules that occur as surfaces become highly curved, such as in a pore or capillary. This phenomenon is referred to as capillary condensation and is described by the Kelvin equation (27):

$$\ln\left(\frac{P_r}{P^0}\right) = -2\gamma V_m r RT \quad (13)$$

where γ is the surface tension of the adsorbed film (assumed equal to that of the bulk liquid), V_m is the molar volume of the liquid, r is the pore radius, R is the gas constant, and T is the temperature. The Kelvin equation has been shown to be applicable to pore radii as low as 5 nm for water adsorption onto mica (28, 29). As mentioned in the section on The Water Sorption Isotherm, capillary condensation will result in a closed hysteresis loop in the adsorption/desorption isotherm of a porous material. Calculating P_r/P^0 by assuming a surface tension of water of 72.8 ergs/cm² and a density of 0.998 g/cm³ at 293 K [this is an assumption for the purposes of this calculation as it has been shown that the density of water is lower in a pore than the bulk value (30)] shows that condensation is predicted at relative pressures of 0.998, 0.989, 0.898, and 0.340 for pore radii of 1000, 100, 10, and 1 nm, respectively. In this regard, it is clear that capillary condensation need only be considered for very small pore dimensions. In practical terms, one should be concerned about this mechanism of water uptake for microporous pharmaceutical powders that exhibit a relatively large specific surface area (i.e., >100 m²/g), as determined from nonpolar gas adsorption studies.

METHODOLOGY

Control of Relative Humidity

Maintenance of constant relative humidity environments is essential for studying water–solid interactions. There are primarily four techniques that are frequently employed to maintain constant relative humidity:

1. Colligative solutions.
2. Temperature modification of an aqueous solution.
3. Control of total pressure over the solid.
4. Mixing wet and dry air streams.

Saturated salt solutions and sulfuric acid solutions establish relative humidity by reducing the vapor pressure above an aqueous solution (a colligative effect). Saturated salt solutions at controlled temperature maintain a constant relative humidity as long as excess salt and bulk solution are present. As water is added or removed from the solution,

moisture from the headspace will either condense or evaporate (as appropriate), with subsequent dissolution or precipitation of salt to maintain the equilibrium vapor pressure. Because the degree of vapor pressure depression is dependent on the number of species in solution and, further, since the solubility of most salts is somewhat dependent on temperature, the relative humidity generated is also temperature dependent. Hence, use of the same salt at different temperatures can result in different relative humidities. Refs. 31–35 can be consulted for specific saturated salt solutions that result in defined relative humidities as a function of temperature. Because relative humidity is dependent on the number of dissolved species, it is essential that saturation be attained prior to beginning experimentation. In this regard, preparing the salt solutions several days before beginning a sorption study is recommended. Sulfuric acid solutions of varying concentration (35) are also used to establish relative humidity. Addition or removal of water from the solution by desorption or sorption of water to the solid, however, will alter the concentration of sulfuric acid (and water) in solution, and thus change the relative humidity of the headspace. This technique for controlling relative humidity in the headspace is practically more useful when small amounts of water are sorbed/desorbed from the solid.

Temperature modification of an aqueous solution can also be used to maintain constant relative humidity in the headspace (19). This technique maintains the solid at one temperature and an aqueous solution connected to the system at another temperature. Due to strong vapor pressure dependence on temperature, very tight temperature control of the aqueous solution and the solid are required to maintain constant relative humidity in the vicinity of the solid by this technique.

Control of the vapor pressure in the headspace over a solid can also be used to maintain a relative humidity over a solid. As shown in Eq. 4, the relative humidity is directly correlated to the partial pressure of water in the vapor phase. To utilize this technique for relative humidity control, the headspace above the sample must be completely evacuated prior to analysis. Pure water vapor can then be carefully admitted to the vapor phase. Because only water vapor is present, the pressure measured over the system is directly related to the relative humidity over the sample (36).

Mixing dry and water vapor saturated air in defined proportions also can be used to generate constant relative humidity. Control of flow rates and the water vapor content of the dry and saturated air are essential to ensure accurate, reproducible relative humidity production (37, 38).

Measurement of Relative Humidity

Measurement of relative humidity depends on the system used. Systems employing vacuum are usually evacuated prior to introduction of water vapor (36, 39). For cases in which a gas-forming reaction is not occurring, measurement of total pressure in the system can be used as a measure of water vapor pressure. Systems in which air is not evacuated require specific measurement of water vapor pressure. (For the latter type of system, caution should be taken to assure that the relative humidity source is in close proximity to the solid, since the diffusion of water vapor through air to the solid is required to maintain a constant relative humidity in the immediate vicinity of the solid.) A wide variety of pressure measuring instrumentation is commercially available with varying accuracy, precision, and cost.

Measurement of the Critical Relative Humidity, RH_0

The relative humidity at which a solid begins to deliquesce, RH_0 , can be determined in three ways: 1) directly, by measuring the relative humidity above a saturated solution of the substance; 2) by determining the relative humidity at which significant moisture uptake and simultaneous dissolution occurs, or; 3) indirectly, by measuring the steady state moisture uptake rate at relative humidities above RH_0 and then extrapolating to the relative humidity at which the moisture uptake rate is zero (1, 40, 41).

Although other techniques can be used to measure the relative humidity above a saturated solution, one relatively simple procedure is to evacuate the headspace (to remove air by vapor phase expansions) and then, with the vacuum pumps isolated and the saturated solution maintained at a constant temperature, to measure water vapor pressure. Water vapor pressure can then be converted to relative humidity by dividing by P^0 , the vapor pressure above pure water at the temperature of interest (42).

Measurement of Moisture Uptake (Kinetics of Deliquescence)

The rate of moisture uptake above RH_0 requires maintenance and measurement of a range of relative humidities, and the capability of measuring the moisture content of the solid over time. Use of a vacuum system can minimize vapor diffusion through the headspace, thus maintaining constant relative humidity in the vicinity of the sample. Also, because the most reliable estimate of the steady-state moisture uptake rate is when the integrity of the solid is intact and the film of sorbed moisture is thin

(and saturation most likely), it is advisable to determine the moisture uptake rate at early time periods. In this regard, it is also helpful to be able to view the solid during the experiment to verify that integrity is maintained and excess solid remains (41, 43).

Measurement of Equilibrium Moisture Sorption

Generation of water sorption/desorption isotherms in a controlled relative humidity environment can be carried out either gravimetrically or volumetrically. Gravimetric methods require

1. A dry sample weight,
2. Constant temperature of the sample,
3. Maintaining predetermined constant relative humidities in the headspace, and
4. Attaining and measuring an equilibrium weight of sorbed water vapor. Gravimetric measurement of moisture uptake can occur continuously or discontinuously. Continuous measurement usually involves placing a sample on a balance in a temperature- and relative humidity-controlled environment. Microbalances in closed systems have been used successfully for this purpose (37–39, 44), and commercial systems are now available that can accurately and precisely control relative humidity and simultaneously monitor sample weight.

Volumetric methods require

1. A dry sample weight,
2. Constant temperature of the sample,
3. Water vapor pressure measurement in a dosing volume and, later, in the headspace above the equilibrated sample, and
4. Measuring dead volumes of the individual chambers, including the sample chamber. In essence, volumetric methods equilibrate a known headspace dosing volume at a given (measured) water vapor pressure, followed by exposure of the pre-equilibrated sample to this water, with subsequent measurement of the water vapor pressure after equilibration. The mass of water sorbed, Δn (in moles), at the final pressure in the system, P_f , is obtained from the difference, ΔP , between P_f^{calc} , the calculated water vapor pressure at equilibrium, and P_f^{meas} , the final measured water vapor pressure:

$$\Delta n = \frac{\Delta PV}{RT} \quad (14)$$

where V is the final volume, R is the gas constant, and T is the absolute temperature (39).

WATER SORPTION BY CRYSTALLINE SOLIDS

General Model

Fig. 1 schematically describes the important steps in the uptake of water vapor by crystalline water-soluble solids. At low relative humidities, water is adsorbed to the surface of a nonhydrate-forming solid. As the relative humidity is increased, some tendency for multilayer sorption is expected. At some relative humidity (characteristic for a given substance), the solid will begin to dissolve in the sorbed film of water. A saturated solution of solute will most likely exist, and this will cause the vapor pressure over the sorbed film of water to be depressed relative to pure water and to be constant and equal to that above a saturated solution of the substance. This vapor pressure may be expressed as the critical relative humidity, RH_0 . If the relative humidity in the atmosphere is greater than that over the saturated solution (RH_0), water will spontaneously condense on the aqueous film. This will dilute the film allowing more solid to dissolve, which, in turn, will maintain the pressure gradient. The process of water vapor uptake will continue until all the solid has dissolved and further solution dilution has occurred. Only when the relative humidity above the solution is elevated to that of the atmosphere will this process terminate. This phenomenon is called deliquescence. Although hydrates undergo solid state transitions in transforming from the anhydrate to hydrate, as well as from one hydrate species to another, behavior similar to that previously described for nonhydrates is also noted at and above RH_0 for hydrates. In pharmaceutical systems, water-soluble species are frequently encountered in dosage forms. Thus, it is important to understand the conditions responsible for deliquescence and the molecular events occurring at relative humidities below the deliquescence point.

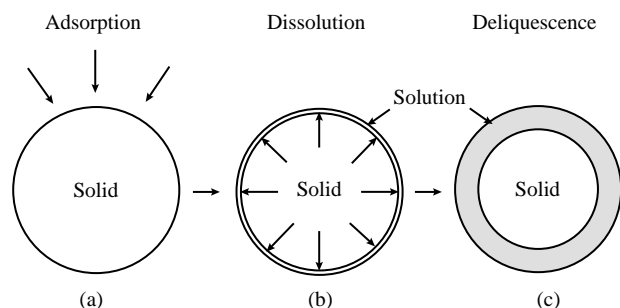


Fig. 1 Water vapor adsorption and deliquescence of a water-soluble solid at (a) atmospheric relative humidity, $RH_i < RH_0$; (b) $RH_i = RH_0$; (c) $RH_i > RH_0$.

Water Sorption onto Nonhydrates below RH_0

The sorption of water vapor onto nonhydrating crystalline solids below RH_0 will depend on the polarity of the surface(s) and will be proportional to surface area. For example, water exhibits little tendency to sorb to nonpolar solids like carbon or polytetrafluoroethylene (Teflon) (29), but it sorbs to a greater extent to more polar materials such as alkali halides (45–48) and organic salts like sodium salicylate (48). Because water is only sorbed to the external surface of these substances, relatively small amounts (i.e., typically less than 1 mg/g) of water are sorbed compared with hydrates and amorphous materials that absorb water into their internal structures.

Unfortunately, the literature is relatively sparse with examples showing the water uptake profile onto crystalline, nonhydrating substances below RH_0 . This is most likely due to the difficulty in accurately measuring the small amounts of water that are sorbed. Alkali halides are an exception, however, likely due to their well-characterized particle morphologies (45–48). Figure 2 shows a water uptake isotherm onto recrystallized sodium chloride (48). Note that the amount of water sorbed as a function of relative humidity is normalized to the specific surface area of the sample. Because water is sorbed only to the external surface of this material, this allows comparison of water uptake data from different lots of material, whereas plotting this data on a “per gram” basis would have little or no meaning. For the sodium chloride sample in Fig. 2 (specific surface area = $0.0875 \text{ m}^2/\text{g}$ from krypton adsorption studies), only $5 \times 10^{-4} \text{ g water/m}^2$ of sodium chloride is sorbed, even up to 70% relative humidity. Also, note the apparent step-like nature of the isotherm. From BET analysis of the sorption data at the lower relative humidities, a W_m value of $7.6 \times 10^{-5} \text{ g/m}^2$ is obtained. This value is only about 0.32 that of the predicted value for monolayer coverage assuming an area per water molecule of 12.5 \AA^2 . This suggests that it is quite meaningless to refer to the number of layers of sorbed water as multiples of W_m , except as a point of reference. Interestingly, the second step plateau in Fig. 2 occurs at about three times the moisture content corresponding to W_m , suggesting that for sodium chloride, the monolayer is actually completed during the second step of the isotherm. Isothermic heat of sorption results for sodium chloride from Barraclough and Hall (45) suggest that the heat of sorption of water up to W_m is invariant, whereas the heat of sorption decreases and becomes constant at about two times W_m . Considering the experimental error involved in obtaining W_m and the isosteric heats of sorption, this suggests that

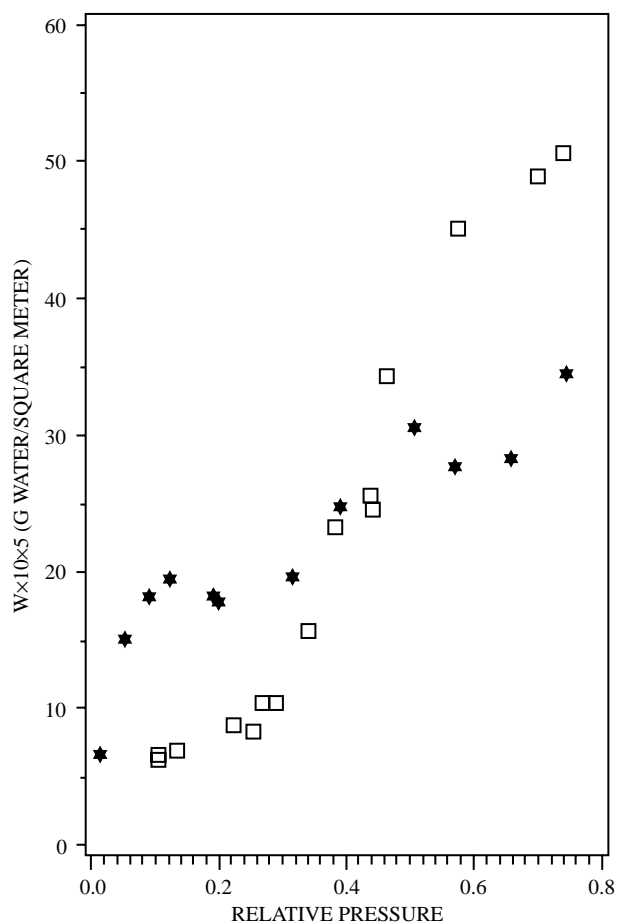


Fig. 2 Water vapor sorption for recrystallized (□) and ground (*) sodium chloride at 20°C. (From Ref. 48.)

water is sorbed with a homogeneous binding energy up to W_m and then interacts to a lesser extent until the monolayer is complete.

As shown in Fig. 2 (48) and also in the work of Barraclough and Hall (45), moisture uptake onto sodium chloride as a function of relative humidity is reversible as long as RH_0 is not attained. This is evidence that actual dissolution of water-soluble crystalline substances does not occur below RH_0 . This is consistent with the thermodynamic rationale that dissolution below RH_0 would require a supersaturated solution (i.e., an increased number of species in solution would be necessary to induce dissolution at a relative humidity below that of the saturated solution, RH_0). In this regard, one should only need to consider the solid state properties of a purely crystalline material below RH_0 . As will be described, other considerations may be warranted for a substance that exists in multiple polymorphic forms or contains amorphous material.

Water Sorption onto Hydrates below RH_0

Many drugs and excipients (cephalexin monohydrate, quinidine sulfate dihydrate, ampicillin trihydrate, codeine sulfate trihydrate, morphine sulfate dihydrate, dicalcium phosphate trihydrate, raffinose pentahydrate, lactose monohydrate) utilize water as an integral part of their crystal structure. Solids that form specific crystal hydrates tend to sorb relatively small amounts of water to their external surface below a characteristic relative humidity, when initially dried to an anhydrous state. Below this characteristic relative humidity, these materials behave similarly to nonhydrates. Once the characteristic relative humidity is attained, addition of more water to the system will not result in a further increase in relative humidity. Rather, this water will be sorbed so that the anhydrate crystal will be converted to the hydrate. The strength of the water–solid interaction depends on the level of hydrogen bonding possible within the lattice (29, 49). In some hydrates (e.g., caffeine and theophylline) where hydrogen bonding is relatively weak, water molecules can aid in hydrate stabilization primarily due to their space-filling role (29, 49, 50).

Since water molecules occupy regular positions within the lattice of a hydrate with a specific stoichiometry (e.g., 1:1 monohydrate, 2:1 dihydrate, 5:1 pentahydrate) to the solid, relatively large quantities of water are sorbed. Fig. 3 shows a moisture uptake isotherm for ipratropium bromide (51). This substance undergoes an apparent hydration of the crystal between 63% and 75% relative humidity. Above 75% relative humidity, approximately 4.6% water is sorbed (theoretical monohydrate is 4.4%). Interestingly, as anhydrous ipratropium bromide is equilibrated for extended time periods (e.g., 2 and 5 months respectively, as shown in Fig. 3), hydration of the crystal appears to occur at 53% and 63% relative humidity. This example clearly shows that a time period of many months may be required to attain a reliable estimate of the equilibrium uptake at select relative humidities. Characteristic of many hydrates, ipratropium bromide exhibits significant hysteresis between the sorption and desorption isotherms. This is attributed to the degree of binding and the physical fit of water in the hydrated lattice.

Nonspecific hydration, or hydration of the lattice without a first-order phase transition, also must be considered. Cox et al. (52) reported the moisture uptake profile of cromolyn sodium, and the related effects on the physical properties of this substance. Although up to nine molecules of water per molecule of cromolyn sodium are sorbed into the crystalline lattice at 90% relative humidity,

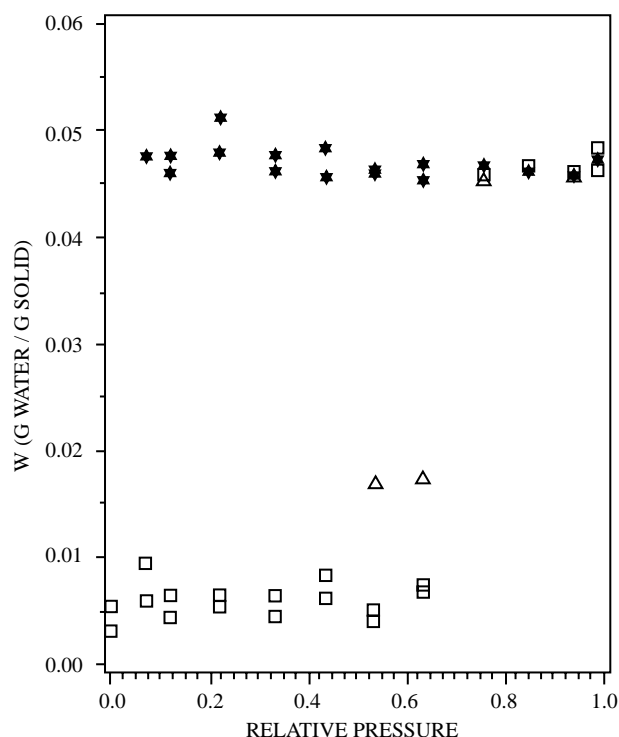


Fig. 3 Water vapor sorption and desorption isotherms for ipratropium bromide at 20°C. (□) 2- month sorption results; (△) 5- month sorption results; (*) 2- and 5- month desorption results. (Note: All 2- month sorption results, except at 53% and 63% relative humidity, were verified at 5 months.)

the sorption profile does not show any sharp plateaus corresponding to fixed hydrates. Rather, the uptake profile exhibits a gradual increase in moisture content as relative humidity increases, which results in marked changes in X-ray diffraction patterns, density, and other physical properties. For this example, moisture uptake onto cromolyn sodium was correlated with expansion of the lattice in the *b* crystallographic direction, which was shown to be reversible on dehydration.

A thorough understanding of the hydration profile for a solid forming a crystal hydrate is important for several reasons. First, because an anhydrate and hydrate(s) are distinct thermodynamic species, they will have different physicochemical properties (e.g., solubility) that may affect dissolution or bioavailability. Second, a desired hydrate species can be formed and used (and retained) simply by controlling the established environmental conditions. Third, because significant quantities of water can be sorbed/liberated as a hydrate becomes hydrated/dehydrated, the physicochemical properties of the immediate system (including other nearby solids) can be markedly affected.

The Critical Relative Humidity, RH_0

Knowledge of RH_0 for each component in a formulation, and for the entire system, is extremely important for predicting relative humidities where gross physical changes of the system are expected due to dissolution of water-soluble components. The value of RH_0 , as a colligative property, is significantly influenced by the number of species in solution. As a rule of thumb, two general comments can be made. First, compounds exhibiting poor water solubility typically have RH_0 values well above 95% relative humidity. Second, as solubility increases, RH_0 decreases. Since nonidealities are introduced as solutions become more and more concentrated, it is not usually possible to use dilute solution models (e.g., Raoult's law) to predict the expected RH_0 for a solute of significant aqueous solubility. Hence, RH_0 should be measured for individual solids. Examples of RH_0 values for single component systems are shown in Table 1.

Values of RH_0 for mixtures, on the other hand, can be calculated from the RH_0 values of single components using an equation developed by Ross (54):

$$\frac{(RH_0)_{\text{mix}}}{100} = \left(\frac{(RH_0)_1}{100} \right) * \left(\frac{(RH_0)_2}{100} \right) * \left(\frac{(RH_0)_3}{100} \right) \cdot \cdot \cdot \quad (15)$$

Table 1 RH_0 values for single component systems at 25°C

Compound	RH_0	Reference
Potassium chloride	84	(41)
Potassium bromide	81	(41)
Potassium iodide	68	(41)
Sodium chloride	75	(41)
Choline iodide	72	(41)
Choline bromide	41	(41)
Choline chloride	23	(41)
Tetrabutylammonium bromide	61	(41)
Potassium acetate	23	(31)
Potassium carbonate	43	(31)
Sucrose	84	(41)
Fructose	64	(41)
Glucose	87	(41)
Sodium salicylate	79	(39)
Sodium benzoate	88	(39)
Salicylic acid	>99	(53)
Benzoic acid	>99	(53)
Malic acid	78	(53)
Tartaric acid	93	(53)
Fumaric acid	98	(53)
Succinic acid	95	(53)

Table 2 RH_0 values for single component systems at 25°C

Mixture	RH_0 calculated	Experiment
Sodium chloride–potassium bromide	61	64
Potassium chloride–sodium chloride	64	67
Potassium chloride–potassium bromide	68	73
Sucrose–potassium bromide	68	66
Sucrose–dextrose monohydrate	69	68
Sucrose–sodium chloride–potassium bromide	51	57
Choline bromide–potassium bromide	33	40
Tetrabutylammonium bromide–potassium bromide	49	57
Tetrabutylammonium bromide–choline bromide	25	34

where $(RH_0)_{\text{mix}}$ is the relative humidity above a saturated solution of the mixture and $(RH_0)_i$ represents the relative humidities of the individual saturated salt solutions. The Ross equation was derived assuming dilute solutions and negligible interaction between the components in solution. The results presented in Table 2 compare RH_0 values obtained by calculating RH_0 values for mixtures from the Ross equation and those obtained experimentally. Agreement is very good, especially considering the high levels of dissolved solute(s) that are attained (i.e., estimated as high as 50 molal for the choline bromide/tetrabutylammonium bromide system) (43).

The Kinetics of Deliquescence above RH_0

Initial work by Edgar and Swan (55), Adams and Merz (56), Prideaux (57), Morkowitz and Boryta (58), and Carstensen (1) suggested that the rate of moisture uptake onto water-soluble solids above RH_0 should depend on the difference between the partial pressure of water in the environment and that of the partial pressure of water above a saturated solution of the water-soluble substance, the temperature, the exposed surface area of the solid, the velocity of movement of the moist air, and a specific reaction constant that is characteristic of the individual solid.

Van Campen et al. (41) developed models describing the rate of moisture uptake above RH_0 that consider both the mass transport of water to the solid substance and the heat transfer away from the surface. For the special case of an environment consisting of pure water (i.e., initial vacuum conditions), the Van Campen et al. model is greatly simplified because vapor diffusion need not be considered. Here, only the rate at which heat is transported away from the surface is assumed to

be an important factor in limiting the sorption rate, W' . For this special case, an expression was derived to express the rate of moisture uptake solely as a function of RH_1 , the relative humidity of the environment, and RH_0 .

This model was shown to be applicable for describing moisture uptake kinetics (in vacuum) above RH_0 for single component systems of alkali halides, sugars, and choline salts (41). The model later was extended to consider the moisture uptake kinetics above RH_0 for multicomponent systems of these substances (43).

WATER SORPTION BY AMORPHOUS SOLIDS

Isotherm Analyses at Ambient Temperatures

The amount of moisture sorbed by amorphous solids is typically much greater than that sorbed by nonhydrating crystalline substances below their critical relative humidities. Typical substances of pharmaceutical interest in this class of solids include celluloses, starches, poly(vinylpyrrolidone), gelatin, and some lyophilized proteins. Although some of these substances exhibit partially crystalline character, they generally contain significant fractions of amorphous material and, thus, fall into this class of substances. A typical isotherm for microcrystalline cellulose is shown in Fig. 4. Note the significant amounts of water that are sorbed over the entire relative humidity range and that both the sorption and desorption isotherms are characterized by the classical sigmoidal shape often observed with the physical adsorption of gases. Also apparent is the hysteresis between the sorption and desorption portions of the isotherm (i.e., the amount of water associated with the solid is greater for the desorption isotherm than the sorption isotherm for a given relative

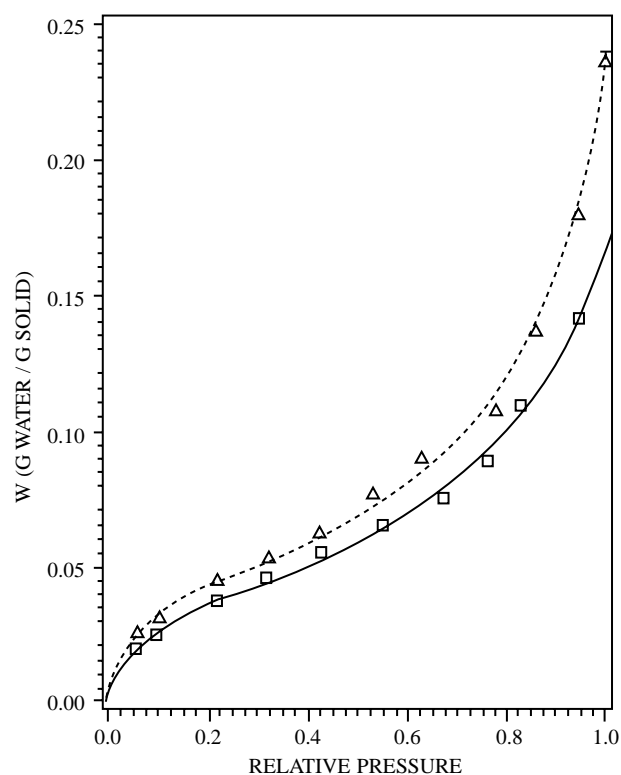


Fig. 4 Water vapor sorption (\square) and desorption (Δ) isotherms for microcrystalline cellulose at 20°C. Solid line: GAB fit to sorption data; Dashed line: GAB fit to desorption data.

pressure). This is typical for these types of materials and is generally attributed to either kinetic effects or to a change in the polymer chain conformation caused by plasticization effects of sorbed water (19, 59–62).

Fig. 4 also shows the excellent fit to the GAB equation (Eq. 10) of the sorption and desorption isotherms for microcrystalline cellulose. In this regard, this equation offers considerable practical utility in fitting isotherms for these types of materials over the entire relative humidity range, especially in contrast to the BET equation, which usually only fits uptake data up to about 40% relative humidity. As previously mentioned, however, this does not in itself confirm the validity of the GAB model for describing moisture sorption data on these materials. Rather, independent confirmation of the physical meaning is necessary.

Considerable physical insight has been gained into the primary binding mechanism of water onto starches and celluloses from isotherm analyses that yield values for W_m (Eqs. 5, 7, and 10). This is illustrated in Table 3, which gives W_m values for three types of starches (61). Table 3 shows that the values for W_m are quite constant despite

Table 3 W_m Values for various starches obtained from BET analysis of moisture uptake isotherms

Starch	W_m (g/g)	Reference
Corn	0.095 ^a	(39)
Corn	0.083	(63)
Potato	0.085	(19)
Wheat	0.080	(19)

^aThis value is taken from the desorption isotherm. Others are from sorption isotherms.

(From Ref. 61.)

significant morphological differences between the various starches. One value of W_m , taken from a fit of desorption data, appears to be slightly higher than values obtained from sorption data. This might be expected if the availability of primary sorption sites had been increased by previous exposure to elevated relative humidities, with subsequent increased levels of water sorption. As shown by Van den Berg et al. (19, 59, 60), these values of W_m are all close to the value of 0.11 g of water per g of starch, calculated by assuming that one water molecule sorbs per anhydroglucose unit. Because this calculation assumes that all anhydroglucose units are available for primary binding, and because this is not likely to be precisely the case, it is not surprising that the values measured for W_m are slightly less than 0.11 g/g.

Zografi et al. (61,64) have extended this analysis to the sorption of water by various celluloses. For celluloses, corrections are necessary because only the amorphous regions of cellulose take up water vapor. Table 4 shows the W_m values obtained from isotherm analyses of several cellulosic materials after accounting for the degree of crystallinity. As expected, celluloses with different degrees of crystallinity exhibit different values of W_m

Table 4 W_m Values for various celluloses obtained from BET analysis of moisture uptake isotherms corrected for degree of crystallinity

Cellulose	Crystallinity %	W_m Corr (g/g)	Reference
Cotton	70	0.093	(65)
Cellophane	40	0.098	(65)
MCC	63	0.095	(66)
MCC ^a	49	0.076	(66)
MCC ^a	38	0.107	(66)
MCC ^a	0	0.086	(66)

^aMCC Ground in a ball mill.

(From Refs. 61 and 66.)

without correction for crystallinity, and all are considerably less than that for the starches. When corrected, however, for the degree of crystallinity, all of the values are in reasonable agreement with each other and with the W_m values obtained for the starches. Especially interesting are the results in Table 4 for microcrystalline cellulose samples having different degrees of crystallinity due to grinding (66). These results suggest that a similar mechanism of water uptake is occurring in starches and the noncrystalline regions of celluloses.

Similar analyses of moisture uptake data available in the literature for other cellulose and starch derivatives used as pharmaceutical excipients are presented in Table 5. Considering the uncertainties associated with the estimated moisture uptake values from published graphs, the values of W_m are all quite consistent with each other and with a stoichiometry of one water molecule per anhydroglucose unit. It is interesting to note that the two samples derived from cellulose, sodium carboxymethylcellulose and sodium croscarmellose, did not require any correction for degree of crystallinity to conform to close to a 1:1 stoichiometry. It appears quite likely, therefore, that the change in chemical structure and the processing of these materials essentially eliminates the crystallinity of cellulose.

The preceding analysis suggests that water, indeed, penetrates throughout the amorphous regions of these materials and undergoes a specific interaction with available sorption sites, most likely the available hydroxyl groups on the anhydroglucose units. Differential heat of sorption results for various starches (19, 63) and cellulose (10,68) support this model. Data gathered indicates that there is a specific water–solid interaction out to a moisture content of at least the equivalent of 3 times W_m . The water present in this system appears to

exist in a more structured state (i.e., reduced mobility) than bulk water over this range. Interestingly, the heats of sorption exhibit discrete breaks corresponding to stoichiometries of one and two water molecules per anhydroglucose unit. Some differential heat of sorption results are nearly constant over the W_m range, suggesting that binding is homogeneous over this range (10, 19). This is, however, not always the case. This is borne out by heat of vaporization data reported by Etzler and Connors on cellulose samples (70). As measured using simultaneous DSC/TGA, the heat of desorption of water from the matrix continuously increases once the moisture content is less than approximately 0.5 g water/g cellulose (about 3 times W_m). These results suggest that the sorbed water exists in multiple “states” that are energetically distinct.

Other supportive evidence for a specific water–solid interaction is available from thermal studies showing the amount of nonfreezeable water (71–73), nuclear magnetic resonance (39, 74–80), and diffusion studies (81, 82). The evidence is less clear, however, concerning whether there is distinct binding of water to sorption sites with discrete energy levels or whether there is a continuum of states where water interacts to a lesser extent with increasing amount sorbed (61, 83). In any event, it is clear that sorbed water behaves with a considerable degree of mobility, and hence, questions the use of the term “bound water” (61, 84).

Isotherm Analyses as a Function of Temperature

Generally speaking, the absorption of water into amorphous solids as a function of relative humidity decreases as the temperature increases, reflective of an overall exothermic process, normally expected with vapor adsorption processes. Such behavior has been observed with cellulose (85), starch (19), poly(vinylpyrrolidone) (13), and poly(methyl methacrylate) (86). In such cases it is often assumed that the dominant factor is the negative heat of absorption arising from the change in the extent of water binding. The process, however, is made much more complex than this because of the changing morphology of the solid and, hence, an entropy change as well. The complexity of the effects of temperature on water vapor absorption and the possible links to the plasticizing effects of water may be observed in the work of Oksanen and Zografi (13), who have reported that the W_m values for poly(vinylpyrrolidone) over the temperature range of -40 – 60°C decrease by a factor of three, suggesting that W_m does not reflect the absolute number of available binding sites on the polymer for directly “bound” water.

Table 5 W_m Values for various pharmaceutical excipients obtained from BET analysis of moisture uptake isotherms

Excipient	W_m (g/g)	Reference
Starch 1500	0.074	(63)
Sodium starch glycolate (Explotab®)	0.081	(67)
Sodium starch glycolate (Primogel®)	0.092	(67)
Crosslinked dextrose (CLD-2)	0.098	(68)
Croscarmellose, sodium (Ac-Di-Sol®)	0.094	(68)
Sodium carboxymethylcellulose	0.103	(69)

(From Ref. 61.)

Rather, W_m appears to be related to $W(T_g = T)$, the amount of water sorbed that will reduce the glass transition temperature, T_g , to the temperature of the sample, as the ratio of $W(T_g = T)/W_m$ remains nearly constant at 3.0 over the entire temperature range.

In summary, it is clear that water absorbs into amorphous polymers to a significant extent. Interaction of water molecules with "available" sorption sites likely occurs via hydrogen bonding such that the mobility of the sorbed water is reduced and the thermodynamic state of water is significantly altered relative to bulk water. Yet accessibility of the water to all potential sorption sites appears to be dependent on the previous history and physicochemical properties of the solid. In this regard, the water–solid interaction in amorphous polymer systems is a dynamic relationship depending quite strongly on water activity and temperature.

The Meaning of Specific Surface Areas Calculated from Water Absorption Studies

Simply calculating specific surface areas from the W_m values in Tables 3–5 leads to "apparent" specific surface areas of approximately 300–500 m²/g (61, 64). Specific surface areas obtained from similar analyses of nonpolar gas (nitrogen or krypton) adsorption studies, however, are typically in the range of 1 m²/g, independent of sample pretreatment.

Interestingly, the ball milling studies of microcrystalline cellulose by Nakai (Table 4) (66) have shown that the W_m values obtained from water sorption studies increase to a much greater extent than the increase in surface area because of comminution of the sample. In fact, as discussed earlier, moisture sorption was shown to be proportional to the amount of amorphous character, suggesting that water is absorbed throughout the amorphous regions of this substance. In this regard, artifactual specific surface areas are obtained if calculated from water absorption data (64) for these types of substances.

The Role of Water as a Plasticizer

Absorption of significant amounts of water into the internal structure of a solid has been shown to influence the properties of the solid. This is apparent, for example, in the hysteresis observed between the sorption and desorption isotherms in Fig. 4. This phenomenon becomes exaggerated to a greater extent for materials that consist of higher proportions of amorphous material. Levine and Slade

(14, 87) have demonstrated that water, with a very low glass transition temperature, can act as a plasticizer, thereby lowering the glass transition temperature, T_g , of amorphous polymers. Similar behavior is observed in amorphous low molecular weight solids. Recognizing that the viscoelastic properties of the solid are altered significantly above T_g (rubbery state) relative to below T_g (glass or vitreous state), it is likely that the solid will undergo changes of its physical properties at distinct moisture contents and defined temperatures as a result of this phenomenon (13, 14). Oksanen and Zografi (13) have shown with poly(vinylpyrrolidone) that the moisture content at which the moisture sorption isotherm begins to increase significantly correlates very well with the moisture content that will reduce T_g to the temperature of the isotherm. This is illustrated in Fig. 5, which shows water absorption isotherms for poly(vinylpyrrolidone) over the temperature range of -40 to 60°C (13, 88). Clearly, the inflection point at which the isotherm begins

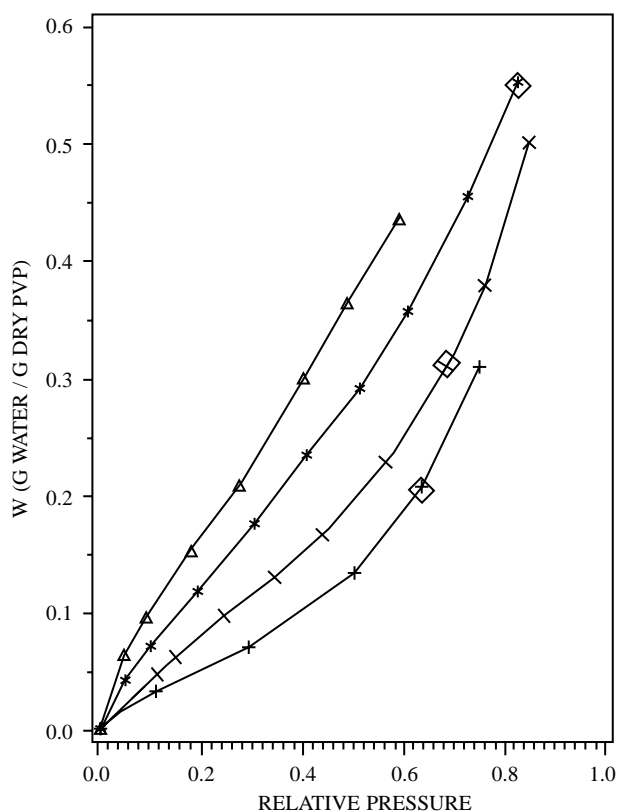


Fig. 5 Water vapor sorption isotherms for poly(vinylpyrrolidone) at 60°C (+); 30°C (x); -20°C (*); -40°C (Δ). Data were taken from Oksanen and Zografi (76, 78). A ◇ represents the calculated water contents necessary to depress T_g to the temperature of the isotherms.

to turn markedly upward shifts to a higher moisture content as the temperature is reduced. To illustrate this, note that the moisture content (0.674 g/g) necessary to reduce T_g to -40°C has not been attained yet in Fig. 5 and the isotherm appears quite linear over the relative humidity range shown. For further clarity, the moisture contents [$W(T_g = T)$] corresponding to T_g at 60, 30, -20°C , and -40°C were shown to be about 0.205, 0.313, 0.553, and 0.674 g/g, respectively. Oksanen and Zografi (13) reported that cellulose and elastin (a protein) exhibit similar relationships, where the glass to rubber transitions correspond to the upward inflections in their respective isotherms.

Subsequent studies by Hancock and Zografi (89) demonstrated that the glass transition temperatures for PVP, hydroxypropyl methylcellulose, and poly(methyl methacrylate) were linearly depressed by the weight fraction of sorbed water, according to the simplified Gordon–Taylor/Kelly–Bueche equation (90):

$$\frac{1}{T_{g(\text{mix})}} = \left(\frac{W_1}{T_{g1}} + \frac{W_2}{T_{g2}} \right) \quad (15)$$

where W_1 and W_2 are the weight fractions of components with T_g values T_{g1} and T_{g2} . The results clearly indicate that the viscoelastic properties of amorphous materials can undergo significant changes as the solid transitions from the glassy to rubbery state. Furthermore, these changes can occur due to elevation of temperature at fixed moisture content or to an increase in moisture content at constant temperature or to a combination of these effects. For dry amorphous substances, molecular mobility of the solid begins to be significantly enhanced relative to the glassy state as low as 50°C below T_g (91). Similar increases in molecular mobility due to the plasticizing effects of absorbed water suggest the need to maintain amorphous systems at least 50°C below the system glass transition temperature to avoid physical, chemical, and/or mechanical property changes over the product shelf life. Some properties that are likely to be affected include tablet compaction (92, 93), gelatin capsule brittleness (14, 94), collapse of lyophilized amorphous powders (88, 95, 96) protein stability (97, 98), and the stability of low molecular weight, moisture sensitive drugs mixed with amorphous polymeric substances (99).

WATER SORPTION BY PHARMACEUTICAL SOLIDS SUBJECTED TO PROCESSING

Understanding the mechanisms of moisture sorption by solids existing in either the crystalline or amorphous

states allows a conceptual estimation of critical points where major changes in physical or chemical properties occur (e.g., RH_0 , a crystal hydration relative humidity, glass transition temperature). Processing (i.e., milling, spray drying, compaction, lyophilization, etc.) of pharmaceutical solids, however, often induces at least partial conversion of most substances to a high energy form (100–107). Such local disorder has been associated with enhanced chemical reactivity (101–108) and increased solubility (15) relative to the thermodynamically favored crystalline state. These regions have been referred to as “hot” spots of the bulk solid and, when present, leave the solid in an “activated state” (15, 100–107, 109, 110–113).

This nonhomogeneity that exists in processed solids complicates the study of moisture sorption phenomena in these materials, as more than one mechanism of uptake must be considered. This is especially difficult, and often frustrating, for cases in which only a small amount of amorphous material is present, as the experimental techniques required to complete these analyses are labor intensive (114, 115). Yet, relatively low percentages of amorphous material can absorb considerable amounts of water into their structure, with these regions undergoing considerable change and a consequent effect on the overall properties of the bulk substance (100). This is especially important for low molecular weight substances that have the ability to readily recrystallize due to their overall greater mobility relative to higher molecular weight polymeric materials. This has been demonstrated for sodium chloride and sodium salicylate ground for 15 min in a mortar and pestle (48). Whereas recrystallized materials exhibited no change in specific surface areas with increasing relative humidities, the ground samples exhibited significant reductions in specific surface areas as relative humidities were increased. Figure 2 illustrates the differing moisture uptake profiles for the recrystallized and ground sodium chloride samples, normalized for specific surface area (48). Whereas the ground material sorbed significantly more water at lower relative humidities than the recrystallized sample, the recrystallized material sorbed greater amounts at higher relative humidities. This relative reduction in sorption capacity of the ground sample is attributed to a reduction in surface area as relative humidity increased, due to the consequent recrystallization of the disordered surface material (48). Fukuoka et al. (116) have demonstrated that a variety of pharmaceutical substances indeed can be made amorphous and, furthermore, exhibit glass transition temperatures over a range from 243 to 354 K. For example, aspirin, progesterone, phenobarbital, and sulfadimethoxine exhibit T_g values of 243, 279, 321, and 339 K, respectively.

Similar to amorphous polymeric systems, low molecular weight amorphous substances also exhibit a reduction in T_g as moisture content increases (117), thereby leading to favorable conditions for recrystallization to occur. Indeed, low molecular weight amorphous solids possess sufficient molecular mobility well below T_g (118). In some systems with multiple solid states possible, the water activity/moisture content can influence the crystallized form of the solid (119). Unfortunately, recrystallization of nonhydrating, low molecular weight amorphous systems can lead to the liberation of significant amounts of water to the headspace (15, 48, 103, 104, 120). Such "moisture dumping" can have additional impact on the physical, chemical, and mechanical properties of the system (17, 48, 100).

To illustrate this more quantitatively, consider the hypothetical sucrose example discussed by Ahlneck and Zografi (100). Assuming that all the sorbed water is taken up by the amorphous portion of material, 0.1% total moisture would correspond to approximately 20%, 10%, 4%, and 2% moisture content in the amorphous material, respectively, for 0.5%, 1%, 2.5%, and 5% of amorphous solid. The glass transition temperatures for the amorphous portions of these systems range from 9 to 49°C, respectively (87,100). Hence, significant changes in the solid state properties are expected at room temperature if relatively small amounts of amorphous material (i.e., < 1%) are initially present. This example illustrates that even for low moisture content materials, significant changes can occur in localized amorphous regions of a solid, which may affect properties of the material influenced by molecular mobility (100).

Inhibition of events resulting from increased molecular mobility due to increased moisture absorption and a subsequent reduction in T_g can be accomplished by formulating such materials with amorphous substances of higher T_g . The net effect is to increase molecular interaction and raise the system T_g to a level where molecular mobility is again sufficiently low (high viscosity) such that the undesired property changes do not occur (107, 121, 122).

The recrystallization of amorphous low molecular weight systems can be convoluted by the impact of structural changes on the material. For example, spray drying α -lactose monohydrate typically produces a material that is completely amorphous as determined by powder X-ray diffraction. However, lactose undergoes anomeric rotation in solution, causing a change in the fundamental structure of the molecule. The impact of this structural change on the uptake and equilibration of water to the amorphous lactose is significant. As expected, an

increase in moisture content will result in the suppression of the glass transition temperature to the point where instantaneous crystallization occurs. However, the material produced is nonuniform as there are two different forms of lactose present, one of which is anhydrous in the crystalline state. Furthermore, at elevated humidity, the crystalline β -anhydrate can undergo structural change in the solid state and produce the α -monohydrate. Evaluation of data gathered needs to be completed in the context of a careful characterization of the possible solid forms (16, 17, 123–125).

TRANSFER OF WATER BETWEEN SOLID COMPONENTS VIA THE HEADSPACE

Combining solids that have previously been equilibrated at different relative humidities results in a system that is thermodynamically unstable because there will be a tendency for moisture to distribute in the system so that a single relative humidity is attained in the headspace. As shown in Fig. 6, moisture will desorb into the headspace from the component initially equilibrated at a higher relative humidity and sorb to the component initially equilibrated at a lower relative humidity. This process will continue until both solids have equilibrated at the final relative humidity. The final relative humidity can be predicted a priori by the sorption–desorption moisture transfer model (SDMT) model (126) if one has moisture uptake isotherms for each of the solid components, their initial moisture contents and dry weights, headspace volume, and temperature. Final moisture contents for each solid can then easily be estimated from the isotherms for the respective solids.

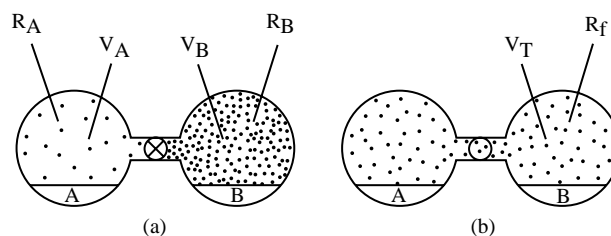


Fig. 6 Schematic representation of moisture transfer between solid components, A and B, respectively. (a) Headspaces isolated from one another; (b) Headspaces allowed to equilibrate. R_A and R_B are initial relative humidities above A and B, respectively; V_A and V_B are headspace volumes above A and B, respectively; R_f and V_T are final relative humidity and headspace volume above A and B, respectively. (From Ref. 126.)

The SDMT model has practical utility in aiding the rational optimization of the initial moisture contents of individual components in a system to attain the final desired relative humidity. Practical applications to date have included adjustment of the initial formulation LODs prior to capsule filling to avoid gelatin capsule brittleness (94, 127), selecting the appropriate formulation moisture content and amount of desiccant to maintain the relative humidity inside a container below a defined value (128), and selection of appropriate dry powder inhaler design and packaging conditions for optimal stability (129).

SUMMARY

Moisture is present in all solid pharmaceutical drugs and dosage forms and in most processing techniques. Understanding where the water resides, its state, and the manner in which it affects the properties of individual materials, their mixtures, and ultimately, final product performance and integrity are essential for the developmental scientist to better understand the role of water in a particular system. Especially important are the kinetics of moisture uptake or loss, "equilibrium" uptake values as a function of relative humidity, whether the water resides externally or is absorbed into the material, its degree of binding with the solid, and the tendency for water to redistribute in a system consisting of more than one solid. Although water-solid interaction(s) can be extremely complex in pharmaceutical systems, application of these fundamental concepts to product development can greatly aid in understanding the role of moisture in affecting the physicochemical properties of solid materials.

ACKNOWLEDGMENTS

The authors thank Professor George Zografi, Dr. Cindy Oksanen, and Dr. Frank Etzler for their technical contributions to this chapter and Ms. Linda Schweikardt for her graphical support.

REFERENCES

- Carstensen, J.T. *Pharmaceutics of Solids and Solid Dosage Forms*; John Wiley & Sons: New York, 1977; 11–15.
- Lewis, G.N.; Randall, M.; Pitzer, K.S.; Brewer, L. *Thermodynamics*; McGraw-Hill: New York, 1961.
- Gal, S. *Vapor Sorption Equilibria and Other Water-Starch Interactions: A Physico-Chemical Approach*; Van den Berg, C., Ed.; Agricultural University: Wageningen, The Netherlands, 1981; 10.
- Brunauer, S.; Emmett, P.H.; Teller, E. J. Am. Chem. Soc. **1938**, *60*, 309.
- Langmuir, I. J. Am. Chem. Soc. **1918**, *40*, 1361.
- Fowler, R.; Guggenheim, E.A. *Statistical Thermodynamics*; Cambridge University Press: Cambridge, UK, 1939.
- Hill, T.L. J. Chem. Phys. **1946**, *14*, 263.
- J. Am. Chem. Soc. **1946**, *68*, 535.
- Cassie, A.B.D. Trans. Faraday Soc. **1945**, *41*, 450.
- Trans. Faraday Soc. **1945**, *41*, 458.
- Lowell, S. *Introduction to Powder Surface Area*; John Wiley & Sons: New York, 1979.
- Hollenback, R.G.; Peck, G.E.; Kildsig, D.O. J. Pharm. Sci. **1978**, *67*, 1599.
- Guggenheim, E.A. *Applications of Statistical Mechanics*; Clarendon Press: Oxford, 1966.
- deBoer, J.H. *The Dynamical Character of Adsorption*, 2nd Ed.; Clarendon Press: Oxford, 1968.
- Oksanen, C.A.; Zografi, G. Pharm. Res. **1990**, *7*, 654.
- Levine, H.; Slade, L. *Water Science Reviews*; Franks, F., Ed.; Cambridge University Press: Cambridge, UK, 1987; 3, 79–185.
- Makower, B.; Dye, W.B. Agr. Fd. Chem. **1956**, *4*, 72.
- Buckton, G.; Darcy, P. Int. J. Pharm. **1996**, *136*, 141.
- Buckton, G.; Darcy, P. Int. J. Pharm. **1997**, *158*, 157.
- Anderson, R.B. J. Am. Chem. Soc. **1946**, *68*, 686.
- Van den Berg, C. *Vapor Sorption Equilibrium and Other Water-Starch Interactions: A Physico-Chemical Approach*; Agricultural University: Wageningen, The Netherlands, 1981.
- Flory, P.J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, New York, 1953.
- Hailwood, A.J.; Horrobin, S. Trans. Faraday Soc. **1946**, *42B*, 84.
- Vrentas, J.S.; Vrentas, C.M. Macromolecules. **1994**, *27*, 4684.
- Vrentas, J.S.; Vrentas, C.M. Macromolecules. **1994**, *27*, 5570.
- Hancock, B.C.; Zografi, G. Pharm. Res. **1993**, *10*, 1262.
- Van den Berg, C.; Bruin, S. *Water Activity: Influences on Food Quality*; Rockland, L.B., Stewart, G.F., Eds.; Academic Press: New York, 1981.
- Venkateswaren, A. Chem. Rev. **1970**, *70*, 619.
- Martin, A.; Swarbrick, J.; Cammarata, A. *Physical Pharmacy*; Lea & Febiger: Philadelphia, 1983; 510–512.
- Fisher, L.R.; Israelachvili, J.N. Chem. Phys. Lett. **1980**, *76*, 325.
- Zografi, G. Drug Devel. Ind. Pharm. **1988**, *14*, 1905.
- Etzler, F.M.; Fagundus, D.M. J. Colloid Interface Sci. **1983**, *93*, 585.
- Greenspan, L. J. Res. N.B.S. **1977**, *81A*, 89.
- Rockland, L.B.; Nishi, S.K. J. Food Technol. **1980**, *34*, 43.

33. Winston, P.W.; Bates, D.H. *Ecology* **1960**, *41*, 232.
34. Stokes, R.H.; Robinson, R.A. *Ind. Eng. Chem.* **1949**, *41*, 2013.
35. Weast, R.C., Ed. *Handbook of Chemistry and Physics*, 6th Ed.; Chemical Rubber Co.: Cleveland, 1986–1987; E-42.
36. Ward, G.H.; Schultz, R.K. *Pharm. Res.* **1995**, *12*, 773.
37. Bergren, M. *Water–Solid Interactions*; AAPS Short-course: Orlando, FL, 1993.
38. Spancake, C.W.; Hastedt, J.E.; Venero, A.F. *Pharm. Res.* **1993**, *10*, S-280.
39. Kontny, M.J. *Water Vapor Sorption Studies on Solid Surfaces*; University of Wisconsin-Madison: Madison, WI, 1985.
40. Van Campen, L.; Zografi, G.; Carstensen, J.T. *Int. J. Pharm.* **1980**, *5*, 1.
41. Van Campen, L.; Amidon, G.L.; Zografi, G. *J. Pharm. Sci.* **1983**, *72*, 1381, 1388, 1394.
42. *Handbook of Chemistry and Physics*; Weast, R.C., Ed.; Chemical Rubber Co.: Cleveland, 1986–1987, D-189–190.
43. Kontny, M.J.; Zografi, G. *J. Pharm. Sci.* **1985**, *74*, 124.
44. Bergren, M. *Int. J. Pharm.* **1994**, *103*, 103.
45. Barraclough, P.B.; Hall, P.G. *Surf. Sci.* **1974**, *467*, 393.
46. Walter, H.U. *Zeitschrift fur Physikalische Chemie Neue Folge Bd.* **1971**, *75*, S287.
47. Ladd, R.A. *Surf. Sci.* **1968**, *12*, 37.
48. Kontny, M.J.; Grandolfi, G.P.; Zografi, G. *Pharm. Res.* **1987**, *4*, 104.
49. Byrn, S.R. *Solid State Chemistry of Drugs*; Academic Press: New York, 1982; 149.
50. Saleki-Gerhardt, A.; Stowell, J.G.; Byrn, S.R.; Zografi, G. *J. Pharm. Sci.* **1995**, *84*, 318.
51. Connors, J.J. Unpublished data.
52. Cox, J.S.G.; Woodard, G.D.; McCrone, W.C. *J. Pharm. Sci.* **1971**, *60*, 1458.
53. Kontny, M.J. Unpublished data.
54. Ross, K.D. *Food Technol.* **1975**, *29*, 26.
55. Edgar, G.; Swan, W.O. *J. Am. Chem. Soc.* **1922**, *44*, 570.
56. Adams, J.R.; Merz, A.R. *Ind. Eng. Chem.* **1929**, *21*, 305.
57. Prideaux, E.B.R. *J. Soc. Chem. Ind.* **1920**, *39*, 182.
58. Markowitz, M.M.; Boryta, D.A. *J. Chem. Eng. Data* **1961**, *6*, 16.
59. Van den Berg, C.; Kaper, F.S.; Weldring, J.A.G.; Wolters, I. *J. Food Technol.* **1975**, *10*, 589.
60. Van den Berg, C. *Water Activity: Influences on Food Quality*; Rockland, L.B., Stewart, G.F., Eds.; Academic Press: New York, 1981; 1–61.
61. Zografi, G.; Kontny, M.J. *Pharm. Res.* **1986**, *3*, 187.
62. Hancock, B.C.; Zografi, G. *Pharm. Res.* **1994**, *11*, 471.
63. Wurster, D.E.; Peck, G.E.; Kildsig, D.O. *Starch* **1984**, *36*, 294.
64. Zografi, G.; Kontny, M.J.; Yang, A.Y.S.; Brenner, G.S. *Int. J. Pharm.* **1984**, *18*, 99.
65. Stamm, A.J. *Wood and Cellulose Science*; The Ronald Press Co.: New York, 1964.
66. Nakai, Y.; Fukuoka, E.; Nakajima, S.; Hasegawa, J. *J. Chem. Pharm. Bull.* **1977**, *25*, 96.
67. Mitrevej, A.; Hollenback, R.G. *Pharm. Tech.* **1982**, *6*, 48.
68. Gordon, R.E.; Peck, G.E.; Kildsig, D.O. *Drug Devel. Ind. Pharm.* **1984**, *10*, 833.
69. Callahan, J.C.; Cleary, G.W.; Elefant, M.; Kaplan, I.; Kensler, T.; Nash, R.A. *Drug Devel. Ind. Pharm.* **1982**, *8*, 355.
70. Etzler, F.M.; Connors, J.J. *Therm. Acta* **1991**, *189*, 185.
71. Rupley, J.A.; Yank, P.-H.; Tollin, G. *Water in Polymers*; Rowland, S.P., Ed.; American Chemical Society: Washington, DC, 1980; 111–132.
72. Nagashima, N.; Suzuki, E.-I. *Appl. Spectroscopy Rev.* **1984**, *20*, 1.
73. Duckworth, R.B. *J. Food Technol.* **1971**, *6*, 317.
74. Hennig, H.J.; Lechert, H. *J. Colloid Interf. Sci.* **1977**, *62*, 199.
75. Mousseri, J.; Steinberg, M.P.; Nelson, A.I.; Wei, L.S. *J. Food Sci.* **1974**, *39*, 114.
76. Tait, M.J.; Ablett, S.; Wood, F.W. *J. Colloid Interf. Sci.* **1972**, *41*, 594.
77. Tait, M.J.; Ablett, S.; Franks, F. *Water Structure at the Water–Polymer Interface*; Jellinek, H., Ed.; Plenum Press: New York, 1972; 29–38.
78. Carles, J.E.; Scallan, A.M. *J. Appl. Poly. Sci.* **1973**, *17*, 1855.
79. Hsi, E.; Voight, G.J.; Bryant, R.G. *J. Colloid Interf. Sci.* **1979**, *70*, 338.
80. Froix, M.F.; Nelson, R. *Macromolecules*. **1975**, *8*, 726.
81. Fish, B.P. *Fundamental Aspects of the Dehydration of Foodstuffs*; Soc. Chem. Ind. (S.C.I.): London, 1958; 143–157.
82. Duckworth, R.B.; Smith, G.M. *Recent Advances in Food Science*; Leitch, J.M., Rhodes, D.N., Eds.; Butterworths: London, 1962; 3, 230–238.
83. Etzler, F.M. *J. Colloid Interf. Sci.* **1983**, *92*, 43.
84. Oksanen, C.A.; Zografi, G. *Pharm. Res.* **1993**, *10*, 791.
85. Urquart, A.R.; Williams, A.M. *J. Textile Inst.* **1924**, *15*, 550.
86. Smith, L.S.A.; Schnitz, V. *Polymer*. **1988**, *29*, 1871.
87. Slade, L.; Levine, H. *Pure Appl. Chem.* **1988**, *60*, 1841.
88. Mackenzie, A.P.; Rasmussen, D.H. *Water Structure at the Water–Polymer Interface*; Jellinek, H.H.G., Ed.; Plenum: New York, 1972; 146–172.
89. Hancock, B.C.; Zografi, G. *Pharm. Res.* **1994**, *11*, 471.
90. Gordon, M.; Taylor, J.S. *J. Appl. Chem.* **1952**, *2*, 493.
91. Hancock, B.C.; Shamblin, S.L.; Zografi, G. *Pharm. Res.* **1995**, *12*, 799.
92. Shukla, A.J.; Price, J.C. *Pharm. Res.* **1991**, *8*, 336.
93. Amidon, G.E.; Houghton, M.E. *Pharm. Res.* **1995**, *12*, 923.
94. Kontny, M.J.; Mulski, C.A. *Int. J. Pharm.* **1989**, *54*, 79.
95. Mackenzie, A.P. *Freeze Drying and Advanced Food Technology*; Goldblith, S.A., Rey, L., Rothmayr, W., Eds.; Academic Press: New York, 1975; 277–307.

96. Costantino, H.R.; Curley, J.G.; Wu, S.; Hsu, C.C. *Int. J. Pharm.* **1998**, *166*, 211.
97. Hageman, M.J. *Drug Devel. Ind. Pharm.* **1988**, *14*, 2047.
98. Hageman, M.J.; Possert, P.L.; Bauer, J.M. *J. Agr. Food Chem.* **1992**, *40*, 342.
99. Kararli, T.T.; Catalano, T. *Pharm. Res.* **1995**, *12*, 923.
100. Ahlneck, C.; Zografi, G. *Int. J. Pharm.* **1990**, *62*, 87.
101. Vadas, E.B.; Toma, P.; Zografi, G. *Pharm. Res.* **1991**, *8*, 148.
102. Levine, H.; Slade, L. *J. Chem. Soc. Faraday Trans. I* **1988**, *84*, 2619.
103. Huttenrauch, R. *Acta Pharm. Technol.* **1988**, *34*, 1.
104. Huttenrauch, R.; Frike, S.; Zeielke, P. *Pharm. Res.* **1985**, *2*, 302.
105. Otsuka, M.; Kaneniwa, N. *Int. J. Pharm.* **1990**, *62*, 65.
106. Hersey, J.A.; Krycer, I. *Int. J. Pharm. Technol. & Prod. Manuf.* **1980**, *1*, 18.
107. Hancock, B.C.; Zografi, G. *J. Pharm. Sci.* **1996**, *85*, 246.
108. Shalaev, E.Y.; Zografi, G. *J. Pharm. Sci.* **1996**, *85*, 1137.
109. Carstensen, J.T.; Van Scoik, K. *Pharm. Res.* **1990**, *7*, 1278.
110. Waltersson, J.-O.; Lundgren, P. *Acta Pharm. Suec.* **1985**, *22*, 291.
111. Prout, E.G.; Tompkins, F.C. *Trans. Faraday Soc.* **1944**, *40*, 489.
112. Ng, W.-L. *Aust. J. Chem.* **1975**, *28*, 1169.
113. Hasegawa, J.; Hanano, M.; Awazu, S. *Chem. Pharm. Bull.* **1975**, *23*, 86.
114. Saleki-Gerhardt, A. *Estimation of Percent Crystallinity in Milled Samples of Sucrose*; University of Wisconsin-Madison: Madison, WI, 1991.
115. Buckton, G.; Darcy, P. *Int. J. Pharm.* **1999**, *179*, 141.
116. Fukuoka, E.; Makita, M.; Yamamura, S. *Chem. Pharm. Bull.* **1989**, *37*, 1047.
117. Saleki-Gerhardt, A.; Zografi, G. *Pharm. Res.* **1994**, *11*, 1166.
118. Andronis, V.; Zografi, G. *Pharm. Res.* **1998**, *15*, 835.
119. Andronis, V.; Yoshioka, M.; Zografi, G. *J. Pharm. Sci.* **1997**, *86*, 346.
120. Lehto, V.-P.; Laine, E. *Pharm. Res.* **2000**, *17*, 701.
121. Yoshioka, M.; Hancock, B.C.; Zografi, G. *J. Pharm. Sci.* **1995**, *84*, 983.
122. Shamblin, S.L.; Zografi, G. *Pharm. Res.* **1999**, *16*, 1119.
123. Angberg, M.; Nystrom, C.; Castensson, S. *Int. J. Pharm.* **1991**, *73*, 209.
124. Briggner, L.-E.; Buckton, G.; Bystrom, K.; Darcy, P. *Int. J. Pharm.* **1994**, *105*, 125.
125. Schmitt, E.A.; Law, D.; Zhang, G.G.Z. *J. Pharm. Sci.* **1999**, *88*, 291.
126. Zografi, G.; Grandolfi, G.P.; Kontny, M.J.; Mendenhall, D.W. *Int. J. Pharm.* **1988**, *42*, 77.
127. Kontny, M.J. *Drug Devel. Ind. Pharm.* **1988**, *14*, 1991.
128. Kontny, M.J.; Koppenol, S.; Graham, E.T. *Int. J. Pharm.* **1992**, *84*, 261.
129. Kontny, M.J.; Connors, J.J.; Graham, E.T. *Proceedings of Respiratory Drug Delivery IV*; Interpharm Press: Englewood, CO, 1994.

WAXES

Roland A. Bodmeier

Freie Universität Berlin, Berlin, Germany

INTRODUCTION

The term wax generally refers to a substance that is a plastic solid at room temperature and a liquid of low viscosity above its melting point. Strictly speaking, a wax is chemically defined as an ester of a monohydric long chain fatty alcohol and a long chain fatty acid. However, generally the term wax has been applied to a broad group of chemically heterogeneous materials. Waxes usually contain a wide variety of materials including glycerides, fatty alcohols, fatty acids, and their esters. In the pharmaceutical literature, the terms waxes, fats, or lipids have often been used interchangeably and no consistent terminology has been established. They have in common their lipophilic character and their insolubility in water and solubility in nonpolar solvents. Besides natural materials, many semisynthetic products such as fatty acids or alcohols or surfactants are derived from lipids.

Waxes have been used by the pharmaceutical industry for many years. Their applications in semisolid preparations, including ointments, creams, or lotions, and in suppositories are well known and numerous publications exist on this topic. Because of their lipophilic properties, waxes have been used in sustained-release single or multiple unit solid dosage forms. This article reviews the different uses of waxes as sustained-release carrier or coating materials.

WAXES IN PHARMACEUTICAL DOSAGE FORMS

Waxes are obtained from various sources and are generally classified into animal, insect, vegetable, mineral, and synthetic waxes (1–7).

The most familiar animal wax is probably lanolin, which is obtained from the wool of the sheep. It consists primarily of esters of C_{18} – C_{26} alcohols and fatty acids, sterols (cholesterol), and terpene alcohols. It is frequently used in topical preparations. Until recently, spermaceti was another commonly used animal wax. Spermaceti is obtained through the precipitation of the head oil from the sperm whale on cooling. It consists primarily of cetyl palmitate. Because of public concerns with animal-derived

products, spermaceti has been replaced with other natural or synthetic products.

The most commonly used insect wax is beeswax. It is obtained from the honeycomb of the bee. White and yellow beeswax are GRAS-listed and consist of mixtures of various esters of straight chain monohydric alcohols with even number carbon chains (C_{24} – C_{36}) esterified with straight chain fatty acids. The major ester is myricyl palmitate. Beeswax also contains free acids and carbohydrates. White wax is obtained through bleaching of yellow wax with oxidizing agents or with sunlight. The National Formulary 18 (NF18) (8) specifications list a melting range of 62–65°C, an acid value of 17–24, and an ester value of 72–79. It is practically insoluble in water, sparingly soluble in ethanol, and soluble in chloroform and various oils. Beeswax is used as a stiffening agent in topical preparations, as a stabilizer of w/o-emulsions, and as a polishing agent in sugar coating.

Carnauba wax is plant-derived and is obtained from the carnauba palm tree, indigenous to Brazil. The wax is obtained from the surface of dried leaves. It is widely used in food, cosmetic, and pharmaceutical products. It consists of a complex mixture of high-molecular-weight esters of acids and hydroxyacids. Carnauba wax is very hard and brittle, and has a high melting point. The NF18 specifications list a melting range of 81–86°C, an acid value of 2–7, and a saponification value of 78–95. It is insoluble in water, slightly soluble in boiling ethanol, and soluble in warm chloroform. Besides the sustained-release applications described later, it is used as a polishing agent in sugar coating because of its high gloss, and in topical preparations. Other, less used vegetable-derived waxes include candelilla wax and castor wax.

Hydrogenated vegetable oils are prepared by hydrogenation of refined vegetable oils. Hydrogenated vegetable oil consists of mixtures of triglycerides, with two types being defined in the USP23. Type II includes partially hydrogenated vegetable oils and has a lower melting range and a higher iodine value than Type I. Type I melts in the range of 57–70°C and has iodine value of 0–5, while Type II has a melting range of 20–50°C and an iodine value of 55–80. They are used as lubricants, as sustained-release matrix materials, as viscosity modifiers in semisolid formulations, to enhance

the solidification of suppositories, and to minimize the sedimentation of dispersed drug.

Two commonly used mineral-derived waxes are petroleum wax, which is microcrystalline, and paraffin wax, which is crystalline. They are both obtained from petroleum: the quality and quantity of the wax depends on the source of the crude oil and the refining process. Microcrystalline wax (petroleum ceresin or wax) consists of straight chain and branched saturated alkanes with a chain length range C_{41} – C_{57} . The NF18 specifications list a melting range of 54–102°C; it comes in plastic and hard grades. It is insoluble in water, slightly soluble in ethanol, and soluble in chloroform. Besides its use as a sustained-release carrier, it is used as a stiffening agent in topical preparations. Because of its high viscosity and melting point, it increases the consistency of creams and ointments. Paraffin wax (hard paraffin) is a mixture of solid straight chain alkanes. It is used in ointments or creams as a base or stiffening agent. It congeals between 47 and 65°C. Various grades with different melting ranges are available. It is insoluble in acetone, ethanol, and water and soluble in chloroform and most warmed fixed oils. Low-molecular-weight polyethylenes ($MW < 10,000$) have wax-like properties and are used in topical preparations, for example, as gelling agents in Plastibase.

CHARACTERIZATION

Because the harvesting of vegetable or insect waxes is often from wild, noncultivated sources and because of their complex composition, it is important to characterize the chemical and physical properties of the waxes (1–6). The composition of natural materials often varies with location, weather, season of harvesting, and age. A good quality control of the raw materials is of utmost importance in order to obtain pharmaceutical products of high quality.

The chemical methods to characterize waxes include the determination of the acid, saponification, iodine, hydroxyl, and peroxide values. Various tests, often yielding different values, are available to measure the melting point of waxes. Since waxes are nonhomogeneous in chemical composition, a melting range rather than a clear melting point is most observed. The melting point of glycerides generally increases with increasing hydroxyl number, decreasing degree of unsaturation, and increasing molecular weight of the fatty acid. The melting point of many waxes can be determined with capillary tubes. The slip point is defined as the temperature at which a column of the testing material starts raising in an open-ended capillary tube, which is dipped in water filled in a beaker

and heated under specific conditions. The drop-point test can be used; however, it is not reliable for more viscous waxes. The congealing point of a wax is the temperature at which the molten wax stops to flow upon cooling. Thermal methods such as differential scanning calorimetry (DSC) are widely used to characterize the heating and cooling profiles of waxes in a qualitative and quantitative manner. Potential polymorphic transitions and recrystallization during processing can be simulated by running different temperature profiles.

The contraction of suppository bases during cooling within the mold is a well-described phenomenon. The expansion or contraction of waxes is also important during the processing of wax melts, for example, during the preparation of microparticles by spray congealing, hot-melt coating, or hot-melt filling of hard gelatin capsules. The dilatation of waxes or thermal expansion during the transition from the solid to the liquid state can be measured with a dilatometer. The hardness of a wax is measured with a penetration test, whereby the depth of penetration of a needle under a given weight is measured, preferably at different temperatures. The viscosity of the molten wax is an important parameter, especially for processes such as hot-melt coating or spray congealing, where wax melts are processed. In an ASTM monograph (D 88), the time that a certain quantity of molten wax requires to flow through an orifice of specified dimensions is measured.

The color of the wax will affect the color of the finished product. A Lovibond Tintometer is often used for color measurements, whereby the color of the raw material is compared against a series of colored standard glasses, under a standard light source. The color of the solidified wax of the same sample may be different depending on the amount of occluded air, the rate of cooling, or surface finish. Therefore, the color of many waxes is best measured in the molten state. Two ASTM color standards are used to measure dark-brown to off-white color and off-white to pure white. The refractive index and the specific gravity are other parameters often determined.

The structural and physical properties, in particular, the solid and liquid state behavior of lipids, and the optical and spectral characteristics of waxes have been described in detail (4).

PHARMACEUTICAL APPLICATIONS

Waxes in Matrix-Type Drug Delivery Systems

The incorporation of drugs into inert matrices is a popular approach to prolong the drug release. Sustained-release wax matrix drug delivery systems include wax granules or

beads prepared by granulation or extrusion/spheronization, tablets, and wax-filled hard gelatin capsules.

Wax Granules and Beads

Drug-containing wax granules were prepared by melt congealing, by congealing in chloroform, by granulation, and by aqueous dispersion (9–11). In the congealing method, the drug was suspended in the molten wax. This suspension was cooled gradually while stirring until a solid mass formed, which was then comminuted into granules. Granules made by congealing in chloroform were prepared similarly, the drug being suspended in a chloroformic solution of the wax. This mixture was agitated until the solvent evaporated, and was then comminuted into granules. In the granulation method, the powdered wax and drug were granulated with chloroform. In the last method, heated water was given to the molten drug–wax mixture until phase inversion occurred. The emulsion was cooled and the particles were separated from the aqueous phase by filtration. The melt-congealing method gave the largest retardation in drug release, probably because of the denser structure of the granules when compared to granules prepared by the other methods.

Sustained-release nitrofurantoin granules containing stearic acid and glycerol monostearate as matrix materials were prepared by either a fusion, solvent evaporation, or melt granulation technique (12). Various channelling agents including Aerosil, Avicel, dibasic calcium phosphate dihydrate (Emcompress), and sodium chloride were investigated in order to increase the drug release. In the fusion method, the lipid was melted and the drug was added to the melt. After cooling, the congealed mass was granulated through standard sieves. In the solvent evaporation method, the drug and the wax carriers were dissolved in dimethyl formamide. The solution was cast and the solvent was evaporated at 70°C. The resulting mass was granulated as described previously. In the melt granulation method, the drug and carriers were mixed at high speeds; the temperature increased by friction and granulation occurred by sintering of the fatty materials near their melting point. The cooled granules were crushed and sieved. Sustained release could only be obtained with granules prepared by the fusion method. The solvent evaporation method and, surprisingly, melt granulation resulted in granules with a fast drug release.

Drug-containing beads were prepared by extrusion/spheronization of powder blends of 10% drug (chlorpheniramine maleate or acetaminophen), 60% Avicel PH-101, and 30% wax (13). Thermally treating the beads at 80°C for 30 min resulted in the melting and recongealing of the

wax within the beads and in a decrease in the drug release, probably because of a densification and redistribution of the wax within the beads. The drug release was dependent on the treatment temperature and the level of the wax, with an increase in both resulting in a decrease in drug release. The beads were subsequently compressed into tablets, which had slower release rates when compared to the release from the corresponding beads. While nontreated compacts mainly disintegrated into the beads, thermal treatment of the compacts resulted in nondisintegrating matrices.

Diclofenac sodium–carnauba wax matrix granules containing various release-controlling agents such as hydroxypropylcellulose, Eudragit L-100, or NaCl were prepared with a twin-screw extruder and evaluated in vitro and in vivo (14, 15). The drug release depended strongly on the composition of the granules: the mechanical strengths of the wetted granules was high, and a good correlation was obtained between in vitro dissolution parameter and in vivo parameters.

Wax Matrix Tablets

Wax matrix tablets are prepared either by compression of the wax granules or beads described in the previous section, or, with the higher melting waxes available in powder form, by direct compression of powder blends. When compared with polymeric matrix materials, the amount of waxes within the tablet is limited because of fusion and sticking to the punches at higher wax concentrations. This problem could possibly be overcome by compression at lower temperatures.

Matrix tablets containing ephedrine hydrochloride and hydrogenated castor oil were prepared either by compression of a physical mixture or by compression of a congealed melt (16–18). In the second method, the drug was added to the molten hydrogenated castor oil at 100°C, this molten mass was poured onto a glass plate and congealed, comminuted and then compressed into a matrix. A surfactant, 0.1% alkyltrimethylammonium bromide, was added to the dissolution medium to enhance the wetting of the wax matrix. The drug release increased with increasing amounts of drug in the matrix because of an increased porosity, and followed the square root of time relationship (Fig. 1). The release was slower from the matrix prepared by the melt method. This was attributed to a higher tortuosity and lower porosity of the melt matrix when compared to the matrix prepared from the physical mixture. Increasing the pressure decreased the release rate with both preparation processes; however, the effect was much more pronounced for the matrix prepared from

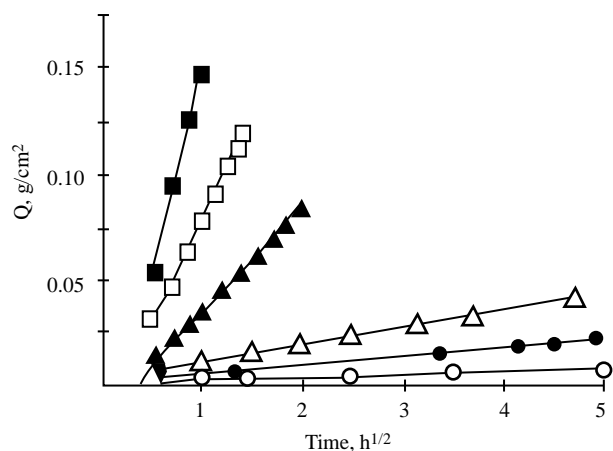


Fig. 1 Effect of concentration of ephedrine hydrochloride on release from matrixes compressed at 7 MPa (48,265 psi). (○) 5%; (●) 10%; (△) 20%; (▲) 30%; (□) 40%; (■) 50%. (From Ref. 17.)

the physical mixture. The processing method also altered the mechanism and rate of release; a matrix diffusion mechanism was dominant with the melt process while a boundary layer diffusion was effective with matrixes prepared by the compression of physical mixtures. In the first case, the drug release was independent of stirring speed, while it was dependent in the second case.

Sustained-release nifedipine tablets based on Gelucires were developed (19). A drug-PVP coprecipitate was either added to a solution of Gelucire in chloroform with the subsequent removal of the solvent or it was added to the molten Gelucire with subsequent cooling of the paste. The mixtures were then granulated through a 0.5-mm sieve and tableted. No difference in the release profiles from the two granulates were observed; however, the melting method was preferred because of the absence of organic solvents. Accelerated stability studies on the tablets as a function of temperature and relative humidity revealed no changes in chemical stability; however, the dissolution profiles changed. The changes occurring after storage at high humidities and temperatures were attributed to the formation of nifedipine microcrystals and to structural changes in the wax vehicle.

A combination of ethylcellulose and paraffin wax or hydrogenated castor oil was used as carrier material in sustained-release aminophylline tablets (20). The tablet granulation was prepared by a wet granulation procedure with hot ethanol. The tablets were further coated with Eudragit RL/RS or HPMC/ethyl cellulose. An annealing step at 70°C significantly decreased the drug release due to fusion within the tablet core. Hydrogenated castor oil was superior to paraffin wax. Increasing the wax content and

decreasing the amount of ethyl cellulose or increasing the drug content resulted in faster drug release.

In order to overcome the disadvantage of hydrophilic matrixes of uncontrollable erosion of the hydrated polymer gel on the tablet surface, a combined polymer-wax carrier material was evaluated by Huang et al. (21). Carnauba wax was combined with the enteric acrylic polymer, Eudragit L 100, and investigated as a carrier material for diphenhydramine HCl. The drug and Eudragit L100 were mixed, added to the molten wax, followed by compression of the congealed granules. The polymer provided an insoluble structure for the wax. The drug release was significantly retarded with increasing amounts of the anionic polymer. The cationic drug, diphenhydramine HCl, interacted with the anionic Eudragit L within the wax matrix and formed a complex after water penetration. While the drug release from tablets prepared from only the drug and the enteric polymer was highly dependent on the pH of the dissolution media, the drug release was almost independent of pH after the inclusion of carnauba wax in the matrix. Besides the retarding effect of the wax, at low pH, the release was retarded through the enteric polymer and at high pH, the drug formed a less soluble complex with the enteric polymer, thus negating the effect of pH on the drug release.

Matrix minitabets based on starch/microcrystalline wax mixtures were prepared by melt granulation in a hot-stage screw extruder, followed by milling and compression (22). This technique was preferred over the production of pellets.

Surfactants were incorporated into wax matrixes in order to increase drug release (23). The drug and surfactant were added to a molten mixture of carnauba wax and stearyl alcohol. Water-insoluble surfactants such as glycerol monostearate had no effect on the dissolution rate, slightly soluble surfactants such as sodium stearate or dioctyl sodium sulfosuccinate moderately increased the drug release, while polyoxyethylene 23 lauryl ether significantly increased the drug release. The drug release occurred via a leaching mechanism; drug diffusion through the matrix did not occur. It was speculated that the surfactant created more channels for the drug to leach into the dissolution medium by increasing the porosity of the matrix.

While the drug release from wax matrix tablets followed the square root of time relationship, approximately zero-order release of ephedrine hydrochloride and procaine hydrochloride could be obtained with multi-layered matrixes of hydrogenated castor oil containing different concentrations of the active compound in each layer (24).

A bioadhesive lozenge containing cetylpyridinium chloride was developed based on a multilayered tablet (25). One layer contained the bioadhesive polymer, carbopol, and the other layer contained the drug and a wax (spermaceti or Precirol ATO-5).

Wax Implants

Besides polymers such as polylactides, waxes have the potential as biocompatible/biodegradable carriers in implants. Standard tableting equipment can be used to prepare wax compacts.

Various lipids including triglycerides (e.g., trilaurin, trimyristin, tripalmitin and tristearin) and fatty acids were evaluated as carrier materials in sustained-release insulin implants (26). The drug/wax powder blend was compressed into a disc and implanted subcutaneously into Wistar rats. Monoglycerides tested eroded too fast and were not suitable. The triglycerides only sustained the insulin release briefly. The best sustained-release properties were obtained with palmitic and stearic acids as the carrier materials.

The drug release of the model protein, bovine serum albumin (BSA), from compressed stearic acid pellets was investigated as function of drug loading, drug and carrier particle size, and compression force (27). At low loadings (5%), the drug release increased with increasing BSA particle size irrespective of the particle size of stearic acid. At high loading (20%), the drug release was higher with larger stearic acid particles. More BSA was released with increasing BSA particle size only when the stearic acid particle size was small. The compression force did not show any effect in the range investigated. In a series of articles, the same research group investigated cholesterol–lecithin implants as a delivery system for antigens (28–31).

Labrafil 1944 CS (derivatized vegetable oil)–Precirol ATO 5 (glyceryl ester of fatty acids) gels were shown to be biocompatible and biodegradable and resulted in controlled release of steroids for prolonged periods of time in vivo (32).

Hard Gelatin Capsules Filled with Waxes

Waxes are difficult to compress at higher levels. The energy imparted during compaction causes melting of the waxes, resulting in sticking and picking of the formulation. This necessitates dilution of the drug-wax granules with inert fillers; high-dose drugs requiring high amounts of wax in order to obtain sustained-release properties are therefore difficult to formulate into tablets. As an alternative to

compressed tablets, hard gelatin capsules have been liquid-filled with solutions or dispersions of drugs in molten waxes. On cooling, drug–wax plugs are obtained. Some of the advantages of liquid filling of hard gelatin capsules, when compared to solid-filled capsules, include a better weight uniformity, and the elimination of dust hazards and cross-contamination via airborne particles. The drug–carrier system should not interact with the gelatin shell in the molten as well as in the congealed state, and the physical state of the drug and wax should not change during storage. The melting point of the wax has to be low enough to avoid degradation of the drug and damage to the capsule shell.

A Zanasi hard gelatin powder filling capsule machine (model LZ64) was modified in order to allow the filling of molten or thixotropic formulations into hard gelatin capsules (33, 34). This technique resulted in excellent fill weight uniformity and overcame many of the problems frequently associated with the filling of conventional capsules.

The release of liquid or deliquescent drugs (benzonatate, nicotinic acid, chloral hydrate, and paramethadione) incorporated into Gelucire bases within hard gelatin capsules was related to the behavior of the carriers in simulated gastric fluid (35). Gelucires are semisynthetic glycerides with varying amphiphilic properties and are derived from natural hydrogenated food grade fats and oils. They are characterized by their melting point (range: 33–64°C) and HLB value (range: 1–13). The drugs were released faster from the Gelucires with the higher HLB value and with lower melting points. The Gelucires either dissolved completely or remained intact but softened.

It is well known that wax-based dosage forms can experience physical instabilities. Suppository bases often show an increase in melting point, accompanied by a hardening process. This hardening can result in a reduced release rate, which could also affect the in vivo performance. Melt-filled hard gelatin capsules were evaluated by DSC, dissolution, and hardness properties as expressed by a relative penetration (36). Ketoprofen dissolved in the wax, Gelucire 50/13, and apparently formed a solid solution at room temperature as indicated by the absence of crystalline drug by DSC and microscopy. The melting point of the wax increased during storage and was accompanied by a hardening process. However, the release rate increased, which was attributed to an increased rate of matrix erosion. The observed in vitro changes did not affect the in vivo performance of the wax matrix.

As an alternative to hot-melt filling of capsules, sustained-release wax matrices were formed in a novel way within hard gelatin capsules through fluidization in a heated air stream within a fluidized bed (37). A drug–wax

powder blend was filled into hard gelatin capsules, which were then suspended in an upward-moving, heated airstream and circulated within the chamber of a fluidized bed unit. The capsules rotated during fluidization at temperatures above the melting point of the wax and centrifugal forces caused the drug-wax melt to flow into the ends of the capsules. The molten mixture then solidified after ending the heating and two solid wax matrices with dissolved/dispersed drug were obtained in the ends of the capsules. Good control over the drug release was obtained by using blends of waxes with different amphiphilic properties (HLB-values). Gelucire 50/13 (m.p. = 50°C, HLB = 13) and Precirol ATO-5 (m.p. = 53°C, HLB = 2) were selected as the drug carriers. The drug release increased with increasing proportion of the more hydrophilic wax. After cooling of the drug-containing wax melt, the drug could be dispersed, dispersed/dissolved, or dissolved in the wax matrix. DSC studies were used to characterize the physical state of the drugs after formation of the wax matrix. A linear relationship existed between the heat of fusion and the amount of drug in the wax matrix (Fig. 2). The solubility of the drug in the matrix at its melting point corresponds to the intercept of the line. Propranolol HCl was insoluble and dispersed in the Precirol ATO-5 matrix while theophylline was partially dissolved in the wax.

Coating with Waxes

Besides their predominant use in matrix preparations, waxes have also been used as coatings for granules and pellets. Solid dosage forms are often coated to sustain the

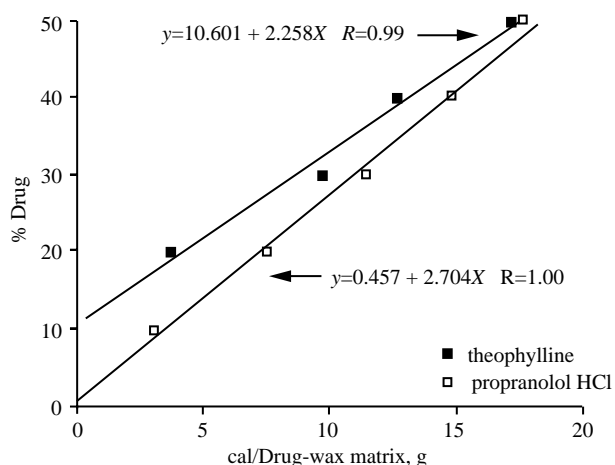


Fig. 2 Relationship of propranolol HCl and theophylline loading and the heat of fusion. (From Ref. 37.)

drug release, to improve the stability, or to mask the taste of poorly tasting drugs. The coating with waxes has various advantages when compared to the coating with polymer solutions or dispersions. The waxes can be applied without organic solvents, and, in the case of hot melts, with a high application rate and therefore shorter processing time. Many food grade natural or semisynthetic waxy materials are available.

Waxes can be applied onto solid dosage forms in the form of hot melts, hot emulsions, aqueous suspensions (colloidal wax particles), or organic solutions. Coating processes include dip coating, pan coating, or fluidized-bed coating. The coating of drug particles by spray congealing is discussed in the section on microencapsulation with waxes.

Coating with Hot Melts

With regard to coating processes, various fluidized-bed techniques and their modifications were evaluated for hot-melt coating (38–40). The fluidized-bed techniques include the top, bottom, and the tangential spray or rotary fluidized-bed modes. The particles to be coated are suspended in a heated high-velocity air stream, and the molten wax is applied in the form of atomized liquid droplets.

The top spray mode, in which the wax melt is sprayed downward on upward moving particles, is the system of choice for hot-melt coating (Fig. 3). The product temperature can be kept closest to the congealing temperature of the wax when compared to the other two spray modes. The wax has to be kept in a molten state in

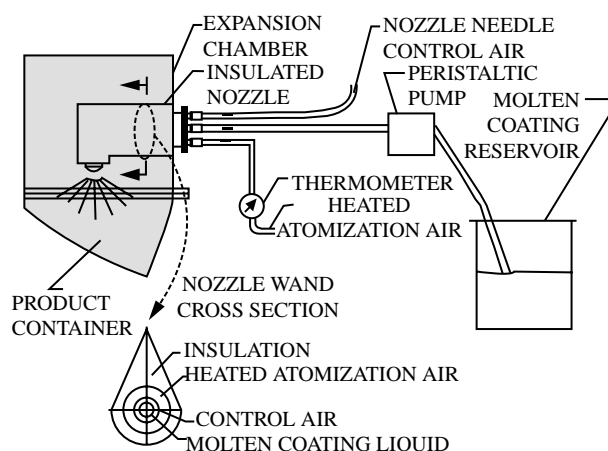


Fig. 3 Insulated nozzle and wand for top-spray hot-melt coating. (From Ref. 38.)

order to be atomized into the fluidized bed. A special nozzle had to be developed. The nozzle wand had a triaxial structure with a center tube for the molten liquid, which is surrounded by a small air space for the delivery of high-pressure, low-volume air to control the valve in the nozzle, which opens when the pump was running. Both of these tubes were surrounded by a larger air space through which the heated atomization air was supplied. The nozzle should be placed as closely as possible to the substrate bed in order to minimize the distance that the molten droplets had to travel prior to contacting the substrate surface.

Substrates with poor fluidization characteristics such as larger particles and/or particles of higher density are difficult to coat with the top spray mode, and the bottom spray mode should be preferred. With the tangential spray process, the product temperatures have to be maintained lower than with the top spray technique in order to avoid adherence of the coated particles to the product container. The tangential spray method is more stressful on the substrates than the other two techniques.

Important process and formulation variables include the product bed temperature, the atomization conditions, the type of substrate, the properties of the coating materials, and the desired release rates (immediate or sustained release). In order to obtain good coatings, the atomization air has to be heated to the same temperature as the molten wax in order to avoid premature congealing. The droplets must remain in a liquid state until they hit the substrate surface. The product bed temperature is very critical to the successful coating of solid dosage forms with molten materials. At low product temperatures, premature congealing of the molten droplets results in poor spreading of the coating on the substrate surface and, in the extreme case, in the failure of the adherence of the coating to the surface. Rough and porous surface structures are obtained, resulting in faster drug release when compared to substrates with smooth coatings. At too high product temperatures, excessive particle agglomeration or clogging of the outlet filter bags is a result of inadequate congealing and hardening of the coating. The product bed temperature can be regulated through the fluidization air temperature. It is recommended to use inlet air temperatures 10–15°C below the melting point of the coating and temperatures for the atomization air and the molten wax of 40–60°C above the melting point.

Droplet size and uniformity are also critical for a successful coating process. The size of the molten droplet is dependent on the viscosity of the melt and the atomization air pressure. Smaller particles require smaller droplet sizes and therefore higher atomization air pressures in order to minimize agglomeration or granulation. In order to obtain small droplets, the viscosity of

the molten material can be decreased by increasing the temperature of the melt. At the same atomization conditions, low feed rates also result in smaller droplets. The spray rate of melts is generally much lower when compared to coating solutions or dispersions. The lower spray rate, however, is offset by the application of pure coating material. The application rate therefore is still higher when compared to polymer solutions or dispersions, with which solvents have to be evaporated.

After the application of the melt, the fluidization is reduced and the product bed is cooled. The cooling circle should be short in order to avoid attrition of the coated product. Rapid cooling, however, may result in cracks in the coating because of contraction of the coating material, and it may also result in unstable polymorphic forms of the wax.

The important variables for the selection of the waxes include the melting point, the melting range, and the viscosity of the melt. The wax should have a melting point of less than 85°C because the melt is usually kept at temperatures of 40–60°C above its melting point. Materials with a broad melting point range can become tacky during spraying because of the broad range of product temperatures and the presence of low melting point fractions. The coating materials include various hydrogenated vegetable oils, beeswax, paraffin wax, carnauba wax, and polyethylene glycol.

Coating with hot emulsions, aqueous suspensions, or organic wax solutions

The coating with hot emulsions has various advantages when compared to the hot-melt process (41, 42). The wax remains in the molten state and premature congealing of the wax could be eliminated because of the presence of hot water. This facilitated the transport of the wax to the spray nozzle and therefore the experimental setup. In addition, the temperatures of the wax emulsion are lower than the comparable wax melts. o/w-Emulsions of various waxes (e.g., glyceryl behenate-Compritol-888 and glyceryl palmitostearate-Precirol-ATO 5) with a solids content of up to 50% were prepared. The emulsions were passed through a microfluidizer to further reduce the particle size of the oil phase. The hot emulsion was then either sprayed directly on the beads or cooled to form a “wax pseudolatex” prior to the coating process. The disadvantages, when compared to hot-melt coating, include the amount of water to be evaporated and therefore longer processing times, and the presence of surfactants in the wax coating. The surfactants, which were needed to stabilize the emulsions, could affect the drug release. The use of liquid surfactants such as various Tween/Span combinations resulted in sticky beads. Solid surfactants such as sodium lauryl sulfate were more suitable. The guaifenesin release from coated

pellets decreased with increasing hydrophobicity of the wax and increasing coating level. The coating conditions (e.g., temperature, spray rate, curing) and the particle size of the emulsion/suspension primarily influenced the microstructure of the wax coatings and hence the drug release. When compared to polymer coatings, thicker coatings have to be applied with waxes to obtain the same sustained release profiles.

Bagaria prepared emulsions with waxes including carnauba, paraffin, ceresin, beeswax, and hydrogenated castor, soybean, or cottonseed oil, which were then coated onto drug-containing nonpareil beads (43). The term emulsions was actually misleading because the final products were aqueous suspensions of the waxes with partially submicron particle size (wax pseudolatex). Similar to aqueous polymer dispersions, the wax dispersions were converted into powders by spray drying. The release from beads coated with the redispersed spray dried powder was then compared to beads coated with the original dispersion. With almost all preparations, 100% drug was released within 2–4 h.

Like with organic polymer solutions, the wax can be dissolved in an organic solvent and then be sprayed onto the solid dosage form. In most studies, waxes such as beeswax, hydrogenated castor oil, microcrystalline wax, or glyceryl mono- and distearate were dissolved in chlorinated organic solvents such as chloroform, carbon tetrachloride, or trichlorethane and applied in coating pans at elevated temperatures (44–46). Mixtures of ethyl cellulose with different waxes such as castor, carnauba, or paraffin wax in chloroform were evaluated as sustained-release coatings (47). A lower ratio of ethyl cellulose to wax was used with drugs of high-molecular weight and/or low solubility and a higher ratio for drugs of low-molecular weight and/or high solubility to achieve the desired release properties at a 10% coating level. As the data of the references shows, the coating with organic solutions is obsolete because of the undesirable use of organic solvents.

Microencapsulation with Waxes

Wax microparticles have been prepared primarily by aqueous and nonaqueous melt dispersion techniques or spray congealing/spray drying. These techniques are briefly reviewed below.

Microparticles prepared by melt-dispersion techniques

In the melt dispersion technique, the drug-containing molten wax phase is emulsified into a heated, emulsifier-containing external phase. Depending on the solubility of the drug, the external phase can be either aqueous (for

water-insoluble drugs) or nonaqueous (for water-soluble drugs). On cooling the emulsion, the liquid droplets congeal and a suspension of the wax microparticles is formed. The microparticles are then separated, mostly by filtration or centrifugation, sometimes washed to remove free drug crystals and surfactants, dried and sized.

Ibuprofen-wax (carnauba wax, paraffin wax, beeswax, Gelucire 64/02, or Precirol ATO5) microparticles were characterized with respect to drug loading and morphological and release properties (48). Microparticles of the more hydrophilic waxes, Gelucire 64/02 and Precirol ATO5, could be prepared without surfactants, while the other waxes rapidly coalesced and formed big lumps on cooling. With the other waxes, increasing the amount of sodium lauryl sulfate in the external aqueous phase decreased the drug loading because of drug solubilization. This was not the case when poly(vinyl alcohol) was used as stabilizer. During emulsification, the drug will partition into the external aqueous phase until its solubility at the emulsification temperature is reached. During cooling of the emulsion, the drug could precipitate in the aqueous phase because of a decreased drug solubility. The type of wax, the rate of cooling, the stirring time, and the temperature of the aqueous phase had no significant effect on the drug loading because of the low solubility of ibuprofen in the external aqueous phase. Actual drug loadings close to 60% could be achieved. The drug release was controlled by the hydrophobicity of the wax (Gelucire 64/02 > Precirol ATO5 > beeswax > carnauba wax > paraffin wax) (Fig. 4). These wax microparticles could be

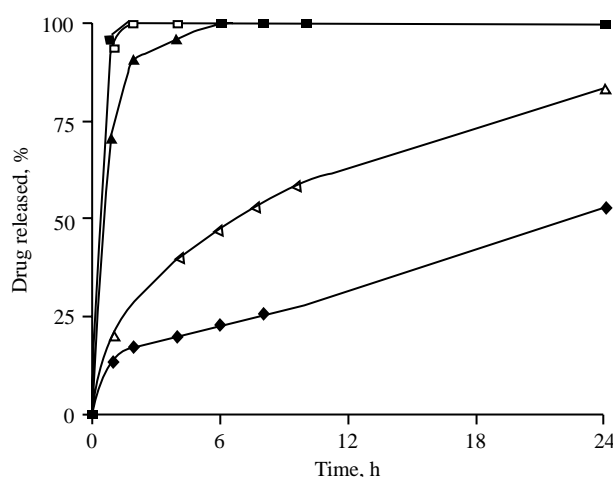


Fig. 4 Ibuprofen release from different wax microparticles (actual drug loading): (■) Precirol ATO5 (35.3%), (□) Gelucire 64/02 (35.4%), (▲) beeswax (37.7%), (△) carnauba wax (37.3%), (◆) paraffin wax (35.0%). (From Ref. 48.)

formulated into aqueous sustained-release oral suspension dosage forms because of the low solubility of the drug.

A modified USP method using mini baskets was used to study the effect of formulation variables, such as type of wax, type of modifier, drug loading, size, on the ibuprofen release from wax microspheres (49). The drug release was in the order of beeswax > ceresine wax > refined paraffin wax > microcrystalline wax. After an initial burst release, the drug release was slow from microparticles prepared with the last two waxes. The dissolution studies were performed in simulated intestinal fluids, because sink conditions could not be maintained in simulated gastric fluids because of the drug solubility. To increase the drug release from the wax microspheres, glycerol monostearate or stearyl alcohol were added prior to the preparation of the microparticles. The addition of these release modifiers also reduced the tendency of the waxes to agglomerate (50). Ibuprofen dissolved in the molten wax and did not crystallize during congealing and microsphere formation, as indicated by DSC studies.

In a modified melt-dispersion method, sulfamethoxazole particles were dispersed in heated water, and powdered beeswax was added. The molten wax droplets collected the drug particles, and spherical agglomerates with sustained-release properties were obtained after cooling. Alternatively, the agglomerates could also be formed at room temperature by using solutions of the wax in a water-immiscible solvent (51).

Water-soluble drugs cannot be encapsulated with the O/W-emulsion technique because the drug would be lost to the external aqueous phase. Two methods have been described for the encapsulation of hydrophilic drugs; one is based on using an external oil phase and the other one on the formation of the microparticles by a w/o/w-melt dispersion technique.

Carnauba-wax microspheres containing 5-fluorouracil for chemoembolization were prepared by a melt dispersion process with an external silicone oil phase (52). The drug was dispersed in the molten carnauba wax and emulsified into silicone oil at temperatures above the melting point of carnauba wax (85°C). The resulting emulsion was cooled through the addition of cold silicone oil and immersion of the beaker in an ice-water bath. After solidification, the microspheres were separated from the oil phase by centrifugation and washed with cyclohexane to remove the silicone oil. 5-Fluorouracil is a hydrophilic drug; an external aqueous phase would have resulted in drug partitioning and therefore low encapsulation efficiencies. However, even with silicone oil, an external phase in which the drug was insoluble, only up to 5% drug could be encapsulated within the carnauba-wax microspheres. This was attributed to the poor wetting of the drug crystals by

the molten wax and therefore the loss of drug crystals to the external silicon oil phase. Various surfactants were added to the wax phase in order to improve the wettability of the drug by the molten wax.

A technique based on the formation of a multiple emulsion with an external aqueous phase was developed for the encapsulation of water-soluble drugs in order to replace the external oil phase (53). Possible unwanted interactions between the oil and the emulsified wax such as swelling or dissolution of the wax, clean-up requirements of the final product, and recovery of the oil phase could be eliminated. In analogy to the encapsulation of water-soluble drugs within polymeric microparticles by a w/o/w-solvent evaporation method, a molten wax phase was used instead of an organic polymer solution. A heated aqueous solution of pseudoephedrine HCl was emulsified into the molten carnauba wax, followed by the emulsification of this w/o-emulsion into a heated external aqueous phase. The temperature of the internal and external aqueous phases had to be kept above the melting temperature of the wax in order to avoid premature congealing and to assure the formation of an emulsion. The microparticles formed after congealing of the wax. A key for high encapsulation efficiencies was the formation of ultrafine internal aqueous phase droplets by sonication. The wax acted as a diffusion barrier between the internal and the external aqueous phases and therefore minimized drug partitioning into the external aqueous phase. Because of the high solubility of the drug in the external aqueous phase, the contact time of the droplets/microparticles with the continuous aqueous phase had to be minimized in order to avoid the drug loss; the microparticles were separated from the aqueous phase within minutes after their formation.

Generally, the wax phase is emulsified into a heated external phase to avoid premature congealing. However, sulfamethazine-japanese synthetic wax particles were prepared by dispersing the drug in the molten wax, followed by slowly pouring the wax phase into a precooled external aqueous phase (54). The microparticles were then separated by filtration after 3 min. Beeswax microparticles were prepared by a phase inversion technique, whereby an aqueous solution of sorbitan monooleate and polysorbate 80 was added to the drug-containing molten wax phase to first form a w/o-emulsion prior to phase inversion. Smaller microparticles were obtained at higher temperatures, with increased amounts of continuous phase, with slower rates of cooling, and with higher speeds of mixing (55).

Suspensions of lipid nanoparticles can be prepared by reducing the size of the molten or dissolved drug containing lipid phase into the colloidal size range by high-pressure homogenization (56, 57). After cooling or solvent evaporation, the nanoparticles are obtained.

The suspensions of the nanoparticles can be converted into powders through freeze or spray drying. The nanoparticles can act as carriers for poorly water-soluble drugs and apparently could result in controlled release over longer periods of time. However, because of the small size and therefore high surface area of the particles, an increase in dissolution rate would be expected unless very low loadings are used.

Microparticles prepared by spray drying and spray congealing

Similar to organic polymer solutions, drug-containing organic wax solutions were spray dried to give sustained-release microparticles (58). The drug could be either dissolved or dispersed in the organic wax solutions. Spray drying is a single step, rapid drying process, which can be scaled up and be used for heat-sensitive drugs. The use of organic solvents, however, is undesirable because of solvent hazards, solvent residuals, and cost. Because of the low melt viscosity, wax microparticles can also be prepared without organic solvents by spray-congealing drug-containing wax melts.

Sulfaethylthiadiazole-hydrogenated castor oil particles were prepared by spray congealing, using a centrifugal wheel atomizer (59). The wax powder was then suspended into water to give an oral sustained-release suspension. In a subsequent article, the effect of various process and formulation variables on the particle size was investigated using a centrifugal wheel atomizer (60). The particle size was directly proportional to the feed rate and inversely proportional to the feed viscosity and the wheel velocity.

In a series of publications, sulfaethylthiadiazole-wax microparticles prepared by spray-congealing were evaluated. The most important factor affecting the drug release was the type of wax used (61). The effect of surfactant on the drug release from spray congealed wax microparticles was investigated by John and Becker (62). White wax-USP, a synthetic wax-like ester, and 1:1 combinations of these two waxes were used as sustained-release matrices. The particle size of the microparticles could be controlled through the nozzle size, with the larger nozzles resulting in larger particles with slower rates of dissolution. Up to 4% sorbitan monooleate apparently softened the particles and promoted wetting, thus resulting in an increase in drug release. However, at a surfactant level of 10%, the slowest release rate was observed. This was partly attributed to the tackiness of the particles, which resulted in agglomeration and a reduction in total surface area available for drug release. The addition of the surfactant allowed the compression of the microparticles into tablets, which was difficult with the surfactant-free microparticles (63). The surfactant-free microparticles adhered or stuck to

the punch surface, and resulted in a high friability of the finished tablets. No additives were incorporated into the tablets; nondisintegrating wax matrix tablets were obtained.

A palatable suspension of the bitter-tasting drug, remoxipride, was developed based on microparticles prepared by spray congealing (64). Because of the high water solubility of the drug, the microparticles were formulated into an external oil phase. Unfortunately, neither the nature of the wax nor of the oily vehicle was revealed.

With lipid drug delivery systems, polymorphic transformations may occur during the preparation of the dosage form and during subsequent storage. The polymorphic behavior of lipid micropellets prepared from glycerides and phospholipids by spray drying or spray congealing and their surface structure were evaluated by DSC and scanning electron microscopy (65–67). The rapid solvent evaporation during spray drying can influence the crystallization of the lipid carrier and different polymorphic structures could be obtained. Similarly, during spray congealing and solidification of the melt, the lipid can crystallize in different polymorphic forms, depending on the composition of the lipid and the cooling rate. The major polymorphic forms of the glycerides are the α -, β -, and β' -forms. Rapid cooling rates generally result in the unstable α -form. The transition of the melt first to the α - and β' -forms and then to the β -form represents the transformation of triglycerides into the most stable form.

The spraying process, especially the spray congealing was simulated by cooling the molten samples rapidly at 320°C/min. A DSC thermogram of pure tristearin heated at 10°C/min revealed a single endothermic peak representing the β -form. After rapid cooling of the melt and reheating, an endothermic peak for the α -form and an exothermic peak indicating the recrystallization of the α -form into the β -form, followed by the endothermic peak for the β -form, were detectable. The DSC thermogram of spray-dried tristearin micropellets was similar to the one of the melt-quenched sample. A crystalline modification therefore took place during spray drying because of the rapid solvent removal. The micropellets were then stored at different temperatures to investigate the effect of storage temperature on the polymorphic transitions. Increasing the storage temperature to 37°C resulted in a complete transformation of the unstable polymorphic form into the stable polymorphic form. The melting endotherm of the α -form disappeared. The effect of various emulsifiers such as lecithin or monoglycerides was investigated in order to prevent or delay the transformation of the unstable form into the stable

β -form. Adding lecithin to the formulation resulted in a delay of the transformation. The type of glyceride (composition and chain length), solvent, and drugs encapsulated affected the polymorphic transformation, its rate of transformation, and also the surface structure of the microparticles. Spray-congealed lipid micropellets showed a similar thermal behavior as the spray-dried pellets. The smooth surface of the sprayed lipid micropellets was attributed to the unstable polymorphic form. The unstable α -form has various thin and small crystals, resulting in a smooth surface of the crystallized sample. The β - and β' -forms have larger crystals, therefore causing irregular structures of the micropellets. During ageing at elevated temperatures, the lipid micropellets lost their smooth surface structures because of polymorphic transformations.

Novel oral controlled release microspheres using polyglycerolesters of fatty acids and hydrogenated cottonseed oil (HCSO), stearic acid, stearyl alcohol, or glycerol monostearate as carriers were prepared by spray chilling using a rotating disc (68). The drug release was related to the hydrophobicity of the wax; the release rate decreased in the following order: stearyl alcohol > stearic acid > glycerol monostearate > carnauba wax > hydrogenated cottonseed oil. Adding increasing amounts of lactose to HCSO increased the drug release as a result of the leaching of lactose.

A new atomizer operating with ultrasonic energy was used as an alternative to traditional atomizers to prepare wax microparticles by spray congealing (69).

REFERENCES

- Bennett, H. *Industrial Waxes, Vol. I—Natural and Synthetic Waxes, Vol. II—Compound Waxes and Technology*; Chemical Publishing Company: New York, 1975.
- Wade, A.; Weller, P.J. *Handbook of Pharmaceutical Excipients*; American Pharmaceutical Association, The Pharmaceutical Press: London, Washington, 1994.
- Lechter, C.S. Waxes. *Encyclopedia of Chemical Technology*; Kirk-Othmer, Ed.; John Wiley & Sons: New York, 1984; 24, 466–481.
- Gunstone, F.D.; Harwood, J.L.; Padley, F.B. *The Lipid Handbook*; Chapman and Hall: London, 1986.
- Knowlton, J.; Pearce, S. *Handbook of Cosmetic Science and Technology*; Elsevier: Amsterdam, 1993; 21–32.
- Warth, A.H. *The Chemistry and Technology of Waxes*; Reinhold Publishing Corporation: New York, 1956.
- Kolattukudy, P.E. *Chemistry and Biochemistry of Natural Waxes*; Elsevier: Amsterdam, 1976.
- In *United States Pharmacopeia, USP23 NF18*; The United States Pharmacopeial Convention: Rockville, MD, 1995.
- Asker, A.F.; Motawi, A.M.; Abdel-Khalek, M.M. A Study of Some Factors Affecting the In-Vitro Release of Drug from Prolonged Release Granulations. Part 1. Effect of Method of Preparation. *Pharmazie* **1971**, 26, 170–172.
- Asker, A.F.; Motawi, A.M.; Abdel-Khalek, M.M. A Study of Some Factors Affecting the In-Vitro Release of Drug from Prolonged Release Granulations. Part 2. Effect of Dissolution Retardant. *Pharmazie* **1971**, 26, 213–214.
- Asker, A.F.; Motawi, A.M.; Abdel-Khalek, M.M. A Study of Some Factors Affecting the In-Vitro Release of Drug from Prolonged Release Granulations. Part 3. Effects of Particle Size, Enzymatic Contents of Pepsin and Pancreatin, Bile and Ionic Concentration. *Pharmazie* **1971**, 26, 215–217.
- El-Shanawany, S. Sustained Release of Nitrofurantoin from Inert Wax Matrixes. *J. Control. Rel.* **1993**, 36, 11–19.
- Ghali, E.S.; Klinger, G.H.; Schwartz, J.B.; *Drug Dev. Ind. Pharm.* **1989**, 15, 1311–1328.
- Miyagawa, Y.; Okabe, T.; Yamaguchi, Y.; Miyajima, M.; Sato, H.; Sunada, H. Controlled-Release of Diclofenac Sodium from Wax Matrix Granule. *Int. J. Pharm.* **1996**, 138, 215–224.
- Miyagawa, Y.; Sato, H.; Okabe, T.; Nishiyama, T.; Miyajima, M.; Sunada, H. In Vivo Performance of Wax Matrix Granules Prepared by a Twin-Screw Compounding Extruder. *Drug Dev. Ind. Pharm.* **1999**, 25, 429–435.
- Foster, T.P.; Parrott, E.L. Effect of Processing on Release from an Inert, Heterogeneous Matrix. *Drug Dev. Ind. Pharm.* **1990**, 16, 1309–1324.
- Foster, T.P.; Parrott, E.L. Release of Highly Water-Soluble Medicinal Compounds from Inert, Heterogenous Matrixes. I: Physical Mixture. *J. Pharm. Sci.* **1990**, 79, 806–810.
- Foster, T.P.; Parrott, E.L. Release of Highly Water-Soluble Medicinal Compounds from Inert, Heterogenous Matrixes. II: Melt. *J. Pharm. Sci.* **1990**, 79, 938–942.
- Remunan, C.; Bretal, M.J.; Nunez, A.; Vila Jato, J.L. Accelerated Stability Study of Sustained-Release Nifedipine Tablets Prepared with Gelucire®. *Int. J. Pharm.* **1992**, 80, 151–159.
- Boles, M.G.; Deasy, P.B.; Donnellan, M.F. Design and Evaluation of Sustained-Release Aminophylline Tablet. *Drug Dev. Ind. Pharm.* **1993**, 19, 349–370.
- Huang, H.-P.; Mehta, S.C.; Radebaugh, G.W.; Fawzi, M.B. Mechanism of Drug Release from an Acrylic Polymer-Wax Matrix Tablet. *J. Pharm. Sci.* **1994**, 83, 795–797.
- De Brabander, C.; Vervaeke, C.; Fiermans, L.; Remon, J.P. Matrix Mini-Tablets Based on Starch/Microcrystalline Wax Mixtures. *Int. J. Pharm.* **2000**, 199, 195–203.
- Dakkuri, A.; Schraeder, H.G.; DeLuca, P.P. Sustained Release from Inert Wax Matrixes. II: Effect of Surfactants on Tripeleminamine Hydrochloride Release. *J. Pharm. Sci.* **1978**, 67, 354–357.
- Foster, T.P.; Parrott, E.L. Constant Release Rate from Inert Heterogeneous Matrixes by Means of Position-Dependent Loading. *Drug Dev. Ind. Pharm.* **1990**, 16, 1633–1648.
- Collins, A.E.; Deasy, P.B. Bioadhesive Lozenge for the Improved Delivery of Cetylpyridinium Chloride. *J. Pharm. Sci.* **1990**, 79, 116–119.

26. Wang, P.Y. Lipids as Excipient in Sustained Release Insulin Implants. *Int. J. Pharm.* **1989**, *54*, 223–230.
27. Kaewvichit, S.; Tucker, I.G. The Release of Macromolecules from Fatty Acid Matrices: Complete Factorial Study of Factors Affecting Release. *J. Pharm. Pharmacol.* **1994**, *46*, 708–713.
28. Khan, M.Z.I.; Tucker, I.G.; Opdebeek, J.P. Cholesterol and Lecithin Implants for Sustained Release of Antigen: Release and Erosion In Vitro, and Antibody Response in Mice. *Int. J. Pharm.* **1991**, *76*, 161–170.
29. Khan, M.Z.I.; Tucker, I.G.; Opdebeek, J.P. Evaluation of Cholesterol-Lecithin Implants for Sustained Delivery of Antigen: Release In Vivo and Single-Step Immunisation of Mice. *Int. J. Pharm.* **1993**, *90*, 255–262.
30. Opdebeek, J.P.; Tucker, I.G. A Cholesterol Implant Used a Delivery System to Immunize Mice with Bovine Serum Albumin. *J. Control. Rel.* **1993**, *23*, 271–279.
31. Walduck, A.K.; Opdebeek, J.P.; Benson, H.E.; Prankerd, R. Biodegradable Implants for the Delivery of Veterinary Vaccines: Design, Manufacture and Antibody Responses in Sheep. *J. Control. Rel.* **1998**, *51*, 269–280.
32. Gao, Z.; Crowley, W.R.; Shukla, A.J.; Johnson, J.R.; Reger, F. Controlled Release of Contraceptive Steroids from Biodegradable and Injectable Gel Formulations: In Vivo Evaluation. *Pharm. Res.* **1995**, *12*, 864–868.
33. Walker, S.E.; Ganley, J.A.; Bedford, K.; Eaves, T. The Filling of Molten and Thixotropic Formulations into Hard Gelatin Capsules. *J. Pharm. Pharmacol.* **1980**, *32*, 389–393.
34. McTaggart, C.; Wood, R.; Bedford, K.; Walker, S.E. The Evaluation of an Automatic System for Filling Liquids into Hard Gelatin Capsules. *J. Pharm. Pharmacol.* **1984**, *36*, 119–122.
35. Doelker, C.; Doelker, E.; Buri, P.; Waginaire, L. *Drug Dev. Ind. Pharm.* **1986**, *12*, 1553–1565.
36. Proceedings of the 15th International Symposium on Controlled Release of Bioactive Materials, Basel, Switzerland, Aug 15–19, 1988, The Controlled Release Society, Illinois 1988, No. 223, 390–391.
37. Bodmeier, R.; Paeratakul, O.; Chen, H.; Zhang, W. Formation of Sustained Release Wax Matrices Within Hard Gelatin Capsules in a Fluidized Bed. *Drug Dev. Ind. Pharm.* **1990**, *16*, 1505–1519.
38. Jones, D.M.; Percel, P.J. Coating of Multiparticulates Using Molten Materials. *Multiparticulate Oral Drug Delivery*; Ghebre-Sellassie, I., Ed.; Marcel Dekker, Inc.: New York, 1994; 113–142.
39. Jozwiakowski, M.J.; Jones, D.M.; Franz, R.M. Characterization of a Hot Melt Fluid Bed Coating Process for Fine Granules. *Pharm. Res.* **1990**, *7*, 1119–1126.
40. Barthelemy, P.; Laforêt, J.P.; Farah, N.; Joachim, J. Compritol® 888 ATO: An Innovative Hot-Melt Coating Agent for Prolonged-Release Drug Formulations. *E. J. Pharm. Biopharm.* **1999**, *47*, 87–90.
41. Bhagwatwar, H.; Bodmeier, R. The Coating of Drug-Loaded Sgura Beads with Various Wax Formulations. *Pharm. Res.* **1989**, *6*, PT 713–S-73.
42. Bhagwatwar, H. *M.S. Thesis*; The University of Texas at Austin: Austin, 1991.
43. Bagaria, S.C., Ph. D. *Dissertation*; Rutgers University, The State University of New Jersey: New Brunswick, 1986.
44. Blythe, R.H. Sympathomimetic Preparation, US Patent 2,738,303, March 13, 1956.
45. Rosen, E.; Swintosky, J.V. Preparation of A ³⁵S Labelled Trimeprazim Tartrate Sustained Action Product for Its Evaluation in Man. *J. Pharm. Pharmacol.* **1960**, *12*, 237T–244T.
46. Heimlich, K.R.; MacDonnell, D.R. Method of Preparing Sustained Release Pharmaceutical Pellets and Product Thereof, US Patent 3,119,742, January 28, 1964.
47. Peters, D.; Goodhart, F.W.; Lieberman, H.A. Sustained Release Dosage in the Pellet Form and Process Thereof. US Patent, 3,492,397, January 27, 1970.
48. Bodmeier, R.; Wang, J.; Bhagwatwar, H. J. Process and Formulation Variables in the Preparation of Wax Micro-particles by a Melt Dispersion Technique I. Oil-in-Water Technique for Water-Insoluble Drugs. *Microencapsulation* **1992**, *9*, 89–98.
49. Adeyeye, C.M.; Price, J.C. Development and Evaluation of Sustained-Release Ibuprofen-Wax Microspheres. II. In Vitro Dissolution Studies. *Pharm. Res.* **1994**, *11*, 575–579.
50. Adeyeye, C.M.; Price, J.C. Development and Evaluation of Sustained-Release Ibuprofen-Wax Microspheres. I. Effect of Formulation Variables on Physical Characteristics. *Pharm. Res.* **1991**, *8*, 1377–1383.
51. Kawashima, Y.; Ohno, H.; Takenaka, H. Preparation of Spherical Matrixes of Prolonged Release Drugs from Liquid Suspensions. *J. Pharm. Sci.* **1981**, *70*, 913–916.
52. Benita, S.; Zouai, O.; Benoit, J.-P. 5-Flourouracil: Carnauba Wax Microspheres for Chemoembolization: An in Vitro Evaluation. *J. Pharm. Sci.* **1986**, *75*, 847–851.
53. Bodmeier, R.; Wang, J.; Bhagwatwar, H.J. Process and Formulation Variables in the Preparation of Wax Micro-particles by a Melt Dispersion Technique. II. W/O/W Multiple Emulsion Technique for Water-Soluble Drugs. *Microencapsulation* **1992**, *9*, 99–107.
54. Kowarski, C.R.; Volberger, B.; Versanno, J.; Kowarski, A. *Am. J. Hosp. Pharm.* **1964**, *21*, 409–410.
55. Draper, E.B.; Becker, C.H. Some Wax Formulations of Sulfaethylthiadiazole Produced by Aqueous Dispersion for Prolonged-Release Medication. *J. Pharm. Sci.* **1966**, *55*, 376–380.
56. Lucks, S.; Müller, R. Medication Vehicles Made of Solid Lipid Particles (Solid Lipid Nanospheres) SLN. PCT Application, WO93/05768 April 1, 1993.
57. Westesen, K.; Siekmann, B. Process for Producing Detergent Tablets. PCT Application, WO94/20072 March 17, 1994.
58. Asker, A.F.; Becker, C.H. Some Spray-Dried Formulations of Sulfaethylthiadiazole for Prolonged-Release Medication. *J. Pharm. Sci.* **1966**, *55*, 90–94.
59. Robinson, M.J.; Swintosky, J.V. *J. Am. Pharm. Assoc.* **1959**, *48*, 473–478.
60. Scott, M.W.; Robinson, M.J.; Pauls, J.F.; Lantz, R.J. Spray Congealing: Particle Size Relationships Using a Centrifugal Wheel Atomizer. *J. Pharm. Sci.* **1964**, *53*, 670–675.

61. Cusimano, A.G.; Becker, C.H. Spray-Congeaed Formulations of Sulfaethylthiadiazole (SETD) and Waxes for Prolonged-Release Medication. *J. Pharm. Sci.* **1968**, *57*, 1104–1112.
62. John, P.M.; Becker, C.H. Surfactant Effects on Spray-Congeaed Formulations of Sulfaethylthiadiazole-Wax. *J. Pharm. Sci.* **1968**, *57*, 584–589.
63. Hamid, I.S.; Becker, C.H. Release Study of Sulfaethylthiadiazole (SETD) from a Tablet Dosage Form Prepared from Spray-Congeaed Formulations of SETD and Wax. *J. Pharm. Sci.* **1970**, *59*, 511–514.
64. Sjoqvist, R.; Graffner, C.; Ekman, I.; Sinclair, W.; Woods, J.P. In Vivo Validation of the Release Rate and Palatability of Remoxipride-Modified Release Suspension. *Pharm. Res.* **1993**, *10*, 1020–1026.
65. Eldem, T.; Speiser, P.; Hincal, A. Optimization of Spray-Dried and -Congeaed Lipid Micropellets and Characterization of the Surface Morphology by Scanning Electron Microscopy. *Pharm. Res.* **1991**, *8*, 47–54.
66. Eldem, T.; Speiser, P.; Altorfer, H. Polymorphic Behavior of Sprayed Lipid Micropellets and Its Evaluation by Differential Scanning Calorimetry and Scanning Electron Microscopy. *Pharm. Res.* **1991**, *8*, 178–184.
67. Eldem, T.; Speiser, P.; Hincal, A. Proceedings of the 15th International Symposium on Controlled Release of Bioactive Materials Basel Switzerland, Aug 15–19, 1988, The Controlled Release Society, Illinois No. 247, 1988, 436–437.
68. Akiyama, Y.; Yoshioka, M.; Horibe, H.; Hirai, S.; Kitamori, N.; Toguchi, H.J. Novel Oral Controlled-Release Microspheres Using Polyglycerol Esters of Fatty Acids. *Control. Rel.* **1993**, *26*, 1–10.
69. Rodriguez, L.; Passerini, N.; Cavallari, C.; Cini, M.; Sancin, P.; Fini, A. Description and Preliminary Evaluation of a New Ultrasonic Atomizer for Spray Congealing Process. *Int. J. Pharm.* **1999**, *183*, 133–143.

WORLD HEALTH ORGANIZATION (WHO) CONTINUES GLOBAL HARMONIZATION OF REQUIREMENTS FOR MEDICINAL PRODUCTS

Juhana E. Idänpään-Heikkilä

*Council for International Organizations of Medical Sciences,
WHO, Geneva, Switzerland*

INTRODUCTION

The World Health Organization (WHO) is an intergovernmental specialized agency of the United Nations with 193 member states. Regarding pharmaceutical and biological products, WHO has the global mandate to develop, establish, and promote quality, safety, and efficacy standards and codes of good practices for all medicinal products. Vigorous harmonization of quality, safety, efficacy, and nomenclature requirements on a worldwide basis has continued in 1998–2000 as the explicit responsibility of WHO. The content of the constitutional mandate, expertise involved, working practices and rules, consultative processes, and role of the governing bodies regarding WHO's normative work was described in detail in the *Encyclopedia of Pharmaceutical Technology*, Vol. 18 (1). This review gives an update on ongoing harmonization activities since 1998 and outlines the new harmonization initiatives.

International Nonproprietary Names (INNs)

The International Nonproprietary Names (INNs) identifies pharmaceutical substances by unique, globally recognized names. A single internationally recognized name for an active drug substance is a starting point for its pharmacopeial monograph, safe prescribing and dispensing, and easy communication among scientists and health professionals worldwide. The INN is for common use without restrictions (except as a trademark); it is distinctive in sound and spelling and not too long. In 1998–1999, WHO published more than 270 new INNs in its quarterly publication, *WHO Drug Information* (2) and issued a revised *Guidelines on the Use of International Nonproprietary Names (INN) for Pharmaceutical Substances* (3). The revision of the procedure included the introduction of a service fee, improvements in Procedures to raise objections to proposed INNs, and the replacement of a recommended INN.

Naming the biotechnology-derived substances and products received increasing attention. Prevention of INNs from using trademarks has now been strengthened in collaboration with the World Intellectual Property Organization (WIPO). A new point of collaboration was the protection of INNs against misuse of Internet domain names (4).

International Pharmacopoeia—50 Years Old

Publication of the International Pharmacopoeia has continued now for 50 years. Its role has been to fulfill a need in developing countries where use of less technically advanced tests using simple instrumentation is common practice for specific substances and preparations. Publishing monographs for finished products has also been useful. In 1999, the Expert Committee recommended that less-advanced methods should be developed in parallel with modern analytical techniques because in some developing countries, more sophisticated methods could be useful (4). Volume 5 of the International Pharmacopoeia is in print. WHO has completed monographs for several antimalarials including artemisin derivatives. Basic tests for pharmaceutical substances, medicinal plant materials, and dosage forms (including 345 tests for substances, 208 for dosage forms, and four for plant materials) were published in 1998 (5). The 69 International Infrared Reference Spectra (IIRS) have been made available from the WHO Collaborating Centre for Chemical Reference Substances (Kungens Kurva, Sweden). The list of reference substances and infrared reference spectra is updated regularly and is now available on the World Wide Web (<http://www.who.int/dmp/irintro.htm>).

A revised guideline on Good Practices for National Pharmaceutical Control Laboratories (GPCL) was published in 1999. It takes into account guidances from ISO 17025, EN 45001, and OECD–GLP and

recommendations published by the Pharmaceutical Inspection Convention (PIC) (4).

WHO GMP and Quality Assurance

In 1998, WHO began a special project to assist in implementing of GMP in member states with limited resources. Countries interested in upgrading their manufacturing have been identified, and training material on basic GMP principles has been prepared together with training modules for advanced GMP topics and inspections. Several regional training workshops for trainers in GMP have been organized, together with missions to concerned countries (4).

The following guidelines have been completed: GMP for Sterile Pharmaceutical Products, Guideline on Pre-Approval Inspection (before granting marketing authorization), Quality System for National GMP Inspectorate, and General Aspects of Packaging Pharmaceuticals (4).

In 1999, WHO published a report on the control and safe trade of starting materials for pharmaceutical products in English, French, Spanish, Russian, Chinese, and Arabic, based on recommendations from a meeting of experts in Geneva in May 1998 (6). The meeting was inspired by several recent incidents of contamination with highly toxic solvent diethylene glycol.

Because of their increasing use worldwide, plant materials used in over-the-counter preparations, home remedies, or as raw materials for pharmaceutical preparations are receiving more and more attention. In 1998, WHO published a book, *Quality Control Methods for Medicinal Plant Materials* to fulfill the needs of quality control laboratories and to provide a basis for the development of national standards (7).

A compendium of WHO guidelines and related materials entitled *Quality Assurance of Pharmaceuticals*, Volume 1, was published (8) and the next volume is in print.

Counterfeit drugs

In 1999, WHO published guidelines on the detection and prevention of counterfeit and substandard products (9). Vigilance and reporting of counterfeit products to WHO were enhanced by setting up a liaison officer network among drug regulatory authorities and WHO.

The WHO Certification Scheme

The WHO Certification Scheme on the Quality of Pharmaceutical Products Moving in International Commerce (1) was used in increasing the number of countries as a model instrument to exchange information among authorities in importing and exporting countries. In 1999,

WHO completed a model certificate of analysis to be used for trade of starting materials intended for drug manufacturing, and manufacturers of pharmaceutical substances, excipients, and medicinal products (4).

Comparator product system for equivalence testing

The use of generic pharmaceutical products has increased in many countries. Consequently, drug regulatory authorities are increasingly involved in the assessment of the scientific documentation submitted for marketing approval of generic medicinal products. WHO has stated that the quality, safety, and efficacy of a generic product must comply with the same requirements applied to the innovator (10). Consequently, the therapeutic equivalency and interchangeability with the innovator must be proven with valid comparative in vivo bioequivalence studies. To assist both manufacturers and drug regulators in this, WHO has developed a list of international comparator products. The information on comparator products was collected first from drug regulatory agencies, and a provisional list of candidate products was consulted with the innovator companies concerned. The companies were requested to designate the market and trademark that best represented the quality, manufacturing, and labeling of the product. The current list thus far contains 147 products from the WHO Model List of Essential Drugs, complemented with detailed instructions for use and a decision tree for special situations (11). The list will be updated periodically by WHO.

The WHO Model List of Essential Drugs

The Model List of Essential Drugs was created in 1975, and the most recent assessment and updating of the list took place in 1997 and 1999 (12,13). The following additions to the list were made including aciclovir for use in herpes infections; amoxicillin + clavulanic acid for the treatment of infections resistant to the production of β -lactamase; dextrometorphan for the treatment of cough; ephedrine for treatment of hypotension in spinal anesthesia during delivery; imipenem + cilastatin for the treatment of *Pseudomonas*; ipratropium bromide for the treatment of asthma; metformin for the treatment of noninsulin-dependent diabetes; triclabendazole for the treatment of liver and lung flukes; and nevirapine and zidovudine to reduce or prevent mother-to-child transmission of HIV infection. The value of lipid-lowering agents was discussed; however, because of their high costs, they were not included in the list, with a recommendation that each country make its own decisions for use in high-risk patient populations (12).

Sale and promotion of medical products on the internet

The Internet is a rapidly expanding medium that has many uses and great potential for disseminating and obtaining information regarding a variety of subjects such as new treatments, institutions offering care, and medical products available. Individuals can get this information more quickly and conveniently than in traditional ways, such as in medical journals and textbooks. Organizations, pharmaceutical companies, and individuals post and exchange information about their aims, activities, or products and sometimes offer to sell their products, including medical products. However, have the products been approved for marketing? Are they safe, effective, properly labeled, and of good quality? Reported cases show that dangerous, harmful, unreliable, low-quality, substandard, fake, counterfeit, and mislabeled products have been available on the Internet (14).

The World Health Assembly (the annual summit of Ministries of Health of 193 member states of WHO) recognized in May 1997 the increasing use of electronic communication by the general public to shop and gather information on health education, diseases, treatments, and medical products (15). The Assembly was concerned about reported cases of inappropriate use of the Internet, which had caused a potential hazard for public health, and decided to convene a WHO ad hoc working group to formulate recommendations for action. The working group met in September 1997 in Geneva and consisted of representatives of health and drug regulatory authorities, consumer groups, professional associations, the pharmaceutical industry, experts in legal and ethical matters, marketing, and other interested parties. The group reviewed the situation and made several recommendations regarding cross-border advertising, promotion, and sale of medical products. Based on these recommendations, in May 1998 (16) the World Health Assembly urged all member states to review existing legislation, regulations, and guidelines to ensure that they are applicable and adequate to cover the promotion and sale of medical products over the Internet. Member states were asked to set up monitoring and surveillance systems for the Internet. Collaboration was recommended among countries to identify difficult cases and disseminate information through WHO. The availability of scientific, validated information to consumers by competent health authorities was determined to be necessary.

The Assembly appealed to the pharmaceutical industry, health professional and consumer organizations, and other interested parties to promote the formulation and use of good information practices consistent with the principles

of the WHO Ethical Criteria for Medicinal Drug Promotion; monitor and report problem cases and aspects of cross-border advertising, promotion, and sale of medical products using the Internet; and maintain legal and ethical standards in these activities.

WHO was requested to encourage the international community to formulate self-regulatory guidelines for good information practices and to develop a model guide for member states. The guide should educate and instruct people using the Internet on how to best obtain reliable, independent, and compatible information on medical products through this medium. The WHO model guide, *Medical Products and the Internet*, was published in 1999 (17).

Biologicals

WHO's normative work for biological substances used in medicinal products, including vaccines, blood products, and diagnostic procedures, has resulted in a number of globally applicable specifications, guidelines, and guiding principles. In 1999, WHO published guidelines for the production and quality control of synthetic peptide vaccines for human use to ensure their consistent safety and efficacy (18). The guidelines address control of starting materials including background data on the synthesis of the peptide of interest and control of the manufacturing process and the final product. WHO requirements for tick-borne encephalitis vaccine (inactivated) were formulated to take into account current manufacturing practices and control procedures in place and to give guidance on how these practices and procedures could be updated (18). Requirements for hepatitis B vaccines made by recombinant DNA techniques have been amended to reflect recent developments in assay methodology, as well as for specifications for *Haemophilus* type b conjugate vaccines. Under development were guidelines for standardization and calibration of cytokine immunoassays, and requirements for tetravalent dengue vaccine (live) and oral poliomyelitis vaccine (18).

In 1999, WHO published requirements for production and control of *Haemophilus influenzae* type b (Hib) conjugate vaccine and the acellular pertussis component of monovalent or combined vaccines. The requirements for oral polio vaccine (OPV) were also revised in 1999, with several additions. WHO published new and replacement International Standards and Reference Materials covering a wide range of products (19).

WHO, ICH, and Regional Harmonization

The role of WHO as an observer of the ICH and its contribution as a global, intergovernmental organization to

tripartite ICH activities involving 17 WHO member states have been described in detail (1).

With strong support from WHO, the ICH Steering Committee decided in 1997 to accept the participation of experts from the International Generic Pharmaceutical Alliance (IGPA) and the World Self-Medication Industry (WSMI) in the designated expert working groups (20). This was a first step toward a more comprehensive harmonization procedure. Since 1998, WHO participated as an observer in drafting a guideline for good manufacturing practices for active pharmaceutical ingredients based on preliminary work by the Pharmaceutical Inspection Co-operation Scheme (PIC/S). Australia, China, and India, known for their extensive production of pharmaceutical substances and excipients, were invited to participate in this exercise. The draft guideline is still in development and was an ICH step 2 document in February 2000.

Because Common Technical Document (CTD) (content and format of a new drug application) was accepted as a new ICH topic in 1998, WHO asked a WHO regional adviser for pharmaceuticals from each of the six WHO regions to coordinate among ICH CTD working groups and drug regulatory authorities WHO regions (20). This is an important development because the CTD will undoubtedly have an impact on new drug applications in non-ICH countries and industries. The CTD expert working groups succeeded in 1999 to sign off as step 2 document tables of contents for all three sections of the CTD, namely quality, safety, and efficacy. Significant work remains to be done, and the progress has been slowest in the quality section.

In 1999, the ICH began a major revision of its existing stability testing guideline. To make this guideline practical and useful for all countries, WHO suggested that it should address not only new molecular entities and associated drug products but also stability testing of generic pharmaceutical products in its new revised form. Furthermore, WHO proposed that it should address, in addition to its current coverage of only climatic zones I and II, the remaining critical climatic zones, namely zone III (hot/dry) and zone IV (hot/humid). The WHO guidelines on stability testing of pharmaceutical products containing well-established (generic) drug substances in conventional dosage forms cover all four climatic zones (21).

REFERENCES

1. Idänpään-Heikkilä, J.E. WHO and the Harmonization of Regulatory Requirements for Medicinal Products.

- Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; suppl. 1. Marcel Dekker, Inc.: New York, 1999; 18, 353–360.
2. WHO. International Nonproprietary Names for Pharmaceutical Substances (INN). *WHO Drug Information*; 13, 1999; 1–4, WHO: Geneva, 1998; 12.
3. WHO. *Guidelines on the Use of International Nonproprietary Names (INN) for Pharmaceutical Substances*; WHO/PHARM S/NOM 1570, WHO: Geneva, 1997.
4. WHO. Technical Report Series. *Report of the WHO Expert Committee on Specifications for Pharmaceutical Preparations*; WHO: Geneva, 2000.
5. WHO. *Basic Tests for Drugs, Pharmaceutical Substances, Medicinal Plant Materials and Dosage Forms*; WHO: Geneva, 1998.
6. WHO. *Starting Materials for Pharmaceutical Products: Control and Safe Trade*; WHO/PHARM/98.605, WHO: Geneva, 1998.
7. WHO. *Quality Control Methods for Medicinal Plant Materials*; WHO: Geneva, 1997.
8. WHO. *Quality Assurance of Pharmaceuticals*; WHO: Geneva, 1997; 1.
9. WHO. *Guidelines for the Development of Measures to Combat Counterfeit Drugs*; WHO/EDM/QSM/99.1, WHO: Geneva, 1999.
10. WHO. *Drug Information* **1999**, 13, 158–162.
11. WHO. International Comparator Products for Bioequivalence Testing. *Drug Information* **1999**, 13, 158–162.
12. WHO. Model List (revised in December 1997). *Drug Information* **1998**, 12, 22–35.
13. WHO. Model List (revised in November 1999). *Drug Information* **1999**, 13.
14. Idänpään-Heikkilä, J.E. Marketing Pharmaceuticals on the Internet. *Medicines on the Internet. Communication from the Industry to the Patients*, Nordic Council on Medicines, 52, NLN: Stockholm, 2000, 63–67.
15. WHO. *Cross-Border Advertising, Promotion and Sale of Medical Products Through the Internet*; WHO/WHA 50.4, WHO: Geneva, 1997.
16. WHO. *Cross-Border Advertising, Promotion and Sale of Medical Products Using the Internet*; WHO/WHA 51.9, WHO: Geneva, 1998.
17. WHO. *Medical Products and the Internet, A Guide to Finding Reliable Information*; WHO/EDM/QSM/99.4, WHO: Geneva, 1999.
18. WHO. Technical Report Series, No. 889. *Expert Committee on Biological Standardization*; WHO: Geneva, 1999.
19. WHO. Progress in Biological Standardization. *Drug Information* **1999**, 13, 86–90.
20. Idänpään-Heikkilä, J.E. ICH and the Common Technical Document (CTD). *WHO Drug Information* **1999**, 13, 78–81.
21. WHO. Technical Report Series No. 863, *Expert Committee on Specifications for Pharmaceutical Preparations*. WHO: Geneva, 1996.
22. Idänpään-Heikkilä, J.E. Pharmaceutical Products: The International Harmonization and Collaboration Perspective. *Focus on Pharmaceutical Research, Policy and Law*; Valverde, J.L., Fracchia, G.N., Eds.; IOS Press, 1999, 25–34.

X-RAY POWDER DIFFRACTOMETRY^a

Raj Suryanarayanan

University of Minnesota, Minneapolis, Minnesota, U.S.A.

Suneel Rastogi

Forest Laboratories, Inc., Inwood, New York, U.S.A.

INTRODUCTION

X-rays are electromagnetic radiation lying between ultraviolet and gamma rays in the electromagnetic spectrum. The wavelength of the X-ray region is considered to be between 0.01 and 100 Å (1).

There are two broad applications of X-rays in the characterization of materials: (i) X-ray spectrometry and (ii) X-ray diffractometry. The former technique is used for chemical analysis and has found only limited use in the characterization of pharmaceuticals. On the other hand, X-ray diffractometry, by providing a means for the study of the structure of crystalline materials, is extensively used to characterize pharmaceutical solids. There are two principal applications of X-ray diffractometry. X-ray crystallography is concerned with the structure determination of crystalline phases. Single crystals are usually used for this purpose. On the other hand, in X-ray powder diffractometry, the sample is usually in the form of a powder. X-ray powder diffractometry is recognized as a powerful technique for the identification of crystalline phases. The technique can also be used for the quantitative analyses of solids. This article will be restricted to the principles and applications of X-ray powder diffractometry (XRD) in the characterization of pharmaceutical solids.

Diffraction is a scattering phenomenon. When X-rays are incident on crystalline solids, they are scattered in all directions. In some of these directions, the scattered beams are completely in phase and reinforce one another to form the diffracted beams (1, 2). The Bragg law describes the conditions under which this would occur. It is assumed that a perfectly parallel and monochromatic X-ray beam,

of wavelength λ , is incident on a crystalline sample at an angle θ . Diffraction will occur if:

$$n\lambda = 2d \sin \theta \quad (1)$$

where d = distance between the successive planes in the crystal lattice, expressed in Å, and n = order of reflection (an integer).

X-ray patterns can be obtained using either a powder diffractometer or a camera. Currently, diffractometers find widespread use in the analysis of pharmaceutical solids. The technique is usually nondestructive in nature. The theory and operation of powder diffractometers is outside the scope of this discussion, but these topics have received excellent coverage elsewhere (1–4). Instead, the discussion will be restricted to the applications of X-ray powder diffractometry (XRD) in the analysis of pharmaceutical solids. The United States Pharmacopeia provides a brief but comprehensive introduction to X-ray diffractometry (5). The use of XRD in the physical characterization of pharmaceutical solids (6) and in the characterization of controlled release delivery systems have been discussed earlier (7).

QUALITATIVE ANALYSIS

Since the X-ray diffraction pattern of every crystalline form of a compound is unique, the technique is widely used for the identification and characterization of solid phases. XRD is the technique of choice to identify different polymorphic forms of a compound (Fig. 1) (8). It can also be used to identify the solvated and unsolvated (anhydrous) forms of a compound, provided their lattice structures are different. The technique can also reveal differences in the crystallinity of compounds. The XRD pattern of an amorphous (noncrystalline) compound will consist of one or more broad diffuse halos (see Fig. 2; upper panel) (9).

The United States Pharmacopeia contains the XRD patterns of two anhydrous forms (Form 1 and Form 2) of

^aSome parts of this article have been reproduced from the chapter, "X-ray Powder Diffractometry," written by R. Suryanarayanan in "Physical Characterization of Pharmaceutical Solids," edited by H. Brittain. This was Vol. 70 of *Drugs and Pharmaceutical Sciences*, Editor: J. Swarbrick, Marcel Dekker, Inc., New York, 1995.

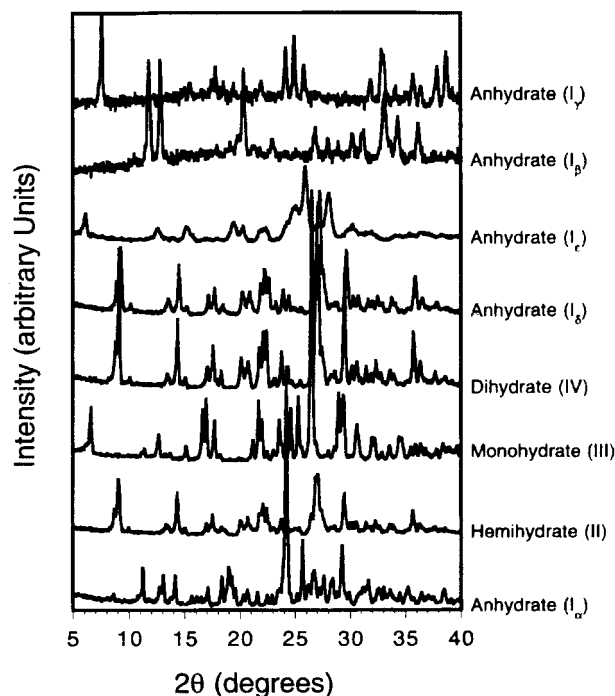


Fig. 1 X-ray powder diffraction (XRD) patterns of the different solid forms of AG337. (From Ref. 8.)

ampicillin, ampicillin trihydrate and amorphous anhydrous ampicillin (5). The powder patterns of these four solid phases reveal pronounced differences, which can form the basis of their identification.

XRD was a powerful tool for the characterization of the different solid phases of AG337 (a 5-substituted quinazolinone; dihydrochloride salt). In addition to several polymorphic forms of the anhydrate ($C_{14}H_{12}N_4OS \cdot 2HCl$), the compound existed as a hemihydrate ($C_{14}H_{12}N_4OS \cdot 2HCl \cdot 0.5H_2O$), a monohydrate ($C_{14}H_{12}N_4OS \cdot 2HCl \cdot H_2O$), and a dihydrate ($C_{14}H_{12}N_4OS \cdot 2HCl \cdot 2H_2O$) (8). The hemihydrate, the dihydrate, and one of the polymorphic forms of the anhydrate (I_8) had identical lattice structures (Fig. 1).

While the amorphous and crystalline solid forms of sucrose can be readily distinguished by XRD (Fig. 2, upper panel), the technique is also capable of revealing the presence of small amounts of crystalline sucrose in the presence of amorphous sucrose (9). This is evident from the XRD patterns of physical mixtures of amorphous and crystalline sucrose wherein the crystalline sucrose content ranged between 1 and 5 wt% (Fig. 2, lower panel).

Six crystalline solid phases of fluprednisolone and an amorphous phase were characterized using XRD, IR spectroscopy, and differential scanning calorimetry (10). Three of these six crystalline phases were anhydrous, two were monohydrates, and one was a tert-butylamine

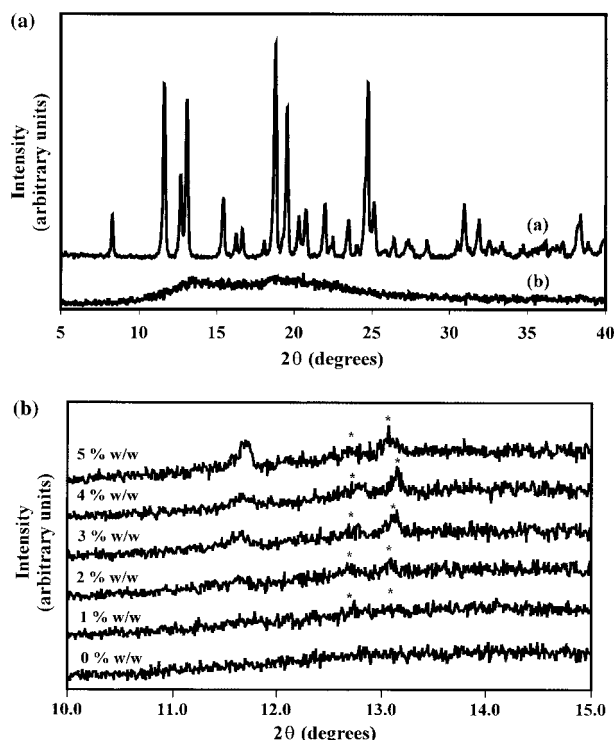


Fig. 2 Upper panel: XRD patterns of (a) crystalline and (b) amorphous sucrose. Lower panel: XRD patterns of the physical mixtures of crystalline and amorphous sucrose. The sucrose content ranged between 0 and 5 wt%. The 12.7 and 13.1°2θ peaks were used for the quantitative analyses. (Bottom panel reproduced from Ref. 9.)

disolvate. The differences in the powder patterns of the phases were readily evident. This study demonstrated the unique ability of XRD for the identification of (a) an anhydrous compound existing in both crystalline and amorphous states, (b) different polymorphic forms of the anhydrate, (c) the existence of solvates, where the solvent of crystallization is water (hydrate) or an organic solvent (in this case, tert-butylamine), and (d) polymorphism in the hydrate.

Reference Diffraction Patterns

The International Centre for Diffraction Data (ICDD, Newtown Square, PA) maintains a collection of single-phase X-ray powder patterns (11). There are separate listings of inorganic and organic compounds. The X-ray diffraction data of ibuprofen, as a representative example is given here (Table 1).

The card pattern contains the Powder Diffraction File (PDF) number (Region 1); quality mark of the data (Region 2); the chemical formula and the specimen name

There are also several indicators of the quality of the data. The highest quality data are given*. To qualify for this mark, the chemistry of the compound must be well characterized. The intensities of the X-ray lines must be measured objectively and instrumentally and there must be no unindexed, space group extinct or impurity lines. Lines with d -spacings of ≤ 2.50 Å must retain at least three significant digits after the decimal point. To qualify for the “i” mark, there can be a maximum of two unindexed, space

group extinct or impurity lines provided none of these belong to the strongest eight. Again there must be no serious systematic errors and lines with d -spacings ≤ 2.00 Å must retain at least three significant digits after the decimal point. If the data are of low precision or if the data are due to a poorly characterized or multiphase system, an “O” mark is assigned. Patterns that do not meet the criteria for “*,” “i,” or “O” are left blank. When the powder pattern is calculated from structural parameters, the pattern is marked “C.” Extensive details about the quality mark guidelines can be found in ICDD publications (12).

Another important database for pharmaceuticals is the Cambridge Structural Database (CSD). Maintained by the Cambridge Crystallographic Data Centre, CSD is a compilation of single-phase data containing structural information including unit cell parameters, crystal system, and space group. The structural information can be used to

Table 1 The card pattern of ibuprofen

1	←	32 - 1723						*	→	2
		C ₁₃ H ₁₈ O ₂	d Å	Int	h k t	d Å	Int	h k t		
3	←	Ibuprofen	14.41 7.24 6.93	85 15 9	100 200 1̄10	2.886 2.792 2.665	4 2 2	3̄13 213 4̄21		
4	←	Rad. CuKα ₁ λ 1.5406 Filter Ni d - sp Cut off Int. Diffractometer I/I _{cor.}	6.33 6.02	15 10	011 1̄11	2.664 2.533	3 6	004 1̄31		
		Ref. Cong. P. Polytechnic Insitute of New York. Brooklyn. New York. USA. JCPDS Grant-in-Aid Report. (1981)	5.34 5.01 4.73	100 45 10	210 2̄11 102	2.504 2.437 2.409	4 3 2	4̄22 3̄23 113		
5	←	Sys. Monoclinic S.G. P2 ₁ /c (14) a 14.667 b 7.899 c 10.731 A 1.8568 C 1.3585 α β 99.46 γ Z 4 mp	4.65 4.55 4.40	15 30 75	2̄02 211 012	2.379 2.280 2.193	4 2 2	502 512 232	→	7
		Ref. lbrd D _x 1.117 D _m SS/FOM F30=33.0 (.0114.80)	4.12 4.06	2 4	3̄10 112					
6	←	Color White Reported by McConnell. Cryst. Struct. Commun., 3 73 (1974) as: a = 14.667, b = 7.886, c = 10.730, β = 99.36, Space Group = P2 ₁ /c, Z=4. Silicon used as internal standard. PSC: mP132. Merck Index. 9th Ed., 4796. To replace 30.1757.	3.973 3.897 3.811 3.664 3.617 3.546 3.466 3.373 3.290 3.219 3.053 3.015	40 4 2 4 4 9 4 1 4 15 3 3	202 3̄02 120 311 400 212 2̄20 2̄21 410 221 320 411					

calculate powder diffraction patterns. The ICDD and the Cambridge Crystallographic Data Centre have a cooperative program. This allows the ICDD to calculate powder patterns for inclusion in the Powder Diffraction File, from the data recorded in the CSD.

Phase Identification in Solid Dosage Forms

In addition to the active ingredient, solid dosage forms usually contain one or more excipients. In such powder mixtures, each crystalline phase produces its pattern independently of the other constituents in the mixture. Thus, the unique advantage of XRD is that it combines absolute specificity with a high degree of accuracy. Shell pioneered the use of XRD in the characterization of pharmaceutical dosage forms (13). The recent advances in instrumentation and software have substantially enhanced the utility of the technique. XRD was used to simultaneously identify the three active ingredients, acetaminophen, aspirin, and caffeine, in a commercially available tablet formulation (Excedrin®). The three active ingredients together constituted 83 wt% of the formulation (14). The XRD pattern contained numerous peaks in the angular range of 7–37° 2 θ (Fig. 3d). In an effort to identify the components in the dosage form, the XRD patterns of acetaminophen, aspirin, and caffeine were obtained (Fig. 3a–c). Acetaminophen could be readily identified by two unique lines with *d*-spacings of 6.48 and 4.92 Å (2 θ values of 13.65 and 18.00°, respectively). At these 2 θ values, the XRD patterns of aspirin and caffeine contained no peaks. Similarly, aspirin could be readily identified by two unique lines with *d*-spacings of 11.54 and

3.95 Å (2 θ values of 7.65 and 22.45°, respectively). Caffeine had an intense line with a *d*-spacing of 7.55 Å (2 θ value of 11.70°) and two intense lines with *d*-spacings of 3.39 and 3.31 Å (2 θ values of 26.25 and 26.85°, respectively). Since at these 2 θ values, peaks due to aspirin and acetaminophen occurred, unambiguous identification of caffeine was not possible. An added complication was that caffeine constituted only 8.9 wt% of the formulation.

In order to identify caffeine, the XRD patterns of acetaminophen and aspirin were selectively subtracted from the XRD pattern of the formulation. The details of the pattern subtraction procedure are described in the literature (14). When the subtracted profile (Fig. 4a) was compared with that of caffeine (Fig. 4b), the high intensity peaks of caffeine at 11.70, 26.25, and 26.85° 2 θ were readily discernible. However, an amorphous halo was observed over the angular range of 18–23° 2 θ . The formulation contained numerous excipients including microcrystalline cellulose, hydroxypropyl methylcellulose, and hydroxypropyl cellulose. The XRD pattern of microcrystalline cellulose exhibited a broad halo over the angular range of 18–25° 2 θ (Fig. 5c). While amorphous halos were also observed in the XRD patterns of hydroxypropyl methylcellulose (Fig. 5b) and hydroxypropyl cellulose (Fig. 5a), their angular range did not match that of the formulation (Fig. 5d). Therefore, microcrystalline cellulose is likely to be the major contributor to the observed halo. Thus, XRD not only permitted simultaneous identification of all the active ingredients in the dosage form, but it also provided information about the excipients in the formulation.

The authors also evaluated the sensitivity of the method, using chlordiazepoxide HCl (hereafter referred to

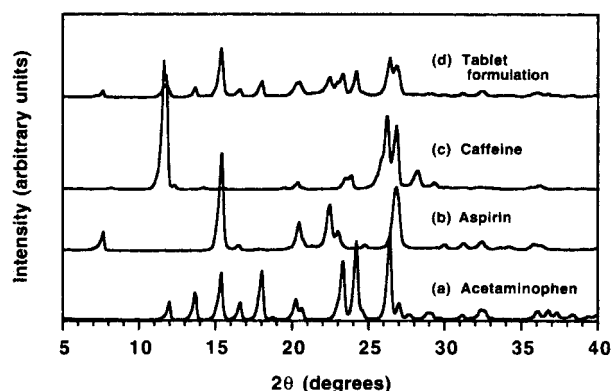


Fig. 3 (a) The XRD pattern of acetaminophen powder. (b) The XRD pattern of aspirin powder. (c) The XRD pattern of caffeine powder. (d) The XRD pattern of a powdered tablet formulation containing acetaminophen, aspirin, and caffeine (Excedrin®). (From Ref. 14.)

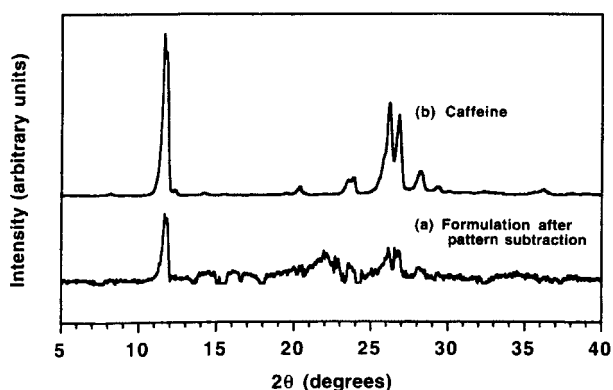


Fig. 4 (a) The residual XRD pattern after proportional subtraction of acetaminophen and aspirin XRD patterns from that of the powdered tablet formulation. (b) The XRD pattern of caffeine powder. (From Ref. 14.)

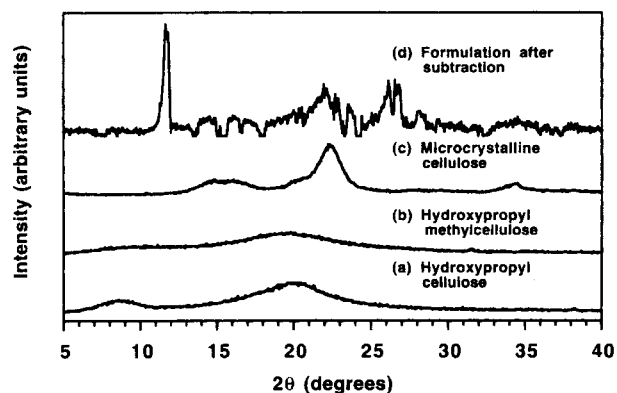


Fig. 5 (a) The XRD patterns of hydroxypropyl cellulose powder. (b) The XRD pattern of hydroxypropyl methylcellulose powder. (c) The XRD pattern of microcrystalline cellulose powder. (d) The residual XRD pattern after proportional subtraction of acetaminophen and aspirin XRD patterns from that of the powdered tablet formulation. (From Ref. 14.)

as chlordiazepoxide) as the model compound (14). While chlordiazepoxide was crystalline (Fig. 6a), microcrystalline cellulose exhibited a broad amorphous halo (Fig. 6c). Identifying chlordiazepoxide was no problem so long as its weight fraction was ≥ 0.10 . When the drug weight fraction was decreased to 0.05, its presence was not readily discernible (Fig. 6b). Using the pattern subtraction technique, the XRD pattern of microcrystalline

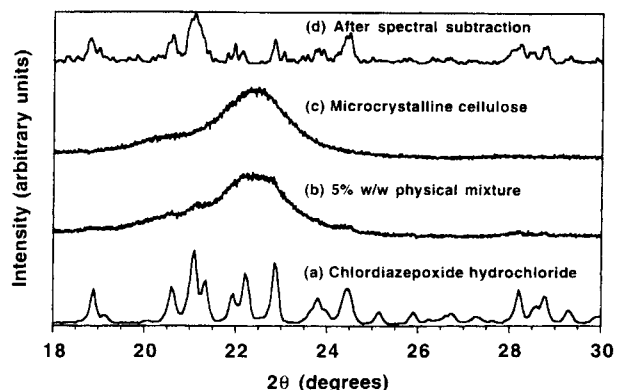


Fig. 6 (a) The XRD pattern of chlordiazepoxide hydrochloride powder. (b) The XRD pattern of a powder mixture of chlordiazepoxide hydrochloride (5 wt%) and microcrystalline cellulose (95 wt%). (c) The XRD pattern of microcrystalline cellulose powder. (d) The residual XRD pattern after proportional subtraction of the microcrystalline cellulose XRD pattern from that of the physical mixture (b). The full scale in this case is different from that of the other three XRD patterns. (From Ref. 14.)

cellulose was subtracted from the XRD pattern of the drug-microcrystalline cellulose mixture (Fig. 6d). This permitted ready identification of chlordiazepoxide (compare Fig. 6d with Fig. 6a).

Phase Transitions Induced During Processing

The components of a dosage form, both active ingredients and excipients, can undergo a variety of physical transformations during pharmaceutical processing and storage. These include polymorphic transformations, alterations in crystallinity, and changes in the state and degree of hydration. XRD is well-suited for analyses of such transformations, provided the analytes are crystalline. In most cases, the dosage forms can be analyzed directly with minimal or no sample pretreatment. There is no need to extract the active ingredient from the dosage form, since it can be usually characterized in presence of the excipient(s). The technique will permit simultaneous identification and quantification of more than one active ingredient in formulations. Finally, XRD can provide quantitative information about the degree of crystallinity.

When theophylline monohydrate was dehydrated, it formed a metastable anhydrous phase, which then transformed to the stable anhydrate (15). The XRD patterns of the two anhydrate phases and the monohydrate were sufficiently different so that all three of them could be simultaneously identified in a sample (Fig. 7; upper panel). Anhydrous theophylline was granulated with an aqueous solution of PVP. During the wet-massing stage, the anhydrate transformed to the monohydrate. When the granules were dried, there was dehydration resulting in the formation of the metastable anhydrate, which then transformed to the stable anhydrate. All three phases were simultaneously monitored by XRD (Fig. 7; lower panel).

DEGREE OF CRYSTALLINITY

Solids may be either crystalline or noncrystalline. The crystalline state is characterized by a perfectly ordered lattice and the noncrystalline (amorphous) state is characterized by a disordered lattice. These represent two extremes of lattice order and intermediate states are possible. The term degree of crystallinity is useful in attempts to quantify these intermediate states of lattice order.

XRD is widely used to determine the degree of crystallinity of pharmaceuticals. The procedure developed by Hermans and Weidinger (16) is based on three assumptions. First, it must be possible to demarcate and

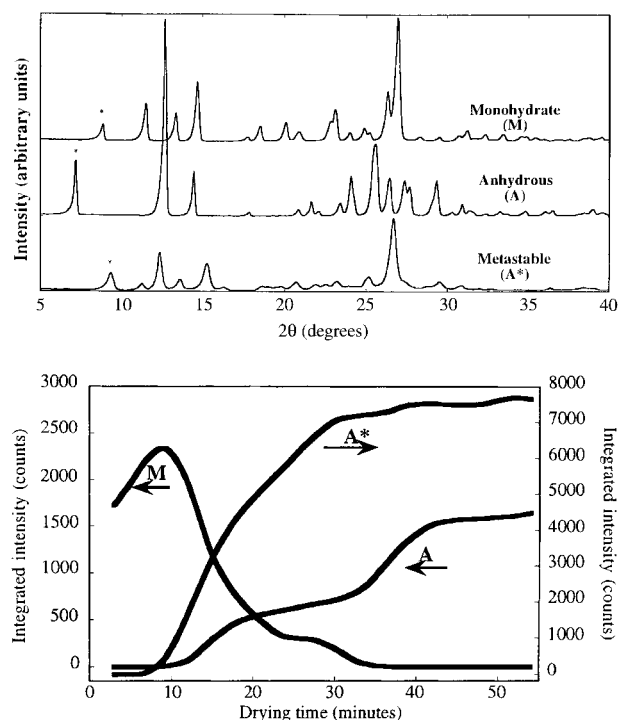


Fig. 7 Upper panel: XRD patterns of theophylline phases. M, A, and A* refer to theophylline monohydrate, stable anhydrous theophylline, and metastable anhydrous theophylline respectively. Lower panel: Phase transitions during the drying of theophylline granules. PVP was the granulating agent. The integrated intensities of the 8.9, 9.4, and 7.0° 2θ peaks unique to M, A*, and A respectively were simultaneously monitored as a function of the drying time. (From Ref. 15.)

measure the crystalline intensity (I_c) and amorphous intensity (I_a) from the powder pattern. Usually, the integrated line intensity (area under the curve), rather than the peak intensity (peak height), is measured. Second, there is a proportionality between the experimentally measured crystalline intensity and the crystalline fraction (x_c) in the sample. Finally, a proportionality exists between the experimentally measured amorphous intensity and the amorphous fraction (x_a) in the sample. The degree of crystallinity (or percent crystallinity), x_{cr} , is given by the expression,

$$x_{cr} = \frac{I_c 100}{I_c + \frac{qI_a}{p}} \quad (2)$$

where p and q are proportionality constants. The values of I_a and I_c can be determined for samples of varying degrees of crystallinity. A plot of the measured values of I_a against those of I_c will result in a straight line, and

the intercepts on the y- and x-axes will provide the intensity values of the 100% amorphous and 100% crystalline materials, respectively. This method was used by Nakai et al. (17) to estimate the degree of crystallinity of lactose that had been milled for various time periods (Fig. 8). If the value of (q/p) is known, the degree of crystallinity of an unknown sample can be calculated from the experimentally determined values of I_c and I_a . The degree of crystallinity of microcrystalline cellulose and digoxin milled for various time periods were also determined (18, 19). The problems and limitations of this method have been discussed in the literature (6).

A number of drugs and excipients exist in the amorphous state. However, the physical instability of amorphous compounds can lead to their crystallization resulting in partially crystalline materials. In recent years, several techniques have been developed to quantify low levels of an amorphous compound in presence of its crystalline counterpart. Using sucrose as a model compound, an XRD method was developed to detect and to quantify crystalline sucrose when it occurred as a mixture with amorphous sucrose (9). The XRD patterns of amorphous and crystalline sucrose were presented earlier (Fig. 2). Standards consisting of amorphous and crystalline sucrose were prepared wherein the crystalline sucrose content ranged between 1 and 5 wt%. The sum of the background subtracted integrated intensities of the 12.7° 2θ (6.94 Å) and 13.1° 2θ (6.73 Å) peaks of sucrose

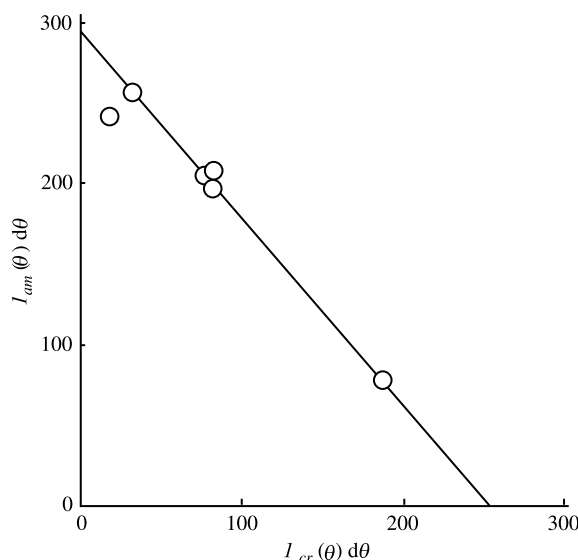


Fig. 8 A plot of the integrated amorphous intensity (I_{am}) as a function of the integrated crystalline intensity (I_c) for lactose samples milled for various time periods. (From Ref.17.)

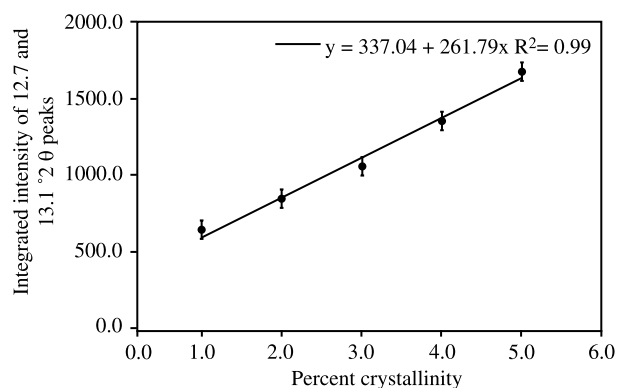


Fig. 9 Plot of the sum of the intensities of the 12.7 and 13.1° 2θ peaks as a function of the weight percent of crystalline sucrose in mixtures of amorphous and crystalline sucrose. (From Ref. 9.)

(these peaks have been marked with an asterisk in Fig. 2) were linearly related to the weight percent of sucrose (Fig. 9).

The limits of detection and quantitation of crystalline sucrose were determined to be 0.9 and 1.8 wt% respectively. Water sorption and FT-Raman spectroscopy also appear to be very sensitive with detection possible down to levels of 1 wt% (20, 21).

PHASE QUANTIFICATION (QUANTITATIVE ANALYSIS)

The Direct Method (No Internal Standard)

Alexander and Klug developed the theoretical basis of the quantitative analyses of powder mixtures (3). Although a powder mixture may be composed of several components, it can be regarded as being composed of just two components: Component 1 (which is the unknown) and the sum of the other components (which is designated as the matrix). The relationship between the intensity of peak *i* of the unknown component (I_{i1}) and its weight fraction (x_1) in the mixture is given by the equation:

$$I_{i1} = \frac{Kx_1}{\rho_1[x_1(\mu_1^* - \mu_M^*) + \mu_M^*]} \quad (3)$$

where K is a constant, ρ_1 is the density of Component 1, and μ_1^* and μ_M^* are the mass attenuation coefficients of Components 1 and the matrix respectively. The mass attenuation coefficient of a substance is the weighted average of the mass attenuation coefficients of its constituent elements.

Let us first consider a two-component mixture, where $\mu_1^* \neq \mu_M^*$. One example is a mixture of the anhydrous

and hydrated forms of a compound. It is first necessary to determine the intensity of peak *i* of a sample consisting of only the analyte $[(I_{i1})_0]$. Next, Equation (3) is modified so that the relative intensity of the XRD peak of Component 1 [expressed as $(I_{i1})/(I_{i1})_0$] is expressed as:

$$\frac{I_{i1}}{(I_{i1})_0} = \frac{x_1\mu_1^*}{x_1(\mu_1^* - \mu_M^*) + \mu_M^*} \quad (4)$$

It is possible to calculate the mass attenuation coefficients of the analyte and the matrix based on their chemical compositions (2). This enables the calculation of the relative intensity $[I_{i1}/(I_{i1})_0]$ as a function of the weight fraction of the analyte in the mixture and thus generate a theoretical curve. This eliminates the need for the preparation of experimental standard curves. An added advantage of this approach is that there is no requirement of an internal standard.

This approach was used to determine the weight fractions of anhydrous carbamazepine and carbamazepine dihydrate when they occurred as mixtures (22). Based on the mass attenuation coefficients of the anhydrate and the dihydrate, the intensity ratios $[I_{i1}/(I_{i1})_0]$ were calculated as a function of the anhydrate content in the mixture (the line in Fig. 10). These were in good agreement with the experimentally obtained values of $[I_{i1}/(I_{i1})_0]$.

A simpler system is a two component mixture, where $\mu_1^* = \mu_M^*$. This occurs when the analyte and the matrix have the same molecular formula, as in polymorphic mixtures. In such cases, Equation (4) reduces to:

$$\frac{I_{i1}}{(I_{i1})_0} = x_1 \quad (5)$$

Therefore a plot of the relative intensity as a function of the analyte weight fraction would be linear. This approach was used to determine the weight fractions of anhydrous α -carbamazepine and β -carbamazepine when they occurred as mixtures (23). While studying the polymorphism of 1,2-dihydro-6-neopentyl-2-oxonicotinic acid, Chao and Vail (24) used XRD to quantify Form I in mixtures of Forms I and II. They estimated that Form I levels as low as 0.5 wt% can be determined by this technique. Similarly the α -inosine content in a mixture consisting of α - and β -inosine was achieved with a detection limit of 0.4 wt% for α -inosine (25).

Internal Standard Method

In this method, an internal standard is added, and its weight fraction is maintained constant in all the mixtures. This method was used to simultaneously quantify the S(+)-enantiomer and the racemic compound of ibuprofen (26).

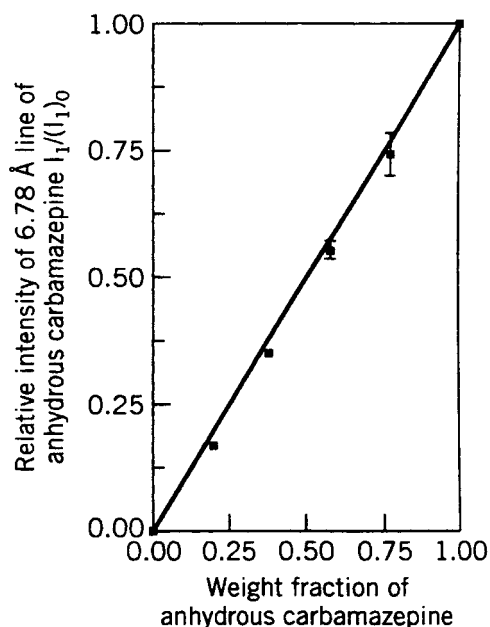


Fig. 10 The relative intensity of the 6.78 Å line of anhydrous carbamazepine (β -form) as a function of its weight fraction in binary mixtures of anhydrous carbamazepine (β -form) and carbamazepine dihydrate. The line is based on theoretical values, while the data points are experimental measurements. (From Ref. 22.)

Selection of Internal Standard

Success in quantitative XRD may hinge on the selection of an appropriate Internal Standard. The following properties are desired in an internal standard (13). (i) The compound must have high crystal symmetry so that strong, but few, diffraction peaks are produced. (ii) The high-intensity lines in the analyte and the internal standard that are to be used for quantitative purposes should not overlap with one another, but they should be close to each other. (iii) The density of the internal standard should be close to that of the system ingredients so that homogeneous mixtures can be prepared for analysis. (iv) The internal standard must be chemically stable in the experimental system. When dealing with organic pharmaceutical systems, an organic internal standard is preferred. Since the XRD patterns of organic compounds usually contain numerous lines, it might not be possible to identify lines unique to the analyte and the internal standard that are completely separated from one another. Therefore, inorganic compounds such as corundum (α - Al_2O_3), silicon, lithium fluoride, and zinc oxide are often used as internal standards. Methenamine, an organic compound, has several of the properties desired in an internal standard (13). However, it has not found widespread use in the analyses of pharmaceuticals.

Whole Powder Pattern Analyses

In the pharmaceutical community, quantitative analyses has conventionally been based on the intensity of a characteristic peak of the analyte. It is now recognized that phase quantification will be more accurate if it is based on the *entire* powder pattern (27, 28). This forms the basis for the whole-powder-pattern analyses method developed in the last few decades. Of the available methods, the Rietveld method is deemed the most powerful since it is based on structural parameters. This is a whole-pattern fitting least-squares refinement technique that has also been extensively used for crystal structure refinement and to determine the size and strain of crystallites.

Simultaneous quantitative analyses, of complex pharmaceutical mixtures, has been accomplished by the Rietveld method (29). Using lithium fluoride as an internal standard, mixtures consisting of anhydrous β -carbamazepine, anhydrous α -carbamazepine, and carbamazepine dihydrate were subjected to quantitative analyses (Table 2). When the analyte concentration was high ($\geq 20\%$), the method was accurate, with a relative error $< \pm 5\%$. However, when the analyte concentration was low ($< 5\%$), the method lacked accuracy.

NONAMBIENT XRD

Conventionally, XRD patterns are obtained at room temperature under ambient conditions. Variable temperature XRD is a technique where XRD patterns are obtained while the sample is subjected to a controlled temperature program. It is also possible to control the environment and maintain the sample under the desired relative humidity. Recent advances in commercially available instrumentation have greatly facilitated the study of pharmaceutical systems under nonambient conditions. Experiments can be carried out both at elevated temperatures and under subambient conditions.

High Temperature XRD

In the characterization of pharmaceutical systems, high temperature XRD has been used extensively. The phase transitions of theophylline monohydrate were discussed earlier in the section dealing with Phase Transitions Induced During Processing (Fig. 7). High temperature XRD was used to investigate the dehydration of theophylline monohydrate (Fig. 11). This technique revealed that dehydration resulted in a metastable

Table 2 Measured composition of carbamazepine mixtures and the residuals obtained by the Rietveld analysis

Actual composition of carbamazepine (wt %):				Measured composition of carbamazepine (wt %):				<i>R</i> ^a _{wp} (%)
α	β	Dihydrate	Lithium fluoride	α	β	Dihydrate	Lithium fluoride	
	80		20		79		21	20.0
	80		20		79		21	22.7
	80		20		80		20	20.7
	79.6	0.4	20		78	1	21	21.5
	79.2	0.8	20		80	1	19	21.4
	79.2	0.8	20		78	2	21	20.9
	79.2	0.8	20		78	2	21	21.4
0.4	79.6		20	<1	79		20	20.7
0.4	79.6		20	<1	79		21	21.3
0.4	79.6		20	<1	78		22	20.9
0.8	79.2		20	1	78		21	20.9
0.4	78.8	0.8	20	<1	77	1	21	21.3
0.8	78.8	0.4	20	<1	78	1	21	20.1
4	72	4	20	3	73	5	19	15.5
4	72	4	20	2	71	7	19	26.9
4	72	4	20	2	71	7	19	26.9
4	40	36	20	2	41	36	20	22.3

^a Weighted profile residual.
(From Ref. 29.)

anhydrous phase, which then transformed to the stable anhydrate (30). Theophylline has been in widespread use for many years and its solid-state properties have been the subject of numerous investigations. Without high temperature XRD, it is unlikely that this metastable phase would have been identified.

Kinetics of Solid-State Reactions

Solid-state reactions in pharmaceuticals can be broadly classified into chemical decompositions and physical transformations. The chemical decomposition of drugs has received adequate attention in the pharmaceutical literature. The active ingredient in a solid dosage form can also undergo a variety of physical transformations including polymorphic transitions, alterations in degree of crystallinity, and alterations in degree of solvation. These transformations can affect pharmaceutically important properties such as solubility, stability, powder flow, and tableting behavior, which ultimately can cause variations in the performance of the product. There are strict pharmacopeial standards, which govern not only the chemical purity of drugs and excipients, but also the drug content in dosage forms. On the other hand, very few compounds have pharmacopeial standards that govern their physical form. However, it is increasingly recognized that the physical form of a compound and physical transformations in a dosage form can profoundly affect product performance. An added complication is that

conventional analytical techniques such as HPLC, which are used for monitoring chemical decomposition reactions are unsuitable to monitor solid-state physical transformations. This is because of their inability to distinguish between the different solid forms of a compound. Both chemical decomposition as well as physical transitions can occur simultaneously during the processing or storage of solid formulations. Therefore, it would be ideal if an analytical technique can simultaneously monitor both types of transformations. XRD is one such technique.

Conventionally, thermoanalytical techniques, specifically differential scanning calorimetry and thermogravimetric analysis, have been used to investigate the kinetics of solid-state reactions. These techniques have several limitations. They do not unambiguously identify crystalline phases and are incapable of providing quantitative information about the degree of crystallinity. Intermediate phases, if formed, may not be readily identified. The techniques may therefore be unsuitable for discerning the reaction mechanism. When two or more gaseous products are simultaneously evolved, thermal analysis alone is not very helpful in the study of reaction kinetics. In such cases, thermal analysis should be used in conjunction with other techniques for evolved gas analysis (e.g. infrared spectroscopy, mass spectroscopy, or gas chromatography). It is often realized that solid-state reactions take place through complex mechanisms involving intermediates and the same end product is obtained through multiple pathways. In such cases, thermoanalytical

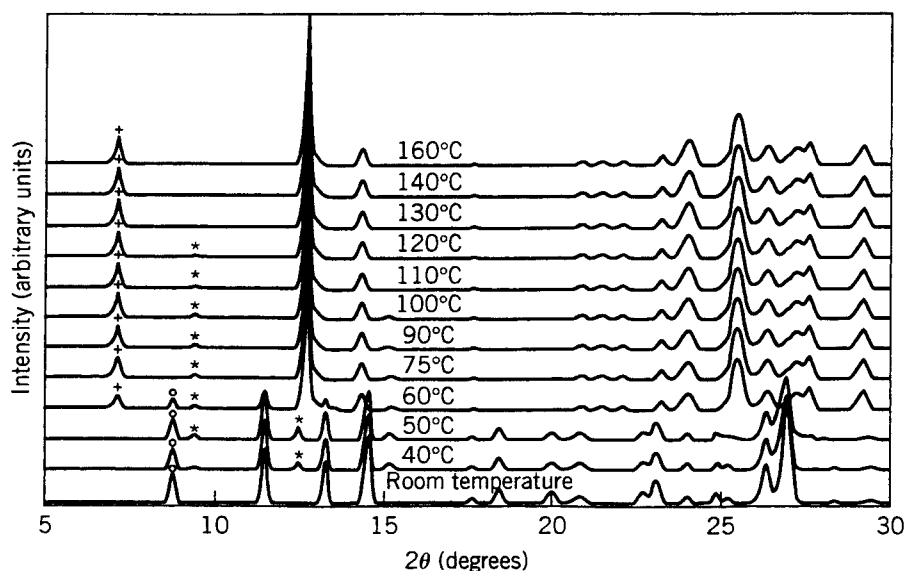


Fig. 11 High temperature XRD of theophylline monohydrate. XRD patterns were obtained at the temperatures indicated in the figure. The “*,” “+,” and “O” marks indicate peaks unique to metastable anhydrous theophylline (referred to as A* in Fig. 7), stable anhydrous theophylline (A in Fig. 7), and theophylline monohydrate (M in Fig. 7) respectively. (Adapted from Ref. 30.)

techniques yield complex profiles because of overlapping thermal events. Often, it is difficult to separate the reaction steps and study the kinetics of the individual steps. XRD does not suffer from these limitations and can therefore serve as an excellent complement to thermoanalytical techniques. Recent advances in instrumentation have enabled simultaneous and independent control of the temperature and the water vapor pressure in the sample chamber.

The dehydration kinetics of theophylline monohydrate ($C_7H_8N_4O_2 \cdot H_2O$) and ampicillin trihydrate ($C_{16}H_{19}N_3O_4S \cdot 3H_2O$) were studied by XRD (31). Dehydration of theophylline monohydrate resulted in a crystalline anhydrous phase, while the ampicillin trihydrate formed an amorphous anhydrate. In the case of theophylline, simultaneous quantification of both the monohydrate and the anhydrate was possible. By carrying out the reaction at several temperatures, the activation energy for the dehydration reaction was determined.

The enantiomers of pseudoephedrine were observed to react in the solid state to form the racemic compound (32). While the powder patterns of the enantiomers were identical, the racemic compound exhibited a different powder pattern. XRD permitted the simultaneous quantification of the disappearance of the enantiomers and the subsequent appearance of the racemic compound. The rate of disappearance of the enantiomers followed a diffusion-controlled reaction model. During the kinetic experiment, the sum of the weight fractions of the enantiomers and the racemic compound progressively decreased from an initial value of unity, suggesting the formation of an intermediate noncrystalline phase. This was confirmed by a steady increase in the background counts in the powder pattern.

The sweetener aspartame exists as a hemihydrate ($C_{14}H_{18}N_2O_5 \cdot 0.5H_2O$; ASH), under ambient conditions. When heated, ASH converted to aspartame anhydrate (ASA), which on further heating decomposed to form a diketopiperazine (DKP) derivative (33). The XRD patterns of ASH, ASA, and DKP showed pronounced differences (Fig. 12). XRD was used to simultaneously quantify the (i) disappearance of ASH and appearance of ASA in the first reaction and (ii) disappearance of ASA and appearance of DKP in the second reaction. For studying the kinetics of the first reaction, the peaks unique to ASH at 15.9 and $16.4^\circ 2\theta$ and the $17.1^\circ 2\theta$ peak of ASA were used. For the second reaction, the sum of the integrated intensities of the peaks at 10.2 , 11.0 , and $11.8^\circ 2\theta$ of ASA and the $13.0^\circ 2\theta$ peak of DKP were used. While the dehydration of ASH appeared to follow first-order kinetics, the cyclization of ASA was a nucleation-controlled process. Figs. 13 and 14 contain the dehydration and cyclization data at 118 and $180^\circ C$ respectively. Since the concentrations of the

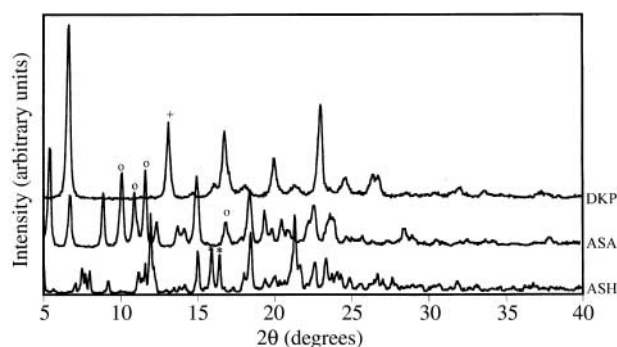


Fig. 12 XRD patterns of aspartame hemihydrate (ASH), aspartame anhydrate (ASA) and diketopiperazine derivative (DKP). “*,” “o” and “+” denote the peaks unique to aspartame hemihydrate, aspartame anhydrate, and diketopiperazine derivative respectively. (From Ref. 33.)

crystalline reactant as well as the product were simultaneously monitored, mass balance calculations of the crystalline phases were possible at each time point. The reaction rate constants were obtained at several temperatures, which permitted the calculation of the activation energies of dehydration and cyclization from the Arrhenius plots.

Drug–Excipient Interactions

XRD is an excellent technique to study drug–excipient interactions, provided the drug and the excipients are crystalline. In a solid mixture, the powder pattern of each crystalline phase is produced independently of the other constituents. Thus the diffraction pattern of a powder mixture will be the summation of the diffraction patterns of the individual constituents. If the drug–excipient interaction results in a crystalline product, this will be characterized by the appearance of new peaks in the powder pattern. However, if the interaction results in an amorphous product, this will become evident from the broad halos in the pattern. Thus, irrespective of the nature of the product phase, XRD is capable of revealing such interactions.

Low Temperature XRD

The use of a cooling accessory permits XRD patterns to be obtained under subambient conditions. In pharmaceutical systems, the greatest utility of the technique is to monitor the crystallization of solutes in frozen solutions. Conventionally, differential scanning calorimetry has been the most popular technique for the characterization of frozen systems. However, as mentioned earlier, this

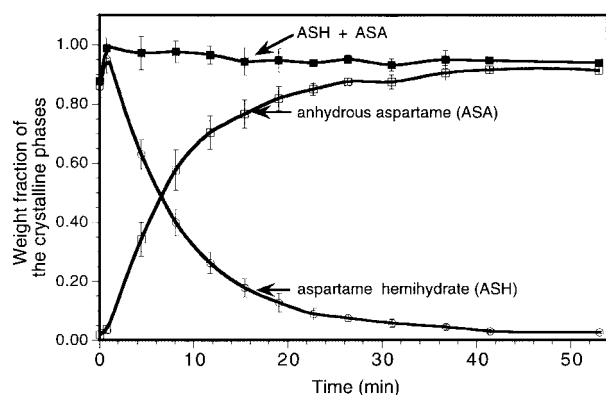


Fig. 13 Weight fractions of ASH and ASA as a function of time following storage of ASH at 118°C. The curves are drawn only to assist in visualizing the trends. (From Ref. 33.)

technique has some drawbacks: (i) It does not enable direct identification of crystalline solid phase(s). Moreover, it is difficult to draw any definitive conclusions about the degree of crystallinity. (ii) The interpretation of DSC curves is very difficult if there are overlapping thermal events. Low temperature XRD was found to be an excellent complement to differential thermal analysis in the characterization of water–glycine–sucrose ternary systems (34).

When sodium nafcillin solutions were frozen, the solute did not crystallize (35). However, annealing caused solute crystallization. It was also possible to monitor the amount of sodium nafcillin crystallizing as a function of annealing time. XRD was also useful to characterize the behavior of buffer solutions. While disodium hydrogen phosphate crystallized when an aqueous solution was frozen, the presence of monosodium hydrogen phosphate inhibited the crystallization of the former.

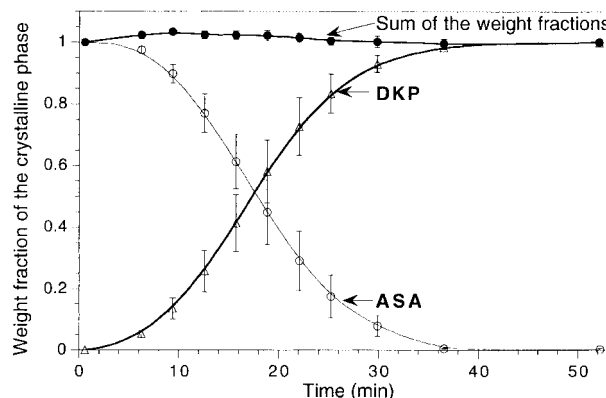


Fig. 14 Weight fractions of ASA and DKP as a function of time, following storage of ASA at 180°C. (From Ref. 33.)

In situ freeze-drying

The low temperature stage of the X-ray powder diffractometer was modified so that the sample chamber could be evacuated. As a result, the entire freeze-drying process was carried out *in situ* in the sample chamber of the XRD. The advantage of such a set-up is that the alterations in the solid state *during* the various stages of the freeze-drying process (cooling, primary, and secondary drying) were monitored (36). As mentioned earlier, annealing of frozen aqueous solutions of sodium nafcillin resulted in the crystallization of sodium nafcillin hydrate (the hydrate stoichiometry was not known). During the primary drying, there was partial dehydration resulting in the formation of partially crystalline sodium nafcillin hemihydrate (Fig. 15). It was possible to simultaneously monitor the disappearance of sodium nafcillin hydrate as well as the appearance of sodium nafcillin hemihydrate (Fig. 16). XRD revealed that this reaction was complete in about 40 min. The technique was also used to monitor the phase transitions during the freeze-drying of mannitol.

Implications in the Design of Freeze-Dried Formulations

Most solid-state characterization studies investigate the starting material and the freeze-dried end product. It will be useful and relevant to study alterations in the solid state *during* the different stages of the freeze-drying process, and this is enabled by low temperature XRD. The physical characterization of the phases at different stages of freeze-drying will not only aid in the optimization of the freeze-drying cycle, but will also facilitate the production of formulations that are physically and chemically stable.

ISSUES AND CHALLENGES IN XRD

Most materials of pharmaceutical interest are organic compounds. Jenkins has comprehensively discussed the problems associated with the X-ray diffractometric analysis of organic materials (37): 1) The unit cells of organic compounds are often large, resulting in low angle lines (large d-spacings). The accuracy of d-spacing data tends to decrease as the d-spacing increases, particularly when the d-spacing value is higher than 8 Å. 2) The unit cells are often of low symmetry resulting in complex powder patterns. 3) Since organic compounds usually have low mass attenuation coefficients, the X-rays can penetrate

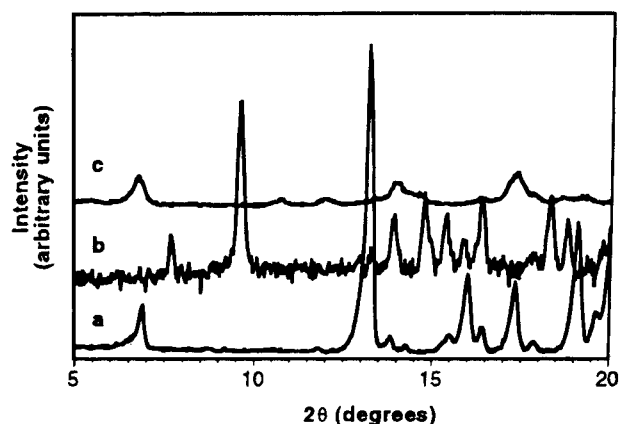


Fig. 15 The XRD patterns of (a) “as is” sodium nafcillin, (b) frozen solution of sodium nafcillin (40 wt%) annealed at -4°C for 1 h, and (c) after freeze-drying the annealed system. The “as is” sodium nafcillin was a monohydrate. Annealing is believed to result in the crystallization of a higher hydrate of sodium nafcillin (unknown stoichiometry). The freeze-dried material was partially crystalline sodium nafcillin hemihydrate. (From Ref. 35.)

deep into the specimen leading to large beam transparency errors. 4) Finally, organic compounds are especially prone to preferred orientation (preferred orientation is discussed in the next section).

Sample Preparation

There are many different methods of sample preparation. Most of the sample holders contain a cavity into which the

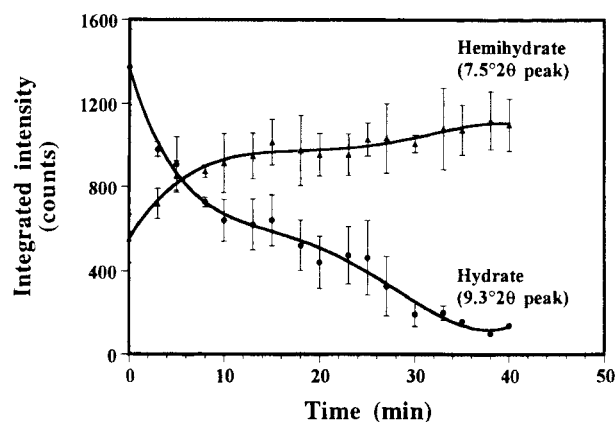


Fig. 16 Real time monitoring of the phase changes during the freeze-drying of sodium nafcillin by in situ XRD. The disappearance of sodium nafcillin hydrate (“Hydrate”) and the appearance of sodium nafcillin hemihydrate (“Hemihydrate”) were simultaneously quantified. The curves were drawn to assist in visualizing the trends. (From Ref. 36.)

powder sample is filled, usually from the top. However, the side-loading (also referred to as side-drift) method is considered the best packing method (1, 38). Unfortunately, many commercial sample holders do not permit sample filling by this method. As a result, holders have been specially fabricated for this purpose (23). Since organic compounds predominantly consist of atoms with low atomic numbers, there will be significant penetration of X-rays. This can result in peak displacement as well as broadening. Detailed information on sample preparation is available in the literature (4).

When only a very small amount of powder sample is available, the use of a low-background (also referred to as zero background) support is advisable. The support is a quartz or silicon plate cut along a nondiffracting crystallographic direction, which is then polished (1). This results in a very low background and improves the signal-to-noise ratio and enables XRD patterns to be obtained with extremely small amounts of sample.

Sources of Error

There are numerous sources of error in XRD. Two of these are discussed hereafter.

Preferred orientation

Reliable and reproducible results can only be obtained if the sample is prepared carefully. When a sample is packed in an X-ray holder, the distribution of crystal orientations can be nonrandom, and this condition is referred to as preferred orientation. Jenkins and Snyder (1) have elegantly illustrated this point in Fig. 17. Atoms bond together, in a unique manner, to form unit cells. If the unit cells are grouped so as to have long-range order, a crystalline material is obtained. In the absence of long-range order, the material is amorphous (noncrystalline). A powder sample is composed of crystallites. When the crystallites in a sample are oriented in some particular (nonrandom) manner, they are said to exhibit preferred orientation (1). Ideally, random orientation of individual crystallites is desired. It is recognized that preferred orientation is a commonly encountered condition in powdered samples (2).

Grinding the sample and thereby reducing the particle size can decrease preferred orientation. However, grinding can cause lattice disorder (decrease in crystallinity) and also induce other undesirable transitions, such as polymorphic transformations. Moreover, when the particle size is very small, particle size induced peak broadening can be observed. This effect, observed when the particle size is $<10,000 \text{ \AA}$ ($1 \mu\text{m}$), is simply a consequence of the

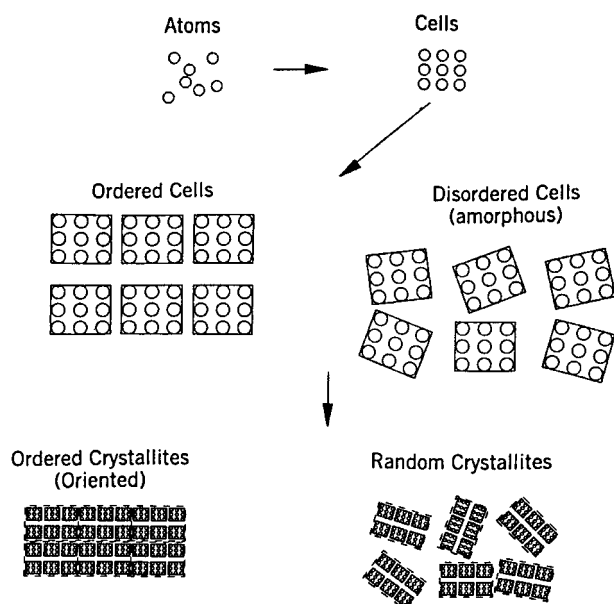


Fig. 17 The structural makeup of solids. (From Ref. 1.)

limited number of planes that are diffracting X-rays (Fig. 18). The particle (crystallite) size is related to the X-ray line width according to the Scherrer equation (1). It is important to recognize the fact that line broadening can be brought about by a loss in crystallinity as well as a decrease in particle size. In the pharmaceutical literature, there is a tendency to automatically attribute line broadening to loss in crystallinity. Such a conclusion, while usually correct, can be occasionally erroneous.

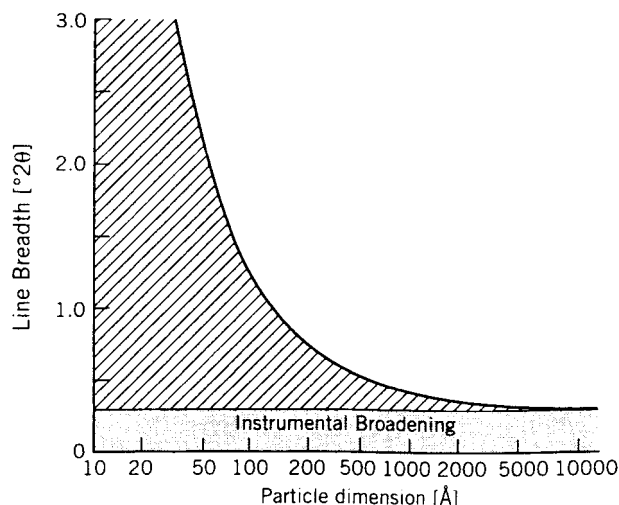


Fig. 18 The width of X-ray lines as a function of particle dimension. (From Ref. 1.)

Measurement of line intensity

When measuring intensity, the integrated line intensity (peak area) and not the maximum intensity (peak height) must be measured. Variations in lattice strain and particle size can significantly influence the line shape, but their effect on the integrated intensity will generally be minimal (2).

Other sources of error

There are numerous other sources of error that are particularly relevant while conducting quantitative XRD studies. These are discussed in detail in the literature (Refs. 1–3, 6).

ACKNOWLEDGMENT

We thank Mr. Rahul Surana for his assistance in the preparation of the manuscript.

REFERENCES

1. Jenkins, R.; Snyder, R.L. *Introduction to X-Ray Powder Diffractometry*; Wiley: New York, NY, 1996.
2. Cullity, B.D. *Elements of X-Ray Diffraction*, 2nd Ed.; Addison-Wesley: Reading, MA, 1978.
3. Klug, H.P.; Alexander, L.E. *X-Ray Diffraction Procedures for Polycrystalline and Amorphous Materials*, 2nd Ed.; Wiley: New York, NY, 1974.
4. Bührke, V.E.; Jenkins, R.; Smith, D.K. *A Practical Guide for the Preparation of Specimens for X-Ray Fluorescence and X-Ray Diffraction Analysis*; Wiley-VCH: New York, NY, 1998.
5. *The United States Pharmacopeia*; 24th Rev. U.S. Pharmacopeial Convention: Rockville, MD, 1999.
6. Suryanarayanan, R. *X-Ray Powder Diffractometry. Physical Characterization of Pharmaceutical Solids*; Brittain, H.G., Ed.; Marcel Dekker, Inc.: New York, NY, 1995.
7. Rastogi, S.; Suryanarayanan, R. Characterization of Delivery Systems—X-Ray Powder Diffractometry. *The Encyclopedia of Controlled Drug Delivery*; Mathiowitz, E., Ed.; Wiley: New York, NY, 1999.
8. Rastogi, S.; Zamansky, I.; Roy, S.; Tyle, P.; Suryanarayanan, R. Solid-State Phase Transitions of AG337, an Antitumor Agent. *Pharm. Dev. Technol.* **1999**, *4* (4), 623–632.
9. Surana, R.; Suryanarayanan, R. Quantitation of Crystallinity in Substantially Amorphous Pharmaceuticals and Study of Crystallization Kinetics by X-Ray Powder Diffractometry. *Powder Diff.* **2000**, *15* (1), 2–6.
10. Halebian, J.K.; Koda, R.T.; Biles, J.A. Isolation and Characterization of Some Solid Phases of Fluprednisolone. *J. Pharm. Sci.* **1971**, *60* (10), 1485–1488.
11. McClune, W.F., Ed. *Powder Diffraction File, Inorganic Phases, Organic and Organometallic Phases Search*

- Manual*; International Centre for Diffraction Data: Newtown Square, PA, 1993.
12. Powder Diffraction File: Set 43. *Inorganic and Organic Databook*; McClune, W.F., Ed.; International Centre for Diffraction Data: Newtown Square, PA, 1993.
 13. Shell, J.W. X-Ray and Crystallographic Applications in Pharmaceutical Research, II. *J. Pharm. Sci.* **1963**, *52*, 24–39.
 14. Phadnis, N.V.; Cavatur, R.K.; Suryanarayanan, R. Identification of Drugs in Pharmaceutical Dosage Forms by X-Ray Powder Diffractometry. *J. Pharm. Biomed. Anal.* **1997**, *15* (7), 929–943.
 15. Tank, J. *Changes in the Solid-State of Theophylline Upon Aqueous Wet Granulation*; M.S. Thesis Department of Pharmaceutics, University of Minnesota: Minneapolis, MN, 1997.
 16. Hermans, P.H.; Weidinger, A. Quantitative X-Ray Investigation on the Crystallinity of Cellulose Fibres. A Background Analysis. *J. Appl. Phys.* **1948**, *19*, 491–506.
 17. Nakai, Y.; Fukuoka, E.; Nakajima, S.; Morita, M. Physicochemical Properties of Crystalline Lactose. I. Estimation of the Degree of Crystallinity and the Disorder Parameter by an X-Ray Diffraction Method. *Chem. Pharm. Bull.* **1982**, *30* (5), 1811–1818.
 18. Nakai, Y.; Fukuoka, E.; Nakajima, S.; Hasegawa, J. Crystallinity and Physical Characteristics of Microcrystalline Cellulose. *Chem. Pharm. Bull.* **1977**, *25* (1), 96–101.
 19. Black, D.B.; Lovering, E.G. Estimation of the Degree of Crystallinity in Digoxin by X-Ray and Infrared Methods. *J. Pharm. Pharmacol.* **1977**, *29* (11), 684–687.
 20. Saleki-Gerhardt, A.; Ahlneck, C.; Zografi, G. Assessment of Disorder in Crystalline Solids. *Int. J. Pharm.* **1994**, *101*, 237–247.
 21. Taylor, L.S.; Zografi, G. Quantitative Analysis of Crystallinity using FT-Raman Spectroscopy. *Pharm. Res.* **1998**, *15*, 755–761.
 22. Suryanarayanan, R. Determination of the Relative Amounts of Anhydrous Carbamazepine and Carbamazepine Dihydrate in a Mixture by Powder X-Ray Diffractometry. *Pharm. Res.* **1989**, *6* (12), 1017–1024.
 23. Suryanarayanan, R. Determination of the Relative Amounts of α -Carbamazepine and β -Carbamazepine in a Mixture by Powder X-Ray Diffractometry. *Powder Diff.* **1990**, *5* (3), 155–159.
 24. Chao, R.S.; Vail, K.C. Polymorphism of 1,2-Dihydro-6-Neopentyl-2-Oxonicotinic Acid: Characterization, Interconversion, and Quantitation. *Pharm. Res.* **1987**, *4* (5), 429–432.
 25. Doff, D.H.; Brownen, F.L.; Corrigan, O.I. Determination of α -Impurities in the β -Polymorph of Inosine using Infrared Spectroscopy and X-Ray Powder Diffraction. *Analyst (London)* **1986**, *111* (2), 179–182.
 26. Phadnis, N.V.; Suryanarayanan, R. Simultaneous Quantification of an Enantiomer and the Racemic Compound of Ibuprofen by X-Ray Powder Diffractometry. *Pharm. Res.* **1997**, *14* (9), 1176–1180.
 27. Anwar, J. Analysis of Time-Resolved Powder Diffraction Data using a Pattern-Decomposition Method with Restraints. *J. Appl. Crystallogr.* **1993**, *26* (3), 413–421.
 28. Rietveld, H.M. Profile Refinement Method for Nuclear and Magnetic Structures. *J. Appl. Crystallogr.* **1969**, *2*, 65–71.
 29. Iyengar, S.S.; Phadnis, N.V.; Suryanarayanan, R. Quantitative Analyses of Complex Pharmaceutical Mixtures by the Rietveld Method. *Powder Diff.* *in press*.
 30. Phadnis, N.V.; Suryanarayanan, R. Polymorphism in Anhydrous Theophylline-Implications for the Dissolution Rate of Theophylline Tablets. *J. Pharm. Sci.* **1997**, *86* (11), 1256–1263.
 31. Shefter, E.; Fung, H.-L.; Mok, O. Dehydration of Crystalline Theophylline Monohydrate and Ampicillin Trihydrate. *J. Pharm. Sci.* **1973**, *62* (5), 791–794.
 32. Duddu, S.P.; Khin-Khin, A.; Grant, D.J.W.; Suryanarayanan, R. A Novel X-Ray Powder Diffractometric Method for Studying the Reaction between Pseudoephedrine Enantiomers. *J. Pharm. Sci.* **1997**, *86* (3), 340–345.
 33. Rastogi, S.; Zakrzewski, M.; Suryanarayanan, R. Investigation of Solid-State Reactions using Variable Temperature X-Ray Powder Diffractometry. I. Aspartame Hemihydrate. *Pharm. Res.* **(2001)**, *18* (3), 267–273.
 34. Shalae, E.Y.; Malakhov, D.V.; Kanev, A.N.; Kosyakov, V.I.; Tuzikov, F.V.; Varaksin, N.A.; Vavilin, V.I. Study of the Phase Diagram Water Fraction of the System Water–Glycine–Sucrose by DTA and X-Ray Diffraction Methods. *Thermochim. Acta* **1992**, *196* (1), 213–220.
 35. Cavatur, R.K.; Suryanarayanan, R. Characterization of Frozen Aqueous Solutions by Low Temperature X-Ray Powder Diffractometry. *Pharm. Res.* **1998**, *15* (2), 194–199.
 36. Cavatur, R.K.; Suryanarayanan, R. Characterization of Phase Transitions during Freeze-Drying by in Situ X-Ray Powder Diffractometry. *Pharm. Dev. Technol.* **1998**, *3* (4), 579–586.
 37. Jenkins, R. New Directions in the X-Ray Diffraction Analysis of Organic Materials. *Adv. X-Ray Anal.* **1992**, *35*, 653–660.
 38. McMurdie, H.F.; Morris, M.C.; Evans, E.H.; Paretzkin, B.; Wong-Ng, W.; Hubbard, C.R. Methods of Producing Standard X-Ray Diffraction Powder Patterns. *Powder Diff.* **1986**, *1* (1), 40–43.

ZETA POTENTIAL

Luk Chiu Li

Abbott Laboratories, Abbott Park, Illinois, U.S.A.

Youqin Tian

Alcon Research, Ltd., Fort Worth, Texas, U.S.A.

INTRODUCTION

Dispersion systems represent an important class of pharmaceutical dosage forms such as emulsions, suspensions, microspheres, liposomes, and nanoparticles. The medium of these systems is mainly aqueous in nature and the dispersed phase can be either solid particles or immiscible liquid droplets. Electrical charges are developed by several mechanisms at the interface between the dispersed phase and the aqueous medium (1). The two most common mechanisms are the ionization of surface functional groups and the specific adsorption of ions. These electrical charges play an important role in determining the interaction between particles of the dispersed phase and the resultant physical stability of the systems, particularly for those in the colloidal size range. For dispersed systems which are used as drug carriers, i.e., liposomes and microparticles, their *in vivo* fate and therapeutic efficacy are both affected by these surface charges. Therefore, the understanding of this electrical phenomenon becomes essential in developing these systems.

The presence of these surface charges influences an uneven distribution of charges (ions) surrounding the particle and the development of an electrical potential (or an electrical field) between the surface and the electrically neutral bulk-solution phase of the system. The surface charge and the counterions in its vicinity give rise to an electric double layer as shown in Fig. 1. The double layer is divided into two parts separated by a plane called the Stern layer which is located at about a hydrated-ion radius from the surface. Usually, the number of ions in the Stern layer is smaller than that needed to achieve neutralization of the surface charge and the balance of the neutralization occurs in a Gouy–Chapman layer outside the Stern layer. The surface potential (ψ_0) and the Stern potential (ψ_d) are not readily measured experimentally; instead, the potential between a stationary fluid layer enveloping the particle and the bulk-solution phase can be determined by measuring the mobility of the particle in an applied electrical field.

The potential between the tightly bound surface liquid layer (shear plane) of the particle and the bulk phase of the solution is called zeta potential (ζ). It can provide a measure of the net surface charge on the particle and potential distribution at the interface. Zeta potential serves as an important parameter in characterizing the electrostatic interaction between particles in dispersed systems and the properties of the dispersion as affected by this electrical phenomenon.

METHOD AND INSTRUMENTATION

Microelectrophoresis

The most common method for determining the zeta-potential is the microelectrophoretic procedure in which the movements of individual particles under the influence of a known electric field are followed microscopically. The zeta potential can be calculated from the electrophoretic velocity of the particles using the Helmholtz–Smoluchowski equation,

$$U_{mp} = V_{mp}E = \zeta \epsilon \eta \quad (1)$$

where U_p is the electrophoretic mobility, V_p is the electrophoretic velocity, E is the electric-field strength, ζ is the zeta potential, ϵ is the permittivity, and η is the viscosity of the medium.

It is important to realize that the observed electrophoretic migration is the sum of two contributions, one of which is the electro-osmotic flow of the medium through the cell. The glass of the sample cell generally bears negative charges, so that a diffuse layer of cations exists adjacent to them. This sets up an electro-osmotic flow through the cell (with or without particles). This flow has its maximum value at the center, since the layer of fluid adjacent to the walls is stationary. The particles tracked at the center of the cell therefore possess the maximum increment in velocity because of the electro-osmotic flow. Since there is no net liquid flow through the closed cell,

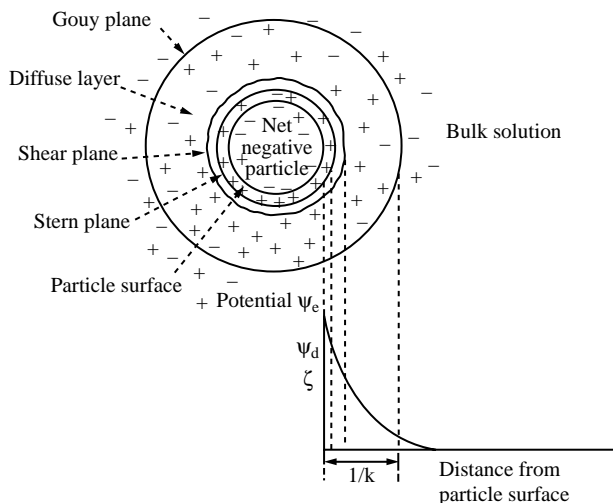


Fig. 1 A schematic representation of the electric double layer around a particle with negative surface charges and electrical potentials surrounding the particle.

a back-pressure builds up which causes fluid flow in the reverse direction in the core of the cell. In the steady state, these flows balance, giving rise to a velocity profile as shown in Fig. 2. Therefore, the tracking of particles at a location where the medium experiences no net flow becomes important in yielding an accurate electrophoretic mobility measurement. A velocity profile has been derived for the liquid contained in a cylindrical electrophoretic cell. The location with zero liquid flow is shown to be 14.6% of the cell diameter inside the surface of the capillary. The location of the surface of zero liquid flow in cells of rectangular cross section has also been determined. For a cell in which the direction of migration is relatively long compared to the width of the cell, the surface with zero liquid flow lies 21.1% of the cell depth above the bottom and below the top of the working compartment. Experimentally, particles tracked at these positions of the sample cell will display their mobility uncomplicated by the electro-osmotic effect.

The rate of particle migration can be determined by measuring with a stopwatch the time required for a particle

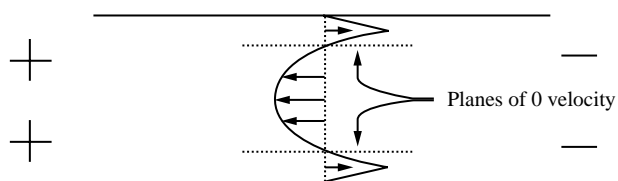


Fig. 2 The velocity profile of fluid flow in an electrophoretic cell.

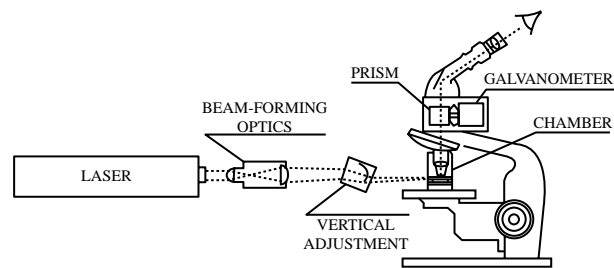


Fig. 3 A diagram of light paths in the rotating prism method of microelectrophoretic measurement.

to travel between the marks of a calibrated graticule in the microscope eyepiece. Since many particles are usually measured in order to get good statistical results, the measurement of a sample may take 30 min or even longer. As the time of measurement increases, however, the susceptibility to errors also increases. For example, a long measurement time will cause convection currents generated by heating of the sample, air-bubble formation, settling of particles, and electrode reactions. These potential errors can be avoided if the measurement is taken much more quickly without loss of accuracy. A rotating prism method employed by the Lazer Zee Meter (Model 50,1 Pen Kem, Inc.), allows the operator to measure more than 10 particles simultaneously. Data are thereby attained more rapidly and, consequently, with fewer time-dependent errors. Fig. 3 shows a diagram of light paths in the rotating prism method (2). The microscope is focused on the stationary layer of the cell (with no fluid flow). Special cylindrical optics compress the laser beam into an illumination sheet of laser light. A vertical adjustment is provided to vary the height so that it coincides with the focal plane of the microscope. A cube prism inside the microscope causes the viewed image to translate at a rate proportional to the prism's speed of rotation. The applied electric field causes the particle to move at an electrophoretic velocity proportional to the mobility and the applied voltage. Adjustment of the potentiometer causes the prism to rotate at a rate proportional to potentiometer voltage times applied voltage. When the image velocity caused by prism rotation is equal and opposite to the mean electrophoretic velocity, the cloud of particles appears stationary. The applied voltage is therefore proportional to the mean particle electrophoretic mobility. The mobility value is scaled by an appropriate constant in the digital meter and the computed zeta potential is displayed on the 3-digit readout.

In spite of a major improvement gained from the rotating prism method, there are still serious limitations in the use of microelectrophoresis in determining zeta potential. First, this technique cannot be applied to dense dispersions, i.e., the number density of the particles must

be very low. This can sometimes, but not always, be remedied by filtering the sample and diluting a small amount of the original dispersion with a large volume of filtrate. Second, it cannot be applied to particles less than about 0.2 μm in diameter because they are not easily visible in the ultramicroscope, though the particle-size range may be extended by adsorbing the colloidal particles (i.e., protein molecules) on suitable carrier particles. An additional and unresolvable problem with very small particles is the blurring effect imparted by Brownian motion. Furthermore, the microelectrophoretic technique cannot be used to determine mobility distributions or to separate multimodal mobility distributions.

Electrophoretic Light-Scattering (3–5)

Electrophoretic light-scattering is a technique which determines the electrophoretic velocities by measuring the Doppler shifts of scattered laser light. The electrophoretic mobility can be expressed by

$$U_p = \Delta\nu\lambda_0/2nE\sin(\theta/2) \quad (2)$$

where U_p is electrophoretic mobility, $\Delta\nu$ is the Doppler frequency shift, λ_0 is the wavelength of the laser in a vacuum, n is the index of refraction of the medium, E is the electric field applied, and θ is the scattering angle. Since a He–Ne laser light has a frequency of about 6.0×10^{14} , a direct measurement of frequency shifts ranging from zero to several hundred Hz may be problematic because of the difficulty in measuring a small difference between large quantities. However, this can be overcome by using the principle of heterodyning, which measures the difference in frequency resulting from a reference beam and the Doppler-shifted beam.

When the Brownian motion of the particles is negligible, the power spectrum which is generated by the autocorrelation function of the scattered light has a very narrow peak. However, the actual spectral peak has considerable line width attributable to the Brownian motion of the particles and a distribution of electrophoretic mobilities. The separation of these two sources of spectral-peak broadening is critical because, without such a separation, a measurement of electrophoretic mobility at a single angle provides information only about the peak position (Doppler shift). The shape of the peak (Doppler broadening), being attributable to either of the two effects, cannot be interpreted with a single-angle measurement. Simultaneous multiple-angle measurements allow peak shapes to be analyzed to give accurate measurement of the electrophoretic mobility and the subsequent determination

of zeta potential according to the Helmholtz–Smoluchowski equation (Eq. 1).

The DELSA 440 (Coulter Scientific Instruments) is a laser-scattering particle electrophoretic analyzer that measures the electrophoretic mobility distribution and zeta-potential distribution (6). This instrument provides simultaneous measurements of line widths at four different angles so that the motion of the particle can be decomposed with high resolution into components caused by electrophoretic and diffusional motions. Therefore, particle sizing can also be carried out by the same instrument simultaneously with electrophoretic mobility determinations. Fig. 4 presents a schematic diagram of the DELSA 440 system. Other commercially available electrophoretic analyzers using the laser light-scattering principles are the Zetasizer 3 (Malvern Instruments), ZetaPlus (Brookhaven Instruments), System 3000 (PenKem, Inc.), and NICOMP 380/ZLS (Particle Sizing Systems).

Electrokinetic Sonic Analysis (3, 6, 7)

When a colloidal dispersion is subject to ultrasonic waves, a density difference between the dispersed phase and the continuous phase will cause relative motion between the particles and the surrounding fluid. This relative motion means that there will be a periodic interchange of the charged particle surface and the oppositely charged counter-ions in the electric double layer. This interchange results in the development of an alternating dipole at the frequency of the applied ultrasound wave. This effect is termed the ultrasound vibration potential (UVP). The UVP effect in colloids is also referred to as the colloid vibration potential or CVP. An effect opposite that of CVP is generated, when an alternating electric field is applied to a colloidal dispersion, creating an acoustic wave which, by means of a piezoelectric device, may be transduced into an alternating potential of a certain amplitude, depending on the dispersion and the electric double-layer properties. This is called the electrokinetic sonic amplitude (ESA). These two phenomena are both known as electro-acoustic effects.

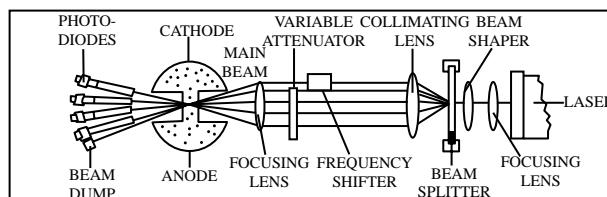


Fig. 4 A schematic diagram of the DELSA 440 system.

The magnitude of these effects is directly proportional to the electrophoretic mobility of the particles. The mobility determined by either electro-acoustic effect is the dynamic or AC mobility of the particle.

The dynamic mobility which is determined by the electro-acoustic effects differs from the low-frequency or DC mobility that is determined by electrophoresis because of the effects of particle inertia which cause particle velocities in an alternate field to become out of phase with the applied field. The higher the frequency, the greater is the phase lag between the particle and the applied field. Dynamic mobility (U_d) is a function of frequency, particle size, particle density, and particle zeta potential. These inertial effects arise in both ESA and CVP measurements. In the case of spherical particles with thin double layers and low zeta potential, the equation below can be applied:

$$U_d = (\epsilon\zeta/\eta)G(\alpha) \quad (3)$$

where $G(\alpha)$ is the inertial term. The formula for dynamic mobility is identical to the well-known Helmholtz–Smoluchowski equation (Eq. 1) for DC electrophoretic mobility except for the $G(\alpha)$ term.

A linear relationship exists between the ESA or CVP amplitude and the volume fraction of the suspended particles. At relatively high-volume fractions, hydrodynamic and electric double-layer interactions lead to a non-linear dependence of these two effects on volume fraction. Generally, non linear behavior can be expected when the electric double-layer thickness is comparable to the interparticle spacing. In most aqueous systems, where the electric double layer is thin relative to the particle radius, the electro-acoustic signal will remain linear with respect to volume fraction up to 10% by volume. At volume-fractions that are even higher, particle-particle interactions lead to a reduction in the dynamic mobility.

The commercial systems for zeta-potential determination using these electro-acoustic effects are typified by the AcoustoSizer 8000 (Matek Applied Sciences) consisting of five main components (Fig. 5). The synthesizer

produces a continuous sinusoidal voltage which feeds into the gated amplifier. This creates a sinusoidal voltage pulse across the cell which contains the dispersion. The resulting ESA sound waves are converted by a pressure transducer in the cell to an electrical signal, which then passes to the signal processing electronics. The electronics extract the amplitude and the phase of the ESA signal, storing them in the computer housed in the AcoustoSizer.

A schematic diagram of the cell is shown in Fig. 6. The dispersion is located in the region between the pair of electrodes. The pulse from the gated amplifier is applied across these electrodes, generating the ESA sound waves. Voltage pulses rather than continuous sinusoids are applied to avoid interference by the sound wave emanating from the electrodes. The use of pulsed signals also avoids electrical heating of the suspension and the complications of multiple reflections of the waves at the electrodes and at the ends of the rods. The voltage pulse produced by the sound wave in the transducer on the right passes into the signal-processing electronics, which measures the amplitude and phase of the sinusoidal component of the pulse. These data are passed to the computer where it is stored for subsequent processing.

Since the ESA signal depends on particle motion and also on the acoustic property of the dispersion, the determination of particle motion from the ESA signal requires knowledge of the acoustic impedance of the dispersion. This parameter is measured using sensors associated with the rod on the left immediately after each ESA measurement. The signals from these two rods are stored in the computer and are subsequently converted to a particle-velocity spectrum, and from that the zeta potential and size information is determined. The AcoustoPhor 8000 (PenKem, Inc.) is another example of a commercial system which determines zeta potential and the particle size of dispersions using these electro-acoustic effects.

APPLICATIONS

The practical significance of zeta-potential measurement lies in the fact that strong empirical correlation exists between the measured zeta potential of the system and the properties of the system which are the manifestation of the electrostatic interfacial phenomenon. Since the measurement of zeta potential can be conveniently performed, it becomes an ideal parameter for use in routine testing. Zeta-potential control has been successfully applied to various technical fields involving colloidal and non-colloidal systems.

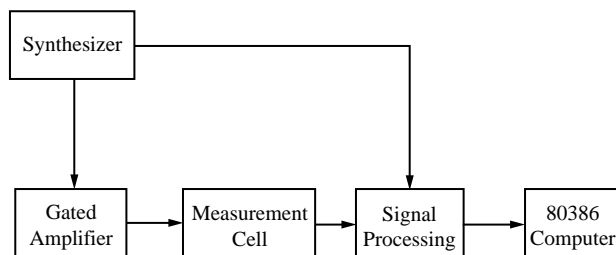


Fig. 5 An AcoustoSizer signal processing block diagram.

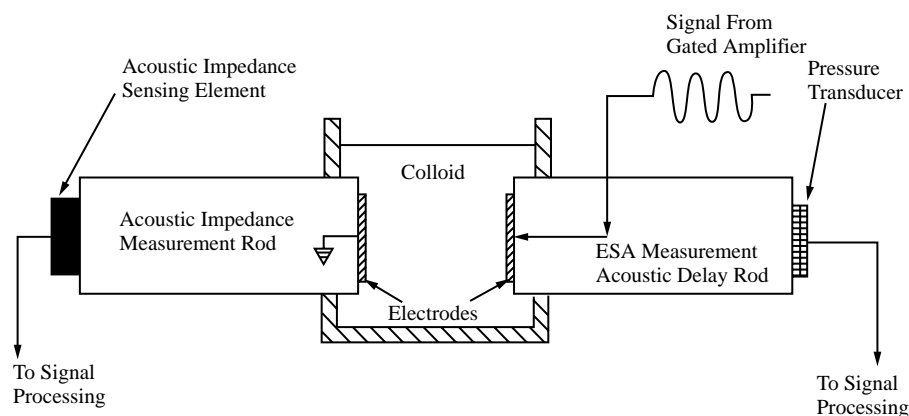


Fig. 6 A schematic diagram of AcoustoSizer measurement cell.

Colloid Stability (3, 8, 9)

The physical stability of a colloidal system is determined by the balance between the repulsive and attractive forces which is described quantitatively by the Deryaguin–Landau–Verwey–Overbeek (DLVO) theory. The electrostatic repulsive force is dependent on the degree of double-layer overlap and the attractive force is provided by the van der Waals interaction; the magnitude of both are a function of the separation between the particles. It has long been realized that the zeta potential is a good indicator of the magnitude of the repulsive interaction between colloidal particles. Measurement of zeta potential has therefore been commonly used to assess the stability of colloidal systems.

Electrostatic repulsive energy, V_R , at a given inter-particle distance is the work which must be performed to bring the particles to a specific point. Using the Debye–Huckel low-potential approximation and assuming equal spheres, V_R can be described by

$$V_R = 2\pi\epsilon a\psi_d^2 \ln(1 + \exp[-\kappa H]) \quad (4)$$

where a is the particle radius, H is the distance of separation between the two particles, and κ is the Debye parameter. The magnitude of V_R decreases in an approximately exponential fashion with increasing H and that its range decreases by increasing κ (i.e., by increasing electrolyte concentration and/or counter-ion charge number). V_R can be also influenced by other specific factors. Counter-ion adsorption in the Stern layer may cause a reversal of charge, so that V_R will be zero at the reversal-of-charge concentration and will be positive (repulsion) at both below and above this concentration.

The attractive-force component between particles in a colloidal system is developed by summation of the London

dispersion forces between all atom pairs in the particles. Neglecting retardation effects, the expression for the attractive energy V_A between two particles of radius a at a distance of separation H_0 , for $a \gg H_0$, is

$$V_{rma} = A, a12H_0 \quad (5)$$

where A is the effective Hamaker constant describing the attraction between the particles and the dispersion medium. It is apparent that V_A decreases as an inverse power of the distance between the particles.

According to DLVO theory, the total energy of interaction between colloidal particles is given by the sum of the attraction (V_A) and repulsion (V_R) energies:

$$V_T = V_R + V_A \quad (6)$$

Fig. 7a shows the total interaction energy curve for a colloidal system. At very close distances between the particles, the Born repulsion between adjoining electron clouds predominates, and the net interaction is repulsion. However, at slightly greater distances, the van der Waals attraction predominates over the electrostatic repulsion, and the net interaction is strong attraction as shown by the deep potential energy minimum V_p . At even greater distances, the net interaction is moderate attraction (a shallow secondary minimum V_s) because with increasing distance between two particles, electrostatic repulsion weakens more rapidly than the van der Waals attraction energy. At intermediate distances, the electrostatic repulsion predominates and the net interaction becomes repulsion, with an energy barrier, V_m .

The attraction energy arising from van der Waals forces does not vary significantly with changes in the surface potential and electric double-layer properties of the colloidal particles. By contrast, changes in these two parameters have significant impact on the repulsion energy

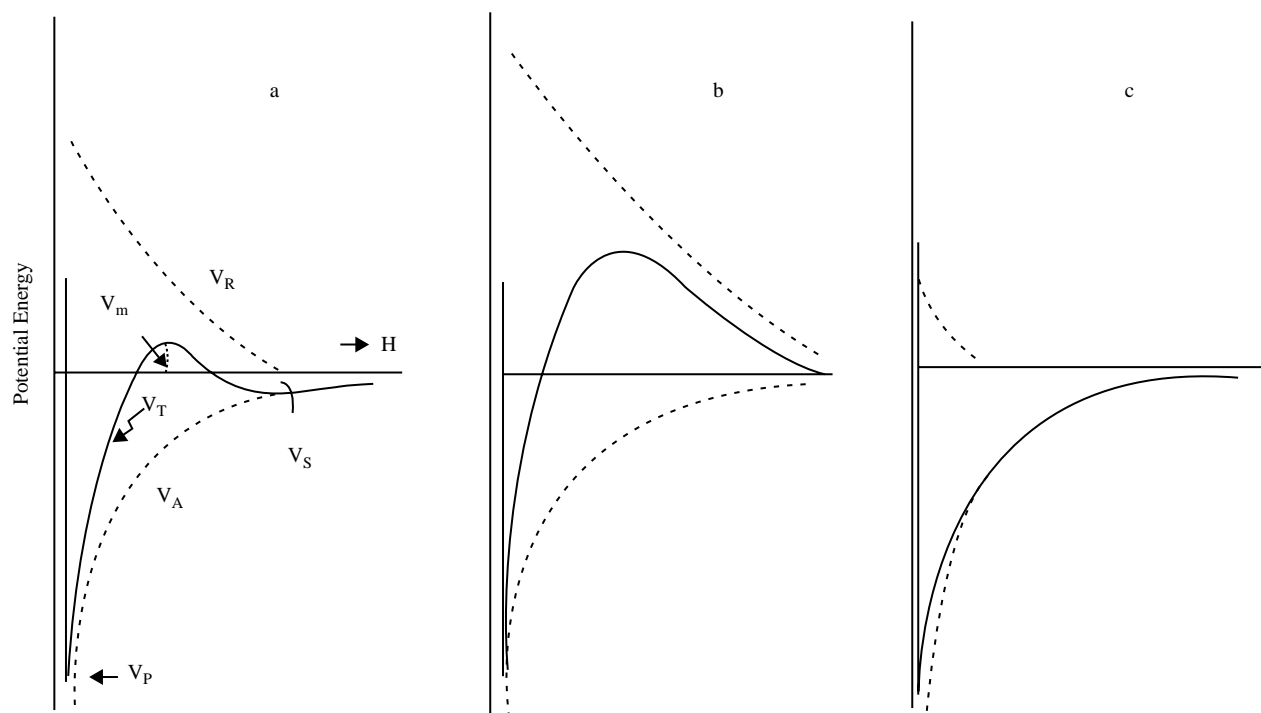


Fig. 7 Variation of the attraction and repulsion energies and the total energy of interaction between two colloidal particles with interparticle distance.

component of the total interaction energy. Therefore, at greater distances, a further increase in electrostatic repulsion will result in a total repulsive net interaction curve without the secondary minimum (Fig. 7b). However, as the repulsive component of the interaction energy diminishes (i.e., the electrolyte concentration increases), the interaction curve becomes totally attractive and rapid coagulation (controlled by particle-diffusion rate) of the colloidal particles takes place (Fig. 7c).

An empirical relationship was developed between the zeta potential and the coagulation behavior of a variety of systems. Deryaguin shows that a rapid decrease in stability can be expected when the following conditions are met (3).

$$4\pi\epsilon\xi^2/\kappa[A] \quad (7)$$

For a given system, A and ϵ are fixed; therefore, varying the electrolyte (in type and concentration) should produce instability at a particular value of ξ^2/κ . The validity of using zeta potential in describing the coagulation process should come as no surprise because it characterizes the potential of the diffuse part of the electric double layer which is involved in double-layer overlap and which plays an important role in the coagulation process. When the zeta potential is sufficiently high to produce a potential

barrier opposing coagulation, the rate of coagulation is slowed by a factor W , called the stability ratio, defined as the ratio of the total number of particle collisions to the number of collisions forming permanent doublets. Wiese and Healy reported an excellent correlation between the theoretical values for W (calculated from the zeta potential) and the experimental values for TiO_2 and Al_2O_3 sols as a function of pH (Fig. 8) and electrolyte (KNO_3) concentration (10). Rapid coagulation, whether induced by pH or KNO_3 concentration changes, occurs at a value of $|\zeta| \leq 14 \pm 4$ mV, corresponding to $\xi^2/\kappa = 6 \times 10^{-4} (\text{mV})^2 \text{ cm}$.

Pharmaceutical Systems

Most of the pharmaceutical dispersions can be classified as colloidal systems, although this depends upon the size of their dispersed phase. Therefore, in general, the theoretical and empirical relationships between zeta potential and colloid stability are applicable to the physical stability of these pharmaceutical systems. However, the practical aspects of applying zeta potential to these systems may vary because of differences in the chemical composition and physical state of their dispersed and continuous phases. In addition, the practical significance of zeta-potential

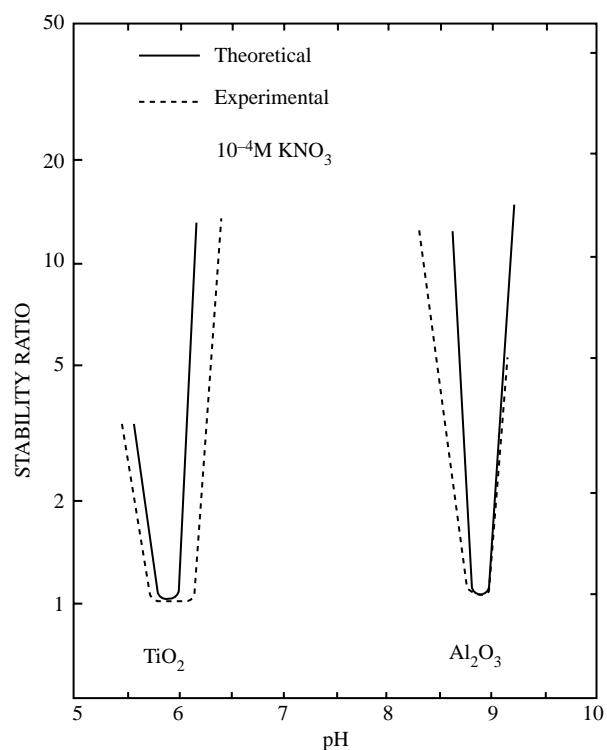


Fig. 8 Comparison of theoretical and experimental stability ratios for TiO_2 (0.05 g/L) and Al_2O_3 (0.15 g/L) colloids in 10^{-4}M KNO_3 .

measurement may be different for products which differ in their clinical requirements such as route of administration and consequent in vivo disposition.

Emulsions

Emulsions have been widely used as vehicles for oral, topical, and parenteral delivery of medications. Although the product attributes of an emulsion dosage form are dependent on the route of administration, a common concern is the physical stability of the system, in particular the coalescence of its dispersed phase and the consequent alteration in its particle-size distribution and phase separation. The stabilization mechanism(s) for an emulsion is mainly dependent on the chemical composition of the surfactant used. Electrostatic stabilization as described by DLVO theory plays an important role in emulsions (O/W) containing ionic surfactants (11). For O/W emulsions with low electrolyte content in the aqueous phase, a zeta potential of 30 mV is found to be sufficient to establish an energy maximum (energy barrier) to ensure emulsion stability (11). For emulsions containing nonionic surfactants (polymers), the principal stabilization mechanism is the repulsion between the adsorbed polymer chains on

the surface of the oil globules called steric stabilization (11). In such a case, zeta-potential measurement may have limited values with respect to emulsion stability.

In the past, injectable emulsions were administered intravenously to provide patients with vegetable oils as a major energy source in nutrition therapy (12). In recent years, injectable emulsions have been found to be very effective vehicles for the parenteral delivery of water-insoluble drugs (13). Because of toxicological considerations, surfactants which are used in the formulation of injectable emulsions have been largely limited to mixtures of phospholipids which are either derived from egg (egg yolk lecithin) or soybean (soybean lecithin). The majority of the phospholipids (80–90%) are phosphatidylcholine (PC) and phosphatidylethanolamine (PE), which are uncharged at physiological pH. But the presence of small quantities (2–5%) of acidic components, largely phosphatidylserin (PS) and phosphatidylglycerol (PG), which are ionized at pH 7, confers a surface charge of approximately -40 to -50 mV on the emulsion particles (14). Therefore, the stability of these emulsions is mainly attributable to electrostatic stabilization. Any factors which alter the surface potential are likely to have an impact on overall emulsion stability. The physical instability of an injectable emulsion as manifested by an increase in particle size ($>5\text{ }\mu\text{m}$) will result in serious clinical consequences such as formation of pulmonary emboli (15). Therefore, the extent of electrostatic stabilization of an injectable emulsion as measured by its zeta potential has been an important parameter in the characterizing and monitoring of its physical stability.

A relative wealth of information relating to the application of zeta potential to injectable emulsions has been documented with respect to the use of total nutrient admixtures (TNA) (16). Total nutrient admixtures are prepared by mixing the lipid emulsion with other components (i.e., dextrose, amino acids, and electrolytes) in a single container prior to administration. Depending on composition, the mixtures vary widely in their stability and may show clinically unacceptable coalescence after different periods of storage time.

When the electrostatic stabilization of the emulsion is considered, the electrolytes (monovalent and divalent) added to the mixture are the major destabilizing species. The zeta potential of the emulsion particles is a function of the concentration and type of electrolytes present. Two types of emulsion particle-electrolyte (ions) interaction are proposed: nonspecific and specific adsorption (16). In nonspecific adsorption the ions are bound to the emulsion particle only by electrical double-layer interactions with the charged surface. As the electrolyte concentration is increased, the zeta potential asymptotes to zero. As

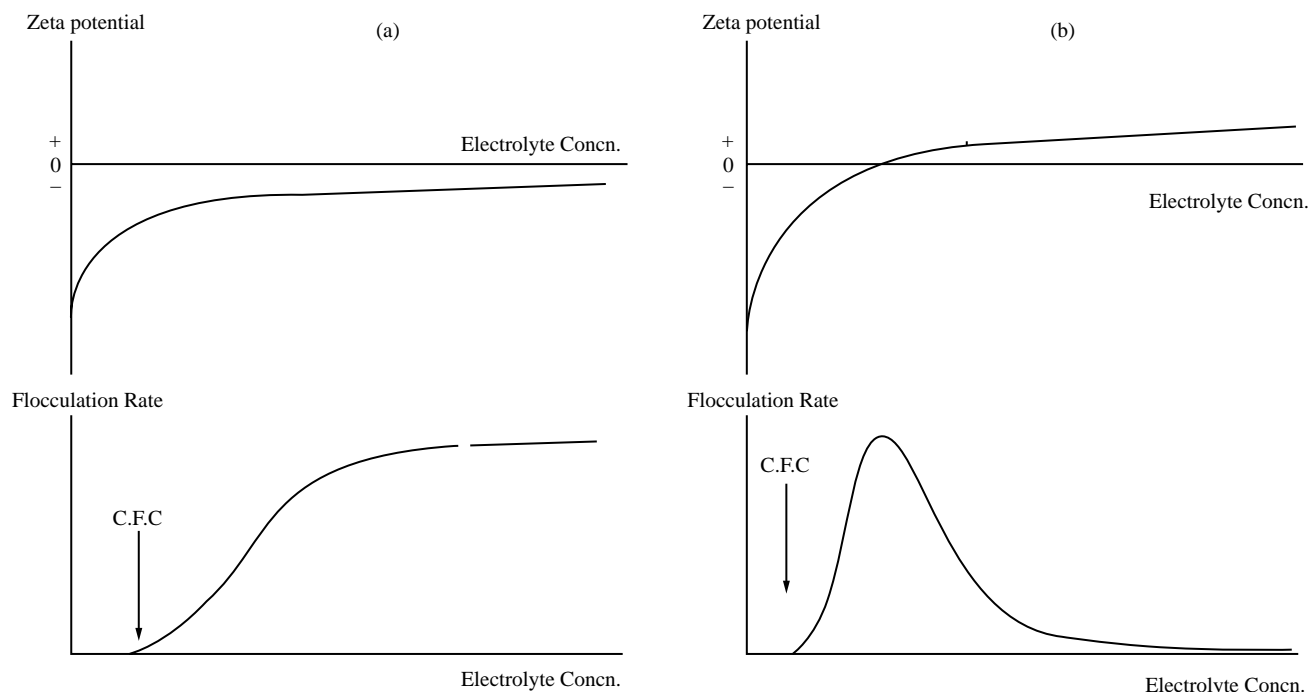


Fig. 9 Zeta potential and flocculation rate of a parenteral emulsion in the presence of a nonspecifically adsorbing electrolyte (a) and a specifically adsorbing electrolyte (b).

the electrostatic repulsion decreases, a point can be found where the attractive van der Waals force is equal to the repulsive electrostatic force and flocculation of the emulsion occurs (Fig. 9a). This point is called the critical flocculation concentration (CFC).

DLVO theory predicts that the CFC should vary inversely with the sixth power of the ion charge. The application of CFC alone fails to correlate the TNA stability with electrolytes in the mixtures. This is because divalent ions (calcium and magnesium) adsorbed specifically on the emulsion particles through their complexation with the negatively charged acidic phospholipids. Fig. 9b demonstrates the relationship between the concentration of a specifically adsorbed ion and the zeta potential and flocculation rate of an injectable emulsion. The increase in electrolyte concentration causes the particle charge to pass through zero; this is referred to as the point of zero charge (PZC). At this point the flocculation rate is at maximum; a further increase in electrolyte concentration causes a surface-charge reversal of the particles and increases in zeta potential (opposite sign) as well as a decrease of the flocculation rate from its maximum value.

The effect of pH on the stability of an injectable emulsion was also followed by measuring its zeta potential (Fig. 10). At pH 7 the ionization of PS and PG imparts a surface negative charge (-30 to -50 mV) to the emulsion

particles. As the pH is reduced, this ionization is suppressed until the charge is zero at a pH of 3.2. Further decreases in pH cause the protonation of the phospholipid and a positive surface charge (positive zeta potential). Significant progress has been made in relating the electrokinetic properties of phospholipid-stabilized emulsions to their instability when influenced by other TNA ingredients (17, 18). Washington and his coworkers have attempted to develop a model to predict emulsion stability in TNAs using DLVO-based methods in the measurement of the zeta potential of the system (19).

Suspensions

Pharmaceutical suspensions are dispersions of solid particles in a suspending medium or vehicle (usually aqueous in nature). When the suspended solids are less than $1\text{ }\mu\text{m}$, the system is referred to as a colloidal suspension. When the particle sizes are greater than about $1\text{ }\mu\text{m}$, the system is called a coarse suspension. The practical upper limit for particles in a coarse suspension is approximately $50\text{--}75\text{ }\mu\text{m}$. Depending on the affinity or interaction between the dispersed phase and the dispersion medium, a colloidal dispersion can be classified as lyophilic (hydrophilic) or lyophobic (hydrophobic) (20).

The dispersion of hydrophilic particulate solids in water occurs spontaneously and gives rise to a thermodynamically

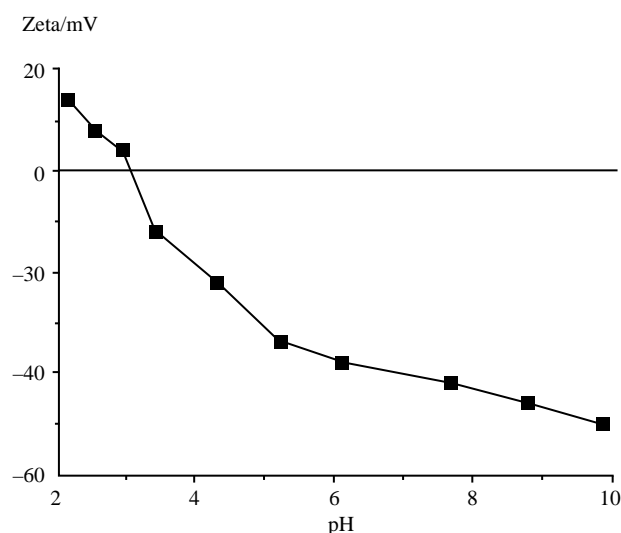


Fig. 10 Zeta potential of a parenteral emulsion as a function of pH.

stable system such as colloidal silicone dioxide or microcrystalline cellulose. Hydrophobic solids are not easily wetted and are not dispersed spontaneously in water as sulfur, clays, and most nonpolar organic compounds. The van der Waals attractive forces between particles cause them to aggregate, since the solvation forces which promote dispersal in water are weak. Therefore, aqueous dispersions of hydrophobic solids can only be stabilized kinetically to resist flocculation and coagulation, which can be defined by referring to the total interaction energy curve (Fig. 7). Flocculation indicates the association of particles in the secondary minimum and coagulation is the aggregation of particles in the primary minimum. Since the attractive forces between particles in a flocculated system are relatively weak, they can be redispersed and the process is reversible. In contrast, coagulation is irreversible as attractive forces operating between particles in the primary minimum are difficult to overcome.

When electrostatic repulsion is predominant between particles in a suspension, it is called a deflocculated system. For particles with a diameter of 2–5 μm , Brownian movement counteracts sedimentation to a measurable extent at room temperature by keeping the dispersed particles in random motion. This results in a system which is physically stable with respect to sedimentation and flocculation (20). However, in a deflocculated suspension consisting of larger particles (coarse), gravitational force counteracts Brownian movement and sedimentation occurs (21). At the bottom of the container, a compact sediment with strong attractive forces between the particles is formed and it cannot be easily

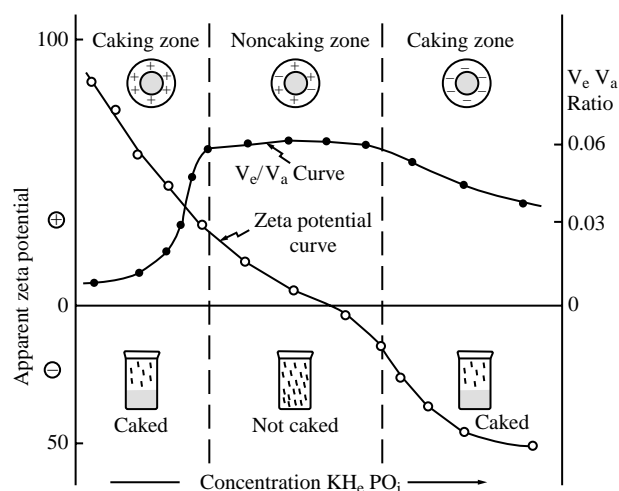


Fig. 11 A caking diagram showing the flocculation of a bismuth subnitrate suspension by means of the flocculating agent, potassium monobasic phosphate.

redispersed. This phenomenon is referred to as caking. The prevention of caking has been one of the main objectives in preparing stable coarse suspensions (21, 22). Although flocculation is referred to as a physical instability of colloids, flocculated suspensions are pharmaceutically acceptable since they are noncaking. In a flocculated suspension, particles attract each other (at the secondary minimum) loosely to form flocs which tend to settle together, creating a distinct boundary between the sediment and the supernatant. Particles in flocs are thereby easily redispersed after settling.

Controlled flocculation is a process by which flocculation of particles in a suspension is purposely produced by adding flocculating agents. One of the most effective methods of achieving controlled flocculation is to reduce the electrostatic repulsive forces between particles (21, 22). Electrolytes have been shown to be effective flocculating agents. Zeta-potential measurement has been a valuable tool in monitoring and evaluating the influence of various electrolytes on the flocculation phenomena in coarse suspensions (23). Sedimentation volume F , which is the ratio of the volume of sediment to the original volume of the suspension, is a useful parameter in measuring the extent of flocculation of the suspension. An excellent correlation between the zeta potential, sedimentation volume, and caking was found for a series of bismuth subnitrate suspensions containing various concentrations of potassium phosphate monobasic (Fig. 11) (24). The addition of potassium phosphate monobasic causes a decrease in the zeta potential of the suspension because of partial neutralization of the positive

surface charge on the bismuth subnitrate particles. With the decrease in the zeta potential (positive), the suspension becomes flocculated and noncaking, accompanied by an increase in the sedimentation volume. The continued addition of electrolyte causes the zeta potential to fall to zero and then increases in a negative direction. The suspension remains flocculated until the zeta potential becomes sufficiently negative to effect deflocculation of the system, which is shown by a decrease in the sedimentation volume and an increase in caking tendency.

Zeta potential has also found an application in relating the surface-charge characteristics of a suspension to properties other than physical stability. Heyd and Dhabhar improved the fluidity of concentrated antacid suspensions by adding a colloidal polyelectrolyte (carrageenan sodium) which is called a fluidizing agent (25). As the concentration of the fluidizing agent increases, the zeta potential of the suspension changes from positive to negative, accompanied by a reduction in viscosity of the suspension. The mechanism for this viscosity-thinning phenomenon is believed to be the selective adsorption of the negatively charged fluidizing agent onto the positively charged antacid particles. This imparts an electronegative charge (as measured by the zeta potential) onto the antacid particles, decreasing the demand for water.

In the literature, the reported use of zeta potential measurement for nonaqueous suspensions is relatively infrequent because nonaqueous suspensions only represent a small percentage of all medicated suspensions. Su and others evaluated the flocculation–deflocculation behavior of cefazolin sodium in nonaqueous media and the effect of surfactants as measured by zeta potential along with sedimentation and porosity measurements (26). A significant difference in zeta potential was observed when the particles were dispersed in peanut oil and ethyl oleate. The addition of lecithin reduced the zeta potential of cefazolin sodium, resulting in a deflocculated state accompanied by a decrease in sedimentation volume. The effect of surfactant excipient on the zeta potential of salbutamol in trichlorofluoroethane (P113) for metered-dose inhalation was determined (27). Although the zeta potential of the particles is decreased (more negative) by the addition of oleic acid, the overall zeta potential is too low to be effective in providing the system with adequate electrostatic stabilization, particularly in media with a low dielectric constant (organic solvents).

Microparticulate drug-delivery systems, including microspheres and nanoparticles, are suspensions when they are administered in a liquid medium. Microspheres are solid particles containing dispersed drug in either solution or microcrystalline form and range in size from

1 to 1000 μm . Although most microspheres are outside the conventional colloidal size range, in pharmaceutical literature microspheres with sizes up to approximately 15 μm are still considered as colloidal delivery systems (28). Nanoparticles are similar to microspheres but have particle sizes in the range of 0.01–1 μm ; therefore, they form colloidal suspensions when dispersed in a liquid medium (29). Zeta-potential measurement is an important surface-characterization technique which provides information regarding the surface charge of these microparticulate systems. The effect of surface charge (zeta potential) on the physical stability of these systems (i.e., aggregation) can be predicted using the DLVO theory, although steric stabilization plays an important role in systems with surface-adsorbed macromolecules.

When the microparticles are administered intravenously, surface charge is an important parameter affecting the interaction between the microparticles and other biological components of the body and their subsequent *in vivo* distribution and disposition (30). Tabata and Ikada reported that phagocytosis of polymer microspheres by macrophages is enhanced as the absolute value of zeta potential increases for both the negatively charged and the positively charged surfaces, while the lowest phagocytosis is shown for the surface with a zeta potential of zero (31). The uptake of charged serum components by nanoparticles (opsonization) in blood is considered to be the first step leading to phagocytosis by liver and spleen macrophages. Muller and his coworkers determined the increase in zeta potential of nonparticles in serum as a measure of the extent of opsonization and used this as selection criteria for *i.v.* drug carriers avoiding liver/spleen uptake (32).

Liposomes

Liposomes are phospholipid-based vesicles which have been studied as delivery systems to target drugs to specific sites in the body (33–35). The surface of a liposome can be neutral or negatively or positively charged, depending on the composition of lipids used. For charged liposomes, their surface potential is a critical factor affecting their physical stability and *in vivo* performance. Instead of their surface potential, zeta-potential measurement is frequently used in the investigation of the behavior of liposomes as influenced by their surface charge. Liposomes in aqueous dispersions have a tendency to aggregate and subsequently fuse on storage. This physical instability potentially limits their application as dosage forms.

Crommelin incorporated charge-inducing components in liposome bilayers to produce electrostatic repulsion between liposomes, thus increasing the shelf life of the dispersion (36). He investigated the effect of including

stearylamine or phosphatidylserin on the zeta potential of phosphatidylcholine-cholesterol-containing liposomes. A correlation between the aggregation stability of negatively charged liposomes and the increments in the surface-charge density at both low and high ionic strength was established, whereas the positive liposome dispersions were found to be unstable.

Carrión and his coworkers investigated the effect of incorporating phosphatidic acid on the zeta potential and aggregation of large unilamellar vesicle liposomes of phosphatidylcholine in the presence of neutral electrolytes (37). Their results show that increasing concentrations of phosphatidic acid in lipid bilayers resulted in high zeta potential and enhanced physical stability of the liposomes. The destabilizing effect of divalent cations (i.e., calcium ions) on charged liposomes has been well documented and is attributable to the reduced surface potential of the liposomes as the result of cation binding (38–40).

The effect of calcium ions on the aggregation behaviors of neutral liposomes composed of egg phosphatidylcholine and dipalmitoylphosphatidylcholine was investigated by Mosharraf and coworkers (41). Both systems displayed a high negative zeta potential in deionized water, probably the result of adsorption of hydroxyl ions. A reasonable correlation was found in both cases between particle-size increase and zeta potential decrease of the system, suggesting that the mechanism by which the aggregation takes place is related to the surface-charge characteristics. Although it has long been established that liposomes composed of neutral phospholipids acquire negative surface charge via anion adsorption, Makino and his group in Japan found that neutral liposomes exhibit negative zeta potential in buffer solution containing no chloride ion (42). They also reported that changes in the ionic strength and temperature causes the zeta potential of neutral liposomes to reverse sign. They proposed that the reversal of zeta potential is triggered by changes in the direction of the dipole connecting the negative charge of the phosphatidyl group and the positive charge of the choline group consisting the head group of a lipid molecule.

The study of the effect of drug-loading on the zeta potential of liposomes provides information regarding the drug-liposome interaction. Lawrence and others reported that the loading of polymyxin B, a polycationic antibiotic in negatively charged liposomes, results in the decrease of zeta potential, indicating electrostatic attraction between the positive drug and the negative lipid bilayer (43). That the highest drug-loading is found in negatively charged liposomes in comparison to positively charged and neutral liposomes further supports this electrostatic drug-liposome interaction. Beschiaschvili and Seelig investigated the

binding of a cyclic somatostatin analogue, a positively charged peptide, to negatively charged multilamellar liposomes (44). From the binding isotherm and the zeta-potential measurement, they were able to describe a partition equilibrium between the peptide and the negatively charged membrane with a surface-partition constant. They concluded that most of the peptide molecules are embedded in the headgroup region with little penetration into the lipid core.

The measurement of the zeta potential of liposomes provides valuable information relating to their *in vivo* performance because, in addition to size, the surface charge of liposomes is an important determinant of their clearance from the general circulation and their tissue disposition after parenteral administration. Large multilamellar liposomes are cleared from the circulation much more rapidly than small unilamellar liposomes. Among small unilamellar liposomes, however, those with a negative surface charge are cleared rapidly, but positively charged or uncharged liposomes remain in the circulation for longer periods (45). Rahman and his coworkers compared the pharmacological and therapeutic effects of adriamycin entrapped in positively charged and negatively charged liposomes in rats (46). They found that less cardiotoxicity was associated with the drug entrapped in the positively charged liposomes which exhibited lower *in vivo* uptake by cardiac tissue. When liposomes are employed to target drugs to lymph nodes, their surface charge has been shown to affect their lymphatic uptake from subcutaneous and intraperitoneal injection sites (47). Patel showed that negatively charged liposomes yield the highest localization in the lymph nodes, followed by positively charged and neutral ones. The drainage of negatively charged liposomes has also shown to be faster than that of positive liposomes after intraperitoneal administration (48).

SUMMARY

Zeta-potential measurement has long been recognized as an excellent tool for characterizing colloidal systems. Its practical significance stems from the fact that it allows the quantitation of the electrostatic repulsion between particles, which is one of the most important forces governing the behavior and physical stability of colloidal systems. In recent years, the concept of zeta potential has been applied to areas beyond classical colloidal sciences and industrial processes. Pharmaceutical sciences are the new fields to which the application of zeta potential has been extended. The expanding role of zeta potential in

these fields is attributable to the advance in modern instrumentation of zeta potential measurement, the rapid development of colloidal drug-delivery systems, and the emphasis on interdisciplinary basic research. While good correlation has been demonstrated for zeta potential and in vitro properties of colloidal drug-carrier systems, the understanding of the impact of zeta potential in organ distribution and in vivo clearance of these systems has been the focus of additional research efforts.

REFERENCES

- Shaw, D.J. *Introduction to Colloid and Surface Chemistry*, 4th Ed.; Butterworth Heinemann Ltd.: Oxford, 1992.
- Goetz, P.J.; Penniman, J.G. A New Technique for Microelectrophoretic Measurements. *Jr. Am. Lab. Oct. 1976*.
- Hunter, R.J. *Zeta Potential in Colloid Science, Principles and Applications*, Academic Press: London, 1981.
- Uzgiris, E.E. *Progress in Surface Science*, Pergamon Press: New York, 1981; X.
- Oja, T.; Bott, S.; Sugrue, S. Doppler Electrophoretic Light Scattering Analysis Using the Coulter DELSA 440. Coulter Electronics: Luton, 1989.
- O'Brien, R.W. *J. Fluid Mech.* **1988**, *190*, 71–86.
- Babchin, A.J.; Chow, R.S.; Sawatzky, R.P. Electrokinetic Measurements by Electroacoustical Methods. *Adv. Colloid Interface Sci.* **1989**, *111–151*, 30.
- Hiemenz, P.C. *Principles of Colloid and Surface Chemistry*, 2nd Ed.; Marcel Dekker, Inc.: New York, 1986.
- Hirtzel, C.S.; Rajagopalan, R. *Colloidal Phenomena, Advanced Topics*, Noyes Publications: New Jersey, 1985.
- Wiese, G.R.; Healy, T.W. Coagulation and Electrokinetic Behavior of TiO_2 And Al_2O_3 Colloidal Dispersions. *J. Colloid Interface Sci.* **1975**, *51*, 427–433.
- Friberg, S.E.; Goldsmith, L.B.; Hilton, M.L. *Theory of Emulsion in Pharmaceutical Dosage Forms: Dispersion Systems*, Lieberman, H.A., Reiger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1988; I.
- Wretling, A.J. Development of Fat Emulsions. *Parenter. Enteral. Nutr.* **1981**, *5*, 230–235.
- Davis, S.S.; Washington, C.; West, P.; Illum, L.; Liversidge, G.; Sternson, L.; Kirsh, R. *Ann. N.Y. Acad. Sci.* **1987**, *507*, 75–88.
- Washington, C.; Chawla, A.; Christy, N.; Davis, S.S. The Electrokinetic Properties of Phospholipid-Stabilized Fat Emulsions. *Int. J. Pharm.* **1989**, *54*, 191–197.
- Burnham, W.R.; Hansrani, P.K.; Knott, C.E.; Cook, J.A.; Davis, S.S. Stability of a Fat Emulsion Based Intravenous Feeding Mixture. *Int. J. Pharm.* **1983**, *13*, 9–22.
- Washington, C. The Stability of Intravenous Fat Emulsions in Total Parenteral Nutrition Mixtures. *Int. J. Pharm.* **1990**, *66*, 1–21.
- Washington, C.; Athersuch, A.; Kynoch, D.J. The Electrokinetic Properties of Phospholipid Stabilized Fat Emulsions. IV. The Effect of Glucose and of pH. *Int. J. Pharm.* **1990**, *64*, 217–222.
- Washington, C.; Connolly, M.A.; Manning, R.; Skerratt, M.C.L. The Electrokinetic Properties of Phospholipid Stabilized Fat Emulsions. V. The Effect of Amino Acids on Emulsion Stability. *Int. J. Pharm.* **1991**, *77*, 57–63.
- Washington, C.; Ferguson, J.A.; Irwin, S.E. Computational Prediction of the Stability of Lipid Emulsions in Total Nutrient Admixtures. *J. Pharm. Sci.* **1993**, *82*, 808–812.
- Zograff, G.; Schott, H.; Swarbrick, J. *Dispersion Systems. Remington's Pharmaceutical Sciences*, 18th Ed.; Mack Publishing Company: Pennsylvania, 1990.
- Hiestand, E.N. Theory of Coarse Suspension Formulation. *J. Pharm. Sci.* **1964**, *53*, 1–18.
- Falkiewicz, M.J. Theory of Suspensions in Theory of Emulsion. *Pharmaceutical Dosage Forms: Dispersion System*, Lieberman, H.A., Reiger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1988; I.
- Nash, R.A.; Haeger, B.E. Zeta Potential in the Development of Pharmaceutical Suspension. *J. Pharm. Sci.* **1966**, *55*, 829–837.
- Martin, A.; Swarbrick, J. *American Pharmacy*, 6th Ed.; Lippincott: Philadelphia, 1966.
- Heyd, A.; Dhabhar, D.J. Enhancing Fluidity of Concentrated Antacid Suspensions. *J. Pharm. Sci.* **1975**, *64*, 1697–1699.
- Su, K.S.E.; Quay, F.; Campanale, K.M.; Stucky, J.F. Nonaqueous Cephalosporin Suspension for Parenteral Administration: Cefazolin Sodium. *J. Pharm. Sci.* **1984**, *73*, 1601–1602.
- Clarke, J.G.; Wicks, S.R.; Farr, S.J. Surfactant Mediated Effects in Pressurized Metered Dose Inhalers Formulated as Suspensions. I. Drug/Surfactant Interactions in a Model Propellant Systems. *Int. J. Pharm.* **1993**, *93*, 221–231.
- Burgess, D.J.; Hickey, A.J. Microsphere Technology and Applications. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc. New York, 1994; 10.
- Kreuter, J. Nanoparticles. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1994; 10.
- Wilkins, D.J.; Meyers, P.A. Studies on the Relationship Between the Electrophoretic Properties of Colloids and their Blood Clearance and Organ Distribution in the Rat. *Br. J. Exp. Pathol.* **1966**, *47*, 568–576.
- Tabata, Y.; Ikada, Y. Effect of the Size and Surface Charge of Polymer Microspheres on their Phagocytosis by Macrophage. *Biomaterials* **1988**, *9*, 356–362.
- Müller, R.H.; Wallis, K.H.; Tröster, S.D.; Kreuter, J. In Vitro Characterization of Poly(Methyl-Methacrylate) Nanoparticles and Correlation to their In Vivo Fate. *J. Contr. Rel.* **1992**, *20*, 237–246.
- Riaz, M.; Weiner, N.; Martin, F. Liposomes in Theory of Emulsion. *Pharmaceutical Dosage Forms: Dispersion Systems*, Lieberman, H.A., Reiger, M.M., Banker, G.S., Eds.; Marcel Dekker, Inc.: New York, 1989; 2.
- Gregoriadis, G.; Florence, A.T. Liposomes in Drug Delivery. *Drugs* **1993**, *45*, 15–28.
- Barenholz, Y.; Crommelin, D.J.A. Liposomes as Pharmaceutical Dosage Forms. *Encyclopedia of Pharmaceutical Technology*, 1st Ed.; Swarbrick, J., Boylan, J.C., Eds.; Marcel Dekker, Inc.: New York, 1994; 9.

36. Crommelin, A.J.A. Influence of Lipid Composition and Ionic Strength on the Physical Stability of Liposomes. *J. Pharm. Sci.* **1984**, *73*, 1559–1563.
37. Carrion, F.J.; De La Maza, A.; Parra, J.L. The Influence of Ionic Strength and Lipid Bilayer Charge on the Stability of Liposomes. *J. Colloid Interface Sci.* **1994**, *164*, 78–87.
38. Düzgünes, N.; Sur, S.; Wilschut, J.; Bentz, J.; Newton, C.; Portis, A.; Papahadjopoulos, D. Calcium-and Magnesium-Induced Fusion of Mixed Phosphatidylserine/Phosphatidylcholine Vesicles: Effect of Ion Binding. *J. Membr. Biol.* **1981**, *59*, 115–125.
39. Hope, M.J.; Walker, D.C.; Cullis, P.R. Ca^{2+} and pH Induced Fusion of Small Unilamellar Vesicles Consisting of Phosphatidylethanolamine and Negatively Charged Phospholipids: A Freeze Fracture Study. *Biochim. Biophys. Res. Commun.* **1983**, *110*, 15–22.
40. Minami, H.; Inoue, T.; Shimozawa, R. Aggregation Kinetics of Dimyristoylphosphatidylglycerol Vesicles Induced by Divalent Cations. *J. Colloid Interface Sci.* **1993**, *158*, 460–465.
41. Mosharraf, M.; Taylor, K.M.G.; Craig, D.Q.M. Effect of Calcium Ions on the Surface Charge and Aggregation of Phosphatidylcholine Liposomes. *J. Drug Target.* **1995**, *2*, 541–545.
42. Makino, K.; Yamada, T.; Kimura, M.; Oka, T.; Ohshima, H. Temperature and Ionic Strength-Induced Conformational Changes in the Lipid Head Group Region of Liposomes as Suggested by Zeta-Potential Data. *Biophys. Chem.* **1991**, *41*, 175–183.
43. Lawrence, S.M.; Alpar, H.O.; McAllister, S.M.; Brown, M.R.W. Liposomal (MLV) Polymyxin B: Physicochemical Characterization and Effect of Surface Charge on Drug Association. *J. Drug Target.* **1993**, *1*, 303–310.
44. Beschiaschvilli, G.; Seelig, J. Peptide Binding to Lipid Bilayers. Binding Isotherms and ζ -Potential of a Cyclic Somatostatin Analogue. *Biochemistry* **1990**, *29*, 10995–11000.
45. Juliano, R.L.; Stamp, D.; McCullough, N. Pharmacokinetics of Liposome-Encapsulated Antitumor Drugs and Implications for Therapy. *Ann. N.Y. Acad. Sci.* **1978**, *308*, 411–425.
46. Rahman, A.; Kessler, A.; More, N.; Sikic, B.; Rowden, G.; Woolley, P.; Schein, P.S. Liposomal Protection of Adriamycin-Induced Cardiotoxicity in Mice. *Cancer Research* **1980**, *40*, 1532–1537.
47. Hawley, A.E.; Davis, S.S.; Illum, L. Targeting of Colloids to Lymph Nodes: Influence of Lymphatic Physiology and Colloidal Characteristics. *Adv. Drug Delivery Rev.* **1995**, *17*, 129–148.
48. Patel, H.M.; Boodle, K.M.; Vaughan-Jones, R. Assessment of the Potential Uses of Liposomes for Lymphoscintigraphy and Lymphatic Drug Delivery. *Biochim. Biophys. Acta* **1984**, *801*, 76–86.

about the first edition . . .

"The depth and breadth of coverage...make it an excellent source for academic and special libraries supporting research programs in the health sciences, pharmacology, and chemical engineering."
—*American Reference Books Annual*

about the second edition . . .

Building on the success of the best-selling first edition, the *Encyclopedia of Pharmaceutical Technology, Second Edition* focuses on the discovery, development, design, manufacture, testing, regulation, and commercialization of drugs and dosage forms—emphasizing pharmaceuticals, pharmacokinetics, analytical chemistry, quality assurance, drug safety, and manufacturing processes.

New topics in the **Second Edition** include advertising and promotion of prescription and over-the-counter drug products...drug delivery via liquid inhalation...mixing and segregation in tumbling blenders...outsourcing of pharmaceuticals...paperless documentation systems...scale-up and postapproval changes...x-ray powder diffractometry...and more.

about the editors . . .

JAMES SWARBRICK is President of PharmaceuTech, Inc., Pinehurst, North Carolina, as well as Vice President for Scientific Affairs at aaiPharma, Inc., Wilmington, North Carolina. Prior to joining aaiPharma in 1993, he was Professor and Chairman of the Division of Pharmaceutics at the University of North Carolina at Chapel Hill School of Pharmacy. The coeditor of the first edition of the *Encyclopedia of Pharmaceutical Technology, Vols. 1-17 (with supplements)*, he is a Fellow of the American Association of Pharmaceutical Scientists (AAPS), the Academy of Pharmaceutical Sciences, the Royal Society of Chemistry, and the Royal Pharmaceutical Society of Great Britain. Dr. Swarbrick received the B.Pharm. degree (1960) from London University, England, the Ph.C. diploma (1961) from the Royal Pharmaceutical Society of Great Britain, and the Ph.D. degree (1964) in pharmaceutics and the D.Sc. degree (1972) in surface and physical chemistry from London University, England.

Printed in the United States of America

MARCEL DEKKER, INC. • NEW YORK • BASEL

ISBN 0-8247-2152-7



9 780824 721527